Logo

Description automatically generated

**Abhinav Nathari, Hansika Jaryal, Ninad Patil, Sahil Labhe**

# **Project - Phase I: Planning**

**“**

**Due Date: April 09, 2023**

**SECTION 1: DATASET DESCRIPTION**

# **Introduction to Dataset:**

**Name of the Dataset: Online Automotive Sales Statistics'23 (Volkswagen)**

There are a number of publicly available datasets that provide statistics on the Automotive Sales (on Kaggle and related websites), however there aren't many Volkswagens specific datasets available online. The largest online real estate and automobile sales platform in Turkey provided the data for this dataset. The information only includes vehicles under the Volkswagen brand and spans the three months from January 1 to March 31, 2023.

This dataset includes data that was scraped from 7216 customers in order to assess the automotive market. For various vehicle categories, costs, color, and locations, a row in the dataset correlates to other columns. Numerous null items are included in the columns for Date and Fuel and can be pre-processed before real visualizations are created. The dataset offers insightful information on Volkswagen automobile sales and advertising patterns in Turkey during the first quarter of 2023. The information may be used to spot patterns and trends in customer behavior, such as the most popular models, transmission styles, and fuel types. The information may also be used to determine which cities have the most demand for Volkswagen vehicles and to assess the success of advertising initiatives.

# **Row and column counts:**

There are 7216 data points (rows) and 13 attributes (Columns) in the Online Automotive Sales Statistics'23 (Volkswagen) dataset.

# **Specification of Data Types for a column:**

|  |  |
| --- | --- |
| **Attribute Name** | **Data Type** |
| Customer ID | Ordinal |
| Advertisement number | Categorical |
| Brand | Categorical |
| Model | Categorical |
| Variant | Categorical |
| Year | Interval |
| Kilometer | Categorical |
| Color | Categorical |
| Transmission | Categorical |
| Fuel | Categorical |
| City | Categorical |
| AD Date | Interval |
| Price | Ordinal |

* Other data scientists have used this dataset for data analysis purposes, such as exploratory data analysis and data cleansing.
* This dataset can visualize:
  1. several applications for business team.
  2. the top five sold car variants in a certain category.
  3. the dataset's top 5 priciest cars.
  4. The number of cars across various categories
  5. Kilometers based on vehicle fuel type.
  6. Sales based on the advertisements.

# **Description of an attribute:**

We chose the “Online Automotive Sales Statistics'23 (Volkswagen) dataset”. This dataset contains multiple columns containing information about various Volkswagen car sales available on the platform. These columns aid in the interpretation of various information about the sales, distance, date, and car category. The attributes described in the dataset are listed below.

* **Customer ID:** A special identification number for the advertiser's client.
* **Advertisement number:** The advertisement's special identification number.
* **Brand:** The name of the automobile (in this dataset, it is always Volkswagen).
* **Model:** The car's make and model (e.g., Golf, Polo, Passat, etc.).
* **Variant:** The car's version (e.g., 1.6 FSI Midline, 2.0 TDI Comfortline, etc.)
* **Year:** the year when the vehicle was produced.
* **Kilometer:** The length of time the vehicle has been on the road (in kilometers).
* **Color:** The car's color.
* **Transmission:** If it is (manual or automatic).
* **Fuel:** The kind of gasoline that the vehicle uses. (e.g., petrol, diesel, hybrid, etc.)
* **City:** The location of the advertising.
* **Date:** The date that the advertising was posted, or "ad\_date"
* **price:** The car's asking price.

# **The domain of Columns:**

|  |  |
| --- | --- |
| **Column Name** | **Interval range** |
| Customer ID | Integer values [0 - 7216] |
| Advertisement number | Integer values [195460036 – 777835401] |
| Brand | Accepts a String value. |
| Model | Accepts a String value. |
| Variant | Accepts a String value. |
| Year | It is a date-taking interval attribute. |
| Kilometer | Accepts an Integer value and that value varies |
| Color | Accepts a string value |
| Transmission | Accepts string either Manual or Automatic. |
| Fuel | Accepts string either Gasoline, Gas or Diesel. |
| City | Location in the turkey and it accepts a String value. |
| AD Date | It is a date-taking interval attribute. |
| Price | Integer values [100-2109] |

**“SECTION 2: PROSPECTIVE DASHBOARD USERS.”**

**“This dataset's potential dashboard users include:”**

“Below are the prospective dashboard users of this dataset:”

1. **“Sales Managers:** Sales managers can use a dashboard to monitor the performance of their sales team, track sales metrics, and identify trends to make data-driven decisions.”
2. **“Marketing Managers:** Marketing managers can use a dashboard to analyze marketing campaigns, track leads, and measure the effectiveness of different marketing channels.”
3. **“Operations Managers:** Operations managers can use a dashboard to monitor inventory levels, track orders, and manage logistics to ensure timely delivery of vehicles.”
4. **“Customer Service Representatives:** Customer service representatives can use a dashboard to manage customer inquiries, track customer interactions, and identify opportunities to improve customer satisfaction.”
5. **“Finance Managers:** Finance managers can use a dashboard to track revenue and expenses, monitor financial performance, and analyze profitability.”
6. **“Business Owners:** Business owners can use a dashboard to get a high-level view of their business operations, identify areas for improvement, and make strategic decisions based on data insights.”
7. **“Data Analysts:** Data analysts can use a dashboard to collect and analyze data, create visualizations, and generate reports to support business decisions.”
8. **“Customers:** Customers can use a dashboard to search for cars, compare prices, and make purchases online. They can also use the dashboard to track their orders, view order history, and receive updates on the status of their purchases.”
9. **“Data Scientists:** The rating is determined by data scientists based on user reviews and feedback.”

**SECTION 3:LIST OF USER REQUIREMENTS & POTENTIAL QUESTIONS**

**List of user requirements:**

1. Analysts in the automotive sector can use the dataset to evaluate the sales trends of Volkswagen automobiles in Turkey during the first quarter of 2023 and gain a better understanding of the market demand for various car models and variants.

2. The dataset can be utilized by marketing and advertising teams to assess the efficiency of the advertising campaigns they run for Volkswagen automobiles in various cities and to make adjustments to such ads as necessary in order to appeal to a larger audience.

3. The dataset can be used by car dealers and sellers to determine the price that should be asked for Volkswagen automobiles based on the model, year, and mileage of the automobiles, along with other aspects that can affect the price.

4. The dataset can help automobile makers gain a deeper understanding of the preferences and requirements of customers, allowing for the creation of new Volkswagen car models and variants that are better suited to meet those requirements.

5. The dataset can be utilized by financial analysts to forecast the future sales patterns of Volkswagen automobiles in Turkey as well as estimate the income that is generated by the sales of automobiles.

6. The dataset can be used by car rental companies to determine the demand for various Volkswagen car models and variants in different cities, allowing the companies to adjust the size of their fleets and the prices they charge accordingly.

7. Researchers and policymakers in Turkey can use the dataset to investigate the effects that increased sales of Volkswagen automobiles have had on the country's environment, fuel consumption, and transportation infrastructure.

8. Students and researchers can make use of the dataset to investigate a variety of statistical and machine learning strategies for data analysis and to gain practical experience in applications that are relevant to the real world.

9. Data scientists can use the dataset to build predictive models, which can then be used to forecast the demand for various Volkswagen car models and variants based on a number of different factors, such as price, year, kilometer, city, and the type of transmission.

**Potential Questions:**

1. How does the average price of cars vary by brand and model?
2. What is the distribution of car prices by year of manufacture?
3. How does the number of kilometers driven affect the price of a car?
4. Which cities have the highest and lowest average car prices?
5. What is the most popular color of cars being advertised in the dataset?
6. How does the distribution of car prices vary by fuel type?
7. How does the distribution of car prices vary by transmission type?
8. How has the number of car advertisements posted on each date changed over time?
9. How does the number of car advertisements posted vary by city?
10. How does the average price of cars vary by both year and brand?
11. How does the average price of Volkswagen cars vary by city and fuel type?
12. What is the relationship between the number of previous owners and the price of Volkswagen cars?
13. Which model and year of Volkswagen cars have the highest and lowest resale values?
14. How does the distribution of kilometers driven vary by transmission type for Volkswagen cars?
15. Is there a relationship between the mileage of Volkswagen cars and their age?
16. Which cities have the highest and lowest demand for Volkswagen cars, and what factors may be contributing to these differences?

**SECTION 4: REFERENCES**

**Dataset Link:** <https://www.kaggle.com/datasets/bimervos/online-automotive-sales-statistics-volkswagen>

**Mural Link:** <https://app.mural.co/invitation/mural/dvelesquad8353/1680730295576?sender=u2142599d86f2cb51dcf58730&key=e7d80a8e-8ccc-447e-8bc3-8c4f60bedb60>