**Group 4:**

Sushant Burde

Swarnim Jambhule

Kathryn Mak

Pratik Parmar

Monish Phatarpekar

Smit Shiroya

**Topic** : Prediction of Bitcoin Currency

**Abstract:**

A cryptocurrency is a digital money designed to work as a medium of exchange in an electronic payment system in which payments are validated by a decentralized network of system users and cryptographic protocols instead of by a centralized intermediary (such as a bank). Since 2009, cryptocurrencies have gone from being academic concept, niche technological curiosities to rapidly proliferating financial instruments that are the subject of intense public interest. Recently, they have been incorporated into a variety of other financial transactions and products. For example, cryptocurrencies have been sold to investors to raise funding through initial coin offerings (ICOs), and the terms of certain derivatives are now based on cryptocurrencies. Some government central banks have examined the possibility of issuing cryptocurrencies or other digital currency. Like any investment, these can carry risk. Media coverage of cryptocurrencies has been widespread, and observers have characterized cryptocurrencies as either the future of monetary and payment systems that will displace cash, government-backed currencies, or a fad with little real value.

**Background:**  
The world of cryptocurrency and blockchain are constantly evolving and fascinating to many. As one of the most valuable virtual currency that is completely paperless and autonomous, Bitcoin has been the topic of discussion among financial investors, stock traders, software programmers, and the public in general. For centuries in history, people rely on printed money as the major medium of transacting business. Since its first introduction by a Japanese person who uses the pseudonym of Satoshi Nakamoto, Bitcoin and others forms of cryptocurrency open up new spectrum through this form of digital money to provide business transactions that are relatively secure, trusted, reliable. The nature by which cryptocurrency work makes the system less prone to human errors and malicious activities. The system of cryptocurrency as a form of digital money is based upon the idea of blockchain, which can be understood as a ledger that is publicly available and can be monitored simultaneously by hundreds of computer users. The information contained within these blockchains can be verified among the public, thus potentially bypassing a centralized system operated by banks and other financial institutions. The Bitcoin, the Ethereum, and other similar cryptocurrencies presented a relatively volatile fluctuation in its face value. Over time, the market favored a rise in the values of these digital currencies. Although some financial and academic leaders have raised concerns to the reliability of the cryptocurrency system, further research is required to better understand and improve any shortcomings of the cryptocurrency and blockchain system.

1). Data Set: You may refer to http://www.calstatela.edu/centers/hipic/related-site (Links to an external site.)Links to an external site.

External Link : https://www.kaggle.com/jessevent/all-crypto-currencies

2). Data Set Size: over 50MB: One file size or You can join more than one data sets to make the size greater than 50MB

3). What you are going to do with this data set.

We are going to do some analysis and visualization of trending bitcoin currency with the help of elasticsearch and kibana and will apply some deep learning method(If Possible) using keras library and build a neural network by doing some python programming to predict the future price of a particular currency which is bitcoin.

Visualization :

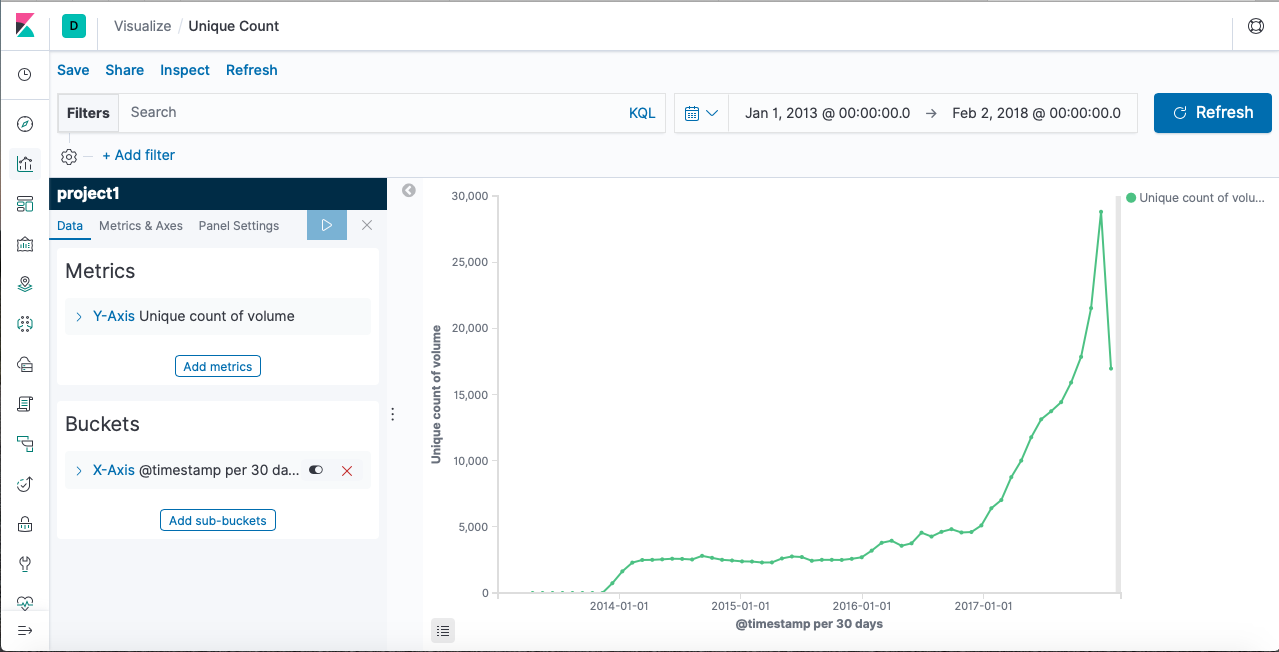
1) Using line graph we are going to display all the price chart on yearly basis.

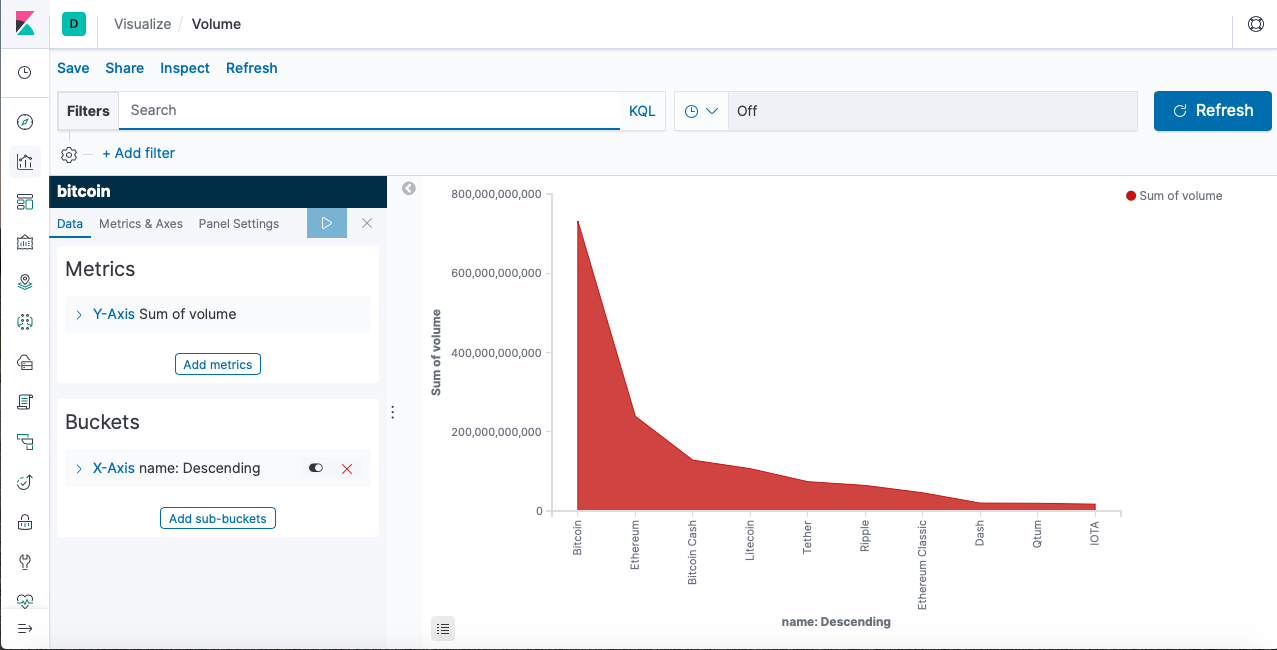
2) Using pie chart we are going to distribute the volume like which year has the highest volume.

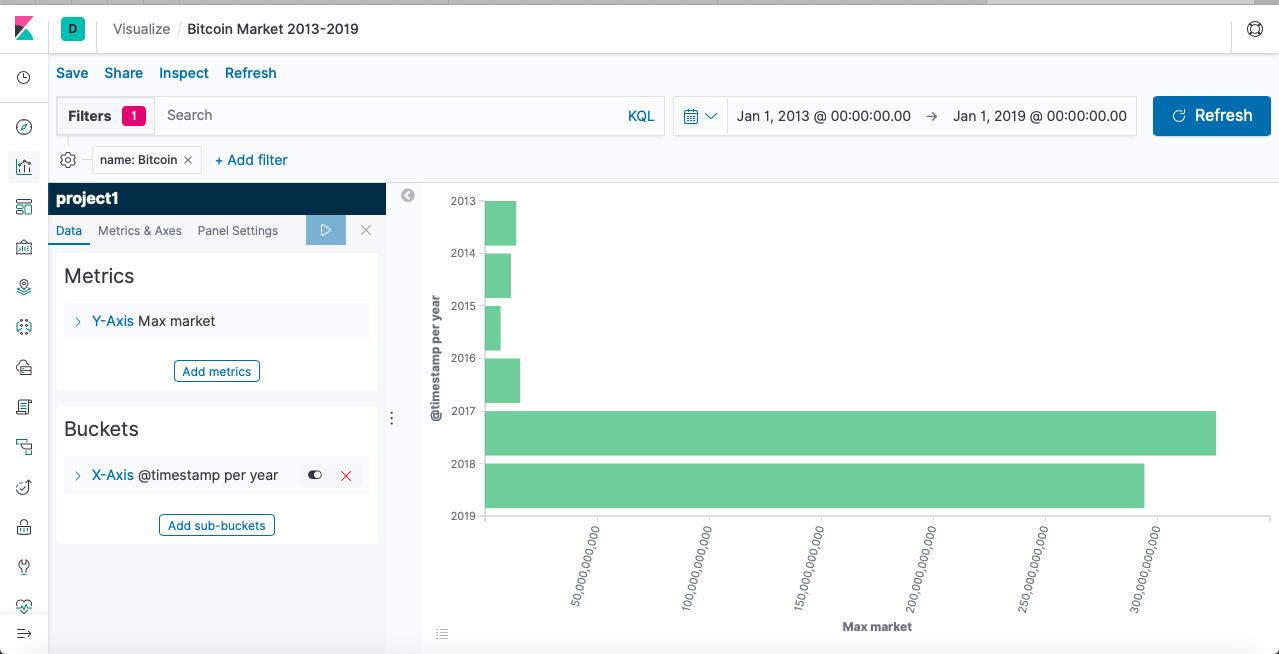
3) Comparison between the opening and the closing price.

Term project should be similar to what you have done at the lab: Collecting Data, Analyze, and Visualize using ElasticSearch (ES) - Optionally Deep Learning (DL) and Prediction using ES or DL - and present the insights you find out.

NOTE: DL and prediction is optional as this is only 10 weeks class. But, if you add it to your presentation and term paper, you will get an extra credit (20% extra of the term project).







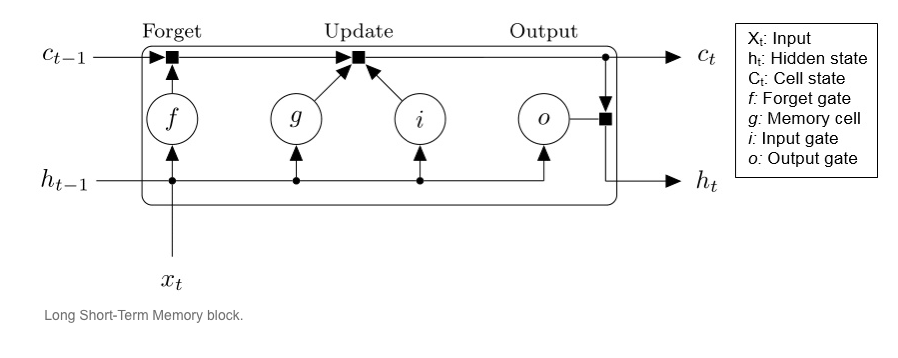
Applying Deep learning to predict one particular currency (“Bitcoin”) :

-> From the dataset we extract the bitcoin data and applied deep learning with the help of python programming.

-> Libraries : Numpy, Pandas , Sklearn , Keras

-> Long Term Short Memory Network : Long short-term memory (LSTM) is an artificial recurrent neural network (RNN) architecture used in the field of deep learning. A common LSTM unit is composed of a cell, an input gate, an output gate and a forget gate. The cell remembers values over arbitrary time intervals and the three *gates* regulate the flow of information into and out of the cell.LSTM networks are well-suited to classifying, processing and making predictions based on time series data, since there can be lags of unknown duration between important events in a time series. LSTMs were developed to deal with the exploding and vanishing gradient problems that can be encountered when training traditional RNNs.

Architecture of LSTM Network:



Output :



------------------------------------------------------------------------------------------------------------------------

------------------------------------------------------------------------------------------------------------------------

------------------------------------------------------------------------------------------------------------------------

**Team Presentation Rubric: 100%**

1. Presentation Slide Format: 30%

a. Font Size (at least 28 font size): 9%

b. Data Size (should be at least 2GB; or need permission to get full credit): 7%

c. Data Source URLs: 6%

d. H/W experimental Specifications (Your server’s OS, memory size, CPU speed): 8%

2. Originality: 20%

a. How unique your idea (10%)

b. your deliverable (10%)

§ Workflow or architecture charts, you may see the example workflow or architecture (2% of 10%)

3. Relevance with the topic in the class: 35%

a. Completeness Visualization in location and time info (Geo-Spatial Visualization) if your dataset has them: 16%

b. Implementation (Flow Chart that shows the architecture of your implementation, for example,<https://goo.gl/3AnnPN>,<https://www.rroij.com/articles-images/IJIRCCE-264-g002.GIF>, ): 9%

c. Github link that has all the codes and documents: 10%

4. Communicate with the instructor about the topic to get approval about the topic: 15%

**NOTE:**

1. **Peer Evaluation (Optional)**: You have to email to the instructor the peer evaluation about your team members for the term project and the presentation. It should be composed of:

Section #, Group name, Your Name, Team Member Name, Team Members’ Scores out of 100%, The reason you evaluate the member(s) with the score(s). If you don’t email me peer evaluation, I assume, all of you contribute the work fairly well. **For example,** your team score is 95% and your peer evaluation by your team members are 100%, your score is 95 (= 95 x 100%)

2. **Plagiarism**: If you make a copy of others, it should violate the academic integrity so that you should get 0 in the term project or F in the course in the worst case.

**Term Paper Rubric: 100%**

It should be almost same as the team presentation. But, mostly, I will take a look at if you revise the content per my comment at the presentation. Thus, any penalty at the presentation can be recovered. You have to follow the format of the paper template that the instructor shares.

The term paper should have a section of Related Work after the section of Introduction. It should include 2 – 3 existing works and you need to describe what the existing works are and the difference between yours and the existing works in 2 or 3 paragraphs.

You also need to email the instructor the **peer evaluation** for the term paper. If you don’t email me peer evaluation, I assume, all of you contribute the work fairly well.

**For example,** your team score is 95% and your peer evaluation by your team members are 100%, your score is 95 (= 95 x 100%)

1). You can use your term paper template “termPaperForm11212016.doc” that I put to dropbox or box.com.

2). Term paper should be 3 - 4 pages. If less than or more than it, you will not get the good grade.

3) Term paper requires to follow the structured in order to get the good score as below:

**NOTE**: I prefer the diagram and table that are drawn not the screenshot

**Abstract:** one or two paragraphs summary of your work

**Introduction:** why you choose this topic and why your work/topic should be important. And what is the background of your work

**Related work:** Describe 1 paragraph of 2 or 3 papers that adopt the similar works as you have done; Then, present the difference between yours and theirs.

**Background/existing work:** Detailed background and existing work that your work is based on

**Your work:** Illustrate your work; you’d better present what insights you look for and how do you find out. You have to present a diagram of your work which is similar to ElasticSearch’s query commands and visualization (or Deep Learning model NN, epoch, paramters and its accuracy ).

**Conclusion:** summarize what you have did, why your work is interesting or important, what you have built, how accurate your prediction is and what you learned from the work

**References:** Papers, articles, URLs (your github, data source,…) that you referred to.

------

|  |  |
| --- | --- |
| **Instructions for CIS graduate class Term Paper Manuscript** | |
| **Abstract:** This document is an example of what your term paper manuscript of CIS510 should look like. Authors are asked to conform to the directions repo­­­rted in this document.  **1. Introduction**  This document is a version of the instructions for preparing copies for the final term paper of CIS510. The format here described allows for a graceful transition to the style required for that publication.  This document is an example of what your term paper of CIS510 should look like. Authors are asked to conform to the directions reported in this document.    **2. Related Work**  The maximum length of a manu­script is 4 pages, printed single-sided. Print all text, including section titles and figures, in two-column format where each column is 8.5 cm by 24.5 cm (3.35 in by 9.65 in) and there is a 0.6 cm (0.24 in) space between the two columns. Excep­tions to the two-column format include the title at the top of the first page and any full-width figures or tables. Start all pages directly under the top margin. Text should be centered on each page.    **3. General Instructions**  On A4 paper, this roughly means leaving 1.7 cm (0.67 in) margins on left and right sides of each page as well as a 2.5 cm (1 in) margin on the top and bottom of each page. Type single-spaced. Indent when starting a new paragraph. Use standard fonts such as Times New Roman or Computer Modern Roman, 10 points for text, 11 points (bold) sub­section headings, 12 points (bold) for section headings, 14 points (bold) for title, 11 points for authors’ names, and 10 points for their affiliations.    **3.1 The First Page**  Center the title across both columns. Use the two-column format only when you begin the abstract.  **Title:** Place the title at the top of the first page, followed by the authors’ names and their affiliations. Long title should be typed on two lines without a blank line intervening. Leave approximately 1 cm (0.39 in) between the title and the body of the first page.  **Abstract:** Type the abstract at the beginning of the first column. The abstract should be no longer than 200 words.  **Text:** Begin typing the main body of the text immediately after the abstract, observing the two-column format as shown in this example. | **3.2 Sections**  **Headings:** Type and label section and subsection headings in the style shown on these pages. Use numbered sections, in order to facilitate cross references.  **References:** Citations within the text appear in brackets as [ref. number]. Gather the full set of references together under the heading **References**; place the section before any **Appendices**, unless they contain references. Arrange the references in the order that they are cited in the text. Provide as com­plete a citation as possible, using a consistent format.  **Appendixes:** Appendixes, if any, directly follow the text and the references (but see above). Letter them in sequence and provide an informative title: **Appendix A Title of Appendix.**    **3.3 Footnotes**  Put footnotes at the bottom of the page. They may be numbered or referred to by asterisks or other symbols.[1] Footnotes should be separated from the text by a line.[2]    **3.4 Graphics**  **Illustrations:** Place figures, tables, and photographs in the paper near where they are first discussed, rather than at the end, if possible. Wide illustrations may run across both columns.  **Captions:** Provide a caption for every illus­tration; number each one sequentially in the form: "Figure 1. Caption of the Figure." "Table 1. Caption of the Table." Type the captions for figures below the figures. Type the captions for tables above the tables.    **4. Length of Camera-ready Manuscript**  For the length of camera-ready manuscripts, **a paper is limited up to 4 pages**. All illustrations, references, and appendices must be accommodated within these page limits. Any extra page beyond the first four pages will be deleted. **Please DO NOT put a page number in each pag**    **5. Submission Process**  1. Format your paper using this template.  2. Turn the hardcopy by Dec 4th before the lecture starts     **References** [1] T.A. Jones, “Writing a good paper,” *IEEE Trans. on General Writing*, Vol. 1, no. 2, pp.1-10, May 2002.  [2] K. Hwang, *Computer Arithmetic*, John Wiley, 1997.  [1] This is how a footnote should appear  [2] Note the line separating the footnotes from the text |

-------

**Term Project Tutorial Rubric: 100%**

Anyone can easily reproduce your work by following your tutorial as the lab tutorial that the instructor shared.

1. Materials Available (30%)

a. If Data Set can be downloadable per the direction (15%)

b. If source code is downloadable per the direction (15%)

2. Completeness (70%)

a. If each step is clear to follow (15%)

b. If the source (query or python) code is correct (15%)

c. If the source codes are executable or possible to copy/paste to execute (15%)

d. If the visualization using tools are easy to follow or clearly executable (15%)

e. If the geo-spatial visualization is clear – when the dataset has it (10%)

**NOTE**: You also need to email the instructor the **peer evaluation** for the term paper (Optional). If you don’t email me peer evaluation, I assume, all of you contribute the work fairly well.

**For example,** your team score is 95% and your peer evaluation by your team members are 100%, your score is 95 (= 95 x 100%)

---

**Lab Tutorial**

yourname (yourname@calstatela.edu)

06/10/2016

**Yelp Data Analysis using ElaticSearch and Tensorflow (your Title)**

**Objectives**

**List what your objectives are.** In this hands-on lab, you will learn how to:

* Get data manually using REST API
* Create Spark cluster
* Train NLP system
* SQL commands to perform the analysis.
* Visualization

**Platform Spec**

· IBM Bluemix BigInsights

· CPU Speed: ?

· # of CPU cores: ?

· # of nodes: ?

· Total Memory Size: ?

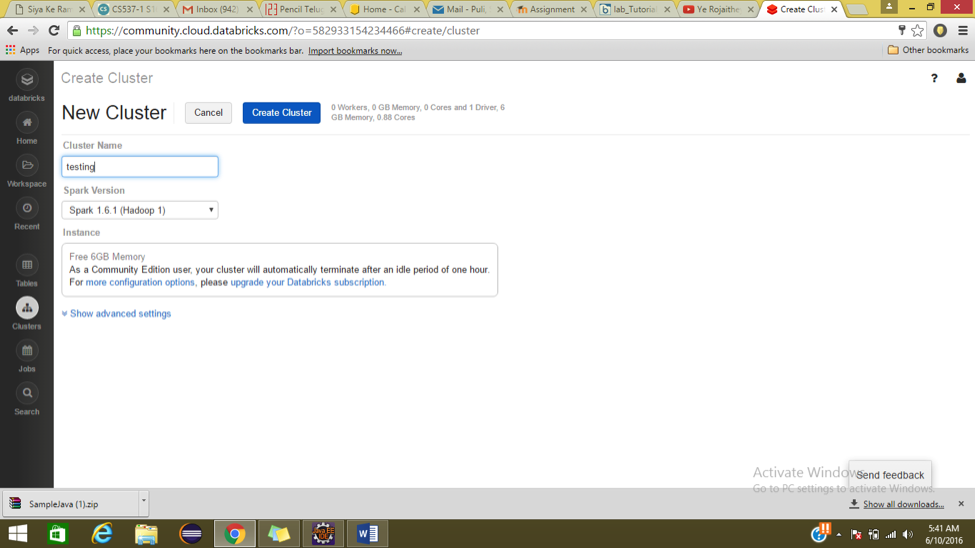
Step 1: Get data manually using REST API

**Explain what this step is for.** This step is to get data manually….

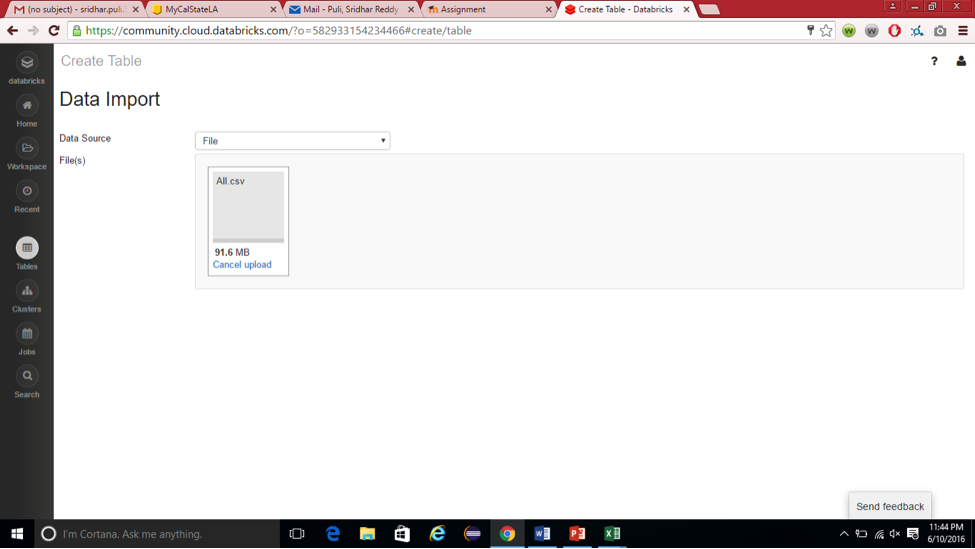
1. Create Google API keys at https://develop:

2. Sign into your databricks account.

3. Go to Clusters option on the left and click on create cluster.

4. Give the cluster name and click create cluster.

5. Under tables section click on create table and select the file to upload.



Step 2: Train NLP

**Explain what this step is for.** This step is to …

**Code should be in the following format and indent:**

import org.apache.spark.ml.feature.RegexTokenizer

val tokenizer = new RegexTokenizer()

.setPattern("\\p{L}+").setMinTokenLength(3)

.setGaps(false)

.setInputCol("text")

.setOutputCol("words")

val tokenized\_df=tokenizer.transform(splits(0))

vi) Use the below code to remove stop words

Run them in separate cells for better understanding

%sh wget http://ir.dcs.gla.ac.uk/resources/linguistic\_utils/stop\_words -O /tmp/stopwords

%fs cp file:/tmp/stopwords dbfs:/tmp/stopwords

val stopwords = sc.textFile("/tmp/stopwords").collect()

import org.apache.spark.ml.feature.StopWordsRemover

// Set params for StopWordsRemover

val remover = new StopWordsRemover()

.setStopWords(stopwords) // This parameter is optional

.setInputCol("words")

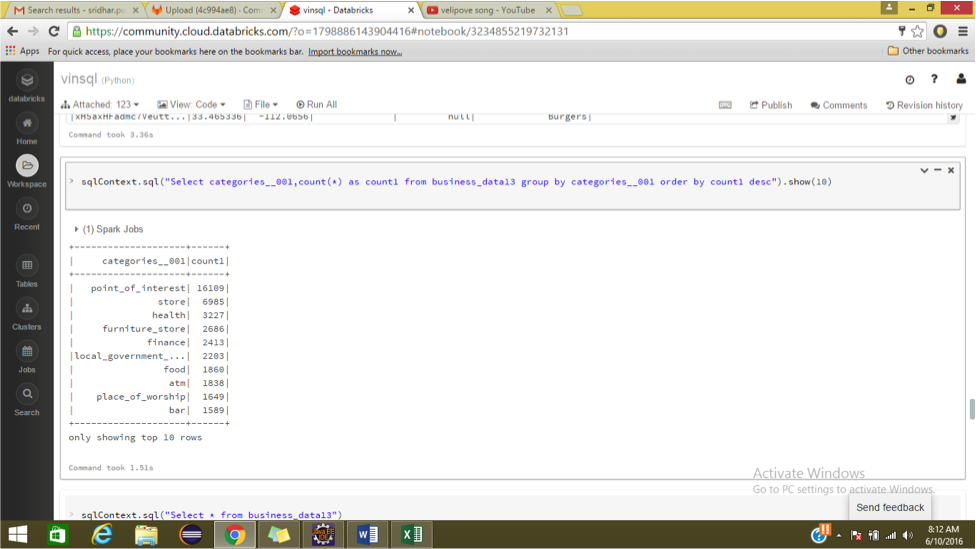
.setOutputCol("filtered")

// Create new DF with Stopwords removed

val filtered\_df = remover.transform(tokenized\_df)

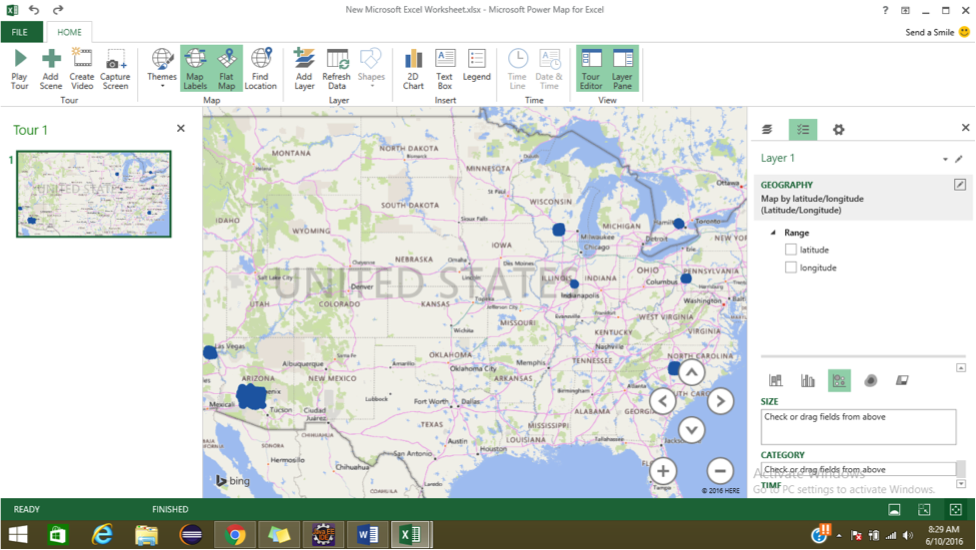
1. To show top ten categories

sqlContext.sql("Select categories\_\_001,count(\*) as count1 from business\_data13 group by categories\_\_001 order by count1 desc").show(10)



Step 3: Visualization

**Explain what this step is for.** This step is to…



1. To visualize location type of results on map, convert csv file to excel and click on map button under insert tab.

References

1. URL of Data Source,<http://www.calstatela.edu>

2. URL of your Github

3. URL of References

4.

---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------