ISCS 540 – Big Data Analytics, Spring 2022
Group Project – Deliverable 3: Simple Data Analysis and Visualization in Tableau

For this deliverable of the group project, you will use Tableau Public to perform simple data analysis and visualization on the data collected about your topic. Almost all of the tasks for this assignment have been covered previously in in-class assignments.

You will be submitting your group's screenshots for this deliverable in a single Word document named **GroupName_Deliverable3.docx**

Following are the requirements for this deliverable:

1. In RStudio, export your data frame containing the all of your tweets that have been cleaned (with the additional day/date columns added) to a tab separated value file. Make sure this data frame **does not have all tweets without a location removed** (one of the steps in Group Project Deliverable 2). You can use a command similar to:
   ```
   write.table(your_tweet_df, "filename.txt", sep = '\t', row.names = F)
   ```

   Note that exporting the data and loading it into Tableau may be time consuming.

2. Start Tableau Public (Desktop Edition). Under Connect on the left pane, select "Text file". Navigate to the location where you saved your file and select it to import your data.

3. For this step, you will view the distribution of tweets you collected on a world map. Create a new sheet called **"symbol map"**. Add the **Country** dimension to the columns (Add all members), and add the **Number of Records** (filename.txt(Count)) measure to the rows. From the Show Me menu in the top right corner select symbol maps. In the Marks box, Ctrl + drag (Cmd + drag for Mac users) CNT(filename.txt) to the Color option. Take a screenshot and add it to your Word document.
   [Screenshot 1]

4. Create a duplicate of the current sheet by right-clicking your sheet's tab, Duplicate. Rename the new sheet to **"map"**. Select the map option from the Show Me menu. Take a screenshot and add it to your Word document.
   [Screenshot 2]

5. For this step, you will view the distribution of tweets you collected by day and hour in area charts. Create a new sheet called **"area discrete"**. Click the down arrow for the **Created At** dimension, select Change Data Type, then select Date & Time. Add the **Created At** dimension to the columns (Add all members), and add the **Number of Records** (filename.txt(Count)) measure to the rows. From the Show Me menu in the top right corner select **area charts (discrete)**. Click the down arrow for YEAR(Created At), go to More and select **Weekday**. Take a screenshot and add it to your Word document.
   [Screenshot 3]

Click the down arrow for WEEKDAY(Created At), go to More and select **Hour**. Take a screenshot and add it to your Word document.

From the list of dimensions, drag the Country dimension to Color under Marks. Exclude Null, NA, and any other values that should not be considered countries. Take a screenshot and add it to your Word document.

6. For this step, you will view the distribution of number of tweets based on usernames. Create a new sheet called **"users"**. Add the **ScreenName** dimension to the rows. When you get a warning popup select "Add all members". Add the **Number of Records** (filename.txt (Count)) measure to the columns. Click the sort button next to the Number of Records axis title to sort in descending order by number of tweets. Exclude Null. Take a screenshot and add it to your Word document.

7. For this step, you will view the number of tweets by language. Create a new sheet called "**treemap**". Add the **Lang** dimension to Color under Marks. Add the **Number of Records** (filename.txt (Count)) measure to Size under Marks. In the Marks box, Ctrl + drag (Cmd + drag for Mac users) Lang to the Label option to add the tweet language as a label to the graph. Exclude Null, NA, and all non-language values. It may be easiest to exclude all, then re-add the two-letter language values. Take a screenshot and add it to your Word document.

8. Convert the **Followers Count** dimension to a measure. Change the data type of the Followers Count to **Number (whole)**. Remove the **Lang** dimension from Color under Marks and add the **Followers Count** measure to Color instead to color the graph by Followers Count. Take a screenshot and add it to your Word document.

9. Convert the **Favorites Count** dimension to a measure. Change the data type of the Favorites Count to **Number (whole)**. Remove the **Followers Count** measure from Color under Marks and add the **Favourites Count** measure to Color instead to color the graph by Favourites Count. Take a screenshot and add it to your Word document.

10. For this step, you will view the distribution of tweets by day and hour in bar charts. Create a new sheet called "**barchart**". Add the **Created At** dimension to the columns, and add the **Number of Records** (filename.txt (Count)) measure to the rows. Click the down arrow for YEAR(Created At), go to More and select Weekday. From the Show Me menu in the top right corner select horizontal bars. Exclude Null. Take a screenshot and add it to your Word document.

11. Click the down arrow for WEEKDAY(Created At), go to Exact Date. Drag the **Created At** measure to Label under Marks. Click the down arrow for the Created At label and select Day. This will add labels indicating the days on which the tweets were streamed. Take a screenshot and add it to your Word document.
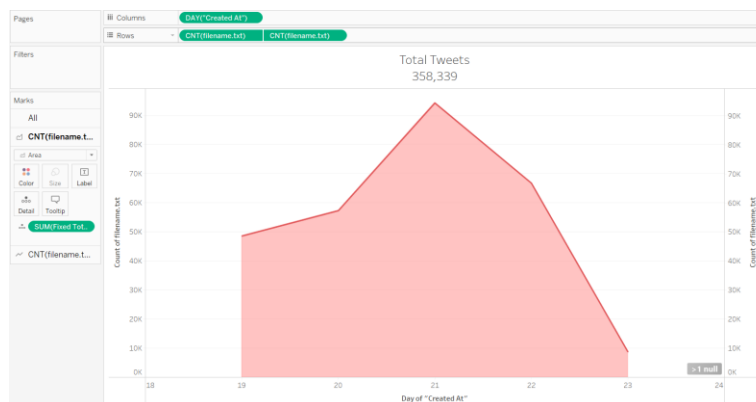   [Screenshot 11]

12. Create a **Calculated Field** called **Fixed Total Tweets**. This will be a fixed field that will display the total number of tweets in your data file. You can use the following calculation:
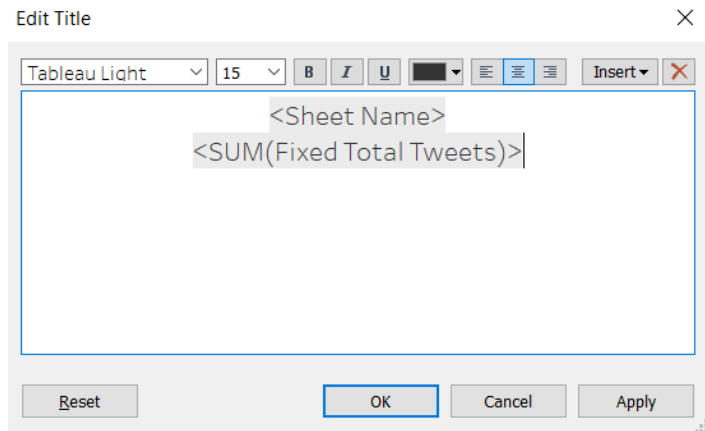   `{ FIXED:COUNT([filename.txt])}`

Take a screenshot of your Calculated Field edit window and add it to your Word document.
[Screenshot 12]

Drag the **Created At** dimension to columns, and the **Number of Records** (filename.txt (Count)) measure to rows. Drag the **Number of Records** (filename.txt (Count)) measure to rows again so that you have it repeated twice. Click on the down arrow of the second occurrence and select **Dual Axis**. You now have two charts listed under marks. Set one of them to **Line** and the other to **Area**. Use appropriate colors to make it where both are easily visible (two shades of the same color are fine). Your spreadsheet should look similar to the image below.

Edit the title of the sheet by double clicking on it. Under the sheet name placeholder click the Insert drop-down menu and select SUM(Fixed Total Tweets).

Edit Title   ✕

| Tableau Light | ⌄ | 15 | ⌄ | **B** | *I* | U | ⬛ ▾ | ☰ | ☰ | ☰ | Insert ▾ | ✕ |

<Sheet Name>
<SUM(Fixed Total Tweets)>

| Reset | | | OK | Cancel | Apply |

Take a screenshot of your Edit Title window and add it to your Word document.
[Screenshot 13]

Take a screenshot of the Total Tweets spreadsheet and add it to your Word document.
[Screenshot 14]

Submit your work in the form of **a Word document** using the following format:
TeamName_Deliverable3.docx
Your Word document should contain 14 screenshots.

Only one person per group needs to submit the zip file.