# Analysis of Space Missions and Budgets of Space Agencies

**Author: Srinath Begudem**

## Motivation/Rationale

**Objective**: The objective of this project is to gather data from diverse and intricate space-related websites to construct a comprehensive space mission Data Frame. 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.

**Key questions and interesting facts 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.

Understanding the allocation of budgets, success rates, and mission focus of space agencies is crucial as humanity continues to explore the vastness of space. This project's exploration into the intricate details of space missions and budgetary allocations sheds light on how agencies prioritize their resources in pursuit of their goals. Space exploration has captured the imagination of our generation, driven by the quest to discover what lies beyond our observable universe and to potentially find life or new resources on other planets. Such discoveries could have profound implications for the future of humanity.

The questions proposed in this study are particularly interesting because they provide insights into the operational efficiencies and strategic directions of space agencies. By analyzing the relationship between budget allocations and mission outcomes, we can identify which agencies are achieving high success rates and how they manage their investments in technology and human capital. This analysis is vital for understanding which strategies contribute to successful missions and how financial planning influences these outcomes.

Moreover, by comparing the long-term planning and strategic objectives across different agencies, the study reveals trends in budgetary shifts and their alignment with global space exploration goals. It also highlights the impact of technological advancements on the cost and frequency of missions, offering a view into how innovation drives the industry forward.

The project's visualizations illustrate these dynamics, showing changes in budget allocations over time, mission success rates, and the evolving landscape of international collaboration. These visualizations not only enhance our understanding of historical data but also help predict future trends in space exploration.

In conclusion, this analysis is not just about fiscal figures and mission counts; it is about understanding the broader implications of space exploration for society. It informs stakeholders and the public about how effectively each dollar is utilized towards pushing the boundaries of human knowledge and capability in space. It also provides a framework for future studies on the sustainability and policy impacts of space missions, thereby contributing to a more informed approach to global space governance.

## Description of data sources:

### Datasets and challenges faced:

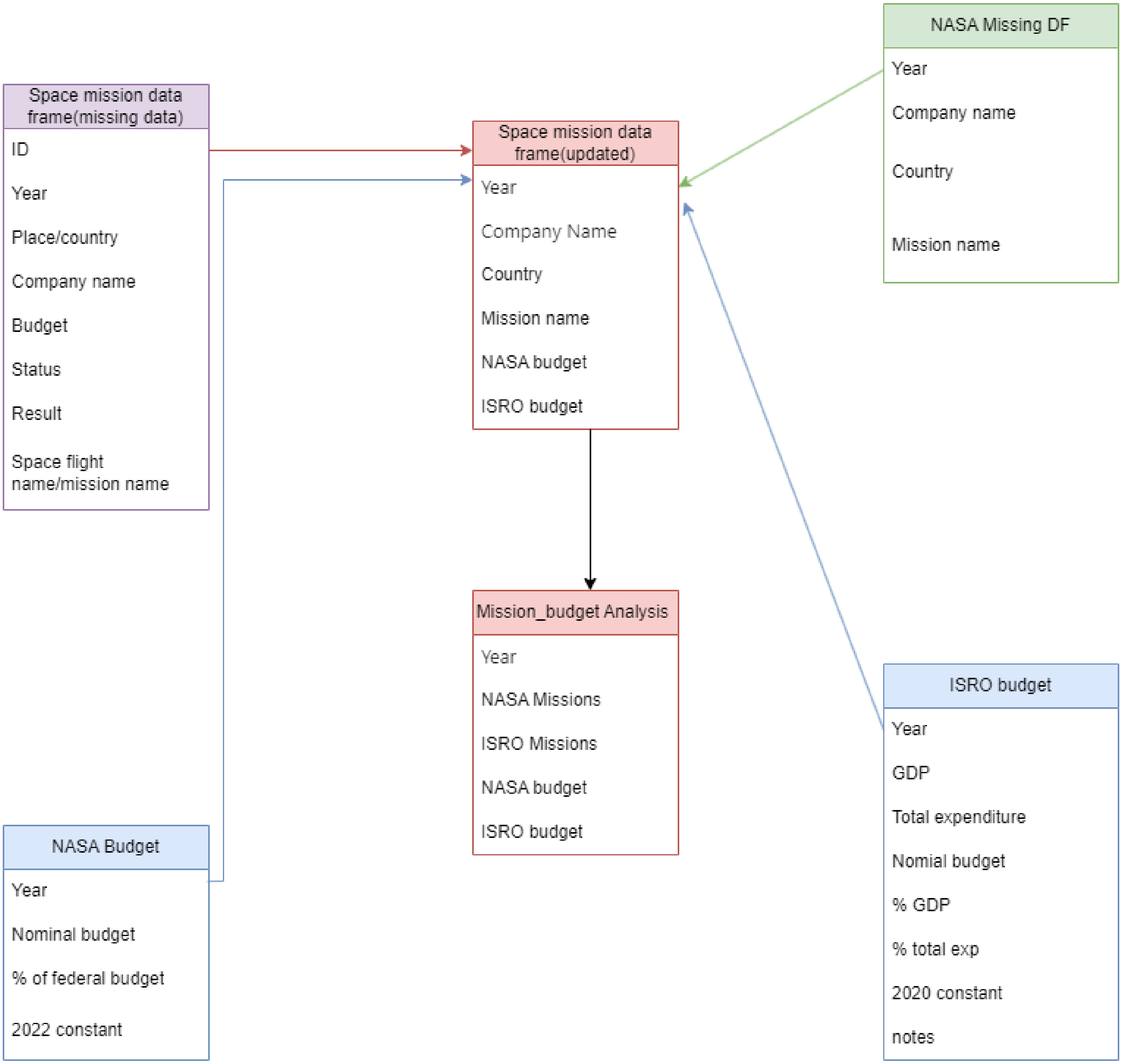
• The core dataset on space missions was meticulously extracted from NextSpaceFlight, a resource with a non-tabular format spread over 226 sub-pages, each rich with detailed information. The complexity of the website architecture, which included nested child pages within each main page, posed significant challenges. Utilizing Python's Beautiful Soup library, I conducted an intensive web scraping operation that lasted approximately 5-6 hours. This process involved not only harvesting data from the primary pages but also diving into the secondary pages linked within them, to gather comprehensive details on launch dates, mission statuses, budgets, success rates, and locations. The dual-level scraping was computationally demanding and required careful handling to maintain data integrity and structure.

• To ensure the completeness of our dataset, particularly regarding NASA's recent mission activities, we supplemented our primary data by incorporating additional mission details from 2010 to the present. This information was sourced from a series of Wikipedia pages, most notably the "Timeline of Space Exploration." These pages provided essential data on missions that were not initially covered in our primary source, filling gaps and enhancing the robustness of our analysis. The integration of these details was crucial for maintaining the accuracy and comprehensiveness of our study on space missions, enabling a more nuanced understanding of NASA's activities over the past decade. This approach not only strengthened our dataset but also ensured that our analyses reflected the most current and complete information available.

• In our project, a significant emphasis was placed on analyzing the annual budgetary allocations of NASA, one of the world's leading space agencies, alongside ISRO, known for its cost-effective and successful missions. To facilitate this analysis, detailed budget information for both agencies was meticulously extracted from Wikipedia. Specifically, data concerning NASA's financial allocations were sourced from the "Budget of NASA" page, while budgetary details for ISRO were drawn from the "ISRO Statistics" page. This extraction process was critical in providing a comprehensive understanding of the financial trends and resource allocations that support the varied and ambitious space missions undertaken by these two prominent space entities.

## Integrated Data Model

## ER Diagram




The data integration for this project was structured to create a comprehensive and cohesive model, enabling detailed analyses of space missions and financial trends across different space agencies. The integration process utilized a relational approach with "Year" serving as the primary key, aligning all datasets chronologically.

1. **Space Mission Data Frame (Updated):** This central dataset amalgamates comprehensive mission details including company names, countries, specific mission names, and budget information. The data was initially extracted from Next Spaceflight and enriched with additional mission and budget details from Wikipedia and other credible sources.
2. **NASA Missing Data Frame:** To rectify the absence of recent NASA mission data, particularly from 2010 onward, this dataset was incorporated. It includes data fields such as year, company name, country, and mission name, which were meticulously sourced from Wikipedia's detailed records of space exploration events.

3. **NASA and ISRO Budget Data Frames:** These datasets provide financial details for NASA and ISRO, extracted from Wikipedia. They encompass nominal budget figures, inflation-adjusted values, and budgetary data as percentages of national expenditures, crucial for assessing financial commitments over the years.

The integration process involved the following steps:

- **Merging Data on 'Year':** By merging data based on the 'Year' column, we ensured that mission data were directly comparable with financial data across the same timeline, facilitating trend analysis and correlation studies between mission activities and budgetary allocations.
- **Filling Data Gaps:** Where primary data sources were incomplete, especially for recent years, supplementary data from additional sources such as Wikipedia were integrated to provide a more complete overview.
- **Data Normalization and Adjustment:** To ensure comparability and accuracy in analysis, all financial figures were normalized to a consistent currency format, and inflation adjustments were applied where necessary. This standardization was crucial for the analytical modeling and predictive analyses conducted later in the project.

This integrated model provided a robust foundation for analyzing the dynamics of space missions in conjunction with financial data, enabling a comprehensive exploration of how budget allocations impact mission planning and execution across key space agencies like NASA and ISRO.

## Analyses/Visualizations

### Data Analysis

**Descriptive Statistics and Data Distribution:**

1. **Number of Space Flights Each Year**:
   - **Method**: Utilized the **value_counts()** function to tally space flights per year, followed by **sort_index()** to organize the data chronologically.
   - **Visualization**: A simple output displaying the count of space flights for each year.
   - **Purpose**: This analysis helps to identify trends in space mission frequency over the years, providing a clear view of how active certain periods were in terms of space exploration.

2. **Number of Space Flights by Company**:
   - **Method**: Again, **value_counts()** was applied to the 'company_name' column to count the number of missions per company, highlighting the most to least active companies in space missions.
   - **Visualization**: Presented as a straightforward list, this method highlights the companies with the highest number of space missions, indicating their significant roles in the aerospace industry.

3. **Average Budget for Space Flights Per Company**:
   - **Method**: Employed the **groupby()** function to group data by 'company_name', then calculated the mean of the 'budget (millions of dollars)' for each company and sorted the results in descending order to highlight companies with the highest average budgets.
   - **Visualization**: Displayed as a list, it shows average spending on space flights by each company, providing insights into financial investments relative to their number of missions.

**Budget Trends Analysis:**

4. **Trend of Budget Over the Years**:

- **Method**: Used **groupby()** to group the data by 'Year' and then calculated the mean of 'budget (millions of dollars)' to understand the average budget per year.
- **Visualization**: The result is presented as a simple list showing the average budget for each year.
- **Purpose**: This helps in analyzing how budget allocations for space missions have fluctuated over the years, offering insights into possible economic impacts or shifts in space exploration priorities.

These methodologies and their corresponding outputs provide foundational insights into the operational and financial aspects of space missions conducted by various agencies. The straightforward presentation of data through basic counts, averages, and lists serves as a preliminary step in identifying key trends and patterns that merit further exploration through more complex analyses and visualizations.

## Statistical Analysis

For the statistical analysis of the project's data, I employed descriptive statistics and distribution analysis to understand the variability and central tendency of the space missions' budget data.

1. **Summary Statistics**: I began by calculating the summary statistics for the budget data using the **.describe()** method in pandas. This provided a comprehensive view of the budget distribution, including the mean, standard deviation, minimum, maximum, and quartile values. Such statistics are crucial for understanding the scale of investment in space missions and the variation across different missions and agencies.
2. **Skewness and Kurtosis**: To further understand the distribution characteristics of the budget data, I calculated the skewness and kurtosis. Skewness measures the asymmetry of the data distribution around its mean, and kurtosis indicates the tail's heaviness. These metrics help in identifying the outliers and the peak of the data distribution, which are essential for normalizing data before any advanced statistical analysis or machine learning application.
3. **Correlation Matrix**: I also generated a correlation matrix to explore the relationships between the numeric variables, particularly between the year and budget. This analysis helps in identifying trends over time and the influence of time on budget allocations.
4. **Yearly Budget Variation Analysis**: By grouping the budget data by year and calculating descriptive statistics for each group, I provided insights into how the budget distribution has evolved annually. This is particularly useful for spotting trends in budget increases or decreases and planning future budget allocations based on past patterns.

These statistical methods enabled a deep dive into the budgetary aspects of space missions, facilitating a thorough understanding of how financial resources are allocated and how they fluctuate over time. Each of these analyses provided a foundational understanding necessary for more complex predictive modeling and trend analysis.

## Data Visualization

For the visual representation of our dataset, various graphical methods were employed to enhance the understanding and interpretation of the data. Here’s a summary of the techniques used and the insights they provided:

1. **Distribution of Space Flights Over the Years**: Using seaborn's **countplot**, we visualized the distribution of space flights over the years, which clearly displayed the evolution and growth in space mission launches over time. This visualization was vital in showing trends and peak periods of activity, using a vibrant **viridis** color palette to enhance visual perception.

2. **Top 20 Companies by Space Flights**: To understand the key players in space exploration, a seaborn **countplot** was utilized to rank the top 20 companies based on the number of space flights. This plot highlighted dominant space agencies and commercial entities, such as SpaceX and Roscosmos, allowing for an analysis of the competitive landscape in space mission launches.
3. **Animated Bar Charts using Plotly**: Advanced visualizations included animated bar charts created with Plotly, which provided an interactive view of space flights and budget allocations over the years for the top companies. These animations offered a dynamic way to observe changes over time and engage more deeply with the data. The bar charts were oriented horizontally and updated through the years, reflecting changes in the number of flights and budget variations in a compelling narrative form.
4. **Trend Analysis of Space Flight Budgets**: For a deeper financial analysis, a line plot was crafted to illustrate the trend of average budgets over the years. This visualization, made with matplotlib, used a red color to denote fluctuations, thereby facilitating the identification of years with significant budget increases or cuts.
5. **Success Rate of Space Flights by Company**: The success rate of space missions was crucial for assessing operational efficiency and reliability. A bar chart represented the success rates across different companies, which was essential for understanding which companies had higher operational success and could be considered more reliable or successful in their missions.

Further, we consolidated and cleaned all the gathered data to facilitate comprehensive joint analysis. We merged the datasets based on the year, organizing and refining them to enable extraction of valuable insights effectively.

## Joint Analysis

The integration and subsequent analysis of the data involved consolidating multiple sources and aligning them by the 'Year' of mission or budget data. This alignment facilitated a detailed exploration of the temporal patterns in space missions and budget allocations among the major space agencies—NASA, ISRO, and Roscosmos.

1. **Data Integration and Cleaning**: Initially, data from distinct sources were meticulously cleaned and merged based on the year. The merging process was crucial to provide a unified view that encompasses both the number of missions and the budgetary outlines of the agencies involved.
2. **Mission Analysis by Agency**: Using this integrated data, we created various visualizations to compare the total missions per year by each agency. The visualizations provided clear insights into how active each agency was in any given year and how their activities fluctuated over time. These analyses were visualized through line plots and bar charts, emphasizing the comparative mission counts across years.
3. **Automated Plotly Graphs**: For dynamic and interactive visual analysis, Plotly was employed to create animated graphs that showcase the trends over the years. These animations provide a more engaging way to see the progression of each space agency in terms of mission counts over the years.
4. **Statistical Distributions and Comparisons**: Further analysis included the use of boxplots to depict the distribution of missions across agencies, highlighting variations and outliers in mission activities. These plots are instrumental in identifying the consistency of mission launches and any significant deviations that might indicate special focus years or changes in agency strategies.

Overall, this joint analysis not only streamlined the vast amounts of data into coherent insights but also highlighted the operational trends of the major space agencies, thereby enabling a deeper understanding of their historical performance and strategic priorities.

## Budget Analysis: Mission Funding for ISRO and NASA (2004-2021)

In an in-depth analysis of the mission funding for ISRO and NASA from 2004 to 2021, significant insights were derived from the data which revealed trends in budget allocation in relation to the number of missions conducted by each agency.

**Key Observations:**

1. **NASA's Budget Trends**:
   - NASA's nominal budget has shown a steady increase over the years, reflecting a strong financial commitment towards space exploration.
   - Notably, in the years when NASA undertook more missions, there was generally an uptick in the budget, demonstrating a correlation between mission frequency and budget increments.
2. **ISRO's Budget Trends**:
   - ISRO's budget, while smaller in comparison to NASA's, also exhibited increases aligning with years of more intensive mission activity.
   - The budget allocations for ISRO show a significant upward trend as the number of missions increased, particularly in years like 2014 and 2015.

**Budget Utilization and Mission Execution:**

- Both agencies displayed a pattern where increased budgets were often correlated with years of higher mission activity. This suggests that both NASA and ISRO scale their budget allocations based on the planned mission load for each year.

**Comparative Analysis:**

- When comparing the two agencies, NASA consistently had a higher budget, which correlates with its broader mission scope and larger scale operations.
- The budgets for both agencies show peaks and troughs that align with key missions or program phases, underlining the strategic financial planning tailored to mission objectives.

**Efficiency Insights:**

- A box plot analysis of the budgets versus the total missions reveals the variability and distribution of spending per mission. NASA shows a wider distribution, indicating variability in mission costs, whereas ISRO's spending is more concentrated, suggesting a more uniform cost structure per mission.

This analysis not only underscores the financial strategies employed by each agency but also highlights the impact of budgeting on mission planning and execution. The correlation between budgets and missions provides an essential insight into the operational efficiencies and strategic priorities of these space agencies over the analyzed period.

## Machine Learning Predictions for Future Budgets and Missions of NASA and ISRO

Based on the machine learning analysis performed on the budgets and missions data of NASA and ISRO from 2004 to 2021, here are the key findings and predictions:

1. **ISRO Missions Prediction**:
   - The regression analysis predicts a slight increase in the number of ISRO's missions in the coming years. From 2022 onwards, the predictions suggest a steady number of missions per year, with a slight uptick observed as we move closer to 2025 and 2026.
2. **NASA Missions Prediction**:
   - For NASA, the trend in total missions per year is expected to remain relatively stable, with a very slight decline. This trend reflects a stabilization in NASA's operational mission planning.
3. **NASA Budget Prediction**:
   - NASA's budget predictions show a consistent upward trend, with the budget expected to increase annually. This reflects ongoing and future planned expansions in NASA's space exploration and research initiatives.
4. **ISRO Budget Prediction**:

- Similarly, ISRO’s budget is also predicted to increase steadily over the next few years. This aligns with India’s growing investment in space technology and its ambitions to expand its space capabilities.

5. **Future Budgets Overview**:

- By 2026, NASA’s budget is expected to reach approximately $23.9 billion, reflecting significant growth. In contrast, ISRO’s budget is projected to be around $1.73 billion, which, while much smaller than NASA’s, indicates substantial growth from its current levels.

These predictions are based on linear regression models that use past data to forecast future trends. However, it's important to note that these predictions could be influenced by numerous factors such as changes in governmental policies, technological advancements, and unexpected global economic conditions which are not accounted for in the models. Thus, while these forecasts provide a useful guideline, they should be considered with an understanding of their inherent uncertainties.

## Conclusions

Our extensive data analysis and machine learning predictions reveal that both NASA and ISRO are expected to experience an increase in their budgets over the next five years, reflecting a global escalation in space exploration investments. Specifically, ISRO is predicted to slightly increase its mission counts, suggesting an emphasis on expanding its space activities. In contrast, NASA's mission trends do not show significant changes, potentially indicating a focus on more substantial, cost-intensive projects rather than a higher frequency of missions. This strategic difference highlights ISRO's predictable growth in missions, whereas NASA prioritizes impactful projects, which may not follow a predictable trend. Our visual analysis supports these conclusions, showing clear trends in budget increases and mission planning that align with each agency's strategic objectives.

## Future Work:

Given more time and resources, several avenues could be explored to enhance this project:

1. **Integration of Additional Data:** Incorporating variables such as technological advancements, international space policies, and economic indicators could refine the predictions and offer more nuanced insights into space exploration trends.
2. **Advanced Modeling Techniques:** Employing more sophisticated modeling approaches, such as time series forecasting, Random Forest, or Neural Networks, might improve the accuracy of our predictions for NASA and ISRO's future budgets and mission plans.
3. **Real-Time Data Integration:** Developing a system to integrate real-time data updates would allow for dynamic predictions, adapting to the latest trends and information, providing a continuously updated outlook on space agency activities.
4. **Collaborative Studies:** Partnering with space agencies to access more detailed internal data could provide deeper insights into budget allocations and mission planning processes, enhancing the validity and applicability of our findings.
5. **Impact Assessment:** Conducting analyses on the socio-economic impacts of space missions and budget changes could offer a holistic view of the strategic importance of space exploration on technology, education, and international relations.
6. **Complex Machine Learning Techniques:** Utilizing more complex machine learning algorithms could help uncover the underlying patterns in NASA’s budget and mission trends that are not captured by simple linear regression.
7. **Expanding the Scope:** Extending the machine learning analysis to include other space agencies could provide comparative insights and broaden the understanding of global space exploration efforts.
8. **Enhanced Visualizations:** Developing more detailed visualizations to better communicate the findings and trends observed in the data, making the insights accessible to a broader audience.

These enhancements could significantly improve the project’s depth and impact, providing clearer guidance for future space-related decision-making processes.