Assignment on Uber dataset

# Load the dataset

In [1]: **import** os

**import** pandas **as** pd

ud **=** pd**.**read\_csv("Uber.csv") print(os**.**path**.**exists(r'C:\Users\CVR\Downloads\Uber.csv'))

True

# Display basic info about dataset

In [2]: print(type(ud))

<class 'pandas.core.frame.DataFrame'>

In [3]: ((ud**.**describe))

Out[3]: <bound method NDFrame.describe of START\_DATE\* END\_DATE\* CATEGORY\* START\* \

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 0 | 1/1/2016 21:11 | 1/1/2016 21:17 | Business | Fort | Pierce |
| 1 | 1/2/2016 1:25 | 1/2/2016 1:37 | Business | Fort | Pierce |

2 1/2/2016 20:25 1/2/2016 20:38 Business Fort Pierce

3 1/5/2016 17:31 1/5/2016 17:45 Business Fort Pierce

4 1/6/2016 14:42 1/6/2016 15:49 Business Fort Pierce

... ... ... ... ...

1151 12/31/2016 13:24 12/31/2016 13:42 Business Kar?chi

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 1152 | 12/31/2016 | 15:03 | 12/31/2016 | | 15:38 | Business | Unknown Location |
| 1153 | 12/31/2016 | 21:32 | 12/31/2016 | | 21:50 | Business | Katunayake |
| 1154 | 12/31/2016 | 22:08 | 12/31/2016 | | 23:51 | Business | Gampaha |
| 1155 | Totals | | NaN | | | NaN | NaN |
|  |  | STOP\* | MILES\* | PURPOSE\* | | | |
| 0 | Fort | Pierce | 5.1 | Meal/Entertain | | | |
| 1 | Fort | Pierce | 5.0 | NaN | | | |
| 2 | Fort | Pierce | 4.8 | Errand/Supplies | | | |
| 3 | Fort | Pierce | 4.7 | Meeting | | | |
| 4 | West Palm Beach | | 63.7 | Customer Visit | | | |
| ... | ... | | ... | ... | | | |
| 1151 | Unknown Location | | 3.9 | Temporary Site | | | |
| 1152 | Unknown Location | | 16.2 | Meeting | | | |
| 1153 | Gampaha | | 6.4 | Temporary Site | | | |
| 1154 | Ilukwatta | | 48.2 | Temporary Site | | | |
| 1155 | NaN | | 12204.7 | NaN | | | |

[1156 rows x 7 columns]>

In [4]: (ud**.**info)

Out[4]:

|  |  |  |  |
| --- | --- | --- | --- |
| <bound method DataFrame.info of  END\_DATE\* CATEGORY\* START\* \ | START\_DATE\* |  | |
| 0 1/1/2016 21:11 1/1/2016 21:17 | Business | Fort | Pierce |
| 1 1/2/2016 1:25 1/2/2016 1:37 | Business | Fort | Pierce |

2 1/2/2016 20:25 1/2/2016 20:38 Business Fort Pierce

3 1/5/2016 17:31 1/5/2016 17:45 Business Fort Pierce

4 1/6/2016 14:42 1/6/2016 15:49 Business Fort Pierce

... ... ... ... ...

1151 12/31/2016 13:24 12/31/2016 13:42 Business Kar?chi

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 1152 | 12/31/2016 | 15:03 | 12/31/2016 | | 15:38 | Business | Unknown Location |
| 1153 | 12/31/2016 | 21:32 | 12/31/2016 | | 21:50 | Business | Katunayake |
| 1154 | 12/31/2016 | 22:08 | 12/31/2016 | | 23:51 | Business | Gampaha |
| 1155 | Totals | | NaN | | | NaN | NaN |
|  |  | STOP\* | MILES\* | PURPOSE\* | | | |
| 0 | Fort | Pierce | 5.1 | Meal/Entertain | | | |
| 1 | Fort | Pierce | 5.0 | NaN | | | |
| 2 | Fort | Pierce | 4.8 | Errand/Supplies | | | |
| 3 | Fort | Pierce | 4.7 | Meeting | | | |
| 4 | West Palm Beach | | 63.7 | Customer Visit | | | |
| ... | ... | | ... | ... | | | |
| 1151 | Unknown Location | | 3.9 | Temporary Site | | | |
| 1152 | Unknown Location | | 16.2 | Meeting | | | |
| 1153 | Gampaha | | 6.4 | Temporary Site | | | |
| 1154 | Ilukwatta | | 48.2 | Temporary Site | | | |
| 1155 | NaN | | 12204.7 | NaN | | | |

[1156 rows x 7 columns]>

In [5]: print((ud**.**dtypes)) START\_DATE\* object

END\_DATE\* object

CATEGORY\* object

START\* object

STOP\* object

MILES\* float64

PURPOSE\* object dtype: object

In [6]: print((ud**.**columns))

Index(['START\_DATE\*', 'END\_DATE\*', 'CATEGORY\*', 'START\*', 'STOP\*', 'MILES\*',

'PURPOSE\*'],

dtype='object')

In [7]: print((ud**.**shape))

(1156, 7)

# Check for missing values

In [8]: print(ud**.**isnull()**.**sum()) print(ud)

|  |  |
| --- | --- |
| START\_DATE\* | 0 |
| END\_DATE\* | 1 |
| CATEGORY\* | 1 |
| START\* | 1 |
| STOP\* | 1 |
| MILES\* | 0 |
| PURPOSE\* | 503 |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| dtype: | int64  START\_DATE\* | | END\_DATE\* | | | CATEGORY\* | START\* \ |
| 0 | 1/1/2016 21:11 | | 1/1/2016 21:17 | | | Business | Fort Pierce |
| 1 | 1/2/2016 1:25 | | 1/2/2016 1:37 | | | Business | Fort Pierce |
| 2 | 1/2/2016 | 20:25 | 1/2/2016 | | 20:38 | Business | Fort Pierce |
| 3 | 1/5/2016 | 17:31 | 1/5/2016 | | 17:45 | Business | Fort Pierce |
| 4  ... | 1/6/2016 | 14:42  ... | 1/6/2016 | | 15:49  ... | Business  ... | Fort Pierce  ... |
| 1151 | 12/31/2016 | 13:24 | 12/31/2016 | | 13:42 | Business | Kar?chi |
| 1152 | 12/31/2016 | 15:03 | 12/31/2016 | | 15:38 | Business | Unknown Location |
| 1153 | 12/31/2016 | 21:32 | 12/31/2016 | | 21:50 | Business | Katunayake |
| 1154 | 12/31/2016 | 22:08 | 12/31/2016 | | 23:51 | Business | Gampaha |
| 1155 |  | Totals |  | | NaN | NaN | NaN |
|  |  | STOP\* | MILES\* | PURPOSE\* | | | |
| 0 | Fort | Pierce | 5.1 | Meal/Entertain | | | |
| 1 | Fort | Pierce | 5.0 | NaN | | | |
| 2 | Fort | Pierce | 4.8 | Errand/Supplies | | | |
| 3 | Fort | Pierce | 4.7 | Meeting | | | |
| 4 | West Palm Beach | | 63.7 | Customer Visit | | | |
| ... | ... | | ... | ... | | | |
| 1151 | Unknown Location | | 3.9 | Temporary Site | | | |
| 1152 | Unknown Location | | 16.2 | Meeting | | | |
| 1153 | Gampaha | | 6.4 | Temporary Site | | | |
| 1154 | Ilukwatta | | 48.2 | Temporary Site | | | |
| 1155 | NaN | | 12204.7 | NaN | | | |

[1156 rows x 7 columns]

# Drop rows with missing values

In [9]: ud **=** ud**.**dropna() print(ud)

START\_DATE\* END\_DATE\* CATEGORY\* START\* \

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 0 | 1/1/2016 | 21:11 | 1/1/2016 | | 21:17 | Business | Fort Pierce |
| 2 | 1/2/2016 | 20:25 | 1/2/2016 | | 20:38 | Business | Fort Pierce |
| 3 | 1/5/2016 | 17:31 | 1/5/2016 | | 17:45 | Business | Fort Pierce |
| 4 | 1/6/2016 | 14:42 | 1/6/2016 | | 15:49 | Business | Fort Pierce |
| 5  ... | 1/6/2016 | 17:15  ... | 1/6/2016 | | 17:19  ... | Business  ... | West Palm Beach  ... |
| 1150 | 12/31/2016 1:07 | | 12/31/2016 1:14 | | | Business | Kar?chi |
| 1151 | 12/31/2016 | 13:24 | 12/31/2016 | | 13:42 | Business | Kar?chi |
| 1152 | 12/31/2016 | 15:03 | 12/31/2016 | | 15:38 | Business | Unknown Location |
| 1153 | 12/31/2016 | 21:32 | 12/31/2016 | | 21:50 | Business | Katunayake |
| 1154 | 12/31/2016 | 22:08 | 12/31/2016 | | 23:51 | Business | Gampaha |
|  | STOP\* | | MILES\* | PURPOSE\* | | | |
| 0 | Fort Pierce | | 5.1 | Meal/Entertain | | | |
| 2 | Fort Pierce | | 4.8 | Errand/Supplies | | | |
| 3 | Fort Pierce | | 4.7 | Meeting | | | |
| 4 | West Palm Beach | | 63.7 | Customer Visit | | | |
| 5 | West Palm Beach | | 4.3 | Meal/Entertain | | | |
| ... | ... | | ... | ... | | | |
| 1150 | Kar?chi | | 0.7 | Meeting | | | |
| 1151 | Unknown Location | | 3.9 | Temporary Site | | | |
| 1152 | Unknown Location | | 16.2 | Meeting | | | |
| 1153 | Gampaha | | 6.4 | Temporary Site | | | |
| 1154 | Ilukwatta | | 48.2 | Temporary Site | | | |

[653 rows x 7 columns]

# fill missing values (propose column with unknown value)

In [12]: ud**.**loc[:, 'PURPOSE\*'] **=** ud['PURPOSE\*']**.**fillna('Unknown') print(ud)

print(" ")

missing\_values\_in\_purpose **=** ud['PURPOSE\*']**.**isnull()**.**sum() print (missing\_values\_in\_purpose)

START\_DATE\* END\_DATE\* CATEGORY\* START\* \

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 0 | 1/1/2016 | 21:11 | 1/1/2016 | | 21:17 | Business | Fort Pierce |
| 2 | 1/2/2016 | 20:25 | 1/2/2016 | | 20:38 | Business | Fort Pierce |
| 3 | 1/5/2016 | 17:31 | 1/5/2016 | | 17:45 | Business | Fort Pierce |
| 4 | 1/6/2016 | 14:42 | 1/6/2016 | | 15:49 | Business | Fort Pierce |
| 5  ... | 1/6/2016 | 17:15  ... | 1/6/2016 | | 17:19  ... | Business  ... | West Palm Beach  ... |
| 1150 | 12/31/2016 1:07 | | 12/31/2016 1:14 | | | Business | Kar?chi |
| 1151 | 12/31/2016 | 13:24 | 12/31/2016 | | 13:42 | Business | Kar?chi |
| 1152 | 12/31/2016 | 15:03 | 12/31/2016 | | 15:38 | Business | Unknown Location |
| 1153 | 12/31/2016 | 21:32 | 12/31/2016 | | 21:50 | Business | Katunayake |
| 1154 | 12/31/2016 | 22:08 | 12/31/2016 | | 23:51 | Business | Gampaha |
|  | STOP\* | | MILES\* | PURPOSE\* | | | |
| 0 | Fort Pierce | | 5.1 | Meal/Entertain | | | |
| 2 | Fort Pierce | | 4.8 | Errand/Supplies | | | |
| 3 | Fort Pierce | | 4.7 | Meeting | | | |
| 4 | West Palm Beach | | 63.7 | Customer Visit | | | |
| 5 | West Palm Beach | | 4.3 | Meal/Entertain | | | |
| ... | ... | | ... | ... | | | |
| 1150 | Kar?chi | | 0.7 | Meeting | | | |
| 1151 | Unknown Location | | 3.9 | Temporary Site | | | |
| 1152 | Unknown Location | | 16.2 | Meeting | | | |
| 1153 | Gampaha | | 6.4 | Temporary Site | | | |
| 1154 | Ilukwatta | | 48.2 | Temporary Site | | | |
| [653 | rows x 7 columns] | |  |  | | | |
| 0 |  | |  |  | | | |

# Check and remove duplicates

In [13]: duplicates **=** ud**.**duplicated()#.sum() print(duplicates) print(ud[duplicates])

print(" ")

0 False

1. False
2. False
3. False
4. False

...

1150 False

1151 False

1152 False

1153 False

1154 False

Length: 653, dtype: bool

START\_DATE\* END\_DATE\* CATEGORY\* START\* STOP\* MILES\*

PURPOSE\*

492 6/28/2016 23:34 6/28/2016 23:59 Business Durham Cary 9.9

Meeting

In [14]: ud\_clean**=**ud**.**drop\_duplicates() print(ud\_clean**.**shape[0])

652

# Convert START\_DATE and END\_DATE to datetime

In [17]: ud**.**loc[:, 'START\_DATE\*'] **=** pd**.**to\_datetime(ud['START\_DATE\*']) In [20]: ud**.**loc[:,'END\_DATE\*'] **=** pd**.**to\_datetime(ud['END\_DATE\*'])

ud**.**dtypes

Out[20]: START\_DATE\* datetime64[ns] END\_DATE\* datetime64[ns]

CATEGORY\* object

START\* object

STOP\* object

MILES\* float64

PURPOSE\* object

dtype: object

# Total number of rides per category:

In [21]: rides\_per\_categoty**=** ud**.**groupby('CATEGORY\*')**.**size() print(rides\_per\_categoty)

CATEGORY\*

Business 647

Personal 6

dtype: int64

# Total miles traveled for each purpose:

In [22]: total\_miles **=** ud**.**groupby('PURPOSE\*')['MILES\*']**.**sum() print(total\_miles)

PURPOSE\*

Airport/Travel 16.5

Between Offices 197.0

Charity ($) 15.1

Commute 180.2

Customer Visit 2089.5

Errand/Supplies 508.0

Meal/Entertain 911.7

Meeting 2851.3

Moving 18.2

Temporary Site 523.7

Name: MILES\*, dtype: float64

# Average distance for business vs. personal rides:

In [23]: avg\_dist**=** ud**.**groupby('PURPOSE\*')['MILES\*']**.**mean() print(avg\_dist)

|  |  |
| --- | --- |
| PURPOSE\* |  |
| Airport/Travel | 5.500000 |
| Between Offices | 10.944444 |
| Charity ($) | 15.100000 |
| Commute | 180.200000 |
| Customer Visit | 20.688119 |
| Errand/Supplies | 3.968750 |
| Meal/Entertain | 5.698125 |
| Meeting | 15.247594 |
| Moving | 4.550000 |
| Temporary Site | 10.474000 |

Name: MILES\*, dtype: float64

# Add a column for cost estimation (assuming $2 per mile):

In [25]: ud**.**loc[:, 'COST\_ESTIMATION\*'] **=** ud['MILES\*'] **\*** 2 print(ud[['START\_DATE\*', 'MILES\*', 'COST\_ESTIMATION\*']]**.**head())

START\_DATE\* MILES\* COST\_ESTIMATION\*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 0 | 2016-01-01 | 21:11:00 | 5.1 | 10.2 |
| 2 | 2016-01-02 | 20:25:00 | 4.8 | 9.6 |
| 3 | 2016-01-05 | 17:31:00 | 4.7 | 9.4 |
| 4 | 2016-01-06 | 14:42:00 | 63.7 | 127.4 |
| 5 | 2016-01-06 | 17:15:00 | 4.3 | 8.6 |

# Filter rides longer than 50 miles:

In [26]: long\_rides **=** ud[ud['MILES\*'] **>** 50] print(long\_rides)

START\_DATE\* END\_DATE\* CATEGORY\* START\* \

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 4 | 2016-01-06 | | 14:42:00 | 2016-01-06 | | 15:49:00 | Business | Fort Pierce |
| 232 | 2016-03-17 | | 12:52:00 | 2016-03-17 | | 15:11:00 | Business | Austin |
| 251 | 2016-03-19 | | 19:33:00 | 2016-03-19 | | 20:39:00 | Business | Galveston |
| 268 | 2016-03-25 | | 13:24:00 | 2016-03-25 | | 16:22:00 | Business | Cary |
| 269 | 2016-03-25 | | 16:52:00 | 2016-03-25 | | 22:22:00 | Business | Latta |
| 270 | 2016-03-25 | | 22:54:00 | 2016-03-26 | | 01:39:00 | Business | Jacksonville |
| 295 | 2016-04-02 | | 12:21:00 | 2016-04-02 | | 14:47:00 | Business | Kissimmee |
| 296 | 2016-04-02 | | 16:57:00 | 2016-04-02 | | 18:09:00 | Business | Daytona Beach |
| 297 | 2016-04-02 | | 19:38:00 | 2016-04-02 | | 22:36:00 | Business | Jacksonville |
| 298 | 2016-04-02 | | 23:11:00 | 2016-04-03 | | 01:34:00 | Business | Ridgeland |
| 299 | 2016-04-03 | | 02:00:00 | 2016-04-03 | | 04:16:00 | Business | Florence |
| 559 | 2016-07-17 | | 12:20:00 | 2016-07-17 | | 15:25:00 | Personal | Boone |
| 869 | 2016-10-28 | | 15:53:00 | 2016-10-28 | | 17:59:00 | Business | Cary |
| 870 | 2016-10-28 | | 18:13:00 | 2016-10-28 | | 20:07:00 | Business | Winston Salem |
| 871 | 2016-10-28 | | 20:13:00 | 2016-10-28 | | 22:00:00 | Business | Asheville |
| 1088 | 2016-12-21 | | 20:56:00 | 2016-12-21 | | 23:42:00 | Business | Rawalpindi |
|  |  | STOP\* | | MILES\* | PURPOSE\* | | COST\_ESTIMATION\* | |
| 4 | West | Palm Beach | | 63.7 | Customer Visit | | 127.4 | |
| 232 |  | Katy | | 136.0 | Customer Visit | | 272.0 | |
| 251 |  | Houston | | 57.0 | Customer Visit | | 114.0 | |
| 268 |  | Latta | | 144.0 | Customer Visit | | 288.0 | |
| 269 Jacksonville 310.3 Customer Visit 620.6 | | | | | | | | |
| 270 | Kissimmee | | | 201.0 | Meeting | | 402.0 | |
| 295 | Daytona Beach | | | 77.3 | Customer Visit | | 154.6 | |
| 296 | Jacksonville | | | 80.5 | Customer Visit | | 161.0 | |
| 297 | Ridgeland | | | 174.2 | Customer Visit | | 348.4 | |
| 298 | Florence | | | 144.0 | Meeting | | 288.0 | |
| 299 | Cary | | | 159.3 | Meeting | | 318.6 | |
| 559 | Cary | | | 180.2 | Commute | | 360.4 | |
| 869 | Winston Salem | | | 107.0 | Meeting | | 214.0 | |
| 870 | Asheville | | | 133.6 | Meeting | | 267.2 | |
| 871 | Topton | | | 91.8 | Meeting | | 183.6 | |
| 1088 | Unknown Location | | | 103.0 | Meeting | | 206.0 | |

# Filter by specific purpose (e.g., meetings):

In [27]: meeting\_rides **=** ud[ud['PURPOSE\*'] **==** 'Meeting'] print(meeting\_rides)

START\_DATE\* END\_DATE\* CATEGORY\* START\* \

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 3 | 2016-01-05 | 17:31:00 | 2016-01-05 | | 17:45:00 | | Business | Fort Pierce |
| 6 | 2016-01-06 | 17:30:00 | 2016-01-06 | | 17:35:00 | | Business | West Palm Beach |
| 7 | 2016-01-07 | 13:27:00 | 2016-01-07 | | 13:33:00 | | Business | Cary |
| 8 | 2016-01-10 | 08:05:00 | 2016-01-10 | | 08:25:00 | | Business | Cary |
| 10  ... 1142 | 2016-01-10  2016-12-29 | 15:08:00  ... 20:15:00 | 2016-01-10  2016-12-29 | | 15:51:00  ... 20:45:00 | | Business  ...  Business | New York  ...  Kar?chi |
| 1144 | 2016-12-29 | 23:14:00 | 2016-12-29 | | 23:47:00 | | Business | Unknown Location |
| 1148 | 2016-12-30 | 16:45:00 | 2016-12-30 | | 17:08:00 | | Business | Kar?chi |
| 1150 | 2016-12-31 | 01:07:00 | 2016-12-31 | | 01:14:00 | | Business | Kar?chi |
| 1152 | 2016-12-31 | 15:03:00 | 2016-12-31 | | 15:38:00 | | Business | Unknown Location |
| 3 | Fort | STOP\*  Pierce | MILES\*  4.7 | PURPOSE\*  Meeting | | COST\_ESTIMATION\*  9.4 | | |
| 6 | Palm Beach | | 7.1 | Meeting | | 14.2 | | |
| 7 | Cary | | 0.8 | Meeting | | 1.6 | | |
| 8 | Morrisville | | 8.3 | Meeting | | 16.6 | | |
| 10 | Queens | | 10.8 | Meeting | | 21.6 | | |
| ... | ... | | ... | ... | | ... | | |
| 1142 | Kar?chi | | 7.2 | Meeting | | 14.4 | | |
| 1144 | Kar?chi | | 12.9 | Meeting | | 25.8 | | |
| 1148 | Kar?chi | | 4.6 | Meeting | | 9.2 | | |
| 1150 | Kar?chi | | 0.7 | Meeting | | 1.4 | | |
| 1152 | Unknown Location | | 16.2 | Meeting | | 32.4 | | |
| [187 | rows x 8 columns] | |  |  | |  | | |

# What is the total number of business trips versus personal trips?

In [28]: category\_counts **=** ud['CATEGORY\*']**.**value\_counts()

print(category\_counts) CATEGORY\*

Business 647

Personal 6

Name: count, dtype: int64

# What percentage of trips are business versus personal?

In [29]: category\_counts **=** ud['CATEGORY\*']**.**value\_counts() total\_trips **=** len(ud)

In [30]: business\_percentage **=** (category\_counts**.**get('Business', 0) **/** total\_tr personal\_percentage **=** (category\_counts**.**get('Personal', 0) **/** total\_tr print(business\_percentage)

print(personal\_percentage)

99.08116385911178

0.9188361408882083

# Display records with Miles greater than 10 and city Cary or New York

In [ ]: df1**=**ud**.**loc[(ud['MILES\*']**>**10) **&**

ud['START\*']**.**isin(['Cary','New York'])]

print(df1)

In [33]: df1**.**count()

Out[33]: START\_DATE\* 50

END\_DATE\* 50

CATEGORY\* 50

START\* 50

STOP\* 50

MILES\* 50

PURPOSE\* 50

COST\_ESTIMATION\* 50

dtype: int64

In [38]: uf **=** pd**.**read\_csv("Uber.csv")

In [39]: uf**.**sort\_values(by**=**'MILES\*', ascending**=True**)

Out[39]: **START\_DATE\* END\_DATE\* CATEGORY\* START\* STOP\* MILES\* PURP**

**420** 6/8/2016 17:16 6/8/2016

17:18

Business Soho Tribeca 0.5 Er

Sup

**44** 1/26/2016

17:27

**120** 2/17/2016

16:38

**1111** 12/25/2016

0:10

**1110** 12/24/2016

22:04

1/26/2016

17:29

2/17/2016

16:43

12/25/2016

0:14

12/24/2016

22:09

Business Cary Cary 0.5 Er

Sup

Business Katunayaka Katunayaka 0.5 Er

Sup

Business Lahore Lahore 0.6 Er

Sup

Business Lahore Lahore 0.6 Er

Sup

**...** ... ... ... ... ... ...

|  |  |
| --- | --- |
| 9/27/2016  21:01 | 9/28/2016  2:37 |
| 10/30/2016 | 10/30/2016 |
| 15:22 | 18:23 |
| 3/25/2016 | 3/26/2016 |
| 22:54 | 1:39 |
| 3/25/2016 | 3/25/2016 |

**776**

Business Unknown Location

Unknown Location

195.6

**881**

**270**

**269**

16:52

Business Asheville Mebane 195.9

Business Jacksonville Kissimmee 201.0 Me 22:22 Business Latta Jacksonville 310.3 Cust

**1155** Totals NaN NaN NaN NaN 12204.7

## 1156 rows × 7 columns

In [400]: uf**.**sort\_values(by**=**['START\_DATE\*','MILES\*'],ascending**=**[**True**,**False**])

Out[40]: **START\_DATE\* END\_DATE\* CATEGORY\* START\* STOP\* MILES\* PURPOSE**

1/1/2016 21:11 1/1/2016

|  |  |
| --- | --- |
|  | **0** |
| **9** |
| **10** |
| **11** |
| **12**  **... 779** |
| **780** |
|  | **750** |

21:17

Business Fort Pierce

Fort Pierce

* 1. Mea

Entertai

1/10/2016

12:17

1/10/2016

15:08

1/10/2016

18:18

1/10/2016

12:44

1/10/2016

15:51

1/10/2016

18:53

Business Jamaica New York 16.5 Custom

Vis

Business New York Queens 10.8 Meetin

Business Elmhurst New York 7.5 Meetin

1/10/2016

19:12

1/10/2016

19:32

Business Midtown East Harlem

6.2 Meetin

... ... ... ... ... ... .

9/30/2016

17:39

9/30/2016

20:20

Business Islamabad Islamabad 37.7 Na

9/30/2016

20:59

9/30/2016

22:34

Business Islamabad Unknown

Location

16.7 Na

9/5/2016 10:25 9/5/2016

10:44

Business Unknown

Location

R?

walpindi

17.2 Na

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **751** 9/6/2016 17:49 | | 9/6/2016  17:49 | Business | Unknown Location | Unknown Location | 69.1 | Na |
| **1155** | Totals | NaN | NaN | NaN | NaN | 12204.7 | Na |

## 1156 rows × 7 columns

In [50]: **import** numpy **as** np

**import** pandas **as** pd

uber **=** pd**.**read\_csv("Uber.csv")

uber['MILES\_CAT'] **=** np**.**where(uber['MILES\*'] **>** 100, 'Long trip', 'Sho (uber**.**head())

Out[58]: **START\_DATE\* END\_DATE\* CATEGORY\* START\* STOP\* MILES\* PURPOSE\* MILE**

**0** 1/1/2016 21:11 1/1/2016

21:17

Business Fort Pierce

Fort Pierce

5.1 Meal/

Entertain

Sh

**1** 1/2/2016 1:25 1/2/2016

1:37

Business Fort Pierce

Fort Pierce

5.0 NaN Sh

**2** 1/2/2016 20:25 1/2/2016

20:38

Business Fort Pierce

Fort Pierce

4.8 Errand/ Supplies

**3** 1/5/2016 17:31 1/5/2016

Sh

17:45

Business Fort Pierce

Fort Pierce

4.7 Meeting Sh

**4** 1/6/2016 14:42 1/6/2016

15:49

Business Fort Pierce

West Palm Beach

63.7 Customer Visit

Sh

In [59]: uber**.**tail()

Out[59]: **START\_DATE\* END\_DATE\* CATEGORY\* START\* STOP\* MILES\* PURPOS**

**1151** 12/31/2016

13:24

12/31/2016

13:42

Business Kar?chi Unknown

Location

3.9 Tempora S

**1152** 12/31/2016

15:03

12/31/2016

15:38

Business Unknown Location

Unknown Location

16.2 Meeti

**1153** 12/31/2016

21:32

12/31/2016

21:50

Business Katunayake Gampaha 6.4 Tempora

S

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **1154** | 12/31/2016  22:08 | 12/31/2016  23:51 | Business | Gampaha | Ilukwatta | 48.2 | Tempora  S |
| **1155** | Totals | NaN | NaN | NaN | NaN | 12204.7 | N |

In [61]: uber['nc']**=**10 uber

Out[61]: **START\_DATE\* END\_DATE\* CATEGORY\* START\* STOP\* MILES\* PURPOS**

**0** 1/1/2016 21:11 1/1/2016

21:17

**1** 1/2/2016 1:25 1/2/2016

1:37

**2** 1/2/2016 20:25 1/2/2016

20:38

**3** 1/5/2016 17:31 1/5/2016

17:45

**4** 1/6/2016 14:42 1/6/2016

15:49

Business Fort Pierce Fort Pierce

Business Fort Pierce Fort Pierce

Business Fort Pierce Fort Pierce

Business Fort Pierce Fort Pierce

West Business Fort Pierce Palm

Beach

5.1 Me

Entert

5.0 N

4.8 Erran Suppli

4.7 Meeti

63.7 Custom Vi

**...** ... ... ... ... ... ...

**1151** 12/31/2016

13:24

12/31/2016

13:42

Business Kar?chi Unknown

Location

3.9 Tempora S

**1152** 12/31/2016

15:03

12/31/2016

15:38

Business Unknown Location

Unknown Location

16.2 Meeti

**1153** 12/31/2016

21:32

12/31/2016

21:50

Business Katunayake Gampaha 6.4 Tempora

S

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **1154** | 12/31/2016  22:08 | 12/31/2016  23:51 | Business | Gampaha | Ilukwatta | 48.2 | Tempora  S |
| **1155** | Totals | NaN | NaN | NaN | NaN | 12204.7 | N |

## 1156 rows × 9 columns

In [80]: uber['Trip']**=**np**.**where(

uber['MILES\*'] **<=**100,

"Short trips",

np**.**where(uber['MILES\*']**<=**200, "Medium Trip","Lo

uber

Out[80]: **START\_DATE\* END\_DATE\* CATEGORY\* START\* STOP\* MILES\* PURPOS**

**0** 1/1/2016 21:11 1/1/2016

21:17

**1** 1/2/2016 1:25 1/2/2016

1:37

**2** 1/2/2016 20:25 1/2/2016

20:38

**3** 1/5/2016 17:31 1/5/2016

17:45

**4** 1/6/2016 14:42 1/6/2016

15:49

Business Fort Pierce Fort Pierce

Business Fort Pierce Fort Pierce

Business Fort Pierce Fort Pierce

Business Fort Pierce Fort Pierce

West Business Fort Pierce Palm

Beach

5.1 Me

Entert

5.0 N

4.8 Erran Suppli

4.7 Meeti

63.7 Custom Vi

**...** ... ... ... ... ... ...

**1151** 12/31/2016

13:24

12/31/2016

13:42

Business Kar?chi Unknown

Location

3.9 Tempora S

**1152** 12/31/2016

15:03

12/31/2016

15:38

Business Unknown Location

Unknown Location

16.2 Meeti

**1153** 12/31/2016

21:32

12/31/2016

21:50

Business Katunayake Gampaha 6.4 Tempora

S

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **1154** | 12/31/2016  22:08 | 12/31/2016  23:51 | Business | Gampaha | Ilukwatta | 48.2 | Tempora  S |
| **1155** | Totals | NaN | NaN | NaN | NaN | 12204.7 | N |

## 1156 rows × 10 columns

In [82]: medium\_trips **=** uber[uber['Trip'] **==** 'Medium Trip'] medium\_trips

Out[82]: **START\_DATE\* END\_DATE\* CATEGORY\* START\* STOP\* MILES\* PURPOS**

3/17/2016

|  |  |
| --- | --- |
|  | **232** |
| **268** |
| **297** |
| **298** |
| **299** |
| **546** |
| **559** |
| **727** |
| **776** |
| **788** |
| **869** |
| **870** |
| **881** |
|  | **1088** |

12:52

3/25/2016

13:24

3/17/2016

15:11

3/25/2016

16:22

Business Austin Katy 136.0 Custom

V

Business Cary Latta 144.0 Custom

V

4/2/2016 19:38 4/2/2016

22:36

4/2/2016 23:11 4/3/2016

1:34

4/3/2016 2:00 4/3/2016

4:16

Business Jacksonville Ridgeland 174.2 Custom

V

Business Ridgeland Florence 144.0 Meeti

Business Florence Cary 159.3 Meeti

7/14/2016

16:39

7/14/2016

20:05

Business Morrisville Banner

Elk

195.3 N

7/17/2016

12:20

7/17/2016

15:25

Personal Boone Cary 180.2 Comm

8/27/2016

16:15

9/27/2016

21:01

8/27/2016

19:13

9/28/2016

2:37

Business Unknown Location

Business Unknown Location

Unknown 156.9 N Location

Unknown 195.6 N Location

10/6/2016

17:23

10/28/2016

15:53

10/6/2016

17:40

10/28/2016

17:59

Business R?walpindi Unknown 112.6 N

Location

Business Cary Winston 107.0 Meeti

Salem

10/28/2016

18:13

10/28/2016

20:07

Business Winston Salem

Asheville 133.6 Meeti

10/30/2016

15:22

10/30/2016

18:23

Business Asheville Mebane 195.9 N

12/21/2016

20:56

12/21/2016

23:42

Business Rawalpindi Unknown

Location

103.0 Meeti

In [78]: medium\_trips**.**count()

Out[78]: START\_DATE\* 14

|  |  |
| --- | --- |
| END\_DATE\* | 14 |
| CATEGORY\* | 14 |
| START\* | 14 |
| STOP\* | 14 |
| MILES\* | 14 |
| PURPOSE\* | 9 |
| MILES\_CAT | 14 |
| nc | 14 |
| Trip | 14 |
| dtype: int64 |  |

In [64]: long\_trips**=**uber[uber['Trip']**==**'Long Trip'] long\_trips

Out[64]: **START\_DATE\* END\_DATE\* CATEGORY\* START\* STOP\* MILES\* PURP**

**269** 3/25/2016

16:52

**270** 3/25/2016

22:54

3/25/2016

22:22

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Business | Latta | Jacksonville | 310.3 | Cust |
| Business | Jacksonville | Kissimmee | 201.0 | Me |
| NaN | NaN | NaN | 12204.7 |  |

3/26/2016

1:39

**1155** Totals NaN

In [79]: long\_trips**.**count()

Out[79]: START\_DATE\* 3

|  |  |
| --- | --- |
| END\_DATE\* | 2 |
| CATEGORY\* | 2 |
| START\* | 2 |
| STOP\* | 2 |
| MILES\* | 3 |
| PURPOSE\* | 2 |
| MILES\_CAT | 3 |
| nc | 3 |
| Trip | 3 |
| dtype: int64 |  |

In [69]: short\_trips**=**uber[uber['Trip']**==**'Short trips'] short\_trips

Out[69]: **START\_DATE\* END\_DATE\* CATEGORY\* START\* STOP\* MILES\* PURPOS**

**0** 1/1/2016 21:11 1/1/2016

21:17

**1** 1/2/2016 1:25 1/2/2016

1:37

**2** 1/2/2016 20:25 1/2/2016

20:38

**3** 1/5/2016 17:31 1/5/2016

17:45

1/6/2016 14:42 1/6/2016

|  |  |
| --- | --- |
|  | **4** |
| **...**  **1150** |
| **1151** |
| **1152** |
| **1153** |
|  | **1154** |

15:49

Business Fort Pierce Fort Pierce

Business Fort Pierce Fort Pierce

Business Fort Pierce Fort Pierce

Business Fort Pierce Fort Pierce

West Business Fort Pierce Palm

Beach

5.1 Me

Enterta

5.0 Na

4.8 Erran Suppli

4.7 Meeti

63.7 Custom Vi

... ... ... ... ... ...

12/31/2016

1:07

12/31/2016

1:14

Business Kar?chi Kar?chi 0.7 Meeti

12/31/2016

13:24

12/31/2016

13:42

Business Kar?chi Unknown

Location

3.9 Tempora Si

12/31/2016

15:03

12/31/2016

15:38

Business Unknown Location

Unknown Location

16.2 Meeti

12/31/2016

21:32

12/31/2016

22:08

12/31/2016

21:50

12/31/2016

23:51

Business Katunayake Gampaha 6.4 Tempora

Si

Business Gampaha Ilukwatta 48.2 Tempora

Si

## 1139 rows × 10 columns

In [83]: short\_trips**.**count()

Out[83]: START\_DATE\* 1139

|  |  |
| --- | --- |
| END\_DATE\* | 1139 |
| CATEGORY\* | 1139 |
| START\* | 1139 |
| STOP\* | 1139 |
| MILES\* | 1139 |
| PURPOSE\* | 642 |
| MILES\_CAT | 1139 |
| nc | 1139 |
| Trip | 1139 |
| dtype: int64 |  |

In [85]: all**=** uber['Trip']**.**value\_counts() all

Out[85]: Trip

Short trips 1139

Medium Trip 14

Long Trip 3

Name: count, dtype: int64

In [87]: uber**.**groupby('START\*')['MILES\*']**.**agg('mean')

Out[87]: START\*

Agnew 2.775000

Almond 15.200000

Apex 5.341176

Arabi 17.000000

Arlington 4.900000

...

West University 2.200000

Weston 4.000000

Westpark Place 2.182353

Whitebridge 4.020588

Winston Salem 133.600000

Name: MILES\*, Length: 177, dtype: float64

# Find Average Miles of each purpose

In [960]: grouped **=** uber**.**groupby('CATEGORY\*')['MILES\*']**.**agg(['sum', 'mean', ' print(grouped)

|  |  |  |  |
| --- | --- | --- | --- |
| CATEGORY\* | sum | mean | max |
| Business | 11487.0 | 10.655844 | 310.3 |
| Personal | 717.7 | 9.320779 | 180.2 |

In [980]: grouped **=** uber**.**groupby('START\*')['nc']**.**agg(['sum', 'mean', 'max']) grouped

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Out[98]: | **START\*** | **sum** | **mean** | **max** |
|  | **Agnew** | 40 | 10.0 | 10 |
|  | **Almond** | 10 | 10.0 | 10 |
|  | **Apex** | 170 | 10.0 | 10 |
|  | **Arabi** | 10 | 10.0 | 10 |
|  | **Arlington** | 10 | 10.0 | 10 |
|  | **...** | ... | ... | ... |
|  | **West University** | 20 | 10.0 | 10 |
|  | **Weston** | 20 | 10.0 | 10 |
|  | **Westpark Place** | 170 | 10.0 | 10 |
|  | **Whitebridge** | 680 | 10.0 | 10 |
|  | **Winston Salem** | 10 | 10.0 | 10 |

## 177 rows × 3 columns

In [ ]: