

## Quality Assurance Report

Carl Chen

Given the tables given by Mr. Tao Chen, I imported 3 tables (Fig. 1) into MySQL and used **union** to make one working sheet (Fig. 2) for future data analysis.

uber-raw-data-may14	
Date/Time	Date
Lat	FLOAT(53)
Lon	FLOAT(53)
Base	varchar (20)

uber-raw-data-jun14	
Date/Time	Date
Lat	FLOAT(53)
Lon	FLOAT(53)
Base	varchar (20)

uber-raw-data-jul14	
Date/Time	Date
Lat	FLOAT(53)
Lon	FLOAT(53)
Base	varchar (20)

Fig. 1 Raw Data in 3 tables

Then, based on Python platform, I did the ETL process for data analysis of Uber, over the three months.

uber-raw-data-3-month	
Date/Time	Date
Lat	FLOAT(53)
Lon	FLOAT(53)
Base	varchar (20)

Fig. 2 Union Table

According to the cumulative hourly data in Fig. 3 (a), the Uber trips decreases after about 5:00 pm and then start increasing after 2 am and the trips keep rising till 5:00 pm such that 5:00 pm is the busiest hour for Uber then the trips start decreasing. In July, 5 pm is the busiest day for Uber at 5 pm, compared with the lowest consumption in May at 2 am.

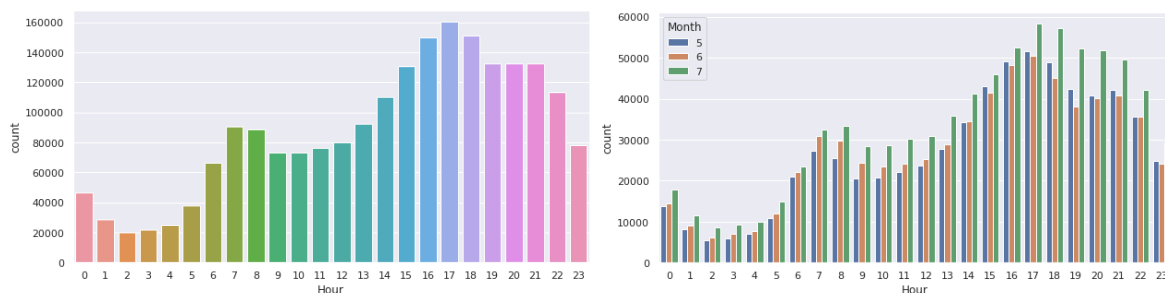


Fig. 3 (a) Count of Uber Trips per hour over 3 months

(b) Count of Uber Trips per hour for each month

By looking at the daily trips shown in Fig.4 (a) (b), we can say that the Uber trips are rising on the working days (Monday to Friday) and decreases on the weekends (Saturday and Sunday).

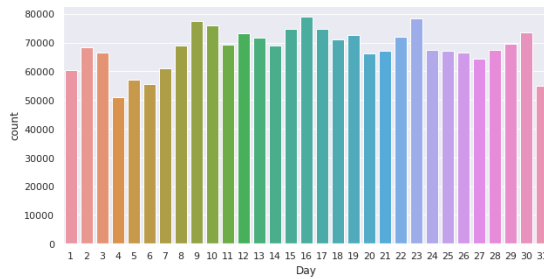
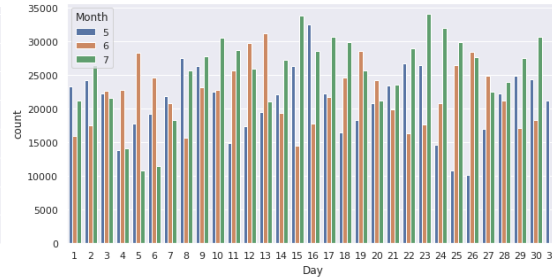


Fig. 4 (a) Count of Uber Trips per day over 3 months



(b) Count of Uber Trips per day for each month

Now let us analyze the Uber trips according to the weekdays. In Fig. 5 (a), bar 0 indicates Sunday, on Sundays the Uber trips and more than Saturdays (bar 6). On Saturdays, the Uber trips are the lowest. But on Thursday, people preferred to take Uber for their trips, with the largest consumption of Uber trips. Thus, we can say people also use Uber for working rather than for outing on weekends. In Fig. 5 (b), the consumptions in May on Saturday were lowest with about 60000 trips, but those in March on Thursday were highest with more than 140000 trips. Therefore, Thursday is the most profitable day for Uber.

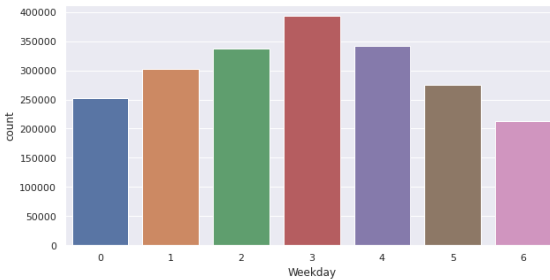
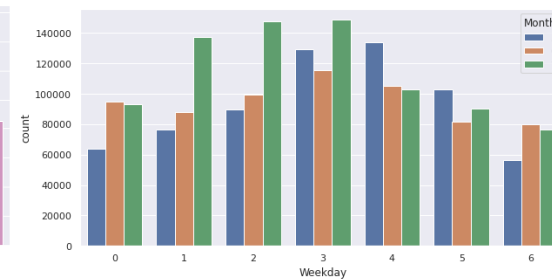


Fig. 5 (a) Count of Uber Trips per weekday over 3 months



(b) Count of Uber Trips per weekday for each month

Furthermore, we can easily find that 5 pm on Thursday is the busiest time shown in Fig. 6 (a), and the consumption of Uber in the according days was biggest in 5 pm, which is shown in Fig. 6 (b).

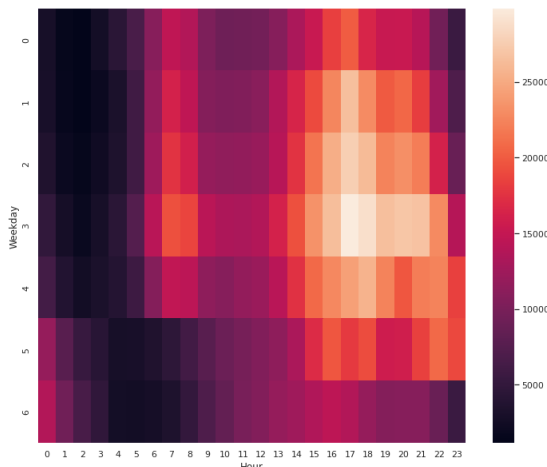
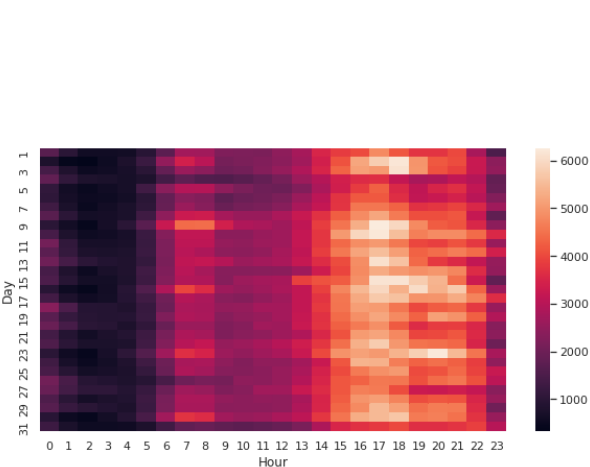


Fig. 6 (a) Heat Map By Weekday and Hour



(b) Heat Map By Day and Hour

As for the column 'base', Fig 7 (a) shows the biggest consumptions were in the bases of B02598, B02617, and B02682, and in July, the number of the trips for B02617 was largest with more than 300000 trips shown in Fig 7 (b).

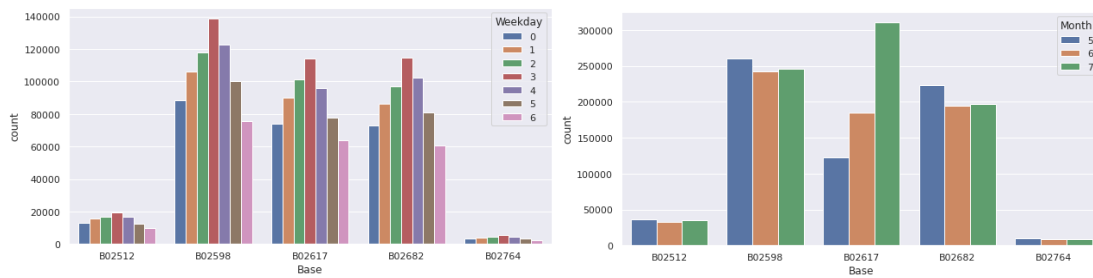


Fig. 7 (a) Count of Uber Trips per base in each weekday

(b) Count of Uber Trips per base for each month

In Fig. 8, it is obvious that most of the Uber trips originate near the Manhattan region in New York. So, there was a big demand in these areas.

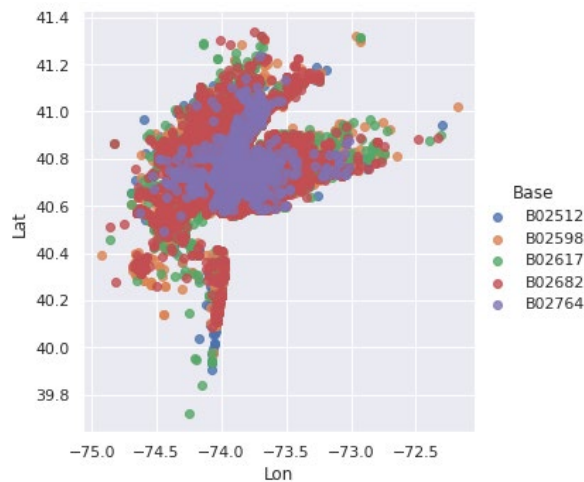


Fig. 8 GPS Plotting

According to the given weather table for the relative months, we can easily join tables on the date variable. Using python, data analysis of the new table with weather information can be obtained and visualized in Fig. 9. In Fig. 9, we can see that customers chose Uber for their trips when it was rainy. So, more Uber vehicles could be dispatched for rainy day.

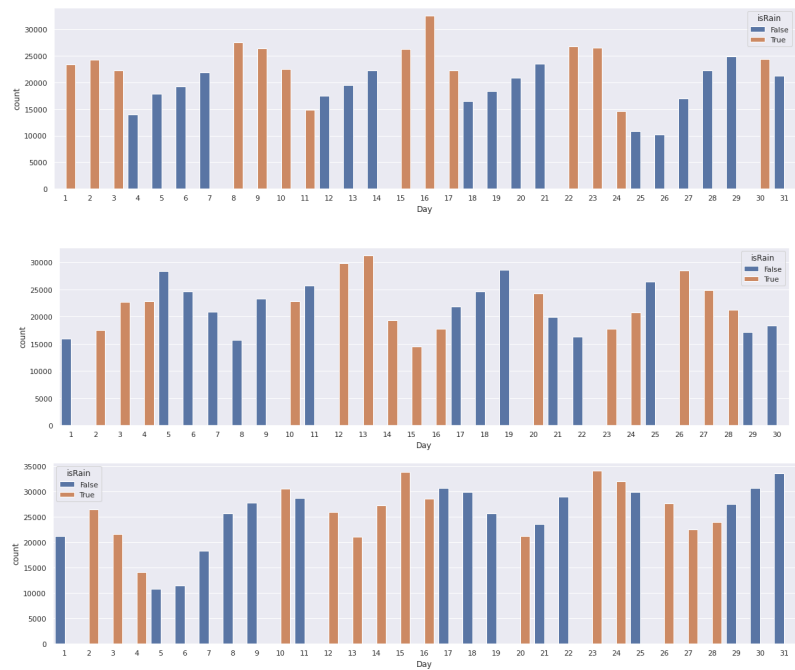


Fig. 9 Uber Trips for Each Day Labelled by Weather Information  
(Top one is for May, Middle one is for June, Last one is for July)