IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Alexander
Dec 2021

Outline

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

Executive Summary

- Summary of methodologies
- Summary of all results

Introduction

- This presentation is made to complete the IBM Data Science Capstone
- We use different methods that we have learned in Data Science to a data-set of rocket launches

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Describe how data was collected
- Perform data wrangling
  - Describe how data was processed
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection

Data of rocket launches are collected from two sources:

- Requests to the SpaceX API
- Wikipedia through webscraping

Data Collection – SpaceX API

https://github.com/a-schier/IBM-DS-Capstone/blob/59c7f4df57531547a00c4ba59588563f19239bb1/jupyter-labs-spacex-data-collection-api.ipynb

Request to SpaceX API
V
Response
V
Decode content as JSON to Pandas DataFrame
V
Data Wrangling

Data Collection – Scraping and Wrangling

https://github.com/a-schier/IBM-DS-Capstone/blob/59c7f4df57531547a00c4ba59588563f19239bb1/jupyter-labs-webscraping.ipynb

Request content from Wikipedia w/ a URL

V

Find HTML tables in content

V

Decode table to Pandas DataFrame

V

Data wrangling

GitHub Links

https://github.com/a-schier/IBM-DS-Capstone/blob/1c54e12fb3b32e8b5cdb2d536ba8c8482891e55c/jupyter-labs-eda-dataviz.ipynb

https://github.com/a-schier/IBM-DS-Capstone/blob/59c7f4df57531547a00c4ba59588563f19239bb1/jupyter-labs-eda-sql-coursera.ipynb

https://github.com/a-schier/IBM-DS-Capstone/blob/59c7f4df57531547a00c4ba59588563f19239bb1/lab_jupyter_launch_site_location.ipynb

https://github.com/a-schier/IBM-DS-Capstone/blob/59c7f4df57531547a00c4ba59588563f19239bb1/spacex_dash_app.py

https://github.com/a-schier/IBM-DS-Capstone/blob/59c7f4df57531547a00c4ba59588563f19239bb1/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

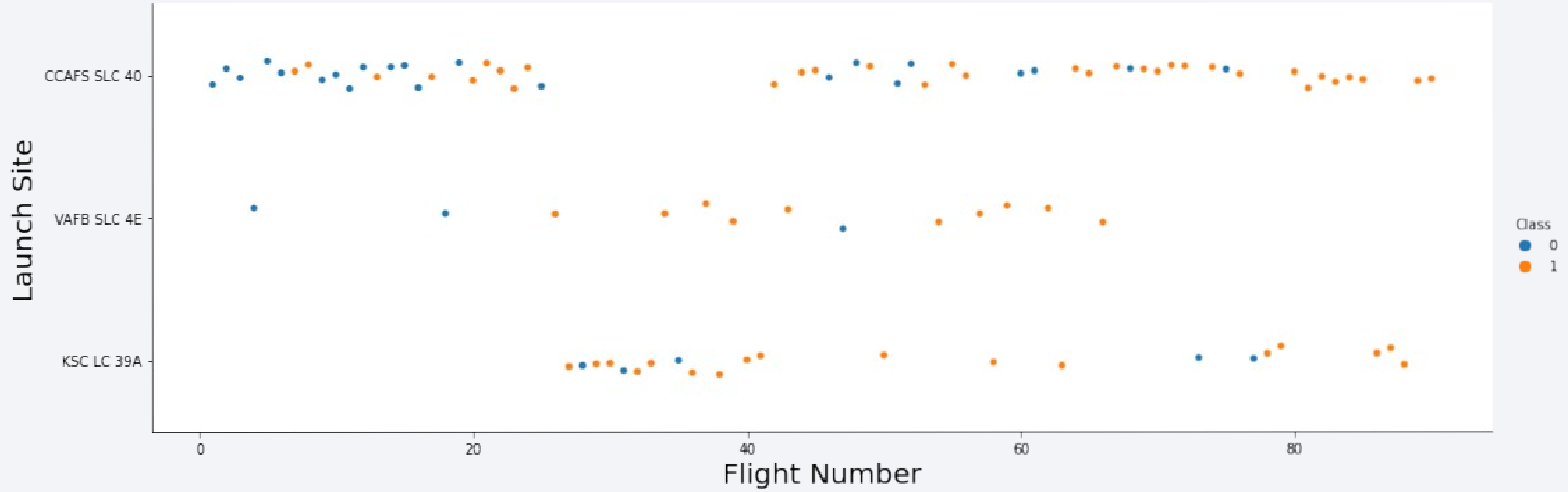
Results

- Exploratory data analysis results
  - The success rate for number of successful landings have increased markedly in recent years
- Interactive analytics demo in screenshots
- Predictive analysis results
  - The success of a mission is predicted using machine learning methods

Section 2

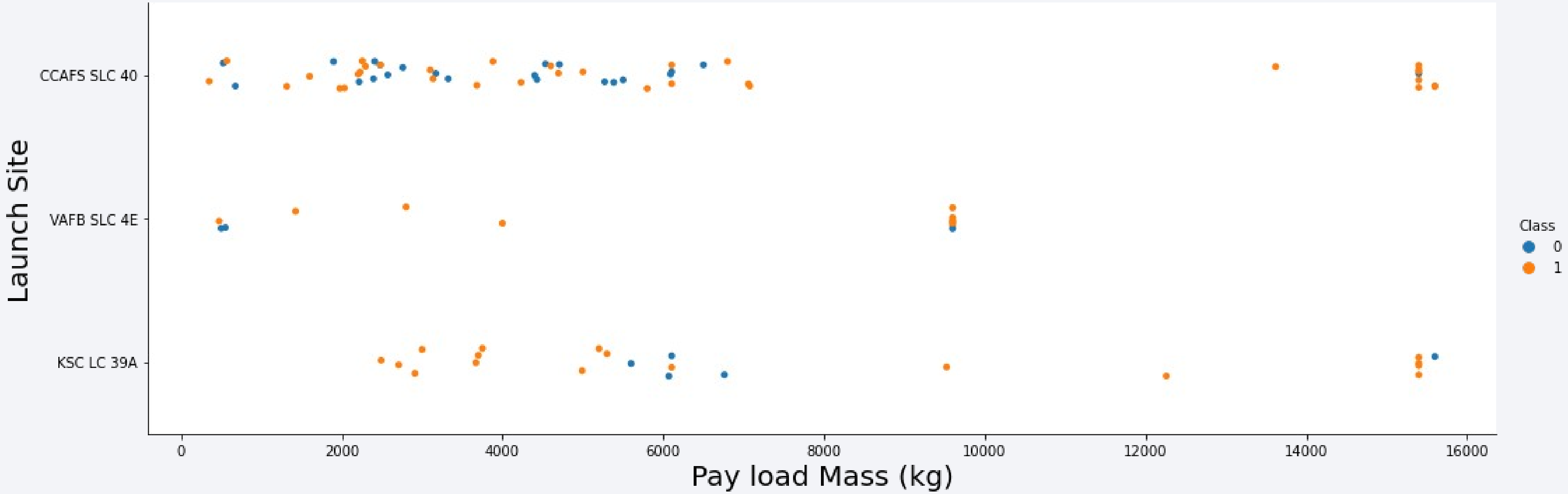
Insights drawn from EDA

Flight Number vs. Launch Site



- The above graph shows three takeaways:
  - The success rate increases with flight number overall
  - The success rate is the higher for KSC LC 39A and VAFB SLC 4E than CCAFS SLC 40
  - Most launches happen from CCAFS SLC 40 with a break in the middle

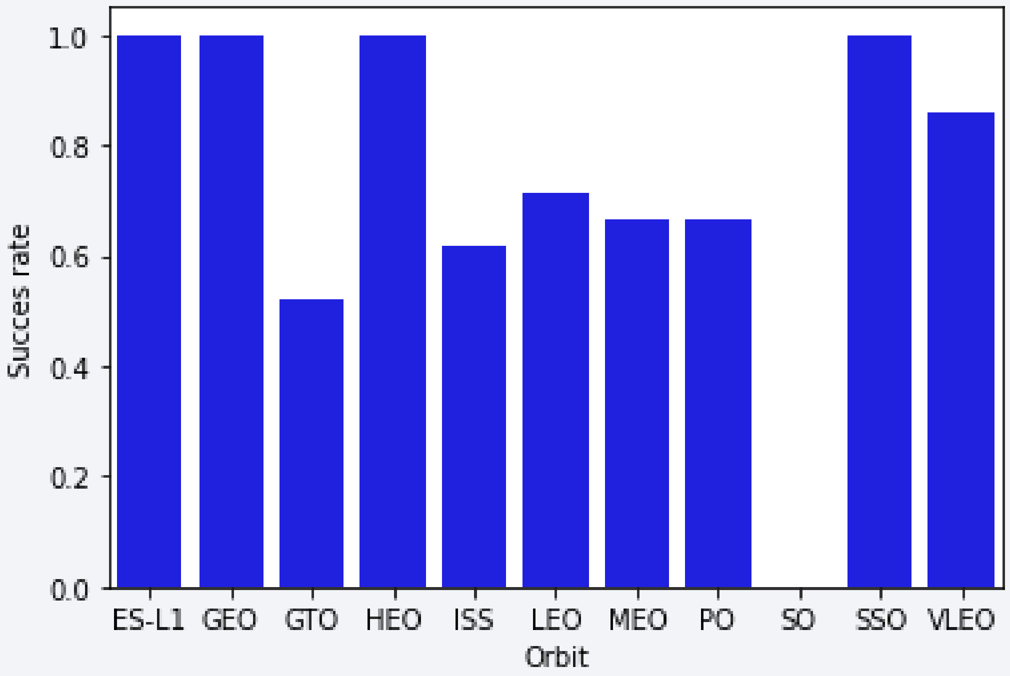
Payload vs. Launch Site



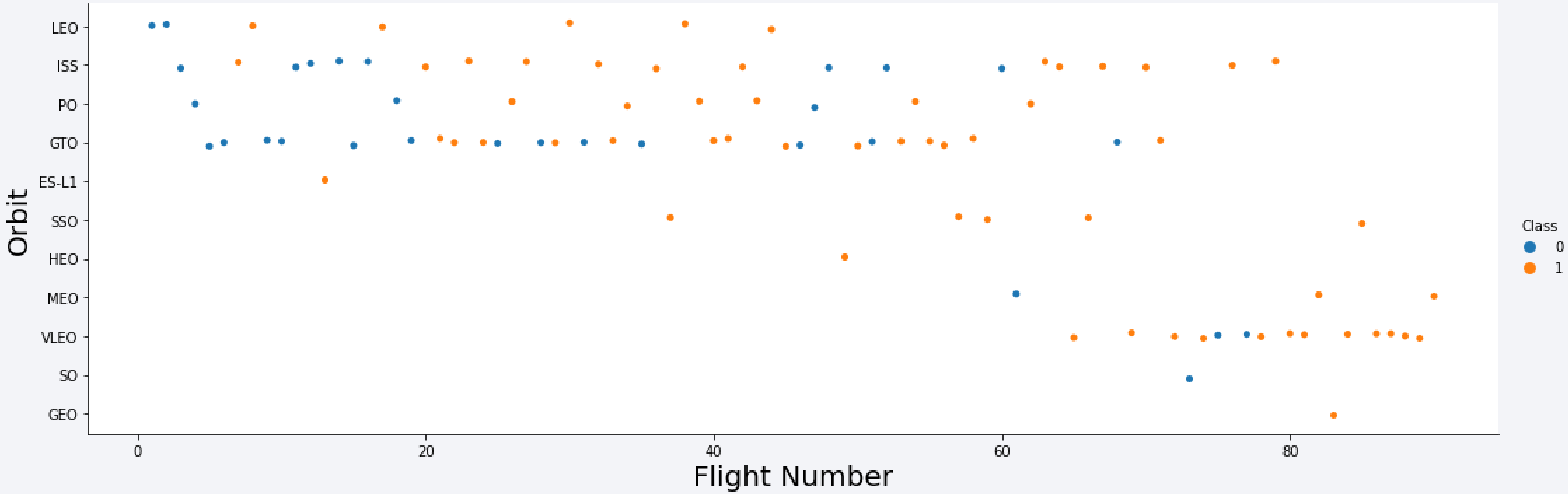
- The payload varies for each flight and the maximum payload is different for each launch site (VAFB SLC 4E w/ ~950 kg vs CCAFS SLC 40 and KSC LC 39A w/ ~1550 kg)

Success Rate vs. Orbit Type

- Some orbit types have higher success rates than others:
  - ES-L1, GEO, HEO and SSO all have a 100% success rate

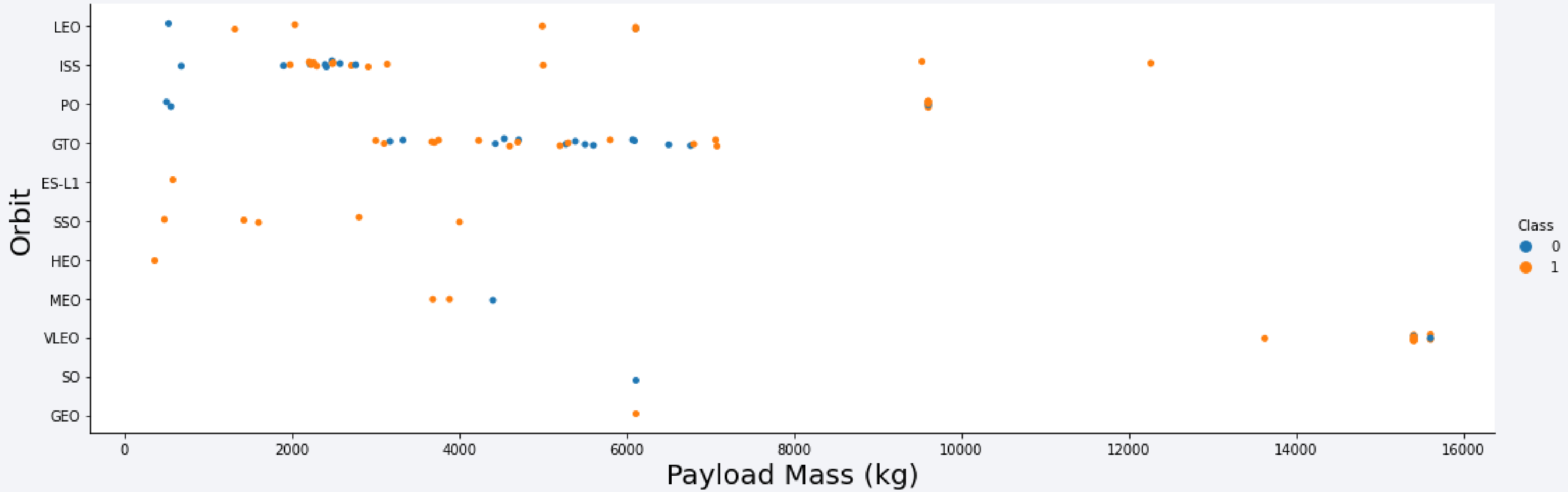


Flight Number vs. Orbit Type



- Over time the type of mission has be changed from LEO/ISS/PO/GTO orbits to HEO/MEO/VLEO orbits
  - There is a positive correlation between success rate of the LEO orbit and flight number, but no such relationship exists for GTO – i.e. LEO launches get better over time but GTO launches do not

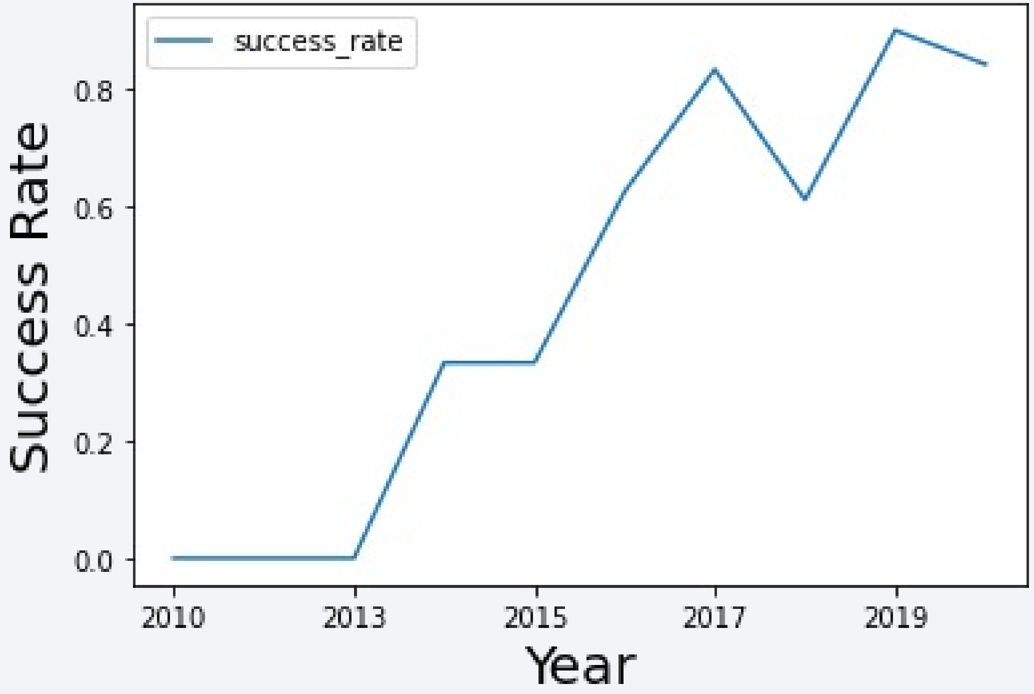
Payload vs. Orbit Type



- With heavy payloads the successful landing rates are more for Polar, LEO and ISS.
- However for GTO we cannot distinguish this as both succesful- and unsuccessful missions are both there here.

Launch Success Yearly Trend

- It is seen that there exists an overall trend that the success rate of the launches increase year by year



All Launch Site Names

Find the names of the unique launch sites

The SQL query to find these are:

SELECT DISTINCT(LAUNCH_SITE) FROM SPACEXTBL;

See output of the query on the right

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

Launch Site Names Begin with 'CCA'

Find 5 records where launch sites begin with `CCA`

The SQL query to find these are:

SELECT * FROM SPACEXTBL WHERE LAUNCH_SITE LIKE 'CCA%'
LIMIT 5;

See output of the query on the next slide

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	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	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 SQL query to find these are:

SELECT SUM(PAYLOAD_MASS__KG_) AS TOTAL_PAYLOAD_MASS FROM SPACEXTBL WHERE CUSTOMER = 'NASA (CRS)';

See output of the query on the right

total_payload_mass
45596

Average Payload Mass by F9 v1.1

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

The SQL query to find these are:

SELECT AVG(PAYLOAD_MASS__KG_) AS AVG_PAYLOAD_MASS FROM SPACEXTBL WHERE BOOSTER_VERSION LIKE '%F9 v1.1%';

See output of the query on the right

avg_payload_mass
2534

First Successful Ground Landing Date

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

The SQL query to find these are:

SELECT MIN(DATE) as FIRST_SUCESSFUL_DATE FROM SPACEXTBL
WHERE LANDING__OUTCOME = 'Success (ground pad)';

See output of the query on the right

first_sucessful_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

The SQL query to find these are:

SELECT BOOSTER_VERSION FROM SPACEXTBL 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

See output of the query on the right

Total Number of Successful and Failure Mission Outcomes

Calculate the total number of successful and failure mission outcomes

The SQL query to find these are:

SELECT MISSION_OUTCOME, COUNT(*) FROM SPACEXTBL GROUP BY MISSION_OUTCOME;

See output of the query on the right

mission_outcome	2
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Boosters Carried Maximum Payload

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

The SQL query to find these are:

SELECT BOOSTER_VERSION, PAYLOAD_MASS__KG_ FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL);

See output of the query on the right

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

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

The SQL query to find these are:

SELECT DATE, LANDING__OUTCOME, BOOSTER_VERSION, LAUNCH_SITE FROM SPACEXTBL WHERE LANDING__OUTCOME = 'Failure (drone ship)' AND YEAR(DATE) = 2015;

See output of the query on the right

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

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 SQL query to find these are:

SELECT LANDING__OUTCOME, COUNT(*) AS LANDINGS FROM SPACEXTBL WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY LANDING__OUTCOME ORDER BY LANDINGS DESC;

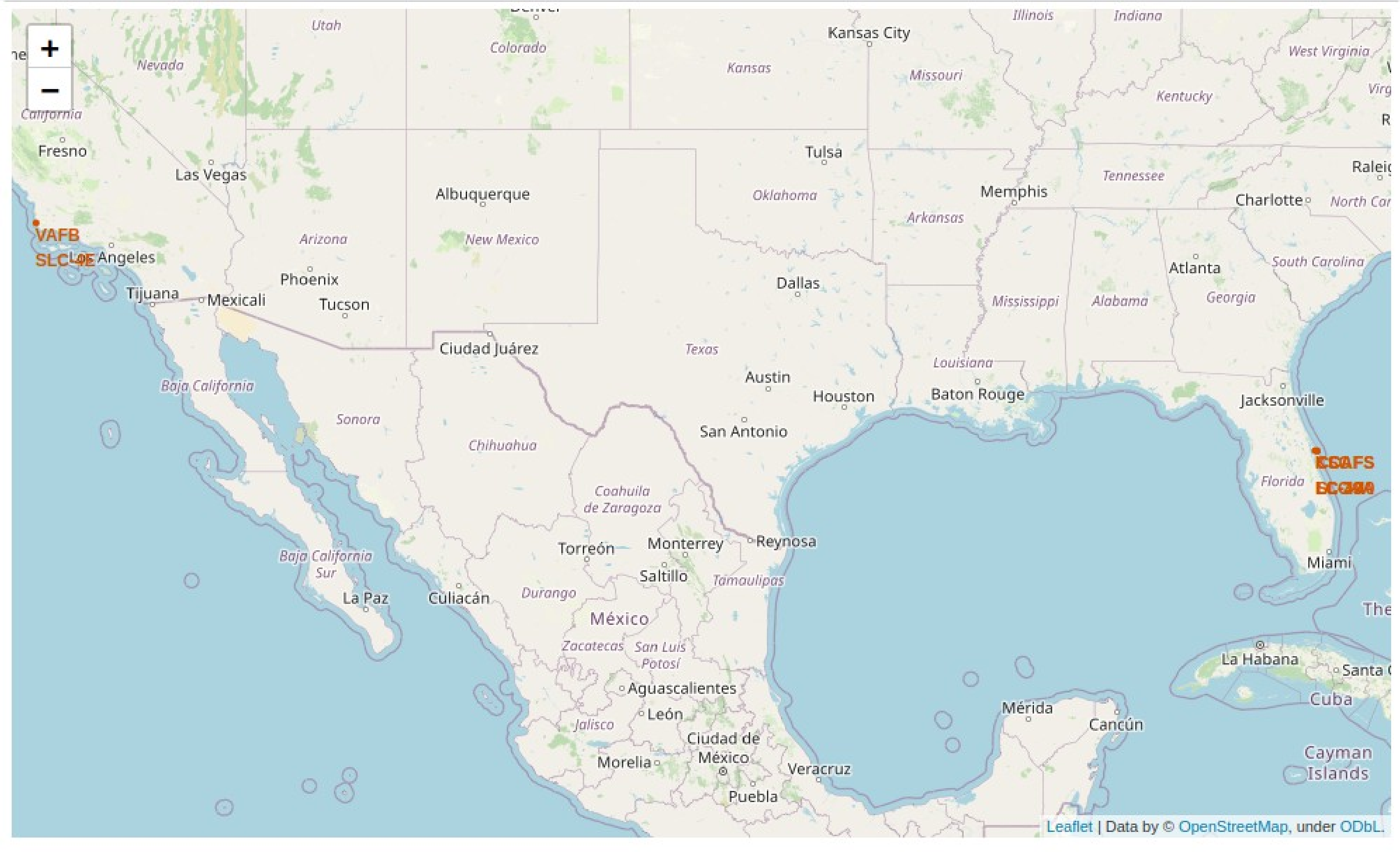
See output of the query on the right

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

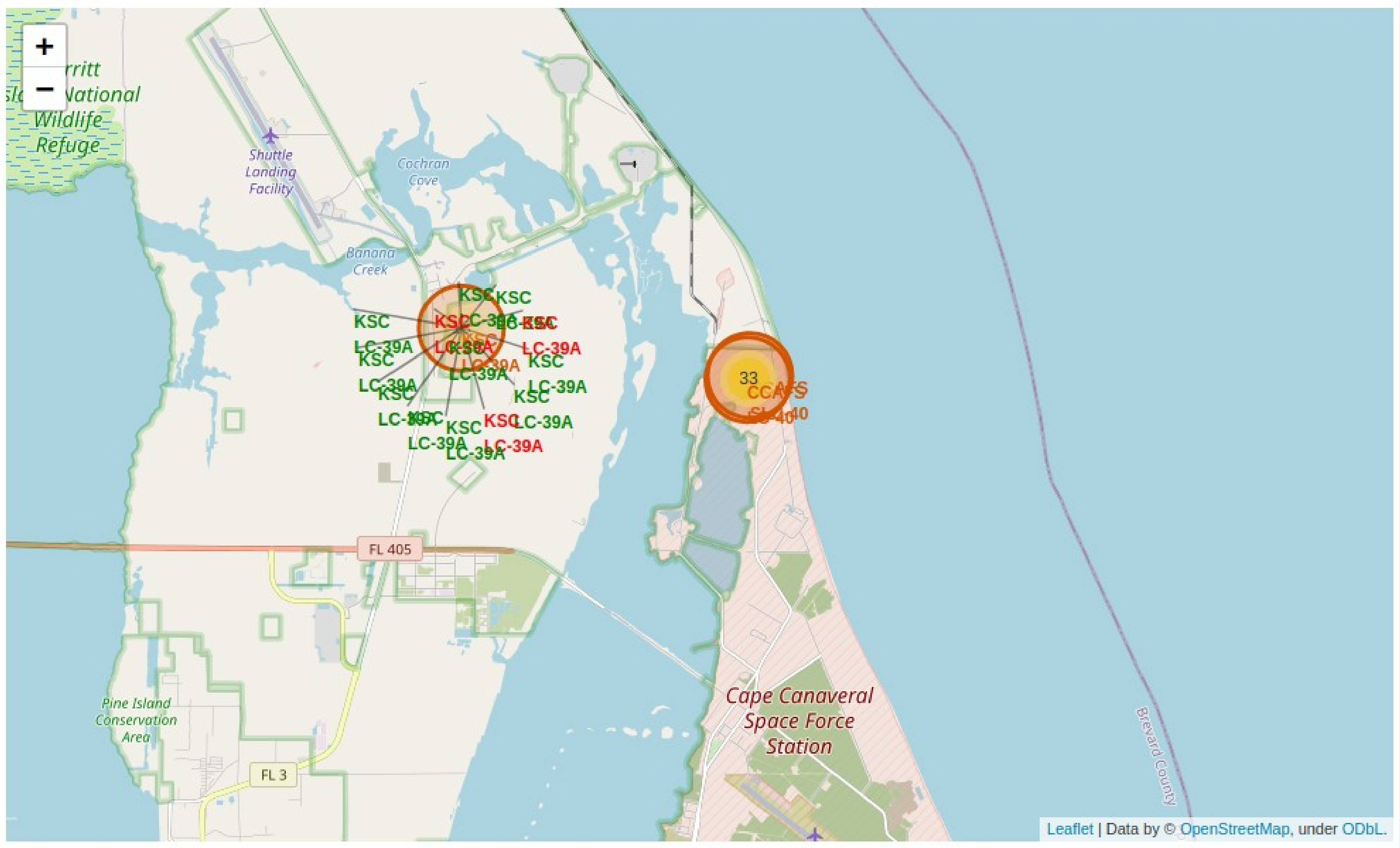
Section 4

Launch Sites Proximities Analysis

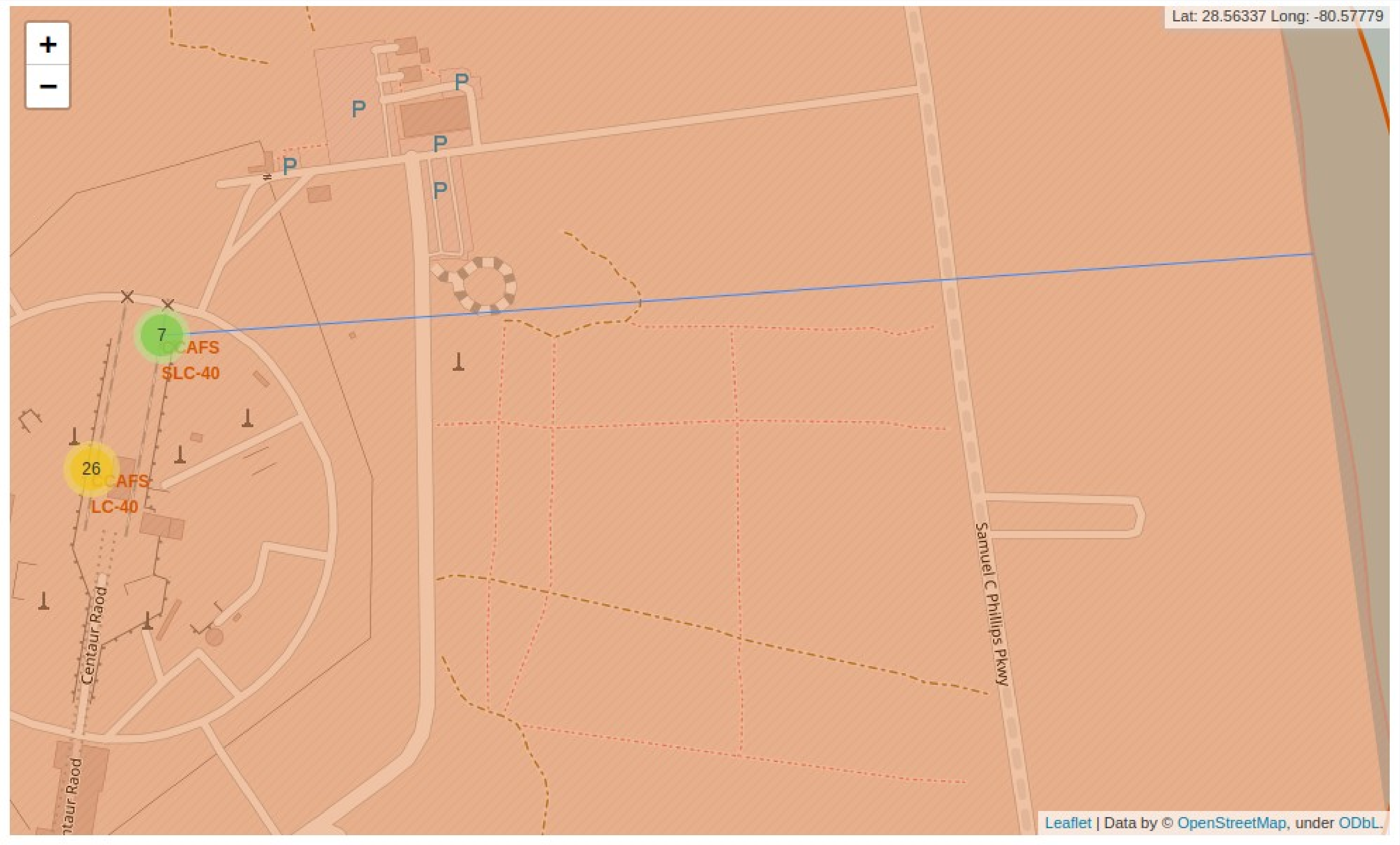
Launch Site Locations



Map with Launch Outcomes at Each Site (example)



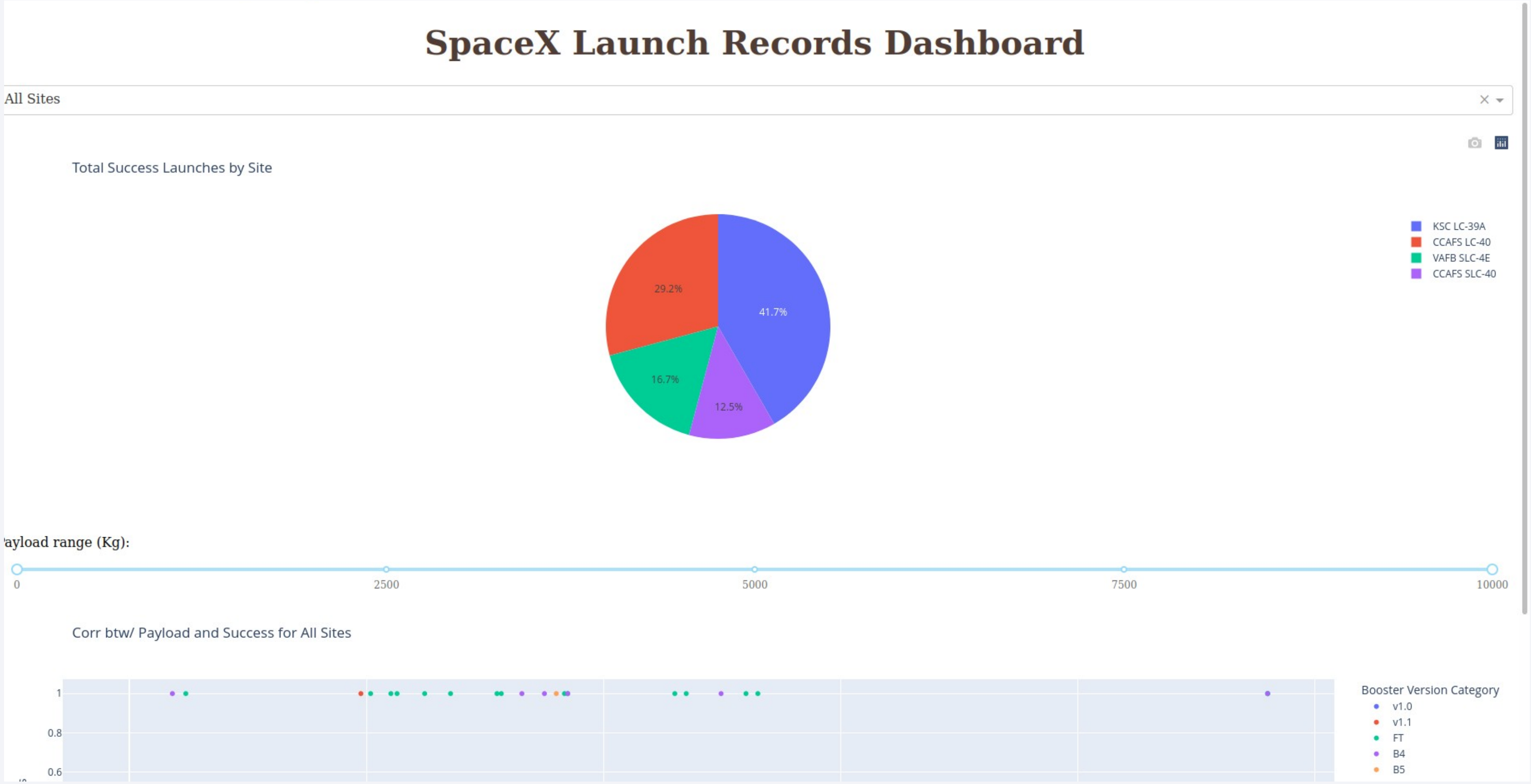
Proximity to Coastline



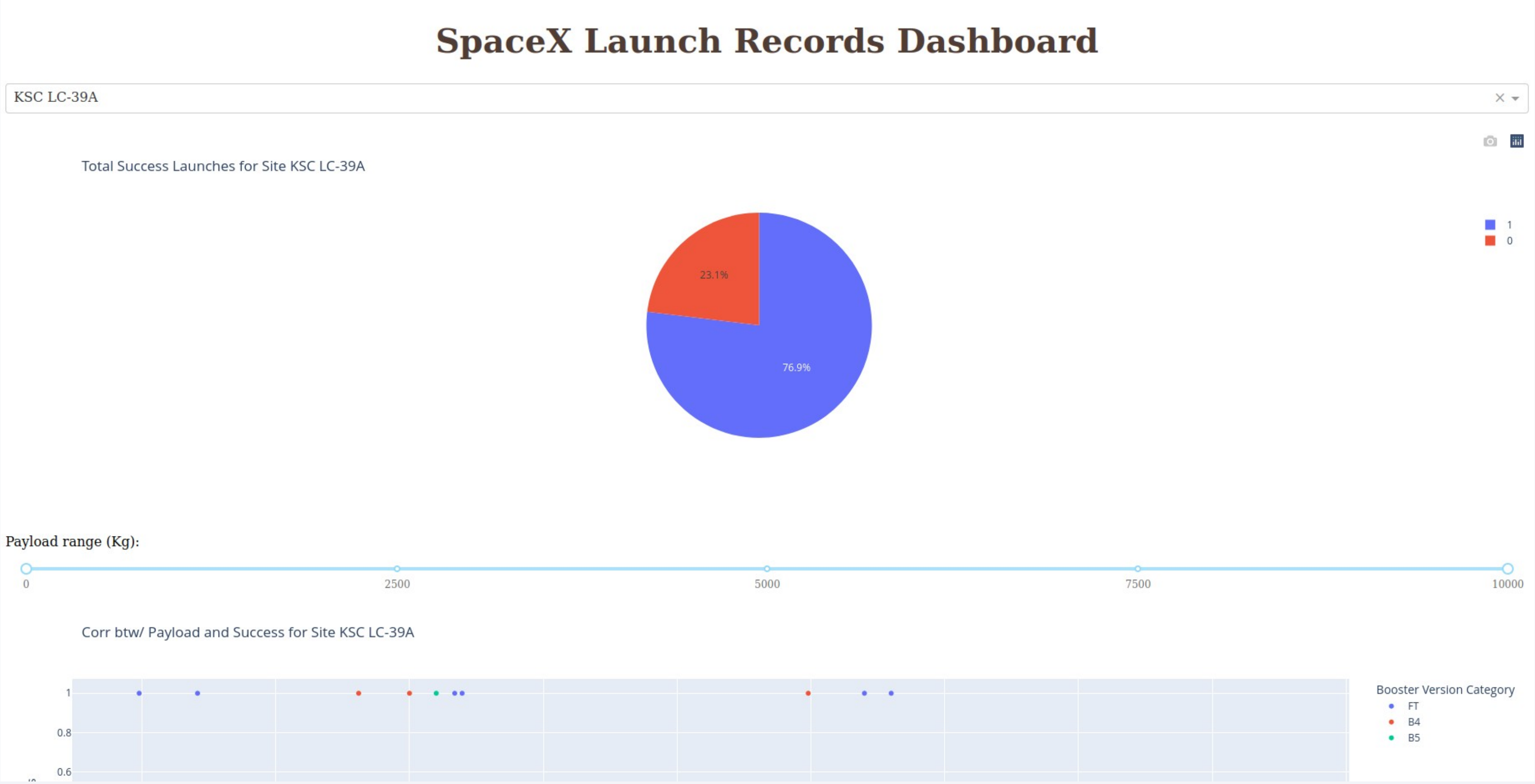
Section 5

Build a Dashboard with Plotly Dash

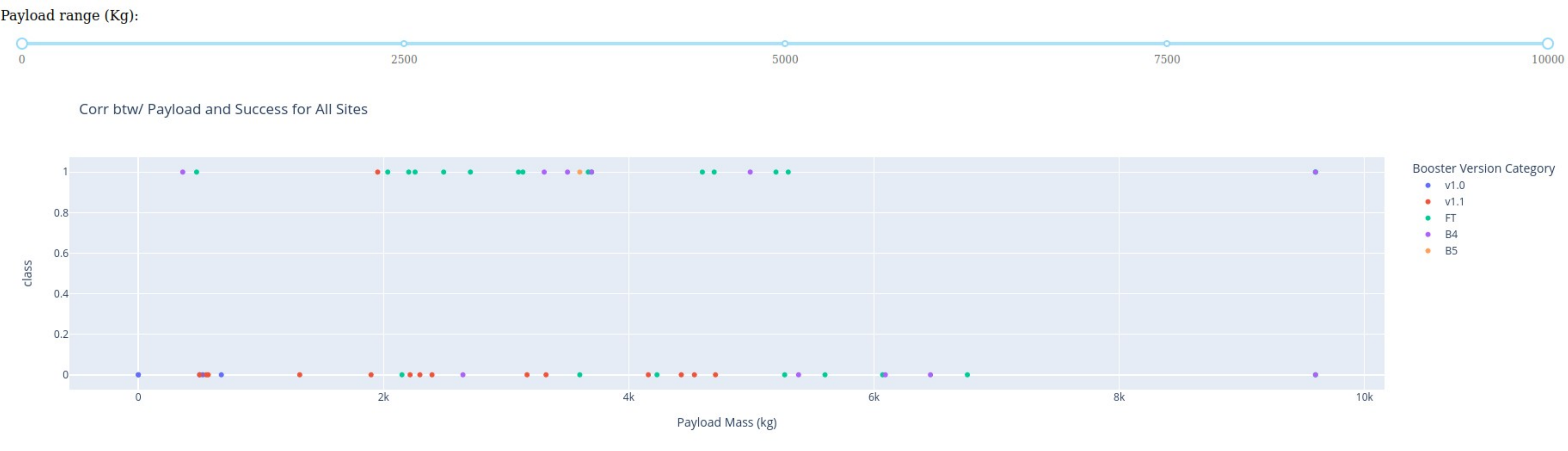
Launch success for all sites



Launch Site with Highest Success Ratio



Payload vs Launch Outcome for All Sites



Section 6

Predictive Analysis (Classification)

Classification Accuracy

The most acurate nodel is the KNN

Confusion Matrix

Thank you!