MS&E231: Assignment 2

Cyprien de Lichy, Louis Eugene, Guillaume Rostaing

October 28, 2015

1 Introduction

In this assignment, we propose to use Amazon Elastic MapReduce to investigate the reasons behind the apparent difficulty to catch a cab in the rain in New York City. Various theories try to explain this fact.

For Camerer, drivers set themselves an income target every day and stop working when their goal is reached. According to this theory, when it rains the hourly income should increase and thus drivers should quit earlier as they reach their target faster, and this explains a fall in supply.

More recently, Farber suggested that the answer is less clear. If demand could increase when it rains, resulting in higher taxi occupancy rates, driving conditions deteriorate and this could result in lower earnings in the end. Drivers would then be less willing to work for lower or same money in more difficult conditions, as neoclassical economics predicts.

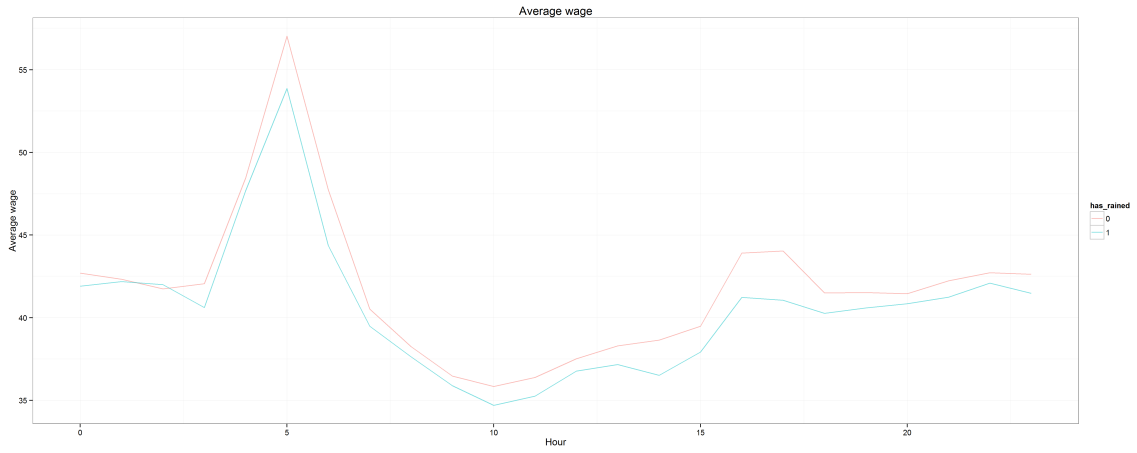
2 Methodology

To perform the assignment, we used the provided set of around 700 million NYC taxi rides between 2010 and 2013. We removed some improbable cases, namely trip s with incorrect GPS coordinates, improper earnings or unrealistic speeds. Trips longer than 40 miles or 3 hours were also discarded. Because the data is quite large (approximately 100 GB), we used 3 MapReduce processes on Amazon Elastic MapReduce to finally get for each date and hour several quantities such as the total amount of money earned during the hour by all the drivers (tip excluded) or the total distance traveled. We then joined this data with hourly precipitation data from NOAA collected at Central Park. From this we were able to compute various indicators and quantities to try to corroborate one theory or the other.

3 Analysis

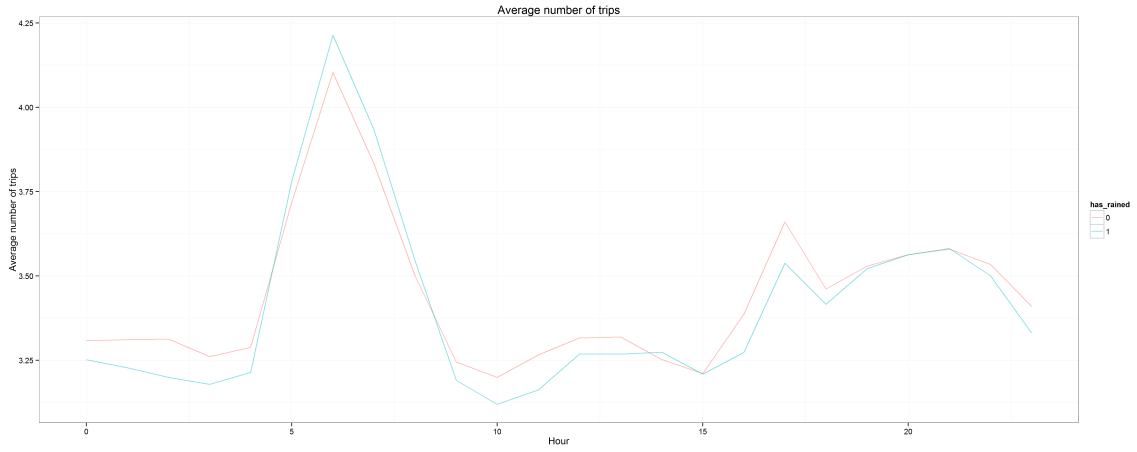
To start the analysis, we have computed the average hourly wage for each hour of the day depending on the precipitation. We have assumed that under 8 hundredth of inches of

precipitation[1], the effect of rain was negligible and thus we assimilated small precipitations with dry conditions. The average wage seems to be lower under rainy conditions. See the following figure:



Then, we computed the average earning over an entire day. We found that the average hourly wage is 39.56$ when it rains, and 40.73$ when it doesn't. As there is only a slight drop in the average wage when it rains, Camerer's correlation between target earnings and the difficulty of finding a cab when it rains is invalidated.

To analyse the effect of rain on driving conditions, we computed the average number of trips per hour on duty. We found that the average number of trips is 3.39 when it rains, and 3.42 when it doesn't. Interesingly the number of trips is not worst when it rains in the middle of the night / very early morning. This could be explained by a low congestion at that time of the day.



[1]Based on http://www.currentresults.com/Weather/New-York/average-yearly-precipitation.php, we found that it rains about 50 inches a year for about 120 days or rain, for an average of 42 hundredth of inches per raining day. Thus we think that 8 hundredth of inches is a good threshold for one hour

We also calculated the average distance a driver on duty drives per hour. When it rains, this distance is 9.00 miles compared to 9.62 miles when the road is dry (thus a decrease of about 5%). Those two results combined show that the driving conditions are worse when it rains: fewer trips and less distance per hour. This indicates that a single trip takes more time to be satisfied when it rains as the driving conditions are less favorable.

After this assessment we decided to analyze supply and demand for taxis to check if Farber's theory is holding.

- Supply: The average number of drivers on duty per hour under the rain is 7602 while it is 7987 when it does not rain. This is a decrease of around 5% that proves that supply is higher when it doesn't rain. These results are in line with Farber's (who found a decrease of 7.1%). Driving under the rain doesn't improve the wage (and in fact it slightly reduces them as we have calculated it) and result in harsher conditions, therefore fewer drivers accept to drive. This is coherent with neoclassical economics' predictions.

- Demand: We observe that drivers are more occupied when it's raining: the occupancy rate ($\frac{t_{\text{occupied}}}{t_{\text{onduty}}}$) is equal to 74% when it is raining and 70% otherwise. This represents an increase of 5.7% and is in line with Farber's finding of 4.8%. This result corroborates the hypothesis of a surge in demand when it is raining.

These observations are interesting because as demand is increasing, supply is dropping because of harsh conditions and similar wages. To balance these contradictory effects, applying additional charges to passengers when it rains would help limit the drop in supply and reduce demand, these two trends tending towards a point of equilibrium.

4 Conclusion

Through the analysis of a large data set of taxi rides from 2010 to 2013 in New-York City, we have been able to investigate the question of the difficulty to find a cab in the rain. We have found that Camerer's target earnings theory cannot explain the phenomenon. However, our findings are in line with Farber's. If the demand for taxis increases when it rains, the supply falls due to similar earnings and harsher driving conditions. A solution to this problem could be to introduce a price surge when it rains, intended to motivate drivers and cut the demand in order to reach an equilibrium (like Uber is doing).

5 References

Farber, Henry S. *Why You Can't Find a Taxi in the Rain and Other Labor Supply Lessons from Cab Drivers*. No. w20604. National Bureau of Economic Research, 2014.

Camerer, Colin, et al. "Labor supply of New York City cabdrivers: One day at a time." *The Quarterly Journal of Economics* (1997): 407-441.