

Winning Space Race with Data Science

Surender Kannah
12/9/2025



Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion

Executive Summary

Executive Summary

This project analyzes SpaceX launch data to uncover patterns influencing mission outcomes and booster landings. The workflow integrates data engineering, visualization, and machine learning to deliver actionable insights.

Methodology Overview

1. **Data Acquisition**
2. **Data Preparation**
3. **Feature Engineering**
4. **Interactive Visualization**
5. **Model Development & Evaluation**

Key Results

- **Data Insights**
- **Model Accuracy**
- **Strategic Findings**

Introduction

Project Overview

This capstone initiative explores the predictive modeling of rocket booster landings, focusing on SpaceX's Falcon 9 missions. The goal is to understand the operational dynamics that enable reusable launch systems, which are central to SpaceX's cost advantage in the aerospace market. By analyzing launch data and applying machine learning techniques, the project aims to generate insights that can inform strategic decisions for competitors and stakeholders in the commercial space industry.

Core Objectives

- **Identify operational drivers** that contribute to successful first-stage landings.
- **Develop predictive models** that estimate landing outcomes based on mission parameters.

Section 1

Methodology

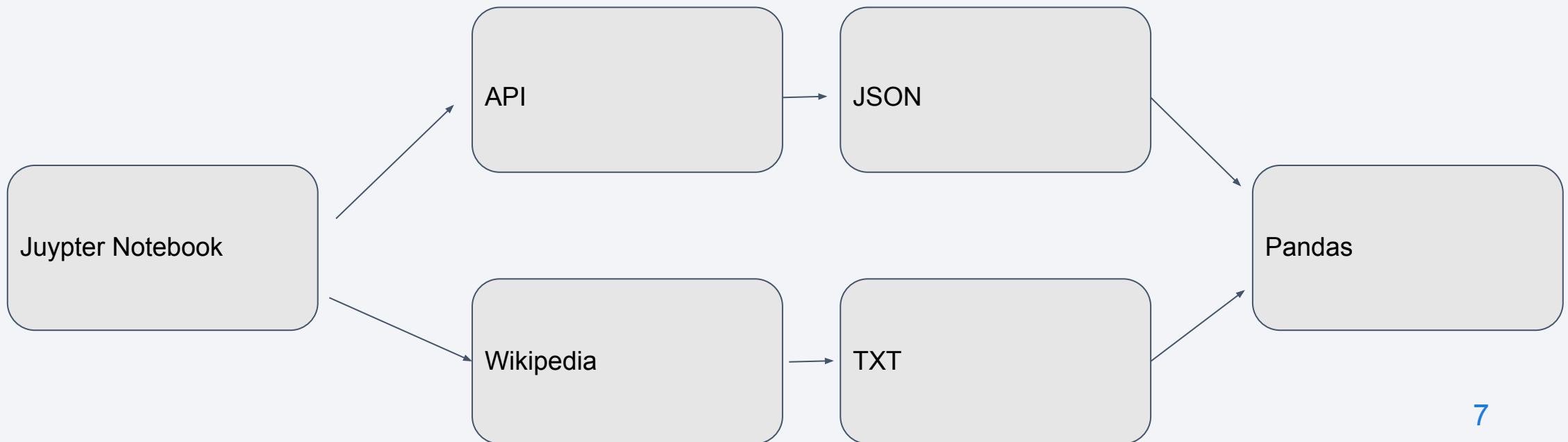
Methodology

Executive Summary

- Data collection methodology:
 - SpaceX API
- Perform data wrangling
 - Cleaned for missing values, formats, consistency
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
 - Folium and Plotly Dash

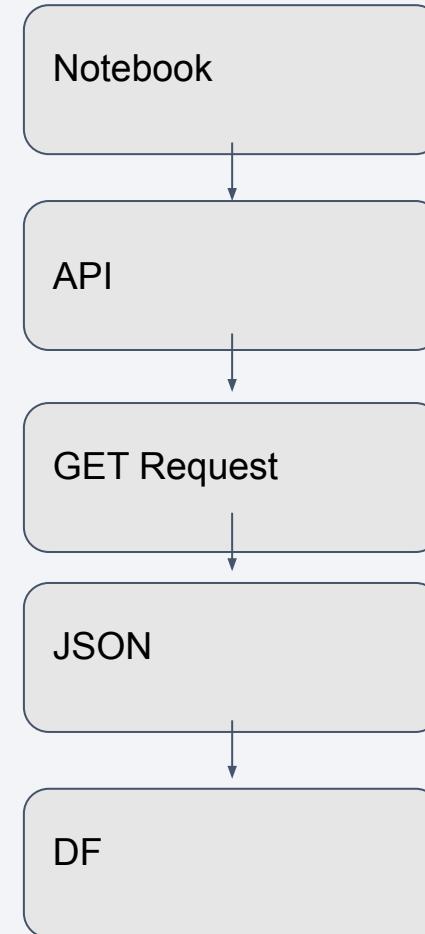
Data Collection

- Spacex API
- Wikipedia



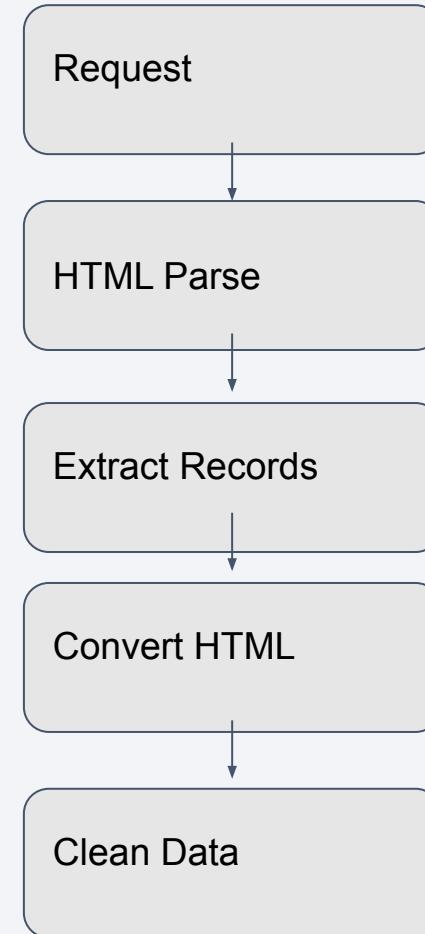
Data Collection – SpaceX API

- SpaceX API
 - import Pandas, Numpy, and Request
 - Get request
 - Geospatial
 - Type
 - Orbit
 - Flight Number
- [https://github.com/skannah/IBM-Capstone/blob/main/jupyter-labs-spacex-data-collection-api%20\(2\).ipynb](https://github.com/skannah/IBM-Capstone/blob/main/jupyter-labs-spacex-data-collection-api%20(2).ipynb)



Data Collection - Scraping

- Step 1 - Web Scrape
 - Step 2 - Parse HTML
 - Step 3 - Convert to DF
-
- [https://github.com/skannah/IBM-Capstone/blob/main/jupyter-labs-webscraping%20\(1\).ipynb](https://github.com/skannah/IBM-Capstone/blob/main/jupyter-labs-webscraping%20(1).ipynb)



Data Wrangling

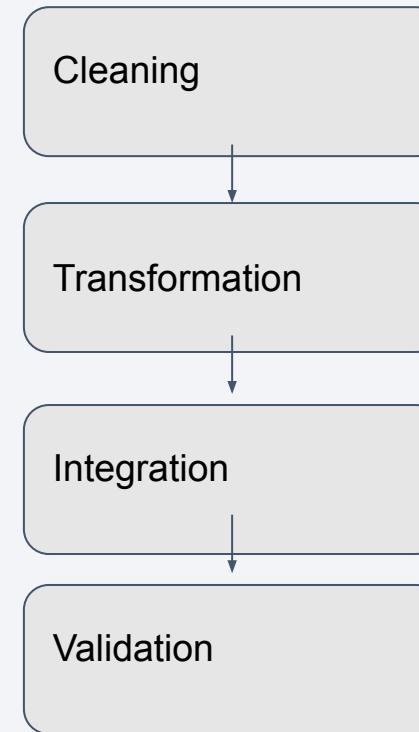
Step 1 - Data Cleaning

Step 2 - Data Transformation

Step 3 - Data Integration

Step 4 - Data Validation

- <https://github.com/skannah/IBM-Capstone/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb>



EDA with Data Visualization

- Histograms - central tendency of data
 - Bar charts - proportions of categorical data
 - Line charts - reveal temporal patterns
 - Scatter plots - visualize how one variable changes over another
 - Heatmaps - help identify correlations
 - Box plots - visualize the spread of data
-
- <https://github.com/skannah/IBM-Capstone/blob/main/edadataviz.ipynb>

EDA with SQL

- Aggregate
 - Join
 - Filtering
 - Sorting
 - Subqueries
-
- [https://github.com/skannah/IBM-Capstone/blob/main/jupyter-labs-eda-sql-courseera_sqlite%20\(1\).ipynb](https://github.com/skannah/IBM-Capstone/blob/main/jupyter-labs-eda-sql-courseera_sqlite%20(1).ipynb)

Build an Interactive Map with Folium

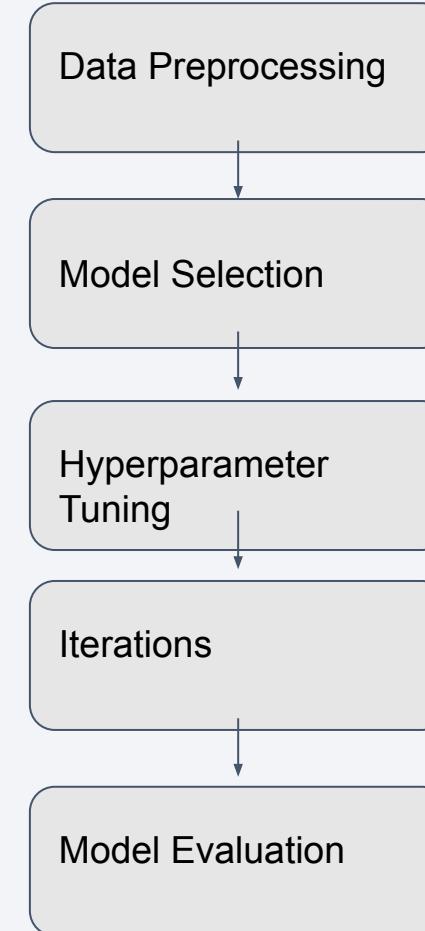
- Markers - pinpoint exact launch locations
 - Circles - illustrates the impact zone
 - Lines - connection between launch sites
-
- https://github.com/skannah/IBM-Capstone/blob/main/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- Pie Chart - overview of mission success rates
- Scatter Plot - correlations between payload and outcome
- Launch Site Drop Down - allows to focus analysis
- Range Slider - interactive exploration of payload mass and success
- <https://github.com/skannah/IBM-Capstone/blob/main/lab.py>

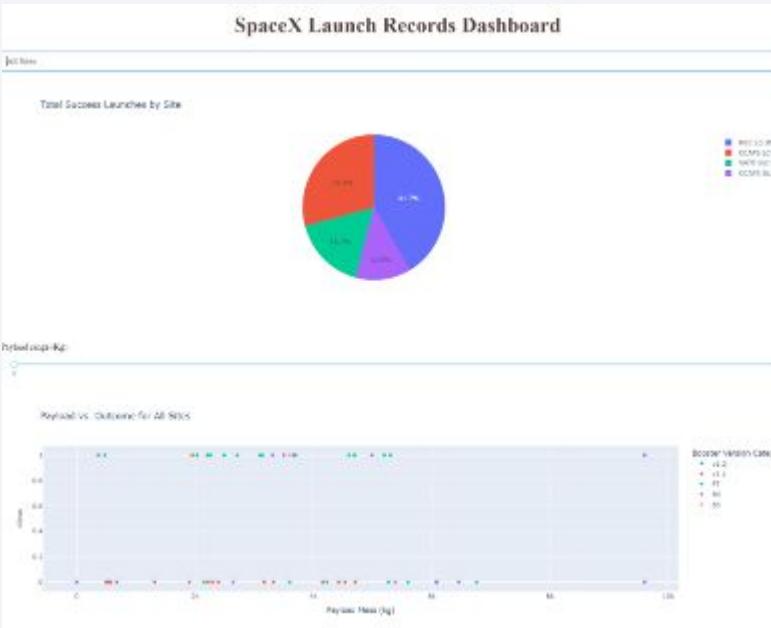
Predictive Analysis (Classification)

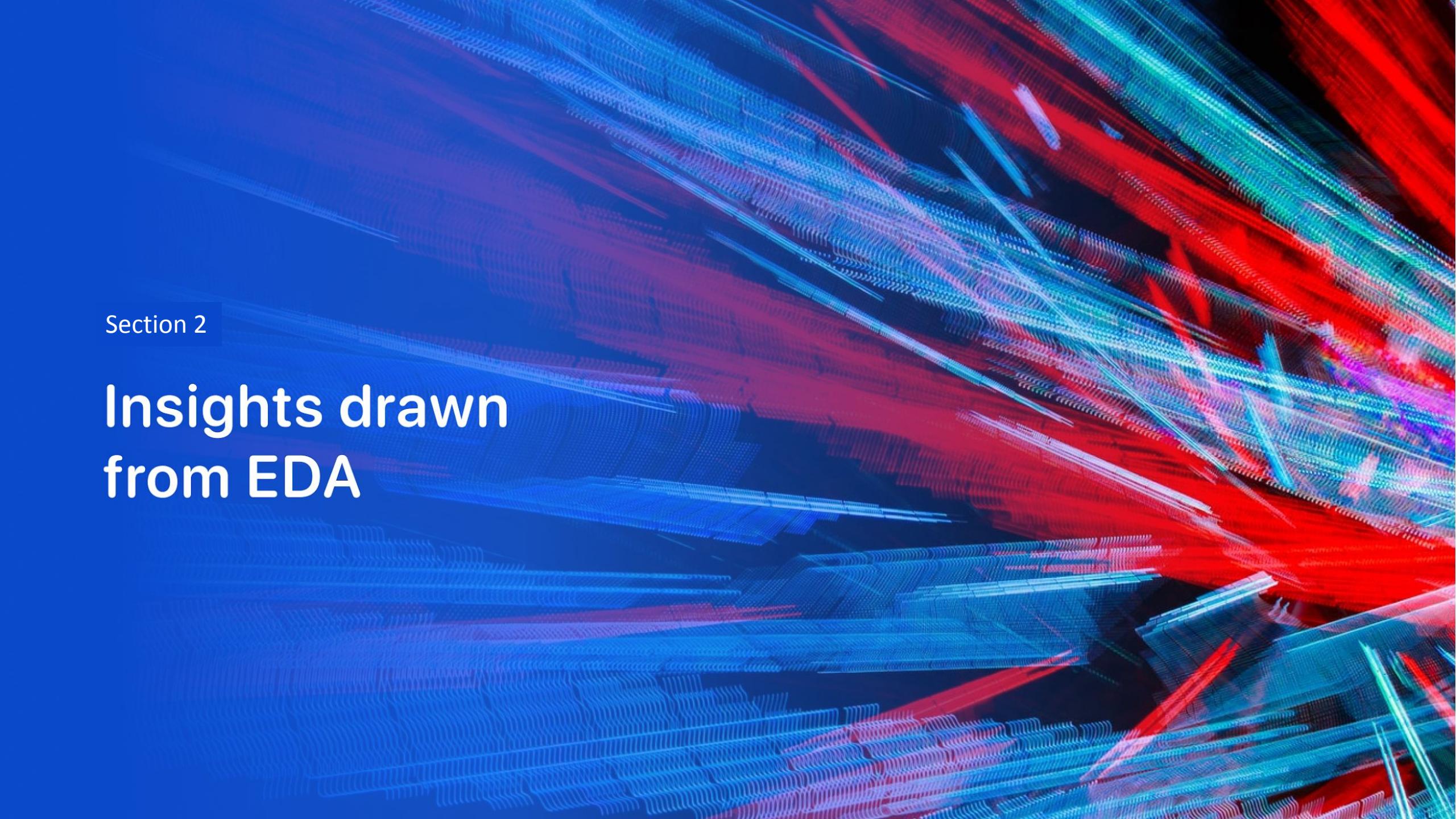
- Data Preprocessing
 - Model Selection
 - Hyperparameter Tuning
 - Iterations
 - Model Evaluation
-
- [https://github.com/skannah/IBM-Capstone/
blob/main/SpaceX_Machine_Learning_Prediction.ipynb](https://github.com/skannah/IBM-Capstone/blob/main/SpaceX_Machine_Learning_Prediction.ipynb)



Results

- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results



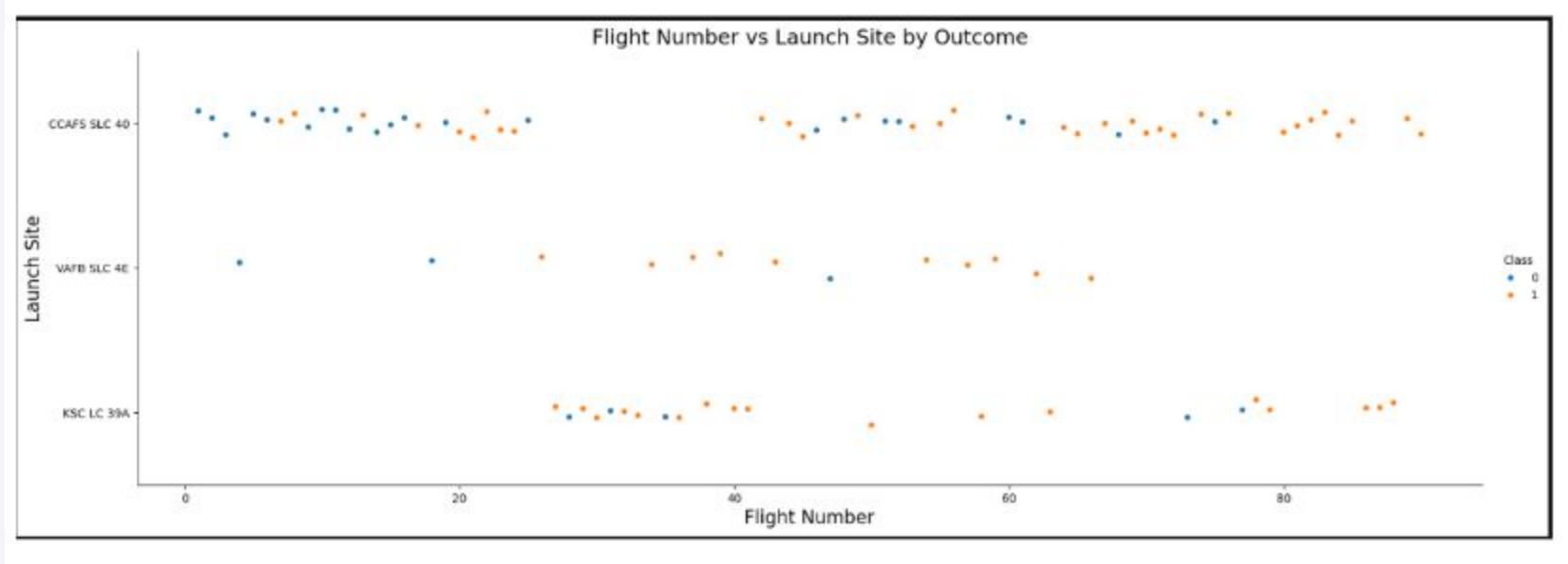
The background of the slide features a complex, abstract pattern of glowing lines. These lines are primarily blue and red, creating a sense of depth and motion. They appear to be composed of numerous small, glowing particles or dots, giving them a textured, almost liquid-like appearance. The lines converge and diverge, forming various shapes and directions across the dark, solid-colored background.

Section 2

Insights drawn from EDA

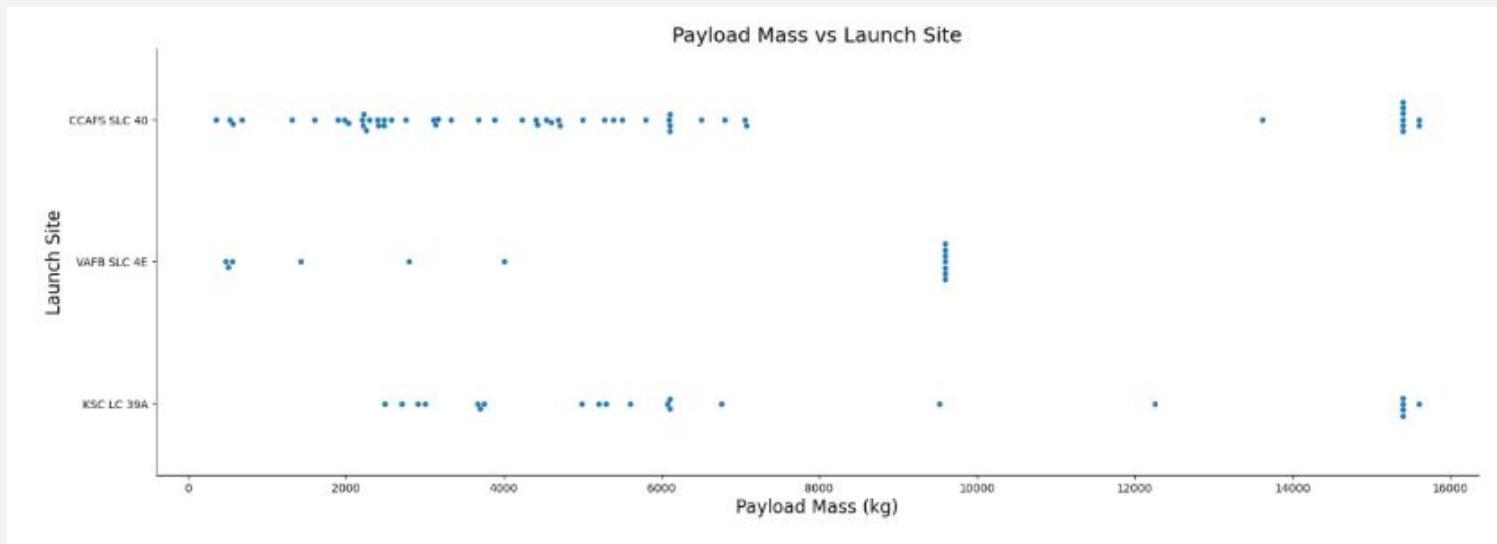
Flight Number vs. Launch Site

- Mixed outcomes at major launch sites
- Consistent activity across flight numbers



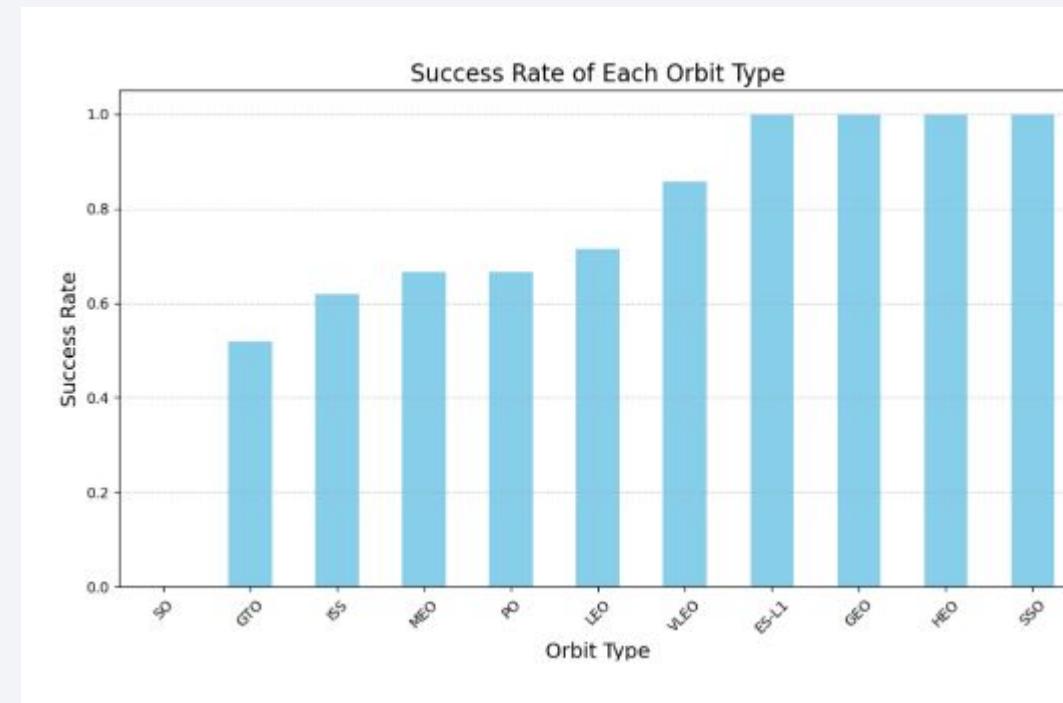
Payload vs. Launch Site

- Payload distribution
- High capacity launches



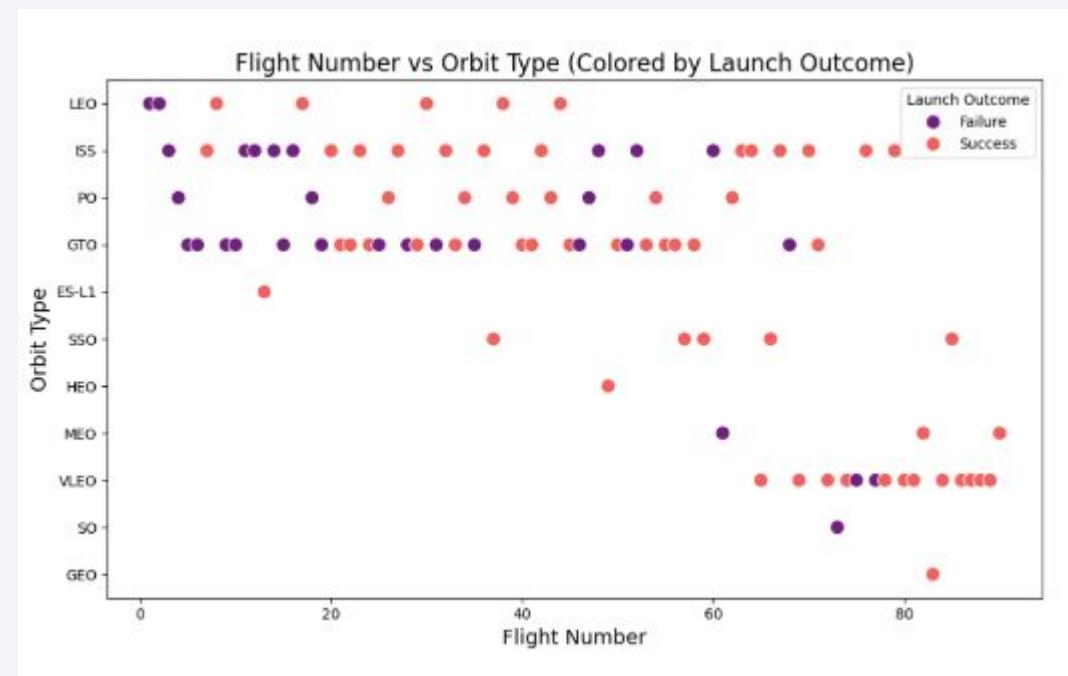
Success Rate vs. Orbit Type

- High success rate
- Lower success rate for GTO



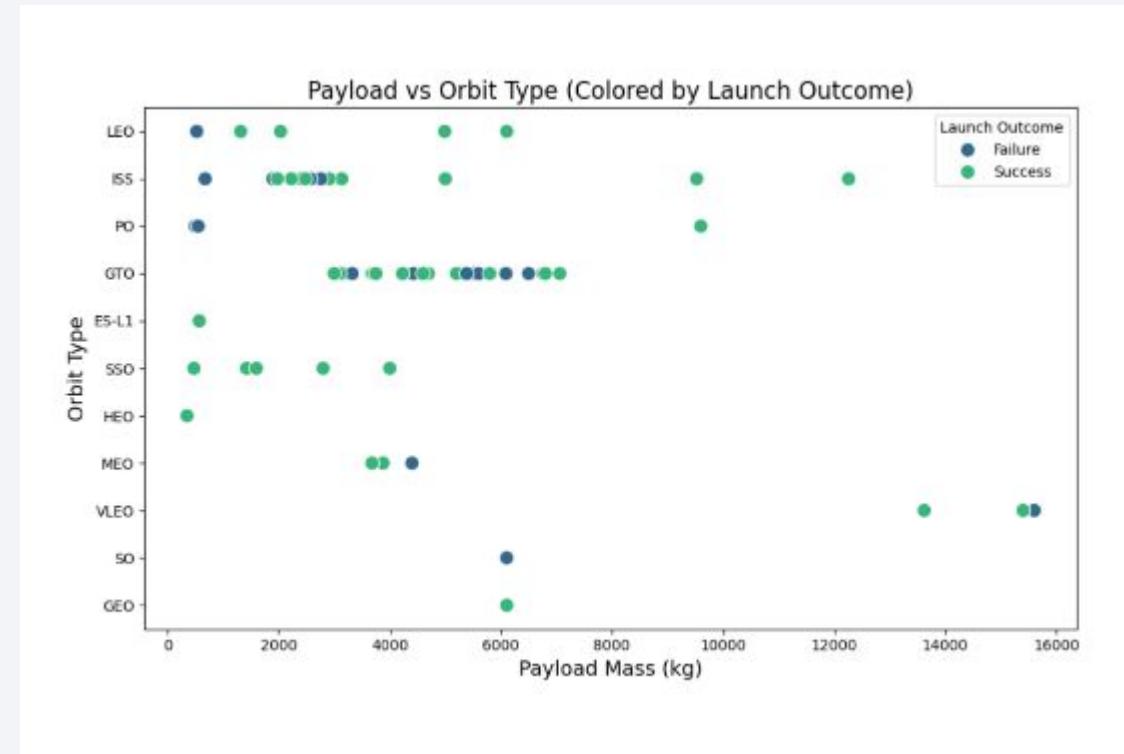
Flight Number vs. Orbit Type

- Increased success over time
- Orbit specific performance



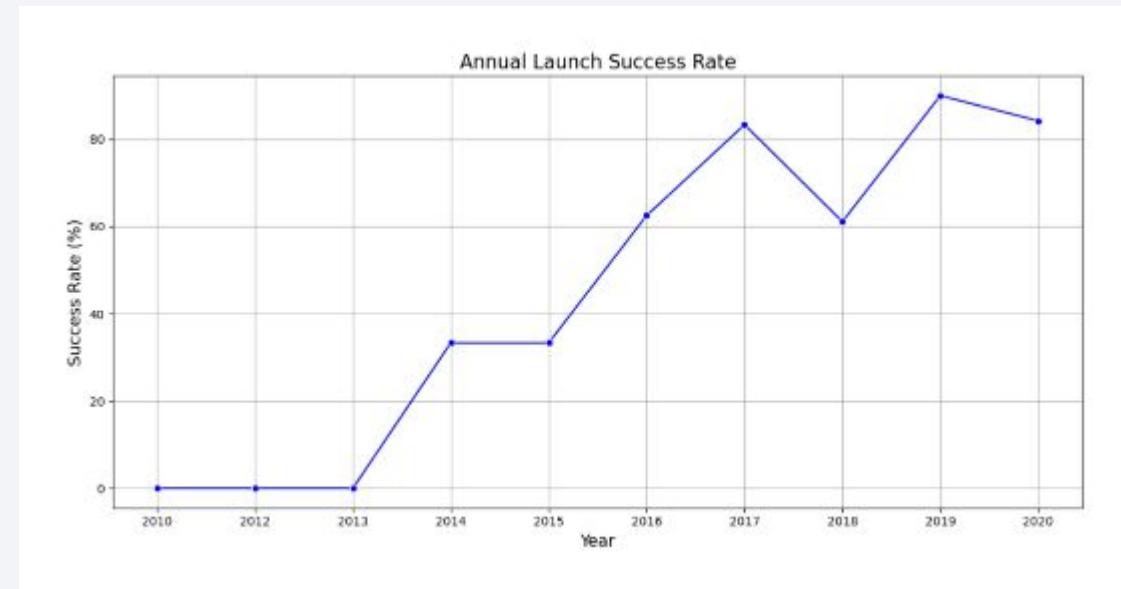
Payload vs. Orbit Type

- Successful landings less than 6000kg
- Increased difficulty with heavier payloads



Launch Success Yearly Trend

- Success rate improves from 2013 onwards
- Dip in 2018, but overall trend indicates increasing reliability



All Launch Site Names

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

Launch Site Names Begin with 'CCA'

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS_KG	Orbit	Customer	Mission_Outcome	Landing_Outcome
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-12-08	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

SUM(PAYLOAD_MASS_KG_)
45596

Average Payload Mass by F9 v1.1

AVG(PAYLOAD_MASS_KG_)
2928.4

First Successful Ground Landing Date

MIN(Date)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

Booster_Version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

Mission_Outcome	Total
Success	98

Boosters Carried Maximum Payload

Booster_Version
F9 B5 B1048.4
F9 B5 B1049.4
F9 B5 B1051.3
F9 B5 B1056.4
F9 B5 B1048.5
F9 B5 B1051.4
F9 B5 B1049.5
F9 B5 B1060.2
F9 B5 B1058.3
F9 B5 B1051.6
F9 B5 B1060.3
F9 B5 B1049.7

2015 Launch Records

Month_Name	Mission_Outcome	Booster_Version	Launch_Site
January	Success	F9 v1.1 B1012	CCAFS LC-40
February	Success	F9 v1.1 B1013	CCAFS LC-40
March	Success	F9 v1.1 B1014	CCAFS LC-40
April	Success	F9 v1.1 B1015	CCAFS LC-40
April	Success	F9 v1.1 B1016	CCAFS LC-40
June	Failure (in flight)	F9 v1.1 B1018	CCAFS LC-40
December	Success	F9 FT B1019	CCAFS LC-40

Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

Landing_Outcome	Count
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

The background of the slide is a photograph taken from space at night. It shows the curvature of the Earth's horizon against a dark blue sky. City lights are visible as small white dots, with larger clusters of lights indicating major urban areas. In the upper right corner, there is a faint, greenish glow of the aurora borealis or a similar atmospheric phenomenon.

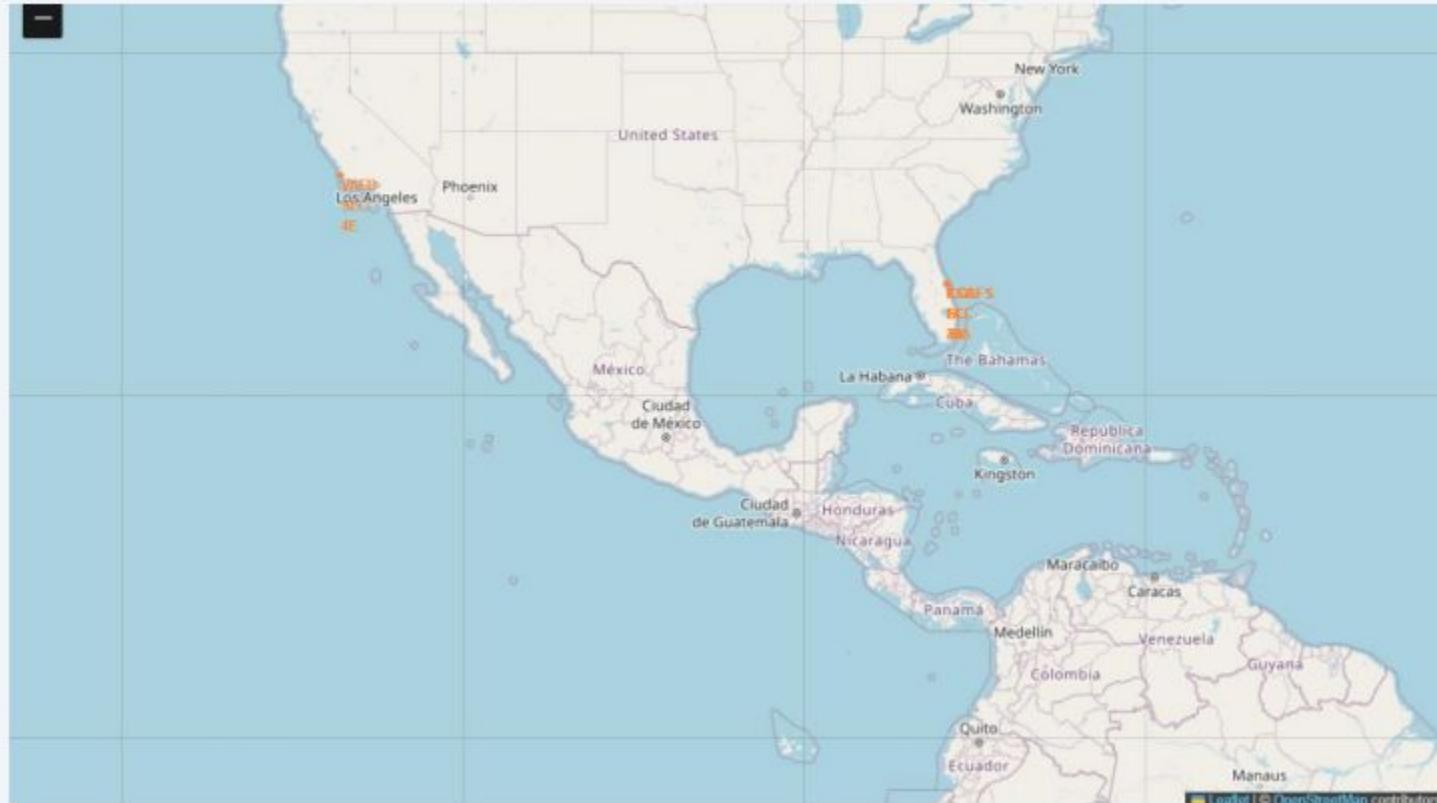
Section 3

Launch Sites Proximities Analysis

Task 1

- 1. Equator Line

- 2. Proximity to Coast

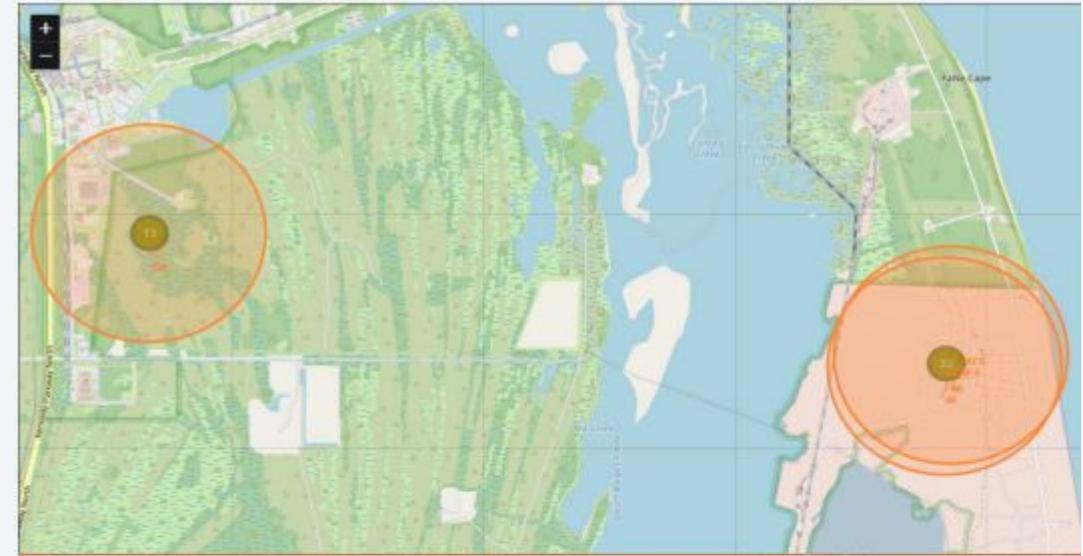


Task 2

- 19 red markers
- 7 green markers

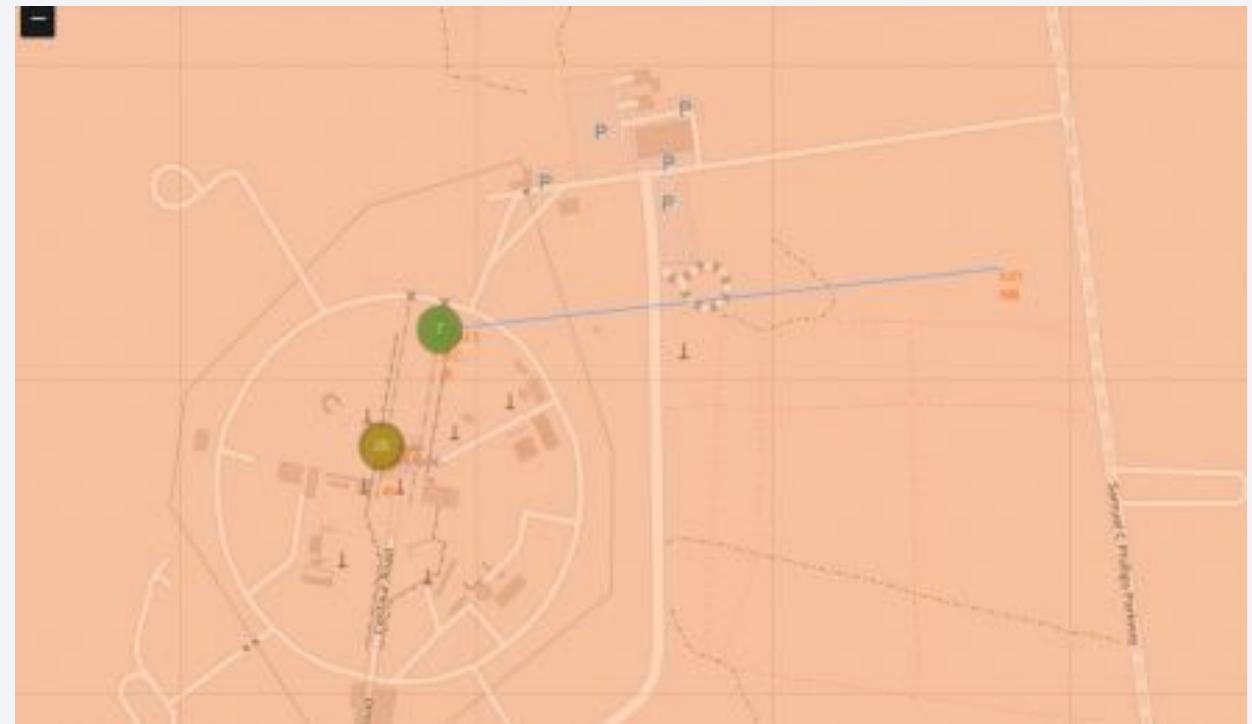
Red is unsuccessful

Green is successful



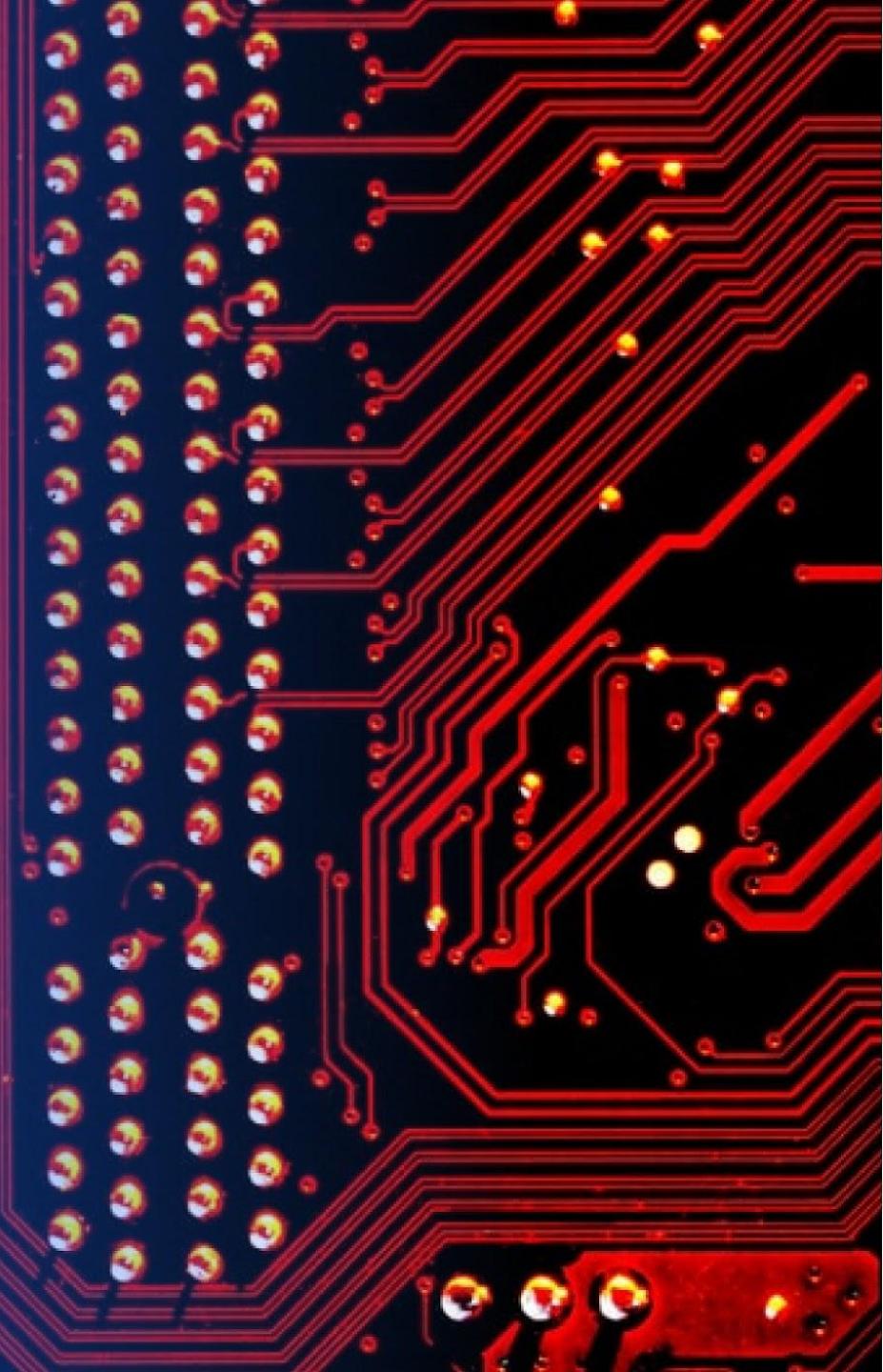
Task 3

- Proximity to coastline
- 0.51 km distance
- Safe recovery operations



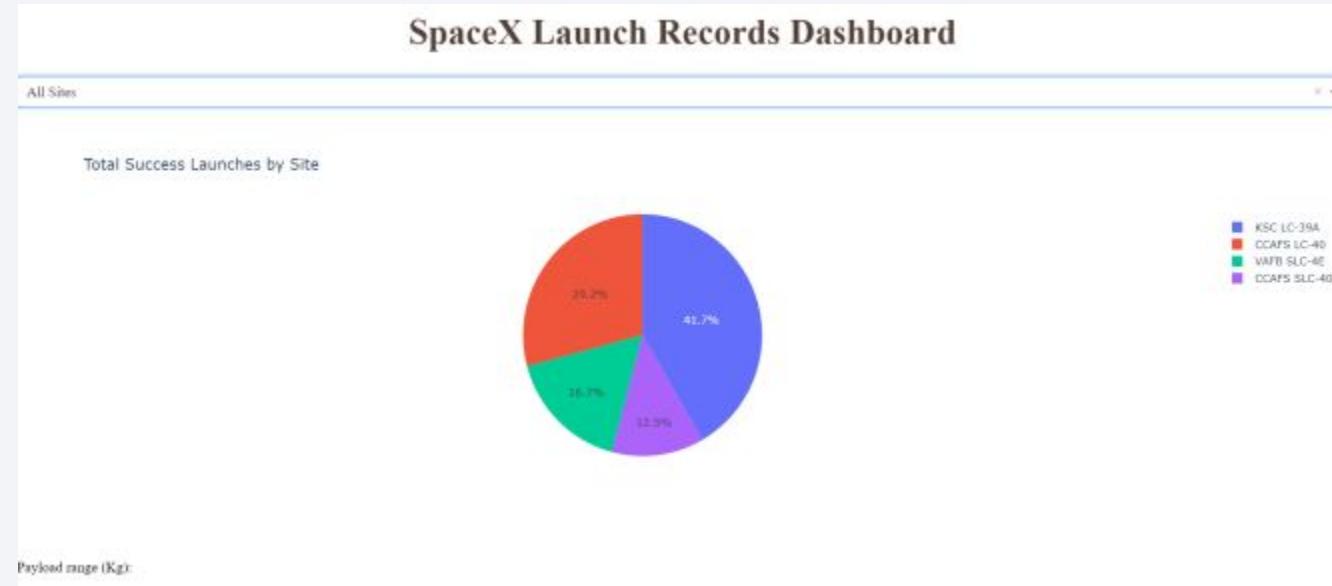
Section 4

Build a Dashboard with Plotly Dash



Launch Success

- CCAFS LC 40 - 29.2%
- CCAFS SLC 40 - 12.5%
- VAFB SLC 4E - 16.7%
- KSC LC 39A - 41.7%

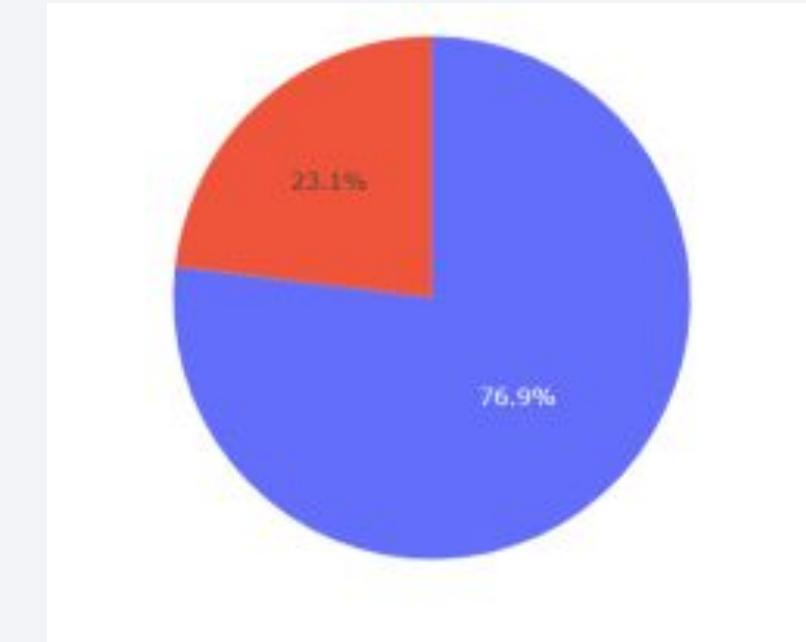


Pie chart Highest Success

Class 1 - 76.9%

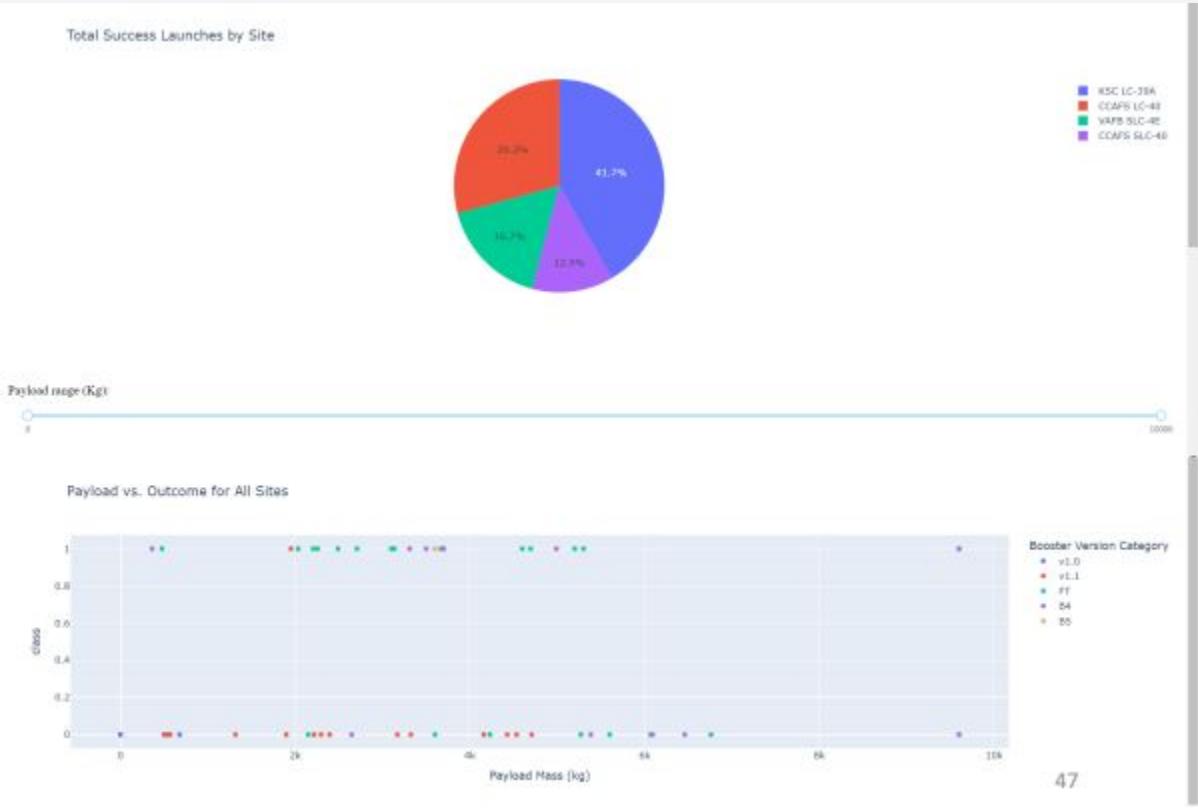
Class 0 - 23.1%

High success rate



Key Insights

- Launch Site Success Rate
 - CCAFS LC-40 - most reliable
- Booster Version
 - FT - More frequent
 - v1.0 - Fewer

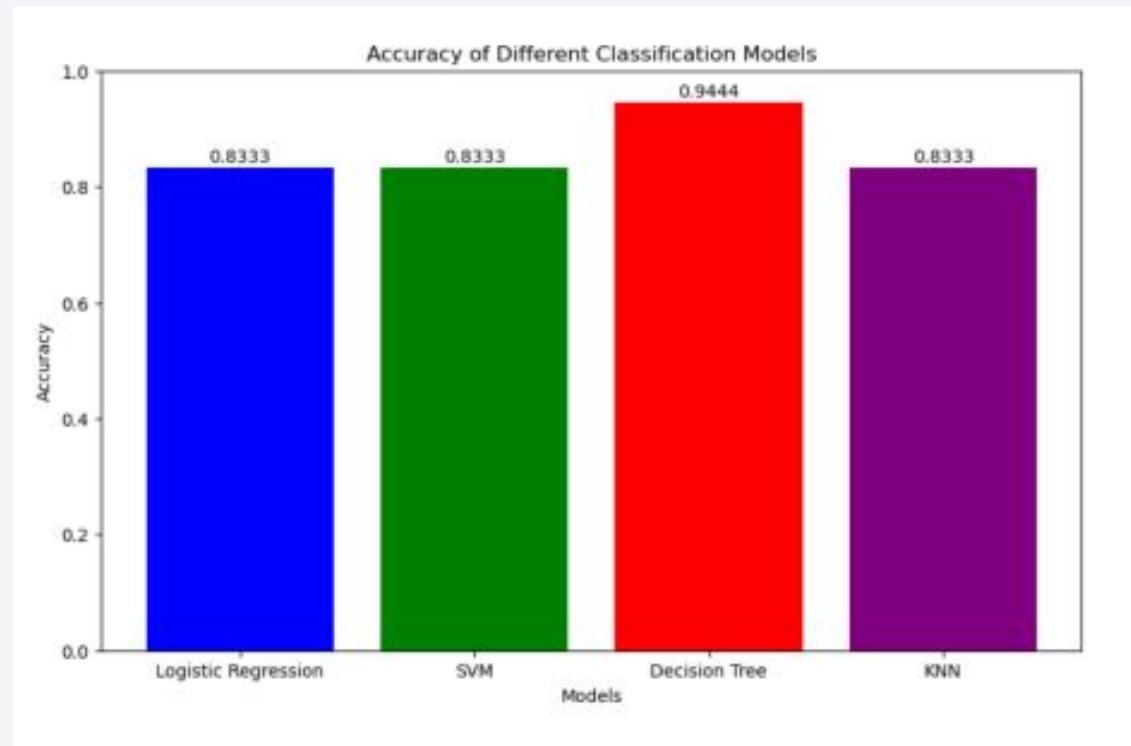


Section 5

Predictive Analysis (Classification)

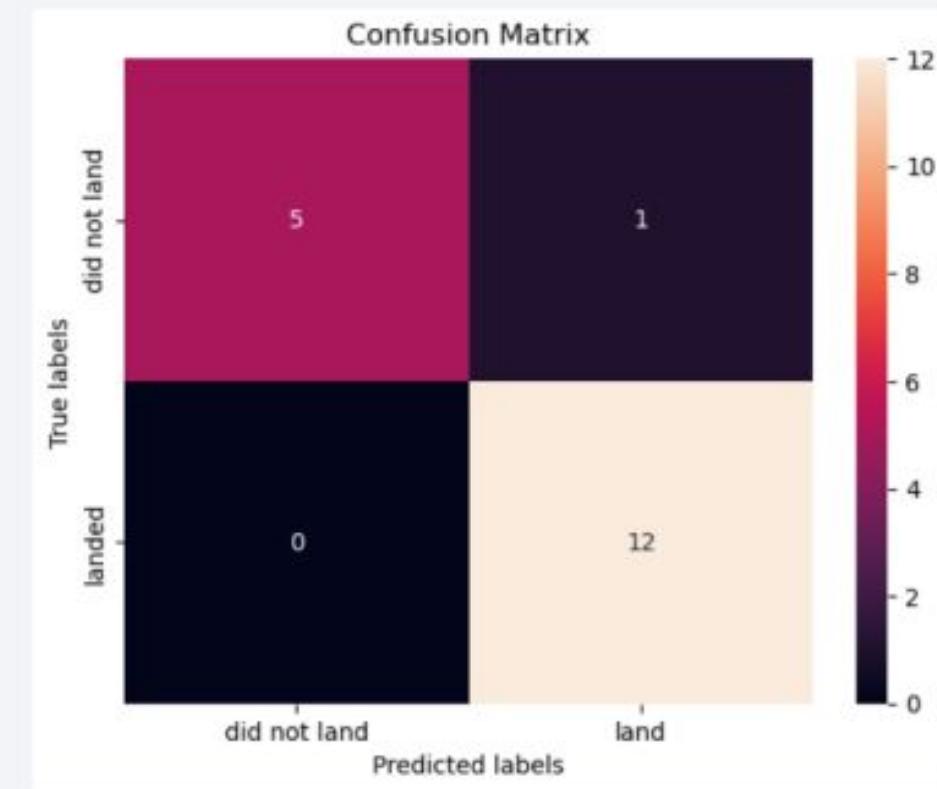
Classification Accuracy

- Decision Tree - 0.944
- Logistic/SVM/KNN - 0.8333



Confusion Matrix

High Accuracy
No False Negatives
Few False Positive
Balanced



Conclusions

- No clear pattern for higher payload to lower success rate
- FT booster has a high success rate
- CCAFS LC-40 has highest success rate
- Interactive visualization provide valuable insights

Thank you!

