The Unfolding of a Burrito Tycoon's Rapid Expansion

Finding the right place to charge extra for guac

Jean-Luc Dinello

Problem Statement: Chipotle Mexican Grill has gone through some dramatic changes in the past couple of years. Up until 2015 the company saw tremendous growth. However, after their virus outbreaks in 2015, the company's stock dropped nearly 45%. After making several internal changes within the company such as hiring a new CEO and safety consultant the company has bounced back. Their CEO has recently said that he wants to double the amount of stores in the US[1]. Because all of Chipotle's locations are company-owned, each store opening has far more risk than other chains that are owned by franchisees. This business model also gives them far more control into where they choose to put new locations. However the company needs to attain insights into what locations would provide a high chance of success for this aggressive expansion.

Value to Client: This project aims to present several key findings regarding customer data. Things such as what brands are closely tied to theirs, distance the average customer travels to their successful locations, and census data on the area of their successful locations. By the end of the project the company should have a roadmap in place for determining the viability of a new location

Datasets:

1. The first dataset that I found for this project comes from a company called SafeGraph. The Neighborhood Consumer Insights & Movement Patterns dataset takes in data from a Census Block Group (CBG) which represents a neighborhood of ~1500 people and combines this with anonymized GPS movement data to show to popularity of a CBG and distances traveled. They also combine the data with their private SafeGraph Places data to determine visits to places like stores or restaurants and show top brands. **https://www.kaggle.com/safegraph/visit-patterns-by-census-block-group**
2. The second dataset that I plan on using is the Census Block Group American Community Survey Data that I also got from SafeGraph. This dataset shows demographic data from the American Community Survey(2016) with 5 year estimates. This data comes in the form of an geoJSON file. https://www.kaggle.com/safegraph/census-block-group-american-community-survey-data

Solution Methodology: I want to create a model using the K-Nearest Neighbor approach to create a cluster of census blocks that would be optimal for the next locations. I also would like to try other classifications to test and find models with high levels of recall or precision to come to

---

[1] https://www.bloomberg.com/news/articles/2019-03-14/chipotle-becomes-another-boring-chain-and-that-s-a-good-thing

make a stronger recommendation. I then want to use these clusters to find potential blocks in the census that would make strong candidates for further expansion.

Data Cleaning: Because both of these datasets come from SafeGraph, a data company, I suspect that they will be very clean. Nonetheless, I plan on examining the sets for missing values, extreme outliers, and duplicates. I also will need to determine which features from the census data are relevant to the goal of the project.

Data Wrangling: The contents of the Census Data come in the form of an archive file that will need to be extracted into the notebook. I will need to merge several of tables together regarding geographic data for mapping visualizations, the customer patterns data and the census data. Luckily most of the tables contain the census block group ID making this project relatively simple. The geographic information comes in the form of a geojson file that will need to be wrangled into a pandas dataframe.

Exploration and Analysis: I first plan on exploring the data related to Census Blocks where the company is successful. Showing key demographic information such as Age, Persons in Family, Race, Sex, Empolyment, Industry, and Military Personel along with visualizations depicting correlations, linearity, with popularity of the company's brand. I then plan to follow this with an analysis on the top associated brands and conduct a similar study on the same demographic factors for each associated brand.

Deliverables:

Jupyter Notebook: I will construct a Notebook Using Jupyter Lab to showcase my process and display the code along with markdowns describing each step taken along the way. This can also assist in creating the framework for the story I am trying to tell with this data.

Tableau Story: With the code that I will have made with the notebook, I will be able to further customize my visualizations in tableau and implement even more interactive elements. I plan to compile these visualizations into a tableau story to show the transformation sequences of these visualizations to enhance the story of the project.

Presentation: With the enhanced visualizations, I can create a structure that perfectly lends itself to a presentation to showcase the work that was put into the project. I can potentially do this for the program and even put a recording of it on my website.

Write-Up: I will create a brief write-up that details the findings of the project. This write up will describe the methodology, process, and summarize the conclusions made from my analysis. This report will be an easy way for interested parties to look into what I have found whether or not they are a data specialist.