Exploratory Analysis of Airbnb Data in Victoria, BC

STATS 744 Project

Chenxi Yu, Frank Cai, Yusang Yu

McMaster university

Table of contents

Introduction

Introduction

- The market of Airbnb expanded rapidly in recent years. It is a good idea to make an investment in properties in Victoria and then rent it out through Airbnb.
- Benefits such as free listing, safe payment method, free registration, host protection insurance attract people to invest in this new business.
- Victoria is the capital city of British Columbia and also a popular destination for travellers.
- we are going to do some analysis based on a dataset sourced from Inside Airbnb website which provides listings details in Victoria for 2018-09 to 2019-09.
- The main purpose of our project is to provide some information to investors who are thinking about to rent out their properties as a host on airbnb.

Questions we are trying to answer

Questions we are trying to answer

- What is the rental price for each neighbourhood region in Victoria?
- What is the rental price for different room types in each region?
- How the rental price changed though the whole year (2018.09-2019.09)?
- Is the price change different for different room type or neighbourhood?

Data Description

Data Description

Our data set is from the Inside Airbnb website. This data set contains 80 attributes. And listing information from 2018.09 to 2019.09.
In our project, we focus on following attributes:

- Neighbourhood_group_cleansed: which is a categorical variable, it indicates which neighbourhood region the listing belongs to.
- Room_type: which is a categorical variable and has 4 categories: private room, entire home, hotel room or shared room.
- Property_type: this variable shows the type of property for rent.
- Price: the rental price.
- Longitude latitude: which indicate the exact location of the property.
- Date: the last scraped date.

First Plot

First Plot

Goal: show the average daily price for different property type in each neighborhood.
Dataset Clean: combines 12 dataset into one and select five most commom property types.
Method:

- Package: gganimate, gglot2
- Bar plot: geom_col, position: dodge
- animate graph: library(gifski), transit_reveal()

Second Plot

Second Plot

Goal: showing monthly average price fluctuation for each room type for each neighbourhood.
Dataset Clean: removes some extremely high price, removes room type:'hotel room' which first appeared in August, 2019.
Packages: shiny, plotly.
Method

- inputs from users: "room type" and "neighborhood".
- line plot: x-axis: time, y-axis: average price, color: neighbourhood, line type: room type, easy to observe the trend of price fluctuation.
- bar plot: x-axis: neighbourhood, y-axis: log-price difference (%), position: 'dodge',color: room type.

Third Plot

Third Plot

Goal: show the distribution of different listings for different neighborhoods and the average price for different room type.
Dataset clean: select neighbourhood_group_cleansed, latitude, longitude, room_type, bathrooms, bedrooms, price. convert price from factor to numerical.
Packages: Plotly, Leaflet.
Method:

- interactive plot:displays listings according to the user input, a heat map show the distribution of rooms.
- calculated average price for different room type based on the input neighbourhood.
- calculate the average price for different room_type, I first filter the data by room_type and then compute price mean.
- x-axis: room type.
- y-axis: average price.

Future work

Future Work

The main drawback of our project is that we do not provide any information about the cost of the investment. If the information of the property price is provided, we can analyse monthly benefits based on setting of mortgage.


Thank
You!