Project Report for BDAD Project

**Social media analytics using knowledge graph**

Submitted By:

In fulfilment for the **Big Data Application Development** (Semester - 7)

Of

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE AND ENGINERRING (BDA)

at

[](https://www.google.com/url?sa=i&rct=j&q=&esrc=s&source=images&cd=&ved=2ahUKEwjNlteg5_XhAhWGfH0KHWOXBNcQjRx6BAgBEAU&url=https://www.ict.gnu.ac.in/&psig=AOvVaw2_w3VyDOiqJFS36G1Ege7D&ust=1556644622566621)

2020-2021

Table of Contents

[Abstract 2](#Abstract)

[Introduction 4](#Introduction)

[Project Scope 5](#_Toc54717188)

[Software and Hardware Requirement 6](#Software_and_Hardware_requirements)

[Minimum Hardware Requirement: 6](#Minimum_hr)

[Minimum Software Requirement: 6](#Minimum_sr)

[Project Plan 7](#Project_Plan)

[System Design 8](#System_Design)

Designing Graph Database & User Interface 9

[Implementation Details 11](#Implementation_Details)

[Flowchart: 11](#Flowchart)

[Actual Program Code Snapshots: 11](#Actual_Code)

[Testing 14](#Testing)

[User Manual 15](#User_Manual)

[For Colab Notebook 15](#Colab_Notebook)

[For Neo4j Desktop 15](#Neo4j)

[Conclusion and Future Work 18](#Conclusion_and_Futurework)

[References: 19](#References)

**Abstract**

The aim of this project is to provide a complete solution for understanding customer feedbacks through twitter data and knowledge graph to decide various strategies and changes a brand/ company/ organization can make to their product/service for better customer satisfaction.

**Introduction**

This project concerns about scraping data from twitter and visualising it so as to form speculations. The amount of data scrapped is in conjunction with reflections customers and user convey through the Tweets.

The analysis done can give insights such as pin pointing the direct relationships between twitter users, content of tweet and information regarding that. Based on Tweets, insights drawn are easy to understand.

The certain important keywords which convey and imparts meaning to the tweets can be sorted out with this project. Such as location the users are mentioning in their tweets regarding certain issue or product.

**Project Scope**

With this project, user can search up twitter regarding certain keywords and sort the information fetched based on desired input keyword or in a sense re-filter the information. With such good quality data, one can easily draw conclusions and relationships based on the visualizations present with.

**Software and Hardware requirements**

**Minimum Hardware requirements:**

* OS: Windows, Linux, Mac
* Minimum 2 GB Ram
* 2 GHz Processor
* Minimum of 5 GB Storage

**Minimum Software requirements:**

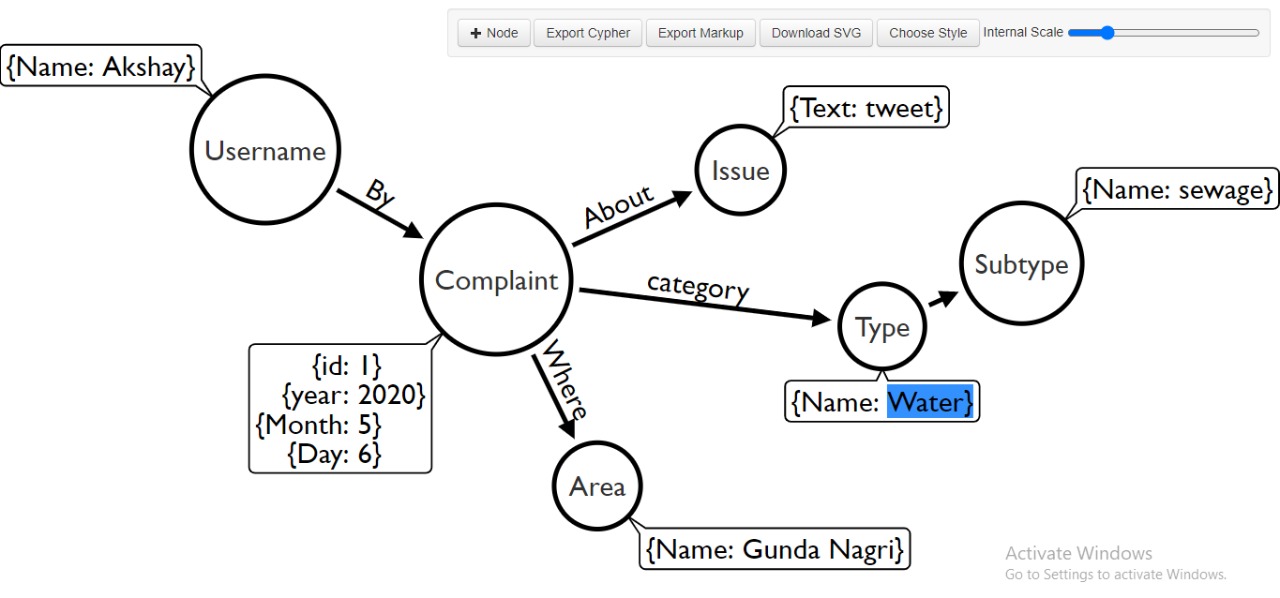
* Platform: Colab Notebook
* For Tweet Extraction: Python, Twint, PySpark, Optimus
* For Knowledge Graphs: Neo4j

**Project Plan**

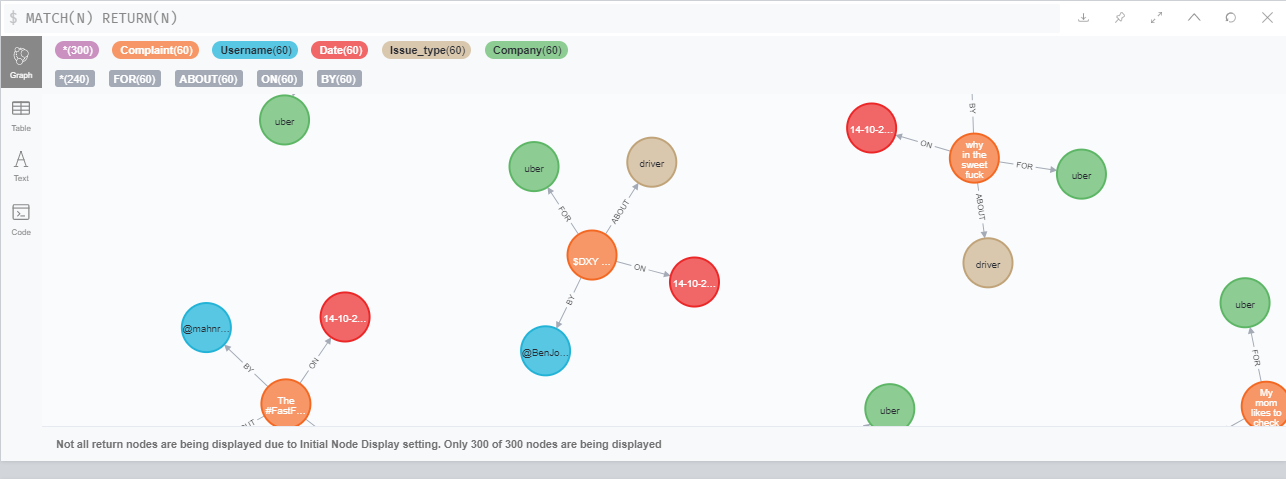
* Twint Setup and Understanding: 1 Week
* Cleaning and fetching twitter data: 2 Weeks
* Neo4j Setup and Understanding: 2 Weeks
* Designing graph database: 1 Week
* Filtering, Quering and Generating Final Output: 2 Weeks

**System Design**

**∙** **Designing Graph Database**

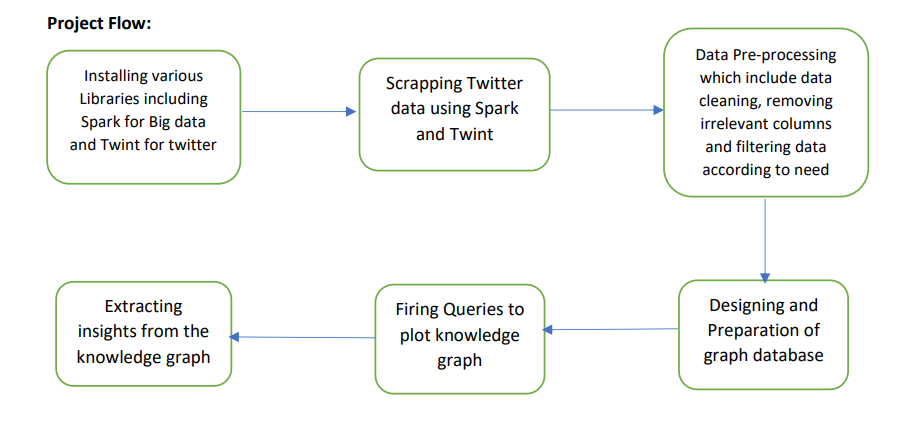


**∙** **User interface design**

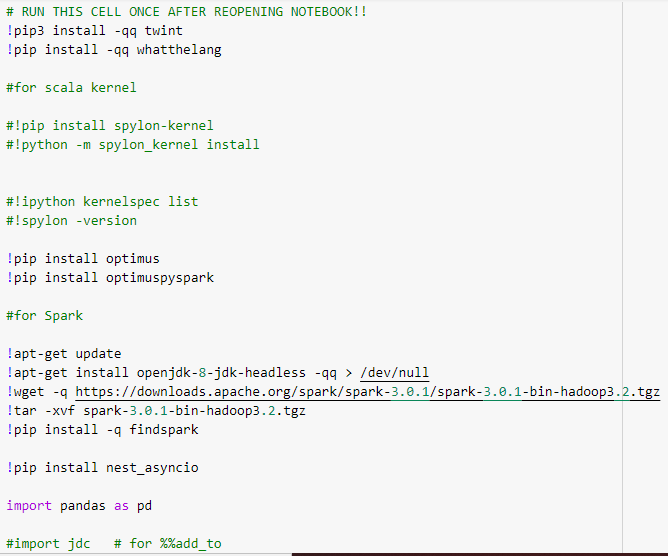


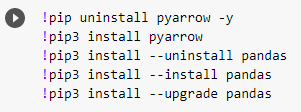
**Implementation Details**

**∙** **Flowchart of Implementation**

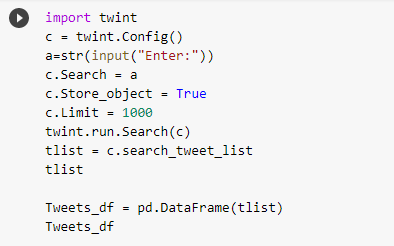


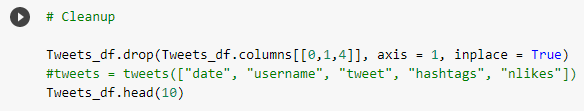
**∙** **Actual Program Code snapshots**

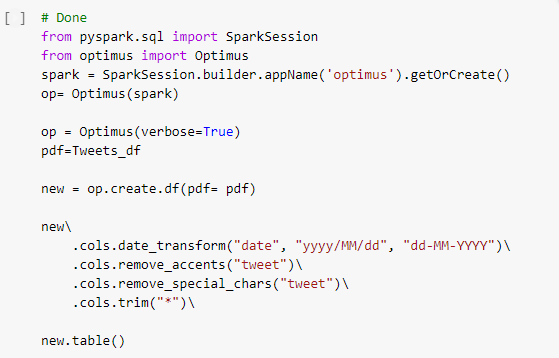


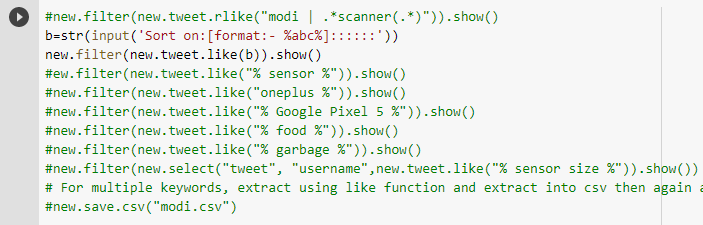


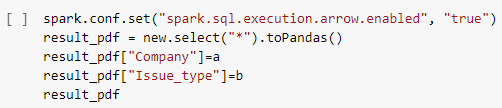






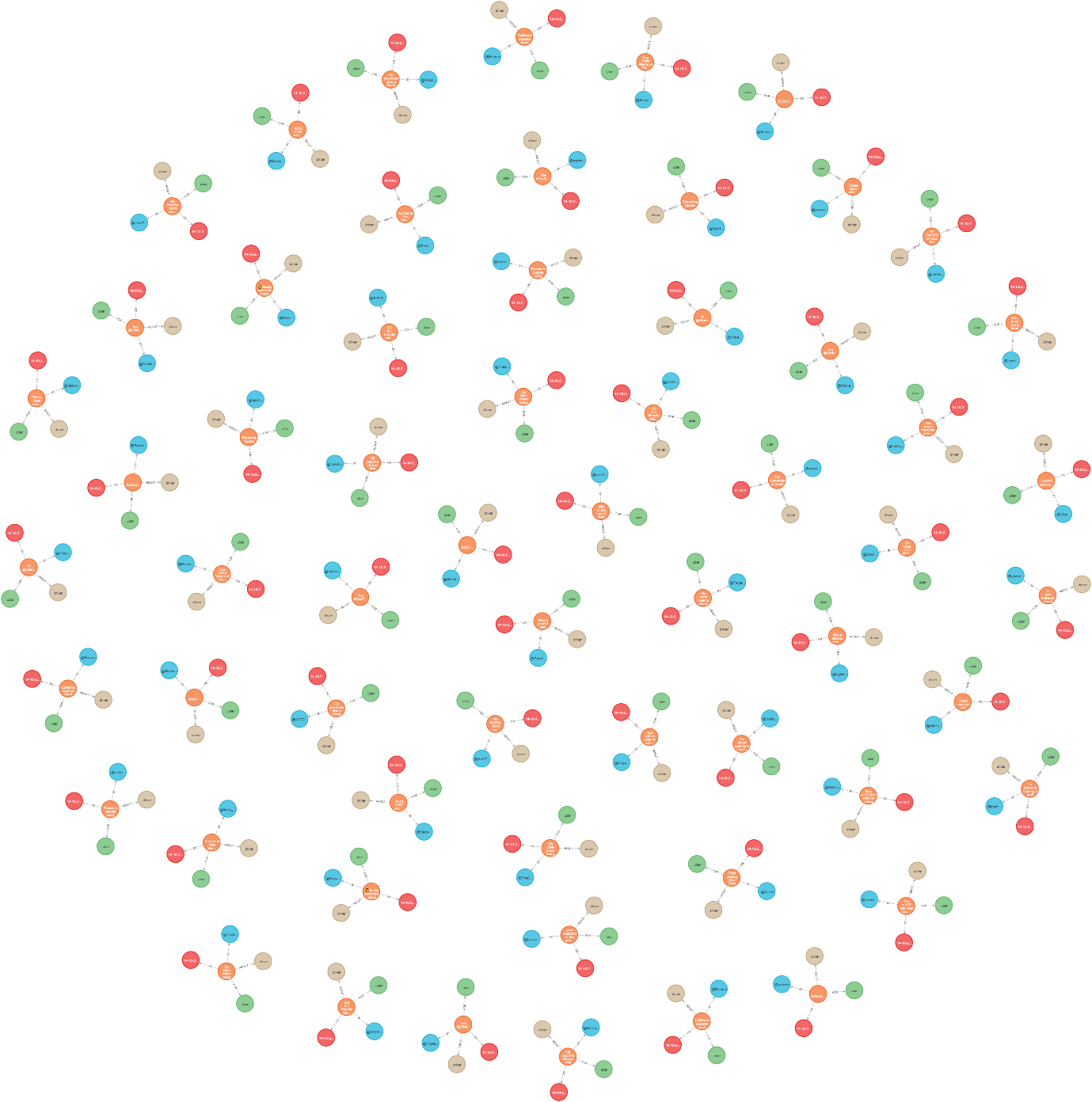








**Testing**

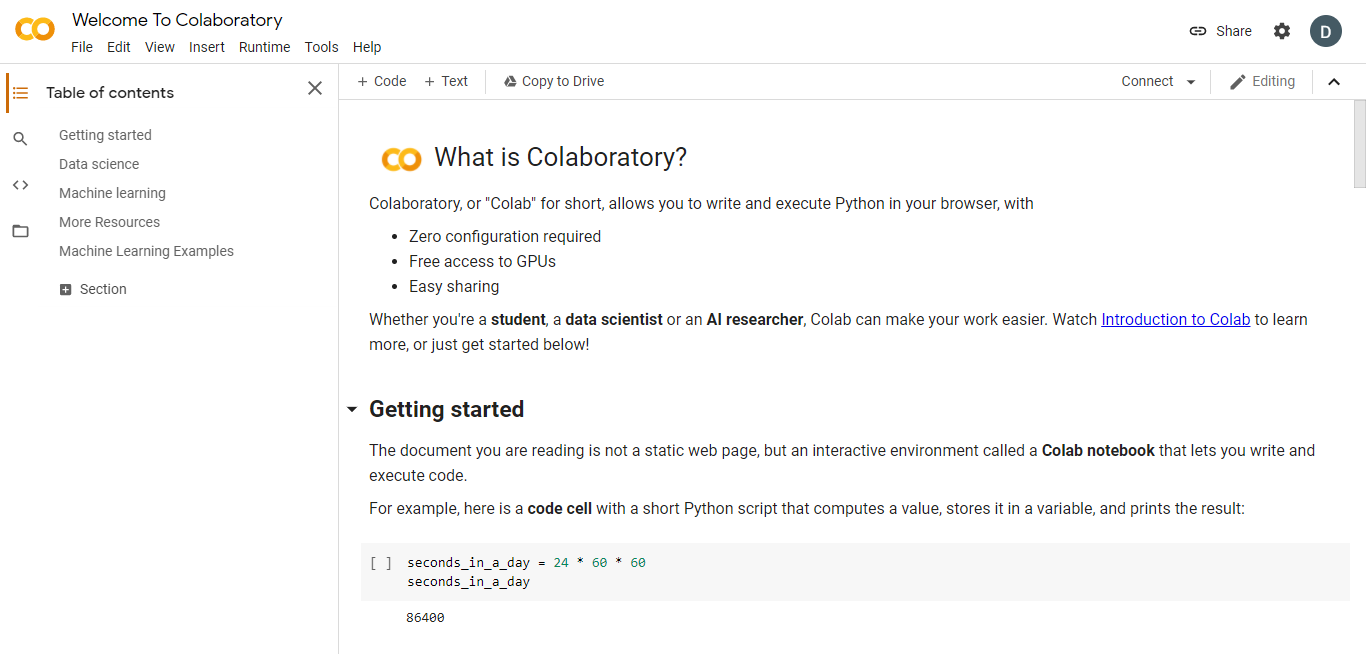


**User manual**

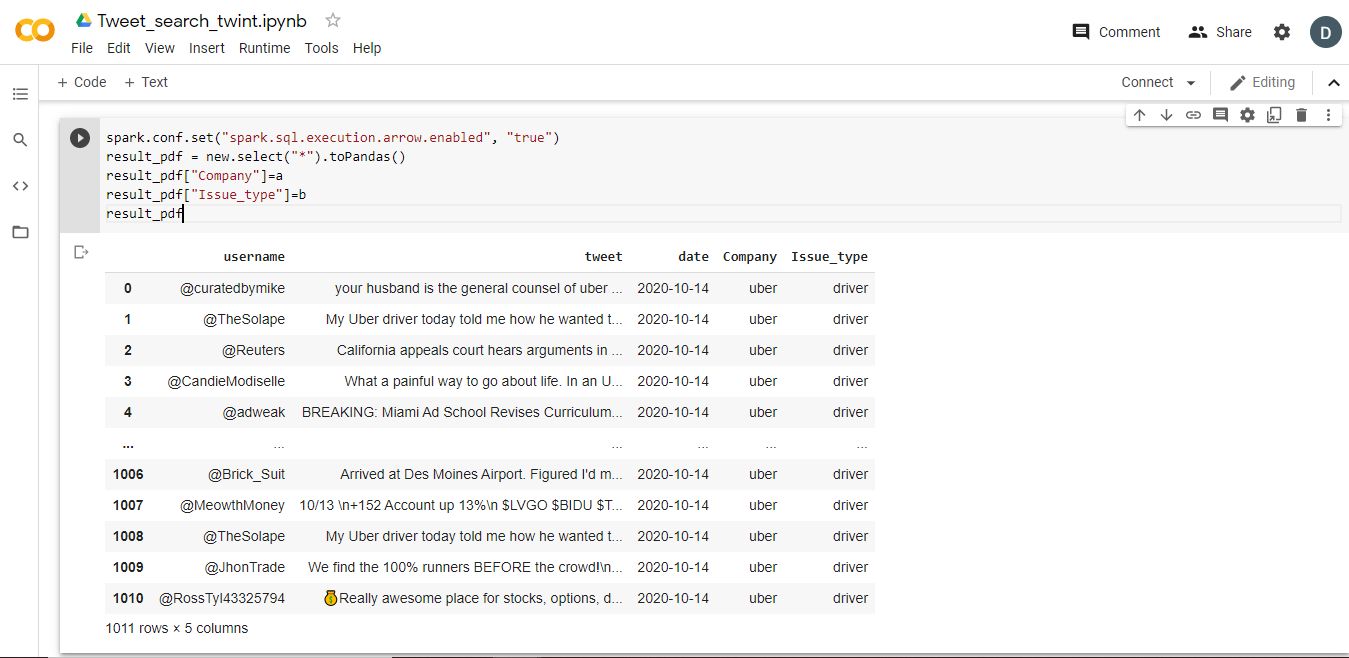
**For Colab Notebook**

**Step 1:** Create new Notebook in the Google Research Collaboratory

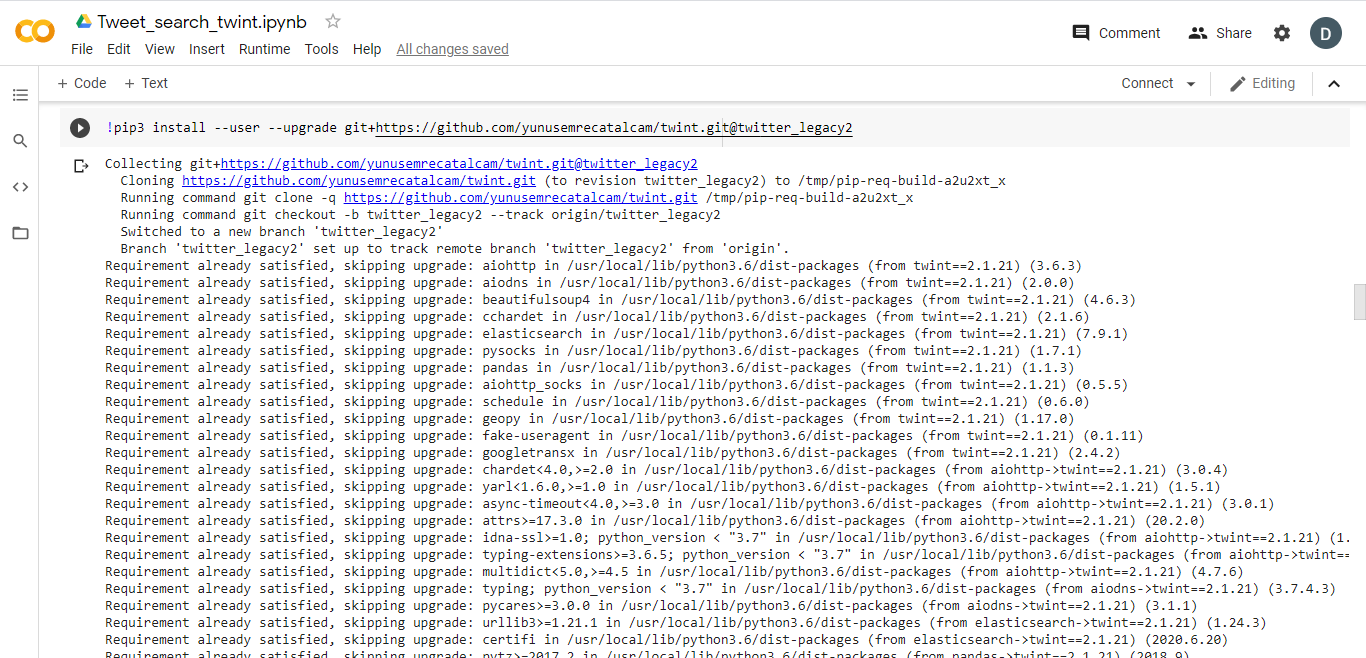
Open https://colab.research.google.com/ , create a new notebook and switch Runtime to GPU.

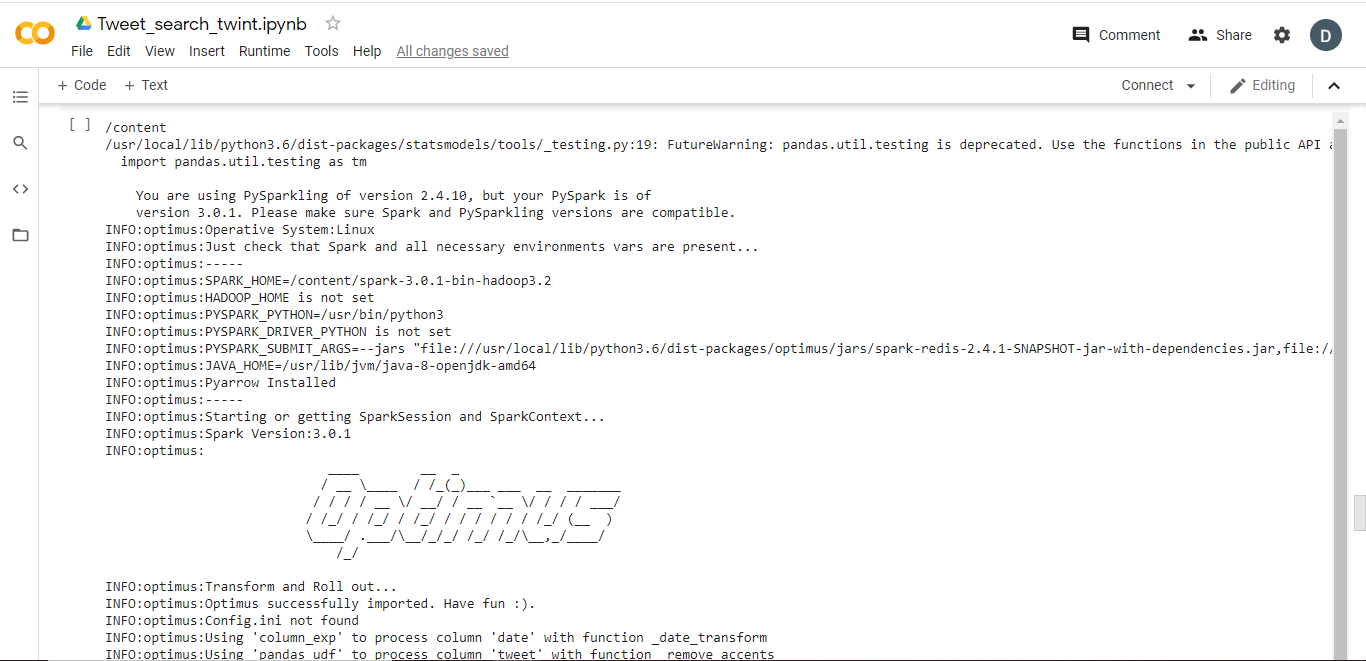


**Step 2:** Create a new Code Cell, with the following code



**Step 3:** Use Python environment normally and install necessary libraries like Spark & Optimus.



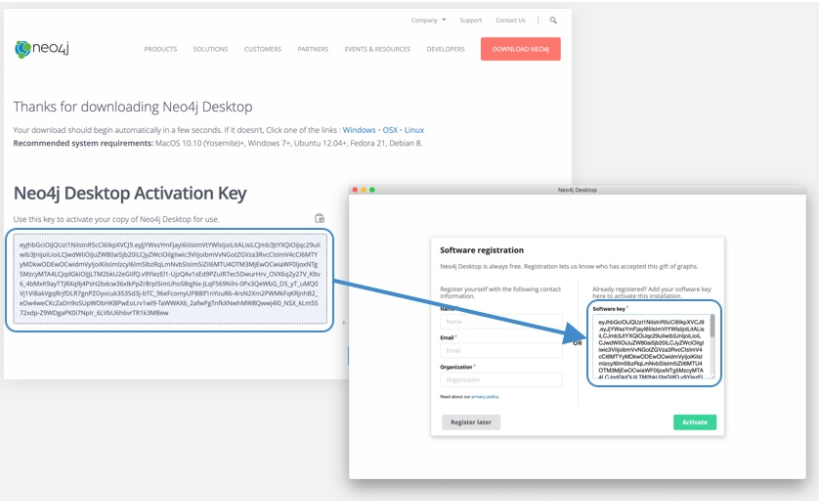


**For Neo4j Desktop**

**Step 1:** Go to Neo4j website <https://neo4j.com/> to download Neo4j desktop

**Step 2:** Opening Neo4j Desktop for the First Time

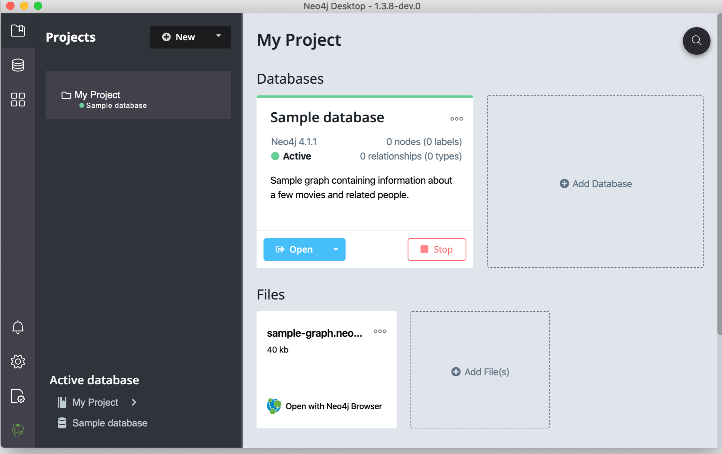
If you are opening Neo4j Desktop for the first time, it should ask you to register the software with an activation key. This activation key is generated when you first download Neo4j Desktop and will be displayed on the download confirmation page. Keep it somewhere safe.



Copy this activation key into the Software registration form and click Activate to continue.

**Step 3:** Clicking the Browser Guide link for the Sample database will open up the Neo4j Browser with a Browser Guide that will guide you through your first queries on this Neo4j database.

The sample database is created using a randomly generated password, which is stored within Neo4j Desktop and used by Neo4j Browser to authenticate against the database. You can reset this password in the Administration tab on the Manage screen.



**Conclusion and Future work**

**Conclusion:**

Thus, we were successful in implementing knowledge graph using Neo4j on Twitter data to get insights of what are the relationships between various columns generated from python code.

**Future Work:**

Implementing GUI through Django framework.

**References**

**List of references**

1. <https://pypi.org/project/twint/>
2. <https://neo4j.com/>
3. <https://twitter.com/>