

Project Proposal

Urban Analytics: Crime, Community Services, and Airbnb Pricing in Seattle

Team member: Chenchen Jiang, Hui Du, Jianjian Liu, Jiayang Liu

Background

The datasets we choose are from data.seattle.gov and insideairbnb.com :

- Current City Properties: the properties currently owned by the City of Seattle.
- SPD Crime Data, 2008-Present: including incident date and time, expanded victim types, location data etc.
- Emergency Food and Meals Seattle and King County : a list of emergency food (meals, food banks, etc.) available in Seattle and King County.
- Seattle Airbnb Activity: the price and the income for each listing for the last 12 months.

There are several use cases for the datasets: crime condition and social impact analysis, homestay market research, and Airbnb rate prediction.

This project represents an innovative and beneficial use of data science because it seeks to intersect public safety with economic insights by correlating crime rates and emergency food services with Airbnb pricing dynamics in Seattle. By utilizing advanced analytics, the project aims to reveal patterns and causal relationships that can inform community support measures to potentially reduce crime. Additionally, it empowers residents and visitors to make data-informed decisions regarding their accommodation choices based on safety and affordability, enhancing the living and travel experience in the Seattle area.

Business Goal and Purpose

The primary objective of this project is to predict the listing distribution and the price fluctuation of Airbnb in Seattle. This exploration will be accomplished by analyzing the Crime situation and the Emergency Food and Meals distribution in different neighborhoods, considering factors such as victim location, offense types, emergency food location. This analysis and application will provide valuable insights for guiding neighborhood security plans, devising effective marketing strategies and enhancing homestay sales predictions.

Methodology

Our analysis will rely on several key datasets, including Airbnb listings, crime reports, and locations of emergency food supplies. To gain a comprehensive understanding of the data and establish relationships between variables, we will employ data mining techniques and exploratory data analysis. This will encompass data cleansing and preprocessing, cluster and correlation analysis, the application of data mining models, and model evaluation and validation and more. Below are the modeling and user interface proposals:

Data Mining Model

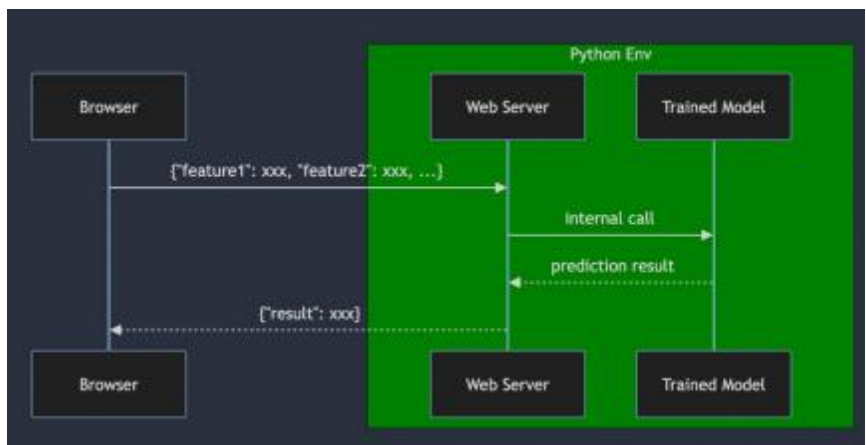
1. Classification Models: employ algorithms like Decision Trees and Random Forest to classify neighborhoods based on crime rates and analyze their relationship to rental prices. Classify crime

- types, such as distinguishing between violent and property crimes and further categories like robbery, sexual offenses, etc., to study their impact on listing distribution and pricing.
2. Regression Models: apply models like Linear Regression and Ridge Regression to explore how the abundance of emergency food supply points might impact crime rates, and how this relationship in turn affects rental prices.
3. Clustering Models: utilize K-Means or to cluster Airbnb listings to identify potential patterns or hotspot areas of listing distribution. And cluster neighborhoods based on crime rates and Airbnb listing features to analyze which characteristics correlate with higher or lower rental prices.

User Interface

We will build a web application for user interaction using Bootstrap as frontend toolkit and Flask as backend framework. We choose Bootstrap 5.0 version to make our website more responsive and extensible, and Flask will help us build a web server which will be responsible for handling user requests and modeling.

- User interface: users will interact with our prediction model through the website we will deploy. Features to predict the result will be submitted through the form provided on our website.
- Data schema: request and response will be encoded and transmitted in JSON format.
- Backend API: RESTful backend APIs are to be provided to handle user requests and return the prediction result in response to features requested by users.



The sketch of our frontend website may look like:

The screenshot shows a web form titled "Let's do it" with a dark header bar containing "Demo project" and "Predict Model Group". The form has three input fields: "Zip code" with the value "98109", "Latitude" with the value "47.62337196100074", and "Longitude" with the value "-122.33645841668736". A blue "Submit" button is at the bottom.

Users will provide information like post code, latitude, longitude or other features (to be determined...) to get the prediction of house price from our trained model.