

Coursera Capstone

IBM Applied Data Science Capstone

Investing in a property in Lisbon, Portugal



SARA NUNES

APRIL 2020

Introduction

Portugal has seen a huge increase in tourists in the past few years, especially in Lisbon, the biggest city in the country that has been competing with Porto as the number one destination in Portugal. Due to this growth in the tourism sector in Lisbon, more and more investors are now looking into investing in real estate in this city. Now they are looking to identify those neighborhoods that have the most interesting venues that makes them better for living and therefore a better investment and potential residency.

Business Problem

The objective of this project is to identify and outline the best neighborhoods in the city of Lisbon to invest in a property. Using data analysis tools and machine learning algorithms like clustering and data from the city it is possible to identify and group those neighborhoods into categories.

Data

The data needed to solve this problem is:

- List of neighborhoods in Lisbon. We get the list on https://en.wikipedia.org/wiki/Category:Parishes_of_Lisbon and create our neighborhoods data frame.
- Coordinates of the neighborhoods. These are obtained with `geopy` library based on district names and added to the data frame.
- Information about the venues and landmarks in the neighborhoods. The Foursquare API is used to collect this information.

Methodology

The goal of this project is to group the different neighborhoods in Lisbon by their characteristics such as venues and landmarks.

For this, it is necessary to analyze the most five most common venues in each district. After K-means clustering is used to put the districts that have similar most common venues together. After each cluster is analyzed to try to determine which one is more attractive.