



**ΙΟΝΙΟ
ΠΑΝΕΠΙΣΤΗΜΙΟ**

ΔΙΑΤΜΗΜΑΤΙΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ

**“Ψηφιακές Τεχνολογίες στη Διοίκηση της Φιλοξενίας και του
Τουρισμού”**

Τμήμα Πληροφορικής - Τμήμα Τουρισμού

**ΑΝΑΠΤΥΞΗ ΜΟΝΤΕΛΩΝ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ ΓΙΑ ΤΗΝ
ΔΙΑΧΕΙΡΙΣΗ ΚΡΑΤΗΣΕΩΝ ΣΕ ΞΕΝΟΔΟΧΕΙΑΚΕΣ ΜΟΝΑΔΕΣ**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Της φοιτήτριας

ΜΑΖΙΩΤΗ ΘΕΟΔΩΡΑΣ

(ΑΜ: trs.dit2221)

ΤΡΙΜΕΛΗΣ ΕΞΕΤΑΣΤΙΚΗ ΕΠΙΤΡΟΠΗ:

1. Έξαρχος Θεμιστοκλής – Επιβλέπων
2. Κερμανίδου Κάτια – Λήδα
3. Βραχάτης Αριστείδης

Σεπτέμβριος 2024

Την παρούσα διπλωματική εργασία αφιερώνω στην οικογένειά μου και, ιδιαίτερα, στο παιδί μου. Εύχομαι μέσα από τη δική μου προσπάθεια να διδαχθεί την αξία της διαρκούς μάθησης και της προσωπικής εξέλιξης, σύμφωνα με το ρητό «Γηράσκω αεί διδασκόμενος».

Περιεχόμενα

1. Abstract	4
2. Λέξεις – Κλειδιά / Keywords	4
3. Εισαγωγή	5
4. Βιβλιογραφική Επισκόπηση	6
5. Περιγραφή Δείγματος.....	23
6. Προεπεξεργασία Δεδομένων.....	38
7. Μεθοδολογία.....	45
8. Αποτελέσματα Πειραμάτων	46
9. Συζήτηση – Συμπεράσματα Έρευνας.....	59
10. Γενικά συμπεράσματα και Προτάσεις	64
11. Προτάσεις για μελλοντική έρευνα	66
12. Βιβλιογραφία	67

1. Abstract

Οι ακυρώσεις των κρατήσεων έχουν σημαντικό αντίκτυπο στις αποφάσεις διαχείρισης των διοικήσεων των ξενοδοχειακών μονάδων. Για την αντιμετώπιση των επιπτώσεων των ακυρώσεων, οι διοικήσεις εφαρμόζουν αυστηρές πολιτικές ακύρωσης και υπερκρατήσεων, γεγονός το οποίο αδιαμφισβήτητα μπορεί να έχει αρνητικό αντίκτυπο στις μελλοντικές κρατήσεις, τα έσοδα αλλά και τη φήμη του ξενοδοχείου.

Σκοπός της παρούσας εργασίας είναι αφενός η ανάδειξη ενός αποτελεσματικού και αξιόπιστου μοντέλου πρόβλεψης με την χρήση της μηχανικής μάθησης για την πρόβλεψη των ακυρώσεων των κρατήσεων στις Ξενοδοχειακές μονάδες με όσο το δυνατόν μεγαλύτερη ακρίβεια και αφετέρου η ανάδειξη εκείνων των παραγόντων που επηρεάζουν την ακύρωση της κράτησης.

Για την ανάδειξη του μοντέλου η παρούσα μελέτη ανέλυσε δευτερογενή δεδομένα κρατήσεων 2 ξενοδοχείων της Πορτογαλίας χρησιμοποιώντας τους αλγόριθμους IBK, J48 και Random Forest και την διασταυρούμενη επικύρωση 10 fold Cross Validation. Το μοντέλο που αναδείχθηκε στην παρούσα εργασία κρίνεται πολύ αποτελεσματικό καθώς έφθασε 90,39% στην ακρίβεια ενώ ταυτόχρονα με την βοήθεια του αλγόριθμου J48 παρουσιάζονται τα χαρακτηριστικά βαρύνουσας σημασίας για την ακύρωση της κράτησης.

Με την ανάδειξη ενός τέτοιου μοντέλου δύνανται τα ξενοδοχεία να προβλέπουν τα ποσοστά ακύρωσης και να ελαχιστοποιήσουν τις επιπτώσεις λανθασμένων πολιτικών και είναι δυνατόν να επιτρέψει στα ξενοδοχεία να προβλέψουν την καθαρή ζήτησή τους και έτσι να λάβουν καλύτερες αποφάσεις σχετικά με το ποιες κρατήσεις τελικά θα πρέπει να δεχτούν και ποιες να απορρίψουν, σχετικά με τις τιμολογήσεις των δωματίων αλλά και τα ποσοστά υπερκρατήσεων.

2. Λέξεις – Κλειδιά / Keywords

Decision support systems, predictive analytics,
Machine learning, Cancellation prediction, Revenue Management
Big data
Predictive modelling
Hotel booking cancellation

3. Εισαγωγή

Ο τουρισμός είναι ένας από τους ταχύτερα αναπτυσσόμενους κλάδους στον κόσμο και η σημασία του στην παγκόσμια οικονομία είναι αδιαμφισβήτητος. Ωστόσο, το περιβάλλον όπου το «φαινόμενο» τουρισμός αναπτύσσει τη δραστηριότητά του διακατέχεται από συνεχή αβεβαιότητα και διαρκώς μεταβαλλόμενες συνθήκες. Υπό αυτήν την έννοια, οι ξενοδοχειακές μονάδες διαρκώς «απειλούνται» από πολλούς εξωτερικούς παράγοντες που μπορεί να έχουν σημαντικό αντίκτυπο στην λειτουργία τους και στο μέλλον τους.

Μέσα σε αυτό το περιβάλλον αβεβαιότητας και διαρκούς μεταβλητότητας κατά το οποίο η ζήτηση μεταβάλλεται συνεχώς, η επιτυχημένη διαχείριση κρατήσεων αποτελεί τον πιο σημαντικό και κρίσιμο παράγοντα βιωσιμότητας και κερδοφορίας για τις ξενοδοχειακές μονάδες. Η ικανότητα ενός ξενοδοχείου να προβλέπει με ακρίβεια τις ακυρώσεις είναι καίριας σημασίας για τη βελτίωση της πληρότητας, την αποφυγή ακυρώσεων και τη μεγιστοποίηση των εσόδων.

Μέχρι πρόσφατα οι ξενοδοχειακές μονάδες υιοθετούσαν τις παραδοσιακές μεθόδους διαχείρισης των κρατήσεων χρησιμοποιώντας τα δεδομένα τους και την πρόβλεψη με την βοήθεια της στατιστικής. Και ενώ η στατιστική με την χρήση των δεδομένων αδιαμφισβήτητα βοήθησαν στην λειτουργία των ξενοδοχείων εντούτοις η εξέλιξη της επιστήμης με την τεχνητή νοημοσύνη έχει να προσφέρει αδιαμφισβήτητα ένα μεγάλο πλεονέκτημα στις διοικήσεις των ξενοδοχείων για την αντιμετώπιση της πολυπλοκότητας της διαχείρισης των κρατήσεων. Σήμερα, η μηχανική μάθηση προσφέρει νέες δυνατότητες και αποτελεσματικότητα στη βελτιστοποίηση της διαχείρισης κρατήσεων, επιτρέποντας την ανάλυση μεγάλων όγκων δεδομένων και την ανάπτυξη μοντέλων πρόβλεψης με μεγάλη ακρίβεια.

Η παρούσα διπλωματική εργασία στοχεύει στην ανάπτυξη και αξιολόγηση μοντέλων μηχανικής μάθησης με απώτερο σκοπό την ανάδειξη του αποτελεσματικότερου μοντέλου στην πρόβλεψη της ακύρωσης της κράτησης για την επιτυχή διαχείριση κρατήσεων σε ξενοδοχειακές μονάδες. Συγκεκριμένα, θα αναπτυχθούν πολλά και διαφορετικά μοντέλα μηχανικής μάθησης για την πρόβλεψη της ακύρωσης της κράτησης και στην συνέχεια θα παρουσιαστούν τα αποτελέσματά τους. Στην συνέχεια, μέσα από την ανάλυση των αποτελεσμάτων και από την εξέταση της ακρίβειας θα επιδιωχθεί να προσδιοριστεί το αποτελεσματικότερο μοντέλο το οποίο θα μπορεί με ακρίβεια να προβλέπει την ακύρωση ή όχι της κράτησης παρέχοντας με αυτόν τον τρόπο στις ξενοδοχειακές μονάδες ένα σημαντικό εργαλείο για την διαχείριση των κρατήσεων και θα αποτελέσει χρήσιμο σύμβουλο των διοικήσεων των ξενοδοχείων στην λήψη αποφάσεων.

Η παρούσα έρευνα αναμένεται να συμβάλει αφενός στην κατανόηση της δυσκολίας διαχείρισης των κρατήσεων και να αναδείξει τη δυναμική της μηχανικής μάθησης για τη διαχείριση κρατήσεων στον ξενοδοχειακό κλάδο, παρέχοντας χρήσιμες προτάσεις για τη βελτίωση της λειτουργίας των ξενοδοχειακών μονάδων. Τα αποτελέσματα της παρούσας έρευνας ανέδειξαν μοντέλα που κρίνονται ιδιαίτερα αποτελεσματικά και φθάνουν σε ακρίβεια 90,39%.

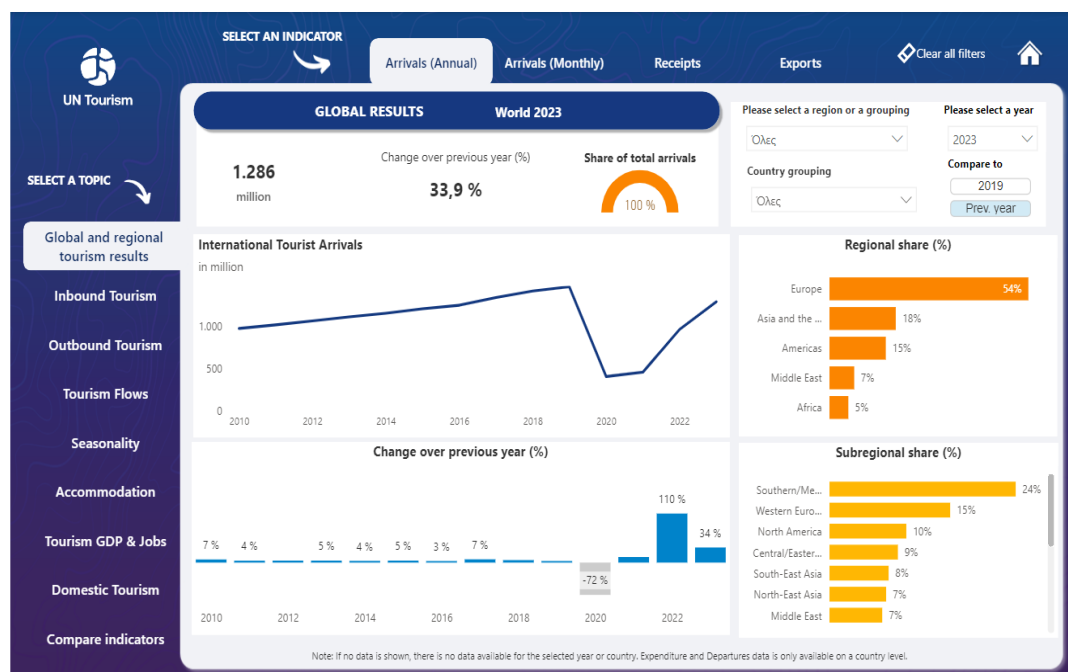
4. Βιβλιογραφική Επισκόπηση

4.1 Τουριστική Βιομηχανία

Στις μέρες μας η τουριστική βιομηχανία διαδραματίζει καθοριστικό ρόλο στην οικονομική ανάπτυξη των κρατών και συμβάλλει στην διατήρηση την ειρήνης σε παγκόσμιο επίπεδο καθώς δημιουργεί θέσεις εργασίας και εισόδημα στα έθνη μέσω της κατανάλωσης αγαθών και υπηρεσιών από τους τουρίστες, βελτιώνει τις συνθήκες διαβίωσης των ανθρώπων και προάγει την «πολιτιστική» και «πολιτισμική» ανταλλαγή μεταξύ των διαφορετικών εθνών - πολιτισμών. Επιπρόσθετα, η βιομηχανία του τουρισμού είναι μία από τις σημαντικότερες στον τομέα των υπηρεσιών, καθώς καταλαμβάνει μία στις δέκα θέσεις εργασίας παγκοσμίως και συνεισφέρει περισσότερο από το 10% του ΑΕΠ της παγκόσμιας οικονομίας. [1]

Αξίζει να σημειωθεί ότι εκατοντάδες εκατομμύρια άνθρωποι ανά τον κόσμο εξαρτώνται από τον τουριστικό τομέα για την επαγγελματική τους απασχόληση και σε αρκετές νησιωτικές οικονομίες ο τουρισμός δεν είναι απλώς ο μεγαλύτερος εργοδότης αλλά ουσιαστικά ο μοναδικός εργοδότης στον οποίο στηρίζονται αποκλειστικά για την διαβίωσή τους [2]. Είναι πλέον γεγονός αδιαμφισβήτητο ότι ο τουρισμός συμβάλλει στην δημιουργία και την διατήρηση βιώσιμων οικονομιών σε όλο τον πλανήτη.

Η τουριστική βιομηχανία αποτελείται από εκατομμύρια εταιρείες και εργοδότες, από τους μεγαλύτερους παγκόσμιους ταξιδιωτικούς πράκτορες μέχρι τις μικρότερες ξενοδοχειακές μονάδες και κάθε είδους μικρή τουριστική επιχείρηση. Ταυτόχρονα είναι ένας από **ταχύτερα αναπτυσσόμενους τομείς στον κόσμο**, με την οικονομική ανάπτυξη το 2017 (4,6%) να ξεπερνά αυτή της παγκόσμιας οικονομίας (3%) για έβδομη συνεχή χρονιά, καθώς και όλων των άλλων μεγάλων βιομηχανικών κλάδων. [2]. Το 2017, υπήρξαν 1,337 δισεκατομμύρια διεθνείς τουριστικές αφίξεις - αύξηση κατά 7,4% σε σύγκριση με το προηγούμενο έτος. [3] Το 2018, σύμφωνα με τον UNWTO οι διεθνείς τουριστικές αφίξεις αυξήθηκαν κατά 5,6% και έφτασαν τα 1,412 δισεκατομμύρια με αποκορύφωμα το 2019 όπου υπήρξε επιπλέον αύξηση της τάξεως του 3,5% με τις διεθνείς αφίξεις να φτάνουν στα 1,462 δισεκατομμύρια. Μετά την σημαντική πτώση στα Ταξίδια παγκοσμίως λόγω του Covid - 19 παρατηρήθηκε εκ νέου σταθερή άνοδος και σύμφωνα με τον UNWTO οι διεθνείς τουριστικές αφίξεις το 2023 έφτασαν τα 1,286 δισεκατομμύρια.



Εικόνα 1 Στοιχεία Παγκόσμιου οργανισμού Τουρισμού [3]

Η τουριστική βιομηχανία θα μπορούσαμε να πούμε ότι έχει σαφή θετικό αντίκτυπο στις οικονομίες των κρατών ανά τον κόσμο και θα μπορούσαμε να ισχυριστούμε ότι έχει βελτιώσει τις ζωές των ανθρώπων προωθώντας την οικονομική ανάπτυξη και την ειρήνη αλλά και μειώνοντας τη φτώχεια με την δημιουργία νέων θέσεων εργασίας. Ο τομέας έχει επιδείξει συνεχή ανάπτυξη και κρίνεται αναγκαίο η επιστήμη να βοηθήσει περαιτέρω με επιστημονικές μελέτες το φαινόμενο του τουρισμού το οποίο κατά πως φαίνεται θα συνεχίσει να αυξάνεται σε μέγεθος και σημασία.

4.2 Το πρόβλημα των ακυρώσεων στις Ξενοδοχειακές Μονάδες

Ένας από τα σημαντικότερους πυλώνες της τουριστικής βιομηχανίας είναι οι ξενοδοχειακές μονάδες καθώς τα ταξίδια και ο τουρισμός βασίζονται στην παροχή καταλυμάτων για τους ταξιδιώτες. Ωστόσο, οι ξενοδοχειακές μονάδες αλλά και γενικότερα το τουριστικό προϊόν αναπτύσσουν την δραστηριότητα τους σε καθεστώς συνεχούς αβεβαιότητας εφόσον απειλούνται μόνιμα από εξωτερικούς παράγοντες – κινδύνους. **Ένας από τους σημαντικότερους κινδύνους των Ξενοδοχείων είναι ότι η υπηρεσία που προσφέρουν δεν αποθηκεύεται και κάθε μη κατειλημμένο δωμάτιο στο τέλος της ημέρας συνεπάγεται απώλεια εσόδων.**[4]

Επιπλέον, η κράτηση ως «προϊόν» έχει πολλές ιδιαιτερότητες σε σχέση με ένα κλασσικό προϊόν προς πώληση. Το προϊόν το οποίο δεν πωλείται σε κάποιον αγοραστή θα πωληθεί σε έναν άλλο υποψήφιο αγοραστή ενώ σε κάποιες περιπτώσεις το ίδιο το προϊόν μπορεί να επιστραφεί με ταυτόχρονη επιστροφή χρημάτων. Αντίθετα, η κράτηση αποτελεί ουσιαστικά μία σύμβαση μεταξύ του καταλύματος και του υποψήφιου ταξιδιώτη για μελλοντική παροχή υπηρεσίας και η ακύρωση μιας κράτησης είναι στην πραγματικότητα η καταγγελία μιας σύμβασης ή η αποποίηση της δέσμευσης για την αγορά του δωματίου του ξενοδοχείου. [5] Ακόμη και μια πρόωρη αποχώρηση από ξενοδοχείο δεν πρόκειται για επιστροφή μιας υπηρεσίας, αλλά μάλλον για παραβίαση ενός συμβολαίου για κατανάλωση της υπηρεσίας τις υπολειπόμενες ημέρες.[5]

Τα καταλύματα έχουν περιορισμένη χωρητικότητα αλλά προβαίνουν σε κρατήσεις εκ των προτέρων με την εγγύηση όμως της μελλοντικής εξυπηρέτησης του φιλοξενούμενου. [6] Εύλογα οι υποψήφιοι φιλοξενούμενοι αλλά και οι επιχειρήσεις φιλοξενίας αισθάνονται αβεβαιότητα λόγω του χρονικού διαστήματος που μεσολαβεί μεταξύ του χρονικού σημείου που γίνεται η κράτηση και του σημείου που τελικά παρέχεται η υπηρεσία.[6]

Στις μέρες μας παρατηρείται πολύ συχνά οι πελάτες να κάνουν κρατήσεις πριν μάθουν την ακριβή ανάγκη τους για εξυπηρέτηση [6] και τελικά να μην τηρούν τις κρατήσεις τους είτε επειδή ακυρώνουν την τελευταία στιγμή είτε επειδή δεν εμφανίζονται. Το γεγονός αυτό αποτελεί ένα αρκετά συνηθισμένο φαινόμενο και είναι δυνατόν να οδηγήσει σε **μεγάλη απώλεια εσόδων για τις ξενοδοχειακές μονάδες**. Οι ανεκπλήρωτες κρατήσεις έχουν ως άμεσο αποτέλεσμα δωμάτια κενά με μη πώληση της προσφερόμενης υπηρεσίας και ως έμμεσο αποτέλεσμα την απώλεια σημαντικών εσόδων.



Οι συνηθέστεροι λόγοι ακύρωσης είναι ξαφνικές ασθένειες, ατυχήματα, αλλαγές στο πρόγραμμα του ταξιδιώτη, απροσδόκητες οικογενειακές ή επαγγελματικές υποχρεώσεις, φυσικές καταστροφές ή ακόμη και πόλεμος. [7] Άλλοι παράγοντες οι οποίοι φαίνεται να επηρεάζουν την ακύρωση της κράτησης είναι η επιλογή του καναλιού διανομής, η οποία δεν είναι τυχαία αντίθετα καθορίζεται από το ίδιο το ταξίδι που επιθυμεί ο ταξιδιώτης και τις ανάγκες που προκύπτουν από αυτό. [7]

Στο παρελθόν, αξίζει να σημειωθεί ότι **οι ακυρώσεις ήταν συχνά δύσκολες** και αρκετά περίπλοκες καθώς τα ξενοδοχεία είτε επέβαλαν χρέωση υψηλού τέλους ακύρωσης είτε δεν παρείχαν την δυνατότητα της ακύρωσης. [8] Ωστόσο, οι ξενοδοχειακές μονάδες τα τελευταία τα τελευταία χρόνια **έχουν υιοθετήσει πιο ευέλικτες πολιτικές ακύρωσης** για την καλύτερη εξυπηρέτηση των επισκεπτών. [5] Ένας από τους σημαντικότερους λόγους που έχουν οδηγήσει τις διοικήσεις τους σε αυτή την αλλαγή είναι η εξαιρετική άνοδος των διαδικτυακών κρατήσεων και ο μεγάλος ανταγωνισμός καθώς ένα σημαντικό ποσοστό των ακυρώσεων προέρχεται από πελάτες που αναζητούν την καλύτερη υπηρεσία στην πιο συμφέρουσα τιμή. [8] Ένας άλλος σημαντικός παράγοντας ήταν η ανάπτυξη των διαδικτυακών πλατφορμών της οικονομίας διαμοιρασμού, όπως η Airbnb οι οποίες λόγω της φιλοσοφίας τους αυξάνουν σε μεγάλο βαθμό τον ανταγωνισμό προσφέροντας πιο ευέλικτες πολιτικές ακύρωσης ασκώντας με αυτόν τον τρόπο πίεση στις ξενοδοχειακές μονάδες οι οποίες τελικώς αναγκάζονται να υιοθετήσουν παρόμοιες πολιτικές με απώτερο σκοπό να παραμείνουν ανταγωνιστικές. [8] Επιπλέον, η πανδημία του COVID-19 είχε πρόσφατα σημαντικό αντίκτυπο στον ξενοδοχειακό κλάδο, όπου τα ξενοδοχεία έπρεπε να προσφέρουν πιο ευέλικτες πολιτικές ακύρωσης για να εξυπηρετήσουν τα μεταβαλλόμενα ταξιδιωτικά σχέδια σε συνθήκες αβεβαιότητας [9]

Εν συνεχεία, στις μέρες μας το διαδίκτυο έχει μεταβάλει καθοριστικά το περιβάλλον των κρατήσεων και μπορεί να έχει πολλά πλεονεκτήματα, αλλά μπορεί επίσης να οδηγήσει σε παρορμητικές κρατήσεις και να οδηγήσει σε υψηλότερα ποσοστά ακύρωσης. [7] Οι διαδικτυακές ιστοσελίδες κρατήσεων διευκολύνουν την κράτηση αλλά και την ακύρωση, κάτι που ενθαρρύνει τους ανθρώπους να κάνουν πολλές κρατήσεις σε παρόμοιες ημερομηνίες σε διαφορετικά ξενοδοχεία, αναζητώντας τις πιο βολικές επιλογές, για να επιλέξουν τελικά μία από αυτές και να ακυρώσουν τις υπόλοιπες. [4] Κατά συνέπεια, η ζήτηση που εκτελείται από ιστότοπους φαίνεται να αυξάνεται αλλά ταυτόχρονα αυξάνονται και τα ποσοστά ακύρωσης. [4]

Αξίζει να σημειωθεί, επίσης, ότι με την εξέλιξη της τεχνολογίας των πληροφοριών έχει παρατηρηθεί αλλαγή στην συμπεριφορά των ταξιδιωτών και κατά συνέπεια έχει καταστεί ακόμη πιο δύσκολη η πρόβλεψη μελλοντικής ζήτησης και των ποσοστών ακύρωσης. [4] Οι ταξιδιώτες πια έχουν περισσότερες πληροφορίες σχετικά με τις εγκαταστάσεις και τις υπηρεσίες που προσφέρουν τα ξενοδοχεία και έχουν την δυνατότητα να διαβάσουν προηγούμενες εμπειρίες πελατών και να επηρεαστούν θετικά ή αρνητικά γεγονός που διευκολύνει τη σύγκριση διαφορετικών προσφορών ενώ αρκετές φορές οι υποψήφιοι πελάτες οδηγούνται και σε ακύρωση. [4]

Τέλος, ένα άλλο φαινόμενο που έχει επηρεάσει αρνητικά την διαχείριση των κρατήσεων και της καθαρής ζήτησης είναι η αύξηση των κρατήσεων της «τελευταίας στιγμής», όπου οι ταξιδιώτες αναβάλλουν την κράτηση στην προσπάθειά τους να επωφεληθούν από κάποια έκπτωση της τελευταίας στιγμής. [4] Τέτοιες πολιτικές εκπτώσεων των διοικήσεων των ξενοδοχειακών μονάδων οδηγούν σε αύξηση της πιθανότητας ακύρωσης από επισκέπτες που έχουν ήδη κάνει κράτηση καθώς μπορεί να είναι διατεθειμένοι να ακυρώσουν την ήδη υπάρχουσα κράτηση και να επιλέξουν να προβούν σε κράτηση σε μία άλλη -πιο οικονομική- ξενοδοχειακή μονάδα. [4] Με την υιοθέτηση αυτών των πολιτικών προωθώντας ευκαιρίες της τελευταίας στιγμής από το ίδιο το ξενοδοχείο ή από ανταγωνιστικές εταιρείες λίγες ημέρες πριν από την ώρα της υπηρεσίας τελικά επιτυγχάνεται αύξηση των ποσοστών των ακυρώσεων. [4]

4.3 Διαχείριση των κρατήσεων στις Ξενοδοχειακές μονάδες

Η διαχείριση των κρατήσεων αποτελεί μία από τις μεγαλύτερες προκλήσεις για τις ξενοδοχειακές μονάδες και τις διοικήσεις τους εφόσον οι ακυρώσεις των κρατήσεων για τις Ξενοδοχειακές μονάδες συνεπάγονται μειωμένη αποτελεσματικότητα στην διαχείριση των εσόδων και μπορούν επίσης να επηρεάσουν τον προγραμματισμό των πόρων και τη διαχείριση αποθεμάτων. [8] Οι ακυρώσεις έχουν σημαντικό αντίκτυπο στις επιδόσεις των ξενοδοχείων καθώς είτε οδηγούν σε απώλεια εσόδων λόγω μη πώλησης του δωματίου είτε σε μειωμένα έσοδα καθώς αυτό μπορεί να διατεθεί σε χαμηλότερη τιμή σε μία “πώληση της τελευταίας στιγμής”. [8] Επιπρόσθετα, τα μη προβλεπόμενα υψηλά ποσοστά ακύρωσης μπορεί να οδηγήσουν σε υπεράριθμο προσωπικό, το οποίο με τη σειρά του δημιουργεί περιττά έξοδα και έχει παρατηρηθεί ότι μπορεί να επηρεάσουν την εξυπηρέτηση πελατών. [8] Στην πραγματικότητα οι ακυρώσεις δημιουργούν τεράστιο πρόβλημα στις ξενοδοχειακές μονάδες καθώς έχει υπολογιστεί ότι τα ξενοδοχεία τελικά χάνουν το ένα πέμπτο των εσόδων τους ως αποτέλεσμα των ακυρώσεων δωματίων. [8] Ένα επιτυχημένο σύστημα διαχείρισης εσόδων των ξενοδοχείων πρέπει να λαμβάνει υπόψη την πιθανότητα ακύρωσης μιας κράτησής ή την πιθανότητα ο πελάτης της κράτησης να μην εμφανιστεί κατά τη στιγμή της υπηρεσίας. [10]

Ο στόχος της διαχείρισης των εσόδων είναι η αύξηση των εσόδων και επιτυγχάνεται μέσω αποφάσεων διαχείρισης της ζήτησης και των ακυρώσεων. [11] Μια συνηθισμένη πρακτική διαχείρισης κρατήσεων των διοικήσεων προκειμένου να ληφθούν υπόψη οι ακυρώσεις είναι η προσπάθεια υπολογισμού της λεγόμενης **“καθαρής ζήτησης”**. [10] Καθαρή ζήτηση ορίζεται ως ο αριθμός των αιτημάτων ζήτησης μείον τον αριθμό των ακυρώσεων. [10] Εναλλακτικά, οι διοικήσεις διαχειρίζονται τις κρατήσεις έχοντας υπόψη την **“εικονική χωρητικότητα”**, η οποία είναι η πραγματική χωρητικότητα συν ένα ρυθμιστικό περιθώριο που αντιπροσωπεύει τον αριθμό των κρατήσεων που αναμένεται να ακυρωθούν. [10] **Και στις δύο πρακτικές της διοίκησης τα ακριβή ποσοστά ακύρωσης είναι ζωτικής σημασίας για το σύστημα διαχείρισης των κρατήσεων.** [10]

Η επιτυχημένη διαχείριση κρατήσεων στις επιχειρήσεις φιλοξενίας είναι υψίστης σημασίας. Υπολογίζεται ότι οι ακυρώσεις επηρεάζουν το 20% έως 60% των κρατήσεων των ξενοδοχείων. [11] Είναι γεγονός ότι οι κρατήσεις επιτρέπουν στους πελάτες να ακυρώσουν μια υπηρεσία με ή χωρίς κυρώσεις πριν από την παροχή της και συνεπώς τα ξενοδοχεία πρέπει να αναλάβουν τον κίνδυνο της εγγύησης δωματίων για τους πελάτες που τιμούν τις κρατήσεις τους, ωστόσο, ταυτόχρονα, τα ξενοδοχεία πρέπει να υποστηρίξουν το κόστος της ύπαρξης κενών δωματίων όταν οι πελάτες ακυρώνουν ή δεν εμφανίζονται [12] Για να μετριάσουν τους κινδύνους που προαναφέρθηκαν οι ξενοδοχειακές μονάδες εφαρμόζουν πολιτικές υπερκρατήσεων και περιοριστικές πολιτικές ακύρωσης. [12]

Στα πλαίσια μιας προληπτικής πολιτικής ακύρωσης, η υπερκράτηση είναι μια πρακτική των διοικήσεων για τη μείωση του κινδύνου που δημιουργείται από απροσδόκητες μη εμφανίσεις. [12] Άλλες ξενοδοχειακές πολιτικές επιβάλλουν περιορισμούς στην ακύρωση για την αποφυγή ζημιών από τυχαίες ακυρώσεις. [8] Οι πολιτικές ακύρωσης είναι σχεδιασμένες με σκοπό την αύξηση των εσόδων με την επιβολή ποινών ή την ενθάρρυνση των πελατών να ακυρώνουν νωρίτερα, ώστε τα ξενοδοχεία να έχουν αρκετό χρόνο για να πουλήσουν τα δωμάτια τους σε άλλους επισκέπτες. [13]

Σπάνια, συναντάται στον τουριστικό τομέα ξενοδοχεία τα οποία μπορεί να έχουν επιλέξει να εφαρμόσουν μια πολιτική η οποία δεν επιτρέπει στους υποψήφιους επισκέπτες να ακυρώσουν σε καμία περίπτωση τις κρατήσεις τους. [8] Τα αποτελέσματα της αυστηρής αυτής πολιτικής είναι διττά και μάλλον περισσότερο με αρνητικά αποτελέσματα παρά θετικά. Στα θετικά στοιχεία της πολιτικής αυτής συγκαταλέγεται η

μείωση του κινδύνου των κενών δωματίων και των χαμένων εσόδων ενώ αντίθετα στα αρνητικά μπορεί να είναι η απογοήτευση των επισκεπτών που ενδεχομένως να έχουν απροσδόκητες αλλαγές στα σχέδιά τους ή απρόβλεπτες καταστάσεις με αποτέλεσμα την μη ικανοποίηση του πελάτη, την δημιουργία αρνητικής κριτικής και την ανεπανόρθωτη βλάβη στη φήμη του ξενοδοχείου. [8] Επίσης, στα αρνητικά αποτελέσματα αυτής της ανελαστικής πολιτικής συγκαταλέγεται και η απώλεια εσόδων από δυνητικούς - μελλοντικούς επισκέπτες που μπορεί να επιλέξουν να μην κάνουν κράτηση στο ξενοδοχείο και τελικά να περιορίσει την ευελιξία του ξενοδοχείου να προσαρμόζει τις τιμές του στις μεταβαλλόμενες συνθήκες της ζήτησης και της προσφοράς. [8]. Γενικά, θα μπορούσαμε να ισχυριστούμε ότι όσο πιο άκαμπτη είναι η πολιτική ακύρωσης ενός ξενοδοχείου, τόσο πιο αρνητικές επιπτώσεις μπορεί να έχει στη συνολική απόδοση του ξενοδοχείου. [8]

Η Διαχείριση των κρατήσεων έχει συνεπώς δύο συνιστώσες: τιμολόγηση και υπερκράτηση. [6] Όσον αφορά την τιμολόγηση, η επιχείρηση χρεώνει ένα τέλος κράτησης (κατά τη στιγμή της κράτησης) και μια τιμή υπηρεσίας (κατά τη στιγμή της υπηρεσίας) και παρατηρείται γενικότερά μεγάλη ποικιλία στις πρακτικές της διοίκησης των ξενοδοχειακών μονάδων. Τέτοιες πρακτικές μπορεί να είναι είτε η δωρεάν κράτηση άλλα με ταυτόχρονη χρέωση μιας μη επιστρεπτέα προκαταβολής είτε η πλήρη και μη επιστρεπτέα πληρωμή κατά την κράτηση. [6] Εν συνεχεία, **όσον αφορά τις υπερκρατήσεις**, οι επιχειρήσεις πρέπει να αποφασίσουν πόσες υπερκρατήσεις θα δεχθούν πέραν του αριθμού που μπορεί να φιλοξενήσει η χωρητικότητα.

Ωστόσο, τόσο οι υπερκρατήσεις όσο και οι περιοριστικές πολιτικές ακύρωσης ως στρατηγική μπορούν να έχουν αρνητικές επιπτώσεις στις επιδόσεις των ξενοδοχείων ενώ δεν είναι βέβαιο ότι επιτυγχάνουν τον στόχο τους. Οι διοικήσεις, στην προσπάθειά τους να ελαχιστοποιήσουν τις απώλειες και εφαρμόζοντας περιοριστικές πολιτικές ακύρωσης και τακτικές υπερκρατήσεων έρχονται αντιμέτωπες με τον κίνδυνο μείωσης των κρατήσεων και κατ' επέκταση των εσόδων. [11] Η υπερκράτηση μπορεί να οδηγήσει στην εξαιρετικά δυσάρεστη κατάσταση για τον επισκέπτη της άρνησης εξυπηρέτησής του και σε διαδικτυακά παράπονα και γενικότερα στη δημιουργία αρνητικού αντίκτυπου όσον αφορά την φήμη του Ξενοδοχείου. [14] Ειδικά στις μέρες μας που η δημοσίευση κριτικών μέσω κοινωνικών δικτύων είναι αρκετά διαδεδομένη είναι εύκολα αντιληπτό ότι τέτοιου είδους πρακτικές μπορεί να έχουν ολέθρια αποτελέσματα στην φήμη ενός ξενοδοχείου και μπορεί να οδηγήσει σε απώλεια μελλοντικών εσόδων καθώς υπάρχει πιθανότητα οι δυσαρεστημένοι πελάτες να μην κάνουν ξανά κράτηση στο ίδιο ξενοδοχείο αλλά μήτε και τρίτα άτομα -υποψήφιοι μελλοντικοί πελάτες- που έλαβαν υπόψιν την αρνητική κριτική. Βέβαια, μια άλλη αρνητική επίπτωση είναι η απώλεια που προκύπτει ως αποτέλεσμα της υποχρέωσης του ξενοδοχείου να αποζημιώσει τον πελάτη, συμπεριλαμβανομένου και του κόστους εύρεσης δωματίου σε άλλο ξενοδοχείο. [15] Από την άλλη πλευρά, οι περιοριστικές πολιτικές ακύρωσης, ιδίως οι πολιτικές που απαιτούν μη επιστρεπτέες προκαταβολές και προθεσμίες ακύρωσης μεγαλύτερες των 48 ωρών, μπορεί να οδηγήσουν τόσο σε μείωση του αριθμού των κρατήσεων όσο και σε μείωση των εσόδων λόγω των σχετικών εκπτώσεων τιμών.

Αυτή η απώλεια εσόδων που προκύπτει από τις ακυρώσεις είναι ένα μεγάλο πρόβλημα για τα ξενοδοχεία και εκτιμάται ότι περίπου το 20% των εσόδων τελικά χάνεται επειδή δεν εξετάζονται οι ακυρώσεις ως μέρος των συστημάτων διαχείρισης κρατήσεων δωματίων. [4] Τέλος, αξίζει να σημειωθεί ότι είναι πολύ σημαντικό για αυτά να υιοθετήσουν μια ακριβή μέθοδο προσδιορισμού της αποτελεσματικής πληρότητας του ξενοδοχείου προκειμένου να υλοποιήσουν μια επιτυχημένη διαχείριση εσόδων [4].

4.5 Η δύναμη της εξόρυξης δεδομένων και της πρόβλεψης με την χρήση της μηχανικής μάθησης

Η διαχείριση των εσόδων (Revenue Management) μπορεί να αυξήσει κατακόρυφα τα έσοδα μιας επιχείρησης μέσω αποφάσεων διαχείρισης της ζήτησης. Για τις διοικήσεις των ξενοδοχειακών μονάδων οι ακυρώσεις κρατήσεων εμποδίζουν την παραγωγή ακριβών προβλέψεων, οι οποίες αποτελούν κρίσιμο εργαλείο στον κλάδο της φιλοξενίας και μπορούν να βοηθήσουν σημαντικά στην μείωση της απώλειας των εσόδων. Ένα σύστημα προβλέψεων μπορεί να επιτρέψει στα ξενοδοχεία να προβλέψουν την καθαρή ζήτησή τους και έτσι να λάβουν καλύτερες αποφάσεις σχετικά με το ποιες κρατήσεις να δεχτούν και να απορρίψουν, ποιες τιμές να κάνουν και σε ποιο βαθμό θα προβούν σε υπερκρατήσεις.[16]

Συνεπώς, η διαχείριση εσόδων αφορά τις μεθοδολογίες και τα συστήματα που απαιτούνται για τη λήψη αποφάσεων σχετικά με τη ζήτηση και τις πιθανές ακυρώσεις. Είναι γεγονός ότι οι ακυρώσεις, ενδέχεται να μειώσουν την ακρίβεια των προβλέψεων ζήτησης και, κατά συνέπεια, να μειώσουν την αποτελεσματικότητα των πρακτικών διαχείρισης εσόδων διότι ο βέλτιστος έλεγχος των τιμών και των αποθεμάτων εξαρτάται από ακριβείς προβλέψεις.[5] Η απόδοση των προβλέψεων είναι ένα κρίσιμο εργαλείο για τα συστήματα διαχείρισης εσόδων (Revenue Management System) καθώς δίχως ακριβείς προβλέψεις, οι συστάσεις του τμήματος διαχείρισης εσόδων για τις τιμές και τη διαθεσιμότητα θα ήταν πιθανώς εξαιρετικά ανακριβείς.[16] Στην πραγματικότητα, η εκτίμηση και η πρόβλεψη αποτελούν βασικά τμήματα της γνωστής κυκλικής διαδικασίας διαχείρισης εσόδων: συλλογή δεδομένων, εκτίμηση και πρόβλεψη, βελτιστοποίηση και έλεγχος [11]

Στις μέρες μας με την εξέλιξη της τεχνολογίας εν γένει και ειδικότερα με την εξέλιξη της τεχνητής νοημοσύνης και την χρήση της μηχανικής μάθησης είναι γεγονός ότι η υφιστάμενη αλληλεπίδραση συνεργασίας των διοικήσεων των Ξενοδοχειακών μονάδων με την τεχνολογία έχει αλλάξει τα δεδομένα και παρατηρείται μια συνεχής προσαρμογή του ξενοδοχειακού κλάδου στις τεχνολογικές εξελίξεις. [1] **Η Τεχνητή Νοημοσύνη και ειδικότερα η μηχανική μάθηση δύναται να ενσωματωθεί στις πρακτικές της διοίκησης ως στρατηγικό εργαλείο πρόβλεψης για τον τουρισμό.**

Πιο συγκεκριμένα η τουριστική βιομηχανία και τα Ξενοδοχεία τα τελευταία χρόνια έχουν στραφεί στην **προγνωστική μοντελοποίηση η οποία είναι μια καινοτόμος ερευνητική μέθοδος στο πλαίσιο της φιλοξενίας και του τουρισμού που κάνει προβλέψεις για το μέλλον με βάση ιστορικά δεδομένα.** [8] **Η προγνωστική μοντελοποίηση με την χρήση της μηχανικής μάθησης αποτελεί μια βιώσιμη εναλλακτική λύση λόγω της μοναδικής της ικανότητας να εντοπίζει σχέσεις σε σύνολα δεδομένων για τα οποία δεν θα μπορούσαν να προσδιοριστούν ακριβή μοντέλα με παραδοσιακά μέσα.** [16] [4] **Η μηχανική μάθηση**, ως υποπεδίο της τεχνητής νοημοσύνης περιλαμβάνει την εκπαίδευση αλγορίθμων για την ανάδειξη των ακριβέστερων και αποτελεσματικότερων μοντέλων ώστε να επιτυγχάνονται προβλέψεις. [8] Στα πλεονεκτήματα της μηχανικής μάθησης περιλαμβάνονται οι ανώτερες επιδόσεις της στην επεξεργασία και την ανάλυση μεγάλων δεδομένων (Big Data) ενώ υπάρχει δυνατότητα υιοθέτησης υβριδικής μηχανικής μάθησης με την οποία χρησιμοποιούνται πολλαπλές τεχνικές μηχανικής μάθησης προκειμένου να επιτευχθεί αύξηση της ακρίβειας και δημιουργία ενός πιο ισχυρού μοντέλου πρόβλεψης. [8]

Η μηχανική μάθηση μπορεί να βοηθήσει ουσιαστικά στην πραγματοποίηση προβλέψεων για τη ζήτηση δωματίων ξενοδοχείων αν λάβουμε υπόψιν ότι η ζήτηση για δωμάτια ξενοδοχείων δεν είναι απλώς μια συνάρτηση του αριθμού των κρατήσεων, αλλά περιλαμβάνει επίσης εκτιμήσεις για το ποσοστό των ακυρώσεων οι οποίες επηρεάζονται σημαντικά από εξωγενείς παράγοντες που δεν ελέγχονται από το ξενοδοχείο καθώς και από την πολιτική του ίδιου του Ξενοδοχείου. [8] Αδιαμφισβήτητα οι ακυρώσεις παίζουν

σημαντικό ρόλο στον καθορισμό των καθαρών εσόδων και συνεπώς κρίνεται σημαντικό οι Διοικήσεις των Ξενοδοχείων να κατανοήσουν τη «συμπεριφορά» των υποψήφιων πελατών που τελικά ακυρώνουν και να οργανώσουν την στρατηγική διαχείρισης των εσόδων τους κατανοώντας αυτήν την συμπεριφορά.[8]

Γίνεται εύκολα αντιληπτό ότι πλέον υπάρχει ισχυρή συσχέτιση μεταξύ της πρόβλεψης της ζήτησης και της διαχείρισης των εσόδων των ξενοδοχειακών μονάδων. Η πρόβλεψη της ζήτησης είναι ένα από τα σημαντικότερα εργαλεία για τα όργανα της διοίκησης προκειμένου να κατανοήσουν τους λόγους της ζήτησης, να ανταποκριθούν στην αυξημένη ζήτηση και να μπορέσουν να προβλέψουν τη μελλοντική ζήτηση. [1] Ειδικότερα, η αξιόπιστη πρόβλεψη ακύρωσης ξενοδοχείων δύναται να βοηθήσει στον καθορισμό κατάλληλων επιχειρησιακών στρατηγικών για τη διαχείριση των κρατήσεων των ξενοδοχείων. [17] Η δυνατότητα ορθής πρόβλεψης για τις επιχειρήσεις είναι κρίσιμη καθώς μπορούν έτσι να καθορίσουν προσαρμοσμένες πολιτικές τιμολόγησης, κατανομής αποθεμάτων ή ωρών εργασίας. [1]

Για την κατασκευή αυτών των μοντέλων πρόβλεψης χρησιμοποιούνται αποθηκευμένα δεδομένα εκτός σύνδεσης στη βάση δεδομένων των ξενοδοχείων και στη συνέχεια τα δεδομένα αυτά επεξεργάζονται και αναλύονται με την βοήθεια της μηχανικής μάθησης. Κάθε σύστημα διαχείρισης εσόδων των Ξενοδοχείων συλλέγει και αποθηκεύει αρχεία κρατήσεων και πληροφορίες αγοράς και τις χρησιμοποιεί για να προβλέπει τη ζήτηση και να μαθαίνει τη συμπεριφορά των πελατών. [10] Υπό αυτό το πρίσμα, τα δεδομένα κρατήσεων επισκεπτών τα **“Personal Name Record (PNR)”** μπορεί να είναι η πιο αντιπροσωπευτική πηγή πληροφοριών για τις εργασίες πρόβλεψης. [17] **Τα αρχεία πληροφοριών των PNR περιέχουν τις βασικές λεπτομέρειες μιας κράτησης ενός ξενοδοχείου και έχουν αποδειχθεί ιδιαίτερα πολύτιμες για τη διαχείριση των ξενοδοχείων.** [17] Τα τελευταία χρόνια έχει υιοθετηθεί η προγνωστική μοντελοποίηση με την χρήση λεπτομερών δεδομένων κρατήσεων - αντί για συγκεντρωτικά ιστορικά δεδομένα χρονοσειρών - σε συνδυασμό με αλγόριθμους ταξινόμησης επειδή είναι πιο ακριβείς και μπορούν να αναπτύξουν μοντέλα πρόβλεψης ταξινόμησης [8] Ωστόσο, τα αρχεία αυτά - όντας προσωπικά δεδομένα - προστατεύονται από την νομοθεσία και συνεπώς δεν επιτρέπεται στους ενδιαφερόμενους φορείς της φιλοξενίας να μοιράζονται αυτά τα πολύτιμα δεδομένα. [17]

Οι προβλέψεις σχετικά με τις ακυρώσεις των ξενοδοχειακών κρατήσεων αποτελούν κρίσιμο εργαλείο των διοικήσεων στον κλάδο της φιλοξενίας και φαίνεται να είναι η αποτελεσματικότερη λύση που θα καθοδηγήσει τις διοικήσεις στην ανάπτυξη κατάλληλων επιχειρησιακών στρατηγικών για την αποφυγή των δυσμενών επιπτώσεων [17] Σήμερα, με την εξέλιξη της τεχνολογίας και τους προηγμένους αλγορίθμους μηχανικής μάθησης είναι δυνατή η δημιουργία μοντέλων με υψηλή ακρίβεια για την πρόβλεψη της πιθανότητας ακύρωσης κρατήσεων. [17]

Με την χρήση της μηχανικής μάθησης και την ανάλυση των δεδομένων πελατών μπορεί να επιτευχθεί ακριβής πρόβλεψη των ποσοστών ακύρωσης και συνεπώς οι διοικήσεις των ξενοδοχείων να λάβουν αποτελεσματικότερες αποφάσεις σχετικά με την προσαρμογή των τιμών των δωματίων, τα επίπεδα πληρότητας και τις στρατηγικές μάρκετινγκ μεγιστοποιώντας με αυτόν τον τρόπο τα έσοδά των ξενοδοχείων και ταυτόχρονα ελαχιστοποιώντας τυχόν πιθανές απώλειες.[4] Η πρόβλεψη της πραγματικής ζήτησης μπορεί να μειωθεί εάν δεν ληφθούν υπόψη τα ποσοστά ακύρωσης. [4] Επιπλέον, είναι εξαιρετικά σημαντικό για τις ξενοδοχειακές επιχειρήσεις η πρόβλεψη της ακύρωσης και η διαχείριση των ακυρώσεων έγκαιρα και αποτελεσματικά για την δημιουργία καλών σχέσεων με τους πελάτες και την μείωση των παραπόνων των πελατών.[8]

Τέλος, αν και η έρευνα σχετικά με την πρόβλεψη των ποσοστών ακύρωσης με την χρήση μηχανικής μάθησης στο πλαίσιο της φιλοξενίας είναι αρκετά περιορισμένη οι πρόσφατες

τεχνολογικές καινοτομίες και οι αλγόριθμοι πρόβλεψης που χρησιμοποιούν διαφορετικές μεθόδους με λεπτομερή δεδομένα κρατήσεων ξενοδοχείων έχουν κερδίσει δημοτικότητα τα τελευταία χρόνια επειδή είναι πιο ακριβείς και μπορούν να αναπτύξουν μοντέλα πρόβλεψης ταξινόμησης. [8]

4.6 Μελέτες προϋφιστάμενες της παρούσας εργασίας

Οι μελέτες που έχουν γίνει προγενέστερα είναι σχετικά περιορισμένες ωστόσο ήδη από **το 2010 οι Morales και Wang** προσπάθησαν να μελετήσουν το πρόβλημα της πρόβλεψης των ποσοστών ακύρωσης χρησιμοποιώντας μεθόδους εξόρυξης δεδομένων. Η μελέτη χρησιμοποίησε για την ανάλυση 14 μεταβλητές με δεδομένα κρατήσεων, όπως η διάρκεια της διαμονής, το κανάλι διανομής, το αν η κράτηση είναι με επιστροφή χρημάτων, ο τύπος δωματίου και το σύστημα κράτησης που χρησιμοποιήθηκε στην κράτηση, και εξέτασε την απόδοση οκτώ μεθόδων εξόρυξης δεδομένων. (ιδέ φωτογραφία...)

Πίνακας 1 Μέθοδοι εξόρυξης Δεδομένων Morales και Wang (2010) [10]

Abbreviation	Description
AVG	Average cancellation rate among all training data
SAVG	Seasonally averaged cancellation rate among all training data
LR	Logistic Regression
C4.4	C4.4 probability estimation tree (Provost and Domingos, 2003)
MSEE	Minimum Squared Expected Error tree (Nielsen, 2004)
RF	Random Forest (Breiman, 2001)
SVM	Support Vector Machine (Cristianini and Shawe-Taylor, 2000)
KLR	Kernel Logistic Regression (Wahba, 1999)

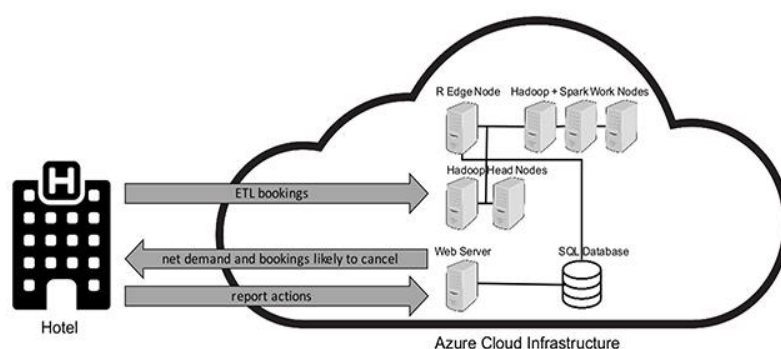
Σύμφωνα με την έρευνα η πρόβλεψη των ακυρώσεων που συμβαίνουν κατά τη διάρκεια των διαφόρων σταδίων του ορίζοντα κράτησης είναι ένα πολύπλοκο πρόβλημα. Συγκεκριμένα, στα διαφορετικά στάδια του ορίζοντα το σύνολο των παραγόντων που επηρεάζουν την πιθανότητα ακύρωσης μιας κράτησης είναι πολύ διαφορετικό. Η μελέτη τους αντιμετώπισε τα ποσοστά ακύρωσης ως πρόβλημα παλινδρόμησης και διαπίστωσαν ότι οι μέθοδοι estimation tree και kernel είναι ευνοϊκές για την πρόβλεψη των ποσοστών ακύρωσης. Τέλος, οι Morales και Wang πρότειναν να δημιουργηθούν είτε πολλαπλά μοντέλα για τα διαφορετικά στάδια του ορίζοντα της κράτησης είτε ένα ενιαίο μοντέλο που να λαμβάνει υπόψη τη χρονική εξάρτηση. [10]

Στη συνέχεια, σε πιο πρόσφατη έρευνα **του 2017 των Nuno Antonio, Luis Nunes και Ana de Almeida** μελέτησαν σύνολα δεδομένων των ετών 2013 έως 2015 από τέσσερα πολυτελή ξενοδοχεία (θέρετρα) μεσαίας χωρητικότητας της Πορτογαλίας και αντιμετώπισαν την πρόβλεψη ακύρωσης κρατήσεων ως πρόβλημα ταξινόμησης. Τα δεδομένα των ξενοδοχείων αντλήθηκαν από το σύστημα διαχείρισης κρατήσεων του ξενοδοχείου και με την χρήση του εργαλείου Microsoft Azure Machine Learning Studio αναπτύχθηκαν νέα μοντέλα με τη χρήση διαφορετικών αλγορίθμων ταξινόμησης Boosted Decision Tree, Decision Forest, Decision Jungle, Locally Deep Support Vector Machine και Neural Network των οποίων η απόδοση αξιολογήθηκε με την διασταυρούμενη επικύρωση 10 fold cross-validation. Οι ερευνητές απέδειξαν ότι είναι δυνατή η δημιουργία μοντέλων για την πρόβλεψη ακύρωσης κρατήσεων καθώς οι περισσότεροι αλγόριθμοι παρουσίασαν μέσες τιμές ακρίβειας άνω του 90% με τον αλγόριθμο Decision Forest να παρουσιάζει τις καλύτερες επιδόσεις όσον αφορά την ακρίβεια. [18]

Σε μεταγενέστερη έρευνα που δημοσιεύτηκε **την 29.5.2019 οι ίδιοι ερευνητές Nuno Antonio, Luis Nunes και Ana de Almeida** μελέτησαν δεδομένα κρατήσεων από οκτώ πορτογαλικά ξενοδοχεία εκ των οποίων τα 4 ήταν ξενοδοχεία θέρετρα και τα υπόλοιπα 4 ήταν ξενοδοχεία πόλης. Τα σύνολα των δεδομένων αφορούσαν το χρονικό διάστημα από 1.1.2016 έως 20.11.2017 και αντλήθηκαν από τις βάσεις δεδομένων του συστήματος διαχείρισης κρατήσεων αλλά και από άλλες πηγές δεδομένων όπως το Weather Underground για παγκόσμια δεδομένα καιρικών συνθηκών, κατασκευασμένοι αυτοματοποιημένοι web scrapers για αντληση παγκόσμιων δεδομένων εκδηλώσεων, τοπικών και εθνικών εορτών αλλά και τις ιστοσελίδες Booking.com και Tripadvisor.com για την άντληση δεδομένων κριτικών και κοινωνικής φήμης. Για την εν λόγω έρευνα χρησιμοποιήθηκε ο βραβευμένος αλγόριθμος μηχανικής μάθησης XGBoost. Σε σύγκριση με τις προηγούμενες δημοσιευμένες έρευνες, η σημαντικότερη καινοτομία που εισήχθη με την έρευνα αυτή στη μοντελοποίηση ήταν ο συνδυασμός του XGBoost με τον τρόπο με τον οποίο εξήχθησαν τα δεδομένα από το PMS και άλλες πηγές και τον τρόπο με τον οποίο χωρίστηκαν τα σύνολα δεδομένων για την εκπαίδευση με το XGBoost. Όλες οι μετρικές παρουσίασαν καλά αποτελέσματα όσον αφορά την απόδοση πρόβλεψης χρησιμοποιώντας το σύνολο επικύρωσης. Η ακρίβεια “Accuracy” κυμαινόταν από 0,7775 έως 0,8928, το “Precision” κυμαινόταν από 0,5724 έως 0,8033 και η AUC κυμαινόταν από 0,7864 έως 0,9307. Από την έρευνα τους προέκυψε ότι η χρήση περισσότερων δεδομένων δεν παράγει πάντα καλύτερα μοντέλα πρόβλεψης και ταυτόχρονα η χρήση περισσότερων δεδομένων από την ίδια πηγή μπορεί να μην οδηγεί σε μοντέλα με καλύτερες επιδόσεις και αποδείχθηκε ότι με τη χρήση προηγμένων αλγορίθμων μηχανικής μάθησης, είναι δυνατόν να προβλεφθεί η πιθανότητα ακύρωσης κάθε κράτησης. Αποδείχθηκε, επίσης ότι τα μοντέλα πρόβλεψης ταξινόμησης που χρησιμοποιούν λεπτομερή δεδομένα κρατήσεων, σε σύγκριση με τα μοντέλα παλινδρόμησης και τα μοντέλα που χρησιμοποιούν ιστορικά δεδομένα, είναι πολύ πιο αποτελεσματικά στην ανάπτυξη ολοκληρωμένων μοντέλων. Διαπίστωσαν επίσης ότι, τα μοντέλα πρόβλεψης ταξινόμησης μπορούν να χρησιμοποιηθούν για τη δημιουργία προβλέψεων σε πολύ λεπτομερές επίπεδο (ανά κράτηση), καθώς και για την πρόβλεψη της καθαρής ζήτησης σε παγκόσμια ή συγκεντρωτικά επίπεδα, όπως τμήμα αγοράς, κανάλι διανομής και ταξιδιωτικό γραφείο. Απέδειξαν ότι για προβλήματα πρόβλεψης ακύρωσης κράτησης, τα δεδομένα κράτησης θα πρέπει να περιλαμβάνουν λεπτομέρειες κράτησης πριν από το αποτέλεσμα της ακύρωσης (ημερομηνία άφιξης ή ακύρωσης) και ο αριθμός των τροποποιήσεων που σχετίζονται με μια κράτηση είναι ένας σημαντικός παραγοντας ακύρωσης. Κατέληξαν στο συμπέρασμα ότι η ζήτηση και οι ακυρώσεις μπορεί να διαφέρουν ανάλογα με το ξενοδοχείο, τον πελάτη ή την κράτηση ή λόγω εξωτερικών παραγόντων και κατασκεύασαν ένα συγκεκριμένο μοντέλο για κάθε ξενοδοχείο και κατά συνέπεια επιβεβαιώθηκαν τα ευρήματα προηγούμενων μελετών σύμφωνα με τα οποία παράγοντες όπως ο χρόνος παράδοσης, η χώρα, η διάρκεια διαμονής, το τμήμα της αγοράς και το κανάλι διανομής έχουν μεγάλη σημασία για την πρόβλεψη ακυρώσεων για οποιοδήποτε ξενοδοχείο αλλά ταυτόχρονα ότι αυτή η σημασία μπορεί να ποικίλλει για διαφορετικά ξενοδοχεία. Ένα από τα σημαντικότερα ευρήματα της έρευνας ήταν ότι η απόδοση των μοντέλων δεν βελτιώθηκε ουσιαστικά με τη συμπερίληψη χαρακτηριστικών από άλλες πηγές και κανένα από τα χαρακτηριστικά από πηγές δεδομένων εκτός PMS δεν είχε ιδιαίτερη προγνωστική σημασία επιβεβαιώνοντας έτσι τους Falk και Vieru (2018) ότι οι ειδικές εκδηλώσεις και οι δείκτες εμπιστοσύνης πελατών δεν εξηγούν τα μοτίβα ακύρωσης. [11]

Σε έρευνα που δημοσιεύτηκε **την 8.7.2019 οι ίδιοι ερευνητές Nuno Antonio, Luis Nunes και Ana de Almeida** προσπάθησαν να αξιολογήσουν την αποτελεσματικότητα των μοντέλων πρόβλεψης με την αξιοποίηση της μηχανικής μάθησης σε πραγματικό περιβάλλον με την κατασκευή ενός πρωτότυπου που αναπτύχθηκε σε δύο πορτογαλικά ξενοδοχεία τα οποία ήταν ένα θέρετρο και ένα ξενοδοχείο πόλης. Το πρωτότυπο κατασκευάστηκε με την χρήση της μεθολογίας CRISP-DM και βασίστηκε σε ένα αυτοματοποιημένο σύστημα μηχανικής μάθησης σχεδιασμένο να μαθαίνει συνεχώς, το

οποίο σχεδιάστηκε για να καταγράφει τις τροποποιήσεις στα πρότυπα ακυρώσεων με την πάροδο του χρόνου και να μαθαίνει από τις επιτυχίες και τα σφάλματα των προβλέψεων των προηγούμενων ημερών. Για την κατασκευή των μοντέλων χρησιμοποίησαν **τον αλγόριθμο XGBoost** για τη δημιουργία μοντέλων ταξινόμησης για την πρόβλεψη της ακύρωσης κάθε κράτησης. Το σύστημα κατασκευάστηκε στην πλατφόρμα cloud Microsoft Azure με χρήση πολλών στοιχείων ανοιχτού κώδικα και τεχνολογίες που διατίθενται ως υπηρεσίες στην συγκεκριμένη πλατφόρμα, ένα σύμπλεγμα, δηλαδή, βασισμένο σε HDInsight Linux, Hadoop και Spark με διακομιστή R που επέτρεψε την επεξεργασία μεγάλων δεδομένων. Το πρωτότυπο αποθήκευε επίσης όλα τα αποτελέσματα πρόβλεψης με ενέργειες εκπαιδευμένων χρηστών σε έναν δυναμικό ιστότοπο με την βοήθεια ενός διακομιστή. Σε αυτόν τον ιστότοπο υπήρχε η δυνατότητα οι χρήστες να μπορούν να συμβουλευονται τη ζήτηση, τις προβλέψεις και να αναφέρουν τις πρακτικές που υιοθετήθηκαν για τις κρατήσεις που έχουν εντοπιστεί ως πιθανό να ακυρωθούν.



Εικόνα 2 Διάγραμμα του συστήματος των ερευνητών Nuno Antonio, Luis Nunes και Ana de Almeida [16]

Bookings Cancellation Prototype											
Planning											
Processing date: 2017-08-15 (Tue)			Lookup days: Next 30 days			Pickup analysis: 2017-08-14 (Mon)			Search:		
Folios	Lookup Date	A	B	C	D	E	F	G	H	TOTAL	Pickup
66	2017-08-15 (Tue)	68	1	11	43	29	9	7	3	171	90.4%
		67	1	11	43	29	9	7	3	170	92.9%
71	2017-08-16 (Wed)	49	1	12	49	30	10	7	2	182	99.5%
		66	1	11	49	30	10	7	1	175	95.6%
Folio	Arrival	Nights	Departure	Adults	Children	Babies	ADR	Room revenue	Frequency	Action executed	
45838	2017-08-16 (Wed)	6	2017-08-22 (Tue)	2	0	0	241.67 EUR	1,450.02 EUR	22.7% (17/75)	✗	
46221	2017-08-16 (Wed)	5	2017-08-21 (Mon)	3	1	0	388.00 EUR	1,940.00 EUR	60.3% (38/63)	✓	
46422	2017-08-16 (Wed)	5	2017-08-21 (Mon)	2	0	0	260.00 EUR	1,300.00 EUR	100.0% (6/6)	✗	
4 other bookings were hidden in the control group											
66	2017-08-17 (Thu)	60	0	12	48	28	9	8	3	177	96.7%
		60	0	11	48	27	9	6	1	162	88.5%
69	2017-08-18 (Fri)	61	1	12	48	30	8	2	0	180	96.4%
		61	1	11	47	29	10	6	0	165	90.2%
70	2017-08-19 (Sat)	61	1	11	47	28	8	6	0	162	88.5%
		61	1	11	47	28	8	6	0	162	88.5%
66	2017-08-20 (Sun)	60	1	11	44	24	7	5	0	152	83.1%
		60	1	11	44	24	7	5	0	152	83.1%
70	2017-08-21 (Mon)	62	1	11	50	27	8	7	2	176	96.2%
		62	1	11	47	25	8	6	0	170	87.4%
71	2017-08-22 (Tue)	64	0	11	50	28	9	8	3	180	96.4%
		64	0	10	48	24	8	8	0	162	88.5%
69	2017-08-23 (Wed)	59	0	12	48	30	9	8	3	179	97.8%
		59	0	11	47	25	8	8	1	159	86.9%
70	2017-08-24 (Thu)	60	0	12	52	27	10	7	3	181	96.9%
		60	0	10	51	23	9	7	1	161	88.0%

Πίνακας 2 Κεντρική Οθόνη Πρωτοτύπου των ερευνητών Nuno Antonio, Luis Nunes και Ana de Almeida[16]

Επιπλέον, αξίζει να αναφερθεί ότι, με δεδομένο ότι τα πρότυπα ακύρωσης αλλάζουν με την πάροδο του χρόνου και επειδή το σύστημα είχε σχεδιαστεί να μαθαίνει συνεχώς δημιουργήσαν ένα νέο μέτρο την “**Ελάχιστη συχνότητα**” (Minimum Frequency) για τη μέτρηση της ακρίβειας των προβλέψεων με την πάροδο του χρόνου. Τελικώς, το πρωτότυπο απέδειξε την αποτελεσματικότητά του, με αποτελέσματα που ξεπέρασαν το 84% στην ακρίβεια, το 82% στην ακρίβεια και το 88% στην περιοχή κάτω από την καμπύλη (AUC) και συνέβαλε ώστε τα ξενοδοχεία να μπορούν να προβλέψουν την καθαρή ζήτησή τους και έτσι να λάβουν καλύτερες αποφάσεις σχετικά με το ποιες κρατήσεις να δεχτούν και να απορρίψουν, την τιμολόγηση και σε τι ποσοστό θα προβούν σε υπερκρατήσεις. Η συστηματική πρόβλεψη των κρατήσεων με υψηλή πιθανότητα ακύρωσης επέτρεψε στα ξενοδοχεία να μειώσουν τις ακυρώσεις κατά 37 ποσοστιαίες μονάδες, ενεργώντας ώστε να αποφύγουν την ακύρωσή τους.[16]

Το 2020 οι **Agustín J. Sánchez-Medina, Eleazar C-Sánchez** μελέτησαν και αυτοί το πρόβλημα των ακυρώσεων των κρατήσεων και προσπάθησαν να αναπτύξουν ένα μοντέλο για την πρόβλεψη ακυρώσεων ξενοδοχειακών κρατήσεων χρησιμοποιώντας δεδομένα κρατήσεων ενός ξενοδοχείου που βρίσκεται στην Γκραν Κανάρια της Ισπανίας. Το σετ δεδομένων αποτελούνταν από 13 ανεξάρτητες μεταβλητές οι οποίες συμπίπτουν με αυτές που ζητούνται συχνότερα από τους πελάτες όταν κάνουν κράτηση καθώς κύριος στόχος αυτής της έρευνας ήταν η δημιουργία μοντέλου με την χρήση των πιο κοινών μεταβλητών που ζητούνται από τους πελάτες όταν προβαίνουν σε κράτηση. Για τον λόγο αυτό εξετάστηκαν οι πιο δημοφιλείς διαδικτυακές πύλες κρατήσεων όπως η Booking, το Tripadvisor ή η Trivago, καθώς και μεγάλες ξενοδοχειακές αλυσίδες όπως η Radisson Hotels & Resorts, η Riu Hotels & Resorts ή η Meliá Hotel Resorts μεταξύ άλλων.

Name	Description	Type
Status	Booking status: on place, cancelled	Categorical
Adults	Number of adults	Numeric
Entity	Enterprises through which customers may book rooms	Categorical
Nationality	Nationality of the guest	Categorical
Advance payment	If require advance payment or not	Categorical
Nights	Number of nights to be spent in the hotel	Numeric
Notice period	Difference between booking date and arrival date	Numeric
Day of creation	Day in which booking was created	Numeric
Month of creation	Month in which booking was created	Numeric
Day of check in	Effective check in day	Numeric
Month of check in	Effective check in month	Numeric
Mean price	Room mean price	Numeric
Sales channel	Sales channels used for the reservation. Entities are organised into 9 channels (e.g. business to business, hotel website)	Categorical
Weekend	Number Saturdays and Sundays during the stay	Numeric

Πίνακας 3 Μεταβλητές μελέτης Agustín J. Sánchez-Medina, Eleazar C-Sánchez [4]

Για την ανάπτυξη του μοντέλου των **Agustín J. Sánchez-Medina, Eleazar C-Sánchez** υιοθετήθηκε η διαδικασία CRISP-DM (Cross Industry Standard Process for Data Mining) σύμφωνα με την οποία, πριν από οποιαδήποτε προετοιμασία δεδομένων, είναι απαραίτητη η κατανόηση της επιχείρησης και των ίδιων των δεδομένων, και στη συνέχεια μπορούν να κατασκευαστούν και να δοκιμαστούν μοντέλα. Για την έρευνα χρησιμοποιήθηκαν δεδομένα κρατήσεων δύο ετών με περισσότερες από 10.000 κρατήσεις, συμπεριλαμβανομένων όλων των κρατήσεων του έτους 2016 έως και τον Απρίλιο του 2018. Για την ανάπτυξη του μοντέλου χρησιμοποιήθηκε το στατιστικό

λογισμικό R (R Core Team, 2013) και εφαρμόστηκαν διαφορετικοί αλγόριθμοι μάθησης με επίβλεψη. Υπό το πλαίσιο αυτό, χρησιμοποιήθηκαν διάφορα πακέτα: **δύο αλγόριθμοι που βασίζονται σε δέντρα απόφασης, C5.0, random forest, support vector machine (SVM), τεχνητά νευρωνικά δίκτυα (ANN) και γενετικός αλγόριθμος (GA).** [4]

	Accuracy	Precision	F1 score	Specificity	Recall	AUC
Random forest	0.804	0.813	0.806	0.809	0.799	0.804
Support vector machine	0.753	0.733	0.748	0.743	0.764	0.820
C 5.0	0.790	0.818	0.796	0.807	0.774	0.864
ANN (GA optimized)	0.980	0.972	0.979	0.972	0.987	0.989

Πίνακας 4 Επιδόσεις Αλγορίθμων μελέτης Agustín J. Sánchez-Medina, Eleazar C-Sánchez [4]

Η έρευνα κατέδειξε **πρώτον** ότι είναι δυνατή η πρόβλεψη ακυρώσεων με υψηλό επίπεδο ακρίβειας **χρησιμοποιώντας τα δεδομένα PNR και με μειωμένο αριθμό μεταβλητών** σε σχέση με προηγούμενες έρευνες, **δεύτερον** οι αλγόριθμοι που βασίζονται σε δενδρικές αποφάσεις ή ο SVM είχαν χαμηλότερες επιδόσεις, **τρίτον** ο καθορισμός της αρχιτεκτονικής του τεχνητού νευρωνικού δικτύου είναι ένα πολύπλοκο πρόβλημα και δεν υπάρχει συστηματική διαδικασία για την εκτέλεση αυτού του έργου, γεγονός που αναγκάζει τη ρύθμιση με δοκιμή και σφάλμα και **τέταρτον** ότι η χρήση γενετικών αλγορίθμων επιτρέπει την επίτευξη ακρίβειας 98% ακόμη και όταν χρησιμοποιούνται μόνο 13 ανεξάρτητες μεταβλητές. [4]

Σε έρευνα που δημοσιεύτηκε το 2020 οι **Md. Shahriare Satu, Khair Ahammed και Mohammad Zoynul Abedin** προσπάθησαν να διερευνήσουν τις επιδράσεις των διαφόρων μεθόδων μηχανικής μάθησης στη διαδικασία διαχείρισης ξενοδοχειακών κρατήσεων. Για την έρευνα χρησιμοποιήθηκε ένα σύνολο δεδομένων κρατήσεων ξενοδοχείων από το αποθετήριο δεδομένων της Kaggle. Τα δεδομένα αυτά περιείχαν περιπτώσεις κρατήσεων τριών ετών από την 1η Ιουλίου 2015 έως την 31η Αυγούστου 2017, συλλέχθηκαν από τους Antonio, Almeida και Nunes, περιείχαν πληροφορίες για κρατήσεις ξενοδοχείων από ένα θέρετρο (H1) και ένα ξενοδοχείο της πόλης (H2) και αποτελούνταν από 31 μεταβλητές. Στη συνέχεια, εφαρμόστηκαν διαφορετικές τεχνικές μετασχηματισμού χαρακτηριστικών στο πρωτογενές σύνολο δεδομένων και δημιουργήθηκαν μετασχηματισμένα σύνολα δεδομένων. Για την ανάπτυξη μοντέλων χρησιμοποίησαν διαφορετικούς ταξινομητές όπως **Gradient Boosting (GB), Random Forest (RF), Xgboost (XGB), Decision Tree (DT), Logistic Regression (LR), K-nearest Neighbour (KNN) και Gaussian Naive Bayes (GNB)** και αναλύθηκε η απόδοση των επιμέρους ταξινομητών για την αξιολόγησή τους και την δημιουργία του πιο αποτελεσματικού μοντέλου. Περαιτέρω, αφαιρέθηκαν οι ασήμαντες μεταβλητές χρησιμοποιώντας μεθόδους επιλογής χαρακτηριστικών και χρησιμοποιήθηκαν οι ταξινομητές σε πρωτογενή και παραγόμενα υποσύνολα. Οι ερευνητές κατέληξαν στο συμπέρασμα ότι είναι πραγματικά δύσκολο να εντοπιστεί ο καλύτερος αλγόριθμος για όλες τις περιπτώσεις και τα υποσύνολα ωστόσο η επιλογή χαρακτηριστικών παρείχε καλά αποτελέσματα μειώνοντας τα ανεπιθύμητα χαρακτηριστικά. [19]

CT	Acc	AUROC	F-M	Acc	AUROC	F-M	Acc	AUROC	F-M	Acc	AUROC	F-M
	Baseline			Min-Max			Z-Transformion			Square-root		
	CFS			CFS			CFS			CFS		
GB	0.7630	0.6835	0.7288	0.7451	0.6750	0.7183	0.7451	0.6750	0.7183	0.7451	0.6750	0.7183
RF	0.7627	0.6828	0.7281	0.5529	0.5236	0.5540	0.5533	0.5239	0.5542	0.5529	0.5235	0.5539
XGB	0.7628	0.6827	0.7281	0.6870	0.6567	0.6841	0.6870	0.6567	0.6841	0.6870	0.6567	0.6841
DT	0.7627	0.6828	0.7281	0.5111	0.4933	0.5172	0.5113	0.4936	0.5174	0.5132	0.4954	0.5192
LR	0.7550	0.6717	0.7163	0.7370	0.6966	0.7291	0.7362	0.6986	0.7296	0.7687	0.7305	0.7618
KNN	0.5245	0.5954	0.5021	0.4901	0.4754	0.4974	0.5870	0.5638	0.5892	0.6164	0.5900	0.6167
GNB	0.5629	0.5986	0.5645	0.7273	0.6647	0.7059	0.7438	0.6692	0.7129	0.6950	0.6508	0.6854
	GRAE			GRAE			GRAE			GRAE		
GB	0.7750	0.7086	0.7525	0.7419	0.6750	0.7183	0.7451	0.6750	0.7183	0.7451	0.6750	0.7183
RF	0.7559	0.7262	0.7522	0.5529	0.5236	0.5540	0.5529	0.5236	0.5540	0.5529	0.5235	0.5539
XGB	0.7921	0.7521	0.7281	0.6870	0.6567	0.6841	0.6870	0.6567	0.6841	0.6870	0.6567	0.6841
DT	0.7194	0.6996	0.7195	0.5135	0.4956	0.5195	0.5131	0.4951	0.5191	0.5119	0.4937	0.5179
LR	0.7722	0.7250	0.7612	0.7370	0.6966	0.7291	0.7370	0.6966	0.7291	0.7687	0.7305	0.7618
KNN	0.6697	0.6341	0.6649	0.4901	0.4754	0.4974	0.4901	0.4754	0.4974	0.6164	0.5900	0.6167
GNB	0.6119	0.6572	0.6111	0.7273	0.6647	0.7059	0.7273	0.6647	0.7059	0.6950	0.6508	0.6854
	IGAE			IGAE			IGAE			IGAE		
GB	0.7690	0.7102	0.7514	0.7690	0.7102	0.7514	0.7690	0.7102	0.7514	0.7690	0.7102	0.7514
RF	0.7561	0.7263	0.7524	0.7560	0.7263	0.7523	0.7560	0.7261	0.7523	0.7552	0.7253	0.7515
XGB	0.7915	0.7524	0.7843	0.7915	0.7524	0.7843	0.7915	0.7524	0.7843	0.7915	0.7524	0.7843
DT	0.7189	0.6982	0.7187	0.7180	0.6968	0.7176	0.7190	0.6984	0.7188	0.7204	0.6996	0.7202
LR	0.7688	0.7211	0.7575	0.7691	0.7175	0.7557	0.7691	0.7213	0.7578	0.7805	0.7361	0.7709
KNN	0.6730	0.6373	0.6681	0.7590	0.7324	0.7563	0.7542	0.7270	0.7514	0.7393	0.7144	0.7375
GNB	0.7447	0.6857	0.7263	0.6202	0.6629	0.6206	0.5312	0.6103	0.5017	0.6390	0.6775	0.6410

Πίνακας 5 Αποτελέσματα Αλγορίθμων Md. Shahriare Satu, Khair Ahammed και Mohammad Zoynul Abedin [19]

Το 2020 οι **Zharfan Akbar Andriawan, Satriawan Rasyid Purnama, Adam Sukma Darmawan, Ricko, Adi Wibowo, Aris Sugiharto, Feri Wijayanto** δημοσίευσαν την έρευνα τους σχετικά με την πρόβλεψη της ακύρωσης της κράτησης με τη βοήθεια της επιστήμης δεδομένων. Με τη χρήση συνόλων δεδομένων που σχετίζονται με αιτήματα κρατήσεων ξενοδοχείων πραγματοποίησαν μια προγνωστική ανάλυση χρησιμοποιώντας το πλαίσιο Cross-Industry Standard Process for Data Mining (CRISP-DM) με το οποίο δημιουργούνται οι πιο κατάλληλες λειτουργίες για το μοντέλο μηχανικής εκμάθησης. Εκτελώντας πρώτα ορισμένες διαδικασίες προετοιμασίας δεδομένων επιλέχθηκαν και εφαρμόστηκαν παραλλαγές στα μοντέλα μηχανικής μάθησης ώστε να αναπτυχθεί το καλύτερο μοντέλο. Το μοντέλο που χρησιμοποίησαν είναι ένας αλγόριθμος που βασίζεται σε δέντρα που περιλαμβάνει XGBoost (XGB), Catboost, Light Gradient Boosting Machine (LGBM) και Random Forest (RF). Η αξιολόγηση των αλγορίθμων βασίστηκε στις τιμές που λαμβάνουν στο accuracy, το recall, το precision και τις τιμές F1 που παράγονται από κάθε μοντέλο. [20]

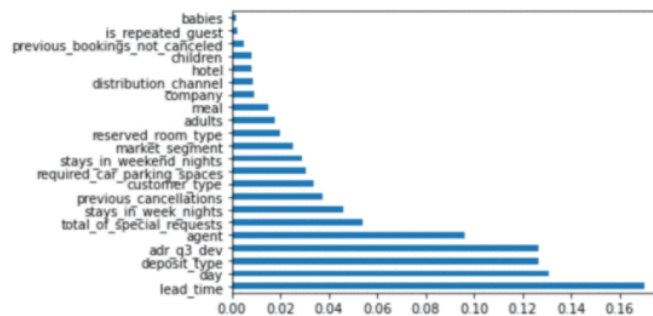
No.	Modelling Result				
	Model	Accuracy	Recall	Precision	F1
1.	XGB	0.8227	0.6141	0.8688	0.7195
2.	Catboost	0.8484	0.7105	0.8557	0.7763
3.	LGBM	0.8380	0.6739	0.8580	0.755
4.	RF	0.8717	0.7750	0.8646	0.8173

Πίνακας 6 Αξιολόγηση αλγορίθμων Zharfan Akbar Andriawan, Satriawan Rasyid Purnama, Adam Sukma Darmawan, Ricko, Adi Wibowo, Aris Sugiharto, Feri Wijayanto [20]

Οι συγγραφείς χρησιμοποίησαν 10-fold Cross Validation, μια τεχνική αξιολόγησης μοντέλων που είναι αρκετά δημοφιλής και ευρέως χρησιμοποιούμενη. Στη συνέχεια, αφού τα αποτελέσματα έδειξαν ότι το μοντέλο Random Forest είναι το καλύτερο μοντέλο ρυθμίστηκε η υπερπαραμέτρος του μοντέλου για την λήψη της καλύτερης βαθμολογίας λαμβάνοντας τιμή 0,8725. [20]

Modelling Result				
Model	Accuracy	Recall	Precision	F1
RF	0.8717	0.7750	0.8646	0.8173
After Tuning RF	0.8725	0.7826	0.8606	0.8197

Πίνακας 7 Αξιολόγηση Random Forest σύμφωνα με την μελέτη των Zharfan Akbar Andriawan, Satriawan Rasyid Purnama, Adam Sukma Darmawan, Ricko, Adi Wibowo, Aris Sugiharto, Feri Wijayanto [20]



Εικόνα 3 Παράγοντες επιρροής ακύρωσης σύμφωνα με τους Zharfan Akbar Andriawan, Satriawan Rasyid Purnama, Adam Sukma Darmawan, Ricko, Adi Wibowo, Aris Sugiharto, Feri Wijayanto [20]

Τέλος, σύμφωνα με τους ερευνητές ο χρόνος που μεσολαβεί μεταξύ της στιγμής πραγματοποίησης της κράτησης και της ημέρας της άφιξης είναι ο πιο σημαντικός παράγοντας επιρροής της ακύρωσης και το πιο σημαντικό χαρακτηριστικό στην πρόβλεψη της ακύρωσης.[20]

Σε μελέτη που πραγματοποιήθηκε πρόσφατα και συγκεκριμένα το 2023 οι **Myongjee Yoo, Ashok K. Singh, Noah Loewy** προσπάθησαν με την χρήση της μηχανικής μάθησης να αναπτύξουν ένα μοντέλο που μπορεί να προβλέψει με ακρίβεια τις ακυρώσεις δωματίων ξενοδοχείου και να εντοπίζει τους βασικούς παράγοντες ακύρωσης. Για την μελέτη χρησιμοποίησαν δευτερογενή δεδομένα για δύο ξενοδοχεία της Πορτογαλίας εκ των οποίων το ένα είναι θέρετρο και το άλλο είναι ξενοδοχείο πόλης. Και τα δύο σύνολα δεδομένων των δύο ξενοδοχείων είχαν ακριβώς την ίδια δομή και αποτελούνται συνολικά από 31 μεταβλητές. Τα δεδομένα που αναλύθηκαν αφορούσαν κρατήσεις με αφίξεις τη χρονική περίοδο από 1 Ιουλίου 2015 έως 31 Αυγούστου 2017. Για την ανάπτυξη του μοντέλου χρησιμοποίησαν τρεις διαφορετικούς αλγόριθμους ταξινόμησης και συγκεκριμένα το **XGBoost**, **Random Forest** καθώς και το **Support Vector Machine** και συνδυάζοντας τις μεθόδους αυτές με έναν αλγόριθμο ψηφοφορίας ανέπτυξαν έναν **υβριδικό ταξινομητή** που χαρακτηρίζεται ως **μέθοδος συνόλου** με απώτερο σκοπό την αύξηση της απόδοσης. Για το ξενοδοχείο με τα στοιχεία H1 (θήρετρο), ο ταξινομητής Random Forest εμφανίζει χαμηλή ανάκληση και ακρίβεια, αλλά τόσο το Support Vector Machine όσο και το XGBoost παρέχουν ταξινόμηση με υψηλή ακρίβεια ενώ και η μέθοδος του συνόλου αποδίδει επίσης καλά. Για το H2 – ξενοδοχείο πόλης, ο ταξινομητής XGBoost δείχνει υψηλή ανάκληση και ακρίβεια, το Support Vector Machine όσο και το Random Forest αποδίδουν αρκετά καλά αλλά και η μέθοδος συνόλου αποδίδει την καλύτερη συνολική απόδοση.

Σύμφωνα με την ίδια μελέτη σχετικά με τους παράγοντες που τελικώς επηρεάζουν την ακύρωση τα αποτελέσματα έδειξαν ότι αυτοί είναι :

- **LeadTime**: ο αριθμός των ημερών που μεσολαβεί μεταξύ της Κράτησης και της ημέρας Άφιξης
- **ADR (average daily rate)** : Μέση τιμή δωματίου ανά ημέρα

- **Χώρα προέλευσης:** Επισκέπτες που προέρχονται από την Πορτογαλία είναι πιο πιθανό να ακυρώσουν
- **MarketSegment:** Επισκέπτες που ανήκουν στην Κατηγορία “Online Travel Agents” – Διαδικτυακοί τουριστικοί πράκτορες είναι επίσης πιθανό να ακυρώσουν
- **Επισκέπτες με περισσότερες προηγούμενες ακυρώσεις κρατήσεων και εκείνου που δίνουν προτεραιότητα στις μη επιστρεπτές προκαταβολές είναι πιο πιθανό να ακυρώσουν τις κρατήσεις δωματίων τους.**

Τέλος, η μελέτη απέδειξε ότι οι επισκέπτες που εμπίπτουν στην κατηγορία MarketSegments “Direct and Offline TA (Travel agents)/TO (Tour Operators)”, οι επισκέπτες με μεγαλύτερο αριθμό TotalSpecialRequests (αριθμός συνολικών ειδικών αιτημάτων που υπέβαλε ο πελάτης- π.χ. ψηλός όροφος ή διπλό κρεβάτι) είναι λιγότερο πιθανό να ακυρώσουν. [8]

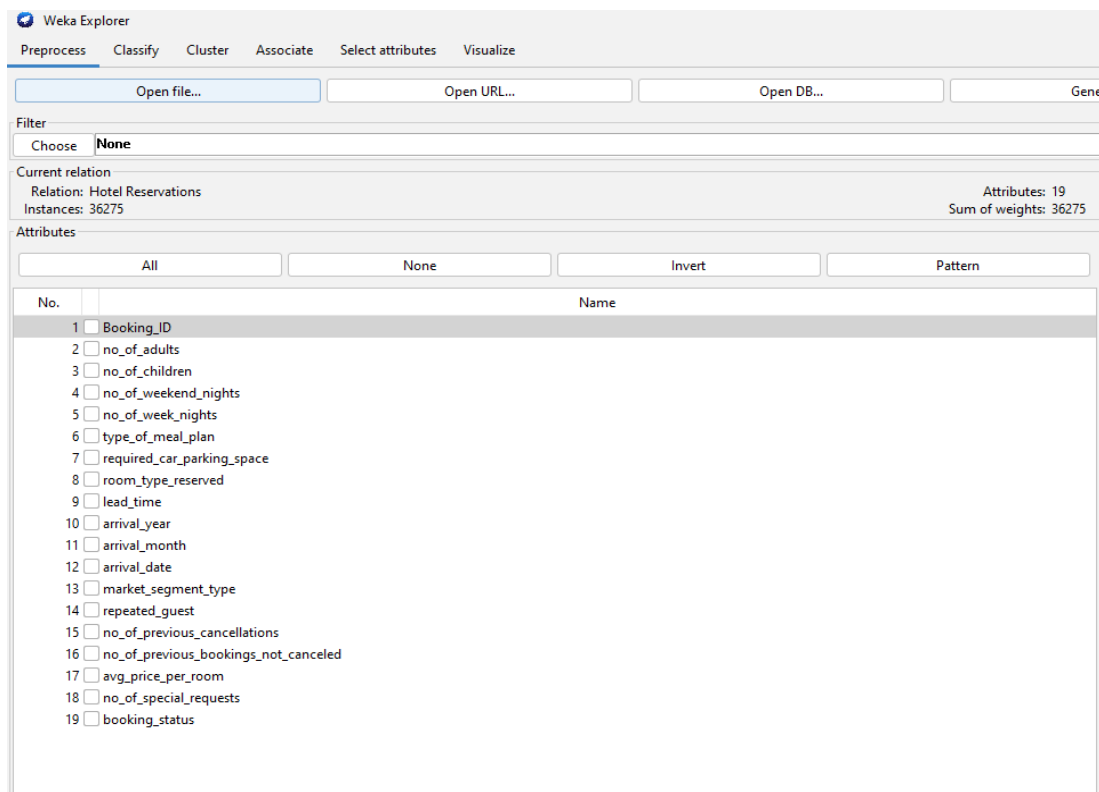
5. Περιγραφή Δείγματος

Το δείγμα της παρούσας εργασίας αποτελούν δευτερογενή δεδομένα τα οποία αντλήθηκαν από το αποθετήριο δεδομένων της ιστοσελίδας Kaggle[21]. Από την σελίδα άντλησης των δεδομένων προκύπτει ότι τα δεδομένα τα οποία αντλήθηκαν και πρόκειται να αναλύσουμε στην παρούσα μελέτη αποτελούν μέρος ενός συνόλου δεδομένων που χρησιμοποιήθηκαν στα πλαίσια δημοσιευμένης έρευνας με τίτλο “Hotel booking demand datasets” κατά το έτος 2019 των Antonio Nuno, Ana de Almeida και Nunes Luis. Το σύνολο των δεδομένων αποτελείται από 31 μεταβλητές με δεδομένα κρατήσεων ξενοδοχείων και αφορούν δύο ξενοδοχεία τα οποία βρίσκονται στην Πορτογαλία. Το ένα από τα ξενοδοχεία (H1) είναι ένα ξενοδοχείο θέρετρο και το άλλο είναι ένα ξενοδοχείο πόλης (H2).

Για την παρούσα μελέτη το σετ δεδομένων μας αποτελείται από τα 18 χαρακτηριστικά (εκ του συνόλου των 31) τα οποία αναλύονται κατωτέρω και έχουν επεξεργασθεί και εξαχθεί σε αρχείο csv από τον διαθέτη της ιστοσελίδας kaggle. Κάθε γραμμή του αρχείου αντιπροσωπεύει μια κράτηση ξενοδοχείου. Το σετ δεδομένων που θα αναλυθεί με την παρούσα μελέτη περιλαμβάνει στοιχεία 36.275 κρατήσεων που επρόκειτο να φτάσουν μεταξύ της 1ης Ιανουαρίου 2017 έως και της 31ης Δεκεμβρίου 2018, συμπεριλαμβανομένων των κρατήσεων που πράγματι έφτασαν και των κρατήσεων που ακυρώθηκαν. Τέλος, καθώς πρόκειται για πραγματικά δεδομένα ξενοδοχείων λόγω προστασίας των προσωπικών δεδομένων διαγράφηκαν όλα τα στοιχεία δεδομένων που αφορούν την ταυτοποίηση του ξενοδοχείου ή του πελάτη της κράτησης.

Το σετ δεδομένων αποτελείται από δεδομένα διακριτά και αριθμητικά.

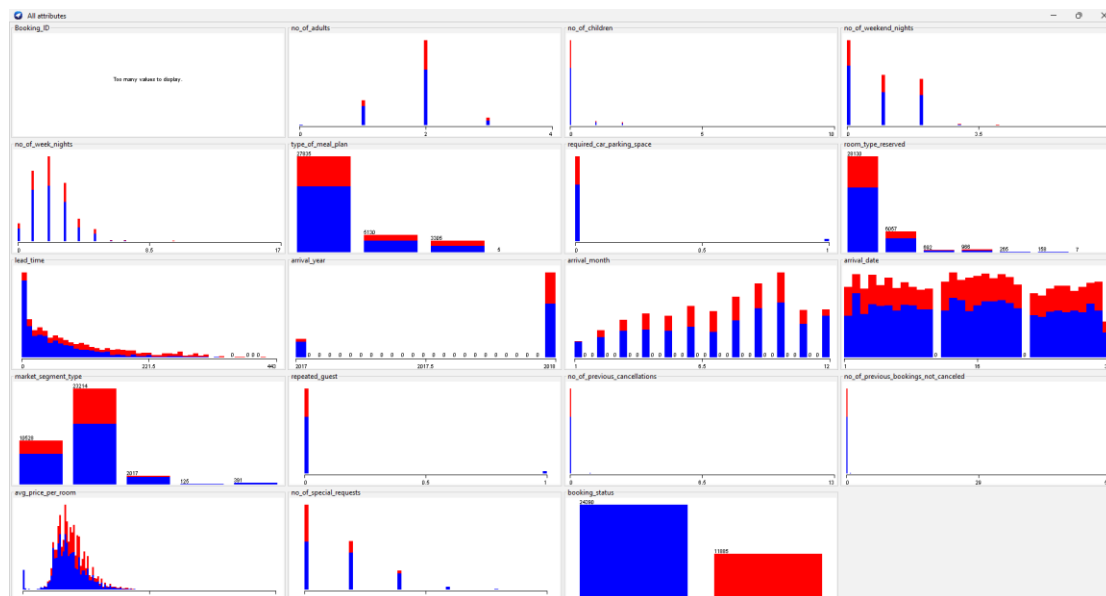
Συγκεκριμένα το δείγμα αποτελείται από τα εξής χαρακτηριστικά:



Εικόνα 4 Χαρακτηριστικά Δείγματος παρούσας μελέτης

1. **Booking_ID:** Μοναδικός Κωδικός Ταυτοποίησης Κάθε Κράτησης
2. **no_of_adults:** Αριθμός ενηλίκων κράτησης.
3. **no_of_children:** Αριθμός των παιδιών κράτησης.
4. **no_of_weekend_nights:** Αριθμός νυκτών Σαββατοκύριακου (Σαββάτου ή Κυριακής) στις οποίες ο πελάτης έμεινε ή έκλεισε προς διανυκτέρευση στο Ξενοδοχείο
5. **no_of_week_nights:** Αριθμός καθημερινών νυκτών (Δευτέρα έως Παρασκευή) στις οποίες ο πελάτης έμεινε ή έκλεισε προς διανυκτέρευση στο Ξενοδοχείο.
6. **type_of_meal_plan:** Τύπος διατροφής που έκλεισε ο πελάτης
7. **required_car_parking_space:** απαίτηση πελάτη για χώρο στάθμευσης (0 - No, 1- Yes)
8. **room_type_reserved:** Τύπος δωματίου Κράτησης.
9. **lead_time:** Αριθμός ημερών μεταξύ της ημερομηνίας καταχώρισης της κράτησης στην πλατφόρμα PMS του ξενοδοχείου και της ημερομηνίας άφιξης
10. **arrival_year:** Έτος Άφιξης
11. **arrival_month:** Μήνας άφιξης
12. **arrival_date:** Ημέρα μήνα άφιξης
13. **market_segment_type:** Προσδιορισμός τμήματος της αγοράς.
14. **repeated_guest:** Προσδιορισμός πελάτη ως επαναλαμβανόμενος επισκέπτης ή μη (0=No, 1= Yes)
15. **no_of_previous_cancellations:** Αριθμός προηγούμενων κρατήσεων που ακυρώθηκαν από τον πελάτη πριν από την τρέχουσα κράτηση.
16. **no_of_previous_bookings_not_canceled:** Αριθμός προηγούμενων κρατήσεων που δεν ακυρώθηκαν από τον πελάτη πριν από την τρέχουσα κράτηση.
17. **avg_price_per_room:** Μέση τιμή ανά ημέρα κράτησης σε ευρώ
18. **no_of_special_requests:** Συνολικός αριθμός ειδικών αιτημάτων που υποβλήθηκαν από τον πελάτη (π.χ. ψηλός όροφος, θέα από το δωμάτιο κ.λπ.)
19. **booking_status:** Ακυρωμένη ή μη ακυρωμένη κράτηση. (Έξοδος των Πειραμάτων της παρούσας μελέτης)

Ακολουθεί συνολική εικόνα από τα γραφήματα των χαρακτηριστικών του weka:



Εικόνα 5 Γραφήματα WEKA

Αναλυτικότερα:

Το Booking ID είναι μοναδικός κωδικός της κάθε κράτησης και δεν αποτελεί χρήσιμο στοιχείο στην παρούσα μελέτη.

Τα υπόλοιπα στοιχεία των κρατήσεων αναλύονται ως ακολούθως:

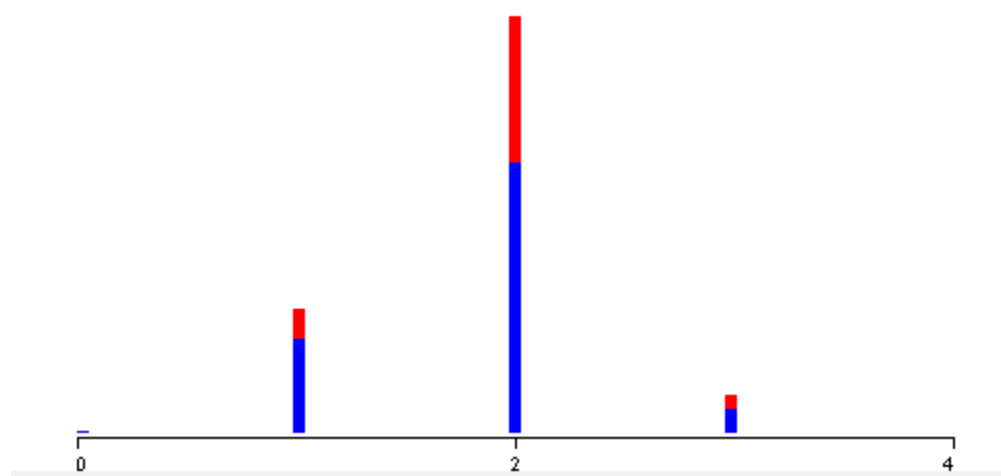
1. Αριθμός ενηλίκων:

Στο παρόν δείγμα ο μικρότερος αριθμός ενηλίκων της Κράτησης είναι το 0 και ο μεγαλύτερος το 4.

No of adults	Value
MINIMUM	0
MAXIMUM	4
MEAN	1,845
STDDEV	0,519

Πίνακας 8 Αριθμός Ενηλίκων Κράτησης

Ακολουθεί γραφική παράσταση από το Weka:



Εικόνα 6 Γραφική Παράσταση WEKA αριθμού ενηλίκων κράτησης

Missing Values	Most Frequent Value
0 (0%)	[1.973 , 2.027] (26108)

Πίνακας 9 Missing Values / most Frequent Value αριθμού ενηλίκων κράτησης

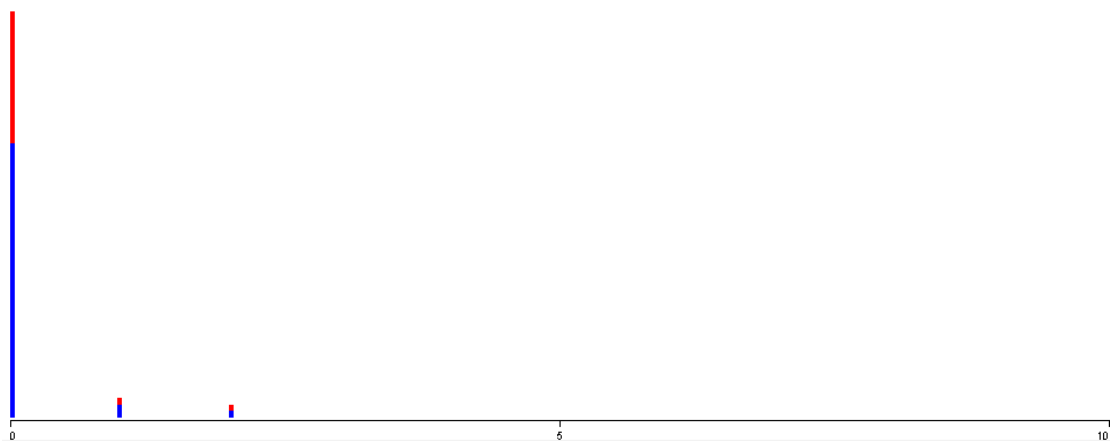
2. Αριθμός των παιδιών

Μέγιστος αριθμός των παιδιών που παρατηρούνται στις Κρατήσεις είναι το 10 και μικρότερος το 0.

No of children	Value
MINIMUM	0
MAXIMUM	10
MEAN	0,105
STDDEV	0,403

Πίνακας 10 Αριθμός παιδιών κράτησης

Ακολουθεί γραφική παράσταση από το Weka:



Εικόνα 7 Γραφικής Παράσταση WEKA Αριθμού παιδιών κράτησης

Missing Values	Most Frequent Value
0 (0%)	[0 , 0.042] (33577)

Πίνακας 11 Missing Values / most Frequent Value αριθμού παιδιών κράτησης

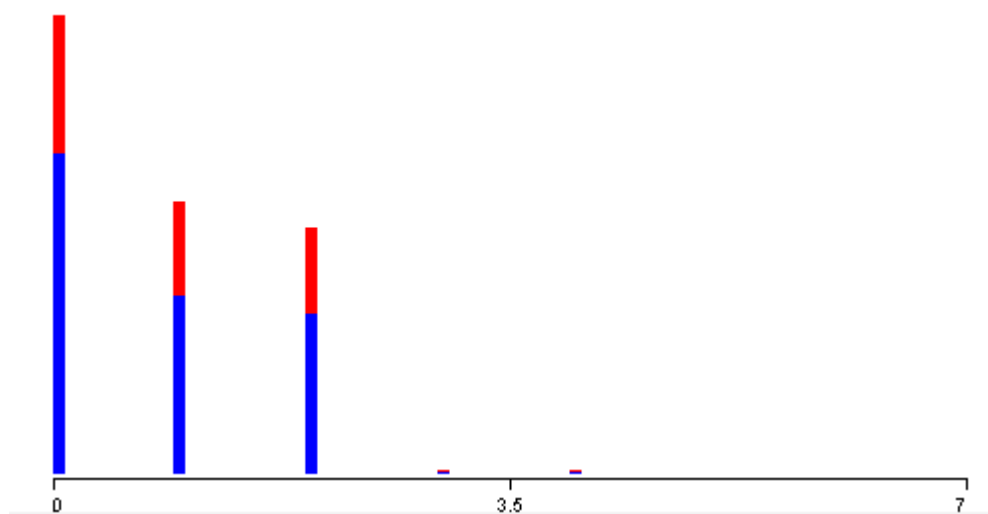
3. Αριθμός νυκτών Σαββατοκύριακου (Σαββάτου ή Κυριακής) στις οποίες ο πελάτης έμεινε ή έκλεισε προς διανυκτέρευση στο Ξενοδοχείο

Μέγιστος αριθμός νυκτών Σαββατοκύριακου 7 και ελάχιστος το 0.

No of weekend nights	Value
MINIMUM	0
MAXIMUM	7
MEAN	0,811
STDDEV	0,871

Πίνακας 12 Αριθμός νυκτών Σαββατοκύριακου (Σαββάτου ή Κυριακής) στις οποίες ο πελάτης έμεινε ή έκλεισε προς διανυκτέρευση στο Ξενοδοχείο

Ακολουθεί γραφική παράσταση από το Weka:



Εικόνα 8 Γραφική παράσταση αριθμού νυκτών Σαββατοκύριακου (Σαββάτου ή Κυριακής) στις οποίες ο πελάτης έμεινε ή έκλεισε προς διανυκτέρευση στο Ξενοδοχείο

Missing Values	Most Frequent Value
0 (0%)	[0, 0.092] (16872)

Πίνακας 13 Missing Values / most Frequent Value αριθμού νυκτών Σαββατοκύριακου (Σαββάτου ή Κυριακής) στις οποίες ο πελάτης έμεινε ή έκλεισε προς διανυκτέρευση στο Ξενοδοχείο

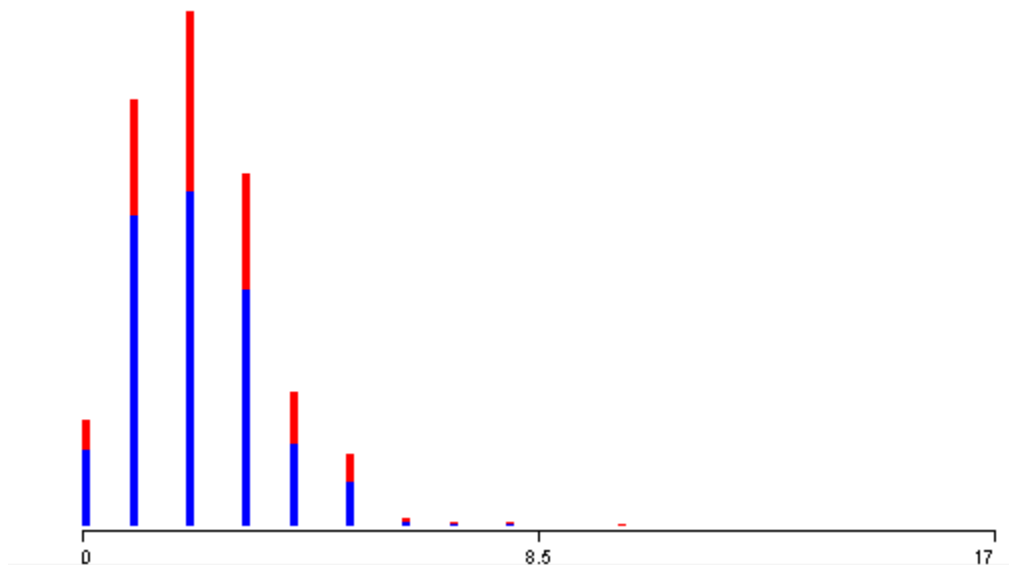
4. Αριθμός καθημερινών νυκτών (Δευτέρα έως Παρασκευή) στις οποίες ο πελάτης έμεινε ή έκλεισε προς διανυκτέρευση στο Ξενοδοχείο.

Μέγιστος αριθμός καθημερινών νυκτών 17 και ελάχιστος το 0.

No of week nights	Value
MINIMUM	0
MAXIMUM	17
MEAN	2,204
STDDEV	1,411

Πίνακας 14 Αριθμός καθημερινών νυκτών (Δευτέρα έως Παρασκευή) στις οποίες ο πελάτης έμεινε ή έκλεισε προς διανυκτέρευση στο Ξενοδοχείο

Ακολουθεί γραφική παράσταση από το Weka:



Εικόνα 9 Γραφική Παράσταση WEKA αριθμού καθημερινών νυκτών (Δευτέρα έως Παρασκευή) στις οποίες ο πελάτης έμεινε ή έκλεισε προς διανυκτέρευση στο Ξενοδοχείο

Missing Values	Most Frequent Value
0 (0%)	[1.939, 2.088] (11444)

Πίνακας 15 Missing Values / most Frequent Value Αριθμού καθημερινών νυκτών (Δευτέρα έως Παρασκευή) στις οποίες ο πελάτης έμεινε ή έκλεισε προς διανυκτέρευση στο Ξενοδοχείο

5. Τύπος διατροφής που έκλεισε ο πελάτης

Ο τύπος του διατροφικού πλάνου που έχει κλείσει ο πελάτης με την Κράτηση διακρίνεται σε 4 τιμές, οι οποίες είναι οι εξής:

- Meal Plan 1
- Meal Plan 2
- Meal Plan 3
- Not Selected

Από την διαδικτυακή σελίδα άντλησης των δεδομένων καθώς και από την δημοσιευμένη έρευνα προκύπτει ότι:

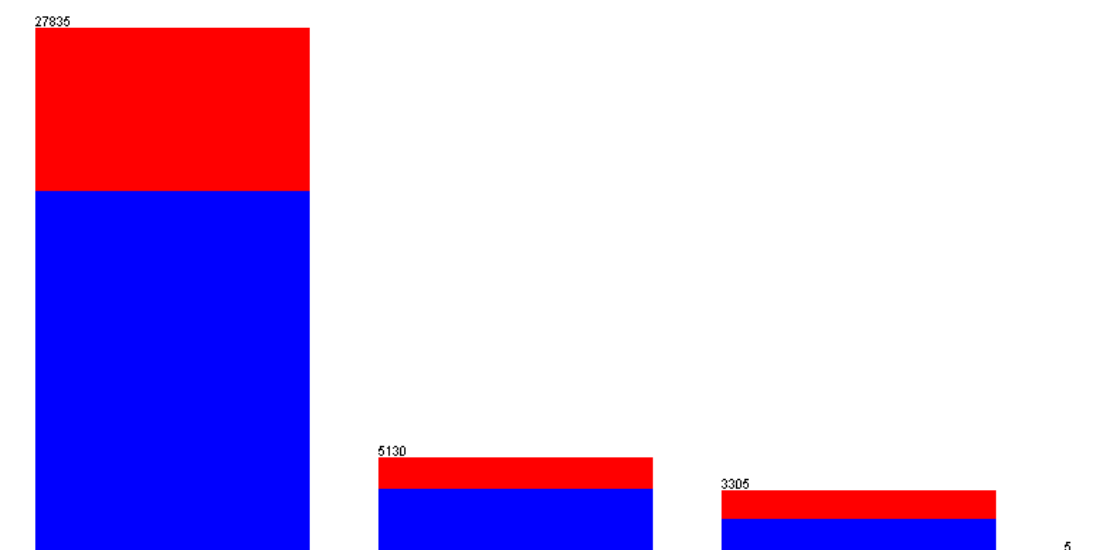
- Το Meal Plan 1 είναι Κράτηση με παροχή πρωινού (Bed and Breakfast)
- Το Meal Plan 2 είναι Κράτηση με Ημιδιατροφή (Half Board)
- Το Meal Plan 3 είναι Κράτηση με Πλήρη διατροφή (Full Board)
- Το Not Selected συναντάται στο δείγμα όταν δεν έχει επιλεγθεί meal plan από τον πελάτη

Στο παρόν δείγμα 27.835 Κρατήσεις έχουν επιλέξει το Meal Plan 1, 5.130 δεν έχουν επιλέξει κάποιο διατροφικό πλάνο με την διαμονή τους, 3.305 έχουν επιλέξει Meal Plan 2 και μόλις 5 Κρατήσεις επέλεξαν το Meal Plan 3.

Type_of_meal_plan	Count
Meal Plan 1	27,835
Not Selected	5,130
Meal Plan 2	3,305
Meal Plan 3	5

Πίνακας 16 Τιμές τύπου διατροφής

Ακολουθεί γραφική παράσταση από το Weka:



Εικόνα 10 Γραφική Παράσταση τιμών τύπου διατροφής

Missing Values	Most Frequent Value
0 (0%)	Meal Plan 1 (27835)

Πίνακας 17 Missing Values / most Frequent Value τιμών τύπου διατροφής

6. Απαίτηση πελάτη Κράτησης για χώρο στάθμευσης

Το χαρακτηριστικό του δείγματος σχετικά με την απαίτηση του πελάτη για χώρο στάθμευσης λαμβάνει τιμές 0 όταν ο πελάτης δεν ζητά την παραχώρηση χώρου στάθμευσης και 1 όταν ζητά.

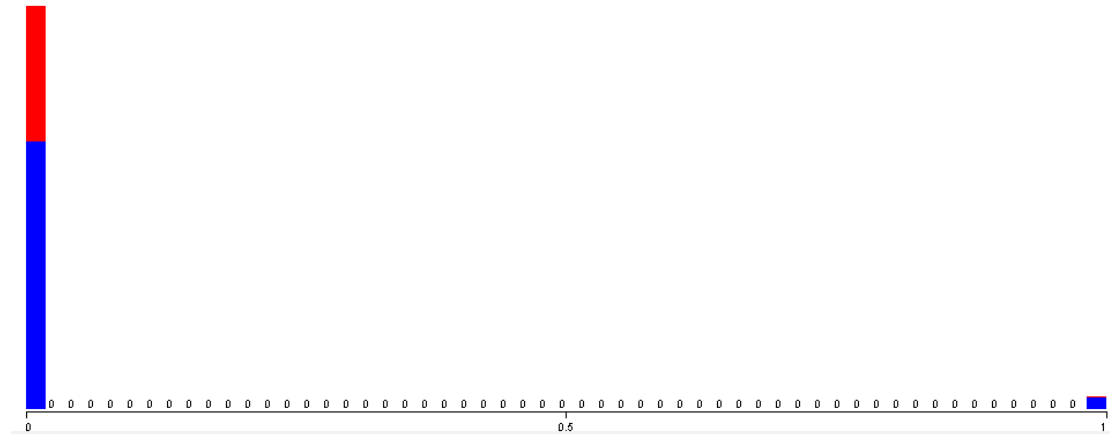
0 = Όχι

1 = Ναι

required_car_parking_space	Value
MINIMUM	0
MAXIMUM	1
MEAN	0.031
STDDEV	0.173

Πίνακας 18 Τιμές που λαμβάνει το αίτημα για parking

Ακολουθεί γραφική παράσταση από το Weka:



Εικόνα 11 Γραφική Παράσταση WEKA για τις τιμές του αιτήματος parking

Missing Values	Most Frequent Value
0 (0%)	[0,0.018](35151)

Πίνακας 19 Missing Values / most Frequent Value τιμών για το αίτημα για parking

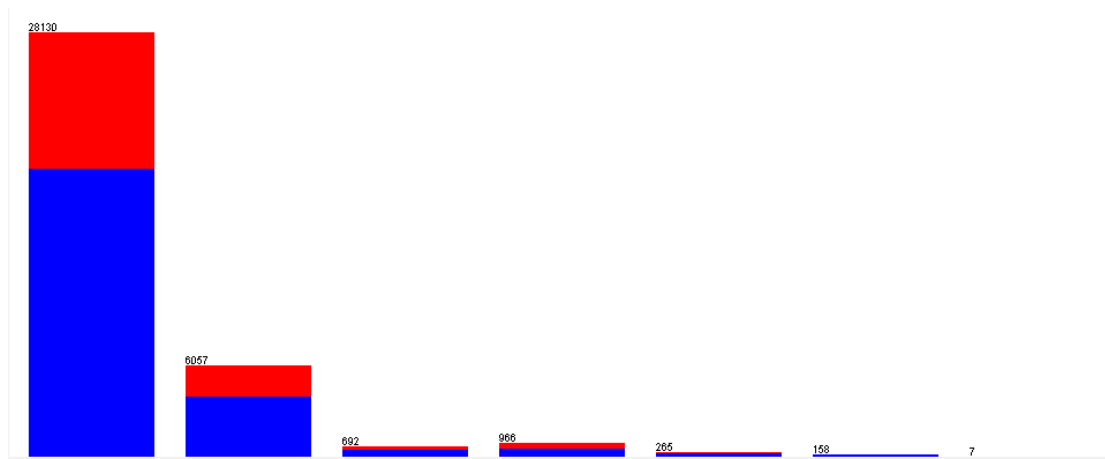
7. Τύπος δωματίου Κράτησης

Για λόγους προστασίας των δεδομένων και μη ταυτοποίησης των Ξενοδοχείων δεν διευκρινίζεται μήτε από την ιστοσελίδα άντλησης των δεδομένων (Kaggle) μήτε από τους συντάκτες του επιστημονικού άρθρου ο τύπος του δωματίου.

Room_type_reserved	Count
Room Type 1	28,130
Room Type 4	6,057
Room Type 2	692
Room Type 6	966
Room Type 5	265
Room Type 7	158
Room Type 3	7

Πίνακας 20 Τύποι Δωματίου - τιμές

Ακολουθεί γραφική παράσταση από το Weka:



Εικόνα 12 Γραφική Παράσταση WEKA για τις τιμές που λαμβάνουν οι τύποι δωματίων

Missing Values	Most Frequent Value
0 (0%)	Room Type 1 (28,130)

Πίνακας 21 Missing Values / most Frequent Value τιμών τύπων δωματίου

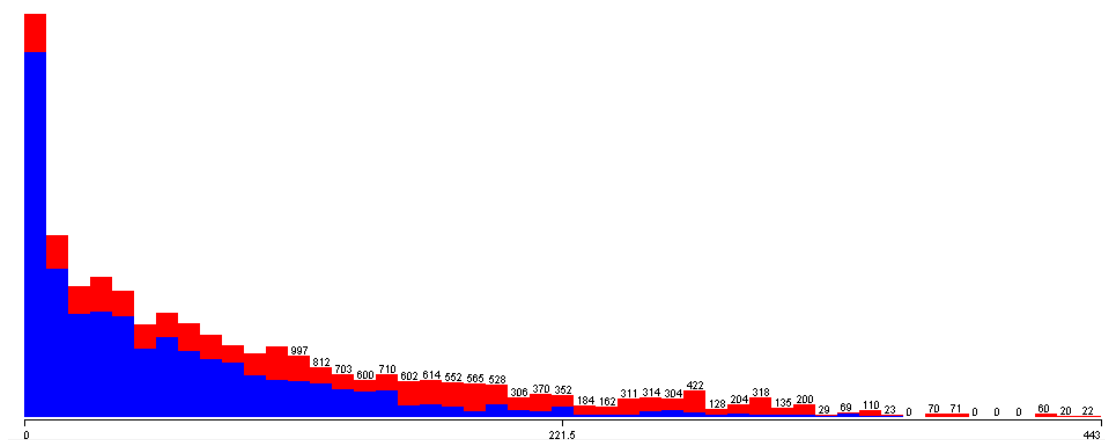
8. Αριθμός ημερών μεταξύ της ημερομηνίας καταχώρισης της κράτησης στην πλατφόρμα PMS του ξενοδοχείου και της ημερομηνίας άφιξης

Από το δείγμα προκύπτει ότι η ελάχιστη τιμή είναι 0 ημέρες και η μέγιστη τιμή 443 από τη στιγμή που καταχωρείται η Κράτηση στην πλατφόρμα κρατήσεων του ξενοδοχείου έως και την ημερομηνία της Άφιξης.

Lead_time	Value
MINIMUM	0
MAXIMUM	443
MEAN	85.233
STDDEV	85.931

Πίνακας 22 Αριθμός ημερών μεταξύ της ημερομηνίας καταχώρισης της κράτησης στην πλατφόρμα PMS του ξενοδοχείου και της ημερομηνίας άφιξης

Ακολουθεί γραφική παράσταση από το Weka:



Εικόνα 13 Γραφική Παράσταση αριθμού ημερών μεταξύ της ημερομηνίας καταχώρισης της κράτησης στην πλατφόρμα PMS του ξενοδοχείου και της ημερομηνίας άφιξης

Missing Values	Most Frequent Value
0 (0%)	[0,9.041] (6569)

Πίνακας 23 Missing Values / most Frequent Value αριθμού ημερών μεταξύ της ημερομηνίας καταχώρισης της κράτησης στην πλατφόρμα PMS του ξενοδοχείου και της ημερομηνίας άφιξης

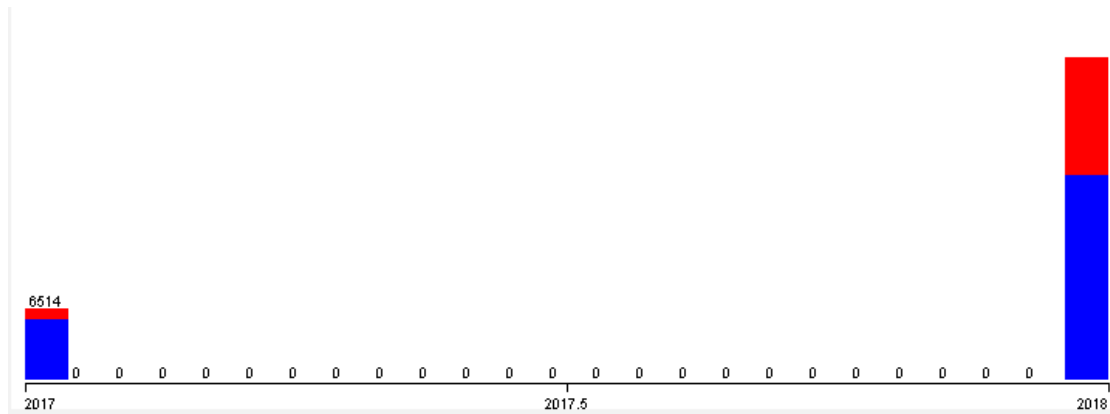
9. Έτος Άφιξης

Το δείγμα μας περιλαμβάνει δεδομένα Κρατήσεων κατά τις οποίες έτη άφιξης είναι τα 2017 και 2018.

No of adults	Value
MINIMUM	2017
MAXIMUM	2018
MEAN	2017.82
STDDEV	0.384

Πίνακας 24 Τιμές έτους άφιξης

Ακολουθεί γραφική παράσταση από το Weka:



Εικόνα 14 Γραφική Παράσταση WEKA για τις τιμές του έτους άφιξης

Missing Values	Most Frequent Value
0 (0%)	[2017.96,2018] (29761)

Πίνακας 25 Missing Values / most Frequent Value τιμών έτους άφιξης

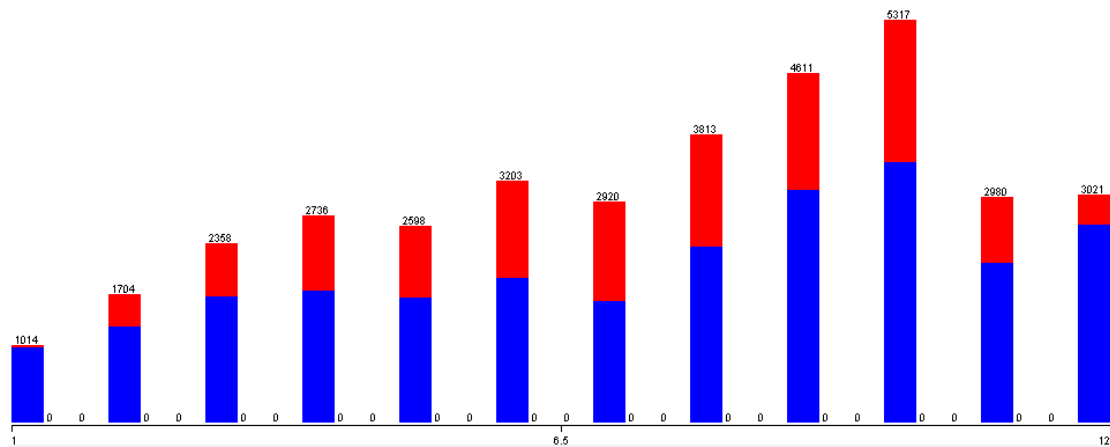
10. Μήνας άφιξης

Το δείγμα που μελετάται στην παρούσα εργασία αφορά κρατήσεις με άφιξη από 1/1/2017 έως 31/12/2018.

Arrival_Month	Value
MINIMUM	1
MAXIMUM	12
MEAN	7.424
STDDEV	3.07

Πίνακας 26 Τιμές μήνα άφιξης

Ακολουθεί γραφική παράσταση από το Weka:



Εικόνα 15 Γραφική Παράσταση WEKA για τις τιμές του μήνα της άφιξης

Missing Values	Most Frequent Value
0 (0%)	[9.735,10.059] (5317)

Πίνακας 27 Missing Values / most Frequent Value τιμών μήνα άφιξης

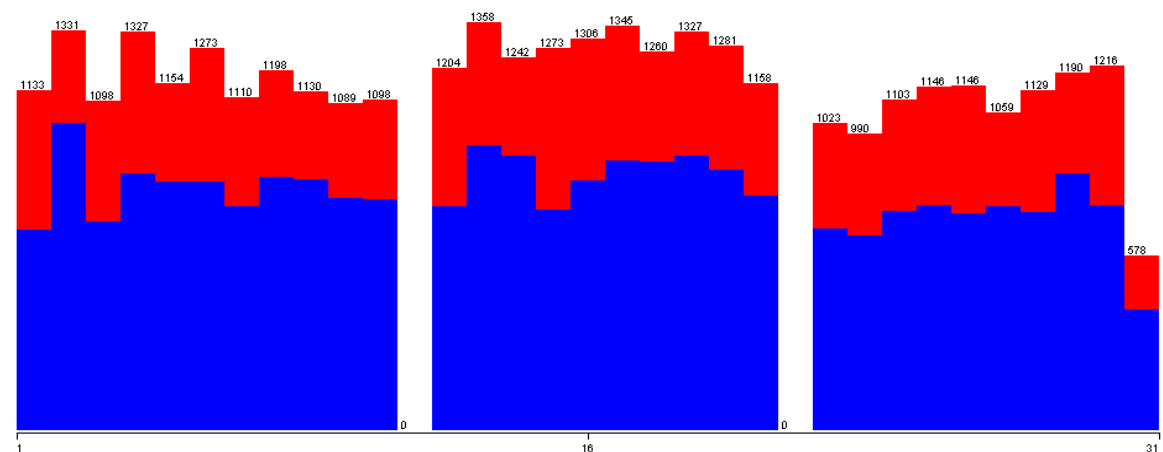
11. Ημέρα μήνα άφιξης

Το χαρακτηριστικό της ημέρας λαμβάνει τιμές από 1 έως 31 όσες και οι μέρες του μήνα.

arrival_date	Value
MINIMUM	1
MAXIMUM	31
MEAN	15.597
STDDEV	8.74

Πίνακας 28 Τιμές ημέρας άφιξης

Ακολουθεί γραφική παράσταση από το Weka:



Εικόνα 16 Γραφική παράσταση WEKA τιμών ημέρας άφιξης

Missing Values	Most Frequent Value
0 (0%)	[12.818,13.727] (1358)

Πίνακας 29 Missing Values / most Frequent Value τιμών ημέρας άφιξης

12. Προσδιορισμός τμήματος της αγοράς

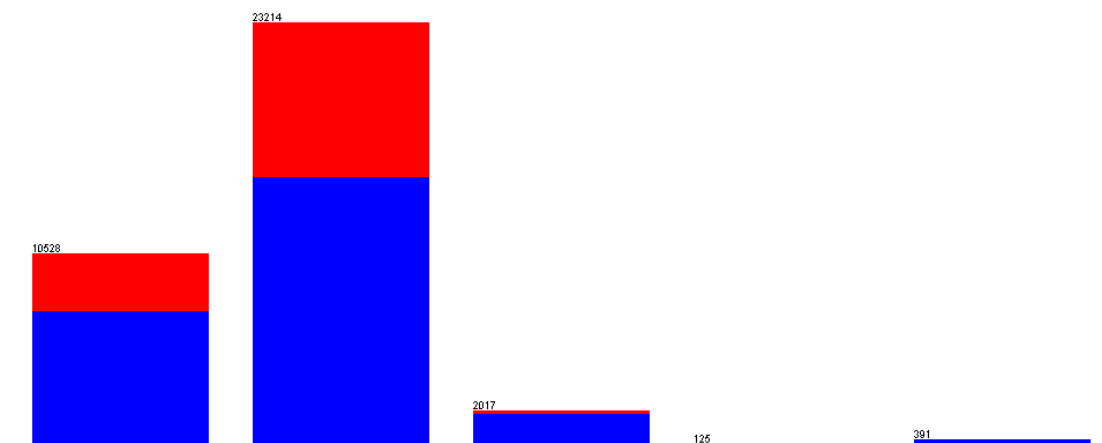
Το χαρακτηριστικό **market segment type** προσδιορίζει το τμήμα της αγοράς από την οποία προέρχεται η Κράτηση. Συγκεκριμένα:

- Online Κρατήσεις αφορούν τις κρατήσεις που γίνονται διαδικτυακά μέσω της ιστοσελίδας του ξενοδοχείου
- Offline αφορά τις απευθείας κρατήσεις που γίνονται τηλεφωνικά ή με επίσκεψη στο Ξενοδοχείο.
- Corporate λογίζονται οι Κρατήσεις που γίνονται μέσω εταιρειών
- Aviation είναι οι Κρατήσεις που γίνονται μέσω αεροπορικών εταιρειών
- Complementary είναι οι κρατήσεις χωρίς χρέωση ή συμπληρωματικές κρατήσεις

market_segment_type	Count
Offline	10528
Online	23214
Corporate	2017
Aviation	125
Complementary	391

Πίνακας 30 Τιμές τμήματος αγοράς

Ακολουθεί γραφική παράσταση από το Weka:



Εικόνα 17 Γραφική παράσταση WEKA τιμών τμήματος αγοράς

Missing Values	Most Frequent Value
0 (0%)	Online (23214)

Πίνακας 31 Missing Values / most Frequent Value τιμών τμήματος αγοράς

13. Προσδιορισμός πελάτη ως επαναλαμβανόμενου επισκέπτη ή μη

Το χαρακτηριστικό αυτό του δείγματος προσδιορίζει εάν ο πελάτης της κράτησης είναι επαναλαμβανόμενος ή όχι.

0: Μη επαναλαμβανόμενος επισκέπτης

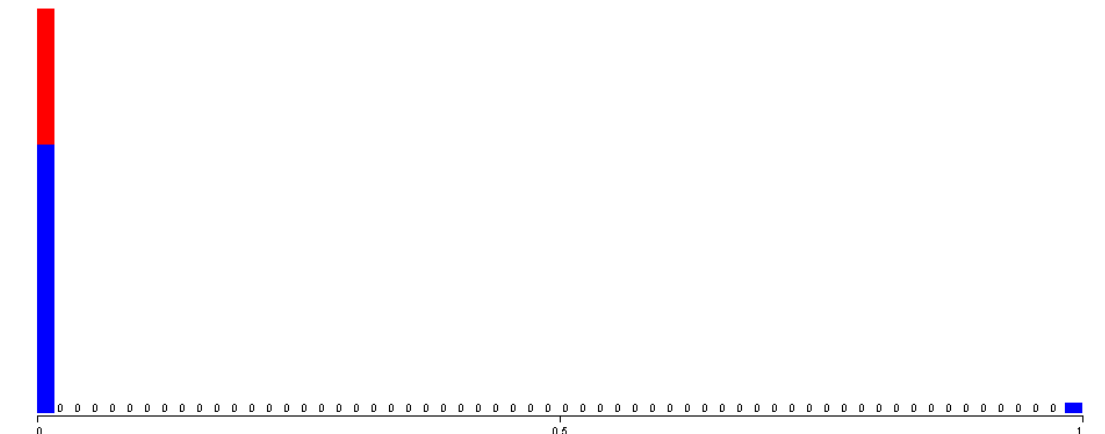
1: Επαναλαμβανόμενος Επισκέπτης

Η πλειοψηφία του παρόντος δείγματος είναι μεμονωμένοι επισκέπτες.

Repeated_guest	Value
MINIMUM	0 (No)
MAXIMUM	1 (Yes)
MEAN	0.026
STDDEV	0.158

Πίνακας 32 Τιμές επαναλαμβανόμενου επισκέπτη

Ακολουθεί γραφική παράσταση από το Weka:



Εικόνα 18 Γραφική Παράσταση τιμών επαναλαμβανόμενου επισκέπτη

Missing Values	Most Frequent Value
0 (0%)	[0, 0.017] (35345)

Πίνακας 33 Missing Values / most Frequent Value τιμών επαναλαμβανόμενου επισκέπτη

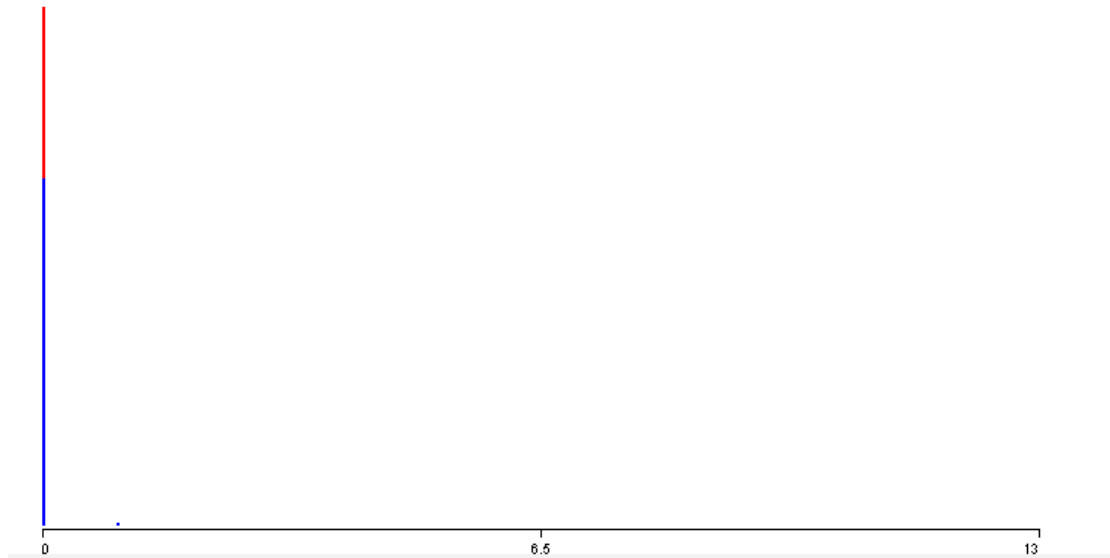
14. Αριθμός προηγούμενων κρατήσεων που ακυρώθηκαν από τον πελάτη πριν από την τρέχουσα κράτηση.

Το χαρακτηριστικό περί ακύρωσης προηγούμενων κρατήσεων λαμβάνει ελάχιστη τιμή 0 και μέγιστη το 13. Η πλειοψηφία των κρατήσεων στο παρόν δείγμα δεν έχει ακυρώσει προηγούμενη Κράτηση.

No_of_previous_cancellations	Value
MINIMUM	0
MAXIMUM	13
MEAN	0.023
STDDEV	0.368

Πίνακας 34 Αριθμός προηγούμενων κρατήσεων που ακυρώθηκαν από τον πελάτη πριν από την τρέχουσα κράτηση

Ακολουθεί γραφική παράσταση από το Weka:



Εικόνα 19 Γραφική Παράσταση WEKA αριθμού προηγούμενων κρατήσεων που ακυρώθηκαν από τον πελάτη πριν από την τρέχουσα κράτηση

Missing Values	Most Frequent Value
0 (0%)	[0, 0.039] (35937)

Πίνακας 35 Missing Values / most Frequent Value αριθμού προηγούμενων κρατήσεων που ακυρώθηκαν από τον πελάτη πριν από την τρέχουσα κράτηση

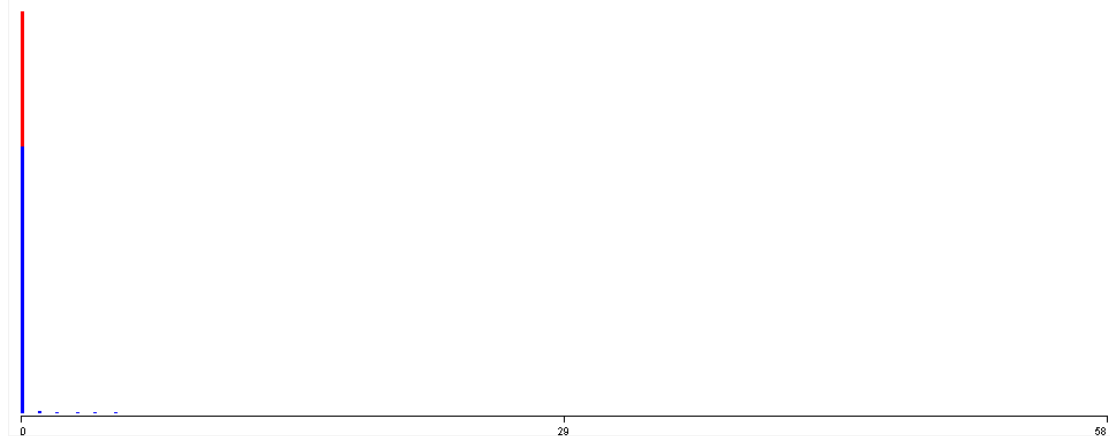
15. Αριθμός προηγούμενων κρατήσεων που δεν ακυρώθηκαν από τον πελάτη πριν από την τρέχουσα κράτηση.

Το χαρακτηριστικό λαμβάνει τιμές από 0 έως 58.

No_of_previous_bookingsnot cancelled	Value
MINIMUM	0
MAXIMUM	58
MEAN	0.153
STDDEV	1.754

Πίνακας 36 Αριθμός προηγούμενων κρατήσεων που δεν ακυρώθηκαν από τον πελάτη πριν από την τρέχουσα κράτηση

Ακολουθεί γραφική παράσταση από το Weka:



Εικόνα 20 Γραφική Παράσταση WEKA αριθμού προηγούμενων κρατήσεων που δεν ακυρώθηκαν από τον πελάτη πριν από την τρέχουσα κράτηση

Missing Values	Most Frequent Value
0 (0%)	[0, 0.185] (35463)

Πίνακας 37 Missing Values / most Frequent Value αριθμού προηγούμενων κρατήσεων που δεν ακυρώθηκαν από τον πελάτη πριν από την τρέχουσα κράτηση

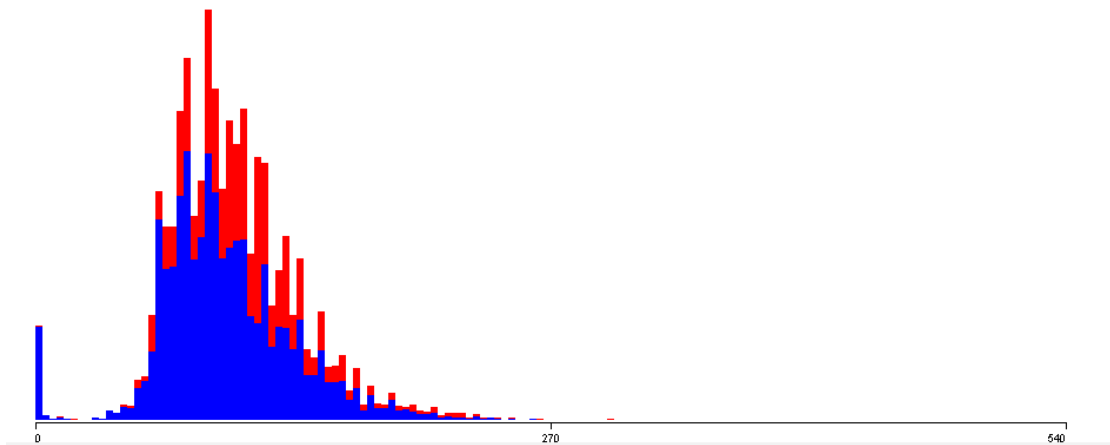
16. Μέση τιμή ανά ημέρα κράτησης

Το χαρακτηριστικό αυτό μας δίνει την πληροφορία της μέση τιμή του δωματίου υπολογισμένη σε €. Οι τιμές που λαμβάνει το χαρακτηριστικό είναι από 0 (για τις complementary κρατήσεις) έως 540€.

Avg_price_per_room	Value
MINIMUM	0
MAXIMUM	540
MEAN	103.424
STDDEV	35.089

Πίνακας 38 Μέση τιμή ανά ημέρα κράτησης

Ακολουθεί γραφική παράσταση από το Weka:



Εικόνα 21 Γραφική Παράσταση WEKA μέσης τιμής ανά ημέρα κράτησης

Missing Values	Most Frequent Value
0 (0%)	[88.767, 92.466] (2451)

Πίνακας 39 Missing Values / most Frequent Value μέσης τιμής ανά ημέρα κράτησης

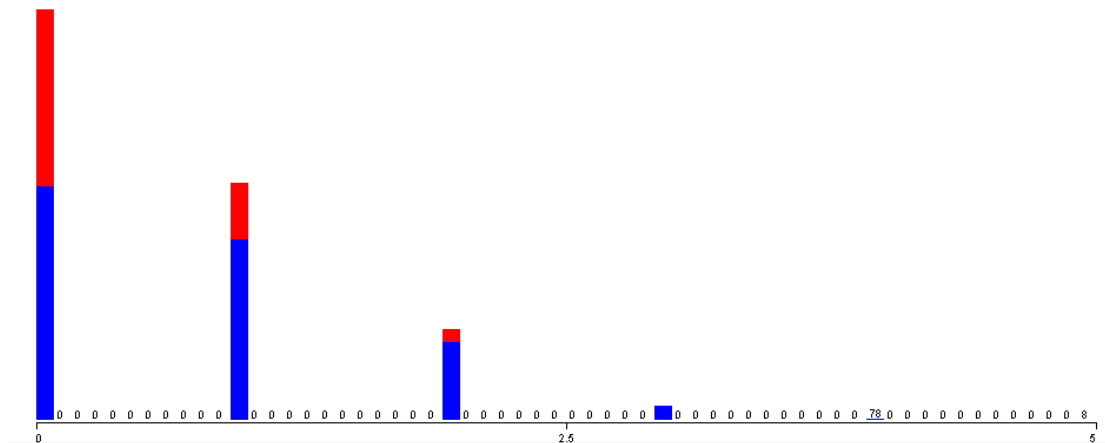
17. Συνολικός αριθμός ειδικών αιτημάτων που υποβλήθηκαν από τον πελάτη.

Το χαρακτηριστικό των ειδικών αιτημάτων που έχουν υποβληθεί από τον πελάτη κάθε Κράτησης λαμβάνει τιμές από 0 έως 5.

No_of_special_requests	Value
MINIMUM	0
MAXIMUM	5
MEAN	0.62
STDDEV	0.786

Πίνακας 40 Συνολικός αριθμός ειδικών αιτημάτων που υποβλήθηκαν από τον πελάτη

Ακολουθεί γραφική παράσταση από το Weka:



Εικόνα 22 Γραφική Παράσταση WEKA συνολικού αριθμού ειδικών αιτημάτων που υποβλήθηκαν από τον πελάτη

Missing Values	Most Frequent Value
0 (0%)	[0, 0.083] (19777)

Πίνακας 41 Missing Values / most Frequent Value συνολικού αριθμού ειδικών αιτημάτων που υποβλήθηκαν από τον πελάτη

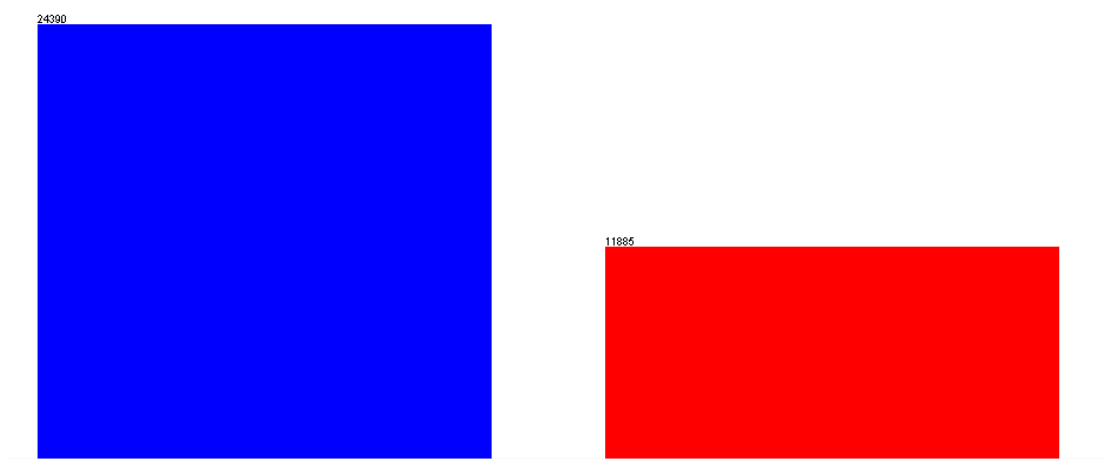
18. Ακύρωση ή όχι της Κράτησης

Έξοδος στο παρόν δείγμα είναι η ακύρωση ή όχι της κράτησης. Στην πλειοψηφία του δείγματος η Κράτηση δεν ακυρώθηκε.

Booking_status	count
Not_cancelled	24390
Cancelled	11885

Πίνακας 42 Τιμές Ακύρωσης ή όχι της Κράτησης

Ακολουθεί γραφική παράσταση από το Weka:



Εικόνα 23 Γραφική Παράσταση WEKA τιμών ακύρωσης ή όχι της Κράτησης

Missing Values	Most Frequent Value
0 (0%)	Not cancelled 24390

Πίνακας 43 Missing Values / most Frequent Value τιμών ακύρωσης ή όχι της Κράτησης

6. Προεπεξεργασία Δεδομένων

Αρχικά, από το δείγμα αφαιρείται το Booking ID το οποίο είναι μοναδικός κωδικός της κάθε κράτησης και δεν αποτελεί στοιχείο το οποίο είναι χρήσιμο στην ανάλυση. Επίσης λόγω μη ικανοποιητικής κατανομής των τιμών διαγράφουμε το χαρακτηριστικό **No_of_previous_bookings not cancelled** που αφορά στον αριθμό προηγούμενων κρατήσεων που δεν ακυρώθηκαν από τον πελάτη πριν από την τρέχουσα κράτηση καθώς και το **No_of_previous_cancellations** το οποίο και αφορά στον **αριθμό προηγούμενων κρατήσεων που ακυρώθηκαν από τον πελάτη πριν από την τρέχουσα κράτηση**. Οι τιμές που λαμβάνουν τα δύο αυτά χαρακτηριστικά δεν βοηθούν τους αλγόριθμους στην ταξινόμηση και συνεπώς δεν προσφέρουν κάποια αξία στην παρούσα ανάλυση.

Επιπροσθέτως, το έτος της άφιξης της κράτησης λαμβάνει τιμές τα 2017 και 2018 αλλά κρίνεται ότι δεν προσφέρει αξία σαν χαρακτηριστικό στην παρούσα ανάλυση καθώς η ανάπτυξη μοντέλων μηχανικής μάθησης για την διαχείριση κρατήσεων στις ξενοδοχειακές μονάδες απαιτείται να είναι διαχρονική και συνεπώς διαγράφεται από το τελικό δείγμα.

Στην συνέχεια, με το φίλτρο **numeric to binary** μετατρέπουμε τις τιμές από αριθμητικές σε δυαδικές στα εξής χαρακτηριστικά:

- **required_car_parking_space**
- **repeated_guest**



Εικόνα 24 Μετατροπή του χαρακτηριστικού *repeated guest*



Εικόνα 25 Μετατροπή του χαρακτηριστικού αιτήματος parking

Μετά την προεπεξεργασία του δείγματος και την διαγραφή ορισμένων χαρακτηριστικών **το τελικό δείγμα – full data set** αποτελείται από 14 χαρακτηριστικά και την Έξοδο και συνολικά από **36275 παραδείγματα**. Το full data set παρουσιάζεται κατωτέρω:

No.	Name
1	☐ no_of_adults
2	☐ no_of_children
3	☐ no_of_weekend_nights
4	☐ no_of_week_nights
5	☐ type_of_meal_plan
6	☐ required_car_parking_space_binarized
7	☐ room_type_reserved
8	☐ lead_time
9	☐ arrival_month
10	☐ arrival_date
11	☐ market_segment_type
12	☐ repeated_guest_binarized
13	☐ avg_price_per_room
14	☐ no_of_special_requests
15	☒ booking_status

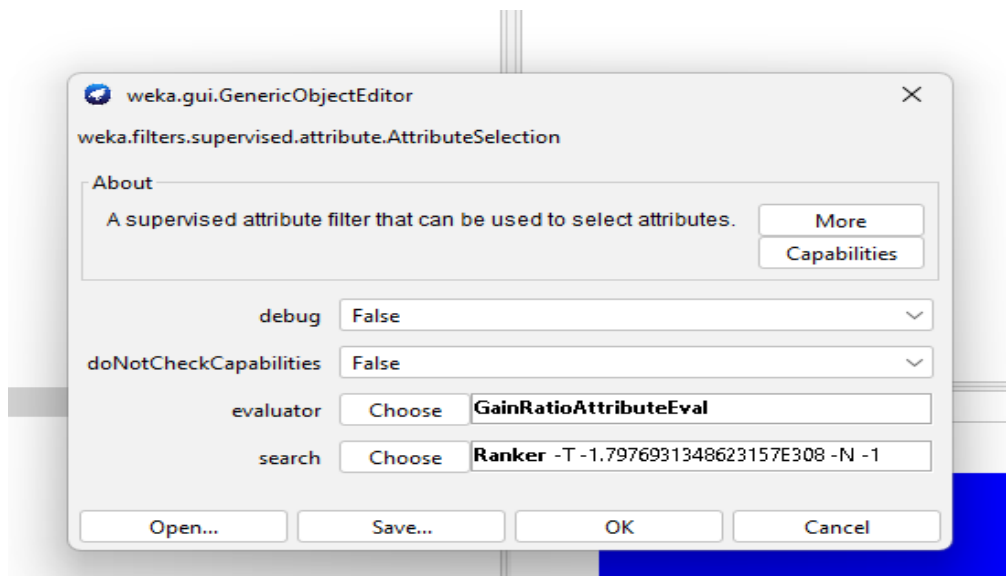
Εικόνα 26 Τελικό Δείγμα



Εικόνα 27 Έξοδος Πειραμάτων

Στην συνέχεια, στην προσπάθεια ανάδειξης του αποτελεσματικότερου μοντέλου εφαρμόστηκαν και φίλτρα επεξεργασίας του δείγματος για την υλοποίηση πληθώρας πειραμάτων. Αρχικά, θα εφαρμοστεί το φίλτρο **Attribute Selection** με τις εξής παραμέτρους:

- Evaluator: GainRatioAttributeEval
- Search: Ranker

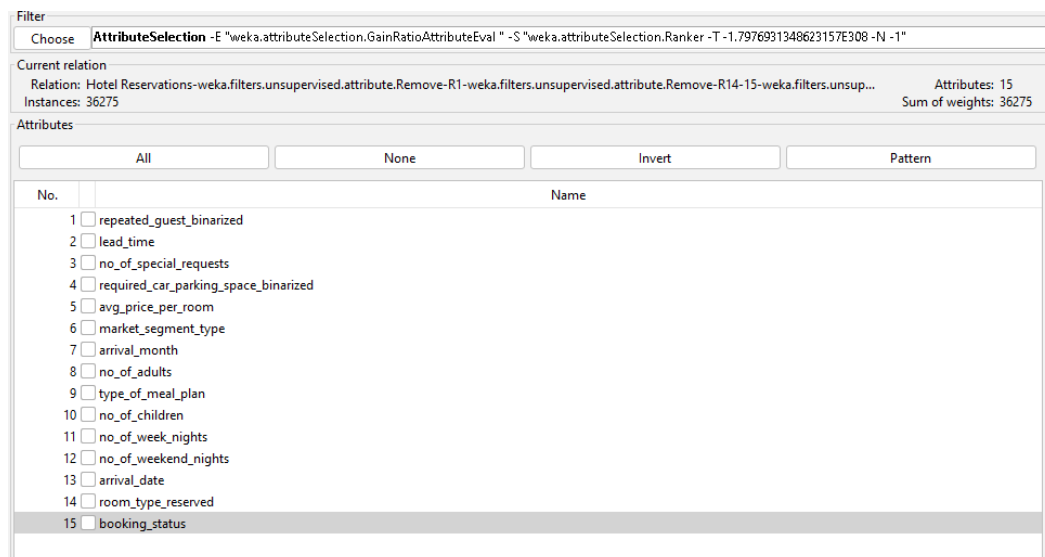


Εικόνα 28 Παράμετροι εφαρμογής φίλτρου Attribute Selection

Με την εφαρμογή του φίλτρου **Attribute Selection** και την ρύθμιση με τις παραμέτρους που παρουσιάζονται τα χαρακτηριστικά του δείγματος μπαίνουν σε κατάταξη αφού αξιολογούνται από τον αλγόριθμο.

Η κατάταξη που πραγματοποιείται στο WEKA με την εφαρμογή του φίλτρου **Attribute Selection** έχει ως εξής:

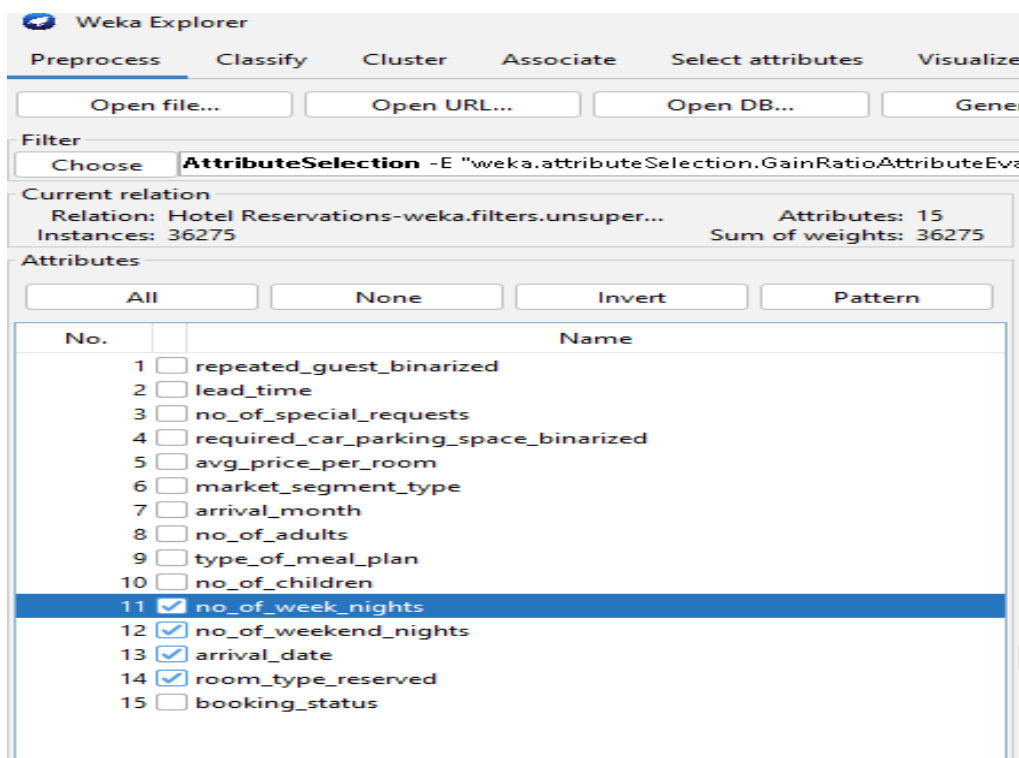
1. repeated_guest
2. lead_time
3. no_of_special_requests
4. required_car_parking_space
5. avg_price_per_room
6. market_segment_type
7. arrival_month
8. no_of_adults
9. type_of_meal_plan
10. no_of_children
11. no_of_week_nights
12. no_of_weekend_nights
13. arrival_date
14. room_type_reserved



Εικόνα 29 Κατάταξη Χαρακτηριστικών με την εφαρμογή του φίλτρου Attribute Selection

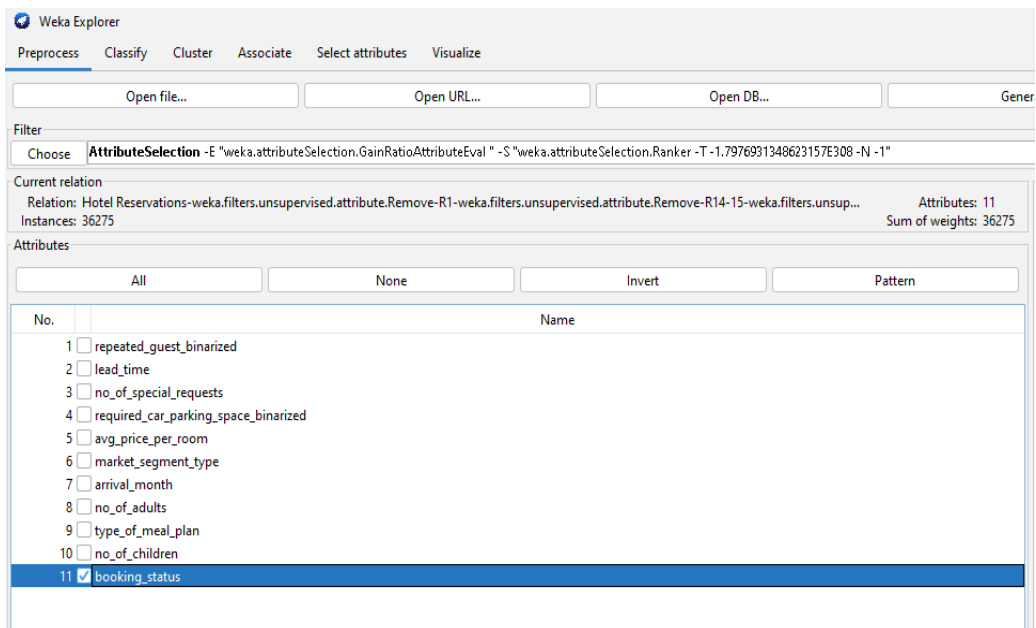
Στην συνέχεια, για την πραγματοποίηση πειραμάτων θα διαγραφούν τα 4 τελευταία χαρακτηριστικά σύμφωνα με την κατάταξη του φίλτρου **Attribute Selection**:

- no_of_week_nights
- no_of_weekend_nights
- arrival_date
- room_type_reserved



Εικόνα 30 Διαγραφή 4 χαρακτηριστικών

Το δείγμα για την πραγματοποίηση μέρους των πειραμάτων θα αποτελείται από 10 χαρακτηριστικά.

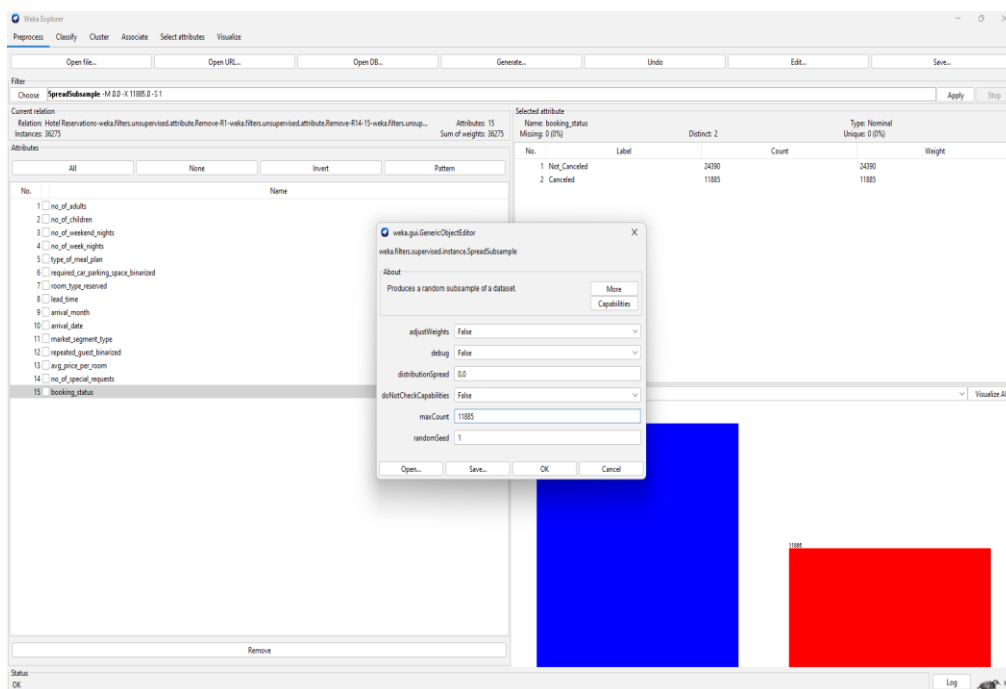


Εικόνα 31 Δείγμα για μέρος πειραμάτων με 10 χαρακτηριστικά

Επίσης, για την εφαρμογή μέρους των πειραμάτων χρησιμοποιήθηκε το φίλτρο **SpreadSubSample** με το οποίο οι τιμές της εξόδου γίνονται ισορροπημένες. Συγκεκριμένα, το συνολικό δείγμα αποτελείται από 36.275 Παραδείγματα τα οποία στην Έξοδο κατανέμονται ως εξής:

CANCELLED	NOT CANCELLED
11885 instances	24390 instances

Πίνακας 44 Τιμές εξόδου

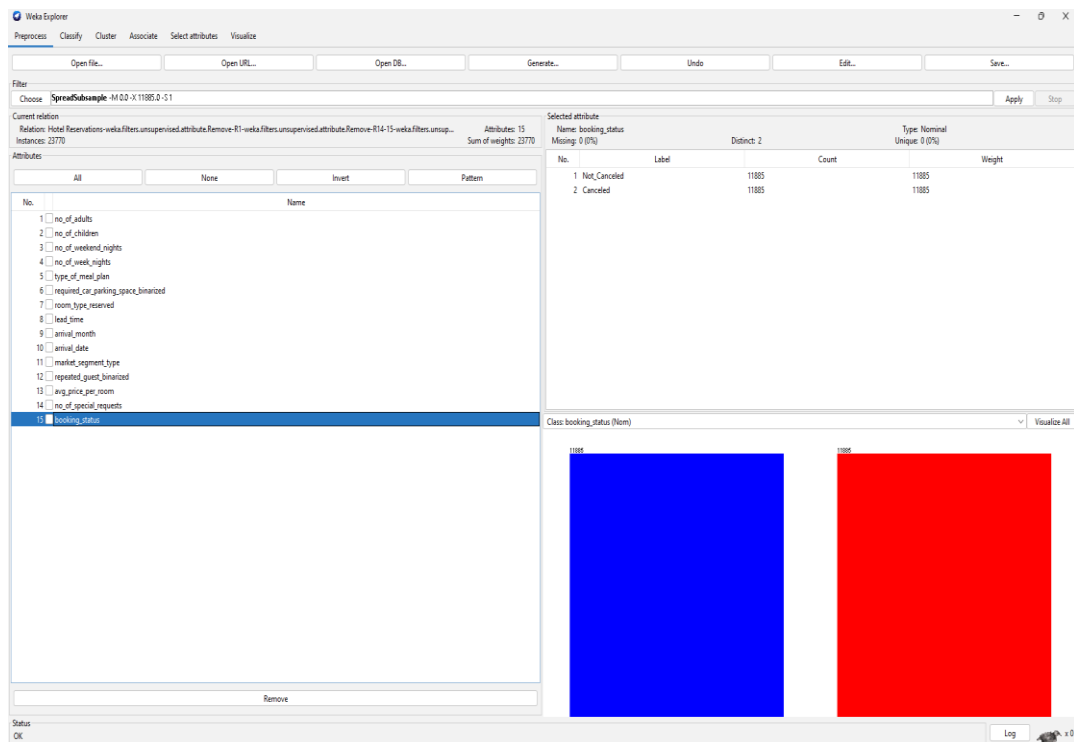


Εικόνα 32 Εφαρμογή φίλτρου Spread Sub Sample

Με την εξισορρόπηση του δείγματος που επιτυγχάνεται με το φίλτρο **SpreadSubSample** το δείγμα για μέρος των πειραμάτων αποτελείται συνολικά από 23770 Παραδείγματα και οι τιμές της εξόδου και για τις δύο κατηγορίες είναι ο αριθμός

11885, καθώς αυτός είναι ο αριθμός των παραδειγμάτων της Κατηγορίας με τα λιγότερα δεδομένα.

Εξισορροπημένο δείγμα:



Εικόνα 33 Εξισορροπημένο δείγμα

7. Μεθοδολογία

Για την εφαρμογή των Πειραμάτων της παρούσας μελέτης χρησιμοποιήθηκε το **Waikato Environment for Knowledge Analysis (Weka)** το οποίο διαθέτει ελεύθερο λογισμικό μηχανικής μάθησης και ανάλυσης δεδομένων γραμμένο σε Java, που αναπτύχθηκε στο Πανεπιστήμιο του Waikato της Νέας Ζηλανδίας.

Με την βοήθεια των Ταξινομητών Αλγόριθμων πλησιέστερου γείτονα IBK, του δέντρου απόφασης J48 και του Random Forest θα επιχειρηθεί με την παρούσα εργασία να αναπτυχθούν μοντέλα μηχανικής μάθησης **με απώτερο στόχο την ανάπτυξη του αποτελεσματικότερου μοντέλου προκειμένου να μπορεί να προβλεφθεί από τις ξενοδοχειακές μονάδες η ακύρωση ή όχι της κράτησης με όσο το δυνατόν μεγαλύτερη ακρίβεια και να παρουσιαστούν τα χαρακτηριστικά που επηρεάζουν την ακύρωση κράτησης.** Για την παρούσα μελέτη όλα τα πειράματα που πραγματοποιήθηκαν στο WEKA έγιναν με την επιλογή του τεστ Cross – Validation Folds 10, το οποίο αποτελεί μια λειτουργία τεστ με διασταυρούμενη επικύρωση 10 επαναλήψεων, όπου, για κάθε επανάληψη, ο αλγόριθμος 10-fold αξιολογεί το μοντέλο χρησιμοποιώντας 10% δεδομένα δοκιμής και 90% δεδομένα εκπαίδευσης.[20]

Στην προσπάθεια ανάδειξης του ακριβέστερου μοντέλου θα παρουσιαστούν **τέσσερις (4) πίνακες** στους οποίους περιλαμβάνονται τα αποτελέσματα 12 Πειραμάτων των αλγορίθμων με τις διάφορες παραμέτρους. Ο πρώτος πίνακας θα παρουσιάζει αποτελέσματα ύστερα από ανάλυση όλου του δείγματος που επιλέχθηκε για την παρούσα εργασία και περιλαμβάνει 14 χαρακτηριστικά. Στη συνέχεια, ο δεύτερος πίνακας θα παρουσιάζει τα αποτελέσματα ύστερα όμως από την εφαρμογή του φίλτρου Attribute Selection ενώ ο τρίτος με Attribute Selection και αφαίρεση των 4 τελευταίων χαρακτηριστικών σύμφωνα με την κατάταξη. Τέλος, στον τελευταίο πίνακα παρουσιάζονται τα αποτελέσματα ύστερα από την εφαρμογή του φίλτρου Spread Subsample.

Στην παρούσα μελέτη παρουσιάζονται τα αποτελέσματα που πραγματοποιήθηκαν από συνολικά 48 Πειράματα.

8. Αποτελέσματα Πειραμάτων

Στην παρούσα ενότητα θα παρουσιαστούν και θα αναλυθούν τα αποτελέσματα των πειραμάτων σε 4 πίνακες που θα περιέχουν τον προσδιορισμό του Data set, το είδος του αλγόριθμου, το είδος του test, τις διάφορες παραμέτρους του πειράματος, το TP Rate, το Precision και το Accuracy.

ΠΙΝΑΚΑΣ Ι Αποτελέσματα πειραμάτων στο Weka με Full Data Set								
Data Set	Αλγόριθμος Πειράματος/ Ταξινομητής	Test	Παράμετροι Πειράματος	TP RATE Not Cancelled	TP RATE Cancelled	PRECISION Not Cancelled	PRECISION Cancelled	Accuracy
Full data Set -36275 Instances- 14 attributes	IBK	Cross - Validation folds 10	K =1	0,893	0,775	0,890	0,779	85,4252%
Full data Set -36275 Instances- 14 attributes	IBK	Cross - Validation folds 10	K =3	0,900	0,743	0,878	0,762	84,8105%
Full data Set -36275 Instances- 14 attributes	IBK	Cross - Validation folds 10	K =5	0,903	0,735	0,875	0,788	84,827%
Full data Set -36275 Instances- 14 attributes	IBK	Cross - Validation folds 10	K =7	0,904	0,726	0,871	0,787	84,5762%
Full data Set -36275 Instances- 14 attributes	IBK	Cross - Validation folds 10	K =11	0,904	0,710	0,865	0,783	84,0469%
Full data Set -36275 Instances- 14 attributes	IBK	Cross - Validation folds 10	K =13	0,903	0,707	0,863	0,781	83,887%

Full data Set -36275 Instances- 14 attributes	J48	Cross - Validation folds 10	Unpruned	0,912	0,807	0,906	0,816	87,7243%
Full data Set -36275 Instances- 14 attributes	J48	Cross - Validation folds 10	Pruned	0,927	0,796	0,903	0,842	88,4218%
Full data Set -36275 Instances- 14 attributes	J48	Cross - Validation folds 10	Pruned, binary splits, minNumObj=20	0,928	0,758	0,887	0,836	87,2198%
Full data Set -36275 Instances- 14 attributes	J48	Cross - Validation folds 10	Pruned, binary splits, minNumObj=20 reduced error pruning	0,926	0,746	0,882	0,831	86,6988%
Full data Set -36275 Instances- 14 attributes	J48	Cross - Validation folds 10	Pruned, minNumObj=100 reduced error pruning	0,918	0,724	0,872	0,812	85,4666%
Full data Set -36275 Instances- 14 attributes	Random Forest	Cross - Validation folds 10	P 100 -I 100 -num- slots 1 -K 0 -M 1.0 - V 0.001 -S 1	0,949	0,812	0,912	0,885	90,3928%

Πίνακας 45 Αποτελέσματα πειραμάτων στο Weka με Full Data Set

Συμπεράσματα – σχόλια Πίνακα Ι

Από τα πειράματα που έγιναν με full data set προκύπτει ότι όλοι οι αλγόριθμοι έχουν καλά αποτελέσματα στην ακρίβεια με ποσοστά που κυμαίνονται από 83,89% έως 90,39%. Ο πλησιέστερος γείτονας IBK έχει ικανοποιητική απόδοση καθώς τα ποσοστά ακρίβειας κυμαίνονται από 83,89% έως 85,43% με το καλύτερα αποτελέσματα να τα έχουμε στο πείραμα όπου $K=1$. Και ενώ ο IBK έχει μια καλή απόδοση δεν είναι τόσο αποτελεσματικός όσο τα δέντρα απόφασης J48 όπου τα ποσοστά ακρίβειας κυμαίνονται από 85,47% έως 88,42%. Αξίζει να σημειωθεί μάλιστα ότι το δέντρο με την παράμετρο pruned να αναδεικνύει τα καλύτερα αποτελέσματα. **Ωστόσο, ο Random Forest είναι με διαφορά ο πιο αποτελεσματικός αλγόριθμος καθώς έχει εξαιρετική ακρίβεια της τάξεως του 90,39%.**

ΠΙΝΑΚΑΣ II Αποτελέσματα πειραμάτων στο Weka με εφαρμογή φίλτρου Attribute Selection								
Data Set	Αλγόριθμος Πειράματος/ Ταξινομητής	Test	Παράμετροι Πειράματος	TP RATE Not Cancelled	TP RATE Cancelled	PRECISION Not Cancelled	PRECISION Cancelled	Accuracy
Full data Set -36275 Instances-14 attributes	IBK	Cross - Validation folds 10	K =1	0,893	0,775	0,890	0,779	85,4252%
Full data Set -36275 Instances-14 attributes	IBK	Cross - Validation folds 10	K =3	0,900	0,743	0,878	0,783	84,8105%
Full data Set -36275 Instances-14 attributes	IBK	Cross - Validation folds 10	K =5	0,903	0,735	0,875	0,788	84,827%
Full data Set -36275 Instances-14 attributes	IBK	Cross - Validation folds 10	K =7	0,904	0,726	0,871	0,787	84,5762%
Full data Set -36275 Instances-14 attributes	IBK	Cross - Validation folds 10	K =11	0,904	0,710	0,865	0,783	84,0469%
Full data Set -36275 Instances-14 attributes	IBK	Cross - Validation folds 10	K =13	0,903	0,707	0,863	0,781	83,887%
Full data Set -36275 Instances-14 attributes	J48	Cross - Validation folds 10	Unpruned	0,912	0,807	0,906	0,816	87,7243%
Full data Set -36275 Instances-	J48	Cross - Validation	Pruned	0,927	0,796	0,903	0,842	88,4328%

14 attributes		n folds 10						
Full data Set -36275 Instances- 14 attributes	J48	Cross - Validatio n folds 10	Pruned, binary splits, minNumObj=20	0,928	0,758	0,887	0,836	87,2198%
Full data Set -36275 Instances- 14 attributes	J48	Cross - Validatio n folds 10	Pruned, binary splits, minNumObj=20 reduced error pruning	0,924	0,746	0,882	0,828	86,6051%
Full data Set -36275 Instances- 14 attributes	J48	Cross - Validatio n folds 10	Pruned, minNumObj=10 0 reduced error pruning	0,917	0,706	0,865	0,806	84,7939%
Full data Set -36275 Instances- 14 attributes	Random Forest	Cross - Validatio n folds 10	P 100 -I 100 - num-slots 1 -K 0 -M 1.0 -V 0.001 - S 1	0,948	0,812	0,912	0,884	90,3598%

Πίνακας 46 Αποτελέσματα πειραμάτων στο Weka με εφαρμογή φίλτρου Attribute Selection

Συμπεράσματα – σχόλια Πίνακα II

Από τα πειράματα που έγιναν με full data set αλλά με την εφαρμογή του φίλτρου Attribute Selection προκύπτει ότι όλοι οι αλγόριθμοι έχουν καλά αποτελέσματα στην ακρίβεια με ποσοστά που κυμαίνονται από 83,89% έως 90,36% και δεν παρουσιάστηκαν σημαντικές διαφορές. Ο πλησιέστερος γείτονας IBK έχει ικανοποιητική απόδοση καθώς τα ποσοστά ακρίβειας κυμαίνονται από 83,89% έως 85,43% με το καλύτερα αποτελέσματα να τα έχουμε και εδώ στο πείραμα όπου $K=1$. Και ενώ ο IBK έχει μια καλή απόδοση τα δέντρα απόφασης J48 και εδώ αποδίδουν καλύτερα όπου τα ποσοστά ακρίβειας κυμαίνονται από 84,80% έως 88,43%. Παρατηρήθηκαν διαφορές στα αποτελέσματα των δέντρων απόφασης μικρής ωστόσο σημασίας. Αξίζει να σημειωθεί μάλιστα ότι το δέντρο με την παράμετρο pruned να αναδεικνύει και σε αυτό το σετ πειραμάτων τα καλύτερα αποτελέσματα. **Αποδεικνύεται και εδώ ότι ο Random Forest είναι με διαφορά ο πιο αποτελεσματικός αλγόριθμος καθώς έχει εξαιρετική ακρίβεια της τάξεως του 90,36% αλλά η απόδοσή του παρουσίασε μια μικρή πτώση.**

ΠΙΝΑΚΑΣ ΙΙΙ Αποτελέσματα πειραμάτων στο Weka με εφαρμογή φίλτρου Attribute Selection (10 attributes)								
Data Set	Αλγόριθμος Πειράματος/ Ταξινομητής	Test	Παράμετροι Πειράματος	TP RATE Not Cancelled	TP RATE Cancelled	PRECISION Not Cancelled	PRECISION Cancelled	Accuracy
Full data Set -36275 Instances-10 attributes	IBK	Cross - Validation folds 10	K =1	0,902	0,787	0,897	0,787	86,4369%
Full data Set -36275 Instances-10 attributes	IBK	Cross - Validation folds 10	K =3	0,910	0,764	0,888	0,806	86,244%
Full data Set -36275 Instances-10 attributes	IBK	Cross - Validation folds 10	K =5	0,909	0,759	0,886	0,803	86,0096%
Full data Set -36275 Instances-10 attributes	IBK	Cross - Validation folds 10	K =7	0,911	0,754	0,884	0,806	86,0041%
Full data Set -36275 Instances-10 attributes	IBK	Cross - Validation folds 10	K =11	0,913	0,747	0,881	0,808	85,8911%
Full data Set -36275 Instances-10 attributes	IBK	Cross - Validation folds 10	K =13	0,913	0,746	0,881	0,807	85,8249%
Full data Set -36275 Instances-10 attributes	J48	Cross - Validation folds 10	Unpruned	0,926	0,780	0,896	0,836	87,785%
Full data Set -36275 Instances-	J48	Cross - Validation	Pruned	0,932	0,773	0,894	0,846	87,9669%

10 attributes		n folds 10						
Full data Set -36275 Instances- 10 attributes	J48	Cross - Validatio n folds 10	Pruned, binary splits, minNumObj=20	0,927	0,748	0,883	0,833	86,8422%
Full data Set -36275 Instances- 10 attributes	J48	Cross - Validatio n folds 10	Pruned, binary splits, minNumObj=20 reduced error pruning	0,920	0,745	0,881	0,819	86,255%
Full data Set -36275 Instances- 10 attributes	J48	Cross - Validatio n folds 10	Pruned, minNumObj=10 0 reduced error pruning	0,911	0,693	0,859	0,791	83,9421%
Full data Set -36275 Instances- 10 attributes	Random Forest	Cross - Validatio n folds 10	P 100 -I 100 - num-slots 1 -K 0 -M 1.0 -V 0.001 - S 1	0,934	0,802	0,906	0,855	89,0586%

Πίνακας 47 Αποτελέσματα πειραμάτων στο Weka με εφαρμογή φίλτρου Attribute Selection (10 attributes)

Συμπεράσματα – σχόλια Πίνακα ΙΙΙ

Από τα πειράματα που έγιναν με full data set αλλά με την εφαρμογή του φίλτρου Attribute Selection και την αφαίρεση των 4 χαρακτηριστικών των χαμηλότερων στην κατάταξη. Από τα αποτελέσματα προκύπτει το εξής παράδοξο τα αποτελέσματα του IBK βελτιώθηκαν ελαφρώς και κυμαίνονται μεταξύ των ποσοστών 85,82% έως 86,44% με τα καλύτερα αποτελέσματα να τα έχουμε και εδώ στο πείραμα όπου $K=1$. Γενικότερα, προκύπτει ότι τα αποτελέσματα όλων των αλγόριθμων ενώ είναι καλά είναι χαμηλότερα σε σχέση με τους δύο πρώτους πίνακες στην ακρίβεια με ποσοστά που κυμαίνονται από 83,94% έως 89,06%. Από τον πίνακα αξιολόγησης ο IBK έχει βελτιωμένη απόδοση τα δέντρα απόφασης J48 παρουσιάζουν μικρή πτώση στην ακρίβεια και τα ποσοστά ακρίβειας κυμαίνονται από 83,94% έως 87,97%. Αξίζει να σημειωθεί μάλιστα ότι το δέντρο με την παράμετρο pruned να παρουσιάζει και σε αυτό το σετ πειραμάτων τα καλύτερα αποτελέσματα. **Ο Random Forest ενώ σε αυτό το σετ πειραμάτων έχει πτωτική πορεία στα αποτελέσματα και πάλι αναδεικνύεται ως ο πιο αποτελεσματικός αλγόριθμος καθώς έχει εξαιρετική ακρίβεια της τάξεως του 89,06%.**

ΠΙΝΑΚΑΣ IV Αποτελέσματα πειραμάτων στο Weka με εφαρμογή φίλτρου SpreadSubSample								
Data Set	Αλγόριθμος Πειράματος/ Ταξινομητής	Test	Παράμετροι Πειράματος	TP RATE Not Cancelled	TP RATE Cancelled	PRECISION Not Cancelled	PRECISION Cancelled	Accuracy
23770 Instances-14 attributes	IBK	Cross - Validation n folds 10	K =1	0,821	0,832	0,830	0,823	82,6167%
23770 Instances-14 attributes	IBK	Cross - Validation n folds 10	K =3	0,814	0,823	0,822	0,816	81,8805%
23770 Instances-14 attributes	IBK	Cross - Validation n folds 10	K =5	0,810	0,828	0,825	0,813	81,8973%
23770 Instances-14 attributes	IBK	Cross - Validation n folds 10	K =7	0,805	0,827	0,823	0,810	81,6113%
23770 Instances-14 attributes	IBK	Cross - Validation n folds 10	K =11	0,797	0,827	0,822	0,803	81,1737%
23770 Instances-14 attributes	IBK	Cross - Validation n folds 10	K =13	0,795	0,826	0,821	0,801	81,0475%
23770 Instances-14 attributes	J48	Cross - Validation n folds 10	Unpruned	0,852	0,858	0,857	0,853	85,4859%
23770 Instances-14 attributes	J48	Cross - Validation n folds 10	Pruned	0,856	0,863	0,862	0,857	85,9529%

		n folds 10						
23770 Instances- 14 attributes	J48	Cross - Validatio n folds 10	Pruned, binary splits, minNumObj=20	0,855	0,847	0,848	0,854	85,0989%
23770 Instances- 14 attributes	J48	Cross - Validatio n folds 10	Pruned, binary splits, minNumObj=20 reduced error pruning	0,859	0,827	0,832	0,854	84,2575%
23770 Instances- 14 attributes	J48	Cross - Validatio n folds 10	Pruned, minNumObj=10 0 reduced error pruning	0,839	0,788	0,798	0,830	81,342%
23770 Instances- 14 attributes	Random Forest	Cross - Validatio n folds 10	P 100 -I 100 - num-slots 1 -K 0 -M 1.0 -V 0.001 - S 1	0,898	0,872	0,876	0,895	88,5065%

Πίνακας 48 Αποτελέσματα πειραμάτων στο Weka με εφαρμογή φίλτρου SpreadSubSample

Συμπεράσματα – σχόλια Πίνακα IV

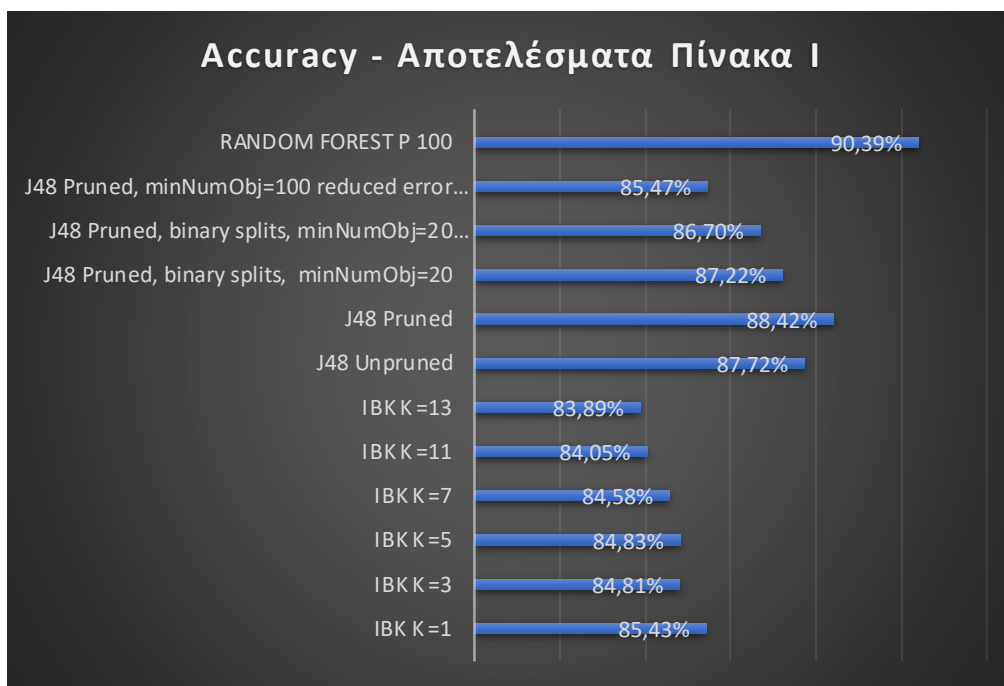
Από τα πειράματα που έγιναν με την εφαρμογή του φίλτρου SpreadSubSample επιτυγχάνοντας την εξισορρόπηση του δείγματος κρατώντας 23.770 παραδείγματα τα αποτελέσματα όλων των αλγορίθμων παρουσιάζουν σημαντική πτώση. Του IBK κυμαίνονται από 81,05% έως 82,62% με το καλύτερα αποτελέσματα να τα έχουμε και εδώ στο πείραμα όπου $K=1$. Τα αποτελέσματα των δέντρων απόφασης J48 κυμαίνονται πτωτικά από 81,34% έως 85,96% και μάλιστα αξίζει να σημειωθεί ότι το δέντρο με την παράμετρο pruned να παρουσιάζει και σε αυτό το σετ πειραμάτων τα καλύτερα αποτελέσματα μεταξύ των δέντρων. **Τέλος, ο Random Forest ενώ έχει πτωτική πορεία στα αποτελέσματα αυτού του σετ πειραμάτων και πάλι αναδεικνύεται ως ο πιο αποτελεσματικός αλγόριθμος με ακρίβεια της τάξεως του 88,51%.**

9. Συζήτηση – Συμπεράσματα Έρευνας

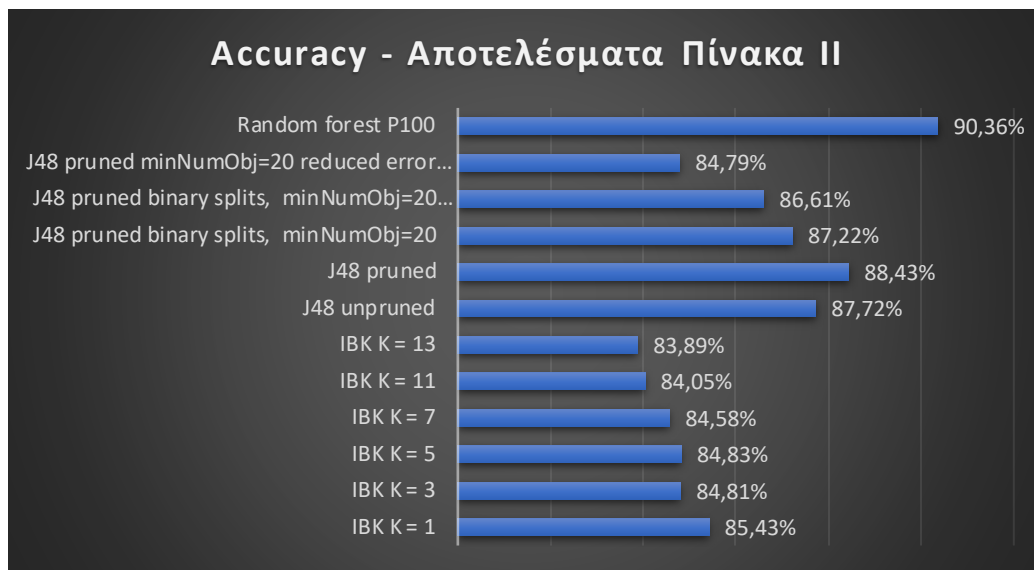
Στην παρούσα έρευνα πραγματοποιήθηκαν συνολικά 48 πειράματα με full data set χωρίς φίλτρα, με full data set και την εφαρμογή του φίλτρου attribute selection, με full data set με την εφαρμογή του φίλτρου attribute selection και την διαγραφή τεσσάρων χαρακτηριστικών αλλά και με την εφαρμογή του φίλτρου SpreadSubSample. Η επιλογή του **WEKA** ως εργαλείου για την ανάπτυξη μοντέλων μηχανικής μάθησης αποδείχθηκε ιδιαίτερα χρήσιμη, καθώς προσφέρει εύχρηστα εργαλεία ανάλυσης μεγάλων δεδομένων για την εκπαίδευση αλγορίθμων. Αξιοποιώντας τους αλγορίθμους IBK, J48, και Random Forest, τα πειράματα επικεντρώθηκαν στην ανάπτυξη μοντέλων που προβλέπουν με υψηλή ακρίβεια την πιθανότητα ακύρωσης κρατήσεων. Η μέθοδος Cross-Validation Folds 10 που επιλέχθηκε για την παρούσα μελέτη προσέφερε αξιόπιστη αξιολόγηση της απόδοσης των μοντέλων.

Από τα αποτελέσματα των πειραμάτων συμπεραίνουμε ότι με full data set δίχως εφαρμογή φίλτρων έχουμε τα βέλτιστα αποτελέσματα γενικότερα. Το φίλτρο attribute selection, η αφαίρεση χαρακτηριστικών και το spread sub sample δεν βελτίωσαν τα αποτελέσματα αλλά αντίθετα ανέδειξαν την σημασία του full data set.

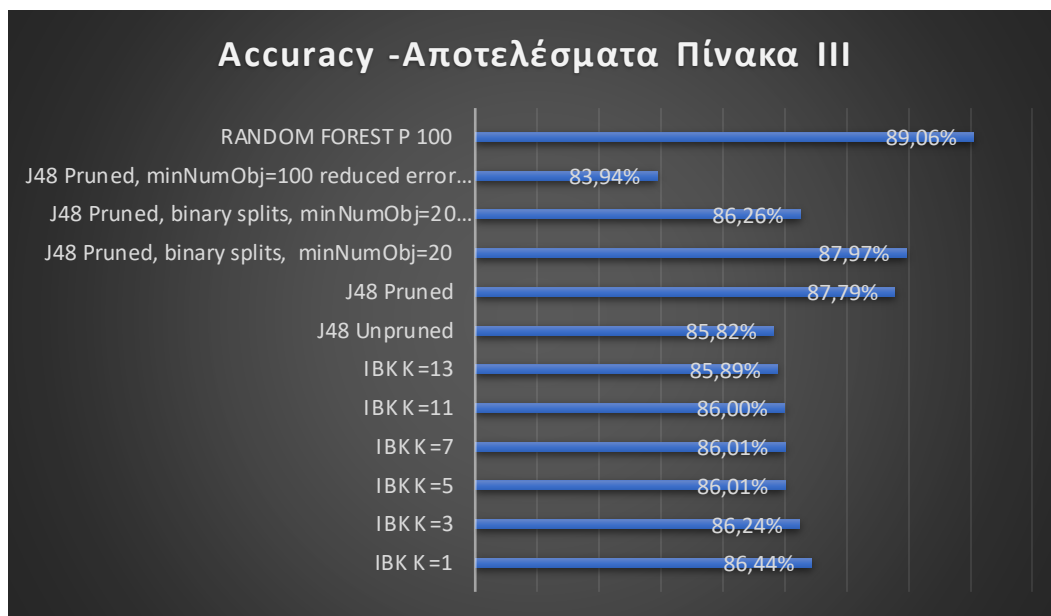
Ακολουθούν όλοι οι συγκριτικοί πίνακες των αποτελεσμάτων:



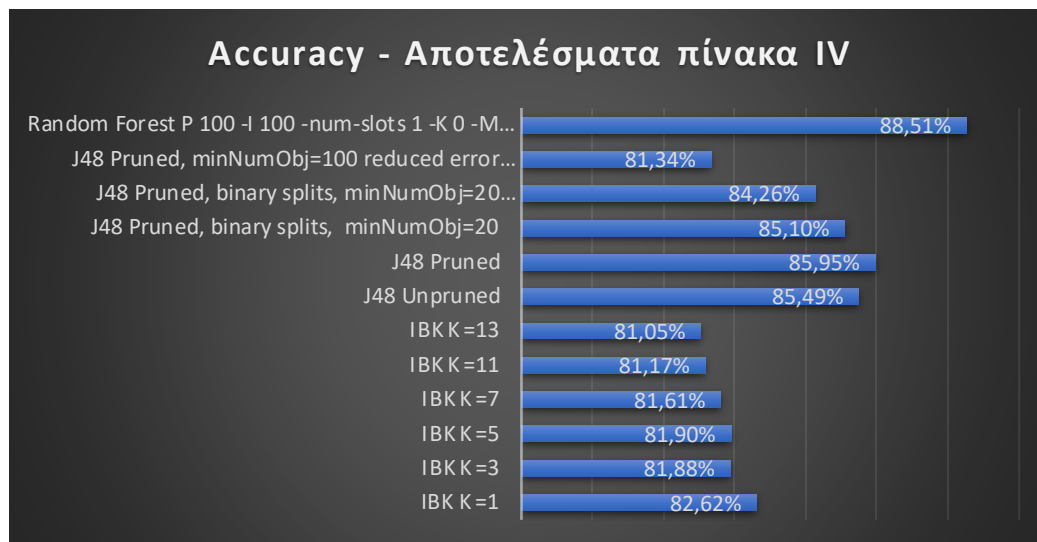
Πίνακας 49 Γράφημα σύγκρισης αλγορίθμων πειραμάτων Πίνακα I με full data set



Πίνακας 50 Γράφημα σύγκρισης αλγορίθμων πειραμάτων Πίνακα II με full data set και attribute selection



Πίνακας 51 Γράφημα σύγκρισης αλγορίθμων πειραμάτων Πίνακα III με attribute selection και συνολικά 10 χαρακτηριστικά



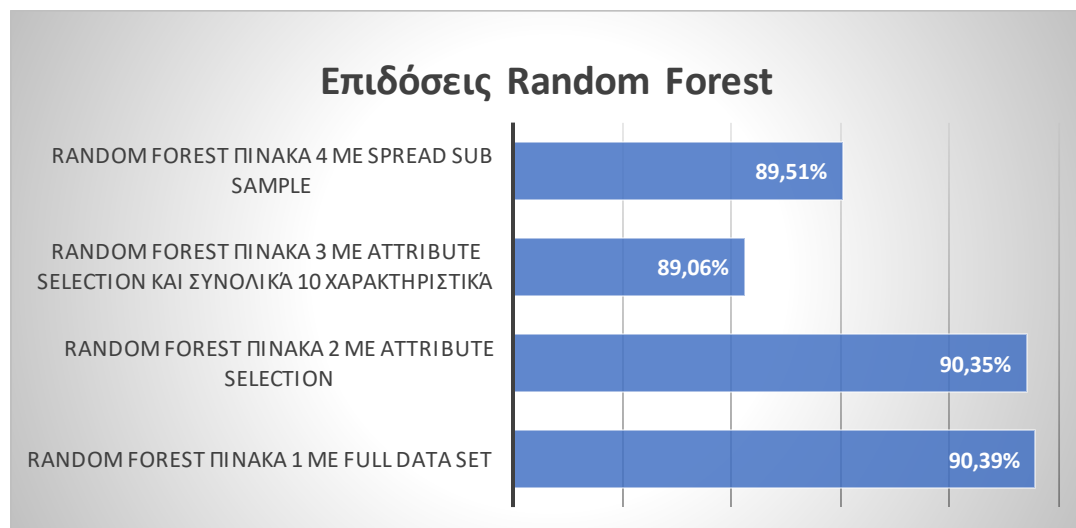
Πίνακας 52 Γράφημα σύγκρισης αλγορίθμων πειραμάτων Πίνακα IV με SpreadSubSample

Συνολικά, όλοι οι αλγόριθμοι παρουσίασαν ικανοποιητικά αποτελέσματα με ακρίβεια που κυμάνθηκε από 81,07% έως 90,39%.

Ο **IBK**, αν και είχε γενικά μια καλή συνολική απόδοση, υπολειπόταν έναντι των άλλων αλγορίθμων και οι καλύτερες επιδόσεις του ήταν όπου $K=1$.

Τα δέντρα απόφασης **J48** παρουσιάζουν ελαφρώς καλύτερη απόδοση, με την παράμετρο pruned να δίνει τα βέλτιστα αποτελέσματα.

Αξίζει να αναφερθεί ότι ο Random Forest αναδείχθηκε ως ο πιο αποτελεσματικός αλγόριθμος της παρούσας μελέτης με ακρίβεια που κυμάνθηκε από 89,51% έως 90,39%, υποδεικνύοντας την υψηλή του ικανότητα στη ανάλυση σύνθετων προβλημάτων.



Πίνακας 53 Γράφημα σύγκρισης επιδόσεων Random Forest

Η ανάλυση δείχνει ότι ο Random Forest υπερτερεί ως προς την ακρίβεια και την ανθεκτικότητα, ακόμα και όταν εφαρμόζονται φίλτρα στο δείγμα. Από την παρούσα μελέτη προκύπτει ότι το αποτελεσματικότερο μοντέλο είναι εκείνο στο οποίο χρησιμοποιήθηκε το full data set.

Αποτελέσματα αλγορίθμου J48 Έρευνας - Ανάδειξη χαρακτηριστικών με βαρύνουσα σημασία.

Στην παρούσα μελέτη χρησιμοποιήθηκε όπως είδαμε και ο αλγόριθμος J48 – δέντρο απόφασης ο οποίος μπορεί να μας καθοδηγήσει σχετικά με τα χαρακτηριστικά εκείνα τα οποία έχουν βαρύνουσα σημασία στην πρόβλεψη. Ακολουθούν πειράματα J48 σε μορφή text αλλά και δέντρου.

```
lead_time <= 151.0
| no_of_special_requests <= 0.0
| | market_segment_type = Online
| | | arrival_month <= 1.0: Not_Canceled (182.0/12.0)
| | | arrival_month > 1.0
| | | | lead_time <= 9.0: Not_Canceled (1037.0/236.0)
| | | | lead_time > 9.0
| | | | | arrival_month <= 11.0
| | | | | | avg_price_per_room <= 111.76
| | | | | | type_of_meal_plan = Meal Plan 1
| | | | | | | lead_time <= 44.0: Not_Canceled (360.0/155.0)
| | | | | | | lead_time > 44.0
| | | | | | | | arrival_month <= 5.0
| | | | | | | | | lead_time <= 88.0: Canceled (247.0/102.0)
| | | | | | | | | lead_time > 88.0: Not_Canceled (210.0/91.0)
| | | | | | | | | arrival_month > 5.0: Canceled (328.0/121.0)
| | | | | | | | | type_of_meal_plan != Meal Plan 1: Canceled (667.0/236.0)
| | | | | | | | | avg_price_per_room > 111.76: Canceled (1670.0/432.0)
| | | | | | | | | arrival_month > 11.0: Not_Canceled (274.0/96.0)
| | | | | | | | | market_segment_type != Online
| | | | | | | | | | lead_time <= 90.0: Not_Canceled (4034.0/334.0)
| | | | | | | | | | lead_time > 90.0
| | | | | | | | | | | lead_time <= 117.0
| | | | | | | | | | | | avg_price_per_room <= 93.56: Not_Canceled (401.0/129.0)
| | | | | | | | | | | | avg_price_per_room > 93.56: Canceled (277.0/100.0)
| | | | | | | | | | | | lead_time > 117.0: Not_Canceled (483.0/81.0)
| | | | | | | | | | | no_of_special_requests > 0.0: Not_Canceled (9257.0/1122.0)
lead_time > 151.0
| avg_price_per_room <= 100.02
| | no_of_special_requests <= 0.0
| | | market_segment_type = Online: Canceled (607.0/11.0)
| | | market_segment_type != Online
| | | | no_of_adults <= 1.0: Not_Canceled (392.0/58.0)
| | | | no_of_adults > 1.0
| | | | | avg_price_per_room <= 88.5
| | | | | | lead_time <= 238.0: Not_Canceled (232.0/65.0)
| | | | | | lead_time > 238.0: Canceled (296.0/106.0)
| | | | | | avg_price_per_room > 88.5: Canceled (320.0/27.0)
| | | | | no_of_special_requests > 0.0: Not_Canceled (764.0/217.0)
| | | | avg_price_per_room > 100.02: Canceled (2146.0/95.0)
```

Από το δέντρο απόφασης που παρατίθεται, τα πιο σημαντικά χαρακτηριστικά για την πρόβλεψη της ακύρωσης των κρατήσεων είναι:

Lead Time: Ο χρόνος που μεσολαβεί μεταξύ της κράτησης και της ημέρας άφιξης είναι το πιο σημαντικό χαρακτηριστικό το οποίο όπως βλέπουμε είναι το χαρακτηριστικό εκείνο που αποτελεί κορυφή στο δέντρο μας αλλά ταυτόχρονα συναντάται συχνά στο δέντρο, υποδεικνύοντας τη μεγάλη επιρροή του στην πρόβλεψη της ακύρωσης.

No of Special Requests: Ο αριθμός των ειδικών αιτημάτων των πελατών φαίνεται να έχει μεγάλη σημασία, καθώς επηρεάζει την πιθανότητα ακύρωσης, με περισσότερα αιτήματα να συνδέονται με χαμηλότερα ποσοστά ακύρωσης.

Market Segment Type: Το είδος του καναλιού διανομής παίζει επίσης κρίσιμο ρόλο, ιδιαίτερα το κανάλι "Online," που σχετίζεται με υψηλότερα ποσοστά ακύρωσης.

Avg Price Per Room: Η μέση τιμή ανά δωμάτιο αποδεικνύεται σημαντικός παράγοντας με υψηλότερες τιμές να σχετίζονται συνήθως με μεγαλύτερη πιθανότητα ακύρωσης.

Arrival Month: Ο μήνας άφιξης επηρεάζει την απόφαση ακύρωσης, ιδιαίτερα σε συγκεκριμένες περιπτώσεις, όπως οι αφίξεις πριν τον Ιούνιο που συνδέονται με ακυρώσεις σε συνδυασμό με άλλους παράγοντες.

Type of Meal Plan: Το είδος του προγράμματος γευμάτων, ειδικά αν είναι ή όχι το "Meal Plan 1", επηρεάζει σημαντικά την πιθανότητα ακύρωσης.

Αυτά τα χαρακτηριστικά αποτελούν τα κύρια κριτήρια που καθορίζουν την πρόβλεψη της ακύρωσης σύμφωνα με το δέντρο απόφασης και υποδεικνύουν ποιες πληροφορίες είναι κρίσιμες για τη λήψη αποφάσεων σχετικά με την πιθανότητα ακύρωσης των κρατήσεων.

Συνοψίζοντας θα μπορούσαμε να πούμε τα εξής:

- Οι κρατήσεις με χαμηλό κόστος και μικρό χρονικό διάστημα που μεσολαβεί μεταξύ κράτησης και άφιξης είναι λιγότερο πιθανό να ακυρωθούν, ειδικά αν δεν είναι διαδικτυακές ή αν περιλαμβάνουν ειδικά αιτήματα.
- Η υψηλή τιμή δωματίου και το μεγάλο χρονικό διάστημα που μεσολαβεί μεταξύ κράτησης και άφιξης αυξάνουν την πιθανότητα ακύρωσης, ιδιαίτερα για online κρατήσεις.
- Οι κρατήσεις που περιλαμβάνουν ειδικά αιτήματα είναι λιγότερο πιθανό να ακυρωθούν.
- Κρατήσεις με μικρό χρονικό διάστημα που μεσολαβεί μεταξύ κράτησης και άφιξης και με χαμηλή τιμή δωματίου είναι πιο πιθανό να μην ακυρωθούν, ειδικά αν γίνονται εκτός διαδικτύου ή έχουν ειδικά αιτήματα.
- Αν η τιμή δωματίου είναι υψηλή και οι μέρες που μεσολαβούν μεταξύ κράτησης και άφιξης είναι πολλές, οι πιθανότητες ακύρωσης αυξάνονται, ιδιαίτερα για διαδικτυακές κρατήσεις.

10. Γενικά συμπεράσματα και προτάσεις

Η παρούσα διπλωματική επικεντρώθηκε στην ανάλυση δεδομένων κρατήσεων ξενοδοχείων, στην ανάπτυξη μοντέλων μηχανικής μάθησης για την πρόβλεψη της ακύρωσης κρατήσεων αλλά και στην ανάδειξη των χαρακτηριστικών – στοιχείων της κράτησης που αποδεδειγμένα επηρεάζουν την ακύρωση ή μη της κράτησης. Με την χρήση του λογισμικού WEKA και των αλγορίθμων IBK, J48 και Random Forest πραγματοποιήθηκαν 48 πειράματα με διαφορετικά σύνολα δεδομένων και τεχνικές φιλτραρίσματος. Από την ανάλυση προέκυψε ότι τα πλήρη σύνολα δεδομένων χωρίς την εφαρμογή φίλτρων απέδωσαν τα βέλτιστα αποτελέσματα, υποδεικνύοντας την αξία της αξιοποίησης όλων των διαθέσιμων πληροφοριών για την πρόβλεψη των ακυρώσεων.

Οι αλγόριθμοι IBK, J48 και Random Forest γενικά είχαν ικανοποιητικές αποδόσεις και αποδείχθηκε ότι είναι δυνατή η δημιουργία μοντέλων για την πρόβλεψη ακύρωσης κρατήσεων καθώς οι περισσότεροι αλγόριθμοι παρουσίασαν μέσες τιμές ακρίβειας άνω του 85%, με τον Random Forest να αναδεικνύεται ως ο πιο αποτελεσματικός, προσφέροντας ακρίβεια έως και 90,39% επιβεβαιώνοντας τους Nuno Antonio, Luis Nunes και Ana de Almeida (2017) αλλά και τους Zharfan Akbar Andriawan, Satriawan Rasyid Purnama, Adam Sukma Darmawan, Ricko, Adi Wibowo, Aris Sugiharto, Feri Wijayanto (2020). Δεδομένης της υψηλής απόδοσης του συγκεκριμένου αλγορίθμου συνιστάται η χρήση του για την πρόβλεψη της ακύρωσης κρατήσεων. Η σταθερή υψηλή του απόδοση και η ανθεκτικότητά του σε διαφορετικές παραλλαγές των δεδομένων τον καθιστά ικανό για εφαρμογή σε πραγματικές συνθήκες πρόβλεψης. Αυτό υποδεικνύει ότι οι δενδρικοί ταξινομητές φαίνεται να είναι πιο αποτελεσματικοί από τις μεθόδους παλινδρόμησης για την πρόβλεψη της ακύρωσης.

Εν συνεχεία, ο αλγόριθμος J48, με την παράμετρο pruned, επίσης παρουσίασε καλή απόδοση και ανέδειξε τα πιο σημαντικά χαρακτηριστικά που επηρεάζουν τις ακυρώσεις, με ποιο σημαντικό χαρακτηριστικό τον χρόνο που μεσολαβεί μεταξύ τη κράτησης και της ημέρας άφιξης (Lead Time) επιβεβαιώνοντας τους Zharfan Akbar Andriawan, Satriawan Rasyid Purnama, Adam Sukma Darmawan, Ricko, Adi Wibowo, Aris Sugiharto, Feri Wijayanto (2020). Σύμφωνα με τα αποτελέσματα του δέντρου απόφασης J48 άλλοι σημαντικοί παράγοντες που αναδείχθηκαν με την παρούσα έρευνα και επηρεάζουν την ακύρωση ή όχι της κράτησης είναι οι μέρες που μεσολαβούν μεταξύ της Κράτησης και της ημέρας άφιξης, ο αριθμός των τυχόν ειδικών αιτημάτων, το είδος καναλιού διανομής, ο συμφωνημένος τύπος διατροφής και η μέση τιμή ανά δωμάτιο. Η ανάλυσή τους από τις διοικήσεις ξενοδοχείων κρίνεται αναγκαία και απαραίτητη και θα πρέπει να ενσωματωθεί σε διαδικασίες διαχείρισης κρατήσεων προκειμένου να εντοπίζονται οι πιθανές ακυρώσεις.

Η μελέτη έδειξε ότι η χρήση πλήρων συνόλων δεδομένων χωρίς φίλτρα για την ανάλυση είναι πιο αποτελεσματική αλλά και αξιόπιστη. Το γεγονός αυτό κατέδειξε τη σημασία της διατήρησης όλων των χαρακτηριστικών στα σύνολα δεδομένων, ακόμα και αυτών που κρίνονται λιγότερο σημαντικά, για την αποτελεσματικότερη πρόβλεψη. Αυτό ενισχύει τα ευρήματα της έρευνας των Nuno Antonio, Luis Nunes και Ana de Almeida (2019) ότι η ζήτηση και οι ακυρώσεις μπορεί να διαφέρουν ανάλογα με το ξενοδοχείο, τον πελάτη ή την κράτηση ή λόγω εξωτερικών παραγόντων και συνεπώς όλα τα διαθέσιμα δεδομένα κρίνονται σημαντικά για την ακρίβεια της πρόβλεψης.

Τέλος, προτείνεται η διαρκής αξιολόγηση των αλγορίθμων με νέα δεδομένα προσαρμοσμένα σύμφωνα με τις μεταβαλλόμενες συνθήκες της αγοράς προκειμένου να διατηρηθεί η ακρίβεια των προβλέψεων. Η ενσωμάτωση τέτοιων αποτελεσματικών μοντέλων πρόβλεψης σε υπάρχουσες πλατφόρμες διαχείρισης κρατήσεων μπορεί να δώσει την ικανότητα της αυτόματης αναγνώρισης πιθανών ακυρώσεων και τη λήψη ορθών προληπτικών μέτρων από τα ξενοδοχεία.

Η παρούσα μελέτη ανέδειξε την σημασία της χρήσης της μηχανικής μάθησης από τις ξενοδοχειακές μονάδες για την πρόβλεψη ακυρώσεων και επιβεβαίωσε προηγούμενες μελέτες που ανέδειξαν τους αλγορίθμους βασισμένους σε δέντρα αποφάσεων και ειδικά τον Random Forest ως τους πλέον αποτελεσματικούς για τον σκοπό αυτόν ανεξάρτητα από την ακριβή δομή του συνόλου των δεδομένων.

11. Προτάσεις για μελλοντική έρευνα

Όπως σε κάθε μελέτη, έτσι και στην παρούσα υπάρχουν περιορισμοί. Τα ευρήματα της παρούσας μελέτης προέρχονται από έναν μόνο προορισμό – την Πορτογαλία. Κρίσιμο θα ήταν αντίστοιχες έρευνες να γίνουν με δεδομένα προερχόμενα από ξενοδοχεία ανά την υφήλιο αλλά ίσως και να επικεντρωθούν τοπικά όπως για παράδειγμα Κεντρική Ευρώπη ή Μεσόγειο όπου υπάρχουν κοινά χαρακτηριστικά στο φαινόμενο του τουρισμού. Συνιστάται η επανάληψη της παρούσας μελέτης με τη συλλογή δεδομένων από μια ευρεία γεωγραφική περιοχή με μια ποικιλία ενδεχομένως και άλλων παραγόντων που μπορούν να επηρεάσουν τα ποσοστά ακύρωσης. Μελλοντικές έρευνες μπορούν να εξετάσουν την επίδραση ίσως και νέων χαρακτηριστικών - παραγόντων που δεν συμπεριλήφθηκαν στην παρούσα έρευνα, όπως η εμπειρία του πελάτη με το ξενοδοχείο, οι κριτικές από το διαδίκτυο αλλά και παράγοντες εξωτερικοί όπως οι καιρικές συνθήκες και οι τοπικές εκδηλώσεις, που μπορεί να επηρεάσουν την απόφαση ακύρωσης.

12. Βιβλιογραφία

- [1] A. B. Pinheiro, A. S. Pinto, A. Abreu, E. Costa, and I. Borges, 'The Impact of Artificial Intelligence on the Tourism Industry: A Systematic Review', in *Advances in Tourism, Technology and Systems*, J. V. de Carvalho, Á. Rocha, P. Liberato, and A. Peña, Eds., in Smart Innovation, Systems and Technologies. Singapore: Springer, 2021, pp. 458–469. doi: 10.1007/978-981-33-4256-9_42.
- [2] S. Hadad, Y. Hadad, M. Malul, and M. Rosenboim, 'The Economic Efficiency of the Tourism Industry: A Global Comparison', *Tour. Econ.*, vol. 18, no. 5, pp. 931–940, Oct. 2012, doi: 10.5367/te.2012.0165.
- [3] 'Global and regional tourism performance'. Accessed: Feb. 27, 2024. [Online]. Available: <https://www.unwto.org/tourism-data/global-and-regional-tourism-performance>
- [4] A. J. Sánchez-Medina and E. C-Sánchez, 'Using machine learning and big data for efficient forecasting of hotel booking cancellations', *Int. J. Hosp. Manag.*, vol. 89, p. 102546, Aug. 2020, doi: 10.1016/j.ijhm.2020.102546.
- [5] A. Riasi, Z. Schwartz, and C.-C. Chen, 'A proposition-based theorizing approach to hotel cancellation practices research', *Int. J. Contemp. Hosp. Manag.*, vol. 30, no. 11, pp. 3211–3228, Jan. 2018, doi: 10.1108/IJCHM-10-2017-0616.
- [6] J. Oh and X. Su, 'Optimal Pricing and Overbooking of Reservations', *Prod. Oper. Manag.*, vol. 31, no. 3, pp. 928–940, Mar. 2022, doi: 10.1111/poms.13583.
- [7] M. Falk and M. Vieru, 'Modelling the cancellation behaviour of hotel guests', *Int. J. Contemp. Hosp. Manag.*, vol. 30, no. 10, pp. 3100–3116, Jan. 2018, doi: 10.1108/IJCHM-08-2017-0509.
- [8] M. Yoo, A. K. Singh, and N. Loewy, 'Predicting hotel booking cancelation with machine learning techniques', *J. Hosp. Tour. Technol.*, vol. 15, no. 1, pp. 54–69, Jan. 2023, doi: 10.1108/JHTT-07-2022-0227.
- [9] E. J. Kim, E. L. Kim, M. Kim, and S. Tanford, 'Post-Pandemic hotel cancellation policy: Situational cues as perceived risk triggers', *J. Hosp. Tour. Manag.*, vol. 55, pp. 153–160, Jun. 2023, doi: 10.1016/j.jhtm.2023.03.019.
- [10] D. Romero Morales and J. Wang, 'Forecasting cancellation rates for services booking revenue management using data mining', *Eur. J. Oper. Res.*, vol. 202, no. 2, pp. 554–562, Apr. 2010, doi: 10.1016/j.ejor.2009.06.006.
- [11] N. Antonio, A. de Almeida, and L. Nunes, 'Big Data in Hotel Revenue Management: Exploring Cancellation Drivers to Gain Insights Into Booking Cancellation Behavior', *Cornell Hosp. Q.*, vol. 60, no. 4, pp. 298–319, Nov. 2019, doi: 10.1177/1938965519851466.
- [12] G. J. van Ryzin and K. T. Talluri, 'An Introduction to Revenue Management', in *Emerging Theory, Methods, and Applications*, in INFORMS TutORials in Operations Research. , INFORMS, 2005, pp. 142–194. doi: 10.1287/educ.1053.0019.
- [13] C.-C. Chen, Z. Schwartz, and P. Vargas, 'The search for the best deal: How hotel cancellation policies affect the search and booking decisions of deal-seeking customers', *Int. J. Hosp. Manag.*, vol. 30, no. 1, pp. 129–135, Mar. 2011, doi: 10.1016/j.ijhm.2010.03.010.
- [14] X. Guo, Y. Dong, and L. Ling, 'Customer perspective on overbooking: The failure of customers to enjoy their reserved services, accidental or intended?', *J. Air Transp. Manag.*, vol. 53, pp. 65–72, Jun. 2016, doi: 10.1016/j.jairtraman.2016.01.001.
- [15] B. M. Noone and C. H. Lee, 'Hotel Overbooking: The Effect of Overcompensation on Customers' Reactions to Denied Service', *J. Hosp. Tour. Res.*, vol. 35, no. 3, pp. 334–357, Aug. 2011, doi: 10.1177/1096348010382238.
- [16] N. Antonio, A. de Almeida, and L. Nunes, 'An Automated Machine Learning Based Decision Support System to Predict Hotel Booking Cancellations', *Data Sci. J.*, vol. 18, pp. 32–32, Jul. 2019, doi: 10.5334/dsj-2019-032.

- [17] S. Chen, E. W. T. Ngai, Y. Ku, Z. Xu, X. Gou, and C. Zhang, 'Prediction of hotel booking cancellations: Integration of machine learning and probability model based on interpretable feature interaction', *Decis. Support Syst.*, vol. 170, p. 113959, Jul. 2023, doi: 10.1016/j.dss.2023.113959.
- [18] N. Antonio, A. de Almeida, and L. Nunes, 'Predicting hotel booking cancellations to decrease uncertainty and increase revenue', *Tour. Manag. Stud.*, vol. 13, no. 2, Art. no. 2, Apr. 2017.
- [19] Md. S. Satu, K. Ahammed, and M. Z. Abedin, 'Performance Analysis of Machine Learning Techniques to Predict Hotel booking Cancellations in Hospitality Industry', in *2020 23rd International Conference on Computer and Information Technology (ICCIT)*, Sep. 2020, pp. 1–6. doi: 10.1109/ICCIT51783.2020.9392648.
- [20] Z. A. Andriawan *et al.*, 'Prediction of Hotel Booking Cancellation using CRISP-DM', in *2020 4th International Conference on Informatics and Computational Sciences (ICICoS)*, Aug. 2020, pp. 1–6. doi: 10.1109/ICICoS51170.2020.9299011.
- [21] 'Hotel Reservations Dataset'. Accessed: Sep. 09, 2024. [Online]. Available: <https://www.kaggle.com/datasets/ahsan81/hotel-reservations-classification-dataset>