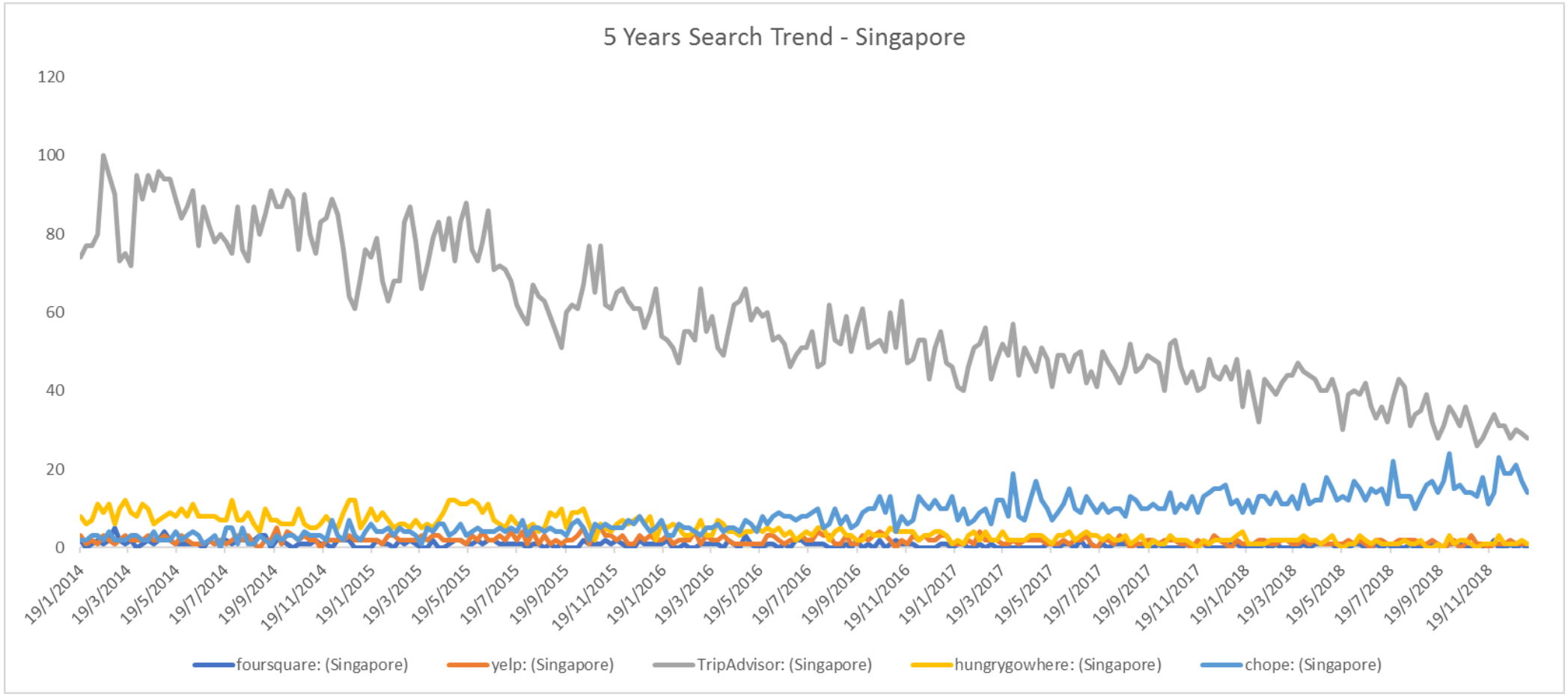
Using Foursquare Results – Relevancy in Local Context?

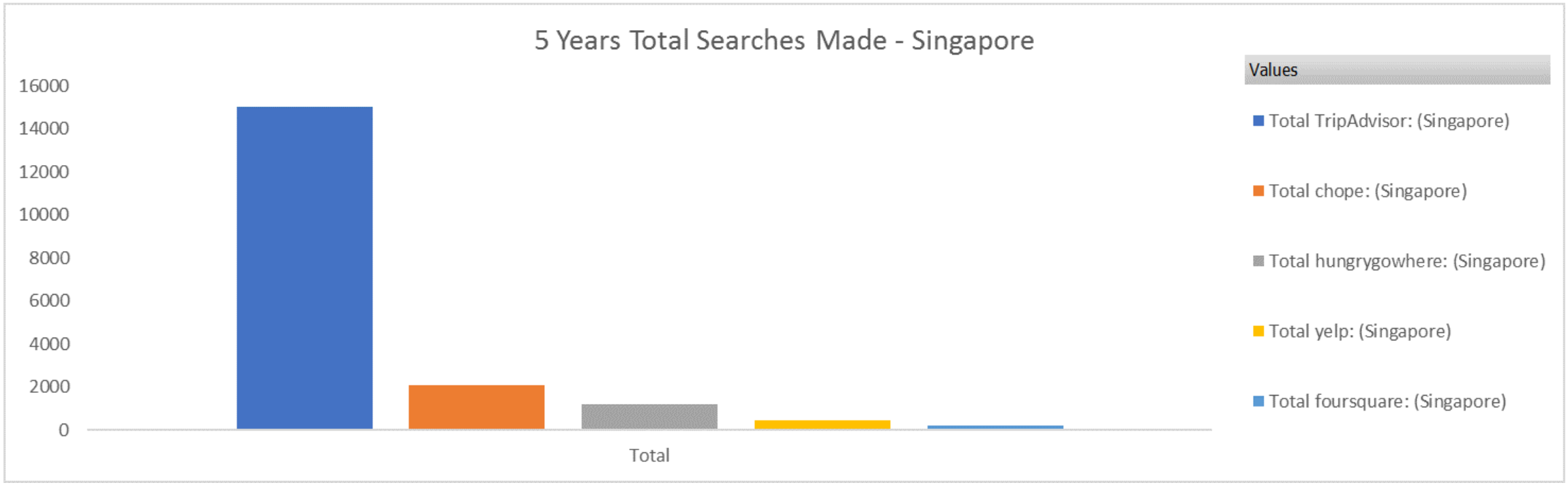
IBM Capstone Project

Jan 2019

Search Trend for Past 5 Years - Singapore



Total Searches Made for Past 5 Years - Singapore



Relevancy of Foursquare Results in Local Context

- Given the low search results, are Foursquare results relevant for local context?
- Does low search results means lower usage?
- A location intelligence service will provide better results if there are more users on the platform.

Relevancy of Foursquare Results in Local Context

- Will attempt to answer the relevancy question by obtaining answers for:
  - If the neighbourhood is a known tourist spot, does it return more results?
  - Do the results reflect the nature or characteristics of the neighbourhood?

Methodology – 1/2

- Select 5 local neighbourhoods, of which, some are known tourist spots
- Google search their coordinates and update in Excel spreadsheet
- Include 5 random postal codes from Toronto geospatial file
- Prepare analysis using Python, a programming language, in Juypter Notebook.
- Read Excel file and perform pre-processing work (e.g. remove unnecessary columns and rows, converting to a Panda DateFrame).
- Load local map into Juypter Notebook and includes the 5 neighbourhoods.

Methodology – 2/2

- Randomly select 1 neighbourhood and obtain returned results from Foursquare to have a feel of the returned dataset.
- As the returned dataset is satisfactory (i.e. no abnormalities spotted), extend to all neighbourhoods.
- Group the venues by neighbourhood to determine the number of venues per neighbourhood.
- Perform one-hot encoding to facilitate computation of the frequency of unique categories.
- Rank in descending order for the 15 most frequent venues for each neighbourhood.
- Using Clustering, an unsupervised machine learning algorithm, group the processed dataset into 4 clusters.

Results & Discussion – 1/3

Number of Venues per Neighbourhood

Neighbourhood	Number of Venues
Marina Bay	100
Sentosa	96
Orchard	84
Jurong East	80
Yishun	57

If the neighbourhood is a known tourist spot, does it return more results?

- From the table in the right, it appears that neighbourhoods which are known tourist spots (Marina Bay, Sentosa and Orchard), returned more results

As the free plan is used, there are some constraints on the returned results (limit of 100). But the table showed tourist spots generally have higher results returned.

Results & Discussion – 2/3

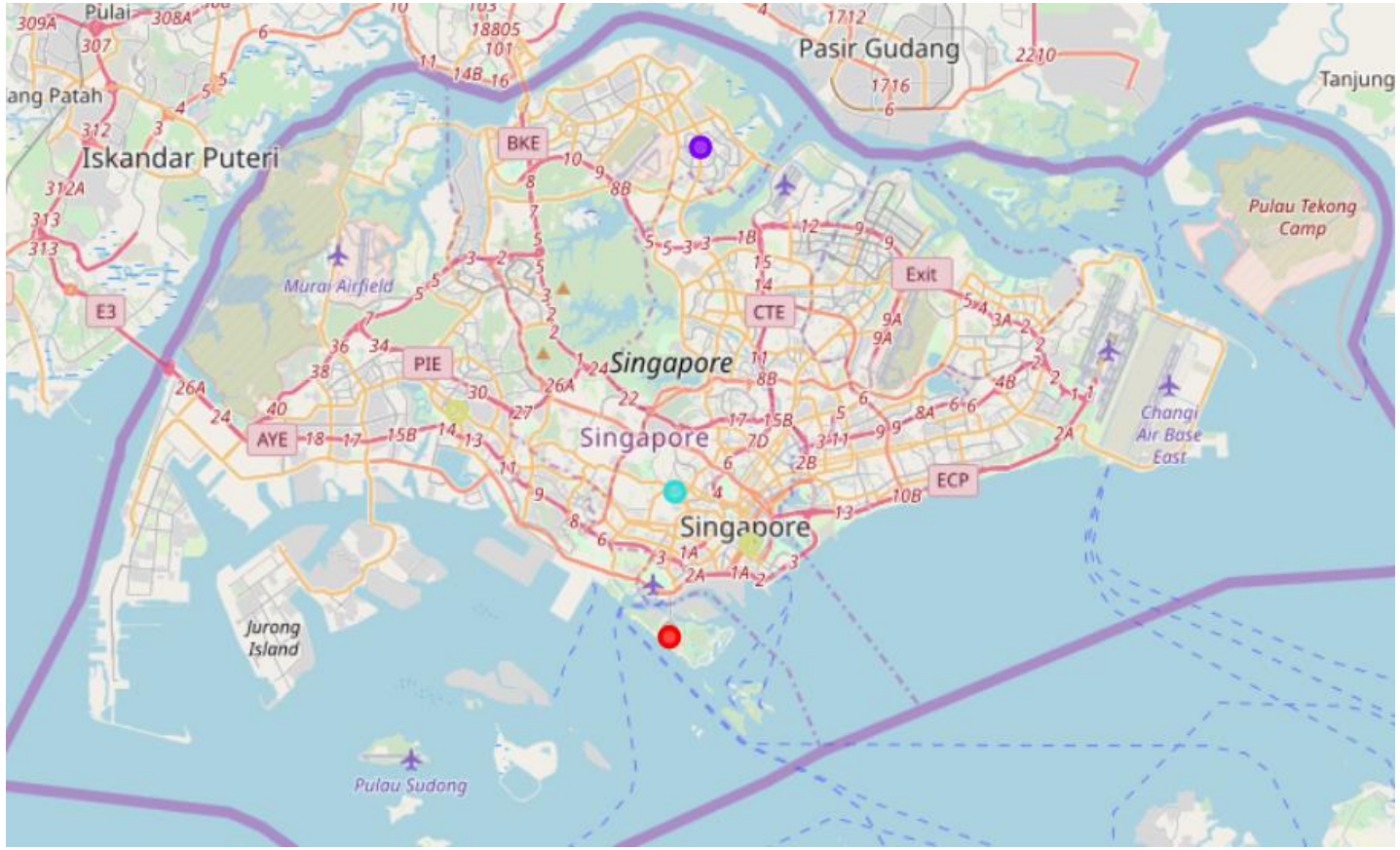
Top 10 Venues per Neighbourhood

	Neighbourhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Jurong East	Japanese Restaurant	Shopping Mall	Chinese Restaurant	Food Court	Café	Department Store	Sandwich Place	Bubble Tea Shop	Bakery	Clothing Store
1	Marina Bay	Waterfront	Japanese Restaurant	Hotel	Salad Place	Cocktail Bar	Italian Restaurant	Sandwich Place	Gym / Fitness Center	Gym	Coffee Shop
2	Orchard	Hotel	Chinese Restaurant	Japanese Restaurant	Café	Boutique	Supermarket	Miscellaneous Shop	Cosmetics Shop	Coffee Shop	Cocktail Bar
3	Sentosa	Theme Park Ride / Attraction	Theme Park	Fast Food Restaurant	Café	Restaurant	Hotel	Beach	Food Court	Italian Restaurant	Resort
4	Yishun	Coffee Shop	Fast Food Restaurant	Indian Restaurant	Food Court	Grocery Store	Café	Supermarket	Shopping Mall	Chinese Restaurant	Ramen Restaurant

The results show the nature / characteristics of the neighbourhood. For example, Marina Bay has a beautiful Waterfront while Sentosa has a theme park.

Results & Discussion – 3/3

Group Neighbourhoods into 4 Clusters



	Neighbourhood	Country	Latitude	Longitude	Cluster Labels
0	Marina Bay	Singapore	1.2840	103.8535	3
1	Sentosa	Singapore	1.2494	103.8238	0
2	Jurong East	Singapore	1.3329	103.7436	3
3	Yishun	Singapore	1.4304	103.8354	1
4	Orchard	Singapore	1.3030	103.8258	2

Jurong East and Marina Bay are grouped as one cluster when we use K-mean clustering with K=4 clusters.

Conclusion

Though Foursquare is the lowest searched term, the results returned from the API search call are fairly consistent with the visitors' makeup (tourist / local) of the neighbourhood and highlighted its nature / characteristics.

It can be gathered that Foursquare results are fairly relevant for local context, with an inclination towards a tourist visitor base (i.e. higher returned results). Whether there are better alternatives (other than Foursquare) will have to be further explored.

Note: the free plan is used for this analysis, which placed some constraints on the results returned but this does not significantly skewed the observations.

Next Step

Some possibilities:

1) Subscribe to the Premium plan offered by Foursquare to validate the results
2) As Yelp also provides API calls, we could make use of its services to do a further analysis
3) Check with local location intelligence services / apps for access to database to conduct same analysis

The python codes used for this analysis could be found in this link:
https://github.com/fongwc/Foursquare-Results-Relevancy-in-Local-Context-/blob/master/Foursquare%20Results%20-%20Relevancy%20in%20Local%20Context%20(Submission).ipynb