Ice Cream Truck New Business Evaluation

Phoenix, Arizona

Abstract

Comprehensive evaluation utilizing Python, Data Visualization and Machine Learning that evaluates if Phoenix, Arizona would be a lucrative market to operate an Ice Cream Truck Business.

James Martin

January 2020

Introduction

For the Capstone Project, Battle of the Neighborhoods, I decided to use data science and machine learning methods to analyze if the city of Phoenix would be a good market to start and operate an Ice Cream Truck business.

Problem Statement: A local investor is trying to decide if he should open an Ice Cream Truck business within the City of Phoenix. The business would focus on driving an Ice Cream Truck to various local parks to provide a convenient way for parents to purchase ice cream for their children.

The investor believes Phoenix is an ideal market for an Ice Cream Truck business due to its warm weather and growing population. However, before launching the business and making a capital investment, the investor wants to validate:

1. That most local parks in Phoenix do not already have Ice Cream options nearby and
2. That Phoenix has a large population of children to consume the ice cream.

Solving the Problem and Providing Insights:

The investor is seeking data to answer these questions, but does not know how to find and analyze the necessary data for insights and answers.

The investor is a personal friend of mine and reached out for assistance as he knows I have a passion for data and recently acquired new data analytic skills through IBM's Data Science Certification courses on Coursera.

To answer these questions, I leveraged publically available data sources, which I then analyzed using Data Science, Data Visualization and Machine Learning techniques.

The following report, along with accompanying notebook, contains the necessary data analysis to provide insights into Phoenix Parks and Phoenix Demographics. The analysis can be leveraged to make a data driven decision regarding launching an Ice Cream Truck Business in Phoenix.

Data

We need to obtain data in order to answer the Investor's questions:

1. Do local parks in Phoenix already have Ice Cream options nearby?
2. Does Phoenix have a large population of children to consume the ice cream?

To answer #1 we will use two data sources:

A. **Phoenix Park Data** - Fortunately, Phoenix has publically available data about the location of all its parks available at https://phoenixopendata.com/dataset/parks/resource/4dedd0ad-ea1e-4000-aec4-e4feca296ea1
   - This data is in a CSV file and contains the Longitude of the Park (X), the Latitude of the Park (Y), The Park Name (PlaceName), Park Address (AddressFull), City (AddressCity), State (AddressState), Zip (AddressZip) and Park Details (ParkWeb).
   - This data can be placed into a pandas dataframe for indepth analysis and we can also use the Folium library to visualize the data on a map.

B. **Foursquare API** – Foursquare is a technology company that has a large amount of accurate location data. Foursquare has an API that provides access to this data that over 100,000 developers leverage to analyze location data. With Foursquare, we can evaluate what types of venues are nearby to a specific location. It allows us to get details and even user reviews on these venues. This is extremely valuable for our evaluation as it will enable us to determine if there are yogurt or ice cream shops nearby Phoenix Parks. It will also allow us to further segment and cluster Phoenix Parks using a Machine Learning approach called K-Means. Foursquare can be accessed at www.Foursquare.com.

To answer #2 we will use two public data sources from Phoenix's City Website.

A. **Phoenix Household Demographics** – Phoenix has publically available data about its household demographics available at https://www.phoenixopendata.com/dataset/33ab1c58-d8a9-46fc-b65e-50a62d8eb928/resource/29c9d1b5-3620-4622-909c-e3b2eab59a35/download/phoenix-az-household-type.csv.
   This CSV file has a wealth of information about the number of Phoenix households and the type of households. This is helpful as it shows data on Family vs non-family households in Phoenix. For our Ice Cream shop business, the more family households the better! With this data, we can input it into a Panda Dataframe and conduct several data visualization approaches such as Line, Bar, Pie, Area, Histogram and Box Plot to further analyze the data and obtain insights.

B. **Phoenix Age Data** - Phoenix has publically available data about its Age demographics available at https://www.phoenixopendata.com/dataset/fd0400b1-f62f-4d4e-9a88-253798891158/resource/6f460cd1-d0aa-4005-aadb-

c371772cbd7b/download/phoenixazdemographic.csv. What we are interested in with this data is what % of the population is under 18. Unfortunately, this data does not give this specific answer, so additional data manipulation and calculations will be necessary to obtain this answer. Fortunately, there are python and panda scrips we can run to get the answers we are looking for.