IBM Data Science Certification

Capstone Report

Brianna Shannon

Dec. 2020

I. Introduction

Rome is the second most visited city in Europe after Paris for tourists. Because Rome has such a deep, rich history dating back to 735 BC, there's a lot to see and do. Ancient Rome was setup on the seven hills of Rome, which segmented much of Romans' activities and interactions, so many tourist attractions are also segmented into distinct areas. Because of this segmentation and the number of things to see, the goal of this assignment is to cluster neighborhoods by tourist attractions, food, and nightlife to help tourists maximize their visit. By doing so, these clusters will help tourists plan out which attractions to see on the same day and where to eat on the same day.

II. Data

The subdivisions of Rome will be pulled from the following link: https://www.timetomomo.com/en/destination/rome/. The subdivisions are not in a table but rather a picture, which is shown below in Figure 1. Using the list compiled from this image, the coordinates will be pulled from google maps. Using the location data for each of these subdivisions, the Foursquare API will be used to pull tourist attractions and food locations. The Foursquare data will be combined with the subdivision and K-means will be used to cluster it.

Figure 1: Subdivisions of Rome