TECH SAKSHAM

CAPSTONE PROJECT REPORT

"AGRICULTURAL RAW MATERIAL ANALYSIS"

"COLLEGE OF ENGINEERING GUINDY"

NM ID	NAME
au2021109011 5006DF3AB5BC804A3BD49783EB884BAB	GOPIKA A

TRAINER NAME

RAMAR BOSE

Sr. AI MASTER TRAINER

ABSTRACT

This study delves into the analysis of agricultural raw material prices over several years, employing exploratory data analysis (EDA) techniques. The primary objectives include identifying high and low-range raw materials based on their prices, pinpointing materials with the highest and lowest percentage changes, and assessing the range of price fluctuations over time. Additionally, the study aims to map the correlation between various raw materials using a heatmap visualization, providing insights into interdependencies within the agricultural market. Through comprehensive data analysis, this research contributes to understanding the dynamics and trends of agricultural raw material prices, which are crucial for stakeholders in the industry.

INDEX

CHAPTER 1

INTRODUCTION

1.1 PROBLEM STATEMENT

With the increasing complexity of the agricultural market, understanding the dynamics of raw material prices has become paramount for stakeholders. This study aims to address key questions regarding agricultural raw material prices over time through exploratory data analysis (EDA). Specifically, the research seeks to identify high and low-range raw materials based on price fluctuations, pinpoint materials exhibiting the highest and lowest percentage changes, and assess the range of price variations across different years. Additionally, the study endeavors to map correlations between various raw materials using a heatmap visualization, providing insights into interdependencies within the agricultural market. By tackling these objectives, this research aims to contribute valuable insights to support decision-making processes and strategic planning in the agricultural sector.

1.2 PROPOSED SOLUTION

To address the task of analyzing agricultural raw material prices dataset over the years and performing exploratory data analysis(EDA) tasks, including finding 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, the proposed solution outlines : data loading and preprocessing, identifying high range and low range raw materials, identifying high and low % change materials, determining range of prices changed over the years, mapping correlation between raw materials using heatmap, interpretation and insights, visualization and reporting, documentation and code reusability. By following this proposed solution, stakeholders can gain valuable insights into the trends, variation, and correlations within the agricultural raw material prices dataset, facilitating informed decision making and strategic planning in the agricultural industry.

1.3 FEATURE

- Real time analysis
- Price range analysis
- Percentage change analysis
- Price fluctuation analysis
- Correlation analysis
- Summary statistics
- Trend analysis
- Outlier detection
- Comparison analysis
- Spatial analysis
- Forecasting

1.4 ADVANTAGES

- **Data-Driven Decision Making:** EDA provides stakeholders with data-driven insights into the agricultural raw material market, enabling informed decision-making based on empirical evidence rather than intuition or guesswork.
- **Identification of Market Trends:** Analysis of price fluctuations over the years helps identify long-term market trends, enabling stakeholders to anticipate future market conditions and adjust strategies accordingly.
- **Strategic Planning:** EDA serves as a foundation for strategic planning in the agricultural industry, helping stakeholders develop robust business strategies, set realistic goals, and allocate resources effectively to achieve their objectives.
- **Improved Stakeholder Communication:** Visualizations such as heatmaps and summary statistics enhance communication and collaboration among stakeholders by providing clear and intuitive representations of complex data patterns and relationships.

1.5 SCOPE

The project aims to conduct exploratory data analysis (EDA) on agricultural raw material prices dataset over multiple years. This provides insights into price trends, variations, and correlations within the agricultural market, facilitating informed decision making for stakeholders.

CHAPTER 2
SERVICE AND TOOLS REQUIRED

2.1 SERVICES USED

In Google Colab, a cloud-based Jupyter notebook environment, various services and features are available to perform the specified task of analyzing agricultural raw material prices dataset. Some of the services that can be utilized include:

- **Data Loading and Management:** Google Colab provides access to Google Drive, allowing users to upload datasets and manage files easily. Additionally, users can import data from external sources using libraries like pandas or directly from Google Cloud Storage.
- **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.
- **Collaboration and Sharing:** Colab allows users to share and collaborate on notebooks with others in real-time. It also integrates with Google Drive, enabling easy sharing and access to notebooks and datasets.
- **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.
- **Seaborn:** Seaborn is a statistical data visualization library that works on top of Matplotlib and provides a high-level interface for creating attractive and informative statistical graphics. It can be used to create more complex and visually appealing plots, including heatmaps for correlation analysis.
- **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.

By leveraging these tools and software in Google Colab, users can effectively analyze agricultural raw material prices dataset, conduct exploratory data analysis tasks, and derive valuable insights to support decision-making processes.

CHAPTER 3

PROJECT ARCHITECTURE

3.1 ARCHITECTURE

Here's a high-level project architecture for the task of analyzing agricultural raw material prices dataset over the years:

Data Acquisition and Preprocessing:

- **Data Collection**: Obtain the agricultural raw material prices dataset from reliable sources or databases.
- **Data Cleaning:** Handle missing values, outliers, and inconsistencies in the dataset. Standardize or normalize the data if necessary.
- **Data Transformation:**Convert the dataset into a suitable format for analysis, such as a pandas DataFrame in Python.

Exploratory Data Analysis (EDA):

- **Descriptive Statistics:** Calculate summary statistics (e.g., mean, median, standard deviation) for prices of each raw material.
- **Price Range Analysis:** Identify high-range and low-range raw materials based on their prices.
- **Percentage Change Analysis:** Compute the percentage change in prices for each raw material over consecutive years to identify high and low %Change materials.
- **Price Fluctuation Analysis:** Determine the range of price changes for each raw material across different years.
- **Correlation Analysis:** Map correlations between raw materials using a heatmap.

Visualization:

Generate visualizations to present the analysis results, including histograms, line plots, scatter plots, and heatmaps.

Visualize the distribution of raw material prices, percentage changes, and correlations between raw materials using appropriate plots and charts.

Modeling (Optional):

Develop predictive models or forecasting techniques to predict future prices of agricultural raw materials based on historical data and trends. This step is optional but can provide additional insights for stakeholders.

Documentation and Reporting:

Document the analysis process, methodologies, and findings in a comprehensive report.

Summarize key insights, trends, and recommendations for stakeholders to facilitate decision-making.

Deployment and Integration (Optional):

Deploy the analysis pipeline or models to production environments for ongoing monitoring and analysis.

Integrate the analysis results with existing systems or dashboards for easy access and visualization by stakeholders.

Maintenance and Iteration:

Regularly update and maintain the analysis pipeline to incorporate new data and insights. Iterate on the analysis process based on feedback and evolving requirements to improve the accuracy and effectiveness of the analysis.

This project architecture provides a structured approach to analyzing agricultural raw material prices dataset, conducting exploratory data analysis, and deriving actionable insights to support decision-making processes in the agricultural industry

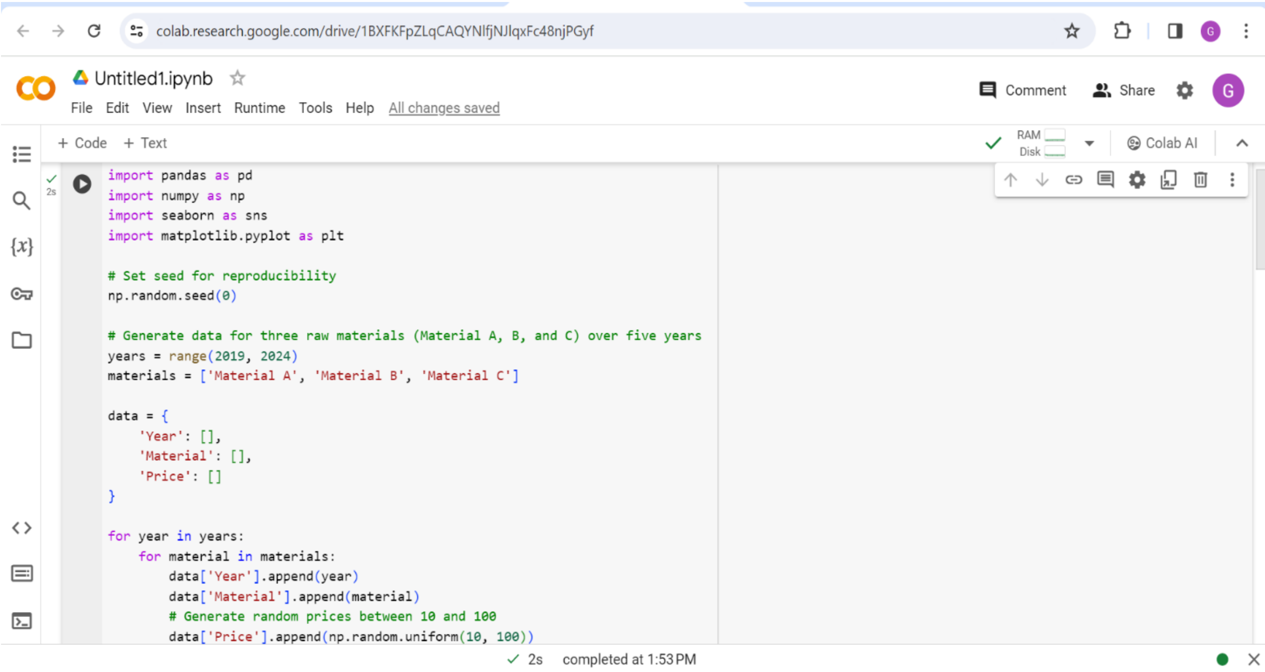
CHAPTER 4

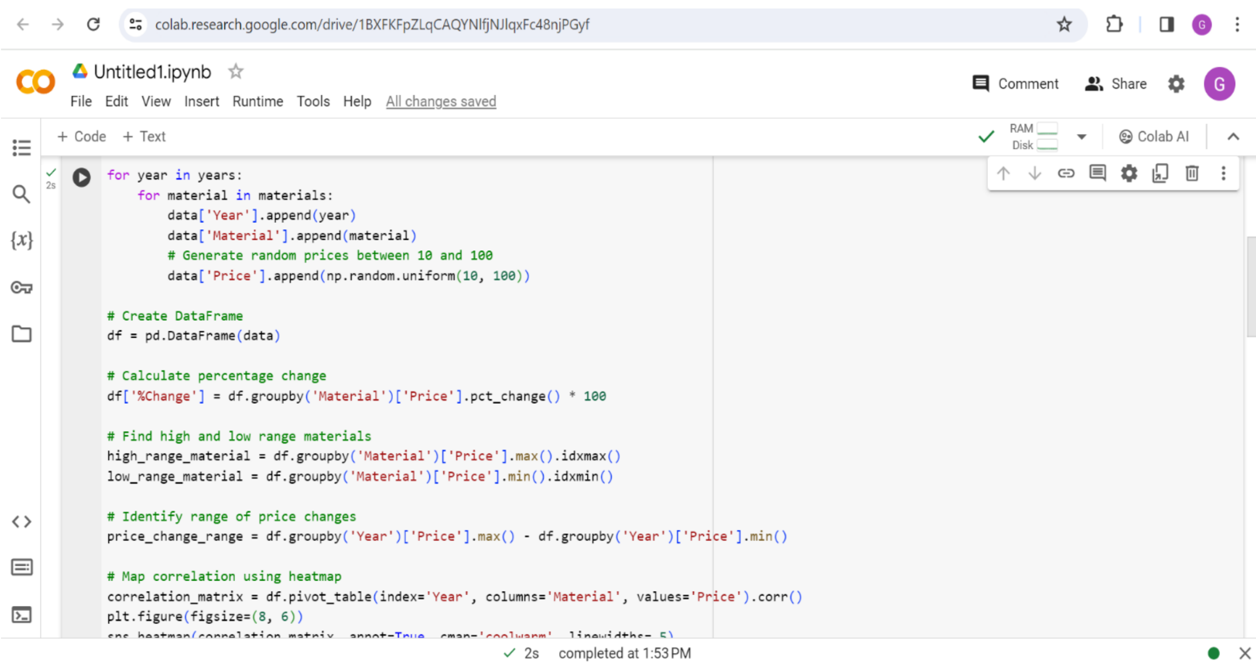
MODELING AND PROJECT OUTCOME

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.

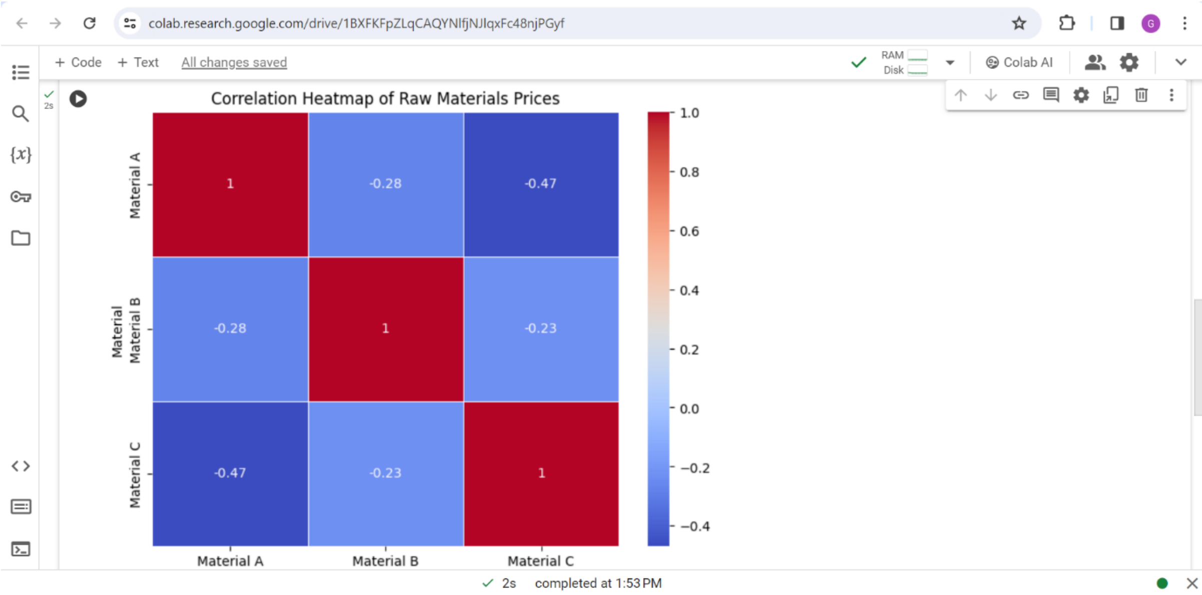
PROGRAM DONE IN GOOGLE COLAB:

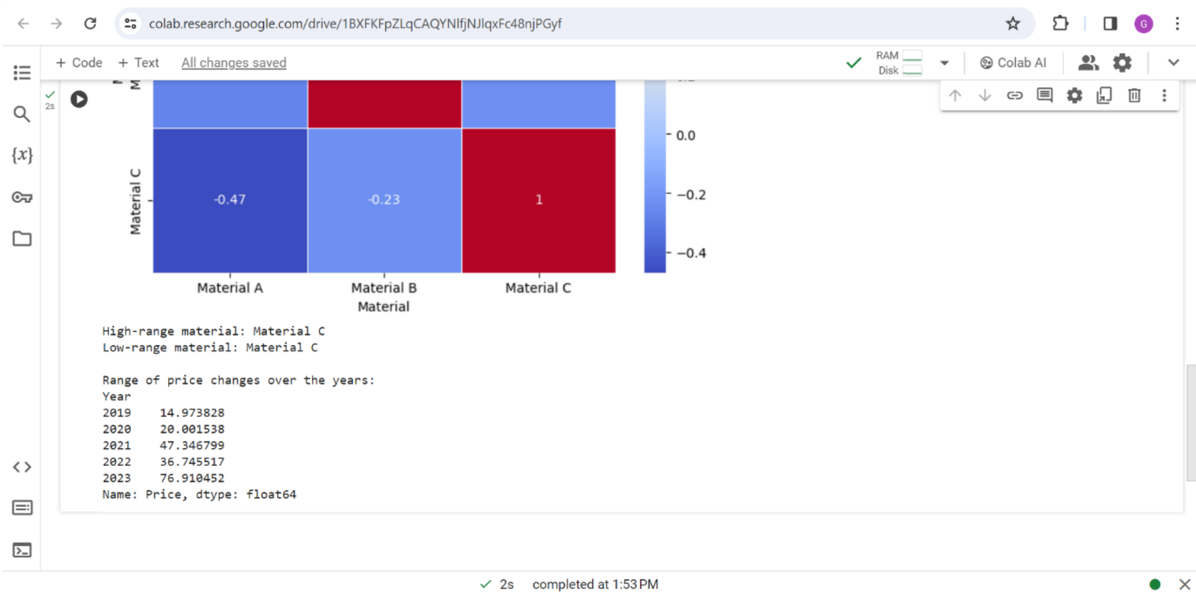





colab.research.google.com/drive/1BXFKFpZLqCAQYNIfjNJlqxFc48njPGyf
Untitled1.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
+ Code + Text
RAM
Disk
Colab AI
# Map correlation using heatmap
correlation_matrix = df.pivot_table(index='Year', columns='Material', values='Price').corr()
plt.figure(figsize=(8, 6))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', linewidths=.5)
plt.title('Correlation Heatmap of Raw Materials Prices')
plt.show()
print("High-range material:", high_range_material)
print("Low-range material:", low_range_material)
print("\nRange of price changes over the years:")
print(price_change_range)
Correlation Heatmap of Raw Materials Prices
Material A
1
-0.28
-0.47
1.0
0.8
0.6
2s completed at 1:53 PM

PROGRAM OUTPUT FROM THE ABOVE PROGRAM:





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.

Benefits of Google Colab:

Google Colab offers several benefits that make it a popular choice among data scientists, researchers, and machine learning practitioners. **Key features of Google Collaboratory notebook include:**

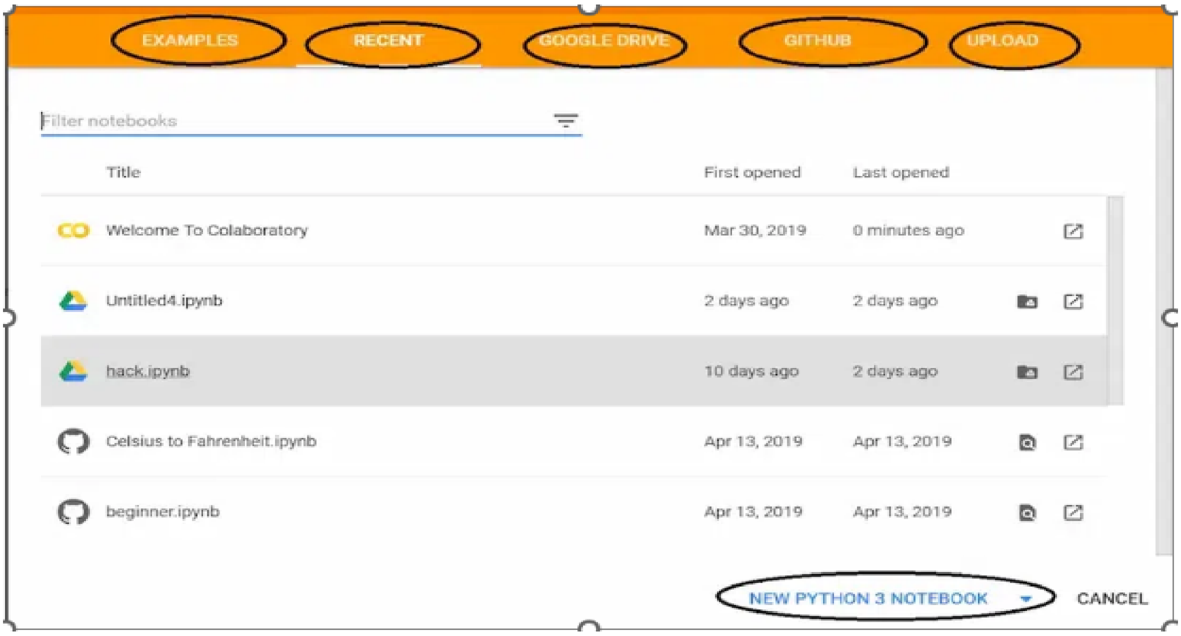
- **Free Access to GPUs:** Colab offers free GPU access, which is particularly useful for training machine learning models that require significant computational power.
- **No Setup Required**: Colab runs in the cloud, eliminating the need for users to set up and configure their own development environment. This makes it convenient for quick coding and collaboration.
- **Collaborative Editing:** Multiple users can work on the same Colab notebook simultaneously, making it a useful tool for collaborative projects.
- **Integration with Google Drive**: Colab is integrated with Google Drive, allowing users to save their work directly to their Google Drive account. This enables easy sharing and access to notebooks from different devices.
- **Support for Popular Libraries**: Colab comes pre-installed with many popular Python libraries for machine learning, data analysis, and visualization, such as TensorFlow, PyTorch, Matplotlib, and more.
- **Easy Sharing**:Colab notebooks can be easily shared just like Google Docs or Sheets. Users can provide a link to the notebook, and others can view or edit the code in real-time.

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.

OPEN COLLABORATORY NOTEBOOK

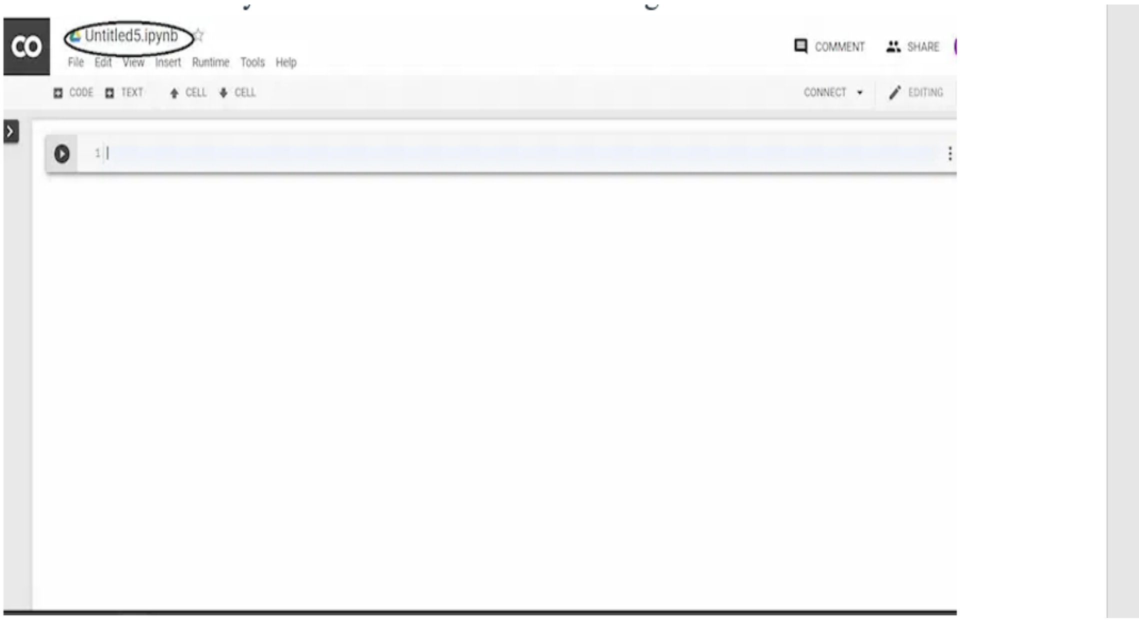
On opening the website you will see a pop-up containing the following tabs –



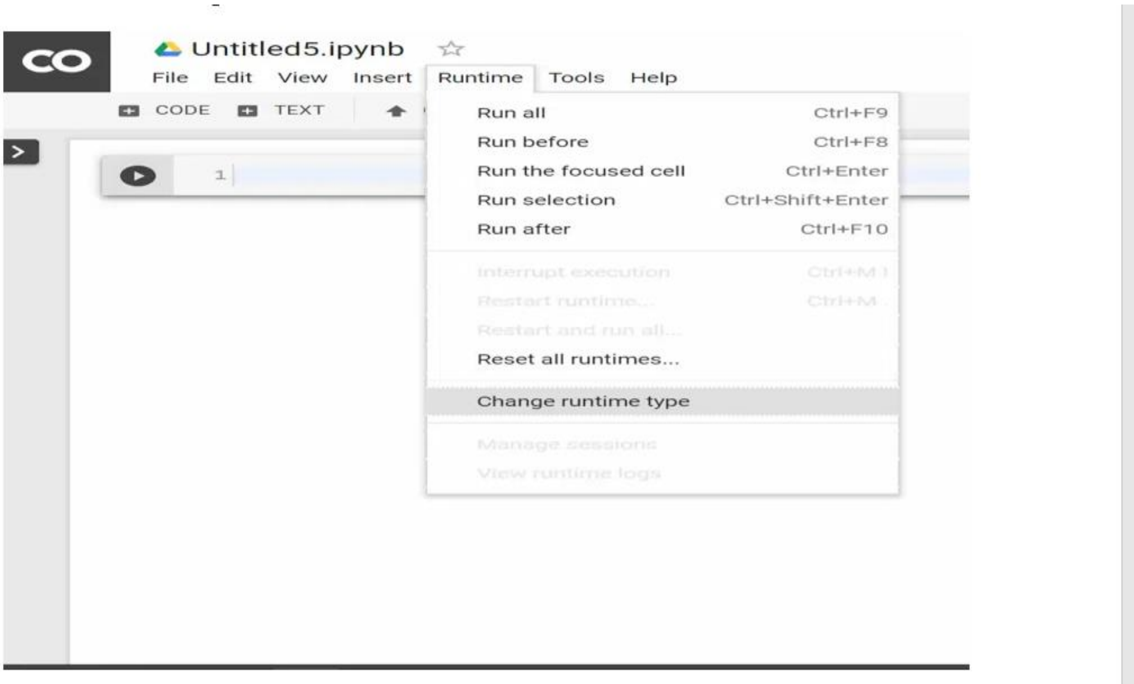
- **EXAMPLES:** Contain a number of Jupyter notebooks of various examples.
- **RECENT:** Jupyter notebook you have recently worked with.
- **GOOGLE DRIVE:** Jupyter notebook in your google drive.
- **GITHUB:** You can add Jupyter notebook from your GitHub but you first need to connect Colab with GitHub.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

Python

```
! pip install pandas
```

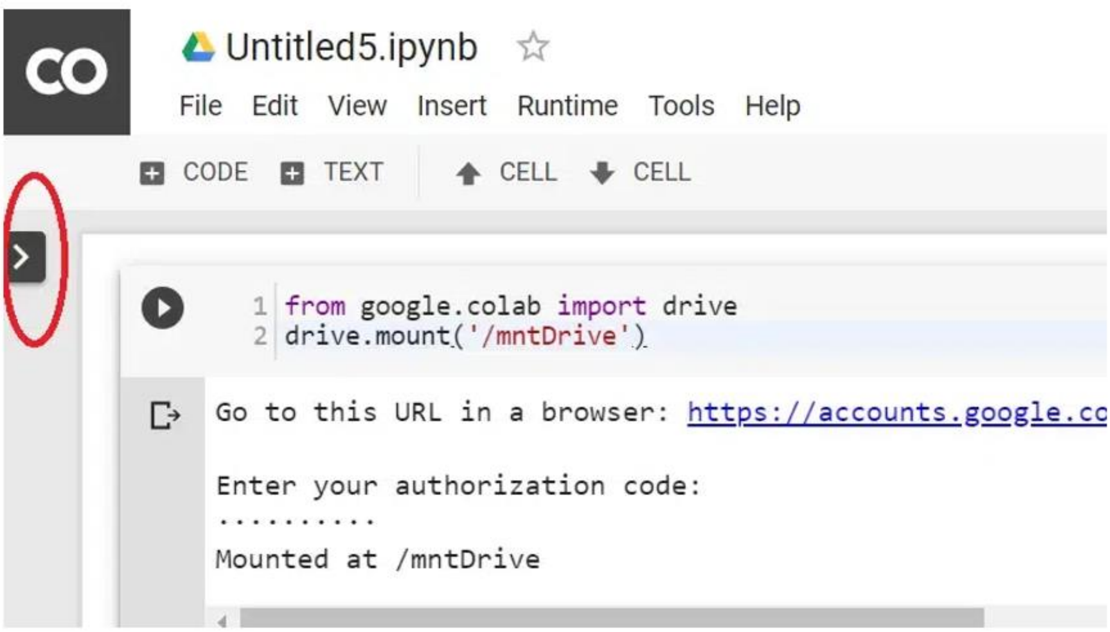
UPLOAD FILE ON GOOGLE COLAB

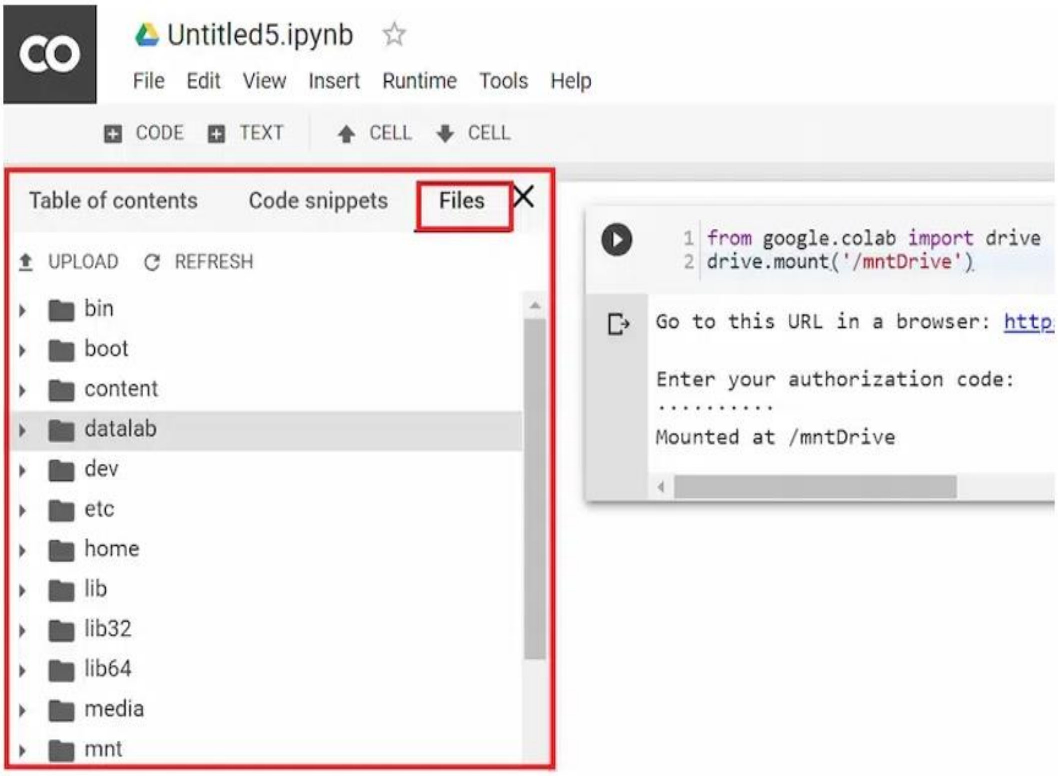
Python

```
from google.colab import files
uploaded = files.upload()
```

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).





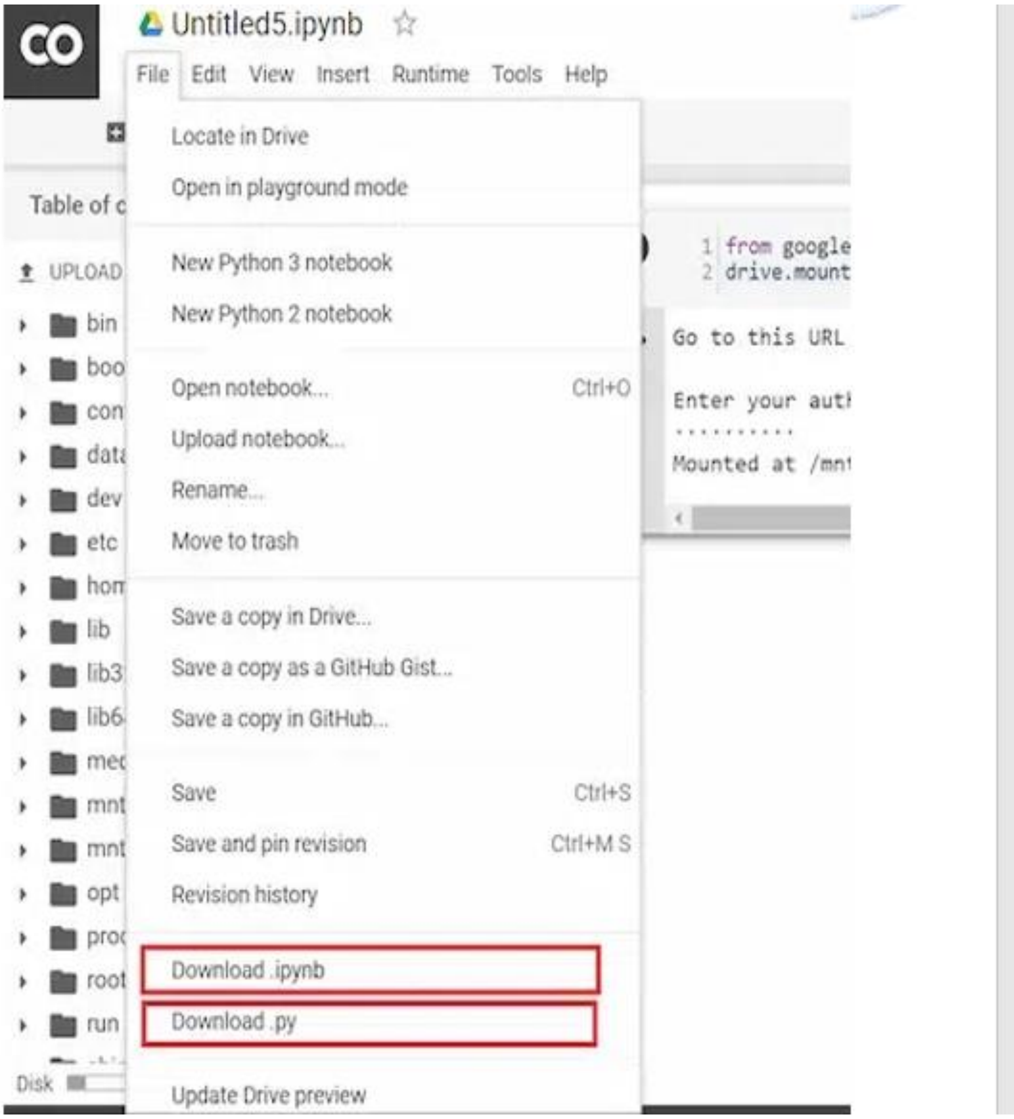
DOWNLOAD FILES FROM GOOGLE COLLAB

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:

Python

```
cp file_name.csv "/mntDrive/My Drive/data/renamed_file_name.csv"
```

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”**



co
Untitled5.ipynb
File
Edit
View
Insert
Runtime
Tools
Help
Locate in Drive
Open in playground mode
New Python 3 notebook
New Python 2 notebook
Open notebook... Ctrl+O
Upload notebook...
Rename...
Move to trash
Save a copy in Drive...
Save a copy as a GitHub Gist...
Save a copy in GitHub...
Save Ctrl+S
Save and pin revision Ctrl+M S
Revision history
Download .ipynb
Download .py
Update Drive preview
UPLOAD
bin
etc
lib
opt
root
run
Disk
Go to this URL
Mounted at

CONCLUSION

In conclusion of this project, the analysis of agricultural raw material prices dataset through exploratory data analysis (EDA) has provided valuable insights into the dynamics of the agricultural market. By examining price fluctuations over the years, we were able to identify both high-range and low-range raw materials based on their prices. Furthermore, the assessment of percentage changes in prices highlighted materials with the highest and lowest percentage changes, offering additional perspectives on market trends. Additionally, the range of price changes observed over the years underscored the volatility and variability inherent in the agricultural sector, shedding light on the challenges and opportunities faced by stakeholders.

The mapping of correlations between different raw materials using a heatmap visualization revealed potential interdependencies and relationships within the market, paving the way for a deeper understanding of market dynamics and trends. Overall, this analysis serves as a foundation for informed decision-making and strategic planning in the agricultural industry. By leveraging these insights, stakeholders can adapt their strategies, optimize resource allocation, and mitigate risks to navigate the complexities of the agricultural market effectively. Moving forward, further exploration and enhancement of analytical techniques and data sources will continue to drive innovation and value creation in the agricultural sector.

FUTURE SCOPE

The analysis of agricultural raw material prices dataset offers several avenues for future exploration and enhancement:

- **Machine Learning Predictive Models:** Implement machine learning algorithms to forecast future prices of agricultural raw materials based on historical data. Models such as time series forecasting (e.g., ARIMA, SARIMA) and regression analysis can provide valuable insights for stakeholders to make informed decisions.
- **Advanced Statistical Analysis: Explore** advanced statistical techniques to uncover hidden patterns and trends within the dataset. This could involve time series decomposition, cluster analysis, or advanced correlation analysis methods to delve deeper into the relationships between raw materials.
- **Integration with External Data Sources:** Incorporate additional datasets, such as weather data, commodity market trends, or geopolitical factors, to enhance the analysis and provide a more comprehensive understanding of the factors influencing agricultural raw material prices.
- **Real-Time Data Analysis:** Implement mechanisms to ingest and analyze real-time or streaming data on agricultural raw material prices. This enables stakeholders to react quickly to market changes and adapt their strategies accordingly.
- **Predictive Analytics for Risk Management:** Utilize predictive analytics to assess and mitigate risks associated with fluctuations in agricultural raw material prices. By identifying potential risks and developing risk mitigation strategies, stakeholders can safeguard their operations and investments.
- **Collaborative Research and Data Sharing:** Foster collaboration with research institutions, government agencies, and industry partners to leverage collective expertise and data resources. Sharing insights and best practices can drive innovation and improve decision-making across the agricultural sector.

REFERENCES

- https://towardsdatascience.com/exploratory-data-analysis-8fc1cb20fd15
- https://seaborn.pydata.org/introduction.html
- https://machinelearningmastery.com/time-series-data visualization-with-python/
- https://towardsdatascience.com/heatmap-basics-with-pythons-seaborn-fb92ea280a6c

LINKS

https://github.com/GopikaArul/NaanMudhalvanProject.git