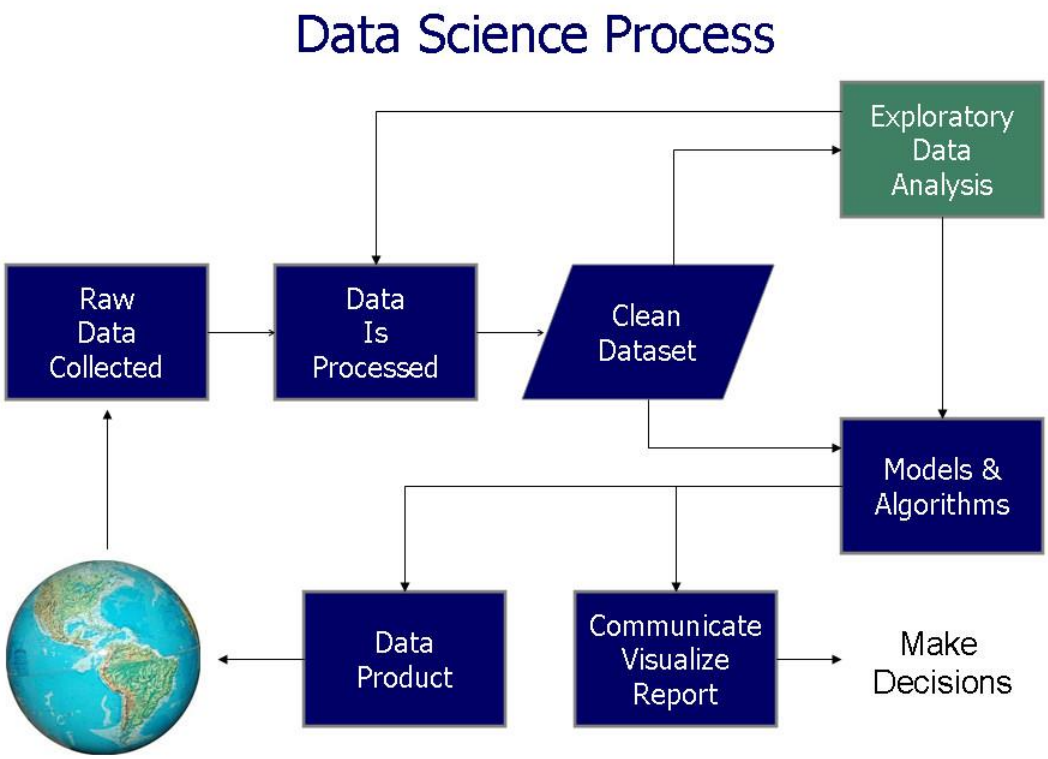


Using Foursquare Results – Relevancy in Local Context?

CAPSTONE PROJECT

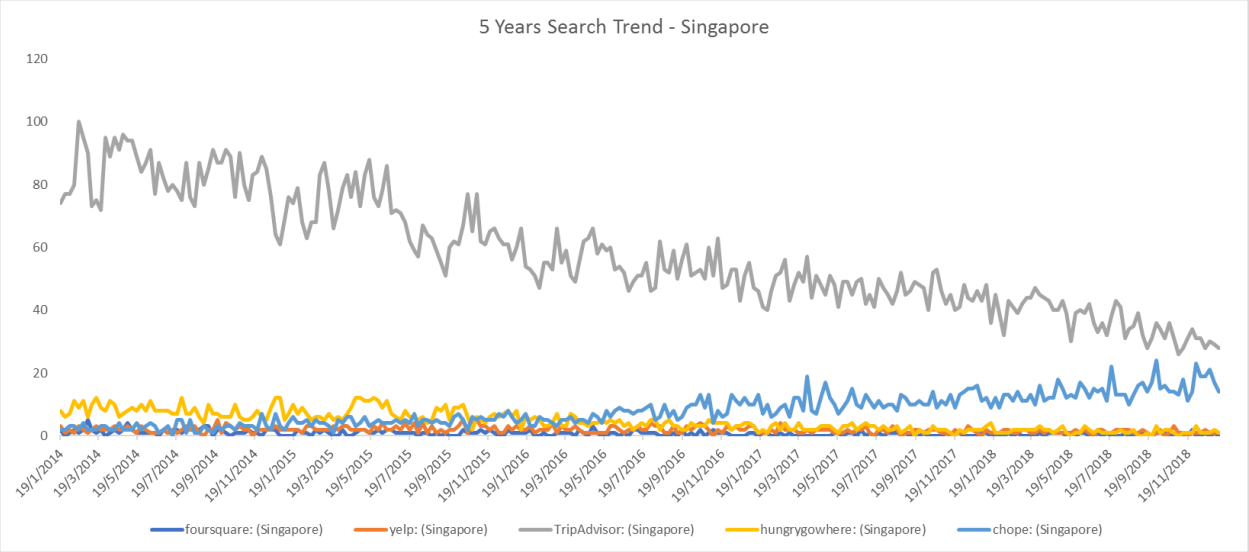
IBM DATA SCIENCE | Jan 2019

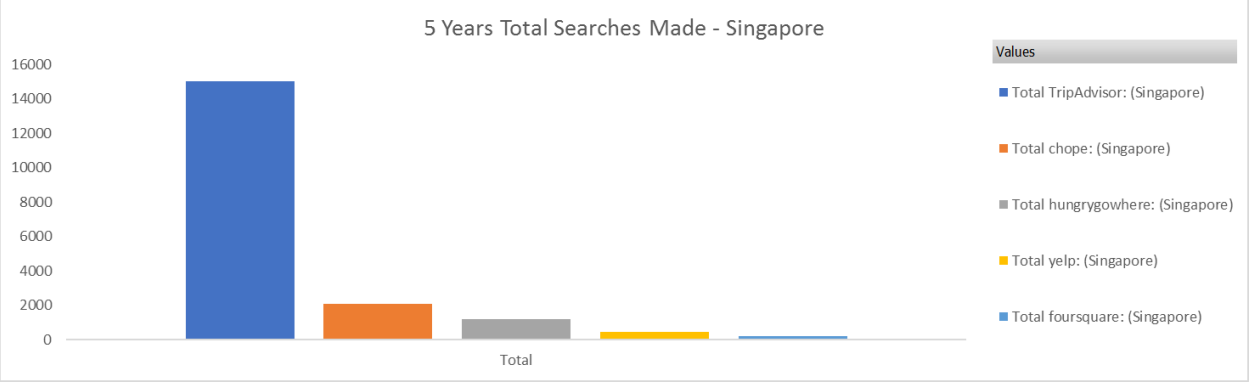
Introduction

As stated in its website, Foursquare is a technology company that uses location intelligence to build meaningful consumer experience and business solutions. Foursquare also provides developers free and premium access to its huge database to develop innovative business solutions or to check out nearby venues for making personal (e.g. where to dine) or business (e.g. where to set up shop) decisions.

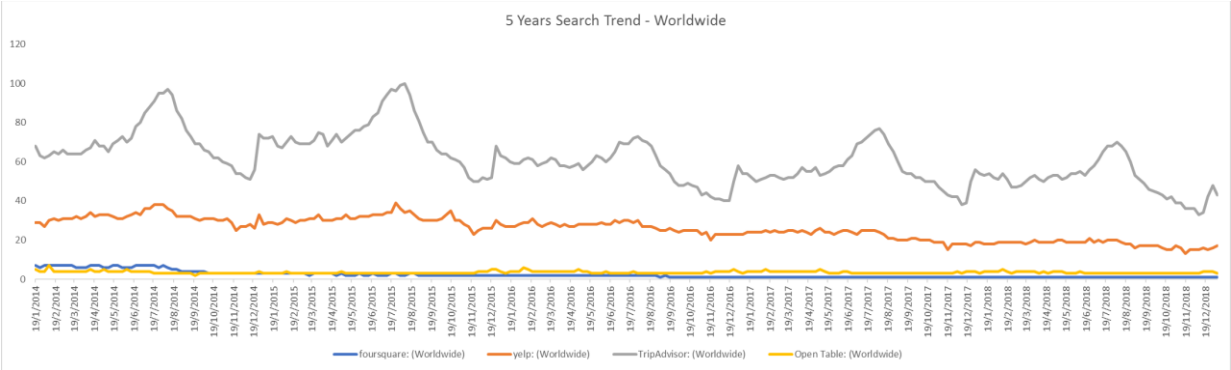
Other than Foursquare, there are also other international (e.g. Yelp, Trip Advisor) or local (e.g. Hungry Go Where, Chope) companies or apps which provide location-based advice / suggestion.

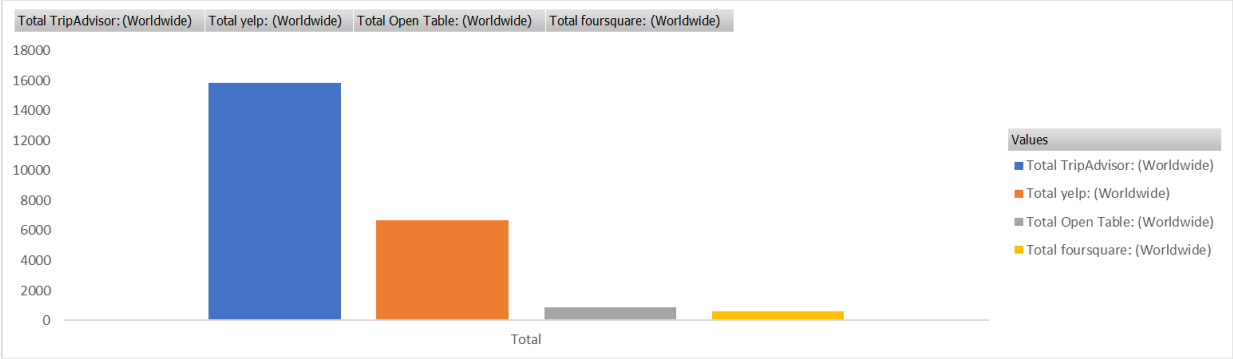
Performing a Google trend on the above search terms for the past 5 years, it appears that Foursquare is the least searched term in local context, while Trip Advisor is a highly searched term but on a downward trend. Local location-based app, Chope, has seen a rise in the number of searches over the past 5 years.





Given this background, we performed another Google trend analysis for the search terms: Foursquare, Yelp, Trip Advisor and Open Table on a worldwide basis.





The information from the worldwide trend analysis collaborated with that of the local analysis and Foursquare is the least searched term amongst the selected or targeted terms. Though this does not necessarily mean lowest usage nor market penetration, but it does give a hint to its popularity or market awareness. If it is not a frequently searched term, could this have an impact on usage and thus affecting the location-based intelligence it provides since the algorithm works best when there are many users providing feedback, comments, etc. on the platform? Can we still place relevance on the advice / suggestion it provides given there are other local-based apps or services available?

With the above as the back drop, we will attempt to explore:

- Does Foursquare return more venues in neighbourhoods where there is a higher portion of tourists compared to locals on the basis that locals are more familiar with homegrown apps?

- Does the results reflective of the nature or characteristics of the neighbourhood?

DATA

Based on domain / business knowledge, 5 neighbourhoods have been selected for analysis on the relevancy of the returned results provided by Foursquare:

1. Marina Bay: waterfront / Marina Bay Sands, mixture of tourists and locals
2. Sentosa: tourist attraction / themed park, mixture of tourists and locals
3. Jurong East: transport interchange / shopping mall, more locals
4. Yishun: local neighbourhood, predominantly locals
5. Orchard: shopping belt, good mixture of tourists and locals

After selecting these 5 neighbourhoods, performed a google search to obtain the latitudes and longitudes of these locations and transcribed them into an Excel spreadsheet. Also, randomly select 5 postal codes and coordinates from the geospatial_data.csv file for Toronto, Canada and include in the same spreadsheet. The inclusion of the postal codes is to demonstrate that the data received is not always what we desired, and much time is spent on cleaning and getting the data ready (pre-processing work).

We will use the Foursquare API (Application Programming Interface) to extract nearby venues for the 5 neighbourhoods with a radius of 800 meters and limit on returned results of 1,000 through python programming. The returned results will show the name of establishment, the category which the establishment is in (e.g. hotel, Chinese restaurant, Korean restaurant, etc.), its latitude and longitude. Thereafter, the number of establishments / venues will be grouped by neighbourhood to provide an indication whether those neighbourhoods that a good tourist base has higher returned results.

We will then perform one-hot encoding on the unique categories returned to obtain the frequency a category / venue appears and sort this frequency in descending order (from highest to lowest) for each neighbourhood to have a feel whether the information corresponds to the characteristics of the neighbourhood.

Finally, we will apply clustering, an unsupervised machine learning algorithm, to the data obtained.