ONLINE VEHICLE BOOKING MARKET

KUSHI SAHU
ISHA TADAS
RAKESH SINGH RAWAT
HARSHIT RAIZADA
AAYUSH SOBTI

Github Link: https://github.com/Isha-1708/Online-Vehicle-Booking

Abstract:

This project mainly deals with online vehicle booking with help of machine learning algorithm.
The online taxi services market covered in this report is segmented by service type into ride-hailing and ride-sharing. Ride-sharing refers to an arrangement or instance involving the sharing of rides in a motor vehicle with other people, mainly commuters. The different types of vehicles include motorcycles and cars and involve various payment modes such as online and cash. The online taxi services market consists of sales of taxi services and related goods by entities (organizations, sole traders, and partnerships) that provide passenger transportation by automobile or van through online reservations or on a booking basis. This industry comprises establishments primarily engaged in providing passenger transportation by taxi that are booked by using a mobile application. These companies use online platforms and mobile applications to offer their services to customers. Examples include Uber and Ola. Taxis booked online through an app are mentioned in the report as ride-hailing services. This does not include taxis stopped on roadsides. Only goods and services traded between entities or sold to end consumers are include.

INTRODUCTION:

It is 21st century, nearly everything we use is being digitized, cab booking is no exception. Meru cab services started in 2004 in Indian metro cities and then almost after six years, app based rental cab service was introduced in the Indian market. The app based cab services revolutionized the cab service market. Among various transportation modes, cab service gained popularity because of its advantage of door to door service and now because of technological advancement customers were able to book cabs at competitive prices in just one click using their smartphones. In densely populated countries like India where parking is a major problem because of space crunch as well as public transports are over-crowded during peak hours, these app based services gained tremendous popularity. Before these app based cab services came on the market, a cab used to run 30% of its time without passengers. The companies who run these aggregated cab services mainly don't own any car but they are tied up with the drivers running local cab services who agrees to register with the organization, and the agency simply acted as a buffer or middleman between the drivers and customers by charging a commission only for their "match-making" services while developing the cab application. These app based services need an app which is easy to navigate and needs good android/ IOS development skills. Also, a very strong and protected cloud architecture which is easily scalable is needed because lots of users are going to use it and hence the cloud infrastructure really needs to support thousands of connections simultaneously with an effective well-structured database. There are a number of applications providing such services in India such as Ola, Uber etc. Such companies provide commercial services. On the other hand, there are few carpooling applications such as quick ride, who are on a mission to remove traffic congestion from the road. It is definite to say that carpooling app services are removing carbon emission on a daily basis and is having a better impact on the society, which is motivating. But the question arises whether a person is more likely to use carpooling or is concerned about his/her privacy and going to choose a ride alone? Is it the price factor which is in the customers mind? Or, is it environment friendliness that leads a person to decide which ride to choose on a daily basis? Not only for customer's personal commute, but businesses are also choosing cab booking apps for their business to develop at a faster pace than others.

PROBLEM STATEMENT:

Online Vehicle booking Market

Data Sources:

Importing the Dataset: We will import the dataset that we need to use. So here, we are using the Car Rental Data. It can be imported using the below code

```
df=pd.read_csv("CarRentalData.csv")
```

Data Pre-processing: (steps and libraries used)

Importing Libraries:

We will import the libraries for our model, which is part of data pre-processing. The code is given below:

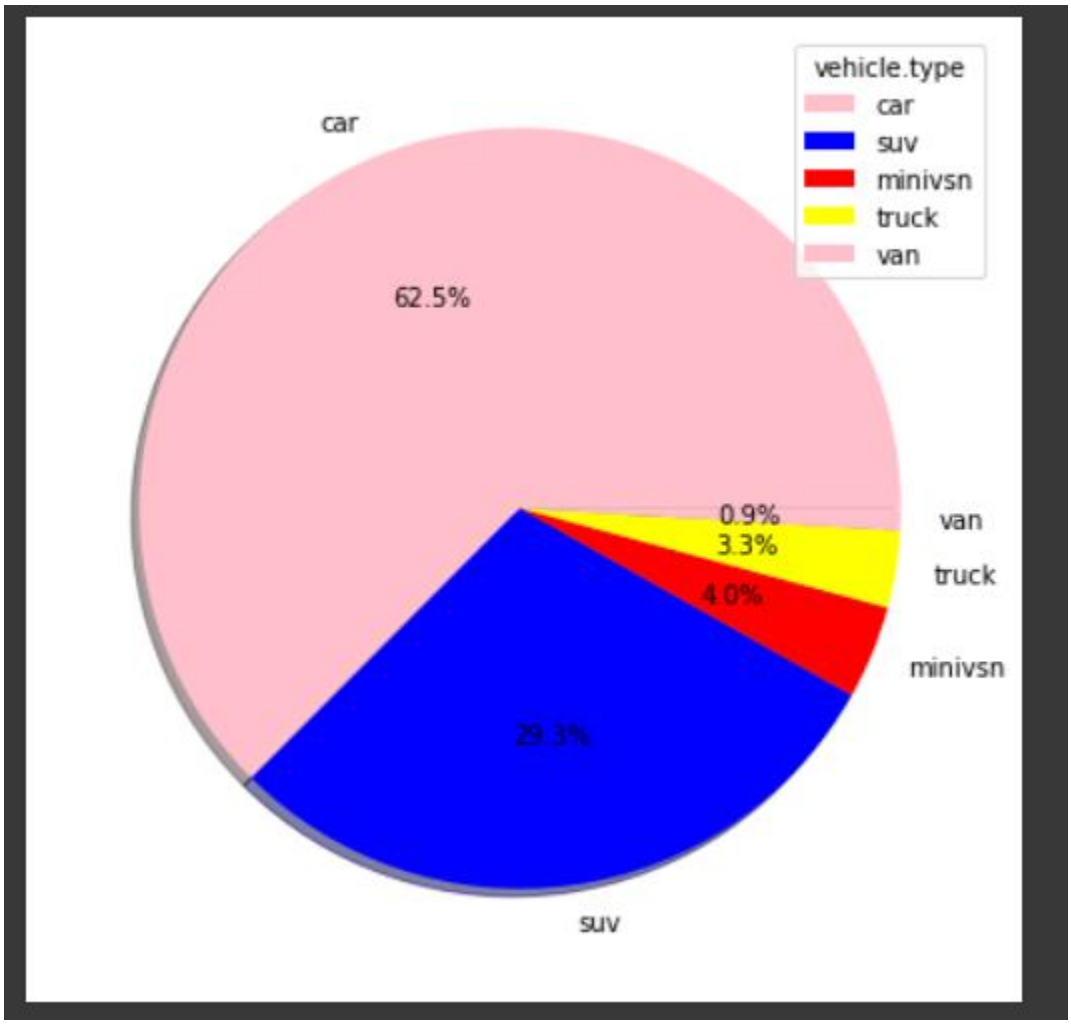
```
Import pandas as pd
Import numpy as np
Import matplotlib.pyplot as plt
Import seaborn as sns
```

- Numpy we have imported for the performing mathematics calculation.
- Matplotlib is for plotting the graph, and pandas are for managing the dataset.
- Seaborn is for data visualization library, it is based on matplotlib.

Exploratory Data Analysis:

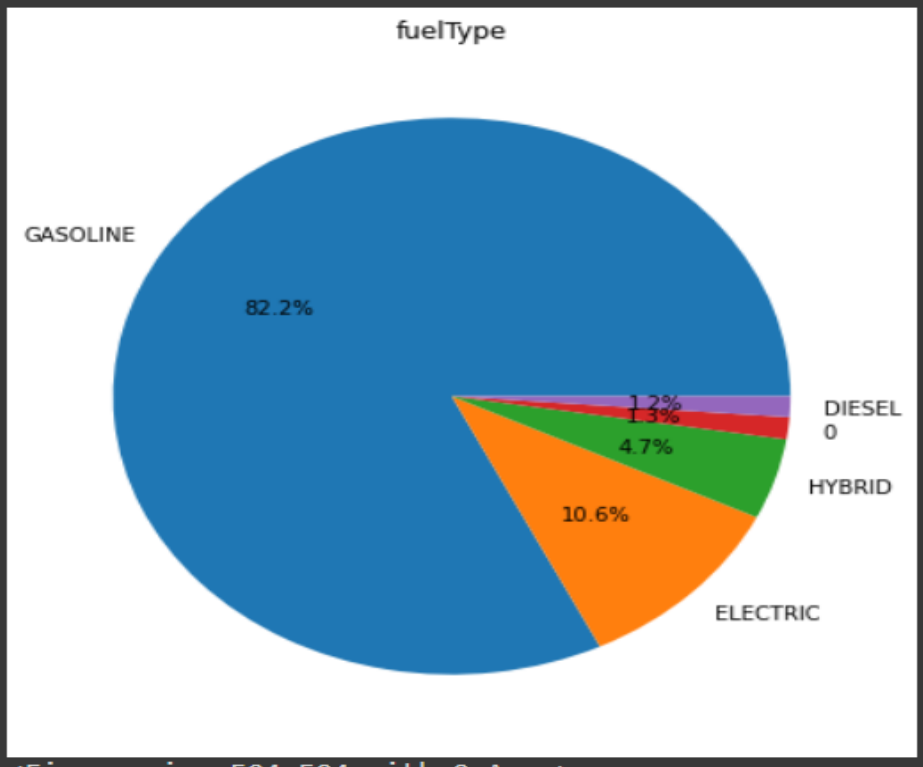
An Exploratory Data Analysis, or EDA is a thorough examination meant to uncover the underlying structure of a data set and is important for a company because it exposes trends, patterns, and relationships that are not readily apparent.

```
[ ]  labels = ['car', 'suv','minivsn','truck','van']
     size = df['vehicle.type'].value_counts()
     explode = [0,0.2]
     colors = ['Pink', 'blue','red','yellow'] ## color Genders

     plt.rcParams['figure.figsize'] = (7,7)
     plt.pie(size, labels=labels, autopct='%1.1f%%', shadow=True, colors=colors,startangle = 0)
     plt.legend(title="vehicle.type",fontsize= 10)
     plt.show()
```



This shows the classification of vehicle type. As we can see car type is most used and then suv followed by minivsn, truck, van.

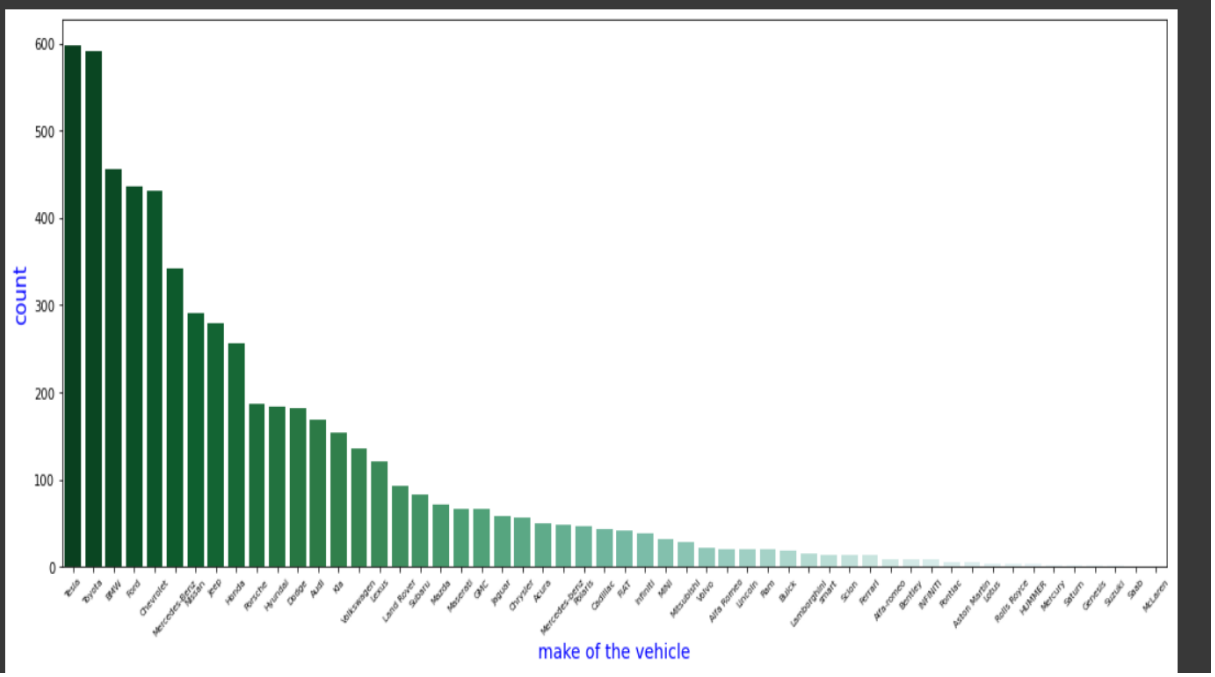
```
[ ]  labels=df['fuelType'].value_counts().index
     values=df['fuelType'].value_counts().values

     #visualization
     plt.figure(figsize=(7,7))
     plt.pie(values ,labels = labels ,autopct='%1.1f%%')
     plt.title('fuelType')
     plt.show()
     plt.savefig('Fuel Type.png', format='png')
```




```
<Figure size 504x504 with 0 Axes>
```

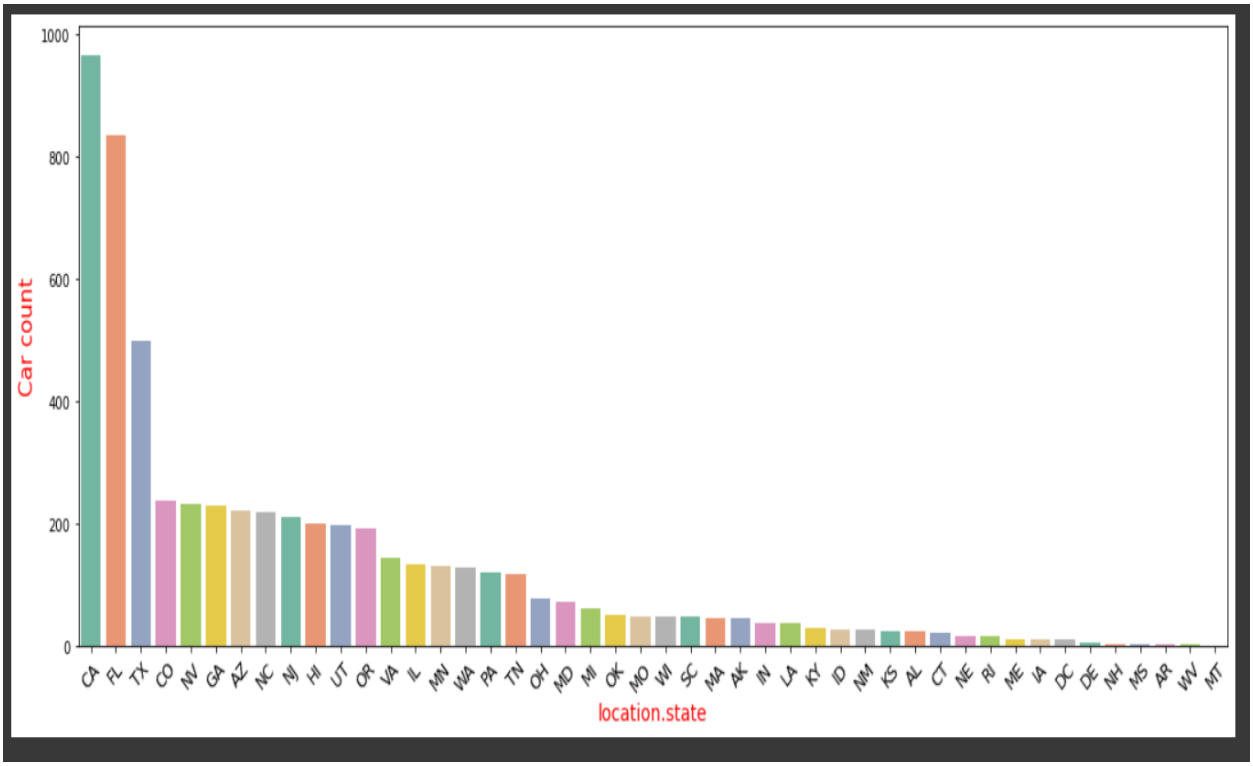
This shows the classification regarding the fuel type. As we can see Gasoline type of fuel is used the most.

```
[ ] labels=dt1['vehicle.make'].value_counts().index
    f, ax = plt.subplots(figsize=(18, 7))
    sns.countplot(x='vehicle.make', data=dt1,
                  order = labels,
                  #hue='vehicle.year'
                  palette="BuGn_r"
                 )
    plt.xticks(rotation= 45,fontsize=7 )
    ax.set_ylabel('count', fontsize=15, color='b')
    ax.set_xlabel('make of the vehicle', fontsize=14, color='b')
    #plt.savefig('make of the vehicle.png', format='png')
    plt.savefig('myimage.svg', format='svg', dpi=1200)
```



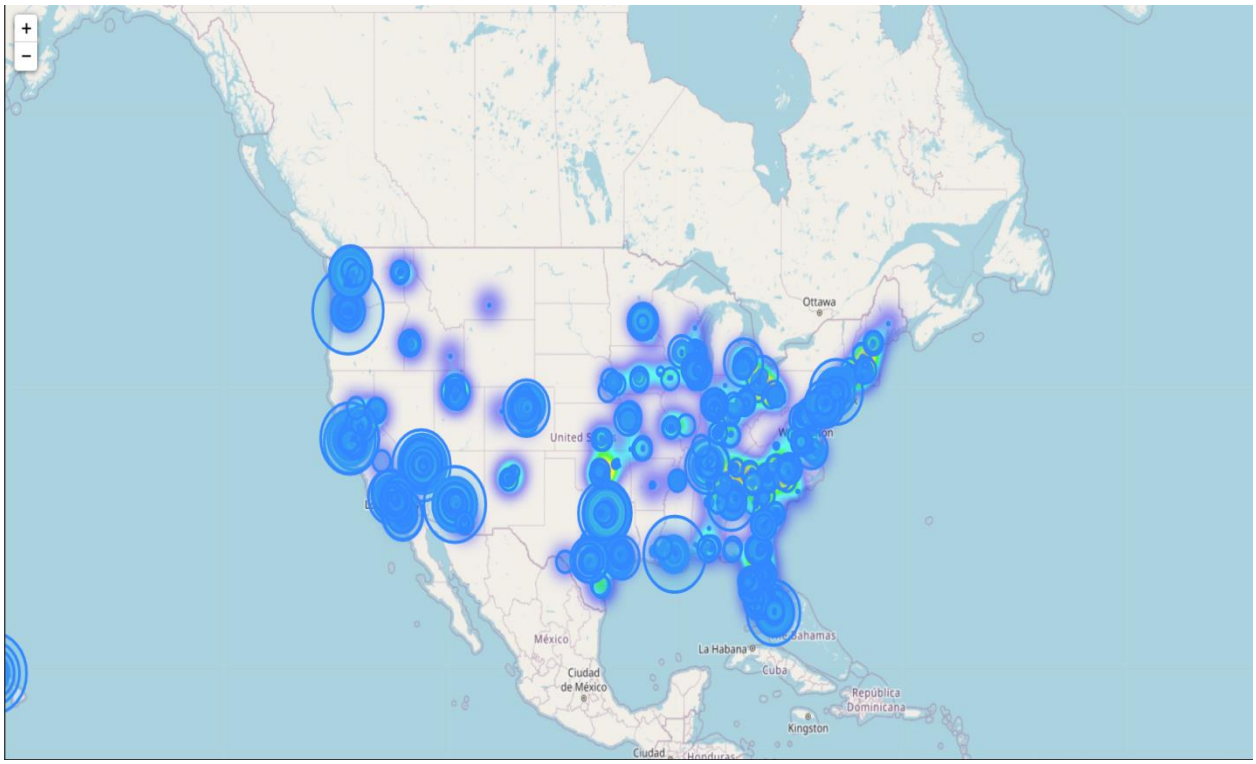
This Bar graph shows the diversity of the data geographically. We can see that we have
the maximum number of data of vehicle Tesla and Toyota; and minimum number of data
for Saab, McLaren and Suzuki.

```
[ ]  labels=dt1['location.state'].value_counts().index
     f, ax = plt.subplots(figsize=(18, 7))
     sns.countplot(x='location.state', data=dt1,
                   order = labels,
                   #hue='vehicle.year'
                   palette="Set2"
                  )
     plt.xticks(rotation= 45,fontsize=12 )
     ax.set_ylabel('Car count', fontsize=15, color='r')
     ax.set_xlabel('location.state', fontsize=14, color='r')
     #plt.savefig('make of the vehicle.png', format='png')
     plt.savefig('Car count per state', format='svg', dpi=1200)
```



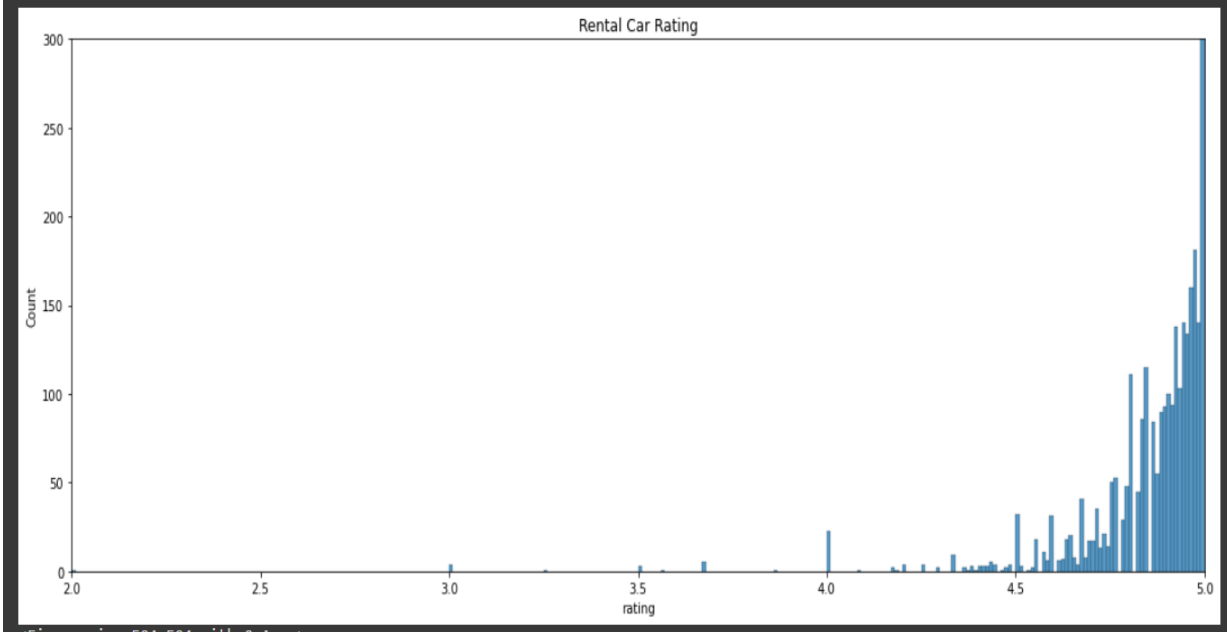
This Bar Chart shows the location (state) where the most of the cars are hire or taken for the rent in the US country. According to the graph we can see that CA state has more number of count than state MT and WW.

```
[ ]  import folium
     from folium.plugins import HeatMap
     center = [35.582889, -99.632773]  #data.describe(mean)
     m = folium.Map([dt1.latitude.mean(), dt1.longitude.mean()], zoom_start=4,center=center)
     for index, row in dt1.iterrows():
         folium.CircleMarker([row['latitude'], row['longitude']],
                             radius=row['renterTripsTaken']/10,
                             fill_color="#3db7e4",
                            ).add_to(m)

     points = dt1[['latitude', 'longitude']].values
     m.add_children(HeatMap(points, radius=15)) # plot heatmap
     m.save('map.html')
     m
```



The Image shoe the longitude and latitude of the country US according to the given dataset.

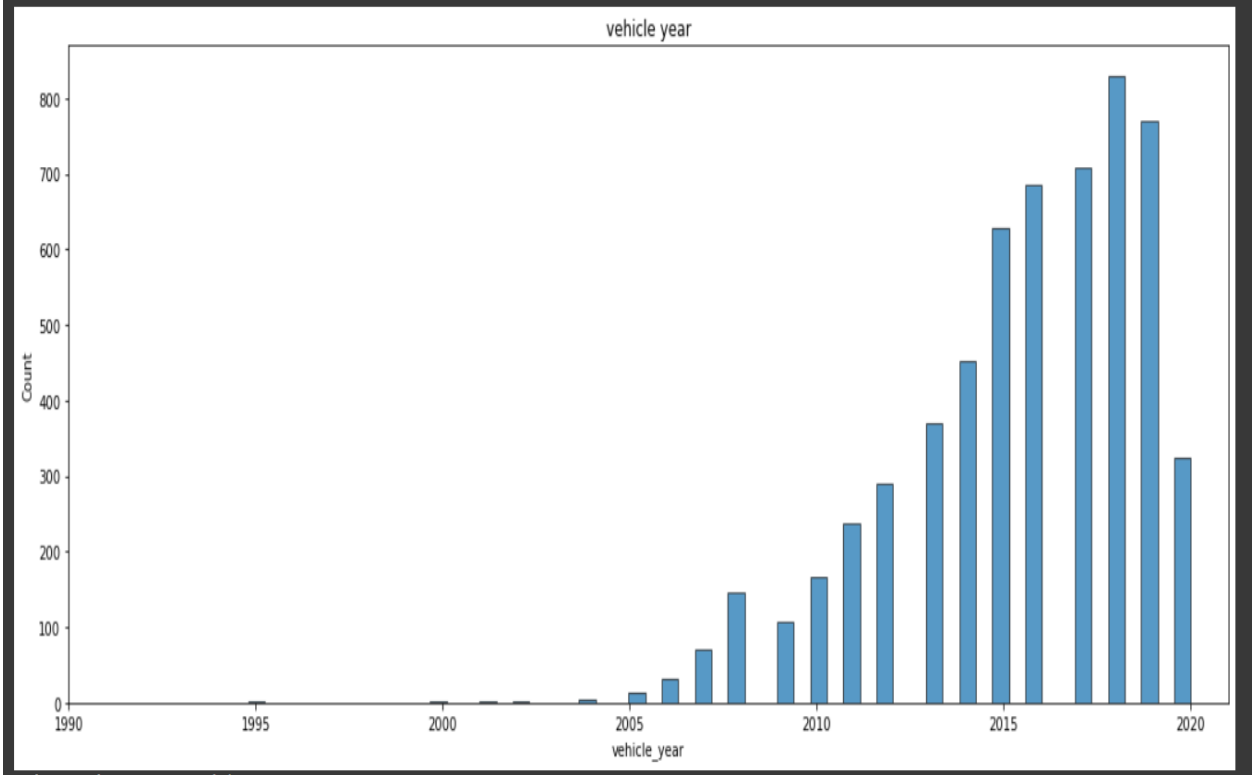
```
f, ax = plt.subplots(figsize=(18, 7))
sns.histplot(data=dt1, x="rating", binwidth=.01)
ax.set_ylim(0,300)
ax.set_xlim(2,5)
plt.title('Rental Car Rating')
plt.show()
plt.savefig('Rental Car Rating.png', format='png')
```




```
<Figure size 504x504 with 0 Axes>
```

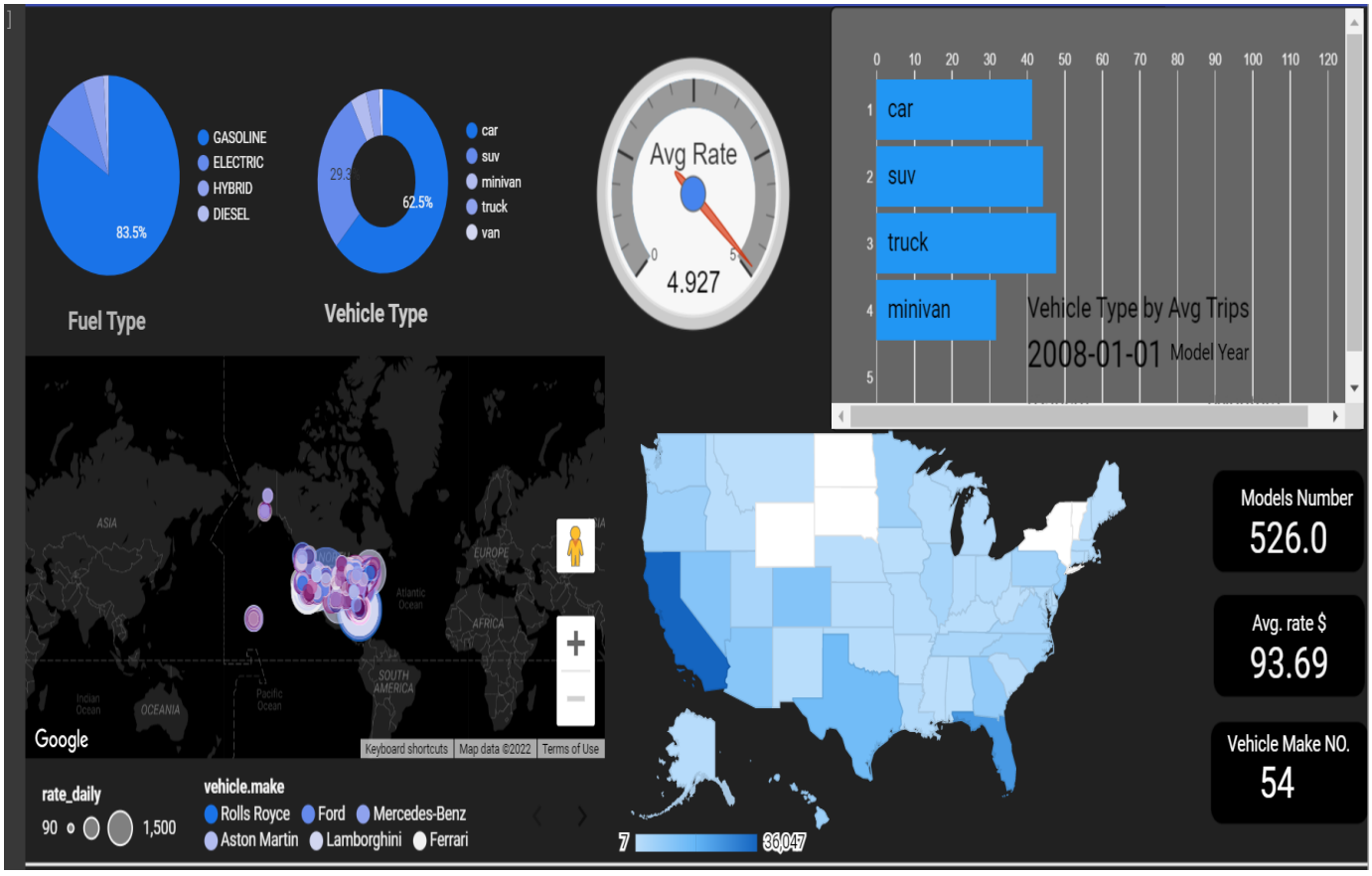
This graph shows the rating given to the particular car according to the given dataset.

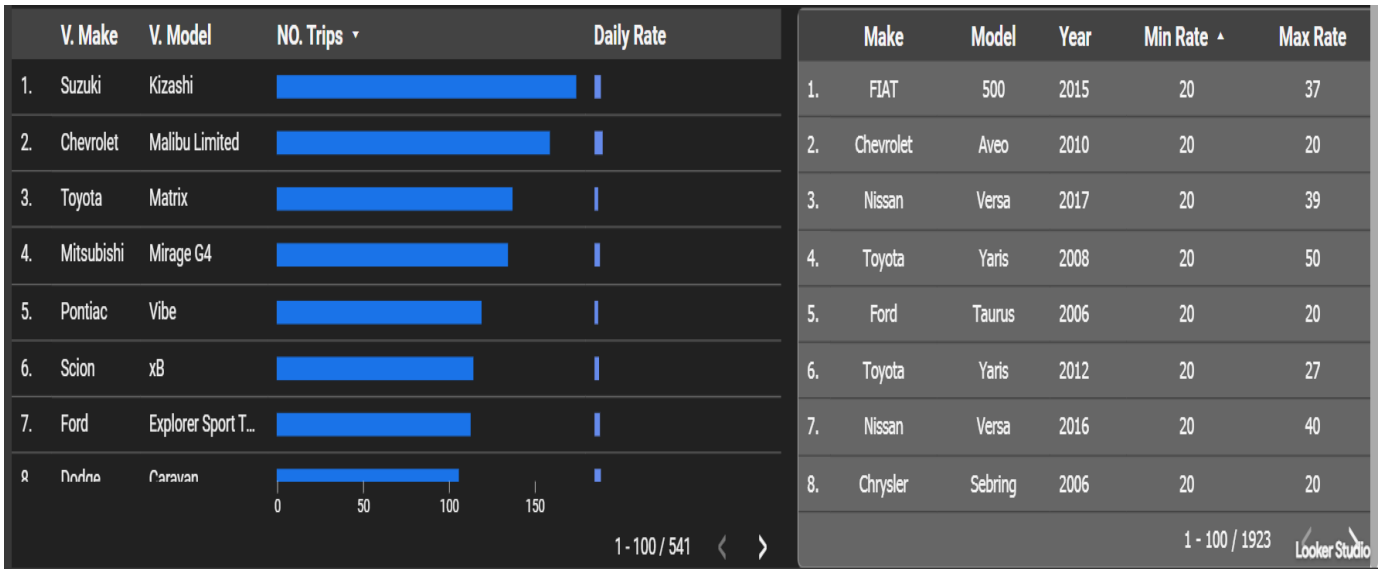
```
f, ax = plt.subplots(figsize=(18, 7))
sns.histplot(data=dt1, x="vehicle_year")
ax.set_xlim(1990,2021)
plt.title('vehicle year')
plt.show()
plt.savefig('vehicle year.png', format='png')
```




```
<Figure size 504x504 with 0 Axes>
```

This graph shows the count of car rental per year. As we can see that from year 2015 to 2020 the count is highest as of year 2005.





	V. Make	V. Model	NO. Trips	Daily Rate
1.	Suzuki	Kizashi		
2.	Chevrolet	Malibu Limited		
3.	Toyota	Matrix		
4.	Mitsubishi	Mirage G4		
5.	Pontiac	Vibe		
6.	Scion	xB		
7.	Ford	Explorer Sport T...		
8.	Dodge	Caravan		

0 50 100 150

1 - 100 / 541

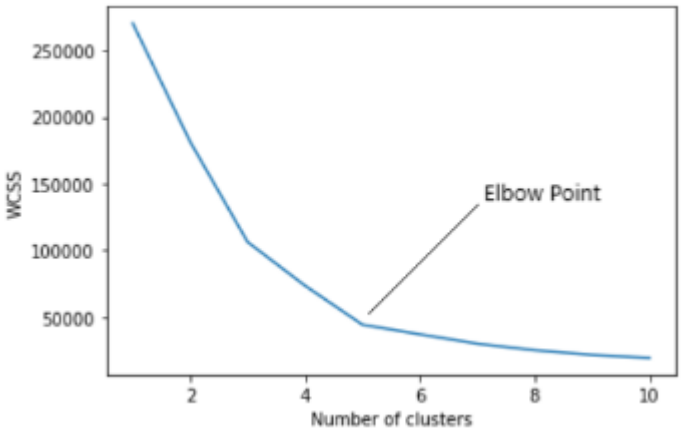
	Make	Model	Year	Min Rate	Max Rate
1.	FIAT	500	2015	20	37
2.	Chevrolet	Aveo	2010	20	20
3.	Nissan	Versa	2017	20	39
4.	Toyota	Yaris	2008	20	50
5.	Ford	Taurus	2006	20	20
6.	Toyota	Yaris	2012	20	27
7.	Nissan	Versa	2016	20	40
8.	Chrysler	Sebring	2006	20	20

1 - 100 / 1923

This shows overall display of the model.

Segment Extraction

K means is one of the most popular Unsupervised Machine Learning Algorithms Used for Solving Classification Problems. K Means segregates the unlabeled data into various groups, called clusters, based on having similar features, common patterns. Suppose we have N number of Unlabeled Multivariate Datasets of various features like water-availability, price, city etc. from our dataset. The technique to segregate Datasets into various groups, on the basis of having similar features and characteristics, is called Clustering. The groups being Formed are known as Clusters. Clustering is being used in Unsupervised Learning Algorithms in Machine Learning as it can segregate multivariate data into various groups, without any supervisor, on the basis of a common pattern hidden inside the datasets.
In the Elbow method, we are actually varying the number of clusters (K) from 1 – 10. For each value of K, we are calculating WCSS (Within-Cluster Sum of Square). WCSS is the sum of squared distance between each point and the centroid in a cluster. When we plot the WCSS with the K value, the plot looks like an Elbow.



As the number of clusters increases, the WCSS value will start to decrease. WCSS value is largest when K = 1. When we analyze the graph we can see that the graph will rapidly change at a point and thus creating an elbow shape. From this point, the graph starts to move almost parallel to the X-axis. The K value corresponding to this point is the optimal K value or an optimal number of clusters.

Analysing Market Segments

Geographic Segmentation: Geographic segmentation divides a target market by location so marketers can better serve customers in a particular area. This type of market segmentation is based on the geographic units themselves (countries, states, cities, etc.), but also on various geographic factors, such as climate, cultural preferences, populations, and more. Geographic segmentation involves segmenting your audience based on the region they live or work in. This can be done in any number of ways: grouping customers by the country they live in, or smaller geographical divisions, from region to city, and right down to postal code.
Geographic segmentation might be the simplest form of market segmentation to get

your head around, but there are still plenty of ways it can be used that companies never think about. The size of the area you target should change depending on your needs as a business. Generally speaking, the larger the business the bigger the areas you'll be targeting. After all, with a wider potential audience, targeting each postcode individually simply won't be cost-effective.

Demographic Segmentation: Demographic segmentation is a market segmentation technique where an organization's target market is segmented based on demographic variables such as age, gender, education, income, etc. It helps organizations understand who their customers are so that their needs can be addressed more effectively.

Instead of reaching an entire market, companies can use demographic segmentation to focus their time and resources on those segments that have customers who are most likely to make purchases, and are therefore most valuable to them.

There are several different variables by which demographic segmentation is done:

a. Age: Age is one of the most important variables used within demographic segmentation as consumers' preferences and needs differ significantly based on the age group they fall under. When an organization wants to target young adults or teenagers, digital marketing campaigns may prove to be most effective as they appeal to this age group. However, older adults often prefer traditional marketing methods, such as television and magazine advertisements.

b. Income: Income levels have a significant effect on consumer purchasing decisions. Those with higher-income levels may prefer high-end and luxury products. Conversely, individuals with lower income levels may prefer to get products at the best deal and are likely to choose inexpensive products/services.

c. Gender: Individuals may identify with different areas of the gender spectrum, like feminine or masculine, and this will have a significant effect on their preferences and purchasing decisions. By understanding which gender your product or service appeals to, you can tailor your marketing campaigns accordingly to meet the needs of your consumers better.

Psychographic Segmentation : Psychographic segmentation is the research methodology used for studying consumers and dividing them into groups using psychological characteristics including personality, lifestyle, social status, activities, interests, opinions, and attitudes.

Psychographic marketing enables you to engage with multiple target audiences in the ways that will make the biggest impact for each one. This approach saves time and money on approaches that might fall flat and makes it easier to relate to the groups you care about.

We can use psychographics for market segmentation to understand:

a. How consumers really perceive your products and services
b. What consumers really want—and why
c. Gaps or pain points with your current products or services
d. Opportunities for future engagement

e. How to better communicate with your target audience.

Behavioral Segmentation: Behavioral segmentation refers to a process in marketing which divides customers into segments depending on their behavior patterns when interacting with a particular business or website.

These segments could include grouping customers by:
a. Their attitude toward your product, brand or service;
b. Their use of your product or service,
c. Their overall knowledge of your brand and your brand's products,
d. Their purchasing tendencies, such as buying on special occasions like birthdays or holidays only, etc.

Behavioral segmentation offers marketers and business owners a more complete understanding of their audience, thus enabling them to tailor products or services to specific customer needs.

Customizing the Market Mix

The marketing mix refers to the set of actions, or tactics, that a company uses to promote its brand or product in the market. The 4Ps make up a typical marketing mix - Price, Product, Promotion and Place.

a. Price: refers to the value that is put for a product. It depends on costs of production, segment targeted, ability of the market to pay, supply - demand and a host of other direct and indirect factors. There can be several types of pricing strategies, each tied in with an overall business plan
b. Product: refers to the item actually being sold. The product must deliver a minimum level of performance; otherwise even the best work on the other elements of the marketing mix won't do any good.
c. Place: refers to the point of sale. In every industry, catching the eye of the consumer and making it easy for her to buy it is the main aim of a good distribution or 'place' strategy. Retailers pay a premium for the right location. In fact, the mantra of a successful retail business is 'location, location, location'.
d. Promotion: this refers to all the activities undertaken to make the product or service known to the user and trade. This can include advertising, word of mouth, press reports, incentives, commissions and awards to the trade. It can also include consumer schemes, direct marketing, contests and prizes.

All the elements of the marketing mix influence each other. They make up the business plan for a company and handle it right, and can give it great success. The marketing mix needs a lot of understanding, market research and consultation with several people, from users to trade to manufacturing and several others.