

UNIVERSITY OF MALTA

L-Università ta' Malta

MSc in Strategic Management and Digital Marketing

MRK5812 - Digital Analytics for Marketing Management

Karsten Guenther (0295697)

Jashid Munshi (1905943)

Abhishek Shah (1906057)

Assignment Part 1 – Word Clouds



Figure 1

An online article from Times of Malta website was selected so that it would provide us with a data set which can be utilised for the purpose of creating a word cloud chart (Figure 1).

The data is copied from the downloaded article and pasted into Note++ since we need to do some data pre-processing so that we could utilise it in Tableau. This step is required so that I can transform raw data into more appropriate refined data which will be used for analysis purposes and chart mapping

In this step I convert all text to lowercase and removed punctuation and split text by space which allows me to transform the data from paragraph form into a list which makes it easier for tableau to access and process. This text file saved as 5GTOM.txt and opened with tableau to start the process of creating a word count. Figure 2& Figure 3.

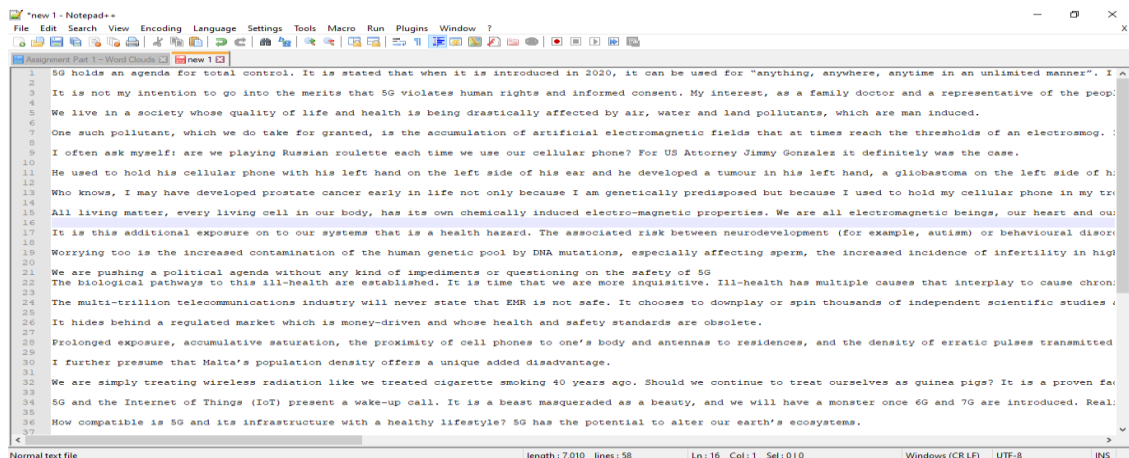


Figure 2

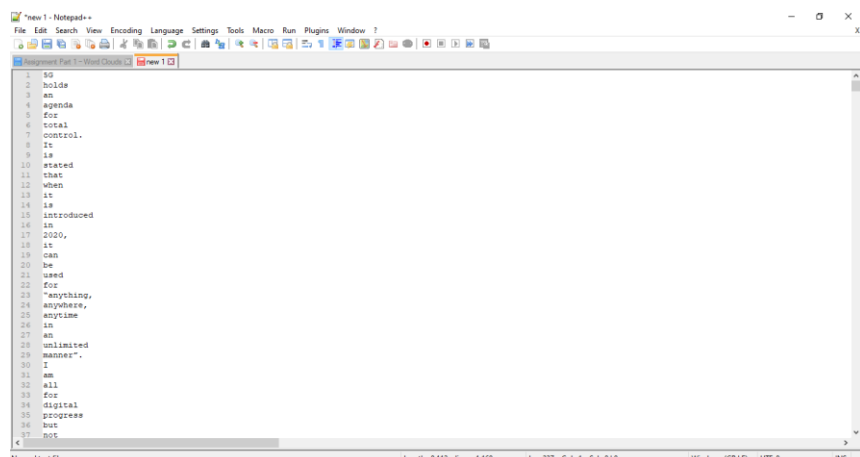


Figure 3

As we can see from Figure 4, I opened the text file 5GTOM.txt on which the pre-processed data is available (data source) and opened a sheet on which I shall start the word cloud. moving to sheet one I started dragging '5G' which marks the data to be used to 'Text' and 'Size' into the 'Marks Box'. I did this step since I need the Words to be displayed and their relative cumulative appearances. I right clicked on the second '5G' and choose Measure and then Count. This ensures that we are showing the words by count.

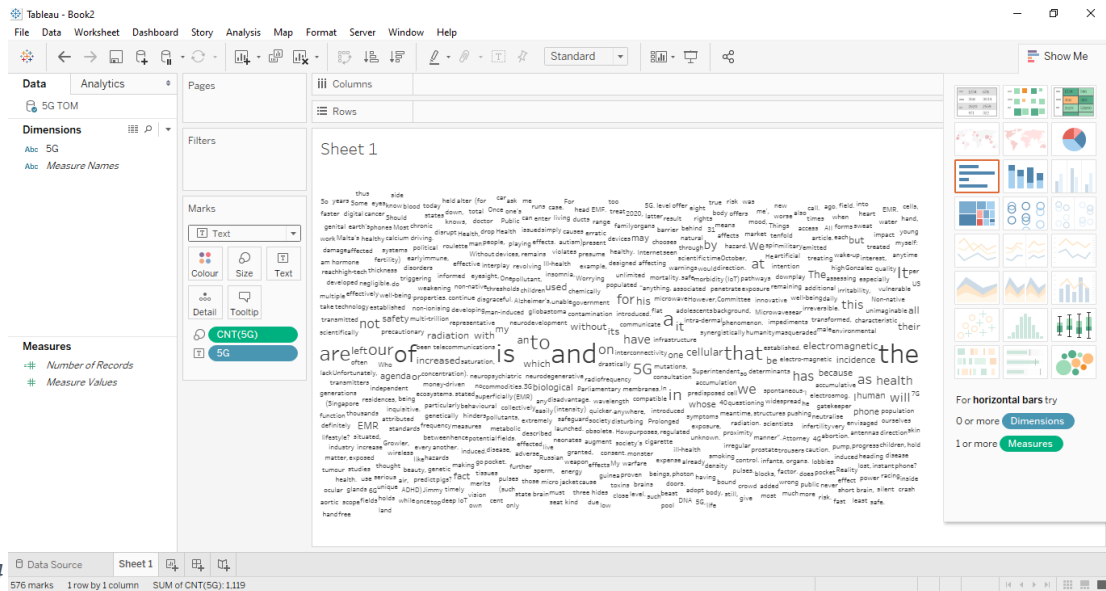


Figure 4

In Figure 5 we can see that I dragged the data source '5G' into the 'Filters Box'. I edited the filter to display words only that appear more than 4 times in the document. Then I dragged '5G' to colour and represent each word frequency with a different shade of blue. By changing the measure to 'count'. The darker the blue is the more frequent the word was in the data.

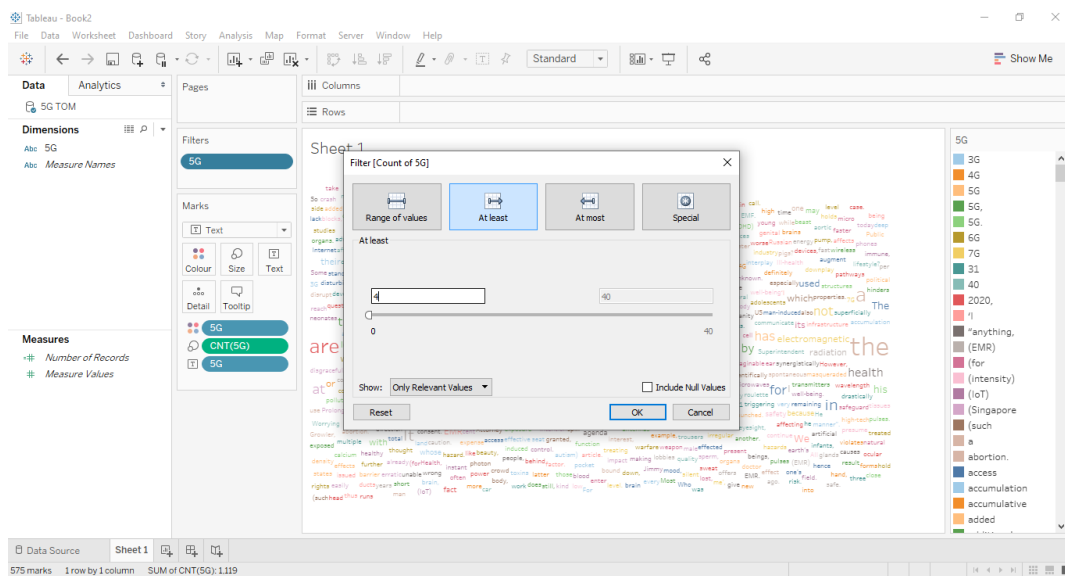


Figure 5

In Figure 6 we can see that I created a new data source 'StopWords' which consists of common stop words that need to be eliminated so that we have a more robust word cloud. This excel file was created and the stop words were typed and saved to the file. The 'Stop Words' was dragged to 'Filters' and I edited the filter to exclude the words that were both in that word cloud and in 'StopWords'. This gave me an error and told me that I had to blend the data sources in order to be able to use the data together. This step is shown in Figure 6. I went to 'Data' and 'Blend Relationships' and selected both sources to be utilised for this procedure.

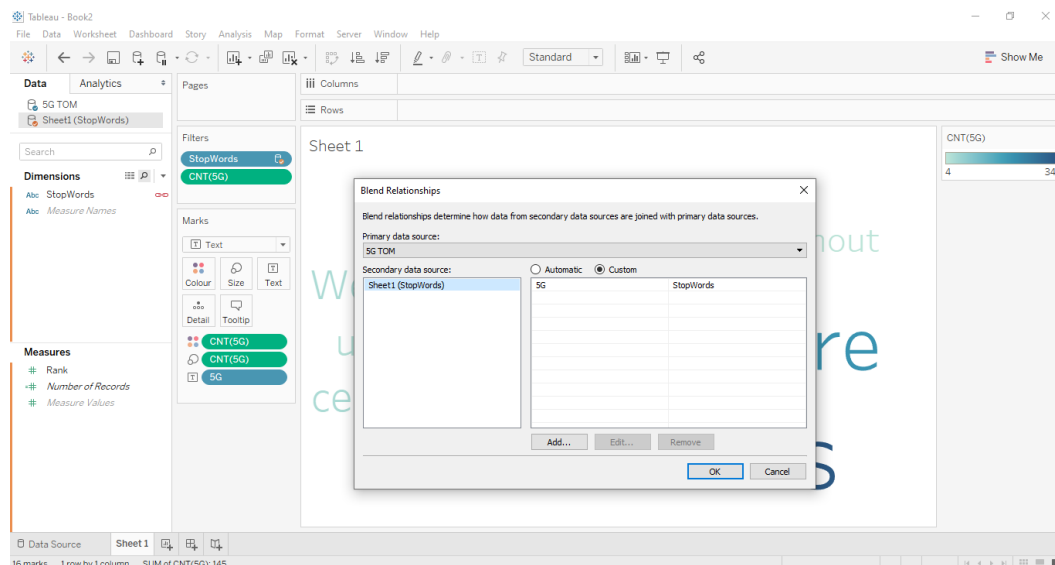


Figure 6

A couple of finishing touches for the sake of representation the colour of '5G' in the 'Marks Area' is edited to count so that it reflects the number of appearances in the data source.

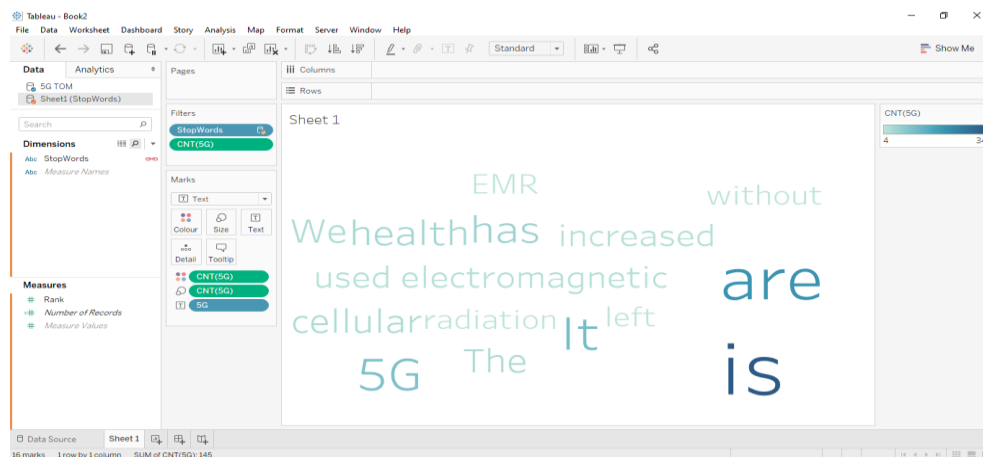
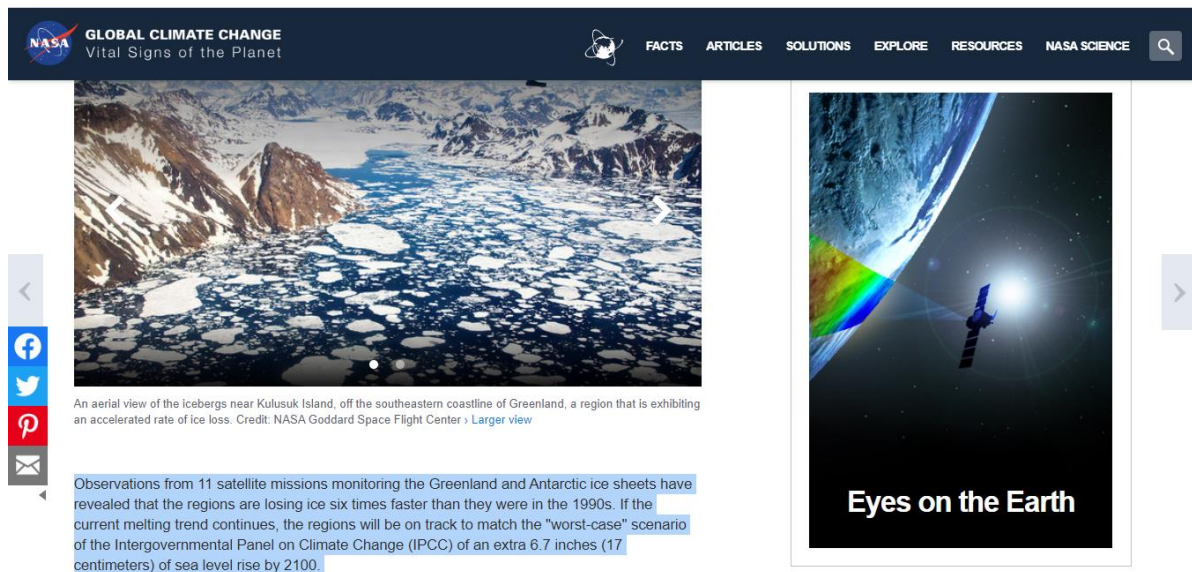


Figure 7

Summary:	
StopWords/Filters	Data Source used to eliminate useless words from our Main Data Source
CNT(5G)/Filters	Filters and displays words that appear more than 4 times
CNT(5G) Colour/Marks	Colour Codes the words by number of appearances
CNT(5G) Size/Marks	Displays the words in the size according to their appearances
5G Text/Marks	Displays the words

Comparing 2 Articles:

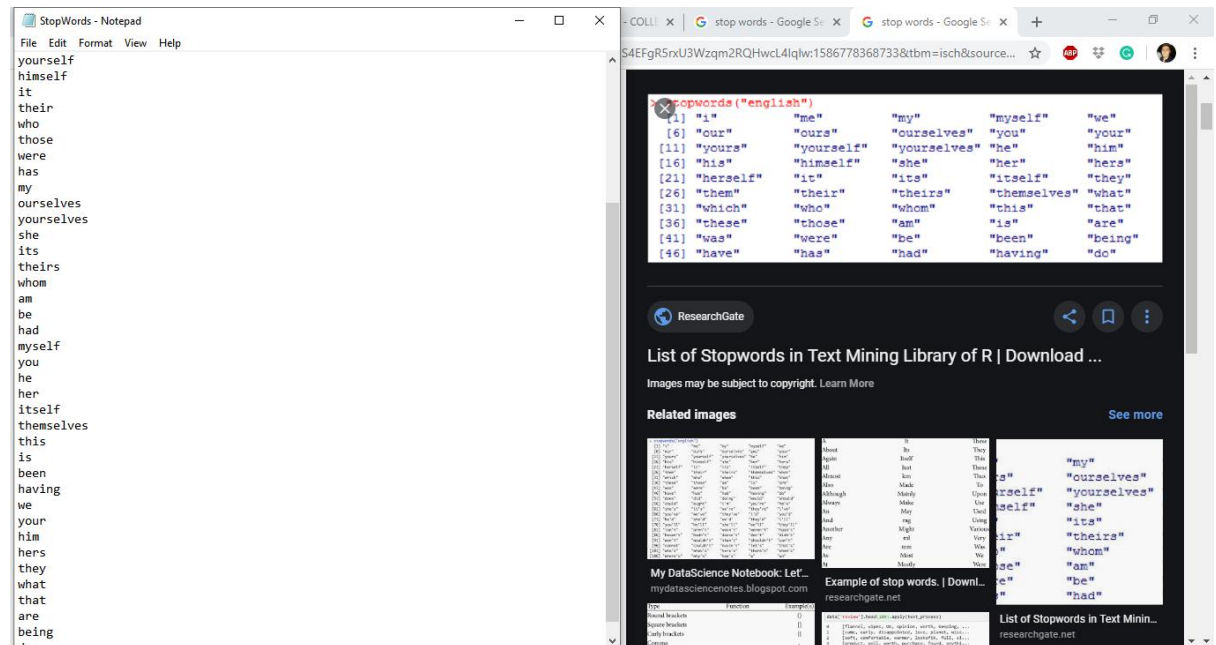
1. 2 articles from climate.nasa.gov were selected so that we can compare the word cloud that they produce.



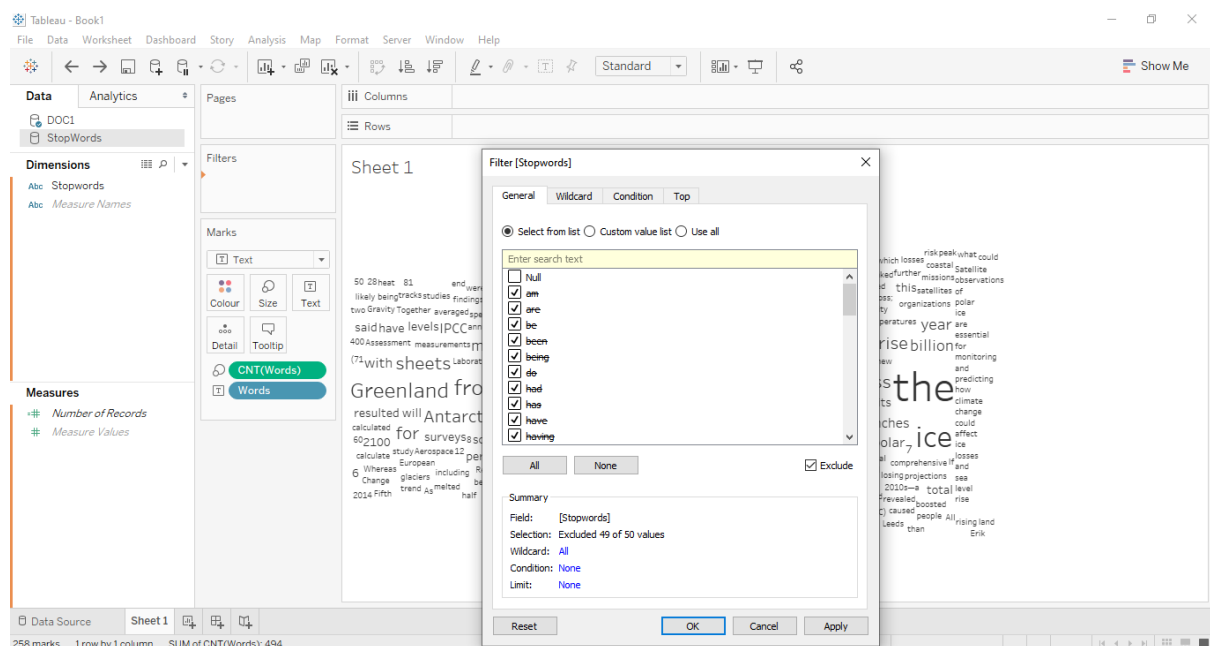
2. The contents are then pasted into a Note++ file so that we can prepare the data for processing in tableau.
 - This step involves converting every word as a separate cell and clean the data by replacing any spaces with a new line and cleaning the data from any fonts and punctuation.

```
*C:\Users\Karsten\Desktop\DOC1 - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
DOC1 DOC2
1 observations
2 from
3 11
4 satellite
5 missions
6 monitoring
7 the
8 greenland
9 and
10 antarctic
11 ice
12 sheets
13 have
14 revealed
15 that
16 the
17 regions
18 are
19 losing
20 ice
21 six
22 times
23 faster
24 than
25 they
26 were
27 in
28 the
29 1990s
30
31 if
32 the
33 current
34 melting
35 trend
36 continues
37
Normal text file length: 3,969 lines: 606 Ln: 1 Col: 1 Sel: 0 | 0 Windows (CR LF) UTF-8 INS
```

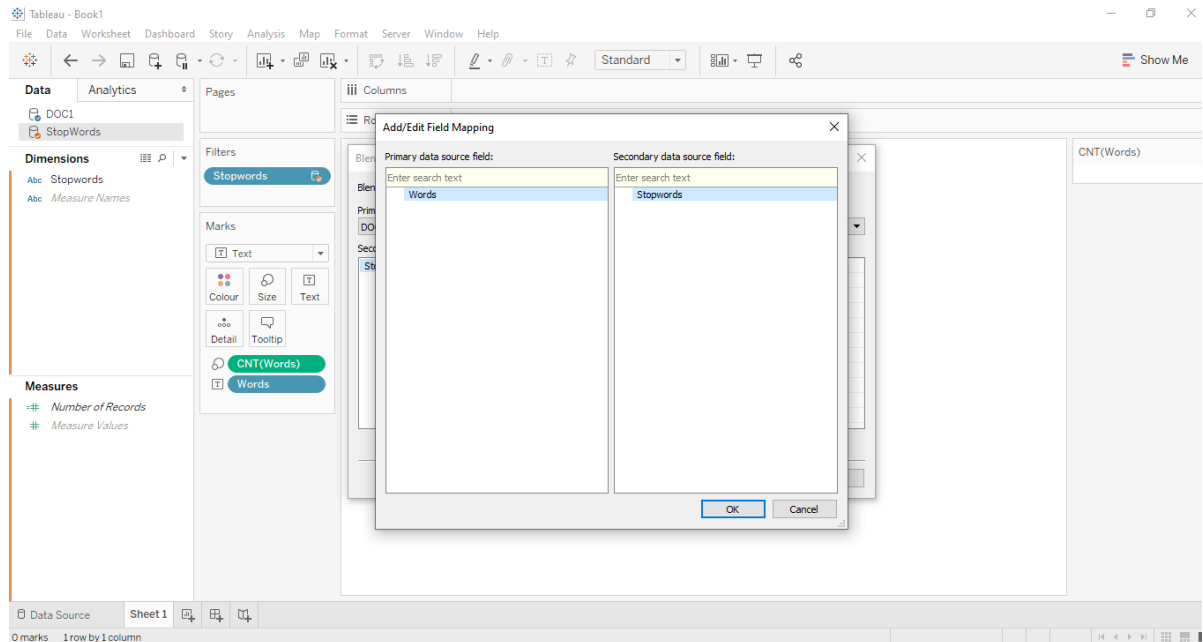

4. A list of stop words was created by searching for stop words on the internet and copying them into a text file so that they can be used when processing the data. This file was added to the worksheet as another data source which will serve to filter out any unnecessary words in the word cloud.



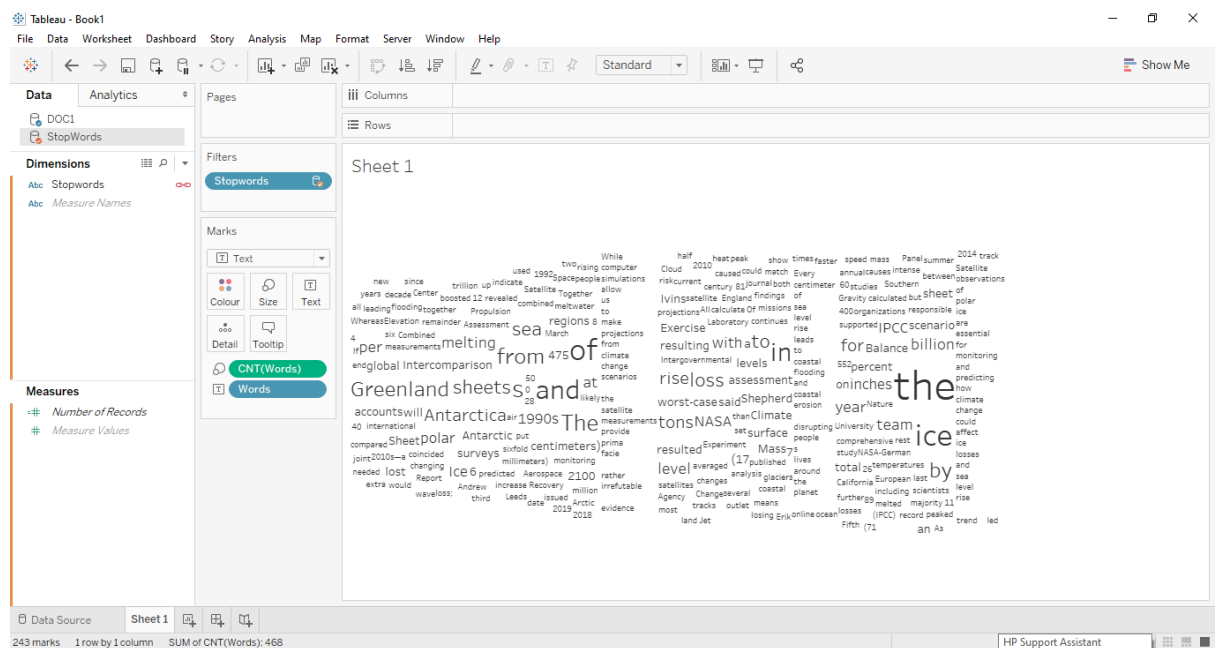
5. In this step I dragged 'StopWords' to 'Filters' selected all the values except for Null to exclude all the words from the cloud and I applied the filter



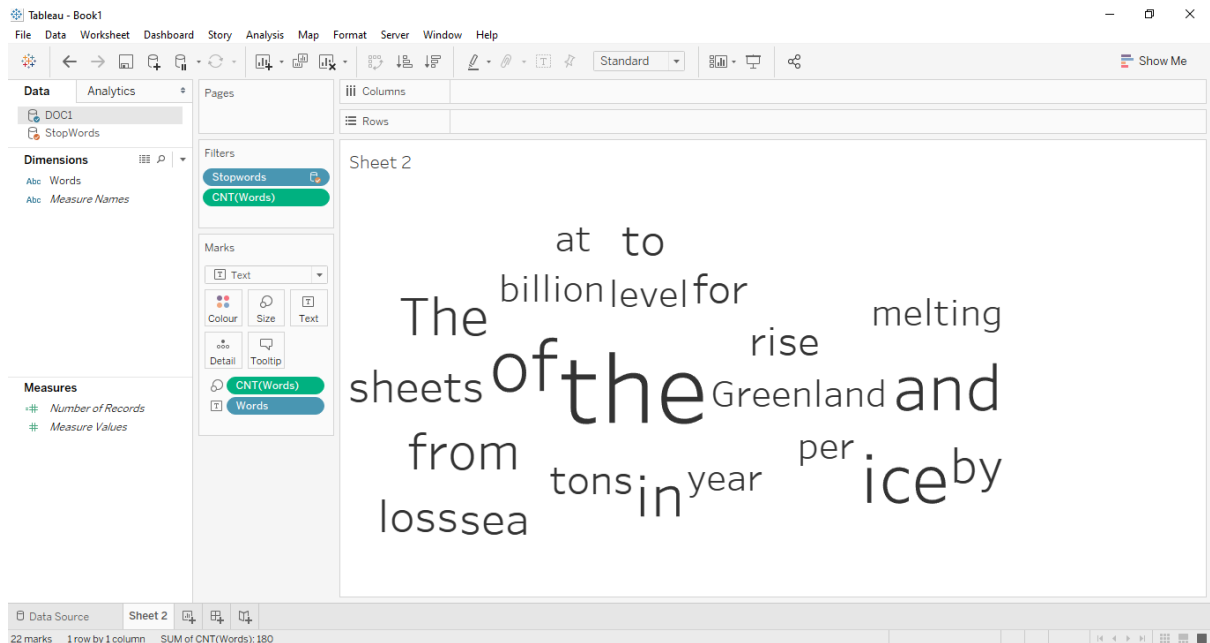
- An error occurred since I had to blend the data between the two data sources so that they can be used together before trying to filter out the stop words. To overcome this problem, I went to Data – Edit Blend Relationships - Custom – Ok.



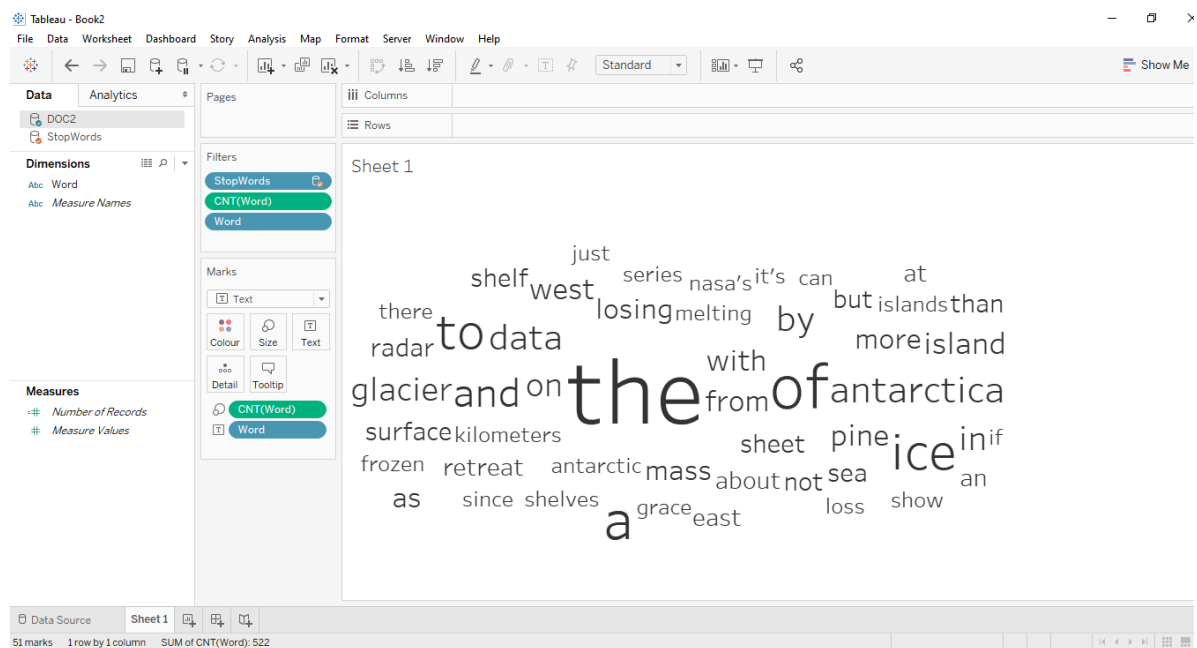
- The Result of applying the 'StopWord' filter



7. After applying the 'StopWord' Filter we had to reduce the words to those that appear more than 4 times in the data source. This was done by selecting 'CNT(Words)' from 'Marks' and 'Edit Filter' – 'At Least' - and set it to 4 meaning the number of occurrences in the document.
 - The following is the result of the first document what was processed, and a word cloud was visualised.



- This is the result the 2nd Document that needed to be processed and a word cloud to be formed.
- In this document I had a hyphen which was occurred 4 times in the document, so I had to exclude it manually since it did not add any value to the word cloud.

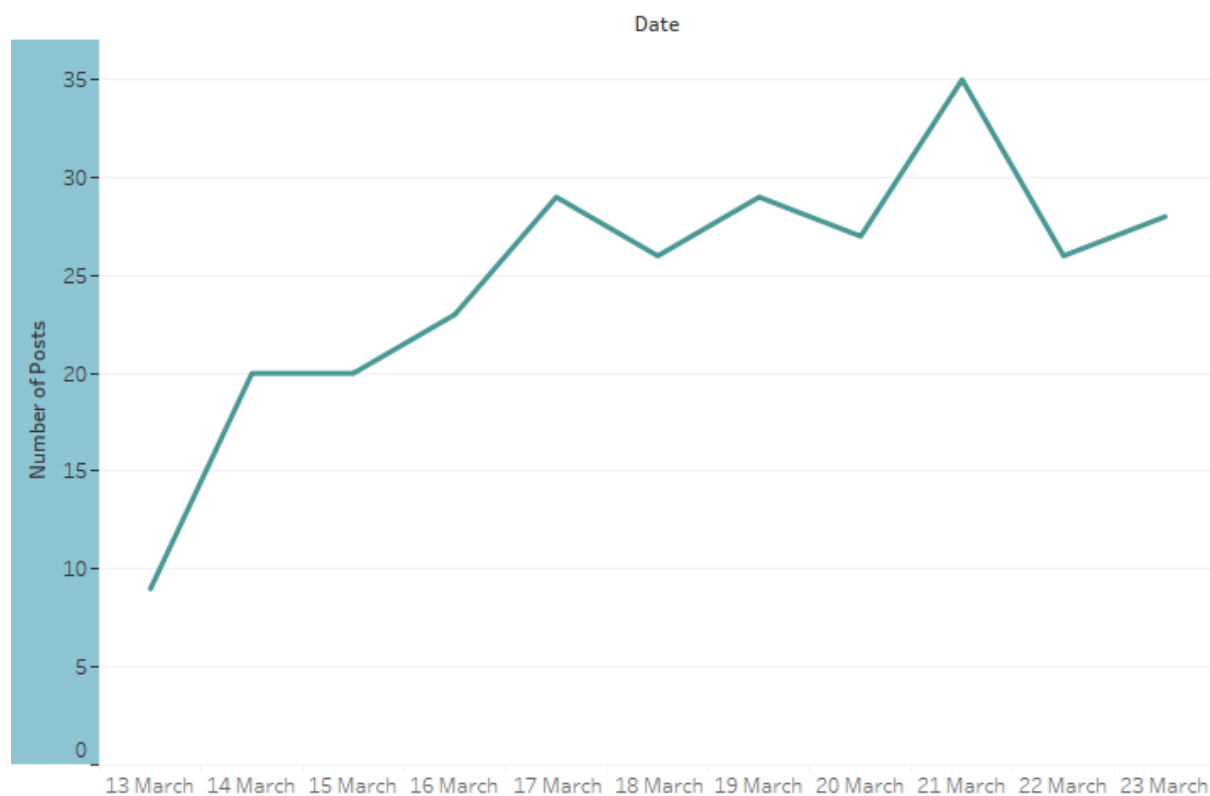


The main differences between the 2 documents are that the more words the data source has the more populated the word cloud is. The first document had a total of 557 words whilst the second one had a total of 1281 this increases the instances where each word can be used. The most common words we visible with a larger font and those were 'the' and 'of' 'to' and 'ice'. Both word clouds' topics were evident since some words were repeated enough to give the viewer enough words to understand what the word cloud is about.

Assignment Part 2:

@timesofmalta Facebook page is the official Facebook news platform for the media company Times of Malta and its page is followed by over 200,000 followers and in this report we will analyse key metrics which will give us interesting insights over the 10 day period from Friday 13th March 2020 to Monday 23rd March 2020. The key metrics to be analysed are:

Number of Posts for March 13 - March 23



The trend of sum of Number of Posts for Date. The view is filtered on Date, which excludes Null.

The number of posts on March 13 amounts to 9 due to the fact the Facebook page was scraped from 17:30 onwards and thus only captured 9 posts on that day. Nonetheless we can see that the frequency of posts increased steadily as the weekend was approaching. The 'weekend posting' is a great way to generate content and organic traffic thanks to the Facebook page since people will be more active during the weekend. The daily analysis of the page posting is analysed next in the next section where we can see at what times the posts generated most traction. As the analyst I assumed that people commenting on a post is more interactive than just sharing the post since people would have read and contributed to the page by means of their opinion, therefore I generated a chart that visualises the time of the post with the associated number of shares and comments.

Since in the data set had its limitations hindered me from analysing the true representation of the actual activity on the page by collating the number of comments and other interactions at time of occurrence I analysed what time the post with the most interactions was posted.

The problem that I faced when analysing the comments is that I assume that the web scraping tool used combines comments at the same time under one time and merged the time for the comments, when coming to analyse every single comment unmerging them would have created empty cells in the document. This would have to be done for all the comments posted at a certain time.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Time (Comments)		Comment													
2	Friday, 13 March 2020 at 17:33	17:33	not really anything they can do about it they loose business short term. else 1000's get sick and they don't have workers to operate and then loose business...													
3	Friday, 13 March 2020 at 17:35	17:35	so what ? are these hoteliers expecting them to run around the streets, how about they invite them home for dinner and a glass of wine as well ?? at this moment every													
4	Friday, 13 March 2020 at 17:36	17:36	They shouldn't be! Health is the priority at this stage, measures are taken hour by hour. Only the irresponsible travel for holidays during this critical situation. People are													
5	Friday, 13 March 2020 at 17:36	17:36	Have they not had it so very good for a long time, what do they expect to infect more people to make them richer?													
6	Friday, 13 March 2020 at 17:38		It's immediate national security actions. What prior notice? Did schools, beauticians, gyms, restaurants, workers had prior notice?!													
7	Friday, 13 March 2020 at 17:39		Health of the nation comes first. No time for discussions!													
8	Friday, 13 March 2020 at 17:40		It-turists ser jehluha ukoll il-multa ta f'elf euro? Ghax filkas mil Birgu biss ghandu fejn idahhal mal-miljun euro													
9	Friday, 13 March 2020 at 17:40		Good AND bad timing for tourists about to travel from Bristol Many dismayed tourists choosing not to travel													
10	Friday, 13 March 2020 at 17:41		I truly depend on tourism actually I am jobless RightNow and for the time being yes it's the right decision													
11	Friday, 13 March 2020 at 17:41		It's not about the money, but doing quarantine in hotels is hard for the staff & workers working inbound hotels too. They're risking to contract													
12	Friday, 13 March 2020 at 17:42		Health first I'm sure you've had good times, many of us r out of work just suck it up and be thankful when u wake up healthy every morning!													
13	Friday, 13 March 2020 at 17:42		Really?! Business over national security?! Stop all nonessential activities and services. Now. Before we lose control over the situation. Stay inside. #stamatalaq#													
14	Friday, 13 March 2020 at 17:43		Tough luck! Our health comes first and foremost!													
15	Friday, 13 March 2020 at 17:43		No time for discussion in such situations!! Such directives are needed and tougher ones are coming													
16	Friday, 13 March 2020 at 17:45		After year after year boasting incredible success and tourism bla bla then they throw a tantrum because some necessary measures had to be taken.													
17	Friday, 13 March 2020 at 17:49		Health is a priority, the more discipline we are the more we get rid of the virus fast and get back on track! No time for discussions sorry!!													
18	Friday, 13 March 2020 at 17:50		Tourists must be informed before they board the plane, so that they have the option not to come at all. They are likely to be very aggressive towards the hotel front line st													

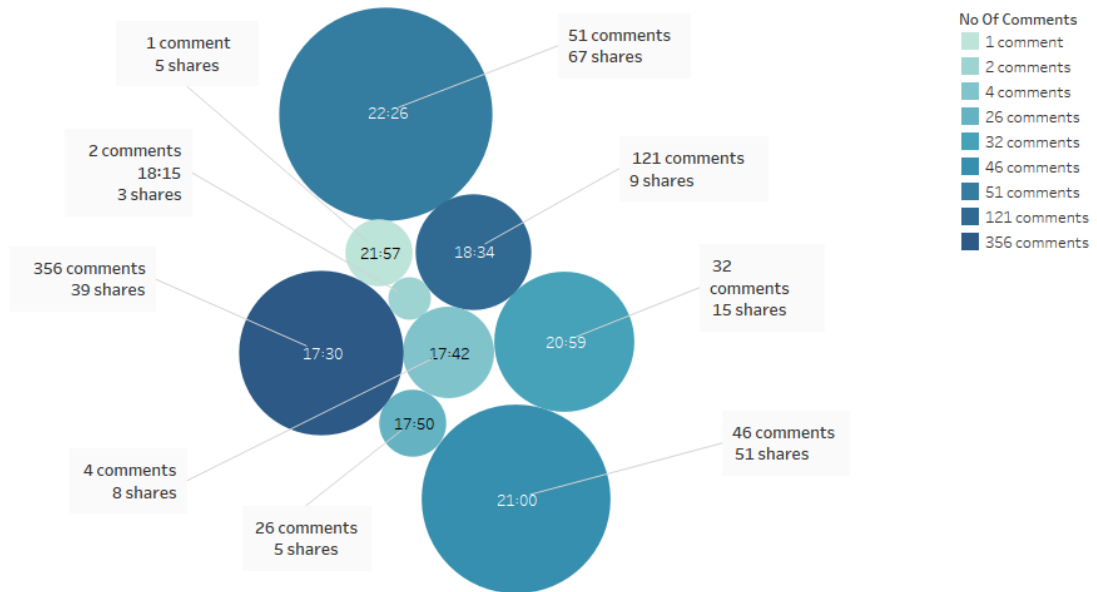
Figure 9 Merged Sheet

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Time (Comments)		Comment													
2	Friday, 13 March 2020 at 17:33	17:33	not really anything they can do about it they loose business short term. else 1000's get sick and they don't have workers to operate and then loose business...													
3	Friday, 13 March 2020 at 17:35	17:35	so what ? are these hoteliers expecting them to run around the streets, how about they invite them home for dinner and a glass of wine as well ?? at this moment every													
4	Friday, 13 March 2020 at 17:36	17:36	They shouldn't be! Health is the priority at this stage, measures are taken hour by hour. Only the irresponsible travel for holidays during this critical situation. People are													
5	Friday, 13 March 2020 at 17:36	17:36	Have they not had it so very good for a long time, what do they expect to infect more people to make them richer?													
6	Friday, 13 March 2020 at 17:38	17:38	It's immediate national security actions. What prior notice? Did schools, beauticians, gyms, restaurants, workers had prior notice?!													
7	Friday, 13 March 2020 at 17:39	17:39	Health of the nation comes first. No time for discussions!													
8	Friday, 13 March 2020 at 17:40	17:40	It-turists ser jehluha ukoll il-multa ta f'elf euro? Ghax filkas mil Birgu biss ghandu fejn idahhal mal-miljun euro													
9	Friday, 13 March 2020 at 17:40	17:40	Good AND bad timing for tourists about to travel from Bristol Many dismayed tourists choosing not to travel													
10	Friday, 13 March 2020 at 17:41	17:41	I truly depend on tourism actually I am jobless RightNow and for the time being yes it's the right decision													
11	Friday, 13 March 2020 at 17:41	17:41	It's not about the money, but doing quarantine in hotels is hard for the staff & workers working inbound hotels too. They're risking to contract													
12	Friday, 13 March 2020 at 17:42	17:42	Health first I'm sure you've had good times, many of us r out of work just suck it up and be thankful when u wake up healthy every morning!													
13	Friday, 13 March 2020 at 17:42	17:42	Really?! Business over national security?! Stop all nonessential activities and services. Now. Before we lose control over the situation. Stay inside. #stamatalaq#													
14	Friday, 13 March 2020 at 17:43	17:43	Tough luck! Our health comes first and foremost!													
15	Friday, 13 March 2020 at 17:43	17:43	No time for discussion in such situations!! Such directives are needed and tougher ones are coming													
16	Friday, 13 March 2020 at 17:45	17:45	After year after year boasting incredible success and tourism bla bla then they throw a tantrum because some necessary measures had to be taken.													
17	Friday, 13 March 2020 at 17:49	17:49	Health is a priority, the more discipline we are the more we get rid of the virus fast and get back on track! No time for discussions sorry!!													
18	Friday, 13 March 2020 at 17:50	17:50	Tourists must be informed before they board the plane, so that they have the option not to come at all. They are likely to be very aggressive towards the hotel front line st													

Figure 8 Unmerged Sheet

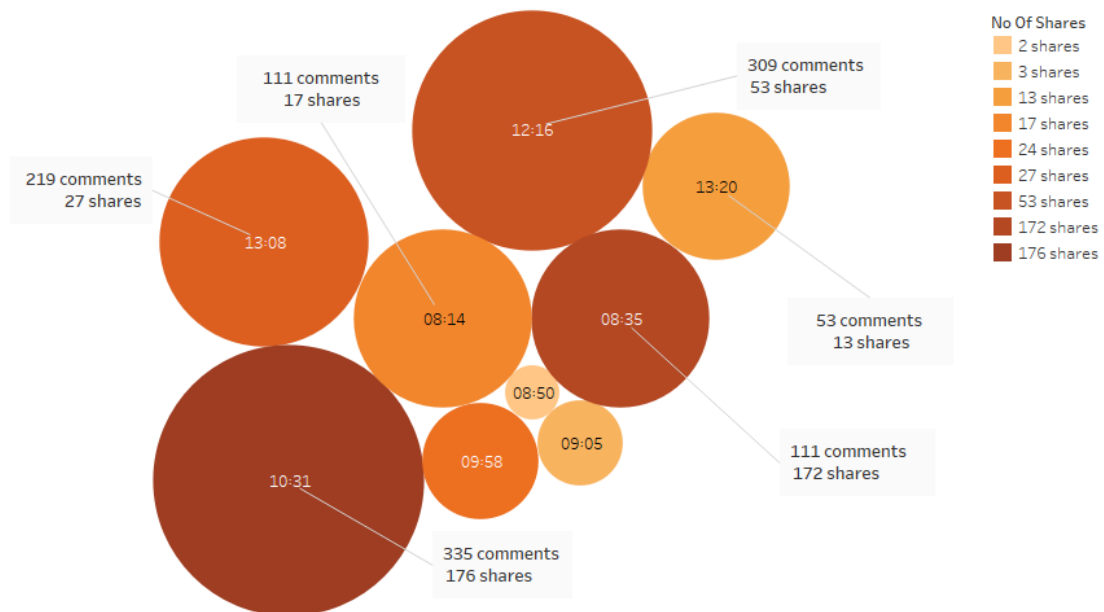
The daily Interactions on the posts are found in this section:

Friday 13th 2020



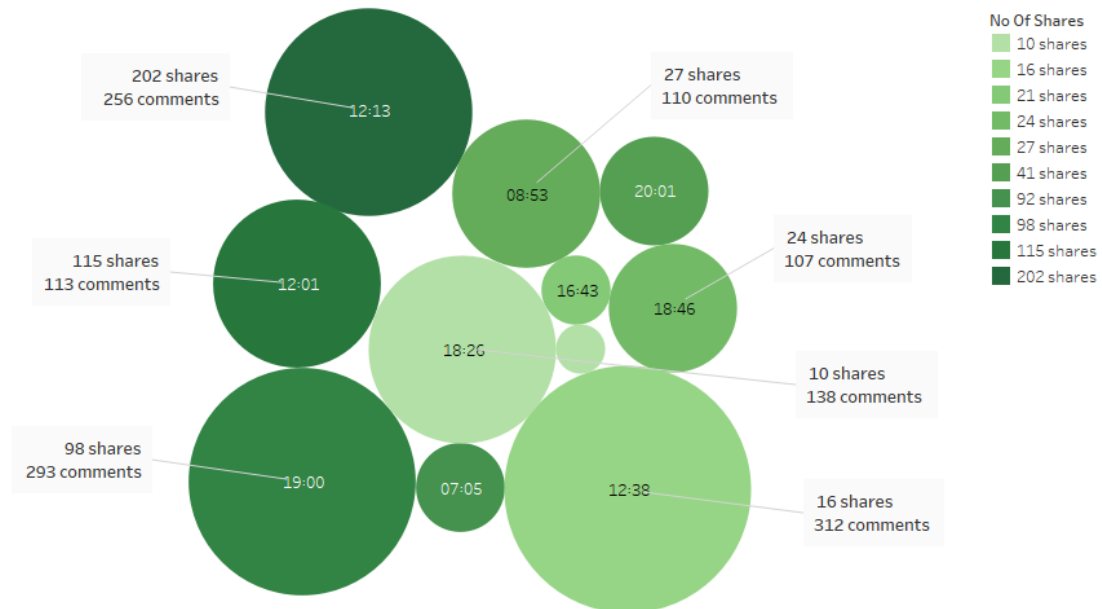
Time. Colour shows details about No Of Comments. Size shows details about No Of Shares as an attribute. The marks are labelled by Time.

Saturday 14th 2020



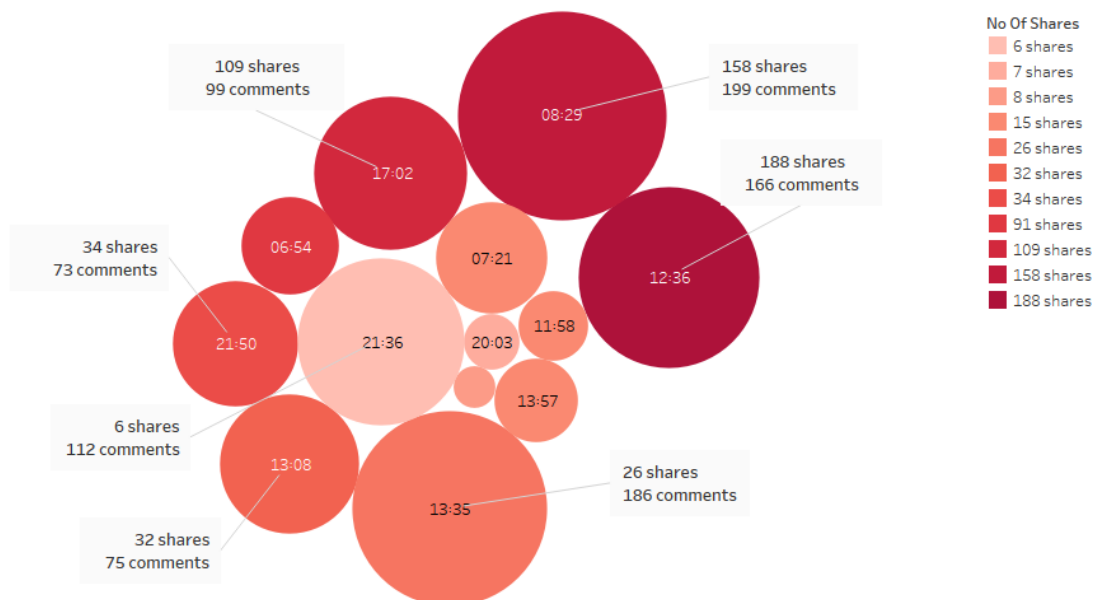
Time. Colour shows details about No Of Shares as an attribute. Size shows details about No Of Comments as an attribute. The marks are labelled by Time. The view is filtered on No Of Shares as an attribute, which excludes Null.

Sunday 15th 2020



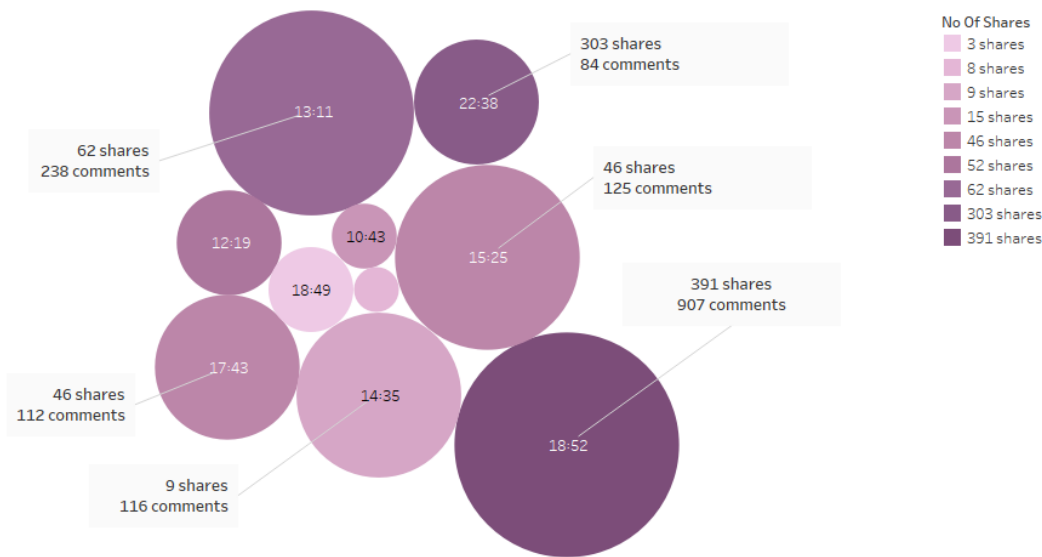
Time. Colour shows details about No Of Shares. Size shows details about No Of Comments as an attribute. The marks are labelled by Time. The view is filtered on No Of Shares, which excludes Null, 1 share, 12 shares and 3 shares.

Monday 16th 2020



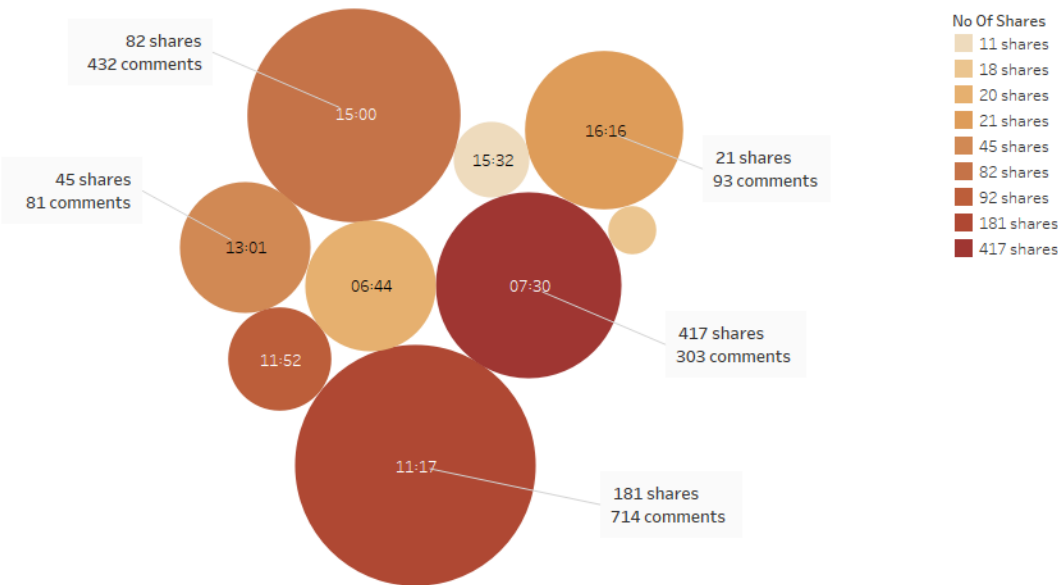
Time. Colour shows details about No Of Shares. Size shows details about No Of Comments as an attribute. The marks are labelled by Time. The view is filtered on No Of Shares, which excludes Null, 1 share, 2 shares, 3 shares and 4 shares.

Tuesday 17th 2020



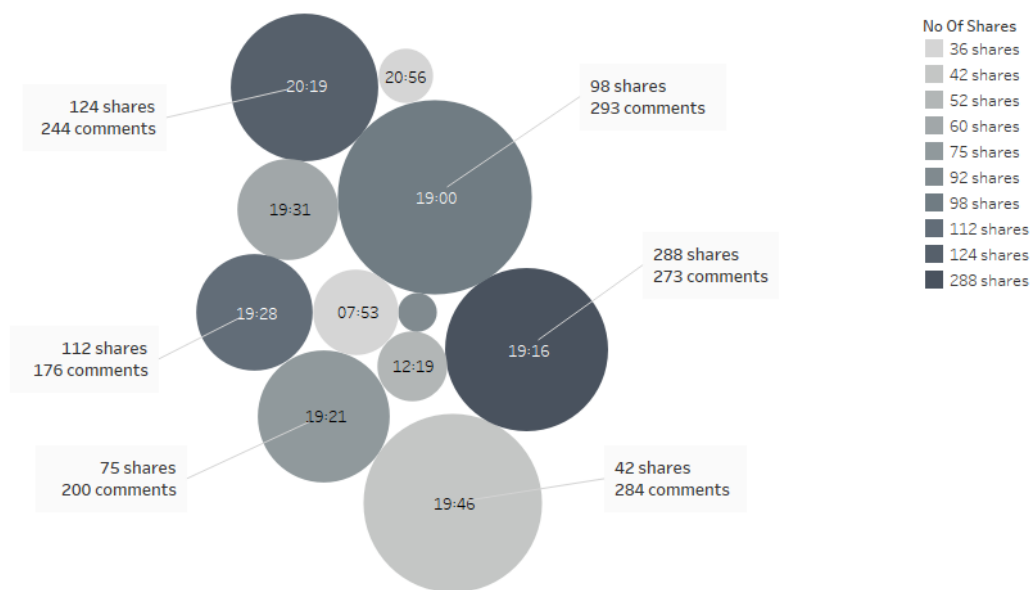
Time. Colour shows details about No Of Shares. Size shows details about No Of Comments as an attribute. The marks are labelled by Time. The data is filtered on No Of Comments and Exclusions (No Of Comments,Time). The No Of Comments filter keeps 10 of 22 members. The Exclusions (No Of Comments,Time) filter keeps 28 members.

Wednesday 18th 2020



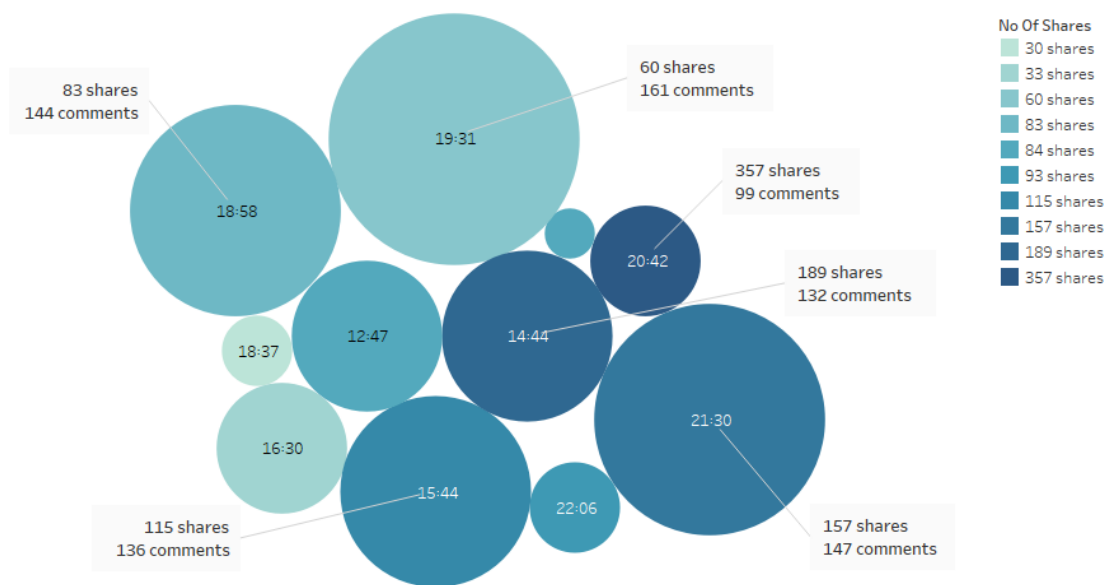
Time. Colour shows details about No Of Shares. Size shows details about No Of Comments as an attribute. The marks are labelled by Time. The data is filtered on No Of Comments, which excludes 10 members. The view is filtered on No Of Shares, which excludes 7 shares, 8 shares and 9 shares.

Thursday 19th 2020



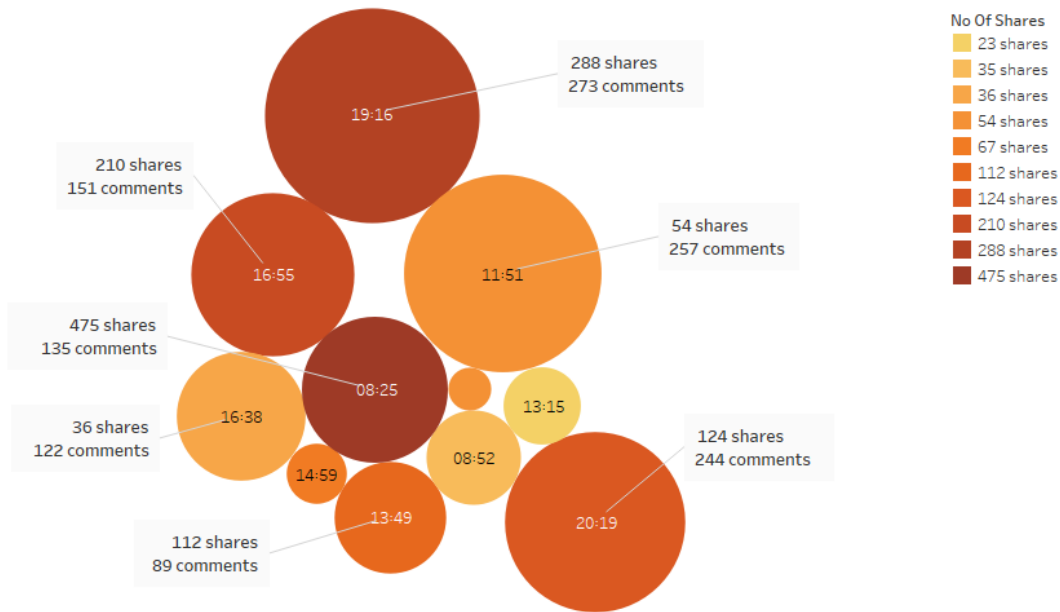
Time. Colour shows details about No Of Shares. Size shows details about No Of Comments as an attribute. The marks are labelled by Time. The view is filtered on No Of Shares, which keeps 10 of 31 members.

Friday 20th 2020



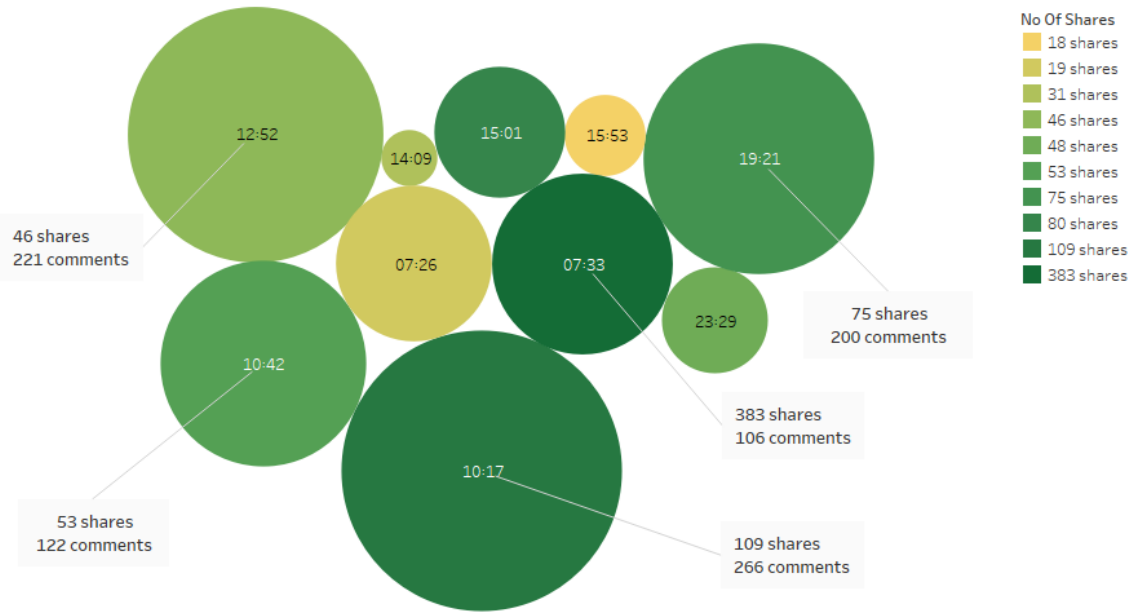
Time. Colour shows details about No Of Shares. Size shows details about No Of Comments as an attribute. The marks are labelled by Time. The view is filtered on No Of Shares, which keeps 10 of 22 members.

Saturday 21 2020



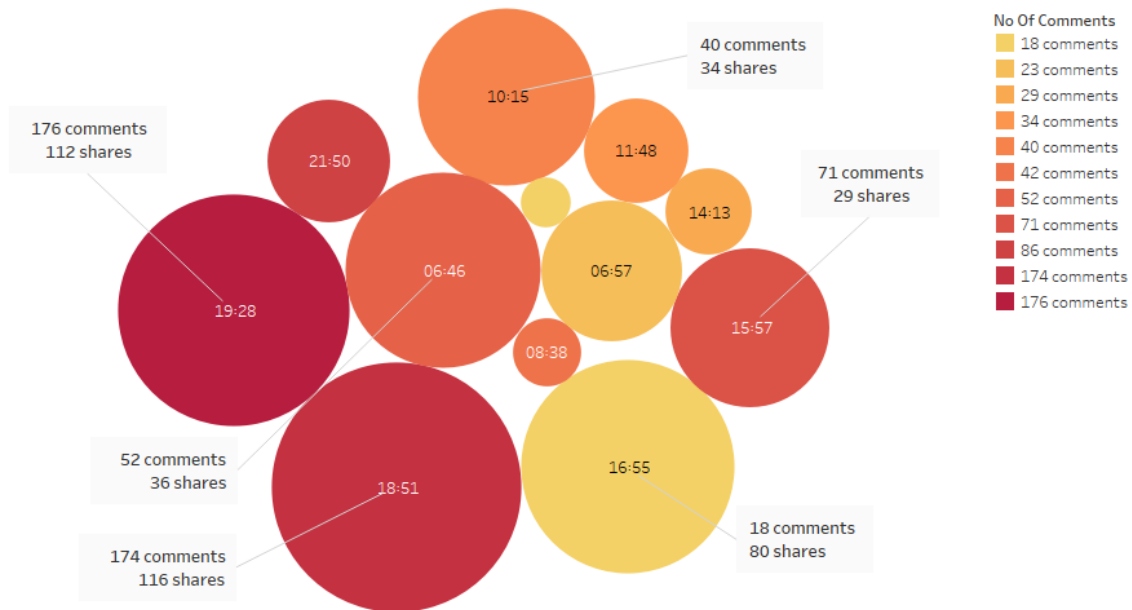
Time. Colour shows details about No Of Shares. Size shows details about No Of Comments as an attribute. The marks are labelled by Time. The view is filtered on No Of Shares, which keeps 10 of 21 members.

Sunday 22nd 2020



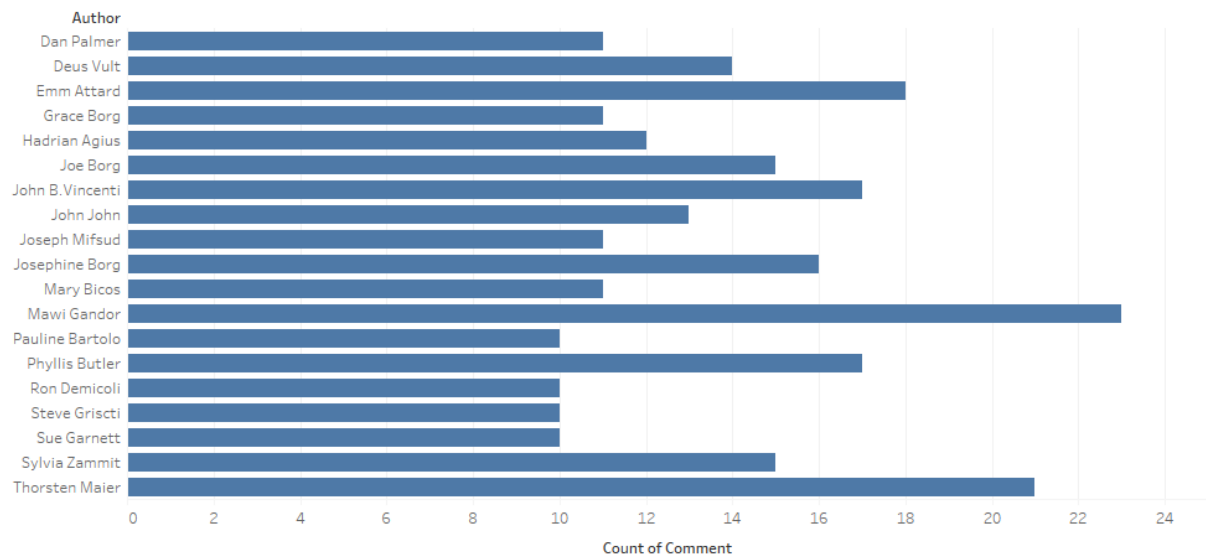
Time. Colour shows details about No Of Shares. Size shows details about No Of Comments as an attribute. The marks are labelled by Time. The view is filtered on No Of Shares, which keeps 10 of 21 members.

Monday 23rd 2020



Time. Colour shows details about No Of Comments. Size shows details about No Of Shares as an attribute. The marks are labelled by Time. The view is filtered on No Of Comments, which excludes 9 members.

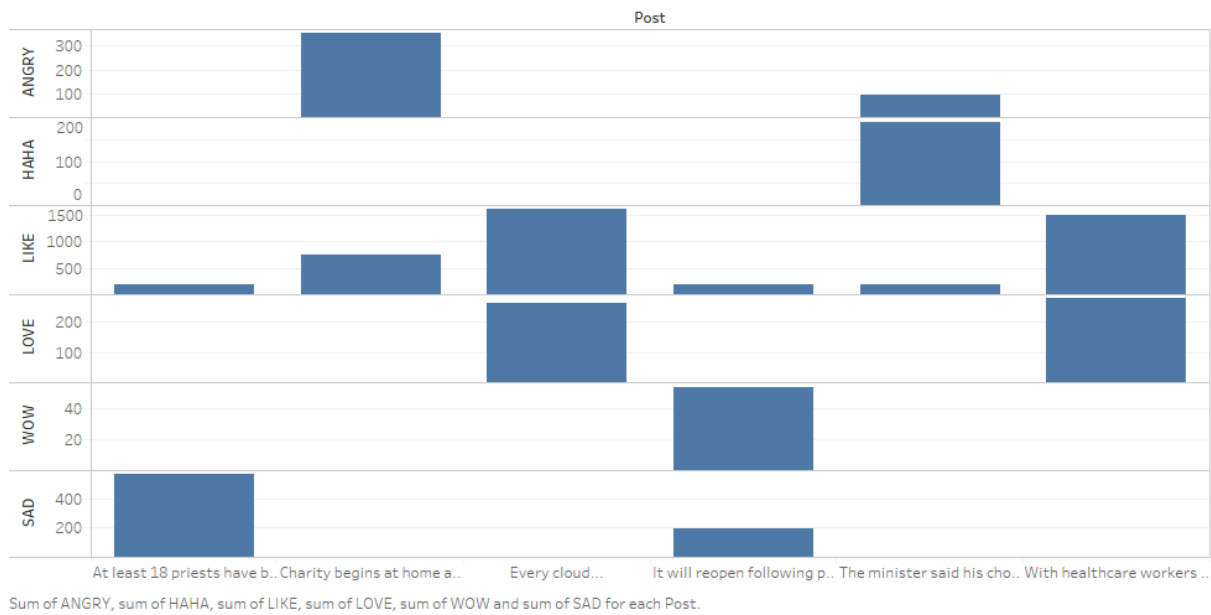
Top Commentors for the Period 13/03/2020 to 23/03/2020



Count of Comment for each Author. The view is filtered on Author and count of Comment. The Author filter excludes Null. The count of Comment filter includes values greater than or equal to 10.

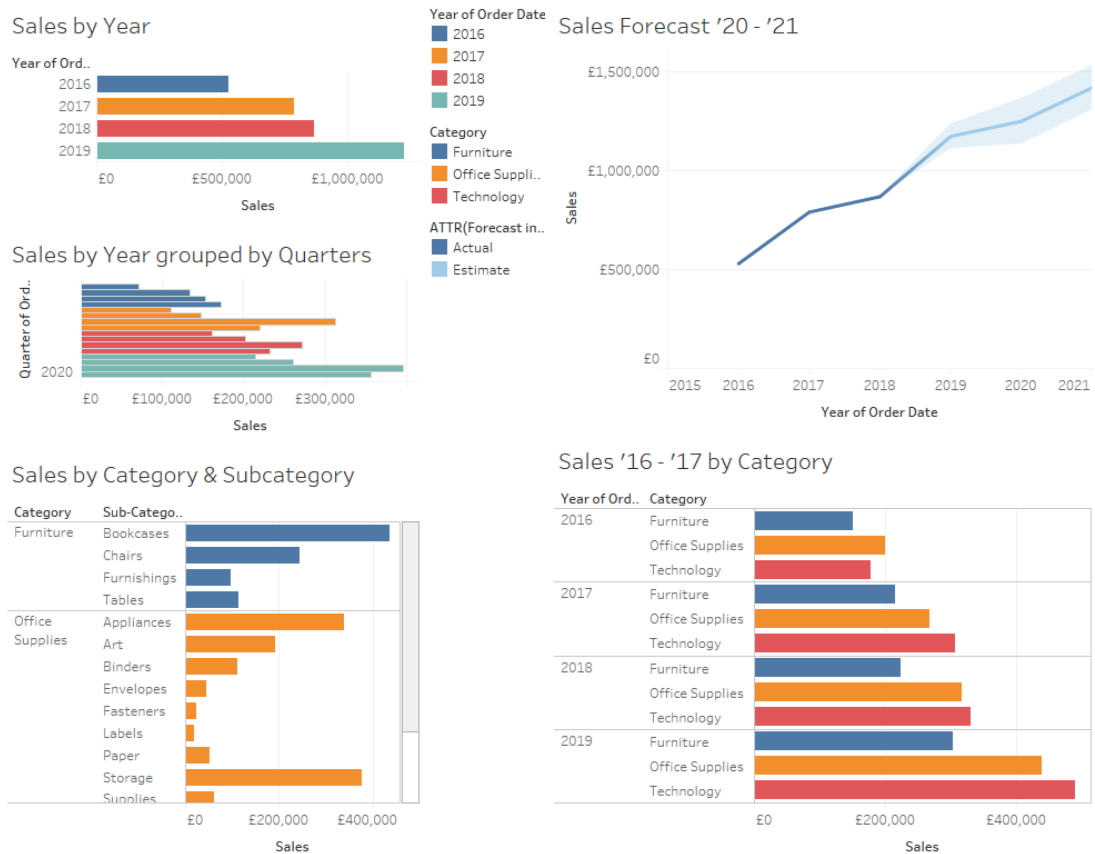
The last two graphs show the most active follower which is Mawi Gandor and the posts with the greatest number of different emotions are found in the last graph.

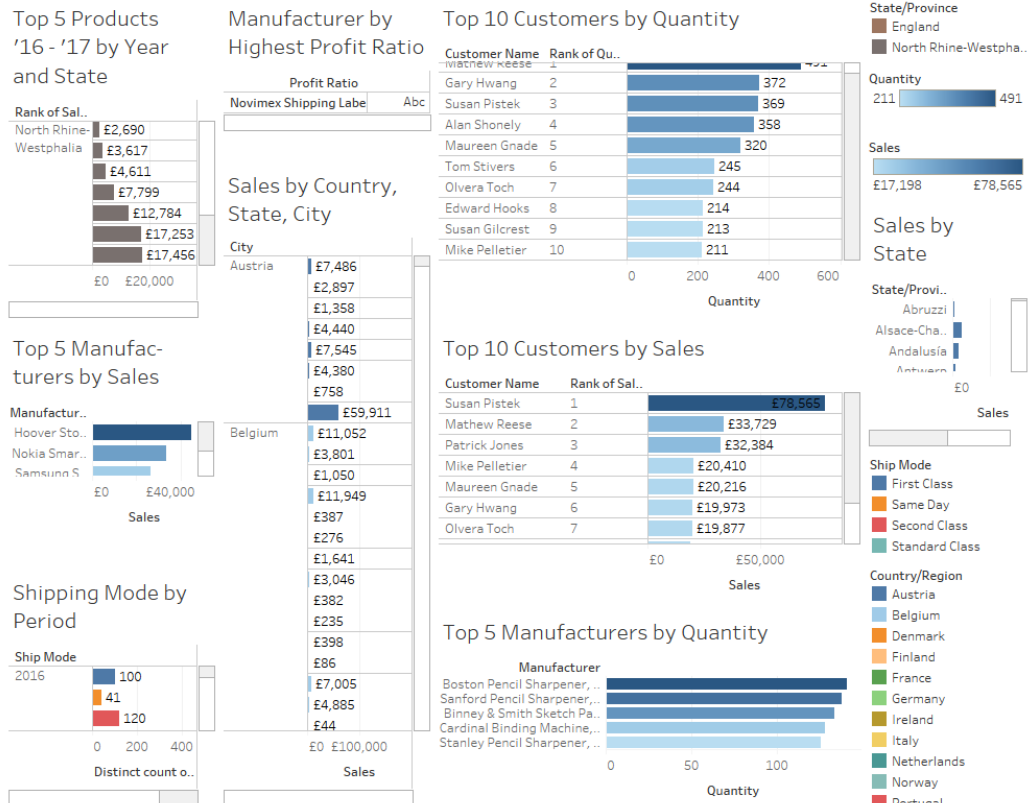
Top Posts By Emoticon



Section 3:

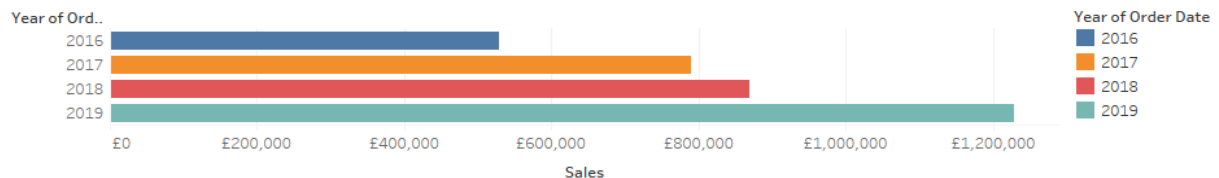
A web-based corporate dashboard for the queries using data provided (Sample – Superstore) is presented below.





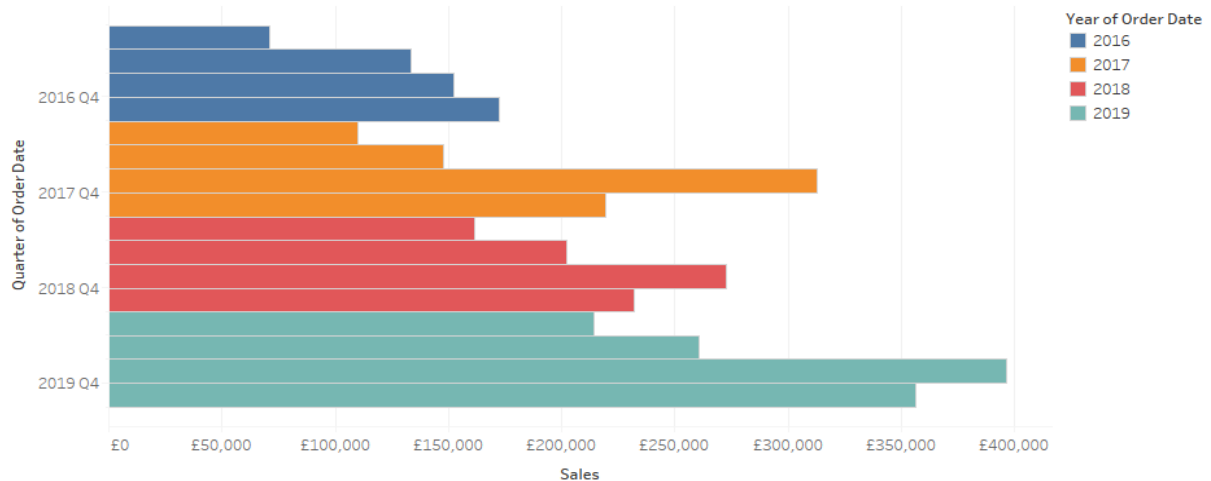
Data worksheets for the dashboard for reference are presented below:

Sales by Year



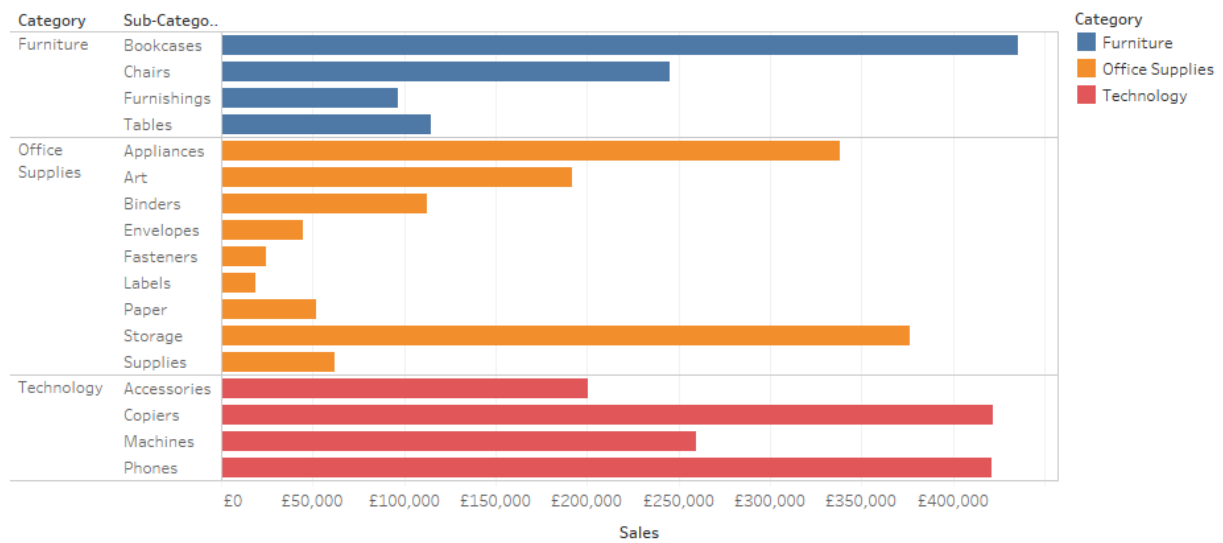
Sum of Sales for each Order Date Year. Colour shows details about Order Date Year.

Sales by Year grouped by Quarters

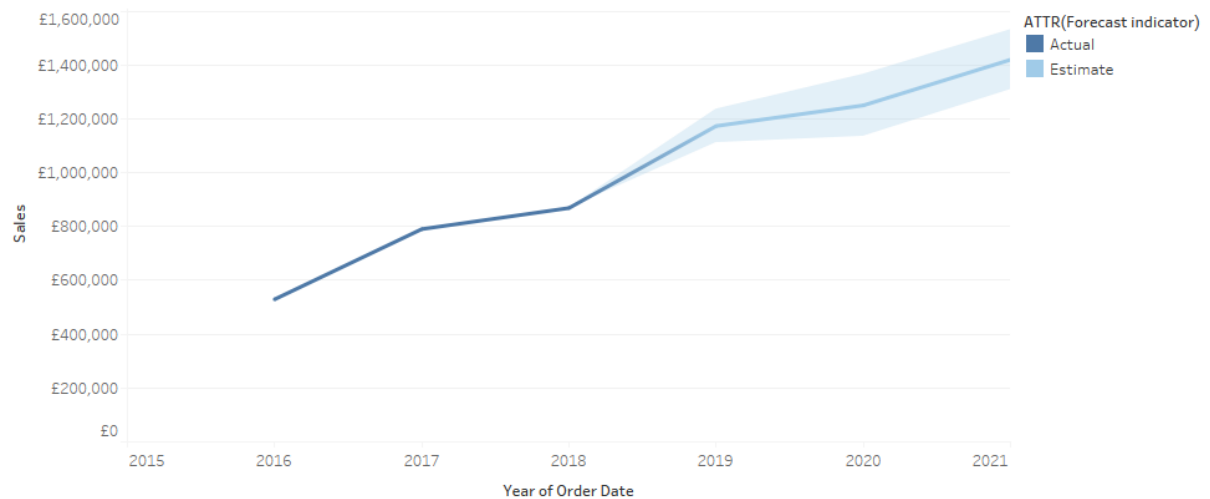


The plot of sum of Sales for Order Date Quarter. Colour shows details about Order Date Year.

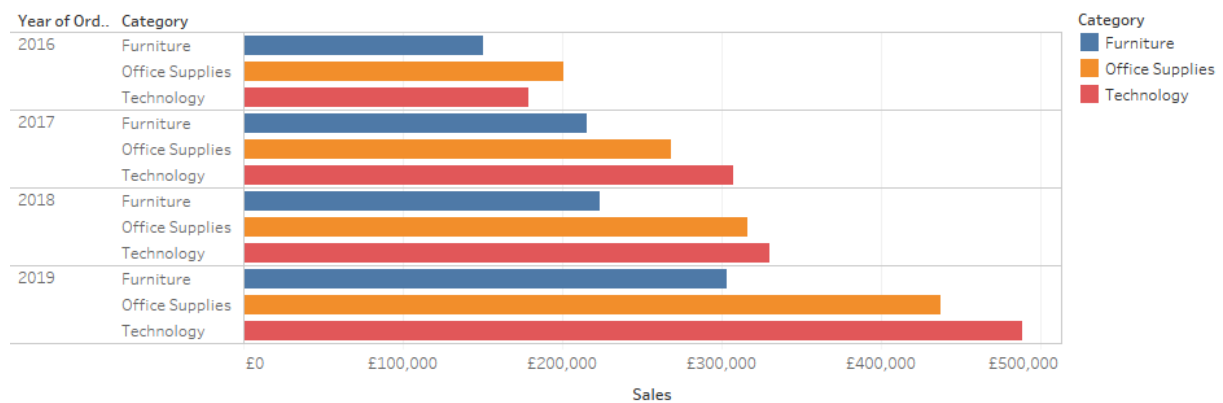
Sales by Category & Subcategory



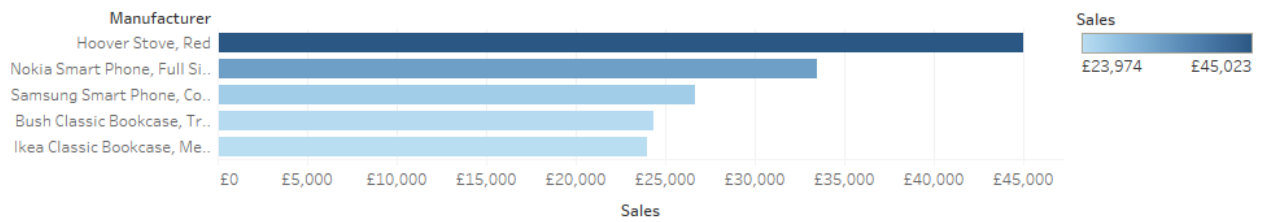
Sales Forecast '20 - '21



Sales '16 - '17 by Category

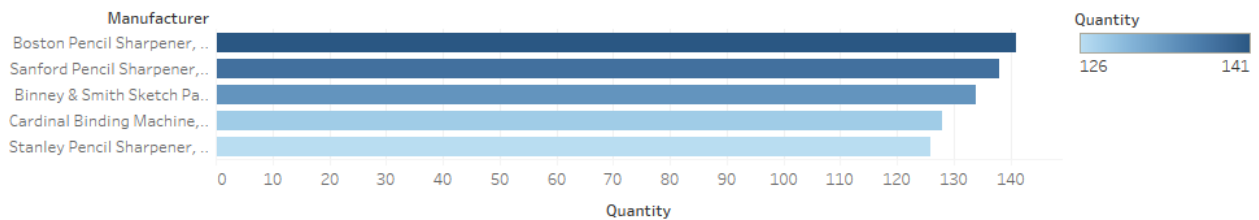


Top 5 Manufacturers by Sales



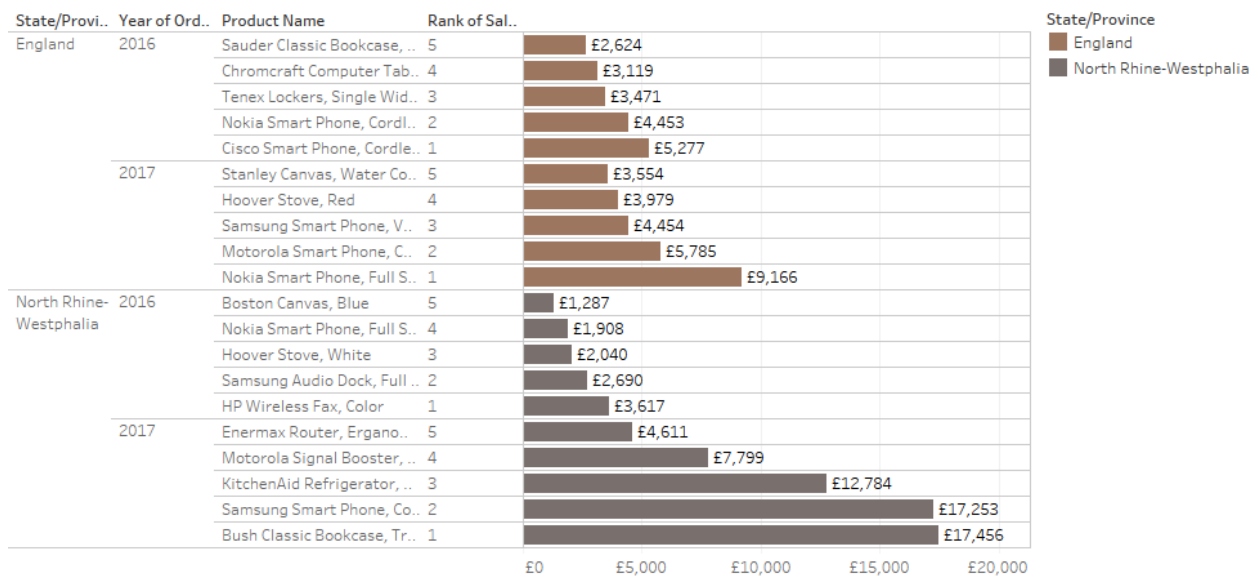
Sum of Sales for each Manufacturer. Colour shows sum of Sales. The view is filtered on Manufacturer, which has multiple members selected.

Top 5 Manufacturers by Quantity



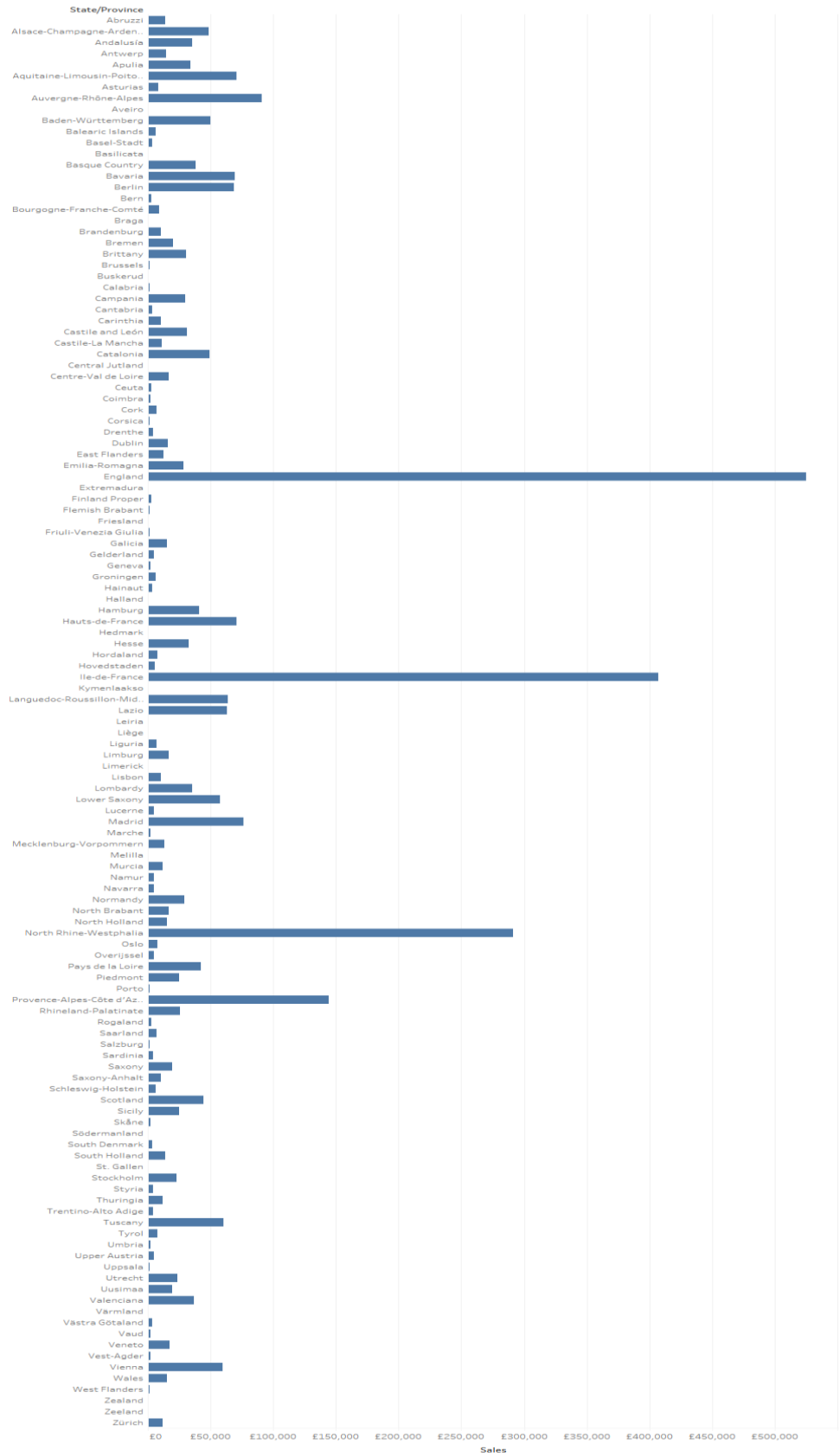
Sum of Quantity for each Manufacturer. Colour shows sum of Quantity. The view is filtered on Manufacturer, which has multiple members selected.

Top 5 Products '16 - '17 by Year and State

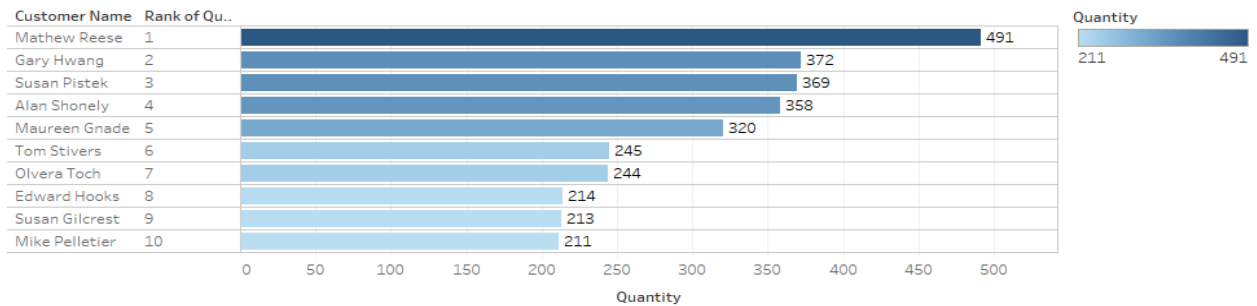


Sum of Sales for each Rank of Sales broken down by State/Province, Order Date Year and Product Name. Colour shows details about State/Province. The marks are labelled by sum of Sales. The view is filtered on Order Date Year, State/Province and Rank of Sales. The Order Date Year filter keeps 2016 and 2017. The State/Province filter keeps England and North Rhine-Westphalia. The Rank of Sales filter keeps 1, 2, 3, 4 and 5.

Sales by State

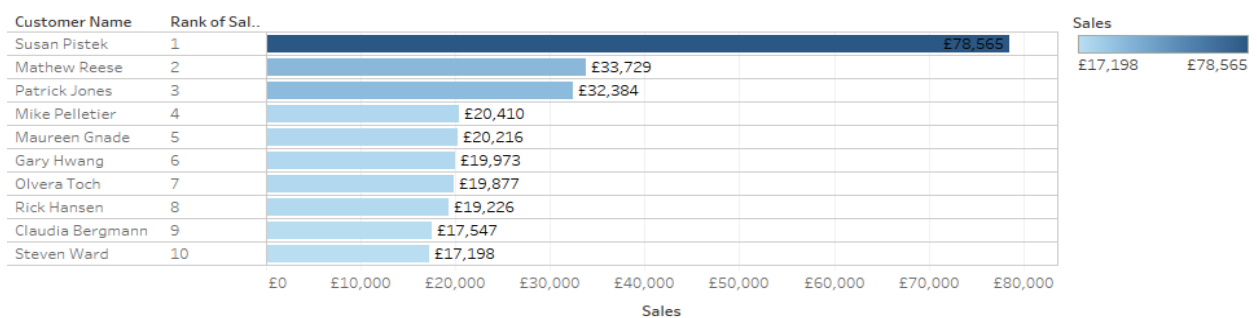


Top 10 Customers by Quantity



Sum of Quantity for each Rank of Quantity broken down by Customer Name. Colour shows sum of Quantity. The marks are labelled by sum of Quantity. The view is filtered on Rank of Quantity, which keeps 10 members.

Top 10 Customers by Sales



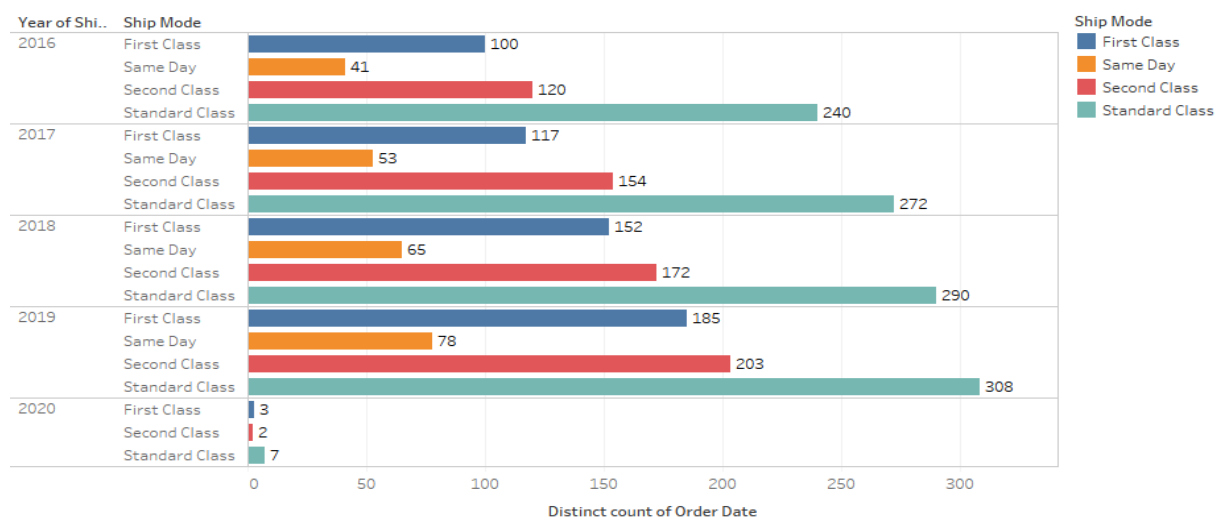
Sum of Sales for each Rank of Sales broken down by Customer Name. Colour shows sum of Sales. The marks are labelled by sum of Sales. The view is filtered on Rank of Sales, which keeps 10 members.

Manufacturer by Highest Profit Ratio

Manufacturer	Product Name	Rank of Profit Ratio	Profit Ratio	
Novimex Shipping Labels, Alphabetical	Novimex Shipping Labels, Alphabetical	1	50%	Abc

The view is broken down by Manufacturer, Product Name, Rank of Profit Ratio and Profit Ratio. The view is filtered on Profit Ratio and Rank of Profit Ratio. The Profit Ratio filter keeps 6 members. The Rank of Profit Ratio filter keeps 1.

Shipping Mode by Period



Distinct count of Order Date for each Ship Mode broken down by Ship Date Year. Colour shows details about Ship Mode. The marks are labelled by distinct count of Order Date.