IBM Data Science Professional Certificate

Applied Data Science Capstone
The Battle of Neighborhoods (Week 1)

Business Case: Best location for a new Breakfast Restaurant in Dubai

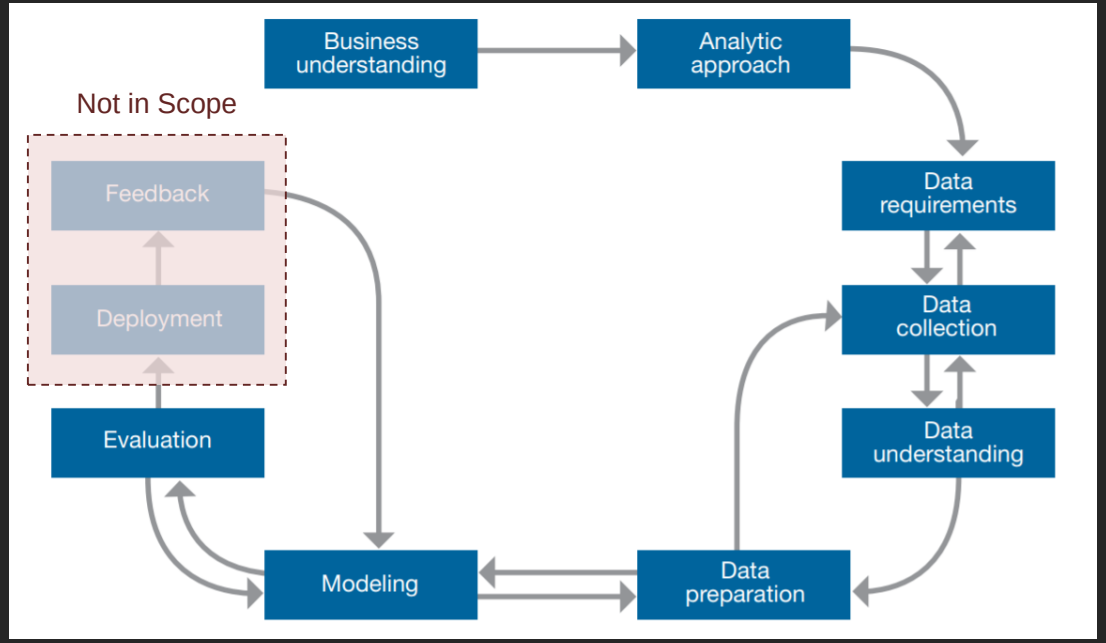
Anthony Lapierre
March 2019

Business Problem

- We were approached by a popular Breakfast Restaurant Chain to identify the best neighborhood to launch their brand in Dubai.
- The company is new to Dubai and want to understand the neighborhoods to support their decision making process.

Methodology

- We will be using the *"IBM Foundational Methodology for Data Science"* throughout this case study.
- This presentation will provide an overview of our strategy for all steps with exception of *"Deployment"* and *"Feedback"*.



Link to: IBM Foundational Methodology for Data Science

Business Understanding

Our Business Sponsor already started to study the Dubai Restaurant Market and provided us with the following context to develop our Business Understanding:

- It is a well know fact that the choice of location for a Restaurant in Dubai is a key success factor: accessibility, parking space, visibility, high footfall, high urban commercial/resident population within a one-kilometer radius are all important factors.
- Hence, this is why we must develop a Data Model that will allow us to analyze those aspects.

Analytic Approach and Data Requirements

1. The analysis must be based on data from the Foursquare API Venue Database.
2. The business sponsor wants the restaurant to be profitable as soon as possible; we must avoid isolated areas or areas under construction, which will be removed from our list of neighborhoods.

Methodology

1. Export Neighborhoods and Breakfast Spots from Foursquare API using our Jupyter notebook.
2. There might be duplicates, inaccuracies or inconsistencies for neighborhood data. We will be using QGIS to make corrections to the data. We will also use QGIS to visualize the Breakfast Spots results and develop our understanding of the data. We will then export the neighborhood locations to GeoJSON format into our Jupyter notebook.
3. We will build our Data Model by perform a train-test-split and KNN Classifier to find the best K. We will then use K Means to cluster the neighborhoods.

Methodology