IBM Data Science Capstone Project

1. Introduction & Business Problem

1.1 Introduction

The capital of Pakistan, Islamabad, considered as the second most beautiful capital of the world is home to some of the most favorite restaurants across the country. Every year millions of tourists both national and international visit here to enjoy the natural beauty of the city while feasting on the variety of foods it has to offer. But with many restaurants to choose from, it often fills the new tourists with overload of choices and choosing a bad restaurant can lead to a miserable experience for the tourists.

1.2 Problem

For a travel advisor, when making profile for a city, it is important to know which areas of the city provide the best food experience for dining out so that they can guide their customers accordingly. This is what this project aims to find out for Islamabad.

2. Data Acquisition & Cleaning

2.1 Data Acquisition

For this project, we will acquire the data of Sectors of Islamabad from Wikipedia here. Then we will use Foursquare API to extract all the venues within those sectors along with their location coordinates, venue categories, and venue IDs. We will use venue categories to filter our data to contain only restaurants in venue column and then use venue IDs to find the count of likes for each restaurant which will determine the popularity of that restaurant.

2.2 Data Cleaning

The data acquired from Wikipedia had few names of sectors that did not exist and hence were removed. Apart from that, the Foursquare API could not locate the coordinates of few sectors which was identified after looking at their returned addresses. Hence, they were removed as well. Finally, the data was plotted on a map using Folium library to confirm that the coordinates received from API were correct.

2.3 Feature Engineering

The count of likes for a restaurant obtained from foursquare API was used to calculate the total number of likes of all the restaurants in a sector and was named “Total Likes per Sector”. Furthermore, restaurants for each sector were counted to make a new feature “No. of Venues per Sector”.