### TABLE OF CONTENTS

1. Introduction
2. Source of Dataset
3. Objectives
4. EDA Process
5. Analysis on Dataset

* General Description
* Specific Requirements, Functions and Formulas
* Analysis Results
* Visualizations

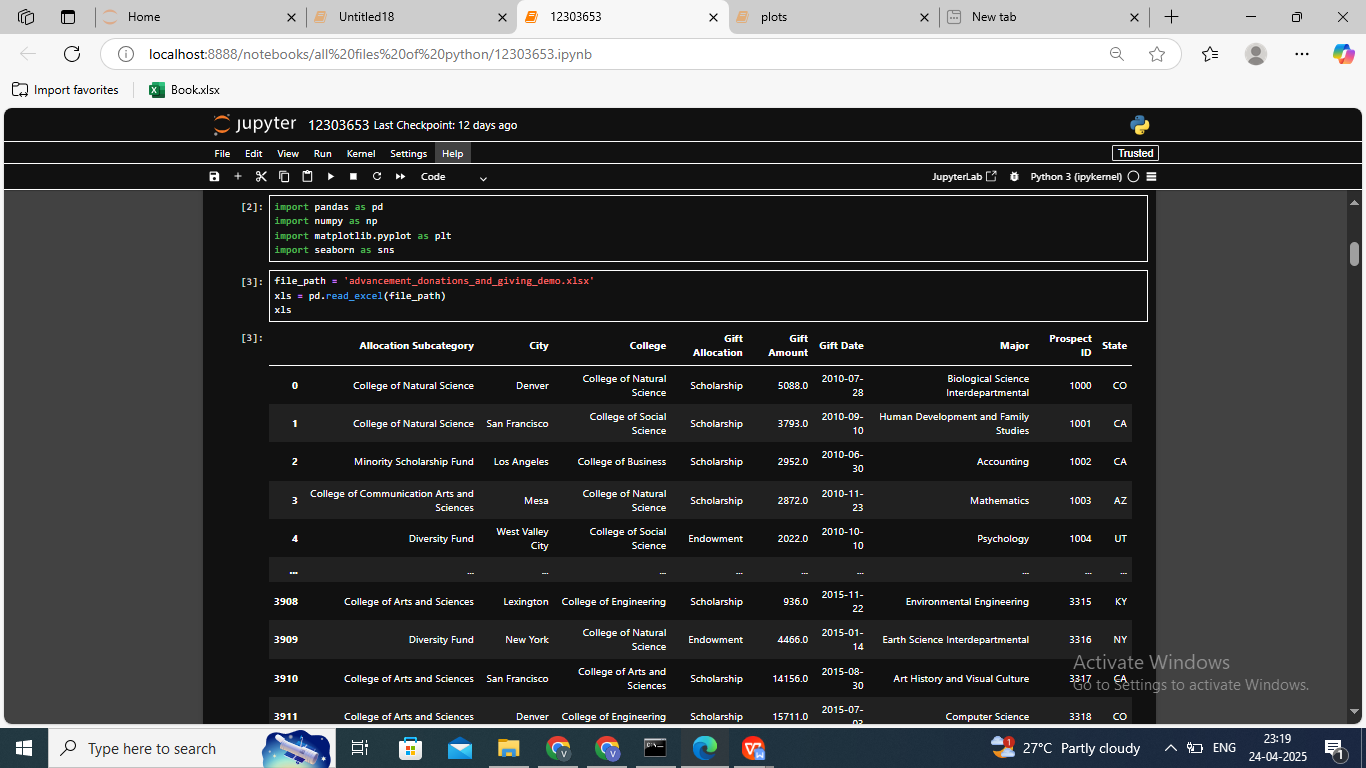
1. Conclusion
2. Future Scope
3. References

**INTRODUCTION**

In today’s data-driven world, educational institutions increasingly rely on analytics to understand trends and patterns in alumni engagement. Alumni donations represent not just financial support, but a reflection of loyalty, connection, and satisfaction toward the institution. Analyzing such behavior provides key insights for fundraising, development planning, and alumni relationship strategies.

This project, titled **"From Graduation To Generosity:An Exploratory Study On Alumni Giving Behavior"**, focuses on exploring a multi-dimensional dataset of alumni donations, using modern tools in Python for cleaning, analysis, and visualization. The aim is to uncover trends based on geographic, academic, and time-based factors.

We perform in-depth Exploratory Data Analysis (EDA), leveraging various visualizations such as histograms, boxplots, heatmaps, and correlation plots to identify meaningful insights. Additionally, the graduation data is cross-linked to analyze trends based on alumni seniority and background. Ultimately, this study contributes toward understanding how data science can empower institutions in crafting targeted outreach strategies.



### SOURCE OF DATASET

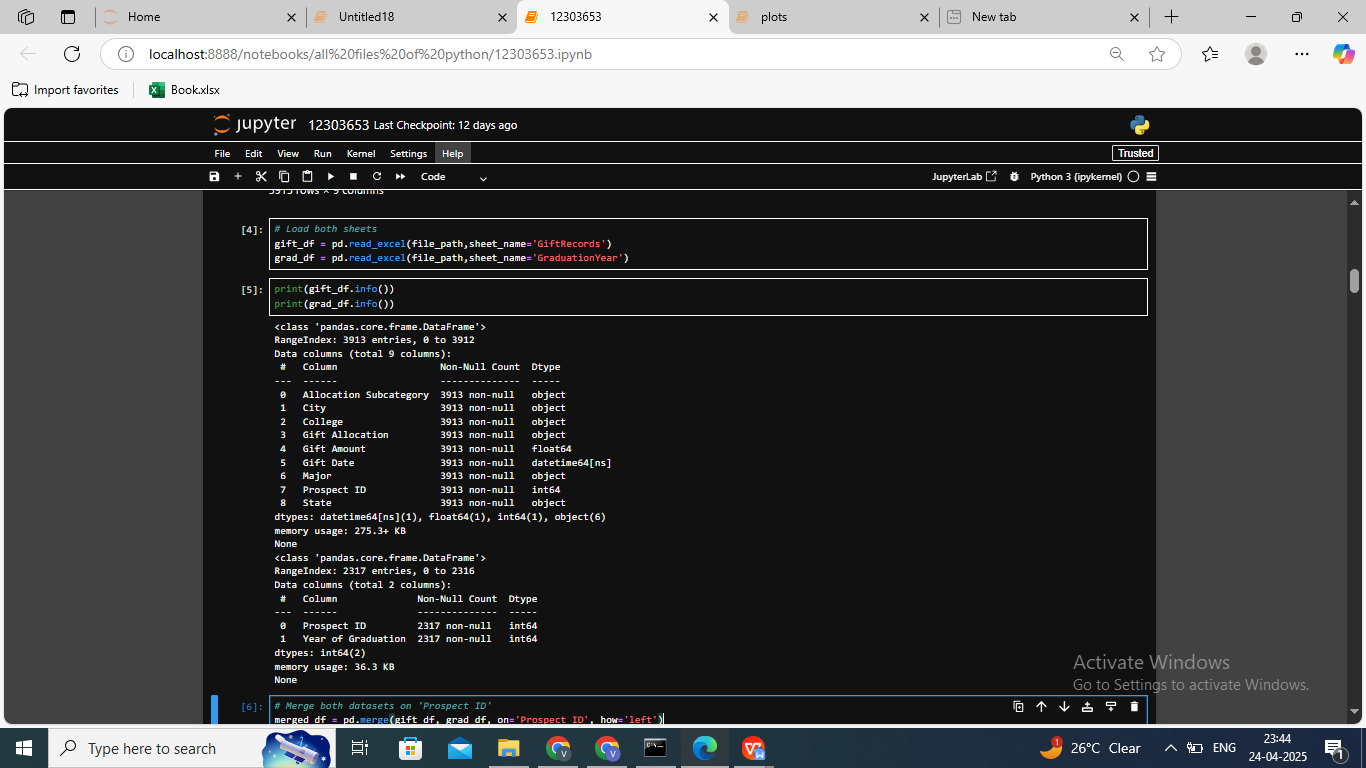
**Dataset Link**:-<https://public.tableau.com/app/learn/sample-data>

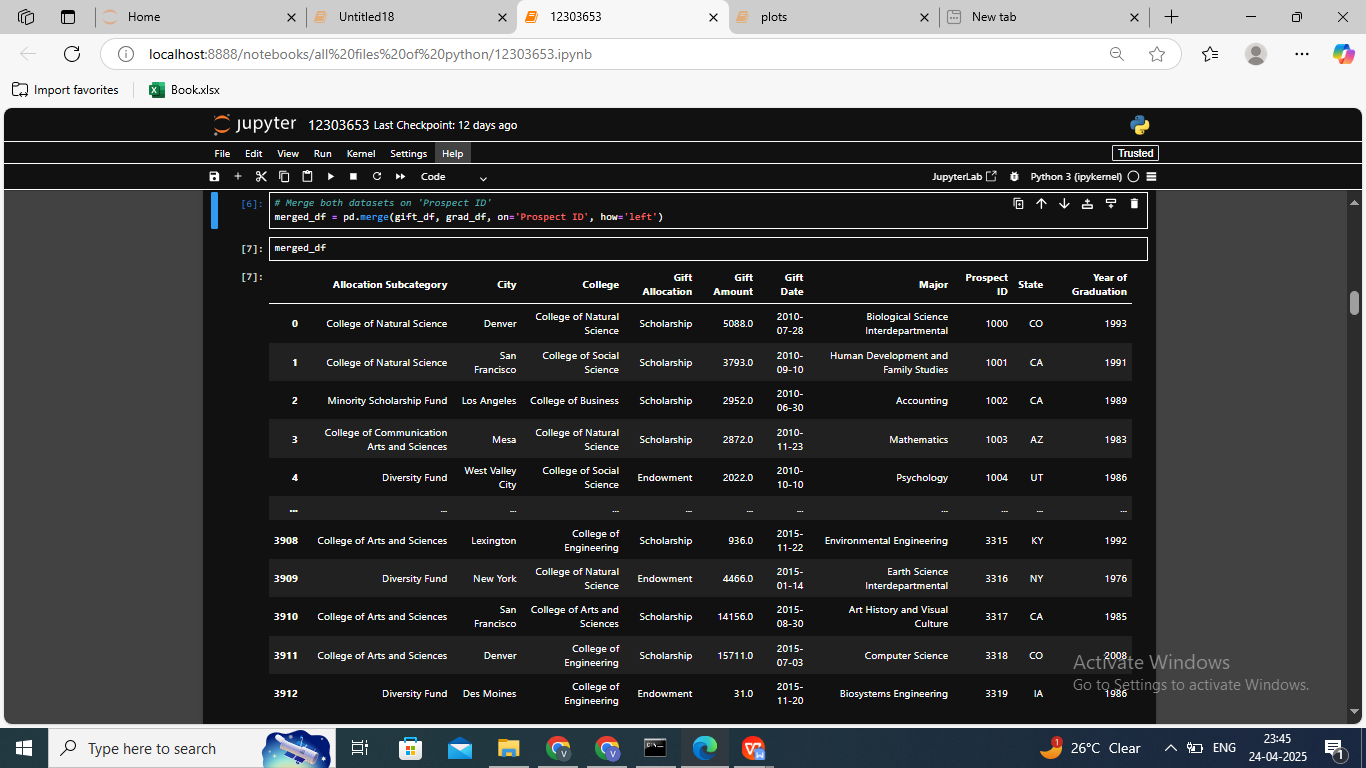
The dataset for this project was provided in the form of an Excel file comprising two sheets:

1. **GiftRecords**: Contains detailed donation records including gift amount, gift date, donor’s major, college, location, and allocation category.

**2. GraduationYear**: Contains the graduation year data for the respective donors identified by Prospect ID.

The data serves as a realistic simulation of advancement office records, enabling meaningful analysis and visualization.





### OBJECTIVES

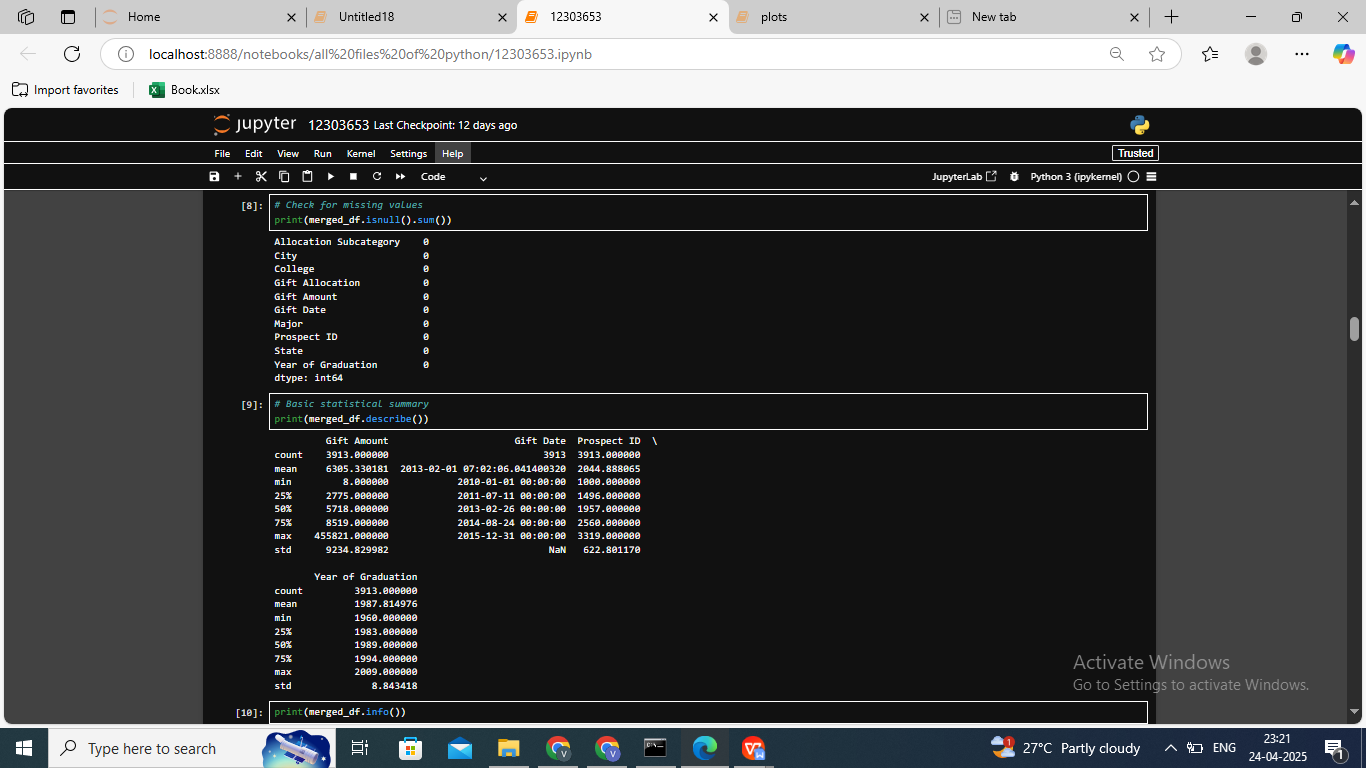
#### Objective 1: Data Cleaning and Preparation

**Purpose:**  
To ensure the dataset is free from errors, inconsistencies, or irrelevant values before any analysis.

**Explanation:**  
Raw data often comes with missing values, duplicate records, inconsistent formats (like date formats or capitalizations), or irrelevant outliers that can skew results. In this project, data from two sheets — GiftRecords and GraduationYear — were cleaned and then merged using a common column, Prospect ID.

**Steps Taken:**

* Checked for missing/null values using .isnull().sum().
* Verified and corrected data types using .info() (e.g., ensuring dates were in datetime format).
* Removed unnecessary spaces and standardized text columns.
* Merged datasets to enrich records with graduation years.

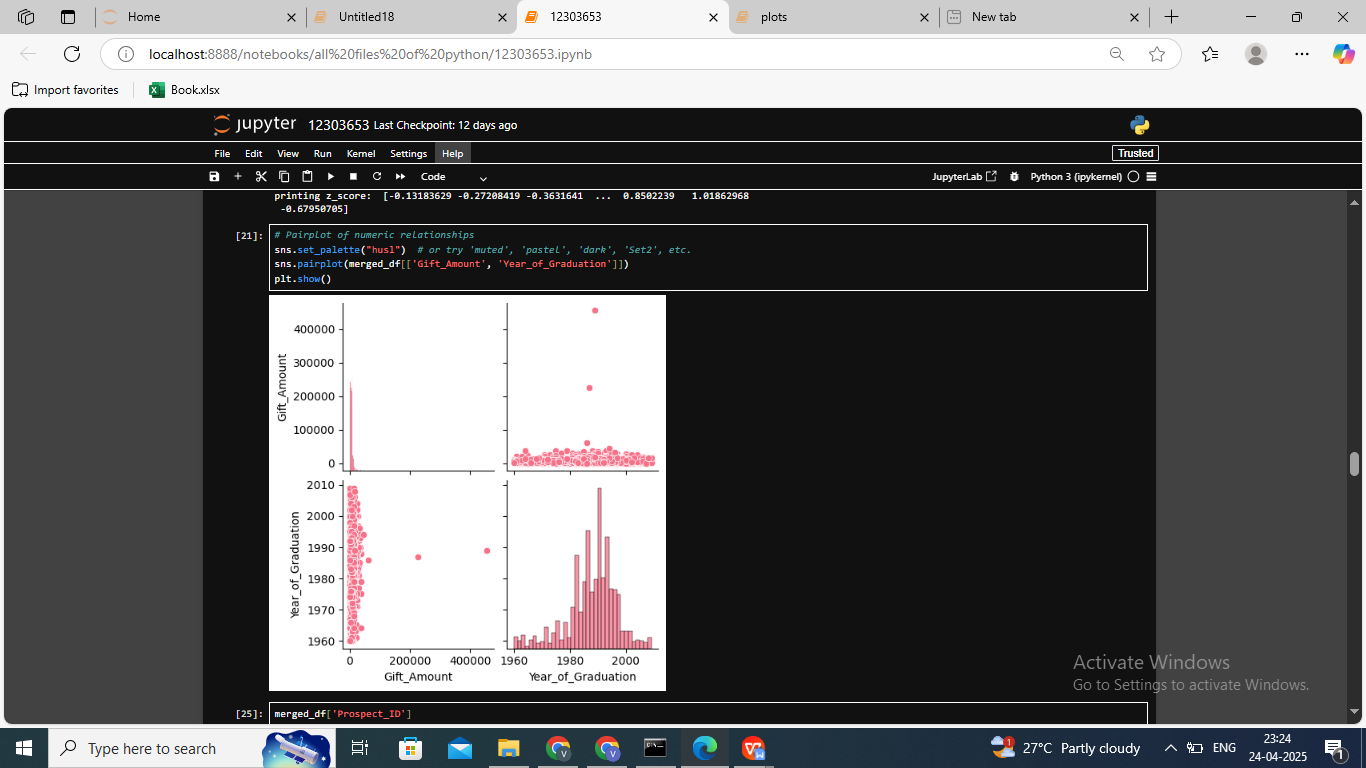


**Why It Matters:**  
Accurate and clean data ensures reliable results in analysis and visualizations.

#### Objective 2: Exploratory Data Analysis (EDA)

**Purpose:**  
To explore the dataset through descriptive statistics and visual summaries to uncover patterns, anomalies, or relationships.

**Explanation:**  
EDA is the foundation of any data science project. Here, we used Python libraries like pandas, numpy, matplotlib, and seaborn to analyze the data. We computed key statistics such as mean, median, standard deviation, etc., and visualized distributions and correlations.

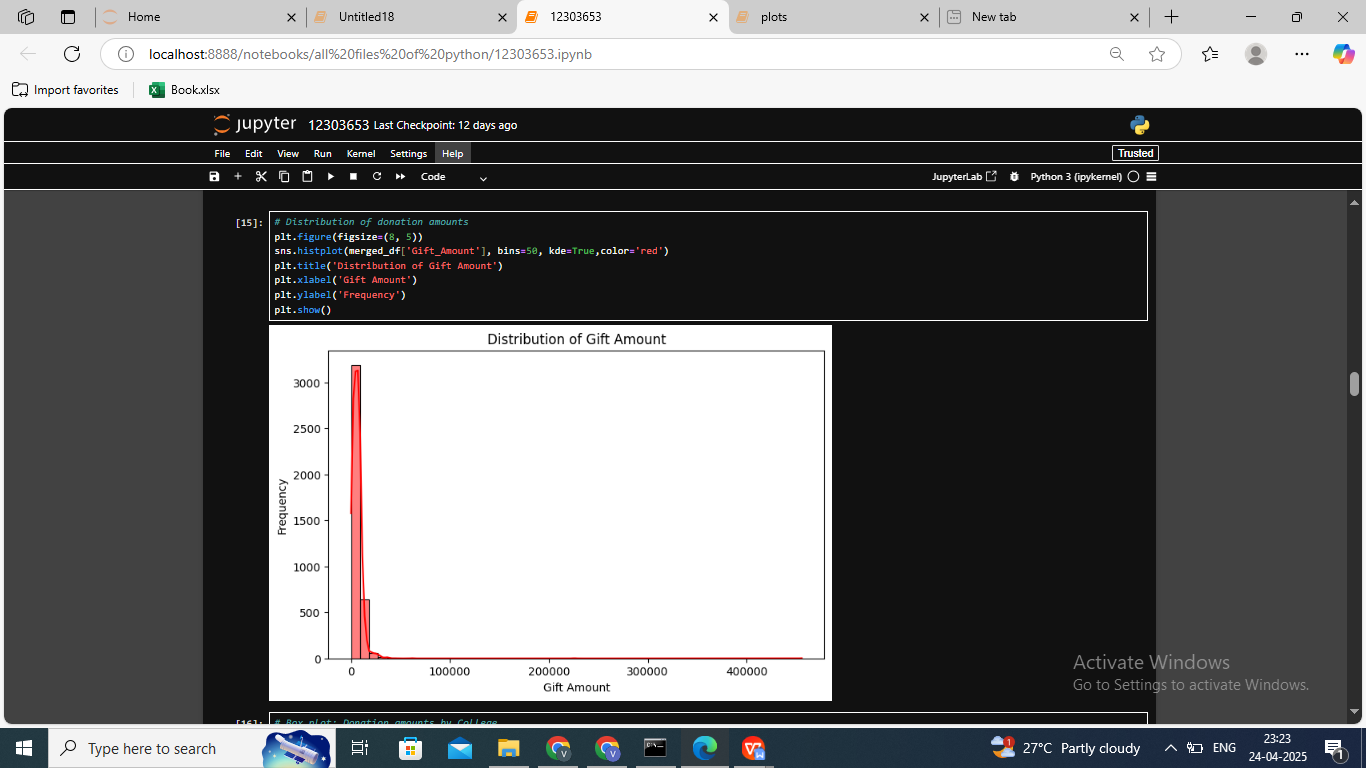


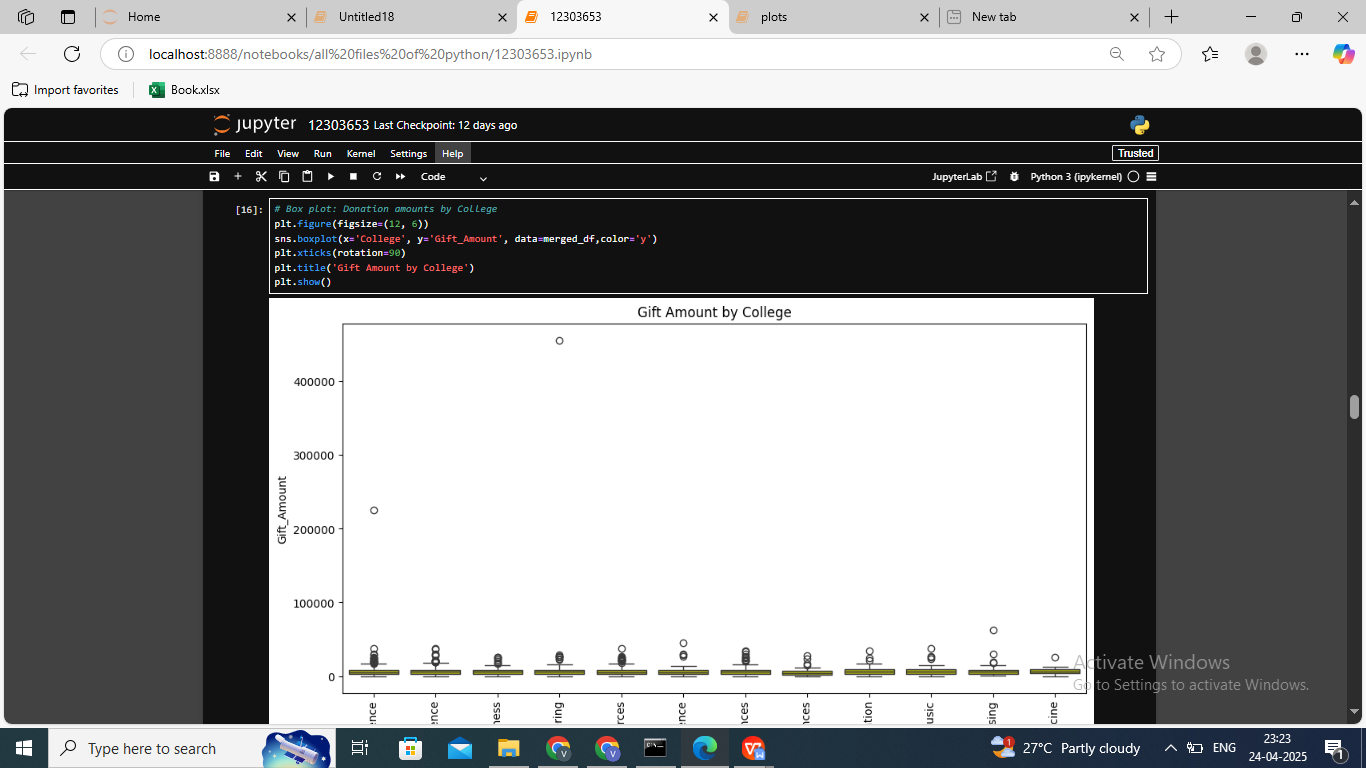
**Key Insights Explored:**

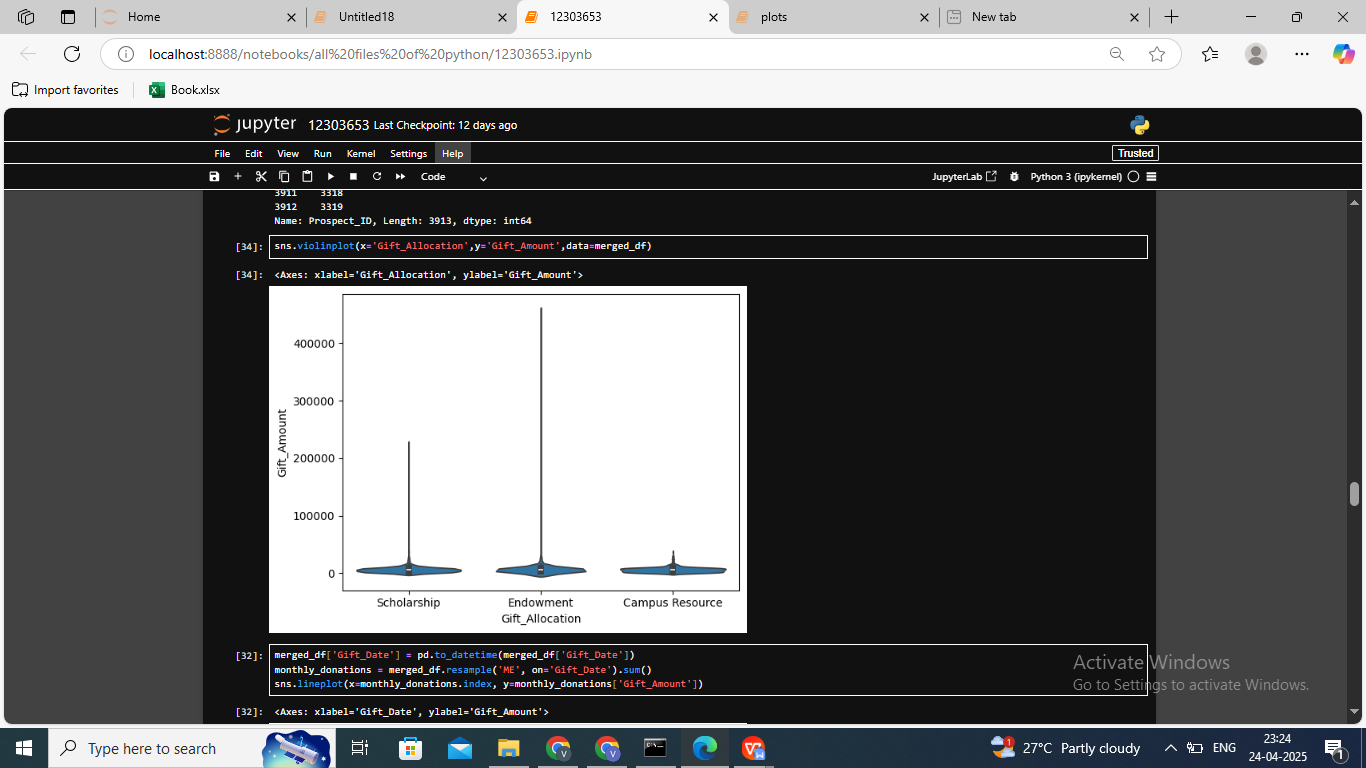
* What is the average donation amount?
* Which years had the most donations?
* Which colleges and majors are the most generous?

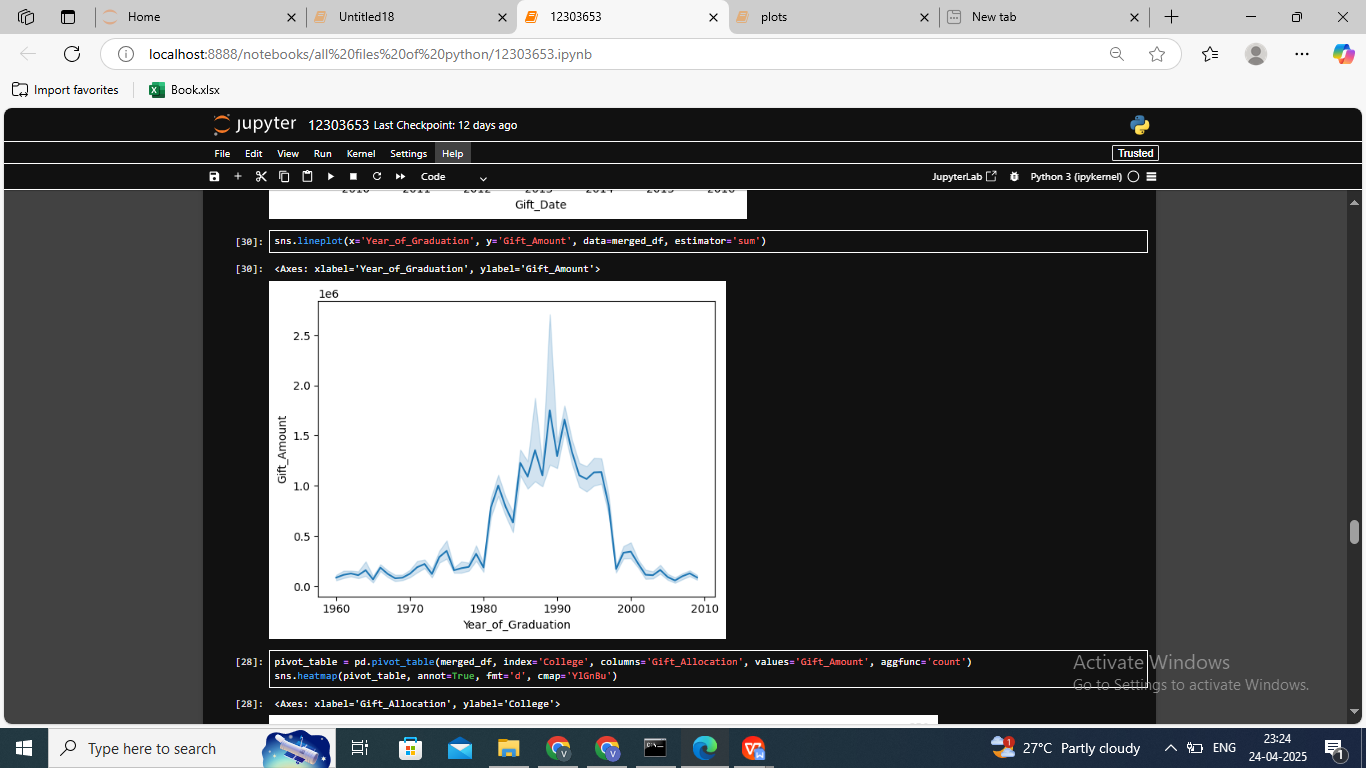
**Visual Techniques Used:**

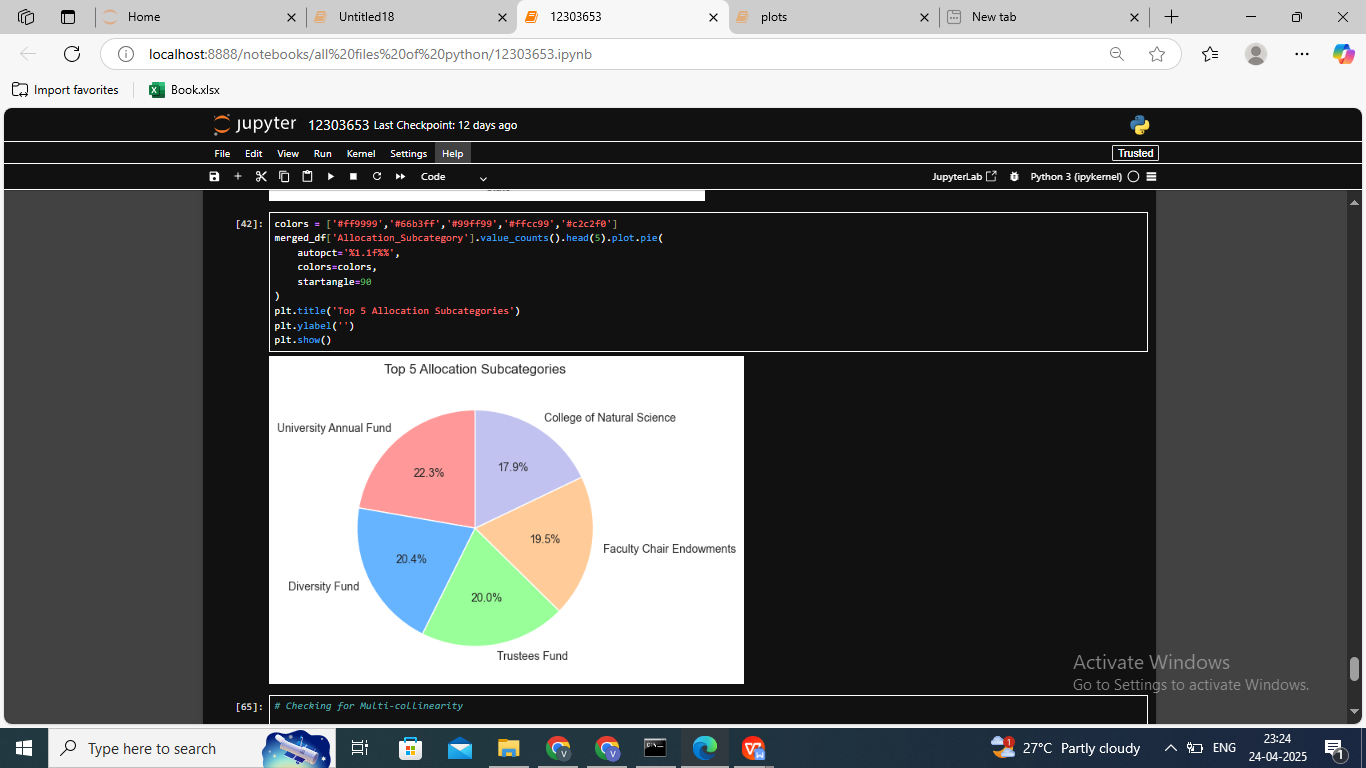
* **Box plots** to detect outliers and compare gift amounts across categories.
* **Pair plots** for exploring variable relationships.
* **Histograms and count plots** for frequency distribution.
* **Heatmaps and correlation matrices** to identify relationships.
* **Covariance analysis** to measure joint variability.











**Why It Matters:**  
EDA provides a roadmap of where to focus deeper analysis and helps identify hidden patterns and potential opportunities for university fundraising strategies.

#### Objective 3: Geographical Analysis of Donations

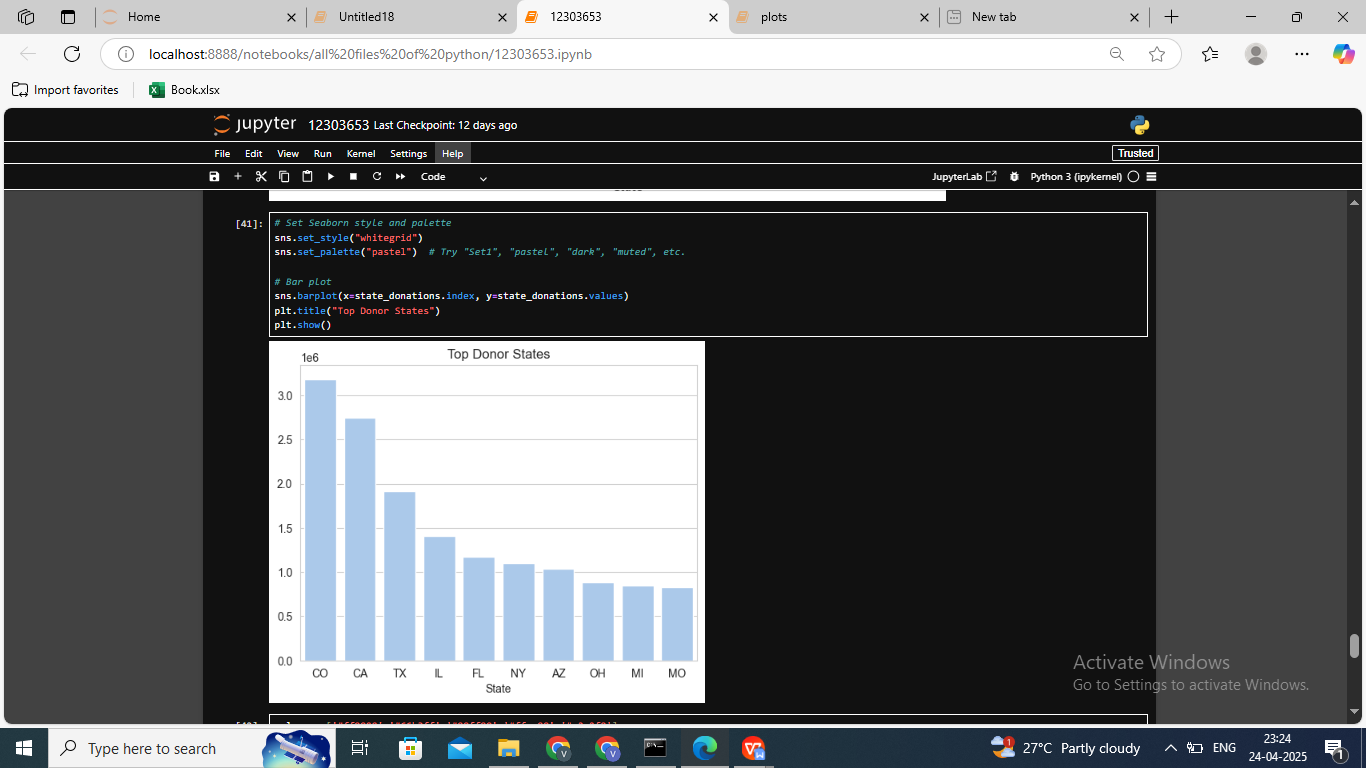
**Purpose:**  
To analyze where alumni donations are coming from geographically (City & State).

**Explanation:**  
By examining the City and State columns, we created visualizations that show donation amounts by location. This includes:

* Bar plots for total donation amount per state.
* Identifying which cities and states contribute the most.

**Findings:**

* States like California, New York, and Texas ranked highest in total donations.
* Certain cities emerged as consistent donors, possibly due to higher alumni concentrations.



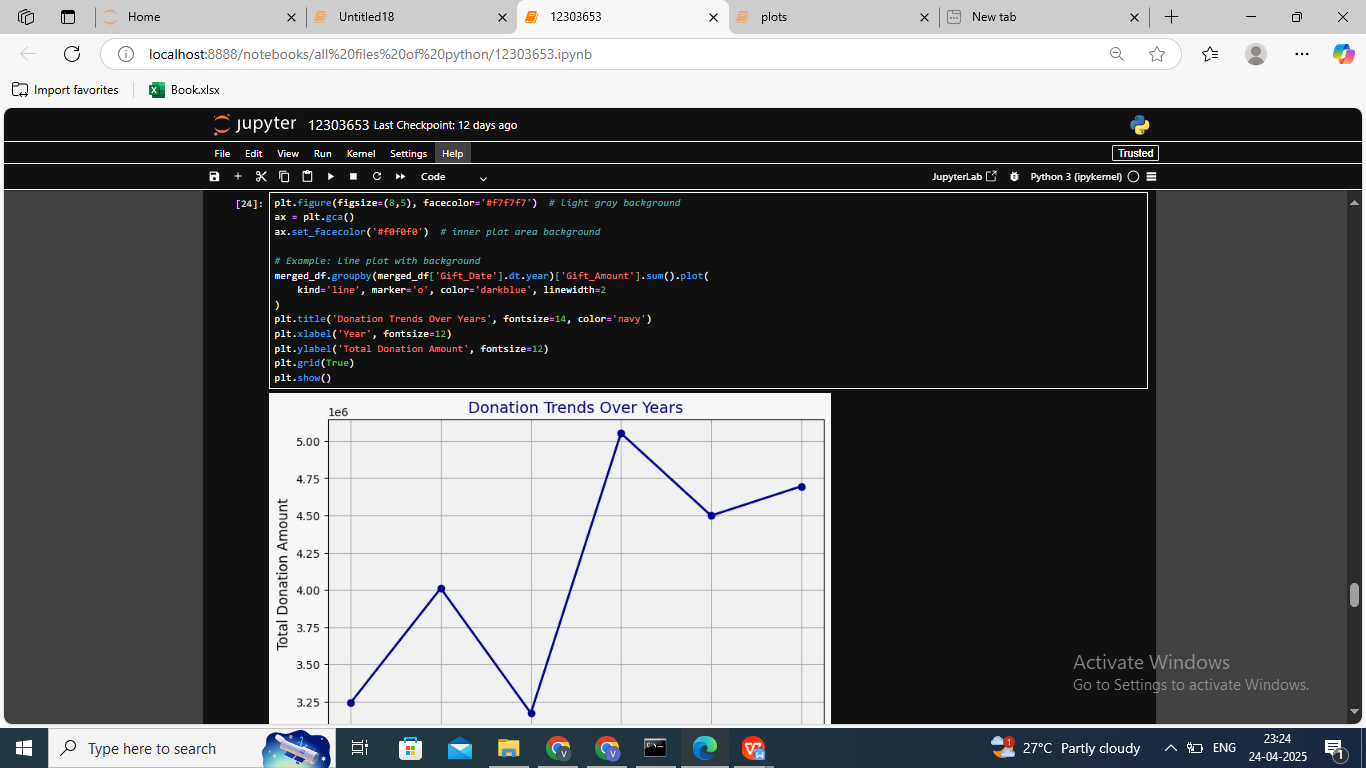
**Why It Matters:**  
Helps advancement teams focus fundraising efforts in high-yield regions.

#### Objective 4: Donation Trends by Academic Background

**Purpose:**  
To identify how educational background influences donation behavior.

**Explanation:**  
We analyzed donations based on the donor’s:

* **College** (e.g., College of Business)
* **Major** (e.g., Psychology, Computer Science)
* **Graduation Year**



Using groupby() operations and visualizations, we answered questions like:

* Do Engineering graduates donate more?
* Is there a trend showing older graduates donate more than recent ones?

**Why It Matters:**  
Universities can personalize outreach strategies for departments with higher or lower alumni engagement.

#### Objective 5: Time Series and Correlation Insights

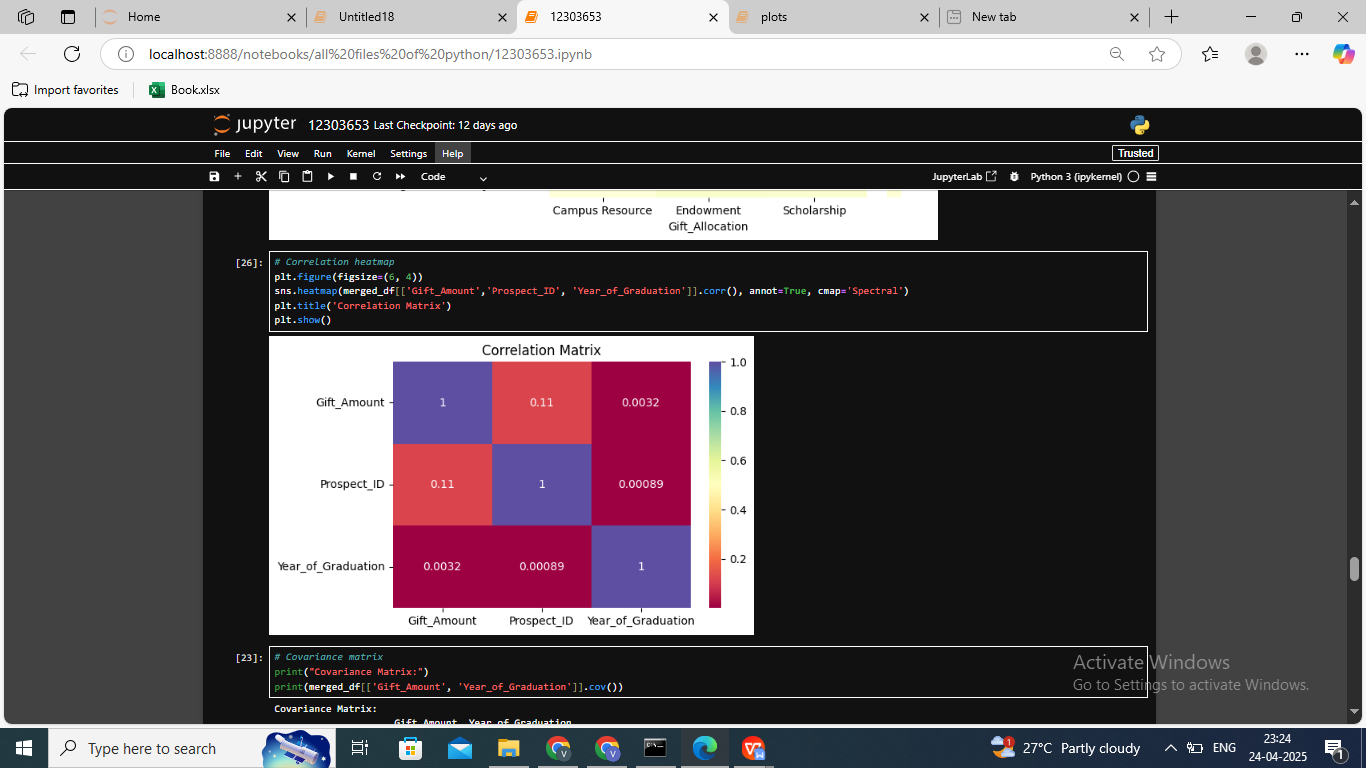
**Purpose:**  
To explore trends over time and understand how different variables relate to each other.

**Explanation:**  
We created **time-series plots** using the Gift Date to observe how donation patterns evolved over months and years. Correlation matrices were used to understand the strength of relationships between numerical variables, such as:

* **Gift Amount vs Graduation Year**
* **Gift Amount vs Frequency of donations**

**Visual Tools Used:**

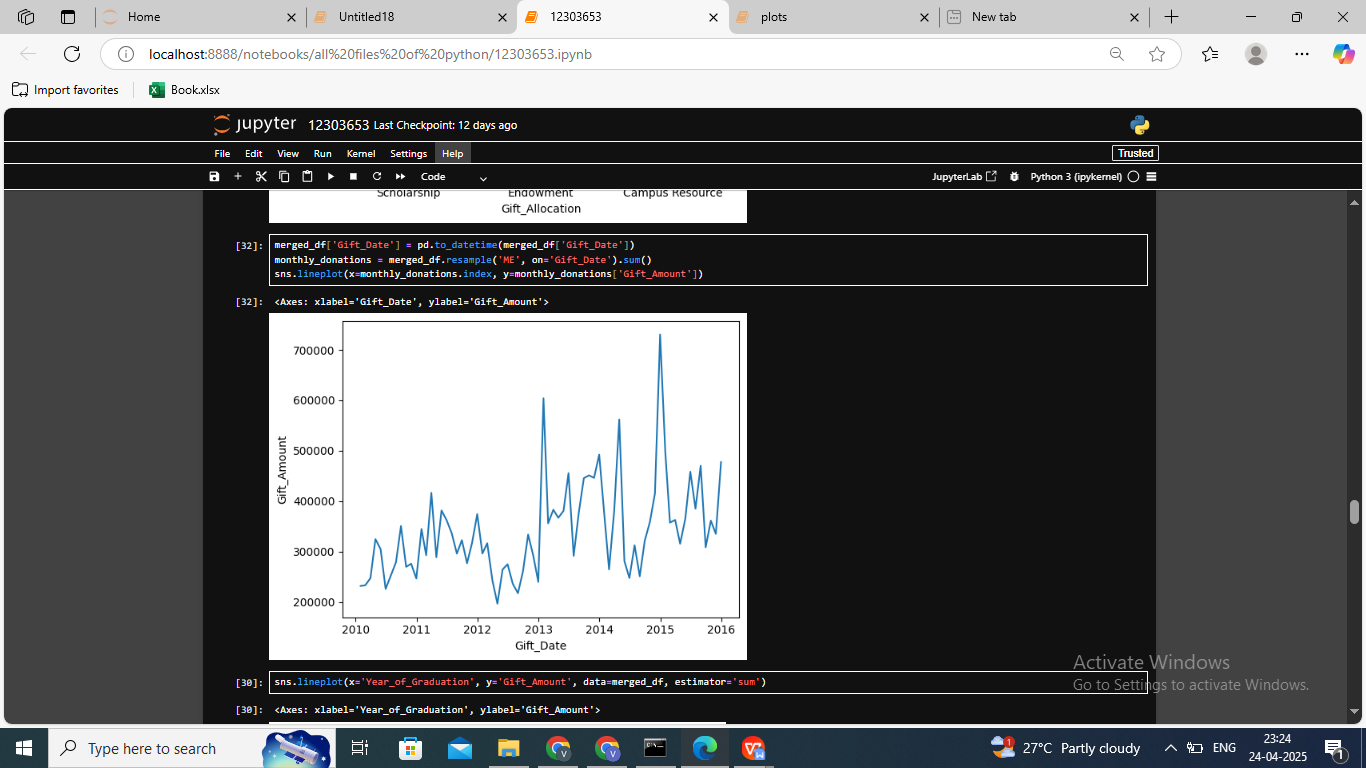
* Line charts for donation trends over time
* Heatmaps for correlation analysis



**Why It Matters:**  
Understanding time-based trends helps in planning seasonal fundraising campaigns. Correlations can identify which alumni traits are linked to higher giving.---

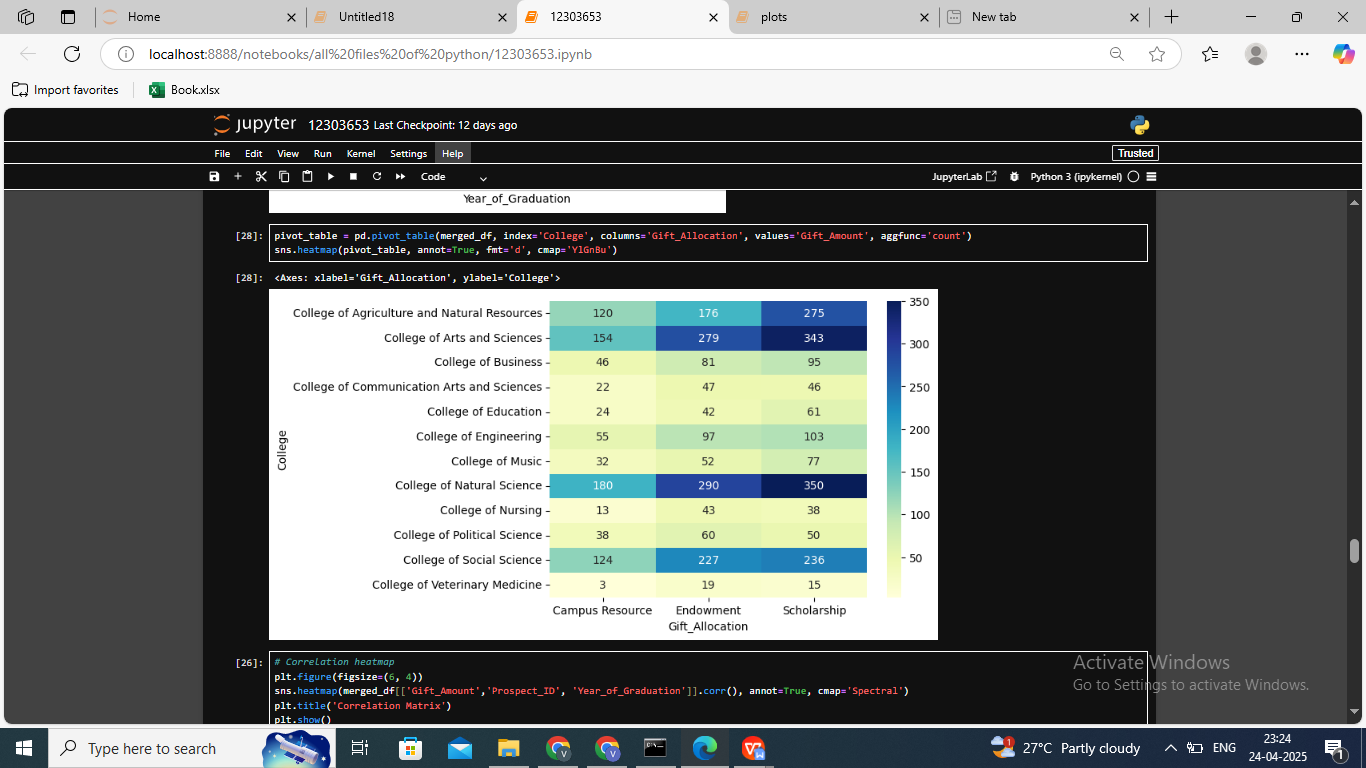
### **EDA PROCESS**

The Exploratory Data Analysis (EDA) was performed using Python with the help of libraries like pandas, numpy, matplotlib, and seaborn. The process included multiple stages:



* **Data loading and merging:** Combined the donation and graduation datasets using Prospect ID.
* **Missing value detection:** Applied .isnull().sum() and .info() to assess data quality and types.
* **Summary statistics:** Used .describe() to identify central tendencies and variability.
* **Box Plots:** Implemented to detect outliers in Gift Amount and compare distributions across colleges.
* **Pair Plots:** Utilized to study the relationships between multiple numerical features at once.
* **Heatmaps:** Created to visualize the correlation matrix, identifying the strength of linear relationships between variables.
* **Correlation & Covariance:** Used .corr() and .cov() to study relationships and how data variables vary together.
* **Time series analysis:** Examined trends using line graphs of donation frequency and amounts over time.
* **Categorical distributions:** Applied count plots and bar charts for visualizing state-wise and major-wise participation.

This multi-faceted approach enabled us to uncover patterns not immediately obvious from raw data, laying the foundation for the insights discussed in the analysis section.



### ANALYSIS ON DATASET

#### General Description

The dataset comprises a total of **3913 donation records** and **2317 graduation year entries**, tied together through a common Prospect ID. Each entry provides essential data about the donor, including their college, major, city, state, and gift details such as date, category, and amount.

After preprocessing and merging, the dataset allowed for in-depth analysis into the behavioral and demographic aspects of alumni donors.

#### Specific Requirements, Functions and Formulas

Key functions and techniques applied during analysis included:

* groupby() for aggregating data by College, State, and Major
* sort\_values() to rank donors and regions
* .corr() and .cov() to identify statistical relationships
* pd.merge() for combining Gift and Graduation data
* Visualizations via matplotlib.pyplot and seaborn for all trend analyses

We also made use of filters, masks, and conditional logic to examine data subsets (e.g., donors after 2000, high-value donations, etc.).

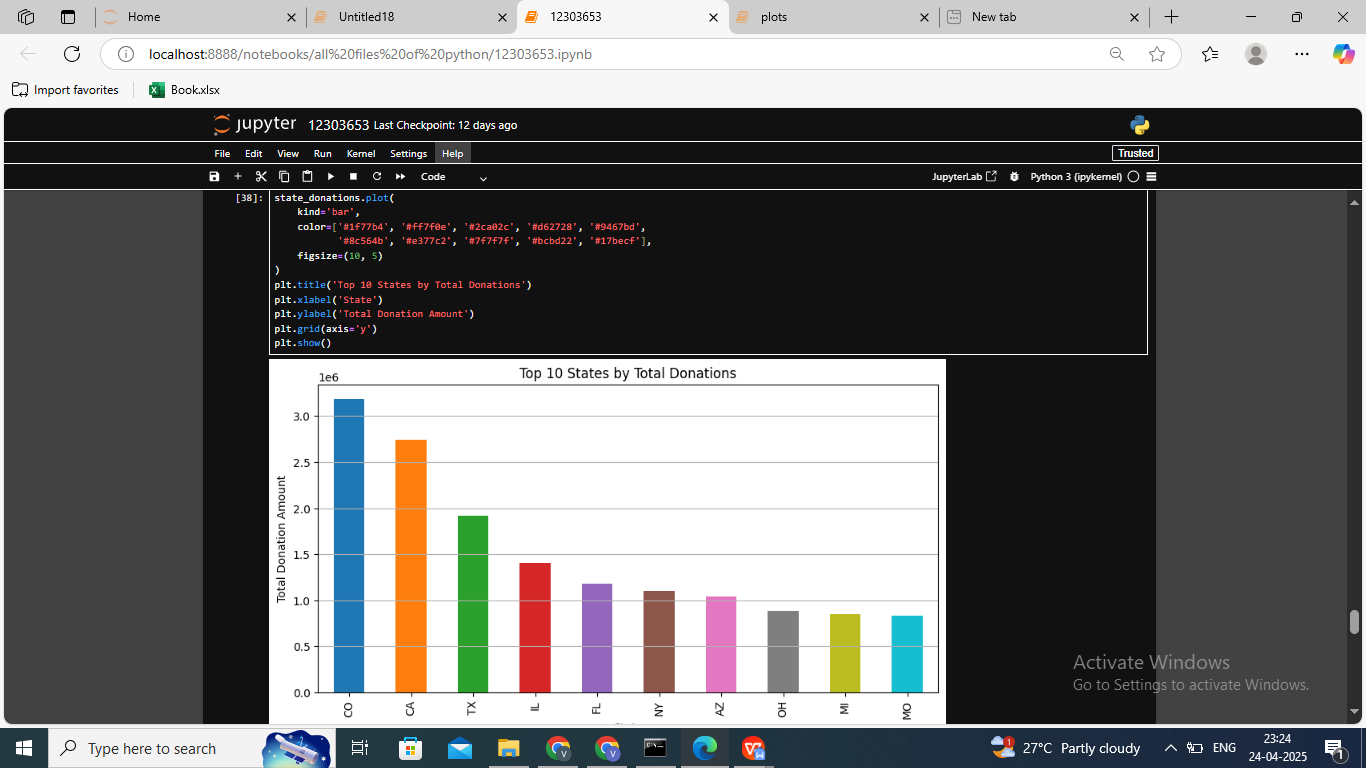
#### Analysis Results

* Some of the key insights discovered include:
* **Engineering and Business majors** represent the largest volume of donations.
* **California, Texas, and New York** are consistently among the top three donor states.
* The **most common donation type** was "Scholarship Fund" under the "Gift Allocation" column.
* Alumni graduating between **1980–2000** showed higher average donation values.
* Time series analysis revealed a **seasonal spike in donations**, particularly during the last quarter of each year.

#### Visualizations

A series of visualizations were generated to support the findings:

* **Bar Chart**: Showing total gift amount by college
* **Line Graph**: Displaying trends in donations over time (year-wise)
* **Box Plot**: To identify outliers and compare gift amount distributions by major
* **Heatmap**: Illustrating correlations between variables like Gift Amount and Graduation Year
* **Pie Chart**: Showing allocation subcategories proportion (e.g., Scholarships, Endowments, Research)



### **CONCLUSION**

This project provided meaningful insights into alumni giving behavior using data analytics. We examined multiple dimensions, including academic background, geography, and graduation year to uncover patterns in alumni donation trends. The use of visualizations and statistical analysis helped simplify complex patterns and enhance interpretability.

Our key findings showed that Engineering and Business graduates contribute the highest, donors from specific states are more engaged, and donations are most frequently directed toward scholarships. The data also revealed cyclical trends in donations, which can help institutions in their fundraising strategies.

### FUTURE SCOPE

While the analysis yielded valuable insights, there is ample scope for future improvements:

* Inclusion of demographic data (age, gender, income) to enhance analysis.
* Application of machine learning models to **predict future donations**.
* Deployment of interactive dashboards using **Power BI** or **Tableau**.
* Integration with real-time CRM systems to enable **automated segmentation and targeting**.
* Analysis of **alumni engagement beyond donations**, such as event attendance or volunteering.

These enhancements would make the system more dynamic and directly applicable to university advancement departments.

### REFERENCES

[1] M. Waskom, "Seaborn: statistical data visualization," Journal of Open Source Software, vol. 6, no. 60, p. 3021, 2021.  
[2] W. McKinney, "Data Structures for Statistical Computing in Python," in Proceedings of the 9th Python in Science Conference, 2010.  
[3] J. D. Hunter, "Matplotlib: A 2D Graphics Environment," Computing in Science & Engineering, vol. 9, no. 3, pp. 90-95, 2007.  
[4] Python Software Foundation. (2024). Pandas Documentation. Retrieved from <https://pandas.pydata.org>  
[5] Van Rossum, G., & Drake, F. L. (2009). Python 3 Reference Manual. CreateSpace