Market Perception of Banks in London
Oscar J. Urizar

November 1, 2020

1. Introduction
In this section we provide a description of the problem and a discussion of the background

1.1   Background
Retail banking, also known as consumer banking or personal banking, is banking that provides financial services to consumers as individuals not businesses. Retail banking is a way for individual consumers to manage their money, have access to credit, and deposit their money in a secure manner. Services offered by retail banks include checking and savings accounts, mortgages, personal loans, credit cards, and certificates of deposit (CDs).

1.2 Problem
Retail banking remains a highly competitive business, with banks aiming to gain any edge on the competition. Although several services are conducted online, many other services are provided to customers at office branches. The perception of customers towards their bank is highly important in assessing the performance of banks around cities where they operate in order to plan future improvements. We propose a project to gain insights in the market perception of major banks in London by performing a sentiment analysis on the text feedback provided by customers visiting various banks of this branch.

2. Data Acquisition and Preprocessing

2.1 Data Sources
For this project we are required to find the location of office branches of the various banks operating in London. Furthermore we need text feedback/reviews from people who have visited these venues. In summary we require two main data components:

- Location of office branches
- Text reviews from customers

The data provider FourSquare contains a robust database covering the requirements for this project. Firstly, It provides data of multiple types of venues, including banks. Secondly, It also provides tips as plain text provided from people who have actually visited these venues.

2.2 Data Acquisition
To obtain this data we will be using the FourSquare API. The API implements REST operations to query the required data, for example, using the venue category id

Category Id: 4bf58dd8d48988d10a951735
Category Name: Banks

We can implement a REST request to obtain information about the venues. For this project we will use the following data attributes

Attribute	Description
id	Unique identifier for this venue
name	Recorded name of the venue
Distance	Distance (m) from the provided location
Lat	Latitude coordinate of the venue
Lng	Longitude coordinate of the venue

Once we have identified the venues of interest, we can use the id to make a REST request and retrieve the following tip's data attributes

Attribute	Description
Venue id	Unique identifier for this venue
Id	Unique identifier for this tip
Created at	Datetime stamp of creation
Text	Tip's text
Lang	Language of the tip's text

2.3 Data Cleaning
The data downloaded is stored in CSV files for further use in this project. The files are loaded into pandas dataframes and data is manipulated to obtain a form suitable for this project.

The cleaning process for the venue's data involves checking a valid venue's name is provided, and grouping the venues by the bank they belong to. All the banks with only one venue will be grouped into 'others'.

Cleaning venue's tips encompasses several other steps are performed in the tip's text, including:

- Remove tips not in English language
- Remove invalid characters
- Remove punctuation
- Remove stopwords

3. Further Steps

In the final week of this course, the data collected and preprocessed will be used to conduct a sentiment analysis. The pre-trained model provided by the package TextBlob will be used to assign a sentiment score to each tip's text. The mean sentiment score of the tips for each bank will be reported as the market perception metric. A series of plots will be provide it to understand the insights found throughout the project.