Springboard Data Science Intensive

Capstone Project Milestone Report

Jonathan Calindas

4/16/2017

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: 1/1/2016 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:**

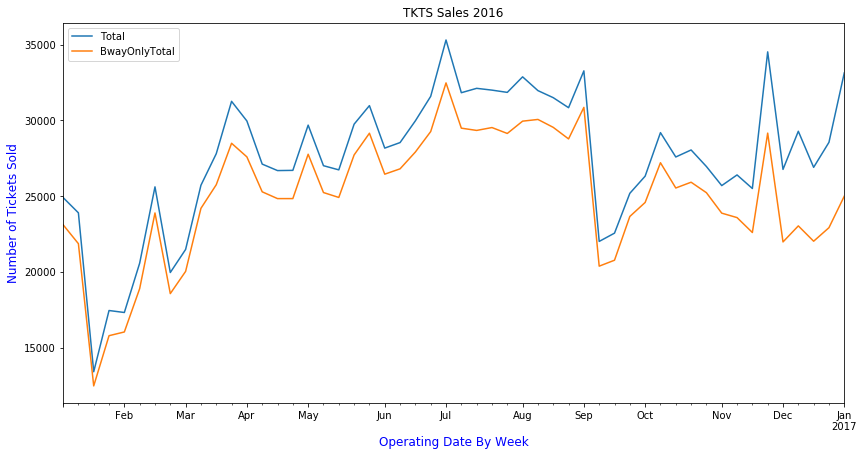
1. Tickets sales at each of the booths by week and by show.
   1. Sales figures are maintained weekly (not daily) and are kept in individual spreadsheets for each week of operation. Data had to manually be compiled into one spreadsheet that could be imported as a whole table.
2. Ticket sales for each show for all of Broadway as published by Variety.
3. TKTS App logs
   1. Overall activity – this is measured by the volume of calls to the API that supplies the app with booth data. Each call to refresh the list of shows in the app is measured towards this number. Calls to other parts of the app, for example, FAQs about the booths will not be included in this tally.
   2. Popularity of shows viewed. Users can view the details of a particular show in the app. They can do this by scrolling on the list of available shows at the booth, or by searching for the show and viewing its details. Details include the show description, the address of the theatre, and other information.
4. Weather (booths are located outdoors). Data is downloaded from the NOAA for the year 2016.
5. Number of shows playing on Broadway
   1. Number sold at the booth. This is taken from sales data as well as from the software used by the sellers at the booth to post info to the app.
   2. Number of shows that sell at full price and don’t appear at the booth. In addition to shows sold at the booth, a select number of shows sell very well and do not sell discounted tickets at the booth. A list of full show openings and closings are published at IBDB.com.
   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)

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.

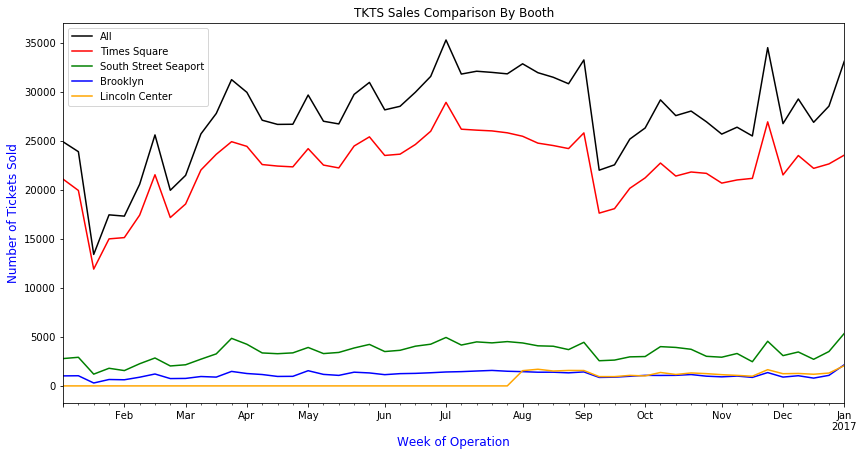
# TKTS Sales

Before we examine the various factors that may influence the rise and fall of sales at the booth, it is important to establish a baseline with regards to the natural changes in booth sales.

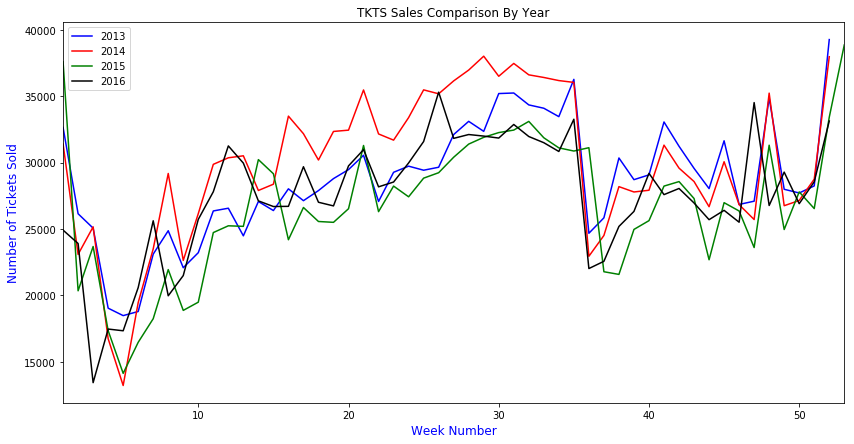


The blue line shows the total sales in all the booths for both Broadway and Off Broadway shows. The orange line shows only Broadway sales. We can see that Broadway represents the large percentage of shows sold at the booth and the gap between them only widens as more tickets are sold. The only exception to this is near the end of the year, in the months of November and December. This is attributed to the big show, Radio City Christmas Spectacular that is sold only at the end of the year. This is classified as an Off Broadway show, even though it is a big venue. Otherwise, we can also see the both lines dip and fall in the same way so factors that affect sales at the booth general affect both Broadway and Off.

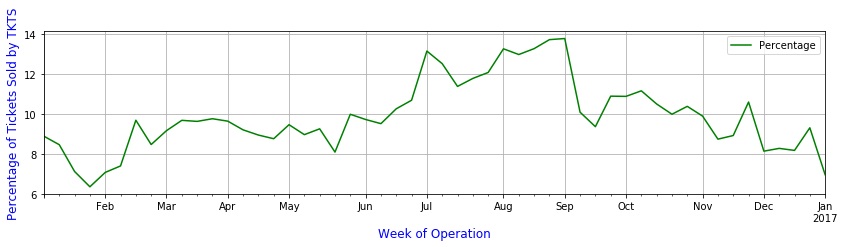
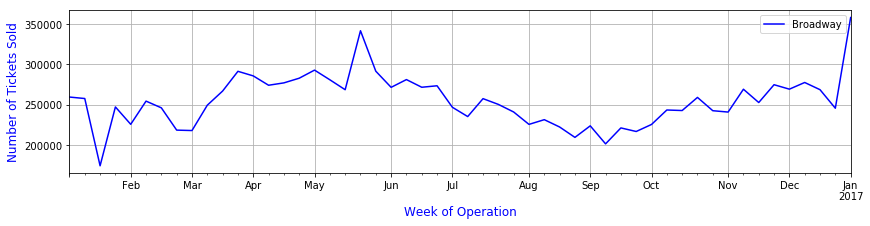
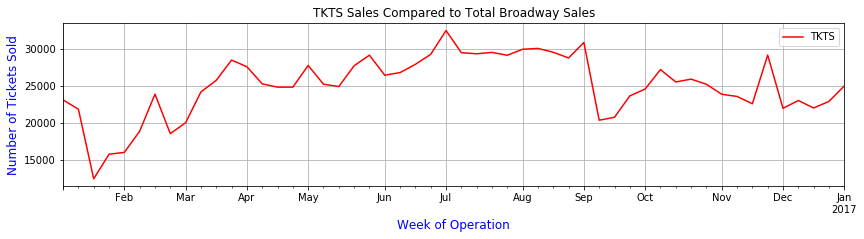
We can also see that the months of January and February and the worst months of the year with the summer months being the best months for the booth. We also see a precipitous drop in the first week of September, probably attributed to the start of the school year. We also see great spikes in sales in the week of Valentine’s Day, the week of the Fourth of July, the week of Thanksgiving, and the week of Christmas.



The chart above shows ticket sales by booth. We can see here that the vast majority of ticket sales are sold at the Times Square location. Note that the Lincoln Center location did not open until late July, which is why it is zero up to that point. We can see that every location generally follows the rise and fall in sales proportionally in tandem, so all booths see the peaks and valleys at the same time. If it doesn’t seem to rise as much in the chart above, it’s only because the scale is much smaller.



When we compare the trends compared to previous years, we can better determine the natural rise and fall of sales throughout the year. The minor shifts are attributed to the migration of days as they fall within the week number. For example, in January 2016, week 1 ended in Jan 10, whereas in other years week 1 ended much earlier. We see that the lowest point in sales always occur in the January after the week of New Years. Then there is a gradual rise in the spring months, and peak sales in the summer months. There is always a precipitous drop in the first week of September, a rise in sales in the fall, and a sudden spike in sales during special holidays, such as Valentines Day (Feb 14), the Fourth of July, the week of Thanksgiving, and the week of Christmas and New Years.

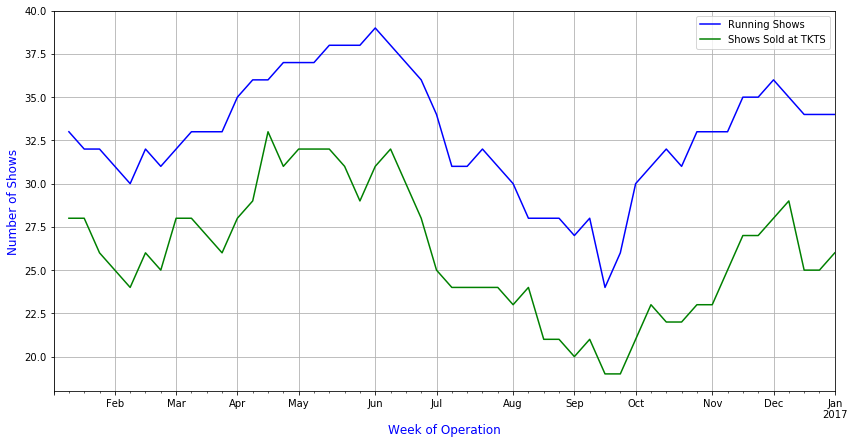
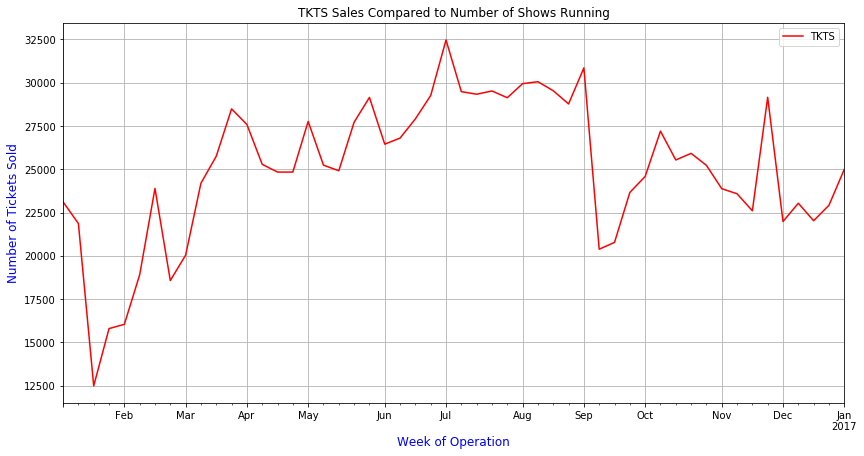


Now we will compare TKTS sales compared to all of Broadway. The first chart contains Broadway only sales for the booths. The second chart shows sales for all of Broadway, which includes full price tickets sales and sales through other ticket outlets. We need a different graph because the scale is much higher in the second chart. The third chart shows the percentage of total ticket sales sold through TKTS (market share).

We can see that Broadway sales peak during the spring and the fall months. This is generally when new shows open and close. We see a sharp rise in sales in the end of May to the beginning of June, which is attributed to the end of the Broadway season and the Tony Awards. Shows see a rise in sales due to the excitement of the Tony awards and it is the when Tony voters come to see the shows.

As we compare Broadway versus TKTS, we also see the rise in sales during the fall and spring months, but sales continue to rise into the summer for TKTS, while the rest of the industry falls during the summer. Both charts see the lowest point of sales in January, but Broadway is generally steady during holidays, with the exception of the Christmas/New Years holiday.

From the third chart, we see that TKTS has the highest market share during the summer months, with peaks during specific holidays.



In the charts above, we compare TKTS sales to the number of shows playing on Broadway in any given time. The blue line indicates the total number of shows, while the green line indicates the number of shows that are sold at the booths. The slow rises during the spring and fall months show the season as shows open during the fall and spring months. Shows try to stay open until the Tony awards in June, but see some shows close during the winter months of January and February, and a large number of shows close right after the Tonys. They begin to open again in the fall from September on.

When we compare the blue line to the green line, we generally see that about the same five or six blockbuster shows sell at full price and not at the booths, with the exception of a few shows that will sell really well at the start and then begin to sell at the booth after a while.

We see from the chart, however, that the number of shows generally do not affect the short term level of sales at the booths, although sales are generally high (along with the industry) during the spring and fall months. The booth still does well, however, with the shows that remain after the Tonys and the fewer number of shows still playing.

**Other Notes About Sales Data**

In compiling the sales figures above, I chose to only use the number of tickets sold rather than the actual dollar amounts in sales. This simplifies the data a great deal, but brings with it some important caveats. Some shows will sell at the booths a smaller discount percentage than other shows, so their dollar amounts may be higher even if they sell fewer tickets. Also, during peak Broadway season, Tony voters are usually comped, so even though tickets are high, actual dollar figures will probably be lower or remain even. Further investigations into these figures should yield more insight into the behavior of ticket buyers.

# Milestone Report

Now that we have an idea of how a typical year for the booth looks like, the next sections of this report will investigate the effect of weather on sales at the booth, and then we will investigate how the TKTS App, available on iOS and Android, affects the sales at the booth (if at all) and how app activity correlates with the levels of activity at the booth. We will also investigate if interest in specific shows in the app correlate to sales at the booths.

The trends in these figures may not be as readily apparent, so we will use some inferential statistics to compute the level of correlation, if any,