# Ultimate Take Home Challenge ([Click For Notebook](#))

Exploratory Data Analysis (EDA)
The EDA we conducted on the data provided yield some interesting results in terms of interday and intraday patterns. One note of concern was that the year for the data was "1970" which we were not sure if it was accurate given that electronic records from that time period are somewhat suspect. That said, we will move forward assuming all the information is correct.

Interday Analysis
The char above shows the number of logins that occur on a given day for the provided timeframe of the data. One of the clear things we can see with the data is that there is a cyclical pattern of the data increasing over a seven day period before falling and
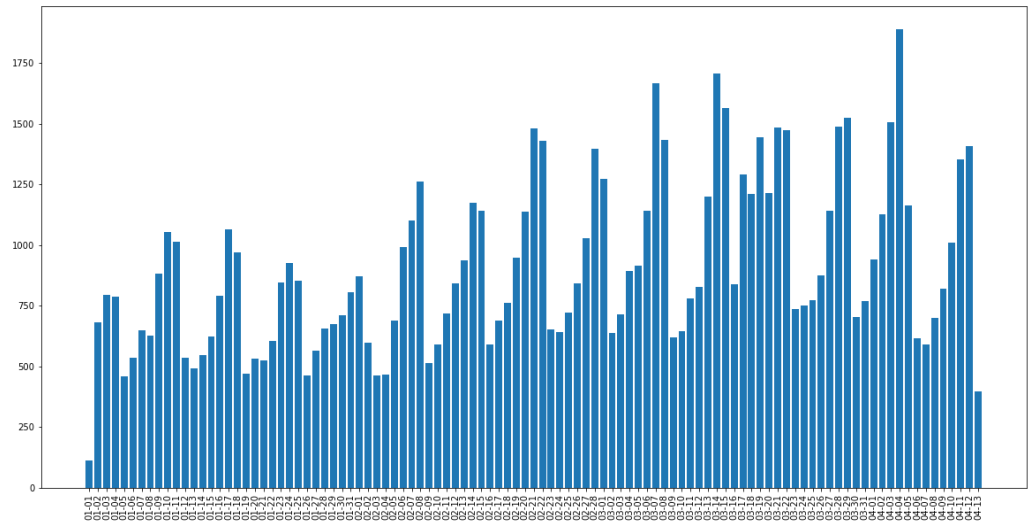
repeating the cycle. In addition, we conduct a Dickey-Fuller test to evaluate whether the data followed a random walk, and the results indicated that the above series <u>is not</u> a random walk. Therefore we can assume there is an underlying pattern as the visualization would imply.

<u>Intraday Analysis (15-minute Intervals)</u>
The variation within a day is also revealing. This graph shows the average logins for each 15 minute time internal. We can see that there is high activity between 2000 through 0500 and
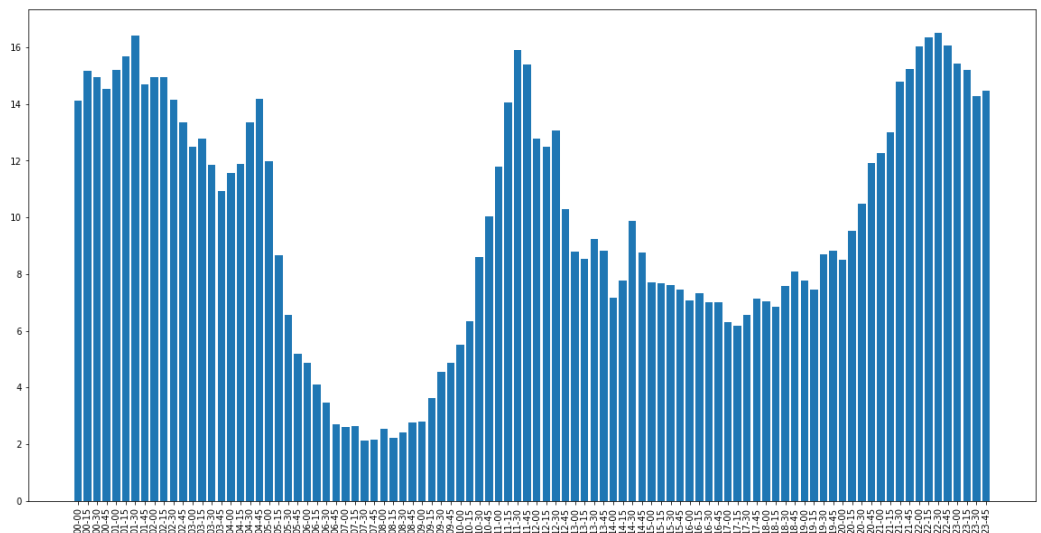
then decreasing activity from 0500 to 0800 and then it increases again from 0800 to roughly 1200 before going through another cycle from 1200 to 2000 but not going as low as the morning hours. This would imply that activity is greatest during the night, lowest during the morning, peaks around noon, and then decreases during the afternoon and evening.

Based on the EDA, if a model was needed to potentially analyze this data, an ARIMA may make the most sense.

Experiment & Metrics Design
The following are my answers to the questions asked:

Key Measure of Success (2.1)
I will work off the assumption that both city managers assume that more inter-city travel results in economic activity gains that are worth the cost of paying the toll bridge tolls (mainly because that would be a different question to test). As a key measure of success, I would work off 'average crossings per day' as the key metric because it would allow for practical testing, and is relatively easy to track as we know tolls are collected and therefore data should be trackable. The key difference would be to consider a "before" level and an "after" level. One important aspect to consider is whether the sponsors have a "level" of success that needs to be considered as there is a difference between "an average increase per day" versus "an average increase of at least 100 or more bridge crossings per day. For now, we will assume that they simply want to see a statistically significant increase in bridge crossings.

Practical Experiment & Implementation (2.2 & 2.3)
I would perceive this as a "A/B" style of experiment where we need to see what the baseline looks like ("A") and then determine what the post-policy data looks like ("B"). In order to establish a baseline, it would in essence be a sampling of past data to establish a "pre-policy" baseline for the number of 'average crossings per day.' One key thing to consider is that the data used should be *before this policy was made public* for two main reasons. First, If the public is made aware that a policy like this is being considered, they may alter their driving behavior in anticipation. Second, if the policy is announced to be implemented on a specific date, then it is reasonable to assume drivers will definitely alter their behavior. As such, you'd want a baseline that wouldn't be affected by news. That said, you could use this baseline to test pre and post announcement to see if there is any change. Once you establish this baseline, you have a means of testing the new levels with some means of measuring the change. As a note - given that we are told that each city has certain characteristics (e.g., Gotham active during the day), we could establish baselines for specific time frames. The ones that seem applicable given the information would be during the week during the day(Monday through Friday from 08:00 to 18:00), during the week during the night (Monday through Friday from 18:00 to 08:00), and then during the weekends given that

both cities are active. Once the policy is implemented successfully (we'll assume operational competency), you would collect the data and have to go through the traditional Data Science Pipeline of Data Wrangling and Exploratory Data Analysis into Statistical Analysis.

Statistical Test(s) (2.4)
This seems like a situation where you can use a *one-sided t-test* quite effectively to test changes between two groups as we are specifically testing to see if there is an increase. This would be standard test you could run since it would work off averages. There are some other tests you could run as well to help verify the benefits. One test we could also consider is linear regression where we have features that predict bridge crossings per day (e.g., day of the week, day of the month, auto-regressive terms, etc.) where we have a "pre" policy and "post" policy period and see if a dummy variable is that "0" at the "pre" era and "1" for the "post" policy era to see if it has an explanatory power. The coefficient (beta) of a variable like this would give us an indication of the measured effect. The last test we could conduct (depending on the level of data we collect) would be a "best match" analysis where we compare "apples to apples" changes and then average them together. For example, let's say we had one-year of "pre" data and one-year of "post" data. What we could do is compare the difference for each commensurate day (so February 15th pre policy and February post policy) to measure the difference. Then, you can average those differences to get a confidence interval to measure the difference. This allows you to compare "apples to apples" as directly as possible. Using all of these tests together would allow you to build a robust measure regarding the change that occurred.

Recommendations & Caveats (2.5)
The biggest drivers of recommendations for a program like this would have to be based on some threshold that needs to be met as I'm assuming both cities have verified that paying the cost of tolls is worth the associated economic gains. At first, we can use the results of a "one-sided t-test" to verify the aggregate change. If a t-test like this shows no significant change, it would be strong evidence that the suggested policy is most likely ineffective and probably should be dropped. If we find changes associated with different time segments (e.g., Weekdays during the day, Weekdays during the night, Weekends, etc.) to be significant for some and not significant for others, recommendations would need to be tailored to the associated gains (and troubles). For example, if we find that the only period that changes would be weekdays during the night, but we also find that the cities have zero confidence it would be worth it to implement the policy for economic reasons, then we'd have to consider not recommending the policy.

In terms of caveats to consider from a statistical perspective, the biggest one would be tied to how long they are willing to let the policy run before the analysis. If the sponsor wants results to be processed quickly (i.e., before a lot of data is collected), it might be

tough to be confident that an increase is actually caused by a policy change. In addition, we would potentially need to consider the "exploration bump" that may occur when people travel to a different city due to it being free. This is common with restaurants where a new restaurant opens and owners get a false sense of demand being high due to initial numbers. This is due to people being open to trying new restaurants but trying a new restaurant *rarely* translates to a restaurant entering the normal basket of a restaurants a customer returns to consistently. Thus, as time passes, demand plummets, because the new restaurant is no longer new and people don't return. We could have a similar issue with a policy like this as we see an initial bump due to people exploring "the other city" but then finding they prefer the options in their own city which results in a reversion to previous numbers. Thus, you'd need enough time to pass to be confident that the "exploration bump" had passed. In the event that you find an "exploration bump" as opposed to a consistent change, it may be evidence to consider doing a periodic deal (e.g., twice a year) to reimburse travel costs to take advantage of people wanting to go back and explore.

Predictive Modeling

We find that four main features are the most predictive for determined if a user is retained:

- Average Rating By the Driver
- Weekday Percentage
- Surge Percentage
- Average Surge

The implication here would be that if the organization wants to leverage the most impactful factors of getting users to be retained it should focus on these main factors. The most important is the Average Rating By the Driver[1] which makes sense given that a rider is more likely to take rides if they have a positive experience. As such, the organization should focus intensely on making sure that any "bad experiences" are dealt with quickly and effectively (e.g., culling out bad drivers, following up with riders who have a bad experience, etc.) In addition, having riders take trip during the week seem to be of importance, which can make sense if the weekends tend to be overcrowded for the service and it becomes a less enjoyable experience. It may make sense to offer some specials and "nudges" to get riders to use the service during the weekday which may help get more riders to become retained. The next two were based on surge pricing frequency the multiplier for surge pricing. This may warrant further exploration to consider how to balance the (expected) increased profits of using surge pricing against the possible effect on converted riders into retained riders. If short profits increase at the expense of "shrinking" the possible customer base, it may not be

---
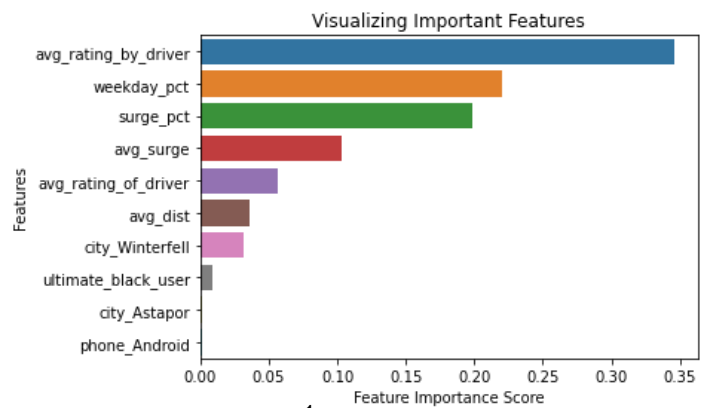
[1] The rider's average rating over all of their trips

in the long-term interest of the organization to always use surge pricing when not appropriate. At the very least, the organization should be judicious in how the utilize surge pricing.

The final model we went with was a Random Forest Classifier Model with a precision & accuracy score of 72.08% and a recall score of 100%.

Thank you for the opportunity! Cheers!
Emre