



The Cuban Meal Crisis

Capstone Project for
IBM Applied Data Science Course

June 2020

By Krishna Prasad S Rao

Introduction

Los Angeles (LA) is one of the most populous and multi-cultural city in the west coast of the USA. Just a few hundred miles south of LA is the country of Mexico, with a populace that majorly speaks Spanish. Cuba, on the other hand, is a small island country located not far away from the USA, in Atlantic Ocean. Cuba is just under 100 miles from Miami, another metropolitan city within the USA.

Although both Cuba and Mexico are Spanish speaking nations, culturally they are significantly different, which also reflects in their unique colorful and spicy foods. Thus, although Mexican restaurants are aplenty in LA, authentic Cuban food is hard to come by. This creates an opportunity, especially in a metropolitan city like LA.

This project is inspired from the 2014 movie, “Chef”. The title of the report is a word play on the geopolitical event, famously known as the ‘Cuban Missile Crisis’ that occurred in the cold-war era between the US and the Soviet Republic.

Business Problem

The goal of this project is to methodically explore the best possible locations of selling authentic Cuban food, in a food-truck, within the populous LA. Food trucks enjoy the unique advantage of mobility, allowing varied locations to be targeted for selling food. Using data science methods and predictive statistical techniques such as clustering, the project aims at providing answers to the following questions:

- a. What are the ideal locations for selling Cuban food, in the city of Los Angeles, in a food truck?*
- b. What are the ideal times (of the day) for selling Cuban food at the locations obtained from (a)?*

Target Audience

The report is of key interest to the proprietor of the Cuban Food truck. Besides, LA being a densely populated city with good amount of locals and tourists, a food truck can create traffic congestion as well. Thus, the findings of the report can be key to the city's traffic and police authorities as well.

Data & Sources

To answer the questions put forth in the 'business problem' section, the following data sets are essential:

- a. List of businesses to identify commercial / residential locations which can be useful for targeting the right crowd at varied times of the day. The dataset was found on [kaggle.com](https://www.kaggle.com), at this [particular URL](#), consists of fields like business name, co-ordinates, zip-codes and business category.*
- b. The above dataset does not contain 'neighborhood names'. To obtain those, [this URL](#) is scraped and then merged with the above data set, with Zip Code as the key.*
- c. While the combination of (a) and (b) yields information about businesses, Foursquare API data is used to additionally provide locations of competing Cuban restaurants, so that other locations can be targeted for selling Cuban food. To achieve this goal, the venue and search APIs of [foursquare.com](https://www.foursquare.com) is utilized.*

Sources mentioned in (a) and (b) above are available as 'csv' file and html respectively, while the data source in (c) is available in json format, which are all directly readable by various APIs of Python's Pandas package.