IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Tom Rowley
07/08/2024

Outline

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

Executive Summary

- The data was collected using APIs and web scraping. Data wrangling was the performed followed by EDA and visuals were created to provide insights. Classification models were then used to determine success rate.
- The outcome of the launch is highly dependent on the orbit type, payload mass and flight number.
- The launch outcome is highly dependent on the launch site, booster type and payload mass.

Introduction

- Space X has many accomplishments in comparison to its competitors
  - Sending a spacecraft to the international space station
  - Starlink, a satellite internet constellation providing satellite internet access
  - Sending manned missions to space
- Space X can reuse first stage of Falcon 9 which keeps the cost far lower than competitors ($62m compared to $165m)
- First stage does majority of work
- Space X reuse the first stage by recovering it
- A machine learning model and public information was used to predict if SpaceX will reuse the first stage to see how competitors can keep costs low.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Describe how data was collected
- Perform data wrangling
  - Describe how data was processed
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection – SpaceX API

- A SpaceX API was used to collect some of the data
- This data was then cleaned using various python functions.
- [Applied-Data-Science-Capstone/jupyter-labs-spacex-data-collection-api.ipynb at main · trowley21/Applied-Data-Science-Capstone (github.com)]()

```
flowchart TD
A[Start] --> B[Client Request]
B --> C[Server Receives Request]
C --> D[Authentication & Authorization]
D --> E[Process Request]
E --> F[Send Response]
F --> G[Client Receives Response]
G --> H[End]
```

Data Collection - Scraping

- Falcon 9 records were extracted using web scraping
- These tables were then parsed and converted into a pandas data frame.
- [Applied-Data-Science-Capstone/jupyter-labs-webscraping.ipynb at main · trowley21/Applied-Data-Science-Capstone (github.com)]()

```
flowchart TD
A[Start] --> B[Send HTTP Request]
B --> C[Receive HTTP Response]
C --> D[Parse HTML Content]
D --> E[Handle Data (optional)]
E --> F[Save Data]
F --> G[End]
```

Data Wrangling

- Data Wrangling was then used to perform Exploratory Data Analysis (EDA)
- This found some patterns in the data and determined what would be the label for training supervised models.
- [Applied-Data-Science-Capstone/labs-jupyter-spacex-Data wrangling.ipynb at main · trowley21/Applied-Data-Science-Capstone (github.com)]

flowchart TD

A[Start] --> B[Understanding the Data]

B --> C[Data Assessment]

C --> D[Data Cleaning]

D --> E[Data Transformation]

E --> F[Data Integration]

F --> G[Data Validation]

G --> H[Documentation]

H --> I[End]

EDA with Data Visualization

The charts plotted were:

- Flight Number vs Payload Mass
- Flight Number vs Launch Site
- Payload Mass vs Launch Site
- The relationship between success rate of each orbit type
- Flight Number vs Orbit Type
- Payload vs Orbit Type
- Launch Success Yearly Trent

These charts were plotted to allow insights to be made into the effect of specific variables on landing success rate

[Applied-Data-Science-Capstone/jupyter-labs-eda-sql-coursera_python.ipynb at main · trowley21/Applied-Data-Science-Capstone (github.com)]

EDA with SQL

- The SQL queries were used to make calculations about the payload mass and insights into the missions performed along with outcomes.
- Queries were used to gain insights into the Launch sites.
- They were also used to learn more information about the relationship between booster versions used and payload mass.

- Applied-Data-Science-Capstone/jupyter-labs-eda-sql-coursera_sqllite.ipynb at main · trowley21/Applied-Data-Science-Capstone (github.com)

Build an Interactive Map with Folium

- Markers were placed on the map to indicate where each launch site is located.
- These were followed by marking the successful/failed launches for each launch site.
- The distances between the launch sites and its proximities, such as coast lines, railways, cities and highways, were then calculated and marked.
- These objects were added to gain better insights into the location of launch sites and its surroundings.
- Applied-Data-Science-Capstone/lab_jupyter_launch_site_location.ipynb at main · trowley21/Applied-Data-Science-Capstone (github.com)

Build a Dashboard with Plotly Dash

- A Dashboard was created with Plotly dash.
- Initially a Launch Site Drop-down Input Component was added so the user could choose which launch site to look at.
- A callback function was added to render a success pie chart based on the dropdown selection.
- A range slider was added to select payload to find if variable payload is correlated to mission outcome.
- Another callback function was added to render a scatter plot with variables mission outcome and payload mass to visually observe how payload may be correlated with mission outcomes for selected site(s

- Applied-Data-Science-Capstone/spacex_dash_app.py at main · trowley21/Applied-Data-Science-Capstone (github.com)

Predictive Analysis (Classification)

- Predictive analysis was performed to gain insights from the data and testing which model gave the best prediction.
- The data was initially normalized and then split into training and testing sets
- Each model was then tested with different parameters to see which would output the highest testing accuracy
- The models were then compared to see which gave the highest accuracy to predict landing success/failure.

- Applied-Data-Science-Capstone/SpaceX_Machine Learning Prediction_Part_5.ipynb at main · trowley21/Applied-Data-Science-Capstone (github.com)

A[Start] --> B[Define Objectives]

B --> C[Select Evaluation Metrics]

C --> D[Split Data]

D --> E[Train Models]

E --> F[Evaluate Models]

F --> G[Compare Performance]

G --> H[Select Best Model]

H --> I[End]

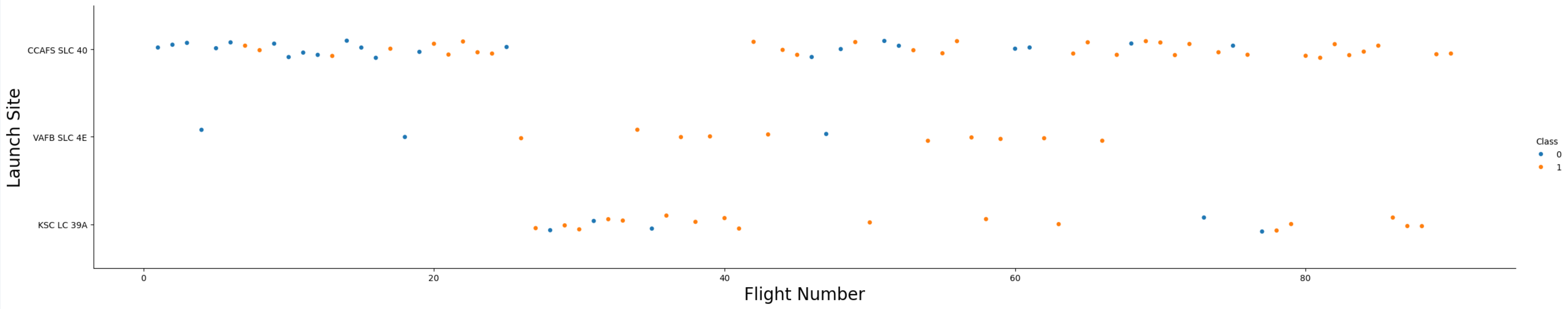
Results

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

Section 2

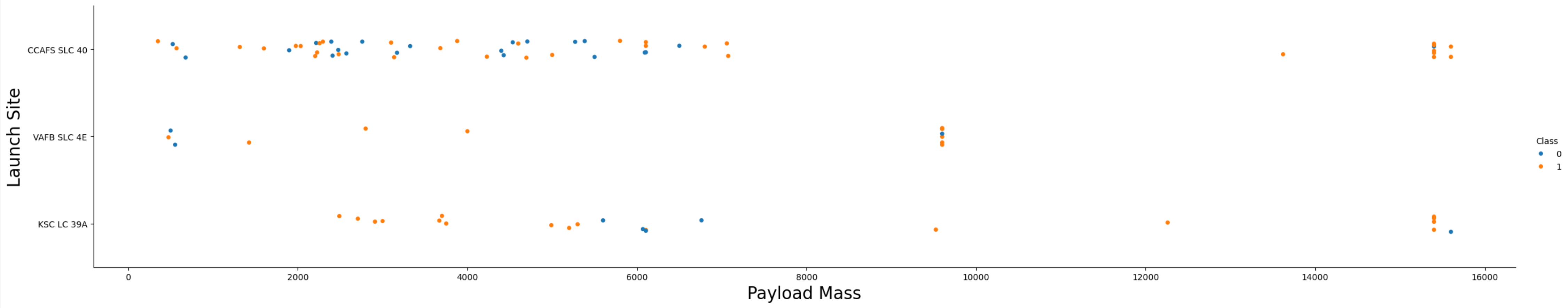
Insights drawn from EDA

Flight Number vs. Launch Site



- We can see that as the flight number increases, the first stage is more likely to land successfully regardless of Launch Site.
- We can also see that Launch Site 'VAFB SLC 4E' has the fewest launches.
- It can also be seen that this site does not have any flight numbers over 70.

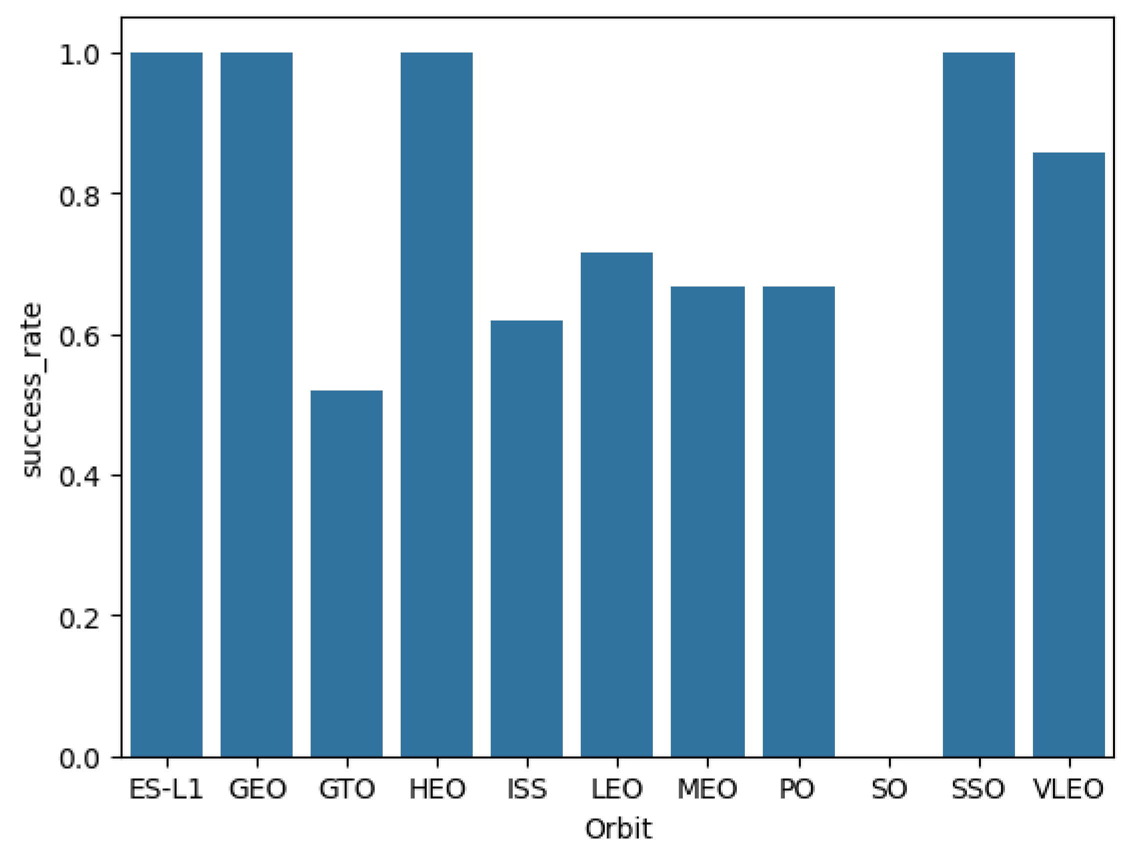
Payload vs. Launch Site



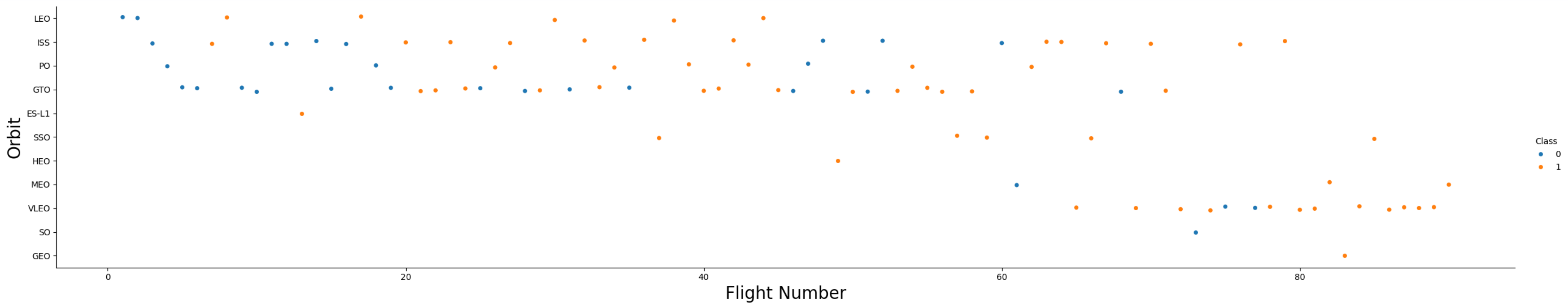
- It can be observed on the Payload Mass Vs. Launch Site scatter point chart that for the VAFB-SLC launch site there are no rockets launched for heavypayload mass(greater than 10000).
- Also it appears that a greater Payload Mass results in a greater chance of a successful launch.

Success Rate vs. Orbit Type

- The graph shows that ES-L1, GEO, HEO and SSO orbits have a 100% success rate.
- SO has a 0% success rate.
- GTO has the second lowest success rate with around 50%.
- The other orbit types range around 60%-70%.

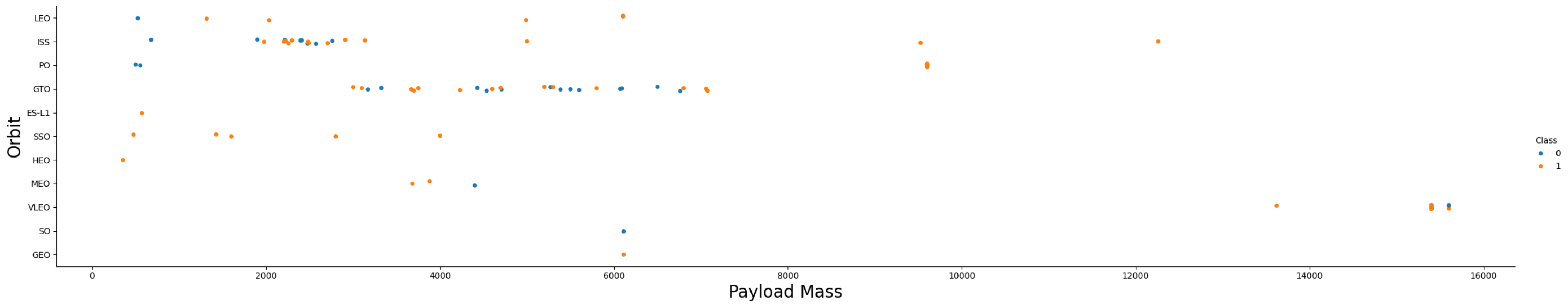


Flight Number vs. Orbit Type



- It can be observed that in the LEO orbit, success seems to be related to the number of flights.
- Conversely, in the GTO orbit, there appears to be no relationship between flight number and success.

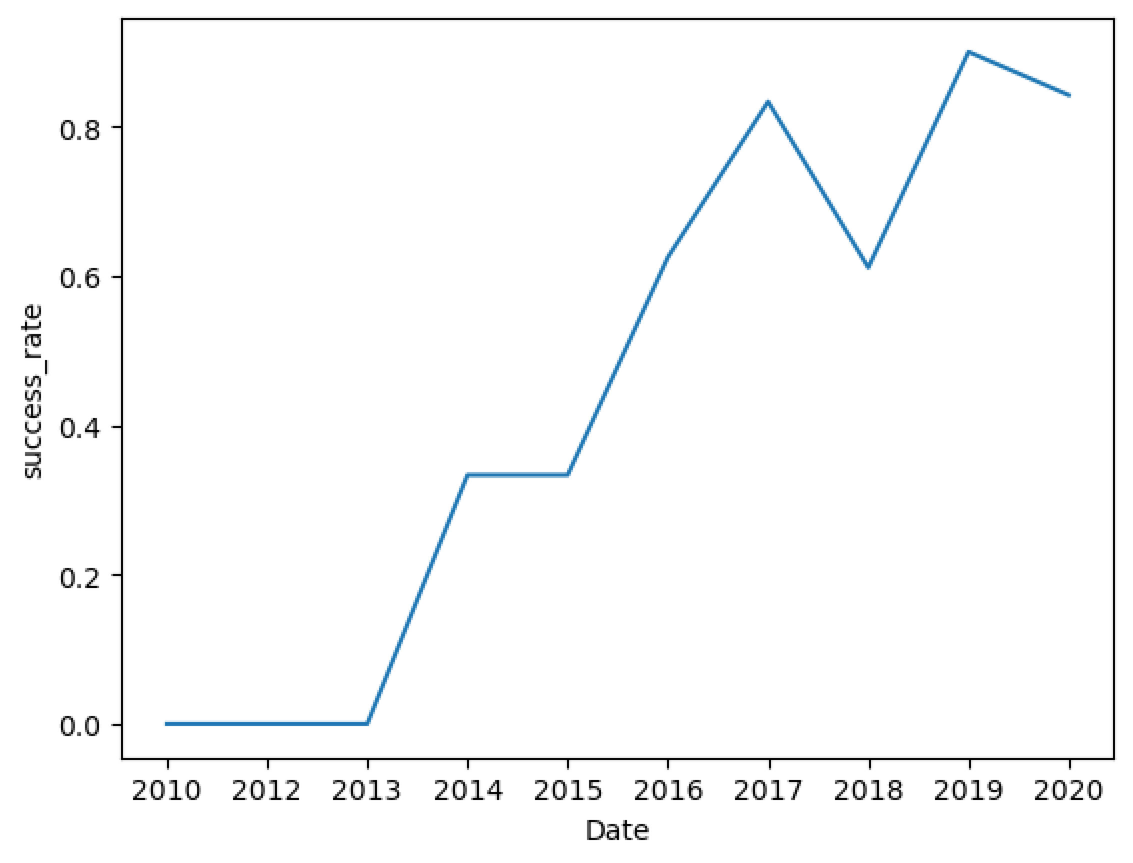
Payload vs. Orbit Type



- With heavy payloads the successful landing or positive landing rate are more for Polar, LEO and ISS.
- However, for GTO, it's difficult to distinguish between successful and unsuccessful landings as both outcomes are present.

Launch Success Yearly Trend

- It can be observed that the success rate has increased since 2013.
- However, there was a slight dip in 2018.



All Launch Site Names

- It can be seen that there are only 4 launch sites used for the Falcon 9 launches.

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

Launch Site Names Begin with 'CCA'

- It can be seen that 5 instances from launch sites beginning with 'CCA' were all LEO and the landing outcome was failure. However the mission outcome was success.

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

Total Payload Mass

- The total amount of payload mass used by NASA was calculated to be 45596kg.

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

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

total_amount
45596

Average Payload Mass by F9 v1.1

- The average Payload mass carried by the F9 v1.1 booster version was calculated to be 2928.4kg

```
%%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

First Successful Ground Landing Date

- The first successful ground landing date was found out to be 2015-12-22.

```
%%sql SELECT min(Date)
From SPACEXTBL
where Landing_Outcome = 'Success (ground pad)';
```

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

min(Date)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- The names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000 are listed in the table to the right of the page.

```
%%sql SELECT distinct(Booster_Version)
From SPACEXTBL
where PAYLOAD_MASS__KG_ > 4000 AND PAYLOAD_MASS__KG_ < 6000;
```

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

Booster_Version
F9 v1.1
F9 v1.1 B1011
F9 v1.1 B1014
F9 v1.1 B1016
F9 FT B1020
F9 FT B1022
F9 FT B1026
F9 FT B1030
F9 FT B1021.2
F9 FT B1032.1
F9 B4 B1040.1
F9 FT B1031.2
F9 B4 B1043.1
F9 FT B1032.2
F9 B4 B1040.2
F9 B5 B1046.2
F9 B5 B1047.2
F9 B5B1054
F9 B5 B1048.3
F9 B5 B1051.2
F9 B5B1060.1
F9 B5 B1058.2
F9 B5B1062.1

Total Number of Successful and Failure Mission Outcomes

- The total number of mission successes was 100 and the total number of failures was 1.

```
%%sql SELECT
    COUNT(CASE WHEN Mission_Outcome LIKE '%success%' THEN 1 END) AS total_success,
    COUNT(CASE WHEN Mission_Outcome LIKE '%failure%' THEN 1 END) AS total_failure
FROM SPACEXTBL;
```

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

total_success	total_failure
100	1

Boosters Carried Maximum Payload

- The names of the booster which have carried the maximum payload mass are listed on the table to the right of the page.

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

```
 * 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

2015 Launch Records

- The failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015 are listed in the table to the bottom right of this page.

```
%%sql SELECT
    CASE SUBSTR(Date, 6, 2)
        WHEN '01' THEN 'January'
        WHEN '02' THEN 'February'
        WHEN '03' THEN 'March'
        WHEN '04' THEN 'April'
        WHEN '05' THEN 'May'
        WHEN '06' THEN 'June'
        WHEN '07' THEN 'July'
        WHEN '08' THEN 'August'
        WHEN '09' THEN 'September'
        WHEN '10' THEN 'October'
        WHEN '11' THEN 'November'
        WHEN '12' THEN 'December'
    END AS month_name,
    Landing_Outcome,
    Booster_Version,
    Launch_Site
FROM SPACEXTBL
WHERE
    SUBSTR(Date, 1, 4) = '2015' AND
    Landing_Outcome = 'Failure (drone ship)';
```

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

month_name	Landing_Outcome	Booster_Version	Launch_Site
January	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
April	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

- A Ranked 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 are listed in the table at the bottom right of this page.
- No attempt is the highest with 10 and precluded (drone ship) is bottom with 1.

```
%%sql WITH OutcomeCounts AS (
    SELECT
        Landing_Outcome,
        COUNT(*) AS count
    FROM SPACEXTBL
    WHERE Date BETWEEN '2010-06-04' AND '2017-03-20'
    GROUP BY Landing_Outcome
),
RankedOutcomes AS (
    SELECT
        Landing_Outcome,
        count,
        RANK() OVER (ORDER BY count DESC) AS rank
    FROM OutcomeCounts
)
SELECT
    Landing_Outcome,
    count,
    rank
FROM RankedOutcomes
ORDER BY rank;
```

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

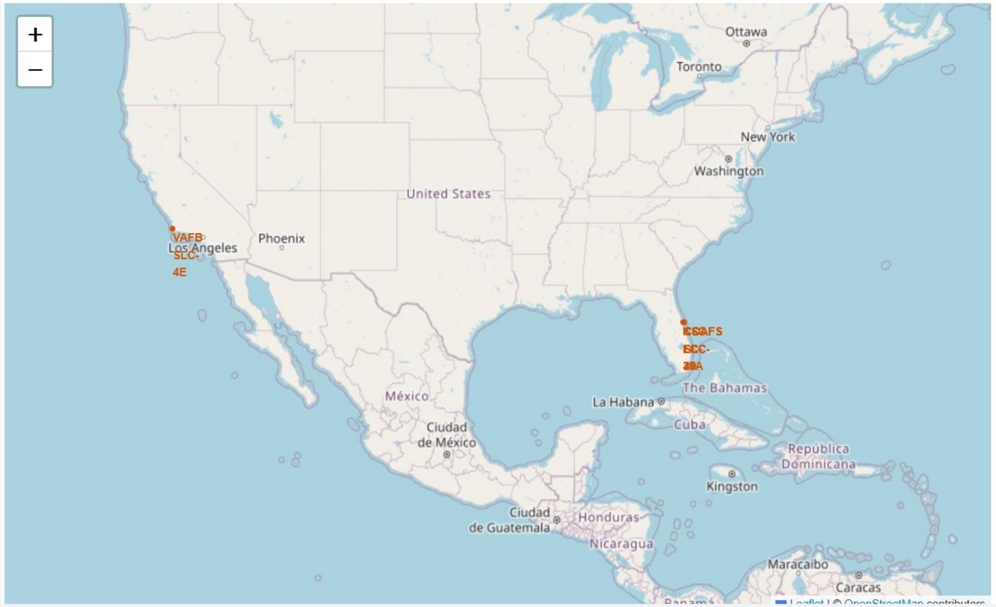
Landing_Outcome	count	rank
No attempt	10	1
Failure (drone ship)	5	2
Success (drone ship)	5	2
Controlled (ocean)	3	4
Success (ground pad)	3	4
Failure (parachute)	2	6
Uncontrolled (ocean)	2	6
Precluded (drone ship)	1	8

Section 3

Launch Sites Proximities Analysis

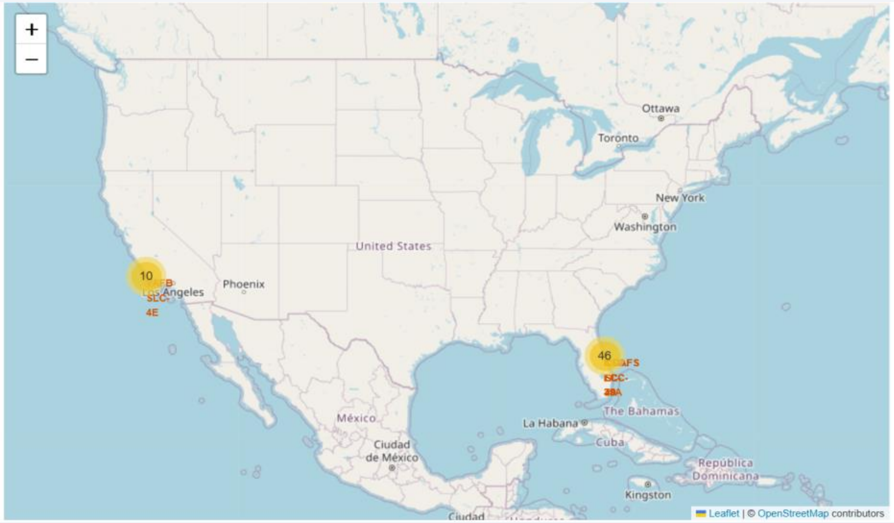
Map Showing Labels of Launch Sites

- All of the launch sites are in the southern reigon of USA which means they wre close in proximity to the equator line.
- Also, all of the launch sites are in very close proximity to the coast being in the states of Florida and California.



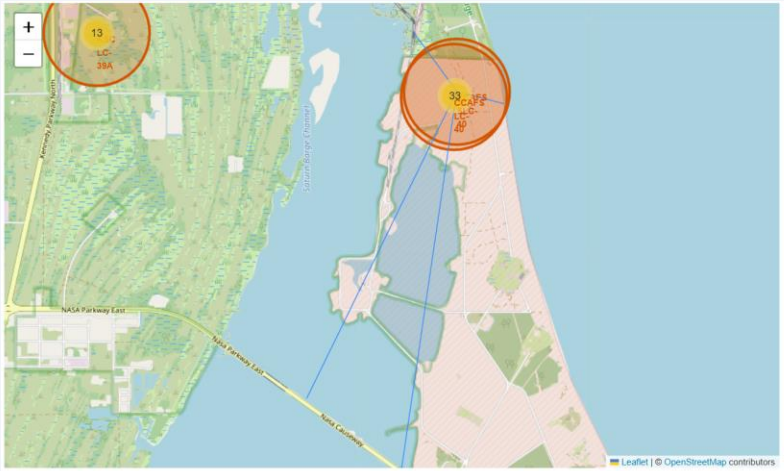
A Map Showing Successful and Failed Launch Attempts

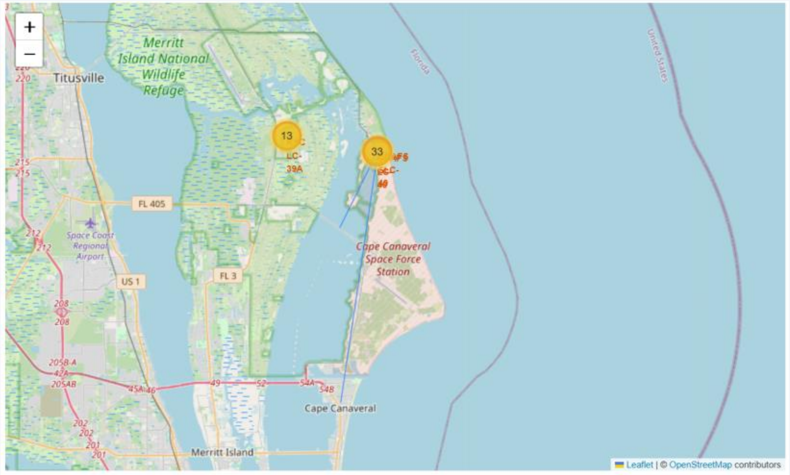
- The maps on the left show the locations of the launch sites.
- This is along with the successful and failed launch attempts.
- It can be seen that CCAFS SLC-40 has more failed launches than successful launches.



A Map to Show Launch Sites and their Proximities

- The nearest railway is called the 'NASA Railcard' and it could be assumed that ,during a launch, this railway would not be in use. However, as it is in close proximity to the launch site, there is potential for it to be damaged during a failed launch.
- The closest highway is the 'NASA Parkway East' and ,once again, it could be assumed that this would be under careful watch during a launch. However, it is a considerable distance away from the launch site so will probably be safe regardless of if there is a launch or not.
- The closest city is Cape Canaveral. Cape Canaveral is a considerable distance away from the launch site and also has 'Rocket Launch Viewpoints' which implies that it is a safe enough distance away for people during a launch.



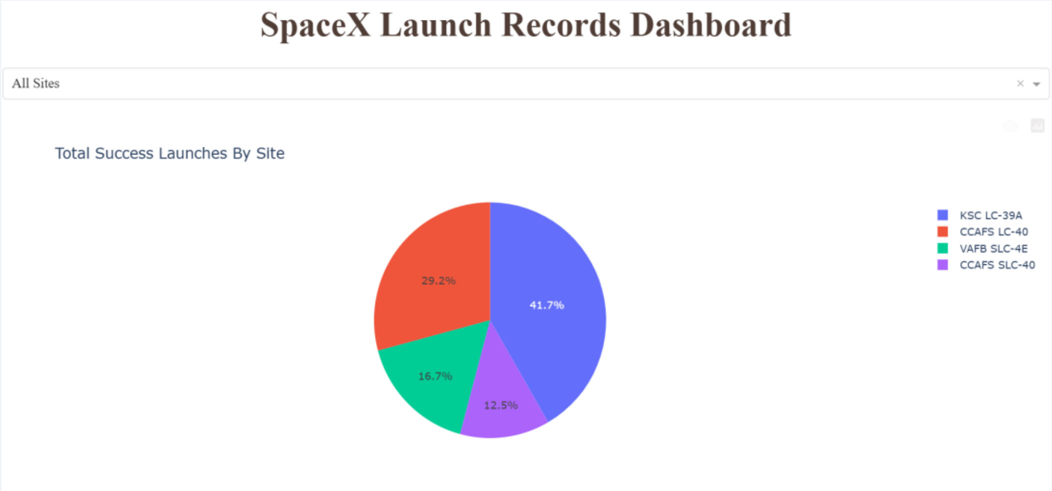


Section 4

Build a Dashboard with Plotly Dash

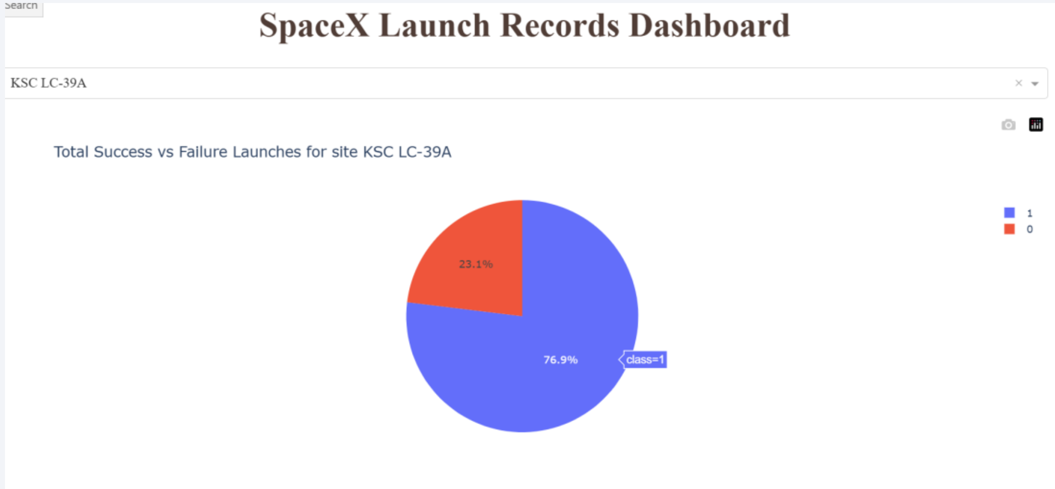
Launch Success Count Piechart

- It can be seen from the screenshot to the right of the page that KSC LC-39A has the highest percent of total successful launches with 41.7%
- It can also be seen that CCAFS SLC-40 has the lowest percent with 12.5%.



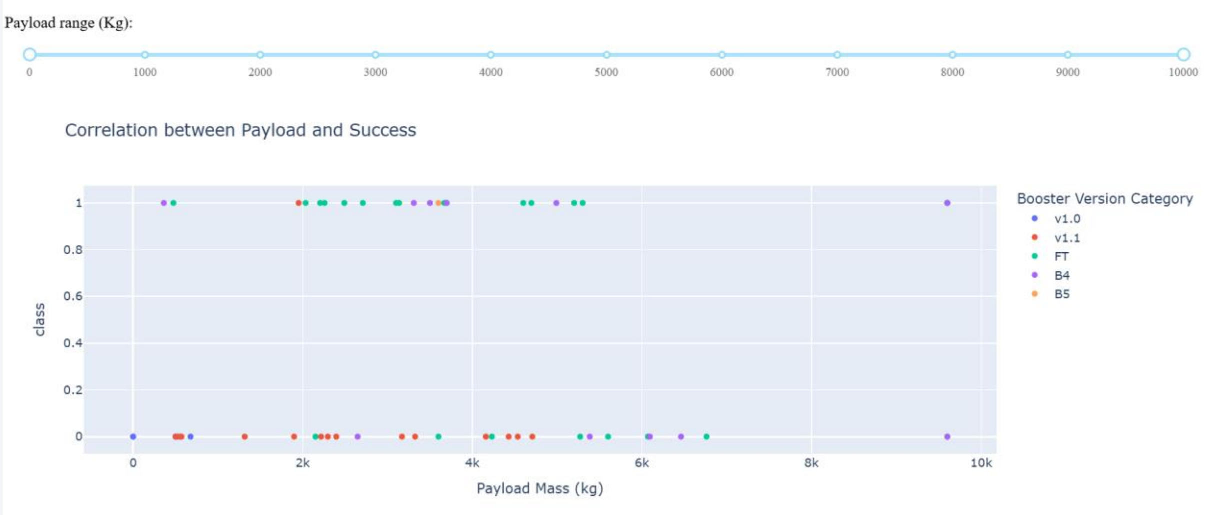
Launch Success from KSC LC-39A

- It can be seen from the pie chart to the right of the page that the success percentage from KSC LC-39A is 76.9%



Payload vs. Launch Outcome scatter plot for all sites

- It can be seen from the graph below that the majority of launches have payload masses below 6000kg
- It can also be seen that there are more failed launches than successful ones.
- B5 has 1 launch which was successful
- V1.1 has 15 total launches and 1 successful
- V1 has 0 successful launches out of 4 attempts
- FT has more successful launches than failed launches.

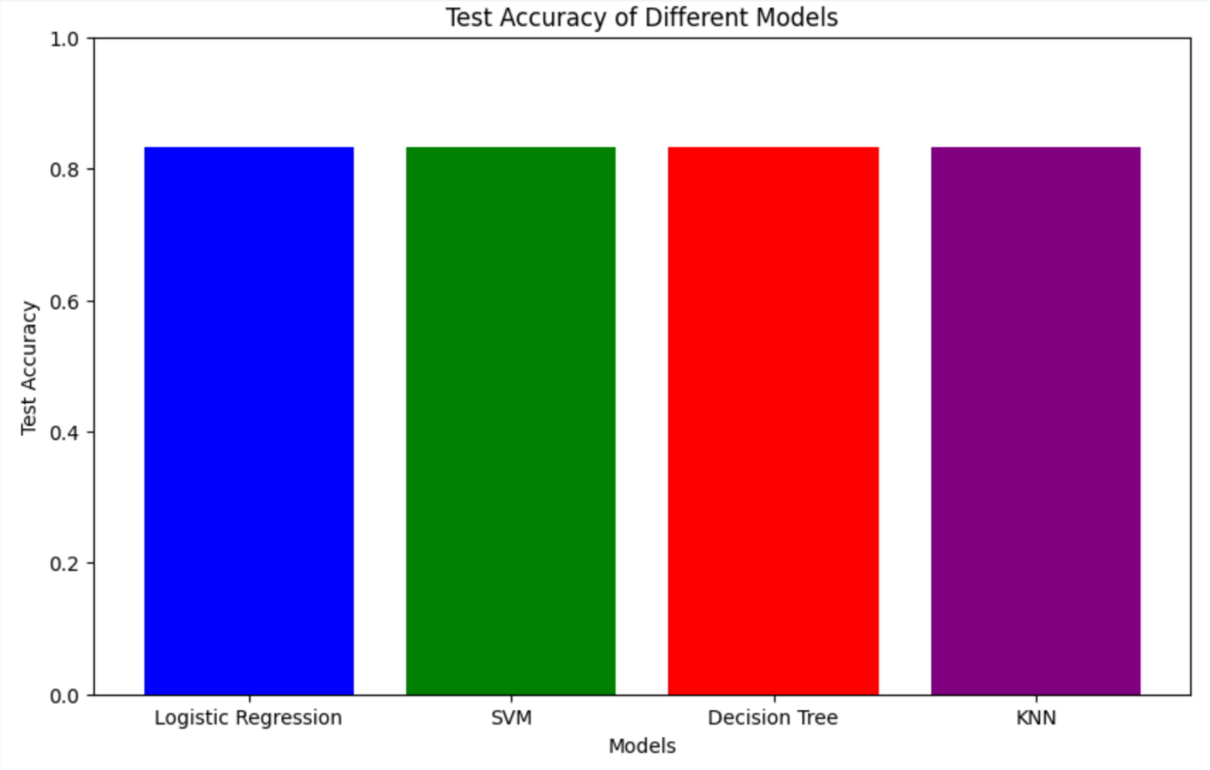


Section 5

Predictive Analysis (Classification)

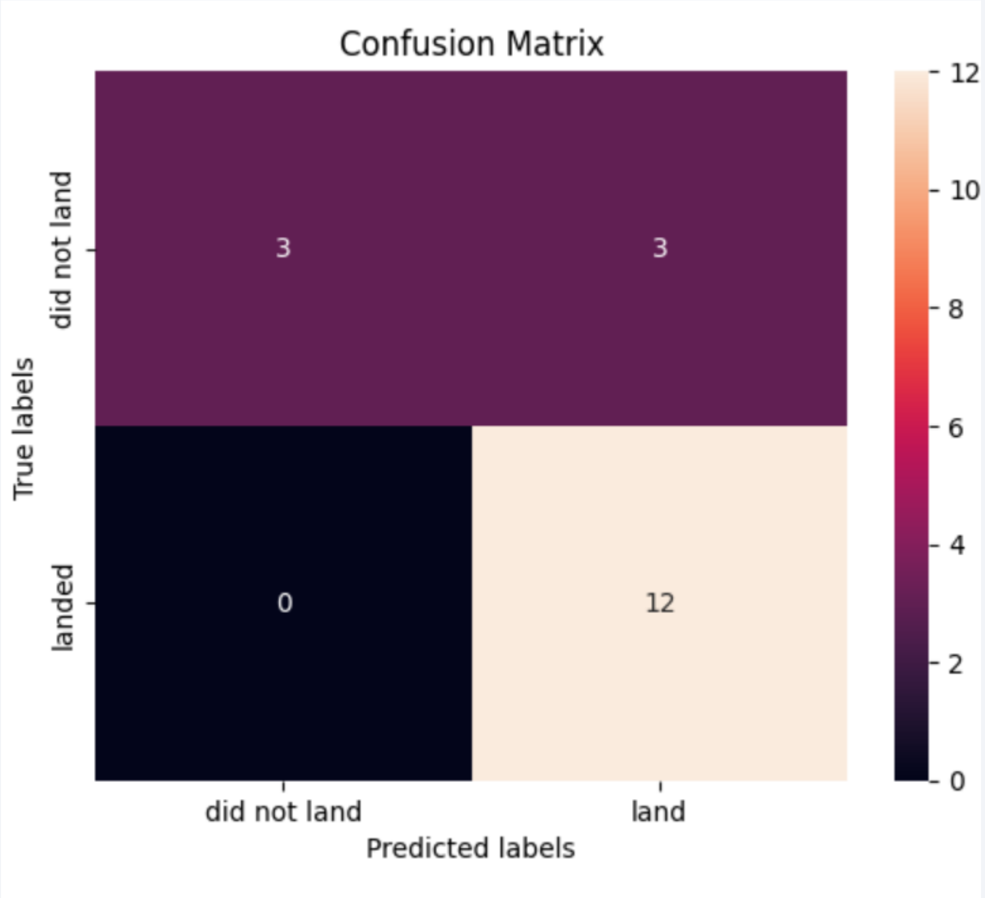
Classification Accuracy

- In the bar chart it can be seen that all of the models have the same testing accuracy of 0.8334.



Confusion Matrix

- To the right is the confusion matrix of the Logistic Regression used on the testing data.
- It can be seen that the false negatives were 0
- However, the false positives were 3.



Conclusions

- The success rate is improving for the landing of space crafts
- KSC LC-39A is the best booster to use with the highest launch success
- A higher flight number will provide a higher success rate
- ES-L1, GEO, HEO and SSO orbits are the optimal orbits to use.

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!