Boston Airbnb Open Data Analysis

Tian Gao, Xinyue Li

Introduction

Airbnb has successfully disrupted the traditional hospitality industry as more and more travelers decide to use Airbnb as their primary accommodation provider. Since its beginning in 2008, Airbnb has seen an enormous growth, with the number of rentals listed on its website growing exponentially each year. However, we have no idea about **how Airbnb is really used in Boston**. Therefore, we want to explore the public data set to help users get more information behind Airbnb.

Technique Selection

- **Programming Language:** Python 3.7
- **APIs:**
  1. Google Cloud Translation API
- **Algorithms & Methods:**
  1. TF-IDF
  2. Classifier
  3. Sentimental Analysis
- **Visualization Library:**
  1. Seaborn
  2. Wordcloud
  3. Folium

Our Work

1. Data Preparation

- Data Cleansing
- Feature Engineering

2. Exploratory Data Analysis

- Example: How does the number of listings changes from 2018 to 2019?

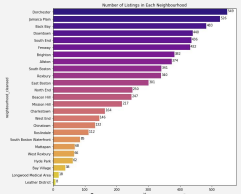


Figure 1 2018

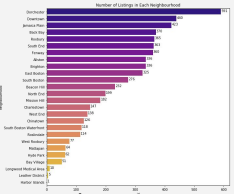


Figure 2 2019

3. Neighborhood Vibe Analysis

- Use "*neighborhood_overview*" provided by host in order to describe the vibe of each neighborhood
- Build a model with TF-IDF vectorizer and SGD classifier

4. Sentiment Analysis

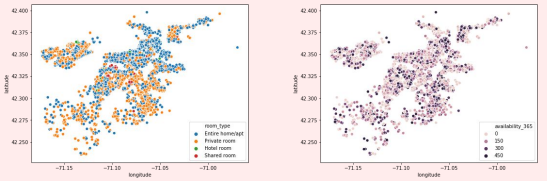
- Join existed data set (listings and reviews)
- Perform translation and generate a new data set
- Perform sentiment analysis to grade listings

5. Visualization

- Word cloud
- Interactive web

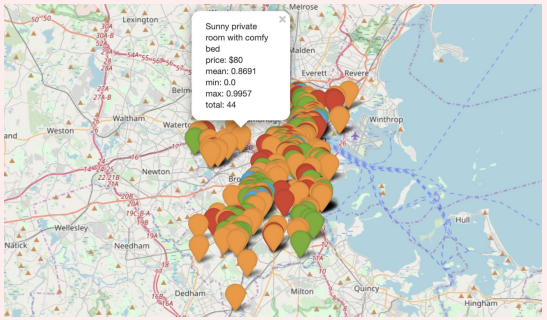
Result

- **Plot**



- **Word Cloud**

- **Interactive Web**



Github: https://github.com/feiyue33/cs506-final-project