Kara Cushman  
DS-SEA-3  
September 7, 2016

**Problem Statement and Hypothesis**

Liquor stores are a common sight in many distressed neighborhoods. But does the presence of liquor stores actually cause crime? In this paper, I use data on the locations of alcohol outlets in the City of Seattle, merged with detailed incident crime reports, to evaluate the effects of alcohol outlet openings and closings on local crime rates.

Main question - Can we predict higher crimes rates in beats with liquor stores?

My hypothesis is that we will be able to predict high crimes rates in Beats with liquor stores.

**Description of the Data**

Like many other major US cities, the City of Seattle has an online open data portal (http://data.seattle.gov). This portal currently hosts 3172 different datasets across a wide range of topics and date ranges (e.g. Code Violations, Sold Fleet Equipment, Building Permits).

I used two primary datasets for this project: Seattle Police Department Police Report Incident & 2016 Active Business License Data.

Crime Data

The City currently has several datasets of all crime incidents, going back as far as 1990, as well as all 911 incident responses. For the purposes of my project, I chose to use the Seattle Police Department Police Report Incidents.

1. This dataset contained the following 18 columns:
2. RMS CDW ID
3. General Offense Number
4. Offense Code
5. Offense Type
6. Summary Offense
7. Summarized Offense Description
8. Date Reported
9. Occurred Date or Date Range Start
10. Occurred Date Range End
11. Hundred Block Location
12. District/Sector
13. Zone/Beat
14. Census Tract 2000
15. Longitude
16. Latitude
17. Location
18. Month
19. Year

The dataset had 661,553 rows, with each row representing a single crime incident.

*Crime Data – Data Cleaning Process*

While the crime dataset was almost all complete, there was some basic cleaning required. I decided to filter the dataset to crimes occurring between the years of June 30, 2015 and June 30, 2016. I did this to correspond with the dates in the active business license dataset to ensure I wasn’t evaluating crimes around liquor stores that were not yet open or had closed.

I also filtered the crime dataset to only include offenses tagged as Murder, Assault, Robbery, Disorderly Conduct, Property Damage, Car Prowl, and Car Theft.

The crimes I filtered out were less dependent on location (for examples Forgery), and since my question is heavily based on location, I decided to eliminate them.

Business License Data

This dataset contained the following 7 columns:

1. Business Legal Name
2. Ownership Type
3. Trade Name
4. NAICS Code
5. NAICS Description
6. License Start Date
7. City, State, Zip
8. The dataset had 92,671 rows, with each representing a single business license.

*Business License Data – Data Cleaning Process*

The business license dataset was also all complete, but, again, there was some basic cleaning required.

The dataset contains all active business licenses, so I had to filter the data to only list liquor stores. This field was under the column NAICS Description labeled “Beer, Wine, and Liquor Stores.” I also filtered to ensure the location was listed to be in “Seattle.”

Merging the Data

I created two different data frames that merged the two datasets. The first was based on a groupby I performed. The DF included beat and then total count of liquor stores and total count of crimes per beat. The second data frames I created was grouped by Beat with total count of liquors stores per beat and total count for of individual crimes.

**Background on “Beats”**

The Seattle Police Department operates within a framework that divides the city into five geographical areas called "precincts." Under this model, neighborhood-based enforcement personnel in each precinct assume responsibility for public safety management within their geographic area. These officers serve as crime-prevention and law enforcement resources in the communities they serve.

Seattle is divided into five geographic areas. Within those areas are the 5 precincts or police stations: North, East, South, West and Southwest. Precinct boundaries were determined through consideration of neighborhood boundaries, geographic and other natural boundaries.

Each precinct contains smaller geographic areas called Sectors. There are 17 sectors total in the city.

Each of these Sectors are divided into between 3 smaller sections called Beats (i.e. Ocean sector has three beats O1, O2, O3) These are the areas that individual patrol officers are assigned responsibility for.

I evaluated total crime and liquor stores based on these Beats.

**Descriptive and Exploratory Analyses and Visuals**

****

****

**Modeling and Model Evaluation**

Clustering

Random Forests

**Conclusion and Key Takeaways**

My hypothesis was incorrect in that liquor stores do not seem to affect crime in a notable ways. As it turns out, the City of Seattle has surprisingly few liquor stores, and many that were included in the dataset are wineries or breweries.

In the future it would be interesting to examine liquor licenses and crime and do so in smaller geographic areas such as square block.