

Karanjot Singh

Q1. From the attached data analyse overall which is a bigger issue among – Non availability of Cars, Trips cancelled?

In [1]:

```
1 import pandas as pd
2 import seaborn as sns
3 import matplotlib.pyplot as plt
4 data = pd.read_csv(r"cabs-data.csv")
5 count = data['Status'].value_counts()
6 print(count)
7 #print(data)
8 sns.countplot(x = 'Status', data = data)
```

```
Trip Completed      2831
No Cars Available   2650
Cancelled            1264
Name: Status, dtype: int64
```

Out[1]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x22cd8e9a108>
```

Q2. Divide the request time slot into six slots: Early morning (1-4), morning (5-8), Late Morning (9-12), Noon(13-16), Evening(17-21) & Night (21-24)

In [14]:

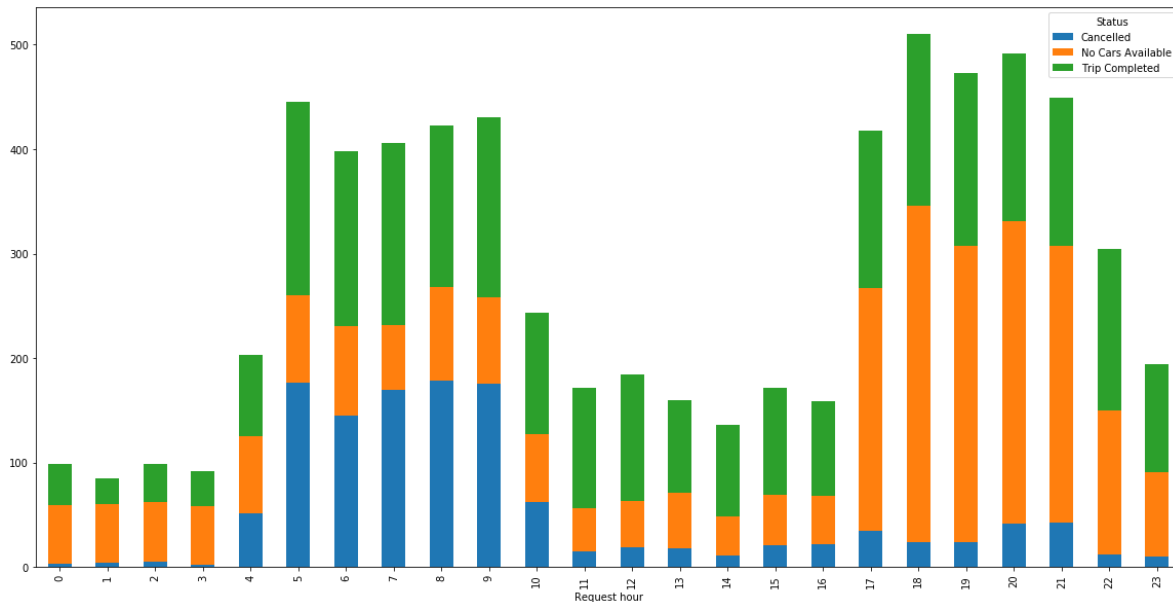
```

1 data['Request timestamp'] = pd.to_datetime(data['Request timestamp'])
2 data['Request hour'] = data['Request timestamp'].dt.hour
3 data.groupby(['Request hour', 'Status']).size().unstack()\
4 .plot(kind = 'bar', stacked = True, figsize = (20,10))

```

Out[14]:

<matplotlib.axes._subplots.AxesSubplot at 0x22cd78cb348>



In [13]:

```

1 data = data.assign(TD=pd.cut(data['Request hour'],[1,4,8,12,16,20,24],labels = \
2     ['Early Morning', 'Morning', 'Late Morning', 'Noon', 'Evening', 'Night']))
3 data.head()

```

Out[13]:

	Request id	Pickup point	Driver id	Status	Request timestamp	Drop timestamp	Request hour	TD	completed
0	619	Airport	1.0	Trip Completed	2016-11-07 11:51:00	11/7/2016 13:00	11	Late Morning	Trip Completed
1	867	Airport	1.0	Trip Completed	2016-11-07 17:57:00	11/7/2016 18:47	17	Evening	Trip Completed
2	1807	City	1.0	Trip Completed	2016-12-07 09:17:00	12/7/2016 9:58	9	Late Morning	Trip Completed
3	2532	Airport	1.0	Trip Completed	2016-12-07 21:08:00	12/7/2016 22:03	21	Night	Trip Completed
4	3112	City	1.0	Trip Completed	2016-07-13 08:33:16	13-07-2016 09:25:47	8	Morning	Trip Completed

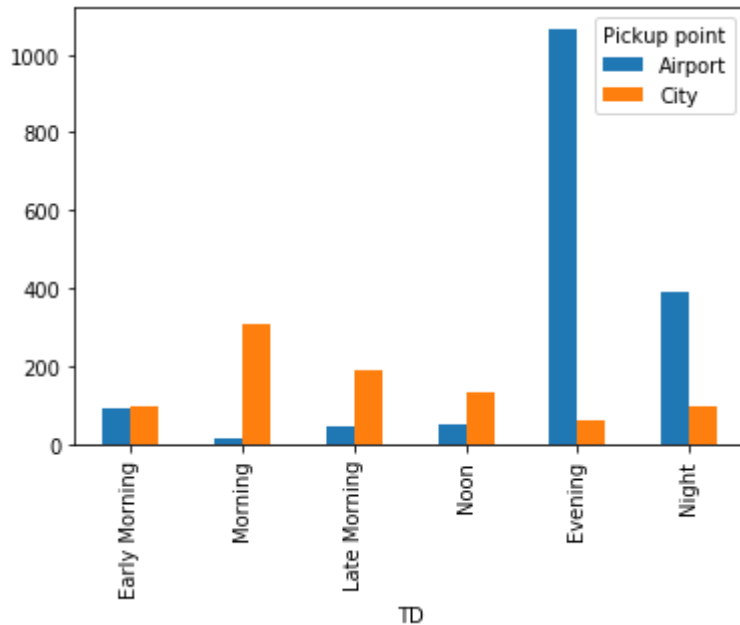
Q3. In which time slot and pickup point cars are not available (highest) ?

In [9]:

```
1 data[data['Status']=='No Cars Available'].groupby(['TD','Pickup point'])\  
2 .size().unstack().plot(kind = 'bar')
```

Out[9]:

<matplotlib.axes._subplots.AxesSubplot at 0x22cd955f548>



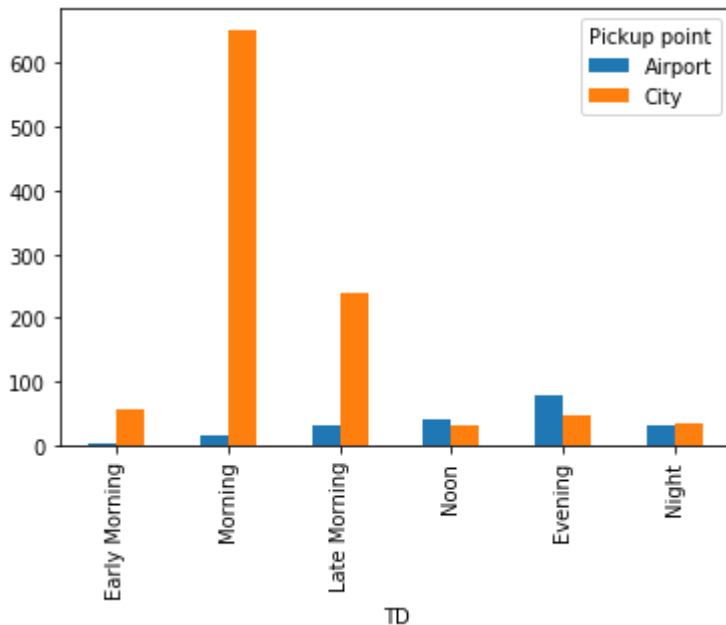
Q4. In which time slot and pickup point trips are mostly cancelled?

In [10]:

```
1 data[data['Status']=='Cancelled'].groupby(['TD','Pickup point'])\  
2 .size().unstack().plot(kind = 'bar')
```

Out[10]:

<matplotlib.axes._subplots.AxesSubplot at 0x22cd95ea048>



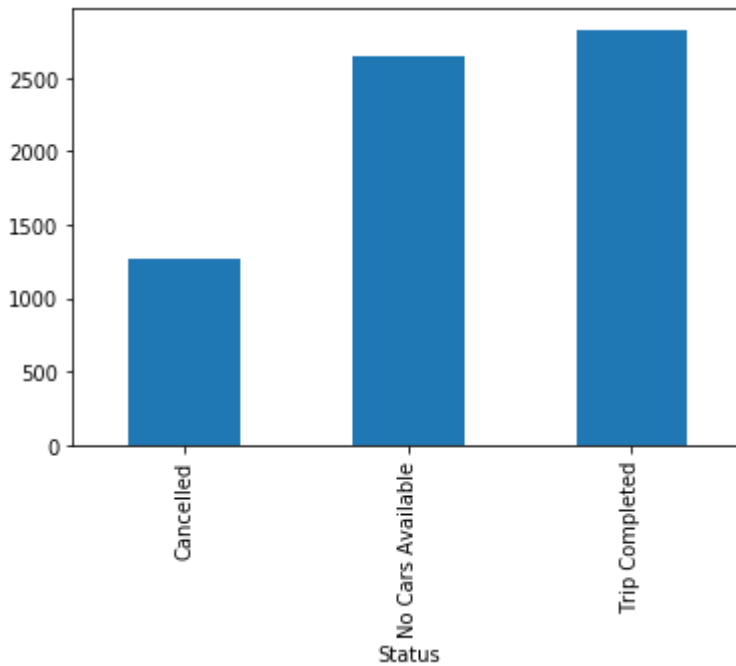
Q5. Count the no. of cases of trips completed and trips not completed and to compare, plot both cases.

In [6]:

```
1 data[data['Status']!='Trip completed'].groupby(['Status']).size().plot(kind = 'bar')
```

Out[6]:

<matplotlib.axes._subplots.AxesSubplot at 0x22cd9460c48>



Q6. In which time slot and pickup point trips are not completed (highest).

In [12]:

```
1 data['completed'] = [x if x=='Trip Completed' else 'Trip not completed'\
2                       for x in data['Status']]
3 #print(data)
4 data[data['completed']=='Trip not completed'].groupby(['TD','Pickup point'])\
5 .size().unstack().plot(kind='bar')
```

Out[12]:

<matplotlib.axes._subplots.AxesSubplot at 0x22cd78cb248>

