IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Mason Flannery

June 29, 2023

Outline

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

Executive Summary

- This project sets out to predict if a given Falcon 9 first stage rocket will land successfully–without all the rocket science gobbledygook. With the dramatic decrease in cost largely due to the reusability of the rockets, we set out to determine the cost of the launch. To achieve this, we collected data from several reliable sources, thoroughly explored the data through descriptive EDA, prepared the data for analysis through feature engineering, and analyzed the data using advanced machine learning techniques, including support vector machines and decision tree classification. The results were finding a robust model that could predict if a given rocket would successfully land 84% of the time.

Introduction

- Space travel, once a thing for only mega rich corporations and governments, has become something far more obtainable by individuals–with an average rocket costing about 165 million dollars each. However, through Elon Musk's efforts with his company SpaceX, a trip to space can cost as low as 65 million dollars. The dramatic reduction in cost is largely tied to the reusability of Musk's rockets. Thus, we set out to see if the rockets land consistently. If they don't, then we expect alternate companies can still outbid SpaceX for a rocket launch.
- In this project, we set out to solve the following problems:
    - Can we predict how often a SpaceX rocket will land successfully?
    - What are the best predictors of a successful landing?

Section 1

Methodology

Methodology

Executive Summary

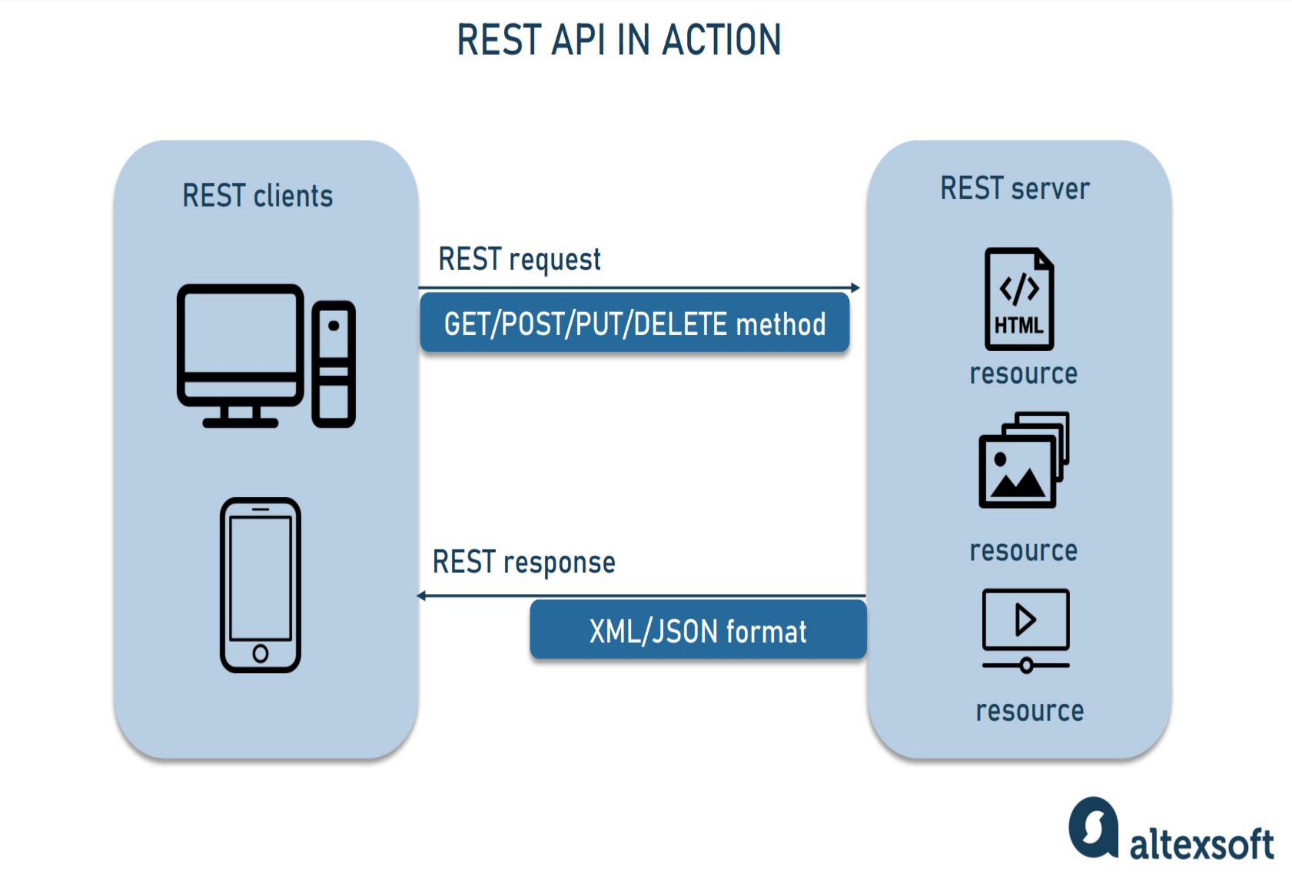
- Data collection methodology:
  - Data was collected in 2 ways: web scraping and SpaceX's REST API
- Perform data wrangling
  - Data was processed using the pandas library
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Multiple models were thoroughly tested until the best was selected

Data Collection

- The data was collected from the two following resources:
    - SpaceX’s official API
    - Web scraping Wikipedia data on launch details
- Through combining these two resources we were able to gather sufficient data to accurately predict the landing state of the rockets

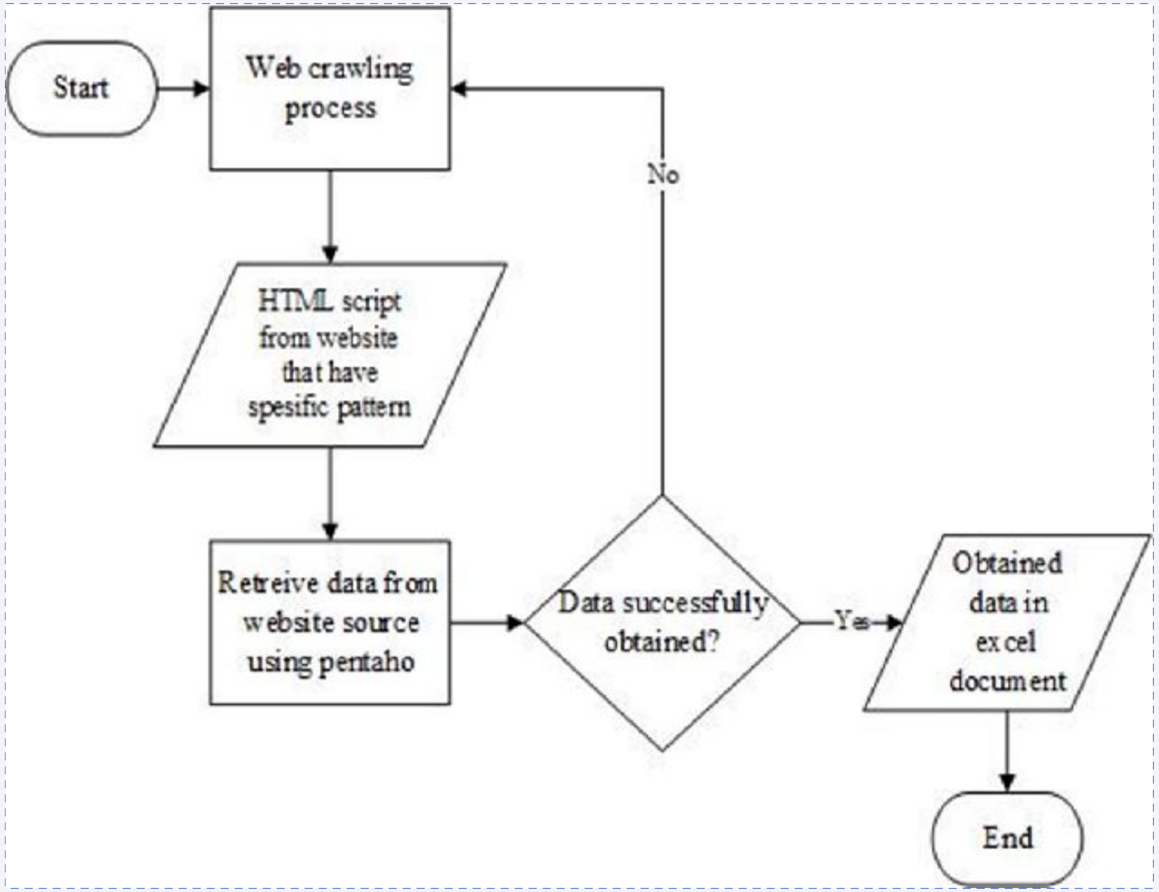
Data Collection – SpaceX API

- Much of the data came from the SpaceX official API. This resource provided reliable data on every launch. The data was filtered for those pertaining to the Falcon 9 rockets, and more can be found on the specifics at the following link:
- https://github.com/Ddotsam/IBM-Certification/blob/main/ds_capstone/work_books/collect-data-api.ipynb



Data Collection - Scraping

- Some data was also scraped from wikipedia using BeautifulSoup in Python. This data was rigorously searched to ensure validity, and it was found to be accurate in all ways pertaining to this project. More on the specific ifcs can be found at the link below:
- https://github.com/Ddotsam/IBM-Certification/blob/main/ds_capstone/work_books/collect-data-webscrape.ipynb



Data Wrangling

- In order to feed the data into the predictive models, the data must first be turned into a form that is conducive to machine learning. To do this, we thoroughly wrangled the data. This task included
    - Filtering for those tuples that were of interest
    - Organizing the data into vectorized objects using numpy and pandas in Python
    - Analyzing what data we had to ensure that there was sufficient
- More can be found at the following link
- https://github.com/Ddotsam/IBM-Certification/blob/main/ds_capstone/work_books/data-wrangling.ipynb

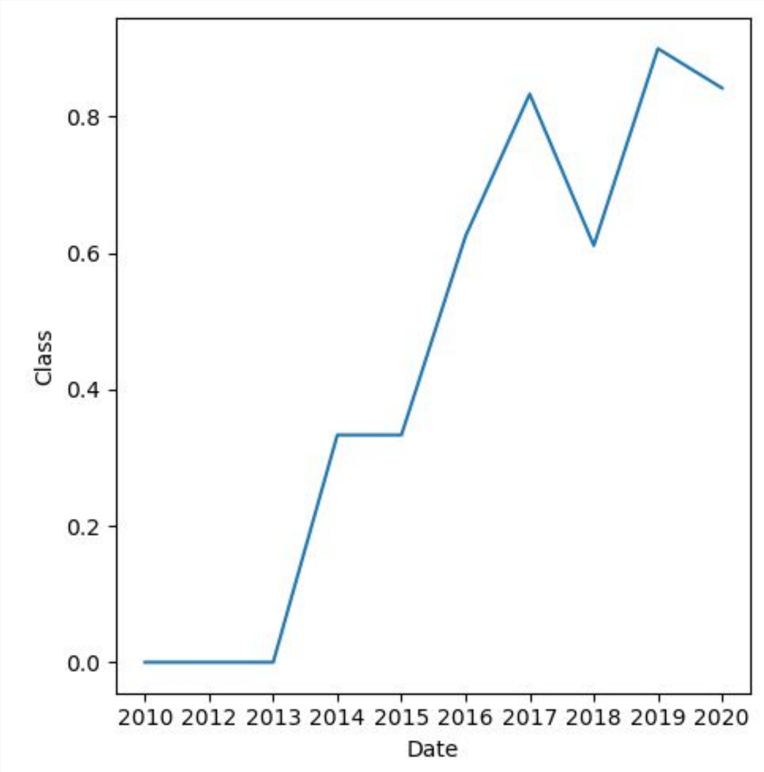
EDA with Data Visualization

- To explore the data, many charts were built, scatter plots and line plots in particular. It was discovered that reusability increased with time, that is, as the years passes, SpaceX only got better at landing their rockets.

- It was also discovered that certain orbits and launch sites were better at landing rockets, which suggests that decision tree classifiers or support vector machines would fit the data well.

- More can be found at the following link

- https://github.com/Ddotsam/IBM-Certification/blob/main/ds_capstone/work_books/EDA-mpl.ipynb



EDA with SQL

- We also sought to understand the data in a tabular format using Structured Query Language (SQL).
- This also put the data in a friendly format for relational use later

- More can be found at the following link
- https://github.com/Ddotsam/IBM-Certification/blob/main/ds_capstone/work_books/EDA-sql.ipynb

Build an Interactive Map with Folium

- Though EDA with SQL, we discovered that the launch sites shared a one-to-many relationship with the landing sites. That is, every mission launched from just one site, but it could land at any given return site. Since some of these were on the ocean, we set out to map the launch sites and their class with where they launched and where they landed using the python package folium.
- Interestingly, almost no relationship was found
- More can be found at the link
- https://github.com/Ddotsam/IBM-Certification/blob/main/ds_capstone/work_books/Analysis-folium.ipynb

Build a Dashboard with Plotly Dash

- Using Plotly and Dash, we created an interactive dashboard web application to further explore the data.
- We discovered insights on the success rate of various boosters when launched at various places. For example, FT version of the Falcon 9 Rocket was nearly always successful when launched from CCAFS LC-40
- More can be found at the following link:
- https://github.com/Ddotsam/IBM-Certification/blob/main/ds_capstone/work_books/Analysis-dashboard.py

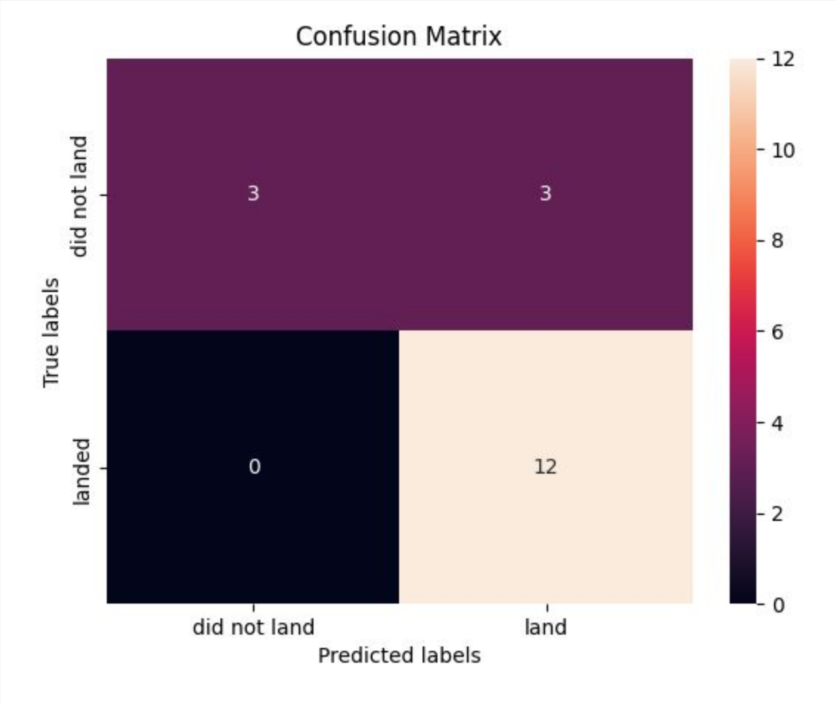
Predictive Analysis (Classification)

- Many models were tested to see which performed the best. We used many accuracy metrics, including F1 score, jaccard index, and log loss. Hyperparameter spaces were explored thoroughly using Grid Search methodologies.

- more on the specifics can be found at the following link
- https://github.com/Ddotsam/IBM-Certification/blob/main/ds_capstone/work_books/prediction-ML.ipynb

Results

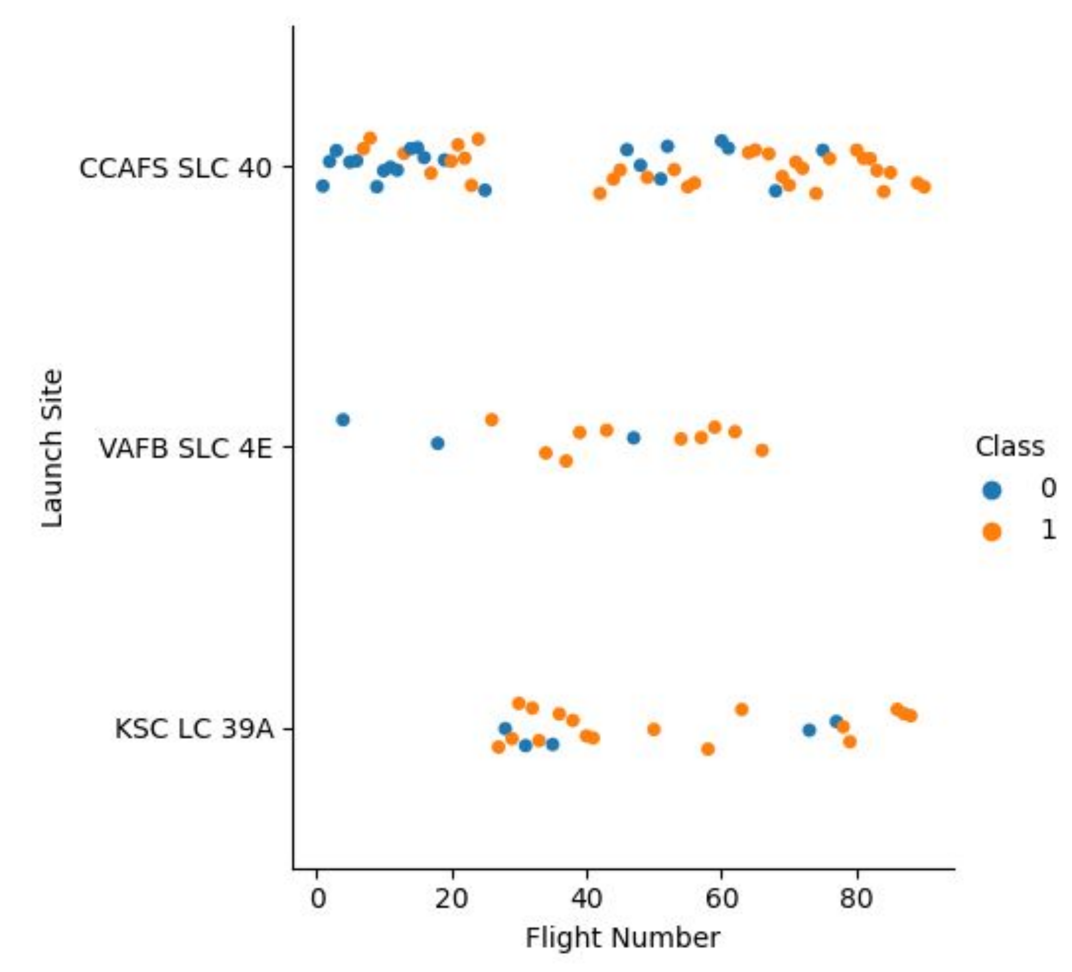
- In the end, we build a robust model that was able to predict the outcome of a given launch with 84% accuracy. This will help in the decision process of choosing a space travel provider



Section 2

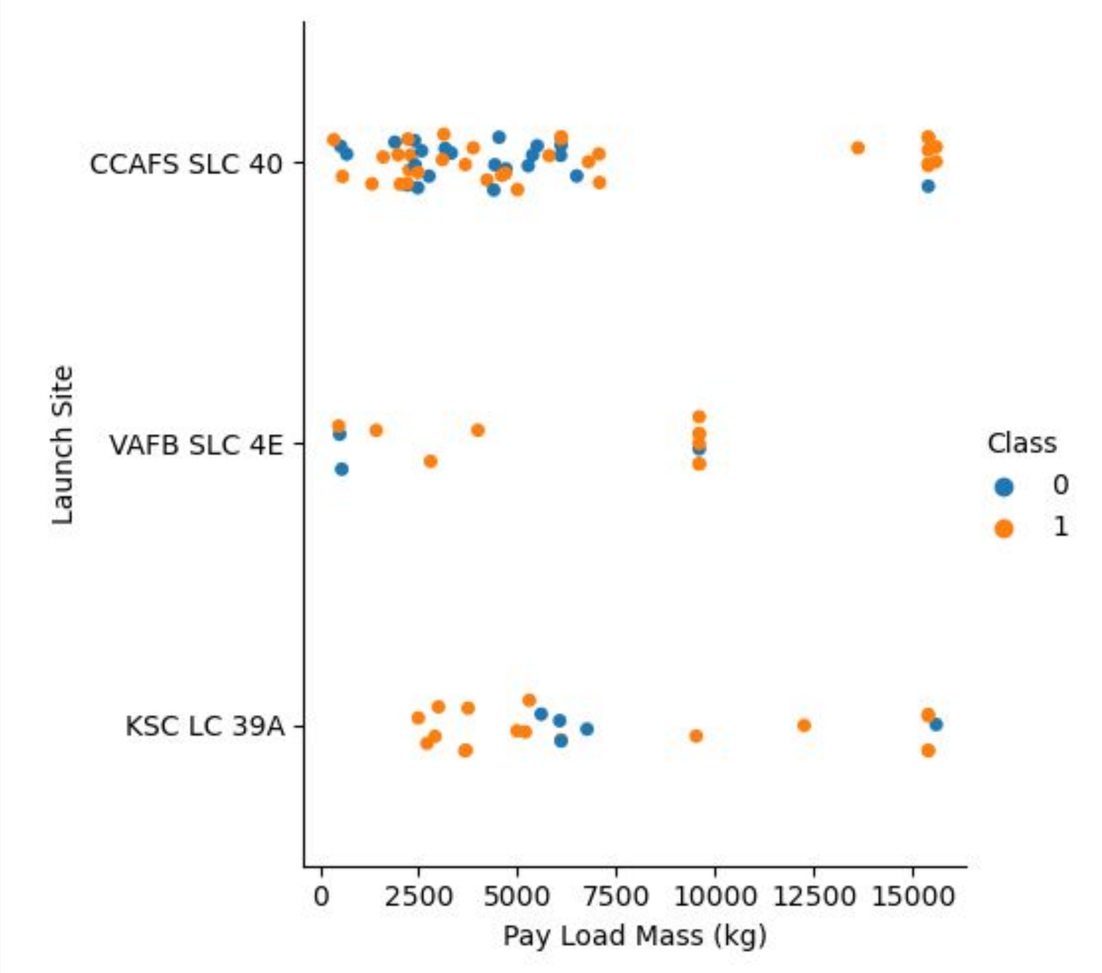
Insights drawn from EDA

Flight Number vs. Launch Site



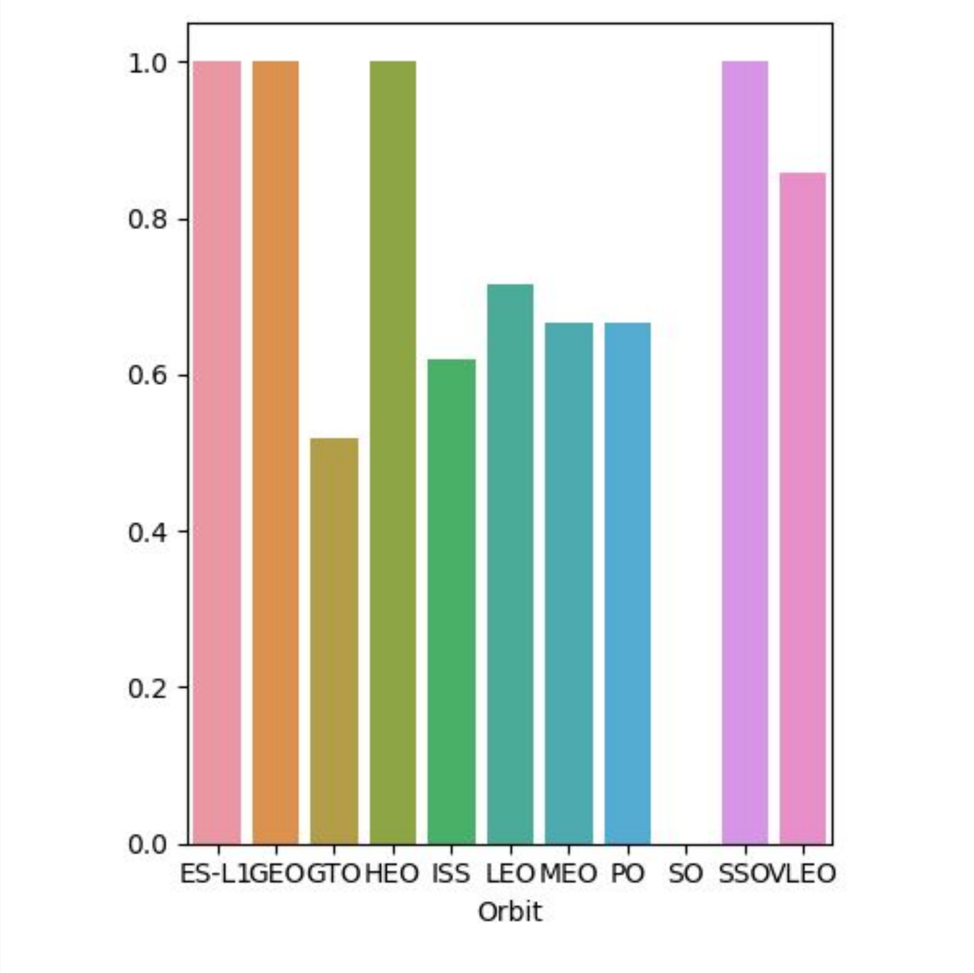
- As you can see, the successful landings (orange dots) seem to be disconnected from the unsuccessful landings (blue dots) in many cases. This shows that machine learning algorithms that focus on euclidean separation such as decision trees or support vector machines could be suitable

Payload vs. Launch Site



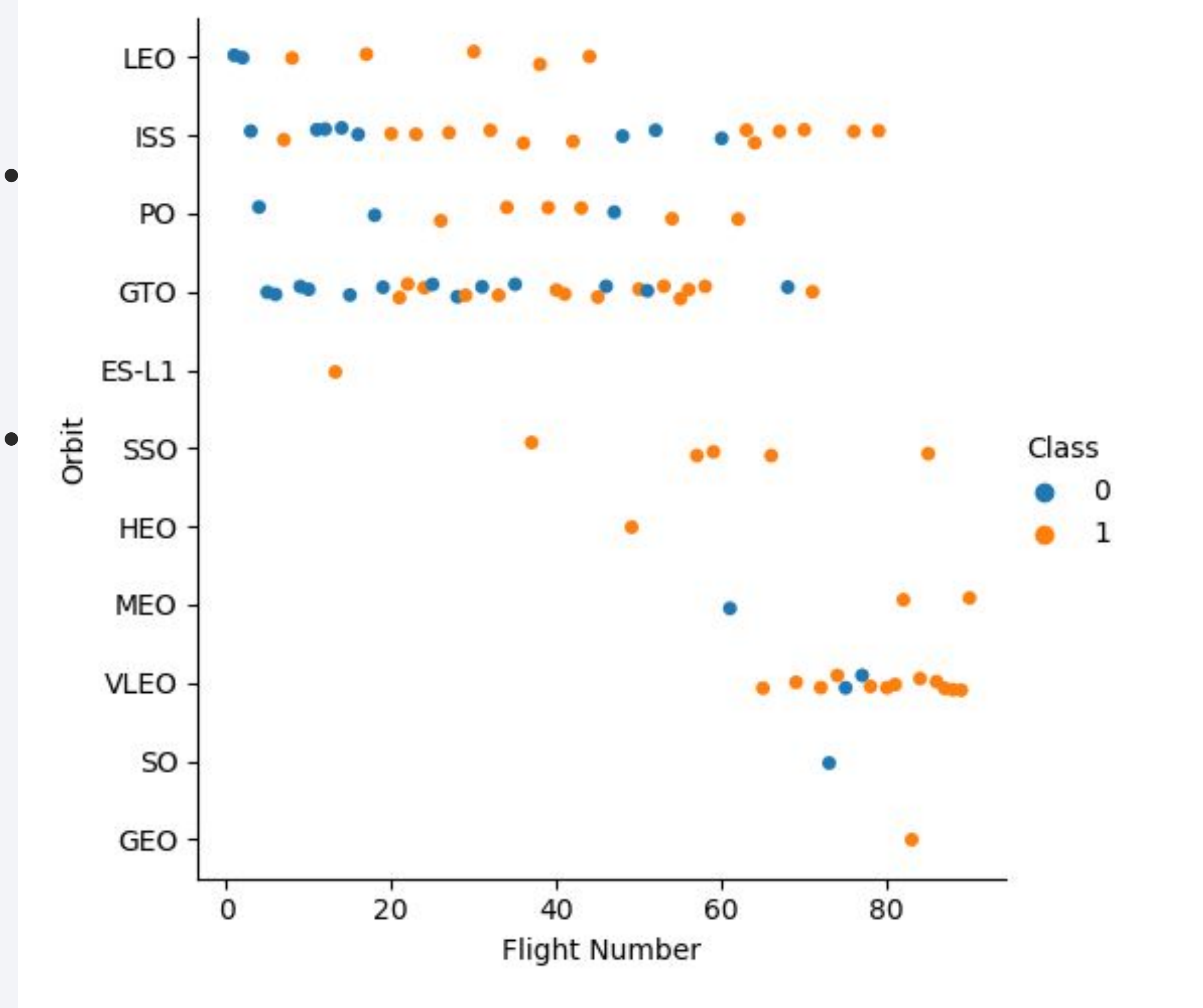
- The vertical lines that seem to form in the launch site sections suggest that the launch sites have a maximum payload mass. Additionally, it seems that as we increase the mass, the likelihood of the rocket successfully landing increases

Success Rate vs. Orbit Type



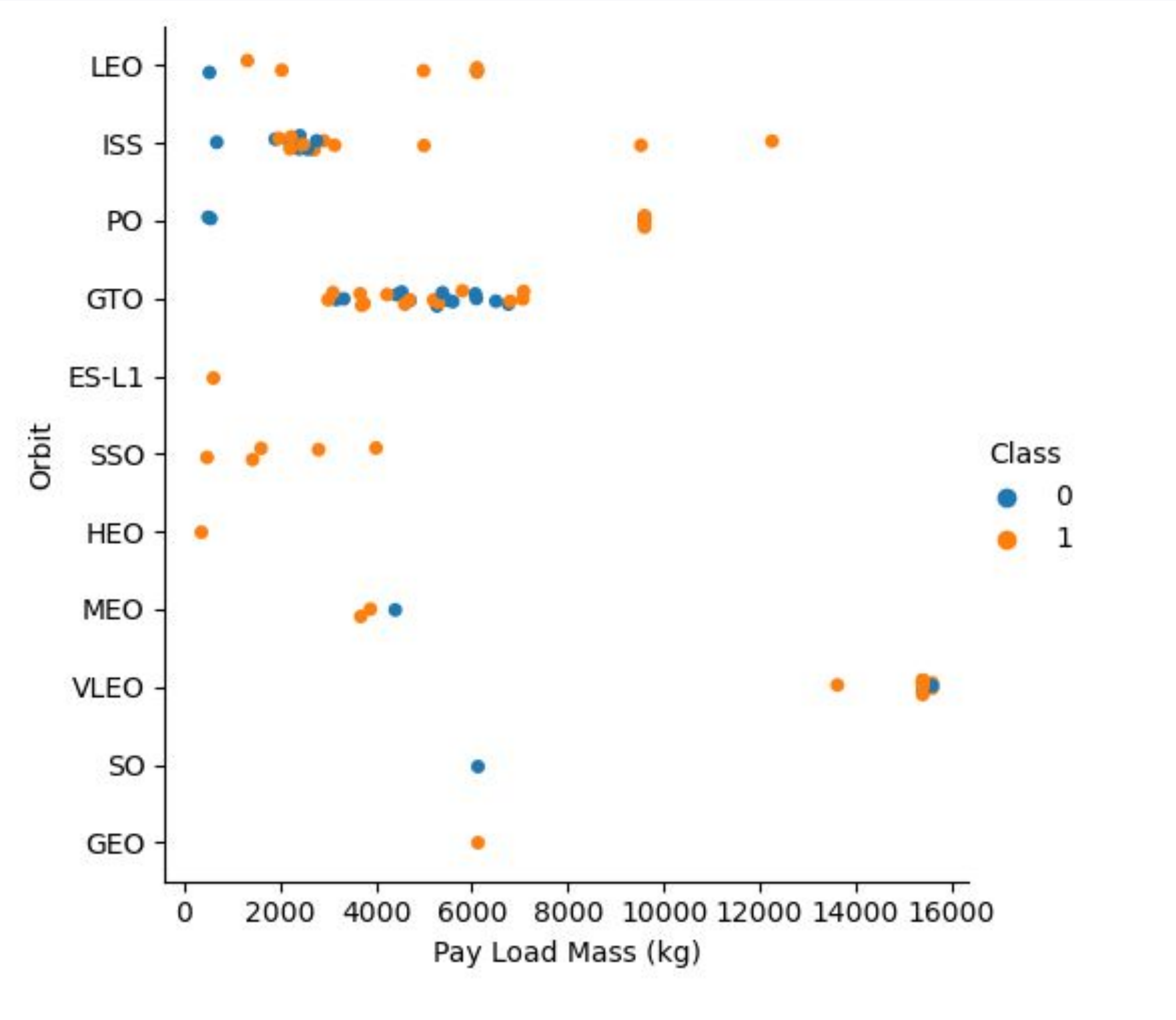
- here we see the success rate (bar height) plotted against the orbit type (x axis).
- The orbit types GEO, SO, HEO, and ES-L1 all had only 1 launch, so their relative outstanding success and failure rates are to be taken with a grain of salt.
- SSO orbits had multiple successful launches.

Flight Number vs. Orbit Type



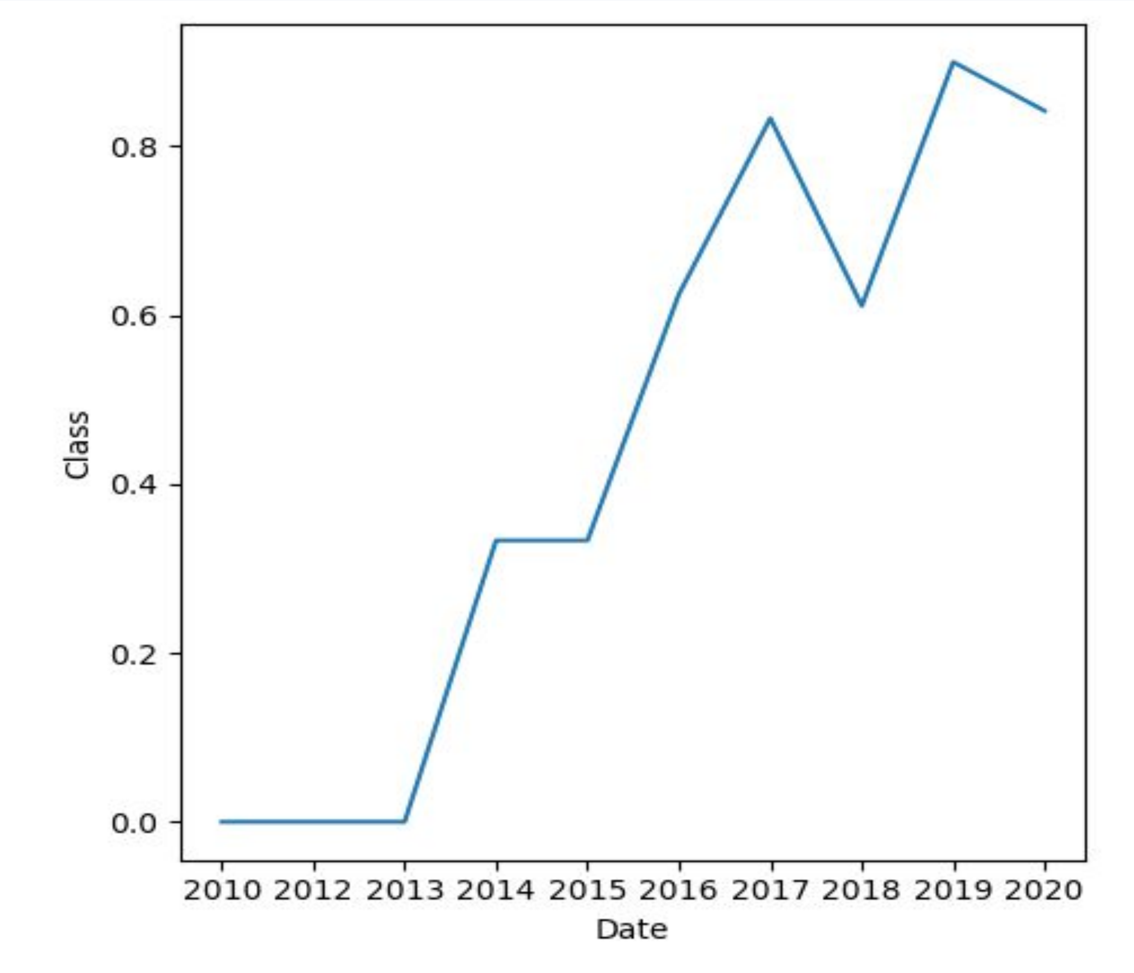
- As flight number increases, so does time. As you can see, landings were largely failures in the early stages of SpaceX, but outcomes have improved with time
- Certain orbits tend to have troubles all throughout, while other orbits are better in general

Payload vs. Orbit Type



- Here is the payload mass (x-axis) plotted against the type of orbit (y-axis).
- A similar trend of likelihood of a successful landing increasing with payload mass is shown, but this graph also shows that perhaps SpaceX is simply better at landing from certain orbit heights, regardless of payload mass

Launch Success Yearly Trend



- As you can see, SpaceX has gotten better every year with landing their rockets.
- comparing the success with other metrics, it seems that time is perhaps the best predictor of landing success
  - as time increases, we reliably predict a positive outcome.
  - as time decreases, we reliably predict a negative outcome

All Launch Site Names

```
%sql SELECT DISTINCT(Launch_Site) FROM SPACEXTBL
```

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

Launch_Site
CCAFS LC-40
VAFB SLC-4E
KSC LC-39A
CCAFS SLC-40
None

- these are the launch sites considered under this analysis
- all Falcon 9 launches occurred at one of these sites.

Launch Site Names Begin with 'CCA'

```
%sql SELECT Launch_Site FROM SPACEXTBL WHERE Launch_Site LIKE 'CCA%' LIMIT 5
```

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

Launch_Site
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40

Here are 5 of the launches that took place at a site beginning with the string “CCA”

Total Payload Mass

```
%sql SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE Customer = 'NASA (CRS)'
```

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

SUM(PAYLOAD_MASS__KG_)
45596.0

The total payload mass in kilograms carried by SpaceX rockets for NASA was discovered to be 45596 kilograms

Average Payload Mass by F9 v1.1

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

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

AVG(PAYLOAD_MASS__KG_)
2928.4

The average payload mass in kilograms carried by Falcon 9 Rockets (Version 1.1 only) was discovered to be 2928.4 kilograms

First Successful Ground Landing Date

```
%sql SELECT MIN(Date) FROM SPACEXTBL WHERE Landing_Outcome = 'Success (ground pad)'
```

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

MIN(Date)
01/08/2018

The first successful landing on a ground pad took place on August 1st, 2018

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%sql SELECT * FROM SPACEXTBL WHERE Landing_Outcome = 'Success (drone ship)' AND (PAYLOAD_MASS__KG_ > 4000 AND PAYLOAD_MAS
```

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

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outco
05/06/2016	5:21:00	F9 FT B1022	CCAFS LC-40	JCSAT-14	4696.0	GTO	SKY Perfect JSAT Group	Success	Success (dr s
14/08/2016	5:26:00	F9 FT B1026	CCAFS LC-40	JCSAT-16	4600.0	GTO	SKY Perfect JSAT Group	Success	Success (dr s
30/03/2017	22:27:00	F9 FT B1021.2	KSC LC-39A	SES-10	5300.0	GTO	SES	Success	Success (dr s
10/11/2017	22:53:00	F9 FT B1031.2	KSC LC-39A	SES-11 / EchoStar 105	5200.0	GTO	SES EchoStar	Success	Success (dr s

These are the only successful drone ship landings that had a payload between 4000 kilograms and 6000 kilograms

Total Number of Successful and Failure Mission Outcomes

```
%sql SELECT mission_outcome, COUNT(*) FROM SPACEXTBL GROUP BY mission_outcome
```

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

Mission_Outcome	COUNT(*)
None	898
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Here are the number of successful and failure mission outcomes

Boosters Carried Maximum Payload

```
%sql SELECT booster_version FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAY
```

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

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

These are the specific versions of the boosters that carried the maximum payload mass

2015 Launch Records

```
%sql SELECT substr(Date, 4, 2) MONTH,  substr(Date,7,4) YEAR, Booster_Versior
```

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

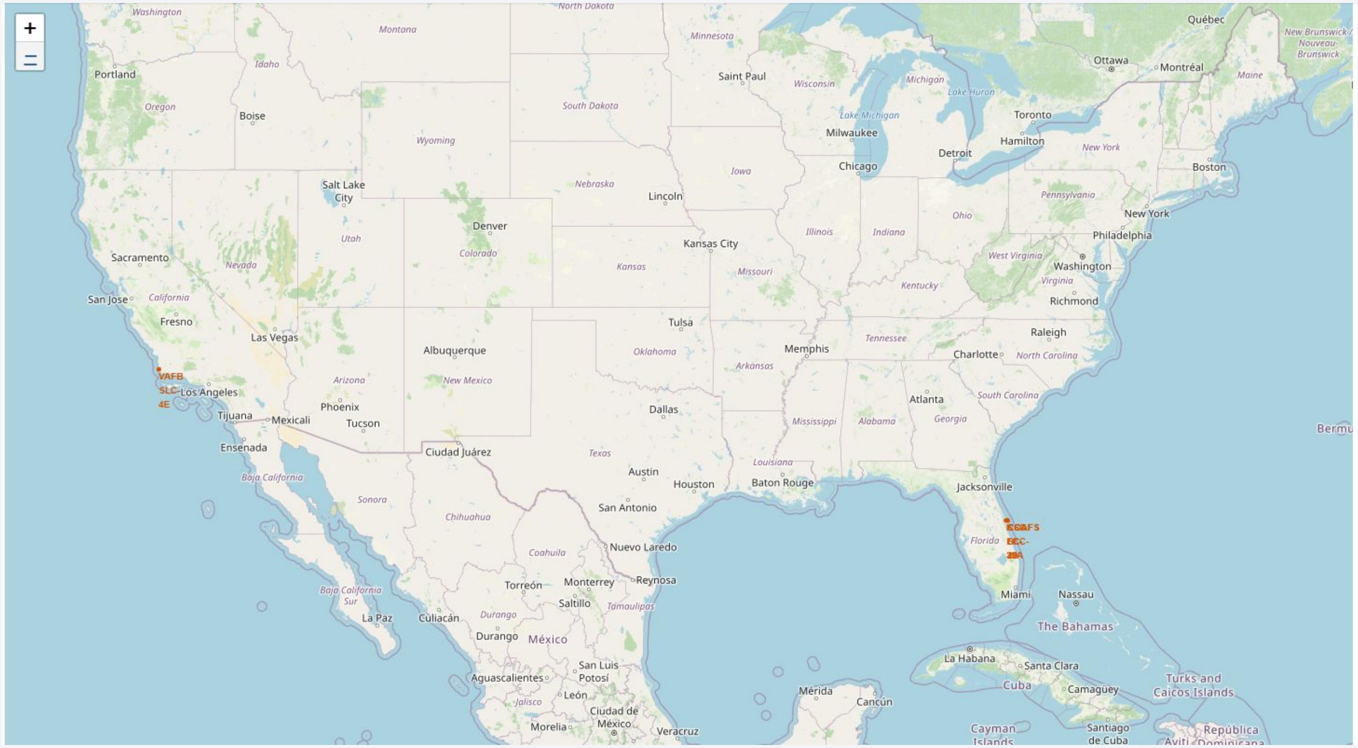
MONTH	YEAR	Booster_Version	Launch_Site	Landing_Outcome
10	2015	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
04	2015	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

These are all the launches that attempted to land on a drone ship in the year 2015.

Section 3

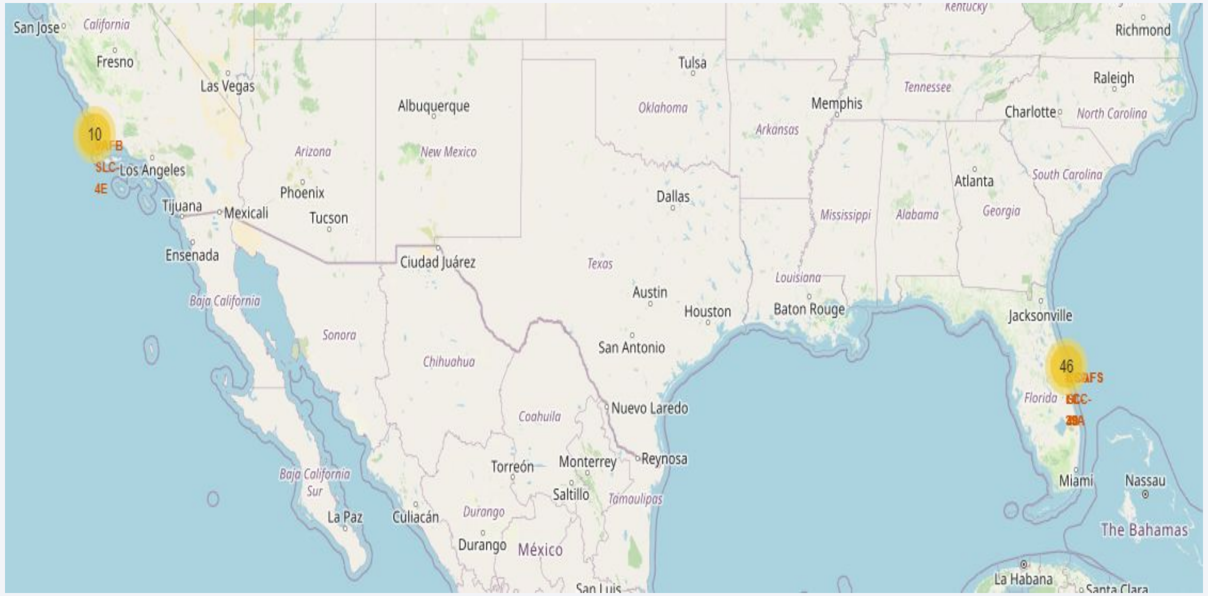
Launch Sites Proximities Analysis

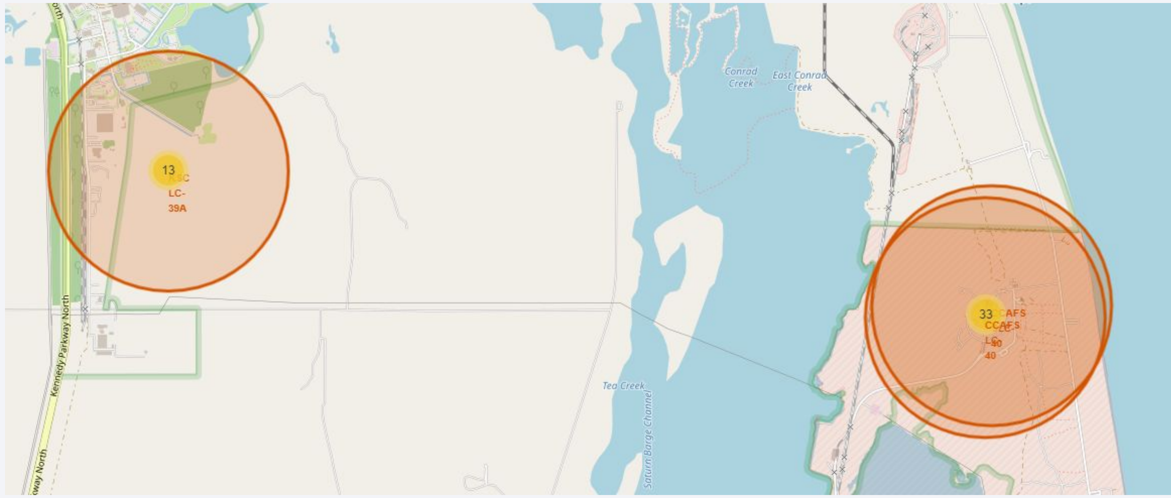
Map of Launch Sites



Here you can see the various launch sites that the Falcon 9 rockets launched from

Grouped Launch Sites





Here, we grouped the launches onto the launch locations so that we could more easily see how the outcome was related to the launch location

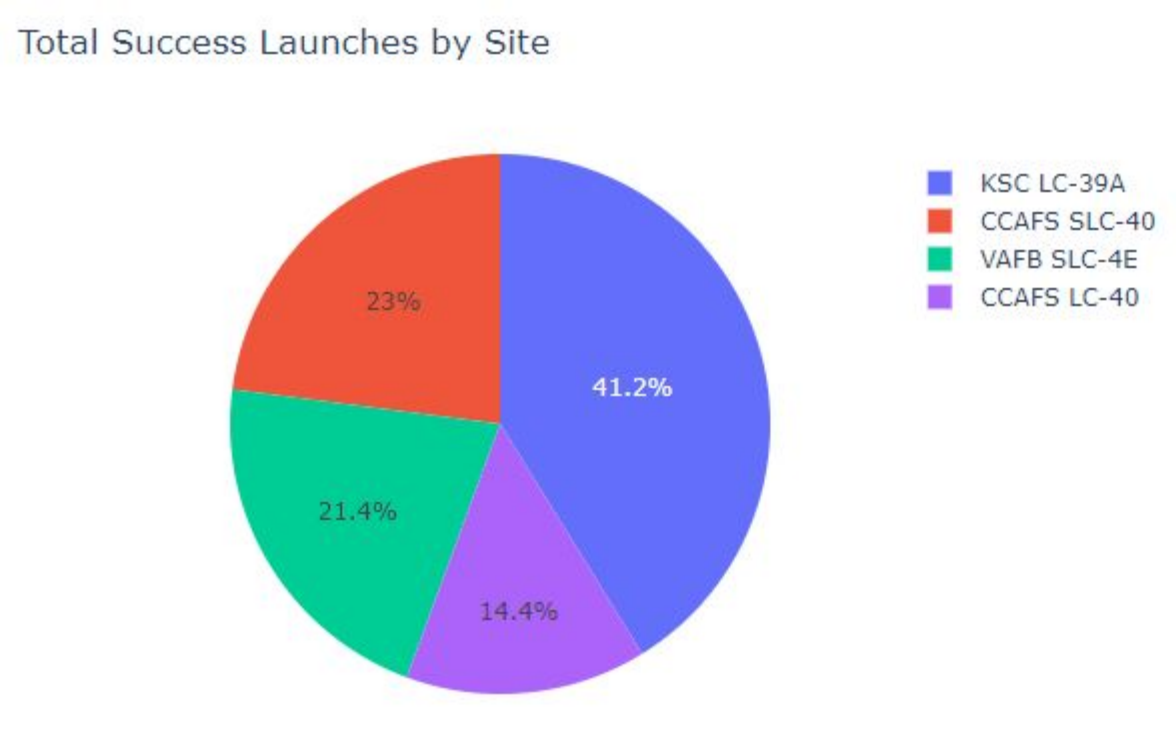
Distance from Shoreline

Finally, we colored the locations according to their success and failure rate, and then we found the distance from the nearest shore for one of the launch locations, as shown by the polyline annotation here

Section 4

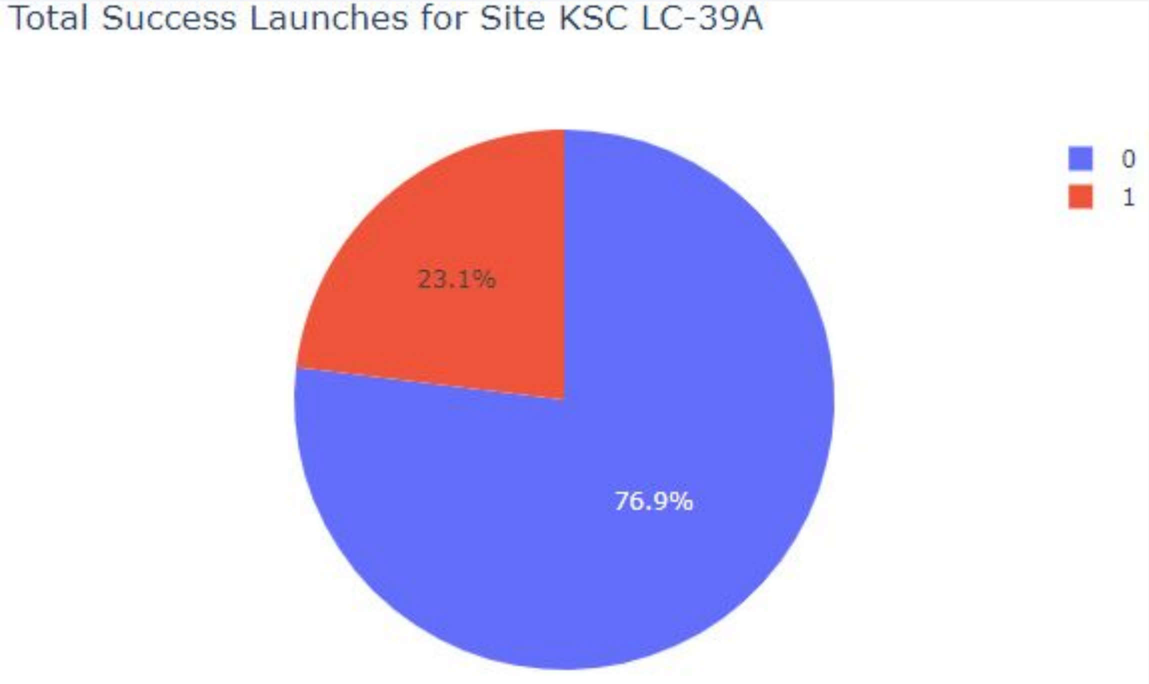
Build a Dashboard with Plotly Dash

Total Success Rate by Site



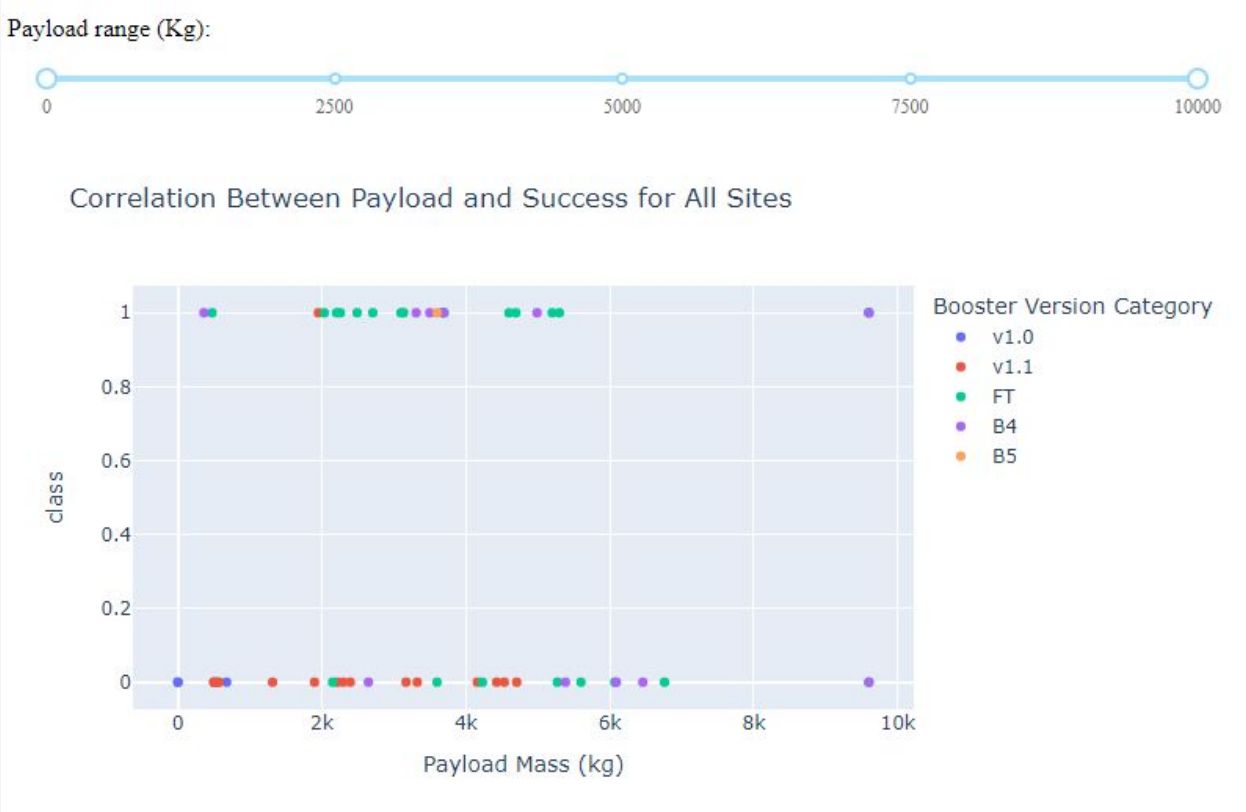
- Here, we see the success rate of various launch sites.
- Clearly, the KSC LC-39A site has the largest success rate by nearly double when compared to all the other sites.

Most Successful Launch Site



- Here is the breakdown for the launch site with the highest success rate
- Clearly, launching from this site is probably the best bet as far as saving some money on launching rockets into space

Payload Mass and Success Rate

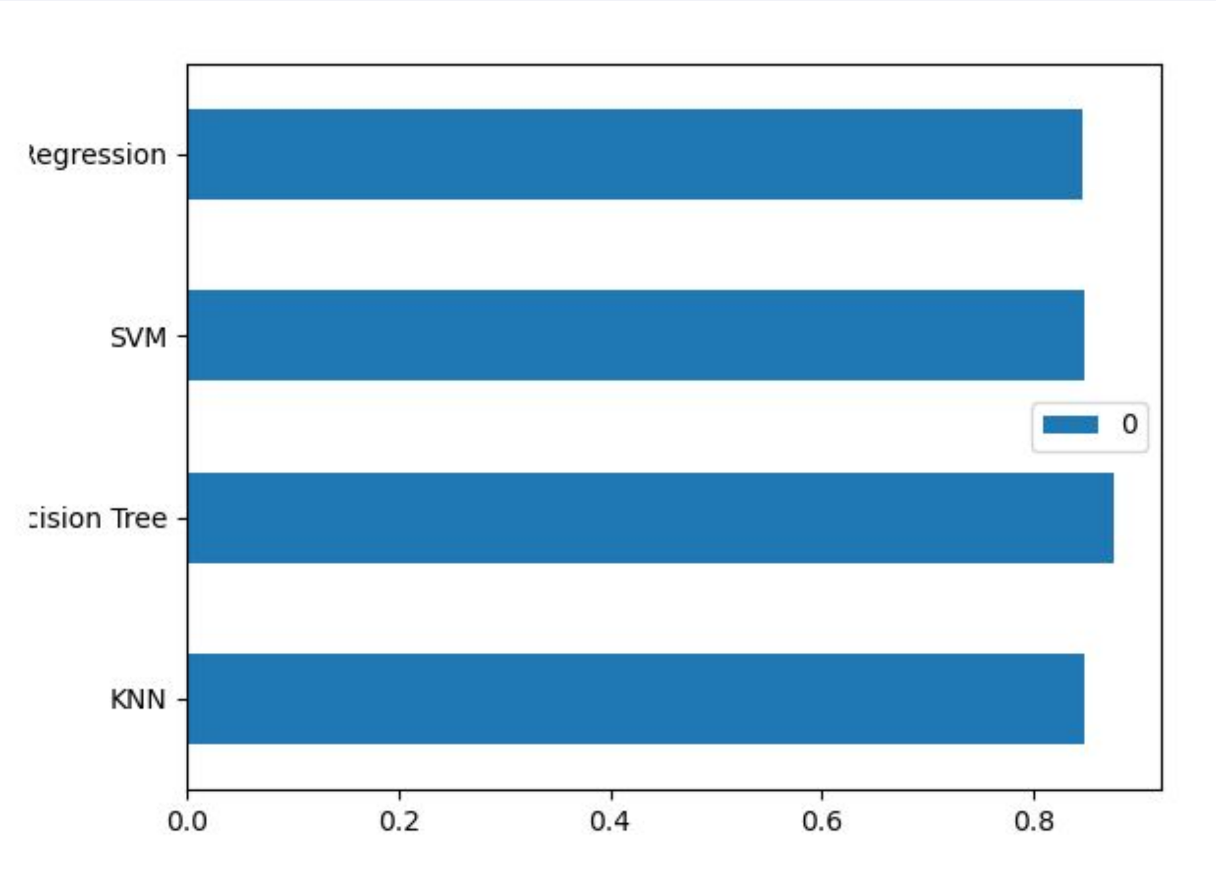


- This is a representation of the launches from all sites, the payload mass of each launch, and the success of each launch. The color of the dot indicates the specific booster version
- As mentioned before the FT booster seems to be the most successful

Section 5

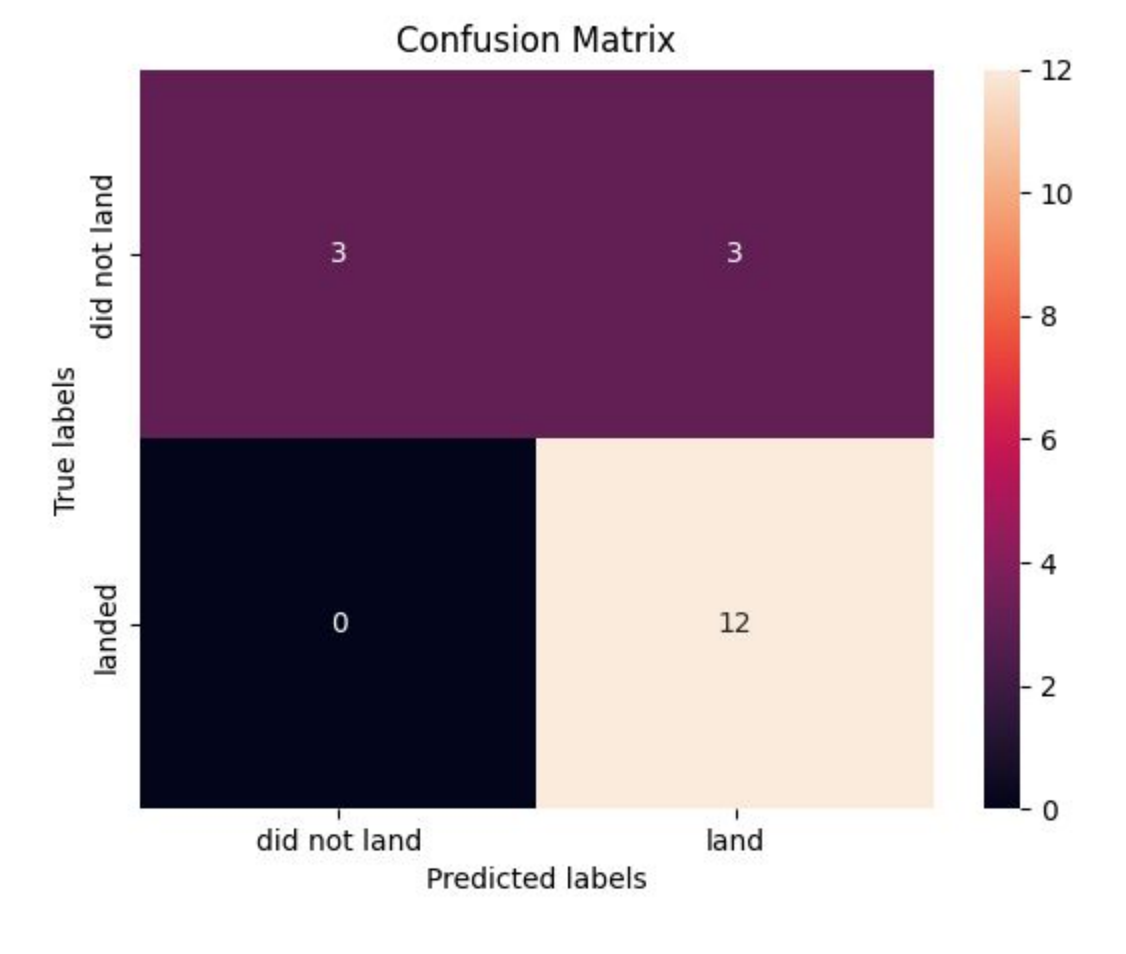
Predictive Analysis (Classification)

Classification Accuracy



Here, you can see that the Decision tree preformed the best out of all the other classification algorithms

Confusion Matrix



- Here is the confusion matrix of the Classification Tree.
- It performed very well at correctly classifying successful flights, but it was just guessing that every flight would land successfully, which was not the case

Conclusions

- From this we can see that the algorithm that performed the best on average is the one that simply guessed every flight would land successfully!
- Especially considering what we know about how launches that occurred later in time are more likely to land successfully in general, this shows that we could trust SpaceX to offer the cheapest experience in space flight on average

Thank you!