In [1]:

```
import pandas as pd
import numpy as np
```

In [2]:

```
weather = pd.read_csv("data/weather.csv")
weather.head()
```

Out[2]:

	year	month	element	day1	day2	day3	day4	day5	day6	
0	2018	1	max	17.573016	19.796815	22.412495	17.813163	20.165825	17.060539	22.
1	2018	1	min	22.725760	21.007865	17.730792	18.045290	20.766734	18.656651	22.
2	2018	2	max	19.015120	19.261805	17.510713	21.080425	17.915749	19.082145	18.
3	2018	2	min	18.653843	22.818600	21.842673	21.958159	22.523078	18.535469	19.
4	2018	3	max	20.741115	19.704016	17.039811	20.703908	22.714125	17.205000	19.

5 rows × 34 columns

In [3]:

```
weather.shape
```

Out[3]:

```
(22, 34)
```

In [4]:

```
weather.stack()
```

Out[4]:

```
0   year            2018
    month              1
    element          max
    day1        17.573016
    day2        19.796815
                   ...
21  day27       18.744409
    day28       20.488991
    day29       21.662129
    day30       20.137534
    day31       18.775539
Length: 748, dtype: object
```

In [ ]:

```
```

In [7]:

```
weather_melt = pd.melt(weather, id_vars=['year', 'month', 'element'], var_name="day"

weather_melt
```

Out[7]:

	year	month	element	day	temp
0	2018	1	max	day1	17.573016
1	2018	1	min	day1	22.725760
2	2018	2	max	day1	19.015120
3	2018	2	min	day1	18.653843
4	2018	3	max	day1	20.741115
...	...	...	...	...	...
677	2018	10	min	day31	21.691537
678	2018	11	max	day31	20.750438
679	2018	11	min	day31	18.939767
680	2018	12	max	day31	19.648924
681	2018	12	min	day31	18.775539

682 rows × 5 columns

In [10]:

```
weather_tidy =  weather_melt.pivot_table(index= ['year', 'month', 'day'], columns='e
weather_tidy
```

Out[10]:

		element	max	min
year	month	day		
2018	1	day1	17.573016	22.725760
		day10	19.067288	19.931129
		day11	19.361002	22.598325
		day12	20.982134	17.715137
		day13	21.668005	17.940334
	...	...	...	...
	12	day5	21.375349	20.865535
		day6	17.992885	20.310116
		day7	19.683359	20.531823
		day8	20.477046	19.310346
		day9	20.210640	22.820992

341 rows × 2 columns

In [12]:

```
weather_tidy.loc[(2018,1, 'day1')]['max']
```

Out[12]:

```
17.573016045594
```

In [13]:

```
weather_tidy = weather_tidy.reset_index()
weather_tidy
```

Out[13]:

element	year	month	day	max	min
0	2018	1	day1	17.573016	22.725760
1	2018	1	day10	19.067288	19.931129
2	2018	1	day11	19.361002	22.598325
3	2018	1	day12	20.982134	17.715137
4	2018	1	day13	21.668005	17.940334
...	...	...	...	...	...
336	2018	12	day5	21.375349	20.865535
337	2018	12	day6	17.992885	20.310116
338	2018	12	day7	19.683359	20.531823
339	2018	12	day8	20.477046	19.310346
340	2018	12	day9	20.210640	22.820992

341 rows × 5 columns

In [14]:

```
weather_tidy.loc[0]
```

Out[14]:

```
element
year          2018
month            1
day           day1
max       17.573016
min        22.72576
Name: 0, dtype: object
```

In [ ]:

```

```

In [15]:

```
flights = pd.read_csv('https://raw.githubusercontent.com/mwaskom/seaborn-data/master
flights.shape
```

Out[15]:

```
(144, 3)
```

In [16]:

```
flights.head()
```

Out[16]:

	year	month	passengers
0	1949	January	112
1	1949	February	118
2	1949	March	132
3	1949	April	129
4	1949	May	121

In [20]:

```
flight_pivot = flights.pivot_table(index='year', columns='month', values='passengers
flight_pivot
```

Out[20]:

month	year	April	August	December	February	January	July	June	March	May	November
0	1949	129	148	118	118	112	148	135	132	121	104
1	1950	135	170	140	126	115	170	149	141	125	114
2	1951	163	199	166	150	145	199	178	178	172	146
3	1952	181	242	194	180	171	230	218	193	183	172
4	1953	235	272	201	196	196	264	243	236	229	180
5	1954	227	293	229	188	204	302	264	235	234	203
6	1955	269	347	278	233	242	364	315	267	270	237
7	1956	313	405	306	277	284	413	374	317	318	271
8	1957	348	467	336	301	315	465	422	356	355	305
9	1958	348	505	337	318	340	491	435	362	363	310
10	1959	396	559	405	342	360	548	472	406	420	362
11	1960	461	606	432	391	417	622	535	419	472	390

In [22]:

```
pd.melt(flight_pivot, id_vars=['year'], value_name='passangers')
```

Out[22]:

	year	month	passangers
0	1949	April	129
1	1950	April	135
2	1951	April	163
3	1952	April	181
4	1953	April	235
...	...	...	...
139	1956	September	355
140	1957	September	404
141	1958	September	404
142	1959	September	463
143	1960	September	508

144 rows × 3 columns

convert to a csv file

In [27]:

```
weather_tidy.head()
```

Out[27]:

	element	year	month	day	max	min
0	2018	1	day1	17.573016	22.725760
1	2018	1	day10	19.067288	19.931129
2	2018	1	day11	19.361002	22.598325
3	2018	1	day12	20.982134	17.715137
4	2018	1	day13	21.668005	17.940334

In [28]:

```
# create output folder first.
weather_tidy.to_csv("output/preprocessed.csv", index=False)
```

In [29]:

```
pd.read_csv('output/preprocessed.csv')
```

Out[29]:

	year	month	day	max	min
0	2018	1	day1	17.573016	22.725760
1	2018	1	day10	19.067288	19.931129
2	2018	1	day11	19.361002	22.598325
3	2018	1	day12	20.982134	17.715137
4	2018	1	day13	21.668005	17.940334
...	...	...	...	...	...
336	2018	12	day5	21.375349	20.865535
337	2018	12	day6	17.992885	20.310116
338	2018	12	day7	19.683359	20.531823
339	2018	12	day8	20.477046	19.310346
340	2018	12	day9	20.210640	22.820992

341 rows × 5 columns

In [ ]:

```

```

In [30]:

```
weather_tidy
```

Out[30]:

element	year	month	day	max	min
0	2018	1	day1	17.573016	22.725760
1	2018	1	day10	19.067288	19.931129
2	2018	1	day11	19.361002	22.598325
3	2018	1	day12	20.982134	17.715137
4	2018	1	day13	21.668005	17.940334
...	...	...	...	...	...
336	2018	12	day5	21.375349	20.865535
337	2018	12	day6	17.992885	20.310116
338	2018	12	day7	19.683359	20.531823
339	2018	12	day8	20.477046	19.310346
340	2018	12	day9	20.210640	22.820992

341 rows × 5 columns

In [35]:

```
pd.cut(weather_tidy['max'], bins=[14,18,20,22,25]).value_counts()
```

Out[35]:

```
(20, 22]    121
(18, 20]    107
(22, 25]     68
(14, 18]     45
Name: max, dtype: int64
```

In [ ]:

```

```

Uber Data Analysis

In [36]:

```
data = pd.read_csv('data/UberData.csv')
```

In [37]:

```
data.head()
```

Out[37]:

	START_DATE*	END_DATE*	CATEGORY*	START*	STOP*	MILES*	PURPOSE*
0	1/1/2016 21:11	1/1/2016 21:17	Business	Fort Pierce	Fort Pierce	5.1	Meal/Entertain
1	1/2/2016 1:25	1/2/2016 1:37	Business	Fort Pierce	Fort Pierce	5.0	NaN
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 17:45	Business	Fort Pierce	Fort Pierce	4.7	Meeting
4	1/6/2016 14:42	1/6/2016 15:49	Business	Fort Pierce	West Palm Beach	63.7	Customer Visit

In [38]:

```
data.shape
```

Out[38]:

```
(1156, 7)
```

In [39]:

```
data.columns
```

Out[39]:

```
Index(['START_DATE*', 'END_DATE*', 'CATEGORY*', 'START*', 'STOP*', 'MI
LES*',
       'PURPOSE*'],
      dtype='object')
```

In [41]:

```
data.tail(10)
```

Out[41]:

	START_DATE*	END_DATE*	CATEGORY*	START*	STOP*	MILES*	PURPOSE*
1146	12/30/2016 11:31	12/30/2016 11:56	Business	Kar?chi	Kar?chi	2.9	Errand/Supplies
1147	12/30/2016 15:41	12/30/2016 16:03	Business	Kar?chi	Kar?chi	4.6	Errand/Supplies
1148	12/30/2016 16:45	12/30/2016 17:08	Business	Kar?chi	Kar?chi	4.6	Meeting
1149	12/30/2016 23:06	12/30/2016 23:10	Business	Kar?chi	Kar?chi	0.8	Customer Visit
1150	12/31/2016 1:07	12/31/2016 1:14	Business	Kar?chi	Kar?chi	0.7	Meeting
1151	12/31/2016 13:24	12/31/2016 13:42	Business	Kar?chi	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
1155	Totals	NaN	NaN	NaN	NaN	12204.7	NaN

In [42]:

```
data.drop(1155, inplace=True)
```

In [43]:

```
data.tail()
```

Out[43]:

	START_DATE*	END_DATE*	CATEGORY*	START*	STOP*	MILES*	PURPOSE*
1150	12/31/2016 1:07	12/31/2016 1:14	Business	Kar?chi	Kar?chi	0.7	Meeting
1151	12/31/2016 13:24	12/31/2016 13:42	Business	Kar?chi	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

In [44]:

```
data.info()

# data.dtypes
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1155 entries, 0 to 1154
Data columns (total 7 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   START_DATE*  1155 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*       1155 non-null   float64
 6   PURPOSE*     653 non-null    object 
dtypes: float64(1), object(6)
memory usage: 72.2+ KB
```

Rename the columns

In [ ]:

```
# data.columns = ['START_DATE', 'END_DATE', 'CATEGORY', 'START', 'STOP', 'MILES','PU
```

In [48]:

```
data.columns = [col.replace("*", "") for col in data.columns]
```

In [49]:

```
data.head()
```

Out[49]:

	START_DATE	END_DATE	CATEGORY	START	STOP	MILES	PURPOSE
0	1/1/2016 21:11	1/1/2016 21:17	Business	Fort Pierce	Fort Pierce	5.1	Meal/Entertain
1	1/2/2016 1:25	1/2/2016 1:37	Business	Fort Pierce	Fort Pierce	5.0	NaN
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 17:45	Business	Fort Pierce	Fort Pierce	4.7	Meeting
4	1/6/2016 14:42	1/6/2016 15:49	Business	Fort Pierce	West Palm Beach	63.7	Customer Visit

In [50]:

```
# data.rename()
```

In [58]:

```
data['STOP'].apply(lambda x :len(x))
```

Out[58]:

```
0       11
1       11
2       11
3       11
4       15
        ..
1150     7
1151    16
1152    16
1153     7
1154     9
Name: STOP, Length: 1155, dtype: int64
```

In [62]:

```
data['STOP'].str.len()
```

Out[62]:

```
0       11
1       11
2       11
3       11
4       15
        ..
1150     7
1151    16
1152    16
1153     7
1154     9
Name: STOP, Length: 1155, dtype: int64
```

In [63]:

```
data['STOP'].str.lower()
```

Out[63]:

```
0             fort pierce
1             fort pierce
2             fort pierce
3             fort pierce
4          west palm beach
               ...
1150              kar?chi
1151    unknown location
1152    unknown location
1153              gampaha
1154            ilukwatta
Name: STOP, Length: 1155, dtype: object
```

In [65]:

```
data['STOP'].str.split()
```

Out[65]:

```
0             [Fort, Pierce]
1             [Fort, Pierce]
2             [Fort, Pierce]
3             [Fort, Pierce]
4       [West, Palm, Beach]
                ...
1150              [Kar?chi]
1151    [Unknown, Location]
1152    [Unknown, Location]
1153              [Gampaha]
1154            [Ilukwatta]
Name: STOP, Length: 1155, dtype: object
```

In [ ]:

```

```

Datetime

In [69]:

```
data.head()
```

Out[69]:

	START_DATE	END_DATE	CATEGORY	START	STOP	MILES	PURPOSE
0	1/1/2016 21:11	1/1/2016 21:17	Business	Fort Pierce	Fort Pierce	5.1	Meal/Entertain
1	1/2/2016 1:25	1/2/2016 1:37	Business	Fort Pierce	Fort Pierce	5.0	NaN
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 17:45	Business	Fort Pierce	Fort Pierce	4.7	Meeting
4	1/6/2016 14:42	1/6/2016 15:49	Business	Fort Pierce	West Palm Beach	63.7	Customer Visit

In [71]:

```
data['START_DATE'] =  pd.to_datetime(data['START_DATE'])
data['END_DATE'] =  pd.to_datetime(data['END_DATE'])
```

In [72]:

```
data.head()
```

Out[72]:

	START_DATE	END_DATE	CATEGORY	START	STOP	MILES	PURPOSE
0	2016-01-01 21:11:00	2016-01-01 21:17:00	Business	Fort Pierce	Fort Pierce	5.1	Meal/Entertain
1	2016-01-02 01:25:00	2016-01-02 01:37:00	Business	Fort Pierce	Fort Pierce	5.0	NaN
2	2016-01-02 20:25:00	2016-01-02 20:38:00	Business	Fort Pierce	Fort Pierce	4.8	Errand/Supplies
3	2016-01-05 17:31:00	2016-01-05 17:45:00	Business	Fort Pierce	Fort Pierce	4.7	Meeting
4	2016-01-06 14:42:00	2016-01-06 15:49:00	Business	Fort Pierce	West Palm Beach	63.7	Customer Visit

In [73]:

```
data.dtypes
```

Out[73]:

```
START_DATE    datetime64[ns]
END_DATE      datetime64[ns]
CATEGORY              object
START                 object
STOP                  object
MILES                float64
PURPOSE               object
dtype: object
```

In [76]:

```
ts = data['START_DATE'][0]
ts
```

Out[76]:

```
Timestamp('2016-01-01 21:11:00')
```

In [77]:

```
ts.year
```

Out[77]:

```
2016
```

In [78]:

```
ts.month
```

Out[78]:

```
1
```

In [79]:

```
ts.day
```

Out[79]:

```
1
```

In [81]:

```
ts.month_name()
```

Out[81]:

```
'January'
```

In [82]:

```
ts.day_name()
```

Out[82]:

```
'Friday'
```

In [83]:

```
ts.day_of_week
```

Out[83]:

```
4
```

In [85]:

```
ts.hour
```

Out[85]:

```
21
```

In [86]:

```
ts.minute
```

Out[86]:

```
11
```

In [93]:

```
data['YEAR'] = data['START_DATE'].dt.year
data['MONTH'] = data['START_DATE'].dt.month
data['DAY'] = data['START_DATE'].dt.day
```

In [94]:

```
data.head()
```

Out[94]:

	START_DATE	END_DATE	CATEGORY	START	STOP	MILES	PURPOSE	YEAR	MONTH
0	2016-01-01 21:11:00	2016-01-01 21:17:00	Business	Fort Pierce	Fort Pierce	5.1	Meal/Entertain	2016	1
1	2016-01-02 01:25:00	2016-01-02 01:37:00	Business	Fort Pierce	Fort Pierce	5.0	NaN	2016	1
2	2016-01-02 20:25:00	2016-01-02 20:38:00	Business	Fort Pierce	Fort Pierce	4.8	Errand/Supplies	2016	1
3	2016-01-05 17:31:00	2016-01-05 17:45:00	Business	Fort Pierce	Fort Pierce	4.7	Meeting	2016	1
4	2016-01-06 14:42:00	2016-01-06 15:49:00	Business	Fort Pierce	West Palm Beach	63.7	Customer Visit	2016	1

In [ ]:

```

```

In [96]:

```
data.describe()
```

Out[96]:

	MILES	YEAR	MONTH	DAY
count	1155.000000	1155.0	1155.000000	1155.000000
mean	10.566840	2016.0	6.982684	16.309091
std	21.579106	0.0	3.544915	8.713157
min	0.500000	2016.0	1.000000	1.000000
25%	2.900000	2016.0	3.500000	9.000000
50%	6.000000	2016.0	7.000000	17.000000
75%	10.400000	2016.0	10.000000	24.000000
max	310.300000	2016.0	12.000000	31.000000

In [98]:

```
data['MILES'].mean()
```

Out[98]:

```
10.566839826839812
```

In [ ]:

```
```

In [95]:

```
data['MONTH'].value_counts()
```

Out[95]:

```
12    146
8     133
11    122
2     115
3     113
7     112
6     108
10    106
1      61
4      54
5      49
9      36
Name: MONTH, dtype: int64
```

In [ ]:

```
```

In [102]:

```
data[data['MILES']<5].shape[0]/1155
```

Out[102]:

```
0.42943722943722945
```

In [103]:

```
data.head()
```

Out[103]:

	START_DATE	END_DATE	CATEGORY	START	STOP	MILES	PURPOSE	YEAR	MONTH
0	2016-01-01 21:11:00	2016-01-01 21:17:00	Business	Fort Pierce	Fort Pierce	5.1	Meal/Entertain	2016	1
1	2016-01-02 01:25:00	2016-01-02 01:37:00	Business	Fort Pierce	Fort Pierce	5.0	NaN	2016	1
2	2016-01-02 20:25:00	2016-01-02 20:38:00	Business	Fort Pierce	Fort Pierce	4.8	Errand/Supplies	2016	1
3	2016-01-05 17:31:00	2016-01-05 17:45:00	Business	Fort Pierce	Fort Pierce	4.7	Meeting	2016	1
4	2016-01-06 14:42:00	2016-01-06 15:49:00	Business	Fort Pierce	West Palm Beach	63.7	Customer Visit	2016	1

In [106]:

```
data['KMS'] = np.round(data['MILES']*1.609,1)
```

In [107]:

```
data.head()
```

Out[107]:

	START_DATE	END_DATE	CATEGORY	START	STOP	MILES	PURPOSE	YEAR	MONTH
0	2016-01-01 21:11:00	2016-01-01 21:17:00	Business	Fort Pierce	Fort Pierce	5.1	Meal/Entertain	2016	1
1	2016-01-02 01:25:00	2016-01-02 01:37:00	Business	Fort Pierce	Fort Pierce	5.0	NaN	2016	1
2	2016-01-02 20:25:00	2016-01-02 20:38:00	Business	Fort Pierce	Fort Pierce	4.8	Errand/Supplies	2016	1
3	2016-01-05 17:31:00	2016-01-05 17:45:00	Business	Fort Pierce	Fort Pierce	4.7	Meeting	2016	1
4	2016-01-06 14:42:00	2016-01-06 15:49:00	Business	Fort Pierce	West Palm Beach	63.7	Customer Visit	2016	1

In [ ]:

```

```

In [ ]:

```
```

In [123]:

```
def convertMinutes(x):
    return x.total_seconds()/60
```

In [125]:

```
(data['END_DATE'] -  data['START_DATE']).apply(lambda x: convertMinutes(x))
```

Out[125]:

```
0          6.0
1         12.0
2         13.0
3         14.0
4         67.0
        ...
1150       7.0
1151      18.0
1152      35.0
1153      18.0
1154     103.0
Length: 1155, dtype: float64
```

In [126]:

```
data['DURATION'] = (data['END_DATE'] -  data['START_DATE']).apply(lambda x: convertM
```

In [127]:

```
data.head()
```

Out[127]:

	START_DATE	END_DATE	CATEGORY	START	STOP	MILES	PURPOSE	YEAR	MONTH
0	2016-01-01 21:11:00	2016-01-01 21:17:00	Business	Fort Pierce	Fort Pierce	5.1	Meal/Entertain	2016	1
1	2016-01-02 01:25:00	2016-01-02 01:37:00	Business	Fort Pierce	Fort Pierce	5.0	NaN	2016	1
2	2016-01-02 20:25:00	2016-01-02 20:38:00	Business	Fort Pierce	Fort Pierce	4.8	Errand/Supplies	2016	1
3	2016-01-05 17:31:00	2016-01-05 17:45:00	Business	Fort Pierce	Fort Pierce	4.7	Meeting	2016	1
4	2016-01-06 14:42:00	2016-01-06 15:49:00	Business	Fort Pierce	West Palm Beach	63.7	Customer Visit	2016	1

In [ ]:

```
```

In [137]:

```
data['SPEED_KMPH'] = data['KMS']/(data['DURATION']/60)
```

In [139]:

```
data.head()
```

Out[139]:

	START_DATE	END_DATE	CATEGORY	START	STOP	MILES	PURPOSE	YEAR	MONTH
0	2016-01-01 21:11:00	2016-01-01 21:17:00	Business	Fort Pierce	Fort Pierce	5.1	Meal/Entertain	2016	1
1	2016-01-02 01:25:00	2016-01-02 01:37:00	Business	Fort Pierce	Fort Pierce	5.0	NaN	2016	1
2	2016-01-02 20:25:00	2016-01-02 20:38:00	Business	Fort Pierce	Fort Pierce	4.8	Errand/Supplies	2016	1
3	2016-01-05 17:31:00	2016-01-05 17:45:00	Business	Fort Pierce	Fort Pierce	4.7	Meeting	2016	1
4	2016-01-06 14:42:00	2016-01-06 15:49:00	Business	Fort Pierce	West Palm Beach	63.7	Customer Visit	2016	1

In [140]:

```
data['SPEED_KMPH'].mean()
```

Out[140]:

```
inf
```

In [143]:

```
data[data['SPEED_KMPH'] > 10000]
```

Out[143]:

	START_DATE	END_DATE	CATEGORY	START	STOP	MILES	PURPOSE	YEAR	MO
751	2016-09-06 17:49:00	2016-09-06 17:49:00	Business	Unknown Location	Unknown Location	69.1	NaN	2016	
761	2016-09-16 07:08:00	2016-09-16 07:08:00	Business	Unknown Location	Unknown Location	1.6	NaN	2016	
798	2016-10-08 15:03:00	2016-10-08 15:03:00	Business	Karachi	Karachi	3.6	NaN	2016	
807	2016-10-13 13:02:00	2016-10-13 13:02:00	Business	Islamabad	Islamabad	0.7	NaN	2016	

Observation: 43.129974 km/hr is the mean if we remove inf

In [ ]:

```

```

In [145]:

```
data.isna().sum()
```

Out[145]:

```
START_DATE      0
END_DATE        0
CATEGORY        0
START           0
STOP            0
MILES           0
PURPOSE       502
YEAR            0
MONTH           0
DAY             0
KMS             0
DURATION        0
SPEED_KMPH      0
dtype: int64
```

In [147]:

```
data['PURPOSE'].value_counts(dropna=False)
```

Out[147]:

```
NaN                 502
Meeting             187
Meal/Entertain      160
Errand/Supplies     128
Customer Visit      101
Temporary Site       50
Between Offices      18
Moving                4
Airport/Travel        3
Charity ($)           1
Commute               1
Name: PURPOSE, dtype: int64
```

In [ ]:

```
data.drop(data['SPEED']==float(inf))
```

In [151]:

```
(data['SPEED_KMPH']==np.inf).sum()
```

Out[151]:

```
4
```

In [156]:

```
data[data['SPEED_KMPH']<100000]
```

Out[156]:

	START_DATE	END_DATE	CATEGORY	START	STOP	MILES	PURPOSE	YEAI
0	2016-01-01 21:11:00	2016-01-01 21:17:00	Business	Fort Pierce	Fort Pierce	5.1	Meal/Entertain	201
1	2016-01-02 01:25:00	2016-01-02 01:37:00	Business	Fort Pierce	Fort Pierce	5.0	NaN	201
2	2016-01-02 20:25:00	2016-01-02 20:38:00	Business	Fort Pierce	Fort Pierce	4.8	Errand/Supplies	201
3	2016-01-05 17:31:00	2016-01-05 17:45:00	Business	Fort Pierce	Fort Pierce	4.7	Meeting	201
4	2016-01-06 14:42:00	2016-01-06 15:49:00	Business	Fort Pierce	West Palm Beach	63.7	Customer Visit	201
...	...	...	...	...	...	...	...	.
1150	2016-12-31 01:07:00	2016-12-31 01:14:00	Business	Kar?chi	Kar?chi	0.7	Meeting	201
1151	2016-12-31 13:24:00	2016-12-31 13:42:00	Business	Kar?chi	Unknown Location	3.9	Temporary Site	201
1152	2016-12-31 15:03:00	2016-12-31 15:38:00	Business	Unknown Location	Unknown Location	16.2	Meeting	201
1153	2016-12-31 21:32:00	2016-12-31 21:50:00	Business	Katunayake	Gampaha	6.4	Temporary Site	201
1154	2016-12-31 22:08:00	2016-12-31 23:51:00	Business	Gampaha	Ilukwatta	48.2	Temporary Site	201

1151 rows × 13 columns

In [ ]:

```
```