FICTIONAL BUSINESS OWNER: SAMANTHA STEVENS

*Launching a New Business—Samantha has recently moved to Chicago and is keen on opening a new business in **Sports Adventure & Travel** recreation within the prime real estate of Chicago.*

Samantha has had a few previous related venture startups internationally and is now looking to test the market domestically in the States. She is keen on finding the proper environment that fits her business and attracts clients.

Based on her market research and experience abroad, she expects her revenue to come during/from:

- *Holidays and Weekends*
- *April – August*
- *Possibly Seasonal Business*
- *Non-Corporate Enthusiasts*

During the non-peak periods, she may invest in pop-up stores in other locations, depending on how the business progresses.

Her clients will most probably be walk-in visitors with occasional appointments. One of her priorities is to be central to the CTA (Chicago Transit Authority) Stops, such that the business can be easily reached via foot.

Possible criteria include:

- *Connecting Stations*
- *Central Stations (Hubs)*
- *Demographic Neighborhood Fit*
- *Dense vs Sparse Environments*

If the initial pilot goes as planned, she would be willing to enter into a longer lease term.

POTENTIAL QUESTIONS TO EXPLORE/ANSWER

BASED ON CTA DATA

Initial Insights:

Trends:
Are people taking the CTA "L" as the primary form of transportation, or are other transportations means more frequent. What has been the trend over the last couple years?

Frequency:
Are certain Station Stops being visited more than others? Could this possibly be because of which Line (e.g. RED Line) a Station is connected to?

Temporal Patterns:

Type of Day:
Are their certain Stations being visited more based on the type of day, e.g. Weekday, Weekend, Holiday? This could impact the amount of traffic the business receives.

Seasonal:
Are there any Seasonal related patterns/trends? As the primary business revenue is expected to be between April – August.

Volatility:
Are there particular bursty or drop periods of activity? Ideally we would prefer a stable amount of traffic.

Connections:

Central Hubs:
What is the impact for Station Stops with many junction points (Stations from different lines). For example, Stops with at least 3 Junction Points.

Reliability, Data Skew:
Do we have enough data from all the different lines and Stations, or is the the data coming from only the more frequent (majority) Stations and Lines.
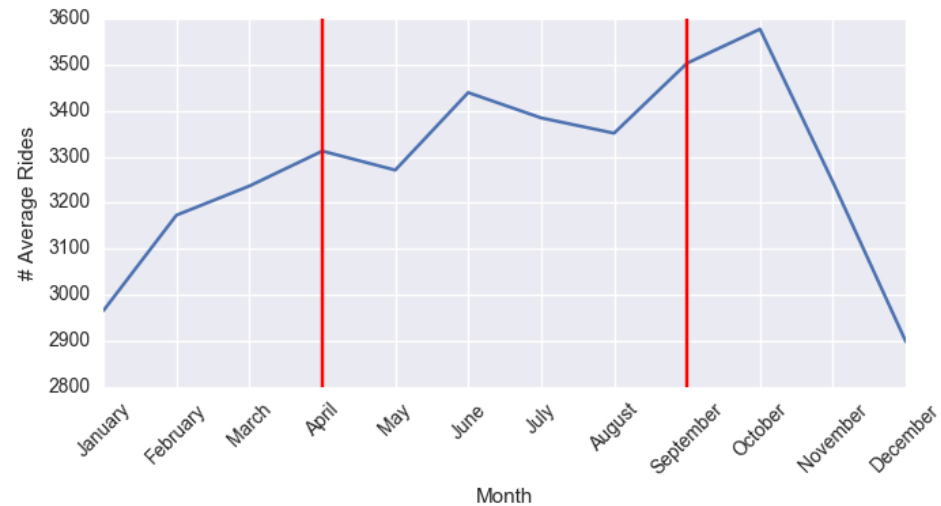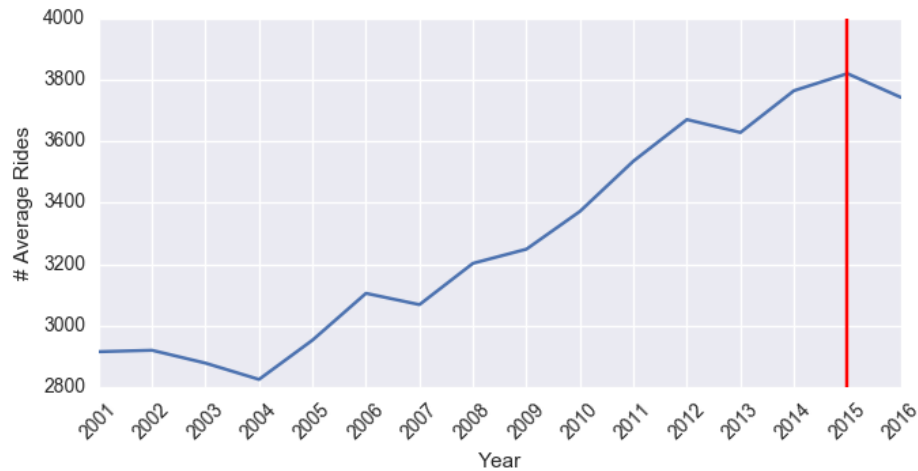
Environment:

Demographic:
Does the projected neighborhood represent the business clientele and/or can we expect clients to visit the neighborhood? Can this possibly be representative of the CTA location stops?

Sparsity:
On average, does the traffic represent very dense or sparse traffic footfall patterns?

TRENDS

There has been an increasing trend of patrons taking CTA transit. In addition, the months of April – August look to be in good shape, particularly in contrast to the Winter months. Samantha could possibly even extend till October.

It should be noted that in 2016, data was only included through July 31rst, so this could have an impact at the latter months. Alternatively this data can be filtered up till 2015.
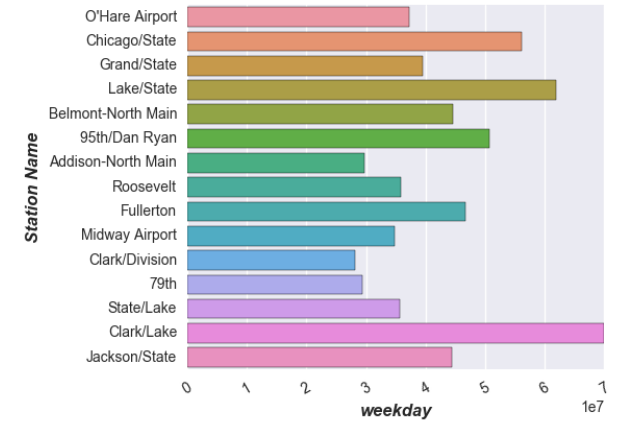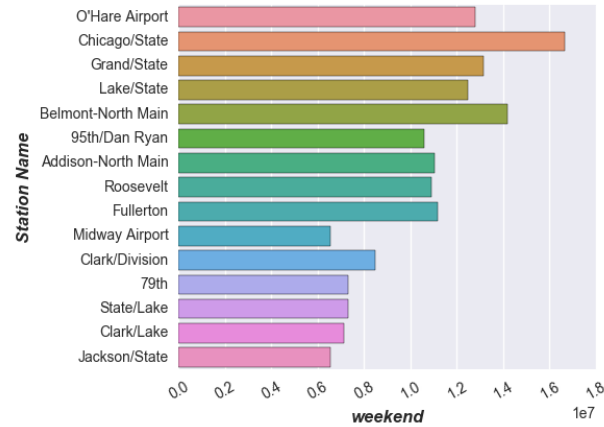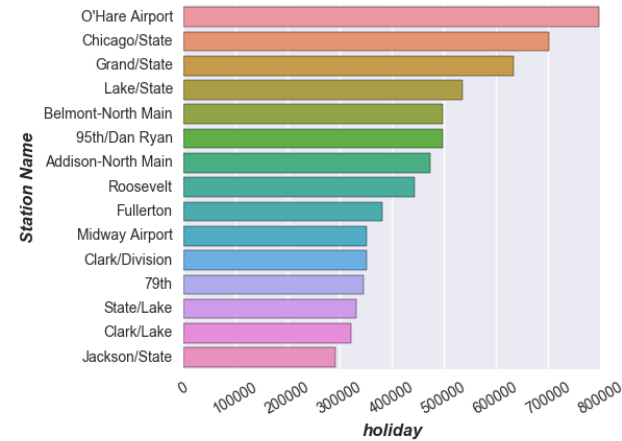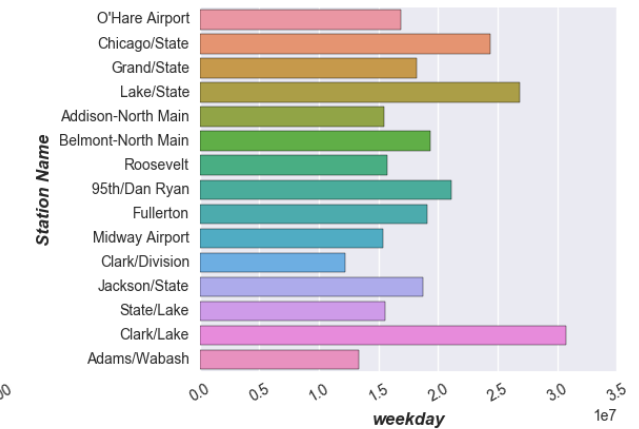
LINE RIDERSHIP

Line Ridership (April-August)
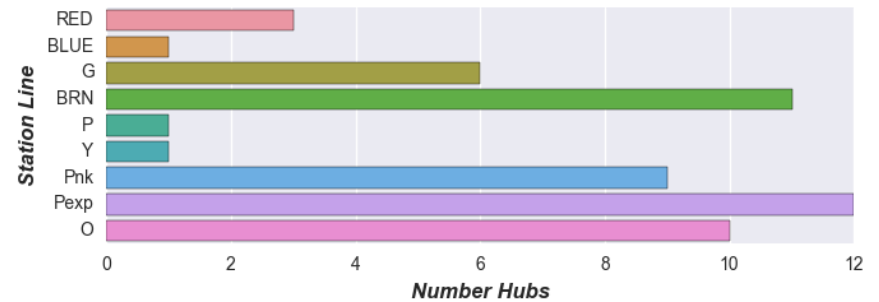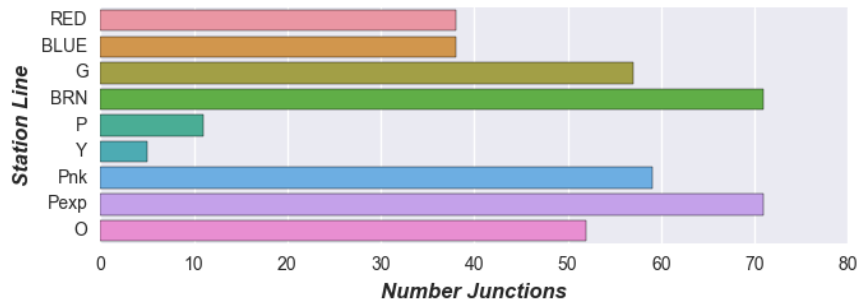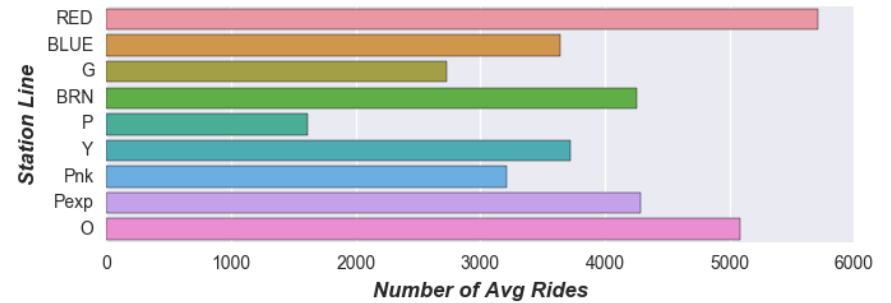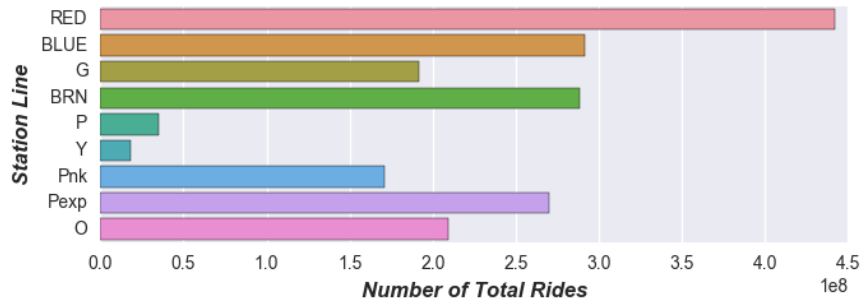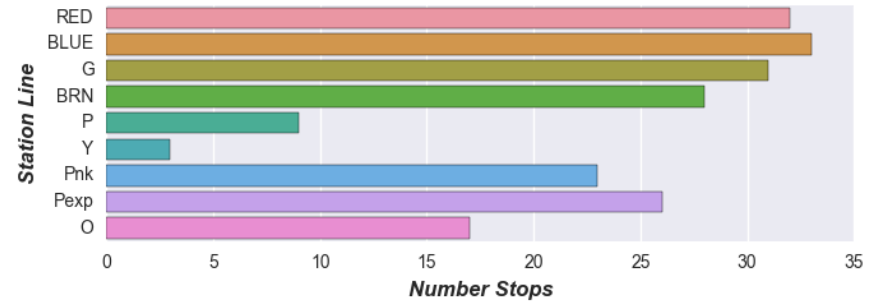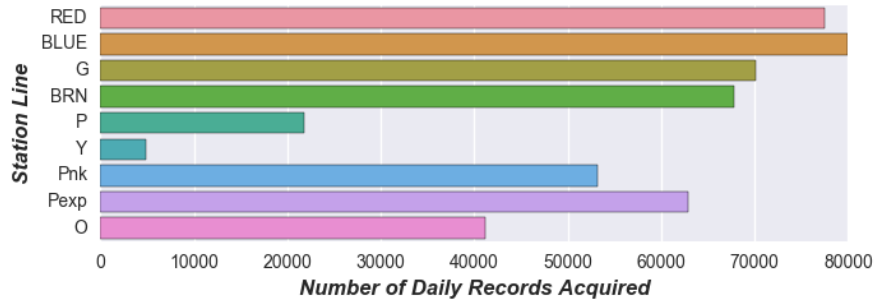
STATION RIDERSHIP PATTERNS

Overall Station Stop Distribution

April-August Stop Distribution
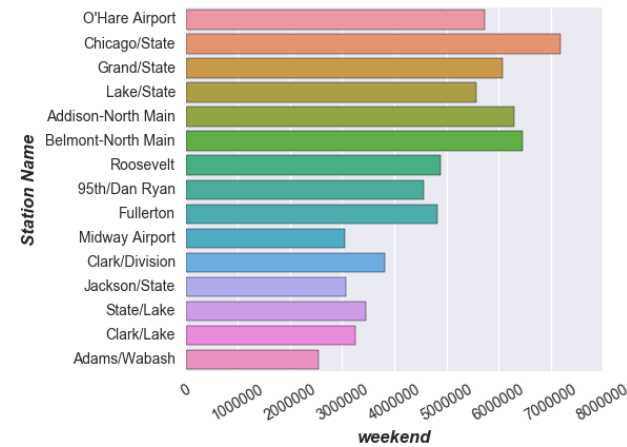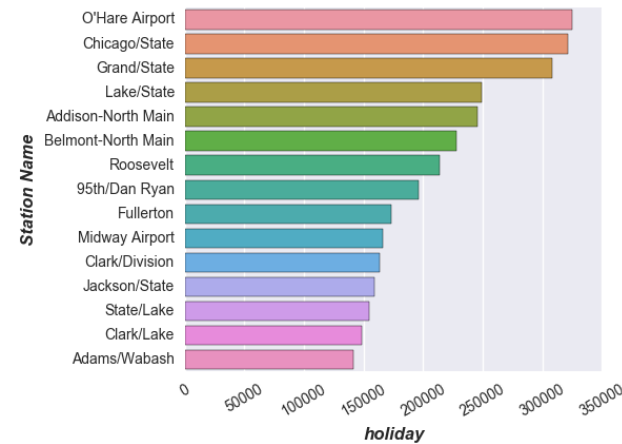
SELECT STATION VOLATILITY
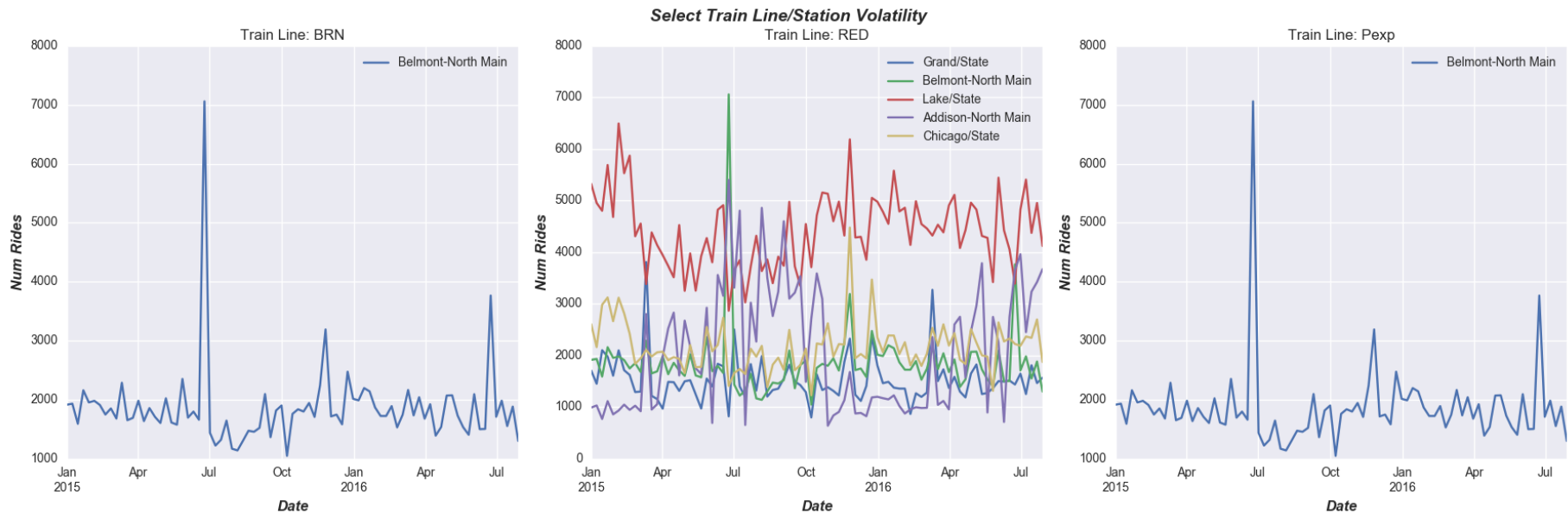
Narrowing down on just the select Train Lines of Interest (RED, BRN, Pexp) , as well as Train Stations around the <u>State and Main</u> *Station regions based on business owner criteria and gained insights.*

GEOGRAPHICAL VIEW

April-August Holiday Rideship
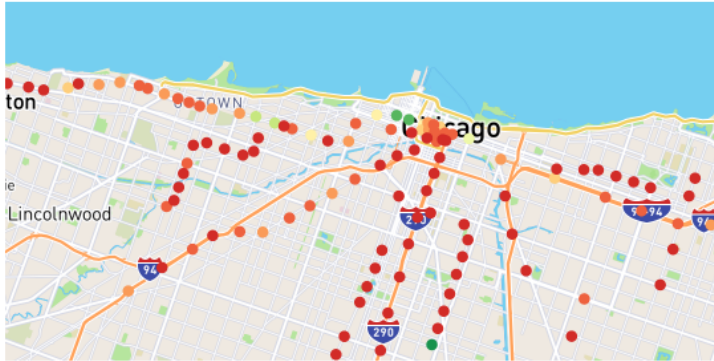
April-August Weekend Rideship

April-August Weekday Rideship

April-August Low Volatility Index Rideship

0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0

INSIGHTS

WHAT JUMPED OUT ?

Line Ridership:

Line Skew:
There are a few Train Lines where we did not gather enough daily records from (e.g. P and Y Lines). In the absence of more data, we could attempt to resample the data (e.g. SMOTE).

High Positive Correlations:

- *# Stops vs # Rides*
- *# Daily Records vs # Rides*
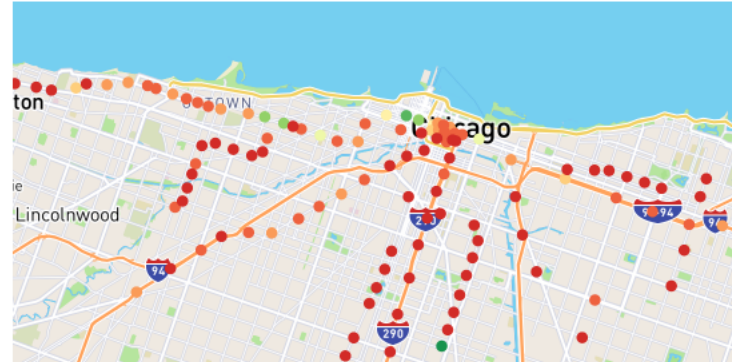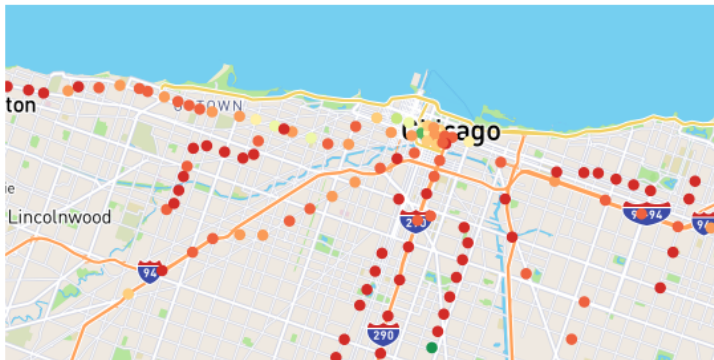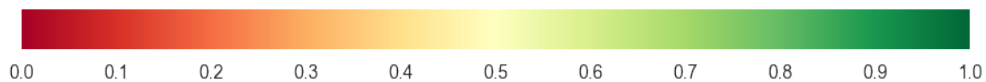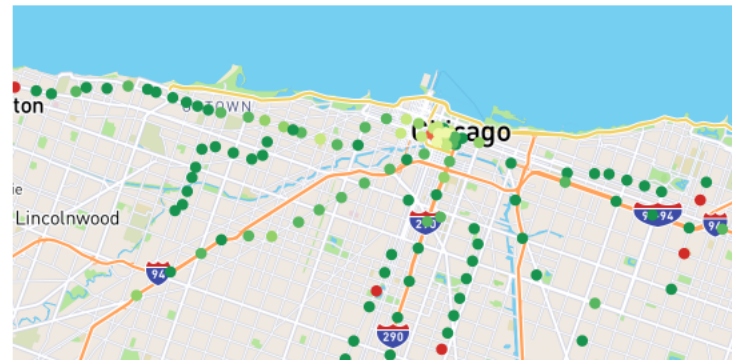- *# Junctions vs # Rides*

We didn't gather a high correlation with the *Number of Hubs* on a Train Line and Ridership in April - August.

Initially focus on Lines:
RED, BLUE, BRN, Pexp

Station/Stop Ridership:

Type of Days:
During the prime periods of April – August and Weekend & Holidays, we have a few leaders in the pack to look at. We decide to focus our attention on *State and Main* related area Stations.

These regions seem to fit our selection within the RED, BRN, Pexp Lines, while still yet close in proximity to other junctions.

In addition, these regions do not exhibit a large drop-off outside of Weekend, Holidays or April-August periods.

Alternatively we could have employed a SARIMAX model for seasonal effects.

Volatility:

Stability:
Low Volatility (higher stability) is of preference.   While all Train Lines seem to have exhibited some volatility, the RED, BRN, Pexp seem to be on par with the others during prime periods of revenue.

Grand/State on average seems to have the lowest volatility in contrast to Lake/State which has the highest volatility of the selected Stations on the RED Line.

ADDITIONAL DATA

GATHER MORE INFORMED INSIGHTS

Line/Station Analysis:

Lack of Daily Records:
As noted in *Line Ridership* plots, some Lines have significantly fewer records collected (e.g. P, Y). For this reason, we may decide to pursue other Train Lines at this point. Although these Lines have *limited number* of Stops.

Timestamps:
All data is Daily recorded, it may be beneficial to have Timestamps recorded hourly, such that *Congestion Periods* with a finer granularity can be determined and binned*.*

Contextual:
Contextual analysis via Special Events, Holiday Trains.

Proximity:

Degrees of Separation:
How many Station Stop *Hops* are the major Stations from the potential nearby CTA stop. This could impact willingness of patrons to visit the location.

Distance to Business:
How far is the potential business from a CTA stop? Are their potential Lease Options already being looked into?

Convenience:
Collecting Survey information based on how convenient potential locations could be. It is possible some potential clients may not extend outside their comfort zones.

Aggregation:

Modes of Transportation:
How does the CTA information compare with other modes of Transportation, e.g. walking, bicycling, buses.

Segmentation:
Based on particular events or trips, Samantha may prefer to have particular segments of clients to cater to, offering a more *personalized and recommended* service.

Competition:
Aggregate information based on nearby similar competitors.

Economy:

Projections:
What do experts believe the projection over the next 1-2 years will be. As it will take some time to setup the business, the next few months are less important than the future growth. Furthermore, we can employ our own predict models, e.g. ARIMA, Linear Regression/OLS.

Current Events:
With Donald Trump as the new president, should we expect any changes that could affect Samantha's business, e.g. possibly new tax codes, propensity of clients to spend.

REPRESENTATION OF ANALYSIS

DRIVE DECISIONS

Audience:

Summarization:
Providing a high level summary of the analysis, rather than drilling into the specific details. Some examples are Highlights, Heatmaps, Geographical Region overlays.

Interactivity:
Having a hands on approach that caters towards questions, e.g. if the threshold were changed in some manner, how would this affect the analysis. This could be applied via sliders and tools like d3.js.

Eye Grabbing:
Is the analysis "Sexy" enough to grab attention?

Criteria:

Criteria:
Assess whether the initial criteria set by the business owner, Samantha can help her explore particular regions.

Tradeoffs:
Would certain criteria be prioritized over others, and can these decisions be explored from the analysis.

Visualizations:

Forms of Analysis:
Different visual forms can present the data in a variety of ways, which can have an impact on what decisions are made, e.g.:

- *Interactive (d3.js)*
- *Sliders*
- *Heatmaps*
- *Geographical Overlays*
- *Time Series*
- *Large Deviations/Outliers*
- *Normalcy/Skews*

Ratings:

Index:
Providing a single metric that encapsulates a Rating metric of performance or recommendation could be helpful. For example in Football, we have the Quarterback Rating (QBR).

RESOURCES

Visualization Links via Plotly:

Day Types:
http://plot.ly/~ari.kamlani/4/april-august-weekday-rideship/
http://plot.ly/~ari.kamlani/2/april-august-weekend-rideship/
http://plot.ly/~ari.kamlani/0/april-august-holiday-rideship/

Volatility Index:
https://plot.ly/~ari.kamlani/8/april-august-low-volatility-index-rideship/

CONTACT

Connect with Me:

arikamlani.com
415-926-1221
akamlani@gmail.com

LinkedIn:
@akamlani

Twitter:
@akamlani