IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Johann Drayne
10 - 10 - 2024

Outline

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

Executive Summary

- This report explores SpaceY's space industry competitor SpaceX.
- SpaceX has developed the Falcon 9 which has the ability to reuse the first stage leading to cost-reduction and increased profits.
- By leveraging machine learning models and public data, the success of a launch is predicted up to 83.3% accuracy using parameters such as payload mass and desired orbit.
- Space Y can use these insights to out price SpaceX.

Introduction

BLUE ORIGIN

The commercial space industry is rapidly evolving, with companies like Virgin Galactic, Rocket Lab, Blue Origin, and SpaceX making space travel more accessible and affordable.

SpaceX, in particular has found significant launch cost reduction through the reuse of the Falcon 9's first stage, making space missions more economical and profitable.

This project aims to predict the likelihood of a successful first stage rocket landing.

This can be used as information to calculate the coast of a launch.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
- 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

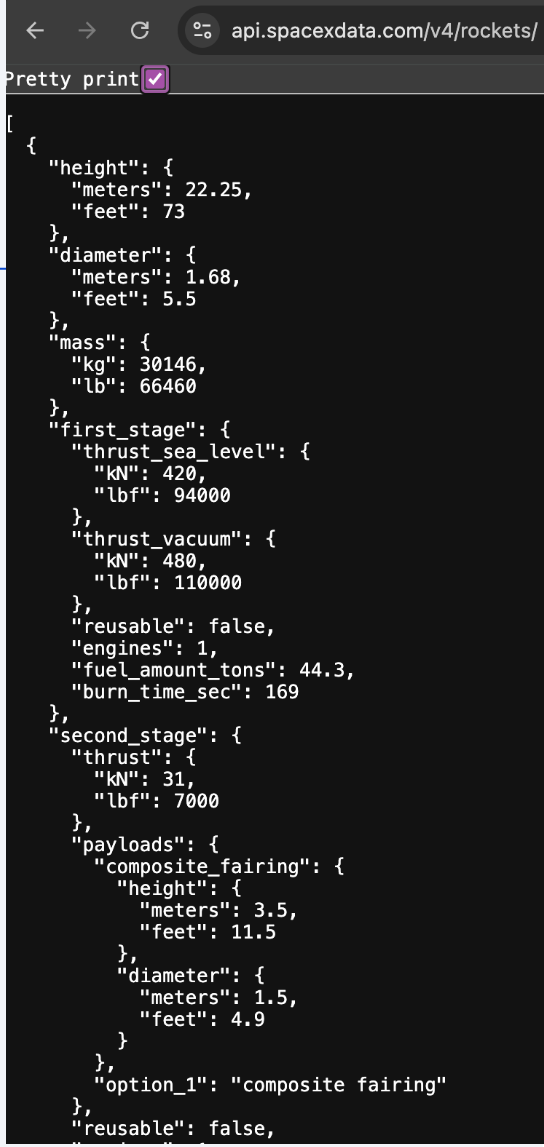
Data Collection – SpaceX API

- API : Open Source REST API for SpaceX

e.g. https://api.spacexdata.com/v4/rockets/

GET request → Only include Falcon 9 launches → Replace missing payload mass with mean

- Github : https://github.com/johann997/IBMDataScience/blob/main/C10-Capstone/M1/jupyter-labs-spacex-data-collection-api.ipynb



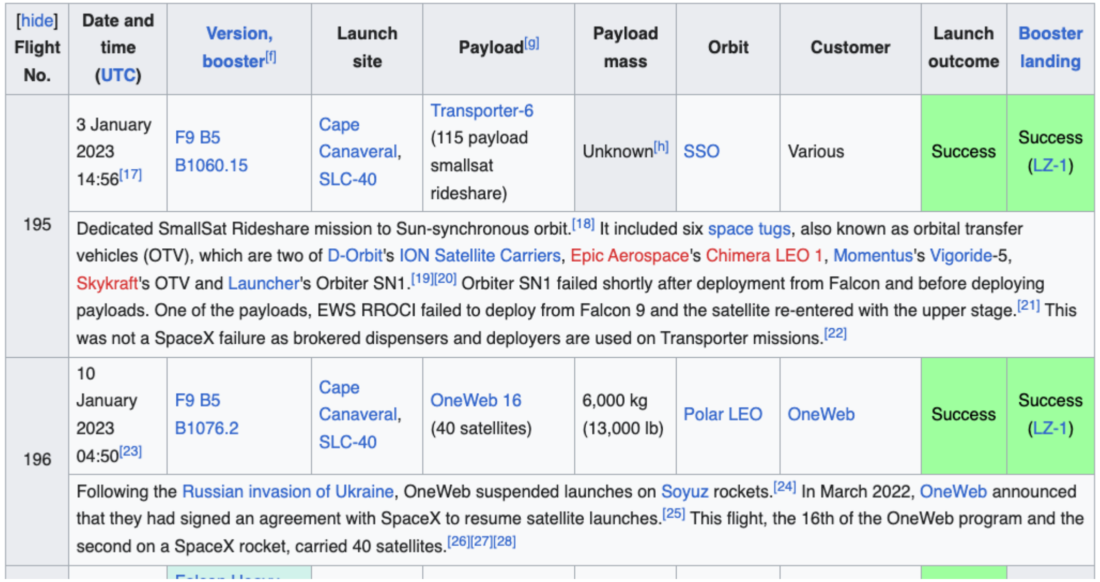

```
[
  {
    "height": {
      "meters": 22.25,
      "feet": 73
    },
    "diameter": {
      "meters": 1.68,
      "feet": 5.5
    },
    "mass": {
      "kg": 30146,
      "lb": 66460
    },
    "first_stage": {
      "thrust_sea_level": {
        "kN": 420,
        "lbf": 94000
      },
      "thrust_vacuum": {
        "kN": 480,
        "lbf": 110000
      },
      "reusable": false,
      "engines": 1,
      "fuel_amount_tons": 44.3,
      "burn_time_sec": 169
    },
    "second_stage": {
      "thrust": {
        "kN": 31,
        "lbf": 7000
      },
      "payloads": {
        "composite_fairing": {
          "height": {
            "meters": 3.5,
            "feet": 11.5
          },
          "diameter": {
            "meters": 1.5,
            "feet": 4.9
          }
        },
        "option_1": "composite fairing"
      },
      "reusable": false,
```

Data Collection - Scraping

- Web Scraping : Wiki page 'List of Falcon 9 and Falcon Heavy launches'

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

GET request → Parse with Beautiful Soup



[hide] Flight No.	Date and time (UTC)	Version, booster[f]	Launch site	Payload[g]	Payload mass	Orbit	Customer	Launch outcome	Booster landing
195	3 January 2023 14:56[17]	F9 B5 B1060.15	Cape Canaveral, SLC-40	Transporter-6 (115 payload smallsat rideshare)	Unknown[h]	SSO	Various	Success	Success (LZ-1)
	Dedicated SmallSat Rideshare mission to Sun-synchronous orbit.[18] It included six space tugs, also known as orbital transfer vehicles (OTV), which are two of D-Orbit's ION Satellite Carriers, Epic Aerospace's Chimera LEO 1, Momentus's Vigoride-5, Skykraft's OTV and Launcher's Orbiter SN1.[19][20] Orbiter SN1 failed shortly after deployment from Falcon and before deploying payloads. One of the payloads, EWS RROCI failed to deploy from Falcon 9 and the satellite re-entered with the upper stage.[21] This was not a SpaceX failure as brokered dispensers and deployers are used on Transporter missions.[22]								
196	10 January 2023 04:50[23]	F9 B5 B1076.2	Cape Canaveral, SLC-40	OneWeb 16 (40 satellites)	6,000 kg (13,000 lb)	Polar LEO	OneWeb	Success	Success (LZ-1)
	Following the Russian invasion of Ukraine, OneWeb suspended launches on Soyuz rockets.[24] In March 2022, OneWeb announced that they had signed an agreement with SpaceX to resume satellite launches.[25] This flight, the 16th of the OneWeb program and the second on a SpaceX rocket, carried 40 satellites.[26][27][28]								

- Github : https://github.com/johann997/IBMDataScience/blob/main/C10-Capstone/M1/jupyter-labs-webscraping.ipynb

Data Wrangling

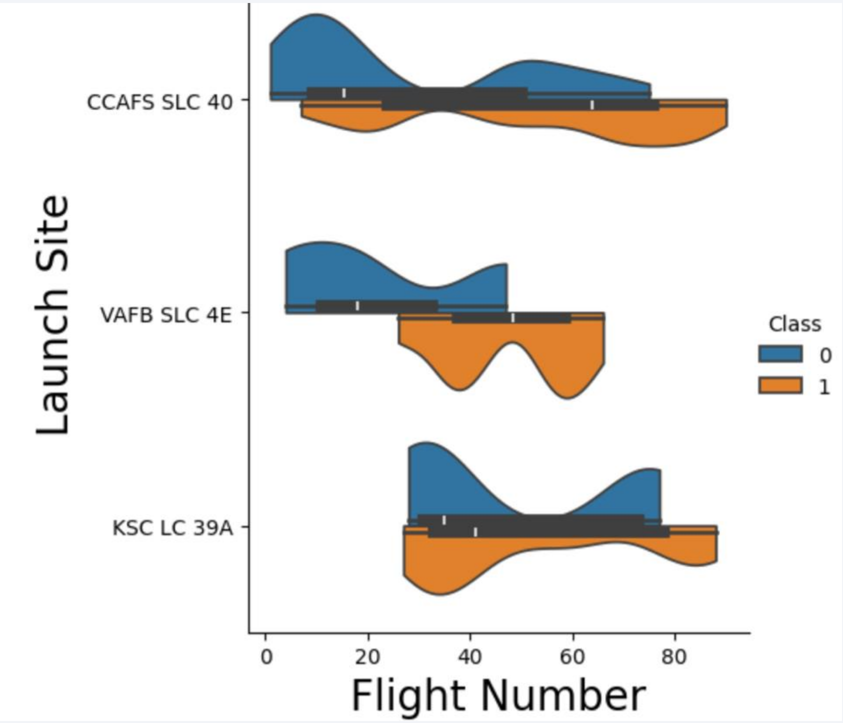
- Exploratory analysis
    - Identify categorical vs numerical columns
    - Calculate number of launch sites per site
    - Calculate occurence per orbit
    - Calculate occurence of mission outcomes
    - Categorise mission outcomes in successs or failure by adding column 'class' with values 0 or 1.

- Red = class 0
- Green = class 1

True ASDS
None None
True RTLS
False ASDS
True Ocean
False Ocean
None ASDS
False RTLS

- Github : https://github.com/johann997/IBMDataScience/blob/main/C10-Capstone/M1/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization



- Types of plots
- Launch site vs. Flight Number (separated by class)
- Launch site vs. Payload Mass (separated by class)
- Success Rate vs. Orbit Type
- Orbit Type vs. Flight Number (separated by class)
- Orbit Type vs. Payload Mass (separated by class)
- Success Rate vs. Year

Want to understand what parameters affect the success rate of each flight.

- It seems Flight Number, Launch site, Orbit Type, Payload Mass all affect the success rate in different ways.

- Github : https://github.com/johann997/IBMDataScience/blob/main/C10-Capstone/M2/edadataviz.ipynb

EDA with SQL

Loaded data into an IBD database instance. Used sql magic to find information on:

- Found unique launch site names.
- Total payload mass carried by boosters launched by NASA (CRS)
- Average payload mass carried by booster version F9 v1.1
- Date when the first successful landing outcome in ground pad was achieved.
- Total number of successful and failure mission outcomes.
- Names of the booster versions which have carried the maximum payload mass.
- List the records which will display the month names, failure landing outcomes in drone ship, booster versions, launch site for the months in year 2015.
- Github : https://github.com/johann997/IBMDataScience/blob/main/C10-Capstone/M2/jupyter-labs-eda-sql-coursera_sqllite.ipynb

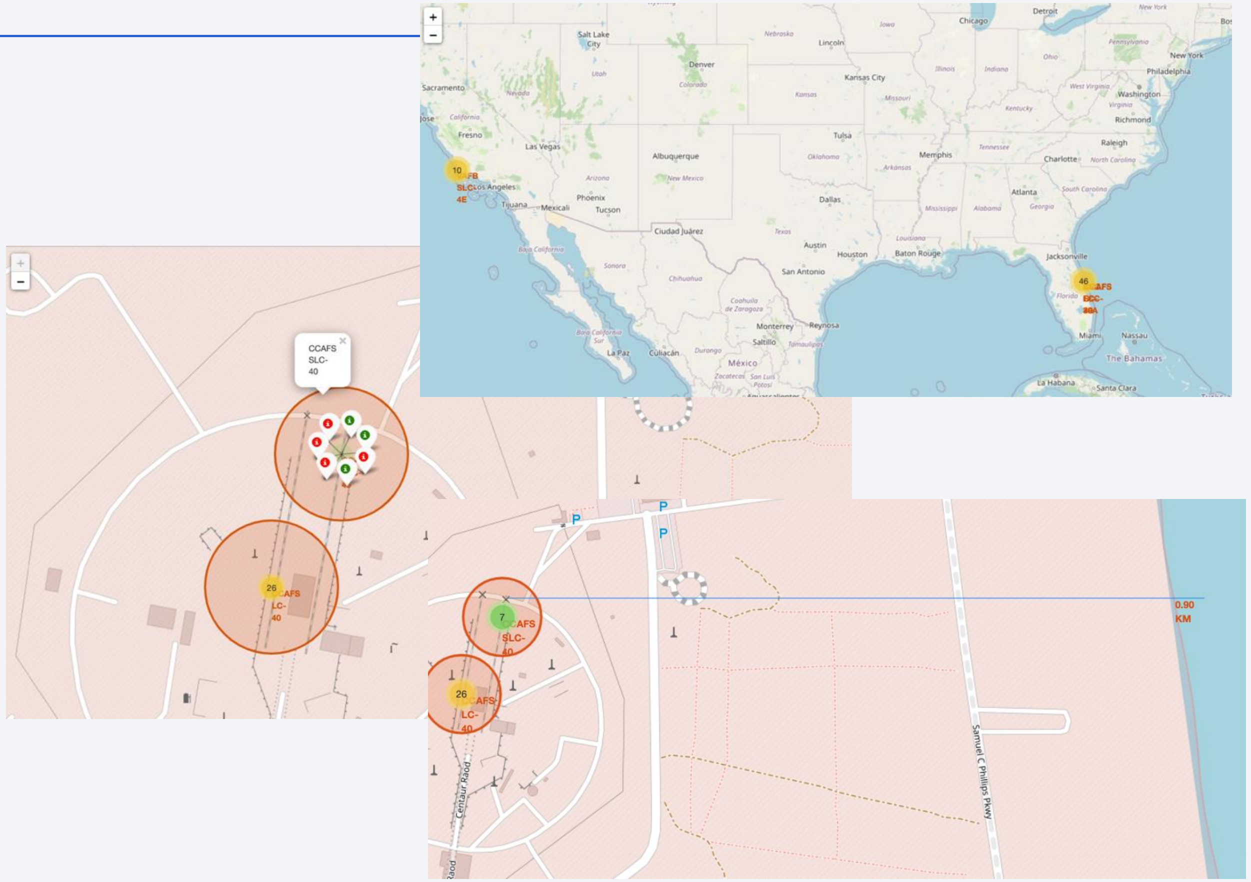
```
%sql SELECT DISTINCT "launch_site" FROM SPACEXTABLE;
```

```
* sqlite:///my_data1.db
Done.
```

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

Build an Interactive Map with Folium

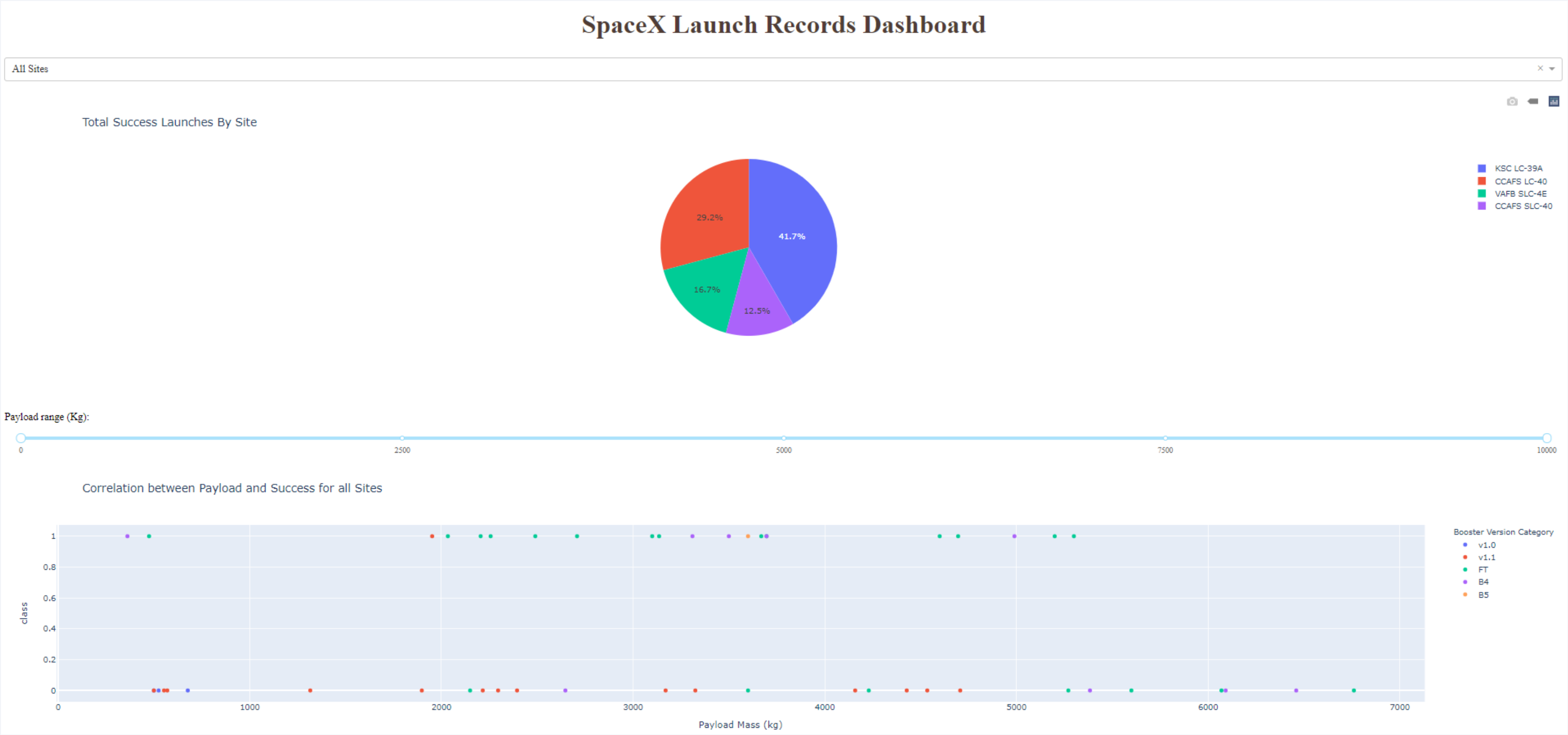
- Added launch sites as a circle and marker, launch attempts with coloured markers and also polylines connecting launch site to nearby infrastructure with calculated distance.
- Launch attempts visually showed how successful a launch site was. And polylines showed how close the site were to coastline, road and rail networks and far from cities for safety.



- Github : https://github.com/johann997/IBMDataScience/blob/main/C10-Capstone/M3/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- Dropdown to include ‘all’ launch site or choose a specific one.
- Pie graph showing success share between ‘all’ sites or success/failure for a specific site.
- Range slide to choose payload range.
- Scatter plot of class vs. payload mass split up by booster version categories.
- This helped quickly show differences between launch site success rates and optimal payload ranges.



- Github : https://github.com/johann997/IBMDataScience/blob/main/C10-Capstone/M3/spacex_dash_app.py

Predictive Analysis (Classification)

- Classification was chosen to predict as a means to predict whether a rocket launch will succeed or not.
- Data was standardised and split into train / test cohorts.
- Models tested:
  - SVM
  - Classification Trees
  - Logistic Regression
  - KNN

Hyperparameters were tuned using a cross-validated grid search

Standardised independent data. → Split data into train / test 80 / 20 → Set up classification model → Use GridSearch to iterate over params → Evaluate with accuracy and plot confusion matrix

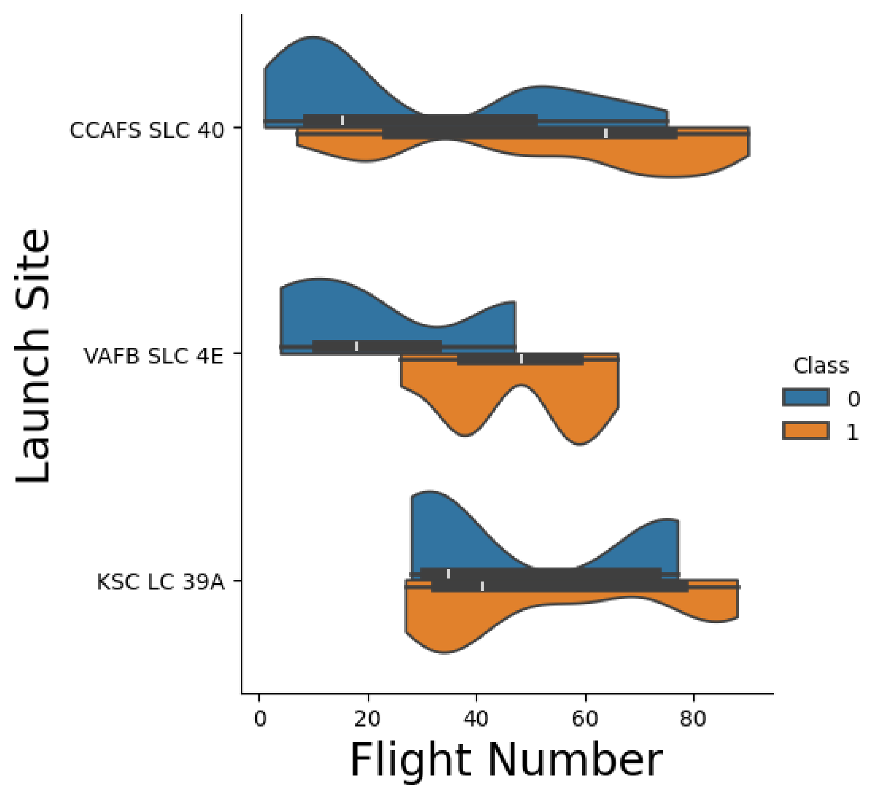
- Github : https://github.com/johann997/IBMDataScience/blob/main/C10-Capstone/M4/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Section 2

Insights drawn from EDA

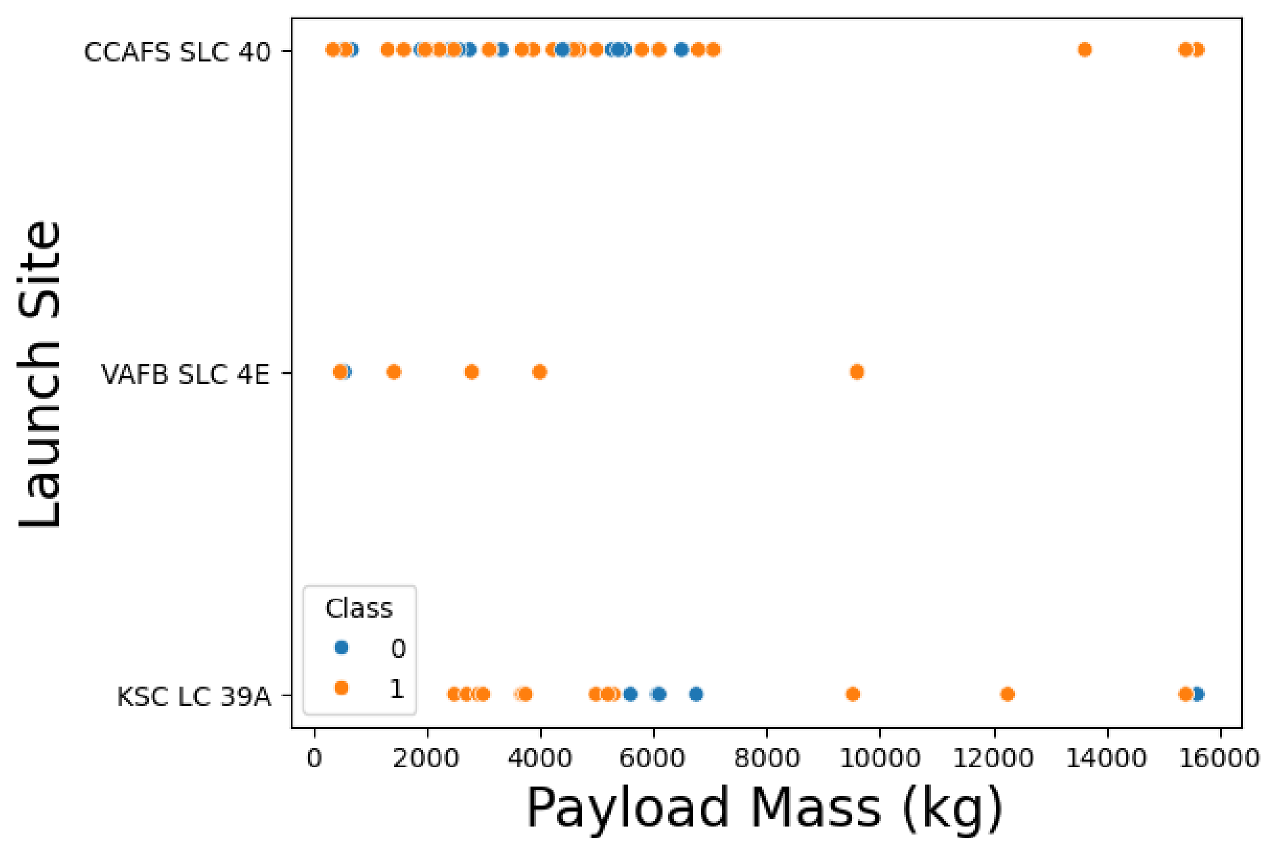
Flight Number vs. Launch Site

- A scatter plot was not specified so I choose a violin plot.
- Here we clearly see the mean flight number for successful classes being larger than for unsuccessful classes. i.e. flight become more successful after further iterations.
- The difference is most significant in launch sites
  - CCAFS SLC 40
  - VAFB SLC4E



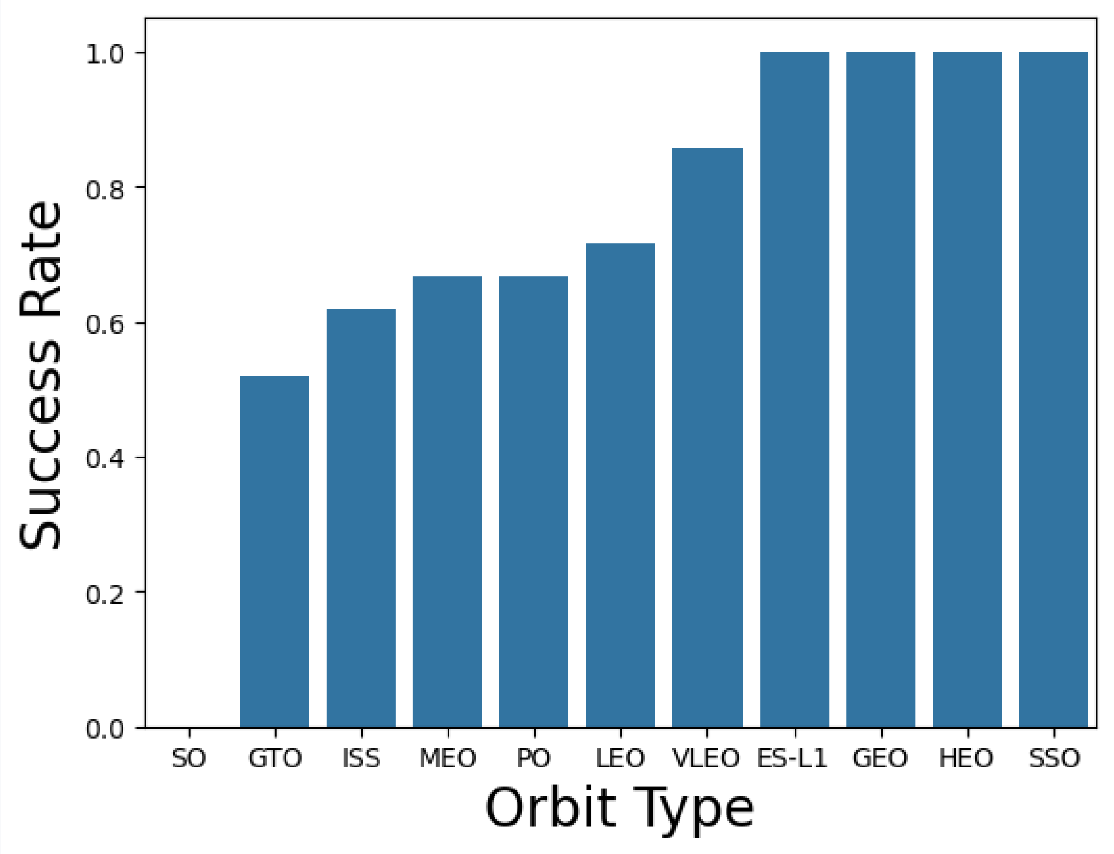
Payload vs. Launch Site

- For the VAFB-SLC launch site there are no rockets launched for heavy payload mass(greater than 10000).



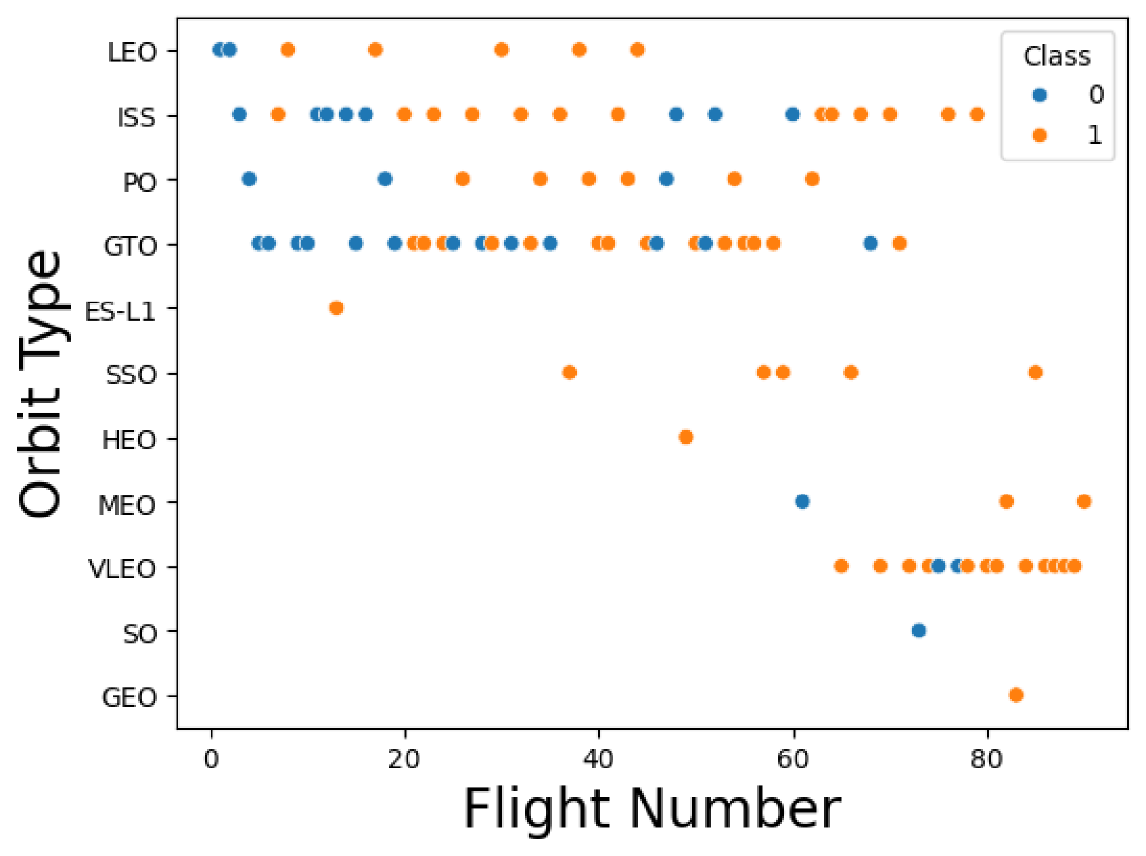
Success Rate vs. Orbit Type

- Shows that no flights were successful at SO orbit.
- Orbits SSO, HEO, GEO and ES-L1 have 100% success rate.



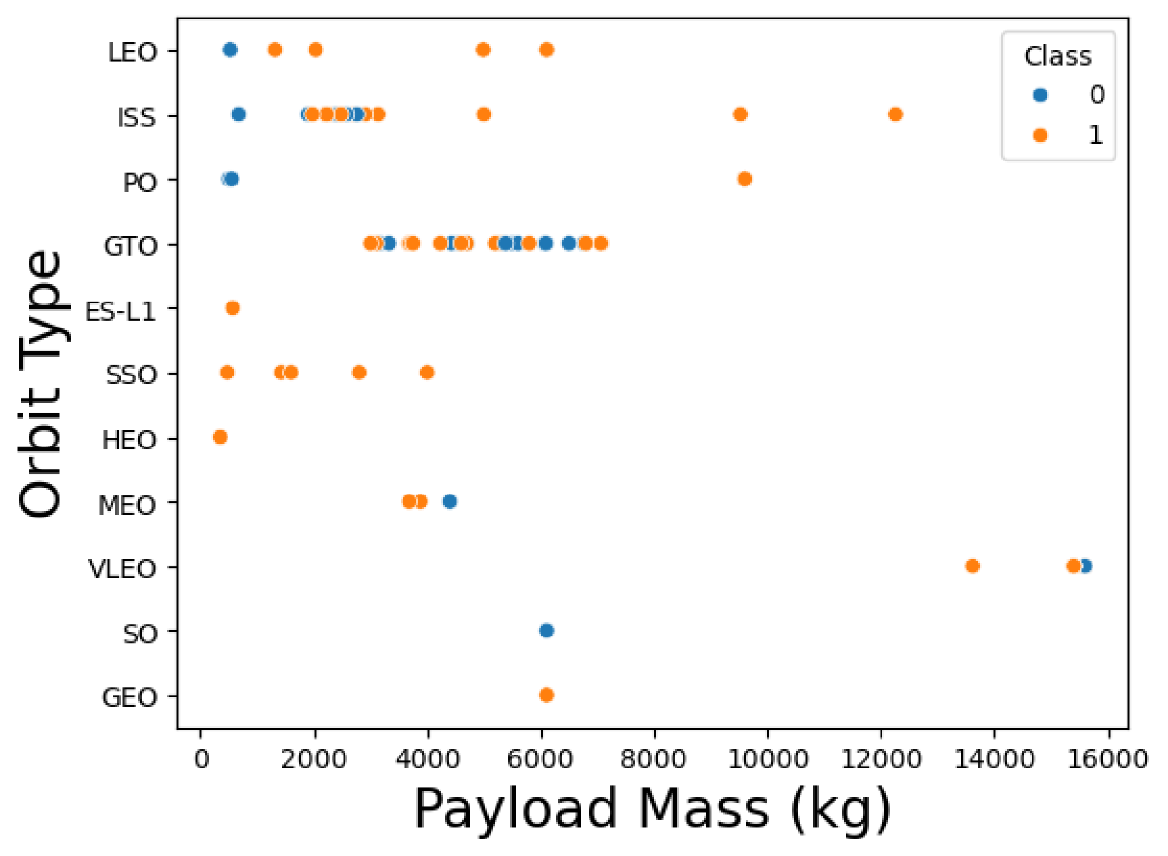
Flight Number vs. Orbit Type

- In the LEO orbit, success seems to be related to the number of flights.
- In the GTO orbit, there appears to be no relationship between flight number and success.



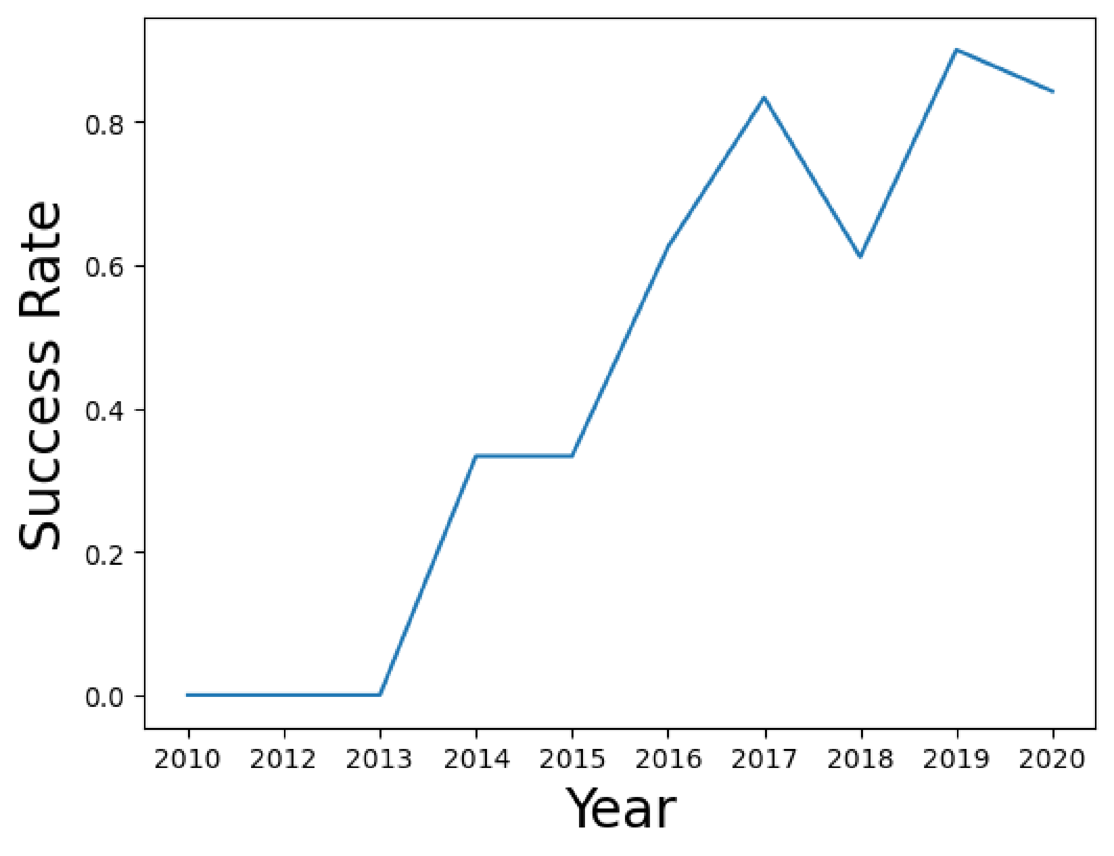
Payload vs. Orbit Type

- With heavy payloads, the successful landing or positive landing rate are more for Polar, LEO and ISS.
- For GTO, it's difficult to distinguish between successful and unsuccessful landings, as both outcomes are present.



Launch Success Yearly Trend

- The success rate was zero until 2013 and then kept increasing till 2020.



All Launch Site Names

%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'

%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_
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (p
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 (p
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	N
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	N
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	N

Total Payload Mass

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

AVG(PAYLOAD_MASS__KG_)
2928.4

Average Payload Mass by F9 v1.1

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

AVG(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

```sql
%sql SELECT MIN(DATE) FROM SPACEXTABLE WHERE Landing_Outcome == 'Success (ground pad)';
```

MIN(DATE)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

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

```
%%sql
SELECT
SUM(CASE WHEN Landing_Outcome LIKE '%success%' THEN 1 ELSE 0 END) AS total_successful,
SUM(CASE WHEN Landing_Outcome NOT LIKE '%success%' THEN 1 ELSE 0 END) AS total_failed
 FROM SPACEXTABLE;
```

total_successful	total_failed
61	40

Boosters Carried Maximum Payload

```
%%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

```
%%sql
SELECT
CASE
WHEN SUBSTR(Date, 6, 2) = '01' THEN 'January’
WHEN SUBSTR(Date, 6, 2) = '02' THEN 'February’
WHEN SUBSTR(Date, 6, 2) = '03' THEN 'March’
WHEN SUBSTR(Date, 6, 2) = '04' THEN 'April’
WHEN SUBSTR(Date, 6, 2) = '05' THEN 'May’
WHEN SUBSTR(Date, 6, 2) = '06' THEN 'June’
WHEN SUBSTR(Date, 6, 2) = '07' THEN 'July’
WHEN SUBSTR(Date, 6, 2) = '08' THEN 'August’
WHEN SUBSTR(Date, 6, 2) = '09' THEN 'September’
WHEN SUBSTR(Date, 6, 2) = '10' THEN 'October’
WHEN SUBSTR(Date, 6, 2) = '11' THEN 'November’
WHEN SUBSTR(Date, 6, 2) = '12' THEN 'December' END AS month_name, Landing_Outcome, booster_version, launch_site
FROM SPACEXTABLE
WHERE
 SUBSTR(Date, 0, 5) = '2015’
AND Landing_Outcome NOT LIKE '%success%’
AND Landing_Outcome LIKE '%Drone Ship%';
```

month_name	Landing_Outcome	Booster_Version	Launch_Site
January	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
April	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40
June	Precluded (drone ship)	F9 v1.1 B1018	CCAFS LC-40

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

```
%%sql
SELECT
   Landing_Outcome,
   COUNT(*) AS outcome_count
FROM
   SPACEXxaTABLE
WHERE
   Date BETWEEN '2010-06-04' AND '2017-03-20'
GROUP BY
   Landing_Outcome
ORDER BY
   outcome_count DESC;
```

Landing_Outcome	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

Section 3

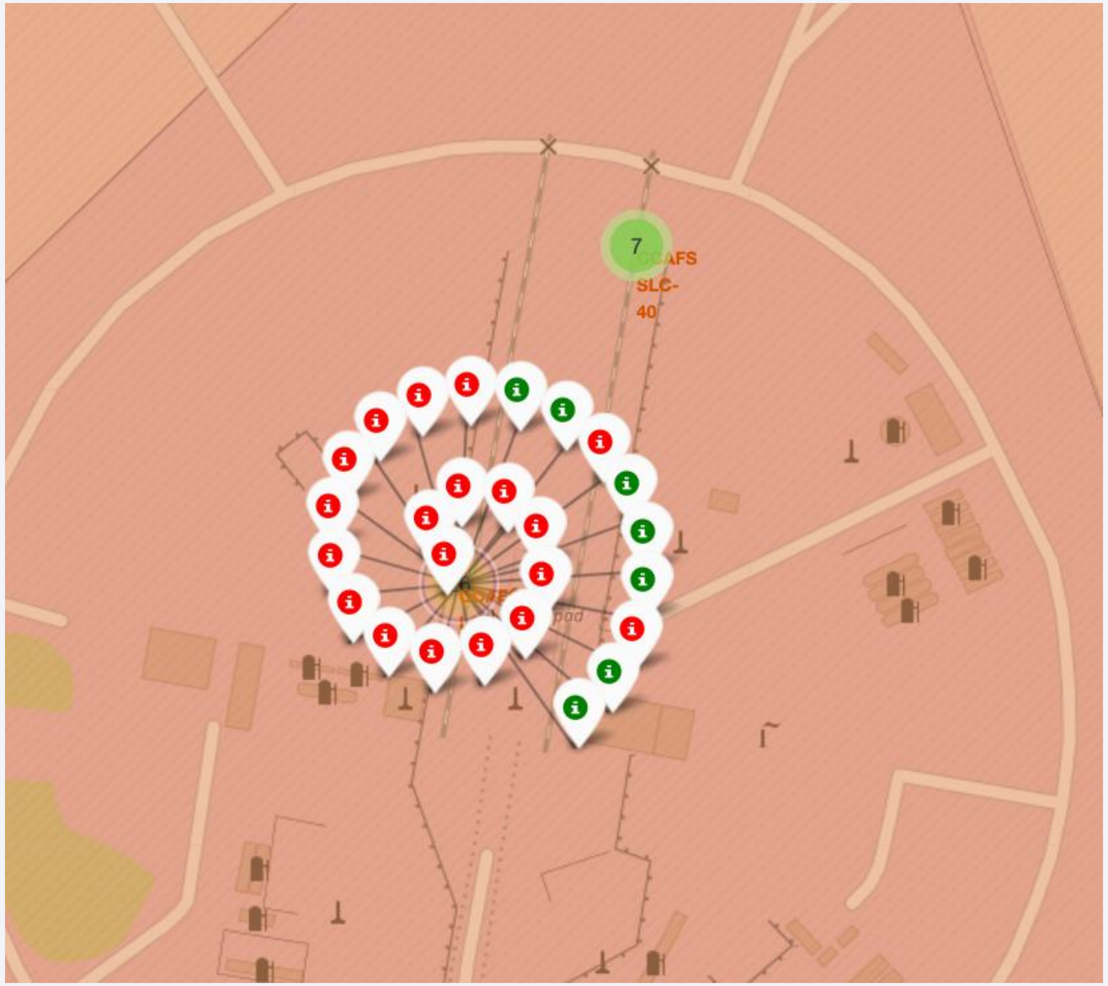
Launch Sites Proximities Analysis

Visualising the Launch Site Locations

- The map visually shows how close the launch sites are to the equator and coastline.
- Proximity to is due to ease of transport.
- Proximity to equator is due to fuel saving when launching rockets into space.

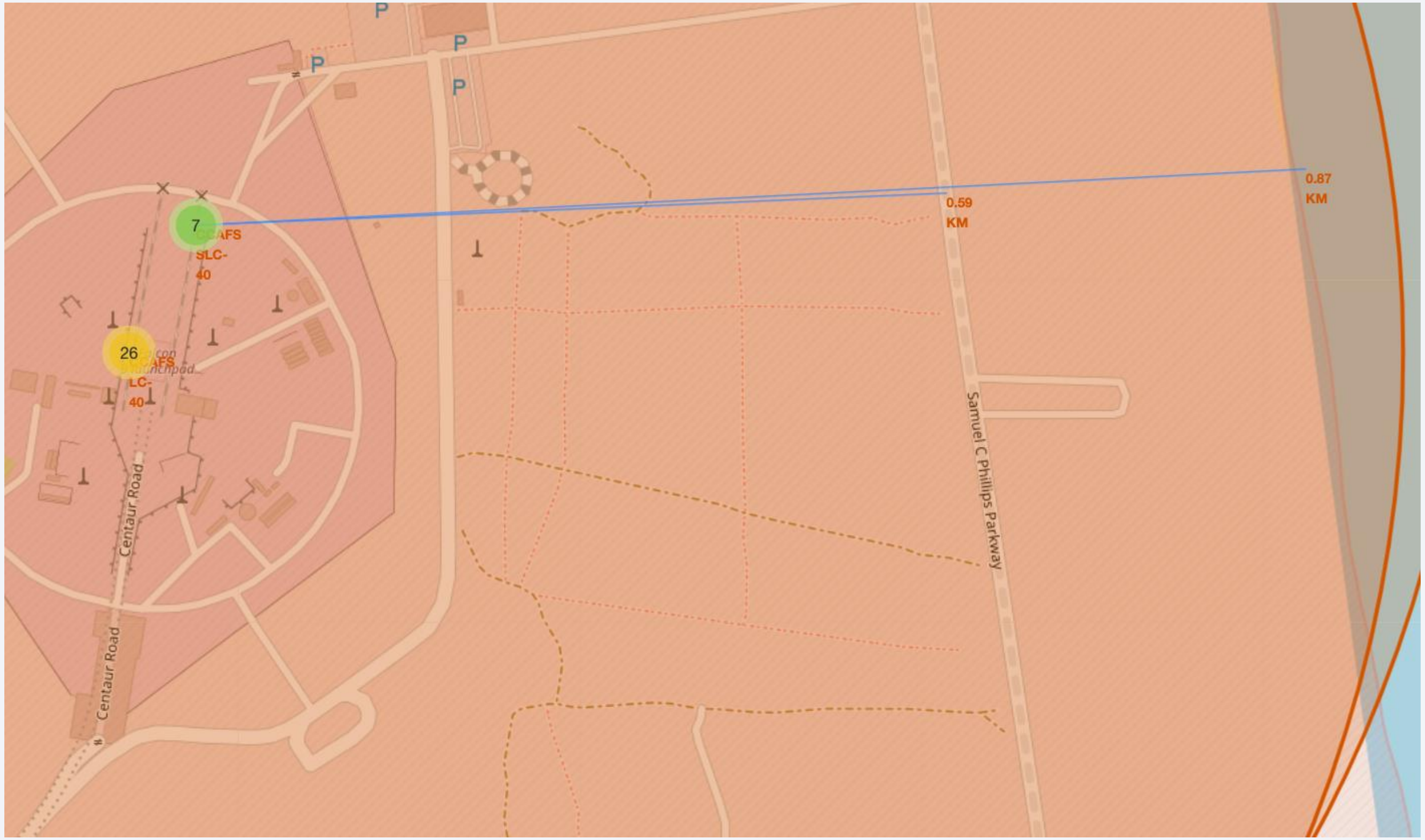
Visualising Launch Success vs. Failure

- At each launch site a separate marker was added for each attempted launch.
- Green for success
- Red for failure



Visualising proximity to Access Routes

- Here we added polylines from a launch site to a nearby road and coastline with calculate distance.
- This shows how close rail, road and coastline are to the launch sites.
- Conversely cities are far away.

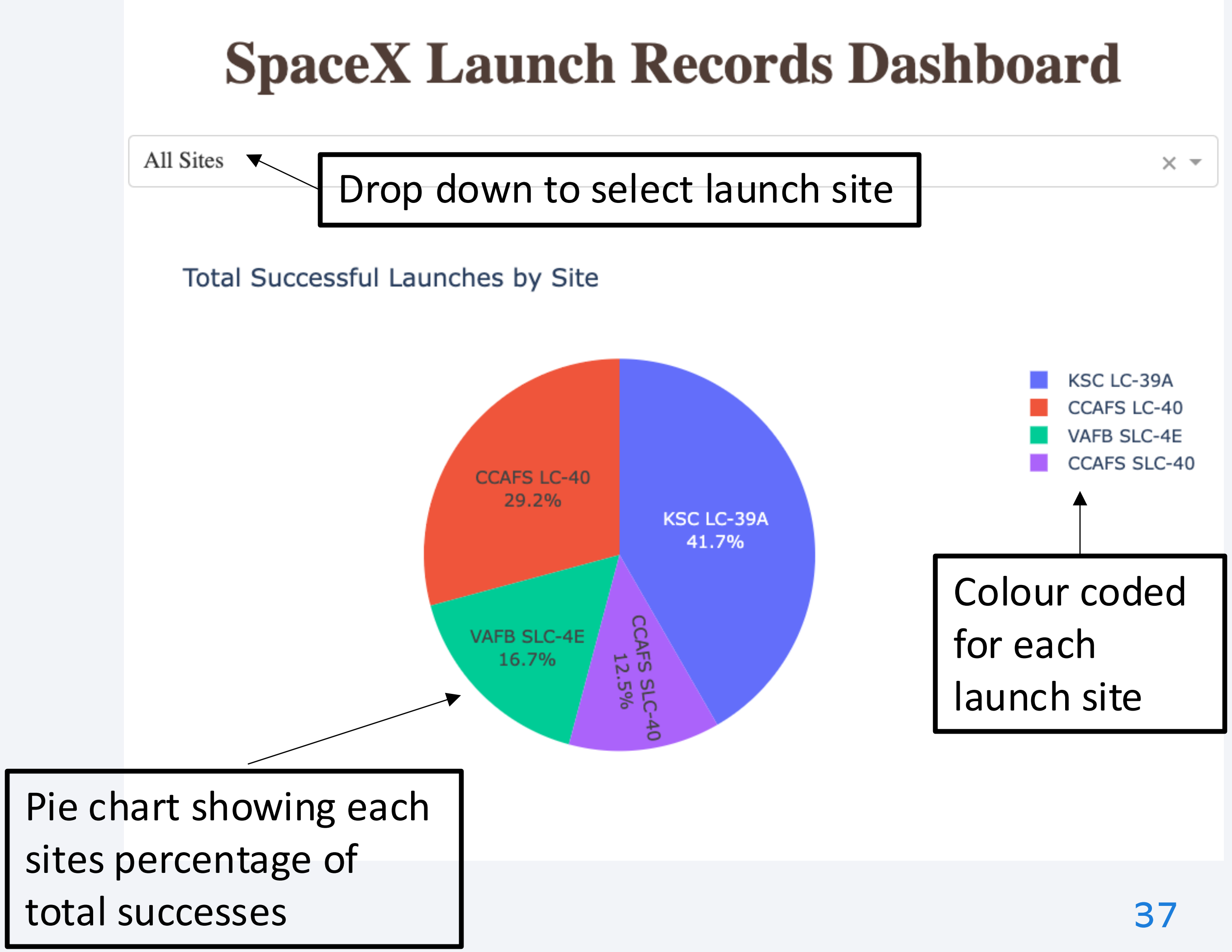


Section 4

Build a Dashboard with Plotly Dash

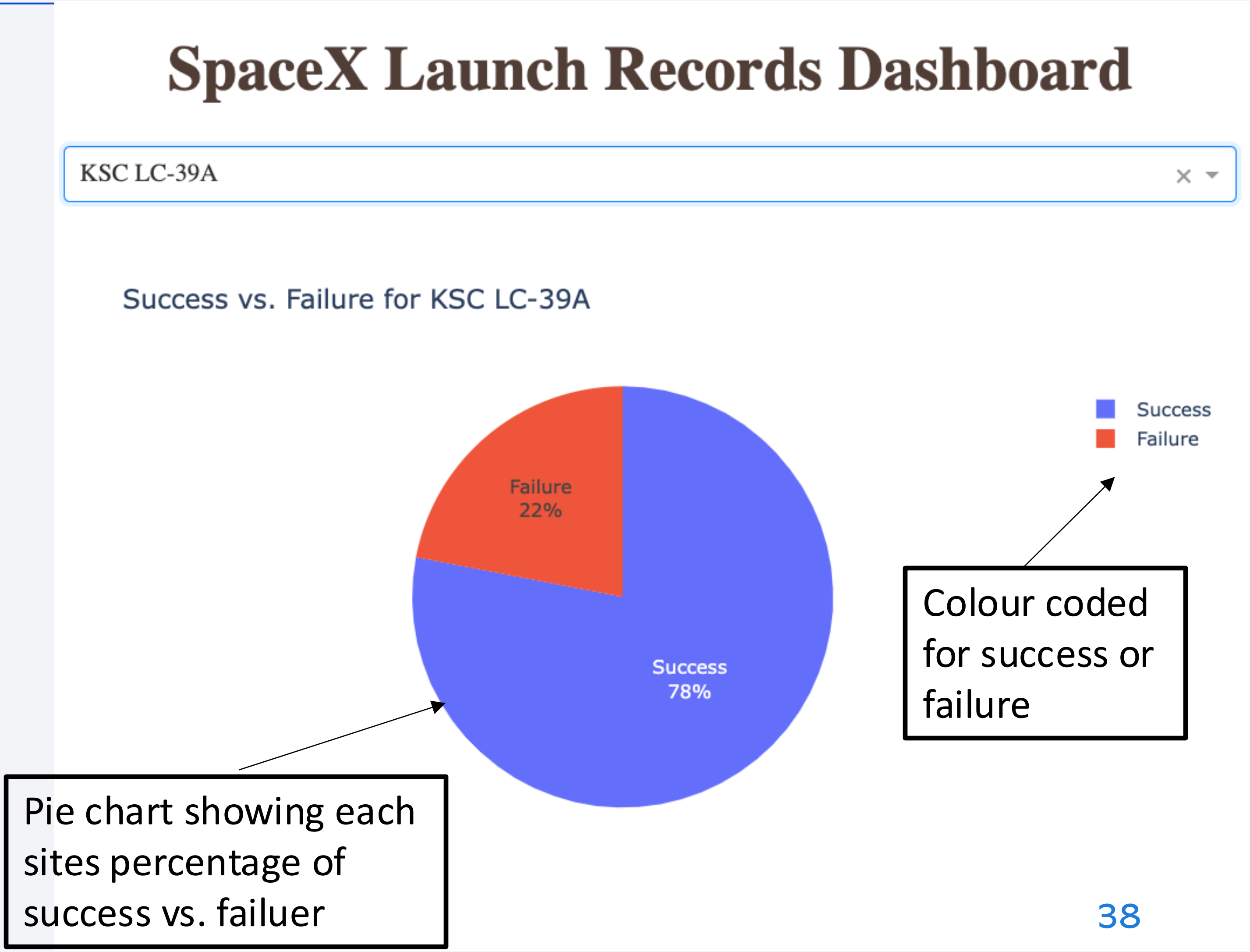
Launch Dashboard : Most Successful Launch Sites

- KSC LC-39A has the most successful launches.
- CCAFS SLC-40 has the least successful launches.
- But is this due to number of flights or flight success ratio?



Launch Dashboard : Highest Launch Success Ratio

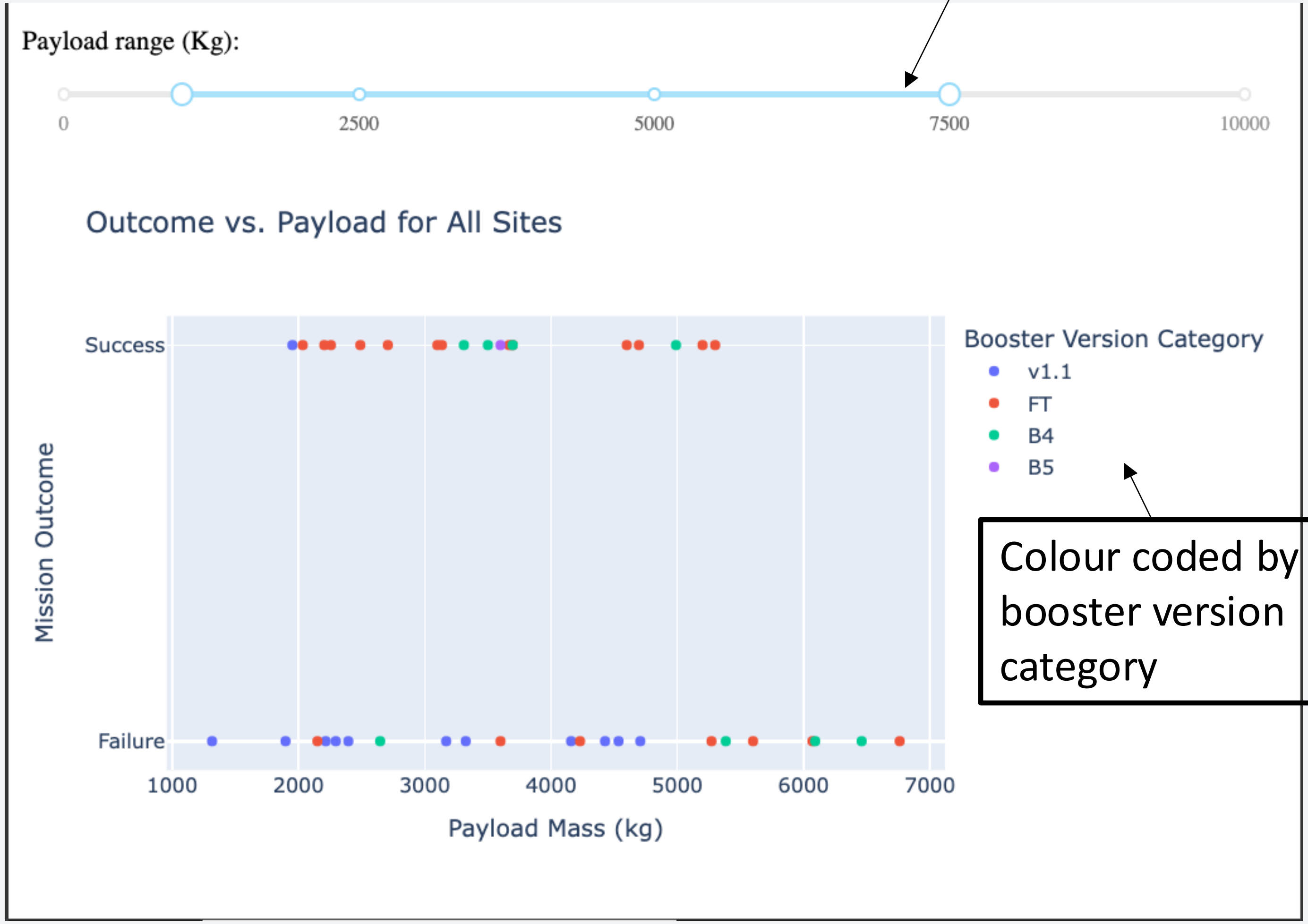
- KSC LC-39A has the highest ratio of success to unsuccessful launches.
- In the previous graph we see it also has the most successful launches. Making it a good candidate / location for a launch site.



Launch Dashboard : Payload vs. Outcome

Range slider allowing for choice of payload.

- Scatter graph of mission outcome vs. payload mass for different booster versions.
- The most successful launches lie around 2000 – 5300 kg.
- Booster Version v1.0 has had 4 failures and no successes.
- Booster Version v1.1 has had many failures and 1 success.
- Booster Version B5 has never failed but only had one attempt



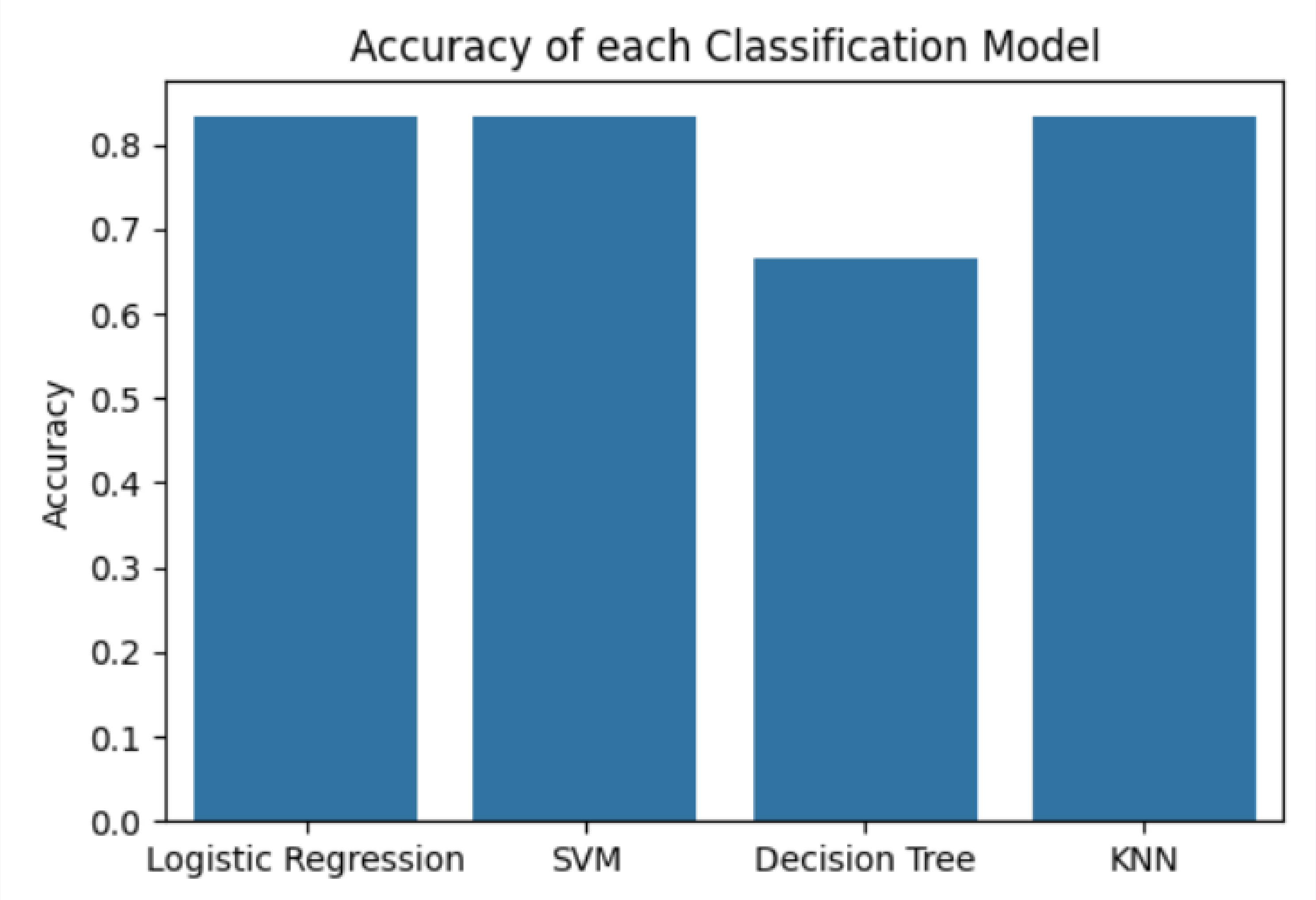
Colour coded by booster version category

Section 5

Predictive Analysis (Classification)

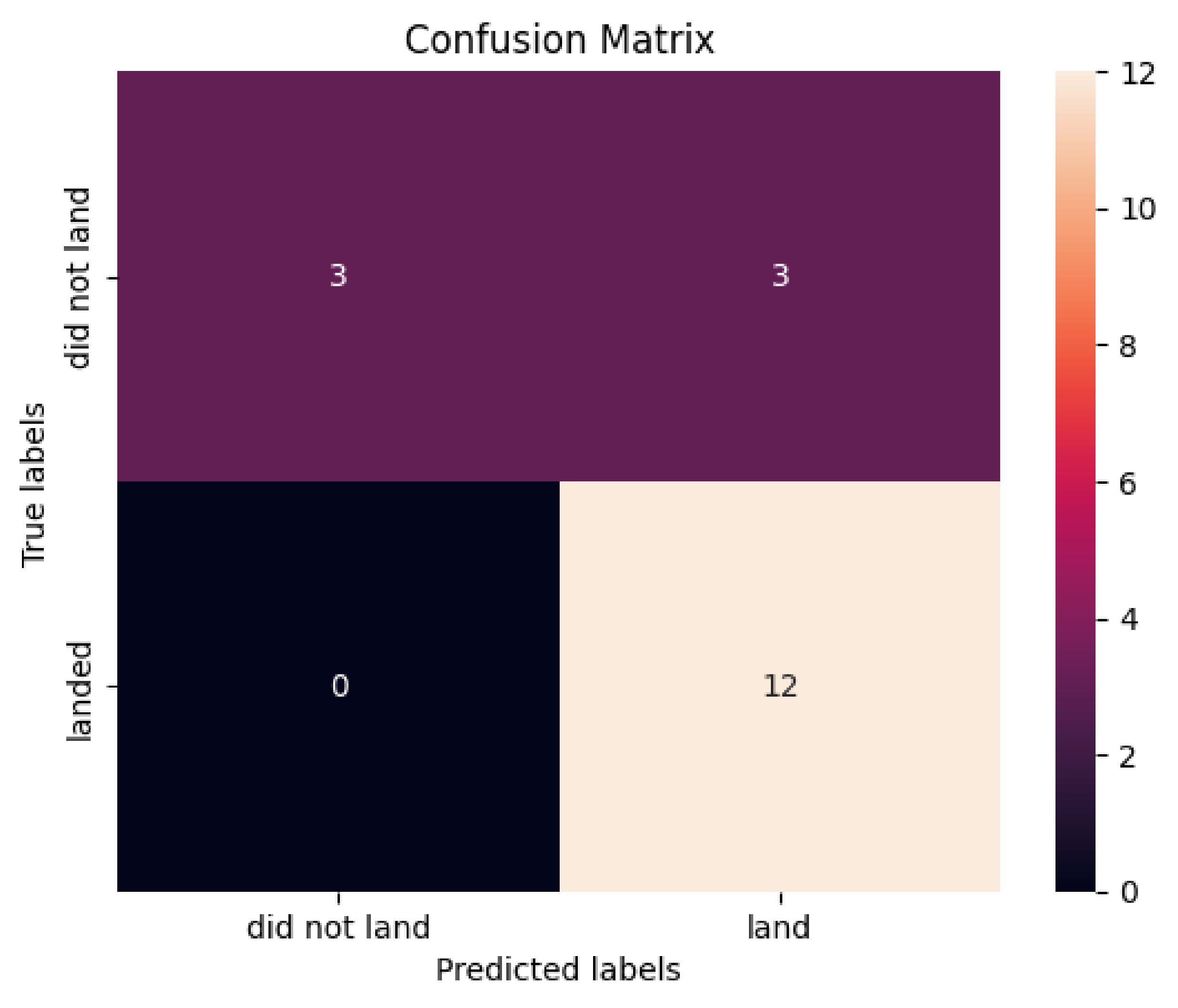
Classification Accuracy

- We find that Decision Tree has the lowest accuracy with 66.67%
- Logistic Regression, SVM and KNN have equal accuracy with 83.3%



Confusion Matrix

- This is the confusion matrix for SVM.
- If the rocket is to successfully land the model will predict it to land successfully 100% of the time.
- If the rocket is to unsuccessfully land, the model will predict it to land unsuccessfully 50% of the time.
- True Positive - 12
- False Positive - 3
- True Negative - 3



Conclusions

- Many factors affect the outcome of a rocket launch. E.g. payload, orbit type, booster version, flight number and launch site. Making it hard as humans to predict what are the important factors that determine success.

- Using a SVM classification model we can predict with 83.3% accuracy whether a launch will be successful or not. Aiding SpaceY in their pricing for a rocket launch.

- Next Steps:
  - Addition of other factors such as weather and location into SVM model to improve accuracy.
  - Perhaps a regression model could be used to understand what parameters need tweaked to improve chances of a successful launch.

Thank you!