**Kara Cushman  
DS-SEA-3  
August 15, 2016**

**DRAFT PAPER**

**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.

*Can we predict an increase in crime in Seattle near 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.

This dataset contained the following 18 columns:

1. RMS CDW ID
2. General Offense Number
3. Offense Code
4. Offense Type
5. Summary Offense
6. Summarized Offense Description
7. Date Reported
8. Occurred Date or Date Range Start
9. Occurred Date Range End
10. Hundred Block Location
11. District/Sector
12. Zone/Beat
13. Census Tract 2000
14. Longitude
15. Latitude
16. Location
17. Month
18. 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 January 1, 2016 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 out crimes described as “Animal Complaint”, “Extortion”, “Embezzle.” “Forgery”, “Harbor Calls”, “False Report”, “Illegal Dumping”, “Stay out of Area of Drugs”, “Reckless Burning”, “Theft of Services”, and “INC - CASE DC USE ONLY”.

Most of these crimes are not location-dependent like a robbery or an assault, and therefore I did not want them to interfere with the model, which is heavily based on location.

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

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.”

This narrowed the data to a mere 59 rows.

**Conclusion**

Does the presence of alcohol outlets actually cause crime or are alcohol outlets more likely to open in declining neighborhoods? Hopefully I will find out.