Analysis of Space Missions and Budgets of Space Agencies

Author: Srinath Begudem

Objective: The objective of this project is to gather data from diverse and intricate space-related websites to construct a comprehensive space mission DataFrame. Additionally, the project aims to extract budget information from the official websites of NASA and ISRO. The overarching goal is to analyze and visualize the budget allocations, success rates, and number of missions conducted by each space agency, while also examining the current status of their rockets. Furthermore, the project seeks to conduct a detailed analysis of the budgets of the top two space agencies, ISRO and NASA, and ultimately apply machine learning techniques to predict their future budget trends.

Datasets:

- The primary data set on space missions was obtained from NextSpaceFlight, a non-tabular data website with a complex structure of 225 sub-pages. It provided comprehensive information on past space missions, including launch date, status, and location. To handle this non-tabular data, I converted the text paragraphs into tabular data from scratch, utilizing Beautiful Soup for web scraping.
- Additional mission details for NASA were collected from relevant Wikipedia links, such as the Timeline of space exploration page, to fill in any missing information.
- Budget details for NASA and ISRO were sourced from Wikipedia articles on their budgets, with NASA's data from the Budget of NASA page and ISRO's from the ISRO Statistics page.
- RSA budget data, covering 2005 to 2021, was obtained from The Space Report.
- This comprehensive approach ensures a robust dataset for analyzing space missions, including mission details, budget allocations, and historical trends.

Integration of Data and Key Questions: The datasets are integrated based on the year of launch and the space agency columns, with the year serving as the primary key. While additional columns may be used for further analyses, the year remains the central integration factor.

Key questions to be explored include:

- **Evolution of Space Missions**: Explore the temporal trends of space missions for each space agency.
- **Correlation between Budget and Missions**: Investigate the relationship between budget allocations and the number of missions launched by each agency.
- **Analysis of Mission Types**: Identify any patterns in the types of missions conducted by each agency over time.
- **Predictive Analysis of Budgets**: Assess the feasibility of using budget data to predict future allocations for space agencies.

Methods and Expected Outcome: The project utilizes web scraping with Beautiful Soup and requests for data collection, alongside pandas, NumPy, and advanced data cleaning methods for manipulation. Visualization techniques such as plots and graphs are employed, and simple linear regression analysis is applied to predict future budget allocations. The expected outcome is to gain insights into space mission trends and agency budgets, contributing to discussions on space research while offering predictions for future budget trends.