Coursera IBM DATA SCIENCE FINAL PROJECT

Laurence Lin

2020

1. Introduction

Selecting the appropriate location for a retail store is an important task. Time and money spent searching for location of store, surveying the market demand is an non-neglectable cost, and will result in huge loss if the location is decided in an improper place. Proprietor and investors may want to find a way to intelligently select an optimal location for a new retail store, reducing the risk of financial loss. Past research applies geographical data and mobile data to model the anticipate popularity of a new opening store. Foursquare API provides rich geographic location data for various venues, which helps data scientist to apply location data on multiple application field. In this project, we exploit Foursquare API to access geographic data and data from the other location API to analyze the factors that influence a retail store's popularity and rating, and build prediction model that could estimate the popularity and ratings of a new retail store for a location. We believe that Foursquare API geographic data could help investors and retailers to select the optimal places for a new retail store.

2. Data Description

Locational based service has provided informations for data scientist to analyze tasks around a certain area. In the retail store location selection task, business analyst should consider factors such as demography, incoming flow, cuisines habit, competitiveness and lots of elements. Foursquare API enables us to scrape geographic data around certain venue and human activities, past researchers have achieve reliable analysis and receive valuable results. We select some geographic features obtained by Foursquare API and another Zomato API location data for retail store popularity as prediction target. Our goal is to analysis the location data features that is correlated to a retail store's popularity, and build promising model.

Features from Foursquare API is selected as follows: