

Applied Data Science Capstone project - Dubai Neighborhood Analysis

By: Giridhar Reddy Kolan

II: DATA SOURCES

Based on our business problem statement we will need following parameters data:

1. **Tourists footfall:** What are top destinations of tourists in Dubai city with locations
2. **Population density:** Dubai neighborhoods and population counts / Densities
3. **Number of Venues** in the neighborhood to understand the commercial traffic
4. **Existing restaurants** and their types around each neighborhood
5. **Availability competition** (existing Japanese restaurants) in the neighborhoods

We can collect the first two datasets from the following sources:

Top Tourist Destinations in Dubai City:

- <https://www.globalmediainsight.com/blog/dubai-tourism-statistics/>
- <https://www.planetware.com/tourist-attractions-/dubai-uae-dub-dubai.htm>



Number of Estimated Population Dubai City.

- <https://www.dsc.gov.ae/en-us/Themes/Pages/Population-and-Vital-Statistics.aspx?Theme=42>

The screenshot shows the Dubai Statistics Center website. The header includes the Government of Dubai logo, the year of tolerance (2020), and the Dubai Statistics Center logo. The main navigation bar includes links to Home, About Us, Methodologies, Statistical Dashboards, Publications, Infographs, and Contact Us. The page is titled "Themes > Population and Vital Statistics".

On the left, there is a list of themes. The "Population and Vital Statistics" theme is selected and highlighted in red. The main content area for this theme contains a description of the data, a list of reports, and a table of publications.

Reports:

- Population by Gender (2018, 67 KB)
- Population by Gender (2018, 60 KB)
- Number of Estimated Population by Sector and Community (2018, 470 KB)

Remaining data sets from **3 to 5** can be collected / extracted and generated from **Foursquare API**.

The screenshot shows the Foursquare Developers website. The header includes the Foursquare logo, the word "DEVELOPERS", and links to Products, Docs, and Log-in. The main content area is titled "Places API" and describes the API's capabilities.

Key Features:

Feature	Description
Access to Foursquare's Global Database	Get real-time access to over 105MM places available across 190 countries and 50 territories.
Power App Experiences	Use our custom API endpoints to power geo-tagging, venue search, venue recommendations, and more in your app
Descriptive Place Profiles	Leverage 70+ venue attributes and 900+ categories, sourced by the Foursquare consumer community.
Rich User Content	Create engaging location experiences with access to user-generated tips, tastes, photos & more.

Data Exploration and Analysis:

Above datasets will be collected, cleaned and transformed into suitable format for further data exploration analysis. The sites with high tourist's footfall (top tourist destinations) will be mapped by collecting their latitude and longitude values. Once we have the locations of these sites, we will spatially check them against **population density** and **number of venues** data to understand population density and commercial activity traffic around each site.

If any destinations are falling in very low population density with very low commercial activity traffic (venue counts), then those destinations will be omitted from further analysis as they do not represent prospective sites for new restaurant. The final shortlisted sites will be used in machine learning algorithms such as K-means clustering to understand the overall restaurants distribution and in particular the Japanese restaurants.

The top tourist's destination sites with high population density, high commercial activity traffic and with low Japanese restaurants distributions will be selected as best location to open a Japanese restaurant in Dubai.