



ΧΑΡΟΚΟΠΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ

ΧΑΡΟΚΟΠΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΜΑΤΙΚΗΣ

Σχεδιασμός αναλυτικών ταμπλό για επιχειρησιακά δεδομένα
Πτυχιακή εργασία

Σταυρούλα Πράττη

Αθήνα, 2024



HAROKOPIO UNIVERSITY
HAROKOPIO UNIVERSITY OF ATHENS
INFORMATICS AND TELEMATICS DEPARTMENT

Designing analytical dashboards for business data
Thesis

Stavroula Pratti

Athens, 2024



ΧΑΡΟΚΟΠΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ

ΧΑΡΟΚΟΠΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΜΑΤΙΚΗΣ

Τριμελής Εξεταστική Επιτροπή

Ηρακλής Βαρλάμης (Επιβλέπων)
Καθηγητής

Τμήματος Πληροφορικής και Τηλεματικής,
Χαροκόπειο Πανεπιστήμιο

Δημήτριος Μιχαήλ
Αναπληρωτής Καθηγητής

Τμήματος Πληροφορικής και Τηλεματικής,
Χαροκόπειο Πανεπιστήμιο

Κλεοπάτρα Μπαρδάκη
Επίκουρη Καθηγήτρια

Τμήματος Πληροφορικής και Τηλεματικής,
Χαροκόπειο Πανεπιστήμιο

Η Σταυρούλα Πράττη

δηλώνω υπεύθυνα ότι:

- 1)** Είμαι ο κάτοχος των πνευματικών δικαιωμάτων της πρωτότυπης αυτής εργασίας και από όσο γνωρίζω η εργασία μου δε συκοφαντεί πρόσωπα, ούτε προσβάλει τα πνευματικά δικαιώματα τρίτων.
- 2)** Αποδέχομαι ότι η ΒΚΠ μπορεί, χωρίς να αλλάξει το περιεχόμενο της εργασίας μου, να τη διαθέσει σε ηλεκτρονική μορφή μέσα από τη ψηφιακή Βιβλιοθήκη της, να την αντιγράψει σε οποιοδήποτε μέσο ή/και σε οποιοδήποτε μορφότυπο καθώς και να κρατά περισσότερα από ένα αντίγραφα για λόγους συντήρησης και ασφάλειας.

Αφιέρωση

Αφιερώνω την εργασία αυτή στον αδερφό μου Παναγιώτη, που με ενέπνευσε να ασχοληθώ με τον μαγικό κόσμο της πληροφορικής και που ήταν πάντα δίπλα μου από την πρώτη μέρα.

Εύχομαι επίσης στην μικρή μας αδερφή, τη Δέσποινα, να γίνει καλύτερη προγραμματίστρια κι από εμένα και από τον αδερφό μας.

Ευχαριστίες

Με την εκπόνηση αυτής της πτυχιακής εργασίας, συγκεκριμένες δεξιότητές μου στις Βάσεις και Ανάλυση Δεδομένων αναπτύχθηκαν. Πλέον για μένα το όραμά μου να ασχοληθώ με αυτόν τον κλάδο είναι ξεκάθαρο. Φυσικά δεν βρέθηκα σε αυτό το στάδιο χωρίς υποστήριξη και βοήθεια.

Με το πέρας των σπουδών μου λοιπόν, θα ήθελα να εκφράσω την ευγνωμοσύνη μου για το Χαροκόπειο Πανεπιστήμιο. Θα ήθελα να ευχαριστήσω όλο το προσωπικό του Πανεπιστημίου, από τη Γραμματεία μέχρι και το Γραφείο Διασύνδεσης και τους υπεύθυνους του Erasmus. Είχα την τύχη να μάθω από σπουδαίους καθηγητές που με καθοδηγούσαν συνεχώς.

Ιδιαίτερα ευχαριστώ τον κ. Βαρλάμη, ο οποίος τη δύσκολη περίοδο της καραντίνας με ενέπνευσε με το μάθημα των Βάσεων Δεδομένων, μου μετέδωσε τον ενθουσιασμό του για τον κλάδο αυτόν και μου εμπιστεύτηκε μια από τις σημαντικότερες εργασίες στην ακαδημαϊκή μου καριέρα.

Ταυτόχρονα, νιώθω τυχερή για τους φίλους που γνώρισα αυτά τα τέσσερα χρόνια, για τη βοήθειά τους τόσο σε ακαδημαϊκό όσο και σε προσωπικό επίπεδο. Ο Άγγελος, η Κωνσταντίνα, η Κάλλη, η Φαίη, ο Σπύρος και πολλοί άλλοι μου απέδειξαν ότι το Πανεπιστήμιο είναι κάτι παραπάνω από γνώσεις : είναι σχέσεις.

Τέλος οι γονείς μου θα είναι πάντα πρότυπο γιατί μου έδειξαν από μικρή με τις πράξεις τους τη σημασία της εκπαίδευσης. Κατάφεραν και μετέδωσαν σε όλα τους τα παιδιά την αγάπη τους για την Επιστήμη της Πληροφορικής και μας προσέφεραν όλα τα εφόδια που χρειαζόμασταν όλα αυτά τα χρόνια.

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

Περίληψη στα Ελληνικά	9
Περίληψη στα Αγγλικά	10
ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ	11
ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ	13
ΣΥΝΤΟΜΟΓΡΑΦΙΕΣ	14
ΕΙΣΑΓΩΓΗ: Σκοπός και δομή της εργασίας	15
ΚΕΦΑΛΑΙΟ 1: Data Warehouse.....	16
1.1 Εισαγωγή.....	16
1.1.1 Στόχοι του Data Warehouse	17
1.1.2 Τύποι του Data Warehouse.....	17
1.2 Αρχιτεκτονική	18
1.3 Components	19
1.4 OLAP & OLTP	21
1.5 Schema.....	21
1.6 Fact table.....	22
1.7 Dimension tables	22
1.8 Σχέση μεταξύ fact & dimension tables	23
1.9 Primary key	23
1.10 Foreign key	24
ΚΕΦΑΛΑΙΟ 2: Business Intelligence	25
2.1 Εισαγωγή.....	25
2.1.1 «Προσωποποιημένη» Επιχειρηματική Ευφυΐα.....	26
2.2 Πλεονεκτήματα	26
2.3 Μειονεκτήματα	27
2.4 Gartner Magic Quadrant BI tools	28
ΚΕΦΑΛΑΙΟ 3: Οπτικοποίηση δεδομένων	29
3.1 Εισαγωγή.....	29
3.2 Dashboards	29
3.2.1 Στρατηγικά dashboards.....	29
3.2.2 Λειτουργικά dashboards	30
3.2.3 Αναλυτικά dashboards.....	31
3.2.4 Τακτικά dashboards	31
3.2.5 Χρώματα σε dashboards	31

3.2.6 Σύγκριση dashboards.....	32
3.3 Γραφήματα.....	33
ΚΕΦΑΛΑΙΟ 4: Δημιουργία του Data Warehouse	38
4.1 Εισαγωγή.....	38
4.2 Excel	38
4.3 Data Load	39
4.4 Data Cleaning	40
4.5 Ένωση όμοιων πινάκων	42
4.5.1 Επαληθεύσεις	43
4.6 Fact και Dimension πίνακες.....	44
4.7 Σχεδίαση στο Tableau.....	49
ΚΕΦΑΛΑΙΟ 5: Υλοποίηση στο Tableau.....	51
5.1 Εισαγωγή.....	51
5.1.1 Πλεονεκτήματα & μειονεκτήματα του Tableau	52
5.1.2 Groups.....	53
5.1.3 Ιεραρχίες	54
5.2 Star Schema.....	55
5.3 Στρατηγικό κυβερνητικό dashboard.....	57
5.3.1 Forecast μοντέλο	61
5.3.2 Calculated field	62
5.4 Λειτουργικό dashboard για real estate επενδυτές	64
5.4.1 Slicer.....	65
5.4.2 Clusters.....	67
ΚΕΦΑΛΑΙΟ 6: Συμπεράσματα.....	71
6.1 Εισαγωγή.....	71
6.2 Στρατηγικό κυβερνητικό dashboard.....	71
6.3 Λειτουργικό dashboard για real estate επενδυτές	73
Επίλογος	77
ΒΙΒΛΙΟΓΡΑΦΙΑ	78

Περίληψη στα Ελληνικά

Η εργασία αυτή πραγματεύεται την ανάλυση και οπτικοποίηση δεδομένων που αντλούνται από την πλατφόρμα της Airbnb και που αφορούν ακίνητα από τέσσερις μεγάλες περιοχές της χώρας μας: την Αθήνα, Θεσσαλονίκη, Κρήτη και τα νησιά του νοτίου Αιγαίου. Συγκεκριμένα, το τελικό αποτέλεσμα στοχεύει στον σχεδιασμό δύο διαφορετικών ειδών dashboards: ένα strategic dashboard για το Υπουργείο Οικονομικών μιας χώρας κι ένα operational dashboard για real-estate επενδυτές. Κύριος σκοπός των ταμπλό είναι η παρακολούθηση της διακύμανσης των τιμών των ακινήτων.

Για την υλοποίηση αυτής της σχεδίασης, έγινε ανάλυση των διαφορετικών CSV αρχείων που παρέχονται από την Airbnb, με έμφαση στα δεδομένα σχετικά με τις καταχωρήσεις ακινήτων, τις τιμές, τη διαθεσιμότητα, καθώς και τις αξιολογήσεις. Έγινε επίσης επιλογή των στηλών που θα διατηρηθούν στους πίνακες της Βάσης Δεδομένων, λαμβάνοντας υπόψη τα ζητούμενα των δύο τελικών χρηστών.

Για τη δημιουργία και διαχείριση της βάσης χρησιμοποιήθηκε το SQL Server Management Studio, ενώ η οπτικοποίηση και η αναπαράσταση των δεδομένων έγινε δυνατή με το Tableau, ένα ισχυρό εργαλείο ανάλυσης δεδομένων με αμέτρητες δυνατότητες που εξερεύνησα.

Σε ένα data-driven κόσμο τα dashboards και τα διαγράμματα αποτελούν σημαντικά εργαλεία για την ανάγνωση και κατανόηση των μεγάλων και περίπλοκων δεδομένων. Με την κατάλληλη επιλογή τους ο χρήστης μπορεί να αντιληφθεί και να απομνημονεύσει ευκολότερα τη ζητούμενη πληροφορία.

Η εργασία αναδεικνύει μια βελτιστοποιημένη προσέγγιση για την ανάλυση και τη δημιουργία Βάσεων Δεδομένων από δεδομένα της Airbnb, προσφέροντας ένα ισχυρό εργαλείο για την κατανόηση και παρακολούθηση της αγοράς ακινήτων σε διαφορετικά επίπεδα. Οι πίνακες, όπως αυτοί για τις καταχωρήσεις, τις τιμές και τις αξιολογήσεις, παρέχουν μια ολοκληρωμένη προοπτική της αγοράς.

Στην παρούσα πτυχιακή, πέρα από την αρχιτεκτονική του συστήματος και τα αναλυτικά βήματα που ακολουθήθηκαν για την υλοποίησή της, θα αναλυθεί επίσης κι ένα μεγάλο κομμάτι της θεωρίας του Data Analytics.

Λέξεις κλειδιά: Tableau, SSMS, Dashboard, SQL, Data Analysis

Περίληψη στα Αγγλικά

This thesis deals with the analysis and visualization of data drilled from the Airbnb platform and concerning properties from four major regions of our country: Athens, Thessaloniki, Crete, and the islands of the southern Aegean. Specifically, the result aims to design two different types of dashboards: a strategic dashboard for a country's Ministry of Finance and an operational dashboard for real-estate investors. The main purpose of the dashboards is to monitor the fluctuation of real estate prices.

To implement this design, I analyzed the different CSV files provided by Airbnb, with a focus on data on property listings, prices, availability, and ratings. The columns to be maintained in the database tables were also selected, considering the requests of the two end users.

SQL Server Management Studio was used to create and manage the database and data visualization and representation was made by Tableau, a powerful data analysis tool with countless possibilities that I explored.

In a data-driven world, dashboards and charts are important tools for reading and understanding big and complex data. With appropriate selections, the user can more easily understand and memorize the requested information.

The thesis highlights an optimized approach for analyzing and building databases from Airbnb data, offering a powerful tool for understanding and monitoring the real estate market at different levels. Tables, such as those for listings, prices, and ratings, provide a comprehensive view of the market.

In this thesis, apart from the architecture of the system and the analytical steps followed for its implementation, a large part of the theory of Data Analytics will also be analyzed.

Keywords: Tableau, SSMS, Dashboard, SQL, Data Analysis

ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ

- Εικ.1: Three tier αρχιτεκτονική
- Εικ.2: Components
- Εικ.3: Star schema
- Εικ.4: Παράδειγμα fact table
- Εικ.5: Παράδειγμα dimension table
- Εικ.6: Σχέση μεταξύ fact & dimensional tables
- Εικ.7: Προσεγγίσεις BI
- Εικ. 8: Gartner Magic Quadrant BI tools
- Εικ. 9: Στρατηγικό dashboard
- Εικ. 10: Χρώματα σε dashboards
- Εικ. 11: Απλό κείμενο
- Εικ. 12: Πίνακας
- Εικ. 13: Heatmap
- Εικ. 14: Scatter plot
- Εικ. 15: Line graph
- Εικ. 16: Slopegraph
- Εικ. 17: Vertical bar chart
- Εικ. 18: Horizontal bar chart
- Εικ. 19: Stacked horizontal bar chart
- Εικ. 20: Waterfall chart
- Εικ. 21: Reviews
- Εικ. 22: Calendars
- Εικ. 23: Listings
- Εικ. 24: Data Dictionary
- Εικ. 25: Διαγραφή στηλών
- Εικ. 26: Μετονομασία στηλών
- Εικ. 27: Union πινάκων
- Εικ. 28: Επαληθεύσεις
- Εικ. 29: FactReviews
- Εικ. 30: FactCalendar
- Εικ. 31: DimLocation, Dim.PropertyType, DimRoomType
- Εικ. 32: DimHost
- Εικ. 33: FactListings
- Εικ. 34: Τελικοί πίνακες
- Εικ. 35: Προσχέδιο για το strategic dashboard
- Εικ. 36: Προσχέδιο για το operational dashboard
- Εικ. 37: Χρόνος στο Tableau
- Εικ. 38: Επαληθεύσεις μεταξύ ΒΔ και Tableau
- Εικ. 39: Location groups
- Εικ. 40: Property type groups
- Εικ. 41: Χρονική ιεραρχία
- Εικ. 42: Τοπική ιεραρχία
- Εικ. 43: Property type ιεραρχία
- Εικ. 44: Star schema
- Εικ. 45: Star schema από Tableau

- Εικ. 46: KPIs για στρατηγικό dashboard
Εικ. 47: Map of listings
Εικ. 48: Avg. price per day
Εικ. 49: Listings per location
Εικ. 50: Αναλυτικά listings per location
Εικ. 51: Annual availability
Εικ. 52: Annual avg. price & forecasting
Εικ. 53: Αναλυτικό annual avg. price & forecasting
Εικ. 54: Calculated field
Εικ. 55: Room type percentage
Εικ. 56: Expected daily total price per location
Εικ. 57: Strategic Government Dashboard
Εικ. 58: KPIs για λειτουργικό dashboard
Εικ. 59: Map
Εικ. 60: Expected monthly total price per location
Εικ. 61: Αναλυτικό expected monthly total price per location
Εικ. 62: Monthly avg. price per location
Εικ. 63: Price per type of review
Εικ. 64: Avg. beds and price per property type
Εικ. 65: Superhost statistics
Εικ. 66: Avg. price/host total listings per host year since
Εικ. 67: Count of listings rated per time
Εικ. 68: Operational Real Estate Dashboard
Εικ. 69: Συνοπτικές απαντήσεις για το strategic dashboard
Εικ. 70: Συνοπτικές απαντήσεις για το operational dashboard

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

Πίν.1: Διαφορές dashboards

Πίν.2: Σύγκριση ΒΔ

ΣΥΝΤΟΜΟΓΡΑΦΙΕΣ

SSMS	SQL Server Management Studio
BI	Business Intelligence
DW	Data Warehouse
ΒΔ	Βάση/εις Δεδομένων
PK	Primary Key
FK	Foreign Key

ΕΙΣΑΓΩΓΗ: Σκοπός και δομή της εργασίας

Ο ραγδαίος ρυθμός της ψηφιακής παραγωγής δεδομένων που παρατηρείται στις μέρες μας, καθιστά τις παραδοσιακές μεθόδους επεξεργασίας πληροφοριών ανεπαρκείς. Η έλλειψη συστηματικής ανάλυσης δεδομένων οδηγεί μάλλον στην αδυναμία αντιμετώπισης των πελώριων και ποικίλων συνόλων δεδομένων.

Έχοντας συνειδητοποίησει τις παραπάνω προκλήσεις, αντιλαμβάνομαι ότι η ανάλυση και η οπτικοποίηση αποτελούν το κλειδί για την κατανόηση των πελατών και την πρόβλεψη των τάσεων στην αγορά. Η εξαγωγή ευφυών αποφάσεων βασιζόμενη σε δεδομένα γίνεται εφικτή με την οπτικοποίηση, προσφέροντας καλύτερη ενημέρωση και ευανάγνωστα γραφήματα. Συνολικά, η αξιοποίηση αυτών των διαδικασιών συντελεί στην επίτευξη πιο ευέλικτων και ανταγωνιστικών επιχειρηματικών στρατηγικών.

Η παρούσα πτυχιακή εργασία είναι μια προσπάθεια υλοποίησης ενός Data Warehouse με δεδομένα αντλούμενα από την Airbnb, που αφορούν ακίνητα στην Ελλάδα. Αρχικό και κύριο βήμα αποτελεί το Data Cleaning στα εισαγόμενα στη βάση CSVs. Αφού δημιουργηθούν οι επιθυμητοί πίνακες με SQL queries που θα αναλυθούν περαιτέρω, είναι επόμενο να διαγραφούν οι στήλες οι οποίες δεν είναι απαραίτητες. Υστερα εκτελείται μετονομασία των στηλών, όπου αυτές δεν είναι νοηματικά ξεκάθαρες, με σκοπό τη διευκόλυνση στην κατανόηση των περιεχομένων τους. Δημιουργώ τις σχέσεις, τους νέους πίνακες και επαληθεύω, εκτελώντας queries.

Καταλήγω -με τη βοήθεια του Tableau- στον σχεδιασμό δύο διαφορετικών ειδών dashboards: ένα strategic dashboard για το Υπουργείο Οικονομικών μιας χώρας με στόχο την παρακολούθηση της εξέλιξης των τιμών και των ενοικιάσεων στις πόλεις της χώρας ανά τον χρόνο κι ένα operational dashboard για real-estate επενδυτές που αγοράζουν και πουλούν ακίνητα προς χρήση στην πλατφόρμα Airbnb. Συνεπώς και για τους τελευταίους, είναι σημαντική η καταγραφή της διακύμανσης των τιμών αλλά και ο εντοπισμός ευκαιριών.

Συγκεκριμένα, η πτυχιακή εργασία είναι δομημένη ως εξής:

- Στο κεφάλαιο 1 γίνεται επεξήγηση του Data Warehouse και των βασικών εννοιών του
- Στο κεφάλαιο 2 αναλύεται η έννοια του Business Intelligence
- Στο κεφάλαιο 3 αναλύεται ο τύπος των dashboards και των διαγραμμάτων
- Στο κεφάλαιο 4 επεξηγείται η σχεδίαση και η υλοποίηση του DW
- Στο κεφάλαιο 5 γίνεται η σχεδίαση και η υλοποίηση στο Tableau

- Στο κεφάλαιο 6 πραγματοποιείται μια περιεκτική παρουσίαση των συμπερασμάτων και ο επίλογος

ΚΕΦΑΛΑΙΟ 1: Data Warehouse

Μέρος της παρούσας εργασίας αποτελεί η συγκέντρωση, ο συντονισμός και η ανάλυση τεράστιων ποσοτήτων δεδομένων από την πλατφόρμα της Airbnb. Επομένως, η επιλογή μιας αρχιτεκτονικής Data Warehouse είναι αναπόφευκτη, σε αντίθεση με τις σχεσιακές βάσεις δεδομένων που εξυπηρετούν στη διαχείριση και ταχεία ανάκτηση δεδομένων για τις καθημερινές λειτουργίες μιας εφαρμογής.

Σε γενικές γραμμές, μια ΒΔ επικεντρώνεται στην ενημέρωση δεδομένων σε πραγματικό χρόνο, ενώ ένα DW έχει ευρύτερο πεδίο εφαρμογής, καταγράφοντας τρέχοντα και ιστορικά δεδομένα για προγνωστική ανάλυση, μηχανική μάθηση και άλλους προηγμένους τύπους ανάλυσης.

1.1 Εισαγωγή

Το DW αποτελεί ένα εξειδικευμένου τύπου data management σύστημα, που προορίζεται για επεξεργασία και ανάλυση ενοποιημένων δεδομένων (Widom, 1995a). Σκοπός του είναι η ενσωμάτωση δεδομένων από πολλαπλές, ενδεχομένως πολύ μεγάλες, κατανεμημένες, ετερογενείς βάσεις δεδομένων και άλλες πηγές πληροφοριών (Widom, 1995b). Έχει σχεδιαστεί δηλαδή για να επιτρέπει και να υποστηρίζει δραστηριότητες BI. Αυτή η συνολική προσέγγιση στη διαχείριση δεδομένων ενισχύει την ικανότητα των οργανώσεων να προβλέπουν τάσεις, να αντιλαμβάνονται τη συμπεριφορά της αγοράς και να λαμβάνουν ουσιαστικές αποφάσεις.

Κατά αυτήν την προσέγγιση, αυτές είναι οι δυο εναλλακτικές που ακολουθούνται:

- Επιλεγμένες πληροφορίες από κάθε πηγή εξάγονται εκ των προτέρων, μεταφράζονται και φιλτράρονται ως απαιτείται, συγχωνεύονται με σχετικές πληροφορίες από άλλες πηγές και αποθηκεύονται στο DW.
- Οι πληροφορίες αξιολογούνται απευθείας στο DW με τη χρήση queries, χωρίς πρόσβαση στις αρχικές πηγές.

Γενικά όταν προκύπτουν αλλαγές στα δεδομένα, αυτές πρέπει να γίνονται εξίσου και στο DW. Υπάρχουν διαφορετικές τακτικές που μπορούν να εφαρμοστούν γι' αυτό, οι οποίες εξαρτώνται σε μεγάλο βαθμό κι από τις ανάγκες του πελάτη. Οι τεχνικές σταδιακής ενημέρωσης μπορεί να είναι πιο αποτελεσματικές, ειδικά όταν επηρεάζουν μόνο ένα μικρό μέρος της ΒΔ, όπως συμβαίνει συνήθως. Μια βασική τεχνική είναι η εφαρμογή των SQL queries.

Πλεονεκτήματα του DW¹:

- Καλύτερη ποιότητα δεδομένων
- Γρηγορότερα προσβάσιμες πληροφορίες
- Εξυπνότερες αποφάσεις για τις επιχειρήσεις
- Αύξηση του ανταγωνιστικού πλεονεκτήματος του οργανισμού

1.1.1 Στόχοι του Data Warehouse

Οι 6 κύριοι στόχοι του DW (Kimball and Ross, 2011):

- Το DW να καθιστά τις πληροφορίες ενός οργανισμού εύκολα προσβάσιμες
- Το DW να είναι αξιόπιστο και να παρουσιάζει με συνέπεια τις πληροφορίες του οργανισμού
- Το DW να είναι προσαρμοστικό και ανθεκτικό στις αλλαγές
- Το DW να είναι ασφαλές και να προστατεύει τα στοιχεία των πληροφοριών
- Το DW να χρησιμεύει ως βάση για βελτιωμένες αποφάσεις
- Το DW να γίνεται αποδεκτό από την επιχειρηματική κοινότητα εάν πρόκειται να κριθεί επιτυχές

1.1.2 Τύποι του Data Warehouse

- **Cloud data warehouse:** σχεδιασμένο να λειτουργεί σε ένα περιβάλλον υπολογιστικού νέφους, αυτό το είδος DW αξιοποιεί τους πόρους και τις υπηρεσίες του cloud για να παρέχει αποθήκευση και δυνατότητες επεξεργασίας δεδομένων που είναι κλιμακούμενες, ευέλικτες και οικονομικά αποδοτικές, ιδιαίτερα για μεγάλους όγκους δεδομένων.
- **Data warehouse software (on-premises/license):** αποτελεί δαπανηρότερη επιλογή από το cloud, αφού οι επιχειρήσεις χρειάζεται να αγοράσουν άδεια και να υλοποίησουν τη δική τους υποδομή, όμως σε περιπτώσεις που χρειάζεται περισσότερος έλεγχος ή συμμόρφωση με πρότυπα ασφαλείας αποτελεί την καλύτερη επιλογή.
- **Data warehouse appliance:** είναι ένα προ-ενσωματωμένο πακέτο υλικού και λογισμικού (CPUs, αποθήκευση, λειτουργικό σύστημα και λογισμικό DW) που μια επιχείρηση μπορεί να συνδέσει στο δίκτυό της και να ξεκινήσει να χρησιμοποιεί όπως είναι (Kimball and Ross, 2011).

¹ [What Is a Data Warehouse | Oracle](#)

1.2 Αρχιτεκτονική

Συνήθως, η αρχιτεκτονική ενός DW αποτελείται από τρία βασικά στρώματα (Vishesh et al., 2017). Σπανιότερα μπορεί να αποτελείται από ένα ή δύο στρώματα:

- **Bottom Tier**

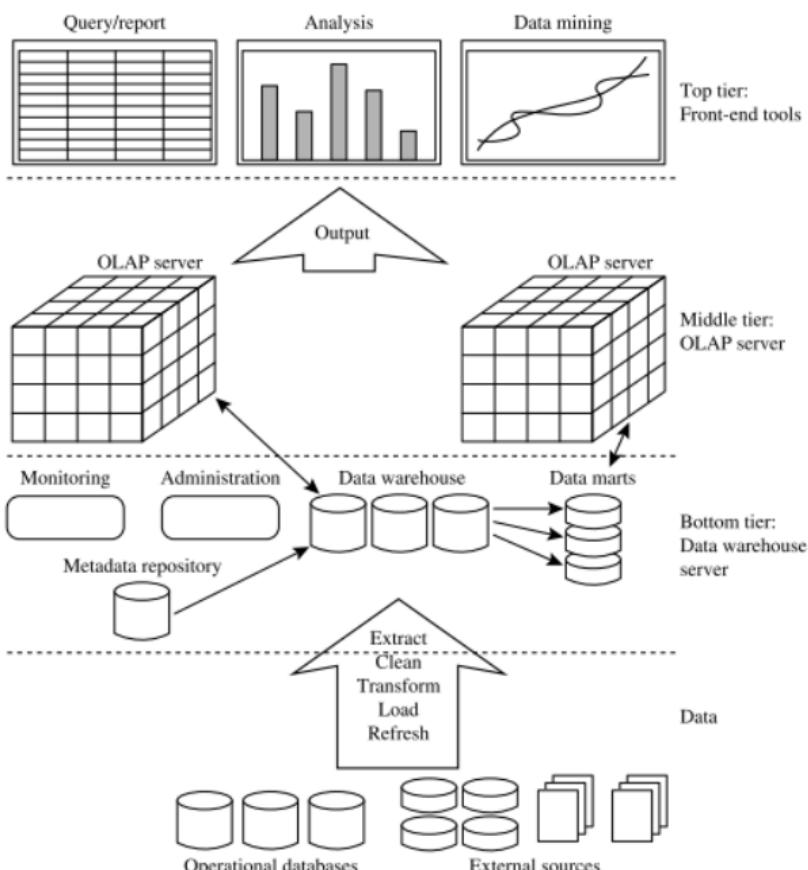
Αποτελείται από έναν server, συνήθως μια σχεσιακή ΒΔ, ο οποίος συγκεντρώνει, καθαρίζει και μετατρέπει τα δεδομένα από πολλαπλές πηγές.

- **Middle Tier**

Αποτελείται από έναν OLAP server, ο οποίος επιτρέπει τις υψηλές ταχύτητες στα queries. Σε αυτό το επίπεδο μπορούν να χρησιμοποιηθούν τρία είδη μοντέλων OLAP, τα οποία είναι γνωστά ως ROLAP, MOLAP και HOLAP. Ο τύπος του μοντέλου OLAP που χρησιμοποιείται εξαρτάται από τον τύπο της ΒΔ που υπάρχει.

- **Top Tier**

Αντιπροσωπεύεται από ένα front-end user interface ή reporting εργαλείο, που επιτρέπει στον χρήστη να διεξάγει ανάλυση δεδομένων στα δικά του επιχειρησιακά δεδομένα



Εικ.1: Three tier αρχιτεκτονική ('Architecture of the BI support technology infrastructure', Bernardino, Ferreira και Pedrosa, 2017)

1.3 Components

Ο σχεδιασμός ενός DW αποτελείται από 6 κύρια components (Jameel et al., 2022):

- **ETL**

Το ETL βρίσκεται στο Bottom Tier σε σχέση με την αρχιτεκονική που αναλύθηκε στο προηγούμενο υποκεφάλαιο. Είναι η διαδικασία που ακολουθείται από έναν αναλυτή δεδομένων όταν πρόκειται να μεταφέρει δεδομένα από μια πηγή στο DW. Αποτελεί τα αρχικά του Extract, Transform και Load, που αποδεικνύει ακριβώς τα βήματα που ακολουθούνται. Ουσιαστικά, η ETL διαδικασία μεταφέρει τα data σε ένα νέο σχήμα κατάλληλο για επεξεργασία και ανάλυση.

Πιο συγκεκριμένα, και για τα πλαίσια αυτής της εργασίας, όταν η Airbnb θα διαθέσει στην πλατφόρμα της τα επόμενα, ανανεωμένα CSVs, η διαδικασία ETL θα λειτουργήσει ως εξής:

- *Εξαγωγή (Extract)*: Πρώτο βήμα είναι η εξαγωγή των νέων δεδομένων από την Airbnb, με τη λήψη των κατάλληλων CSVs.
- *Μετασχηματισμός (Transform)*: Τα νέα δεδομένα που εξάγονται από την Airbnb πρέπει να υποστούν διάφορες μετασχηματιστικές διαδικασίες. Αυτό μπορεί να περιλαμβάνει την ενσωμάτωσή τους στους κατάλληλους πίνακες του DW, τον έλεγχο για διπλότυπες καταχωρήσεις, τον υπολογισμό ενδεχόμενων νέων μετρικών ή οποιαδήποτε άλλη επεξεργασία απαιτείται για την προσαρμογή των δεδομένων στις ανάγκες του συστήματος.
- *Φόρτωση (Load)*: Τα μετασχηματισμένα δεδομένα φορτώνονται στη βάση δεδομένων του DW, με τις απαραίτητες SQL εντολές.

Αφού τα νέα δεδομένα φορτωθούν με επιτυχία στη βάση δεδομένων, το Tableau, το οπόιο είναι συνδεδεμένο με τη βάση δεδομένων, ανανεώνεται αυτόματα για να αντικατοπτρίζει τα νέα δεδομένα. Ο χρήστης μπορεί να έχει πρόσβαση στα dashboards και να δει τις πιο πρόσφατες πληροφορίες χωρίς να απαιτείται κάποια επιπλέον παρέμβαση από τον αναλυτή δεδομένων.

- **Metadata**

Τα μεταδεδομένα είναι δεδομένα σχετικά με άλλα δεδομένα. Περιγράφουν όλα τα data που αποθηκεύονται σε ένα σύστημα, με σκοπό να τα καθιστά αναζητήσιμα. Είναι ουσιαστικά σαν τίτλοι μιας στήλης σε ένα φύλλο καταγραφής. Τα metadata βοηθούν στην οργάνωση και συντελούν στη δημιουργία πινάκων ελέγχου και αναφορών.

- **Επεξεργασία ερωτημάτων SQL**

Η SQL αποτελεί τη γλώσσα που χρησιμοποιούν οι αναλυτές για την ανάκτηση δεδομένων, χρησιμοποιώντας προηγμένες τεχνολογίες επεξεργασίας SQL που συνδέονται στενά με τον υπολογιστικό τομέα. Αυτό επιτρέπει την επίτευξη υψηλής απόδοσης κατά τη διαδικασία της ανάλυσης.

- **Επίπεδο δεδομένων (data layer)**

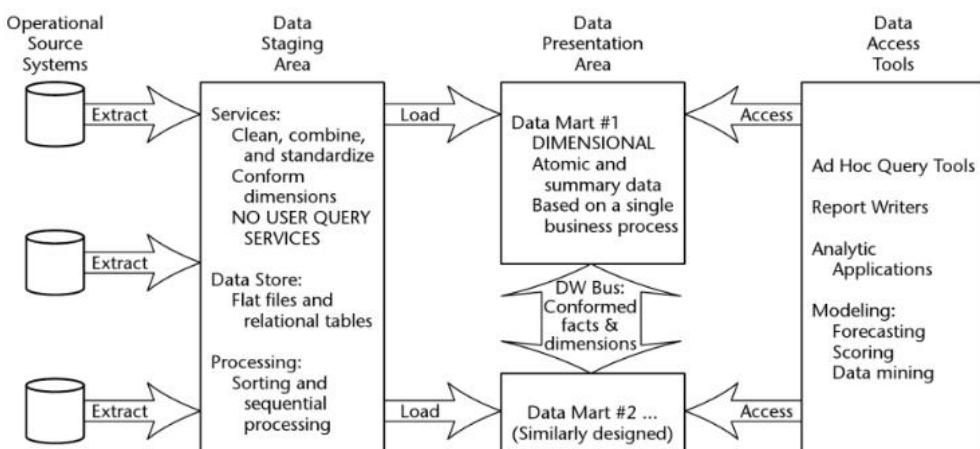
Αποτελεί το επίπεδο πρόσβασης που επιτρέπει στους χρήστες να έχουν πραγματική πρόσβαση στα δεδομένα. Συνήθως, εδώ βρίσκεται ένα data mart (υποσύνολο ενός DW που περιέχει δεδομένα ειδικά για ένα συγκεκριμένο επιχειρηματικό τμήμα). Αυτό το επίπεδο διαχωρίζει τμήματα των δεδομένων, επιτρέποντας πολύ λεπτομερείς ρυθμίσεις σε όλο τον οργανισμό.

- **Διακυβέρνηση και Ασφάλεια**

Αυτό συνδέεται με το προηγούμενο component, καθώς ο οργανισμός χρειάζεται να είναι σε θέση να παρέχει λεπτομερείς πολιτικές πρόσβασης και ασφαλείας σε όλα τα δεδομένα του. Συνήθως, τα DW διαθέτουν εξαιρετικές ενσωματωμένες δυνατότητες διακυβέρνησης και ασφάλειας.

- **DW εργαλεία πρόσβασης**

Ενώ τα εργαλεία αυτά θεωρούνται εξωτερικά του DW, θεωρούνται φιλικές προς τον χρήστη προσθήκες του. Αυτά μπορεί να είναι εργαλεία αναφοράς και οπτικοποίησης, τα οποία χρησιμοποιούνται από τους αναλυτές για να αλληλεπιδρούν με τα δεδομένα και να δημιουργούν διαγράμματα που μπορούν να αξιοποιηθούν. Χαρακτηριστικό παράδειγμα αποτελεί το Tableau, το οποίο χρησιμοποιείται και στην παρούσα πτυχιακή.



Εικ.2: Components ('Basic elements of the data warehouse', Kimball and Ross, 2011)

1.4 OLAP & OLTP

Όπως περιγράφεται και στην παράγραφο 1.2, το Middle Tier του Data Warehouse περιλαμβάνει έναν OLAP server (Online Analytical Processing). Αυτός ο OLAP server επιτρέπει την εκτέλεση πολυδιάστατης ανάλυσης με υψηλές ταχύτητες, ιδίως για μεγάλους όγκους δεδομένων. Η λειτουργία του συνίσταται στη δυνατότητα εκτέλεσης πραγματικού χρόνου ανάλυσης για ένα μεγάλο αριθμό συναλλαγών από πολλά άτομα. Συνήθεις χρήσεις περιλαμβάνουν την εξόρυξη δεδομένων, σύνθετους αναλυτικούς υπολογισμούς και προγραμματισμό προβλέψεων (Gupta, Harinarayan και Quass, 1995).

Ωστόσο, σε πολλούς οργανισμούς υπάρχει και διαφορετικό λογισμικό επεξεργασίας δεδομένων, το οποίο επικεντρώνεται στη διαχείριση και εκτέλεση των καθημερινών συναλλαγών μιας επιχείρησης σε πραγματικό χρόνο, γνωστό ως OLTP (Online Transaction Processing). Όπως υποδηλώνει και το όνομά του, το OLTP υποστηρίζει τη γρήγορη, αποτελεσματική και ασφαλή εκτέλεση των συναλλαγών, όπως παραγγελίες σε πραγματικό χρόνο, πληρωμές, ενημερώσεις και διαγραφές.

1.5 Schema

Tα schemes είναι τρόποι με τους οποίους τα δεδομένα οργανώνονται σε μια ΒΔ (Lechtenbörger, 2001). Υπάρχουν δύο κύριοι τύποι δομών, το star schema και το snowflake schema.

- **Star schema**

Αυτή η προσέγγιση αποτελείται από έναν fact table, ο οποίος μπορεί να ενωθεί με έναν αριθμό αποκανονικοποιημένων dimension tables. Θεωρείται ο πιο απλός και κοινός τύπος, αφού καθιστά την απόδοση των queries πολύ γρήγορη.



Εικ.3: Star schema ('Measured facts in the retail sales schema', Kimball and Ross, 2011)

- **Snowflake schema**

Εδώ ο fact table συνδέεται με έναν αριθμό κανονικοποιημένων dimension tables, οι οποίοι με τη σειρά τους έχουν θυγατρικούς πίνακες. Δεν είναι τόσο ευρέως υιοθετημένο αλλά μπορεί κανείς να επωφεληθεί από τα χαμηλά επίπεδα

πλεονασμού δεδομένων. Παρόλα αυτά υφίσταται κόστος για την απόδοση των queries (Lechtenbörger, 2001).

1.6 Fact table

Ένας fact table είναι ο κεντρικός πίνακας σε ένα DW, που περιέχει δεδομένα μέτρησης, συχνά αριθμητικά, που αντιπροσωπεύουν κάποιο είδος επιχειρησιακού γεγονότος. Είναι ουσιαστικά οι πίνακες που χρησιμοποιούνται κατά την ανάλυση και τη δημιουργία αναφορών, όπως μαρτυρεί και το όνομά του: πίνακας γεγονότων (Kimball and Ross, 2011).

Daily Sales Fact Table
Date Key (FK)
Product Key (FK)
Store Key (FK)
Quantity Sold
Dollar Sales Amount

Εικ.4: Παράδειγμα fact table ('Sample fact table', Kimball and Ross, 2011)

Όπως αναφέρθηκε και στην παρ. 1.5, ο fact table συνδέεται με διαστάσεις, που αυτές αντιπροσωπεύουν τους τρόπους με τους οποίους θέλουμε να κατηγοριοποιήσουμε ή να αναλύσουμε τα δεδομένα. Για τον λόγο αυτόν οι fact tables συνήθως περιλαμβάνουν ξένα κλειδιά (foreign keys) που συνδέονται με τις διαστάσεις για την ανάκτηση σχετικών πληροφοριών.

1.7 Dimension tables

Όπως εξηγήθηκε και παραπάνω, οι dimension tables αποτελούν αναπόσπαστες σχέσεις του fact table. Περιέχουν πληροφορίες για τις διαστάσεις ή τους τρόπους κατηγοριοποίησης του. Μπορούν να χαρακτηριστούν ως περιγραφικοί, αφού προσθέτουν συνεκτικότητα στα δεδομένα του fact table, επιτρέποντας στους χρήστες να αναλύουν και να κατανοούν τις επιχειρησιακές δραστηριότητες με πιο εξελιγμένο τρόπο. Οι dimension tables περιλαμβάνουν επίσης ένα πρωτεύον κλειδί (primary key), που αναπαριστά το μοναδικό αναγνωριστικό για κάθε εγγραφή στους πίνακες αυτούς (Kimball and Ross, 2011).

Product Dimension Table
Product Key (PK)
Product Description
SKU Number (Natural Key)
Brand Description
Category Description
Department Description
Package Type Description
Package Size
Fat Content Description
Diet Type Description
Weight
Weight Units of Measure
Storage Type
Shelf Life Type
Shelf Width
Shelf Height
Shelf Depth
... and many more

Εικ.5: Παράδειγμα dimension table ('Sample dimension table', Kimball and Ross, 2011)

1.8 Σχέση μεταξύ fact & dimension tables

Η σχέση μεταξύ fact table και dimension tables αποτελεί το κεντρικό στοιχείο στον σχεδιασμό ενός συστήματος DW. Το πιο σημαντικό είναι ότι συνδέονται μέσω κλειδιών. Συγκεκριμένα, το primary key στο dimension table συνδέεται με το foreign key στο fact table, που περιέχει τα μετρήσιμα γεγονότα. Αυτή η σχέση επιτρέπει τη συνδεσιμότητα και τη συσχέτιση των δεδομένων μεταξύ των διαστάσεων και των γεγονότων (Kimball and Ross, 2011).



Εικ.6: Σχέση μεταξύ fact & dimensional tables ('Measured facts in the retail sales schema', Kimball and Ross, 2011)

1.9 Primary key

Το πρωτεύον κλειδί σε μια ΒΔ είναι ένα μοναδικό αναγνωριστικό που αντιστοιχεί σε κάθε εγγραφή σε έναν πίνακα. Η μοναδικότητα του εξασφαλίζεται, επιτρέποντας την αναζήτηση και την ταυτοποίηση δεδομένων. Συχνά, είναι αυτόματα αυξομειούμενο και χρησιμοποιείται ως σύνδεσμος σε άλλους πίνακες. Αποτελεί κρίσιμο στοιχείο για τη διαχείριση και ανάκτηση δεδομένων, καθιστώντας το απαραίτητο για την ορθή λειτουργία της βάσης δεδομένων (Thalheim, 1987).

1.10 Foreign key

Το ξένο κλειδί σε μια βάση δεδομένων είναι ένα σημαντικό στοιχείο που επιτρέπει τη δημιουργία σχέσεων μεταξύ διάφορων πινάκων. Αυτό το κλειδί λειτουργεί σαν γέφυρα μεταξύ πινάκων, συνδέοντας ένα πεδίο σε έναν πίνακα με το πρωτεύον κλειδί ενός άλλου. Με αυτόν τον τρόπο, επιτυγχάνεται η συσχέτιση δεδομένων μεταξύ διαφορετικών πινάκων, εξασφαλίζοντας τη συνέπεια και την ακεραιότητα των δεδομένων (Thalheim, 1987).

ΚΕΦΑΛΑΙΟ 2: Business Intelligence

2.1 Εισαγωγή

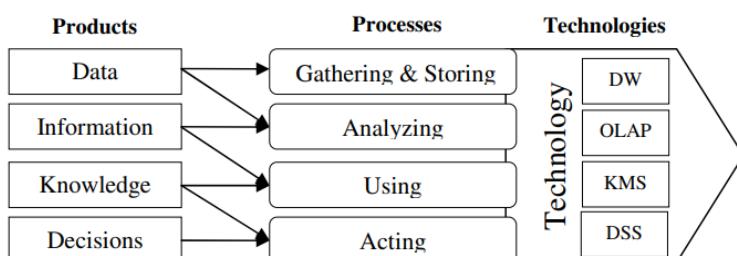
Σύμφωνα με τον πατέρα της επιχειρηματικής νοημοσύνης, Howard Dresner, ο όρος BI αποτελεί μια ευρεία κατηγορία λογισμικού και λύσεων για συγκέντρωση, ενοποίηση, ανάλυση και παροχή πρόσβασης σε δεδομένα με τρόπο που επιτρέπει στους εταιρικούς χρήστες να λαμβάνουν καλύτερες επιχειρηματικές αποφάσεις (Sharda, Delen and Turban, 2007).

Ανά τα χρόνια διάφοροι επιστήμονες προσπάθησαν να τελειοποιήσουν τον ορισμό αυτόν, όπως ο Luhn, ωστόσο η κοινότητα έχει καταλήξει στις εξής τρεις προσεγγίσεις (Shollo and Kautz, 2010):

Η προσέγγιση με εστίαση στη **διαδικασία** επικεντρώνεται στη συστηματική συλλογή δεδομένων από διάφορες εσωτερικές και εξωτερικές πηγές. Αυτές οι πηγές μπορεί να περιλαμβάνουν εταιρικά συστήματα, δεδομένα από την αγορά, κοινωνικά μέσα και άλλες πλατφόρμες. Ο στόχος είναι η ανάλυση αυτών των δεδομένων με σκοπό τη δημιουργία πληροφοριών που θα χρησιμοποιηθούν για τη βελτίωση της διαδικασίας λήψης αποφάσεων. Οι συγγραφείς που έχουν ασχοληθεί με αυτό το θέμα περιλαμβάνουν τους Whitehorn, Moss & Hoberman, Hostmann, Turban, Markarian και Brobst & Bedell.

Η **τεχνολογική** προσέγγιση εστιάζει στα εργαλεία, τις τεχνολογίες και τις μεθόδους που χρησιμοποιούνται για την ανάπτυξη, εφαρμογή και εκτέλεση ενός συστήματος ή μιας διαδικασίας. Συχνά περιλαμβάνει τους υπολογιστές, το λογισμικό, δίκτυα και άλλες τεχνολογικές υποδομές που απαιτούνται για την υλοποίηση μιας συγκεκριμένης λειτουργίας ή διαδικασίας. Οι συγγραφείς Moss & Atre, Adelman & Moss, Turban και Horstman εξετάζουν αυτήν την πτυχή.

Η περιγραφή του BI ως **προϊόν** αναδύεται από μια λεπτομερή και εμβαθυμένη ανάλυση των επιχειρηματικών δεδομένων, που έχει ως αποτέλεσμα τη δυνατότητα εξαγωγής συμπερασμάτων, προβλέψεων και ενδεχόμενων μελλοντικών τάσεων. Επίσης, αναφέρεται στις πρακτικές αναλύσεις που πραγματοποιούνται χρησιμοποιώντας BI εργαλεία. Οι Chang, Gangadharan & Swami, Kulkarni & King, Turban είναι τα κύρια πρόσωπα αυτής της θεωρίας.



Εικ.7: Προσεγγίσεις BI ('Conceptualizing BI', Shollo and Kautz, 2010)

2.1.1 «Προσωποποιημένη» Επιχειρηματική Ευφυΐα

Στη σημερινή εποχή, η τάση στον χώρο της ανάλυσης δεδομένων είναι να δίνεται μεγαλύτερη αυτονομία στους αναλυτές δεδομένων, απελευθερώνοντάς τους από την ανάγκη συνεχούς συνεργασίας με την κεντρική διεύθυνση μηχανογραφημένων εφαρμογών. Αυτό δεν σημαίνει ότι οι υπολογιστικές εφαρμογές δεν χρησιμοποιούνται για την επεξεργασία και ανάλυση δεδομένων για τη λήψη αποφάσεων, αλλά ότι οι αναλυτές δεδομένων αποκτούν μεγαλύτερη ελευθερία και ευελιξία (Στρατής, 2023).

Σε αυτό το νέο περιβάλλον, οι αναλυτές δεδομένων απαιτούν πλέον ευρύτερο φάσμα γνώσεων και δεξιοτήτων. Εκτός από τις παραδοσιακές δεξιότητες πληροφορικής και διαχείρισης βάσεων δεδομένων, απαιτείται και η κατανόηση των μαθηματικών και της στατιστικής. Επιπλέον, είναι αναγκαία η ικανότητα κριτικής σκέψης όσον αφορά τα δεδομένα που επεξεργάζονται, καθώς και η κατανόηση του τρόπου λειτουργίας της επιχειρησης ή υπηρεσίας που εξυπηρετούν.

2.2 Πλεονεκτήματα

Τα συστήματα Business Intelligence επιτρέπουν στην εταιρεία να παίρνει αποφάσεις με μεγαλύτερη αποτελεσματικότητα και βεβαιότητα, ενώ ταυτόχρονα συνδράμουν στον εντοπισμό και την αξιοποίηση των ευκαιριών που προσφέρει η εταιρεία. Επιπλέον, συντελούν στη μείωση των προσπαθειών στον τομέα του μάρκετινγκ, στη βελτίωση των σχέσεων με τους πελάτες, στον αρμονικό συντονισμό των προσπαθειών με τη στρατηγική της εταιρείας και στην ενίσχυση των εσόδων και των κερδών (Williams & Williams, 2010). Επίσης, τα συστήματα Business Intelligence βοηθούν τις επιχειρήσεις να αξιολογήσουν ποσοτικά την αξία των σχέσεων τους με προμηθευτές και πελάτες, δίνοντάς τους επιπλέον επιρροή κατά τη διάρκεια των διαπραγματεύσεων. Χωρίς την υιοθέτηση της επιχειρηματικής νοημοσύνης, μια εταιρεία αντιμετωπίζει τον κίνδυνο να λαμβάνει κρίσιμες αποφάσεις βασισμένες είτε σε ανεπαρκείς είτε σε εσφαλμένες πληροφορίες.

Σε γενικές γραμμές, μια καλά υλοποιημένη στρατηγική BI βοηθά τις εταιρείες με τους εξής τρόπους:

- Γρήγορη ανίχνευση και αντίδραση σε τάσεις του επιχειρηματικού περιβάλλοντος
- Ενδυνάμωση του προσωπικού με χρήσιμες πληροφορίες και αναφορές που αφορούν τις τάσεις
- Δημιουργία λεπτομερών αναφορών για το οικονομικό, τις λειτουργίες, τους πελάτες και τους προμηθευτές
- Εύκολη προβολή, χειρισμός, ανάλυση και διανομή αναφορών με τη χρήση διάφορων καταξιωμένων εργαλείων

- Ανάκτηση στιγμιαίων συνολικών συνόψεων, ομάδων λογαριασμών ή λεπτομερών συναλλαγών
- Συγχώνευση δεδομένων από πολλαπλές εταιρείες, τμήματα και βάσεις δεδομένων
- Ελαχιστοποίηση της ανθρώπινης εργασίας και των επαναλαμβανόμενων διαδικασιών

2.3 Μειονεκτήματα

Σύμφωνα με έρευνα της εταιρείας Garner Analytics, το 70% έως το 80% των επιχειρηματικών έργων Business Intelligence (BI) αντιμετωπίζουν αποτυχία. Οι εταιρείες αντιμετωπίζουν πολλές προκλήσεις, με τις κυριότερες να είναι η αντίσταση των χρηστών για την υιοθέτηση των στρατηγικών BI και η κακή ποιότητα δεδομένων (Djerdjouri, 2020).

Συγκεκριμένα, η αντίσταση των χρηστών για την υιοθέτηση της επιχειρηματικής νοημοσύνης αποτελεί σημαντικό μειονέκτημα κατά την ανάπτυξη και υλοποίηση μιας στρατηγικής BI. Η αρνητική αντίδραση των χρηστών μπορεί να αποτελέσει εμπόδιο για την επιτυχή υιοθέτηση του συστήματος, καθώς οι χρήστες συχνά δυσκολεύονται να αποχωριστούν τις υφιστάμενες διαδικασίες και να προσαρμοστούν σε νέες τεχνολογίες. Η παρούσα αντίσταση μπορεί να οδηγήσει σε χαμηλά επίπεδα χρήσης του BI και να μειώσει την αποτελεσματικότητα του συστήματος, καθώς οι χρήστες ενδέχεται να μην αξιοποιούν πλήρως τα διαθέσιμα εργαλεία και λειτουργίες. Επιπλέον, η παραμονή των χρηστών εκτός του νέου συστήματος μπορεί να οδηγήσει σε απώλεια ευκαιριών για βελτιώσεις και καινοτομίες που προσφέρει το BI.

Εξίσου μεγάλη πρόκληση αποτελεί η κακή ποιότητα των δεδομένων. Χωρίς την αξιόπιστη συλλογή, αποθήκευση και πρόσβαση σε εγκυρα δεδομένα, η εταιρεία είναι ανίκανη να αποκτήσει ουσιαστική και ακριβή κατανόηση της επιχείρησής της και του επιχειρηματικού περιβάλλοντός της. Τα δεδομένα αποτελούν τον πυρήνα κάθε συστήματος BI και η πρόκληση εδώ είναι για την εταιρεία να διασφαλίσει την εύρυθμη λειτουργία των συστημάτων αποθήκευσης δεδομένων προτού αρχίσει την ανάκτηση και την αξιοποίηση των ερευνών. Το ρίσκο που διατρέχει είναι ότι, εάν αυτό δεν γίνει σωστά, θα ληφθούν κρίσιμες αποφάσεις βασιζόμενες σε αναξιόπιστες πληροφορίες. Η εταιρεία πρέπει, συνεπώς, να επιδιώξει και να διατηρήσει υψηλό επίπεδο ποιότητας δεδομένων για να διασφαλίσει την επιτυχή λειτουργία του συστήματος BI.

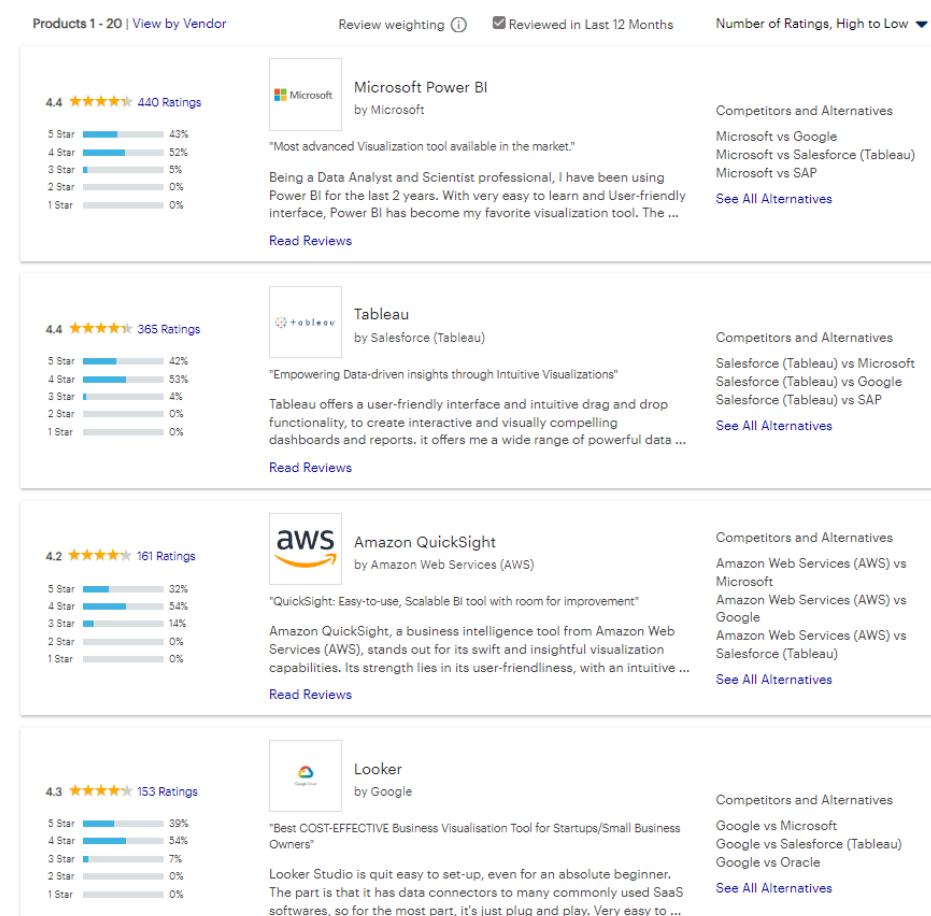
Άλλες προκλήσεις περιλαμβάνουν τον αντικατακλυσμό των τμηματικών γνωστικών σιλοκατασκευών, την ενσωμάτωση των εργαλείων BI με άλλα λειτουργικά, συστήματα διαχείρισης απόδοσης και συναλλαγών, τον μετασχηματισμό του χώρου εργασίας από μια κουλτούρα βασισμένη στην εμπειρία σε μια άλλη βασισμένη σε δεδομένα λήψη αποφάσεων, καθώς και την εξασφάλιση

της υποστήριξης και της απαραίτητης οικονομικής ενίσχυσης από την εκτελεστική γηγεσία.

2.4 Gartner Magic Quadrant BI tools

Το Magic Quadrant (MQ) είναι μια σειρά ερευνητικών αναφορών που δημοσιεύονται από την συμβουλευτική εταιρεία, Gartner. Αυτές οι αναφορές βασίζονται σε αποκλειστικές μεθόδους ποιοτικής ανάλυσης δεδομένων για την απεικόνιση των τάσεων της αγοράς. Οι αναλύσεις αυτές πραγματοποιούνται για διάφορους κλάδους τεχνολογίας και ενημερώνονται κάθε χρόνο. Η λίστα αυτή ενδέχεται να απαιτεί επικαιροποίηση σε τακτά χρονικά διαστήματα, καθώς κατά καιρούς κάποιοι εισέρχονται στην αγορά ενώ σταδιακά άλλοι αποχωρούν.

Μεταβαίνοντας στη σελίδα του οργανισμού², συναντάμε μια έρευνα σχετικά με τη δημοφιλία των BI πλατφορμών στην αγορά. Φίλτραροντας τα ratings των τελευταίων 12 μηνών, παίρνουμε τις εξής κορυφαίες πλατφόρμες:



Εικ. 8: Gartner Magic Quadrant BI tools ([Best Analytics and Business Intelligence Platforms Reviews 2023 | Gartner Peer Insights](#))

² [Best Analytics and Business Intelligence Platforms Reviews 2023 | Gartner Peer Insights](#)

ΚΕΦΑΛΑΙΟ 3: Οπτικοποίηση δεδομένων

3.1 Εισαγωγή

Η οπτικοποίηση των δεδομένων είναι ένα ισχυρό εργαλείο που χρησιμοποιεί γραφικές αναπαραστάσεις για να καθιστά πιο κατανοητά τα δεδομένα. Αντί να επικεντρώνεται σε αριθμητικές τιμές, προσφέρει μια ευανάγνωστη εικόνα των μοτίβων, των τάσεων και των συσχετίσεων προσιτή προς όλους. Ιδιαίτερα στις μέρες μας που η πληροφορία μας κατακλύζει περισσότερο από ποτέ, η αποτελεσματική αναπαράσταση της αποτελεί μονόδρομο.

3.2 Dashboards

Ένα dashboard αντιπροσωπεύει ένα γραφικό περιβάλλον που προβάλλει και οργανώνει κατανοητά τα κύρια δεδομένα και τις μετρικές για μια επιχείρηση, ένα έργο κλπ. Είναι ένα σημαντικό εργαλείο διαχείρισης πληροφοριών.

Ένας πλήρης κύκλος ζωής των dashboards περιλαμβάνει τη δημιουργία, την ανάλυση (ερμηνεία και εξήγηση) και την επικοινωνία (δημιουργία ιστοριών). Βασικά τους στοιχεία αποτελούν τα εξής (Elias, 2012):

- Συγκεντρωτική προβολή δεδομένων
- Διαδραστικότητα
- Στοιχεία πρόβλεψης και τάσης
- Εξατομικευμένες ρυθμίσεις
- Ενσωματωμένες πηγές δεδομένων

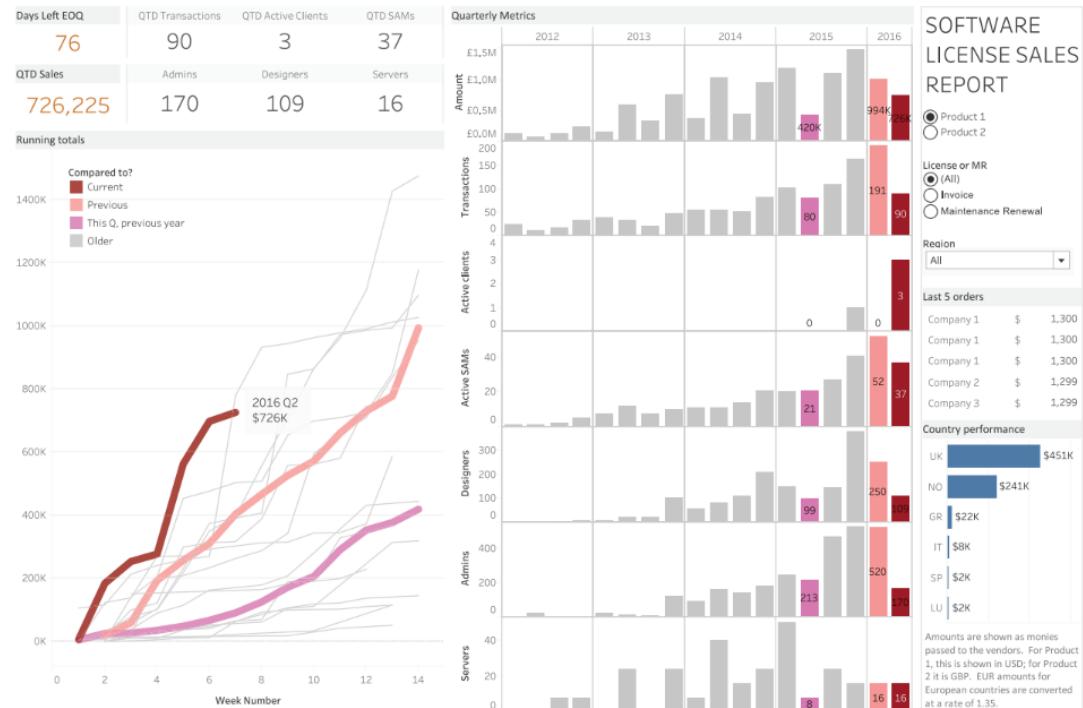
Υπάρχουν πολλά είδη dashboards, όμως αναλύονται τα παρακάτω βασικότερα τέσσερα είδη:

- **Στρατηγικά:** εστιάζουν σε μακροχρόνιες στρατηγικές και μετρήσεις υψηλού επιπέδου
- **Λειτουργικά:** εστιάζουν σε μικρότερα χρονικά πλαίσια και λειτουργικές διεργασίες
- **Αναλυτικά:** περιέχουν τεράστιο όγκο δεδομένων και προέρχονται από αναλυτές
- **Τακτικά:** χρησιμοποιούνται για την ιχνηλάτιση επίδοσης

3.2.1 Στρατηγικά dashboards

Τα στρατηγικά dashboards είναι εργαλεία οπτικοποίησης δεδομένων που σχεδιάζονται για να βοηθήσουν ανώτερα στελέχη και διαχειριστές να κατανοήσουν

την επίδοση, τις τάσεις και τα στρατηγικά στοιχεία της επιχείρησής τους. Παρέχουν μια συνολική, ενοποιημένη άποψη των κύριων μετρικών και δεικτών που αντικατοπτρίζουν τους στόχους και την επίδοση της επιχείρησης. Τα δεδομένα στρατηγικών dashboard δεν αλλάζουν συχνά, καθώς δεν είναι σε πραγματικό χρόνο. Επίσης εφαρμόζονται σε ποικίλους τομείς, όπως ο έλεγχος της παραγωγής, οι πωλήσεις, η εξυπηρέτηση πελατών και οι ανθρώπινοι πόροι, που παρουσιάζουν σημαντικά δεδομένα για τη λήψη κρίσιμων αποφάσεων (Cotgreave, Shaffer και Wexler, 2017).



Εικ. 9: Στρατηγικό dashboard (Cotgreave, Shaffer και Wexler, 2017)

3.2.2 Λειτουργικά dashboards

Τα λειτουργικά dashboards αποτελούν εξίσου ισχυρά εργαλεία οπτικοποίησης δεδομένων που παρέχουν στους χρήστες μια πλήρη επισκόπηση των κύριων πληροφοριών και μετρικών τους. Η κύρια λειτουργία τους είναι να διευκολύνουν τη λήψη αποφάσεων με τρόπο γρήγορο και ευανάγνωστο. Παρουσιάζουν αναλυτικά τα δεδομένα με χρήση γραφικών αναπαραστάσεων, επιτρέποντας την ευελιξία και προσαρμογή στις ατομικές προτιμήσεις και εστιάζουν στην καθημερινή λειτουργία και απόδοση. Επιπλέον, υποστηρίζουν τόσο τη συνδυασμένη προβολή πολλαπλών στοιχείων όσο και την αλληλεπίδραση του χρήστη με τα δεδομένα για μια ολοκληρωμένη και εξατομικευμένη ανάλυση αντίστοιχα (Cotgreave, Shaffer και Wexler, 2017).

3.2.3 Αναλυτικά dashboards

Τα αναλυτικά dashboards αποτελούν εξειδικευμένα εργαλεία που επιτρέπουν τη λεπτομερή εξέταση και ανάλυση δεδομένων. Προσφέρουν εξελιγμένα γραφήματα, στατιστικά μοντέλα και εξατομικευμένες επιλογές, επιτρέποντας στους χρήστες να προβαίνουν σε προβλέψεις και να εξερευνούν τις λεπτομέρειες. Με τη δυνατότητα χρονοσειριακής ανάλυσης, ασχολούνται με την εξέλιξη των δεδομένων στον χρόνο, ενισχύοντας την κατανόηση των χρηστών για τις τάσεις και τις αλλαγές. Αυτά τα dashboards αποτελούν ισχυρό εργαλείο για τη λήψη εμβληματικών αποφάσεων και την ανάπτυξη συνολικής κατανόησης των δεδομένων (Cotgreave, Shaffer και Wexler, 2017).

3.2.4 Τακτικά dashboards

Τα τακτικά dashboards προσφέρουν συνοπτική και περιεκτική παρουσίαση των κύριων επιδόσεων και μετρήσεων μιας οργάνωσης ή συνήθως ενός τμήματος. Είναι σχεδιασμένα για γρήγορη και εύκολη κατανόηση, καθώς παρέχουν βασικές πληροφορίες που αφορούν την καθημερινή λειτουργία και τις τρέχουσες επιδόσεις. Αυτά τα dashboards συνήθως περιλαμβάνουν βασικά γραφήματα, διαγράμματα και μετρήσεις, παρουσιάζοντας τις πληροφορίες συνολικά, χωρίς να επικεντρώνονται σε λεπτομέρειες. Οι χρήστες των τακτικών dashboards είναι συνήθως επικεφαλείς επιχειρήσεων, που χρειάζονται γρήγορη επισκόπηση για την παρακολούθηση των βασικών επιδόσεων και τη λήψη στρατηγικών αποφάσεων (Cotgreave, Shaffer και Wexler, 2017).

3.2.5 Χρώματα σε dashboards

Τα χρώματα σε dashboards αποτελούν κρίσιμο στοιχείο για την αποτελεσματική οπτικοποίηση των δεδομένων και την επικοινωνία των πληροφοριών. Η χρήση κατάλληλων χρωμάτων μπορεί να ενισχύσει την κατανόηση, να προβάλει προτεραιότητες και να διευκολύνει τη λήψη αποφάσεων. Ορισμένες βασικές αρχές για τη χρήση χρωμάτων σε dashboards περιλαμβάνουν τη συνέπεια, την αντίθεση, τα ποσοστά, την κατηγοριοποίηση κλπ (Cotgreave, Shaffer και Wexler, 2017).

USE OF COLOR IN DATA VISUALIZATION

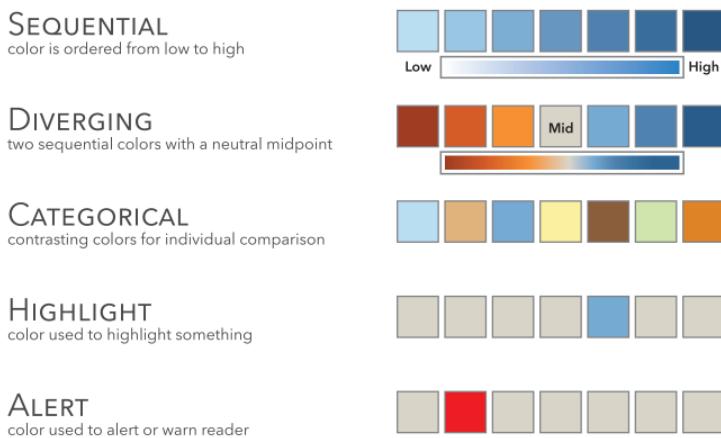


FIGURE 1.16 Use of color in data visualization.

Εικ. 10: Χρώματα σε dashboards (Cotgreave, Shaffer και Wexler, 2017)

Τα σειριακά χρώματα χρησιμοποιούν ένα μόνο χρώμα από ανοιχτό προς σκούρο για να απεικονίσουν μια κλίμακα. Τα διαφοροποιημένα χρώματα χρησιμοποιούνται για να αναδείξουν ένα εύρος που αποκλίνει από ένα κεντρικό σημείο. Τα κατηγορικά χρώματα διακρίνουν διάφορες κατηγορίες. Το χρώμα επισήμανσης χρησιμοποιείται για να ξεχωρίσει κάτι χωρίς να είναι έντονο. Το χρώμα ειδοποίησης χρησιμοποιείται όταν χρειάζεται να προσελκυστεί γρήγορα η προσοχή του αναγνώστη σε κάτι. Σε αυτήν την περίπτωση, είναι συχνά καλύτερο να χρησιμοποιούνται έντονα χρώματα.

3.2.6 Σύγκριση dashboards

Τα λειτουργικά dashboards παρέχουν βασικές πληροφορίες με μια σύντομη ματιά, κυρίως στους διαχειριστές οργανισμών, επιτρέποντας την άμεση κατανόηση των σημαντικότερων πτυχών. Τα στρατηγικά dashboards, από την άλλη πλευρά, είναι εξελιγμένα dashboards που παρέχουν πληροφορίες όπως ιστορικό και συγκρίσεις με στόχους. Τα αναλυτικά dashboards προσφέρουν λεπτομερείς αναφορές και είναι πιο κατάλληλα για εξειδικευμένη ανάλυση δεδομένων. Τα τακτικά dashboards παρέχουν τακτικές ενημερώσεις και παρακολούθηση καθημερινών δραστηριοτήτων (Eckerson, 2010):

	<i>Operational</i>	<i>Tactical</i>	<i>Strategic</i>
Purpose	Control operations	Optimize processes	Manage strategy
Scope	Operational	Departmental	Enterprise
Users	Staff	Managers	Executives
Primary activity	Act	Analyze	Review

Focus	Current	Past	Future
Data refresh	Daily	Daily/Weekly	Monthly
Information	Detailed	Detailed/Summary	Summary
Architecture	Core systems	Data Warehouse	Excel/Data mart
Metrics	Drivers	Drivers/Outcomes	Outcomes
Looks like...	Dashboard	Metrics portal	Scorecard

Πίν.1: Διαφορές dashboards (Eckerson, 2010)

3.3 Γραφήματα

Μέσω των γραφημάτων, μπορούμε να αναδείξουμε πρότυπα, τάσεις και συσχετίσεις στα δεδομένα, καθιστώντας πιο εύκολη την λήψη αποφάσεων και την επικοινωνία των αποτελεσμάτων. Από τα απλά διαγράμματα μπάρας και γραμμής έως τα πιο σύνθετα γραφήματα πίτας και διασποράς, η οπτική αναπαράσταση των δεδομένων σε γραφήματα ενισχύει την κατανόηση και την αποτελεσματική ερμηνεία της πληροφορίας.

Πιο συγκεκριμένα τα βασικότερα είναι τα παρακάτω (Knafllic, 2015). Τα τρία πρώτα δεν αποτελούν ακριβώς γραφήματα αλλά είναι συνήθη μέσα για την αναπαράσταση πληροφορίας. Από την άλλη τα γραφήματα αποτελούνται συνήθως από σημεία, γραμμές, bars και μια περιοχή (area). Επιπλέον τα pie charts και τα donut charts καλό θα ήταν να αποφεύγονται, καθώς παρουσιάζουν παραπλανητικά τα δεδομένα, όπως επίσης κάθε 3D γράφημα.

- **Απλό κείμενο:** Όταν ο αναλυτής έχει έναν ή δύο αριθμούς να μοιραστεί, η απλή κειμενική μορφή μπορεί να αποτελέσει έναν εξαιρετικό τρόπο επικοινωνίας. Χρησιμοποιείται αποκλειστικά ο αριθμός -καθιστώντας τον όσο το δυνατόν πιο έντονο- με λίγες υποστηρικτικές λέξεις για σαφήνεια.



Εικ. 11: Απλό κείμενο ('Simple text', Knafllic, 2015)

- **Πίνακας:** Ο πίνακας είναι ιδανικός για την επικοινωνία με ένα κοινό που κάθε μέλος του θα ψάξει για τη γραμμή του ενδιαφέροντός του. Εάν χρειάζεται να μεταφερθούν πολλές διαφορετικές μονάδες μέτρησης, αυτό

είναι συνήθως πιο εύκολο με έναν πίνακα παρά με ένα γράφημα. Η χρήση ενός πίνακα κατά τη διάρκεια μιας ζωντανής παρουσίασης είναι σπάνια μια καλή ιδέα, διότι χάνεται η προσοχή του κοινού. Επίσης τα δεδομένα πρέπει να ξεχωρίζουν, κι όχι τα όρια του πίνακα.

	A	B	C
Category 1	15%	22%	42%
Category 2	40%	36%	20%
Category 3	35%	17%	34%
Category 4	30%	29%	26%
Category 5	55%	30%	58%
Category 6	11%	25%	49%

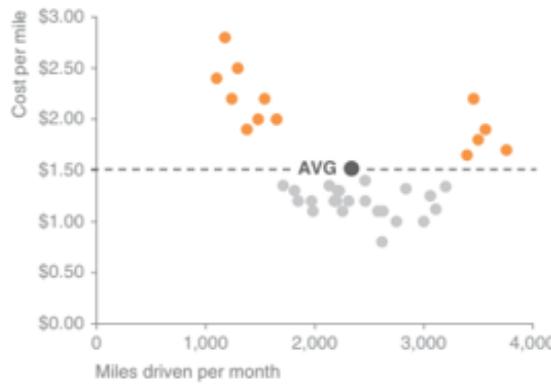
Εικ. 12: Πίνακας ('Table', Knaflic, 2015)

- **Heatmap:** Μια προσέγγιση για τον συνδυασμό λεπτομερειών που μπορεί κανείς να συμπεριλάβει σε έναν πίνακα, ενώ ταυτόχρονα να χρησιμοποιεί οπτικά στοιχεία, είναι μέσω ενός "heatmap" (χάρτης θερμότητας). Αποτελεί έναν τρόπο οπτικοποίησης δεδομένων σε μορφή πίνακα, όπου αντί για τους αριθμούς, χρησιμοποιούνται χρωματισμένα κελιά που μεταδίδουν τη σχετική μεγέθυνση των αριθμών.

	A	B	C
Category 1	15%	22%	42%
Category 2	40%	36%	20%
Category 3	35%	17%	34%
Category 4	30%	29%	26%
Category 5	55%	30%	58%
Category 6	11%	25%	49%

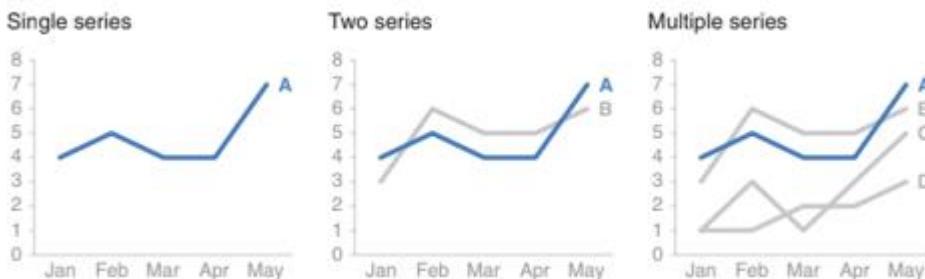
Εικ. 13: Heatmap ('Heatmap', Knaflic, 2015)

- **Scatter plot:** Τα διαγράμματα διασποράς μπορούν να είναι χρήσιμα για να δείξουν τη σχέση μεταξύ δύο μεγεθών, καθώς επιτρέπουν την κωδικοποίηση δεδομένων ταυτόχρονα σε έναν οριζόντιο άξονα x και έναν κατακόρυφο άξονα y για να δει κανείς αν υπάρχει και ποια σχέση υπάρχει. Συνήθως χρησιμοποιούνται πιο συχνά στους επιστημονικούς τομείς.



Εικ. 14: Scatter plot ('FIGURE 2.7 Modified scatter plot', Knafllic, 2015)

- **Line graph:** Χρησιμοποιούνται κυρίως για να αναπαραστήσουν συνεχείς δεδομένα, καθώς για τα σημεία που συνδέονται μέσω της γραμμής, υπονοείται μια σχέση μεταξύ των σημείων. Συχνά, τα συνεχή δεδομένα που αναπαρίστανται είναι σε κάποια μονάδα χρόνου: ημέρες, μήνες, τρίμηνα ή έτη. Σε ορισμένες περιπτώσεις, η γραμμή στο γράφημα γραμμών μπορεί να αντιπροσωπεύει ένα στατιστικό συνολικό στοιχείο, όπως η μέση τιμή ή η εκτίμηση ενός προγνωστικού σημείου.



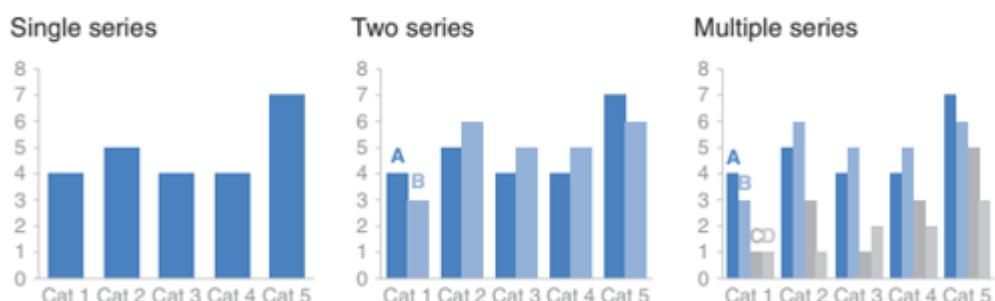
Εικ. 15: Line graph ('FIGURE 2.8 Line graphs', Knafllic, 2015)

- **Slopegraph:** Τα slopegraphs μπορούν να είναι χρήσιμα όταν υπάρχουν δύο χρονικές περίοδους ή σημεία σύγκρισης και χρειάζεται να αναδειχτούν γρήγορα οι σχετικές αυξήσεις και μειώσεις ή οι διαφορές σε κατηγορίες μεταξύ των δύο σημείων δεδομένων. Είναι ακόμα χρήσιμο για να επισημανθεί μια μόνο κατηγορία.

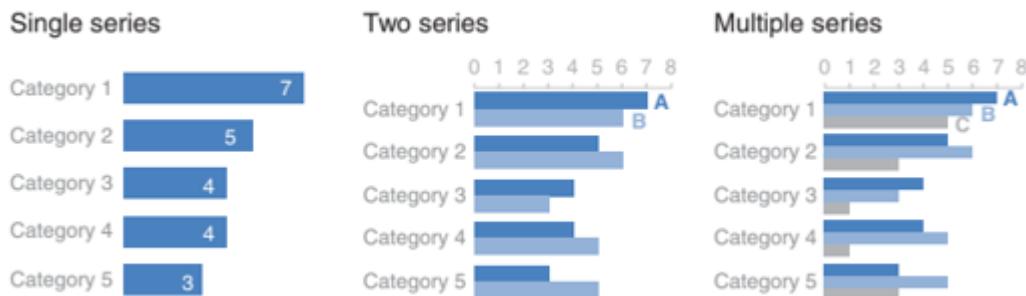


Εικ. 16: Slopegraph ('Modified slopegraph', Knafllic, 2015)

- Bars:** Χαρακτηριστικό είναι πως οι πληροφορίες οργανώνονται σε ομάδες. Τα μάτια συγκρίνουν τα άκρα των bars, κάτι που διευκολύνει την άμεση αντίληψη της μεγαλύτερης και της μικρότερης κατηγορίας, καθώς και της σταδιακής διαφοροποίησης μεταξύ των κατηγοριών. Μπορεί να είναι είτε κάθετο bar chart είτε οριζόντιο. Το τελευταίο είναι ιδιαίτερα χρήσιμο όταν τα ονόματα των κατηγοριών είναι μεγάλα, καθώς το κείμενο γράφεται από αριστερά προς τα δεξιά, όπως περισσότεροι αναγνώστες συνήθως διαβάζουν, καθιστώντας το γράφημα ευανάγνωστο για το κοινό. Ειδική κατηγορία αποτελεί ακόμα το **stacked horizontal bar chart**, για την οπτικοποίηση δεδομένων έρευνας που συλλέγονται κατά μήκος μιας κλίμακας Likert (μια κλίμακα που συνήθως χρησιμοποιείται σε έρευνες και κυμαίνεται συνήθως από Απολύτως Διαφωνώ έως Απολύτως Συμφωνώ).

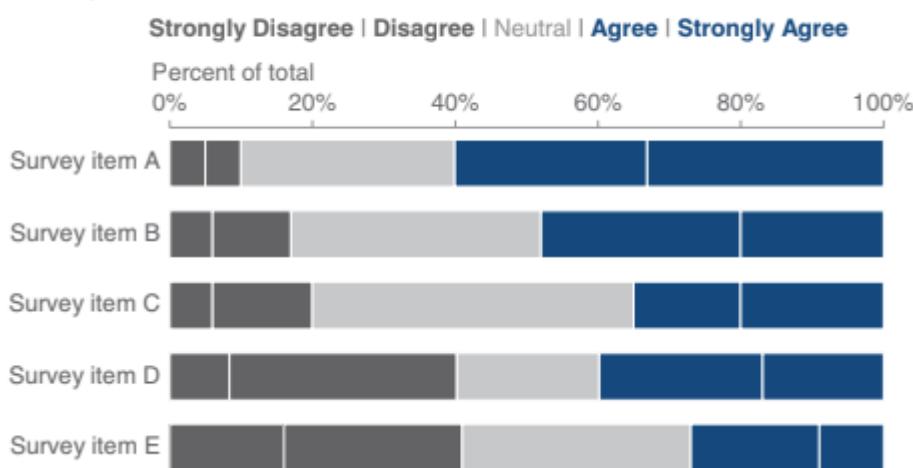


Εικ. 17: Vertical bar chart ('Bar charts', Knafllic, 2015)



Εικ. 18: Horizontal bar chart ('Horizontal bar charts', Knaflic, 2015)

Survey results



Εικ. 19: Stacked horizontal bar chart ('100% stacked horizontal bar chart', Knaflic, 2015)

- **Waterfall chart:** Το γράφημα καταρράκτη μπορεί να χρησιμοποιηθεί για να διαχωρίσει τα κομμάτια ενός γραφήματος στοίβας με σκοπό να επικεντρωθεί σε ένα κάθε φορά, ή για να δείξει ένα αρχικό σημείο, αυξήσεις και μειώσεις, και το τελικό αποτέλεσμα.



Εικ. 20: Waterfall chart ('FIGURE 2.17 Waterfall chart', Knaflic, 2015)

ΚΕΦΑΛΑΙΟ 4: Δημιουργία του Data Warehouse

4.1 Εισαγωγή

Για να δημιουργήσω τη ΒΔ, επέλεξα το Microsoft SQL Server Management Studio, καθώς φημίζεται για την ευελιξία και λειτουργικότητά του, παρέχοντας ένα ολοκληρωμένο περιβάλλον για τη διαχείριση δεδομένων. Για την υλοποίηση της εργασίας αυτής, ήμουν ανάμεσα και σε άλλες databases, όπως PostgreSQL και MongoDB, όμως κατέληξα πως η πρώτη αποτελεί την βέλτιστη επιλογή για ανάλυση δεδομένων (Tauriainen, 2023):

Χαρακτηριστικό	<i>MS SQL Server</i>	<i>PostgreSQL</i>	<i>MongoDB</i>
Τύπος Βάσης Δεδομένων	Σχεσιακή	Σχεσιακή	NoSQL
Σχεδίαση Δεδομένων για Ανάλυση	Ναι	Ναι	Όχι
Υποστήριξη SQL	Ναι	Ναι	Όχι
Συμπεριφορά Αναζήτησης	Ταχύτατη	Αποτελεσματική	Καλή
Υποστήριξη Συγχρονισμού Δεδομένων	Ναι	Ναι	Όχι
Υποστήριξη Συναλλαγών	Υψηλή	Καλή	Βασική
Διαχείριση Μεγάλων Όγκων Δεδομένων	Υψηλή	Υψηλή	Καλή

Πίν.2: Σύγκριση ΒΔ

4.2 Excel

Μεταβαίνοντας στη σελίδα της Airbnb³, κατέβασα τα αρχεία που αφορούν την Ελλάδα και με τα οποία δούλεψα. Από τα διαθέσιμα CSVs επέλεξα τα detailed, καθώς περιέχουν περισσότερη και σημαντικότερη πληροφορία από τα απλά. Επομένως, αρχικό βήμα αποτέλεσε το “τακτοποίημα” των CSVs. Αντιμετώπισα διάφορα προβλήματα, όπως την μη εμφάνιση των ελληνικών χαρακτήρων και τον

³ [Inside Airbnb: Get the Data](#)

διαχωρισμό των στηλών. Οπότε έφτιαξα νέα αρχεία μεταφορτώνοντας τα ήδη κατεβασμένα (Data -> Get Data -> From Text/CSV) στο Excel και αποθηκεύοντας τα αυτή τη φορά ως CSV UTF-8. Μελέτησα τον τύπο των δεδομένων κάθε στήλης, καθώς αυτός ζητείται επακριβώς από το SSMS στο σημείο της φόρτωσης των CSVs, αργότερα στη βάση. Δεν επέλεξα να σβήσω στήλες που μου φάνηκαν περιπτές μέσα από το Excel, αλλά πραγματοποίησα αυτήν τη διαδικασία στον SSMS με SQL queries. Όσο αφορά το CSV Calendar, επειδή ήταν αρκετά μεγάλο αρχείο (λίγο παραπάνω του ενός εκατομμυρίου εγγραφών), ο server έβγαζε system out of memory error. Για τον λόγο αυτό ανανέωσα τα αρχεία, βάζοντας maximum εγγραφές που δέχεται το Excel.

4.3 Data Load

Στη συνέχεια, δημιούργησα τη ΒΔ μου (Databases -> New Database), με το όνομα **Airbnb Listings Greece** και φόρτωσα τα CSVs (Tasks -> Import Flat File):

Κατά τη διαδικασία αυτή και στο βήμα Modify Columns, ασχολήθηκα με τον τύπο των δεδομένων για κάθε γραμμή, επιλέγοντας το Allow Nulls για όλα. Ναι μεν, ήδη το SSMS συμπληρώνει αυτόματα τον τύπο αλλά εγώ επέλεξα να πειράξω μερικά που δεν αντιστοιχούσαν στην πραγματική τιμή με βάση τις τιμές των CSVs και με βάση σχετικά errors που έβγαζε ο server. Παρατίθονται πίνακες με τον τελικό τύπο ανά listings, calendar και reviews, πρωτού το data cleaning.

Reviews:

- ❑ listing_id (bigint, null)
- ❑ review_id (bigint, null)
- ❑ review_date (date, null)

Εικ. 21: Reviews

Calendars:

- ❑ listing_id (bigint, null)
- ❑ date (date, null)
- ❑ available (bit, null)
- ❑ price (money, null)
- ❑ adjusted_price (money, null)
- ❑ minimum_nights (int, null)
- ❑ maximum_nights (int, null)

Εικ. 22: Calendars

Listings:

└ listing_id (bigint, null)	
└ listing_url (nvarchar(50), null)	
└ name (nvarchar(200), null)	
└ description (nvarchar(max), null)	
└ picture_url (nvarchar(150), null)	
└ host_id (int, null)	
└ host_url (nvarchar(50), null)	
└ host_name (nvarchar(50), null)	
└ host_since (date, null)	
└ host_location (nvarchar(100), null)	
└ host_about (nvarchar(max), null)	
└ host_response_time (nvarchar(50), null)	
└ host_response_rate (nvarchar(50), null)	
└ host_acceptance_rate (nvarchar(50), null)	
└ host_is_superhost (bit, null)	
└ host_picture_url (nvarchar(150), null)	
└ host_neighbourhood (nvarchar(50), null)	
└ host_total_listings_count (smallint, null)	
└ host_verifications (nvarchar(50), null)	
└ host_identity_verified (bit, null)	
└ neighbourhood_cleansed (nvarchar(50), null)	
└ latitude (float, null)	
└ longitude (float, null)	└ availability_365 (smallint, null)
└ property_type (nvarchar(50), null)	└ number_of_reviews (smallint, null)
└ room_type (nvarchar(50), null)	└ number_of_reviews_ltm (int, null)
└ accommodates (tinyint, null)	└ number_of_reviews_l30d (int, null)
└ bathrooms (nvarchar(1), null)	└ first_review (date, null)
└ bathrooms_text (nvarchar(50), null)	└ last_review (date, null)
└ bedrooms (tinyint, null)	└ review_scores_rating (float, null)
└ beds (tinyint, null)	└ review_scores_accuracy (float, null)
└ amenities (nvarchar(max), null)	└ review_scores_cleanliness (float, null)
└ price (money, null)	└ review_scores_checkin (float, null)
└ minimum_nights (int, null)	└ review_scores_communication (float, null)
└ maximum_nights (int, null)	└ review_scores_location (float, null)
└ minimum_nights_avg_ntm (float, null)	└ review_scores_value (float, null)
└ maximum_nights_avg_ntm (float, null)	└ instant_bookable (bit, null)
└ calendar_updated (nvarchar(1), null)	└ calculated_host_listings_count (smallint, null)
└ has_availability (bit, null)	└ calculated_host_listings_count_entire_homes (smallint, null)
└ availability_30 (tinyint, null)	└ calculated_host_listings_count_private_rooms (tinyint, null)
└ availability_60 (tinyint, null)	└ calculated_host_listings_count_shared_rooms (tinyint, null)
└ availability_90 (tinyint, null)	└ reviews_per_month (float, null)

Εικ. 23: Listings

4.4 Data Cleaning

Μέρος του Data Cleaning αποτελεί και η διαγραφή των στηλών κάθε πίνακα, που δεν είναι απαραίτητες. Σημαντικό ρόλο στην κατανόηση των στηλών και των πληροφοριών που μελετώ, αποτέλεσε το Data Dictionary που παριείχε η Airbnb (Data -> Data Dictionary):

	A	B	C	D
1	Data Dictionary for listings.csv detailed file			
2				
3	File Name	listings.csv		
4	Version	4.3		
5	Date Introduced	August, 2022		
6	Description			
7				
8	Field	Type	Calculated	Description
9	id	integer		Airbnb's unique identifier for the listing
10	listing_url	text	y	
11	scrape_id	bigint	y	Inside Airbnb "Scrape" this was part of
12	last_scraped	datetime	y	UTC. The date and time this listing was "scraped".
13	source	text		One of "neighbourhood search" or "previous scrape". "neighbourhood search" means that the listing was found by searching the city, while "previous scrape" means that the listing was seen in another scrape performed in the last 65 days, and the listing was confirmed to be still available on the Airbnb site.
14	name	text		Name of the listing
15	description	text		Detailed description of the listing
16	neighborhood_overview	text		Host's description of the neighbourhood
17	picture_url	text		URL to the Airbnb hosted regular sized image for the listing
18	host_id	integer		Airbnb's unique identifier for the host/user
19	host_url	text	y	The Airbnb page for the host
20	host_name	text		Name of the host. Usually just the first name(s).
21	host_since	date		The date the host/user was created. For hosts that are Airbnb guests this could be the date they registered as a guest.
22	host_location	text		The host's self reported location
23	host_about	text		Description about the host
24	host_response_time			
25	host_response_rate			
26	host_acceptance_rate			That rate at which a host accepts booking requests.
27	host_is_superhost	boolean [t=true; f=false]		
28	host_thumbnail_url	text		

Εικ. 24: Data Dictionary ([Inside Airbnb: Get the Data](#))

Συγκεκριμένα στους reviews πίνακες για κάθε περιοχή, αφαίρεσα το όνομα του reviewer, καθώς δεν δίνει κάποια επιπλέον πληροφορία αλλά και τα σχόλια, αφού δεν υπάρχει λόγος να τα αναλύσω. Όσο αφορά τα listings αφαιρέθηκαν διάφορες περιπτές στήλες όπως φαίνονται και στην εικόνα που ακολουθεί με το δεδομένο query:

```
--drop unused columns
alter table [aegean_reviews]
drop column reviewer_id, reviewer_name, comments;

alter table ath_reviews
drop column reviewer_id, reviewer_name, comments;

alter table skg_reviews
drop column reviewer_id, reviewer_name, comments;

alter table creta_reviews
drop column reviewer_id, reviewer_name, comments;

alter table aegean_listings
drop column scrape_id, last_scraped, source, neighborhood_overview,
host_thumbnail_url, host_listings_count, host_has_profile_pic,
neighbourhood, neighbourhood_group_cleansed, minimum_minimum_nights,
maximum_minimum_nights, minimum_maximum_nights, maximum_maximum_nights,
calendar_last_scraped, license
```

```

alter table ath_listings
drop column scrape_id, last_scraped, source, neighborhood_overview,
host_thumbnail_url, host_listings_count, host_has_profile_pic,
neighbourhood, neighbourhood_group_cleansed, minimum_minimum_nights,
maximum_minimum_nights, minimum_maximum_nights, maximum_maximum_nights,
calendar_last_scraped, license

alter table skg_listings
drop column scrape_id, last_scraped, source, neighborhood_overview,
host_thumbnail_url, host_listings_count, host_has_profile_pic,
neighbourhood, neighbourhood_group_cleansed, minimum_minimum_nights,
maximum_minimum_nights, minimum_maximum_nights, maximum_maximum_nights,
calendar_last_scraped, license

alter table creta_listings
drop column scrape_id, last_scraped, source, neighborhood_overview,
host_thumbnail_url, host_listings_count, host_has_profile_pic,
neighbourhood, neighbourhood_group_cleansed, minimum_minimum_nights,
maximum_minimum_nights, minimum_maximum_nights, maximum_maximum_nights,
calendar_last_scraped, license

```

Εικ. 25: Διαγραφή στηλών

Εξίσου απαραίτητη είναι και η μετονομασία στηλών με σκοπό την καλύτερη κατανόησή τους. Έτσι, άλλαξα στα listings της κάθε περιοχής το id σε listing_id, καθώς επίσης και για τους πίνακες reviews μετέτρεψα το id σε review_id. Ακόμα, θέλοντας να δώσω περισσότερο περιγραφικά στοιχεία, στα reviews κάθε περιοχής, μετονόμασα το date σε review_date. Η εντολή sp_rename χρησιμοποιείται για να μετονομάσει στήλες στους πίνακες:

```

sp_rename 'aegean_listings.id', 'listing_id', 'COLUMN';
sp_rename 'aegean_reviews.id', 'review_id', 'COLUMN';
sp_rename 'aegean_reviews.date', 'review_date', 'COLUMN';

sp_rename 'ath_listings.id', 'listing_id', 'COLUMN';
sp_rename 'ath_reviews.id', 'review_id', 'COLUMN';
sp_rename 'ath_reviews.date', 'review_date', 'COLUMN';

sp_rename 'creta_listings.id', 'listing_id', 'COLUMN';
sp_rename 'creta_reviews.id', 'review_id', 'COLUMN';
sp_rename 'creta_reviews.date', 'review_date', 'COLUMN';

sp_rename 'skg_listings.id', 'listing_id', 'COLUMN';
sp_rename 'skg_reviews.id', 'review_id', 'COLUMN';
sp_rename 'skg_reviews.date', 'review_date', 'COLUMN';

```

Εικ. 26: Μετονομασία στηλών

4.5 Ένωση όμοιων πινάκων

Στο σημείο αυτό, έχω φορτώσει όλα τα αρχικά αρχεία και σκοπός μου είναι να καταλήξω σε τρεις βασικούς πίνακες: calendar, listings και reviews. Ο καθένας θα παριέχει για όλες τις περιοχές που μελετούνται (Αθήνα, Θεσσαλονίκη, Κρήτη και Αιγαίο) όλα τα δεδομένα. Ουσιαστικά στοχεύω να συνδυάσω και να ενοποιήσω όλους τους επιμέρους πίνακες σε έναν μαζικό. Δεδομένου ότι φτιάχτηκαν οι τελικοί πίνακες, θα μπορούσα να διαγράψω τους αρχικούς. Ωστόσο επέλεξα να τους κρατήσω, σε περίπτωση λάθους.

Το union όλων των πινάκων σε έναν γίνεται με το παρακάτω query. Χρησιμοποίησα το UNION ALL, καθώς αυτό επιτρέπει να υπάρχουν διπλότυπα. Στο πρώτο query έγραψα το into, το οποίο δημιουργεί έναν πίνακα με τις αντίστοιχες στήλες και τα ίδια data types, πραγματοποιώντας ουσιαστικά concatenate. Αυτό μπορούμε να το επαληθεύσουμε, βλέποντας πως ο κάθε πίνακας έχει ένα εκατομμύριο εγγραφές και ο καινούργιος πίνακας έχει περίπου 4 εκατομμύρια εγγραφές. Αντιστοίχως τρέχω queries για listings και reviews:

```
--CALENDAR          --REVIEWS           --LISTINGS
select *           select *             select *
into airbnb_calendar    into airbnb_reviews   into airbnb_listings
from aegean_calendar   from aegean_reviews  from aegean_listings
UNION ALL          UNION ALL          UNION ALL
select *           select *             select *
from ath_calendar     from ath_reviews    from ath_listings
UNION ALL          UNION ALL          UNION ALL
select *           select *             select *
from skg_calendar     from skg_reviews   from skg_listings
UNION ALL          UNION ALL          UNION ALL
select *           select *             select *
from creta_calendar   from creta_reviews  from creta_listings
```

Εικ. 27: Union πινάκων

4.5.1 Επαληθεύσεις

Μια σημαντική επαλήθευση αποτελεί πως ο τελικός πίνακας (airbnb_listings) περιέχει το ίδιο πλήθος εγγραφών με το πλήθος μοναδικών τιμών στο listing_id, εξασφαλίζοντας πως τα listings μιας περιοχής σε σχέση με μιας άλλης δεν πρόκειται ποτέ να είναι τα ίδια. Όλα τα listings γενικότερα της Airbnb πλατφόρμας παίρνουν μοναδικές τιμές και δεν επαναλαμβάνονται ανά περιοχή. Συγκρίνοντας το πλήθος αυτών των τιμών, βλέπουμε πως ταυτίζονται. Ομοίως και για τα reviews:

```
SELECT count(distinct listing_id), count(*)
FROM [Airbnb Listings Greece].[dbo].[airbnb_listings]
```

	(No column name)	(No column name)
1	76825	76825

```
SELECT count(distinct review_id), count(*)
FROM [Airbnb Listings Greece].[dbo].[airbnb_reviews]
```

	(No column name)	(No column name)
1	1841657	1841657

Ταυτόχρονα πρέπει να ελεγχθεί αν στο airbnb_calendar για κάθε listing και για κάθε ημερομηνία υπάρχει μια μοναδική εγγραφή. Στο πρώτο μέρος του ελέγχου, βρίσκουμε πόσες εγγραφές έχουμε γενικά. Στο δεύτερο μέρος κάνω το ίδιο βήμα, μόνο που θα πραγματοποιήσω group by listing_id και date (PK), καθώς ανάλογα με τις στήλες που έχουν επιλεχθεί, ο server φέρνει unique combination. Άρα αν για

κάποιον λόγο το listing_id με date είναι το ίδιο, ο server δεν θα το φέρει. Τα πλήθη ταυτίζονται:

```
--4 194 300 rows
SELECT count(*)
FROM [Airbnb Listings Greece].[dbo].[airbnb_calendar]

--4 194 300 rows. That means that for each listing and date we have exactly one unique record
SELECT count(1)
FROM [Airbnb Listings Greece].[dbo].[airbnb_calendar]
group by listing_id, date
```

	4194297	1
	4194298	1
	4194299	1
1	4194300	1
	4194300	1

Εικ. 28: Επαληθεύσεις

Τελικά, το PK του airbnb_listings είναι το listing_id, στο airbnb_reviews το review_id και στον airbnb_calendar πίνακα είναι ο συνδυασμός listing_id με date.

4.6 Fact και Dimension πίνακες

Σε αυτό το σημείο δημιουργήθηκε το DW, αφού πραγματοποιήθηκαν κάποια αναγκαία data transformation και data cleaning. Ακολουθεί επεξήγηση στο πώς δημιουργήθηκαν οι Fact και Dimension Tables.

- **FactReviews**

Μια σημαντική συνθήκη που πρέπει να σημειωθεί είναι ότι το listing_id είναι μια τιμή που δεν πρέπει να είναι NULL. Επομένως γράφω την παρακάτω συνθήκη. Η συνάρτηση ISNULL ελέγχει αν η τιμή της στήλης listing_id είναι NULL. Εάν είναι, επιστρέφει την τιμή 0. Αυτό δημιουργεί μια νέα στήλη με όνομα listing_id στην οποία αντιστοιχίζονται οι τιμές της στήλης listing_id από τον πίνακα Airbnb_reviews, με τη διαφορά ότι οι NULL τιμές αντικαθίστανται με το 0. Το review_id -ορθώς- δεν γίνεται ποτέ 0, μετά από σχετικό έλεγχο:

```
--Data Warehouse
--checking column by column to see if there is any null value to replace it accordingly

SELECT review_id, ISNULL(listing_id, 0) AS listing_id, review_date
INTO FactReviews
FROM airbnb_reviews
```

Εικ. 29: FactReviews

- **FactCalendar**

Κι εδώ το listing_id, επειδή θέλω να είμαι σίγουρη ότι δεν παίρνει NULL τιμή, εισάγω μια όμοια συνθήκη. Στα dates, αν δεν υπάρχουν δεδομένα, αντίστοιχα

ορίζω την ημερομηνία 1900-01-01, η οποία ξεχωρίζει. Αντικαθιστώ ακόμα το f και t (false/true) με 0 και 1 για να είναι αριθμός, που πιάνει λιγότερο χώρο. Αργότερα το μετατρέπω από int σε bit, που καταλαμβάνει ακόμα λιγότερη χωρητικότητα. Εξασφαλίζω ότι το price και adjusted_price σε περιπτώσεις που είναι NULL παίρνουν την τιμή 0. Το ίδιο ισχύει και για τις διανυκτερεύσεις, με επιπλέον την πληροφορία (από το Data Dictionary) ότι οι minimum και maximum nights μπορεί να διαφέρουν από αυτές που είναι στα listings, βάζοντας έτσι το calendar μπροστά για να τις προσδιορίζει:

```
--check column by column to see if there is any null value to replace it accordingly
SELECT ISNULL(listing_id, 0) as listing_id, ISNULL(date, '1900-01-01') as date,
ISNULL(price, 0) as price, ISNULL(adjusted_price, 0) as adjusted_price,
ISNULL(minimum_nights, 0) as calendar_minimum_nights,
ISNULL(maximum_nights, 0) as calendar_maximum_nights
INTO FactCalendar
FROM airbnb_calendar
```

Εικ. 30: FactCalendar

- ***DimLocation, Dim.PropertyType, Dim.RoomType***

Για αυτούς τους πίνακες, βρέθηκα στην ανάγκη να δημιουργήσω από μόνη μου το id. Ξεκινώντας με τον DimLocation, έφτιαξα μια στήλη με το όνομα location_id, τύπου integer και χρησιμοποίησα την εντολή identity (1,1). Αυτό σημαίνει πως η βάση ξεκινάει από την πρώτη εγγραφή και της δίνει id = 1 και κάθε φορά που βλέπει μια νέα εγγραφή, την αυξάνει κατά ένα. Το location_id αποτελεί και το PK, αφού παίρνει μοναδική τιμή. Πρόσθεσα και το neighborhood_description με το μέγεθος του. Στη συνέχεια, προσθέτω στον νέο πίνακα όλες τις μοναδικές τιμές (DISTINCT) από το neighborhood_cleansed από τον πίνακα airbnb_listings. Ουσιαστικά αντιστοιχίζονται οι τιμές ανάμεσα σε neighborhood_description και neighborhood_cleansed. Ταυτόχρονα όμως, χρειαζόμουν και μια εγγραφή, που να ανταποκρίνεται στο id = 0 στην περίπτωση που δεν είναι γνωστή η τοποθεσία/ room type/ property type/ host id κλπ, σχετικά με αντίστοιχα κλειδιά που βρίσκονται στον FactListings. Για την περιγραφή (neighborhood_description) επιλέγω το κείμενο ‘Unknown’. Οπότε το (0, ‘Unknown’) αποτελεί από μόνο του μια εγγραφή για την άγνωστη τιμή. Όμως επειδή έχω ορίσει κατά τη δημιουργία του location_id πως παίρνει τιμή αυτόματα, δεν μπορεί να αποδεχτεί το σύστημα να ορίσω εγώ κάποια τιμή. Για τον λόγο αυτό, πρωτού γράψω το query για την εγγραφή αυτή, ενεργοποιώ το identity insert -από τον προγραμματιστή- και στο τέλος το απενεργοποιώ στιγμαία. Με το ίδιο σκεπτικό, δημιουργήθηκαν και οι άλλοι πίνακες, με PK τα property_type_id και room_type_id αντίστοιχα:

```

CREATE TABLE DimLocation (
    location_id int identity(1,1) primary key,
    neighbourhood_description nvarchar(150)
)

INSERT INTO DimLocation
SELECT DISTINCT neighbourhood_cleansed
FROM airbnb_listings

SET IDENTITY_INSERT DimLocation ON

INSERT INTO DimLocation (location_id,neighbourhood_description )
SELECT 0, 'Unknown'

SET IDENTITY_INSERT DimLocation OFF

CREATE TABLE Dim.PropertyType (
    property_type_id int identity(1,1) primary key,
    property_type_description nvarchar(50)
)

INSERT INTO Dim.PropertyType
SELECT DISTINCT property_type
FROM airbnb_listings

SET IDENTITY_INSERT Dim.PropertyType ON

INSERT INTO Dim.PropertyType (property_type_id,property_type_description )
SELECT 0, 'Unknown'

SET IDENTITY_INSERT Dim.PropertyType OFF

CREATE TABLE Dim.RoomType (
    room_type_id int identity(1,1) primary key,
    room_type_description nvarchar(50)
)

INSERT INTO Dim.RoomType
SELECT DISTINCT room_type
FROM airbnb_listings

SET IDENTITY_INSERT Dim.RoomType ON

INSERT INTO Dim.RoomType (room_type_id, room_type_description )
SELECT 0, 'Unknown'

SET IDENTITY_INSERT Dim.RoomType OFF

```

Εικ. 31: DimLocation, Dim.PropertyType, Dim.RoomType

- *DimHost*

Στον συγκεκριμένο πίνακα χρειάστηκε να κάνω αρκετό data transformation, αντικαθιστώντας κυρίως τις NULL τιμές με το ISNULL. Όπως αναφέρθηκε και προηγουμένως, για NULL ημερομηνίες, χρησιμοποιήθηκε η 1900-01-01 και για N/A αριθμούς, έβαλα τον αρνητικό -1, με σκοπό να εξαιρεθούν από το τελικό report. Ελέγχω ανά στήλη τις τιμές, τρέχοντας το query:

```
--check column by column to see if there is any null value to replace it accordingly
select distinct host_identity_verified
from airbnb_listings
```

Ιδιαίτερη προσοχή δίνεται στα ποσοστά (rates). Σκοπός μου ήταν να απομονώσω τον αριθμό. Πρώτο βήμα του query ήταν να αντικαταστήσω το σύμβολο % με το κενό. Έπειτα, κατόπιν αναζήτησης, εντόπισα την τιμή N/A, οδηγώντας με να εκτελέσω εμφωλευμένο replace με -1. Τελικά, με όλα αυτά που έχω καταλήξει, αν η DB βρει NULL, το αντικαθιστά με -1 και έπειτα το μετατρέπω σε smallint (τελευταία queries).

Ο πίνακας αυτός προέρχεται αποκλειστικά από τον airbnb_listings, χρησιμοποιώντας πάλι το DISTINCT. Τα καταλύματα προέρχονται από διάφορους hosts. Ένας host μπορεί να έχει πολλά καταλύματα, επομένως για να καταγράψω τους μοναδικούς hosts, χρησιμοποιώ την εντολή αυτή. Το host_id (που αποτελεί και το PK) ήδη παρέχεται μέσα από τον airbnb_listings, άρα αποφεύγεται η διαδικασία που ακολουθήθηκε με τον DimLocation.

```
SELECT DISTINCT host_id, host_url, ISNULL(host_name, 'Unknown') as host_name,
ISNULL(host_since, '1900-01-01') as host_since, ISNULL(host_location, 'Unknown') as host_location,
host_about, ISNULL(host_response_time, 'N/A') as host_response_time,
isnull(replace(replace(host_response_rate, '%',''),'N/A', '-1'), '-1') as host_response_rate,
isnull(replace(replace(host_acceptance_rate, '%',''),'N/A', '-1'), '-1') as host_acceptance_rate,
ISNULL(host_is_superhost, -1) as host_is_superhost, host_picture_url,
host_total_listings_count, replace(host_verifications, '[]', 'None') as host_verifications,
ISNULL(host_identity_verified, -1) as host_identity_verified
INTO DimHost
FROM airbnb_listings

--check column by column to see if there is any null value to replace it accordingly
select distinct host_identity_verified
from airbnb_listings

ALTER TABLE DimHost
ALTER COLUMN host_response_rate smallint;

ALTER TABLE DimHost
ALTER COLUMN host_acceptance_rate smallint;
```

Εικ. 32: DimHost

- ***FactListings***

Ο FactListings αποτελεί τον πιο σημαντικό πίνακα και βασίζεται εξ ολοκλήρου στον airbnb_listings. Σκοπίμως δημιουργήθηκε τελευταίος γιατί φέρει κλειδιά από τις διαστάσεις που φτιάχτηκαν προηγουμένως. Ο airbnb_listings δεν περιελάμβανε αυτά τα κλειδιά. Χρησιμοποιήθηκε ως αφετηρία, για να χτίσω και να φέρω αυτά τα id, με τη χρήση left joins, διότι μέσω αυτών δεν χάνεται καμία εγγραφή από τον airbnb_listings, αν και εφόσον υπάρχει περίπτωση να μην γίνεται join κάποιος

αριθμός με κάποια στήλη. Left join του airbnb_listings έγινε ανάμεσα στις στήλες του neighborhood_cleansed, property_type και room_type αντίστοιχα με neighborhood_description, property_type_description και room_type_description των πινάκων DimLocation, DimPropertyType και DimRoomType αντίστοιχα, όπως φαίνεται στο query. Δεν χρειάστηκε να εκτελέσω join ανάμεσα στο host_id του airbnb_listings με του DimHost, καθώς ο πρώτος ήδη περιλαμβάνει το πεδίο αυτό. Έτσι, έχω τα κλειδιά του εκάστοτε Dimension πίνακα, με τα joins να χρησιμοποιούνται για την αντιστοίχιση των κλειδιών τους. Ομοίως με τον DimHost, φρόντισα για τις NULL τιμές. Επαλήθευση έγινε με το τελευταίο query, με τα πλήθη να ταυτίζονται ανάμεσα σε airbnb_listings και τον νέο πίνακα:

```

SELECT listing_id, isnull(host_id, 0) as host_id, isnull(location_id, 0) as location_id,
       isnull(property_type_id, 0) as property_type_id, isnull(room_type_id, 0) as room_type_id,
       listing_url, name, description, picture_url, isnull(price, 0) as price, latitude, longitude,
       accommodates, bathrooms,
       isnull(bathrooms_text, '-1 baths') as bathrooms_text, ISNULL(bedrooms, 0) AS bedrooms,
       ISNULL(beds, 0) as beds, amenities, minimum_nights, maximum_nights, minimum_nights_avg_ntm,
       maximum_nights_avg_ntm, has_availability, availability_30, availability_60, availability_90,
       availability_365, instant_bookable, number_of_reviews, number_of_reviews_ltm, number_of_reviews_l30d,
       isnull(first_review, '1900-01-01') as first_review, isnull(last_review, '1900-01-01') as last_review,
       reviews_per_month, review_scores_rating, review_scores_accuracy, review_scores_cleanliness,
       review_scores_checkin, review_scores_communication, review_scores_location, review_scores_value,
       calculated_host_listings_count, calculated_host_listings_count_entire_homes,
       calculated_host_listings_count_private_rooms,
       calculated_host_listings_count_shared_rooms
INTO FactListings
FROM airbnb_listings a
left join DimLocation c on a.neighbourhood_cleansed = c.neighbourhood_description
left join DimPropertyType d on a.property_type = d.property_type_description
left join DimRoomType e on a.room_type = e.room_type_description

```

(76825 rows affected)

Completion time: 2024-01-10T23:09:06.6747167+02:00

```

SELECT COUNT(1)
FROM airbnb_listings

```

	(No column name)
1	76825

Εικ. 33: FactListings

Είναι σημαντικό να σημειωθεί σε αυτό το σημείο ότι συχνά στις πηγές που χρησιμοποιεί ο αναλυτής ενδέχεται να εμπεριέχονται δεδομένα από την πηγή, τα οποία μπορεί να μην ανταποκρίνονται στην πραγματικότητα λόγω της πολύ υψηλότερης τιμής τους σε σχέση με τις άλλες τιμές. Αυτά αποτελούν τα λεγόμενα out layers, τα οποία θεωρώ ότι υπάρχουν καταλάθως. Στη βάση δεδομένων εντόπισα τρεις εγγραφές, τις οποίες και αφαίρεσα.

Έχοντας ήδη σκεφτεί πιθανά διαγράμματα με την έναρξη της εργασίας αυτής, κράτησα από το data cleaning, στήλες που εγώ θεωρούσα σημαντικές. Ναι μεν κάποια fields μπορεί να μην χρησιμοποιηθούν στις απεικονίσεις αλλά προσδίδουν

εξίσου σημαντική πληροφορία για τυχόν μελλοντικές υλοποιήσεις ή αλλαγές. Αυτό δεν σημαίνει βέβαια ότι ανά πάσα στιγμή η δομή του DW δεν μπορεί να αλλάξει.

Οι τελικοί πίνακες είναι οι εξής:

- ⊕ dbo.aegean_calendar
- ⊕ dbo.aegean_listings
- ⊕ dbo.aegean_reviews
- ⊕ dbo.airbnb_calendar
- ⊕ dbo.airbnb_listings
- ⊕ dbo.airbnb_reviews
- ⊕ dbo.ath_calendar
- ⊕ dbo.ath_listings
- ⊕ dbo.ath_reviews
- ⊕ dbo.creta_calendar
- ⊕ dbo.creta_listings
- ⊕ dbo.creta_reviews
- ⊕ dbo.DimHost
- ⊕ dbo.DimLocation
- ⊕ dbo.Dim.PropertyType
- ⊕ dbo.DimRoomType
- ⊕ dbo.FactCalendar
- ⊕ dbo.FactListings
- ⊕ dbo.FactReviews
- ⊕ dbo.skg_calendar
- ⊕ dbo.skg_listings
- ⊕ dbo.skg_reviews

Εικ. 34: Τελικοί πίνακες

4.7 Σχεδίαση στο Tableau

Με βάση τα εναπομείναντα πεδία και τους fact/dimension πίνακες που αναλύθηκαν παραπάνω, γίνεται η υλοποίηση των γραφημάτων στο Tableau. Δυστυχώς, δεν υπάρχουν δεδομένα που να μελετούν μετρικές στο παρελθόν, με σκοπό την εύρεση μοτίβων ή τη μελέτη της εξέλιξης μιας τιμής με την πάροδο του χρόνου. Με τις δεδομένες, πολλαπλές και διαφορετικές πληροφορίες, καλό είναι να γίνει μια σχεδίαση πριν την υλοποίηση. Για τον λόγο αυτό, προσχεδίασα την όψη και τις βασικές πληροφορίες που θα ήθελα να εμπεριέχονται στα dashboards.

Όπως αναλύθηκε και στο θεωρητικό κομμάτι, ένα strategic dashboard διαθέτει βασικές πληροφορίες. Επομένως αποφάσισα να είναι πιο περιεκτικό από το λειτουργικό. Θεωρώ επίσης πως οι επενδυτές δεν ενδιαφέρονται μόνο για την πώληση ακινήτων αλλά και για ενδεχόμενες συνεργασίες, συνενοικιάσεις με οικοδεσπότες της Airbnb. Ακόμα κι αν δεν νοιάζονται για κοινοπραξίες αλλά επιθυμούν να εισαχθούν αυτόνομοι στην πλατφόρμα αυτή μετά από κάποια αγορά τους, είναι σημαντικό να μελετήσουν κι άλλους παράγοντες που καθιστούν

επιτυχημένο ένα κατάλυμα προς ενοικίαση, βασισμένοι σε πραγματικές και επίσημες πληροφορίες. Έτσι, στο operational dashboard επιλέγω να μελετήσω και άλλα πεδία πέρα από την τιμή, τον χρόνο κλπ. Πιο συγκεκριμένα:



Εικ. 35: Προσχέδιο για το strategic dashboard



Εικ. 36: Προσχέδιο για το operational dashboard

ΚΕΦΑΛΑΙΟ 5: Υλοποίηση στο Tableau

5.1 Εισαγωγή

Στην προσπάθειά μου να δημιουργήσω τα dashboards, χρησιμοποίησα όλους τους πίνακες που αναλύθηκαν στο παραπάνω κεφάλαιο, με κυριότερο -όπως είναι φυσικό- τον FactListings. Χρησιμοποιήσα πολλά και διαφορετικά πεδία τους, ιδιαίτερα για το operational dashboard, καθώς χρειάζεται να είναι αναλυτικότερο. Προσπάθησα επίσης να χρησιμοποιήσω ποικίλα γραφήματα, ταϊριαστά κάθε φορά στις ανάγκες της ανάλυσης. Μερικά από τα γραφήματα είναι τα pies, bar charts, gantt charts, treemaps, maps, tables, scatter plots, area charts ή και συνδυασμός αυτών. Σε κάθε περίπτωση, παρακάτω εξηγείται αναλυτικά ποιοι πίνακες και fields τους χρησιμοποιήθηκαν για το κάθε γράφημα.

Τα δεδομένα ξεκινούν χρονικά από τον φθινόπωρο του 2023 και καταλήγουν στο επόμενο φθινόπωρο του 2024. Όταν λοιπόν χρησιμοποιώ το πεδίο αυτό του χρόνου για απεικονίσεις, πάντα επιλέγω το διάστημα που βρίσκεται στο δεύτερο tab σύφωνα με την παρακάτω εικόνα. Αυτό συμβαίνει διότι το πρώτο section αναγνωρίζει τον χρόνο ως dimension, κι όχι ως measure, το οποίο ενοποιεί τους ίδιους μήνες/μέρες κλπ:

Year	2015
Quarter	Q2
Month	May
Day	8
More	▶

● Year	2015
Quarter	Q2 2015
Month	May 2015
Week Number	Week 5, 2015
Day	May 8, 2015
More	▶

Εικ. 37: Χρόνος στο Tableau

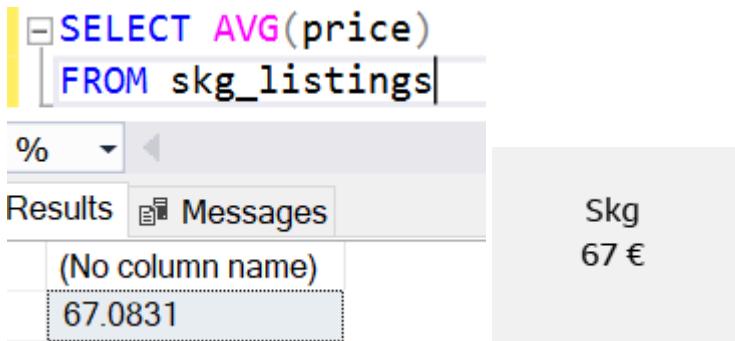
Ανάμεσα στα πεδία price του FactListings και FactCalendar, επιλέγω το πρώτο όταν μελετώ το μέσο όρο, το άθροισμα κλπ αφού το data dictionary αναφέρει πως η τιμή στο listings πρόκειται για το καταγεγραμμένο σε κάθε εγγραφή ποσό, κατά τη στιγμή που η Airbnb εξάγει τις πηγές της. Όταν όμως θέλω να μελετήσω την τιμή σε βάθος χρόνου, πάντα διαλέγω το price του FactCalendar, αφού σε αυτόν τον πίνακα για κάθε ημερομηνία της τρέχουσας χρονιάς αντιστοιχεί και μια τιμή.

Κοινό χαρακτηριστικό και στα δύο dashboards αποτελεί το μαύρο χρώμα για background. Αυτό επιλέχτηκε σκοπίμως, καθώς είναι σύνηθες από τους αναλυτές να το χρησιμοποιούν αφού θυμίζει παλιές κονσόλες, όπου γινόταν η ανάλυση στο παρελθόν. Επίσης θεωρείται ότι προσφέρει καλύτερη ορατότητα και περισσότερη

συγκέντρωση. Το theme χρώμα στο strategic dashboard αποτελεί το κόκκινο, όπως είναι και το σύμβολο της εταιρίας που μελετάται. Στο operational dashboard επέλεξα το μπλε για να υπάρχει διαχωρισμός. Πάντως όταν εμπλέκονται οι τέσσερις βασικές τοποθεσίες (Αθήνα, Κρήτη, Αιγαίο, Θεσσαλονική) και στα δυο dashboards χρησιμοποιούνται πάντα τα ίδια χρώματα για την κάθε μια (ροζ, μωβ, κόκκινο και πορτοκαλί αντίστοιχα), έτσι ώστε να μην υπάρχουν παρανοήσεις (Location field ως χρώμα). Η διαδραστικότητα επίσης αποτελεί αναπόσπαστο κομμάτι των ταμπλό - όπου αυτή είναι δυνατή- καθώς βοηθάει στην ομαδοποίηση, την μελέτη και κατανόηση σε διαφορετικά πεδία ταυτοχρόνως όπως και στην εξάλειψη της πολυτλοκότητας. Επέλεξα να χρησιμοποιήσω highlighter, κι όχι φίλτρο, καθώς το δεύτερο εξαφανίζει τις πληροφορίες που δεν μελετούνται, καθιστώντας έτσι τα σχήματα πολλές φορές παραπλανητικά.

Το tooltip επίσης είναι ένα δυνατό εργαλείο που προσφέρει το Tableau, αφού παραθέτει σε ρορ υπ σύννεφο επιλεγμένες πληροφορίες που θεωρούνται σημαντικές. Πράγματι, σε πολλές περιπτώσεις, δεν μπορούσα ή δεν ήταν χρηστικό και ευανάγνωστο να χρησιμοποιήσω κάποιο δεδομένο σε γράφημα, τοποθετώντας το έτσι στο tooltip. Έγινε επίσης χρήση κι άλλων δυνατοτήτων του Tableau, όπως η δημιουργία hierarchies, groups, clusters, filters και calculated fields.

Δεν παρέλειψα ακόμα την επαλήθευση αποτελεσμάτων ανάμεσα στη ΒΔ και το Tableau:



Εικ. 38: Επαληθεύσεις μεταξύ ΒΔ και Tableau

5.1.1 Πλεονεκτήματα & μειονεκτήματα του Tableau

Πλεονεκτήματα⁴:

- Προσφέρει διαδραστικότητα, με βοήθεια λειτουργιών, όπως filters, highlighters κλπ
- Δεν χρειάζονται ιδιαίτερες προγραμματιστικές ικανότητες κατά τη χρήση του, καθιστώντας το προσιτό προς όλους
- Είναι εύκολο στη χρήση και όμορφο στο αποτέλεσμα
- Παρέχει live σύνδεση μεταξύ του Tableau και της πηγής δεδομένων

⁴ [Online Courses - Learn Anything, On Your Schedule | Udemy](#)

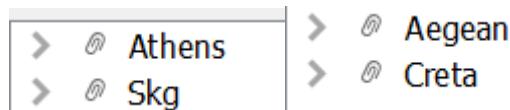
- Είναι εύκολο στην κοινή χρήση
- Agile & high performance

Μειονεκτήματα

- Δεν αποτελεί κάποια ΒΔ, δεν είναι εργαλείο ETL
- Δεν χρησιμοποιείται για λειτουργικό reporting, όπως το excel
- Ο χρήστης χρειάζεται να αποθηκεύσει χειροκίνητα διαφορετικές εκδοχές των workbooks, αν θέλει να πραγματοποιήσει rollback

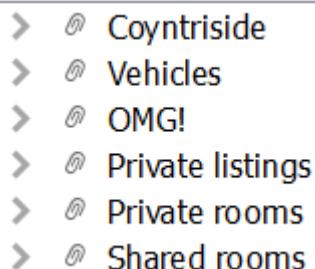
5.1.2 Groups

Τα groups αναφέρονται σε μια δομή, των οποίων οι εγγραφές οργανώνονται με βάση κοινά χαρακτηριστικά τους. Στον DimHost, παρατήρησα ότι οι περιοχές που βρίσκονται τα καταλύματα είναι πολλές, διαφορετικές και σχετικά αναλυτικές αλλά όλες ανήκουν στην Αθήνα, τη Θεσσαλονίκη, το Αιγαίο ή την Κρήτη (Area), που με τη σειρά τους αυτές ανήκουν στην Ελλάδα (Country). Άρα έφτιαξα αυτά τα groups, για να μπορώ να κάνω πιο γενικές απεικονίσεις.



Εικ. 39: Location groups

Ομοίως χρειάστηκε να αναλύσω τον τύπο κατοικίας, πεδίο που αρχικά αποτελούταν από πολλά και διαφορετικά στοιχεία. Έτσι, επειδή δυσκολευόμουν στην οργάνωση, τα ομαδοποίησα με βάση κοινά χαρακτηριστικά: σπίτια στην εξοχή, ιδιωτικά σπίτια, ιδιωτικά δωμάτια, κοινόχρηστα δωμάτια, σπίτια-μεταφορικά μέσα και σπίτια πολυτελείας (ονόμασα αυτό το group OMG!, καθώς υπάρχει ίδια κατηγορία και στην Airbnb). Προφανώς μια κατοικία μπορεί να περιλαμβάνεται σε παραπάνω από ένα group, αλλά το μελέτησα αρκετά για να καταλήξω σε αυτές τις επιλογές:



Εικ. 40: Property type groups

5.1.3 Ιεραρχίες

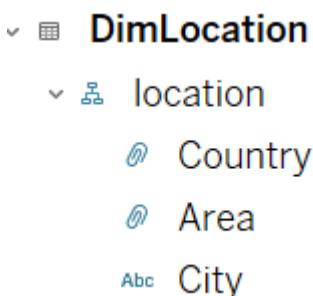
Από την άλλη οι ιεραρχίες αναφέρονται σε μια διάταξη, της οποίας τα στοιχεία κατηγοριοποιούνται σε επίπεδα με βάση τη σημασία, τη σχέση κλπ. Το κατώτερο στοιχείο συνήθως είναι υποκατηγορία του ανώτερου, ενώ το πεδίο που είναι πάνω από όλα αποτελεί την ομπρέλα που καλύπτει τα παρακάτω (ή και αντίστροφα). Χρησιμοποιείται για καλύτερη οργάνωση και ευκολότερη κατανόηση. Επίσης σε πλατφόρμες οπτικοποίησης, ξεκλειδώνει τη λειτουργία του drill down/up κάνοντας το εκάστοτε γράφημα περισσότερο/λιγότερο αναλυτικό αντίστοιχα. Η πιο συνηθισμένη ιεραρχία, είναι αυτή του χρόνου. Σχεδόν πάντα ο αναλυτής απασχολείται με δεδομένα αναφορικά με ημερομηνίες. Οπότε πχ. ξεκινάμε από το έτος -> εποχή -> μήνας -> ημέρα κοκ.

Συγκεκριμένα, στη ΒΔ αναφορικά με την Airbnb, υπάρχουν πράγματι πίνακες με χρονολογία: ο DimHost με το field host_since, ο FactCalendar με το field date, ο FactListings με τα first_review και last_review καθώς και ο FactReviews με το review_date. Το Tableau είναι αρκετά έξυπνο για να προσφέρει από μόνο τους ως προεπιλογή την παρακάτω ιεραρχία:

●	Year	2015
	Quarter	Q2 2015
	Month	May 2015
	Week Number	Week 5, 2015
	Day	May 8, 2015
	More	

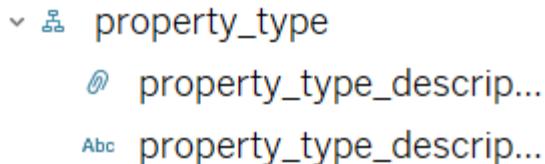
Εικ. 41: Χρονική ιεραρχία

Άλλες ιεραρχίες που εγώ δημιούργησα, είναι αυτή στον DimLocation, με όνομα location. Αφού κατηγοριοποίησα στο group Area τις πόλεις των τεσσάρων περιοχών που μελετούνται (Athens, Creta, Aegean, Skg), και έπειτα αυτές στο group με όνομα Country, δημιούργησα την τοπική ιεραρχία. Ξεκίνησα και πήρα τα πολλά, αναλυτικά και τα έκανα όλο και πιο γενικά, λίγα. Αποφάσισα επίσης να ορίσω το group Area ως ένα ανεξάρτητο field στον πίνακα:



Εικ. 42: Τοπική ιεραρχία

Έπειτα δημιούργησα μια μικρή ιεραρχία στον Dim.PropertyTypeType ανάμεσα στο field property_type_description και το ευρύτερο group που έκανα προηγουμένως:

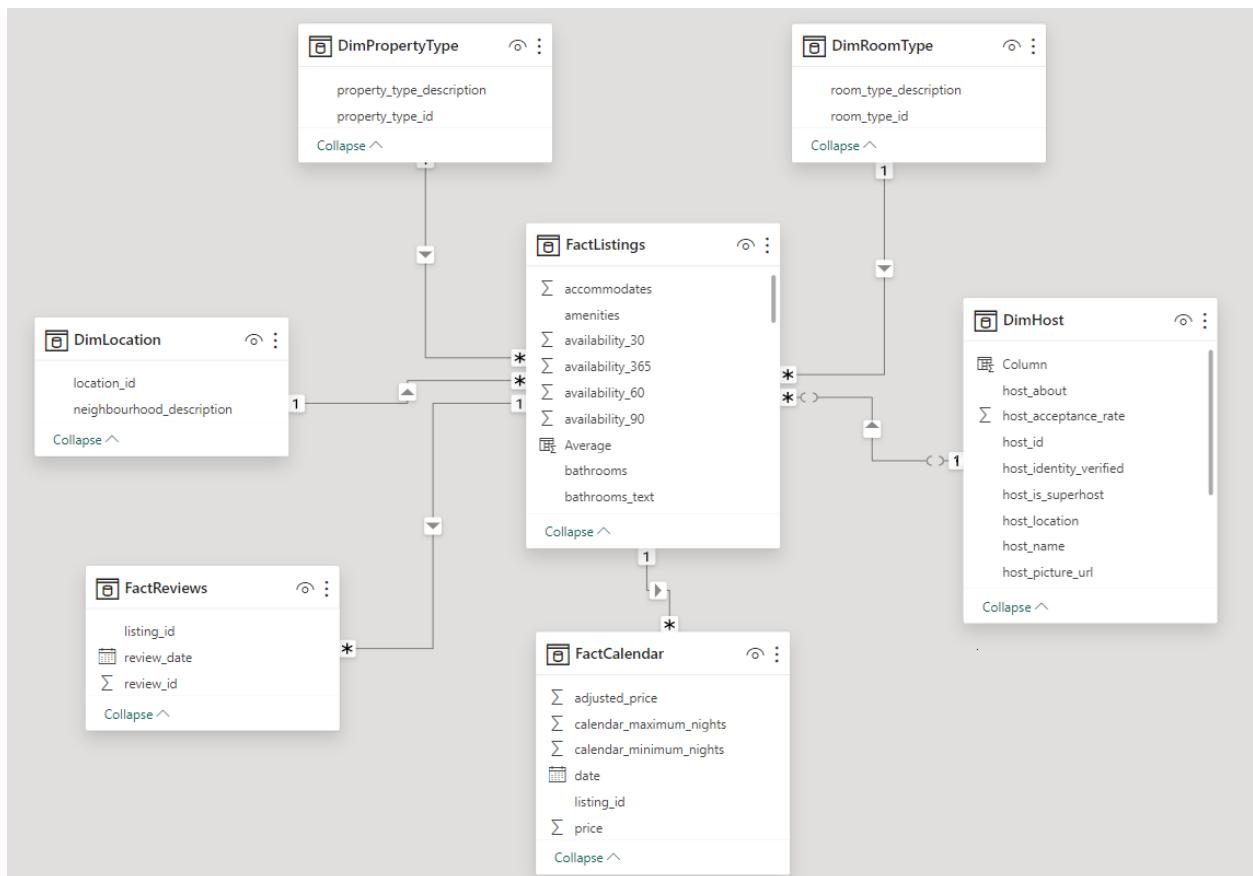


Εικ. 43: Property type ιεραρχία

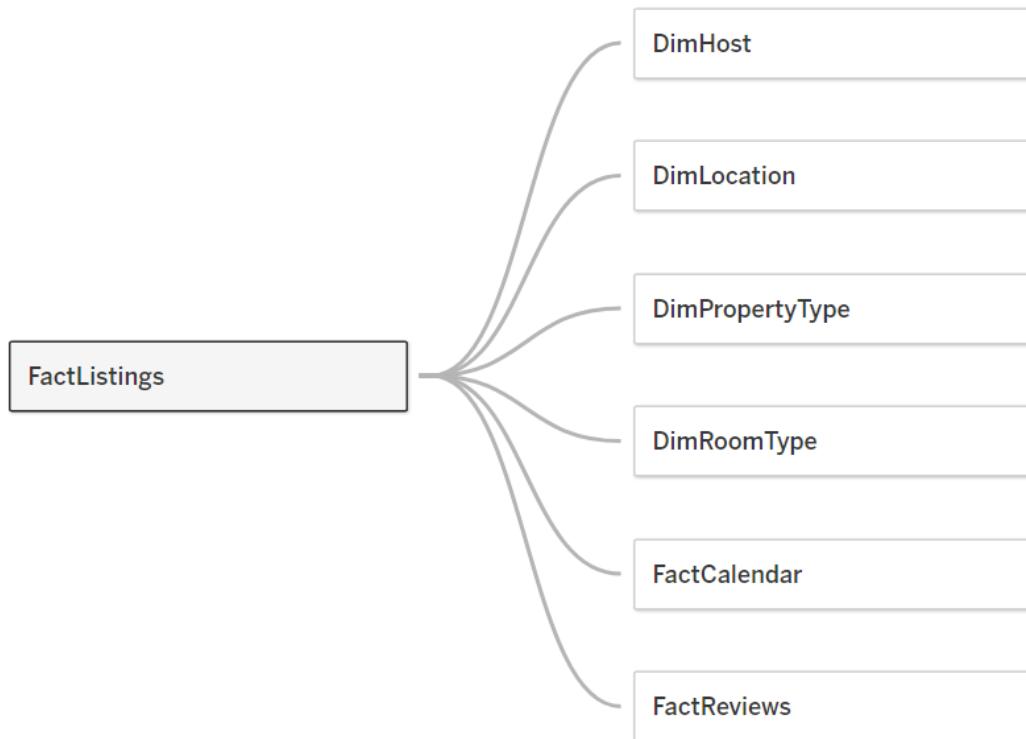
Μελετώντας τους πίνακες και τα δεδομένα, ανακάλυψα κι άλλες δυο ιεραρχίες. Αυτές που περιλαμβάνουν το availability ανά μέρες, ξεκινώντας από την κορυφή με το availability_365 -> availability_90 -> availability_60 -> availability_30. Ομοίως με το number_of_reviews -> number_of_reviews_ltm -> number_of_reviews_l30d, δηλαδή τον αριθμό κριτικών του listing γενικά, για τους τελευταίους 12 μήνες και τις τελευταίες 30 μέρες αντίστοιχα.

5.2 Star Schema

Συνδέοντας τη βάση με το Tableau, το πρώτο πράγμα που πρέπει να γίνεται είναι η σύνδεση των πινάκων με relations, έτσι ώστε να δημιουργηθεί το schema. Ο κύριος πίνακας είναι ο FactListings, ο οποίος περιέχει τα σημαντικότερα αριθμητικά δεδομένα που θα αναλύσω. Επομένως αυτός αποτελεί τον κύριο γνώμονα για τη δημιουργία των σχέσεων. Ο FactListings συνδέεται με τον FactReviews και τον FactCalendar με τη σχέση (1,N). Αυτό συμβαίνει επειδή υπάρχει μόνο ένα, μοναδικό listing_id στον FactListings, το οποίο όμως μπορεί να το συναντήσει κανείς πολλές φορές στους άλλους δυο πίνακες, αφού οι κριτικές είναι παραπάνω από μία, προερχόμενες από παραπάνω από ένα άτομο για το ίδιο κατάλυμα. Αντίστοιχα ο FactCalendar περιλαμβάνει τις διαθέσιμες μέρες και τιμές του κάθε καταλύματος για διάρκεια ενός χρόνου, άρα το id εμφανίζεται πολλαπλές φορές. Δεν είναι τυχαίο πως αυτό το CSV ήταν και το μεγαλύτερο σε μέγεθος. Foreign key αποτελεί το listing_id. Από την άλλη, ξεκινώντας πάλι από τον FactListings, προς τους 3 dimension πίνακες DimLocation, Dim.PropertyType και Dim.RoomType, η σχέση τώρα είναι (M,1) κι αυτό με την αντίστοιχη λογική ότι πολλά και διαφορετικά listings αντιστοιχούν σε μια ίδια τοποθεσία, τύπο ιδιοκτησίας και τύπο δωματίου, με FK το location_id, property_type_id και room_type_id αντίστοιχα. Επιπλέον, στον DimHost βλέπω ότι ένας host μπορεί να έχει παραπάνω από ένα listing του FactListings, δηλαδή είναι δυνατό ένας οικοδεσπότης να έχει παραπάνω από μια ιδιοκτησία του στην πλατφόρμα, με το host_id να συνδέει τους δυο πίνακες με την ίδια σχέση. Τελικά δημιουργείται ένα star schema ως εξής:



Εικ. 44: Star schema



Εικ. 45: Star schema από Tableau

5.3 Στρατηγικό κυβερνητικό dashboard

Για την υλοποίηση ενός στρατηγικού dashboard, γνωρίζω ότι πρόκειται για ένα πιο γενικό ταμπλό, με συνοπτική, ξεκάθαρη και σημαντική πληροφορία, σε ετήσιο επίπεδο. Απευθύνεται στην κυβέρνηση, η οποία θέλει να έχει μια σφαιρική εικόνα για τα καταλύματα της χώρας που είναι εγγεγραμμένα στην Airbnb.

Για αυτόν τον λόγο, αναπόσπαστο κομμάτι του dashboard αποτελούν τα KPIs που αφορούν αριθμούς-κλειδιά. Αποφάσισα να συμπεριλάβω τέσσερις σημαντικούς αριθμούς: το συνολικό πλήθος των καταλυμάτων όλων των τεσσάρων περιοχών, το μέσο όρο ενοικίασής τους ανά τη μέρα και τον μέσο όρο του πλήθους των καταλυμάτων που διαθέτει ένας χρήστης στην πλατφόρμα. Θεωρώ ότι και το τρίτο νούμερο είναι εξίσου σημαντικό, καθώς η κυβέρνηση πιθανόν χρειάζεται να γνωρίζει πόσα καταλύματα διατίθονται από το ίδιο άτομο, σε περίπτωση που θέλει να ασκήσει μελλοντικά βέτο στον αριθμό αυτό.

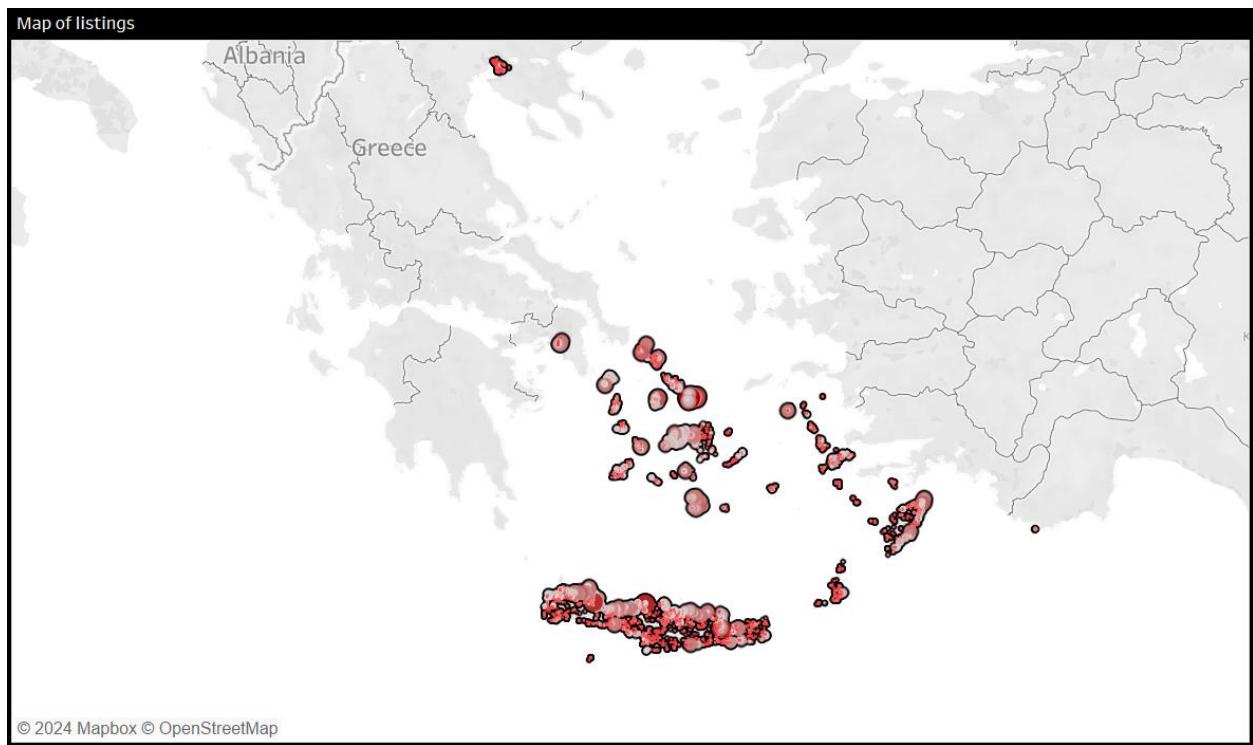
Για τη συγκεκριμένη υλοποίηση, χρησιμοποιήθηκαν τα πεδία listing_id, διαλέγοντας από την επιλογή measure το count και το πεδίο price, επιλέγοντας το average αντίστοιχα. Και τα δύο αυτά πεδία προέρχονται από τον πίνακα FactListings. Το τρίτο νούμερο είναι το average του host_total_listings_count του πίνακα DimHost. Τα KPIs έγιναν με τη χρήση πίνακα.

KPIs		
Count of listing_id	Avg. price	Avg. host_total...
76.822	216 €	6.12

Εικ. 46: KPIs για στρατηγικό dashboard

Εξίσου χρήσιμο visual που υλοποιώ είναι κι ο χάρτης. Αποφάσισα να χρησιμοποιήσω light theme, έτσι ώστε να φαίνονται οι περιοχές προσεγγιστικά και ταυτόχρονα το λευκό χρώμα να ταιριάζει με το λογότυπο της εταιρίας. Παραμετροποίησα το σχήμα και το χρώμα των συντεταγμένων με βάση τη τιμή: όσο μεγαλύτερο και εντονότερο χρώμα έχει ένα σημείο, τόσο πιο ακριβό είναι.

Στο tooltip φαίνεται ακριβώς η εκάστοτε γεωγραφική περιοχή, οι συντεταγμένες και η τιμή. Θεώρησα απαραίτητο να φαίνεται η τιμή, πατώντας πάνω στα σημεία. Για τον χάρτη χρησιμοποιήσα το latitude και longitude του FactListings. Επίσης χρησιμοποίησα το price του ίδιου πίνακα ως attribute για το χρώμα και το σχήμα. Το Location του DimLocation είναι επίσης στο tooltip.



Εικ. 47: Map of listings

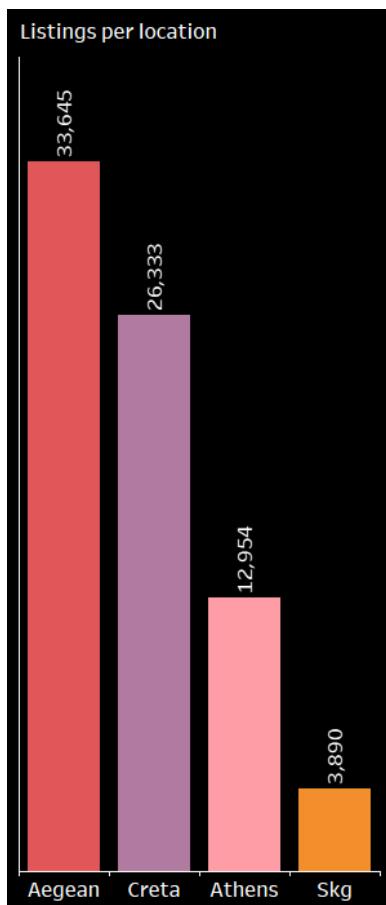
Επόμενο γράφημα που χρησιμοποιείται είναι το treemap, το οποίο περιέχει το μέσο όρο διανυκτέρευσης ανά κάθε περιοχή συγκεκριμένα, χρησιμοποιώντας πάλι το price του FactListings. Αυτό είναι και το πεδίο που καθορίζει το χρώμα και το μέγεθος. Πάλι του tooltip δείχνει το Location του DimLocation. Αυτά τα δύο προστέθηκαν επίσης στην επιλογή label, για να φαίνονται πάνω στο διάγραμμα μονίμως.



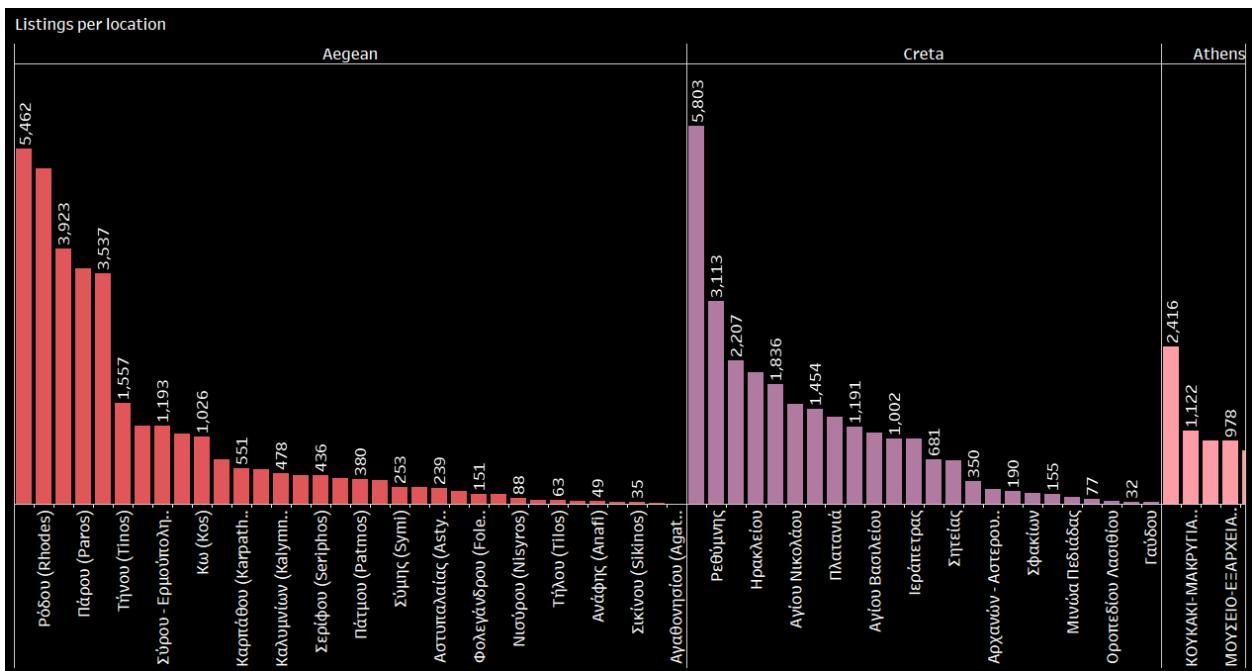
Εικ. 48: Avg. price per day

Με horizontal bars απεικονίζω την κατανομή του πλήθους των καταλυμάτων ανά περιοχή. Για αυτόν τον λόγο χρησιμοποιώ στα rows το count listing_id από τον FactListings και στα columns το πεδίο Area, από την ιεραρχία που δημιούργησα στο DimLocation. Σκοπίμως επέλεξα την ιεραρχία, κι όχι το απλό Location, καθώς δίνεται στον χρήστη η δυνατότητα να κάνει drill down και να δει αναλυτικά πόσα

καταλύματα υπάρχουν σε κάθε συγκεκριμένη περιοχή/δήμο της ευρύτερης γεωγραφικής περιοχής πατώντας το ‘+’ σύμβολο. Κι εδώ με τη βοήθεια του label φαίνεται ο ακριβής αριθμός για κάθε στήλη, ακόμα και όταν γίνεται drill down. Δεν θεωρώ ότι χρησιμεύει κάπου το drill up από τις τέσσερις βασικές περιοχές, καθώς αυτό θα οδηγούσε στη χώρα, δηλαδή την Ελλάδα, για την οποία έχει ήδη δωθεί το πλήθος των listings από τα KPIs. Το drill up θα ήταν χρήσιμο αν μελετούσαμε πολλές χώρες. Το χρώμα για κάθε περιοχή καθορίζεται από την τοποθεσία, όπως έχει αναφερθεί παραπάνω. Στο tooltip επιλέγω να φαίνεται το Location (που δεν είναι στην ιεραρχία) για να μπορεί αυτό το διάγραμμα στη συνέχεια να συνδέεται με τα υπόλοιπα στο dashboard με το highlighter. Ανάμεσα στα γραφήματα του συγκεκριμένου dashboard, είναι το μόνο που επέλεξα να γίνεται drill down (αποτελεί σημαντική πληροφορία), καθώς προσπαθώ να διατηρήσω το dashboard γενικότερο αλλά ταυτόχρονα δεν θέλω να στερήσω τη δυνατότητα να μελετήσει κανείς αναλυτικότερα τις τιμές για κάθε δήμο, αν το επιθυμεί.



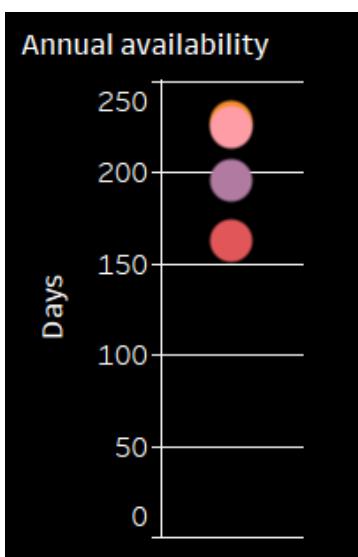
Εικ. 49: Listings per location



Εικ. 50: Αναλυτικά listings per location

Θεωρώ πως ένα σημαντικό metric που θα απασχολούσε την κυβέρνηση είναι η διαθεσιμότητα των καταλυμάτων σε βάθος χρόνου, δηλαδή τις πόσες μέρες από τις 365 το listing είναι διαθέσιμο. Αυτό βέβαια, σύμφωνα και με το data dictionary δεν σημαίνει απαραίτητα ότι είναι κλεισμένο από φιλοξενούμενο αλλά μπορεί ο ίδιος ο ιδιοκτήτης να το διατηρεί μη διαθέσιμο για προσωπικούς λόγους. Με αυτό το διάγραμμα (διάγραμμα απεικόνισης με σημεία-κύκλους) μπορεί κανείς να αντιληφθεί μοτίβα για τη δημοφιλία και τη ζήτηση του κάθε τόπου.

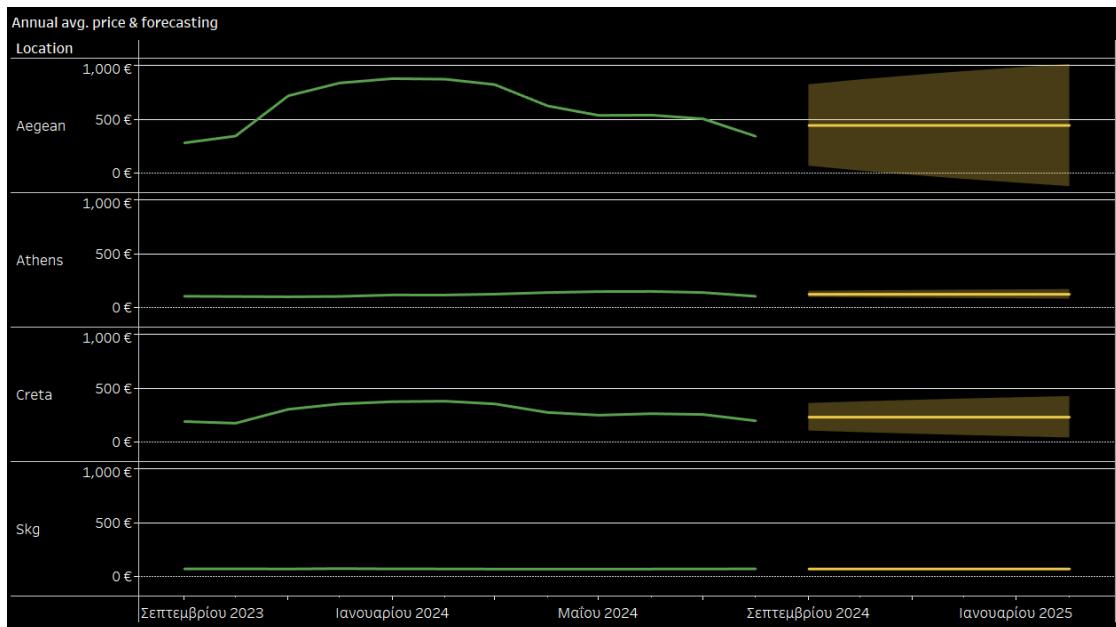
Για το γράφημα αυτό χρησιμοποιήθηκε το average του availability_365 του FactListings. Στο tooltip φαίνεται ακριβώς ο αριθμός των διαθέσιμων ημερών, ενώ το χρώμα των bubbles καθορίζεται από το Location.



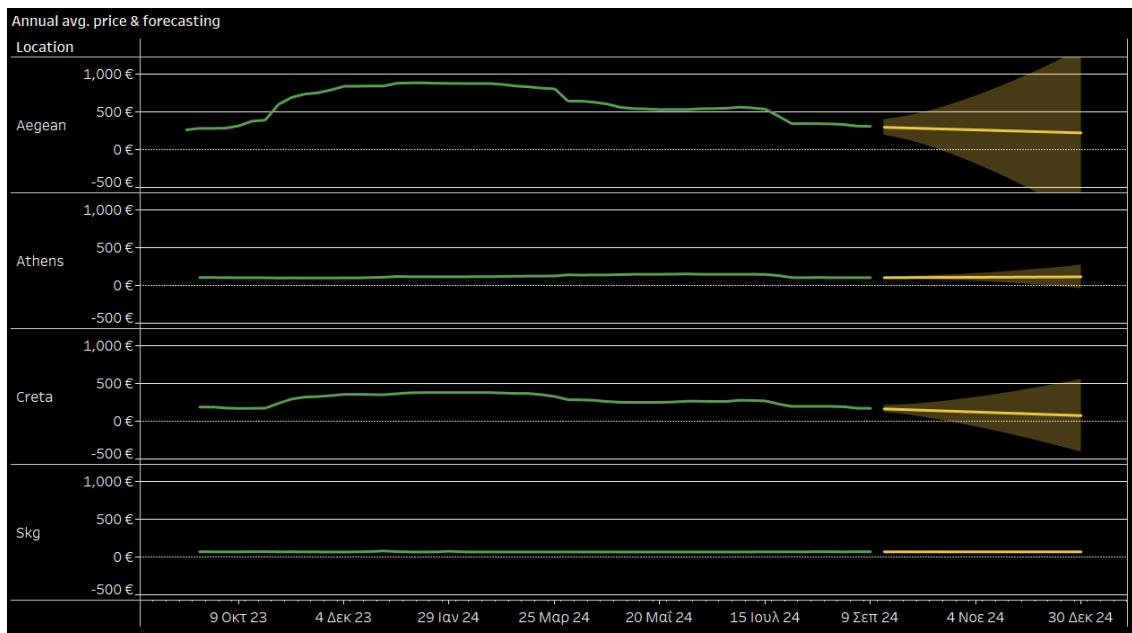
Εικ. 51: Annual availability

5.3.1 Forecast μοντέλο

Ο πίνακας που βοηθάει στη μελέτη της τιμής σε βάθος χρόνου, κι άρα στην εξαγωγή μοτίβων και συμπερασμάτων δεν μπορεί να είναι άλλος πέρα από τον FactCalendar, ο οποίος διαθέτει συγκεκριμένα τις τιμές και ημερομηνίες για έναν χρόνο, για όλες τις εγγραφές. Έτσι, παίρνοντας τα πεδία price ως average και date αυτού, προχωράω στη δημιουργία ενός line chart, το οποίο δείχνει ακριβώς πώς κυμαίνεται ο μέσος όρος των τιμών σε συνάρτηση με τον χρόνο για τα τέσσερις περιοχές (πεδίο Location του DimHost). Η δυνατότητα του Tableau να εκτελέσει κανείς drill down και υπ στον χρόνο μπορεί να προσδιορίσει ακριβώς τις τιμές σε επίπεδο χρόνου, μηνών και ημερών. Οι ημέρες του καλοκαιριού πχ. Θα ήταν χρήσιμο να εξεταστούν, αφού συνήθως τότε ανθίζει ο τουρισμός. By default, χωρίζω τη διάρκεια αυτού του χρόνου σε μήνες για να μελετήσω τις τιμές ανά εποχή μέχρι το φθινόπωρο του 2024. Επιπλέον με τη βοήθεια του Tableau εισάγω στο γράφημα μοντέλο πρόβλεψης (analytics -> model -> forecast), γεγονός το οποίο επεκτείνει το χρόνο μέχρι το χειμώνα του 2024 και προβλέπει πώς θα κυμανθούν οι τιμές τότε. Είναι γνωστό από τη θεωρία ότι ένα στρατηγικό dashboard μπορεί να επικεντρωθεί και στο μέλλον, ωθώντας με να επιλέξω τη λειτουργία αυτή. Αξίζει να σημειωθεί πως το μοντέλο μπορεί να ρυθμιστεί για βάθος χρόνων αλλά δεν αποτελεί επιθυμητή πληροφορία για το δεδομένο dashboard, αφού θέλω να μελετήσω τις τιμές μόνο για το έτος που διανύεται. Με τη βοήθεια του tooltip φαίνεται ακριβώς η περιοχή, ο χρόνος, η μέση τιμή, αλλά και το αν πρόκειται για πραγματική ή forecasting τιμή. Ο διαχωρισμός αυτός επίσης γίνεται κατανοητός και με το πράσινο/κίτρινο χρώμα αντίστοιχα.



Εικ. 52: Annual avg. price & forecasting



Εικ. 53: Αναλυτικό annual avg. price & forecasting

5.3.2 Calculated field

Στην προσπάθεια της κυβέρνησης να γνωρίζει τα είδη των καταλυμάτων που χρησιμοποιούνται για ενοικίαση -κυρίως λόγω προβλήματος εύρεσης και τιμής ενοικιαζόμενων σπιτιών στη χώρα- επέλεξα να φτιάξω ένα ακόμα treemap, αυτή τη φορά με ένα calculated field που θα δείχνει τις τιμές σε ποσοστά, τα οποία ποσοστά είναι συνήθως πιο εύκολα στην κατανόηση. Έχοντας ήδη το field room_type_description του DimRoomType με το listing_id του FactListings, έφτιαξα ένα δικό μου calculated field με όνομα Percentage of room type, παίρνοντας τα πλήθη της κάθε κατηγορίας και μετατρέποντάς τα σε ποσοστά. Για αυτό χρησιμοποίησα το εξής query στο Tableau:

```
:percentage of room type
```

Results are computed along Table (across).

```
COUNTD([listing_id]) / WINDOW_SUM(COUNTD([listing_id]))
```

Εικ. 54: Calculated field

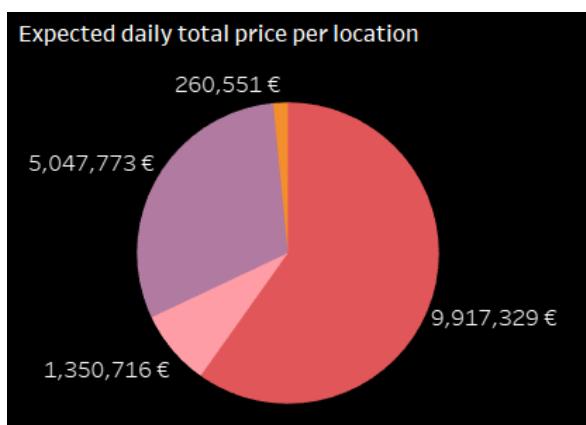
Με στήλη το είδος καταλύματος και γραμμή τις τέσσερις περιοχές δημιούργησα το γράφημα αυτό. Το πλήθος των καταλυμάτων του FactListings χρησιμοποιείται για την ένταση του χρώματος. Στο tooltip αναφέρονται το Location για τη διαδραστικότητα στο dashboard, το room_type_description, το πλήθος σε αριθμό των listings που ανήκουν στη συγκεκριμένη κατηγορία και το αντίστοιχο ποσοστό.

Δεν επέλεξα να απεικονίσω το property_type_description του Dim.PropertyType, καθώς θεωρώ ότι περιέχει πολύ αναλυτικότερη πληροφορία, γεγονός που ταιριάζει καλύτερα σε ένα λειτουργικό dashboard.

Room type percentage					
Location	Entire home/apt	Hotel room	Private room	Shared room	
Aegean	83.49%	2.18%	14.29%	0.04%	
Athens	89.94%	0.81%	8.59%	0.66%	
Creta	88.47%	1.42%	9.93%	0.18%	
Skg	97.30%	0.18%	2.44%	0.08%	

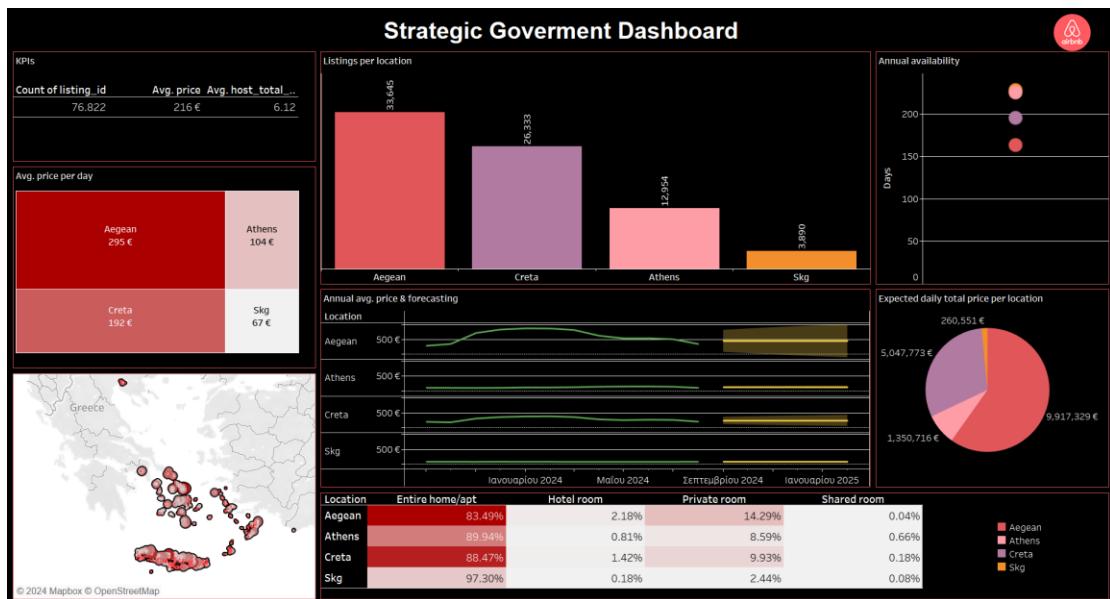
Εικ. 55: Room type percentage

Με διάγραμμα πίτας απεικονίζω το νούμερο που αντιστοιχεί σε κάθε περιοχή αν αθροίσω τις καταγεγραμμένες τιμές των listings (ημερήσια). Έτσι, αντιλαμβάνεται κανείς τα επιθυμητά ποσά που δύναται να κερδίσουν οι οικοδεσπότες ανά περιοχή. Το μέγεθος αντιστοιχεί στο sum του price του FactListings, ενώ το χρώμα πάλι στις τέσσερις περιοχές (πεδίο Location του DimHost). Και τα δύο πεδία φαίνονται στο tooltip.



Εικ. 56: Expected daily total price per location

Με τα παραπάνω γραφήματα και δεδομένα δημιουργώ το στρατηγικό κυβερνητικό ταμπλό. Η διάταξη μεταξύ τους έγινε με τέτοιο τρόπο ώστε να έχουν μια ομαλή ροή και να είναι όμορφα και κατανοητά για τον χρήστη. Έτσι, έβαλα φωτεινά, κοκκινωπά πλαίσια γύρω από το καθένα και φρόντισα να υπάρχει διαδραστικότητα μεταξύ τους. Αυτό σημαίνει πως όταν επιλέγω την τοποθεσία σε ένα γράφημα αυτόματα γίνεται highlighted η ίδια τοποθεσία και στα υπόλοιπα. Όπου θεώρησα ότι ο τίτλος του άξονα κάποιου γραφήματος είναι περιπτώς, το αφαίρεσα όπως στο διάγραμμα Listings per location, αφού επέλεξα να εμφανίζω τον τίτλο και τον ακριβή αριθμό καταλυμάτων πάνω από κάθε bar. Επίσης αφαιρέθηκε ο τίτλος του άξονα γ στο διάγραμμα Annual avg. price, αφού με το σύμβολο του ευρώ, γίνεται εύκολα κατανοητό ότι πρόκειται για χρήματα. Επέλεξα να φαίνεται το ευρώ για κάθε στήλη/γραμμή και δεδομένα που αφορούν χρήματα. Στα ποσοστά καταλύματος ανά περιοχή παρέλειψα τον τίτλο με το ίδιο σκεπτικό. Έτσι, το dashboard απλοποιήθηκε ως εξής:



Εικ. 57: Strategic Government Dashboard

5.4 Λειτουργικό dashboard για real estate επενδυτές

Το λειτουργικό dashboard οφείλει να είναι αναλυτικότερο και να απεικονίζει πολλαπλά και διαφορετικά πεδία. Συνεπώς, και γι' αυτό το ταμπλό, χρησιμοποιήθηκαν τα ίδια KPIs με το στρατηγικό, δηλαδή το συνολικό πλήθος των καταλυμάτων προς ενοικίαση, η μέση τιμή ενοικίασης στην Ελλάδα ανά διανυκτέρευση και ο μέσος όρος του πλήθους των καταλυμάτων που διαθέτει ένας χρήστης στην πλατφόρμα. Χρησιμοποιήθηκαν τα πεδία listing_id, διαλέγοντας από την επιλογή measure το count και το πεδίο price, επιλέγοντας το average αντίστοιχα. Και τα δύο αυτά πεδία προέρχονται από τον πίνακα FactListings. Το τρίτο νούμερο είναι το average του host_total_listings_count του πίνακα DimHost.

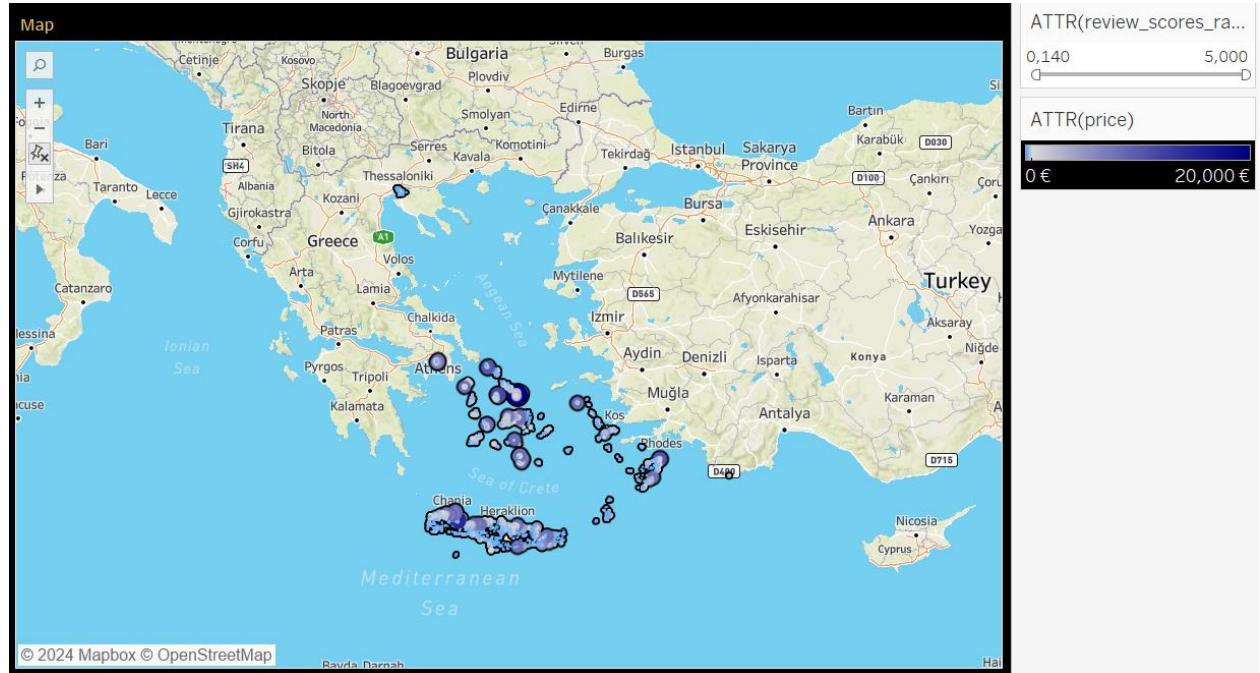
Επιπλέον, πρόσθεσα τον μέσο όρο βαθμολογίας για όλες τις περιοχές και τον μέσο όρο ημερών που είναι διαθέσιμα τα καταλύματα σε βάθος μήνα. Έτσι μπορεί κανείς να αντιληφθεί την ποιότητα και τη ζήτηση αντίστοιχα, πληροφορίες που ενδιαφέρουν αρκετά τυχόν επενδυτές. Για αυτά πήρα τον μέσο όρο των πεδίων review_scores_rating και availability_30 του FactListings.

Count of listing...	Avg. price	Avg. host_total..	Avg. review_sco..	Avg. availability..
76.822	216 €	6.12	4.785	14

Εικ. 58: KPIs για λειτουργικό dashboard

5.4.1 Slicer

Κατά τον ίδιο ακριβώς τρόπο φτιάχτηκε και ο χάρτης. Αυτή τη φορά χρησιμοποιώ πολιτικό χάρτη, καθώς μπορεί κανείς με zoom να βρει ακριβώς τη τοποθεσία του καταλύματος για το οποίο ενδιαφέρεται. Ακόμα πρόσθεσα ένα slicer, με τη βαθμολογία των καταλυμάτων (reviews_scores_rating του FactListings), με σκοπό να μπορεί κανείς να δει τα καταλύματα με βαθμολογία της αρεσκείας του. Έτσι, σύρροντας στο παράθυρο με τις βαθμολογίες, εμφανίζονται και εξαφανίζονται καταλύματα με βάση τον αριθμό που ορίζεται. Είναι σημαντικό για τους επενδυτές να θέλουν να βρουν τις περιοχές πχ. με την υψηλότερη βαθμολογία.



Εικ. 59: Map

Επειδή στους επενδυτές αρέσουν τα νούμερα, ιδίως τα μεγαλύτερα, αποφάσισα να δημιουργήσω treemap με τον ακριβή αριθμό του αθροίσματος των τιμών των καταλυμάτων ανά περιοχή για τον επόμενο χρόνο (αυτή είναι η επιθυμητή συνολική τιμή). By default, ο χρόνος είναι χωρισμένος σε μήνες. Μάλιστα η περιοχή περιλαμβάνει την ιεραρχία από τον DimLocation, δηλαδή πατώντας το +, μπορεί κανείς να δει τις πόλεις και τις τιμές τους, όπως επίσης δίνεται η δυνατότητα αυτή και για τον χρόνο. Αφού εμπλέκεται ο χρόνος, χρησιμοποίησα το date και price του FactCalendar. Στο tooltip φαίνεται και το Location.

Month of date	Area			
	Aegean	Athens	Creta	Skg
Σεπτέμβριος 2023	11.205.146	2.977.381	4.050.698	2.000.789
Οκτώβριος 2023	30.518.397	8.998.888	15.432.154	6.181.084
Νοέμβριος 2023	61.891.050	8.482.944	26.124.967	5.866.751
Δεκέμβριος 2023	74.547.300	9.081.385	31.417.986	6.373.438
Ιανουάριος 2024	78.156.702	10.266.486	33.280.351	6.131.678
Φεβρουάριος 2024	72.667.566	9.587.481	31.510.837	5.710.067
Μάρτιος 2024	73.199.746	10.969.338	31.446.285	5.995.773
Απρίλιος 2024	53.759.610	11.897.486	23.575.424	5.770.514
Μάιος 2024	47.674.636	13.110.497	22.102.344	5.990.569
Ιούνιος 2024	46.343.605	12.798.402	22.594.604	5.821.499
Ιούλιος 2024	44.811.928	12.314.445	22.672.264	6.047.042
Αύγουστος 2024	30.319.768	9.188.408	17.427.050	6.170.140
Σεπτέμβριος 2024	13.380.119	5.619.260	10.491.829	3.764.669

Εικ. 60: Expected monthly total price per location

Month o..	Area / City												Aegean
	Αμοργού (Α.. Ανάφης (Απ.. Άνδρου (Απ.. Αντιπάρου .. Αστυπαλαί.. Θήρας (San.. Ιητών (Ios) Καλυμνίων.. Καρπάθου (.. Κάσου (Kas.. Κέας (Kea) Κιμώλου												
Σεπτέμβριος	61.419	1.050	233.308	143.892	16.326	3.055.770	73.729	52.948	43.796	13.860	339.442	17.	
Οκτώβριος	176.542	1.960	492.347	323.242	33.666	6.332.008	151.732	114.765	83.886	28.334	2.284.228	42.	
Νοέμβριος	341.985	1.800	708.500	311.300	35.100	19.980.367	161.970	104.320	118.160	27.420	2.367.486	46.	
Δεκέμβριος	354.640	1.860	733.752	313.224	36.270	28.968.138	168.513	108.347	120.886	28.334	2.538.997	47.	
Ιανουάριος	355.105	1.860	742.383	313.224	35.650	30.628.185	168.650	106.242	119.205	28.334	2.710.394	47.	
Φεβρουάριος	332.195	1.740	695.630	293.016	33.350	28.785.600	157.774	100.398	117.370	26.506	2.307.931	44.	
Μάρτιος	355.105	1.860	739.007	313.224	37.820	28.147.598	168.668	108.903	123.836	28.890	2.471.302	47.	
Απρίλιος	274.070	1.800	710.460	307.120	36.600	16.603.380	165.882	106.180	121.810	28.200	2.405.381	45.	
Μάιος	117.609	1.860	735.143	344.503	38.750	15.099.880	184.328	109.565	126.184	29.140	2.449.896	47.	
Ιούνιος	117.157	1.800	752.266	333.010	38.100	14.653.546	198.427	106.080	124.625	28.240	2.393.287	48.	
Ιούλιος	123.745	1.860	644.429	372.143	40.920	13.874.404	249.836	113.199	127.571	29.140	2.221.723	50.	
Αύγουστος	127.209	1.860	579.780	373.668	43.090	8.271.582	249.194	113.880	132.144	29.140	855.176	53.	
Σεπτέμβριος	55.266	840	250.915	169.963	19.700	3.923.527	95.045	53.540	62.660	13.662	377.212	23.	

Εικ. 61: Αναλυτικό expected monthly total price per location

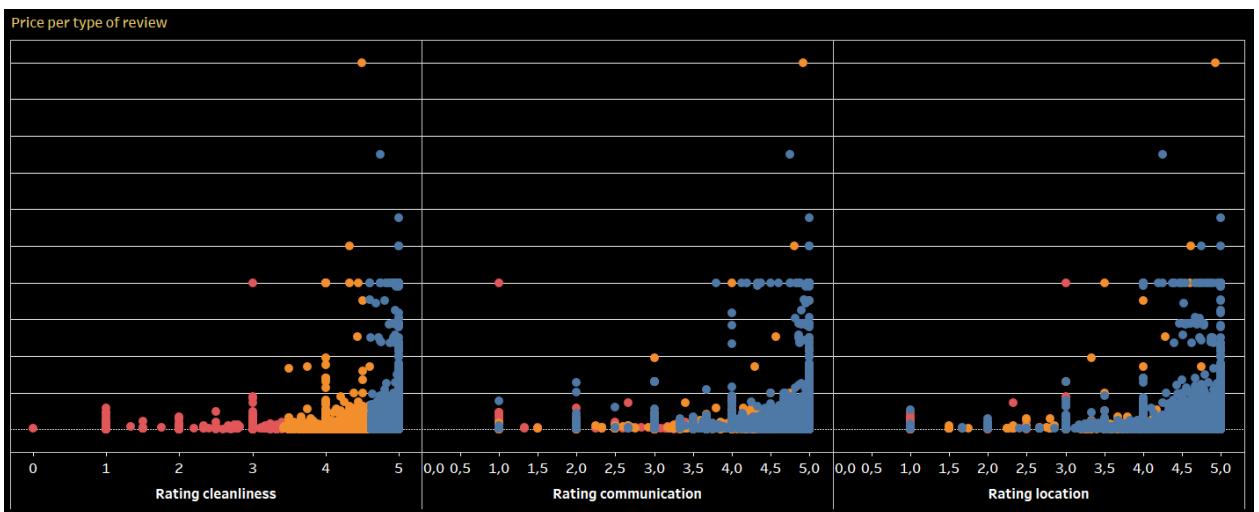
Ένα line chart που δείχνει το μέσο όρο της τιμής (κι όχι το άθροισμα) των καταλυμάτων ανά περιοχή για βάθος χρόνου είναι απαραίτητο. Χρησιμοποιώ το price και date του FactCalendar και το Location για χρώμα. Με drill down και up μπορεί κανείς να εξερευνήσει περισσότερο/λιγότερο αναλυτικά τον χρόνο.



Εικ. 62: Monthly avg. price per location

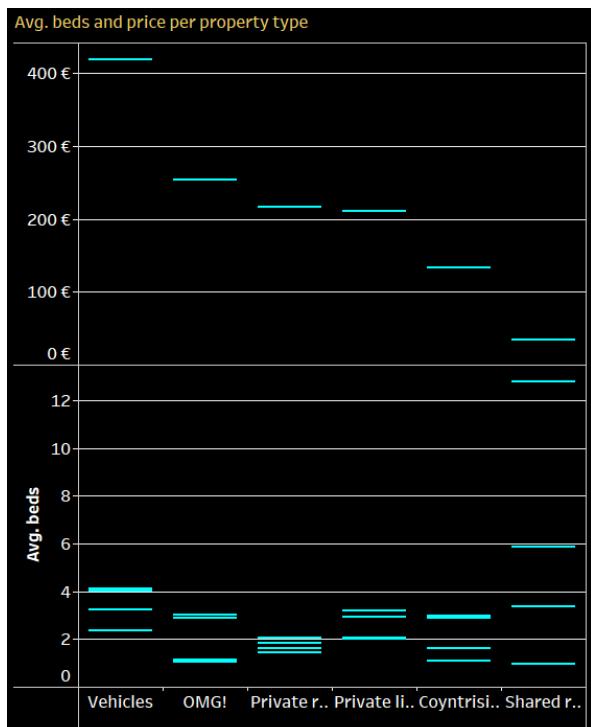
5.4.2 Clusters

Εντοπίζοντας τα εξειδικευμένα πεδία που διαθέτει η Airbnb για τις βαθμολογίες ως προς την καθαριότητα, την επικοινωνία και την τοποθεσία του καταλύματος, δημιούργησα τρία scatter plots για να δω πώς κυμαίνονται αυτά με βάση την τιμή (και τα τέσσερα παραπάνω πεδία προέρχονται από τον FactListings). Επιπλέον επέλεξα τη δυνατότητα του Tableau για k-means clustering, τη δημιουργία δηλαδή ομάδων με βάση κοινά χαρακτηριστικά που εντοπίζει το ίδιο. Δημιουργούνται έτσι clusters διαφορετικών χρωμάτων, με αποτέλεσμα να μπορεί κανείς να συγκρίνει πώς κυμαίνονται οι τιμές των ομάδων, όχι μόνο μεμονωμένα αλλά και μεταξύ των διαφορετικών ειδών ratings.



Εικ. 63: Price per type of review

Ένα σπανιότερο γράφημα επιλέγω για την επόμενη σχέση μεταξύ πολλαπλών πεδίων. Το gantt chart με βοηθάει να απεικονίσω τον μέσο όρο τιμής διανυκτέρευσης και αριθμό κρεβατιών (price και beds του FactListings) ανά τύπο ιδιοκτησίας και συγκεκριμένα τις κατηγορίες του γενικότερου group που δημιούργησα αρχικά. Εκτελώντας drill down, μπορεί κανείς να δει συγκεκριμένα τους ειδικούς τύπους ιδιοκτησίας, αν το επιθυμεί. Επιλέγω να απεικονίσω τον αριθμό κρεβατιών κατά μέσο όρο, γιατί αυτός υποδηλώνει τα άτομα ανά κεφαλή που μπορούν να μείνουν στον συγκεκριμένο τύπο ιδιοκτησίας. Μάλιστα, φαίνονται τα νούμερα για κάθε περιοχή ξεχωριστά, αφού πρόσθεσα το Location ως detail. Έτσι, έχω ένα gantt chart χωρισμένο στη μέση: το πρώτο δείχνει τη σχέση μεταξύ τιμής και τύπο ιδιοκτησίας και το άλλο τον μέσο όρο κρεβατιών/ατόμων σε κάθε περιοχή πάλι ως προς τον τύπο ιδιοκτησίας.



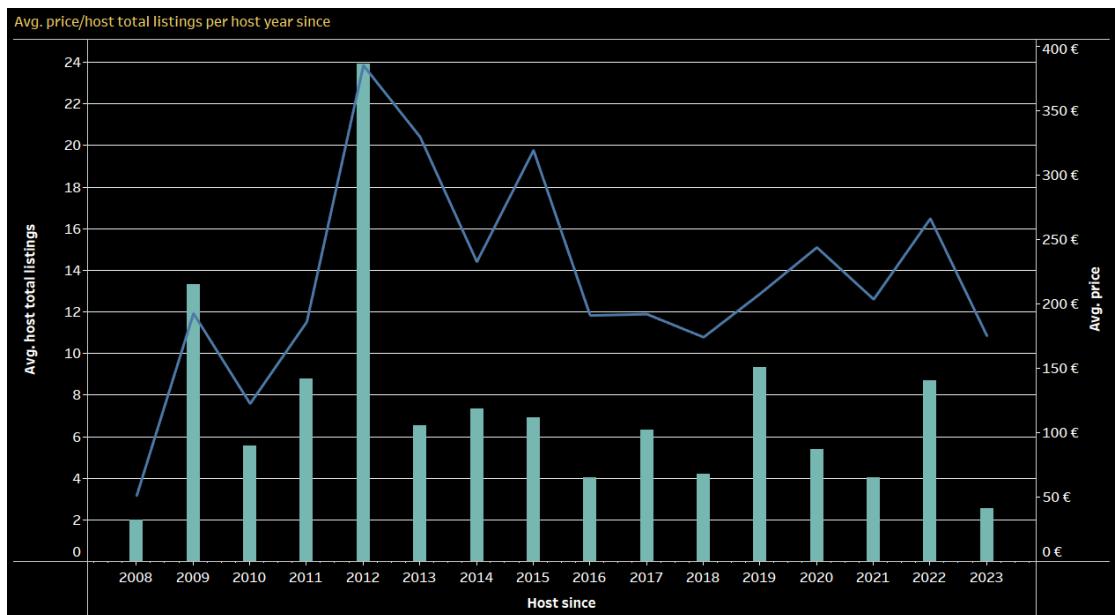
Εικ. 64: Avg. beds and price per property type

Στο επόμενο γράφημα αποφάσισα να συγκρίνω διάφορα σημαντικά νούμερα ανάμεσα στους οικοδεσπότες που έχουν χαρακτηριστεί ως superhosts και σε αυτούς που δεν έχουν αυτόν τον τίτλο. Οι superhosts είναι χρήστες που έχουν υψηλή βαθμολογία, συχνή επισκεψιμότητα, χαμηλό ποσοστό ακύρωσης κλπ. Σκεπτόμενη ότι οι επενδυτές χρειάζεται να γνωρίζουν εάν για τα άτομα που θα προσεγγίσουν έχει σημασία αν είναι superhosts ή όχι και αν ναι, πόσο σημαντική είναι η διαφορά τους, φτιάχνω ένα γράφημα packed bubbles, στο οποίο το μέγεθος της κάθε φούσκας εξαρτάται από τις τιμές που ορίζει ο αναλυτής. Συγκεκριμένα πήρα το πεδίο host_is_superhost του DimHost, το οποίο καθορίζει το χρώμα και το price του FactListings ως average για το μέγεθος. Το κύριο μέγεθος που μελετάται είναι ο μέσος όρος της τιμής ανά διανυκτέρευση ανάμεσα σε αυτές τις δυο κατηγορίες. Ταυτόχρονα συμπλήρωσα στο tooltip και label και τον μέσο όρο βαθμολογίας και διαθεσιμότητας για τριάντα μέρες, πεδία του FactListings επίσης.



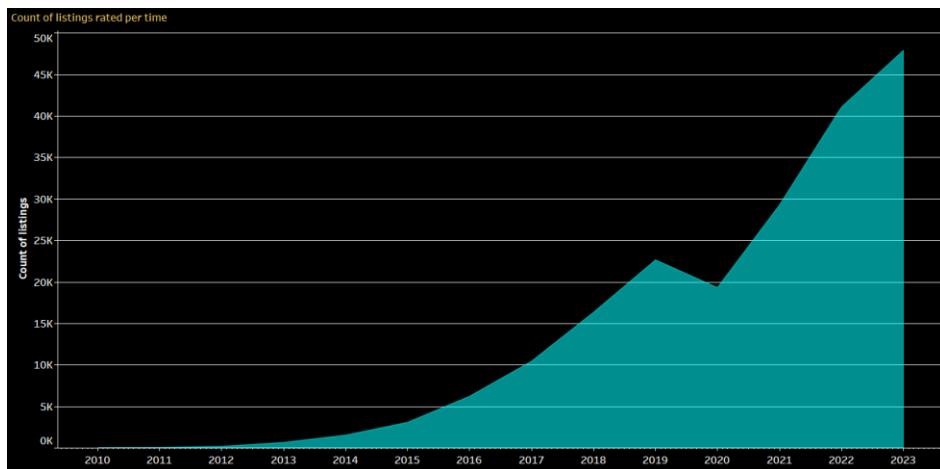
Εικ. 65: Superhost statistics

Ένας ακόμα ενδιαφέρον παράγοντας προς ανάλυση αποτελεί η χρονολογία που ξεκίνησε κανείς να είναι οικοδεσπότης στην Airbnb, πληροφορία που βρίσκεται στο host_since του DimHost. Αποφάσισα να συγκρίνω σε διάγραμμα διπλού συνδυασμού (bars και line), ταυτόχρονα τα συνολικά ακίνητα που έχει ένας άνθρωπος στην πλατφόρμα συγκριτικά με τα χρόνια εμπειρίας του και τον μέσο όρο τιμής ανά διανυκτέρευση που ορίζει για αυτά. Με αυτόν τον τρόπο μπορεί κανείς να μελετήσει τη συμπεριφορά των οικοδεσποτών στο πέρασμα του χρόνου, εξάγοντας συμπεράσματα για τάσεις και μοτίβα. Για τις γραμμές του γραφήματος χρησιμοποιήθηκε το average του host_total_listings_count από τον DimHost και το average του price από τον FactListings.



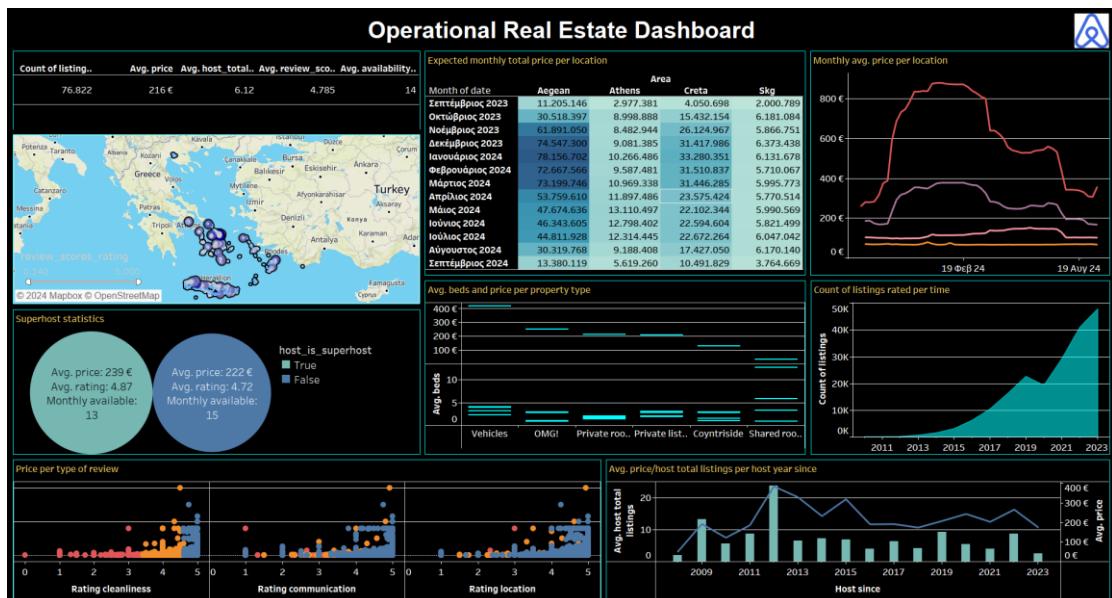
Εικ. 66: Avg. price/host total listings per host year since

Ο πίνακας FactReviews συντελεί στη μελέτη της εξέλιξης του αριθμού των καταλυμάτων που βαθμολογήθηκαν με το πέρασμα του χρόνου. Έτσι, μπορεί κανείς να αναλύσει την ποσότητα των βαθμολογιών, κι άρα τη σημασία της κριτικής, σε σχέση με τον χρόνο. Χρησιμοποιώ το review_date του FactReviews και το count distinct του listing_id από τον FactListings σε ένα area chart. Δεν θεωρώ ότι βοηθάει κάπου να χωρίσω το area chart σε περιοχές, αφού ούτως ή άλλως κάθε περιοχή έχει χιλιάδες παραπάνω καταλύματα από την άλλη, οπότε δεν θα ήταν και τόσο εύστοχο να τις συγκρίνω. Αντίθετα, θέλω να μελετήσω την τάση αυτή γενικότερα. Έτσι, γνωρίζοντας το πλήθος των σύγχρονων συνολικών καταλυμάτων, μπορώ να υπολογίσω πόσα συνολικά καταλύματα βαθμολογήθηκαν τη φετινή χρονιά.



Εικ. 67: Count of listings rated per time

Για το τελικό λειτουργικό dashboard για επενδυτές η διάταξη μεταξύ των γραφημάτων έγινε πάλι με τέτοιο τρόπο ώστε να υπάρχει μια ομαλή ροή. Έβαλα γαλάζια πλαίσια γύρω από το καθένα, αφού αυτή τη φορά διάλεξα το λογότυπο σε μπλε χρώμα. Ομοίως όταν επιλέγει κανείς την τοποθεσία σε ένα γράφημα αυτόματα γίνεται highlighted η ίδια τοποθεσία και στα υπόλοιπα. Αφαίρεσα τον τίτλο στα KPIs και τον χάρτη καθώς είναι περιττά. Επέλεξα επίσης να φαίνεται το σύμβολο του ευρώ για κάθε στήλη/γραμμή και δεδομένα που αφορούν χρήματα. Έτσι το dashboard διαμορφώνεται ως εξής:



Εικ. 68: Operational Real Estate Dashboard

ΚΕΦΑΛΑΙΟ 6: Συμπεράσματα

6.1 Εισαγωγή

Σε γενικές γραμμές, ο αριθμός των 76.882 καταλυμάτων στην Airbnb για τις τέσσερις περιοχές που αναλύονται είναι εντυπωσιακά υψηλός, αν σκεφτεί κανείς ότι δεν πρόκειται για όλη την Ελλάδα. Σύμφωνα με έρευνα που διεξήχθη από την Eteron, με τίτλο “Housing insecurity today”⁵, η Ελλάδα το 2010 ξεκίνησε με 132 ακίνητα στην πλατφόρμα και κατέληξε με πάνω από 125.000 λίγο πριν το ξέσπασμα της πανδημίας. Τα νούμερα αυτά μάλλον όλο και αυξάνονται.

Ο μέσος όρος της τιμής για Αθήνα, Αιγαίο, Θεσσαλονίκη και Κρήτη είναι εξίσου υψηλός, φτάνοντας τα 216€ ανά διανυκτέρευση! Ταυτόχρονα, ο μέσος οικοδεσπότης διαθετεί έξι περίπου καταλύματα στην πλατφόρμα, αριθμός επίσης αυξημένος, που θα πρέπει να απασχολεί την κυβέρνηση. Όπως φαίνεται υπάρχει μια τάση προς την προσωρινή παρά την μόνιμη ενοικίαση, δημιουργώντας προβλήματα όπως τη δυσκολία στην εύρεση κατοικίας.

6.2 Στρατηγικό κυβερνητικό dashboard

Βλέποντας το πρώτο γράφημα, το treemap, είναι αρκετά κατανοητό ότι οι τιμές στο Αιγαίο είναι αρκετά υψηλές, στα 295€. Μάλιστα από τη δεύτερη σε ακρίβεια περιοχή, την Κρήτη, απέχει λίγο πιο πάνω από 100€. Δεύτερη λοιπόν έρχεται η Κρήτη, με μέσο όρο διανυκτέρευσης 192€. Η Αθήνα αγγίζει περίπου τα 100€, αριθμός υποτριπλάσιος του Αιγαίου και μισός της Κρήτης. Η Θεσσαλονίκη με αριθμό διψήφιο, στα 67€, παίρνει την τελευταία θέση. Η ακρίβεια στην τιμή συνεπάγεται συνήθως και τη ζήτηση, επομένως η κατανομή των τιμών δείχνουν μάλλον και τη δημοφιλία των προορισμών. Από τον χάρτη είναι φανερό ότι το πιο ακριβό κατάλυμα βρίσκεται στο Αιγαίο, στη Μύκονο. Τα δύο επόμενα ακριβότερα βρίσκονται στην Κρήτη.

Ακριβώς με τον ίδιο τρόπο κατατάσσονται οι περιοχές κι ως προς το πλήθος των καταλυμάτων, με το Αιγαίο να φτάνει τα 33.645 εγγεγραμμένα καταλύματα και την Κρήτη τα 26.333. Η Αθήνα περιέχει 12.954 καταλύματα και η Θεσσαλονίκη μόλις 3.890. Μεταξύ της κάθε περιοχής φαίνεται να υπάρχει μείωση 10 χιλιάδες κάθε φορά, γεγονός που δικαιολογεί και την αντίστοιχη ζήτηση.

Από το διάγραμμα παρακολούθησης μέσου όρου τιμής σε συνάρτηση με τον χρόνο, αντιλαμβάνεται κανείς ότι όσο πιο αργά γίνεται μια κράτηση τόσο πιο ακριβά κοστίζει και το αντίστροφο. Δεδομένου ότι τα αρχεία ξεκινούν χρονικά από τον Σεπτέμβριο του 2023, υπάρχει μια τάση αύξησης της τιμής μέχρι την άνοιξη, με αποκορύφωμα τον χειμώνα. Τότε ο μέσος όρος της τιμής στο Αιγαίο τριπλασιάζεται

⁵ [Airbnb and the future of housing in Greece - ETERON](#)

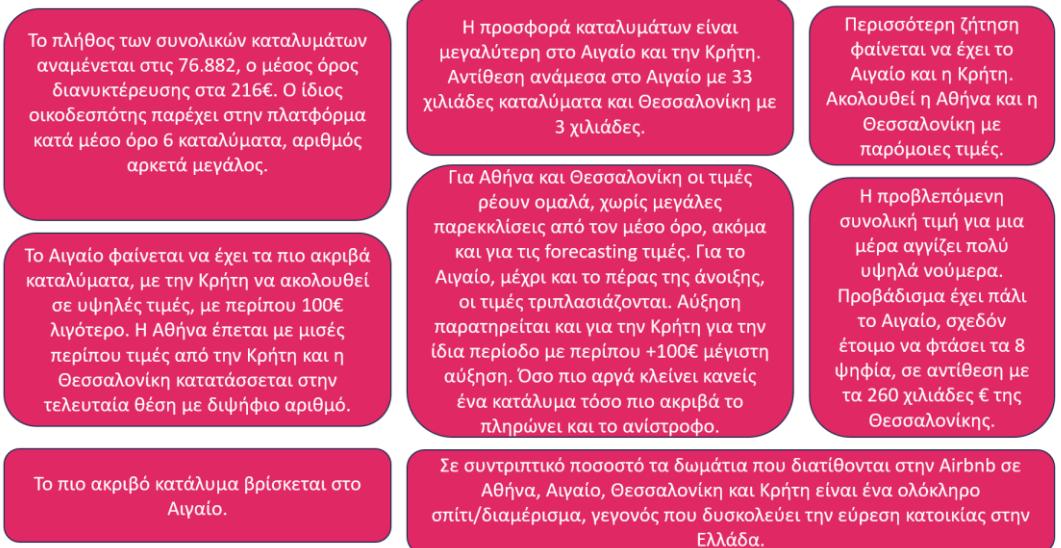
και στην Κρήτη διπλασιάζεται, ενώ στην Αθήνα και τη Θεσσαλονίκη οι τιμές ρέουν ομαλά, με μικρές παρεκκλίσεις. Το γεγονός αυτό δείχνει για άλλη μια φορά τη δημοφιλία των προορισμών. Από την άλλη, για το καλοκαίρι, που θεωρείται εποχή ακμαίου τουρισμού, δεν παρατηρούνται αυξημένες τιμές. Αυτό κατά την άποψη μου συμβαίνει για δύο λόγους: αφενός επειδή χρονικά το φθινόπωρο με το καλοκαίρι απέχουν αρκετά, αφετέρου επειδή πολλοί οικοδεσπότες έχουν επιλέξει συνειδητά να μην έχουν διαθέσιμα τα καταλύματά τους από το φθινόπωρο για το καλοκαίρι. Αυτό συμβαίνει επειδή μπορεί να θέλουν να δουν πώς κυμαίνονται οι τιμές και η ζήτηση ή ακόμα και να εκμεταλλεύονται τις διακοπές “τελευταίας στιγμής”, με αποτέλεσμα να θέτουν αργοπωρημένα τις ιδιοκτησίες τους ανοιχτές προς ενοικίαση με αυξημένες τιμές. Το forecasting δεν μπορεί να συνυπολογίσει αυτήν την οπτική, ωστόσο προβλέπει αύξηση για το επόμενο φθινόπωρο και χειμώνα στο Αιγαίο με μέσο όρο τιμής ανά διανυκτέρευσης στα 442€ δηλαδή 150€ παραπάνω. Η Αθήνα και η Κρήτη αυξάνονται μόνο κατά 20-30€, ενώ η Θεσσαλονίκη παραμένει σταθερή.

Ενδιαφέρον αποτελεί πως και οι τέσσερις περιοχές στην συντριπτική τους πλειοψηφία, με ποσοστό περίπου 89,5%, διαθέτουν προς ενοικίαση ολόκληρα διαμερίσματα ή σπίτια. Δεύτερο σε ποσοστό έρχονται τα ιδιωτικά δωμάτια και σε πολύ μικρό ποσοστό τα ξενοδοχεία. Αυτό σημαίνει ότι η Airbnb απευθύνεται σε ιδιώτες, σε ανθρώπους που έχουν κάποια περιουσία και θέλουν να κερδίσουν λεφτά από αυτήν. Από την άλλη, οι ξενοδόχοι δεν φαίνεται να προτιμούν την πλατφόρμα αυτή. Τα ποσοστά αυτά μαρτυρούν επίσης και το πρόβλημα εύρεσης σταθερής κατοικίας στη χώρα μας. Όταν οι κάτοχοι ακινήτων, σε μια χώρα που φημίζεται για τον τουρισμό της, προτιμούν να διαθέτουν την περιουσία τους σε προσωρινούς ενοικιαστές, κι όταν μάλιστα η τιμή φτάνει τα 216€ για μια μέρα (αντί για έναν μήνα), δεν είναι άξιο απορίας οι άνθρωποι αυτοί να στρέφονται στην Airbnb και αντίστοιχες πλατφόρμες.

Η διαθεσιμότητα σε βάθος χρόνου πάλι ακολουθεί την παραπάνω κατάταξη, με το Αιγαίο να είναι κλεισμένο τις 203 μέρες του χρόνου, ενώ η Θεσσαλονίκη και η Αθήνα 140 μέρες. Βέβαια, σύμφωνα με το data dictionary, δεν σημαίνει απαραίτητα ότι ένα κατάλυμα δεν είναι διαθέσιμο λόγω κράτησης αλλά ο οικοδεσπότης μπορεί να το διατηρεί κλειστό για προσωπικούς λόγους που αναλύθηκαν παραπάνω. Και οι δύο λόγοι ωστόσο, αποδεικνύουν τη ζήτηση στις περιοχές αυτές.

Η πίτα με το αναμενόμενο συνολικό ποσό ανά μέρα για τις τέσσερις περιοχές ναι μεν αποτελεί ουτοπικό σενάριο αλλά χάρη σε αυτή μπορεί κανείς να αντιληφθεί τα υπέρογκα ποσά που μελετούνται. Για άλλη μια φορά το Αιγαίο, κοντεύοντας να φτάσει τα 10 εκατομμύρια, κερδίζει την πρώτη θέση. Ακολουθεί η Κρήτη με τα μισά και η Αθήνα με 1,5 εκατομμύριο. Η Θεσσαλονίκη παραμένει τελευταία με μόνο 260 χιλιάδες ευρώ.

Strategic Government Dashboard



Εικ. 69: Συνοπτικές απαντήσεις για το strategic dashboard

6.3 Λειτουργικό dashboard για real estate επενδυτές

Το λειτουργικό dashboard, απευθυνόμενο σε επενδυτές, μελετάει σε μεγαλύτερο βαθμό μετρικές και παραμέτρους διαφόρων ειδών. Πέρα λοιπόν από τα KPIs που έχουν αναλυθεί και εξηγηθεί παραπάνω, προστέθηκαν δύο επιπλέον, που αφορούν την ποιότητα και τη ζήτηση. Το πρώτο μελετάται με τον μέσο όρο βαθμολογίας, που είναι 4,785 και το δεύτερο με τον μέσο όρο διαθεσιμότητας σε βάθος μήνα, με τις μισές περίπου μέρες να είναι κλεισμένες. Επιπλέον, με βάση τον χάρτη με το φίλτρο βαθμολογιών, παρατηρείται ότι καθώς πλησιάζει κανείς το μέγιστο όριο βαθμολογίας των 5, δεν υπάρχει σημαντική μείωση στον αριθμό των καταλυμάτων. Όλα αυτά τα νούμερα είναι πράγματι υψηλά, υποδηλώνοντας ταυτόχρονα και τον πιθανό ανταγωνισμό στον κλάδο. Το ακριβότερο κατάλυμα βρίσκεται στη Μύκονο.

Ο ανταγωνισμός αυτός δεν φαίνεται να υπάρχει ανάμεσα στους superhosts και στους μη superhosts, αφού οι μετρικές που συγκρίνονται δεν έχουν σημαντική διαφορά. Οι superhosts είναι χρήστες που έχουν υψηλή βαθμολογία, συχνή επισκεψιμότητα, χαμηλό ποσοστό ακύρωσης κλπ. Συγκεκριμένα, συγκρίνω τον μέσο όρο διανυκτέρευσης, τον μέσο όρο βαθμολογίας και τον μέσο όρο διαθεσιμότητας σε βάθος μήνα. Οι superhosts παρατηρείται να έχουν ένα μικρό προβάδισμα. Αυτή είναι μια σημαντική πληροφορία, εάν οι επενδυτές επιθυμούν να προσεγγίσουν κάποιον οικοδεσπότη για να συνεργαστούν, κι όχι να εξαγοράσουν. Σε αυτήν την περίπτωση θα πρέπει να γνωρίζουν ότι οποιοσδήποτε οικοδεσπότης που έχει καλές μετρικές, αποτελεί μάλλον καλή συνεργασία.

Θέλοντας να μελετήσω αναλυτικότερα τις κατηγορίες κριτικών, δηλαδή βαθμολογία ως προς την καθαριότητα, την επικοινωνία και την τοποθεσία, ο αλγόριθμος ομαδοποίησης του Tableau, k-means clustering, συντελεί στην αναγνώριση διακριτών ομάδων με βάση κοινά χαρακτηριστικά. Με τη βοήθεια των

clusters, δημιουργούνται τρεις ομάδες. Η μπλε κατέχει την υψηλότερη βαθμολογία και αντίστοιχη υψηλή τιμή διανυκτέρευσης, η πορτοκαλί με μεσαία βαθμολογία και τιμή και η κόκκινη με τη χαμηλότερη βαθμολογία και τιμή. Παρατηρείται πώς αυτές οι ομάδες συμπεριφέρονται διαφορετικά υπό τις διαφορετικές κατηγορίες κριτικών. Ενδιαφέρον είναι το scatter plot για τη βαθμολογία ως προς την καθαριότητα, όπου οι ομάδες είναι αρκετά διακριτές σε σχέση με τα υπόλοιπα γραφήματα, γεγονός που σημαίνει ότι υπάρχουν κάποια κοινά χαρακτηριστικά ή παράγοντες που επηρεάζουν την καθαριότητα και που έχουν μεγαλύτερη επίδραση σε αυτό το συγκεκριμένο cluster σε σύγκριση με τα άλλα. Με άλλα λόγια, ο αλγόριθμος ομαδοποίησης αναγνωρίζει ένα πιο συγκεκριμένο πρότυπο ή συμπεριφορά στα δεδομένα που αντιστοιχούν στο μπλε cluster. Μάλλον αυτό αποδεικνύει τη σημαντικότητα και αυστηρότητα της βαθμολογίας ως προς την καθαριότητα.

Στο treemap, μπορεί να δει κανείς τις εκτιμώμενες μηνιαίες συνολικές τιμές για κάθε περιοχή, με το Αιγαίο να ξεπερνάει κατά το διπλάσιο την Κρήτη. Οι τιμές είναι ενδεικτικές, διότι δεν είναι δυνατό όλα τα καταλύματα να είναι κλεισμένα όλους τους μήνες, όλες τις μέρες, για όλες τις περιοχές. Όμως όπως φαίνεται και από τη ζήτηση των περιοχών, υφίσταται αυτή η αναλογία ποσών και στην πραγματικότητα.

Χρήσιμο είναι να γνωρίζει κάποιος επενδυτής τόσο τους τύπους κατοικιών που υπάρχουν όσο και τη ζήτησή τους. Έτσι, μετά από ένα grouping που εφάρμοσα στο Tableau, είδα πως υπάρχουν έξι διαφορετικές κατηγορίες: σπίτια στην εξοχή, ιδιωτικά σπίτια, ιδιωτικά δωμάτια, κοινόχρηστα δωμάτια, σπίτια-μεταφορικά μέσα και σπίτια πολυτελείας, εμπνευσμένη από το όνομα OMG!, το οποίο η ίδια εταιρία έχει καθιερώσει για αυτόν τον τύπο κατοικίας. Πιο συγκεκριμένα, αυτό το group περιλαμβάνει σπίτια σε σπηλιά, πύργο, Α-σπίτια κλπ. Τα μέσα μεταφοράς πρόκειται για καράβια, τροχόσπιτα κλπ, τα οποία αποτελούν και την πιο ακριβή επιλογή με γενικό μέσο όρο διανυκτέρευσης 421€. Αυτό είναι δικαιολογημένο, εάν σκεφτεί κανείς πως δεν είναι απλά ένα μέρος διαμονής αλλά κι ένας τρόπος μετακίνησης. Οι πληροφορίες αυτές επαληθεύουν τον αυξημένο μέσο όρο διανυκτέρευσης σε Αιγαίο και Κρήτη, αφού αυτός ο τύπος κατοικίας συναντάται συνήθως σε νησιά, κι όχι στις πόλεις. Στη συνέχεια ακολουθούν τα σπίτια πολυτελείας με μέσο όρο 255€, αριθμός ανάλογος της αρχιτεκτονικής αυτών των καταλυμάτων. Τα ιδιωτικά σπίτια και δωμάτια κυμαίνονται στις ίδιες τιμές, περίπου στα 210€, ενώ μια πιο οικονομική λύση αποτελεί το σπίτι στην εξοχή, όπως σκηνή, φάρμα, ξυλόσπιτο κλπ. Η πιο φθηνή επιλογή είναι σίγουρα τα shared rooms με μέσο όρο μόνο 36€ την ημέρα. Από την άλλη, το πλήθος των κρεβατιών φανερώνει και το πλήθος των ατόμων που χωράνε σε ένα κατάλυμα, με την προσαύξηση του ενούς ατόμου κάθε φορά, αφού μάλλον σε κάθε σπίτι υπάρχει κατά μέσο όρο ένα διπλό κρεβάτι. Αυτός ο αριθμός για κάθε κατηγορία, ειδικότερα για τα ιδιωτικά καταλύματα, και κάθε περιοχή δεν ξεπερνάει τα πέντε, με αποτέλεσμα τα καταλύματα της Airbnb να χαρακτηρίζονται ως family-friendly. Τα OMG! listings δέχονται κατά μέσο όρο 2 άτομα, άρα μάλλον απευθύνονται σε ζευγάρια/δυάδες, όπως και τα σπίτια στην εξοχή και τα ιδιωτικά δωμάτια που πιθανόν προτιμούνται από solo travelers. Για τα shared rooms υπάρχει μια αύξηση με την Κρήτη να φτάνει τα 13 κρεβάτια, αλλά αυτό είναι λογικό διότι πρόκειται συνήθως για hostels. Ο μέσος όρος κρεβατιών σε hostels για Αιγαίο, Θεσσαλονίκη και Αθήνα είναι τρία.

Ενδιαφέρον αποτελεί το ποιες τιμές κατά μέσο όρο ορίζουν στο παρόν περισσότερο και λιγότερο έμπειροι οικοδεσπότες. Οι άνθρωποι που διέθεσαν τις περιουσίες τους από τη γέννηση της εταιρίας, φαίνεται να έχουν επαναπαυτεί, διατηρώντας τις χαμηλές τιμές που ενδεχομένως ήταν για δοκιμή στη νέα αυτή ιδέα. Οι άνθρωποι εγγεγραμένοι αργότερα το 2012 με 2015, έχουν τον υψηλότερο μέσο όρο διανυκτέρευσης με περίπου 300€. Οι χρήστες του 2012 έχουν μάλιστα τον υψηλότερο μέσο όρο διανυκτέρευσης, με 385€. Οι τιμές για άτομα εγγεγραμμένα από το 2016 έως σήμερα μειώνονται και σταθεροποιούνται στα 200€. Μια μικρή αύξηση του μέσου όρου τιμής της κλάσης των 60€ παρατηρείται για τους χρήστες του 2022, οι οποίοι εκμεταλλεύτηκαν μάλλον το τέλος της πανδημίας και την επανέναρξη των ταξιδιών. Οι νέοι χρήστες ορίζουν για τα ακίνητά τους την τιμή των 175€ κατά μέσο όρο, ποσό που -όπως είναι λογικό- είναι χαμηλότερο του συνολικού μέσου όρου, λόγω μάλλον της απειρίας τους. Ο μέσος όρος του πλήθους των καταλυμάτων που διατίθεται από έναν χρήστη είναι γενικά ίσα κατανεμημένος στους οικοδεσπότες όλων των χρονολογιών, με τις χρονολογίες από 2008 έως το 2012 να είναι ελαφρώς μεγαλύτερος. Μεγάλη παρέκκλιση σημειώνουν οι οικοδεσπότες του 2012 που φτάνουν περίπου τα 20 καταλύματα ανά άτομο, σε αντίθεση με τους φετινούς χρήστες που διαθέτουν 2 καταλύματα. Οι άνθρωποι που έγιναν μέλη το 2012 σημειώνουν τον μεγαλύτερο μέσο όρο, κι αυτό μάλλον επειδή βρίσκονται στη μέση χρονολογικά: 5 χρόνια αφότου η ιδέα δοκιμάστηκε και πέτυχε και 10 χρόνια επιπλέον που μπόρεσαν και πρόσθεσαν κι άλλα ακίνητα. Αυτή η κατηγορία μάλιστα κατέχει και τον υψηλότερο μέσο όρο διανυκτέρευσης, όπως αναφέρθηκε παραπάνω. Γενικότερα, δεν υπάρχει κάποιο μοτίβο που να συσχετίζει τα χρόνια χρήσης της πλατφόρμας με το πλήθος των καταλυμάτων για ένα άτομο συγκεκριμένων χρονολογιών. Ούτως ή άλλως, η περιουσία του καθενός είναι ένας τυχαίος παράγοντας, που όμως - όπως φαίνεται - στο ζενίθ της οικονομικής κρίσης έως και σήμερα, μειώθηκε ή χρησιμοποιήθηκε για ιδιοκατοίκηση λόγω οικονομικών δυσκολιών.

Όπως ακριβώς διαπιστώθηκε κι από το στρατηγικό dashboard, οι τιμές φαίνεται να αυξάνονται καθώς πλησιάζει η άνοιξη, με την κορύφωση να σημειώνεται τον χειμώνα. Σε αυτήν την περίοδο, οι τιμές στα νησιά του Αιγαίου τριπλασιάζονται και στην Κρήτη διπλασιάζονται, ενώ στην Αθήνα και τη Θεσσαλονίκη παραμένουν σταθερές, με μικρές παραλλαγές. Αυτή η τάση επαναβεβαιώνει τη δημοφιλία των διαφόρων προορισμών. Κατά τους καλοκαιρινούς μήνες, παρατηρείται σταθερότητα στις τιμές, καθώς οι οικοδεσπότες πιθανότατα διαχειρίζονται επιλεκτικά τη διαθεσιμότητα των καταλυμάτων τους, προσδιορίζοντας τις τιμές με βάση τη ζήτηση. Επιπλέον, οικοδεσπότες ενδεχομένως επιλέγουν να μην προσφέρουν τα καταλύματά τους για κρατήσεις κατά το καλοκαίρι πολύ νωρίτερα χρονικά, καθιστώντας τα διαθέσιμα μόνο λίγους μήνες πριν, με σκοπό την αύξηση των τιμών. Σε γενικές γραμμές, η πρώιμη κράτηση καταλύματος αποτελεί μια οικονομική επιλογή και το αντίθετο.

Η γνώμη του κοινού μετράει και μάλιστα αποτελεί σήμερα τη σημαντικότερη κριτική για να πάρει κανείς αποφάσεις. Είναι πλέον σύνηθες πριν αποφασίσει κανείς να κάνει χρήση μιας υπηρεσίας να ελέγχει πρώτα τις κριτικές από καθημερινούς ανθρώπους. Η κριτική όμως βοηθάει και τους οικοδεσπότες, αφού μόνο με συστηματικό feedback μπορούν να βελτιωθούν. Με το τελευταίο

διάγραμμα θέλω να μελετήσω εάν οι επισκέπτες έχουν αντιληφθεί τη σημαντικότητα της κριτικής κι εάν εκμεταλλεύονται τη δύναμη αυτή που τους δίνεται. Πράγματι επιβεβαιώνεται η θεωρία αυτή, αφού από το 2008 έως σήμερα ο αριθμός των καταλυμάτων που βαθμολογούνται όλο και αυξάνεται. Βέβαια αυτό είναι ανάλογο και του πλήθους των καταλυμάτων που βρίσκονται στην Airbnb κάθε χρονιά. Η κλίση είναι γενικά θετική και για το διάστημα 2014-2017 παρατηρείται εκθετική κατανομή. Εξαίρεση αποτελεί το διάστημα 2019-2020, που όπως είναι λογικό οι επισκέπτες κι άρα οι κριτικές μειώθηκαν αισθητά λόγω της πανδημίας. Τέτοιες καταστάσεις είναι λογικό να επηρεάζουν σε μεγάλο βαθμό διάφορους παράγοντες, αλλά αποτελούν ανακρίβειες όταν πρόκειται για μελέτη τάσεων και συμπεριφορών. Τελικά το 62,3% των καταγεγραμένων καταλυμάτων βαθμολογήθηκε την τρέχουσα χρονιά, αριθμός ικανοποιητικός, σε αντίθεση με τα τέσσερα μόνο καταλύματα που βαθμολογήθηκαν το 2010.



Εικ. 70: Συνοπτικές απαντήσεις για το operational dashboard

Επίλογος

Η παρούσα πτυχιακή εργασία αναδεικνύει την ουσιώδη σημασία της ανάλυσης και οπτικοποίησης δεδομένων στον σύγχρονο ψηφιακό κόσμο. Αρχικά, η δημιουργία ενός Data Warehouse με fact και dimension πίνακες, βασισμένοι σε δεδομένα από την Airbnb αποτελεί στοιχείο πρωταρχικής σημασίας στον αγώνα για την κατανόηση των διαδικασιών και των τάσεων που διαμορφώνουν την αγορά. Στη συνέχεια, μέσω του Tableau, διαφορετικά είδη dashboards που υλοποιούνται, όπως το strategic για υπουργεία και το operational για επενδυτές, προσφέρουν ένα πλαίσιο για τη λήψη ευφυών αποφάσεων. Μέσω αυτών των εργαλείων, οι επαγγελματίες μπορούν να παρακολουθούν την εξέλιξη της αγοράς και να εντοπίζουν αποτελεσματικές επενδύσεις και στρατηγικές ανάπτυξης. Επιπλέον, η διαδικασία ETL εξασφαλίζει τη μελλοντική ενημέρωση και διατήρηση της ακεραιότητας και ποιότητας των δεδομένων από την Airbnb στο εν λόγω σύστημα.

Η ανάλυση αποκαλύπτει έναν εντυπωσιακά μεγάλο αριθμό καταλυμάτων στην πλατφόρμα της Airbnb σε Αθήνα, Αιγαίο, Κρήτη και Θεσσαλονίκη. Με τον μέσο όρο τιμής ανά διανυκτέρευση να φτάνει τα 216€ και την κατανομή των τιμών να αντικατοπτρίζει τη δημοφιλία των προορισμών, παρουσιάζεται ένας δυναμικός τομέας με πολλές προκλήσεις. Η ακριβότερη και δημοφιλέστερη επιλογή φαίνεται να είναι τα νησιά του Αιγαίου, με την Κρήτη να ακολουθεί. Η Θεσσαλονίκη παίρνει την τελευταία θέση. Ταυτόχρονα, οι ιδιοκτήτες προτιμούν να εκμεταλλεύονται ολόκληρα διαμερίσματα και σπίτια προς ενοικίαση, ενώ οι προβλέψεις δείχνουν αυξημένες τιμές και έσοδα για το μέλλον.

Σε βαθύτερη ανάλυση, το operational dashboard για επενδυτές παρουσιάζει επιπλέον μετρήσεις που αφορούν την ποιότητα και τη ζήτηση των καταλυμάτων. Ο συνολικός αριθμός καταλυμάτων που αντιστοιχούν σε υψηλές βαθμολογίες είναι εντυπωσιακά αυξημένος, αφού ο μέσος όρος βαθμολογίας των συνολικών καταλυμάτων είναι 4,785, αναδεικνύοντας τον ανταγωνισμό μεταξύ των οικοδεσποτών. Οι superhosts φαίνεται να μη διαφέρουν σημαντικά από τους μη superhosts σε σχέση με βασικές μετρικές ποιότητας.

Ο αλγόριθμος ομαδοποίησης αναγνωρίζει τρεις ομάδες καταλυμάτων με βάση διαφορετικές κατηγορίες αξιολόγησης. Το treemap αποτυπώνει τα εκτιμώμενα μηνιαία έσοδα για κάθε περιοχή, όπου το Αιγαίο φαίνεται να ξεπερνά κατά πολύ τα υπόλοιπα. Σε εξίσου υψηλές τιμές φτάνει και ο μέσος όρος διανυκτέρευσης για το Αιγαίο σε βάθος χρόνου. Το grouping των καταλυμάτων καταλήγει σε έξι διαφορετικές κατηγορίες, με τα OMG! listings και τα μέσα μεταφοράς να είναι οι ακριβότερες επιλογές και τα shared rooms η πιο οικονομική. Το πλήθος των κρεβατιών αντιστοιχεί στον αριθμό των ατόμων και αποκαλύπτει την οικογενειακή φιλικότητα των καταλυμάτων. Σήμερα, ο μέσος επισκέπτης συνηθίζει να βαθμολογεί και να αξιολογεί την εμπειρία του για τα καταλύματα της Airbnb.

Η ανάλυση αυτή προσφέρει πολύτιμες πληροφορίες για την ανάπτυξη του τομέα των ενοικιαζομένων καταλυμάτων στην Ελλάδα, καθώς και για τις προκλήσεις που αντιμετωπίζουν οι αρχές σχετικά με τη ρύθμιση και διαχείριση αυτής της δραστηριότητας.

ΒΙΒΛΙΟΓΡΑΦΙΑ

Ξενόγλωσση:

- A. Gupta, V. Harinarayan, and D. Quass. (1995). *Aggregate query processing in data warehousing environments*. In Proc. of VLDB.
- Cotgreave A., Shaffer J. and Wexler S. (2017). *The Big Book of Dashboards*. John Wiley & Sons, Inc.
- Eckerson W.W. (2010). *Performance Dashboards: Measuring, Monitoring and Managing Your Business*. Google Books. John Wiley & Sons.
[Performance Dashboards: Measuring, Monitoring, and Managing Your Business - Wayne W. Eckerson - Βιβλία Google](#)
- Kimball R. and Ross M. (2011). *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling*. Google Books. John Wiley & Sons.
[https://books.google.gr/books?hl=el&lr=&id=XoS2oy1lcB4C&oi=fnd&pg=PA1&dq=data+warehouse&ots=1EEamEmMiy&sig=TGInEebUbm1yaRj9L9FeO_jFd8o&redir_esc=y#v=onepage&q=data%20warehouse&f=false](#)
- Knafllic C.N. (2015). *Storytelling with data*. John Wiley & Sons, Inc.
- Lechtenbörger, J. (2001). *Data Warehouse Schema Design*. Google Books. IOS Press.
[https://books.google.gr/books?hl=el&lr=&id=BBOM25evBHKC&oi=fnd&pg=P A1&dq=data+warehouse+schemas&ots=ixX_17b9pw&sig=8ZanVnW6QKuGujH3x8YaCpcWx5I&redir_esc=y#v=onepage&q=data%20warehouse%20schema s&f=false](#)
- Sharda R., Delen D., Turban E. (2007). *BUSINESS INTELLIGENCE AND ALANYTICS. Systems for Decision Support*, Pearson ed.
[business_intelligence20200316-50198-128rsmo.pdf \(d1wqxts1xzle7.cloudfront.net\)](#)
- J. Widom (1995). *Special Issue on Materialized Views and Data Warehousing*. volume 18(2). IEEE, editor. Data Engineering.
- J. Widom (1995). *Research problems in data warehousing*. In Proc. CIKM.
- Williams, S., & Williams, N. (2010). *The Profit Impact of Business Intelligence*. San Francisco: Morgan Kaufmann (Elsevier).

Άρθρα:

- Bernardino J., Ferreira T., Pedrosa I. (2017). *Business Intelligence for E-commerce: Survey and Research Directions*. Isabel & Bernardino, Jorge.
[\(PDF\) Business Intelligence for E-commerce: Survey and Research Directions \(researchgate.net\)](#)
- Djerdjouri M. (2020). *Data and Business Intelligence Systems for Competitive Advantage: prospects, challenges, and real-world applications*. Mercados y Negocios.
[Data and Business Intelligence Systems for Competitive Advantage: prospects, challenges, and real-world applications \(redalyc.org\)](#)
- Elias M. (2012). *Enhancing User Interaction with Business Intelligence Dashboards*. Ecole Centrale Paris.
[Enhancing User Interaction with Business Intelligence Dashboards - TEL - Thèses en ligne \(hal.science\)](#)
- Jameel, Karwan & Yazdeen, Abdulmajeed & Ibrahim, Abass & Abdulrazzaq, Maiwan & Mahmood, Mayyadah (2022). *Analyses the Performance of Data Warehouse Architecture Types*. *Applied Soft Computing*. 3. 45-57. 10.30880/jscdm.
[\(PDF\) Analyses the Performance of Data Warehouse Architecture Types \(researchgate.net\)](#)
- Shollo A., Kautz K. (2010). *Towards an Understanding of Business Intelligence*. ACIS 2010 Proceedings. 86.
[Towards an Understanding of Business Intelligence \(aisnet.org\)](#)
- Tauriainen E. (2023). *Database Utilization in Embedded Software Development*. Oulu University of Applied Sciences.
[Tauriainen_Emma.pdf \(theseus.fi\)](#).
- Thalheim, Bernhard. (1991). *Dependencies in Relational Databases*. 10.1007/978-3-663-12018-6.
[\(PDF\) Dependencies in Relational Databases \(researchgate.net\)](#)
- Vishesh, Siddesh & Srinath, Manu & Kumar, Akshatha & A.S, Nandan (2017). *Data Warehousing Architecture and Pre-Processing*. IJARCCE. 6. 13-18. 10.17148/IJARCCE.2017.6503.
[\(PDF\) Data Warehousing Architecture and Pre-Processing \(researchgate.net\)](#)
- Στρατής Κ.Δ. (2023). *Επεξεργασία και Ανάλυση Δεδομένων στο Power BI*. Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών

Ιστότοποι:

- [What Is a Data Warehouse | Oracle](#)

- [Best Analytics and Business Intelligence Platforms Reviews 2023 | Gartner Peer Insights](#)
- [Inside Airbnb: Get the Data](#)
- [Online Courses - Learn Anything, On Your Schedule | Udemy](#)