

COURSERA CAPSTONE PROJECT REPORT

IBM APPLIED DATA SCIENCE CAPSTONE

Opening a new Café in Delhi, India

Prepared By: Harshit Singhal

Place: India

INTRODUCTION

We all are busy in work and hectic schedule of life. Taking out time to have some fun time enjoy the life we find places to go out have some great time. Café these days have become one of the most common place to hang out with friends and family. A **café** is a type of restaurant which typically serves coffee and tea, in addition to light refreshments such as baked goods or snacks. The term "café" comes from the French word meaning "coffee".

Nowadays café are basically theme based come up with some fantastic concepts. In the city like Delhi having a population size of near about 1.9 crores (as of 2012) it provide a great opportunity to open a new café in such great city. But the question that stuck is “where to open the open”, where the is full with some amazing café.

Business Problem

The objective of this capstone project is to analyse and select the best locations that are available in the Delhi, India to open a new café. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question:- If someone who is looking to open to open a café in Delhi, India, what would be some best location available?

Audience Interested in Project

All the builders, businessman, café chain owners or some great future business tycoon or some entrepreneurs would be interested in knowing the location in the city like Delhi, India which is densely populated and have a great scope to open café and flourish them in near future. Also all the people who are in food industry might also be interested in knowing location.

People would come to this project because many people have land and resources to have business but are not aware which business is best suited for that locality. So such people would be helped a lot.

DATA

Data Description

In order to solve the problem, we will be needing the following data:

- List of neighbourhoods in Delhi. This defines the scope of this project which is confined to the city of Delhi, the capital city of the country India in Asia.
- Latitude and longitude coordinates of those neighbourhoods. This is required in order to plot the map and also to explore and get the venue data of the neighbourhoods.
- Venue data such as venue category or type and number of all such categories, particularly data related to café. We will use this data to perform clustering on the neighbourhoods.

Data Source

We will be taking the neighbourhoods of Delhi data available on the Wikipedia.

Link: https://en.wikipedia.org/wiki/Neighbourhoods_of_Delhi

We will be using the requests library of python in order to scrap the data and beautifulsoap to perform the parsing. For the sake of geographical coordinates of the neighbourhoods we will be using the geopy library.

For the the purpose of exploring the neighbourhoods we will be using the foursquare API. Foursquare has one of the largest database of 105+ million places and is used by over 125,000 developers.