Winning Space Race with Data Science

Vince Conca
3/27/2024

Outline

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

Executive Summary

- Methodology
  - Data collection
  - Data wrangling
  - Exploratory data analysis (EDA)
  - Dashboard creation
  - Predictive modeling
- Results
  - Overall success rate for the Falcon 9 was 67% allowing the option to reuse the first stage of the rocket 2 out of every 3 launches.

Introduction

- The commercial space age is here, companies are making space travel far more affordable.
- One of the most successful is SpaceX. SpaceX's accomplishments include: Sending spacecraft to the International Space Station. And Starlink, a satellite internet constellation providing satellite Internet access.
- SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upwards of 165 million dollars each.
- Much of the savings is because SpaceX can reuse the first stage.
- Therefore, if we can determine if the first stage will land, we can determine the cost of a launch.
- This project aims to predict if the first stage will successfully land and be available for reuse, thus dramatically impacting the cost of a rocket launch.

Section 1

Methodology

Methodology

Executive Summary

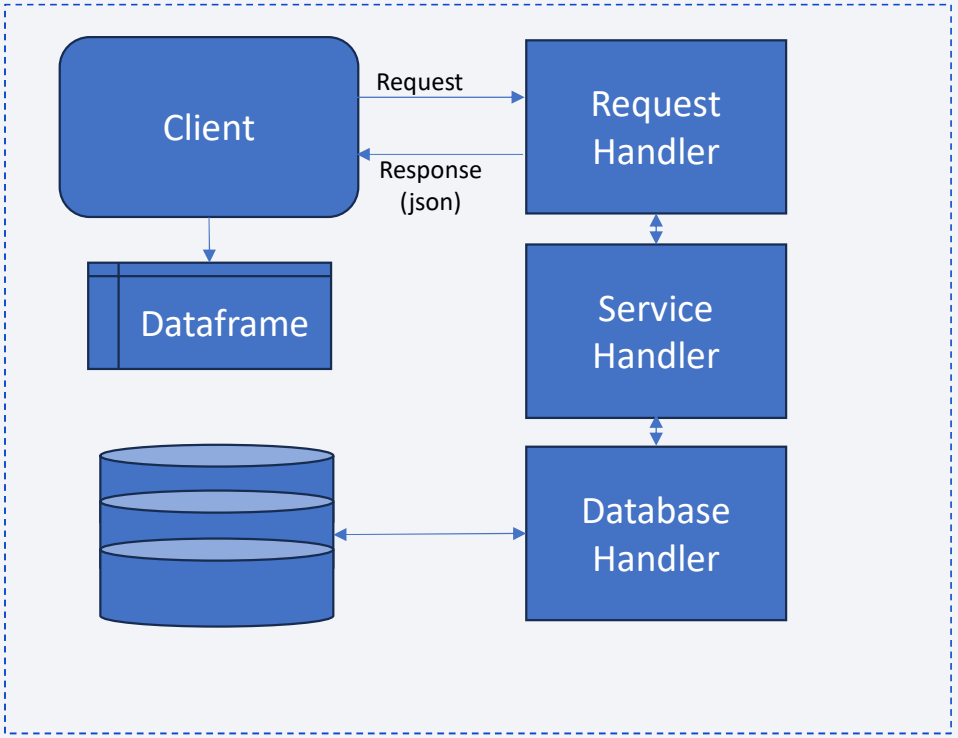
- Data collection methodology:
  - API calls, Web scraping
- Perform data wrangling
  - Broke data down into a relevant features set and a target set of success/failure
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Built 4 models, tuning each for optimal parameters, then reviewed accuracy and confusion matrix for each model

Data Collection

- Two methods used
  - SpaceX launch data REST API pulls general data regarding launches. That dataset is combined with multiple other endpoints:
    - Rocket
      - Adds specific data pertaining to rocket types
    - Payload
      - Adds payload mass and orbit type
    - Cores
      - Adds outcome of the landing, type of landing, number of flights, the landing pad used, the block of the core (which is a number used to separate version of cores), number of times this specific core has been reused, and the serial of the core.
  - Web Scraping from SpaceX Wikipedia

Data Collection – SpaceX API

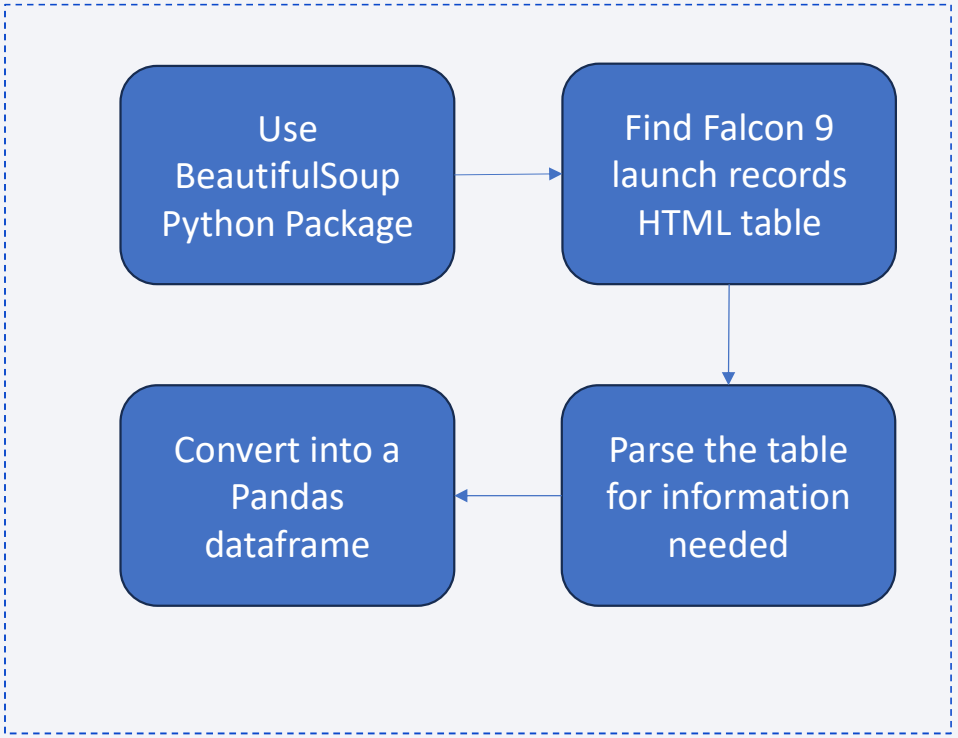
- Parent SpaceX REST call URL: https://api.spacexdata.com/v4
  - Endpoints used:
    - /launches/past
    - /rockets
    - /launchpads
    - /payloads
    - /cores
- GitHub URL of completed SpaceX API calls notebook here.



Data Collection - Scraping

- Web scraping Falcon 9 and Falcon Heavy Launches Records from Wikipedia.

- GitHub URL of completed Web Scraping notebook here.



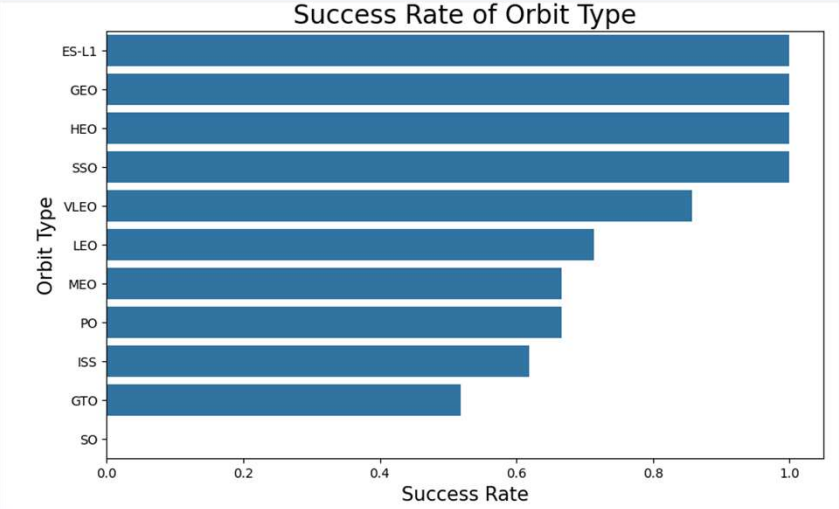
Data Wrangling

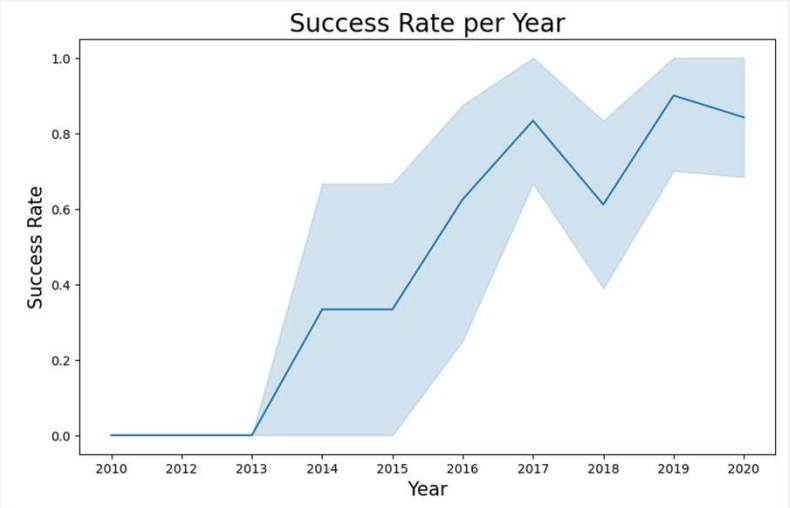
- Reviewed the launch sites and number of launches per site
  - Majority (55) took place in Florida at site: CCAFS SLC 40
- Reviewed all orbit types and counts for each orbit
  - GTO was the favored orbit type with 27 launches of that type.
    - GTO = A geosynchronous orbit is a high Earth orbit that allows satellites to match Earth's rotation
- Reviewed launch outcomes
  - Majority (41) were successfully landed to a drone ship
  - Overall success rate was 67%
- Prepared data for modeling
  - Created a 'Class' column using 1 for success and 0 for fail

- GitHub URL of completed Wrangling notebook here.

EDA with Data Visualization

- Visualizing the Orbit types and launch outcomes over time can be seen below.
- The below visualizations can be found within this Github notebook.





EDA with SQL

- SQL queries were ran to gather further insights. Some highlights below.
  - Launch site names
  - Payload mass carries by boosters launched by NASA
  - Average payload mass for specific boosters
  - Date of first successful launch
  - Total number of successful outcomes
  - Boosters carrying max payload
  - Counts of landing outcomes

- To see the actual queries and results, plus all other queries, refer to this notebook on Github.

Build an Interactive Map with Folium

- Folium was used to plot launch sites and launch outcomes.
  - Markers and marker clusters were used to identify the criteria.
    - Launch locations included three in Florida and one in California.
    - Launch sites were very near the coast and away from major cities.
  - Distance to Miami was calculated to show the distance of a launch site to a major city.
    - A line on the map indicates the relationship.

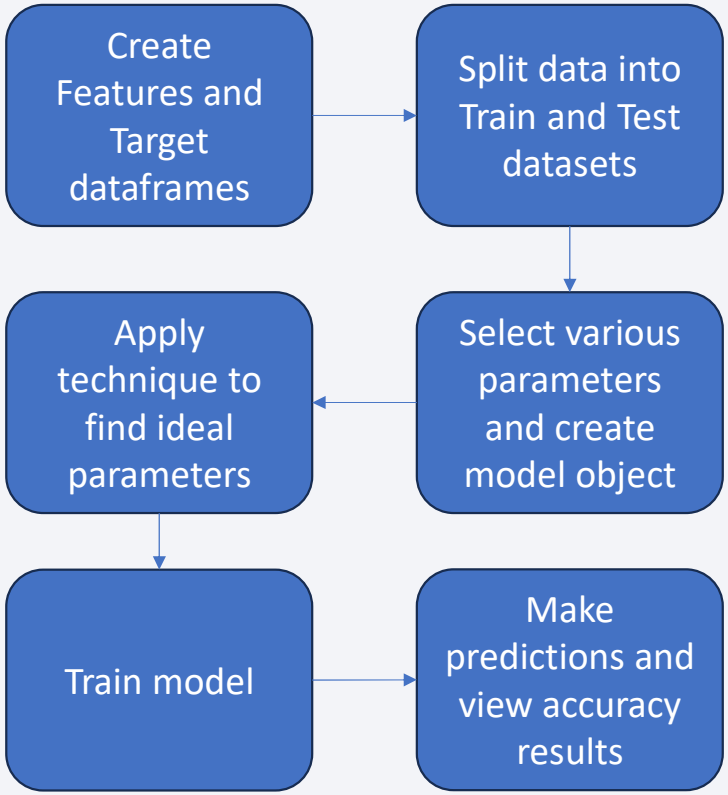
- To view the Folium visualizations, a notebook on Github found here can be reviewed.

Build a Dashboard with Plotly Dash

- A Dashboard was created allowing a user to select a specific site or all launch sites along with payload masses as these two criteria carried a lot of weight in launch success.
- The user could then view the outcomes of launches based on selected criteria.
- The dashboard proved useful in viewing which combinations provided the best launch results.
- The notebook containing the python code and the python executable file can be found here.

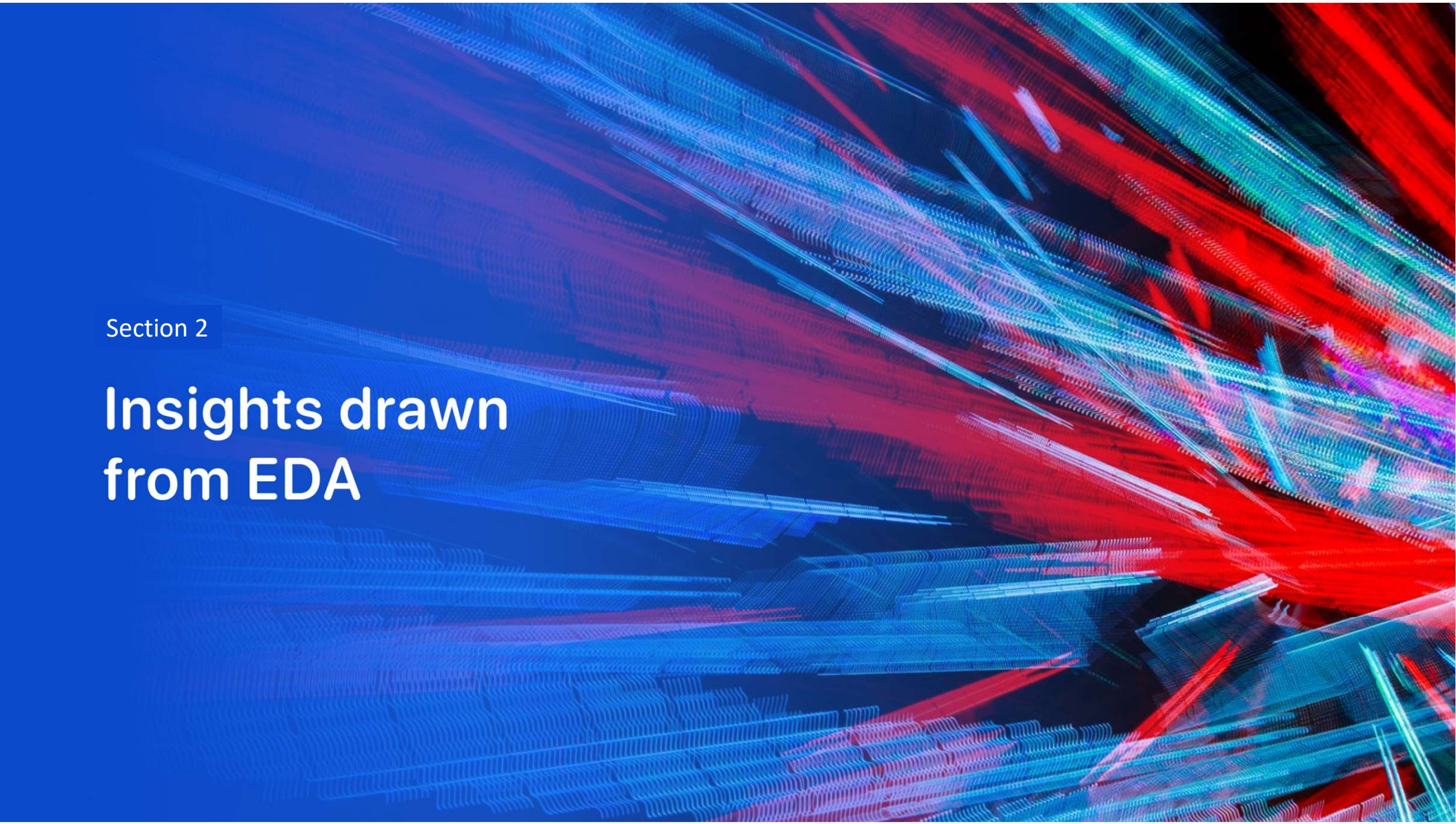
Predictive Analysis (Classification)

- Four models were built: Linear Regression, SVM, Decision Tree, KNN
- Each model used a technique to find ideal parameters to train the model.
- Predictions were made and accuracy calculated.
- Each model was then compared against each other to identify the optimal model.
- The notebook on Github can be found here.



Results

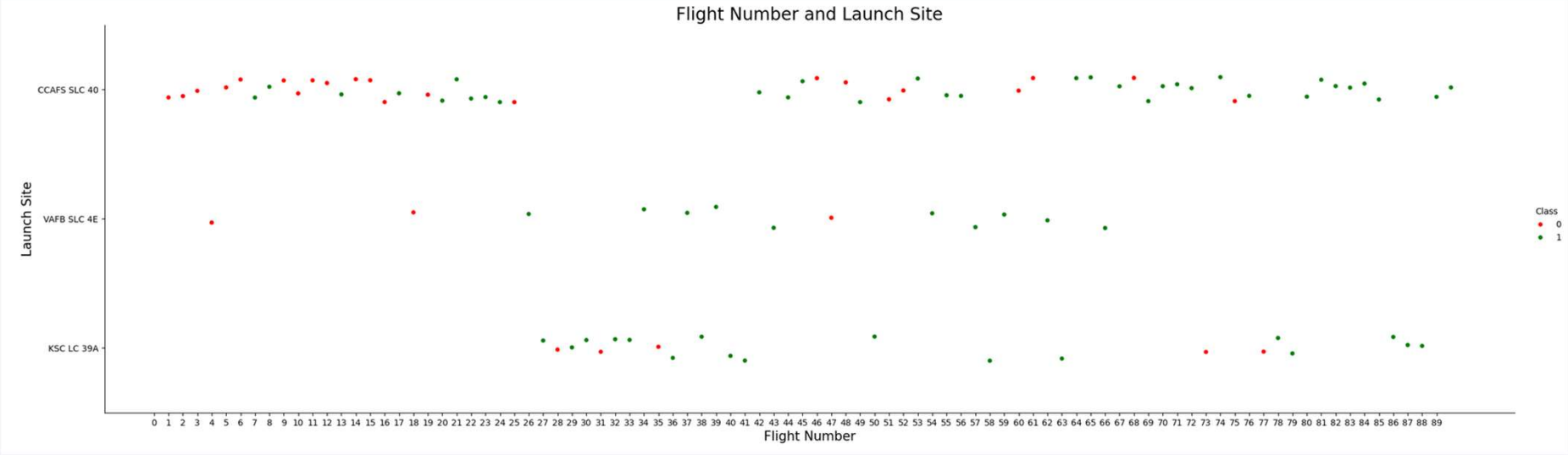
- The below Results are detailed on the remaining slides
  - Exploratory data analysis results
  - Interactive analytics demo in screenshots
  - Predictive analysis results



Section 2

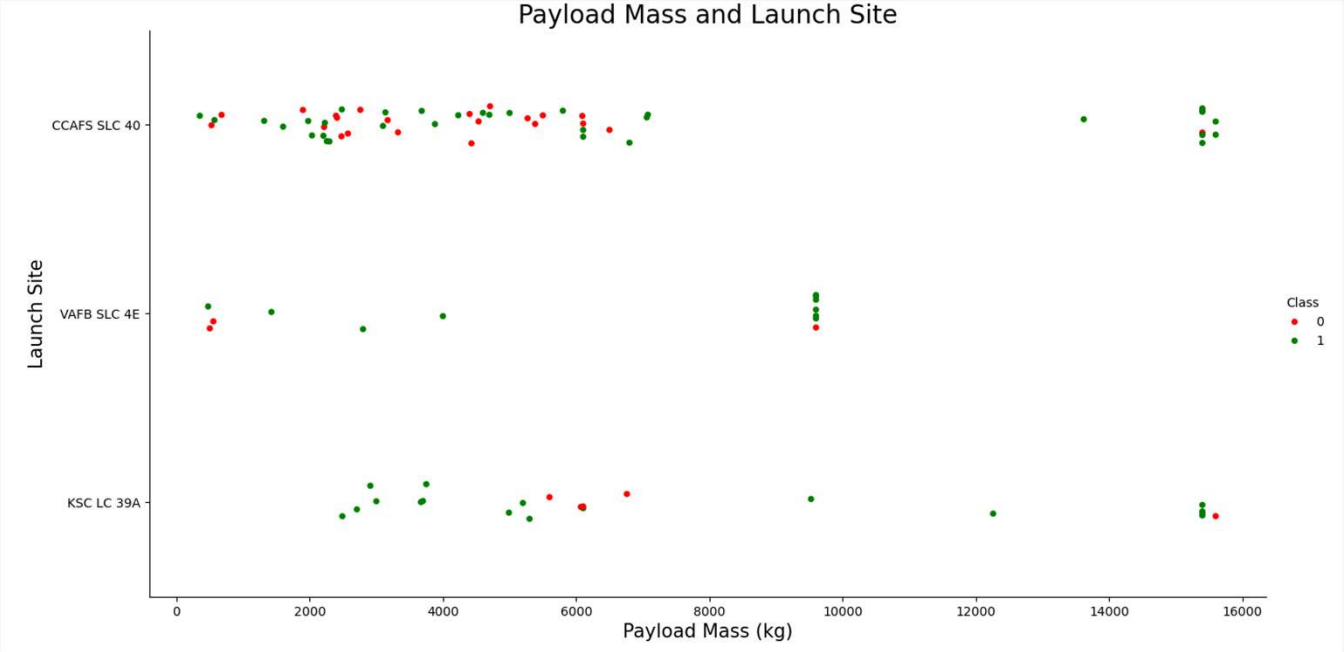
Insights drawn from EDA

Flight Number vs. Launch Site



- 0 (red dots) = failed outcomes; 1 (green dots) = successful outcome
- As the launches increased, the success rate improved. This also points out the high number of fails for location CCAFS SLC 40 is due to that is the only site with early launches when there was a much higher fail rate.

Payload vs. Launch Site

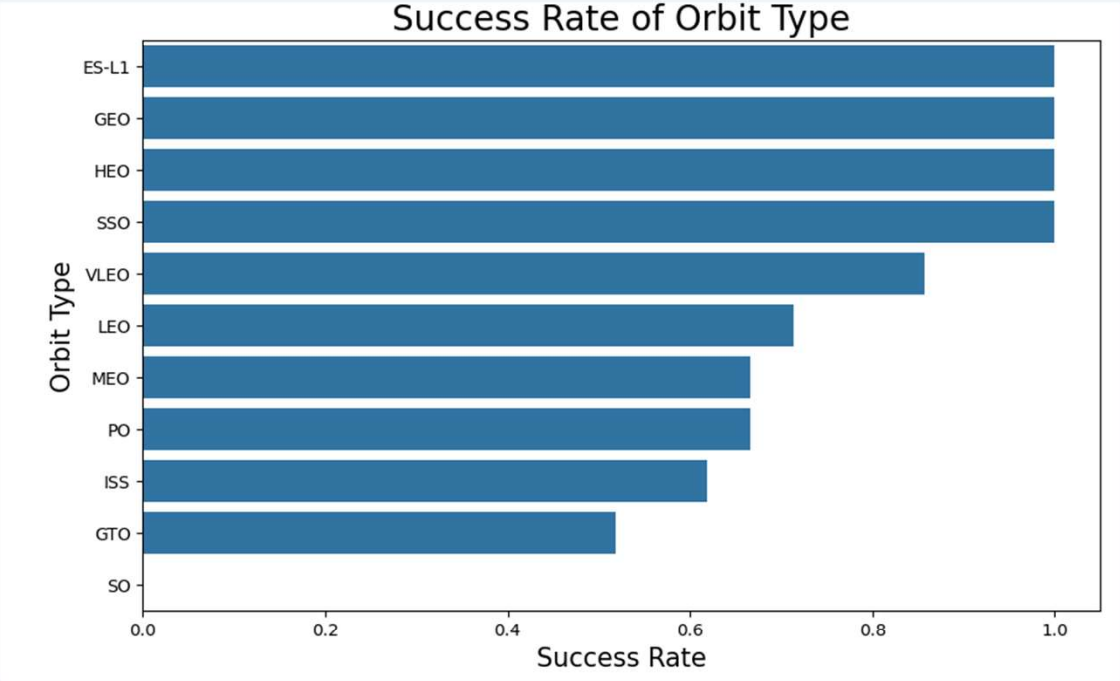


0 (red dots) = failed outcomes

1 (green dots) = successful outcome

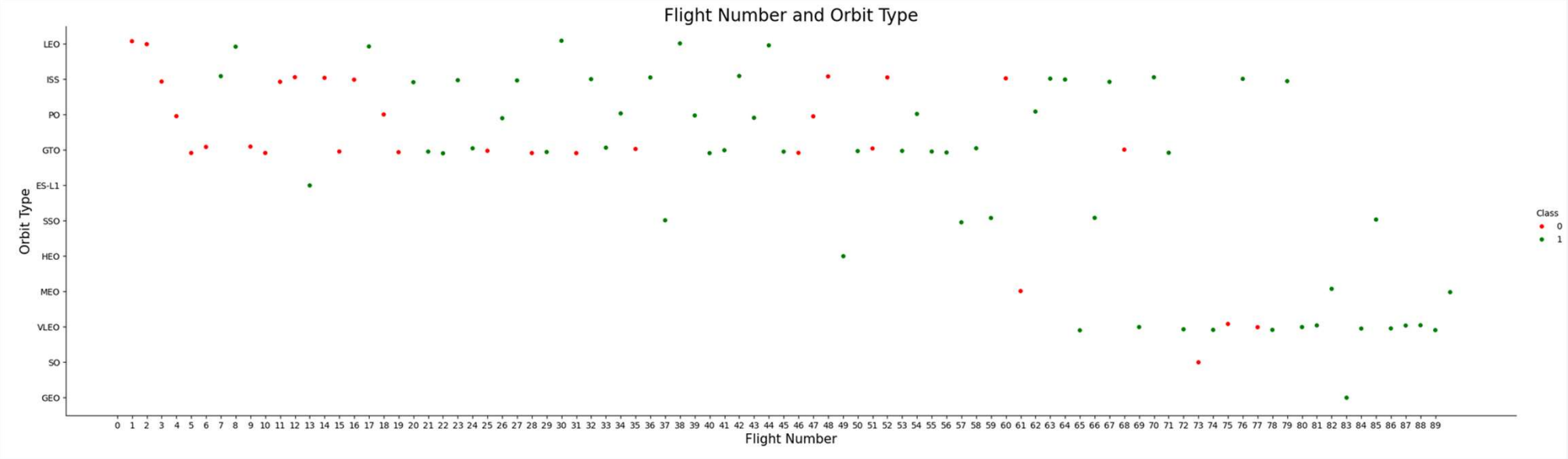
- All sites held launches with payloads under 10,000kg with varying degrees of success.
- Location VAFB SLC 4E did not have any payloads greater than 10,000kg.
- Majority of all launches were under 8,000kg payload

Success Rate vs. Orbit Type



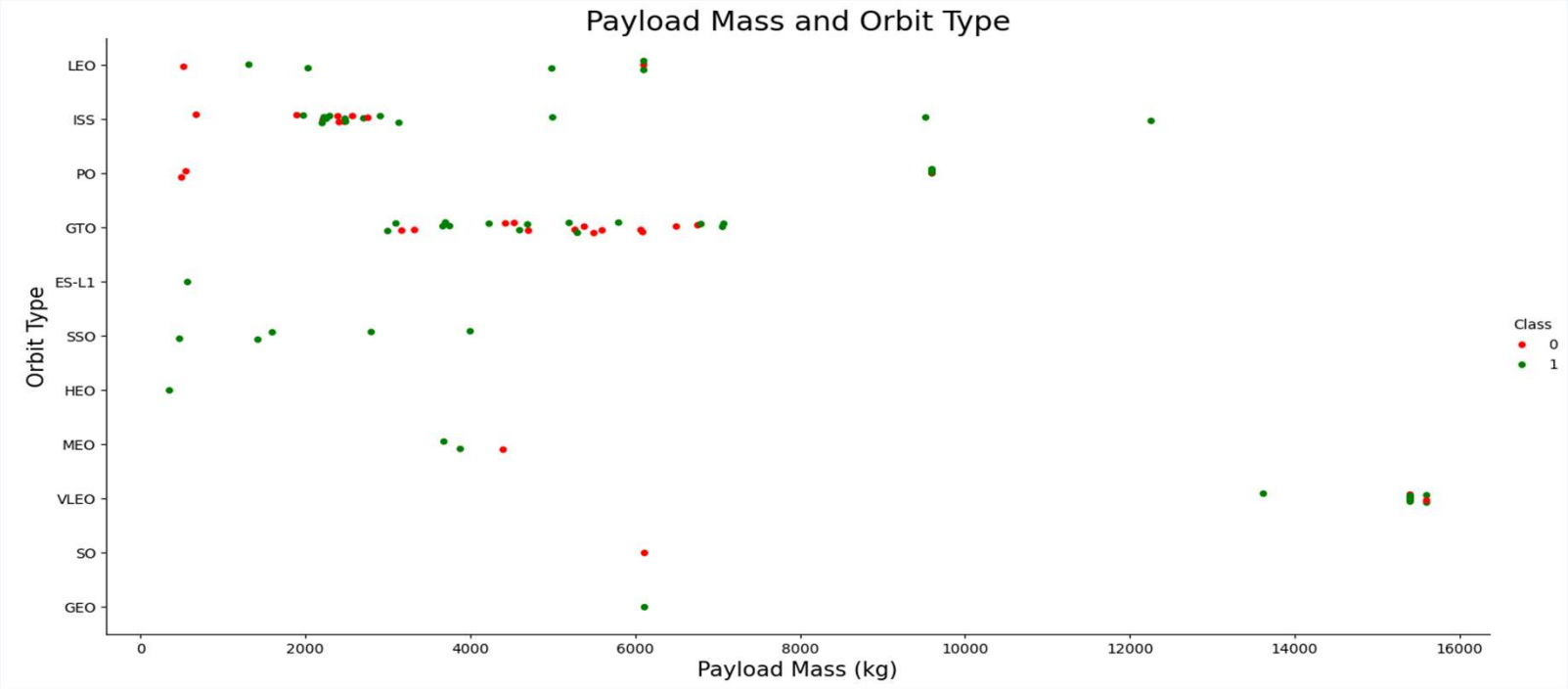
- Orbit types of ES-L1, GEO, HEO, and SSO had 100% success rates
- GTO was the lowest with no orbit types for SO

Flight Number vs. Orbit Type



- VLEO success is more attributed to it being later flights.
- The failure rates of LEO, ISS, PO, and GTO can be more attributed to they were the only orbit types for early flights.
- Since orbit types were not consistently tested from start to finish, it's hard to glean what sort of success or failure can be attributed to orbit type.

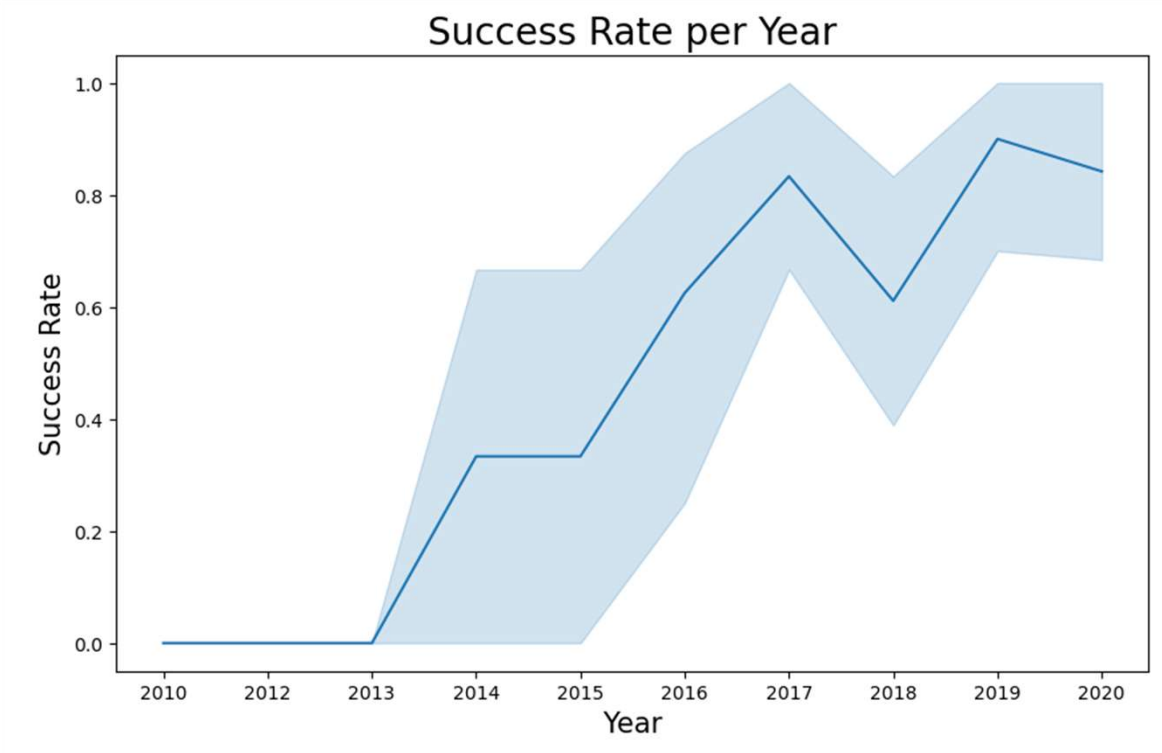
Payload vs. Orbit Type



- GTO had many launches, with no clear success, using payload between ~3,000kg and ~7,000kg

Launch Success Yearly Trend

- Launch success rate has steadily increased from 2013.
- A dip took place in 2018, but recovered to an all time success high in 2019.



All Launch Site Names

- The launch site names are:
  - CCAFS LC-40
  - VAFB SLC-4E
  - KSC LC-39A
  - CCAFS SLC-40

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

- The SQL query used was:

```
SELECT DISTINCT "Launch_Site“
FROM SPACEXTABLE;
```

Launch Site Names Begin with 'CCA'

- Here are 5 records where launch sites 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	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

- The SQL query used was:

```
SELECT *
FROM SPACEXTABLE
WHERE "Launch_Site" LIKE 'CCA%'
LIMIT 5;
```

Total Payload Mass

- The total payload carried by boosters from NASA
  - 99,980kg

NASA Total Payload Mass (KG)
99980

- The SQL query used was:

```
SELECT SUM(PAYLOAD_MASS__KG_) AS 'NASA Total Payload Mass (KG)'
FROM SPACEXTABLE
WHERE Customer LIKE 'NASA%';
```

Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1
  - 2,534.67kg

AVG Payload Mass (KG) Booster F9 v1.1
2534.6666666666665

- The SQL query used was:

```
SELECT AVG(PAYLOAD_MASS__KG_) AS 'AVG Payload Mass (KG) Booster F9 v1.1'
FROM SPACEXTABLE
WHERE Booster_Version LIKE 'F9 v1.1%';
```

First Successful Ground Landing Date

- The date of the first successful landing outcome on ground pad
  - December 22, 2015

Landing_Outcome	First Successful Landing
Success (ground pad)	2015-12-22

- The SQL query used was:

```
SELECT DISTINCT Landing_Outcome, Date AS 'First Successful Landing'
FROM SPACEXTABLE
WHERE Landing_Outcome = 'Success (ground pad)'
ORDER BY date ASC
LIMIT 1;
```

Successful Drone Ship Landing with Payload between 4000 and 6000

- The boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000
  - F9 FT B1022
  - F9 FT B1026
  - F9 FT B1021.2
  - F9 FT B1031.2

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

- The SQL query used was:

```
SELECT Booster_Version
FROM SPACEXTABLE
WHERE Landing_Outcome = "Success (drone ship)"
    AND PAYLOAD_MASS__KG_ > 4000
    AND PAYLOAD_MASS__KG_ < 6000;
```

Total Number of Successful and Failure Mission Outcomes

- The total number of successful and failure mission outcomes
  - Successful: 100
  - Failures: 1

Successes Count
100

Failures Count
1

- 1st SQL query used:

```
SELECT COUNT(Mission_Outcome) AS 'Successes Count'
FROM SPACEXTABLE
WHERE Mission_Outcome LIKE '%Success%';
```

- 2nd SQL query used:

```
SELECT COUNT(Mission_Outcome) AS 'Failures Count'
FROM SPACEXTABLE
WHERE Mission_Outcome LIKE '%Failure%';a
```

Boosters Carried Maximum Payload

- The boosters which have carried the maximum payload mass
  - 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

- The SQL query used was:

```
SELECT DISTINCT Booster_Version AS 'Boosters Carrying Max Payload'
FROM SPACEXTABLE
WHERE PAYLOAD_MASS__KG_ = (
            SELECT MAX(PAYLOAD_MASS__KG_)
            FROM SPACEXTABLE);
```

Boosters Carrying Max Payload
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

- The failed landing outcomes for drone ship, their booster versions, and launch site names for the year 2015
  - January and April

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

- The SQL query used was:

```
SELECT substr(Date, 6,2) AS 'Month', Booster_Version, Launch_Site, Landing_Outcome
FROM SPACEXTABLE
WHERE Landing_Outcome = 'Failure (drone ship)'
        AND substr(Date, 0, 5) = '2015'
```

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

- The count of landing outcomes, fail or success, between June 4th, 2010 and March 20th, 2017
  - Landing_Outcome Counts

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

- The SQL query used was:

```
SELECT Landing_Outcome, COUNT(Landing_Outcome) AS 'Counts'
FROM SPACEXTABLE
WHERE Date BETWEEN '2010-06-04' AND '2017-03-20'
GROUP BY Landing_Outcome
ORDER BY Counts DESC;
```

Landing_Outcome	Counts
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

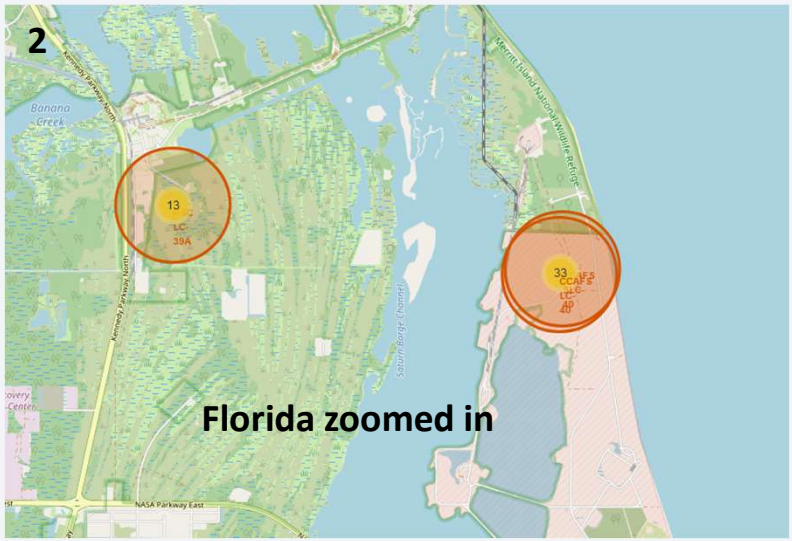
Launch Sites Proximities Analysis

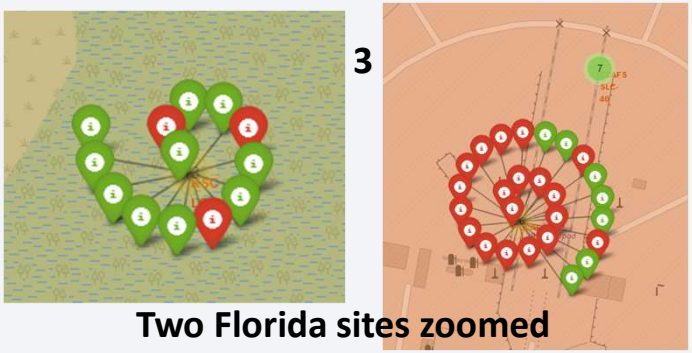
Map of Launch Locations

- The 1st Folium map Shows the launch locations on both the west coast (California) and east cost (Florida)

Map with Launch Counts

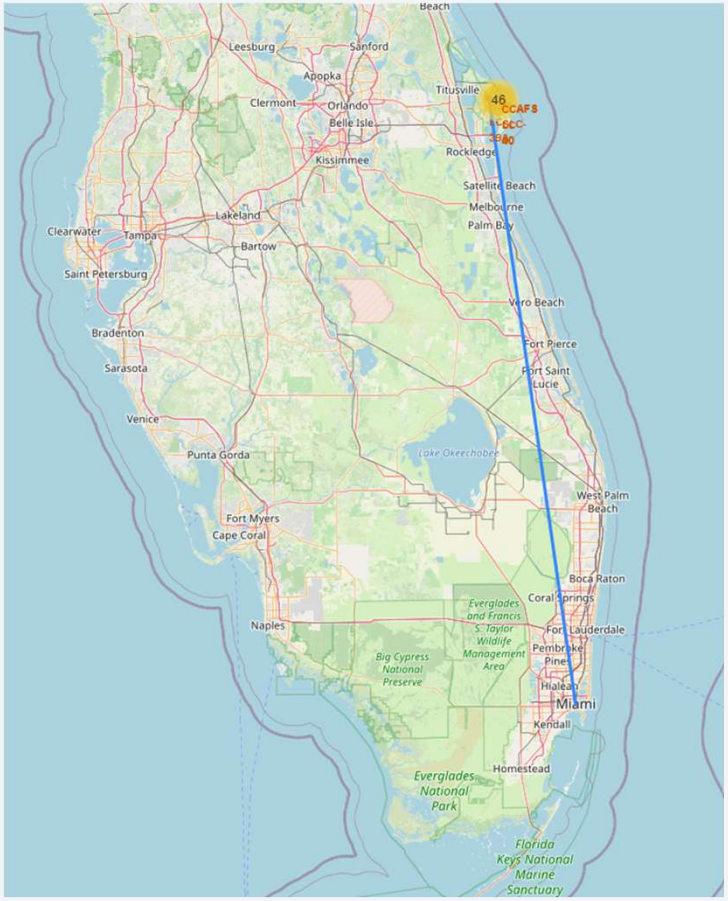
- The 2nd map added counts and clusters when zoomed in.
- Green = success | Red = fail

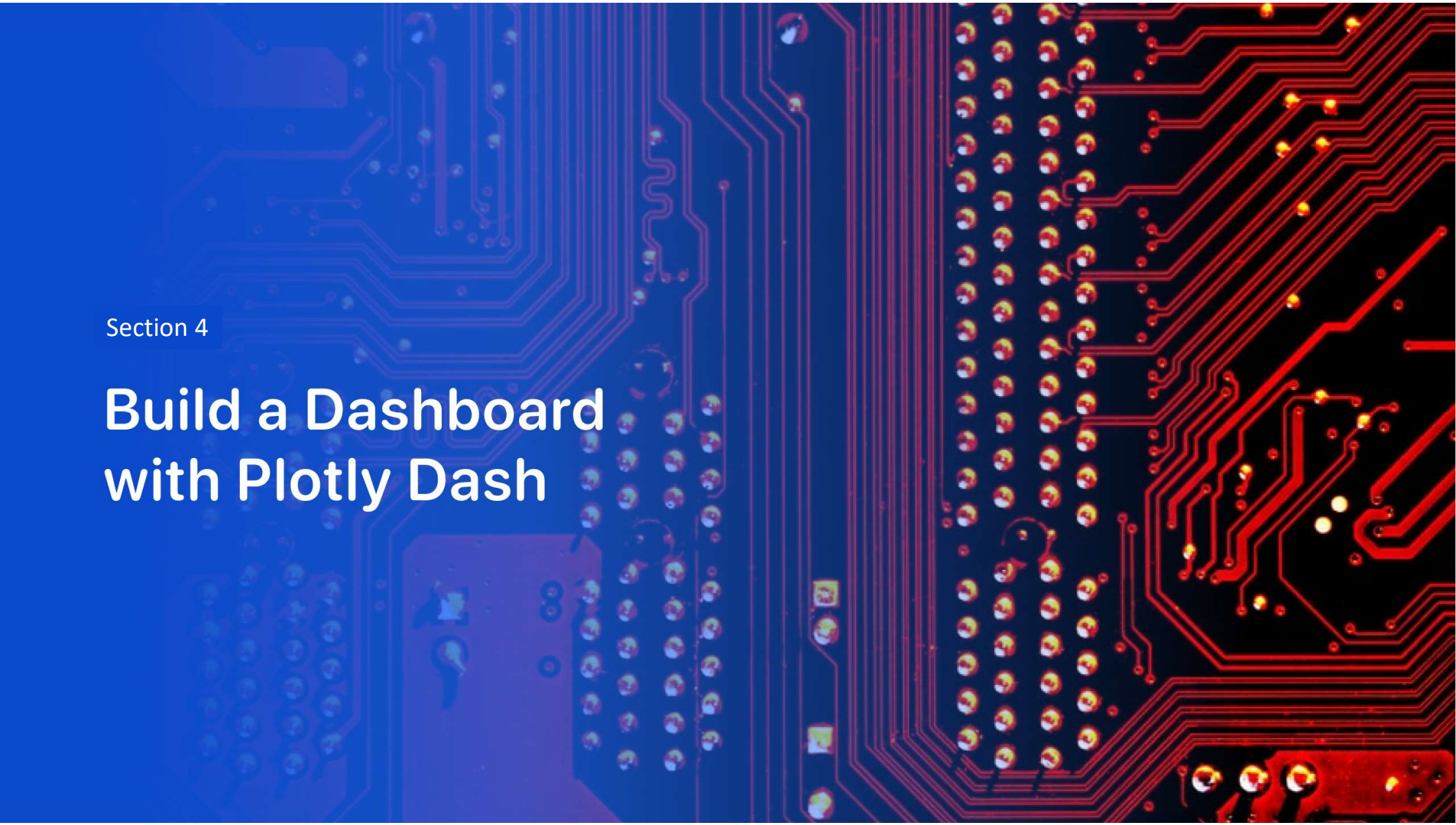




Map of a Launch Site to Miami

- The 3rd map added lines to nearby locations.
- The example on the map to the right is from launch site KSC LC-39A to Miami.
  - 314.77 miles
- Note: Any marker, line, and distance can be added to the map as needed.

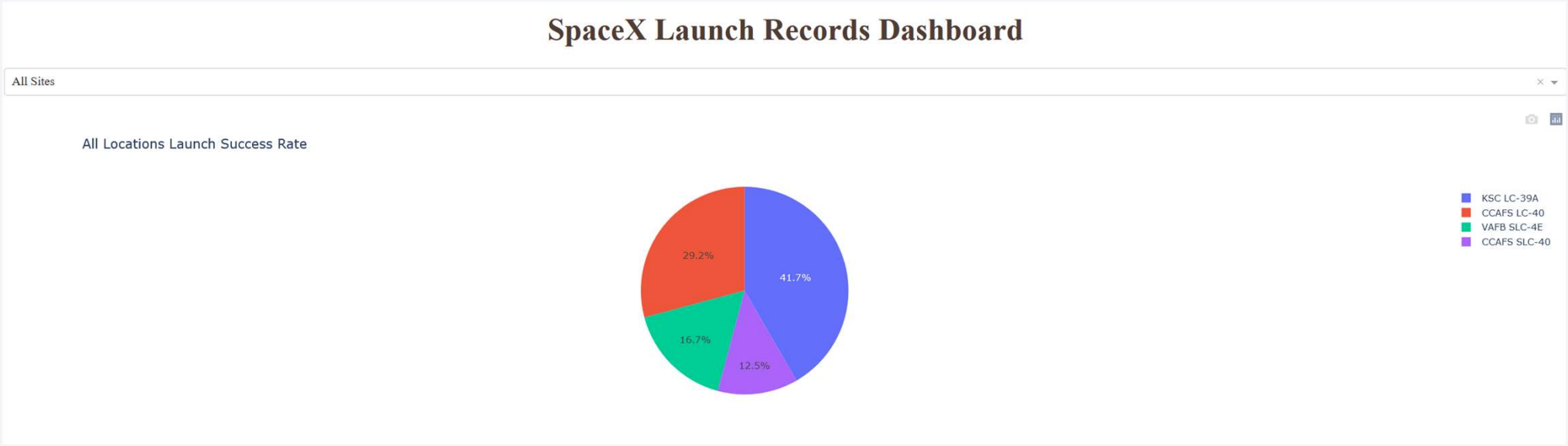




Section 4

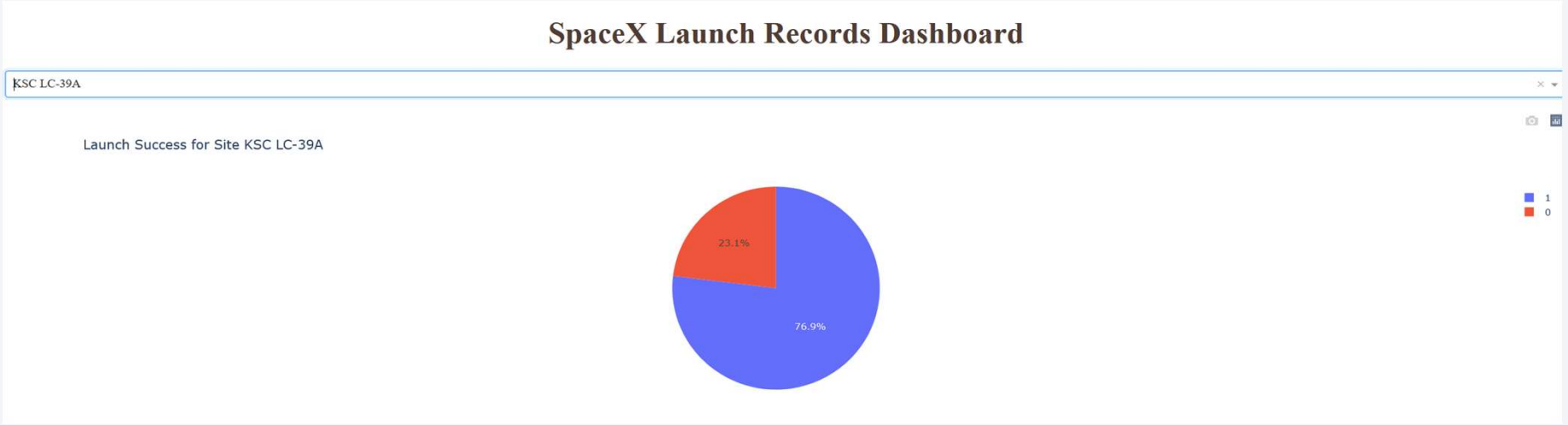
Build a Dashboard with Plotly Dash

All Launch Site Success Rate



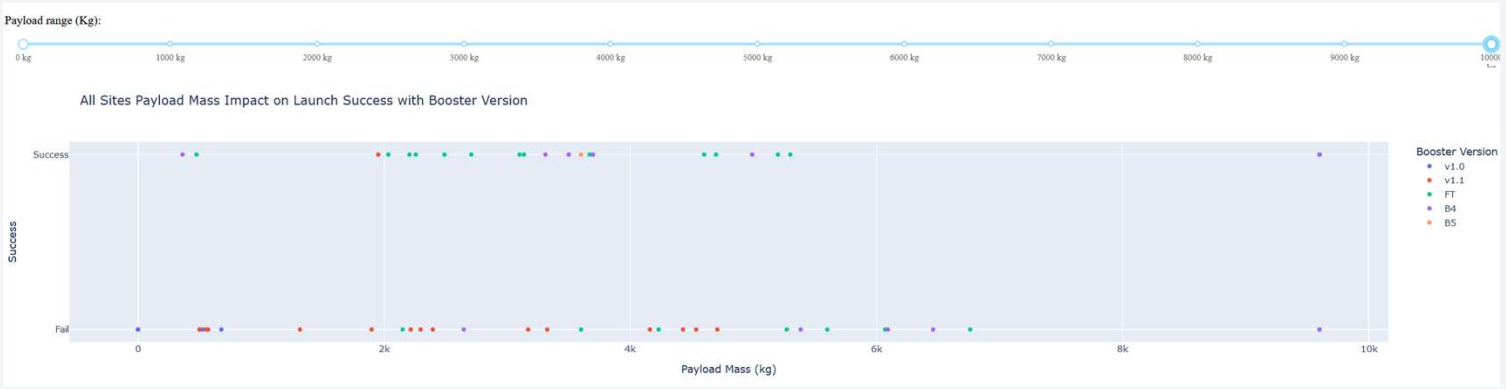
The launch location with the highest success rate is KSC LC-39A with a 41.7% success rate

KSC LC-39A Launch Success Rate



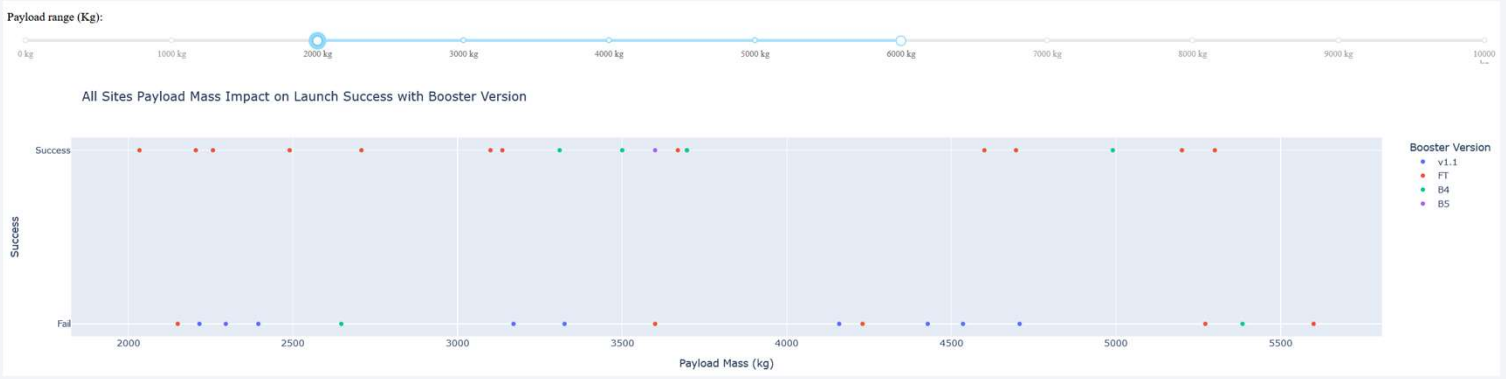
Focusing in on KSC LC-39A, we see the success rate of just that site is 76.9%

Payload Slider with Successes and Failures



All Sites; payload slider set to full (0kg to 10,000kg)

Most successes are between 2,000kg and 6,000kg



All Sites; payload slider set to range (2,000kg to 6,000kg)

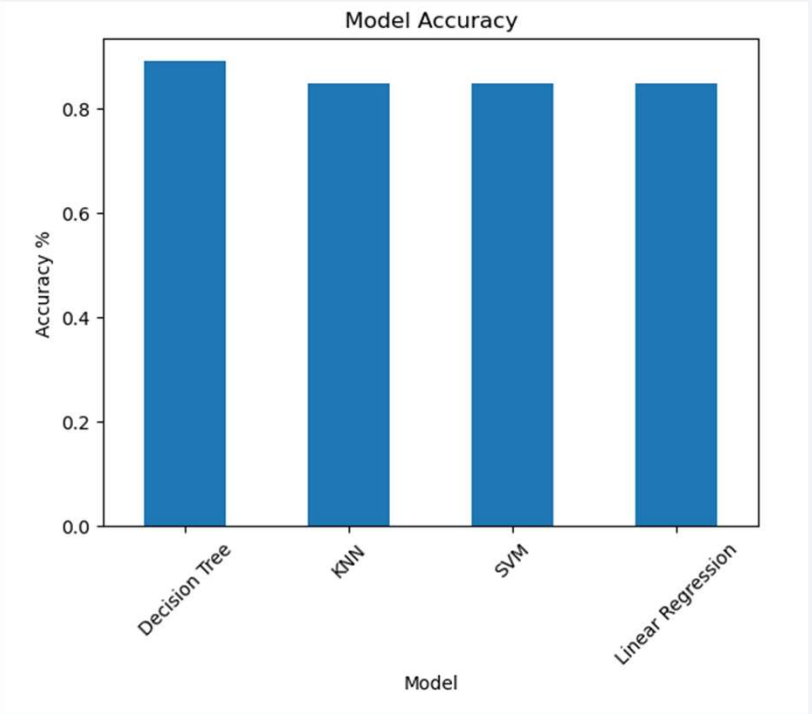
Section 5

Predictive Analysis (Classification)

Classification Accuracy

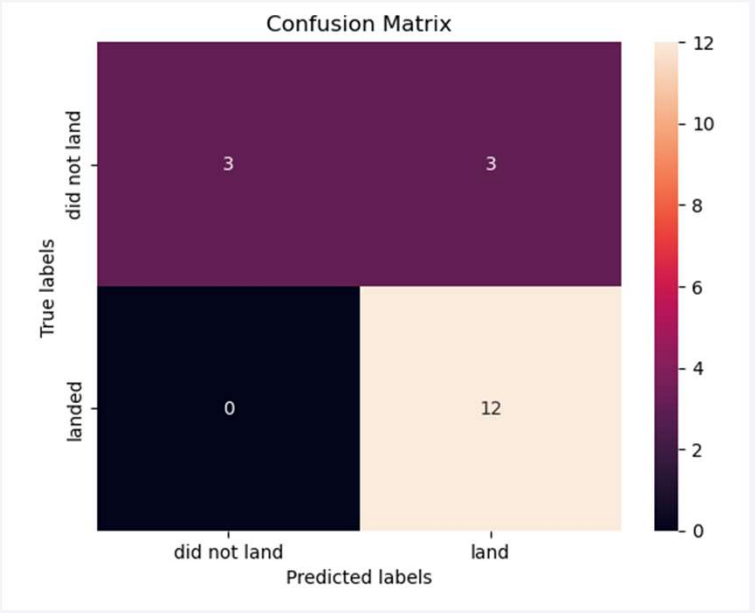
- Of the 4 models built, trained, and tested, the Decision Tree model proved the best.

Model	Accuracy
Decision Tree	0.889286
KNN	0.848214
SVM	0.848214
Linear Regression	0.846429



Confusion Matrix

- The Decision Tree Confusion Matrix shows we had 12 successful landings and 6 failed landings.
- The Decision Tree predicted 15 successes and 3 failures, so 3 true fails were mislabeled as successes, or in other words, the model had 3 false positives.
- This is the model that performed the best, however it was not by a wide margin.



Conclusions

- All Launch Sites have a high success rate
- Payload mass seems to have minimal impact on success rate
- Orbit type appeared to have an impact, however it was more due to later launches were of a different orbit type and later launches were simply more successful
- More than anything, time and experience is what appears to have the greatest impact on success rate. The visualizations, dashboard, and models all indicated a varying degree of success with the various variables and they all improved over time as the number of launches increased. By 100 launches, SpaceX seems got it pretty well figured out, however since they did not test everything equally throughout all 100 launches, it made it a little difficult to discern what led to better results.
- Overall success rate for the Falcon 9 was 67% allowing the option to reuse the rocket 2 out of every 3 launches.

Appendix

- All Jupyter Notebooks, source files, the Dashboard, and presentation can be found at the Github repository found here.

Thank you!