IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Hugo Cea
April 12, 2024

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection using SpaceX API
  - Data Collection with Web Scraping
  - Data Wrangling
  - Exploratory Data Analysis using SQL
  - EDA DataViz Using Python Pandas and Matplotlib
  - Launch Sites Analysis with Folium-Interactive Visual Analytics and Ploty Dash
  - Machine Learning Landing Prediction
- Summary of all results
  - EDA results
  - Visual Analytics and Dashboards
  - Predictive Analysis

Introduction

- Project background and context
  - SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upward 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 information can be used if an alternate company wants to bid against SpaceX for a rocket launch.
- Problems you want to find answers
  - In this capstone, we will predict if the Falcon 9 first stage will land successfully using data from Falcon 9 rocket launches advertised on its website.

Section 1

Methodology

Methodology

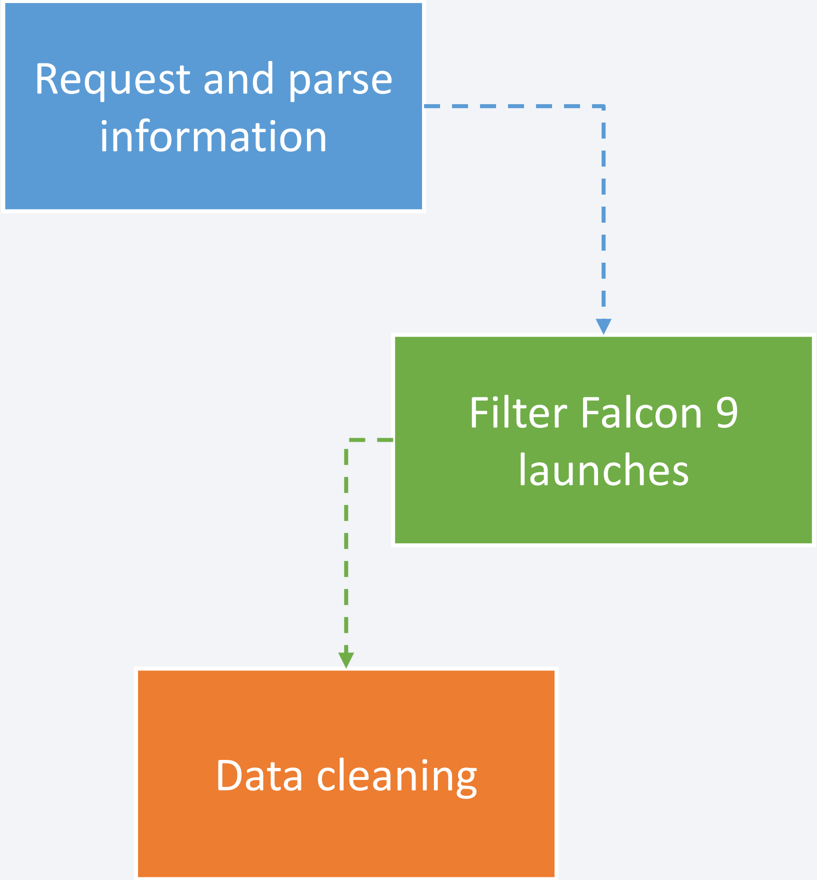
Executive Summary

- Data collection methodology:
  - Space X data was obtained from the following sources:
    - Web Scraping .- Wikipedia (...List_of_Falcon/_9/_and_Falcon_Heavy_launches)
    - Space X API (https://api.spacexdata.com/v4/rockets)
- Perform data wrangling
  - Data Wrangling process applied to calculate the number of launches on each site, number and occurrences of each orbit, the number and occurrence of mission outcome per orbit type and create a landing outcome label from Outcome column.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Data were normalized, splitting in training and test datasets to evaluate under 4 different methods and calculate their accuracy

Data Collection – SpaceX API

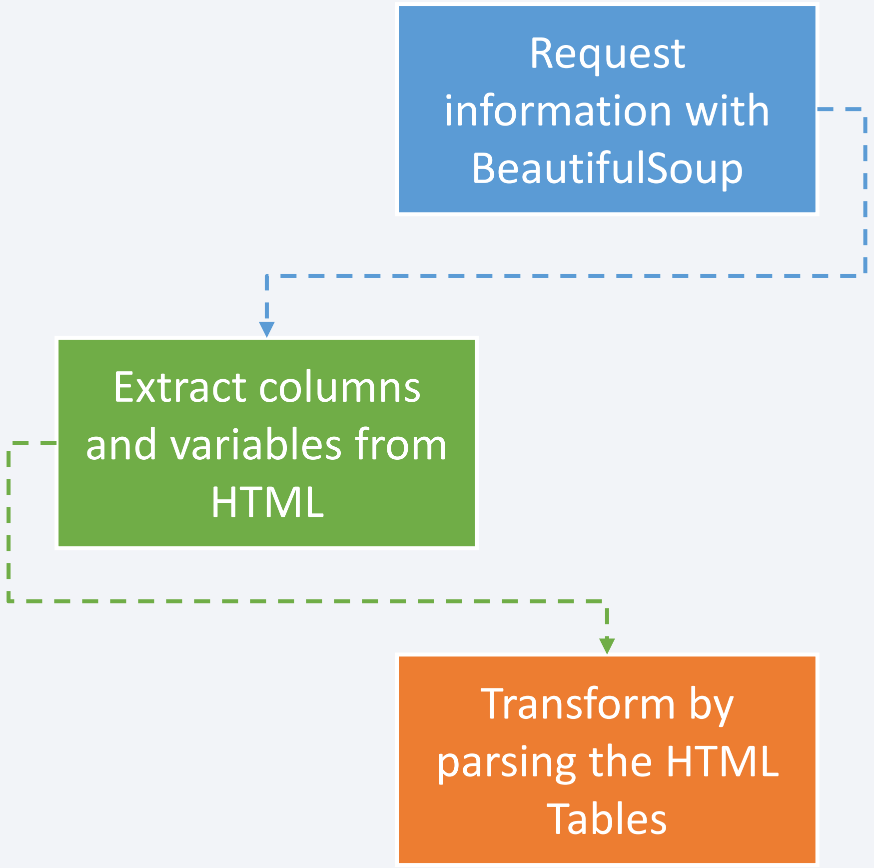
- Data was obtained and parsed from SpaceX API
- Filter was apply to only include Falcon 9 launches
- Deal with missing values

[Data Collection with API (GitHub)]



Data Collection – Web Scraping

- Data from Web Scraping was obtained from Wikipedia (...List_of_Falcon/_9/_and_Falcon_Heavy _launches)

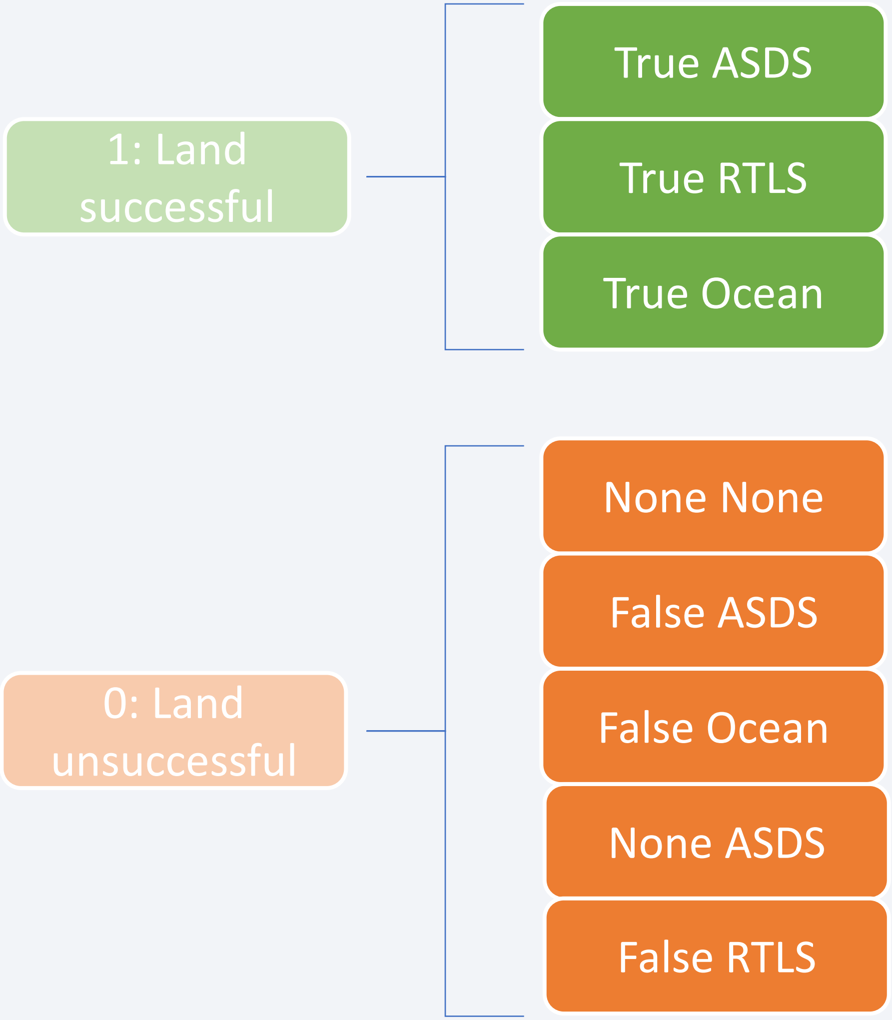


[Web Scraping (GitHub)]

Data Wrangling

- Calculated the number and occurrence of mission outcome of the orbits
- Any ‘True’ landing outcome was successful landings. The rest represent failure to land.
- Then create a landing outcome label from the ‘Outcome’ column using zeroes and ones to categorize.

[Data Wrangling (GitHub)]



EDA with SQL

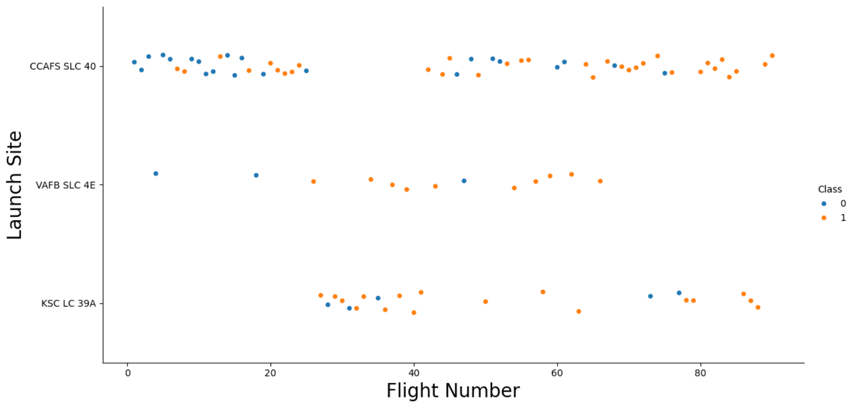
SQL queries performed:

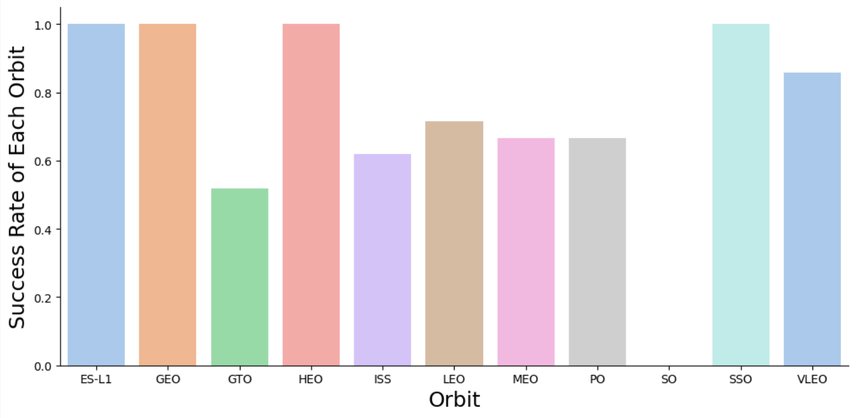
- Names of the unique launch sites in the space mission
- 5 records where launch sites begin with the string 'CCA’
- Total payload mass carried by boosters launched by NASA (CRS)
- Average payload mass carried by booster version F9 v1.1
- Date when the first successful landing outcome in ground pad was achieved
- Names of the boosters which have success in drone ship and have payload mass between 4000 and 6000
- Total number of successful and failure mission outcomes
- Names of the booster versions which have carried the maximum payload mass
- Failed landing outcomes in drone ship, booster versions, and launch site names for the months in year 2015
- Rank of the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20

[EDA with SQL (GitHub)]

EDA with Data Visualization

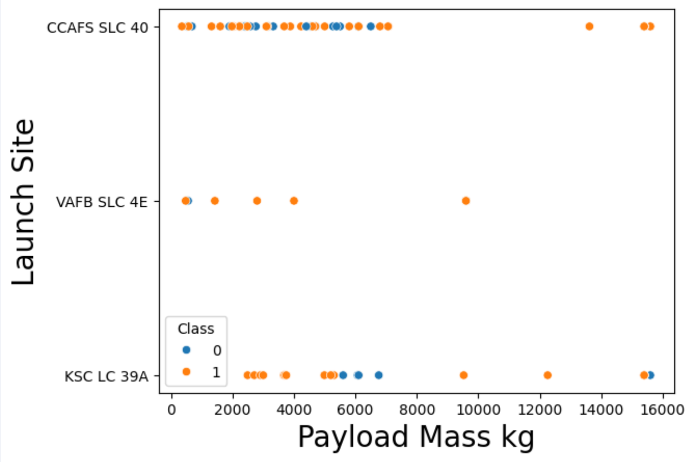
- Relationship between Flight Number and Launch Site





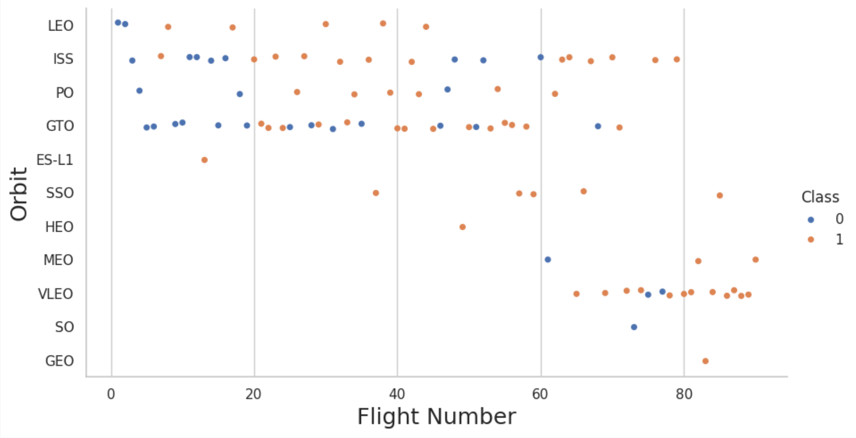
- Relationship between Payload and Launch Site

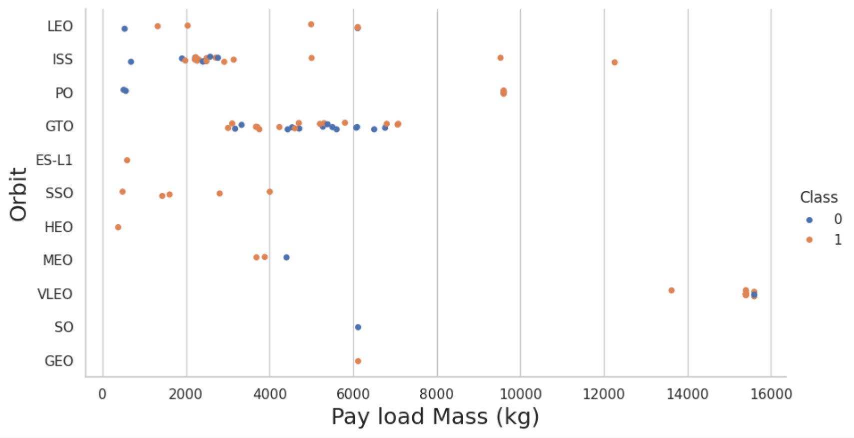
- Relationship between success rate of each orbit type



EDA with Data Visualization

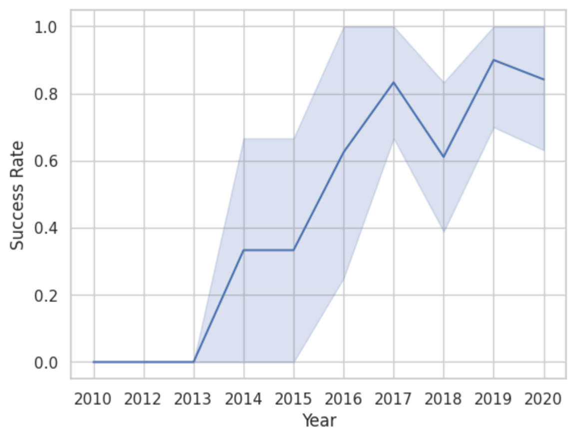
- Relationship between FlightNumber and Orbit type





- Relationship between Payload and Orbit type

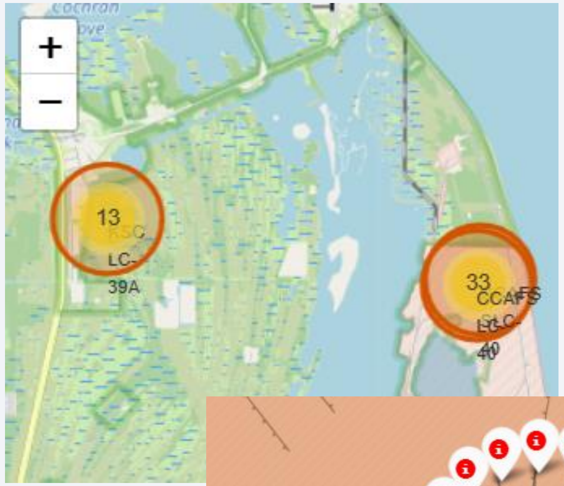
- Visualize the launch success yearly trend



[EDA with Data Visualization (GitHub)]

Build an Interactive Map with Folium

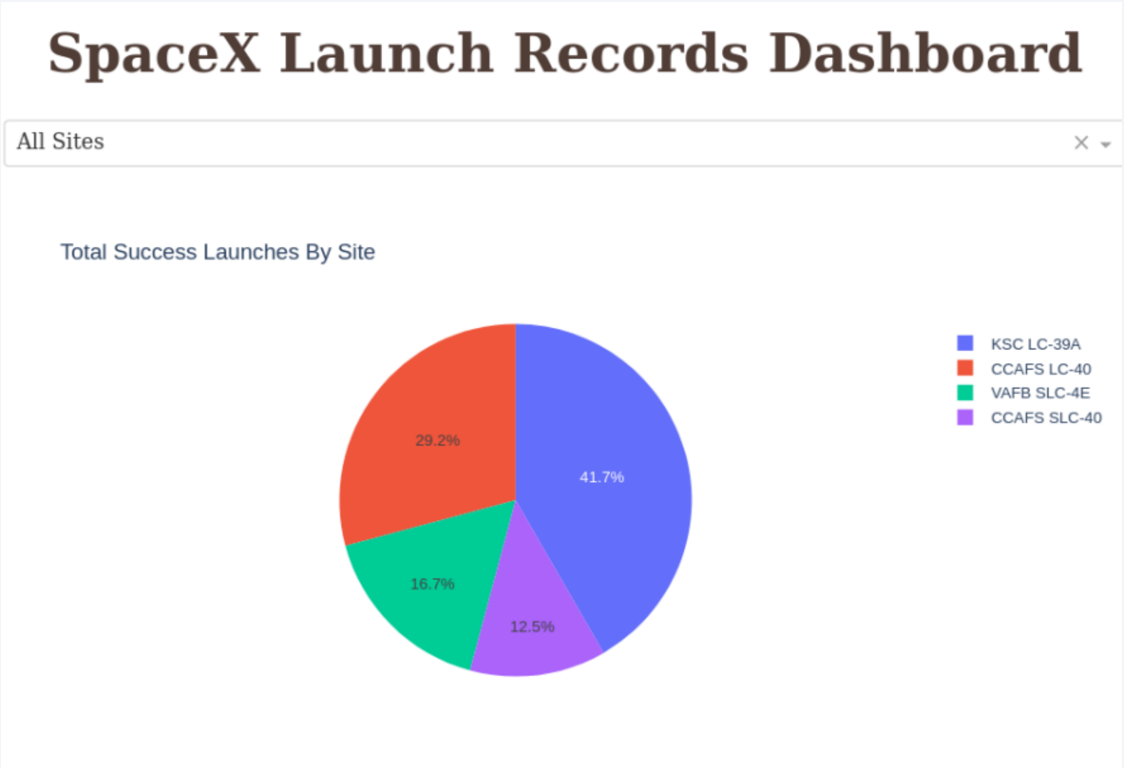
- Map objects used: Markers, marker clusters, circles and polyline (line):
  - Markers indicate points like launch site
  - Circles highlight their areas
  - Marker clusters represent group of events
  - Lines are used to represent distance between points like launch sites and cities



[Visual Analytics with Folium (GitHub)]

Build a Dashboard with Plotly Dash

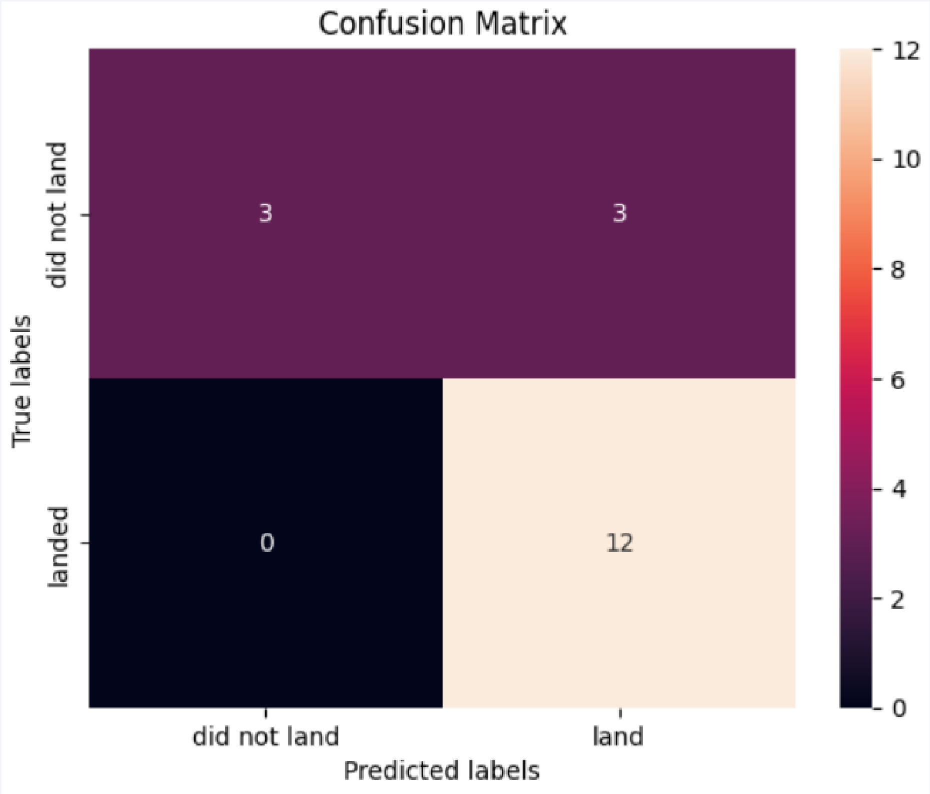
- Plots/graphs and interactions:
  - Launch site dropdown list
  - Range slider to select payload mass
  - Pie chart based on site dropdown selection
  - Scatter plot to show success and failure according to payload and booster version



[Dashboard with Ploty Dash (GitHub)]

Predictive Analysis (Classification)

- Four algorithms were compared:
    - Logistic Regression
    - Support Vector Machine
    - Decision Tree
    - K-Nearest Neighbors
- Dataframe split into training data and test data
- Confusion matrix and score function applied to identify accuracy



Machine Learning (GitHub)

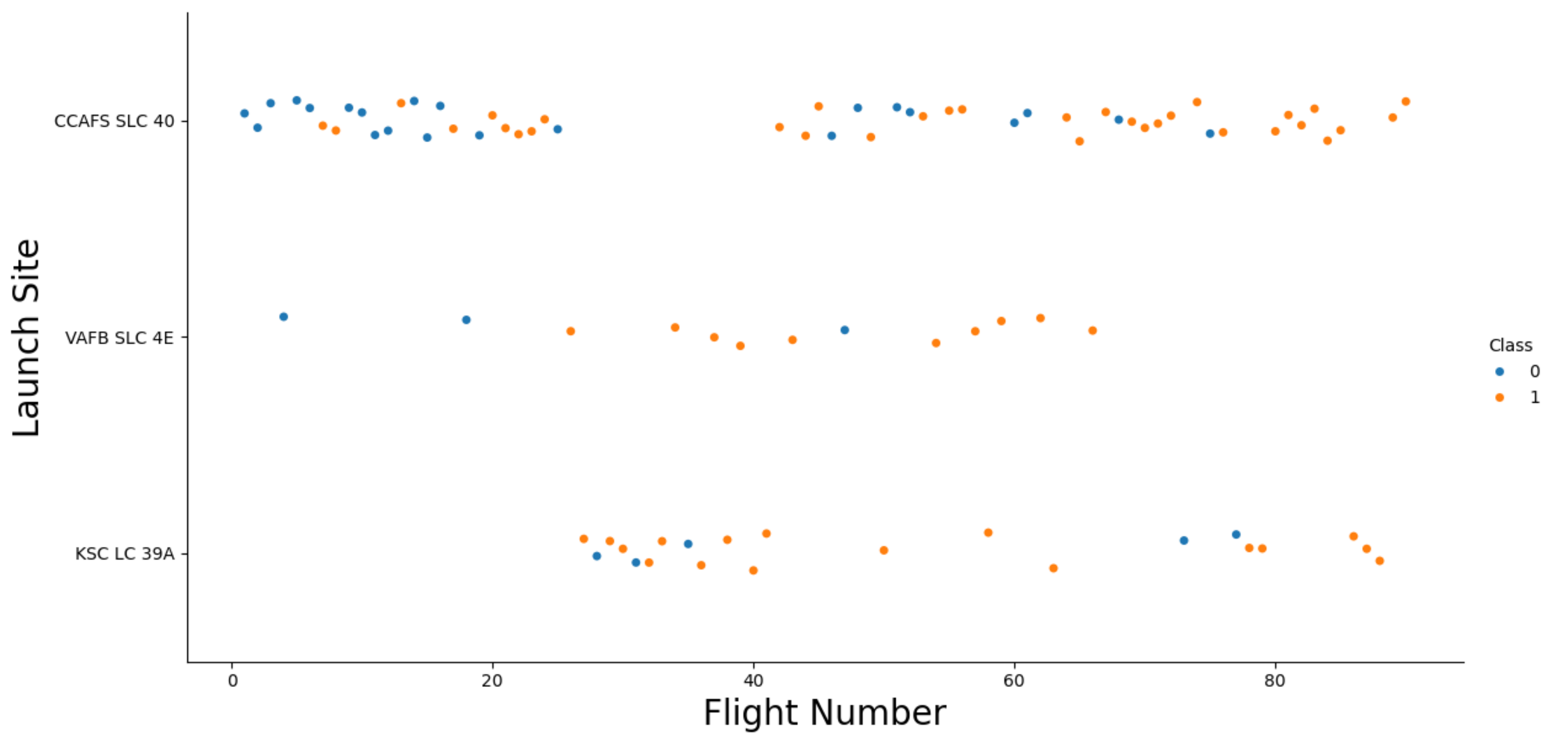
Results

- Exploratory data analysis results
  - High success rate at Falcon 9 booster versions at landing in drone ships
  - High failure rate at Falcon 9 booster versions at landing in drone ships
  - Launch success rate increases over time
  - Higher success rate for higher orbits
- Interactive analytics demo in screenshots
  - Most launches happens at east cost launch sites.
  - Higher success rate for Kennedy Space center
- Predictive analysis results
  - Each method applied had 0.833.. as accuracy score

Section 2

Insights drawn from EDA

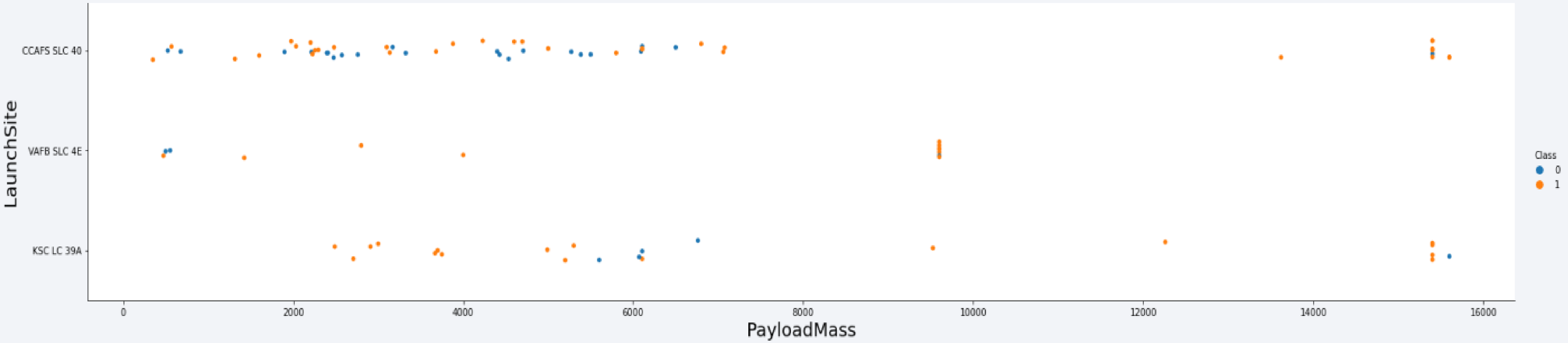
Flight Number vs. Launch Site



Explanation:

- CCAFS SLC 40 has more attempts
- More of the launches at CCAFS SLC 40 were successful

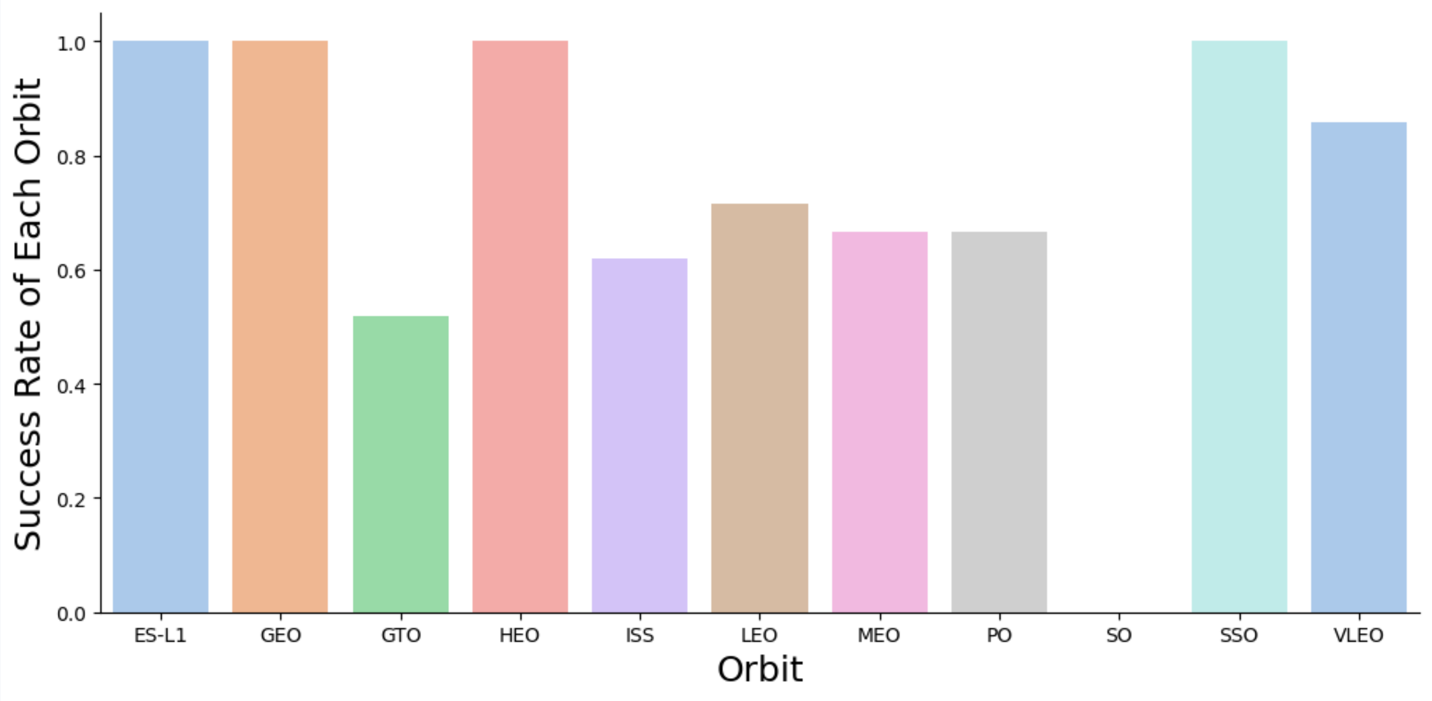
Payload vs. Launch Site



Explanation:

- Payload over 9000 kgs have high success rate
- CCAFS SLC 40 has higher success rate with payloads bigger than 8000 kgs
- No payload over 10000 registered for VAFB SLC 4E

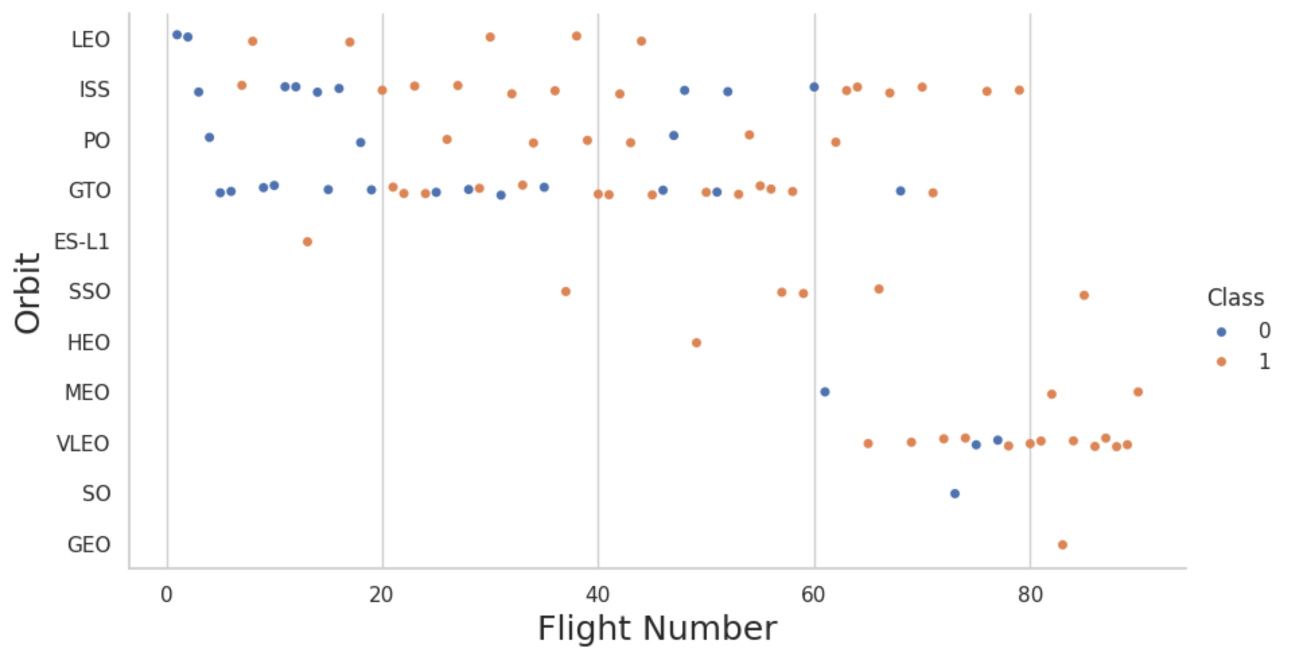
Success Rate vs. Orbit Type



Explanation:

- Best success rate at ES-L1, GEO, HEO and SSO orbit
- Lower success rate at SO. 1 attempt and failed.

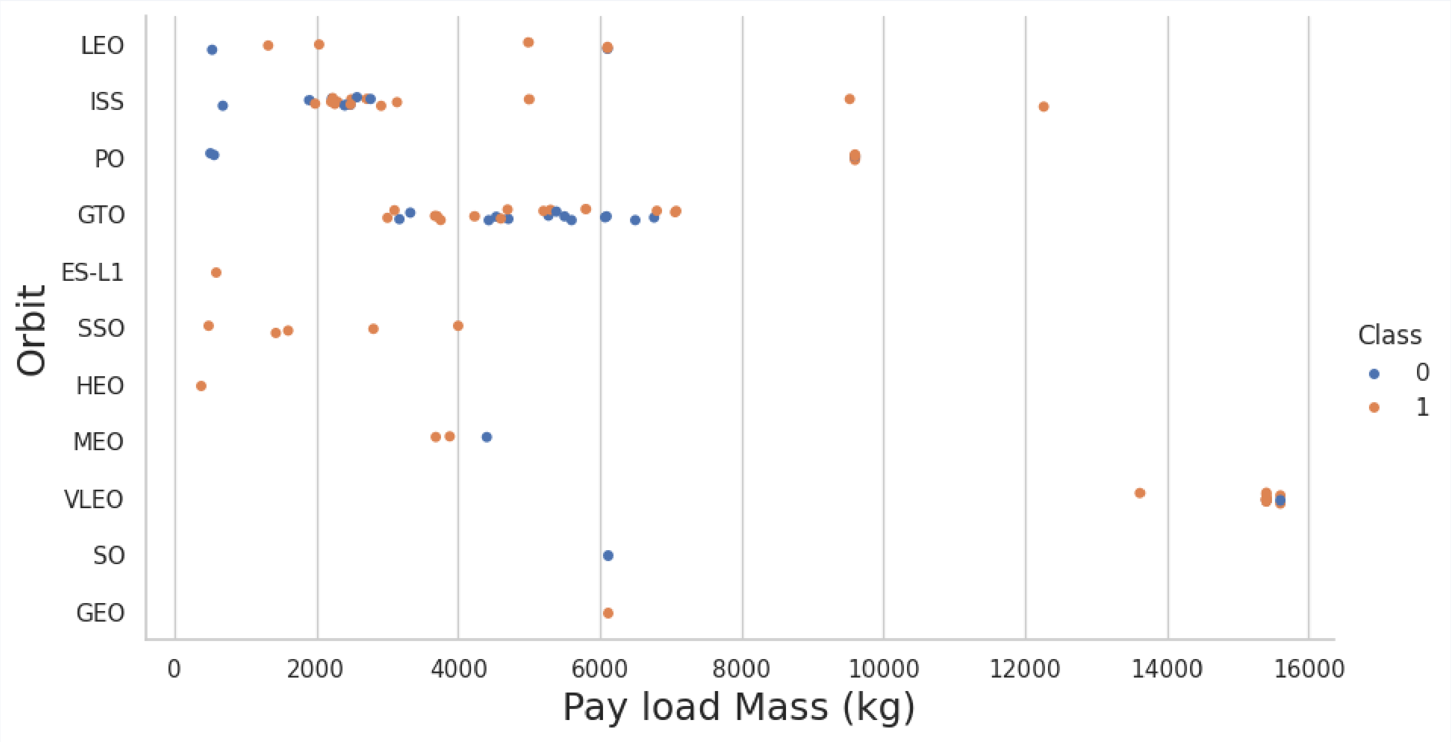
Flight Number vs. Orbit Type



Explanation:

- Success improves over time
- SO records the lowest number of attempts and the highest proportion of failures
- VLEO shows a better chance considering its latest attempts

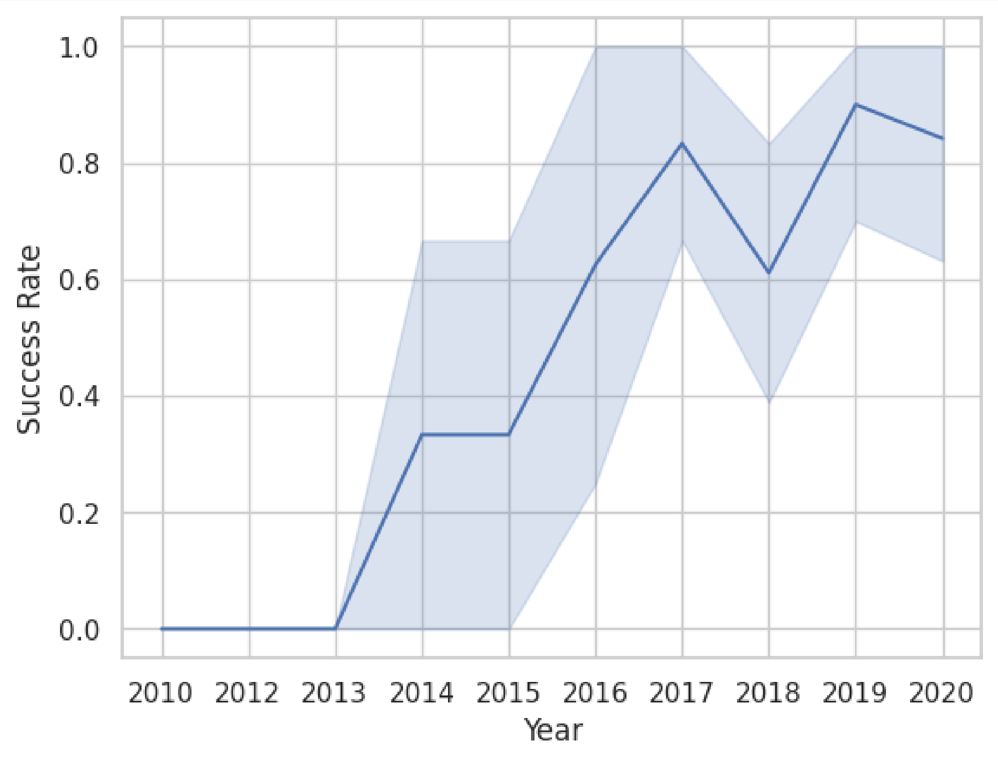
Payload vs. Orbit Type



Explanation:

- Lower success rate ratio between the payload near 6000 and the GTO orbit registered.
- SSO has the best rate of success under 5000

Launch Success Yearly Trend



Explanation:

- Success rate increases over the years
- Success rate grows from 2013 onwards

All Launch Site Names

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

```
%sql SELECT DISTINCT launch_site FROM spacextbl
```

- CCAFS LC-40: Cape Canaveral Launch Complex 40
- CCAFS SLC-40: Cape Canaveral Space Launch Complex 40
- KSC LC-39A: Kennedy Space Center Launch Complex 39
- VAFB SLC-4E: Space Launch Complex 4

Launch Site Names Begin with 'CCA'

```
%sql SELECT * FROM 'spacextbl' WHERE Launch_Site LIKE 'CCA%' LIMIT 5
```

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

```
%sql SELECT SUM(payload_mass__kg_) as "Total Payload Mass Carried", Customer FROM 'spacextbl' WHERE Customer = 'NASA (CRS)'
```

Total Payload Mass Carried	Customer
45596	NASA (CRS)

Average Payload Mass by F9 v1.1

```
%sql SELECT AVG(PAYLOAD_MASS__KG_) as "Average Payload Mass", Customer, Booster_Version FROM 'spacextbl' WHERE Booster_Vers
```

Average Payload Mass	Customer	Booster_Version
2534.6666666666665	MDA	F9 v1.1 B1003

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 SELECT DISTINCT Booster_Version, Payload FROM spacextbl WHERE "Landing_Outcome" = "Success (drone ship)" AND payload_ma
```

Booster_Version	Payload
F9 FT B1022	JCSAT-14
F9 FT B1026	JCSAT-16
F9 FT B1021.2	SES-10
F9 FT B1031.2	SES-11 / EchoStar 105

Total Number of Successful and Failure Mission Outcomes

```
%sql SELECT "Mission_Outcome", COUNT("Mission_Outcome") as Total FROM spacextbl GROUP BY "Mission_Outcome"
```

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

Boosters Carried Maximum Payload

```
%sql SELECT "Booster_Version", "Payload", "Payload_Mass__KG_" FROM SPACEXTBL WHERE "Payload_Mass__KG_" = (SELECT MAX("Payloa
```

Booster_Version	Payload	PAYLOAD_MASS__KG_
F9 B5 B1048.4	Starlink 1 v1.0, SpaceX CRS-19	15600
F9 B5 B1049.4	Starlink 2 v1.0, Crew Dragon in-flight abort test	15600
F9 B5 B1051.3	Starlink 3 v1.0, Starlink 4 v1.0	15600
F9 B5 B1056.4	Starlink 4 v1.0, SpaceX CRS-20	15600
F9 B5 B1048.5	Starlink 5 v1.0, Starlink 6 v1.0	15600
F9 B5 B1051.4	Starlink 6 v1.0, Crew Dragon Demo-2	15600
F9 B5 B1049.5	Starlink 7 v1.0, Starlink 8 v1.0	15600
F9 B5 B1060.2	Starlink 11 v1.0, Starlink 12 v1.0	15600
F9 B5 B1058.3	Starlink 12 v1.0, Starlink 13 v1.0	15600
F9 B5 B1051.6	Starlink 13 v1.0, Starlink 14 v1.0	15600
F9 B5 B1060.3	Starlink 14 v1.0, GPS III-04	15600
F9 B5 B1049.7	Starlink 15 v1.0, SpaceX CRS-21	15600

2015 Launch Records

```
%sql SELECT substr(Date,0,5) as "Year", substr(Date, 6, 2) as "Month", "Booster_Version", "Launch_Site", "Payload", "PAYLOAD
```

Year	Month	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Mission_Outcome	Landing_Outcome
2015	01	F9 v1.1 B1012	CCAFS LC-40	SpaceX CRS-5	2395	Success	Failure (drone ship)
2015	04	F9 v1.1 B1015	CCAFS LC-40	SpaceX CRS-6	1898	Success	Failure (drone ship)

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

```
%sql SELECT * FROM spacextbl WHERE "Landing_Outcome" LIKE 'Success%' AND (Date BETWEEN '2010-06-04' AND '2017-03-20') ORDER
```

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outc
2017-02-19	14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490	LEO (ISS)	NASA (CRS)	Success	Success (gro
2017-01-14	17:54:00	F9 FT B1029.1	VAFB SLC-4E	Iridium NEXT 1	9600	Polar LEO	Iridium Communications	Success	Success (dr
2016-08-14	5:26:00	F9 FT B1026	CCAFS LC-40	JCSAT-16	4600	GTO	SKY Perfect JSAT Group	Success	Success (dr
2016-07-18	4:45:00	F9 FT B1025.1	CCAFS LC-40	SpaceX CRS-9	2257	LEO (ISS)	NASA (CRS)	Success	Success (gro
2016-05-27	21:39:00	F9 FT B1023.1	CCAFS LC-40	Thaicom 8	3100	GTO	Thaicom	Success	Success (dr
2016-05-06	5:21:00	F9 FT B1022	CCAFS LC-40	JCSAT-14	4696	GTO	SKY Perfect JSAT Group	Success	Success (dr
2016-04-08	20:43:00	F9 FT B1021.1	CCAFS LC-40	SpaceX CRS-8	3136	LEO (ISS)	NASA (CRS)	Success	Success (dr
2015-12-22	1:29:00	F9 FT B1019	CCAFS LC-40	OG2 Mission 2 11 Orbcomm-OG2 satellites	2034	LEO	Orbcomm	Success	Success (gro

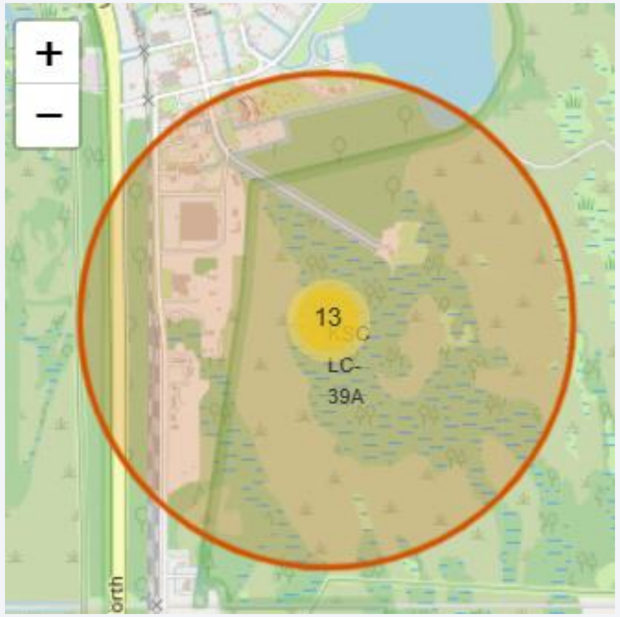
Section 3

Launch Sites Proximities Analysis

Folium Map: Launch Sites

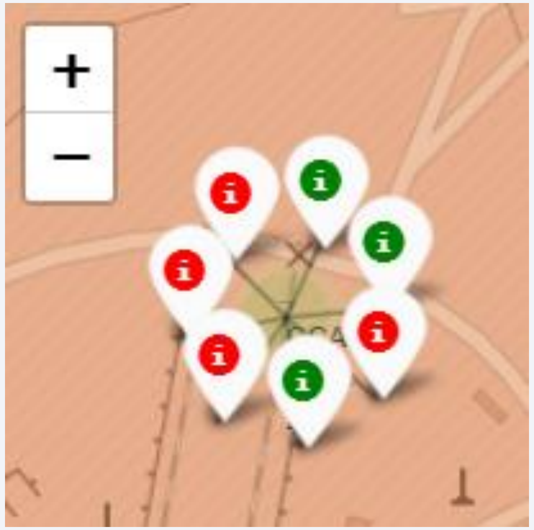
- Launch sites:
  - CCAFS: Cape Canaveral Space Launch Complex
  - KSC: Kennedy Space Center Launch Complex 39
  - VAFB: Vanderberg Space Launch Complex 4

Folium Map: Landing events order by success rate



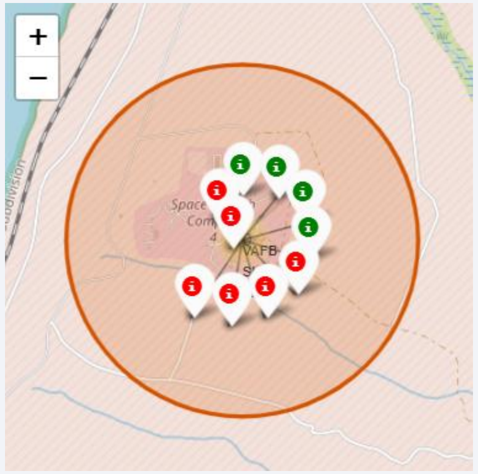
KSC LC-39A

76.92% success



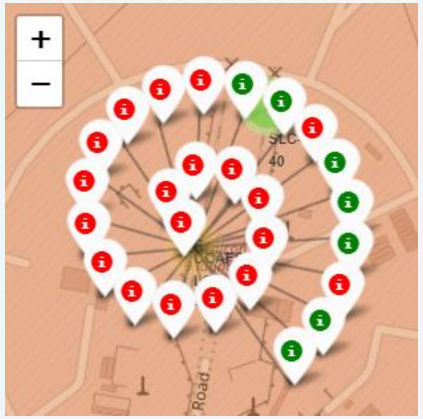
CCAFS SLC-40

42.85% success



VAFB SLC-4E

40.00% success



CCAFS LC-40

26.92% success

Logistics and Safety

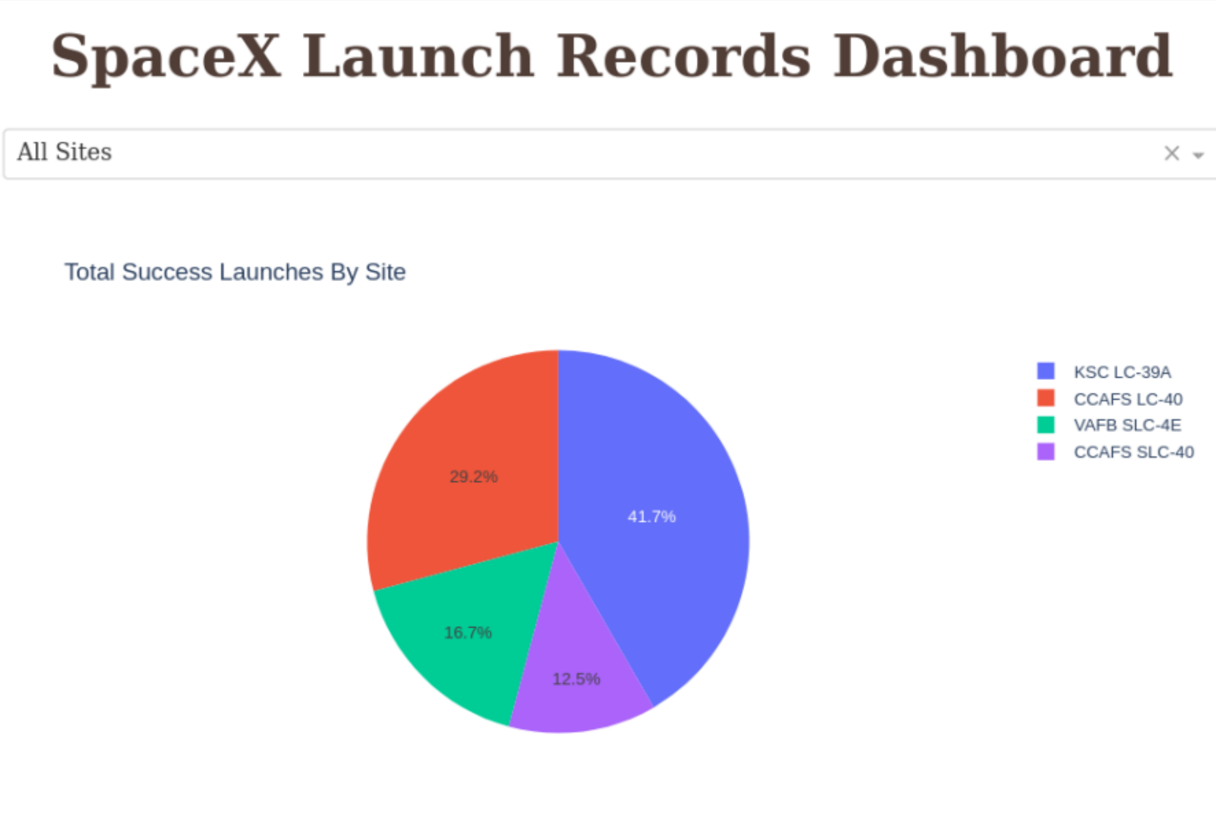
- CCAFS has good logistics since is far from the nearest city and has 2 stations. However, it does not have a high success rate.

Section 4

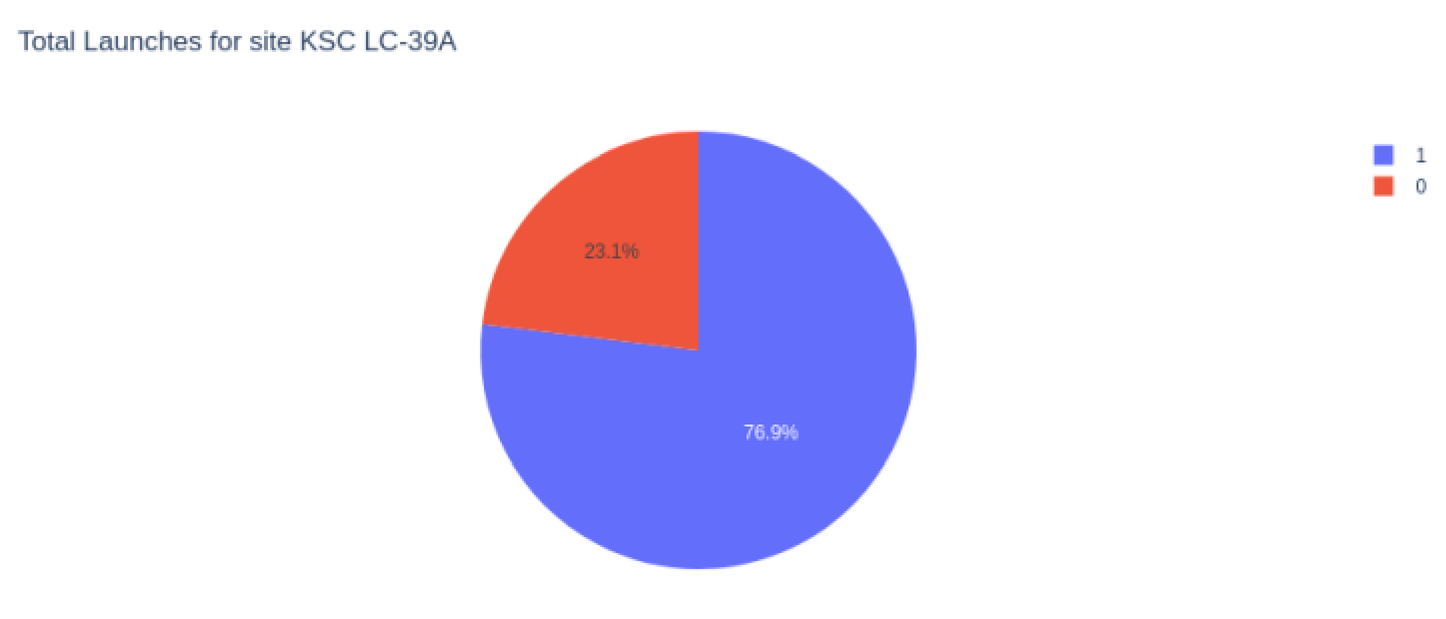
Build a Dashboard with Plotly Dash

Dash: Successful Launches

- KSC LA-39A has the most successful launches
- CCAFS SLC-40 has the least successful launches



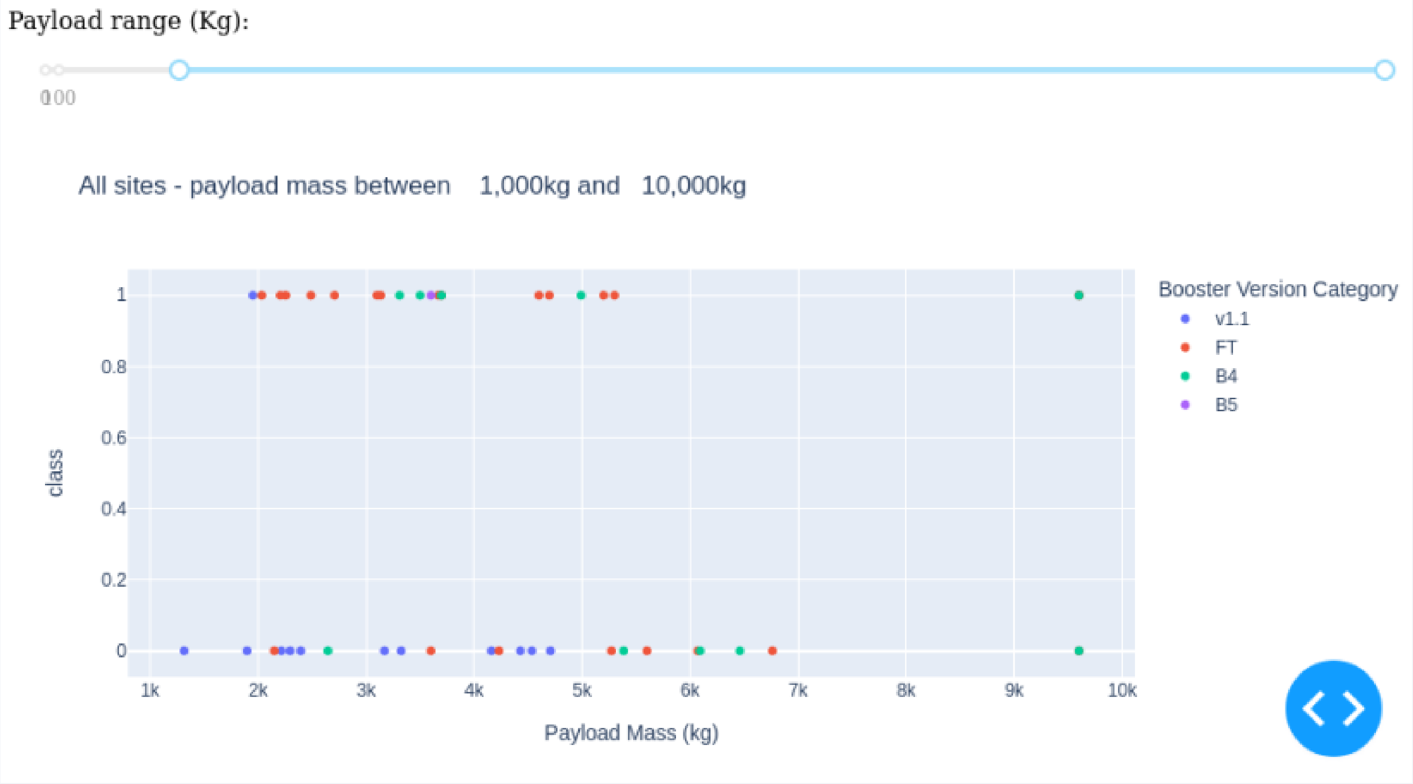
Launch Site with the Highest Launch Success Ratio



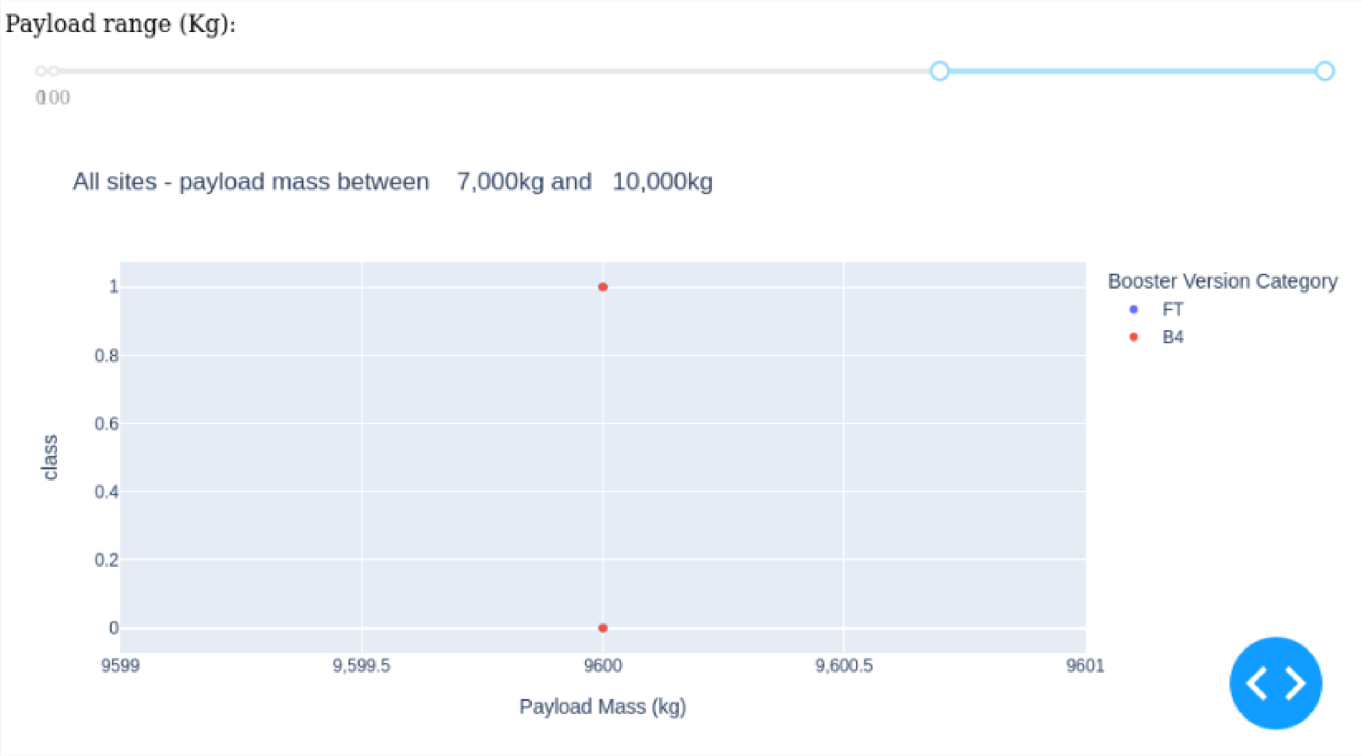
- KSC LC-39A has the highest launch success ratio with a 76.9% according to their launches

Payload Mass vs Launch Outcome

- Payloads under 6000 kgs and FT boosters are the most successful combination.



Payload Mass vs Launch Outcome



- There is not enough data to estimate risk of launches over 7000 kgs

Section 5

Predictive Analysis (Classification)

Classification Accuracy

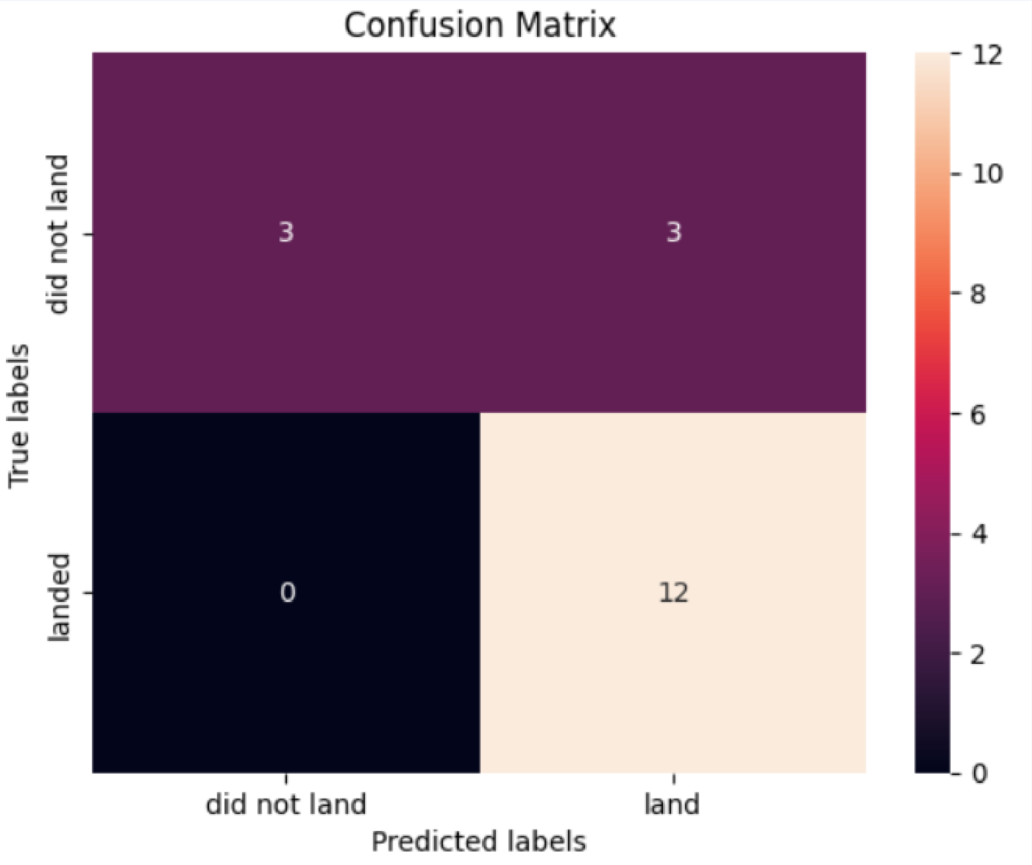
- According to a the accuracy of the test data score, all methods are suitable for the use of prediction.

```
result = pd.DataFrame({'Algorithm Used' : ['Test Accuracy (Score)']})

logisticregression_accuracy=logreg_cv.score(X_test, Y_test)
svm_accuracy=svm_cv.score(X_test, Y_test)
decisiontree_accuracy=tree_cv.score(X_test, Y_test)
knn_accuracy=knn_cv.score(X_test, Y_test)

result['Logistic Regression'] = [logisticregression_accuracy]
result['Support Vector Machine'] = [svm_accuracy]
result['Decision Tree'] = [decisiontree_accuracy]
result['K-Nearest Neighbors'] = [knn_accuracy]

result.transpose()
```

	0
Algorithm Used	Test Accuracy (Score)
Logistic Regression	0.833333
Support Vector Machine	0.833333
Decision Tree	0.833333
K-Nearest Neighbors	0.833333

Confusion Matrix

- The results for the confusion matrix implemented in each of the methods (Logistics regression, SVM, Decision Tree and K-Nearest Neighbors) provided the same results for both correct and incorrect predictions



Conclusions

- OrbitsES-L1, GEO, HEO, SSO, VLEO had the most success rate
- Launch success rate started to increase from 2013 to 2020
- The larger the flight amount at a launch site, the greater the success rate at a launch site.
- All of the methods has the same accuracy

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!