

## 1. Θεματική Περιοχή

### Εφαρμογή:

Στο τουριστικό περιβάλλον, πολλές επιχειρήσεις δεν έχουν τους απαιτούμενους πόρους και την τεχνογνωσία για να αξιοποιήσουν τα δεδομένα από τις καθημερινές δραστηριότητες. Για παράδειγμα, πληροφορίες από συστήματα κρατήσεων, προτιμήσεις πελατών και σχόλια επισκεπτών παραμένουν συνήθως αναξιοποίητα.

### Πρόβλημα:

Χωρίς την εφαρμογή μεθόδων, οι επιχειρήσεις δεν μπορούν να προβλέψουν τη ζήτηση, να εντοπίσουν συγκεκριμένες ανάγκες πελατών τους ή να ανακαλύψουν τάσεις που θα τις βοηθήσουν για καινοτομία. Έτσι, χάνουν ανταγωνιστικό πλεονέκτημα, σε περιόδους αστάθειας όπου η ευελιξία και οι καλές αποφάσεις είναι κρίσιμες.

### Ανάγκη & Καινοτομία:

Η αξιοποίηση τεχνικών μπορεί να βοηθήσει στην πρόβλεψη της ζήτησης, στις προτιμήσεις πελατών, και στις στοχευμένες προσφορές ή πακέτων υπηρεσιών. Οι επιχειρήσεις με περιορισμένους πόρους μπορούν χρησιμοποιώντας cloud υπηρεσίες, να αποκτήσουν πρόσβαση σε εργαλεία που θα τις οδηγήσουν σε καλύτερη οργάνωση, βέλτιστη τιμολόγηση και στρατηγικές μάρκετινγκ.

## 2. Ανάλυση

### Ανάγκες για Δεδομένα:

Οι επιχειρήσεις χρειάζονται δεδομένα για να κατανοήσουν τους πελάτες και την αγορά:

- **Ιστορικά Δεδομένα:** Αριθμός κρατήσεων ανά χρονική περίοδο (μέρα, εβδομάδα, μήνα), διάρκεια παραμονής, εποχιακές διακυμάνσεις, τιμές και είδος καταλύματος. Αυτά τα δεδομένα βοηθούν στον εντοπισμό περιόδων αιχμής, στον καλύτερο προγραμματισμό προσωπικού, καθώς και στην τιμολόγηση.
- **Δημογραφικά Δεδομένα:** Ηλικία, χώρα, εισόδημα και οικογενειακή κατάσταση. Τέτοιες πληροφορίες είναι απαραίτητες για την ταξινόμηση της πελατείας. Μπορούν να δημιουργηθούν ομάδες (π.χ. νέοι & οικογένειες).
- **Ικανοποίηση & Αξιολόγηση:** Σχόλια πελατών από φόρμες, online βαθμολογίες, κριτικές στα social media. Αυτά τα δεδομένα βοηθάνε στον εντοπισμό προβλημάτων και τη βελτίωση των υπηρεσιών.
- **Δεδομένα Εξωτερικών Παραγόντων:** Το κλίμα επηρεάζει τη ζήτηση, καθώς καλός καιρός ισούται με αυξημένες κρατήσεις. Αντίστοιχα, η κακοκαιρία ενδέχεται να αλλάζει τη διαθεσιμότητα υπηρεσιών.

### Πλαίσιο Καταγραφής & Αποθήκευσης Δεδομένων:

Η συλλογή των παραπάνω δεδομένων γίνεται από:

- **Συστήματα Κρατήσεων (CRM):** Τα συστήματα CRM αποθηκεύουν αυτόματα πληροφορίες πελατών και κρατήσεων. Η εξαγωγή δεδομένων μπορεί να γίνεται και σε πραγματικό χρόνο.

- Πλατφόρμες (API): Ενσωμάτωση με APIs πλατφόρμες κρατήσεων και αξιολογήσεων, για αυτόματη λήψη σχολίων, βαθμολογιών και τιμών.
- IoT Αισθητήρες: Σε χώρους τουριστικών εγκαταστάσεων, οι αισθητήρες μπορούν να μετρούν πληρότητα και επισκεψιμότητα στη χρήση εγκαταστάσεων.

Όλα αυτά πάνε σε ένα (Data Warehouse ή Data Lake στο Cloud. Οι διαδικασίες ETL θα βάζουν τα δεδομένα σε κοινά σχήματα, θα τα καθαρίζουν και θα τα φορτώνουν στη βάση.

Ανάγκες για Αποθηκευτικό Χώρο:

- Στα πρώτα στάδια αρκούν GB για ιστορικά δεδομένα και κριτικές.
- Καθώς η επιχείρηση μεγαλώνει και μαζεύονται περισσότερα δεδομένα, ο χώρος μπορεί να φτάσει δεκάδες ή εκατοντάδες GB.
- Το cloud έτσι, επιτρέπει την κλιμάκωση του αποθηκευτικού χώρου ανάλογα με τις ανάγκες, αποφεύγοντας το υψηλό αρχικό κόστος.

Χρήστες & Παροχής Γνώσης:

- Χρήστες: Διευθυντικά στελέχη, τμήμα μάρκετινγκ, τμήμα πωλήσεων, οικονομικοί αναλυτές.
- Στόχος: Χρήση αυτών των πληροφοριών για βελτιστοποίηση πόρων, ανάπτυξη στοχευμένων προσφορών, βελτίωση της εμπειρίας πελατών αλλά ακόμη και σχεδιασμό για επένδυση.

### 3. Προεπεξεργασία

Επειδή τα δεδομένα προέρχονται από διαφορετικές πηγές (κρατήσεις, κριτικές, δημογραφικά στοιχεία, κ.λπ.) και μπορεί να παρουσιάζουν ελλείψεις, λάθος κλίμακα ή ακόμη και θόρυβο, η προεπεξεργασία, που υλοποιείται και στον κώδικα στο τέλος κάτω κάτω, στοχεύει στη βελτίωση της ποιότητας, της χρησιμότητας των δεδομένων αλλά και στην βέλτιστη λειτουργία των αλγόριθμων εξόρυξης.

Ανάγκες για Προεπεξεργασία:

- Ελλιπείς Τιμές: Κάποιες εγγραφές μπορεί να μην έχουν συμπληρωμένα όλα τα πεδία. Αν δεν λυθεί το συγκεκριμένο ζήτημα τότε, θα επηρεαστούν αρνητικά οι αλγόριθμοι εξόρυξης.
- Διαφορετικές Κλίμακες: Χαρακτηριστικά όπως η ηλικία, το εισόδημα, η διάρκεια διαμονής έχουν διαφορετικά εύρη τιμών. Χωρίς κανονικοποίηση, αλγόριθμοι όπως το K-means μπορεί να μην λειτουργούν εύστοχα.
- Πολυπλοκότητα Δεδομένων: Όταν υπάρχουν πολλά χαρακτηριστικά, πρέπει να συμπυκνωθούν σε λιγότερες διαστάσεις, για να γίνει η ανάλυση πιο αποδοτική και οι οπτικοποιήσεις πιο καλές.

Βήματα & Αλγόριθμοι:

1. Καθαρισμός Δεδομένων:  
Στον κώδικα, εισάγονται συγκεκριμένες ελλιπείς τιμές, οι οποίες αντικαθίστανται μετά με τη διάμεσο. Το αποτέλεσμα είναι, ότι διατηρούμε όλες τις εγγραφές και έτσι τα δεδομένα έχουν συνεκτικότητα (σημαντικό για να μην χάσουμε πληροφορία).
2. Κανονικοποίηση (Normalization):  
Με το MinMaxScaler, οι τιμές της ηλικίας, του εισοδήματος και της διάρκειας διαμονής

μετασχηματίζονται σε κλίμακα του  $[0,1]$ . Αυτό βοηθά ώστε κανένα χαρακτηριστικό να μην κυριαρχεί στην ανάλυση εξαιτίας της τάξης του μεγέθους του.

### 3. Μείωση Διαστάσεων (PCA):

Εφαρμόζεται PCA για να μειωθεί η πολυπλοκότητα των δεδομένων, εξάγοντας τις κυριότερες συνιστώσες. Έτσι κατανέμονται οι πελάτες, κάτι που διευκολύνει το clustering.

## 4. Αλγόριθμοι Εξόρυξης Δεδομένων

### 1. Clustering K-means:

- Στόχος: Διαχωρισμός των πελατών ανά clusters με βάση τα χαρακτηριστικά τους (ηλικία, εισόδημα, διάρκεια διαμονής). Έτσι, εντοπίζονται ομάδες με παρόμοια μοτίβα, προτιμήσεις αλλά και συμπεριφορά.
- Γνώση: Η επιχείρηση θα μπορεί να στοχεύσει διαφορετικά clusters με συγκεκριμένες προσφορές. Δηλαδή, ένα cluster μπορεί να αντιπροσωπεύει νέους με χαμηλό εισόδημα που θέλουν οικονομικές λύσεις, ενώ αντιστρόφως ένα άλλο μπορεί να έχει πελάτες μεγάλου εισοδήματος που ψάχνουν ποιοτικές λύσεις.
- Παράδειγμα στον Κώδικα: Χρήση KMeans( $n\_clusters=3$ ). Μετά την προεπεξεργασία, γίνεται clustering και τα γίνεται οπτικοποίηση των αποτελεσμάτων με τις συνιστώσες (PCA).

### 2. Association Rules Apriori:

- Στόχος: Εύρεση κανόνων συσχέτισης μεταξύ υπηρεσιών. Δηλαδή, πελάτες που επιλέγουν συγκεκριμένη δραστηριότητα (π.χ. excursion) μπορεί να θέλουν και spa, αυτό πιθανότατα αποτελεί στοιχείο για δημιουργία προσφορών.
- Γνώση: Η επιχείρηση θα δημιουργήσει 'δελεαστικά πακέτα'. Για παράδειγμα, αν η ανάλυση δείξει ότι οι πελάτες που επιλέγουν excursion διαλέξουν και hiking τότε η επιχείρηση μπορεί να προσφέρει ένα ενιαίο πακέτο με μία έκπτωση.
- Παράδειγμα στον Κώδικα: Χρήση Apriori και συνάρτηση association\_rules. Οι κανόνες που προέκυψαν στον κώδικα είναι:
  - excursion -> spa confidence 0.505
  - excursion -> hiking confidence 0.526 Φαίνεται ότι υπάρχει πιθανότητα όσοι αγοράζουν excursion να αγοράσουν και spa ή hiking.

### 3. Classification Naive Bayes:

- Στόχος: Πρόβλεψη προτίμησης δραστηριότητας βάσει π.χ. δημογραφικών δεδομένων για τους νέους τουρίστες.
- Γνώση: Η επιχείρηση μπορεί να προτείνει κατάλληλα προγράμματα δραστηριοτήτων σε νέους πελάτες. Αν ο ταξιδιώτης είναι 30 ετών με συγκεκριμένο εισόδημα, τότε ο αλγόριθμος μπορεί να προβλέψει εάν θα προτιμήσει sea, mountain ή cultural.
- Παράδειγμα στον Κώδικα: Χρήση Naive Bayes. Για έναν νέο τουρίστα 30 ετών με εισόδημα 3000, το μοντέλο προβλέπει sea.

## 5. Στρατηγικές Αποφάσεις

Σύμφωνα με τα παραπάνω, η επιχείρηση μπορεί:

- Διαχείριση Πόρων: Να ξέρει πότε αυξάνεται η ζήτηση, για να προσλάβει κατάλληλο προσωπικό ή να προμηθευτεί εγκαίρως αγαθά και υπηρεσίες.
- Στοχευμένο Μάρκετινγκ: Αν μια κατηγορία πελατών αγαπά τις περιπέτειες, στέλνονται προσωποποιημένες προσφορές. Αν τώρα μία άλλη ομάδα προτιμά πολιτιστικές δραστηριότητες, τότε ομοίως δημιουργούνται τα αντίστοιχα πακέτα.
- Βελτιστοποίηση Υπηρεσιών: Οι κανόνες συσχέτισης συμβάλλουν στη δημιουργία συμφερουσών προσφορών για αύξηση των αγορών.
- Μακροπρόθεσμος Σχεδιασμός: Με την πρόβλεψη προτιμήσεων, η επιχείρηση μπορεί να επενδύσει σε υπηρεσίες που θα έχουν ζήτηση στο μέλλον, ενισχύοντας την ανταγωνιστικότητά.