VIA Part I

Yijun Wan

1. Propose a metric and/or algorithm to assess the potential efficiency of aggregating rides from many vehicles into one, given the available data. Make realistic assumptions and any necessary simplifications and state them.

Metric: 1- total distance after aggregate vehicles/ original total distance $\in [0,1]$

Algorithm: Order passengers by their pick-up time. Then for each passenger, find its qualified share-ride passenger sequentially. The qualified passenger should meet the time requirement and route requirement.

Time requirement: If the first passenger already gets on car, only passengers with pick-up time before the time the first passenger arrives at the destination will be considered, and the algorithm is on a 'first come, first served' basis.

Route requirement: A is the first passenger, and B is the second passenger. B' pick-up spot should in the rectangular with its diagonal lining by A' pick-up and A's drop-off; B' drop off spot should in the rectangular with its diagonal lining by A' pick-up and A's drop-off or A' drop off spot should in the rectangular with its diagonal lining by B' pick-up and B's drop-off.

Assumptions:

a.) User gets on the vehicle as soon as it arrives, we do not need to worry about the time delay.
b.) Each ride will be shared with no more than 2 groups of people (maximum passenger capacity is 4).
c.) The distance used is calculated in Manhattan Distance.
d.) All streets are two-way street.

2. Implement your proposed method and evaluate Manhattan's overall efficiency using yellow taxi data from the first full week (Monday-Sunday) in June 2016. Discuss how your method would scale with more data; in other words, discuss the complexity of your implementation.

../../../../../../Desktop/VIA_Challenge_Problem/VIA_algorithm.py

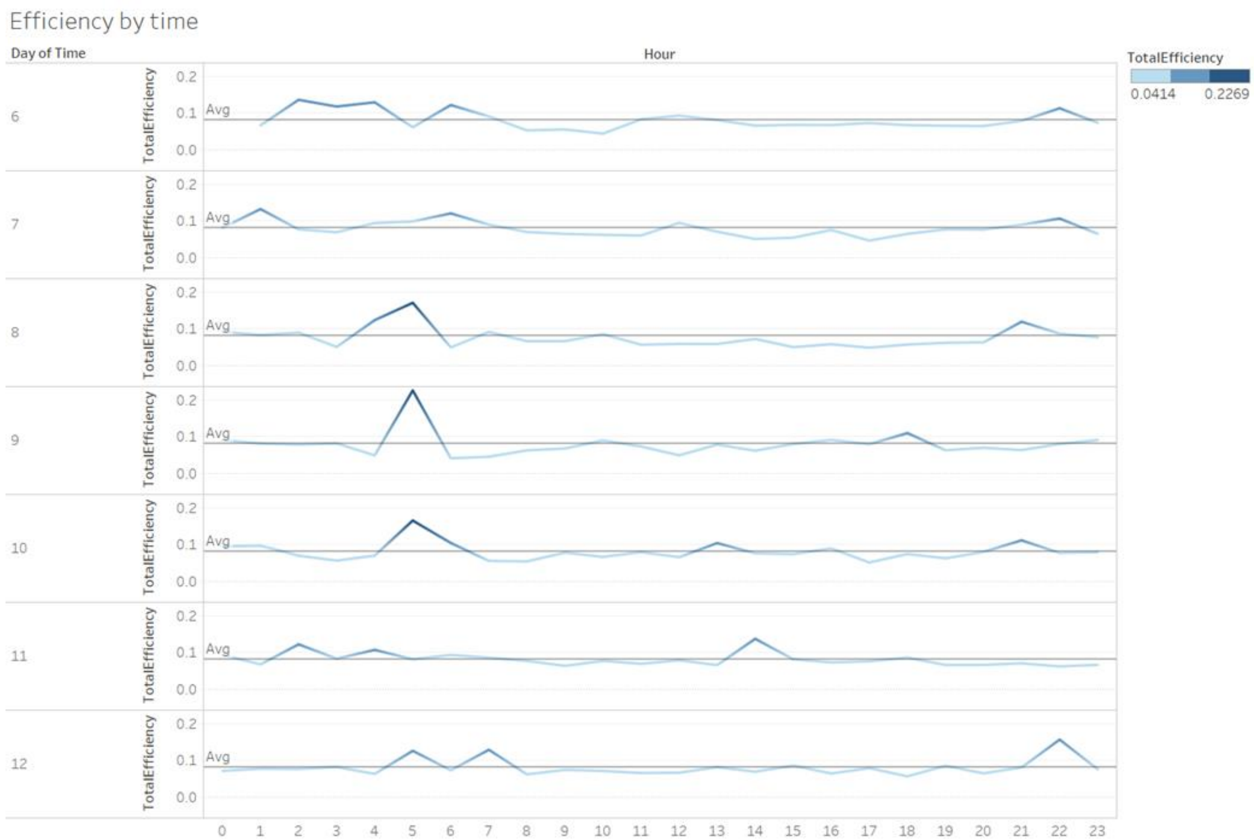
(check code for the algorithm here)

a. In my proposed algorithms, I will make the first passenger with pick-up time qualified, route qualified be the share-ride passenger. So, in order to reduce the complexity of the computing, I first order the data by 'pick-up time'. Time complexity O(n*logn)

b. In general, the time complexity is O(k*n) and k is the number of passengers during certain time periods (number of passengers who meet the time requirement). In worst cases, the time complexity is O(n^2) because when search for share-ride passenger for each passenger, it is possible that all other passengers are checked (k=n).

So the time complexity is O(n*logn + k*n).

3. Based on your implementation in the previous question, use visualizations to show how efficiency varies with time and location. Discuss any potential business implications based on your findings.



The trend of sum of TotalEfficiency for Hour Hour broken down by Time Day. Color shows sum of TotalEfficiency.

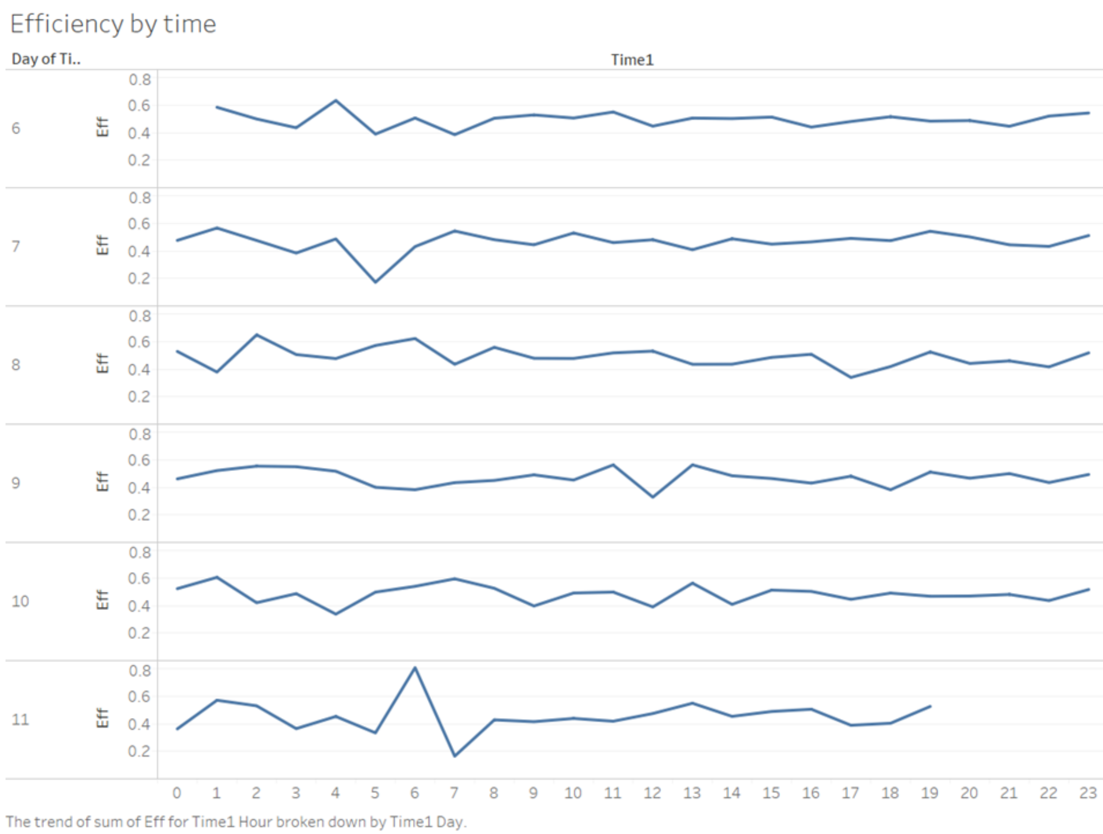
Efficiency by location:

Location	Efficiency
Bronx	12.15%
Staten Island	2.90%
Queens	13.51%
Manhattan	11.32%
Brooklyn	9.88%

Given the result:

a.) The efficiency patterns on weekdays such as Wednesday, Thursday, Friday are similar, while those on weekends are different. When we predict demand, we need to take that into account.
b.) Most efficiency peaks occur during late evenings and early morning, so I think we could encourage people to share rides during these time periods, from 10pm-7am.
c.) The efficiency on different areas are different, we may need to focus on area with high efficiency like Manhattan, Bronx if possible.
d.) Here, I think the efficiency may varies according to the demand size. For example, the more people require for the ride, the higher efficiency aggregating vehicles may achieve. But I have not tested my hypothesis. I do think I need more information to make more realistic and accurate decisions.

I also run another algorithm with different rules for qualified share-ride passengers (the algorithm not finish running yet). We can see the results are a little different, but the overall efficiency of aggregating vehicles increases a lot comparing the one above.



Different methods of aggregating vehicles may achieve different efficiency. Besides, other metrics such as percentage of reduced vehicles after aggregating could be used and result in different efficiency pattern. I think more time and effort is needed to find a more appropriate model and metric.