

## COFFEE SHOPS\_YELP REVIEWS

### RELEVANT JMP PLATFORMS AND STATISTICAL TECHNIQUES

Distribution :	Bar Chart and Frequency Tables
Tabulate :	Tables of Summary Statistics
Text Explorer :	Analysis of Unstructured Text
Graph Builder :	Scatterplot

### PROBLEM STATEMENT

Reviewing customers' experience of cafes and restaurants has become a very common way in which people decide upon which establishment they should go to. Yelp has become one of the most popular platforms used to post and read such reviews. As part of an effort to understand residents' and visitors' experience and sentiment of coffee shops in Austin, TX , a dataset was gathered of 4,466 reviews of 72 establishments [1].



Some goals of the analysis will be to: 1. Develop a general assessment of how patrons feel about these establishments ; 2. Identify common attributes patrons comment on and positive/negative sentiment associated with them ; 3. Identify those coffee shops that receive some of the best/worst rating uncovering some reasons for their rating ; 4. Develop a thorough understanding of the experience and sentiment patrons have of a specific coffee shop.

## DATA SET

### COFFEE\_SHOPS\_YELP\_REVIEWS.JMP

ID	ID of the reviewer
Coffee Shop	Name of the Coffee Shop
Rating	Numerical rating (1,...,5) provided by reviewer
Rating (Ordinal)	Rating (1,...,5) as an Ordinal variable
Reviews	Written reviews posted to Yelp

## EXERCISES

1. Create a Bar Chart and Frequency Table of the Rating. Provide a brief sentence or two describing what is learned from these results. *Hint: Use Distribution platform.*
2. Create a Bar Chart of the Avg Rating for each coffee shop.  
*Hint: Use Graph Builder. Put 'Rating' on the X axis., 'Coffee Shop' on the Y axis, and select Bar Chart Graph. Right click on the Y axis and choose Order > 'Rating' ascending. Choose to 'Label by Value'.*
  - a. How many coffee shops have an average rating of 4.5 or higher?
  - b. How many have an average rating of 4.0 or lower?
3. To see how many reviews each coffee shop received, create a Bar Chart displaying the number of rating for each.  
*Hint: Use Graph Builder. Put 'Coffee Shop' on the Y axis, and select Bar Chart Graph. Right click on the Y axis and choose Order > 'Count' ascending. Choose to 'Label by Value'.*
4. Examine the relationship between the Number of Reviews a Coffee Shop reviews and their Rating via the following set of analyses.
  - a. Create a Table showing the number of reviews and the average rating for each coffee shop. *Hint: Use Analyze > Tabulate. Place the statistics 'N' and 'Mean' in the 'Resulting Cells' zone. Place 'Rating' in the 'Resulting Cells' area. This small table shows the total number of reviews and the average rating for all those reviews. Finish by placing the variable 'Coffee Shop' in the Row Drop Zone.*
  - b. Create a Scatterplot of the Average Rating by Number of Reviews. *Hint: In the table created in a. above, choose 'Make into Data Table' under the Red Triangle. From this new data table, use Graph Builder to create a Scatterplot of 'Mean(Rating)' by 'N(Rating)'.*
  - c. Does there appear to be some level of a correlation between these two variables? In other words, do shops that get the highest and lowest average rating tend to have more/less reviews? Provide a brief sentence or two on what this might mean.
5. Analyze the written reviews by conducting the following set of analyses.

- a. Generate a Table and Bar Graph of the most frequently used words and phrases.  
*Hint: Use Analyze > Text Explorer.*
- b. As many of the words/phrases used in the reviews don't provide any actual assessment of how people feel about these establishments, it's important in analyzing text data to focus on only those words/phrases that do. As a result, a common first step in analyzing text data is to remove those noninformative words. *Hint: Select those noninformative words and phrases in the tables, right-click and choose 'Add Stop Word'. An example would be the most common word that occurs – "coffee". Clearly not helpful in understanding sentiment as reviewers use this word simply because they are reviewing coffee shops. Choose to do this only to words/phrases that occur at least 750 times.*
- c. Examine the top ten most frequently used words? What does this reveal about the themes patrons frequently comment on in their reviews?
- d. Examine a table of these words/phrases for each Rating? *Hint: Choose Local Data Filter under the Red Triangle and choose the variable 'Rating (Ordinal)'.*  
What are the top 3-5 themes reviewers comment on for those that give a rating of a '1'?  
What are the top 3-5 themes reviewers comment on for those that give a rating of a '5'?  
Comment on similarities and differences observed.
- e. Examine the themes most frequently commented on for 2-3 of the coffee shop with the highest average rating and 2-3 with the lowest. *Hint: Click on the 'And' button in the Local Data Filter and add 'Coffee Shop'. Now select the desired Coffee Shop. The analysis will rerun using only the reviews from the chosen shop.*  
What are reviewers predominantly commenting on in each of these shops and what insights does this provide regarding what patrons are and/or are not happy about?  
*Hint: To see the whole review for a given word, right-click on that word in the Terms and Phrases List table and choose 'Show Text'.*
6. Perform a detailed analysis of the reviews for one coffee shop (choose one not examined thus far and with at least 50 reviews). Provide a few brief paragraphs summarizing some of the key findings.  
*Hint: What are some things patrons are happy and not happy about with this coffee shop? How do they feel about the quality of the coffee, food, vibe, service, ambiance? Provide some recommendation on ways they can improve.*

## COMPLIMENTARY MATERIALS

1. Data is derived from data.world 'Austin Coffee Yelp Reviews'.  
<https://data.world/rdowns26/austin-coffee-yelp-reviews>