Investing in a Sushi Restaurant in London

Sukriti Shukla

08th June, 2019

Data Acquisition & Cleaning

For the analysis of Tai Chi's problem, I have tried to understand his situation from a broad perspective while secondly, from a granular perspective. The data acquisition therefore, involves both macroeconomic data sources like the ones about the economy & demographics of London/United Kingdom as well as micro-level data pertaining to the Sushi restaurants in London neighborhoods, their ratings & reviews.

Expenditure & consumption patterns of UK's households like Year-on-year annual household expenditure growth and its region-wise expenditure have both been studied by me in order to understand the purchasing power & disposable incomes of UK households. These act as important determinants for eating out in restaurants & cafes.

Data Sources & Selection –

A. **Macroeconomic data of London/United Kingdom** –

A.1 Demographic indicators –

i. ***Population & Ethnicity in London*** (source link) - I have tried to use Office for National Statistics (ONS) Annual Population Survey to analyze borough-wise share of London population as well as ethnic distribution of population in London. I have restricted my analysis for the year 2017 only as it is more relevant.

ii. ***Population growth in London*** (source link) - These population estimates and projections come from the latest revision of the UN World Urbanization Prospects & represent the Urban agglomeration of London, which typically includes London's population in addition to adjacent suburban areas. Growth rates have been compared for the period 2000-35 showing future projections & past records.

iii. ***Ethnicity in London vs Rest of United Kingdom*** (source link) - These population estimates and projections come from the latest revision of the UN World Urbanization Prospects. I have studied the ethnic mix of population in London vs. other regions in the United Kingdom as per the 2011 Census.

iv. ***Ethnicity in London Boroughs*** (source link) – Wikipedia page here produces data about the Demography of London. I have scraped the data in the table Race by

borough section of this page to show the proportion of races by London borough (as found in 2011 census).

A.2 Economic, Consumption & Expenditure Indicators-

i. ***Number of International Visitors to London*** (source link) - Visitors from overseas broken down by duration, quarter and country of origin 2002 to 2017 (Quarterly data to Q3 2018). Table includes data on visits, nights and spending. All data has been taken from the International Passenger Survey (IPS). I have only used data for Visits (in 000s) by Country and considered only visitors from Japan during recent period (2012-18) to evaluate Japanese tourists visiting London.

ii. ***Year-on-year annual household expenditure growth in UK*** (source link) – I have used data from Office for National Statistics in order to understand consumer trends in UK. Household final consumption expenditure (HHFCE) for the UK, as a measure of economic growth. Here, I have used the UK's data pertaining to chained volume measure (this measure removes the effects of inflation) for the period 2011-18 to show year on year percentage growth in expenditure of households.

iii. ***UK's Household expenditure by region vs. UK average expenditure per week*** (source link) – This data source helps understand region-wise Average weekly household expenditure on goods and services in the UK, for financial year ending 2015 to financial year ending 2017. It tries to compare the £ per week expenditure in each of these regions (including London) with the UK average.

iv. ***Nationwide House Price Index for London and the UK*** (source link) - Nationwide Regional House Price Index (Quarterly since 1973) is one of several indicators for house prices published on a quarterly basis. I have considered housing indicator's data from 2013-18 only for estimating the Annual % change in house prices in London vs. that in the UK.

v. ***How businesses rate London as a place to do business in 2014*** (source link) - The 2014 London Business Survey (LBS) is an innovative survey designed by the Office for National Statistics. This dataset contains information on London as a business location and the factors affecting businesses presented by enterprise size & industry sector as a percentage of business units. I have tried to evaluate the survey results but scraped the data & tried to restrict my study to only Accommodation, food, travel and tourism industry in particular.

vi. ***Factors affecting businesses in London in 2014*** (source link) - The 2014 London Business Survey (LBS) is an innovative survey designed by the Office for National Statistics like the above one & studies the factors affecting businesses including the costs of energy and materials, taxes and business rates, government regulations, travel infrastructure, IT/connectivity. Here too, I have restricted myself to data with respect to only Accommodation, food, travel and tourism industry in particular.

B. **Micro-level data London Neighborhoods, Sushi Restaurants** –

B.1 London Neighborhoods' Data Compilation –

Data related to areas in London along with their neighboring boroughs was found from the below 2 sources on Wikipedia page. London is administered by the City of London and 32 London boroughs.

i. ***List of areas of London*** (source link) – I have tried to scrape through the Wikipedia page to extract table listing down major locations in London with their Post Codes, OS grid reference (Ordnance Survey National Grid reference system is a system of latitude and longitude) & their respective London boroughs (a borough is an administrative division in various English-speaking countries).
ii. ***List of London boroughs*** (source link) – I have combined the data derived from the above source with the list of boroughs data here to enumerate a list of all the London boroughs with their corresponding coordinates.
iii. ***London Neighborhoods Data Visualization*** (Folium Choropleth Maps) – London neighborhoods' location & corresponding latitude, longitude details were compiled from Wikipedia pages above to create a choropleth map. It comprised of all the London borough locations superimposed on the map of London city.

C. **Explore & Cluster Sushi Restaurants (Foursquare API location data)** –
Finally, I have leveraged on the Foursquare location data to execute my search for Sushi restaurants in London neighborhoods around the center of London city. I have explored & clustered all venues, their ratings, tips & reviews to help Tai Chi find the most suitable Sushi restaurant in London neighborhood for his investment.

Data Cleaning –

Data downloaded or scraped from multiple sources & webpages were cleaned & formatted by me as per the desired requirements for my analysis. These were then combined to perform further analysis in Python. I have used Macroeconomic data regarding demographics & economic indicators of United Kingdom & London pertaining to more recent years (mostly 2010 onwards up till 2017-18) in order to make it more relevant. In most cases, data for 2019 was not readily available on the web, hence not included.

There have been some limitations with the datasets which would have made the analysis difficult thus, some adjustments have been made in compilation. Firstly, demographics & economic data for United Kingdom is an important factor determining the growth of restaurant industry. Office for National Statistics, one of the largest independent producers of official statistics therefore, has been primarily used to draw economic & ethnicity-related statistics for this project. Data regarding number of international visitors to London have been compiled for all the countries. However, this data has been cleaned primarily to obtain Japan-specific numbers so as to highlight the changes Japanese immigrants to London over the last few years.

Secondly, the data pertaining to London neighborhoods was scraped from Wikipedia page and converted to data frame tables. However, some Locations had multiple London boroughs against them for example *Acton* location had *Ealing, Hammersmith and Fulham* as its London boroughs therefore, for ease of mapping them to respective coordinates I have selected only one of the London boroughs against these locations. So, now Acton location has its London borough as Ealing and coordinates of Ealing are shown in the latitude & longitude columns.

Thirdly, while using location data services of Foursquare API to search for Sushi restaurant venues around London center there were some duplicate venue names filtered in the data frame. For example, the YO! Sushi restaurant or the Kulu Kulu Sushi restaurants are chains of Sushi restaurants located in different locations across London. Each outlets of these chain of restaurants was regarded as a separate venue. I have tried to compare the ratings of each of these outlets but considered only the highest rated outlet as one venue option representing the entire restaurant chain in my analysis.