Setting up a Bubble Tea Shop in a Singapore Shopping Mall

IBM Data Science Capstone

BY: CHENG SENG TAN

✓A Taiwanese bubble tea brand owner wants to evaluate opportunity to setup a bubble tea shop in a Singapore shopping mall.

✓Is the market saturated in Singapore?

✓Where would be a suitable location to setup shop?

✓This study explores all the districts in Singapore in terms of the Bubble Tea Shops and Shopping Mall.

✓Cluster analysis is applied to group similar district together.

✓Population density is used to select districts within the favourable cluster.

Bubble tea, which started in Taiwan, has won fans worldwide, with an industry that is worth $60 billion.

DATA

Web Scrapping of District Information from Wikipedia page (https://en.wikipedia.org/wiki/Planning_Areas_of_Singapore)

Latitudes and longitudes of these boroughs was obtained using Python Geocoder package.

Venue data for each district gathered using Foursquare API.

Population density obtained from Wikipedia page

Web scraping Wikipedia page for list of district in Singapore

Get latitude and longitude coordinates using Geocoder

Use Foursquare API to get venue data

Group data by district taking the mean of frequency of occurrence of each venue category

Filter venue category by Bubble Tea Shop & Shopping Mall
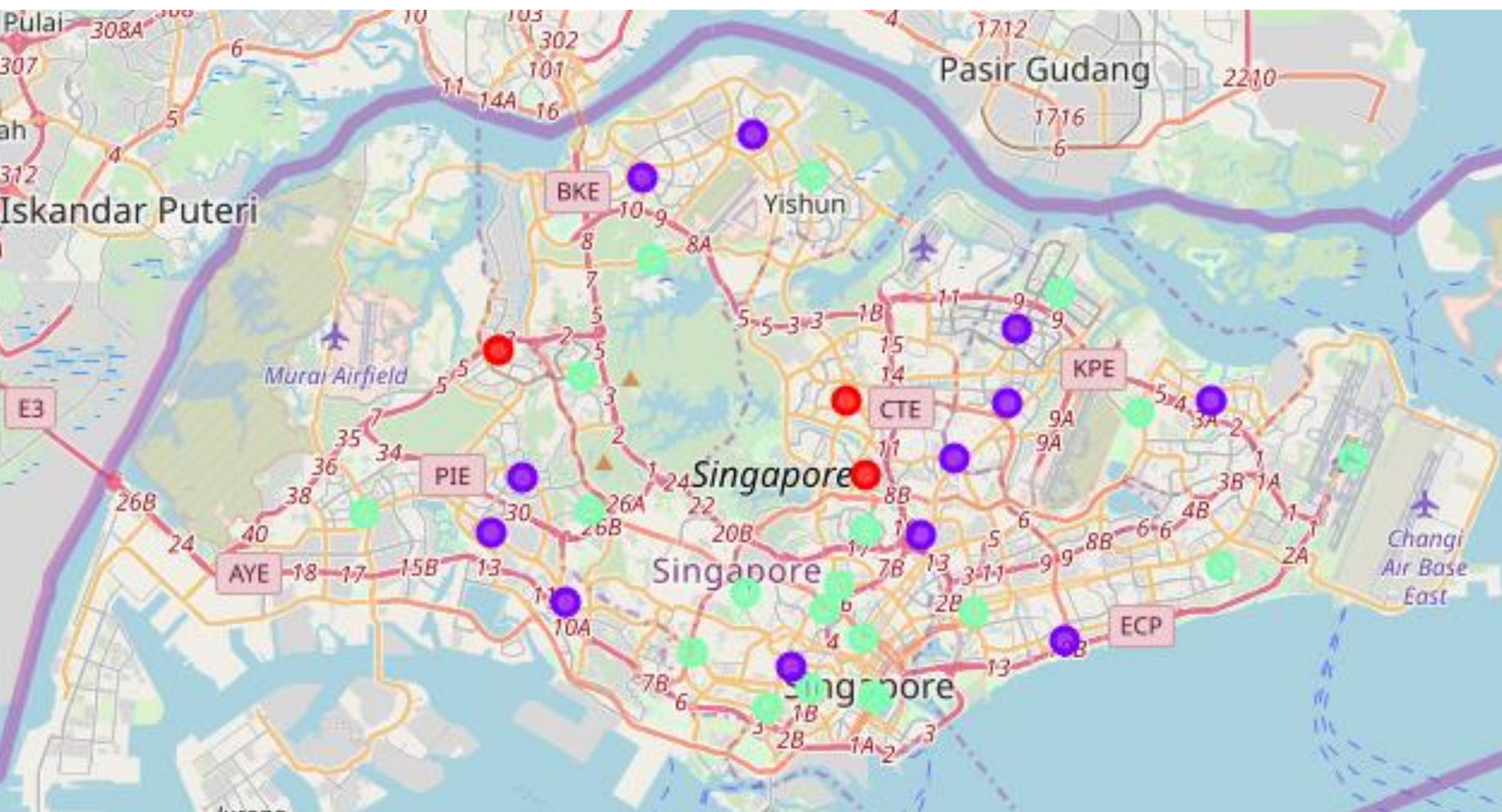
Perform clustering on the data by using K-means clustering

Visualize the clusters in a map usingFolium
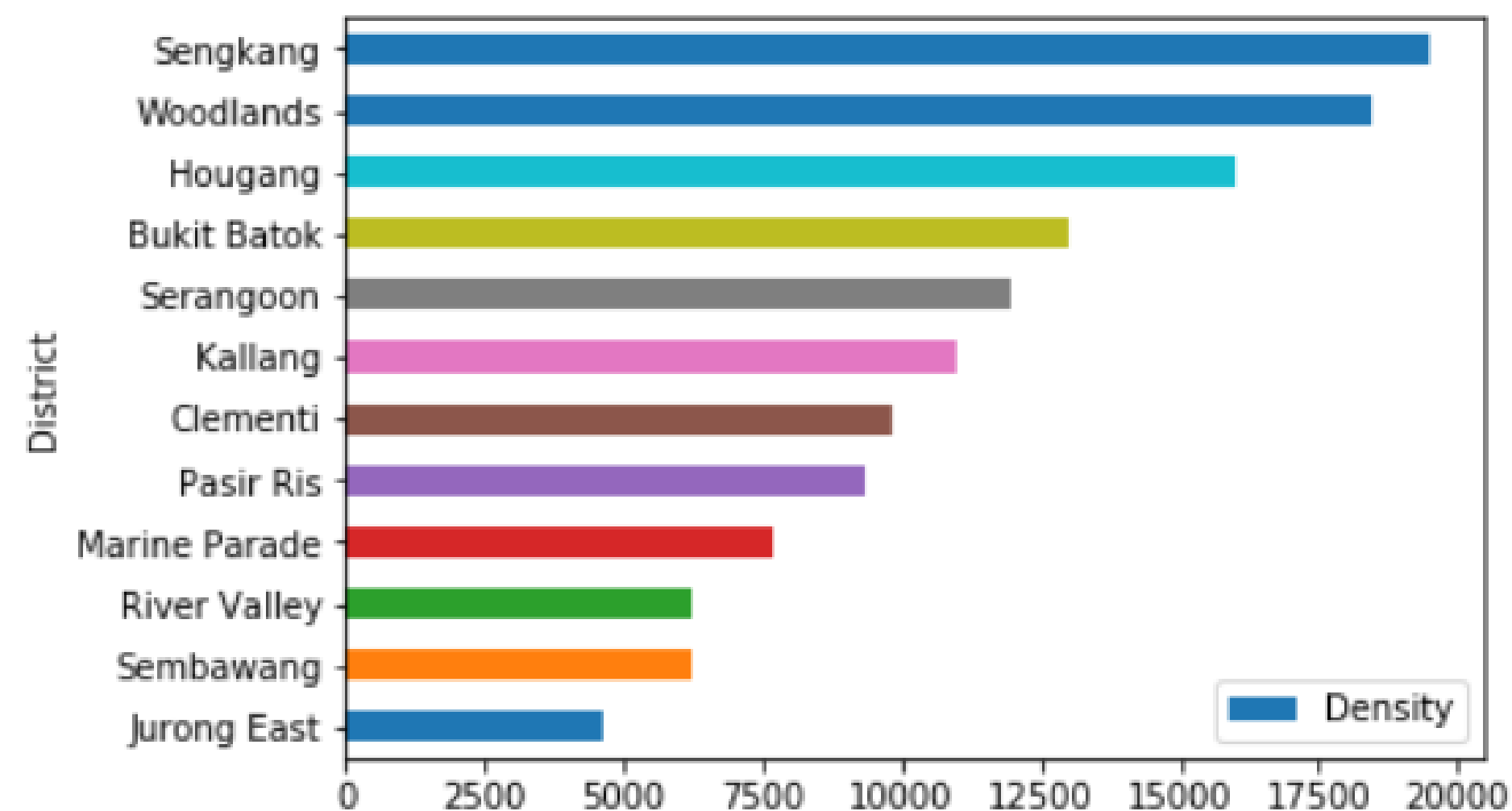
Visualize population density using matlplotlib bar chart

Cluster 1 🔴 **shows districts with high penetration rate for bubble tea shops and presence of shopping malls.**

🏆**Cluster 2** 🟣 **contains districts with presence of shopping malls and almost no (except for Jurong East) presence of bubble tea shops.**

Cluster 3 🟢 **has the districts that have no or low presence of bubble tea shops. These districts are also characterised by no or few shopping malls.**

Sengkang, Woodlands & Hougang have the highest population density in preferred **Cluster 2**

✓ **Cluster 1 has the highest saturation of bubble tea shops housed within shopping malls. Intense competition will be expected in these districts.**

✓ **Cluster 2 & 3 has low presence of bubble tea shop in the districts. The difference in the profiles is that Cluster 2 has the shopping malls whereas Cluster 3 has no or low presence of shopping malls. The client's preference is to setup a bubble tea shop in a high-traffic shopping mall so Cluster 3 will not be suitable.**

Districts in Cluster 2 will be the most suitable to setup a bubble tea shop within an existing shopping mall.
Sengkang, Woodlands or Hougang has the highest population density, being away from the Central District will be affordable in rental

CONCLUSION

The Business Question is Answered.
- The districts of Sengkang, Woodlands & Hougang in Cluster 2 are preferred location to setup a Bubble Tea Shop due to low or no competition.
- Findings based on data evidence

Data science techniques used in arriving at conclusion:.
- Scraping of data from the internet
- Data Wrangling to prepare data
- Using Foursquare API to gather venue information
- Mapping using Folium
- Applying K-means clustering to group similar districts Plotting bar chart to visualise population density

THANK YOU

HAVE A NICE DAY EVERYONE!