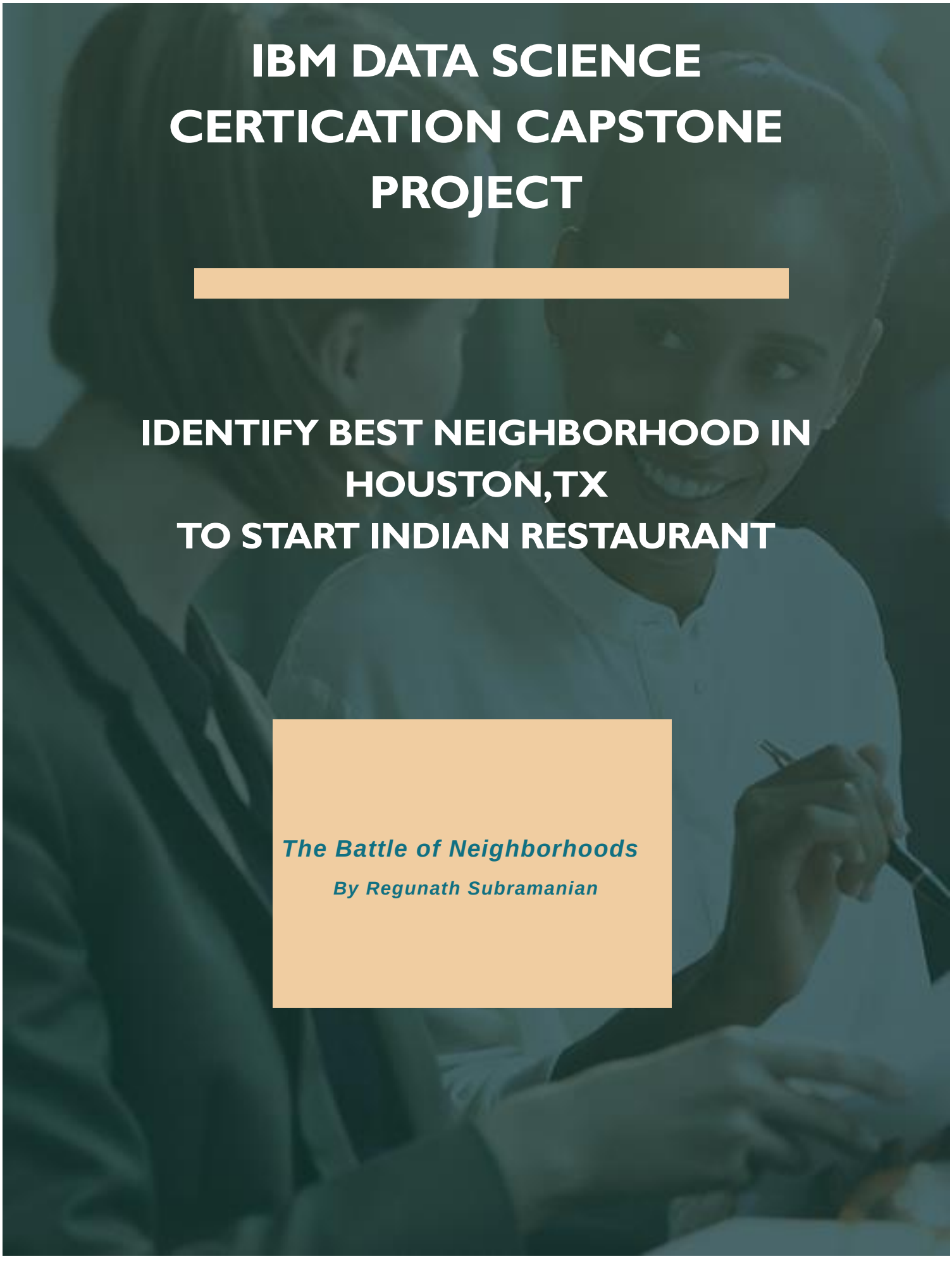

IBM DATA SCIENCE CERTICATION CAPSTONE PROJECT

IDENTIFY BEST NEIGHBORHOOD IN HOUSTON,TX TO START INDIAN RESTAURANT

The Battle of Neighborhoods

By Regunath Subramanian

TABLE OF CONTENTS

INTRODUCTION

Houston is the energy capital of the world; it's the headquarters and intellectual city for virtually all segment of the oil industry including technology, exploration, production, marketing, transmission, and supply. Houston is the fourth largest populous city in the United Status. One of my clients would like to start an Indian restaurant in Houston, TX and they would like to identify the optimal or best neighborhood. Therefore, this project will perform data analysis and try to find the most optimal neighborhood to open the Indian restaurant according to those criteria. It's obvious, that there are many additional factors, such as distance from parking places or distance from the main streets, but this analysis can be done after choosing the neighborhood, and thus will not be performed within the scope of this project. The insights derived from analysis will give good understanding of the business environment which help in strategically targeting the market. This will help in reduction of risk. And the Return on Investment will be reasonable.

1. BUSINESS PROBLEM

Due to oil prices, it is always full of highs, lows and learning for the restaurant industry in Houston. For the past few years, lower oil prices and increased competition have put significant pressure on Houston's more than 12,000 restaurants. Houston restaurant community suffered damages, including lost homes, cars and businesses due to recent hurricanes such as Hurricane Harvey. These unfortunate circumstances highlight the precarious nature of the restaurant business, where margins are slim, and most operators do not have significant sums saved for emergencies. Restaurants in Houston still face many challenges: Oil prices have not fully recovered, competition grows as more concepts open, rents are unreasonably high in many areas, and it's still a challenge to find enough qualified hospitality workers to keep restaurants fully staffed. In spite of these challenges and the ups and downs, Houston's restaurant and bar community will only get bigger and better in 2018.

Due to all these scenarios opening a new Indian restaurant requires serious consideration and is a lot more complicated than it seems. Particularly, the location of the restaurant is one of the most important decisions that will determine whether the restaurant will be a success or a failure. The objective of this capstone project is to analyze the data and select the best locations in the city of Houston, Texas to open a new Indian restaurant. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question to find an optimal neighborhood in the city of Huston, TX to open a new Indian restaurant.

Target Audience of this project

This project is particularly useful to restaurant entrepreneurs and investors looking to open or invest in restaurant in the city of Huston, TX.

2. DATA DESCRIPTION

The data below will be used to analyze this problem and make a recommendation to the client:

- List of neighborhoods in Huston, TX.
  - The below Wikipedia page would contain a list of neighborhoods in Huston, TX.
    - https://en.wikipedia.org/wiki/Category:Neighborhoods_in_Houston
  - Web scraping techniques is used to extract the data from the Wikipedia page, with the help of Python libraries and beautifulsoup packages.

- Latitude and longitude coordinates of those neighborhoods in order to plot the map and also to get the venue data.
  - Geographical coordinates will be retrieved for neighborhoods using Python Geocoder package, which will give us the latitude and longitude coordinates of the neighborhoods.

- Venue data, particularly data related to restaurants and it would be used to perform clustering on the neighborhoods.
  - Foursquare API will be used to get the venue data for these neighborhoods. Foursquare has one of the largest databases of 105+ million places and it had 50 million monthly active users . Foursquare API will provide many categories of the venue data, especially restaurant category.

- The below data science techniques and skills will be used to analyze and solve the business problem.
  - Web scraping Wikipedia Data
  - Using with Foursquare API
  - Data cleaning
  - Data wrangling
  - Machine learning techniques such as K-means clustering
  - Map visualization by using Folium

3. Prepare a report to present the solution or/and recommendation