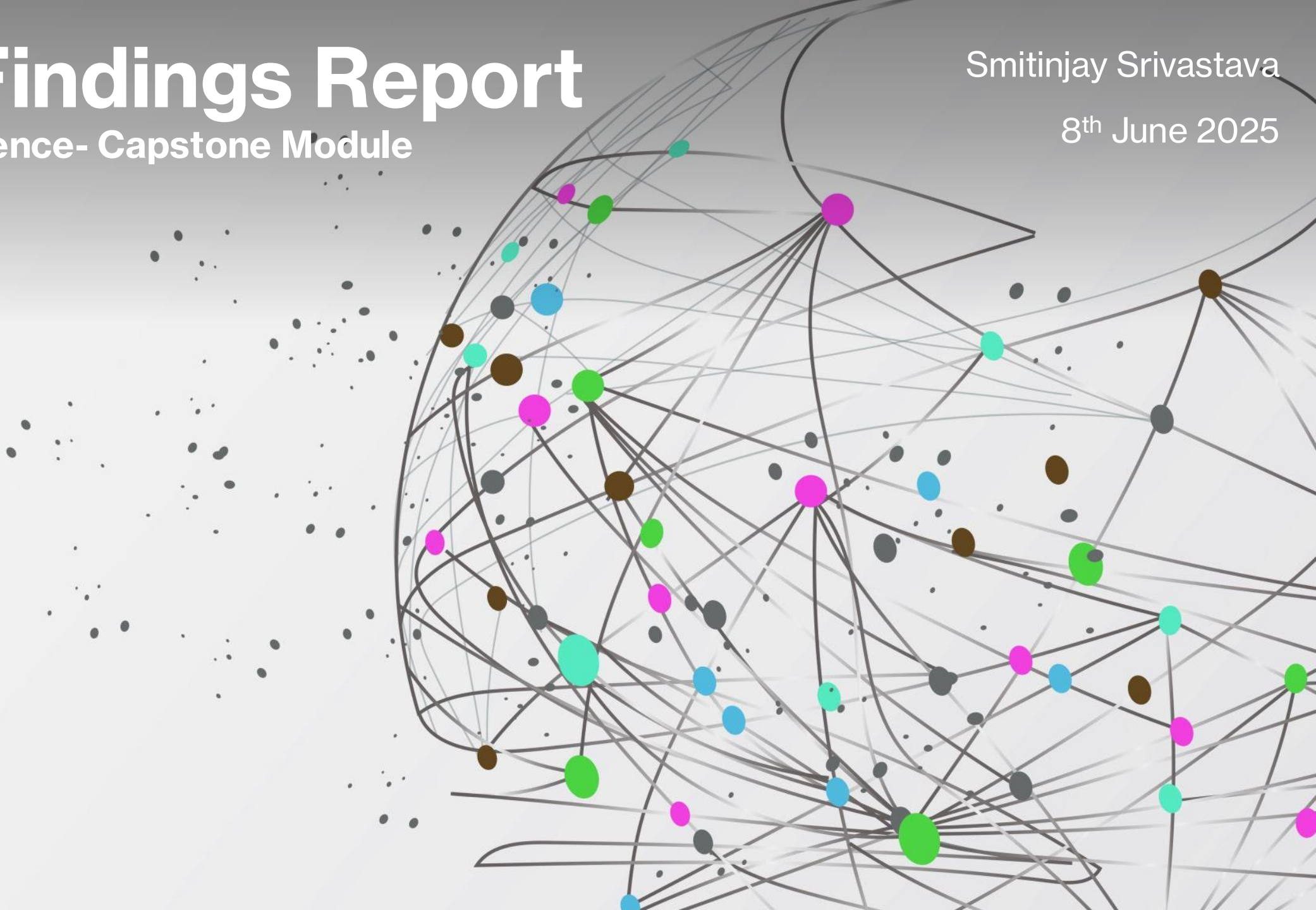# Data Findings Report

IBM-Data Science- Capstone Module

Smitinjay Srivastava

8th June 2025

Table of Contents

Executive Summary

- Examining the rocket company SpaceX and its first-stage rockets, and determining the price of each launch using machine learning algorithms.

- Outlines the findings and implications of:
  - Rocket Testing
  - Rocket Launch

Data Collection & Wrangling Methodology

Data Collection:

Scraped SpaceX Launch Data using APIs.

Combined with publicly available CSV datasets.

Data Wrangling:

Dropped redundant columns.

Handled missing values.

Created new features (Booster Version Category).

Normalized payload range.

EDA & Interactive Visual Analytics Methodology

- **EDA**:
  - Used SQL queries for basic exploratory analysis.
  - Aggregated launch outcomes by site and booster.

- **Visual Analytics**:
  - Dashboards created using Plotly Dash.
  - Interactive dropdowns, sliders, and maps.

EDA with Visualization Results

- Success Launches Pie Chart:
  - Visualizes success rates across all launch sites.
  - Drill-down view for specific sites.
- Payload vs. Launch Outcome Scatter Plot:
  - Payload correlation with success.
  - Booster version category color-coded.

EDA with SQL Results

- **SQL Queries Executed**:
  - Total launches per site.
  - Success vs. failure counts.
  - Average payload mass per booster category.
- **Insights**:
  - KSC LC-39A has the highest success rate.
  - Heavy boosters correlate with heavier payloads.

Interactive Map with Folium Results

- Mapped launch sites with Folium.

- Color-coded markers for successful vs. failed launches.

- Distance measurements from launch sites to nearby coastlines.

Predictive Analysis Methodolgy
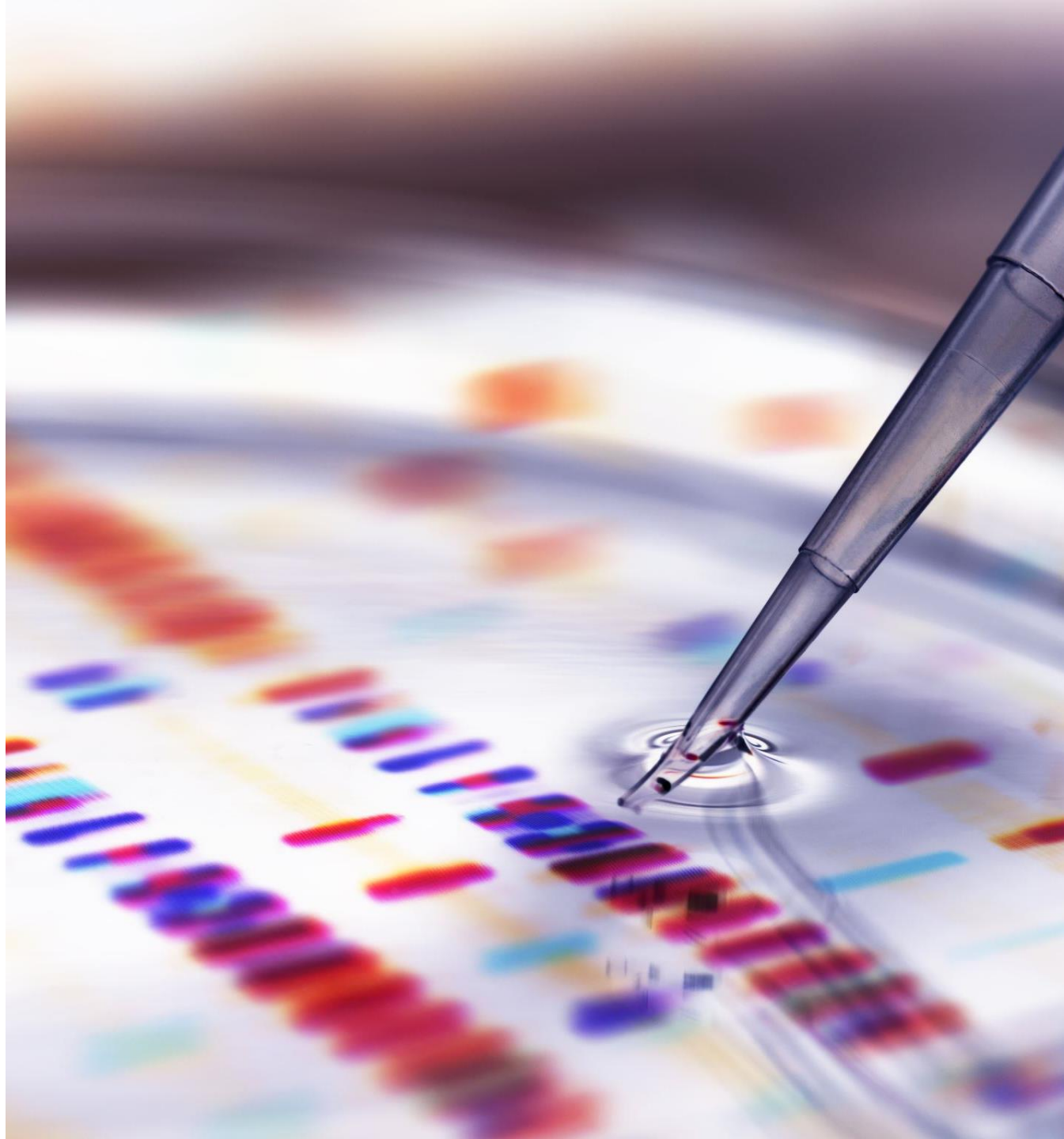
Predictive Analysis Methodolgy

Logistic Regression Model.

Input Features: Payload Mass, Booster Version Category, Launch Site.

Target: Launch success (binary classification).

Standardized features using Scikit-learn.

Predictive Analysis Results
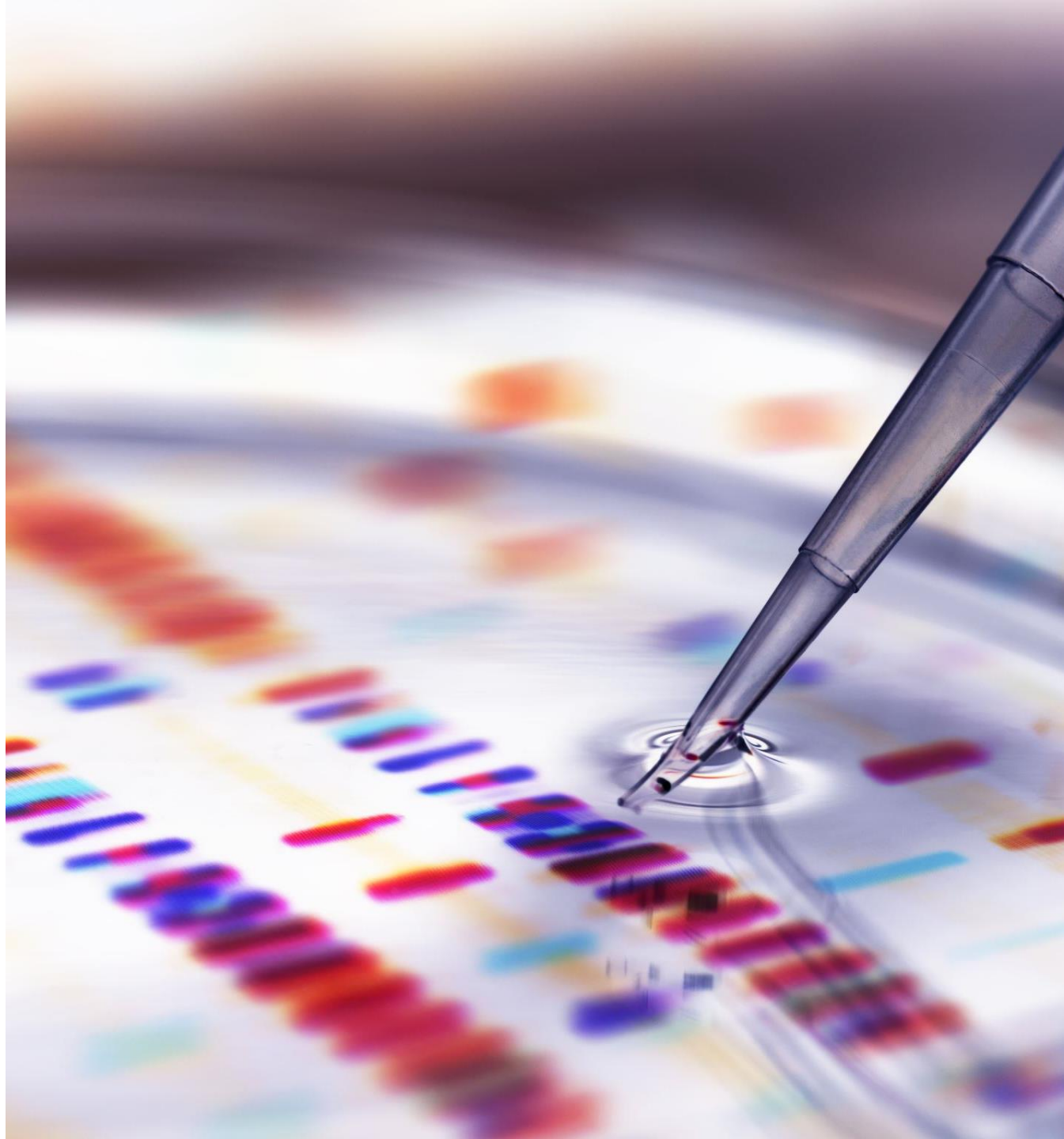
Accuracy:

Train accuracy: ~85%

Test accuracy: ~82%

Confusion Matrix:

High true positive rate.

Few false negatives (missed successes).

Plotly Dashboard: SpaceX_Dashboard Results

- KSC LC 39A had both the largest number of successful launches (76.9% of successful launches) and the highest launch success rate in total (41.7%).

- The payload range that had the highest successful launch rate was 1,1225 to 6,225 range

- The payload range that had the lowest successful launch rate was the 7,500 to 10,000 range

- The F9 booster version that had the highest launch success rate was the FT with 14 successful ranges across the entire payload range.
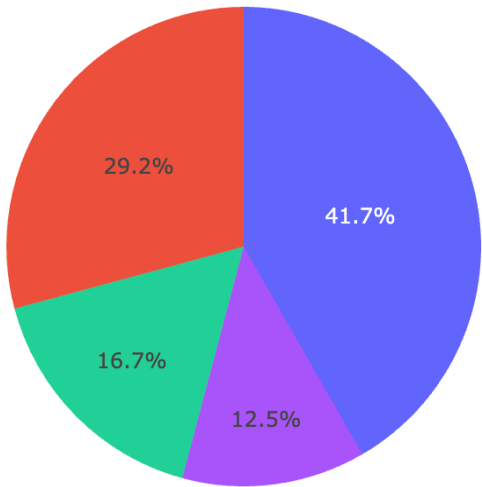
SpaceX Launch Records Dashboard

All Sites ✕ ▾

Total Success Launches by Site

- ■ KSC LC-39A
- ■ CCAFS LC-40
- ■ VAFB SLC-4E
- ■ CCAFS SLC-40

41.7%

29.2%

16.7%

12.5%

Payload range (Kg):

| 0 | 2500 | 5000 | 7500 | 10000 |

Correlation between Payload and Success for All Sites

Conclusion

- Launch site and booster type are significant predictors of success.
- Payload mass shows a mild correlation with success.
- An interactive dashboard provides easy access to insights.
- Future Work: Include weather data, expand booster version details.