Online Vehicle Booking Market

December 2, 2022

[2]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

[6]:
```
df=pd.read_csv("CarRentalData.csv")
```

[15]:
```
df=df.fillna(0)
df.sample(5)
```

[15]:

	fuelType	rating	renterTripsTaken	reviewCount	location.city	location.country	location.latitude	location.longitude	location.state	owner.id	rate.daily	vehicle.make	vehicle.model	vehicle.type	vehicle.year
650	GASOLINE	5.00	26	20	Chattanooga	US	35.023739	-85.155102	TN	5252641	85	Infiniti	G Sedan	car	2013
4315	ELECTRIC	4.89	19	18	San Diego	US	32.921706	-117.139695	CA	8198418	95	Tesla	Model 3	car	2020
5261	ELECTRIC	5.00	8	6	Westminster	US	33.746283	-118.012524	CA	11671249	70	Tesla	Model 3	car	2019
4233	GASOLINE	5.00	5	4	North Miami Beach	US	25.950070	-80.166528	FL	864946	150	BMW	4 Series	car	2015
2028	HYBRID	5.00	55	47	Las Vegas	US	36.093421	-115.296466	NV	2434410	294	BMW	i8	car	2016

[19]:
```
print(df.isnull().sum())
```

```
fuelType              0
rating                0
renterTripsTaken      0
reviewCount           0
location.city         0
location.country      0
location.latitude     0
location.longitude    0
location.state        0
owner.id              0
rate.daily            0
vehicle.make          0
vehicle.model         0
vehicle.type          0
vehicle.year          0
dtype: int64
```

[13]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5851 entries, 0 to 5850
Data columns (total 15 columns):
 #   Column              Non-Null Count  Dtype
---  ------              --------------  -----
 0   fuelType            5851 non-null   object
 1   rating              5851 non-null   float64
 2   renterTripsTaken    5851 non-null   int64
 3   reviewCount         5851 non-null   int64
 4   location.city       5851 non-null   object
 5   location.country    5851 non-null   object
 6   location.latitude   5851 non-null   float64
 7   location.longitude  5851 non-null   float64
 8   location.state      5851 non-null   object
 9   owner.id            5851 non-null   int64
 10  rate.daily          5851 non-null   int64
 11  vehicle.make        5851 non-null   object
 12  vehicle.model       5851 non-null   object
 13  vehicle.type        5851 non-null   object
 14  vehicle.year        5851 non-null   int64
dtypes: float64(3), int64(5), object(7)
memory usage: 685.8+ KB
```

[9]:
```
df.describe()
```

[9]:

	rating	renterTripsTaken	reviewCount	location.latitude	\
count	5851.000000	5851.000000	5851.000000	5851.000000	

```
mean      4.499016         33.477354    28.454794          35.582889
std       1.387888         41.898954    35.136113           6.431408
min       0.000000          0.000000     0.000000          21.272565
25%       4.850000          5.000000     4.000000          30.453623
50%       4.980000         18.000000    16.000000          35.554502
75%       5.000000         46.000000    39.000000          39.996864
max       5.000000        395.000000   321.000000          64.893610

       location.longitude      owner.id   rate.daily  vehicle.year
count         5851.000000  5.851000e+03  5851.000000   5851.000000
mean           -99.632773  6.034830e+06    93.691506   2015.340113
std             20.391476  4.646022e+06    96.080920      4.050813
min           -158.165693  5.105000e+03    20.000000   1955.000000
25%           -117.158285  1.917451e+06    45.000000   2014.000000
50%            -95.673319  4.968749e+06    69.000000   2016.000000
75%            -81.538631  9.657496e+06   110.000000   2018.000000
max            -68.823637  1.581088e+07  1500.000000   2020.000000
```

[16]:
```
df['vehicle.make'].value_counts()
```

[16]:
```
Tesla            598
Toyota           591
BMW              456
Ford             436
Chevrolet        431
Mercedes-Benz    342
Nissan           291
Jeep             279
Honda            257
Porsche          187
Hyundai          183
Dodge            182
Audi             169
Kia              154
Volkswagen       136
Lexus            121
Land Rover        93
Subaru            83
Mazda             71
Maserati          67
GMC               66
Jaguar            58
Chrysler          57
Acura             50
Mercedes-benz     49
Polaris           46
Cadillac          44
```

```
FIAT              42
Infiniti          38
MINI              31
Mitsubishi        29
Volvo             22
Alfa Romeo        21
Lincoln           21
Ram               20
Buick             19
Lamborghini       15
smart             14
Scion             13
Ferrari           13
Alfa-romeo         9
Bentley            9
INFINITI           8
Pontiac            5
Aston Martin       5
Lotus              4
Rolls Royce        3
HUMMER             3
Mercury            2
Saturn             2
Genesis            2
Suzuki             2
Saab               1
McLaren            1
Name: vehicle.make, dtype: int64
```

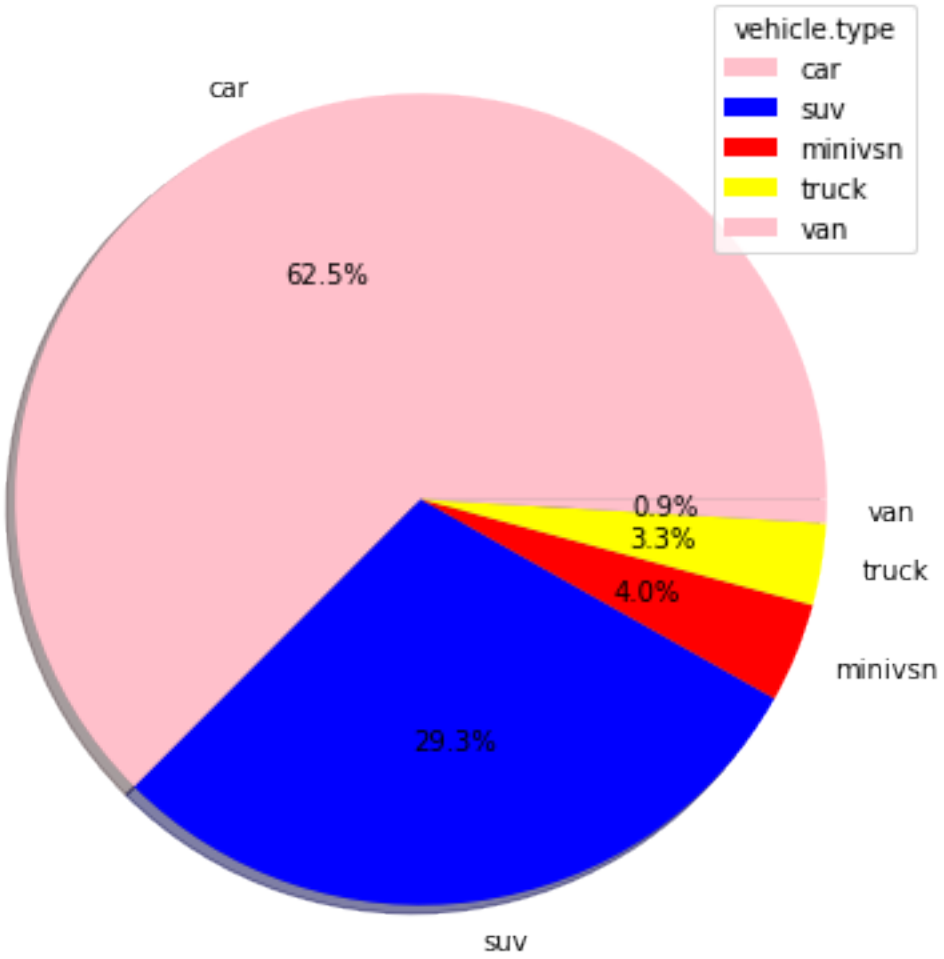
[ ]:

[10]:
```
df['vehicle.type'].value_counts()
```

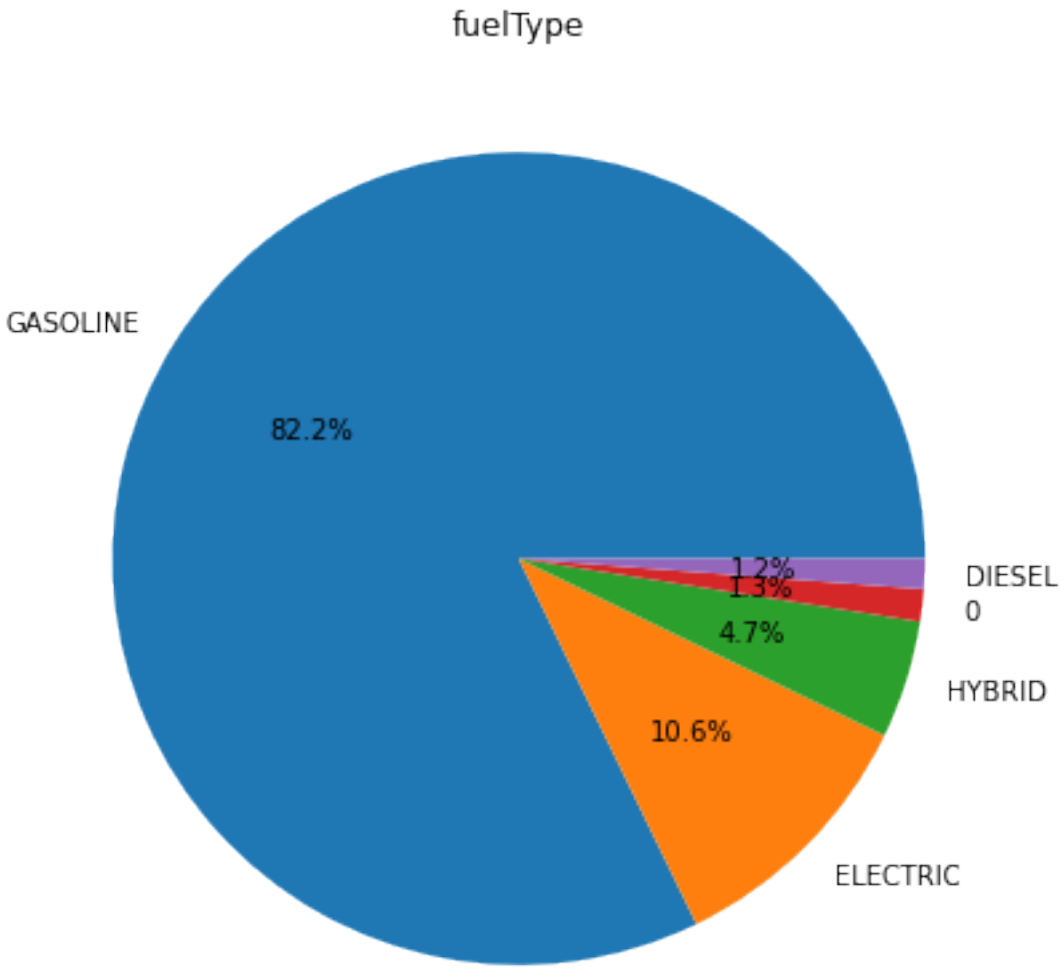
[10]:
```
car        3659
suv        1714
minivan     232
truck       191
van          55
Name: vehicle.type, dtype: int64
```

[11]:
```
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()
```



[46]:
```
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>
```

[68]:
```
df['Location'].value_counts()
```

[68]:
```
Barcelona          83
Madrid             60
Valencia           27
Murcia             27
Málaga             18
Alicante           12
Pontevedra         11
Vizcaya            11
Ciudad Real         7
Cantabria           6
```

```
La Coruña               6
Tarragona               5
Zaragoza                4
Guipúzcoa               4
Córdoba                 3
Islas Baleares          3
Girona                  3
Castellón               3
Palencia                2
Salamanca               2
Asturias                2
Granada                 2
Valladolid              2
Albacete                2
Huesca                  2
Lugo                    2
León                    1
Cádiz                   1
Sevilla                 1
Burgos                  1
Sta. C. de Tenerife     1
Cáceres                 1
Lleida                  1
Toledo                  1
Almería                 1
Ceuta                   1
Navarra                 1
Name: Location, dtype: int64
```

[20]:
```
df["fuelType"].value_counts()
```

[20]:
```
GASOLINE    4810
ELECTRIC     622
HYBRID       274
0             75
DIESEL        70
Name: fuelType, dtype: int64
```

[21]:
```
df["location.country"].value_counts()
```

[21]:
```
US    5851
Name: location.country, dtype: int64
```

[23]:
```
df.drop("location.country", axis=1, inplace=True)
```

[26]:
```
# get number of categories value
print("Number of Categories in: ")
```

```
for ColName in df[['fuelType','location.city','location.state','vehicle.
 ↪make','vehicle.model','vehicle.year']]:
    print("{} = {}".format(ColName,len(df[ColName].unique())))
```

```
Number of Categories in:
fuelType = 5
location.city = 971
location.state = 46
vehicle.make = 54
vehicle.model = 526
vehicle.year = 34
```

[29]:
```
median=df["rating"].median()
df["rating"].fillna(median, inplace=True)
```

[31]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5851 entries, 0 to 5850
Data columns (total 14 columns):
 #   Column              Non-Null Count  Dtype
---  ------              --------------  -----
 0   fuelType            5851 non-null   object
 1   rating              5851 non-null   float64
 2   renterTripsTaken    5851 non-null   int64
 3   reviewCount         5851 non-null   int64
 4   location.city       5851 non-null   object
 5   location.latitude   5851 non-null   float64
 6   location.longitude  5851 non-null   float64
 7   location.state      5851 non-null   object
 8   owner.id            5851 non-null   int64
 9   rate.daily          5851 non-null   int64
 10  vehicle.make        5851 non-null   object
 11  vehicle.model       5851 non-null   object
 12  vehicle.type        5851 non-null   object
 13  vehicle.year        5851 non-null   int64
dtypes: float64(3), int64(5), object(6)
memory usage: 640.1+ KB
```

[32]:
```
dt1 = df.replace(np.nan, 'GASOLINE', regex=True)
```

[33]:
```
dt1 = dt1.rename(columns={'location.latitude': 'latitude', 'location.longitude':
 ↪ 'longitude',
                          'rate.daily': 'rate_daily','vehicle.year':␣
 ↪'vehicle_year'})
dt1.head()
```

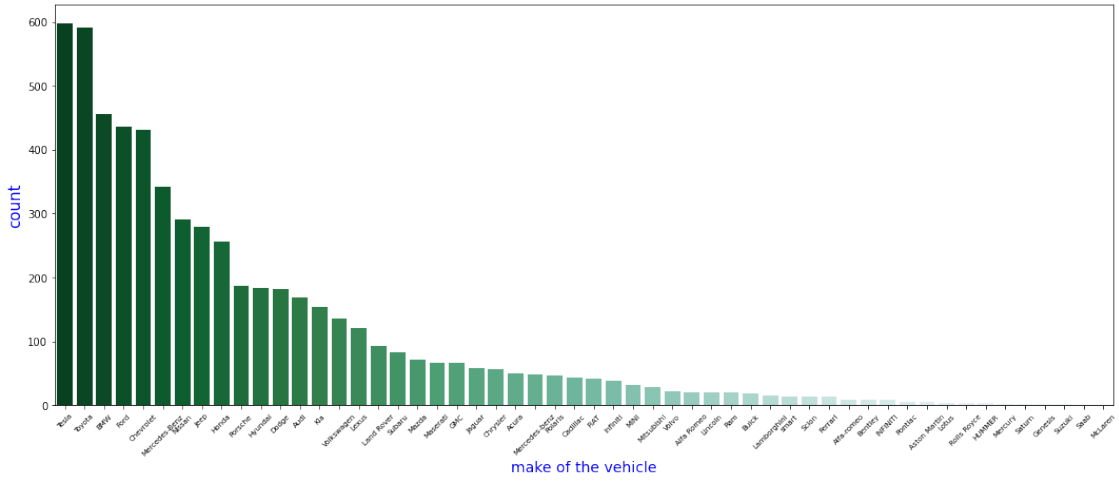
[33]:

	fuelType	rating	renterTripsTaken	reviewCount	location.city	latitude	\
0	ELECTRIC	5.00	13	12	Seattle	47.449107	
1	ELECTRIC	5.00	2	1	Tijeras	35.111060	
2	HYBRID	4.92	28	24	Albuquerque	35.127163	
3	GASOLINE	5.00	21	20	Albuquerque	35.149726	
4	GASOLINE	5.00	3	1	Albuquerque	35.208659	

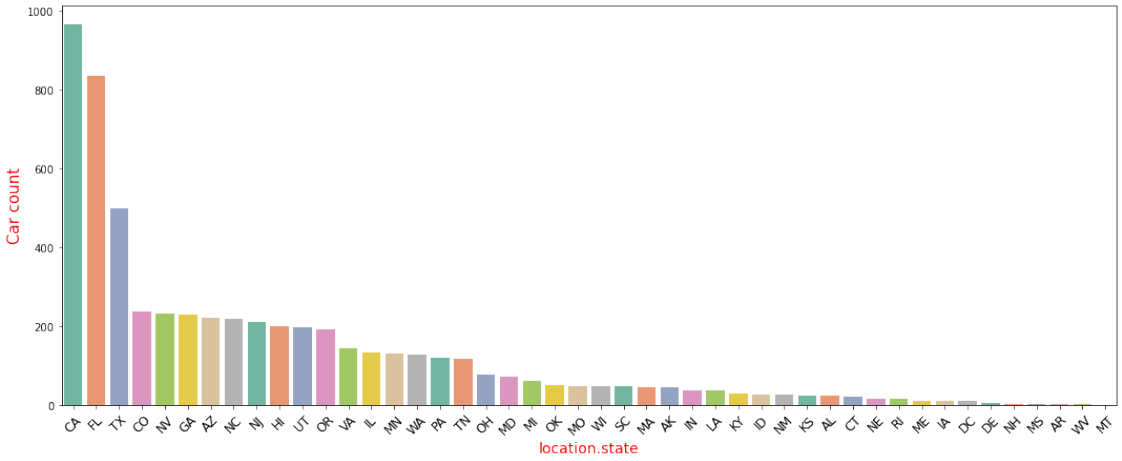
	longitude	location.state	owner.id	rate_daily	vehicle.make	vehicle.model	\
0	-122.308841	WA	12847615	135	Tesla	Model X	
1	-106.276551	NM	15621242	190	Tesla	Model X	
2	-106.566681	NM	10199256	35	Toyota	Prius	
3	-106.711425	NM	9365496	75	Ford	Mustang	
4	-106.601008	NM	3553565	47	Chrysler	Sebring	

	vehicle.type	vehicle_year
0	suv	2019
1	suv	2018
2	car	2012
3	car	2018
4	car	2010

[49]:
```
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)
```



[51]:
```
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)
```



[36]:
```
pip install folium
```

```
Collecting folium
  Downloading folium-0.13.0-py2.py3-none-any.whl (96 kB)
Requirement already satisfied: numpy in c:\users\srakesh\anaconda3\lib\site-
packages (from folium) (1.21.5)
Requirement already satisfied: requests in c:\users\srakesh\anaconda3\lib\site-
packages (from folium) (2.27.1)
Collecting branca>=0.3.0
  Downloading branca-0.6.0-py3-none-any.whl (24 kB)
Requirement already satisfied: jinja2>=2.9 in
c:\users\srakesh\anaconda3\lib\site-packages (from folium) (2.11.3)
Requirement already satisfied: MarkupSafe>=0.23 in
c:\users\srakesh\anaconda3\lib\site-packages (from jinja2>=2.9->folium) (2.0.1)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in
c:\users\srakesh\anaconda3\lib\site-packages (from requests->folium) (1.26.9)
```

```
Requirement already satisfied: charset-normalizer~=2.0.0 in
c:\users\srakesh\anaconda3\lib\site-packages (from requests->folium) (2.0.4)
Requirement already satisfied: idna<4,>=2.5 in
c:\users\srakesh\anaconda3\lib\site-packages (from requests->folium) (3.3)
Requirement already satisfied: certifi>=2017.4.17 in
c:\users\srakesh\anaconda3\lib\site-packages (from requests->folium) (2021.10.8)
Installing collected packages: branca, folium
Successfully installed branca-0.6.0 folium-0.13.0
Note: you may need to restart the kernel to use updated packages.
```

Plotting the data on a map with Folium

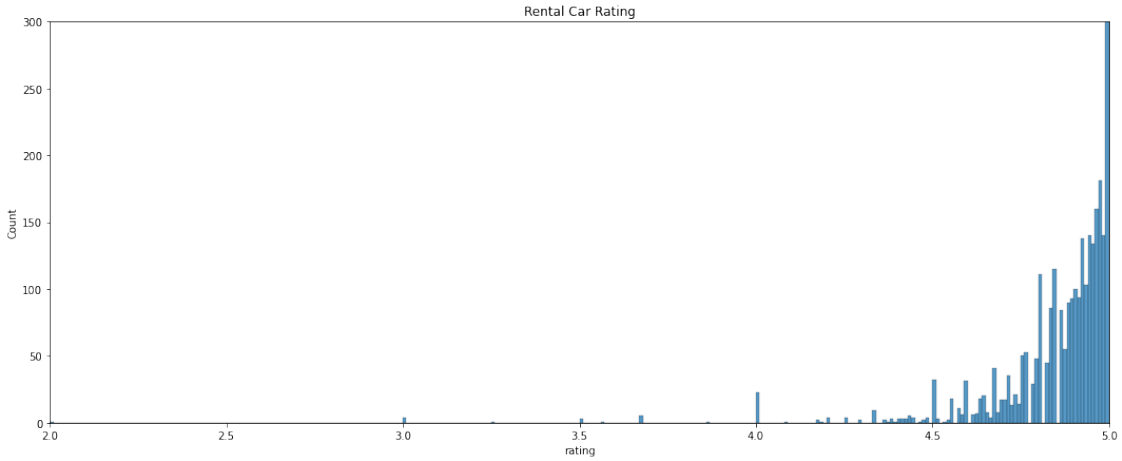
[38]:
```
import folium
from folium.plugins import HeatMap
center = [35.582889, -99.632773]  #data.describe(mean)
m = folium.Map([dt1.latitude.mean(), dt1.longitude.mean()],␣
 ↪zoom_start=4,center=center)
for index, row in dt1.iterrows():
    folium.CircleMarker([row['latitude'], row['longitude']],
                        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
```

```
C:\Users\sRakesh\AppData\Local\Temp\ipykernel_1016\1378005713.py:12:
FutureWarning: Method `add_children` is deprecated. Please use `add_child`
instead.
  m.add_children(HeatMap(points, radius=15)) # plot heatmap
```

[38]:
```
<folium.folium.Map at 0x1139bafeaf0>
```

Histogram of Rental Car Rating

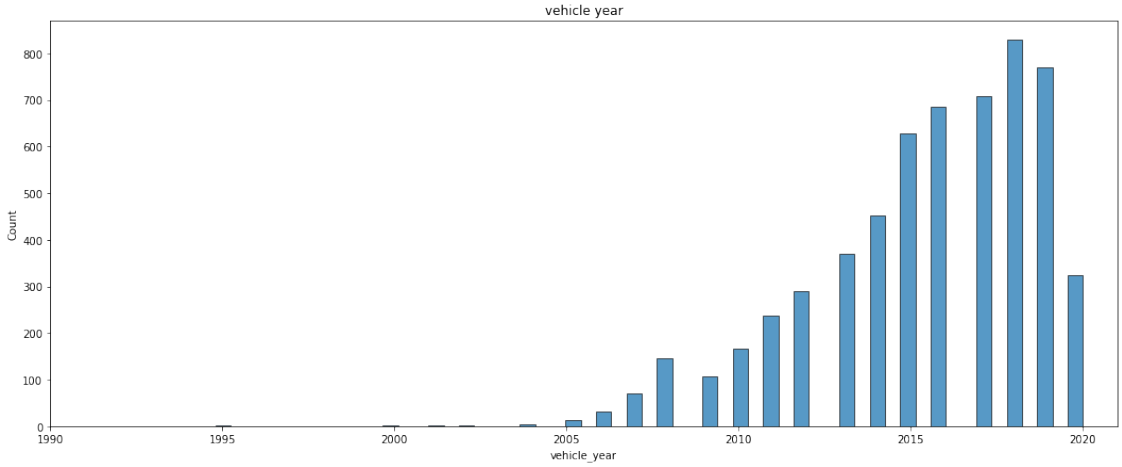
[41]:
```
f, ax = plt.subplots(figsize=(18, 7))
sns.histplot(data=dt1, x="rating", binwidth=.01)
ax.set_ylim(0,300)
ax.set_xlim(2,5)
plt.title('Rental Car Rating')
plt.show()
plt.savefig('Rental Car Rating.png', format='png')
```




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

Histogram of vehicle_year

[43]:
```
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>
```

[53]:
```
pip install plotly_express
```

```
Collecting plotly_express
  Downloading plotly_express-0.4.1-py2.py3-none-any.whl (2.9 kB)
```

```
Requirement already satisfied: statsmodels>=0.9.0 in
c:\users\srakesh\anaconda3\lib\site-packages (from plotly_express) (0.13.2)
Requirement already satisfied: pandas>=0.20.0 in
c:\users\srakesh\anaconda3\lib\site-packages (from plotly_express) (1.4.2)
Requirement already satisfied: scipy>=0.18 in
c:\users\srakesh\anaconda3\lib\site-packages (from plotly_express) (1.7.3)
Requirement already satisfied: numpy>=1.11 in
c:\users\srakesh\anaconda3\lib\site-packages (from plotly_express) (1.21.5)
Requirement already satisfied: plotly>=4.1.0 in
c:\users\srakesh\anaconda3\lib\site-packages (from plotly_express) (5.6.0)
Requirement already satisfied: patsy>=0.5 in
c:\users\srakesh\anaconda3\lib\site-packages (from plotly_express) (0.5.2)
Requirement already satisfied: python-dateutil>=2.8.1 in
c:\users\srakesh\anaconda3\lib\site-packages (from
pandas>=0.20.0->plotly_express) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in
c:\users\srakesh\anaconda3\lib\site-packages (from
pandas>=0.20.0->plotly_express) (2021.3)
Requirement already satisfied: six in c:\users\srakesh\anaconda3\lib\site-
packages (from patsy>=0.5->plotly_express) (1.16.0)
Requirement already satisfied: tenacity>=6.2.0 in
c:\users\srakesh\anaconda3\lib\site-packages (from
plotly>=4.1.0->plotly_express) (8.0.1)
Requirement already satisfied: packaging>=21.3 in
c:\users\srakesh\anaconda3\lib\site-packages (from
statsmodels>=0.9.0->plotly_express) (21.3)
Requirement already satisfied: pyparsing!=3.0.5,>=2.0.2 in
c:\users\srakesh\anaconda3\lib\site-packages (from
packaging>=21.3->statsmodels>=0.9.0->plotly_express) (3.0.4)
Installing collected packages: plotly-express
Successfully installed plotly-express-0.4.1
Note: you may need to restart the kernel to use updated packages.
```

[56]:
```
import plotly_express as px
dt_make_model = dt1.groupby(['vehicle.make', 'vehicle.model']).size().
 ↪reset_index()
dt_make_model.rename(columns = {0:'model_count'}, inplace=True)
dt_make_model['make_count'] = dt_make_model['vehicle.make'].apply(
    lambda x : dt_make_model[dt_make_model['vehicle.make'] == x]['model_count'].
 ↪sum())
dt_make_model.sort_values(by = 'make_count', ascending=False, inplace=True)
fig =px.scatter(dt_make_model[dt_make_model['make_count'] >45],
             x = 'vehicle.make', y='model_count', color = 'vehicle.
 ↪model',width=1100, height=700,
                 title='Make and Model of Top Most Rented Cars')
fig.show()
```

[57]:
```
from IPython.display import IFrame
IFrame('https://datastudio.google.com/embed/reporting/
  ↪6ddc4bc8-e881-4983-a37f-51e057485122/page/OFP8B', width='100%', height=900)
```

[57]:
```
<IPython.lib.display.IFrame at 0x113a2bd0b80>
```

[ ]:

[ ]: