Coursera IBM Data Science Professional Certification Coursera Capstone Project

Instructor: Alex Aklson

Submitted by: Luis Lopez

Table of Content

Contents

Capstone Project – Week 1 Part 1

Assignment Requirements.

Part 1. Submit a description of the problem intended to solve and a discussion of the background information.

The development of this capstone project will proceed along the guidelines and procedures in the methodology presented in course 3 "Data Science Methodology" of "Coursera IBM Data Science Professional Certification" specialization.

This methodology is driven by 10 methodology questions that provide context to the 10 development phases that a typical data science project should adhere in the path toward successful completion.

Assignment Part 1: Introduction/Business Problem.

The context of the problem that this assignment intends to resolve is as follows:

> *A business person, in the city of Sydney, who has been in the area of food service provider for 5 years, is considering to open his own restaurant for the first time.*
>
> *You are the data scientist that this business person has hired to get an objective data-based recommendation on where in the city of Sydney he/she should locate the new restaurant.*

This part of the assignment deal with the following 2 of the 10 crucial questions of the data science methodology:

Question 1: What is the problem that we intent to solve?

Question 2: How can you use data to answer the question?

1.1 Business Understanding Phase.

Methodology Question 1: What is the problem that we intent to solve?

A series of interviews with the business person, **the client** from now on, has uncovered a good deal of the business needs, the context and the requirements.

The client has been in the food services area for about 5 years and he feels quite confident about the concept and brand of the restaurant he is planning on; however, he would like to have his assumptions tested and confirmed by sound quantitative analysis.

The insights gathered during the interviews revealed to following factors will have to be considered when analysing the available data that support the study.

A. Location. To assess where in the city of Sydney the new restaurant should be established.

B. Cuisine. To assess references to food and cooking characteristics of a particular country or region, together with typical dishes and menus.

C. Demographics. To characterise customer base in terms of genre and age.

D. Atmosphere. To assess references to interior decoration details, plants, flowers, lighting, background music, and other ambience details.

1.2 Analytic Approach Phase.

Methodology Question 2: How can you use data to answer the question?

By definition of the assignment, the idea is to leverage the Foursquare location data to come up with a solution to the chosen problem.

From the insights gained in the Business Understanding Phase, we can say that the application of clustering methods & techniques is the appropriate analytic approach to address this problem.

Each of the four factors, Location, Cuisine, Demographics, and Atmosphere, define four different clustering.

The Location clustering has already been resolved in one of the lab of this course; that use, use Postcodes to partition a city in mutually exclusive groups.

For the other factors we need to find a measure of similarity to be able to apply a clustering algorithm.

Once the four clustering has been acomplished we will have a cluster_ID for each of these factors for every restaurant in the city of Sydney.

By the end will have a Restaurant Matrix Score of that will enable us to try various recommendation algorithms.