Caffeine Form Data Analysis Project

Objective: Assist Caffeine Form (CF) with choosing the best brick and mortar shops to pair with to sell their products directly, branching out from purely online sales. CF believes that choosing the store with the most reviews will help them with this goal. They would like to understand the types of shops and if that influences the number of reviews as well.

Question: What are the most common types of shops and which types of shops get the most reviews?

Dataset
Original dataset has 9 columns and 201 rows, including header. Data validation was completed:
- Region, A-J, 10 unique values, no blanks found
- Place Name, variable string, no blanks found, 177 unique values after removing outliers
- Place Type, "Cafe" "Coffee shop" "Espresso bar" "Others", 4 unique values, no blanks found
- Rating, 5 point scale with 1 decimal place, 2 blanks were replaced with "0" in row 51 and 188, 13 unique values
- Reviews, integer of count of reviews, 2 blanks replaced with median value of "271.5" in rows 51 and 188, 156 unique values
- Price, "$" "$$" "$$$", no blanks found, 3 unique values
- Delivery Option, TRUE or FALSE, no blanks found, 2 unique values
- Dine In Option, TRUE or FALSE, missing values replaced with FALSE: a total of 60 rows affected, 2 unique values
- Takeout Option, TRUE or FALSE, missing values replaced with FALSE: a total of 56 rows affected, 2 unique values

After outliers were removed 193 rows (including header) remained.

Data:

What is the most common type of store?

Coffee Shops are most common.
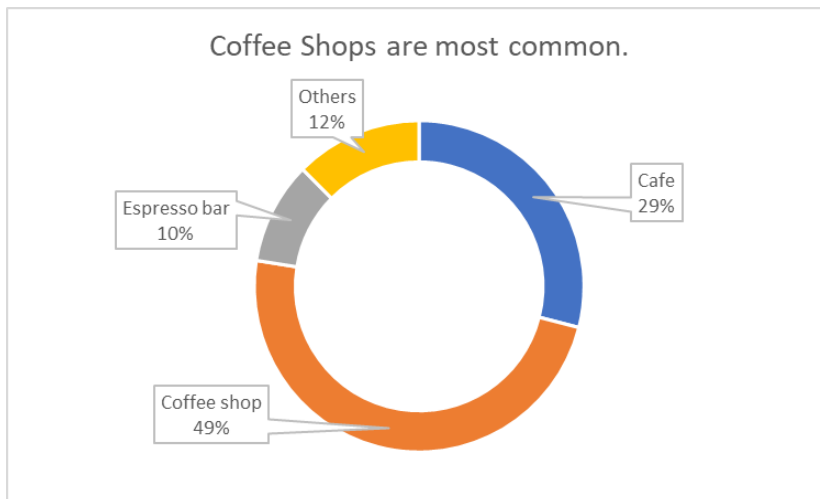
Type of Store	Number	Percent
Cafe	58	29%
Coffee shop	97	49%
Espresso bar	20	10%
Others	25	13%
Grand Total	200	100%

Coffee shops are the most common type of shop, with cafes coming up next. Coffee shops are nearly 50% of the total types of shops. This agrees with the background information that CF currently sells to coffee shops through their online platform. The observations are not equal due to the fact that CF currently seeks out coffee shops to sell their products.

What is the distribution of the number of reviews?

Initial Descriptive Statistics:

Number of Reviews	
MIN	3
MAX	17937
MODE	10
MEAN	618.985
STND DEV	1390.7908

Due to the mode being extremely far from the min and max values, there are likely outliers in this review count.

The z-score was calculated for each data point. The threshold for finding outliers was set at Z-score > 3 or Z-score < -3. A total of 3 iterations of the test was intended to be run.

- Lviv Coffee had a z-score of 12.45 with 17937 reviews, and therefore was removed from the dataset.

Another iteration of the test was ran and the following were removed:

- Traveler's Coffee zscore of 3.308 with 2700 reviews
- Svit Kavy zscore of 3.66 with 2931 reviews
- Вірменка zscore of 3.57 with 2873 reviews
- Kofeyin zscore of 3.63 with 2914 reviews

A final iteration of the test was ran and the following were removed:

- Lviv Croissants zscore of 2.06 with 2236 reviews
- ONE LOVE espresso bar zscore of 3.21 with 2319 reviews
- DoubleDecker Cake and Coffee zscore of 3.18 with 2381 reviews

Final Descriptive Statistics:

Number of Reviews	
MIN	3
MAX	2141
MODE	10
MEAN	455.76042
STND DEV	526.37536

Few Places Get Thousands of Reviews
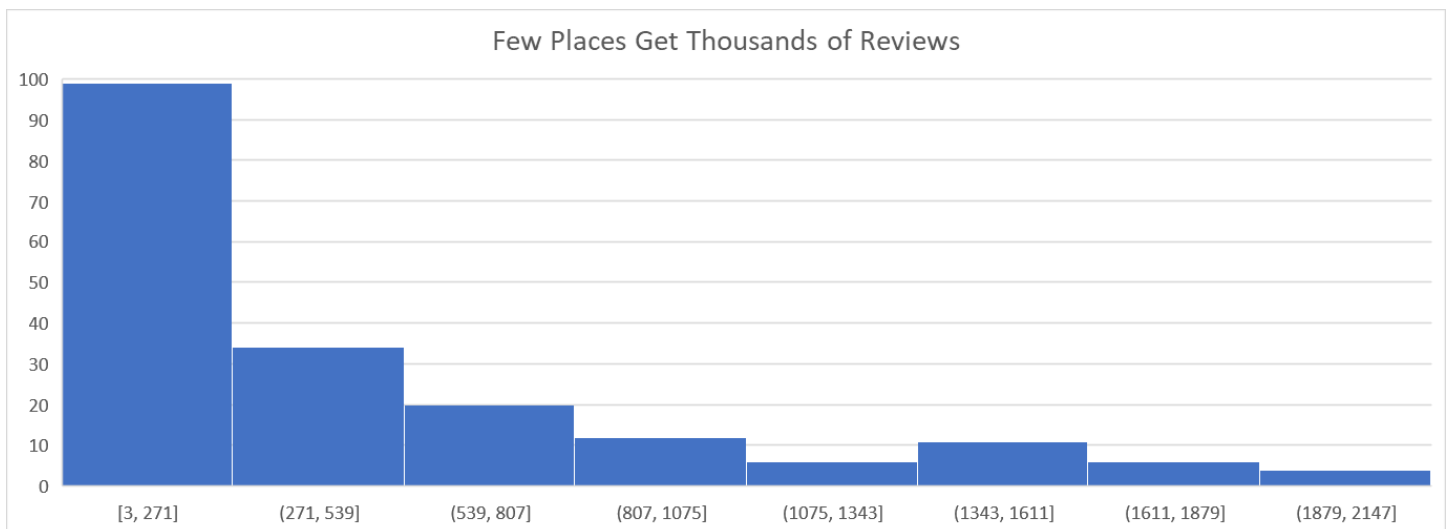
The overall distribution is right skewed. Most of the shops have less than 272 reviews, while very few have review counts leading into the thousands.

What is the relationship between type of store and number of reviews?
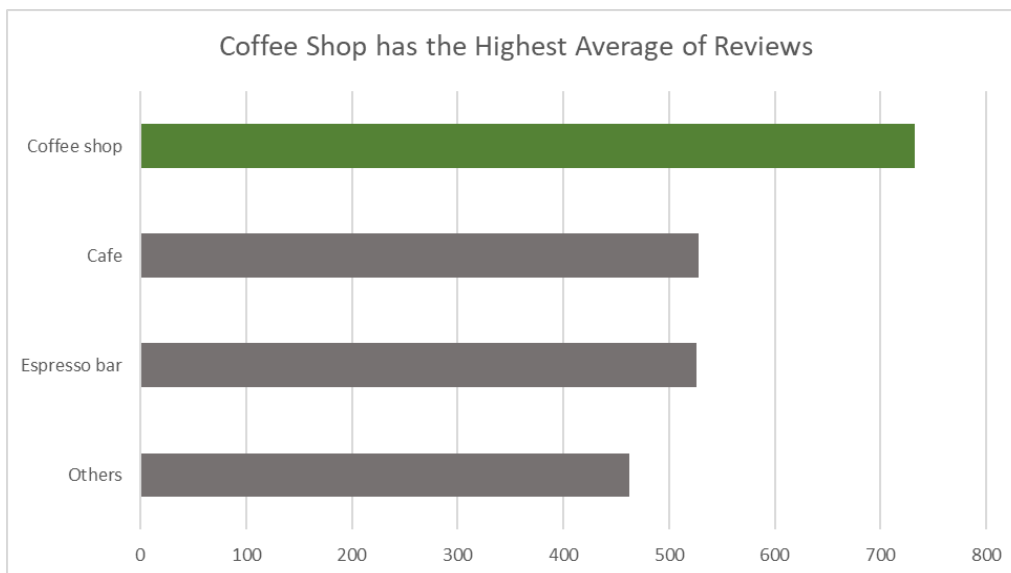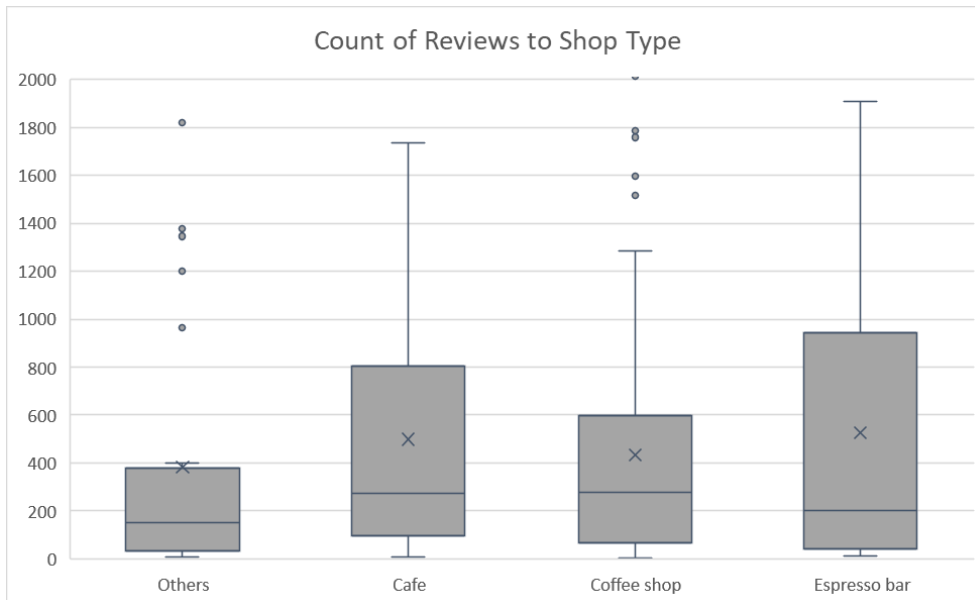
Coffee shops have on average the highest number of reviews counted. This is likely due to the average being pulled up by outliers that we can see in the chart above. Espresso bars have the strongest amount of reviews. This is due to the 3rd IQR being so large, with no outliers.

Final Thoughts:
Based on the analysis above, I suggest Caffeine Forum should focus on coffee shops with large (>1000) reviews to start. This is the most common type of store, and it has a significant amount of reviews. Espresso bars and cafes with large amounts of reviews could be targeted as well.