

# BUSINESS INTELLIGENCE PROJECT

Name: Shaheer Ahmed Farooqui

Erp: 19722

**Dataset:** OLX used car dataset

**Dataset Link:** <https://opendata.com.pk/dataset/pakistan-used-cars>

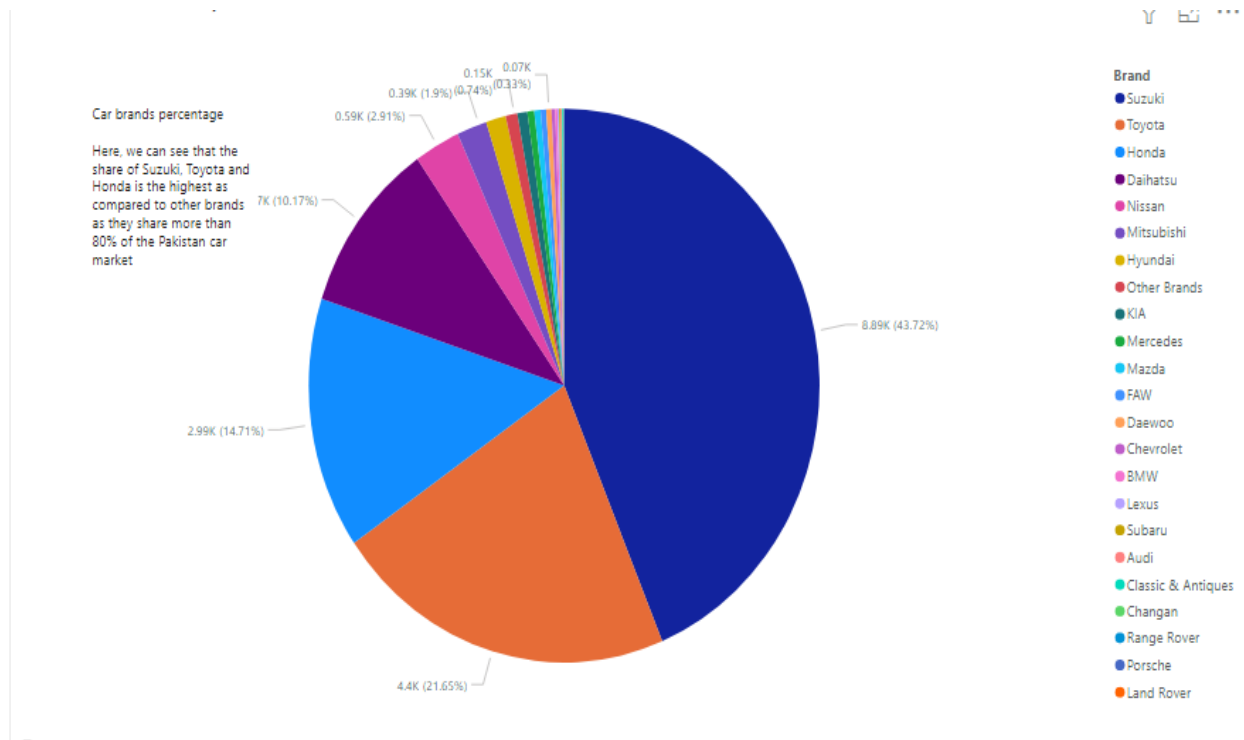
## About the dataset:

This dataset is about the data of old and new cars of Pakistan that were posted on the website. The data has all the details of the car that are important in order to sell a car like the year and model of the car, KMs driven, Brand name, Registration city, price etc.

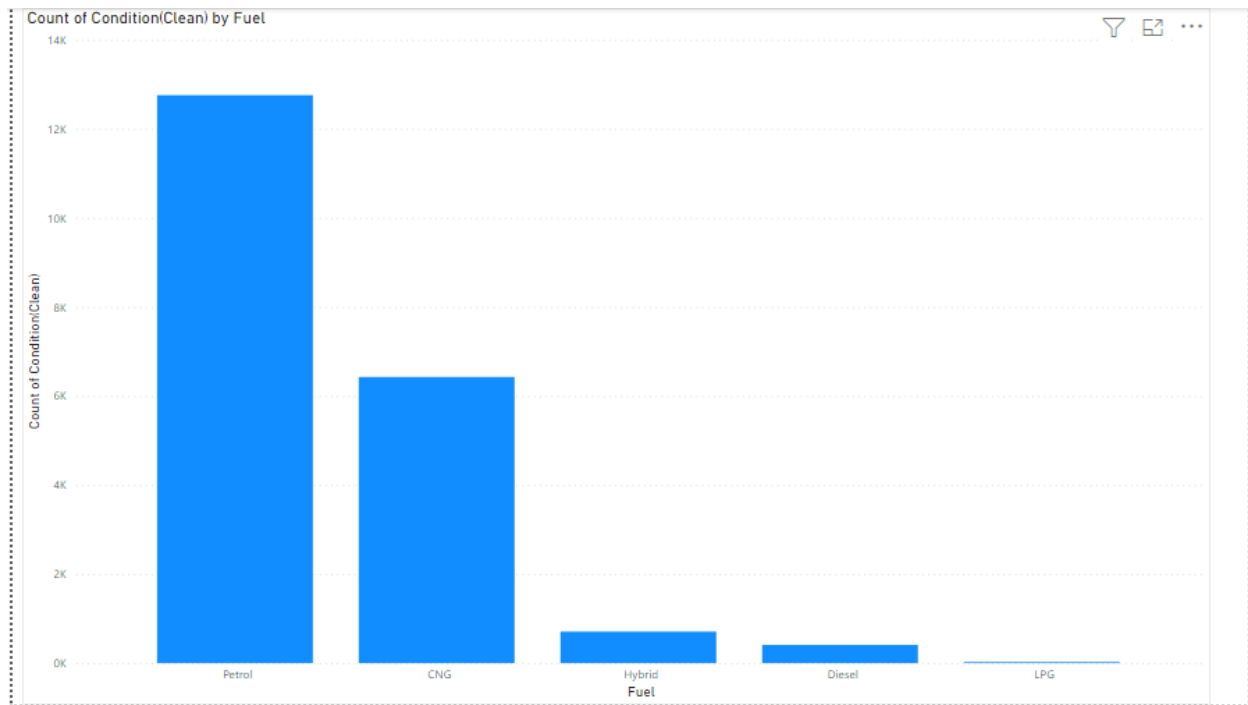
Knowledge of buying and selling of a car is important in order to understand this dataset completely so that the next step of cleaning the data can be started. This is important as a strong knowledge of cars and their prices will help to identify the wrong values mentioned in the dataset.

## DATA WRANGLING STEPS:

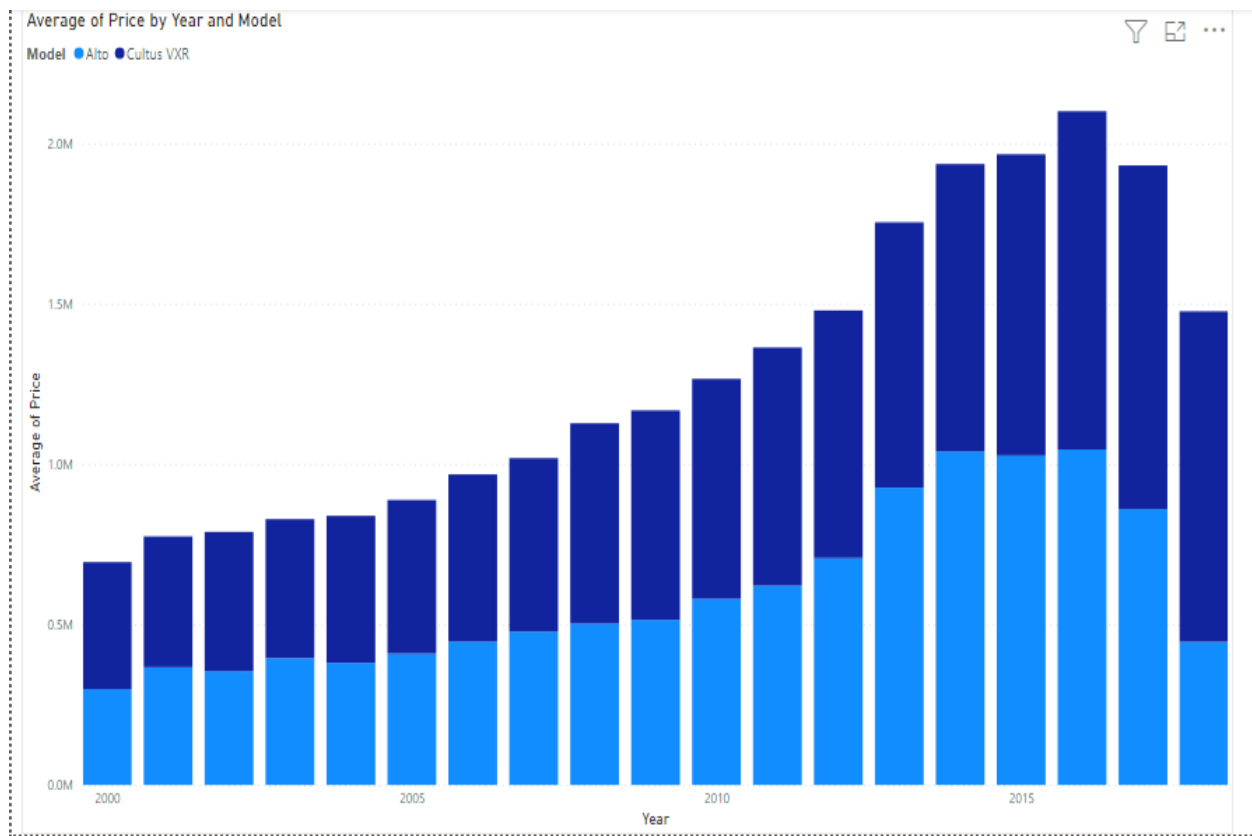
- The null values were removed completely by using the remove Null feature.
- There were more than 1000 cars that had the value of 1 km in the KMs Driven column so the value of 1 was replaced with 0
- Moreover, there were many rows that were showing a car as Used but the value of KMs driven was zero and same for the other way as a car was showing as New but the KMs driven were more than 1000 kms. So, in order to correct it, a range of 0 to 200 kms is labelled as a New Car and KMs driven of more than 200kms is labelled as Used.
- Moreover, the price of cars above 1,00,00,000 were checked carefully which was around 37 rows and the majority of them had the wrong price mentioned in it as For example, a Suzuki Mehran of 2014 had a price of 1,20,00,000 which was obviously incorrect so in order to correct it a zero at the end of the price was removed which made the value of the car a normal one.



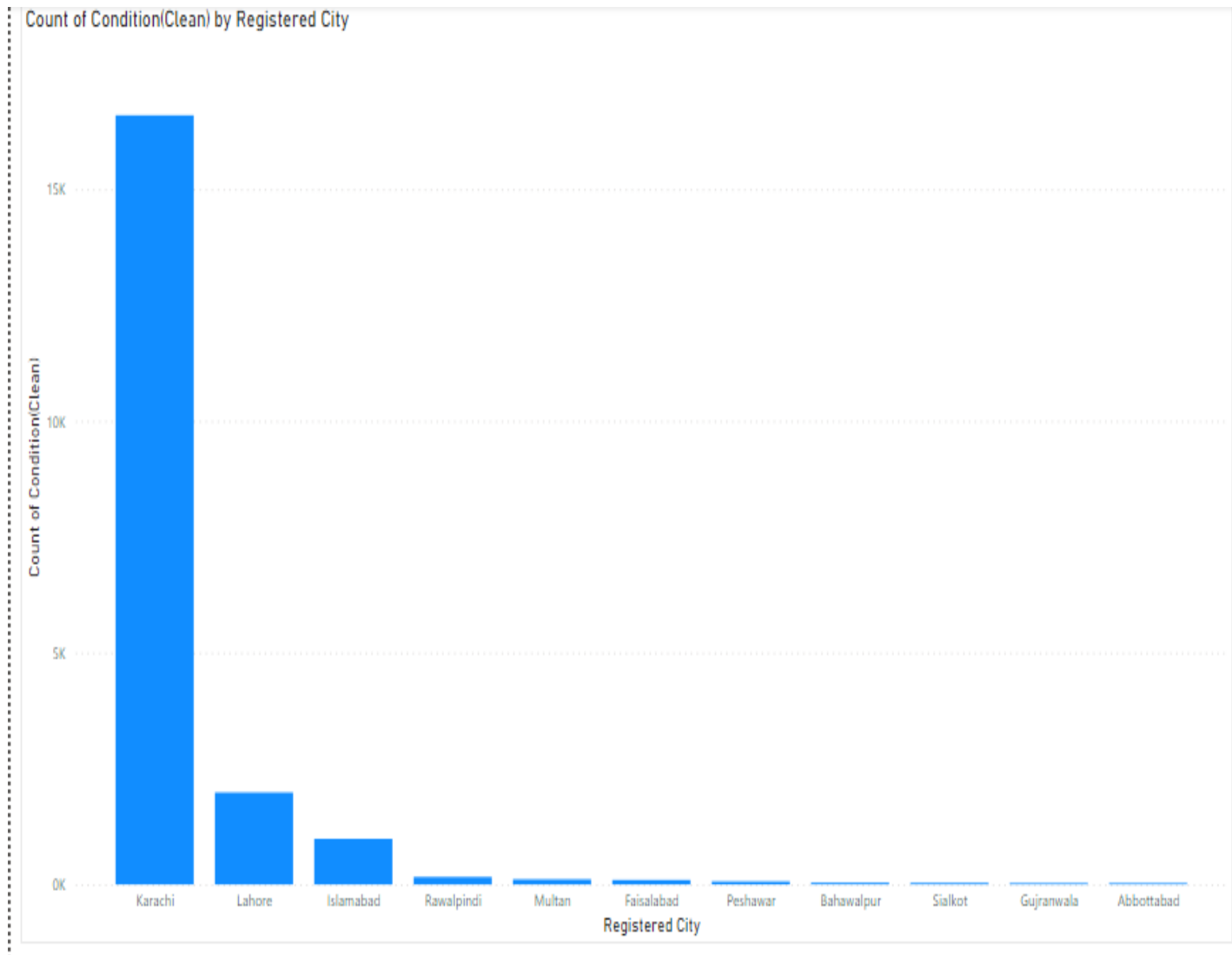
This chart is displaying the market of the brands that were posted on OLX. Here we can clearly see that the brand of Suzuki is the highest brand in the number of vehicles posted for sale as it has around more than 40% of the overall market. After Suzuki, Toyota is present on second number and Honda on third number in the market.



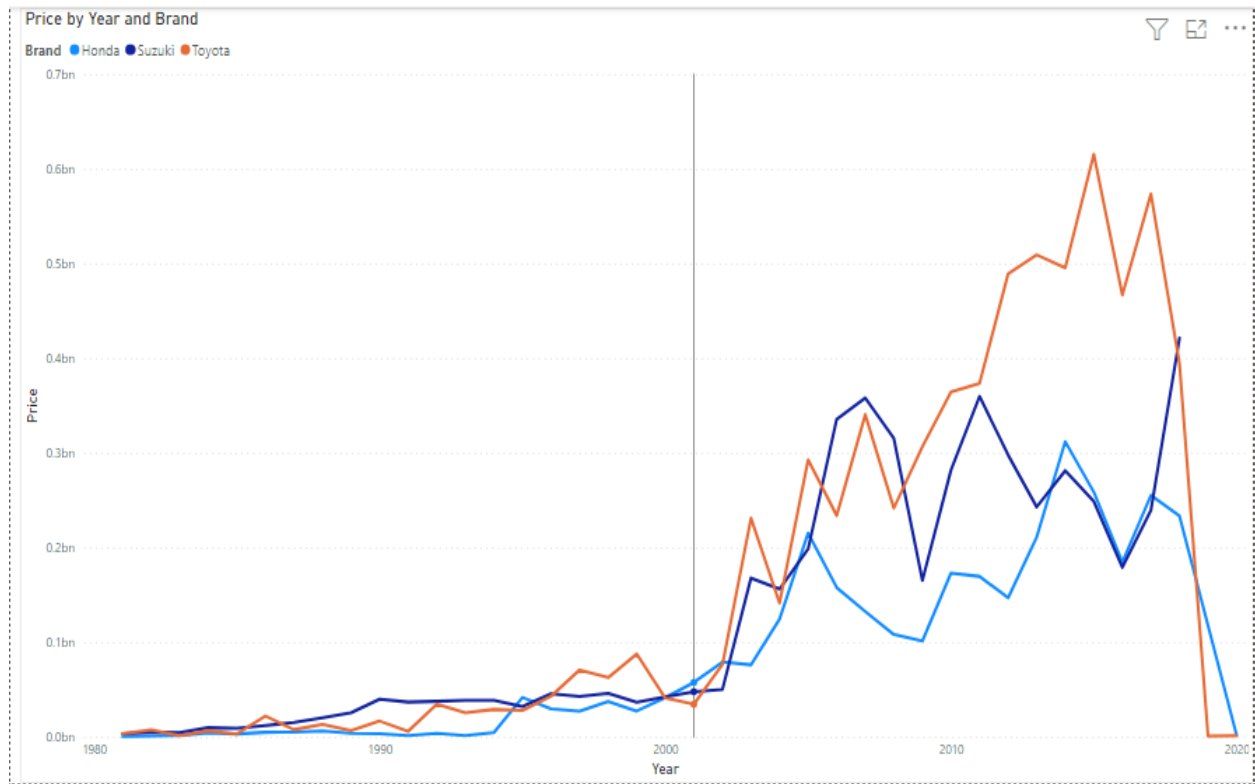
The above chart is showing the type of fuel used by cars that were posted for sale. The data shows that the majority of the cars had Petrol as its fuel and after Petrol, CNG was the second most appeared fuel type in the cars.



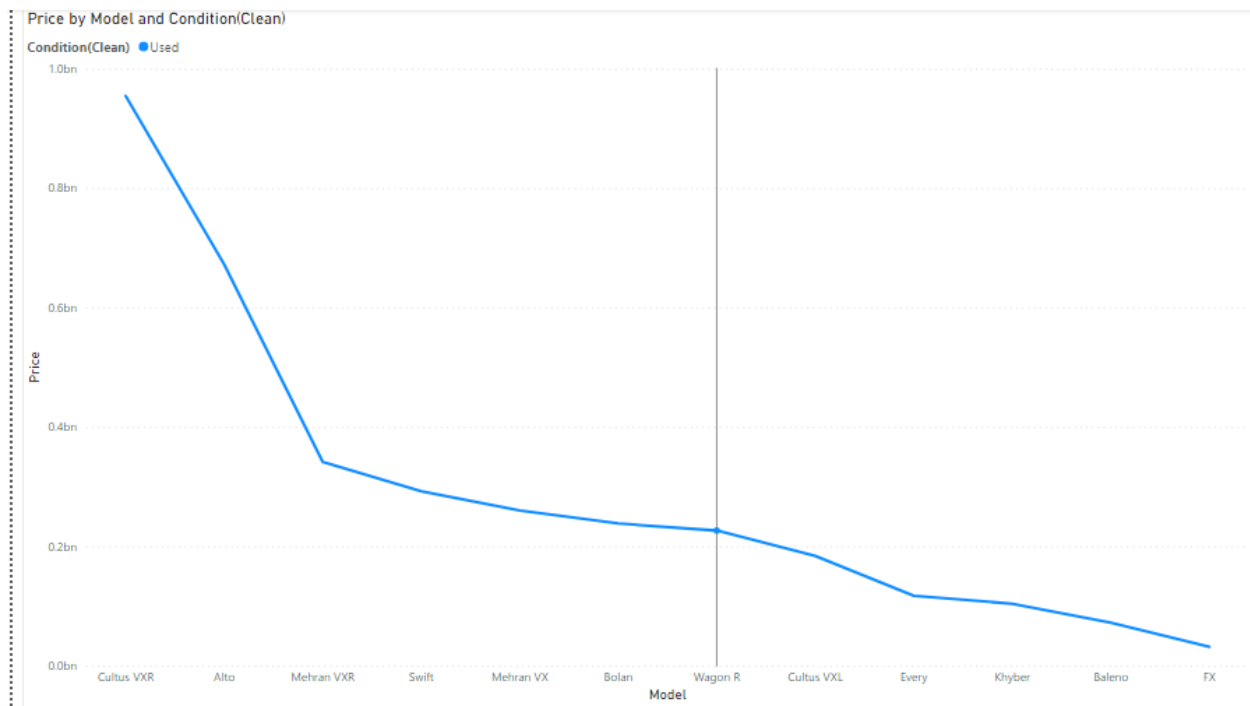
This chart is showing the top 2 cars of Suzuki brand and the 2 cars that are used here are the top 2 most sold cars of Suzuki which are Alto and Cultus. The Average price of these cars are placed on y axis and we can clearly observe that the average price of both the cars are increasing as we move from year 2000 to 2018.



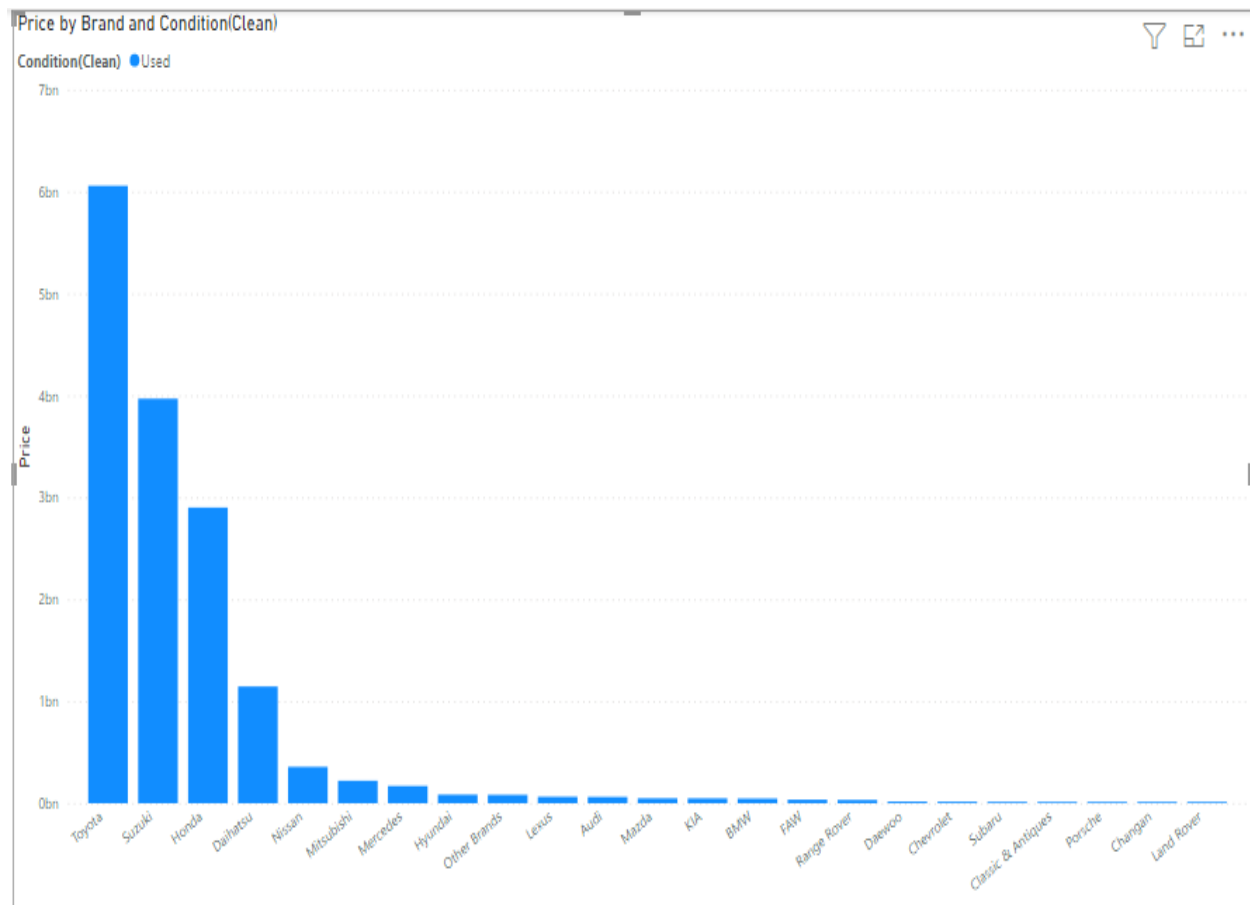
This chart is showing the number of vehicles that were registered in different cities and here we can clearly observe that Karachi city has the highest number of registered vehicles because of its population.



This graph is showing the trend of sales of brands of Suzuki, Toyota and Honda. Here we can observe that despite the fact that Suzuki has the highest number of cars available in market, Toyota has generated the highest sales as compared to Suzuki which means that the price of Toyota is higher as compared to Suzuki and that can be observed with the trend of the graph.



This chart is showing the revenue generated of Suzuki by each model of it and here we have selected the top 12 models of Suzuki.



This chart is showing the total price of used cars by each brand. Here we can see that the Toyota generate the highest sales in used cars and Suzuki is after Toyota in the total sales generated aspect. We can conclude that vehicles of Suzuki are of less price as compared to Toyota and that's why Toyota is above Suzuki in the total Sales revenue factor.