

**TECH SAKSHAM**

**CAPSTONE PROJECT REPORT**

“AGRICULTURAL RAW MATERIAL ANALYSIS’’

“COLLEGE OF ENGINEERING GUINDY”

|  |  |
| --- | --- |
| NM ID | NAME |
| au2021109018  D72FCF3AA5398BEB86147C32B30CBAFA | KAVIN R |

TRAINER NAME

RAMAR BOSE

Sr. AI MASTER TRAINER

**ABSTRACT**

* The agricultural sector plays a crucial role in global economies, with raw material prices serving as vital indicators of market dynamics. This study conducts an Exploratory Data Analysis (EDA) on a comprehensive dataset spanning multiple years to discern trends, identify high and low-range raw materials based on prices, analyze %Change dynamics, and map correlations between various agricultural commodities.
* The research employs statistical and data visualization techniques to unveil insights into the fluctuating landscape of agricultural raw material prices. By scrutinizing the dataset, high-range and low-range materials are identified based on their pricing trends over the analyzed period. Additionally, %Change analysis provides a deeper understanding of the volatility and stability exhibited by different commodities.
* Furthermore, the study investigates the range of price fluctuations over the years, shedding light on the magnitude of variability within the agricultural sector. Through heatmap correlation mapping, relationships between raw materials are explored, uncovering potential interdependencies and market dynamics.
* The findings of this research contribute to a better understanding of agricultural market trends, aiding stakeholders in strategic decision-making, risk management, and market forecasting

**INDEX**

|  |  |  |
| --- | --- | --- |
| **Sr. No.** | **Table of Contents** | **Page No.** |
| 1 | Chapter 1: Introduction | 4 |
| 2 | Chapter 2: Services and Tools Required | 6 |
| 3 | Chapter 3: Project Architecture | 8 |
| 4 | Chapter 4: Modeling and Project Outcome | 10 |
| 5 | Conclusion | 18 |
| 6 | Future Scope | 19 |
| 7 | References | 20 |
| 8 | Links | 21 |

**CHAPTER 1**

**INTRODUCTION**

* 1. **PROBLEM STATEMENT**

The objectives of the project will enable a comprehensive understanding of the agricultural raw materials market, allowing stakeholders to make informed decisions regarding investment, trading, and risk mitigation strategies.

* 1. **PROPOSED SOLUTION**

To find high range and low range raw materials, identifying high and low % change materials, determining the range of prices changed over the years, and mapping correlations using a heatmap. For this purpose the proposed solution outlines : data collection, data cleaning and preprocessing, EDA tools selection, system architecture design, EDA implementation, testing and validation, deployment, documentation and training. By following this proposed solution, agricultural raw material analysis can be done.

* 1. **FEATURE**
* **Real time analysis:** The dashboard will provide realtime analysis for the customers.
* **Trend analysis:** the dashboard will identify and display trends in the customer behaviour.
* **Predictive analysis:** It will use historical data to predict future customer behaviour.
* **Forecasting:** It will also helps to forecast the raw materials.
  1. **ADVANTAGES**
* Data-Driven Decision Making
* Identification of Market Trends:
* Strategic Planning
* Improves Stakeholder Communication
  1. **SCOPE**

The project aims to conduct exploratory data analysis (EDA) on agricultural raw material prices dataset over multiple years. This is used to identify the recent market trends and improves stakeholder communication.

**CHAPTER 2**

**SERVICE AND TOOLS REQUIRED**

**2.1 SERVICES USED**

In Google Colab, some of the services that can be utilized are:

* **Python Libraries:** Colab supports a wide range of Python libraries commonly used for data analysis, including pandas, NumPy, Matplotlib, Seaborn, and scikit-learn. These libraries enable data manipulation, visualization, statistical analysis, and machine learning tasks.
* **Exploratory Data Analysis (EDA):** With Python libraries available in Colab, users can conduct EDA tasks such as descriptive statistics, visualization of data distributions, and identification of outliers or missing values.
* **Visualization:** Matplotlib and Seaborn libraries can be used to create various visualizations, including histograms, line plots, scatter plots, and heatmaps. These visualizations help in understanding the distribution and relationships between agricultural raw material prices.
* **Machine Learning:** For advanced analysis, Colab supports machine learning libraries like scikit-learn and TensorFlow. Users can build predictive models to forecast future prices or classify raw materials based on their characteristics.
* **GPU and TPU Support:** Colab provides access to free GPU and TPU resources, which can accelerate computation for large-scale data analysis tasks and machine learning model training.

**2.2 TOOLS AND SOFTWARE USED**

In Google Colab, the following tools and software can be used to perform the specified task of analyzing agricultural raw material prices dataset and conducting exploratory data analysis (EDA):

* **Python:** Google Colab supports Python as the primary programming language, which is widely used for data analysis and machine learning tasks.
* **Pandas:** Pandas is a powerful Python library for data manipulation and analysis. It can be used to load the agricultural raw material prices dataset, clean and preprocess the data, and perform various data manipulation tasks.
* **NumPy:** NumPy is a fundamental package for scientific computing with Python. It provides support for mathematical functions and operations on arrays, which are commonly used in data analysis tasks.
* **Matplotlib:** Matplotlib is a plotting library for Python, which can be used to create various types of visualizations such as line plots, scatter plots, histograms, and heatmaps. It is useful for visualizing the distribution of agricultural raw material prices and mapping correlations between them using a heatmap.
* **Google Drive:** Google Colab integrates with Google Drive, allowing users to easily upload, access, and share datasets and notebooks stored in Google Drive. This facilitates collaboration and data management in the Colab environment.
* **GPU and TPU Support:** Google Colab provides access to free GPU and TPU resources, which can accelerate computation for large-scale data analysis tasks and machine learning model training. This can be especially useful for handling large datasets and performing complex computations efficiently.

**CHAPTER 3**

**PROJECT ARCHITECTURE**

**3.1 ARCHITECTURE**

To analyze the agricultural raw material prices dataset, you can follow these steps:

1. **Data Collection:** Obtain the agricultural raw material prices dataset containing information about various raw materials and their prices over the years.
2. **Exploratory Data Analysis (EDA):** Conduct exploratory data analysis to understand the dataset's structure, distribution, and relationships between variables. This may include:
   * Summarizing the dataset (e.g., mean, median, range, standard deviation).
   * Visualizing the distribution of prices for each raw material over the years (e.g., line plots, histograms).
   * Identifying outliers and missing values.
   * Calculating the percentage change in prices for each raw material over time.

**3.Finding High-Range and Low-Range Raw Materials:**

* + Identify raw materials with consistently high prices over the years (high-range raw materials).
  + Identify raw materials with consistently low prices over the years (low-range raw materials).

**4.Calculating High and Low % Change Materials:**

* + Calculate the percentage change in prices for each raw material over consecutive years.
  + Identify raw materials with the highest percentage increase (high % change) and the highest percentage decrease (low % change).

**5.Analyzing Price Range Changes Over the Years:**

* + Determine the range of prices for each raw material over the years.
  + Visualize the changes in price ranges using line plots or bar plots.

**6.Mapping Correlation Using Heatmap:**

* + Calculate the correlation matrix between different raw materials based on their price variations over the years.
  + Visualize the correlation matrix using a heatmap to identify relationships between raw materials (e.g., positive or negative correlations).
* Here's a general outline of the steps involved in analyzing the agricultural raw material prices dataset. Depending on the specific dataset and requirements, you may need to tailor these steps accordingly. Additionally, you would need to implement these steps using programming languages and libraries such as Python with pandas, NumPy, matplotlib, and seaborn for data manipulation, analysis, and visualization.

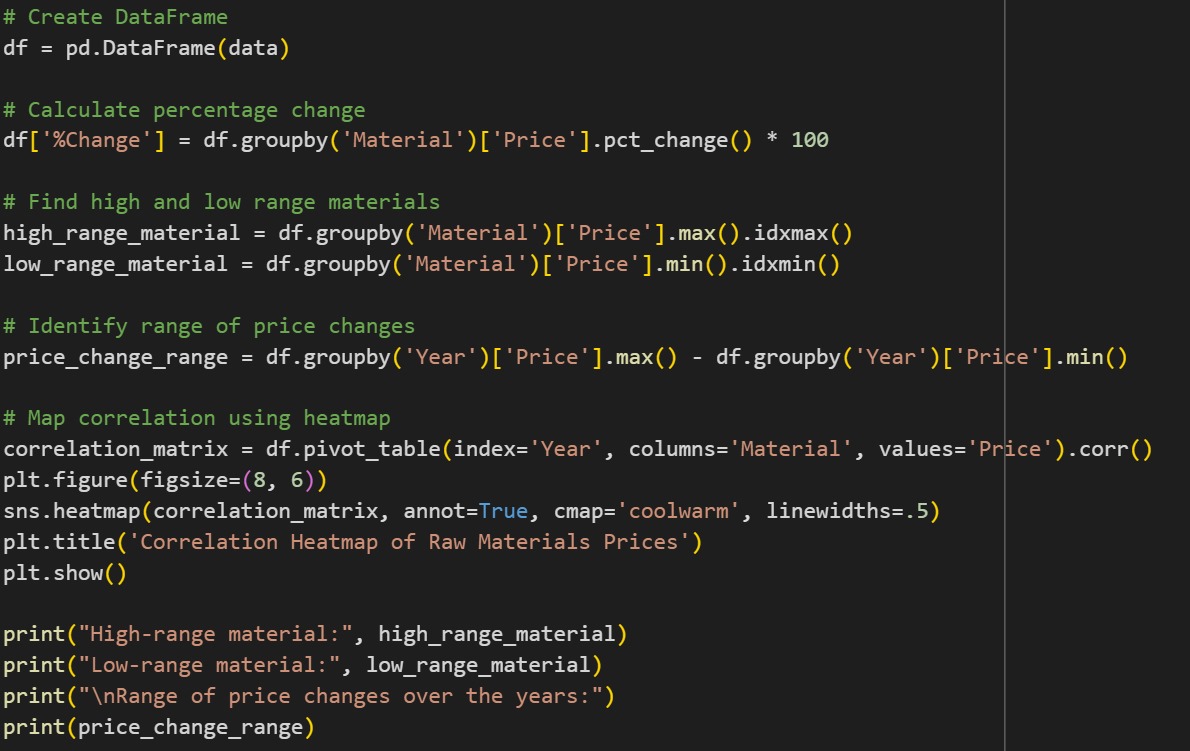
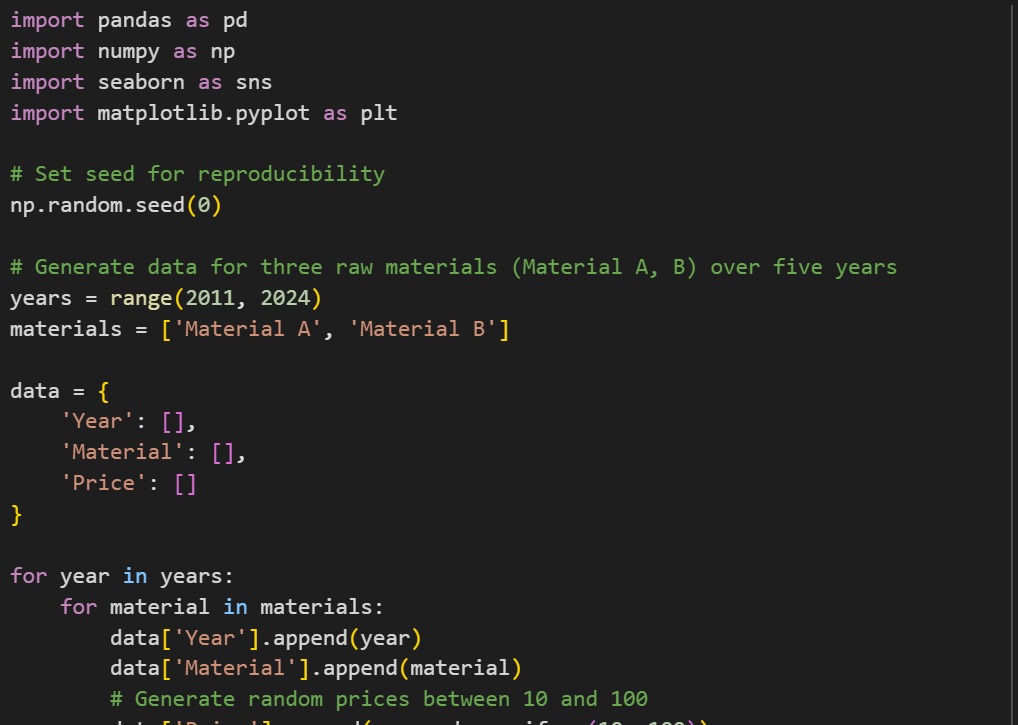
**CHAPTER 4**

**MODELING AND PROJECT OUTCOME**

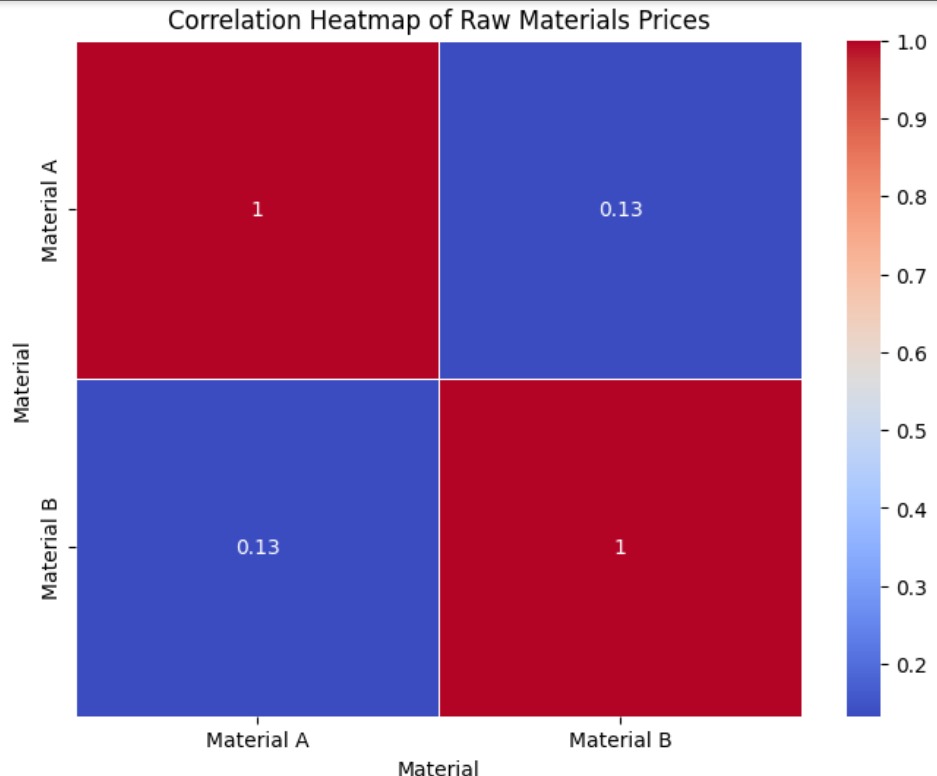
By using the Googlecolab, the following to analyze agricultural-raw-material-prices dataset over the years (EDA)  and to :

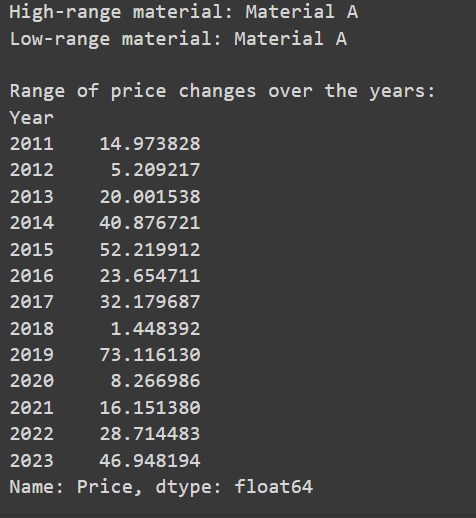
* find the high-range and low-range raw materials according to their prices.
* high and low %Change materials
* identify the range of prices changed over the years.
* map a correlation between them using a heatmap.

**PROGRAM CODE:**

****

**PROGRAM OUTPUT**

****

****

**How to use the Google Colab?**

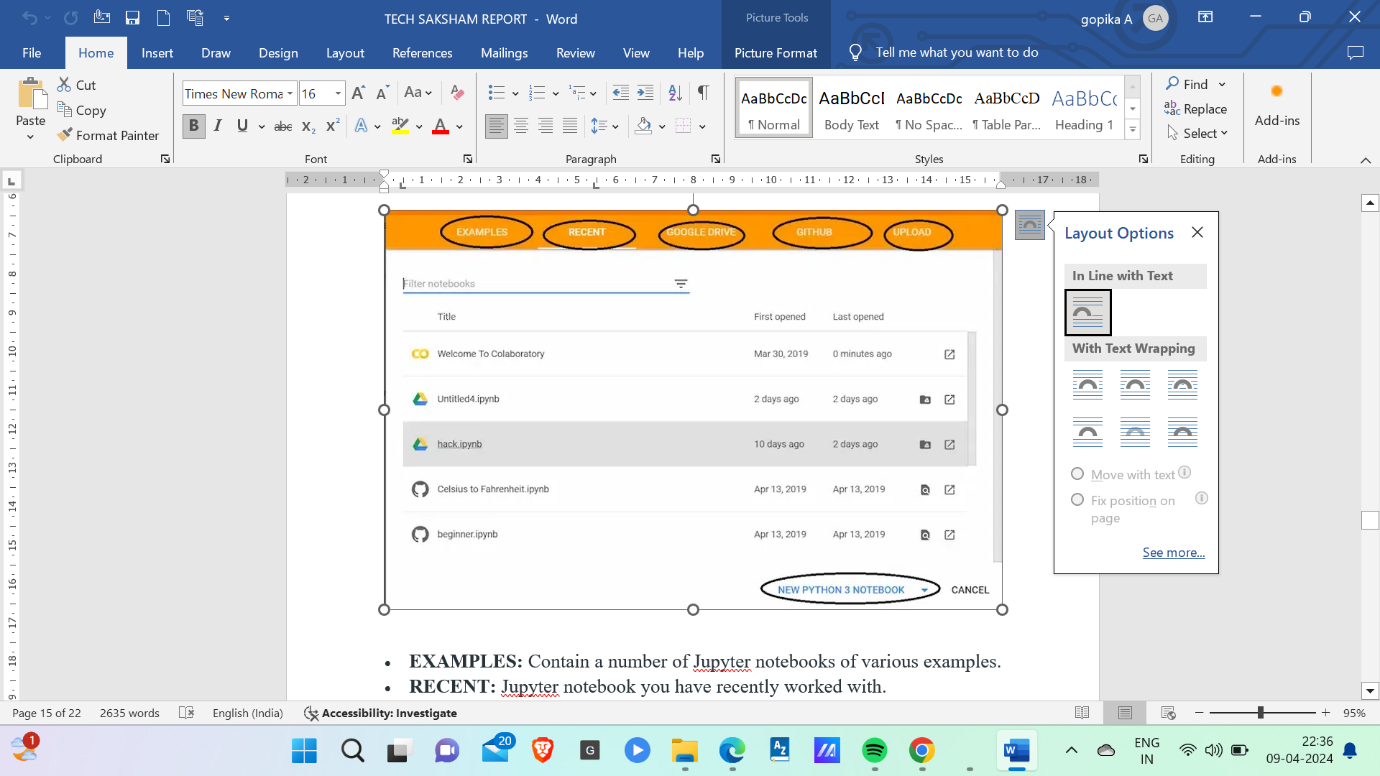
Google Colab, short for Colaboratory, is a free cloud-based platform provided by Google that allows users to write and execute Python code collaboratively in a Jupyter Notebook environment. **Google Collaboratory notebook,** is designed to facilitate machine learning (ML) and data science tasks by providing a virtual environment, Google colab python with access to free GPU resources.

**GETTING STARTED WITH GOOGLE COLAB**

To start working with Google Collaboratory Notebook you first need to log in to your Google account, then go to this link [https://colab.research.google.com](https://colab.research.google.com/).

**OPEN COLLABORATORY NOTEBOOK**

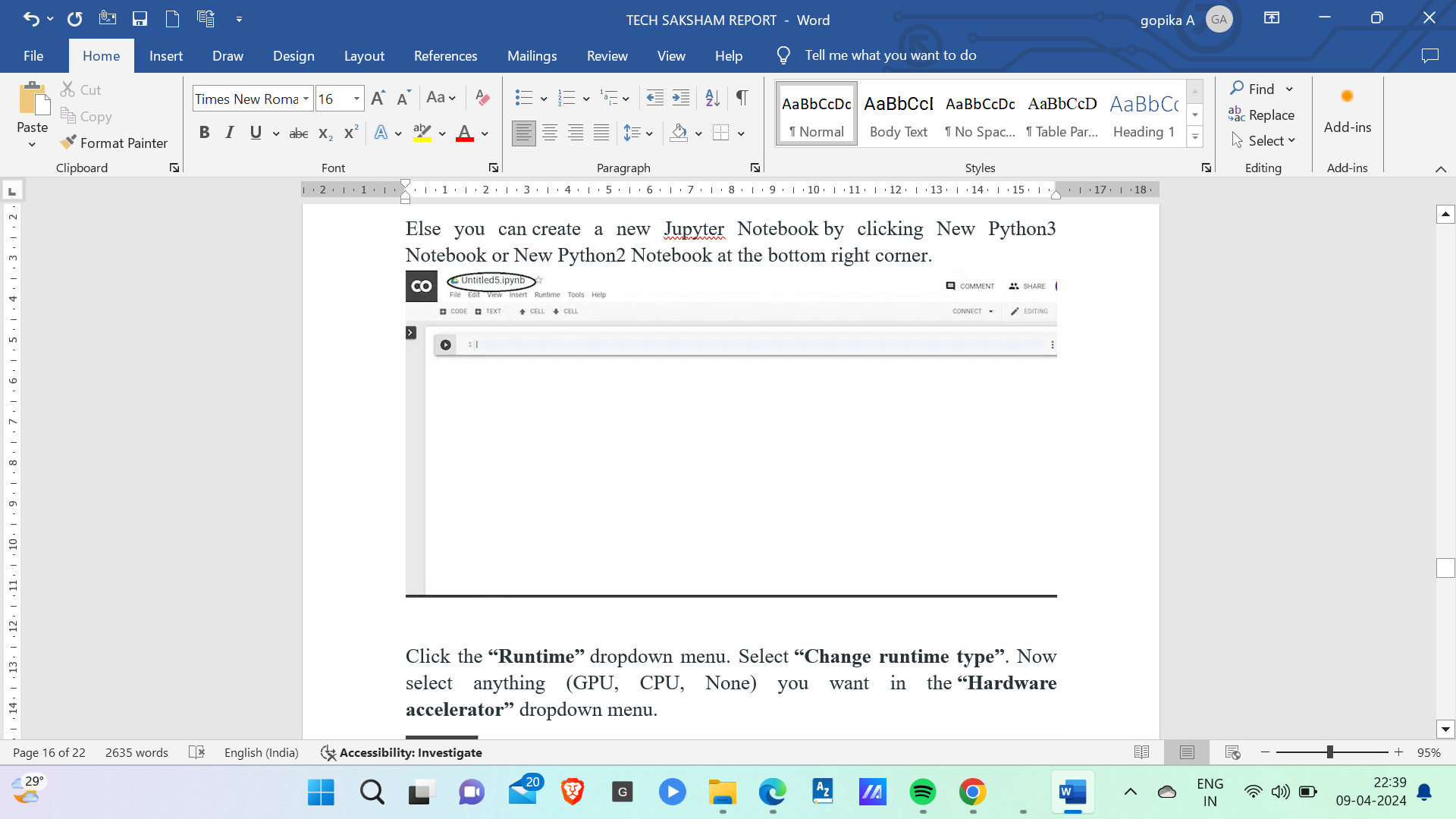
On opening the website you will see a pop-up containing the following tabs –

****

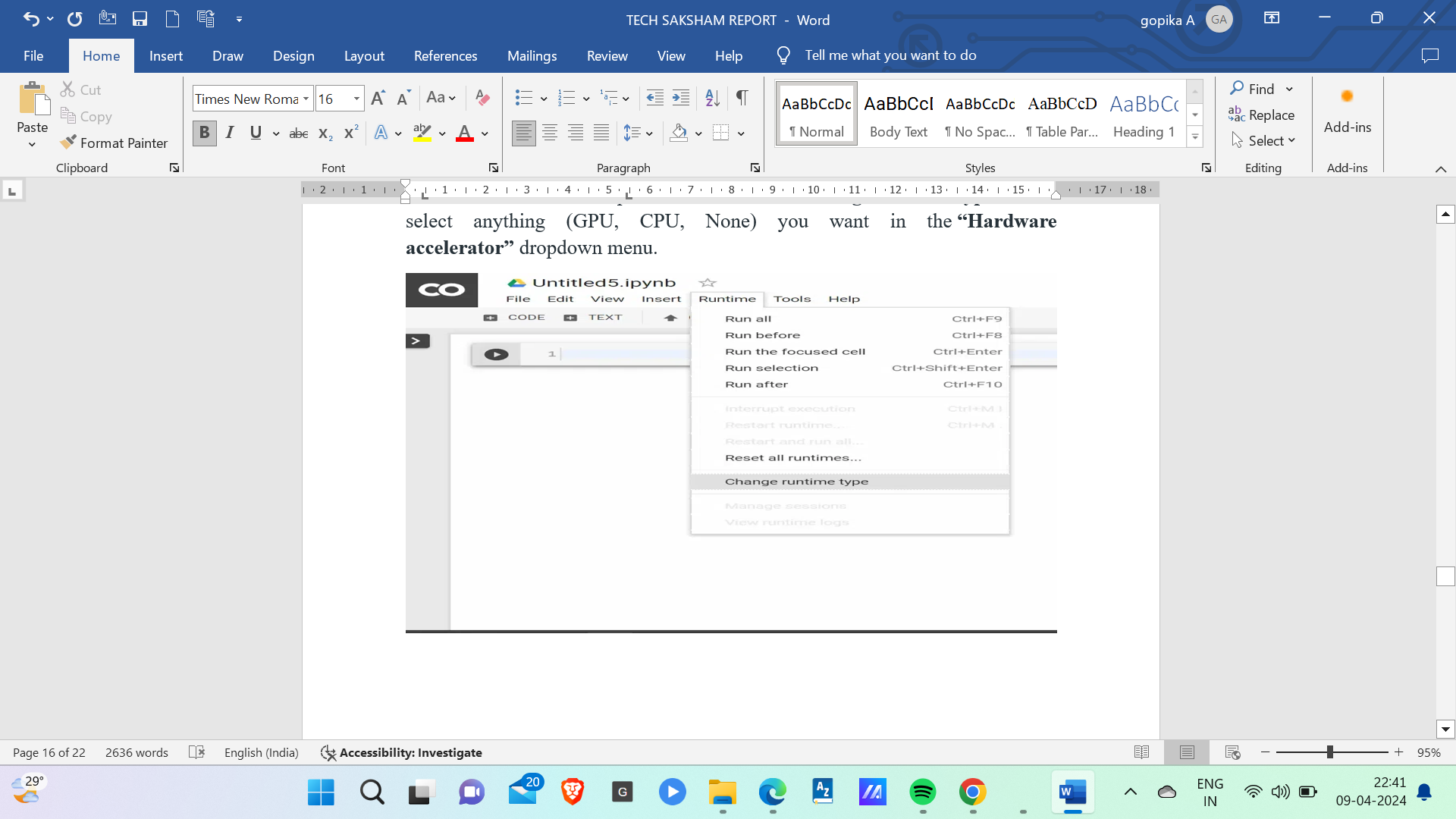
* **EXAMPLES:** Contain a number of Jupyter notebooks of various examples.
* **RECENT:** Jupyter notebook you have recently worked with.
* **GOOGLE DRIVE:** Jupyter notebook in your google drive.
* **GITHUB:** You can add Jupyter notebook from your GitHub but you first need to connect Colab with GitHub.
* **UPLOAD:** Upload from your local directory.

**CREATE COLLABORATORY NOTEBOOK**

Else you can create a new Jupyter Notebook by clicking New Python3 Notebook or New Python2 Notebook at the bottom right corner.

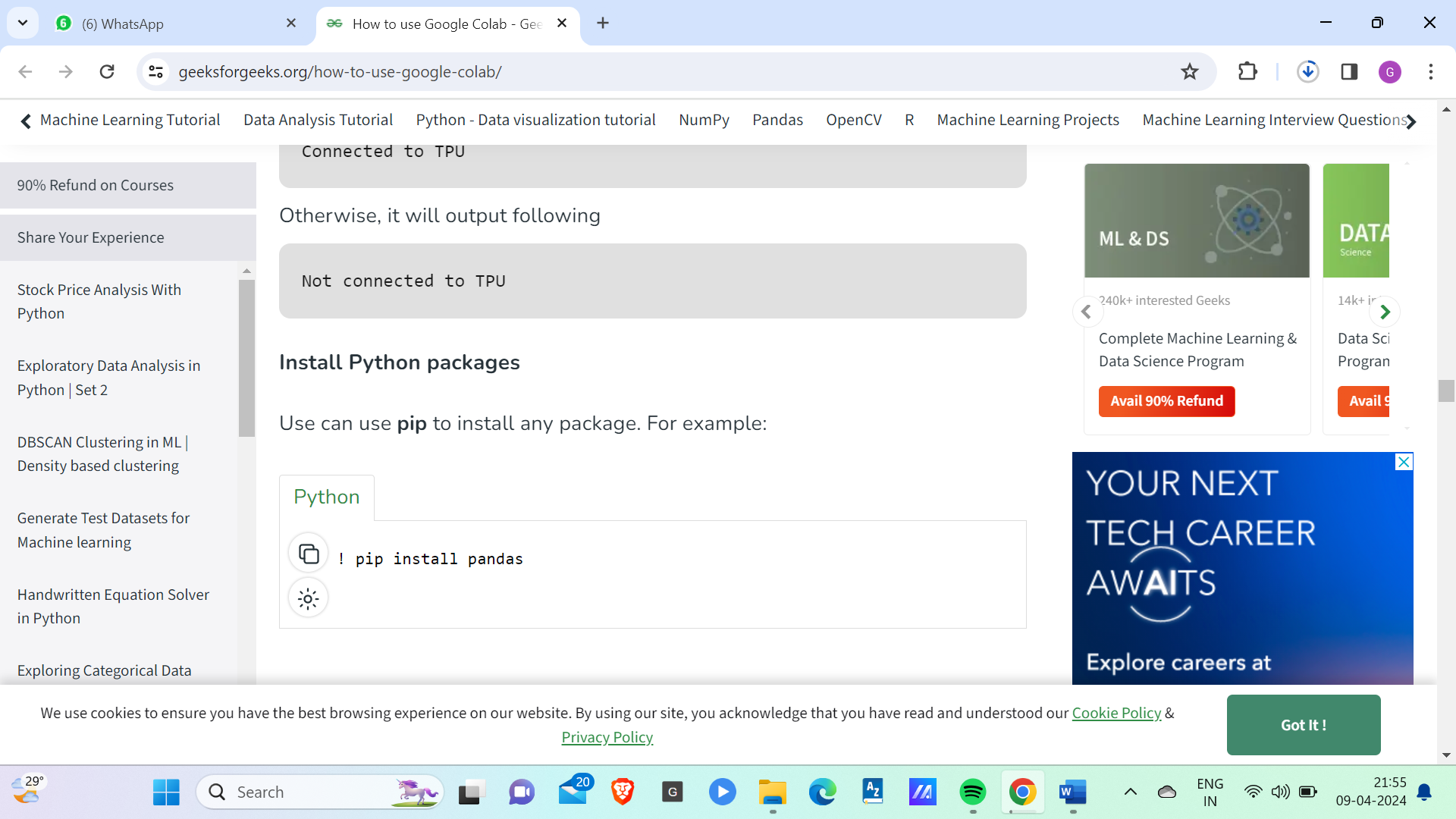


Click the **“Runtime”** dropdown menu. Select **“Change runtime type”**. Now select anything (GPU, CPU, None) you want in the **“Hardware accelerator”** dropdown menu.

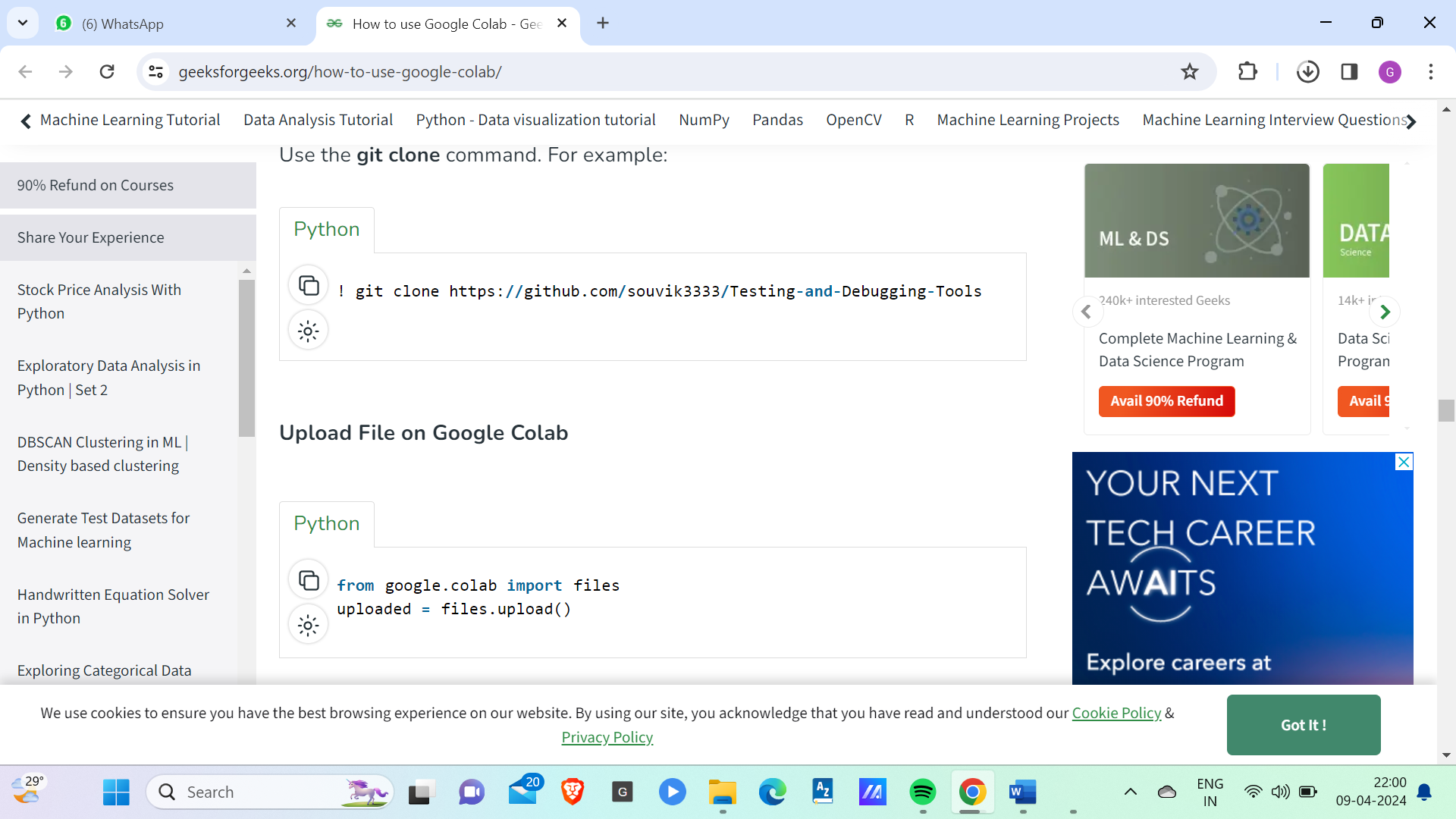
****

**INSTALL PYTHON PACKAGES**

Use can use **pip** to install any package. For example:

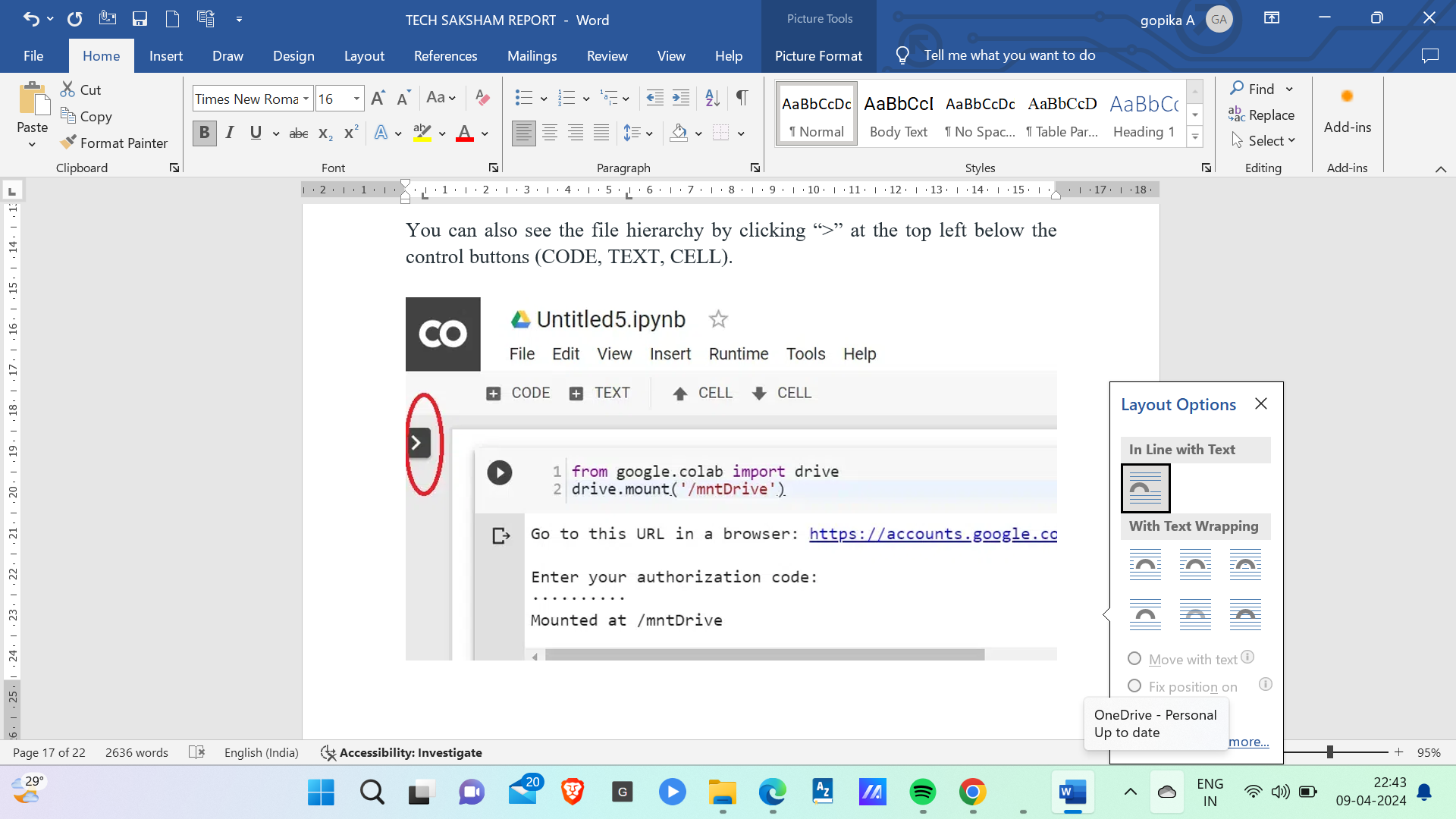


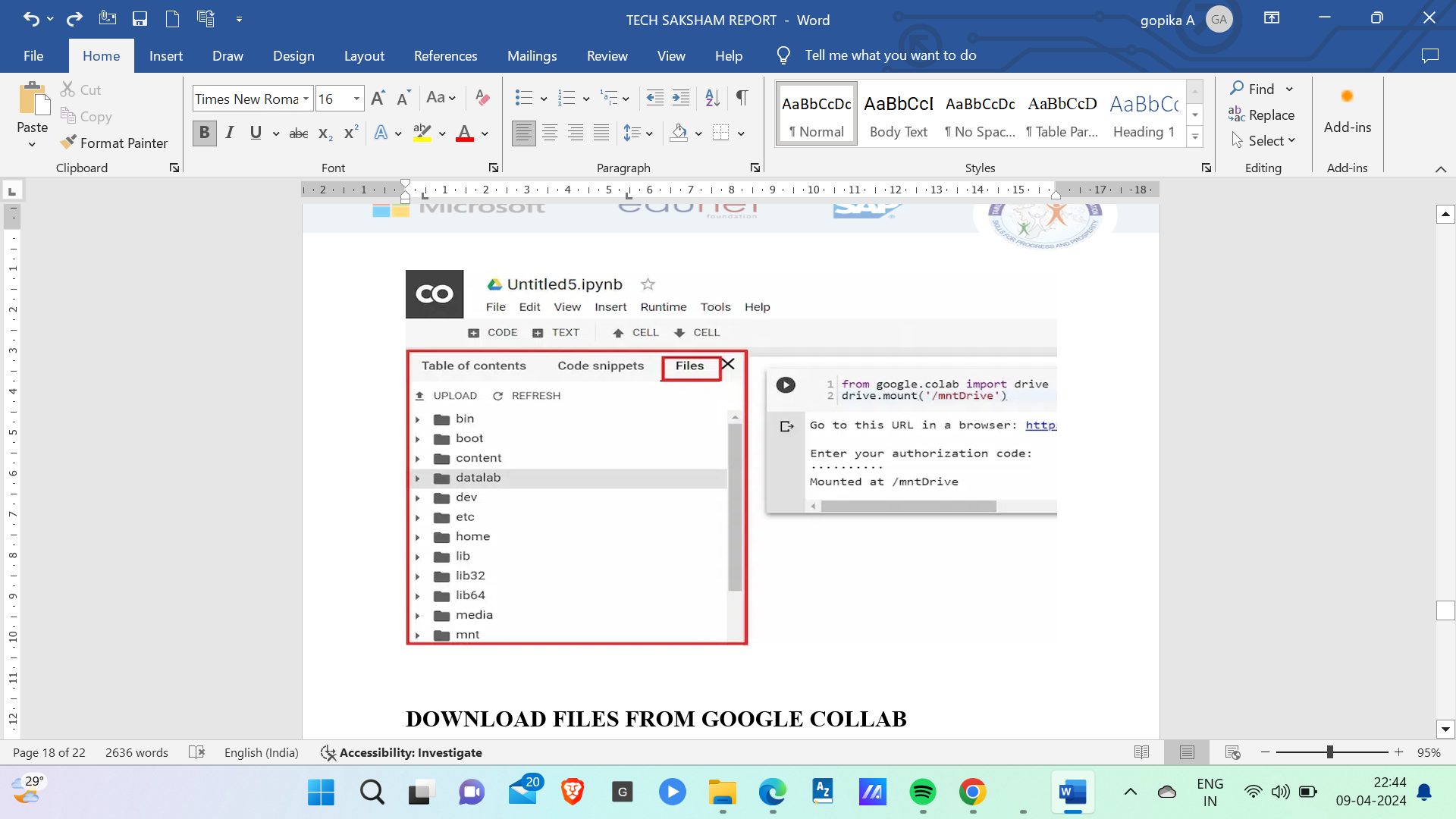
**UPLOAD FILE ON GOOGLE COLAB**

****

**FILE HEIRARCHY IN GOOGLE COLAB**

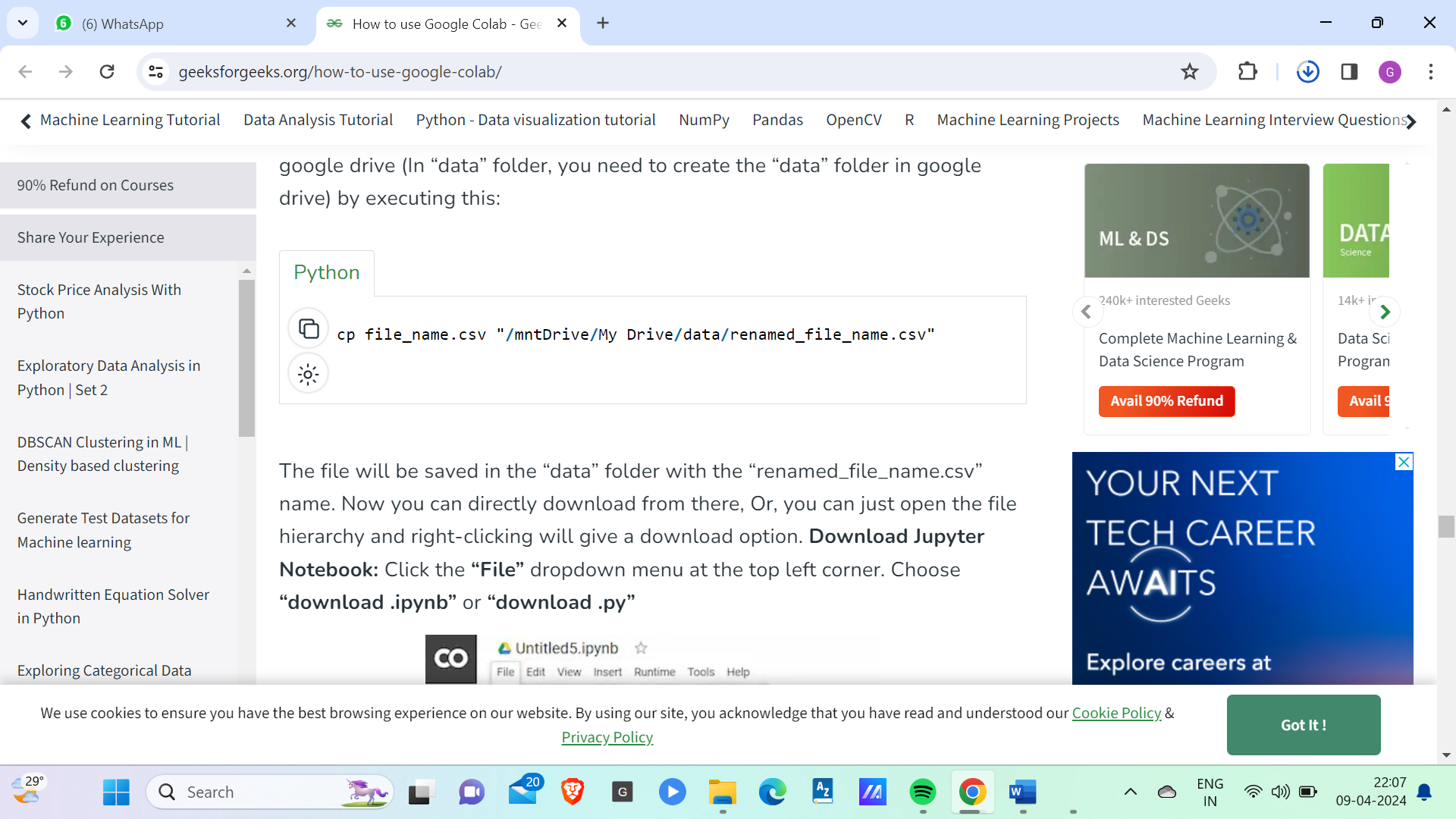
You can also see the file hierarchy by clicking “>” at the top left below the control buttons (CODE, TEXT, CELL).

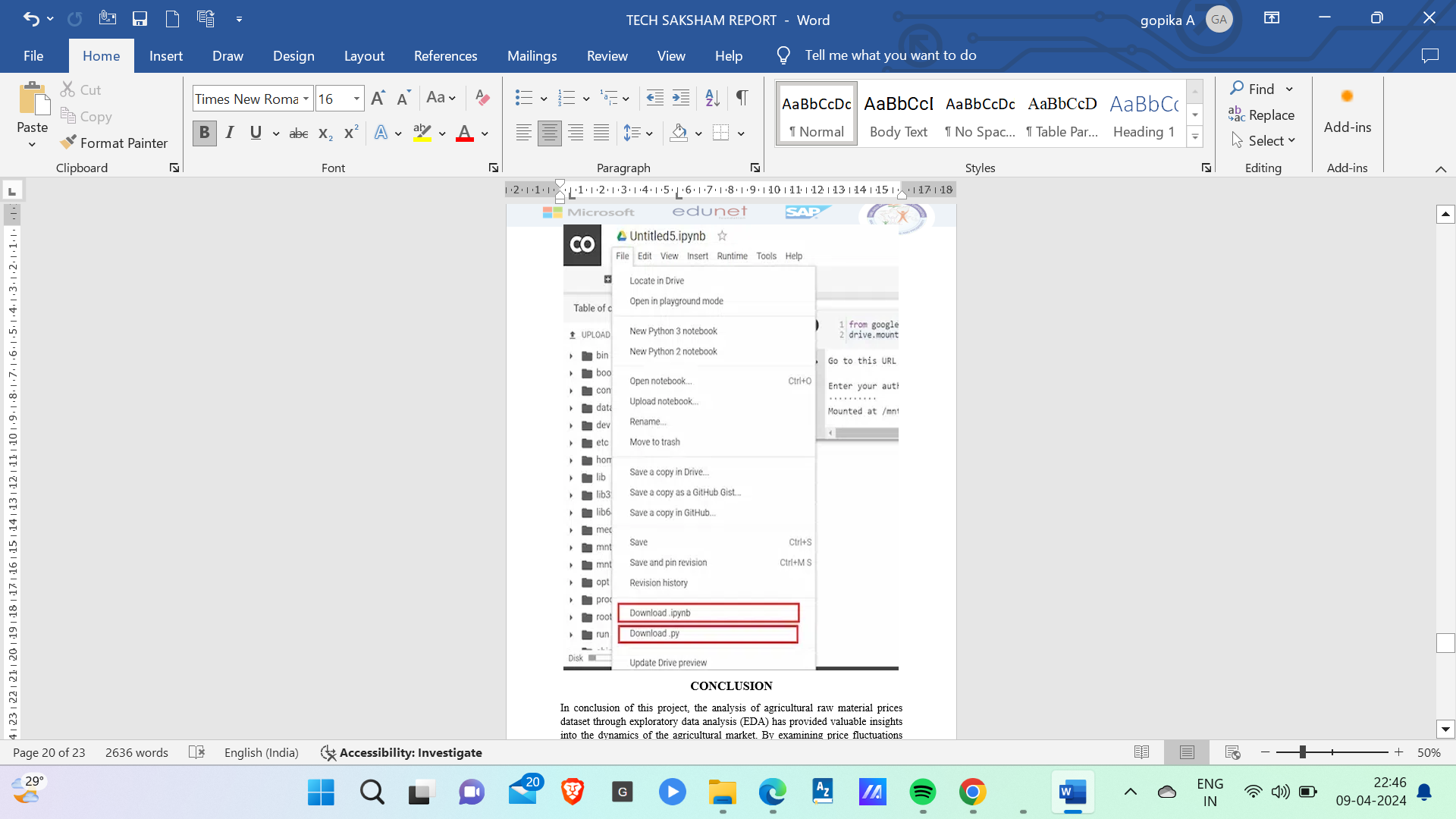


****

**DOWNLOAD FILES FROM GOOGLE COLLAB**

Let’s say you want to download “file\_name.csv”. You can copy the file to your google drive (In “data” folder, you need to create the “data” folder in google drive) by executing this:

 The file will be saved in the “data” folder with the “renamed\_file\_name.csv” name. Now you can directly download from there, Or, you can just open the file hierarchy and right-clicking will give a download option. **Download Jupyter Notebook:**Click the **“File”** dropdown menu at the top left corner. Choose **“download.ipynb”** or **“download .py”**

****

**CONCLUSION**

* By identifying both high-range and low-range raw materials based on their prices, as well as calculating the high and low percentage changes, we have gained a comprehensive understanding of the market dynamics.
* Moreover, examining the range of price changes over the years has shed light on the volatility and stability of different agricultural commodities. This information is crucial for stakeholders to make informed decisions regarding investment, trading, and risk management in the agricultural sector.
* Correlation between various raw materials using a heatmap has allowed us to discern any interdependencies or patterns in their price movements. This correlation analysis aids in identifying potential market trends and opportunities for diversification or hedging strategies.
* Overall, this EDA has not only provided a snapshot of historical price trends but also laid the groundwork for predictive modeling and further analysis to enhance decision-making processes in the agricultural industry.

**FUTURE SCOPE**

The future scope of this project is that it helps in:

1. **Interpreting Results**: Interpret the findings from the analysis to extract meaningful insights for stakeholders. This may involve identifying key drivers of price movements, understanding market dynamics, and assessing the impact of external factors such as weather conditions, geopolitical events, and economic policies.
2. **Recommendations and Future Predictions**: Based on the analysis results, provide recommendations for stakeholders such as farmers, traders, and policymakers. Additionally, explore the potential for developing predictive models to forecast future price movements and mitigate risks associated with agricultural raw material markets.

Thus, analyzing an agricultural raw material prices dataset and conducting exploratory data analysis (EDA) offers valuable insights into market trends, price volatility, and potential correlations between different commodities.

**REFERENCES**

* [**https://www.analyticsvidhya.com/blog/2022/07/step-by-step-exploratory-data-analysis-eda-using-python/**](https://www.analyticsvidhya.com/blog/2022/07/step-by-step-exploratory-data-analysis-eda-using-python/)
* [**https://builtin.com/data-science/EDA-python**](https://builtin.com/data-science/EDA-python)

**LINKS**

[**https://github.com/kavinrajesh15/Kavin**](https://github.com/kavinrajesh15/Kavin)