

Exploratory Data Analysis and Hypothesis Testing on Mission Success

Author Name: **Manas Sanjiv Sawant**

Date: **22/02/2025**

Organization: **Vidyalankar School of Information Technology**

# **Introduction**

As a passionate space enthusiast, I always stay updated on the latest developments in space exploration. Following the successful launch and landing of SpaceX's rocket from its launch pad, my curiosity was piqued: how many rockets have been launched and failed over the years?

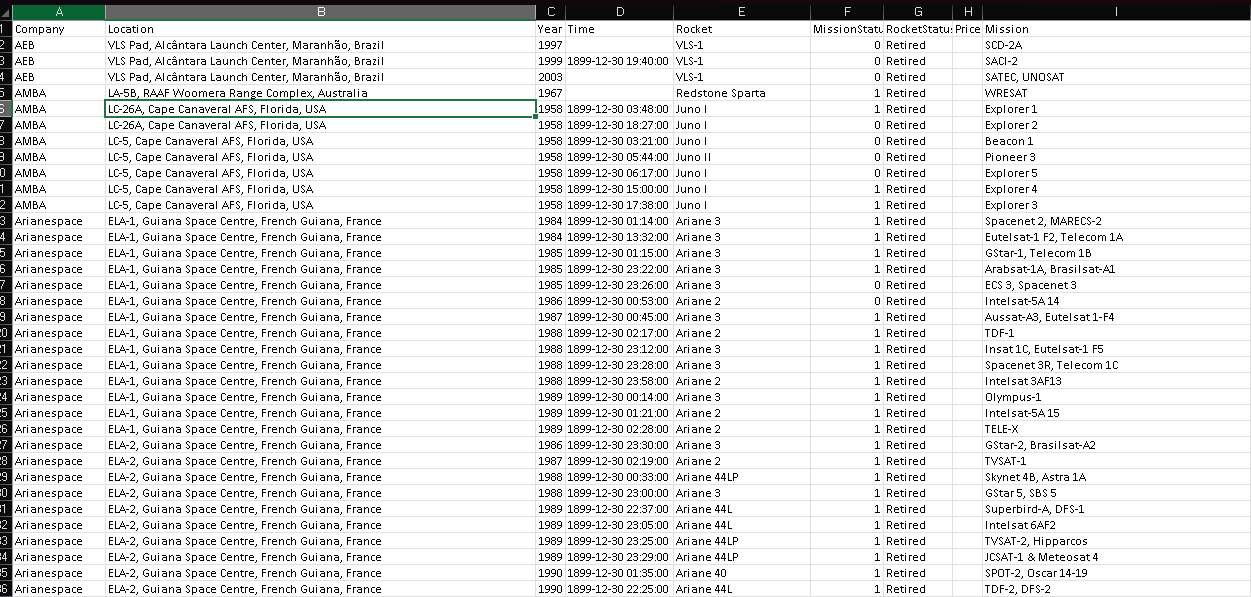
My curiosity led me to an insightful dataset available on Kaggle that documents real-time space missions. Additionally, my Hypothesis Testing professor assigned a mini-project task, making it the perfect opportunity to turn this curiosity into a structured project. This dataset provides valuable insights into space missions, allowing us to analyze trends, success rates, and costs over time.

# **Dataset Overview**

The dataset, available on Kaggle ([Real-Time Space Mission Dataset](https://www.kaggle.com/datasets/abhijitdahatonde/real-time-space-mission-dataset)), contains information on various space missions conducted by different companies. Below is a brief description of the key fields in the dataset:

* **Company**: The organization responsible for the space mission.
* **Location**: The location where the launch took place.
* **Date**: The date of the launch.
* **Time**: The time of the launch (UTC format).
* **Rocket**: The name of the rocket used in the mission.
* **Mission**: The name of the space mission(s) conducted.
* **RocketStatus**: The current operational status of the rocket as of August 2022 (Active or Inactive).
* **Price**: The cost of the rocket in millions of US dollars.
* **MissionStatus**: The outcome of the mission, categorized as:
  + Success
  + Failure

(Data set image)



# **Project Objective**

The main goal of this project is to analyze the trends in space missions by examining the number of launches, their success rates, and financial investments over time. Through statistical analysis and hypothesis testing, we aim to uncover insights such as:

* The frequency of successful vs. failed missions.
* The most reliable space agencies based on their mission success rates.
* The correlation between rocket cost and mission success.
* Identifying trends in rocket reusability and longevity.

By leveraging data analysis techniques, this project seeks to provide a deeper understanding of the advancements and challenges in space exploration.

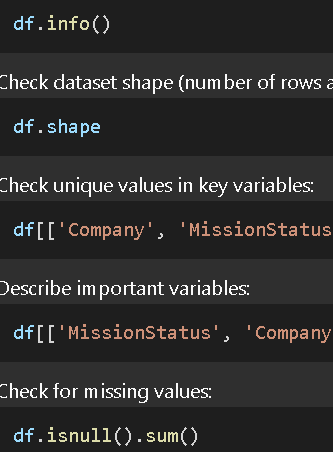
# Exploratory Data Analysis (EDA)

## Importing and Understanding the Date

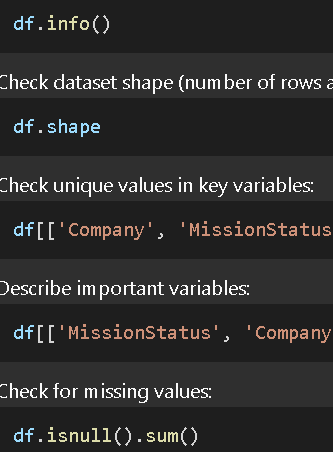


## Basic Data Exploration

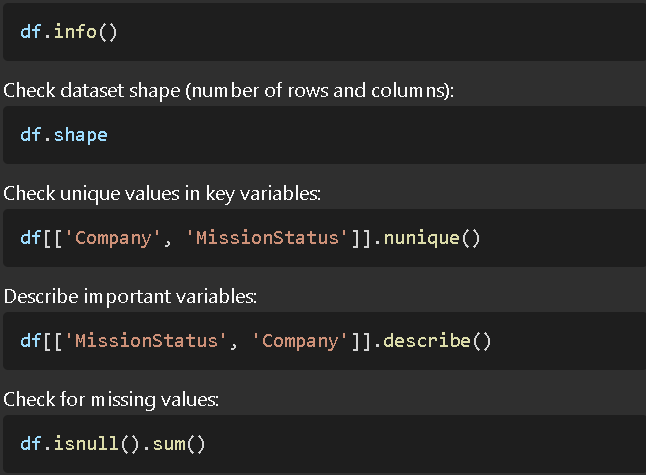
### Check dataset structure:



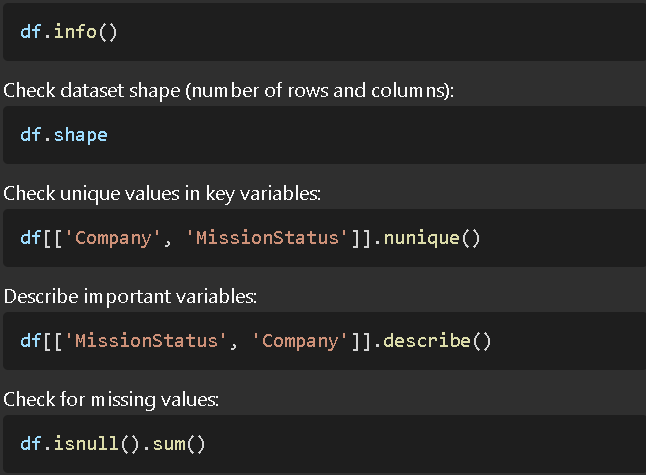
### Check dataset shape (number of rows and columns):



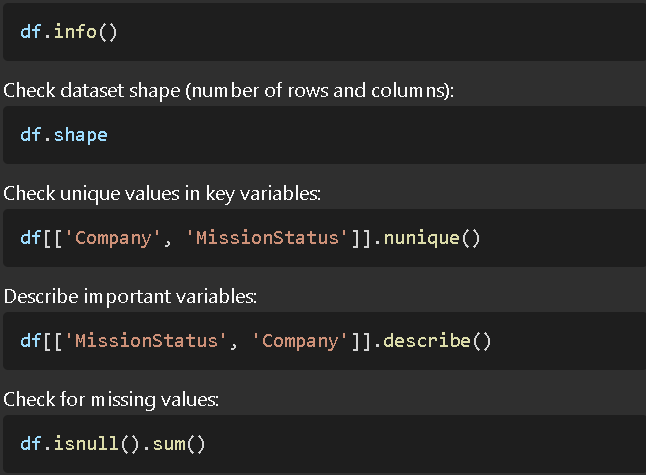
Check unique values in key variables:



### Describe important variables:



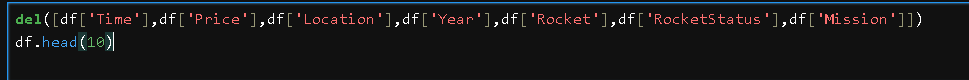
### Check for missing values:



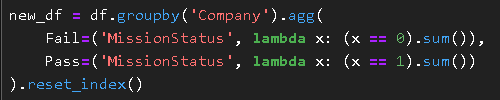
*Since there are no null values in the variables we are using (Company, MissionStatus), we proceed with hypothesis testing.*

## Data Cleaning and Transformation

### Remove Unwanted Columns



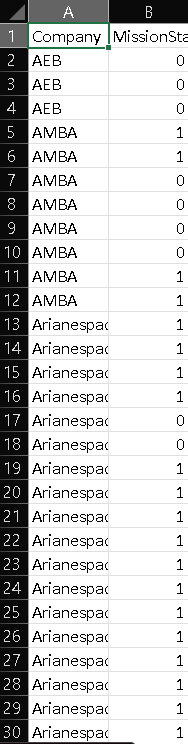
### Create a Transformed DataFrame



### Save Transformed Data



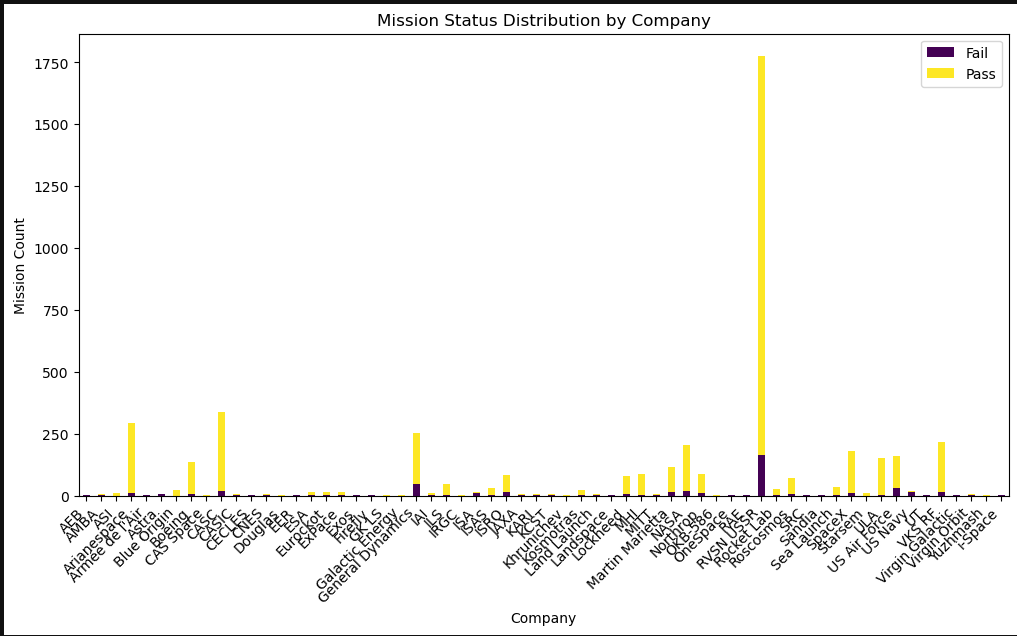
(New Data set image)

****

# Data Visualization

To better understand the relationship between companies and mission success rates, we will create visualizations such as:

* **Bar charts for mission success rates per company**



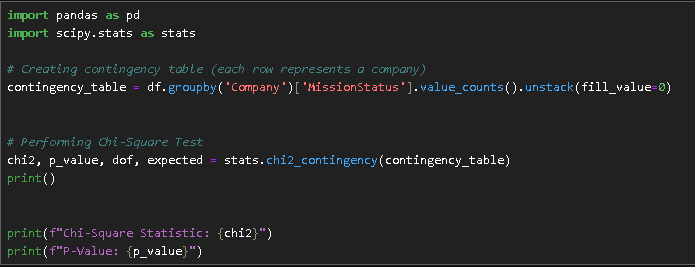
* **Heatmaps for correlation analysis**



# Hypothesis Testing

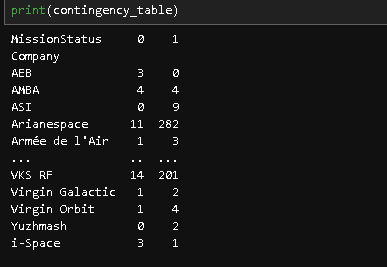
## Chi-Square Test for Independence

To examine whether the success or failure of space missions is dependent on the company responsible for the launch, we will conduct a **Chi-Square Test for Independence**. This statistical test helps determine whether there is a significant relationship between two categorical variables—in this case, **Company** and **Mission Status**.



## Contingency Table

We will first create a contingency table that summarizes the count of successful and failed missions for each company. This table will provide an overview of how mission outcomes vary by company and serve as the basis for the hypothesis test.



## Hypothesis Formulation

### Null Hypothesis (H₀):

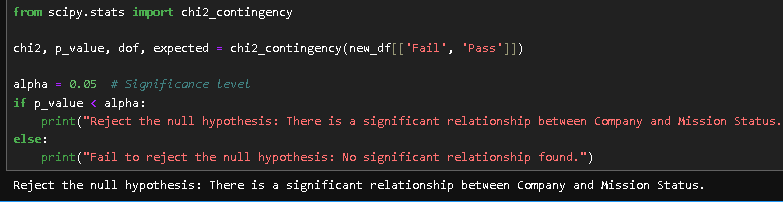
* + There is no relationship between the company conducting the mission and the mission outcome (success or failure). In other words, mission success or failure is independent of the company.

### Alternative Hypothesis (H₁):

* + There is a significant relationship between the company and mission outcome, meaning that certain companies might have different success rates compared to others.

## Performing the Chi-Square Test

Using the contingency table, we will compute the **Chi-Square statistic**, **degrees of freedom**, and **p-value** to assess the relationship between the company and mission success:



## Interpretation of Results

Based on the calculated p-value:

* **If p-value < 0.05**: Reject the null hypothesis, concluding that the company has a statistically significant impact on mission success.
* **If p-value ≥ 0.05**: Fail to reject the null hypothesis, indicating that mission success is likely independent of the company.

This test will help us understand whether mission success rates vary significantly across different space agencies and whether certain companies demonstrate higher reliability than others.

# Git Hub:

Link : <https://github.com/Monike123/Exploratory-Data-Analysis-and-Hypothesis-Testing>