Will Advertising Saturday Drink Specials On thedrinknation.com Increase Customer Satisfaction for Bars in Philadelphia?

brgarm

May 4, 2019

1. **Introduction/Business Problem**

A new bar, ABC Bar, is opening in Philadelphia, PA. It is looking for ways to increase customer satisfaction. One idea they have is to offer drink specials on a Saturday and advertise it on https://philly.thedrinknation.com/specials/Saturday# . To explore whether this would be a useful strategy for increasing customer satisfaction, ABC Bar will perform a data science study to see whether posting a Saturday special on thedrinknation.com correlates with increased Foursquare likes in Philadelphia, PA as an indicator of customer satisfaction.

2. **Data**

Philadelphia Bars Advertising Saturday Specials on thedrinknation.com Dataset

First, all Philadelphia bars advertising Saturday specials on https://philly.thedrinknation.com/specials/Saturday, and their addresses, were scraped using the python packages requests and beautiful soup. The addresses were then used to get the latitude and longitude coordinates of each of these bars by utilizing Nominatim from the python library geopy.geocoders. Bars for which latitude and longitude coordinates could not be found were excluded from the dataset. The latitude and longitude coordinates for each bar was required in order to be able to make "search" calls to the Four Square API, which was done in order to get the FourSquare VenueID for each bar in this dataset. The FourSquare VenueID for each bar in this dataset was then used to make "likes" calls to the Four Square API in order to get the number of Four Square likes for each bar in this data set. The name associated with each VenueID on Four Square, which was included in the "likes" api responses, was also included in the pandas data frame in order to confirm that the correct venues and likes were being retrieved from Four Square. A screenshot of the first few rows of this data set is found in Figure 1.

	BarName	BarAddress	Neighborhood	Latitude	Longitude	VenueID	FS_name	Likes
0	12 Steps Down	831 Christian Street Philadelphia, PA	South Philly	39.9390528	-75.1576839	40b28c80f964a520bdfa1ee3	12 Steps Down	39
1	2nd Story Brewing	117 Chestnut Street Philadelphia, PA	Old City	39.951862	-75.1718895	53b5cade498e7e8bccf28069	2nd Story Brewing Company	129
2	2nd Street Brew House	1700 S 2nd Street Philadelphia, PA	South Philly	39.926601	-75.149315	4d7c115e73ca54810712587c	2nd Street Brewhouse	20
3	a.bar	135 S 18th Street Philadelphia, PA	Rittenhouse	39.9504178	-75.1704316840335	51e4b307498eb7f2d3e8a95b	a.bar	72
4	Abbaye	637 N. 3rd Street Philadelphia, PA	Northern Liberties	39.9622191	-75.1429705	43e9e70bf964a5201f2f1fe3	The Abbaye	116

Figure 1. First few rows of the Philadelphia bars advertising Saturday specials on thedrinknation.com dataset.

Philadelphia Bars Not Advertising Saturday Specials on thedrinknation.com Dataset

 Next, the FourSquare API was called again using the search parameter and "Bar" categoryID to get a list of bars in Philadelphia, PA. Four square limits the number of calls to 50, thus only 50 bars could be collected in this dataset. The API response includes VendorIDs, which were then used to call the Foursquare API again, using the "likes" parameter, in order to get the number of Four Square likes for each of the bars in this dataset. Any bars appearing in both this list and the list of bars advertising on drinknation.com were dropped from the pandas data frame in order to formulate the Philadelphia Bars Not Advertising Saturday Specials on thedrinknation.com dataset. A screenshot of the first few rows of the Philadelphia Bars Not Advertising Saturday Specials on thedrinknation.com data set is found in Figure 2.

	id	location.address	location.lat	location.lng	name	likes
0	3fd66200f964a520e6f11ee3	1310 Drury St	39.950223	-75.162579	McGillin's Olde Ale House	557
1	49ac4499f964a520ae521fe3	1500 Walnut St	39.949334	-75.166176	Butcher and Singer	173
2	4b523b05f964a520627127e3	112 S 13th St	39.949849	-75.162030	Zavino Wine Bar & Pizzeria	155
3	56b3bc63498e0e0d584a1d17	120 S 13th St	39.949710	-75.161997	Double Knot	134
4	53757808498ee116741e16a6	131 S 13th St	39.949493	-75.161790	Charlie Was A Sinner	143

Figure 2. First few rows of the Philadelphia Bars Not Advertising Saturday Specials on thedrinknation.com data set.

Data Processing for Visualization & Linear Regression Analysis

The resulting data sets had 102 (Philadelphia Bars Advertising Saturday Specials) and 46 (Philadelphia Bars Not Advertising Saturday Specials) bars, respectively. These data sets were combined into one pandas data frame. To fill NA values in the "Philadelphia Bars Not Advertising Saturday Specials Likes" column of the resulting pandas dataframe, without biasing the data, the .fillna() pandas function was used to add the mean like value for the "Philadelphia Bars Not Advertising Saturday Specials Likes" column into each cell with an NA value. Additionally, the categorical variables were changed to integers (0 for Philadelphia Bars Advertising Saturday Specials and 1 for Philadelphia Bars Not Advertising Saturday Specials) in order to perform linear regression analysis.

3. **Methodology**

The FourSquare likes data was first explored by visualizing it via scatterplot using the python package matplot lib in order to get a rough idea of whether Bar Status (whether the bar advertises a Saturday Special on thedrink.nation.com or not) correlates with the number of Four Square likes a Philadelphia bar has. Next, the FourSquare likes data set was split into training and test sets for performing linear regression. Roughly, 80% of the bars were included in the training set and 20% were included in the test set.

Linear regression was chosen as a more definitive correlation analysis, compared to scatter plot visualization, due to its simplicity and the fact that it is less subjective since calculations can be performed to determine its accuracy. It is easy to interrupt the results of a linear regression analysis, as well as to implement it using the sklearn python library, which is the tool that was used for developing a linear regression model in this study. To evaluate the accuracy of the linear regression model in this study, the Mean absolute error, Residual sum of squares (MSE), and R2-score were also calculated.

4. Results

Data Exploration: Scatterplot Visualization

The scatter plot visualization of the Foursquare likes data is shown in Figure 3. Shown on the X axis is the Bar Status of the Philadelphia bars (i.e whether the bar advertises Saturday Specials on thedrinknation.com or not), and shown on the y axis is the number of Four Square likes.
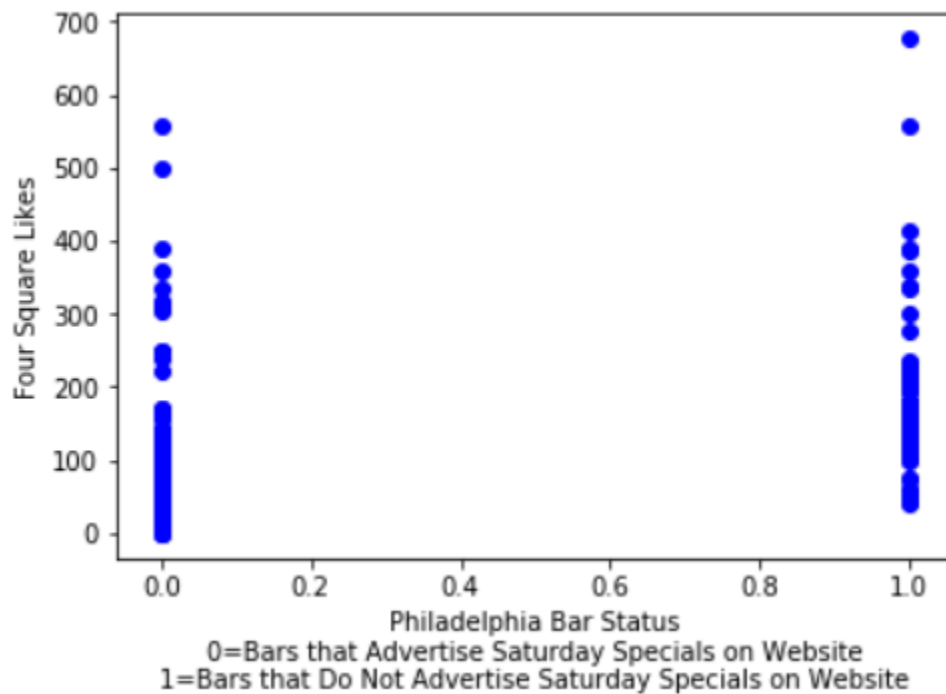

Figure 3. Scatterplot Visualization Comparing the Four Square likes of Philadelphia bars that advertise Saturday Specials on thedrinknation.com vs. Philadelphia bars that do not.

Developing a Linear Regression Model

The linear regression model was developed using the training set (~80% of the data randomly chosen from the data visualized in Figure 3). The coefficient and intercept were 90.65099458

and 101.28571429, respectively. A visualization of the fit line derived from this analysis is shown in red in Figure 4.
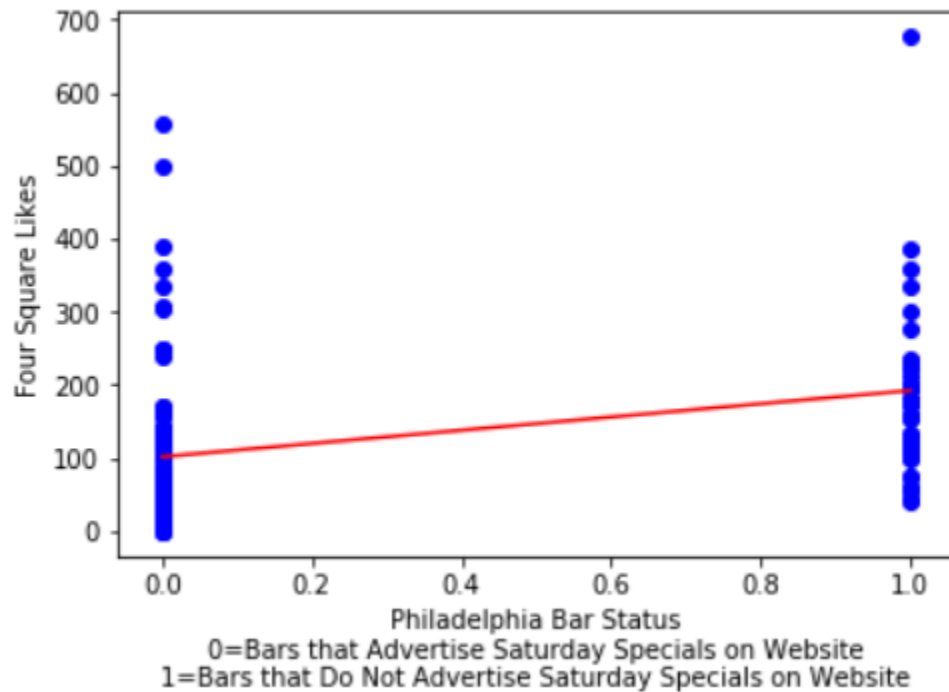

Figure 4. Visualization of the linear regression fit line derived using the training set.

The test set (the remaining data, ~20% of the data from Figure 3, was used to predict number of Four Square likes. The accuracy of the linear regression model was evaluated using three calculations: Mean absolute error, Residual Sum of Squares (MSE), and R^2- score. The values for each of these parameters is found in Table 1.

Table 1. Results of Calculations for evaluating accuracy of linear regression model.

Accuracy Calculation	Value
Mean absolute error	67.13
Residual sum of squares (MSE)	10571.03
R^2-score	-4.22

5. **Discussion**

Just by looking at the initial scatter plot visualization in Figure 3, it is apparent that Philadelphia bar status (those that advertise Saturday specials on thedrinknation.com vs those that do not) does not predict number of Four Square likes. This is because the spread of the data for both groups appear very similar. If Philadelphia bar status did correlate with Four Square likes, one would expect to see many more bars with a higher number of likes for the bars that advertise Saturday specials group and a higher number of bars with a lower number of likes in the group of bars that do not advertise Saturday specials on thedrinknation.com.

To confirm these observations, however, a linear regression analysis was performed (Figure 4). The R^2-score of the model was -4.22. The R^2-score signifies how close the data are to the fitted regression line, with the higher the score, up to a value of 1, the closer the data are to the fitted regression line. That the R^2-score of the model is this study was negative shows how unfit the data are to the fitted regression line. The fact that the data are far away from the fit line confirms that Philadelphia bar status does not predict number of Four Square likes.

Using a data set that isn't restricted to the 50 call limit imposed by Four Square, as well as including likes from multiple sources, as there is inherent bias when only one source is used, may improve the model. Furthermore, how long the bars have been open was not taken into account and is a limitation of the study given that a bar open for less time is more likely to have less likes than a bar that is open for a long time. In a future study, it is advised to consider only bars that have been open for a similar length of time. Additionally, other predictive algorithms could be tested as well in a future study, one of them potentially being decision tree analysis.

6. **Conclusion**

ABC brewing cannot use this analysis to support their hypothesis that posting Saturday specials on thedrinknation.com will increase customer satisfaction as there was no correlation between Philadelphia bar status and Four Square likes. This is supported by the fact that the linear regression model performed very poorly (i.e it was not accurate and the data was far away from the fitted regression line). This does not necessarily mean that posting Saturday specials on thedrinknation.com does not increase customer satisfaction, as there were several limitations to the dataset used as an indicator of customer satisfaction.