IBM - Applied Data Science Capstone
Capstone Project – Battle of the Neighborhoods

Airbnb properties in the District of Porto, Portugal

1. Introduction

1.1 Background

Airbnb is an online-based company that connects people looking for accommodation (Airbnb guests) to people looking to rent their properties (Airbnb hosts) on a short-term or long-term basis. The rentals properties include apartments (dominant), homes, boats, and whole lot more. Renters are presented with a good selection of listings and can filter by criteria like price, number of bedrooms, room type, and more. Tourists are mostly motivated to book Airbnb accommodations because of their low cost, convenient location, and household amenities. They are generally less motivated by the opportunity to interact with the host or other locals, or by the promise of an authentic, local experience. For hosts, participating in Airbnb is a way to earn some income from their property, but with the risk that the guest might do damage to it. For guests, the advantage can be relatively inexpensive accommodations, but with the risk that the property won't be as appealing as the listing made it seem.

1.2 Business Problem

Porto is a city located in the north of Portugal by the outlet of the Douro River. It has elegant neighborhoods and large villas sitting on narrow cobbled streets. There is a lot of properties available in the Airbnb platform in the District of Porto. Exploratory analysis could help determine the best places (cities around, neighborhood) to stay and minimize the options of properties. This project aims to conduct some exploratory data analysis using the Foursquare API and the Airbnb list of properties to produce a small list of properties to be analyzed based on the client requirements.

1.1 Target Audience

I am planning to spend a month or two in the District of Porto (Portugal) and I am looking for a reasonable accommodation from Airbnb. There is a lot of properties (approximately 12,000) and analyze each one available will spend many time. So, the main objective of

this analysis is to produce a small list of properties to be analyzed in the Airbnb web platform, helping the property decision process and saving time.

This would interest anyone who wants to travel to the cities around Porto, Portugal.

2. Dataset

2.1 Data sources

The dataset used for this project comes from Inside Airbnb: http://insideairbnb.com/get-the-data.html.
The dataset that was employed was named *listings.csv*; it is a detailed data set with 106 attributes, a few of the attributes being: price per day (which will hereafter be simply referred to as price), number of beds, property type, neighborhood, cleaning fee, security deposit, host's ratings score, etc.

The data contains a total of 12,005. Each row in the data set is a listing available for rental in Airbnb's site. The columns describe different characteristics of each listing (features). The geographic dataset named *neighborhood.geoson* (available in Inside Airbnb) was used to create exploratory maps of the location. It contains the coordinates of each neighborhood group present in the *listings.csv* file.

I've used the Foursquare API to explore neighborhoods in the District of Porto. The Foursquare explore function will be used to get the most common venue categories in each neighborhood. The following information were retrieved for the 8 top venues from each neighborhood group:

- Venue ID
- Venue Name
- Coordinates
- Categories Name

2.2 Feature selection

Many of the features is not necessary for our analysis, so a selection of the principal features was realized. Out of 106 features, 26 features were selected. A few of the important numerical features are:

- *accommodates*: the number of guests the rental can accommodate
- *bedrooms*: number of bedrooms included in the rental
- *beds*: number of beds included in the rental
- *price*: nightly price for the rental
- *minimum_nights*: minimum number of nights a guest can stay for the rental
- *maximum_nights*: maximum number of nights a guest can stay for the rental
- *review_scores_rating: score of reviews that previous guests have left*

A few of the important categorical features are:

- *property_type*: house, townhouse, apartment, condo, hostel, cabin, etc.
- *room_type*: entire home/apt, private room or shared room
- *neighbourhood_cleansed*: neighborhood e.g. Midtown, Harlem, Murray Hill, etc.
- *cancellation_policy*: 6 categories: super_strict_60, super_strict_30, strict_14_with_grace_period, strict, moderate, and flexible.

3. Requirements

Considering the problem, a list of requirements was created to guide the analysis. The accommodation must be in a great location, considering the proximity to bakeries, restaurants and cafes.

Requirements:

- Entire home/apt
- Price by night bellow $100
- Properties in neighborhoods that presents a good offer of Café, Bakery and Restaurant
- The most common type of properties in the selected neighborhoods
- Only properties with review score rating equal 100
- Properties that accommodate 3 persons with 2 bedrooms