

Pioneering Entrepreneurial Innovation in the Midst of National Crisis

Christopher Fuhrmeister

November 17, 2020

2. Data

2.1. Sources

To provide answers to the above questions posed in sections 1.2 and 1.3, a wealth of data will be used to compare metrics among neighborhoods/districts in Portland to provide valuable insight into and/or predict districts in which a new business might thrive.

Data will be collected from Foursquare and Yelp APIs for use to cluster and analyze business activity within individual neighborhoods, GIS-provided geojson data/shape files for plotting boundaries on maps in Folium (including crime statistics), and an array of ACS data sourced from recent US Census Bureau (including consumer spending, education completion levels, and family habitation etc.).

2.2. Usage

In this project, the above-referenced data sources will be used to provide a data-driven resource tool for an entrepreneur to explore, examine and use findings to determine the location with the highest probability of success for a new craft brewing facility/taproom in the city of Portland with respect to the questions discussed above.

2.3. Features

Features will be selected from merged datasets cleansed of improper values (ex. NaN, None) using Pandas, converted into numerical arrays using NumPy, normalized to enhance comprehension and standardization for machine learning algorithms, and executed within a regression model to produce meaningful inferences.