

Shopping Mall

Liverpool, UK

Rutvi Shah | Coursera Capstone | October 2020

# Introduction:

Liverpool ONE is the largest open-air shopping center in the UK and the [ninth-largest shopping center overall](https://en.wikipedia.org/wiki/List_of_UK_shopping_centres).

Each store was created by a different architect, leading to stark differences between some buildings, one way in which Liverpool ONE differentiates itself from other shopping centers.

* With around 170 stores and services, Liverpool ONE is the largest shopping center in the city.
* For many local people and especially tourists, visiting shopping malls is great enjoy and relaxing. They can buy almost anything they wish in one place, go to restaurants or go to cinema.
* Opening shopping malls, allows property developers to earn consistent rental income. Opening shopping malls requires serious consideration and takes many factors that makes the whole process more complex.
* Location of shopping mall is one of the most important decisions and can determine how successful the project will be; and we will examine the city clusters to decide where the place for shopping mall will be the best for the process of building

## BUSINESS problem:

The goal of this project is to analyze and decide where to build a shopping mall in Liverpool (United Kingdom).

Using Data Science and Machine Learning algorithms, this project try to answer the question is it worth building a new Shopping Mall in Liverpool and especially determine where to build it.

Database:

1. List of neighbors in Liverpool.
2. Latitude and longitude of these neighbors.
3. Venue data, especially the data about shopping malls

The dataset containing its boroughs will be retrieved from Wikipedia [13] by applying Web Scraping to get table from an HTML webpage. Unfortunately, there is no information available about their neighborhoods nor latitude or longitude values.

On this operation, we make calls to its database by sending Request/GET sentences to retrieve this information. It will be stored in a .json file, which will be parsed and converted to a Pandas data frame to start working with its information.

After creating both datasets, we will join them to create a Pandas data frame containing all the information required. The data frame will contain the following columns:

1. Location
2. Latitude
3. Longitude

APIs used to gather venues and their coordinates:

As mentioned before, to gather the coordinates of each shopping mall where this information is missing, we will make calls to the ArcGIS database by its World Geocoding REST API [14] , to get the latitude and longitude values.

On the other hand, to cluster the neighborhoods based on their venues, we will use the Foursquare API [15] to make calls to its database and retrieve a .json file containing the venues of the different area.

Python libraries used for the Capstone:

To import the data, preprocess, make an exploratory analysis and then model it and evaluate the results, we need to import several Python libraries to perform these several tasks that we will carry out during the final project. This is because these libraries contain the functions and methods needed to perform the segmentation of the neighborhoods. The following table summarizes each Python library that will be used in the Final Capstone and its description:

Sources and Methods:

We will be using web scraping tools like Python requests and beautifulsoup packages to extract the data from Wikipedia page("https://en.wikipedia.org/wiki/Category:Areas\_of\_Liverpool") which has 41 neighborhoods. With Python Geocoder we will get longitude and latitude of every neighborhood in the city.

Then we will use Foursquare API, who provide many categories of venue data. This project contains many Data Science skills, from data scrapping, working with Foursquare, data wrangling, data visualization, data cleaning, machine learning & map visualization (Folium).

**We can summarize the tasks to perform in this project as follows:**

1. Business and Data understanding to come up with a problem and define the data requirements to solve it.

2. Data Preparation: wrangling, formatting and preprocessing it to prepare data for further analysis.

3. Exploratory Data Analysis: this step is done to summarize the main venues from each neighborhood to later group them based on the categories of their most common venues. 4. Modelling: with the datasets from Lima and Santiago de Chile, we apply the k-Means Clustering Algorithm per each city to cluster their neighborhoods. This is done to find which ones are more similar based on their venues’ categories.

5. Evaluation: each model is evaluated to segment and label which neighborhoods are the most touristic ones.

6. Deployment: write and show to the potential stakeholders the touristic neighborhoods and governments of Chile and Peru, to improve their touristic packages post COVID-19 pandemic.

After this step, the data frames from the touristic neighborhoods of both cities are merged into one. Then, this data frame is used in the following steps:

7. Re-Modelling: the data frame which groups the touristic neighborhood is used to run again the k-Means Clustering Algorithm to cluster and find which neighborhoods are more similar based on their most popular venues’ categories.

8. Re-Evaluation: from this model, we segment and label the touristic neighborhoods across the cities studied: Lima & Santiago de Chile.

9. Re-Deployment: again, we deliver a stack of proposals to the touristic neighborhoods and governments from Lima and Santiago de Chile. But now, the focus is to propose them cooperative alliances and policies that can help their countries to recover the touristic industry, in a faster and more effectively way, their economic recovering.