Introduction

For many people, visiting shopping malls is one way to unwind during the weekends and vacations. There are a multitude of activities one can do at shopping malls, such as watching a movie, visiting fashion outlets, dining in at restaurants, purchasing groceries or just simply to window shop and look at the latest promotions and discounts that are available. Shopping malls are also a great hang out location to catch up with friends and family. Therefore, the location of a shopping mall is extremely vital to attract people to patronise it. Considering the perspective of retail shop and business owners, they will rent a place at a shopping mall to market their products if they envisage a large crowd and high population density in the area. On the same note, property developers also take advantage of the large crowds that shopping malls pull to set up hotels and residential properties in the area. Therefore, the location of shopping malls plays large part in deciding whether it would make constructing one profitable.

Business Problem

This capstone project will identify the most optimal location in Bangkok, Thailand to construct a new shopping mall. Leveraging on data science and analytics techniques as well as machine learning algorithms, this project aims to find the best solution to tackle the identified business problem: Where in the city of Bangkok will you recommend a property developer set up a new shopping mall?

Scope and Target Audience of Project

The target audience of this project is that of property developers, investors and retailers looking to invest in a new shopping mall in Bangkok, Thailand. Bangkok is one of the world's top tourist destinations with more than 22.7 million arrivals per year. Bangkok also ranks 4th in cross-border spending and has a diverse range of attractions to cater to both international and domestic tourism in the country. As such, it is a prime city for developers to build shopping malls as well as retail owners looking to market their brands by leveraging on the high volume of people in the area.

Data Science Methodology

To tackle this problem, we will need to obtain the following data:

- List of districts in Bangkok, along with the latitude and longitude coordinates of these districts
- Data on the existing shopping malls

Data Sources

The data on different district information can be obtained from Wikipedia (https://en.wikipedia.org/wiki/List_of_districts_of_Bangkok). From this, we will web scrape the page to obtain the relevant information we need.

Thereafter, the Foursquare API will be used to get data on the existing shopping malls in these districts. Upon collection of data, machine learning techniques (k-means clustering) will be employed to determine the optimal location to build the new shopping mall.

Data Science and Analytics Skills

The data science skills employed in this project include web scrapping, utilising an API, cleaning, and wrangling of data, machine learning algorithms (k-means clustering) and data visualisation skills.