IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Pedro Carlos Ronda Pérez
03/15/2025

Outline

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

Executive Summary

- Summary of methodologies
  - Data collection
  - Data Wrangling
  - Exploratory Data Analysis
  - Data Visualization
  - Machine Learning Prediction
- Summary of all results
  - Exploratory Data Analysis result
  - Data Visualization
  - Predictive Analysis result

Introduction

- The objetive: Machine Learning can predict the success of the Falcon 9 landing
- Problems you want to find answers:
  - Identify which factors determine the success of the landing

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - The data has been obtained from SpaceX's website through its API.
  - Data has also been extracted from Wikipedia.
- Perform data wrangling
  - The data corresponding to the Falcon 9 has been selected.
  - It has been cleaned and classified according to the launch site.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Different models such as Decision Tree, Logistic Regression, SVM, and K-Nearest have been evaluated, and the best one has been estimated.

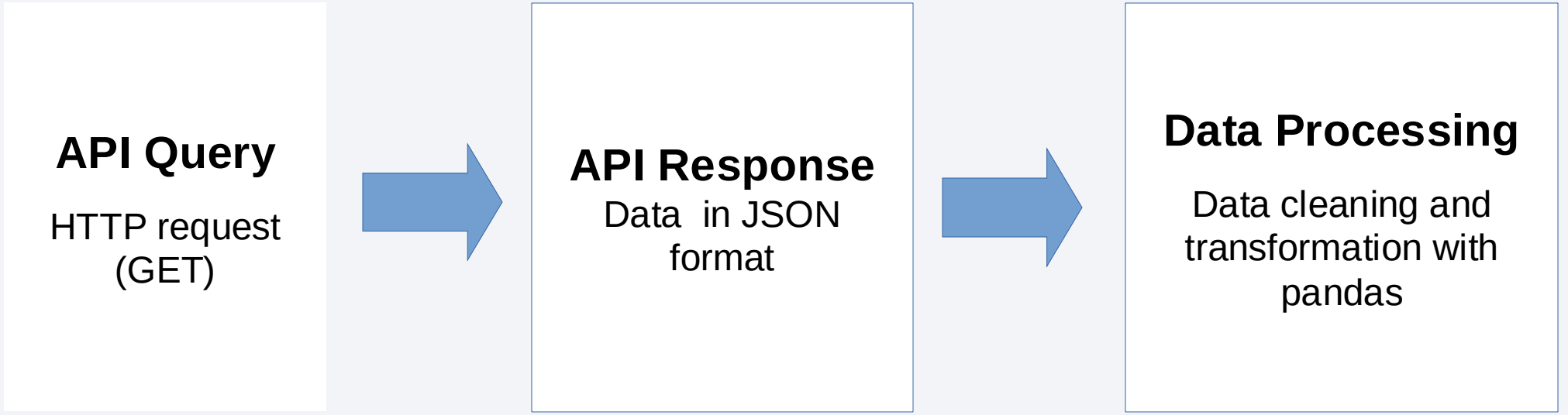
Data Collection

- The data has been obtained from SpaceX's website through an official API.

- Also the data has been obtanined from Wikipedia site.

Data Collection – SpaceX API

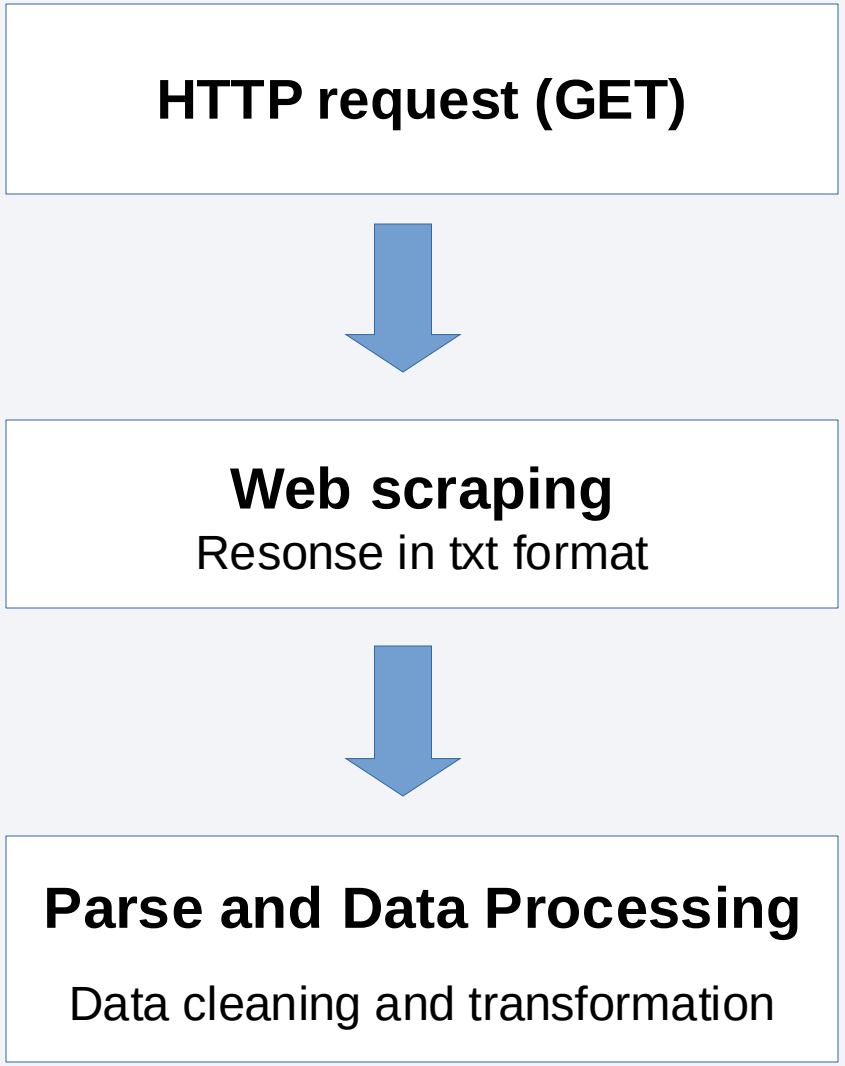
- Present your data collection with SpaceX REST calls using key phrases and flowcharts



- The github link: SpaceX Falcon 9 first stage Landing Prediction

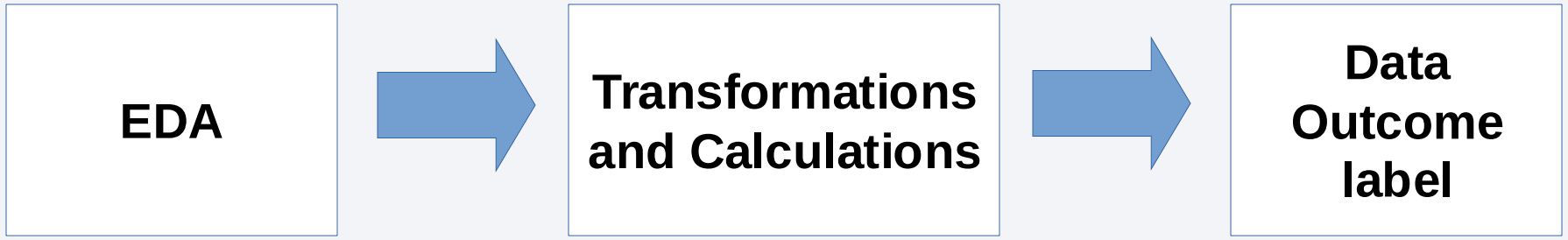
Data Collection - Scraping

- Web scraping proccess
- The github link: Web scraping Falcon 9 and Falcon Heavy Launches Records from Wikipedia



Data Wrangling

- An initial exploratory data exploration (EDA) was performed.
- The number of launches per location and the number of matches per orbit were calculated.
- A landing outcome label was created from outcome column.

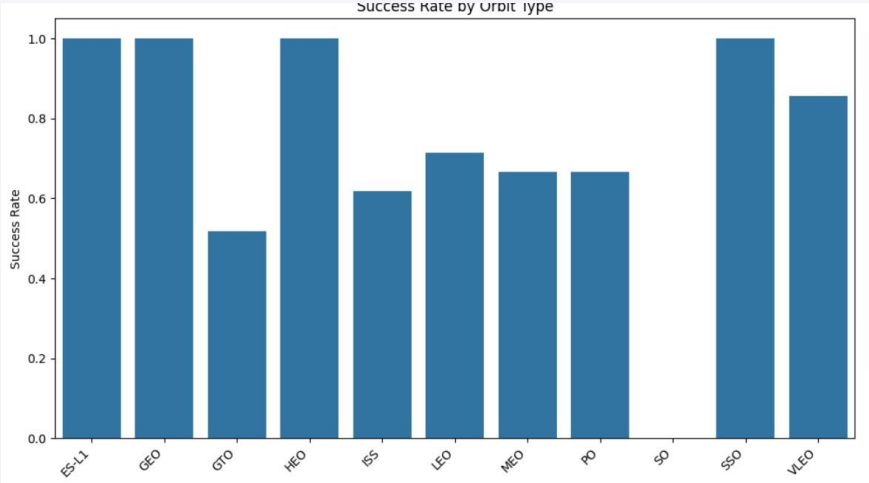


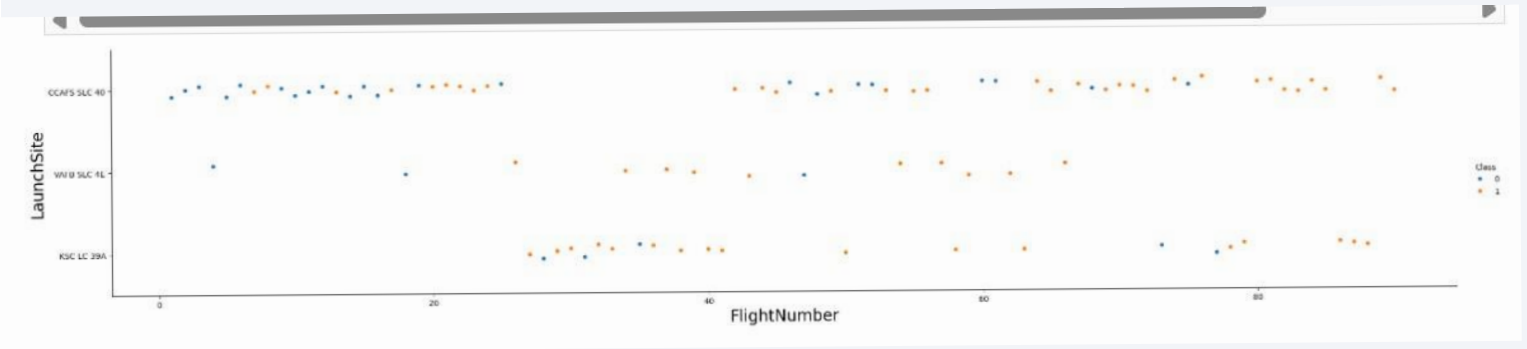
- The github link: Space X Falcon 9 Frist Stage Landing Prediction

EDA with Data Visualization

Bar charts were used to visualize the relationship between success and orbit types.

Scatter charts to observe Playload x Orbit, Launch Site x Flight Number or Launch Site x Playload Mass for example





The github link: [EDA and Data Visualization]

EDA with SQL

With SQL we have obtained next data:

- The names of the unique launch sites in the space mission
- The total payload mass carried by boostes launched by NASA
- The average payload mass carried by booster version F9 v1,1
- The total number of successful and failure missions outcomes
- The names of the booster_versions which have carried the maximum payload mass
- Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order.

The github linnk:EDA with SQLlite

Build an Interactive Map with Folium

- All launch sites have been marked and added map objects such as markers for launch sites, circles for areas around specific coordinates and lines to indicate distances between coordinates.
- The feature launch outcomes have been assigned as class 0 for failure and 1 for success.
- Using the color-labeled marker cluster which launch sites have relatively high success rate.

The github link:Interactive Visual Analytics with Folium

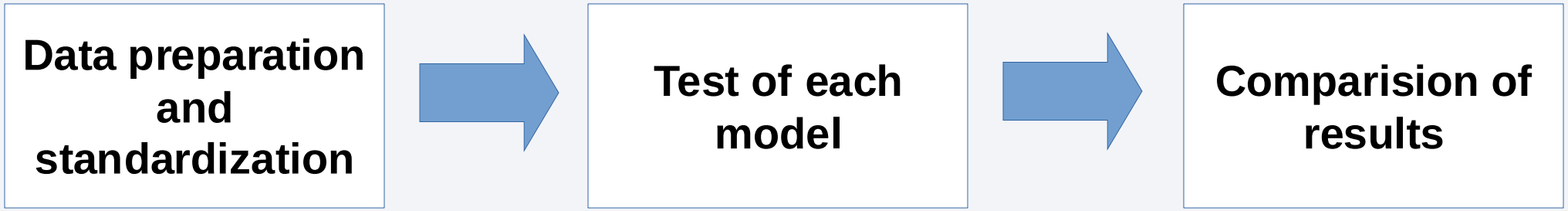
Build a Dashboard with Plotly Dash

- The following graphs were used to visualizate the data:
  - Percentage of launches by site
  - Payload range
- This combination allow analyze the relation between payloads and launch sites, helping to identify where is the best place to launch according to payloads.

The github link: Dash and plot app.

Predictive Analysis (Classification)

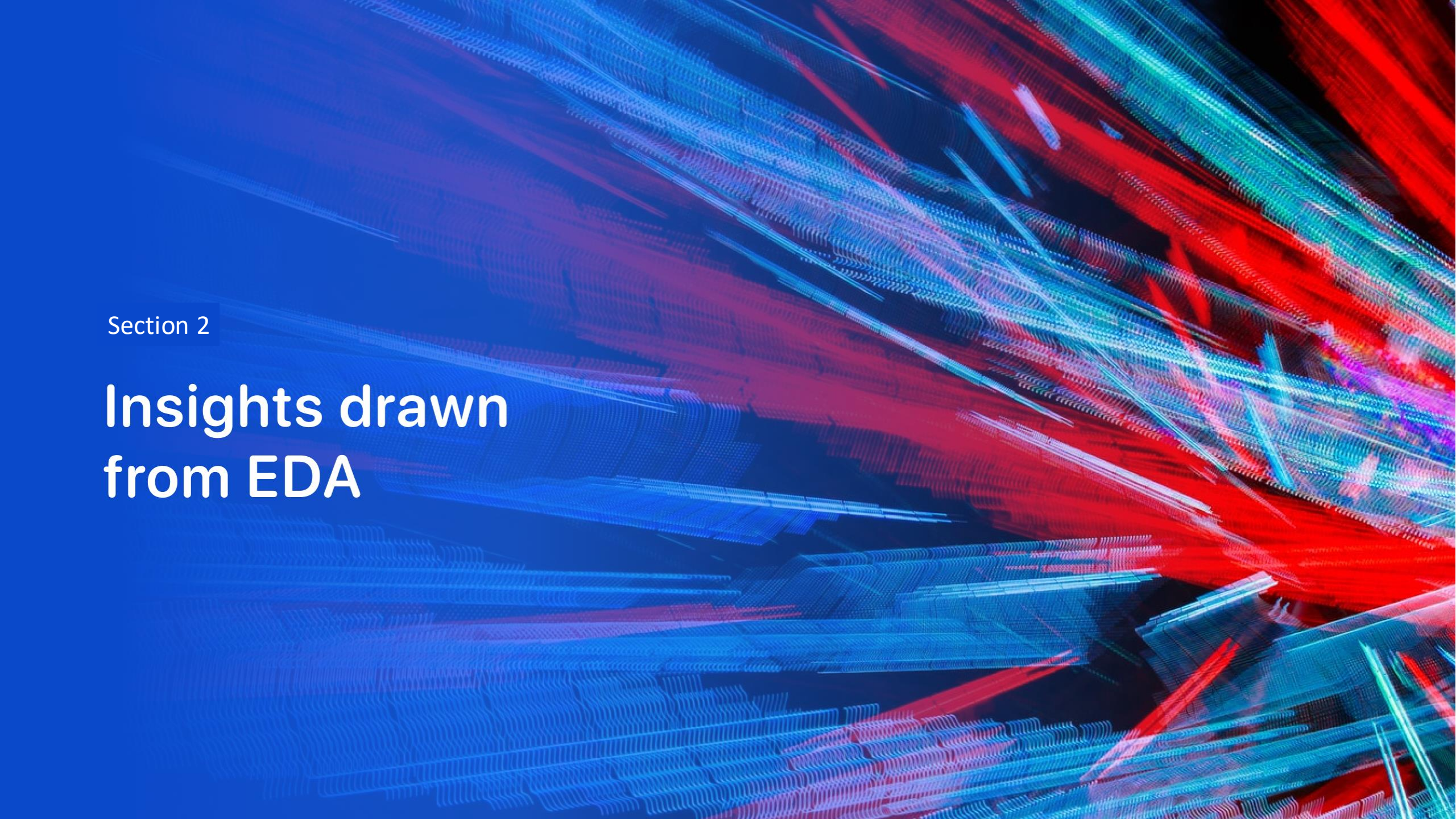
- Four different predictive models have been tested: logarithmic regression, support vector machine, decision tree and k-nearest.
- Different combinations and hyperparameters have been tested and the results have been compared.



The github link: pace X Falcon 9 First Stage Landing Prediction

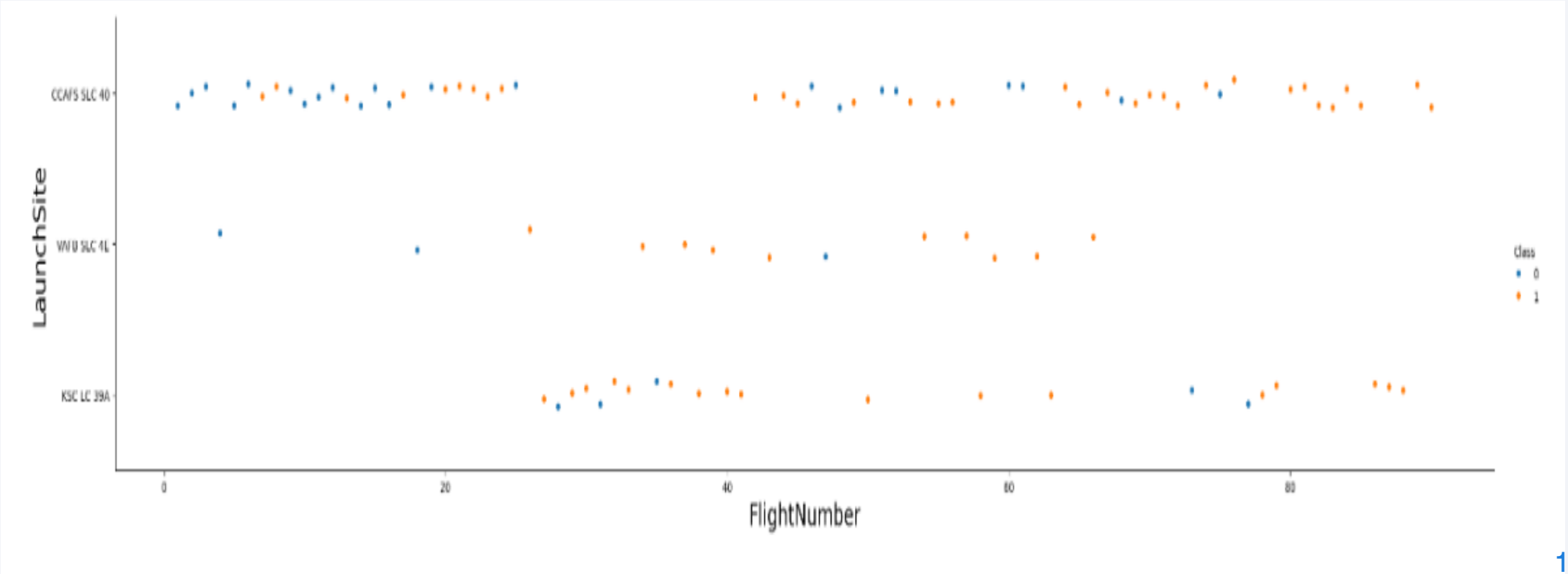
Results

- Exploratory data analysis results:
  - Space X uses 4 different launch sites.
  - The first success landing outcome happened in 2015 five years after the first launch.
  - The number landing outcomes became better as years passed.
  - The best method to predict the success of a landing is Logistic Regression with an accuracy of 83,33%


Section 2
Insights drawn from EDA

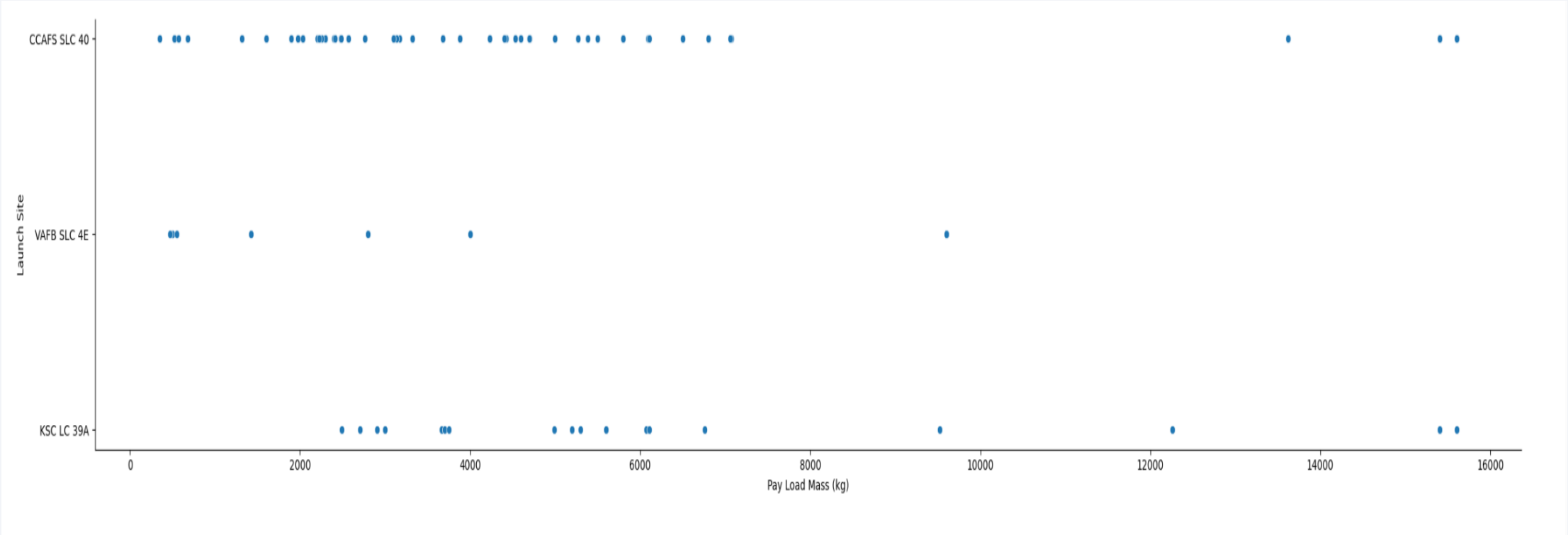
Flight Number vs. Launch Site

- The first flights landed at CCAFS SLC 40, and most flights did too. The second-most landing sites are KSC LS 39A. 0 is failure, 1 is success.

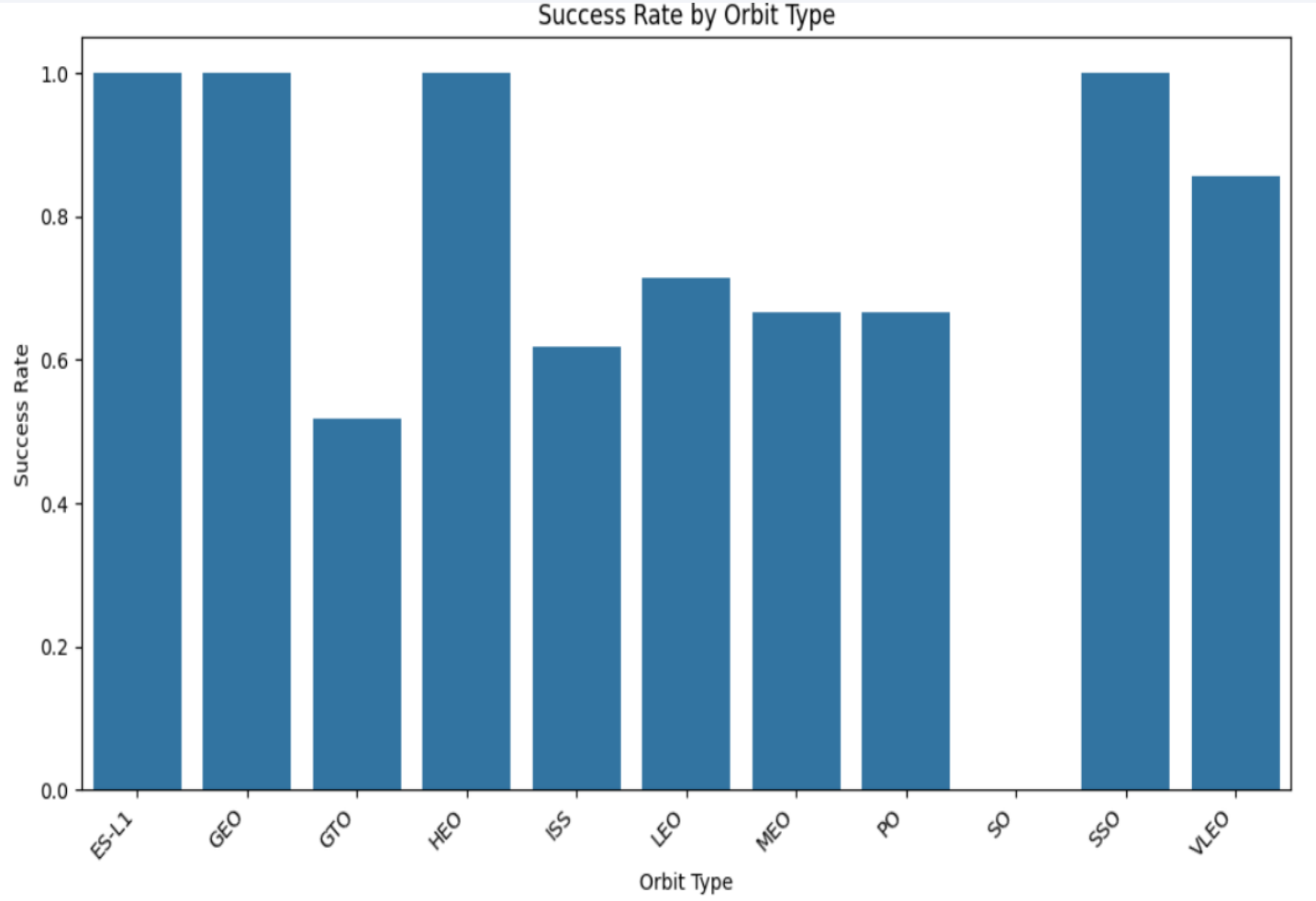


Payload vs. Launch Site

Most of the heavy cargo has been launched from the CCAFS position and only from the VAFB location.



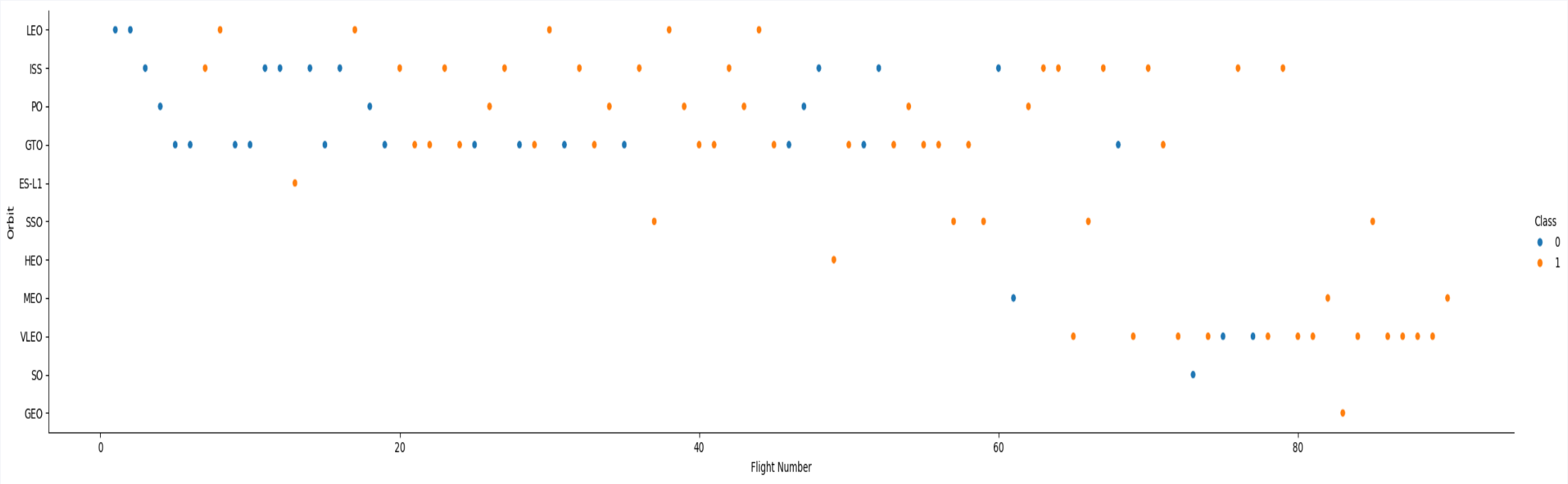
Success Rate vs. Orbit Type



- The orbits with the highest success rates are ES-L1, GEO, HEO and SSO with a 100% success rate

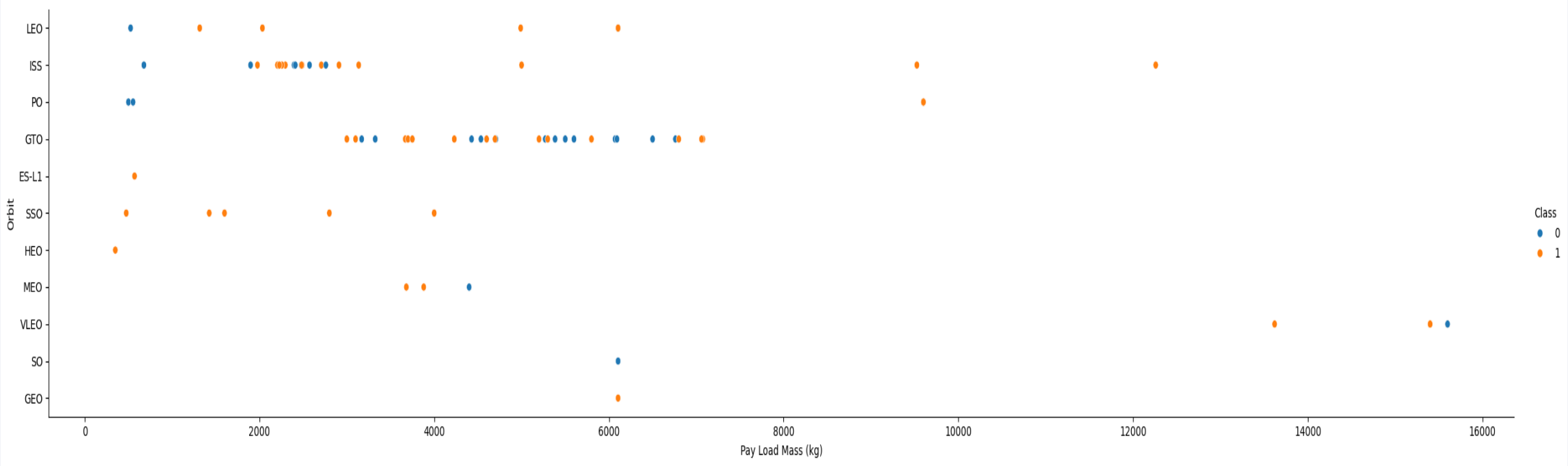
Flight Number vs. Orbit Type

- The largest number of flights went to orbits GTO, PO and ISS. It can also be observed that there is a high success rate in these orbits.



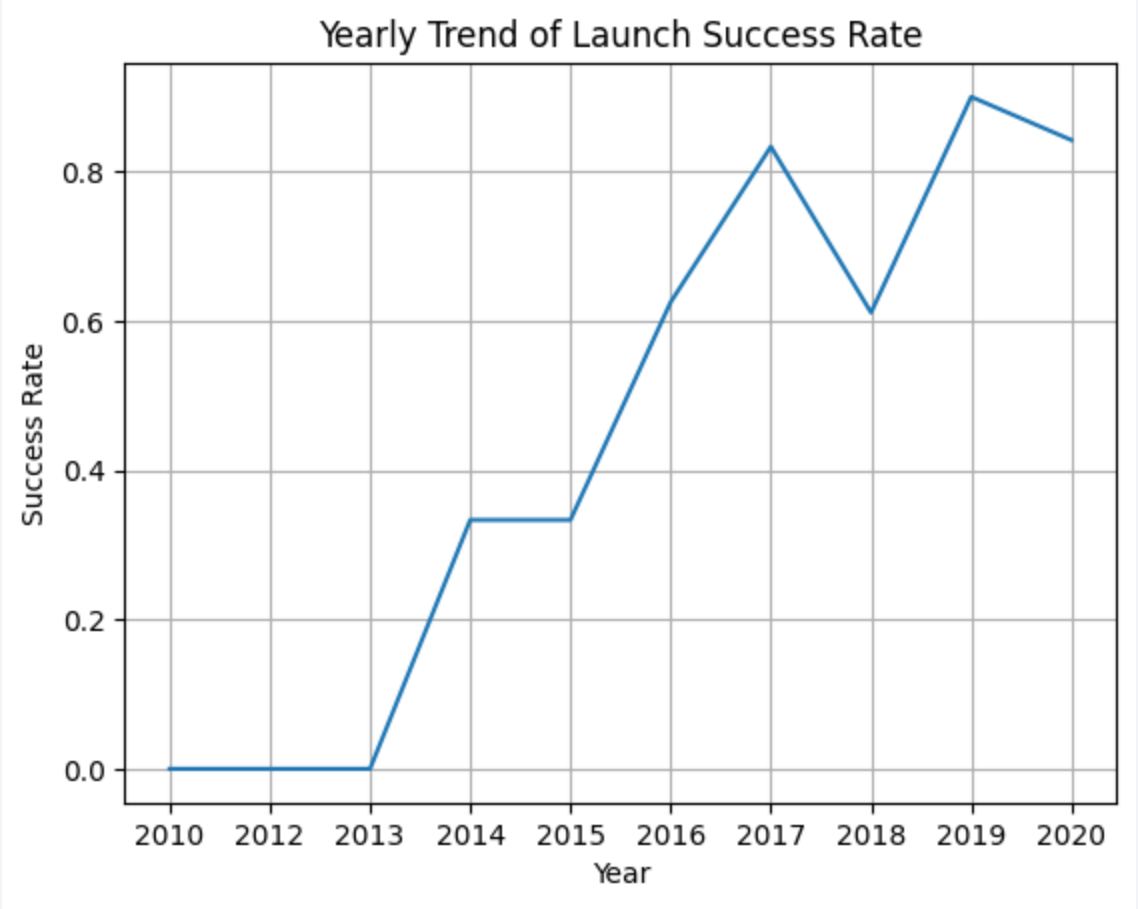
Payload vs. Orbit Type

- The ISS orbit has a high success rate, but smaller payloads are also involved, approximately 2,000 kg.
- The GTO orbit is not entirely clear whether it is very successful, but it has the largest payload volume, between 3,000 and 7,000 kg.
- The largest payloads were launched into the VLO orbit, with values between 15,000 and 16,000 kg, although there were both successes and failures.



Launch Success Yearly Trend

- Since 2013, we've seen a successful launch. In 2019, it achieved its highest success rate.



All Launch Site Names

Find the names of the unique launch sites

The values are selected from the SPACEXTABLE table and the different ones are selected with the DISTINCT command

```
%sql SELECT DISTINCT Launch_Site FROM SPACEXTABLE;
```

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

Launch Site Names Begin with 'CCA'

Find 5 records where launch sites begin with `CCA`.

The values are selected from the SPACEXTABLE table and selected with the WHERE and LIKE commands, with the LIMIT command 5 are chosen.

```
%sql SELECT * FROM SPACEXTABLE WHERE Launch_Site LIKE 'CCA%' LIMIT 5;
```

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

Calculate the total payload carried by boosters from NASA.

The SUM command adds the payload values of the SPACEXTABLE table.

```
%sql SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXTABLE WHERE Customer LIKE
'NASA (CRS)';
```

SUM(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

Calculate the average payload mass carried by booster version F9 v1.1

The AVG command averages the PAYLOAD_MASS_KG_ values, with the WHERE command selecting Booster Version and with LIKE selecting version F9 v1.1

```
%sql SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTABLE WHERE Booster_Version
LIKE 'F9 v1.1';
```

AVG(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

Find the dates of the first successful landing outcome on ground pad

The MIN command selects the date from the Date column and with LIKE that matches Success (ground pad)

```
%sql SELECT MIN(Date) AS FirstSuccessfull_landing_date FROM SPACEXTABLE
WHERE Landing_Outcome LIKE 'Success (ground pad)';
```

FirstSuccessfull_landing_date
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

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

```
%sql SELECT Booster_Version FROM SPACEXTABLE WHERE Landing_Outcome =
'Success (drone ship)' AND PAYLOAD_MASS__KG_ > 4000 AND PAYLOAD_MASS__KG_ <
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

Calculate the total number of successful and failure mission outcomes.

The COUNT command counts all Mission Outcomes

```
%sql SELECT Mission_Outcome, COUNT(*) AS Total FROM SPACEXTABLE GROUP BY
Mission_Outcome;
```

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

Boosters Carried Maximum Payload

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

The SELECT and MAX commands obtain the PAYLOAD_MASS_KG_ data.

```
%sql SELECT Booster_Version FROM SPACEXTABLE
WHERE PAYLOAD_MASS__KG_ = (SELECT
MAX(PAYLOAD_MASS__KG_) FROM SPACEXTABLE);
```

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

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

```
%sql SELECT Booster_Version, Launch_Site, Landing_Outcome FROM SPACEXTABLE
WHERE Landing_Outcome LIKE 'Failure (drone ship)' AND Date BETWEEN '2015-
01-01' AND '2015-12-31';
```

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

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

Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

The DESC command sorts values in descending order.

```
%sql SELECT
Landing_Outcome,
COUNT(Landing_Outcome)
FROM SPACEXTABLE WHERE
DATE BETWEEN '2010-06-04'
AND '2017-03-20' GROUP BY
Landing_Outcome ORDER BY
COUNT(Landing_Outcome)
DESC;
```

Landing_Outcome	COUNT(Landing_Outcome)
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

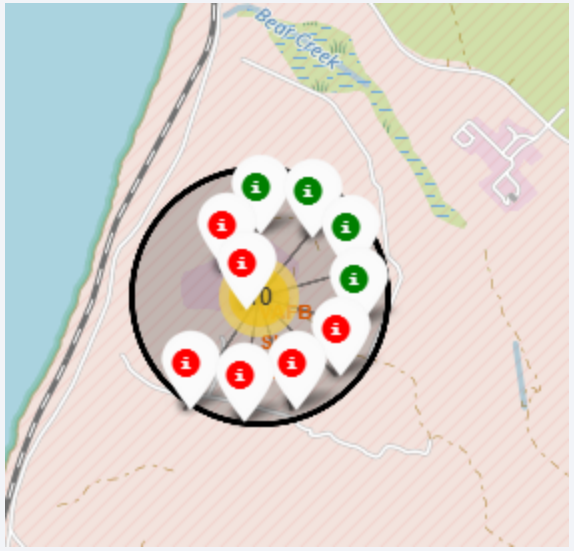

Section 3
Launch Sites
Proximities Analysis

Launch sites with Folium

Landing sites are shown on the east coast in Florida and on the west coast in California.

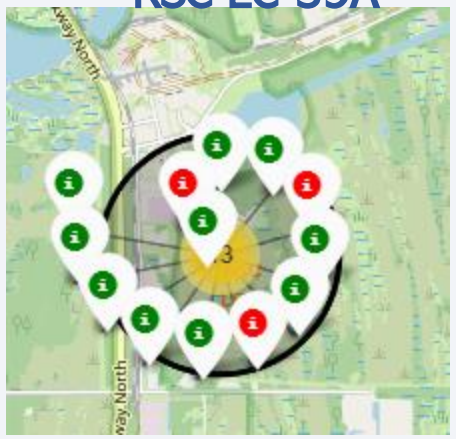
Markers in launch sites with Folium

VAFB SLC-4E California launch site



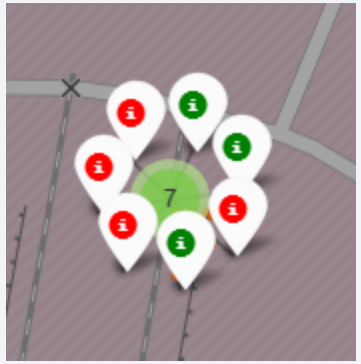
Florida launch site

KSC LC-39A



CCAFS LC-40

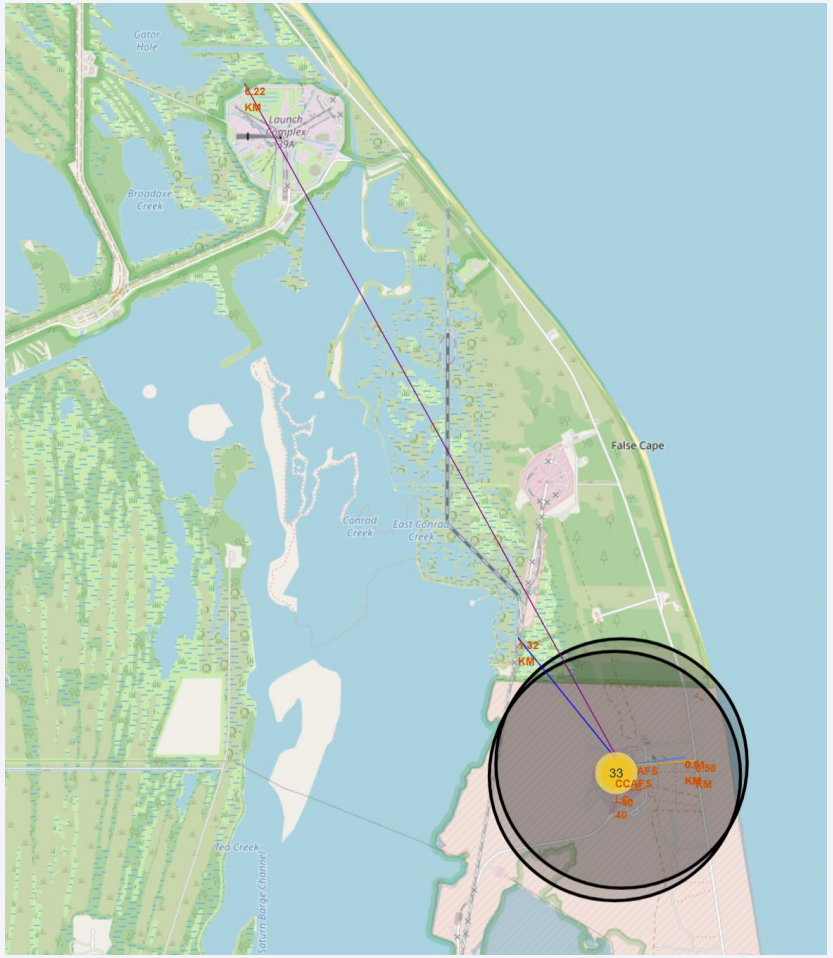
KSC SLC-40

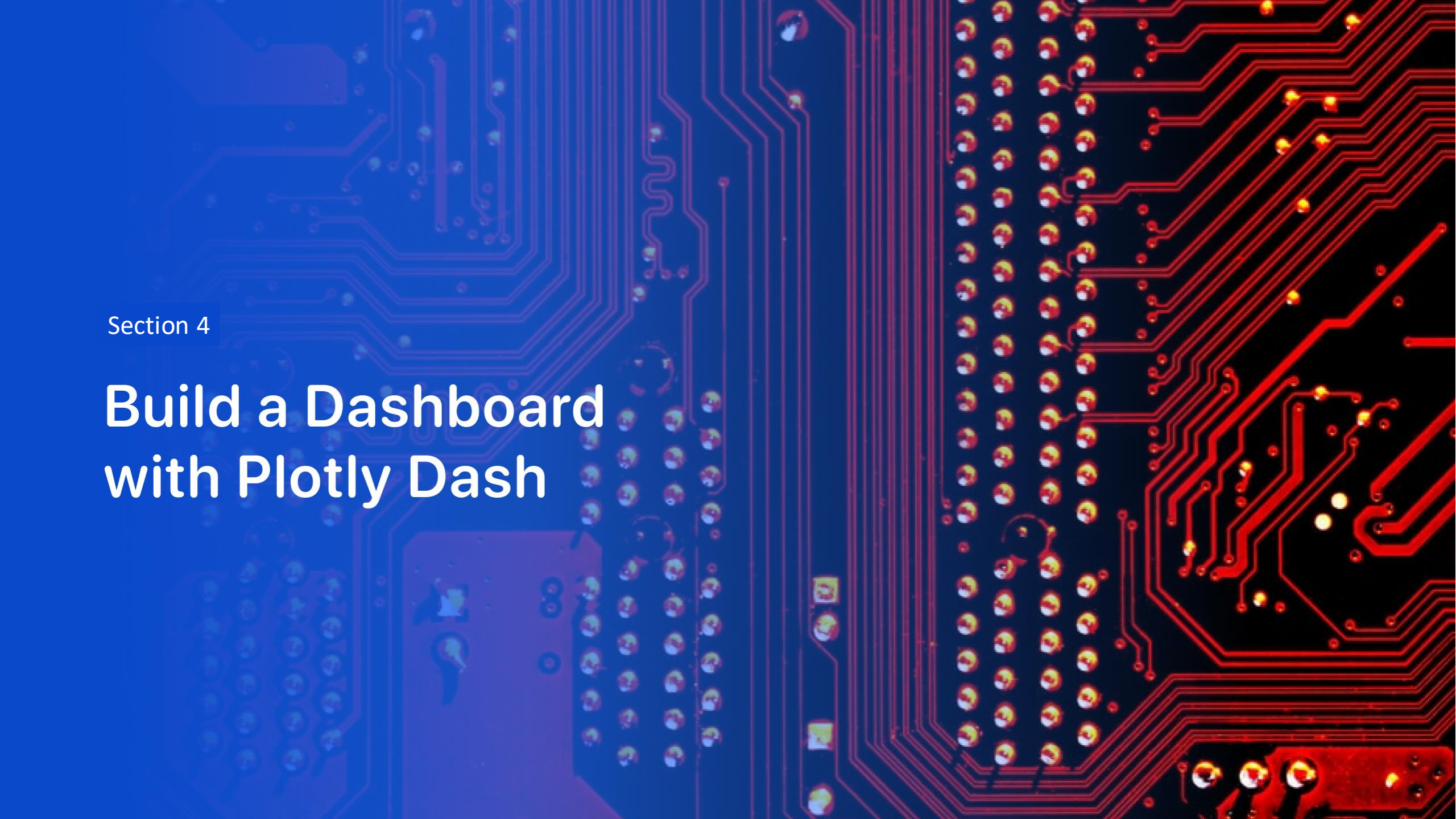


- Green markers are success landings.
- Red markers are failure landings.

Proximity to major infrastructures

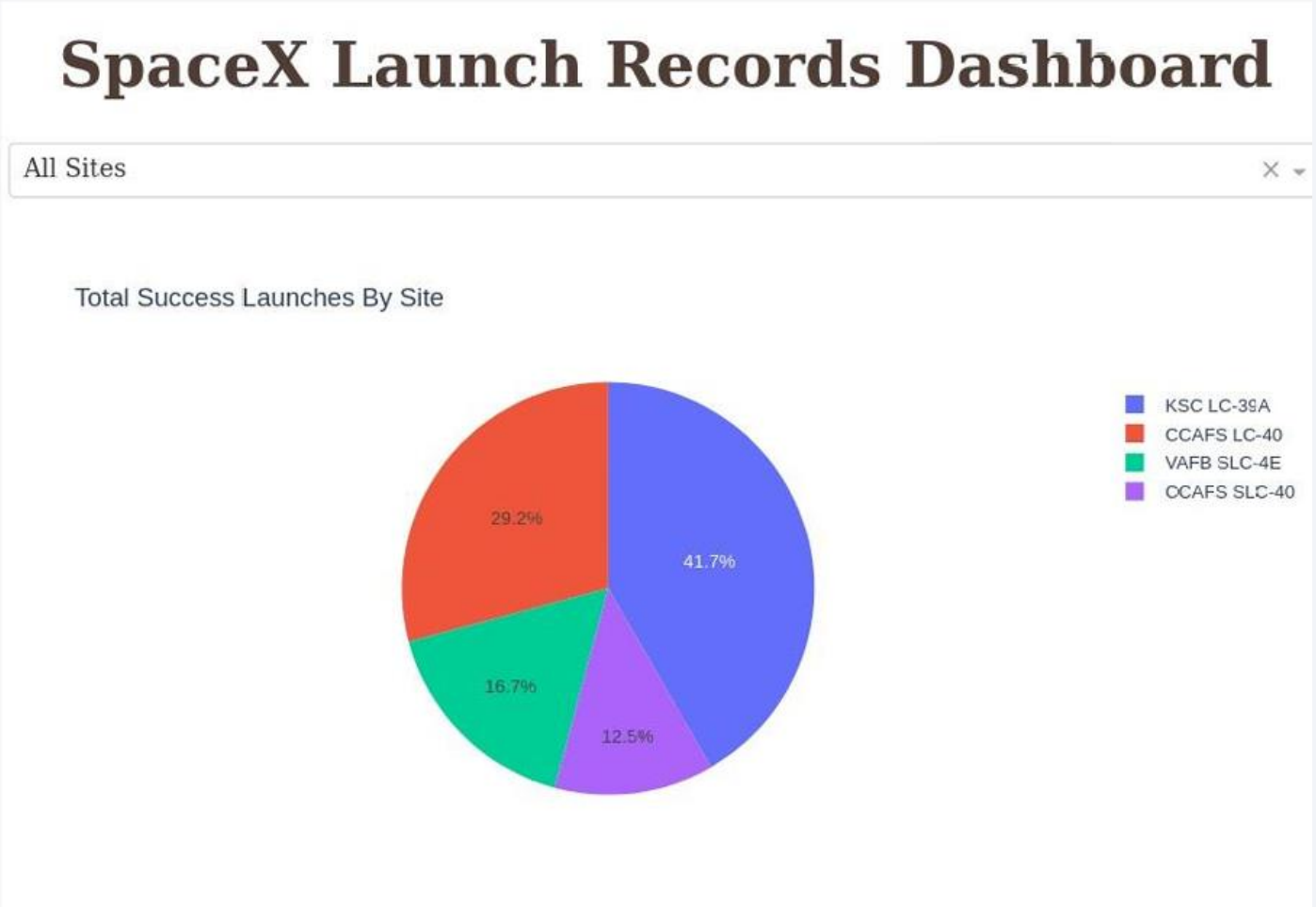
It can be seen that the launch site is not near to major infrastructures so the risk of accident is reduced.




Section 4
Build a Dashboard with Plotly Dash

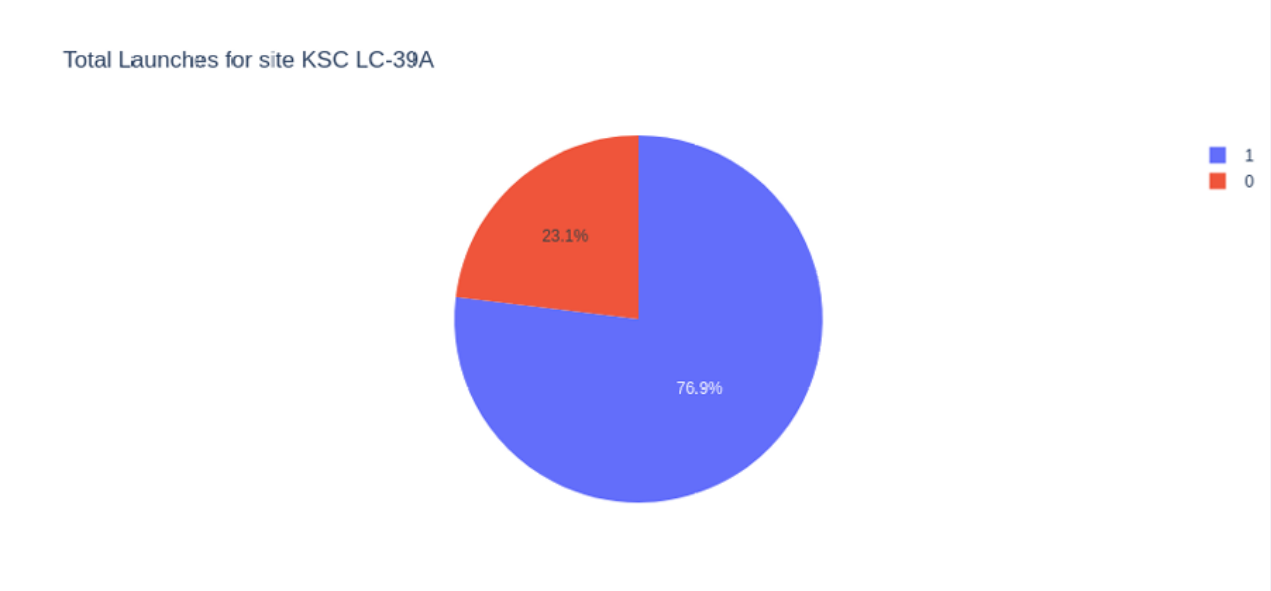
Success launches by site

It can be seen that the site with the highest success rate is KSC LC-39A



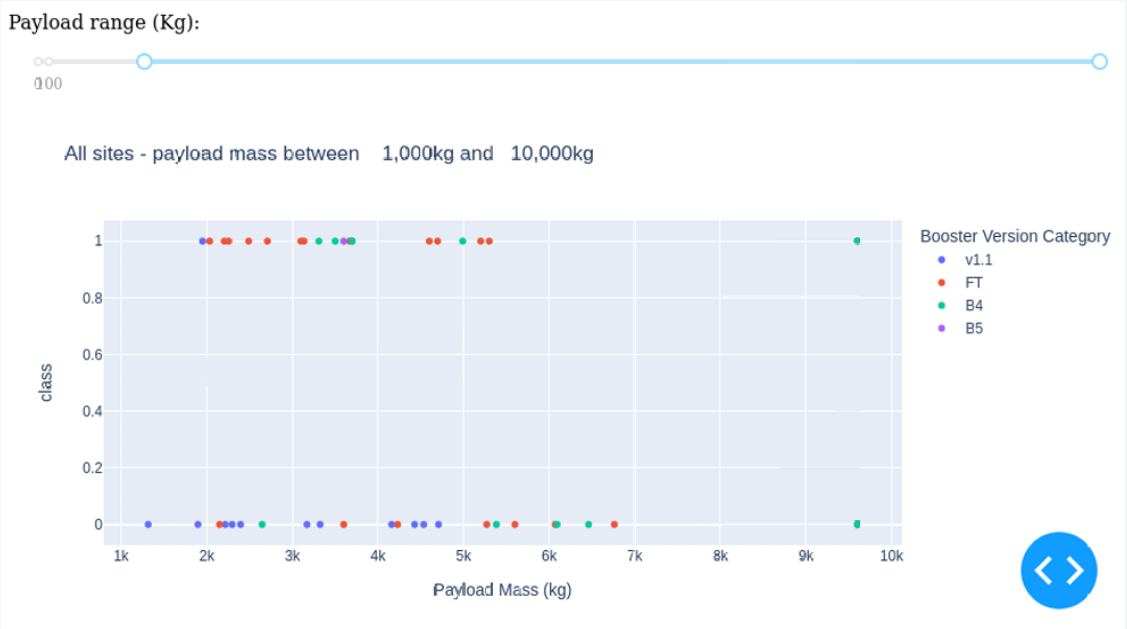
Launch success ratio at KSC LC-39A

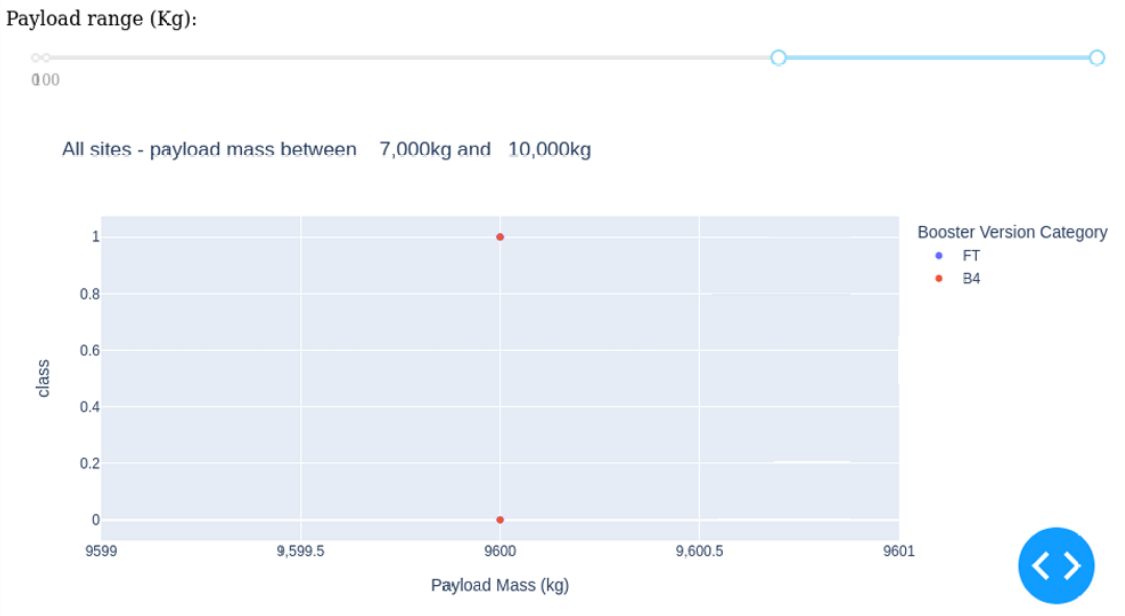
With a 76.9% success rate and 23.1% of failure rate only, KSC LC-39A is the site with the highest launch success rate.



Payload vs Outcome

It can be seen two different scatter plots of different payloads in all launch site.



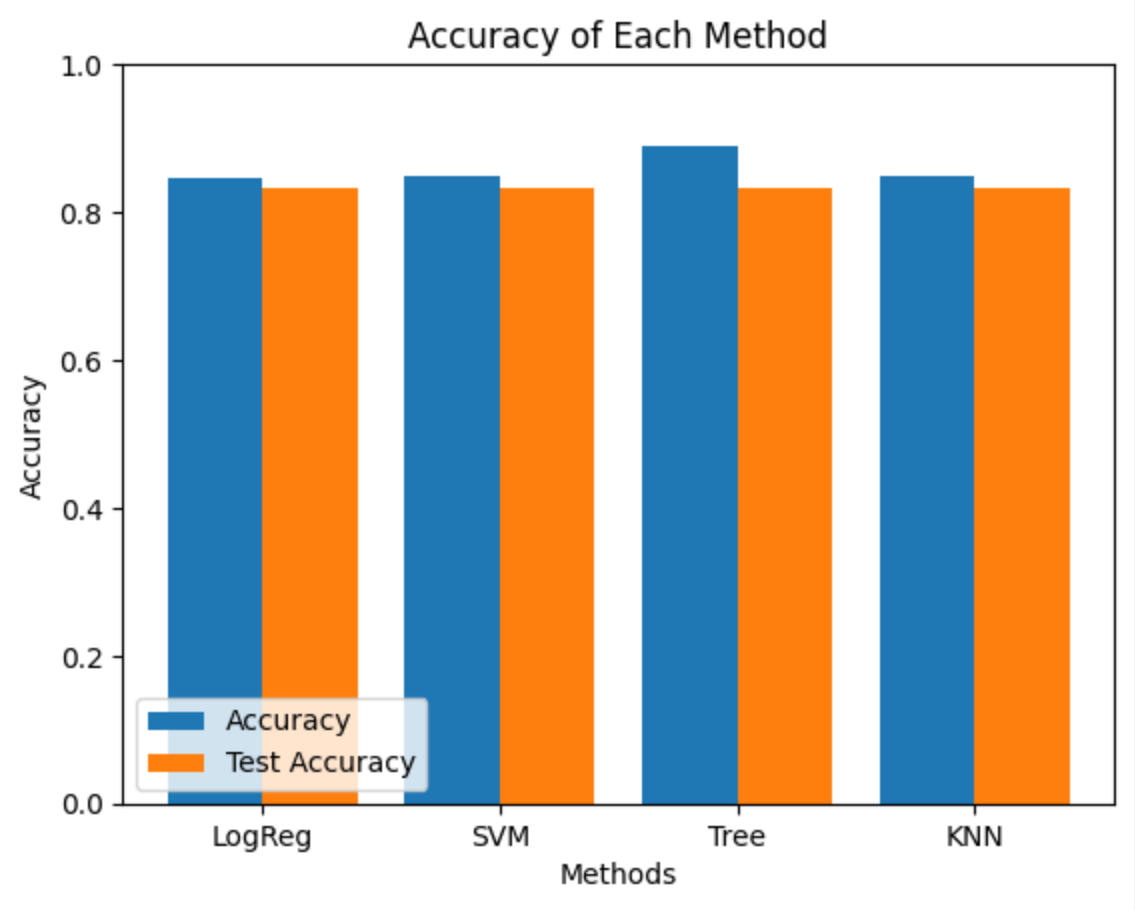


Section 5

Predictive Analysis (Classification)

Classification Accuracy

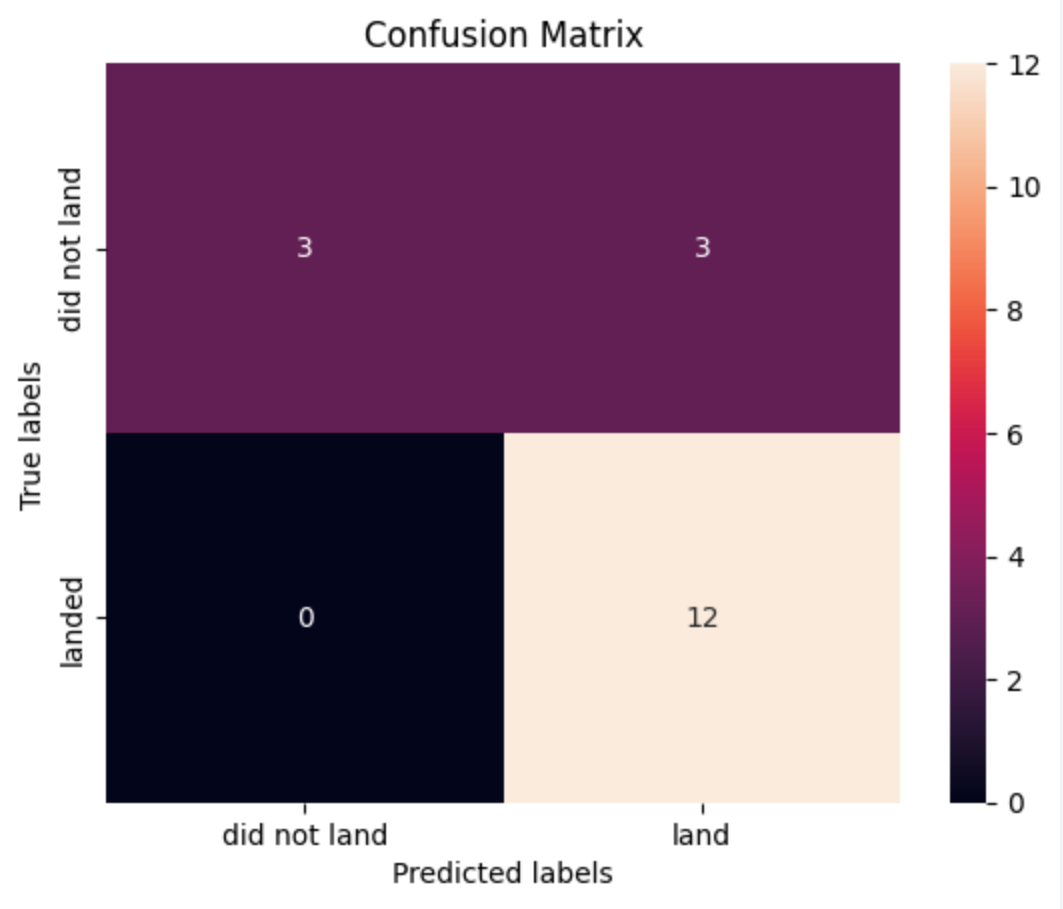
- It can be seen in the graph that the most accurate model is the decision tree with an 88,9% accuracy.



Confusion Matrix

- The rows represent the true labels: "did not land" and "landed".
- The columns represent the predicted labels: "did not land" and "land".

The model correctly predicted 3 instances of "did not land" and 12 instances of "landed". It incorrectly predicted 3 instances as "land" when they did not, and 0 instances as "did not land" when they did.



Conclusions

- Launch success rate started to increase in 2013 till 2020.
- Orbits ES-L1, GEO, HEO, SSO, VLEO had the most success rate.
- The best launch site was KSC LC-39A, it had the most successful launches of any sites.
- The Decision tree classifier is the best machine learning algorithm and have the most accuracy with an 88,9%

Thank you!