

Starbucks Dataset Analysis

Team Starbucks

Abby Herrup, Angie Tran, Chris Haub, and Grant Wooldridge

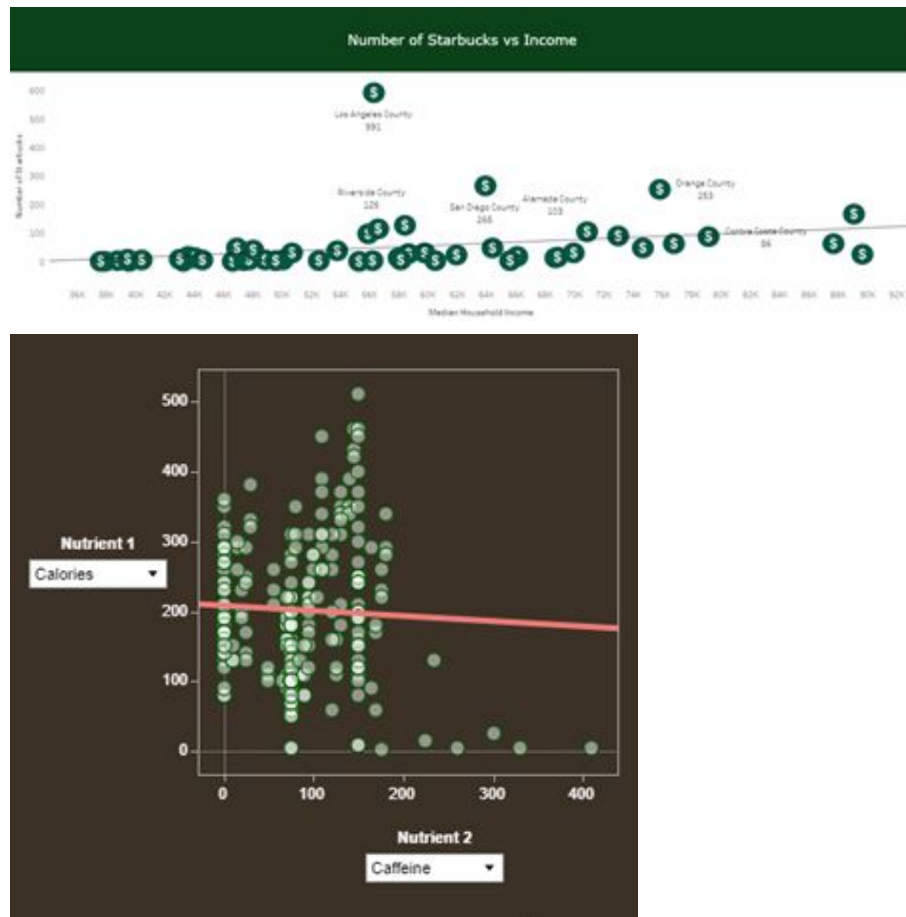
24 February 2021

TABLE OF CONTENTS

Inspiration and Overview	3
Data & Modeling Approach	4
Results of Data Analysis	4
(i) What might Starbucks look at when determining a location for a new store? Are there relationships between a Starbucks location and household income, population, or age?	4
Median Income vs Number of Starbucks	4
Total Population vs Number of Starbucks	5
Top 10 City Table	5
(ii) Are there relationships between beverage types and nutrition? Are there relationships between nutrients themselves?	6
Nutritional Visualization	6
Conclusions	7
Limitations/Bias	7
Future Work Recommendations	7
Works Cited	8

Inspiration and Overview

For our project, we chose to dive into Starbucks. We all really enjoy our morning cup of joe, and we wanted to learn more about the most successful coffee purveyor. We found some cool looking visualizations on tableau public and it gave us some inspiration on how we would go about diving into our own sets of data.



In our analysis of Starbucks, we used 3 different datasets. One, for drink nutrition information, another for store locations, and the last for US Census data. The drink nutrition dataset was a csv file that contained 242 different drink/preparation combinations along with nutritional facts such as fat content, protein, vitamins, etc. The store location dataset was in geojson format and consisted of coordinates, city names, state names, store numbers, and more. Finally the US Census dataset was a csv file that contained over 32,000 rows of data. Data points consisted of things like zip code, population, median age, household income, and more.

Questions to be asked from this dataset include: (i) What might Starbucks look at when determining a location for a new store, (ii) are there relationships between a Starbucks location and household income, population, or age, (iii) are there relationships between beverage types and nutrition, (iv) are there relationships between nutrients themselves?

Data & Modeling Approach

Our original datasets required some cleaning of the data. The Starbucks nutritional dataset had special characters and spaces in the column headings that needed to be fixed. The US Census dataset was pulled via an API call in Jupyter Notebooks, which required a join, and filling of null values with 0 or NaN. It also had some additional calculations performed in order to create rate data such as poverty rate, unemployment rate, and highschool graduate rate.

Results of Data Analysis

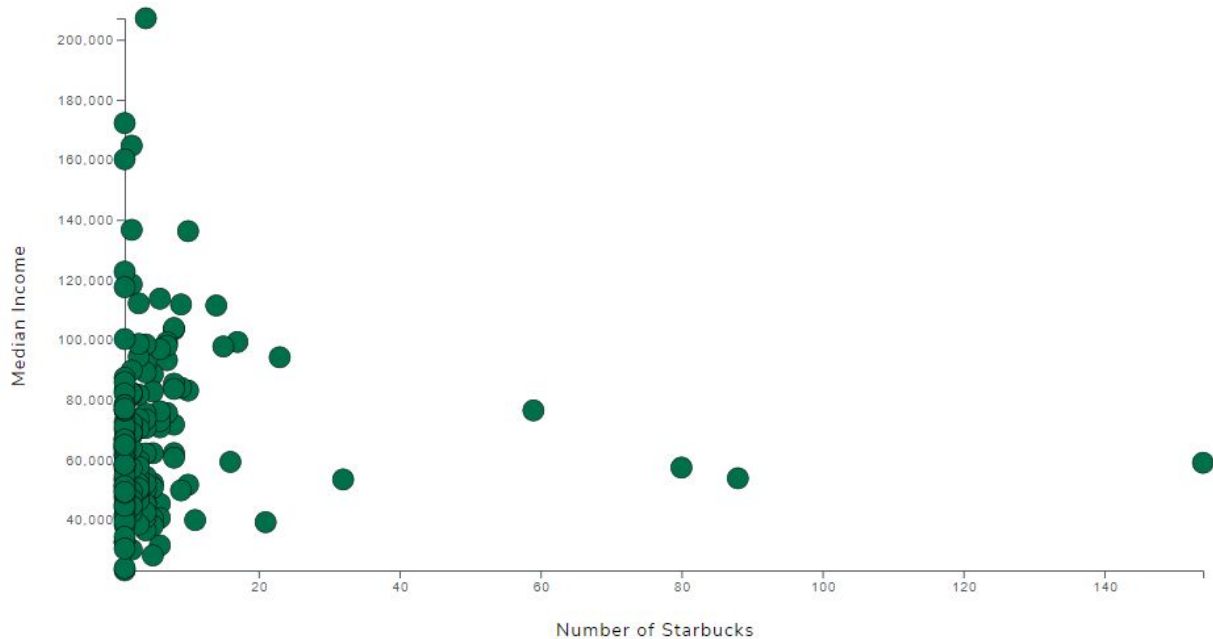
(i) What might Starbucks look at when determining a location for a new store? Are there relationships between a Starbucks location and household income, population, or age?

In order to determine the answers to the above questions, we chose to make a couple of scatter plots as well as a table that would all dynamically update when we chose a specific State in the United States. The scatter plots would compare the number of Starbucks locations to median income and total population, while the data table would display city names and the number of Starbucks, median income, and median age of each city.

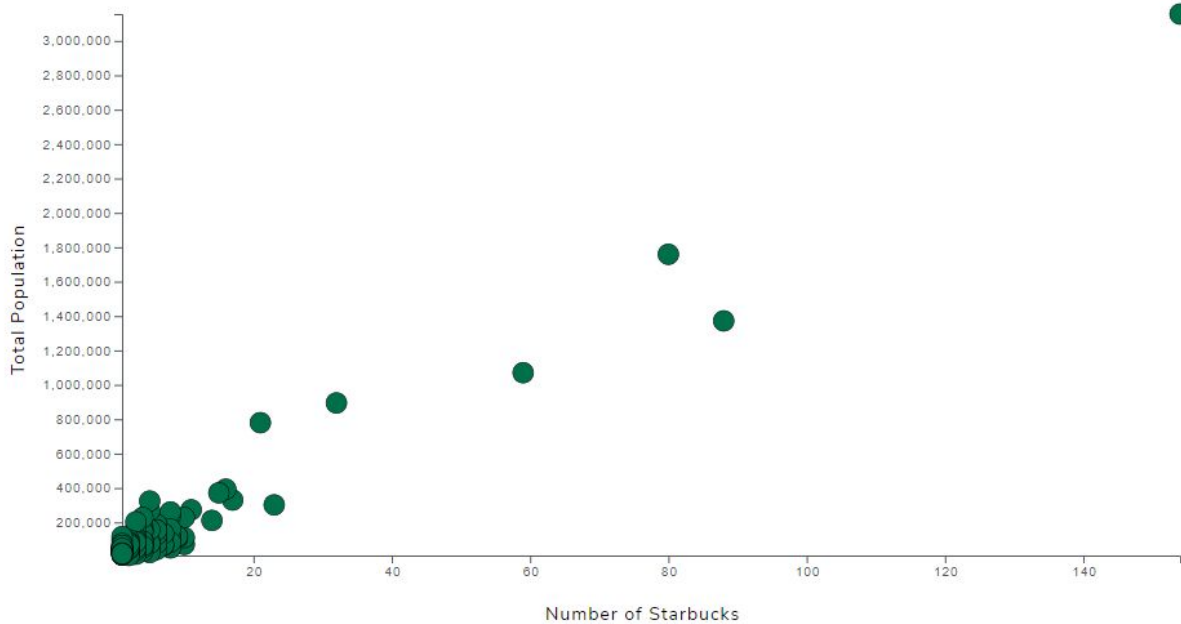
Median Income vs Number of Starbucks

CHOOSE A STATE

Texas ▼



Total Population vs Number of Starbucks



Top 10 City Table

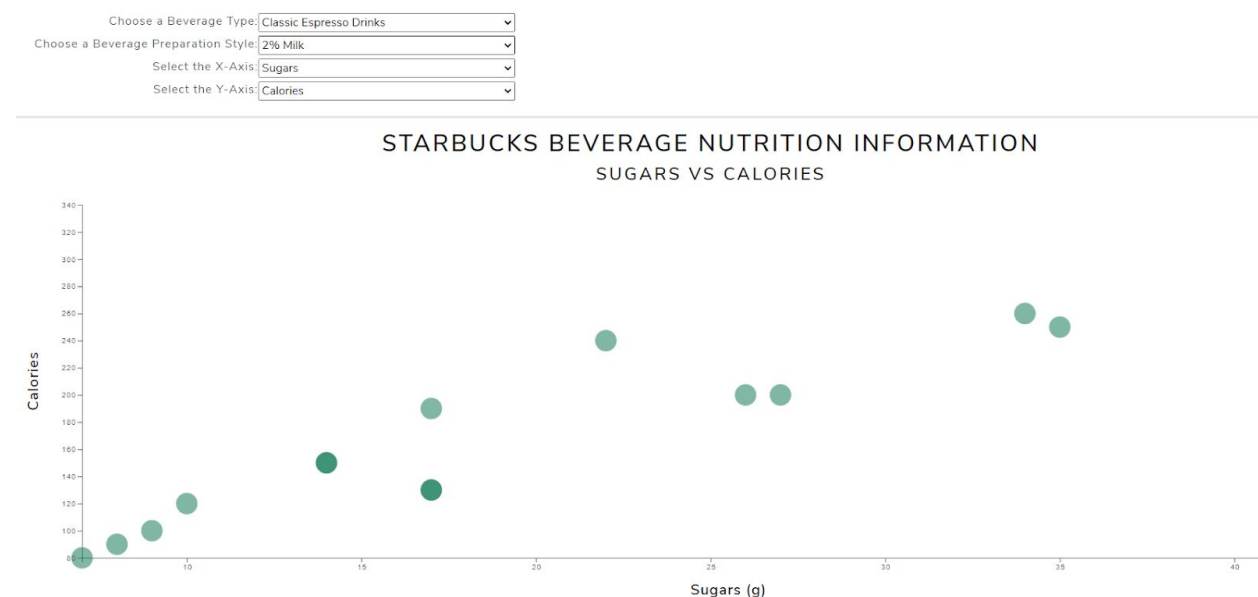
Demographic Info- Top 10 Cities by Number of Starbucks			
CITY	NUMBER OF STARBUCKS	MEDIAN INCOME	MEDIAN AGE
Houston	154	59019	33
Dallas	88	53851	33
San Antonio	80	57419	34
Austin	59	76478	35
Fort Worth	32	53461	32

We found that there really are not strong positive correlations between median income and Starbucks locations, however total population and Starbucks locations more often than not showed a slight positive correlation. It is worth noting that a lot of the positive correlations are skewed by major cities in each state.

(ii) Are there relationships between beverage types and nutrition? Are there relationships between nutrients themselves?

To take a dive into the questions above, we made a scatter plot that would allow the user to select a preferred beverage type, the way the beverage was prepared, and change the x and y axis to compare nutrients.

Nutritional Visualization



We found that there are many positive correlations amongst nutrients. Certain correlations we would expect such as the not-so-good for you nutrients like sugar and fat tend to be positively correlated with one another.

What was surprising to me was how protein fell into the mix. It showed positive correlations with calories, carbohydrates, sodium, and fat. All nutrients that are commonly perceived as bad for you when taken in excess. It also showed positive correlations to things such as vitamin A and calcium.

That being said, the positive correlations would fluctuate based upon beverage and preparation type. For instance, a Classic Espresso Beverage prepared with 2% milk showed a slight

positive correlation between protein and sugar, however when you change the beverage type to Tazo Tea drinks and keep 2% milk, the correlation becomes much more noticeable.

This wasn't just noticeable between types of beverages prepared the same way, but so to with the same beverage prepared differently. For instance Signature Espresso drinks show a slight positive correlation between sodium and sugar when prepared with 2% milk, but when prepared with soymilk there is zero correlation due to several of the beverages having no sodium.

Overall the correlations found are largely impacted by the beverage type, but preparation impacts the results as well. This makes sense when you consider the individual ingredients that make up each beverage. A Frappuccino made with nonfat milk is going to show positive correlations to sugars, fat, cholesterol and the like just because of the core ingredients that make up a Frappuccino. Conversely, even a black cup of coffee with hardly any nutritional value can be skewed with the addition of cream and sugar.

Conclusions

We believe that Starbucks primarily focuses on the population of an area when determining it's next store location.

Nutritional correlations are largely impacted by the type of beverage, but the way a beverage is prepared impacts the correlations as well.

Limitations/Bias

When analyzing Starbucks locations and demographic data, we only were able to perform the comparisons with US locations due to only having access to the US Census data.

Our nutritional data set was severely lacking in data points. It only contained 242 drink combinations when Starbucks offers over 80,000 drink combinations.

The Starbucks datasets we used are outdated, they all came from the year 2017. Starbucks has an API, however you need to be a Starbucks employee to have access to said API.

The data did not include any commercial development information from other companies that would give us insight if Starbucks coattails or teams up with other companies in developing new locations.

Future Work Recommendations

If we had more time we would include demographic data of other countries, not just the United States. This would allow us to get a clearer picture of Starbucks overall strategy when deciding a new location.

We would also work on removing outliers from our location and demographic page in order to see a more accurate picture of correlations between Starbucks locations and demographic data.

We would also have loved to pull data from Starbucks API directly. This would have allowed us to get the latest data, as well as a complete set of nutritional information to play with.

Works Cited

1. "Nutrition facts for Starbucks Menu." Kaggle, 20 July 2017, www.kaggle.com/starbucks/starbucks-menu/metadata.
2. fgregg. "Census." census, 22 Dec. 2020, github.com/datamade/census.
3. United States Census Bureau. "American Community Survey: 5-Year Estimates: Detailed Tables 5-Year." census.gov. 7 Feb. 2021. <https://api.census.gov/data/2017/acs/acs5>