



Πανεπιστήμιο Πειραιώς – Τμήμα Πληροφορικής

Πρόγραμμα Μεταπτυχιακών Σπουδών

«Κυβερνοασφάλεια και Επιστήμη Δεδομένων»

Μεταπτυχιακή Διατριβή

Τίτλος Διατριβής	"Υπολογιστικό Νέφος και Αναλυτική Δεδομένων: Αξιοποίηση της Πλατφόρμας Oracle Cloud στην Επιχειρηματικότητα και Προβλεπτική Ανάλυση Δεδομένων" "Cloud Computing and Data Analytics: Leveraging the Oracle Cloud Platform in Business and Predictive Data Analytics"
Όνοματεπώνυμο Φοιτητή	Κατσικάνης Ευάγγελος
Πατρώνυμο	Αθανάσιος
Αριθμός Μητρώου	ΜΠΚΕΔ21022
Επιβλέπων	Χρήστος Δουληγέρης, Καθηγητής

Ημερομηνία Παράδοσης **Ιούλιος 2023**

Τριμελής Εξεταστική Επιτροπή

(υπογραφή)

(υπογραφή)

(υπογραφή)

Δουληγέρης Χρήστος
Καθηγητής

Κοτζανικολάου Παναγιώτης
Αναπληρωτής Καθηγητής

Ψαράκης Μιχαήλ
Αναπληρωτής Καθηγητής

Περίληψη

Η παρούσα διπλωματική εργασία εστιάζει στον τομέα του Υπολογιστική Νέφους και της Αναλυτικής δεδομένων στο Νέφος, αναλύοντας διάφορες πτυχές και εφαρμογές. Αρχικά, παρουσιάζεται η ιστορία του Υπολογιστική Νέφους και εξετάζονται οι τύποι υπηρεσιών στο σύννεφο, συμπεριλαμβανομένων του Δημοσίου, του Ιδιωτικού και Υβριδικού νέφους. Στη συνέχεια, αναλύονται τα χαρακτηριστικά της Υπολογιστικής Νέφους και παρουσιάζονται τα εργαλεία Αναλυτικής Νέφους και η λειτουργία τους. Εξετάζεται η χρήση του από τις επιχειρήσεις και αναδεικνύονται τα πλεονεκτήματα που προσφέρει αυτή η τεχνολογία. Επιπλέον, παρουσιάζονται οι τάσεις και οι προκλήσεις που αντιμετωπίζει το υπολογιστικό νέφος. Στη συνέχεια, εξετάζεται η Μηχανική Μάθηση και παρουσιάζεται η έννοια της, καθώς και τα οφέλη που προσφέρει στις επιχειρήσεις. Επικεντρωνόμαστε στη σημασία της Μηχανικής Μάθησης στο πλαίσιο του Υπολογιστική Νέφους, αλλά και στους περιορισμούς που μπορεί να αντιμετωπίσει. Στο τελευταίο μέρος της εργασίας, γίνεται ανασκόπηση της πλατφόρμας Oracle Cloud, παρουσιάζοντας τα χαρακτηριστικά της και τα πλεονεκτήματά της. Επίσης, γίνεται σύγκριση της Oracle Cloud Platform με άλλες πλατφόρμες όπως η Amazon Web Services (AWS), το Google Cloud Platform (GCP) και το Microsoft Azure. Τέλος, παρουσιάζεται η δημιουργία ροών δεδομένων και αναλύσεων με το Oracle Αναλυτική Νέφους, εξετάζοντας παραδείγματα προβλέψεων και αλγορίθμους μοντελοποίησης μηχανικής μάθησης.

Abstract

This thesis focuses on the field of Cloud Computing and Cloud Analytics, analyzing various aspects and applications. First, the history of Cloud Computing is presented and the types of cloud services are examined, including Public Cloud, Private Cloud, and Hybrid Cloud. Then, the characteristics of Cloud Computing are analyzed and the Cloud Analytics tools and their operation are presented.

The use of Cloud Analytics by businesses is examined and the advantages offered by this technology are highlighted. In addition, the trends and challenges facing Cloud Computing are presented. Then, Machine Learning is examined and its concept is presented, as well as the benefits it offers to businesses. We focus on the importance of Machine Learning in the context of Cloud Computing, but also the limitations it can face. In the last part of the paper, the Oracle Cloud platform is reviewed, presenting its features and advantages. It also compares Oracle Cloud Platform with other platforms such as Amazon Web Services (AWS), Google Cloud Platform (GCP) and Microsoft Azure. Finally, creating data streams and analytics with Oracle Cloud Analytics is presented, looking at examples of predictions and machine learning modeling algorithms.

Περιεχόμενα

Περίληψη	3
Abstract	4
1. Εισαγωγή	7
2. Υπολογιστική Νέφος	10
2.1. Ιστορία του Υπολογιστική Νέφος	10
2.2. Τύποι σύννεφων	11
2.3. Χαρακτηριστικά της Υπολογιστικής Νέφος	12
2.4. Οφέλη μετακίνησης στο σύννεφο	13
3. Εργαλεία Αναλυτικής Νέφος	14
3.1. Λειτουργία Αναλυτικής Νέφος	14
3.2. Πώς οι επιχειρήσεις χρησιμοποιούν την Αναλυτική Νέφος	14
3.3. Πλεονεκτήματα Αναλυτικής Νέφος	15
3.4. Δυνατότητες των Εργαλείων Αναλυτικής Νέφος	16
3.5. Τάσεις και προκλήσεις του Υπολογιστικού Νέφος	16
3.6. Αναλυτική Δεδομένων ως Υπηρεσία (DAaaS) - Εφαρμογές	18
3.7. Αναλυτική Δεδομένων ως Υπηρεσία (DAaaS) - Περιορισμοί	18
3.8. Προκλήσεις του υπολογιστικού νέφους	19
4. Μηχανική Μάθηση	21
4.1. Τι είναι η Μηχανική Μάθηση;	21
4.2. Οφέλη της Μηχανικής Μάθησης	21
4.3. Επίλυση προβλημάτων με την Μηχανική Μάθηση	22
4.4. Οφέλη της Μηχανικής Μάθησης στο Cloud	23
4.5. Περιορισμοί της Μηχανικής Μάθησης στο Υπολογιστικό Νέφος	23

5. Παρουσίαση της πλατφόρμας Oracle Cloud.....	24
5.1. Χαρακτηριστικά της πλατφόρμας Oracle Cloud	25
5.2. Πλεονεκτήματα της πλατφόρμας Oracle Cloud	25
5.3. Σύγκριση Oracle με άλλες πλατφόρμες:.....	26
5.3.1. Το Oracle Cloud Infrastructure σε σύγκριση με τις Υπηρεσίες Ιστού της.....	26
Amazon.....	26
5.3.2 Το Oracle Cloud Infrastructure σε σύγκριση με το Google Cloud Platform .	28
5.3.2. Το Oracle Cloud Infrastructure σε σύγκριση με το Microsoft Azure	29
6. Σενάριο Ανάλυσης δεδομένων με την πλατφόρμα Oracle Cloud	32
6.1. Δημιουργία ροών δεδομένων.....	32
6.2. Δημιουργία συνόλου δεδομένων βάσει αρχείου	34
6.3. Διαγράμματα	35
6.4. Διάγραμμα Ροής Δεδομένων (Data-FlowDiagram)	39
6.5. Δημιουργία Data Flow.....	40
6.6. Ακολουθίες (Sequences).....	50
6.7. Αναλύσεις στα σύνολα δεδομένων.....	52
7.1. Τι είναι τα προγνωστικά μοντέλα Oracle Analytics;	55
7.2. Αλγόριθμοι μοντελοποίησης μηχανικής μάθησης.....	56
8. Συμπεράσματα	78
Βιβλιογραφία	79

Κεφάλαιο 1:

1. Εισαγωγή

Είναι γεγονός πως η πρόοδος της τεχνολογίας είναι αλματώδης στην εποχή μας. Η Πληροφορική αν και νεότερη επιστήμη κατάφερε σε ελάχιστο χρόνο να αλλάξει τα δεδομένα της οικονομίας αλλά και γενικότερα της ευρύτερης κοινωνίας σε παγκόσμιο επίπεδο. Ο τομέας της Πληροφορικής γνώρισε τεράστια ανάπτυξη τον 21^ο αιώνα ο οποίος δικαίως χαρακτηρίζεται “ο αιώνας της πληροφορίας”. Είναι πραγματικότητα πως ζούμε στην εποχή των δεδομένων. Τα δεδομένα, μας κατακλύζουν παντού χάρη στις ραγδαίες τεχνολογικές εξελίξεις που έχουν σημειωθεί τα τελευταία χρόνια. Ο ρυθμός ψηφιοποίησης έχει αυξηθεί σημαντικά και τώρα δικαίως μιλάμε για «κοινωνίες ψηφιακής πληροφορίας». Αν πριν από 20 ή 30 χρόνια μόνο το 1% των πληροφοριών που παράγονταν ήταν ψηφιακές, τώρα πάνω από το 94% αυτών των πληροφοριών είναι ψηφιακές και προέρχονται από διάφορες πηγές όπως τα κινητά μας τηλέφωνα, διακομιστές, συσκευές αισθητήρων στο Διαδίκτυο των πραγμάτων (IoT), κοινωνικά δίκτυα κ.λπ. [1]. Το έτος 2002 θεωρείται η «αρχή της ψηφιακής εποχής» όπου παρατηρήθηκε μια έκρηξη ψηφιακά παραγόμενου εξοπλισμού και πληροφοριών.

Ο αριθμός και ο όγκος των πληροφοριών που συλλέγονται έχουν αυξηθεί σημαντικά λόγω της αύξησης των συσκευών που συλλέγουν αυτές τις πληροφορίες, όπως κινητές συσκευές, φθηνές και πολυάριθμες συσκευές αισθητήρων στο Διαδίκτυο των πραγμάτων (IoT), τηλεπισκόπηση, αρχεία καταγραφής λογισμικού, κάμερες, μικρόφωνα, RFID αναγνώστες, ασύρματα δίκτυα αισθητήρων κ.λπ. [2]. Σύμφωνα με στατιστικά στοιχεία, ο όγκος των δεδομένων που παράγονται ανά ημέρα είναι περίπου 44 zettabyte (44× 1021 byte). Κάθε δευτερόλεπτο, παράγονται 1,7 MB δεδομένων ανά άτομο [3]. Με βάση τις προβλέψεις του International DataGroup, ο παγκόσμιος όγκος δεδομένων θα αυξηθεί εκθετικά από το 2020 έως το 2025, με μια κίνηση από 44 σε 163 zettabyte [4]. Το Σχήμα 1 δείχνει την ποσότητα των παγκόσμιων δεδομένων που δημιουργούνται, αντιγράφονται και καταναλώνονται.

Για να πάρουμε μια γεύση του όγκου των δεδομένων που παράγονται σε καθημερινή βάση, ας δούμε ένα τμήμα δεδομένων που παράγουν διαφορετικές πλατφόρμες. Στο Διαδίκτυο, υπάρχουν τόσες πολλές πληροφορίες στα χέρια μας. Είναι αλήθεια ότι κάθε φορά που αναζητούμε απαντήσεις από τις μηχανές αναζήτησής μας προσθέτουμε πληροφορίες.

Ως αποτέλεσμα, η Google παράγει πλέον περισσότερες από 500.000 αναζητήσεις κάθε δευτερόλεπτο (περίπου 3,5 δισεκατομμύρια αναζητήσεις την ημέρα) [5]. Τα μέσα κοινωνικής δικτύωσης από την άλλη πλευρά είναι ένας τεράστιος παραγωγός δεδομένων.

Ο «έρωτας» των ανθρώπων με τα μέσα κοινωνικής δικτύωσης σίγουρα τροφοδοτεί τη δημιουργία δεδομένων. Κάθε λεπτό, οι χρήστες του Snapchat μοιράζονται 527.760 φωτογραφίες, περισσότεροι από 120 επαγγελματίες συμμετέχουν στο LinkedIn, οι χρήστες παρακολουθούν 4.146.6000 βίντεο στο Youtube, 456.000 αποστέλλονται στο Twitter και οι χρήστες του Instagram δημοσιεύουν 46.740 φωτογραφίες [5]. Το Facebook παραμένει η μεγαλύτερη πλατφόρμα κοινωνικής δικτύωσης, με πάνω από 300 εκατομμύρια φωτογραφίες να ανεβαίνουν κάθε μέρα με περισσότερα από 510.000 σχόλια και 293.000 καταστάσεις ενημερώνονται κάθε λεπτό. Με την αύξηση του αριθμού και της ποσότητας των δεδομένων, υπήρξαν πλεονεκτήματα αλλά και προκλήσεις καθώς τα συστήματα διαχείρισης σχεσιακών βάσεων δεδομένων και άλλα παραδοσιακά συστήματα έχουν δυσκολίες στην επεξεργασία και ανάλυση αυτής της ποσότητας. Για το λόγο αυτό, ο όρος «μεγάλα δεδομένα» προέκυψε όχι μόνο για να περιγράψει την ποσότητα των δεδομένων αλλά και την ανάγκη για νέες τεχνολογίες και τρόπους επεξεργασίας και ανάλυσης αυτών των δεδομένων. Το Υπολογιστική Νέφος έχει διευκολύνει την αποθήκευση, την επεξεργασία και την ανάλυση δεδομένων.

Χρησιμοποιώντας το Cloud έχουμε πρόσβαση σε σχεδόν απεριόριστο χώρο αποθήκευσης και υπολογιστική ισχύ που παρέχεται από διαφορετικούς προμηθευτές. Τα μοντέλα παράδοσης cloud, όπως τα: IaaS (Υποδομή ως υπηρεσία), PaaS (Πλατφόρμα ως υπηρεσία) μπορούν να βοηθήσουν τους οργανισμούς σε διαφορετικούς τομείς να χειρίζονται τα Μεγάλα Δεδομένα ευκολότερα και



Εικόνα 1: Όγκος δεδομένων/πληροφοριών που δημιουργήθηκαν, καταγράφηκαν, αντιγράφηκαν και καταναλώθηκαν παγκοσμίως από το 2010 έως το 2024 (εκτιμώμενο) [3]

Όπως φαίνεται, τα έτη 2010–2015, ο ρυθμός αύξησης από έτος σε έτος ήταν μικρότερος, ενώ από το 2018 ο ρυθμός αυτός αυξήθηκε σημαντικά, καθιστώντας έτσι την τάση εκθετική [3].

Η παρούσα διπλωματική διατριβή αναλύει το συναρπαστικό πεδίο της Υπολογιστικής Νέφους και της Αναλυτικής Νέφους, αναδεικνύοντας την επαναστατική επίδρασή τους στον τρόπο λειτουργίας των επιχειρήσεων και στην ανάπτυξη της τεχνολογίας. Ξεκινώντας από την ιστορία του υπολογιστικού νέφους και τους διάφορους τύπους συννεφιών, εξερευνούμε τα δημόσια, ιδιωτικά και υβριδικά νέφη, αναδεικνύοντας τα χαρακτηριστικά και τα οφέλη που προσφέρουν στις επιχειρήσεις.

Ένα από τα βασικά οφέλη της μετακίνησης στο σύννεφο είναι η ανάλυση δεδομένων, η οποία παίζει ζωτικό ρόλο στη λήψη αποφάσεων και την ανάπτυξη στρατηγικών για τις επιχειρήσεις. Επιστημαίνουμε τη λειτουργία και τη χρήση εργαλείων αναλυτικής νέφους, καθώς και πώς οι επιχειρήσεις εφαρμόζουν την αναλυτική νέφους για να εξάγουν πολύτιμα εισαγωγές και να βελτιώσουν τις αποτελεσματικές τους διαδικασίες.

Επιπλέον, εστιάζουμε στην Μηχανική Μάθηση και τον ρόλο της στο υπολογιστικό νέφος. Μελετούμε την ουσία της Μηχανικής Μάθησης, τα οφέλη της και τον τρόπο με τον οποίο συμβάλλει στην επίλυση προβλημάτων. Εξετάζουμε επίσης τους περιορισμούς που αντιμετωπίζει η Μηχανική Μάθηση στο υπολογιστικό νέφος και πώς μπορούν να αντιμετωπιστούν.

Το αναλυτικό μας ταξίδι καταλήγει σε μια συζήτηση για τις τάσεις και τις προκλήσεις που συναντά το υπολογιστικό νέφος. Εξετάζουμε ειδικότερα την υπηρεσία ανάλυσης δεδομένων ως υπηρεσία (DAaaS) και τους περιορισμούς που την συνοδεύουν. Τέλος, αναδεικνύουμε τις προκλήσεις που αντιμετωπίζει το υπολογιστικό νέφος και πώς οι επιχειρήσεις μπορούν να τις αντιμετωπίσουν με τη χρήση της πλατφόρμας Oracle Cloud.

Μέσα από αυτήν την πολυποικίλη ανάλυση, η διπλωματική διατριβή αναδεικνύει τον συναρπαστικό κόσμο της Υπολογιστικής Νέφους και της Αναλυτικής Νέφους, προσφέροντας νέες ευρήματα και εισαγωγές για την αξιοποίηση αυτών των τεχνολογιών στον επιχειρηματικό κόσμο.

Συνεχίζοντας την ανάλυση, παρουσιάζουμε την πλατφόρμα Oracle Cloud και αναδεικνύουμε τα χαρακτηριστικά της που την καθιστούν μια ισχυρή επιλογή για τις επιχειρήσεις. Εξετάζουμε τα

πλεονεκτήματα της πλατφόρμας Oracle Cloud και τονίζουμε την αξία που προσφέρει στον τομέα της ανάλυσης δεδομένων και της πρόβλεψης.

Στη συνέχεια, πραγματοποιούμε μια σύγκριση της πλατφόρμας Oracle με άλλες δημοφιλείς πλατφόρμες - ανταγωνιστές. Αναλύουμε το Oracle Cloud Infrastructure και το συγκρίνουμε με τις υπηρεσίες ιστού της Amazon, αναδεικνύοντας τις διαφορές και τα πλεονεκτήματα που προσφέρει. Στη συνέχεια, συγκρίνουμε το Oracle Cloud Infrastructure με το Google Cloud Platform, επισημαίνοντας τις ξεχωριστές χαρακτηριστικές τους και τις επιπτώσεις που έχουν στην επιχειρηματική απόδοση. Τέλος, αναλύουμε το Oracle Cloud Infrastructure σε σύγκριση με το Microsoft Azure, επισημαίνοντας τα διαφορετικά στοιχεία και τις επιλογές που προσφέρουν στις επιχειρήσεις.

Συνεχίζοντας την ανάλυση, στο πλαίσιο της διπλωματικής διατριβής παρουσιάζονται δύο σενάρια ανάλυσης δεδομένων που χρησιμοποιούν την πλατφόρμα Oracle Cloud.

Το πρώτο σενάριο επικεντρώνεται στην ανάλυση δεδομένων με τη χρήση της πλατφόρμας Oracle Cloud. Σε αυτό το σενάριο, παρουσιάζεται η δημιουργία ενός συνόλου δεδομένων βασισμένο σε ένα αρχείο. Στη συνέχεια, πραγματοποιείται απλή ανάλυση των δεδομένων για την εξαγωγή πρώτων εισαγωγών και συμπερασμάτων. Έπειτα, δημιουργείται μια πρώτη ροή δεδομένων και αναλύονται τα αποτελέσματα που προκύπτουν από αυτήν. Στη συνέχεια, παρουσιάζεται η δημιουργία μιας δεύτερης ροής δεδομένων με διακλαδιζόμενο κόμβο, προσφέροντας μια επιπλέον πτυχή ανάλυσης. Στο επόμενο στάδιο, παρουσιάζεται η δημιουργία και εκτέλεση μιας ακολουθίας ροών δεδομένων, καθώς και η ανάλυση των παραγόμενων συνόλων δεδομένων.

Το δεύτερο σενάριο εστιάζει στην ανάλυση δεδομένων με χρήση προγνωστικών μοντέλων. Χρησιμοποιώντας τα ίδια σύνολα δεδομένων, το σενάριο αυτό επικεντρώνεται στη δημιουργία ενός συνόλου δεδομένων πρόβλεψης. Αρχικά, πραγματοποιείται η επεξεργασία και προετοιμασία των δεδομένων. Έπειτα, δημιουργείται ένα τυχαίο δείγμα από το σύνολο δεδομένων που χρησιμοποιείται για την εκπαίδευση ενός προγνωστικού μοντέλου Elastic Net Regression. Στη συνέχεια, πραγματοποιείται η εκπαίδευση του μοντέλου και επανεξετάζεται η ποιότητα του μοντέλου και των συνόλων δεδομένων. Ακολουθεί μια διαδικασία επανάληψης και βελτιστοποίησης της εκπαίδευσης του μοντέλου. Στη συνέχεια, συγκρίνονται τα σενάρια προγνωστικού μοντέλου και των διάφορων τύπων μοντέλων, με έμφαση στην ταξινόμηση του συντελεστή προσαρμογής των μοντέλων παλινδρόμησης. Τέλος, το εκπαιδευμένο μοντέλο χρησιμοποιείται για την πρόβλεψη ενός συνόλου δεδομένων, παρέχοντας μια εφαρμογή πρόβλεψης.

Κεφάλαιο 2:

2. Υπολογιστική Νέφος

Η Υπολογιστική Νέφος (Cloud Computing) είναι μια τεχνολογία πληροφορικής που επιτρέπει στους χρήστες να αποκτούν πρόσβαση σε υπηρεσίες και πόρους υπολογιστικής ισχύος, όπως εφαρμογές, αποθηκευτικό χώρο και υπολογιστική ισχύς μέσω του Διαδικτύου. Οι χρήστες αντί να αγοράζουν και να διατηρούν δικούς τους υπολογιστές και το λογισμικό τους, μπορούν να ενοικιάζουν ή να αγοράζουν την πρόσβαση σε αυτούς τους πόρους από έναν πάροχο Υπολογιστικής Νέφος. Οι υπηρεσίες Υπολογιστικής Νέφος είναι συνήθως διαθέσιμες σε μορφή "πακέτου", όπου οι χρήστες μπορούν να επιλέξουν το επίπεδο της υπηρεσίας που χρειάζονται ανάλογα με τις ανάγκες τους. Οι χρήστες μπορούν να επιλέξουν τον τύπο του Νέφος (Δημόσιο, Ιδιωτικό ή Υβριδικό), το μέγεθος και το επίπεδο της υπηρεσίας που χρειάζονται, καθώς και τον τρόπο πληρωμής. Οι πόροι της Υπολογιστικής Νέφος είναι διαθέσιμοι μέσω διακομιστών που τρέχουν ειδικό λογισμικό που διαχειρίζεται τους πόρους και τις υπηρεσίες.

Η Υπολογιστική Νέφος έχει φέρει επανάσταση στον τρόπο με τον οποίο αφαιρείται και χρησιμοποιείται η υπολογιστική υποδομή. Τα πολλά πλεονεκτήματα της Υπολογιστικής Νέφος, όπως η ελαστικότητα, το pay-as-you-go ή το pay-per-use μοντέλο και η χαμηλή αρχική επένδυση, την έχουν καταστήσει βιώσιμη και επιθυμητή επιλογή για αποθήκευση μεγάλων δεδομένων, τη διαχείριση και ανάλυση τους [7]. Επειδή τα μεγάλα δεδομένα θεωρούνται πλέον ζωτικής σημασίας για πολλούς οργανισμούς και τομείς, οι πάροχοι υπηρεσιών όπως η Amazon, η Google και η Microsoft προσφέρουν τα δικά τους συστήματα μεγάλων δεδομένων με οικονομικά αποδοτικό τρόπο. Τα συστήματα αυτά προσφέρουν επεκτασιμότητα για επιχειρήσεις όλων των μεγεθών. Αυτό οδήγησε στην προβολή του όρου Analytics ως υπηρεσία (AaaS) ως ένας ταχύτερος και αποτελεσματικότερος τρόπος για την ενοποίηση, τον μετασχηματισμό και την οπτικοποίηση διαφορετικών τύπων δεδομένων.

2.1. Ιστορία του Υπολογιστικής Νέφος

Η έννοια του Υπολογιστικής Νέφος αναδείχθηκε αρχικά τη δεκαετία του 1950, όταν εκπαιδευτικά ιδρύματα και εταιρείες χρησιμοποιούσαν κεντρικούς υπολογιστές μεγάλων υπολογιστικών και αποθηκευτικών δυνατοτήτων. Οι χρήστες είχαν πρόσβαση στους υπολογιστές μέσω τερματικών με χαμηλή υπολογιστική ισχύ και μικρές αποθηκευτικές ικανότητες. Τον 1970, η IBM και η Google συνεργάστηκαν για την ανάπτυξη της τεχνολογίας του Υπολογιστικής Νέφος. Αρχικά, η IBM παρουσίασε το λειτουργικό σύστημα εικονικών μηχανών (VMS operating system) που επέτρεπε τη λειτουργία πολλών εικονικών μηχανών στο ίδιο μηχάνημα, με κάθε μηχανή να είναι μία αυτόνομη οντότητα με δικό της λειτουργικό σύστημα και υπολογιστικούς πόρους. Κατά τη δεκαετία του '80, με την εμφάνιση των προσωπικών υπολογιστών, η πρώτη τακτική χρήση συνοδευόταν από την υπόσχεση ότι οι χρήστες θα μπορούσαν να διαχειρίζονται το υπολογιστικό τους περιβάλλον κατά τη δική τους επιλογή. Στη συνέχεια, με την εμφάνιση του Διαδικτύου και του Παγκόσμιου Ιστού, το Υπολογιστικό Νέφος γίνεται ευρέως γνωστό και διαδεδομένο. Τη δεκαετία του '90 εμφανίζονται τα Εικονικά Ιδιωτικά Δίκτυα, ενώ μέχρι εκείνη τη στιγμή, οι εταιρείες τηλεπικοινωνιών είχαν υποστηρίξει τα κυκλώματα δεδομένων σημείο προς σημείο. Το 2000, κολοσσοί της Πληροφορικής, όπως η Amazon, η Microsoft και η Google, είχαν αρχίσει να αναπτύσσουν και να παρέχουν υπηρεσίες Υπολογιστικού Νέφος. Ακολούθησαν συμβάντα που θεωρούνται ευρέως ως ορόσημα στην ιστορία του Υπολογιστικού Νέφος.

Το 2006, η Amazon παρουσίασε την υπηρεσία Ελαστικό Υπολογιστικό Νέφος (Elastic Compute Cloud - EC2), μια εμπορική υπηρεσία που βασιζόταν στον Παγκόσμιο Ιστό και παρείχε στους χρήστες τη δυνατότητα να αναπτύσσουν εφαρμογές σε ενοικιαζόμενα μηχανήματα. Το 2008, κυκλοφόρησε στην αγορά το Eucalyptus, η πρώτη πλατφόρμα ανοικτού κώδικα για την ανάπτυξη ιδιωτικών νεφών (private clouds). Την ίδια χρονιά, η Google κυκλοφόρησε το Google App Engine, μια πλατφόρμα που υποστήριζε διάφορες υπηρεσίες του Υπολογιστικού Νέφος. Το 2011, η IBM

παρουσίασε το IBMSmartCloud και το 2012 κυκλοφόρησε το OracleCloud, συνεχίζοντας τις εξελίξεις στον χώρο του υπολογιστικού νέφους. Τέλος, η Microsoft ανακοίνωσε επίσης ότι θα εισαγάγει το Υπολογιστική Νέφος στην επόμενη μεγάλη ενημέρωση των λύσεων DynamicsERP και θα λειτουργούν μέσω της πλατφόρμας WindowsAzure.

Η έννοια της Αναλυτικής Νέφους αναφέρεται στη διαχείριση και ανάλυση δεδομένων που λαμβάνουν χώρα στο cloud αντί για τοπικές συσκευές, σε ένα επιχειρηματικό περιβάλλον. Τα συστήματα αυτά, που φιλοξενούνται στο σύννεφο, παρέχουν στους χρήστες τη δυνατότητα να έχουν πρόσβαση, να συλλέγουν, να αναλύουν και να χρησιμοποιούν δεδομένα. Με αυτές τις πλατφόρμες, οι χρήστες μπορούν να διαχειρίζονται μεγάλα σύνολα δεδομένων, να εντοπίζουν τάσεις και να ανιχνεύουν περιοχές για βελτίωση σε ολόκληρο τον οργανισμό.

Αν και το πρόγραμμα υπολογιστικών φύλλων του ενός φορητού υπολογιστή ή οι λεπτομέρειες εσωτερικής εγκατάστασης μπορεί να είναι επαρκή για μια λίστα μερικών χιλιάδων σημείων δεδομένων, η ανάλυση πιο σύνθετων συνόλων δεδομένων που αποτελούνται από δεκάδες χιλιάδες έως και εκατομμύρια εισόδους, απαιτεί πιο εξελιγμένες λύσεις. Η ανάλυση στο νέφος επιτρέπει στις εταιρείες να επεξεργάζονται μεγάλα σύνολα δεδομένων με έναν επεκτάσιμο και οικονομικό τρόπο, χωρίς να χρειάζεται να δημιουργούν υποδομές επί τόπου. Αυτός είναι ένας από τους λόγους που συνιστάται η χρήση της δύναμης του cloud για την ανάλυση μεγάλων συνόλων δεδομένων. [10]

2.2. Τύποι σύννεφων

Αν και συνήθως αναφέρουμε το νέφος ως μια άυλη οντότητα, στην πραγματικότητα αναφέρεται σε μεγάλα δίκτυα υπολογιστών που φιλοξενούνται σε ένα ή περισσότερα κέντρα δεδομένων. Ανάλογα με τις ανάγκες και τους στόχους ασφάλειας, απόδοσης και πρόσβασης, μεταξύ άλλων, ένας από τους παρακάτω τύπους μοντέλων νέφους μπορεί να έχει νόημα για την πλατφόρμα αναλυτικών στοιχείων.

2.2.1. Δημόσιο Νέφος (Public Cloud)

Το δημόσιο νέφος είναι μια πλατφόρμα που παρέχεται από τρίτους παρόχους μέσω του δημόσιου διαδικτύου, δίνοντας στους χρήστες τη δυνατότητα να χρησιμοποιήσουν ή να αγοράσουν πόρους. Το δημόσιο cloud χρησιμοποιεί ένα τυπικό μοντέλο υπολογιστικού νέφους για να παρέχει πόρους σε απομακρυσμένους χρήστες, συμπεριλαμβανομένων εικονικών μηχανών, εφαρμογών και αποθήκευσης. Οι περισσότεροι χρήστες δημόσιου νέφους χρησιμοποιούν ιστότοπους, εφαρμογές λογισμικού ως υπηρεσία (SaaS) και άλλες υπηρεσίες που αφορούν τη διαχείριση ηλεκτρονικού ταχυδρομείου, ημερολογίων, διαχείρισης έργων και τραπεζικών λογαριασμών. Οι παρόχοι δημόσιου νέφους παρέχουν συνήθως ευέλικτες λύσεις σε ορισμένο κόστος για τους χρήστες τους, που μπορούν να επιλέξουν τους πόρους που χρειάζονται ανάλογα με τις ανάγκες τους.

2.2.2. Ιδιωτικό Νέφος (Private Cloud)

Ένα ιδιωτικό νέφος αναφέρεται σε ένα υπολογιστικό σύστημα που είναι προσβάσιμο μόνο μέσα στο εσωτερικό δίκτυο ενός οργανισμού και δεν κοινοποιείται με άλλους οργανισμούς. Λειτουργεί παρόμοια με το εταιρικό intranet και άλλους πόρους που είναι προσβάσιμοι μόνο εντός του τοπικού δικτύου ή μέσω ενός εικονικού ιδιωτικού δικτύου (VPN). Όλα τα νέφη γίνονται ιδιωτικά όταν η υποκείμενη υποδομή πληροφορικής αφορά μόνο έναν πελάτη, παρέχοντας πλήρη απομόνωση. Τα ιδιωτικά νέφη παρέχουν επιπλέον χωρητικότητα υποδομής για τη διαχείριση μεγάλων όγκων αποθηκευτικών δεδομένων, υπηρεσίες κατ' απαίτηση και επιπλέον ορατότητα σε πόρους σε ολόκληρη την υποδομή.

Οι ίδιοι πόροι είναι διαθέσιμοι με πρόσθετη ασφάλεια, αλλά μπορεί να υπάρχει υψηλότερο κόστος για την ρύθμιση και τη συντήρηση των δικτύων και των εφαρμογών σε σύγκριση με τις υπηρεσίες που παρέχονται από ένα δημόσιο νέφος.

2.2.3. Υβριδικό Νέφος (Hybrid Cloud)

Το υβριδικό cloud είναι μια επιλογή που συνδυάζει ένα ιδιωτικό cloud με μία ή περισσότερες δημόσιες υπηρεσίες cloud, με σκοπό να παρέχει μια ευέλικτη και αξιόπιστη λύση για τις ανάγκες πληροφορικής μιας επιχείρησης. Αυτή η στρατηγική χρησιμοποιεί ιδιόκτητο λογισμικό για να εξασφαλίσει την αποτελεσματική επικοινωνία μεταξύ των διαφορετικών υπηρεσιών και μπορεί να συνδέσει πολλούς υπολογιστές μέσω ενός δικτύου, να ενοποιήσει πόρους πληροφορικής, να μετακινήσει φόρτους εργασίας σε διαφορετικά περιβάλλοντα και να ενσωματώσει ένα ενιαίο, ενοποιημένο εργαλείο διαχείρισης.

2.3. Χαρακτηριστικά της Υπολογιστικής Νέφους

Η Υπολογιστική Νέφος είναι ένας σχετικά πρόσφατος όρος και η τεχνολογία αυτή έχει προκύψει από τη χρήση και τις τάσεις των υπολογιστικών δικτύων και των σχετικών υπηρεσιών και μοντέλων επιχειρηματικότητας. Σε μελέτες [22]-[25], παρουσιάζονται αρκετά χαρακτηριστικά της Υπολογιστικής Νέφους. Καθώς η τεχνολογία εξελίσσεται, μπορεί να προκύψουν και άλλα χαρακτηριστικά. Ορισμένα από τα κυριότερα χαρακτηριστικά της Υπολογιστικής Νέφους σε αυτήν την ενότητα, τα οποία καθορίζουν και υποστηρίζουν τον πυρήνα της τεχνολογίας και τις προδιαγραφές αποδοχής της.

1. **Δυναμικότητα (Ευελιξία):** οι πλατφόρμες πόρων του cloud είναι δυναμικά επεκτάσιμες, πράγμα που σημαίνει ότι μπορούν να μεγαλώσουν ή να μειωθούν σε μέγεθος βάσει της ζήτησης των χρηστών, μειώνοντας έτσι τον επενδυτικό κίνδυνο για τον χρήστη και ικανοποιώντας τις απαιτήσεις πολλών χρηστών. Η Υπολογιστική Νέφος παρέχει στους χρήστες την αίσθηση ότι ατελείωτοι υπολογιστικοί πόροι είναι διαθέσιμοι για αυτούς.
2. **Εικονικοποίηση:** οι εφαρμογές και οι πλατφόρμες της Υπολογιστικής Νέφους δημιουργούνται βάσει της έννοιας της εικονικοποίησης των πόρων. Η εικονικοποίηση εκτελεί ένα σημαντικό ρόλο στην αποτελεσματικότητα της απόδοσης των πόρων και την αύξηση του επιπέδου ασφάλειας και αξιοπιστίας των υπηρεσιών.
3. **Οικονομία κλίμακας:** η Υπολογιστική Νέφος κυριαρχείται από μεγάλες εταιρείες, όπως η IBM, η Microsoft, η Oracle, η Google και η Amazon, οι οποίες έχουν τη δυνατότητα να αξιοποιήσουν πόρους μεγάλης κλίμακας, επιτρέποντάς τους να ελαχιστοποιήσουν το κόστος μίσθωσης και εκμετάλλευσης. Έτσι, οι εταιρείες – πάροχοι Υπολογιστικής Νέφους μπορούν να ελκύσουν όσους περισσότερους χρήστες είναι δυνατόν. Καθώς υπάρχει ένας μεγάλος αριθμός πιθανών χρηστών που συμμετέχουν σε οποιαδήποτε υπηρεσία βασισμένη στο cloud, γίνεται οικονομικά εφικτό για τον πάροχο υπηρεσίας να προσφέρει την υπηρεσία σε πολύ χαμηλό κόστος.
4. **Υπηρεσία κατόπιν αιτήματος:** οι υπηρεσίες και οι πλατφόρμες της Υπολογιστικής Νέφους αποκτώνται και τιμολογούνται βάσει των πραγματικών αναγκών των χρηστών. Η Υπολογιστική Νέφος εξαλείφει τον κίνδυνο μιας μεγάλης επένδυσης και επιτρέπει στους χρήστες να αποκτήσουν μόνο τους πόρους που χρειάζονται. Κατά συνέπεια, οι υπηρεσίες εξαρτώνται από το κόστος σε βραχυπρόθεσμη βάση (π.χ. με βάση την ώρα), όπου οι χρήστες απελευθερώνουν τους πόρους όταν δεν τους χρειάζονται πλέον.
5. **Δυναμική προσαρμογή:** οι ενοικιαζόμενοι πόροι του νέφους πρέπει να είναι προσαρμόσιμοι σε μεγάλο βαθμό. Για παράδειγμα, στο μοντέλο παροχής IaaS, οι χρήστες έχουν τη δυνατότητα να διαθέσουν εικονικές και εξειδικευμένες συσκευές. Άλλες υπηρεσίες παρέχουν χαμηλή ευελιξία και δεν εφαρμόζονται. Ωστόσο, αναμένεται να παρέχεται σε αυτές τις υπηρεσίες ένα επίπεδο προσαρμογής.
6. **Υψηλή αξιοπιστία:** οι πλατφόρμες της Υπολογιστικής Νέφους πρέπει να εξασφαλίζουν την ασφάλεια των δεδομένων των πελατών, έτσι ώστε η πλατφόρμα εφαρμογών να θεωρείται αξιόπιστη. Γενικά, γίνεται χρήση αντιγράφων ασφαλείας της πλατφόρμας και πολλαπλών δεδομένων για να αυξηθεί η αξιοπιστία της πλατφόρμας. Οι δυναμικές προσεγγίσεις

διαχείρισης δικτύου εφαρμόζονται επίσης από τις πλατφόρμες του Υπολογιστική Νέφους για να επαληθεύσουν την αποτελεσματικότητα και την κατάσταση κάθε κόμβο πόρου. Ο λόγος για αυτό είναι ότι οι κόμβοι μπορούν να μετακινηθούν δυναμικά όταν συναντούν αποτυχία ή χαμηλή αποτελεσματικότητα. Ένας άλλος λόγος είναι να εξασφαλιστεί ότι η απόδοση του συστήματος ως σύνολο δεν επηρεάζεται σε περίπτωση βλάβης.

2.4. Οφέλη μετακίνησης στο σύννεφο

Λόγω της δημοτικότητας του υπολογιστικού νέφους τα τελευταία χρόνια, η τεχνολογία ενσωματώνει μια σειρά από πλεονεκτήματα που διαπιστώνουν το επόμενο επίπεδο κοινής χρήσης δικτύων και εφαρμογών και διανομής υπηρεσιών και πόρων με οικονομικά βιώσιμο και αποδοτικό τρόπο. Ένας αριθμός προηγούμενων μελετών [27]-[32] έχουν επισημάνει ορισμένα από αυτά τα πλεονεκτήματα. Το Υπολογιστική Νέφους επιτρέπει την πρόσβαση στο δίκτυο κατ' απαίτηση και μια σειρά από σχετικές εφαρμογές και υποδομές σε έναν αριθμό προσαρμοσμένων υπολογιστικών πόρων, όπως διακομιστές, εφαρμογές λογισμικού, χώρους αποθήκευσης, υπηρεσίες και άλλα δίκτυα. Σύμφωνα με τα χαρακτηριστικά που προσδιορίστηκαν στην προηγούμενη ενότητα, αυτή η ενότητα επισημαίνει τα κύρια οφέλη και προκλήσεις της μετάβασης στο cloud, τα οποία περιλαμβάνουν κυρίως:

α) Βέλτιστη χρήση πόρων: Δεδομένου ότι το μεγαλύτερο μέρος του υπολογιστικού νέφους βασίζεται στο μοντέλο χρήσης ανά πληρωμή, επομένως οι πόροι απελευθερώνονται μετά από κάθε χρήση. Αυτό έχει ως αποτέλεσμα τη συνολική χρήση όλων των υπολογιστικών πόρων με αποτελεσματικό και βελτιστοποιημένο τρόπο, οδηγώντας σε πράσινους υπολογισμούς.

β) Ενοικίαση κατ' απαίτηση: Ένα άλλο σημαντικό πλεονέκτημα του υπολογιστικού νέφους είναι η διαθεσιμότητα όλων των τύπων πόρων και υπηρεσιών υπολογιστών και υποδομής για όλους, οπουδήποτε και ανά πάσα στιγμή. Το Industry 4.0 [33], [34] παρέχει ευκαιρίες για ανάπτυξη και βιωσιμότητα για όλα τα είδη επιχειρήσεων, είτε μεγάλες είτε μικρές. Οι μελλοντικές επιχειρήσεις βρίσκονται στην καινοτομία που είναι το κλειδί της επιτυχίας για τις ευέλικτες εταιρείες του σήμερα. Οι νεοφυείς επιχειρήσεις μπορούν εύκολα να έχουν πρόσβαση και να χρησιμοποιούν υποδομές και υπολογιστικούς πόρους τελευταίας τεχνολογίας, όπως και όταν χρειάζεται, να ανταγωνιστούν τους γίγαντες του κλάδου, με αποτέλεσμα καλύτερα, φθηνότερα προϊόντα και υπηρεσίες.

γ) Ελαχιστοποιημένο προσωπικό πληροφορικής: Η μετάβαση στο Υπολογιστική Νέφους έχει ως αποτέλεσμα τη μείωση του εσωτερικού προσωπικού πληροφορικής για τη συντήρηση των υπαρχόντων συστημάτων. Ωστόσο, κάποιο τεχνικό προσωπικό εξακολουθεί να απαιτείται να συνεργάζεται με προϋπάρχοντες προμηθευτές, προκειμένου να διαχειρίζονται συγκεκριμένες εφαρμογές που ανατίθενται σε εξωτερικούς συνεργάτες.

δ) Ελαχιστοποιημένη Υποδομή: Η μετεγκατάσταση πόρων στο νέφος ή η πρόσβαση στην πλατφόρμα ως υπηρεσία σημαίνει ότι είναι δυνατή η διατήρηση μιας μικρότερης εσωτερικής υποδομής υλικού.

ε) Διαχειριζόμενα Κόστη: Δεδομένου ότι οι περισσότεροι από τους παρόχους υπηρεσιών υπολογιστικού νέφους αποκτούν εισόδημα με βάση τις οικονομίες κλίμακας και προσπαθούν να περιορίσουν το κόστος, να αυξήσουν τους πελάτες και να διαθέτουν τις πιο πρόσφατες και ενημερωμένες εφαρμογές υλικού και λογισμικού. Επομένως, οι τιμές και η αδειοδότηση μπορούν να ελαχιστοποιηθούν κατά την υιοθέτηση του Υπολογιστική Νέφους. Τα πιο πρόσφατα κόστη βασίζονται σε προκαθορισμένες υπηρεσίες και προκύπτουν από το μοντέλο κόστους που χρησιμοποιεί ο πωλητής.

στ) Βελτιωμένος έλεγχος ασφαλείας: Ο έλεγχος ευπάθειας είναι η ικανότητα παρακολούθησης των δραστηριοτήτων και των αρχείων καταγραφής του συστήματος για την παροχή μεγαλύτερου ελέγχου και την ελαχιστοποίηση του κινδύνου επιθέσεων ανιχνεύοντας και αποτρέποντας την εμφάνισή του πριν συμβεί. Στο οικοσύστημα υπολογιστικού νέφους, ο πάροχος υπηρεσιών παρέχει υπηρεσίες σε πολλούς πελάτες ταυτόχρονα και ως αποτέλεσμα δημιουργούνται μεγάλα έσοδα. Ως εκ τούτου, οι πάροχοι υπηρεσιών διασφαλίζουν ότι όλα τα συστήματα είναι ενημερωμένα με τεχνολογίες αιχμής και ότι δεν υπάρχουν τρωτά σημεία ή απειλές για την ασφάλεια.

3. Εργαλεία Αναλυτικής Νέφους

Υπάρχουν διάφοροι τύποι εργαλείων αναλυτικής νέφους. Πολλά από αυτά είναι εύκολα προσβάσιμα μέσω των προγραμμάτων περιήγησής. Ακολουθούν ορισμένα παραδείγματα μερικών δημοφιλών τύπων εργαλείων ανάλυσης cloud:

Η ανάλυση ιστότοπου (Website Analytics) είναι ένας από τους πιο συνηθισμένους τρόπους ανάλυσης στο νέφος και αφορά τα αναλυτικά στοιχεία επισκεψιμότητας ενός ιστότοπου. Τα εργαλεία ανάλυσης βοηθούν στην κατανόηση της επισκεψιμότητας του ιστότοπου, του ποσοστού μετατροπών, του ποσοστού εγκατάλειψης και πολλών άλλων στοιχείων, προκειμένου να προσαρμοστεί ο ιστότοπος με τρόπο που θα βελτιώσει την εμπειρία του χρήστη, αυξάνοντας τα έσοδα και την κερδοφορία.

Η ανάλυση πωλήσεων (Sales Analytics) είναι μια διαδικασία που βοηθά στη διαχείριση των πελατών, τόσο δυναμικών όσο και υπαρκτών, στην αξιολόγηση των πωλήσεων σε διάφορες γεωγραφικές περιοχές και στην παρακολούθηση της απόδοσης της ομάδας πωλήσεων. Αυτές οι πληροφορίες μπορούν να αποκαλύψουν σημαντικές τάσεις και σημάδια που βοηθούν τις επιχειρήσεις να αναπτύξουν πιο αποτελεσματικές στρατηγικές πωλήσεων.

Η χρηματοοικονομική ανάλυση (Financial Analytics) υπερβαίνει τις απλές οικονομικές καταστάσεις και ασχολείται με την αναγνώριση των τάσεων στα έσοδα και τα έξοδα, καθώς και με την ανάλυση λεπτομερειών στα οικονομικά αποτελέσματα που θα ήταν αδύνατο να ανακαλυφθούν χωρίς τη συνδρομή μιας εξειδικευμένης ομάδας οικονομικών αναλυτών.

Η ανάλυση απόδοσης (Performance Analytics) ασχολείται με την εξέταση των πωλήσεων, της παραγωγής ή άλλων σχετικών δεδομένων, με σκοπό τον εντοπισμό σημείων συμφόρησης, την αναγνώριση πηγών δαπανών και την αναζήτηση ευκαιριών βελτίωσης.

3.1. Λειτουργία Αναλυτικής Νέφους

Η Υπολογιστική Νέφους παρέχει μια εναλλακτική λύση στη δημιουργία και διαχείριση δικών τους κέντρων δεδομένων για πολλές εταιρείες, καθώς αυτά απαιτούν τεράστιες επενδύσεις. Τα κέντρα δεδομένων αναλαμβάνουν την επεξεργασία και την αποθήκευση των πληροφοριών σε απομακρυσμένες τοποθεσίες, εξοικονομώντας την ανάγκη για πολύ μεγάλη υποδομή και εξοπλισμό. Έτσι, οι εταιρείες μπορούν να αποκτήσουν πρόσβαση σε αποτελεσματικούς υπολογιστικούς πόρους μειώνοντας σημαντικά το κόστος και την πολυπλοκότητα.

Οι υπηρεσίες Αναλυτικής Νέφους επιτρέπουν στις εταιρείες να αξιοποιήσουν απομακρυσμένα κέντρα δεδομένων και να επεξεργαστούν μεγάλους όγκους πληροφοριών, παρέχοντας ταυτόχρονα αυξημένη παραγωγικότητα και βελτιωμένα, ακριβέστερα αποτελέσματα στους εργαζόμενους. Αυτή η λύση αποτελεί ένα κέρδος για τις περισσότερες επιχειρήσεις, καθώς καθιστά την επεξεργασία των δεδομένων προσιτή και εύκολα προσβάσιμη.

Τα κύρια πλεονεκτήματα της Αναλυτικής Δεδομένων στο Cloud σε σύγκριση με άλλες λύσεις είναι η επιπλέον η κοινή χρήση και η συνεργασία, η βελτιωμένη ασφάλεια, το χαμηλότερο κόστος και η τεράστια επεκτασιμότητα. Με τις πληροφορίες που αποκτούν από την ανάλυση δεδομένων, οι εταιρείες μπορούν να συγκεντρώσουν μετρήσεις που παρέχουν μια καλύτερη εικόνα για τη συνολική λειτουργία των διαφόρων τομέων, όπως οι λειτουργίες, το μάρκετινγκ, τα οικονομικά και άλλα, επιτρέποντας την καλύτερη λήψη αποφάσεων.

3.2. Πώς οι επιχειρήσεις χρησιμοποιούν την Αναλυτική Νέφους

Όπως υπάρχει ένας σχεδόν άπειρος αριθμός τρόπων για τη λειτουργία μιας επιχείρησης, το ίδιο ισχύει και για τους τρόπους χρήσης των αναλυτικών στοιχείων νέφους. Για παράδειγμα, μια εταιρεία που δραστηριοποιείται στον χώρο της ψηφιακής μουσικής θα έχει διαφορετικές απαιτήσεις ανάλυσης cloud σε σχέση με μια εταιρεία λιανικής. Συνήθως, ο κύριος στόχος των αναλύσεων cloud είναι η βελτίωση της κερδοφορίας, αλλά οι επιχειρήσεις μπορεί να έχουν πιο προσαρμοσμένες ανάγκες, όπως την αύξηση της κερδοφορίας από υπάρχοντες πελάτες ή την αύξηση των διεθνών πωλήσεων.

«"Υπολογιστικό Νέφος και Αναλυτική Δεδομένων: Αξιοποίηση της Πλατφόρμας Oracle Cloud στην Επιχειρηματικότητα και Προβλεπτική Ανάλυση Δεδομένων"»

Οι αναφορές και τα εργαλεία οπτικοποίησης μπορούν να αποκαλύψουν πεδία υψηλής ή χαμηλής απόδοσης μιας εταιρείας, προκειμένου να διαμορφωθούν πιο εξειδικευμένες στρατηγικές. Μικρές και μεγάλες επιχειρήσεις μπορούν να αξιοποιήσουν αναλυτικά στοιχεία νέφους για να εξερευνήσουν δεδομένα πελατών και επιχειρήσεων και να αποκτήσουν βασικές πληροφορίες.

Οι εταιρείες που διαθέτουν μεγάλη ομάδα πωλήσεων μπορούν να επωφεληθούν από την αναλυτική νέφους για να κατανοήσουν καλύτερα τις περιοχές όπου έχουν υψηλή απόδοση σε σύγκριση με άλλες περιοχές. Με τη χρήση αναλυτικών στοιχείων, μπορούν να ανακαλύψουν τις περιοχές όπου κατέχουν τον ηγετικό ρόλο και να επιδιώξουν τη διπλασιασμένη προσέγγιση ή να εστιάσουν στις περιοχές όπου οι ανταγωνιστές αντιμετωπίζουν δυσκολίες στην επίτευξη συμφωνιών. Για παράδειγμα, ένας Influencer μέσω κοινωνικής δικτύωσης μπορεί να ανακαλύψει τις τάσεις στους τύπους περιεχομένου που συγκεντρώνουν τα περισσότερα "likes". Επίσης, ένας επιστήμονας γενετικής θα μπορούσε να αξιοποιήσει την ανάλυση του στο νέφος για να αναπαραστήσει την αλληλουχία των εκατομμυρίων νουκλεοτιδίων που συνθέτουν ένα γονίδιο. Αυτά αποτελούν μόνο μερικά παραδείγματα από μια ατελείωτη λίστα πιθανών περιπτώσεων χρήσης.

3.3. Πλεονεκτήματα Αναλυτικής Νέφους

Τα αναλυτικά στοιχεία του Νέφους προσφέρουν σημαντικά πλεονεκτήματα σε σύγκριση με παραδοσιακές μεθόδους, όπως υπολογιστικά φύλλα και άλλες λύσεις επιτραπέζιου υπολογιστή ή εγκατάστασης εντός της εταιρείας. Τα πλεονεκτήματα της αναλυτικής νέφους το καθιστούν μια εμφανή επιλογή. Ακολουθούν επτά πλεονεκτήματα των αναλυτικών στοιχείων cloud που πρέπει να λάβουν υπόψη όλοι οι ιδιοκτήτες και οι διαχειριστές επιχειρήσεων όταν επιλέγουν μια λύση ανάλυσης.

Ανάπτυξη και επεκτασιμότητα: Οι πλατφόρμες νέφους παρέχουν ισχυρές δυνατότητες κατά απαίτηση με σχεδόν απεριόριστη ευελιξία. Αυτοί οι πόροι είναι διαθέσιμοι κατά παραγγελία για κλιμακωπή ή αποκλιμακωπή, χωρίς την ανάγκη αγοράς, διαμόρφωσης ή διατήρησης διακομιστών ή άλλους πόρους, καθώς οι ανάγκες επέκτασης εξελίσσονται με την ανάπτυξη.

Ταχύτητα και ευελιξία: Τα ιδιωτικά νέφη προσφέρουν μια αποτελεσματική λύση για τη χρήση αναλυτικών στοιχείων μεγάλων δεδομένων και συμπληρώνουν τους εσωτερικούς πόρους μέσω δημόσιων εγκαταστάσεων. Η ευελιξία που παρέχεται από το νέφος επιτρέπει στους χρήστες να αναβαθμίζουν την υποδομή τους καθώς αλλάζουν οι όγκοι των δεδομένων.

Ενιαία προσέγγιση από οπουδήποτε: Όταν οι ομάδες οικονομικών, πληροφορικής, μάρκετινγκ και πωλήσεων διαχειρίζονται τις δικές τους βάσεις δεδομένων και χρησιμοποιούν διάφορα εργαλεία ανάλυσης, οι εργαζόμενοι μπορεί να αναγκαστούν να εκπαιδευτούν σε πολλά συστήματα. Ως αποτέλεσμα, απαιτείται χρονοβόρα αναζήτηση δεδομένων από διάφορα συστήματα ή ακόμα και εκπαίδευση σε πολλές διαφορετικές λύσεις. Τα αναλυτικά στοιχεία νέφους που προέρχονται από το ERP σύστημα μιας εταιρείας παρέχουν σε ολόκληρο το εργατικό της δυναμικό μια ενιαία πηγή πληροφοριών και αναλυτικών στοιχείων, ανεξάρτητα από την τοποθεσία τους.

Ελεύθερη ροή πληροφοριών: Όταν ολόκληρο το προσωπικό της εταιρείας χρησιμοποιεί ένα κοινό σύστημα, διευκολύνεται η συνεργασία μεταξύ διαφορετικών τμημάτων. Παρόλο που οι χρήστες έχουν πρόσβαση μόνο στις πληροφορίες που απαιτούν, οι ομάδες μπορούν πιο εύκολα να επικοινωνούν μεταξύ τους για να ανταλλάξουν πιο χρήσιμες και πολύτιμες πληροφορίες.

Μείωση κόστους: Τα έξοδα για τη φιλοξενία και συντήρηση εσωτερικών συστημάτων περιλαμβάνουν το προσωπικό της τεχνολογίας πληροφορικής, τον εξοπλισμό και την ανάπτυξη - η αναλυτική νέφους απαιτεί μόνο έναν λογαριασμό και, σε πολλές περιπτώσεις, οδηγεί σε συνολικά μειωμένα έξοδα. Η χρήση αναλυτικών στοιχείων βασισμένων σε μεγάλα δεδομένα στο cloud μπορεί να οδηγήσει σε σημαντική μείωση του κόστους σχετικά με την αποθήκευση μεγάλων όγκων δεδομένων και παρέχει δημιουργικούς τρόπους για επιχειρηματική δραστηριότητα.

Εύρεση γρήγορων απαντήσεων: Οι ισχυρότεροι διακομιστές που είναι διαθέσιμοι μέσω της υπολογιστικής νέφους, σε αντίθεση με τα συστήματα εσωτερικής εγκατάστασης ή τους φορητούς υπολογιστές των εργαζομένων, επιτρέπουν ταχύτερη επεξεργασία των δεδομένων.

Αυξημένη κοινή χρήση και συνεργασία: Το «σύννεφο» είναι κατάλληλο για την κοινή χρήση. Αντί να ανταλλάσσονται αρχεία μέσω ηλεκτρονικού ταχυδρομείου, κοινόχρηστου δικτύου ή της σύγχυσης που προκύπτει από πολλαπλές εκδόσεις αρχείων, οι εργαζόμενοι μπορούν να εργάζονται σε κοινά σύνολα δεδομένων και να μοιράζονται αναφορές με ελάχιστα κλικ.

Βελτιωμένη ασφάλεια: Με την χρήση αναλυτικών στοιχείων στο σύννεφο, τα δεδομένα δημιουργούνται περιοδικά αντίγραφα ασφαλείας σε διακομιστές που βρίσκονται σε πολλές τοποθεσίες, εξασφαλίζοντας προστασία κατά την περίπτωση πυρκαγιάς ή φυσικής καταστροφής. Εφόσον δεν αποθηκεύονται τοπικά πληροφορίες, δεν υπάρχουν τοπικοί σκληροί δίσκοι που να είναι ευάλωτοι σε κλοπή, και τα ευαίσθητα δεδομένα δεν μεταδίδονται μέσω μη ασφαλών μεθόδων, όπως το email ή οι αποσπώμενες μονάδες flash. Όλα τα δεδομένα προστατεύονται με κωδικό πρόσβασης, οι χρήστες έχουν πρόσβαση μόνο σε αυτά που απαιτούν και τα αρχεία καταγραφής ελέγχου παρέχουν ορατότητα για το ποιος είχε πρόσβαση σε ποια δεδομένα και πότε, καθώς και τι ενέργειες πραγματοποιήσε.

3.4. Δυνατότητες των Εργαλείων Αναλυτικής Νέφους

Οι πλατφόρμες ανάλυσης προσφέρουν κοινά χαρακτηριστικά που μπορούν να βελτιώσουν την παραγωγικότητα και τα αποτελέσματα των εταιρειών, συμπεριλαμβανομένων:

Βελτιωμένη διαδικασία λήψης αποφάσεων: Η αξιοποίηση αναλύσεων μνήμης σε συνδυασμό με την ικανότητα αξιολόγησης διαφορετικών πηγών δεδομένων βοηθά στην ανάλυση πληροφοριών και επιτρέπει την ταχεία λήψη αποφάσεων με βάση την ανάλυση των δεδομένων. Η Αξιοποίηση της αξίας των μεγάλων δεδομένων από επιχειρήσεις που επικεντρώνονται στον προϋπολογισμό τους για την υπηρεσία Αναλυτικής Δεδομένων ως υπηρεσία (AaaS), η οποία υποστηρίζεται από τα τρία μοντέλα νέφους, δηλαδή το δημόσιο, ιδιωτικό και υβριδικό μοντέλο.

Δεδομένα μοντέλων: Η αναζήτηση λύσεων που περιλαμβάνουν μοντέλα δεδομένων και η ανάλυση μεγάλων δεδομένων επιτρέπει στις επιχειρήσεις να αναπτύσσουν νέα προϊόντα και υπηρεσίες που ανταποκρίνονται στις ανάγκες και επιθυμίες των πελατών, αναλύοντας τις απαιτήσεις και τη συμπεριφορά τους.

Επεξεργασία εφαρμογών: Η επεξεργασία των δεδομένων και η τροποποίησή τους, οδηγούν σε νέα μοτίβα, χρήσεις και πρότυπα πληροφοριών.

Κοινή χρήση δεδομένων: Ο διαμοιρασμός πληροφοριών μεταξύ των μελών μιας ομάδας μπορεί να βελτιώσει την παραγωγικότητα και την αποδοτικότητα σε όλη την εταιρεία.

Αποθήκευση δεδομένων: Οι υπηρεσίες αποθήκευσης στο νέφος επιτρέπουν στις εταιρείες να αυξομειώνουν εύκολα τον χώρο αποθήκευσης ανάλογα με τις ανάγκες τους. Αυτό σημαίνει ότι μπορούν να αποκτήσουν ή να απελευθερώσουν χώρο αποθήκευσης όποτε το απαιτεί η επιχειρηματική τους ανάπτυξη, χωρίς την ανάγκη για αγορά νέου εξοπλισμού.

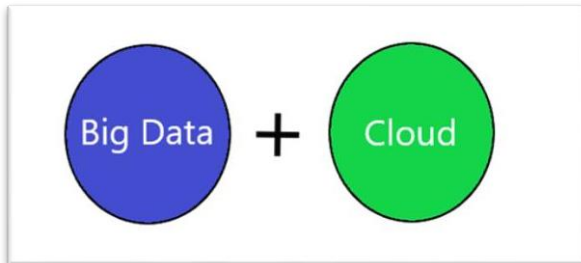
Αναλυτικά μοντέλα: Υπάρχει η δυνατότητα για χρήση προκατασκευασμένων μοντέλων αναλυτικών στοιχείων που έχουν σχεδιαστεί για διάφορες περιπτώσεις χρήσης ή οι χρήστες να δημιουργήσουν τα δικά τους μοντέλα.

Αναφορές: Δυνατότητα δημιουργίας ακριβών και λεπτομερών αναφορών που μπορούν να αξιοποιηθούν κατάλληλα σε παρουσιάσεις στο ανώτατο διοικητικό επίπεδο, στο διοικητικό συμβούλιο, στους δανειστές ή στους επενδυτές μιας εταιρείας.[11]

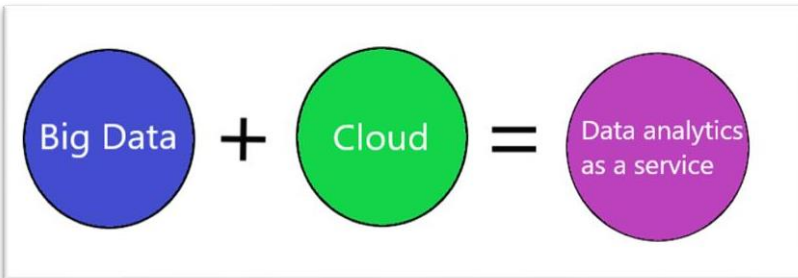
3.5. Τάσεις και προκλήσεις του Υπολογιστικού Νέφους

Οι εταιρείες αξιοποιούν τα αναλυτικά δεδομένα μεγάλου όγκου για να ανακαλύψουν επιχειρηματικές τάσεις και πληροφορίες από την υπερπληθώρα δεδομένων που δημιουργούνται [12]. Η χρήση του υπολογιστικού νέφους παρέχει ευελιξία και τεχνογνωσία για την πρόσβαση σε δεδομένα και την απόσπαση κέρδους από αυτά. Οι επιχειρήσεις μπορούν να εκμεταλλευτούν την ιδιωτική υποδομή του νέφους για να μειώσουν τον κίνδυνο και να διατηρήσουν τον έλεγχο, αναλύοντας τον όγκο δεδομένων, το κόστος, την ασφάλεια και άλλες παραμέτρους. Η ανάλυση μεγάλων δεδομένων

προσφέρει πολλαπλά οφέλη, συμβάλλοντας στην ανάπτυξη της επιχείρησης, τη μείωση του κόστους, την αύξηση των εσόδων και τη βελτίωση του μάρκετινγκ για τους οργανισμούς. Οι υπηρεσίες αναλυτικών στοιχείων, όταν αξιοποιούνται από μια επιχείρηση, ενισχύουν την επεκτασιμότητά της. Οι επαγγελματίες χρήστες έχουν τη δυνατότητα να επεξεργαστούν δεδομένα με τη χρήση εργαλείων συγκριτικής αξιολόγησης τα οποία μπορούν να ενσωματωθούν με βασικές βάσεις δεδομένων για την ανάλυση τόσο δομημένων όσο και μη δομημένων δεδομένων και για τη δημιουργία προηγμένων απεικονίσεων 3D.



Εικόνα 2: Συνδυασμός Μεγάλων Δεδομένων και Υπολογιστικού νέφους Data Analytics As A Service (Daaas) [13]



Εικόνα 3: Ανάλυση μεγάλων δεδομένων σαν υπηρεσία αποτελεί συνδυασμό αναλυτικής μεγάλων δεδομένων και του υπολογιστικού νέφους. [13]

Η ανάλυση μεγάλων δεδομένων (Big Data Analytics) [13] αναφέρεται στη διαδικασία εξέτασης τεράστιων συνόλων δεδομένων με σκοπό την ανακάλυψη κρυφών μοτίβων, άγνωστων συσχετίσεων, τάσεων της αγοράς, προτιμήσεων πελατών και άλλων χρήσιμων πληροφοριών. Η αναλυτική δεδομένων ως υπηρεσία (Daaas) αποτελεί μια επεκτάσιμη αναλυτική πλατφόρμα που παρέχεται μέσω ενός μοντέλου παράδοσης βασισμένου στο νέφος. Αυτή η πλατφόρμα παρέχει διάφορα εργαλεία για ανάλυση δεδομένων και μπορεί να προσαρμοστεί από τον χρήστη για την αποτελεσματική επεξεργασία και ανάλυση τεράστιων ποσοτήτων ετερογενών δεδομένων. Ο συνδυασμός των μεγάλων δεδομένων (Big Data) με το υπολογιστικό νέφος μπορεί να βελτιώσει την υιοθέτηση αναλυτικών δυνατοτήτων σε μεγάλο όγκο ετερογενών πηγών δεδομένων που οι επιχειρήσεις δυσκολεύονται να χειριστούν.

Οι εταιρείες επωφελούνται σε μεγάλο βαθμό από την αξία που προκύπτει από την ανάλυση μεγάλων δεδομένων. Η αξία βρίσκεται στα ίδια τα δεδομένα και όχι στην τεχνολογία που χρησιμοποιείται σε αυτά. Οι εταιρείες πρέπει να έχουν πρόσβαση σε υπηρεσίες που παρέχουν εκ των προτέρων αναλυτικά στοιχεία μεγάλων δεδομένων για το μέλλον. Πρέπει επίσης να αναπτύξουν καλή υποδομή και τις σχετικές επενδύσεις κεφαλαίου. Το «σύννεφο» μπορεί να είναι καλό μέσο για τους αναλυτές λόγω της ιδιότητας της ελαστικότητας. Το μοντέλο λογισμικού ως υπηρεσία (SaaS) προσαρμόστηκε από παρόχους υπηρεσιών ανάλυσης για να βοηθήσει το νέφος να επιλύει προβλήματα μεγάλων δεδομένων. Για παράδειγμα, η Emcien Corp παρέχει λογισμικό ανίχνευσης προτύπων ως υπηρεσία που εκτελείται στο Amazon Elastic Compute Cloud (EC2).

Εξυπηρετεί πελάτες όπως μεγάλους λιανοπωλητές, παρόχους τηλεπικοινωνιών και υπηρεσίες πληροφοριών. Στο μοντέλο SoftwareAsAService [14], ο χρήστης δεν χρειάζεται να αγοράσει όλο αυτό το υλικό, αλλά χρειάζεται μόνο να έχει ένα πρόγραμμα περιήγησης ιστού (Browser). Τα μεγάλα δεδομένα είναι εξελιγμένες πλατφόρμες επειδή για να λάβουν οικονομικά αποδοτικές υπηρεσίες από προμηθευτές αναλυτικών στοιχείων, ο καταναλωτής πρέπει να εγκαταστήσει πρόσθετο λογισμικό πόρους και άλλα βασικά στοιχεία για καλύτερη απόδοση.

3.6. Αναλυτική Δεδομένων ως Υπηρεσία (DAaaS) - Εφαρμογές

Η ανάλυση δεδομένων ως υπηρεσία αποτελεί μια αναδυόμενη τάση που έχει ευρεία εφαρμογή σε πολλούς κλάδους. Οι εταιρείες αποκτούν ανταγωνιστικό πλεονέκτημα μέσω της προσαρμογής και εφαρμογής της υπηρεσίας ανάλυσης δεδομένων ως υπηρεσίας (DAaaS). Η προσέγγιση ανάλυσης διαφέρει ανάλογα με τον κλάδο δραστηριότητας. Οι εφαρμογές του DAaaS περιλαμβάνουν τα εξής:

Κοινωνικά δίκτυα: Η ανάλυση δεδομένων στον χώρο του υπολογιστικού νέφους έχει μεγάλη σημασία για τα κοινωνικά δίκτυα, όπως το Facebook, το Twitter και άλλα. Οι υπηρεσίες νέφους επιτρέπουν την παράλληλη εξέταση των δεδομένων των κοινωνικών δικτύων, ενώ τα αποτελέσματα μπορούν να αναλυθούν σε πραγματικό χρόνο.

Ιατρική: Οι τεχνολογίες ανάλυσης μεγάλων δεδομένων, και αναλυτικών πληροφοριών μπορούν να συμβάλουν στη βελτίωση της ιατρικής πρακτικής, επιτρέποντας στους γιατρούς να λαμβάνουν καλύτερες αποφάσεις με βάση πραγματικού χρόνου ανάλυση του ιστορικού του ασθενούς, των γενετικών του χαρακτηριστικών, των εξετάσεων αίματος, της ανίχνευσης παρόμοιων ιατρικών περιπτώσεων και της πρόσβασης σε ερευνητικά άρθρα από όλο τον κόσμο. Οι γιατροί μπορούν να αξιοποιήσουν αναλυτικά εργαλεία, τεχνικές και να εκμεταλλευτούν την προηγμένη τεχνολογία, μειώνοντας έτσι το κόστος των θεραπειών.

Έξυπνες πόλεις: Η υπηρεσία ανάλυσης δεδομένων που βασίζεται στο υπολογιστικό σύννεφο μπορεί να συμβάλει στην ανάπτυξη έξυπνων πόλεων (Smart Cities) μέσω της έξυπνης χρήσης των μεγάλων δεδομένων. Πληροφορίες σε πραγματικό χρόνο σχετικά με μεταφορές, ενέργεια, έξυπνο φωτισμό, ηχορύπανση, ενημέρωση των πολιτών, διαχείριση απορριμμάτων, διαχείριση νερού κ.λπ. μπορούν να συλλέγονται και να αποθηκευτούν στο νέφος για ανάλυση. Η διοίκηση της πόλης μπορεί να αξιοποιήσει δεδομένα από διάφορες πηγές για να προσφέρει καλύτερη διακυβέρνηση, σχεδιασμό και διαχείριση. Πτυχές της αστικής ζωής, όπως η υγεία, η ρύπανση και η ατμοσφαιρική ποιότητα μπορούν να βελτιωθούν και να επωφεληθούν από τις δυνατότητες ανάλυσης μεγάλων δεδομένων στο σύννεφο.

Προτίμηση παρακολούθησης και προϊόντα: Η Amazon χρησιμοποιεί υπηρεσίες ανάλυσης δεδομένων σε προγράμματα οδήγησης cloud για να παρακολουθεί το προϊόν σε ολόκληρη την αποθήκη της. Οι ιστότοποι παρέχουν προτάσεις στους χρήστες βασιζόμενοι στο προηγούμενο ιστορικό αναζήτησής τους. Οι προτιμήσεις των χρηστών αποθηκεύονται στο cloud για μελλοντική αναφορά. Ο αναλυτής των κοινωνικών μέσων αναλύει δεδομένα σε όλο το δίκτυο για να προσφέρει πρότυπα αγορών και να αξιολογήσει την επιτυχία τους.

Τήρηση αρχείων πελατών: Ο ιστότοπος αποθηκεύει το προφίλ του χρήστη στο cloud για μελλοντική αναφορά και επεξεργάζεται δεδομένα σε τοπικούς διακομιστές σε πραγματικό χρόνο. Οι περισσότερες πληροφορίες του χρήστη αποθηκεύονται στο cloud, επιτρέποντας τη συνέχιση των προτιμήσεών του όταν αλλάζει συσκευή.

3.7. Αναλυτική Δεδομένων ως Υπηρεσία (DAaaS) - Περιορισμοί

Η πλατφόρμα DAaaS έχει κατασκευαστεί για οριζόντια πλατφόρμα που μπορεί να χρησιμοποιηθεί σε διαφορετική κάθετη πλατφόρμα. Μπορεί να σχεδιαστεί ώστε να είναι καθολική καθώς και επεκτάσιμη πλατφόρμα. Καθώς το DAaaS είναι μια καθολική πλατφόρμα που έχει ζητήματα αποθήκευσης και μοντελοποίησης δεδομένων και δεν υποστηρίζει ανάλυση δεδομένων σε πραγματικό χρόνο. Για παράδειγμα, το πλαίσιο MapReduce βασίζεται σε τρόπους λειτουργίας

ομαδικής εκτέλεσης. Αυτό επιβάλλει φόρτο εργασίας στο σύστημα εάν χρησιμοποιούνται τρόποι λειτουργίας σε πραγματικό χρόνο. Η θεμελιώδης ιδιότητα των μεγάλων δεδομένων είναι να μπορεί να αντιπροσωπεύει και να αποθηκεύει διαφορετικά είδη δεδομένων και να αποθηκεύει τεράστιο όγκο δεδομένων με κατανεμημένο τρόπο. Οι περιορισμοί DAaaS είναι οι εξής:

Αποθήκευση: Ο χώρος αποθήκευσης επηρεάζεται από φορτία εικονικοποίησης, με αποτέλεσμα να επηρεάζει την απόδοση σε ένα εξαιρετικά εικονικοποιημένο και κατανεμημένο σύννεφο είναι επικίνδυνο.

Ζητήματα δικτύωσης: Η ανάλυση μεγάλων δεδομένων στο cloud προκαλεί προβλήματα δικτύωσης για τους παρόχους υπηρεσιών. Απαιτούνται τεχνικές γνώσεις για την αποτελεσματική μεταφορά δεδομένων σε μια ομάδα ανάλυσης.

Αναλυτική Γνώση: Η ανάλυση δεδομένων απαιτεί καταρτισμένο επιστήμονα δεδομένων, ο οποίος είναι ενημερωμένος με τις πιο πρόσφατες προηγμένες τεχνικές ανάλυσης.

Ασυνέπεια στη συλλογή δεδομένων: Τα εργαλεία που χρησιμοποιούνται για τη συλλογή μεγάλων συνόλων δεδομένων είναι μερικές φορές ανακριβή.

Ασφάλεια και απόρρητο: Οι επιχειρήσεις μπορεί να ανησυχούν ότι τα δεδομένα τους αποθηκεύονται σε έναν εικονικό διακομιστή που μεταβιβάζει τα δεδομένα τους σε τρίτους. Το απόρρητο μπορεί να επηρεάσει τα αναλυτικά στοιχεία του cloud, καθώς τα δεδομένα γίνονται δημόσια μετά την ανάλυση.

Όγκος δεδομένων: Η μεταφορά δεδομένων στο σύννεφο για υπολογισμό γίνεται επίπονη εργασία, αλλά το αντίστροφο δεν ισχύει. [15]

3.8. Προκλήσεις του υπολογιστικού νέφους

Η χρήση και η μετάβαση στο σύννεφο έχει πάρα πολλά οφέλη. Ωστόσο, δεν είναι ούτε η βέλτιστη λύση ούτε ακίνδυνη, ιδιαίτερα όταν τα δεδομένα δεν είναι προσβάσιμα από τους χρήστες. Άλλα μειονεκτήματα της μετάβασης στο νέφος είναι τα ζητήματα αξιοπιστίας και η απόδοση του συστήματος, καθώς οι χρήστες εξαρτώνται πλήρως από τους πόρους του. Για παράδειγμα, όταν κάποιος έχει πρόσβαση σε αυτό για να αναζητήσει μια υπηρεσία, ο χρόνος που απαιτείται για την εκτέλεση της εργασίας (ο χρόνος μετ' επιστροφής ή RTT) θα μπορούσε να είναι ένα πρόβλημα για τους χρήστες. Αυτό μπορεί να επιδεινωθεί εάν το cloud είναι απασχολημένο με την εξυπηρέτηση άλλων παρουσιών ή η κυκλοφορία έχει ήδη συμφορηθεί. Το απόρρητο και η ασφάλεια είναι άλλοι περιορισμοί που είναι ευρέως γνωστό ότι καθιστούν το Υπολογιστική Νέφος που αποτελεί πρόκληση για τους χρήστες. Κάποια από τις κύριες προκλήσεις που αντιμετωπίζει το Υπολογιστική Νέφος παρέχονται ως εξής:

A) Αξιοπιστία: Μια πλατφόρμα υπολογιστικού νέφους πρέπει να εγγυάται την αξιοπιστία της πλατφόρμας εφαρμογής και την ακεραιότητα των δεδομένων των πελατών. Όταν υπάρχει εμπειρία ενός συστήματος μεγάλης κλίμακας, αναμένεται μια αποτελεσματική λύση προκειμένου να ληφθεί υψηλό επίπεδο αξιοπιστίας. Επιπλέον, ένα δυναμικό σύστημα διαχείρισης δικτύου ελέγχει την αποτελεσματικότητα και την κατάσταση των κόμβων πόρων, οπότε οι αναποτελεσματικοί ή αποτυχημένοι κόμβοι μετεγκαθίστανται δυναμικά. Κατά συνέπεια, ολόκληρη η απόδοση του συστήματος δεν επηρεάζεται από αυτούς τους κόμβους. Η διασφάλιση ότι όλα τα συστήματα λειτουργούν τέλεια και αξιόπιστα θέτει συνεχείς προκλήσεις τους παρόχους υπηρεσιών νέφους.

B) Προμήθεια και Προγραμματισμός Πόρων: Η δυναμική κατάργηση και η επέκταση των πόρων βασίζεται στις απαιτήσεις των χρηστών, παρουσιάζοντας έτσι νέες προκλήσεις για τα συστήματα διαχείρισης και τις πλατφόρμες υπολογιστικού νέφους. Όσον αφορά την παροχή πόρων, είναι εξαιρετικά απαραίτητος ένας αποτελεσματικός αλγόριθμος παροχής πόρων που να κάνει καλύτερη χρήση και κατανομή τους, να μειώνει τον χρόνο απόκρισης και να έχει στιβαρότητα καθώς και δυνατότητες ανοχής σφαλμάτων. Ομοίως, ο προγραμματισμός για τις απαιτήσεις πόρων κατ' απαίτηση ή η μακροπρόθεσμη κράτηση πόρων είναι μια πρόκληση, ειδικά όταν ο αριθμός των πόρων και ο αριθμός των χρηστών είναι εξαιρετικά μεγάλοι.

Γ) Θέματα Διαχείρισης: Η διαδικασία διαχείρισης μιας πλατφόρμας υπολογιστικού νέφους είναι εξαιρετικά περίπλοκη. Ειδικά, η ενοποίηση των πόρων είναι ένας από τους βασικούς τομείς της έρευνας και έχει κερδίσει ουσιαστική προσοχή από την ερευνητική κοινότητα τα τελευταία χρόνια. Περιλαμβάνει τη διαχείριση των μέσων για τον αποτελεσματικό έλεγχο των πόρων του συστήματος, την ανάπτυξη και τον προγραμματισμό διαφορετικών πόρων με δυναμικό τρόπο και τη διαχείριση των πελατών, των συστημάτων χρέωσης και των συμφωνιών παροχής υπηρεσιών. Ωστόσο, η εφαρμογή της προσέγγισης της ύπαρξης ενός παρόχου υπηρεσιών δημιουργεί εμπόδια, όπως τα ακόλουθα: (i) πολλή ενέργεια αξιοποιείται μέσω ενός τεράστιου κέντρου δεδομένων προκειμένου να λειτουργήσει. (ii)

πολλή ενέργεια εκμεταλλεύεται ένα τεράστιο κέντρο δεδομένων για να είναι λειτουργικό και (iii) τα κέντρα δεδομένων είναι γεωγραφικά απομακρυσμένα από τους χρήστες τους και τα δεδομένα πρέπει να μετακινηθούν από την πηγή τους προκειμένου να υποστούν επεξεργασία. Αυτό σημαίνει ότι τα προσωπικά ή ευαίσθητα δεδομένα που δημιουργούνται μέσω της χρήσης διαφορετικών εφαρμογών φυλάσσονται σε τοποθεσία διαφορετική από εκεί που παράγονται.

Δ) Ανοχή σφαλμάτων: Η ανοχή σφαλμάτων αναφέρεται στη συνέχεια των υπηρεσιών νέφους ακόμη και στην ύπαρξη οποιασδήποτε δυσλειτουργίας υλικού ή λογισμικού. Σε περίπτωση αστοχίας τέτοιων εξαρτημάτων, είναι μεγάλη πρόκληση να διατηρηθεί όλο το σύστημα σε λειτουργία και χωρίς υποβάθμιση της απόδοσης παρουσία σφάλματος.

Ε) Απόρρητο και Διαφάνεια: Το απόρρητο και η διαφάνεια των δεδομένων των χρηστών και των υπηρεσιών αντίστοιχα είναι πολύ κρίσιμες σε κάθε σύστημα υπολογιστικού νέφους.

Προκειμένου να εξασφαλισθεί η εμπιστοσύνη σε μια υπηρεσία που βασίζεται σε σύννεφο, όπου τα δεδομένα των χρηστών και όλα τα σχετικά διαπιστευτήρια αποθηκεύονται σε ένα εικονικό περιβάλλον, το απόρρητο των δεδομένων είναι πολύ σημαντικό. Ομοίως, η διαφάνεια των υπηρεσιών cloud και η εικονικοποίηση όλων των συστημάτων και υποδομών είναι πρωταρχικής σημασίας. Οι πάροχοι υπηρεσιών πρέπει να ενημερώνουν τους πελάτες για τον τρόπο με τον οποίο θα διατηρούνται, θα αποθηκευτούν και θα μεταδοθούν τα δεδομένα τους. Επομένως, τα συστήματα ασφάλειας και απορρήτου που αναπτύσσονται και ποιες εσωτερικές πολιτικές και τεχνολογίες εφαρμόζονται, είναι πολύ σημαντικά από την άποψη ενός πελάτη, ειδικά όταν ο πελάτης είναι ένας μεγάλος οργανισμός.

ΣΤ) Ασφάλεια: Η ασφάλεια του Υπολογιστική Νέφους είναι κατά πολύ η μεγαλύτερη πρόκληση σε αυτήν την τεχνολογία. Αντιμετωπίζει κάθε είδους προκλήσεις που σχετίζονται με την ασφάλεια των δεδομένων, την ασφάλεια των πληροφοριών, την ακεραιότητα των δεδομένων και την εμπιστευτικότητα. Η ασφάλεια είναι μία από τις κύριες προκλήσεις και εμπόδια όσον αφορά την εφαρμογή και την υιοθέτηση του υπολογιστικού νέφους. Ως εκ τούτου, μια λεπτομερής συζήτηση για αυτό και τα σχετικά θέματα είναι αναπόφευκτη για την πληρότητα αυτής της εργασίας. [26]

4. Μηχανική Μάθηση

Επισκόπηση μηχανικής μάθησης: Η μηχανική μάθηση είναι ένα υποσύνολο της Τεχνητής Νοημοσύνης (AI) που εστιάζει στην κατασκευή συστημάτων που μαθαίνουν ή βελτιώνουν την απόδοση με βάση τα δεδομένα που καταναλώνουν. [16]

4.1. Τι είναι η Μηχανική Μάθηση;

Η μηχανική μάθηση είναι μια τεχνική που ανακαλύπτει προηγουμένως άγνωστες σχέσεις στα δεδομένα. Η μηχανική μάθηση και η τεχνητή νοημοσύνη συζητούνται συχνά μαζί. Μια σημαντική διάκριση είναι ότι παρόλο που όλη η μηχανική μάθηση είναι AI, δεν είναι όλη η τεχνητή νοημοσύνη μηχανική μάθηση. Η μηχανική μάθηση αναζητά αυτόματα δυνητικά μεγάλα αποθέματα δεδομένων για να ανακαλύψει μοτίβα και τάσεις που υπερβαίνουν την απλή στατιστική ανάλυση. Η μηχανική εκμάθηση χρησιμοποιεί εξελιγμένους αλγόριθμους που προσδιορίζουν μοτίβα στα μοντέλα δημιουργίας δεδομένων. Αυτά τα μοντέλα μπορούν να χρησιμοποιηθούν για να κάνουν προβλέψεις και να κατηγοριοποιήσουν δεδομένα.

Τα βασικά χαρακτηριστικά της μηχανικής μάθησης είναι:

- Αυτόματη ανακάλυψη μοτίβων
- Πρόβλεψη πιθανών αποτελεσμάτων
- Δημιουργία ενεργών πληροφοριών
- Δυνατότητα ανάλυσης δυνητικά μεγάλου όγκου δεδομένων

Η μηχανική μάθηση μπορεί να απαντήσει σε ερωτήσεις που δεν μπορούν να αντιμετωπιστούν μέσω παραδοσιακών τεχνικών απαγωγικών ερωτημάτων και αναφοράς.

4.2. Οφέλη της Μηχανικής Μάθησης

Η μηχανική μάθηση αποτελεί μια ισχυρή τεχνολογία που μπορεί να βοηθήσει κάποιον να ανακαλύψει μοτίβα και σχέσεις στα δεδομένα του.

Βρίσκοντας τάσεις και μοτίβα: η μηχανική εκμάθηση ανακαλύπτει κρυφές πληροφορίες στα δεδομένα. Με την πάροδο του χρόνου, μπορεί να υπάρχουν ήδη σημαντικά μοτίβα που έχουν προκύψει από την ανάλυση των δεδομένων. Η μηχανική μάθηση μπορεί να επιβεβαιώσει ή να καταχωρίσει αυτές τις εμπειρικές παρατηρήσεις, αλλά μπορεί επίσης να ανακαλύψει νέα πρότυπα που δεν είναι εμφανή μέσω της απλής παρατήρησης. Μέσω της μηχανικής μάθησης, μπορούν να αναδειχθούν προγνωστικές σχέσεις που δεν είναι αιτιώδεις, και αυτές οι πληροφορίες μπορούν να χρησιμοποιηθούν για την ανάπτυξη μιας στρατηγικής μάρκετινγκ. Η μηχανική μάθηση μπορεί να αντιμετωπίσει μεγάλα σε όγκο δεδομένα και να χρησιμοποιηθεί σε οικονομική ανάλυση. Οι εφαρμογές της περιλαμβάνουν την πρόβλεψη τιμών μετοχών, την αλγοριθμική συναλλαγή και τη διαχείριση χαρτοφυλακίου. Αυτές οι τεχνικές επιτρέπουν στις εταιρείες να αναγνωρίσουν μοτίβα στις αγορές και να προβλέψουν τις τάσεις των τιμών, παρέχοντας έτσι ένα πλεονέκτημα στον ανταγωνιστικό χώρο.

Λήψη αποφάσεων βάσει δεδομένων: Πολλές εταιρείες διαθέτουν μεγάλα δεδομένα και η εξαγωγή σημαντικών πληροφοριών από αυτά τα δεδομένα είναι σημαντική για τη λήψη επιχειρηματικών αποφάσεων που βασίζονται σε δεδομένα. Αξιοποιώντας τους αλγόριθμους μηχανικής μάθησης, οι οργανισμοί είναι σε θέση να μετατρέψουν τα δεδομένα σε γνώση και ευφυΐα που μπορεί να λειτουργήσει. Με τις μεταβαλλόμενες απαιτήσεις, οι εταιρείες είναι σε θέση να λαμβάνουν καλύτερες αποφάσεις πιο γρήγορα χρησιμοποιώντας τεχνικές μηχανικής μάθησης.

Συνολικά, η μηχανική μάθηση και η ανάλυση δεδομένων παρέχουν στις εταιρείες μια ισχυρή εργαλειοθήκη για την αντιμετώπιση πολύπλοκων προκλήσεων και την επίτευξη επιτυχημένων στρατηγικών λήψης αποφάσεων.

Προτάσεις λύσεων – προϊόντων: Τα αποτελέσματα της μηχανικής εκμάθησης μπορούν να χρησιμοποιηθούν για να επηρεάσουν τις αποφάσεις των πελατών, προωθώντας ή προτείνοντας

προϊόντα που είναι σχετικά και χρήσιμα για αυτούς. Με βάση τα πρότυπα συμπεριφοράς των πελατών στο διαδίκτυο ή την ανταπόκρισή τους σε μια καμπάνια μάρκετινγκ, η μηχανική εκμάθηση μπορεί να αναλύσει τα δεδομένα και να παράσχει εξατομικευμένες προτάσεις προϊόντων. Με τη χρήση αυτών των τεχνικών, οι εταιρείες μπορούν επίσης να κατανοήσουν καλύτερα τις ανάγκες και τις προτιμήσεις των πελατών τους, προσαρμόζοντας και βελτιώνοντας τις προσφερόμενες προϊόντα και υπηρεσίες.

Εντοπισμός απάτης, ανωμαλιών και κινδύνων ασφάλειας: - Ο τομέας των Χρηματοοικονομικών Υπηρεσιών έχει επωφεληθεί από αλγόριθμους και τεχνικές μηχανικής μάθησης ανακαλύπτοντας ασυνήθιστα μοτίβα ή απάτες και ανταποκρινόμενος σε νέες συμπεριφορές απάτης πολύ πιο γρήγορα. Σήμερα οι εταιρείες και οι κυβερνήσεις διεξάγουν επιχειρηματικές δραστηριότητες και μοιράζονται πληροφορίες στο διαδίκτυο. Σε τέτοιες περιπτώσεις, η ασφάλεια του δικτύου είναι μια ανησυχία. Η μηχανική μάθηση μπορεί να βοηθήσει στον εντοπισμό ανώμαλης συμπεριφοράς και να λάβει αυτόματα διορθωτικές ενέργειες.

Ανάλυση πωλήσεων: Η μηχανική μάθηση βοηθά στην ανάλυση των μοτίβων αγορών πελατών για την παροχή προωθητικών προσφορών για πελάτες - στόχους. Αυτή η υπηρεσία εξασφαλίζει ανώτερη εμπειρία πελατών και βελτιώνει την αφοσίωση των πελατών.

Υγειονομική περίθαλψη: Η μηχανική μάθηση στην ιατρική χρήση γίνεται κοινή, βοηθώντας ασθενείς και γιατρούς. Οι προηγμένες τεχνικές μηχανικής μάθησης χρησιμοποιούνται στην ακτινολογία για τη λήψη μιας έξυπνης απόφασης εξετάζοντας εικόνες όπως ακτινογραφίες, CT, MRI, εικόνες PET και αναφορές ακτινολογίας. Αναφέρεται ότι η αυτόματη ανίχνευση και διάγνωση με βάση τη μηχανική μάθηση είναι ισοδύναμη ή καλύτερη από τη διάγνωση ενός πραγματικού ακτινολόγου. Ορισμένες από τις εφαρμογές μηχανικής εκμάθησης είναι εκπαιδευμένες για την ανίχνευση του καρκίνου του μαστού. Ορισμένες εφαρμογές που χρησιμοποιούν μηχανική εκμάθηση μπορούν επίσης να υποδείξουν τον κίνδυνο του ασθενούς για διάφορες καταστάσεις όπως εγκεφαλικό επεισόδιο, διαβήτη, στεφανιαίες παθήσεις και νεφρική ανεπάρκεια και να συστήσουν φαρμακευτική αγωγή ή διαδικασία που μπορεί να είναι απαραίτητη.

Συνοψίζοντας, η μηχανική μάθηση μπορεί να:

- αναγνωρίζει εύκολα τάσεις και μοτίβα.
- απλοποιεί την πρόβλεψη μάρκετινγκ και πωλήσεων προϊόντων.
- διευκολύνει την έγκαιρη ανίχνευση ανωμαλιών.
- ελαχιστοποιεί τη χειρωνακτική παρέμβαση μέσω της εκπαίδευσης και χειρίζεται πολυδιάστατα δεδομένα.

4.3. Επίλυση προβλημάτων με την Μηχανική Μάθηση

Οι επιχειρήσεις αντιμετωπίζουν προβλήματα όπως η ταξινόμηση εγγράφων, η πρόβλεψη των οικονομικών αποτελεσμάτων, ο εντοπισμός κρυφών προτύπων και ανωμαλιών κ.λπ. Η μηχανική μάθηση μπορεί να βοηθήσει στην επίλυση τέτοιων προβλημάτων, υπό την προϋπόθεση ότι έχει ξεκάθαρη κατανόηση του επιχειρηματικού προβλήματος με αρκετά δεδομένα και μάθει να κάνει τις σωστές ερωτήσεις για να έχει ουσιαστικά αποτελέσματα. Απαιτούνται δεξιότητες στην προετοιμασία δεδομένων, την εφαρμογή τεχνικών μηχανικής μάθησης και την αξιολόγηση των αποτελεσμάτων.

Τα μοτίβα που βρίσκονται μέσω της μηχανικής εκμάθησης μπορεί να είναι πολύ διαφορετικά ανάλογα με τον τρόπο που διατυπώνεται το πρόβλημα. Για παράδειγμα, αντί να προσπαθεί να μάθει πώς να "βελτιώσει την ανταπόκριση σε μια καμπάνια απευθείας αλληλογραφίας", μπορεί να προσπαθήσει να βρει τα χαρακτηριστικά των ατόμων που έχουν απαντήσει στις καμπάνιες του στο παρελθόν. Στη συνέχεια, μπορεί να ταξινομήσει εάν ένα δεδομένο προφίλ ενός υποψήφιου πελάτη θα ανταποκρινόταν σε μια καμπάνια απευθείας email. Πολλές μορφές μηχανικής μάθησης είναι προγνωστικές. Για παράδειγμα, ένα μοντέλο μπορεί να προβλέψει το επίπεδο εισοδήματος με βάση την εκπαίδευση και άλλους δημογραφικούς παράγοντες. Ορισμένες μορφές προγνωστικής μηχανικής μάθησης δημιουργούν κανόνες, οι οποίοι είναι συνθήκες που υποδηλώνουν ένα δεδομένο αποτέλεσμα. Για παράδειγμα, ένας κανόνας μπορεί να ορίζει ότι ένα άτομο που έχει

πτυχίο και ζει σε μια συγκεκριμένη γειτονιά είναι πιθανό να έχει εισόδημα μεγαλύτερο από τον περιφερειακό μέσο όρο. Άλλες μορφές μηχανικής μάθησης προσδιορίζουν ομαδοποιήσεις στα δεδομένα. Για παράδειγμα, ένα μοντέλο μπορεί να προσδιορίσει το τμήμα του πληθυσμού που έχει εισόδημα εντός ενός καθορισμένου εύρους, που έχει καλό ιστορικό οδήγησης και που μισθώνει ένα νέο αυτοκίνητο σε ετήσια βάση.

4.4. Οφέλη της Μηχανικής Μάθησης στο Cloud

Πολλοί οργανισμοί είναι σε θέση να δημιουργήσουν εσωτερικά μοντέλα μηχανικής μάθησης, χρησιμοποιώντας πλαίσια ανοιχτού κώδικα όπως το ScikitLearn, το TensorFlow ή το PyTorch. Ωστόσο, ακόμα κι αν οι εσωτερικές ομάδες είναι σε θέση να δημιουργήσουν αλγόριθμους, συχνά θα δυσκολεύονται να αναπτύξουν μοντέλα στην παραγωγή και να τα κλιμακώσουν σε πραγματικούς φόρτους εργασίας, κάτι που συχνά απαιτεί μεγάλα συμπλέγματα υπολογιστών.

Υπάρχουν πολλά εμπόδια στην είσοδο για την ανάπτυξη δυνατοτήτων μηχανικής εκμάθησης σε εταιρικές εφαρμογές. Η τεχνογνωσία που απαιτείται για την κατασκευή, την εκπαίδευση και την ανάπτυξη μοντέλων μηχανικής εκμάθησης προσθέτει στο κόστος εργασίας, ανάπτυξης και υποδομής, μαζί με την ανάγκη αγοράς και λειτουργίας εξειδικευμένου εξοπλισμού υλικού.

Πολλά από αυτά τα προβλήματα μπορούν να αντιμετωπιστούν με την Υπολογιστική Νέφος. Τα δημόσια σύννεφα και οι υπηρεσίες Τεχνητής Νοημοσύνης ως Υπηρεσία AaaS (Artificial Intelligence as a Service) βοηθούν τους οργανισμούς να αξιοποιήσουν τις δυνατότητες μηχανικής εκμάθησης για την επίλυση επιχειρηματικών προβλημάτων χωρίς να χρειάζεται να αναλάβουν το τεχνικό βάρος. Τα βασικά οφέλη του υπολογιστικού νέφους για φόρτους εργασίας μηχανικής εκμάθησης μπορούν να συνοψιστούν ως εξής:

- Τα μοντέλα τιμολόγησης κατ' απαίτηση καθιστούν δυνατή την ανάληψη πρωτοβουλιών Μηχανικής Μάθησης χωρίς μεγάλη επένδυση κεφαλαίου.
- Το υπολογιστικό νέφος παρέχει την ταχύτητα και την απόδοση των GPU και των FPGA χωρίς να απαιτείται επένδυση σε υλικό.
- Το νέφος επιτρέπει στις επιχειρήσεις να πειραματίζονται εύκολα με τις δυνατότητες μηχανικής μάθησης και να κλιμακώνονται καθώς τα έργα προχωρούν στην παραγωγή και η ζήτηση για αυτές τις δυνατότητες αυξάνεται.
- Το νέφος επιτρέπει την πρόσβαση σε δυνατότητες ML χωρίς προηγμένες δεξιότητες στην τεχνητή νοημοσύνη ή στην επιστήμη δεδομένων.

4.5. Περιορισμοί της Μηχανικής Μάθησης στο Υπολογιστικό Νέφος

Η μηχανική εκμάθηση στο σύννεφο έχει τρεις βασικούς περιορισμούς:

Δεν αντικαθιστά τους ειδικούς - Παρά το γεγονός ότι η μηχανική μάθηση (ML), ακόμη και όταν λειτουργεί στο cloud, εξακολουθεί να απαιτεί ανθρώπινη εποπτεία και βελτιστοποίηση. Υπάρχουν πρακτικά όρια για το τι μπορεί να επιτύχει η τεχνητή νοημοσύνη χωρίς τη συνεχή παρακολούθηση και παρέμβαση του ανθρώπου. Οι αλγόριθμοι δεν κατανοούν πλήρως μια κατάσταση και δεν γνωρίζουν πώς να αντιδράσουν σε κάθε δυνατή είσοδο.

Κινητικότητα δεδομένων— Κατά την εκτέλεση μοντέλων μηχανικής μάθησης στο νέφος, η μετάβαση από ένα νέφος ή μια υπηρεσία σε μια άλλη μπορεί να αποτελέσει πρόκληση. Αυτό απαιτεί τη μεταφορά των δεδομένων με τρόπο που να μην επηρεάζει την απόδοση του μοντέλου. Τα μοντέλα μηχανικής μάθησης είναι συχνά ευαίσθητα σε μικρές αλλαγές στα δεδομένα εισόδου. Για παράδειγμα, ένα μοντέλο μπορεί να αντιμετωπίσει προβλήματα εάν αλλάξει η μορφή ή το μέγεθος των δεδομένων.

Ανησυχίες για την ασφάλεια— Η μηχανική μάθηση που λειτουργεί μέσω υπολογιστικού νέφους αντιμετωπίζει τις ίδιες ανησυχίες με οποιαδήποτε άλλη πλατφόρμα νέφους. Τα συστήματα μηχανικής μάθησης που βασίζονται στο νέφος εκτίθενται συχνά σε δημόσια δίκτυα και μπορεί να επέλθουν παραβιάσεις από εισβολείς, οι οποίοι μπορεί να παραπλανήσουν τα αποτελέσματα της

μηχανικής μάθησης ή να αυξήσουν το κόστος της υποδομής. Επιπλέον, τα μοντέλα μηχανικής μάθησης που βασίζονται σε υπολογιστικό νέφος είναι ευάλωτα σε επιθέσεις άρνησης υπηρεσίας (DoS). Αυτοί οι κίνδυνοι συνήθως δεν υφίστανται όταν τα μοντέλα αναπτύσσονται πίσω από εταιρικά τείχη προστασίας. [35]

5. Παρουσίαση της πλατφόρμας Oracle Cloud

Η Oracle Cloud Platform ανήκει στις κορυφαίες 7 λύσεις στις Υπηρεσίες PaaS. Το Oracle Cloud Platform συγκρίνεται πιο συχνά με το Microsoft Azure. Η Oracle Cloud Platform είναι δημοφιλής στο τμήμα μεγάλων επιχειρήσεων, αντιπροσωπεύοντας το 65% των χρηστών σύμφωνα με το PeerSpot.

Το Oracle Cloud Platform είναι μια λύση Product as Service Cloud που αποτελείται από δύο κύριες κατηγορίες προϊόντων – Oracle Cloud Infrastructure και Oracle Cloud Applications. Μέσω των πολλών προϊόντων της, η Oracle Cloud Platform προσφέρει μια κοινή κλιμακούμενη πλατφόρμα για την ενοποίηση υπαρχουσών εφαρμογών και ανάπτυξη νέων εφαρμογών. Το Oracle Cloud Infrastructure (OCI) είναι το τμήμα της Oracle Cloud Platform για την εκτέλεση εφαρμογών με ταχύτερο και ασφαλέστερο τρόπο. Προσφέρει υπηρεσίες για μετεγκατάσταση, δημιουργία και λειτουργία IT και επιτρέπει στους χρήστες να κάνουν αυτά τα πράγματα από τον υπάρχοντα φόρτο εργασίας της επιχείρησης έως τις νέες εγγενείς εφαρμογές cloud και πλατφόρμες δεδομένων. Αυτό το τμήμα της Oracle Cloud Platform υποστηρίζει εφαρμογές Oracle, προσαρμοσμένες εφαρμογές και εφαρμογές ISV. Μερικές από τις δυνατότητες του OCI περιλαμβάνουν:

- Τεχνητή νοημοσύνη (AI) και μηχανική μάθηση (ML)
- Αναλυτική Δεδομένων και Επιχειρησιακή Νοημοσύνη
- Μεγάλα Δεδομένα
- Περιοχές Νέφους
- Υπολογισμοί
- Συναρτήσεις και λειτουργίες
- Υπηρεσίες Βάσης Δεδομένων
- DevOps
- Κυβερνητικό σύννεφο
- Υβριδικό σύννεφο
- Ενσωμάτωση
- Δικτύωση
- Παρατήρηση και Διαχείριση

Μέρος του τμήματος εφαρμογών Oracle Cloud Platform είναι το Oracle Enterprise Resource Planning (ERP). Αυτή είναι μια πλατφόρμα που παρέχει στους χρήστες προηγμένες δυνατότητες όπως AI, αναλυτικά στοιχεία και αυτόματες ενημερώσεις. Αυτές οι λειτουργίες διευκολύνουν πολλές διαδικασίες για τους χρήστες, συμπεριλαμβανομένης της αυτοματοποίησης των χειροκίνητων διαδικασιών για εξοικονόμηση χρόνου, της προσαρμογής στις αλλαγές της αγοράς σε πραγματικό χρόνο και της απόκτησης ανταγωνιστικού πλεονεκτήματος μέσω της επιτάχυνσης της ροής εργασίας. Η πλατφόρμα περιλαμβάνει χαρακτηριστικά για:

- Χρηματοοικονομική Διαχείριση
- Προμήθειες
- Διαχείριση Έργου
- Διαχείριση Κινδύνου και Συμμόρφωση
- Διαχείριση Επιχειρηματικής Απόδοσης

Μια άλλη πτυχή των δυνατοτήτων των εφαρμογών Oracle Cloud Platform είναι η Oracle Supply Chain Management (SCM). Αυτό το σύνολο προϊόντων επιτρέπει στους οργανισμούς να ανταποκρίνονται γρήγορα στις μεταβαλλόμενες συνθήκες της αγοράς, τη ζήτηση και την προσφορά

συνδέοντας την αλυσίδα εφοδιασμού τους σε ένα δίκτυο που έχει δημιουργηθεί ειδικά για αυτούς. Αυτό το μέρος του προϊόντος περιλαμβάνει:

- Σχεδιασμό εφοδιαστικής αλυσίδας
- Διαχείριση αποθεμάτων
- Μεταποίηση
- Συντήρηση Διαχείρισης Αποθεμάτων

Ένα άλλο σύνολο προϊόντων από αυτήν τη λύση είναι αυτά στο Oracle Advertising and Customer Experience (CX). Αυτή είναι μια συνδεδεμένη σουίτα εφαρμογών που αναβαθμίζει τη βασική διαχείριση πελατειακών σχέσεων (CRM) και βοηθά τις εταιρείες να δημιουργούν, να διαχειρίζονται και να εξυπηρετούν αποτελεσματικές σχέσεις με τους πελάτες. Μέσω αυτών των προϊόντων, οι πελάτες της Oracle μπορούν να διαχειριστούν:

- Διαφήμιση
- Μάρκετινγκ
- Πωλήσεις
- Service
- Περιεχόμενο

Το τελικό σύνολο που ολοκληρώνει την Oracle Cloud Platform ονομάζεται Oracle Human Capital Management (HCM). Αυτή είναι μια ολοκληρωμένη λύση cloud που συνδέει κάθε διαδικασία ανθρώπινου δυναμικού στις επιχειρήσεις των πελατών. Αυτό επιτυγχάνεται μέσω της διαχείρισης:

- Ανθρώπινου Δυναμικού
- Διαχείριση Ταλέντων

5.1. Χαρακτηριστικά της πλατφόρμας Oracle Cloud

Αυτό το προϊόν έχει διάφορες δυνατότητες, σε όλες τις λειτουργίες που προσφέρει. Οι κύριες δυνατότητες της Oracle Cloud Platform μπορούν να ομαδοποιηθούν με τους εξής τρόπους:

- Διαθεσιμότητα: Μέσω της Oracle Cloud Platform, οι χρήστες μπορούν να επωφεληθούν από λειτουργίες για υψηλή διαθεσιμότητα δεδομένων, η οποία είναι ζωτικής σημασίας για εφαρμογές σε πραγματικό χρόνο. Αυτό διασφαλίζει την υψηλή απόδοση των εφαρμογών και ελαχιστοποιεί τους χρόνους διακοπής λειτουργίας και τις αστοχίες.
- Δημιουργία αντιγράφων ασφαλείας και ανάκτηση: Το προϊόν προσφέρει διάφορες δυνατότητες για τη δημιουργία αντιγράφων ασφαλείας και την ανάκτηση δεδομένων.
- Επεκτασιμότητα: Υπάρχουν διάφορες δυνατότητες στα προϊόντα Oracle που βελτιώνουν την απόδοση και κάνουν τις εφαρμογές πιο επεκτάσιμες.
- Ασφάλεια: Αυτή η πλατφόρμα θέτει την ασφάλεια ως μία από τις κορυφαίες προτεραιότητές της και την διασφαλίζει παρέχοντας αρκετούς μηχανισμούς για τον έλεγχο της πρόσβασης και της χρήσης των δεδομένων.

5.2. Πλεονεκτήματα της πλατφόρμας Oracle Cloud

Η Oracle Cloud Platform έχει πολλά οφέλη για τους χρήστες της σε διαφορετικούς κλάδους. Μερικά από τα μεγαλύτερα πλεονεκτήματα της χρήσης αυτής της λύσης περιλαμβάνουν:

- Χρησιμοποιεί διάφορες μεθοδολογίες και αρχές για την αύξηση της απόδοσης. Η λύση υποστηρίζει τη διαχείριση πολλαπλών βάσεων δεδομένων, επιτρέποντας μεγαλύτερη ευελιξία στους χρήστες της.
- Έχει διαφορετικές εκδόσεις, που καλύπτουν τις ανάγκες διαφορετικών πελατών, γεγονός που τους επιτρέπει να αγοράζουν μόνο αυτό που πραγματικά χρειάζονται.
- Η λύση εξασφαλίζει υψηλή διαθεσιμότητα δεδομένων χρησιμοποιώντας συμπλέγματα.
- Διαθέτει ένα ισχυρό σύστημα ανάκτησης αστοχιών που διασφαλίζει ότι τα αρχεία είναι ασφαλή ακόμα και σε περιόδους διακοπής λειτουργίας.

- Η λύση αποτελείται από πολλά προϊόντα τα οποία είναι οργανωμένα ανά περιπτώσεις χρήσης, επιτρέποντας στους χρήστες να έχουν εύκολη πρόσβαση σε αυτά και να λαμβάνουν τις πληροφορίες που χρειάζονται για την εργασία τους.
- Πελάτες από όλα τα επίπεδα, από προσωπικό έως επιχειρηματικό, μπορούν να χρησιμοποιήσουν τις δυνατότητες αυτής της πλατφόρμας, καθιστώντας την ιδιαίτερα προσβάσιμη.

5.3. Σύγκριση Oracle με άλλες πλατφόρμες:

Η Oracle Cloud Platform είναι μια από τις μεγαλύτερες πλατφόρμες υπολογιστικού νέφους στην αγορά, που προσφέρει ένα ευρύ φάσμα υπηρεσιών για επιχειρήσεις και οργανισμούς. Ακολουθεί μια σύγκριση μεταξύ της Oracle Cloud Platform και ορισμένων από τους σημαντικότερους ανταγωνιστές της:

5.3.1. Το Oracle Cloud Infrastructure σε σύγκριση με τις Υπηρεσίες Ιστού της

Amazon

Το Oracle Cloud Infrastructure έχει δημιουργηθεί για επιχειρήσεις που αναζητούν υψηλότερες επιδόσεις, χαμηλότερο κόστος και ευκολότερη μετάβαση στο σύννεφο για τις εφαρμογές τους. Οι πελάτες επιλέγουν το Oracle Cloud Infrastructure έναντι του AWS για διάφορους λόγους: Πρώτον, μπορούν να καταναλώνουν υπηρεσίες νέφους στο δημόσιο νέφος ή στο δικό τους κέντρο δεδομένων. Δεύτερον, μπορούν να μεταφερθούν και να εκτελέσουν οποιοδήποτε φόρτο εργασίας όπως συμβαίνει στο Oracle Cloud, συμπεριλαμβανομένων βάσεων δεδομένων και εφαρμογών Oracle, VMware ή διακομιστών. Τρίτον, οι πελάτες μπορούν εύκολα να εφαρμόσουν ελέγχους ασφαλείας και αυτοματισμούς για την πρόληψη σφαλμάτων λανθασμένης διαμόρφωσης και την εφαρμογή βέλτιστων πρακτικών ασφαλείας. Τέταρτον, έχουν χαμηλότερους κινδύνους με τα end-to-end SLA της Oracle που καλύπτουν την απόδοση, τη διαθεσιμότητα και τη διαχειριστικότητα των υπηρεσιών. Τέλος, ο φόρτος εργασίας τους επιτυγχάνει καλύτερη απόδοση με σημαντικά χαμηλότερο κόστος με το Oracle Cloud Infrastructure από το AWS. [17]

1. **Ανάπτυξη του ίδιου δημόσιου νέφους στις εγκαταστάσεις:** Οι επιχειρήσεις επιλέγουν υποδομή εσωτερικής εγκατάστασης για να συμβάλουν στην κάλυψη των ρυθμιστικών απαιτήσεων και των απαιτήσεων κυριαρχίας δεδομένων, να ελαχιστοποιήσουν τον λανθάνοντα χρόνο και να διασφαλίσουν τον τοπικό έλεγχο των πόρων. Θέλουν να αξιοποιήσουν τα χαρακτηριστικά και τα οφέλη του δημόσιου νέφους, ενώ τα συνδυάζουν με την απομόνωση και την ασφάλεια της εσωτερικής υποδομής. Η Oracle δίνει στους πελάτες την επιλογή να καταναλώνουν τις υπηρεσίες νέφους τους στο δημόσιο σύννεφο ή στο δικό τους κέντρο δεδομένων. Σε αντίθεση με τα AWS Outposts, η Oracle φέρνει το πλήρες χαρτοφυλάκιο των δημόσιων υπηρεσιών cloud στο κέντρο δεδομένων, ώστε να μειωθεί το κόστος, να αναβαθμιστούν παλαιού τύπου εφαρμογές με σύγχρονες υπηρεσίες και να αντιμετωπιστούν πιο απαιτητικές ανάγκες δεδομένων και καθυστερήσεων.
2. **Μεταφορά του φόρτου εργασίας της Oracle με σιγουριά:** Οι πελάτες επιλέγουν να μεταφέρουν τις επιχειρησιακές τους εφαρμογές και τις βάσεις δεδομένων της Oracle στο σύννεφο προκειμένου να επιτύχουν αποδοτικότητα και να βελτιώσουν την απόδοση και τη διαθεσιμότητα των εφαρμογών τους. Η Oracle αναγνωρίζει τη σημασία της απρόσκοπτης μετάβασης και παρέχει τη δυνατότητα μεταφοράς χωρίς αλλαγές στο λογισμικό ή τις διαδικασίες τους. Αντίθετα με το AWS, η Oracle διευκολύνει τη μετακίνηση των εφαρμογών. Το Oracle Cloud προσφέρει εξαιρετικές επιλογές ανάπτυξης, υψηλή απόδοση, βέλτιστη διαθεσιμότητα και χαμηλότερο κόστος για βάσεις δεδομένων και φορείς εργασίας της Oracle. Όσον αφορά τις επιχειρησιακές εφαρμογές όπως ERP, HCM, κλπ., η Oracle παρέχει στους πελάτες την επιλογή να μεταβούν σε ένα μοντέλο SaaS, κάτι που δεν είναι διαθέσιμο με το AWS, καθώς δεν παρέχει εφαρμογές SaaS για επιχειρήσεις.

3. **Μετακίνηση και εκτέλεση οποιουδήποτε φόρτου εργασίας ως έχει, συμπεριλαμβανομένων εκείνων που χρειάζονται γυμνό μέταλλο ή VMware:** Οι πελάτες μπορούν να εκτελούν φορτία εργασίας απευθείας σε φυσικούς διακομιστές γυμνού μετάλλου, όταν απαιτείται για λόγους απόδοσης, χρόνου απόκρισης ή ασφάλειας, και να απολαμβάνουν τις ίδιες επιλογές στο δημόσιο cloud. Πολλές επιχειρήσεις έχουν επίσης εγκατεστημένο το VMware Sphere στα κέντρα δεδομένων τους και επιθυμούν να μεταφέρουν τους φορτίους εργασίας που εκτελούνται σε αυτούς τους φυσικούς διακομιστές και περιβάλλοντα VMware, χωρίς να απαιτείται επανασχεδίαση εφαρμογών, αλλαγές λογισμικού ή διαδικασιών IT, ή εισαγωγή περαιτέρω πολυπλοκότητας. Η Oracle προσφέρει εύκολες λύσεις για αυτό, ενώ το AWS δεν παρέχει αντίστοιχες δυνατότητες.
4. **Απλοποίηση και αυτοματοποίηση της ασφάλειας με την προσέγγιση μηδενικής εμπιστοσύνης:** Το OCI έχει σχεδιαστεί με αρχιτεκτονική μηδενικής εμπιστοσύνης και δίνει έμφαση στην ασφάλεια. Αντίθετα από το AWS, παρέχει στους χρήστες εύχρηστους ελέγχους ασφαλείας και αυτοματισμούς για την πρόληψη σφαλμάτων εσφαλμένης διαμόρφωσης και την εφαρμογή βέλτιστων πρακτικών ασφαλείας. Μέσω ενός ποικίλου χαρτοφυλακίου υπηρεσιών ασφαλείας, όπως διαχείριση ταυτότητας και πρόσβασης, προστασία δεδομένων και κρυπτογράφηση, ενοποιημένοι έλεγχοι ασφαλείας, προστασία υποδομής, αναγνώριση περιστατικών, αποκατάσταση και υπηρεσίες συμμόρφωσης, οι χρήστες μπορούν να προστατεύσουν τον φόρτο εργασίας και τις εφαρμογές τους στο OCI.
5. **Μείωση του κινδύνου με τις μοναδικές SLA της Oracle:** Οι επιχειρήσεις αναζητούν περισσότερα από απλή διαθεσιμότητα στο cloud για την υποδομή τους. Επιπλέον, οι κρίσιμοι φόρτοι εργασίας απαιτούν σταθερή απόδοση και τη δυνατότητα διαχείρισης, παρακολούθησης και αλλαγής των πόρων που εκτελούνται στο cloud ανά πάσα στιγμή. Μόνο η Oracle προσφέρει ολοκληρωμένες συμφωνίες επιπέδου υπηρεσίας (SLA) που καλύπτουν την απόδοση, τη διαθεσιμότητα και τη δυνατότητα διαχείρισης των υπηρεσιών της.
6. **Αποφυγή εκπλήξεων κόστους AWS:** Μία από τις πιο συνηθισμένες ανησυχίες που αντιμετωπίζουν οι πελάτες της AWS είναι η απροσδόκητη χρέωση του AWS, όπου ένας πελάτης λαμβάνει έναν λογαριασμό και διαπιστώνει ότι χρωστά περισσότερα από ό,τι περίμενε - και δυσκολεύεται να κατανοήσει τον λόγο. Αυτό συμβαίνει επειδή η τιμολόγηση της AWS είναι πολύπλοκη, με μεγάλη κλίμακα και τιμωρεί τους πελάτες για δυσδιάκριτα στοιχεία χρήσης, όπως η απόδοση μεταφοράς δεδομένων και αποθήκευσης, που περιλαμβάνει παράγοντες όπως το μέγεθος του όγκου, οι προβλεπόμενες επιδόσεις IOP και η απόδοση για κάθε όγκο. Η AWS προσφέρει εκπτώσεις για τριετείς δεσμεύσεις σε προϊόντα όπως ο υπολογισμός, αλλά αυτές οι δεσμεύσεις είναι συνδεδεμένες με συγκεκριμένα σχήματα υπολογιστών, λειτουργικά συστήματα και περιοχές. Αν και η έκπτωση μπορεί να βοηθήσει να αποκτηθούν επιχειρηματικές ευκαιρίες, οι εξαρτήσεις από τις υπηρεσίες της AWS συνεπάγονται ότι δεν υπάρχει εγγύηση ότι η τιμολόγηση θα παραμείνει σε αυτά τα επίπεδα στο μέλλον. Η Oracle Cloud Infrastructure παρέχει καλύτερη απόδοση με σημαντικά χαμηλότερο κόστος σε σύγκριση με την AWS. Η δομή τιμολόγησής είναι απλή, προβλέψιμη και διαφανής, χωρίς καμία από τις αναπάντεχες εκπλήξεις που συνδέονται με την AWS. Επιπλέον, η OCI χρεώνει την ίδια σταθερή τιμή για όλες τις περιοχές σε ολόκληρο τον κόσμο, εξαλείφοντας κρυφές αποκλίσεις κόστους εάν μια επιχείρησή διευρύνεται παγκοσμίως. Η Oracle προσφέρει επίσης τα Universal Credits, τα οποία επιτρέπουν στους πελάτες να χρησιμοποιούν οποιαδήποτε υπηρεσία της Oracle Cloud Infrastructure ή πλατφόρμα σε οποιαδήποτε περιοχή με σημαντική εξοικονόμηση, με προβλέψιμη δέσμευση μηνιαίων δαπανών. Αλλά παρόλο που η δαπάνη είναι μηνιαία, η χρήση δεν περιορίζεται σε αυτό το διάστημα - αυτές οι πιστώσεις μπορούν να χρησιμοποιηθούν οποιαδήποτε στιγμή κατά τη διάρκεια των 12 μηνών συμβολαίου τους.

5.3.2 Το Oracle Cloud Infrastructure σε σύγκριση με το Google Cloud Platform

Η υποδομή Oracle Cloud συνδυάζει την ελαστικότητα και τη χρησιμότητα του δημόσιου νέφους με τον λεπτομερή έλεγχο, την ασφάλεια και την προβλεψιμότητα της εσωτερικής υποδομής. Οι πελάτες επιλέγουν το Oracle Cloud Infrastructure έναντι της πλατφόρμας Google Cloud για διάφορους λόγους: Πρώτον, οι πελάτες μπορούν να μετεγκαταστήσουν κάθε φόρτο εργασίας και να δημιουργήσουν ταχύτερα στο Oracle Cloud. Δεύτερον, οι πελάτες μπορούν να καταναλώνουν τις υπηρεσίες νέφους τους στο δημόσιο νέφος ή στο δικό τους κέντρο δεδομένων με το Oracle Dedicated Region Cloud Customer. Τρίτον, οι πελάτες μπορούν εύκολα να εφαρμόσουν βέλτιστες πρακτικές ασφάλειας, ελέγχους και αυτοματισμούς για την αποφυγή σφαλμάτων εσφαλμένης διαμόρφωσης. Τέταρτον, οι πελάτες έχουν χαμηλότερους κινδύνους με τις end-to-end SLA της Oracle που καλύπτουν την απόδοση, τη διαθεσιμότητα και τη διαχειρισσιμότητα των υπηρεσιών.

Τέλος, ο φόρτος εργασίας τους επιτυγχάνει καλύτερη απόδοση με σημαντικά χαμηλότερο κόστος με το Oracle Cloud Infrastructure [18]. Οι δυνατότητες του Oracle Cloud Infrastructure και το καθιστούν καλύτερη πλατφόρμα από το Google Cloud είναι:

1. **Μετακίνηση κάθε φόρτου εργασίας και γρήγορη δημιουργία στο cloud:** Οι πελάτες μετακινούν όλο και περισσότερες επιχειρηματικές εφαρμογές και βάσεις δεδομένων στο cloud για να επιτύχουν βελτιωμένη απόδοση και διαθεσιμότητα. Η Oracle Cloud Infrastructure προσφέρει μια απλή και αποτελεσματική λύση για αυτήν τη μετάβαση. Οι πελάτες μπορούν να μεταφέρουν τους φορτούς εργασίας τους από φυσικούς διακομιστές στο Oracle Cloud χωρίς να χρειάζεται να αλλάξουν την αρχιτεκτονική των εφαρμογών τους ή να εισάγουν πολυπλοκότητα. Επιπλέον, η Oracle προσφέρει υψηλή απόδοση και αξιοπιστία για βάσεις δεδομένων και φορτία εργασίας. Ειδικότερα, για επιχειρησιακές εφαρμογές όπως ERP, CRM και HCM, η Oracle παρέχει επιλογές SaaS που δεν είναι διαθέσιμες στο Google Cloud. Με την Oracle Cloud Infrastructure, οι πελάτες μπορούν να επωφεληθούν από τη συνολική λύση που προσφέρει υψηλή απόδοση, αξιοπιστία και ευελιξία σε συνδυασμό με ανταγωνιστικό κόστος.
2. **Ανάπτυξη ίδιας δημόσιας ισχύος νέφους στις εγκαταστάσεις:** Οι επιχειρήσεις επιλέγουν υποδομή εσωτερικής εγκατάστασης για να συμβάλουν στην κάλυψη των ρυθμιστικών απαιτήσεων και των απαιτήσεων κυριαρχίας δεδομένων, ελαχιστοποιώντας τον λανθάνοντα χρόνο και διασφαλίζοντας τον τοπικό έλεγχο των πόρων. Οι επιχειρήσεις επιθυμούν να αξιοποιήσουν τα χαρακτηριστικά και τα οφέλη του δημόσιου νέφους, συνδυάζοντάς τα με την απομόνωση και την ασφάλεια της εσωτερικής υποδομής. Με το Oracle Dedicated Region Cloud Customer, η Oracle δίνει στους πελάτες την επιλογή να καταναλώνουν τις υπηρεσίες νέφους είτε στο δημόσιο νέφος είτε στο δικό τους κέντρο δεδομένων. Σε αντίθεση με το Google Anthos, η Oracle φέρνει το πλήρες χαρτοφυλάκιο των δημόσιων υπηρεσιών νέφους στο κέντρο δεδομένων τους, επιτρέποντας τη μείωση του κόστους, την αναβάθμιση παλαιού τύπου εφαρμογών με σύγχρονες υπηρεσίες και την αντιμετώπιση των πιο απαιτητικών απαιτήσεων όσον αφορά την κυριαρχία δεδομένων και την καθυστέρηση. Το Dedicated Region Cloud Customer προσφέρει στους πελάτες το καλύτερο αποτέλεσμα.
3. **Απλοποίηση και αυτοματοποίηση της ασφάλειας με την προσέγγιση μηδενικής εμπιστοσύνης:** Το OCI έχει σχεδιαστεί με αρχιτεκτονική μηδενικής εμπιστοσύνης, με έμφαση στην ασφάλεια. Σε αντίθεση με το Google Cloud, το OCI παρέχει εύχρηστους ελέγχους ασφαλείας και αυτοματισμούς για την αποφυγή προβλημάτων από εσφαλμένη διαμόρφωση και την εφαρμογή βέλτιστων πρακτικών ασφαλείας. Οι χρήστες μπορούν να προστατεύσουν τον φόρτο εργασίας και τις εφαρμογές τους στο OCI μέσω ενός ποικίλου χαρτοφυλακίου υπηρεσιών ασφαλείας, περιλαμβανομένων διαχείρισης ταυτότητας και πρόσβασης, προστασίας δεδομένων και κρυπτογράφησης, ενοποιημένων ελέγχων ασφαλείας, προστασίας υποδομής, αναγνώρισης περιστατικών, αποκατάστασης και υπηρεσιών συμμόρφωσης.

4. **Μείωση του κινδύνου με τις μοναδικές SLA της Oracle:** Οι επιχειρήσεις απαιτούν «"Υπολογιστικό Νέφος και Αναλυτική Δεδομένων: Αξιοποίηση της Πλατφόρμας Oracle Cloud στην Επιχειρηματικότητα και Προβλεπτική Ανάλυση Δεδομένων"»

περισσότερα από τη διαθεσιμότητα από την υποδομή τους στο σύννεφο. Τα κρίσιμα φορτία εργασίας απαιτούν επίσης σταθερή απόδοση και τη δυνατότητα διαχείρισης, παρακολούθησης και τροποποίησης των πόρων που εκτελούνται στο cloud ανά πάσα στιγμή. Μόνο η Oracle προσφέρει end-to-end SLA που καλύπτουν την απόδοση, τη διαθεσιμότητα και τη διαχειρισσιμότητα των υπηρεσιών της.

5. **Βελτιστοποίηση των δαπανών μέσω της πλατφόρμας Google Cloud:** Το Oracle Cloud Infrastructure παρέχει καλύτερη απόδοση με σημαντικά χαμηλότερο κόστος από το Google Cloud. Η δομή τιμολόγησής του είναι απλή, προβλέψιμη και διαφανής, χωρίς καμία έκπληξη που σχετίζεται με το Google Cloud. Επιπλέον, η OCI χρεώνει την ίδια σταθερή τιμή για όλες τις περιοχές σε όλο τον κόσμο, εξαλείφοντας τυχόν κρυφές αποκλίσεις κόστους εάν η επιχείρηση επεκταθεί παγκοσμίως. Η Oracle παρέχει επίσης Universal Credits, τα οποία επιτρέπουν στους πελάτες να χρησιμοποιούν οποιαδήποτε υπηρεσία Oracle Cloud Infrastructure ή πλατφόρμα σε οποιαδήποτε περιοχή, με σημαντική εξοικονόμηση και προβλέψιμη μηνιαία δέσμευση δαπανών. Και παρόλο που η δαπάνη στο cloud είναι μηνιαία, οι πιστώσεις Universal Credits δεν λήγουν στο τέλος κάθε μήνα, αλλά μπορούν να χρησιμοποιηθούν οποιαδήποτε στιγμή κατά τη διάρκεια του 12μηνου του συμβολαίου.

5.3.2. Το Oracle Cloud Infrastructure σε σύγκριση με το Microsoft Azure

Το Oracle Cloud Infrastructure έχει δημιουργηθεί για επιχειρήσεις που αναζητούν υψηλότερες επιδόσεις, χαμηλότερο κόστος και ευκολότερη μετάβαση στο νέφος για τις εφαρμογές τους. Οι πελάτες επιλέγουν το Oracle Cloud Infrastructure έναντι του Microsoft Azure για διάφορους λόγους: Πρώτον, η Oracle προσφέρει στους πελάτες μια υπολογιστική πλατφόρμα υψηλής απόδοσης για τις εφαρμογές τους με την μεγαλύτερη ένταση πόρων και την απόδοση. Δεύτερον, οι πελάτες μπορούν να καταναλώνουν τις υπηρεσίες νέφους τους στο δημόσιο νέφος ή στο δικό τους κέντρο δεδομένων με το Oracle Dedicated Region Cloud Customer. Τρίτον, οι πελάτες μπορούν να μεταφέρουν και να εκτελέσουν οποιοδήποτε φόρτο εργασίας όπως συμβαίνει στο Oracle Cloud, συμπεριλαμβανομένων των βάσεων δεδομένων Oracle, του VMware ή των διακομιστών γυμνού μετάλλου. Τέταρτον, οι πελάτες μπορούν εύκολα να εφαρμόσουν βέλτιστες πρακτικές ασφάλειας, ελέγχους και αυτοματισμούς για την αποφυγή σφαλμάτων εσφαλμένης διαμόρφωσης. Τέλος, ο φόρτος εργασίας τους, επιτυγχάνει καλύτερη απόδοση με σημαντικά χαμηλότερο κόστος με το Oracle Cloud Infrastructure. Ρίξτε μια ματιά στο τι κάνει το Oracle Cloud Infrastructure καλύτερη πλατφόρμα νέφους από το Microsoft Azure.[19]

1. **Υπολογιστική πλατφόρμα cloud υψηλής απόδοσης:** Οι εταιρικοί πελάτες και ο πιο απαιτητικός φόρτος εργασίας τους έχουν ωθήσει την Oracle να υιοθετήσει μια διαφορετική προσέγγιση στον σχεδιασμό της πλατφόρμας νέφους της. Το Oracle Cloud Infrastructure συνδυάζει την απόδοση των εσωτερικά εγκατεστημένων περιβαλλόντων με την ελαστικότητα και το κόστος του υπολογιστικού νέφους που βασίζεται στην κατανάλωση. Η πλατφόρμα Oracle Cloud Infrastructure περιλαμβάνει στιγμιότυπα υπολογιστών γυμνού μετάλλου, συστήματα Exadata, ευέλικτες δομές υπολογισμού και αποθήκευσης, δίκτυα συμπλέγματος χαμηλής καθυστέρησης, λύσεις αποθήκευσης και συστήματα αρχείων υψηλής απόδοσης και απομόνωσης κυκλοφορίας δικτύου. Όλα αυτά έχουν κατασκευαστεί από την αρχή με στόχο τη βελτίωση της καθυστέρησης και τη μείωση της μεταβλητότητας απόδοσης για τις εφαρμογές των πελατών με τη μεγαλύτερη απαίτηση πόρων και απόδοσης.
2. **Ανάπτυξη του ίδιου δημοσίου νέφους στις εγκαταστάσεις:** Οι επιχειρήσεις επιλέγουν υποδομή εσωτερικής εγκατάστασης για να ανταποκριθούν στις ρυθμιστικές απαιτήσεις και τις απαιτήσεις κυριαρχίας δεδομένων, να ελαχιστοποιήσουν τον χρόνο αναμονής και να διασφαλίσουν τον τοπικό έλεγχο των πόρων. Αναζητούν την αξιοποίηση των χαρακτηριστικών και των οφελών του δημόσιου νέφους, ενώ ταυτόχρονα επιδιώκουν την απομόνωση και την ασφάλεια της εσωτερικής υποδομής. Η

Oracle προσφέρει στους πελάτες τη δυνατότητα να καταναλώνουν τις υπηρεσίες του cloud είτε στο δημόσιο νέφος είτε στο δικό τους κέντρο δεδομένων μέσω του Oracle Dedicated Region Cloud Customer. Σε αντίθεση με το Azure Stack Hub, η Oracle φέρνει το πλήρες χαρτοφυλάκιο των δημόσιων υπηρεσιών νέφους στο κέντρο δεδομένων του πελάτη, επιτρέποντάς του να μειώσει το κόστος, να αναβαθμίσει παλαιότερες εφαρμογές με σύγχρονες υπηρεσίες και να ανταποκριθεί στις πιο απαιτητικές απαιτήσεις κυριαρχίας δεδομένων και καθυστέρησης. Το Dedicated Region Cloud Customer προσφέρει στους πελάτες τον συνδυασμό των καλύτερων χαρακτηριστικών από τους δύο κόσμους.

3. **Μετακίνηση και εκτέλεση οποιουδήποτε φόρτου εργασίας ως έχει:** συμπεριλαμβανομένων εκείνων που χρειάζονται γυμνό μέταλλο ή VMwareOn-premises, οι φυσικοί διακομιστές επιτρέπουν στους πελάτες να εκτελούν φόρτους εργασίας απευθείας σε υλικό διακομιστή γυμνού μετάλλου όταν το απαιτεί η απόδοση, ο λανθάνοντας χρόνος ή η ασφάλεια και θέλουν τις ίδιες επιλογές στο δημόσιο νέφος. Οι περισσότερες επιχειρήσεις τρέχουν επίσης VMwareonSphere στα κέντρα δεδομένων τους. Θέλουν να μετεγκαταστήσουν τους φόρτους εργασίας που εκτελούνται σε αυτούς τους φυσικούς διακομιστές και Περιβάλλοντα VMware χωρίς επαναρχιτεκτονικές εφαρμογές, αλλάζουν οικείο λογισμικό ή διαδικασίες πληροφορικής ή εισάγουν πολυπλοκότητα στις λειτουργίες τους. Η Oracle το κάνει αυτό εύκολο ενώ το Azure όχι.
4. **Μεταφορά του φόρτου εργασίας της Oracle με σιγουριά:** Οι πελάτες μεταφέρουν ολοένα και περισσότερο κρίσιμες για τις επιχειρήσεις εταιρικές εφαρμογές και βάσεις δεδομένων στο cloud για να επιτύχουν αποτελεσματικότητα λειτουργίας και να βελτιώσουν την απόδοση και τη διαθεσιμότητα της εφαρμογής. Οι πελάτες θέλουν να εκτελέσουν αυτές τις μετεγκαταστάσεις χωρίς να αλλάξουν οικείο λογισμικό ή διαδικασίες. Σε αντίθεση με το Microsoft Azure, η Oracle διευκολύνει τις μετεγκαταστάσεις. Το Oracle Cloud παρέχει τις καλύτερες επιλογές ανάπτυξης, την υψηλότερη απόδοση, τη βέλτιστη διαθεσιμότητα και το χαμηλότερο κόστος για βάσεις δεδομένων και φόρτους εργασίας της Oracle. Στην περίπτωση των εταιρικών εφαρμογών όπως ERP, HCM, κ.λπ., η Oracle προσφέρει στους πελάτες τη δυνατότητα να φτάσουν σε ένα μοντέλο SaaS.
5. **Απλοποίηση και αυτοματοποίηση της ασφάλειας:** Το OCI έχει σχεδιαστεί με αρχιτεκτονική μηδενικής εμπιστοσύνης, με προτεραιότητα την ασφάλεια. Σε αντίθεση με το Azure, το OCI σάς παρέχει εύχρηστους ελέγχους ασφαλείας και αυτοματισμούς για την αποφυγή σφαλμάτων λανθασμένης διαμόρφωσης και την εφαρμογή βέλτιστων πρακτικών ασφαλείας. Οι φόρτοι εργασίας και οι εφαρμογές στο OCIμπορούν να ασφαλιστούν μέσω ενός ποικίλου χαρτοφυλακίου υπηρεσιών ασφαλείας, συμπεριλαμβανομένων των στοιχείων ταυτότητας και διαχείριση πρόσβασης, προστασία δεδομένων και κρυπτογράφηση, ενοποιημένοι έλεγχοι ασφαλείας, προστασία υποδομής, αναγνώριση περιστατικών, αποκατάσταση και υπηρεσίες συμμόρφωσης.
6. **Μείωση των κινδύνων με τις μοναδικές SLA της Oracle:** Οι επιχειρήσεις απαιτούν περισσότερα από τη διαθεσιμότητα από την υποδομή τους στο cloud. Οι κρίσιμοι φόρτοι εργασίας απαιτούν επίσης σταθερή απόδοση και τη δυνατότητα διαχείρισης, παρακολούθησης και τροποποίησης πόρων που εκτελούνται στο cloud ανά πάσα στιγμή. Μόνο η Oracle προσφέρει end-to-end SLA που καλύπτουν την απόδοση, τη διαθεσιμότητα και τη διαχειρισσιμότητα των υπηρεσιών.
7. **Βελτιστοποίηση του κόστους με εγγυημένη τιμολόγηση χωρίς κρυφές χρεώσεις:** Ένα από τα πιο συνηθισμένα ζητήματα που αντιμετωπίζουν οι πελάτες του Azure είναι η έκπληξη χρέωσης Azure, κατά την οποία ένας πελάτης λαμβάνει έναν λογαριασμό και διαπιστώνει ότι χρωστάει περισσότερα από όσα περίμενε - και δεν μπορεί εύκολα να καταλάβει γιατί. Συμβαίνει επειδή η τιμολόγηση του Azure είναι πολύπλοκη, δαπανηρή σε κλίμακα και τιμωρεί τους πελάτες για δύσκολα εκτιμήσιμα στοιχεία χρήσης, όπως η απόδοση μεταφοράς δεδομένων και αποθήκευσης, η οποία περιλαμβάνει τη λήψη

υπόψη του μεγέθους όγκου, των παρεχόμενων IOP και της απόδοσης για τόμους όπως το Azure Ultra Disk. Το Oracle Cloud Infrastructure παρέχει καλύτερη απόδοση με σημαντικά χαμηλότερο κόστος από το Azure. Η δομή τιμολόγησης μας είναι απλή, προβλέψιμη και διαφανής, χωρίς καμία έκπληξη να σχετίζεται με το Azure. Επιπλέον, η OCI χρεώνει την ίδια σταθερή τιμή για όλες τις περιοχές σε όλο τον κόσμο, πράγμα που σημαίνει ότι δεν υπάρχουν κρυφές αποκλίσεις κόστους εάν η επιχείρησή σας επεκταθεί παγκοσμίως. Η Oracle παρέχει επίσης Universal Credits, οι οποίες επιτρέπουν στους πελάτες να χρησιμοποιούν οποιαδήποτε υπηρεσία Oracle Cloud Infrastructure ή πλατφόρμα σε οποιαδήποτε περιοχή με σημαντική εξοικονόμηση, με προβλέψιμη δέσμευση μηνιαίων δαπανών. Αλλά μόνο και μόνο επειδή η δαπάνη σας είναι μηνιαία, δεν σημαίνει ότι η χρήση σας πρέπει να είναι: αυτές οι πιστώσεις δεν λήγουν στο τέλος του μήνα — μπορούν να χρησιμοποιηθούν ανά πάσα στιγμή κατά τη διάρκεια του συμβολαίου τους.

6. Σενάριο Ανάλυσης δεδομένων με την πλατφόρμα Oracle Cloud

Η παρούσα μεταπτυχιακή διατριβή επικεντρώνεται στην ανάλυση δεδομένων μέσω της πλατφόρμας Oracle Cloud. Η ανάλυση δεδομένων αποτελεί ένα κρίσιμο εργαλείο για τις επιχειρήσεις, καθώς παρέχει πολύτιμες πληροφορίες και εισαγωγές για τη λήψη αποφάσεων. Η χρήση του cloud computing σε συνδυασμό με την αναλυτική δεδομένων αποτελεί μια ελκυστική επιλογή για τις επιχειρήσεις, καθώς παρέχει ευελιξία, κλιμακική κλίμακα και αποθήκευση των δεδομένων σε ασφαλές περιβάλλον.

Στο πλαίσιο αυτής της διατριβής είναι να παρουσιαστεί ένα σενάριο ανάλυσης δεδομένων με τη χρήση της πλατφόρμας Oracle Cloud. Αρχικά, πραγματοποιείται η δημιουργία ενός συνόλου δεδομένων βασισμένο σε ένα αρχείο. Έπειτα, εκτελείται μια απλή ανάλυση στο σύνολο δεδομένων προκειμένου να εξαχθούν πρώτες εισαγωγές και συμπεράσματα. Ακολουθεί η δημιουργία μιας πρώτης ροής δεδομένων και η ανάλυση των αποτελεσμάτων που προκύπτουν από αυτήν. Στη συνέχεια, πραγματοποιείται η δημιουργία μιας δεύτερης ροής δεδομένων με διακλαδιζόμενο κόμβο, προσφέροντας μια επιπλέον πτυχή ανάλυσης. Στο επόμενο στάδιο, παρουσιάζεται η δημιουργία και εκτέλεση μιας ακολουθίας ροών δεδομένων, καθώς και η ανάλυση των παραγόμενων συνόλων δεδομένων. Τέλος, πραγματοποιείται η εξαγωγή και εισαγωγή δεδομένων ροής, συμπληρώνοντας έτσι το σενάριο ανάλυσης δεδομένων μέσω της πλατφόρμας Oracle Cloud.

6.1. Δημιουργία ροών δεδομένων

1. Δημιουργία συνόλου δεδομένων με βάση το αρχείο.
2. Εκτέλεση απλής ανάλυσης στο σύνολο δεδομένων
3. Δημιουργία πρώτης ροής δεδομένων
4. Ανάλυση της εξόδου (output) της πρώτης ροής δεδομένων
5. Δημιουργία δεύτερης ροής δεδομένων με διακλαδιζόμενο κόμβο
6. Δημιουργία και εκτέλεση μιας Ακολουθίας ροών δεδομένων
7. Ανάλυση των συνόλων δεδομένων που προκύπτουν
8. Εξαγωγή και εισαγωγή δεδομένων Ροής

Στο παρακάτω σενάριο, δημιουργούμε μια ροή δεδομένων με ένα αρχείο υπολογιστικού φύλλου (excel) που περιέχει δείγματα δεδομένων σχολικής δωρεάς (donation.xlsx) και ένα αρχείο δεδομένων στατιστικών στοιχείων ταχυδρομικού κώδικα (zip_stats.xlsx). Στη ροή δεδομένων, γίνεται τροποποίηση των συνόλων δεδομένων των δωρεών φιλτράροντας δεδομένα σε μια στήλη και στη συνέχεια γίνεται προσθήκη νέων στηλών. Από τα δύο σύνολα δεδομένων, δημιουργείται ένα επιμελημένο σύνολο δεδομένων που μπορεί να χρησιμοποιηθεί σε αναλύσεις και δημιουργία γραφημάτων.

Το Donations.xlsx περιέχει σύνολα δεδομένων δωρεών σχολικών έργων καθώς έχει 150.000 σειρές πληροφοριών σχετικών με έργα σχολικής δωρεάς κατανεμημένες σε 5 χρόνια.

PROJECTID	TEACHER_A	SCH_LATITU	SCH_LONGI	SCH_CITY	SCH_STATEZ	SCH_COUNT	SCH_CHART	TCHR_TEAC	PRIMARY_F	SECONDARY	SECONDARY	POVERTY_L	GRADE_LEV	STUDENTS	TOTAL_DON	NUM_DON	ACT_MATC	ELIGIBLE_D				
P100000	c613d9f2d f5a61953a	35,17894	-80,7679	Charlotte	NC-28212	urban	Mecklenbu	f	Mrs.	t	Health & L	Math & Sc	Literacy	Literacy & Technolog	D: Highest	C: Grades	200	844,81	6	t	co	
P100001	19a38092f 525e2e3dc	35,13882	-89,9012	Memphis	TN-38117	urban	Shelby	f	Mrs.	f	Literacy	Literacy & Language	Books	D: Highest	A: Grades	20	180,37	7	t	co		
P100003	5b3fbc26 57a2826ac	41,93702	-72,795	Granby	CT-6035	suburban	Hartford	f	Ms.	f	Literacy	Literacy & Language	Other	A: Low po	A: Grades	22	0	2	t	ex		
P100005	26f737c80 64c5221dc	29,87183	-95,3363	Houston	TX-77093	urban	Harris	f	Ms.	f	Mathemat Math & Sc	Early Deve	Applied Le	Supplies	D: Highest	A: Grades	25	368,18	2	t	co	
P100006	fdb490ff: 3f465c777	28,07272	-82,6316	Tampa	FL-33626	suburban	Hillsborou	f	Mrs.	f	Literacy	Literacy & Language	Supplies	A: Low po	A: Grades	18	254,92	7	t	co		
P100015	70a7954e: f6efe0b57	35,03216	-106,709	Albuquerque	NM-87105	suburban	Bernalillo	f	Mrs.	f	Early Deve	Applied Le	Health & V	Health & S	Supplies	D: Highest	A: Grades	22	511,89	13	f	co
P100019	fed6745ae 89f101b7c	35,01014	-90,0139	Memphis	TN-38116	urban	Shelby	f	Ms.	f	College & i	Applied Le	Mathemat Math & Sc	Supplies	D: Highest	C: Grades	140	559,92	36	t	co	
P100024	4ea38e80: cebfbc783	25,89163	-97,44	Brownsville	TX-78521	urban	Cameron	f	Mrs.	f	Literacy	Literacy & Language	Technolog	C: High po	B: Grades	600	481,59	2	t	co		
P10003	53f9b103 1842e4f5a	46,14432	-122,935	Longview	WA-98632	urban	Cowlitz	f	Mr.	f	Special Ne	Special Ne	Literacy	Literacy & Books	C: High po	D: Grades	15	861,34	4	f	co	
P10004	bfcba8dc 667b31bbf	27,96603	-82,2755	Warren	MI-48091	urban	Macomb	f	Mrs.	f	Literacy	Literacy & ESL	Literacy & Technolog	C: High po	B: Grades	22	0	0	f	ex		
P100048	876e08cd 667b31bbf	27,96603	-82,2755	Seffner	FL-33584	suburban	Hillsborou	f	Ms.	f	Special Ne	Special Ne	Literacy	Literacy & Supplies	C: High po	B: Grades	14	0	3	t	ex	
P100049	27a278ff9 d700e50d5	36,18383	-115,221	Las Vegas	NV-89108	urban	Clark	f	Mrs.	f	Literacy	Literacy & Visual Arts	Music & TI	Technolog	D: Highest	A: Grades	20	1488,75	3	t	co	
P100053	de2163a13 3277565a5	36,92722	-121,752	Watsonville	CA-95076	urban	Santa Cruz	f	Ms.	t	Special Ne	Special Ne	Literature	Literacy & Supplies	D: Highest	B: Grades	45	0	0	f	ex	
P100057	bc3fa226f 98986c041	40,17211	-79,8657	Donora	PA-15033	suburban	Washington	f	Ms.	f	Mathemat Math & Sc	History & (History & (Technolog	D: Highest	B: Grades	180	485,15	9	f	co	
P100059	bc415090f 7f6f0c30	34,03297	-117,798	Diamond E	CA-91765	suburban	Los Angele	f	Mrs.	f	Health & L	Math & Sc	Environme	Math & Sc	Supplies	B: Modera	A: Grades	24	0	1	t	ex
P100064	c08ba26c 1542be34f	31,97775	-81,1057	Savannah	GA-31406	urban	Chatham	f	Mrs.	f	Applied Sci	Math & Science	Supplies	B: Modera	B: Grades	32	0	0	t	ex		
P100066	e8f3f0f6c 667b31bbf	27,96603	-78,3672	Strasburg	VA-22657	rural	Shenandoa	f	Ms.	f	Mathemat Math & Science	Technolog	B: Modera	C: Grades	65	365,75	2	t	co			
P100068	e655d1c9: 111337ccf	38,8314	-90,3135	Florissant	MO-63031	suburban	St Louis	f	Mrs.	f	Visual Arts	Music & The Arts	Other	C: High po	D: Grades	500	286,72	3	t	co		
P100069	7ae054f8e 24bc55b9f	37,00798	-121,585	Gilroy	CA-95020	suburban	Santa Clar	f	Mrs.	f	Sports	Health & S	Gym & Fitr	Health & S	Supplies	D: Highest	A: Grades	40	0	0	f	ex
P10007	6d869a61: e1fe8481e	35,98739	-86,5077	Smyrna	TN-37167	suburban	Rutherford	f	Mr.	f	Literacy	Literacy & Language	Books	D: Highest	B: Grades	20	0	0	f	ex		
P100071	63d0576ef f41c88fab	33,58027	-81,6901	Aiken	SC-29801	suburban	Aiken	f	Mrs.	f	Literacy	Literacy & Literature	Literacy & Supplies	D: Highest	A: Grades	16	549,07	10	f	co		
P100073	4e4b6d32f b98cb74f6	40,29669	-109,996	Roosevelt	UT-84066	rural	Duchesne	f	Mrs.	f	Performing Music & TI	Music	Music & TI	Technolog	C: High po	C: Grades	130	0	4	f	ex	
P100078	355ac6cd7 84ad2745c	33,92292	-118,274	Los Angele	CA-90061	urban	Los Angele	f	Mr.	f	Foreign La	Literacy & Performing Music & TI	Technolog	D: Highest	D: Grades	87	938,92	28	f	co		
P100083	b715f4fd8 c586eed7d	35,24248	-81,1282	Gastonia	NC-28054	urban	Gaston	f	Mrs.	f	Literacy	Literacy & ESL	Literacy & Books	C: High po	D: Grades	50	395,21	4	t	co		
P100101	3d311d63: 3277f9e6f	33,99029	-118,226	Huntingtor	CA-90255	suburban	Los Angele	t	Ms.	f	Literacy	Literacy & Literature	Literacy & Technolog	C: High po	D: Grades	329	228,11	3	f	co		
P100102	bc7f71a19 3927f9e6f	33,99029	-118,226	Huntingtor	CA-90255	suburban	Los Angele	t	Mr.	f	Foreign La	Literacy & Language	Books	C: High po	D: Grades	130	0	0	f	ex		
P100105	161e6037: 70a4c107f	36,14952	-115,194	Las Vegas	NV-89102	urban	Clark	f	Ms.	t	Literacy	Literacy & ESL	Literacy & Books	D: Highest	B: Grades	45	1058,13	5	t	co		
P100107	511d524cc d6b2e0cc	33,00438	-117,116	San Diego	CA-92127	rural	San Diego	f	Mrs.	f	Mathemat Math & Science	Technolog	A: Low po	B: Grades	35	339,06	2	t	co			
P100110	84a720ccc 8c91b3d7e	40,05679	-75,1609	Philadelph	PA-19138	urban	Philadelph	f	Ms.	t	Special Ne	Special Ne	Literacy	Literacy & Technolog	D: Highest	D: Grades	20	532,71	7	f	co	
P100111	d621b6ad: 304b4b63f	38,67794	-121,163	Folsom	CA-95630	urban	Sacrament	f	Ms.	f	Special Ne	Special Ne	Applied Sci	Math & Sc	Technolog	C: High po	B: Grades	32	0	0	f	ex
P100113	7a14876d: 3a0ac6f7d	41,82040	-71,4504	Providence	RI-02008	urban	Providence	f	Mr.	f	Mathemat Math & Sc	Literacy	Literacy & Supplies	D: Highest	A: Grades	26	302,29	4	t	co		

Εικόνα 4: Μέρος του αρχείου Donations.xlsx με ένα δείγμα των γραμμών και στηλών του.

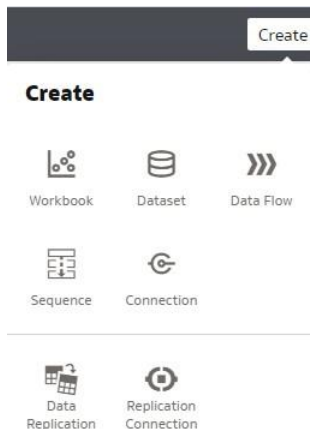
Το Zip_Stats.xlsx είναι το σύνολο δεδομένων στατιστικών και έχει περίπου 32 χιλιάδες σειρές πληροφοριών εισοδήματος και πληθυσμού σε επίπεδο ταχυδρομικού κώδικα.

	A	B	C	D	E	F	G
1	Zip	Median	Mean	Population			
2	39358	33	54	3			
3	79848	71	83	2			
4	83127	91	95	5			
5	84534	258	342	16			
6	32531	332	413	40			
7	54470	323	432	16			
8	79356	316	439	60			
9	62433	346	480	10			
10	79529	331	483	14			
11	81641	451	505	35			
12	67838	519	612	9			
13	79843	447	632	42			
14	2568	516	691	31			
15	6752	450	693	8			
16	3256	585	695	16			
17	49309	551	729	31			
18	76055	587	745	38			
19	15870	744	820	23			
20	48897	754	827	21			
21	79528	846	961	10			
22	13740	841	1.022	10			
23	48891	893	1.104	78			
24	38337	910	1.241	29			
25	35470	864	1.253	216			
26	27305	1.018	1.296	33			
27	84063	1.038	1.347	6			
28	79248	991	1.451	24			
29	49884	1.079	1.487	28			
30	29046	1.186	1.493	10			
31	10464	1.520	1.673	109			
32	98621	1.452	1.763	9			
33	13459	1.575	1.767	68			
34	28310	1.799	1.785	237			

Εικόνα 5: Μέρος του αρχείου Zip_Stats.xlsx

6.2. Δημιουργία συνόλου δεδομένων βάσει αρχείου

Αρχικά δημιουργούμε τα σύνολα δεδομένων που απαιτούνται για αυτό το σενάριο. Θα χρησιμοποιήσουμε αυτά τα σύνολα δεδομένων σε περαιτέρω βήματα. Δημιουργούμε ένα σύνολο δεδομένων και κάνουμε φόρτωση το αρχείο Donations.xlsx.



Εικόνα 6: Επιλογές για δημιουργία νέου έργου από το μενού της πλατφόρμας της Oracle.

Το γεωγραφικό πλάτος (Latitude) και το γεωγραφικό μήκος (Longitude) έχουν επισημανθεί ως μετρήσεις καθώς είναι αριθμοί. Κάνουμε την αλλαγή την αντιμετώπιση των δύο αυτών στηλών από μετρήσεις σε χαρακτηριστικά.

SCH_LONGITUDE

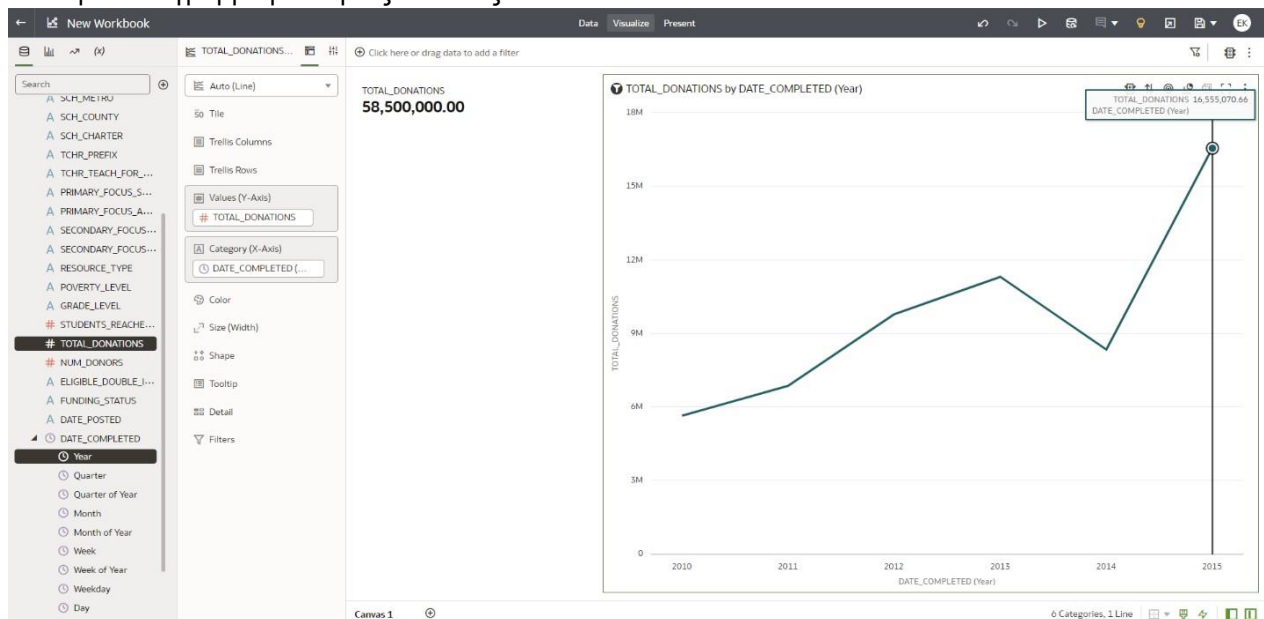
SCH_LATITUDE

Treat as: από metrics σε attributes

Δημιουργώντας ένα Project για το συγκεκριμένο σύνολο δεδομένων που μόλις δημιουργήσαμε, μπορούμε να κάνουμε μερικές αναλύσεις.

6.3. Διαγράμματα

A. Χρονοδιάγραμμα με δωρεές ανά έτος.



Εικόνα 7: Χρονοδιάγραμμα με τις συνολικές δωρεές ανά έτος ολοκλήρωσης

Η μεγαλύτερη Δωρεά έγινε το 2015 με συνολικά 16.555.070,66 \$

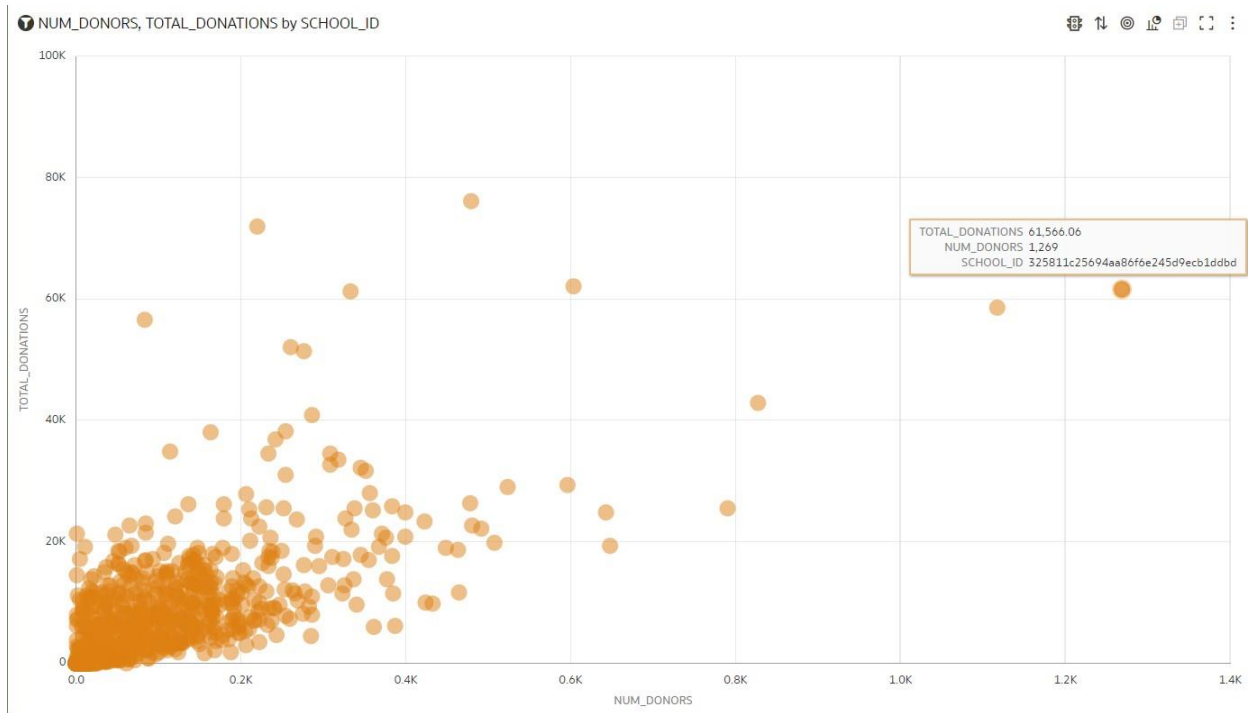
Β. Ας δημιουργήσουμε τώρα ένα scatter που δείχνει όλες τις πόλεις για την αξία της δωρεάς τους και τον αντίστοιχο αριθμό δωρητών.

Παρατηρούμε ότι στο Λος Άντζελες παρουσιάζεται το μεγαλύτερο ποσό δωρητών (39.769) αλλά και το μεγαλύτερο ποσό δωρεών (2.326.812,73 \$).



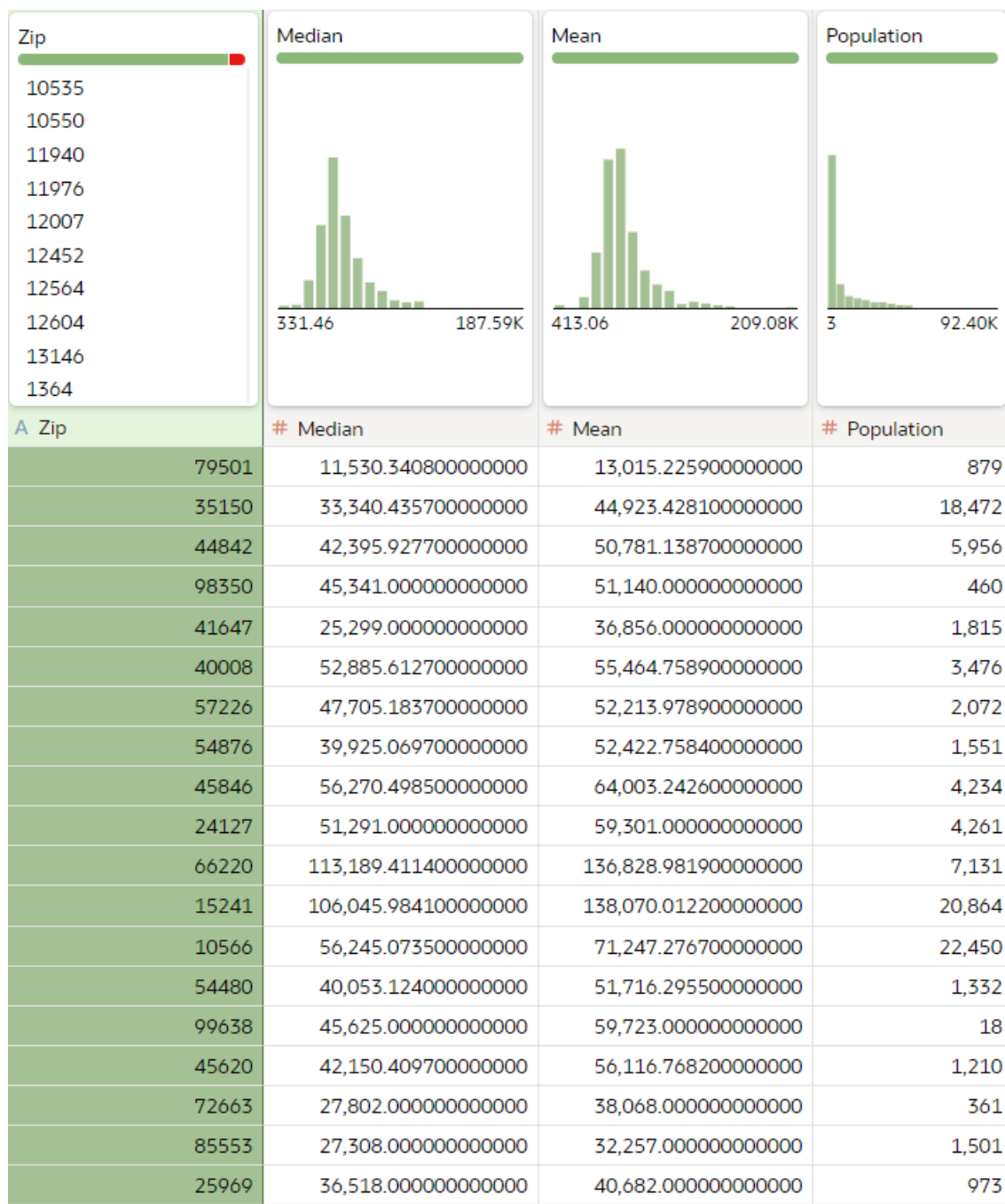
Εικόνα 8-9: Scatter Plot συνολικού αριθμού δωρεών ανά αριθμό δωρητών για κάθε πολιτεία.

Το Σικάγο είναι η δεύτερη στην σειρά πόλη με τον μεγαλύτερο αριθμό δωρητών (23.080) με το συνολικό ποσό που χορηγήθηκε να ανέρχεται στα 2.001.690,73\$.



Εικόνα 10: Scatter Plot συνολικού αριθμού των δωρητών και του συνολικού ποσού δωρεάς ανά σχολείο.

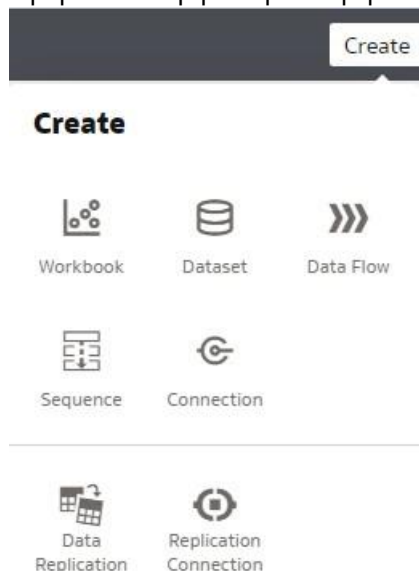
Το Zip (Ταχυδρομικός Κώδικας) έχει οριστεί ως μέτρηση από προεπιλογή καθώς είναι αριθμός. Θα αλλάξει και αυτό σε χαρακτηριστικό (attribute). Το συγκεκριμένο σύνολο δεδομένων αποτελείται από την στήλη του ταχυδρομικού κώδικα και του πληθυσμού αλλά και περιέχει δύο ακόμη στήλες που έχουν μετρήσεις σχετικές με τον πληθυσμό όπως του Median και του Mean.



Εικόνα 11: Αναπαράσταση του συνόλου δεδομένων Ταχυδρομικού κώδικα.

Στη συνέχεια θα ενώσουμε αυτά τα δύο σύνολα δεδομένων και να δούμε τη μέση σχολική δωρεά ανά εισόδημα, τη δωρεά ανά πληθυσμό και άλλες αναλύσεις. Για να τα επιτύχουμε αυτά, θα μπορούσαμε να βασίσουμε τους υπολογισμούς στο έργο ή μια άλλη επιλογή είναι να

δημιουργήσουμε μια ροή δεδομένων για να εξαγάγουμε ένα νέο σύνολο δεδομένων που θα μας βοηθήσει να συγκρίνουμε διαφορετικά σχολεία με διάφορες μετρήσεις.



Εικόνα 12: Μενού δημιουργίας νέων έργων μεταξύ αυτών είναι η δημιουργία συνόλου δεδομένων, Διαγράμματος Ροής, Ακολουθίας, Σύνδεσης

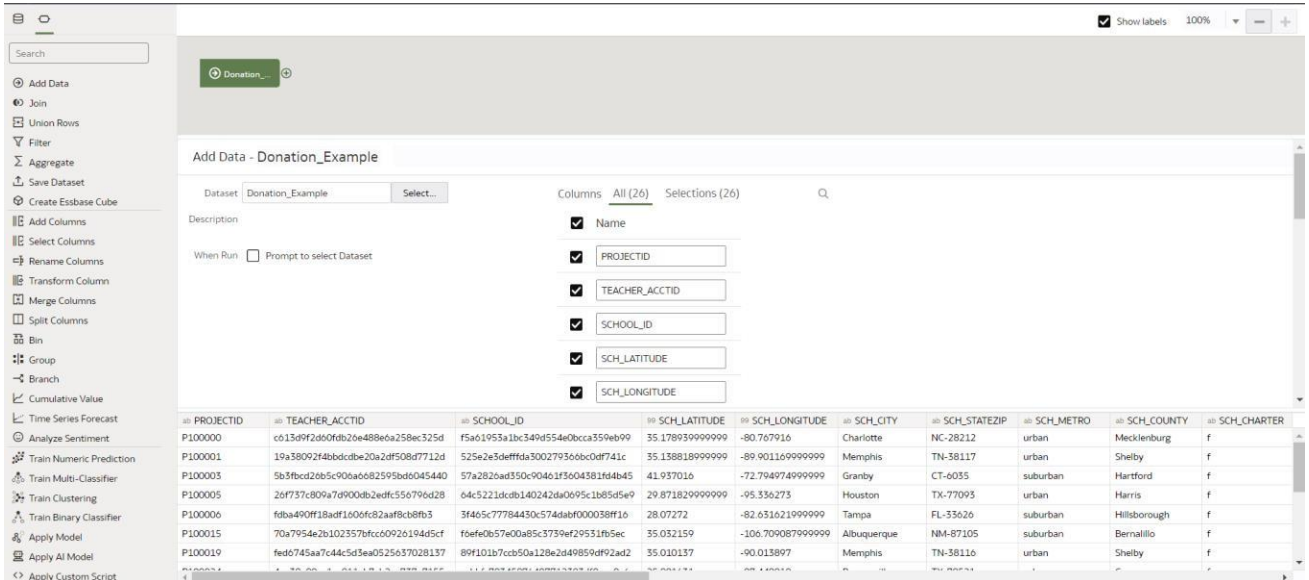
6.4. Διάγραμμα Ροής Δεδομένων (Data-FlowDiagram)

Ένα Διάγραμμα Ροής Δεδομένων (Data Flow Diagram DFD) είναι ένας τρόπος αναπαράστασης της ροής των δεδομένων μέσω μιας διαδικασίας ή ενός συστήματος, συνήθως ενός συστήματος πληροφοριών. Το DFD παρέχει επίσης πληροφορίες σχετικά με τις εισόδους και εξόδους κάθε οντότητας και την ίδια τη διαδικασία. Αντίθετα με άλλα διαγράμματα, ένα DFD δεν περιλαμβάνει ροή ελέγχου, όπως αποφάσεις και βρόχους. Αντί αυτού, επικεντρώνεται στην αναπαράσταση λειτουργιών που βασίζονται στα δεδομένα μέσω ενός διαγράμματος ροής. [8]

Για κάθε ροή δεδομένων, τουλάχιστον ένα από τα τελικά σημεία (πηγή ή/και προορισμός) πρέπει να ανήκει σε μια διαδικασία. Μπορούμε να αναπαραστήσουμε μια πιο λεπτομερή διαδικασία μέσω ενός διαγράμματος ροής δεδομένων που διαιρεί την αρχική διαδικασία σε μικρότερες υποδιαδικασίες.

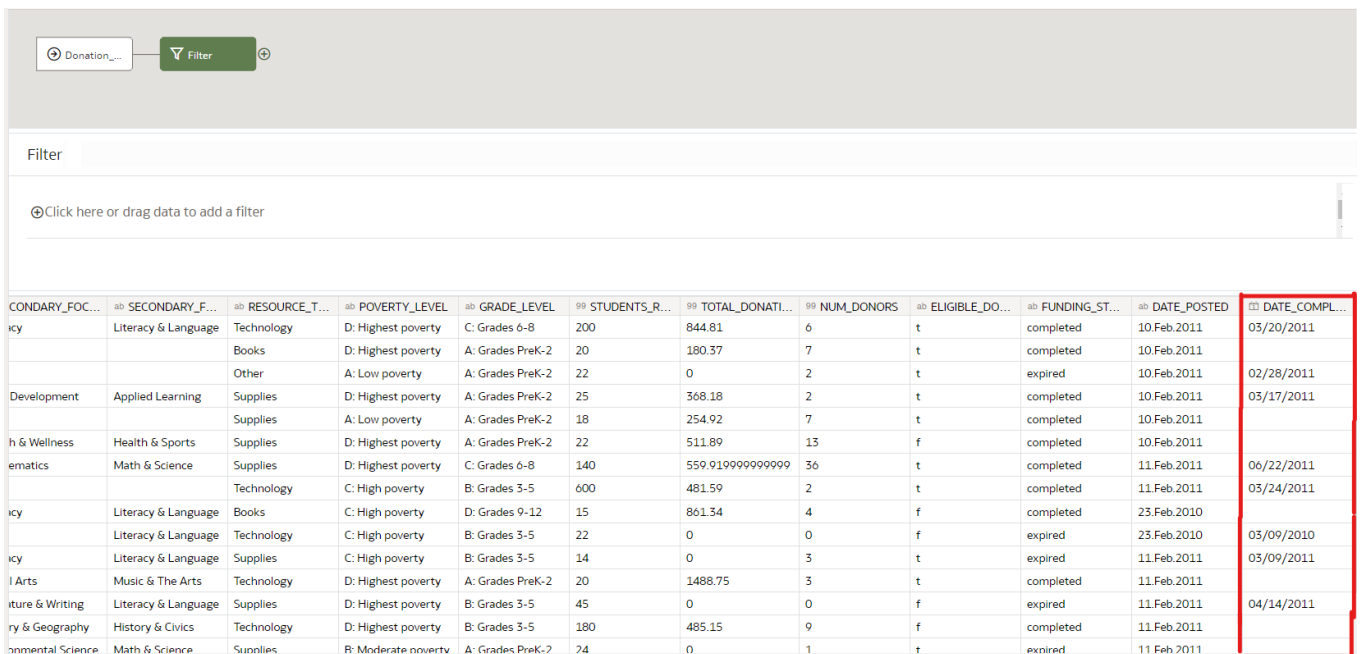
Ένα διάγραμμα ροής δεδομένων είναι ένα εργαλείο που χρησιμοποιείται στη δομημένη ανάλυση και μοντελοποίηση δεδομένων. Στο πλαίσιο του UML, συνήθως χρησιμοποιείται το διάγραμμα δραστηριότητας για την αναπαράσταση της ροής δεδομένων. Ένα ειδικό είδος διαγράμματος ροής δεδομένων είναι το τοποθεσίας-προσανατολισμένο διάγραμμα ροής δεδομένων.[9]

6.5. Δημιουργία Data Flow



Εικόνα 13: Δημιουργία αρχικού κόμβου της ροής δεδομένων.

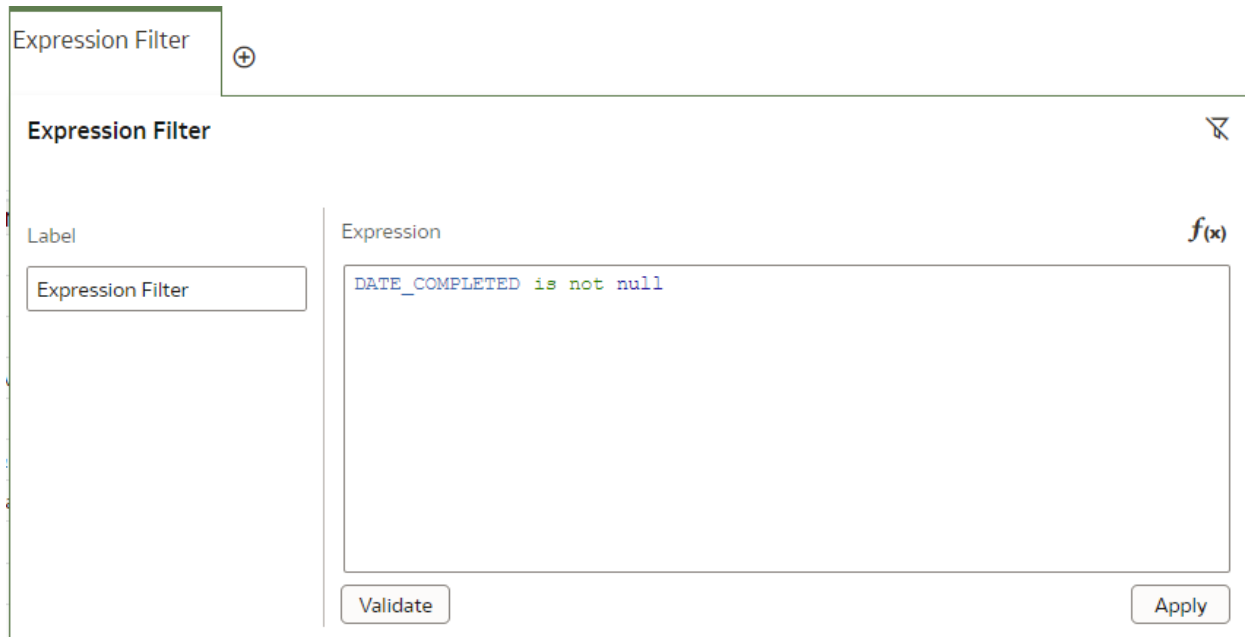
Στο Donations.xlsx παρατηρούμε ότι η στήλη Date_Completed περιέχει κενές τιμές για κάποιες από τις εγγραφές του αρχείου. Έτσι μπορούμε να δούμε ορισμένα έργα που βρίσκονται σε εξέλιξη των οποίων η ημερομηνία ολοκλήρωσης είναι μηδενική. Γι' αυτό φιλτράρουμε αυτές τις εγγραφές.



Εικόνα 14: Μέρος του συνόλου δεδομένων με εστίαση στην στήλη Ημερομηνίας Ολοκλήρωσης.

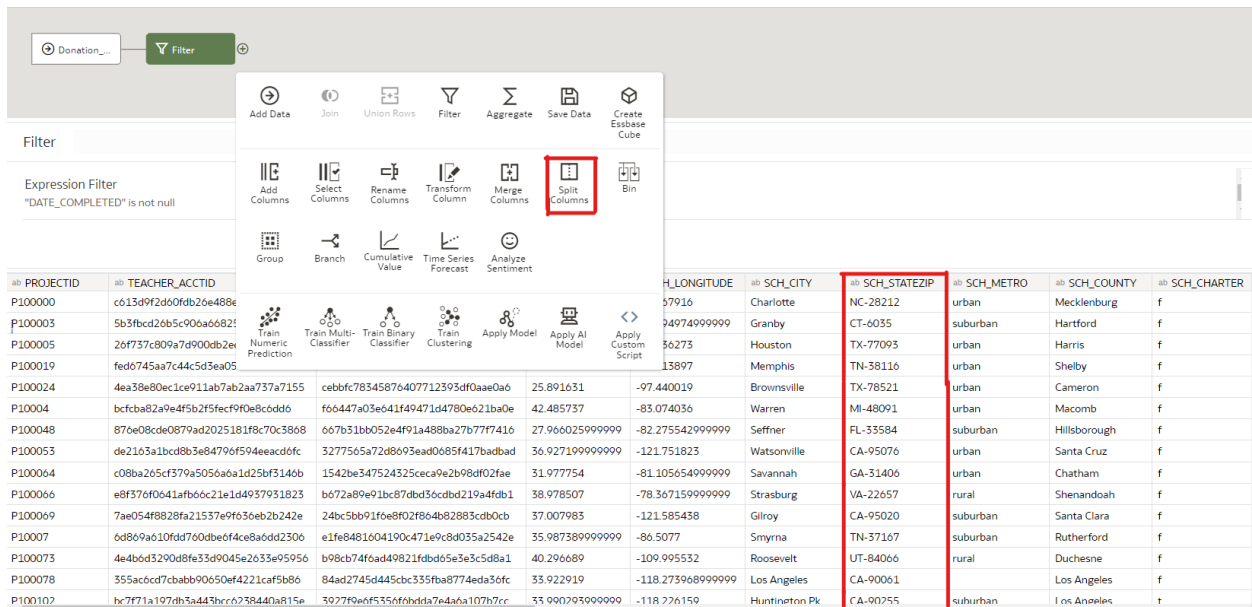
Στο βήμα φίλτρου, έχουμε την επιλογή είτε να σύρουμε/αποθέσουμε μια στήλη και να φιλτράρουμε τις τιμές, είτε να χρησιμοποιήσουμε την επιλογή φίλτρου έκφρασης εάν επιθυμούμε να φιλτράρουμε τις τιμές με βάση μια συνθήκη.

Έπειτα θα χρησιμοποιήσουμε ένα φίλτρο έκφρασης για να αφαιρέσουμε τις τιμές NULL στη στήλη ημερομηνία. Για να το επιτύχουμε αυτό προσθέτουμε στο Dataflow ένα πεδίο φίλτρου με έκφραση τέτοια ώστε οι ημερομηνίες ολοκλήρωσης του έργου να μην είναι κενή.



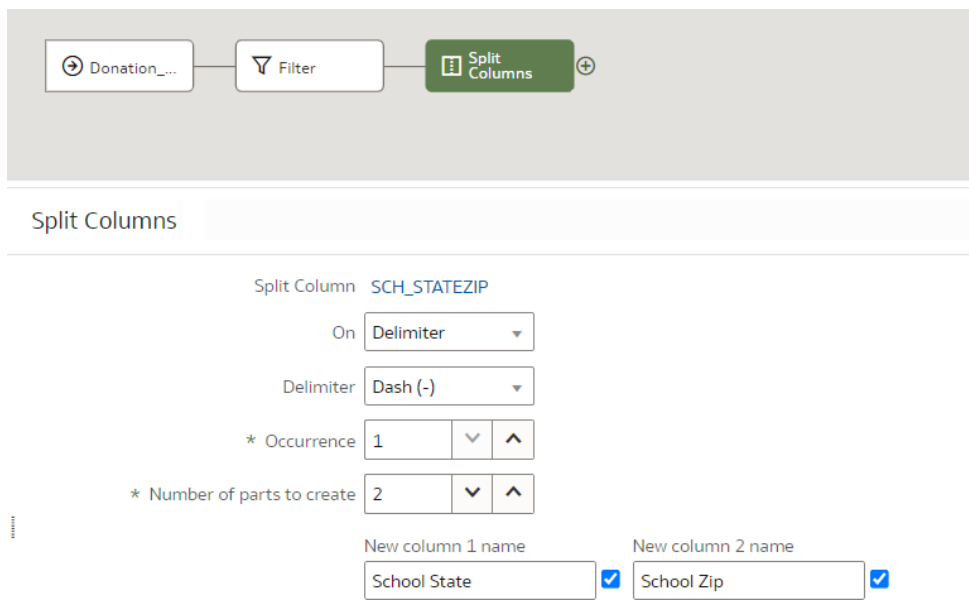
Εικόνα 15: Πεδίο φίλτρου με την χρήση έκφρασης / ελέγχου σε επιλεγμένη στήλη Ημερομηνία Ολοκλήρωσης.

Αν παρατηρήσουμε τα δυο αυτά dataset, θα δούμε ότι το σύνολο δεδομένων του Donation έχει την κατάσταση του σχολείου και τον ταχυδρομικό κώδικα ως μέρος της ίδιας στήλης. Στο σύνολο δεδομένων Zip_Stats, υπάρχει η στήλη του ταχυδρομικού κώδικα μαζί με τις υπόλοιπες πληροφορίες. Γι' αυτό θα χρησιμοποιήσουμε το βήμα διαχωρισμένων στηλών για να εξαγάγουμε πληροφορίες κατάστασης και Ταχυδρομικού κώδικα. Θα δημιουργηθούν δύο διαφορετικές στήλες όπου η μια θα έχει τα προθέματα των κωδικών ενώ η άλλη θα έχει τους αριθμούς των διαφορετικών κωδικών. Με αυτόν τον τρόπο η στήλη με τα TK θα είναι κοινή μεταξύ των δύο σύνολα δεδομένων και θα συμβάλει στην ένωση τους.



Εικόνα 16: Τρίτος κόμβος μετά το φίλτρο με την χρήση της επιλογής διαχωρισμού στήλης SCH_STATEZIP.

Γι' αυτό χρησιμοποιούμε τον διαχωρισμό (Split) της στήλης SCH_STATEZIP καθώς η πληροφορία διασπάται σε δύο στήλες με το χώρισμα τους να είναι η παύλα (Dash).

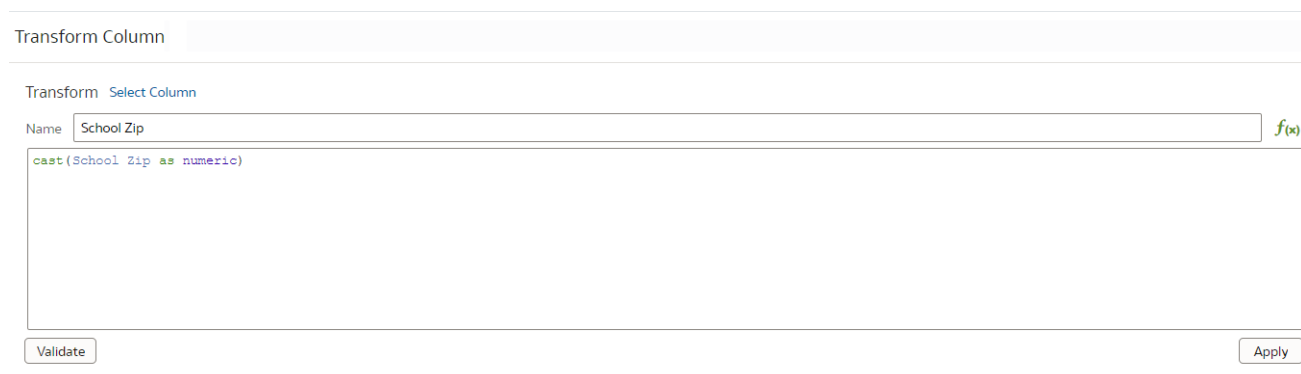


Εικόνα 17: Παραμετροποίηση του κόμβου Διαχωρισμού στήλης SCH_STATEZIP.

ab School State	ab School Zip
NC	28212
CT	6035
TX	77093
TN	38116
TX	78521
MI	48091
FL	33584
CA	95076

Εικόνα 18: Οι νέες στήλες που δημιουργήθηκαν μετά τον διαχωρισμό της στήλης SCH_STATEZIP

Τώρα, μπορεί να δει κανείς τις νέες στήλες SchoolState και SchoolZip που προστέθηκαν στο τέλος. Το βήμα διαίρεσης στήλης δημιουργεί τις νέες στήλες με τον τύπο δεδομένων του γονέα. Το SchoolState και το Zip είναι πεδία κειμένου. Καθώς θα χρησιμοποιήσουμε το SchoolZip για να ενώσουμε στήλη με το σύνολο δεδομένων ZipStat, ας το αλλάξουμε σε αριθμητικό πεδίο χρησιμοποιώντας το βήμα Μετασχηματισμού στήλης.



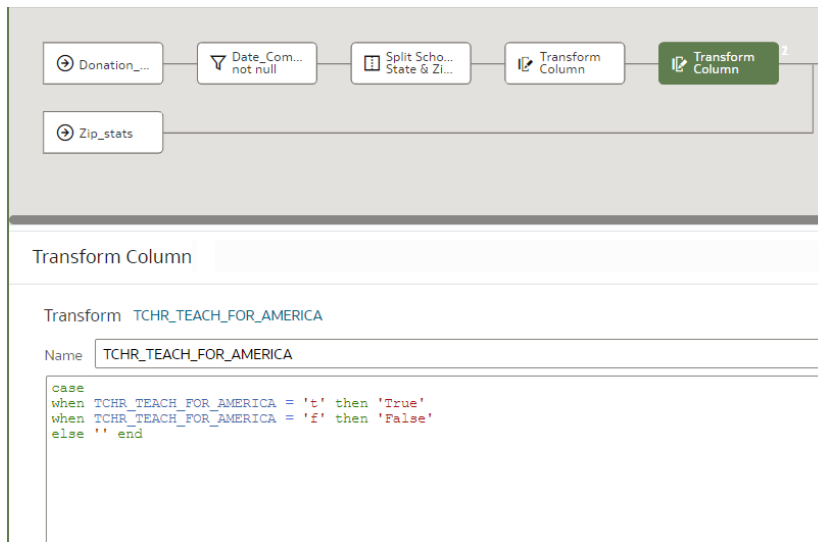
Transform Column

Transform [Select Column](#)

Name f(x)

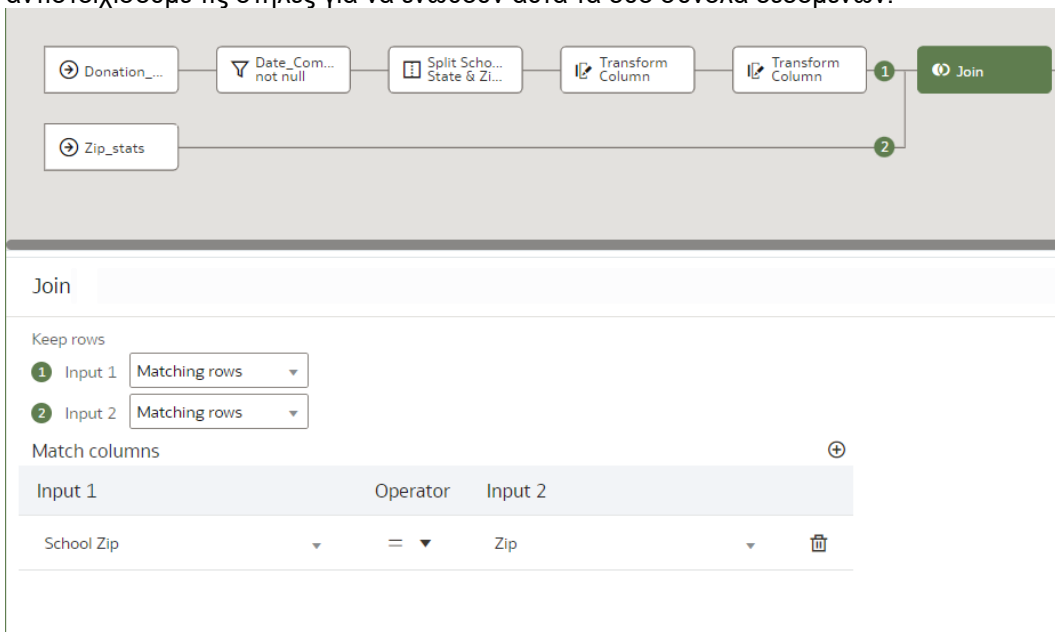
```
cast(School Zip as numeric)
```

Εικόνα 19: Παραμετροποίηση μεταβολής χαρακτηριστικών στήλης σε αριθμητική μονάδα.



Εικόνα 20: Φιλτράρισμα και μετατροπή στοιχείων στήλης από t σε True (Αληθές) και από f σε False (Ψευδές).

Στη συνέχεια, ας προσθέσουμε το σύνολο δεδομένων ZipStats σε αυτήν τη ροή δεδομένων και ας αντιστοιχίσουμε τις στήλες για να ενωθούν αυτά τα δύο σύνολα δεδομένων.



Εικόνα 21: Παραμετροποίηση αντιστοίχισης δύο στηλών.

Υπάρχουν πολλές στήλες στο συνδυασμένο σύνολο δεδομένων μας και δεν τις θέλουμε όλες, θα χρησιμοποιήσουμε το βήμα επιλογής στηλών και θα επιλέξουμε μόνο λίγα σχετικά χαρακτηριστικά και πεδία. Ορισμένα από αυτά είναι:

- PROJECT_ID
- PRIMARY_FOCUS_SUBJECT

- RESOURCE_TYPE
- POVERTY_LEVEL
- GRADE_LEVEL
- STUDENTS_REACHED
- TOTAL_DONATIONS
- NUM_DONORS
- SCHOOL_STATE
- POPULATION
- MEDIAN
- ZIP
- SCHOOL_ID

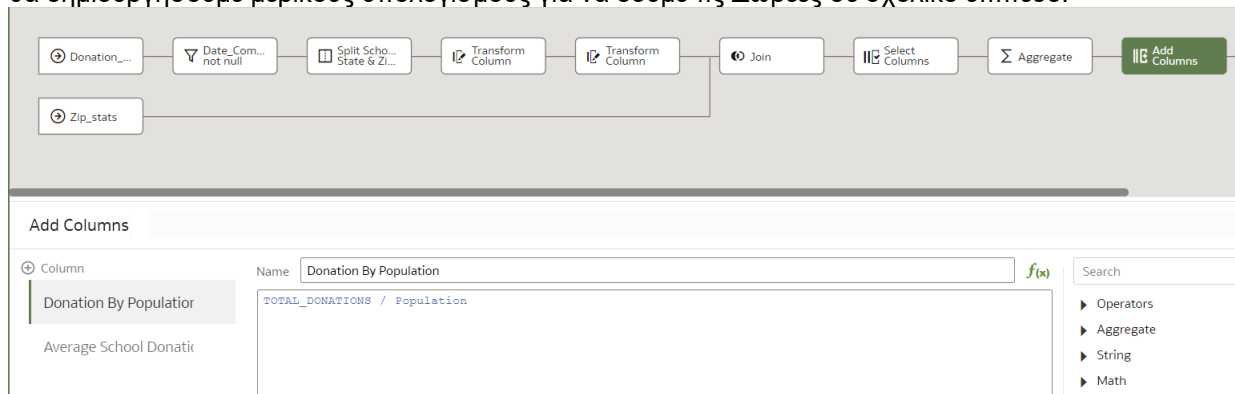
Το επόμενο βήμα είναι να συγκεντρώσουμε το σύνολο δεδομένων μας χρησιμοποιώντας το βήμα συγκέντρωσης (Aggregate Step). Επίσης έχουμε την δυνατότητα να προσθέσουμε και να αφαιρέσουμε στήλες στην Aggregate list. Μπορούμε να προσθέσουμε την ίδια στήλη με μία ή περισσότερες συναρτήσεις. Γι' αυτό θα προσθέσουμε το άθροισμα και τον μέσο όρο για την στήλη Students_reached.

Aggregate	Function	New column name
STUDENTS_REACHED	Average	STUDENTS_REACHED
TOTAL_DONATIONS	Average	TOTAL_DONATIONS
NUM_DONORS	Average	NUM_DONORS
Population	Average	Population
Median	Average	Median
Zip	Average	Zip
PROJECTID	Count	PROJECTID

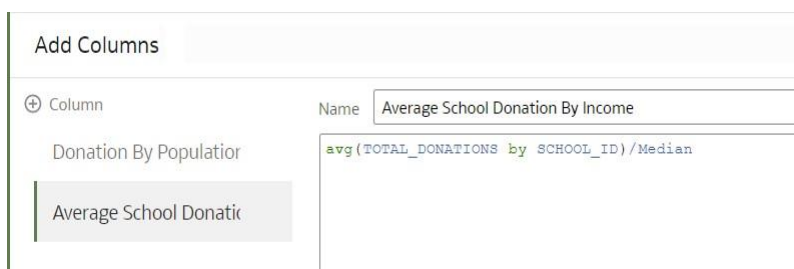
Group by
PRIMARY_FOCUS_SUBJECT
RESOURCE_TYPE
POVERTY_LEVEL
GRADE_LEVEL
School State
SCHOOL_ID
+ Add Group

Εικόνα 22: Παραμετροποίηση στηλών με χρήση συνάρτησης υπολογισμού Μέσου Όρου και Αθροίσματος.

Τώρα που το σύνολο δεδομένων μας συγκεντρώνεται ανά σχολική μονάδα, θα προσθέσουμε και θα δημιουργήσουμε μερικούς υπολογισμούς για να δούμε τις Δωρεές σε σχολικό επίπεδο.



Εικόνα 23: Δημιουργία κόμβου και δημιουργία στήλης Δωρεών ανά πληθυσμό.



Εικόνα 24: Δημιουργία στήλης Μέσου όρου δωρεών ανά σχολείο.

Το σύνολο δεδομένων εξόδου μπορεί είτε να αποθηκευτεί στην αποθήκευση συνόλου δεδομένων είτε ως πίνακας βάσης δεδομένων. Σε περίπτωση βάσης δεδομένων, πρέπει να επιλέξουμε μια σύνδεση με τη βάση δεδομένων. Για κάθε σύνολο δεδομένων στη ροή δεδομένων, μπορούμε να το αλλάξουμε και να το αντικαταστήσουμε με ένα διαφορετικό σύνολο δεδομένων με τον ίδιο ορισμό. Σημειώνουμε ότι αυτό το σύνολο δεδομένων εισόδου αλλάζει κατά το χρόνο σχεδιασμού και είναι διαφορετικό από την αλλαγή του συνόλου δεδομένων εισόδου κατά το χρόνο εκτέλεσης.

The screenshot displays a data pipeline with the following steps: Donation..., Date_Com... not null, Split Scho... State & Zi..., Transform Column, Transform Column, Join, Select Columns, Aggregate, Add Columns, Rename Columns, and Save Data. Below the pipeline is the 'Save Dataset' configuration panel.

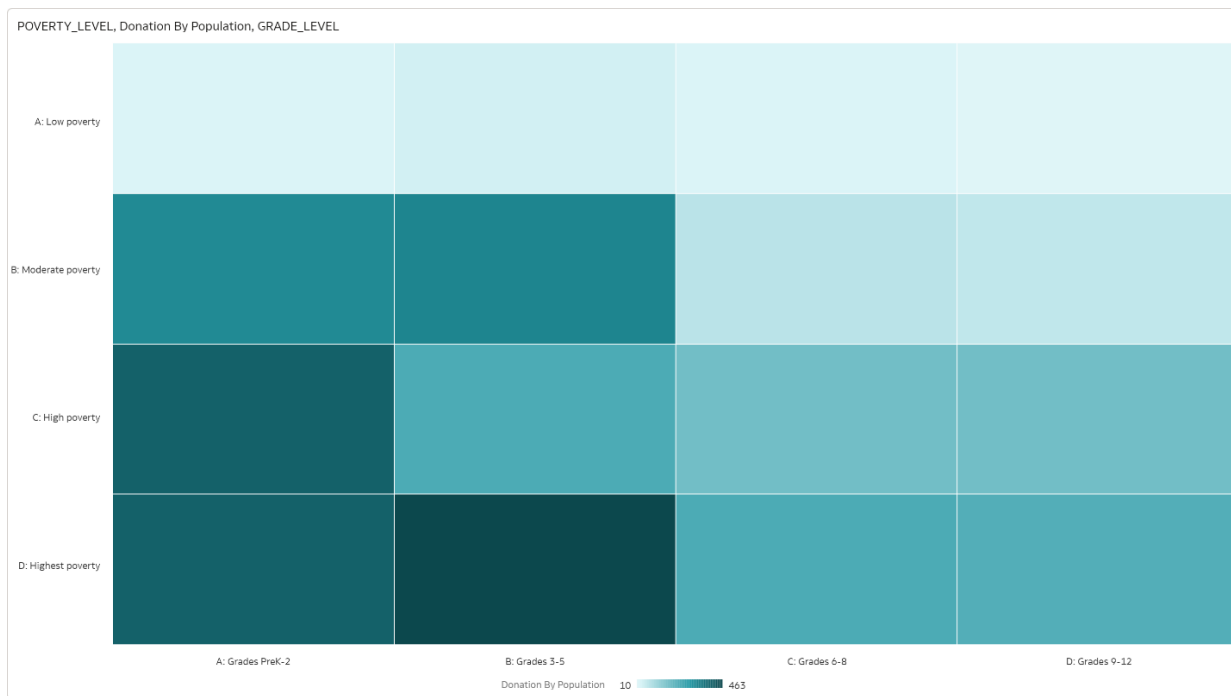
Save Dataset Configuration:

- Dataset: DF - DONATION BY SCHOOL
- Dataset Table: [Empty]
- Description: [Empty]
- Save data to: Dataset Storage
- When Run: Prompt to specify Dataset

Columns Configuration Table:

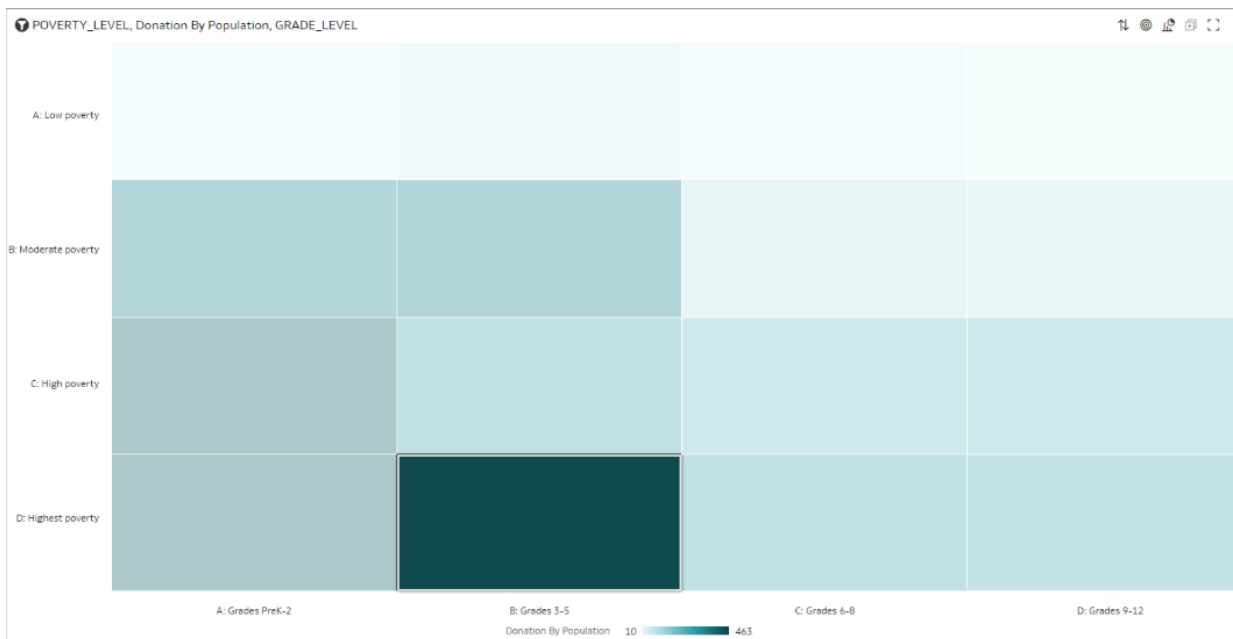
Name	Treat As	Default Aggregation
PRIMARY_FOCUS_SUBJECT	Attribute	
RESOURCE_TYPE	Attribute	
POVERTY_LEVEL	Attribute	
GRADE_LEVEL	Attribute	
School State	Attribute	
SCHOOL_ID	Attribute	
STUDENTS_REACHED	Measure	Average
TOTAL_DONATIONS	Measure	Average
NUM_DONORS	Measure	Average
Population	Measure	Average

Εικόνα 25: Αποθήκευση Δεδομένων



Εικόνα 26: Grid Visualization ανα κατηγορία οικονομικής κατάστασης που ανήκει ο κάθε μαθητής αλλά και την επίδοση του στα μαθήματα του σχολείου.

Παρατηρούμε ότι η υψηλότερη δωρεά γίνεται στους μαθητές που ανήκουν σε πιο δυσχερή οικονομική κατάσταση και παρουσιάζουν από τις χαμηλότερες επιδόσεις στο σχολείο.

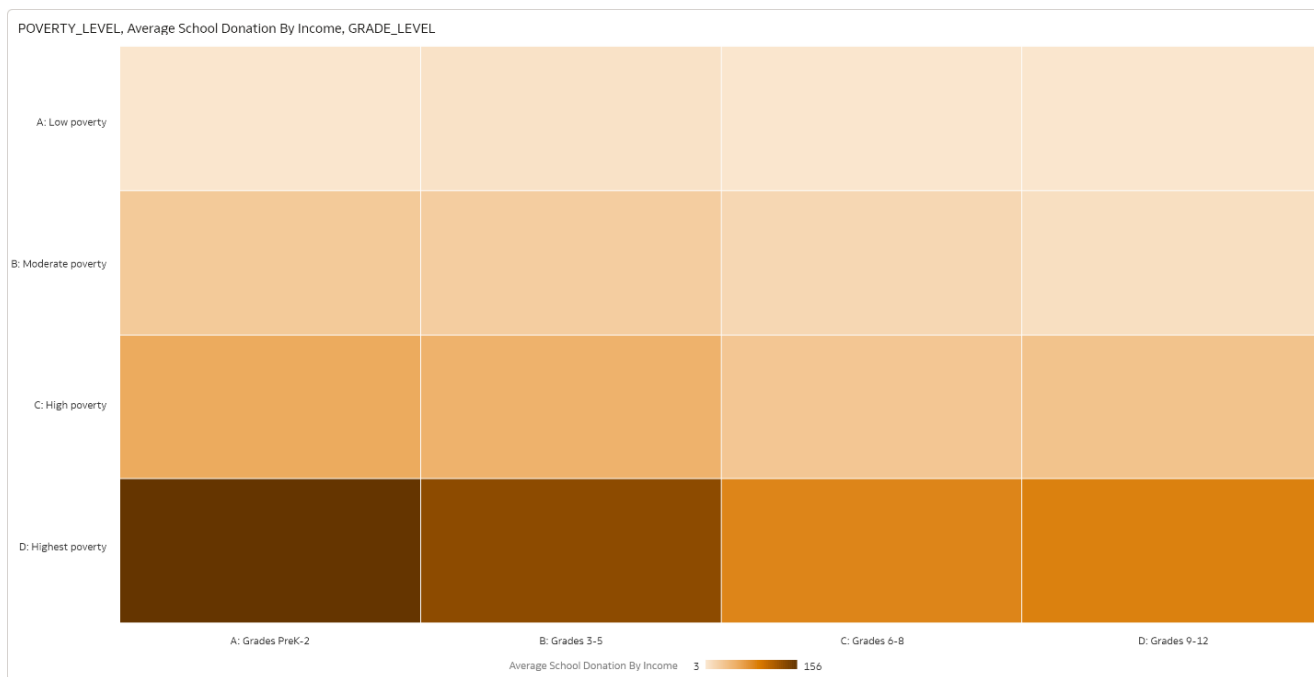


Εικόνα 27: Υψηλότερη δωρεά ανά πληθυσμό.

Στην εικόνα 27, παρατηρούμε την Υψηλότερη δωρεά ανά πληθυσμό με τιμή 462.79 και το Επίπεδο φτώχειας ανήκει στο επίπεδο Δ το οποίο αποτελεί και το Υψηλότερο Επίπεδο Φτώχειας.

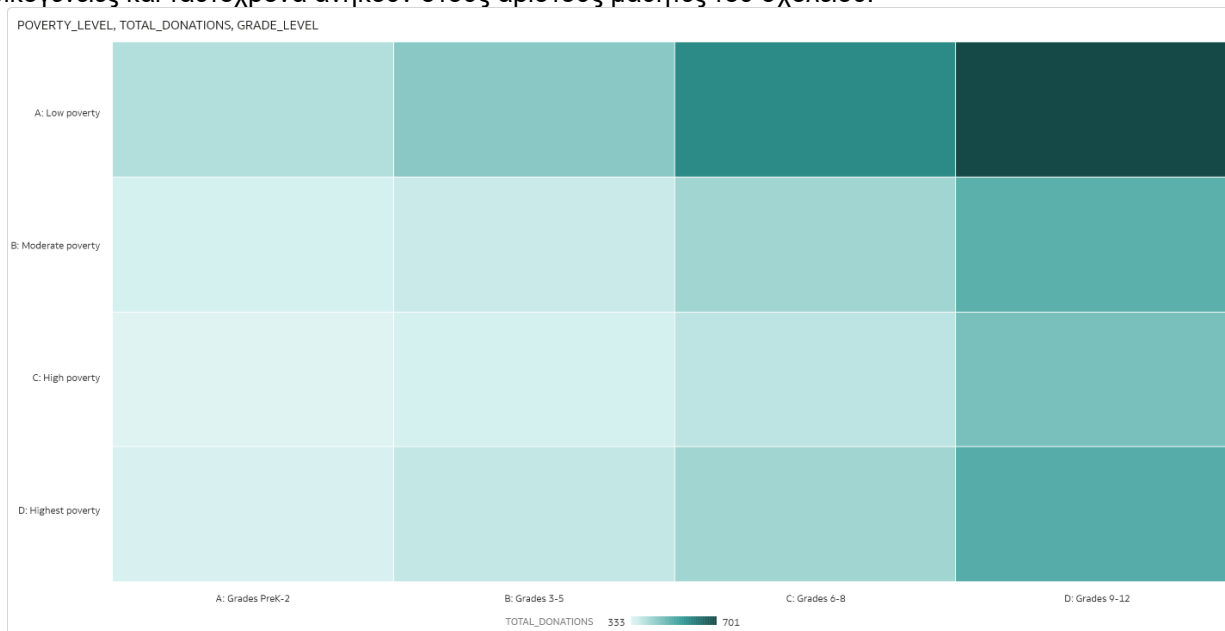
Αλλάζοντας το κριτήριο χρώματος από το Δωρεές ανά πληθυσμό (Donation by population) σε Μέσο όρο δωρεών σχολείων ανά συνολικό εισόδημα (Average School Donation by income), παρατηρούμε ότι ο μεγαλύτερος μέσος όρος δωρεών παρουσιάζεται σε μαθητές που ανήκουν στην χαμηλότερη βαθμολογία αλλά και με χαμηλότερο εισόδημα.

Επίπεδο βαθμολογίας: B: Βαθμολογία 3-5



Εικόνα 28: Απεικόνιση με μορφή πλέγματος με κατακόρυφο άξονα το επίπεδο φτώχειας των οικογενειών, τον οριζόντιο άξονα τις βαθμολογίες των μαθητών και ως χρωματικό κριτήριο τον μέσο όρο των σχολικών δωρεών.

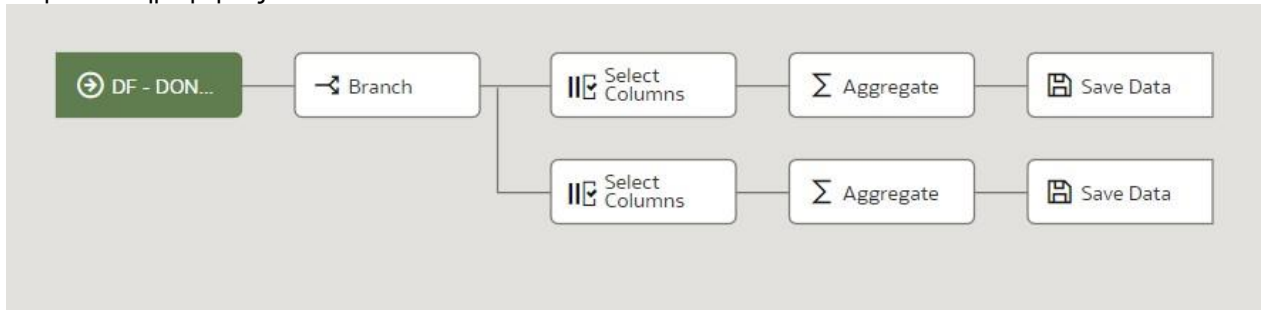
Αν αλλάξουμε ξανά το κριτήριο του χρώματος, από σε Μέσο όρο δωρεών σχολείων ανά συνολικό εισόδημα (Average School Donation by income) σε Συνολικές Δωρεές (Total Donations), παρατηρούμε ότι οι περισσότερες δωρεές γίνονται από μαθητές που ανήκουν σε ευκατάστατες οικογένειες και ταυτόχρονα ανήκουν στους άριστους μαθητές του σχολείου.



Εικόνα 29: Απεικόνιση με μορφή πλέγματος με κατακόρυφο άξονα το επίπεδο φτώχειας των οικογενειών, τον οριζόντιο άξονα τις βαθμολογίες των μαθητών και ως χρωματικό κριτήριο τον συνολικό αριθμό δωρεών που πραγματοποιήθηκαν.

6.6. Ακολουθίες (Sequences)

Στην συνέχεια της ανάλυσης των Dataset θα παρουσιάσουμε την λειτουργικότητα της Διακλάδωσης με αποτέλεσμα να δημιουργούνται στο τέλος ξεχωριστά datasets με συγκεκριμένα δεδομένα πληροφορίας.



Εικόνα 30: Ακολουθία με διακλάδωση.

Στο πρώτο κλαδί θα επιλέξουμε τις εξής στήλες:

1. Primary_focus_subject
2. Donation by Population
3. Average School Donation By Income

Aggregate 1:

Aggregate	Function	New column name	Group by
Donation By Population	Average	Donation By Population	PRIMARY_FOCUS_SUBJECT
Average School Donation By Inco	Average	Average School Donation By Inco	

⊕ Add Aggregate

⊕ Add Group

Εικόνα 31: Το πρώτο dataset θα ονομαστεί ως Donation by primary subject area

Save Dataset

Dataset: Columns

Dataset Table:

Description:

Save data to:

When Run Prompt to specify Dataset

Name	Treat As	Default Aggregation
PRIMARY_FOCUS_SUBJECT	Attribute ▼	
Donation By Population	Measure ▼	Average ▼
Average School Donation By Income	Measure ▼	Average ▼

ab PRIMARY_FOCUS...	99 Donation By Popu...	99 Average School D...
Character Education	0.0434068682202116	0.0102746873596386
College & Career Prep	0.0340577405291344	0.0100578357010667
Extracurricular	0.0708475292430722	0.0092861894034742

Εικόνα 32: Στο δεύτερο κλαδί επιλέγουμε τις στήλες School State, Donation By Population, Average School Donation By Income

Aggregate 2:

Aggregate

Aggregate	Function	New column name	Group by
<input type="text" value="Donation By Population"/>	<input type="text" value="Average"/> ▼	<input type="text" value="Donation By Population Average"/>	<input type="text" value="School State"/>
<input type="text" value="Average School Donation By Incor"/>	<input type="text" value="Average"/> ▼	<input type="text" value="Average School Donation By Incor"/>	<input type="text" value="Add Group"/>

Εικόνα 33: Υπολογισμός Μέσου όρου των επιλεγμένων στηλών σε νέα στήλη.

Save Dataset

Dataset: Columns

Dataset Table:

Description:

Save data to:

When Run Prompt to specify Dataset

Name	Treat As	Default Aggregation
School State	Attribute ▼	
Donation By Population Average	Measure ▼	Average ▼
Average School Donation By Income Average	Measure ▼	Average ▼

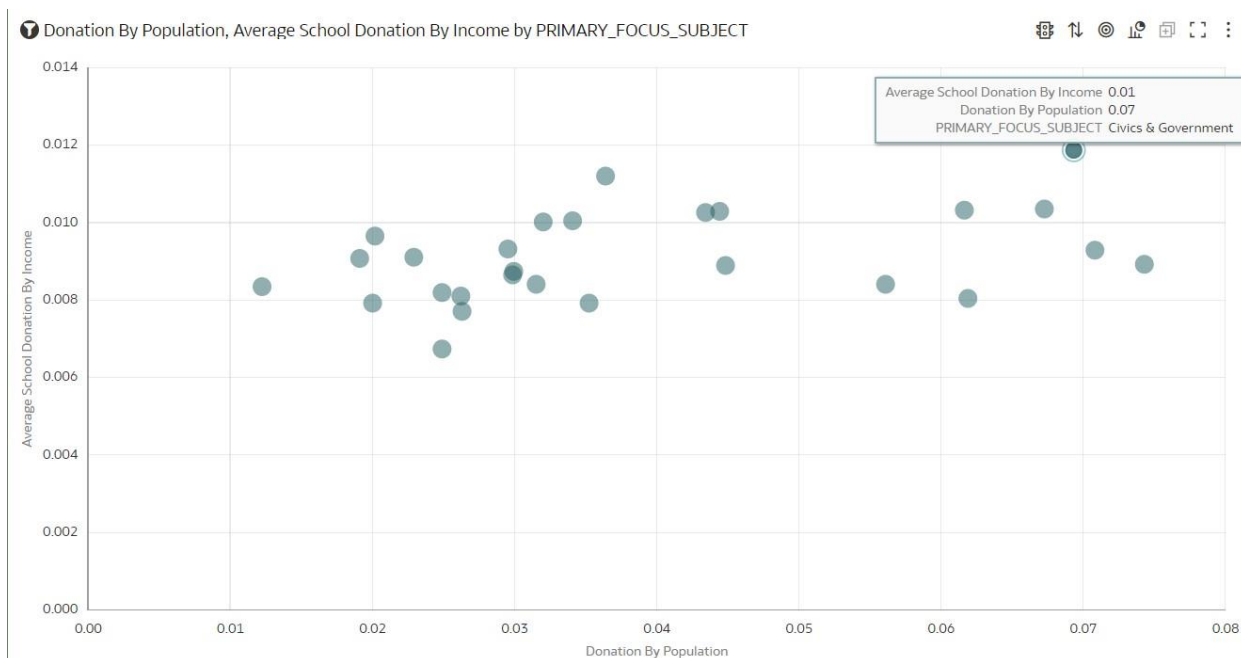
ab School State	99 Donation By Popu...	99 Average School D...
AK	0.1678560969504440	0.0082175937103320
AZ	0.0223838544313457	0.0080381433456848
IA	0.0520133821538000	0.0061145703990299

Εικόνα 34: Δημιουργία του συνόλου δεδομένων DonationByState με τις στήλες το School State, DonationByPopulation και Average School Donation by Income Average.

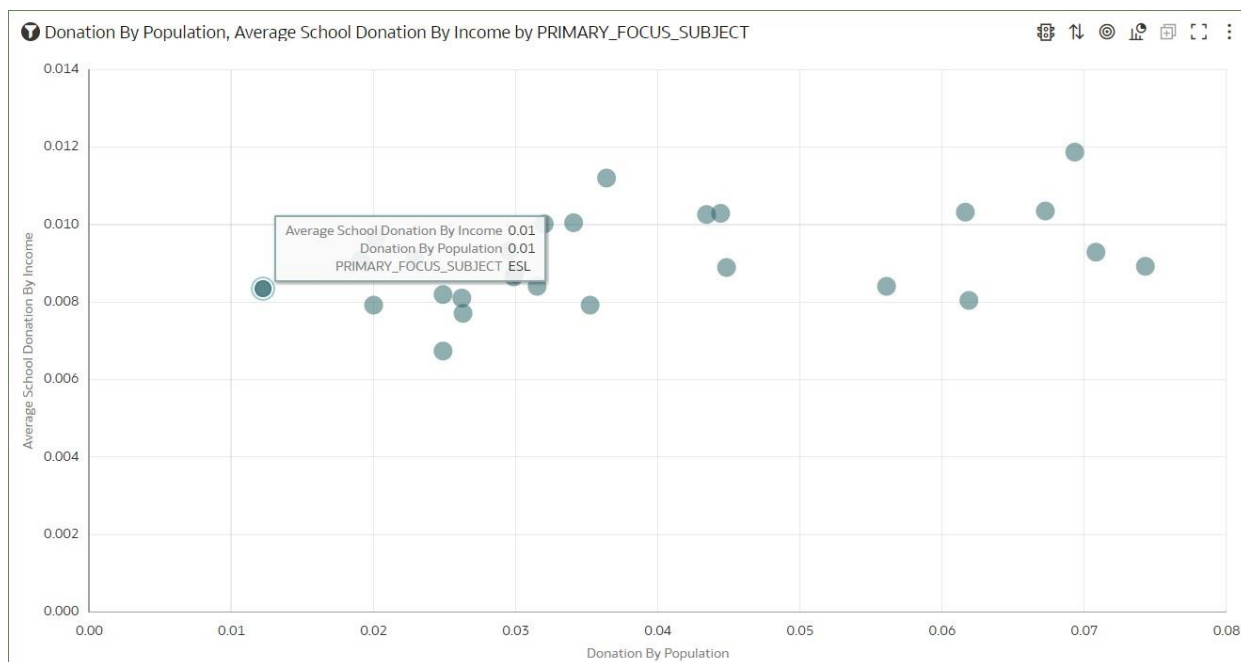
Οι ακολουθίες είναι «πακέτα» που μπορούν να εκτελέσουν πολλαπλές ροές δεδομένων και άλλες ακολουθίες η μία μετά την άλλη με συγκεκριμένη σειρά. Γι' αυτό στη συνέχεια δημιουργούμε μια ακολουθία για την εκτέλεση των δύο ροών δεδομένων η μία μετά την άλλη. Έπειτα από την εκτέλεση της Ακολουθίας (Sequence) δημιουργούνται 3 νέα Dataset.

6.7. Αναλύσεις στα σύνολα δεδομένων

Δημιουργούμε ένα Διάγραμμα Διασποράς (Scatter Plot) που στον οριζόντιο άξονα του είναι το (Donation by Population) και στον κατακόρυφο άξονα το Average School Donation by income και συγκέντρωση το κύριο μάθημα ενδιαφέροντος των μαθητών. Παρατηρούμε ότι οι υψηλότερες τιμές παρουσιάζονται στους μαθητές οι οποίοι παρακολουθούν το μάθημα Πολίτες και Κυβέρνηση (Civics and Government).

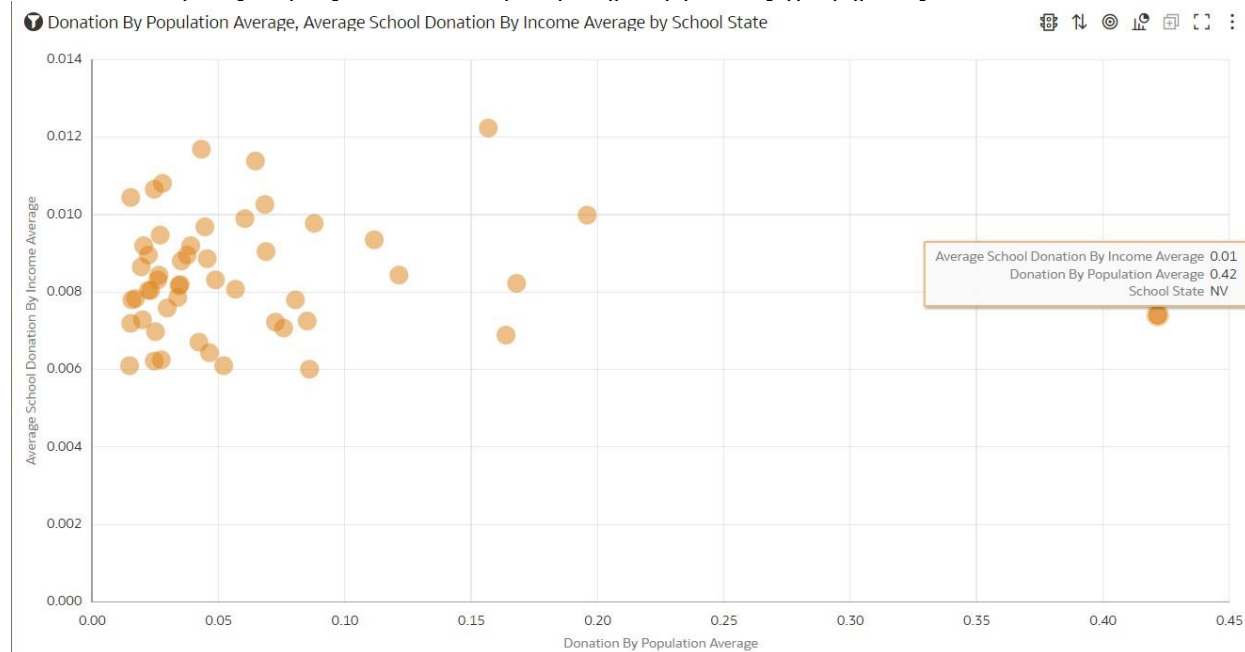


Εικόνα 35: Απεικόνιση με μορφή Διαγράμματος Διασποράς με κατακόρυφο άξονα τον μέσο αριθμό δωρεών ανά εισόδημα, τον οριζόντιο άξονα τις δωρεές ανά πληθυσμό και ως χρωματικό κριτήριο το κύριο μάθημα ενδιαφέροντος των μαθητών.



Εικόνα 36: Οι χαμηλότερες τιμές παρουσιάζονται από τους μαθητές που έχουν ως κύριο μάθημα ενδιαφέροντος το ESL.

Οπτικοποιούμε τις δωρεές ανά πολιτεία με την δημιουργία ενός γραφήματος.



Εικόνα 37: Απεικόνιση με μορφή Διαγράμματος Διασποράς με κατακόρυφο άξονα τον μέσο αριθμό δωρεών ανά μέσο εισόδημα, τον οριζόντιο άξονα τις δωρεές ανά πληθυσμό και ως χρωματικό κριτήριο το κύριο μάθημα ενδιαφέροντος των μαθητών.

7. Σενάριο ανάλυσης δεδομένων με χρήση προγνωστικών μοντέλων

Αξιοποιώντας τα ίδια σύνολα δεδομένων που χρησιμοποιήθηκαν και για το προηγούμενο σενάριο, θα εστιάσουμε σε μια πιο προηγμένη ανάλυση που στοχεύει στη δημιουργία ενός συνόλου δεδομένων πρόβλεψης. Συγκεκριμένα, το σενάριο ξεκινά με τη δημιουργία ενός συνόλου δεδομένων βάσει αρχείων και την ελέγχου και προετοιμασία των δεδομένων. Έπειτα, δημιουργείται ένα τυχαίο δείγμα του συνόλου δεδομένων που θα χρησιμοποιηθεί για την εκπαίδευση της μηχανής μάθησης. Στο επόμενο βήμα, εκπαιδεύεται η μηχανή μέσω του προγνωστικού μοντέλου Elastic Net Regression. Στη συνέχεια, δημιουργείται ένα νέο έργο και γίνεται επιθεώρηση της ποιότητας του μοντέλου και των σχετικών συνόλων δεδομένων. Ακολουθεί μια διαδικασία επανάληψης και βελτιστοποίησης της εκπαίδευσης του μοντέλου. Στη συνέχεια, πραγματοποιείται σύγκριση των σεναρίων προγνωστικού μοντέλου και των μοντέλων διαφορετικών τύπων, με έμφαση στην ταξινόμηση του συντονισμού μοντέλων δέντρων παλινδρόμησης. Τέλος, το εκπαιδευμένο μοντέλο χρησιμοποιείται για τη βαθμολόγηση ενός συνόλου δεδομένων, προσφέροντας μια εφαρμογή πρόβλεψης. Μέσω αυτού του σεναρίου, θα εξετάσουμε την αποτελεσματικότητα της μηχανικής μάθησης και τη συνεισφορά της στη δημιουργία ενός συνόλου δεδομένων πρόβλεψης. Τα αποτελέσματα και οι συμπεράσματα αυτής της ανάλυσης αναμένεται να προσφέρουν νέα ευρήματα και εισαγωγές για την αξιοποίηση της μηχανικής μάθησης στο πεδίο της ανάλυσης δεδομένων και της πρόβλεψης στις επιχειρήσεις.

7.1. Τι είναι τα προγνωστικά μοντέλα Oracle Analytics;

Ένα μοντέλο πρόβλεψης στο Oracle Analytics εφαρμόζει έναν συγκεκριμένο αλγόριθμο σε ένα σύνολο δεδομένων για να προβλέψει τιμές, κλάσεις ή ομάδες στα δεδομένα. Η πλατφόρμα περιλαμβάνει αλγόριθμους που βοηθούν τους χρήστες να εκπαιδεύσουν μοντέλα πρόβλεψης για διάφορους σκοπούς, όπως τα δέντρα ταξινόμησης και παλινδρόμησης (CART), η λογιστική παλινδρόμηση και ο αλγόριθμος k-means. Χρησιμοποιώντας τον επεξεργαστή ροής δεδομένων, αρχικά δημιουργείται μια ροή και προστίθεται το σύνολο δεδομένων που θα χρησιμοποιηθεί για την εκπαίδευση του μοντέλου. Αυτό το σύνολο δεδομένων εκπαίδευσης περιλαμβάνει τα δεδομένα που πρόκειται να προβλέψουμε, όπως οι πωλήσεις ή η ηλικία. Εάν είναι απαραίτητο, μπορούμε να χρησιμοποιήσουμε το πρόγραμμα επεξεργασίας ροής δεδομένων για να επεξεργαστούμε το σύνολο δεδομένων προσθέτοντας ή επιλέγοντας στήλες, ενώνοντας κ.λπ. Αφού επιβεβαιώσουμε ότι τα δεδομένα είναι αυτά που θέλουμε να χρησιμοποιήσουμε για την εκπαίδευση, προσθέτουμε ένα βήμα εκπαίδευσης στη ροή δεδομένων και επιλέγουμε έναν αλγόριθμο ταξινόμησης, παλινδρόμησης ή συμπλέγματος για την εκπαίδευση του μοντέλου. Στη συνέχεια, ονομάζουμε το μοντέλο που δημιουργήθηκε, αποθηκεύουμε τη ροή δεδομένων και την εκτελούμε για να εκπαιδευτεί και να δημιουργηθεί το μοντέλο. Έπειτα, μπορούμε να αξιολογήσουμε τις ιδιότητες του μοντέλου μηχανικής εκμάθησης για να προσδιορίσουμε την ποιότητά του. Εάν είναι απαραίτητο, μπορούμε να επαναλάβουμε τη διαδικασία εκπαίδευσης μέχρι το μοντέλο να φτάσει στην επιθυμητή ποιότητα. Τέλος, μπορούμε να χρησιμοποιήσουμε το εκπαιδευμένο μοντέλο για να αξιολογήσουμε άγνωστα δεδομένα και να δημιουργήσουμε ένα σύνολο δεδομένων στη ροή δεδομένων ή να προσθέσουμε οπτικοποίηση των προβλέψεων σε ένα βιβλίο εργασίας. [20]

Παραδείγματα Προβλέψεων

1. Ας υποθέσουμε ότι θέλουμε να δημιουργήσουμε και να εκπαιδεύσουμε ένα μοντέλο πολλαπλών ταξινομήσεων για να προβλέψουμε ποιοι ασθενείς έχουν υψηλό κίνδυνο να αναπτύξουν καρδιακή νόσο. Παρέχουμε ένα σύνολο δεδομένων εκπαίδευσης που περιέχει χαρακτηριστικά για μεμονωμένους ασθενείς, όπως η ηλικία, το φύλο και εάν έχουν βιώσει ποτέ πόνο στο στήθος, και μετρήσεις όπως η αρτηριακή πίεση, το σάκχαρο στο αίμα, η χοληστερόλη και ο μέγιστος καρδιακός ρυθμός. Το σύνολο δεδομένων εκπαίδευσης

- περιέχει επίσης μια στήλη με το όνομα "Πιθανότητα" στην οποία εκχωρείται μία από τις ακόλουθες τιμές: απουσιάζει, λιγότερο πιθανό, πιθανό, πολύ πιθανό ή υπάρχει.
2. Επιλέγουμε τον αλγόριθμο CART (Δέντρο αποφάσεων) επειδή αγνοεί περιττές στήλες που δεν προσθέτουν αξία για την πρόβλεψη και προσδιορίζει και χρησιμοποιεί μόνο τις στήλες που είναι χρήσιμες για την πρόβλεψη του στόχου. Όταν προσθέτουμε τον αλγόριθμο στη ροή δεδομένων, επιλέγουμε τη στήλη Likelihood για να εκπαιδεύσουμε το μοντέλο. Ο αλγόριθμος χρησιμοποιεί μηχανική εκμάθηση για να επιλέξει τις στήλες του προγράμματος οδήγησης που χρειάζεται για να εκτελέσει και να εξάγει προβλέψεις και σχετικά σύνολα δεδομένων.
 3. Επιθεωρούμε τα αποτελέσματα και συντονίζουμε με ακρίβεια το μοντέλο εκπαίδευσης και, στη συνέχεια, εφαρμόζουμε το μοντέλο σε ένα μεγαλύτερο σύνολο δεδομένων για να προβλέψουμε ποιοι ασθενείς έχουν μεγάλη πιθανότητα να έχουν ή να αναπτύξουν καρδιακή νόσο.

7.2. Αλγόριθμοι μοντελοποίησης μηχανικής μάθησης

Το Oracle Analytics παρέχει αλγόριθμους για οποιαδήποτε από τις ανάγκες σε μοντελοποίηση μηχανικής μάθησης: αριθμητική πρόβλεψη, πολυ-ταξινομητή, δυαδικό ταξινομητή και ομαδοποίηση (numeric prediction, multi-classifier, binary classifier, clustering). Η λειτουργία μηχανικής εκμάθησης της Oracle απευθύνεται σε προχωρημένους αναλυτές δεδομένων που έχουν μια ιδέα για το τι αναζητούν στα δεδομένα τους, είναι εξοικειωμένοι με την πρακτική της προγνωστικής ανάλυσης και κατανοούν τις διαφορές μεταξύ των αλγορίθμων. Το Oracle Analytics διαθέτει πολλούς αλγόριθμους μηχανικής εκμάθησης για κάθε είδος πρόβλεψης ή ταξινόμησης. Με αυτούς τους αλγόριθμους, οι χρήστες μπορούν να δημιουργήσουν περισσότερα από ένα μοντέλα ή να χρησιμοποιήσουν διαφορετικές βελτιωμένες παραμέτρους ή να χρησιμοποιήσουν διαφορετικά σύνολα δεδομένων εκπαίδευσης εισόδου και στη συνέχεια να επιλέξουν το καλύτερο μοντέλο. Ο χρήστης μπορεί να επιλέξει το καλύτερο μοντέλο συγκρίνοντας και αξιολογώντας τα μοντέλα με τα δικά του κριτήρια. Για να προσδιορίσουν το καλύτερο μοντέλο, οι χρήστες μπορούν να εφαρμόσουν το μοντέλο και να οπτικοποιήσουν τα αποτελέσματα των υπολογισμών για να προσδιορίσουν την ακρίβεια ή μπορούν να ανοίξουν και να εξερευνήσουν τα σχετικά σύνολα δεδομένων για τα οποία η Oracle Analytics χρησιμοποίησε το μοντέλο.

- **CART:** Χρησιμοποιεί δέντρα απόφασης για να προβλέψει τόσο διακριτές όσο και συνεχείς τιμές. Χρησιμοποιείται σε μεγάλα σύνολα δεδομένων.
- **Elastic Net Linear Regression:** Προηγμένο μοντέλο παλινδρόμησης. Παρέχει πρόσθετες πληροφορίες (κανονικοποίηση), εκτελεί επιλογή μεταβλητών και εκτελεί γραμμικούς συνδυασμούς. Ποινές των μεθόδων παλινδρόμησης Lasso και Ridge. Χρησιμοποιείται με μεγάλο αριθμό χαρακτηριστικών για την αποφυγή της συγγραμμικότητας (όπου πολλαπλά χαρακτηριστικά συσχετίζονται τέλεια) και την υπερπροσαρμογή.
- **Ιεραρχική:** Δημιουργεί μια ιεραρχία ομαδοποίησης χρησιμοποιώντας είτε από κάτω προς τα πάνω (κάθε παρατήρηση είναι το δικό της σύμπλεγμα και μετά συγχωνεύεται) είτε από πάνω προς τα κάτω (όλες οι παρατηρήσεις ξεκινούν ως ένα σύμπλεγμα) και μετρήσεις απόστασης. Χρησιμοποιείται όταν το σύνολο δεδομένων δεν είναι μεγάλο και ο αριθμός των συμπλεγμάτων δεν είναι γνωστός εκ των προτέρων.
- **K-Means:** Επαναληπτικά χωρίζει τις εγγραφές σε k συμπλέγματα όπου κάθε παρατήρηση ανήκει στο σύμπλεγμα με τον πλησιέστερο μέσο όρο. Χρησιμοποιείται για ομαδοποίηση στηλών μέτρησης και με καθορισμένη προσδοκία για τον αριθμό των απαιτούμενων συστάδων. Λειτουργεί καλά με μεγάλα σύνολα δεδομένων. Τα αποτελέσματα είναι διαφορετικά σε κάθε τρέξιμο.
- **Γραμμική παλινδρόμηση:** Γραμμική προσέγγιση για μια σχέση μοντελοποίησης μεταξύ της μεταβλητής στόχου και άλλων χαρακτηριστικών στο σύνολο δεδομένων.

Χρησιμοποιείται για την πρόβλεψη αριθμητικών τιμών όταν τα χαρακτηριστικά δεν συσχετίζονται τέλεια.

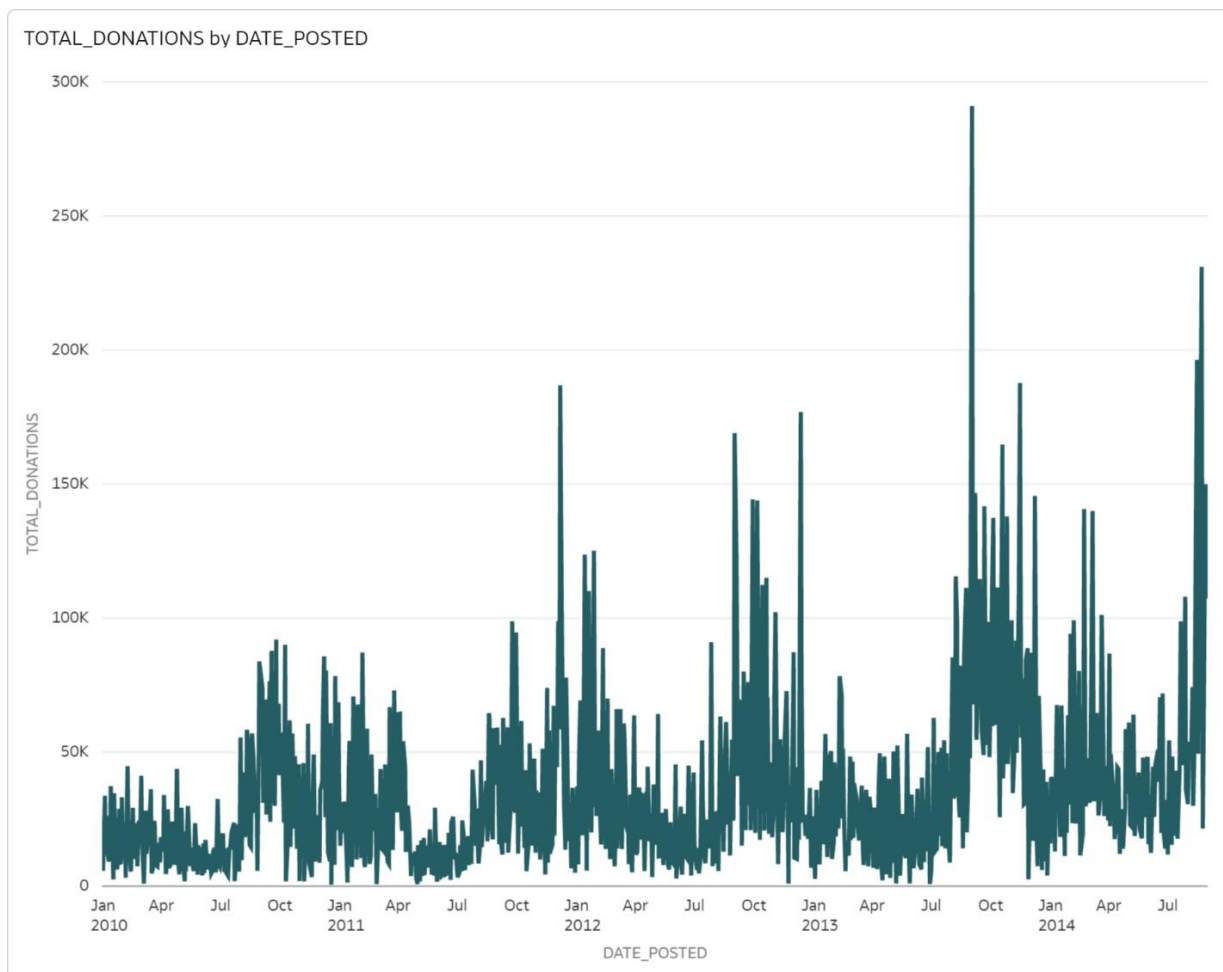
- **Logistic Regression:** Χρησιμοποιείται για την πρόβλεψη της τιμής μιας κατηγορηματικά εξαρτημένης μεταβλητής. Η εξαρτημένη μεταβλητή είναι μια δυαδική μεταβλητή που περιέχει δεδομένα κωδικοποιημένα σε 1 ή 0.
- **Naive Bayes:** Πιθανοτική ταξινόμηση βασισμένη στο θεώρημα του Bayes που δεν προϋποθέτει καμία εξάρτηση μεταξύ των χαρακτηριστικών. Χρησιμοποιείται όταν υπάρχει μεγάλος αριθμός διαστάσεων εισόδου.
- **Νευρωνικό Δίκτυο:** Επαναληπτικός αλγόριθμος ταξινόμησης που μαθαίνει συγκρίνοντας το αποτέλεσμα ταξινόμησης με την πραγματική τιμή και το επιστρέφει στο δίκτυο για να τροποποιήσει τον αλγόριθμο για περαιτέρω επαναλήψεις. Χρήση για ανάλυση κειμένου.
- **RandomForest:** Μια μέθοδος εκμάθησης συνόλου που κατασκευάζει πολλαπλά δέντρα αποφάσεων και εξάγει την τιμή που αντιπροσωπεύει συλλογικά όλα τα δέντρα αποφάσεων. Χρησιμοποιείται για την πρόβλεψη αριθμητικών και κατηγορικών μεταβλητών.
- **SVM:** Ταξινομεί εγγραφές χαρτογραφώντας τις στο διάστημα και κατασκευάζοντας υπερεπίπεδα που μπορούν να χρησιμοποιηθούν για ταξινόμηση. Νέες εγγραφές (δεδομένα βαθμολογίας) χαρτογραφούνται στο χώρο και προβλέπεται ότι ανήκουν σε μια κατηγορία, η οποία βασίζεται στην πλευρά του υπερεπίπεδου όπου πέφτουν.

Με τα ίδια σύνολα δεδομένων που εργαστήκαμε για το προηγούμενο σενάριο. Θα εφαρμόσουμε ένα νέο σενάριο που χρησιμοποιεί την μηχανική μάθηση και να δημιουργήσει ένα σύνολο δεδομένων πρόβλεψης.

1. Δημιουργία συνόλου δεδομένων βάσει αρχείων
2. Έλεγχος & προετοιμασία των δεδομένων
3. Δημιουργία τυχαίου δείγματος του συνόλου δεδομένων
4. Εκπαίδευση μηχανής με το προγνωστικό μοντέλο Elastic Net Regression
5. Δημιουργία σεναρίου σε ένα νέο project
6. Επιθεώρηση της ποιότητα μοντέλου και των σχετικών συνόλων δεδομένων
7. Σύγκριση σεναρίων προγνωστικού μοντέλου
8. Σύγκριση Μοντέλων διαφορετικών τύπων.
9. Ταξινόμηση συντονισμού μοντέλων δέντρων παλινδρόμησης.
10. Χρήση μοντέλου για τη βαθμολόγηση ενός συνόλου δεδομένων (Εφαρμογή πρόβλεψης)

Στο συγκεκριμένο σύνολο δεδομένων το σύνολο των δωρεών ανέρχεται στα 58 εκατομμύρια δολάρια για χρονικό διάστημα τεσσάρων ετών.

Στην συνέχεια παρουσιάζεται η γραφική αναπαράσταση της συνολικής εικόνας των δωρεών για το σύνολο του dataset που χρησιμοποιήθηκε για το σενάριο μας.

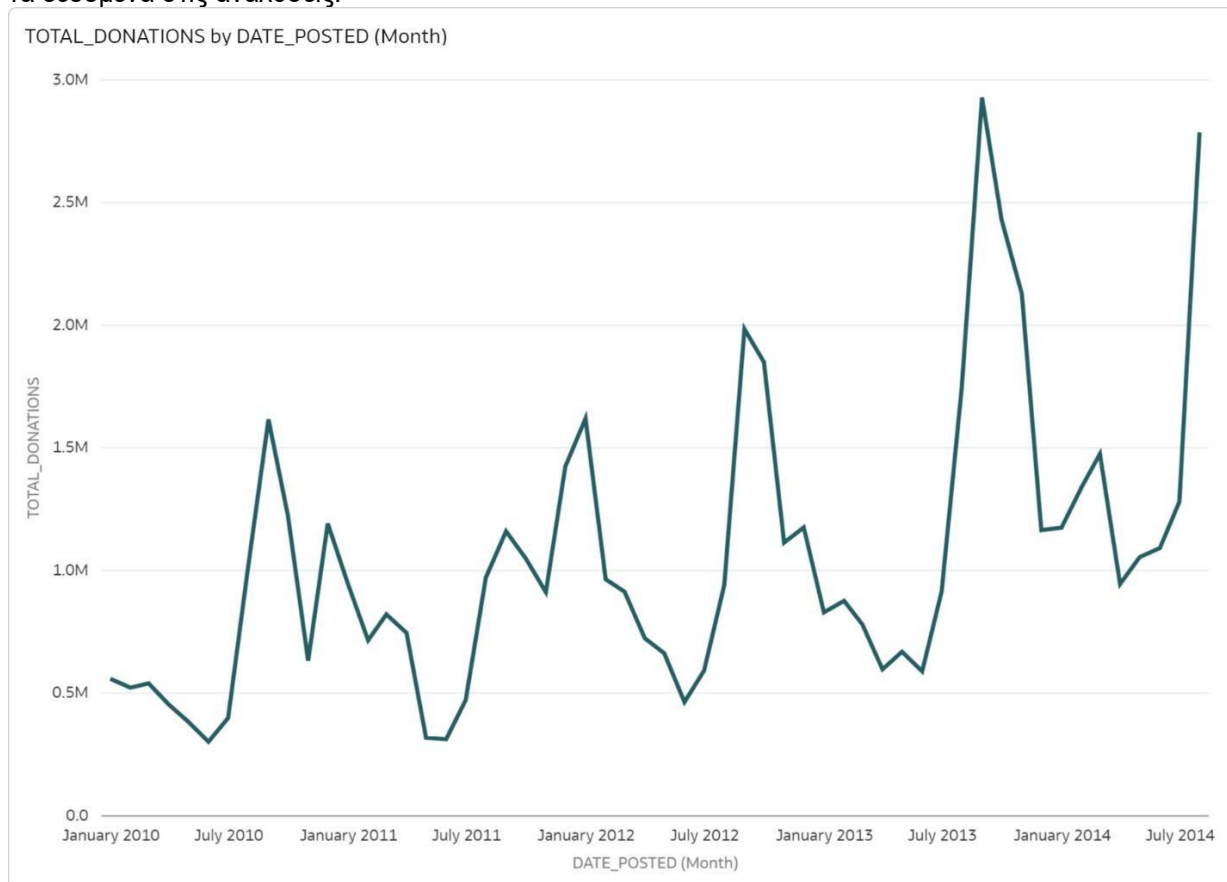


Εικόνα 38: Διάγραμμα συνολικών δωρεών ανά μήνα καταχώρησης.

PROJECTID
150,000
TOTAL_DONATIONS 58,500,000.00

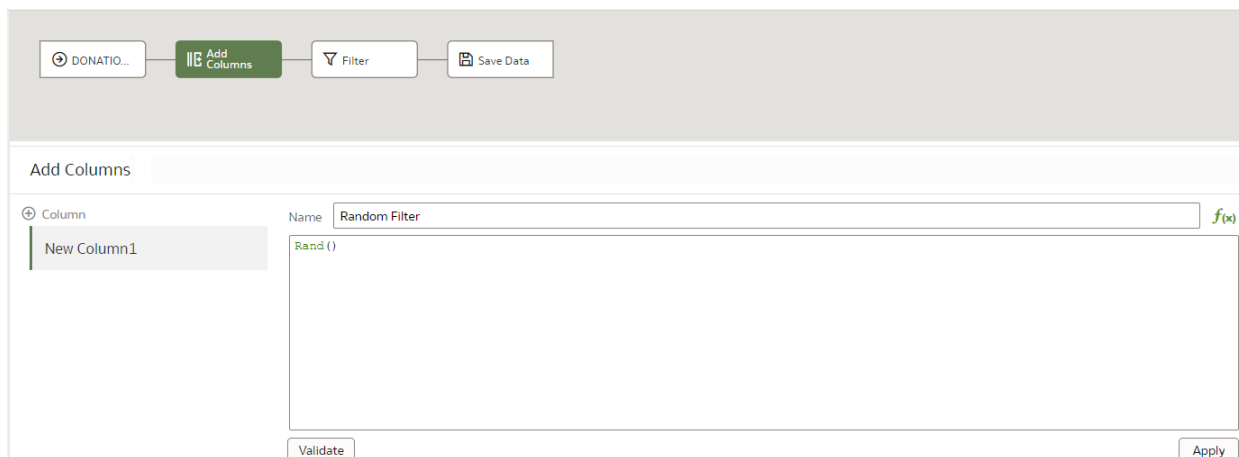
Εικόνα 39: Ο Συνολικός αριθμός δωρεών ανέρχεται στα 58,5 εκατομμύρια δολάρια για 150 χιλιάδες γραμμές του συνόλου δεδομένων.

Η εξέταση των τάσεων σε μηνιαίο επίπεδο βοηθάει να δούμε τα διάφορα μοτίβα που παρουσιάζουν τα δεδομένα στις αναλύσεις.



Εικόνα 40: Διάγραμμα linechart Συνολικού αριθμού δωρεών ανά μήνα δημοσίευσης.

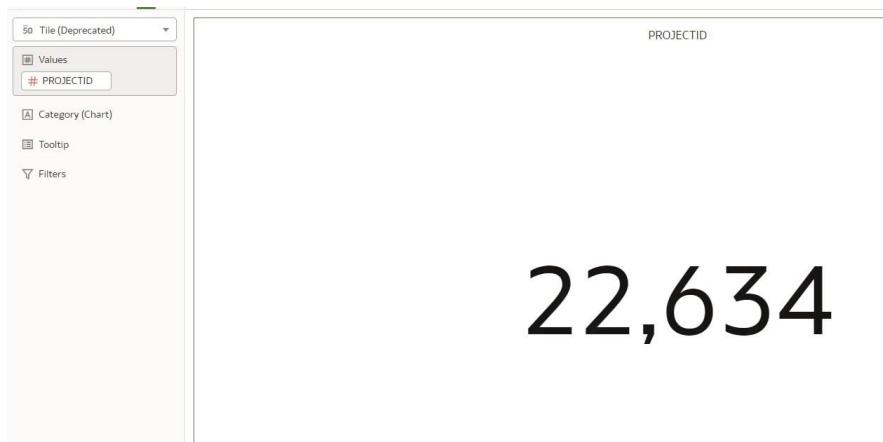
Σε αυτό το σενάριο, θα κατασκευάσουμε μοντέλα για να προβλέψουμε τη συνολική αξία των δωρεών. Με βάση την εποχικότητα που μπορούμε να δούμε, θα διατηρήσουμε τον Ιούνιο του 2012 έως τον Ιούνιο του 2014 ως περίοδο αναφοράς για την εκπαίδευση των μοντέλων. Στη συνέχεια, θα χρησιμοποιήσουμε μια ροή δεδομένων για να δημιουργήσουμε ένα υποσύνολο του αρχικού μας συνόλου 150.000 σειρών, με βάση αυτήν την περίοδο. Αρχικά φιλτράρουμε για να διατηρήσουμε μόνο τη διετία Ιούνιος 2012 έως Ιούνιος 2014. Επίσης, θα μειώσουμε αυτά τα δεδομένα περαιτέρω σε ένα υποσύνολο τους για να τα χρησιμοποιήσουμε στα μοντέλα μηχανικής μάθησης. Για να γίνει αυτό, θα εισαγάγουμε μια τυχαία στήλη και θα τη φιλτράρουμε με την δημιουργία ενός φίλτρου τυχειότητας.



Εικόνα 41: Παραμετροποίηση φίλτρου τυχαιότητας.

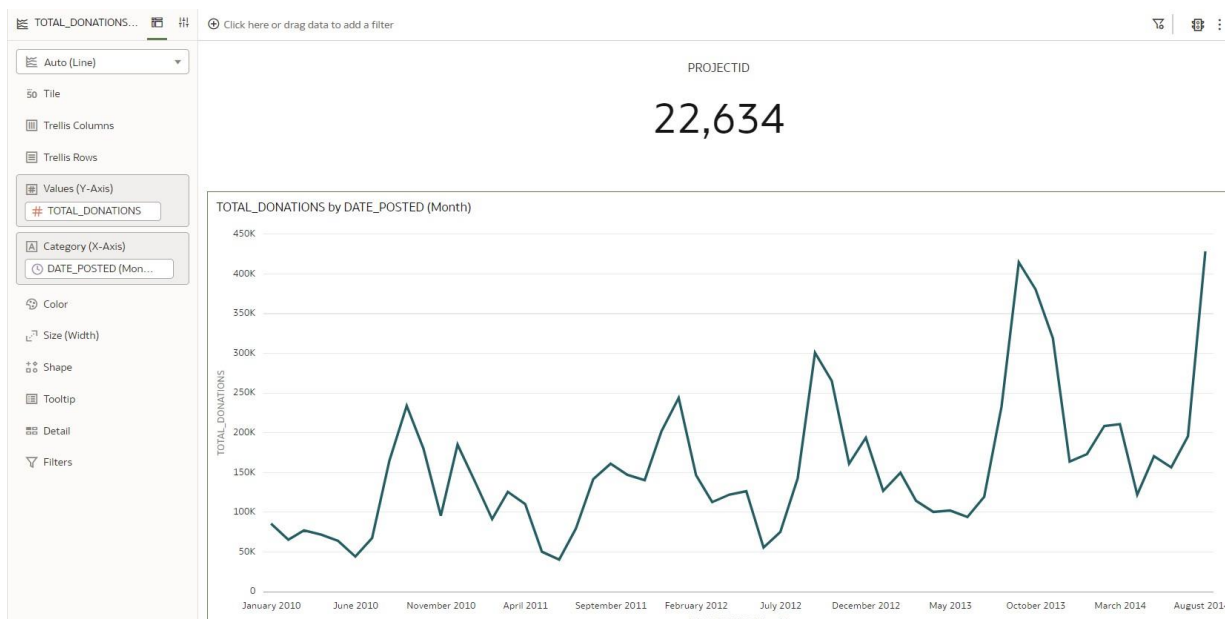
Το φιλτράρισμα στο $\text{Rand}() < 0.15$ ισοδυναμεί με την επιλογή τυχαίου δείγματος 15% του συνόλου δεδομένων. Έπειτα, θα αφαιρέσουμε την τυχαία στήλη μετά το φιλτράρισμα οπότε θα δημιουργηθεί ένα δείγμα από το αρχικό σύνολο δεδομένων.

Μετά την εφαρμογή του φίλτρου το δείγμα του συνόλου δεδομένων απαριθμεί 22.634 γραμμές από τις 150 χιλιάδες που ήταν στην αρχή.



Εικόνα 42: 22.634 σειρές, το 15% των δεδομένων της διαιτίας που επιλέξαμε.

Στην συνέχεια μπορούμε να ορίσουμε περισσότερη συγκέντρωση σε μερικές μετρήσεις, προτού εκπαιδύσουμε μοντέλα μηχανικής μάθησης σε αυτά τα δεδομένα.



Εικόνα 43: Διάγραμμα linechart Συνολικού αριθμού δωρεών ανά μήνα δημοσίευσης μετά το φιλτράρισμα.

Με βάση τα παραπάνω δεδομένα που δημιουργήθηκαν, μπορούν να δημιουργηθούν μοντέλα μηχανικής μάθησης και στην συνέχεια να εκτελεστούν για να δημιουργηθούν νέα δεδομένα.

Η Εκπαίδευση ενός προγνωστικού μοντέλου ελαστικής καθαρής παλινδρόμησης μπορεί να είναι η πρώτη ενέργεια για το σενάριο πρόβλεψης. Αυτή η ενέργεια περιλαμβάνει την εκπαίδευση ενός αριθμητικού μοντέλου από μια ροή δεδομένων για την πρόβλεψη της αξίας των συνολικών δωρεών.

Αρχικά, θα δοκιμαστεί το μοντέλο ελαστικής καθαρής γραμμικής παλινδρόμησης. Το επιδιωκόμενο είναι η εκπαίδευση ενός μοντέλου αριθμητικής πρόβλεψης, με στόχο να προβλεφθεί η συνολική αξία των δωρεών (TOTAL_DONATIONS). Η στήλη στόχος (TOTAL_DONATIONS) αποτελεί την μεταβλητή που θα προσπαθήσουμε να προβλέψουμε. Όλες οι υπόλοιπες παράμετροι του μοντέλου είναι ευέλικτες και μπορούν να τροποποιηθούν, αλλά για λόγους απλότητας, αποφασίσαμε να διατηρήσουμε τις προεπιλεγμένες τους τιμές στην περίπτωση μας.

Model Training Script: Elastic Net Linear Regression for model training

* Target: TOTAL_DONATIONS
target, the target(label) to learn/predict

L1 Ratio: 0.5

L1 Ratio

L2 Ratio: 0.5

L2 Ratio

Categorical Column Imputation: Most Frequent
The mode method for categorical features to fill NA. Two options: most frequent and least frequent. Default is most frequent.

Numerical Column Imputation: Mean
The mode method for numeric features to fill NA. Four options: mean, max, min, median. Default is mean.

Categorical Encoding Method: Indexer
Encoding method.

Maximum Null Value Percent: 80
Maximum Null Value Percent

Train Partition Percent: 80
The percentage of original data used for training, default is 80%.

Standardization: False
Standarize data before training.

Εικόνα 44: Παραμετροποίηση της στήλης – στόχου Total_Donations με χρήση του αλγορίθμου Elastic Net Linear Regression.

Μετά την εκτέλεση, μπορεί να ελεγχθεί το μοντέλο που μόλις δημιουργήθηκε στο μενού ML. Για να χρησιμοποιηθεί αυτό το μοντέλο, μπορεί να εισαχθεί ένα Σενάριο στο έργο και να χρησιμοποιηθεί με το μοντέλο που δημιουργήθηκε.

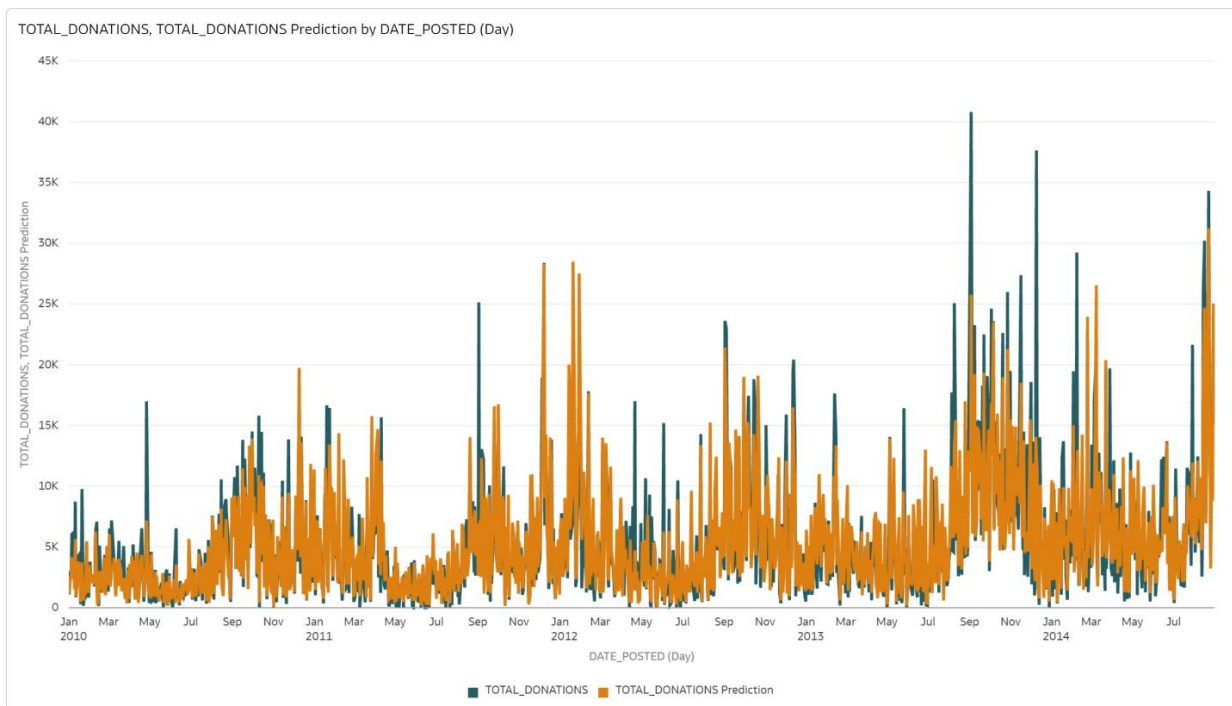
Συνολικές δωρεές Elastic 1 κατά ημερομηνία (ημέρα)

Το Elastic Net Regression αντιπροσωπεύει ένα μοντέλο παλινδρόμησης που συνδυάζει τα χαρακτηριστικά των μοντέλων Lasso και Ridge Regression. Αυτός ο συνδυασμός αποτελεί μια βελτιωμένη έκδοση της γραμμικής παλινδρόμησης, ιδιαίτερα σχεδιασμένη για να αντιμετωπίσει τις προκλήσεις που προκύπτουν από υψηλές διαστάσεις και συσχετίσεις μεταξύ των μεταβλητών.

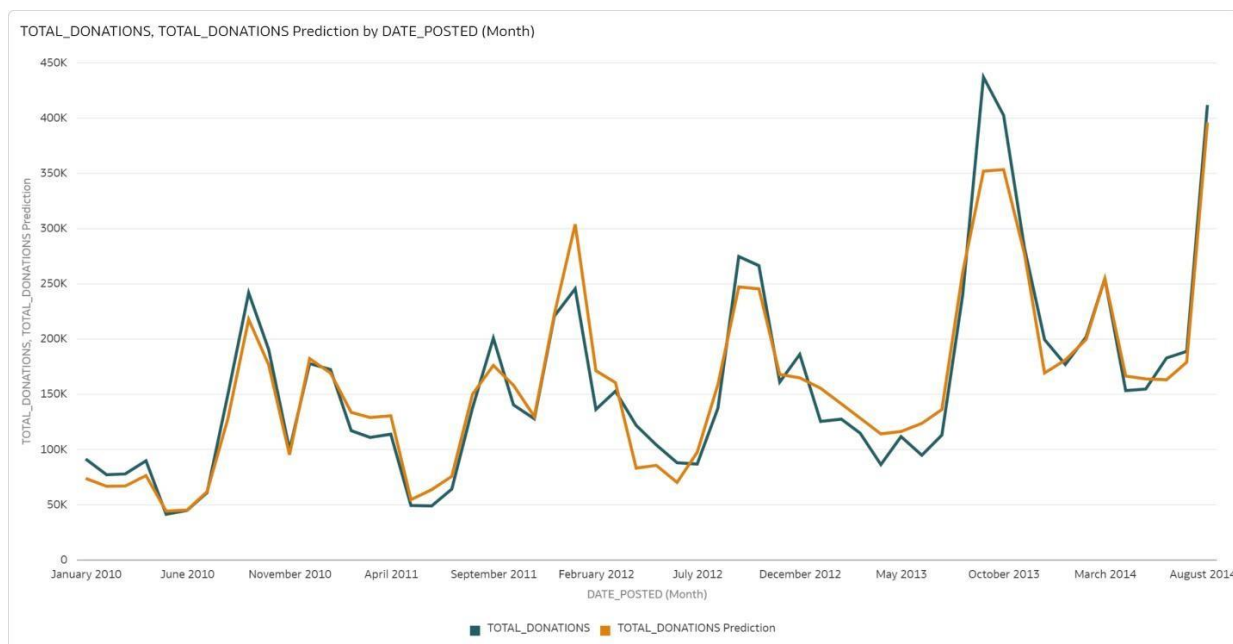
Το μοντέλο ElasticNet επιλύει το πρόβλημα της επιλογής χαρακτηριστικών (feature selection), επιτρέποντας την αυτόματη επιλογή των σημαντικών μεταβλητών και την απόρριψη των ασήμαντων. Επιπλέον, εξασφαλίζει την αυτόματη κανονικοποίηση των βαρών του μοντέλου, βοηθώντας στον έλεγχο της υπερπροσαρμογής (overfitting) και τη βελτίωση της γενικότητας των προβλέψεων.

Η μέθοδος Elastic Net λαμβάνει υπόψη την αναλογία των βαρών των όρων L1 (Lasso) και L2 (Ridge) κατά την εκτέλεση της παλινδρόμησης. Αυτό επιτρέπει την αποτελεσματική διαχείριση των προβλημάτων που σχετίζονται με την Lasso, που τείνει να επιλέγει μόνο ένα υποσύνολο από μεταβλητές, και τη Ridge, που διατηρεί όλες τις μεταβλητές. Με τη χρήση του Elastic Net, μπορεί να επιλεγεί ο κατάλληλος συνδυασμός των δύο όρων, προσαρμόζοντας το επίπεδο κανονικοποίησης και την επιλογή χαρακτηριστικών ανάλογα με το εκάστοτε πρόβλημα που αντιμετωπίζεται.

Στον τομέα της στατιστικής και της μηχανικής μάθησης, ο όρος "Lasso" (γνωστός επίσης ως "χειριστής ελάχιστης απόλυτης συρρίκνωσης και επιλογής" ή "LASSO") αναφέρεται σε μια μέθοδο παλινδρόμησης που συνδυάζει την επιλογή χαρακτηριστικών και την κανονικοποίηση, με σκοπό να βελτιώσει την ακρίβεια των προβλέψεων και την ερμηνευτικότητα του στατιστικού μοντέλου που προκύπτει. Αρχικά αναπτύχθηκε για γεωφυσικές εφαρμογές και αργότερα εισήχθη από τον Robert Tibshirani, ο οποίος δημιούργησε επίσης τον όρο "Elastic Net". [21]



Εικόνα 45: Σύγκριση αποτελεσμάτων πρόβλεψης και αρχικής ανάλυσης. Διάγραμμα συνολικών δωρεών και συνολικών δωρεών (πρόβλεψη) ανά ημέρα.



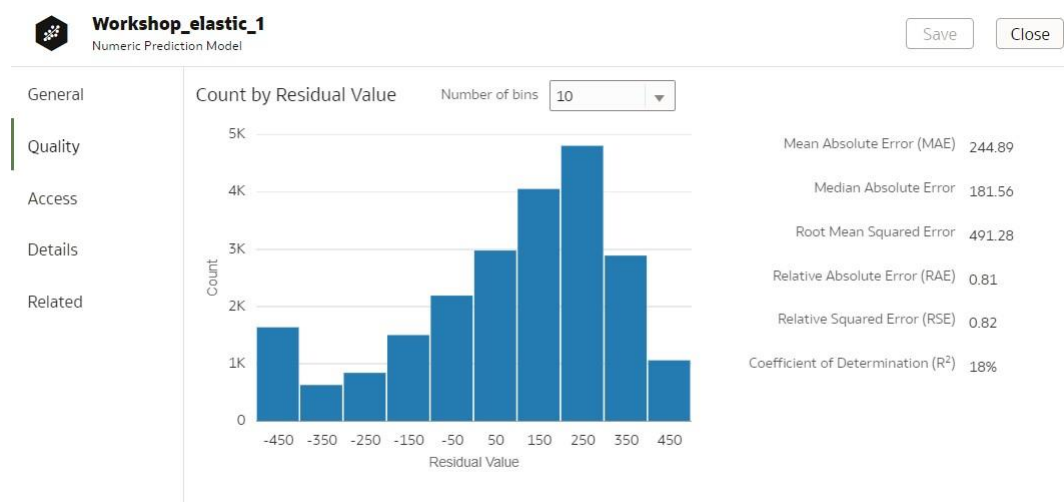
Εικόνα 46: Σύγκριση αποτελεσμάτων πρόβλεψης και αρχικής ανάλυσης. Διάγραμμα συνολικών δωρεών και συνολικών δωρεών (πρόβλεψη) ανά μήνα.

Η πορτοκαλί γραμμή στην εικόνα 46 δείχνει την πρόβλεψη σε σύγκριση με την πραγματική μπλε γραμμή για το επιλεγμένο μοντέλο. Από την ανάλυση του Σεναρίου, μπορούν να παρατηρηθούν ορισμένες λεπτομέρειες σχετικά με το μοντέλο. Είναι σημαντικό να σημειωθεί πως οι εισροές των μοντέλων μπορεί να διαφέρουν κατά την εκτέλεση του εργαστηρίου, καθώς η δειγματοληψία και η εκπαίδευση του μοντέλου δημιουργούν μοναδικά πλαίσια και συνθήκες.

Κατά την επιθεώρηση της ποιότητας του μοντέλου και των σχετικών συνόλων δεδομένων, η καρτέλα ποιότητας παρέχει σημαντικές πληροφορίες. Συγκεκριμένα, απεικονίζει την κατανομή των υπολειμματικών τιμών, δηλαδή τις διαφορές μεταξύ των πραγματικών τιμών και των προβλέψεων που παράγει το μοντέλο. Η κατανομή αυτή μας βοηθά να αξιολογήσουμε την ακρίβεια και την απόδοση του μοντέλου και να εντοπίσουμε πιθανά σημεία υπερπροσαρμογής ή υποπροσαρμογής.

Επιπλέον, η καρτέλα ποιότητας παρέχει τα συνολικά στατιστικά στοιχεία για τα δεδομένα του μοντέλου. Με βάση αυτά τα στοιχεία, μπορούμε να εξετάσουμε τη μέση τιμή, το εύρος, τις τυπικές αποκλίσεις και άλλες χρήσιμες μετρικές για να κατανοήσουμε την κατάσταση των δεδομένων μας.

Τα στατιστικά στοιχεία μας δίνουν σημαντικές πληροφορίες για τυχόν ανωμαλίες, εκτροπές ή ασυνήθιστα μοτίβα στα δεδομένα, που μπορεί να επηρεάζουν την ακρίβεια του μοντέλου μας. Κρίνοντας από την καρτέλα ποιότητας, μπορούμε να αναγνωρίσουμε τις αναγκαίες προσαρμογές που απαιτούνται στο μοντέλο μας, προκειμένου να βελτιώσουμε την απόδοση και την αξιοπιστία των προβλέψεών του. Οι αναλύσεις αυτές αποτελούν ουσιαστικό βήμα για την συνεχή βελτίωση και τον εκσυγχρονισμό του μοντέλου μας, προσφέροντας πιο αξιόπιστες και ακριβείς προβλέψεις.



Εικόνα 47: Στατιστικά ποιότητας των τιμών και της κατανομής τους για το μοντέλο Workshop_elastic_1.

Η σχετική καρτέλα παρουσιάζει διάφορα σύνολα δεδομένων που αναλύουν τα μεταδεδομένα (metadata) του μοντέλου. Μετά από αυτό, θα πραγματοποιήσουμε έλεγχο στο σύνολο δεδομένων προγραμμάτων. Σε αυτό το σημείο, μπορούμε να εντοπίσουμε ποιες στήλες επηρεάζουν το μοντέλο που εκπαιδεύσαμε, και συνεπώς, ποιοι παράγοντες είναι σημαντικοί για τις προβλέψεις του.

Είναι σημαντικό να σημειωθεί ότι τα προγράμματα οδήγησης μπορεί να διαφέρουν ελαφρώς κατά την εκτέλεση του εργαστηρίου, λόγω της διαφορετικής δειγματοληψίας και εκπαίδευσης που πραγματοποιούνται. Αυτό σημαίνει πως τα διαθέσιμα δεδομένα και οι παράμετροι της εκπαίδευσης μπορεί να δημιουργούν μοναδικότητα και διαφορές στην τελική μορφή των προγραμμάτων οδήγησης. Ωστόσο, αυτή η μοναδικότητα δεν επηρεάζει την αξιοπιστία των αποτελεσμάτων, καθώς οι αναλύσεις προβλέψεων και παραγωγής μοντέλων παραμένουν ακριβείς και αξιόπιστες για τις συγκεκριμένες συνθήκες και σύνολα δεδομένων.

Coefficient by Driver Name, Correlation

Driver Name	Correlation	Coefficient
RESOURCE_TYPE	Positive	34.63
GRADE_LEVEL	Positive	18.39
NUM_DONORS	Positive	16.23
SCH_METRO	Positive	6.08
TCHR_TEACH_FOR_AMERICA	Positive	5.60
PRIMARY_FOCUS_AREA	Positive	4.95
SECONDARY_FOCUS_AREA	Positive	4.86
SCH_LATITUDE	Positive	2.91
STUDENTS_REACHED	Positive	0.12
DATE_POSTED	Positive	0.08
SCH_CITY	Positive	0.00
SCH_CHARTER	Positive	0.00
SCHOOL_ID	Negative	0.00
SCH_STATEZIP	Negative	-0.01
DATE_COMPLETED	Negative	-0.02
SCH_COUNTY	Negative	-0.06
SCH_LONGITUDE	Negative	-1.41
ELIGIBLE_DOUBLE_IMPACT_MATCH	Negative	-2.16
SECONDARY_FOCUS_SUBJECT	Negative	-2.84
POVERTY_LEVEL	Negative	-4.55
TCHR_PREFIX	Negative	-4.73
PRIMARY_FOCUS_SUBJECT	Negative	-4.85
FUNDING_STATUS	Negative	-136.89

Εικόνα 48: Στήλες που έχουν ισχυρή επιρροή στο μοντέλο, θετική ή αρνητική. Στήλες που έχουν μικρή επιρροή στην πρόβλεψη του μοντέλου.

Κοιτάζοντας πιο προσεκτικά, παρατηρούμε ότι ορισμένες από αυτές τις στήλες δεν πρέπει να αντιμετωπίζονται ως μεταβλητές μοντέλου. Είτε επειδή δεν είναι ρεαλιστικές ως εισροές (Num_Donors, Students_reached), είτε είναι πολύ λεπτομερείς για την πρόβλεψή μας (Ημερομηνία, Αναγνωριστικό έργου, Πλάτος/Μήκος)

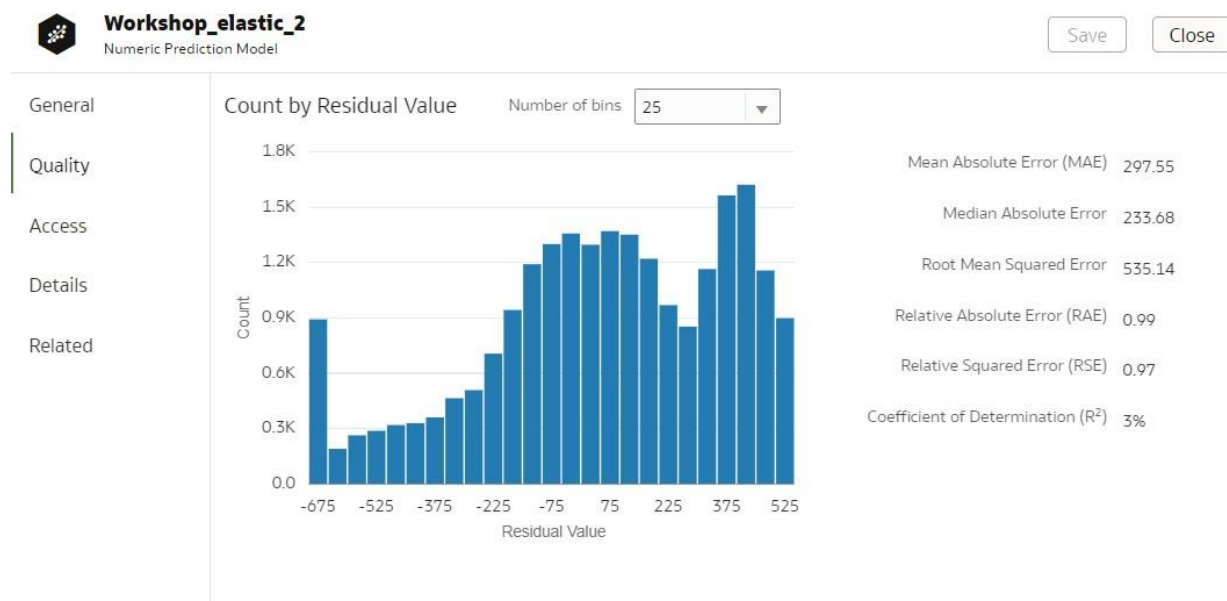
Driver Name	Correlation	Coefficient ▼
GRADE_LEVEL	Positive	18.39
NUM_DONORS	Positive	16.23
SCH_METRO	Positive	6.08
TCHR_TEACH_FOR_AMERICA	Positive	5.60
PRIMARY_FOCUS_AREA	Positive	4.95
SECONDARY_FOCUS_AREA	Positive	4.86
SCH_LATITUDE	Positive	2.91
STUDENTS_REACHED	Positive	0.12
DATE_POSTED	Positive	0.08
SCH_CITY	Positive	0.00
SCH_CHARTER	Positive	0.00
SCHOOL_ID	Negative	0.00
SCH_STATEZIP	Negative	-0.01
DATE_COMPLETED	Negative	-0.02
SCH_COUNTY	Negative	-0.06
SCH_LONGITUDE	Negative	-1.41
ELIGIBLE_DOUBLE_IMPACT_MATCH	Negative	-2.16
SECONDARY_FOCUS_SUBJECT	Negative	-2.84
POVERTY_LEVEL	Negative	-4.55
TCHR_PREFIX	Negative	-4.73
PRIMARY_FOCUS_SUBJECT	Negative	-4.85
FUNDING_STATUS	Negative	-136.89

Εικόνα 49: Επιλογή στηλών για εξαίρεση από το μοντέλο πρόβλεψης

Μετά την ανάλυση, αποφασίστηκε να εξαιρεθούν ορισμένες στήλες από το σύνολο δεδομένων και να εκπαιδευτεί μια νέα έκδοση του μοντέλου. Η διαδικασία αυτή θα γίνει εντός της ροής δεδομένων, όπου αρχικά εκπαιδεύτηκε το μοντέλο WorkshopElastic 1. Για να επιτευχθεί αυτό, προστέθηκε ένας κόμβος "επιλογή στηλών" αντίθετα προς το βήμα εκπαίδευσης, ώστε να αφαιρεθούν οι στήλες που δεν επιθυμούν να χρησιμοποιηθούν κατά την εκπαίδευση του μοντέλου.

Συγκεκριμένα, επιλέχθηκε η αφαίρεση των στηλών που περιλαμβάνουν λιγότερο σημαντικές πληροφορίες για το μοντέλο, όπως ο αριθμός έτους και μήνα. Αυτή η επιλογή αναμένεται να διασφαλίσει πιο συγκροτημένους υπολογισμούς και βελτιστοποιημένες προβλέψεις κατά την εκπαίδευση του μοντέλου. Με αυτόν τον τρόπο, οι πιο ασήμαντες παράμετροι δεν θα επηρεάσουν το μοντέλο, και θα επιτραπεί η έμφυτη κατανόηση των πιο σημαντικών παραγόντων που επηρεάζουν τις προβλέψεις. Οι βελτιώσεις αυτές αναμένεται να ενισχύσουν την ακρίβεια και την αποτελεσματικότητα του μοντέλου στην αριθμητική πρόβλεψη των αποτελεσμάτων.

Κατά τη διαδικασία της εκπαίδευσης του μοντέλου βρόχου και του συντονισμού, ελήφθη υπόψη η επιλογή των νέων στηλών που προτάθηκαν προηγουμένως. Προτού ξεκινήσει η εκπαίδευση, ελήφθη μέριμνα για να βεβαιωθούμε ότι αυτές οι νέες στήλες έχουν επιλεγεί με βάση τα κριτήρια που έχουν τεθεί, έτσι ώστε να παράσχουν τις πιο σημαντικές πληροφορίες για την εκπαίδευση του μοντέλου.



Εικόνα 50: Στατιστικά ποιότητας των τιμών και της κατανομής τους για το νέο μοντέλο Workshop_elastic_2.

Η κατανομή της υπολειμματικής αξίας φαίνεται να έχει ελαφρές διαφορές σε σύγκριση με το προηγούμενο μοντέλο. Με βάση αυτή την παρατήρηση, αποφασίστηκε να προχωρήσουμε στην προσθήκη ενός δεύτερου σεναρίου στο έργο μας, χρησιμοποιώντας το νέο μοντέλο.

Το νέο μοντέλο χρησιμοποιεί διαφορετικές στήλες ως εισόδους σε σχέση με το προηγούμενο, και επίσης απαιτεί δεδομένα για το Έτος και τον Μήνα. Προκειμένου να το χρησιμοποιήσουμε, θα πρέπει να δημιουργήσουμε αυτές τις στήλες στο σύνολο δεδομένων μας. Η προσθήκη αυτών των στηλών θα διαμορφώσει το σύνολο δεδομένων έτσι ώστε να περιλαμβάνει τις απαιτούμενες πληροφορίες που αναφέρονται στο νέο μοντέλο.

Είναι σημαντικό να σημειωθεί πως, κατά την εκτέλεση του Project, ενδέχεται να υπάρχουν μικρές διαφορές στα δεδομένα ή στις διαδικασίες, καθώς η δειγματοληψία και η εκπαίδευση δημιουργούν μοναδικότητα. Ωστόσο, με την κατάλληλη προσθήκη των απαιτούμενων στηλών στο σύνολο δεδομένων, εξασφαλίζεται η ομαλή εκτέλεση του νέου μοντέλου και η αξιόπιστη αξιοποίηση των δεδομένων για τις ανάγκες του εργαστηρίου.

Create Scenario - Map Your Data ✕

Select which Data Set you want to use with the Model

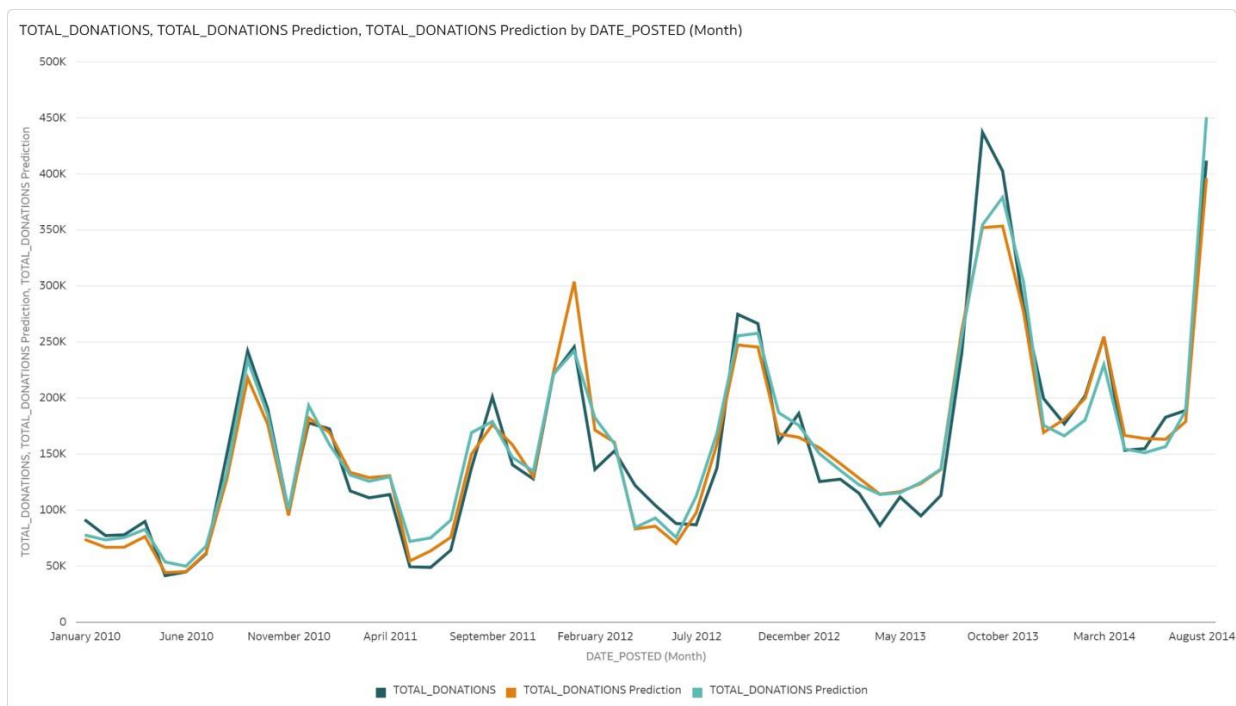
Data Set

For each model input listed on the left, select a corresponding data element from your workbook

Model Input	Map To
ELIGIBLE_DOUBLE_IMPACT_MATCH	* ELIGIBLE_DOUBLE_IMPACT_MA...
GRADE_LEVEL	* GRADE_LEVEL
Posted Month	* Calc Year Posted
Posted Year	* Calc Month Posted
POVERTY_LEVEL	* POVERTY_LEVEL
PRIMARY_FOCUS_AREA	* PRIMARY_FOCUS_AREA
PRIMARY_FOCUS_SUBJECT	* PRIMARY_FOCUS_SUBJECT
RESOURCE_TYPE	* RESOURCE_TYPE
SCH_CHARTER	* SCH_CHARTER

* Required Fields

Εικόνα 51: Επιλογή στηλών για χρήση στο μοντέλο πρόβλεψης



Εικόνα 52: Γράφημα με τις Συνολικές δωρεές, συνολικές προβλεπόμενες δωρεές (δύο περιπτώσεις πρόβλεψων) ανά μήνα δημοσίευσης.

Οι δύο προβλέψεις των μοντέλων παρουσιάζουν σημαντική ομοιότητα παρά το γεγονός ότι χρησιμοποιούν διαφορετικές μεταβλητές εισόδου. Προκειμένου να αναλύσουμε περαιτέρω τις πληροφορίες προγραμμάτων οδήγησης για το νέο μοντέλο που δημιουργήσαμε, είναι αναγκαίο να

εξετάσουμε τις δύο στήλες που προσθέσαμε στο σύνολο δεδομένων.

Coefficient by Driver Name, Correlation, Model Name, Target

Driver Name	Correlation	Model Name	Target	Coefficient
ELIGIBLE_DOUBLE_IMPACT_MATCH	Positive	Workshop_elastic_2	TOTAL_DONATIONS	0.00
GRADE_LEVEL	Positive	Workshop_elastic_2	TOTAL_DONATIONS	22.23
POVERTY_LEVEL	Negative	Workshop_elastic_2	TOTAL_DONATIONS	-4.58
PRIMARY_FOCUS_AREA	Positive	Workshop_elastic_2	TOTAL_DONATIONS	9.39
PRIMARY_FOCUS_SUBJECT	Negative	Workshop_elastic_2	TOTAL_DONATIONS	-5.54
Posted.Month	Positive	Workshop_elastic_2	TOTAL_DONATIONS	7.29
Posted.Year	Positive	Workshop_elastic_2	TOTAL_DONATIONS	18.13
RESOURCE_TYPE	Positive	Workshop_elastic_2	TOTAL_DONATIONS	31.88
SCHOOL_ID	Negative	Workshop_elastic_2	TOTAL_DONATIONS	0.00
SCH_CHARTER	Positive	Workshop_elastic_2	TOTAL_DONATIONS	2.12
SCH_CITY	Positive	Workshop_elastic_2	TOTAL_DONATIONS	0.01
SCH_COUNTY	Negative	Workshop_elastic_2	TOTAL_DONATIONS	-0.07
SCH_METRO	Positive	Workshop_elastic_2	TOTAL_DONATIONS	12.14
SCH_STATEZIP	Negative	Workshop_elastic_2	TOTAL_DONATIONS	-0.02
SECONDARY_FOCUS_AREA	Positive	Workshop_elastic_2	TOTAL_DONATIONS	3.20
SECONDARY_FOCUS_SUBJECT	Negative	Workshop_elastic_2	TOTAL_DONATIONS	-2.56
TCHR_PREFIX	Negative	Workshop_elastic_2	TOTAL_DONATIONS	-6.84
TCHR_TEACH_FOR_AMERICA	Positive	Workshop_elastic_2	TOTAL_DONATIONS	6.15

Εικόνα 53: Επιλογή στήλης για εξαίρεση από το μοντέλο πρόβλεψης

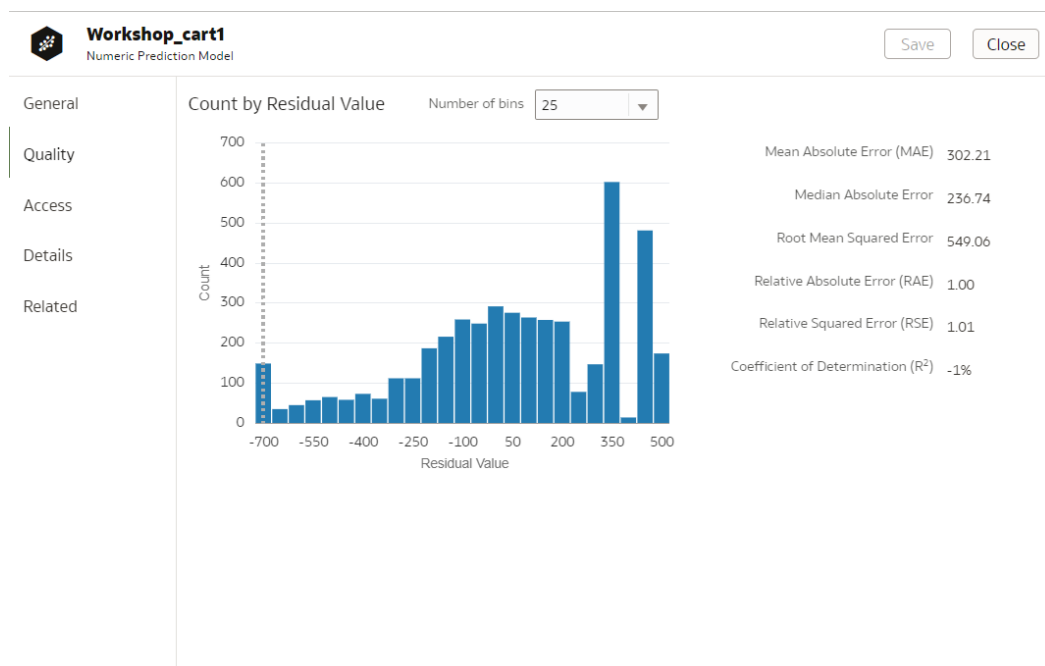
Σε αυτό το στάδιο, δημιουργείται ένα τρίτο μοντέλο πρόβλεψης, το οποίο διαφέρει από τα προηγούμενα μοντέλα καθώς χρησιμοποιεί διαφορετικό τύπο αλγορίθμου από το ElasticNet Regression. Συγκεκριμένα, αυτήν τη φορά επιλέγεται ένα Δέντρο Παλινδρόμησης Ταξινόμησης, το οποίο προσεγγίζει την πρόβλεψη των εξόδων με διαφορετικό τρόπο.

Το τρίτο μοντέλο που προτείνεται, εφαρμόζει τα δέντρα παλινδρόμησης ταξινόμησης προκειμένου να αξιολογήσει τις κατηγορίες ορισμένων εισόδων και να αντιστοιχίσει τιμές πρόβλεψης για κάθε κατηγορία. Σε αντίθεση με το ElasticNet Regression, το οποίο χρησιμοποιεί γραμμική παλινδρόμηση, το Δέντρο Παλινδρόμησης Ταξινόμησης βασίζεται στην ανάλυση της δομής των δεδομένων και την κατηγοριοποίηση τους.

Κάθε ένας από τους διαφορετικούς τύπους μοντέλων έχει τα πλεονεκτήματα και τις περιορισμένες τους δυνατότητες, και η σύγκρισή τους θα βοηθήσει να κατανοήσουμε καλύτερα την απόδοση και την ακρίβεια της πρόβλεψής τους για τα δεδομένα μας. Η επιλογή του κατάλληλου μοντέλου εξαρτάται από τα χαρακτηριστικά των δεδομένων, τον στόχο της πρόβλεψης και τις απαιτήσεις του συγκεκριμένου προβλήματος.

Select Train Numeric Prediction Model Script

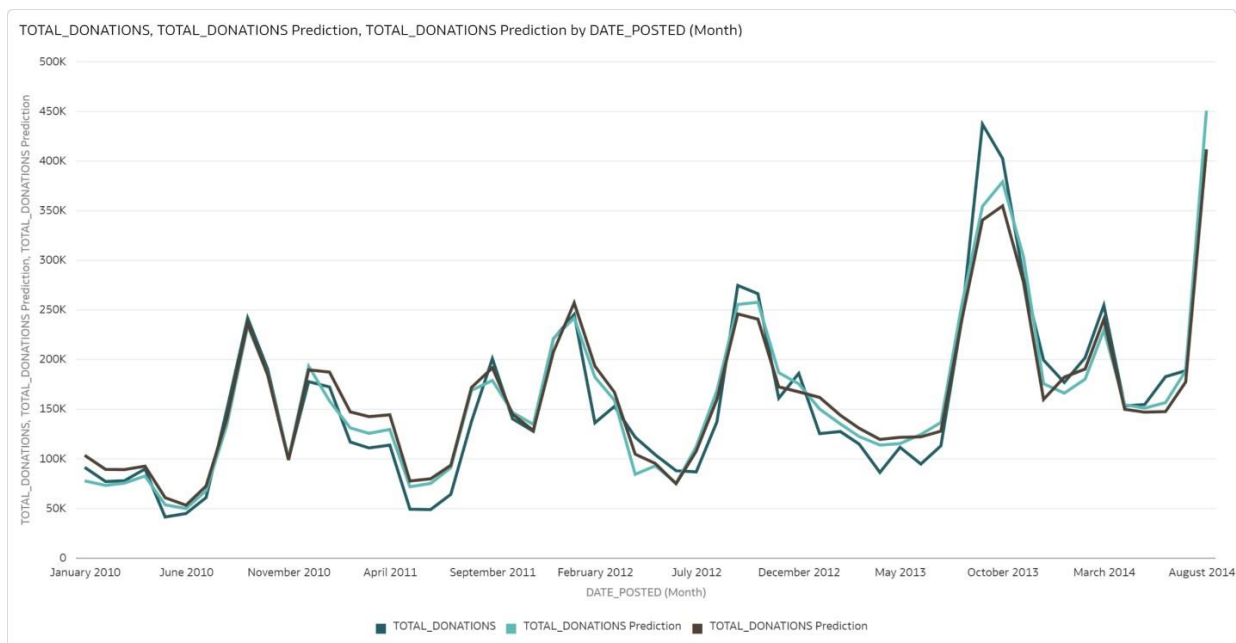




Εικόνα 54: Στατιστικά ποιότητας των τιμών και της κατανομής τους για το νέο μοντέλο Workshop_cart1.

Πρόκειται να προστεθεί ένα νέο σενάριο στο έργο, το οποίο θα βασίζεται στο μοντέλο καλαθιού. Μετά την προσθήκη αυτού του νέου σεναρίου, το πρώτο σενάριο, που είναι παρόμοιο με το δεύτερο, μπορεί να αφαιρεθεί για να μειωθεί ο υπολογιστικός φόρτος. Το νέο μοντέλο χρησιμοποιεί λιγότερες μεταβλητές και εισροές για να δημιουργήσει ένα δέντρο ταξινόμησης των εγγραφών.

Η πρόσθεση αυτού του σεναρίου αποτελεί μια πρακτική λύση, καθώς μειώνει τις απαιτήσεις των εισροών και απλοποιεί το μοντέλο, παρέχοντας ταυτόχρονα ακριβή προβλέψεις. Εάν το νέο μοντέλο επιδειχθεί ακριβές και αξιόπιστο στις προβλέψεις των δωρεών, θα μπορούσε να είναι πιο πρακτική επιλογή για την χρήση στον τομέα των προβλέψεων, καθώς απαιτεί λιγότερους υπολογιστικούς πόρους και προσφέρει αποδοτική χρήση των δεδομένων. Ωστόσο, η αξιολόγηση και η σύγκριση του νέου μοντέλου με τα προηγούμενα μοντέλα θα είναι απαραίτητη για να διαπιστωθεί η βέλτιστη επίδοση και να ληφθούν ενδεχόμενες αποφάσεις σχετικά με την τελική επιλογή του μοντέλου πρόβλεψης που θα χρησιμοποιηθεί.

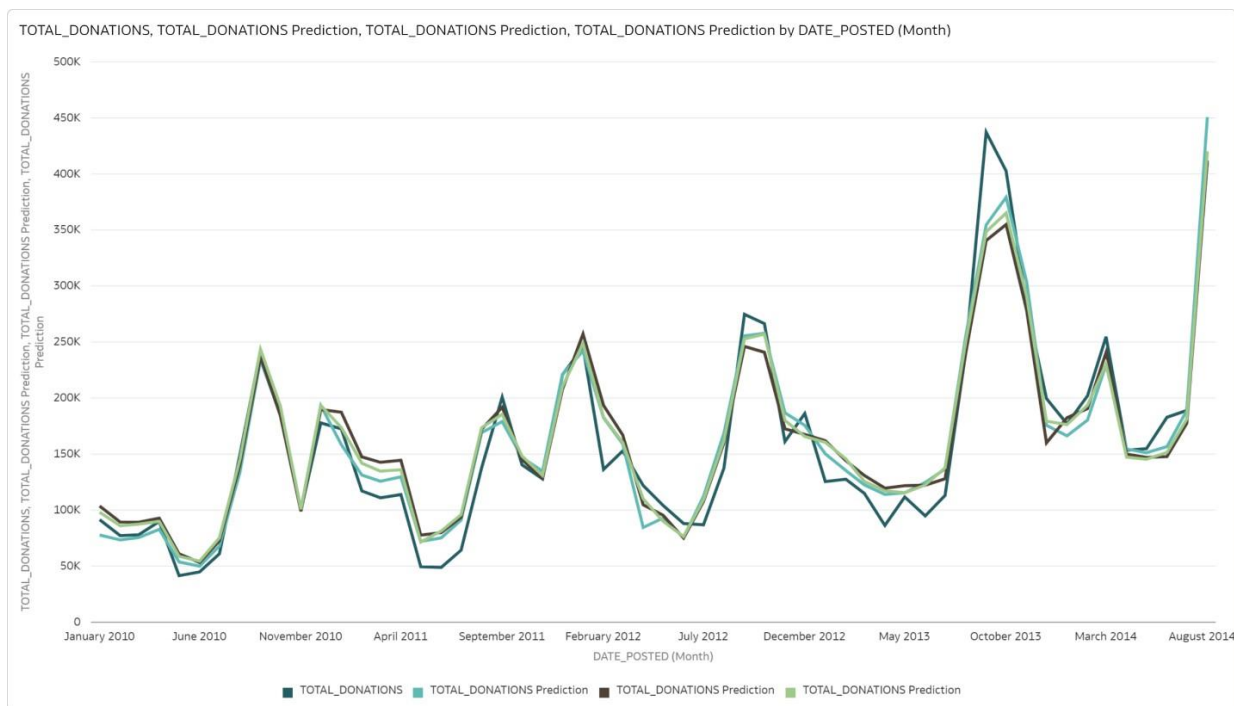


Εικόνα 55: Γράφημα με τις Συνολικές δωρεές, συνολικές προβλεπόμενες δωρεές CART και πρόβλεπόμενη δωρεά Elastic Net ανά μήνα δημοσίευσης.

Το νέο μοντέλο CART φαίνεται εξίσου ακριβές με τα Elastic Net Regression μοντέλα πρόβλεψης που εξετάστηκαν παραπάνω.

Για να βελτιωθεί το μοντέλο δέντρων παλινδρόμησης ταξινόμησης, αποφασίστηκαν αλλαγές στη ροή δεδομένων εκπαίδευσης. Ειδικότερα, αυξήσαμε την τιμή της παραμέτρου MAXDepth, που αντιπροσωπεύει το μέγιστο αριθμό διαδοχικών επιπέδων διακλάδωσης που επιτρέπονται στο δέντρο. Με αυτήν την αλλαγή, το νέο μοντέλο CART τώρα λειτουργεί με βάθος 7 επιπέδων, αντί των προηγούμενων 5 επιπέδων.

Τα αποτελέσματα δείχνουν ότι οι προβλέψεις από το νέο μοντέλο CART βάθους 7 επιπέδων είναι συγκρίσιμες με τις προηγούμενες, παρόλο που υιοθετεί μεγαλύτερο βάθος. Η αύξηση του βάθους επιτρέπει στο δέντρο να κάνει περισσότερες διακλαδώσεις και να λαμβάνει υπόψη περισσότερες λεπτομέρειες των δεδομένων, προσαρμόζοντας τις προβλέψεις του ανάλογα. Ωστόσο, η επιλογή του βάθους εξαρτάται από το συγκεκριμένο πρόβλημα και τον τύπο των δεδομένων, επομένως, μπορεί να απαιτηθούν περαιτέρω δοκιμές και αξιολογήσεις πριν αποφασιστεί η τελική ρύθμιση του μοντέλου.

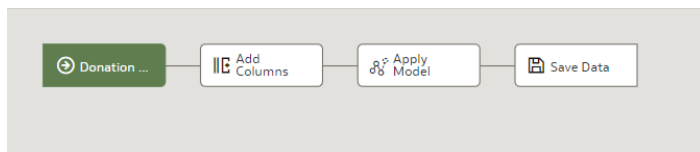


Εικόνα 56: Γράφημα με τις Συνολικές δωρεές, συνολικές προβλεπόμενες δωρεές CART και Elastic Net (δύο περιπτώσεις) ανά μήνα δημοσίευσης.

Στο τελικό στάδιο του εργαστηρίου, χρησιμοποιείται ένα εκπαιδευμένο μοντέλο για την πρόβλεψη σε ένα σύνολο δεδομένων. Για να επιτευχθεί αυτό, δημιουργείται μια ροή δεδομένων που εφαρμόζει το προβλεπτικό μοντέλο στα δεδομένα και στη συνέχεια αποθηκεύει τα αποτελέσματα.

Για λόγους απλότητας, στο παρόν εργαστήριο, το μοντέλο εφαρμόζεται στο ίδιο δείγμα δεδομένων που χρησιμοποιήθηκε για την εκπαίδευσή του. Αυτό επιτυγχάνεται προσθέτοντας έναν κόμβο εφαρμογής μοντέλου στη ροή δεδομένων και επιλέγοντας το προτιμώμενο μοντέλο από τον προηγούμενο συντονισμό.

Κατόπιν, επιλέγονται οι στήλες εξόδου που θα παραμείνουν στο τελικό σύνολο δεδομένων, καθώς και οι στήλες εισόδου που απαιτούνται για τη λειτουργία του μοντέλου. Παρόμοια με τα προηγούμενα σενάρια, εφαρμόζεται μια νέα ροή δεδομένων με τη χρήση του μοντέλου CART 1, δημιουργώντας ένα νέο επίπεδο ανάλυσης του συνόλου δεδομένων.

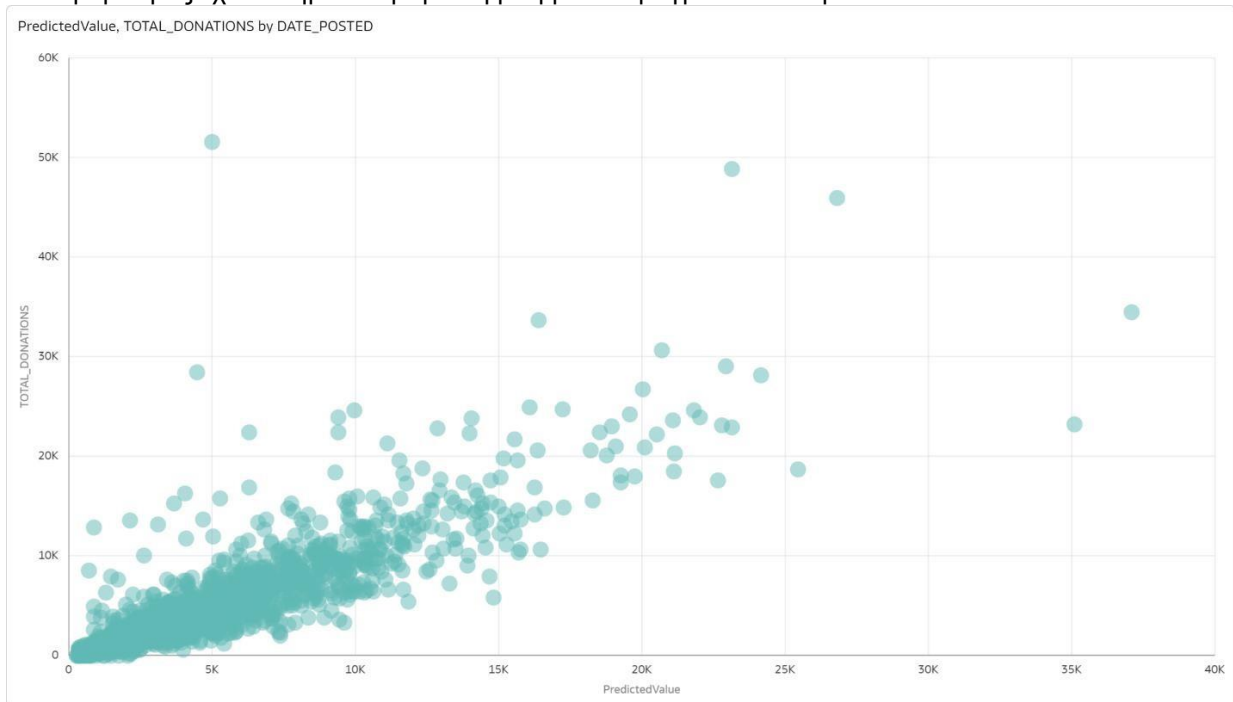


Εικόνα 57: Ροή δεδομένων πρόβλεψης

Μετά την εκτέλεση, δημιουργείται ένα νέο dataset με την ονομασία "WorkshopScoring". Αυτό το νέο σύνολο δεδομένων περιλαμβάνει τα αποτελέσματα των προβλέψεων και τις πραγματικές μετρήσεις δωρεών. Για να πραγματοποιήσουμε μια σύγκριση μεταξύ των προβλέψεων και των πραγματικών τιμών, ορίζουμε έναν κανόνα συγκέντρωσης (μπορεί να είναι, για παράδειγμα, μια

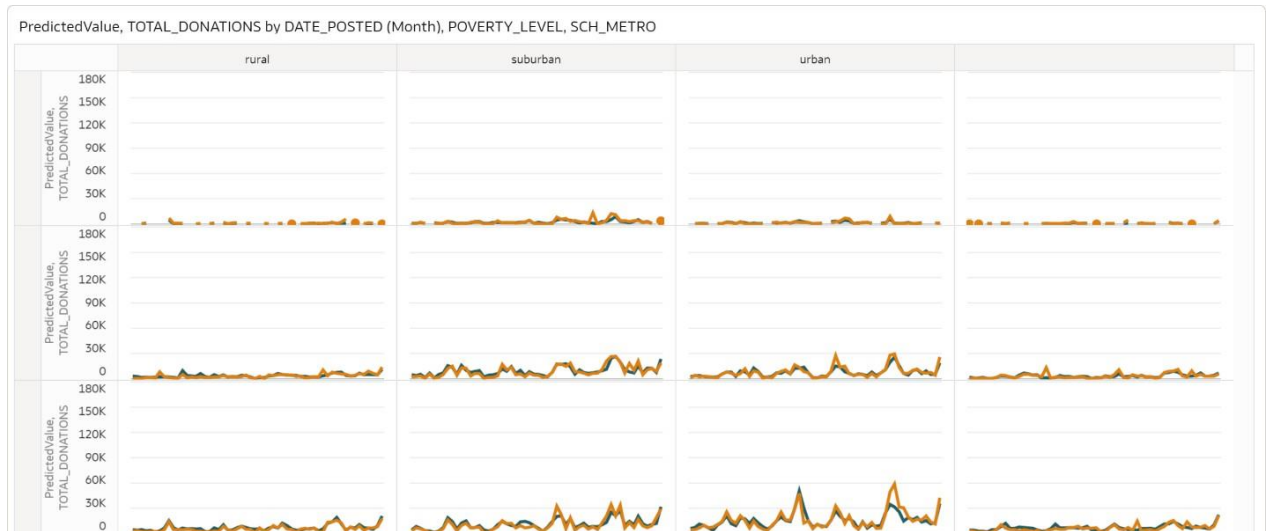
μετρική αξιολόγησης όπως η απόλυτη διαφορά μεταξύ των προβλέψεων και των πραγματικών τιμών).

Έχοντας το νέο προγνωστικό μέτρο στη διάθεσή μας, μπορούμε τώρα να εκτελέσουμε όλες τις αναλύσεις και τις οπτικοποιήσεις που απαιτούν τα δεδομένα πρόβλεψης, συμπεριλαμβανομένων και άλλων χαρακτηριστικών μετρήσεων συνόλων δεδομένων. Αυτό μας επιτρέπει να αξιολογήσουμε την απόδοση του μοντέλου και να εξάγουμε συμπεράσματα για την ακρίβεια των προβλέψεων. Παρακάτω παρουσιάζονται γραφήματα που αντιπροσωπεύουν τη σύγκριση μεταξύ των προβλεπόμενων τιμών δωρεών και των πραγματικών δωρεών. Τα γραφήματα αποδεικνύουν ότι οι προβλέψεις έχουν σημαντική προσέγγιση με τα πραγματικά δεδομένα.

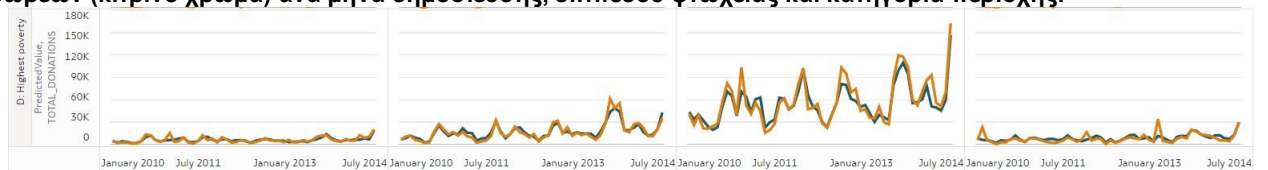


Εικόνα 58: Συγκέντρωση δεδομένων πρόβλεψης δωρεών σε αντίθεση με τον πραγματικό αριθμό δωρεών.

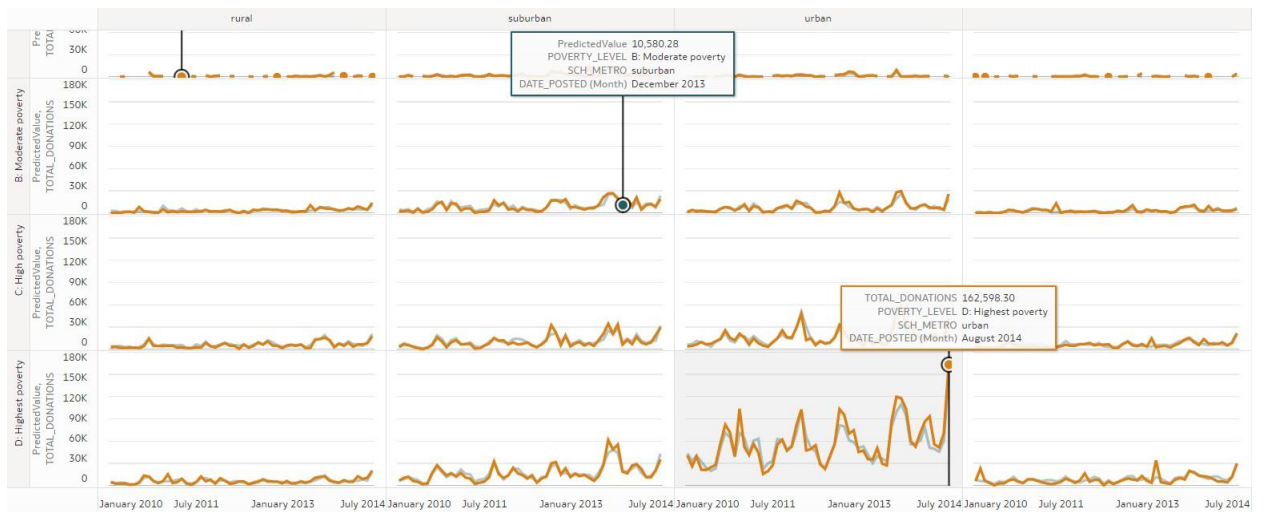
Το συγκεκριμένο Scatterplot παρουσιάζει την συγκέντρωση με τα δεδομένα πρόβλεψης δωρεών σε συνδυασμό με τα πραγματικά δεδομένα του Total_Donations.



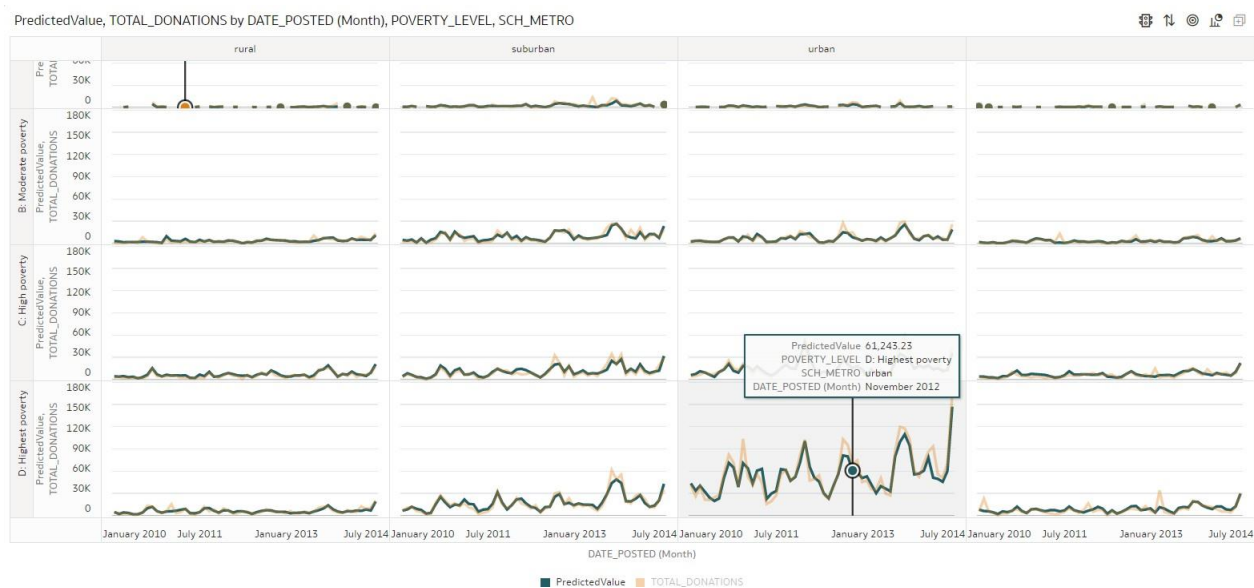
Εικόνα 59: Παρουσίαση πολλαπλών γραφημάτων προβλέψεων δωρεών (μαύρο χρώμα) και αριθμού δωρεών (κίτρινο χρώμα) ανά μήνα δημοσίευσης, επιπέδου φτώχειας και κατηγορία περιοχής.



Εικόνα 60: Παρουσίαση πολλαπλών γραφημάτων σύγκρισης προβλέψεων δωρεών και πραγματικών δωρεών ανά μήνα δημοσίευσης στο χαμηλότερο επίπεδο φτώχειας και ανά κατηγορία περιοχής.



Εικόνα 61: Παρουσίαση πολλαπλών γραφημάτων συνολικών δωρεών ανά μήνα δημοσίευσης, επιπέδου φτώχειας και κατηγορία περιοχής.



Εικόνα 62: Παρουσίαση πολλαπλών γραφημάτων προβλέψεων δωρεών ανά μήνα δημοσίευσης, επιπέδου φτώχειας και κατηγορία περιοχής.

Ολοκληρώνοντας το σενάριο πρόβλεψης στα σύνολα δεδομένων που εξετάσαμε, παρατηρούμε ότι τα αποτελέσματα των προβλέψεων προσεγγίζουν αρκετά τα πραγματικά δεδομένα με μια μικρή φυσιολογική απόκλιση.

8. Συμπεράσματα

Η μεταπτυχιακή διατριβή παρουσιάζει μια συστηματική και λεπτομερή μελέτη για το υπολογιστικό νέφος και την αναλυτική δεδομένων, εξετάζοντας εκτενώς τις τάσεις, προκλήσεις και οφέλη που παρέχονται στον κλάδο της επιχειρηματικότητας. Η διατριβή ξεκινά με μια λεπτομερή εισαγωγή στον κόσμο του υπολογιστικού νέφους, αναδεικνύοντας τους διάφορους τύπους νεφών (δημόσιο, ιδιωτικό, υβριδικό) και τα χαρακτηριστικά τους.

Ένα κύριο κομμάτι της διατριβής επικεντρώνεται στην μηχανική μάθηση, αναδεικνύοντας τις αρχές της, τα οφέλη που προσφέρει και την επίλυση προβλημάτων μέσω αυτής της τεχνολογίας στο υπολογιστικό νέφος. Επιπλέον, αναλύει εκτενώς την πλατφόρμα Oracle Cloud και τα πλεονεκτήματά της σε σύγκριση με άλλες πλατφόρμες, όπως η Amazon, Google και Microsoft.

Η μεταπτυχιακή διατριβή παρουσιάζει με περιεκτικότητα το υπολογιστικό νέφος και την αναλυτική δεδομένων, με εξισορροπημένη έμφαση στην πλατφόρμα Oracle Cloud. Μέσω της μηχανικής μάθησης και των προγνωστικών μοντέλων, εξετάζεται λεπτομερώς η δυνατότητα ανάλυσης δεδομένων με υψηλή ακρίβεια, χρησιμοποιώντας τα προηγμένα εργαλεία και τις δυνατότητες που προσφέρει η πλατφόρμα Oracle Cloud για τη δημιουργία ενημερωμένων ροών δεδομένων και την αξιοποίηση των αποτελεσμάτων.

Οι εκτενείς παρατηρήσεις και αναλύσεις που πραγματοποιούνται στο σενάριο ανάλυσης δεδομένων αποκαλύπτουν την αξία και την αποτελεσματικότητα της μηχανικής μάθησης στην αντιμετώπιση προβλημάτων και την πρόβλεψη μελλοντικών τάσεων. Αυτά τα ευρήματα προσφέρουν πολύτιμες πληροφορίες και νέα δεδομένα για τον τρόπο που η μηχανική μάθηση μπορεί να αξιοποιηθεί στον κόσμο της ανάλυσης δεδομένων και των προβλέψεων στον επιχειρηματικό τομέα. Μέσω αυτών των ευρημάτων, ανοίγονται νέες δυνατότητες για τη βελτίωση της λήψης αποφάσεων, την ανάπτυξη πιο αποτελεσματικών στρατηγικών και την επίτευξη της ανταγωνιστικότητας στον επιχειρηματικό κόσμο. Η μηχανική μάθηση αποδεικνύεται καθοριστική για την πρόοδο και την εξέλιξη της επιχειρηματικότητας και της ανάλυσης δεδομένων.

Βιβλιογραφία

1. Hillbert M, Lopez P (2011) The world's technological capacity to store, communicate and compute information. *Science* III:62–65
2. J. Hellerstein, "Gigaom Blog," 2019. Available: <https://gigaom.com/2008/11/09/mapreduce-leads-the-way-for-parallel-programming/>. Accessed 20 Jan 2021
3. Statista, "Statista," 2020. Available: <https://www.statista.com/statistics/871513/worldwide-data-created/>. Accessed 21 Jan 2021
4. Reinsel D, Gantz J, Rydning J (2017) *Data age 2025: the evolution of data to-life critical*. International Data Corporation, Framingham
5. Forbes, "Forbes", 2020. Available: <https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mindblowing-stats-everyone-should-read/?sh=5936b00460ba>
6. Big data analytics in Υπολογιστική Νέφος: an overview Blend Berisha, Endrit Mëziu and Isak Shabani. Available: [Big data analytics in Υπολογιστική Νέφος: an overview | Journal of Υπολογιστική Νέφος | Full Text \(springeropen.com\)](#)
7. Yadav S, Sohal A (2017) Review paper on big data analytics in Υπολογιστική Νέφος. *Int J Comp Trends Technol (IJCTT)* IX. 49(3);156-160
8. [Data-flow diagram - Wikipedia](#)
9. Kalliopi Kanaki (April 2015): (4) (PDF) [ΥΠΟΛΟΓΙΣΤΙΚΟ ΝΕΦΟΣ \(researchgate.net\)](#)
10. [Υπολογιστικό νέφος - Βικιπαίδεια \(wikipedia.org\)](#)
11. [What Is Αναλυτική Νέφος? How it Works & Benefits Businesses | NetSuite](#)
12. Celestino Güemes, Jordan Janeczko, Thierry Caminel and Matthew Roberts bill, Atos Data Analytics as a Service: unleashing the power of Cloud and Big Data.
13. SAS: Analytics, Business Intelligence and Data Management, http://www.sas.com/en_in/home.html
14. Big data Analytics <http://searchbusinessanalytics.techtarget.com/definition/big-data-analytics>.
15. Affreen Ara, Aftab Ara, Cloud for Big Data Analytics Trends (May 2016), *IOSR Journal of Computer Engineering*(4) (PDF) [Cloud for Big Data Analytics Trends \(researchgate.net\)](#)
16. Machine Learning Overview: [Machine Learning Overview \(oracle.com\)](#)
17. Why Oracle Cloud Infrastructure over Amazon Web Services: [Oracle Cloud \(OCI\) vs AWS](#)
18. Why Oracle Cloud Infrastructure over Google Cloud Platform: [Oracle Cloud vs Google Cloud](#)
19. OCI over Microsoft Azure: [Oracle Cloud vs Microsoft Azure](#)
20. [Visualizing Data and Building Reports in Oracle Analytics Cloud](#)
21. Hui Zou and Trevor Hastie, Regularization and Variable Selection via the Elastic Net, *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, Vol. 67, No. 2 (2005), pp. 301-320.
22. (PDF) Υπολογιστική Νέφος: A review of the Concepts and Deployment Models. https://www.researchgate.net/publication/317413701_Cloud_Computing_A_review_of_the_Concepts_and_Deployment_Models (accessed Jan. 03, 2023).
23. The Role of Υπολογιστική Νέφος in the Development of Information Systems for SMEs, IBIMA Publishing. <https://ibimapublishing.com/articles/JCC/2017/736545/> (accessed Jan. 03, 2023).

24. R. O. Aburukba, M. AliKarrar, T. Landolsi, and K. El-Fakih, —Scheduling Internet of Things requests to minimize latency in hybrid Fog—Υπολογιστική Νέφος, *Future Gener. Comput. Syst.*, vol. 111, pp. 539–551, Oct. 2020, doi: 10.1016/j.future.2019.09.039.
25. An IoT-based task scheduling optimization scheme considering the deadline and cost-aware scientific workflow for Υπολογιστική Νέφος | *EURASIP Journal on Wireless Communications and Networking* | Full Text. <https://jwcneurasipjournals.springeropen.com/articles/10.1186/s13638-019-1557-3> (accessed Jan. 03, 2023).
26. (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 14, No. 3, 2023 329 | Page www.ijacsa.thesai.org Current Development, Challenges and Future Trends in Υπολογιστική Νέφος: A Survey Hazzaa N. Alshareef College of Computing and Informatics, Saudi Electronic University, Riyadh, Kingdom of Saudi Arabia.
27. D. Widyastuti and I. Irwansyah, —Benefits And Challenges Of Υπολογιστική Νέφος Technology Adoption In Small And Medium Enterprises (SMEs), Jan. 2018. doi: 10.2991/bcm-17.2018.46.
28. S. A. Bello et al., —Υπολογιστική Νέφος in construction industry: Use cases, benefits and challenges, *Autom. Constr.*, vol. 122, p. 103441, Feb. 2021, doi: 10.1016/j.autcon.2020.103441.
29. T. Vasiljeva, S. Shaikhulina, and K. Kreslins, —Υπολογιστική Νέφος: Business Perspectives, Benefits and Challenges for Small and Medium Enterprises (Case of Latvia), *Procedia Eng.*, vol. 178, pp. 443–451, Jan. 2017, doi: 10.1016/j.proeng.2017.01.087.
30. A. Aljumah and T. A. Ahanger, —Cyber security threats, challenges and defence mechanisms in Υπολογιστική Νέφος, *IET Commun.*, vol. 14, no. 7, pp. 1185–1191, 2020, doi: 10.1049/iet-com.2019.0040.
31. M. Humayun, —Role of Emerging IoT Big Data and Υπολογιστική Νέφος for Real Time Application, *Int. J. Adv. Comput. Sci. Appl. IJACSA*, vol. 11, no. 4, Art. no. 4, Jun. 2020, doi: 10.14569/IJACSA.2020.0110466.
32. M. Ghobakhloo, —Industry 4.0, digitization, and opportunities for sustainability, *J. Clean. Prod.*, vol. 252, p. 119869, Apr. 2020, doi: 10.1016/j.jclepro.2019.119869.
33. T. Masood and P. Sonntag, —Industry 4.0: Adoption challenges and benefits for SMEs, *Comput. Ind.*, vol. 121, p. 103261, Oct. 2020, doi: 10.1016/j.compind.2020.103261.
34. [Machine Learning in the Cloud: Complete Guide \[2023\] \(run.ai\)](#)