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

Contents

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.

Methodology

To conduct our data analysis, I created a Jupiter notebook leveraging the IBM Watson platform. Once complete, the notebook was uploaded to my github account.

The first step was importing all the necessary tools for the data analysis. This included:

- Tools for python such as pandas, numpy, and JSON
- Data visualization tools such as geocoders, folium and matplotlib
- Machine learning tools like sklearn, which will be used for K- Means Clustering

Part 1: Creating a Dataframe of Phoenix Parks

The next step was to create a panda dataframe of the Phoenix Park's data from https://www.phoenixopendata.com/dataset/6a6b86bc-94b3-4995-9128-7a3e60042e1a/resource/4dedd0ad-ea1e-4000-aec4-e4feca296ea1/download/parks20171024.csv. This was accomplished using Read_CSV code and then running a .head() command to see the first 5 rows of the data and ensure the data was imported correctly.

Once imported there was some minor data cleanup required such as renaming columns and dropping a column. I also ran a command to see the shape of the data.

Part 2: Visualization of Phoenix Parks

Next, I used the geopy library to get the latitude and longitude of Phoenix, Arizona. I then used the Folium Library to provide a map of Phoenix with the Phoenix parks layered on top of it. The map is fully interactive with labels for the parks and ability to zoom in and zoom out.

Part 3: Using Foursquare API to Segment and Explore Parks

Now we are ready for the foursquare API call. First, I needed to define my Foursquare credentials and version. I was then able to get the longitude and latitude of the first park and conduct a GET request URL to look at the top 10 venues nearby the first park.

The next step was to analyze the **get_category_type** function to determine what type of venues are nearby and then input this into a Pandas Dataframe. I then repeated this approach for all parks and analyzed the type of categories nearby.

Next, I created a Dataframe grouped by parks that showed how many of each venue categories were nearby. I ran a describe command that showed the mean for each venue category along with some other statistical details.

Further, I created smaller dataframes isolating details for the “Ice Cream Shop” and “Yogurt Shop” categories. Then for each park, I printed the top 5 closest venues nearby,

Part 4: Cluster Parks

Next we used K-Means to cluster the Parks into 5 clusters. K-Means is a form of unsupervised machine learning that is used to find ways to group the data. For our analysis, I decided to cluster the data into 5 groups or clusters. Further I was able to visualize the clusters over a map of Phoenix using Folium. This would have been impossible to do without Machine Learning.

Part 5: Examine Clusters

Once we established the clusters, we were able to examine each cluster and determine the categories that distinguished it, using a similar approach as in Part 4.

Part 6: Examine Phoenix Household Demographics

Now, we move our focus to Phoenix Demographic data. Our first data source is the Phoenix Household data. First, I created a dataframe from the Phoenix Household CSV data that is publically available.

I had to do some cleanup on the data such as removing columns and changing the index to be the year. I was then able to use Matplotlib to visualize the data. I used Area graphs, line graphs, histograms and box plot charts.

Part 7: Examine Phoenix Age Demographics

Finally, I obtained Phoenix Age Demographic information from Phoenix’s Public Website. This data showed how many residents Phoenix had from 2008-2015 based on many age groups. The challenge here is that the data does not clearly show how many

residents are under 18, which is the question I am after. It also separated the data by male and female, which is irrelevant to our needs. So once I created the dataframe, I had to manipulate the data frame to sort out fields and columns that were not needed. Further I then had to use basic math to calculate the percentage of residents under the age of 18.

Results

Part 1: Creating a Dataframe of Phoenix Parks

A dataframe was successfully created, showing 223 Phoenix Parks.

]:

	Longitude	Latitude	Park	Address	City	State	Zip
0	-112.049636	33.763867	Apache Wash Trailhead	1600 E SONORAN DESERT DR	PHOENIX	AZ	85085
1	-112.090026	33.476654	Encanto Park Clubhouse	2605 N 15TH AVE	PHOENIX	AZ	85007
2	-112.065841	33.449188	Thomas House	609 E ADAMS ST	PHOENIX	AZ	85004
3	-112.065662	33.449587	Teeter House	622 E ADAMS ST	PHOENIX	AZ	85004
4	-112.065416	33.449587	Silva House	628 E ADAMS ST	PHOENIX	AZ	85004

We now have a dataframe showing the Phoenix Park Data. Let's also validate all the data was imported

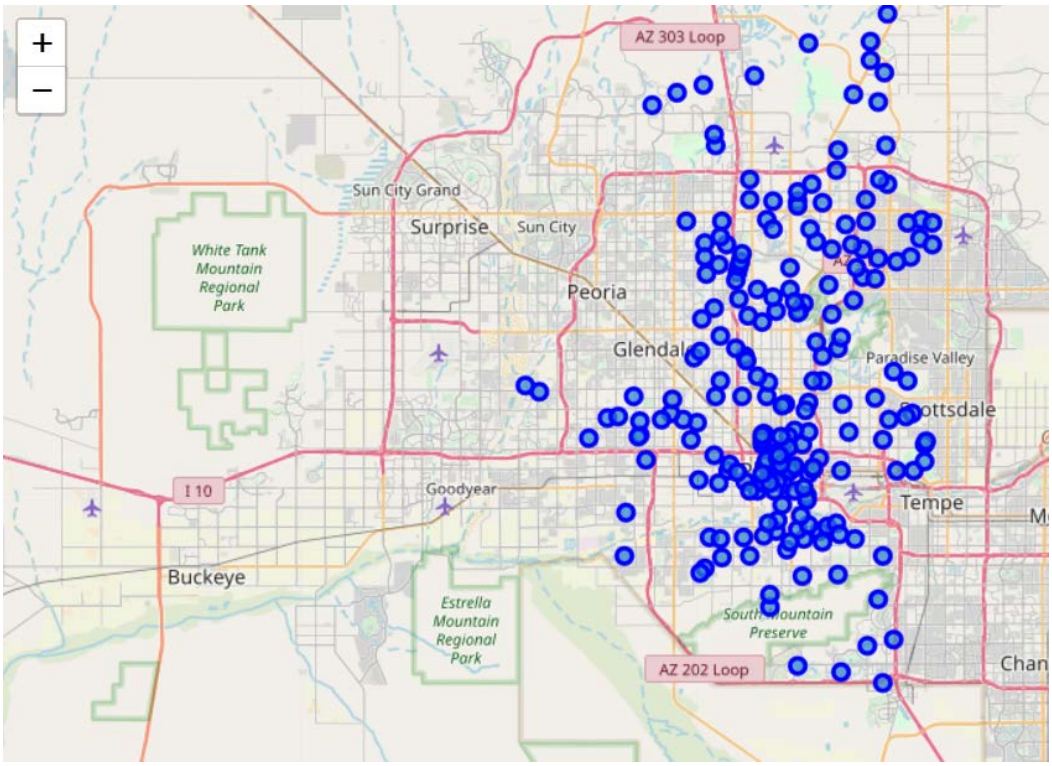
]:
```
Park_data.shape
```

]:
```
(223, 7)
```

There are 223 Phoenix Parks

Part 2: Visualization of Phoenix Parks

Folium provided a beautiful visualization of a map of Arizona with Parks layered on top of it.



Part 3: Using Foursquare API to Segment, Explore and Analyze Parks

- Sample of results from the GET request for the first Park:

```
{'meta': {'code': 200, 'requestId': '5e20b401df2774001ce10870'},
 'response': {'suggestedFilters': {'header': 'Tap to show:',
   'filters': [{'name': '$-$$$$', 'key': 'price'},
    {'name': 'Open now', 'key': 'openNow'}]},
  'headerLocation': 'Phoenix',
  'headerFullLocation': 'Phoenix',
  'headerLocationGranularity': 'city',
  'totalResults': 142,
  'suggestedBounds': {'ne': {'lat': 33.85386671000009,
    'lng': -111.94157803535826},
   'sw': {'lat': 33.67386652999991, 'lng': -112.15769316464171}},
  'groups': [{'type': 'Recommended Places',
```

- Panda's dataframe showing venue categories nearby

	name	categories	lat	lng
0	Apache Wash Trailhead	Trail	33.767786	-112.044700
1	City Of Phoenix Sonoran Preserve Dixie Mountai...	Trail	33.741135	-112.097467
2	Barro's Pizza	Pizza Place	33.765427	-111.983815
3	El Encanto Dos	Mexican Restaurant	33.800032	-112.058591
4	Fry's Marketplace	Grocery Store	33.786054	-112.119405

- Through the analysis we were able to determine there were 225 unique venue categories nearby, two of them being “Ice Cream Shop” and “Frozen Yogurt Shop” which we care the most about.
- On average the “Ice Cream Shop” and “Frozen Yogurt Shop” are only near a park less than 1% of the time!

Now lets look at the Ice Cream Column

```
In [32]: park_grouped["Ice Cream Shop"].mean()
Out[32]: 0.008071748878923767

In [33]: park_grouped["Frozen Yogurt Shop"].mean()
Out[33]: 0.001345291479820628
```

- We also were able to determine the 5 closest venues for each park!

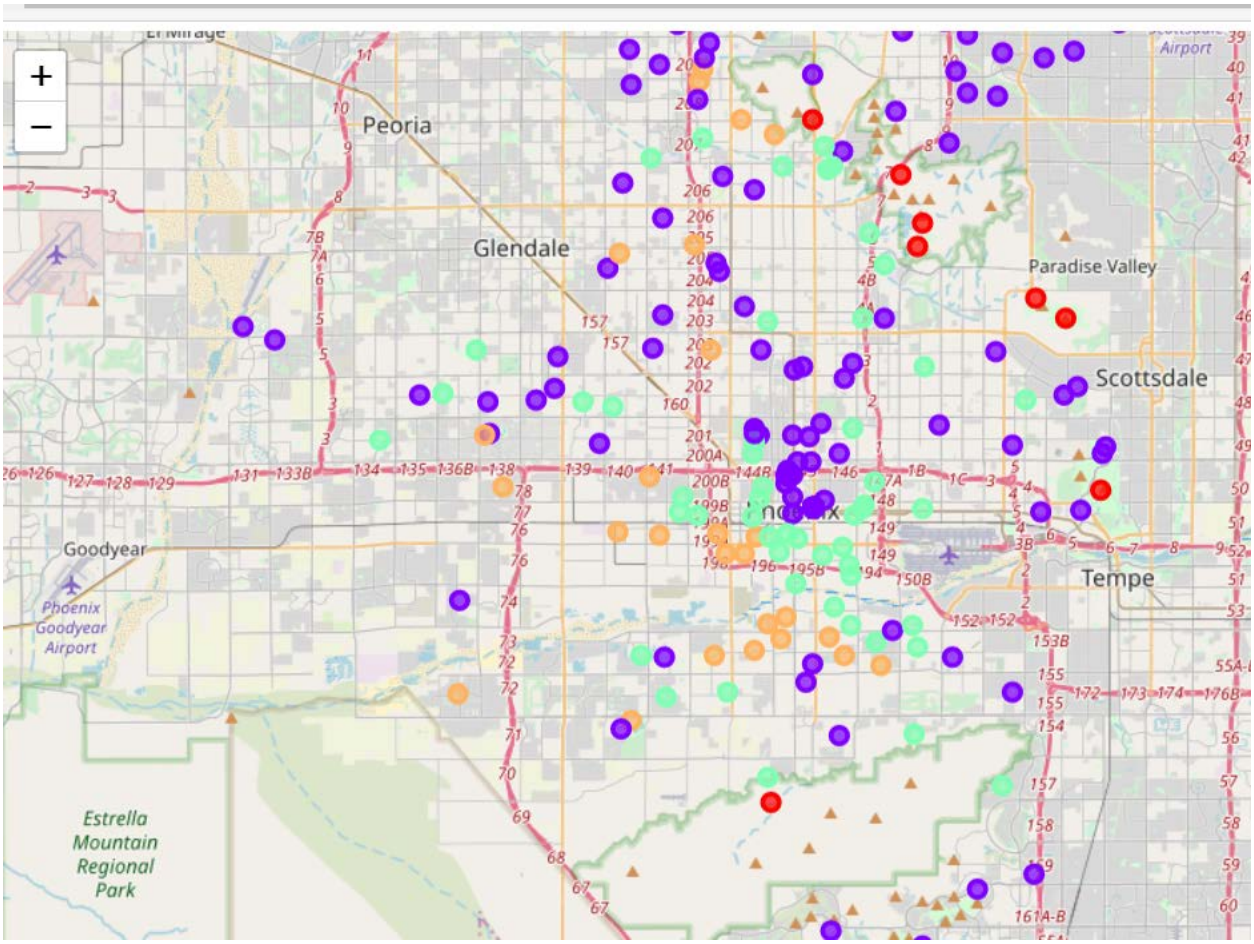
```
----10th Street Wash Park----
                venue  freq
0  Mexican Restaurant   0.2
1   Convenience Store   0.1
2           BBQ Joint   0.1
3             Brewery   0.1
4       Bowling Alley   0.1
```

- Along with a list of the top closest venues by park!

	Park	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	10th Street Wash Park	Mexican Restaurant	Bowling Alley	BBQ Joint	Greek Restaurant	Gym / Fitness Center
1	Acacia Park	Electronics Store	Hardware Store	Pizza Place	Fast Food Restaurant	Sandwich Place
2	Acoma Park	Cosmetics Shop	Home Service	Salon / Barbershop	Fast Food Restaurant	Greek Restaurant
3	Adobe Mountain Park	Gym	Convenience Store	Trail	Cocktail Bar	Pharmacy
4	Alicia Park	Grocery Store	Sandwich Place	Convenience Store	Discount Store	Rental Car Location

Part 4: Cluster Parks

We were successfully able to cluster the parks into 5 key clusters and visualize using Folium.



Part 5: Examine Clusters

We came up with 5 clusters which I named Trails, Hobbies, Scenery, Mexican Food and Fast Food. Here is an example of cluster 5. And the good news is no clusters were centered around Ice Cream or Yogurt!

```
park_merged.loc[park_merged['Cluster Labels1'] == 4, park_merged.columns[[1] + list(range(5, park_merged.shape[1]))]]
```

	Latitude	State	Zip	PlaceWeb	Cluster Labels1	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6 C
11	33.459857	AZ	85009	https://www.phoenix.gov/parks/parks/alphabetic...	4	Fast Food Restaurant	Convenience Store	Fried Chicken Joint	Mexican Restaurant	Arts & Crafts Store	Vid
19	33.398717	AZ	85040	https://www.phoenix.gov/parks/parks/alphabetic...	4	Convenience Store	Bakery	Cosmetics Shop	Coffee Shop	Fish & Chips Shop	I
23	33.440353	AZ	85009	https://www.phoenix.gov/parks/parks/alphabetic...	4	Fast Food Restaurant	Intersection	Diner	American Restaurant	Discount Store	Conv
34	33.405236	AZ	85040	https://www.phoenix.gov/parks/parks/alphabetic...	4	Convenience Store	Chinese Restaurant	Intersection	Construction &	Mexican Restaurant	Fa Re

Part 6: Examine Phoenix Household Demographics

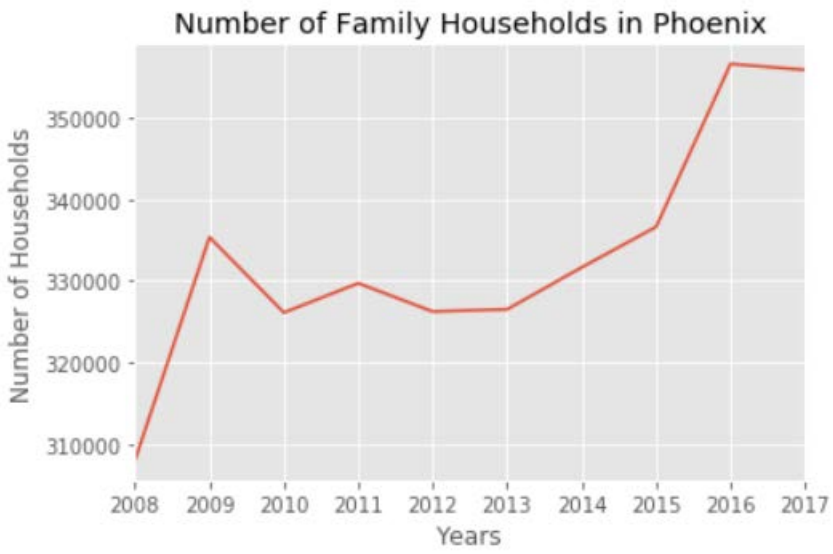
Successfully created a dataframe of Phoenix Household demographics.

]:

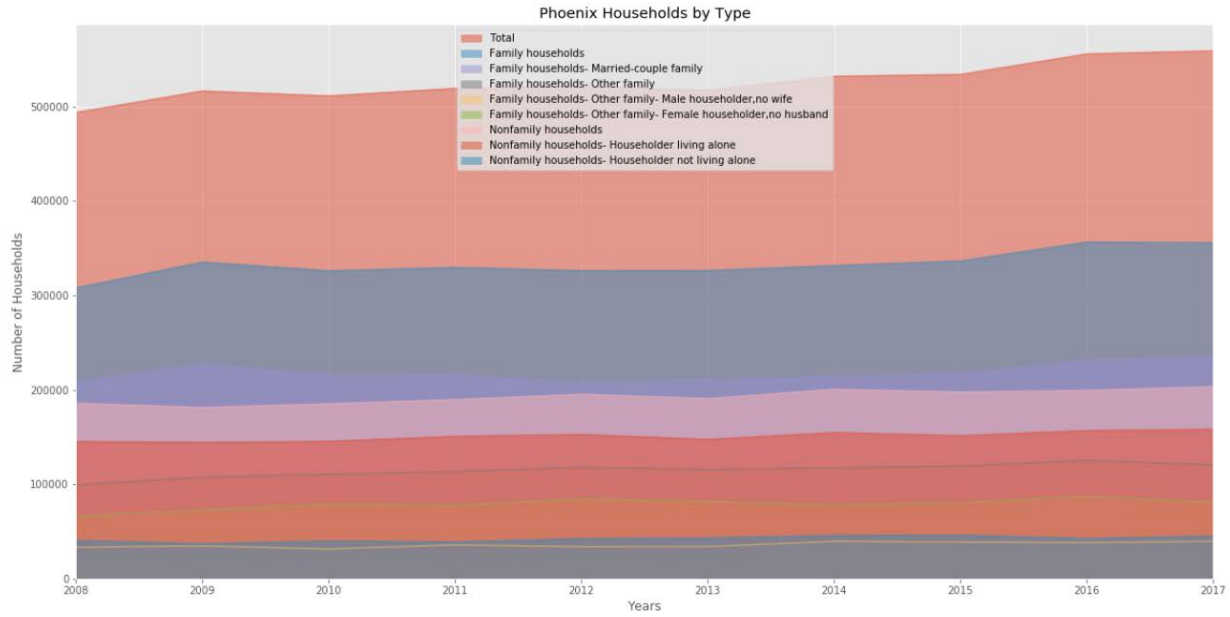
year	Total	Family households	Family households- Married-couple family	Family households- Other family	Family households- Other family- Male householder,no wife	Family households- Other family- Female householder,no husband	Nonfamily households
2008	493898	308067	208802	99265	33169	66096	185831
2009	516655	335343	227967	107376	34635	72741	181312
2010	511432	326132	215746	110386	31407	78979	185300
2011	519369	329717	216329	113388	35719	77669	189652
2012	521555	326255	208149	118106	33709	84397	195300

Was able to visualize the data in several ways to demonstrate that the number of Phoenix households is growing.

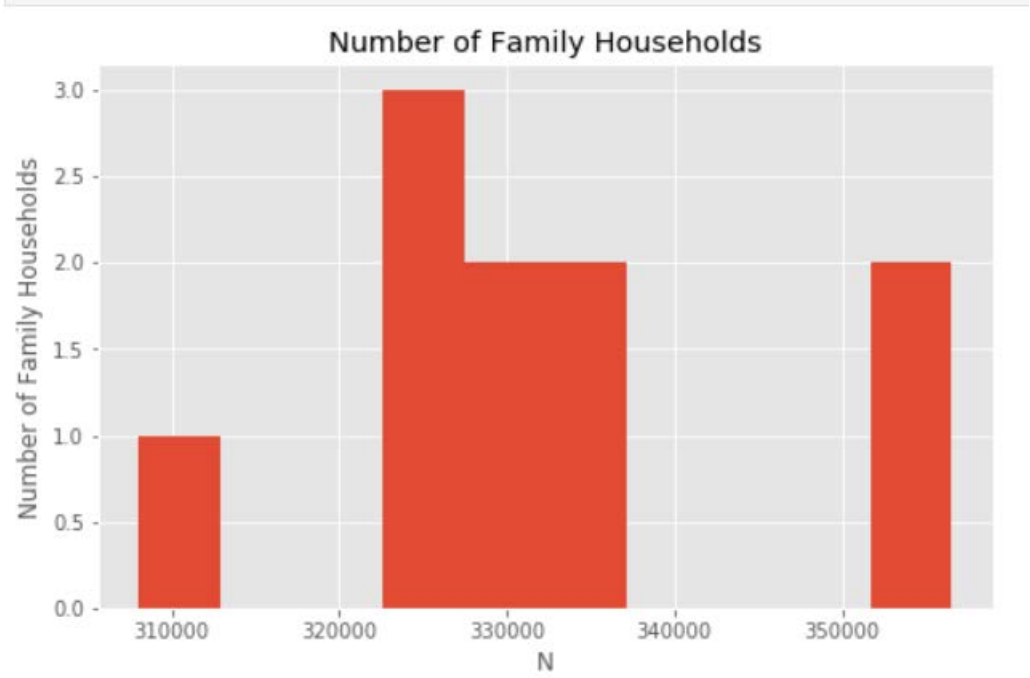
Line Chart



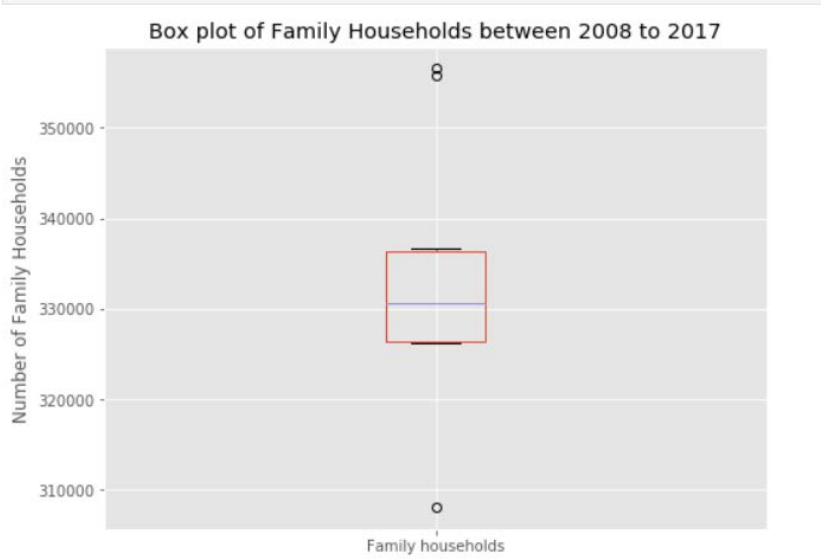
Area Graph



Histogram



Box Plot



Part 7: Examine Phoenix Age Demographics

Good news! We were able to determine ~26% of the population in Phoenix is under 18! Although Phoenix is thought of as a retirement destination, it appears there are plenty of young residents and young families to eat ice cream!

Discussion

I believe the investor will be very pleased with the results of the analysis. First, the investor will be impressed by the wealth of data that is available for free on public web sites.

To the investor's first question: **Do local parks in Phoenix already have Ice Cream options nearby?**

We were able to clearly show, using Phoenix Parks Data and the Foursquare API that no, parks in the Phoenix area DO NOT already have Ice Cream or Yogurt options nearby.

The investor's second question: **Does Phoenix have a large population of children to consume the ice cream?**

We were able to use Phoenix census data to show that the number of Households in Phoenix is growing steadily. We were also able to show that 26% of the population is under 18 and thus there will be many children wanting Ice Cream in Phoenix's hot weather!

Conclusion

We were able to successfully use data along with Data Analysis, Visualization and Machine learning techniques to answer the investor's questions. Although there are many factors to consider when opening a new Ice Cream Truck business, based on our analysis, Phoenix is a prime area to start an Ice Cream Truck business focused on public parks. Public Parks in the Phoenix area do not have Ice Cream or Yogurt shops nearby and there is a large population of kids who make for an ideal customer base.