

UBER ASSIGNMENT

#CLEANING THE DATA AND MAKING IT MORE UNDERSTABLE

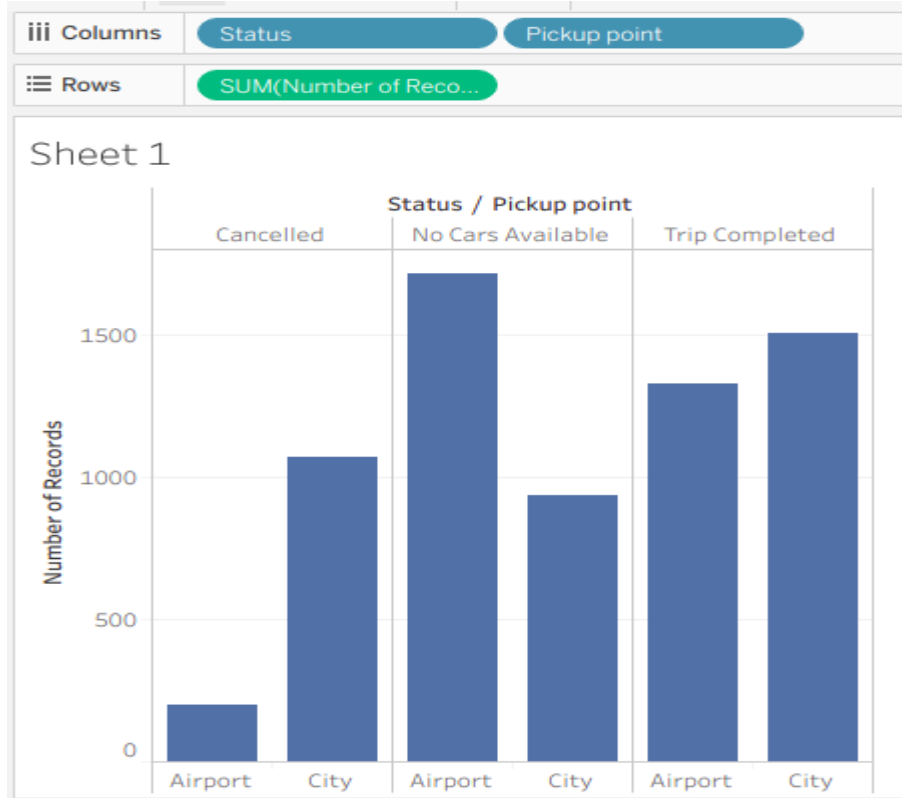
```
In [3]: new=uber['Request timestamp'].str.split(" ",n=1,expand=True)
uber['Request date']=new[0]
uber['Request time']=new[1]
new1=uber['Drop timestamp'].str.split(" ",n=1,expand=True)
uber['Drop date']=new1[0]
uber['Drop time']=new1[1]
```

THE COLUMN REQUEST TIMESTAMP AND DROP TIMESTAMP IS SPLITTED INTO TWO COLUMNS WHICH CONSISTS OF DATE AND OTHER CONSISTS OF TIME SO IT MAKES EASY TO VISUALIZE THE DATASET

NEW DERIVED COLUMNS ARE REQUEST TIME ,REQUEST DATE AND DROP TIME AND DROP COLUMN

```
#CHANGING THE DATATYPE TO (datetime)
uber1=pd.to_datetime(uber['Request time'])
```

CHANGING THE DATATYPE MAKES IT EASIER TO VISUALLIZE THAT COLUMN



THE COMPANY IS NOT SUCCESSFUL TO HANDLE THE DEMAND FROM THE AIRPORT. WE CAN SEE THIS FROM STATUS (NO CARS AVAILABLE) FROM THE TABLEAU DIAGRAM ALSO. MOST OF THE CARS HAVE BEEN CANCELLED FROM THE CITY SIDE. SO THE SOLUTION FOR THIS CAN BE NO OF CARS SHOULD BE INCREASED ON THE AIRPORT SIDE AND THE DRIVERS SHOULD CONFIRM FROM THE CUSTOMER ON THE CITY SIDE BEFORE CONFIRMING THE RIDE BECAUSE ON CANCELLING THE RIDE THE COMPANY FACES LOSS OF BOTH TIME AND MONEY.

```
In [4]: #THESE COLUMNS ARE OF NO USE NOW
uber=uber.drop(['Drop timestamp'],axis=1)
uber=uber.drop(['Request timestamp'],axis=1)
```

THE COLUMN [REQUEST TIMESTAMP] AND [DROP TIMESTAMP] IS DROPPED
BECAUSE IT WAS OF NO USE AFTER DERIVING NEW COLUMNS

```
In [5]: #AS WE HAVE NO ROLE OF DROP DATE AND DROP TIME IN VISUALLIZING THE MAIN FACTORS WHICH ARE
#AFFECTING THE FINANCIAL CONDITION OF THE COMPANY SO THESE COLUMNS SHOULD BE
#DROPPED
uber=uber.drop(['Drop time'],axis=1)
uber=uber.drop(['Drop date'],axis=1)
```

OUR MAIN AIM IS TO FIND THE REASONS OF NO CARS AVAILABILITY AND WHY THE TRIPS ARE CANCELLED BY THE CUSTOMERS THESE ARE THE MAIN FACTORS WHICH AFFECTS THE FINANCIAL CONDITION OF THE COMPANY SO THE COLUMN GIVING DETAILS REGARDING DROP TIME AND DROP DATE IS OF NO USE AND SHOULD BE DROPPED

```
In [7]: table9 = pd.crosstab(index=uber["Status"],  
                             columns=uber["Driver id"].loc[uber["Driver id"]==[84]])
```

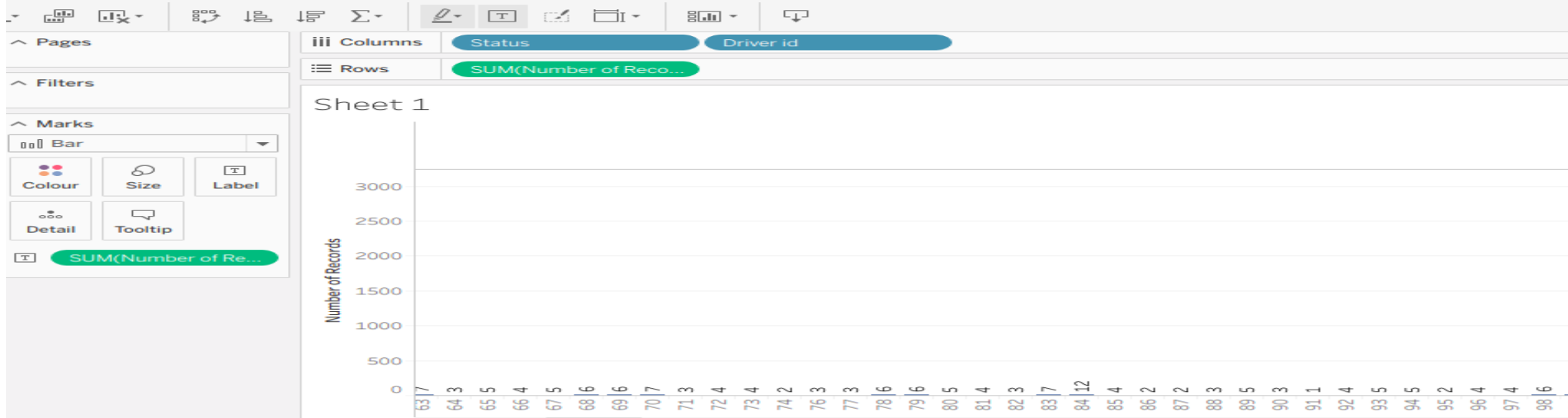
```
In [9]: table9  
#WE CAN SEE THAT MOST OF THE CARS CANCELLED IS FROM (DRIVER ID ==84) ,THERE CAN BE SOME  
#REASONS LIKE HE TAKES MUCH TIME TO PICK UP THE CUSTOMER ,HIS BEHAVIOUR ,SO HE SHOULD BE  
#MONITORED
```

Out[9]:

Driver id 84.0	
Status	
Cancelled	12
Trip Completed	9

FROM THE RECORD WE CAN OBSERVE THAT MOST OF THE TRIPS CANCELLED IS IN THE ACCOUNT OF DRIVER ID = 84 ie 12 TRIPS IS CANCELLED ON HIS RECORD WHICH IS MAXIMUM . THERE CAN BE MANY BEHAVIOURAL ISSUES WITH THE DRIVER SO THE COMPANY SHOULD CHECK THE RECORDS OF THIS DRIVER AND TAKE APPROPRIATE ACTION

WE CAN ALSO CHECK THE DETAILS OF OTHER DRIVERS BY TABLEAU:-



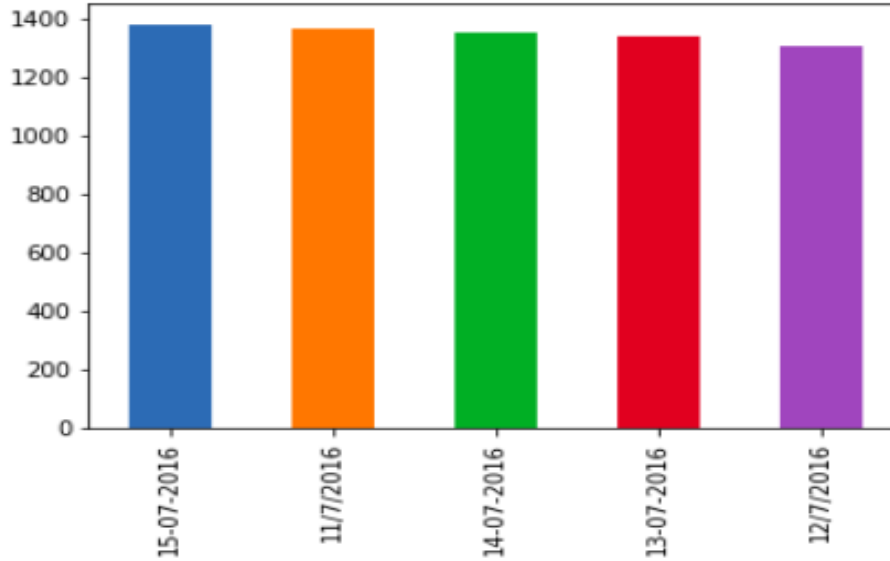
CHECK THE NO OF CARS CANCELLED FOR DRIVER ID ==84 (MAXIMUM)

```
In [19]: #WE ARE FOCUSSING ON THE POINT WHY AND WHERE MOST OF THE CABS WERE NOT AVAILABLE
        #SO THE COLUMN [DRIVER ID] IS OF NO USE AND WE SHOULD DROP IT
        uber=uber.drop(['Driver id'],axis=1)
```

WE HAVE USED THE COL DRIVER ID AND DERIVED THE INFORMATION WHICH WAS USEFUL AND NOW THIS COLUMN SHOULD BE DROPPED AS IT IS OF NO USE

```
In [20]: uber['Request date'].value_counts().plot.bar()  
#WE CAN OBSERVE FROM THE COUNT PLOT THAT THE CUSTOMERS HAVE REQUESTED FOR UBER MOSTLY  
#ON DATE 11/7/2016 AND 15/07/2016 SO THE COMPANY SHOULD TAKE THIS AS A NOTE AND MORE  
#CARS SHOULD BE AVAILABE ON THESE DATES
```

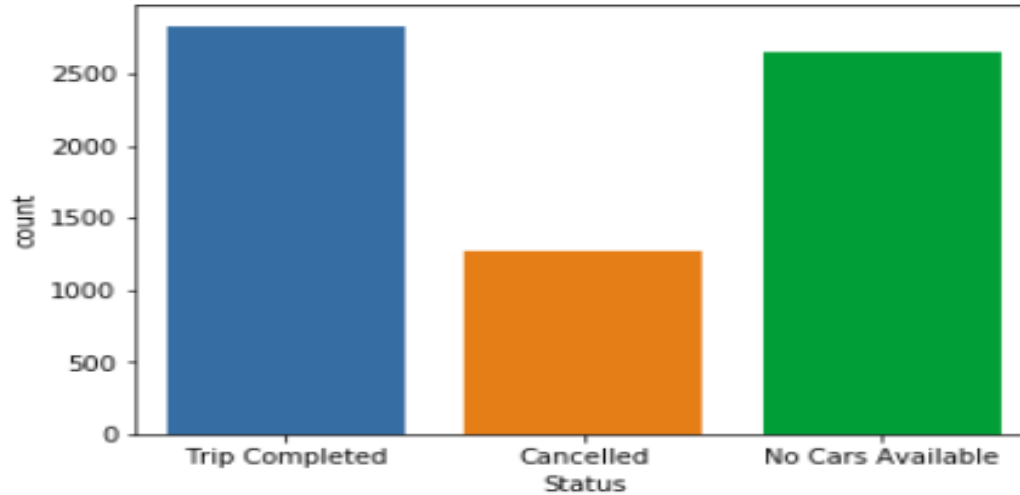
```
Out[20]: <matplotlib.axes._subplots.AxesSubplot at 0x7f370837b748>
```



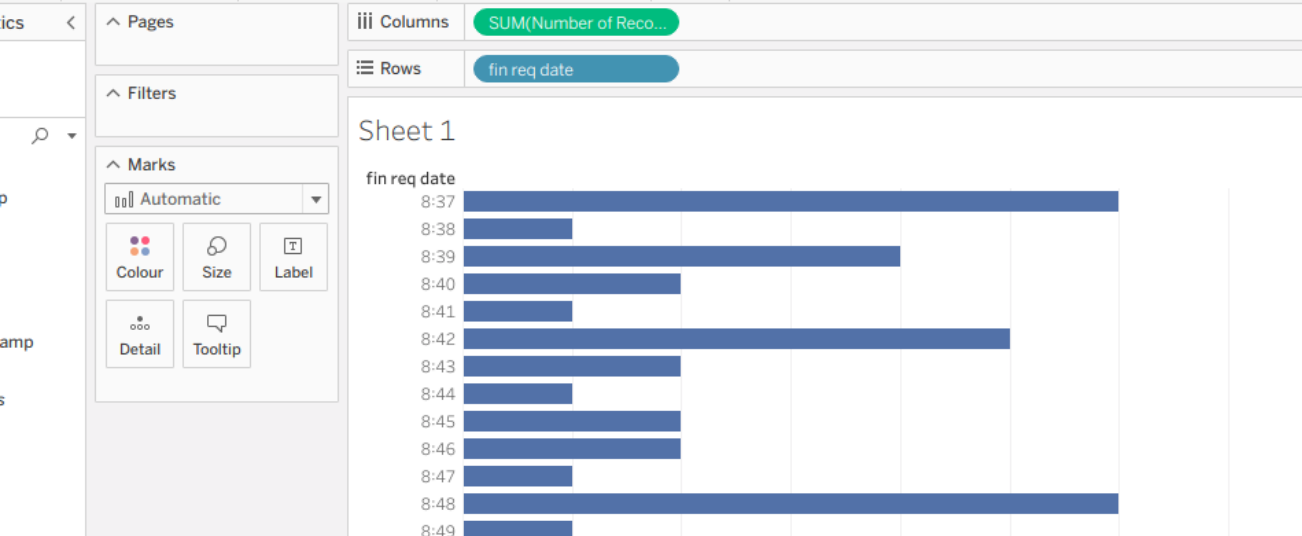
MOST DEMANDS FOR THE CABS WERE OBSERVED FOR THE DATES 15,11,14
SO THE MORE NO OF CABS SHOULD BE AVAILABLE ON THESE DATES

```
In [6]: sns.countplot(x='Status',data=uber)  
#BY THIS PLOT WE CAN CONCLUDE THAT COMPANY IS SUFFERING FROM A HUGE LOSS BECAUSE OF  
#STATUS(NO CARS AVAILABLE) SO THE UBER COMPANY HAD TO INCREASE THE NO OF CARS
```

```
Out[6]: <matplotlib.axes._subplots.AxesSubplot at 0x7f370f4bfb00>
```



BY THIS PLOT THIS SHOULD BE NOTED THAT THE COMPANY IS SUFFERING FROM A GREAT LOSS BECAUSE OF NON AVAILABILITY OF CARS SO COMPANY SHOULD INCREASE THE NO OF CABS TO MATCH UP THE DEMANDS OF THE CUSTOMER



MOST OF THE REQUEST HAVE BEEN DONE IN MORNING TIME IE 5AM TO 10 AM SO NO OF CABS SHOULD BE AVAIL AT THIS TIME INTERVAL

```
In [27]: table3= pd.crosstab(index=uber["Pickup point"],
                             columns=uber["Request time"])
```

```
In [26]: table3
#FROM TABLE3 WE CAN CONCLUDE THAT MOST OF THE PASSENGERS NEED CAB AT THE MORNING TIME IE
#FROM 5AM TO 10 AM FROM THE AIRPORT MAY BE SOME PLANE ARRIVAL TIME CAN BE IN THESE TIME
#INTERVALS SO CARS SHOULD BE AVAILABLE IN THE AIRPORT SIDE IN THE MORNING TIME
```

```
Out[26]:
```

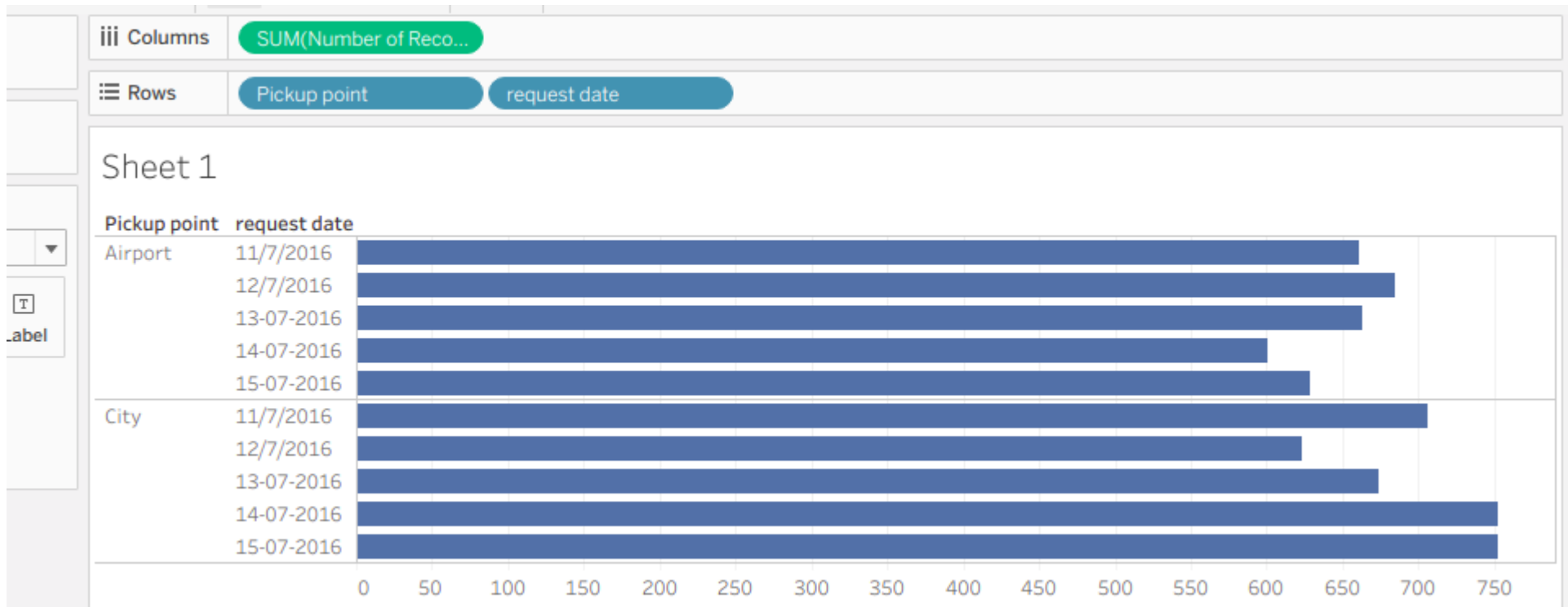
	3:40	00:03:46	00:03:52	00:04:44	00:05:00	00:06:12	00:06:34	...	9:49	9:50	9:51	9:52	9:53	9:54	9:55	9:56	9:57	9:59
1	0	1	0	1	0	1	...	0	1	0	0	0	0	1	2	0	1	
0	1	0	1	0	1	0	...	1	3	1	3	1	4	2	2	2	4	


```
table = pd.crosstab(index=uber["Status"],
                    columns=uber["Request date"])
```

table
#BY CROSSTAB WE CAN SEE THAT THE COMPANY IS FACING MAXIMUM LOSS ON THE DATE ON 14 AND 15
#AS MAXIMUM NO OF CARS WERE NOT AVAILABLE ON THESE DAYS SO WE SHOULD AVAIL MORE CARS ON
#THESE DAYS MOST OF THE TRIP HAD BEEN CANCELLED ON THE DATE 11 AND 13 WHICH RESULT IN LOSS
#OF TIME AND MONEY

Request date	11/7/2016	12/7/2016	13-07-2016	14-07-2016	15-07-2016
Status					
Cancelled	262	240	270	252	240
No Cars Available	504	505	490	571	580
Trip Completed	601	562	577	530	561

A HUGE DIFFERENCE IS SEEN IN THE THE NO OF NO CARS AVAILABLE IN THE DATES 14 AND 15 AS COMPARED TO OTHER DATES .WE HAVE ALSO SEEN MOST OF THE DEMANDS WERE FOR DATES 11 ,14,15 AND THE COMPANY IS NO ABLE TO MATCH UP THE DEMAND ON THE DATES 14 AND 15 SO THE COMPANY SHOULD INCREASE THE NO OF CABS ON THESE TWO DAYS



WE CAN SEE THE HIGH DEMANDS FROM THE AIRPORT SIDE IS ON 11,12,13 DATES WHEREAS ON THE DATES 14 AND 15 HIGH DEMAND HAVE BEEN SEEN FROM THE CITY SIDE SO THE CABS SHOULD BE ACCORDINGLY DISTRIBUTED ACCORDING TO THIS OBSERVATION