ONLINE-VEHICLE BOOKING MARKET ANALYSIS

11/06/2023

Team

Pawan Kumar (Project Lead)

Nanditha Ranjit Nambiar

Snehashish Mahapatra

Aditya Grover

Aman Singh

Problem statement

The problem at hand is to conduct a thorough market analysis of the online vehicle booking sector, incorporating the dynamic variables of location, season, weather, demographics etc. By understanding the interplay of these factors, we aim to identify key trends, patterns, and preferences that influence customer behaviour and decision-making.

Github links

1.	https://github.com/SNEHASHISHMAHAPATRA1993/Uber-Ride-Analysis/tree/main
2.	https://github.com/nanditha48/FeynnLab_OBS
3.	https://github.com/Aditya7-grover/FeynnLabs/tree/main
4.	https://github.com/Pawan12345kr/Feynn_labs_2nd_project
5.	https://github.com/na20b005/online-booking-data-analyse

Abstract

Cab booking is a common kind of transportation provided by several transportation companies in each city. Most individuals rely on cab services for their everyday transportation requirements. The firm must be registered and meet all the transportation department's criteria and security standards. The Online Cab Booking System is a web-based platform that enables the clients to book taxis and executive taxis from the convenience of their own home or workplace. The platform should have an administrative interface via which the taxi business can control the content as well as access all reservations and customer data. The analysis delves into the various segments within the online vehicle booking market, including ride-hailing services, car-sharing platforms, and peer-to-peer car rental platforms. It evaluates the factors influencing consumer behaviour and preferences, such as convenience, cost-effectiveness, safety concerns, and environmental sustainability. The study begins by examining the market size and growth rate of the online vehicle booking industry, considering factors such as increasing urbanization, rising disposable incomes, and the growing adoption of smartphones and internet connectivity. Furthermore, it explores the competitive landscape, identifying major players

and their market shares, as well as emerging startups disrupting the traditional vehicle rental market.

Online Ride-Hailing: Market Overview

The online cab booking apps have transformed the fleet industry as it is one of the best technological advancements that great ideas and vision have offered. No matter where you are going, taxi booking apps got it all covered.

The online ride-hailing market is expected to grow at a CAGR of 10.08% during the period 2020-2025. Low taxi fare and ease of booking through websites and apps are the major factors that have captured online taxi booking segments. Especially if we talk about urban people, they believe that it would be easy for them to book a cab online rather than waiting at the public sport for vehicles.

With the advance and famous ride-hailing applications such as Uber. Lyft, Didi, etc. have been one of the game-changing innovations of the fleet sector.

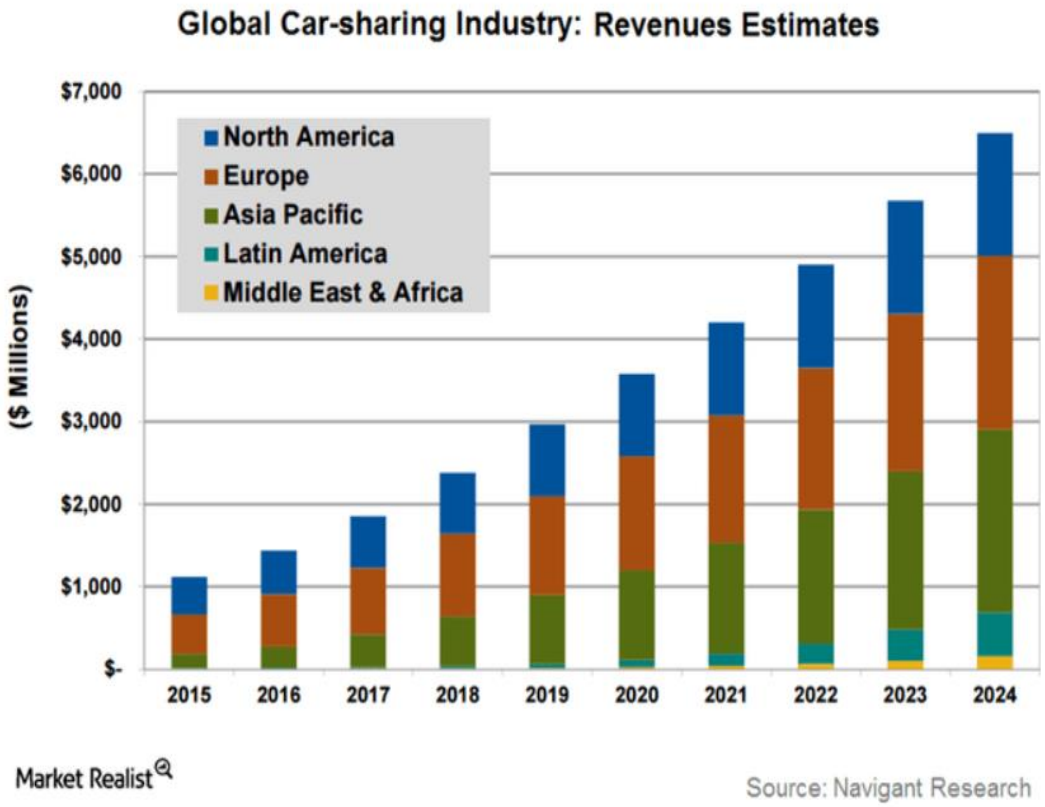


Fig.1

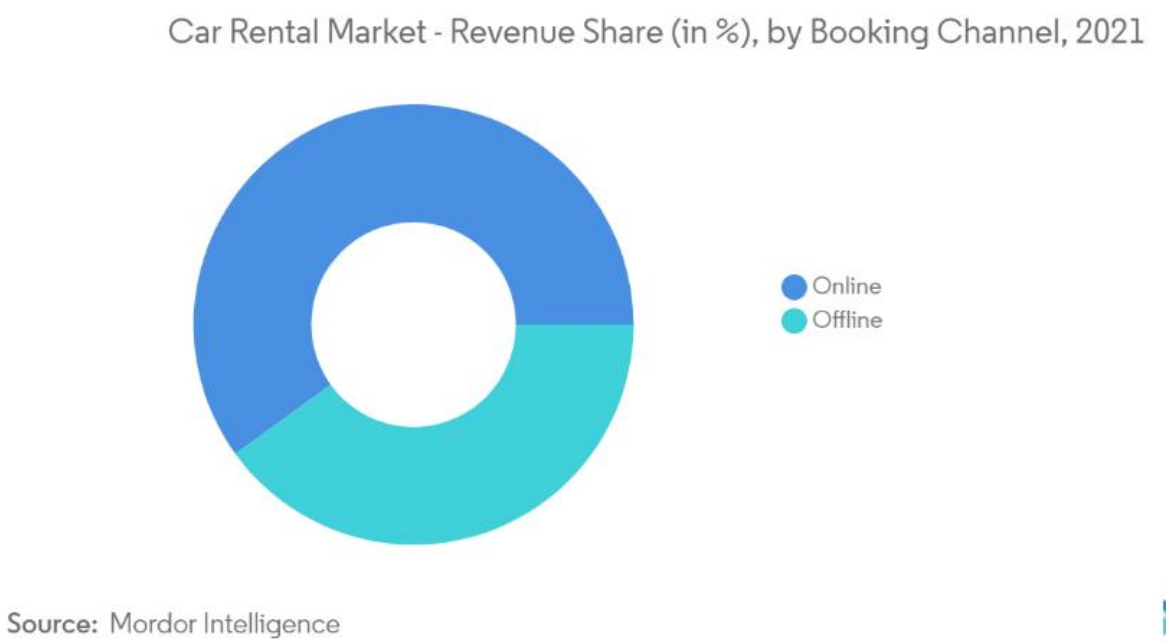


Fig.2

Fig.2 shows the comparison between online and offline car booking. It is clearly observed that online booking is more than the other. From fig.1 Market share of cab sharing has been increasing since almost a decade.

About the data

We have gathered data manually from different sources. And each analysis is based on the dataset which you can find on the above GitHub links. Each of us has worked on different dataset so there will be vast analysis of this online vehicle booking market.

Data Preprocessing and analysis

Data pre-processing refers to the steps and techniques applied to raw data before it can be used for analysis or machine learning tasks. It involves transforming and cleaning the data to ensure its quality, consistency, and suitability for further processing.

Analysis 1

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

Importing Libraries The analysis will be done using the following libraries :

Pandas: This library helps to load the data frame in a 2D array format and has multiple functions to perform analysis tasks in one go

Numpy: Numpy arrays are very fast and can perform large computations in a very short time.

Matplotlib / Seaborn: This library is used to draw visualizations. To importing all these libraries, we can use the below code :

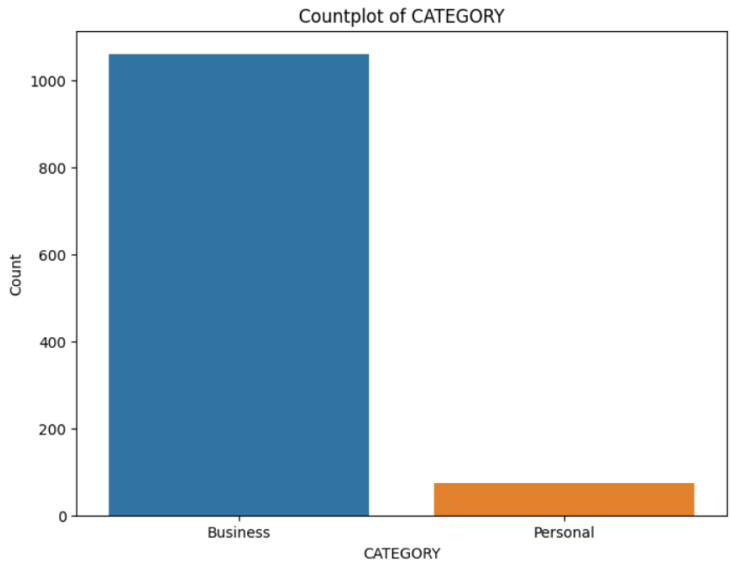
Fig.3

```
dataset = pd.read_csv("UberDataset.csv")
dataset.head()
```

	START_DATE	END_DATE	CATEGORY	START	STOP	MILES	PURPOSE
0	01-01-2016 21:11	01-01-2016 21:17	Business	Fort Pierce	Fort Pierce	5.1	Meal/Entertain
1	01-02-2016 01:25	01-02-2016 01:37	Business	Fort Pierce	Fort Pierce	5.0	NaN
2	01-02-2016 20:25	01-02-2016 20:38	Business	Fort Pierce	Fort Pierce	4.8	Errand/Supplies
3	01-05-2016 17:31	01-05-2016 17:45	Business	Fort Pierce	Fort Pierce	4.7	Meeting
4	01-06-2016 14:42	01-06-2016 15:49	Business	Fort Pierce	West Palm Beach	63.7	Customer Visit

To find the shape of the dataset, we can use dataset.shape

Fig.4



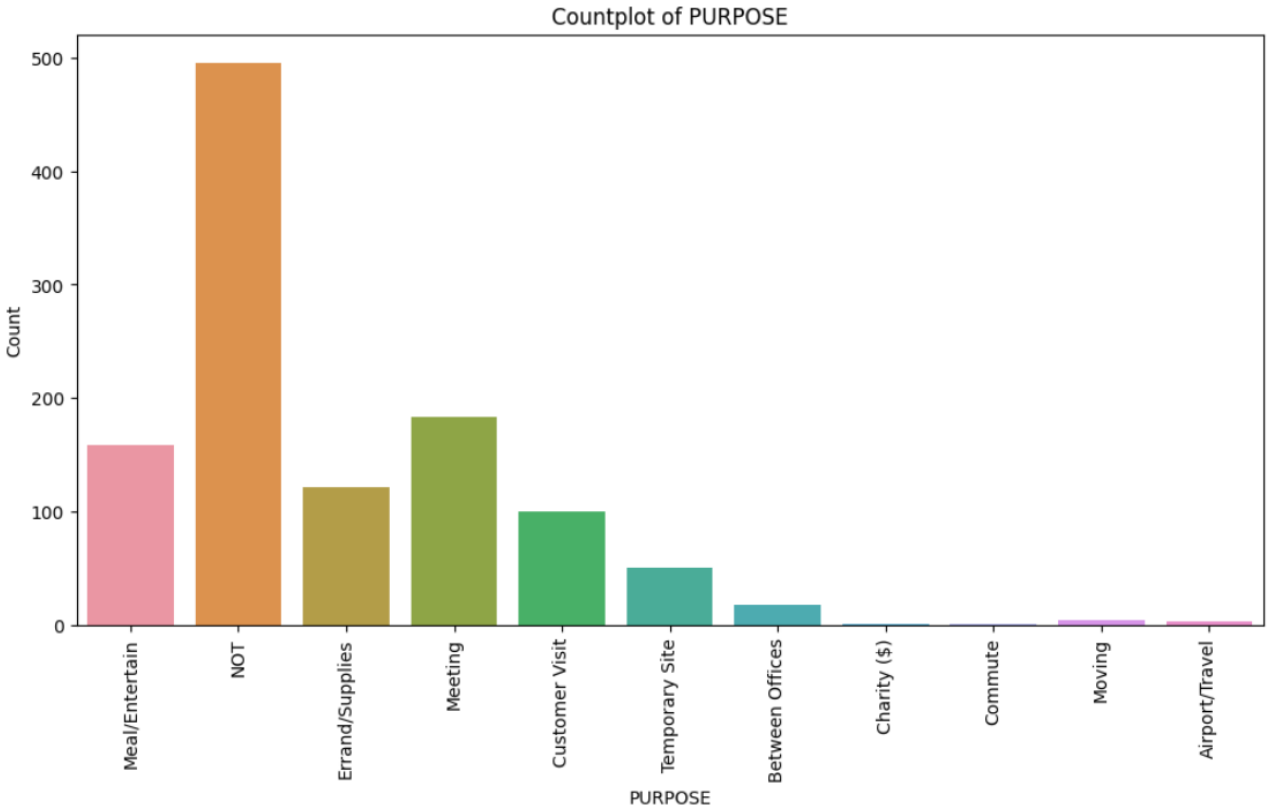

Countplot of PURPOSE
Count
500
400
300
200
100
0
Meal/Entertain
NOT
Errand/Supplies
Meeting
Customer Visit
Temporary Site
Between Offices
Charity ($)
Commute
Moving
Airport/Travel
PURPOSE

Fig.5

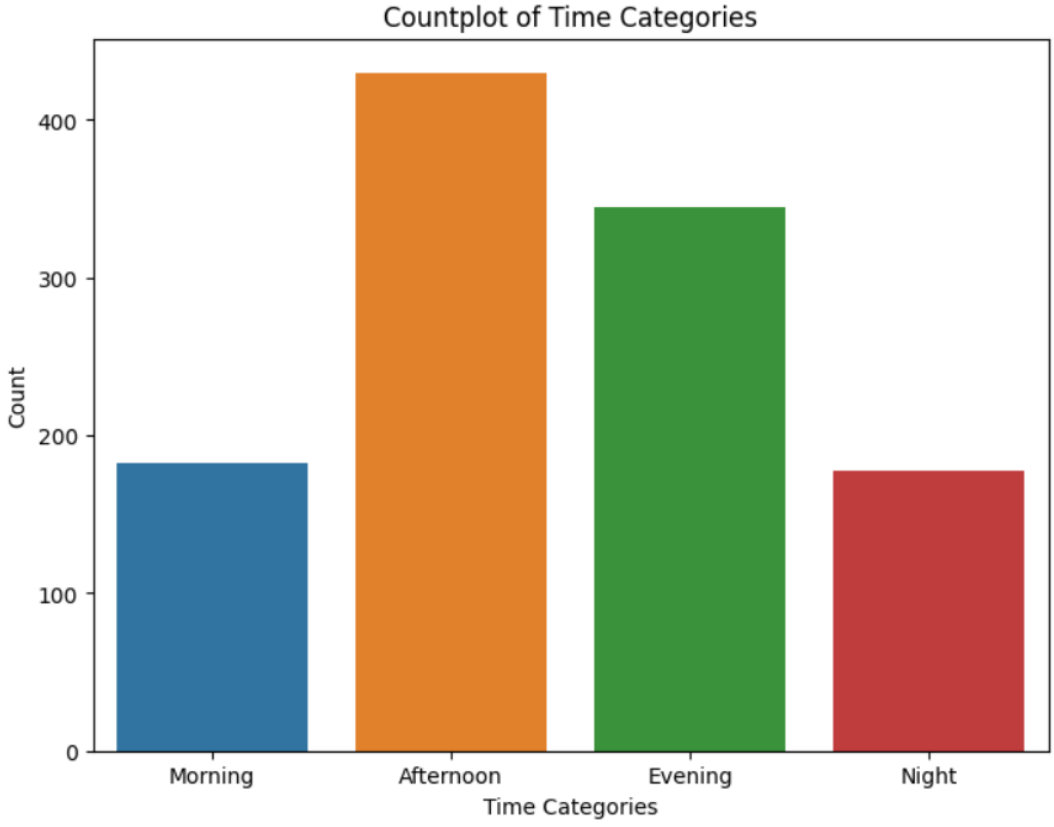

Countplot of Time Categories
Count
400
300
200
100
0
Morning
Afternoon
Evening
Night
Time Categories

Fig.6

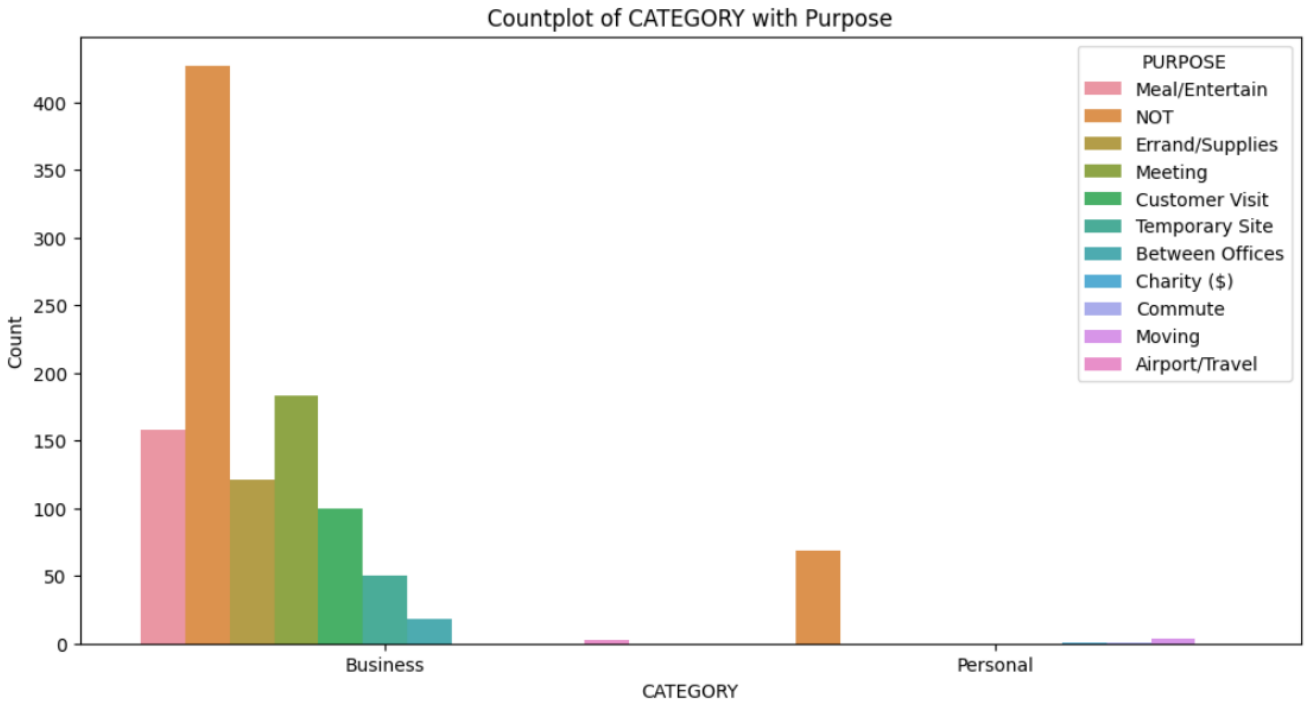


Fig.7

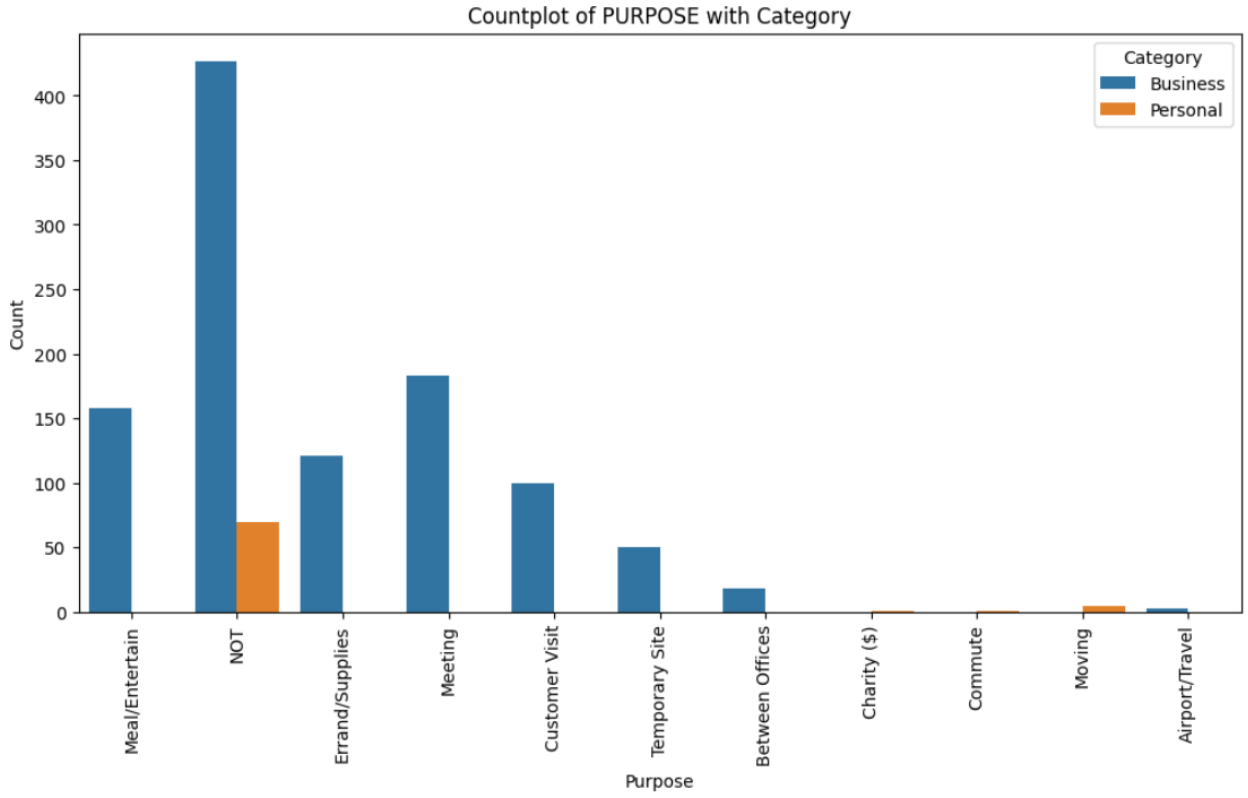


Fig.8

Insights from the above count-plots:

Most of the rides are booked for business purpose.

Most of the people book cabs for Meetings and Meal / Entertain purpose.

Most of the cabs are booked in the time duration of 10am-5pm (Afternoon).

As we have seen that CATEGORY and PURPOSE columns are two very important columns. So now we will be using OneHotEncoder to categories them

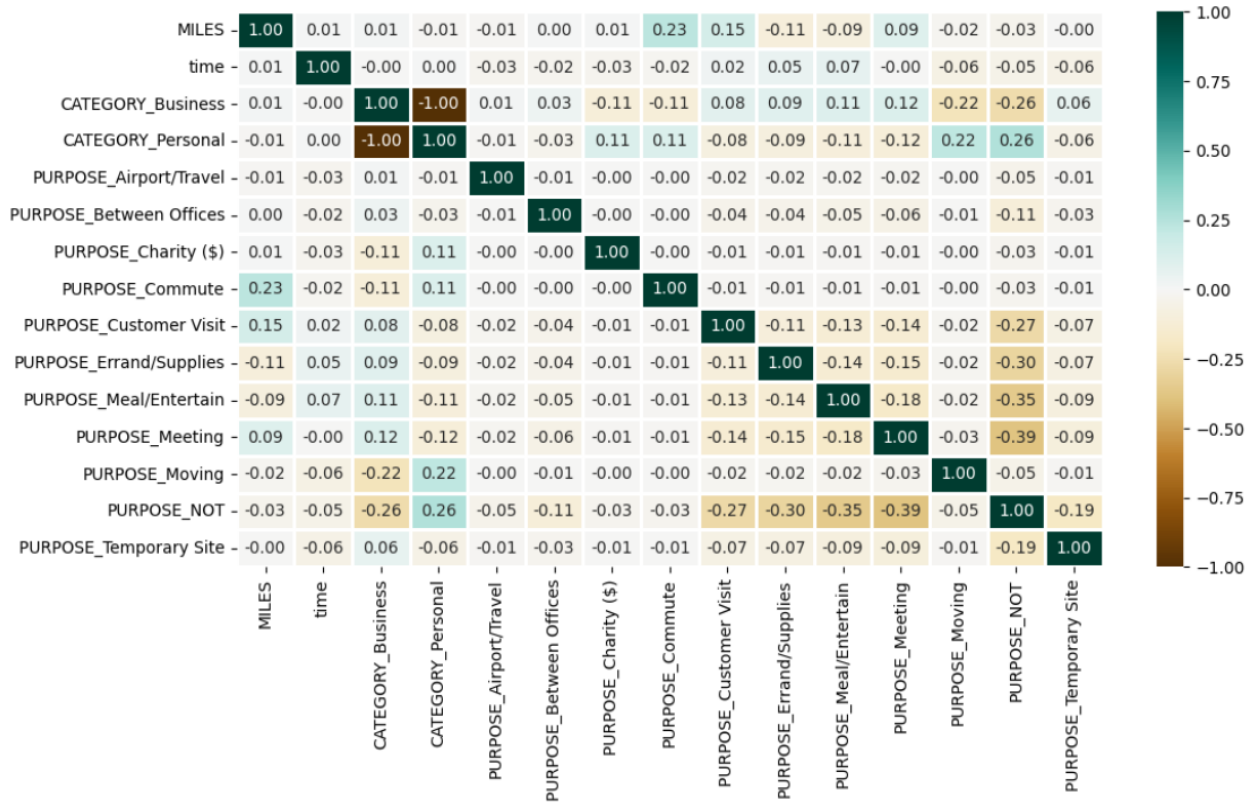


Fig.9

Business and Personal Category are highly negatively correlated, this have already proven earlier. So, this plot, justifies the above conclusions. There is not much correlation between the features.

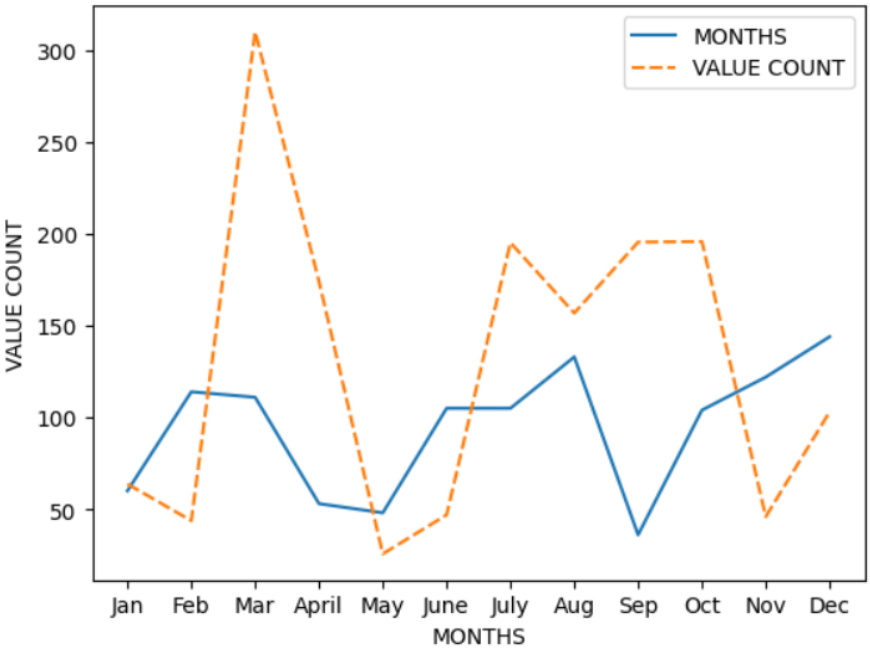


Fig.10

The counts are very irregular.

Still it's very clear that the counts are very less during Nov, Dec, Jan, which justifies the fact that time winters are there in Florida, US.

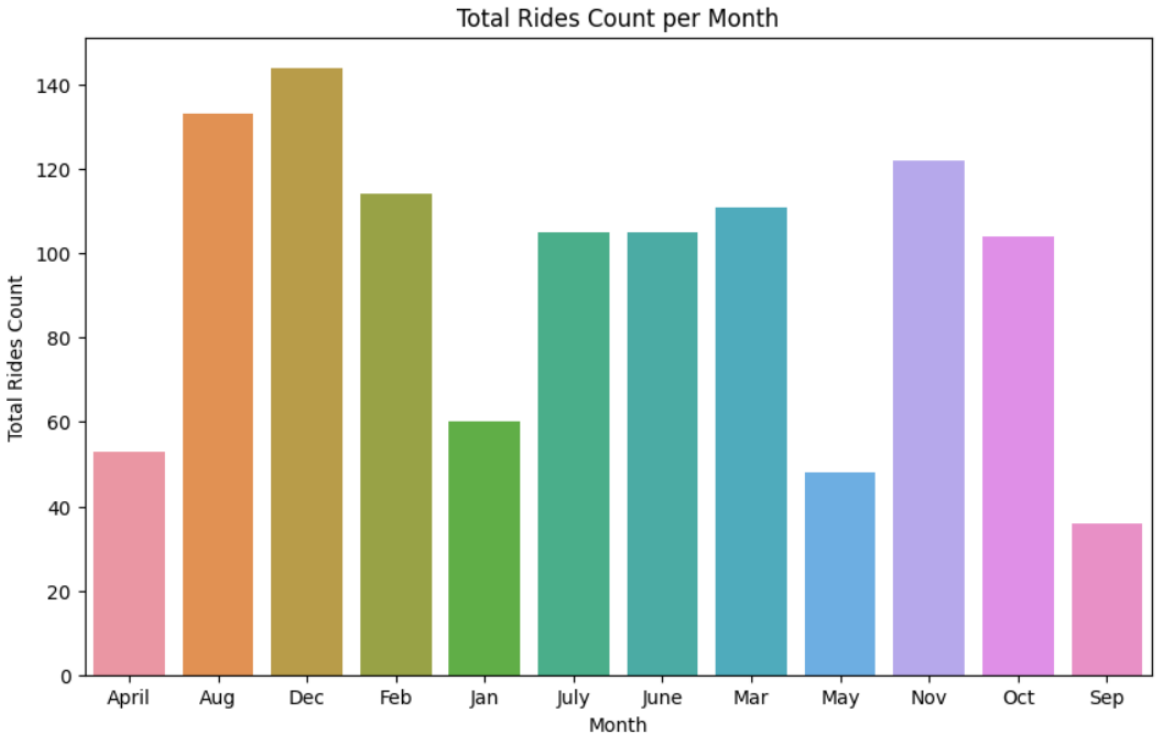


Fig.11

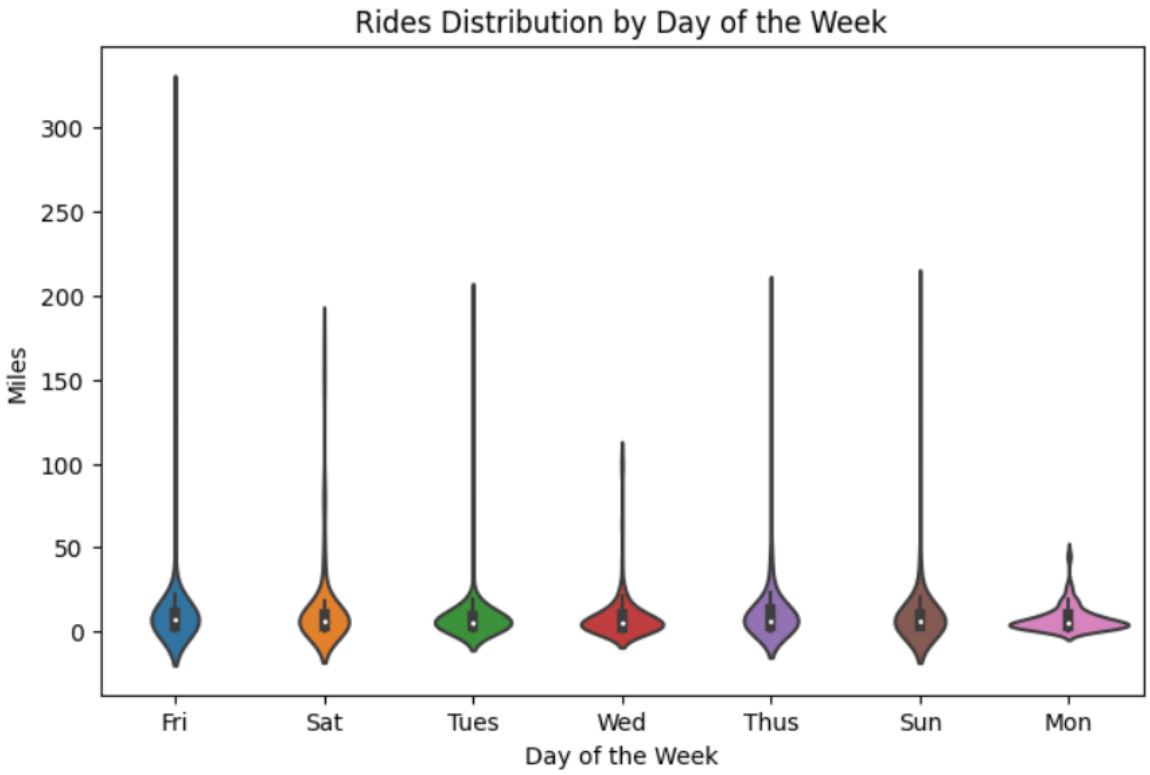


Fig.12

Analysis 2

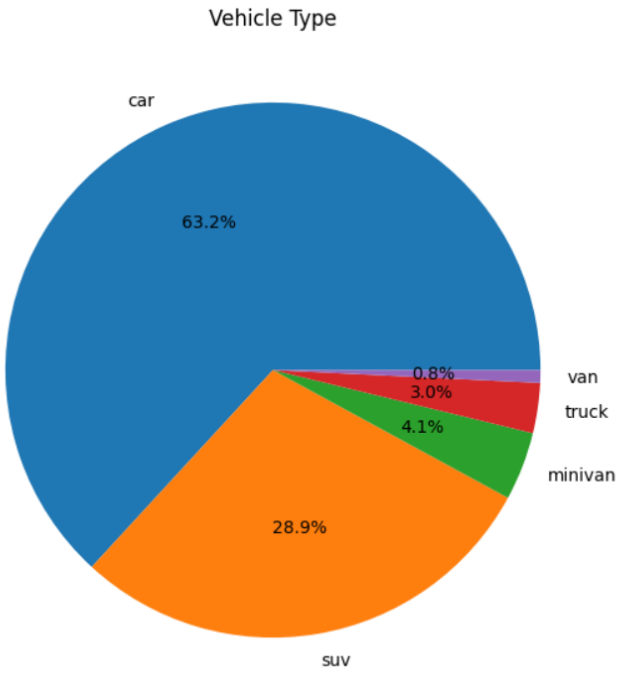


Fig.13

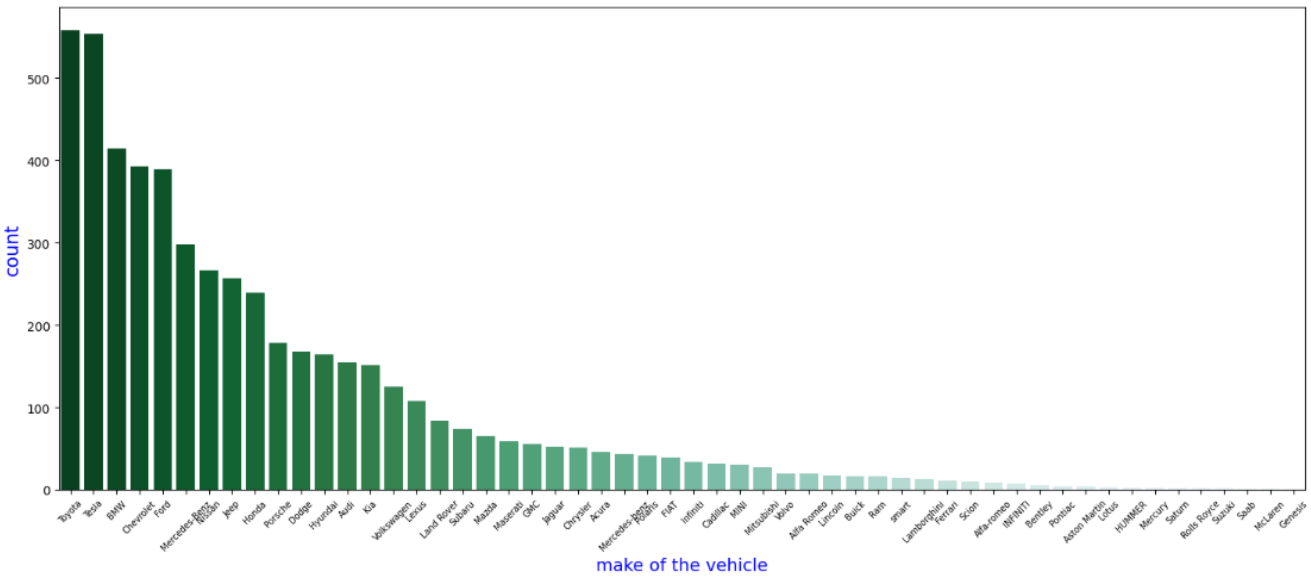


Fig.14

It is clearly visible that Toyota is mostly preferred by people.

Analysis 3

```
train
```

	datetime	season	holiday	workingday	weather	temp	atemp	humidity	windspeed	Total_booking
0	5/02/2012 19:00	Summer	0	1	Clear + Few clouds	22.14	25.760	77	16.9979	504
1	9/05/2012 4:00	Fall	0	1	Clear + Few clouds	28.70	33.335	79	19.0012	5
2	1/13/2011 9:00	Spring	0	1	Clear + Few clouds	5.74	6.060	50	22.0028	139
3	11/18/2011 16:00	Winter	0	1	Clear + Few clouds	13.94	16.665	29	8.9981	209
4	9/13/2011 13:00	Fall	0	1	Clear + Few clouds	30.34	33.335	51	19.0012	184
...	...	...	...	...	...	...	...	...	...	...
8703	1/16/2012 6:00	Spring	1	0	Clear + Few clouds	4.10	6.820	54	6.0032	13
8704	11/10/2011 1:00	Winter	0	1	Mist + Cloudy	16.40	20.455	87	0.0000	11
8705	4/12/2011 3:00	Summer	0	1	Mist + Cloudy	23.78	27.275	56	8.9981	1
8706	11/07/2012 1:00	Winter	0	1	Mist + Cloudy	11.48	13.635	61	16.9979	92
8707	1/10/2011 10:00	Spring	0	1	Mist + Cloudy	5.74	6.060	50	19.9995	31

8708 rows × 10 columns

Fig.15

```
plt.figure(figsize=(10,5))
sns.barplot(x = 'weather', y = 'Total_booking', data = train,palette='rainbow')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x475c596bc8>
```

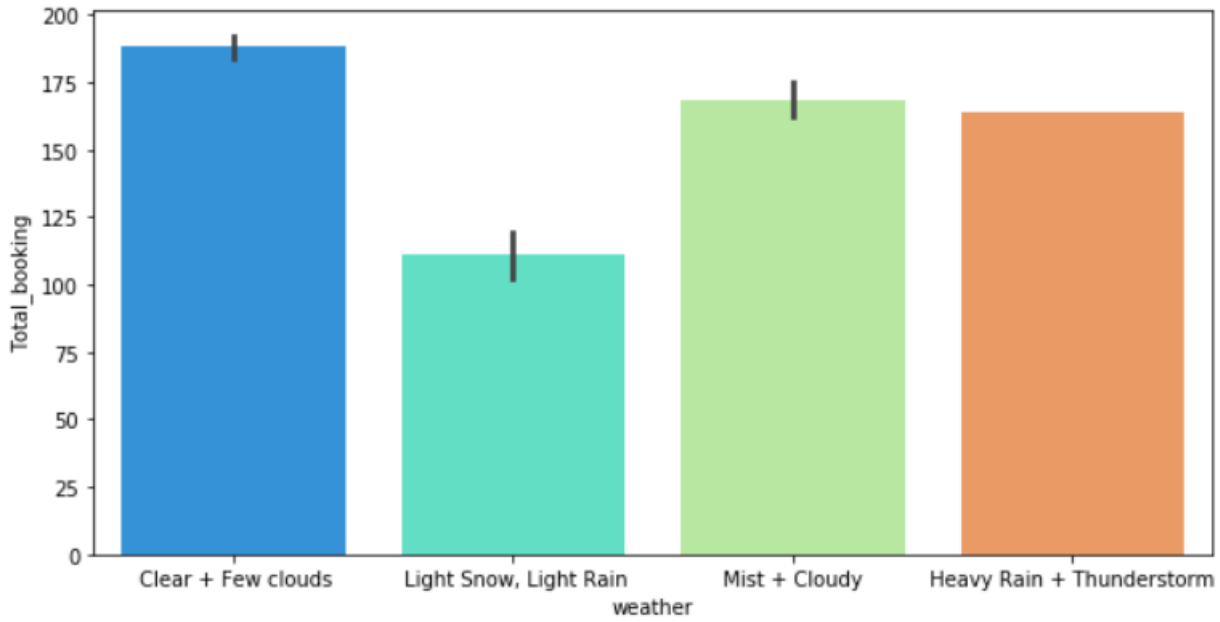


Fig.16

Above figure shows the weather in which corresponding numbers of cabs are booked.

```
plt.figure(figsize=(10,5))
sns.barplot(x = 'weekday', y = 'Total_booking', data = train)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x475c57fc48>
```

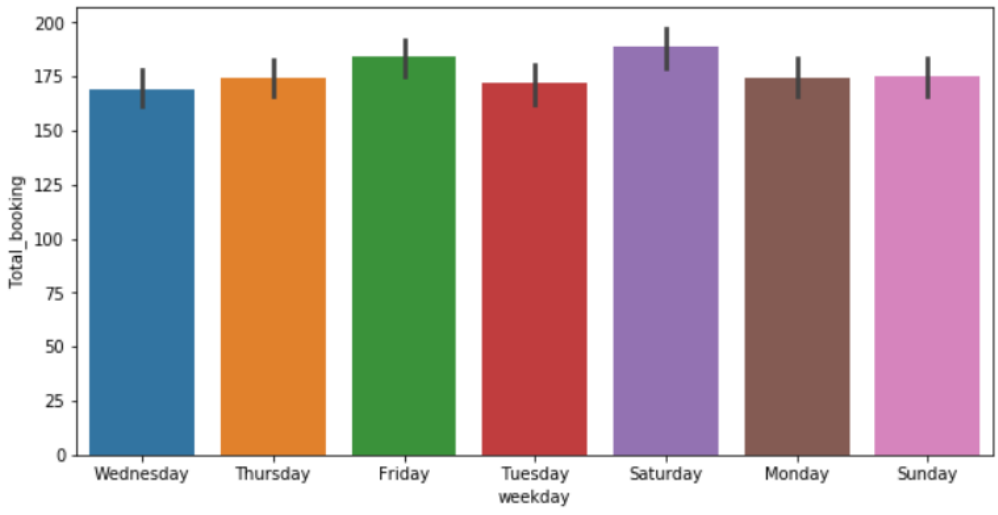


Fig.17

Fig.17 shows that Saturday witnesses the most bookings.

```
plt.figure(figsize=(10,5))
sns.barplot(x = 'season', y = 'Total_booking', data = train)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x475c677f88>
```

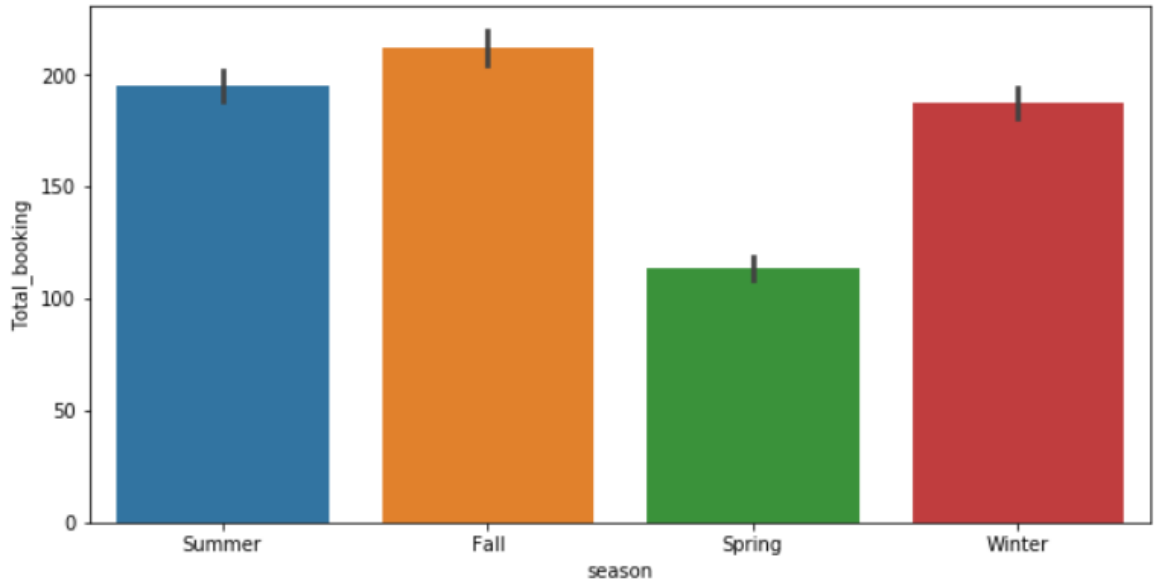


Fig.18

Analysis 4

```
df.head()
```

	VendorID	lpep_pickup_datetime	lpep_dropoff_datetime	store_and_fwd_flag	RatecodeID	PULocationID	DOLocationID	passenger_c
0	1.0	2021-07-01 00:30:52	2021-07-01 00:35:36	N	1.0	74	168	
1	2.0	2021-07-01 00:25:36	2021-07-01 01:01:31	N	1.0	116	265	
2	2.0	2021-07-01 00:05:58	2021-07-01 00:12:00	N	1.0	97	33	
3	2.0	2021-07-01 00:41:40	2021-07-01 00:47:23	N	1.0	74	42	
4	2.0	2021-07-01 00:51:32	2021-07-01 00:58:46	N	1.0	42	244	

Fig.19

Which is the busiest day ?

```
df['day_of_week'] = df['lpep_pickup_datetime'].dt.day_name()
```

```
plt.figure(figsize= (18, 7))
sns.countplot(y= 'day_of_week', data= df)
plt.ylabel('');
```

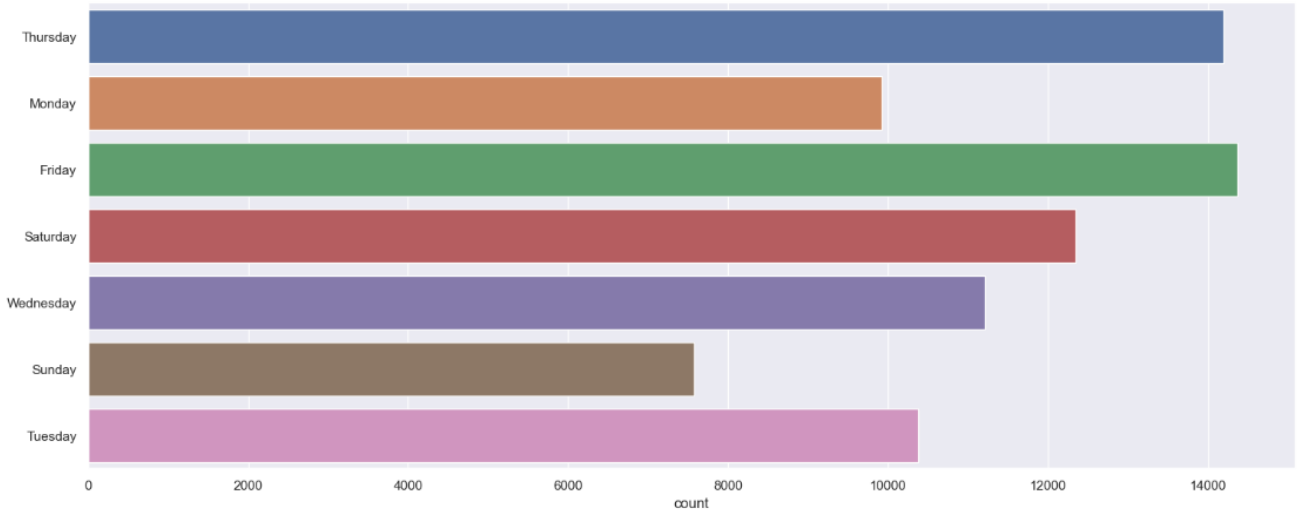


Fig.21

With respect to the data in **Analysis 4**, Friday seems to be the busiest day.

Which is the busiest hour ?

```
plt.figure(figsize= (18, 7))
sns.countplot(x= df['lpep_pickup_datetime'].dt.hour, data= df, color= 'goldenrod')
plt.ylabel('')
plt.xlabel('Hour of Day');
```

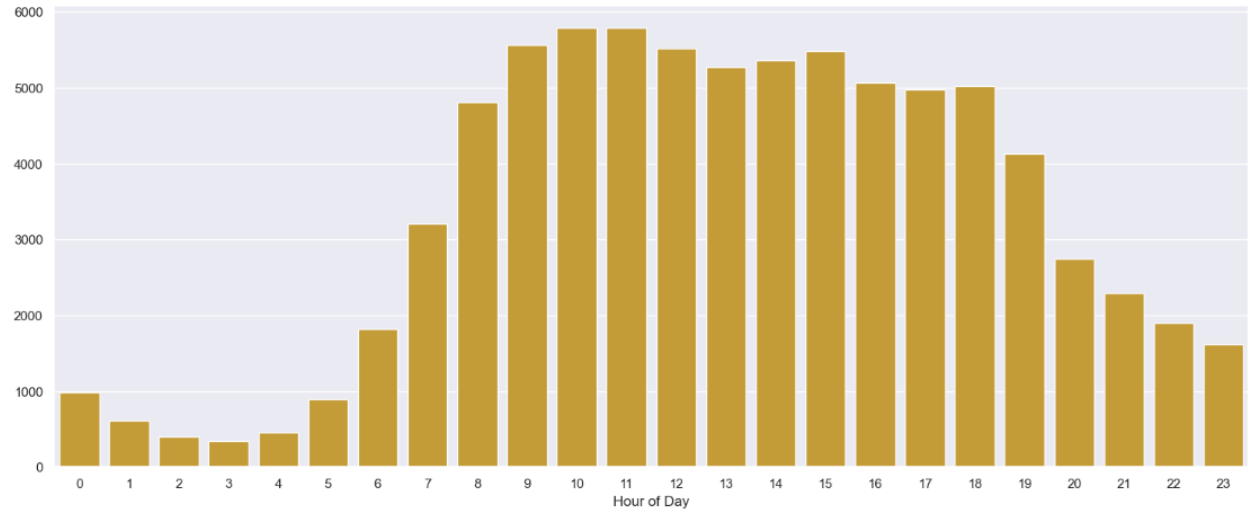


Fig.22

Most bookings are done from 8 AM to 6 PM.

Analysis 5

Exploratory Data Analysis:

Exploratory data analysis, or EDA, is a detailed analysis intended to reveal a data set's underlying structure. It is significant for a business because it reveals trends, patterns, and linkages that are not immediately obvious.

```
[ ]  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()
```

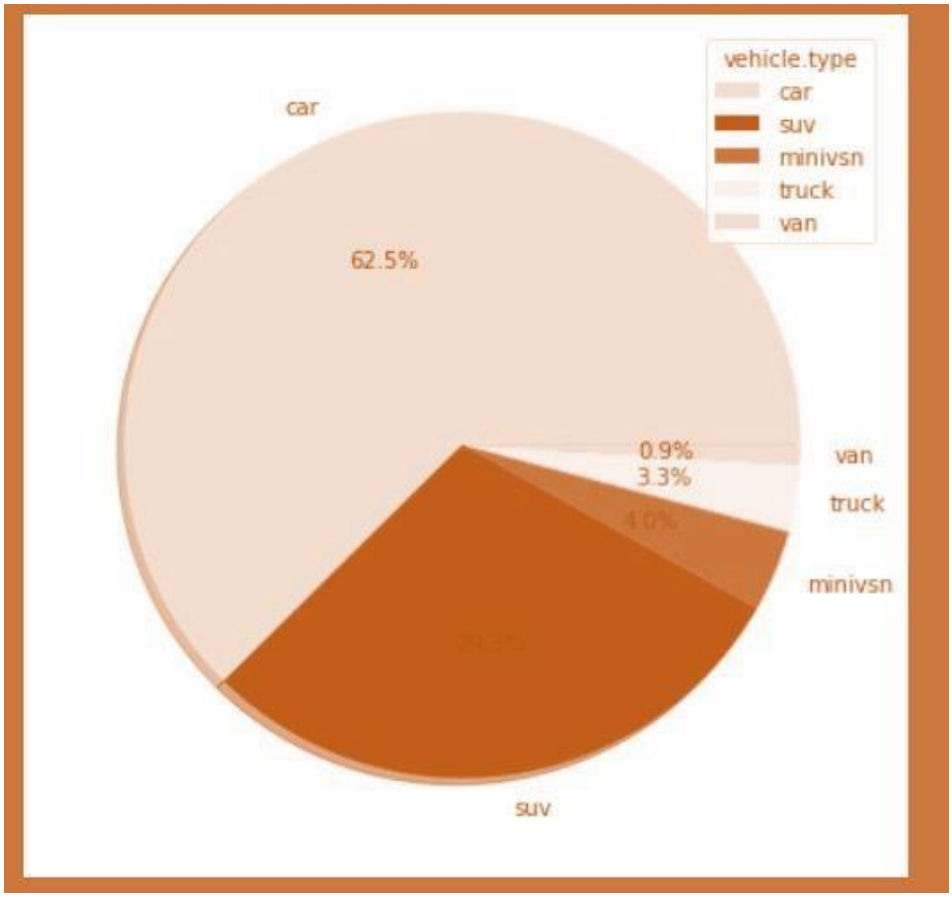
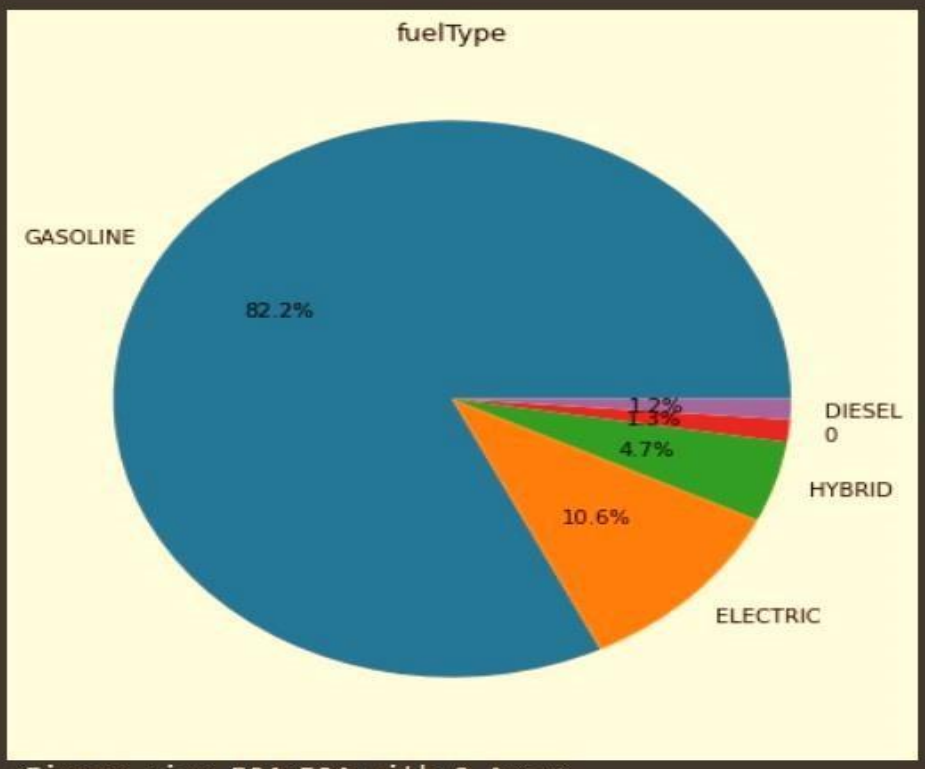


Fig.23

The classification of vehicle type is displayed here. As we can see, cars are the most popular variety, followed by SUVs, minivans, trucks, and vans.

```
[ ]  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>
```

Fig.24

This demonstrates how the fuel types are categorised. As we can see, gasoline is the most common type of fuel.

```
[ ] 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)
```

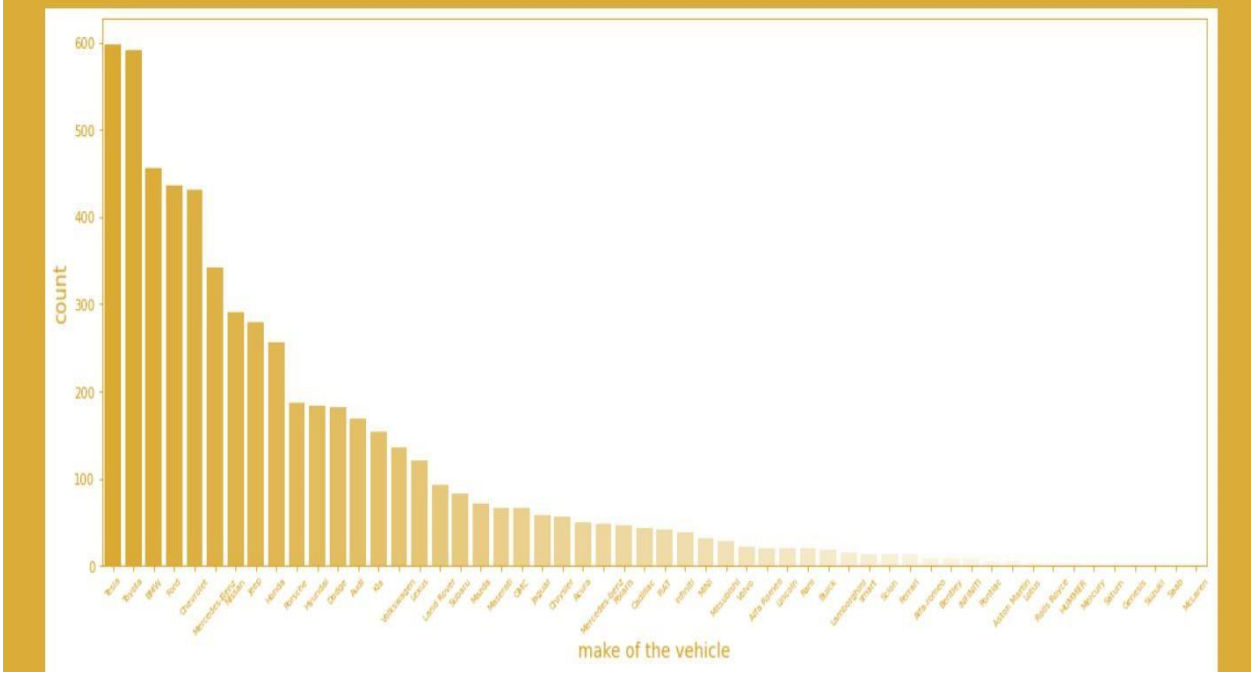


Fig.25

The geographic diversity of the data is depicted in this bar graph. We can observe that the number of data for Tesla and Toyota is at its maximum, while Saab, McLaren, and Suzuki have the lowest number of data.

```
[ ]  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)
```

Fig.26

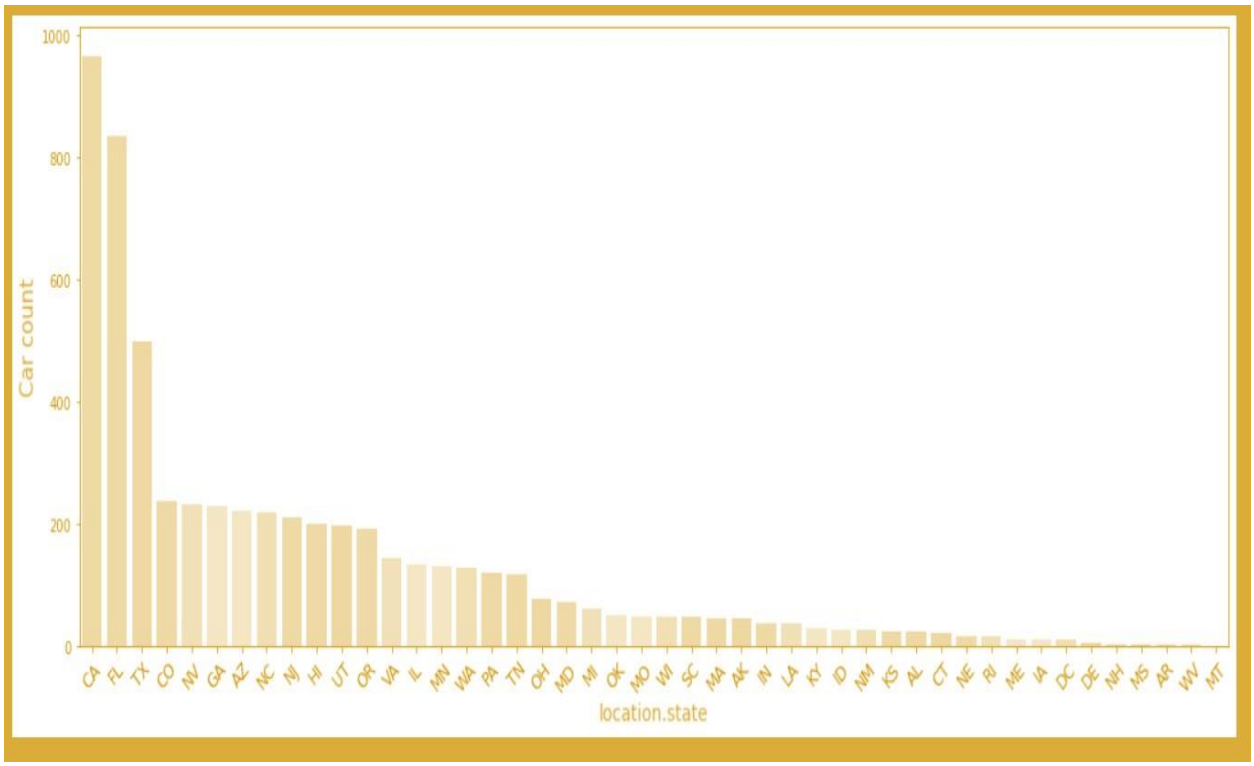


Fig.27

The location (state) where the majority of cars are rented or leased in the US is depicted in this bar chart. The graph shows that the state of CA has a higher count than the states of 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
```

Fig.28

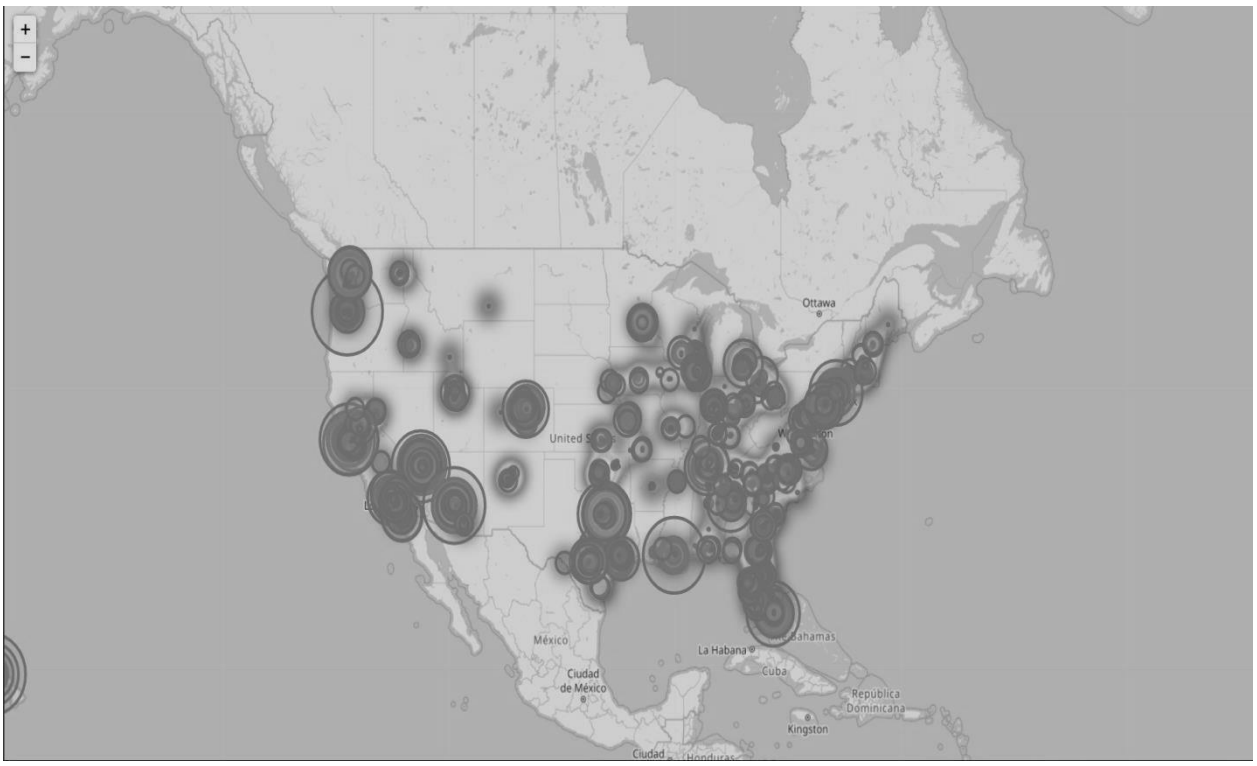
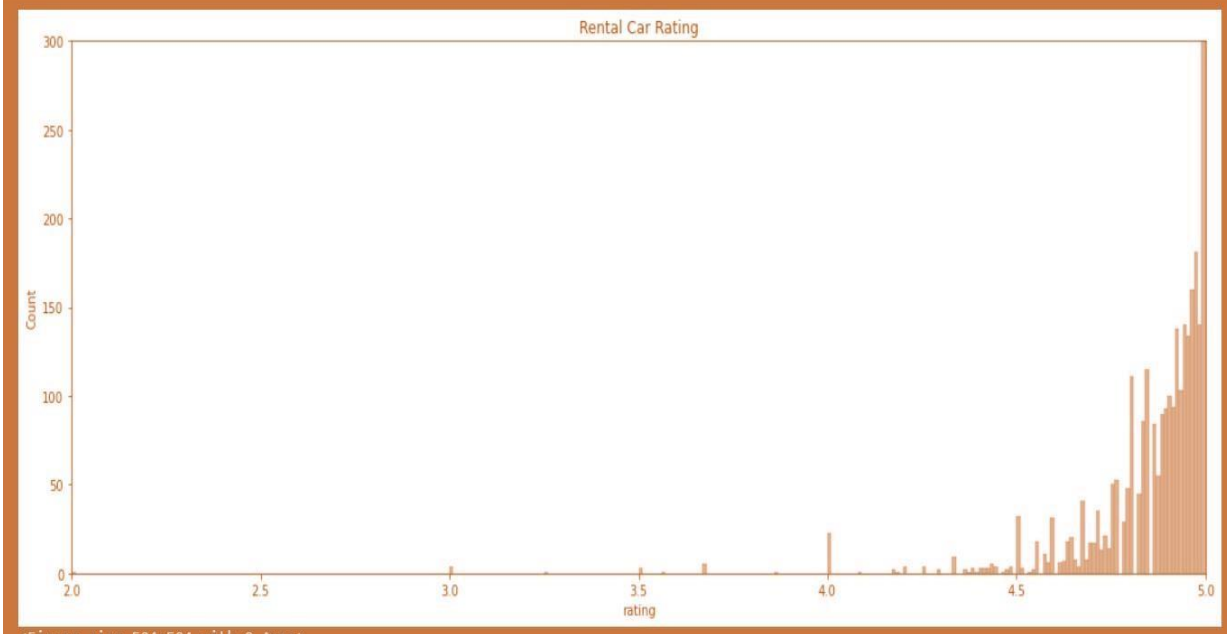


Fig.29

According to the provided dataset, the image shows the longitude and latitude of the US.

```
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')
```

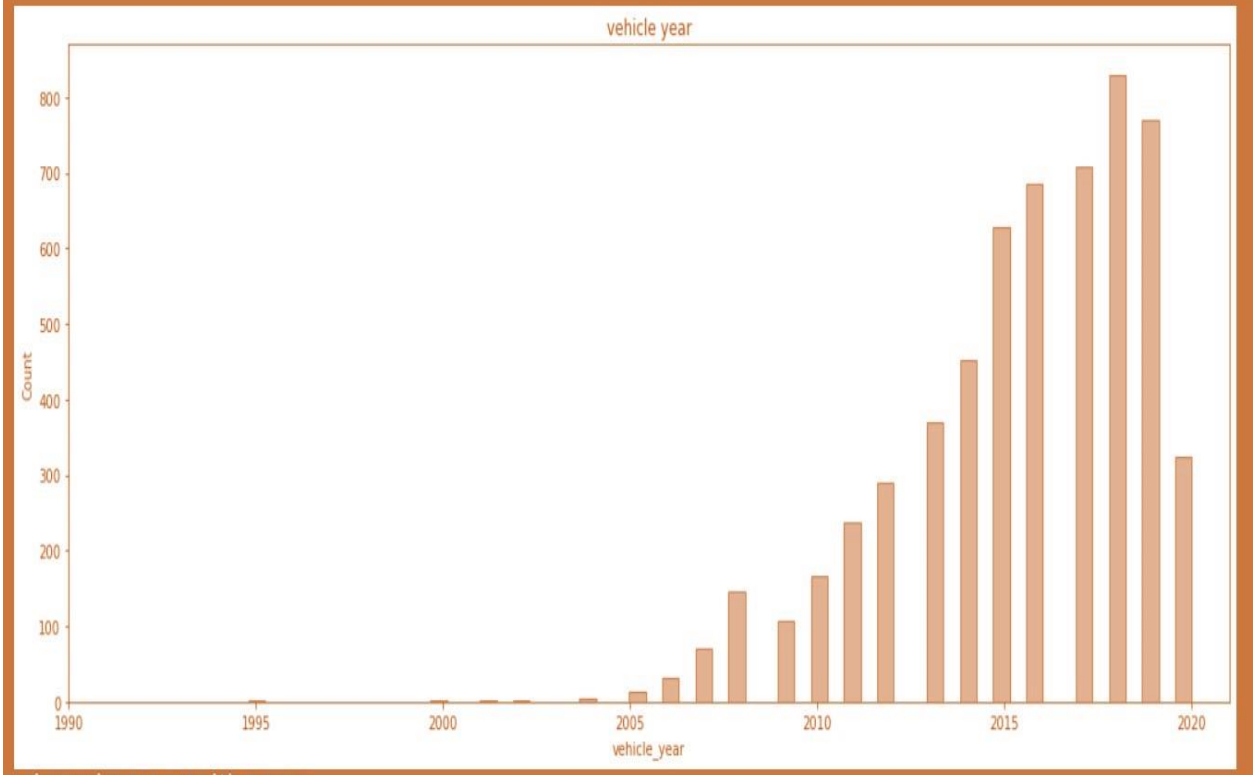



```
<Figure size 504x504 with 0 Axes>
```

Fig.30

This graph displays the rating given to a specific car based on the provided 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>
```

Fig.31

This graph displays the number of annual car rentals. As we can see, the count is at its peak between 2015 and 2020 compared to 2005.

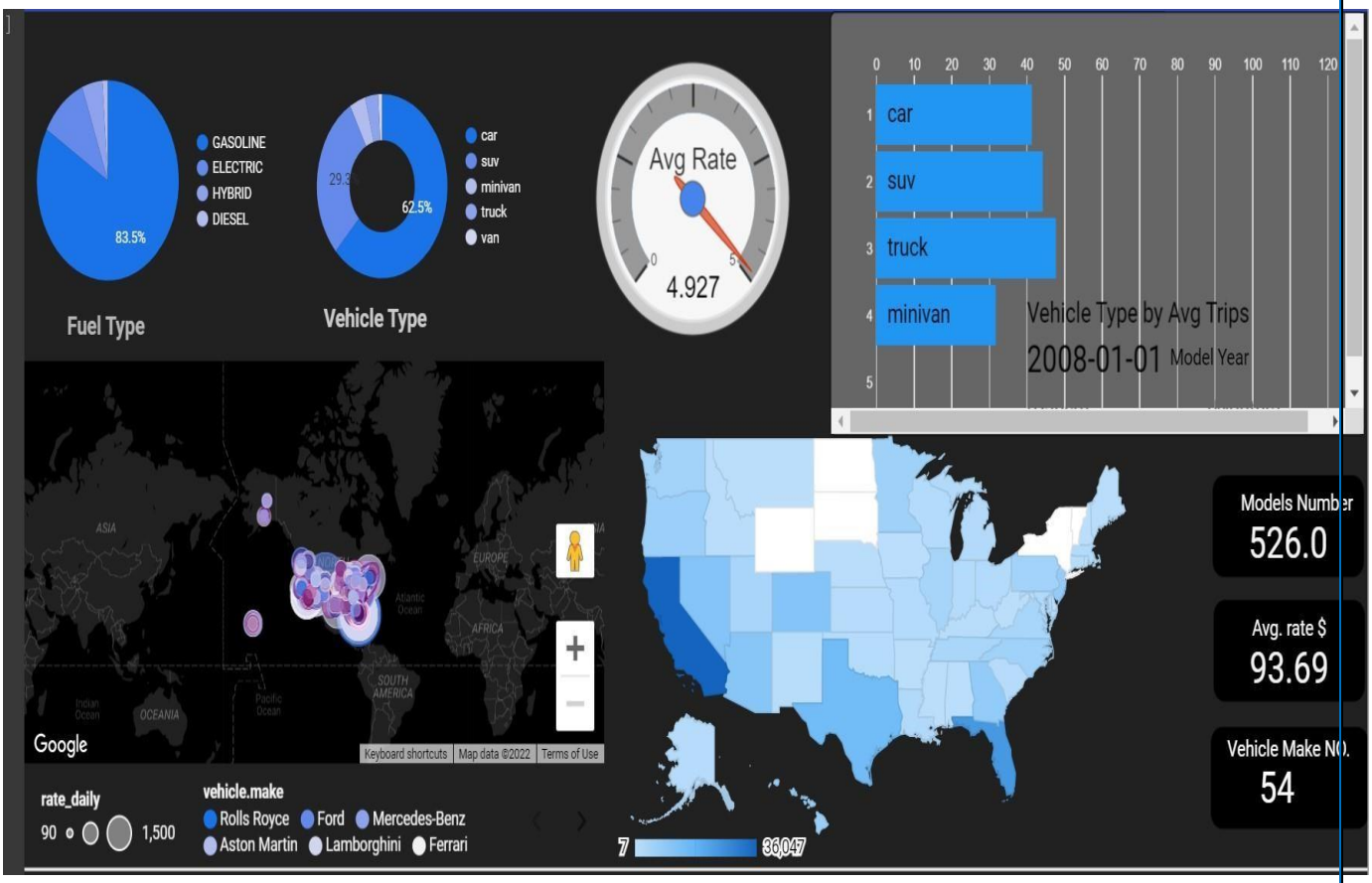


Fig.32

	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

Fig.33

This shows overall display of the model.

Segment Extraction

One of the most widely used unsupervised machine learning algorithms for solving classification issues is K means. K Means divides the unlabeled data into multiple clusters based on shared characteristics and patterns. Assume we have N unlabeled multivariate datasets with different attributes from our dataset, such as water availability, price, city, etc. Clustering is a method for categorising datasets into different groups based on shared traits and characteristics. Clusters are the entities that are forming these groups. Unsupervised learning algorithms in machine learning use clustering because it may divide multivariate data into different groups without the need for a supervisor based on a common pattern concealed within the datasets.

In the Elbow approach, the number of clusters (K) is truly variable and ranges from 1 to 10. We are calculating WCSS for each value of K. (Within-Cluster Sum of Square). The sum of the squared distances between each point and the cluster's centroid is known as WCSS. The plot of the WCSS with the K value resembles an elbow.

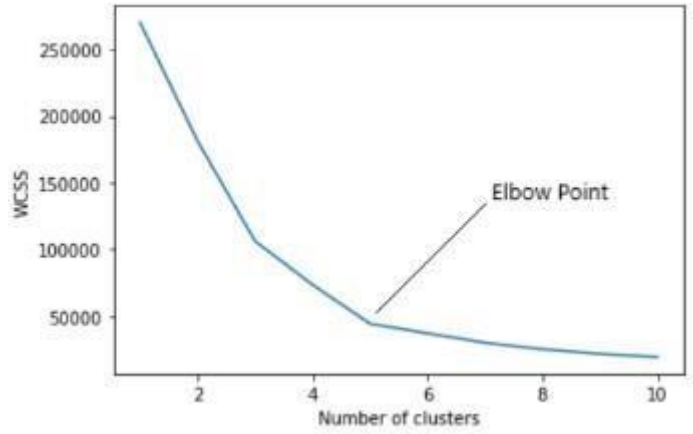


Fig.34

The WCSS value will begin to drop as the number of clusters rises. The highest WCSS value is at K = 1. When we examine the graph, we can observe that it abruptly changes at one point, forming an elbow. The graph then begins to travel nearly parallel to the X-axis from this point on. The best K value, or the most clusters, is the one that corresponds to this location.

Analysing Market Segments

Geographic Segmentation: To better serve customers in a certain region, marketers segregate a target market according to geography. This kind of market segmentation is based on the geographical entities themselves (countries, states, cities, etc.), as well as a number of other geographic elements, including climate, cultural preferences, population, and more. Geographic segmentation entails dividing up your audience according to the area in which they reside or are employed. Customers can be categorised in a variety of ways, including by their nation of residence, or by more specific geographic divisions, such as city, region, or even postal code.

Even if geographic segmentation is the easiest type of market segmentation to understand, there are still many applications for it that businesses never consider. Depending on your needs as a firm, the area you target should have a different size. Generally speaking, the locations you'll be targeting will be larger the larger the firm. After all, it won't be economical to target each postcode individually with a larger potential audience.

Demographic Segmentation: An organization's target market is segmented using the market segmentation technique known as demographic segmentation based on demographic factors like age, gender, education, income, etc. It aids businesses in comprehending their clientele so that their requirements can be better met.
Companies can utilise demographic segmentation to concentrate their time and resources on those groups that have consumers who are most likely to make purchases and are therefore most valuable to them rather than trying to reach the entire market.

The demographic segmentation process involves several different factors, including: 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 : Using psychological factors such as personality, lifestyle, social status, hobbies, interests, opinions, and attitudes, psychographic segmentation is a research tool for researching consumers and classifying them into groups.

You can interact with several target audiences using psychographic marketing in the methods that will have the greatest influence on each of them. This strategy avoids wasting time and money on unproductive strategies and facilitates communication with the groups you care about.

For market segmentation, we can utilise psychographics to comprehend: 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 alsoinclude 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.