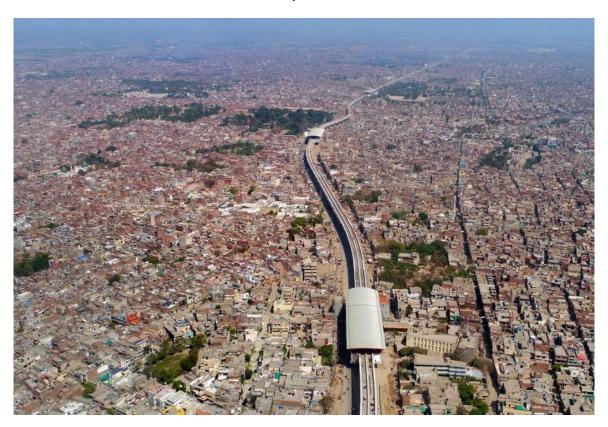# IBM Applied Data Science Capstone

# Opening Shopping Malls for tourists in Lahore

July 2020

# Introduction

## Business Problem

The purpose of this capstone project is to find the neighborhoods in Lahore, Pakistan where there are not enough shopping malls close to the hotels so the tourists don't have to move across taking taxi or buses to do shopping and utilize the time they save exploring the landmarks of the city for which it is gaining popularity day by day. This initiative can definitely please the tourists and the more tourists are pleased with the place the more people will visit in the future and this will eventually help the tourism industry to grow in Lahore. Moreover, if the venture gets successful in Lahore this model can be replicated to the other major cities in Pakistan.

## Target Audience

The project is being made for the tourism ministry of Punjab but it directly affects the tourists whether they have come from other cities of Pakistan or they travelled from other countries. For inter-city tourism the main targets are the people who travel to Lahore for business trips and they do not have a lot of time to travel far for shopping, the couples who stay at Lahore for a day or two before going to the northern areas of Pakistan to spend their Honeymoon, and the travellers who usually go for a complete Pakistan tour. The travellers from different countries come to visit to the traditional landmarks built during the Mughal Empire reign going back to the 15th century and they want to spend as much of their time, energy, and money to make that happen and shopping centers close to the places they are staying can easily let them shop, watch movies, eat, and have indoor gaming activities all at one place whenever they are free from their main purpose to visit.

# Data

To solve our business problem we will use a combination of two datasets:

1. Lahore suburbs (neighborhoods) dataset from Mapcrow
2. Venues dataset from Foursquare location API

## Lahore suburbs (neighborhoods) dataset from Mapcrow

Link: http://www.mapcrow.info/Lahore-PK-suburbs

The link contains the list of all the suburbs of Lahore District. There are a total of 159 unique suburbs in Lahore District but when we look at the Lahore City we limit the total suburbs to 80. First, BeautifulSoup and regex will be used to do web scraping to get the name and coordinates of all the neighborhoods in Lahore District. Python Geocoder Library will be used to get the coordinates of Lahore City. After that, Python Math Library will be used to calculate the distance between the center of Lahore City and all the suburbs. Using this method we will be able to filter all the neighborhoods in Lahore City.

## Venues dataset from Foursquare location API

Foursquare is one of the most used location API in the world and it stores data for millions of places throughout along with details of each place. In this project foursquare API will be used to get the list of venues and their categories in 3km radius of each neighborhood. From there onwards, the list of venues which fall under hotels or shopping malls will be used and clustered to further investigate our business problem.