**Springboard Data Science Intensive**

Capstone Project Proposal

**Jonathan Calindas**

2/21/2017

**Factors that Affect Sales at the TKTS Booths in New York City**

**Background:** The Theatre Development Fund (TDF) operates the tourist landmark TKTS Booths in New York City that sell same day tickets to Broadway and Off Broadway shows for up to a 50% discount. There are four TKTS booth location throughout the city. The busiest location is Times Square, at the heart of the theatre district. Other locations are the South Street Seaport, Brooklyn, and Lincoln Center.

A mobile app has been developed for iOS and Android devices that will allow the user to see the shows that are listed at the booths without having to be there in person (they will still have to purchase tickets in person). The mobile app also serves as a comprehensive index of all the shows currently playing in New York City, including shows that are not on sale at the booth.

**The Problem:** Sales at the booths are subject to some fluctuation, but TDF management do not always know the cause of rises and dips in sales. They will make various assumptions, but there is no hard data to back those assumptions up. With a greater understanding of the forces that affect sales at the booths, management can make better decisions and can be better prepared for a dip in sales or to take advantage of a rise in sales.

**Study Parameters:**

1. Study dates: 7/1/2016 to 12/31/2016
2. Although the booth sells Off-Broadway shows, we will only consider Broadway Shows.
3. To minimize the complexity of the natural cycle of activity within one week, we will examine the data by week.

**Datasets to be studied and compared:**

1. Tickets sales at each of the booths by week and by show.
2. TKTS App logs
   1. Overall activity
   2. Popularity of shows viewed
3. Weather (booths are located outdoors)
4. Number of shows playing on Broadway
   1. Number sold at the booth
   2. Number of shows that sell at full price and don’t appear at the booth
   3. Shows will be ranked according to three criteria:
      1. Actual sales
      2. Demand for the show (as determined by app views)
      3. Frequency of show offered at the booths (occasionally, frequently, etc)
5. Tourism Level
   1. Number of tourists in the city per week.
   2. Ratio of Tourists vs. Locals attending shows
      1. Tourists generally attend well known shows that have been playing a long time, as well as shows with name actors in the cast.
      2. Locals generally attend shows that are newer because they have already seen long playing shows. They also tend to frequent the Brooklyn and Lincoln Center booths.

Much of the data will already be stored within TDF’s databases. However, I will procure data for weather and tourism from the open web.

**Methodology**: The bulk of the work will be to assemble and shape the data in a way that can be examined weekly. These plots will then be compared to the sales plot to determine whether there is any correlation between the two.

**Deliverables**:

1. A Jupyter Notebook containing the code used to assemble the datasets.
2. A Jupyter Notebook plotting the various datasets and containing analyses interpreting the data.
3. A Powerpoint Slideshow that combines all the insights and charts from #2 into a format suitable for presentation.