

Text Analytics and Sentiment Analysis using SAP HANA

Venkata Sarath Chandra Muktevi (VXM180001)

Contents:

Objective	1
Create an App for Twitter Account	1
Configure the Twitter adapter for SAP HANA system	5
Loading the twitter data using Virtual Function	7
Rate Limit Status table	16
Full text index	19
Data Cleaning	26
Sentiment Analysis and Data visualization using SAP Lumira	30
Create an Analytical view	30
Create a Tag Cloud with SAP Lumira Server	34
Create a Network Chart with SAP Lumira Server	38
Create a Pie Chart with SAP Lumira Server	41
Exporting the file	45

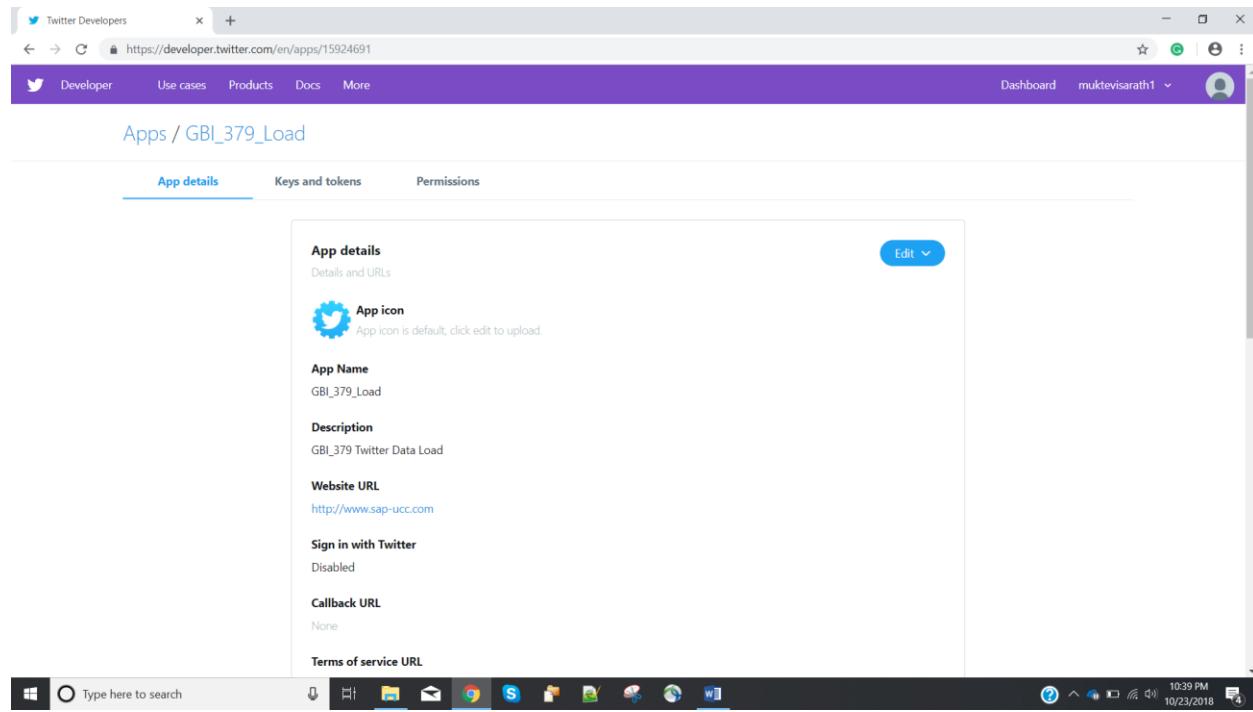
Objective

To perform Text Analytics and Sentiment Analysis using SAP HANA and data visualization using SAP Lumira.

Every day, Global Bike Inc. produces thousands of bikes. Currently, the Shop Floor workers have to walk to the working stations, which are fixed to a certain place, in order to scan their products and to confirm that a new bike was completed. This is not only a waste of time which costs a lot of money, but also an exhausting part of the workers' daily routine, which can make up to several kilometers a day just to confirm their charges. To counteract these vice president (Operations) engages production planner to do some research about various mobile devices as an alternative to the fixed working stations. Production planner is really enthusiastic about this idea and being a passionate Twitter user, he decides to do some analysis on Twitter data first, to get to know the opinions of users about the new iPad Pro and the Surface Pro 4. At lunch he tells his systems & database administrator about his ideas. Database administrator is always up to date on the newest technologies and suggests production planner to use GBI's SAP HANA system for analyzing the data, as there is a brand new functionality for loading Twitter Data into the database.

1) Create an App for Twitter Account

After creating an App for Twitter Account, we get the following confirmation:



The screenshot shows the Twitter Developers App Details page for an application named 'GBI_379_Load'. The page is divided into sections: App details, Keys and tokens, and Permissions. Under 'App details', there is an 'Edit' button, an 'App icon' placeholder (a blue Twitter bird), an 'App Name' field containing 'GBI_379_Load', a 'Description' field containing 'GBI_379 Twitter Data Load', a 'Website URL' field with the value 'http://www.sap-ucc.com', a 'Sign in with Twitter' section set to 'Disabled', a 'Callback URL' field with 'None', and a 'Terms of service URL' field. The browser's address bar shows the URL 'https://developer.twitter.com/en/apps/15924691'. The operating system taskbar at the bottom includes icons for Start, Task View, File Explorer, Mail, Google Chrome, and others, along with a system clock showing '10:39 PM 10/23/2018'.

We need to go to tab "Keys and Access Tokens" to create the Access Tokens for our app.

We can see that a Consumer Key, also known as API Key, and a Consumer Secret, also known as API Secret, have already been created. For accessing Twitter data via the Twitter Adapter we also need an Access Token and Access Key.

To create those scroll to the bottom of the page and click on **Create my access token**.

[Details](#)[Settings](#)[Keys and Access Tokens](#)[Permissions](#)

Application Settings

Keep the "Consumer Secret" a secret. This key should never be human-readable in your application.

Consumer Key (API Key)

LVgd2[REDACTED]

Consumer Secret (API Secret)

gDFJ7[REDACTED]

Access Level

Read and write ([modify app permissions](#))

Owner

[REDACTED]

Owner ID

[REDACTED]

Application Actions

[Regenerate Consumer Key and Secret](#)[Change App Permissions](#)

4

5

Your Access Token

You haven't authorized this application for your own account yet.

By creating your access token here, you will have everything you need to make API calls right away. The access token generated will be assigned your application's current permission level.

Token Actions

[Create my access token](#)

We will get a confirmation that the access tokens have been created.

The screenshot shows the Twitter Developers website interface. At the top, there's a navigation bar with links for Developer, Use cases, Products, Docs, More, Dashboard, and a user profile for 'muktevisarath1'. Below the navigation is a sub-header 'Apps / GBI_379_Load'. Underneath, there are tabs for 'App details' (which is currently inactive) and 'Keys and tokens' (which is active). The 'Keys and tokens' section contains two main sections: 'Consumer API keys' and 'Access token & access token secret'. Under 'Consumer API keys', there are two entries: '8eM4ewVkc978VLMIhohulwOazh (API key)' and 'nMt5AbZxig6NR6OBLVgpwsMFMNbdx1FCTesqubG3LLUNrIxVG (API secret key)'. Each entry has a 'Regenerate' button. Under 'Access token & access token secret', there is one entry: '3288964958-IclazPbZvWLRyv1RCkuXGVSnGHlyzpHfqoVB (Access token)' and 'iqGr3mqDRW9PfFDzvdxOQiCn5mMzkAKqZuWuZHhBk2 (Access token secret)'. This entry also has 'Regenerate' and 'Revoke' buttons. At the bottom of the page, there are links for 'Developer policy and terms', 'Follow @twitterdev', and 'Subscribe to developer news'. The browser's address bar shows the URL 'https://developer.twitter.com/en/apps/15924691'. The operating system taskbar at the bottom shows various pinned icons and the date/time '10:51 PM 10/23/2018'.

2) Configure the Twitter adapter for SAP HANA system

For loading data from twitter into the SAP HANA database, we need to configure a Remote Source for the Twitter adapter.

Open the Catalog of your system in a browser using the Web IDE.

<http://hn1.hana1.ucc.uwm.edu:8000/sap/hana/ide/catalog/>

Then go to:

Provisioning ► Remote Sources.

Right click on folder Remote Sources and choose **New Remote Source**.

A new screen for configuring a remote source opens.

Enter the following data:

Source Name	GBI_379
Adapter Name	TwitterAdapter
Location	agent
Agent Name	dpagent

Now clicking on the + symbol next to API Key, API Secret, Access Token and Access Token Secret to elapse the fields for entering the keys.

We need to go to your Twitter app and copy and paste the API Key, API Secret, Access Token and Access Token Secret into the corresponding fields.

Clicking on Save will create the remote source. Our remote source was created successfully, and got a confirmation and it is shown on the left side under folder “Remote Sources”.

The screenshot shows the configuration of a remote source named 'GBI_379' using the 'TwitterAdapter'. The 'Credentials Mode' is set to 'Technical User'. The 'API Keys' section is expanded, showing four optional fields: 'API Key (Consumer Key)', 'API Secret (Consumer Secret)', 'Access Token (Oauth Token)', and 'Access Token Secret (Oauth Secret)'. The status bar at the bottom displays log messages indicating missing mandatory fields for Server, Port, and Database Name, followed by a success message for saving the source.

When you elapse your Remote Source now, you can see that there are two tables for loading data, Status and Rate_Limit_Status.

The table “Status” contains your home_timeline and can be used for requesting user’s timelines or searching via the Twitter API.

The table “Rate_Limit_Status” gives you information about your rate limits. In order to control the use of the Twitter API Twitter has a limit per user how often the API can be used per hour. The limit is divided into 15 minute intervals. Depending on the method you are using, you can call it up to 15 times per 15 minutes or up to 180 times per 15 minutes. Strategies to avoid exceeding the rate limit are caching, prioritizing active users or using an application-only auth.

3) Loading the twitter data using Virtual Function

We need to get an insight as to how the Twitter Adapter of SAP HANA Data Provisioning works. Therefore, load the user timeline of Microsoft to find out how the data looks like.

We want to have a look on the timelines of Apple Inc. and Microsoft to find information that could be interesting for Sentiment Analysis.

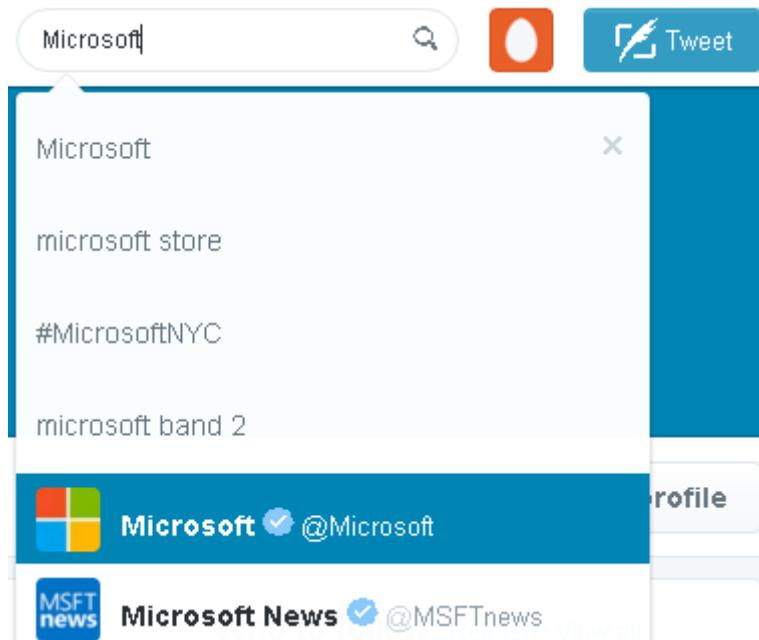
Go to twitter.com and login to your account. In the upper right corner there is a search field to search for users or for any terms. Type in “Apple” or “Microsoft” to find Apple’s and Microsoft’s official user timeline.

When you type in “Apple”, you will find some results, but please be careful. Apple has an official account on twitter with user name “@Apple”, however, this isn’t broadly used and has just a few followers. Apple owns several Twitter accounts that represent different divisions of the company, e.g. @iTunes or @AppStore. That’s why you won’t be able to find a timeline for the whole company Apple Inc. with useful data.

The next screenshot shows the result that you will receive when searching for the term “Apple”.

The screenshot shows the Twitter search interface. At the top, there is a search bar with the word "Apple" typed into it, followed by a magnifying glass icon and a blue "Tweet" button. Below the search bar is a list of search suggestions: "Apple", "Apple TV", "apple tv order", and "Apple Watch". The main search results are displayed below these suggestions. The first result is "Apple Music" (@AppleMusic), which is verified (@verified). The second result is "Apple" (@APPLEOFFICIAL), which is not verified (@unverified). The third result is "christina applegate" (@1capplegate), which is verified (@verified). The fourth result is "Applebee's" (@Applebees), which is verified (@verified). The fifth result is "Apple News" (@applenws), which is not verified (@unverified). At the bottom of the screen, there is a link to "Search all people for Apple".

The first entry with the name “@Microsoft” is the official twitter account of Microsoft. Click on it to open the user timeline.



The screenshot on the next page shows Microsoft's timeline.



A diver wearing a helmet and goggles, looking up from under water towards the camera.

Microsoft 

@Microsoft
The official Twitter page for Microsoft consumer products and your source for major announcements and events.
📍 Redmond, WA
🔗 facebook.com/Microsoft
Joined September 2009

TWEETS 11.2K **FOLLOWING** 1,662 **FOLLOWERS** 7.24M **FAVORITES** 152 **LISTS** 4

Follow

Tweets **Tweets & replies** **Photos & videos**

 Microsoft Retweeted
Xbox @Xbox · Oct 24
Sixty bucks to play.
Zero bucks to watch.
Join #Halo5  [T] LIVE on 10/26 at 3pm PT: xbl.lv/1hGwrbX

 **Xbox**



Who to follow · Refresh · View all

 heise Top-News @heise_to...
Follow

 openSAP @openSAP
Follow

 TechStage  @techstagede
Follow

Find friends

Trends · Change

#earthquake
#LouisYouMakeUsFeelLikeHome
Chhota Rajan
#Stevens
Peshawar
#qanda
Srinagar
#fbpokal

Scroll down to find some interesting tweets about the Surface Pro, which is shown in the screenshot below.

 Microsoft Retweeted
CNET @CNET · Oct 21
The new Surface Pro 4 keeps Microsoft at the top of the tablet hill (rated review) cnet.co/1PA4tEP



◀ ▶ 271 ⭐ 347 ...

To get an insight how the loaded data looks like, you decide to load the Microsoft's user timeline into your database.

Open the SAP HANA Web IDE in your browser. For this, please adapt the following link to your system:

<http://hn1.hana1.ucc.uwm.edu:8000/sap/hana/ide/editor/>

Login with your user credentials.

Surface Pro

We will first list down the tweets which are desired or not. We are trying to find the tweets for Surface Pro 4.

Desired	Undesired
#SurfacePro4	#SurfacePro3
#surfacepro4	#Surface
#SURFACEPRO4	Surface
SurfacePro4	#Pro4
	Pro4
	#SurfacePro
	SurfacePro

Right click on the Tweets under Search and click on New Virtual Function.

The success message after creating virtual function GetSearchTweets is:

The screenshot shows the SAP HANA Web-based Development Workbench Catalog interface. On the left, there is a tree view of database objects, with a red box highlighting the 'Tweets' node under the 'Search' category. The main panel displays the configuration for a virtual function named 'GBI_379'. The 'Source Name' is set to 'SURFACEPRO4_379.sql', the 'Adapter Name' is 'TwitterAdapter', the 'Location' is 'agent', and the 'Agent (Group) Name' is 'dpagent'. The 'Credentials' section shows 'Technical User' selected for 'Credentials Mode'. A message at the bottom of the screen states: '11:52:17 PM (Catalog) Created new virtual function 'GBI_379'.'GetSearchTweets' successfully'.

Now click on the SQL Icon.

The SQL console will get open.

Copy paste the below given SQL query on the query console.

```
CREATE COLUMN TABLE "GBI_379"."SURFACEPRO4_379" (Id BIGINT PRIMARY KEY, ScreenName NVARCHAR(256), Tweet NVARCHAR(256), Source NVARCHAR(256), Truncated TINYINT, InReplyToStatusId BIGINT, InReplyToUserId BIGINT, InReplyToScreenName NVARCHAR(256), Favorited TINYINT, Retweeted TINYINT, FavoriteCount INTEGER, Retweet TINYINT, RetweetCount INTEGER, RetweetedByMe TINYINT, CurrentUserRetweetId BIGINT, PossiblySensitive TINYINT, isoLanguageCode NVARCHAR(256), CreatedAt TIMESTAMP, Latitude DOUBLE, Longitude DOUBLE, Country NVARCHAR(256), Place_name NVARCHAR(256), Place_type NVARCHAR(256), UserId BIGINT, UserName NVARCHAR(256), UserUrl NVARCHAR(256));
```

Save the query under gbi-student-379 as given below. Name the query as SURFACEPRO4_379.sql.

Click on the execute button.

The success message after saving the SURFACEPRO4_379.sql query and clicking on execute button is:

```
11:58:41 PM (SQL File) File 'SURFACEPRO4_379.sql' has been saved.  
11:58:48 PM (SQL Editor) Statement 'CREATE COLUMN TABLE "GBI_379"."SURFACEPRO4_379" (Id BIGINT PRIMARY KEY, ScreenName NVARCHAR(256), ...'  
successfully executed in 7 ms.
```

Now we need to get the search tweet into the SURFACEPRO4_379 table from GetSearchTweets table.

For that we will create the query based on the Desired and Undesired tweets which we have listed down.
Please find below the query:

```
UPSERT "GBI_379"."SURFACEPRO4_379" SELECT * FROM "GBI_379"."GetSearchTweets"('#SurfacePro4 OR #surfacepro4 OR #SURFACEPRO4 OR SurfacePro4 -#SurfacePro3 -#Surface -Surface -#Pro4 -Pro4 -#SurfacePro -SurfacePro', 1500, null, null, null, null, null, null, null);
```

Now we will save this query. Save the query under gbi-student-379. Name of the query should be SURFACEPRO4_GETRESULT_379.sql.

Now execute the query by clicking on Run button.

The success message will come.

Right click on the table and select Open Content.

The below given table will come:

After opening content of SURFACEPRO4_379 table, the screenshot is:

The screenshot shows the SAP HANA Web-based Development Workbench interface. The left sidebar displays the database catalog structure, including the GBI_379 schema and its various tables like GBI_379, SURFACEPRO4_379, and SURFACEPRO4_GETRESULT_379. The main area is a data grid showing the results of the SURFACEPRO4_379 table. The columns are: ID, SCREENNAME, TWEET, SOURCE, TRUNCATED, INREPLYTOSTA..., INREPLYTOUS..., and INREPLYTOS. There are 19 rows of data, each representing a tweet. The data includes various users like AldinGargari, mikeymca, christibzco, etc., with their respective tweets and source URLs. The interface also shows system status like v 1.120.30 | Help | GBI_379 | HN1 | HN1 (hana1 00) |

ID	SCREENNAME	TWEET	SOURCE	TRUNCATED	INREPLYTOSTA...	INREPLYTOUS...	INREPLYTOS
1	1055244399954877695	AldinGargari	CHOP. Inktober day 24. # <a href="http://twitter.com/	1	-1	-1	?
2	1055137154047922176	mikeymca	As an artist you can only z <a href="http://twitter.com/	1	-1	-1	?
3	105512547183347712	christibzco	Hey microsoft x microsft <a href="http://instagram.c	1	-1	-1	?
4	1054954399440943424	SamTheOilMan	My replacement surfaceq <a href="http://twitter.com/	1	-1	-1	?
5	1054948355053441026	k1numata	SurfacePro4に慣れてしま <a href="http://twitter.com/	0	-1	-1	?
6	105472572250222976	AldinGargari	MUDGY. Inktober day 23. <a href="http://twitter.com/	1	-1	-1	?
7	1054706331969487617	akirap72	@h_natano 私のSurfaceP <a href="http://twitter.softa	0	105491430952312832	1407863244	h_natano
8	105467071659509056	eloytne	RT @ukyone: New Digital <a href="http://twitter.com/	0	-1	-1	?
9	1054615914417717248	ukyone	New Digital Painting #Anr <a href="http://twitter.com/	0	-1	-1	?
10	1054599494665719296	BTFCX59999931	SurfacePro4でProに少 <a href="http://twitter.com/	0	-1	-1	?
11	105439973179412480	shichi_6	@SurfaceJP SurfacePro4 <a href="http://twitter.com/	1	-1	1201010054	SurfaceJP
12	105422242107392000	sketchcharitymarc	Solo... #sketch_illustrat <a href="http://twitter.com/	1	-1	-1	?
13	1054398693111597058	AldinGargari	EXPENSIVE. Inktober day <a href="http://twitter.com/	1	-1	-1	?
14	1054039437288517637	ImAndreiMcKey	If you're into #Futur or # <a href="http://twitter.com/	1	-1	-1	?
15	1054038311499694080	AldinGargari	DRAIN. Inktober day 21. # <a href="http://twitter.com/	1	-1	-1	?
16	1054024605281673216	tatataata0405	@th_dendon 中古のSurfa <a href="http://twitter.com/	0	1054024000832061440	1005059134033498112	th_dendon
17	1053996819954704385	mgmrgfuits	高台からノイ ラ撮影メキ <a href="http://twitter.com/	0	-1	-1	?
18	105397197571936000	keruru	あら、うちのSurfacePro4 <a href="https://about.twimg	0	-1	-1	?
19	1053924546121140224	akirap72	@yodafox SurfacePro4で <a href="https://about.twimg	0	1053924238908768256	128816967	yodafox

IPad Pro

We will first list down the tweets which are desired or not. We are trying to find the tweets for IPad Pro.

Desired	Undesired
#IPadPro	#IPad
#ipadpro	IPad
#IPADPRO	#Pro
IPadPro	Pro
	AppleIPad

Right click on the Tweets under Search and click on New Virtual Function. Choose the name of virtual function as GetSearchTweetsIPad.

The success message after creating virtual function GetSearchTweetsIPad is:

ID	SCREENNAME	TWEET	SOURCE	TRUNCATED	INREPLYTOSTA...	INREPLYTOUS...	INREPLYTOS
1	AldinGargari	CHOP. Inktober day 24. #	<a href="http://twitter.com"	1	-1	-1	?
2	mikeymcna	As an artist you can only a	<a href="http://twitter.com"	1	-1	-1	?
3	chrisblazeezo	Hey microsoft x micros	<a href="http://instagram.c	1	-1	-1	?
4	SamTheOilMan	My replacement surface#	<a href="http://twitter.com"	1	-1	-1	?
5	k1nmuta	SurfacePro4からPro6に	<a href="http://twitter.com"	0	-1	-1	?
6	AldinGargari	MUDGY. Inktober day 23.	<a href="http://twitter.com"	1	-1	-1	?
7	akirap72	@h_natano 私のSurfaceP	<a href="http://twitter.sofa	0	1054591430952312832	1407653244	h_natano
8	elodyne	RT @ukyoner: New Digital	<a href="http://twitter.com"	0	-1	-1	?
9	RT@ukyoner	New Digital Painting #Art	<a href="http://twitter.com"	0	-1	-1	?
10	SurfacePro4からPro6に	SurfacePro4からPro6に	<a href="http://twitter.com"	0	-1	-1	?
11	shichi_6	@SurfaceJP SurfacePro4	<a href="http://twitter.com"	1	-1	1201010054	SurfaceJP
12	sketchbymarc	Solo... #sketch #illustrati	<a href="http://twitter.com"	1	-1	-1	?
13	AldinGargari	EXPENSIVE. Inktober da	<a href="http://twitter.com"	1	-1	-1	?
14	ImAndreiMcKay	If you're into #OfferUp or #	<a href="http://twitter.com"	1	-1	-1	?
15	DRAIN. Inktober day 21. #	DRAIN. Inktober day 21. #	<a href="http://twitter.com"	1	-1	-1	?
16	tatataaaaa0405	@th_dendon 中古のSurface	<a href="http://twitter.com"	0	1054024000832061440	1005059134033496112	th_dendon
17	mgmghuts	高台からノイタマ撮影メキ	<a href="http://twitter.com"	0	-1	-1	?
18	keruru	あら、うちのSurfacePro4	<a href="https://about.twimg	0	-1	-1	?
19	akirap72	@yodafox SurfacePro4で	<a href="https://about.twimg	0	105392423890768256	128816967	yodafox

We will create the table for IPad Pro.

Now click on the SQL Icon.

The SQL console will get open.

Copy paste the below given SQL query on the query console.

```
CREATE COLUMN TABLE "GBI_379"."IPADPRO_379" (Id BIGINT PRIMARY KEY, ScreenName NVARCHAR(256), Tweet NVARCHAR(256), Source NVARCHAR(256), Truncated TINYINT, InReplyToStatusId BIGINT, InReplyToUserId BIGINT, InReplyToScreenName NVARCHAR(256), Favorited TINYINT, Retweeted TINYINT, FavoriteCount INTEGER, Retweet TINYINT, RetweetCount INTEGER, RetweetedByMe TINYINT, CurrentUserRetweetId BIGINT, PossiblySensitive TINYINT, isoLanguageCode NVARCHAR(256), CreatedAt TIMESTAMP, Latitude DOUBLE, Longitude DOUBLE, Country NVARCHAR(256), Place_name NVARCHAR(256), Place_type NVARCHAR(256), UserId BIGINT, UserName NVARCHAR(256), UserUrl NVARCHAR(256));
```

Save the query under gbi-student-379 as given below. Name the query as IPADPRO_379.sql.

Click on the execute button.

The success message after saving the IPADPRO4_379.sql query and clicking on execute button is:

```
12:06:31 AM (Catalog) Created new virtual function 'GBI_379'."GetSearchTweetsIPad" successfully  
12:08:19 AM (SQL File) File 'IPADPRO_379.sql' has been saved  
12:09:45 AM (SQL Editor) Statement 'CREATE COLUMN TABLE "GBI_379"."IPADPRO_379" (Id BIGINT PRIMARY KEY, ScreenName NVARCHAR(256), Tweet ...'  
successfully executed in 10 ms.
```

Now we need to get the search tweet into the IPADPRO_379 table from GetSearchTweetsIPad table.

For that we will create the query based on the Desired and Undesired tweets which we have listed down.

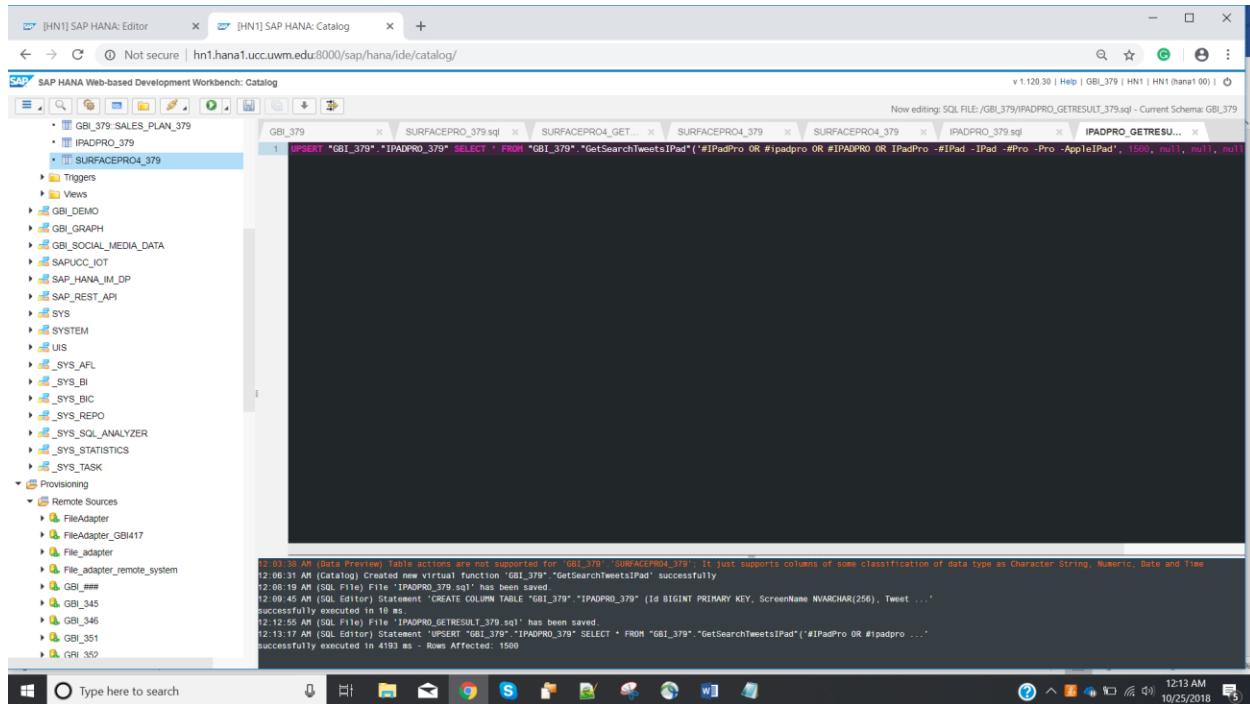
Please find below the query:

```
UPSERT "GBI_379"."IPADPRO_379" SELECT * FROM "GBI_010"."GetSearchTweetsIPad"("#IPadPro OR  
#ipadpro OR #IPADPRO OR IPadPro -#IPad -IPad -#Pro -Pro -AppleIPad", 1500, null, null, null, null,  
null, null);
```

Now we will save this query. Save the query under gbi-student-xxx. Name of the query should be IPADPRO_GETRESULT_379.sql.

Now execute the query by clicking on Run button.

The success message after running the SQL query is:



Right click on the table and select Open Content.

The screenshot after opening the content of IPADPRO_379 table is:

ID	SCREENNAME	TWEET	SOURCE	TRUNCATED	INREPLYTO... ID	INREPLYTO... SCREENNAME	INREPLYOTOS...
1	u2_solutions	The world of reality has its	<a href="http://twitter.com	1	-1	-1	?
2	clygtf99	Halloween is Coming #Cr	<a href="http://twitter.com	1	-1	-1	?
3	steveisaacs	Naomi Watts @CinephileC	<a href="http://twitter.com	1	-1	-1	?
4	DOMzGjn4K9auc	RT @soukun1717. ブレ	<a href="http://twitter.com	0	-1	-1	?
5	isseki	@acsemble Pro3 同じ	<a href="http://twitter.com	0	105491849406422528	763690651786767360	acsemble
6	sapphicjubilee	RT @kevwmada: Mysiqu	<a href="http://twitter.com	0	-1	-1	?
7	takenoko_10	RT @soukun1717. ブレ	<a href="http://twitter.com	0	-1	-1	?
8	akio_fukuda	Lots of leaf. #wpw #tmx	<a href="http://instagram.c	0	-1	-1	?
9	kondii	@99AUO 下取り込むけ	<a href="http://twitter.com	0	1055322367012962304	1939912112	99AUO
10	naosusukino	ほんやり思ってたんだけと	<a href="http://twitter.com	0	-1	-1	?
11	picnic_gold	最近わたしの筆早いの	<a href="http://twitter.com	0	-1	-1	?
12	alexandrapq148	Intobter dia 25 #intobte	<a href="http://instagram.c	1	-1	-1	?
13	mokir1968	VRCは無料だけ #VR&PC	<a href="http://twitter.com	0	-1	-1	?
14	lipopolis	RT @kevwmada: Mysiqu	<a href="http://twitter.com	0	-1	-1	?
15	altpa2ca3	RT @mccain: Plain air p	<a href="http://twitter.com	0	-1	-1	?
16	hamato1082161	RT @eww2823: iPhone8	<a href="http://twitter.com	0	-1	-1	?

4) Rate Limit Status table

Loading data using the Virtual Table

Go to the Editor and open the Provisioning and then go to Application. Right click on Rate_Limit_Status table and then select New Virtual Table.

Click on New Virtual Table, the pop up will come.

Choose the table name as Rate_Limit_Status_379.

The table will get created.

Right click on the table and click Open Content.

In the result table you can see various rate limits. The column “endpoint” contains the type of data for that you call the Twitter API. The column “limit” contains the number of calls you are allowed to do for that type. The column “remaining” contains the number of remaining calls you have. The column “secondsUntilReset” contains the time in seconds until your remaining calls are reset to the rate limit.

In this case you can see that you have just called the function “application/rate_limit_status” once, when opening this table to look up your rate limit.

The screenshot after opening the content of Rate_Limit_Status_379 table is:

The screenshot shows the SAP HANA Web-based Development Workbench interface. The left sidebar displays a tree view of database objects under the 'Catalog' node, including 'Tables' which lists 'Rate_Limit_Status_379'. The main area is a grid table titled 'Rate_Limit_Status_379' with the following data:

	endpoint	limit	remaining	resetTimeInSeconds	secondsUntilReset
1	/application/rate_limit_status	180	179	1540445643	899
2	/account/settings	15	15	1540445643	899
3	/search/tweets	180	180	1540445643	899
4	/statuses/home_timeline	15	15	1540445643	899
5	/statuses/user_timeline	900	900	1540445643	899

At the bottom of the table, there is a message: '12:18:52 AM (Data Preview) Table actions are not supported for 'GBI_379'.'SURFACEPRO4_379'. It just supports columns of some classification of data type as Character, String, Numeric, Date and Time' and '12:18:52 AM (Data Preview) Table actions are not supported for 'GBI_379'.'IPADPRO_379'. It just supports columns of some classification of data type as Character, String, Numeric, Date and Time'.

Now right click on IPADPRO_379 table and choose open content again.

The screenshot after opening the content of IPADPRO_379 table is:

ID	SCREENNAME	TWEET	SOURCE	TRUNCATED	INREPLYTOSTA...	INREPLYTOUS...	INREPLYOTOS
1	10553259937502277632	u2_solutions	The world of reality has its <a href="http://twitter.com/	1	-1	-1	?
2	10553257361259593344	cthy99	Halloween is Coming. #Cr <a href="http://twitter.com/	1	-1	-1	?
3	1055324905784653825	steveisaacs	Naomi Watts @CinephileC <a href="http://twitter.com/	1	-1	-1	?
4	1055324557412724737	DOMzGjn4K9auc	RT @soukun1717: ブレ(<a href="http://twitter.com/	0	-1	-1	?
5	1055324436297994240	issekik	@acsemble Pro3と同じ <a href="http://twitter.com/	0	1054918494405422528	763690661786767360	acsemble
6	1055323958487195647	sapphicjubilee	RT @kevnmwda: Mystigo <a href="http://twitter.com/	0	-1	-1	?
7	1055323846797634560	takenoko_10_	RT @soukun1717: ブレ(<a href="http://twitter.com/	0	-1	-1	?
8	105532353281158913	ako_fukuda	Lots of leafs. #wp #insta <a href="http://instagram.c	0	-1	-1	?
9	1055323071567978496	kondili	@99AUO 下船歌入りむけ <a href="http://twitter.com/	0	1055322367012962304	1939912112	99AUO
10	1055322971433132037	naosusukino	(ぼんやり思ってたんだけど <a href="http://twitter.com/	0	-1	-1	?
11	1055322028297773058	picnic_gold	最近わたしの筆が早いのは <a href="http://twitter.com/	0	-1	-1	?
12	1055319465684633344	alexandrapq148	Inktober dia 25 #inktober <a href="http://instagram.c	1	-1	-1	?
13	105531541761595904	moki1968	VRCは無料だけLV&PC <a href="http://twitter.com/	0	-1	-1	?
14	105531885637191424	lippolito	RT @kevnmwda: Mystigo <a href="http://twitter.com/	0	-1	-1	?
15	1055316863087853568	altpazca3	RT @mmccain: Plein air p <a href="http://twitter.com/	0	-1	-1	?
16	1055316540722036737	hamama01082161	RT @ewel2823: iPhone8: <a href="http://twitter.com/	0	-1	-1	?

11:10:52 AM (Data Preview) Table actions are not supported for 'GBI_379'.'SURFACEPRO4_379'. It just supports columns of some classification of data type as Character String, Numeric, Date and Time
11:10:52 AM (Data Preview) Table actions are not supported for 'GBI_379'.'IPADPRO_379'. It just supports columns of some classification of data type as Character String, Numeric, Date and Time
12:18:54 AM (Data Preview) Result Limited to 1000 row(s) due to value configured in the Catalog settings
12:20:57 AM (Data Preview) Table actions are not supported for 'GBI_379'.'IPADPRO_379'. It just supports columns of some classification of data type as Character String, Numeric, Date and Time
12:20:57 AM (Data Preview) Result Limited to 1000 row(s) due to value configured in the Catalog settings

Please open again the content of Rate_Limit_Status table. You will notice that the remaining call value is reduced to 178.

The screenshot after opening the contents of Rate_Limit_Status_379 table again is as follows. The number of remaining calls now is **178**.

endpoint	limit	remaining	resetTimeInSeconds	secondsUntilReset
/application/rate_limit_status	180	178	1540445643	706
/account/settings	15	15	1540445836	899
/search/tweets	180	180	1540445836	899
/statuses/home_timeline	15	15	1540445836	899
/statuses/user_timeline	900	900	1540445836	899

Similarly, we will create a table for **Samsung Galaxy S 8**.

The create command will be:

```
CREATE COLUMN TABLE "GBI_379"."GALAXYS8_379" (Id BIGINT PRIMARY KEY, ScreenName NVARCHAR(256), Tweet NVARCHAR(256), Source NVARCHAR(256), Truncated TINYINT, InReplyToStatusId BIGINT, InReplyToUserId BIGINT, InReplyToScreenName NVARCHAR(256), Favorited TINYINT, Retweeted TINYINT, FavoriteCount INTEGER, Retweet TINYINT, RetweetCount INTEGER, RetweetedByMe TINYINT, CurrentUserRetweetId BIGINT, PossiblySensitive TINYINT, isoLanguageCode NVARCHAR(256), CreatedAt TIMESTAMP, Latitude DOUBLE, Longitude DOUBLE, Country NVARCHAR(256), Place_name NVARCHAR(256), Place_type NVARCHAR(256), UserId BIGINT, UserName NVARCHAR(256), UserUrl NVARCHAR(256));
```

We will upload the data into this table with the help of **GetSearchTweets** virtual function which we have used to upload the Surface Pro 4 table.

Please find below the command to upload the data into Samsung Galaxy S 8 table:

```
UPSERT "GBI_379"."SURFACEPRO4_379" SELECT * FROM "GBI_010"."GetSearchTweets"("#GalaxyS8 OR #galaxys84 OR #GALAXYS8 OR GalaxyS8 -#Galaxy -#S8 -Galaxy -S8 -GalaxyS -#GalaxyS", 1500, null, null, null, null, null, null, null);
```

The screenshot after running queries of GALAXYS8_379.sql and GALAXY_GETRESULT_379.sql table, and after opening the contents of Rate_Limit_Status_379 table is as follows.

endpoint	limit	remaining	resetTimeInSeconds	secondsUntilReset
/application/rate_limit_status	180	179	1540446964	899
/account/settings	15	15	1540446964	899
/search/tweets	180	174	1540446936	871
/statuses/home_timeline	15	15	1540446964	899
/statuses/user_timeline	900	900	1540446964	899

12:32:49 AM (Data Preview) Table actions are not supported for 'GBI_379"."SURFACEPRO4_379'. It just supports columns of some classification of data type as Character String, Numeric, Date and Time
12:32:49 AM (Data Preview) Result Limited to 1000 rows(s) due to value configured in the Catalog settings
12:32:50 AM (Data Preview) Table actions are not supported for 'GBI_379"."GALAXYS8_379'. It just supports columns of some classification of data type as Character String, Numeric, Date and Time
12:40:21 AM (SQL File) File 'GALAXYS8_GETRESULT_379.sql' has been saved.
12:40:35 AM (SQL Editor) Statement 'UPSERT 'GBI_379"."GALAXYS8_379" SELECT * FROM "GBI_379"."GetSearchTweets"("#GalaxyS8 OR #galaxyS8 ..." successfully executed in 1352 ms - Rows Affected: 463

The number of calls that are left for endpoint “/search/tweets” are **174**.

5) Fulltext Index

Objective

Get familiar with the creation of fulltext indexes and preparation of Twitter data for further analysis.

We will create a Fulltext Index on the tables that have been filled with Twitter Data in the previous exercise. Then we will have a look at the analysis results that have been created with the fulltext index.

To create a fulltext index you need to execute a SQL statement. This creates a table containing the Sentiment Analysis results.

Open the Catalog of your system in a browser using the Web IDE.

<http://hn1.hana1.ucc.uwm.edu:8000/sap/hana/ide/catalog/>

This statement creates a fulltext index for column “Tweet” for the table GBI_379.IPADPRO_379. Because of the keywords “TEXT ANALYSIS ON” the system recognizes that you want to do text analysis.

After the keyword “CONFIGURATION” you can add different configuration options. As we want to analyze customers’ opinion about the iPad Pro and the Surface Pro 4, we add the option “EXTRACTION_CORE_VOICEOFCUSTOMER”.

The keyword “LANGUAGE COLUMN” identifies the column which holds a language indicator if available. If you use this addition the SAP HANA database can analyze data in various languages (e.g. English, German, Portuguese, Japanese and much more).

The keyword “LANGUAGE DETECTION” tells the preprocessor server to use the given languages for the analysis.

When executing the statement, the preprocessor server of the SAP HANA system searches for different keywords that identify to which sentiment a Tweet belongs.

Open the following path:

Catalog ► GBI_379 ► Tables

You can see that a new table with the name “\$TA_iPadPro_Sentiment_379” was created. The beginning “TA” stands for Text Analysis. The table itself contains the analysis results. Right click on the table and choose **Open Content**.

The table contains a lot of different columns with different information that was found during analysis:

TA_COUNTER: This counter counts all tokens in a document. In our case a “document” is one row in the database because each row represents one Tweet.

TA_TOKEN: Represents the token that was identified. That can be only one word or more and depends on the type that was identified.

TA_LANGUAGE: Is the language code for the document, i.e. the language of the Tweet that was taken out of the original database table.

TA_TYPE: Contains the linguistic or semantic type of the token. The types of interest for us are for example “WeakPositiveSentiment”, “PositiveSentiment”, “NegativeSentiment” or “MinorProblem” because these types tell us that the identified token is a sentiment, i.e. the opinions of the user who send the Tweet.

TA_CREATED_AT: Is the creation time of the index and only used for administrative tasks, e.g. database reorganization.

TA_OFFSET: Stores the offset of characters relative to the beginning of the document.

TA_PARENT: Stores the TA_COUNTER of the parent token. If it has no parent token the value is null. If, for example, a token that was identified includes two words, there can be two additional rows with the single words.

The screenshot after opening the contents of \$TA_iPadPro_Sentiment_379 table is:

The screenshot shows the SAP HANA Web-based Development Workbench interface. The left sidebar displays the Catalog structure, including the \$TA_IPadPro_Sentiment_379 table. The main area shows a table editor for the \$TA_IPadPro_Sentiment_379 table, which has 1000 rows. The columns are: ID, TA_RULE, TA_COUNTER, TA_TOKEN, TA_LANGUAGE, TA_TYPE, TA_NORMALIZED, and TA_STEM. The data consists of various entities extracted from text, such as 'Entity Extraction', 'RT', 'en', 'ORGANIZATION/MEDIA', '?', and '?'. A status bar at the bottom provides log information about data preview limitations.

If you have a deeper look into the data, you will recognize that the term “Pro” is sometimes recognized as a weak positive sentiment. Actually this is not a sentiment, but the second word of the term “iPad Pro”.

That means we have to find a way how to tell the SAP HANA database, that “iPad Pro” is the name of a product.

Now, repeat the aforementioned steps to create a Fulltext index for **Surface Pro 4**.

The screenshot after opening the contents of \$TA_SurfacePro4_Sentiment_379 table is:

ID	TA_RULE	TA_COUNTER	TA_TOKEN	TA_LANGUAGE	TA_TYPE	TA_NORMALIZED	TA_STEM
1	1055244399934877696	Entity Extraction	1	inktober day	en	PROP_MISC	?
2	1055244399934877696	Entity Extraction	2	#inktober	en	SOCIAL_MEDIA/TOPIC_1	?
3	1055244399934877696	Entity Extraction	3	inktober2018	en	SOCIAL_MEDIA/TOPIC_1	?
4	1055244399934877696	Entity Extraction	4	illustration	en	SOCIAL_MEDIA/TOPIC_1	?
5	1055244399934877696	Entity Extraction	5	digitalpainting	en	SOCIAL_MEDIA/TOPIC_1	?
6	1055244399934877696	Entity Extraction	6	cincinnati	en	SOCIAL_MEDIA/TOPIC_1	?
7	1055244399934877696	Entity Extraction	7	ohio	en	SOCIAL_MEDIA/TOPIC_1	?
8	1055244399934877696	Entity Extraction	8	gargardesign	en	SOCIAL_MEDIA/TOPIC_1	?
9	1055244399934877696	Entity Extraction	9	https://it.co/b5W8uktmw	en	URI/URL	?
10	1055137154047922176	Entity Extraction	1	admire	en	Sentiment	?
11	1055137154047922176	Entity Extraction	2	admire	en	WeakPositiveSentiment	?
12	1055137154047922176	Entity Extraction	3	short time	en	NON_GROUP	?
13	1055137154047922176	Entity Extraction	4	https://it.co/mVRIT5cjq	en	URI/URL	?
14	1055125447183347712	Entity Extraction	1	microsoft	en	ORGANIZATION/COMME	?
15	1055125447183347712	Entity Extraction	2	x microsoftstore	en	NON_GROUP	?
16	1055125447183347712	Entity Extraction	3	https://it.co/dOffKLWQ3w	en	URI/URL	?
17	1054954399443943424	Entity Extraction	1	#surfacepro4	en	SOCIAL_MEDIA/TOPIC_1	?

11:41:03 AM [Data Preview] Table actions are not supported for 'GBI_379'-'\$TA_SURFACEPRO4_379'. It just supports columns of some classification of data type as Character, String, Numeric, Date and Time
11:41:04 AM [Data Preview] Table actions are not supported for 'GBI_379'-'\$TA_IPADPRO_379'. It just supports columns of some classification of data type as Character, String, Numeric, Date and Time
11:41:05 AM [Data Preview] Result limited to 1000 row(s) due to value configured in the Catalog settings
11:41:06 AM [Data Preview] Table actions are not supported for 'GBI_379'-'\$TA_IPADPro_Sentiment_379'. It just supports columns of some classification of data type as Character, String, Numeric, Date and Time
11:41:07 AM [Data Preview] Result limited to 1000 row(s) due to value configured in the Catalog settings
11:41:08 AM [Data Preview] Table actions are not supported for 'GBI_379'-'\$TA_SurfacePro4_Sentiment_379'. It just supports columns of some classification of data type as Character, String, Numeric, Date and Time
11:41:17 AM [Data Preview] Table actions are not supported for 'GBI_379'-'\$TA_SurfacePro4_Sentiment_379'. It just supports columns of some classification of data type as Character, String, Numeric, Date and Time

If you carefully study the data, you will recognize that the term “Surface Pro” or “Surface Pro 4” is recognized as a topic but should be recognized as a product.

That means we have to find a way how to tell the SAP HANA database, that “Surface Pro 4” is the name of a product.

Custom Dictionary

The iPad Pro and the Surface Pro 4 are not reliably identified as products. We have to find a solution for that problem. We can resolve that issue by adding a custom dictionary for unknown terms in the SAP HANA system.

Open the Editor of your system in a browser using the Web IDE.

<http://hn1.hana1.ucc.uwm.edu:8000/sap/hana/ide/editor/>

To create a custom dictionary, you need a folder to store all necessary configuration files.

Right click on your package “gbi-student-379” and create a new package with the name “CustomDictionaries”.

Now right click on the new package and choose

New ► File

As a file extension for custom dictionaries, you generally enter “.hdbtextdict”. Furthermore, a convention is to add the language of the dictionary in the beginning of the file name, separated by a hyphen. If you don't add any language, the file will be used for all languages. First we want to create a custom dictionary for the iPad Pad. Name the file “iPadPro_379.hdbtextdict”.

The editor for the file will open. Write the following code into the editor:

```

CUSTOMER_DIM_CV... x PRODUCT_DIM_CV... x SALES_CUBE_CV_01... x iPadPro_010.hdbtext...
1 <?xml version="1.0" encoding="UTF-8"?>
2 <dictionary xmlns="http://www.sap.com/ta/4.0">
3 <entity_category name="AppleTabs">
4 <entity_name standard_form="IPad Pro">
5 <variant name = "iPad" />
6 <variant name = "iPadPro" />
7 <variant name = "#iPadPro" />
8 <variant name = "Apple iPad Pro" />
9 </entity_name>
10 </entity_category>
11 </dictionary>

```

Save the file. It will be saved and activated on the server.

The screenshot of the dictionary alongwith the status message is:

[HANA] SAP HANA: Catalog [HANA] SAP HANA: Editor

v 1.120.30 | Help | GBL_379 | HN1 | HN1 (hana1 00) |

Content [package] GBL_379 iPadPro_379.hdbtextdict

Now editing: /GBL_379/CustomDictionaries/iPadPro_379.hdbtextdict

```

[package] GBL_379 x iPadPro_379.hdbtextdict x
1 <?xml version="1.0" encoding="UTF-8"?>
2 <dictionary xmlns="http://www.sap.com/ta/4.0">
3 <entity_category name="AppleTabs">
4 <entity_name standard_form="IPad Pro">
5 <variant name = "iPad" />
6 <variant name = "iPadPro" />
7 <variant name = "#iPadPro" />
8 <variant name = "Apple iPad Pro" />
9 </entity_name>
10 </entity_category>
11 </dictionary>

```

11:43:49 Package /GBL_379/CustomDictionaries created successfully.
11:44:40 File /GBL_379/CustomDictionaries/iPadPro_379.hdbtextdict created successfully.
11:49:05 File /DB_379/CustomDictionaries/iPadPro_379.hdbtextdict saved & activated successfully.

The next step is to create a configuration file for the analysis. The easiest way to do this is to copy a pre-delivered configuration. All configuration files that are pre-delivered when installing a SAP HANA system are stored under:

Content ► sap ► hana ► ta ► config

As we want to implement Sentiment Analysis, we choose the configuration file "EXTRACTION_CORE_VOICEOFCUSTOMER.hdbtextconfig". Right click on this file and choose Copy.

Right click on your package "CustomDictionaries" and click on Paste.

Now you have a copy of the configuration file to which you can add your custom dictionary, but first we want to give the new file a sensible name.

Right click on the file "EXTRACTION_CORE_VOICEOFCUSTOMER.hdbtextconfig" and choose Rename.

As we want to use the configuration for both analysis, name it "CustomConfig_379.hdbtextconfig".

Open this file and scroll down to the line with the dictionaries (that should be around line 135).

```

*CustomConfig_000.h... ✘

137      <!-- List of Text Analysis extraction dictionaries for Sentiment Analysis. -->
138      <property name="Dictionaries" type="string-list">
139          <string-list-value>english-tf-voc-AmbigProfanity.nc</string-list-value>
140          <string-list-value>english-tf-voc-UnambigProfanity.nc</string-list-value>
141          <string-list-value>french-tf-voc-AmbigProfanity.nc</string-list-value>
142          <string-list-value>french-tf-voc-UnambigProfanity.nc</string-list-value>
143          <string-list-value>german-tf-voc-AmbigProfanity.nc</string-list-value>
144          <string-list-value>german-tf-voc-UnambigProfanity.nc</string-list-value>
145          <string-list-value>italian-tf-voc-AmbigProfanity.nc</string-list-value>
146          <string-list-value>italian-tf-voc-UnambigProfanity.nc</string-list-value>
147          <string-list-value>portuguese-tf-voc-AmbigProfanity.nc</string-list-value>
148          <string-list-value>portuguese-tf-voc-UnambigProfanity.nc</string-list-value>
149          <string-list-value>russian-tf-voc-AmbigProfanity.nc</string-list-value>
150          <string-list-value>russian-tf-voc-UnambigProfanity.nc</string-list-value>
151          <string-list-value>simplified-chinese-tf-voc-AmbigProfanity.nc</string-list-value>
152          <string-list-value>simplified-chinese-tf-voc-UnambigProfanity.nc</string-list-value>
153          <string-list-value>spanish-tf-voc-AmbigProfanity.nc</string-list-value>
154          <string-list-value>spanish-tf-voc-UnambigProfanity.nc</string-list-value>
155          <string-list-value>traditional-chinese-tf-voc-AmbigProfanity.nc</string-list-value>
156          <string-list-value>traditional-chinese-tf-voc-UnambigProfanity.nc</string-list-value>
157          <string-list-value>sap.hana.ta.voc::english-tf-voc-thesaurus.hdbtextdict</string-list-value>
158          <string-list-value>sap.hana.ta.voc::french-tf-voc-thesaurus.hdbtextdict</string-list-value>
159          <string-list-value>sap.hana.ta.voc::german-tf-voc-thesaurus.hdbtextdict</string-list-value>
160          <string-list-value>sap.hana.ta.voc::italian-tf-voc-thesaurus.hdbtextdict</string-list-value>
161          <string-list-value>sap.hana.ta.voc::portuguese-tf-voc-thesaurus.hdbtextdict</string-list-value>
162          <string-list-value>sap.hana.ta.voc::russian-tf-voc-thesaurus.hdbtextdict</string-list-value>
163          <string-list-value>sap.hana.ta.voc::simplified-chinese-tf-voc-thesaurus.hdbtextdict</string-list-value>
164          <string-list-value>sap.hana.ta.voc::spanish-tf-voc-thesaurus.hdbtextdict</string-list-value>
165          <string-list-value>sap.hana.ta.voc::traditional-chinese-tf-voc-thesaurus.hdbtextdict</string-list-value>
166          <string-list-value>gb1-student-000.CustomDictionaries::iPadPro_000.hdbtextdict</string-list-value>
167      </property>
168  </configuration>
169</tasdk-configuration>
170
171

```

This adds your custom dictionary to the configuration. Save the file.

The status message after saving the CustomConfig_379.hdbtextconfig is:

Now, repeat the aforementioned steps to create a custom dictionary for Surface Pro 4.

Save the file. It will be checked for any syntax errors and afterwards it will be activated on the SAP HANA system.

[11:56:37] File /GBI_379/CustomDictionaries/CustomConfig_379.hdbtextconfig saved & activated successfully.

Now you can use the configuration file to do an improved analysis on your twitter data.

To create the tables “\$TA_iPadPro_Sentiment_379” and “\$TA_SurfacePro4_Sentiment_379” you have to drop the fulltext index on the original tables first. To do this, go to the Catalog in the WebIDE of your SAP HANA system:

<http://hn1.hana1.ucc.uwm.edu:8000/sap/hana/ide/catalog/>

The screenshot of the configuration file for SurfacePro4 alongwith success message is:

```

<!-- List of Text Analysis extraction dictionaries for Sentiment Analysis. -->
<property name="Dictionaries" type="string-list">
<string-list-value>dutch-tf-voc-AmbigProfanity_nc</string-list-value>
<string-list-value>dutch-tf-voc-UnameigProfanity_nc</string-list-value>
<string-list-value>english-tf-voc-AmbigProfanity_nc</string-list-value>
<string-list-value>english-tf-voc-UnameigProfanity_nc</string-list-value>
<string-list-value>french-tf-voc-UnameigProfanity_nc</string-list-value>
<string-list-value>french-tf-voc-UnameigProfanity_nc</string-list-value>
<string-list-value>german-tf-voc-UnameigProfanity_nc</string-list-value>
<string-list-value>italian-tf-voc-UnameigProfanity_nc</string-list-value>
<string-list-value>italian-tf-voc-UnameigProfanity_nc</string-list-value>
<string-list-value>portuguese-tf-voc-UnameigProfanity_nc</string-list-value>
<string-list-value>portuguese-tf-voc-UnameigProfanity_nc</string-list-value>
<string-list-value>russian-tf-voc-UnameigProfanity_nc</string-list-value>
<string-list-value>spanish-tf-voc-UnameigProfanity_nc</string-list-value>
<string-list-value>spanish-tf-voc-UnameigProfanity_nc</string-list-value>
<string-list-value>chinese-tf-voc-UnameigProfanity_nc</string-list-value>
<string-list-value>chinese-tf-voc-UnameigProfanity_nc</string-list-value>
<string-list-value>english-tf-voc-theaurus_nc</string-list-value>
<string-list-value>french-tf-voc-theaurus_nc</string-list-value>
<string-list-value>italian-tf-voc-theaurus_nc</string-list-value>
<string-list-value>portuguese-tf-voc-theaurus_nc</string-list-value>
<string-list-value>russian-tf-voc-theaurus_nc</string-list-value>
<string-list-value>spanish-tf-voc-theaurus_nc</string-list-value>
<string-list-value>GBI_379_CustomDictionaries:iPadPro_379.hdbtextdict</string-list-value>
<string-list-value>GBI_379_CustomDictionaries:SurfacePro4_379.hdbtextdict</string-list-value>

```

[43:44:0] File GBI_379_CustomDictionaries:CustomConfig_379.hdbtextconfig saved successfully.
[43:44:2] Error while activating GBI_379_CustomDictionaries:CustomConfig_379.hdbtextconfig:
GBI_379_CustomDictionaries:CustomConfig_379.hdbtextconfig: Text Analysis configuration syntax error: <expected '>>
[43:34:48] File GBI_379_CustomDictionaries:CustomConfig_379.hdbtextconfig saved & activated successfully.

Note: In case if you are not getting the into the sentiment table after running the full text query, please check your package structure.

Please find below the steps to check the package structure.

First, please check where you CustomConfig.hdbtextdict, IPadPro_379.hdbtextdict and SurfacePro4_379.hdbtextdict file are present.

Suppose, the files are present in the below given location:

In this case, your package structure in the CustomConfig file for the IPadPro_379.hdbtextdict file will be:

Similarly, the full text index query should also have the correct package structure for CustomConfig_379.hdbtextdict file.

Data Cleaning

Delete terms from the result tables that are recognized as a sentiment, but are, in fact, a part of a product. These terms can, for example, contains the following expressions: “Pro:”, “Pro –” or “Pro –”.

Now the sentiments have been analyzed by the search engine. If you have a deeper look into the tables “\$TA_iPadPro_Sentiment_379” and “\$TA_iPadPro_Sentiment_379”, you will recognize that the term “Pro” followed by signs like “-” or “.” is still categorized as “asSentiment”. That is why we need to do some data cleansing to filter out wrong sentiments.

Open the Catalog of your system in a browser using the Web IDE.

<http://hn1.hana1.ucc.uwm.edu:8000/sap/hana/ide/catalog/>

Go to Catalog ► GBI_379 ► Tables

and right click on the table “\$TA_iPadPro_Sentiment_379”. Choose **Open Content**.

Click on the column “TA_TYPE” and enter “Weak” into the filter field. Press Enter.

Open a SQL console and copy and paste the following statement.

If you now refresh the content view in the WebIDE, you will receive the information that there are no rows in the table containing TA_TOKEN as 'fun' and TA_TYPE as 'WeakPositiveSentiment'.

The screenshot after deleting the rows containing 'fun' and TA_TYPE as 'WeakPositiveSentiment' is:

Now editing: GBI_379/\$TA_IpadPro_Sentiment_379
untitled2.sql
1000 row(s) |

ID	TA_RULE	TA_COUNTER	TA_TOKEN	TA_LANGUAGE	TA_TYPE	TA_NORMALIZED	TA_STEM

No data

2:52:51 PM (Data Preview) Table actions are not supported for "GBI_379"."\$TA_SurfacePro4_Sentiment_379". It just supports columns of view classification of data type as Character String, Numeric, Date and Time.
12:56:14 PM (SQL Editor) Could not execute 'delete from "GBI_379"."\$TA_IpadPro_Sentiment_379" where "TA_TOKEN" = 'fun' AND "TA_TYPE" = ...'
Error (dberror) 260 Invalid column name 'TA_TOKEN' line 1 col 95 (at pos 58)
12:56:26 PM (SQL Editor) Statement 'delete from "GBI_379"."\$TA_IpadPro_Sentiment_379" where "TA_TOKEN" = 'fun' AND "TA_TYPE" = ...'

Go to

Catalog ► GBI_379 ► Tables

and right click on the table “\$TA_SurfacePro4_Sentiment_379”. Choose Open Content.

Click on the column “TA_TYPE” and enter “Weak” into the filter field. Press Enter.

The result will be:

	ID	TA_RULE	TA_COUNTER	TA_TOKEN	TA_LANGUAGE	TA_TYPE
455	846038035686936576	Entity Extraction	6	terrible	en	StrongNegativeSentiment
886	846861225317220352	Entity Extraction	2	wtf	en	StrongNegativeSentiment
905	846610439475650561	Entity Extraction	6	better	en	WeakNegativeSentiment

If you now refresh the content view in the WebIDE, you will receive the information that there are no rows in the table containing TA_TYPE as ‘WeakNegativeSentiment’ and ‘StrongNegativeSentiment’.

The screenshot after rows containing TA_TYPE as ‘WeakNegativeSentiment’ and ‘StrongNegativeSentiment’ is:

SAP HANA Web-based Development Workbench: Catalog

v 1.120.30 | Help | GBI_379 | HNT | HN1 (hana1 00) | [Logout](#)

Now editing: GBI_379/STA_SurfaceProd4_Sentiment_379

0 row(s) | [Filter](#) | [Search](#) | [Reset](#) | [Save](#)

PRO_379 X Rate_Limit_Status_379 X GALAXYS8_379 X STA_IPadPro_Sentime... X untitled2.sql X STA_SurfaceProd4_Se... X untitled3.sql X

ID TA_RULE TA_COUNTER TA_TOKEN TA_LANGUAGE TA_TYPE TA_NORMALIZED TA_STEM

Type to filter

No data

Tables

- STA_IPadPro_Sentiment_379
- STA_SurfaceProd4_Sentiment_379
- GALAXYS8_379
- GBI_379_COUNTRY_CODE_DESC_379
- GBI_379_CUSTOMER_ATTR_379
- GBI_379_CUSTOMER_SEGMENT_ATTR_379
- GBI_379_PRODUCT_ATTR_379
- GBI_379_PRODUCT_CATEGORY_ATTR_379
- GBI_379_PRODUCT_LINE_ATTR_379
- GBI_379_PRODUCT_SUBCATEGORY_ATTR_379
- GBI_379_SALES_379
- GBI_379_SALES_PLAN_379
- IPADPRO_379
- Rate_Limit_Status_379
- SURFACEPRO4_379

Triggers

Views

GBI_DEMO

GR1_GRAPH

successfully executed in 0 ms.

10:03:26 PM (Data Preview) Table actions are not supported for 'GBI_379' 'STA_SurfaceProd4_Sentiment_379'. It just supports columns of some classification of data type as Character String, Numeric, Date and Time.

10:03:24 PM (SQL Editor) Statement 'delete from "GBI_379"."STA_SurfaceProd4_Sentiment_379" where ("TA_TOKEN" = 'terrible' OR ...)

Sentiment Analysis and Visualization using SAP Lumira:

After having the Twitter data loaded and prepared for the Sentiment Analysis, production planner wants to present his findings to vice president operations. Production Planner knows that “a picture is worth a thousand words”, i.e. the best way to explain the findings is to present them graphically. Always striving for the best performance and newest technologies, he decides to use SAP Lumira Server for the visualization of the acquired data. As it provides multifarious possibilities to analyze and represent the data in an easy way.

1) Create an Analytic View

We need to go to the Editor and right click on our package and create New -> Package and name the package “TwitterLoad”.

Right click on TwitterLoad package.

Choose New and then select Calculation View.

Enter the details in the pop up as given above. The name of the analytical view should be **AV_IPADPRO_379**. Click on Create.

The analytical view will get created.

Drag the projection on to the canvas.

Click on the + sign on the projection.

Search for the table - **\$TA_IPadPro_Sentiment_379** which we have created earlier.

Select the table and click on OK.

Add the following columns to the output:

- Id
- TA_TOKEN
- TA_LANGUAGE
- TA_TYPE

Join the Projection 1 to the Star Join

Click on the Star Join and then go to mapping. We need to perform the mapping.

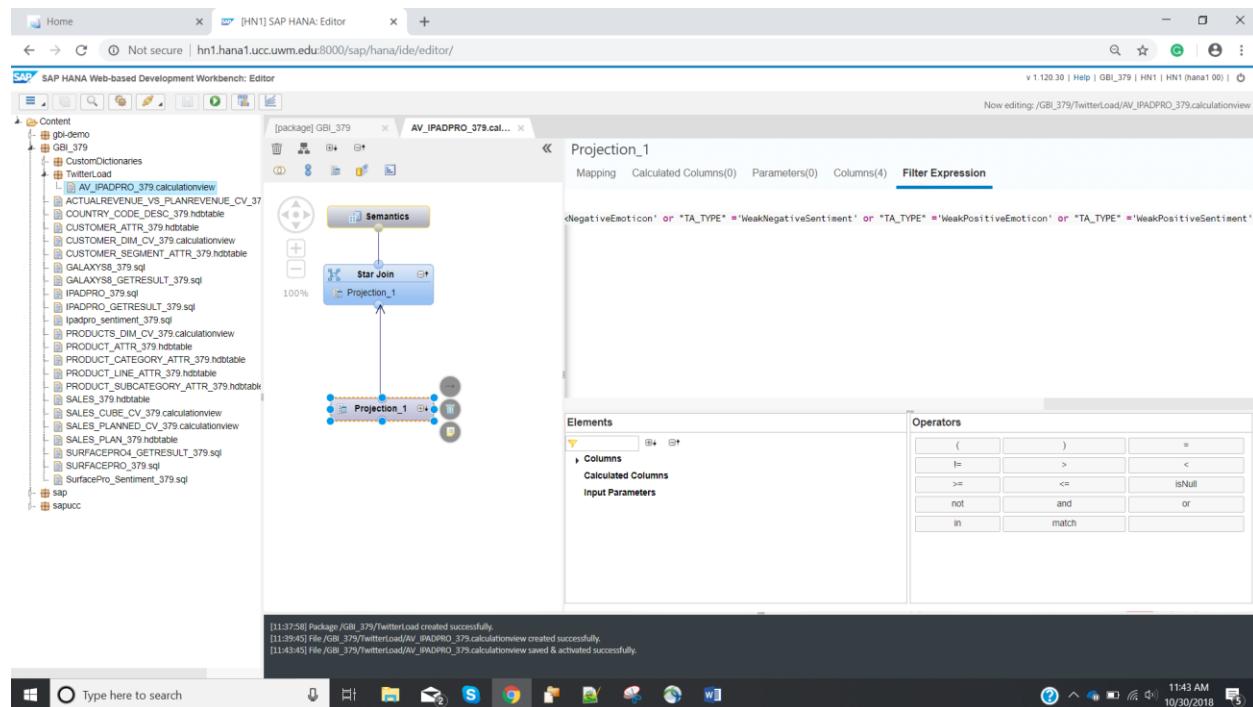
Now go to the Semantics and then change the Aggregation from Sum to Count.

Click on the Projection_1 and then go to filter expression.

Type the below given filter expression:

```
"TA_TYPE" ='MajorProblem' or "TA_TYPE" ='MinorProblem' or "TA_TYPE" ='NeutralEmoticon' or  
"TA_TYPE" ='NeutralSentiment' or "TA_TYPE" ='StrongNegativeEmoticon' or "TA_TYPE"  
='StrongNegativeSentiment' or "TA_TYPE" ='StrongPositiveEmoticon' or "TA_TYPE"  
='StrongPositiveSentiment' or "TA_TYPE" ='WeakNegativeEmoticon' or "TA_TYPE"  
='WeakNegativeSentiment' or "TA_TYPE" ='WeakPositiveEmoticon' or "TA_TYPE"  
='WeakPositiveSentiment'
```

The Screenshot after saving the Analytical view for iPadPro



Now clicking on the execute button

Close the calculation view right away.

Screenshot after executing:

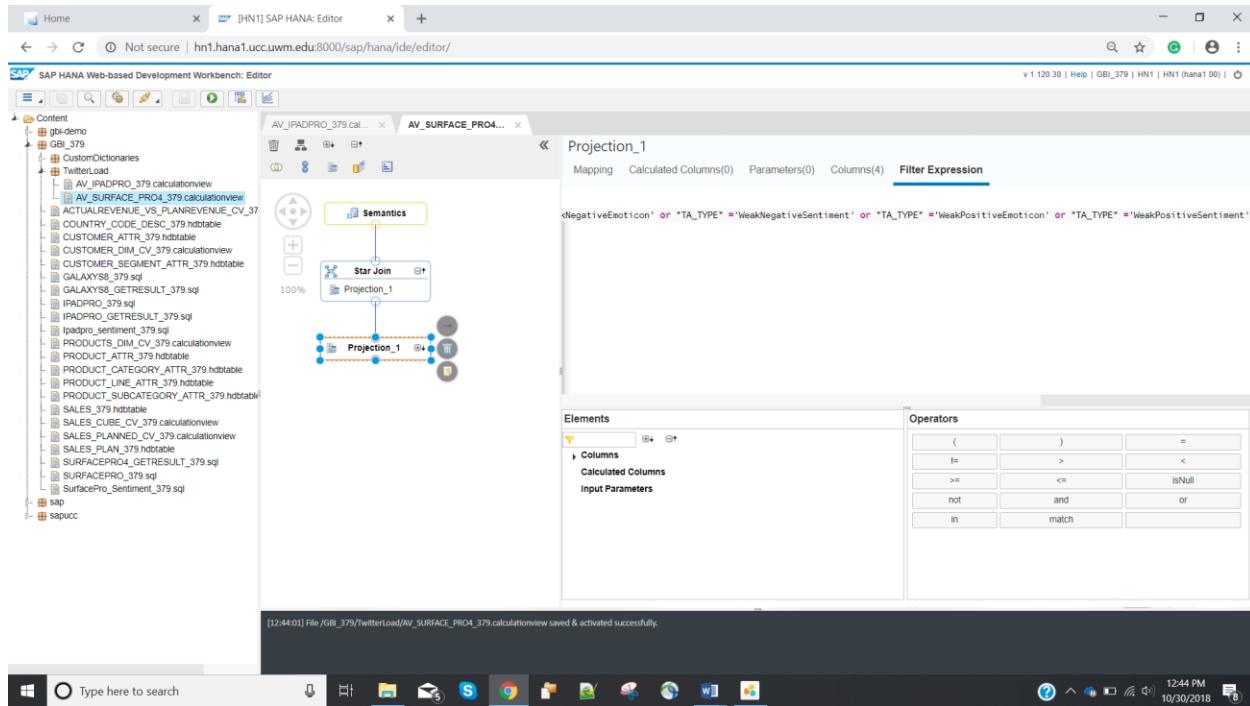
The screenshot shows the SAP HANA Web-based Development Workbench. On the left, there's a navigation tree with packages like 'gb1-demo' and 'GBI_379'. The main area displays a table titled '[package] GBI_379' with the sub-name 'GBI_379.TwitterLoad...'. The table has four columns: 'ID', 'TA_LANGUAGE', 'TA_TOKEN', and 'TA_TYPE'. There are 110 rows of data. The data includes various words and their corresponding sentiment types. A message bar at the bottom of the screen shows log entries related to package creation and view saving.

ID	TA_LANGUAGE	TA_TOKEN	TA_TYPE
1	en	inspired	WeakPositiveSentiment
2	en	boundless	WeakPositiveSentiment
3	en	successful	StrongPositiveSentiment
4	en	easy	WeakPositiveSentiment
5	en	like	WeakPositiveSentiment
6	en	:)	WeakPositiveEmoticon
7	en	Enjoy	WeakPositiveSentiment
8	en	favs	StrongPositiveSentiment
9	en	understandable	NeutralSentiment
10	en	excited	StrongPositiveSentiment
11	en	dedicated	WeakPositiveSentiment
12	en	proper	WeakPositiveSentiment
13	pt	Amo	StrongPositiveSentiment
14	es	orgullo	WeakPositiveSentiment
15	en	lovely	StrongPositiveSentiment
16	en	😊	WeakPositiveEmoticon
17	en	couldn't find	MinorProblem
18	en	😢	WeakPositiveEmoticon
19	en	Can't wait!	StrongPositiveSentiment
20	en	Awesome	StrongPositiveSentiment

[11:37:58] Package /GBI_379/twitterLoad created successfully.
[11:39:45] File /GBI_379/twitterLoad/AV_IPADPRO_379.calculationview created successfully.
[11:40:01] File /GBI_379/twitterLoad/AV_IPADPRO_379.calculationview saved & activated successfully.
[11:46:04] File /GBI_379/twitterLoad/AV_SURFACEPRO4_379.calculationview saved & activated successfully.
[11:48:17] File /GBI_379/twitterLoad/AV_SURFACEPRO4_379.calculationview saved & activated successfully.

Repeat the steps from this scenario for the table **\$TA_SurfacePro4_Sentiment_379**. Name the analytic view **AV_SURFACE_PRO4_379**.

Screenshot after saving the analytical view of Surface_Pro4 is:



Screenshot after executing:

ID	TA_LANGUAGE	TA_TOKEN	TA_TYPE
1	en	good	StrongPositiveSentiment
2	en	crappy	StrongNegativeSentiment
3	en	stupid	StrongNegativeSentiment
4	en	Sucks	StrongNegativeSentiment
5	en	😊	WeakPositiveEmoticon
6	en	brand new	WeakPositiveSentiment
7	en	misplaced	MinorProblem
8	en	Can't wait	StrongPositiveSentiment
9	en	happy	StrongPositiveSentiment
10	en	best	StrongPositiveSentiment
11	es	gran	StrongPositiveSentiment
12	en	sucks	StrongNegativeSentiment
13	en	Nice	WeakPositiveSentiment
14	en	Workhorse	StrongPositiveSentiment
15	de	funktioniert	MinorProblem
16	en	Good	WeakPositiveSentiment
17	en	special	WeakPositiveSentiment
18	en	happy	WeakPositiveSentiment
19	en	Slow	MinorProblem
20	en	interesting	WeakPositiveSentiment

[12:44:01] File /GBI_379/TwitterLoad/AV_SURFACE_PRO4_379.calculationview saved & activated successfully.
[12:44:22] [Data Preview] Result limited to 1000 row(s) due to value configured in the Catalog settings
[12:44:46] File /GBI_379/TwitterLoad/AV_SURFACE_PRO4_379.calculationview saved & activated successfully.

We need to make work around here. This workaround is for the current version of SAP HANA. In the future, this step may not be required.

Right click on the cube created and then open with Text Editor.

We need to go to second line and delete the expression **dimensionType="STANDARD"**

Save the view.

Similarly repeat the steps for Analytical view created for Surface Pro4.

2) Create a Tag Cloud with SAP Lumira Server

A tag cloud is a nice tool to find out which terms are used very frequently in a given data set. Jun Lee wants to get an overview about the sentiments. He thinks it is a good starting point to use a tag cloud for visualization.

Opening the SAP Lumira Desktop Application. Go to file and click on New.

Now select Connect to SAP HANA and click on Next.

Now make the selection

Connect to: HANA

Server: hana1.ucc.uwm.edu

Instance: 30041

User Name:

Password:

After that click on Connect.

We can see the analytical view which we have created in the above screenshot.

Click on the view for the table “\$TA_SurfacePro4_Sentiment_379” and then click on Next.

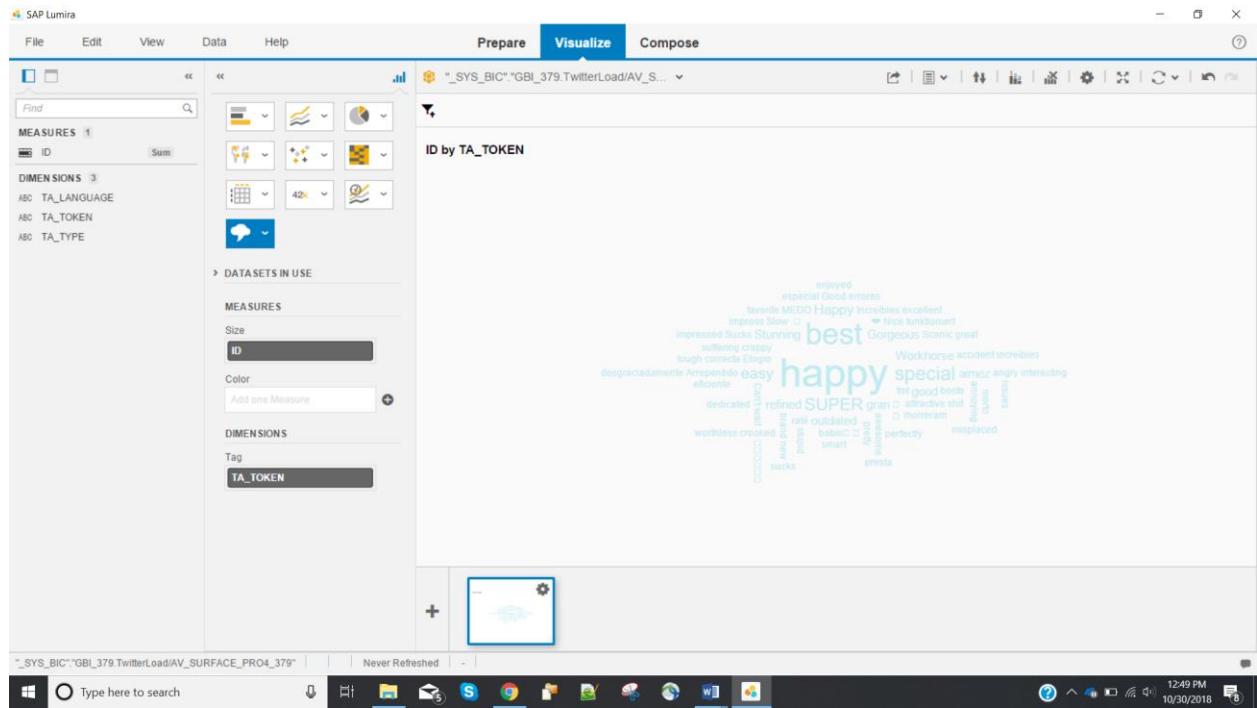
Click on Create.

The Visualize perspective opens. Drag and drop the measure “Id” to the field “X Axis” under “MEASURES” on the right of the screen. Then drag and drop the dimension “TA_TOKEN” to the field “Y Axis” under “DIMENSIONS” on the right of the screen. The data will be shown as a bar diagram.

Change the type of the visualization to tag cloud. You can do this by clicking on the icon for others charts. Then, choose Tag Cloud.

The tag cloud visualizes all sentiments that are stored in the database table. Terms that have been used more frequently are displayed in a larger font size than the others.

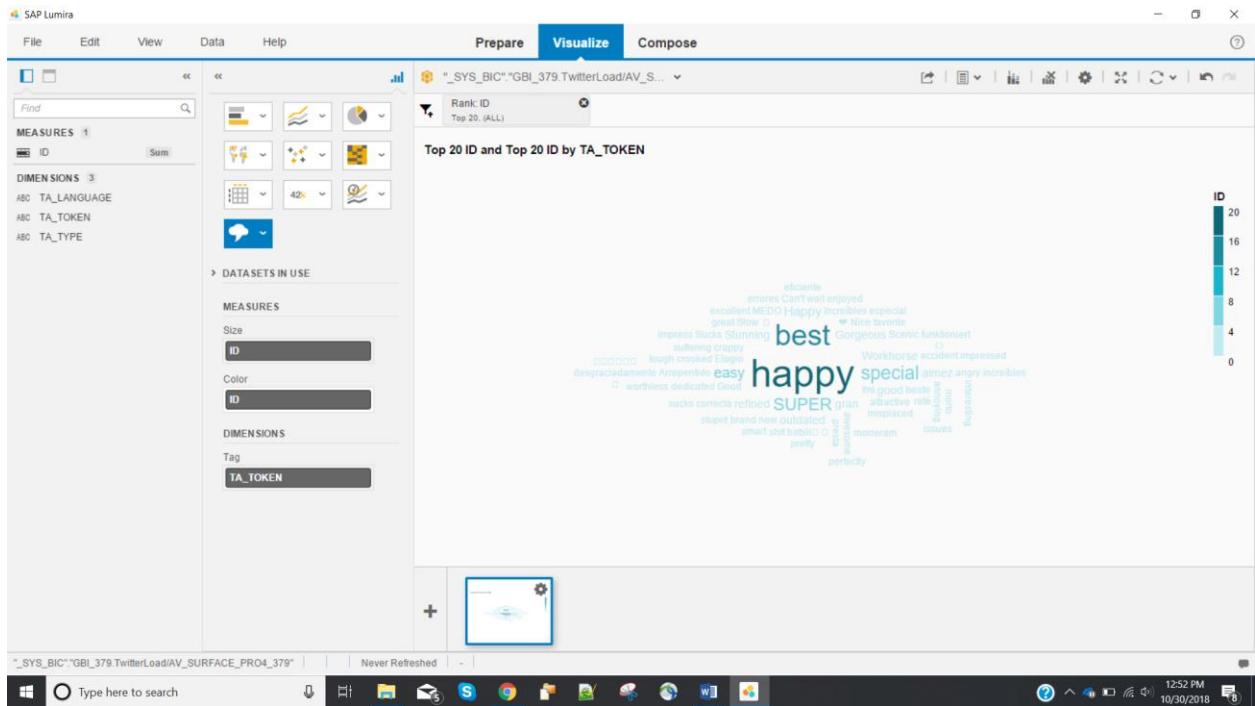
Screenshot after changing the type of visualization to tag cloud for Surface_Pro4 is:



Now, drag and drop the measure “Id” to the field “Color” under “MEASURE” on the right of the screen. Terms that have been used very often will be displayed in a darker color. On the right of the tag cloud legends with the values will be shown.

The tag cloud is still not really readable because there are so many terms in it. To solve this issue, you can add a ranking for the column “Id”. Therefore, click on the ranking symbol at the top of the tag cloud. Enter 20 into the field with the number to show the top 20 sentiments.

The screenshot to show the top 20 sentiments for Surface Pro4:



Now add a filter for one language. Therefore, click on the filter symbol on the top of the table. Choose the dimension “TA_LANGUAGE”.

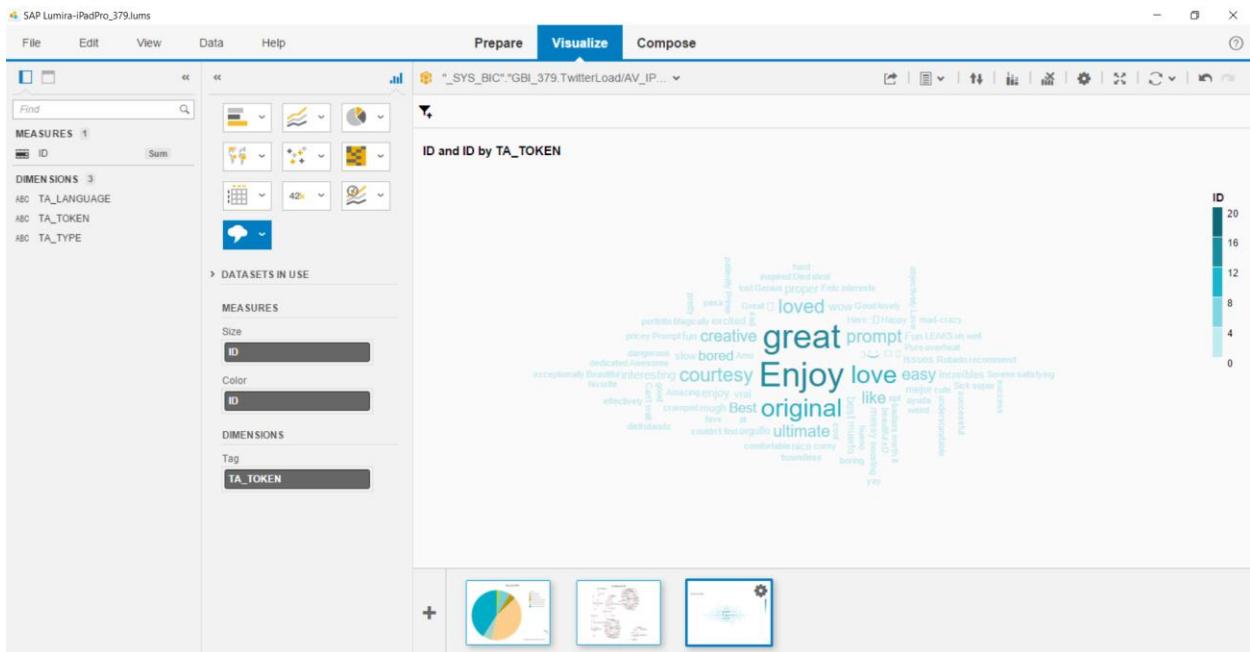
Choose one of the languages from the value list (e.g. “es” for Spanish) and click on **Apply**. If you don't have ‘es’ in your dataset, choose any language except for English. Likewise, replace the chosen language in place of Spanish in the further steps and for repeating it for the other product.

Now your tag cloud changes again and only show terms that were written in Spanish.

Adjust the diagram for your manager. Then click on **File** and then select **Save (Ctrl + S)**. Enter SurfacePro4_379 into the field "Name" and click on **OK**.

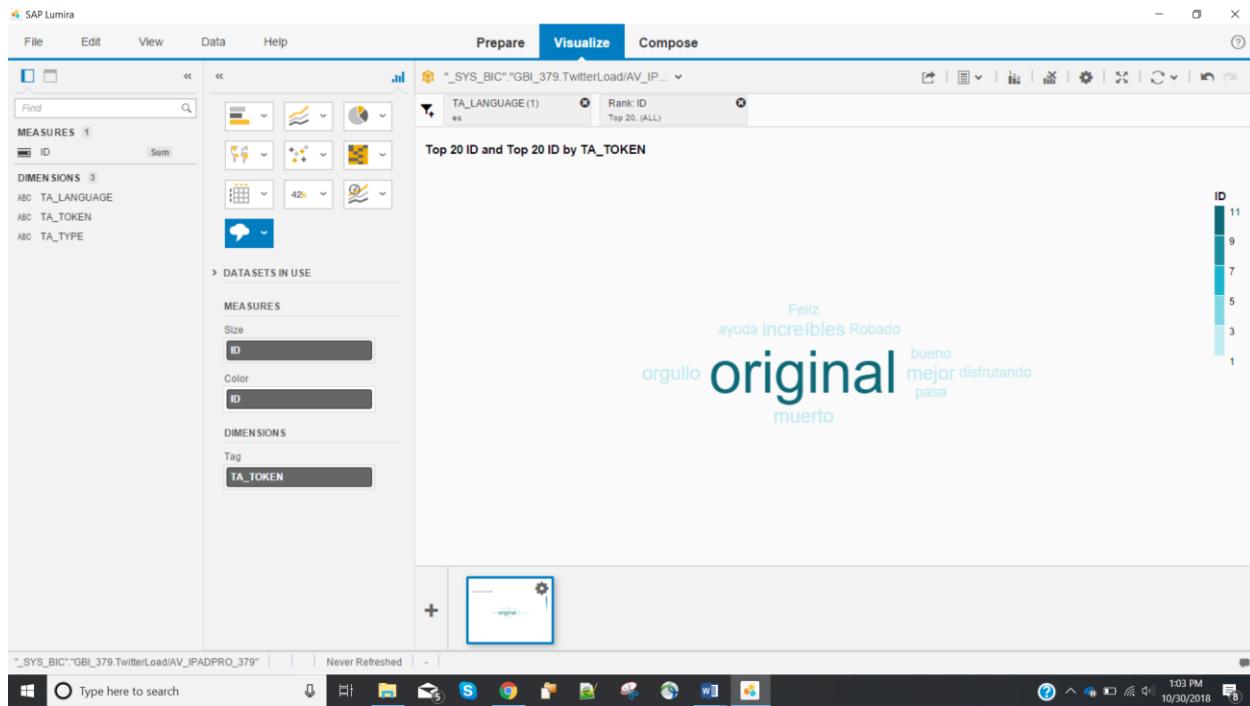
Now you can repeat the steps for the table with the **iPad Pro** analysis results. In our case, the result for the tag cloud without any additional configurations or filters looks like shown in the screenshot below.

Screenshot after changing the type of visualization to tag cloud for iPad_Pro and without any filters is:



If we filter the analytic view for language “Spanish” and display only the Top 20 ranks by ID, we receive the following result for the iPad Pro.

The screenshot after filtering the language for Spanish and top 20 sentiments is:



After filtering based on the language and listing the top 20 sentiments, it contains positive, negative and neutral sentiments.

These are a few of the positive sentiments:

Happy, Incredibles, Good, Best, enjoying

These are a few of the negative sentiments:

Pride, stolen, dead

These are a few of the neutral sentiments:

Original, help, pass

Save the configuration under the name "iPadPro_###".

3) Create a Network Chart with SAP Lumira Server

After creating the tag cloud production planner notices that it doesn't provide any information about the type of the sentiments. He wants to know if there are more positive or negative sentiments. To analyze the data, he creates a Network Chart.

Reopen the SAP Lumira and then double click on the **iPadPro_379**

The “Visualize” perspective opens and shows your tag cloud. You can edit the view if you want. We want to create an additional chart. Click on the **plus** symbol on the bottom of the chart.

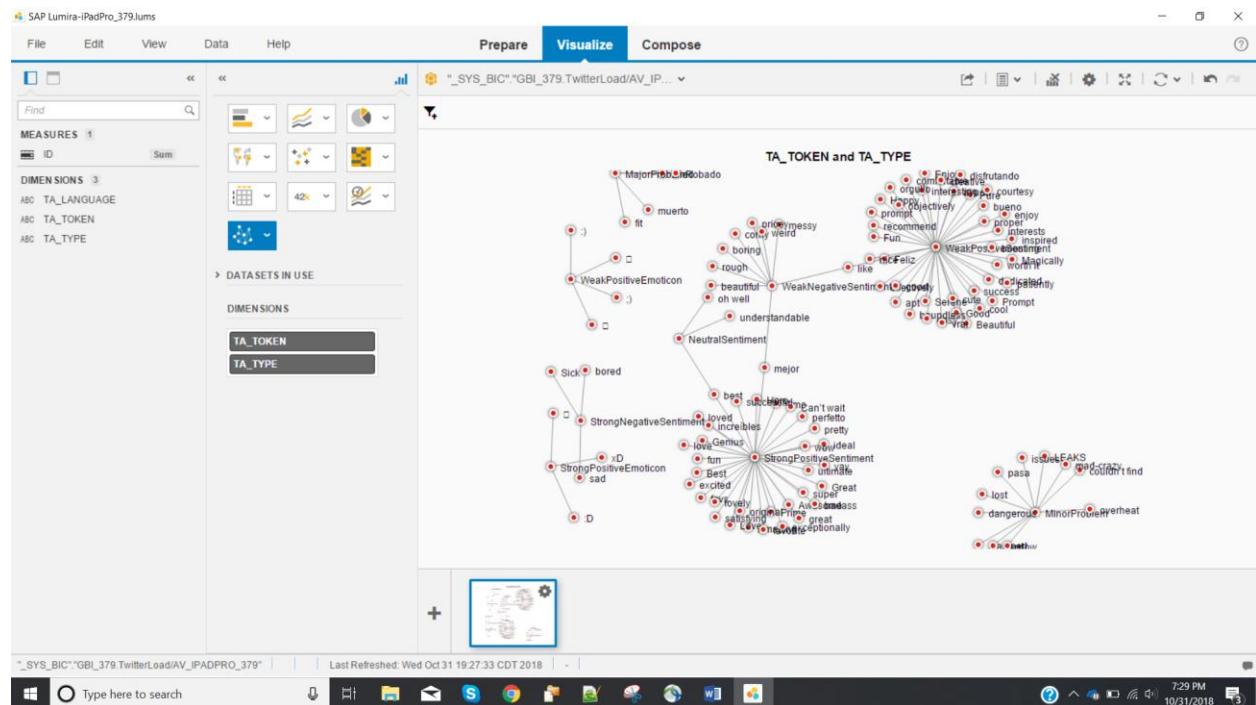
An empty chart will open.

Change the visualization type to “Network Chart”. Therefore, click on the icon for others charts and choose “Network Chart”.

Network charts use dimensions only. Drag and drop the dimension “TA_TOKEN” to the field under “DIMENSIONS” on the right of the screen. Note that the chart in the middle of the screen asks you to add more dimensions to visualize the data.

That means that we have to add a second dimension. Drag and drop the dimension “TA_TYPE” to the field under “DIMENSIONS” on the right of the screen.

The screenshot of the Network Chart of the iPadPro is:

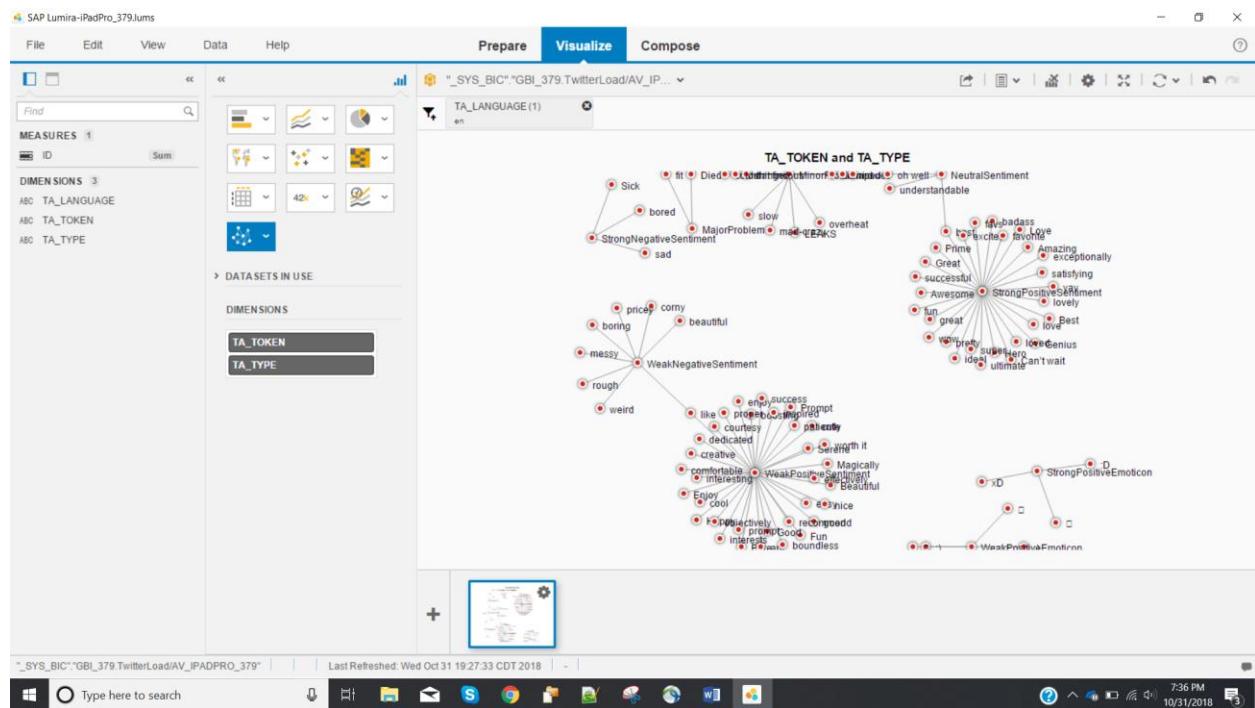


If in case the network chart doesn't look really readable yet then, we need to configure the charts first. Click on the filter symbol at the top of the chart and choose the dimension "TA_LANGUAGE". Add one language to the list, e.g. "en" for English.

If you click on one spot in the middle of one network segment, you can drag and drop it to another place in the chart. The other network segments will organize themselves. Organize the segments that the positive sentiments are on the right side and the negative sentiments are on the left one.

In the screenshot above you can see that there are lots of minor problems but just a few major problems. You can also see that there are lots of positive statements about the iPad Pro and some positive emoticons.

The screenshot after filtering based on language and arranging the positive sentiments on right and negative sentiments on left is:

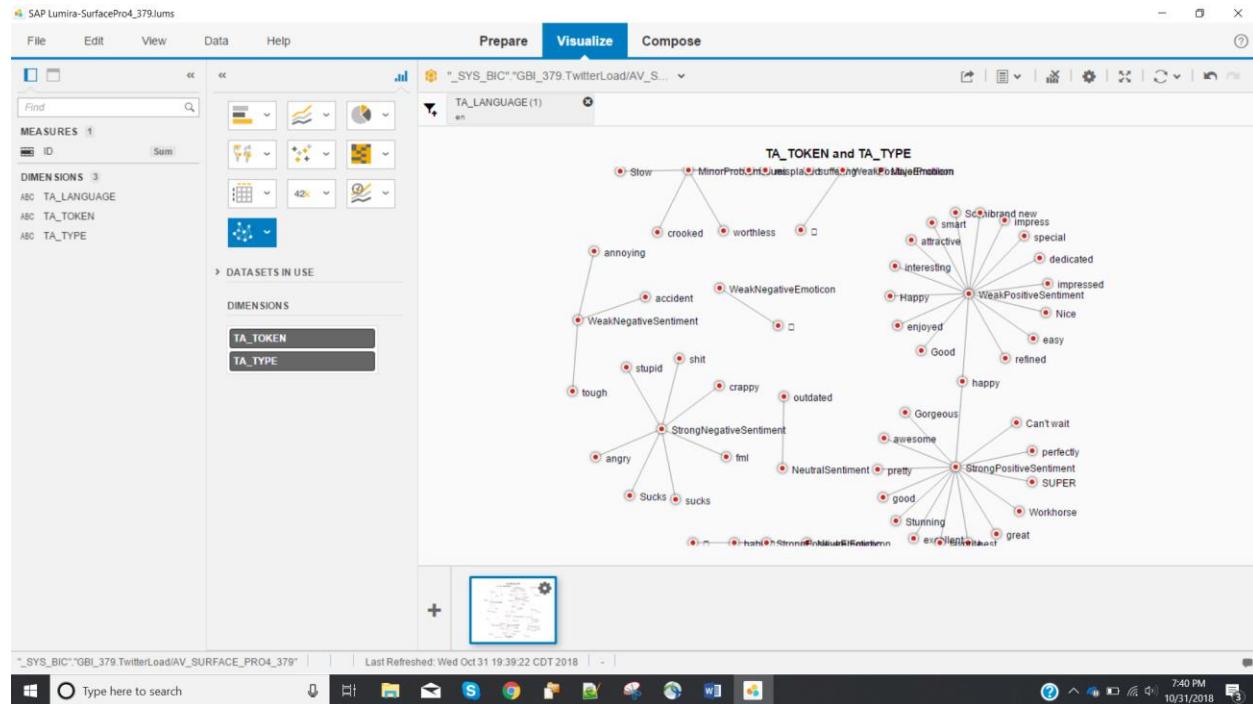


Weak Positive Sentiment has the maximum number of nodes.

Click on the **save** button in the upper right corner to save the configuration. You can repeat the steps for the view with the data of the **Surface Pro 4** (Ensure you include the language filter for English). In our case the result looks like shown in the screenshot on the next page. The graph is filtered for

In this case you can see that there are much more positive sentiments than negative sentiments.

The network chart for Surface Pro4 after filtering the language and arranging positive and negative sentiments on opposite sides is:



Next, we want to know what that means in numbers.

4) Create a Pie Chart with SAP Lumira Server

After creating the network chart production planner needs some real numbers to show to the manager. He creates a Pie Chart to see if there are more positive or negative sentiments. Afterwards he creates the presentation for the manager.

Restart the SAP Lumira. Double click on **iPadPro_379** table. Enter the password.

The “Visualize” perspective opens and shows your charts. You can edit the view if you want. We want to create an additional chart. Click on the plus symbol on the bottom of the chart.

An empty chart will open. Change the visualization type to “Pie Chart”. Therefore, click on the icon for pie charts and choose “Pie Chart”.

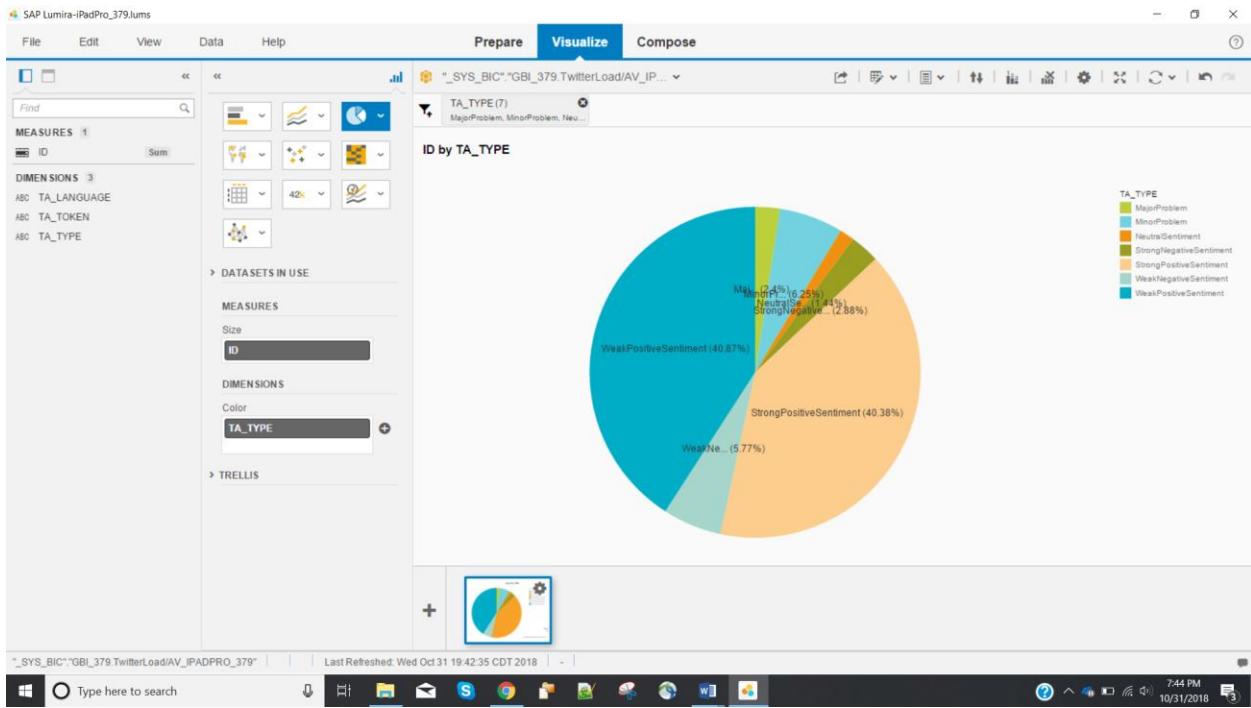
Drag and drop the dimension “TA_TYPE” to the field “Colors” under “DIMENSIONS” on the right of the screen. Then, drag and drop the measure “Id” to the field “Size” under “MEASURES” on the right of the screen.

If you want, you can filter out the emoticons, as they don’t influence the overall pie chart that much. Click on the filter symbol at the upper left corner and choose TA_TYPE.

Now choose “Is in List” from the dropdown field and select “Major Problem”, “Minor Problem”, “Neutral Sentiment”, “WeakPositiveSentiment”, “StrongPositiveSentiment”, “WeakNegativeSentiment” and “StrongNegativeSentiment”. Click on **Apply** to set the filter.

To enable the data labels for the pie chart, right click on the chart and choose **Show Data Labels**.

The percentages for all sentiments are shown in the pie chart now. The screenshot of pie chart for iPadPro is:



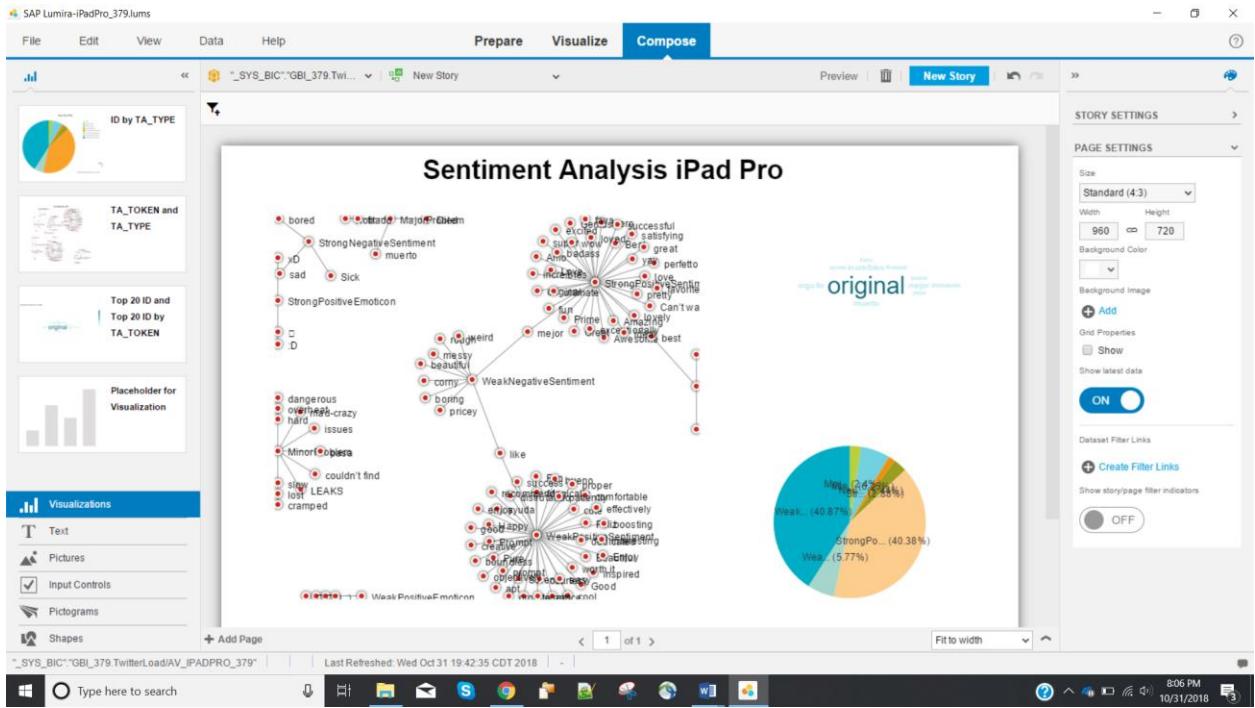
Save your visualization by clicking on the **Save** button on the upper right corner. Next, we want to create a presentation for the meeting with the manager. **Therefore, change to the view “Compose” at the top of the screen. Choose the blank template.**

Our charts are shown on the right side of the view. Drag and drop the charts to the white page in the middle of the screen to add them to the presentation. Arrange the charts to present the content vividly. In the next screenshot, you can see one example how they can be arranged.

For the presentation we don't need the titles of the charts. Click on one element, e.g. the network chart. Then, click on the “explore” symbol on the upper right corner of the element. A new screen to explore and change the chart opens. Right click on the element and choose **Show Title**. Uncheck the Show Title.

Then click on the button **Update** to save your changes. Do the same for the pie chart and the tag cloud. If you want, you can also disable the legends for the pie chart and the tag cloud. Therefore, go to the explore mode, right click on the element and choose **Show Legend**.

To add a title to the slide, go to “Text” on the lower left corner. Drag and drop the tile “Title” to the slide in the middle of the screen and change the text to “Sentiment Analysis iPad Pro”.



Go to file and select Export as File.

Click on it.

Our slide will be downloaded. Now you can open it to show the presentation to the manager. Repeat the steps for this scenario for **Surface Pro 4** to compare the results. (Ensure we have the filter for TA_TYPE on the Pie Chart)

Screenshot after exporting Surface Pro4 as a pdf is:

