

```
[In [47]: import pandas as pd
import numpy as np
import scipy as sc
import matplotlib as pyplot
import seaborn as sns

In [48]: uber_data=pd.read_csv('uberdrive.csv')

Out [49]:
Out [3]:
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
...  ...  ...  ...  ...  ...  ...
1151 12/31/2016 13:24  12/31/2016 13:42  Business  Karachi  Unknown Location  3.9  Temporary Site
1152 12/31/2016 15:03  12/31/2016 15:38  Business  Unknown Location  Unknown Location  16.2  Meeting
1153 12/31/2016 21:32  12/31/2016 21:50  Business  Katunayake  Gampaha  6.4  Temporary Site
1154 12/31/2016 22:08  12/31/2016 23:51  Business  Gampaha  Ilukwatta  48.2  Temporary Site
...  ...  ...  ...  ...  ...  ...
Totals  NaN  NaN  NaN  NaN  12204.7  NaN
1156 rows x 7 columns

In [4]: uber_data.shape
Out[4]: (1156, 7)

In [5]: uber_data.size
Out[5]: 8092

In [7]: uber_data.head(10)
Out[7]:
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  NaN
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  4.7  Meeting
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
5  01-06-2016 17:15  01-06-2016 17:19  Business  West Palm Beach  West Palm Beach  4.3  Meal/Entertain
6  01-06-2016 17:30  01-06-2016 17:35  Business  West Palm Beach  Palm Beach  7.1  Meeting
7  01-07-2016 13:27  01-07-2016 13:33  Business  Cary  Cary  0.8  Meeting
8  01-10-2016 08:05  01-10-2016 08:25  Business  Cary  Morrisville  8.3  Meeting
9  01-10-2016 12:17  01-10-2016 12:44  Business  Jamaica  New York  16.5  Customer Visit

In [8]: uber_data.describe()
Out[8]:
MILES*
count    1156.000000
mean       21.115398
std       359.299007
min         0.500000
25%       2.900000
50%       6.000000
75%      10.400000
max     12204.700000

In [10]: uber_data.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1156 entries, 0 to 1155
Data columns (total 7 columns):
#   Column      Non-Null Count  Dtype
---  ---
0   START_DATE*  1156 non-null    object
1   END_DATE*    1155 non-null    object
2   CATEGORY*    1155 non-null    object
3   START*       1155 non-null    object
4   STOP*        1155 non-null    object
5   MILES*       1156 non-null    float64
6   PURPOSE*     653 non-null     object
dtypes: float64(1), object(6)
memory usage: 63.3+ KB

In [14]: # To know the count of Unique start points
len(uber_data["START*"].unique())
Out[14]: 177

In [12]: # List of unique start points
start=uber_data["START*"].dropna()
unique_start=set(start)

In [13]: unique_start
Out[13]: {'Agnew', 'Almond', 'Apex', 'Arabi', 'Arlington', 'Arlington Park at Amberly', 'Asheville', 'Austin', 'Banner Elk', 'Bellevue', 'Berkeley', 'Boone', 'Briar Meadow', 'Bryson City', 'Burtrose', 'CBD', 'Capitol One', 'Cary', 'Central', 'Chalmette', 'Chapel Hill', 'Chessington', 'College Avenue', 'Colombo', 'Columbia Heights', 'Congress Ave District', 'Connecticut Avenue', 'Convention Center District', 'Cory', 'Couples Glen', 'Covington', 'Coville', 'Coville', 'Daytona Beach', 'Downtown', 'Durham', 'Eagan Park', 'Eagle Rock', 'East Austin', 'East Elmhurst', 'East Harlem', 'Eastgate', 'Edgehill Farms', 'El Cerrito', 'Elk Park', 'Elmhurst', 'Emeryville', 'Fairmont', 'Farmington Woods', 'Fayetteville Street', 'Financial District', 'Flatiron District', 'Florence', 'Fort Pierce', 'Fuquay-Varina', 'Galveston', 'Gampaha', 'Georgian Acres', 'Gulfton', 'Hayesville', 'Hazelwood', 'Hell's Kitchen', 'Heritage Pines', 'Holly Springs', 'Houston', 'Hudson Square', 'Huntington Woods', 'Ingleside', 'Islamabad', 'Jackson Heights', 'Jacksonville', 'Jamaica', 'Jamestown Court', 'K Street', 'Kalamora Triangle', 'Karachi', 'Katunayaka', 'Katunayake', 'Katy', 'Kenner', 'Kilarny Woods', 'Kissimmee', 'Kissimmee', 'Krendle Woods', 'Lahore', 'Lake Reams', 'Lake Wellington', 'Lakeview', 'Latta', 'Lexington Park at Amberly', 'Long Island City', 'Lower Garden District', 'Lower Manhattan', 'Mandeville', 'Marigny', 'Mcav', 'Mebane', 'Medical Centre', 'Menlo Park', 'Meredith', 'Meredith Townes', 'Metairie', 'Midtown', 'Midtown East', 'Morrisville', 'Mountain View', 'NOMA', 'New Orleans', 'New York', 'Newark', 'Newland', 'Noorpur Shahan', 'North Austin', 'North Berkeley Hills', 'Northwoods', 'Nugegoda', 'Oakland', 'Old City', 'Orlando', 'Palo Alto', 'Parkway', 'Parkwood', 'Ponchartrain Beach', 'Ponchartrain Shores', 'Port Bolivar', 'Potrero', 'R'walpindi', 'Raleigh', 'Rawalpindi', 'Red River District', 'Redmond', 'Renaissance', 'Ridgeland', 'Rose Hill', 'SOMISSPO', 'San Francisco', 'San Jose', 'Sand Lake Commons', 'Santa Clara', 'Savon Height', 'Seaport', 'Seattle', 'Sharpsport', 'Sky Lake', 'Soho', 'South', 'South Berkeley', 'South Congress', 'Southside', 'St. Thomas', 'Stoney Brook', 'Storyville', 'Sugar Land', 'Sumnerland', 'Sunnyvale', 'Sunnyvale', 'Tanglewood', 'Tenderloin', 'The Drag', 'Topton', 'Townes at Everett Crossing', 'Tribeca', 'University District', 'Unknown Location', 'Wake Co.', 'Wake Forest', 'Washington', 'Washington Avenue', 'Waverly Place', 'Wayne Ridge', 'West Berkeley', 'West End', 'West Palm Beach', 'West University', 'Weston', 'Westpark Place', 'Whitebridge', 'Winston Salem'}

In [16]: #To know the count of unique stop
len(uber_data["STOP*"].unique())
Out[16]: 188

In [17]: stop=uber_data["STOP*"].dropna()
unique_stop=set(stop)

Out[17]: {'Agnew', 'Alief', 'Almond', 'Apex', 'Arabi', 'Arlington', 'Arlington Park at Amberly', 'Arts District', 'Asheville', 'Banner Elk', 'Bay Farm Island', 'Bellevue', 'Berkeley', 'Boone', 'Briar Meadow', 'Bryson City', 'Burtrose', 'Bywater', 'CBD', 'Capitol One', 'Cary', 'Cedar Hill', 'Central', 'Chalmette', 'Chapel Hill', 'Chessington', 'College Avenue', 'Colombo', 'Columbia Heights', 'Congress Ave District', 'Convention Center District', 'Cory', 'Covington', 'Coville', 'Coxville', 'Daytona Beach', 'Depot Historic District', 'Downtown', 'Durham', 'Eagle Rock', 'East Elmhurst', 'East Harlem', 'Edgehill Farms', 'El Cerrito', 'Elk Park', 'Emeryville', 'Farmington Woods', 'Faubourg Marigny', 'Financial District', 'Florence', 'Fort Pierce', 'French Quarter', 'Fuquay-Varina', 'Galveston', 'Gampaha', 'Georgetown', 'Georgetown Greenpoint', 'Gulfton', 'Harden Place', 'Hayesville', 'Hazelwood', 'Hell's Kitchen', 'Heritage Pines', 'Holly Springs', 'Houston', 'Hudson Square', 'Huntington Woods', 'Ilukwatta', 'Ingleside', 'Islamabad', 'Isles of Buena Vista', 'Jackson Heights', 'Jacksonville', 'Jamaica', 'Jamestown Court', 'K Street', 'Kalamora Triangle', 'Karachi', 'Katunayaka', 'Katy', 'Kenner', 'Kilarny Woods', 'Kissimmee', 'Kissimmee', 'Kissimmee', 'Lahore', 'Lake Reams', 'Lake Wellington', 'Lakeview', 'Latta', 'Lexington Park at Amberly', 'Long Island City', 'Lower Garden District', 'Lower Manhattan', 'Mandeville', 'Marigny', 'Mcav', 'Mebane', 'Medical Centre', 'Menlo Park', 'Meredith', 'Meredith Townes', 'Metairie', 'Midtown', 'Midtown East', 'Midtown West', 'Morrisville', 'Mountain View', 'New Orleans', 'New York', 'Newark', 'Newland', 'NoMad', 'Noorpur Shahan', 'North Austin', 'North Berkeley Hills', 'Northwest Rectangle', 'Northwoods', 'Nugegoda', 'Oakland', 'Orlando', 'Palm Beach', 'Palo Alto', 'Parkway', 'Parkway Museums', 'Parkwood', 'Ponchartrain Beach', 'Ponchartrain Shores', 'Port Bolivar', 'Potrero Flats', 'Potrero', 'Queens', 'Queens County', 'Queens County', 'R'walpindi', 'Raleigh', 'Rawalpindi', 'Red River District', 'Redmond', 'Renaissance', 'Ridgeland', 'Ridgeland', 'SOMISSPO', 'San Francisco', 'San Jose', 'Sand Lake Commons', 'Santa Clara', 'Savon Height', 'Seattle', 'Sharpsport', 'Sky Lake', 'Soho', 'South', 'South Berkeley', 'South Congress', 'Southside', 'Southwest Berkeley', 'St. Thomas', 'Stonewater', 'Storyville', 'Sugar Land', 'Sumnerland', 'Sunnyvale', 'Sunnyvale', 'Tanglewood', 'Tenderloin', 'The Drag', 'Topton', 'Tribeca', 'University District', 'Unknown Location', 'Viata East', 'Wake Forest', 'Walnut Terrace', 'Washington', 'Washington Avenue', 'Waverly Place', 'Wayne Ridge', 'West Berkeley', 'West Palm Beach', 'West University', 'Westpark Place', 'Whitebridge', 'Williamsburg Manor', 'Winston Salem'}

In [24]: uber_data[uber_data["STOP*"].isin(['Katy','Berkeley']) | uber_data["START*"].isin(['San Francisco','Central'],'Arlington')]
Out[24]:
START_DATE*  END_DATE*  CATEGORY*  START*  STOP*  MILES*  PURPOSE*
232  3/17/2016 12:52  3/17/2016 15:11  Austin  Katy  136.0  Customer Visit
362  05-09-2016 14:39  05-09-2016 15:06  Business  San Francisco  Palo Alto  20.5  Between Offices
436  6/13/2016 18:08  6/13/2016 18:47  Business  Emeryville  Berkeley  3.9  Meal/Entertain
440  6/14/2016 16:09  6/14/2016 16:39  Business  San Francisco  Emeryville  11.6  Meeting
831  10/18/2016 19:03  10/18/2016 19:31  Business  Emeryville  Berkeley  3.0  NaN
836  10/19/2016 14:02  10/19/2016 14:31  Business  San Francisco  Berkeley  10.8  NaN
840  10/20/2016 11:26  10/20/2016 11:34  Business  Emeryville  Berkeley  3.1  NaN
913  11-06-2016 16:27  11-06-2016 17:28  Business  Santa Clara  Berkeley  43.9  Customer Visit
917  11-07-2016 19:17  11-07-2016 19:57  Business  San Francisco  Berkeley  13.2  Between Offices
919  11-08-2016 12:16  11-08-2016 12:49  Business  San Francisco  Berkeley  11.3  Meeting
921  11-08-2016 16:21  11-08-2016 16:34  Business  Emeryville  Berkeley  3.0  NaN
927  11-09-2016 18:40  11-09-2016 19:17  Business  Oakland  Oakland  12.7  Customer Visit
928  11-09-2016 20:52  11-09-2016 21:02  Business  Oakland  Berkeley  2.6  NaN
929  11-09-2016 21:56  11-09-2016 22:02  Business  Central  Central  1.1  NaN
933  11-10-2016 15:17  11-10-2016 15:22  Business  San Francisco  Oakland  9.9  Temporary Site
934  11-10-2016 15:30  11-10-2016 15:53  Business  Oakland  Berkeley  6.0  Meeting
939  11-11-2016 14:39  11-11-2016 15:46  Business  Menlo Park  Berkeley  36.6  Customer Visit
940  11-11-2016 18:30  11-11-2016 18:43  Business  Central  College Avenue  2.9  NaN
942  11-12-2016 10:32  11-12-2016 10:52  Business  Central  South  2.3  NaN
945  11-12-2016 13:46  11-12-2016 13:59  Business  Central  West Berkeley  3.7  Errand/Supplies
949  11-12-2016 15:40  11-12-2016 15:59  Business  Emeryville  Berkeley  0.7  Errand/Supplies
950  11/13/2016 8:54  11/13/2016 9:02  Business  Emeryville  Central  2.3  NaN
951  11/13/2016 9:21  11/13/2016 9:53  Business  Central  Central  2.6  NaN
952  11/13/2016 10:37  11/13/2016 10:37  Business  Central  Southside  1.9  NaN
958  11/13/2016 15:14  11/13/2016 15:24  Business  Central  Southside  1.9  NaN
961  11/14/2016 13:40  11/14/2016 14:33  Business  Mountain View  Berkeley  43.6  Customer Visit
963  11/14/2016 20:19  11/14/2016 20:30  Business  Emeryville  Berkeley  3.7  Errand/Supplies
966  11/15/2016 20:44  11/15/2016 21:00  Business  San Francisco  Berkeley  11.8  Temporary Site
968  11/16/2016 22:52  11/16/2016 23:02  Business  El Cerrito  Berkeley  3.1  Meal/Entertain

In [28]: # To know the most popular starting point
uber_data["START*"].value_counts().head(10)
Out[28]:
Cary    201
Unknown Location    148
Morrisville    85
Whitebridge    68
Islamabad    57
Durham    37
Lahore    36
Karachi    31
Raleigh    28
Westpark Place    17
Name: START*, dtype: int64

In [36]: #To know the Most popular stopping pint
uber_data["STOP*"].value_counts().head(10)
Out[36]:
Cary    203
Unknown Location    149
Morrisville    84
Whitebridge    65
Islamabad    58
Durham    36
Lahore    36
Raleigh    29
Karachi    28
Apex    17
Name: STOP*, dtype: int64

In [41]: #The most popular trip
df = uber_data.dropna()
df = pd.DataFrame(df.groupby(['START*', 'STOP*']).size())
df = df.rename(columns = {'Count':''})
df = df.sort values(['Count'], ascending = False)
df.loc[df['Count'] == max(df['Count'])]
Out[41]:
COUNT
START*  STOP*
Cary  Morrisville    52

In [42]: len(df["PURPOSE*"])
Out[42]: 8

In [45]: uber_data["PURPOSE*"].value_counts().head(100)
Out[45]:
Meeting    187
Meal/Entertain    160
Errand/Supplies    128
Customer Visit    101
Temporary Site    50
Between Offices    18
Moving    4
Airport/Travel    3
Charity ($)    1
Commuter    1
Name: PURPOSE*, dtype: int64

In [57]: # Rename the column with a special character to a simplified name
df = pd.DataFrame(uber_data[["MILES*"]].groupby(df["PURPOSE*"]).sum())
df = df.reset_index()
sns.barplot(x = df['MILES*'], y = df['PURPOSE*'])
Out[57]:
<Axes: xlabel='MILES*', ylabel='PURPOSE*'>


In [58]: df
Out[58]:
PURPOSE*  MILES*
0  Airport/Travel    16.5
1  Between Offices    197.0
2  Charity ($)    15.1
3  Commute    180.2
4  Customer Visit    2085.5
5  Errand/Supplies    508.0
6  Meal/Entertain    911.7
7  Meeting    2851.3
8  Moving    18.2
9  Temporary Site    523.7

In [65]: uber_data.head()
df = pd.DataFrame(uber_data[["CATEGORY*"]].value_counts())
df.reset_index()
df.plot(kind = 'bar')
plt.show()
NameError: name 'plt' is not defined
Traceback (most recent call last)
Cell [In(65), line 7]
4 df.reset_index()
5 df.plot(kind = 'bar')
----> 7 plt.show()
NameError: name 'plt' is not defined

In [ ]:
```