**Data Analysis Project**

**on**

**German used cars**

**Name: B Ganga Vignesh**

**Date: 23-12-2023**

**Introduction:**

“eBay” is an e-commerce site where used products and retail products can be found. Analysis was performed on used “German cars” by capturing the data points from “eBay” site.

**Overview:**

The dataset has 371528 entries and 20 columns.

1. dateCrawled: Date and time when the car was launched.
2. Name: Name of the car.
3. Seller: Here there are 2 categories, if the vehicle is a private vehicle or commercial.
4. offerType: If it is to be purchased through “Auction” or through “petition”.
5. Price: Cost of the car.
6. abTest: Method of testing the cars. There are two types. The Control and Test.
7. vehicleType: vehicles are categorised based on structure and its design. There are 8 types.
8. yearOfRegistration: The year on which car was purchased and first registered.
9. Gearbox: The model of shifting the gears. If Manual shifting or Automatic shifting.
10. powerPS: Power of car engine.
11. Model: Model of the car.
12. Kilometer: Kilometers driven.
13. monthOfRegistration: On which month of the year the car was registered.
14. FuelType: Type of fuel the car runs.
15. Brand: The Company of car manufactured.
16. notRepairedDamages: If the car damage is repaired or not.
17. dateCreated: The date on which the post was listed.
18. Nrofpictures: Number of picture posted in the add.
19. Postalcode: The area code of the car.
20. Lastseen: Date and Time the car add was last seen.

**Analysis 1:**

1. Perform general Data analysis.

* Dropped duplicates.
* Treated missing values.
* Performed typecasting
* Treated structural errors.
* Treated Outliers.

## 2. Can you tell me the Distribution of Vehicles based on Year of Registration with the help of a plot.

## 

## Summary:

## The Distribution is left skewed which means most of points are on the right side of the plot. Hence Maximum cars are registered between 2000 and 2020.

## 3. Create a plot based on the Variation of the price range by the vehicle type.

## 

## Summary:

## Price Range differs based on the vehicle type with “SUV” type having the highest and “kleinwagen” having the lowest.

## 

## 4. Find out Total count of vehicles by type available on ebay for sale.As well as create a visualization for the client.

## 

## Summary:

Count of cars available based on type of vehicle. Highest number of vehicles available belong to “limousine” type and least belong to “andere”.

## 5. Is there any relationship between dollar\_price and kilometer? (Explain with appropriate analysis)

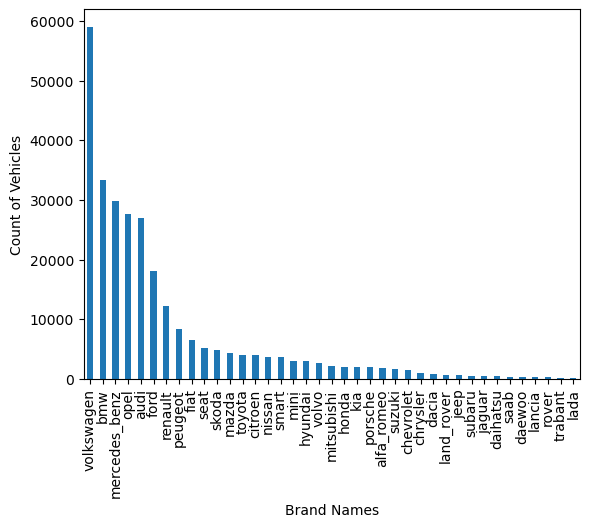
## 

**Summary:**

There is not relation between Dollar\_price and Kilometers. Which means Price of the car is not calculated or quoted based Kilometers driven by car.

**Analysis 2:**

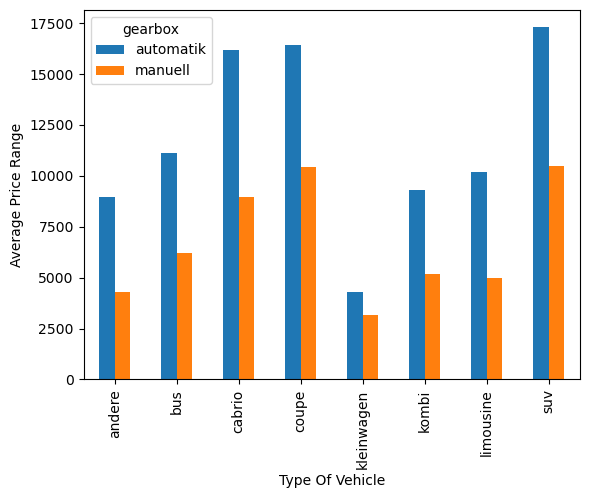
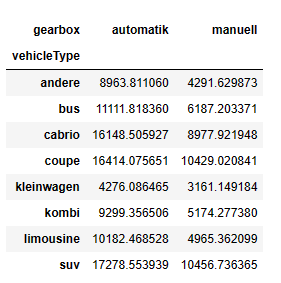
## 1. Can you tell me No of Vehicles by Brand Available on ebay for sale with the help of visualization.



**Summary:**

These are different car manufacturing companies and count of cars available on eBay of those companies. Most cars post belongs to “Volkswagen” and least belong to “lada”.

## 2. What is the Average price for vehicles based on the type of vehicle as well as on the type of gearbox. Explain me with both numerical and visualization analysis.



**Summary:**

Type of vehicle further segregated by gear box and their average price range. “SUV” has highest average price in “Automatic” segment and “Coupe” type has highest average price in “Manual” Segment. “Kleinwagen” has lowest average price in both segments.

## 3. What is the marginal probability of private seller.

Ans) The Marginal probability of Privat seller is 0.999993.

Analysis 3:

## 1. The memory usage of the data is around 6.1 mb.How can we reduce the memory usage of the data set?

## Ans) By removing the unwanted columns, changing the Datatypes and its bytes.

## 2. What is the Average price of vehicle by fuel type and gearbox type. Give a plot.

## 

## Summary:

## Type of fuel further segregated by gear box and their average price range. “Hybrid” fuel has highest average price in both segments and “andere” fuel has lowest average price in “Automatic” segment with “lpg” fuel in “Manual” segment.

## 3. What is the Average power of a vehicle by vehicle type and gearbox type. Give a plot.

## 

## Summary:

Type of vehicle further segregated by gear box and their average power. “Coupe” has highest power in both segments and “Kleinwagen” has lowest power in both segments.

## 4. What is the Average price of a vehicle by brand as well as vehicle type. Use heatmap to explain this.

