IBM Developer
SKILLS NETWORK

Presentation SpaceX Capstone Project

A. Danen Padiachy
13 Mar 2023

Outline

- Executive summary
- Introduction
- Section 1: Methodology
- Section 2: Insights drawn from EDA
    - EDA with Visualization
    - EDA with SQL
- Section 3: Launch sites Proximities Analysis
    - Interactive Maps with Folium
- Section 4: Build a Dashboard with Plotly Dash
    - Plotly Dash Dashboard
- Section 5: Predictive Analytics (Classification)
- Conclusion
- Appendix

Executive Summary

Summary of methodologies

This research has for objective to identify the factors needed for a successful rocket landing. To make this determination, the following methodologies where used:

- Collecting data using SpaceX REST API and web scraping techniques
- Wrangling data to create success/fail outcome variable
- Exploring data with data visualization techniques, considering the following factors: payload, launch site, flight number and yearly trend
- Analyzing the data with SQL, calculating the following statistics: total payload, payload range for successful launches, and total # of successful and failed outcomes
- Exploring launch site success rates and proximity to geographical markers
- Visualizing the launch sites with the most success and successful payload ranges
- Building models to predict landing outcomes using logistic regression, support vector machine (SVM), decision tree and K-nearest neighbor (KNN)

Executive Summary

Summary of all results

Exploratory Data Analysis:

- Improvement in launch success over time.
- Among all landing sites, KSC LC-39A has the highest success rate.
- There is a 100% success rate in regard to orbits ES-L1, GEO, HEO, and SSO.

Visualization/Analytics:

The majority of launch sites are near the equator and they are also close to the coast

Predictive Analytics:

The models performed similarly on the test set. The decision tree model slightly outperformed.

Introduction

Project background / context

SpaceX is a leader in the space industry and strives to make space travel affordable for all. Some of its major achievements include sending spacecraft to the international space station, launching satellites that provides internet access and sending manned missions to space, among others.

SpaceX is able to make such bold progress because the rocket launches are relatively inexpensive ($62 million per launch) due to its novel reuse of the first stage of its Falcon 9 rocket. Other providers, which are not able to reuse the first stage, cost upwards of $165 million each. By determining if the first stage will land, we can determine the price of the launch. To do this, we can use public data and machine learning models to predict whether SpaceX (or a competing company) can reuse the first stage.

Keys points to consider

- How payload mass, launch site, number of flights, and orbits affect first-stage landing success
- Rate of successful landings over time
- Best predictive model for successful landing (binary classification)

Section 1

Methodology

Methodology

Executive Summary

- Data is collected using SpaceX REST API and web scraping techniques
- Data wrangling is done by filtering the data, handling missing values and applying one hot encoding – to prepare the data for analysis and modeling
- Data exploration is done via EDA with SQL and data visualization techniques
- Data is visualized using Folium and Plotly Dash
- Models are built to predict landing outcomes using classification models. The models are tuned and evaluated to find best model and parameters

Data Collection - SpaceX API

Flowchart of SpaceX API calls :

Data request from SpaceX API (rocket launch data)

Response decoding using .json() and convert to a dataframe using .json normalize()

Information request about the launches from SpaceX API using custom functions

Creating dictionary from the data

Creating dataframe from the dictionary

Filtering dataframe to contain only Falcon 9 launches

Replacing missing values of Payload Mass with calculated .mean()

Exporting data to csv file

Github URL: https://github.com/danpaddy01/IBM-Applied-Data-Science-Capstone_ADP/blob/main/Part1_SpaceX_Data_Collection_API.ipynb

Data Collection - Scraping

Flowchart of web scraping:

Data request (Falcon 9 launch data) from Wikipedia

Creating BeautifulSoup object from HTML response

Extracting column names from HTML table header

Collecting data from parsing HTML tables

Creating dictionary from the data

Creating dataframe from the dictionary

Exporting data to csv file

Github URL:
https://github.com/danpaddy01/IBM-Applied-Data-Science-Capstone_ADP/blob/main/Part2_SpaceX_Web_Scraping.ipynb

Data Wrangling

Process::

Perform EDA and determine data labels

Calculate:

- # of launches for each site
- # and occurrence of orbit
- # and occurrence of mission outcome per orbit type]

Create binary landing outcome column (dependent variable)

Export data to csv file

Landing Outcome

We note that landing is not always successful

- **True Ocean:** mission outcome had a successful landing to a specific region of the ocean
- **False Ocean:** represented an unsuccessful landing to a specific region of ocean
- **True RTLS:** meant the mission had a successful landing on a ground pad
- **False RTLS:** represented an unsuccessful landing on a ground pad
- **True ASDS:** meant the mission outcome had a successful landing on a drone ship
- **False ASDS:** represented an unsuccessful landing on drone ship
- **Outcomes converted** into 1 for a successful landing and 0 for an unsuccessful landing

Github URL: https://github.com/danpaddy01/IBM-Applied-Data-Science-Capstone_ADP/blob/main/Part3_SpaceX_Data_Wrangling.ipynb

EDA with Data Visualization

Charts plotted:
• Flight Number vs. Payload
• Flight Number vs. Launch Site
• Payload Mass (kg) vs. Launch Site
• Payload Mass (kg) vs. Orbit type

Usefulness of charts:
• View relationship by using scatter plots. The variables could be useful for machine learning if a relationship exists.
• Show comparisons among discrete categories with bar charts. Bar charts show the relationships among the categories and a measured value.

Github URL: https://github.com/danpaddy01/IBM-Applied-Data-Science-Capstone_ADP/blob/main/Part5_SpaceX_EDA_Data_Visualization.ipynb

EDA with SQL

Queries performed:

Display of query results:
- Names of unique launch sites
- 5 records where launch site begins with 'CCA'
- Total payload mass carried by boosters launched by NASA (CRS)
- Average payload mass carried by booster version F9 v1.1.

List of query results:
- Date of first successful landing on ground pad
- Names of boosters which had success landing on drone ship and have payload mass greater than 4,000 but less than 6,000
- Total number of successful and failed missions
- Names of booster versions which have carried the max payload
- Failed landing outcomes on drone ship, their booster version and launch site for the months in the year 2015
- Count of landing outcomes between 2010 06 04 and 2017 03 20 (desc)

Github URL:
https://github.com/danpaddy01/IBM-Applied-Data-Science-Capstone_ADP/blob/main/Part4_SpaceX_EDA_SQL.ipynb

Map with Folium

Markers Indicating Launch Sites
• Added blue circle at NASA Johnson Space Center's coordinate with a pop-up label showing its name using its latitude and longitude coordinates.
• Added red circles at all launch sites coordinates with a pop-up label showing its name using its name using its latitude and longitude coordinates.

Colored Markers of Launch Outcomes
• Added colored markers of successful (green) and unsuccessful (red) launches at each launch site to show which launch sites have high success rates.

Distances Between a Launch Site to Proximities
• Added colored lines to show distance between launch site CCAFS SLC 40 and its proximity to the nearest coastline, railway, highway, and city.

Github URL: https://github.com/danpaddy01/IBM-Applied-Data-Science-Capstone_ADP/blob/main/Part6_SpaceX_Interactive_Visual_Analytics_Folium.ipynb

Dashboard with Plotly Dash

Dropdown List with Launch Sites
• Allow user to select all launch sites or a certain launch site.

Pie Chart Showing Successful Launches
• Allow user to see successful and unsuccessful launches as a percent of the total.

Slider of Payload Mass Range
• Allow user to select payload mass range.

Scatter Chart Showing Payload Mass vs. Success Rate by Booster Version
• Allow user to see the correlation between Payload and Launch Success.

Github URL:
https://github.com/danpaddy01/IBM-Applied-Data-Science-Capstone_ADP/blob/main/Part7_spacex_dash_app.py

Predictive Analysis (Classification)

Charts

Create NumPy array from the Class column

Standardize the data with StandardScaler. Fit and transform the data.

Split the data using train_test_split

Create a GridSearchCV object with cv=10 for parameter optimization

Apply GridSearchCV on different algorithms: logistic regression (LogisticRegression()), support vector machine (SVC()), decision tree (DecisionTreeClassifier()), K Nearest Neighbor (KNeighborsClassifier())

- Calculate accuracy on the test data using .score() for all models
- Assess the confusion matrix for all models
- Identify the best model using Jaccard_Score, F1_Score and Accuracy

Github URL: https://github.com/danpaddy01/IBM-Applied-Data-Science-Capstone_ADP/blob/main/Part8_SpaceX_Predictive_Analytics.ipynb

Results

Exploratory Data Analysis results

- Launch success has improved over time
- KSC LC-39A has the highest success rate among landing sites
- Orbits ES L1, GEO, HEO and SSO have a 100% success rate

Visual Analytics results

- Most launch sites are near the equator, and all are close to the coast
- Launch sites are far enough away from anything a failed launch can damage (city, highway, railway), while still close enough to bring people and material to support launch activities

Predictive Analytics results

- Decision Tree model is the best predictive model for the dataset

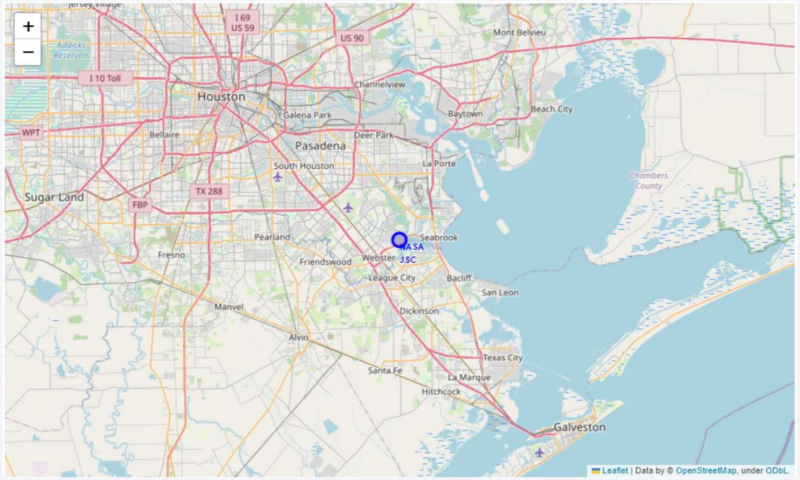
Results

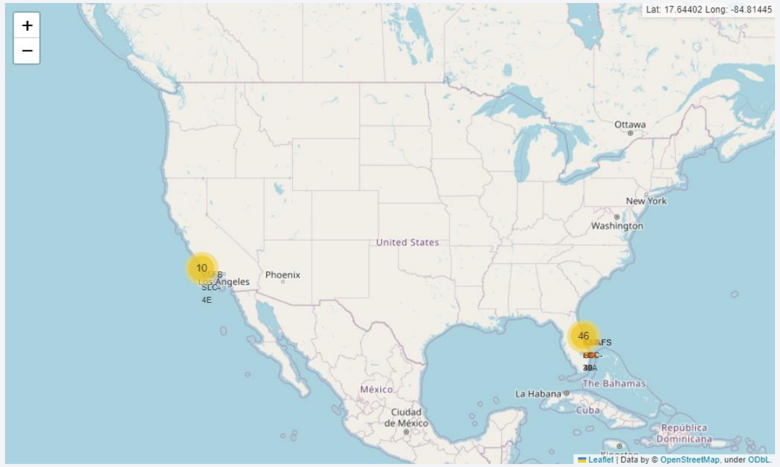
Interactive demo in screenshots

	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

	Launch Site	Lat	Long	class
46	KSC LC-39A	28.573255	-80.646895	1
47	KSC LC-39A	28.573255	-80.646895	1
48	KSC LC-39A	28.573255	-80.646895	1
49	CCAFS SLC-40	28.563197	-80.576820	1
50	CCAFS SLC-40	28.563197	-80.576820	1
51	CCAFS SLC-40	28.563197	-80.576820	0
52	CCAFS SLC-40	28.563197	-80.576820	0
53	CCAFS SLC-40	28.563197	-80.576820	0
54	CCAFS SLC-40	28.563197	-80.576820	1
55	CCAFS SLC-40	28.563197	-80.576820	0

	Launch Site	Lat	Long	class	marker_color
46	KSC LC-39A	28.573255	-80.646895	1	green
47	KSC LC-39A	28.573255	-80.646895	1	green
48	KSC LC-39A	28.573255	-80.646895	1	green
49	CCAFS SLC-40	28.563197	-80.576820	1	green
50	CCAFS SLC-40	28.563197	-80.576820	1	green
51	CCAFS SLC-40	28.563197	-80.576820	0	red
52	CCAFS SLC-40	28.563197	-80.576820	0	red
53	CCAFS SLC-40	28.563197	-80.576820	0	red
54	CCAFS SLC-40	28.563197	-80.576820	1	green
55	CCAFS SLC-40	28.563197	-80.576820	0	red





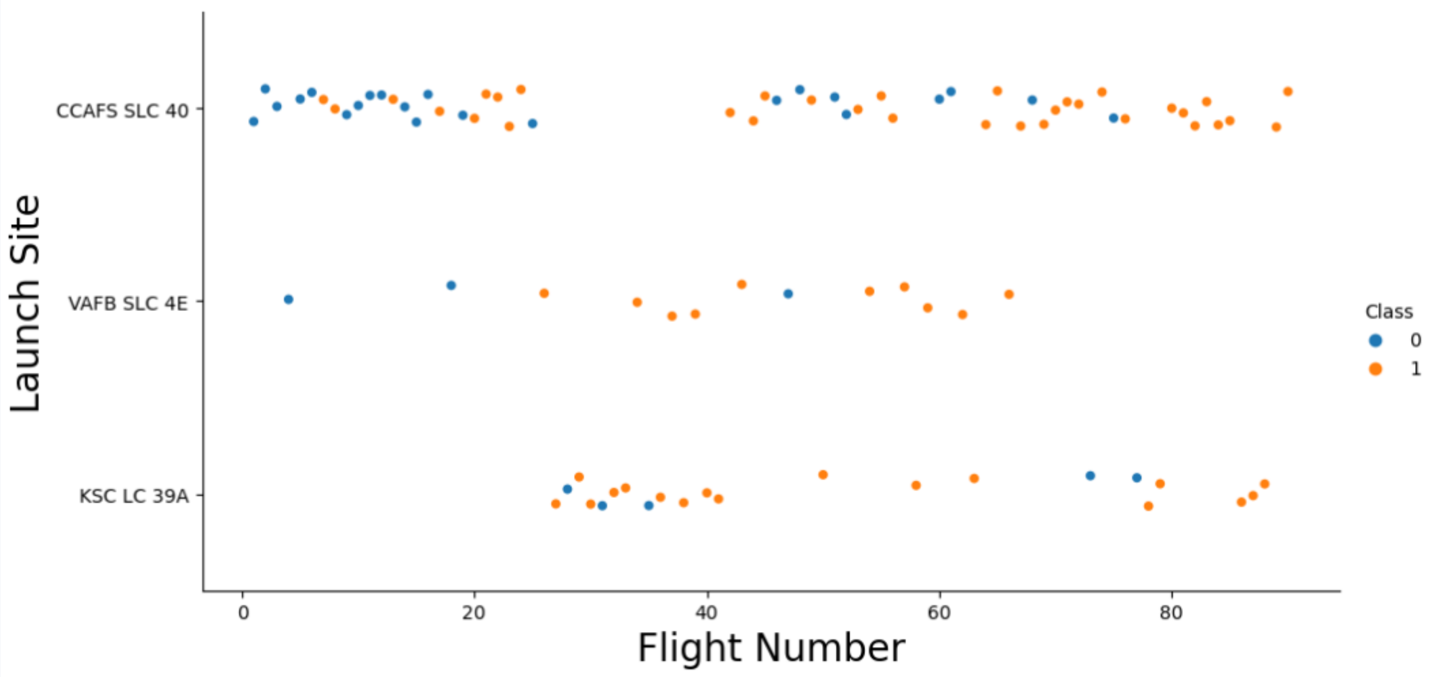
Section 2

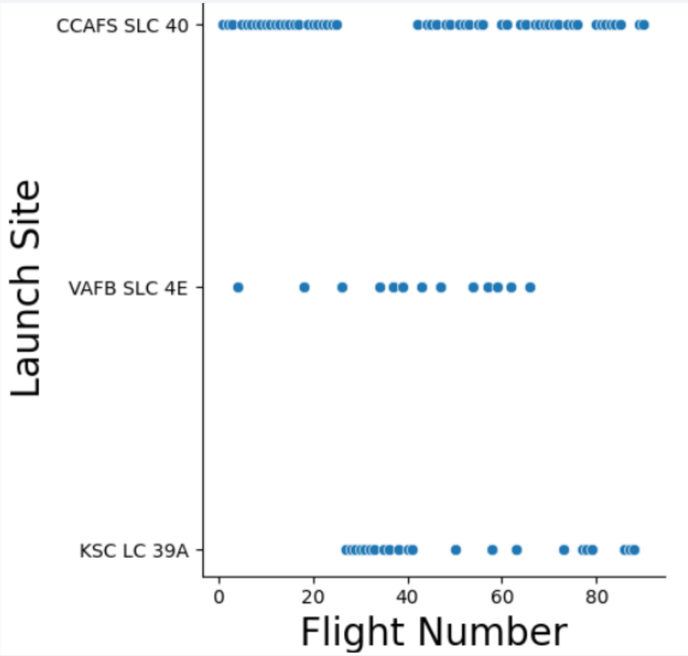
Insights drawn from EDA

Flight Number vs. Launch Site

Exploratory Data Analysis

- Earlier flights had a lower success rate (blue = fail)
- Later flights had a higher success rate (orange = success)
- Around half of launches were from CCAFS SLC 40 launch site
- VAFB SLC 4E and KSC LC 39A have higher success rates
- We can infer that new launches have a higher success rate

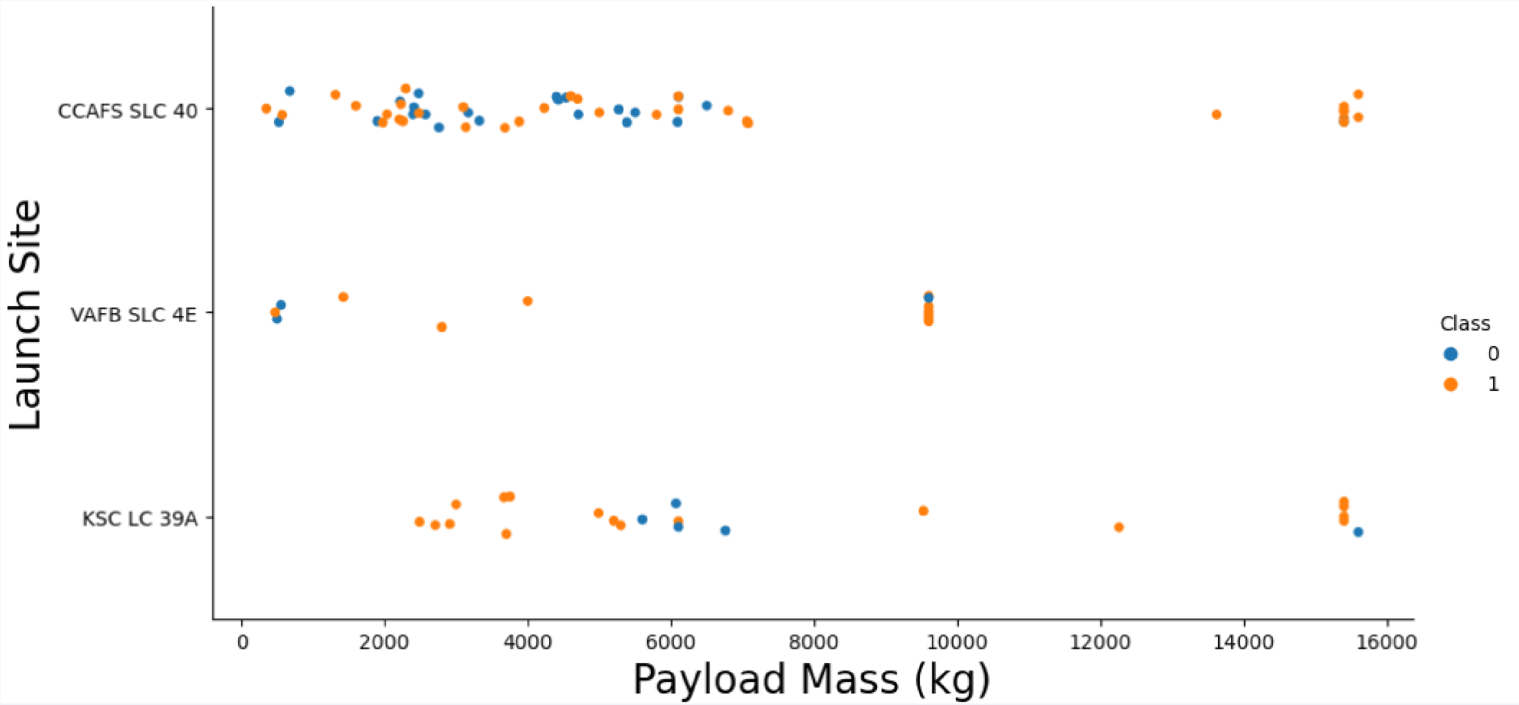


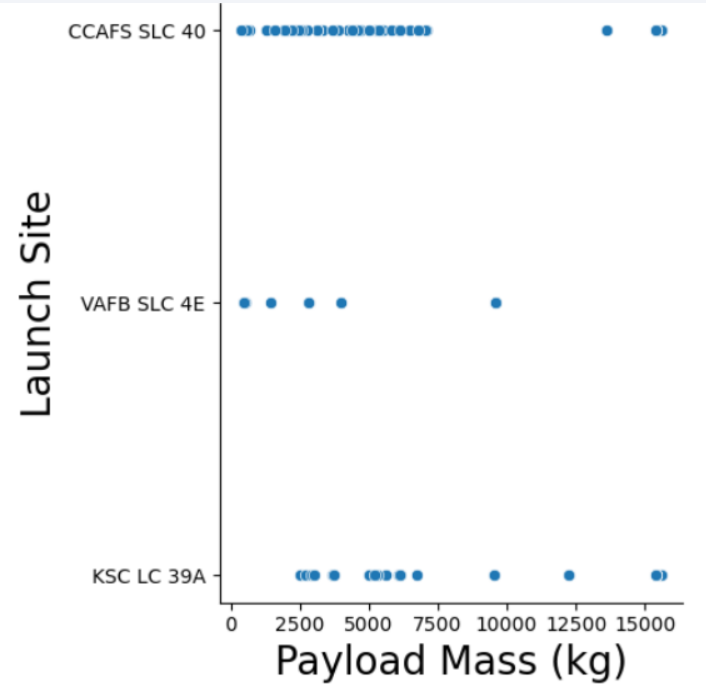


Payload vs. Launch Site

Exploratory Data Analysis

- Typically, the higher the payload mass (kg), the higher the success rate
- Most launces with a payload greater than 7,000 kg were successful
- KSC LC 39A has a 100% success rate for launches less than 5,500 kg
- VAFB SKC 4E has not launched anything greater than ~10,000 kg

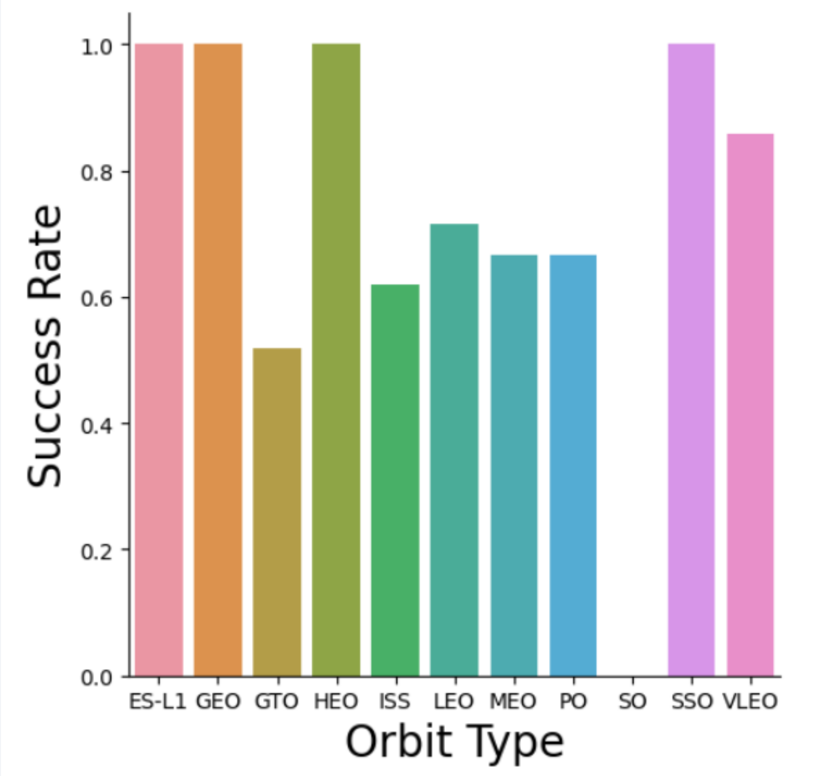




Success Rate vs. Orbit Type

Exploratory Data Analysis

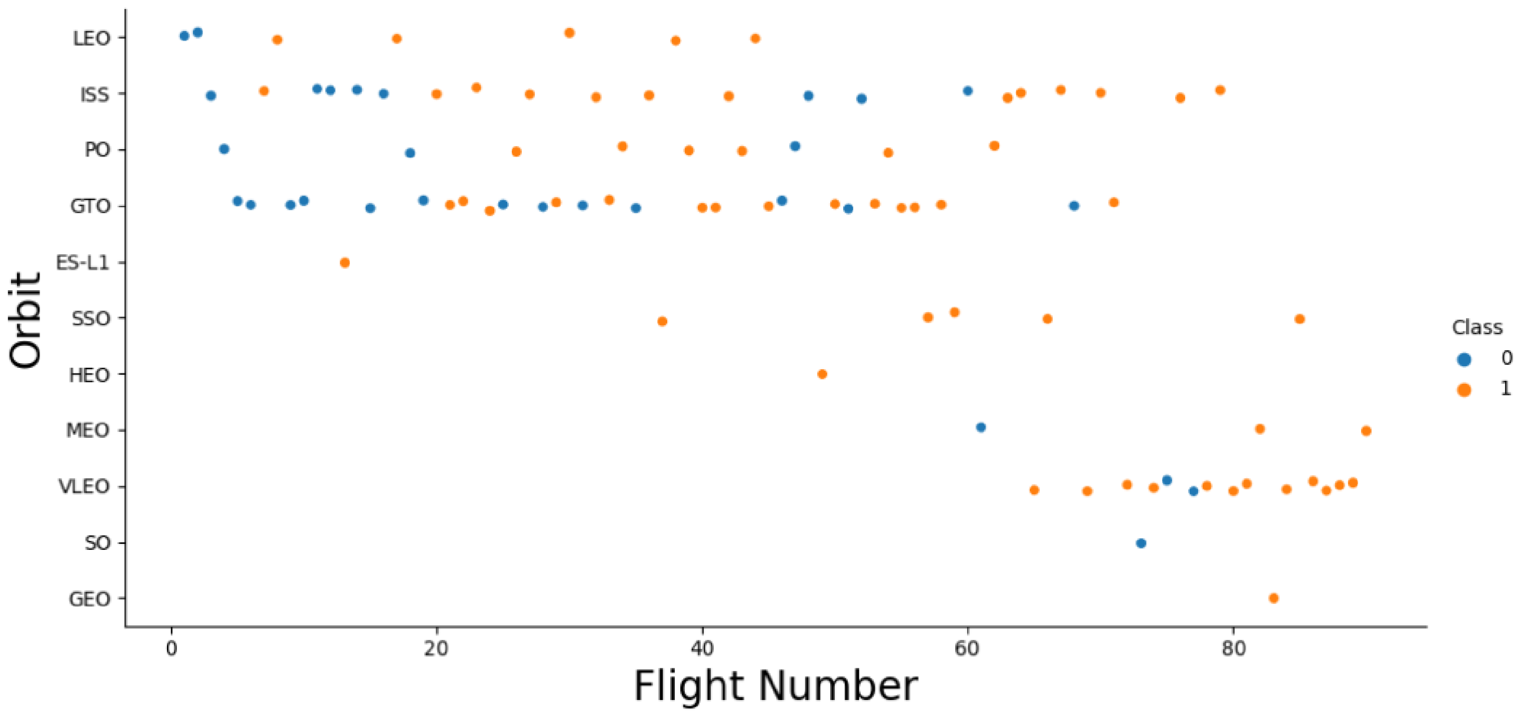
- 100% Success Rate: ES L1, GEO, HEO and SSO
- 50%-80% Success Rate : GTO, ISS, LEO, MEO, PO
- 0% Success Rate: SO

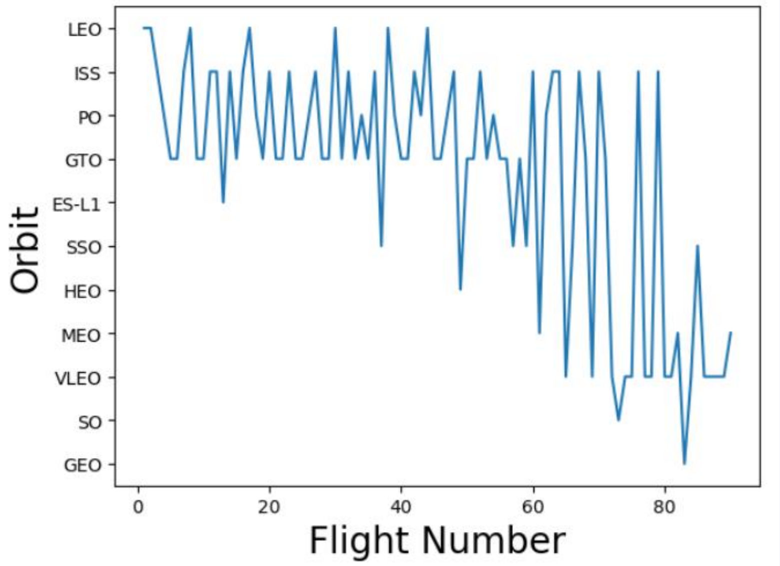


Flight Number vs. Orbit Type

Exploratory Data Analysis

- The success rate typically increases with the number of flights for each orbit
- This relationship is highly apparent for the LEO orbit
- The GTO orbit, however, does not follow this trend

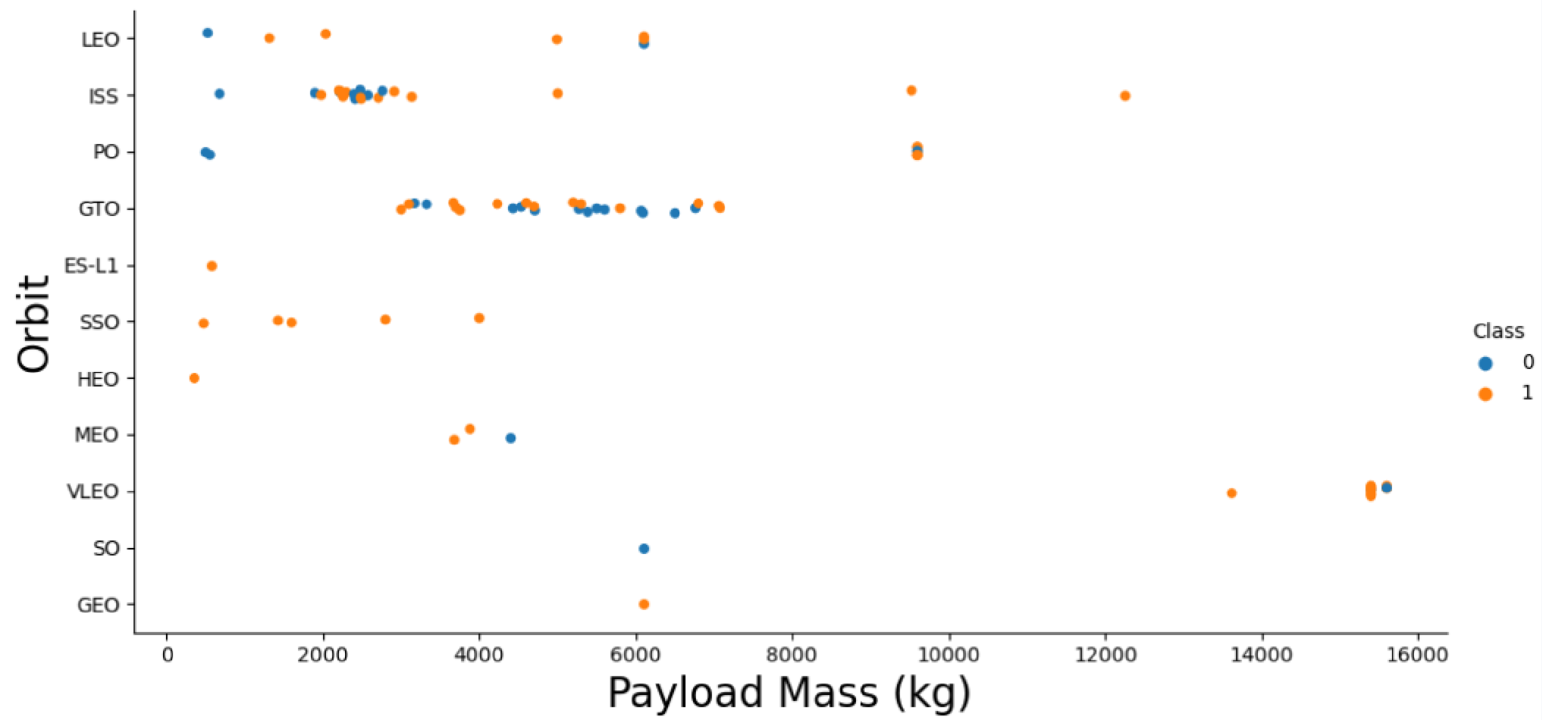




Payload vs. Orbit Type

Exploratory Data Analysis

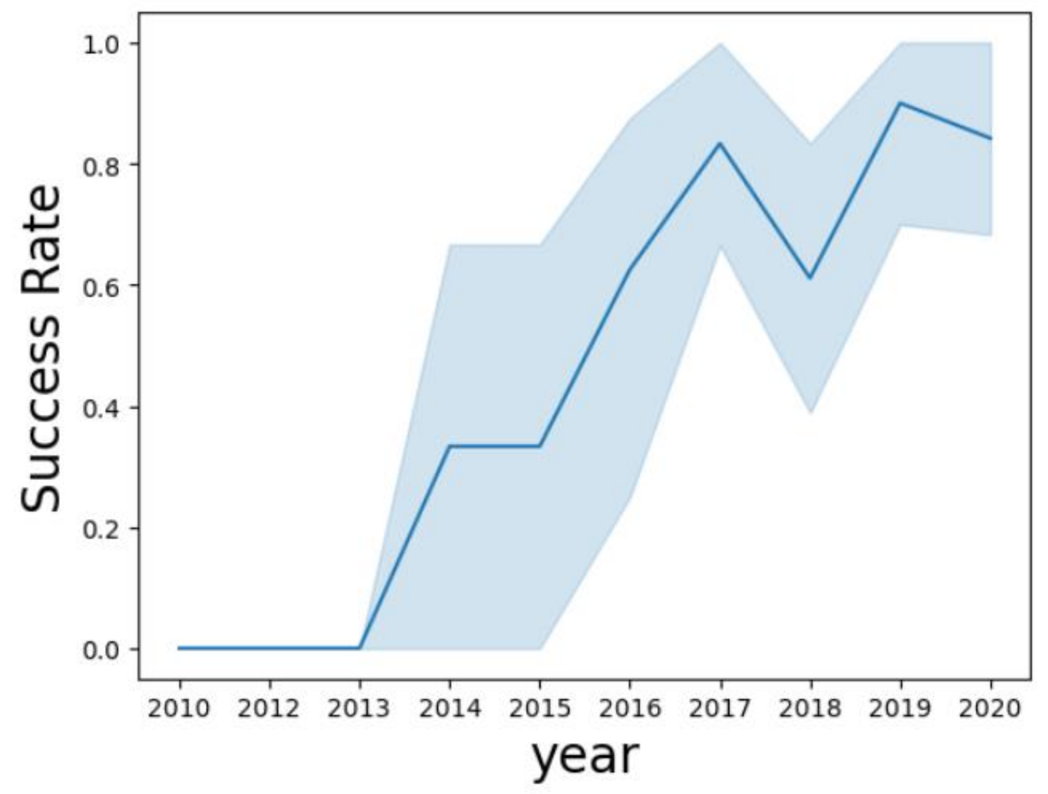
- Heavy payloads are better with LEO, ISS and PO orbits
- The GTO orbit has mixed success with heavier payloads



Launch Success Yearly Trend

Exploratory Data Analysis

- The success rate improved from 2013-2017 and 2018-2019
- The success rate decreased from 2017-2018 and from 2019-2020
- Overall, the success rate has improved since 2013



All Launch Site Names

Launch Site Names
- CCAFS LC-40
- CCAFS SLC-40
- KSC LC-39A
- VAFB SLC-4E

The below code is run to obtain the result:
%sql SELECT DISTINCT LAUNCH_SITE FROM SPACEX_TBL;

```
%sql SELECT DISTINCT LAUNCH_SITE FROM SPACEX_TBL;
```

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

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

Launch Site Names Begin with 'CCA'

Displaying 5 records below:

The code line ‘%sql SELECT FROM SPACEX_TBL WHERE LAUNCH_SITE LIKE ‘CCA%’ LIMIT 5’ is run to obtain the result below:

```
[9] %sql SELECT * \
        FROM SPACEX_TBL \
        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
04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
08-12-2010	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)
22-05-2012	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
01-03-2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

Total Payload Mass:
45,596 kg (total) carried by boosters launched by NASA (CRS)

The below code is run to obtain the result:

%sql SELECT SUM(PAYLOAD_MASS__KG_) \

FROM SPACEX_TBL \

WHERE CUSTOMER = 'NASA (CRS)';'

```
0s [10] %sql SELECT SUM(PAYLOAD_MASS__KG_) \
          FROM SPACEX_TBL \
          WHERE CUSTOMER = 'NASA (CRS)';

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

SUM(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

Average Payload Mass:
2,928 kg (average) carried by booster version F9 v1.1

The below code is run to obtain the result:

```
%sql SELECT AVG(PAYLOAD_MASS__KG_) \
    FROM SPACEX_TBL \
    WHERE BOOSTER_VERSION = 'F9 v1.1';
```

0s

```
%sql SELECT AVG(PAYLOAD_MASS__KG_) \
    FROM SPACEX_TBL \
    WHERE BOOSTER_VERSION = 'F9 v1.1';
```

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

AVG(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

1st Successful Landing in Ground Pad:
12/22/2015

The below code is run to obtain the result:

%sql SELECT MIN(DATE) FROM SPACEX_TBL WHERE [LANDING _OUTCOME] = 'Success (ground pad)'

```
[18] %sql SELECT MIN(DATE) FROM SPACEX_TBL WHERE [Landing _Outcome] = 'Success (ground pad)'

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

MIN(DATE)
01-05-2017

Successful Drone Ship Landing with Payload between 4000 and 6000

Booster Drone Ship Landing

- Booster mass greater than 4,000 but less than 6,000
- JSCAT-14, JSCAT-16, SES-10, SES-11 / EchoStar 105

The below code is run to obtain the result:

%sql SELECT PAYLOAD \
FROM SPACEX_TBL \
WHERE [Landing _Outcome] = 'Success (drone ship)' \
AND PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000;

```
%sql SELECT PAYLOAD \
FROM SPACEX_TBL \
WHERE [Landing _Outcome] = 'Success (drone ship)' \
AND PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000;
```

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

Payload
JCSAT-14
JCSAT-16
SES-10
SES-11 / EchoStar 105

Total Number of Successful and Failure Mission Outcomes

Total Number of Successful and Failed Mission Outcomes

- 1 Failure in Flight
- 99 Success
- 1 Success (payload status unclear)

The below code is run to obtain the result:

```
%sql SELECT MISSION_OUTCOME, COUNT(*) as total_number \
FROM SPACEX_TBL \
GROUP BY MISSION_OUTCOME;
```

We note that success missions have been far higher than failed ones

```
[13] %sql SELECT MISSION_OUTCOME, COUNT(*) as total_number \
     FROM SPACEX_TBL \
     GROUP BY MISSION_OUTCOME;

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

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

Boosters Carried Maximum Payload

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

The below code is run to obtain the result:

```
%sql SELECT BOOSTER_VERSION \
FROM SPACEX_TBL \
WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEX_TBL);
```

```
[14] %sql SELECT BOOSTER_VERSION \
     FROM SPACEX_TBL \
     WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEX_TBL);

 * 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

We note that the booster versions carrying maximum payloads include the B1048, B1049, B1051, B1056, B1058 and B1060 versions.

2015 Launch Records

In 2015

- Showing month, date, booster version, launch site and landing outcome

The below code is run to obtain the result:

%sql SELECT substr(Date,4,2) as month, DATE,BOOSTER_VERSION, LAUNCH_SITE, [Landing _Outcome] \
FROM SPACEX_TBL \
where [Landing _Outcome] = 'Failure (drone ship)' and substr(Date,7,4)='2015';

```
[15] %sql SELECT substr(Date,4,2) as month, DATE,BOOSTER_VERSION, LAUNCH_SITE, [Landing _Outcome] \
     FROM SPACEX_TBL \
     where [Landing _Outcome] = 'Failure (drone ship)' and substr(Date,7,4)='2015';

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

month	Date	Booster_Version	Launch_Site	Landing _Outcome
01	10-01-2015	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
04	14-04-2015	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

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

Ranked Descending

• Count of landing outcomes between 2010 06 04 and 2017 03 20 in descending order

The below code is run to obtain the result:

%sql SELECT [Landing _Outcome], count(*) as count_outcomes \
FROM SPACEX_TBL \
WHERE DATE between '04-06-2010' and '20-03-2017' group by [Landing _Outcome] order by count_outcomes DESC;

```
%sql SELECT [Landing _Outcome], count(*) as count_outcomes \
FROM SPACEX_TBL \
WHERE DATE between '04-06-2010' and '20-03-2017' group by [Landing _Outcome] order by count_outcomes DESC;
```

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

Landing _Outcome	count_outcomes
Success	20
No attempt	10
Success (drone ship)	8
Success (ground pad)	6
Failure (drone ship)	4
Failure	3
Controlled (ocean)	3
Failure (parachute)	2
No attempt	1

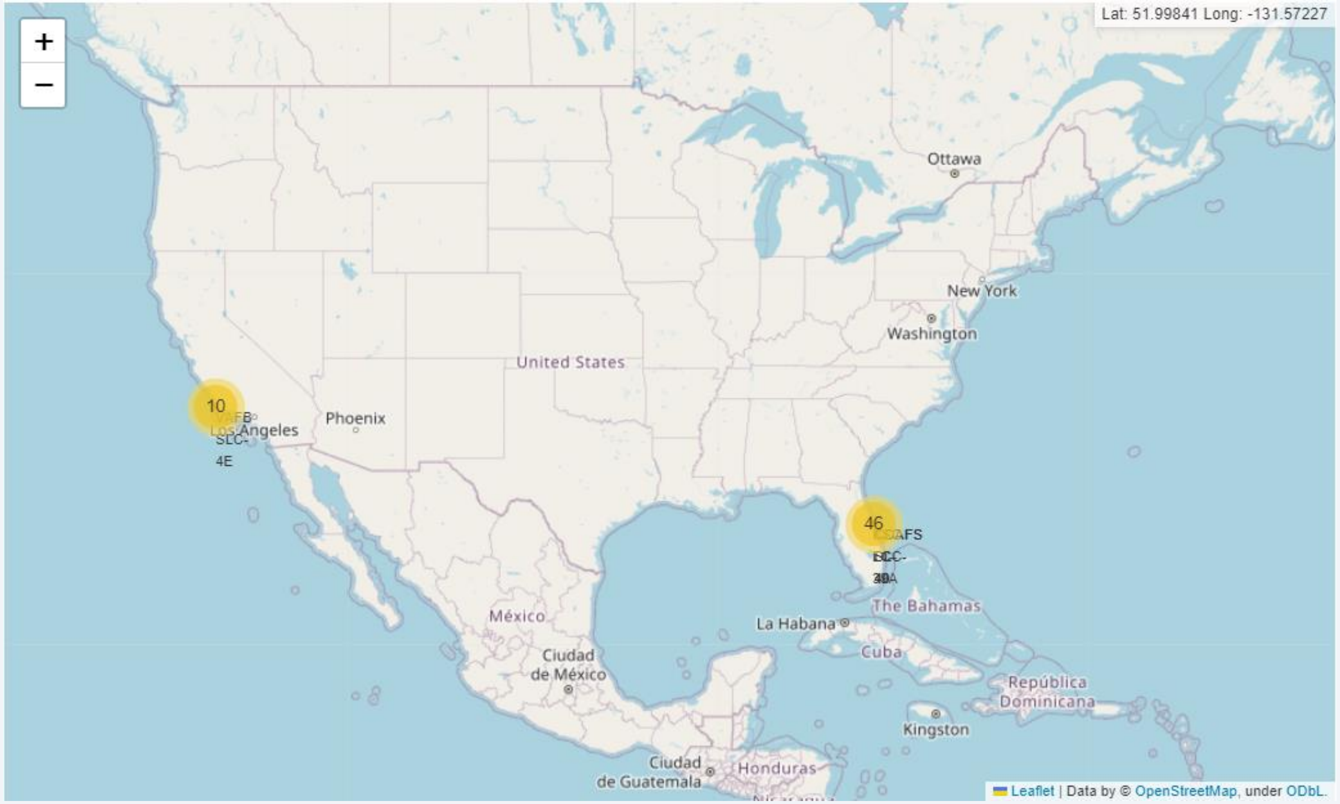
Section 3

Launch Sites Proximities Analysis

Launch Sites

With Markers

Launch sites are near Equator.
The closer the launch site to the equator, the easier it is to launch to equatorial orbit, and the more natural boost obtained from Earth's rotation for a prograde orbit. This helps save the cost of putting in extra fuel and boosters.

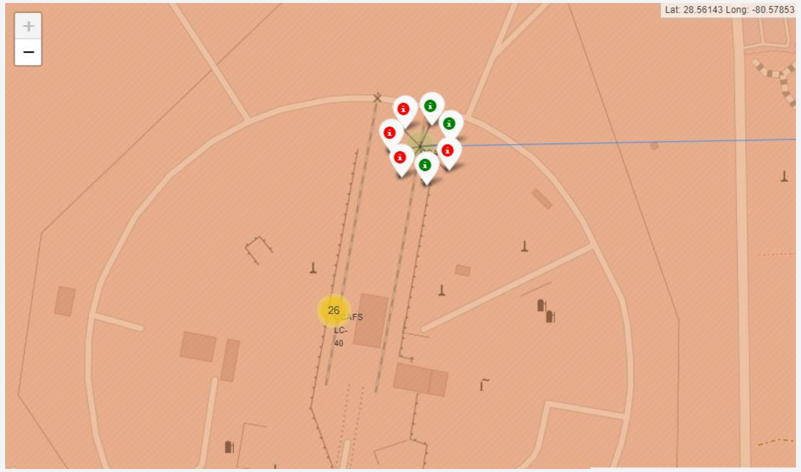


Launch Outcomes

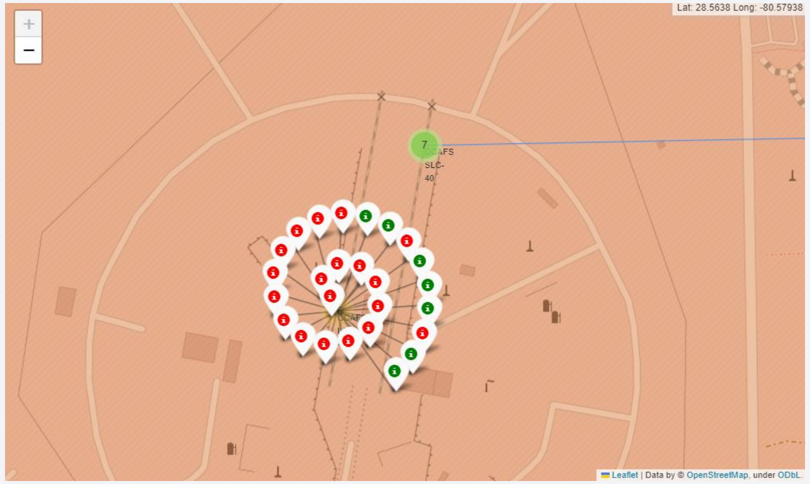
At Each Launch Site

Outcomes:
- Green markers represent successful launches
- Red markers represent unsuccessful launches
- Launch site CCAFS SLC-40 has a 3/7 success rate (42.9%)

Launch Outcomes

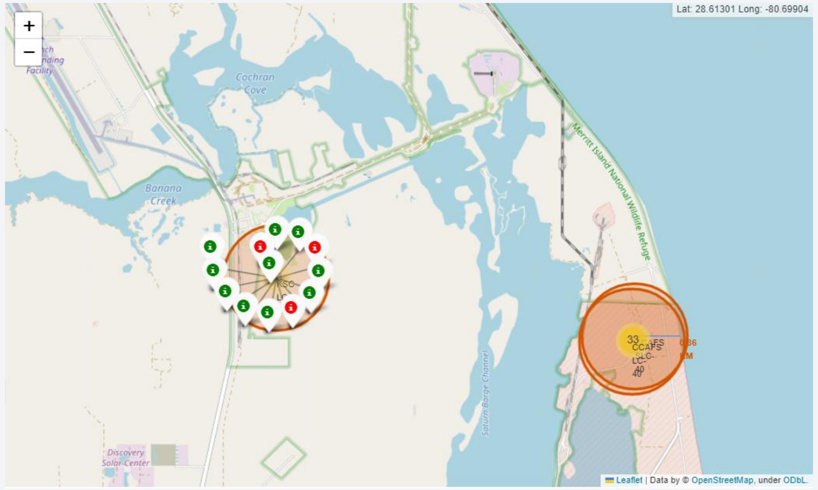


- Site 7, East coast
- We note that unsuccessful launches (4) outweigh the successful ones (3) at this site.

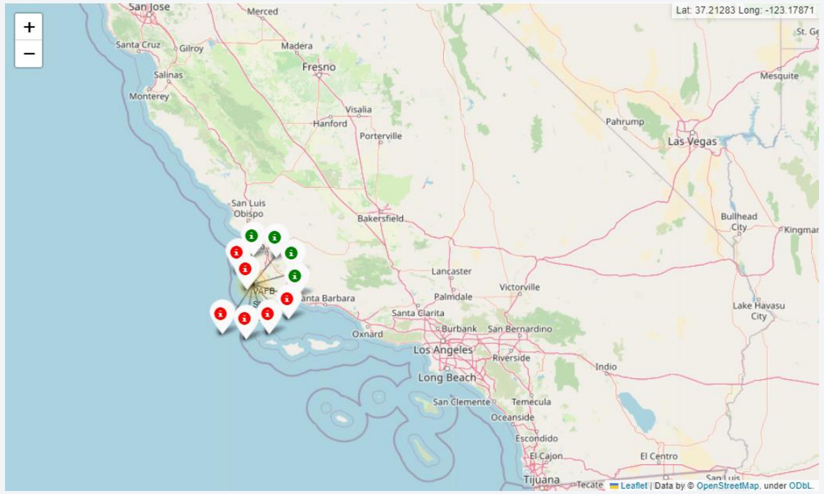


- Site 26, East coast
- We note that unsuccessful launches (19) outweigh the successful ones (7) at this site.

Launch Outcomes



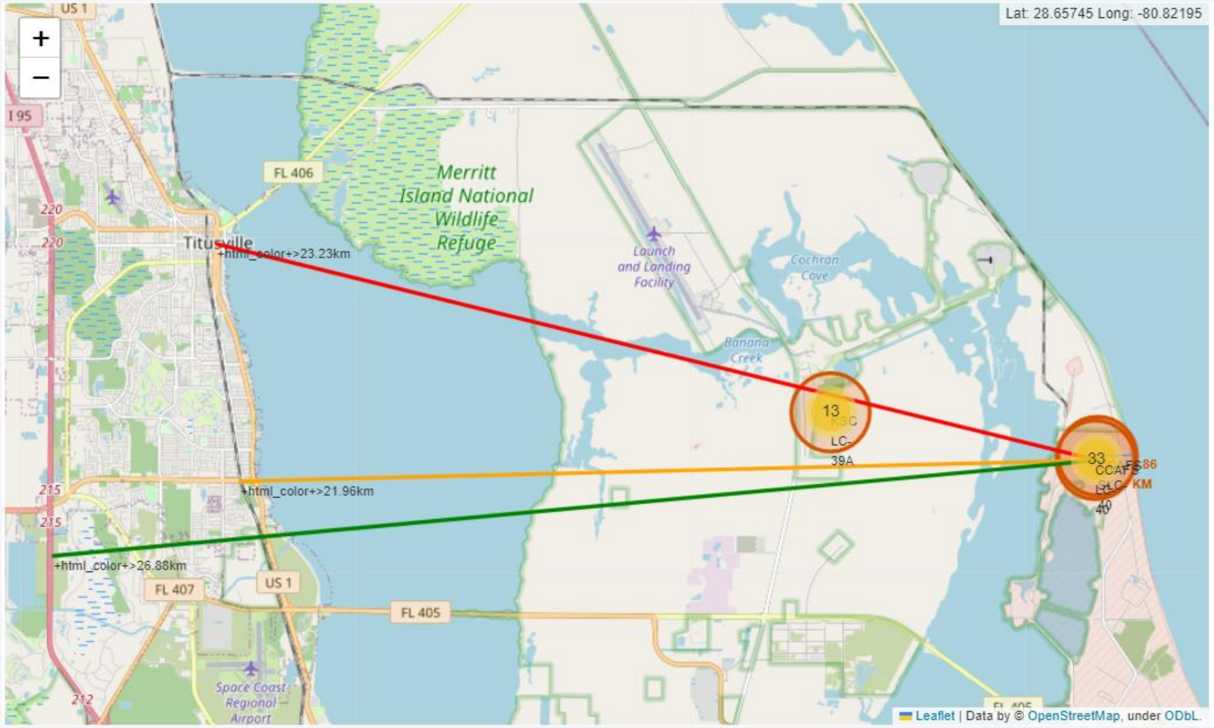
- Site 13, East coast
- We note that successful launches (10) outweigh the unsuccessful ones (3) at this site.



- Site 10, West coast
- We note that unsuccessful launches (6) outweigh the successful ones (4) at this site.

Distance to Proximities

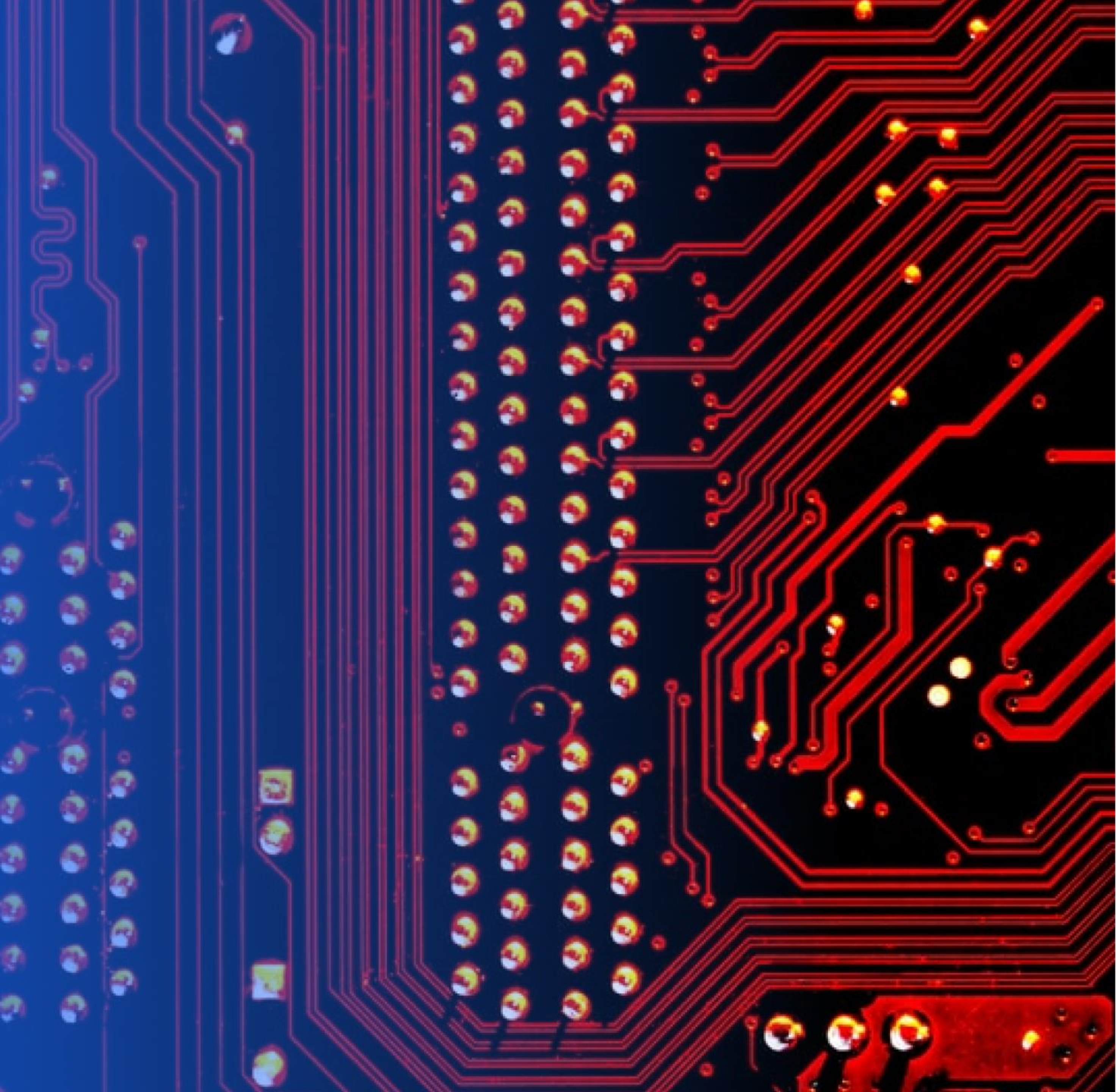
CCAFS SLC-40
- 0.86 km from nearest coastline
- 21.96 km from nearest railway
- 23.23 km from nearest city
- 26.88 km from nearest highway



- Coasts: help ensure that spent stages dropped along the launch path or failed launches don't fall on people or property.
- Safety/Security needs to be an exclusion zone around the launch site to keep unauthorized people away and keep people safe.
- Transportation/Infrastructure and Cities need to be away from anything a failed launch can damage, but still close enough to roads/rails/docks to be able to bring people and material to or from it in support of launch activities.

Section 4

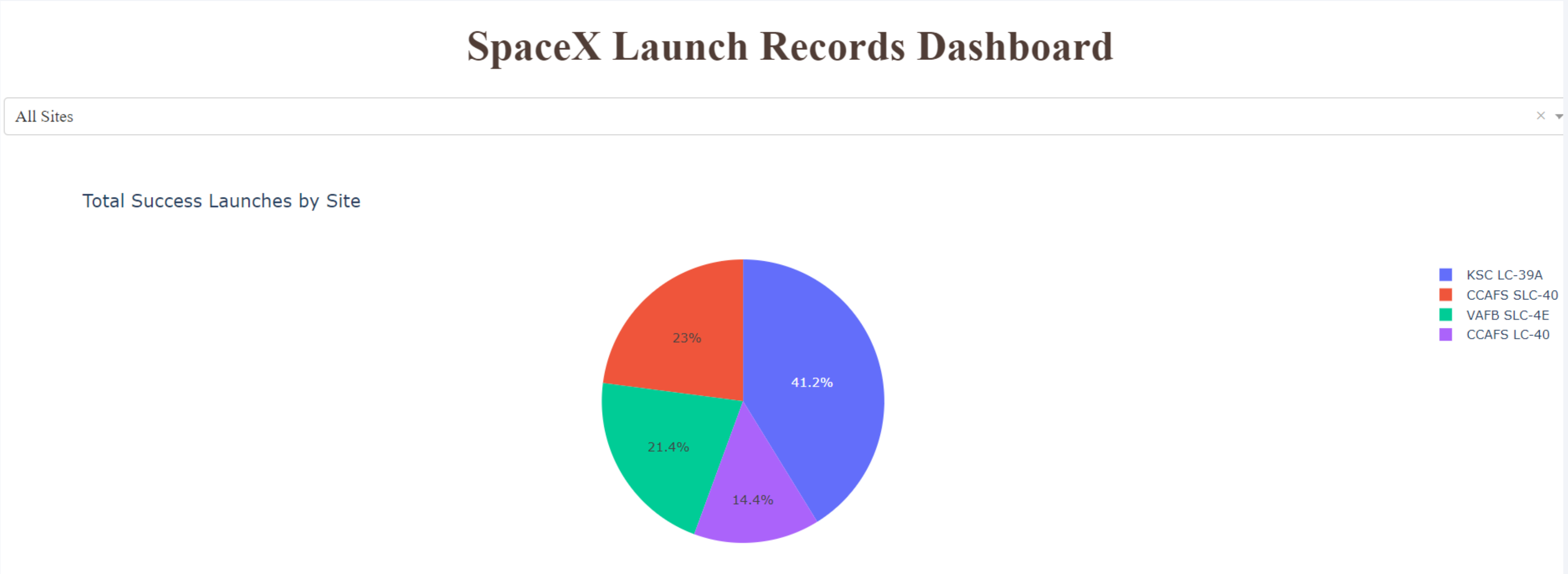
Build a Dashboard with Plotly Dash



Launch Success by Site

Success as percentage of total

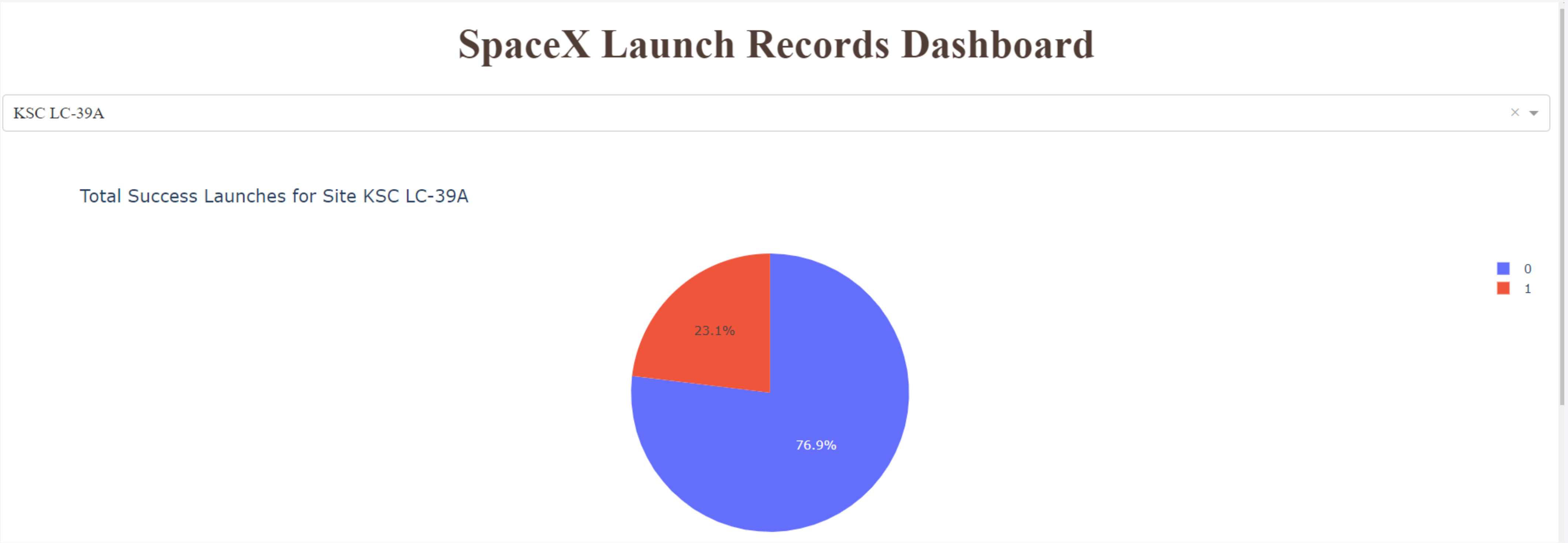
• KSC LC-39A has the most successful launches amongst launch sites (41.2%), followed by CCAFS SLC-40, VAFB SLC-4E and lastly CCAFS LC-40.



Launch Success (KSC LC-29A)

Success as Percent of Total

- KSC LC-39A has the highest success rate amongst launch sites (76.9%)
- 10 successful launches and 3 failed launches



Payload Mass and Success

All booster versions considered: Payloads between 2,000 kg and 5,000 kg have the highest success rate. (1 indicating successful outcome and 0 indicating an unsuccessful outcome)

Payload Mass and Success

	Payload 0-2K		Payload 2-4K		Payload 4-6K		Payload 6-8K		Payload 8-10K	
	Success	Failure	Success	Failure	Success	Failure	Success	Failure	Success	Failure
V1.0	0	3	0	0	0	0	0	0	0	0
V1.1	1	5	0	5	0	4	0	0	0	0
FT	1	0	8	2	4	3	0	2	0	0
B4	1	0	3	1	1	1	0	2	1	1
B5	0	0	1	0	0	0	0	0	0	0

Payload 0-2K
We note higher failures in regard to the V1.0 booster

Payload2-4K
We note higher failures in regard to the V1.1 booster
We note higher success in regard to the FT and B4 booster

Payload 4-6K
We note higher failures in regard to the V1.1 booster
We note slightly higher success in regard to the FT

Payload 6-8K
We note higher failures in regard to the FT and B4 booster

Payload 8-10K
We note the B5 had one success and one failure

Section 5

Predictive Analysis (Classification)

Classification Accuracy

Accuracy

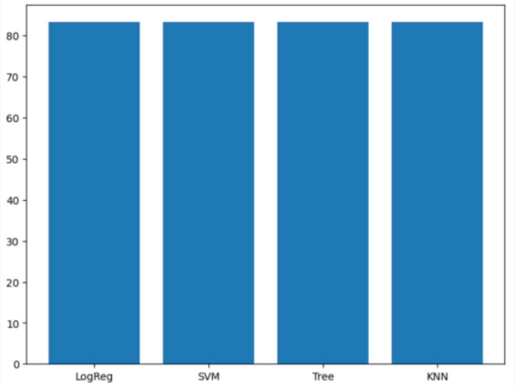
- All the models performed at about the same level and had the same scores and accuracy. This is likely due to the small dataset . The **Decision Tree model** slightly outperformed the rest when looking at .best_score_ (score of 0.875)
- NB: best_score_ is the average of all cv folds for a single combination of the parameters

	LogReg	SVM	Tree	KNN
Jaccard_Score	0.800000	0.800000	0.800000	0.800000
F1_Score	0.888889	0.888889	0.888889	0.888889
Accuracy	0.833333	0.833333	0.833333	0.833333

```
models = {'KNeighbors':knn_cv.best_score_,
              'DecisionTree':tree_cv.best_score_,
              'LogisticRegression':logreg_cv.best_score_,
              'SupportVector': svm_cv.best_score_}

bestalgorithm = max(models, key=models.get)
print('Best model is', bestalgorithm,'with a score of', models[bestalgorithm])
if bestalgorithm == 'DecisionTree':
    print('Best params is :', tree_cv.best_params_)
if bestalgorithm == 'KNeighbors':
    print('Best params is :', knn_cv.best_params_)
if bestalgorithm == 'LogisticRegression':
    print('Best params is :', logreg_cv.best_params_)
if bestalgorithm == 'SupportVector':
    print('Best params is :', svm_cv.best_params_)
```

```
Best model is DecisionTree with a score of 0.875
Best params is : {'criterion': 'gini', 'max_depth': 14, 'max_features': 'sqrt', 'min_samples_leaf': 2, 'min_samples_split': 10, 's
plitter': 'random'}
```



Confusion Matrix

Performance Summary

• A confusion matrix summarizes the performance of a classification algorithm

• All the confusion matrices were identical

• The fact that there are false positives (Type 1 error) is not good

Confusion Matrix Outputs:

• 12 True positive

• 3 True negative

• 3 False positive

• 0 False Negative

Precision = TP / (TP + FP)

• 12 / 15 = .80

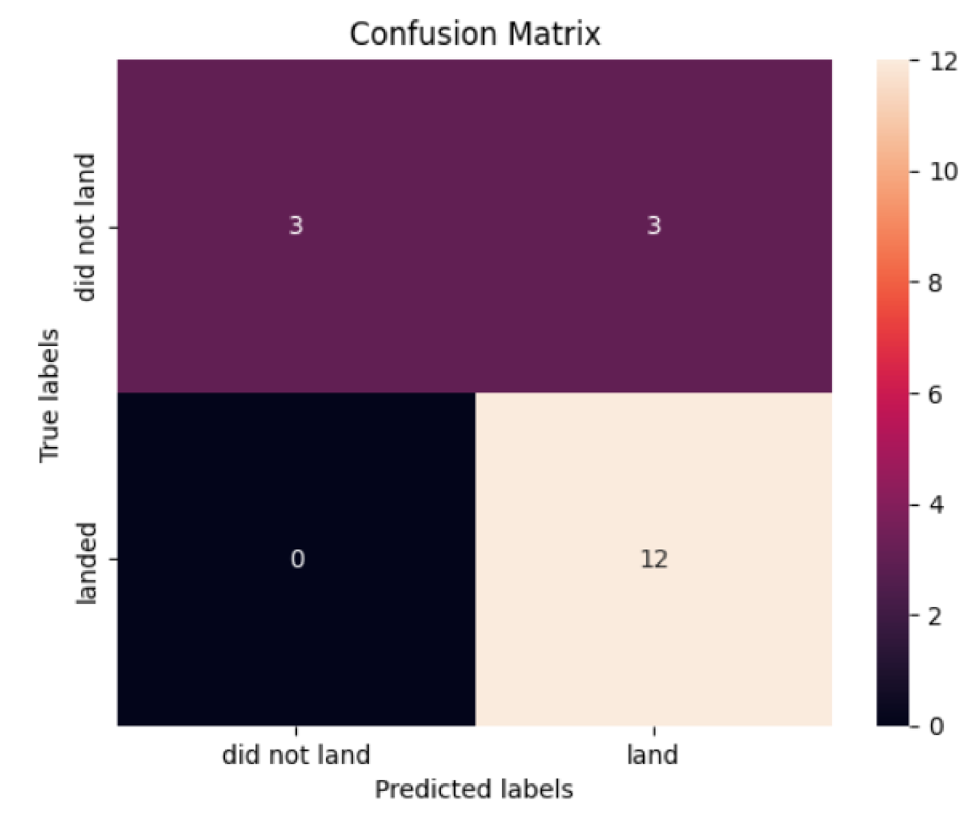
Recall= TP / (TP + FN)

• 12 / 12 = 1

F1 Score = 2 * (Precision * Recall) / (Precision + Recall)

• 2 * (.8 * 1) / (.8 + 1) = .89

Accuracy = (TP + TN) / (TP + TN + FP + FN) = .833



Conclusions

Research conclusion

- The models performed similarly on the test set with the decision tree model slightly outperforming
- Most of the launch sites are near the equator for an additional natural boost due to the rotational speed of earth which helps save the cost of putting in extra fuel and boosters
- All the launch sites are close to the coast
- Launch success increases over time
- KSC LC-39A has the highest success rate among launch sites. Has a 100% success rate for launches less than 5,500 kg
- Orbits ES L1, GEO, HEO, and SSO have a 100% success rate
- Across all launch sites, the higher the payload mass (kg), the higher the success rate

Points to consider

- A larger dataset will help build on the predictive analytics results to help understand if the findings can be generalizable to a larger data set
- Additional feature analysis or principal component analysis should be conducted to see if it can help improve accuracy
- XGBoost is a powerful model which was not utilized in this study. It would be interesting to see if it outperforms the other classification models

Appendix

For other relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets created during this project, please refer to the github links in the presentation.

Thank you!