IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Evan Knep>
<January 24, 2024>

Outline

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

Executive Summary

- Here we utilize publicly available data on SpaceX rocket launches to predict booster landing success likelihoods for future launches.
- We find that although success rate has generally increased with time as the technology matures, success rate is still largely influenced by variables such as where the rocket was launched, booster type, and desired orbit. We also find that using machine learning algorithms, we can predict with high accuracy whether a given rocket launch/landing will be successful.

Introduction

- As the commercial space race continues to grow, a key component of keeping cost down is reusing the first stage of the rocket. Currently this is an area where SpaceX is far ahead of the competition. Predicting whether or not a launch will allow for the successful recovery of the first stage of the rocket allows us to predict the cost of the launch.
- We wish to determine the predictive variables that would allow us to accurately predict the success or failure of a launch/recovery.

Section 1

Methodology

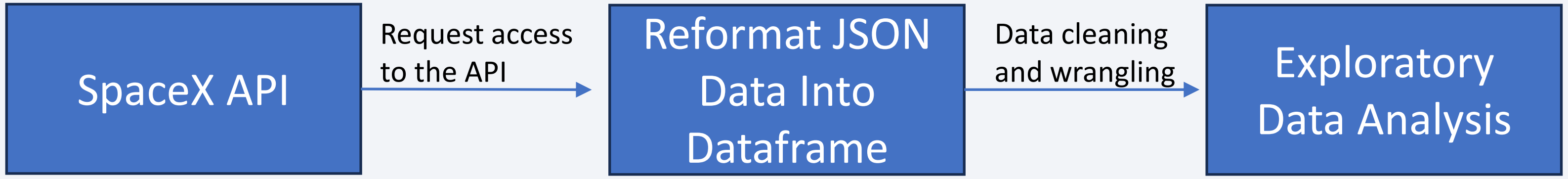
Methodology

Executive Summary

- Data collection methodology:
  - Data was collected from the SpaceX API and Wikipedia webscraping
- Perform data wrangling
  - Data was cleaned, standardized, and filtered using Python.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Logistic Regression, SVM, KNN, and decision tree algorithms were trained on SpaceX data to predict launch/landing success likelihood.

Data Collection

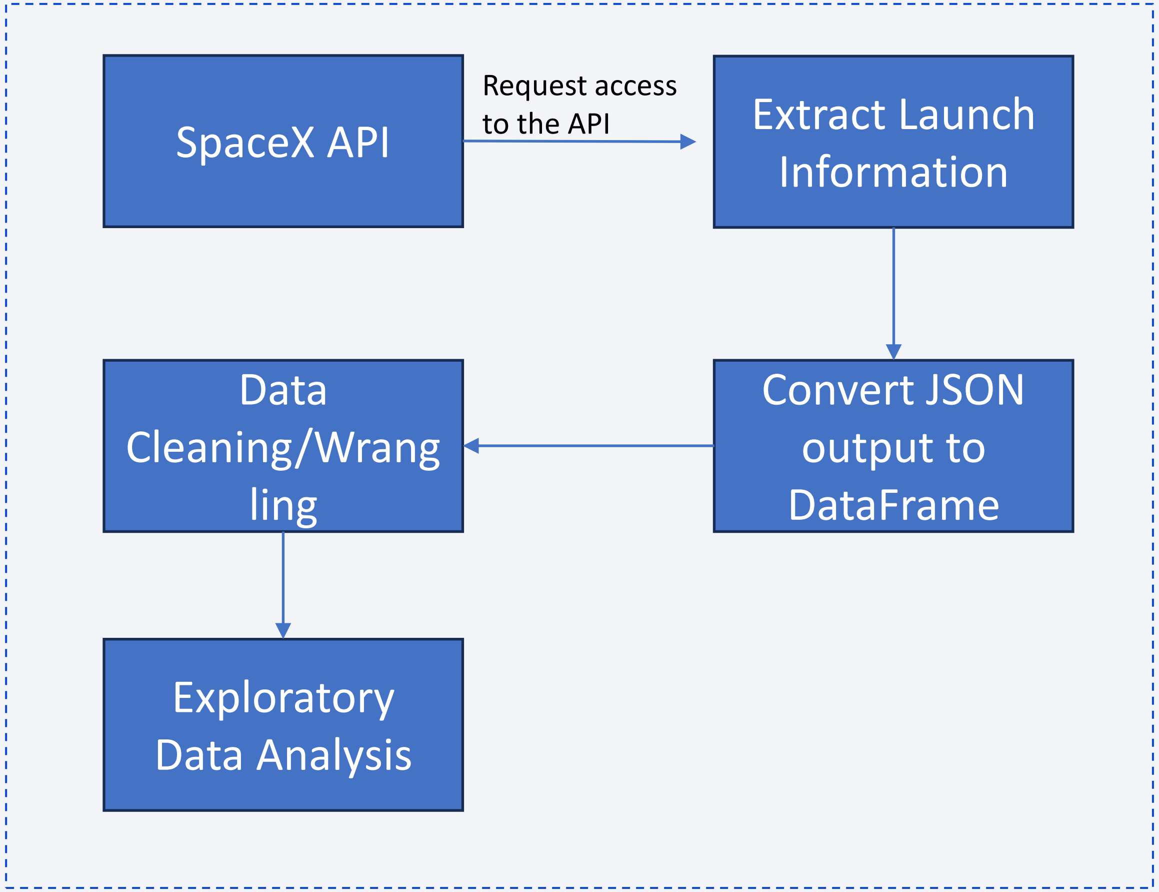
- SpaceX API was used for data collection, containing data for every SpaceX launch since the company's inception. The SpaceX API contains data on launch site, time, equipment, weight, and landing results, as well as more technical data on the rockets that was not incorporated in this project.



Data Collection – SpaceX API

- SpaceX API was used to collect and analyze SpaceX launch data over the last 10-15 years. Data was formatted into accessible dataframes for analysis and modeling.

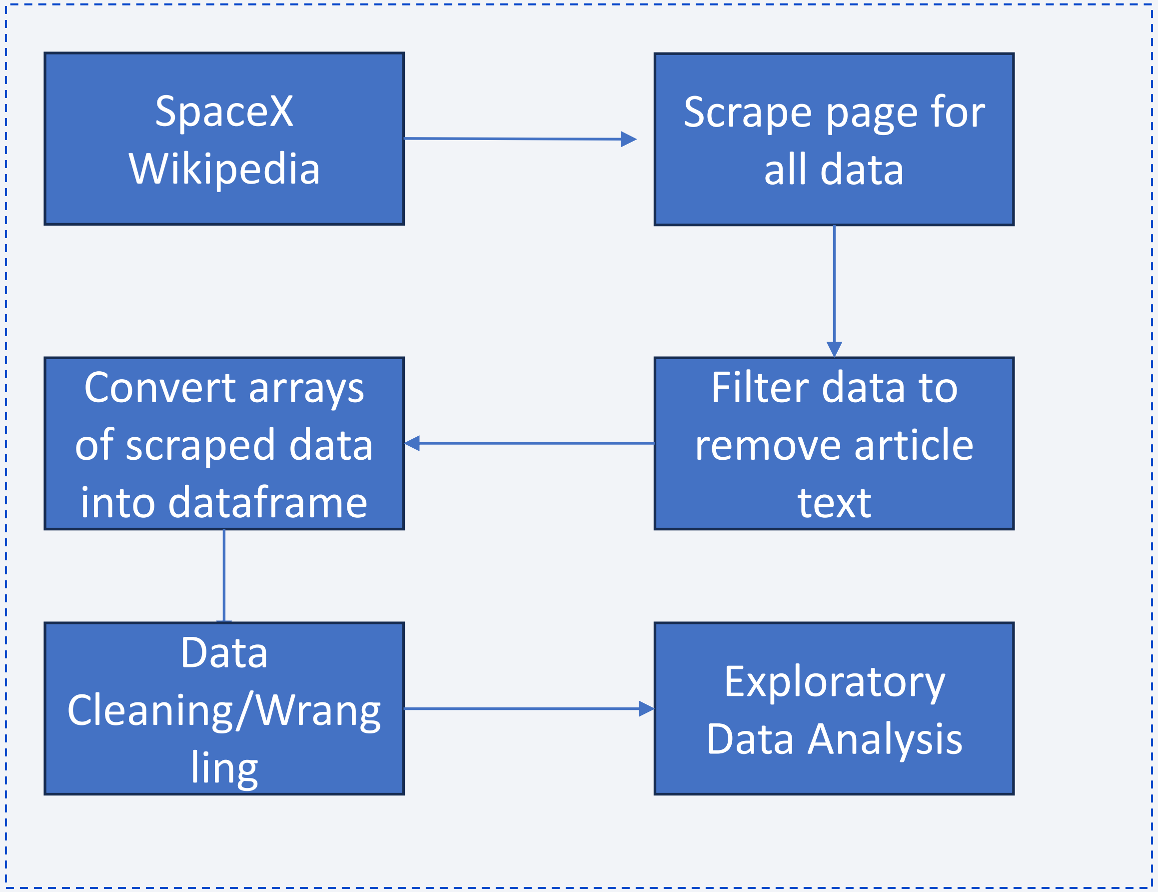
- github link to notebook



Data Collection - Scraping

- Present your web scraping process using key phrases and flowcharts

- Add the GitHub URL of the completed web scraping notebook, as an external reference and peer-review purpose



Data Wrangling

- Following collection, data was cleaned and formatted to ensure proper data types and values. Launches with no planned landing were removed from the data, and missing values for fields such as payload mass were replaced with the mean payload mass to avoid skewing the data.

Reformat JSON to Dataframe → Filter for only launches with planned landing → Remove missing values → Reassign data types prior to analysis

- [Notebook URL]

EDA with Data Visualization

- Scatterplots, histograms, and categorical plots were primarily used for data analysis to visualize correlations between launch variables and successful outcomes. Bar charts were also used to examine how launch site and orbit target related to success.
- Notebook URL

EDA with SQL

- SQL was used to examine booster specific data
- Rate of landing outcomes
- Average Payload mass for different booster types
- Booster type and landing outcome over time

- Notebook URL

Build an Interactive Map with Folium

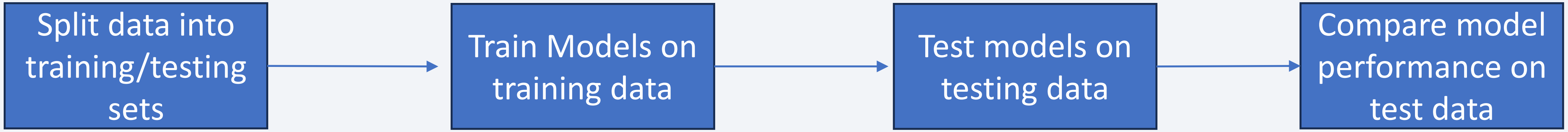
- An interactive map highlighting launch locations and outcome information was built using the Folium package to better visualize the relationship of location and potential environmental factors that may play into success likelihood.
- Notebook URL

Build a Dashboard with Plotly Dash

- Plotly Dash was used to construct a dashboard to communicate findings related to success rate by payload mass and booster type.
- Notebook URL

Predictive Analysis (Classification)

- Summarize how you built, evaluated, improved, and found the best performing classification model



- Notebook URL

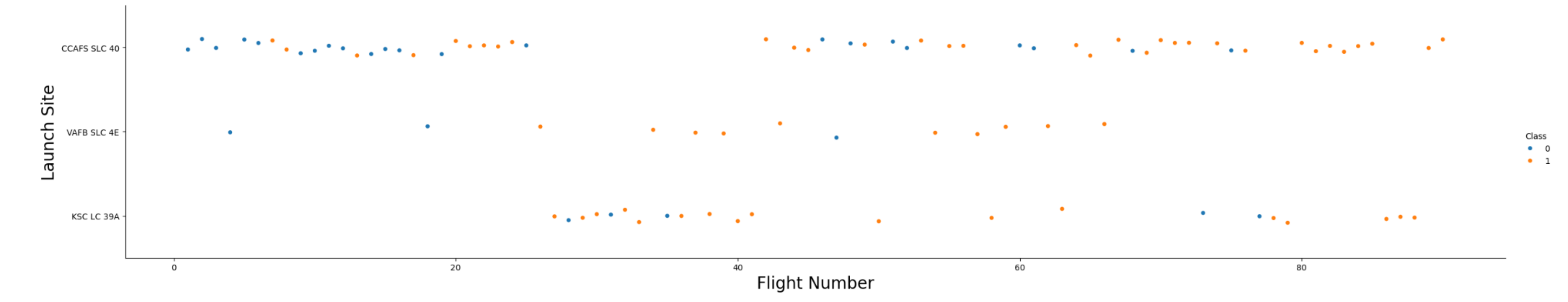
Results

- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

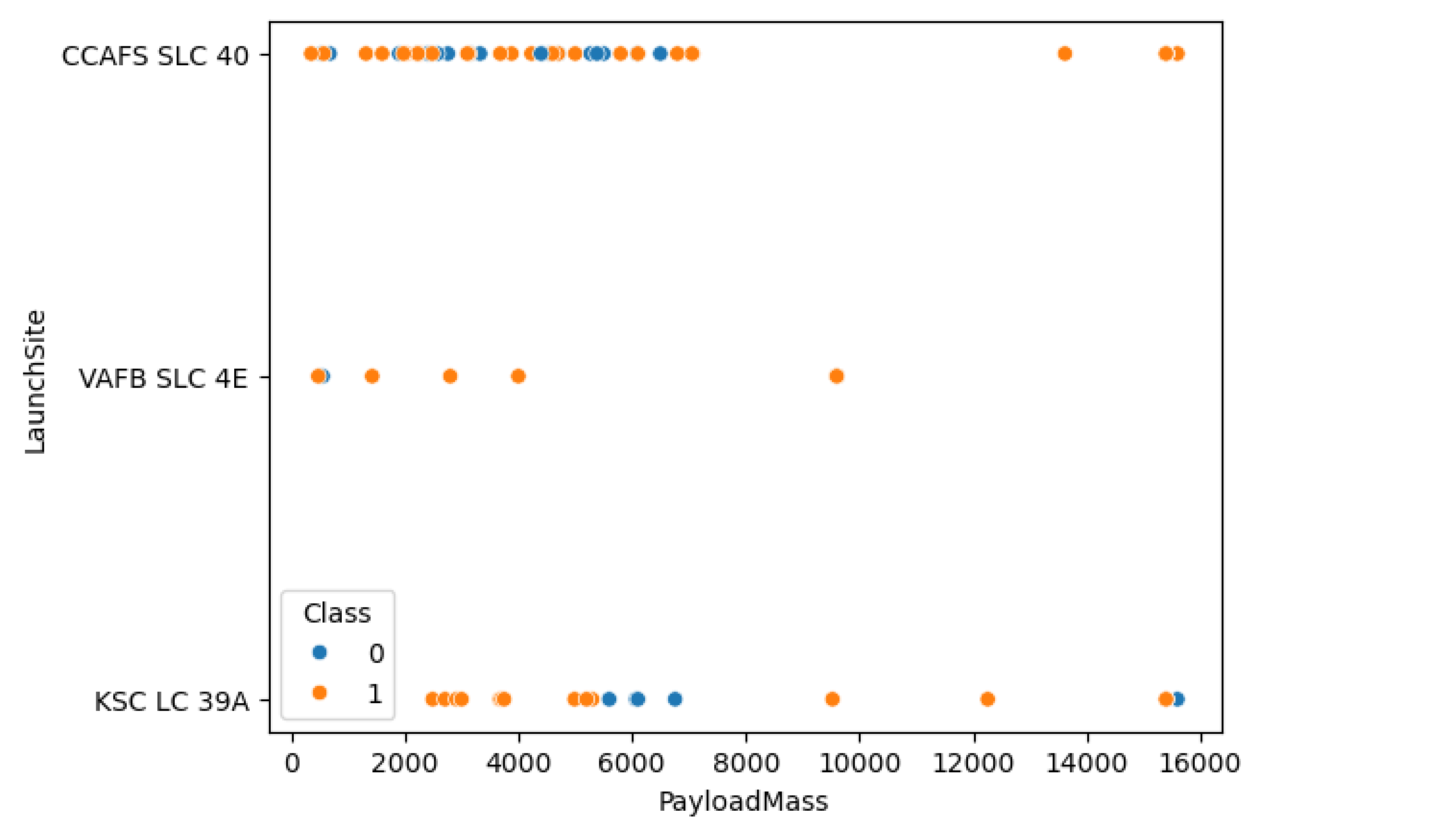
Insights drawn from EDA

Flight Number vs. Launch Site



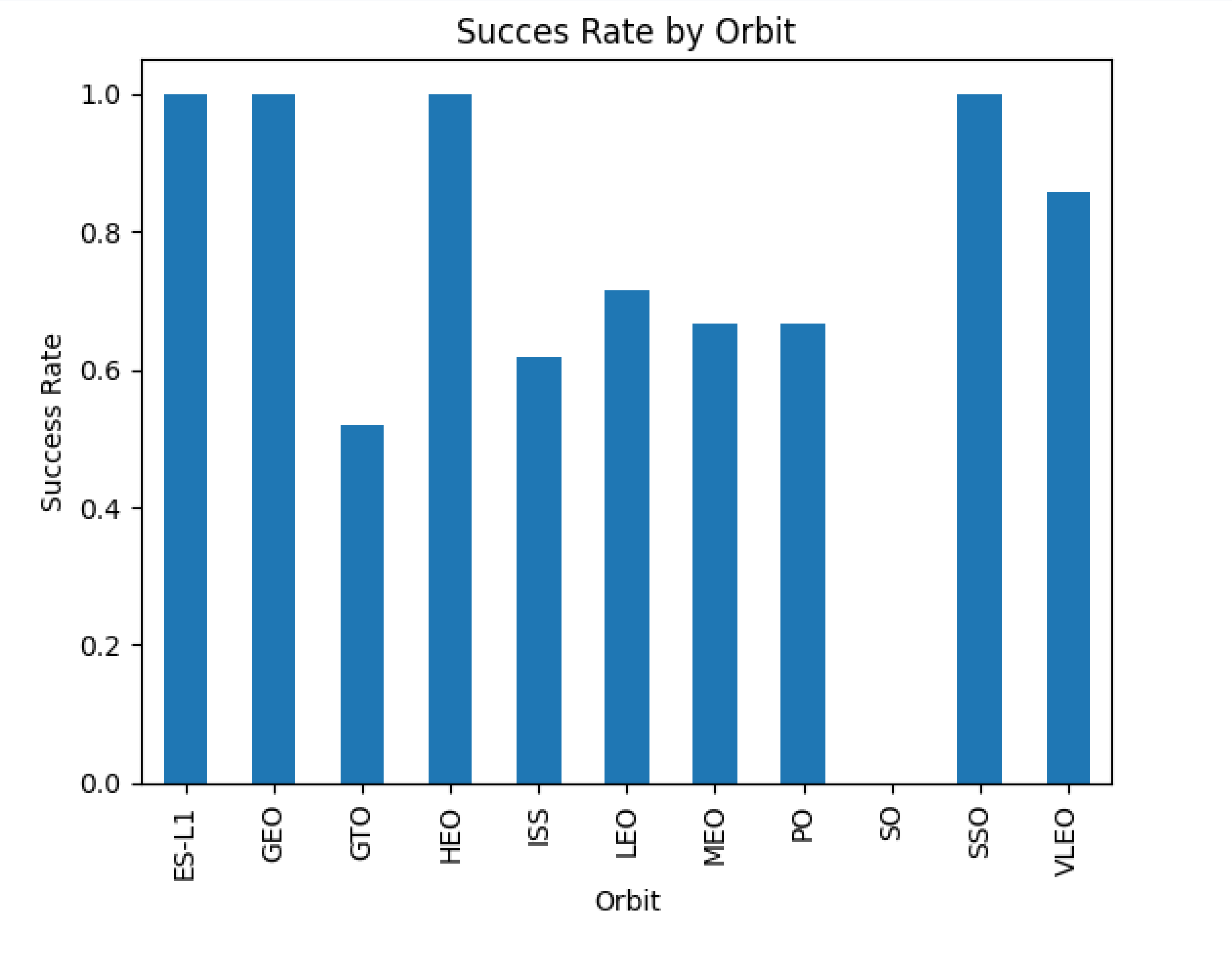
- We can see that as the number of flights increased, the success rate (orange = success) increased. We can also see that site CCAFS SLC-40 had the greatest number of launches and successful launches

Payload vs. Launch Site



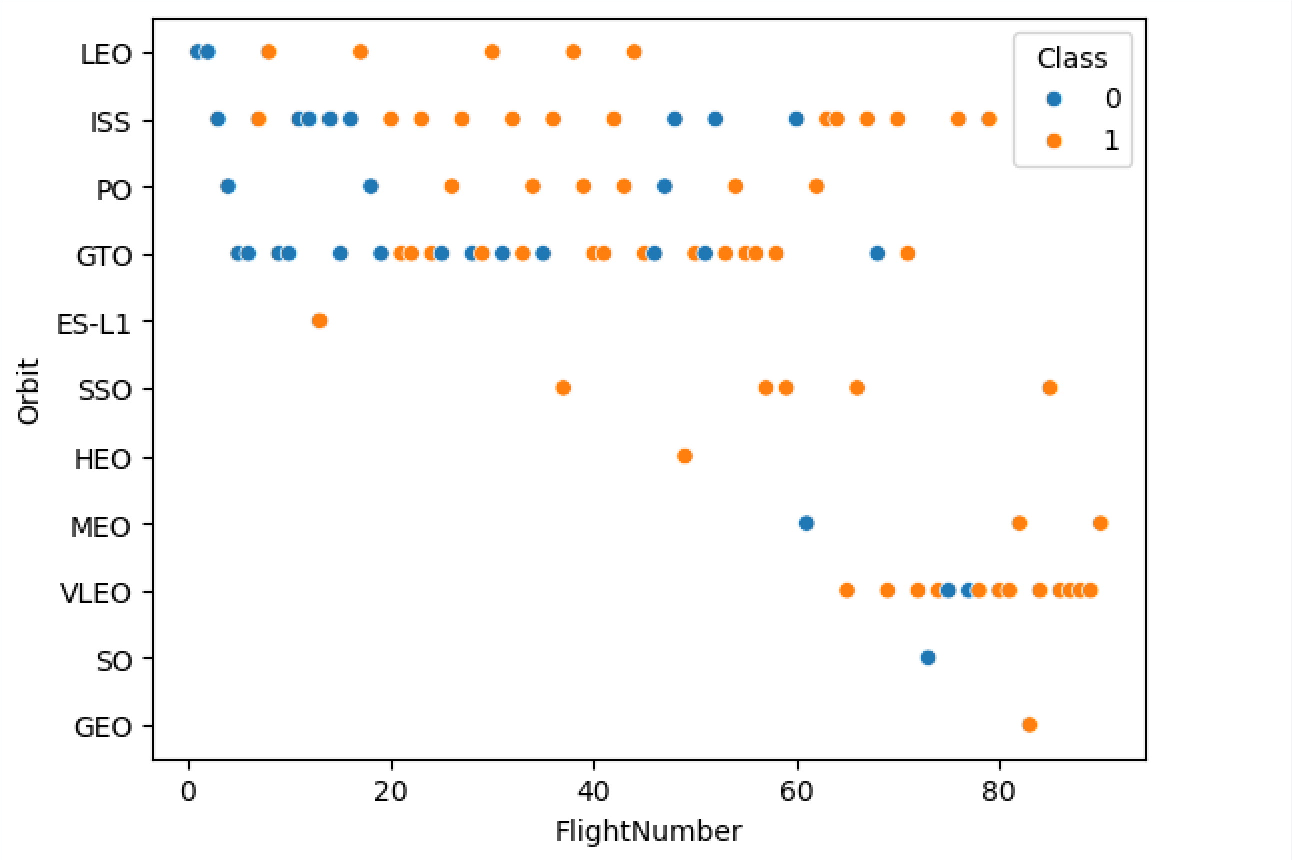
- We can also see that different payload sizes where used at different launch sites, with VAFB SLC 4E being the only site with no launches over 10000kg

Success Rate vs. Orbit Type



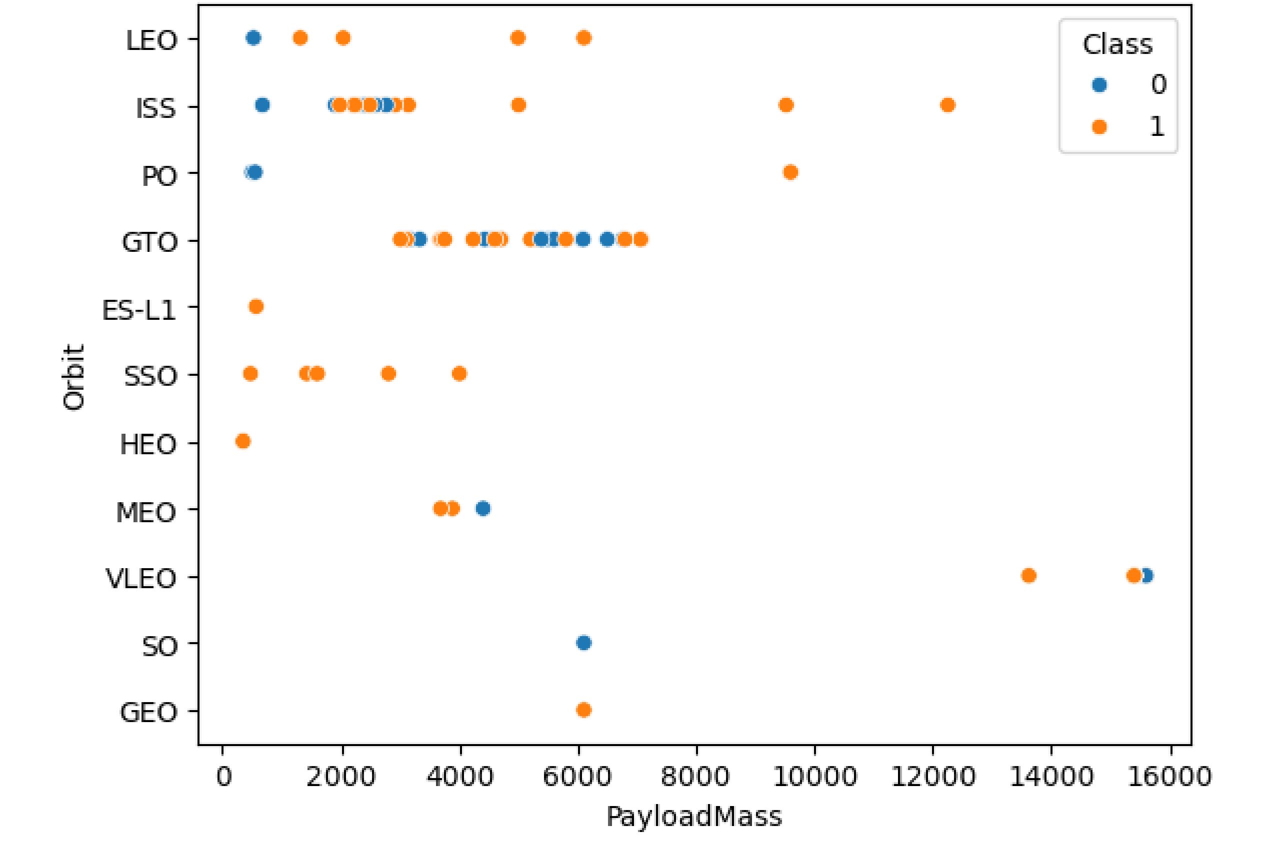
Success rate is largely dependent on desired orbit for the launch

Flight Number vs. Orbit Type



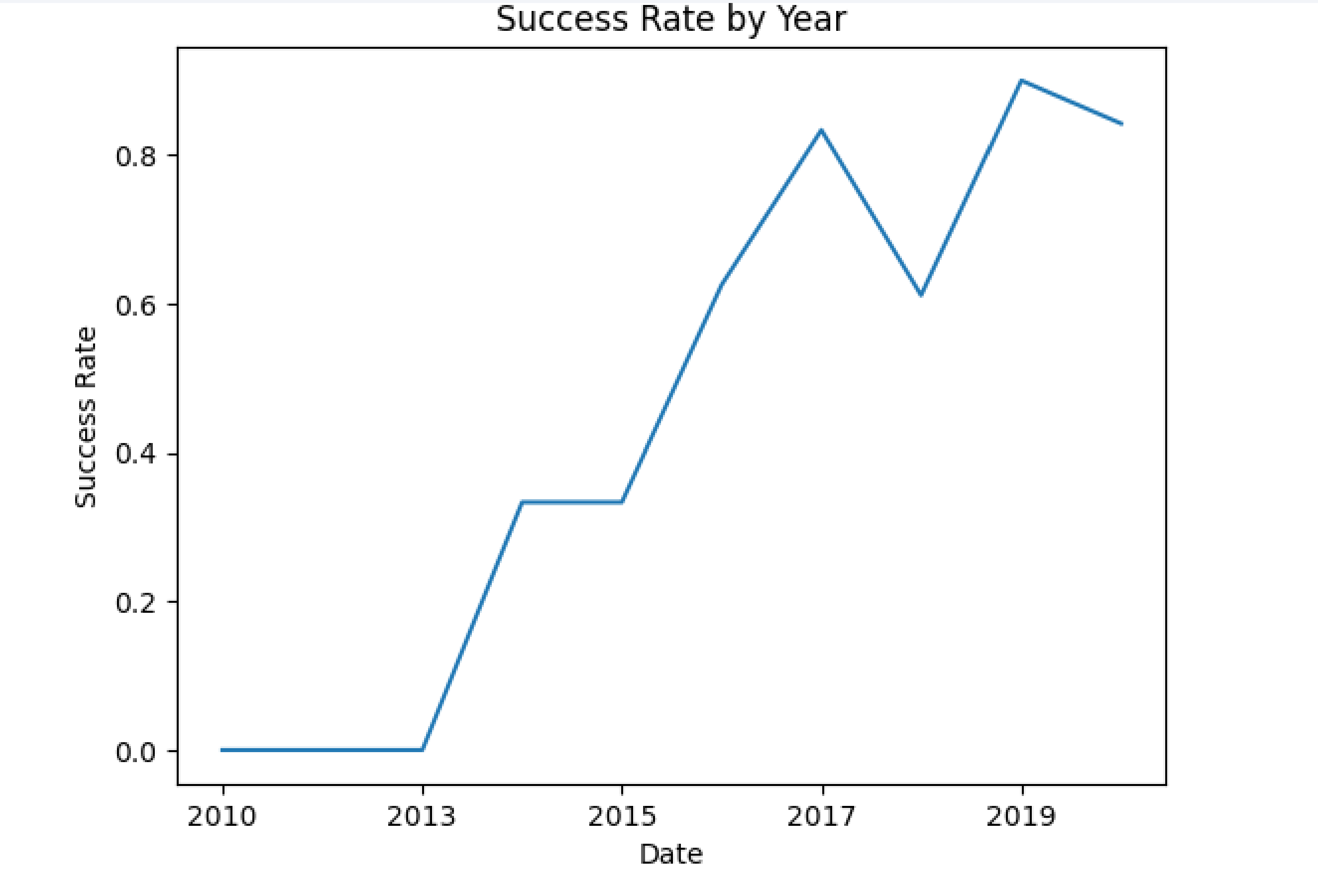
Interestingly, we can see that over time the orbit type shifted dramatically, with a greater number of launches shifting towards orbits with higher success rates and riskier orbits being filtered out

Payload vs. Orbit Type



- Payload mass seemed to have very little influence on orbit type

Launch Success Yearly Trend



Success rate increased dramatically over time

All Launch Site Names

	Launch Site	Lat	Long
0	CCAFS LC-40	28.562302	-80.577356
1	CCAFS SLC-40	28.563197	-80.576820
2	KSC LC-39A	28.573255	-80.646895
3	VAFB SLC-4E	34.632834	-120.610745

- Limited number of launch sites in use for SpaceX, resulting in a variety of launch types at each location

Launch Site Names Begin with 'CCA'

```
%sql select * from SPACEXTABLE where Launch_Site like 'CCA%' limit 5
```

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

- Here we can see the variety of launch types from just one site

Total Payload Mass

```
[10]: %sql select sum(PAYLOAD_MASS__KG_) from SPACEXTABLE where Customer like '%NASA%'

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

[10]:

sum(PAYLOAD_MASS__KG_)
107010

As a whole, NASA launches have carried over 100,000kg

Average Payload Mass by F9 v1.1

Display average payload mass carried by booster version F9 v1.1

```
%sql select avg(PAYLOAD_MASS__KG_) from SPACEXTABLE where Booster_Version like '%F9 V1.1%'
```

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

avg(PAYLOAD_MASS__KG_)
2534.6666666666665

Average payload mass carried by F9 v1.1s was 2534.67 kg

First Successful Ground Landing Date

```
%sql select min(date) from SPACEXTABLE where Landing_Outcome like '%SUCCESS%'
```

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

min(date)
2015-12-22

The first successful ground landing for SpaceX was on 12-22-2015

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%sql select distinct(Booster_Version) from SPACEXTABLE where PAYLOAD_MASS__KG_ between 4000 and 6000 and Landing_Outcome like '%Success (Drone Ship)%'

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

Booster_Version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

The only successful drone ship landings with payload between 4000-6000kg utilized the Boosters listed above

Total Number of Successful and Failure Mission Outcomes

List the total number of successful and failure mission outcomes

```
]: %sql select Landing_Outcome, count(*) as num from SPACEXTABLE group by Landing_Outcome order by num desc
```

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

Landing_Outcome	num
Success	38
No attempt	21
Success (drone ship)	14
Success (ground pad)	9
Failure (drone ship)	5
Controlled (ocean)	5
Failure	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1
No attempt	1

Total number of successful and unsuccessful mission outcomes, ranked by frequency

Boosters Carried Maximum Payload

```
%sql select distinct(Booster_Version) from SPACEXTABLE where PAYLOAD_MASS__KG_ = (select max(PAYLOAD_MASS__KG_) from SPACEXTABLE)
```

```
 * 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

Boosters listed here have carried the maximum payload mass recorded by SpaceX

2015 Launch Records

```
%sql select substr(Date, 6,2) as month, substr(Date, 0,5) as year, Booster_Version, Launch_Site, Landing_Outcome from SPACEXTABLE where year = '2015' and Landing_Outcome = 'Failure (drone ship)'
```

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

month	year	Booster_Version	Launch_Site	Landing_Outcome
01	2015	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
04	2015	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

There were only two Booster Versions that were used in 2015 that resulted in failed launches on drone ships

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(*) as num from SPACEXTABLE where Date between '2010-06-04' and '2017-03-20' group by Landing_Outcome order by num desc
```

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

Landing_Outcome	num
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

Landing outcomes between 2010 and early 2017 are listed here, and we can see that results were very mixed.

Section 3

Launch Sites Proximities Analysis

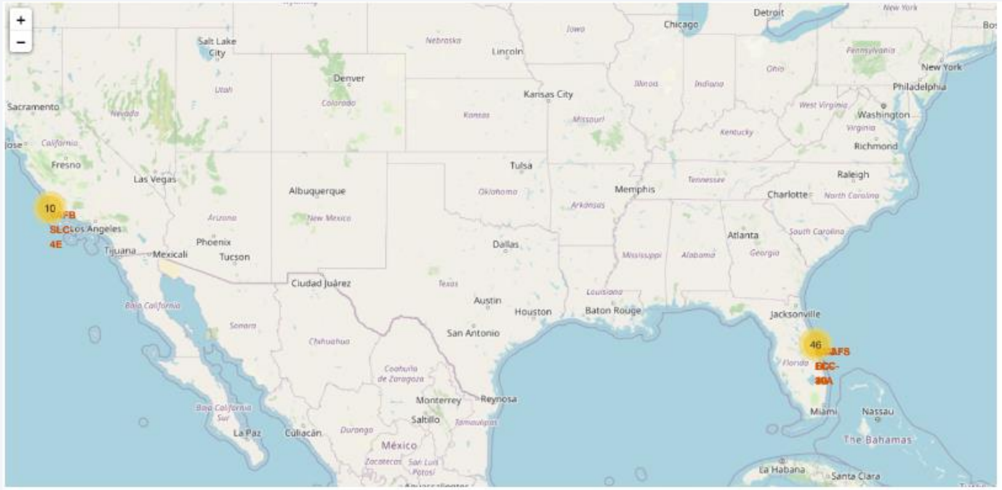
SpaceX Launch Site Locations

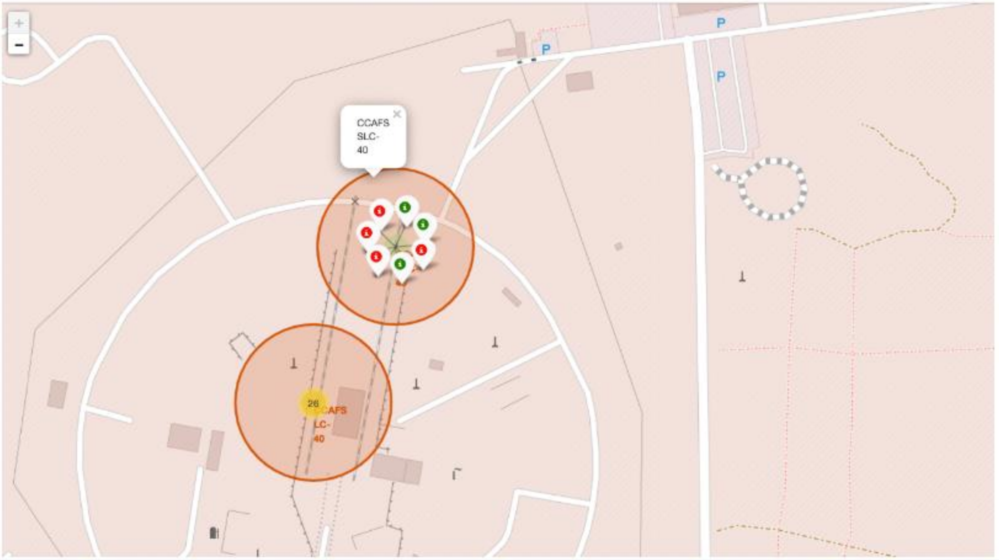
- As a US based company SpaceX only launches from the United States, but we can see that they launch at several coastal locations and no inland locations.

Mapping Outcome By Location

- We can see number of launches and launch success on our Folium visualizations

- (it wasn't letting me trust the notebook so I couldn't display my own version of this even with the proper code. These are the example figures)





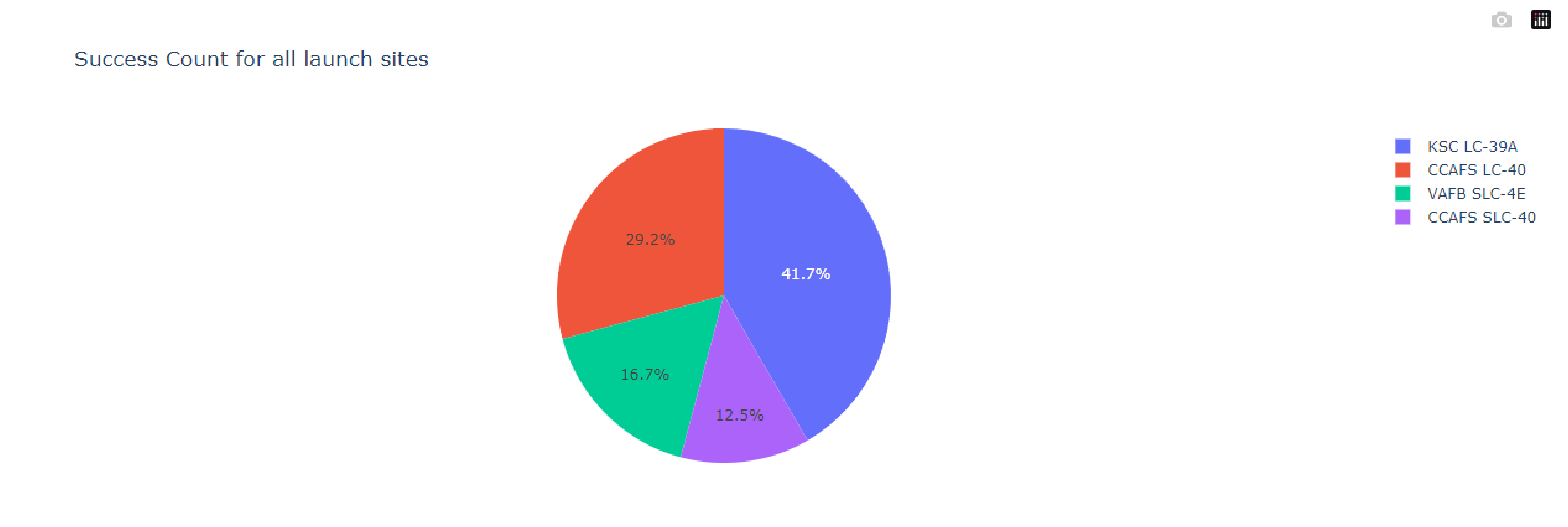
Coastal Proximity of Launch Sites

This example is exemplary of all SpaceX launch sites, with extremely close proximity to the coast. In this case less than 1 km

Section 4

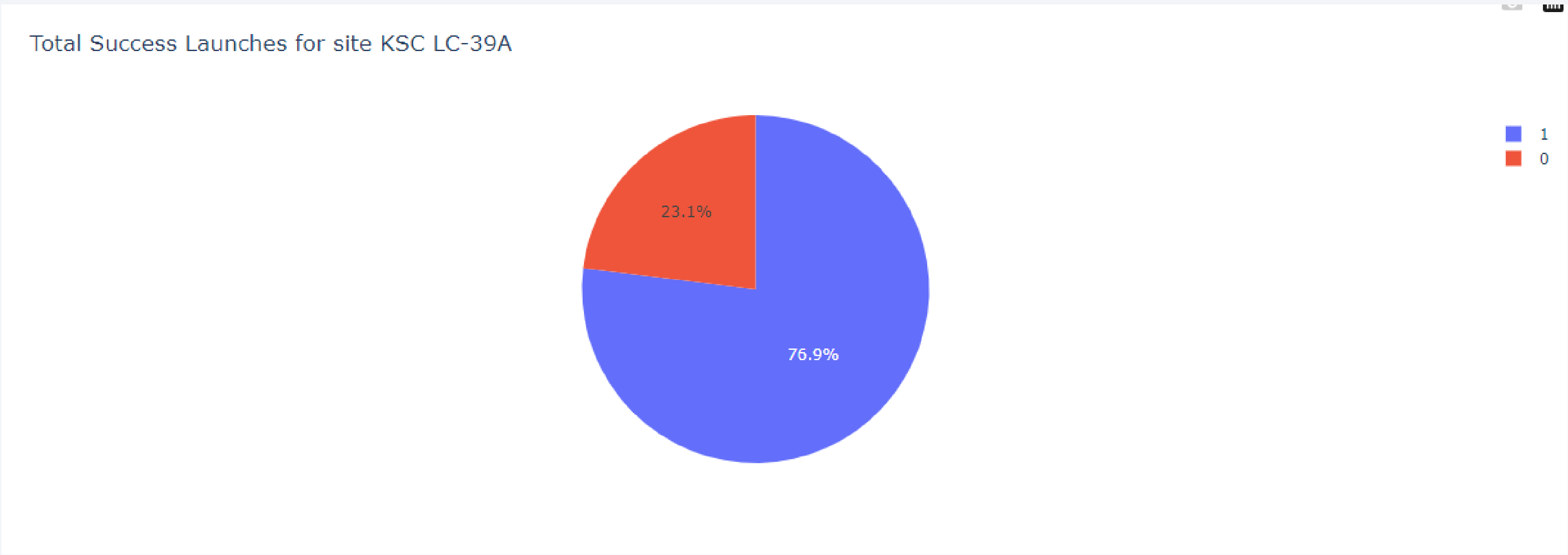
Build a Dashboard with Plotly Dash

Landing Success By Launch Site



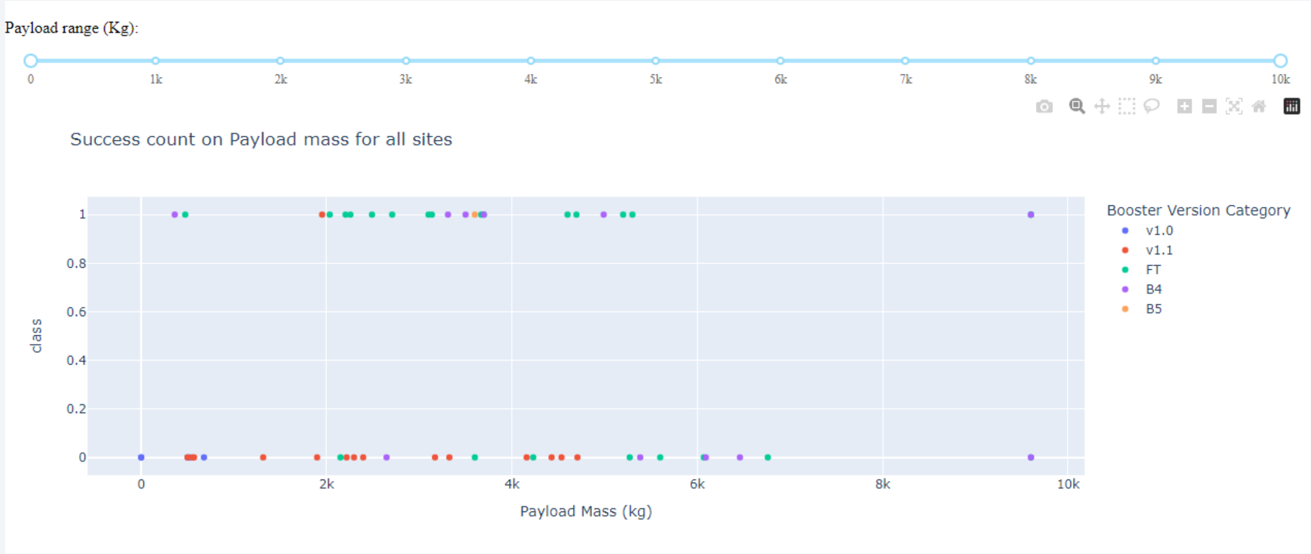
KSC LC-39A showed significantly higher landing success rate than other launch sites

<Dashboard Screenshot 2>



Here we can see that site KSC LC-39A has a 76.9% success rate

<Dashboard Screenshot 3>

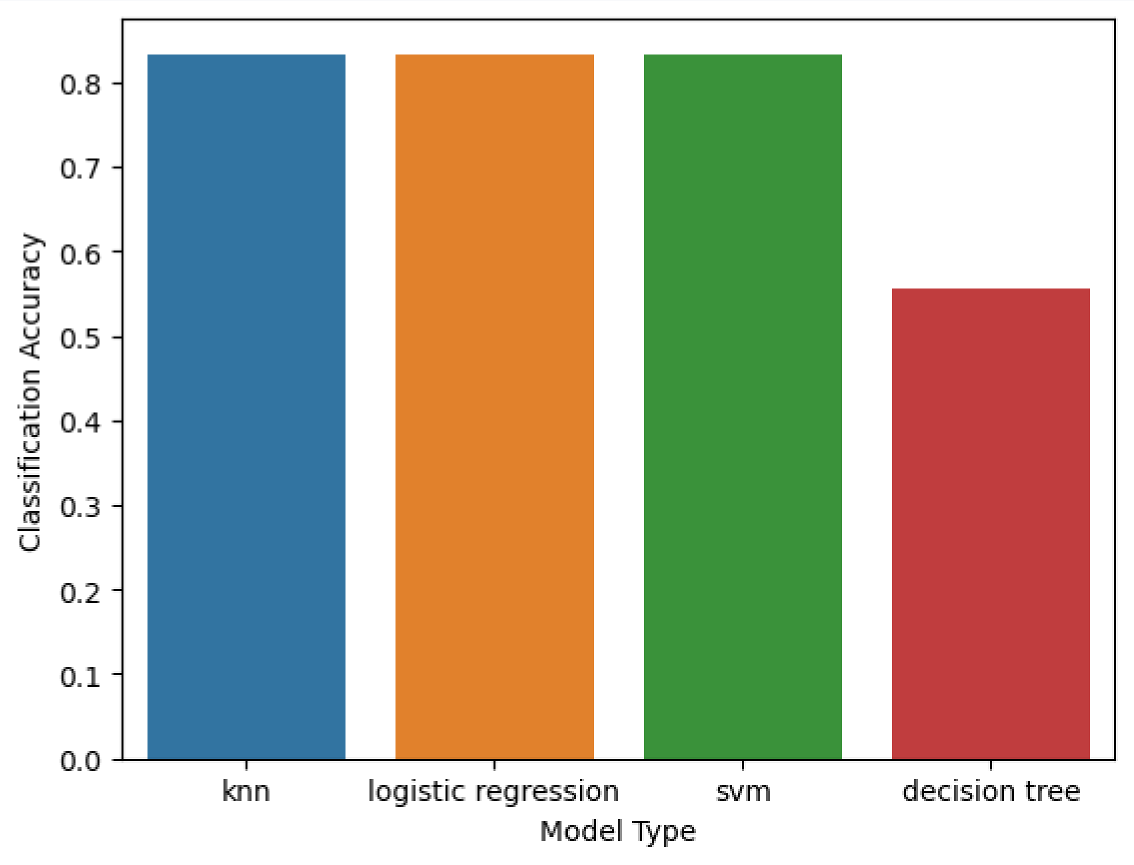


Here we can see how payload mass range influences success rate

Section 5

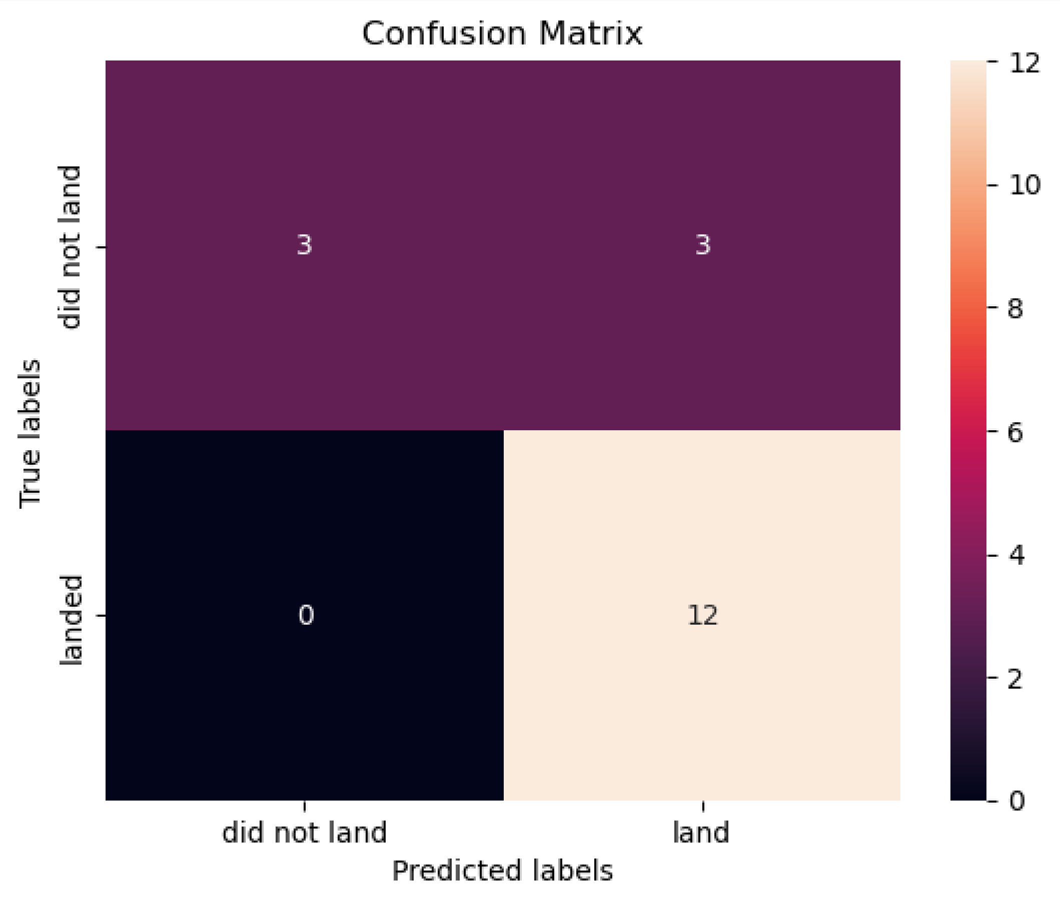
Predictive Analysis (Classification)

Classification Accuracy



- We find nearly identical results for KNN, logistic regression, and SVM, with decision trees performing notably worse than the rest

Confusion Matrix



- Here we have an example confusion matrix for our KNN classifier, where we can see the rate of true and false positives.

Conclusions

- SpaceX landing success rate has gone up as a product of time
- The landing success rate seems to be dependent on launch location, orbit type, and booster type
- Success rate does not appear to be tied to payload mass
- Our models can predict the success of a launch with upwards of 83% accuracy, with KNN, SVM, and a logistic regression model outperforming decision trees.

Appendix

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

Thank you!