Data-Analysis Project-1

(final conclusion(Summary) for each of the analysis)

Submitted to Innomatics Research Labs

Submitted by Noah Shanthi Raj

Analysis 1:

1. Perform general Data analysis

* In the initial general analysis, I examined the dataset and identified some key findings:
* There are a total of 7 types of vehicles, with "limousine" being the most prevalent.
* Out of these, only 3 vehicle types are non-private, while the rest are private.
* The average price of vehicles in the dataset is approximately $17,295.1.
* The highest-priced vehicle in the dataset is "Kaufe\_AUTOS\_Zum\_basteln."

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

* To visualize the distribution of vehicles based on the year of registration, I created a plot. The plot reveals a spike in the number of vehicle registrations around the year 1980, with the highest sales volume being around 16,000 vehicles per year.

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

* A plot was generated to showcase the variation in price range across different vehicle types. It's important to note that most vehicle types exhibit outliers in their price range, with the lowest number of outliers observed in the "suv" category.

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

* The total count of vehicles available for sale on eBay, grouped by type, was determined. Here are the counts:
* Limousine: 95,894
* Kleinwagen: 80,023
* Kombi: 67,564
* Not Known: 37,869
* Bus: 30,201
* Cabrio: 22,898
* Coupe: 19,015
* SUV: 14,707
* Andere: 3,357

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

* A covariance analysis was performed, and a scatter plot was generated to investigate the relationship between the dollar price and the number of kilometers on the vehicles. The analysis revealed that there is no significant relationship between the two variables.

Analysis 2:

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

* Visualizing the number of vehicles by brand available for sale on eBay, I found that there are 40 vechicals by brands with "volkswagen" leads with 79,640 listings, followed by "bmw" with 40,274 listings.

1. 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
2. The average price of vehicles was analyzed based on both vehicle type and gearbox type. Suv, cabrio, bus is expensive vehicle types with less regard to gearbox. Average starting price for the most of vehicles is around $2000. kombi, limousine have observable price variation with regards to gearbox
3. What is the marginal probability of private seller

* The analysis indicates that nearly all sellers are private, with the marginal probability of private sellers being close to 1. This suggests that most listings on eBay are posted by private individuals.

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?

* The dataset's memory usage was known to be around 6.1 MB. To reduce memory usage, it is recommended to change data types from "int64" to "int32" for specific columns.

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

* The average price of vehicles was analyzed with respect to both fuel type and gearbox type. cng fuel type and automatik gearbox is having the highest average price and diseal and elektro fuel type with automatic gearbox are some in top list. elektro has major average price variation with regards to gearbox.   
  Cng and diesel has not much difference in average price in regards to gearbox.

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

* A plot was generated to find the average power of vehicles by their type and gearbox. coupe vehicle type with automatic gearbox have the highest average price and most vehicles have power above 100. least variation with gearbox is seen in andere and bus in regards to average power.

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

* A heatmap was used to visualize the average price of vehicles by brand and vehicle type. It highlights that Kia brand of cabrio vehicle type have average price of 10000, followed by Subaru brand of cabrio vehicle type. Highest average price is seen in cabrio and suv vehicle type.

Most of the analysis with plot or other means was done after cleaning the data as there were outliers, especially in price column. Replaced the values using IQR method of finding outliers.