1. Introduction

1.1 Background

I am a data scientist and I just received a job offer in São Paulo, Brazil. I am very excited for my new job to put my new skills into practice and continue to learn. How can I use my skills to find the best apartment for me to live in? I could certainly search for real estate companies manually, but I want to use what I have learned to help me find it. In order to do this, we need to define some basic rules:

• The apartment must have at least two bedrooms;

• It has to be at least 10km from work place (Rua Girassol, 555, Vila Madalena – São Paulo);

• It has to be near at least 1 km from a subway;

• The rental price cannot exceed R $ 5,000.00 per month;

• The neighborhood must have a market and gym nearby.

1.2 Problem

Data of rental properties such as price and number of bedrooms and location data of each neighborhood of the city. This project aims to find an apartment to rent in São Paulo, Brazil that fit the basic rules described above.

1.3 Interest

Everyone that is considering moving to São Paulo will be interested in find more about the neighborhoods and the economical features of apartments for rent.

2. Data acquisition and cleaning

2.1 Data required

List of subways in São Paulo and their location.

List of venues of each neighborhood in the properties data and their location.

List of apartments for rent and their features such as price, area, number of bedrooms, location.

2.1 Data sources

The data of venues of São Paulo will be collected from Foursquare API. The data of the apartments for rent will be scraped from the website of a major portal of real estate

2.2 How the data will be used to solve the problem

Use Foursquare data to extract the top 10 venues for each neighborhood in the rental properties dataset.

Use Foursquare to extract each subway located near the neighborhoods in the rental properties dataset.

Use geopy and Folium to create a map to show the venues, the subways and the rental properties around 1km.

Create clusters to determine the average rental price per square foot in each cluster in order to find the best cluster to search for properties.