

Business Intelligence

Project Report

Moiz Zulfiqar(24543), Umair Ahmed(24377)

13th May, 2024

Table of Content

Introduction.....	3
Business Problem.....	3
Dataset.....	3
Turkish Market Sales Dataset With 9.000+Items.....	3
Exploratory data analysis (EDA).....	3
Data Wrangling - MVA.....	3
Uni-variate & Bi-variate Analysis.....	4
ANOVA and Tukey's HSD Tests.....	5
Normality Tests.....	6
Brainstorming Before Using Tool.....	6
Identifying Key BI Queries.....	6
Connecting Columns to BI Queries.....	6
Storytelling on Paper.....	8
Visualizations.....	9
Charts Used.....	9
Dashboards Used.....	27
Report Overview Dashboard.....	27
Sales Analysis Dashboard.....	28
Salesmen Performance Dashboard.....	29
Product Insights Dashboard.....	30
Geographic Analysis Dashboard.....	31
Customer Demographics Dashboard.....	32
Solving The Business Problem Using Story.....	33
Revisiting The Business Problem.....	33
Breaking Business Problem Into BI Queries.....	33
1. Key Insights for Swift Decision-Making.....	33
2. Sales Performance Insights.....	33
3. Assessing Sales Team Effectiveness.....	33
4. Product Sales Analysis.....	33
5. Geographical Sales Distribution.....	33
6. Customer Base Analysis.....	33
Sequence of Dashboards To Explain Story.....	34
Team Member Contributions.....	38

Introduction

Business Problem

The core business problem is to assess the company's performance at the end of march, to strategize for the upcoming month. The goal is to identify trends, anomalies, opportunities, and areas needing improvement across various facets of the business, including sales, product performance, salesforce effectiveness, geographic sales distribution, and customer demographics. By analyzing these areas, the company aims to optimize resource allocation, enhance product positioning, improve sales strategies, and better understand customer preferences.

Dataset

Turkish Market Sales Dataset With 9.000+Items

This dataset comprises supermarket data in the Turkish language. It contains over 600,000 sales entries, featuring information such as the names (all of which are artificial), genders, ages, and birthdates of more than 52,000 unique customers. The dataset includes 9,000+ items, categorized and subcategorized into four levels. It also encompasses details from 81 stores and data on over 1,200 salespeople, as well as geographic location information (latitude and longitude). The data represents three months of sales, all of which were artificially generated based on the population distribution of cities. All customer names in the dataset are also artificial.

Link To Dataset:

<https://www.kaggle.com/datasets/omercolakoglu/turkish-market-sales-dataset-with-9000items>

Exploratory data analysis (EDA)

Data Wrangling - MVA

Libraries and Initial Setup: Imported libraries like Pandas, NumPy, Matplotlib, MissingNo, and Scikit-learn's SimpleImputer for handling missing values and visualization. Enabled inline display of plots with %matplotlib inline.

Data Loading: Loaded a dataset from an Excel file located at "./Dataset/MarketSales.xlsx".

Data Inspection:

- Listed all columns in the dataset.

- Checked data types.
- Calculated the sum of missing values for each column.

Visualizing Missing Data:

- Used mano.bar to visualize the completeness of the dataset.
- Used mano.matrix to display the location of missing data points and identify instances of MNAR.
- Identified three examples of Missing Not At Random (MNAR) patterns related to ITEMCODE, ITEMNAME, CLIENTCODE, CLIENTNAME, GENDER, BRAND, and BRANDCODE.
- Plotted a heatmap to show correlations between the presence of missing values across different variables.

Imputation:

- A mapping strategy to impute IITEMNAME and ITEMCODE based on FICHENO
- Implemented mode imputation to fill missing values for CATEGORY_NAME2 using the most frequent value within the column.
- Developed a mapping strategy to impute missing CATEGORY_NAME3 values based on existing CATEGORY_NAME1 values, and used mean imputation where direct mapping was not possible.

Data Cleaning:

- Translated gender values from Turkish ('E' for Male and 'K' for Female) to English.
- Removed rows with missing REGION values to clean the dataset further.

Data Export:

Saved the cleaned and updated dataset to an Excel file named "updated_marketsales.xlsx".

Uni-variate & Bi-variate Analysis

Data Description: Initial exploration using df.describe() to summarize the statistical characteristics of numeric columns, followed by examining the data types with df.dtypes.

Univariate Analysis:

- Descriptive statistics for LINENETTOTAL, AMOUNT, and PRICE columns.
- Calculation of central tendencies, specifically mean and median, to identify distributions and impact of outliers.
- Formatted output for readability, including converting means to integer formats and presenting them in a standardized string format.

Visualizations:

- Histograms and Q-Q plots for LINENET and other financial columns to assess distribution shapes.
- Analysis includes filtering the dataset to remove outliers for clearer distribution visualization.

Correlation Analysis

- Strong Positive Correlation between 'ID' and 'FICHENO' (0.999986), suggesting almost identical or dependent values.
- 'AMOUNT' exhibits strong positive correlations with 'LINENETTOTAL' (0.418553) and 'LINENET' (0.437718), indicating that higher amounts correspond to increased net totals and net values.
- 'PRICE' similarly shows strong positive correlations with 'LINENETTOTAL' (0.489237) and 'LINENET' (0.496955).
- 'LINENETTOTAL' and 'LINENET' have a very strong positive correlation (0.997764), highlighting a likely dependency.
- Geographical Correlations: 'BRANCHNR' shows a positive correlation with 'LATITUDE' (0.132888), while 'LONGITUDE' has a negative correlation with 'BRANCHNR' (-0.135808).
- 'LATITUDE' and 'LONGITUDE' show a weak positive correlation (0.016577), indicating minimal linear relationship.

ANOVA and Tukey's HSD Tests

ANOVA Tests:

- ANOVA for 'LINENET' with respect to 'GENDER'.
- ANOVA for 'LINENET' with respect to 'REGION'.
- ANOVA for 'LINENET' with respect to 'BRAND'.

Tukey's HSD Tests:

- Tukey's test to compare means of 'LINENET' across different 'REGION' categories.
- Tukey's test for 'LINENET' across different 'BRAND' categories.

These statistical tests are designed to determine whether there are significant differences in the 'LINENET' values across various groups defined by 'GENDER', 'REGION', and 'BRAND'. Results from these tests can help identify if any particular group influences the 'LINENET' metric significantly, assisting in decision-making processes or further in-depth analysis.

Normality Tests

Normality Tests:

- Shapiro-Wilk, Anderson-Darling, and D'Agostino's K^2 tests are conducted on filtered data (outliers removed) to determine if data follows a normal distribution.
- Results from these tests indicate significant departures from normality, suggesting non-normal distributions for the financial variables analyzed.

Summary of Normality Tests:

The data is not normally distributed based on the statistical tests conducted. Results and statistics such as the Anderson statistic value are reported in detail, along with critical values for various significance levels.

Brainstorming Before Using Tool

Identifying Key BI Queries

1. Analyzing Overall Sales Performance: Understanding sales trends, daily sales data, and brand performance.
2. Evaluating Sales Team Effectiveness: Assessing the contributions and performance metrics of individual salespeople and the team as a whole.
3. Assessing Product Sales: Focusing on top-selling and most profitable products, and identifying trends.
4. Examining Geographic Sales Distribution: Analyzing how sales distribute across various regions and cities.
5. Diving into Customer Demographics: Understanding customer profiles based on sales data, such as gender distribution and regional preferences.

Connecting Columns to BI Queries

1. Analyzing Overall Sales Performance
 - DATE_: Essential for tracking daily sales and identifying trends over time.
 - AMOUNT: Quantifies the volume of products sold.
 - PRICE: Helps in analyzing the pricing strategies.
 - LINENETTOTAL, LINENET: Reflects total sales revenue which can be crucial for financial analysis.
 - BRAND, BRANDCODE: Useful for evaluating performance by brand.
 - CATEGORY_NAME1, CATEGORY_NAME2, CATEGORY_NAME3: Helps in depth analysis by product categories.
 - ITEMCODE, ITEMNAME: Identifies which specific items are selling.

2. Evaluating Sales Team Effectiveness

- SALESMAN: Directly assesses which salesperson made the sale.
- LINENETTOTAL: Reflects the total value of sales attributed to each salesperson.
- AMOUNT: Shows the quantity sold by each salesperson.

3. Assessing Product Sales

- ITEMCODE, ITEMNAME: Identifies products to assess top-selling and most profitable items.
- AMOUNT, PRICE, LINENET, LINENETTOTAL: Important for profitability analysis.
- CATEGORY_NAME1, CATEGORY_NAME2, CATEGORY_NAME3: For detailed product category trends and performance.

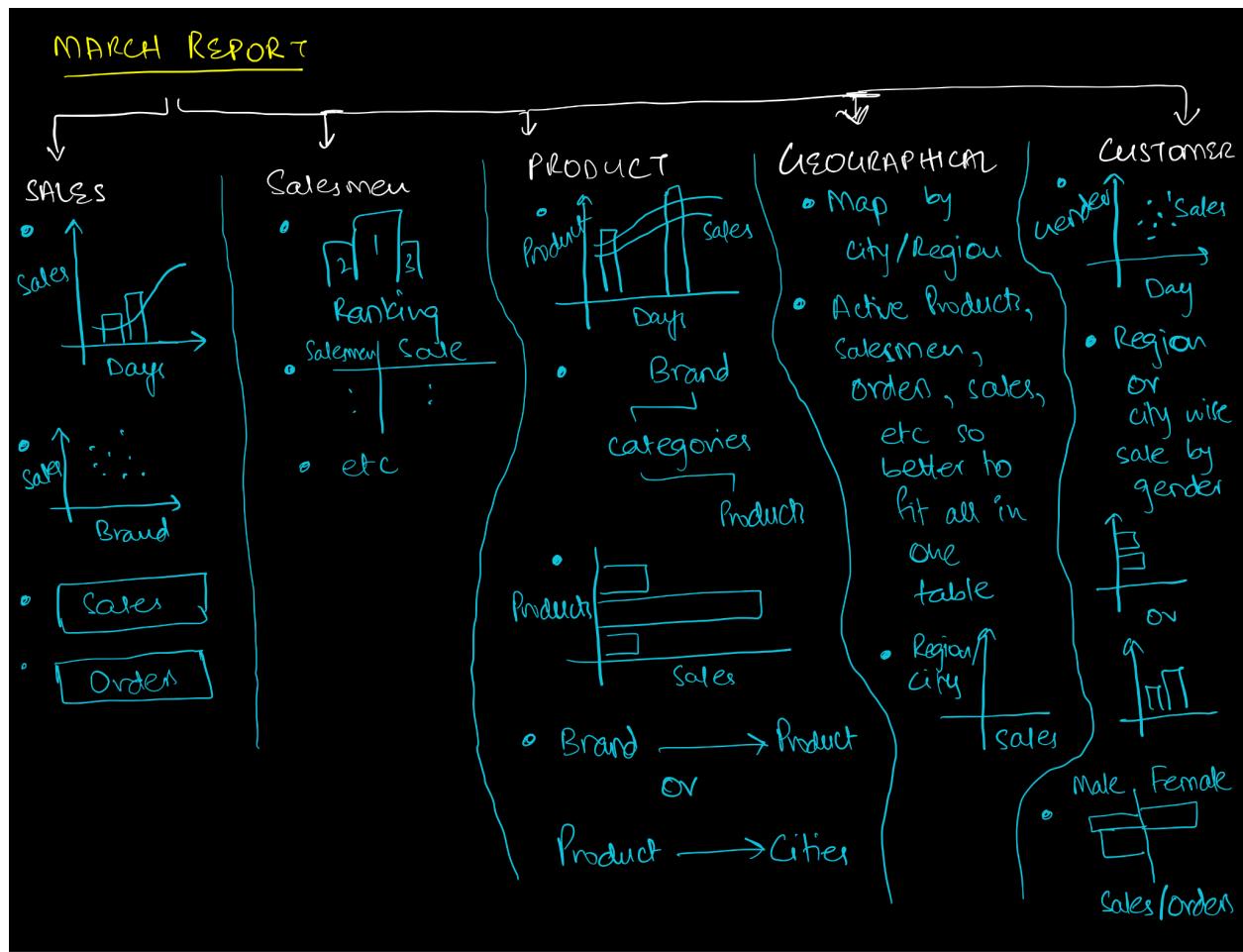
4. Examining Geographic Sales Distribution

- CITY, REGION: Key for regional sales distribution analysis.
- LATITUDE, LONGITUDE: Can be used for more detailed geo-mapping of sales data.
- BRANCHNR, BRANCH: Indicates the location of the sale, which is crucial for regional performance.

5. Diving into Customer Demographics

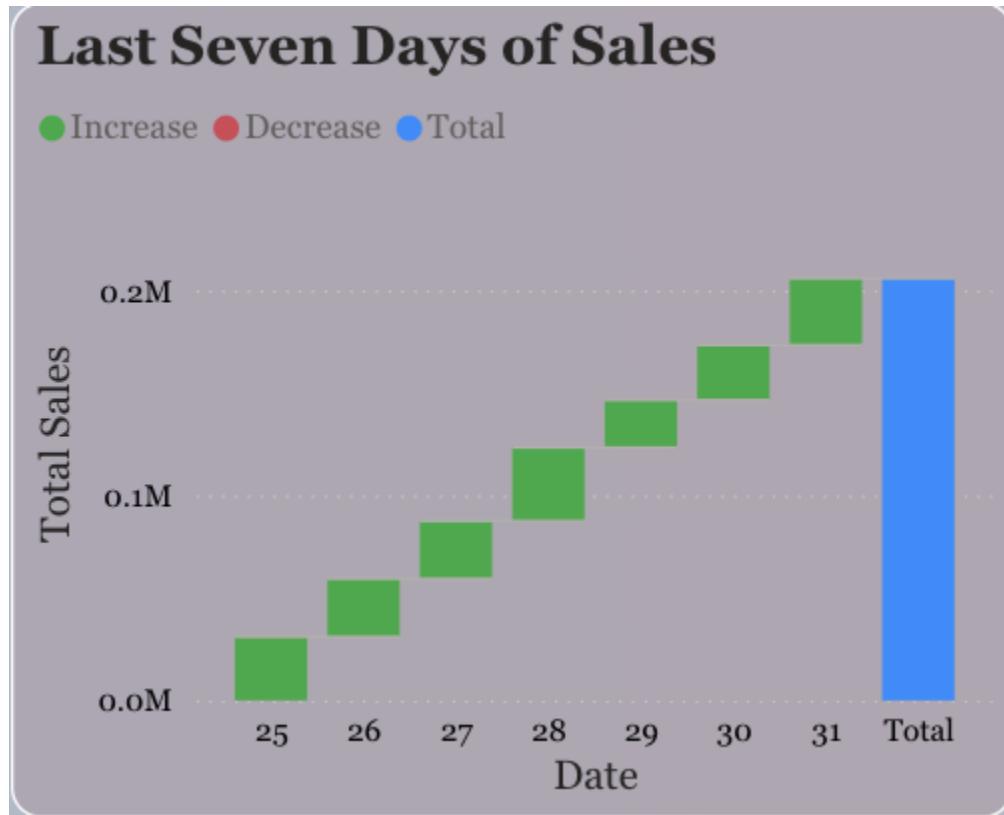
- GENDER: Essential for analyzing sales by gender distribution.
- CLIENTCODE, CLIENTNAME: Helps identify and analyze customer buying patterns.
- CITY, REGION: Offers insights into regional preferences and demographics.

Storytelling on Paper



Visualizations

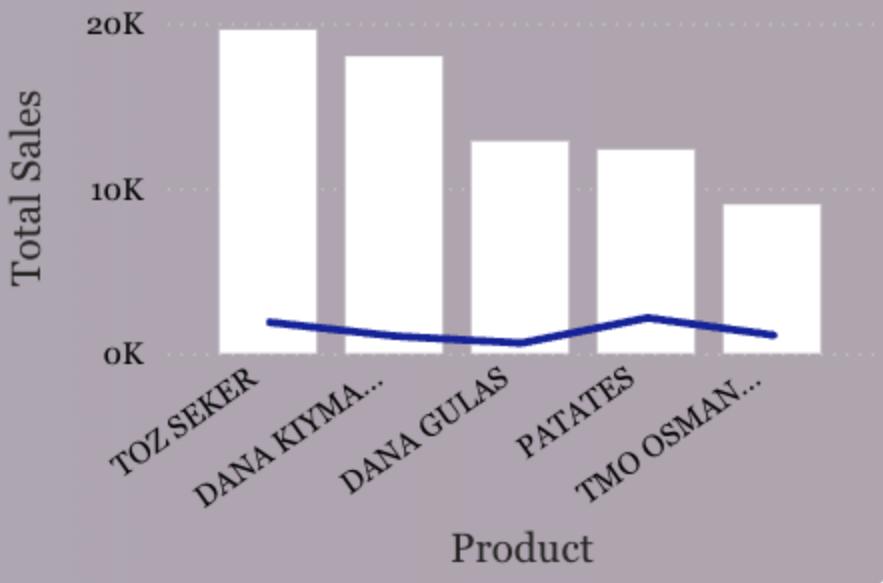
Charts Used



This chart shows the total sales for the last seven days of March, with colors indicating whether sales have increased (green), decreased (red), or the total aggregate (blue bar). It helps quickly assess sales momentum at month-end.

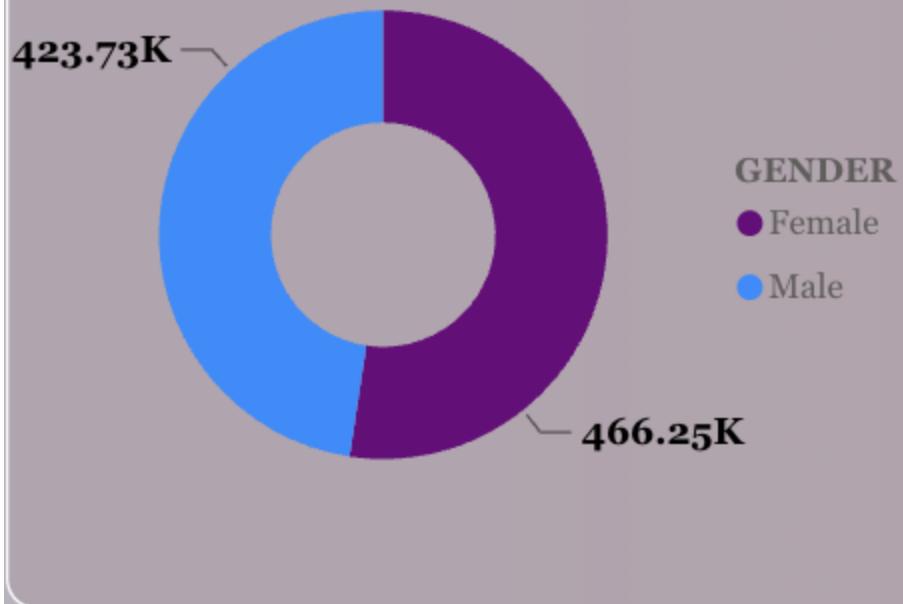
Top 5 Selling Products Sale & Order Count

● Total Sales ● Total Orders



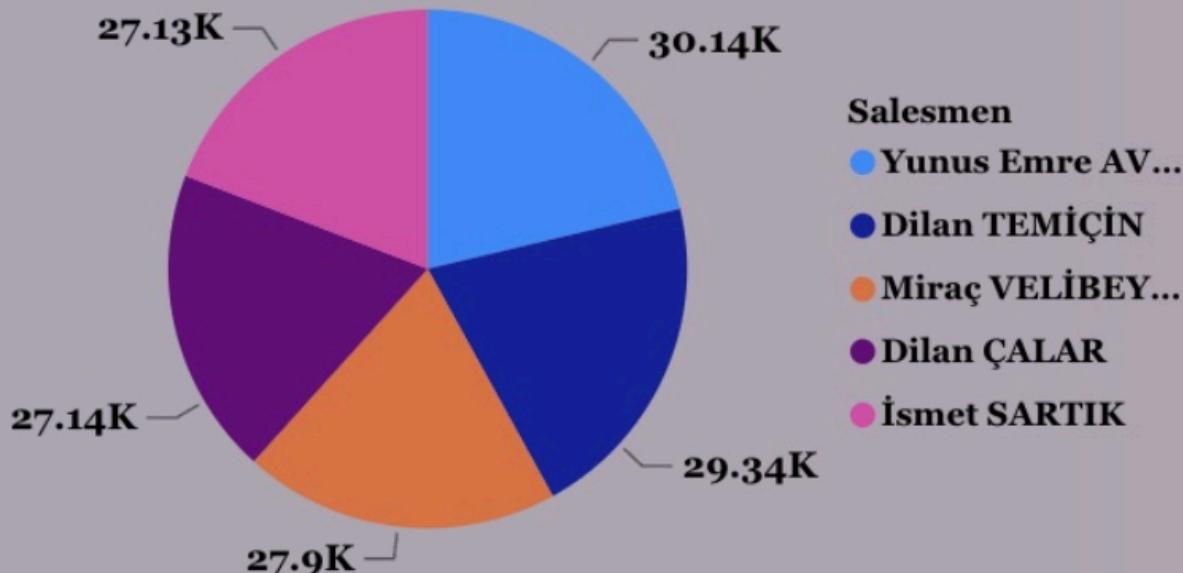
Displays total sales and total order counts for the top 5 selling products, highlighting which products are not only popular in terms of quantity sold but also bring in the most revenue.

Total Sales by GENDER



This chart breaks down the total sales by gender, showing the distribution of sales between male and female customers.

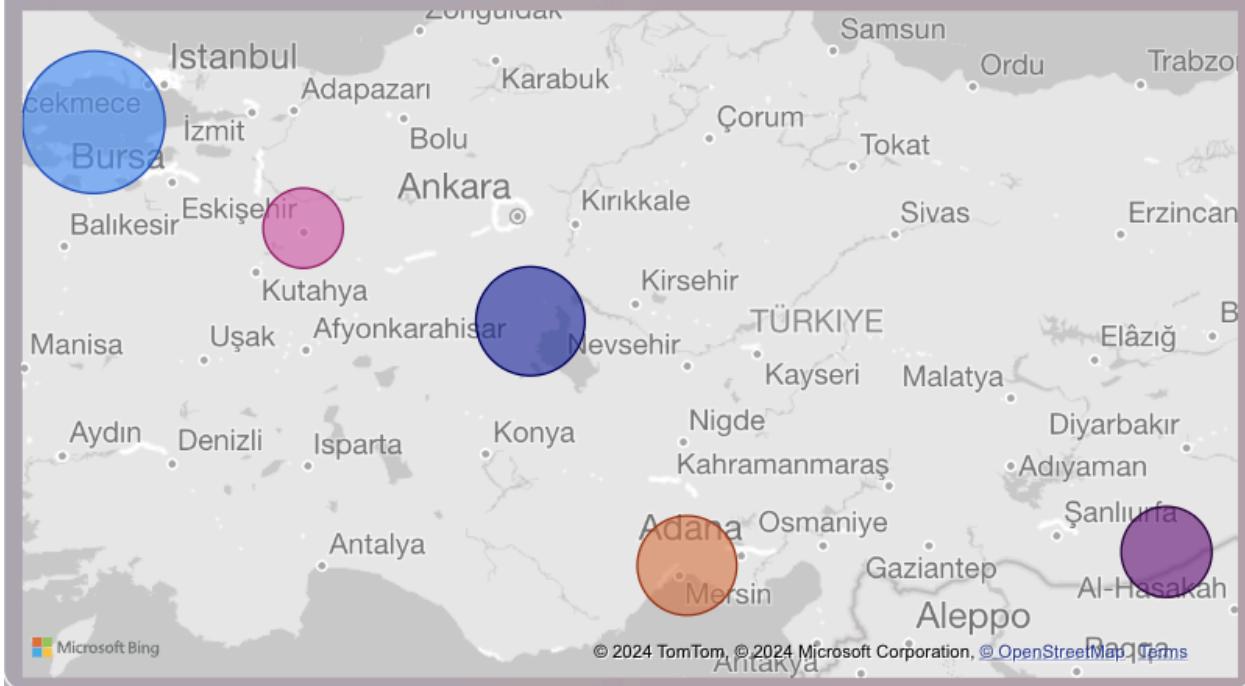
Top 5 Salesmen by Total Sales



Represents sales distribution among the top five salesmen, showing which salesman is leading in sales and how closely others are performing.

Total Sales by Region

REGION ● Marmara ● İç Anadolu ● Akdeniz ● Güneydoğu Anadolu ● Doğu Anadolu



A geographic distribution of sales, where each bubble's size on the map corresponds to the total sales in that region, with the map providing a quick visual reference for regional sales performance.

Total Increase in Sales From Previous Month

919.32K

Displays the total increase in sales from the previous month, quantified at 919.32K. This metric immediately indicates the growth in revenue, helping stakeholders gauge overall sales performance month-over-month.

Total Increase in Orders From Previous Month

226K

Displays the total increase in orders count from the previous month, quantified at 226K. This metric immediately indicates the growth in orders, reflecting on customer demand or marketing effectiveness.

Total Sales by Brand

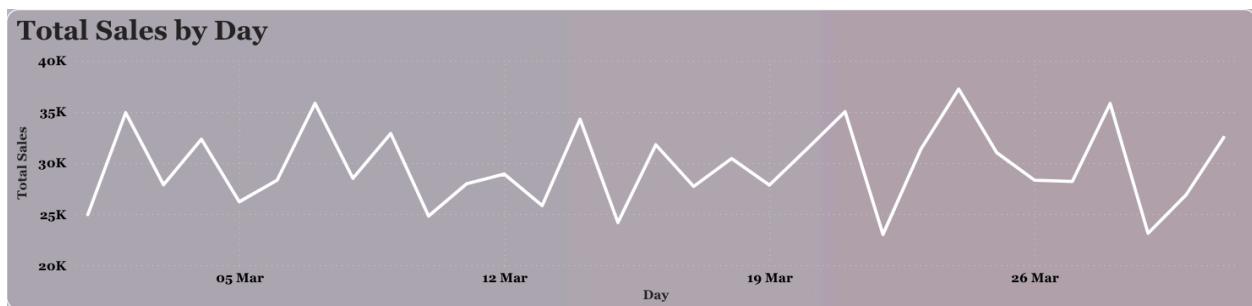


Shows sales figures for various brands, with brands arranged vertically and sales values horizontally. This visual helps in quickly identifying which brands are leading in sales, which are essential for inventory decisions and marketing focus.

Daily Order Counts

Sun		Mon		Tue		March 2017			
						Wed	Thu	Fri	Sat
6812	5	7021	6	8905	7	6312	1	8239	2
7162	12	6590	13	8643	14	6893	8	7911	9
7105	19	7665	20	8750	21	5778	22	7413	23
6930	26	6905	27	8757	28	5639	29	6326	30
								7990	31

Presents a daily breakdown of order counts throughout March, organized by days of the week. This table helps in analyzing day-wise sales volume, useful for understanding customer purchasing patterns.



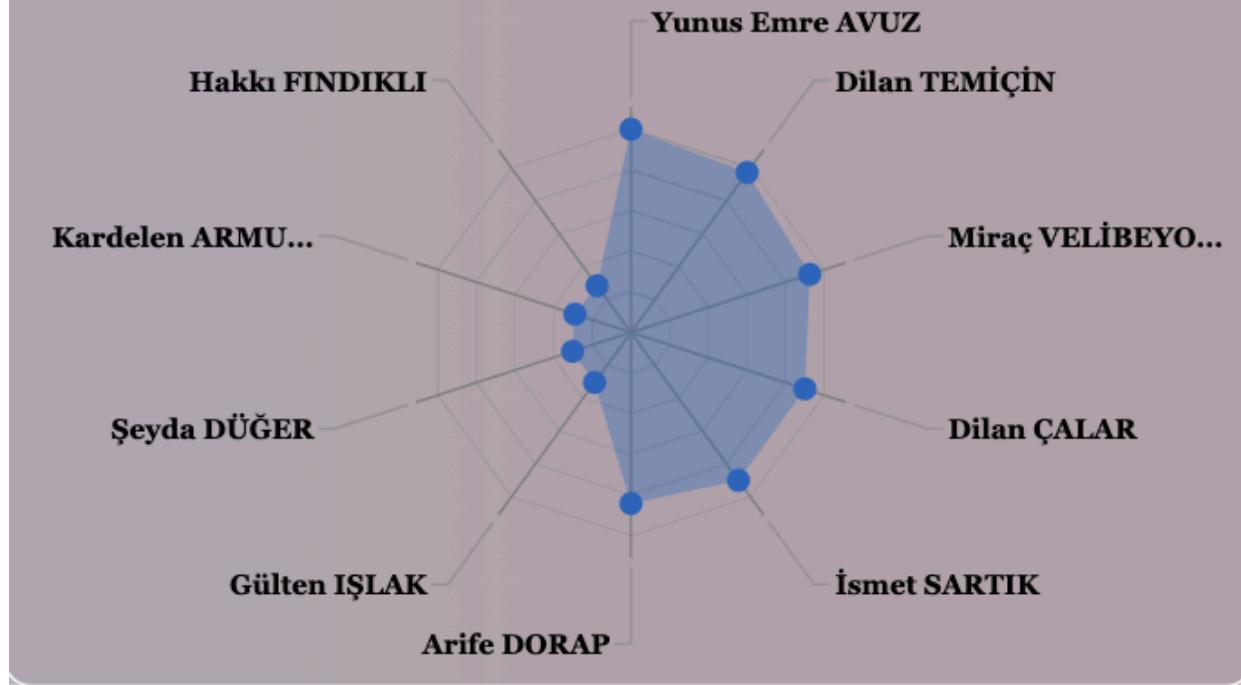
Tracks total sales per day throughout the month of March. The line chart provides a clear visual trend of sales, showing fluctuations and stability over the period.



This chart shows the progress of various salesmen towards their sales targets, with different colors indicating levels of achievement (e.g., red for far from target, yellow for close to target, green for target met).

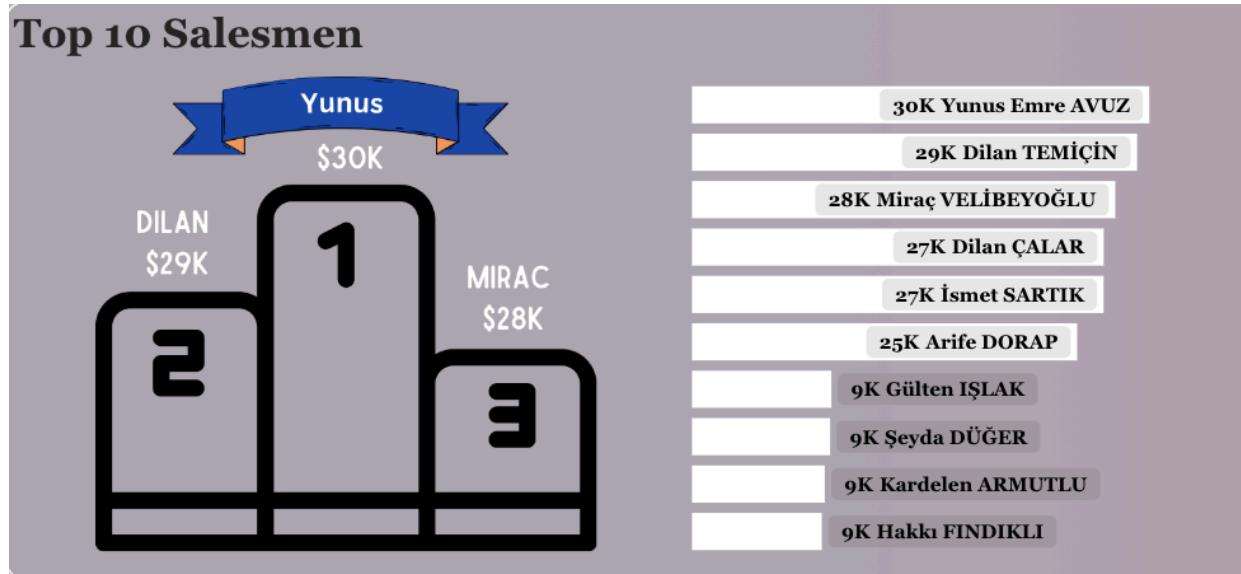
Sales Distribution by Top 10 Salesmen

Axis ● Total Sales



Displays the sales distribution across various performance metrics for the top 10 salesmen, allowing for a comparative analysis of areas like total sales, number of deals closed, and customer reach.

Top 10 Salesmen



Highlights the top three salesmen with their sales figures, presented in a podium format that ranks them first to third based on their sales performance.

SALESMAN	Total Orders	Total Sales
Abdullah BENGİN	337	1,263.25
Abdulsamet TACİM	1887	7,730.12
Abdurrahman ÇAVDARLI	131	434.79
Adil ONDİN	451	1,689.06
Adnan YAGIZ	1769	6,957.45
Aleyna ÇUKUREL	255	1,142.57
Aleyna ERÇETİN	197	730.42
Aleyna KASALAK	153	739.91
Ali Eymen GÖKÇESU	13	85.02
Ali İLTEKİN	88	279.42
Aliye ÇUKURÇAYIR	93	399.95
Aliye GÖKDOĞAN	943	4,029.45
Aliye KÖZEN	57	211.55
Alper FETTAHOĞLU	101	619.40
Alper İNELBARS	260	1,103.09
Alper İZMİRLİ	614	2,258.24
Alya BARAN	180	710.70
Total	225509	919,317.12

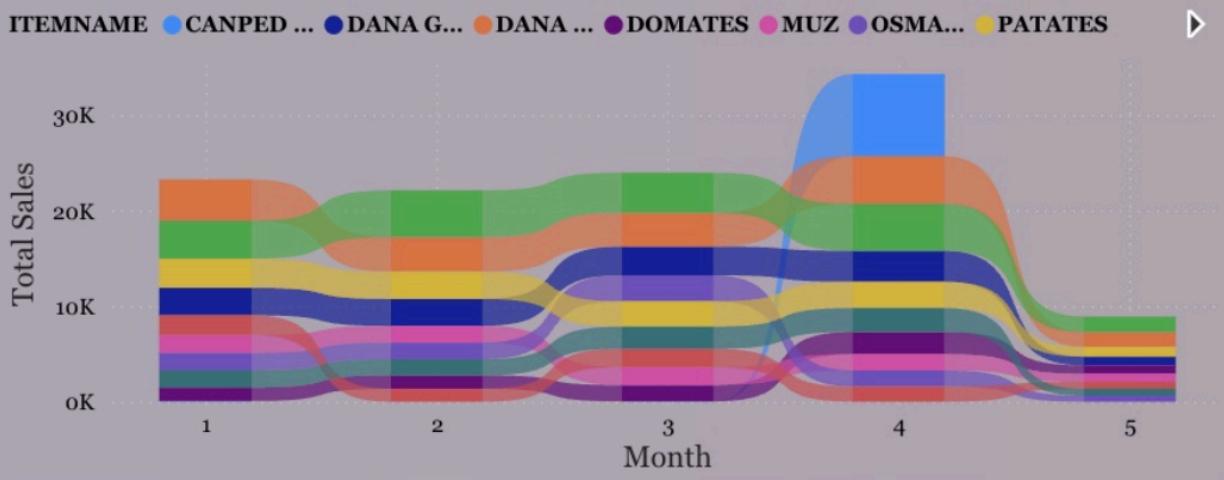
Lists total orders and total sales for each salesmen

Top Profitable Products in Major Cities



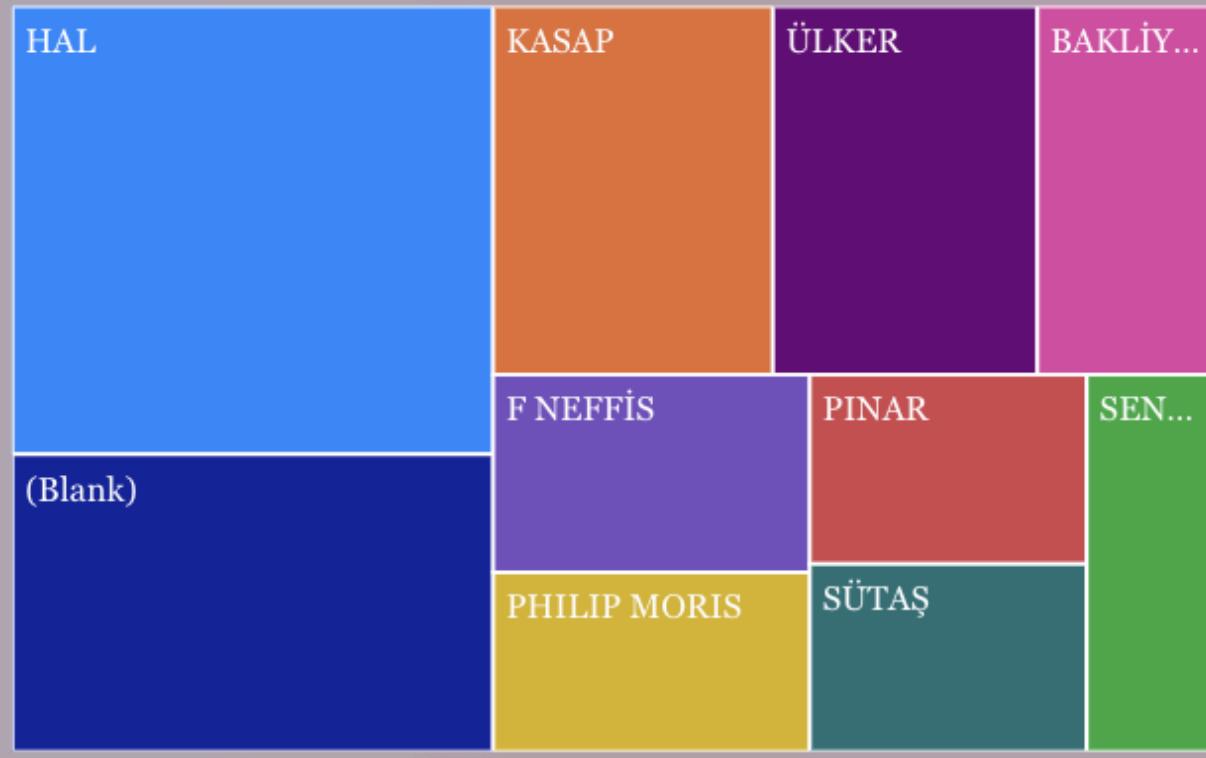
This diagram shows the flow of top profitable products to various major cities, visually representing which products are most successful in specific locations.

Trend of Top 10 Selling Products by Week



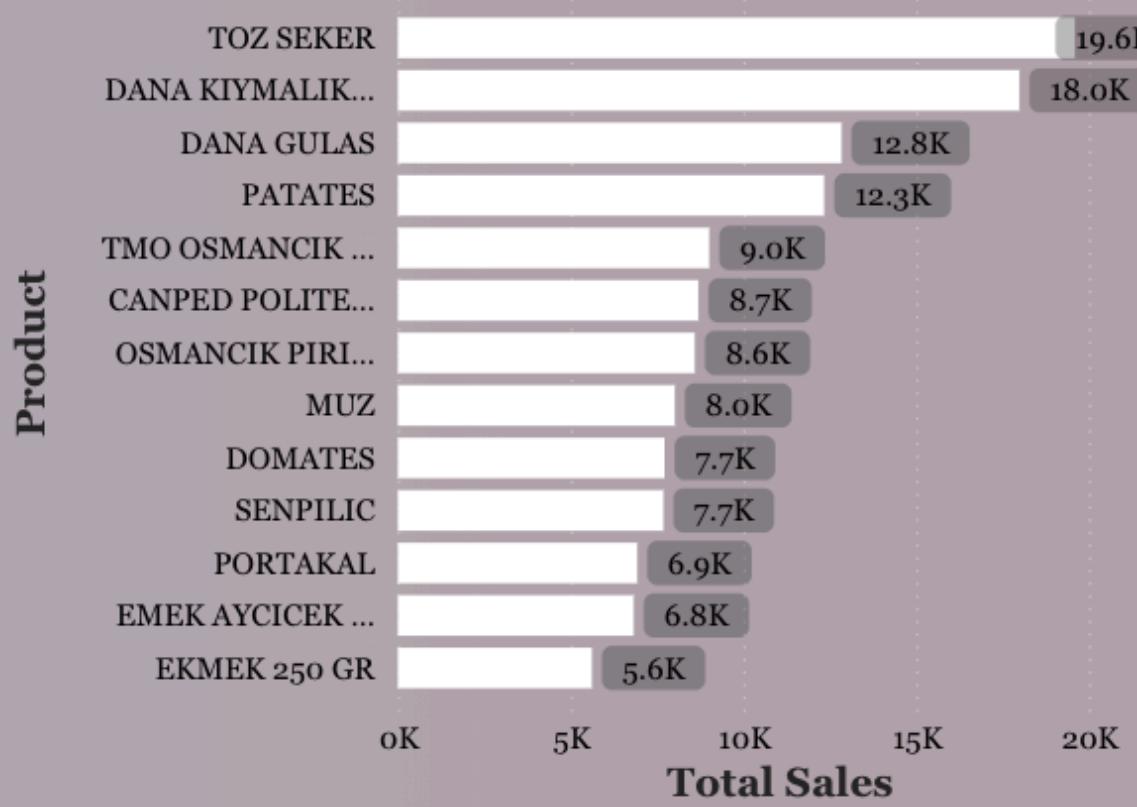
Lists the trend of the top 10 selling products and their sales figures, providing a clear ranking of product popularity based on sales volume.

Sales by Brand, Categories & Products



Displays sales distribution across various brands and their categories, with different colors representing different brands and sizes indicating sales volume.

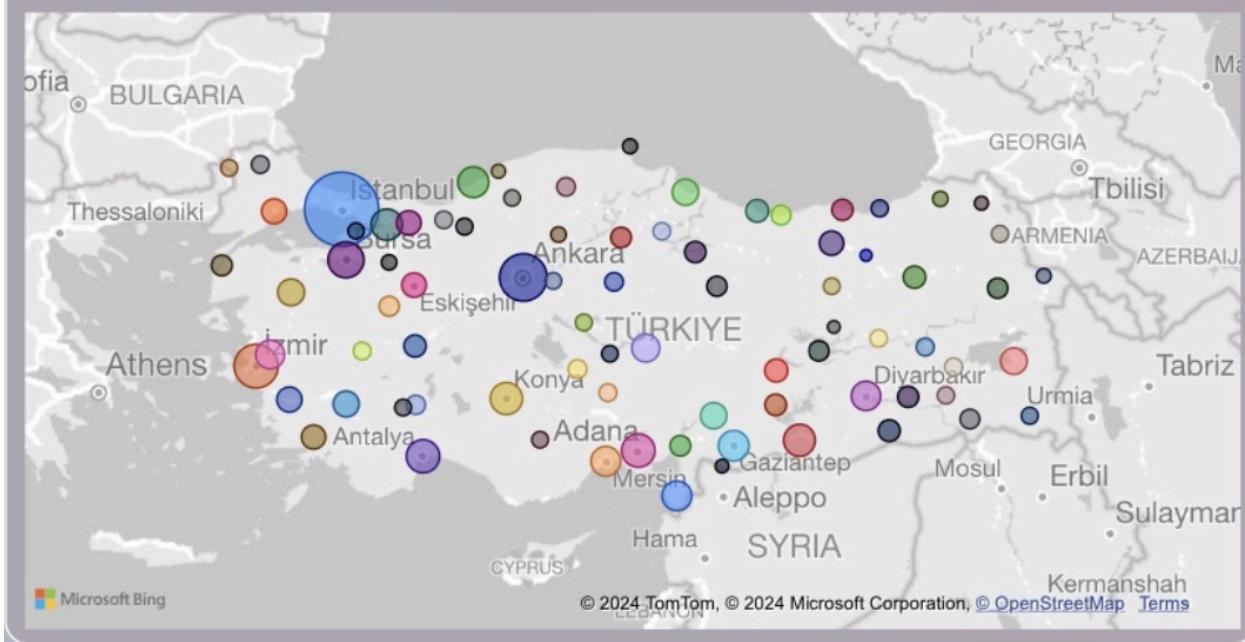
Sales by Top 10 Selling Products



Tracks sales trends of the top 10 selling products over several months, showing how each product's sales volume changes over time.

Sales and Brand Performance by City

CITY • İstanbul • Ankara • İzmir • Bursa • Adana • Antalya • Konya • Şanlıurfa • Kocaeli ▶



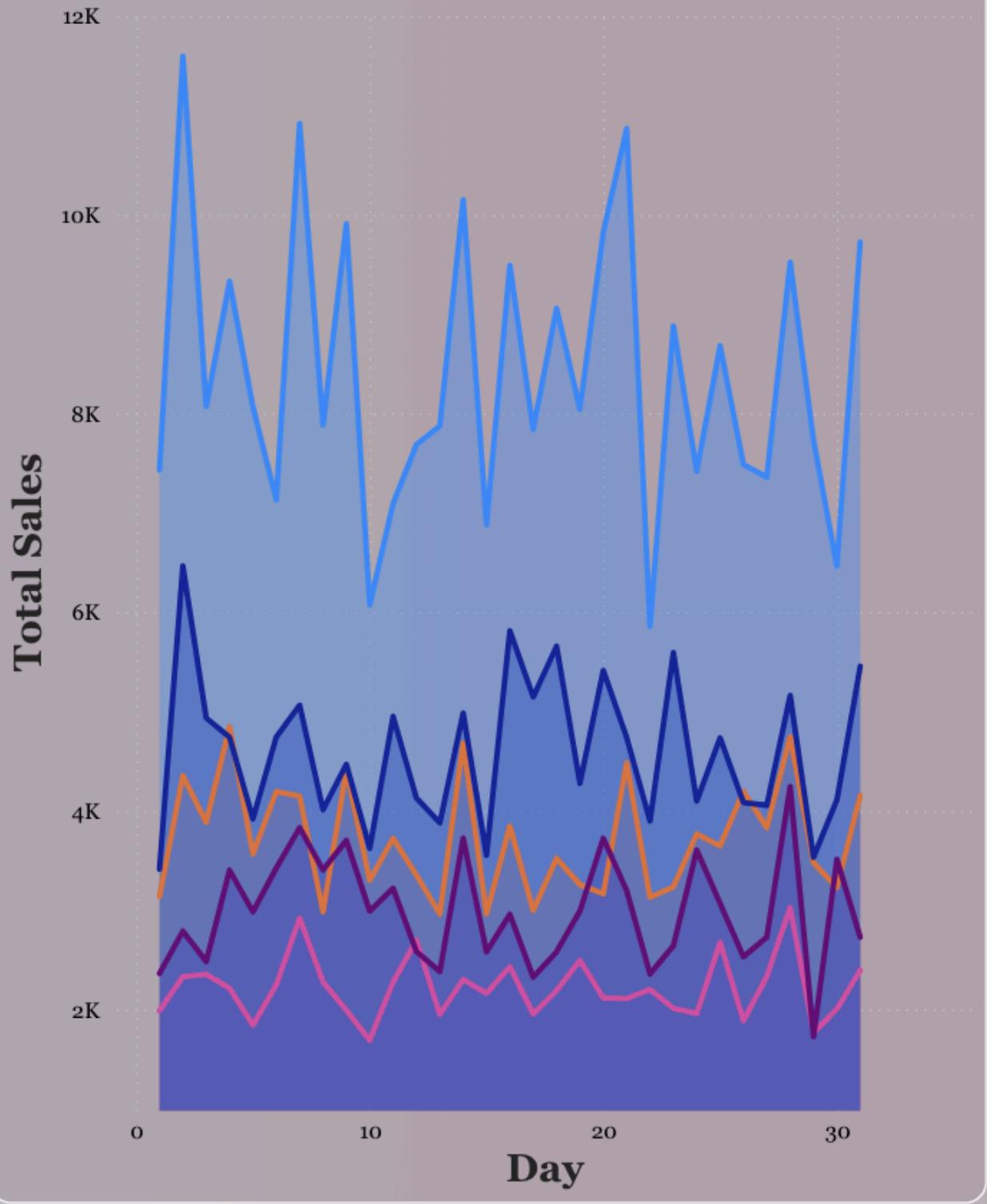
This map displays the geographical distribution of sales across various cities, with bubble sizes representing the order counts, and tooltip showing order counts of the top 10 brands in that city. This helps in visualizing which regions are generating the most orders.

REGION	Active Products	Active Salesmen	Total Orders	Total Sales
✉ Akdeniz	28389	28389	28389	115,157.13
✉ Adana	6235	6235	6235	25,191.18
✉ Antalya	6167	6167	6167	24,897.98
✉ Burdur	819	819	819	3,096.76
✉ Hatay	4682	4682	4682	18,647.58
Hatay Subesi	4682	4682	4682	18,647.58
✉ Isparta	1234	1234	1234	5,067.94
✉ Kahramanmaraş	3107	3107	3107	13,188.26
Kahramanmaraş Subesi	3107	3107	3107	13,188.26
✉ Mersin	4601	4601	4601	18,601.42
✉ Osmaniye	1544	1544	1544	6,466.01
✉ Doğu Anadolu	16959	16959	16959	68,773.06
✉ Ege	32832	32832	32832	133,247.39
✉ Güneydoğu Anadolu	22996	22996	22996	92,747.19
✉ İç Anadolu	34909	34909	34909	142,618.30
✉ Aksaray	1243	1243	1243	4,545.32
Total	225509	225509	225509	919,317.12

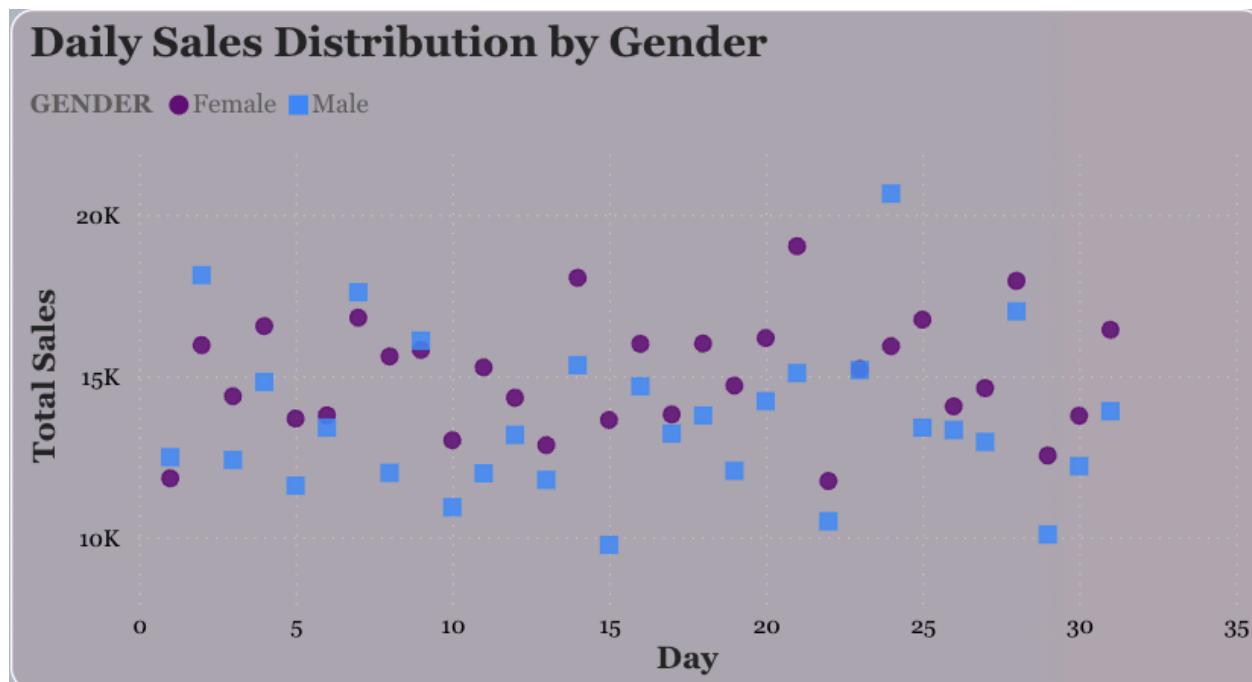
Provides a snapshot of key metrics such as active products, active salesmen, total orders, and total sales for each region listed. This table is instrumental in quickly assessing the scale of operations and performance in various regions.

Distribution of Daily Sale by Region

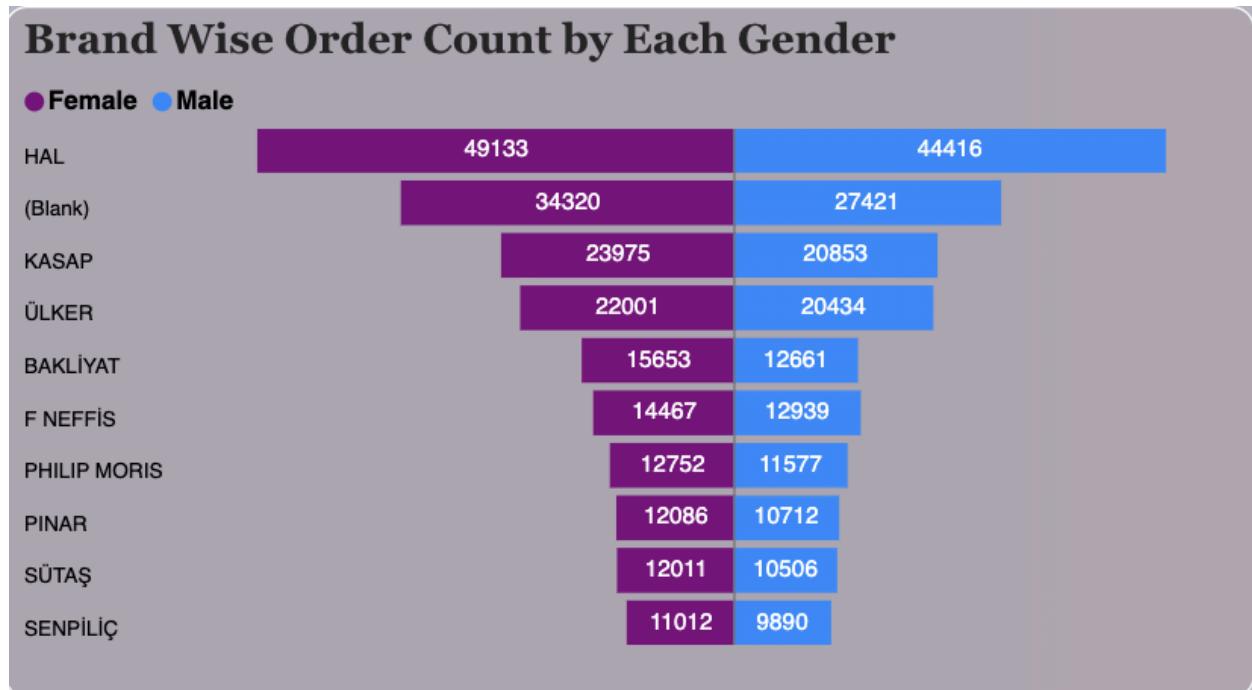
REGION ● Akdeniz ● Doğu Anadolu ● Güneydoğu Anadolu ● İç Anadolu ● Marmara



Shows daily sales trends across different regions, with each line representing a specific region. This visual is crucial for tracking sales fluctuations and identifying trends or seasonal effects in different areas.



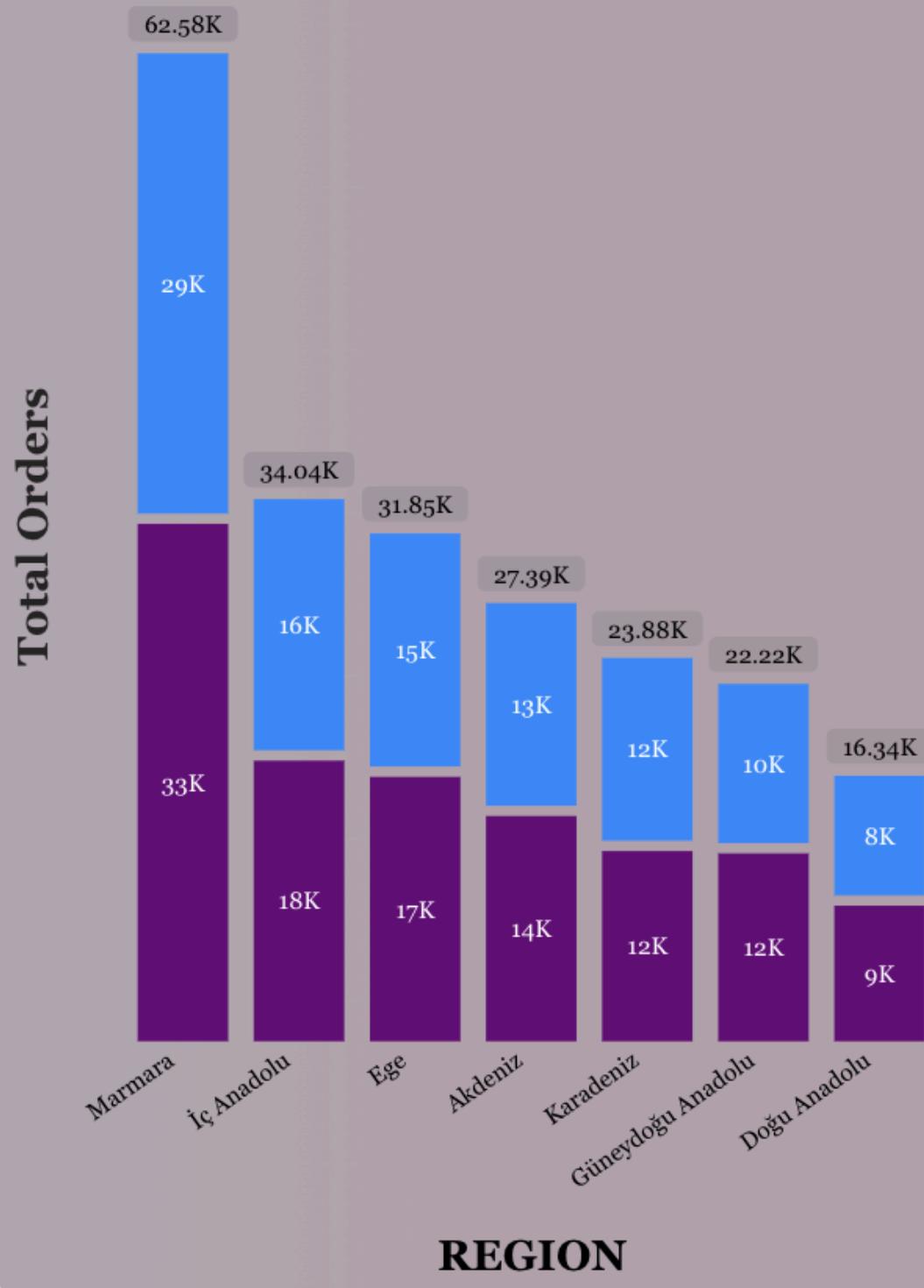
This plot displays daily sales values by gender throughout the month, with different colors representing male and female customers. It illustrates fluctuations in purchasing patterns and highlights any significant gender-based preferences or trends.



Compares the order count for different brands, broken down by gender. This chart helps in evaluating which brands are more popular among male or female customers.

Region Wise Order Count by Each Gender

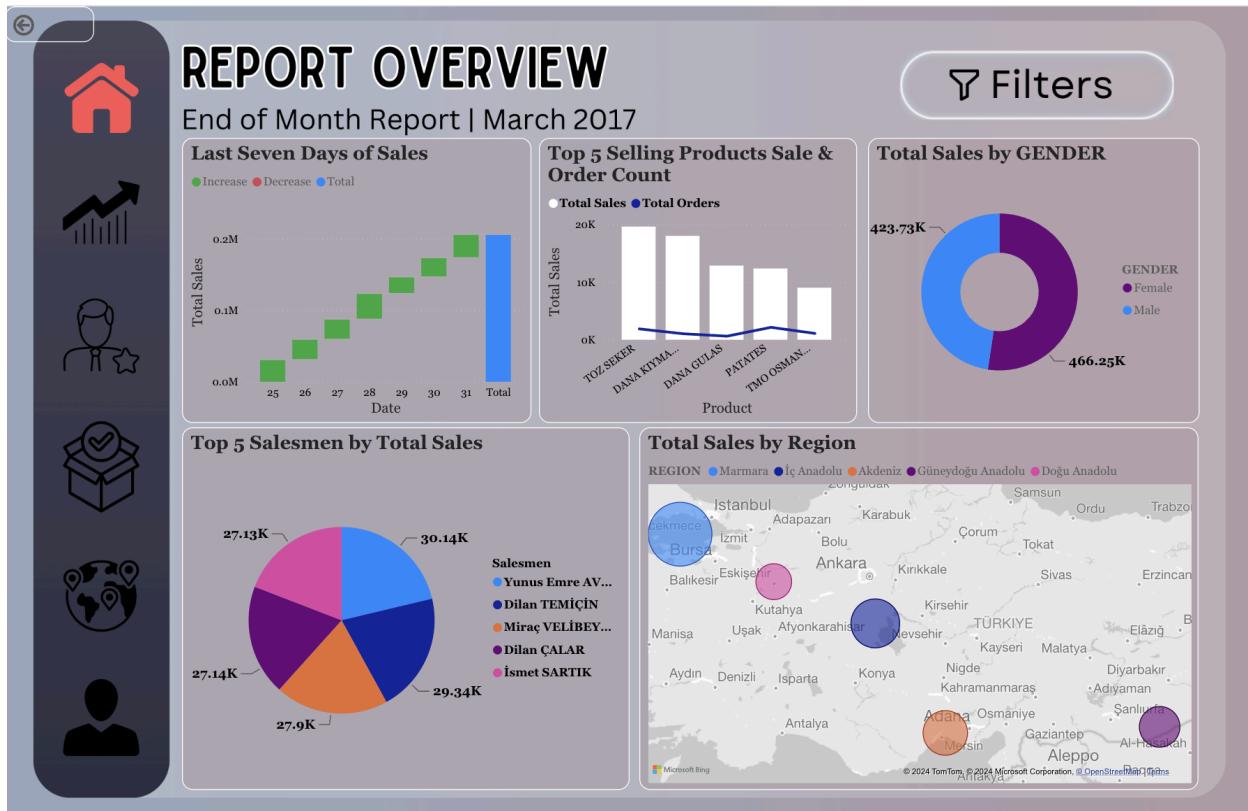
GENDER ● Female ● Male



Shows the total order counts by region, segmented by gender. This visualization is useful for understanding how gender distribution varies across different regions.

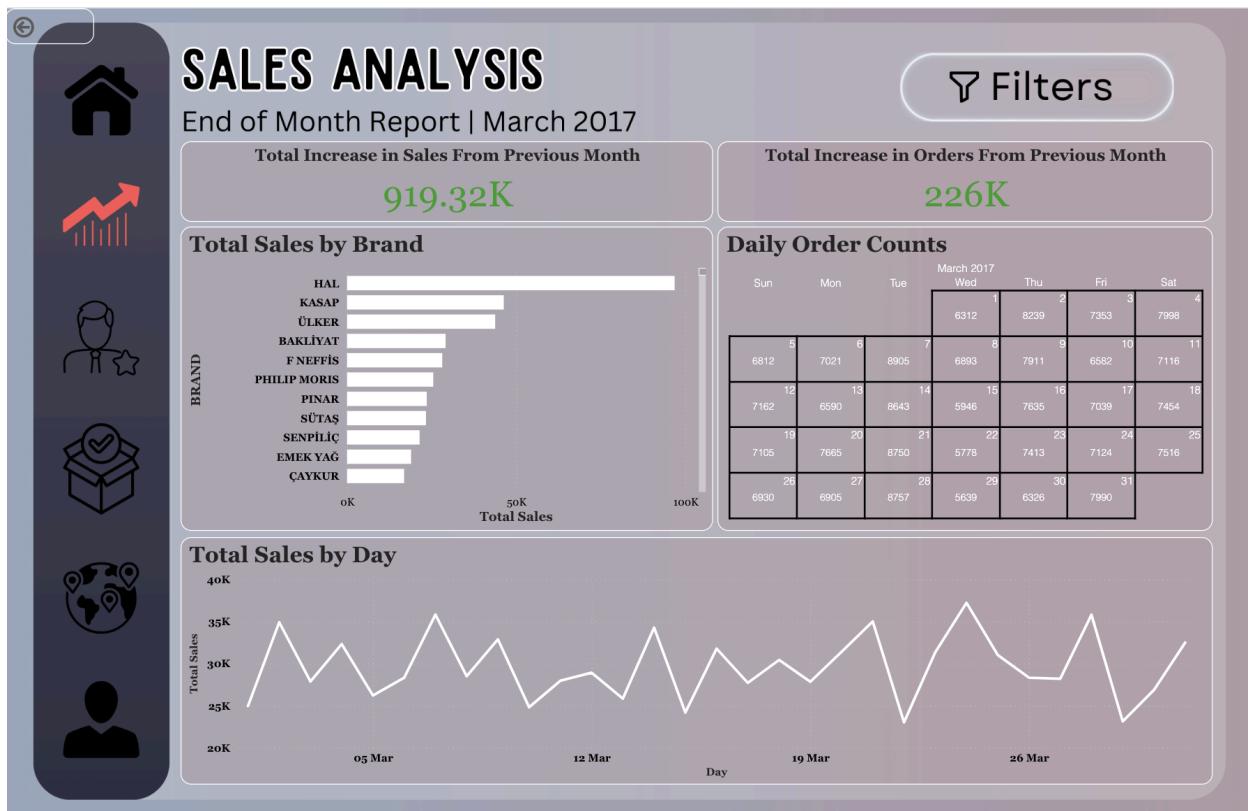
Dashboards Used

Report Overview Dashboard



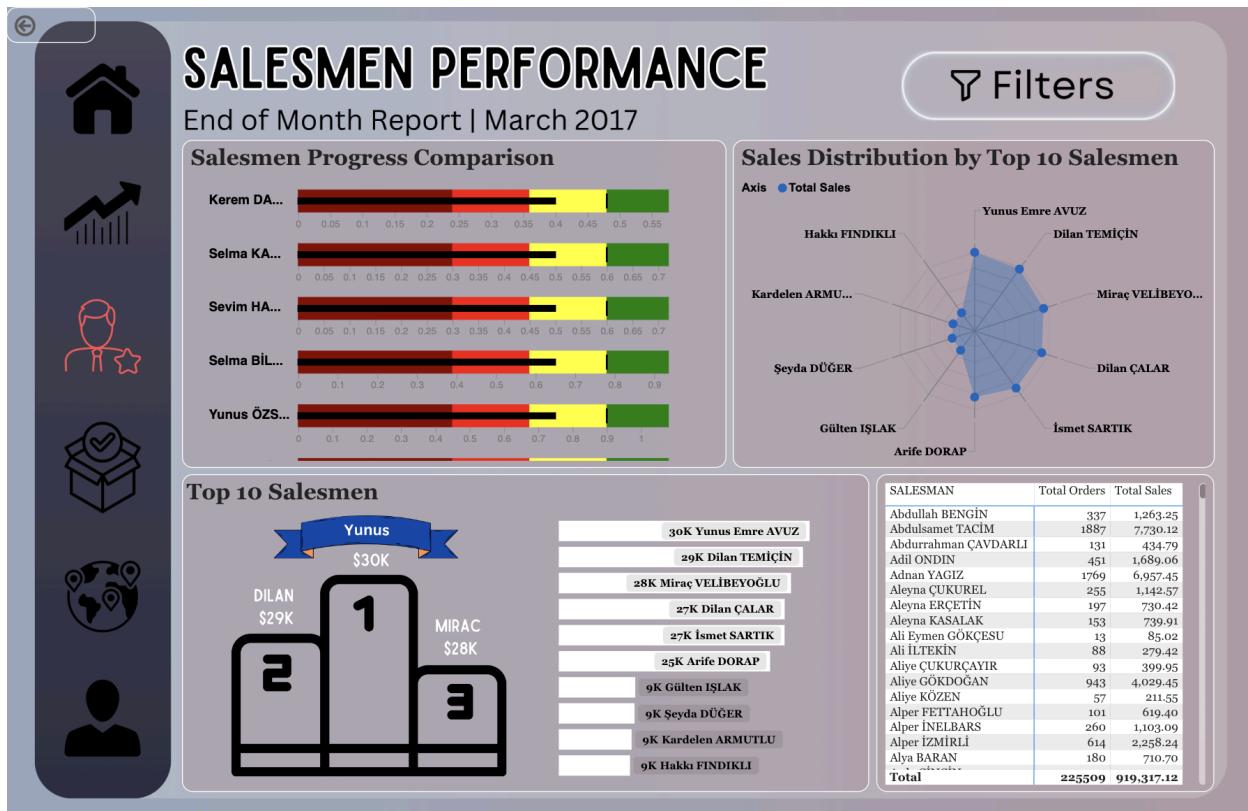
Dashboard Analysis: This dashboard synthesizes data from various business aspects, presenting a cohesive overview to executives for strategic planning and immediate decision-making for the next month.

Sales Analysis Dashboard



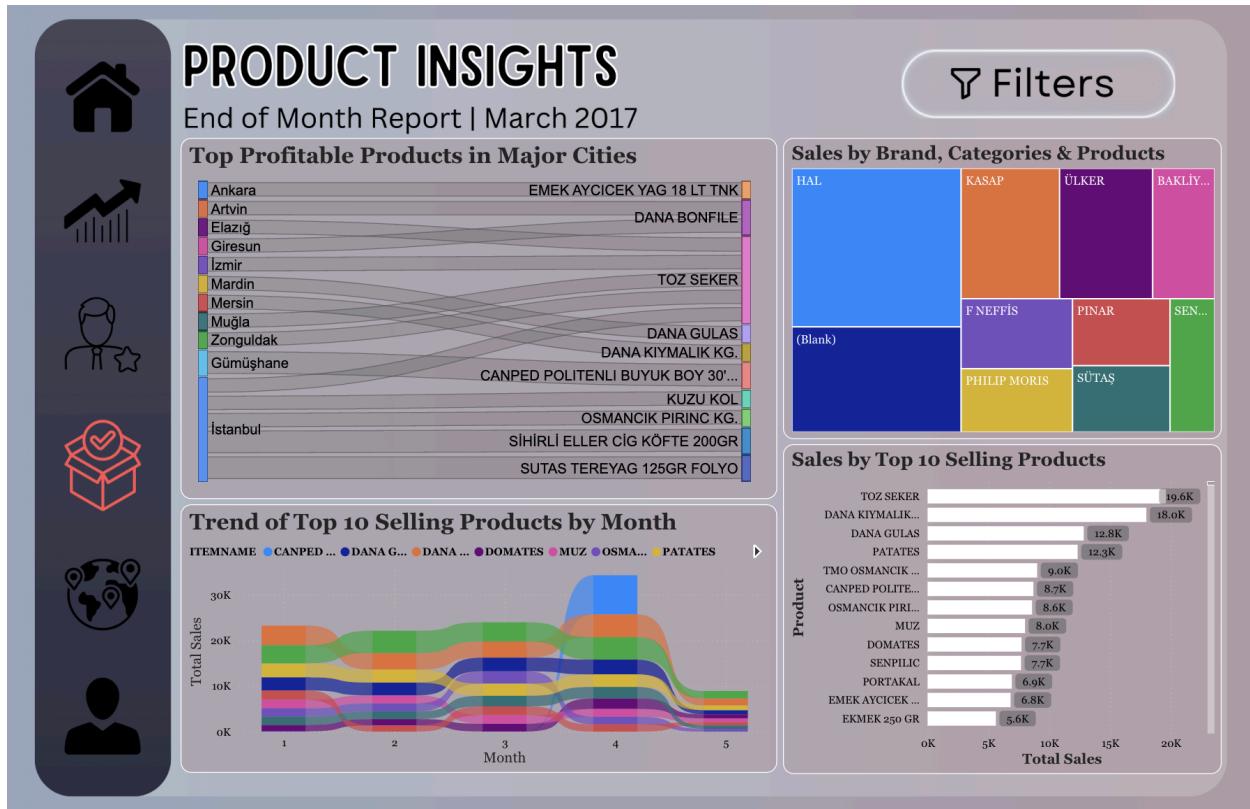
Dashboard Analysis: This dashboard gives a comprehensive view of daily and monthly sales performances and brand contributions, critical for adjusting sales targets and strategies for the next month.

Salesmen Performance Dashboard



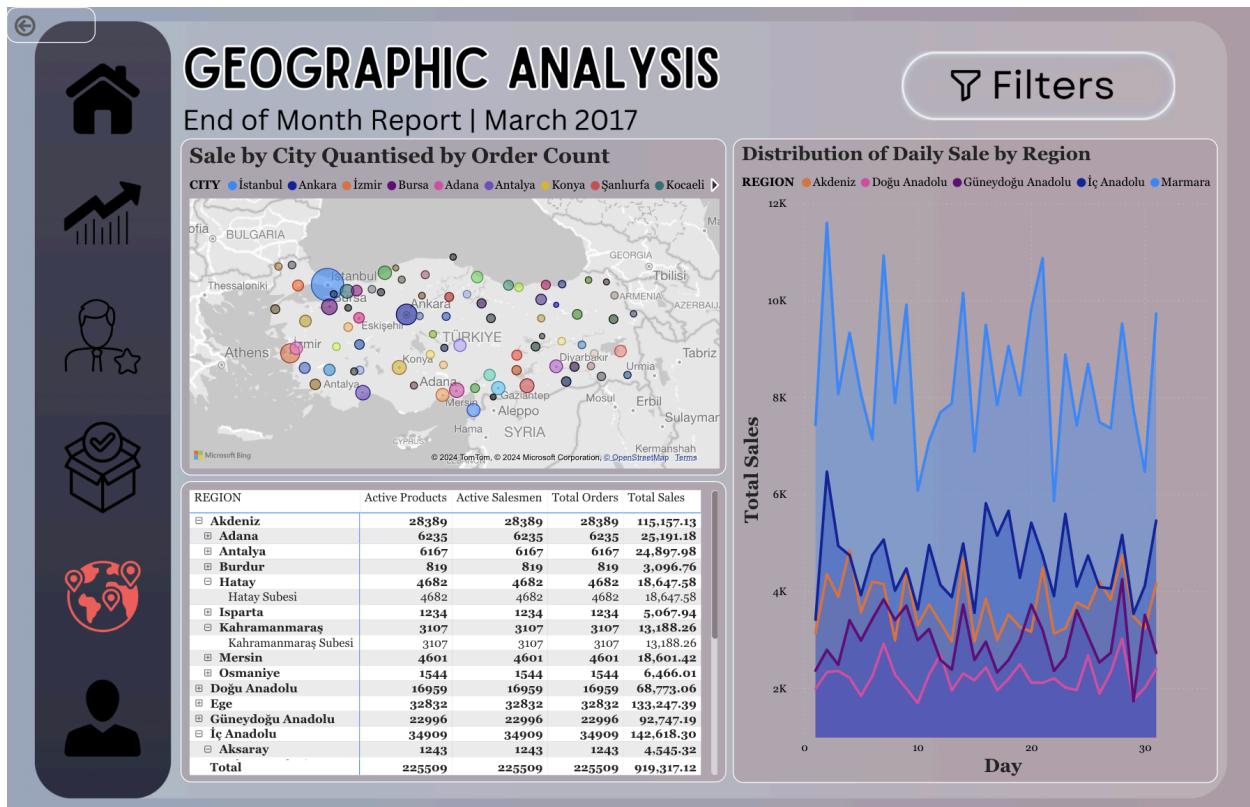
Dashboard Analysis: Highlights the impact of top salesmen and their performance distribution, assisting in training needs assessment and territory assignments.

Product Insights Dashboard



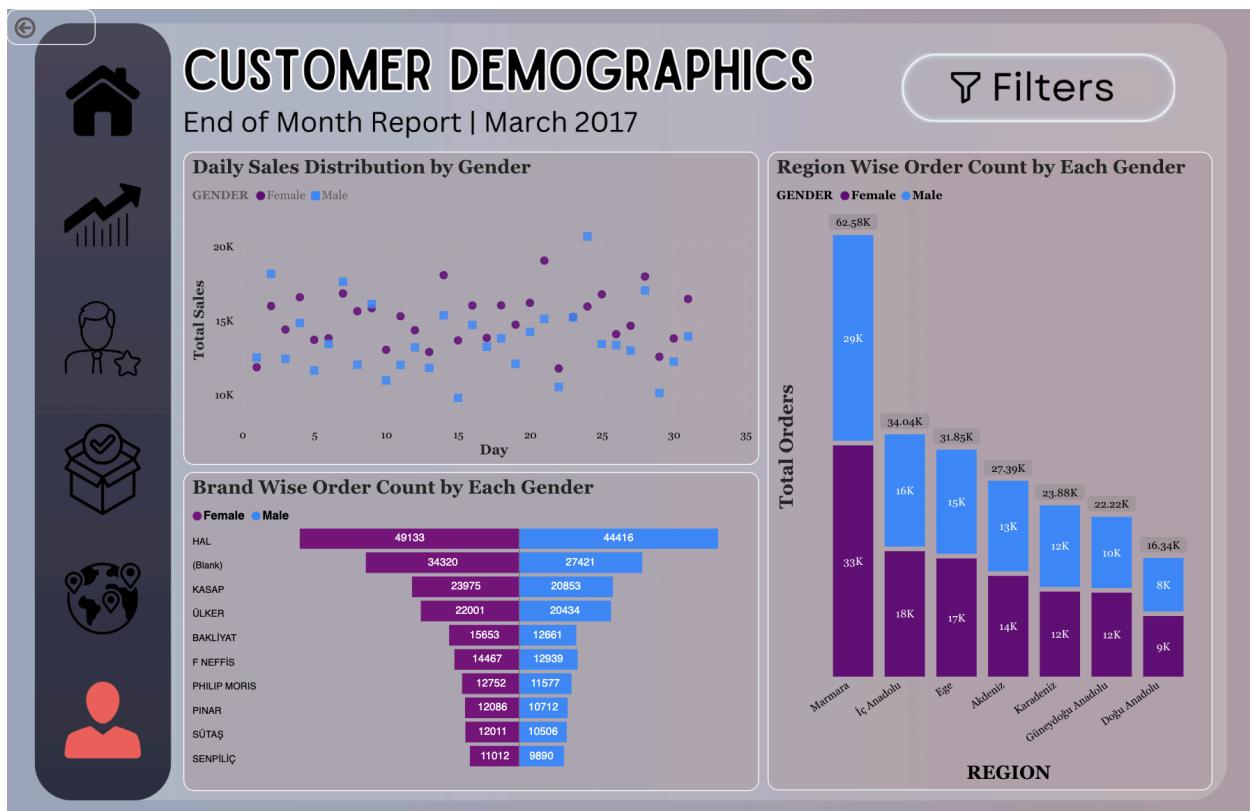
Dashboard Analysis: Provides insights into product popularity and profitability, informing product development, stocking, and promotional activities.

Geographic Analysis Dashboard



Dashboard Analysis: This dashboard is crucial for geographic market analysis, helping to decide where to focus marketing and sales efforts.

Customer Demographics Dashboard



Dashboard Analysis: Offers detailed insights into customer demographics, supporting tailored marketing strategies and product offerings to different demographic groups.

Solving The Business Problem Using Story

Revisiting The Business Problem

The core business problem is to assess the company's performance at the end of march, to strategize for the upcoming month. The goal is to identify trends, anomalies, opportunities, and areas needing improvement across various facets of the business, including sales, product performance, salesforce effectiveness, geographic sales distribution, and customer demographics. By analyzing these areas, the company aims to optimize resource allocation, enhance product positioning, improve sales strategies, and better understand customer preferences.

Breaking Business Problem Into BI Queries

1. Key Insights for Swift Decision-Making

Purpose: Serves as an executive summary, integrating key findings from all other dashboards for quick reference and decision-making.

2. Sales Performance Insights

BI Query: Analyze overall sales performance, focusing on daily sales trends, sales by brand, and the increase in sales from the previous month.

3. Assessing Sales Team Effectiveness

BI Query: Evaluate the effectiveness and performance of the sales team by analyzing individual and collective contributions.

4. Product Sales Analysis

BI Query: Assess product sales, focusing on top-selling and top profitable products, and their trends over recent months.

5. Geographical Sales Distribution

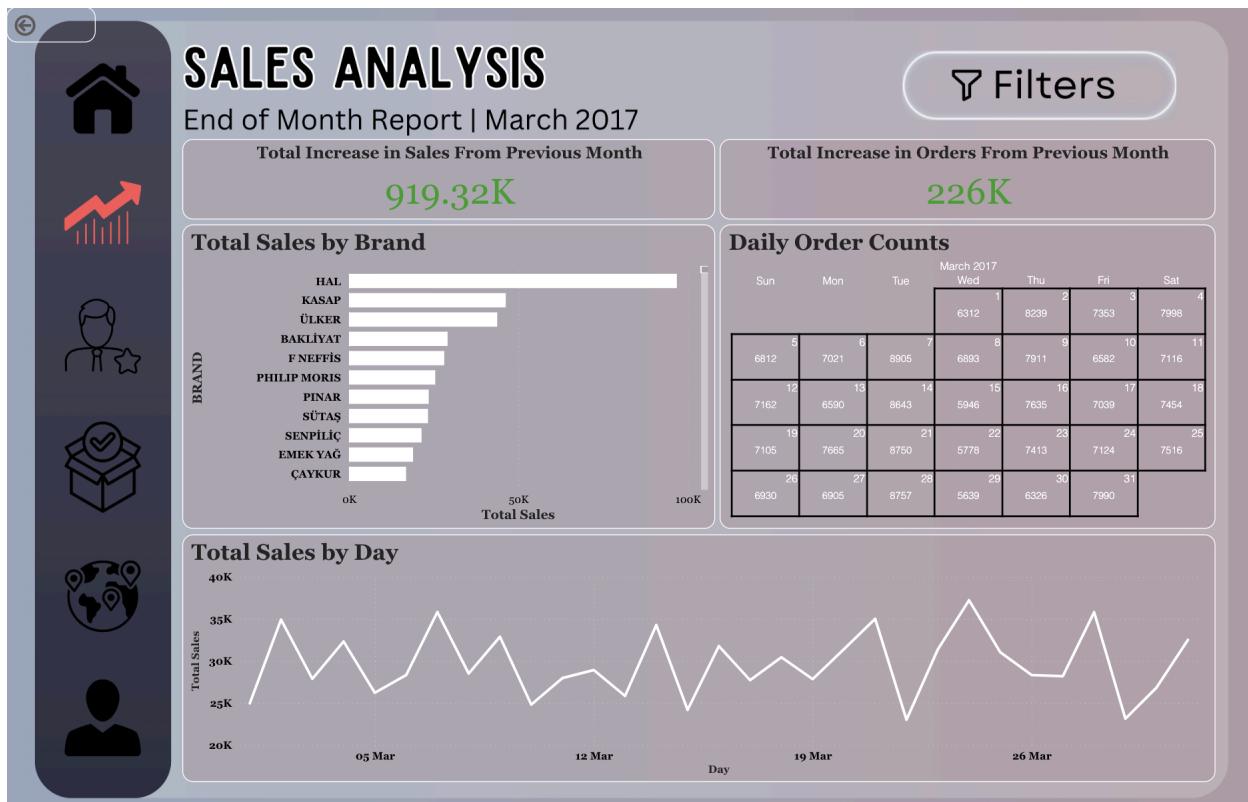
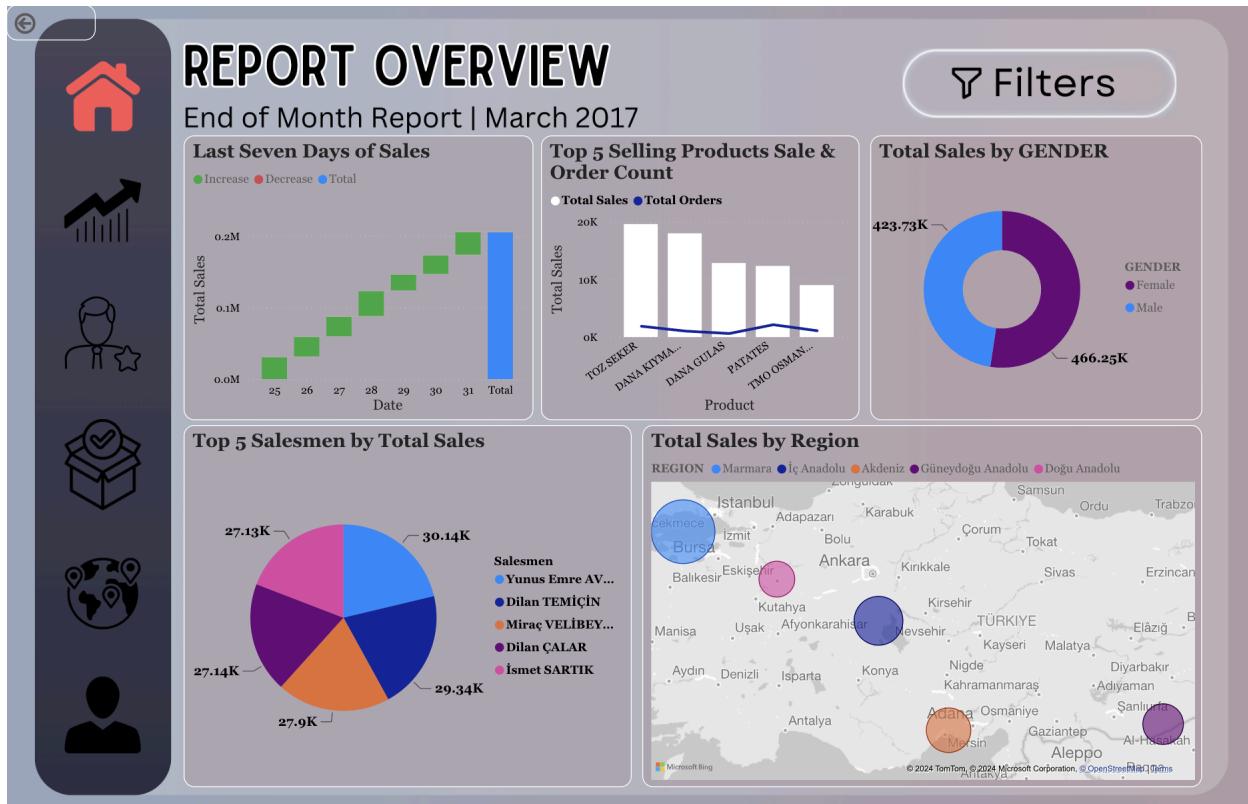
BI Query: Examine sales distribution by region and city, correlating it with order counts and active market parameters.

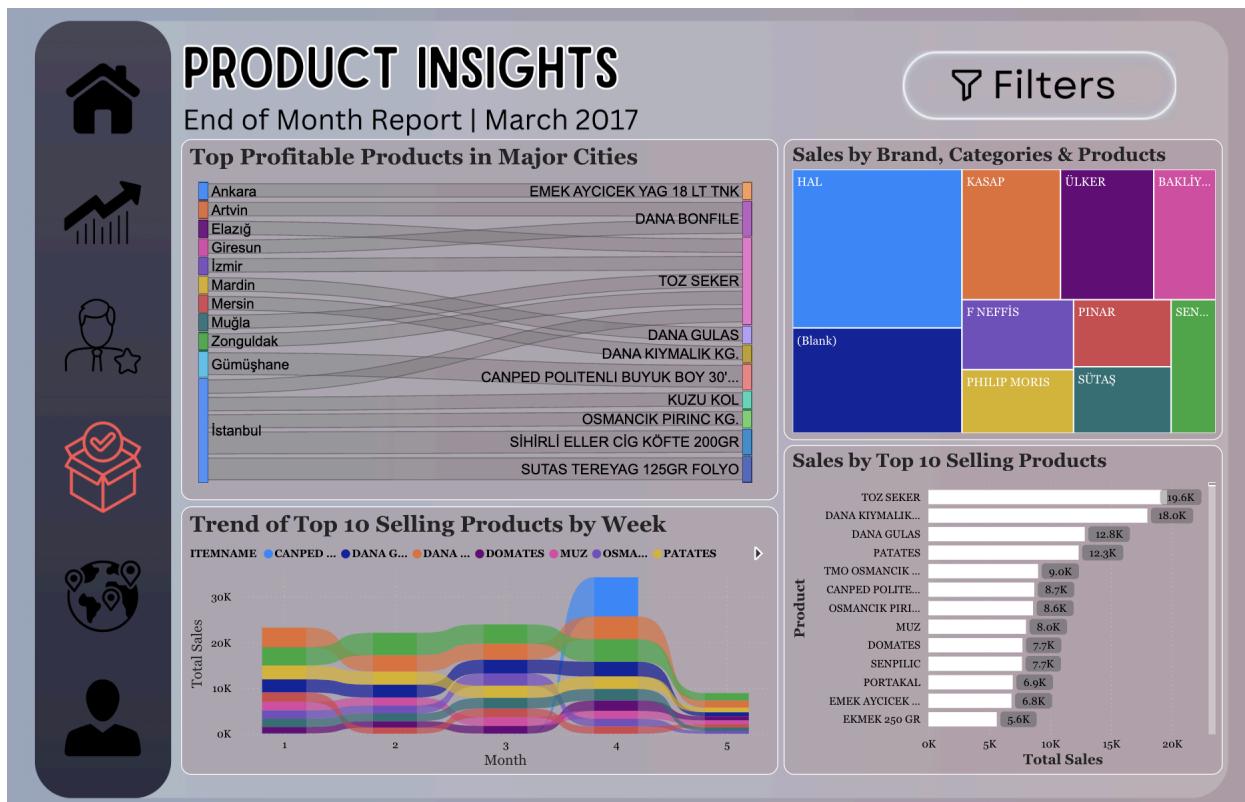
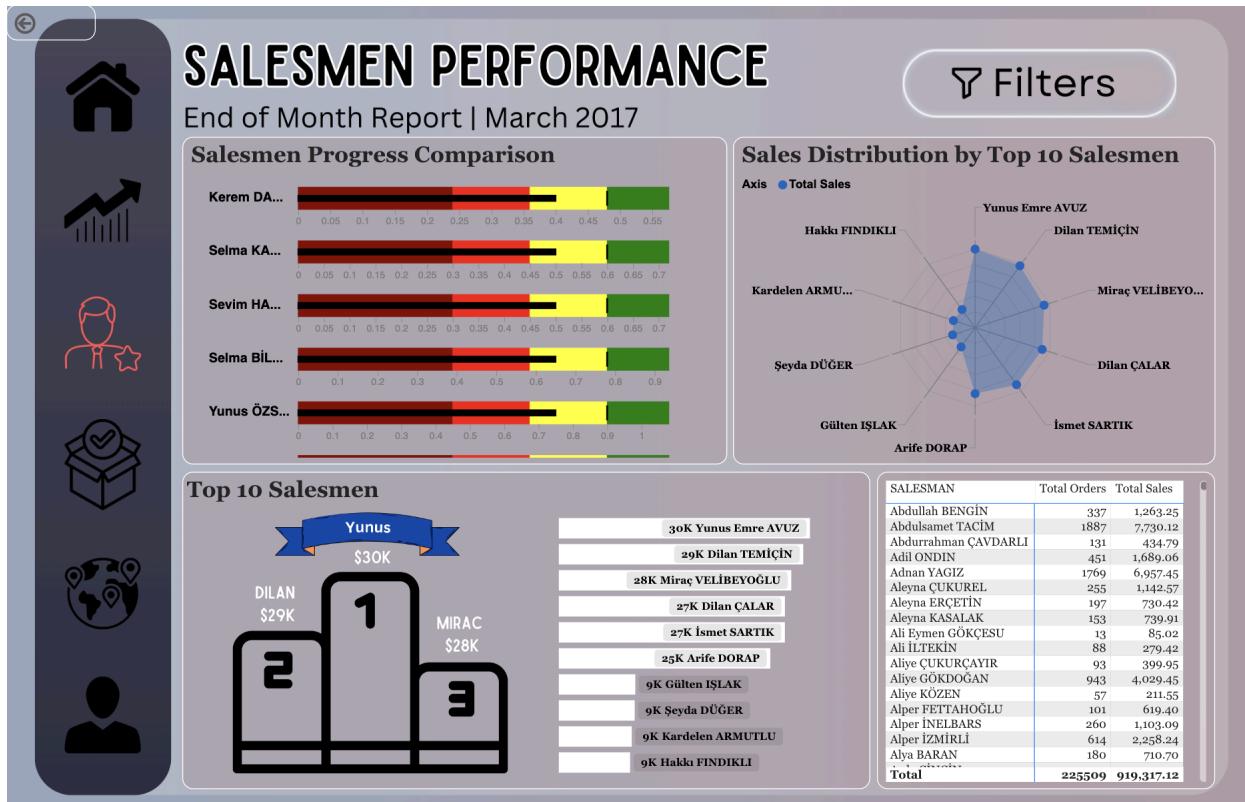
6. Customer Base Analysis

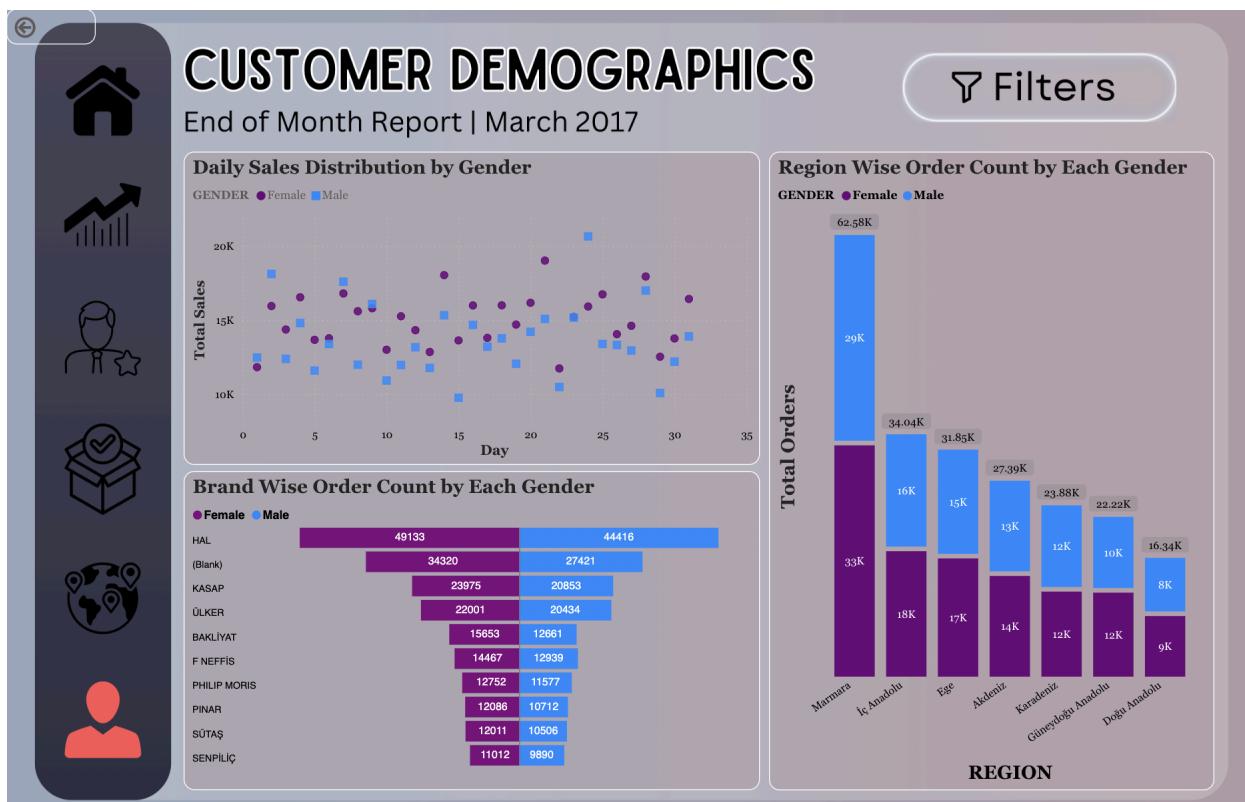
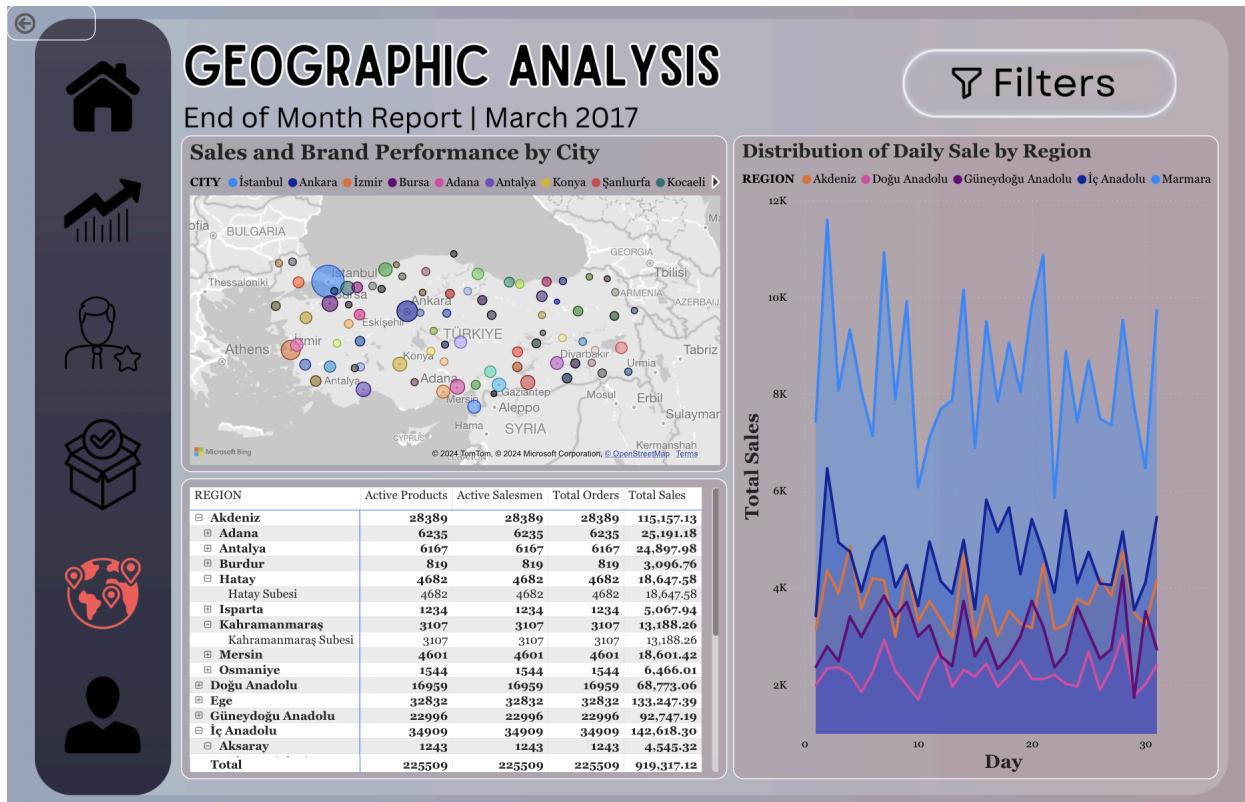
BI Query: Dive deep into the customer base to understand sales distribution by gender and regional preferences.

Sequence of Dashboards To Explain Story









Team Member Contributions

All work was equally divided. Most of the work was done collaboratively, either at the university or remotely using Discord. In each part of the project, the final contribution by each member can be concluded to be 50-50.