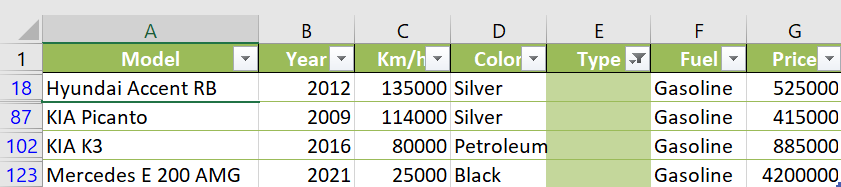
|  |  |  |
| --- | --- | --- |
| Name: TRAN NGOC NGAN IRN: 1832300436 GitHub Portfolio Link: https://github.com/ngantran11 | | Comment |
| **Criteria** | **Score** |  |
| Data Cleaning and Preparation |  |
| Data Analysis and Insight |  |
| GitHub Portfolio Quality |  |
| Total |  |

# 1.Data Overview

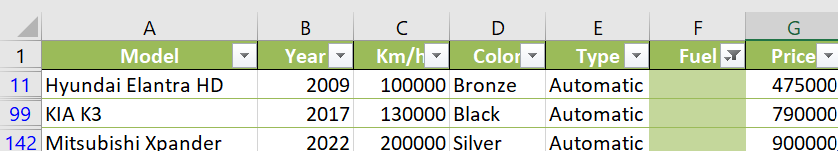
The data source is taken from car sales websites for the years. The dataset contains information on 187 types of cars that are used, with initial data of 202 rows. Key features include production year, kilometers driven (Km), color, transmission type, fuel type, and price. The data has been cleaned to remove duplicates and ensure consistency across entries. Most cars were made around 2016, with the oldest from 2009 and the newest from 2024, giving a good mix of older and newer models. On average, the cars had been driven about 119,000 kilometers, and the majority had mileage below 150,000 km, which suggests they are moderately used. In terms of price, the average was around 1.03 billion VND, with prices ranging from 220 million to 4.8 billion VND. Interestingly, half the cars were priced under 750 million VND, showing that the dataset includes both budget-friendly and higher-end vehicles.

# 2. Data Cleaning - Removed any duplicate rows. https://lh7-rt.googleusercontent.com/docsz/AD_4nXfmuVukfxxF7FQBH5ri90zR31M7xPl5jL2_aPEEP9RaAxPeCD2yOnSgCBaAQ5a4-iE_htVv_PrLA1jJdAJuy8GandhnGF0RXHOc9h2CEAhoNLhcBF7mUzBF_VRf1boO08QKfsrjgQ?key=ByI2FvkVuJ-snnWGeGFn_jS8 *Figure 1: duplicate rows.*

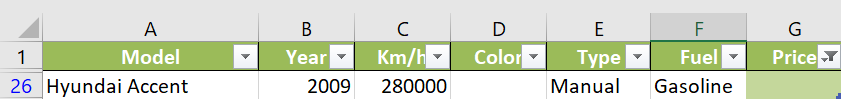
Removing duplicates is essential because repeated entries can distort analysis results.  
  
- Verified absence of missing values in critical fields.  
During data cleaning, we noticed that some records were missing values for Fuel, Type, Color, and Price. Since nearly all cars used gasoline, missing Fuel entries were filled in with "Gasoline." For the Type column, "Automatic"—the most frequent option—was used to replace missing values. Similarly, for Color, common values like "Black" or "Silver" were chosen to fill in the gaps. However, because Price is a critical number for our analysis and guessing it could lead to inaccurate results, any record missing a price was removed. This approach helped us keep as much reliable data as possible while ensuring the quality of our analysis.



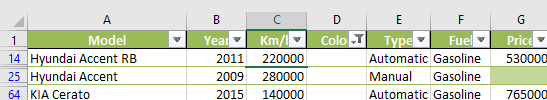
*Table 1: Missing values of type.*



*Table 2: Missing values of flue.*



*Table 3: Missing value of price.*

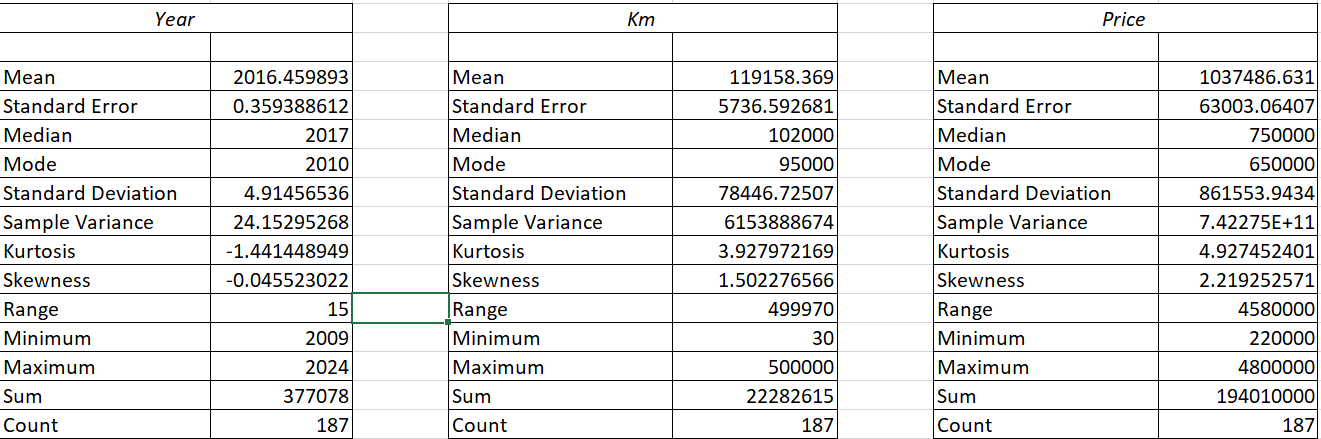


*Table 4: Missing of color.*

# Based on the 4 tables above, to find that

# The missing value in price is 3, value is 1, flue is 3, and type is 4.

# 3. Descriptive Statistics



*Table 5: The descriptive statistics.*

Based on the descriptive statistics from the cleaned dataset, we can draw several clear insights:

Most of the used Hyundai Elantra cars in the dataset were manufactured around the year 2016, which is the average year of production. However, there's a relatively wide spread, with a standard deviation of nearly 5 years, indicating that the dataset includes both older and newer vehicles. Interestingly, the most frequently occurring production year (mode) is 2010, suggesting a notable number of vehicles from that year.

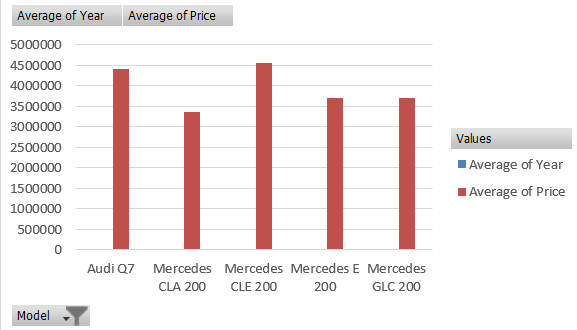
When it comes to mileage, the average distance traveled is approximately 119,000 kilometers. This suggests a moderate level of usage, typical for cars that are around 6 to 10 years old. However, the high standard deviation in mileage shows there is significant variation—some cars may have been used extensively, while others are relatively low-mileage.

As for pricing, the average price is over 1 billion VND. This reflects the general market value for this condition. The highest price reached over 4 billion VND, and the lowest is about 220 million VND.

# 4. Insights from the Data

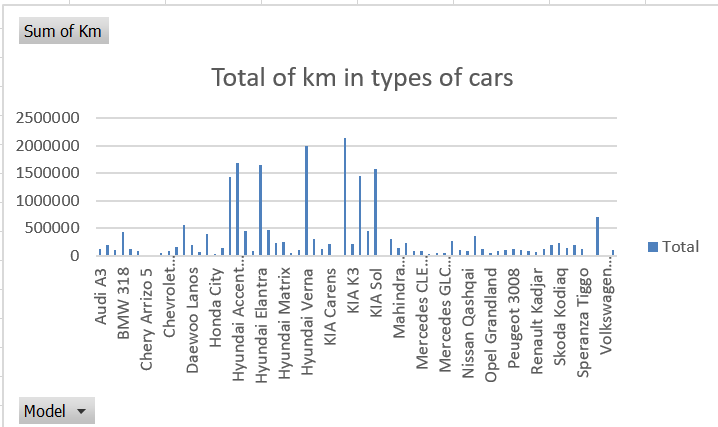
Insight 1: Car Price and Production Year

There is a moderate positive correlation between the production year and price. Newer cars are priced higher, demonstrating the effect of depreciation and consumer preference for recent models.

  
*Figure 2: The model of average cars price and production years.*

The table shows the average year and price for five luxury car models. Among them, the Mercedes CLE 200 stands out with the highest average price at 4.55 billion VND, while the Mercedes CLA 200, despite being one of the newest models (2024), has the lowest average price at 3.35 billion VND. The Audi Q7 is slightly older (over 2021) but still holds a high average price of 4.4 billion VND. Meanwhile, the Mercedes E 200 and GLC 200 have average years of 2021 and 2020, with similar price points of around 3.7 billion VND. On average, across all models, the manufacturing year is around 2022, and the price sits at just over 4 billion VND, showing a range of newer vehicles at varying price levels depending on brand and model.

Insight 2: Car Price and km



*Figure 3: Total of km in types of cars.*

This indicates that vehicles with higher mileage tend to be priced lower, which aligns with general market expectations. The dataset gives a clear picture of how extensively different car models have been used, with a total of over 22 million kilometers recorded across 67 models.