

MIS 0855 – Data Science (Section 006) – Fall 2017

In-Class Exercise (Day 34) – Twitter Sentiment Analysis

Objective: Understand the function and the utility of Twitter sentiment analyses

Learning Outcomes:

- Perform a sentiment analysis of a Twitter stream using an Excel function
- Uncover the patterns in customer sentiments over Twitter

In this exercise, we will use tweets posted by customers of two airlines – United (@united) and Southwest (@SouthwestAir). Twitter users can post tweets about a company by mentioning the company’s Twitter handle (e.g. @companyname). In this way, customers can express their sentiments over the company; in other words, they can complain or compliment about it. We will use a sentiment analysis tool in Excel to analyze the sentiment expressed via tweets.

@united You idiots are the worst airline ever. You don't care about your passengers! Don't fly this garbage airline.

10:27 AM - 6 Nov 2017

Loved using my companion pass to board with my husband today! Thanks Southwest Airlines for trying this program. Keep it going!
@SouthwestAir

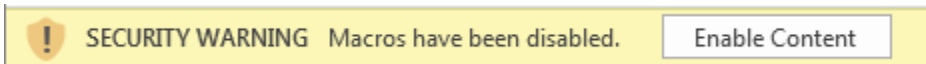
Part 1: Understand the Data

- 1) Download “United Southwest Mentions.xlsx” file from the class site and open it.
- 2) This file has two worksheets. “United” worksheet contains 3,789 tweets that mention @united. This data is collected via Microsoft Flow on November 3-6, 2017. It also includes tweets, locations where each tweet was posted, account names, and date/time of tweets.
- 3) Likewise, “Southwest” worksheet includes 4,108 tweets that mention @SouthwestAir in the same period.

Part 2: Analyze Sentiment Using Excel

- 1) Download “Sentiment Analysis Tools.xlsx” file from the class site and open it.
- 2) Open the Sentiment Analysis Tools workbook. Select the “Sentiment Analysis - United” tab.

3) If you see this warning in Excel:

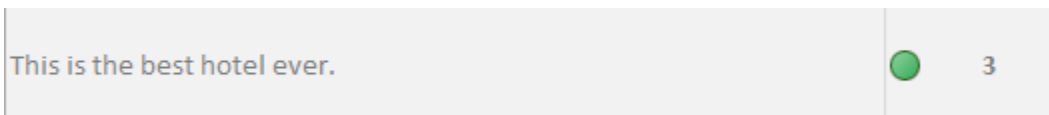


Click “Enable Content”

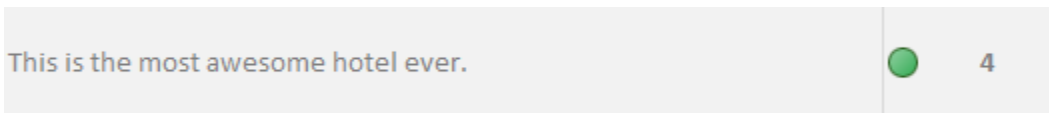
NOTE: This spreadsheet has some embedded code that computes average sentiment using a dictionary of positive and negative words to analyze the text.

It gives an overall score from -5 (extremely negative) to 5 (extremely positive) based on (1) the frequency of positive versus negative words in the text and (2) the strength of the feeling conveyed by those words. For example, “great” is a +3, “amazing” is a +4, and “sucks” is a -3.

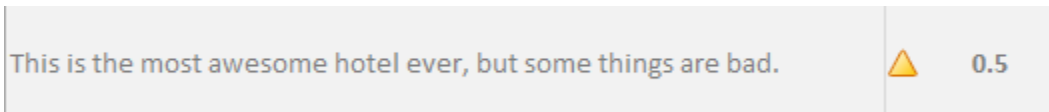
4) To see how it computes sentiment, look at the first sample comment:



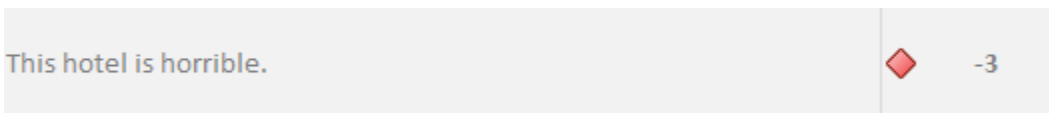
Now change the text to: “This is the most awesome hotel ever.”



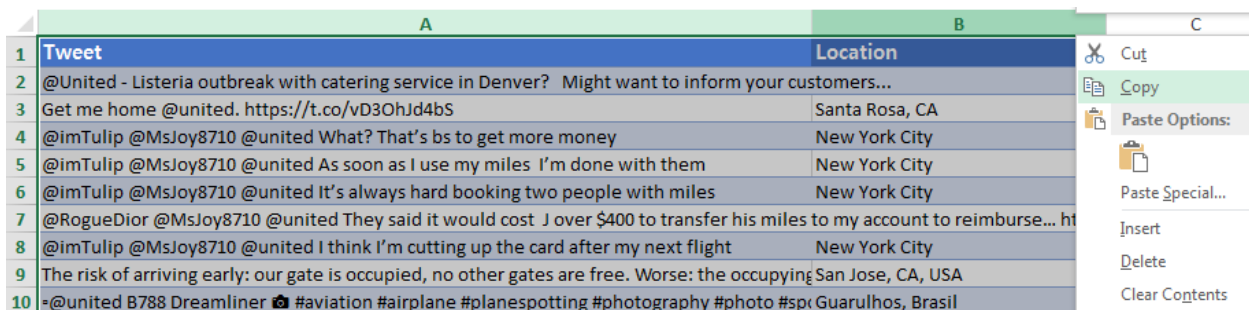
Now change the text to: “This is the most awesome hotel ever, but some things are bad.”



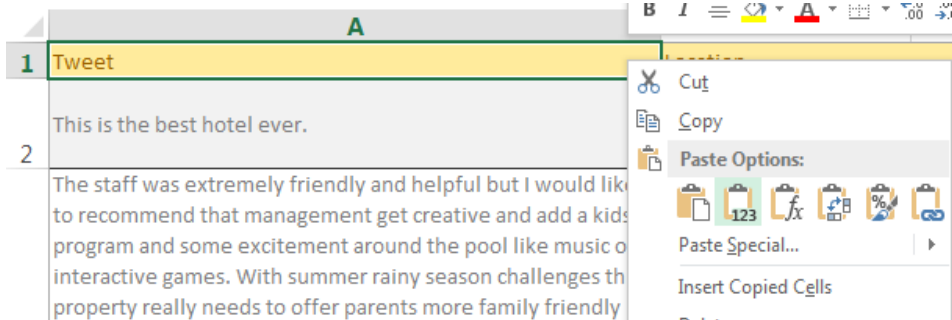
Finally, change the text to: “This hotel is horrible.”



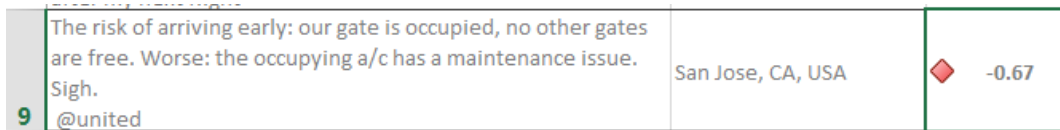
5) In “United Southwest Mentions.xlsx,” from “United” worksheet, select the entire column A and B, right-click them, and click Copy.



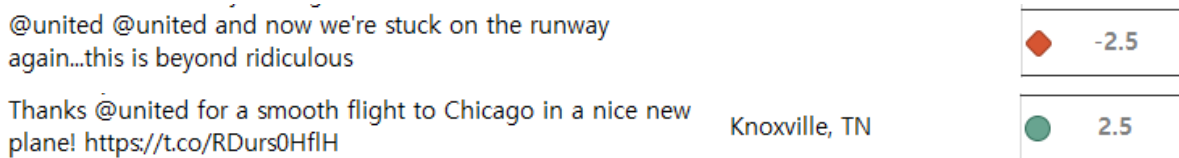
- 6) Switch back to the Sentiment Analysis spreadsheet, and on “Sentiment Analysis – United” worksheet, click Cell A1 and paste the text.



- 7) You’ll notice that it only computed the scores for the first 8 tweets. To compute the scores for the rest of the tweets, move the cursor over the bottom-right corner of Cell C9 (-0.67) and double-click it.



- 8) It might take a few minutes for Excel to calculate the sentiments of all 3,789 tweets.
 9) Look at the text for each tweet (Column A) and the score (Column C). In some cases, it will categorize some tweets in the way you expect. In other cases, you might disagree with some of the score.



- 10) If you want to compute the average sentiment for all of your tweets, click on Cell F2 and type this formula:

=AVERAGE(B:B)

A number greater than 0 means that average sentiment was positive, and a number less than 0 means that average sentiment was negative. Of course, the value matters too. An average value of 0.5 basically means sentiment was either neutral, or there were equal numbers of positive and negative tweets.

- 11) You could also count how many positive comments there were. In Cell F3, type this formula:

=COUNTIF(C:C,">=2")/COUNT(C:C)

This shows the number of the positive tweets (ones that have a sentiment score of 2 or greater) divided by all tweets.

Likewise, in Cell F4, type this formula:

$=\text{COUNTIF}(C:C, "<=-2")/\text{COUNT}(C:C)$

This will calculate the number of the negative tweets (ones that have a sentiment score of -2 or lower) divided by all tweets.

Likewise, in Cell F5, type this formula:

$=\text{COUNTIF}(C:C, "<=-2")/\text{COUNTIF}(C:C, ">=2")$

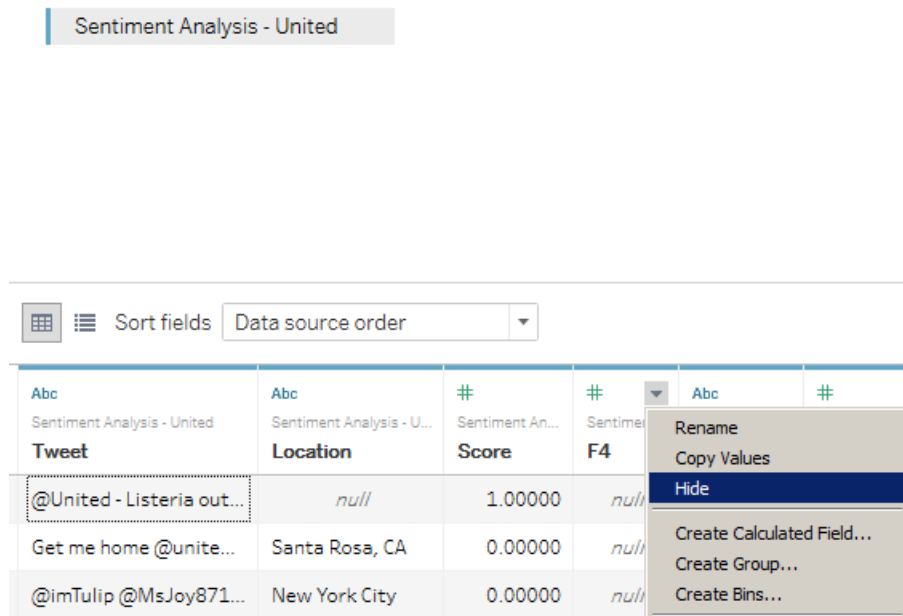
which compares the number of the negative tweets to the positive tweets.

12) Do the same with Southwest tweets on “Sentiment Analysis – Southwest” worksheet. For which airline (United or Southwest) customers are more positive?

13) Save the Excel file.

Part 3: Visualize the Twitter Sentiments

- 1) Start Tableau and connect to your “Sentiment Analysis Tools.xlsx” file.
- 2) Select “Sentiment Analysis – United” worksheet for analysis.
- 3) We only need Tweet, Location, and Score columns for analyses. Let’s hide all the other columns (F4-F8).



4) Go to Sheet 1.

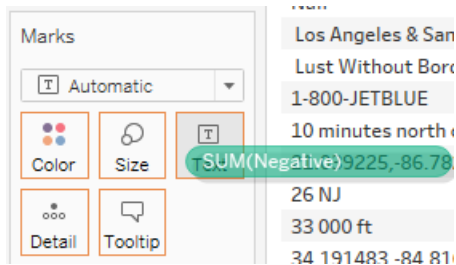
- 5) In this analysis, we will focus on negative tweets, to which companies should pay more attention than to positive ones. For that, create a calculated field as follows.

IIF([Score]<=-2, 1, 0)

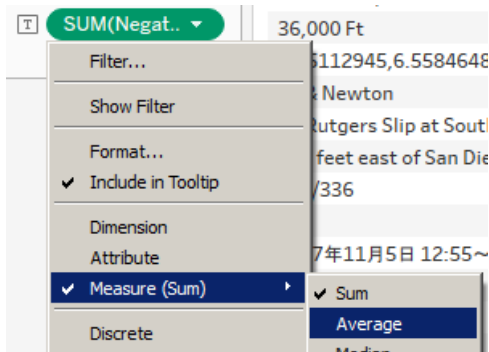


This will create a measure which is 1 if the sentiment score is lower than or equal to -2 and 0 otherwise.

- 6) Drag Location to Columns. Click “Add all members.”
7) Drag Negatives to Text.

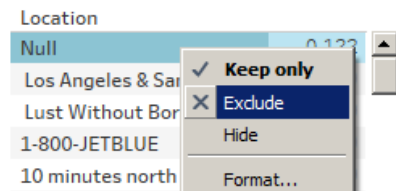


- 8) Change SUM to Average. This will show the share of negative tweets in each location (city).

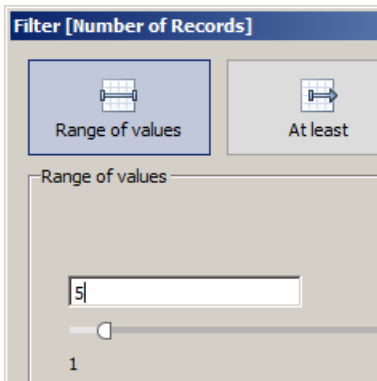


9) Let's exclude "Null" on the first row, which are tweets without location information.

Sheet 1



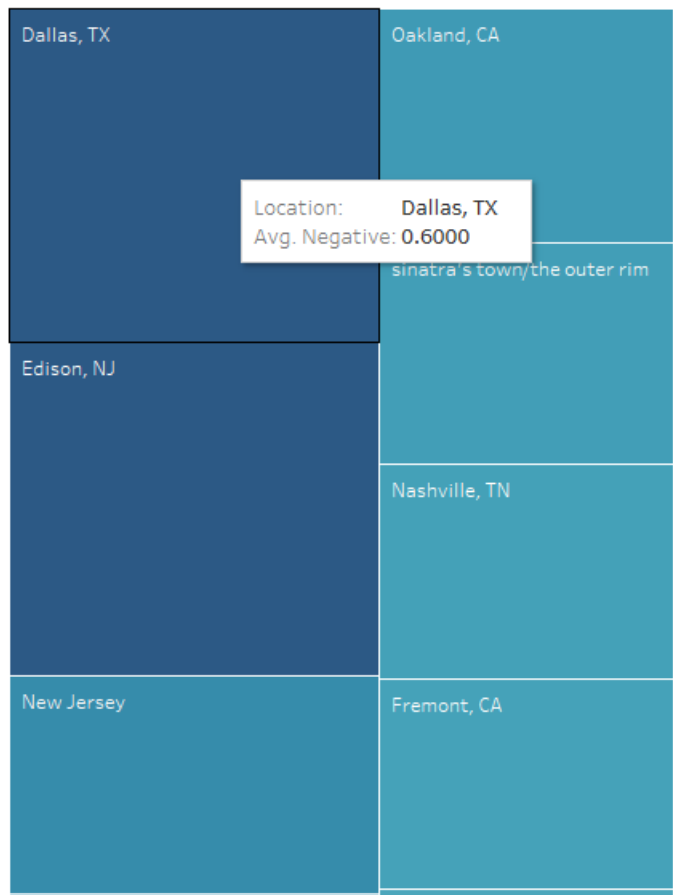
10) Let's also exclude locations where too few tweets were posted. Drag Number of Records to Filters. Select Sum, and set the minimum value to 5.



This will exclude the locations where less than four tweets were posted.

11) Under Show Me, click Treemaps.





12) This Treemap shows that customers in Dallas, Edison, and New Jersey have most negative sentiments over United.

13) Create a similar Treemap with the Southwest tweet analysis.

Part 4: Do It Yourself.

Create a U.S. city map to visualize the average score of tweet sentiments for United. Visualize negative scores with red and positive scores with blue.

