IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Khaled Seleem
11 August 2024

Outline

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

Executive Summary

Summary of methodologies

- Data Preparation: Collect and clean Falcon 9 launch data from SpaceX API and Wikipedia.
- Visualization and Analysis: Conduct exploratory analysis using SQL, Folium for mapping, and Plotly Dash for interactive visualizations.
- Predictive Modelling: Build and refine machine learning models to predict launch outcomes.

Summary of results

- CCAFS SLC 40 Superior with larger payloads while KSC LC 39A More effective with smaller payloads.
- Highest success rates in ES-L1, SSO, HEO, and GEO orbits with marked improvements in LEO over time.
- Notable annual increases in landing success, especially for LEO and ISS orbits with heavy payloads.
- KSC LC-39A leads in launch success, FT boosters and medium payloads (2k-4k kg) show peak performance.
- Random Forest model: High accuracy (0.88), excels in predicting positive outcomes.

Introduction

- Space X advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upward of 165 million dollars each.
- Much of the savings is because Space X can reuse the first stage.
- If we can determine if the first stage will land, we can determine the cost of a launch.

Section 1

Methodology

Methodology

- Data collection:
  - The falcon 9 launch data collected from space x API and Wikipedia
- Perform data wrangling:
  - Clean and process the date to facilitate drawing insights and building prediction models
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models:
  - Build, tune and evaluate machine learning classification models to correctly predict the outcomes

Data Collection

SpaceX API

- Collecting detailed Falcon 9 launch data (2010-2020) from SpaceX API

Web scraping

- Collecting Falcon 9 historical launch records (2010-2021) from Wikipedia

Data Collection – SpaceX API

1. Requesting and parsing the SpaceX launch data using the GET request
2. Filtering the data frame to only include Falcon 9 launches
3. Dealing with Missing Values

The GitHub URL of the SpaceX API calls notebook:

https://github.com/khaled-ma-seleem/SpaceX-Falcon-9-Landing-Analysis/blob/main/1_data_collection_api.ipynb

Data Collection - Scraping

1. • Requesting the Falcon9 Launch Wiki page from its URL
2. • Extracting all column names from the HTML table header
3. • Creating a data frame by parsing the launch HTML tables

The GitHub URL of the web scraping notebook:

https://github.com/khaled-ma-seleem/SpaceX-Falcon-9-Landing-Analysis/blob/main/2_webscraping.ipynb

Data Wrangling

1. • Calculating the number of launches on each site
2. • Calculating the number and occurrence of each orbit
3. • Calculating the number and occurrence of mission outcome of the orbits
4. • Creating a landing outcome label from Outcome column

The GitHub URL of the data processing notebook:

https://github.com/khaled-ma-seleem/SpaceX-Falcon-9-Landing-Analysis/blob/main/3_data_wrangling.ipynb

EDA with Data Visualization

1. Success vs Flight Number and Launch Site
2. Success vs Payload Mass and Launch Site
3. Success rate vs Orbit
4. Success vs Flight Number and Orbit
5. Success vs Payload Mass and Orbit
6. success rate vs year

The GitHub URL of the data visualization notebook:

https://github.com/khaled-ma-seleem/SpaceX-Falcon-9-Landing-Analysis/blob/main/5_eda_data_viz.ipynb

EDA with SQL

The SQL queries performed

- Names of the unique launch sites in the space missions
- Average payload mass carried by booster version F9 v1.1
- Date when the first successful landing outcome in ground pad was achieved
- Boosters which have success in drone ship and have payload mass between 4000 and 6000
- Total number of successful and failure mission outcomes
- Booster versions which have carried the maximum payload mass

The GitHub URL of EDA with SQL notebook:

https://github.com/khaled-ma-seleem/SpaceX-Falcon-9-Landing-Analysis/blob/main/4_eda_sql.ipynb

Build an Interactive Map with Folium

Analysis performed on the map

- Added a circle and a marker for each launch site to show their locations and all similarities between them on the map.
- Added the launch results for each site on the map, with green for success landing and red for failure.
- Calculated the distances between a launch site to its proximities to present which places are near to the launch sites and which are far.

The GitHub URL of the interactive maps with Folium:

https://github.com/khaled-ma-seleem/SpaceX-Falcon-9-Landing-Analysis/blob/main/6_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

Interactive dashboard analysis

- Added pie chart for the distribution of successful landings over launch sites
- Add a pie chart for the success rate of each launch site
- Add scatter plots to show the relation between the success and payload mass and booster version over all launch sites.

The GitHub URL of your completed Plotly Dash python file:

https://github.com/khaled-ma-seleem/SpaceX-Falcon-9-Landing-Analysis/blob/main/7_dash_app.py

Predictive Analysis (Classification)

1. • Built machine learning classification models with logistic regression, SVM, decision tree, random forest and KNN.
2. • Used different hyper parameters with each model.
3. • Trained the models with the training data and tested them on the test data
4. • Calculated the confusion matrix for each model.
5. • Select the random forest model after performing the best on prediction.

The GitHub URL of the predictive analysis notebook:

https://github.com/khaled-ma-seleem/SpaceX-Falcon-9-Landing-Analysis/blob/main/8_prediction.ipynb

Results

Exploratory data analysis results

- Payload Mass: CCAFS SLC 40 excels with larger payloads, while KSC LC 39A performs better with smaller payloads.
- Orbit Performance: Highest success rates found in ES-L1, SSO, HEO, and GEO orbits, with notable improvements over time in LEO orbit success.
- Landing Success Trends: Increased landing success rates over the years, especially for LEO and ISS orbits with heavy payloads.

Interactive analytics results

- Geographic Features: All launch sites are strategically located near the equator, coastlines, railways, and highways while maintaining a safe distance from urban areas.
- Site and Equipment Efficacy: KSC LC-39A shows the highest launch success; FT boosters and payloads between 2k and 4k kg present the best performance.

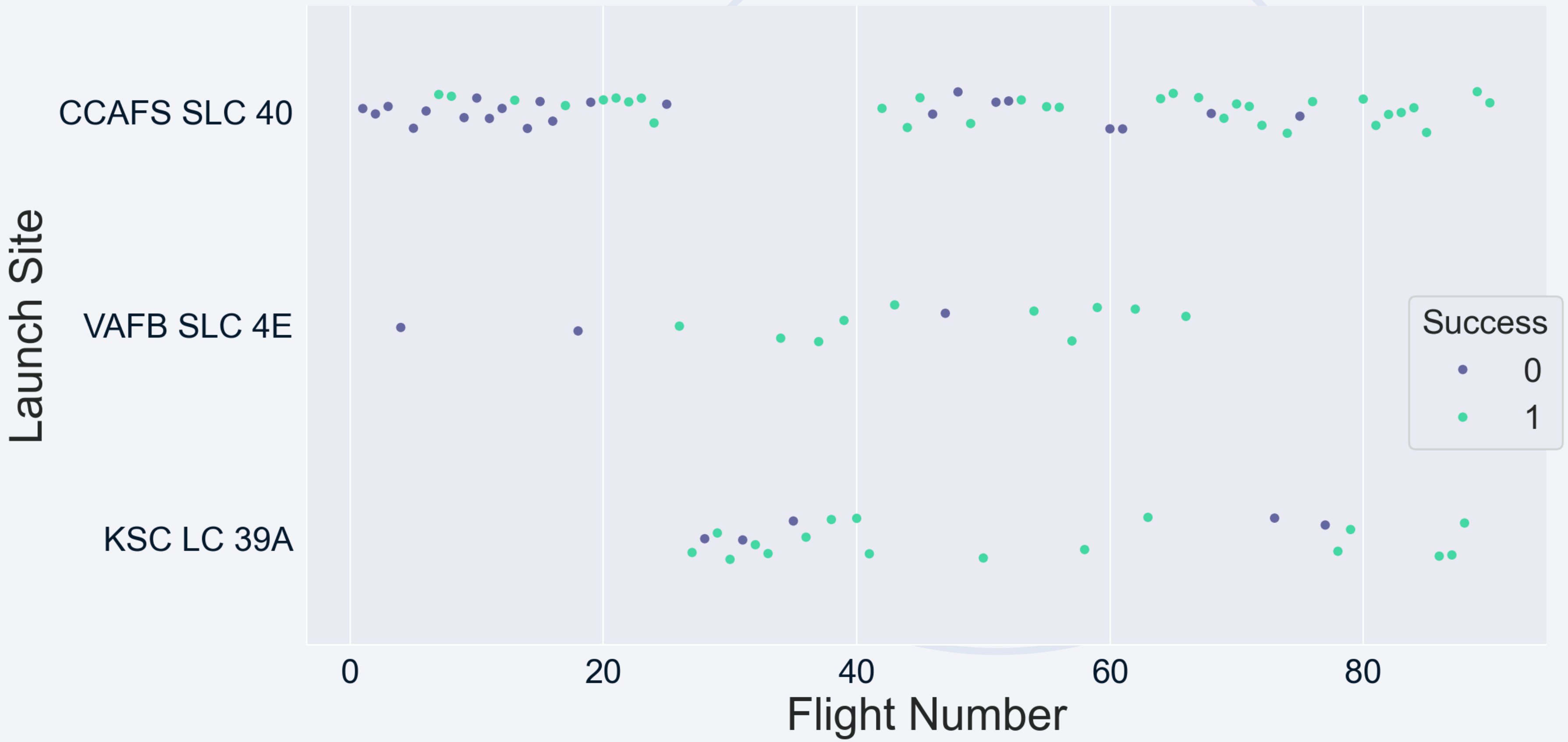
Predictive analysis results

- Model Performance: The Random Forest model achieved the highest accuracy (0.88), excelling in predicting positive outcomes but less so with negative ones.

Section 2

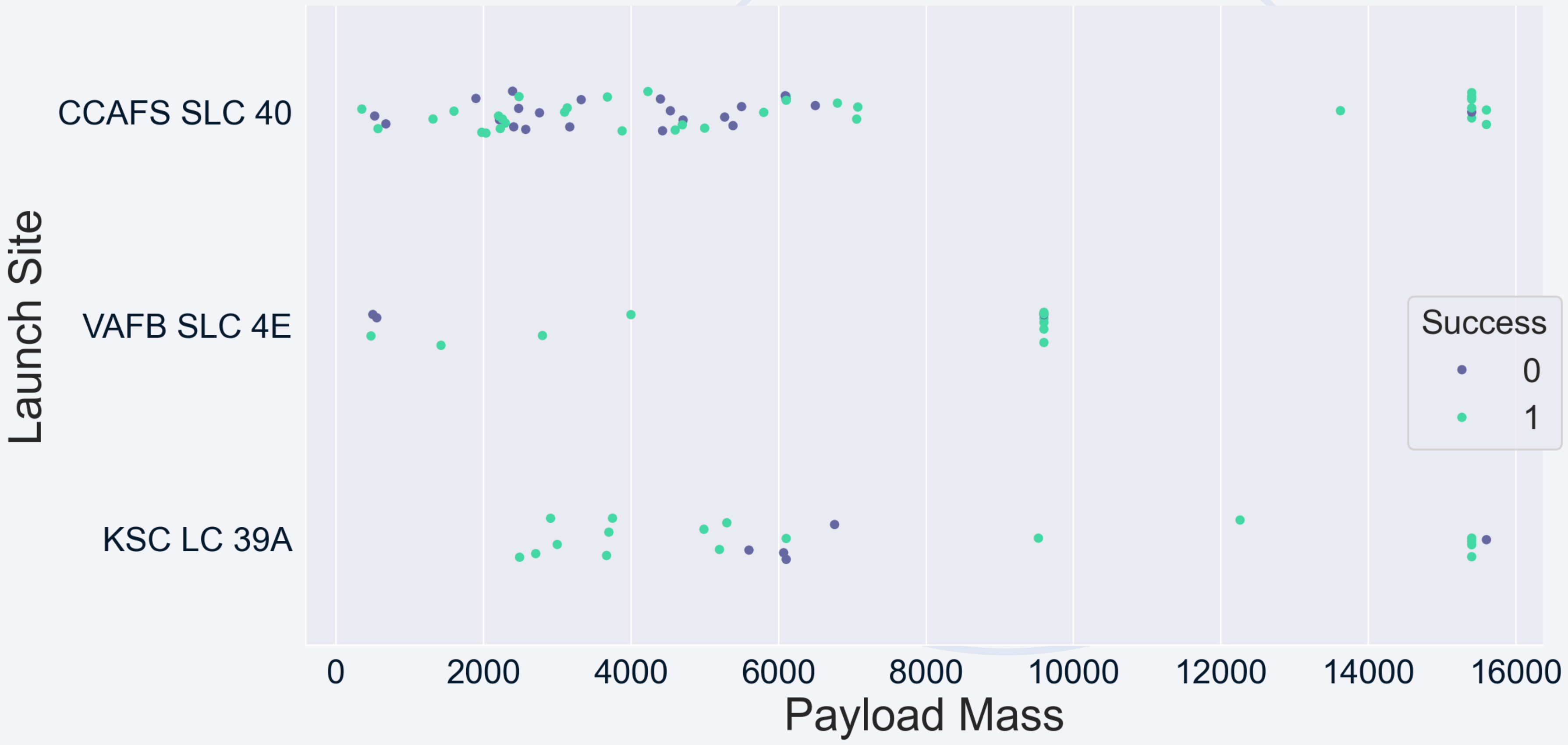
Insights drawn from EDA

Flight Number vs. Launch Site



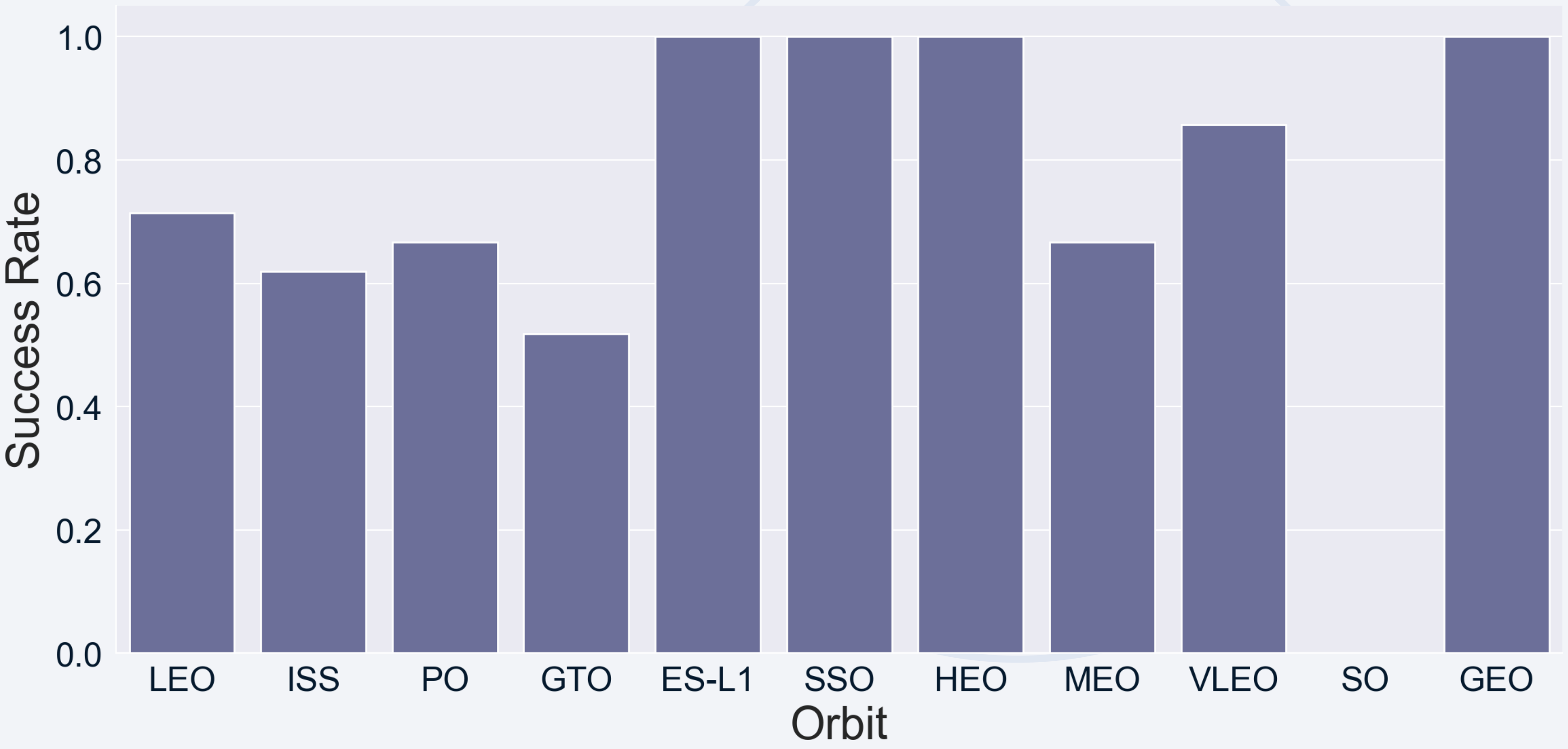
- There is a positive relationship between flight number and success rate for all 3 launch sites.

Payload vs. Launch Site



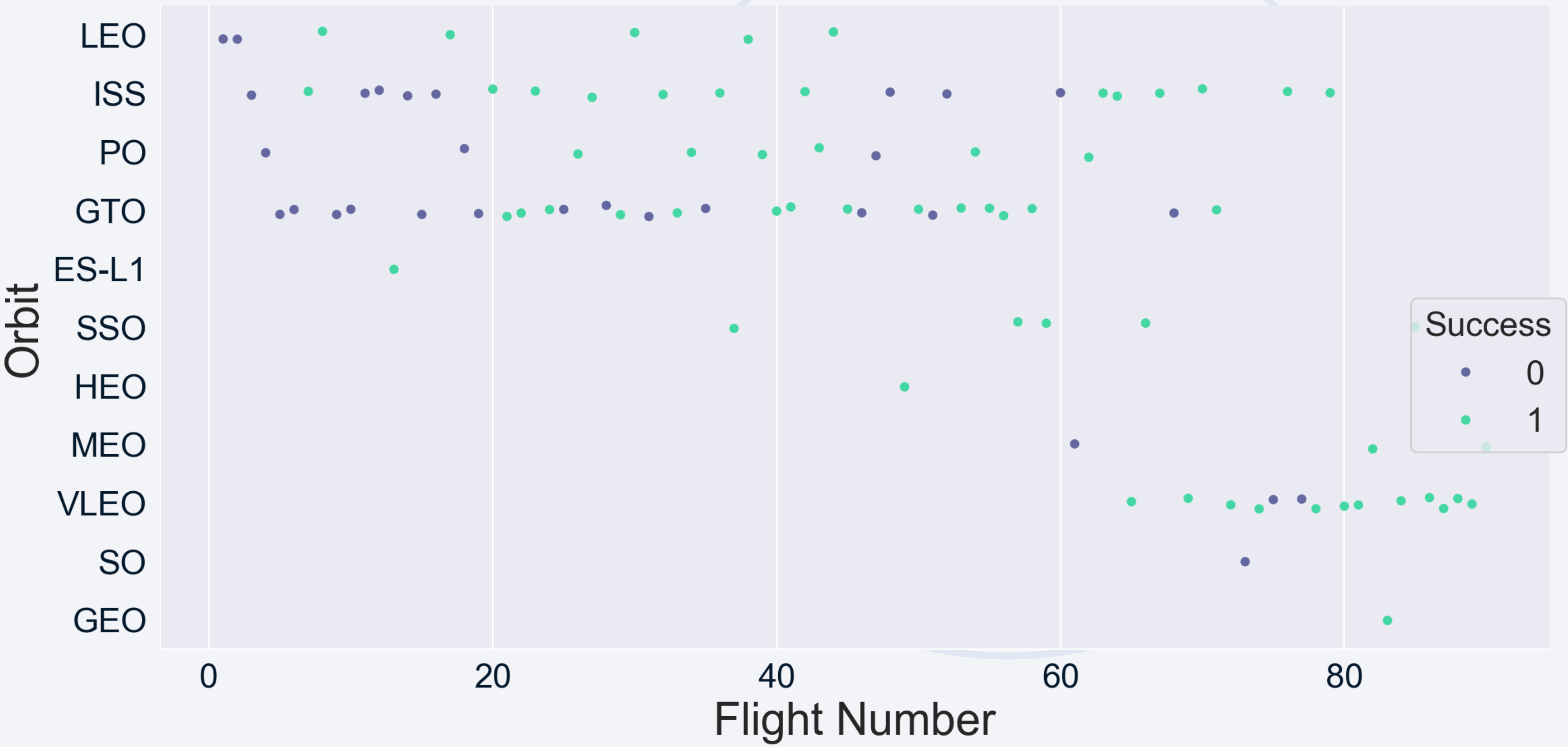
- CCAFS SLC 40 has higher success rate for larger payload masses.
- KSC LC 39A has higher success rate for smaller payload masses.

Success Rate vs. Orbit Type



- The orbits with the highest success rates are ES-L1, SSO, HEO, and GEO.

Flight Number vs. Orbit Type

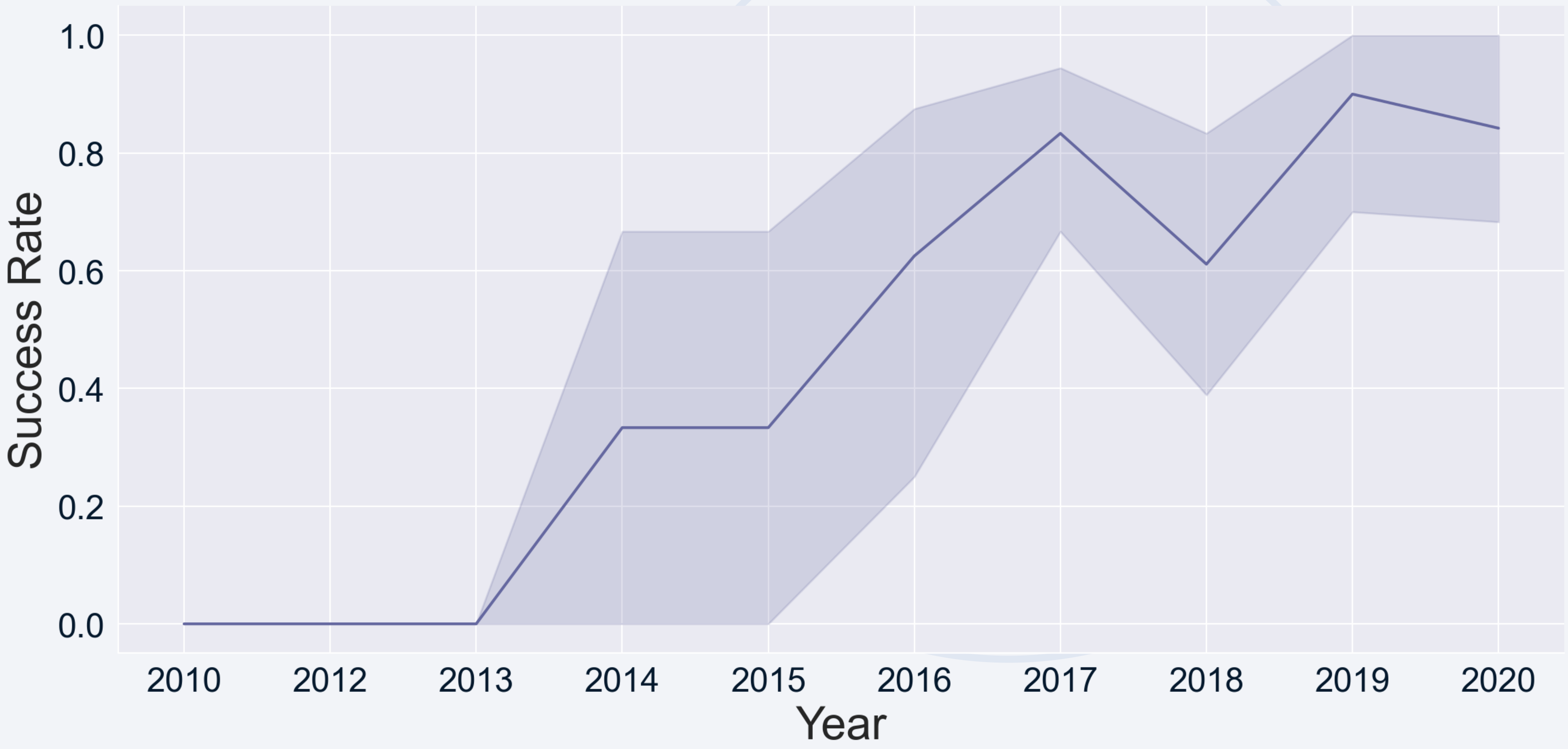


- There is a high improvement in success rate over time for the LEO orbit, while for other orbits like GTO there is less improvement.

Payload vs. Orbit Type

- LEO and ISS orbits have a high landing success rates with heavy payloads, while the landing success rate from GTO has no clear relation with payload mass.

Launch Success Yearly Trend



- The landing success rate is improving over the years.

All Launch Site Names

Launch sites of Falcon 9 missions:

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

Launch Site Names Begin with 'CCA'

Some records of Falcon 9 missions where launch sites begin with `CCA`

Date	Launch Site	Orbit	Mission Outcome	Landing Outcome
2010-06-04	CCAFS LC-40	LEO	Success	Failure (parachute)
2010-12-08	CCAFS LC-40	LEO (ISS)	Success	Failure (parachute)
2012-05-22	CCAFS LC-40	LEO (ISS)	Success	No attempt
2012-10-08	CCAFS LC-40	LEO (ISS)	Success	No attempt
2013-03-01	CCAFS LC-40	LEO (ISS)	Success	No attempt

Total Payload Mass

Total payload mass launched by NASA (CRS):

45,596 Kg

Average Payload Mass by F9 v1.1

The average payload mass carried by booster version F9 v1.1:

2,928.4 Kg

First Successful Ground Landing Date

The date of the first successful landing on ground pad:

2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

The names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 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

Total Number of Successful and Failure Mission Outcomes

The total number of successful and failure mission outcomes

Mission Outcome	Total Number of Missions
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Boosters Carried Maximum Payload

The names of the booster versions which have carried the maximum payload mass

Booster Version	PAYLOAD MASS (KG)
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

2015 Launch Records

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

Date	Landing_Outcome	Booster_Version	Launch_Site
2015-01-10	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
2015-04-14	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

The count of landing outcomes between the date 2010-06-04 and 2017-03-20, ranked in descending order

Landing Outcome	Count
Success	38
No attempt	21
Success (drone ship)	14
Success (ground pad)	9
Failure (drone ship)	5
Controlled (ocean)	5
Failure	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1
No attempt	1

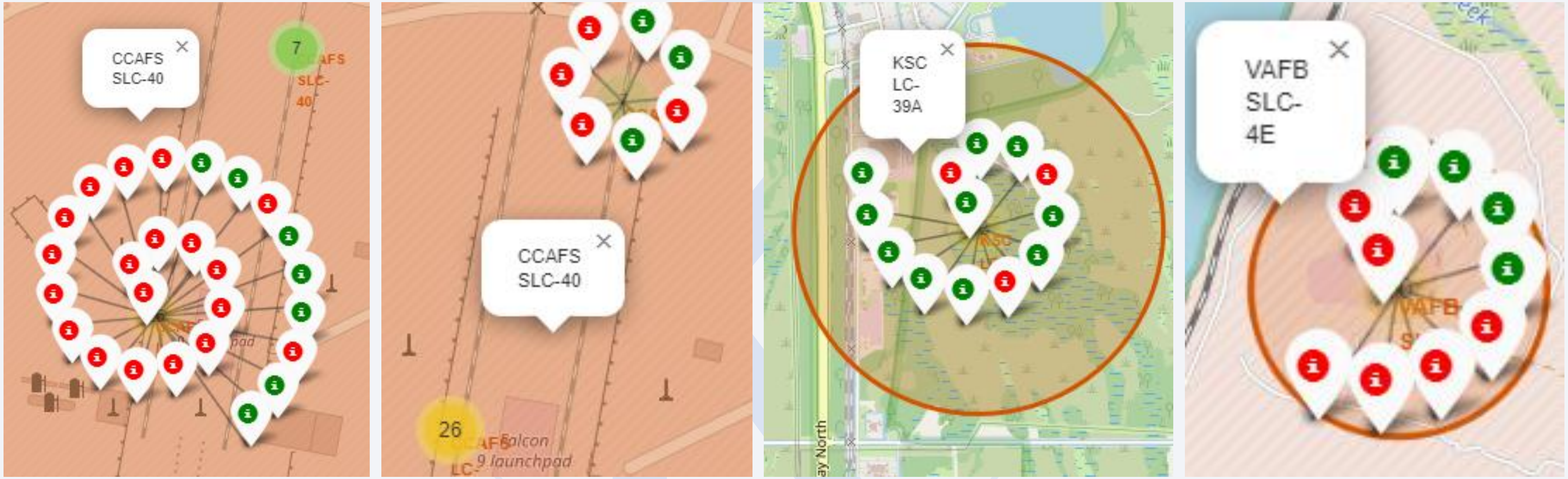
Section 3

Launch Sites Proximities Analysis

Launch site locations in the map

- All launch sites are near to the Equator line
- All launch sites are very close to the coast

Successful and failed launches in each launch site



Some sites are with higher success rates, others with less success rates.

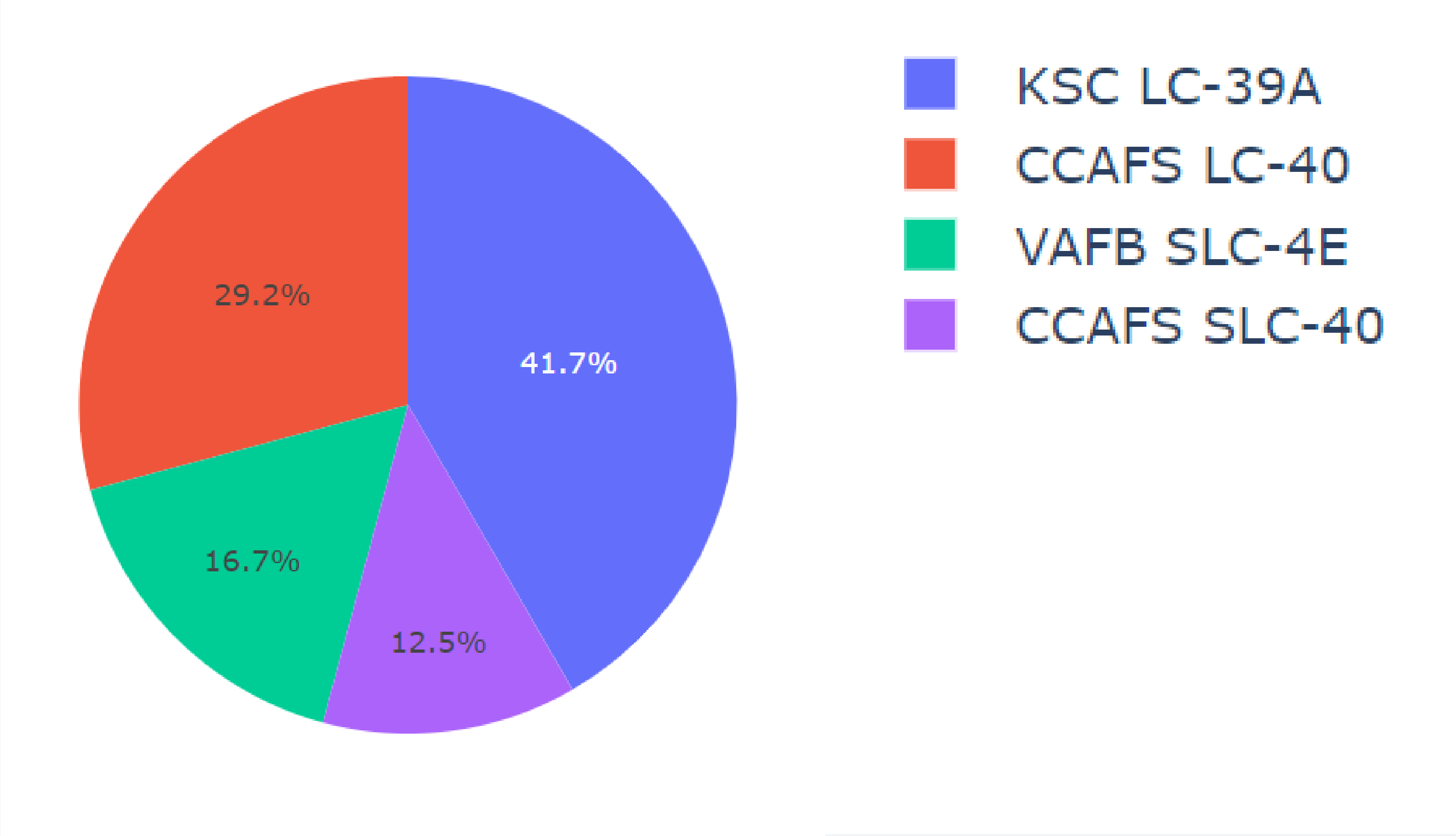
Distance lines of launch sites to the proximities

- Launch site are near to coastlines, railways, and highways, but they keep certain distance away from cities.
- The picture shows the distance between a launch site and the coastline.

Section 4

Build a Dashboard with Plotly Dash

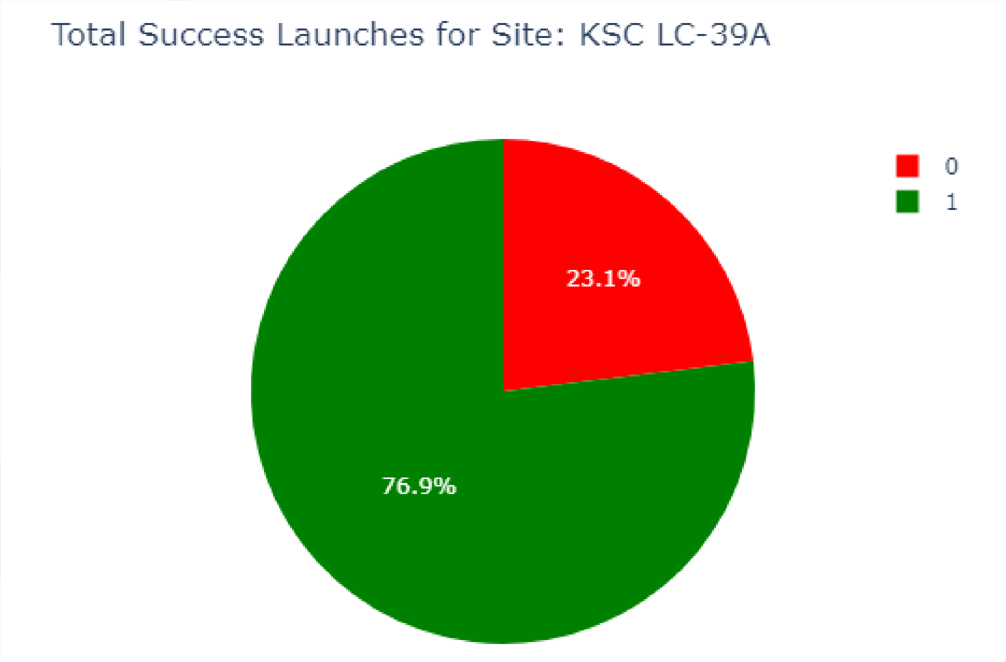
Successful launch distribution over launch sites



KSC LC-39A has the largest number of successful launches.

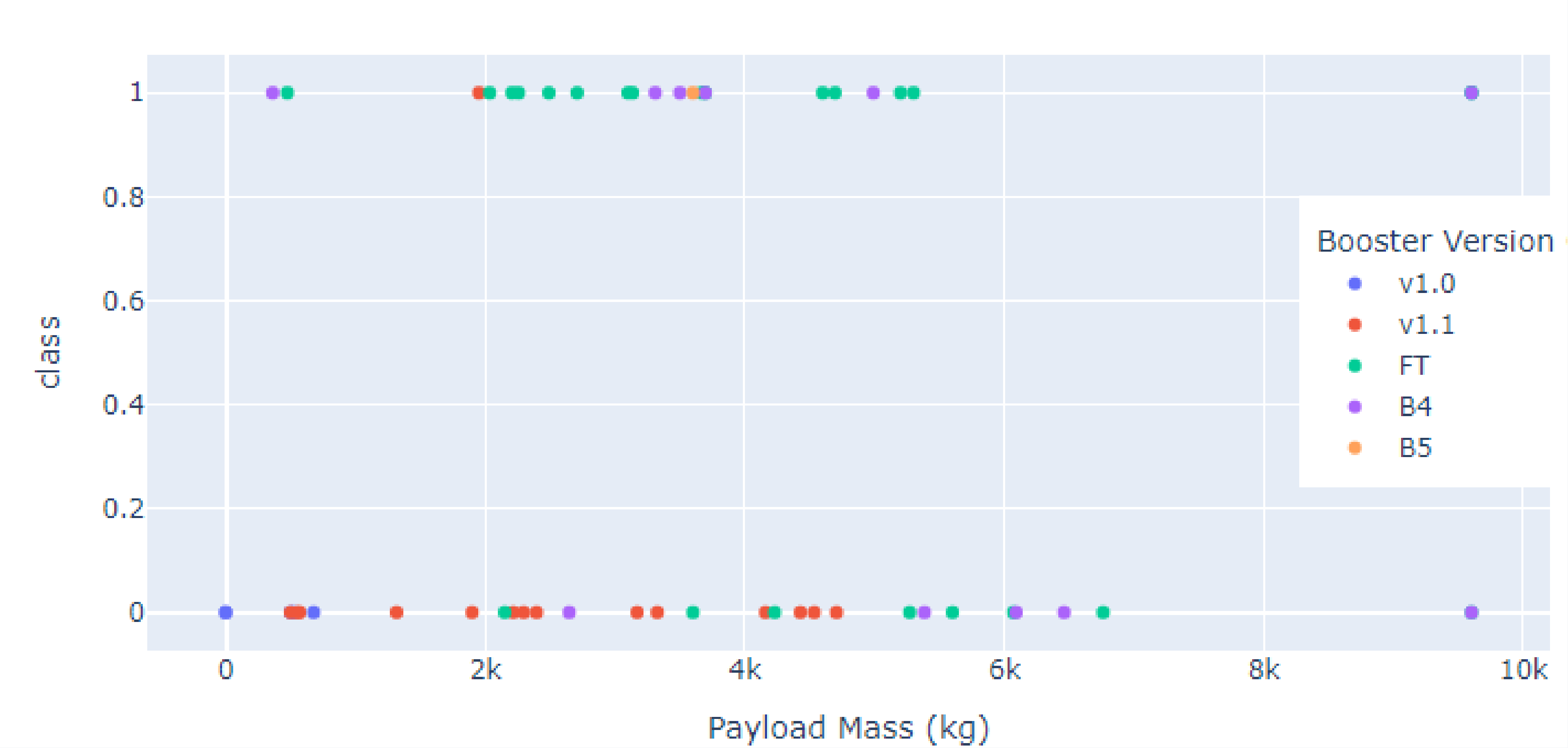
Launch success vs site

Launch Site	Success Rate
KSC LC-39A	76.9%
CCAFS SLC-40	42.9%
VAFB SLC-4E	40.0%
CCAFS LC-40	26.9%



The launch site KSC LC-39A has the largest success rate

Payload vs. Success over booster versions

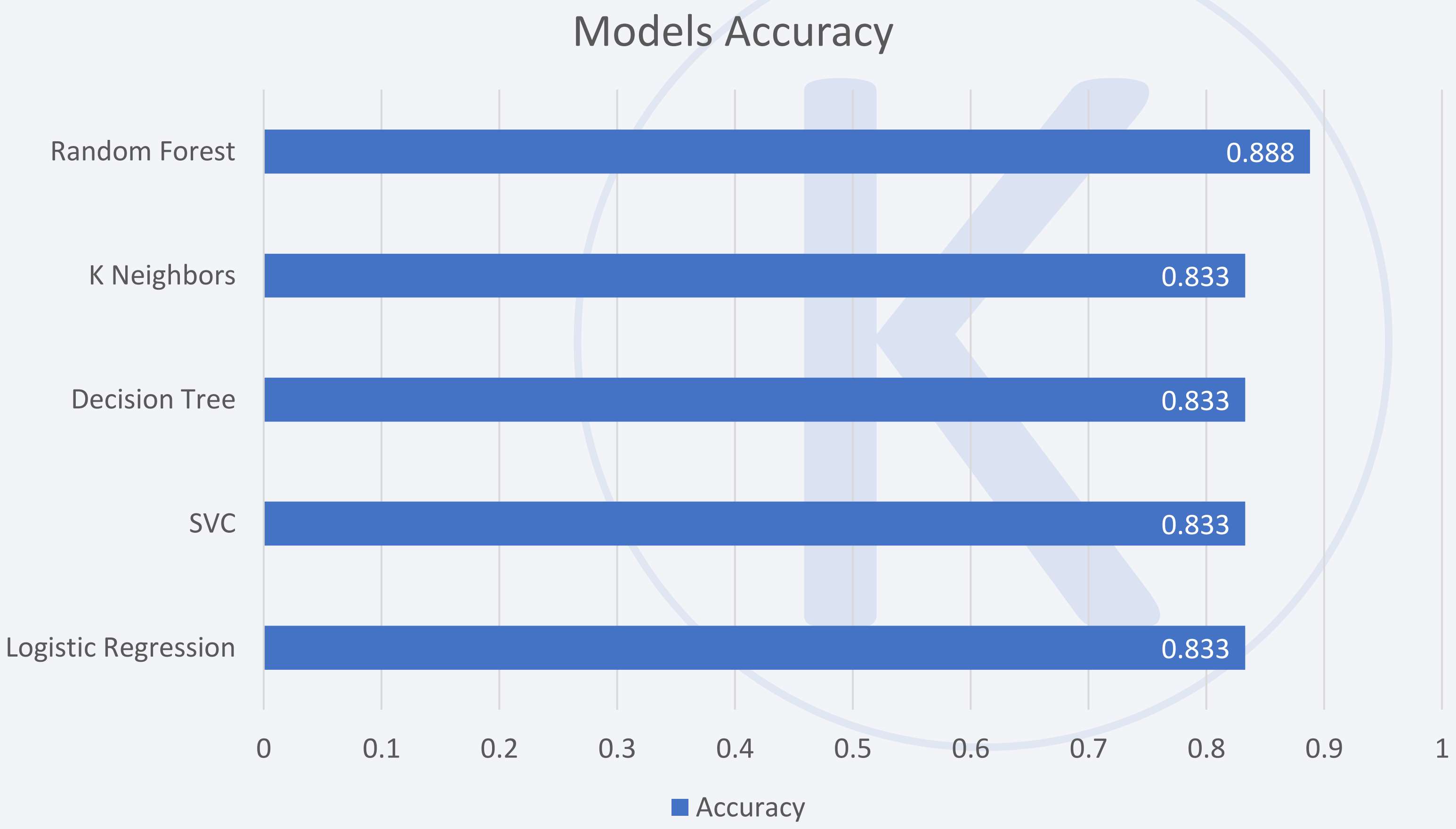


- The booster version with the highest success rate is FT.
- The payload mass rang between 2k and 4k kg has the highest success rate.

Section 5

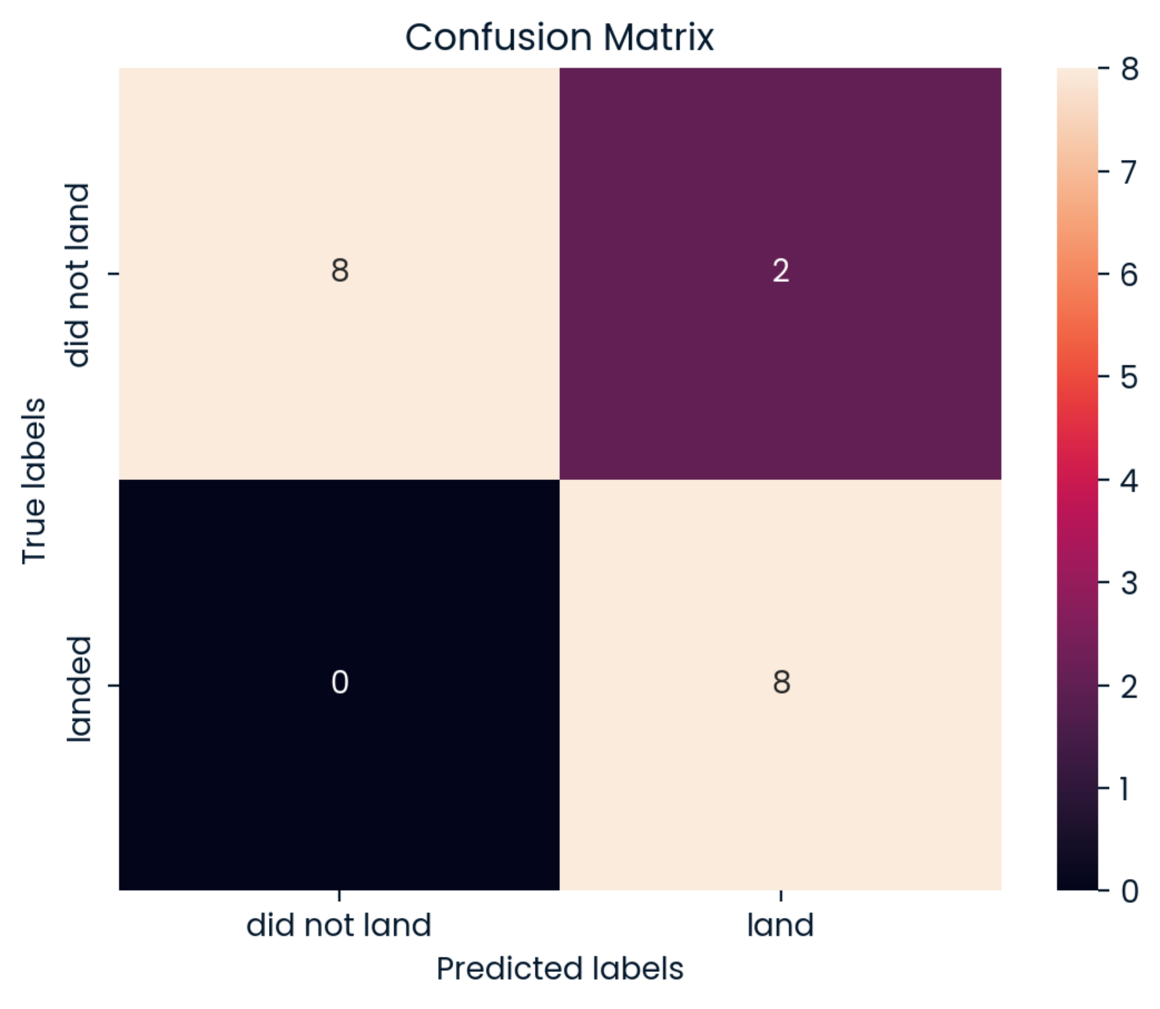
Predictive Analysis (Classification)

Classification Accuracy



The model with the best performance in test data was the random forest.

Confusion Matrix



- The confusion matrix of the random forest model.
- The model preformed well in predicting the positive outcomes but performed less with negative outcomes.

Conclusions

- CCAFS SLC 40: Superior with larger payloads.
- KSC LC 39A: More effective with smaller payloads.
- Highest success rates in ES-L1, SSO, HEO, and GEO orbits with marked improvements in LEO over time.
- Notable annual increases in landing success, especially for LEO and ISS orbits with heavy payloads.
- KSC LC-39A leads in launch success.
- FT boosters and medium payloads (2k-4k kg) show peak performance.
- Random Forest model: High accuracy (0.88), excels in predicting positive outcomes.

Appendix

- SpaceX Data API:

  https://api.spacexdata.com/v4/

- Wikipedia web page:

  https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches

Thank you!