IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Rui Azevedo
2022-05-18

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion

Executive Summary

- We set out to determine what factors will predict whether SpaceX will land the first stage of the Falcon 9 rocket, and to train a model that could predict successful landings
- We used Python to obtain and visualize launch data from the SpaceX API and through webscraping, and to train machine learning (ML) models
- We found that the landing success rate increased over time (consistent with introductions of improved rocket versions), and was dependent on payload mass, and target orbit
- We trained four ML models (KNN, logistic regression, decision tree, SVM), obtaining a best accuracy of 83%

Introduction

- Most rocket manufacturers have, or are working on, reusable rockets
- Companies that have this option, like SpaceX, can offer much better prices than the competition
- The demand for launch vehicles is still extremely high!
- Landing a rocket booster is hard and can only be done under certain conditions
- Can we determine what these conditions are?
- Can we predict whether a company will land a booster, under some launch conditions?

Section 1

Methodology

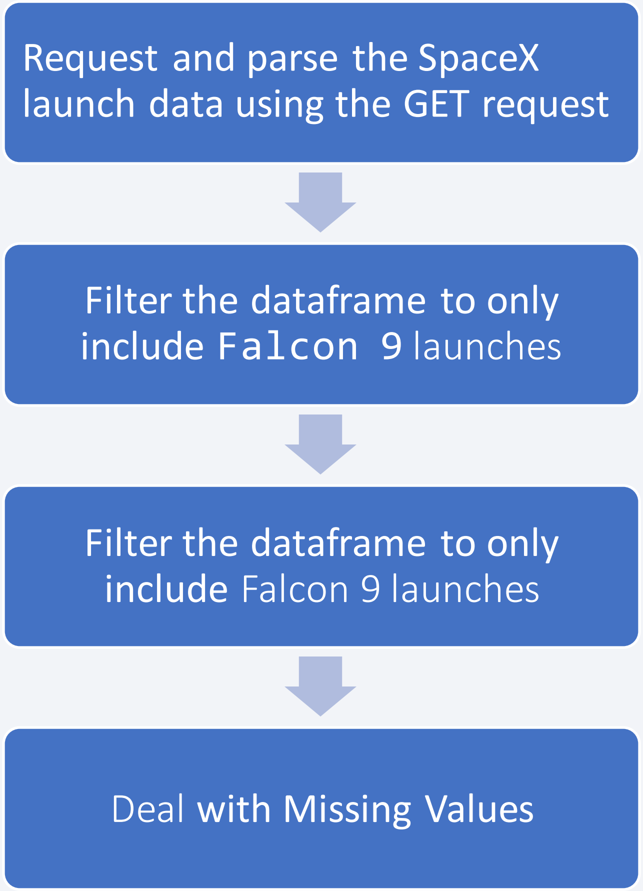
Methodology

Executive Summary

- Data was collected via the SpaceX API and through webscraping, using Python
- Performed data wrangling using Pandas and Numpy, preparing data for visualization and model training
- Perform exploratory data analysis (EDA) using the Python plotting packages matplotlib and seaborn, and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models in SciKit Learn (KNN, logistic regression, decision tree, SVM)

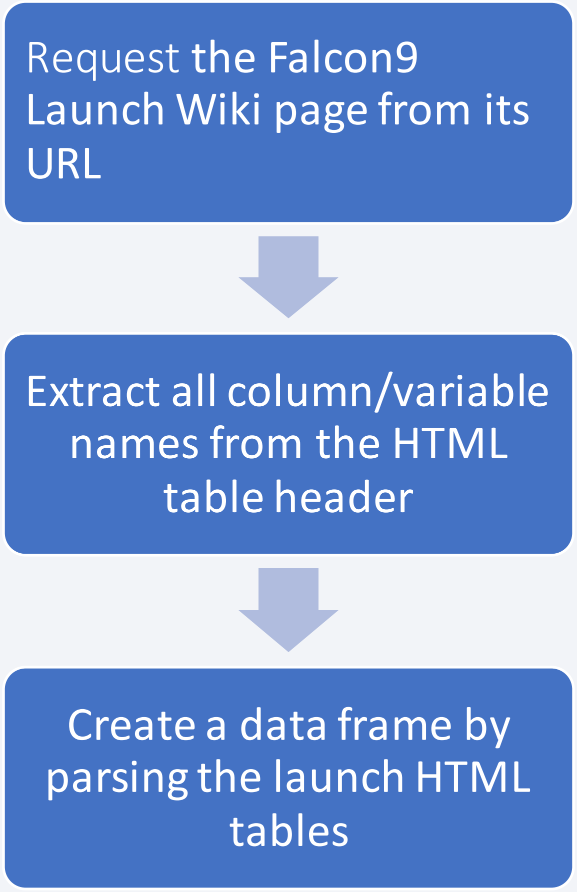
Data Collection – SpaceX API

- The Python "Requests" package simplifies making calling API's
- Pandas was used to clean up the data, removing unwanted values and features
- API Calls Notebook



Data Collection - Scraping

- Requests package Was used to get the HTML source from Wikipedia
- Beautiful Soup package was used to parse the HTML
- Pandas and custom functions ere used to convert the relevant columns into a dataframe



- Webscrapping Notebook

Data Wrangling

- We obtained the number of launches per orbit type, of which the most common were
  - GTO 27
  - ISS 21
  - VLEO 14
- And the number of launches from each launch site:
  - CCAFS SLC 40 55
  - KSC LC 39A 22
  - VAFB SLC 4E 13
- The several landing outcomes were into two binary categories:
  - Successes: 60
  - Failures: 30
- Data Wrangling Notebook

Request the Falcon9 Launch Wiki page from its URL

↓

Extract all column/variable names from the HTML table header

↓

Create a data frame by parsing the launch HTML tables

EDA with Data Visualization

- We used the Seaborn package to plot the outcome of the landing attempts against some of the features of the data set
- We also tried to determine correlations between the features themselves, and how they affect landing outcome
- The feature with the greatest direct impact experience! SpaceX was able to learn from failure and the success rate steadily improve since the first success in 2014

- EDA with Data Visualization notebook

EDA with SQL

We performed the following queries in SQL, using the package "sqlalchemy" on a IBM DB2 database:

- Displayed the names of the unique launch sites
- Displayed 5 records where launch sites begin with the string 'CCA'
- Displayed the total payload mass carried by boosters launched by NASA (CRS)
- Display average payload mass carried by booster version F9 v1.1
- Listed the date when the first successful landing outcome in ground pad was achieved.
- Listed the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- Listed the total number of successful and failure mission outcomes¶
- Listed the names of the booster_versions which have carried the maximum payload mass. Use a subquery
- Listed the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015
- Ranked 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
- EDA with SQL notebook

Build an Interactive Map with Folium

- We added the launch sites to an interactive map, as well, as clusters or markers for the launches from each one
- It is clear that launch sites are located near the coast and away from population centres, where the expended stages of rockets or debris will not fall on any populated areas.
- There are, however, highways and railways nearby, but these are relatively easy to clear for launches

- Interactive map with Folium

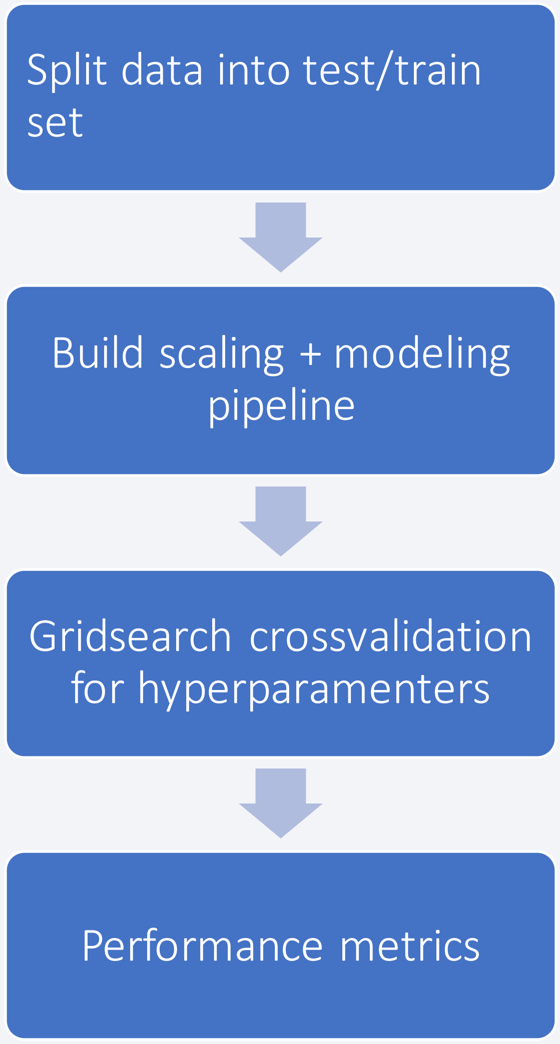
Build a Dashboard with Plotly Dash

- We created interactive chart, in order to visualize the relations each launch site, payload mass, and the number of successful landings
- A pie chart that displays the number of successful landings from each site, or the ratio between successes and failures in each site
- An scatter plot that (for each or all sites) displays the payload mass in a selectable range and the corresponding booster version, and if the landing was successful or not.

- Plotly Dash lab

Predictive Analysis (Classification)

- We used SciKitLearn to perform classification analysis using Decision Trees, KNN, Logistic Regression and SVM
- In order to streamline the process and ensure that the data is scaled correctly, we build a pipeline that included both scaling and model training
- The grid search for hyperparameters and model training was then performed on the pipeline
    - This ensures, for example, that different scalings in train and test sets do not distort our performance metrics
- Completed predictive analysis lab

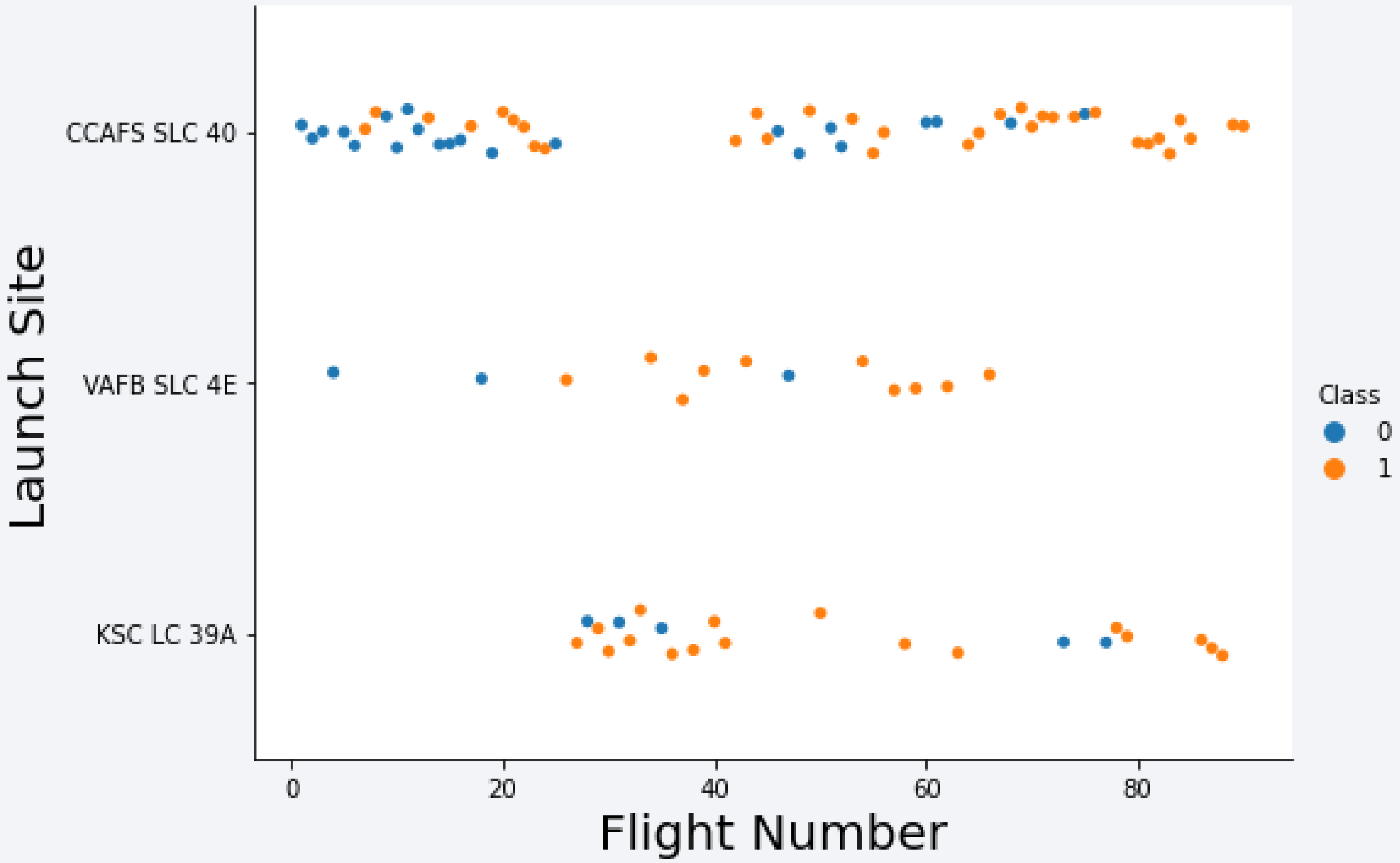


Section 2

Insights drawn from EDA

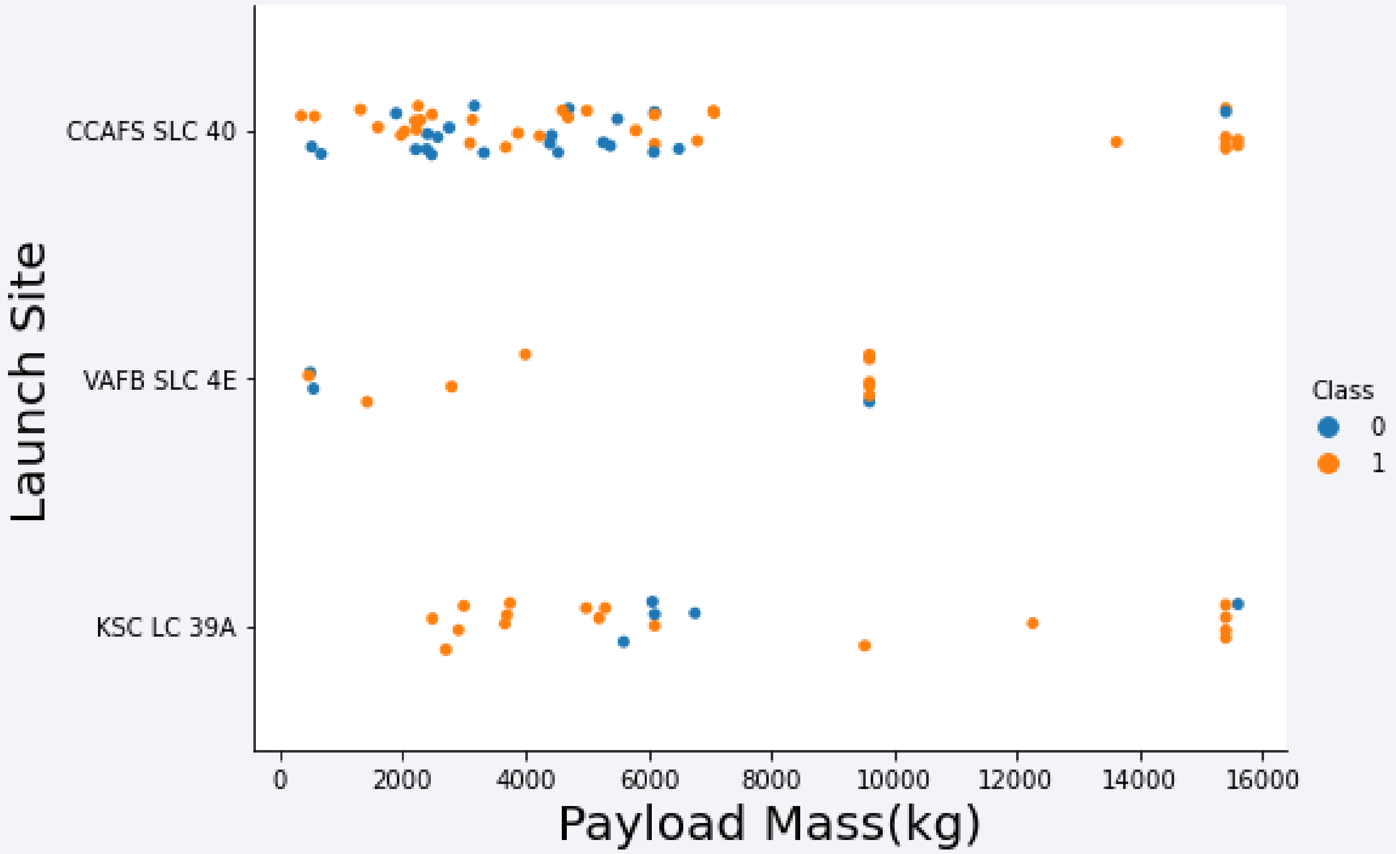
Flight Number vs. Launch Site

- SLC 40 at Cape Canaveral was the initial launch site for SpaceX, and still remains the favoured on
- It is natural that it has a lower average success rate, since it encompasses most of the early launches



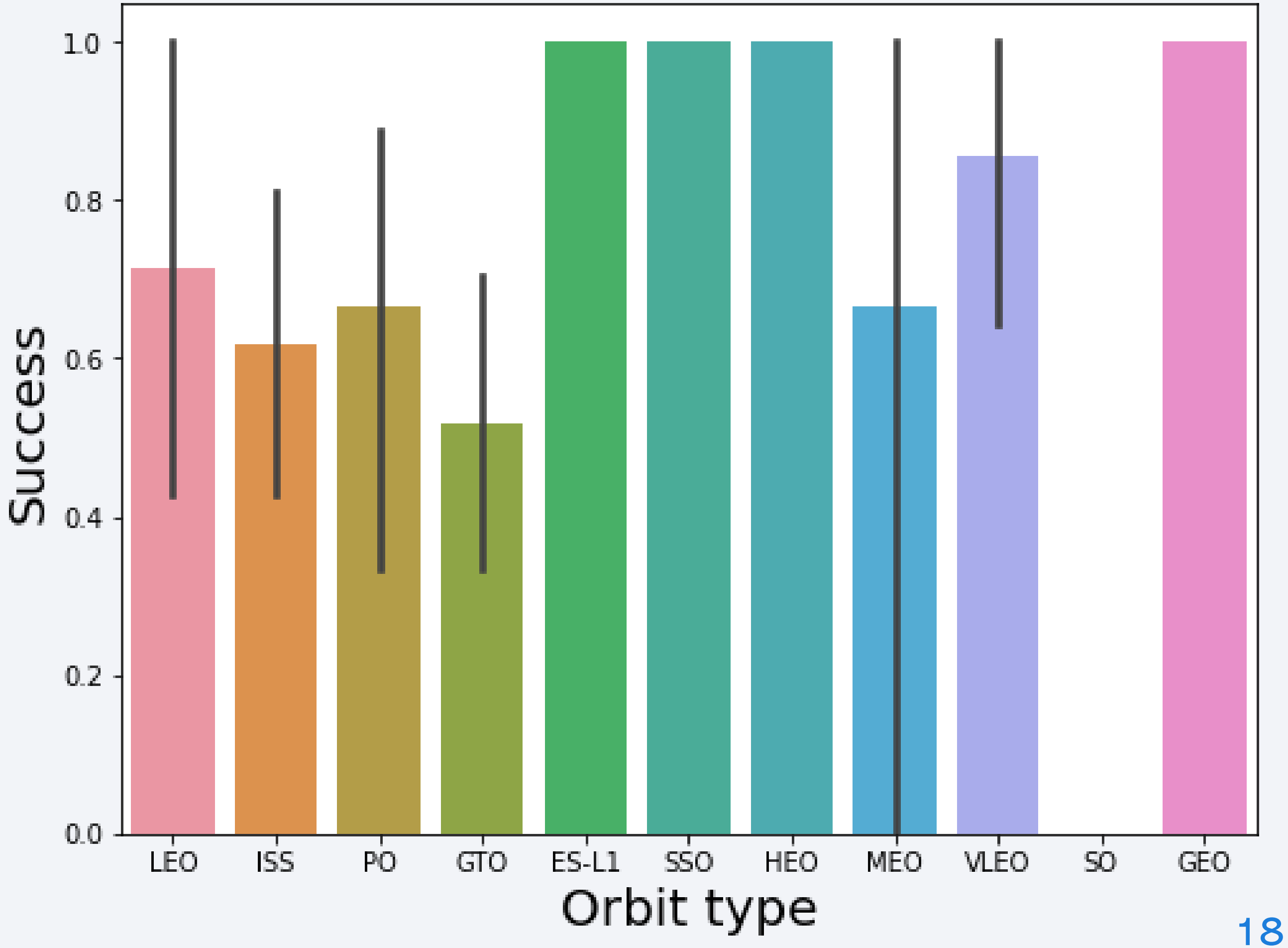
Payload vs. Launch Site

- There seems to be little correlation between payload mass and launch site beyond the launches at 9500kg from Vanderburg Air Force Base.



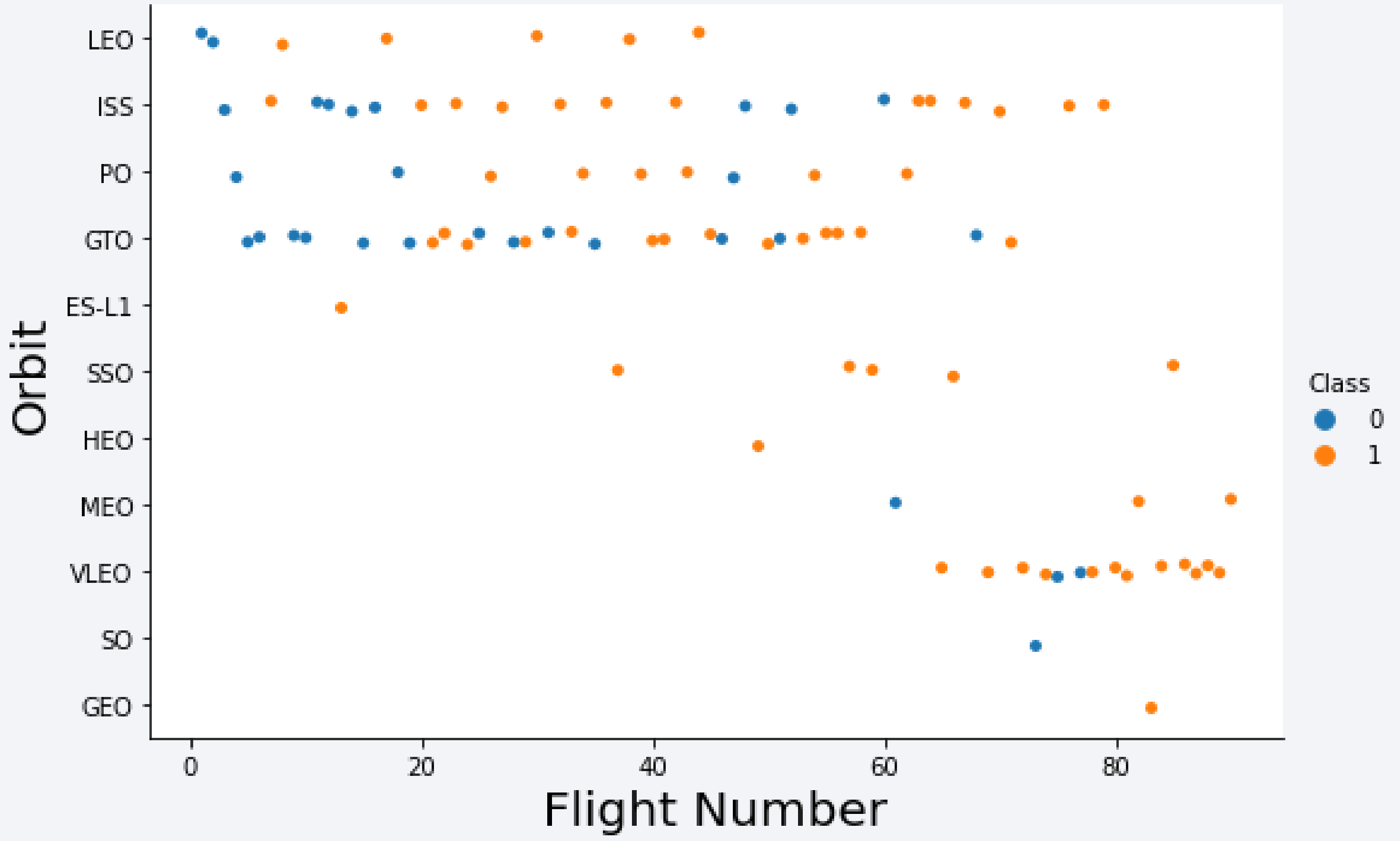
Success Rate vs. Orbit Type

- Success rate depends on the target orbit
- VLEO has higher average success and lower variance, while e.g. launches to the ISS have lower success rate and higher variance
- No meaningful conclusions for ES-L1, SSO, HEO, SO and MEO due to few or no launches into these orbits



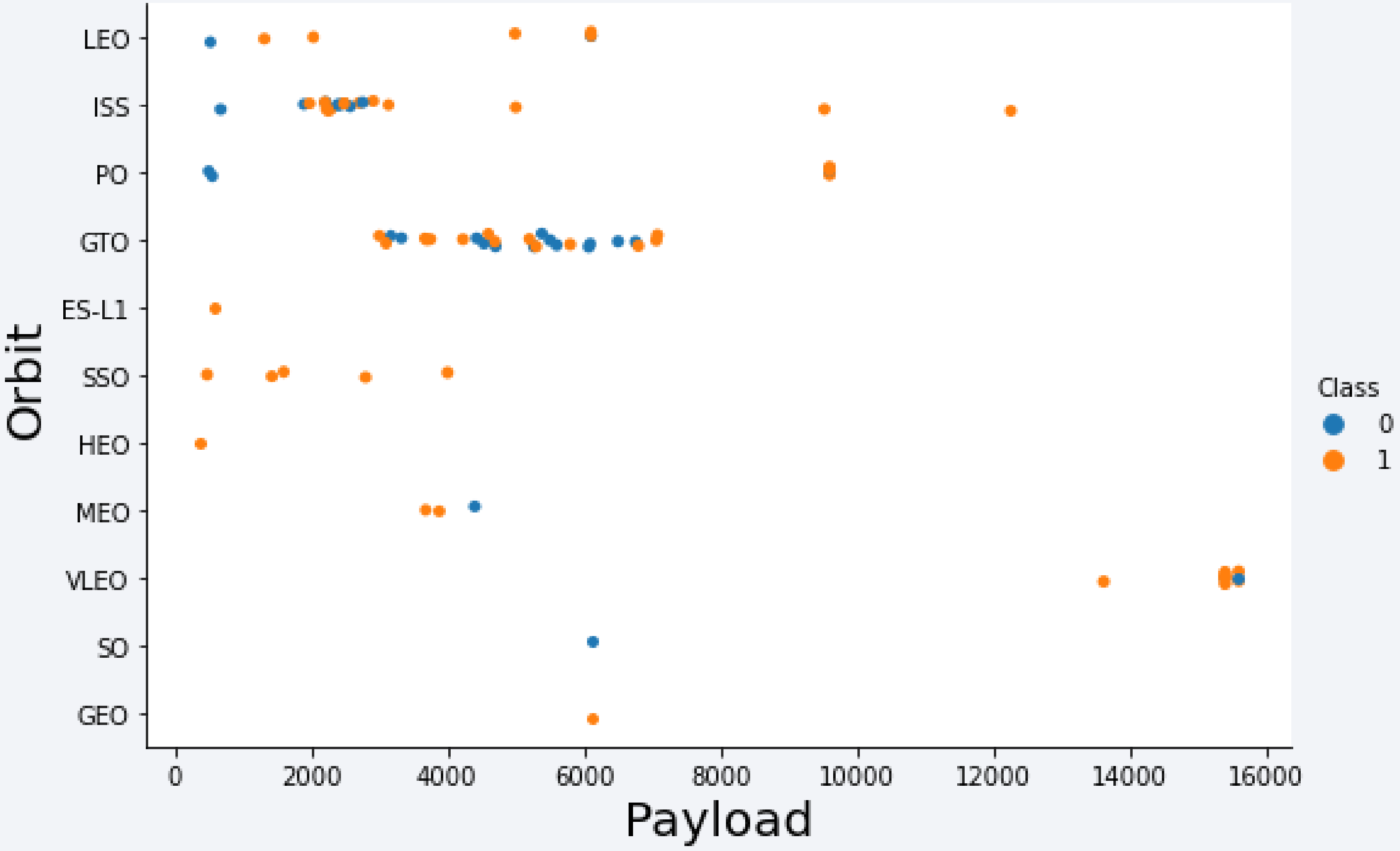
Flight Number vs. Orbit Type

- SpaceX has diversified its target orbits as they gain more experience
- The appearance of VLEO somewhat coincides with the first launches of Starlink, but the dissapearence of LEO might also suggest a reclassification of target orbits



Payload vs. Orbit Type

- Orbit type influences maximum payload: orbits with higher altitude or inclination will require more energy, and therefor allow a lower payload mass.

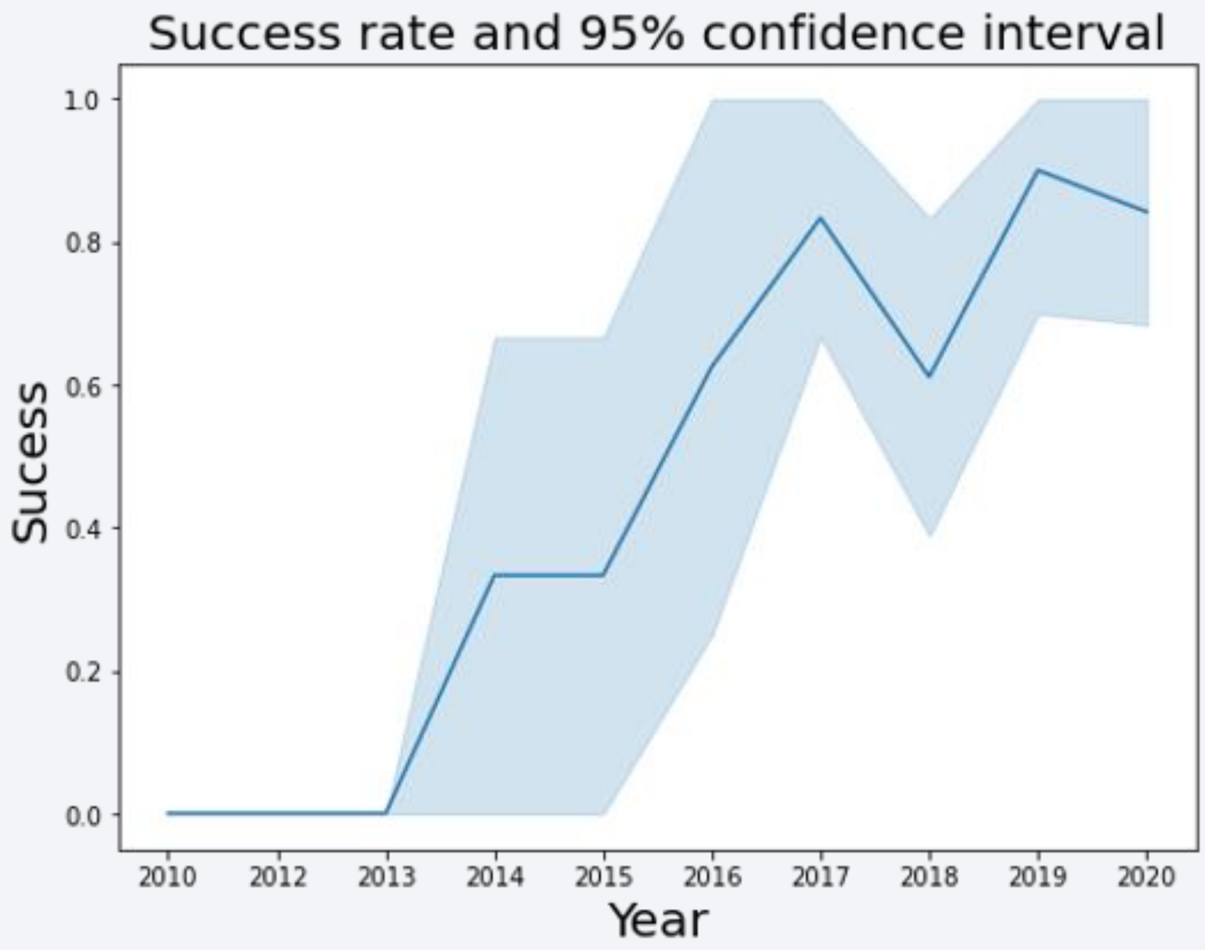


Launch Success Yearly Trend

•The feature with the greatest direct impact experience! SpaceX was able to learn from failure and the success rate steadily improve since the first success in 2014.

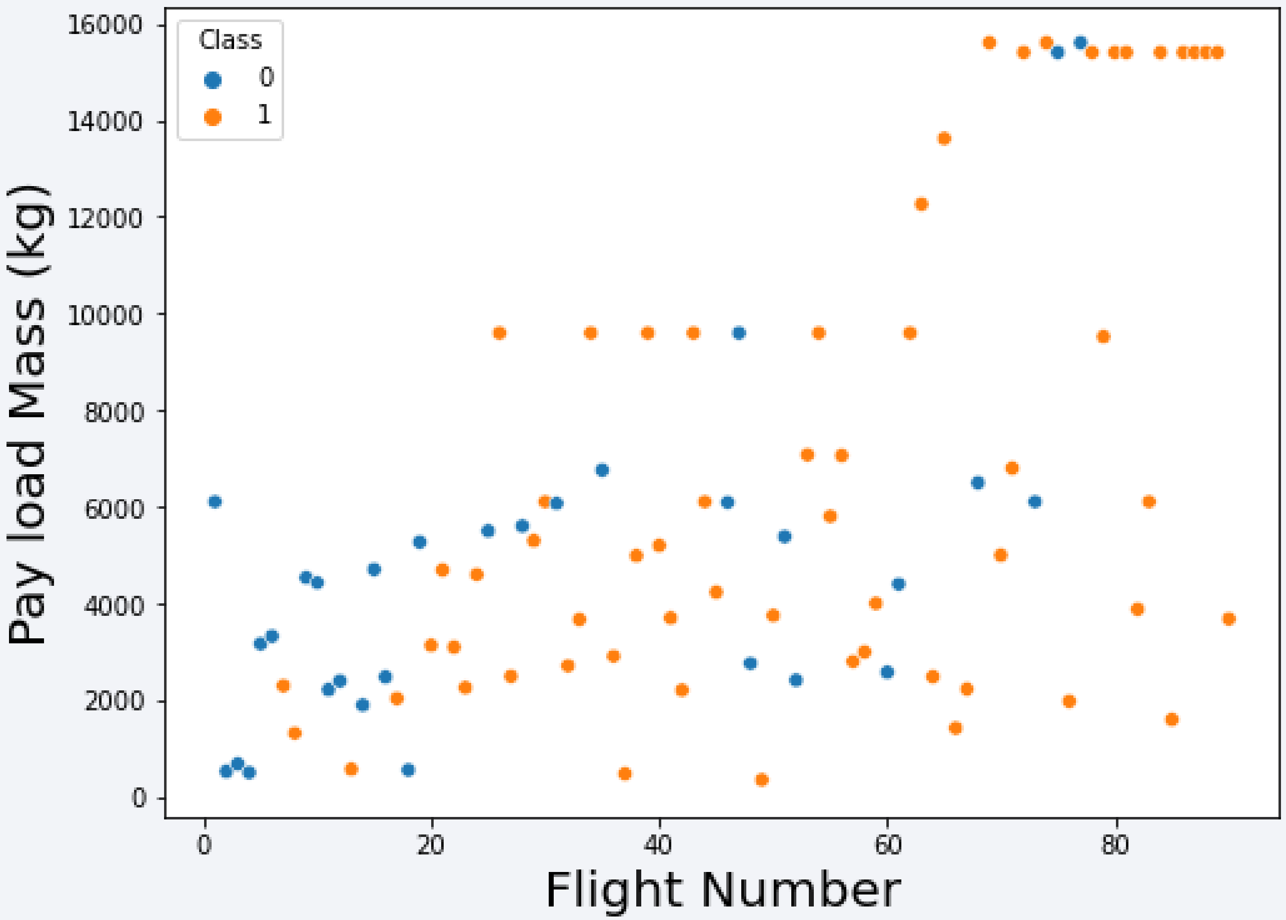
•The variance of the success rate also shrinks with time

•This is probably due to improvements in both software and hardware



Flight number & Payload mass

- It is also possible to see that with experience came a higher average payload mass…
- ….but also a higher success rate for the same payload mass



All Launch Site Names

- Find the names of the unique launch sites

```
%%sql
SELECT UNIQUE(launch_site)
FROM spacextbl;
```

```
 * ibm_db_sa://zwz82067:***
13f33084.bs2io90l08kqb1od8l
30367/bludb
Done.
```

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`

```
%%sql
SELECT *
FROM spacextbl
WHERE launch_site LIKE 'CCA%'
LIMIT 5;
```

```
 * ibm_db_sa://zwz82067:***@815fa4db-dc03-4c70-869a-a9cc13f33084.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:30367/bludb
Done.
```

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

```
%%sql
SELECT SUM(payload_mass__kg_) AS
FROM spacextbl
WHERE customer LIKE 'NASA (CRS)%';
```

```
 * ibm_db_sa://zwz82067:***@815fa4d
bludb
Done.
```

1
48213

Average Payload Mass by F9 v1.1

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

```
%%sql
SELECT AVG(payload_mass__kg_) as avg_payload_mass
FROM spacextbl
WHERE booster_version LIKE 'F9 v1.1%';
```

```
 * ibm_db_sa://zwz82067:***@815fa4db-dc03-4c70-869
bludb
Done.
```

avg_payload_mass
2534

First Successful Ground Landing Date

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

```
%%sql
SELECT MIN(date) as date_first_landing
FROM spacextbl
WHERE landing__outcome='Success (ground pad)';
```

```
 * ibm_db_sa://zwz82067:***@815fa4db-dc03-4c70
pdomain.cloud:30367/bludb
Done.
```

date_first_landing
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 spacextbl
WHERE landing__outcome='Success (drone ship)'
    AND payload_mass__kg_ BETWEEN 4000 AND 6000;
```

```
 * ibm_db_sa://zwz82067:***@815fa4db-dc03-4c70-86
pdomain.cloud:30367/bludb
Done.
```

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

```
%%sql
SELECT mission_outcome, COUNT(*) as number
FROM spacextbl
GROUP BY mission_outcome;
```

```
 * ibm_db_sa://zwz82067:***@815fa4db-dc03-4
pdomain.cloud:30367/bludb
Done.
```

mission_outcome	number
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

```
%%sql
SELECT UNIQUE(booster_version)
FROM spacextbl
WHERE payload_mass__kg_ IN (SELECT MAX(payload_mass__kg_) FROM spacextbl);
```

```
 * ibm_db_sa://zwz82067:***@815fa4db-dc03-4c70-869a-a9cc13f33084.bs2io90l0
pdomain.cloud:30367/bludb
Done.
```

booster_version
F9 B5 B1048.4
F9 B5 B1048.5
F9 B5 B1049.4
F9 B5 B1049.5
F9 B5 B1049.7
F9 B5 B1051.3
F9 B5 B1051.4
F9 B5 B1051.6
F9 B5 B1056.4
F9 B5 B1058.3
F9 B5 B1060.2
F9 B5 B1060.3

2015 Launch Records

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

```
%%sql
SELECT date, booster_version, launch_site, landing__outcome
FROM spacextbl
WHERE YEAR(DATE)=2015;
```

```
 * ibm_db_sa://zwz82067:***@815fa4db-dc03-4c70-869a-a9cc13f3
pdomain.cloud:30367/bludb
Done.
```

DATE	booster_version	launch_site	landing__outcome
2015-01-10	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
2015-02-11	F9 v1.1 B1013	CCAFS LC-40	Controlled (ocean)
2015-03-02	F9 v1.1 B1014	CCAFS LC-40	No attempt
2015-04-14	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)
2015-04-27	F9 v1.1 B1016	CCAFS LC-40	No attempt
2015-06-28	F9 v1.1 B1018	CCAFS LC-40	Precluded (drone ship)
2015-12-22	F9 FT B1019	CCAFS LC-40	Success (ground pad)

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

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

```
 * ibm_db_sa://zwz82067:***@815fa4db-dc03-4c70-869a-a9cc13f
pdomain.cloud:30367/bludb
Done.
```

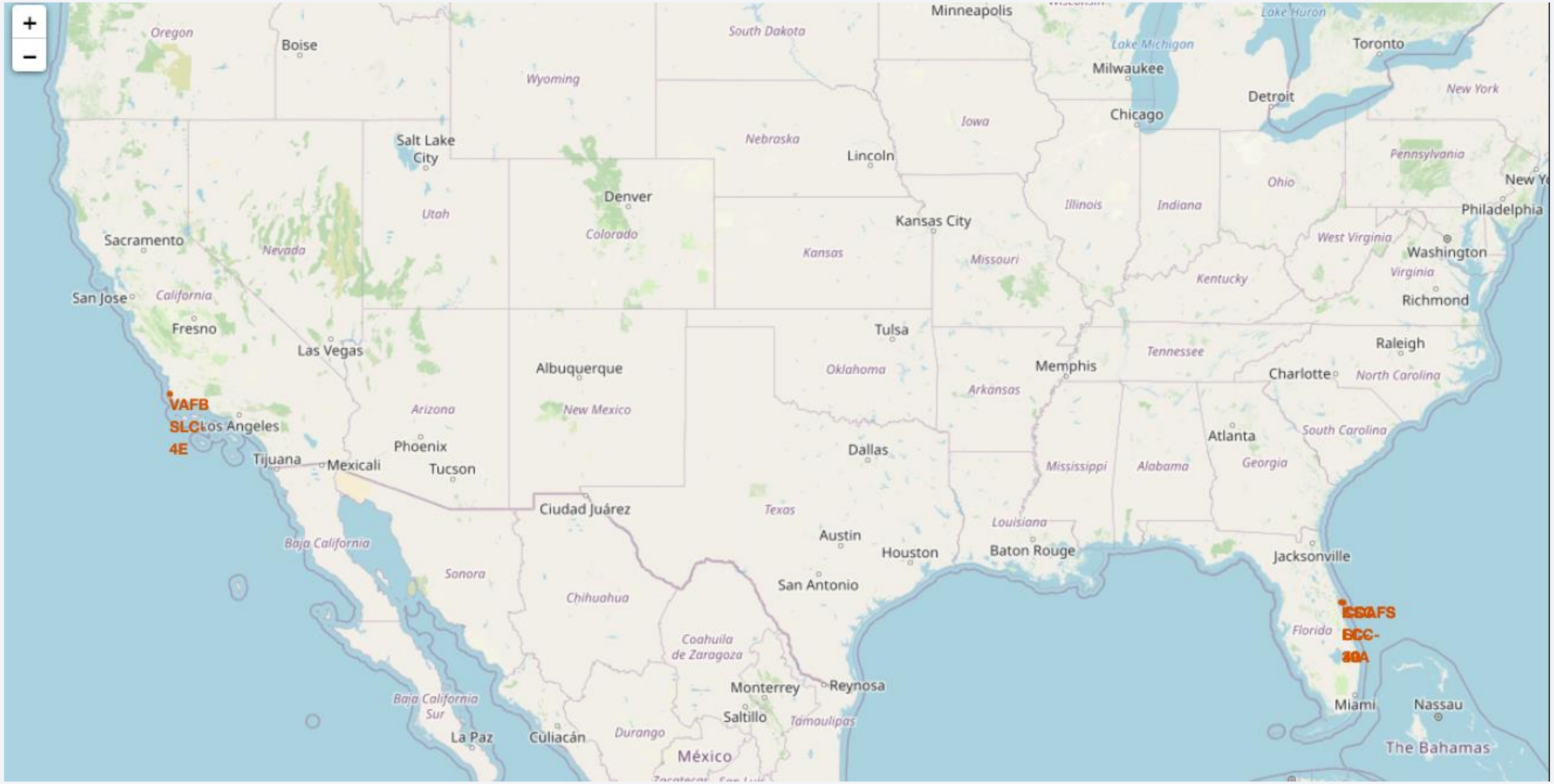
landing_outcome	number
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

SpaceX launch site

- SpaceX launch site are all near the coast and at low longitudes.

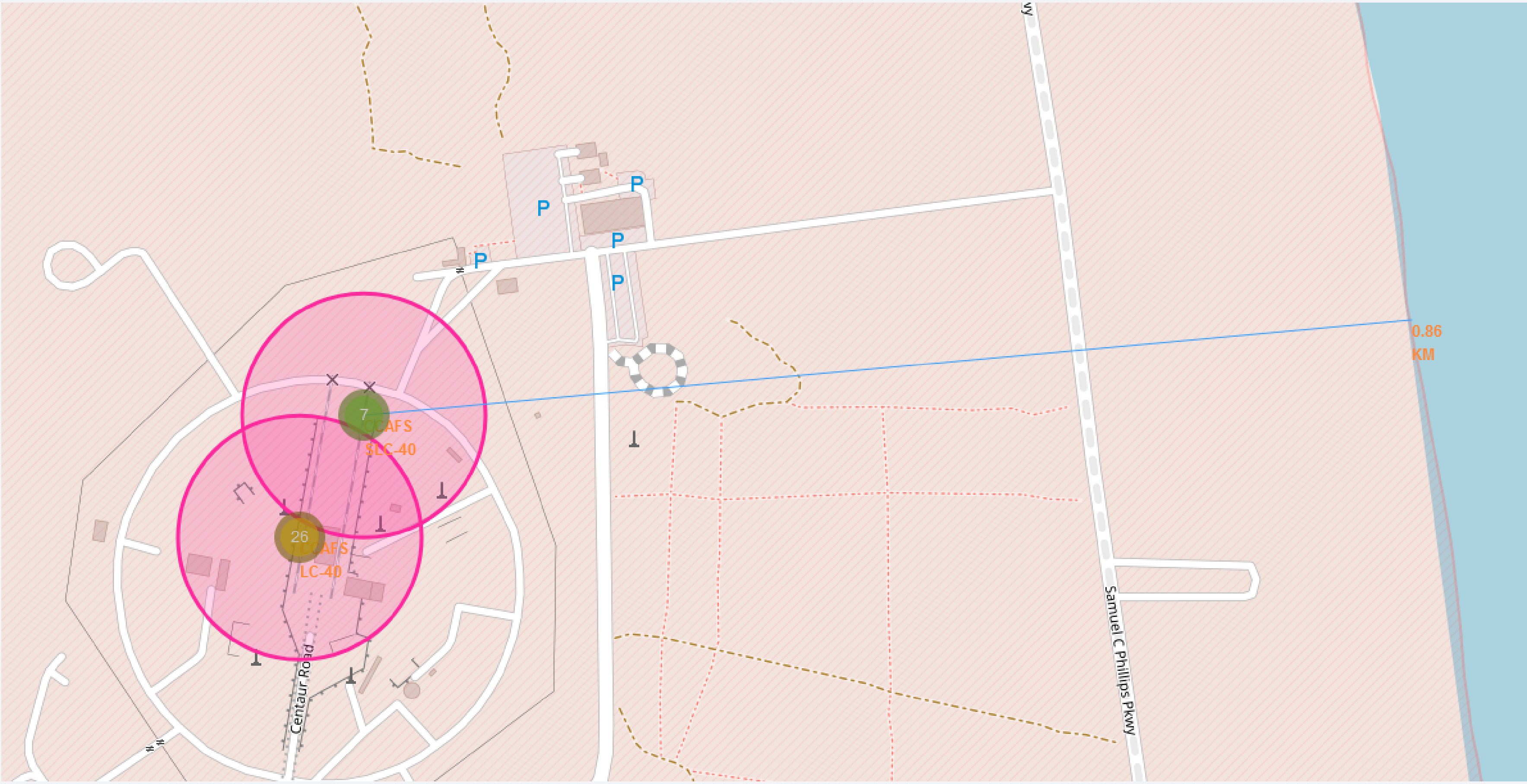


Launches from Vandenberg Space Force Base

- We can see the launches from VSFB (VAFB in the data), green for successful landings, and red for unsuccessful landings.
- VSFB accounts for the lowest success/fail ratio for landings

Distance to coastline

- We can see that the launch sites in Cape Canaveral are less than 1km from the coast

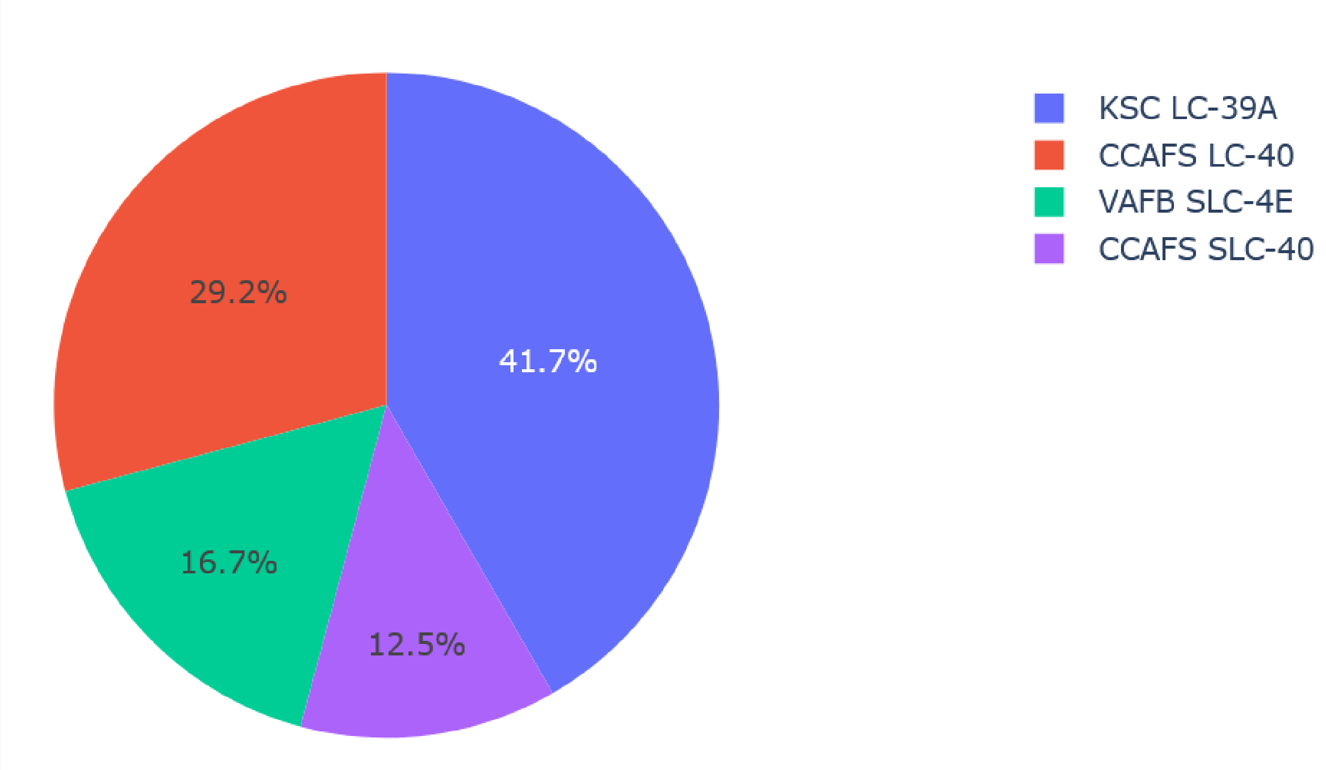


Section 4

Build a Dashboard with Plotly Dash

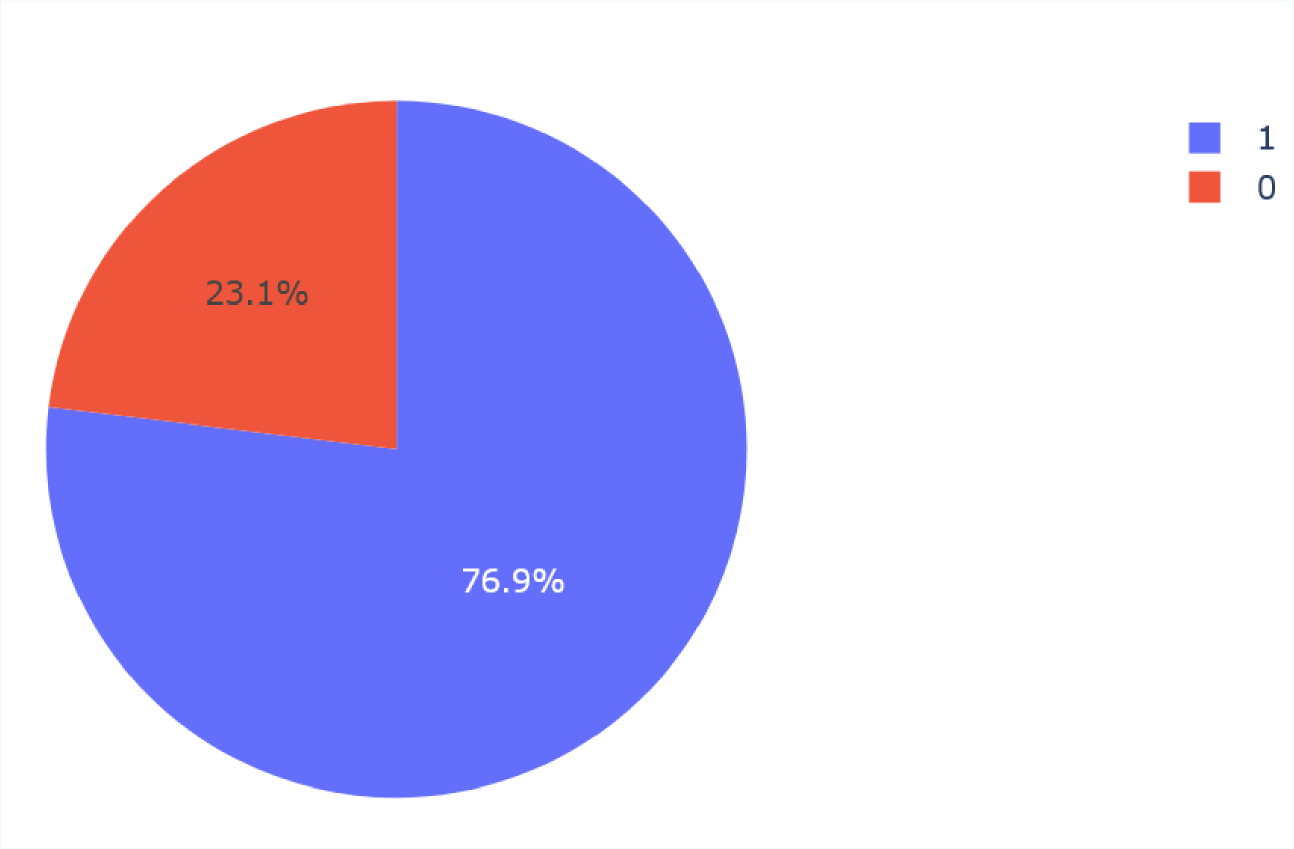
Launch success count for all sites

- We can see that almost half of the successful Landings come from KSC LC-39A



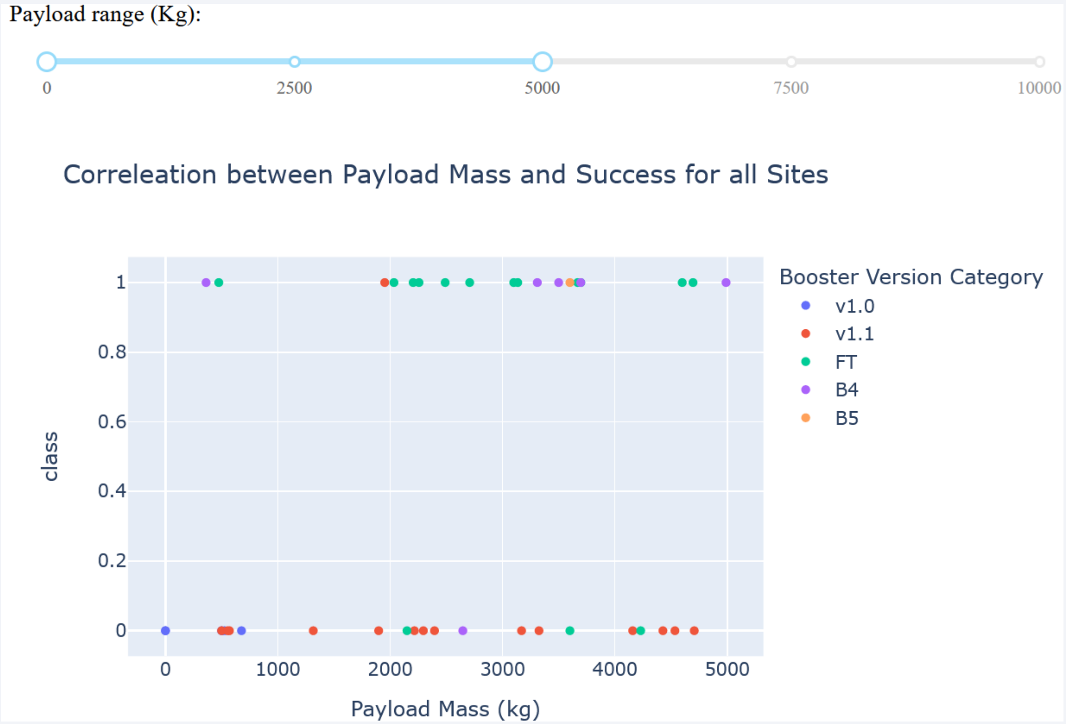
Successful Launches for site KSC LC-39A

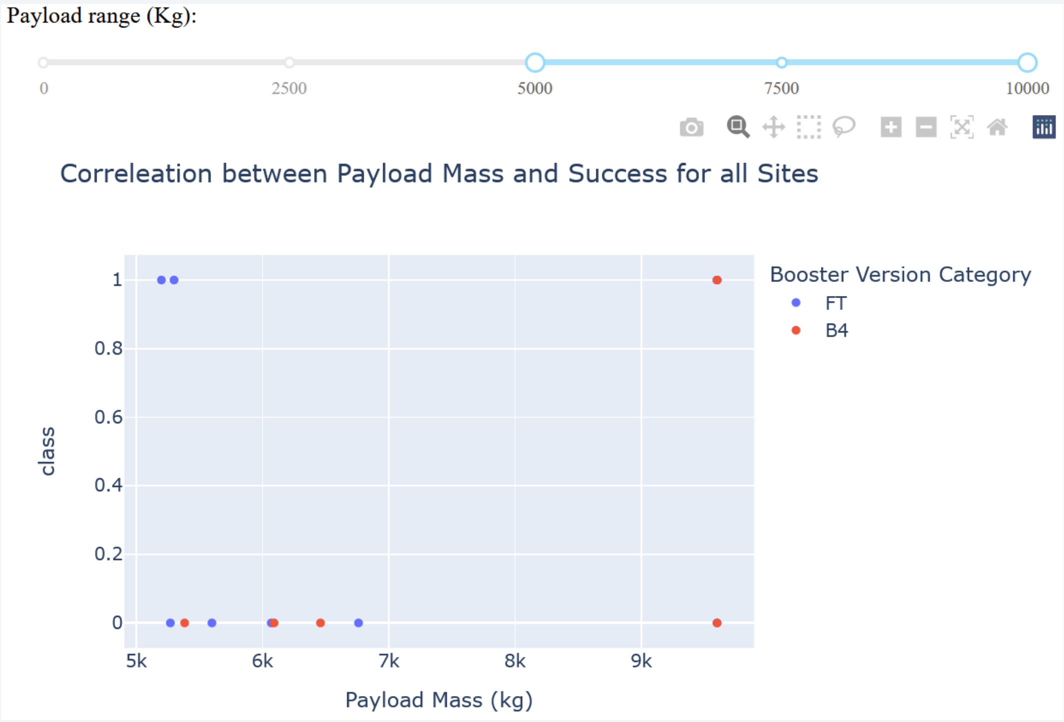
- Site KSC LC-39A accounts for 10 successes and 3 failures



Payload vs. Launch Outcome

- We can see that heavy launches are exclusively done with the FT or B4 boosters, while all of them are used for lighter payloads.
- Heavier payloads seem to have a smaller success rate



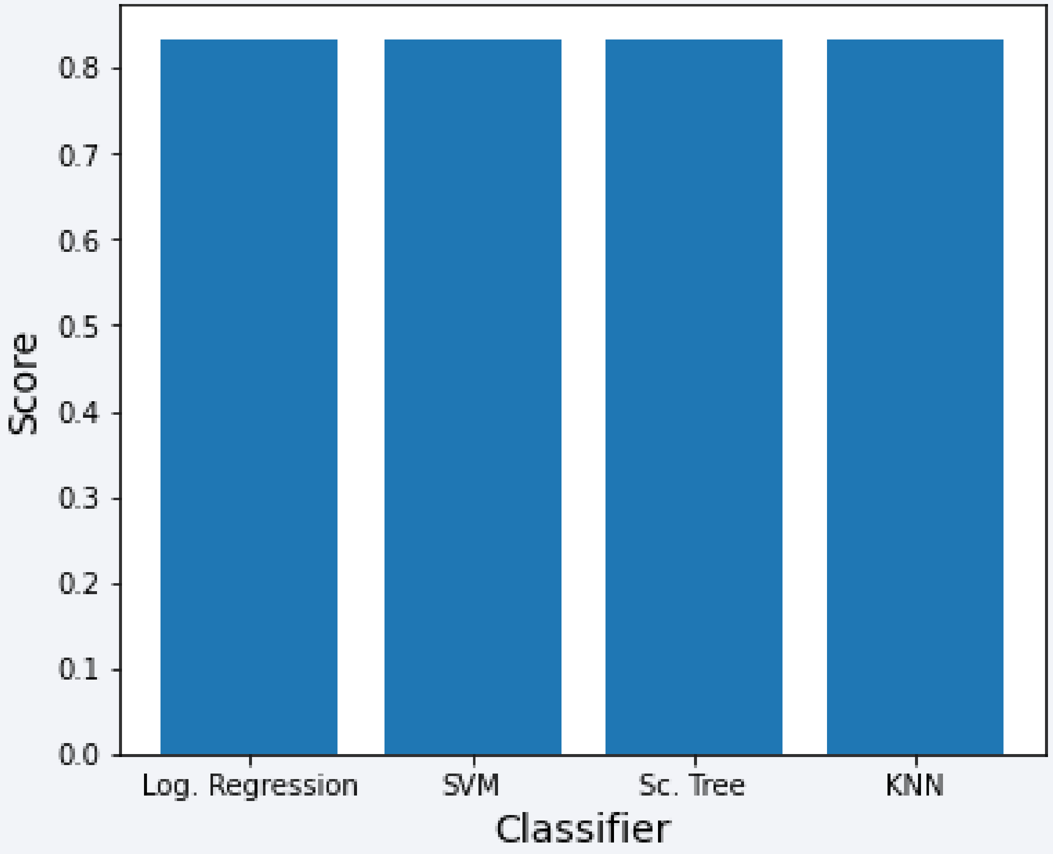


Section 5

Predictive Analysis (Classification)

Classification Accuracy

- All classifiers achieved the same accuracy, at around 83.3%



Confusion Matrix – Decision Tree

- The classifier using logistic regression was able to correctly predict 10 successful landings and 5 failures
- However, it misclassified one failure as a success as 2 successes as failures

Conclusions

- Reusable rocket boosters are the future of the space industry
- Being able to predicted when other companies (like SpaceX) land their boosters can help a newcomer to the business decide which markets to target
- For example, we could develop a better booster than SpaceX to outperform them in a certain sector (payload mass range, orbit, launch site, etc.)
- Our models can predict whether SpaceX will land a booster with 83% accuracy
- Future models could improve this performance by taking other factors into account, such as weather conditions

Thank you!