IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Tahsine K. SALAMI
28/08/2024

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion
- Appendix

Executive Summary

- Summary of methodologies:
  - Data Collection
  - Data Wrangling
  - EDA with Visualisation and SQL
  - Interactive visual analytics with Folium and Ploty Dash
  - Predictive analysis using classification models
- Summary of all results:
  - Exploratory data analysis results
  - Interactive analytics demo
  - Predictive analysis results

Introduction

The commercial space age has arrived, marking a revolution in access to space. Today, companies like Virgin Galactic, Rocket Lab, Blue Origin, and especially SpaceX, are making space travel more affordable and accessible. Among these players, SpaceX stands out for its numerous achievements: sending machines to the International Space Station, development of the Starlink constellation to provide global Internet access, and manned missions in space.

One of the key factors in SpaceX's success lies in the drastic reduction in launch costs. While most space service providers charge more than $165 million per launch, SpaceX offers launches of its Falcon 9 rocket for just $62 million. This impressive savings is largely due to the reuse of the rockets' first stage.

In this context, accurately predicting whether a rocket's first stage will successfully land becomes crucial for estimating the cost of a launch. This is precisely the purpose of our project: to analyze SpaceX data to improve our understanding of these landings and thus refine the evaluation of the associated costs.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - SpaceX API Rest
  - Wikipedia Web Scraping
- Perform data wrangling
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Classification Model Development Process

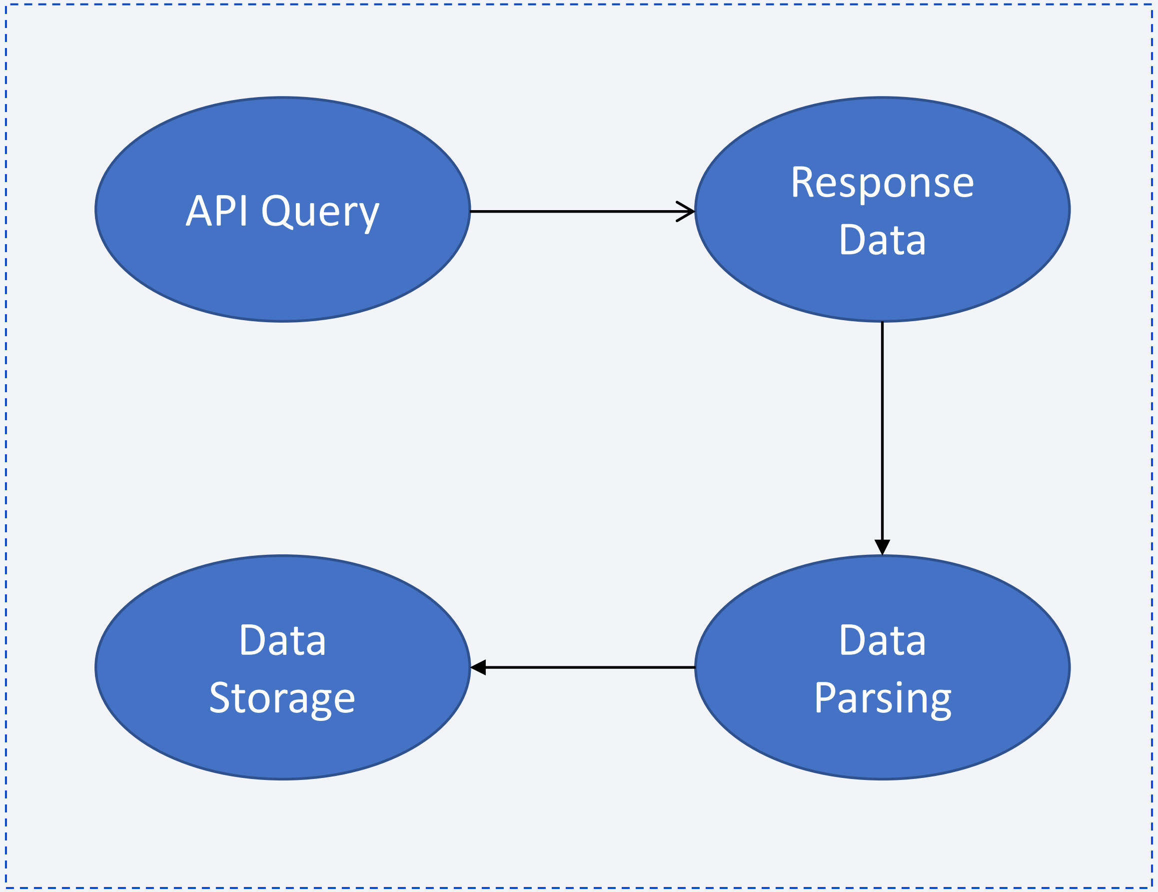
Data Collection

The data collection process for the SpaceX Falcon 9 launch dataset involved two primary methods: API retrieval and web scraping.

By combining data from the SpaceX REST API with hisotical information from Wikipedia, we ensured a comprehensive dataset that captured all revelant aspects of the Falcon 9 launches.

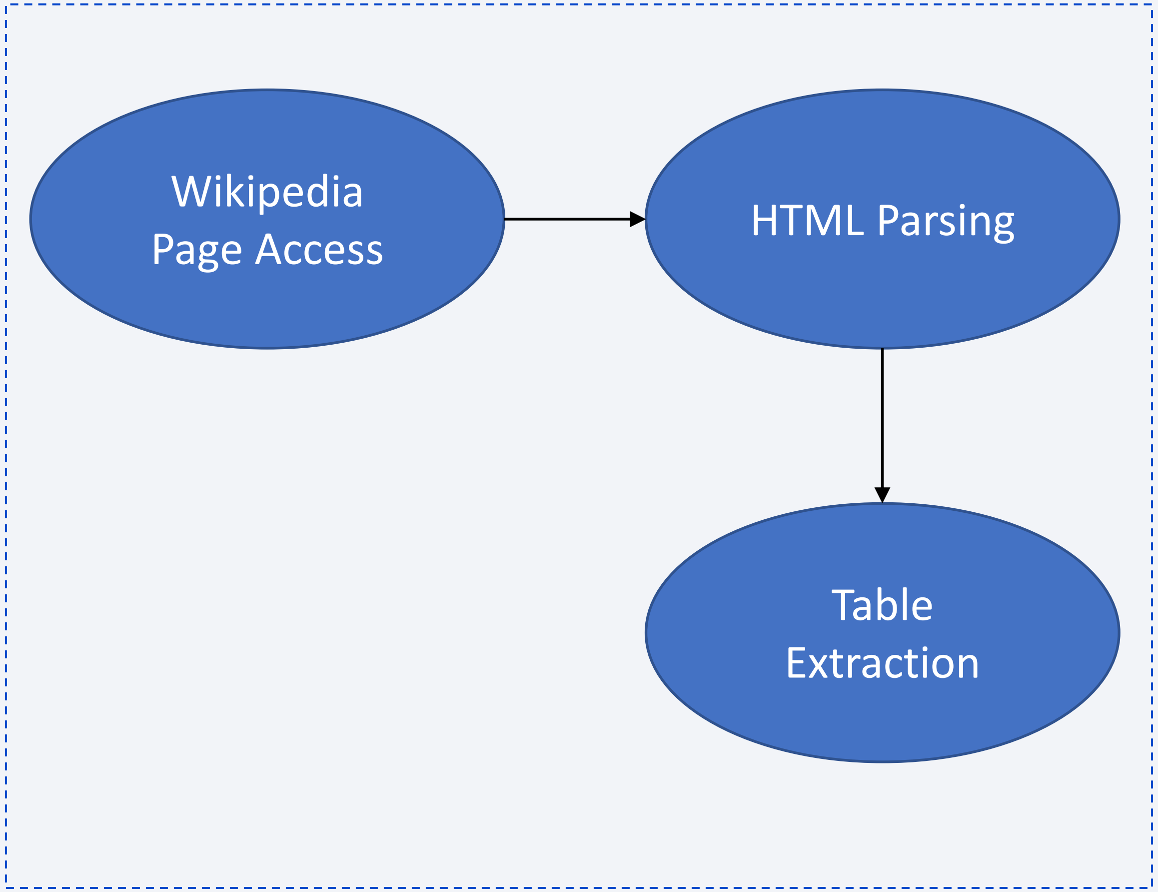
Data Collection – SpaceX API

- **API Request :** We utilized the SpaceX REST API to retrieve detailed information on Falcon 9 Launches,
- **Data Points:** The API provided structured data, including launch dates, payload information, orbit details, launch outcomes, etc,
- **Automated Queries:** Repeated automated queries were performed to fetch the latest data directly from the API
- SpaceX API calls notebook : https://github.com/Tahsine/SpaceX-Capstone/blob/97105c195ca9cd7264a562e4c0a1bd581e933f9c/jupyter-labs-spacex-data-collection-api.ipynb

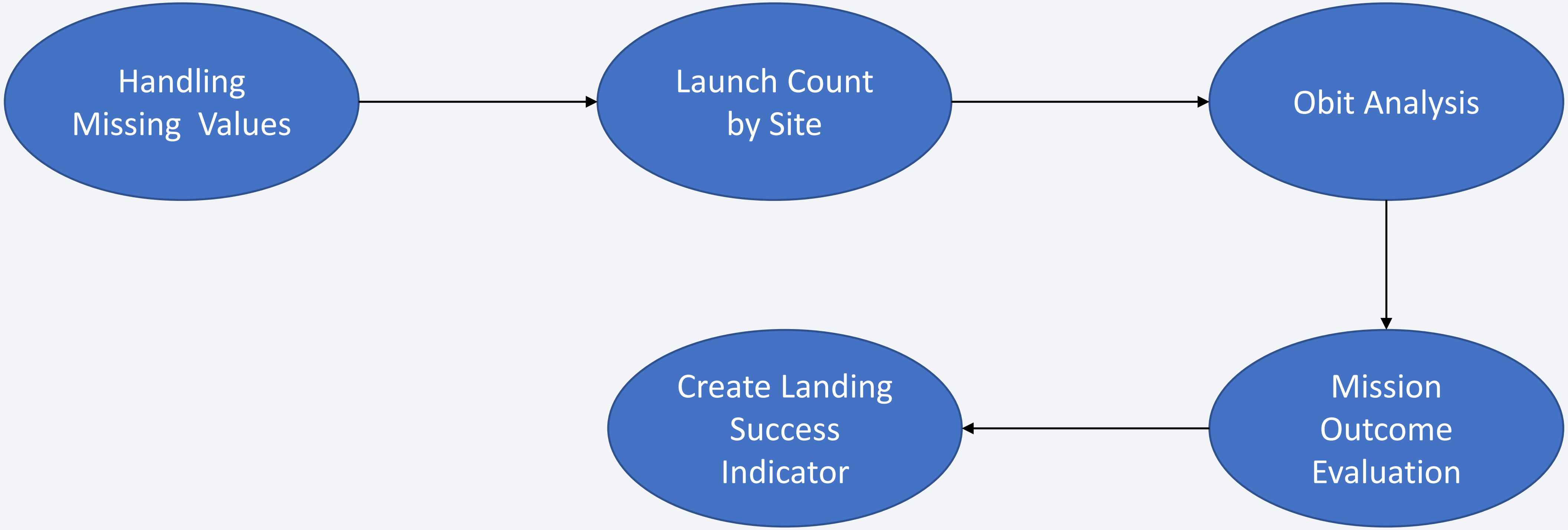


Data Collection - Scraping

- **Web Scraping:** For additional details and historical data on Falcon 9, we implemented web scraping techniques,
- **Target Source:** Wikipedia was selected for its comprehensive tables and consistent formatting related to SpaceX launches,
- **Data Extraction:** HTML parsing was performed to extract tables containing all useful data
- Wikipedia Scraping notebook : https://github.com/Tahsine/SpaceX-Capstone/blob/97105c195ca9cd7264a562e4c0a1bd581e933f9c/jupyter-labs-webscraping.ipynb



Data Wrangling



GitHub URL : https://github.com/Tahsine/SpaceX-Capstone/blob/97105c195ca9cd7264a562e4c0a1bd581e933f9c/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

For visualization, we used:

- Scatter plot:
  - Allow us to explore the relationships between different continuous variables
- Bar Chart
  - Is ideal for comparing success rates between different categories, here orbit types,
- Line Chart
  - Is used to visualize trends over a period of time

GitHub URL : https://github.com/Tahsine/SpaceX-Capstone/blob/97105c195ca9cd7264a562e4c0a1bd581e933f9c/jupyter-labs-eda-dataviz.ipynb

EDA with SQL

SQL queries we used :

- Created a table named SPACEXTABLE to store launch data.
- Display queries
- Query on successful landings
- Query on specific boosters
- Mission Statistics
- Booster analysis:
- Temporal query on landing failures
- Ranking of landing results

GitHub URL : https://github.com/Tahsine/SpaceX-Capstone/blob/97105c195ca9cd7264a562e4c0a1bd581e933f9c/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- We used objects like : Marker, Circle, Lines
  - **Markers** : The markers allowed for a visual distinction between different sites and their outcomes,
  - **Circles** : The size and color of the circles provides a quick visual reference for understanding the significance of each site,
  - **Lines** : These lines helped to visualize the geographical relationships and proximity of key locations to the launch sites,
- GitHub URL : https://github.com/Tahsine/SpaceX-Capstone/blob/97105c195ca9cd7264a562e4c0a1bd581e933f9c/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

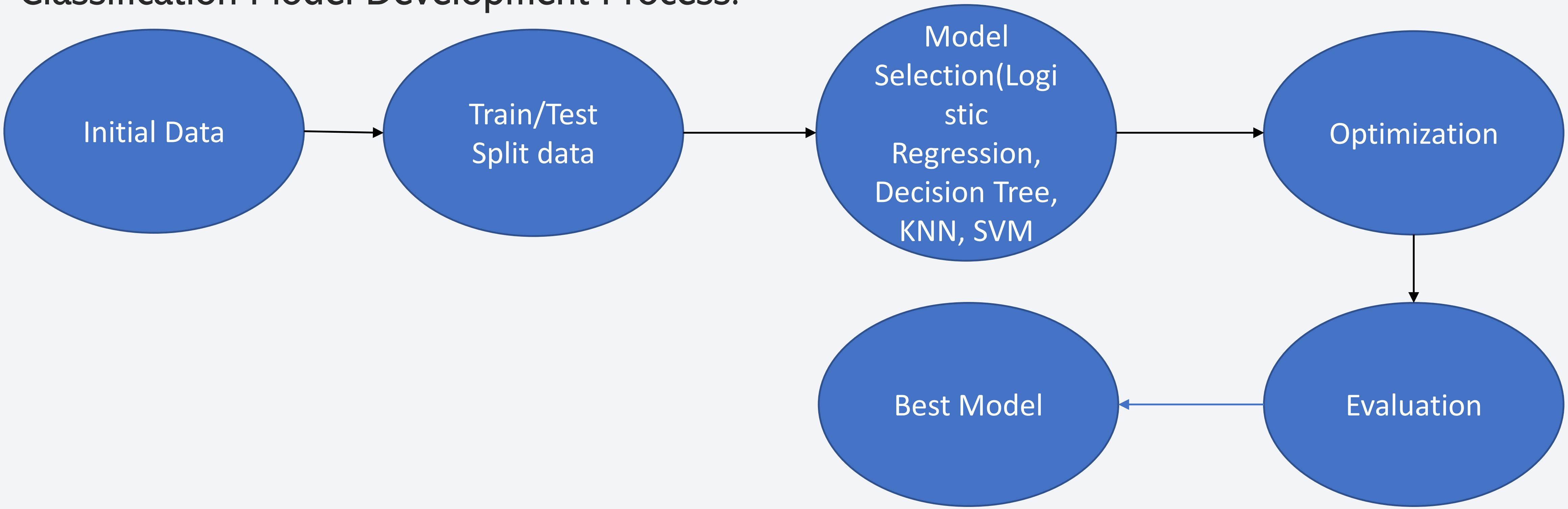
Plots/Graphs we used:

- **Pie Chart** for Successful Launches:
  - The chart updates based on the launch site selected
  - This graph allows us to quickly visualize the distribution of launch successes by site
- **Scatter Plot :**
  - The graph updates based on the selected launch site and user-defined payload range
  - This charts help examine the relationship between payload and launch outcomes,

GitHub URL : https://github.com/Tahsine/SpaceX-Capstone/blob/97105c195ca9cd7264a562e4c0a1bd581e933f9c/spacex_dash_app.py

Predictive Analysis (Classification)

Classification Model Development Process:



GitHub URL: https://github.com/Tahsine/SpaceX-Capstone/blob/97105c195ca9cd7264a562e4c0a1bd581e933f9c/SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

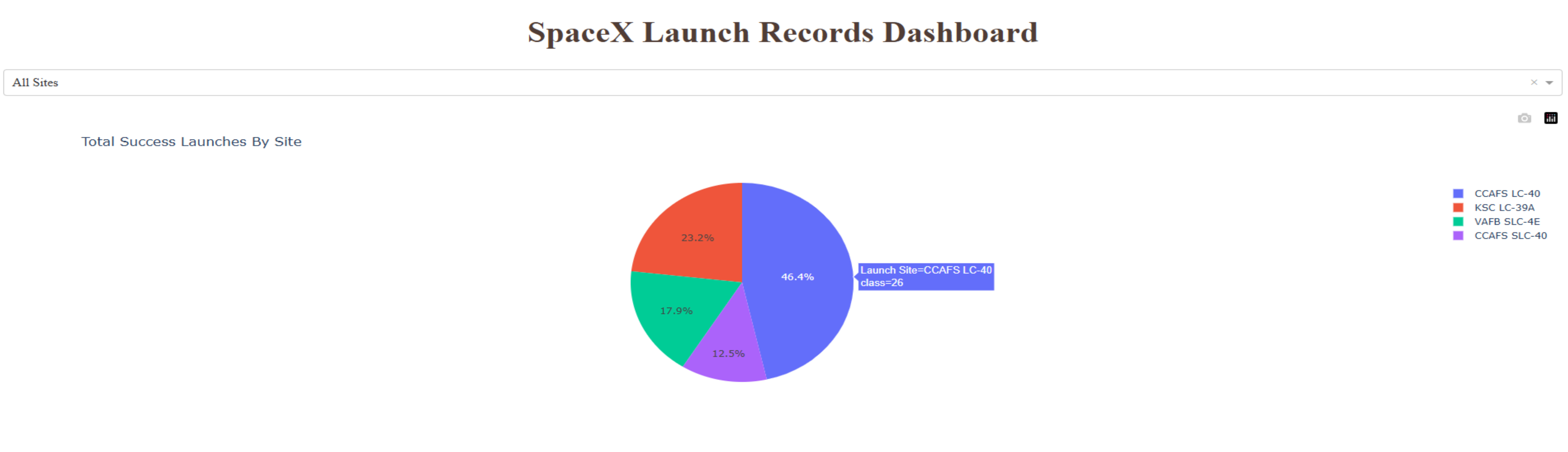
Results

Exploratory data analysis results

- We notice that the success rate is a round 67 %
- Also, the rockets launched from the CCAFS LC-40 launch site are among those with a greater chance of landing success,
- Those coming from the ES-L1, GEO, HEO, SSO orbits are those with the highest landing success rate while those from the SO orbit are almost zero,
- We also observe that the success rate since 2013 kept increasing till 2017,
- The launch sites are far from cities and are closer to the coast so as not to reach civilians,

Results

Interactive analytics demo in screenshots



Results

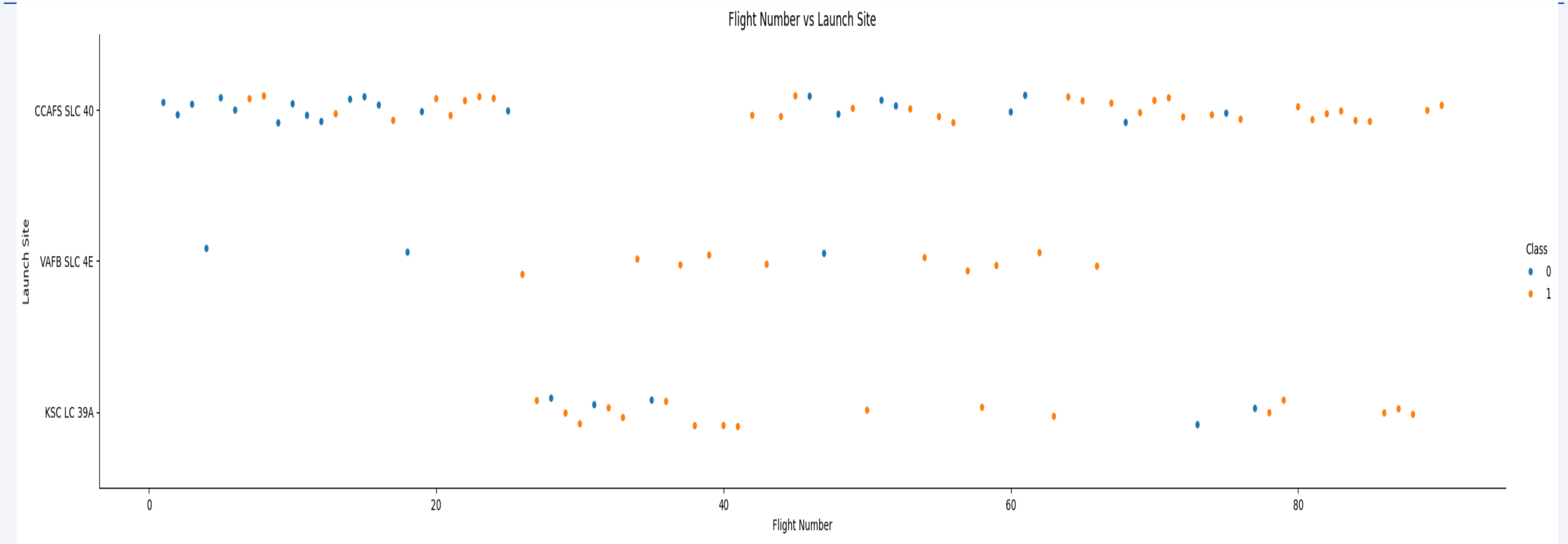
Predictive analysis results

- The four models used all gave good accuracy on the data tested,
- The model chosen is the Decision Tree which presents a slightly better performance than the three others,

Section 2

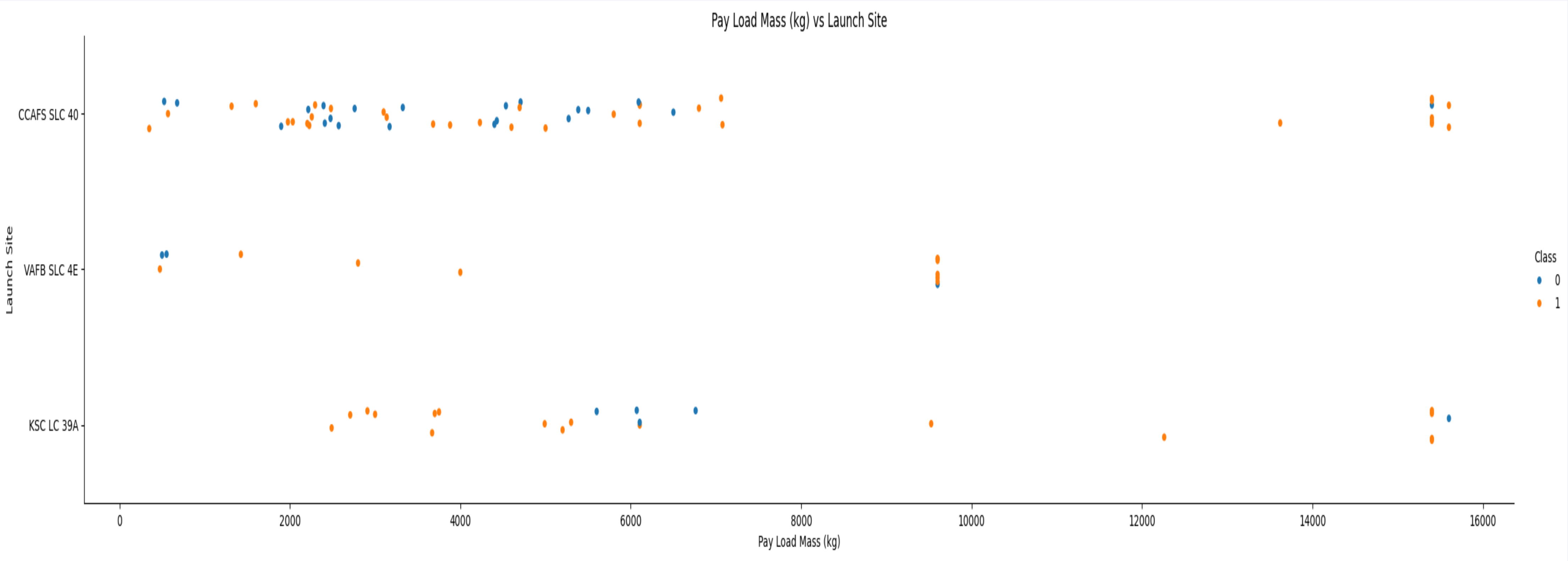
Insights drawn from EDA

Flight Number vs. Launch Site



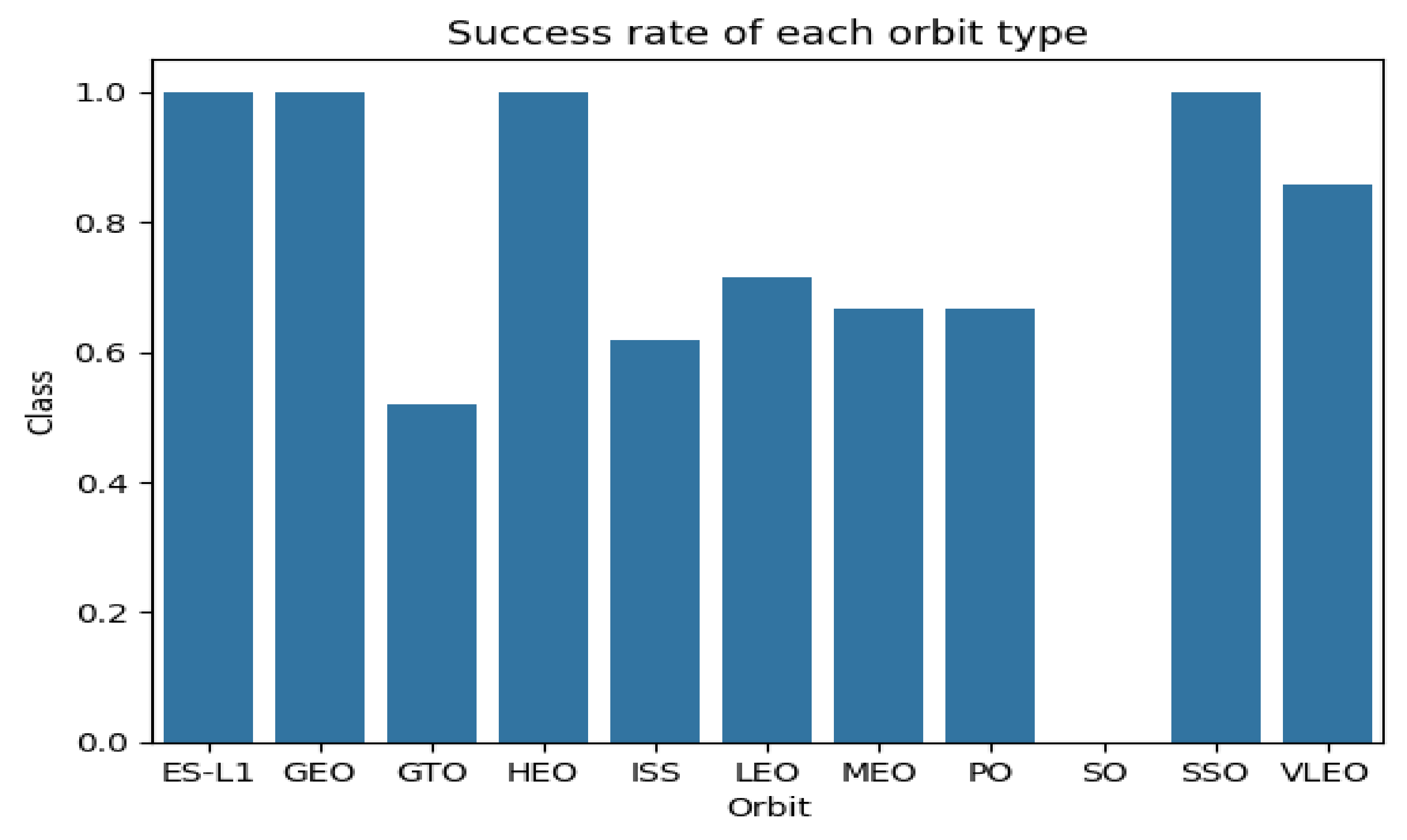
Launch site distribution based on Flight Number. The points are colored to correspond to the two landing outcome categories.

Payload vs. Launch Site



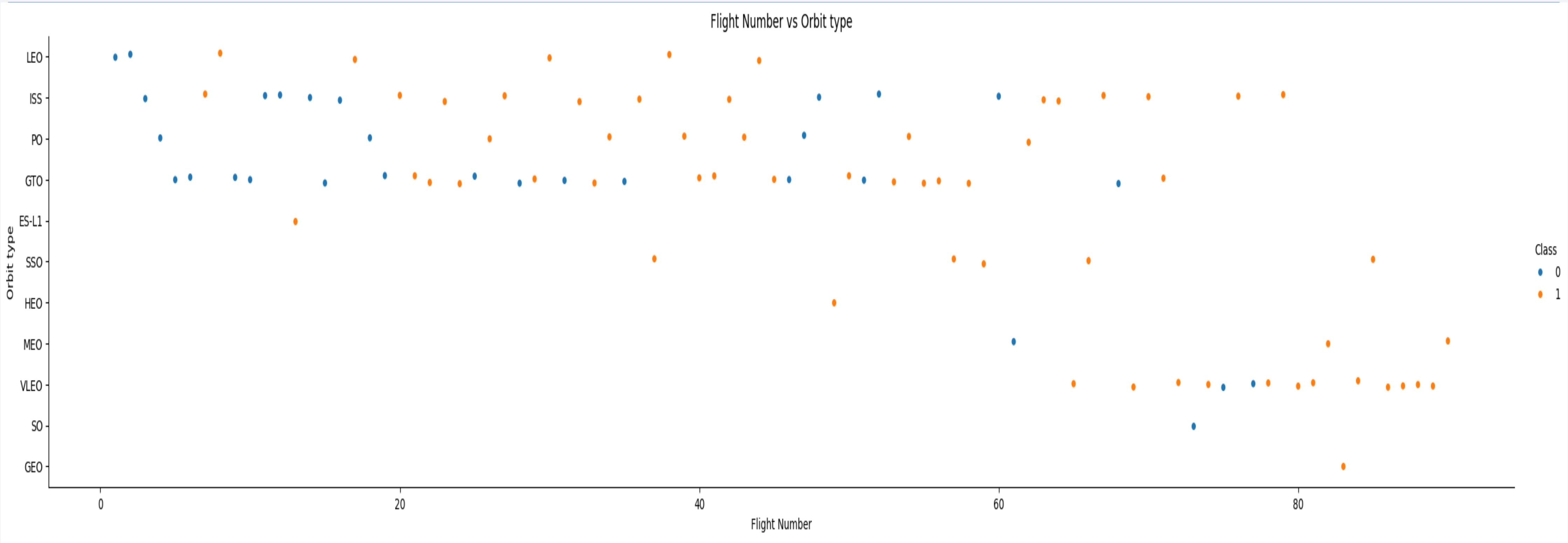
Launch site distribution based on payload mass. The points are colored to correspond to the two categories of landing results.

Success Rate vs. Orbit Type



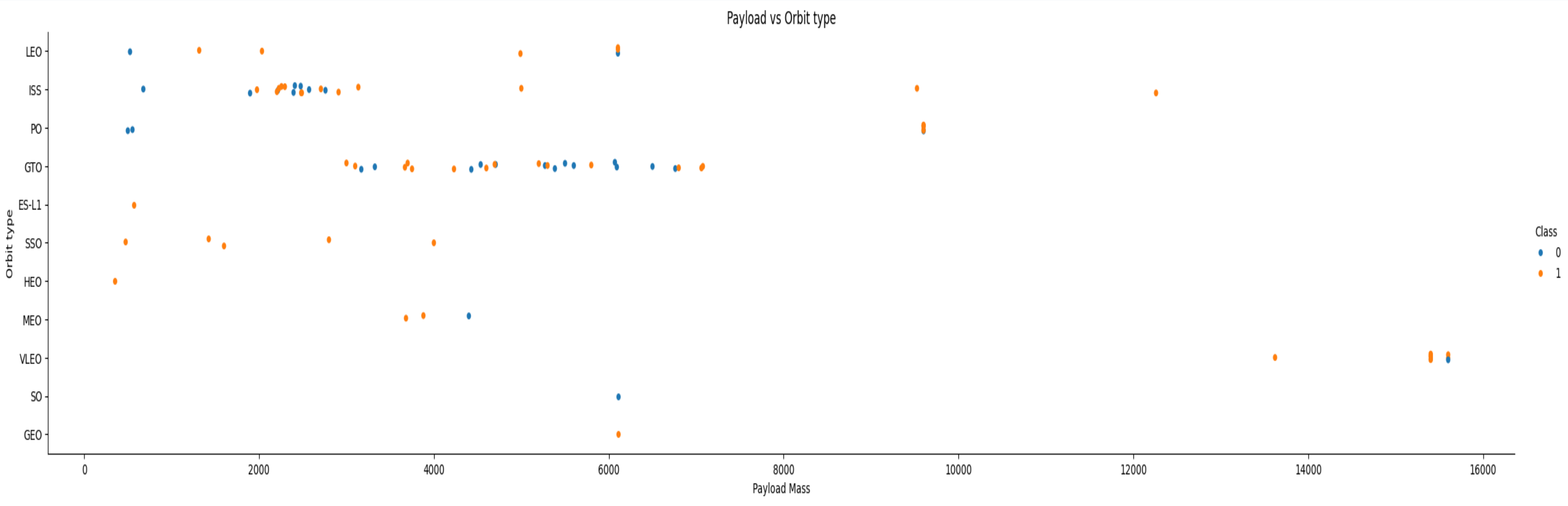
Bar charts showing success rate by orbit type.

Flight Number vs. Orbit Type



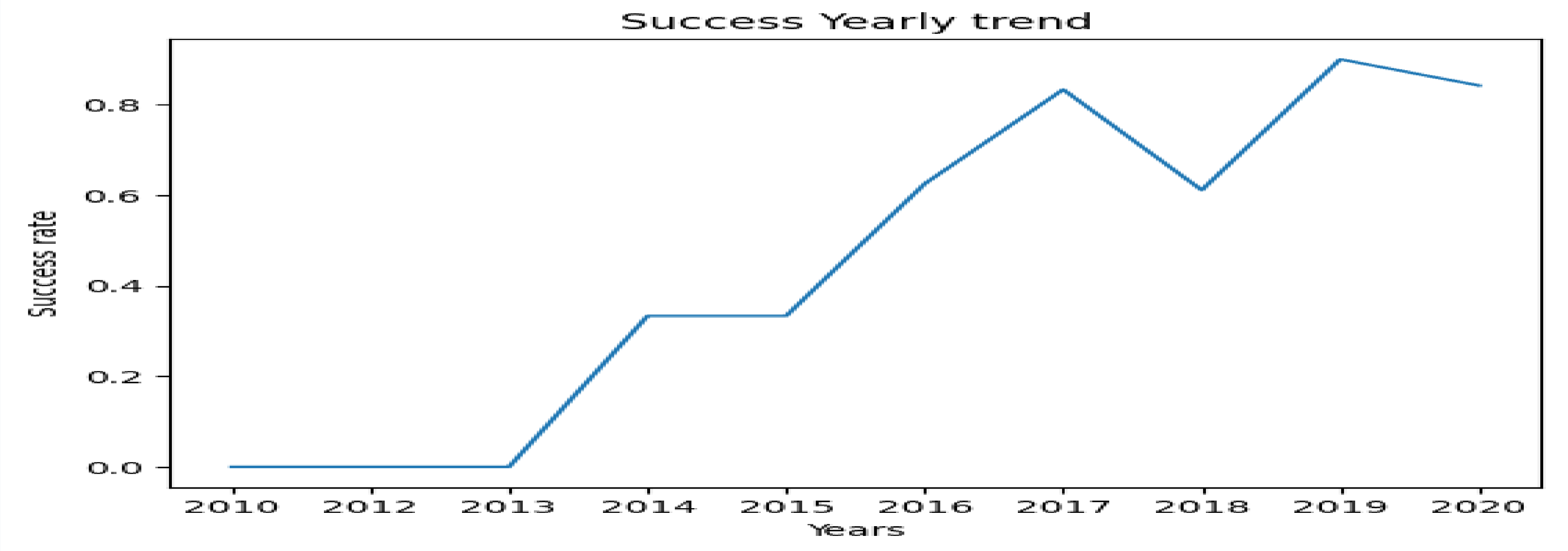
Orbit distribution based on flight number. The points are colored to correspond to the two categories of landing results.

Payload vs. Orbit Type



Orbit distribution based on pay load mass. The points are colored to correspond to the two categories of landing results.

Launch Success Yearly Trend



Graph showing the evolution of the success rate over the years

All Launch Site Names

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

Display of the four launch sites,

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

Display of information concerning launch sites whose name begins with 'CCA'.

Total Payload Mass

Customer	sum(PAYLOAD_MASS__KG_)
NASA (CRS)	45596

Total Payload Mass used by NASA (CRS),

Average Payload Mass by F9 v1.1

Booster_Version	avg(PAYLOAD_MASS__KG_)
F9 v1.1 B1003	2534.6666666666665

Average payload mass by the booster version F9 v1.1

First Successful Ground Landing Date

	Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS_KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
	2015-12-22	1:29:00	F9 FT B1019	CCAFS LC-40	OG2 Mission 2 11 Orbcomm-OG2 satellites	2034	LEO	Orbcomm	Success	Success (ground pad)

Date, and other useful information on the first successful ground landing date

Successful Drone Ship Landing with Payload between 4000 and 6000

Booster_Version	Landing_Outcome	PAYLOAD_MASS__KG_
F9 FT B1022	Success (drone ship)	4696
F9 FT B1026	Success (drone ship)	4600
F9 FT B1021.2	Success (drone ship)	5300
F9 FT B1031.2	Success (drone ship)	5200

Names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000,

Total Number of Successful and Failure Mission Outcomes

Mission_Outcome	count(Mission_Outcome)
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

The total number of successful and failure mission outcomes,

Boosters Carried Maximum Payload

Booster_Version
F9 v1.0 B0003
F9 v1.0 B0004
F9 v1.0 B0005
F9 v1.0 B0006
F9 v1.0 B0007
F9 v1.1 B1003
F9 v1.1
F9 v1.1 B1011
F9 v1.1 B1010

Names of the booster which have carried the maximum payload mass

2015 Launch Records

Month	Year	Booster_Version	Launch_Site	Landing_Outcome
01	2015	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
04	2015	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

List of the failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015,

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

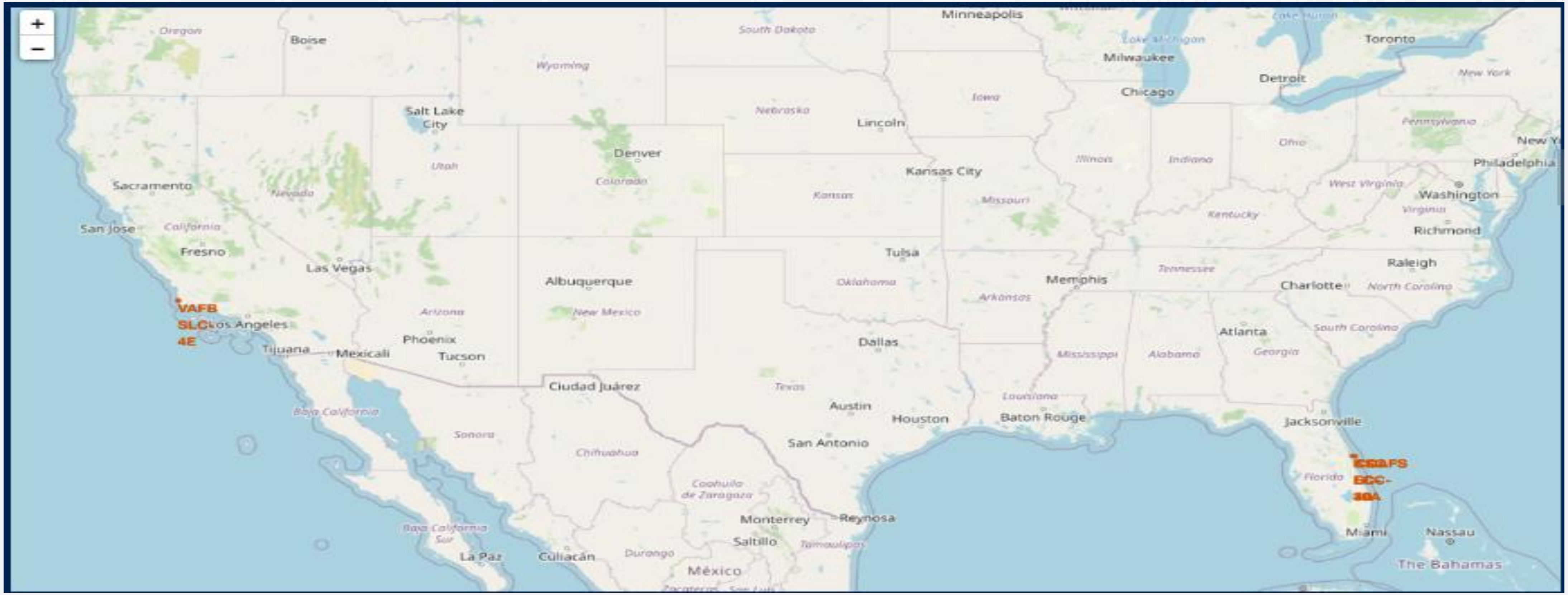
Date	Landing_Outcome	Counts
2012-05-22	No attempt	10
2016-04-08	Success (drone ship)	5
2015-01-10	Failure (drone ship)	5
2015-12-22	Success (ground pad)	3
2014-04-18	Controlled (ocean)	3
2013-09-29	Uncontrolled (ocean)	2
2010-06-04	Failure (parachute)	2
2015-06-28	Precluded (drone ship)	1

Landing outcomes ranking between the date 2010-06-04 and 2017-03-20.

Section 3

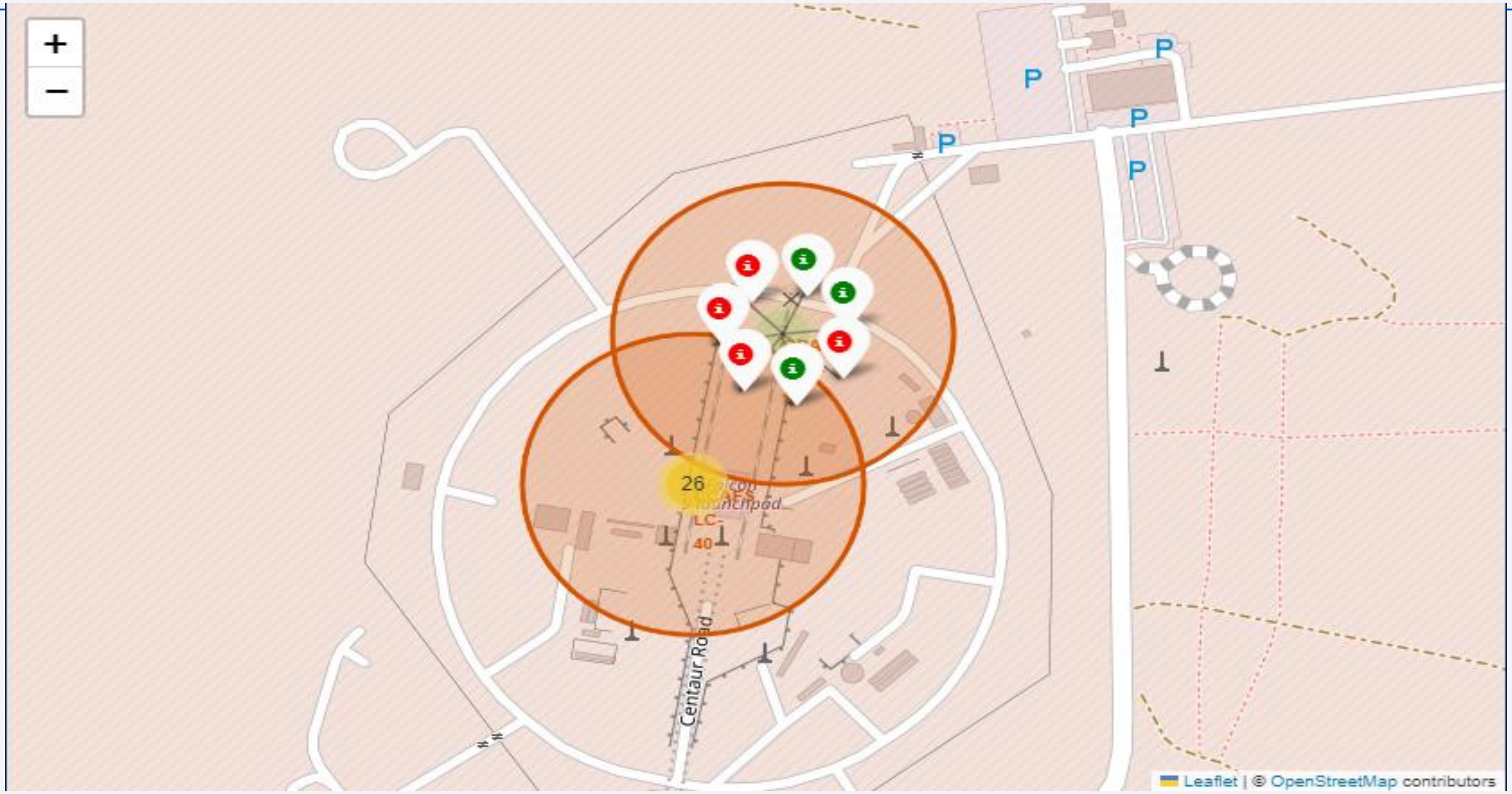
Launch Sites Proximities Analysis

Marked launch sites



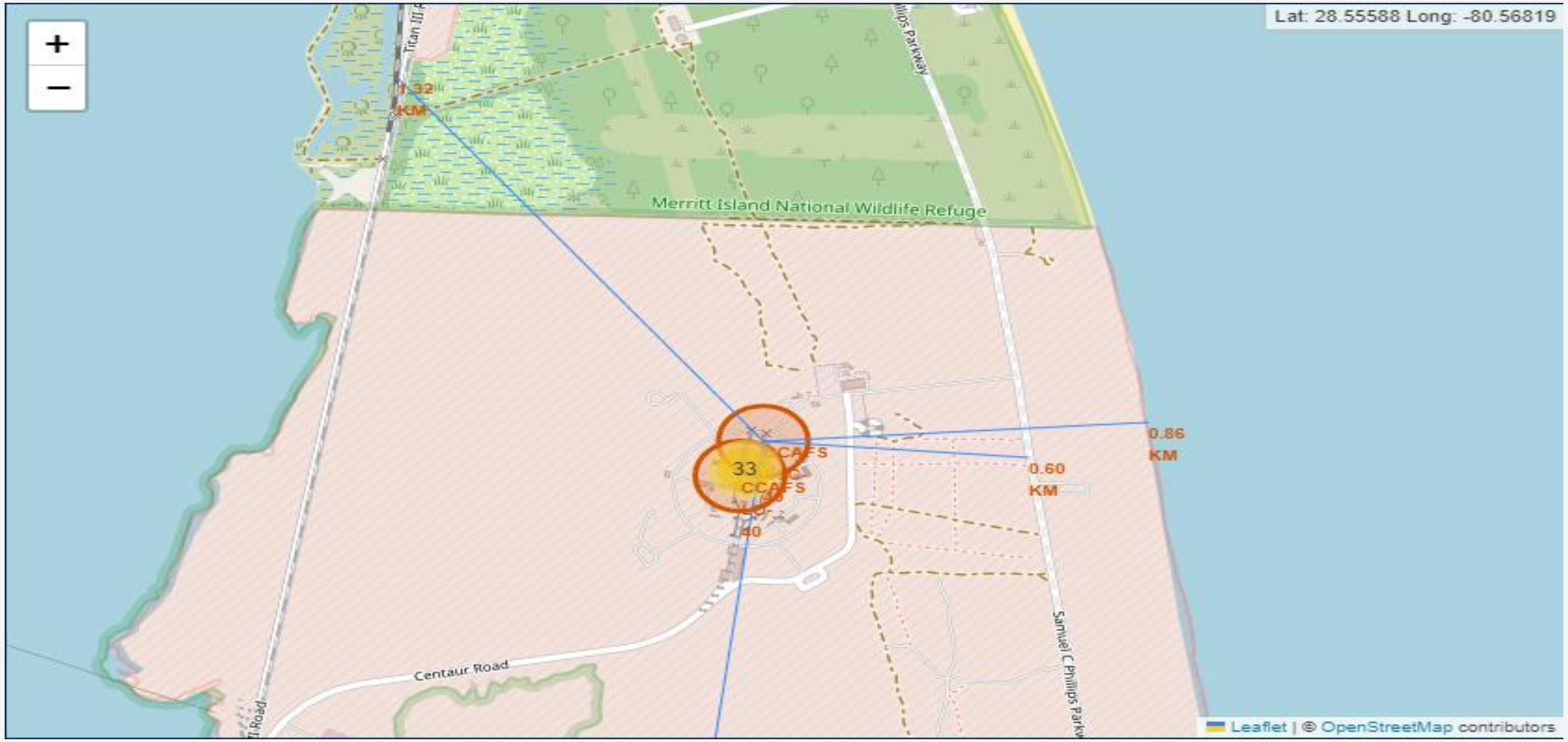
Red marking of launch sites

Successful and Unsuccessful landing zones



Marking of successful landing zones in green and unsuccessful landing zones in red.

Selected launch site to its proximities

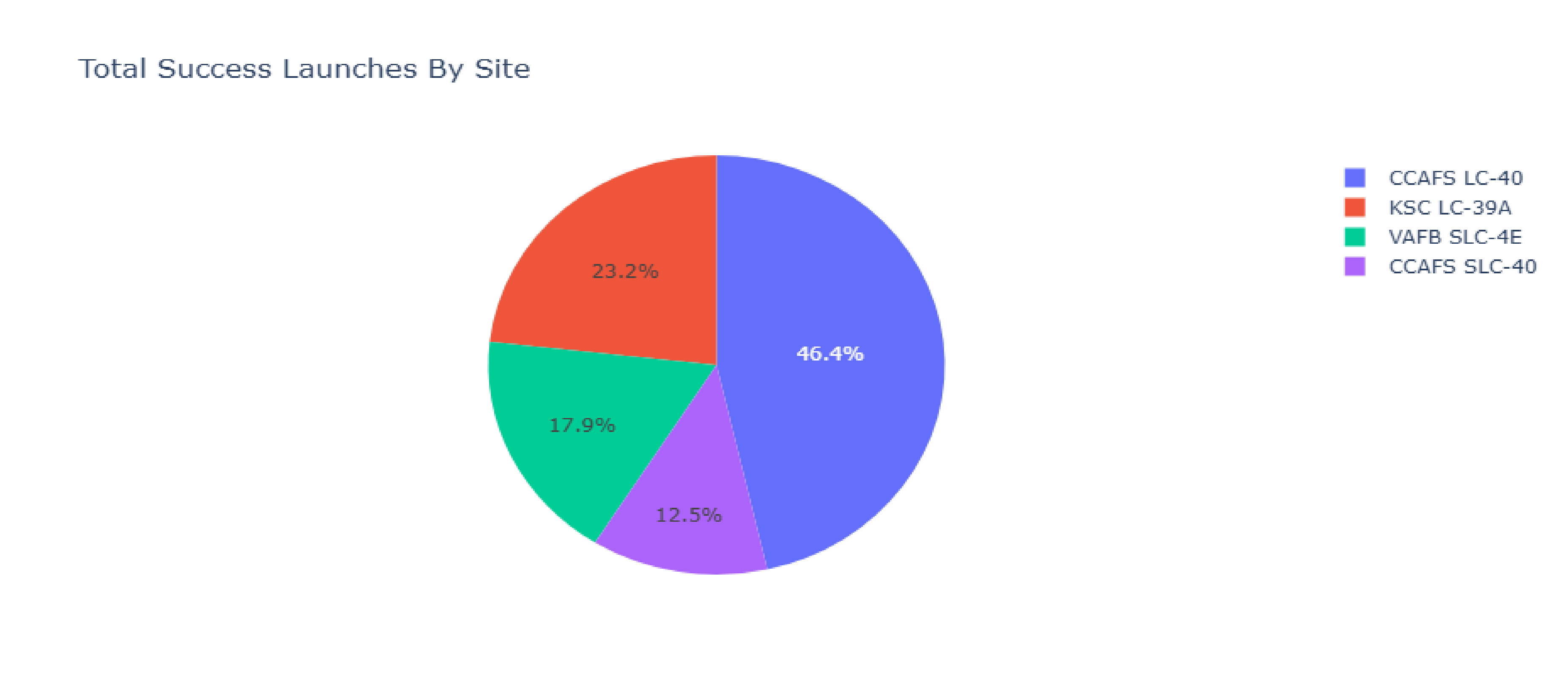


Closest coast, city, railway and highway to the selected launch site.

Section 4

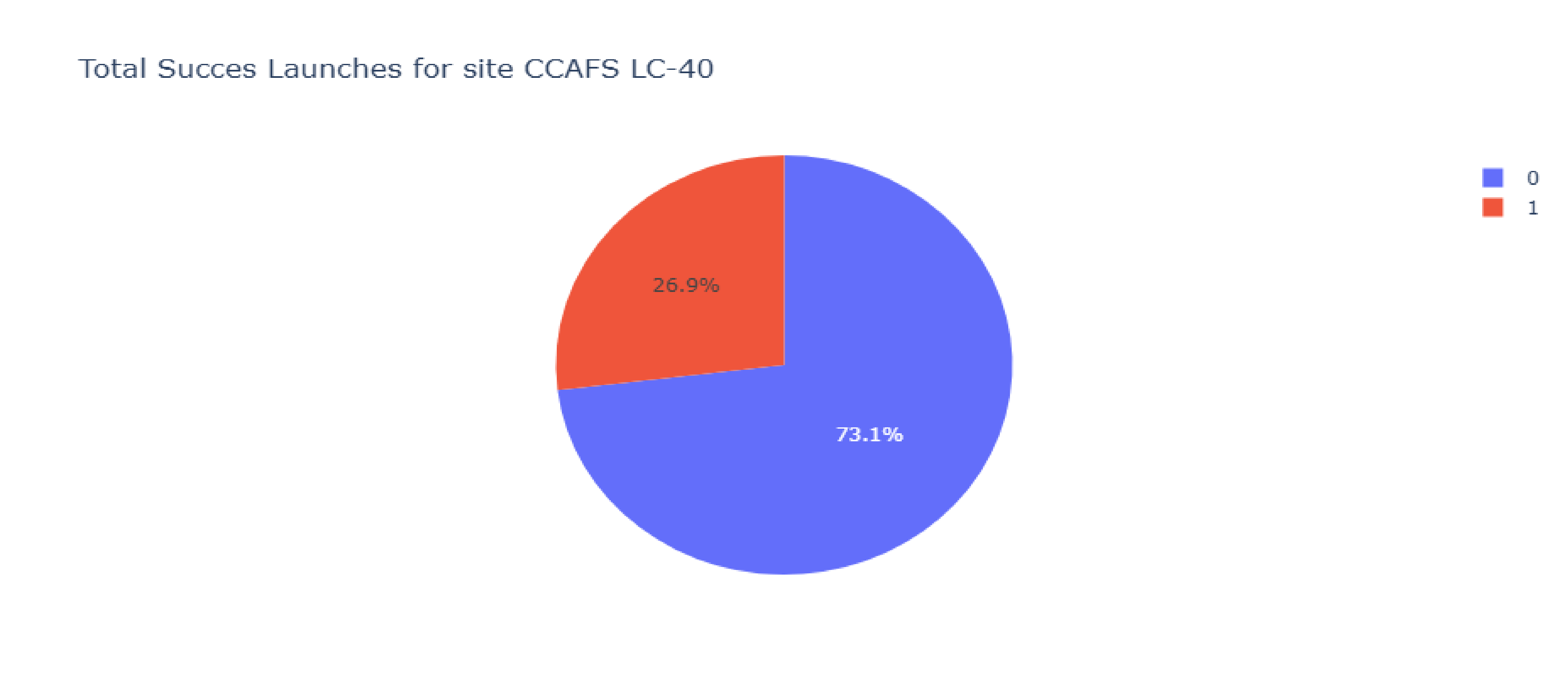
Build a Dashboard with Plotly Dash

Launch Success for All sites



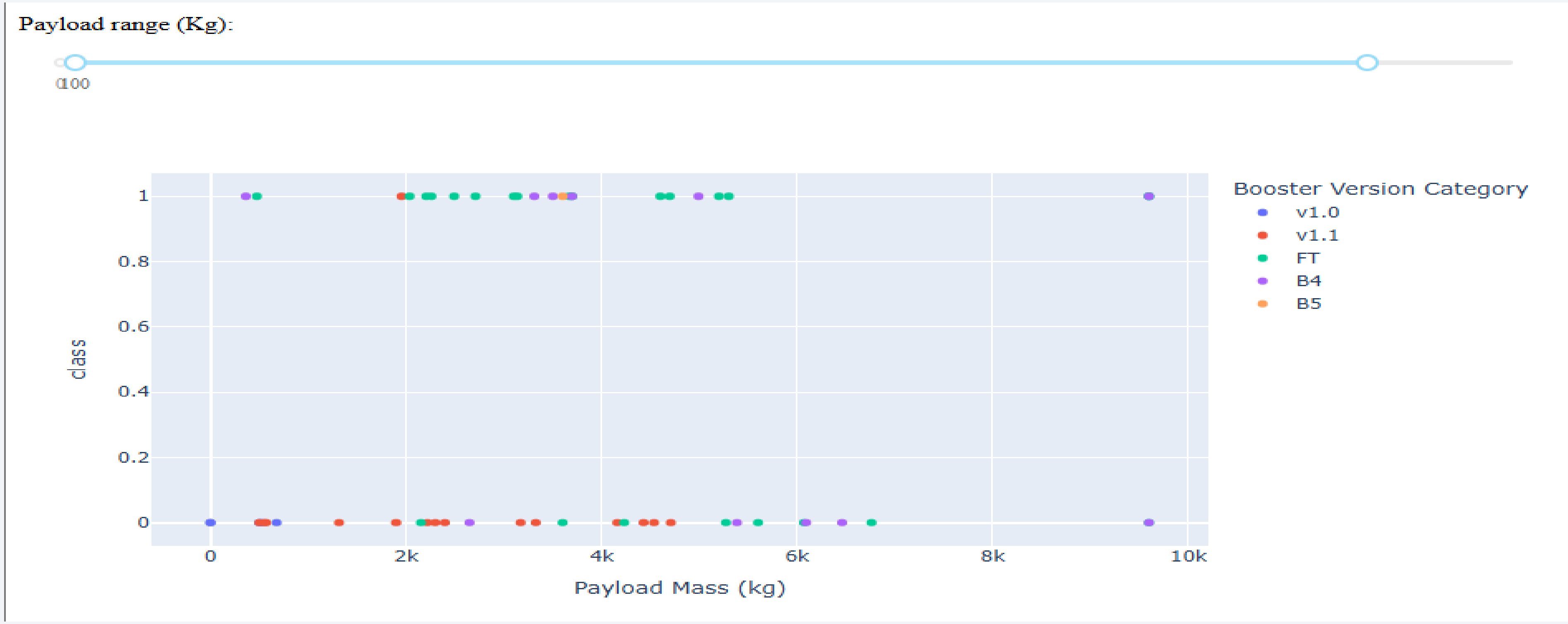
Rockets launched from the CCAFS LC-40 launch site are among those with the greatest chance of landing successfully, and those from CCAFS SLC-40 have the least success.

Launch Site with highest launch success ratio



We observe a successful landing rate of 73.1%. Which is relatively encouraging.

Payload vs Launch Outcome

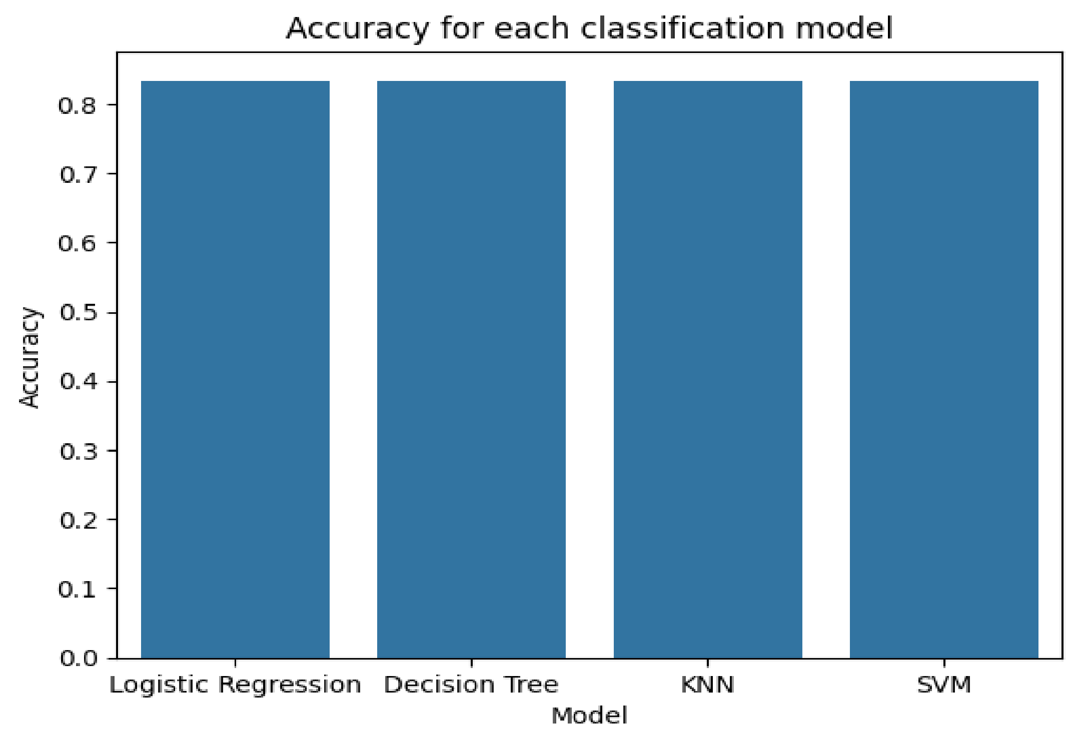


B5 appears to have the highest success rate, while other versions show mixed results.

Section 5

Predictive Analysis (Classification)

Classification Accuracy

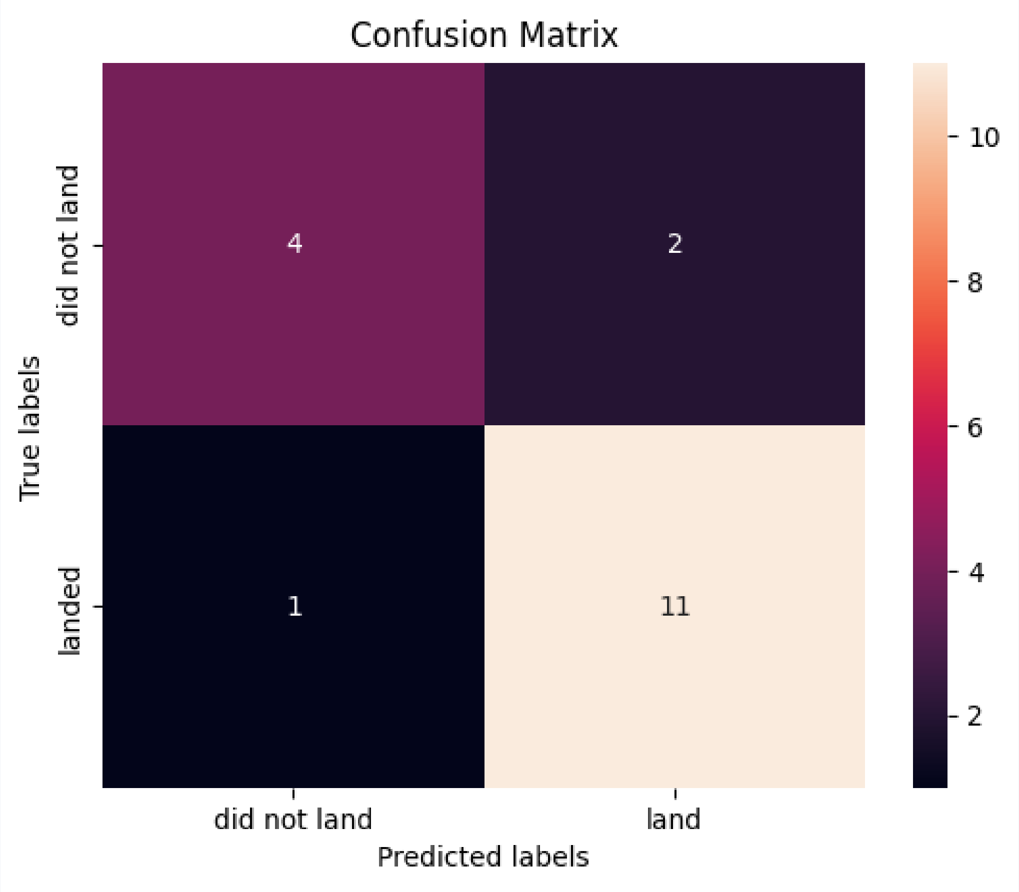


	precision	recall	f1-score	support
0	0.80	0.67	0.73	6
1	0.85	0.92	0.88	12
accuracy			0.83	18
macro avg	0.82	0.79	0.80	18
weighted avg	0.83	0.83	0.83	18

We notice that the Decision Tree model presents the same accuracy as all the other models, but presents a balance in the score of each class. The most efficient model is then the Decision Tree!

Confusion Matrix

Decision Tree Confusion Matrix



- **High True Positive Rate (11/12)**: The model has a good ability to correctly predict when a rocket will land, with only one instance misclassified as "did not land."
- **Reasonable True Negative Rate (4/6)**: The model also has a good capability to predict when a rocket will not land, with two instances misclassified as "landed."
- **Low Error Rates**: The model has relatively low false positive and false negative rates, indicating strong predictive power.
- This confusion matrix reflects a model with a high degree of accuracy and balanced performance, particularly in identifying successful landings, which is crucial for predicting cost savings in SpaceX launches.

Conclusions

Our comprehensive analysis of SpaceX launch data has yielded the following key insights:

1. **Landing Success Probability:** The probability of a successful landing stands at approximately 67%. This metric underscores the advancements SpaceX has made in its landing technology, yet it also points to areas where improvements can still be made to reach higher consistency.
2. **Optimal Launch Site - CCAFS LC-40:** The CCAFS LC-40 launch site emerges as the most reliable option, boasting a significantly higher success rate compared to the other three sites. This finding suggests that CCAFS LC-40 is strategically advantageous for launches, likely due to its infrastructure, location, and operational efficiency. Future missions may benefit from prioritizing this site to maximize the likelihood of successful landings.

Innovative Insights:

- **Site Selection for Maximized Success:** The success rate variation across different launch sites provides a valuable criterion for site selection in future launches. By leveraging this data-driven insight, SpaceX can strategically allocate resources and plan missions to enhance overall mission success.
- **Data-Driven Decision Making:** The use of comprehensive data analytics in evaluating launch outcomes and site performance offers a robust framework for optimizing space missions. This approach not only improves operational efficiency but also contributes to the broader field of aerospace innovation.

Appendix

- https://github.com/Tahsine/SpaceX-Capstone.git
- https://github.com/Tahsine/SpaceX-Capstone/blob/97105c195ca9cd7264a562e4c0a1bd581e933f9c/spacex_launch_dash.csv
- https://github.com/Tahsine/SpaceX-Capstone/blob/97105c195ca9cd7264a562e4c0a1bd581e933f9c/Spacex.csv

Thank you!