INNOMATICS
RESEARCH LABS

About me

Name : Balaji Mummidi
Qualification : B.Tech (Bachelor of Technology), EEE
Work Experience : I am a fresher

INNOMATICS RESEARCH LABS

Web Scraping & EDA on Cars24 website

INNOMATICS
RESEARCH LABS

Problem Statement:

Analyze & visualize data of used cars from Cars24 website

INNOMATICS RESEARCH LABS

Explanation about Use case title and domain

- The title of the project comes under the domain of “Automobile Industry”.
- The main aim of this project is to analyze the data about the used cars based on different features and to give a detailed analysis by visualizing them.
- This analysis will be helpful for the customers who are planning to buy a used car based on different features.

Website for Scraping Data

- The website used in this project to scrape the data is Cars24
- Used cars data extracted from following 5 cities:

1) Hyderabad
2) Delhi
3) Mumbai
4) Bangalore
5) Chennai

Tools & Techniques

Tools

- Selenium – To manipulate the HTML code
- Beautiful Soup – HTML Parser
- Pandas, NumPy – To analyze the data
- Matplotlib, Seaborn – To visualize the data

Techniques

- Regular Expressions to extract required data from a string

Important data for scraping

1) Brand name - To analyze the data based on brands.
2) Model name - To analyze based on different model.
3) Price of the car - To analyse based on price
4) Model Year - To analyse based on model year
5) Location - To analyse car details based on location
6) Fuel Type - Petrol/Diesel/LPG/CNG/Electric
7) Driven (Kms) - Analyse based on distance travelled by car
8) Gear Type - Automatic/Manual car
9) Ownership - Either the car is 1st hand or 2nd and soon
10) EMI (Monthly) - Describes the EMI price for different brands & models

Scraping Methodology

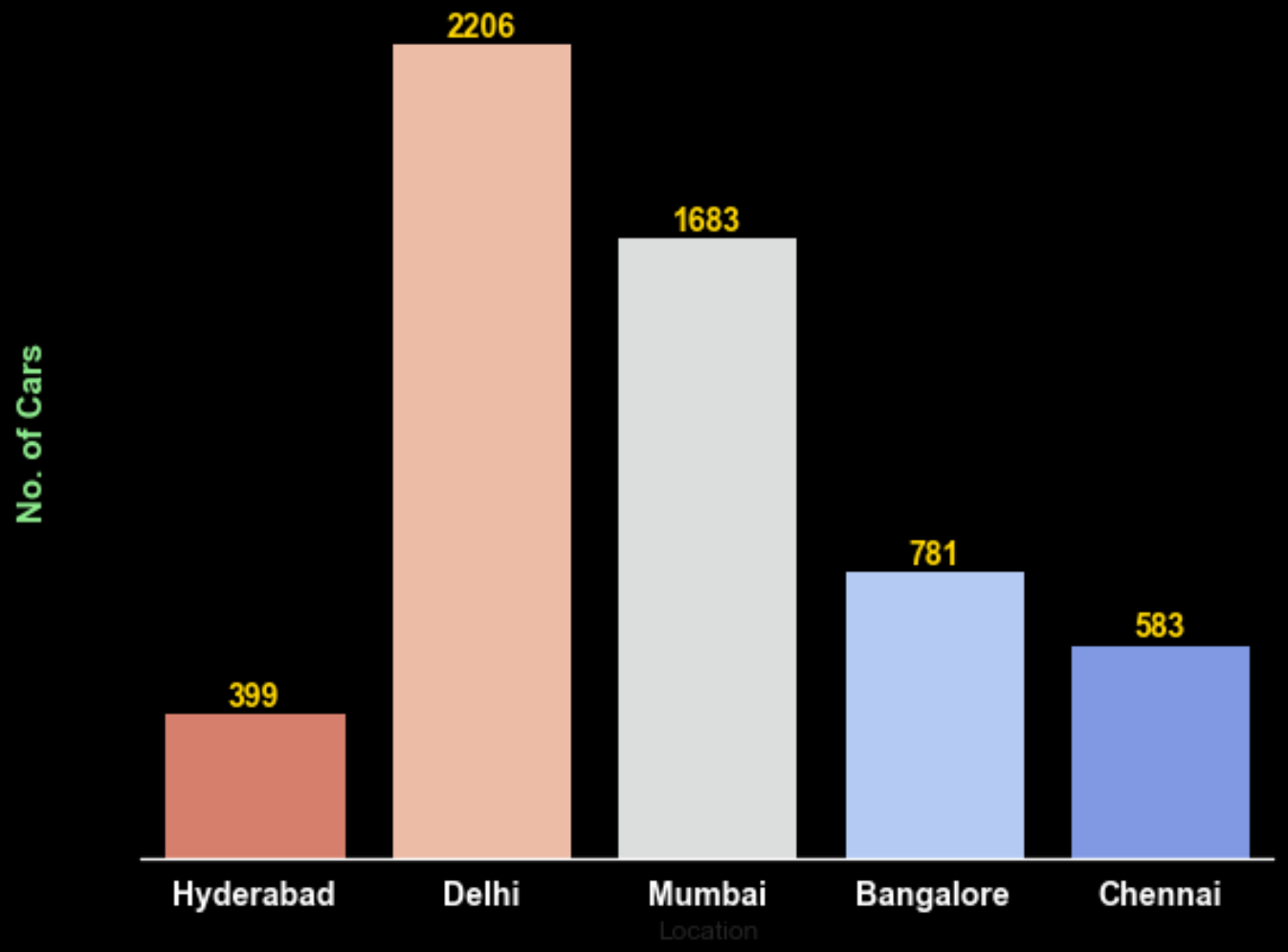
- Collecting the URL's of used cars in different cities. (Hyderabad, Delhi, Mumbai, Bangalore, Chennai)
- Since the data loads upon scrolling the website, Selenium web driver is used to scroll till bottom of the page and to store the html code after manipulating the page source.
- After saving the html file, load the file and parse the file using Beautiful Soup.
- After scraping the data, saving the data in a csv file.

Data Cleaning

- Cleaning poorly formatted columns from the extracted raw data.
- Skipping rows with Nan values.
- Removing duplicated rows.
- Changing the data types for necessary columns.
- Filtering the Data Frame and storing it to a csv file.

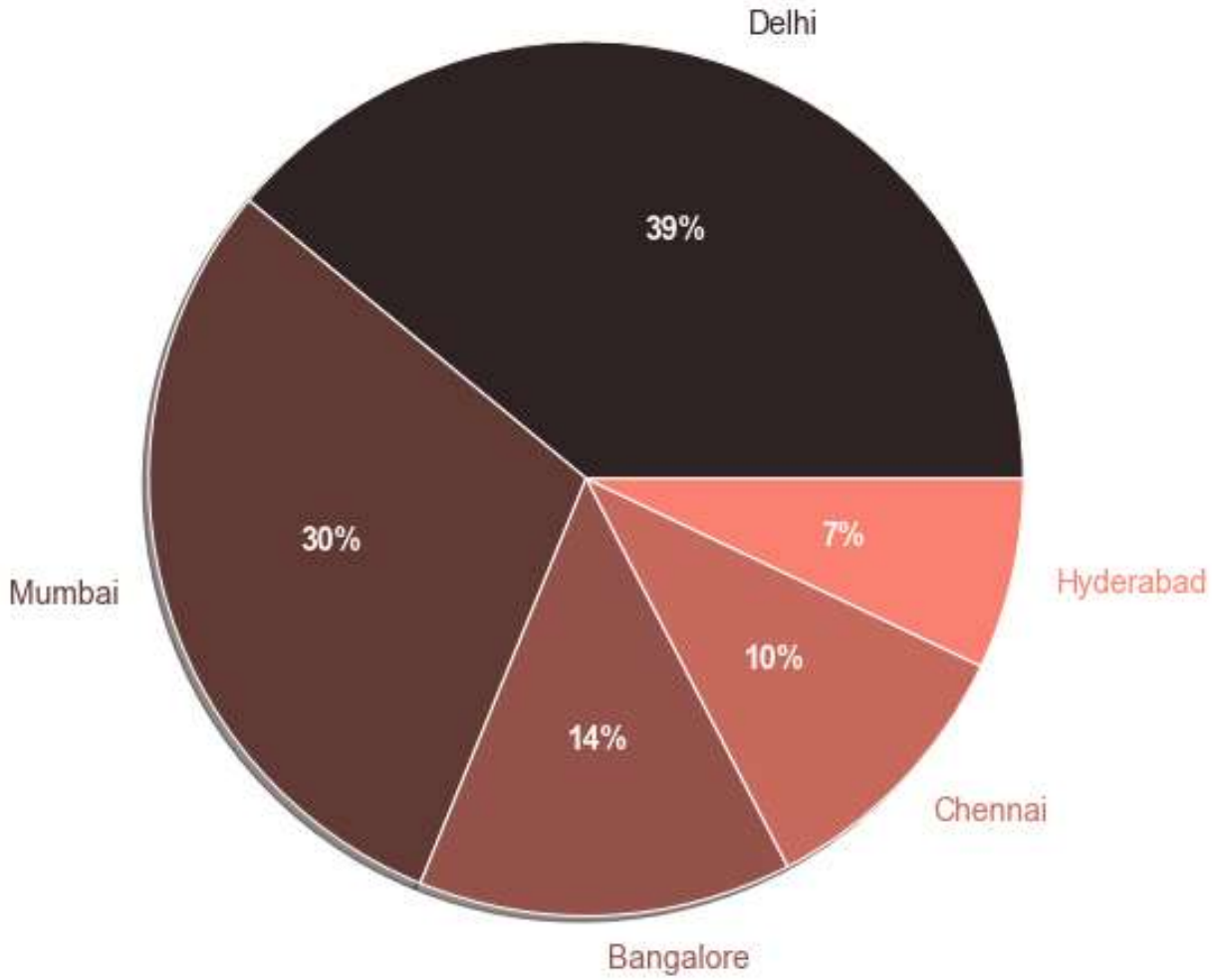
Data Visualization

1) Total number of cars in different cities



- The total number of cars is 5562
- Delhi has the highest number of cars with a total of 2206 cars.
- Hyderabad has the least number of cars when compared to all other cities with a total of 399 cars.

2) Percentage of cars available for sale in different cities



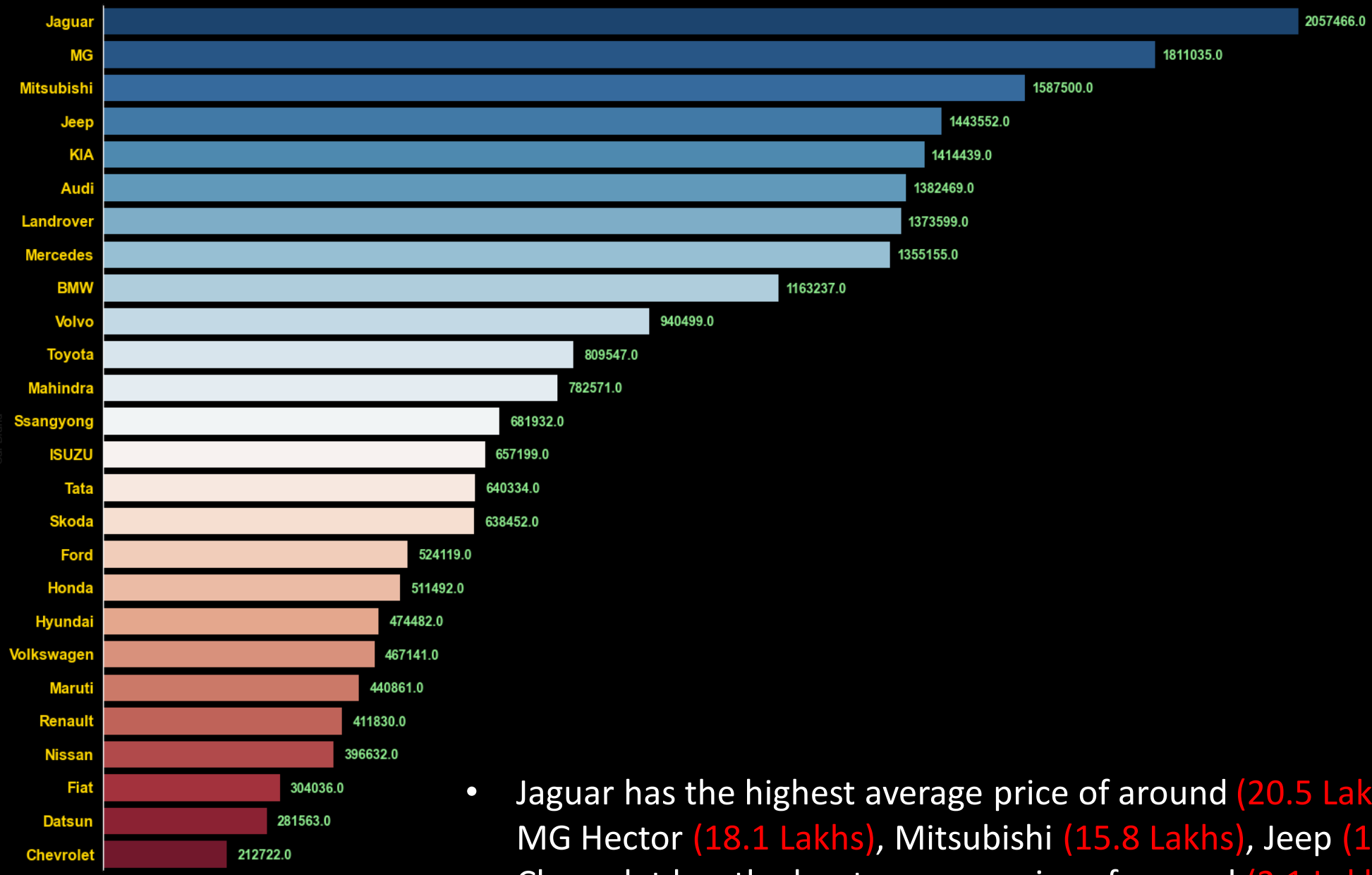
- Among 5562 cars 39% of the used cars are in Delhi is the highest record.
- 30% of used cars are in Mumbai, followed by Bangalore (14%), Chennai (10%) Hyderabad (7%).

INNOMATICS RESEARCH LABS

3) Total number of different branded cars in all cities

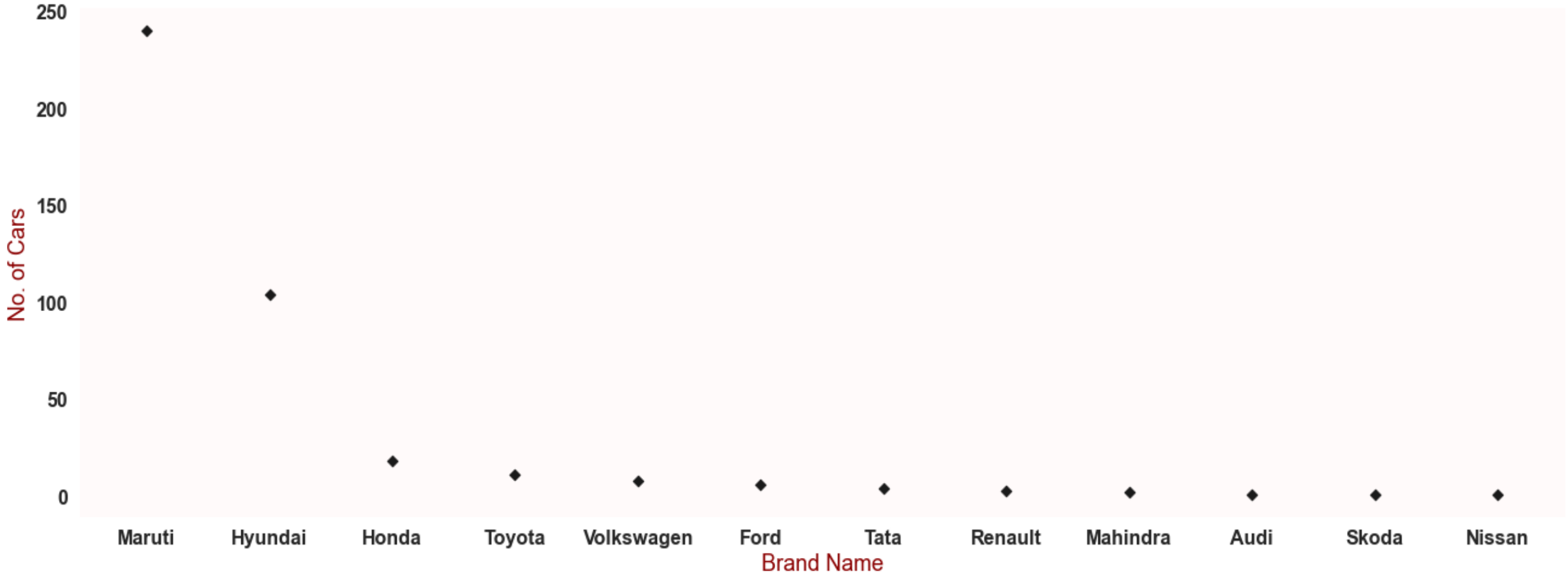
- Maruti Brand cars are mostly available with a count of 2473 cars.
- Hyundai is the second highest brand in availability with a count of 1218 cars.
- Land rover, Jaguar, Mitsubishi, Volvo & ISUZU are the brands which are less available.

4) Average price of all branded cars



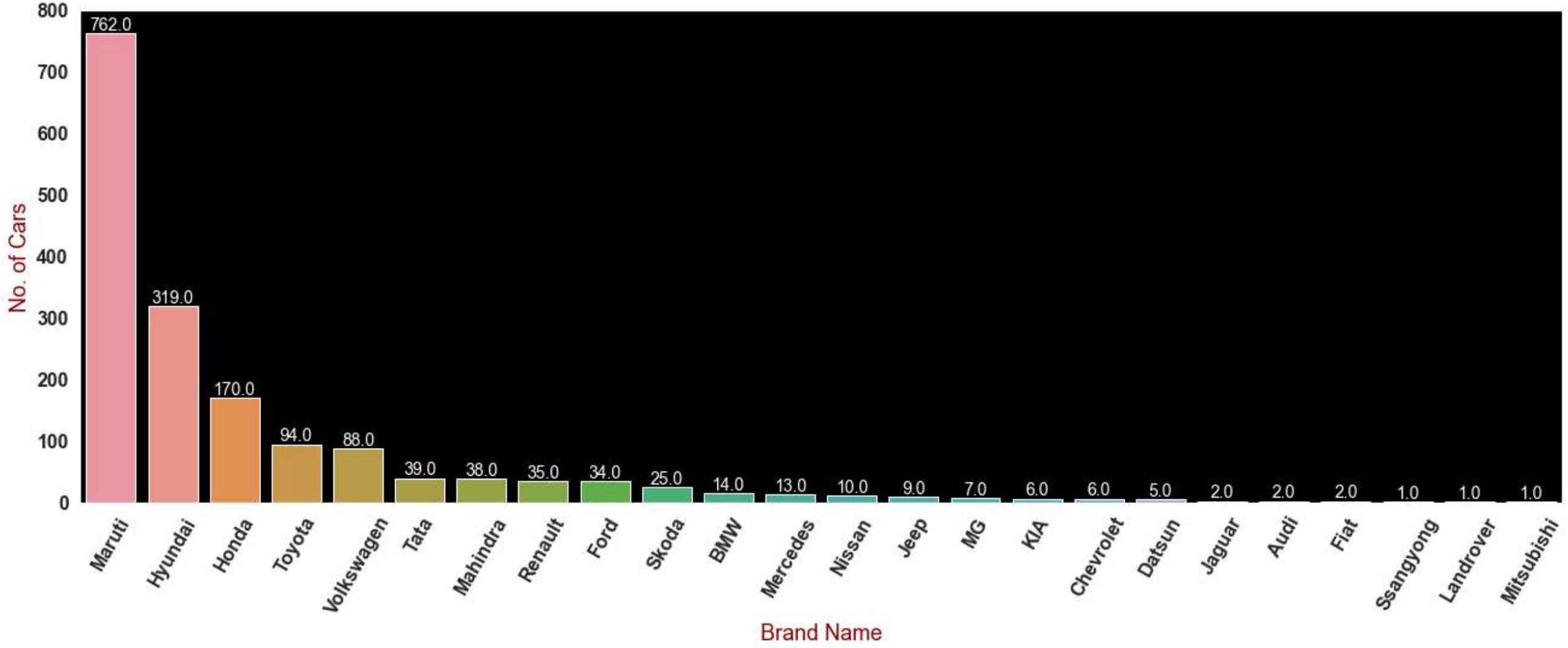
- Jaguar has the highest average price of around (20.5 Lakhs), followed by MG Hector (18.1 Lakhs), Mitsubishi (15.8 Lakhs), Jeep (14.4 Lakhs), KIA (14.1 Lakhs).
- Chevorlet has the least average price of around (2.1 Lakhs).

5) Total number of different car brands in Hyderabad



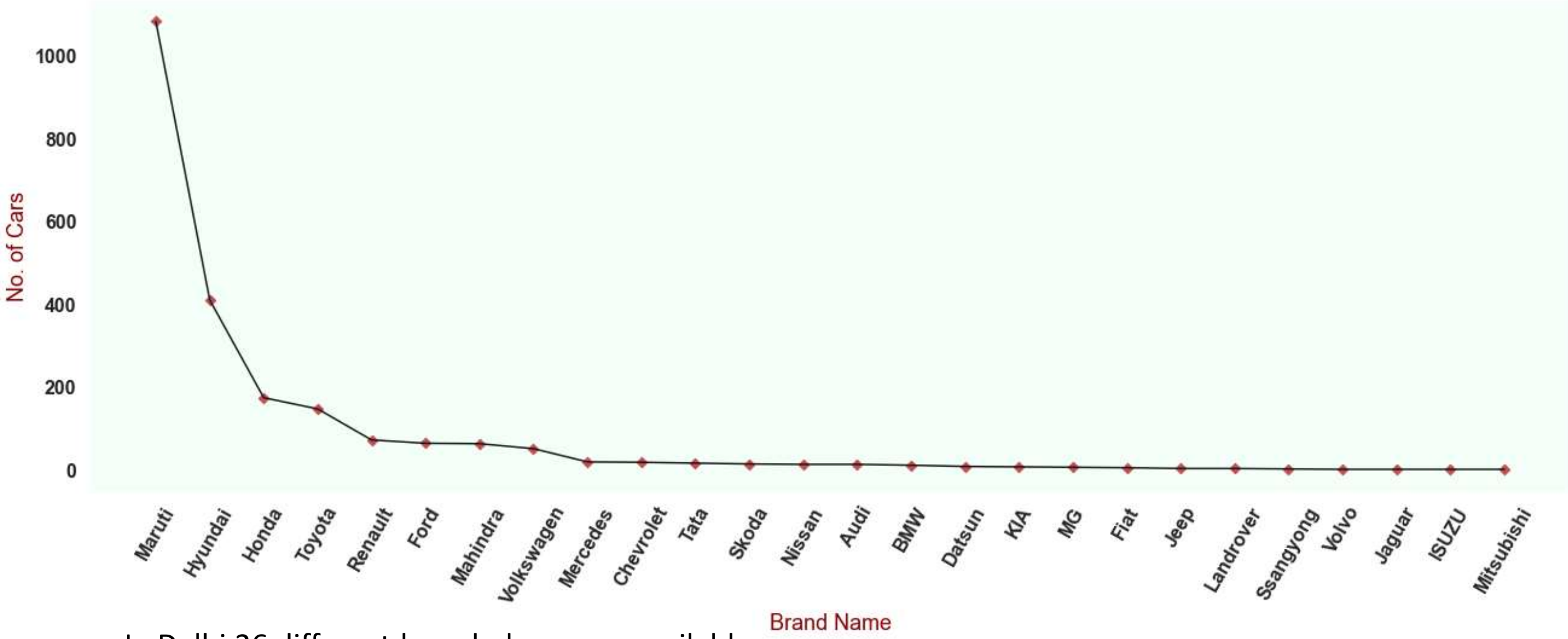
- In Hyderabad 13 different branded cars are available.
- Maruti brand cars are mostly available in Hyderabad with a count of 240 cars, followed by Hyundai (104), Honda(18), Toyota (11)
- Audi (1), Nissan (1), Skoda (1) are the less available cars in Hyderabad.

6) Total number of different car brands in Mumbai



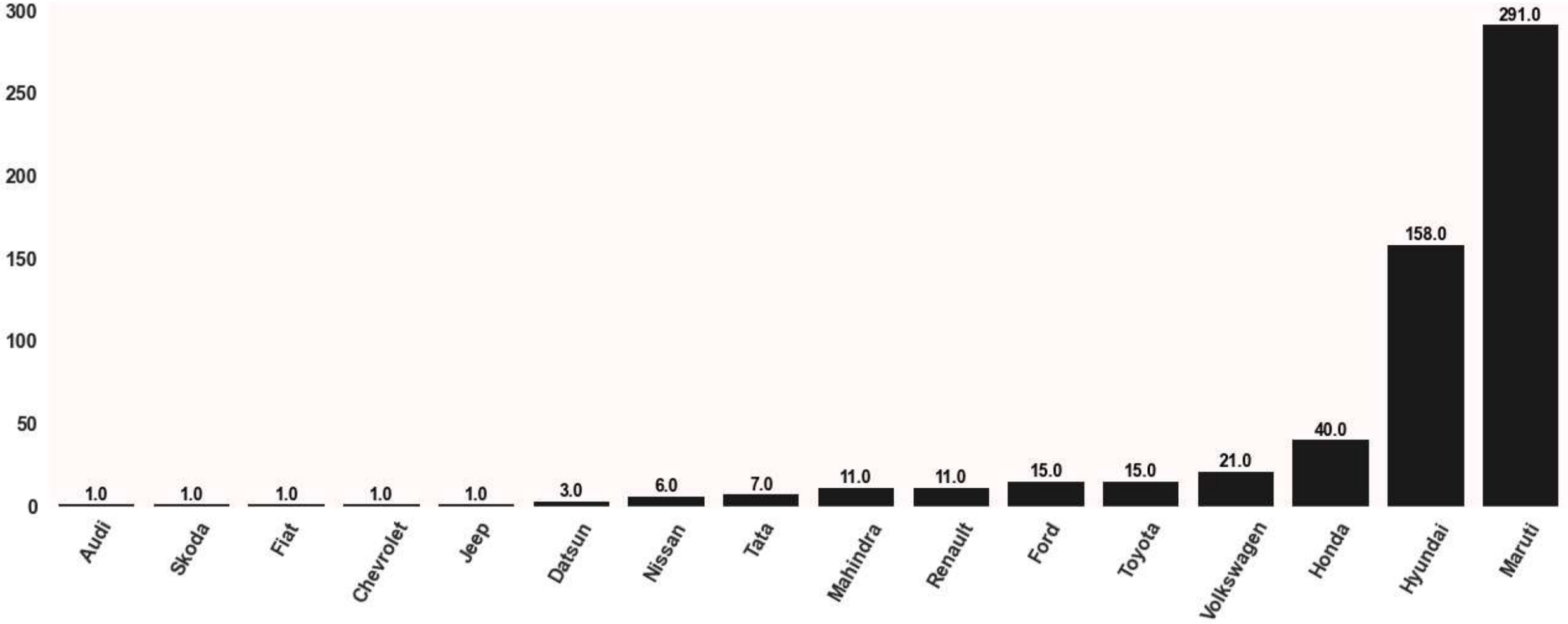
- In Mumbai 24 different branded cars are available.
- Maruti brand cars are mostly available in Mumbai with a count of 762 cars, followed by Hyundai (319), Honda(170), Toyota (94), Volkswagen (88), Tata (39).
- Audi (2), Jaguar (2), Fiat (2) Land rover (1), Mitsubishi (1) are the less available car brands in Mumbai.

7) Total number of different car brands in Delhi



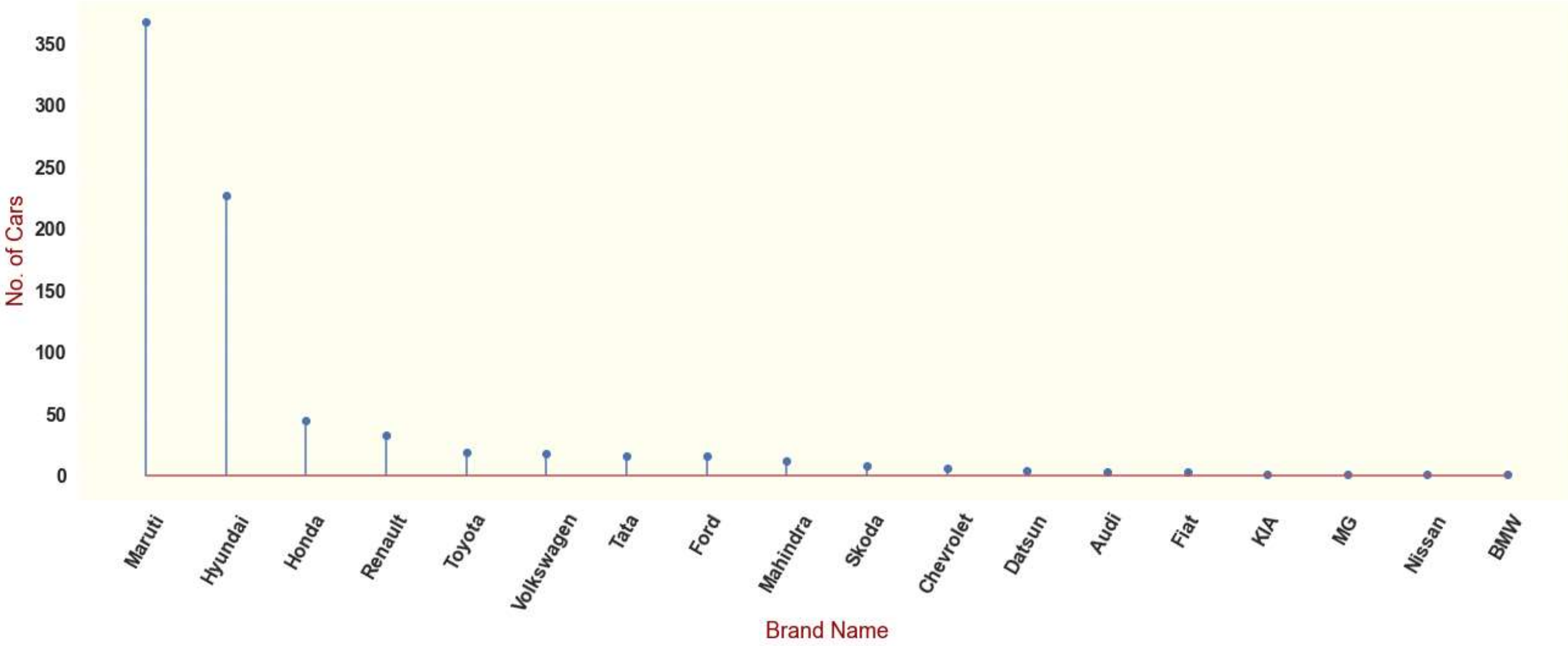
- In Delhi 26 different branded cars are available.
- Maruti brand cars are mostly available in Delhi with a count of 1083 cars, followed by Hyundai (410), Honda(174), Toyota (147), Renault (72), Ford (64).
- Volvo (1), Jaguar (1), ISUZU (1), Mitsubishi (1) are the less available car brands in Delhi.

8) Total number of different car brands in Chennai



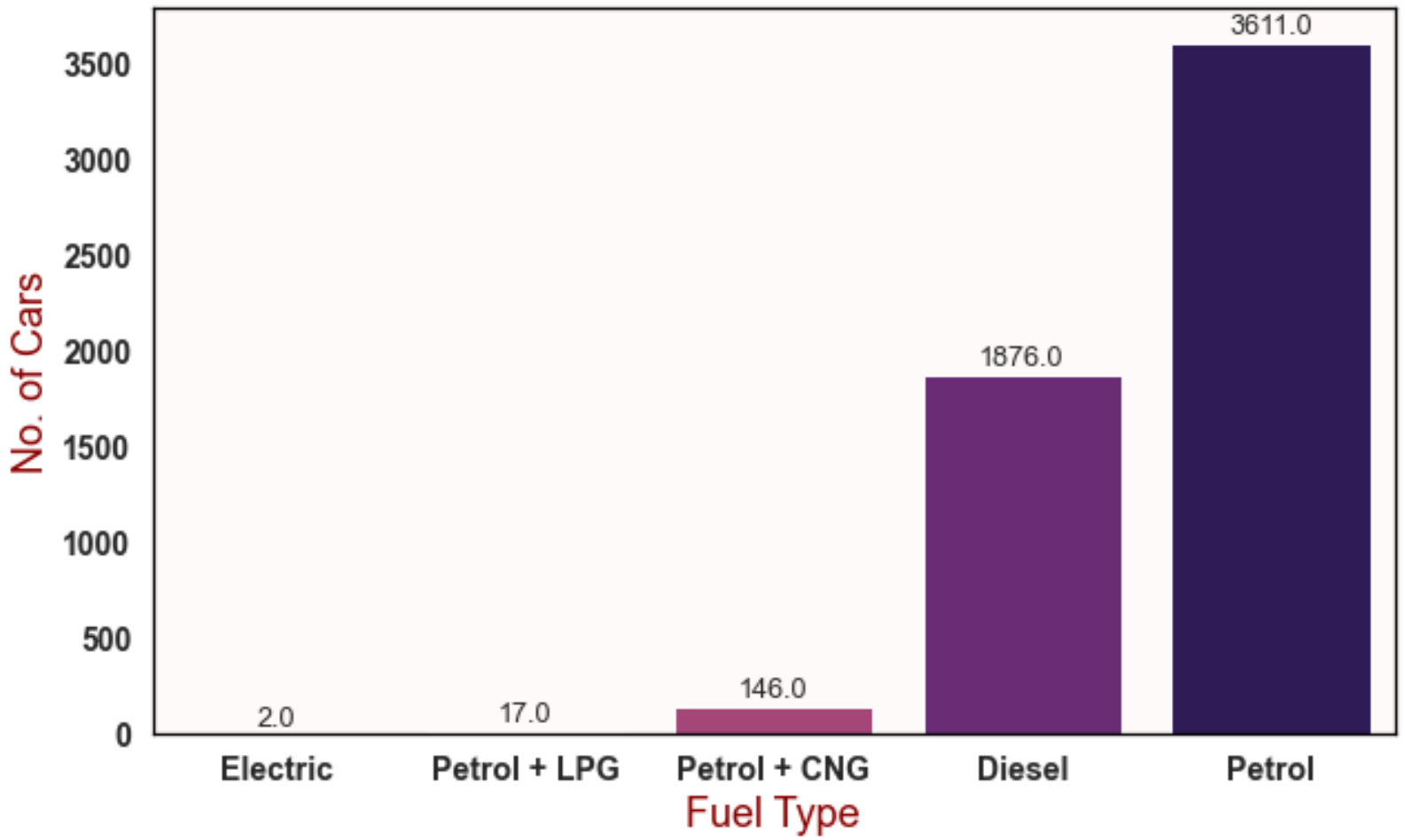
- In Chennai 16 different branded cars are available.
- Maruti brand cars are mostly available in Chennai with a count of 291 cars, followed by Hyundai (158), Honda(40), Volkswagen (21), Toyota (15), Ford (15).
- Audi (1), Skoda (1), Fiat (1), Cheverlot (1), Jeep (1)are the less available car brands in Chennai.

9) Total number of different car brands in Bangalore



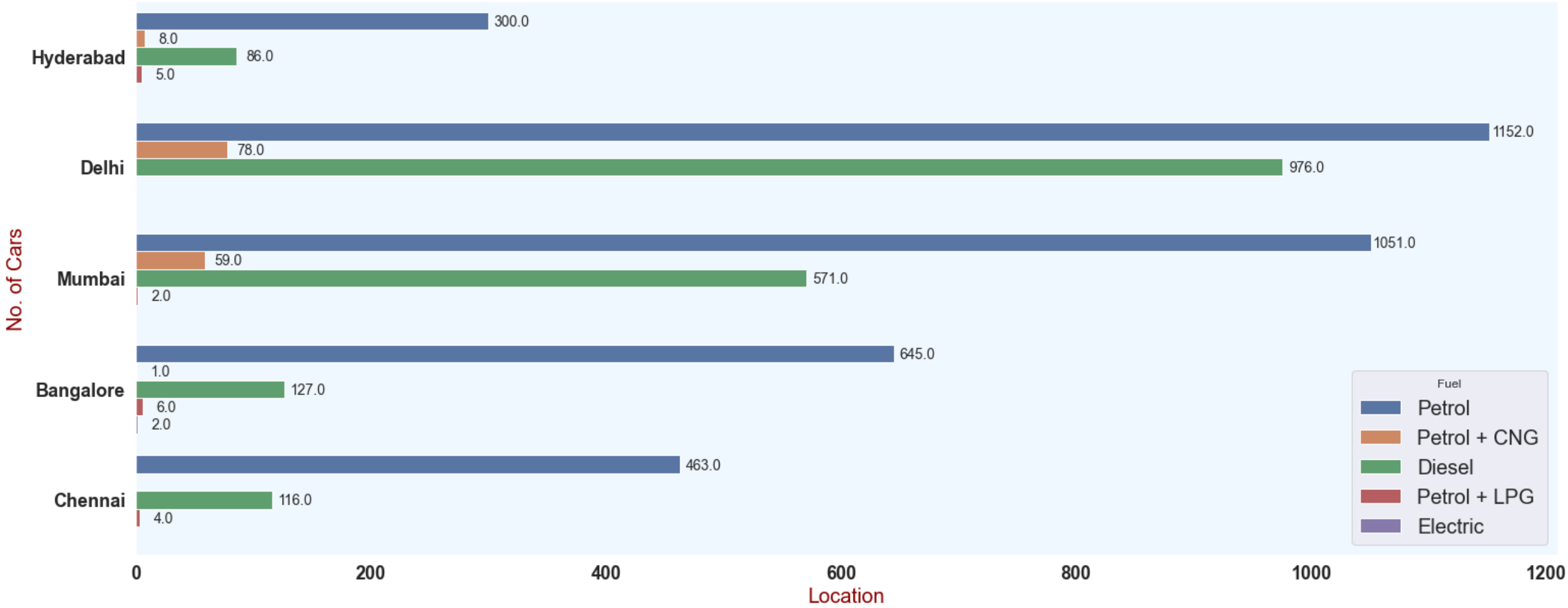
- In Bangalore 18 different branded cars are available.
- Maruti brand cars are mostly available in Bangalore with a count of 762 cars, followed by Hyundai (367), Honda(45), Renault (33), Toyota (19), Volkswagen (18).
- BMW (1) MG Hector (1), Nissan (1) are the less available car brands in Bangalore.

10) Total number of cars brands based on fuel type



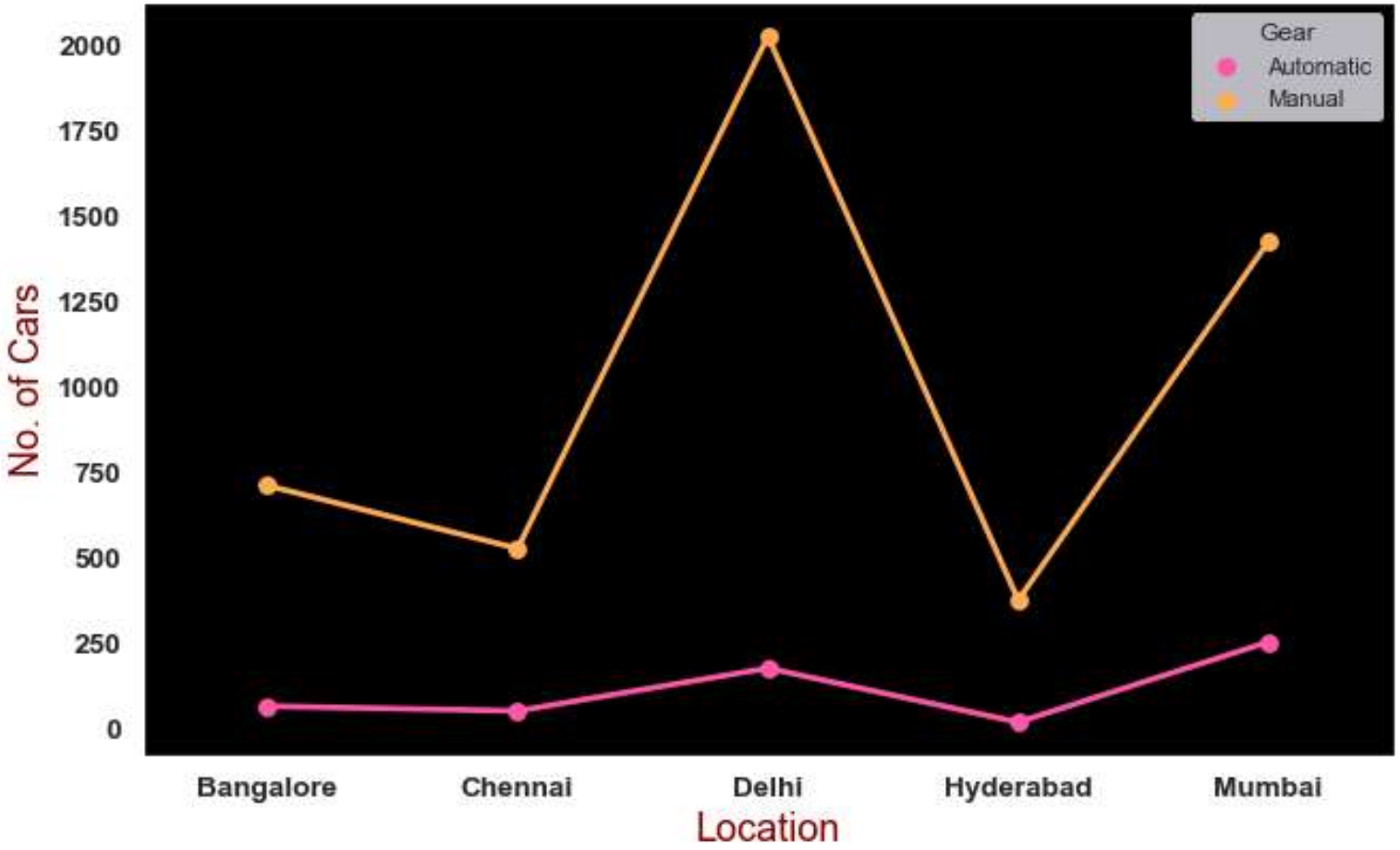
- Among 5562 cars, Petrol cars are mostly available with a count of 3611 cars.
- The total number of Diesel cars available are 1876.
- The total number of cars which runs with Petrol & CNG are 146, followed by Petrol & LPG with a total of 17 cars. There are only 2 Electric cars available.

11) Total number of cars based on fuel type in different cities



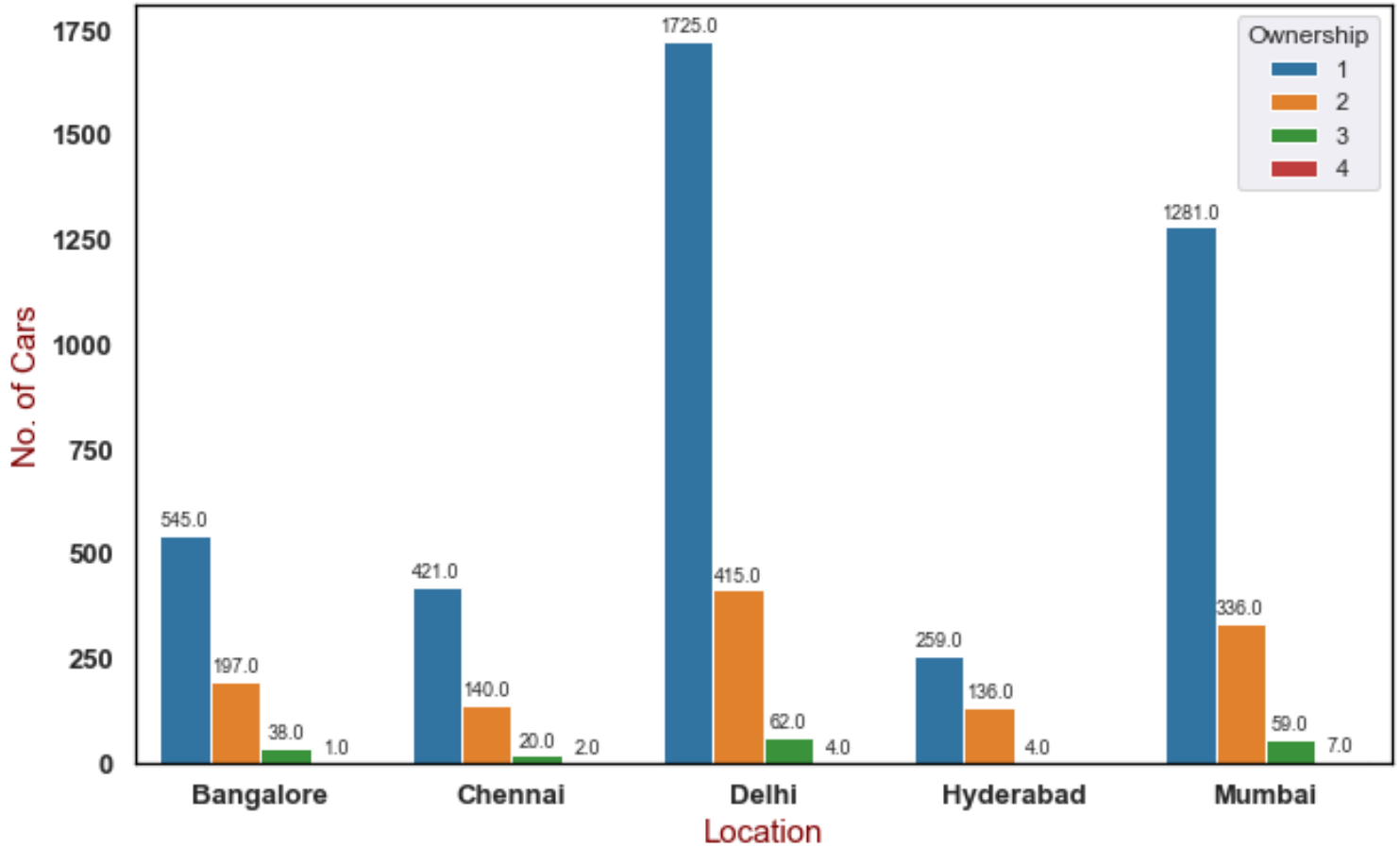
- Petrol & Diesel cars are mostly available in Delhi & less available in Hyderabad.
- Petrol + CNG cars are mostly available in Delhi & less available in Bangalore.
- Petrol + LPG cars are mostly available in Bangalore & less available in Mumbai.
- Only 2 Electric cars are available, they are available in only Bangalore.

12) Total number of cars in different cities based on Gear



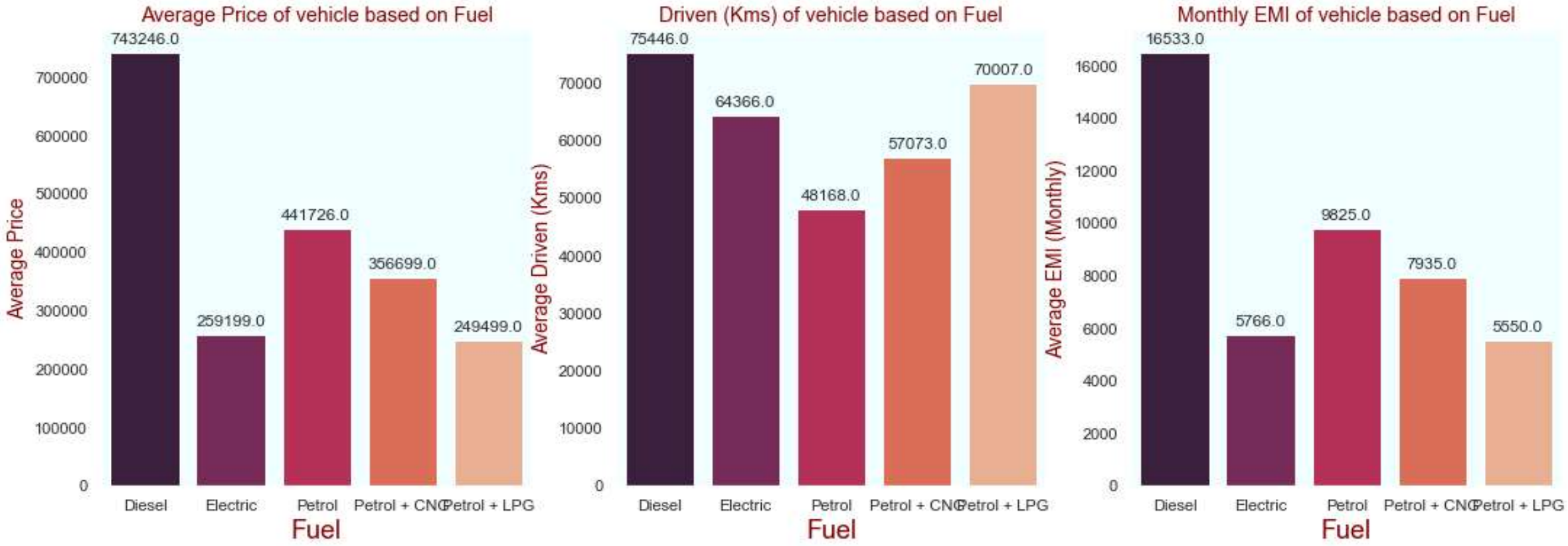
- Mumbai has the highest number of automatic transmission cars with a count of 256 cars.
- Hyderabad has the least number of automatic transmission cars with a total of 21 cars.
- Delhi has highest number of manual transmission cars with a count of 2021 cars.
- Hyderabad has the least number of manual transmission cars with a total of 378 cars

13) Total number of cars in different cities based on Ownership



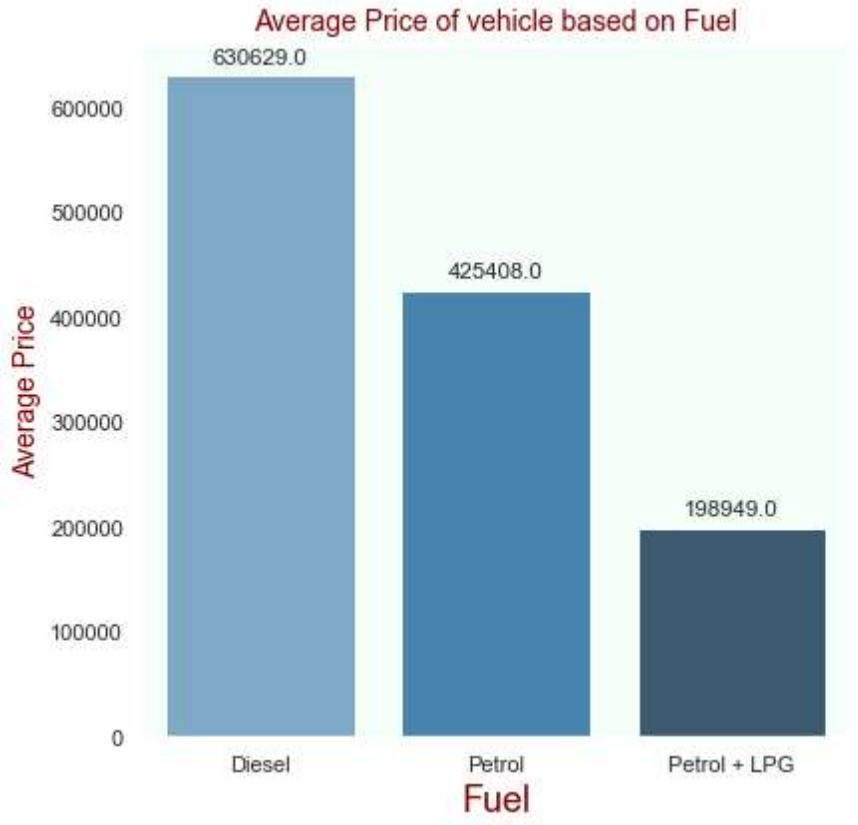
- Delhi city has highest cars with 1st Ownership with a count of 1725 Cars, Hyderabad has the least number of cars with 1st Ownership with a total of 259 Cars.
- Delhi city has highest cars with 2nd Ownership with a count of 415 Cars, Hyderabad has the least number of cars with 2nd Ownership with a total of 136 Cars.
- Mumbai city has highest number of 4th Ownership cars with a count of 7 cars, Bangalore has the least number of 4th Ownership with only 1 car.

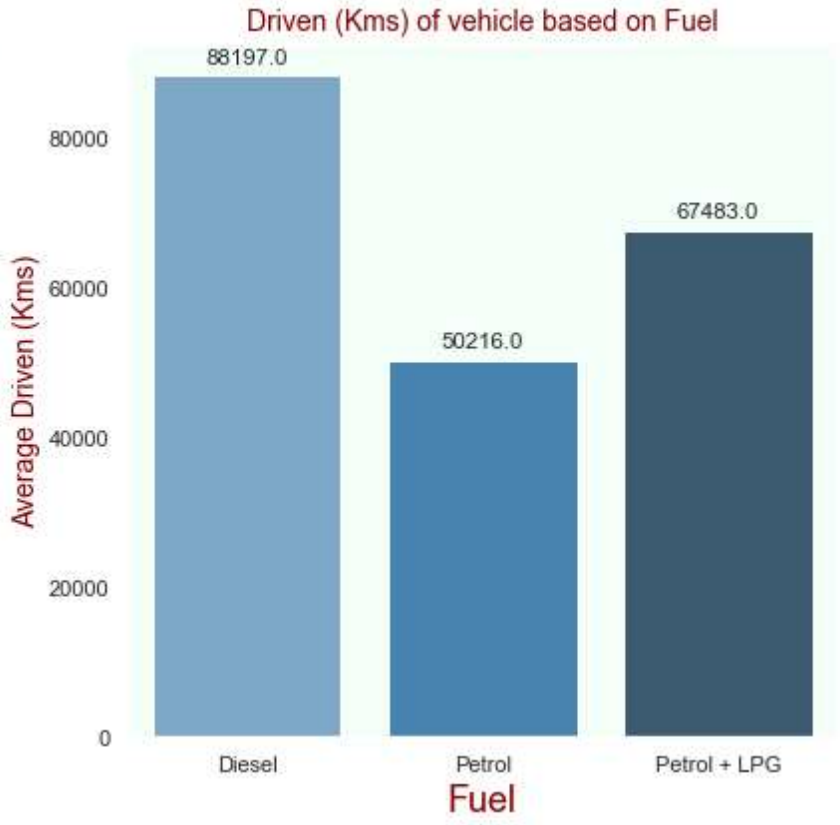
14) Average Price, EMI, Driven(Kms) based on Fuel type in Bangalore

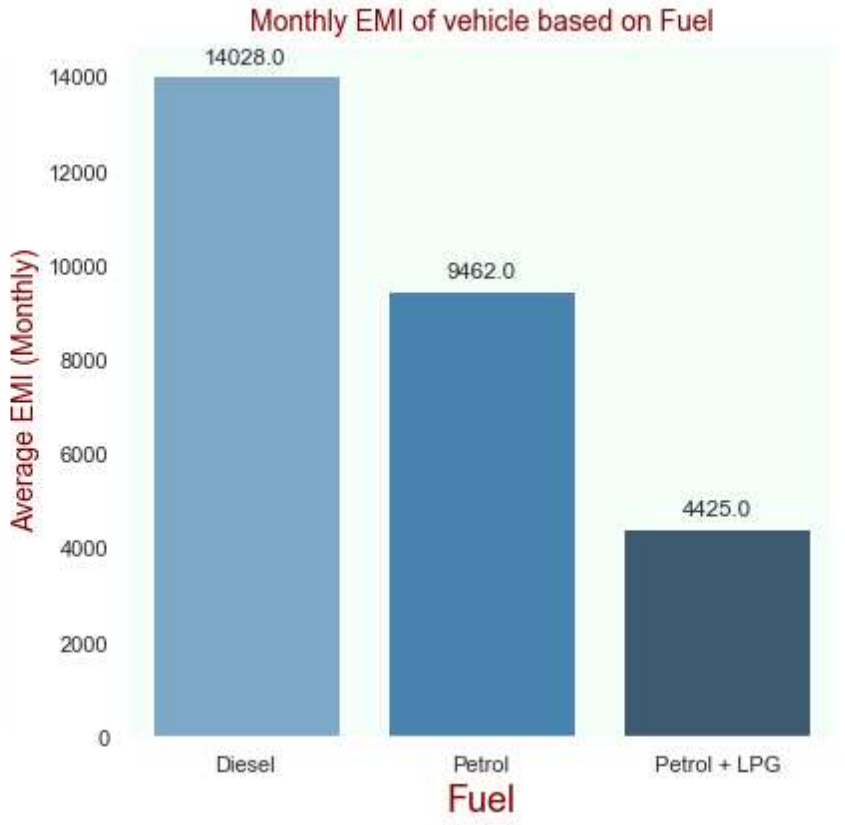


- ✓ **<u>Average Price in (₹):</u>**
  Diesel cars has high average price of 7.5 Lakhs, **Electric** cars have less average price of 2.5 Lakhs
- ✓ **<u>Driven (Kms):</u>**
  Diesel cars have most average driven distance of 75K Kms, **Petrol** cars are least driven with a distance of 48K Kms
- ✓ **<u>Monthly EMI in (₹):</u>**
  Monthly EMI is less for **Petrol + LPG** cars with an avg. of 5K , EMI for **Diesel** cars is high with avg. of 16K

15) Average Price, EMI, Driven(Kms) based on Fuel type in Chennai

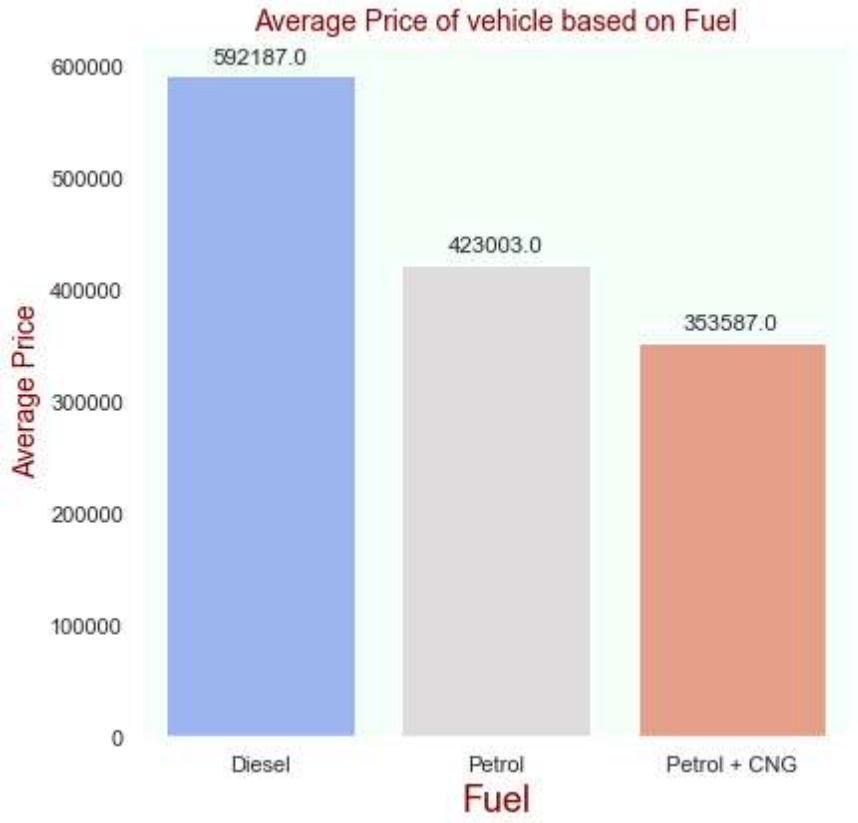


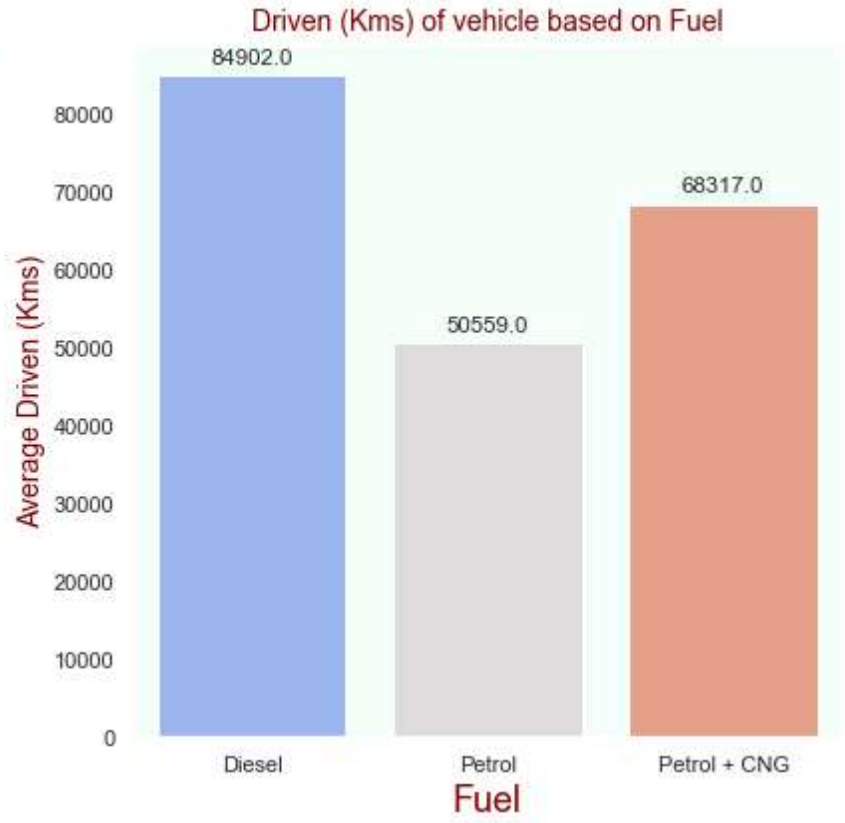


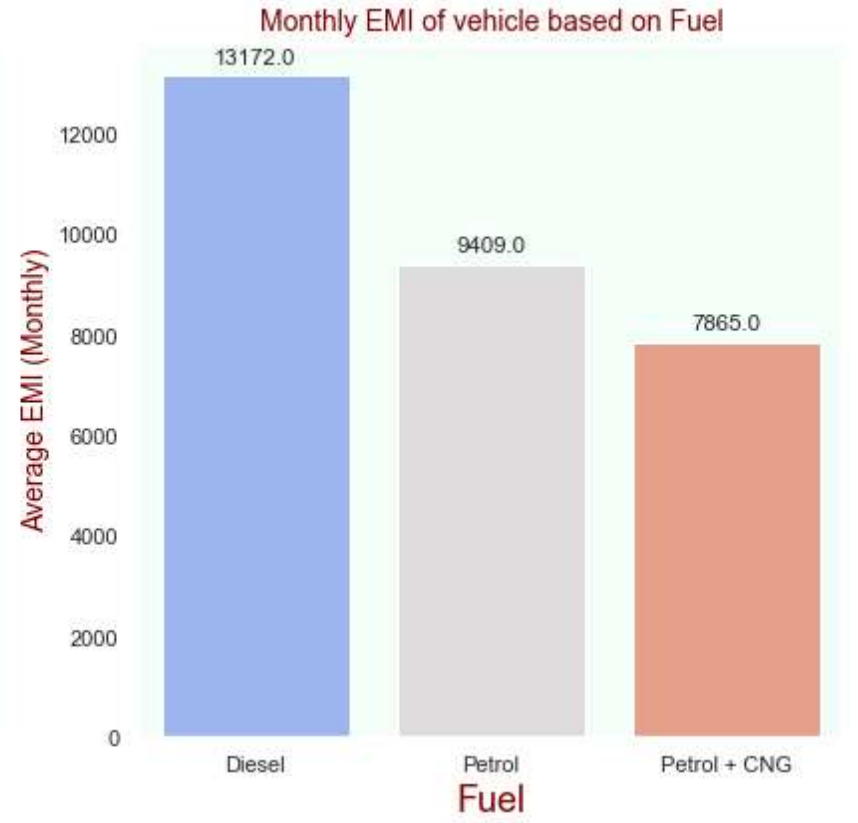


- ✓ **<u>Average Price in (₹):</u>**
  Diesel cars has high average price of 6.3 Lakhs, **Petrol + LPG** cars have less average price of 2 Lakhs
- ✓ **<u>Driven (Kms):</u>**
  Diesel cars have most average driven distance of 88K Kms, **Petrol** cars are least driven with a distance of 50K Kms
- ✓ **<u>Monthly EMI in (₹):</u>**
  Monthly EMI is less for **Petrol + LPG** cars with an avg. of 4.5K , EMI for **Diesel** cars is high with avg. of 14K

16) Average Price, EMI, Driven(Kms) based on Fuel type in Delhi

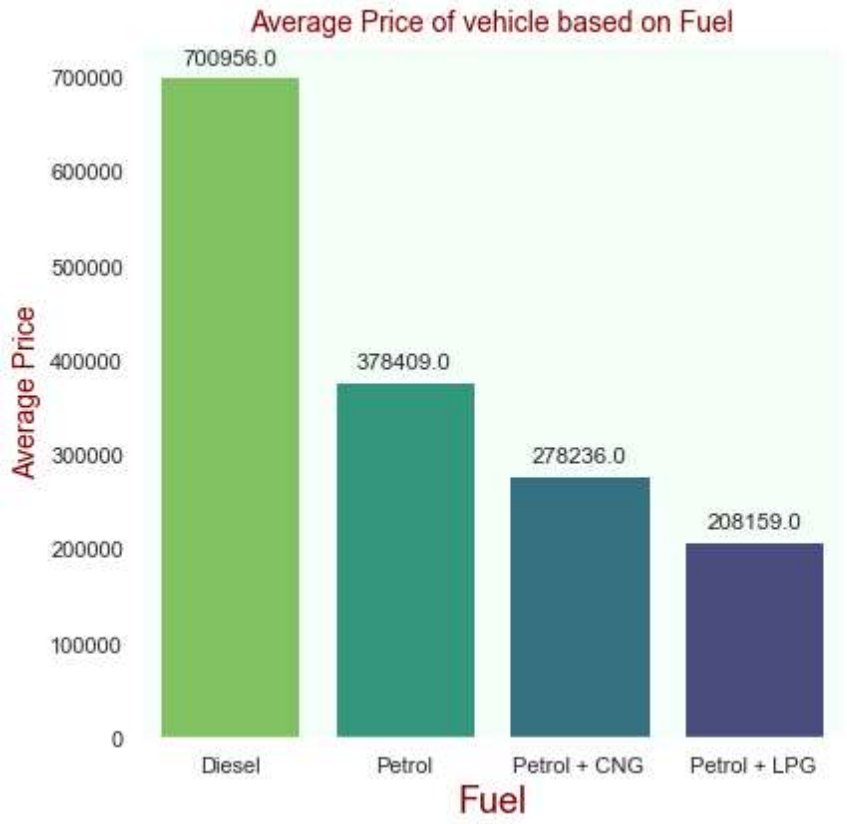


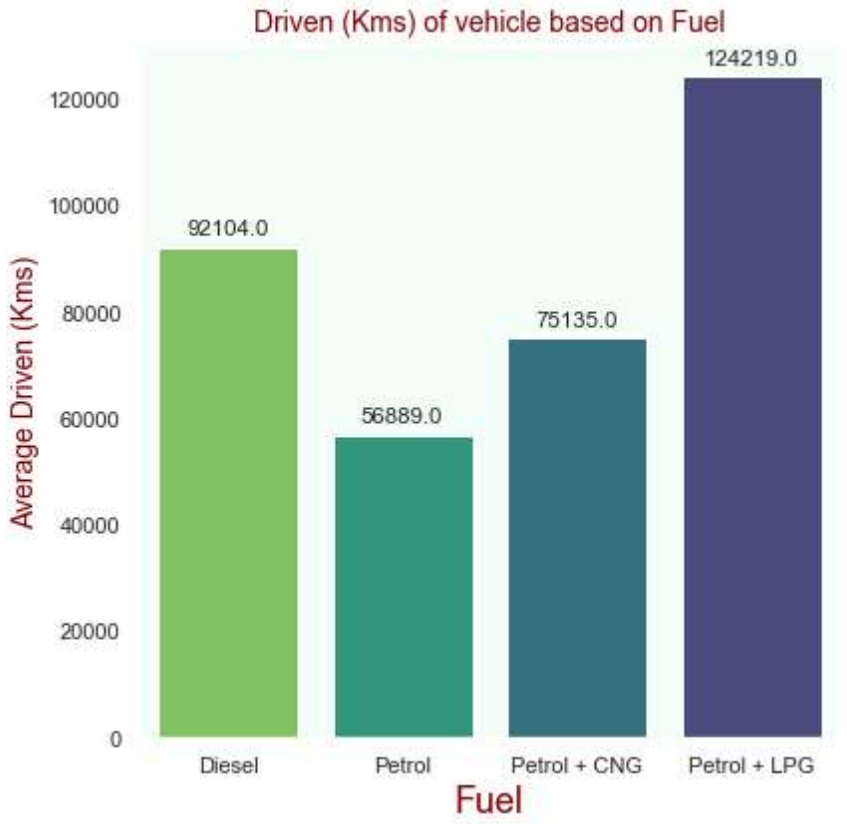


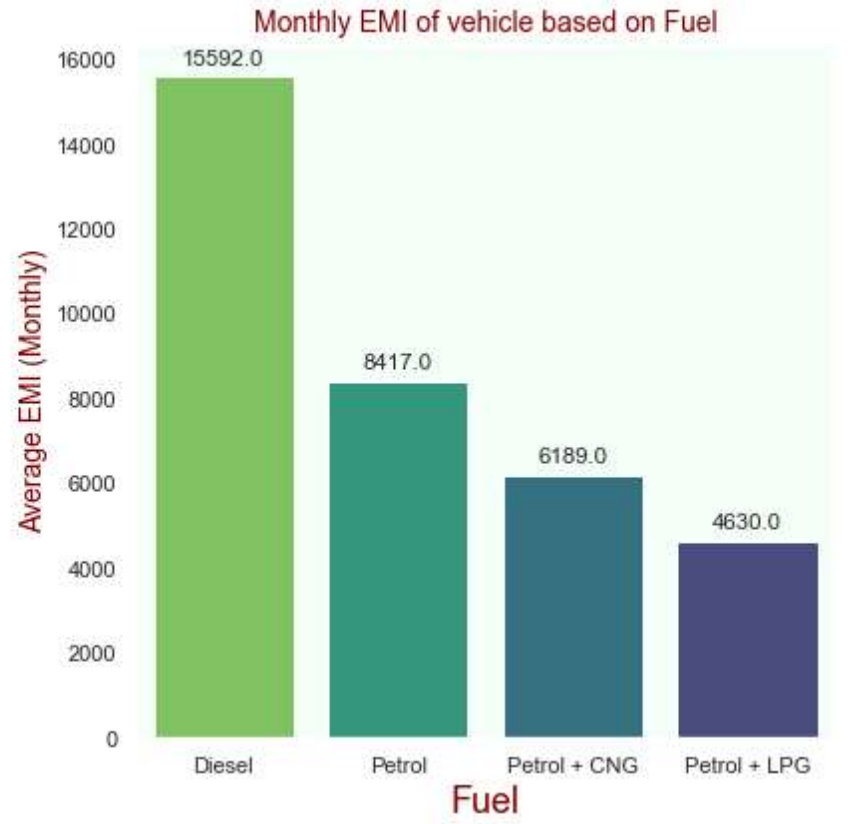


- ✓ **<u>Average Price in (₹):</u>**
  Diesel cars has high average price of 6 Lakhs, **Petrol + CNG** cars have less average price of 3.5 Lakhs
- ✓ **<u>Driven (Kms):</u>**
  Diesel cars have most average driven distance of 85K Kms, **Petrol** cars are least driven with a distance of 50K Kms
- ✓ **<u>Monthly EMI in (₹):</u>**
  Monthly EMI is less for **Petrol + CNG** cars with an avg. of 7.8K , EMI for **Diesel** cars is high with avg. of 13K

17) Average Price, EMI, Driven(Kms) based on Fuel type in Hyderabad

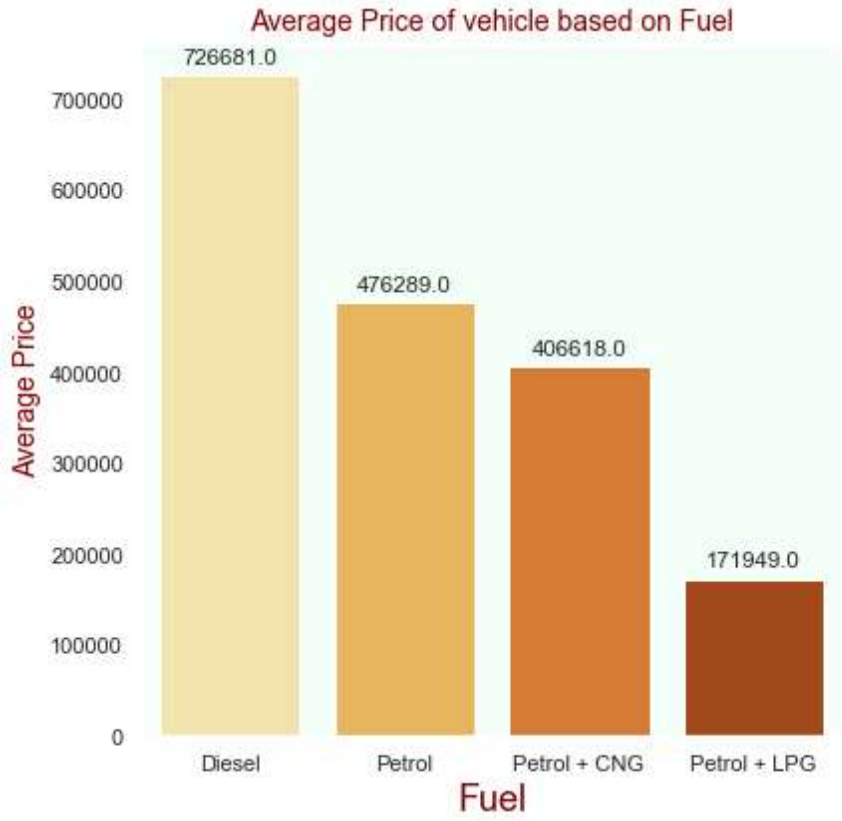


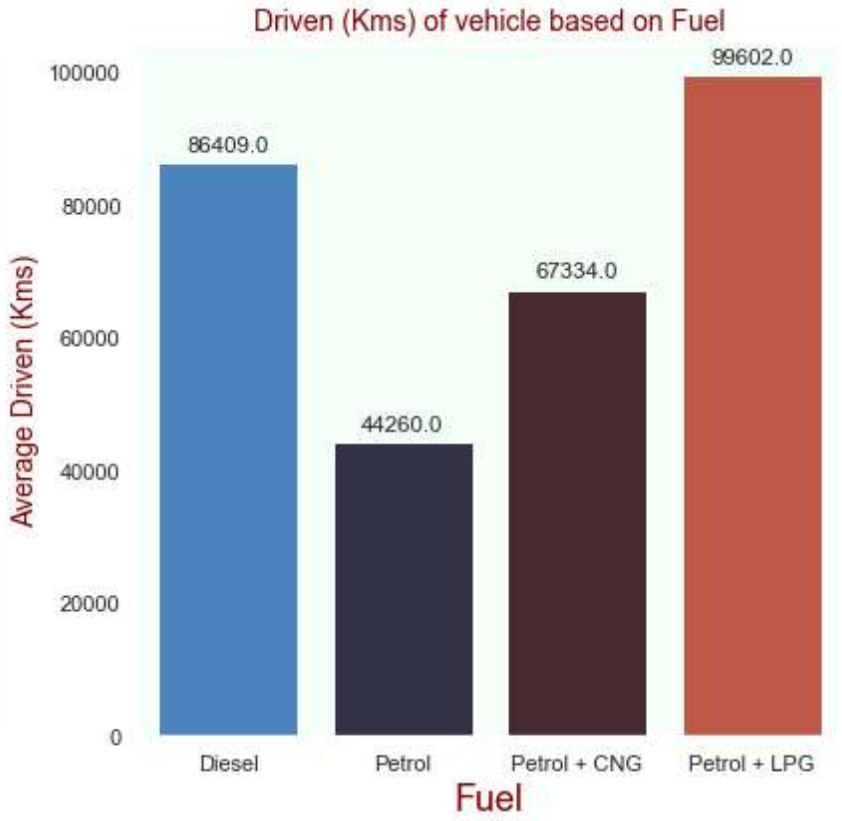


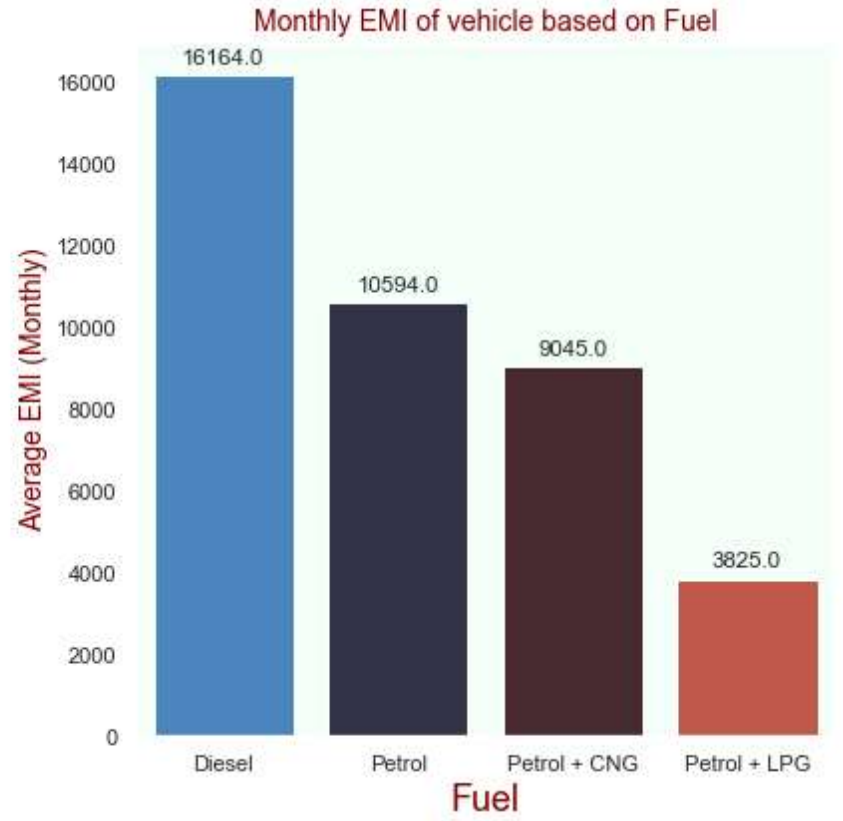


- ✓ **<u>Average Price in (₹):</u>**
  Diesel cars has high average price of 7 Lakhs, **Petrol + LPG** cars have less average price of 2 Lakhs
- ✓ **<u>Driven (Kms):</u>**
  Petrol + LPG cars have most avg driven distance of 1.2L Kms, **Petrol** cars are least driven with a distance of 56K Kms
- ✓ **<u>Monthly EMI in (₹):</u>**
  Monthly EMI is less for **Petrol + LPG** cars with an avg. of 4.6K , EMI for **Diesel** cars is high with avg. of 15.5K

18) Average Price, EMI, Driven(Kms) based on Fuel type in Mumbai

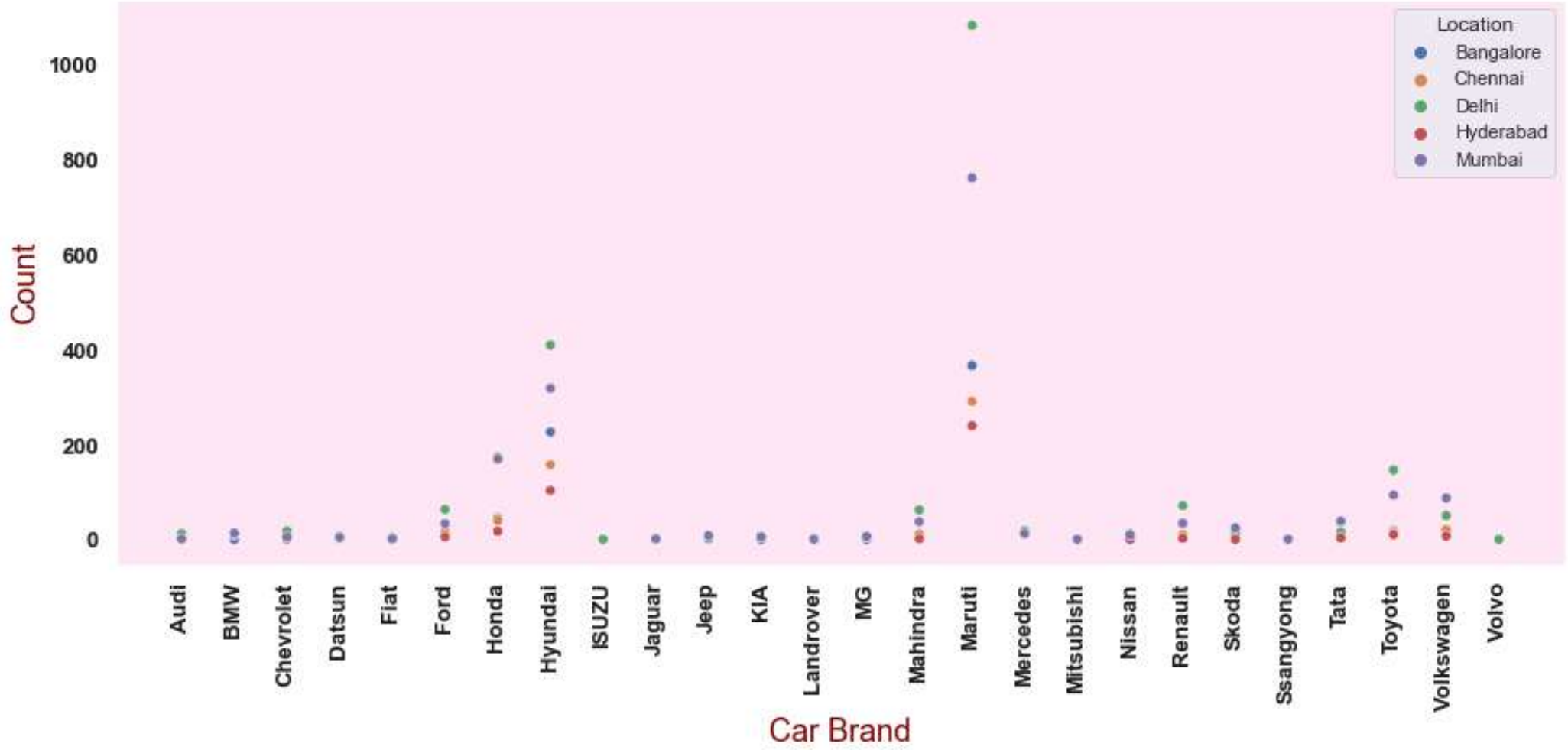






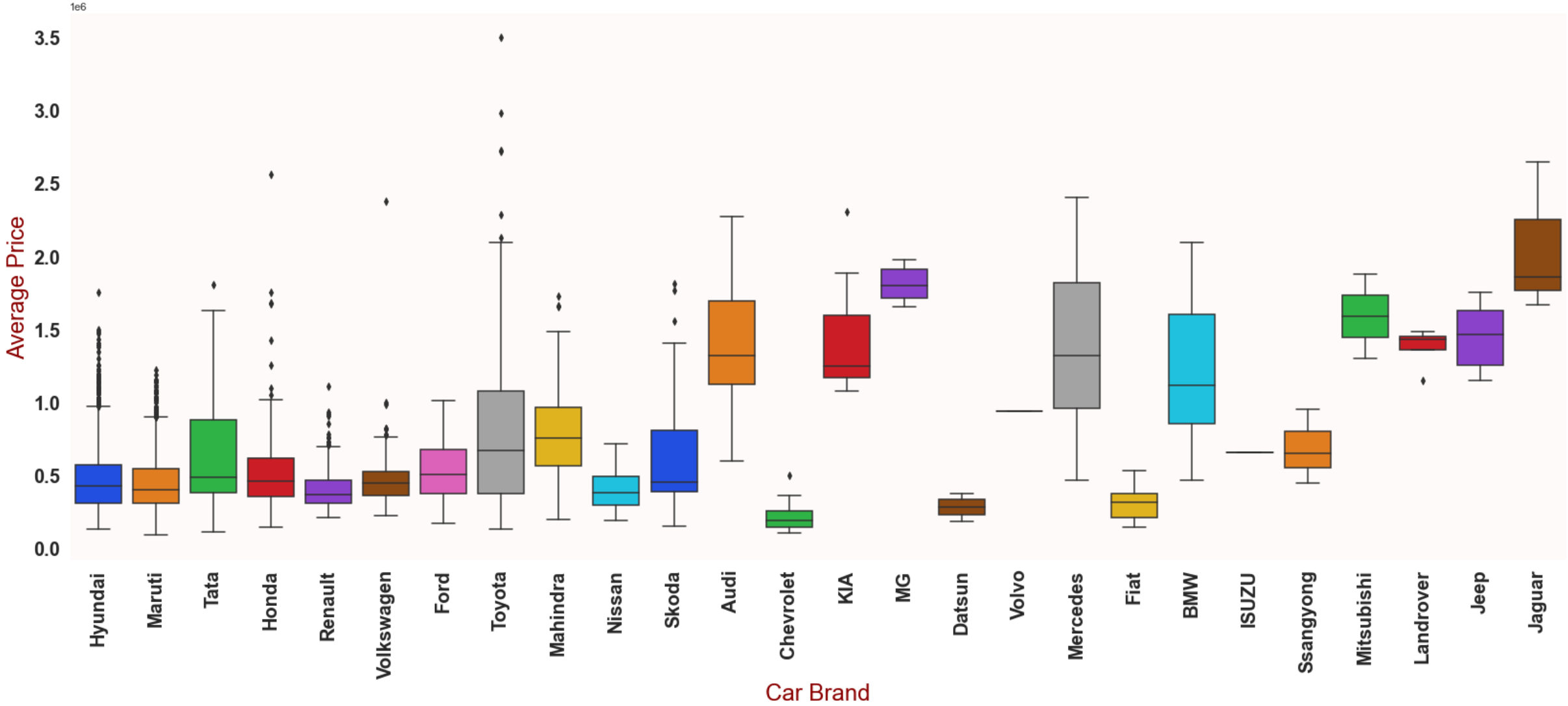
- ✓ **Average Price in (₹):**
  Diesel cars has high average price of 7.2 Lakhs, **Petrol + LPG** cars have less average price of 1.7 Lakhs
- ✓ **Driven (Kms):**
  Petrol + LPG cars have most avg driven distance of 99K Kms, **Petrol** cars are least driven with a distance of 44K Kms
- ✓ **Monthly EMI in (₹):**
  Monthly EMI is less for **Petrol + LPG** cars with an avg. of 3.8K , EMI for **Diesel** cars is high with avg. of 16K

19) Total number of different car brands in different cities



- Maruti, Hyundai, Mahindra brands are available in all cities, they are mostly available in Delhi.
- ISUZU & Volvo are the brands only available in Delhi.
- Datsun, Fiat, Jaguar, Land rover, Mitsubishi are the brands only available in Mumbai.

20) Average Price of each car in all brands



- The brands with more number of outliers are Hyundai, Maruti, Renault, Honda.
- Toyota, Honda brands have long range of outliers.

21) Heatmap to check the correlation

- Correlation ranges from -1 to +1, Values closer to zero means there is no linear relationship between the two variables.
- The values near to 1 are more positively correlated i.e, if one variable increases the other also increases.
- The values near to -1 are more negatively correlated i.e, if one variable increases the other decreases.

Conclusion

- The availability of cars in 'Delhi'(2200+) is the highest among other 4 cities.
- Compared to other 4 cities 'Hyderabad'(415+) has less available cars.
- 'Maruti' brand cars are widely available with a count of around 2800 cars in all the cities, followed by Hyundai(1240+), Honda(449), Toyota(280+)
- Around 3660+ cars runs with petrol.

High budget car among all cities:

- Toyota Land CruiserLC200 VX 2 PREMIUM (3495000/-), 2010 Model, available in Mumbai.

Low budget car among all cities:

- Maruti AltoLX (91000/-), 2008 Model, available in Delhi

My Experience after the project

- Since its an infinite scrolling website, I've faced some difficulties when collecting the data from the website.
- Thus, I've learnt the things required to scrape any kind of website.
- Now, I'm very much confident about web scraping.

INNOMATICS RESEARCH LABS

Future Scope

- Since ‘Cars24’ is a dynamic website, the data in this website changes every minute. To analyze the data, every time the code has to be updated.

- This analysis will be helpful for the customers / dealers while buying or selling a car.

THANK YOU

INNOMATICS
RESEARCH LABS