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.

1.2 Problem

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 (aproximatelly 12,000) and analyze each one will spend many time. Considering this, 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

This project aims to solve this problem, by conducting some exploratory data analysis using the Foursquare API and the Airbnb list of properties.

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.

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 are retrieved:

- 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.