**Data Requirements & Source**

Outlined below are the details of the data, which will be required to carry out this analysis

Source has been identified. Basically, data has to be downloaded from Government sites

data.nsw.gov and abs.gov.au. Refinement on the exact URL will be done in next implementation phase after viewing and analysing each data set

**Suburbs in Sydney Areas**

Source: <https://data.nsw.gov.au/data/dataset/spatial-services-nsw-suburb>

Final data should have attributes of **Local Government** and **Suburbs** in Sydney Area

**Latitude/ Longitude for each suburb**

Use geopy.geolocator libraries in Python to get latitude and longitude values of each suburb

**Indian/ Asian Restaurants in each suburb**

User latitude/ longitude of each suburb and Foursquare API’s search function to get Indian/ Asian restaurants in each suburb

https://api.foursquare.com/v2/venues/**search**?

**NSW Crime Data**

**Source: https://data.nsw.gov.au/data/dataset/8a39cc2d-3c5a-4274-97de-a7f8ce5b24ee**

**Population Density**

Regional Population by Age and Sex, Australia, 2017

Source:

[**http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/3235.02017?OpenDocument**](http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/3235.02017?OpenDocument)

**Parking Data**

Aim is to get data on number of public parking spaces in each Sydney suburb

**Source :** [**https://data.gov.au/dataset?tags=parking**](https://data.gov.au/dataset?tags=parking)

Explicit data set is not available currently for Sydney Area. Though data is available on parking fines. Metrics on parking can be interpolated using data on parking fines. It can be assumed that suburbs with less fines have abundance parking spaces

**House Hold Income**

Source : <http://abs.gov.au/household-income>

This will get average household income suburb wise in Sydney location

**Solution Design**

Get all required data from NSW Govt / Australian Bureau of Statistics sites

Where ever appropriate data cannot be found interpolate on base available data. Make assumptions and document these assumptions in final report

Use Four Square API’s to get Indian/ Asian Restaurant Venues

Create a panda data frame with following features for each suburb in Sydney Area

1. Number of Indian Restaurants
2. Number of Asian Restaurants
3. No crime incidents
4. Population Density
5. Average Family Income
6. Parking Spaces

Use clustering modelling approach and K Means Algorithm to form clusters

Iteratively analyse each cluster and adjust k to get optimum k

Analyse each cluster and identify clusters with following characteristics

1 No Indian Restaurants / 1 Indian Restaurants in high density area

2 Two/ Three Asian Restaurants

4 Affluent Suburbs

5 Less crime rate

6 Ample Parking space available

Assign an appropriate label to each cluster

Select best Cluster and Prepare a report for stakeholders and recommend suburbs suitable for opening a new restaurant in Sydney Area