

Capstone Project

Introduction/Business Problem

Stakeholders have hired me to investigate of into which neighbourhood in the city of Dublin is the best to open Indian Restaurant. As a data scientist, I'll be developing a model to get some recommendation on which are to most ideal locations to invest.

I will be taking a city (Toronto), which is similar to Dublin in many ways. Toronto is also more developed in some categories, which Dublin is hoping to do in coming years. So these are one of the reasons investors are interested in opening a restaurant.

With the growing diverse population, a Indian Restaurant investment will be ideal setup if I am able to pin-point which are the best Neighbourhood to open this venture.

Data

Toronto Dataset:

- Web scrapping to obtain the data that is in the table of postal codes from https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
- Get the latitude and longitude for each postcodes from http://cocl.us/Geospatial_data
- Downloaded Dublin town names with longitude and latitude from <https://www.irelandtownslist.com>
- Using Foursquare API, we will get nearby venues to each neighbourhood and will also group it by 'Indian Restaurant'.
- Create a model that uses number of venues in neighbourhoods except Indian restaurants inputs and number of Indian restaurant as outputs. We'll be building predictive model with SVR algorithm.
- Then use Dublin neighbourhood as target to predict 10 ideal locations to start Indian Restaurant

Methodology

Using Regression modelling, we learn and predict numbers of Indian restaurant in Toronto City.

The following link provides information on how I accessed Toronto post codes and neighbourhood.

https://github.com/johnsg811/GitHub_Sample/blob/master/Capstone_week3.ipynb

In the following link, which contains the main project:

https://github.com/johnsg811/GitHub_Sample/blob/master/Capstone_week3_Part2.ipynb

- The first few lines of code describe how I joined two datasets which will contain latitude and longitude which will help us map the exact geo location.
- Using Foursquare API we extracted neighbourhood venues and venue categories
- We group it by neighbourhood and Indian Restaurant.
- Then, created a model to predict Indian Restaurant.
- Imported Dublin neighbourhoods and explore venues
- Then, using the build model to predict the number of Indian restaurant in the city and compare with existing number of Indian Restaurant.
- This provided an ideal place for our investors to start an Indian Restaurant.

Result/Conclusion

Using this model, it provided best 10 ideal location in the city of Dublin for our investors to start an Indian Restaurants. This is not limited to only finding Indian restaurants, but also works as a system to recommend any type of venues in the given city.

Detailed information and script can be found in the following links:

https://github.com/johnsg811/GitHub_Sample/blob/master/Capstone_Project_Presentation.pdf

Completed notebook:

https://github.com/johnsg811/GitHub_Sample/blob/master/Capstone_week3_Part2.ipynb