**DATA VISUALIZATION CASE STUDY 2 - GROUP 11**

***TOPIC: Analysis of US Elections 2016***

**Introduction:**

We are the data analytics team of a news channel agency- VOX News. We would be presenting our analysis of US Elections 2016 to the editor of VOX News for a news story. The tableau story starts off with an overview of US Elections, campaigning in swing states and highlights the importance of voting giving the message that ‘Every vote matters!’

**Data Sources:**

We have consolidated our data from a variety of sources. There are two main excel workbooks. The workbook- ‘US Elections Dataset’ is a composition of primary and secondary datasets. The second excel workbook- ‘Sentiment Data’ has data imported from SAP HANA. The detail description can be found below:

**I. Primary Data Set:**

* **File Name:** US\_County\_Level\_Presidential\_Re
* **Source:** <https://www.kaggle.com/benhamner/2016-us-election>

<https://github.com/tonmcg/County_Level_Election_Results_12-16>

**II. Secondary Data Set:**

* **File Name:** Promotions
* **Source:** <http://www.nationalpopularvote.com/campaign-events-2016>
* **Information gathering:** We created the excel file by collecting data from the above web page.
* **File Name:** Resolution
* **Source:** <http://www.cnn.com/election>
* **Information gathering:** We created the excel file by picking up data for each state from the above webpage under ‘presidential elects’ title.

**III. Sentiment Analysis:**

* **File Name:** Sentiment Data
* **Source:** SAP HANA and Twitter API
* **Table:** $TA\_Election16

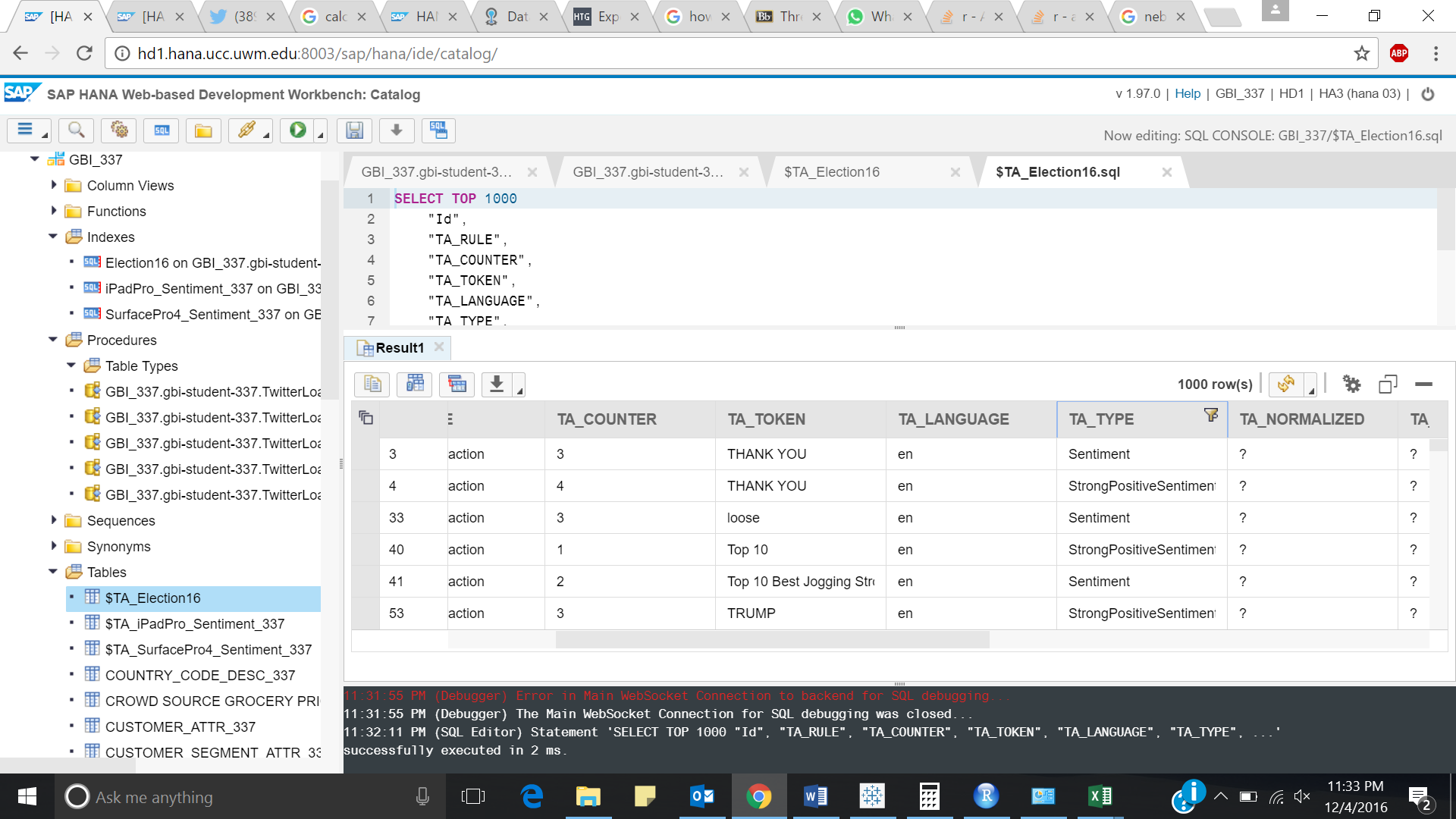


Figure 1: SAP HANA Web IDE

* **Information gathering:** Performed text analysis using twitter adapter and collected the data by filtering the ‘Status’ table with tweet like ‘Election 2016’.

**Key field used in join:**

We have used ‘state code’ to join the primary and secondary datasets using left outer join. The following insights have been obtained from the data: -

**Insights:**

* **Overview of US Elections 2016:**

Comparing the County wise election map of 2012 & 2016, we can see the map of 2016 is more inclined towards the Republicans. As in Pennsylvania, there were some counties with majority of Democrat Votes, but still Electoral votes are gained by Republicans.

* **Campaigning in swing states:**

The following insights were obtained after combining the ‘US\_County\_Level\_Presidential\_Re’ and ‘promotions’ dataset.

1. 2/3rd of the Total Campaigns were run only in 6 states.
2. Republicans focused more on Swing States and won majority of them.

* **Demographic of voters:**

1.Per Capita Income: High per capita Income groups preferred Democrats over Republicans.

2. Sex Ratio: Females seemed to have favored Democrats over Republicans in most of the states.

* **Sentiment Analysis on Social Media-Twitter:**

The twitter sentiment analysis implies that majority of people are indeed happy with their new president. Although there is a bias here as we don’t know whether people who tweeted are from US and voted.

* **Results Using Nebraska and Maine’s “Proportional Method”:**

Proportional Method: In Nebraska and Maine, the winner gets the senate votes and the other votes are distributed proportionally.

After applying the proportional method followed in Nebraska and Maine to all the states, it was found that there was no change in the 2016 Election results.

* **Percentage Voting:**

In the States where Republicans won, majority of the voting percentage was below the average level while in the states where Democrats won, majority of the percentage voting was above the average line.

* **Relation between margin of votes and people who did not vote:**

People who didn’t vote, thinking their vote doesn’t matter, had the power to change the results as the percentage of people who did not cast their vote is higher than percent margin of votes between the two parties, in all the states.

Also, when the margin of votes between two parties is less, votes the vote matters more.

SO, EVERYONE SHOULD CAST THEIR VOTE AS EACH VOTE MATTERS!

**Calculated Fields:**

|  |  |  |  |
| --- | --- | --- | --- |
| 1 | Current Method | IF [Democrat House] > [Republican House] THEN [Electoral Vote] ELSE [Electoral Vote] END | According to the current method, if a candidate wins maximum number of the districts(Houses), He takes all the Electoral Votes of that state. |
| 2 | Dem lose | [Electoral Vote]- [Republican Takes] | In "Proportional Method", if Democrat has maximum number of houses, but Republicans will get the Houses they won and democrats will lose those electoral votes |
| 3 | Democrat Takes | IF [Total Democrat Votes]>[Total Republican Votes] THEN [popular dem vote] ELSEIF [Total Democrat Votes]<[Total Republican Votes] THEN [Electoral Vote]-[popular rep vote] END | To visualize the sheet "Proportional distribution of electoral vote as per new and current methods", this calculated field is used to assign value to the "Dem Wins" row. |
| 6 | Percent Democrat votes | ([Dem vote 2016]/[Adult population])\*100 | percentage of people who voted for Democratic Party |
| 8 | Percent Republican Votes | ([Gop votes 2016]/[Adult population])\*100 | percentage of people who voted for Republican Party. |
| 9 | popular dem vote | [Democrat House]+[Senates] | Total Electoral Votes Republican Party gets if they win according to the "Proportional Method" |
| 10 | popular rep vote | [Republican House]+[Senates] | Total Electoral Votes Democratic Party gets if they win according to the "Proportional Method" |
| 11 | popular vote color | [Democrat Takes]-[Republican Takes] | Used for the purpose of color encoding. |
| 14 | rep lose | [Electoral Vote]-[Democrat Takes] | In ‘Proportional Method’, if Democrat has maximum number of houses, but Republicans will get the Houses they won and democratic will lose those electoral votes |
| 18 | Republican Takes | IF [Total Republican Votes]>[Total Democrat Votes] THEN [popular rep vote] ELSEIF [Total Republican Votes]<[Total Democrat Votes] THEN [Electoral votes]-[popular dem vote] END | To visualize the sheet "Proportional distribution of electoral vote as per new and current methods", this calculated field is used to assign value to the "Dem Wins" row. |
| 19 | Shapes | IF [Total Democrat Votes]>[Total Republican Votes] THEN 1 ELSE 0 END | Binary value Returning Function for Visual Encoding. |
| 20 | Rep Wins | IF [Democrat House]>[Republican House] then [Electoral Vote] ELSEIF [Democrat House]<[Republican House] THEN [Electoral Vote] END | "Winner Takes All", if Republican Party won the Maximum number of houses. |
| 21 | Dem Wins | IF [Democrat House]>[Republican House] then [Electoral Vote] ELSEIF [Democrat House]<[Republican House] THEN [Electoral Vote] END | "Winner Takes All", if Democratic Party won the Maximum number of houses. |
| 23 | Lead | IF [Democrat Takes]>[Republican Takes] THEN [Democrat Takes]-[Republican Takes] ELSE [Democrat Takes]-[Republican Takes] END | Function to return the value of lead of the winning party. |
| 24 | Lead of Democrats | IF ([Democrat Takes]-[Republican Takes])>0 then [Democrat Takes]-[Republican Takes] ELSE 0 END | Function to return the value of lead of democrats |
| 25 | Lead of Republicans | IF [Democrat Takes]>[Republican Takes] then [Republican Takes]-[Democrat Takes] ELSE 0 END | Function to return the value of lead of democrats |
| 26 | Percent Margin | ((ABS([Total Democrat Votes]-[Total Republican Votes]))/[Adult population])\*100 | Absolute value of percentage of difference of votes for the two parties. |
| 27 | Percent Others Votes | ABS ([Total Other Votes]/[Adult population]) | Ratio of eligible population who casted their vote. |
| 28 | Percent People Didn't Vote | (([Adult population] - [Sum of Total\_votes\_2016])/[Adult population])\*100 | Percentage of eligible population who did not cast their vote. |
| 29 | Percent votes casted | [Percent Republican Votes] + [Percent Democrat votes] +[Percent Others Votes] | Percentage of eligible population who casted their vote. |
| 30 | Percent Who didn’t Vote | (([Adult population]-([Total Democrat Votes]+[Total Republican Votes]+[Total Other Votes]))/[Adult population]) | Percentage of eligible population who did not cast their vote. |
| 32 | Total houses | [Democrat House] +[Republican House] | Returns the number of districts. |
| 33 | Vote Matters | [Percent People Didn't Vote]-[Percent Margin] | Percent eligible population greater than the percent margin of the votes to the parties. |