CAPSTONE PROJECT: BATTLE OF THE NEIGHBORHOODS Singapore Visitors and Expatriates Venue Recommendation

1. PURPOSE

This document provides the details of my final peer reviewed assignment for the IBM Data Science Professional Certificate program – Coursera Capstone.

1. INTRODUCTION

Singapore is a small country and one of the most visited countries in Asia. There are a lot of websites where travelers can check and retrieve recommendations of places to stay or visit. However, most of these websites provides recommendation simply based on usual tourist attractions or key residential areas that are mostly expensive or already known for travelers based on certain keywords like "Hotel", or "Backpackers" etc. The intention on this project is to collect and provide a data driven recommendation that can supplement the recommendation with statistical data. This will also be utilizing data retrieved from Singapore open data sources and FourSquare API venue recommendations.

The sample recommender in this notebook will provide the following use case scenario: A person planning to visit Singapore as a Tourist or an Expat and looking for a

• reasonable accommodation. The user wants to receive venue recommendation where he can stay or rent an HDB

• apartment with close proximity to places of interest or search category option. The recommendation should not only present the most viable option, but also present a

• comparison table of all possible town venues.

For this demonstration, this notebook will make use of the following data:

Final Assignment: Capstone Project: Battle of Neighborhoods Singapore Median Rental Prices by town.

• Popular Food venues in the vicinity. (Sample category selection)

• Nightlife

• Nearby Schools, etc.

• I will limit the scope of this search as FourSquare API only allows 50 free venue query limit per day when using a free user access.

1. DATA ACQUISITION

This demonstration will make use of the following data sources: Singapore Towns and median residential rental prices. Data will retrieved from Singapore open dataset from median rent by town and flattype from https://data.gov.sg website.

The original data source contains median rental prices of Singapore HDB units from 2005 up to 2nd quarter of 2018. I will retrieve rental the most recent recorded rental prices from this data source (Q2 2018) being the most relevant price available at this time. For this demonstration, I will simplify the analysis by using the average rental prices of all available flat type. Singapore Towns location data retrieved using Google maps API. Data coordinates of Town Venues will be retrieved using google API. I also make use of MRT stations coordinate as a more important center of for all towns included in venue recommendations. Final Assignment: Capstone Project: Battle of Neighborhoods Singapore Top Venue Recommendations from FourSquare API (FourSquare website: www.foursquare.com) I will be using the FourSquare API to explore neighborhoods in selected towns in Singapore. The Foursquare explore function will be used to get the most common venue categories in each neighborhood, and then use this feature to group the neighborhoods into clusters. The following information are retrieved on the first query:

Venue ID

• Venue Name

• Coordinates : Latitude and Longitude

• Category Name

Another venue query will be performed to retrieve venue ratings for each location. Note that rating information is a paid service from FourSquare and we are limited to only 50 queries per day. With this constraint, we limit the category analysis with only one type for this demo. I will try to retrieve as many ratings as possible for each retrieved venue ID.

1. METHODOLOGY

Singapore Towns List with median residential rental prices. The source data contains median rental prices of Singapore HDB units from 2005 up to 2nd quarter of 2018. I will retrieve the most recent recorded rental prices from this data source (Q2 2018) being the most relevant price available at this time. For this demonstration, I will simplify the analysis by using the average rental prices of all available flat type. Data Cleanup and re-grouping. The retrieved table contains some un-wanted entries and needs some cleanup. The following tasks will be performed: Drop/ignore cells with missing data.

• Use most current data record.

• Fix data types. Post Processed Singapore towns list with and median residential rental

• prices Adding geographical coordinates of each town location.•