**Capstone – Neighbourhoods (Seattle, Washington)**

1. **Project Description:**

The objective of this capstone is to collects various data points from Seattle, Washington as the requirements and main idea to help the stake holders to achieve the desired results, so as to help them spend their valuable time and money in some other productive means rather keep them trapped in an infinite loop of extensive search engines unnecessarily.

This project would help them to take the decision on choosing the best neighbourhood from out of many neighbourhoods. Like: build/buy their houses in the city based on the distribution of various facilities.

We’re using K-mean clustering unsupervised machine learning algorithm to cluster the venues based on the place category

* Restaurants
* Park
* Coffee Shop
* Gym etc.

This would give a better understanding of the similarities and dissimilarities between the two chosen neighbourhoods to retrieve more insights and to conclude with ease which neighbourhood wins over other.

Here we’re going to compare two neighbourhoods to choose one that best matches based on the below factors

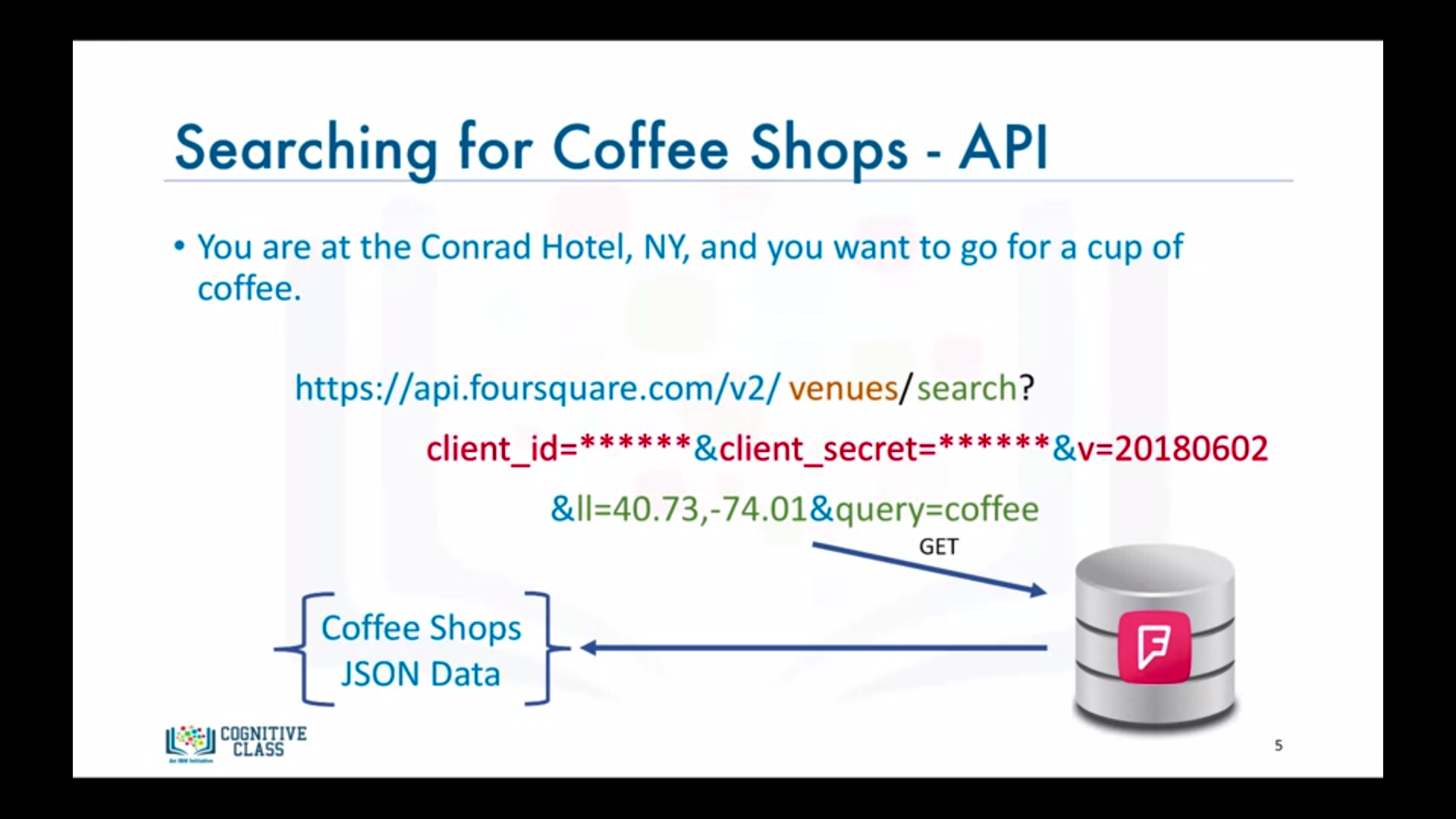
* Housing Price
* Population
* School Rating
* Comfortable Neighbourhoods

1. **Foursquare Data Sets and APIs:**

Foursquare API: We’re going to use in project Four-square API as, prime data gathering source as it has a huge database.

* Its API provides the ability to perform location search, location sharing and details about a Business. Photos, Restaurants, Coffee Shop and etc.,

1. **Work Flow:**

****

1. **Components used in this Analysis**

**Foursquare API:** Search feature would be enabled to collect the nearby places of the neighbourhoods.

**Folium-Python Visualization Library**: would be used to visualize the neighbourhoods cluster distribution of Seattle city over an interactive leaflet map.

**Pandas, NumPy and Scikit-learn:** Python’s scientific libraries are used to drive the desirable insights in the data

**K-mean clustering:** K-Mean machine learning algorithm would be applied to form the clusters of different categories of places residing in and around the neighbourhoods.

These clusters from each of those two chosen neighbourhoods would be analysed individually collectively and comparatively to derive the conclusions.

1. **Key Packages and Dependencies for the implementation:**

* Geopy – To retrieve Location Data
* Requests – Library to handle http requests
* JSON – Library to handle JSON files
* Folium – Map rendering Library
* Pandas
  + Library for Data Analysis
  + NumPy – Library to handle data in a visualization
  + Matplotlib – Python Plotting Module
  + Sklearn – Python machine learning Library

1. **Comparison analysis between neighbourhoods**

|  |  |
| --- | --- |
| **Population Comparison(Indian)** | |
| Bellevue | 157 |
| Kirkland | 366 |
|  | |
| School Rating Comparison | |
| Bellevue | 9 |
| Kirkland | 9 |
|  | |
| Avg Housing Price Comparison | |
| Bellevue | 343500.0 |
| Kirkland | 297900.0 |
|  | |

1. **Conclusion:**

Over all analysis to compare to Bellevue, Kirkland has the higher number of Indian population, good school rating of 9 and a reasonable Avg housing price of around 300k and other reports. So Kirkland is perfect neighbourhoods over Bellevue for Seattle.