Winning Space Race with Data Science

Daniel Favale Espinosa
11/09/2022

Outline

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

Executive Summary

- **Summary of methodologies**
  - Data Collection
  - Data Wrangling
  - EDA with data visualization
  - EDA with SQL
  - Build an interactive map with Folium
  - Building a Dashboard with Plotly Dash
  - Predictive Analysis
- **Summary of all results**
  - Exploratory data analysis results
  - Interactive analysis.
  - Predictive analysis results

Introduction

- In this analysis we will focus on the study of the first stage of the SpaceX Falcon 9 Rocket in order to obtain conclusions that allow us to make cost projections as well as obtain insight on the implications in the area beyond the economic ones.
- We will work on finding the following solutions:
  - How to predict if the rocket will land successfully.
  - What parameters can determine the success rate of a landing successful.
  - Project and determine the costs of future launches.

Section 1

Methodology

Methodology

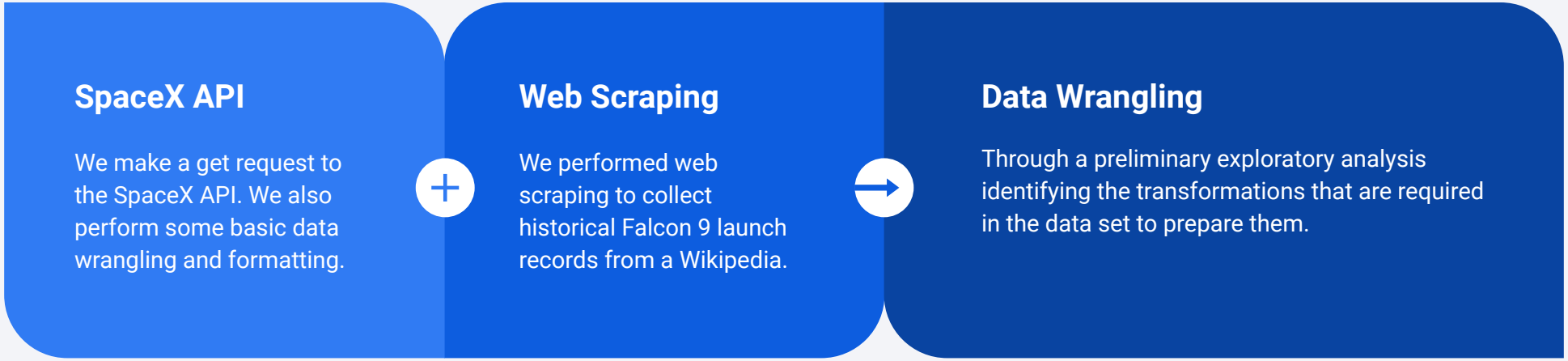
- Data collection methodology:
  - Data was collecting from past SpaceX missions. SpaceX API ***link***
  - Web scraping from Wikipedia ***link***

- Perform data wrangling:
  - Calculated the number of launches on each site
  - Calculated number and occurrences of each orbit
  - Calculated the number and occurrence of mission outcome per orbit type
  - Created a landing outcome label from Outcome column

- Perform exploratory data analysis (EDA) using visualization andSQL
- Perform interactive visual analytics using Folium and Dash.
- Perform predictive analysis using classification models.

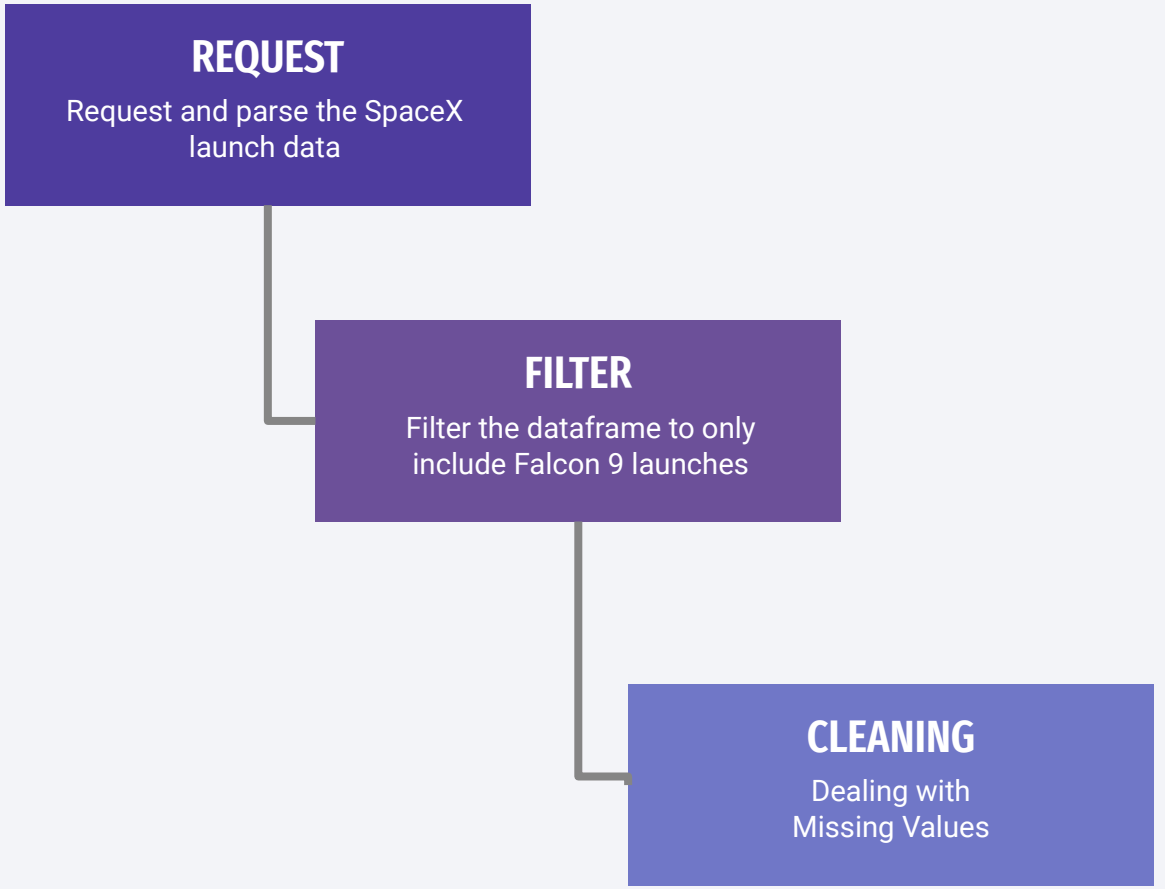
Data Collection

- Data sets were collected from previous SpaceX mission and Wikipedia pages and below processes were obtained to Filter, clean and Transform the data to prepare for Modeling.



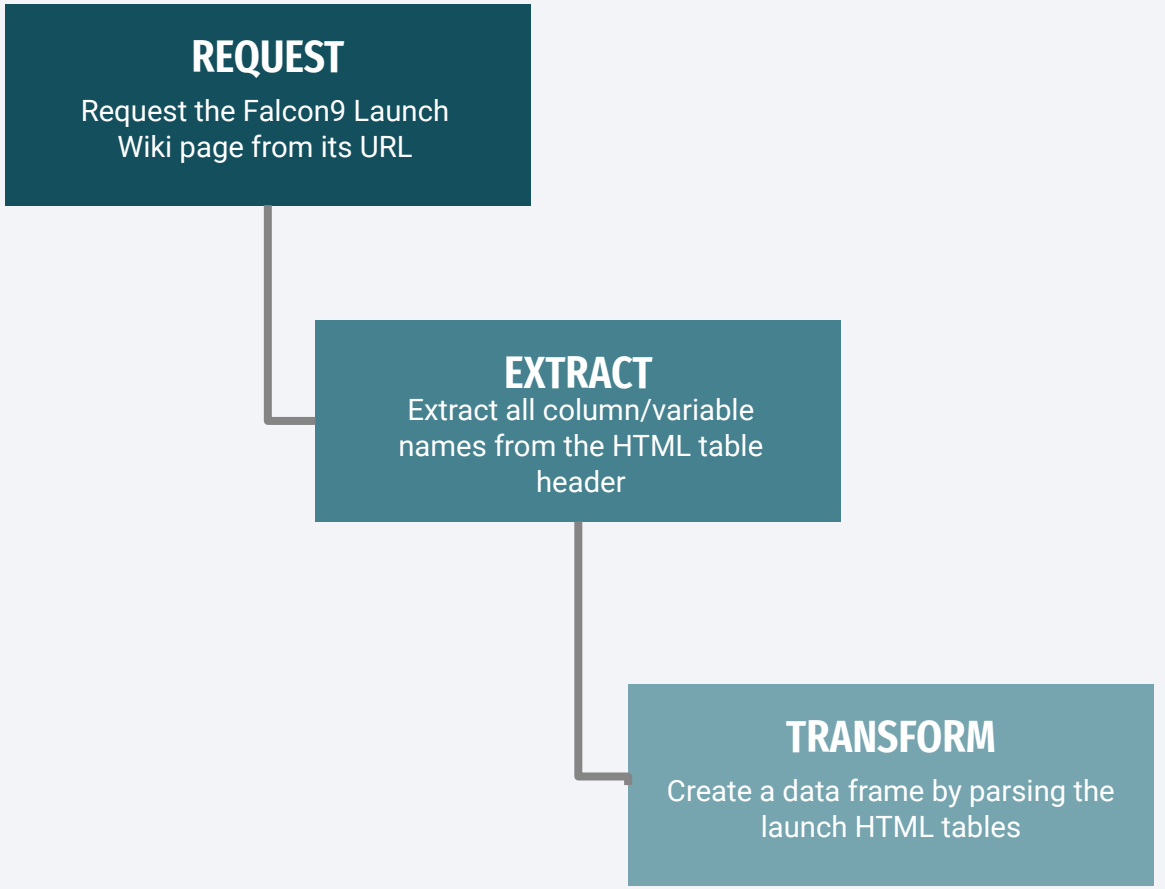
Data Collection – SpaceX API

- We make a get request to the SpaceX API. We also perform some basic data wrangling and formatting.
- The procedure is summarized in *Flowchart Nro.2*.
- *It can be seen in detail in the following* ***link de GitHub*****.**



Data Collection - Web Scraping

- We performed web scraping to collect historical Falcon 9 launch records from a Wikipedia page titled "*List of Falcon 9 and Falcon Heavy launches".*
- The procedure is summarized in *Flowchart Nro.3*.
- *It can be seen in detail in the following* ***link de GitHub.***



Data Wrangling

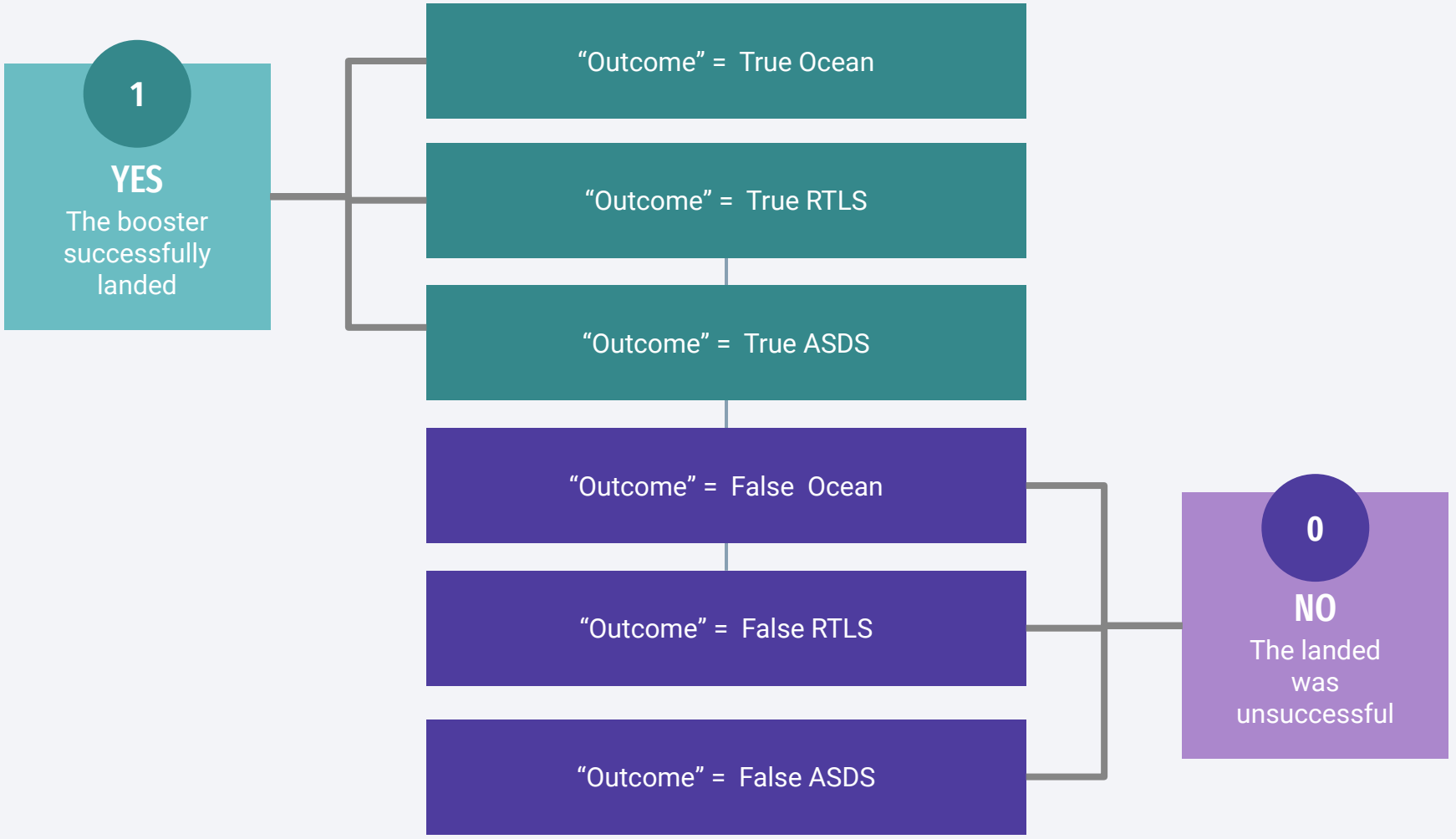
- Through a preliminary exploratory analysis identifying the transformations that are required in the data set to prepare them.

- We will process the landing data into valid tags for training the predictive models later.
    - Training tags with "1" will mean the rocket landed successfully, and "0" means it was unsuccessful.

- *See in detail in GitHub **link.***

See flowchart No.4 on the next slide…

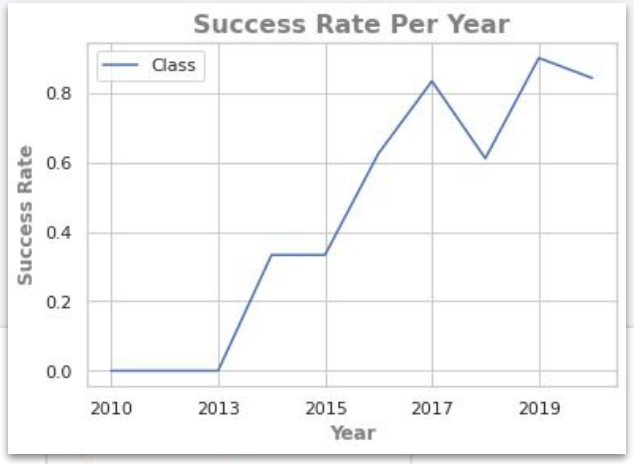
Data Wrangling - Convert outcomes into Training Labels

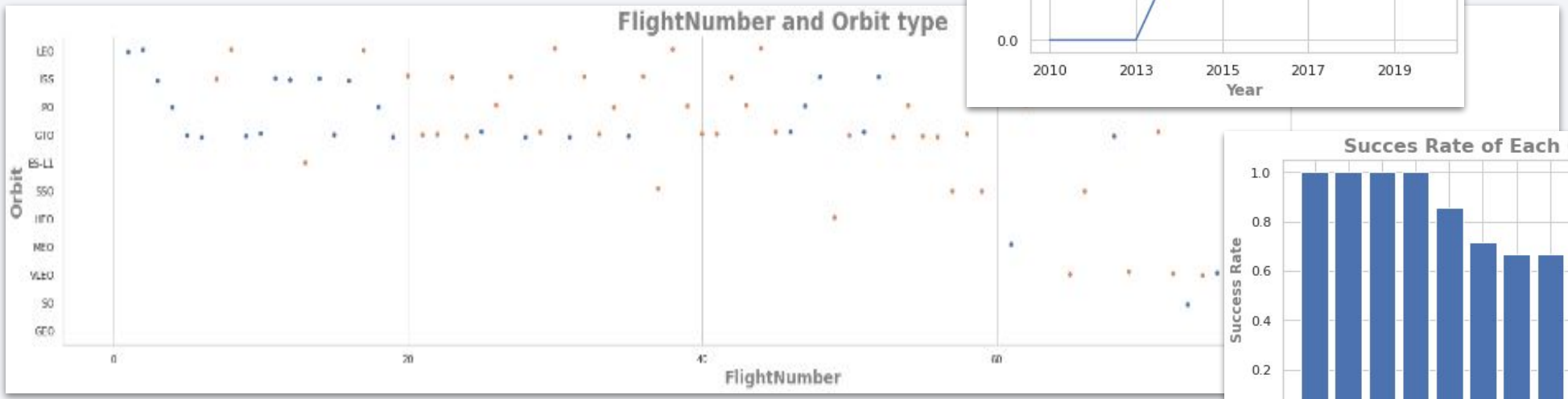


Flowchart 4

EDA with Data Visualization

- Exploratory Data Analysis to visualize the relationship between:
  - Flight Number and Launch Site.
  - Payload and Launch Site.
  - Success rate of each orbit type.
  - Flight Number and Orbit type.
  - Payload and Orbit type.
  - Visualize the launch success yearly trend.





- *See in detail on GitHub **link.***

EDA with SQL

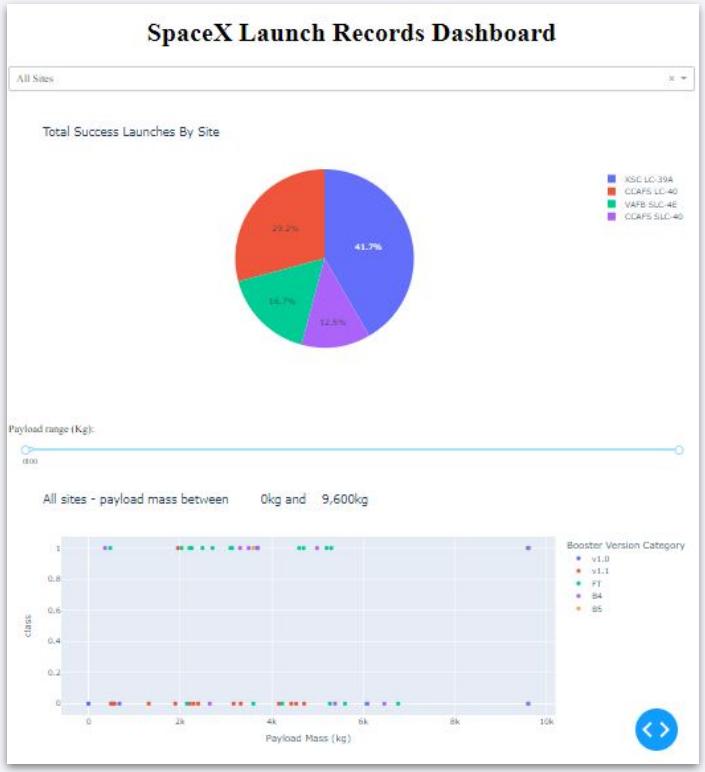
- SQL queries performed:
    - Names of the unique launch sites in the space mission
    - Top 5 launch sites whose name begin with the string 'CCA’
    - Total payload mass carried by boosters launched by NASA (CRS)
    - Average payload mass carried by booster version F9 v1.1
    - Date when the first 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 kg
    - 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, their booster versions, and launch site names for 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
- *Ver en detalle en GitHub **aquí.***

Build an Interactive Map with Folium

- Markers, circles, lines and marker clusters were used with Folium Maps.
- Indications of each element:
  - Markers indicate points like launch sites
  - Circles indicate highlighted areas around specific coordinates, like NASA Johnson Space Center
  - Marker clusters indicates groups of events in each coordinate, like launches in a launch site
  - Lines are used to indicate distances between two coordinates.

- *See in detail on GitHub* ***link.***

Build a Dashboard with Plotly Dash



Link to Dashboard. ***Link to code.***

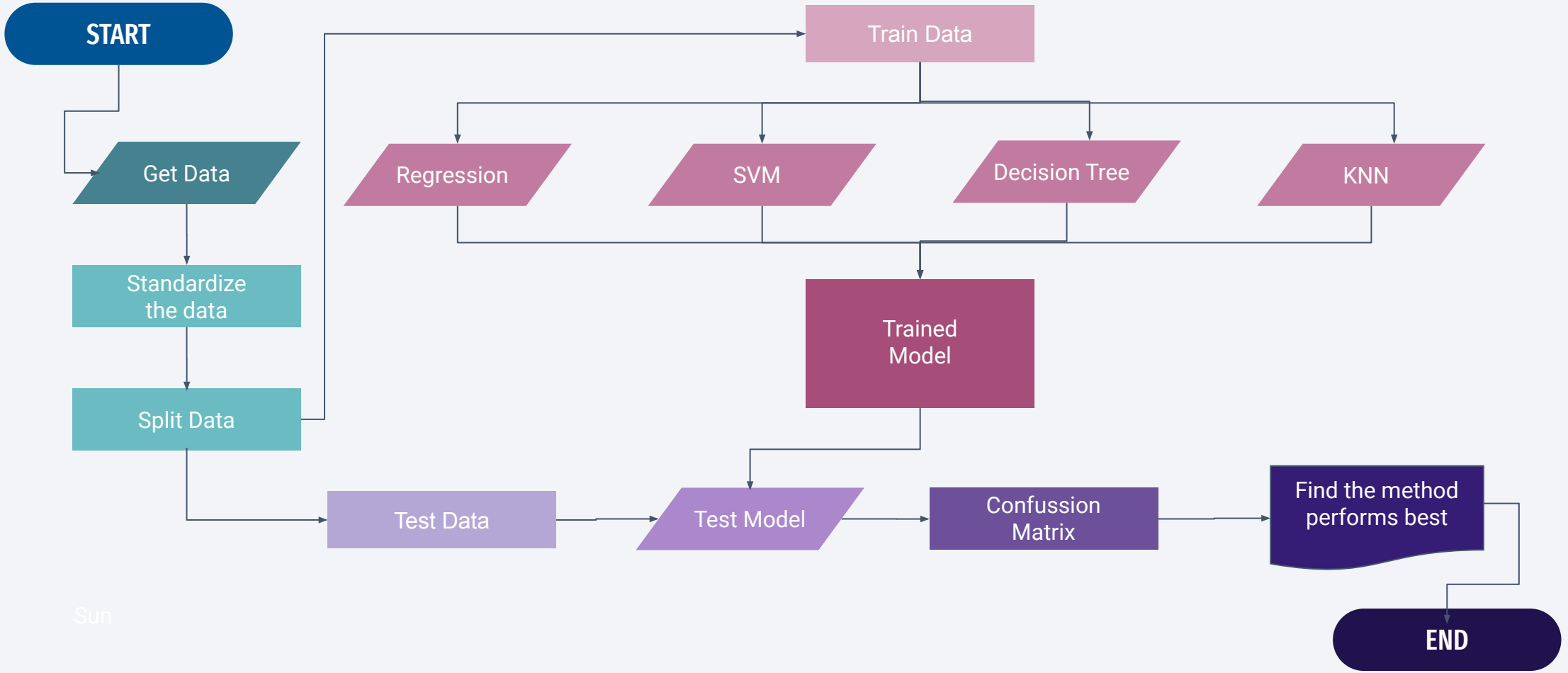
- Elements
  - Dropdown list for the launch site.
  - RangeSlider for selecting the payload mass.
  - PieChart: for showing the success rate of each launch site, or showing the number of successful landing outcomess.
  - Scatterplot: Show success/failure by payload and booster version.
- Findings:
  - Which site has the largest successful launches? KSC LC-39A.
  - Which site has the highest launch success rate? KSC LC-39A (success rate 76.9%).
  - Which payload range(s) has the highest launch success rate? 2000-4000.
  - Which payload range(s) has the lowest launch success rate? 6000-8000.
  - Which F9 Booster version (v1.0, v1.1, FT, B4, B5, etc.) has the highest.
  - launch success rate? B5 (only one successful start), apart from that FT (15 successes, 8 failures).

Predictive Analysis (Classification)

- We create a machine learning pipeline to predict if the first stage will land given the data from the preceding labs.

- Perform exploratory Data Analysis and determine Training Labels
  - Create a column for the class
  - Standardize the data
  - Split into training data and test data
- Find best Hyperparameter for SVM, Classification Trees and Logistic Regression
  - Find the method performs best using test data.

Predictive Analysis (model flowchart)



Sun

Results

Exploratory data analysis results

- Launch success rate increases over time
- Higher success rate for higher orbits

Interactive analytics demo in screenshots

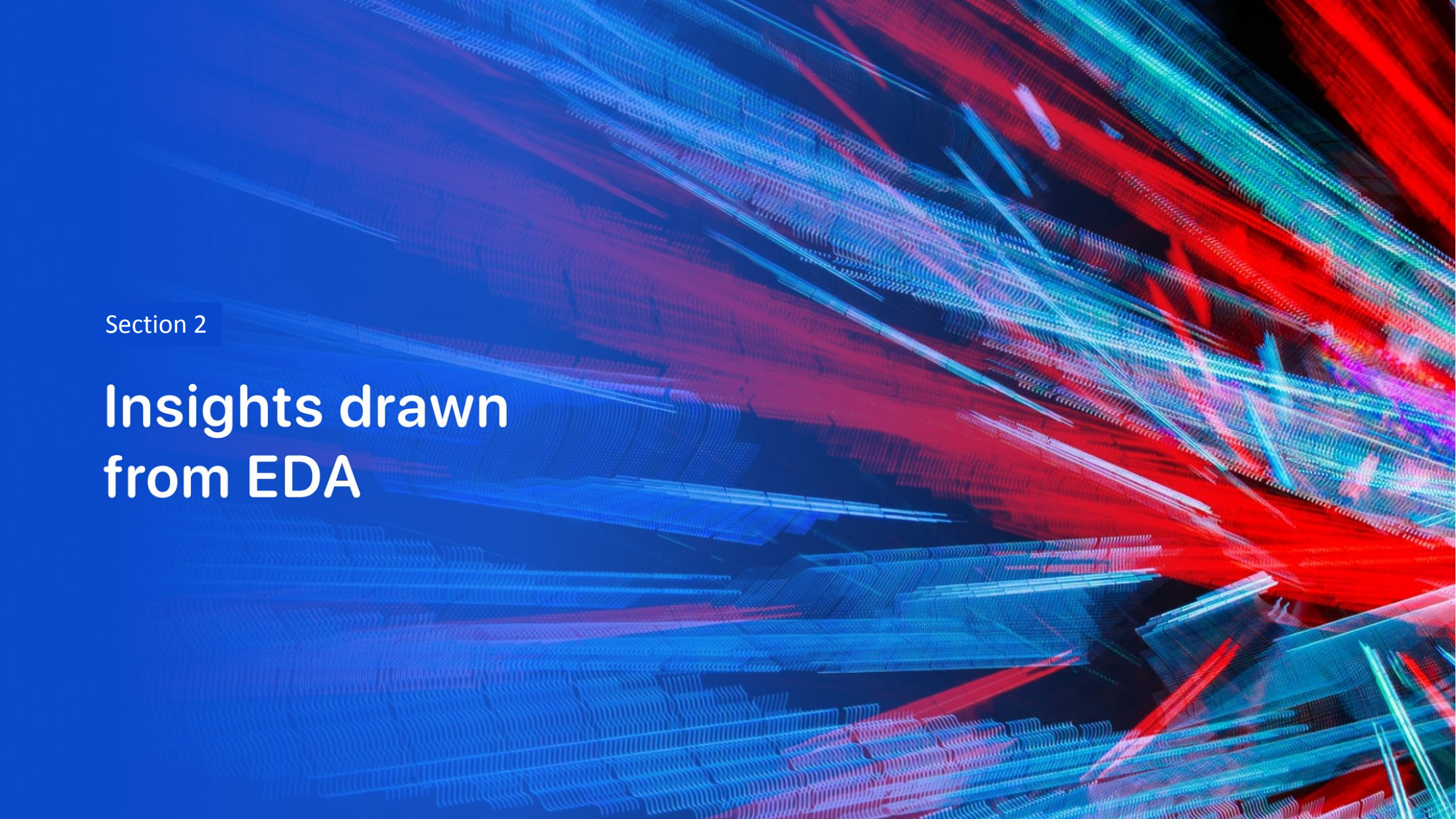
- Higher success rate for higher payload mass
- Low success rate for booster versions v1.0, v1.1, high success rate for FT, B4, B5
- Higher success rate for Kennedy Space center and recent starts at Cape Canaveral

Predictive analysis results

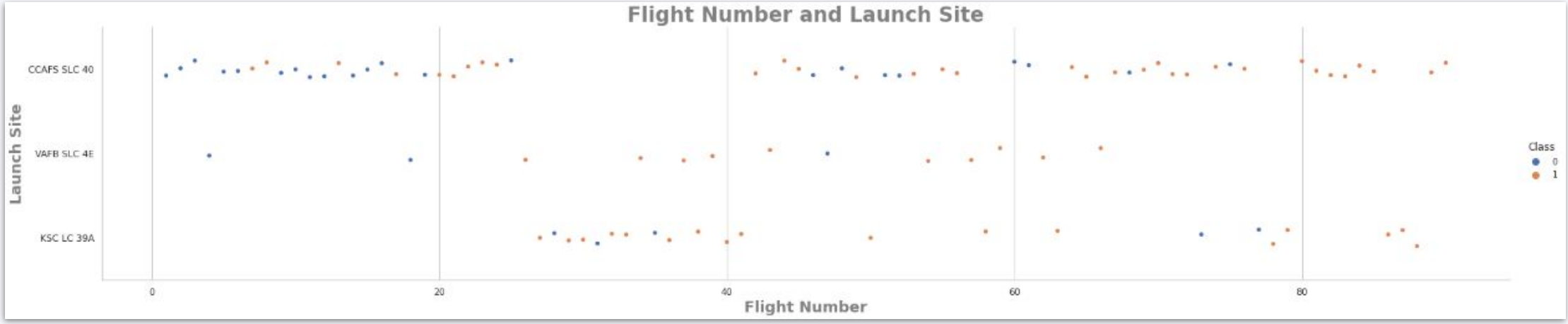
- Best prediction results with Logistic Regression and Support Vector Machine

Section 2

Insights drawn from EDA



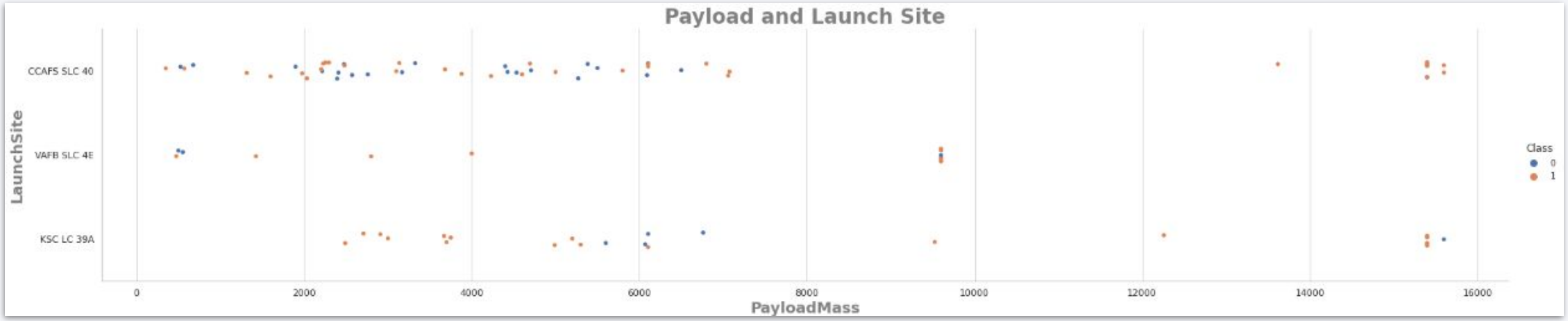
Flight Number vs. Launch Site



Explanations:

- We can see that the CCAFS LC-40 launch site has more attempts than KSC LC-39A and VAFB SLC 4E.We can see that the CCAFS LC-40 launch site has more attempts than KSC LC-39A and VAFB SLC 4E.

Payload vs. Launch Site



Explanations:

- Now if you observe Payload Vs. Launch Site scatter point chart you will find for the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000).

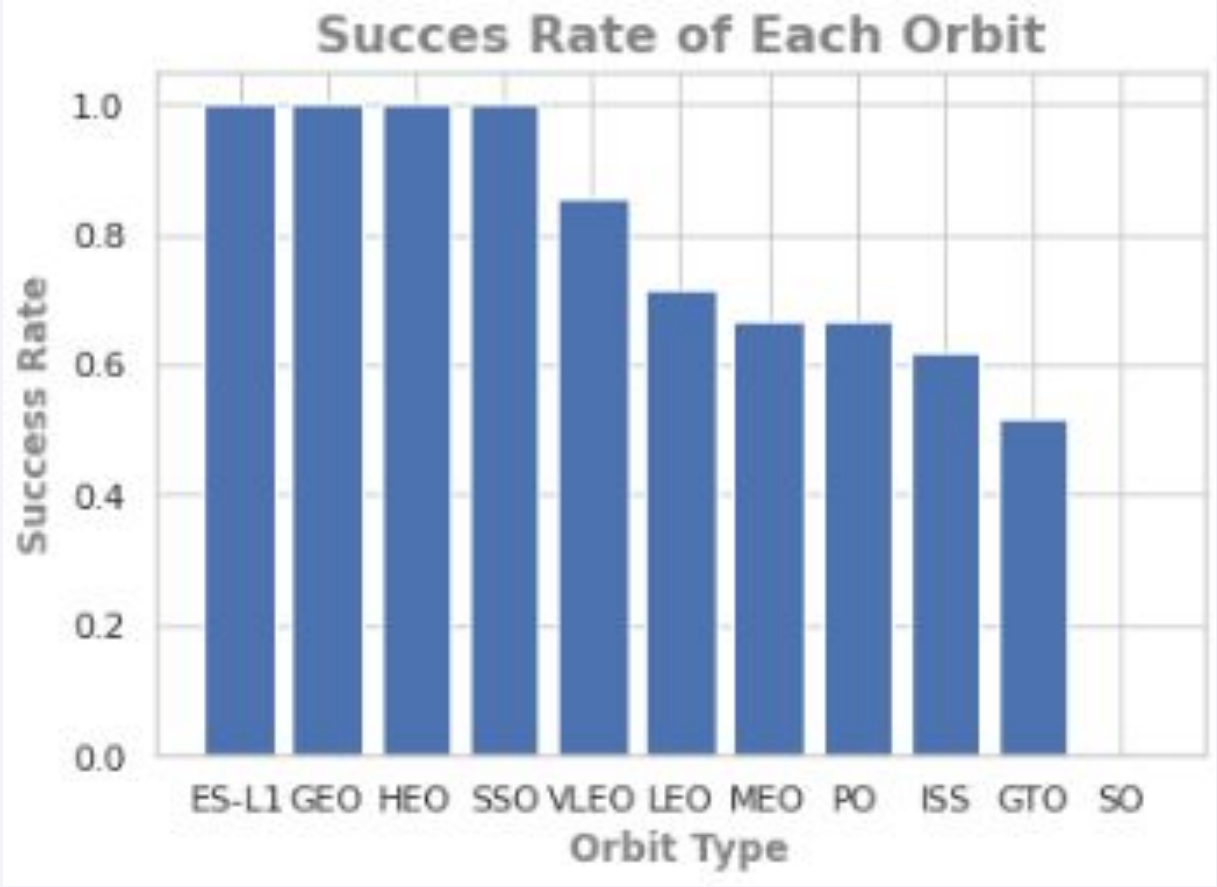
Success Rate vs. Orbit Type

Low Earth Orbits

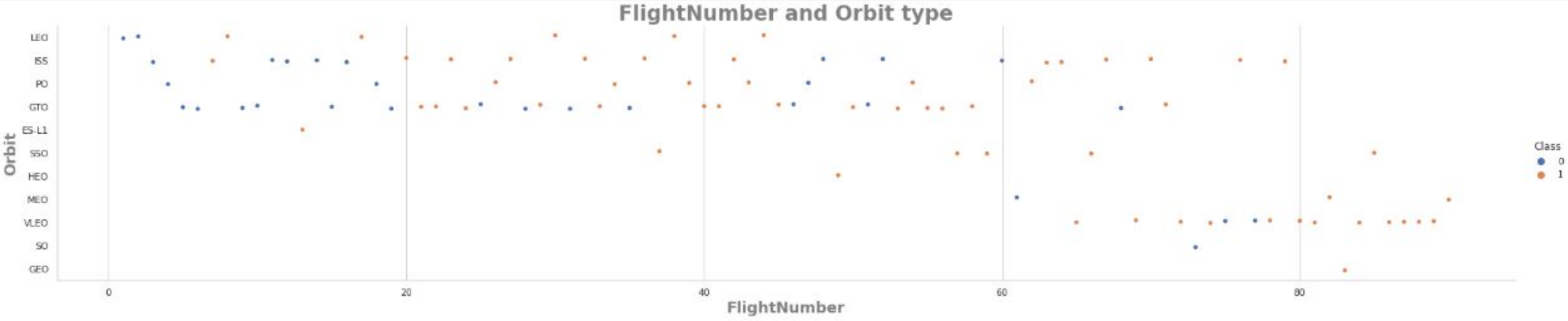
- GTO; ISS; LEO; MEO; PO; VLEO

High Earth Orbits

- ES-L1; GEO; HEO;SSO



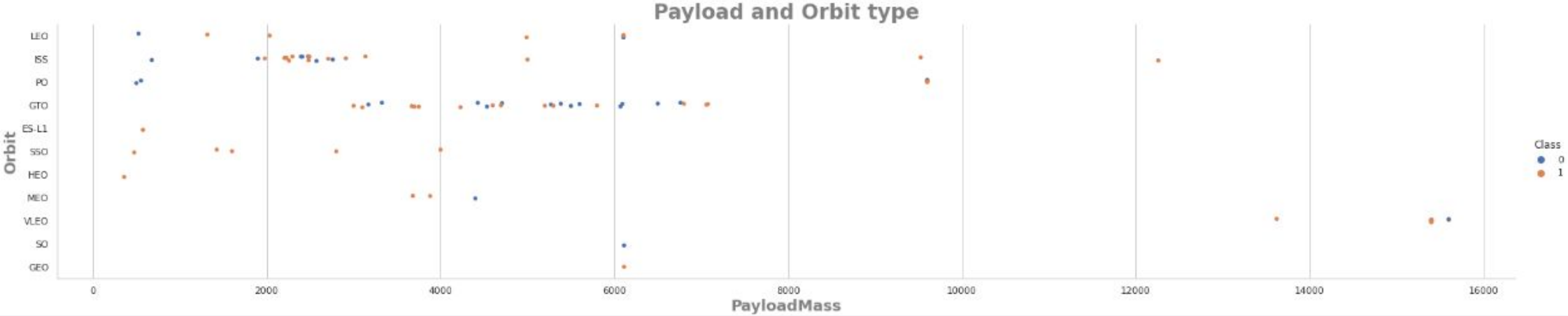
Flight Number vs. Orbit Type



Explanations:

- You should see that in the LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.

Payload vs. Orbit Type

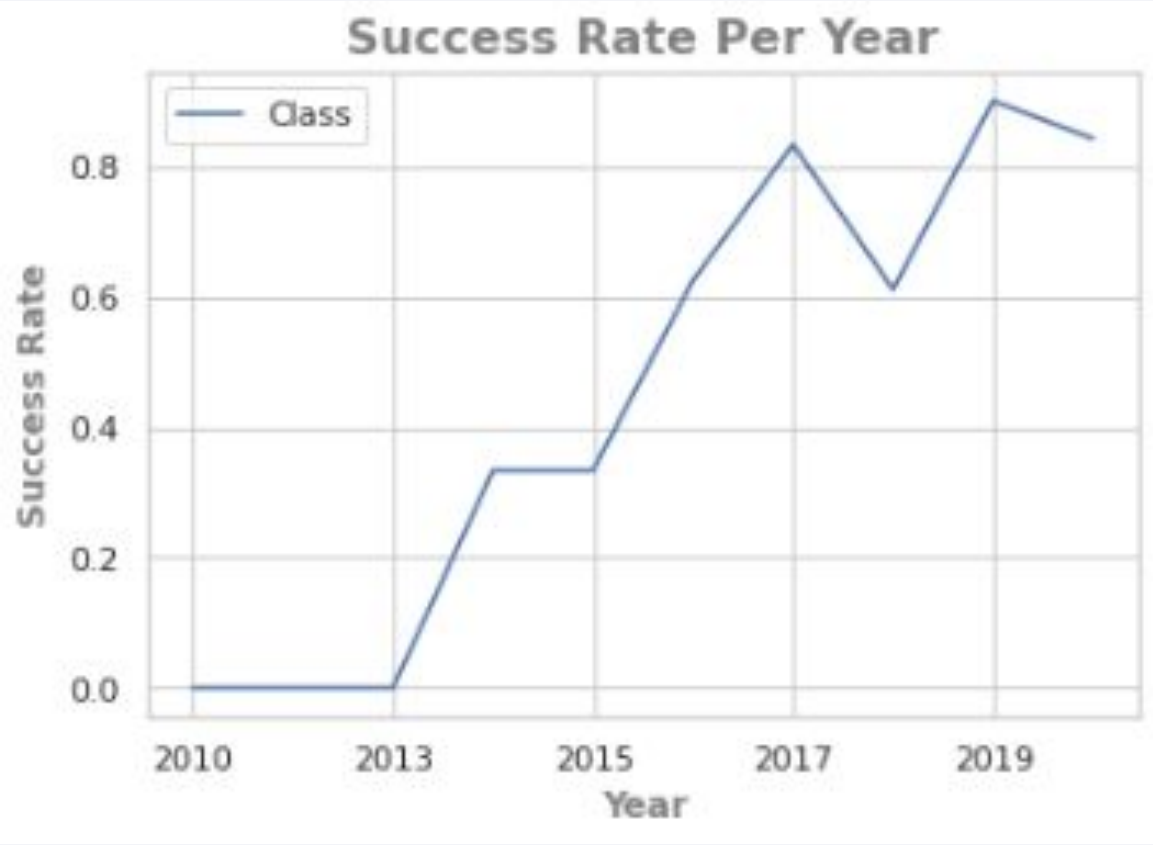


Explanations:

- With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS. However for GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both there here.

Launch Success Yearly Trend

- you can observe that the success rate since 2013 kept increasing till 2020



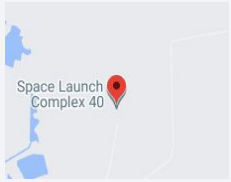
All Launch Site Names

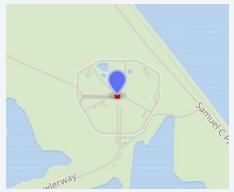
- Query:
  -
```
%sql SELECT DISTINCT LAUNCH_SITE FROM SPACEXDATASET;
```

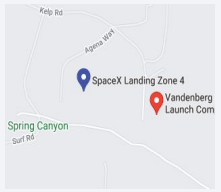
CCAFS Cape Canaveral Space Launch Complex

KSC Kennedy Space Center Launch Complex

VAFB Vandenberg Space Launch Complex



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

Launch Site Names Begin with 'CCA'

- Query:
  -
```
%sql SELECT * from SPACEXDATASET 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	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	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

- Query:
  -
```
%sql SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXDATASET WHERE Customer = 'NASA (CRS)';
```

1
45596

Average Payload Mass by F9 v1.1

- Query:
  -
```
%sql SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXDATASET WHERE Booster_Version LIKE 'F9 v1.0%';
```

1
340

First Successful Ground Landing Date

- Query:
  -
```
%sql SELECT MIN(Date) FROM SPACEXDATASET WHERE Landing__Outcome = 'Success (ground pad)';
```

1
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- Query:
  -
```
%sql SELECT DISTINCT BOOSTER_VERSION FROM SPACEXDATASET WHERE LANDING__OUTCOME = 'Success (drone ship)' AND 4000 < PAYLOAD_MASS__KG_ < 6000;
```

BOOSTER_VERSION
F9 FT B1021.2
F9 FT B1031.2
F9 FT B1022
F9 FT B1026

Total Number of Successful and Failure Mission Outcomes

- Query:
  -
```
%sql SELECT COUNT(MISSION_OUTCOME) FROM SPACEXDATASET;
```

1
101

Boosters Carried Maximum Payload

- Query:
  - ```
    SELECT DISTINCT BOOSTER_VERSION FROM SPACEXDATASET WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_)FROM SPACEXDATASET);
    ```

BOOSTER_VERSION
F9 B5 B1048.4
F9 B5 B1048.5
F9 B5 B1049.4
F9 B5 B1049.5
F9 B5 B1049.7
    ```

2015 Launch Records

- Query:
  -

```
%sql SELECT LANDING__OUTCOME, BOOSTER_VERSION, LAUNCH_SITE FROM SPACEXDATASET WHERE Landing__Outcome = 'Failure (drone ship)' AND YEAR(DATE) = 2015;
```

LANDING__OUTCOME	BOOSTER_VERSION	LAUNCH_SITE
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

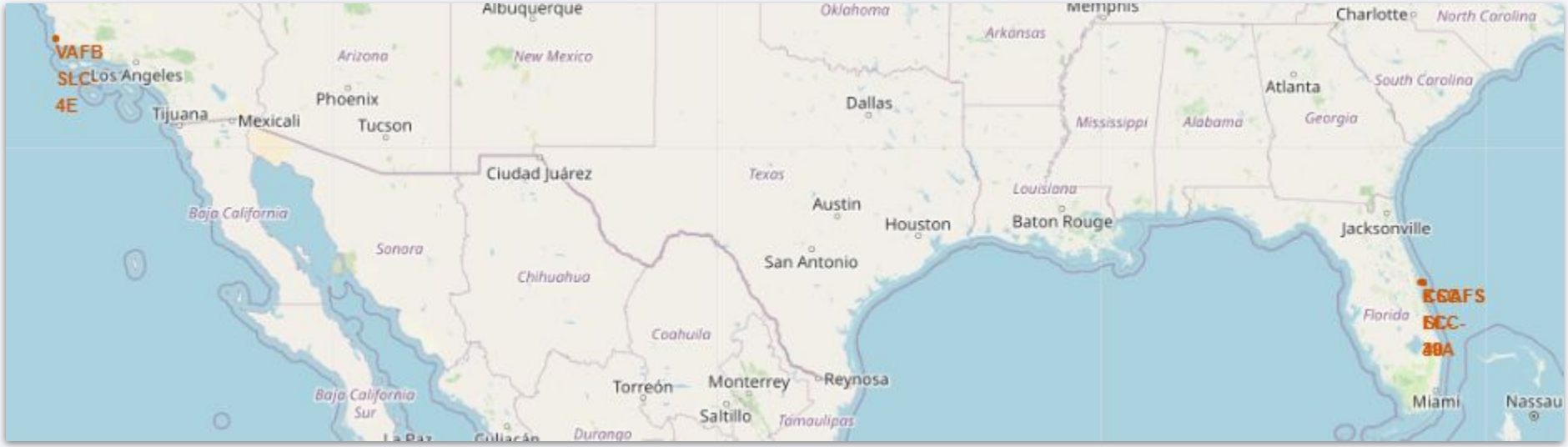
- Query:

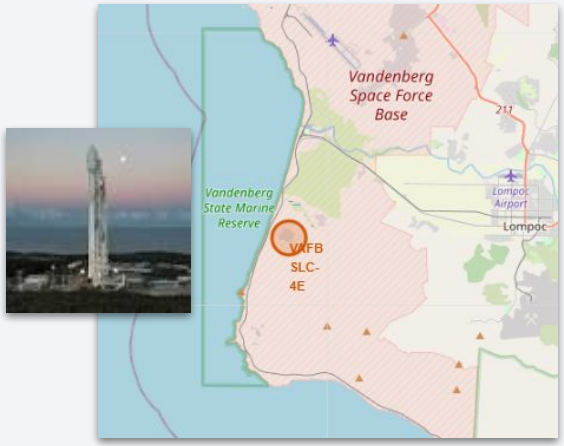
```
%%sql
SELECT LANDING__OUTCOME, COUNT(LANDING__OUTCOME) AS TOTAL_NUMBER FROM SPACEXDATASET
WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY LANDING__OUTCOME ORDER BY TOTAL_NUMBER DESC
```

LANDING__OUTCOME	TOTAL_NUMBER
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3


Section 3
Launch Sites
Proximities Analysis

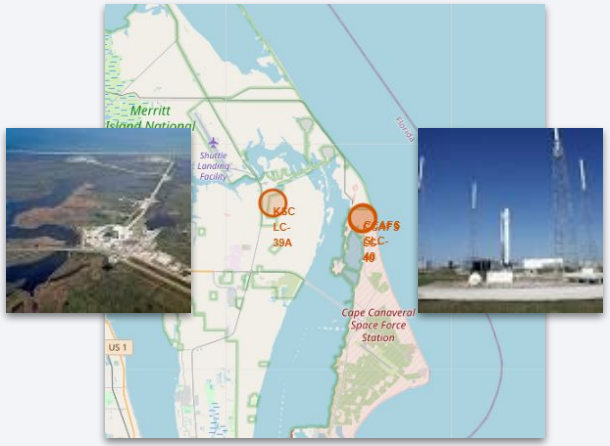
Folium Map: Launch Sites





- Launch sites are at the East and West coast, near the southernmost U.S.mainland area, ;which is Florida and; California

CCAFS Cape Canaveral Space Launch Complex
KSC Kennedy Space Center Launch Complex
VAFB Vandenberg Space Launch Complex

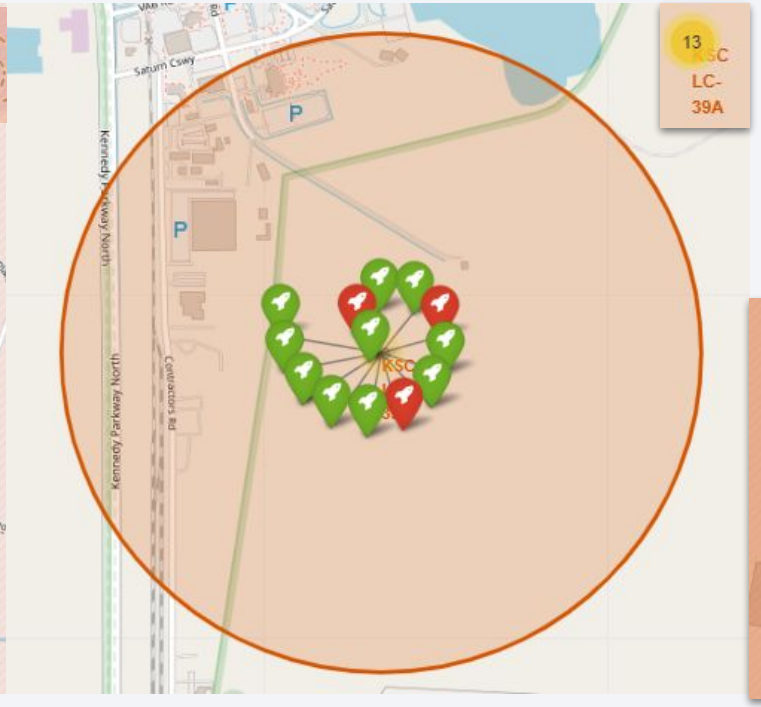


Folium Map: Stage-1 Landing Success by Launch Site

Vandenberg Space Launch Complex

VAFB SLC-4E
40.00% Success

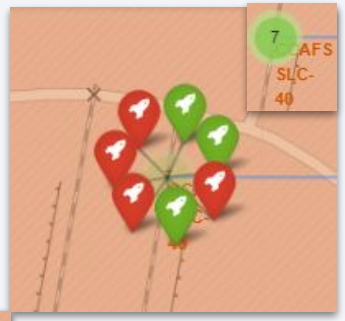
Kennedy Space Center Launch Complex

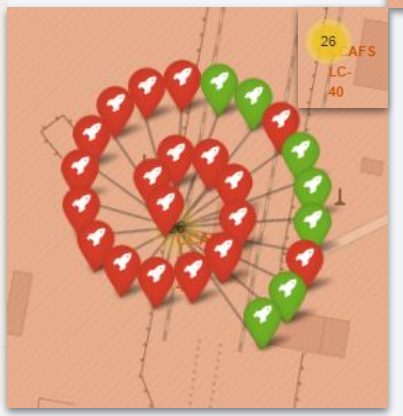


KSC LC-39A
76.92% Success

Cape Canaveral Space Launch Complex

CCAFS SLC-40
42.85% Success

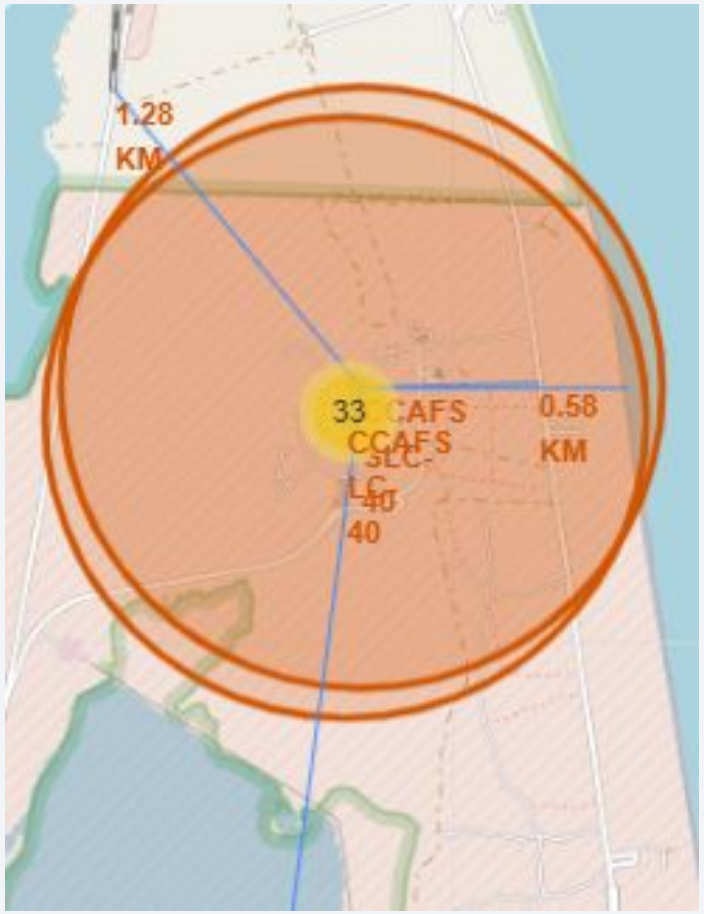


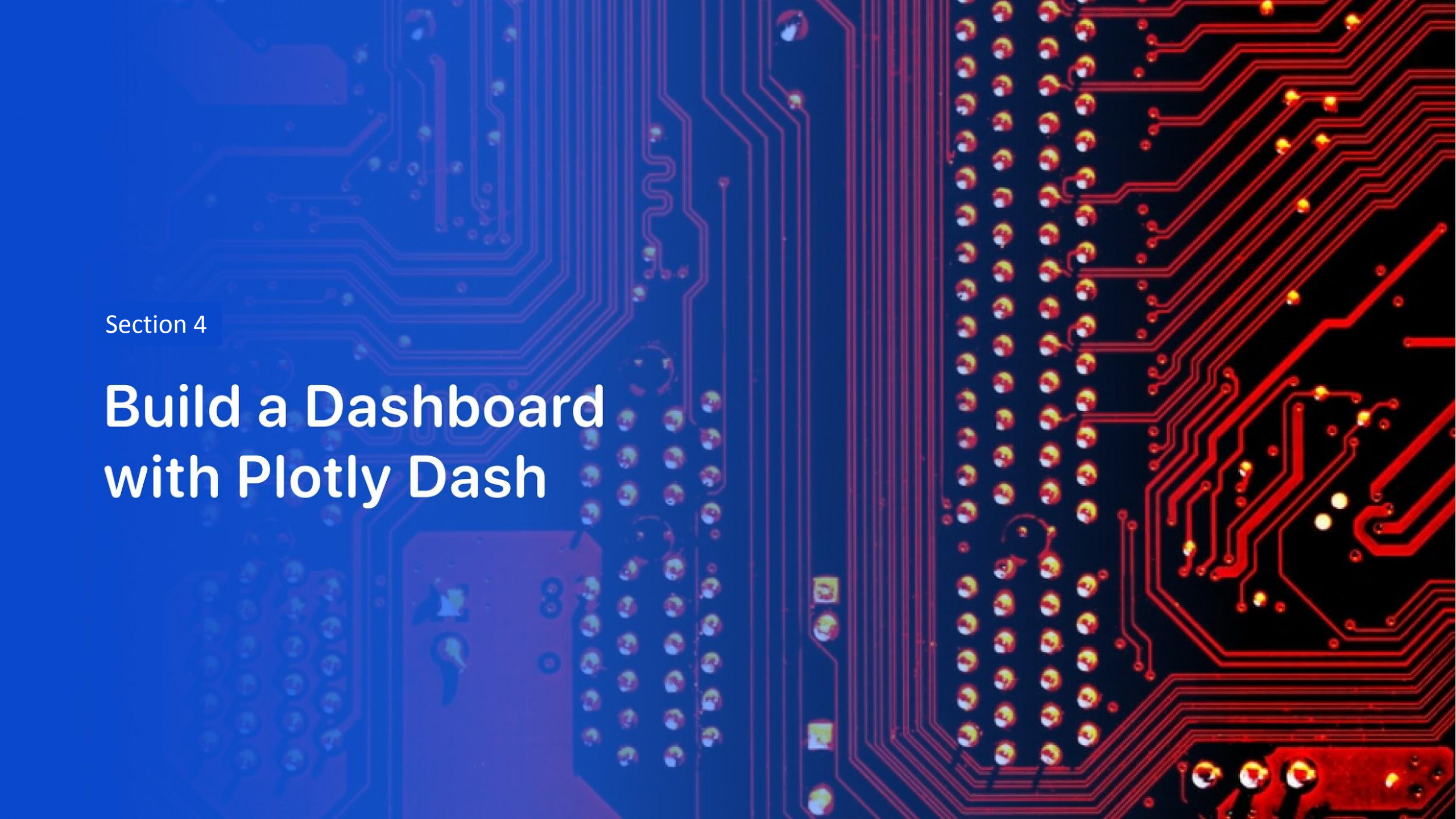


CCAFS LC-40
26.92% Success

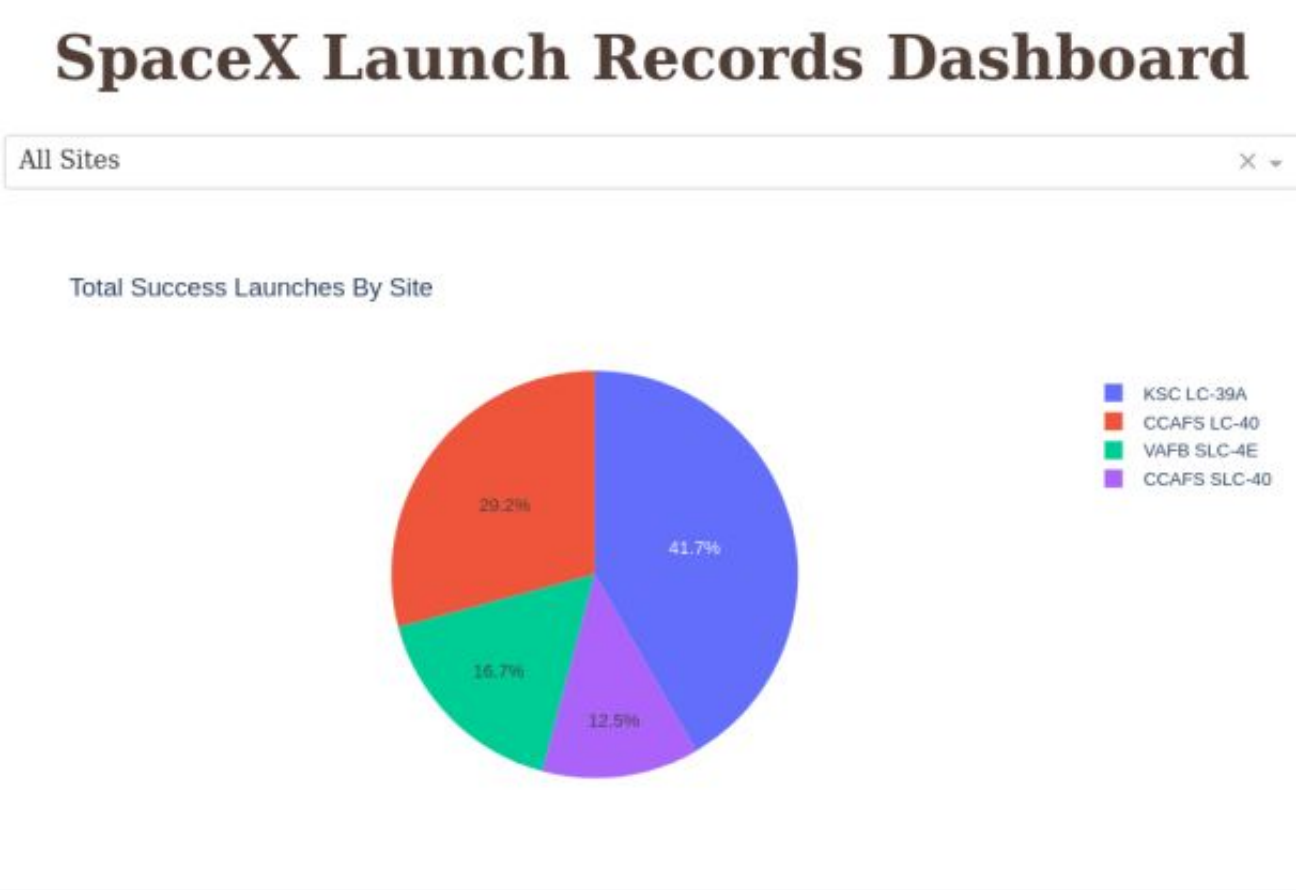
Logistics and Safety

- Launch site KSC LC-39A has good logistics aspects, being near railroad and road and
- relatively far from inhabited areas.



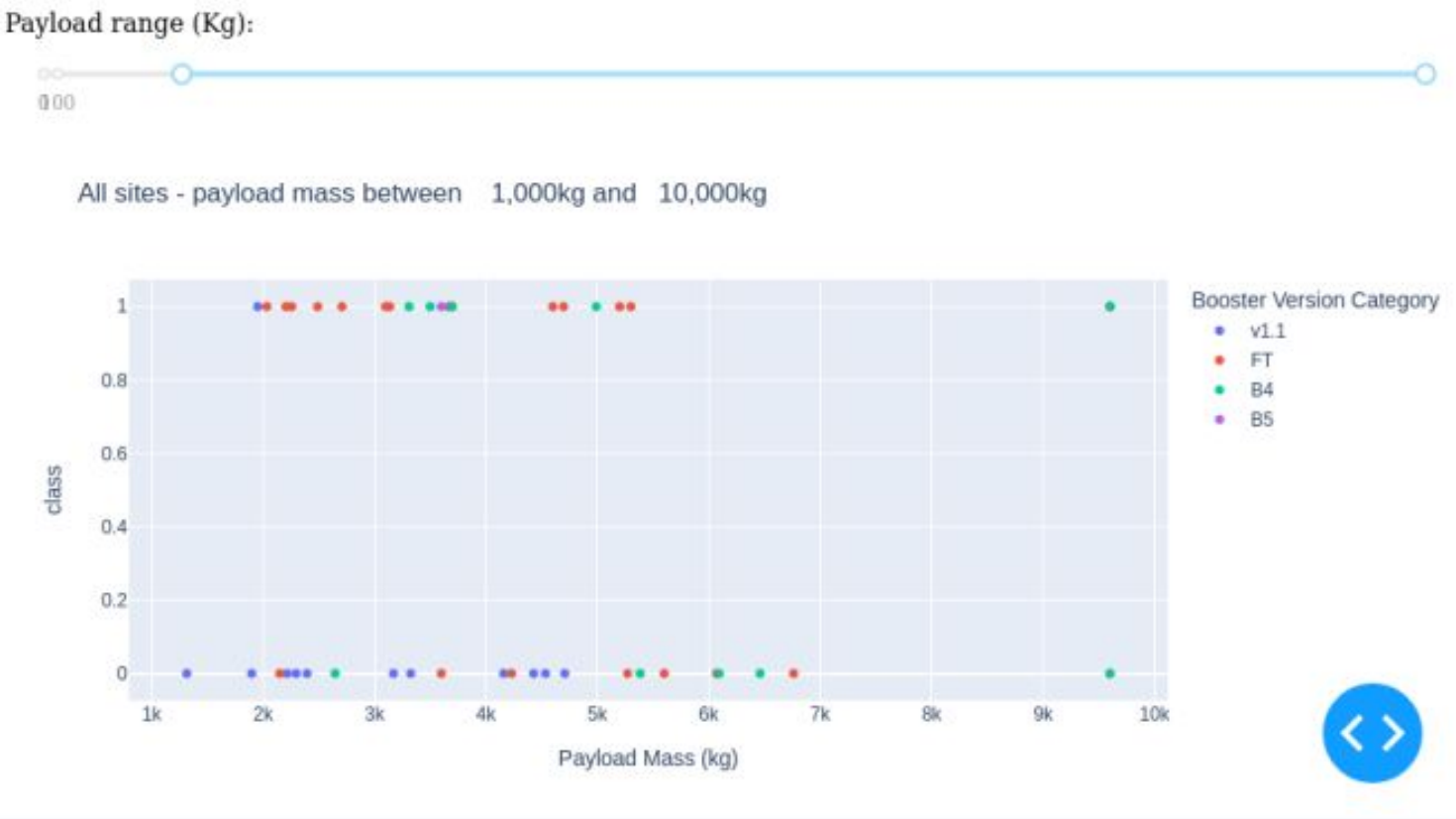

Section 4
Build a Dashboard with Plotly Dash

Dashboard: Launch Success Count For All Sites



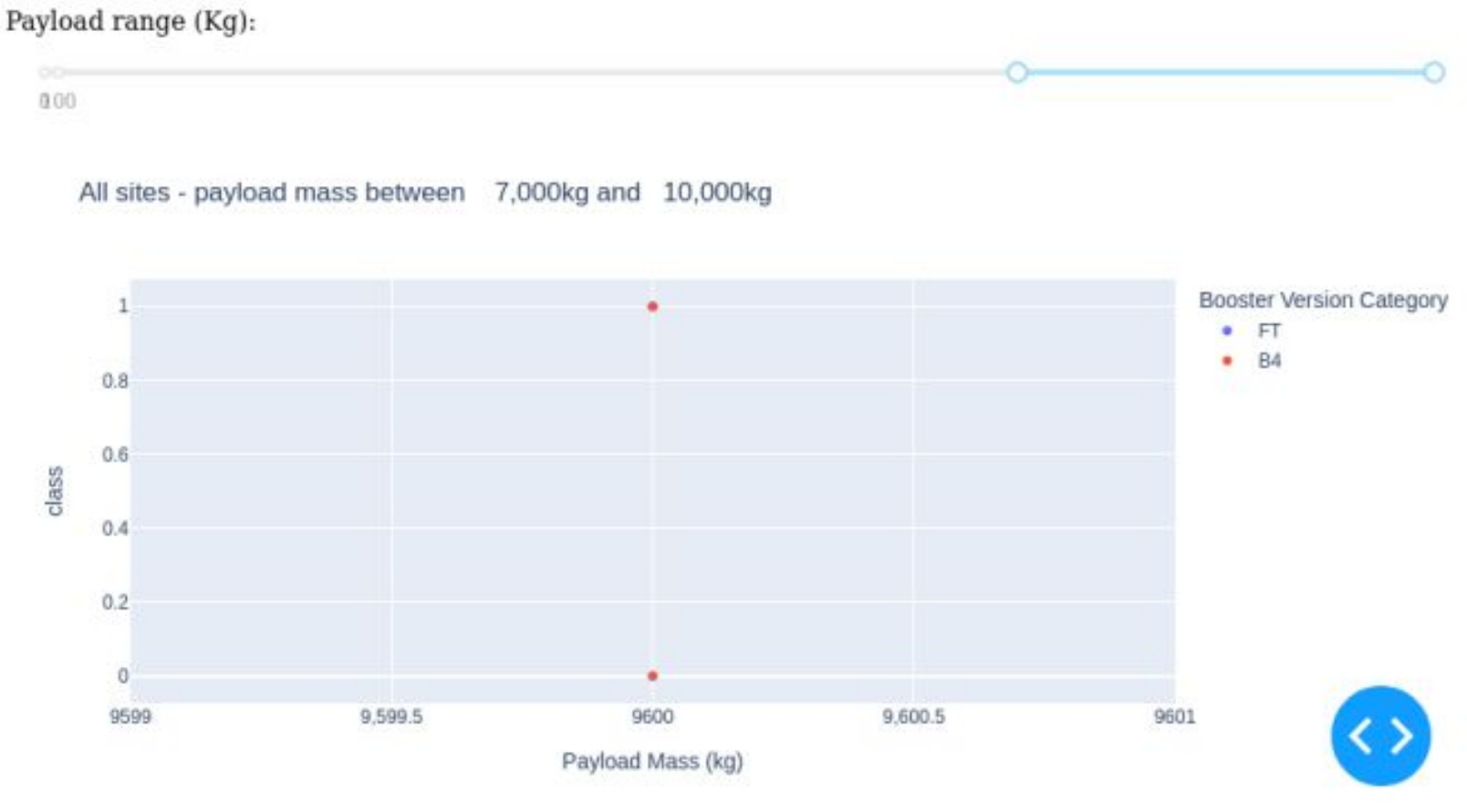
- Kennedy Space Center (KSC LC-39A) hast the most successful stage-1 landings
- Vandenberg Air Force Base (VAFB SLC-4E) has the least number of successful stage-1 landings

Payload vs. Launch Outcome



- Payloads under 6,000kg and FT boosters are the most successful combination.

Payload vs. Launch Outcome

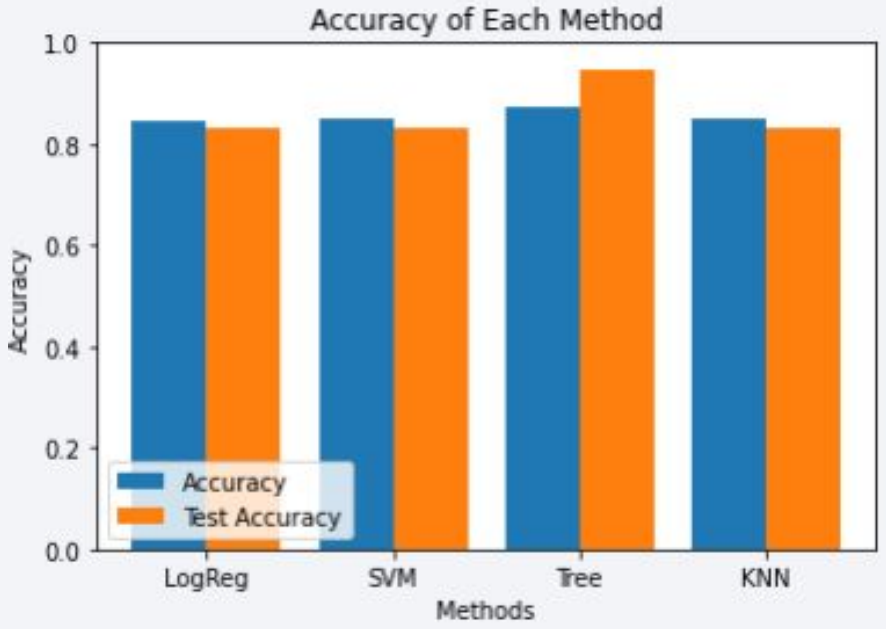


- There's not enough data to estimate risk of launches over 7,000kg

Section 5

Predictive Analysis (Classification)

Classification Accuracy

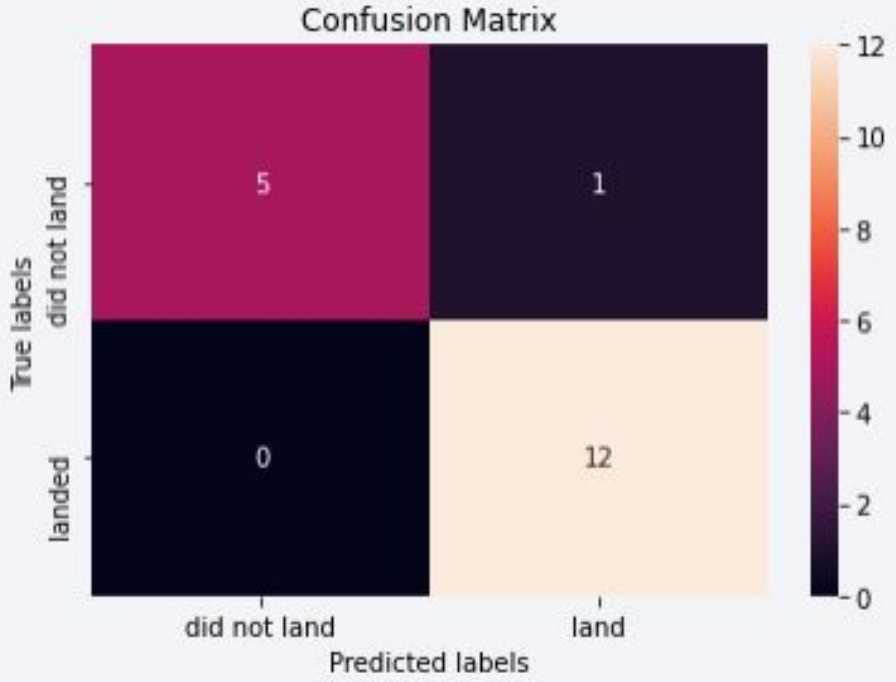


- Four classification models were tested, and their accuracies are plotted beside
- The model with the highest classification accuracy is Decision Tree Classifier, which has accuracies over than 87%.

Confusion Matrix

Confusion matrix of Decision Tree Classifier proves its accuracy by showing the big numbers of true positive and true negative compared to the false ones.

True Positives	12
True Negatives	5
False Positives	1
False Negatives	0



Conclusions

- The best launch site is KSC LC-39A
- Launches above 7,000kg are less risky
- Although most of mission outcomes are successful, successful landing outcomes seem to improve over time, according the evolution of processes and rockets
- None of the models had false negatives
- All models had at least one false positive
- Prediction with Logistic Regression is quite accurate
- Support Vector Machine also provide a good result for predicting the landing outcome

Appendix

- You can see all the references and details of the project at this link.

Thank you!