IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Long Vu
23/10/2023

Outline

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

Executive Summary

Summary of methodologies

This study used a quantitative data analysis approach to assess the cost of Stage 1 in SpaceX's launches and predict the likelihood of them reusing the first stage. Data was collected from 90 SpaceX launches

Summary of all results

- Exploratory Data Analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Introduction

Who are we

Space Y, a space rocket company founded by Allon Musk, poised to outperform our primary rival, SpaceX.

Allon Mask

Our Goal

Assess the cost of Stage 1 in SpaceX's launches and predict the likelihood of them reusing the first stage.

Like how do variables such as payload mass, launch site, number of flights, and orbits affect the success of the first stage landing? And how to predict with success with the highest accuracy.

Section 1

Methodology

Methodology

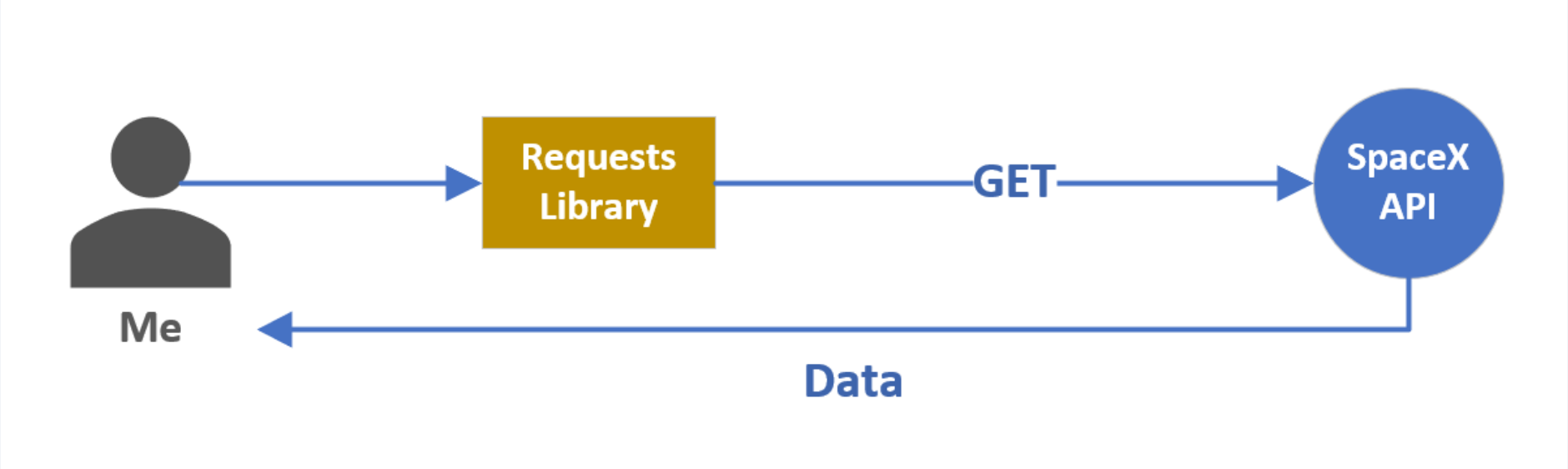
Executive Summary

- Data The data used in this report was collected from SpaceX REST API

- The data was cleaned and preprocessed to ensure that it was consistent and accurate. This involved removing duplicate records, correcting errors, and filling in missing values.

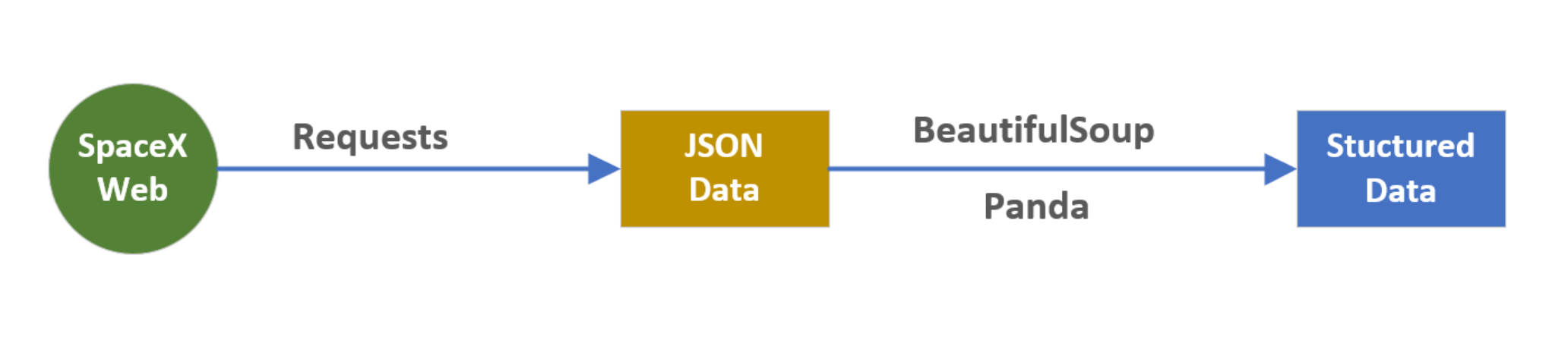
- Once the data was clean, it was analyzed using a variety of methods, including:
  - Perform exploratory data analysis (EDA) using visualization and SQL
  - Perform interactive visual analytics using Folium and Plotly Dash
  - Perform predictive analysis using classification models

Data Collection – SpaceX API



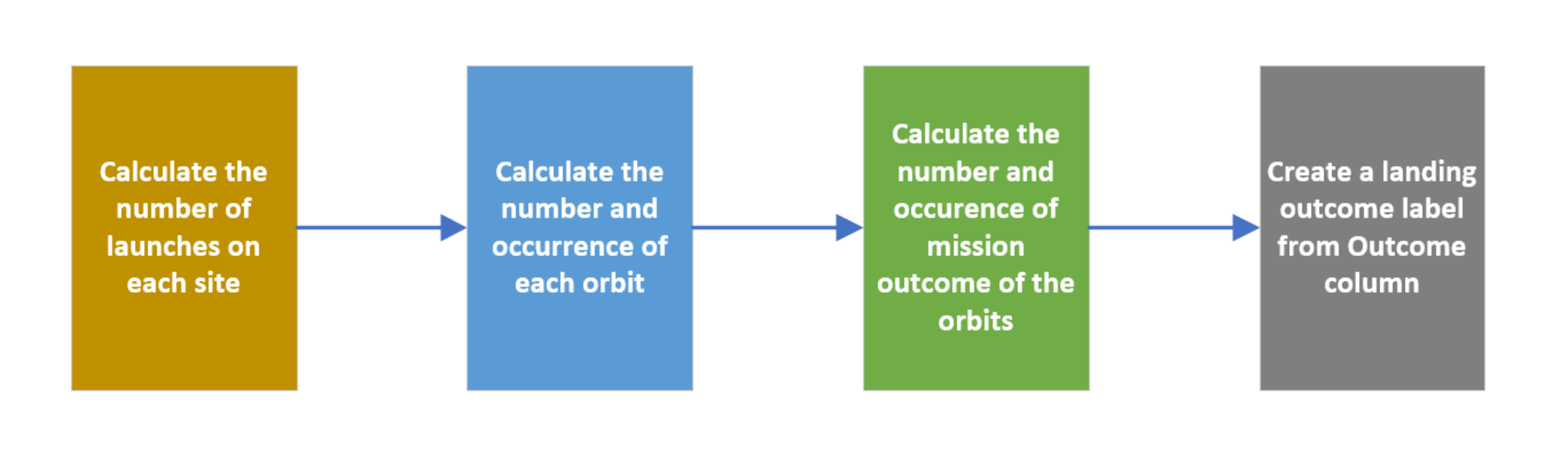
Github link for an external reference and peer-review

Data Collection - Scraping



Github link for an external reference and peer-review

Data Wrangling



Github link for an external reference and peer-review

EDA with Data Visualization

I used 3 types of charts in my analysis:

- Scatter Plot: For visualizing relationships between variables.
- Bar Chart: To compare values between different categories or groups.
- Line Chart: To show trends and changes over time or a continuous range.

Github link for an external reference and peer-review

EDA with SQL

I performed 10 queries to

- Display launch site names and find those starting with 'CCA'.
- Calculate payload mass for NASA (CRS) and the average for F9 v1.1.
- Identify significant events, such as first successful ground pad landing.
- Analyze landing outcomes, ranking them between 2010 and 2017.

Github link for an external reference and peer-review

Build an Interactive Map with Folium

I added various Folium map objects for different purposes:

- **Map Circle**: Represent specific areas of interest with a defined radius.
- **Marker**: Pinpoint important locations on the map.
- **Marker Cluster**: Group nearby points of interest to reduce clutter.
- **Mouse Position**: Display user coordinates for interaction.
- **Polyline**: Show routes or connections between locations.

Github link for an external reference and peer-review

Build a Dashboard with Plotly Dash

The following graphs and plots were used to visualize data

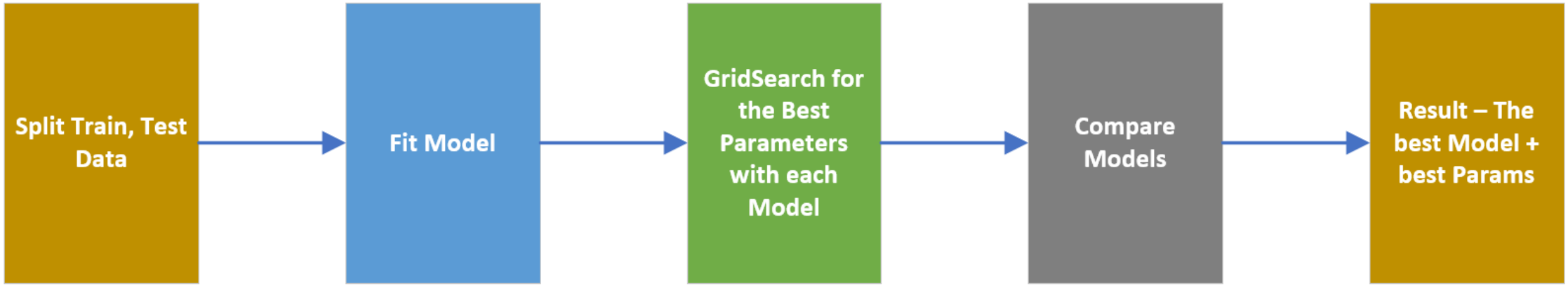
- Percentage of launches by site
- Payload range

This combination allowed to quickly analyze the relation between payloads and launch sites, helping to identify where is best place to launch according to payloads.

Github link for an external reference and peer-review

Predictive Analysis (Classification)

I used GridSearch CV to find the best parameters for KNN, Decision Tree, Logistic Regression and SVM model. Than compare to find the best model.



Github link for an external reference and peer-review

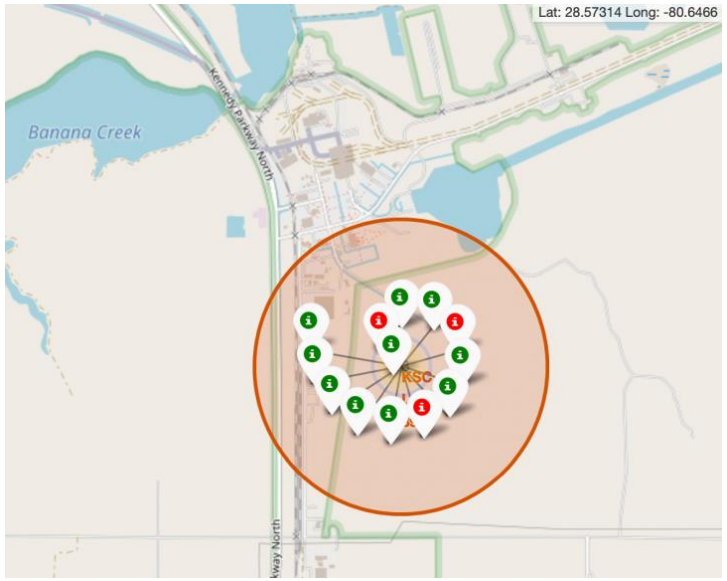
Results

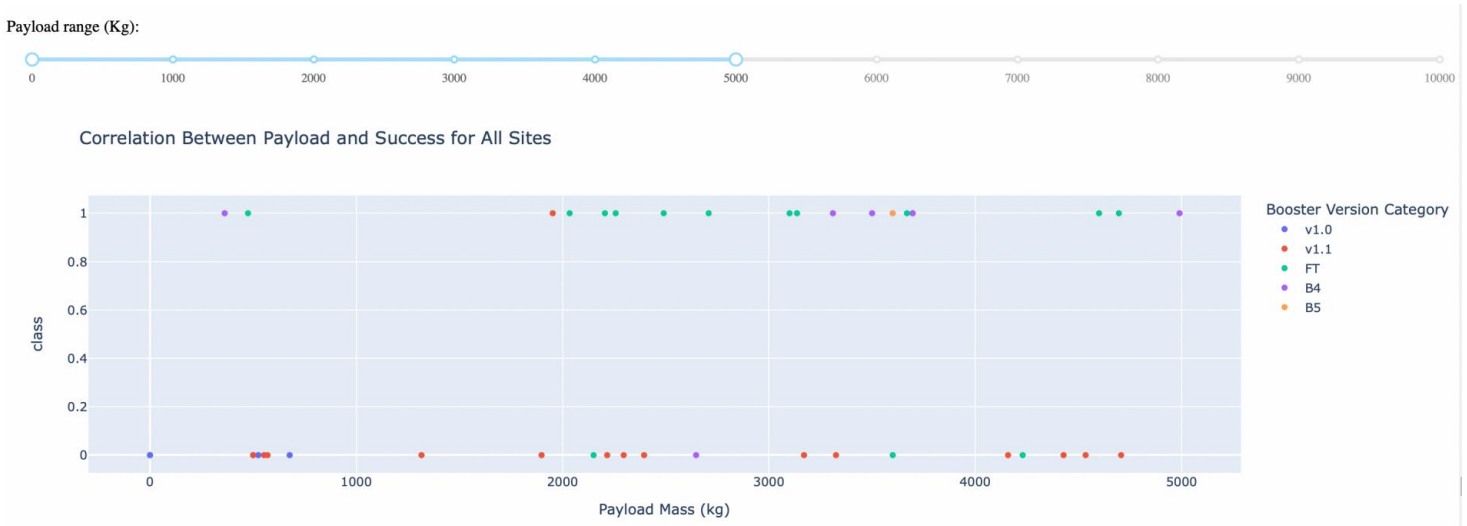
Exploratory data analysis results:

- Space X uses 4 different launch sites;
- The first launches were done to Space X itself and NASA;
- The average payload of F9 v1.1 booster is 2,928 kg;
- The first success landing outcome happened in 2015 fiver year after the first launch;
- Many Falcon 9 booster versions were successful at landing in drone ships having payload above the average;
- Almost 100% of mission outcomes were successful;
- Two booster versions failed at landing in drone ships in 2015: F9 v1.1 B1012 and F9 v1.1 B1015;
- The number of landing outcomes became as better as years passed.

Results

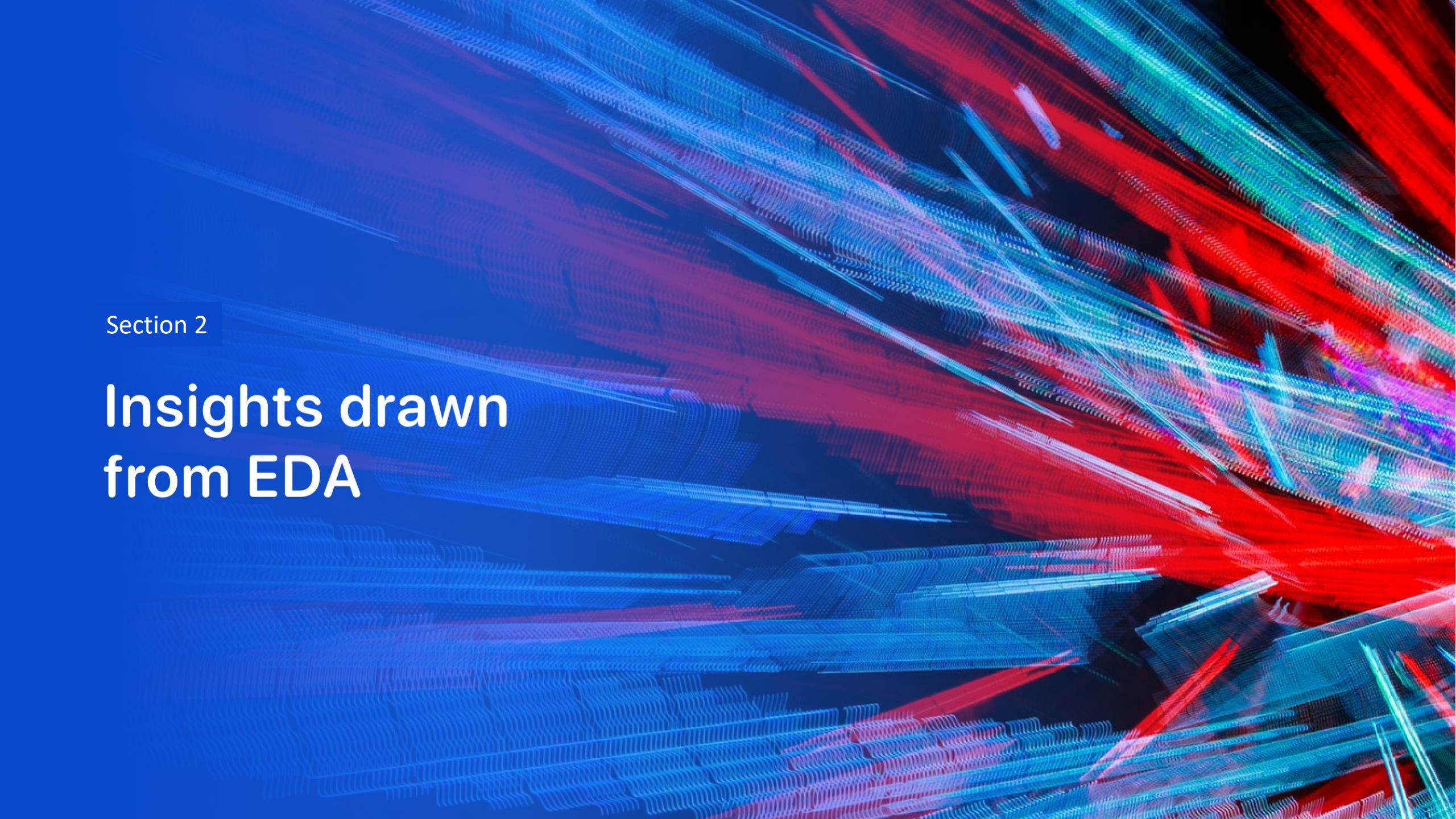
Interactive analytics demo in screenshots:





Predictive analysis results:

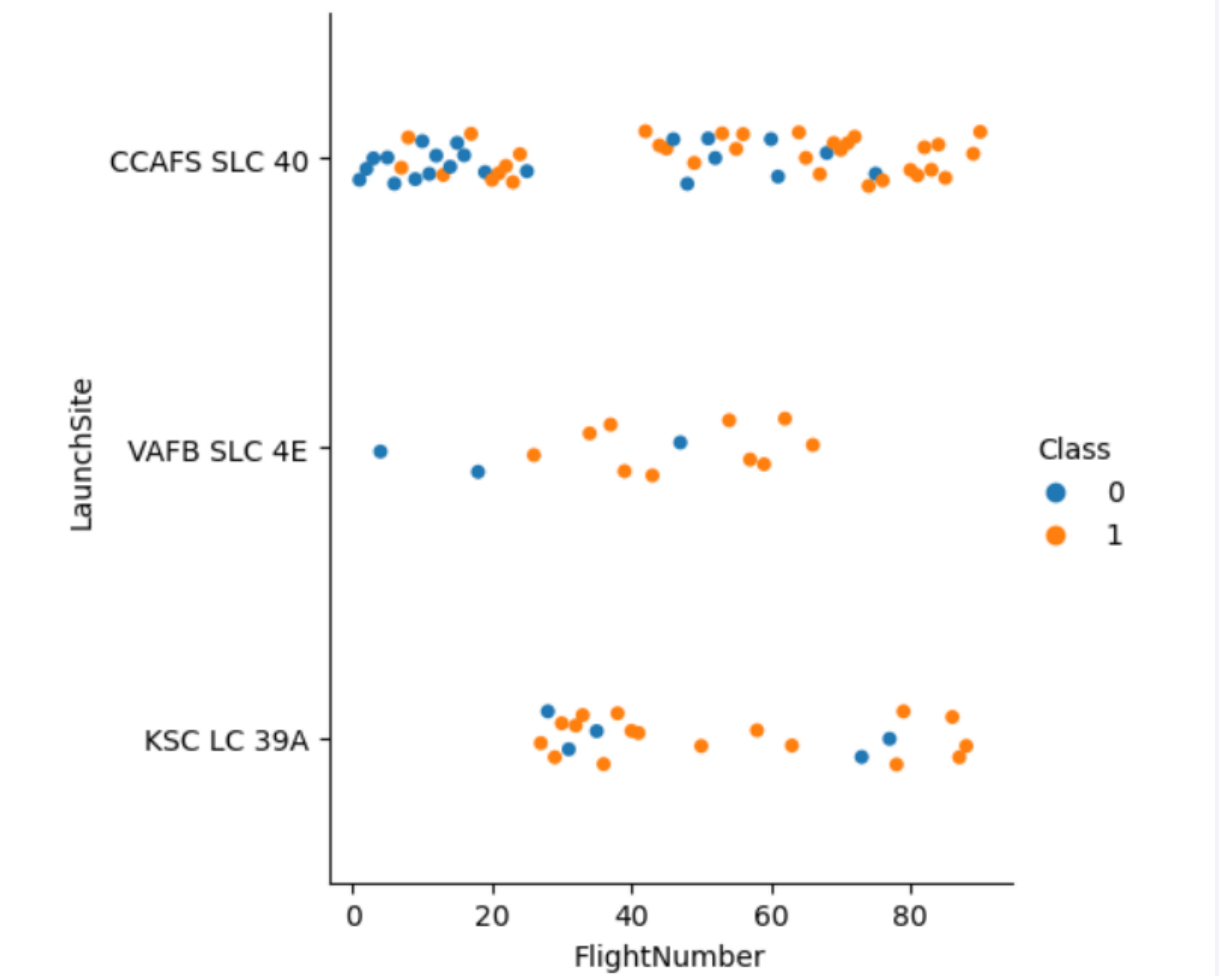
Trees Classification is the model with **highest** accuracy – 87.5%


Section 2
Insights drawn from EDA

Flight Number vs. Launch Site

As can be seen most of the flight is Launched in CCAFS SLC 40

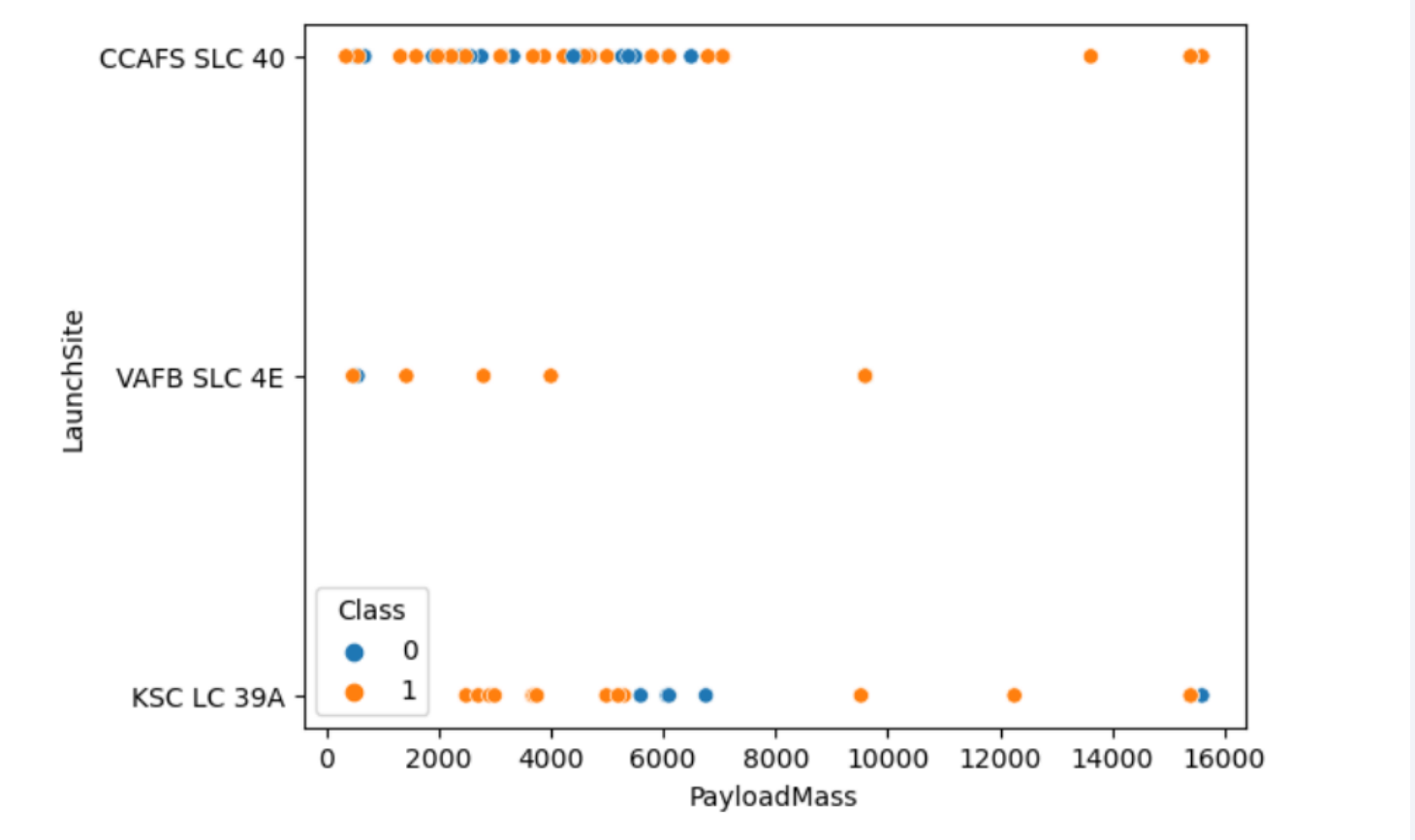
However, VAFB SLCE 4E & KSC LC 39A has the highest successful launch rate



Payload vs. Launch Site

KSC LC 39A has the launches with highest Payload Mass, up to 16,000

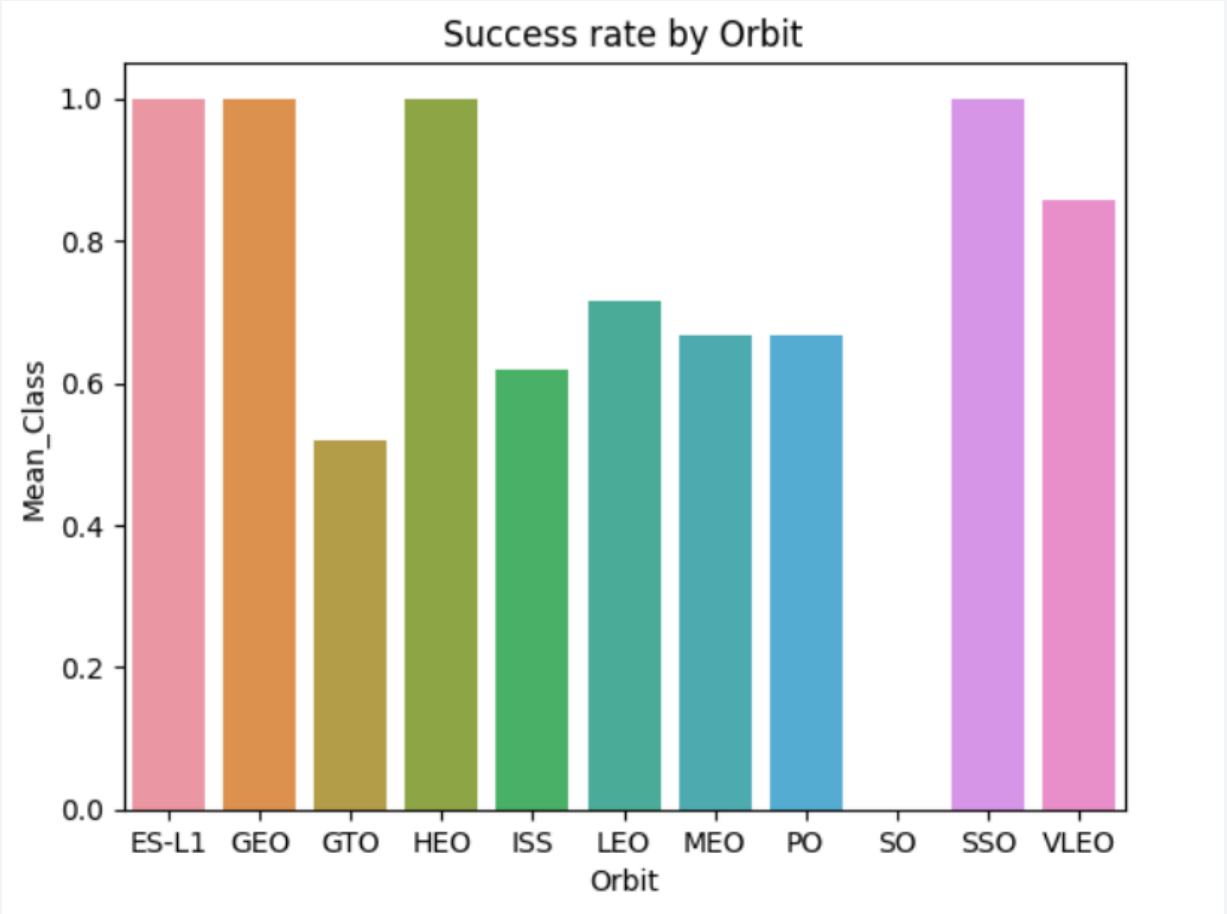
While VAFB SLC 4E has the lowest average Payload mass



Success Rate vs. Orbit Type

Most successful Orbit types are ES-L1, GEO, HEO, SSO with 100%

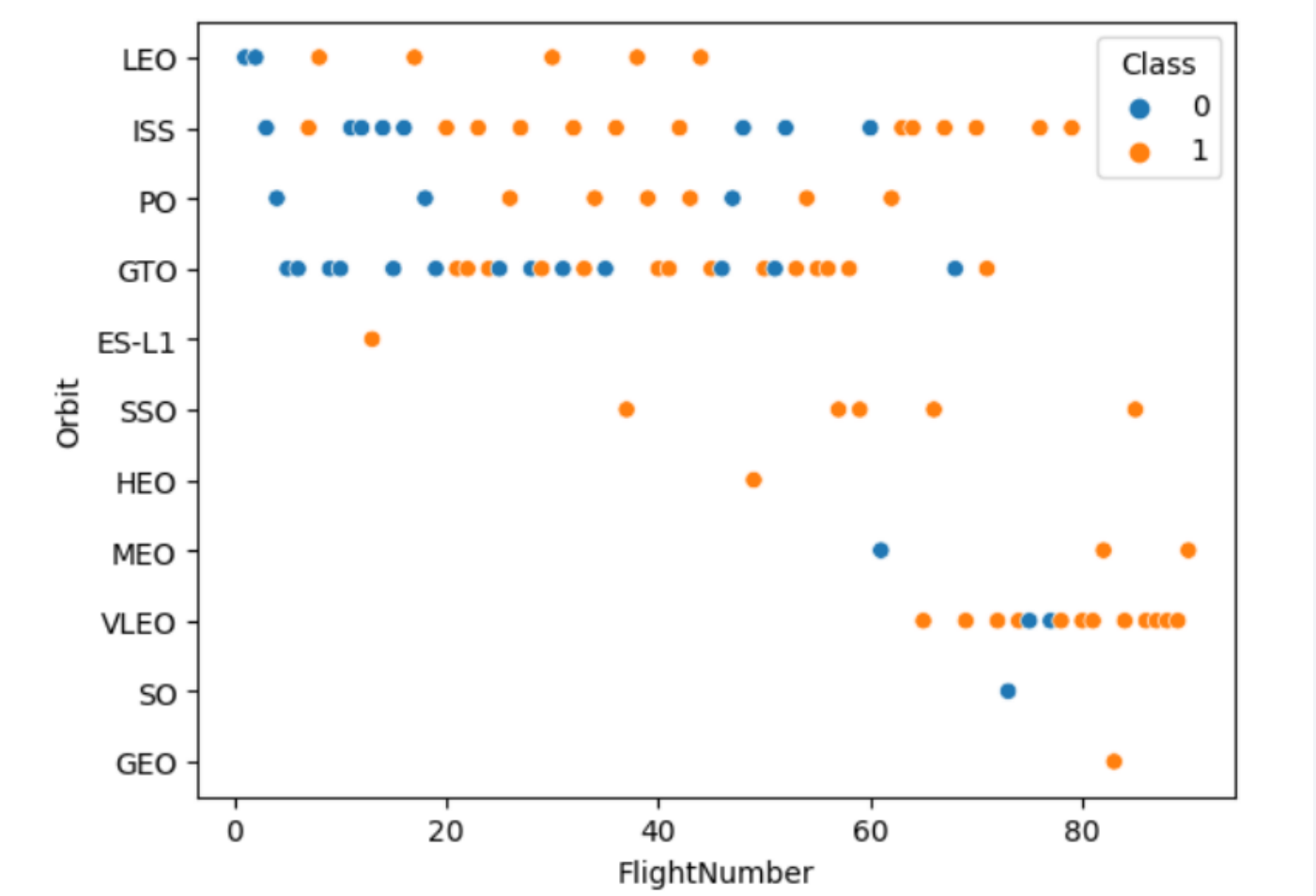
While None of launch with SO type succeeded



Flight Number vs. Orbit Type

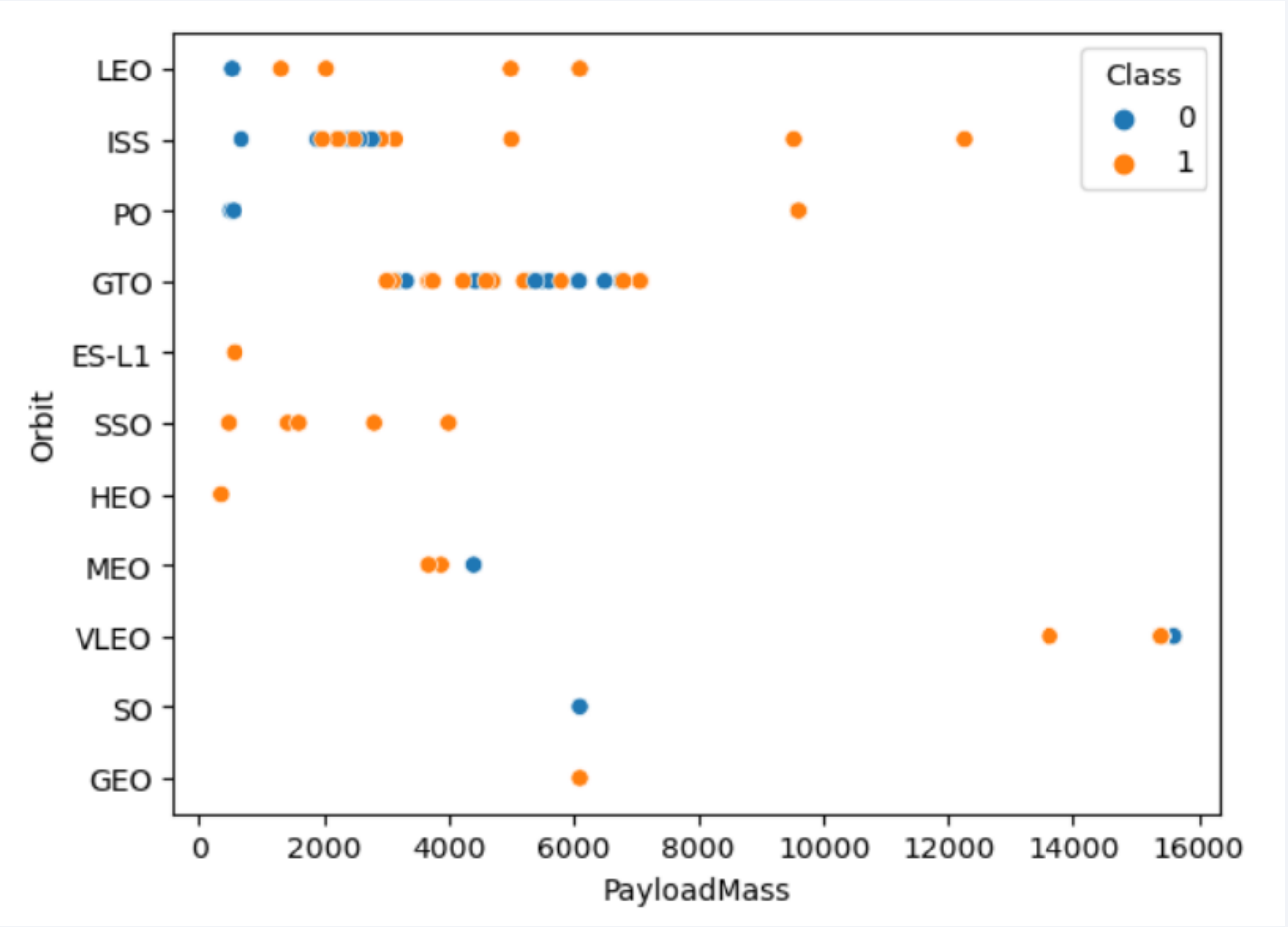
However SO & GEO success rate is not meaningful – only 1 launch.

ISS & GTO are the most popular orbit types for launch



Payload vs. Orbit Type

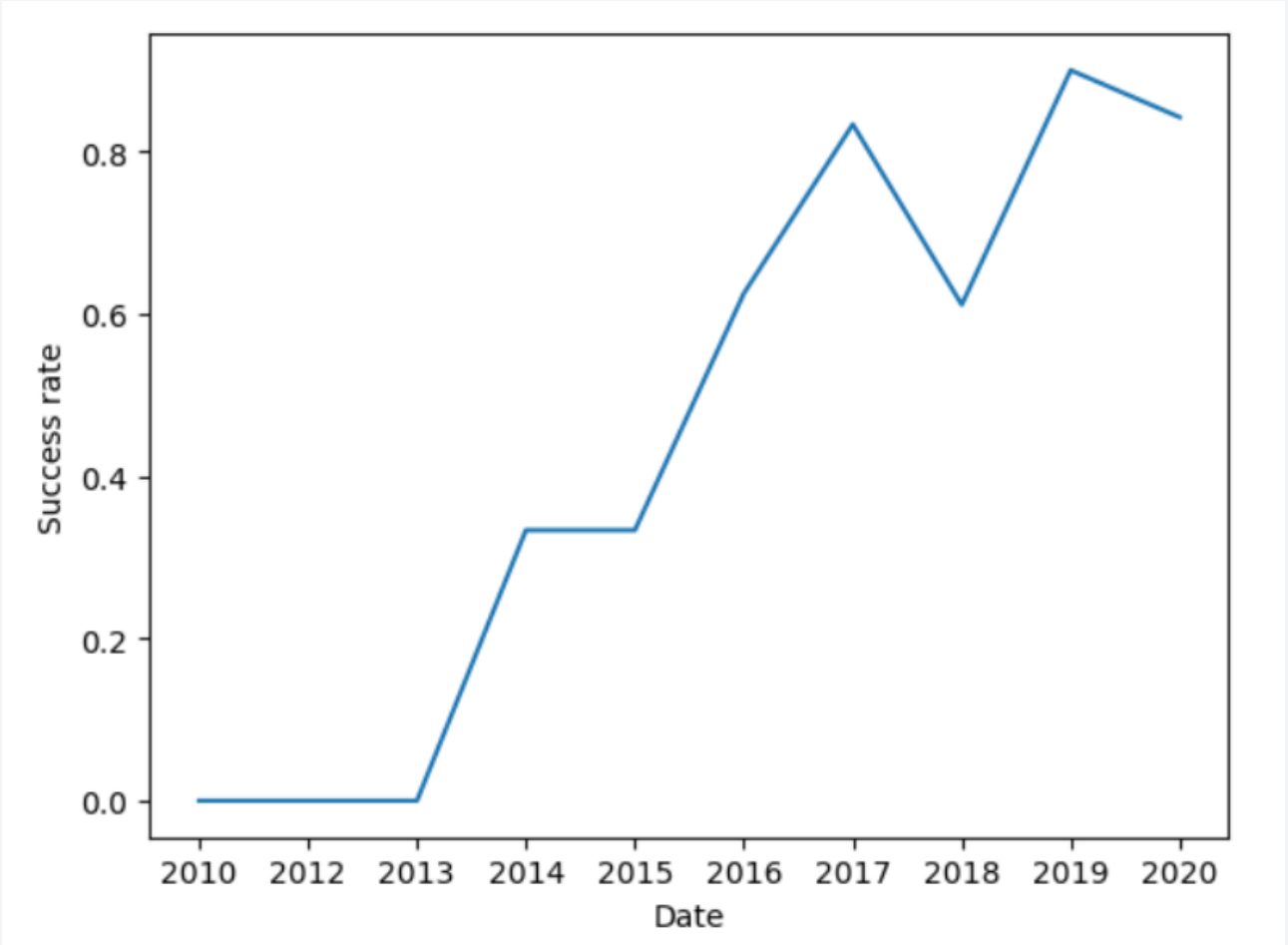
VLEO type has the highest Payloadmass while other tend to have lower Payloadmass.



Launch Success Yearly Trend

Successful rate is getting **Higher** over the years

and is currently **83%**



All Launch Site Names

Launch site: The place from where rockets, missiles, or spacecraft are launched.

👉 Here are all spaceX launch sites:

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
2010-04-06	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-08-12	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	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-08-10	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-01-03	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

5 example records where launch sites begin with `CCA`

Total Payload Mass

The total payload carried by boosters from NASA is 45,596 Kg

Sum_Payload
45596

Average Payload Mass by F9 v1.1

The average payload mass carried by booster version F9 v1.1 is 2,928.4 Kg

AVG(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

The dates of the first successful landing outcome on ground pad is 22/12/2015

Date	Landing_Outcome
2015-12-22	Success (ground pad)

***Payload mass**: The mass of the object being carried into space, such as a satellite, or probe.*

Successful Drone Ship Landing with Payload between 4000 and 6000

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

***Booster**: A rocket stage that provides additional thrust during liftoff.*

Booster_Version
F9 FT B1021.1
F9 FT B1022
F9 FT B1023.1
F9 FT B1026
F9 FT B1029.1
F9 FT B1021.2
F9 FT B1029.2
F9 FT B1036.1
F9 FT B1038.1
F9 B4 B1041.1
F9 FT B1031.2
F9 B4 B1042.1
F9 B4 B1045.1
F9 B5 B1046.1

Total Number of Successful and Failure Mission Outcomes

👉 Most of the outcome is **Success**

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

Boosters Carried Maximum Payload

☞ List of the booster which have carried the maximum payload mass

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

👉 2 failed landing in 2015 are both from CCAFS LC-40

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

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

Beside “No attempt”

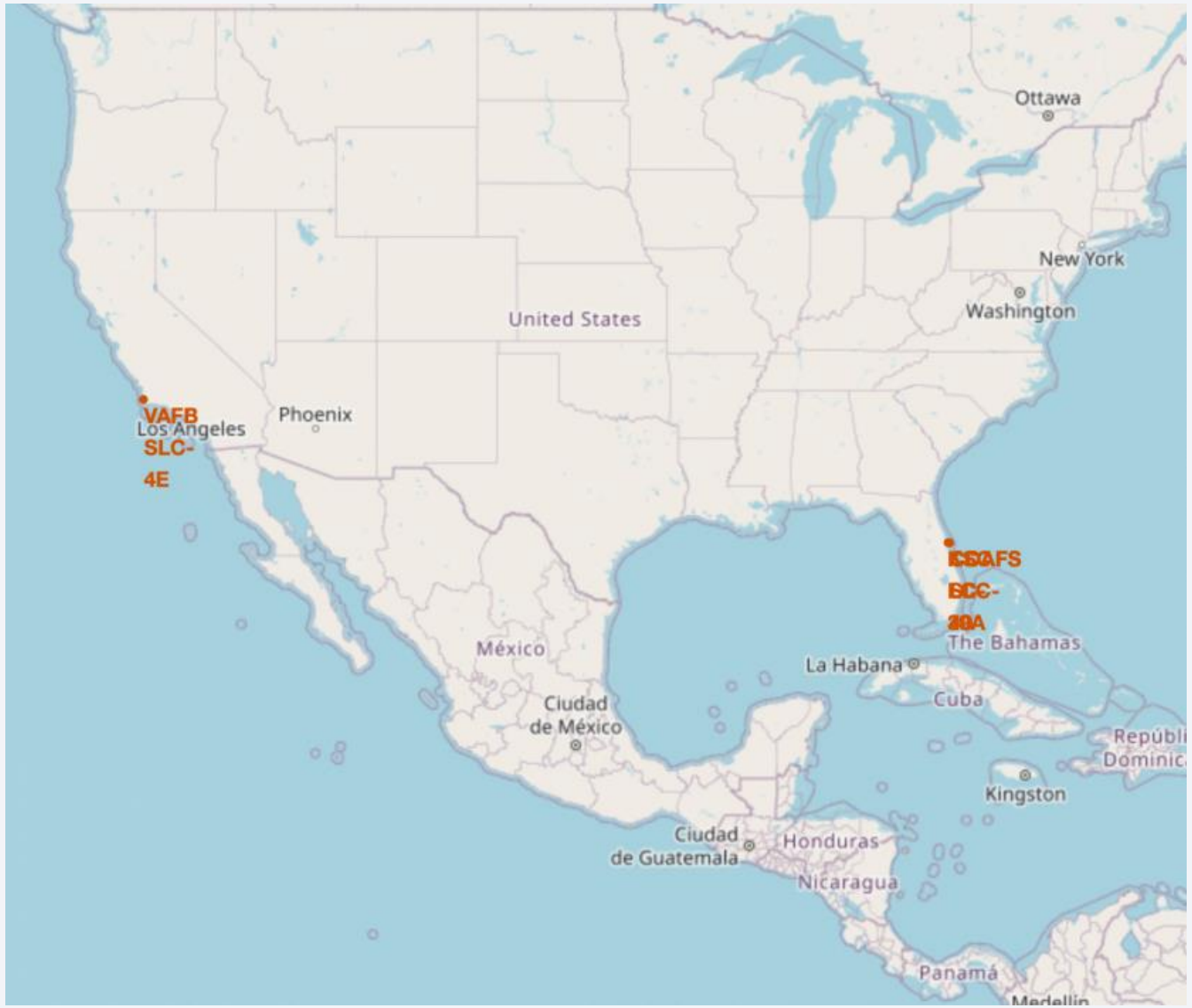
👉 The most common Landing outcome between 2010-06-04 and 2017-03-20 is **Success (ground pad)**

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

Section 3

Launch Sites Proximities Analysis

All Launch Site Locations



☞ Most of Launch sites are in proximity to the Equator line.

- *Due to faster Earth moving speed at the equator.*

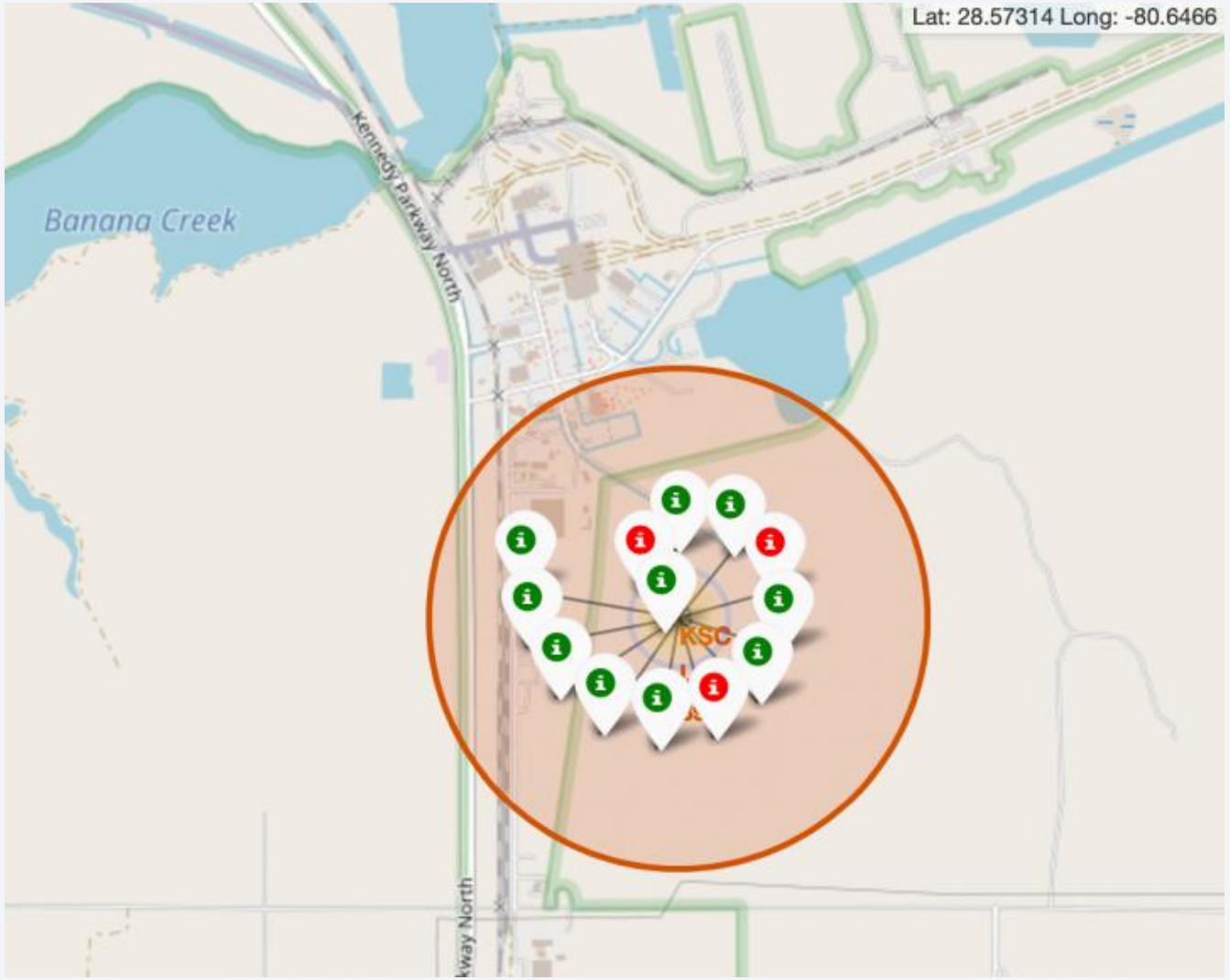
☞ All launch sites are close to the coast.

- *To reduce the risk when rocket fall*

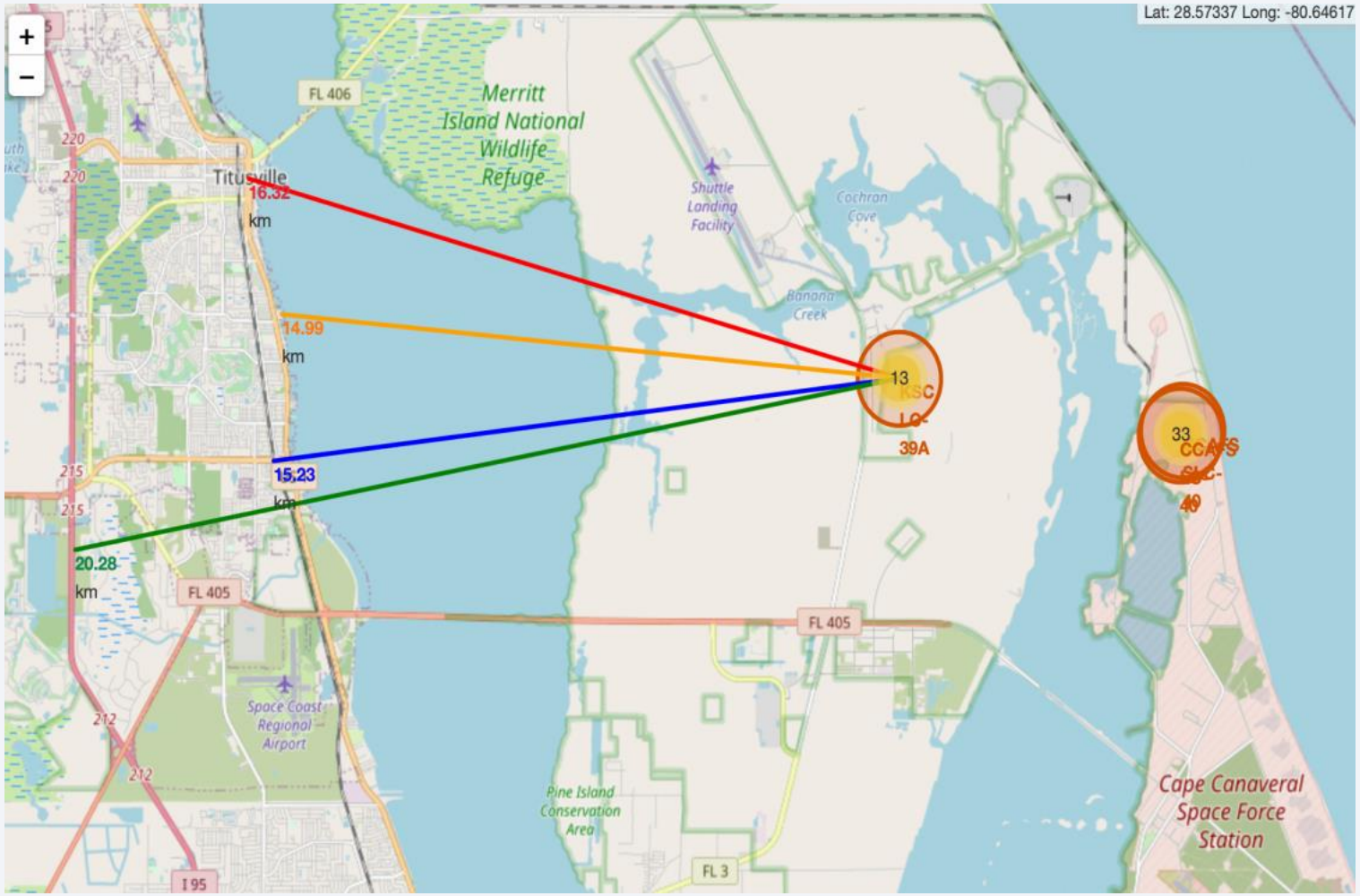
Colour-labeled launch records on the map

☞ Most of the launches are successful

- Green Marker = Successful Launch
- Red Marker = Failed Launch

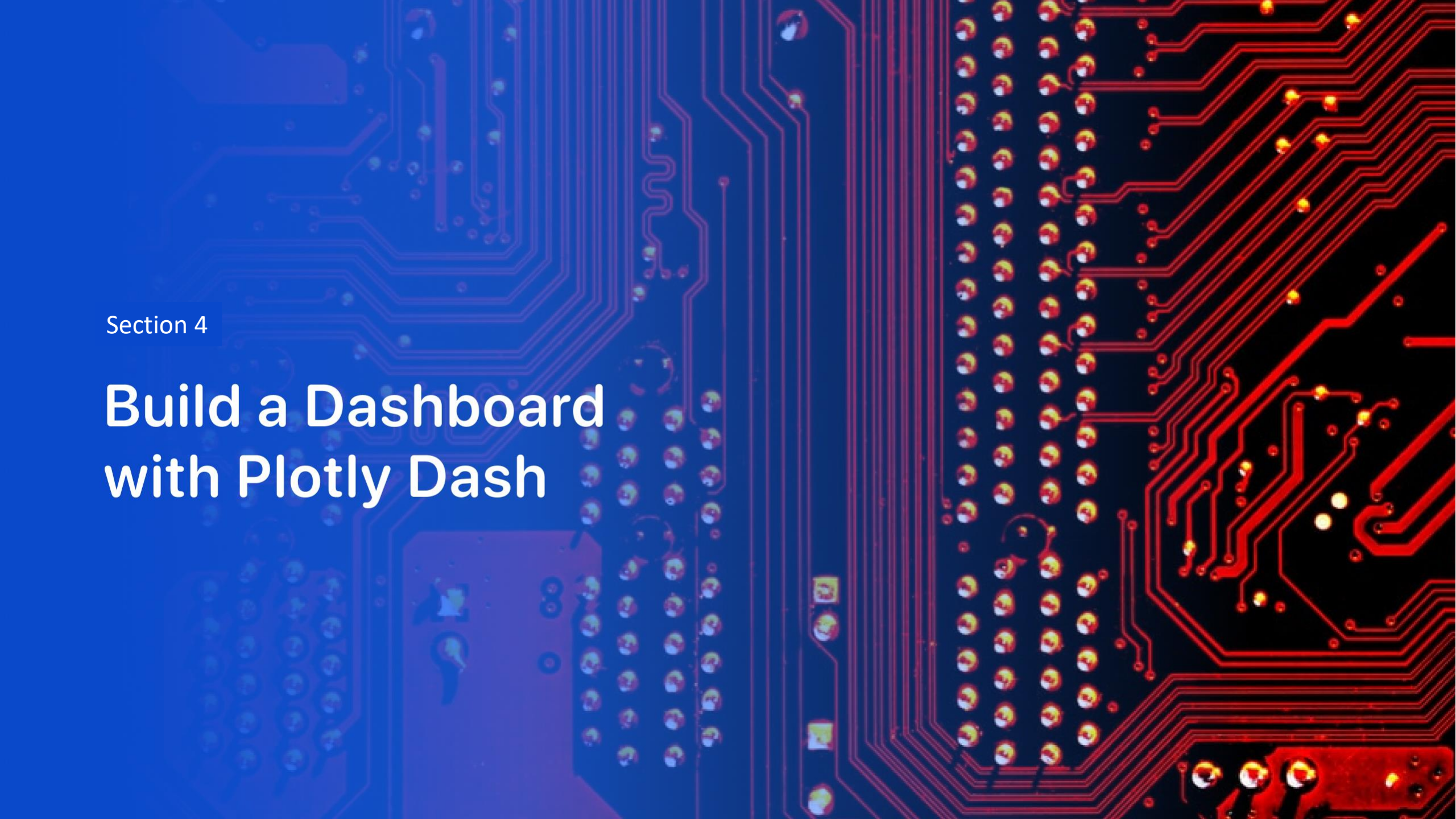


Distance from the launch site KSC LC-39A to its proximities

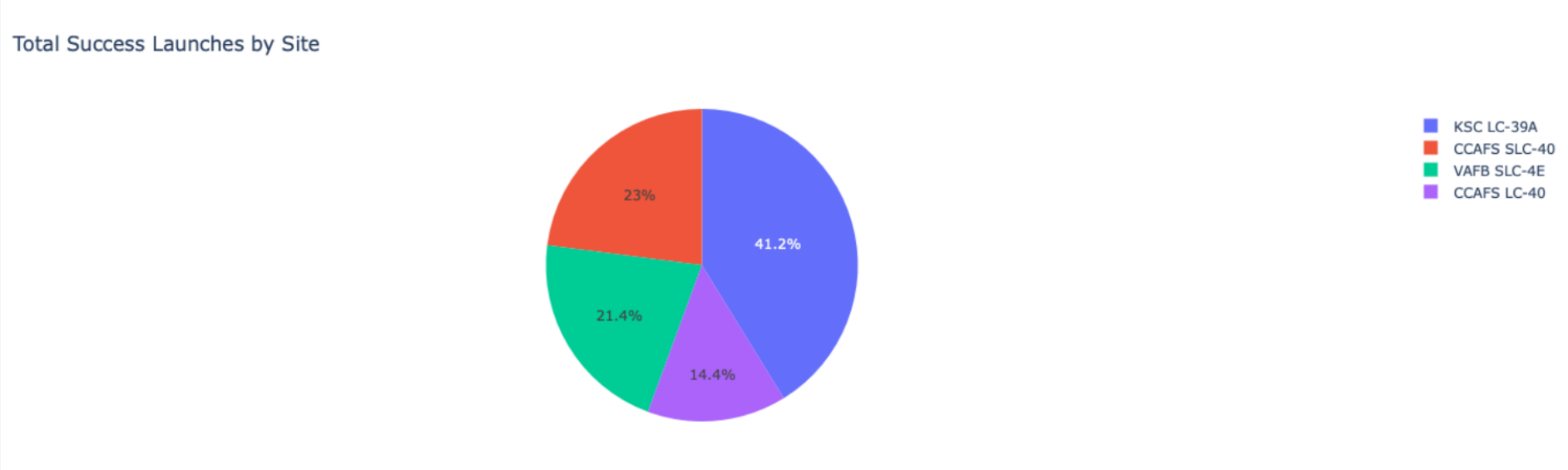


☞ Launch site KSC LC-39A is **relatively close** to:

- Railway (15.23 km)
- Highway (20.28 km)
- Coastline (14.99 km)
- Titusville city (16.32 km)

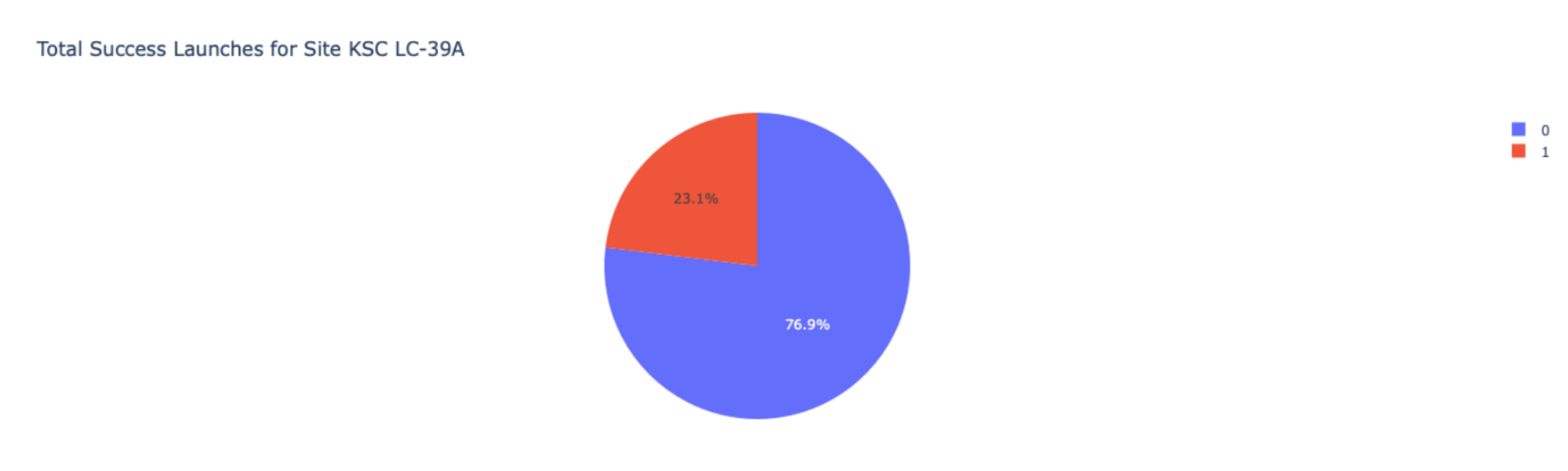

Section 4
Build a Dashboard with Plotly Dash

Success Launches Categorized by Launch Site



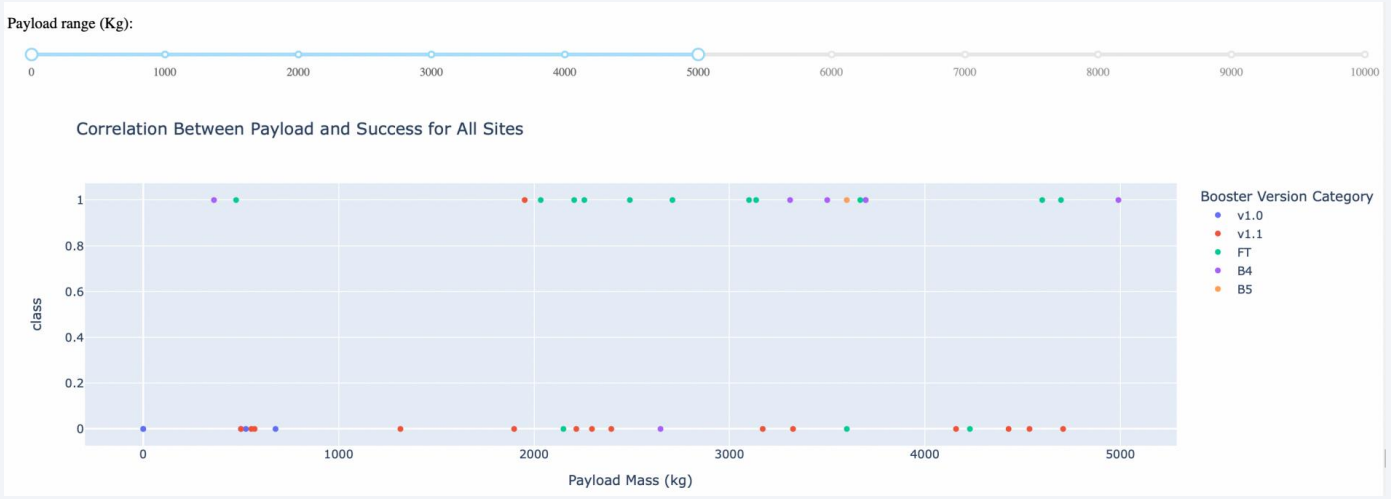
☞ Most success launches are from KSC LC-39A

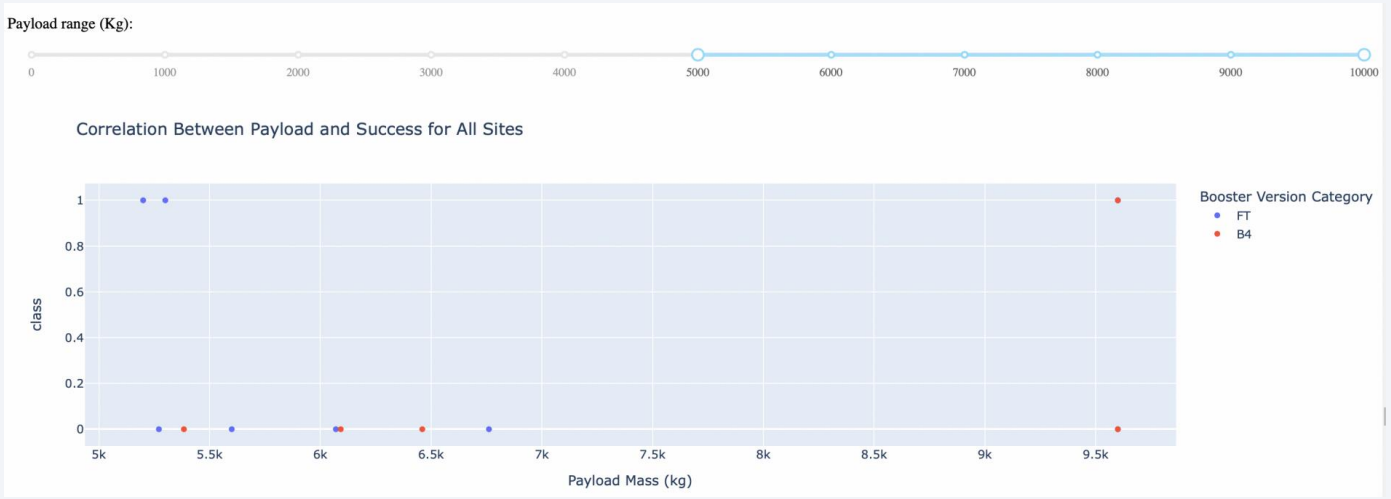
Success Rate of Launch Site KSC LC-39A



KSC LC-39A is **High – 76.9%**

Correlation between Payload & Success Rate



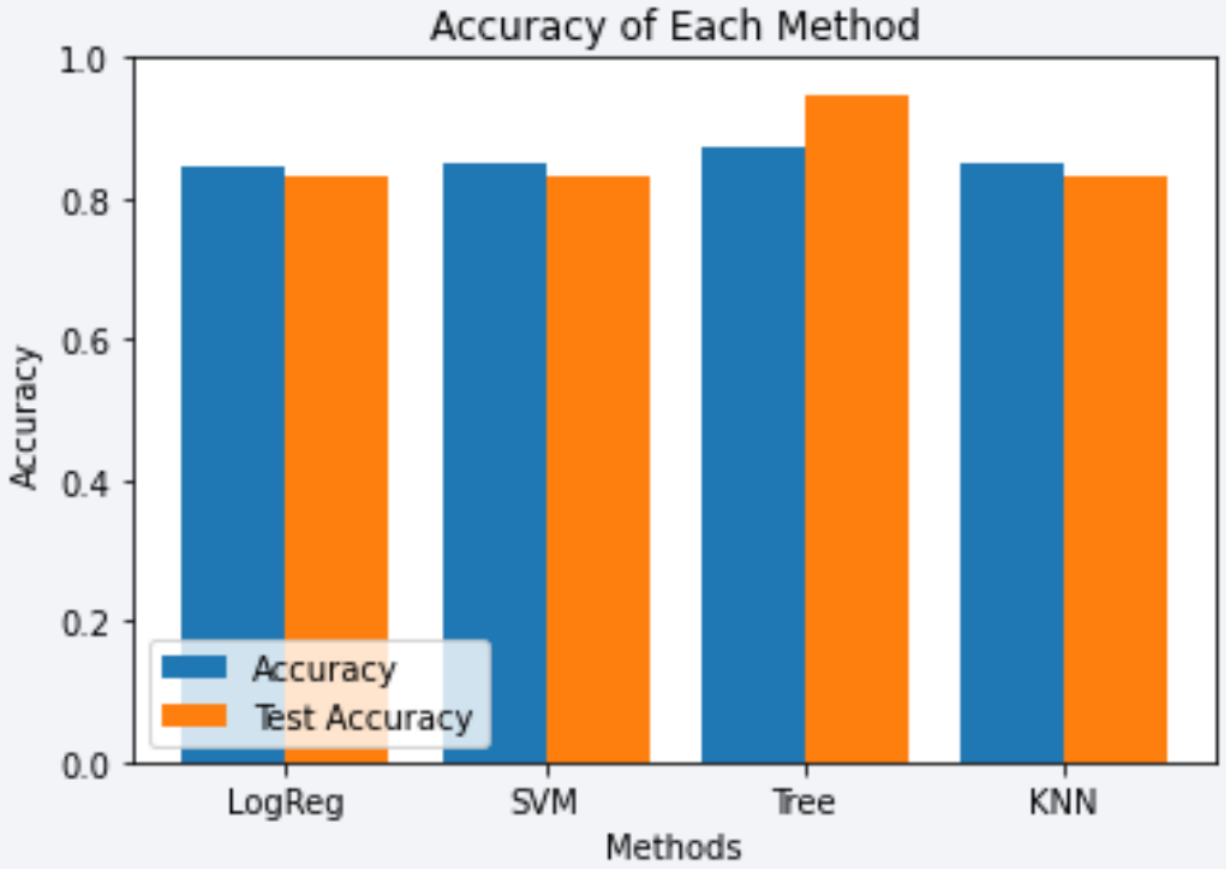


👉 It can be seen that payloads between 2000 - 5500 kg have the highest success rate.

Section 5

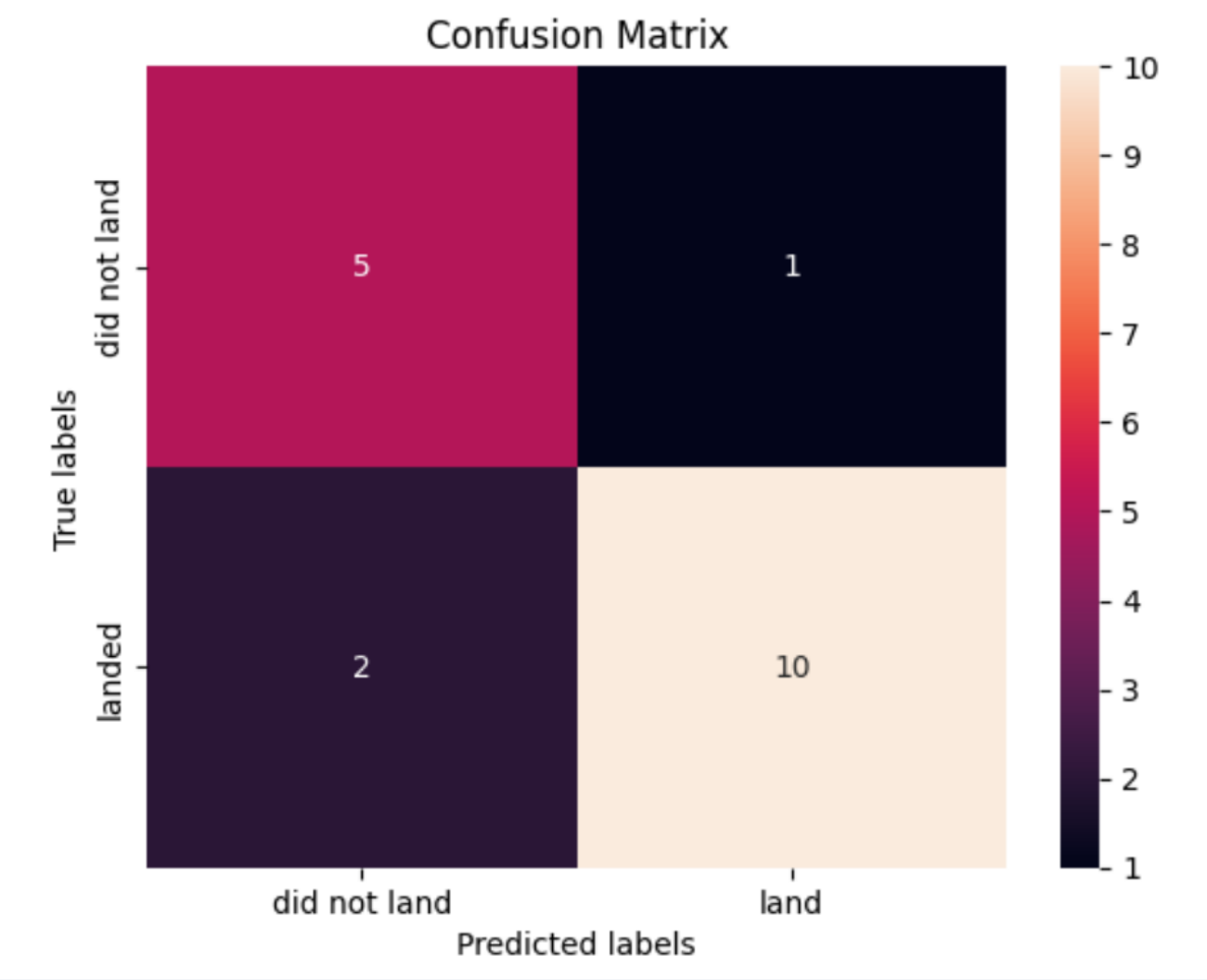
Predictive Analysis (Classification)

Classification Accuracy



👉 Model with **highest** accuracy is Tree Classification – **87.5%**

Confusion Matrix



👉 Decision Tree model is good at predicting true value of both outcomes “Land” and “Did not land”

Conclusions

- Optimal launch site:
  - KSC LC-39A is the best choice with highest success rate.
  - Launches with smaller payloads are more successful.
  - Location: Most launch sites are near the Equator and coast.

- The success rate of launches improves over time.
- Decision Tree Classifier is Best for predicting successful landings
- Reliable orbits: ES-L1, GEO, HEO, and SSO have a 100% success rate.

Appendix

- All datas are origined from **IBM Data Science Professional Certificate** courses
- Jupyter Notebook files are made with Skills Network environment so it might has some differences compare to files made by local Jupyter Notebooks

Thank you!