Project: NYConcerts

Description
The project is essentially a **web scraping** tool to find details on NY concerts. It scrapes the popular website songkick.com and extracts details on every upcoming concert in the NY area. Details on concerts include artist, venue, location and exact address, date, starting ticket price and venue capacity.
The data found is then stored in an **sqlite** database which is later analyzed and visually plotted using **pandas** and **matplotlib**. The attached database(concerts.db) holds information of upcoming concerts over a three week time period, but the program can scrape the web for all the NY concerts as provided by songkick(4 month time period of upcoming concerts).

Code Structure
The code is provided as a notebook file(.ipynb), hence its structure can be easily viewed in Jupyter where comments are also included.

Instructions
1)Save both NYConcerts.ipynb and concerts.db in the same directory.
2)Open NYConcerts.ipynb in Jupyter. Don't run all cells at once. There are two cells that have a "caution!" comment. One deletes the current concerts.db and creates a new empty database. The other cell runs the scraping again which is unnecessary since the concerts.db is already provided.
3)Run the cells one by one from the top(skipping the two cells with "caution!" comments)

Dependencies
Major modules imported:
- BeautifulSoup
- sqlite3
- pandas
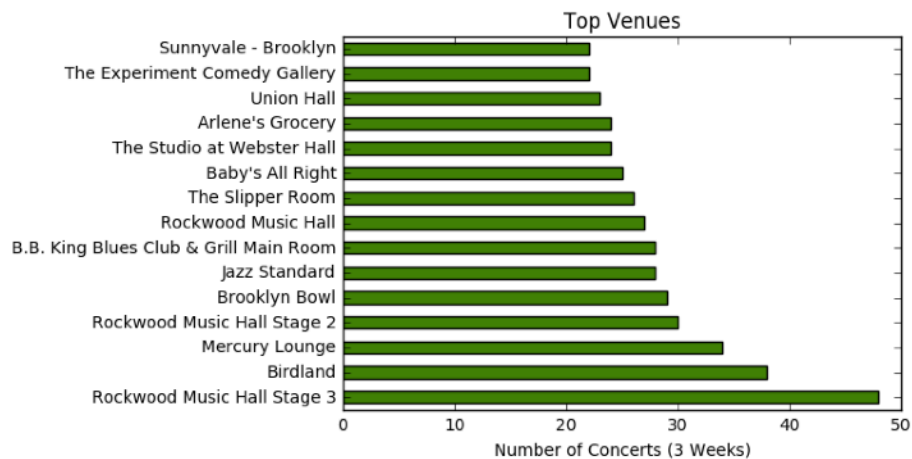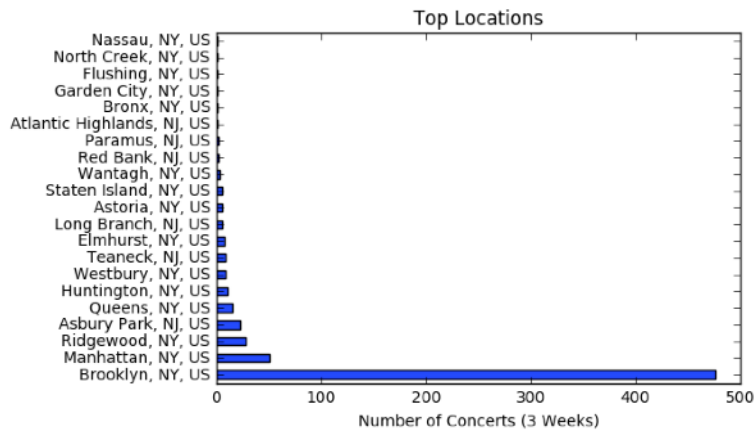- numpy
- matplotlib
Data files:
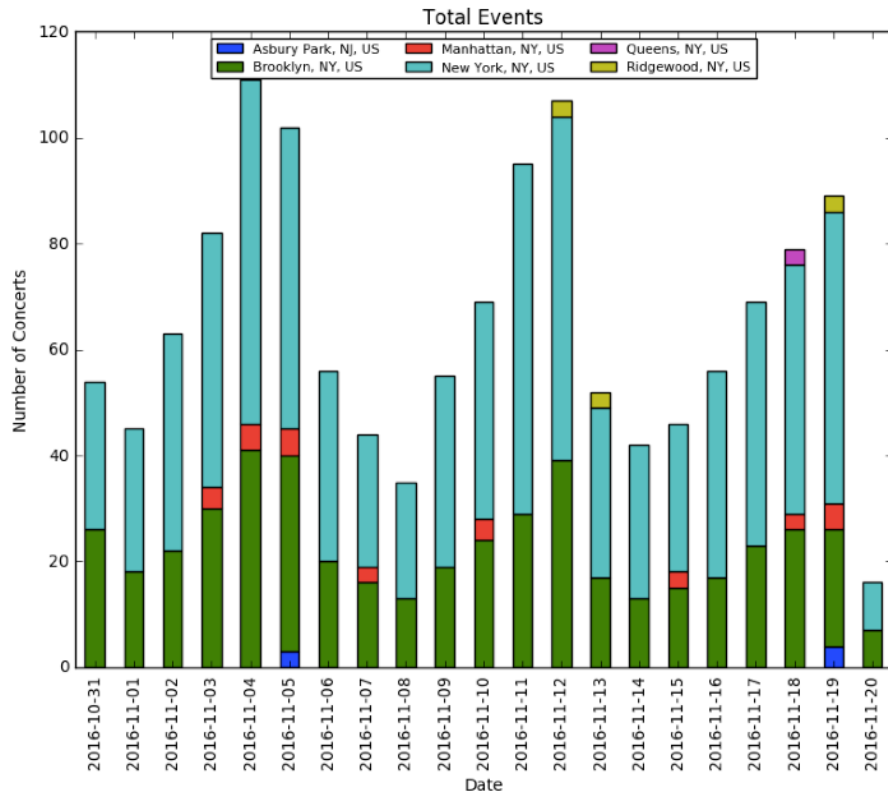- concerts.db

Problems Encountered
- Modifying the data types of the scraped information. Many details were found as object data types on the web source code. For example, it took a while to figure out how to convert date information to proper data types that could then be used in sqlite and pandas.
- Plotting required a lot of modifications to the database such as grouping and counting values of a specific column. Any data that was not numerical was difficult to analyze.

- Scraping multiple web pages for information is time consuming.
- In general, putting together all the scraped information to form a meaningful analysis proved challenging.

Screenshots

More screenshots

Total Events

We can observe that most concerts are held during weekends, and that many concerts take place in Brooklyn, more than any other borough of NY.