A study in Pune City's Population and food-services data.

Coursera IBM Data Science Capstone

Ishan Daga

Introduction

Pune is a fast-growing and upcoming metropolitan city in Maharashtra, India. It is well known as the cultural capital of the state and the de-facto king of the region's Street food.

The sprawling and ever-growing population and city-scape are home to some of the finest joints in the Maharashtra and the number of places to choose from in every locality of the city increases year-on-year.

Owning a joint in the city is a sure shot form of a stable income, even if it were only a food-truck, the people of the city take to fresh ideas quite well and the street-food culture has only added to itself over the years.

This is a serious business and requires careful planning and consideration on the part of the investor/stakeholder to ensure maximum profits from the business.

Business Problem

If a business would like to open a restaurant/brewery/pub/fast food joint in the city, what would be the optimal place to open said service in order to receive maximum revenue/foot-fall?

Target Audience

Future & prospective Food service/restaurant owners and stake holders can get a clear picture of the city's habits and the population density in order to open a new place in a suitable locality.

Data

Data used

To solve this problem, the following data has been leveraged:

- Population Data by Wards (Neighborhoods)
- Ward Offices Map Vector Data (GeoJSON and Shape Vectors)
- Location-wise Venue data for each sub-locality in Pune
- Latitude and Longitude Location Data of Wards and Venues

Acquiring the data

The Population Data was pulled from the Pune Municipal Corporation's opendata webpage

http://opendata.punecorporation.org/Citizen/CitizenDatasets/Index

The Wards and their corresponding ward offices with the populations per ward are listed out in a PDF from the census of 2011 [most recent]. This PDF was then converted to a csv file via the free online tool: PDF to Tables
https://pdftables.com/blog/convert-pdf-to-csv

The resulting CSV file was then manually cleaned up and double-checked in Microsoft Excel, to ease the pre-processing in python. The final CSV file had only the Ward numbers, Ward Names, Ward Offices and Population Data.

The Location data for the Wards was pulled using the Google Maps Geocoding API sequentially and added to the dataframe.

The venue data was procured using the FourSquare API which is simpler to use and has a free option, both of which are not available with the more expansive Google Maps API. Foursquare however does boast more community contributions and is used by 100,000+ developers world-wide.