

Name: Fady Malek Mikhail

ID: 20201470332

Name: Nour Khaled Mohamed

ID: 20201378544

Name: Abdelrahman sabry Ali Ibrahim

ID: 20191480281

Report Social Data Networks (final project)

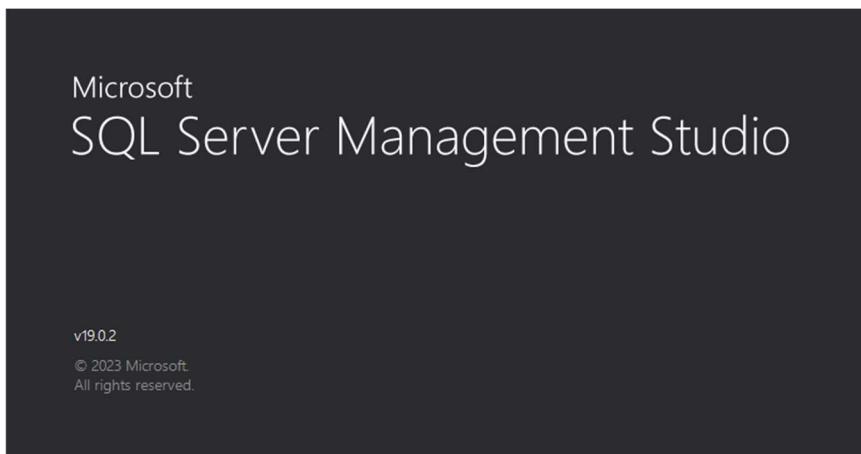
- First, We need to understand the data:

Data contains 15 columns:

- 1) **Tweet_id**: contains the ids of the tweets that we are analyzing (**integer**)
 - 2) **Airline_sentiment**: contains the sentiment of the tweet whether **positive**, **negative**, or **neutral**.
 - 3) **Airline_sentiment_confidence**: contains **the degree of confidence** that these tweet follow these sentiments (**integer from 0 to 1**)
 - 4) **NegativeReason**: contains the reason for the negative responses (**String**)
 - 5) **NegativeConfidence**: contains the degree of confidence that these reasons follow these sentiments (**integer from 0 to 1**)
 - 6) **Airline**: the state which the tweet is talking about (**American – Delta – southwest – united – us airways – virgin America**)
 - 7) **Airline_sentiment_gold**
 - 8) **Name**: name of the user who wrote the tweet(**string**)
 - 9) **NegativeReason_gold**
 - 10) **Retweet_count**: contains number of retweets made by other users on this tweet (**integer**)
 - 11) **Text**: contains the text of the tweet (**string**)
 - 12) **Tweet_coord**: contains **geographical coordinates** where the tweet is written. (like [0,0],[1.3,103.87])
 - 13) **Tweet_created**: contains the **date** and **time** the tweet was created on
 - 14) **Tweet_location**: contains the location which users are tweeting from
 - 15) **Tweet_timezone(user_timezone)**: the time zone which these twitter accounts follow.
-

Build up your data files into a Database; get data from an SQL server; write queries to perform CRUD operations; test data visualizations refreshment

First, we download SQL management Server



SQLQuery1.sql - FADY.Fady (67) - Microsoft SQL Server Management Studio

```
select * from dbo.Worksheet$
```

tweet_id	airline_sentiment	airline_sentiment_confidence	negativereson	negativereson_confidence	airline	airline_sentiment_gold	name	negativereson_gold	retweet_count	text	tweet_coord	tweet_created	tweet_foo
5.70306E+17	neutral	1	null	Virgin America	NULL	jaedino	NULL	0	@VirginAmerica What @shebum said.	NULL	2/24/2015 11:35	NULL	
5.70301E+17	positive	0.3486	null	Virgin America	NULL	yvonnealyne	NULL	0	@VirginAmerica plus you've added commercials to ...	NULL	2/24/2015 11:15	NULL	
5.70301E+17	neutral	0.6837	null	Virgin America	NULL	jaedino	NULL	0	@VirginAmerica I didn't today... Must mean I need...	NULL	2/24/2015 11:15	Lets Play	
5.70301E+17	negative	1	Bad Flight	0.7033	Virgin America	NULL	jaedino	0	@VirginAmerica it's really aggressive to blast abo...	NULL	2/24/2015 11:15	NULL	
5.70301E+17	negative	1	Can't Tell	1	Virgin America	NULL	jaedino	0	@VirginAmerica and it's a really big bummer that it	NULL	2/24/2015 11:15	NULL	
5.70301E+17	negative	1	Can't Tell	0.6842	Virgin America	NULL	jaedino	0	@VirginAmerica I'm gonna have to add a 25% flight for...	NULL	2/24/2015 11:14	NULL	
5.70301E+17	neutral	0.634	null	Virgin America	NULL	ognogains	NULL	0	@VirginAmerica yes, nearly every time I fly Vx this ...	NULL	2/24/2015 11:13	San Franc	
5.70301E+17	positive	0.6559	null	Virgin America	NULL	pilot	NULL	0	@VirginAmerica really missed a prime opportunity...	NULL	2/24/2015 11:12	Los Angel	
5.70295E+17	positive	1	null	Virgin America	NULL	dhepburn	NULL	0	@virginamerica Well, I didn't, but NOW I DO! :-D	NULL	2/24/2015 11:11	NULL	
5.70294E+17	neutral	0.6769	null	Virgin America	NULL	YuptaTate	NULL	0	@VirginAmerica it was amazing, and arrived an ho...	NULL	2/24/2015 10:53	Los Angel	
5.70294E+17	positive	1	null	Virgin America	NULL	idk_but_y...	NULL	0	@VirginAmerica did you know that sometimes the...	NULL	2/24/2015 10:48	1/1 loner t	
5.70287E+17	positive	0.6451	null	Virgin America	NULL	HyperCa	NULL	0	@VirginAmerica It's pretty gracious to see the bet...	NULL	2/24/2015 10:48	NYC	
5.70286E+17	positive	1	null	Virgin America	NULL	moflander...	NULL	0	@VirginAmerica This is such a great deal! Already t...	NULL	2/24/2015 10:30	NYC	
5.70286E+17	negative	0.6842	Late Flight	0.3684	Virgin America	NULL	ajesper...	NULL	@VirginAmerica Thanks!	NULL	2/24/2015 10:15	San Franc	
5.70287E+17	positive	1	null	Virgin America	NULL	smartw...	NULL	0	@VirginAmerica So excited to my first cross count...	NULL	2/24/2015 9:42	palo alto, c	
5.70287E+17	negative	1	Bad Flight	1	Virgin America	NULL	tzbrian...	NULL	@VirginAmerica I flew from NTC to SFO last week ...	NULL	2/24/2015 9:39	west covr	
5.70287E+17	positive	1	Bad Flight	1	Virgin America	NULL	heather...	NULL	0	@VirginAmerica this place i	NULL	2/24/2015 9:39	the place :)

Then we get data from an SQL server

Advanced options

Command timeout in minutes (optional)

SQL statement (optional, requires database)

```
select * from dbo.Worksheet$
```

Include relationship columns

Navigate using full hierarchy

Enable SQL Server Failover support

SQL Server database

Server:

Database (optional):

Data Connectivity mode: Import

DirectQuery

Advanced options

FADY: Projects

tweet_id	airline_sentiment	airline_sentiment_confidence	negativereson	negativereson_confidence	airline	airline_sentiment_gold
5.70306E+17	neutral	1	null	null	Virgin America	
5.70301E+17	positive	0.3486	null	0	Virgin America	
5.70301E+17	neutral	0.6837	null	null	Virgin America	
5.70301E+17	negative	1	Bad Flight	0.7033	Virgin America	
5.70301E+17	negative	1	Can't Tell	1	Virgin America	
5.70301E+17	negative	1	Can't Tell	0.6842	Virgin America	
5.70301E+17	positive	0.6745	null	0	Virgin America	
5.70301E+17	neutral	0.634	null	null	Virgin America	
5.70301E+17	positive	0.6559	null	null	Virgin America	
5.70295E+17	positive	1	null	null	Virgin America	
5.70294E+17	neutral	0.6769	null	0	Virgin America	
5.70294E+17	positive	1	null	null	Virgin America	
5.70294E+17	positive	0.6451	null	null	Virgin America	
5.70287E+17	positive	1	null	null	Virgin America	
5.70287E+17	negative	0.6842	Late Flight	0.3684	Virgin America	
5.70278E+17	positive	1	null	null	Virgin America	
5.70277E+17	negative	1	Bad Flight	1	Virgin America	
5.70271E+17	positive	1	null	null	Virgin America	
5.70268E+17	positive	1	null	null	Virgin America	

The data in the preview has been truncated due to size limits.

Load Transform Data Cancel

Then we open the power Query Editor

The screenshot shows the Power Query Editor interface with a query named "Query1". The data is pulled from a SQL database ("FADV", "Projects", "Query" select * from dbo.Worksheets\$"). The table has 15 columns and 999+ rows. The columns are: tweet_id, airline_sentiment, airline_sentiment_confidence, negativereson, negativereson_confidence, airline, and name. The "APPLIED STEPS" pane shows the "Source" step.

• Usage of powerQuery editor for ETL operands

- 1) Deleting unnecessary columns like **airline_sentiment_gold**, **negativereson_gold** and **tweet_location**

The screenshot shows the Power Query Editor interface with a query named "Tweets". The data is pulled from a table named "#Inserted Absolute Value" with columns: "airline_sentiment_gold", "negativereson_gold", and others. The table has 14 columns and 999+ rows. The columns are: tweet_id, airline_sentiment, airline_sentiment_confidence, negativereson, negativereson_confidence, and others. The "APPLIED STEPS" pane shows the "Removed Columns" step.

Screenshot of the Power Query Editor showing the 'Tweets' query. A context menu is open over the 'tweet_coord' column, specifically targeting the 'Replace Value' option under the 'Text Filters' section. The 'Replacer.ReplaceValue' dialog is displayed, showing the 'Value To Find' field set to "" and the 'Replace With' field set to "[0.0,0.0]". The 'Advanced options' section is collapsed. The right pane shows the 'PROPERTIES' and 'APPLIED STEPS' sections, which include steps like 'Replace Value1' and 'Replace Value2'. The status bar at the bottom right indicates 'PREVIEW DOWNLOADED ON THURSDAY'.

2) Replacing nan values of column tweet_coord with [0.0,0.0]

Screenshot of the Power Query Editor showing the 'Tweets' query. A context menu is open over the 'tweet_coord' column, specifically targeting the 'Replace Value' option under the 'Text Filters' section. The 'Replacer.ReplaceValue' dialog is displayed, showing the 'Value To Find' field set to "" and the 'Replace With' field set to "[0.0,0.0]". The 'Advanced options' section is expanded, showing the 'Replaced Value' step. The right pane shows the 'PROPERTIES' and 'APPLIED STEPS' sections, which include steps like 'Replace Value1' and 'Replace Value2'. The status bar at the bottom right indicates 'PREVIEW DOWNLOADED ON THURSDAY'.

3) And in negative_reason and user_time_zone by not defined

The screenshot shows the Power Query Editor interface with the 'Queries [1]' pane on the left containing a single query named 'Tweets'. The main area displays a table with 14 columns and 999+ rows. A context menu is open over the 'negative_reason' column, with the 'Replace Values' option selected. The 'Replace Values' dialog box is centered, showing the 'Value To Find' field empty and the 'Replace With' field containing 'Not Defined'. Below the dialog, a preview pane shows the first few rows of the table.

4) Splitting the tweet_created column into date,time and id then deleting the id column as it's the same for all rows so it's not needed

The screenshot shows the Power Query Editor interface with the 'Queries [1]' pane on the left containing a single query named 'Tweets'. The main area displays a table with 13 columns and 999+ rows. A context menu is open over the 'tweet_created' column, with the 'Split Column by Delimiter' option selected. The 'Split Column by Delimiter' dialog box is centered, with 'Space' specified as the delimiter and 'Each occurrence of the delimiter' selected. The 'OK' button is highlighted. The right side of the screen shows the 'Properties' and 'Applied Steps' panes for the 'Tweets' query.

The screenshot shows the Power Query Editor interface with the following details:

- File**, **Home**, **Transform**, **Add Column**, **View**, **Tools**, **Help** menu items.
- Toolbars with icons for Transpose, Data Type: Date, Replace Values, Unpivot Columns, Merge Columns, Split Column, Text Column, Number Column, Date & Time Column.
- Queries [1]** pane showing a single query named "Tweets".
- Table** view showing the data structure with columns: tweet_id, tweet_coord, tweet_created_date, tweet_created_time, user_timezone, and tweet_location.
- APPLIED STEPS** pane listing steps like Source, Navigation, Inserted Absolute Value, Removed Columns, Filtered Rows1, Removed Columns1, Trimmed Text, Replaced Value, Split Column by Delimiter, Changed Type, and Renamed Columns.
- Bottom status bar: 15 COLUMNS, 999+ ROWS, PREVIEW DOWNLOADED ON THURSDAY.

5) Relevant DAX expressions; preferably create a measures table including all Measures

1) Creating new column ID that contains the ids without any negative values.

The screenshot shows the Power BI Data View with the following details:

- Structure**, **Formulas**, **Properties**, **Sort**, **Groups**, **Relationships**, **Calculations** tabs.
- Calculated Column** named "ID" with the formula: `= IF(SIGN(Tweets[tweet_id])=-1,Tweets[tweet_id]*-1,Tweets[tweet_id])`.
- Data** pane showing the table structure with columns: text, tweet_coord, tweet_created_date, tweet_created_time, user_timezone, and ID.
- Measures** pane listing various measures such as ID, name, negativerason, negativerason_confidence, retweet_count, text, tweet_coord, tweet_created_date, tweet_created_time, tweet_id, user_timezone, airline, airline_sentiment, and airline_sentiment_confidence.

2) Creating new tables for all columns after cleaning to make it easier for visualization.

text	tweet_coord	tweet_created_date	tweet_created_time	user_timezone	ID
Jr wait times of 2 hours to talk with an exec pt agent, what is going on!!!	[0,0,0]	2/23/2015 12:00:00 AM	12/30/1899 12:02:21 AM	Not Defined	342618111
Jr still waiting on a dm response....#sloooooooverresponses	[0,0,0]	2/23/2015 12:00:00 AM	12/30/1899 12:06:45 AM	Not Defined	1269374976
Jr @Covboyccrone You apologise for HIS frustrations & HIs disappointment but NOT your terrible service? #BlameShiftOverload	[0,0,0]	2/23/2015 12:00:00 AM	12/30/1899 1:24:56 AM	Not Defined	1370034176
Jr love how you can't get an agent on the phone and the automated system hangs up on you	[0,0,0]	2/23/2015 12:00:00 AM	12/30/1899 1:28:17 AM	Not Defined	1646850048
Jr who do I need to talk to to get a full refund on my trip? Your call center offers no assistance.	[0,0,0]	2/23/2015 12:00:00 AM	12/30/1899 2:13:16 AM	Not Defined	1659629568
Jr are you ignoring me on purpose? It was sad enough having your automated system hang up on me ??	[0,0,0]	2/23/2015 12:00:00 AM	12/30/1899 2:14:12 AM	Not Defined	1319886848
Jr waited in the airport for two days to get home, no weather relate flight. Dealt with horrible customer service. Finally switched to	[0,0,0]	2/23/2015 12:00:00 AM	12/30/1899 2:17:21 AM	Not Defined	978628608
Jr @loganex except if you call now you'll get the same message. Trust me, I've tried.	[0,0,0]	2/23/2015 12:00:00 AM	12/30/1899 2:22:08 AM	Not Defined	525643777
Jr @nickcunningham3 Except now there is no wait time, the phone system just hangs up on you. So continuing to hold won't help.	[0,0,0]	2/23/2015 12:00:00 AM	12/30/1899 2:25:29 AM	Not Defined	1366020096
Jr been on hold for over two hours. After waiting 2 hours earlier. Can't get any info on my reservation. How is that ok?	[0,0,0]	2/23/2015 12:00:00 AM	12/30/1899 2:26:37 AM	Not Defined	1549017088
Jr your customer service is inferior to that of a nationalized third world nation's airline. Get it together.	[0,0,0]	2/23/2015 12:00:00 AM	12/30/1899 3:16:41 AM	Not Defined	220794880
Jr i tried all day. Have been disconnected due to "heavy call volume."	[0,0,0]	2/23/2015 12:00:00 AM	12/30/1899 3:39:08 AM	Not Defined	903254016
Jr Have been trying to reach American Airlines since even late evening to Cancelled Flight flights. WORST CUSTOMER SERVICE EVER! What now???	[0,0,0]	2/23/2015 12:00:00 AM	12/30/1899 3:54:45 AM	Not Defined	57217024
Jr you're running 5+ hours behind replying to anyone via Twitter, this is beyond unacceptable. I'm running out of time!	[0,0,0]	2/23/2015 12:00:00 AM	12/30/1899 4:09:31 AM	Not Defined	390010624
Jr no date on when reviewing our money sent a letter in post and several emails no reply?? very frustrated	[0,0,0]	2/23/2015 12:00:00 AM	12/30/1899 4:10:23 AM	Not Defined	1326903296
Jr For two days now I have been trying to reach #AmericanAirlines customer service to change my flight.	[0,0,0]	2/23/2015 12:00:00 AM	12/30/1899 4:32:18 AM	Not Defined	191447040
Jr @emrey35 But it says your agents are too busy and to try back Late flight???	[0,0,0]	2/23/2015 12:00:00 AM	12/30/1899 4:34:17 AM	Not Defined	244764672
Jr All I get is an automated system that hangs up on me. I finally get a message that some will call me back in 2 hours.	[0,0,0]	2/23/2015 12:00:00 AM	12/30/1899 4:34:25 AM	Not Defined	328634368
Jr It's been 5 hours and still no call and now we are back to being hung up on by an automated system. #AmericanAirlines	[0,0,0]	2/23/2015 12:00:00 AM	12/30/1899 4:37:32 AM	Not Defined	86597632
Jr I am still waiting for that call back. Stranded and no one to talk to	[0,0,0]	2/23/2015 12:00:00 AM	12/30/1899 4:52:45 AM	Not Defined	1788157953
Jr stranded in Miami because your automated system keeps hanging up on me for two days. Help !	[0,0,0]	2/23/2015 12:00:00 AM	12/30/1899 4:58:30 AM	Not Defined	141127680
Jr AmericanAir ... Poor customer service 3 trips in a row! @Delta here I come!	[0,0,0]	2/23/2015 12:00:00 AM	12/30/1899 5:28:18 AM	Not Defined	1905696770
Jr Or u cud answer my ? here! 140 is plenty. Infight stealing going on right now & FA won't have the uncomfortable convo w pass.	[0,0,0]	2/23/2015 12:00:00 AM	12/30/1899 6:15:56 AM	Not Defined	2086076416
Jr Filed a PIT with an agent at Indianapolis airport. Unable to pull up information online on your site. Have DMed details.	[0,0,0]	2/23/2015 12:00:00 AM	12/30/1899 6:21:25 AM	Not Defined	233381888
Jr I have still not talked to anyone. You guys should be prepared for these situations. How is this good service? I am so mad!!!	[0,0,0]	2/23/2015 12:00:00 AM	12/30/1899 6:31:40 AM	Not Defined	920035328
Jr I've tried ten times now last night and this morning...how many times should I try? Have meetings to make tonight in ATX.	[0,0,0]	2/23/2015 12:00:00 AM	12/30/1899 6:33:19 AM	Not Defined	795439104
Jr at LAX and your service reps just hand out the 800 number to call. So that's not helpful.	[0,0,0]	2/23/2015 12:00:00 AM	12/30/1899 6:34:40 AM	Not Defined	743731200

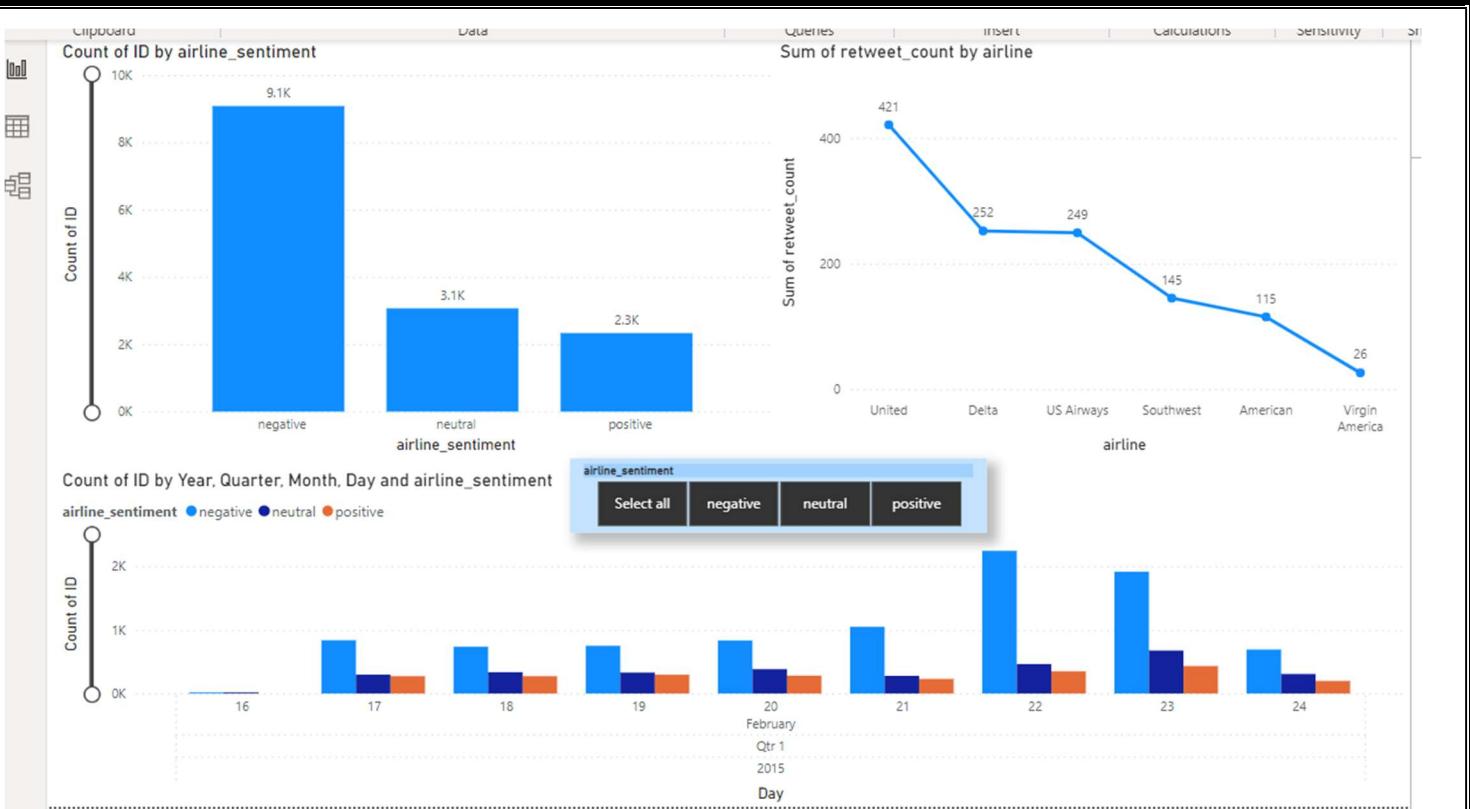
Univariate Graphs with interactive visualization

Univariate graphs are visual representations of data that involve **a single variable**. These graphs provide insights into the distribution, central tendency, variability, and other characteristics of a single variable. Univariate graphs are particularly useful for summarizing and understanding the data's distribution and patterns.

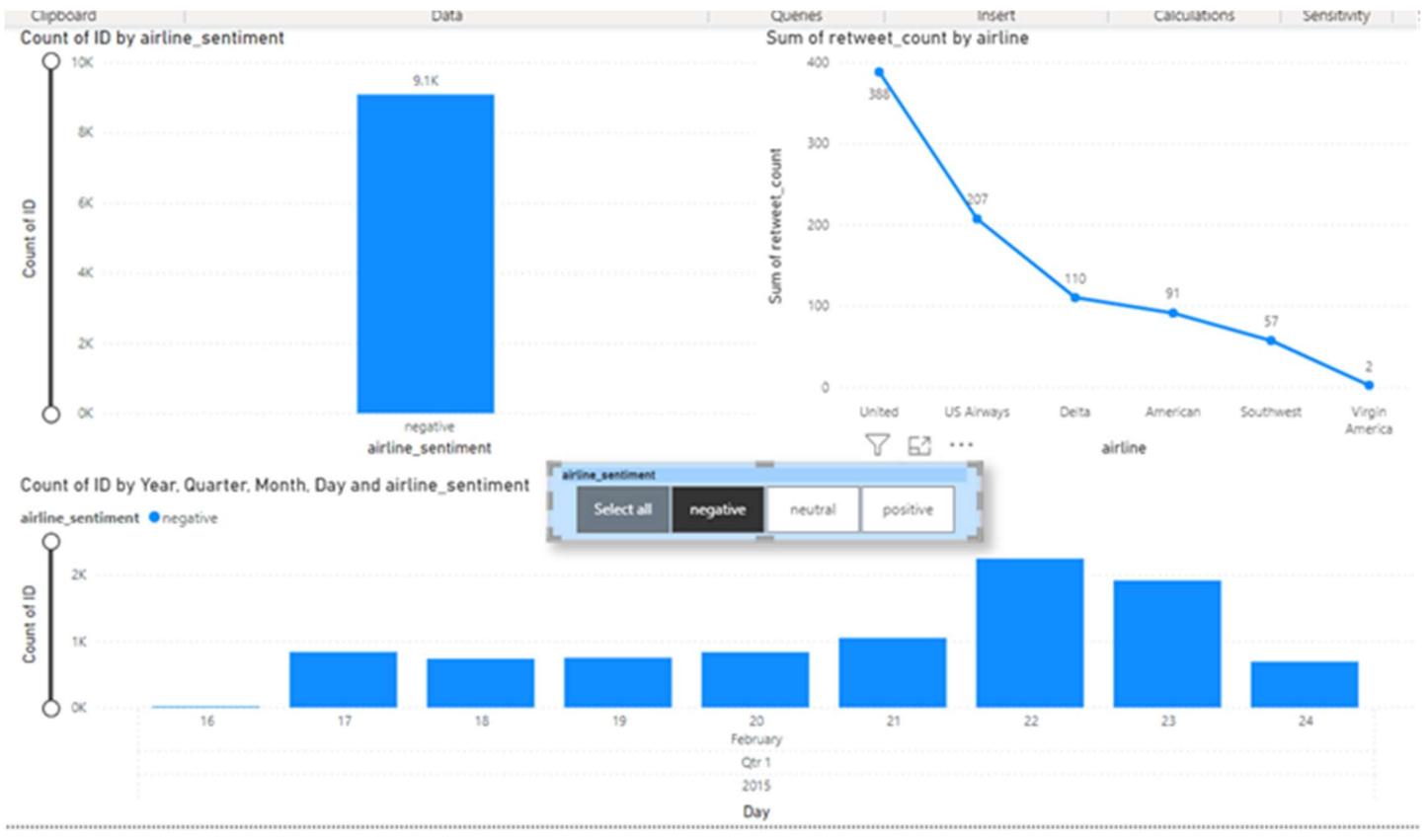
- 1) Relation between tweet id and airline sentiment using bar chart.
- 2) Relation between airline and retweet count using line chart.
- 3) Daily bar chart for the relation between sentiment and id
- 4) Slicer which filters sentiment

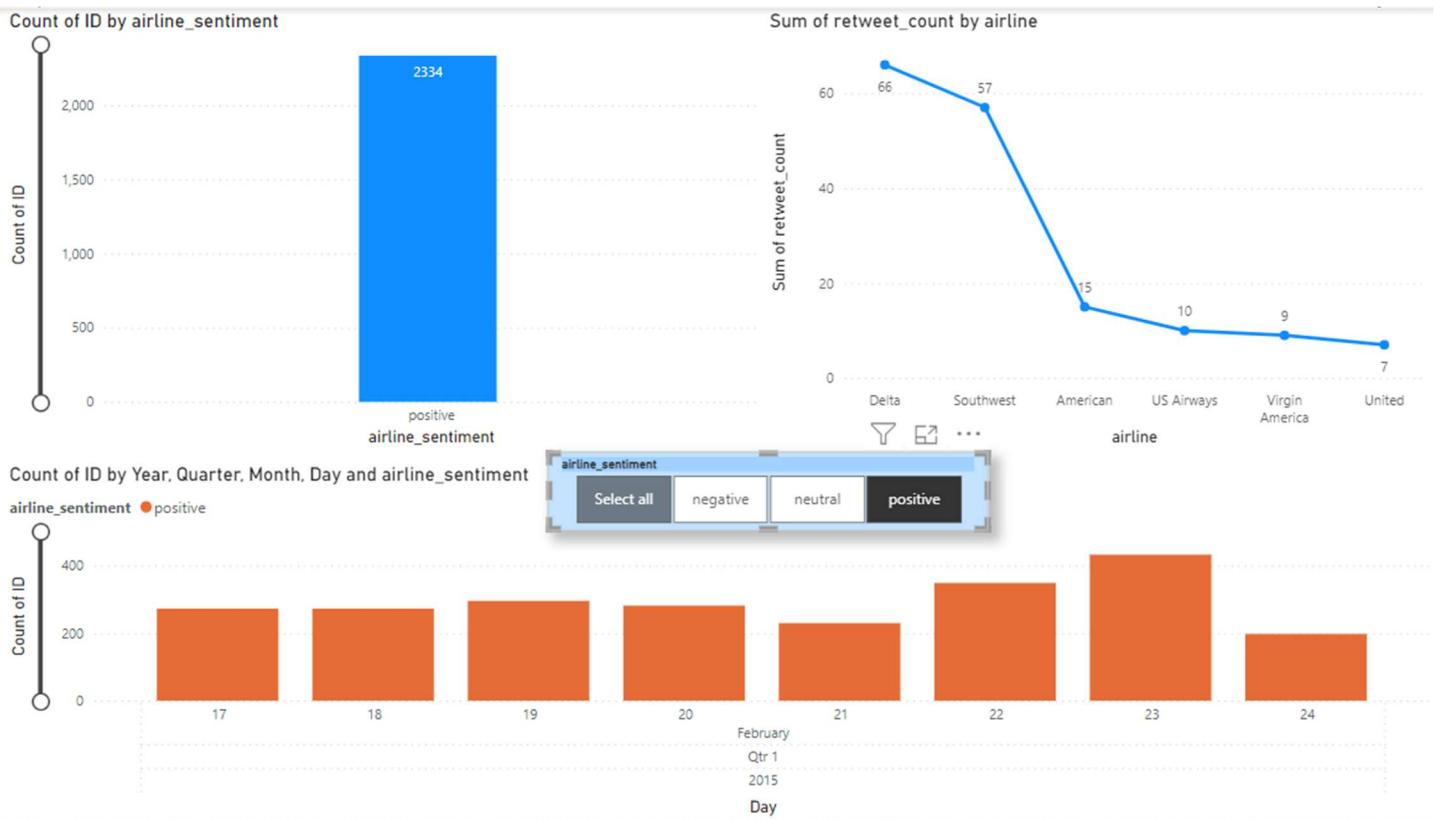
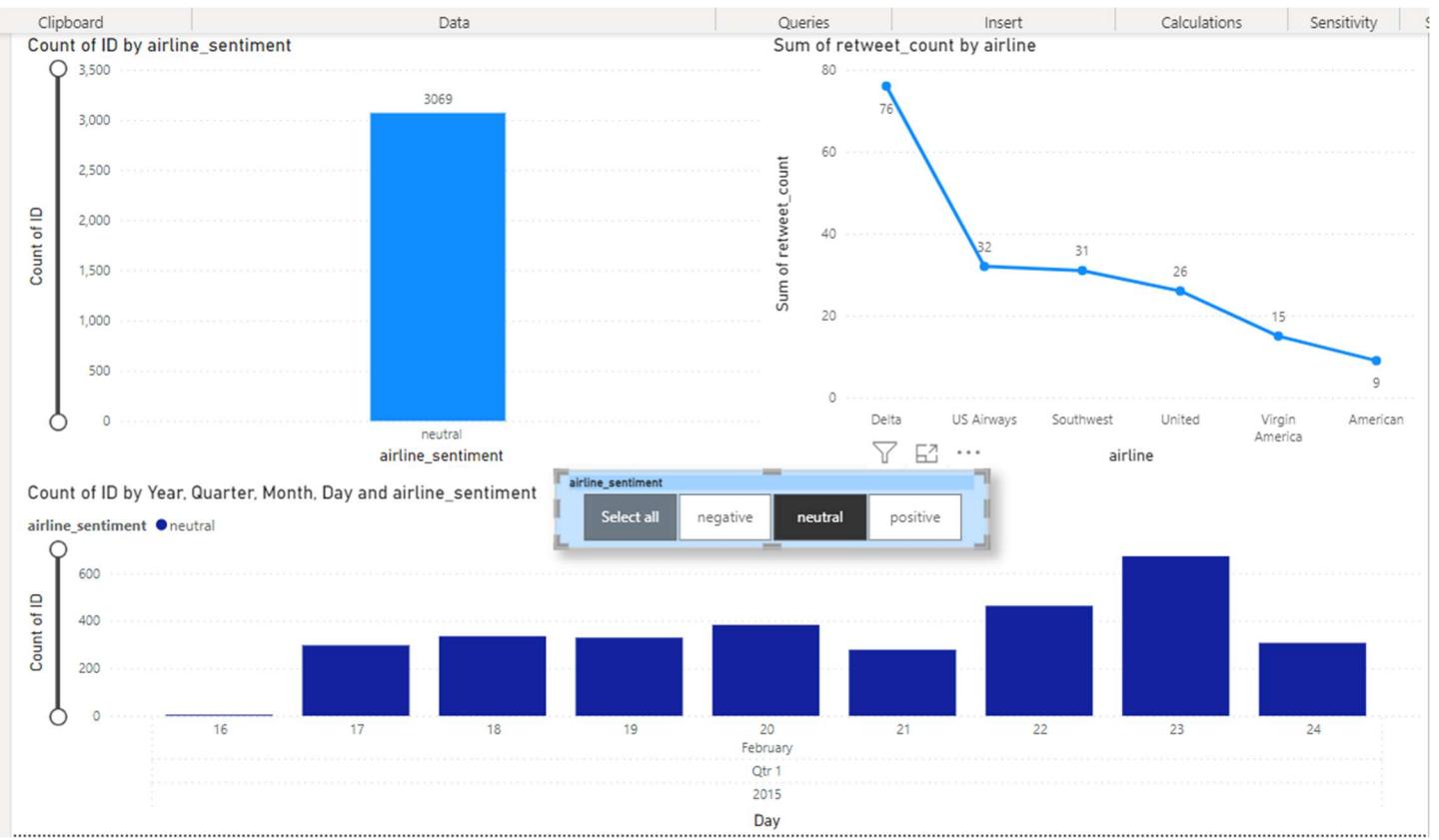
In Power BI, a **slicer** is a visual component that allows users to filter and interact with data in a report or dashboard. It provides a way to slice and dice data by selecting specific values or ranges, enabling users to focus on specific subsets of information.

These are just a few examples of **univariate graphs**. The choice of the graph depends on the type of variable (categorical or continuous) and the specific characteristics of the data. Visualization tools like Power offer a wide range of options for creating **univariate graphs**, allowing for customization and interactive exploration of the data.



From the total dashboard we conclude that negative sentiments are the most and there is daily increase in this sentiment specially at the united air line , so there might be a big problem in united air line that need to be solved and here are some filtered plots





Bi-bivariate Graphs with interactive visualization

Bivariate graphs, also known as scatter plots, are visual representations of data that involve **two variables**. These graphs help to explore the relationship between two continuous variables and provide insights into their correlation, patterns, and trends.

In a **bivariate graph**, each data point is plotted as a single point on the graph, with one variable represented on the x-axis and the other variable represented on the y-axis. The position of each point on the graph corresponds to the values of the two variables for that data point.

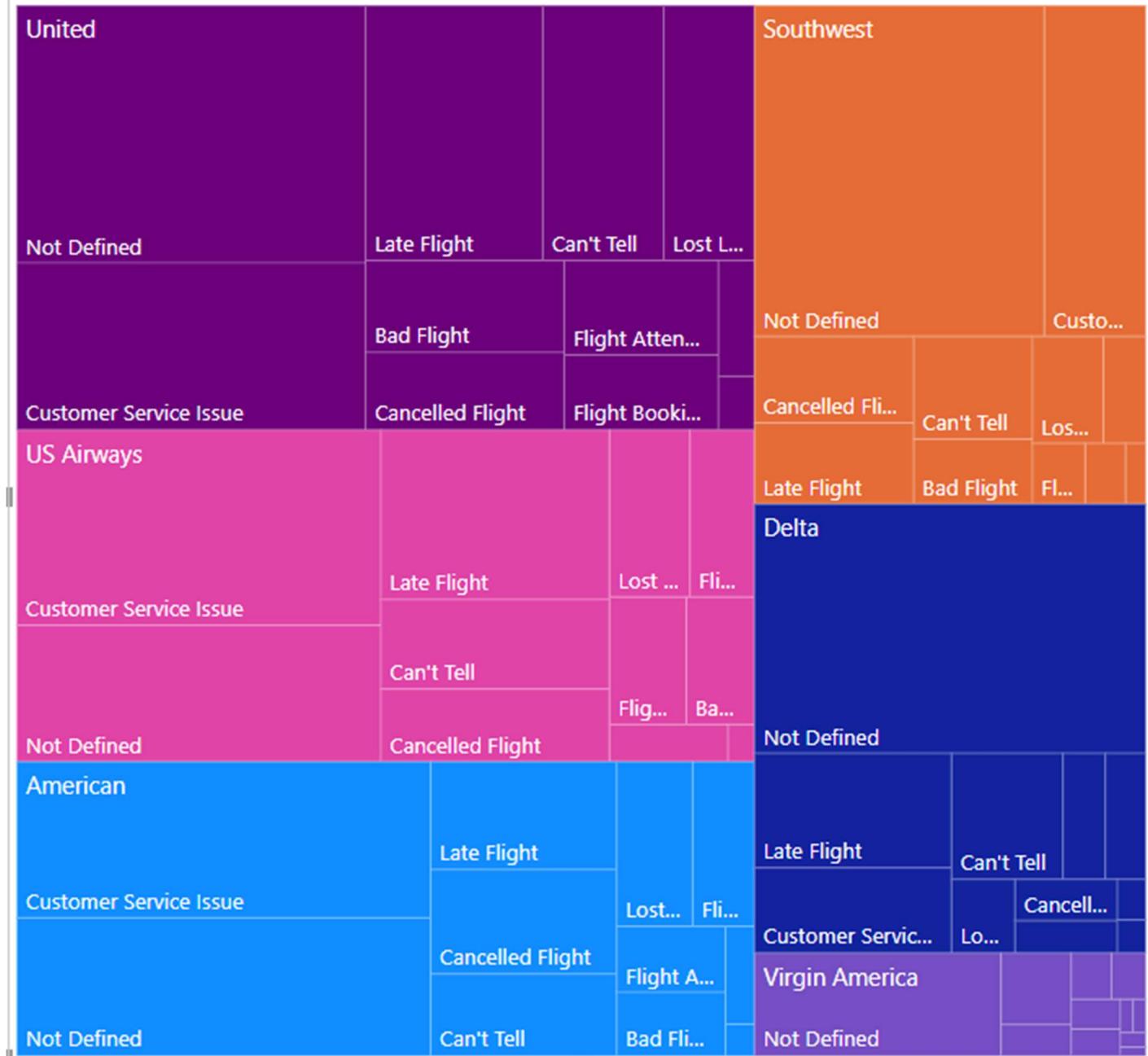
By examining the distribution of points on the graph, you can determine the nature of the relationship between the two variables

Bivariate graphs are useful for visualizing and analyzing relationships between variables, identifying outliers, detecting patterns, and assessing the strength and direction of correlation. They are commonly used in various fields such as statistics, data analysis, social sciences, and business analytics.



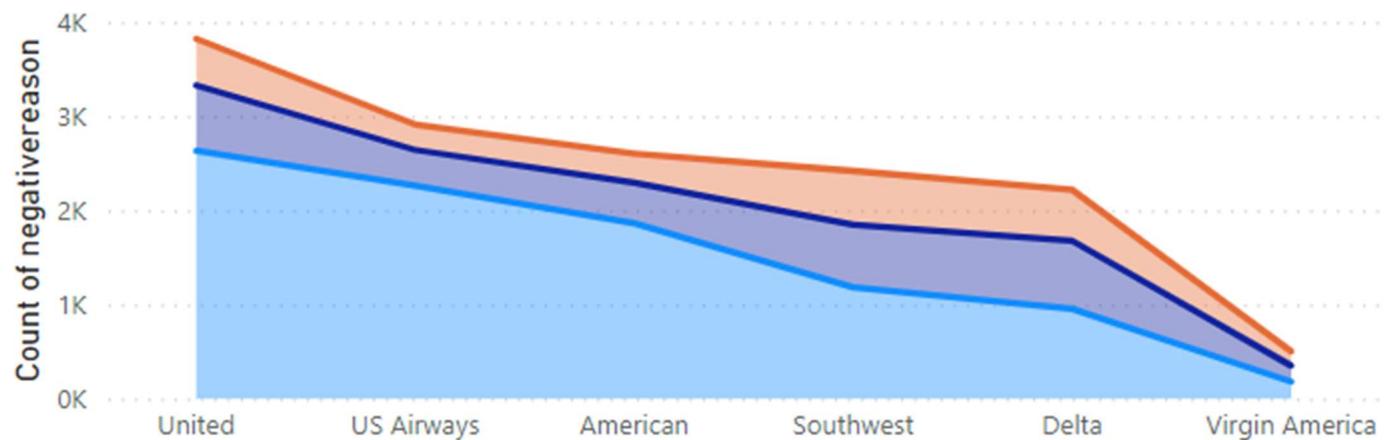
Heat map: A heat map is a bivariate graph that represents the relationship between two categorical variables. It uses color-coding to indicate the frequency or count of combinations of categories. To create a heat map in Power Bi

Sum of airline_sentiment_confidence and Sum of retweet_count by airline and negative reason

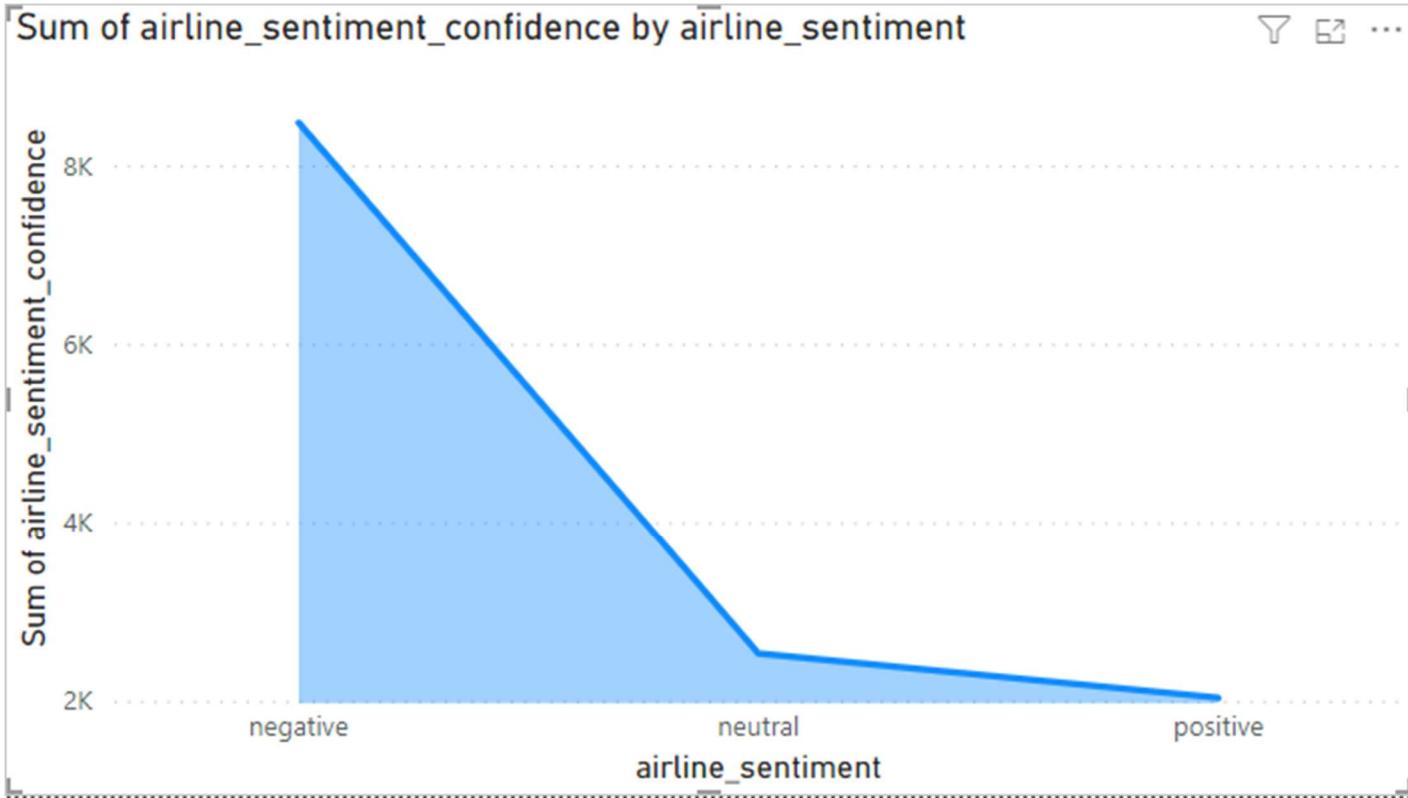


Count of negative reason and Sum of retweet_count by airline and airline_sentiment

airline_sentiment ● negative ● neutral ● positive



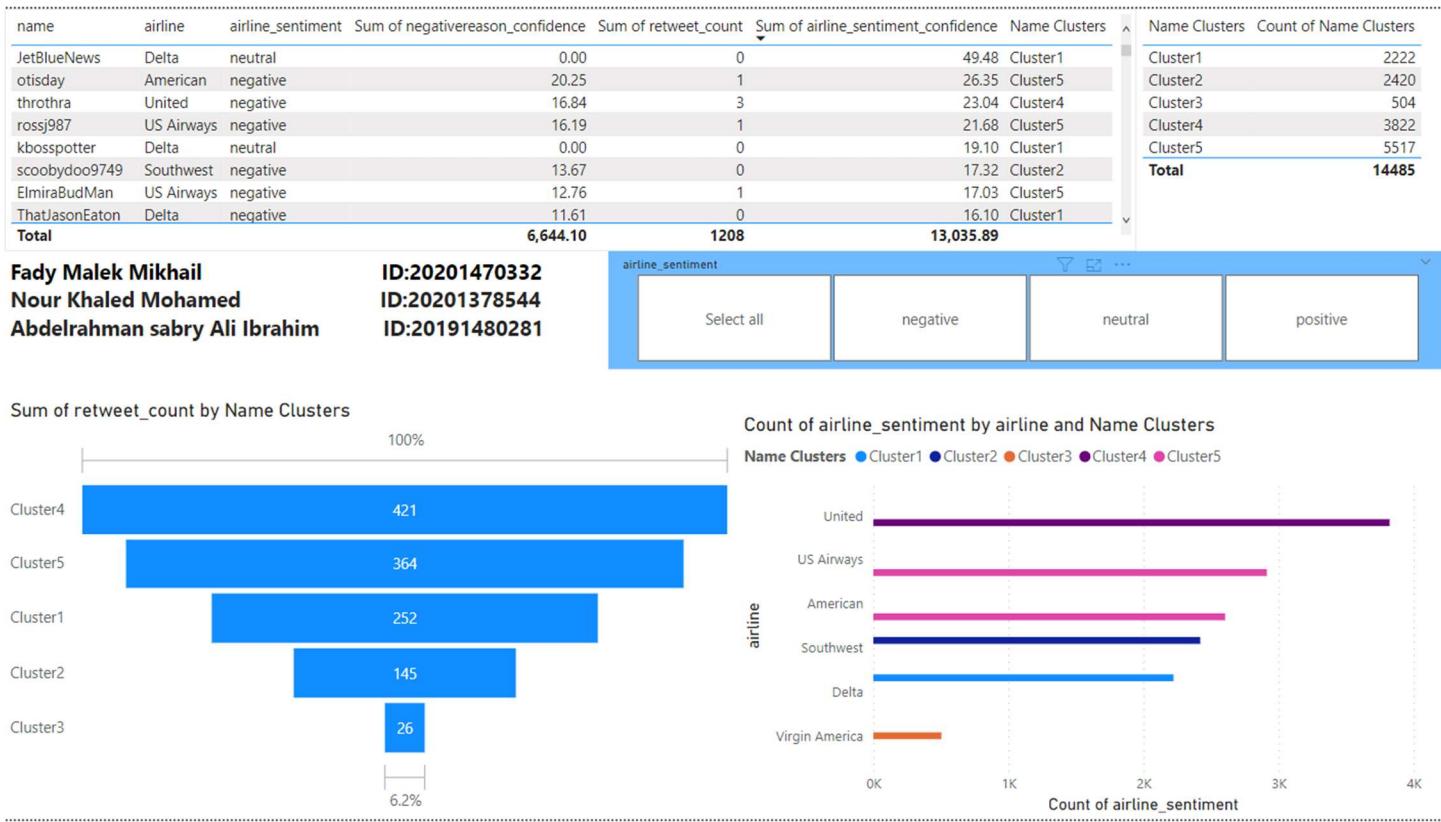
Sum of airline_sentiment_confidence by airline_sentiment



Multivariate Graphs with interactive visualization

Multivariate graphs are visual representations of data that involve three or more variables. These graphs enable the exploration and analysis of relationships and patterns among multiple variables simultaneously. They are particularly useful when examining complex datasets with multiple dimensions.

The choice of graph depends on the types of variables and the analysis objectives. Visualization tools like Power BI provide options for creating multivariate graphs and allow for customization and interactive exploration of the data.



negativereson ×
Cancelled Flight

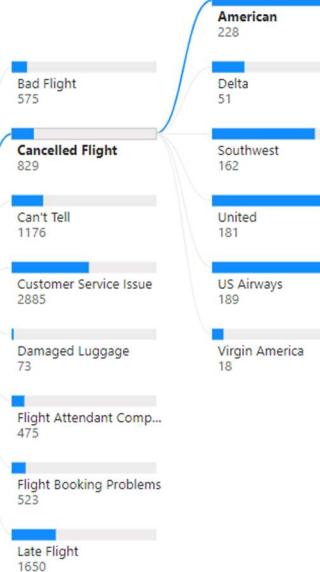
airline ×

Day ×

Fady Malek Mikhail
Nour Khaled Mohamed
Abdelrahman sabry Ali Ibrahim

ID:20201470332
ID:20201378544
ID:20191480281

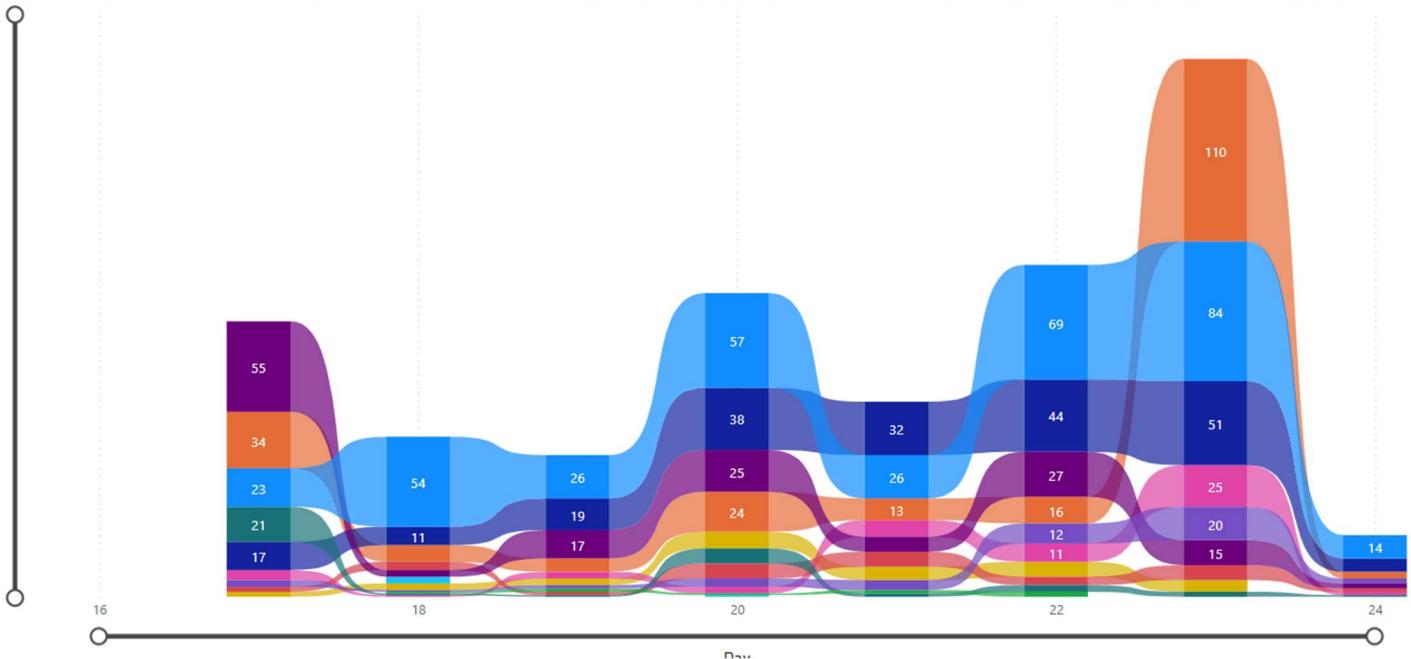
Count of ID
14485



Fady Malek Mikhail ID:20201470332
Nour Khaled Mohamed ID:20201378544
Abdelrahman sabry Ali Ibrahim ID:20191480281

Sum of retweet_count by Day and negativereson

negativereson ● Bad Flight ● Cancelled Flight ● Can't Tell ● Customer Service Issue ● Damaged Luggage ● Flight Attendant Com... ● Flight Booking Pr... ● Late Flight ● longlines ● Lost Luggage ● Not Defined

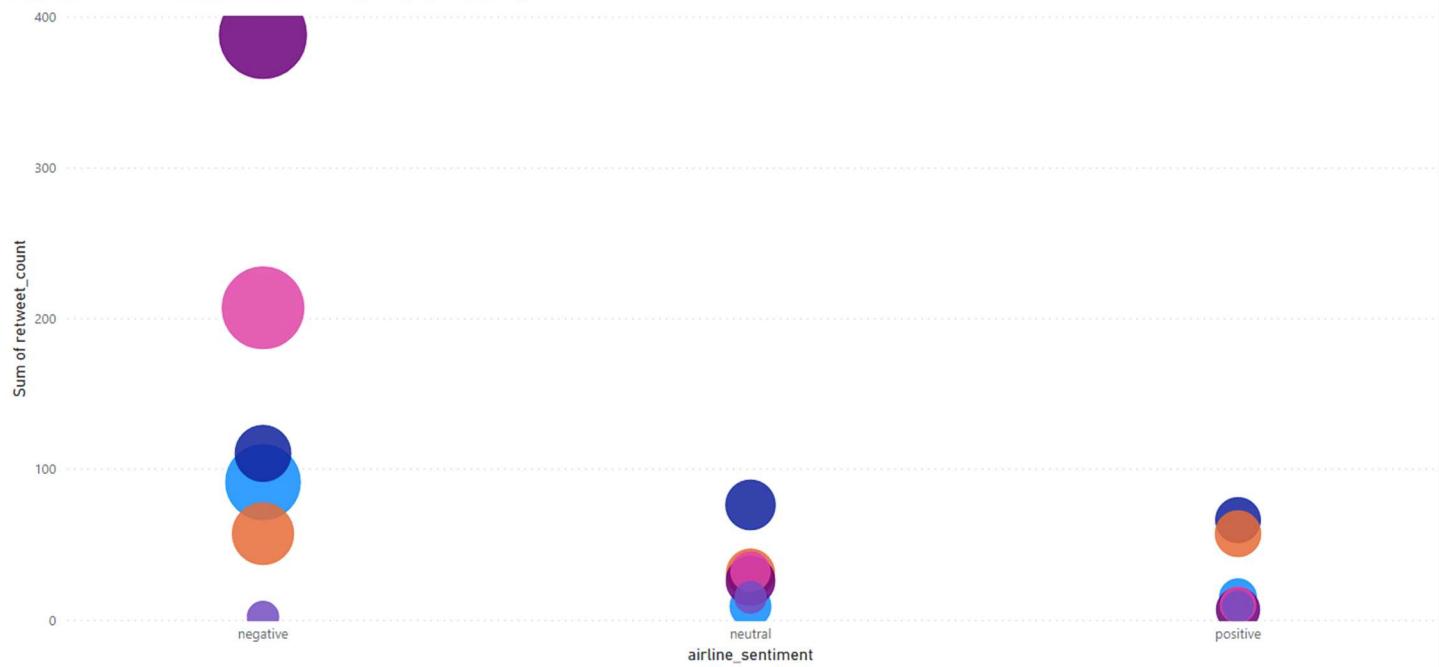


Fady Malek Mikhai
Nour Khaled Mohamed
Abdelrahman sabry Ali Ibrahim

ID:20201470332
ID:20201378544
ID:20191480281

Sum of retweet_count and Count of ID by airline and airline_sentiment

airline ● American ● Delta ● Southwest ● United ● US Airways ● Virgin America



Thank you.