IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Name>
<Date>

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection
  - Data Wrangling
  - Exploratory Data Analysis (EDA) with Visualization and SQL queries
  - Interactive Visuals and Dashboard with Folium and Plotly Dash
  - Predictive Analysis (Classification)
- Summary of all results
  - EDA results
  - Interactive Demo results
  - Predictive Analysis Results

Introduction

- Project background and context
  - SpaceX is the most successful commercial space travel company allowing civilians the ability to go to space. Advertised at 62 million dollars, compared to the cost of over 165 million from other providers SpaceX is considered the more cost-effective option. This is largely due to their ability to reuses first stage rockets. Determining if the first stage lands will help us calculate the cost of a launch. Based on the data obtained and methods used in this project, we will predict if this reuse is possible.
- Questions?
  - How do variables such as launch site location, payload mass, number of launches, and orbits effect success rate of first stage landings?
  - Is the success rate of these landings improving over time?
  - What predictive model produces the best results for the binary classification problem this case presents?

Section 1

Methodology

Methodology

Executive Summary

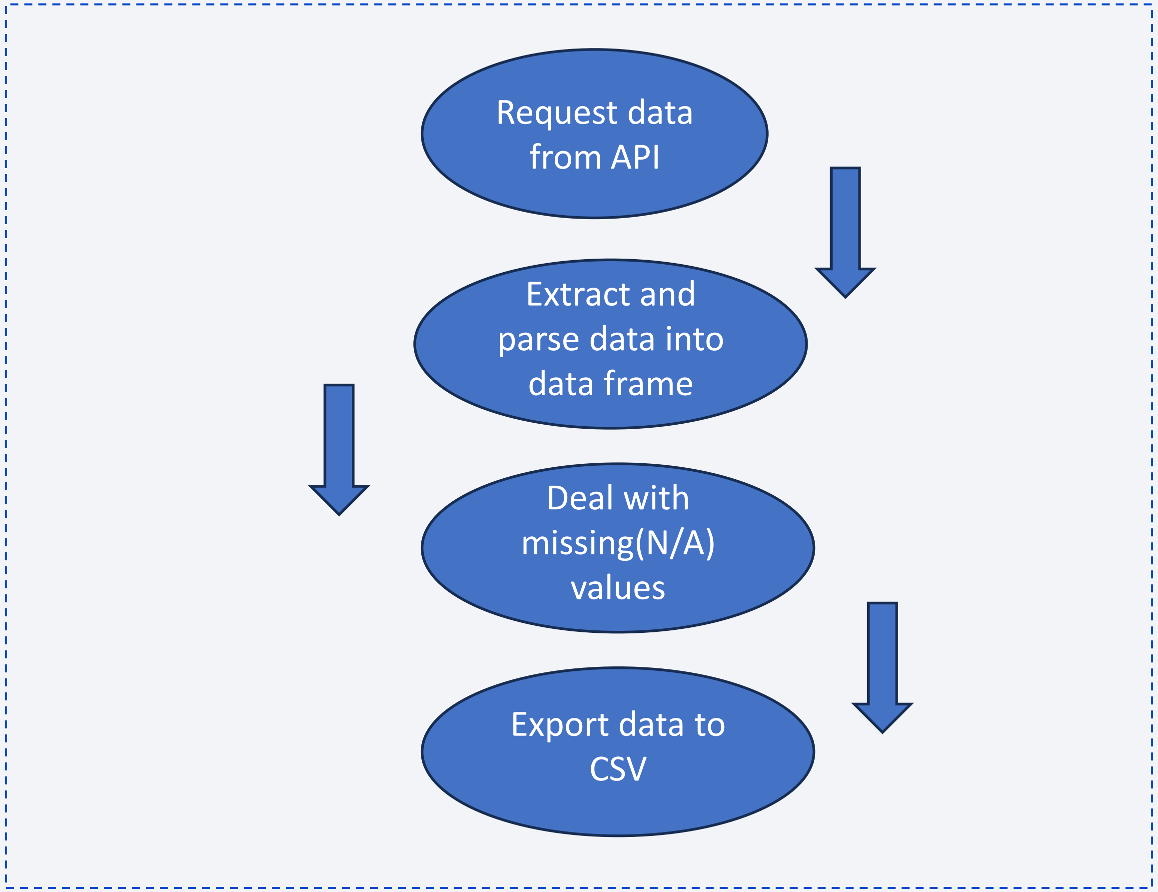
- Data collection methodology:
  - SpaceX API calls and web scraping from the SpaceX Wikipedia
- Perform data wrangling
  - Filter data
  - Deal with missing values
  - Implement One Hot Encoding to prepare data for classification
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models.
  - Building of multiple models, tuning, and evaluating to ensure maximum accuracy.

Data Collection

- Data sets were obtained from the SpaceX API as well as the SpaceX Wikipedia.
- Both data sets, obtained through data collection methods like web scraping, ensured complete information to be used for deatailed analysis.

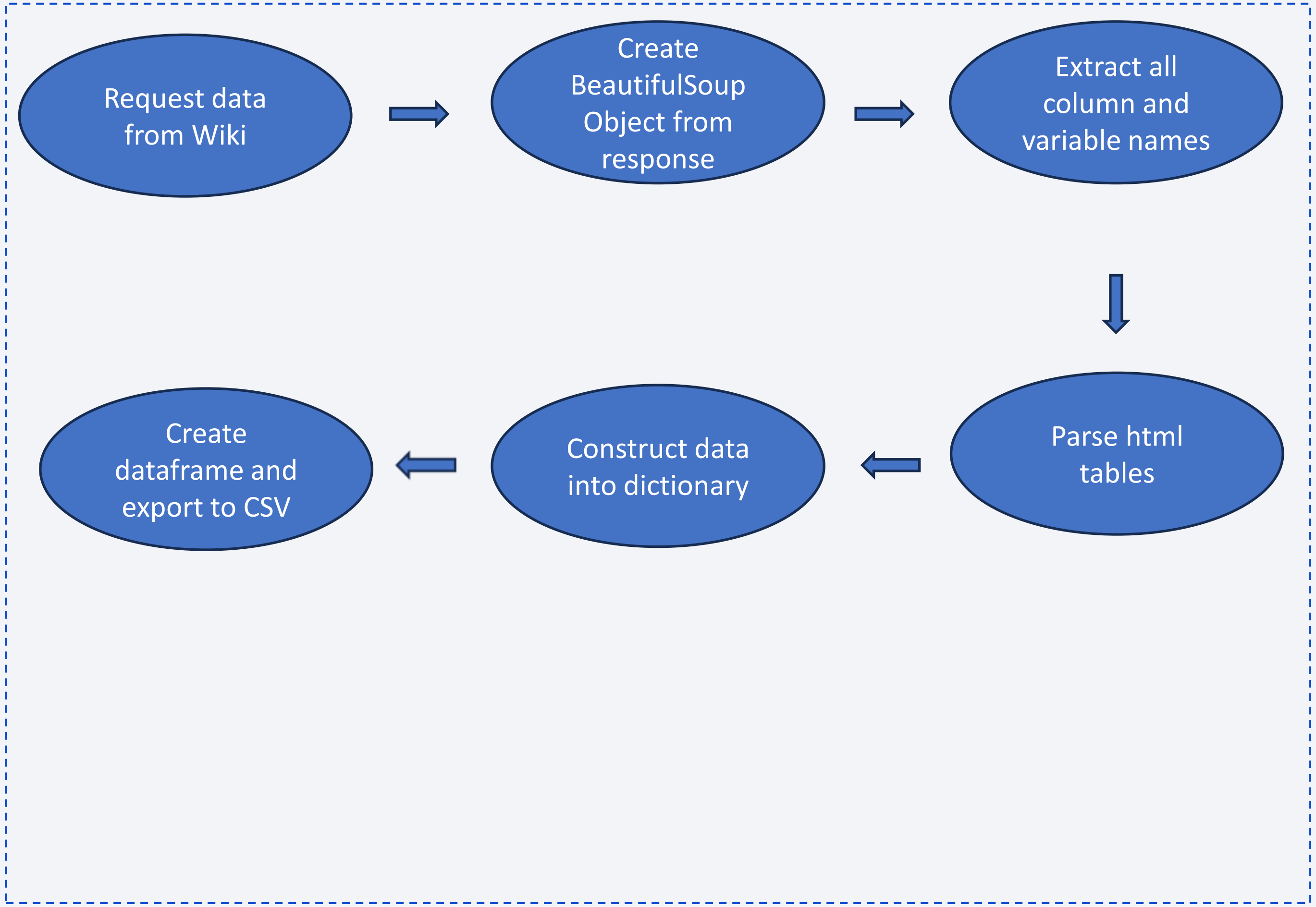
Data Collection – SpaceX API

- The public SpaceX API was used to retrieve data.
- https://github.com/mboettger32/IBM-Applied-Data-Science-Captone/blob/94194dc5cca479a8712ee0d5a4c9d72a717a5d66/jupyter-labs-spacex-data-collection-api%20(1)%20(1).ipynb



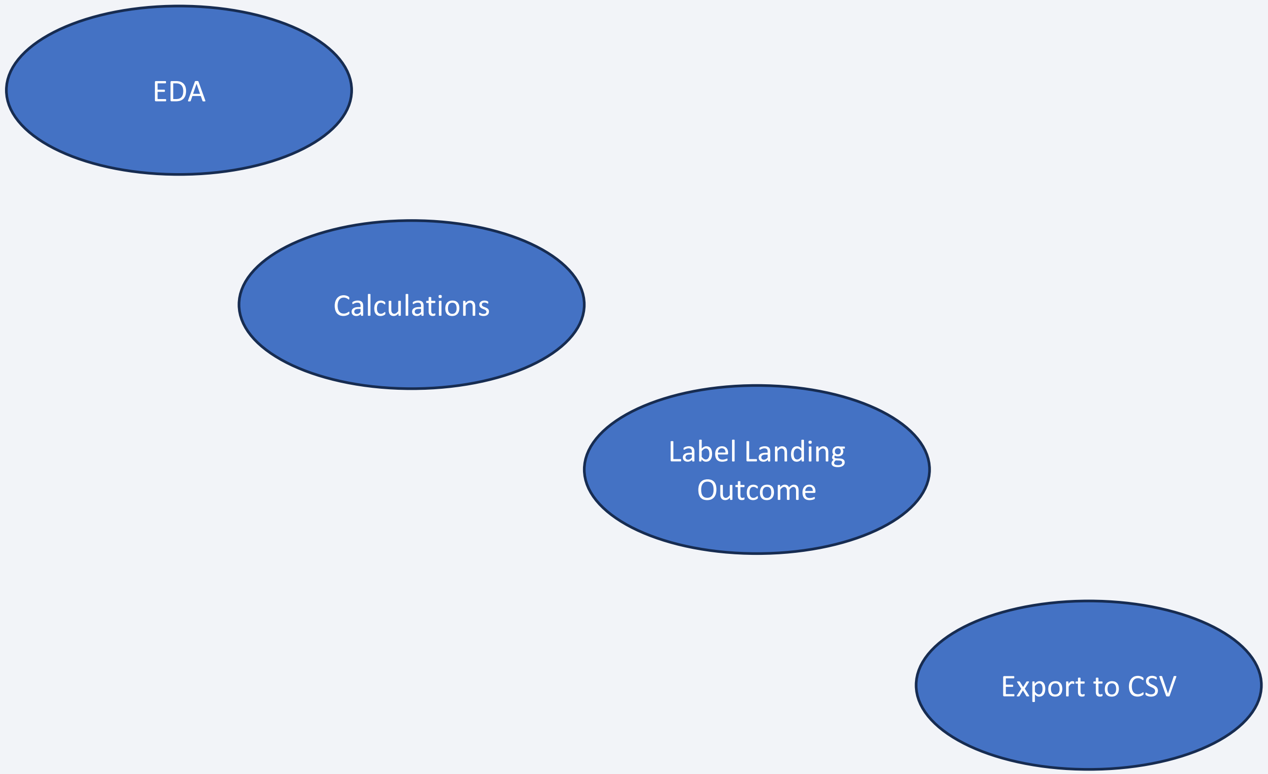
Data Collection - Scraping

- SpaceX launch data obtained from their Wikipedia
- https://github.com/mboettger32/IBM-Applied-Data-Science-Captone/blob/94194dc5cca479a8712ee0d5a4c9d72a717a5d66/jupyter-labs-webscraping%20(1).ipynb



Data Wrangling

- Exploratory Data Analysis was performed on the data.
- Calculated number of launches at each site, the number and occurrence of each orbit, and the number and occurrence of mission outcome per orbit type.
- Create landing outcome label based on success.
- https://github.com/mboettger32/IBM-Applied-Data-Science-Captone/blob/94194dc5cca479a8712ee0d5a4c9d72a717a5d66/3)jupyter-spacex-data_wrangling_jupyterlite.jupyterlite.ipynb



EDA with Data Visualization

- Charts Created
  - Flight Number vs. Payload Mass, Flight Number vs. Launch Site, Payload Mass vs. Launch Site, Orbit Type vs. Success Rate, Flight Number vs. Orbit Type, Payload Mass vs Orbit Type and Success Rate Yearly Trend.
- Scatter Plots
  - To show variable relationships.
- Bar Charts
  - To compare discrete categories based on a measured value.
- Line Charts
  - To visualize trends in data over time.
- https://github.com/mboettger32/IBM-Applied-Data-Science-Captone/blob/94194dc5cca479a8712ee0d5a4c9d72a717a5d66/5)jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb

EDA with SQL

- SQL Queries Performed
  - Displaying the names of the unique launch sites in the space mission
  - Displaying 5 records where launch sites begin with the string ‘CCA’
  - Displaying the total payload mass carried by boosters launched by NASA (CRS)
  - Displaying average payload mass carried by booster version F9 v1.1
  - Listing the date when the first successful landing outcome in ground pad was achieved
  - Listing the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000L
  - Listing the total number of successful and failure mission outcomes
  - Listing the names of the booster versions which have carried the maximum payload mass
  - Listing the failed landing outcomes in drone ship, their booster versions and launch site names for the months in 2015
  - Ranking 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
  - https://github.com/mboettger32/IBM-Applied-Data-Science-Captone/blob/94194dc5cca479a8712ee0d5a4c9d72a717a5d66/4)jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Markers
  - Added markers with circle, popup label and text label of NASA Space Center suing coordinates.
  - Marked all launch sites to show proximity to coastline.
- Colored Markers
  - Green and red markers indicated success rate of launch sites.
- Distance Lines
  - Showed distance between selected launch site to coastlines.
- https://github.com/mboettger32/IBM-Applied-Data-Science-Captone/blob/94194dc5cca479a8712ee0d5a4c9d72a717a5d66/6)jupyter_launch_site_location.jupyterlite.ipynb

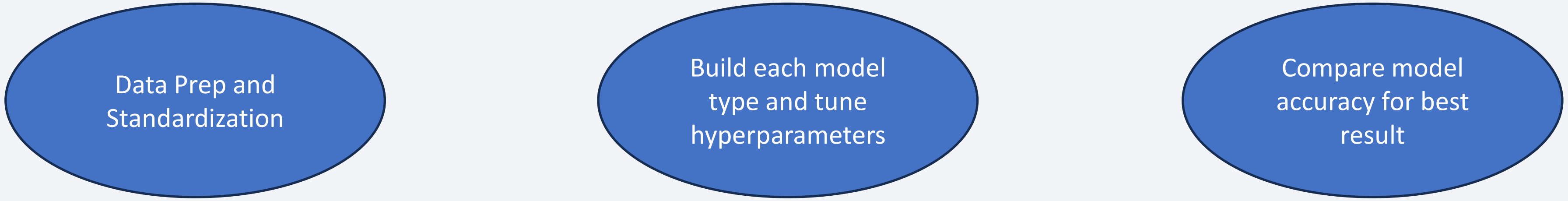
Build a Dashboard with Plotly Dash

- Dropdown List
  - Launch site selection
- Pie Chart
  - Shows total launches as well as success vs failure at selected site.
- Slider
  - To select payload range
- Scatter Chart
  - Correlate payload mass to success rate for various booster versions.

https://github.com/mboettger32/IBM-Applied-Data-Science-Captone/blob/94194dc5cca479a8712ee0d5a4c9d72a717a5d66/7)spacex_dash_app.py

Predictive Analysis (Classification)

- Comparison of four models to achieve best results
  - Logistic Regression
  - Support Vector Machine
  - Decision Tree
  - K Nearest Neighbors



Results

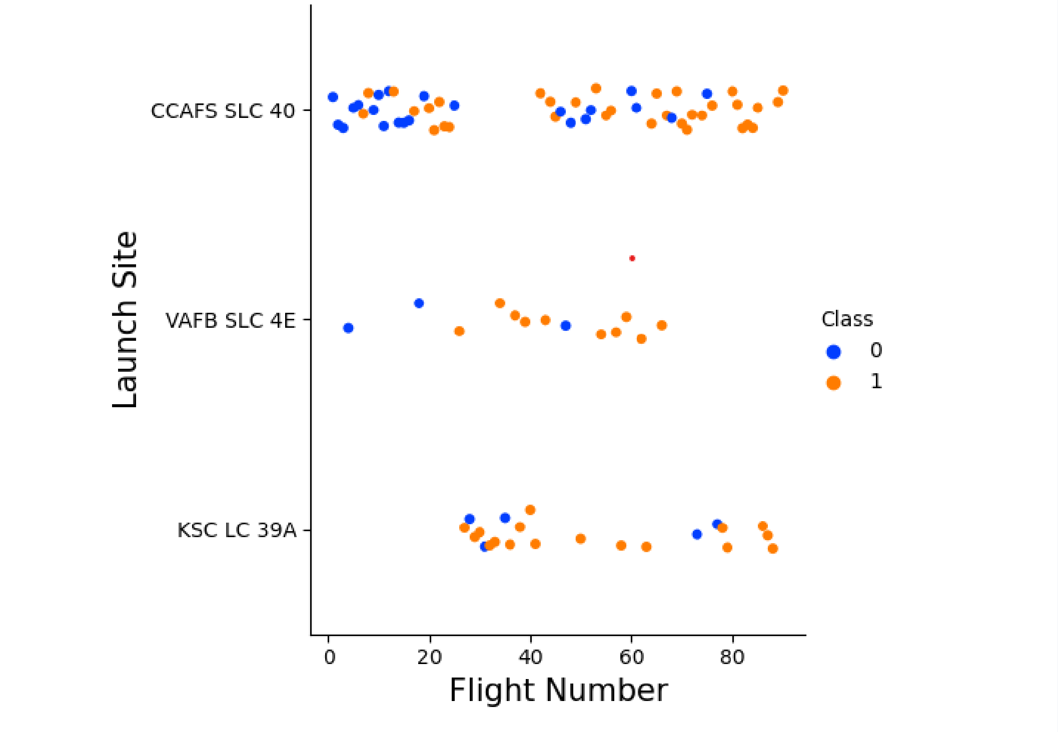
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

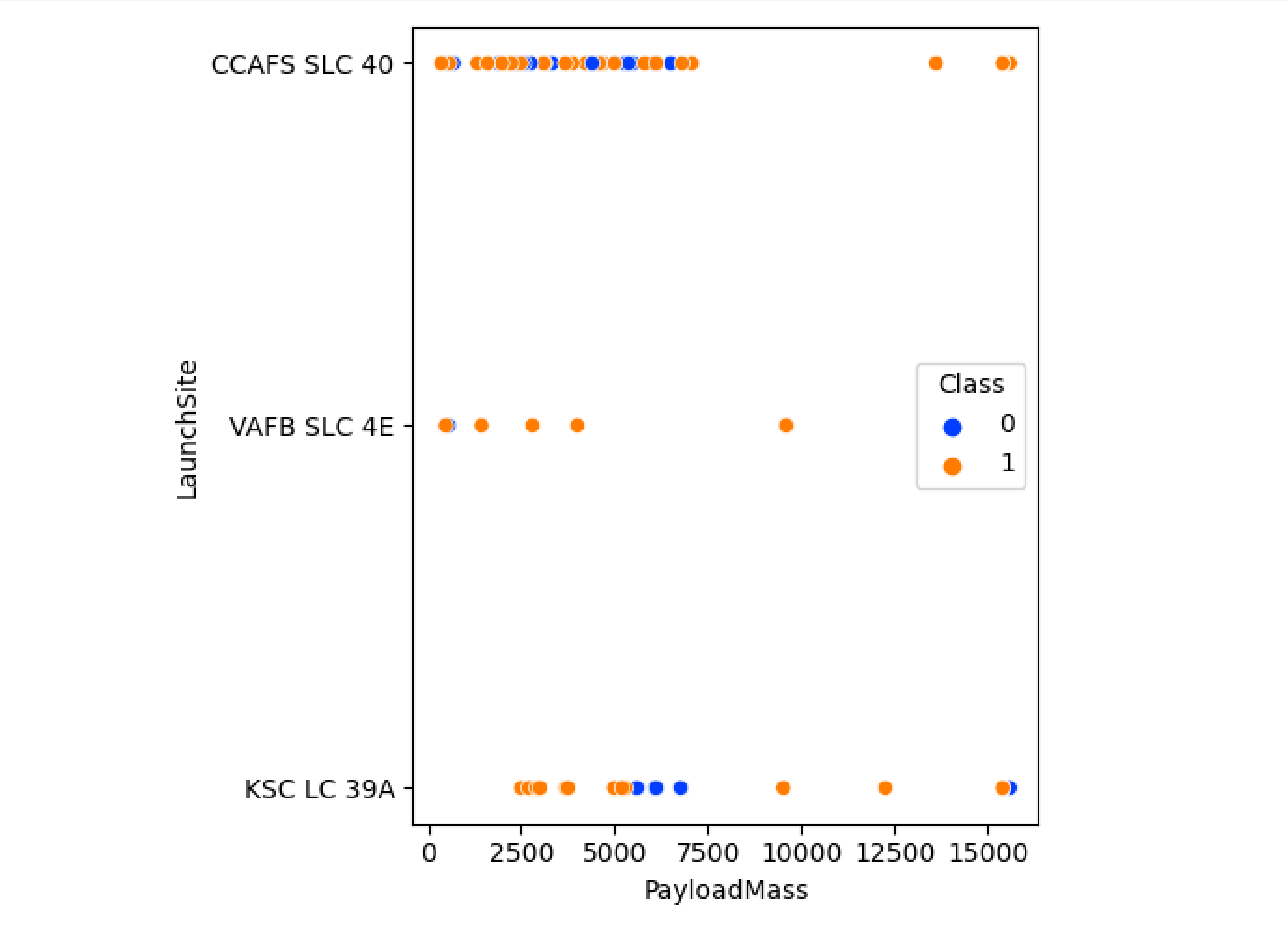
Flight Number vs. Launch Site

- Explanation
  - The earliest flights all failed while the latest flights all succeeded.
  - The CCAFS SLC 40 launch site has about a half of all launches.
  - VAFB SLC 4E and KSC LC 39A have higher success rates.
  - It can be assumed that each new launch will have a higher rate of success.



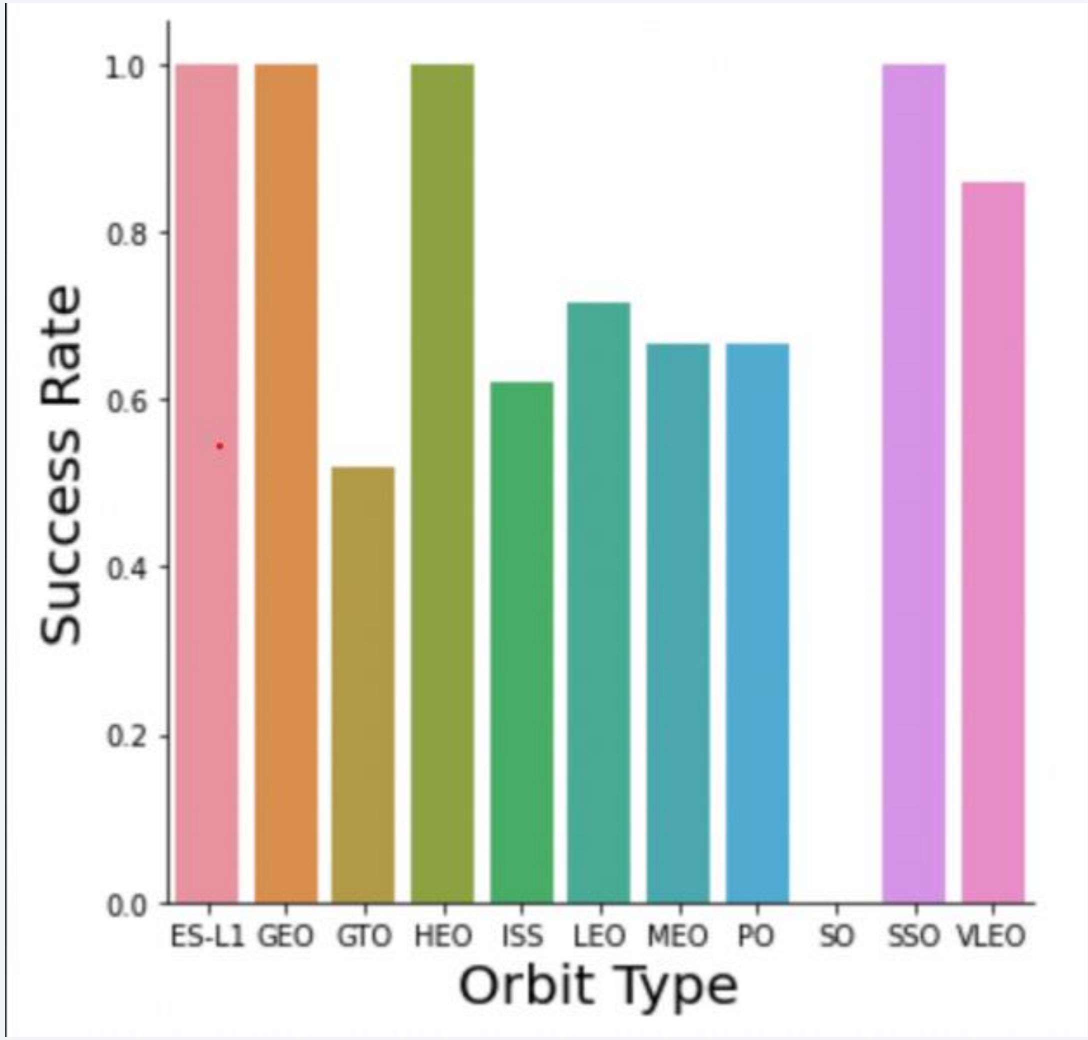
Payload vs. Launch Site

- Explanation
  - CCAFS SLC 40's flights success do not seem directly correlated to payload mass.
  - Higher payload mass may only slightly influence success in a positive way.



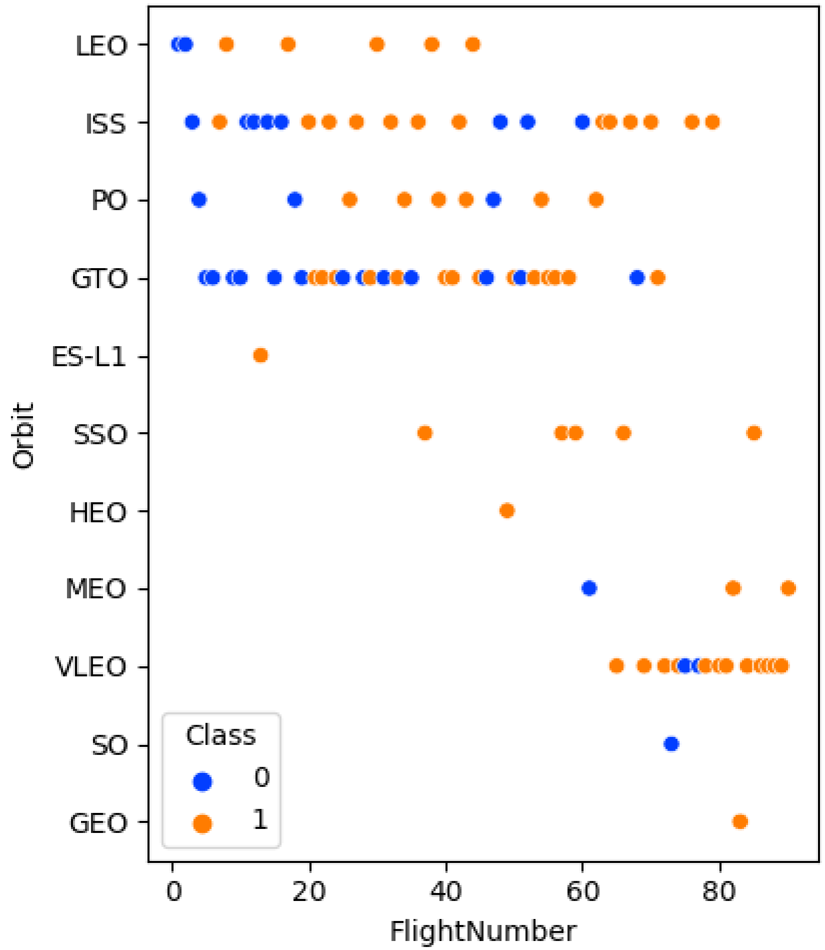
Success Rate vs. Orbit Type

- Orbits with 100% Success Rate
  - ES-L1, GEO, HEO, SSO
- Orbits with 0% Success Rate
  - SO
- All other Orbit Types have Success Rates Between 50% and 85%



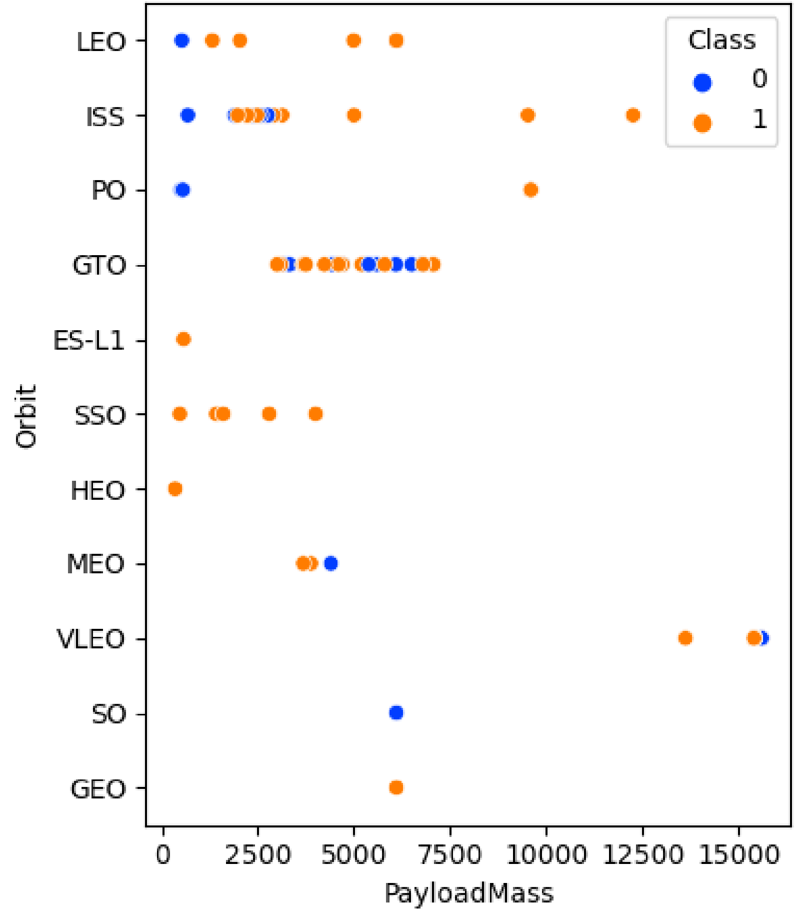
Flight Number vs. Orbit Type

- LEO is the only Orbit Type with a success rate that seems to correlate to flight number.
- All others show no distinct correlation.



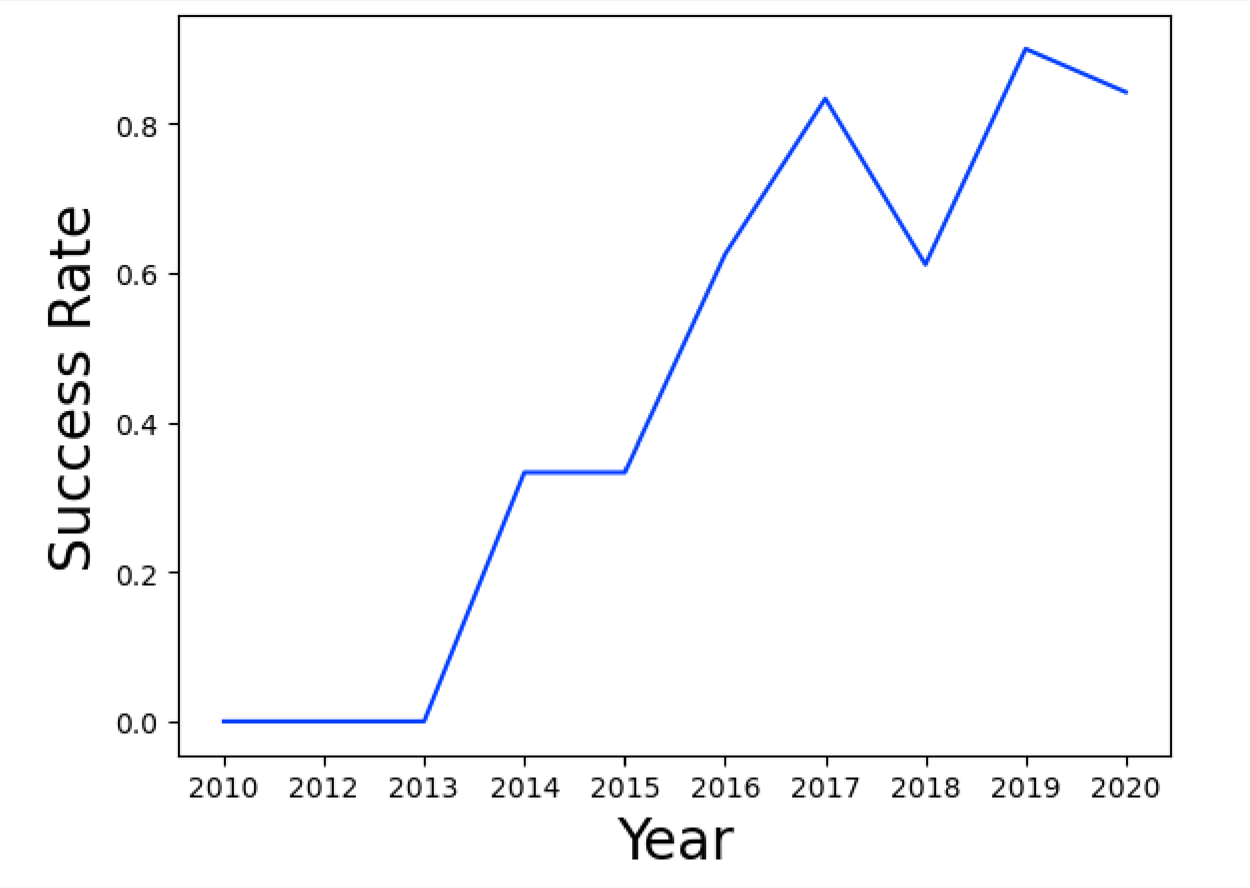
Payload vs. Orbit Type

- Heavier Payloads have higher success with Polar, LEO, and ISS orbit types.
- GTO on the other hand, cannot be distinguished.



Launch Success Yearly Trend

- Explanation
  - There has been an overall positive trend in success rate since 2013.



All Launch Site Names

- The following SQL query was used to find the names of each launch site.

```
[11]: %sql SELECT distinct(LAUNCH_SITE) FROM SPACEXTBL;

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

[11]:

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

Launch Site Names Begin with 'CCA'

- The following query displays 5 launches that where the launch site begins with the string ‘CCA’.

```
[22]: %sql SELECT * FROM SPACEXTBL WHERE LAUNCH_SITE LIKE 'CCA%' LIMIT 5;

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

[22]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
06/04/2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0.0	LEO	SpaceX	Success	Failure (parachute)
12/08/2010	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0.0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
22/05/2012	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525.0	LEO (ISS)	NASA (COTS)	Success	No attempt
10/08/2012	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500.0	LEO (ISS)	NASA (CRS)	Success	No attempt
03/01/2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677.0	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

- Query to calculate total payload mass of all NASA launches.

```
[13]: %sql SELECT SUM(PAYLOAD_MASS__KG_) AS TOTAL_PAYLOAD_MASS FROM SPACEXTBL WHERE CUSTOMER = 'NASA (CRS)';

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

[13]:

TOTAL_PAYLOAD_MASS
45596.0

Average Payload Mass by F9 v1.1

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

```
[14]: %sql SELECT AVG(PAYLOAD_MASS__KG_) AS AVERAGE_PAYLOAD_MASS FROM SPACEXTBL \
          WHERE BOOSTER_VERSION = 'F9 v1.1';

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

[14]:

AVERAGE_PAYLOAD_MASS
2928.4

First Successful Ground Landing Date

- Query to find the date of the first successful landing outcome on ground pad

```
[15]: %sql SELECT MIN(DATE) AS FIRST_SUCCESSFUL_GROUND_LANDING FROM SPACEXTBL \
          WHERE Landing_Outcome = 'Success (ground pad)';

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

[15]:

FIRST_SUCCESSFUL_GROUND_LANDING
01/08/2018

Successful Drone Ship Landing with Payload between 4000 and 6000

- Query lists the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

```
[16]: %sql SELECT BOOSTER_VERSION FROM SPACEXTBL \
          WHERE (Landing_Outcome = 'Success (drone ship)') AND (PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000);

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

[16]:

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

```
[17]: %sql SELECT MISSION_OUTCOME, COUNT(MISSION_OUTCOME) AS TOTAL_NUMBER FROM SPACEXTBL GROUP BY MISSION_OUTCOME;

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

[17]:

Mission_Outcome	TOTAL_NUMBER
None	0
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- Lists the names of the booster which have carried the maximum payload mass

[18]:
```
%sql SELECT DISTINCT(BOOSTER_VERSION) FROM SPACEXTBL \
    WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL);
```

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

[18]:

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

2015 Launch Records

- Lists the failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015

```
[24]: %sql SELECT BOOSTER_VERSION, LAUNCH_SITE, Landing_Outcome, substr(Date, 4, 2) as month FROM SPACEXTBL \
        WHERE (Landing_Outcome = 'Failure (drone ship)') AND (substr(Date,7,4)='2015');

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

[24]:

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

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

- Ranks the count of landing outcomes between the date 2010-06-04 and 2017-03-20, in descending order.

```
[36]: %%sql select Landing_Outcome, count(*) as count_outcomes from SPACEXTBL
        where date between '04/06/2010' and '20/03/2017'
         group by Landing_Outcome
         order by count_outcomes desc;
```

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

[36]:

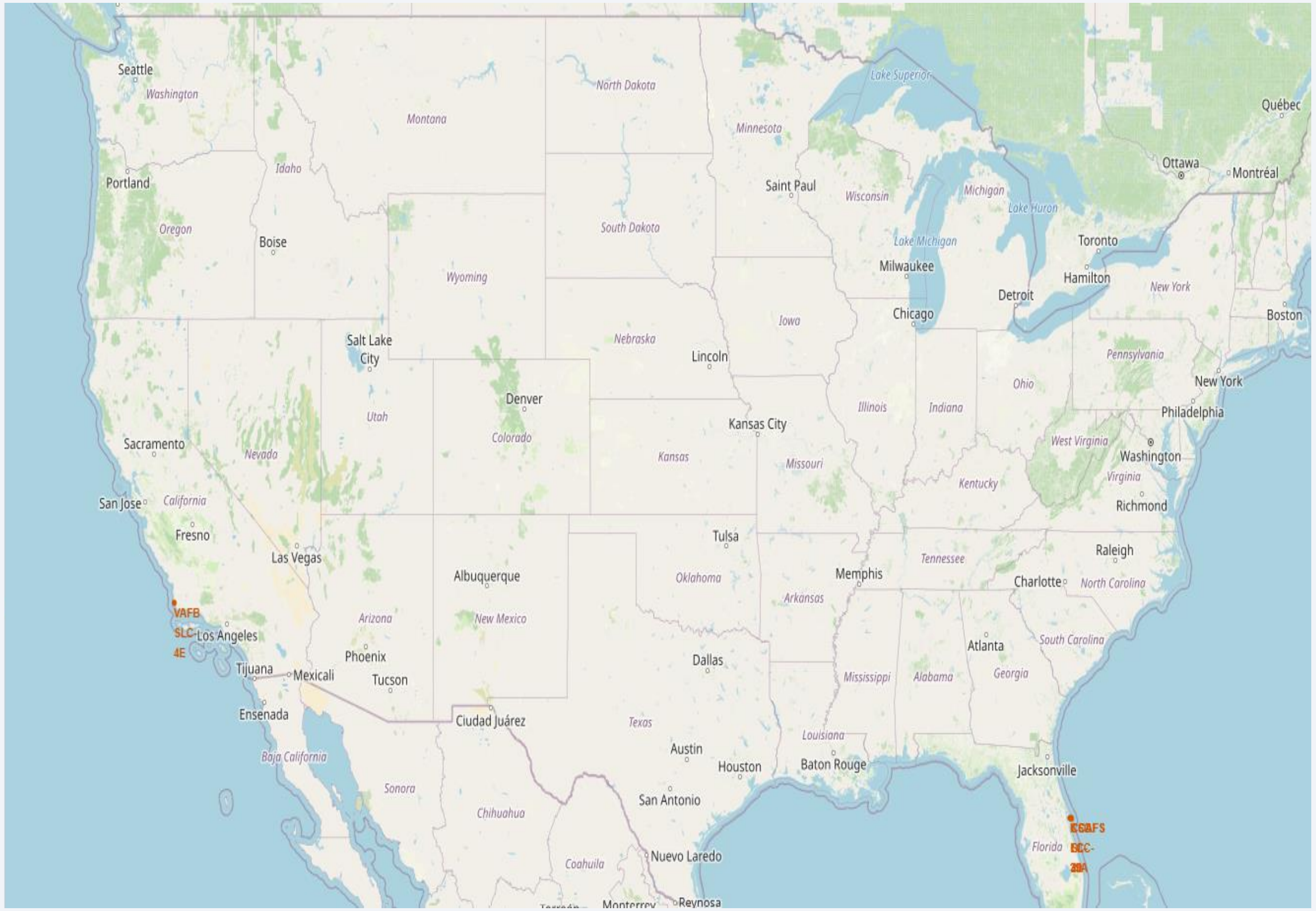
Landing_Outcome	count_outcomes
Success	20
No attempt	9
Success (drone ship)	8
Success (ground pad)	7
Failure (drone ship)	3
Failure	3
Failure (parachute)	2
Controlled (ocean)	2
No attempt	1

Section 3

Launch Sites Proximities Analysis

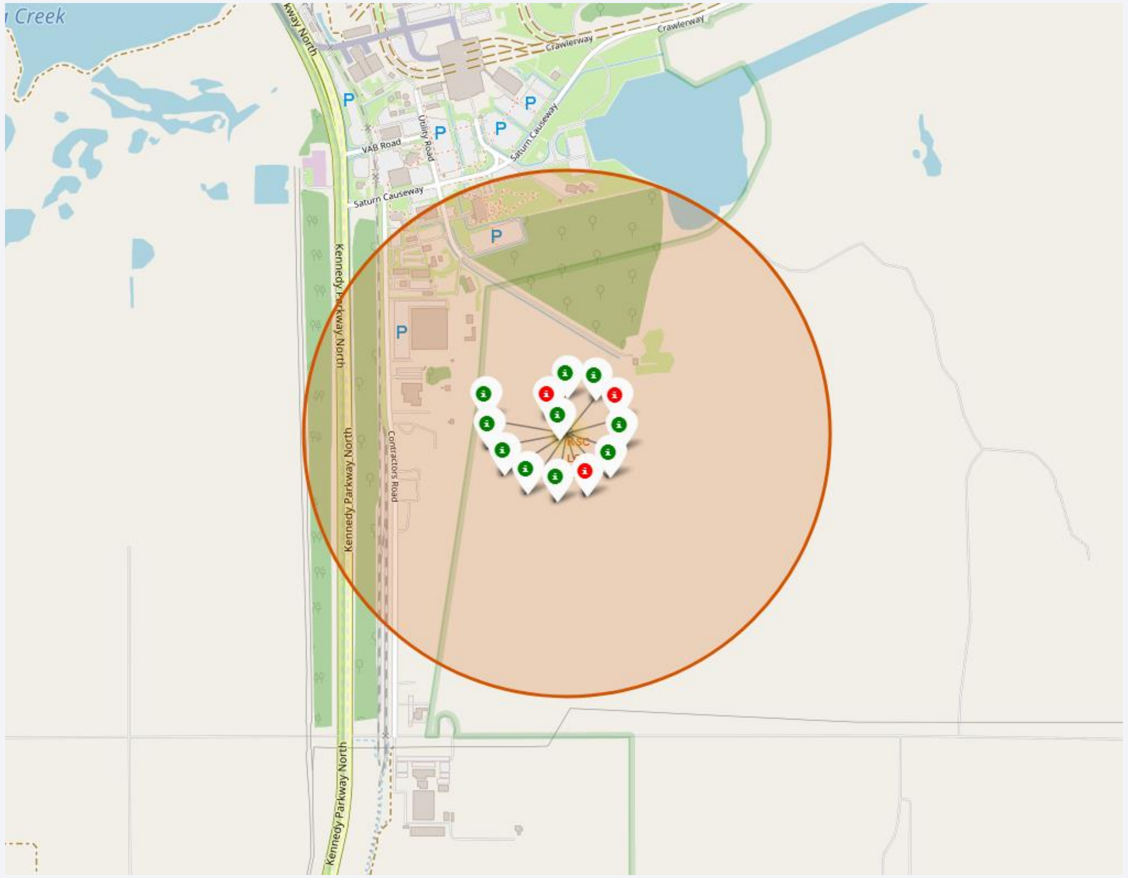
Launch Sites Location Markers

- All launches are in close proximity to the coast and in the southern half of the United States.
- This also makes them closer to the equator, allowing for the space craft to reach desired speed much quicker due to the Earth's rotation and inertia.



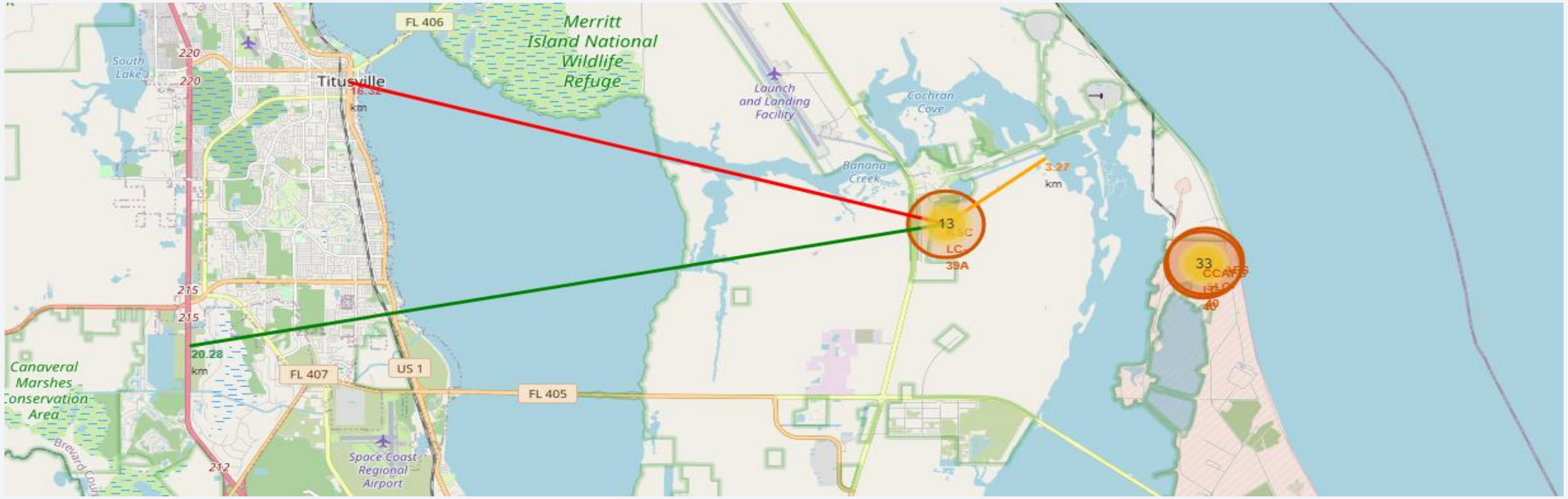
Color Coded Launch Markers

- The color labeled launch marker allow us to quickly identify launch sites with high success rates.
    - The higher the ratio of green markers to red indicates high success rate at a location.
- In this example, we can see launch site KSC LC-39A has a very high success rate.



Distances from Launch Sites

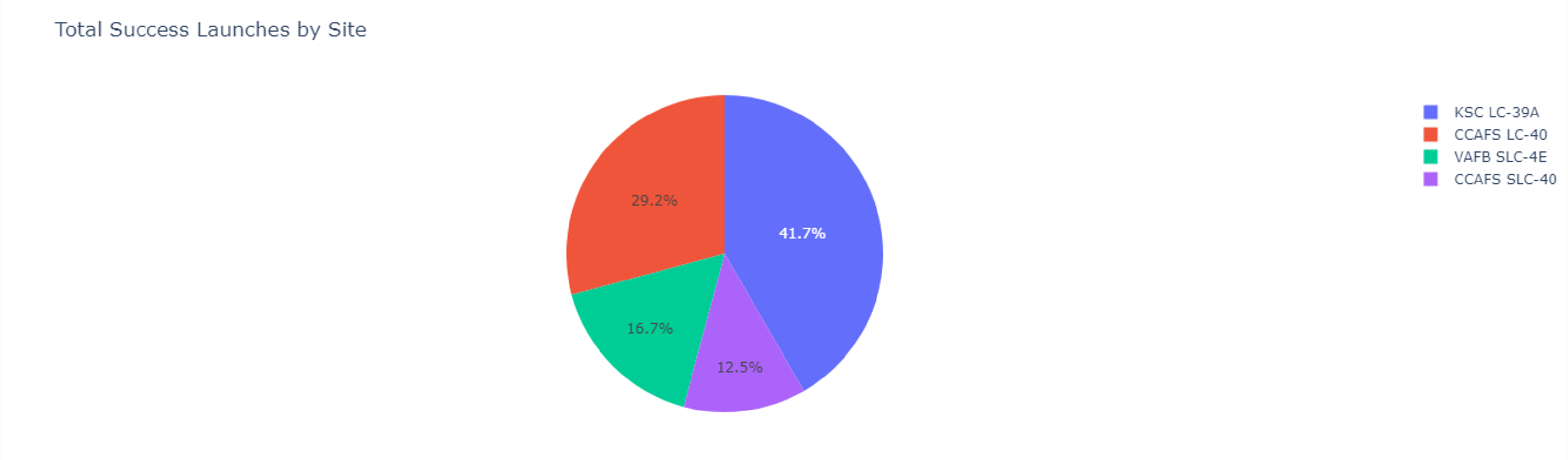
- These launch sites are relatively close to the coast, which allows room for error to the ocean, but they also have proximity to populated areas such as Titusville. This may be dangerous if a launch were to go wrong in the that direction.



Section 4

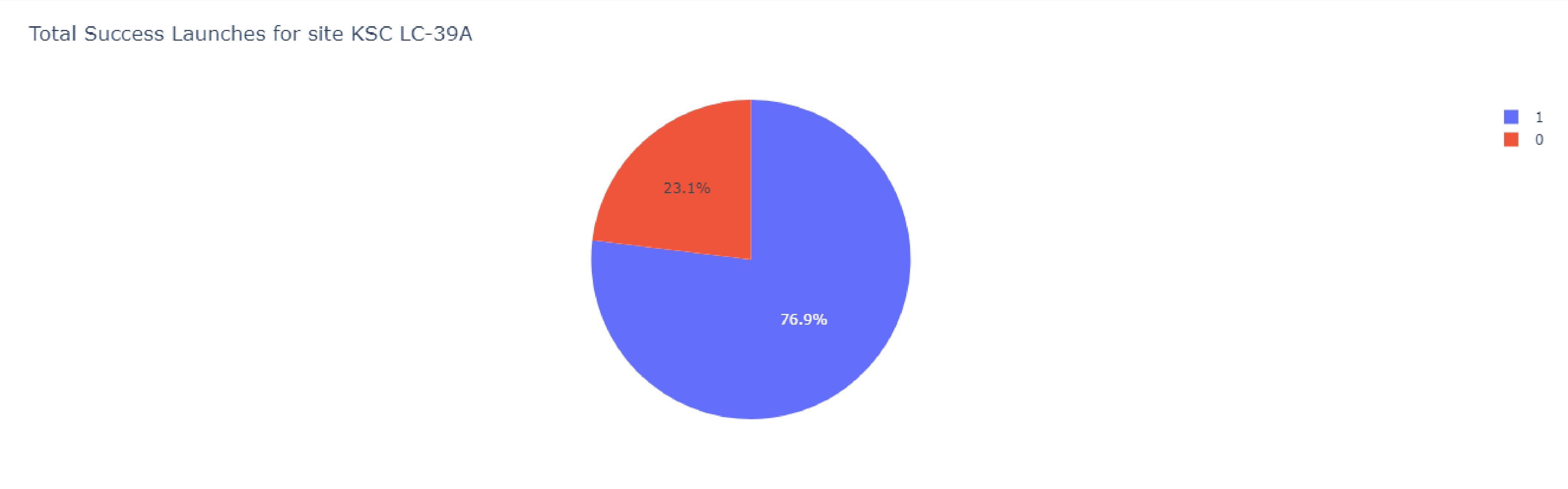
Build a Dashboard with Plotly Dash

Successful Launches by Site



- The pie chart shows that launch site KSC LC-39A the greatest percentage of successful launches out of all the sites.

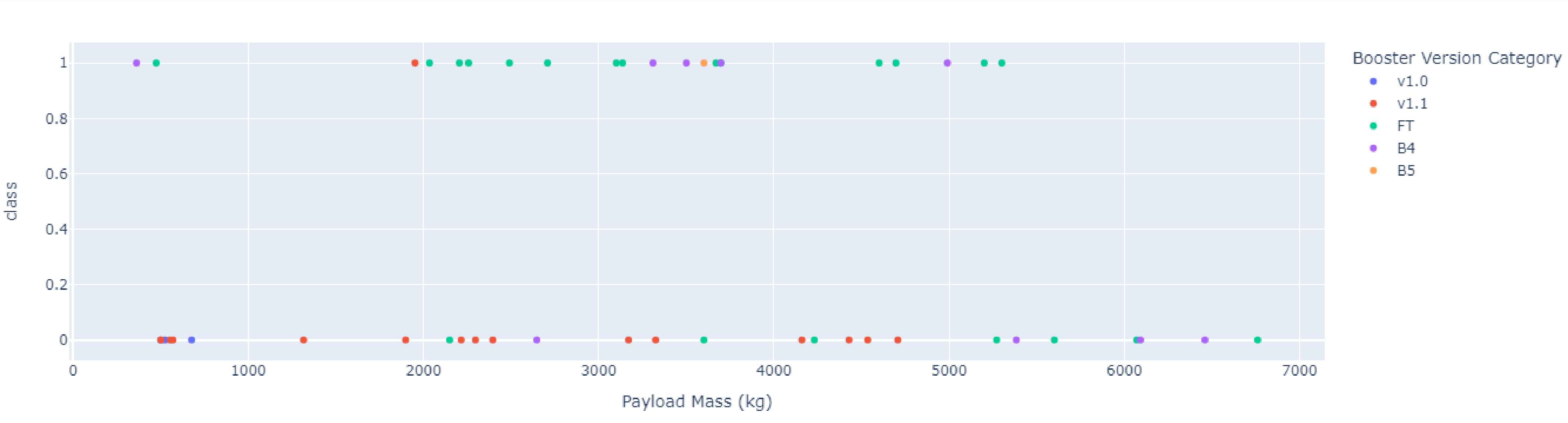
KSC LC-39A Success Ratio



- The launch site with the most successful launches, also has w very high success rate with 76.9% or 10 out of 13 successful launches.

Payload Mass and Success Correlation

- Regardless of Booster Version, the chart shows that the most successful payload range is between 2000 and 5050kg.

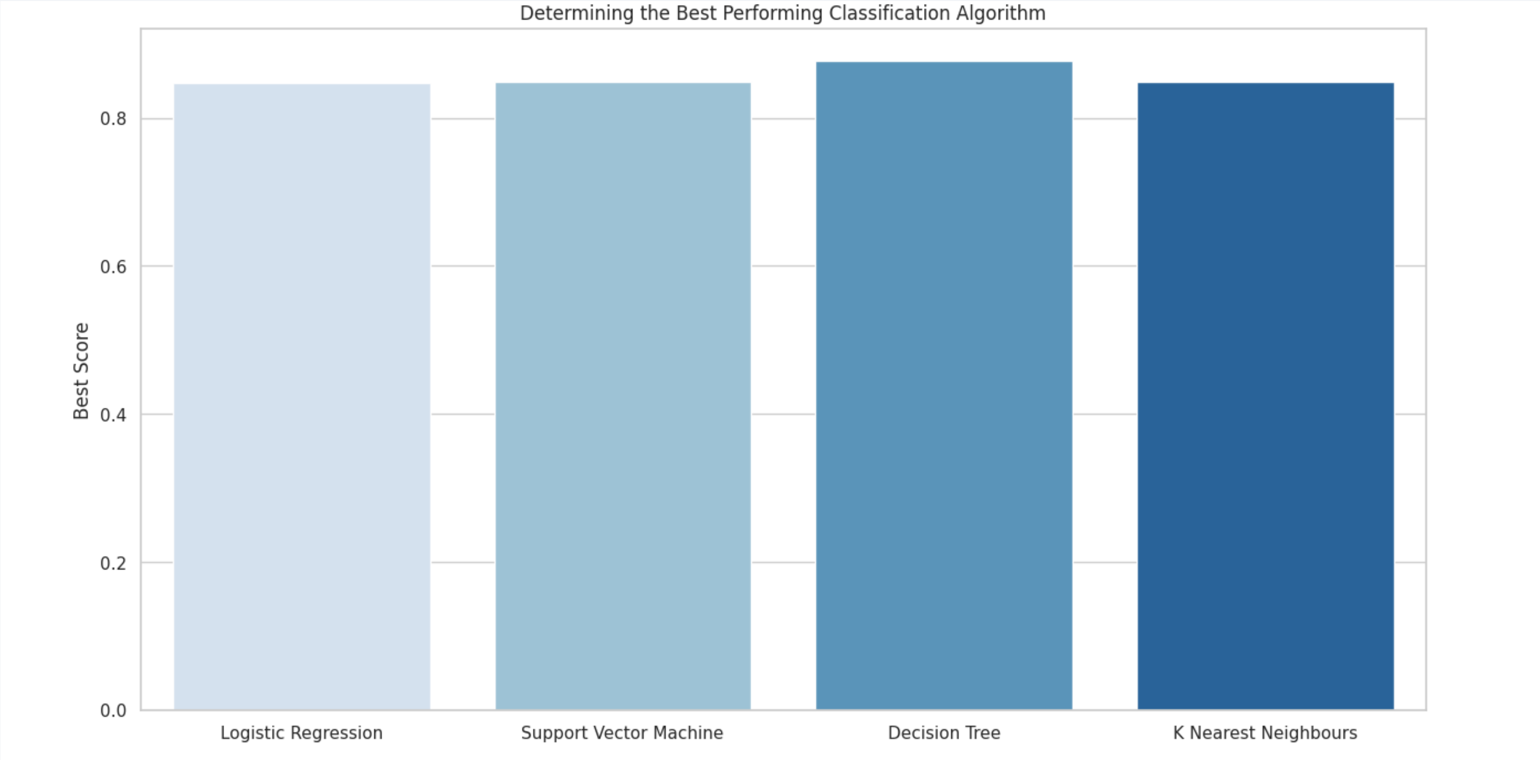


Section 5

Predictive Analysis (Classification)

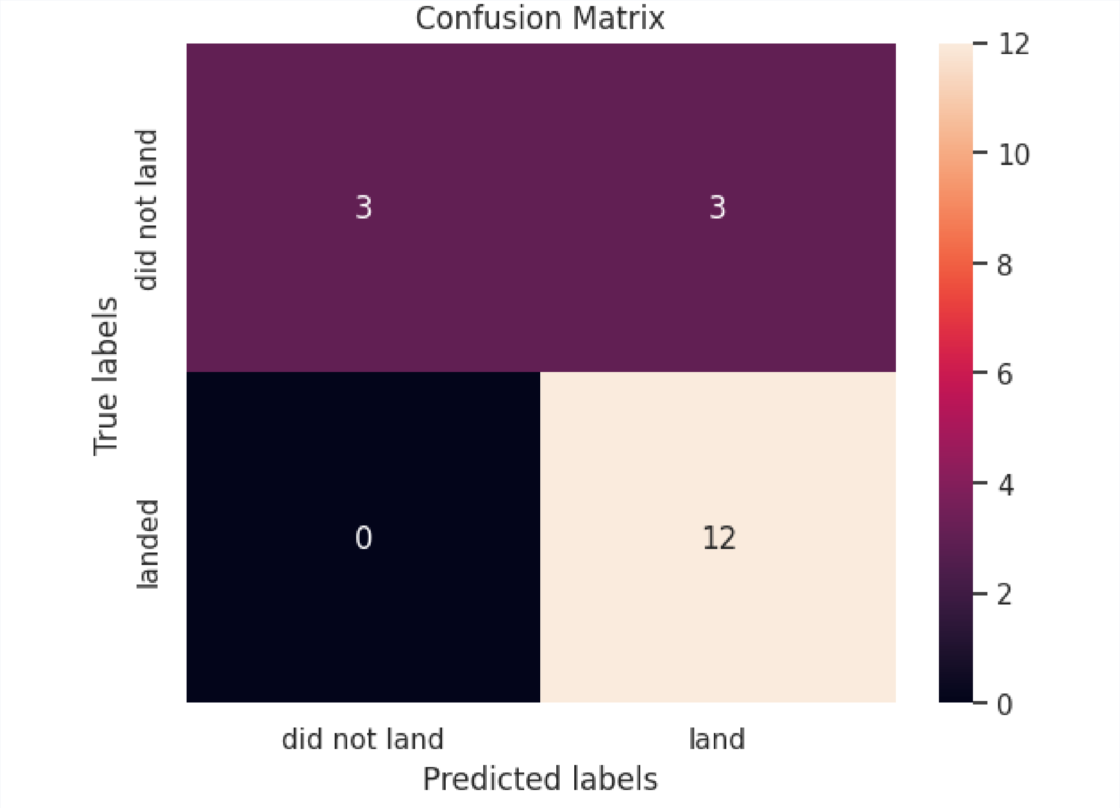
Classification Accuracy

- Based on the bar chart shown we can see that the decision tree model had the best accuracy score of 87.67%.



Confusion Matrix

- We can see that the Decision Tree model is very good at determining True Positives.
- There is however a major issue with false positives.



Conclusions

1. The Decision Tree model was the model that performed best on this set of data.
2. Launches have the highest rate of success with payloads between 2000 and 5050kg.
3. Success rate of launches has steadily increased with time.
4. KSC LC-39A is the most successful launch site to date, possibly due to its proximity to the equator.

Thank you!