APPLIED DATA SCIENCE CAPSTONE

Data Science and SpaceX

By Cody Meuris

June 25, 2024

OUTLINE

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

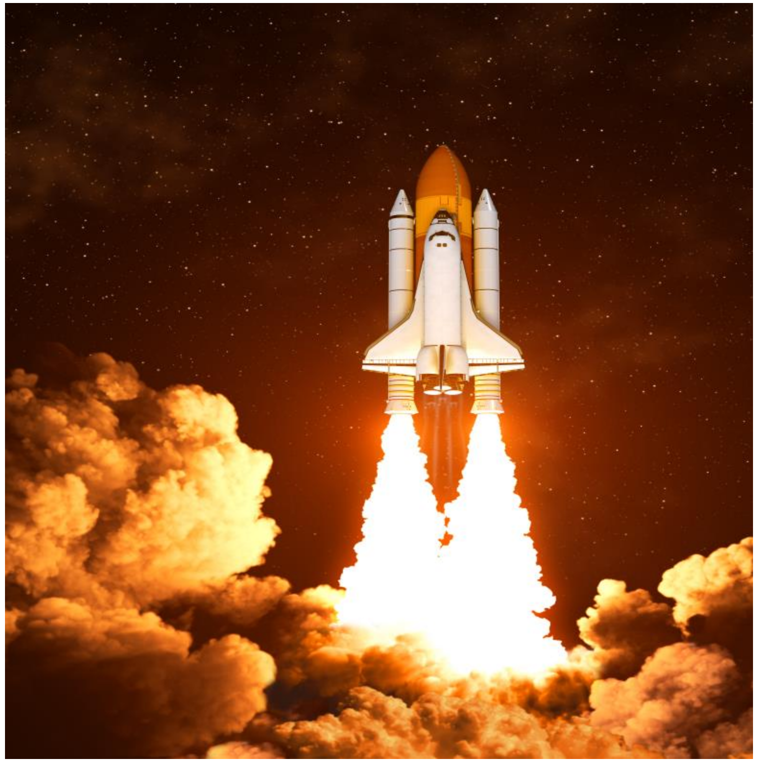
EXECUTIVE SUMMARY

- Methodologies
  - Data Collecting
  - Data Wrangling
  - EDA with Data Visualization
  - EDA with SQL
  - Building an Interactive Map with Folium
  - Building a Dashboard with Plotly Dash
  - Predictive Analysis

Results
  - EDA Results
  - Interactive Analysis
  - Predictive Analysis

INTRODUCTION

- The SpaceX falcon 9 launch has the ability to reuse its first stage, saving it a considerable amount of money compared to its competitors.
- Our goal is to use data science methods to predict the outcome of SpaceY launch by analyzing the SpaceX data available to us.



METHODOLOGY

Data Collection

- Make a get request to the SpeceX API
- Clean the requested data

Data Wrangling

- Exploratory data analysis
- Determine training labels

Exploratory Data Analysis (EDA)

- Visualization and SQL visualization

Interactive Visual Analytics

- Plotly and Dash

Predictive Analytics

- Classification Models
  - Logistic Regression
  - Decision Tree
  - Support Vector Machine
  - K Nearest Neighbors

DATA COLLECTION

- We collected the data using get request to the SpaceX API
- We decoded the response content as a Json using **.json**() function call and turn it into a pandas dataframe using **.json_normalize()**
- We cleaned the data, checked for missing values, and filled in missing values
- We performed web scraping for Falcon 9 launch records with BeautifulSoup

https://github.com/codycoursera/Capstone/blob/Capstone-Main-Branch/jupyter-labs-spacex-data-collection-api.ipynb

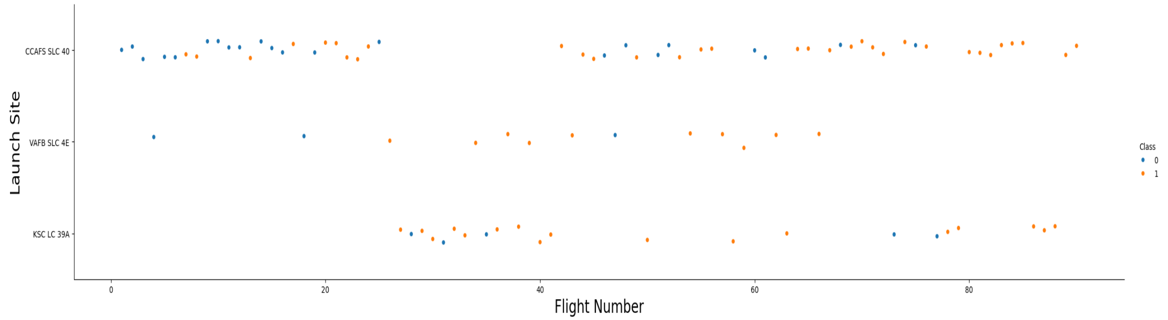
DATA WRANGLING

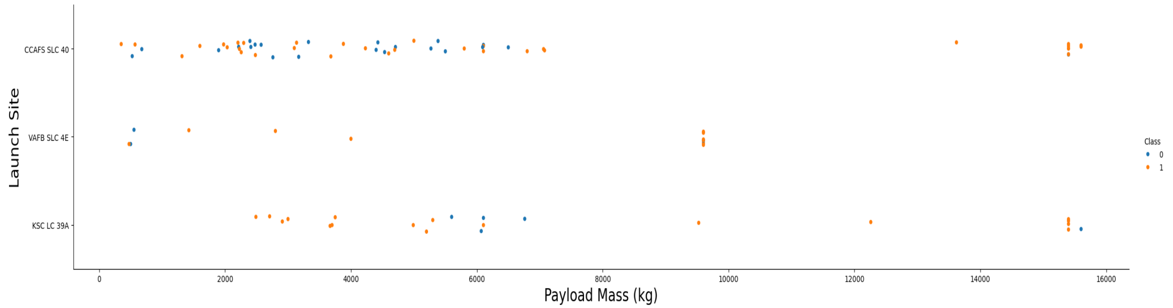
- We calculated the number of launches on each site
- We calculated the number and occurrence of each orbit
- We calculated the number and occurrence of mission outcome of the orbits
- Lastly, we created a landing outcome label from Outcome column

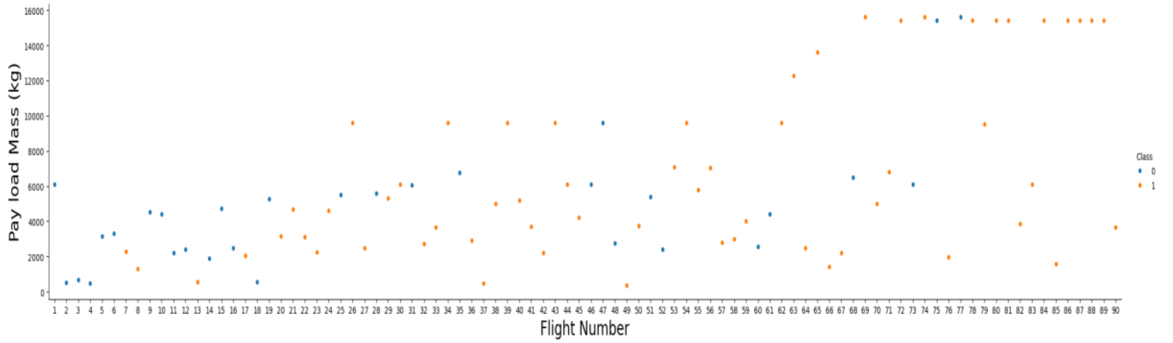
https://github.com/codycoursera/Capstone/blob/Capstone-Main-Branch/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA WITH DATA VISUALIZATION

- Flight Number vs. Payload Mass
- Payload Mass vs. Launch Site
- Flight Number vs. Pay Load Mass



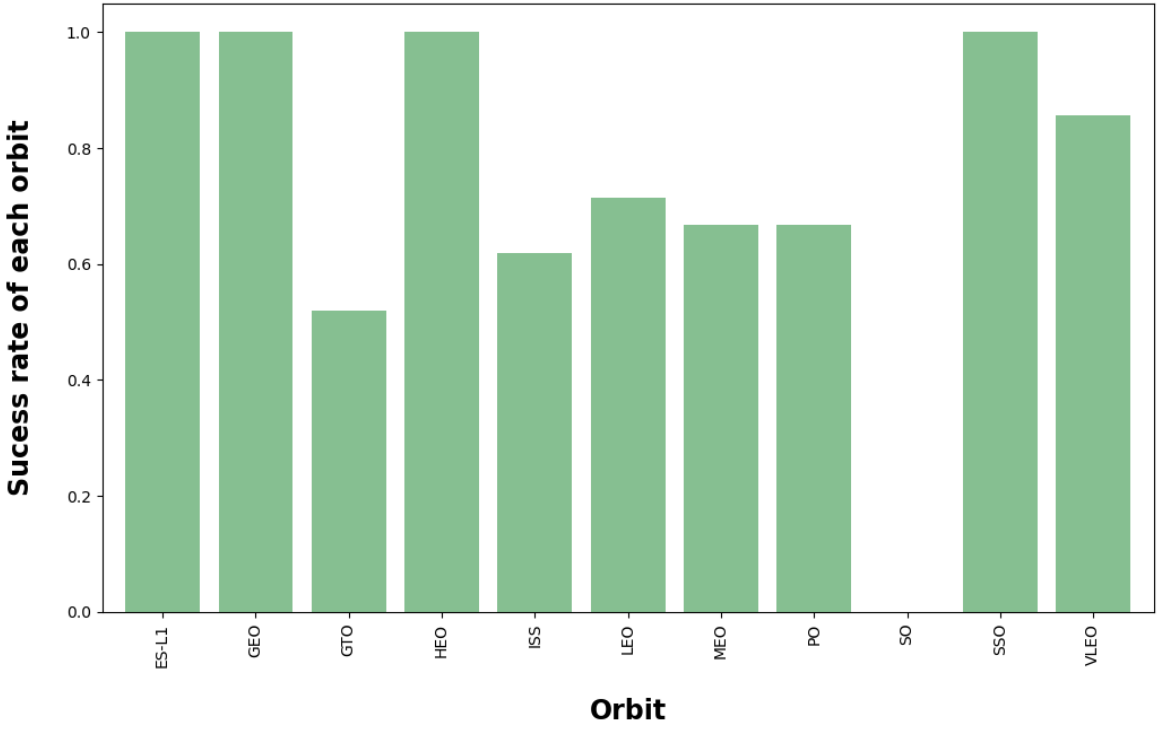


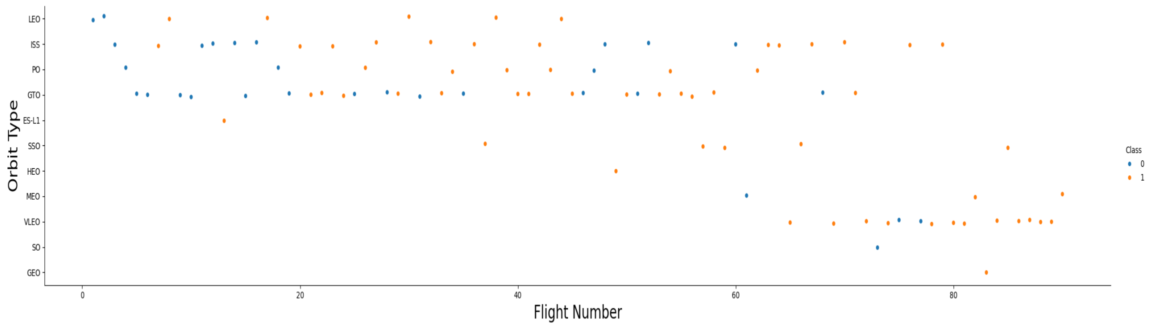


https://github.com/codycoursera/Capstone/blob/Capstone-Main-Branch/edadataviz%20(1).ipynb

EDA WITH DATA VISUALIZATION

- Success Rate of Each Orbit
- Flight Number vs. Orbit Type

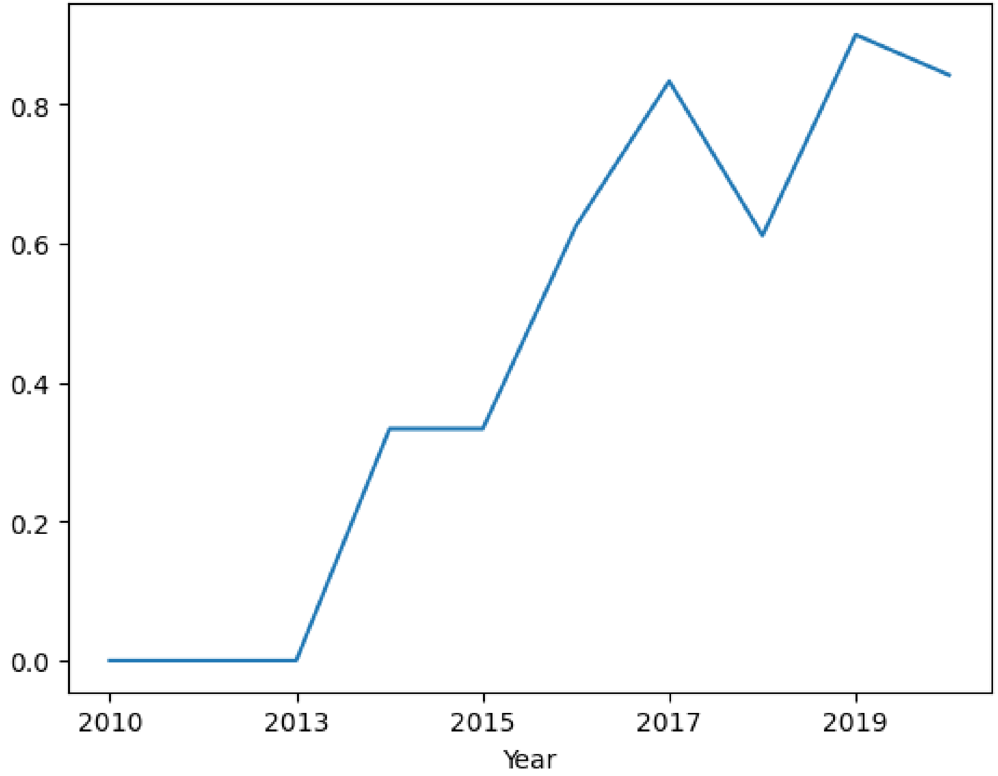


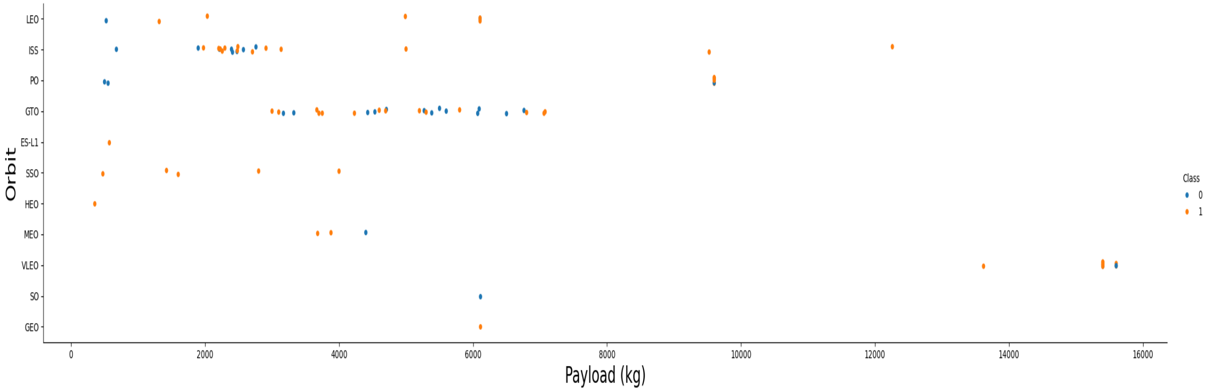


https://github.com/codycoursera/Capstone/blob/Capstone-Main-Branch/edadataviz%20(1).ipynb

EDA WITH DATA VISUALIZATION

- Launch success yearly trend
- Payload vs. Orbit





https://github.com/codycoursera/Capstone/blob/Capstone-Main-Branch/edadataviz%20(1).ipynb

EDA WITH SQL

SQL Queries

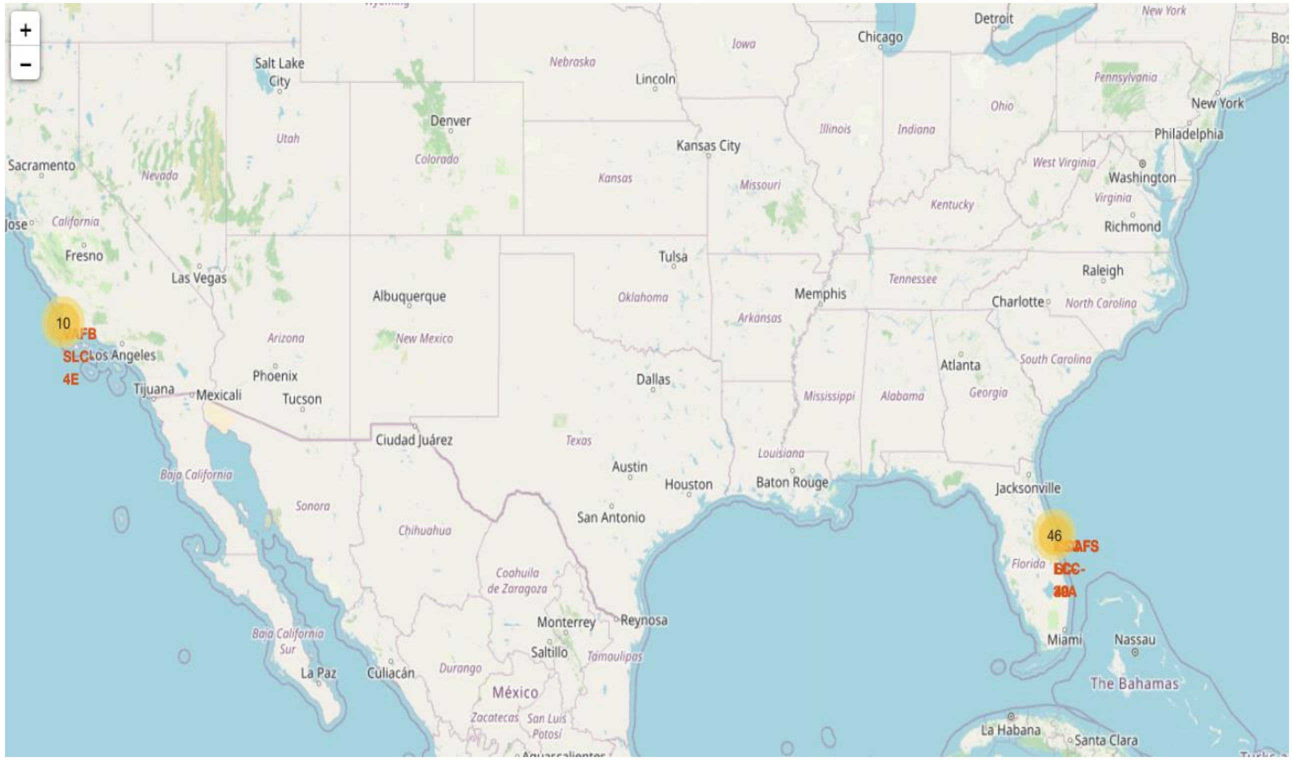
https://github.com/codycoursera/Capstone/blob/Capstone-Main-Branch/jupyter-labs-eda-sql-coursera_sqllite.ipynb

- Display the names of the unique launch sites in the space mission
- Display 5 records where launch sites begin with the string 'CCA'
- Display the total payload mass carried by boosters launched by NASA (CRS)
- Display average payload mass carried by booster version F9 v1.1
- List the date when the first succesful landing outcome in ground pad was acheived
- List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- List the total number of successful and failure mission outcomes
- List the total number of successful and failure mission outcomes
- List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.
- 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.

BUILDING AN INTERACTIVE MAP WITH FOLIUM

- We visualized the data into an interactive map by taking the latitude and longitude coordinates at each launch site and added a circle marker around them with a label of the name and launch site

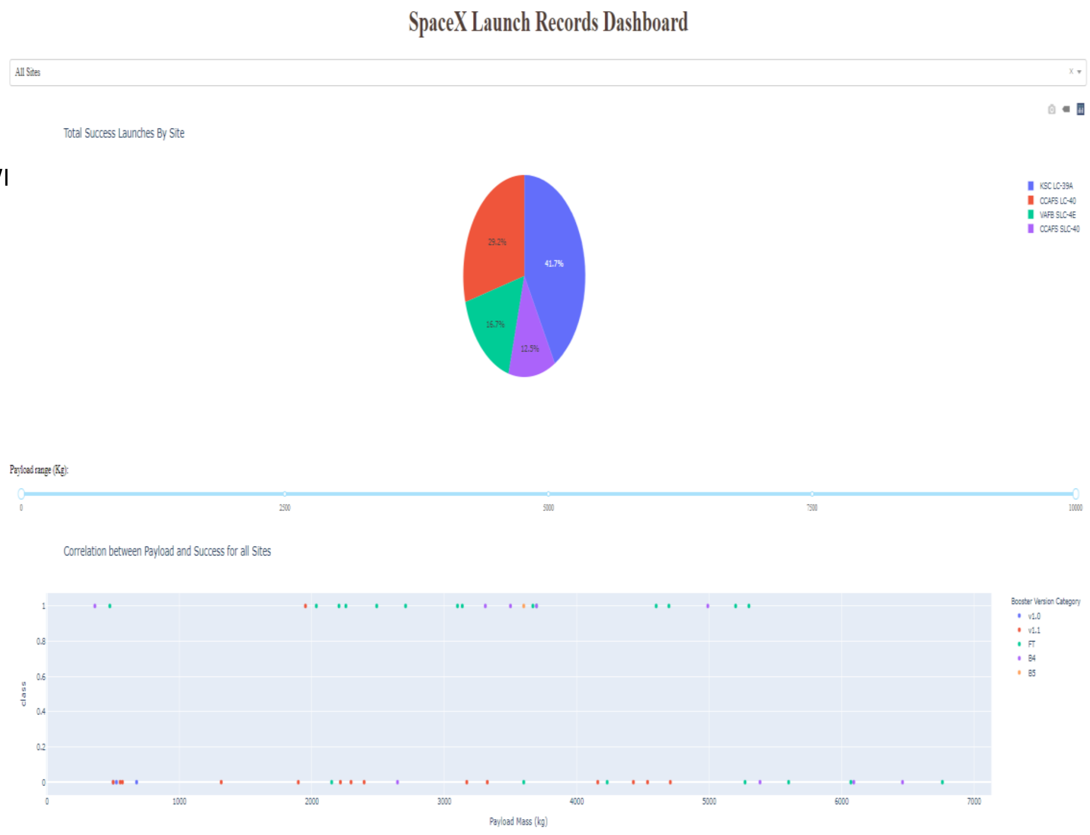
- Then, we marked down a point on the closest coastline using MousePosition and calculated the distance between the coastline point and the launch site



https://github.com/codycoursera/Capstone/blob/Capstone-Main-Branch/lab_jupyter_launch_site_location.ipynb

BUILD A DASHBOARD WITH PLOTLY DASH

- Tasks
  - Add a Launch Site Drop-down Input Component
  - Add a callback function to render success-pie-chart based on selected site dropdown
  - Add a Range Slider to Select Payload
  - Add a callback function to render the success-payload-scatter-chart scatter plot



- Questions Answered

⑩Which site has the largest successful launches?

⑩Which site has the highest launch success rate?

- Which payload range(s) has the highest launch success rate?
- Which payload range(s) has the lowest launch success rate?

⑩Which F9 Booster version (v1.0, v1.1, FT, B4, B5, etc.) has the highest launch success rate?

PREDICTIVE ANALYSIS

- Weperformed exploratory Data Analysis and determine Training Labels
  - We created a column for the class
  - We standardized the data
  - We split into training data and test data

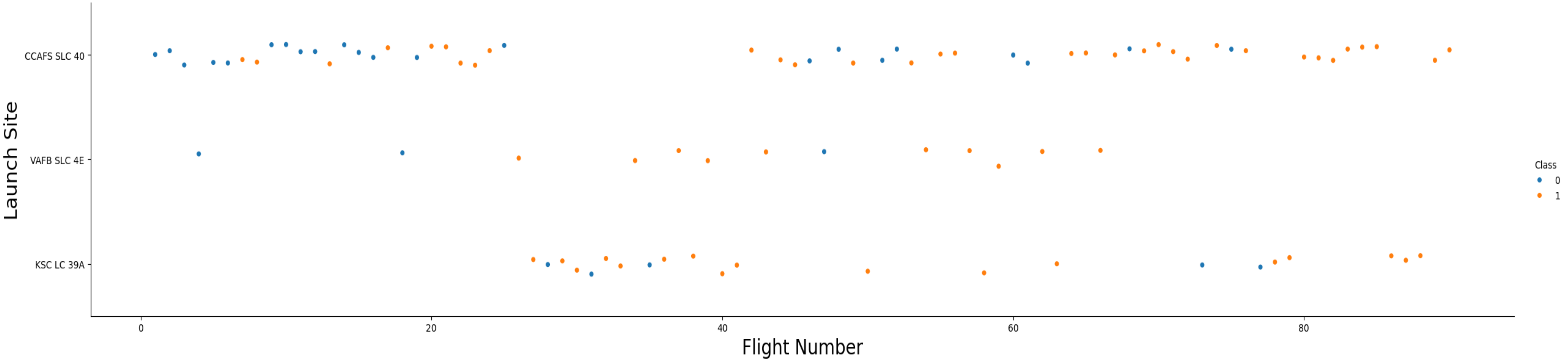
- We found best Hyperparameter for SVM, Classification Trees and Logistic Regression
  - We found the method performs best using test data

https://github.com/codycoursera/Capstone/blob/Capstone-Main-Branch/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

RESULTS

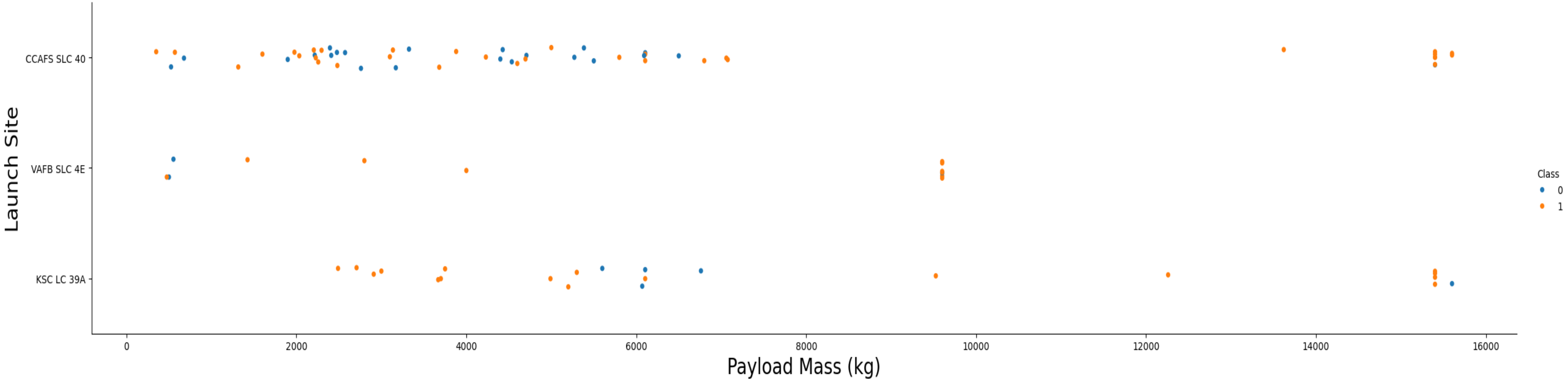
- The SVM, KNN, and Logistic Regression models outperformed the Decision Tree model
- Low weighted payloads performed better than the heavy
- KSC LC 39A had the most successful launches
- Orbit GEO, HEO, SSO, ES-L 1 had the highest success rate
- The GTO had the lowest success rate

FLIGHT NUMBER VS. LAUNCH SITE



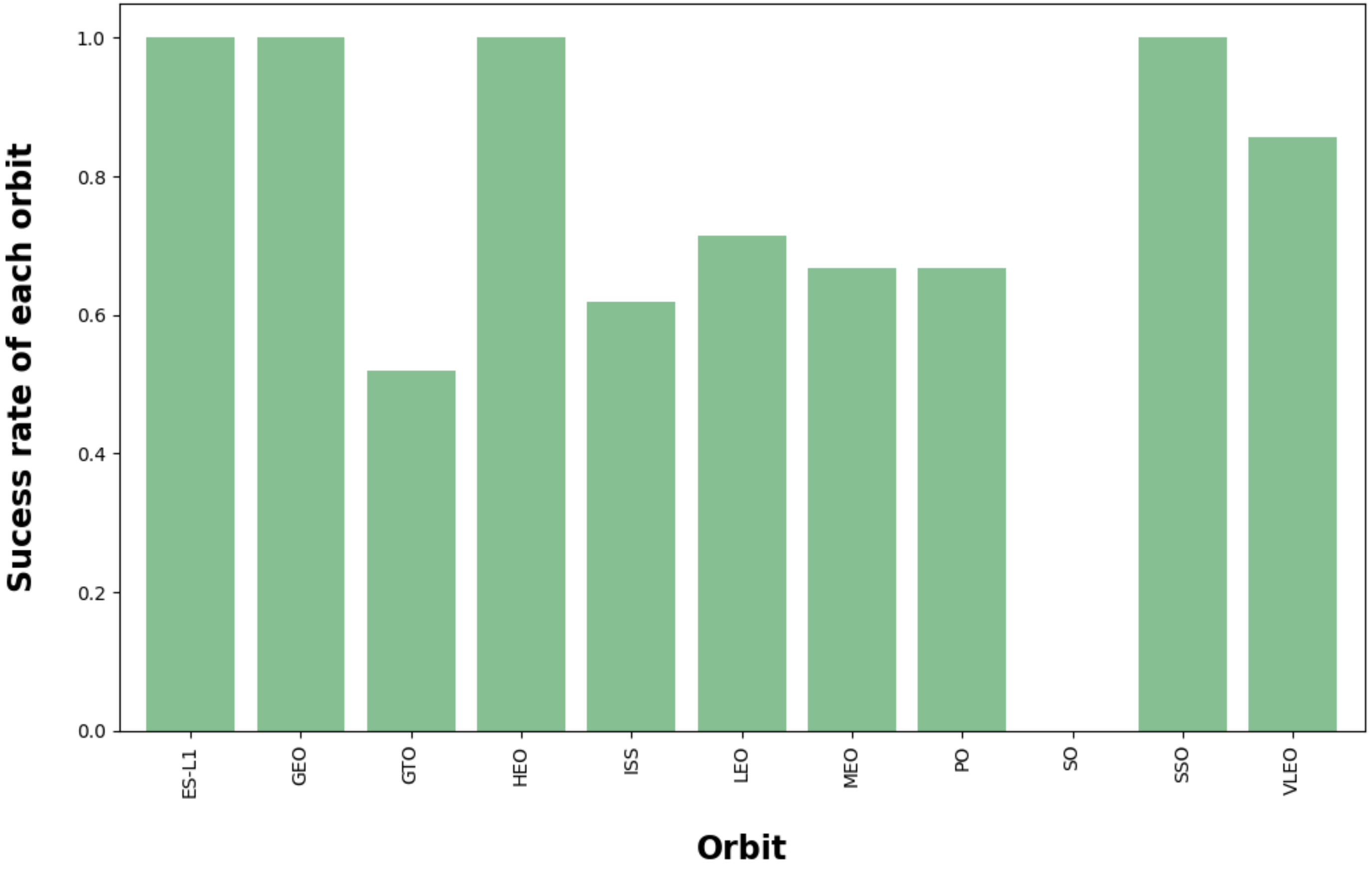
- As the Flight Number increases the success rate increases

PAYLOAD VS. LAUNCH SITE



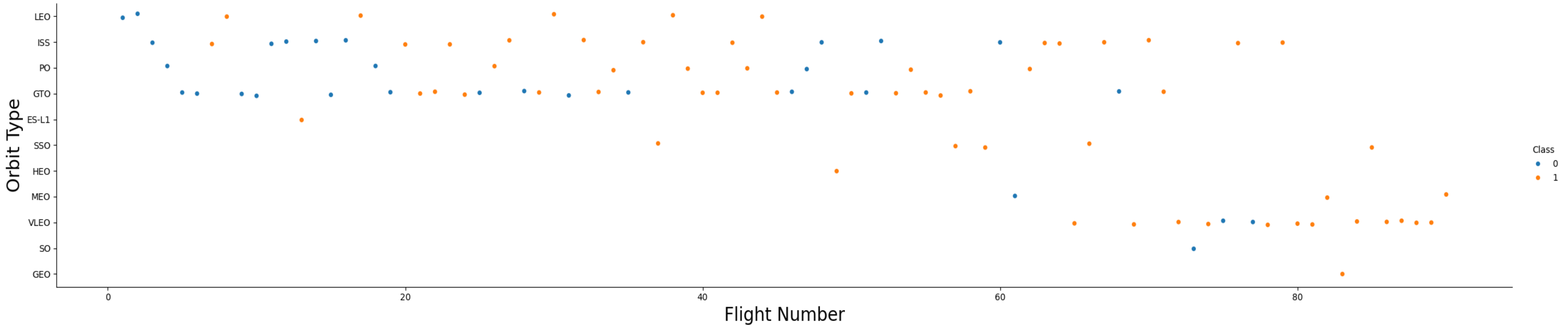
- After 7000kg, the success rate increases
- There is no clear trend per Launch Site

SUCCESS RATE OF EACH ORBIT TYPE



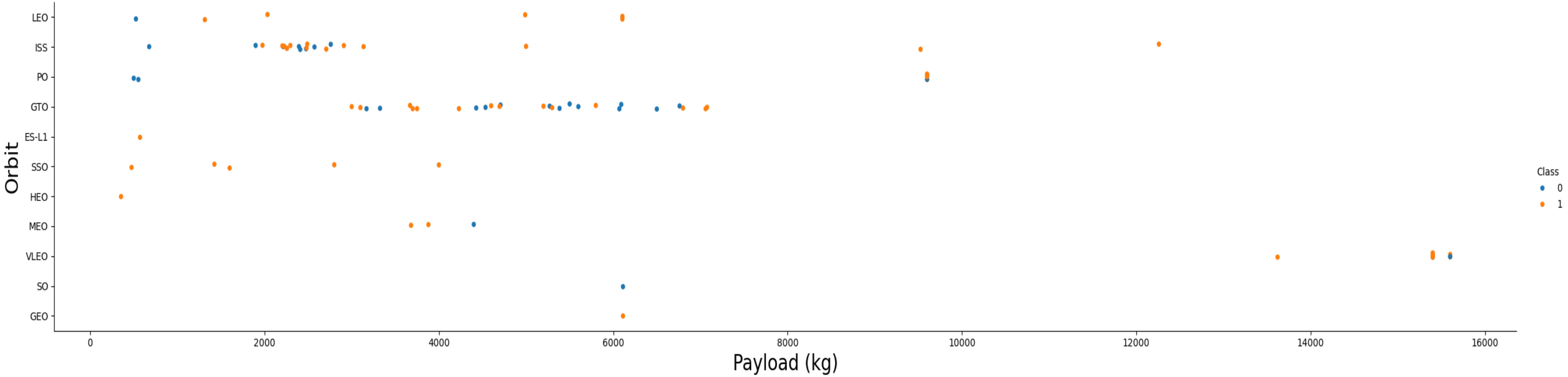
- The ES-L 1, GEO, HEO, and SSO had the highest success rat
- The GTO had the lowest success rate

FLIGHT NUMBER VS. ORBIT TYPE



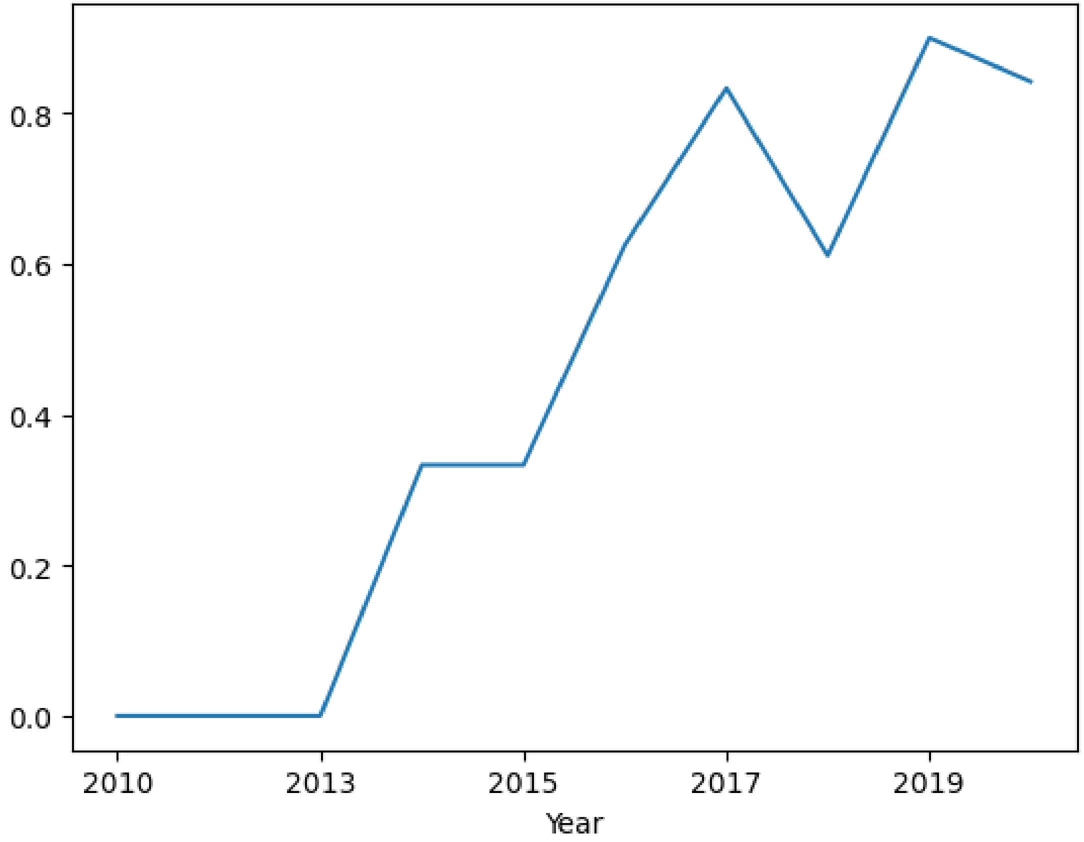
- **As the Flight Number increases, the Orbit Type shifts to GEO**

PAYLOAD VS. ORBIT TYPE



- LEO, ISS, and PO success increases as Payload increases
- GTO and MEO success deacreases as Payload increases

LAUNCH SUCCESS YEARLY TREND



- Launch success has substantially increased over the last decade
- If this trend continues, success rate will taper off into a near 100% success rate

LAUNCH SITE NAMES

%sql SELECT DISTINCT LAUNCH_SITE FROM SPACEXTBL ORDER BY 1;

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

5 LAUNCH SITES BEGINNING WITH 'CCA'

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

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

TOTAL PAYLOAD MASS CARRIED BY BOOSTERS LAUNCHED BY NASA

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

SUM(PAYLOAD_MASS__KG_)
45596

AVERAGE PAYLOAD MASS BY F9 V1.1

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

AVG(PAYLOAD_MASS__KG_)
2534.6666666666665

FIRST SUCCESSFUL GROUND LANDING DATE

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

MIN(Date)
2015-12-22

SUCCESSFUL DRONE SHIP LANDING WITH PAYLOAD BETWEEN 4000 AND 6000

```sql
%%sql
SELECT Booster_Version
FROM SPACEXTBL
WHERE Landing_Outcome = 'Success (drone ship)'
    AND 4000 > Payload_Mass__KG_ < 6000;
```

Booster_Version
F9 FT B1021.1
F9 FT B1022
F9 FT B1023.1
F9 FT B1026
F9 FT B1029.1
F9 FT B1021.2
F9 FT B1029.2
F9 FT B1036.1
F9 FT B1038.1
F9 B4 B1041.1
F9 FT B1031.2
F9 B4 B1042.1
F9 B4 B1045.1
F9 B5 B1046.1

TOTAL NUMBER OF SUCCESSFUL AND FAILURE MISSION OUTCOMES

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

Mission_Outcome	TOTAL_NUMBER
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

BOOSTERS CARRYING MAXIMUM PAYLOAD

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

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

```
%%sql
SELECT substr("Date", 6,2) as Month, LANDING_OUTCOME, BOOSTER_VERSION, LAUNCH_SITE
FROM SPACEXTBL
WHERE Landing_Outcome = 'Failure (drone ship)'
    AND substr("Date",0,5) = '2015';
```

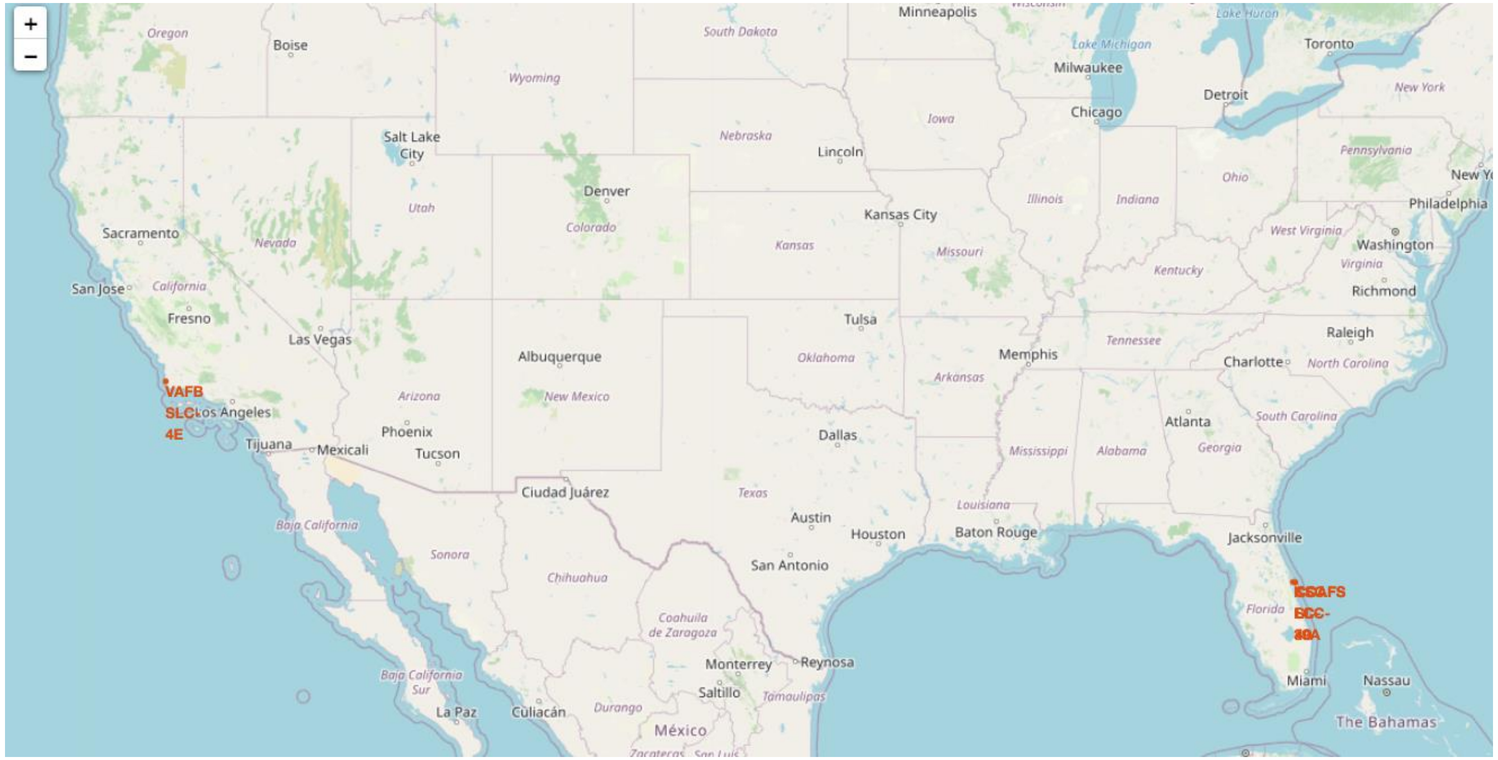
Month	Landing_Outcome	Booster_Version	Launch_Site
01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

LANDING OUTCOMES BY TOTAL NUMBER

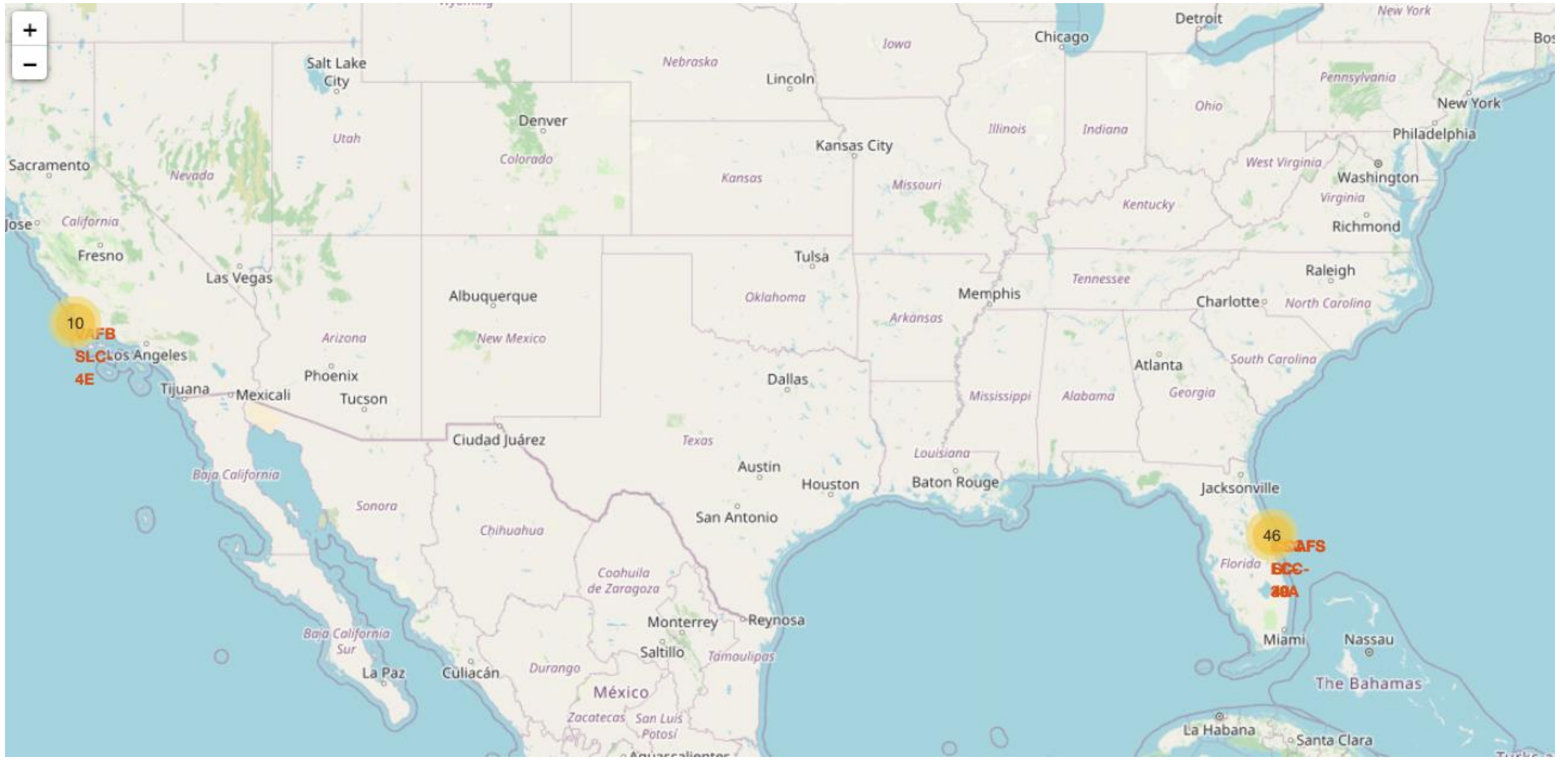
```
%%sql
SELECT Landing_Outcome, COUNT(Landing_Outcome) as Total_Number
from SPACEXTBL
WHERE Date Between '2010-06-04' AND '2017-03-20'
    GROUP BY Landing_Outcome
    ORDER BY Total_Number Desc;
```

Landing_Outcome	Total_Number
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

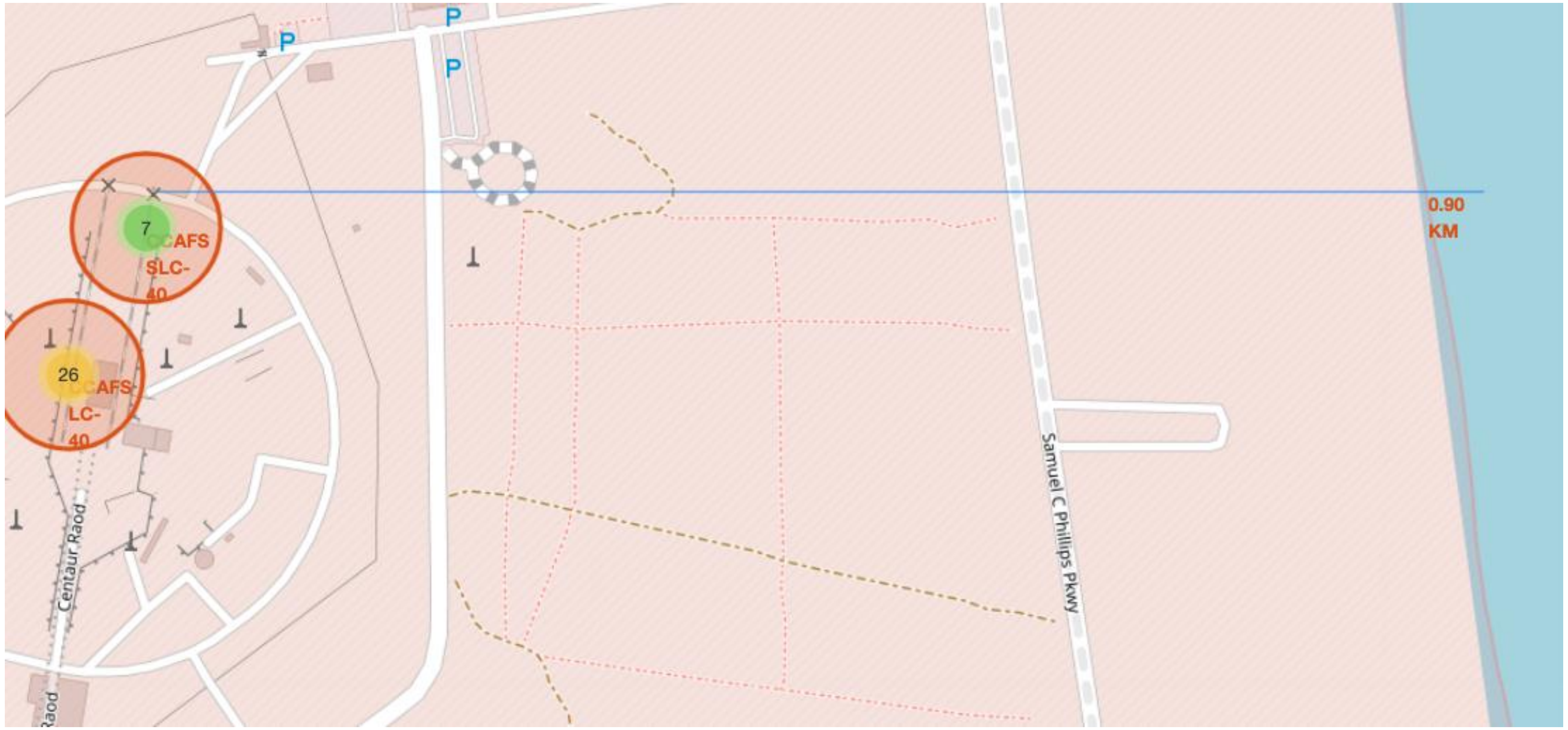
LAUNCH SITES MARKED ON MAP



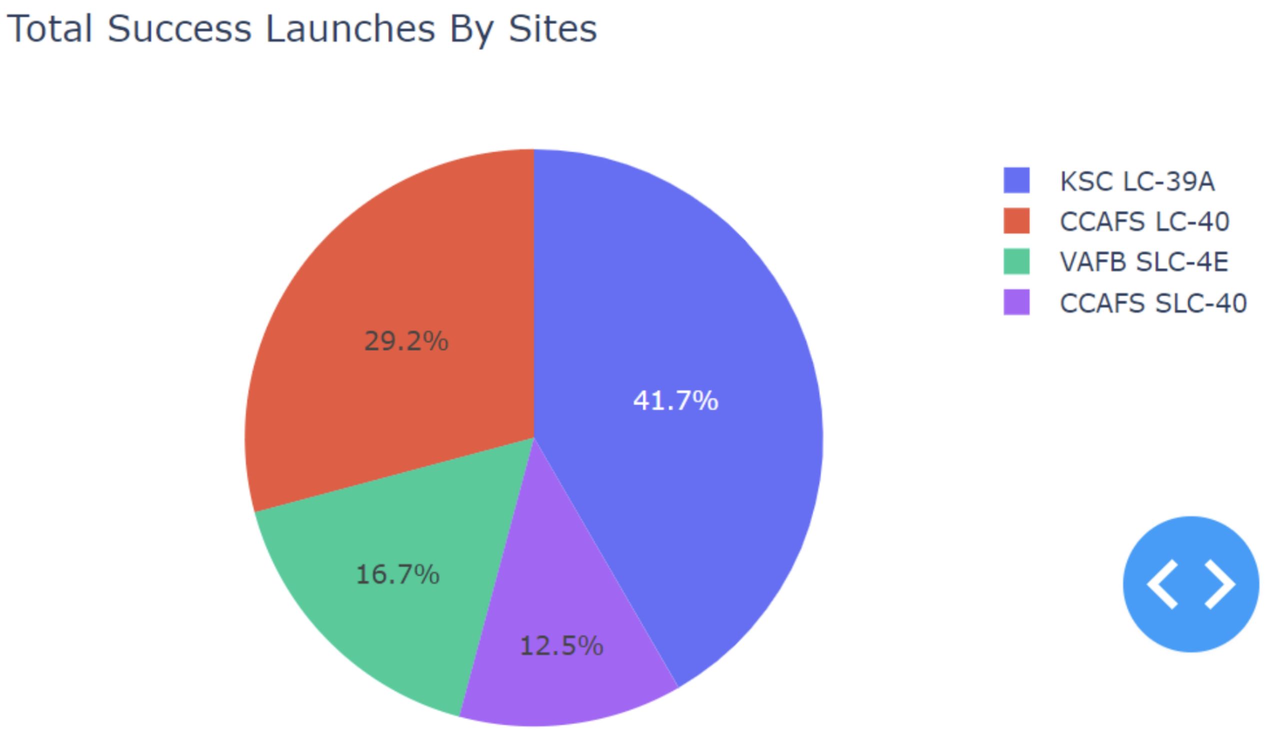
SUCCESS/FAILED LAUNCH SITES MARKED ON MAP



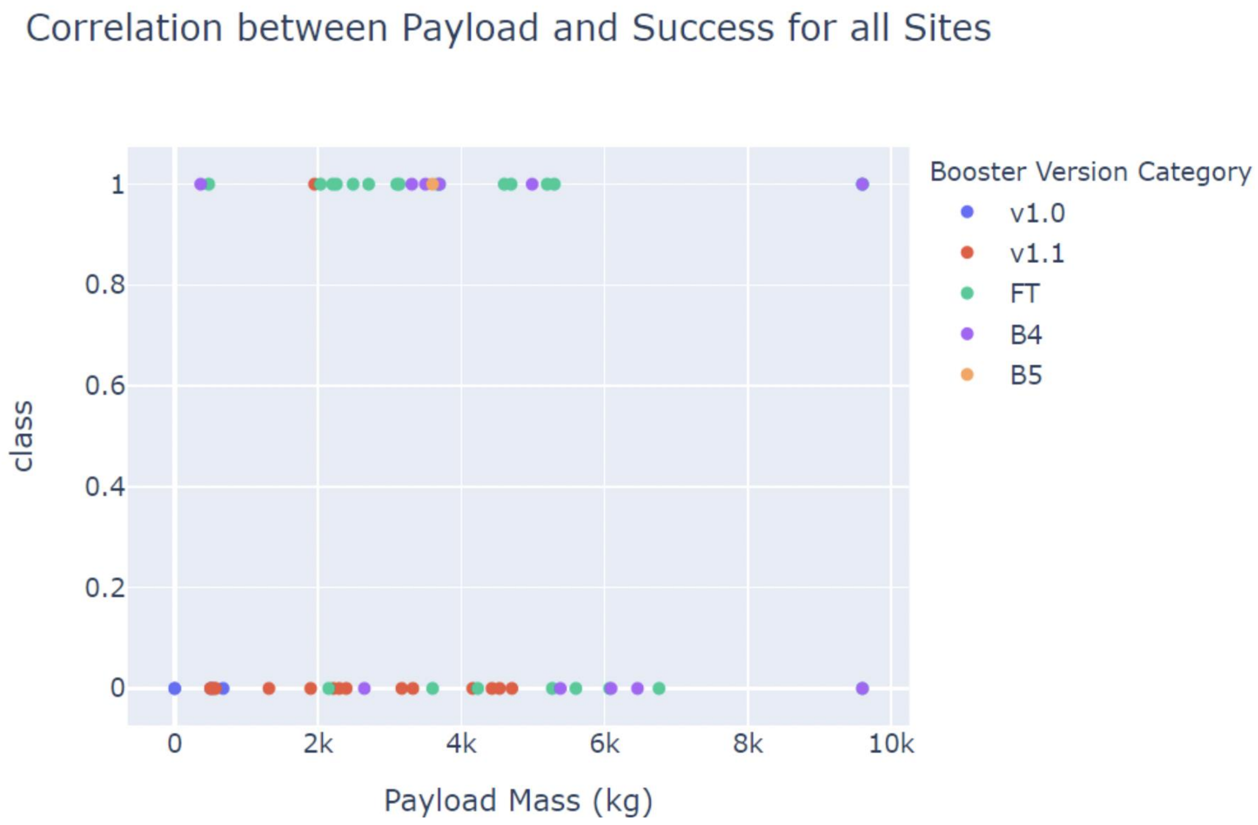
DISTANCES BETWEEN LAUNCH SITES TO PROXIMITIES



TOTAL SUCCESS LAUNCHES BY SITES



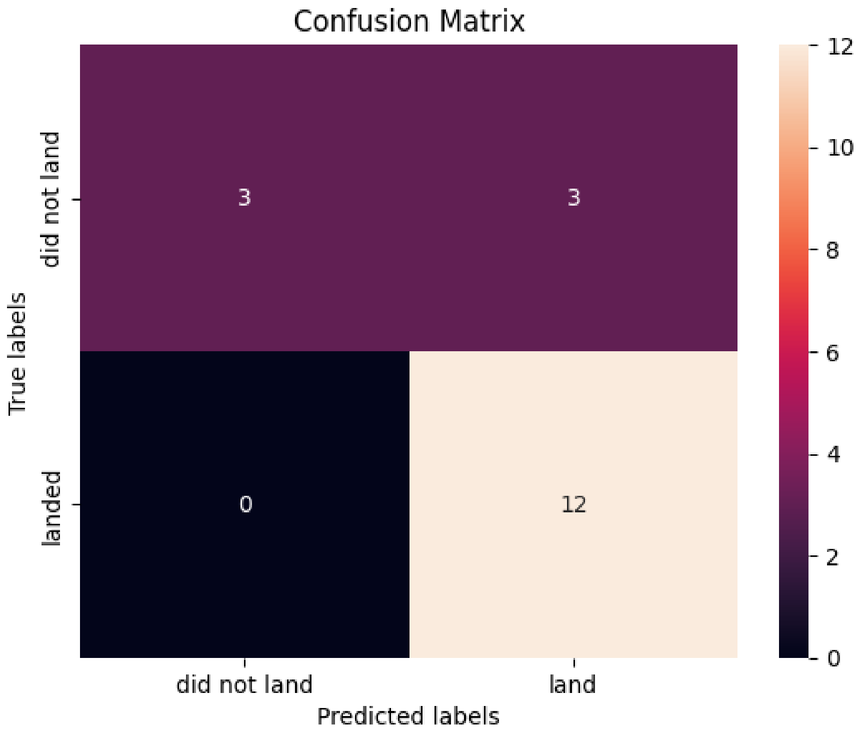
PAYLOAD VS. LAUNCH OUTCOME

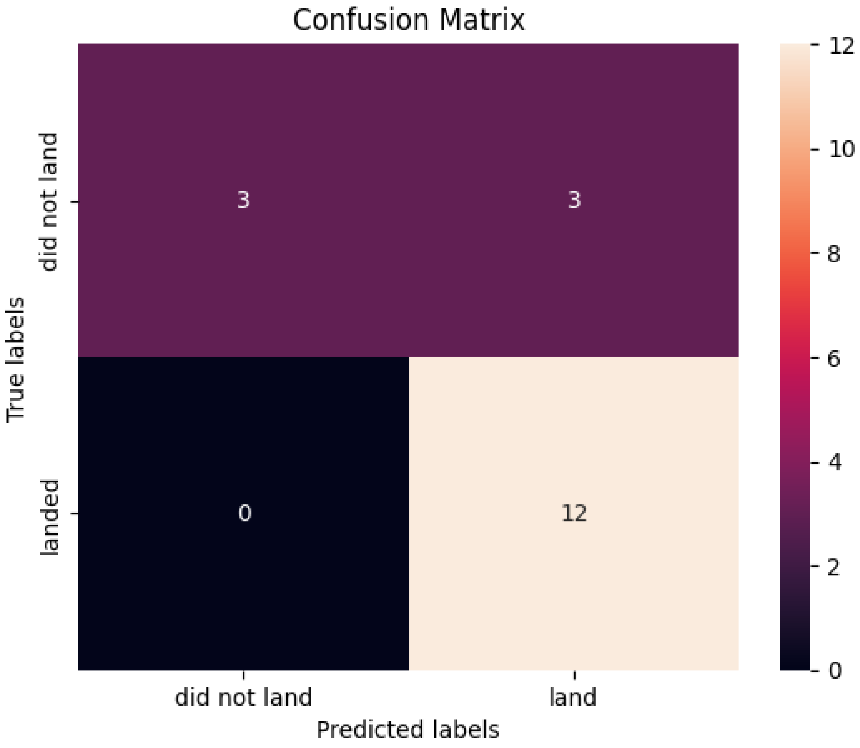


CLASSIFICATION ACCURACY

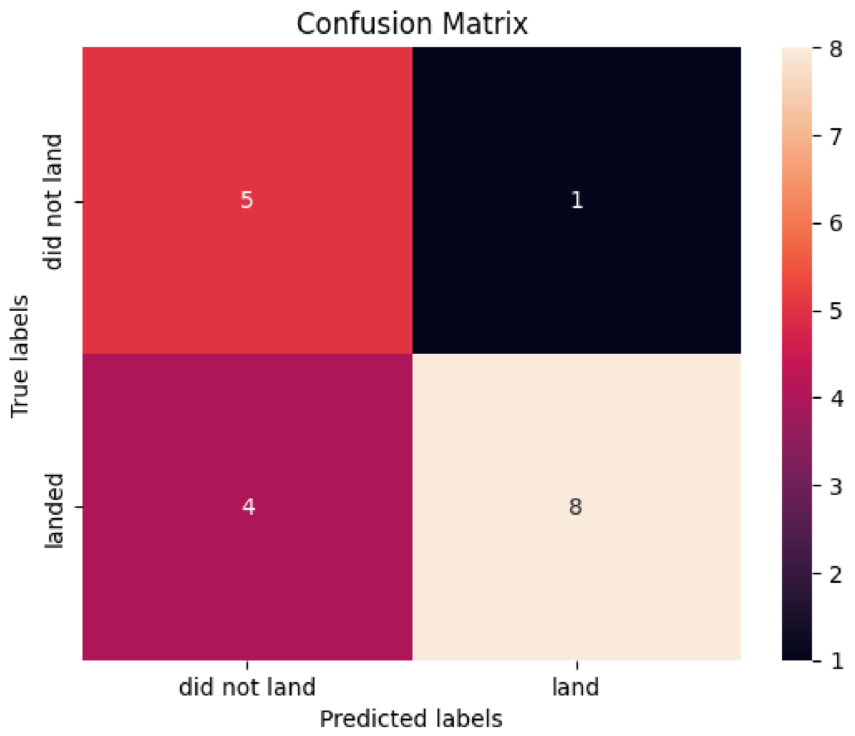
Method	Test Data Accuracy
Logistic_Reg	0.833333
SVM	0.833333
Decision Tree	0.722222
KNN	0.833333

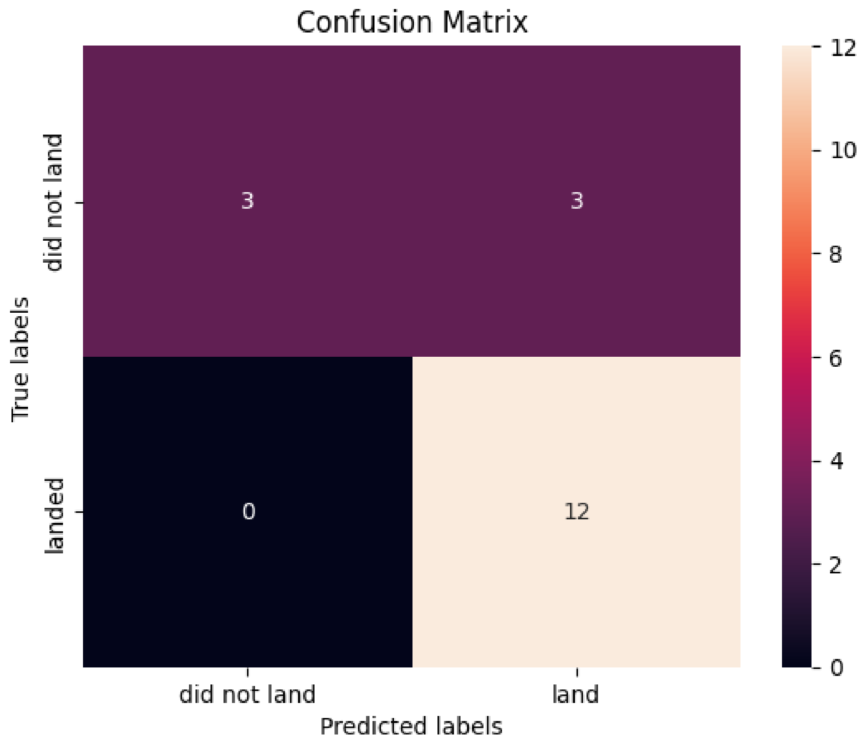
CONFUSION MATRIX





CONFUSION MATRIX CONT.





CONCLUSION

- SVM, KNN, and Logistic Regression models had higher accuracy than the Decision Tree Model
- Lighter payloads outperformed heavier payloads
- Launch success has significantly increased over the last decade
- KSC LC 39A had the most successful launches
- Orbit GEO, HEO, SSO, ES-L 1 had the highest success rate
- The GTO had the lowest success rate


THANK YOU!