**Data Analysis project-1**

**Analysis-1:-**

**Perform data analysis report:-**



Dataset overview:-

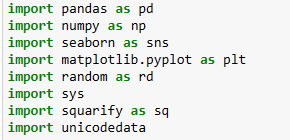
The above dataset is an autos dataset it contains 371,528 rows and 20 columns. It appears to be related to vehicles listing, with information about various outs attributes and transactions. In dataset we do some amputations, data cleaning and data modeling to get important information from it which is help to understand to solve the business problem. Here we explain about the columns of the dataset:-

Column Descriptions:

1. **DateCrawled:** Timestamp when the data was crawled.
2. **Name:** Name of the car.
3. **Seller:** Indicates if the seller is private or commercial.
4. **Offer Type:** Type of the offer.
5. **Price:** Price of the car.
6. **abtest:** A/B test flag.
7. **vehicleType:** Type of the vehicle (e.g., coupe, SUV).
8. **yearOfRegistration:** Year the car was first registered.
9. **gearbox:** Type of transmission (e.g., automatic, manual).
10. **powerPS:** Power of the car in PS.
11. **model:** Model of the car.
12. **kilometer:** Mileage of the car in kilometers.
13. **monthOfRegistration:** Month of registration.
14. **fuelType:** Type of fuel used (e.g., petrol, diesel).
15. **brand:** Brand of the car.
16. **notRepairedDamage:** Indicates if there's unrepaired damage to the car.
17. **dateCreated:** Timestamp when the listing was created.
18. **nrOfPictures:** Number of pictures in the ad
19. **postalCode:** Postal code of the location.
20. **lastSeen:** Timestamp when the listing was last seen.

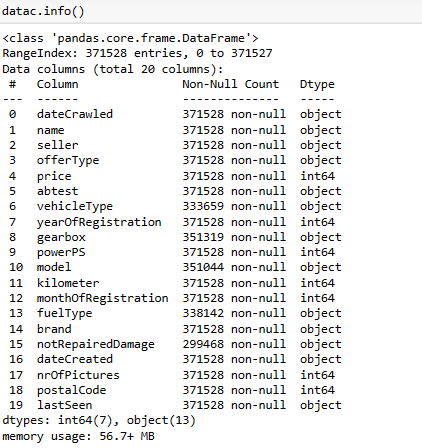
Hear the dataset contains information about used car listing. it includes important columns . it may require cleaning and some manipulations. The next steps are data cleaning in which we address missing values and outliers. Then do EDA means Explore price distributions and key attributes. The last step is do some visualizations for batter understand.

1. **Perform general data analysis:-**

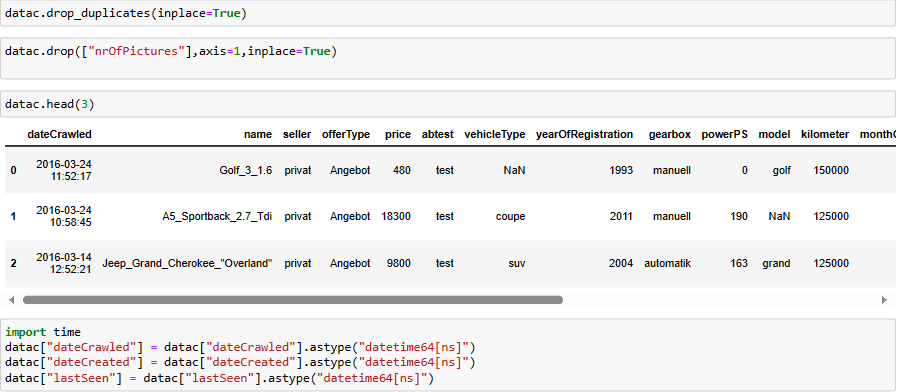
****

We use some libraries for doing analysis of above data.

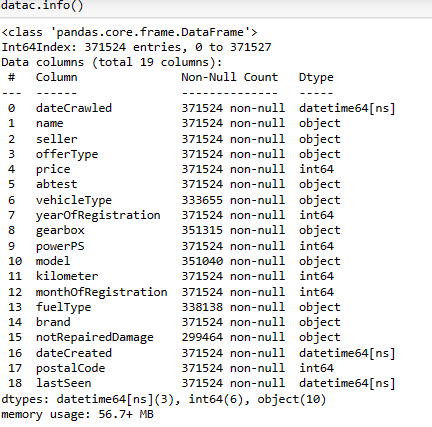
**Data Cleaning:-**



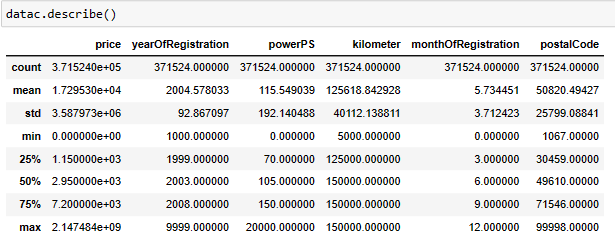
First we fill the all nan values and we’ll change the data type of date crawled, date created and last seen columns to date time data type using this code.



Then we got the info of data set like this.



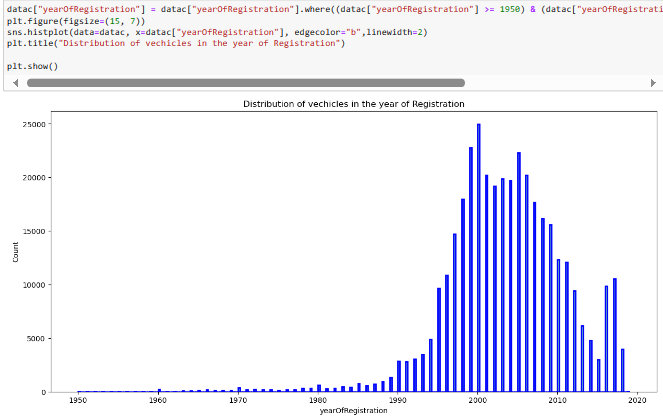
By using Descriptive statistics, we got a summary of our entire data.



From that in the column price we can tell the max price of vehicle which is for sale is 2.14e+09, minimum are 0.0e+00 and standard deviation is 3.5879e+06.

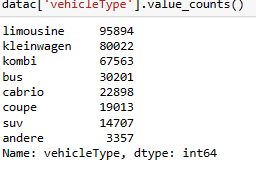
From that in the column powerPS max power is 20K, minimum power is 70 and standard deviation is 192.1395.

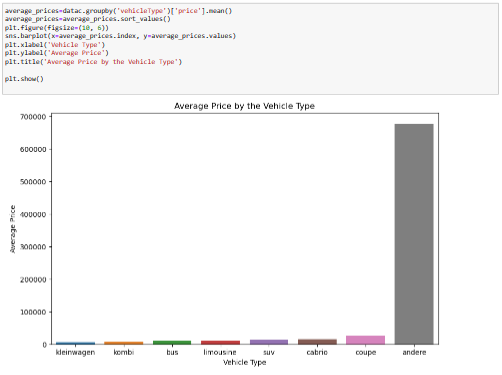
**2.) Can you tell me the Distribution of Vehicles based on Year of Registration with the help of a plot.**

****

Majority of vehicles were registered between 2000-2010. Data cleaned for valid registration years(1950-2023).

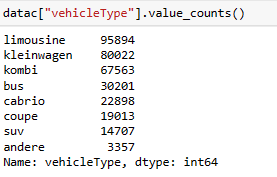
**3. Create a plot based on the Variation of the price range by the vehicle type.**

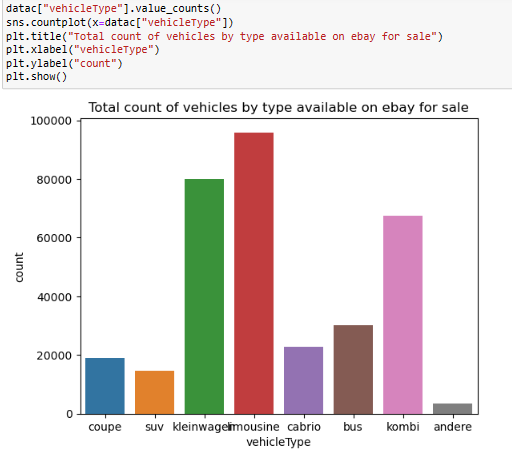
****

****

Compact cars have the lowest average prices, while exotic cars have the highest. Vahicle type significantly affects average car prices.

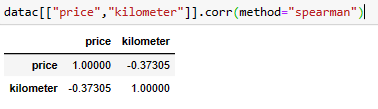
**4. Find out Total count of vehicles by type available on ebay for sale.As well as create visualization for the client.**

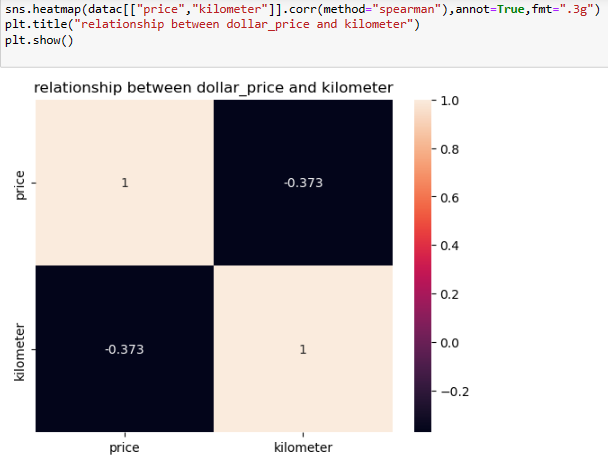
****

****

Most vehicles listed for sale are “limousines” and “small cars”, followed by “ station wagons” and “ buses” limousines dominate the listing.

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

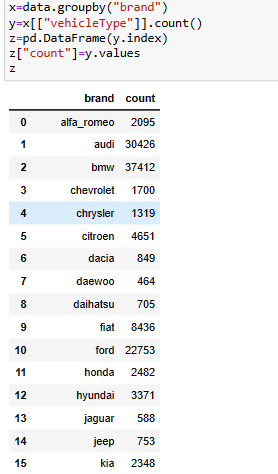
****

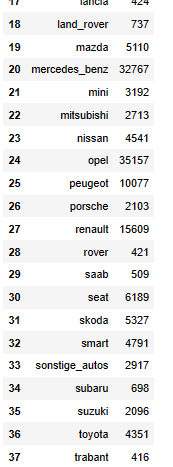
****

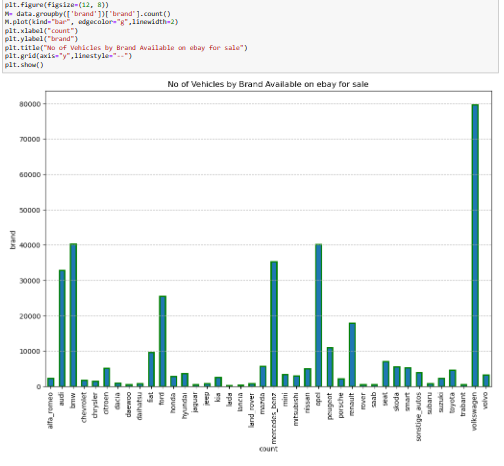
There’s a negative correlation between price and kilometer; as the mileage increases, the price tends to decrease.

**Analysis-2:-**

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

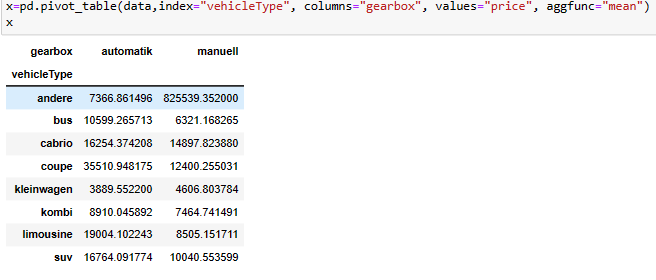
****

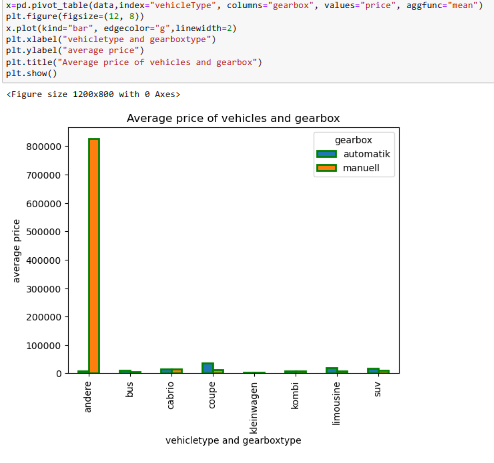
****

****

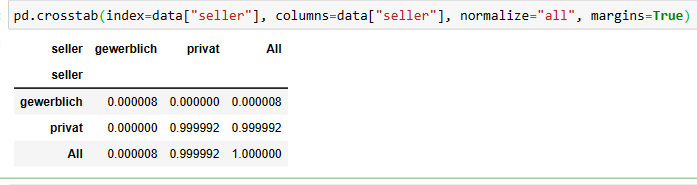
Volkswagen has the most vehicles listed on eBay. Common brands like Volkswagen, BMW and dominate the listing by the bar chart count for each brand.

1. **What is the Average price for vehicles based on the type of vehicle as well as on the type of gearbox.**



 Automatic transmission vehicles tend to have higher prices across different vehicle types. This bar chart illustrates the average prices for vehicle types and gearbox types.

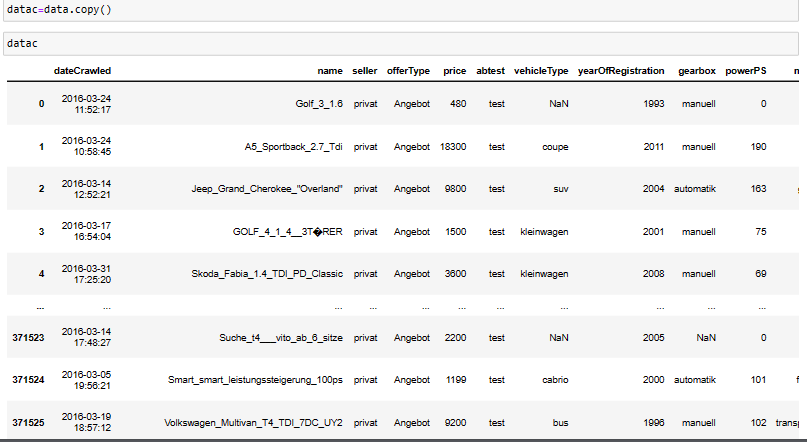
1. **What is the marginal probability of private seller.**

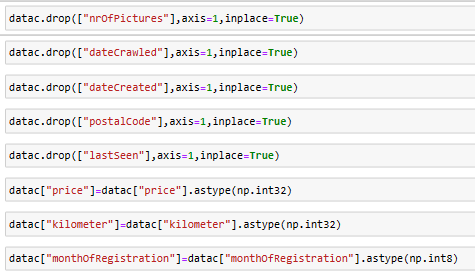
****

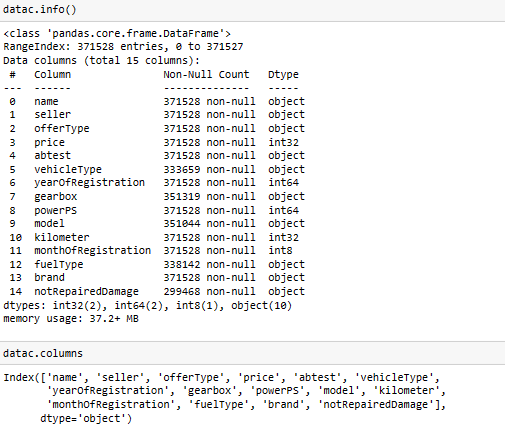
The “seller” column perfectly correlates with itself. This table illustrates the distribution of sellers in the dataset.

**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?**

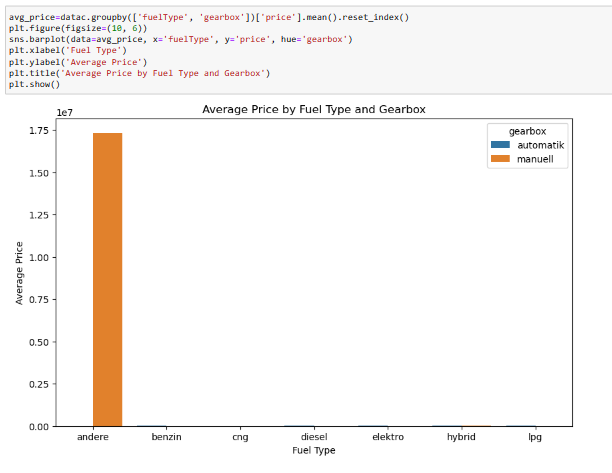
****

****

****

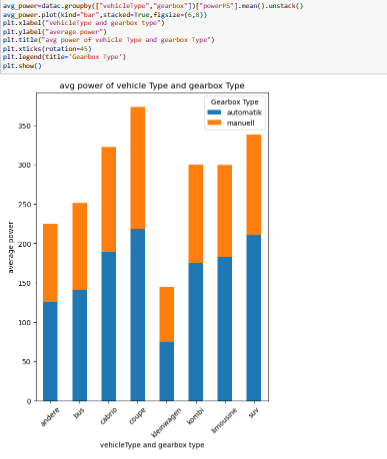
Data cleaning was performed, removing unnecessary columns the converting data types to save memory. Now the dataset contains 15 columns with various data types.

**2.What is the average price of vehicle by fuel type and gearbox type And give a plot.**

****

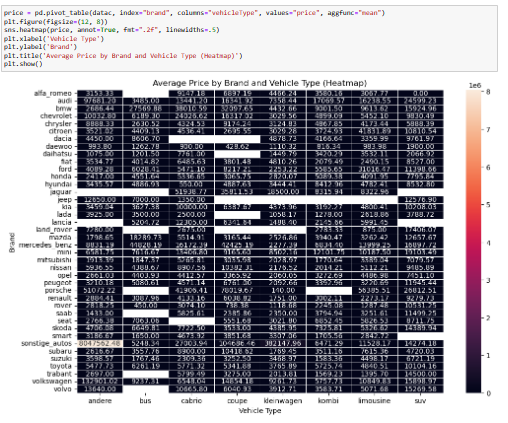
Diesel vehicles with automatic gearboxes tend to have the highest average prices, while manual gasoline cars have lower average prices. By the bar chart displays average prices based on fuel type and gearbox type.

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

****

Automatic transmission vehicles generally have higher average power then manual ones across different vehicle types. The bar chart illustrates average power by vehicle type and gearbox type.

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

****

Luxury brands like Porsche and land Rover have the highest average prices across various vehicle types. By the heatmap displays average prices by brand and vehicle type.

**Takeaways:-**

In above dataset we know the Volkswagen is the most listed brand and Manual transmission compact cars more affordable. Diesel automatic car tend to be expensive, there are some luxury brands like Porsche command high prices. The dataset provides valuable insights into the used car market, helping business understand pricing trends and factors.