**Capstone Project II Final Report: TED Talks - Visualize the Speakers based on the Chosen Topic**

**Baskaran Viswanathan**

**06/26/2018**

**Abstract**

This project shows the options of choosing the category followed by its top speakers based on many aspects of the data collected from TED Talks website until September 21st, 2017.

1. **Introduction**

This is the final report for the capstone project – 02 in Springboard. The dataset used for this project downloaded from Kaggle. Details of the Data set and Analysis are available in GitHub, which link has been given at the end of this document.

1. **Problem Statement**

Giving an option of choosing the topic and total no of top rated speakers.

* + Most popular Talks by category, views, comments, etc.
  + Do certain videos get popular after certain events? (i.e: recession in ‘08)
  + Categories: education, tech, health, love
  + Look at rating information to see what factors contribute to the success of a video

1. **Data Description**

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Description** | **Data Type** |
| Comments | The number of first level comments made on the talk | Numeric |
| Description | A blurb of what the talk is about | String |
| Duration | The duration of the talk in seconds | DateTime |
| Event | The TED/TEDx event where the talk took place | String |
| film\_date | The Unix timestamp of the filming | DateTime |
| languages | The number of languages in which the talk is available | Numeric |
| main\_speaker | The first named speaker of the talk | String |
| name | The official name of the TED Talk. Includes the title and the speaker. | String |
| num\_speaker | The number of speakers in the talk | Numeric |
| published\_date | The Unix timestamp for the publication of the k on TED.com | Numeric |
| ratings | A dictionary of the various ratings given to the talk (inspiring, fascinating, jaw dropping, etc.) | String |
| related\_talks | A list of dictionaries of recommended talks to watch next | String |
| speaker\_occupation | The occupation of the main speaker | String |
| tags | The themes associated with the talk | String |
| title | The title of the talk | String |
| url | The URL of the talk | String |
| views | The number of views on the talk | Numeric |

1. **Data Wrangling Details** 
   1. **Initial findings**
      1. The data set does not have either the image or the URL for the image of the speaker for each talks
      2. The existing data format is in CSV, which cannot be used for any API
      3. Total no of categories’ are huge and cannot be included as part of the user selection, as it will be out of the screen
      4. Extracting the text using the CSV file is very difficult
   2. **Steps taken to overcome the above issues**
      1. To get the URL for images

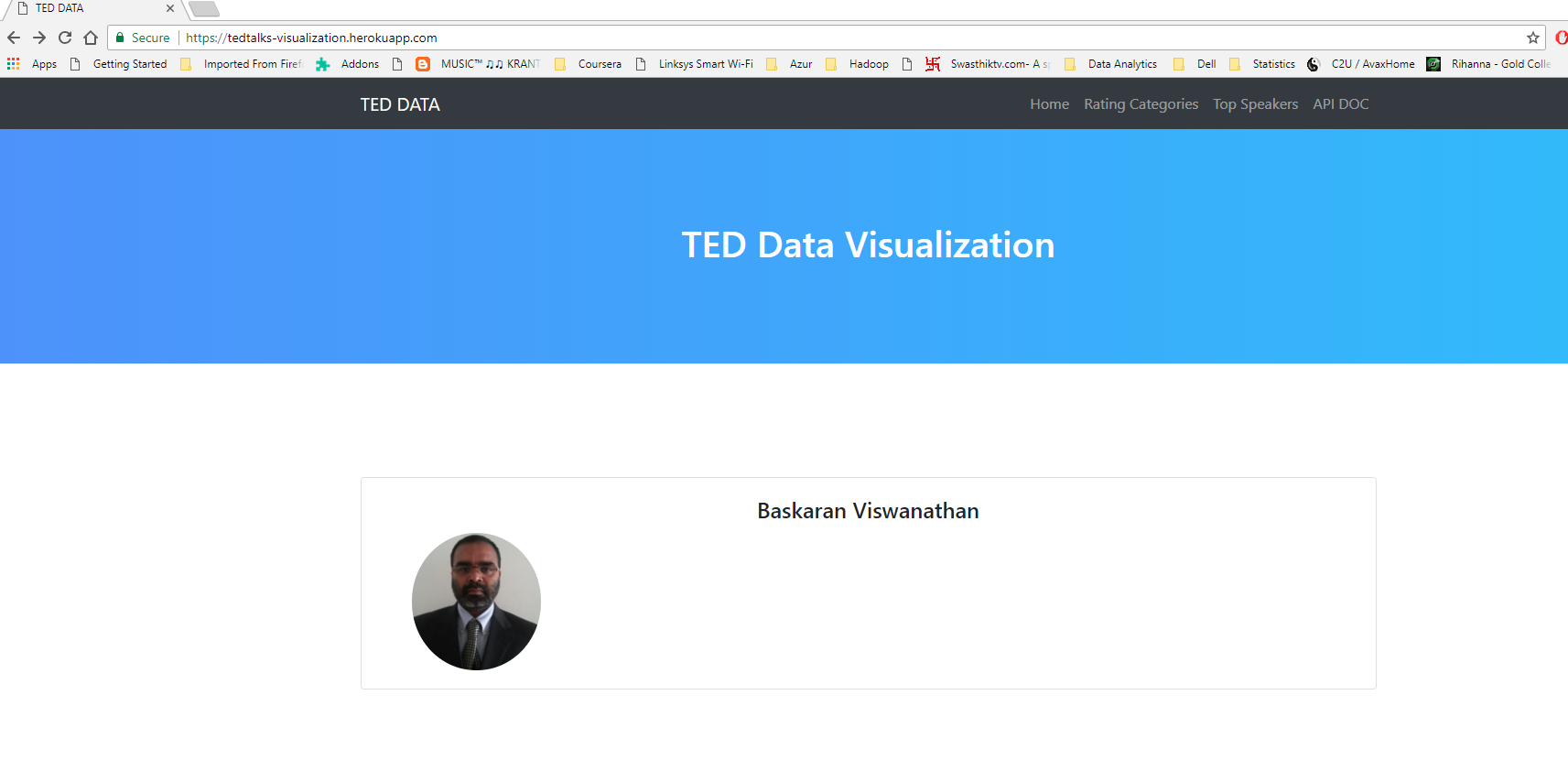
We have collected the image URL of the speaker by collecting their image URL from Ted.com website, by reading each observation, using the Chromedriver, Beutifulsoup Library.

Using Pandas library the collected image URL added to the existing dataset, before uploading the data to the database

* + 1. Upload the details to Mongo DB. This way it would be easy for us to generate the API Data in a JSON Format
    2. The selected Categories data(as mentioned below), will be extracted from the MongoDB, using PyMongo Library

1. **Application Access**
   1. To access the Application, please click the following link.

<https://tedtalks-visualization.herokuapp.com/>



* 1. The link to view the visualization

<https://tedtalks-visualization.herokuapp.com/rating_Categories.html> or

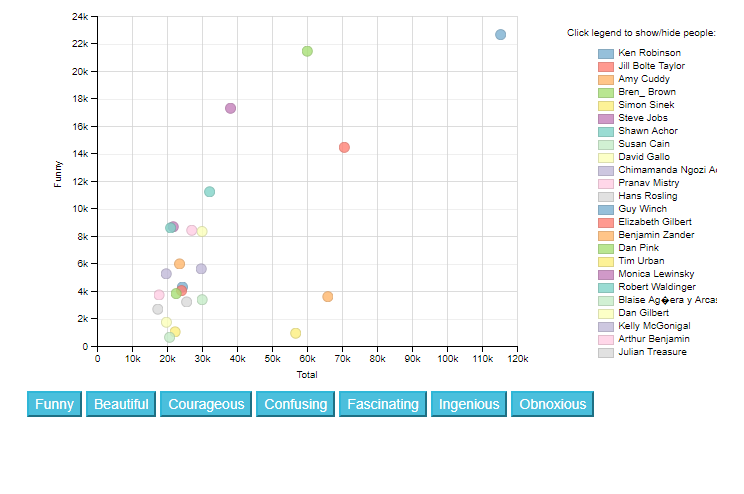
Rating Categories Selection from the above screen

* 1. **Categories Visualization**

Only the following categories are available for this Visualization

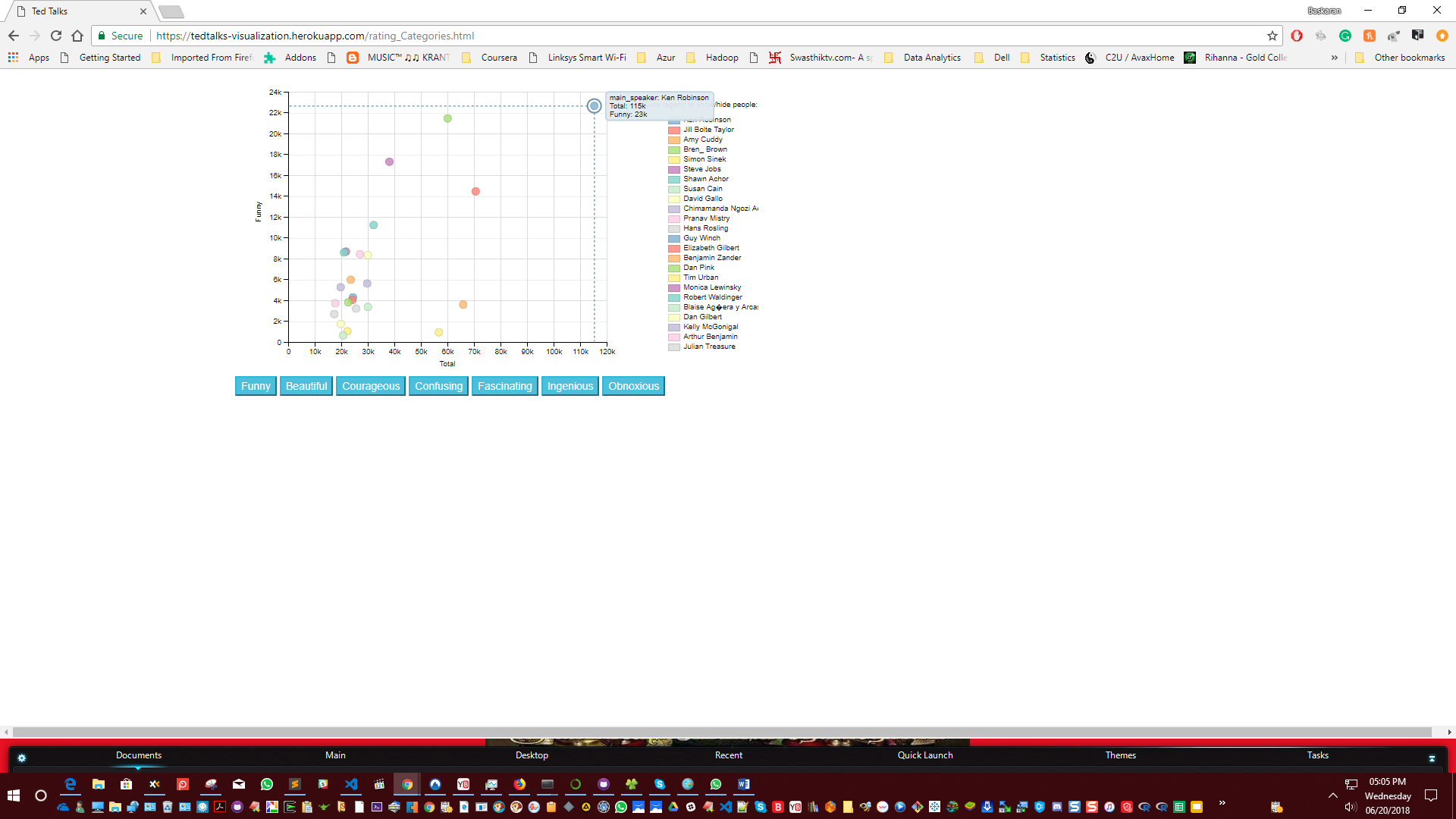
* + 1. [Funny](https://tedtalks-visualization.herokuapp.com/rating_Categories.html)
    2. [Beautiful](https://tedtalks-visualization.herokuapp.com/assets/static/beautiful.html)
    3. [Courageous](https://tedtalks-visualization.herokuapp.com/assets/static/courageous.html)
    4. [Confusing](https://tedtalks-visualization.herokuapp.com/assets/static/confusing.html)
    5. [Fascinating](https://tedtalks-visualization.herokuapp.com/assets/static/fascinating.html)
    6. [Ingenious](https://tedtalks-visualization.herokuapp.com/assets/static/ingenious.html)
    7. [Obnoxious](https://tedtalks-visualization.herokuapp.com/assets/static/obnoxious.html)

The right side is showing the top 24 Top Legendry people, whose videos are watched by many people. The list is in Descending Order



When we click any of the colored circle inside the graph, it will show the Name of the speaker, total no of views and the category, as shown in the below image.

Whenever we click any name in the right, that particular speaker, will be removed from the visual.



***Blue buttons below is show the list of available categories***

* 1. **Top Speakers**

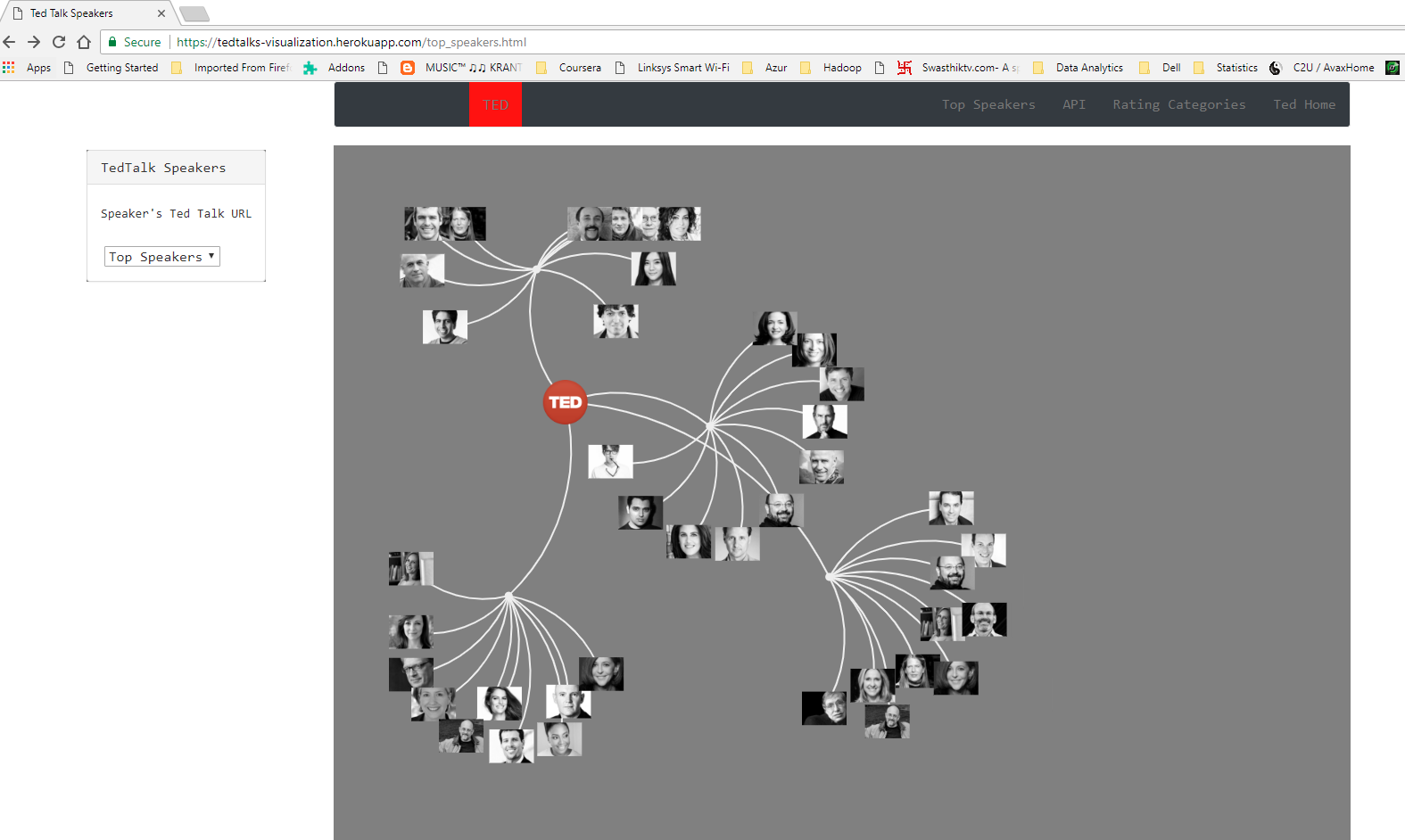
When we click the Top Speakers Link, the below web page will be opened. By default, it will show Tree of Top 10 Speakers for 4 Topics

Science

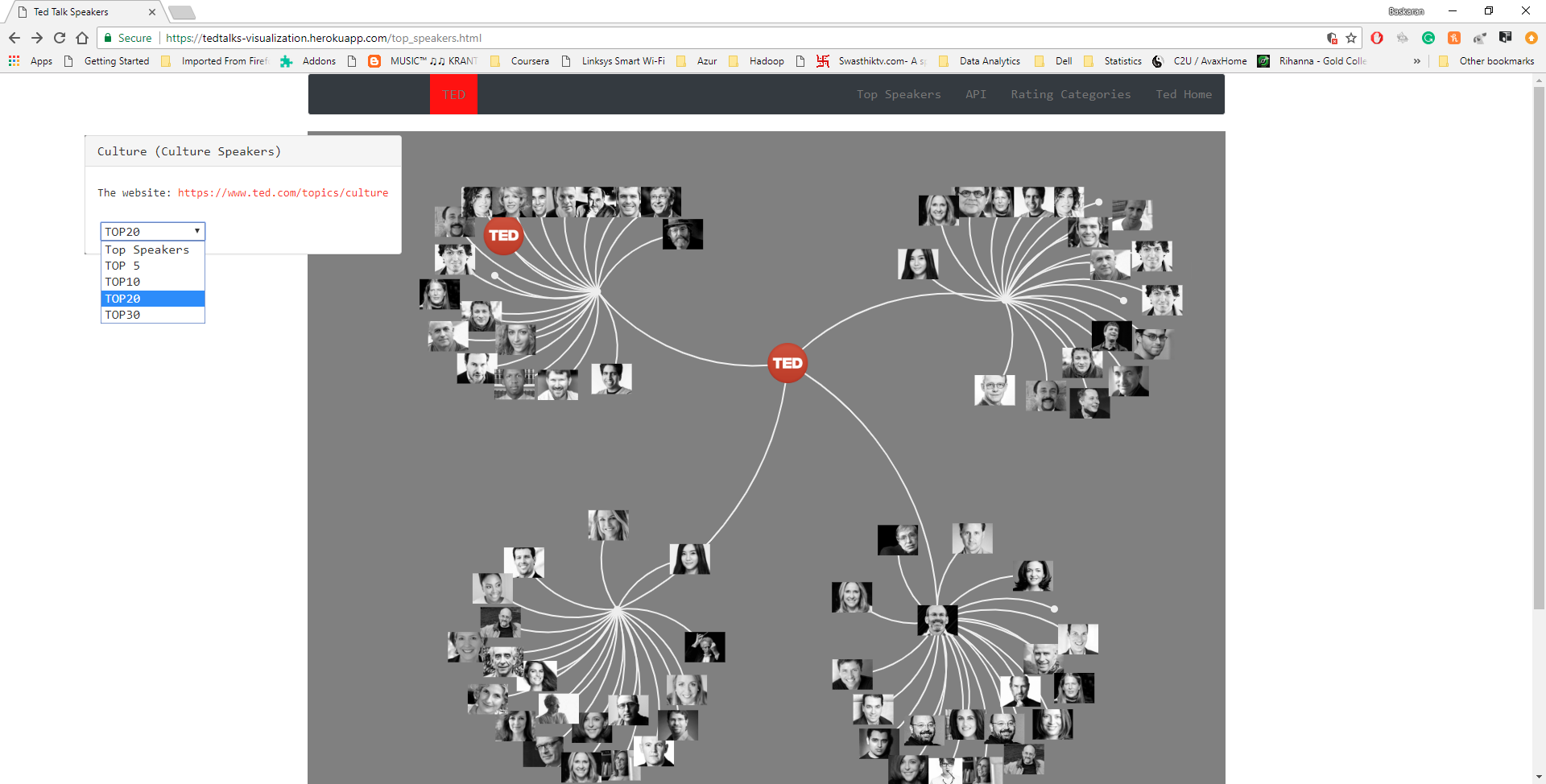
Technology

Global Issues

Culture



**The no of Top Speakers can be changed from Top5 to Top20, as shown in the below picture**



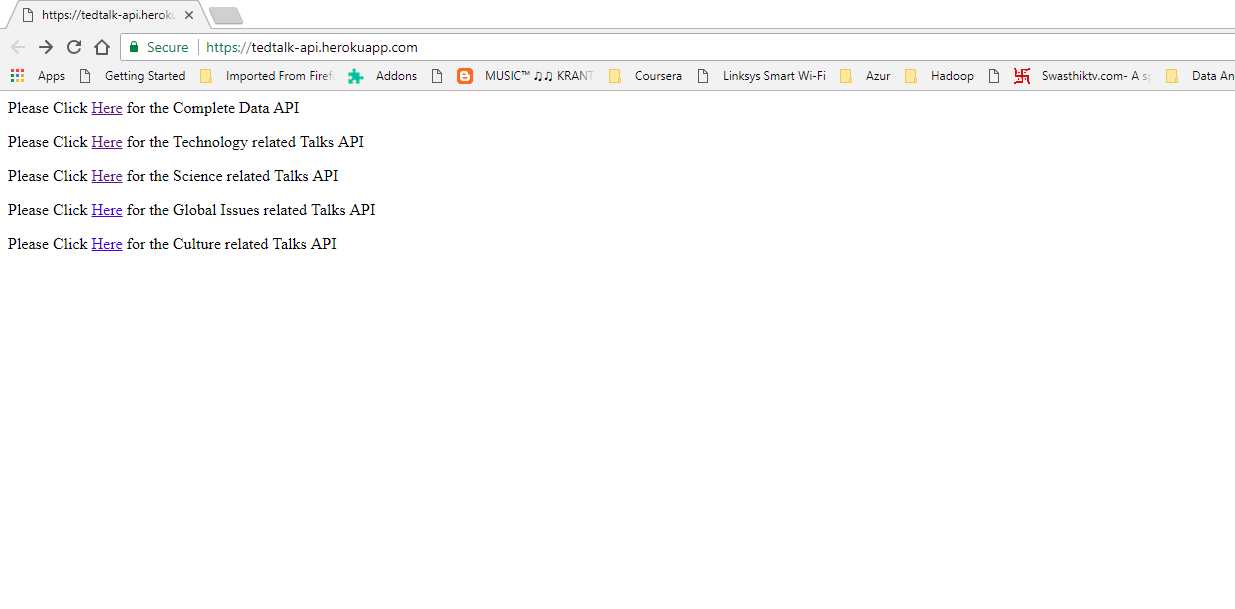
**Based on the selection the right side will change dynamically, for all the 4 topics.**

**When we go to near center point, it will show the topic name of the subtree. If we click the center of any subtree, that topics tree will be removed, it would be back once we click it again.**

**If we click any of the speaker, their talk for that corresponding topic, ted talks link will be displayed on the left. Once we click, it will be opened in a new tab in the same browser.**

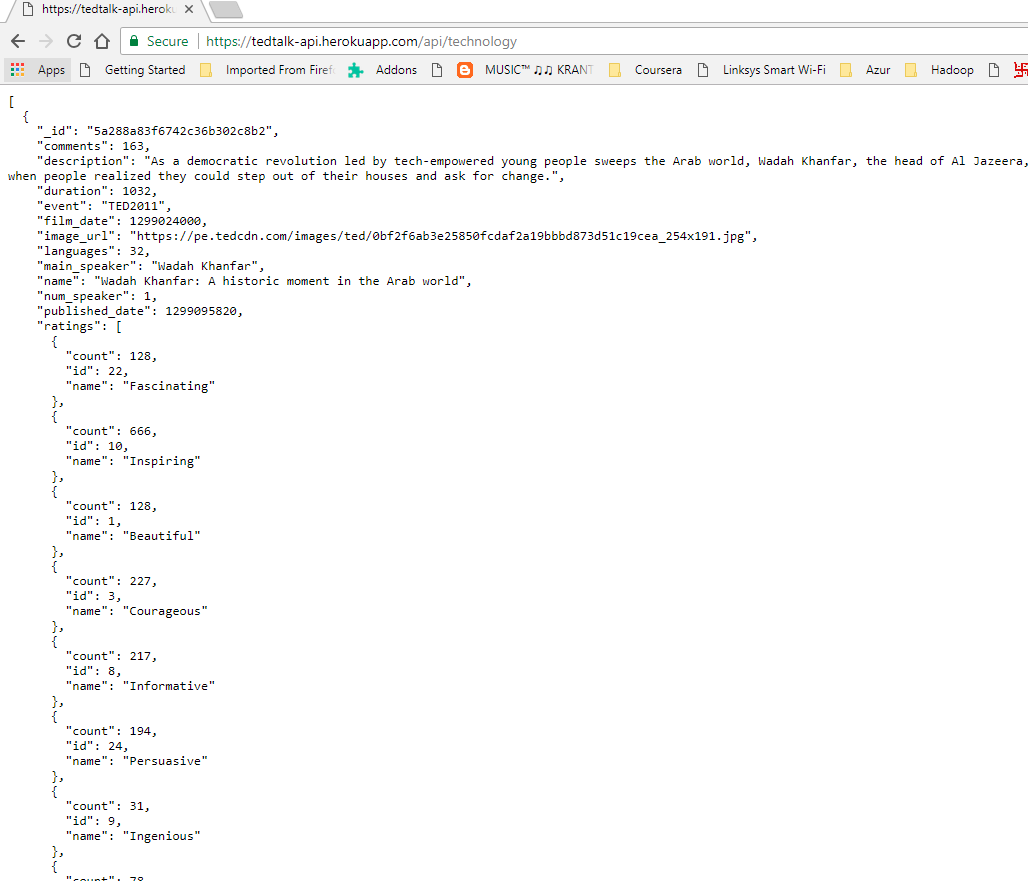
* 1. **API**

**When we click the API link, then it will be opened the below screen.**



**There are 5 links available, as follows**

1. **Complete Data**
2. **Data related to Technology Topic**
3. **Data related to Science Topic**
4. **Data related to Global Issues Topic**
5. **Data related Culture Topic**



**When we right click the webpage we can save the data in a JSON Format file. The same applicable for all the topics.**

1. **Road to reach the destination**

Software Used

Data Wrangling

Python

BeautifulSoup

Pandas

Numpy

PyMongo

Scipy

Web Page

Flask

Visualization

Dimple.js

D3.js

We have uploaded the complete application in Heroku for easy access by the users

The Application Link is

<https://tedtalks-visualization.herokuapp.com/>

1. **Conclusion**

In this Project, we have created an easy way to watch the ted talks video, by selected topics, top speakers with in the selected topics, instead of searching and viewing directly from the Ted.com website. It is just a try, but there are more options can be improved and facilitated.

**Dataset Source**

<https://www.kaggle.com/rounakbanik/ted-talks>

GitHub

<https://github.com/vibaskaran/SpringBoard/tree/master/Capstone-02>