In [20]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import datetime
import calendar
```

In [21]:
```
data = pd.read_csv (r'\Users\m_gad\Downloads\Uber Drives - .csv')
```

In [4]:
```
data.head()
```

Out[3]:

	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 [5]:
```
data.tail()
```

Out[18]:

	START_DATE*	END_DATE*	CATEGORY*	START*	STOP*	MILES*	PURPOSE*
1151	12/31/2016 13:24	12/31/2016 13:42	Business	Kar?ch	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 [14]:
```
data.isnull().any()
```

Out[14]:
```
START_DATE*    False
END_DATE*       True
CATEGORY*       True
START*          True
STOP*           True
MILES*         False
PURPOSE*        True
dtype: bool
```

In [17]:
```
data.isnull().sum()
```

Out[17]:
```
START_DATE*    0
END_DATE*      0
CATEGORY*      0
START*         0
STOP*          0
MILES*         0
PURPOSE*       0
dtype: int64
```

In [18]:
```
data = data.dropna()
```

In [20]:
```
data.dtypes
```

Out[20]:
```
START_DATE*    object
END_DATE*      object
CATEGORY*      object
START*         object
STOP*          object
MILES*        float64
PURPOSE*       object
dtype: object
```

In 1 :
```
data["START_DATE*"] = pd.to_datetime(data["START_DATE*"])
```

In [32]:
```
data.head()
```

Out[32]:

	START_DATE*	END_DATE*	CATEGORY*	START*	STOP*	MILES*	PURPOSE*
0	2016-01-01 21:11:00	1/1/2016 21:17	Business	Fort Pierce	Fort Pierce	5.1	Meal/Entertain
2	2016-01-02 20:25:00	1/2/2016 20:38	Business	Fort Pierce	Fort Pierce	4.8	Errand/Supplies
3	2016-01-05 17:31:00	1/5/2016 17:45	Business	Fort Pierce	Fort Pierce	4.7	Meeting
4	2016-01-06 14:42:00	1/6/2016 15:49	Business	Fort Pierce	West Palm Beach	63.7	Customer Visit
5	2016-01-06 17:15:00	1/6/2016 17:19	Business	West Palm Beach	West Palm Beach	4.3	Meal/Entertain

In [33]:
```
data.dtypes
```

Out[33]:
```
START_DATE*    datetime64[ns]
END_DATE*              object
CATEGORY*              object
START*                 object
STOP*                  object
MILES*                float64
PURPOSE*               object
dtype: object
```

In [38]:
```
pd.to_datetime(data["START_DATE*"])
```

Out[38]:
```
0     2016-01-01 21:11:00
2     2016-01-02 20:25:00
3     2016-01-05 17:31:00
4     2016-01-06 14:42:00
5     2016-01-06 17:15:00
       ...
1150  2016-12-31 01:07:00
1151  2016-12-31 13:24:00
1152  2016-12-31 15:03:00
1153  2016-12-31 21:32:00
1154  2016-12-31 22:08:00
Name: START_DATE*, Length: 653, dtype: datetime64[ns]
```

In [44]:
```
data["START_DATE*"] =pd.to_datetime(data["START_DATE*"])
data["END_DATE*"] =pd.to_datetime(data["END_DATE*"])
```

In [44]:
```
data.dtypes
```

Out[45]:
```
START_DATE*    datetime64[ns]
END_DATE*      datetime64[ns]
CATEGORY*              object
START*                 object
STOP*                  object
MILES*                float64
PURPOSE*               object
I                datetime64[ns]
dtype: object
```

In [48]:
```
hour = []
day = []
month = []
dayofweek = []
weekday=[]

for x in data["START_DATE*"] :
    hour.append(x.hour)
    day.append(x.day)
    month.append(x.month)
    dayofweek.append(x.dayofweek)
    weekday.append(calendar.day_name[dayofweek[-1]])
```

In [49]:
```
data["Hour"]= hour
data["Day"] = day
data["Month"] = month
data["Dayofweek"] = dayofweek
data["Weekday"] = weekday
```

In [50]:
```
data.head()
```

Out[50]:

	START_DATE*	END_DATE*	CATEGORY*	START*	STOP*	MILES*	PURPOSE*	I	Hour	Day	Month	Dayofweek	Weekday
0	2016-01-01 21:11:00	2016-01-01 21:17:00	Business	Fort Pierce	Fort Pierce	5.1	Meal/Entertain	NaT	21	1	1	4	Friday
2	2016-01-02 20:25:00	2016-01-02 20:38:00	Business	Fort Pierce	Fort Pierce	4.8	Errand/Supplies	NaT	20	2	1	5	Saturday
3	2016-01-05 17:31:00	2016-01-05 17:45:00	Business	Fort Pierce	Fort Pierce	4.7	Meeting	NaT	17	5	1	1	Tuesday
4	2016-01-06 14:42:00	2016-01-06 15:49:00	Business	Fort Pierce	West Palm Beach	63.7	Customer Visit	NaT	14	6	1	2	Wednesday
5	2016-01-06 17:15:00	2016-01-06 17:19:00	Business	West Palm Beach	West Palm Beach	4.3	Meal/Entertain	NaT	17	6	1	2	Wednesday

In [51]:
```
data["CATEGORY*"].value_counts()
```

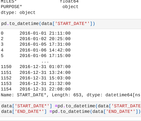
Out[51]:
```
Business    647
Personal      6
Name: CATEGORY*, dtype: int64
```

In [52]:
```
data["PURPOSE*"].value_counts()
```

Out[52]:
```
Meeting            187
Meal/Entertain     160
Errand/Supplies    128
Customer Visit     101
Temporary Site      50
Meeting-Offices     14
Moving               4
Airport/Travel       3
Charity ($)          1
Commute              1
Name: PURPOSE*, dtype: int64
```

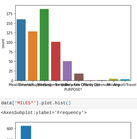
In [22]:
```
# Categories We Have
sns.countplot(x='CATEGORY**', data=data)
```

Out[22]:
```
<AxesSubplot:xlabel='CATEGORY**', ylabel='count'>
```



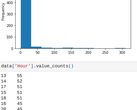
In [29]:
```
sns.countplot(x='PURPOSE**', data= data  )
```

Out[18]:
```
<AxesSubplot:xlabel='PURPOSE*', ylabel='count'>
```



In [56]:
```
data["MILES"].plot.hist()
```

Out[56]:
```
<AxesSubplot:ylabel='Frequency'>
```

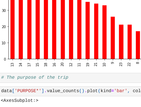


In [59]:
```
data["Hour"].value_counts()
```

Out[59]:
```
13    55
14    54
17    51
15    51
18    51
16    45
12    44
12    43
11    39
19    36
21    34
20    31
9     26
23    21
22    21
8     17
10    13
7     8
1     4
3     3
6     3
0     2
2     1
Name: Hour, dtype: int64
```

In [60]:
```
#what Hour Do Most People Take Uber To Their Destination?
data["Hour"].value_counts().plot(kind ='bar',figsize= (10,5),color= 'red')
```

Out[60]:
```
<AxesSubplot:>
```



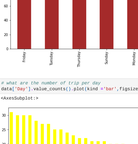
In [67]:
```
# The purpose of the trip
```

In [70]:
```
data["PURPOSE*"].value_counts().plot(kind='bar', color='orange', figsize= (10,5))
```

Out[70]:
```
<AxesSubplot:>
```

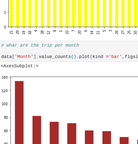
In [72]:
```
# wich day has the haighst number of trip wich day has the haighst number of trip

data["Weekday"].value_counts().plot(kind ='bar',figsize= (10,5),color= 'brown')
```

Out[72]:
```
<AxesSubplot:>
```



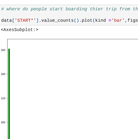
In [75]:
```
# what are the number of trip per day
data["Day"].value_counts().plot(kind ='bar',figsize= (10,5),color= 'yellow')
```

Out[75]:
```
<AxesSubplot:>
```



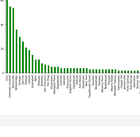
In [76]:
```
# what are the trip per month

data["Month"].value_counts().plot(kind ='bar',figsize= (10,5),color= 'brown')
```

Out[76]:
```
<AxesSubplot:>
```



In [78]:
```
# where do people start boarding thier trip from the most

data["START*"].value_counts().plot(kind ='bar',figsize= (30,15),color= 'green')
```

Out[78]:
```
<AxesSubplot:>
```



In [ ]: