PROJECT

WINNING SPACE RACE

WITH DATA SCIENCE

DATE 04.08.2023

AUTHOR DANIEL KETHER

Outline

1. Executive Summary
2. Introduction
3. Methodology
4. Results
5. Conclusion

Executive Summary

Summary of Methodologies:

- Collected data using SpaceX API and web-scraping Wiki
- Wrangled data by filtering it, handling missing values and applying one hot encoding
- Explored data with SQL and data visualization
- Visualized data using Folium and Plotly Dash
- Built classification models, tuned and evaluated them to find the best model and it's parameters

Summary of All Results:

- To maximize mission success rate:
  launch from CCAFS SLC-40 to SSO orbit with payload mass lower than 5000 kg.
- To maximize prediction of launch outcome:
  use decision tree classifier.
- To maximize model performance:
  consider larger dataset and test other models.

Introduction

Context:

SpaceX is a leader in private sector of space industry. With years of experience and loads of gathered data. It’s novel technology makes it possible to reuse the first stage of Falcon 9 rocket. By predicting if the first stage will land successfully any competing startup can «make more informed bids» against SpaceX for a rocket launch.

Problems:

- What affects successful mission
- How to maximize mission success
- How to predict mission outcome

METHODOLOGY

SECTION №1

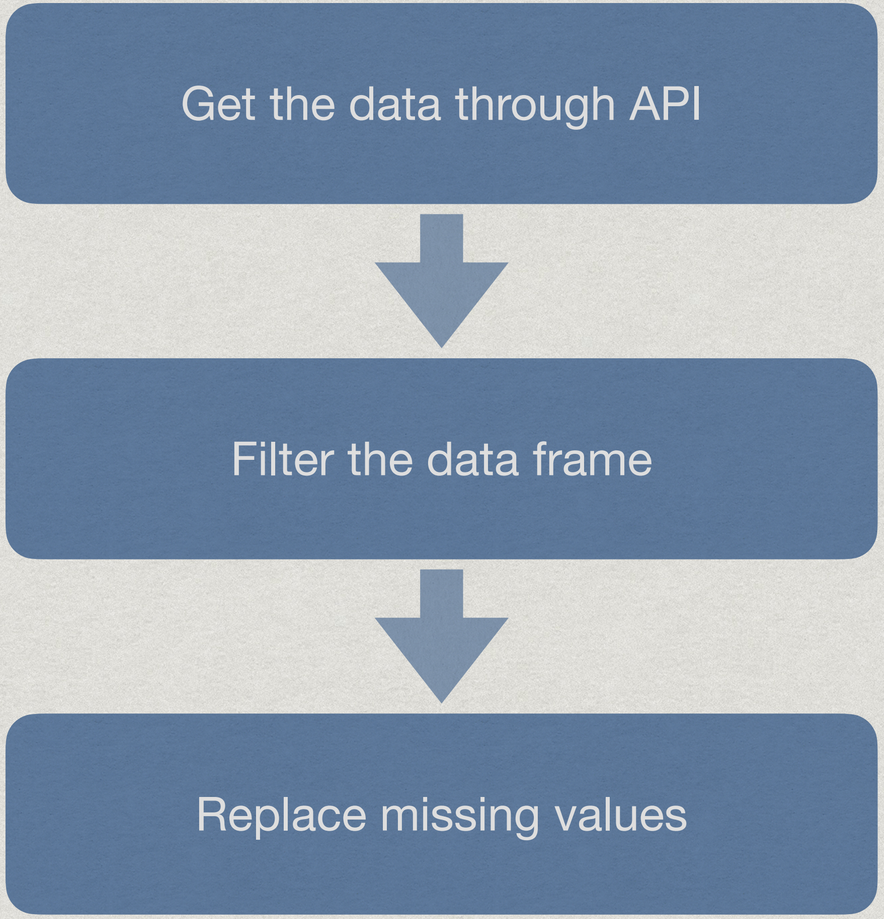
Methodology

1. Collected data using SpaceX API and web-scraping Wiki
2. Wrangled data by filtering it, handling missing values and applying one hot encoding
3. Explored data with SQL and data visualization
4. Visualized data using Folium and Plotly Dash
5. Built classification models, tuned and evaluated them to find the best model and it’s parameters

Data Collection: SpaceX API

1. Requested and parsed the SpaceX launch data using the GET request
2. Filtered the data frame to only include `Falcon 9` launches
3. Dealt with Missing Values

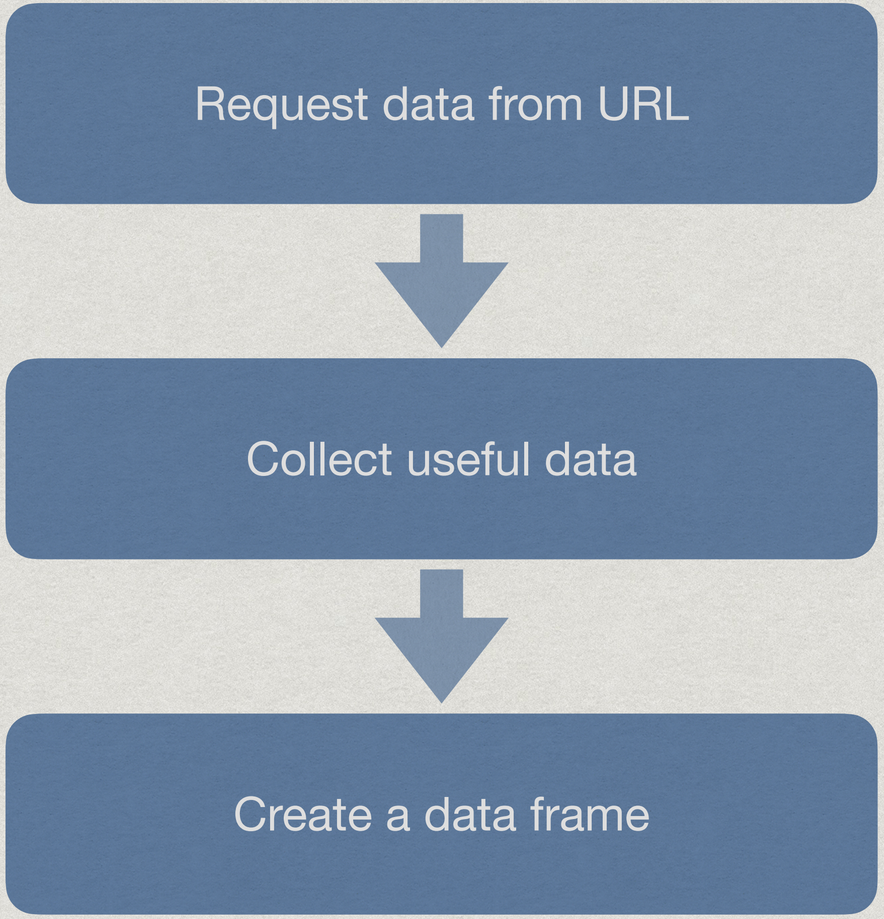
GitHub Link (click)



Data Collection: Web-Scraping

1. Requested the Falcon9 Launch Wiki page from it's URL
2. Extracted all column names from the HTML table header
3. Created a data frame by parsing the launch HTML tables

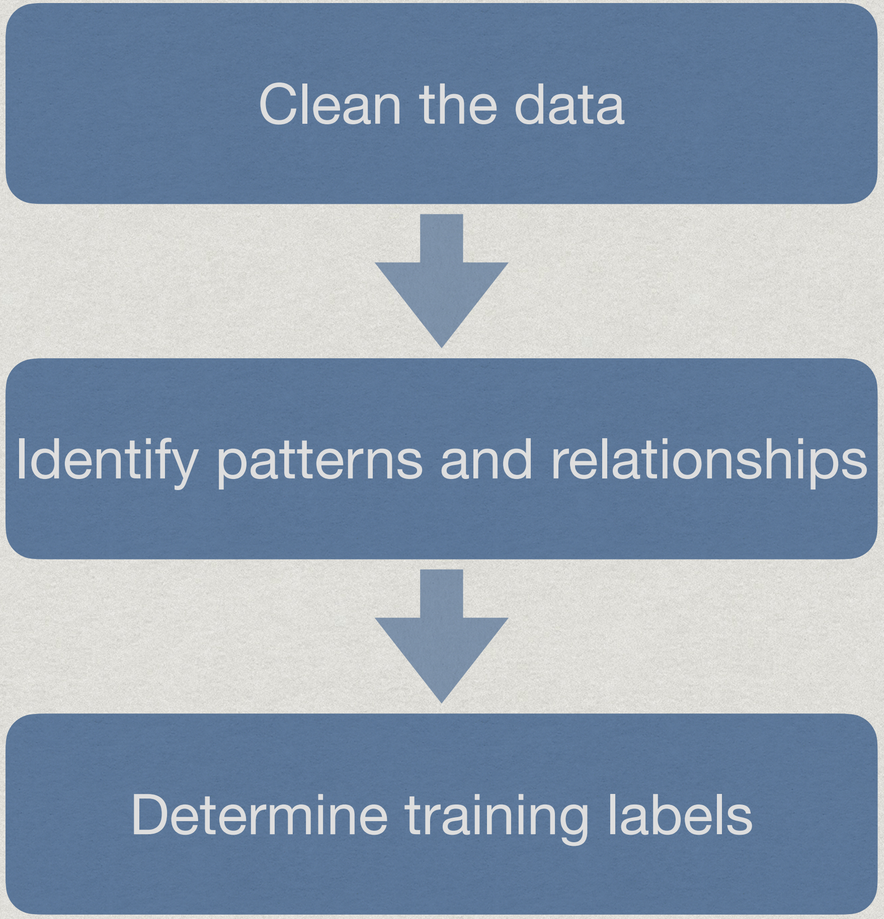
GitHub Link (click)



Data Wrangling

1. Calculated the number of launches on each site
2. Calculated the number and occurrence of each orbit
3. Calculated the number and occurrence of mission outcome of the orbits
4. Created a landing outcome label from Outcome column

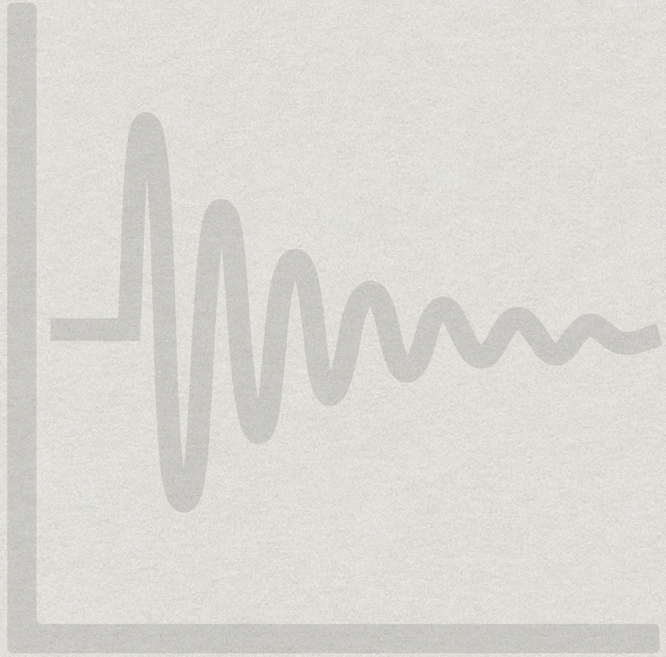
GitHub Link (click)



EDA with Data Visualization

Charts plotted to visualize the relationship:

1. Flight Number vs Launch Site
2. Payload vs Launch Site
3. Success Rate vs Orbits
4. Flight Number vs Orbits
5. Payload vs Orbits
6. Launch Success Yearly Trend



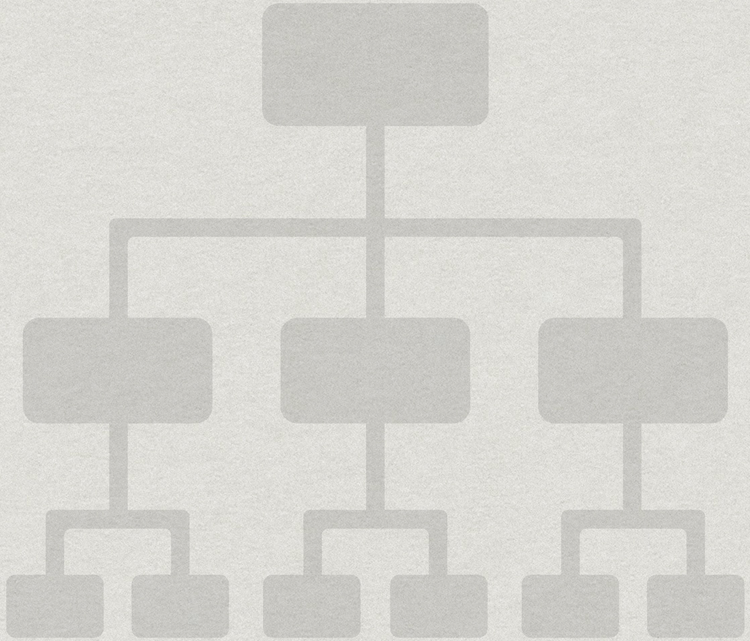
GitHub Link (click)

EDA with SQL

Few of the SQL queries performed:

1. Found the names of the unique launch sites
2. Found 5 records where launch sites begin with `CCA`
3. Calculated the total payload carried by boosters from NASA
4. Calculated the average payload mass carried by booster version F9 v1.1
5. Found the dates of the first successful landing outcome on ground pad

GitHub Link (click)



Interactive Map with Folium

- Circles indicating Launch Sites
  To show where and which launch site is located
- Markers indicating Launch Outcomes
  To show how many successful and failed launches does launch sites have
- Lines indicating proximities
  To show distance between CCAFS SLC-40 site and the nearest coastline, railway, highway, and city

GitHub Link (click)

Dashboard with Plotly Dash

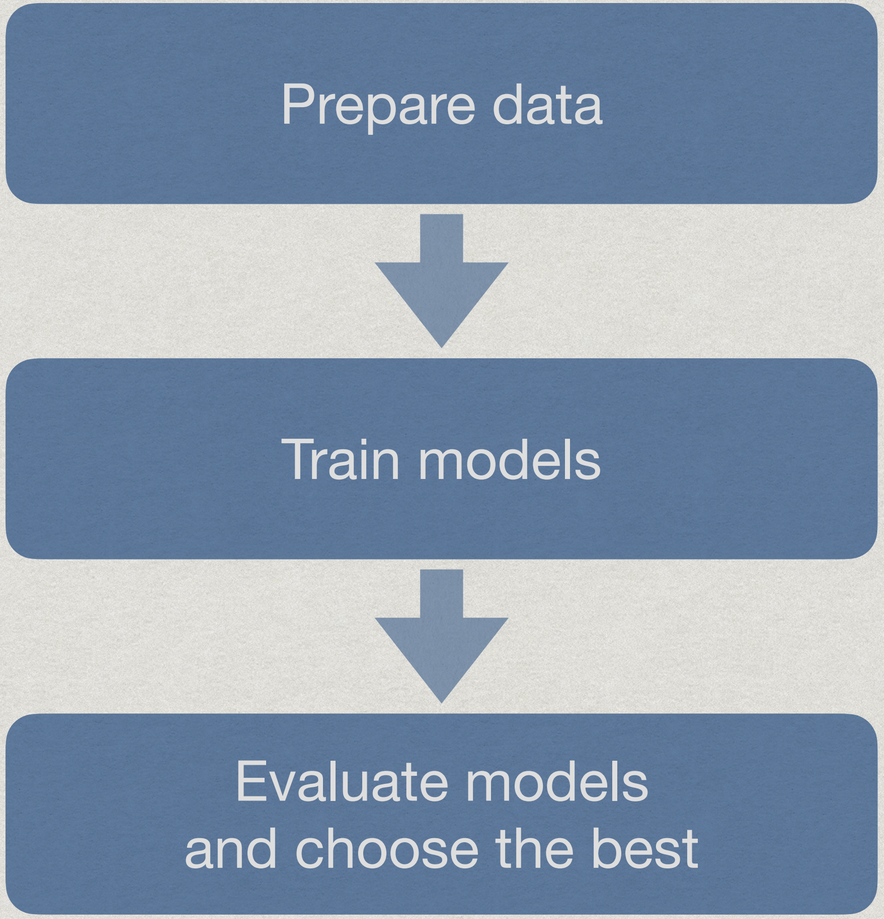
- Dropdown List with Launch Sites
  To select all launch sites or a chosen one
- Pie Chart of Successful Launches
  To show the success ratio of launches
- Slider of Payload Mass
  To select payload mass range
- Scatter Chart for Payload Mass vs Launch Outcome
  To show the relation between Payload and Outcome by Booster Version

GitHub Link (click)

Predictive Analysis: Classification

1. Split the data into X and Y, then into Train and Test
2. Train and evaluate Logistic Regression model
3. Train and evaluate Support Vector Machine model
4. Train and evaluate Decision Tree model
5. Train and evaluate k-Nearest Neighbor model
6. Compare the models

GitHub Link (click)

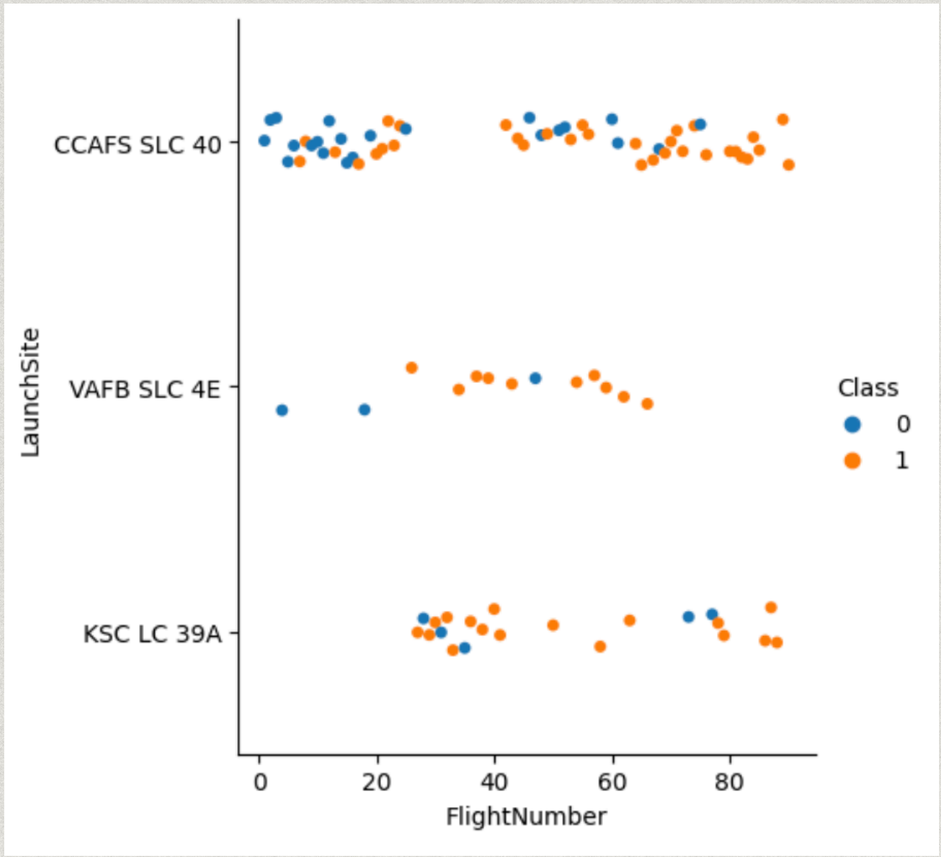


INSIGHTS DRAWN WITH EDA

SECTION №2

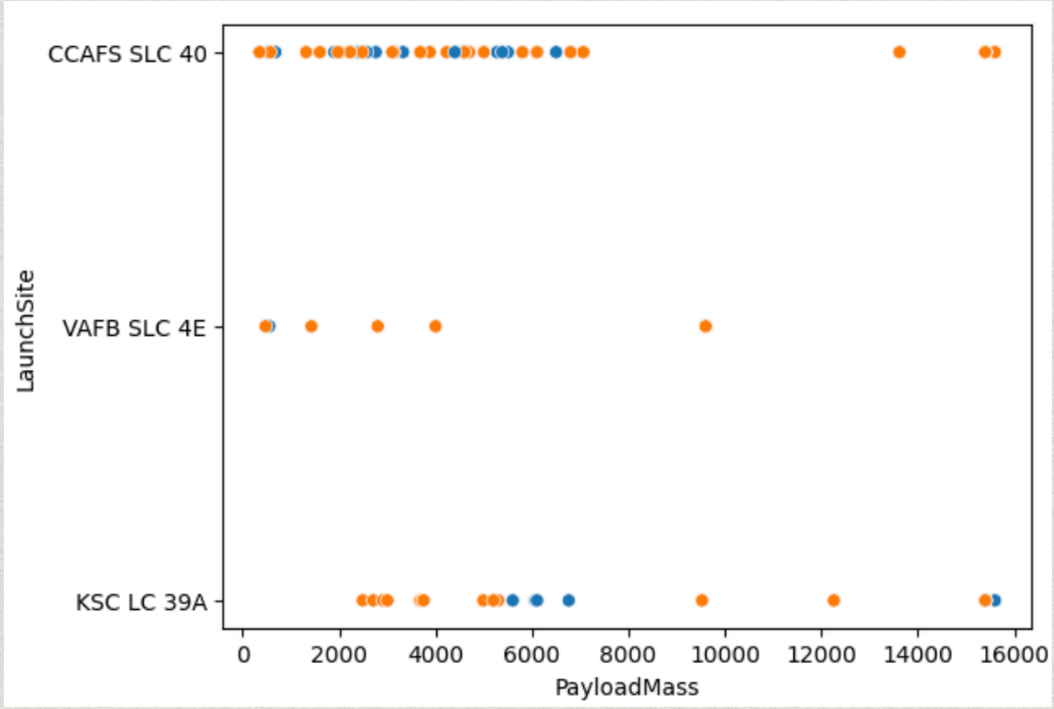
Flight Number vs Launch Site

- Graph shows Flight Number from different Launch Sites. Where Blue means Failed Launch and Orange means Successful Launch.
- Most launches were done from CCAFS SLC 40 site. Also it’s where first starts were made.
- After around 20th launch Success rate increases.



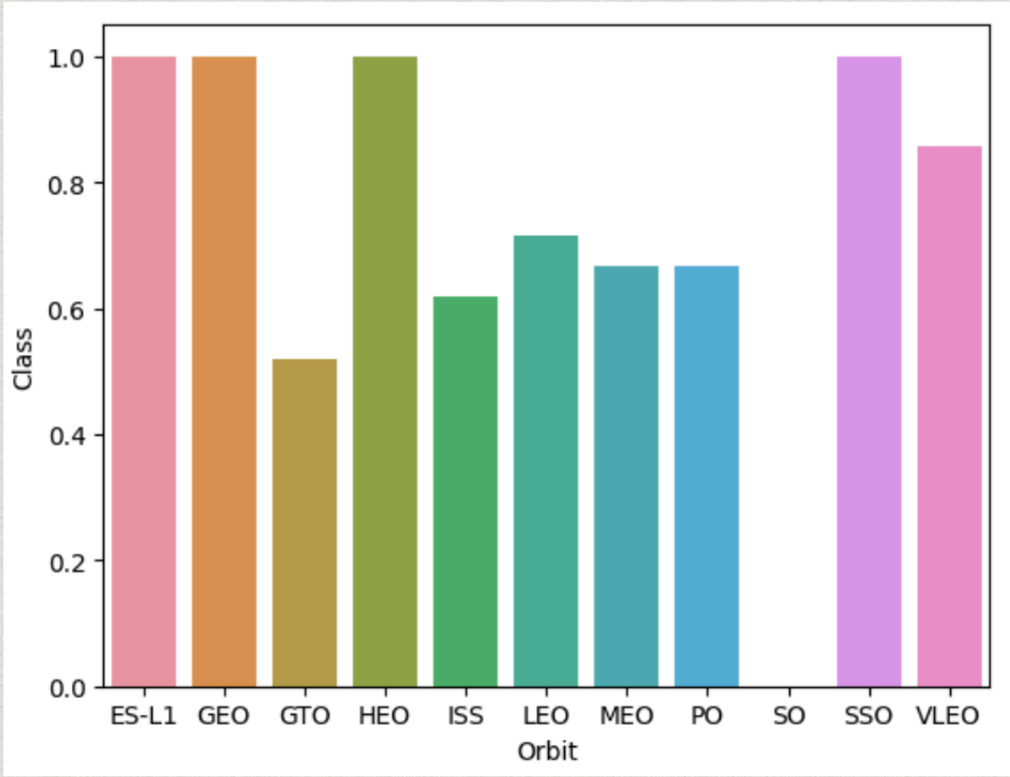
Payload vs Launch Site

- Graph shows Payload Mass launched from different Launch Sites. Where Blue means Failed Launch and Orange means Successful Launch.
- Launches with Payload greater than 7000 kg have higher success rate.
- There were no launches with Payload more than 10000 kg from VAFB SLC 4E site.



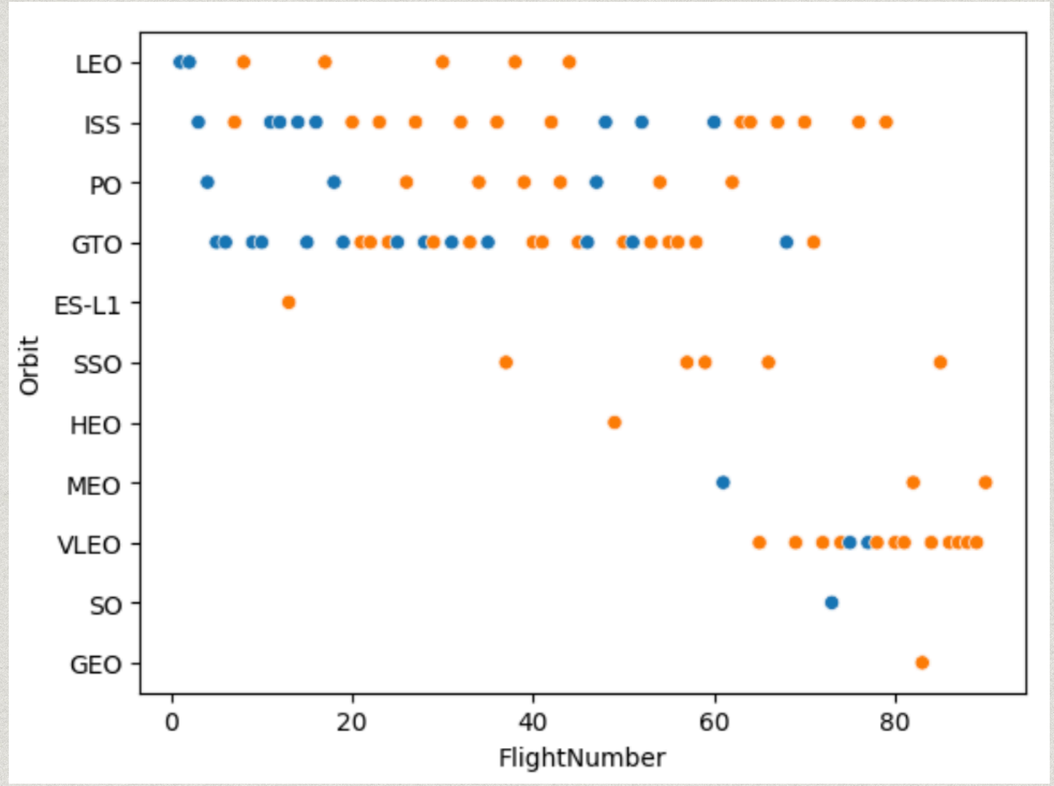
Success Rate vs Orbit Type

- Graph shows Success Rate for different Orbit launches.
- ES-L1, GEO, HEO and SSO orbits have 100% Success Rate.
- SO has the lowest Success Rate of 0%.



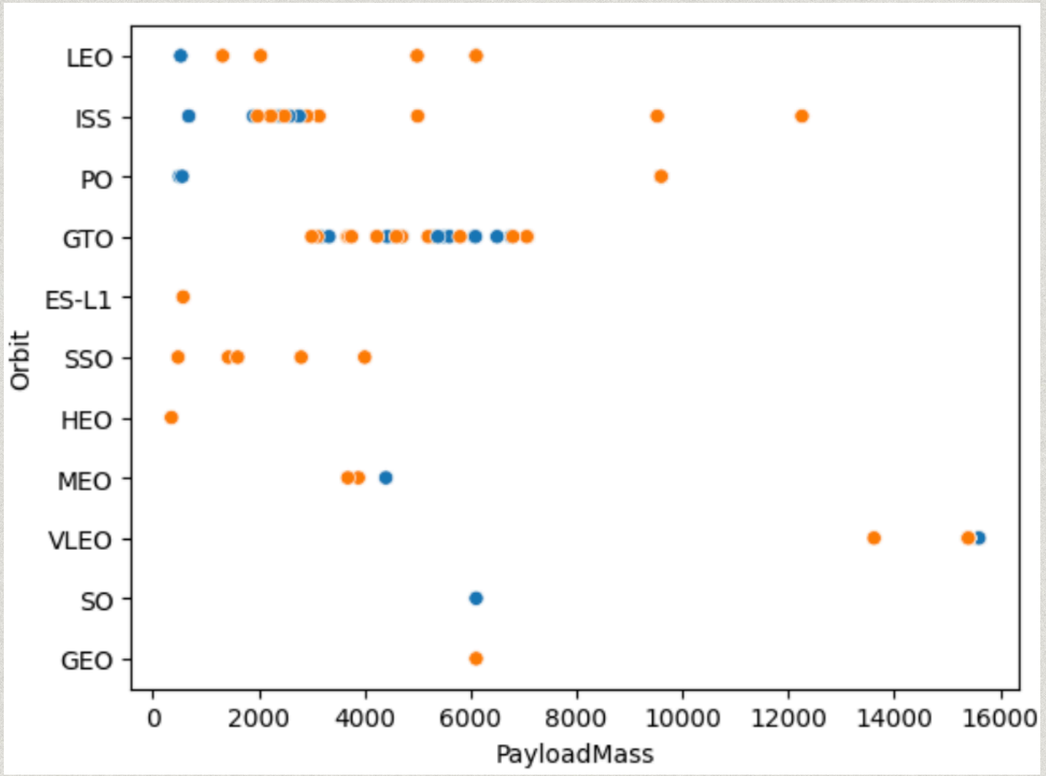
Flight Number vs Orbit Type

- Graph shows Flight Number to different Orbits. Where Blue means Failed Launch and Orange means Successful Launch.
- Orbit choice correlates with Flight Number. There are no launches to MEO, VLEO, SO and GEO before 60th launch.
- More than a half of Failed launches were made before 40th launch, and more than a half of Successful launches were made after it.



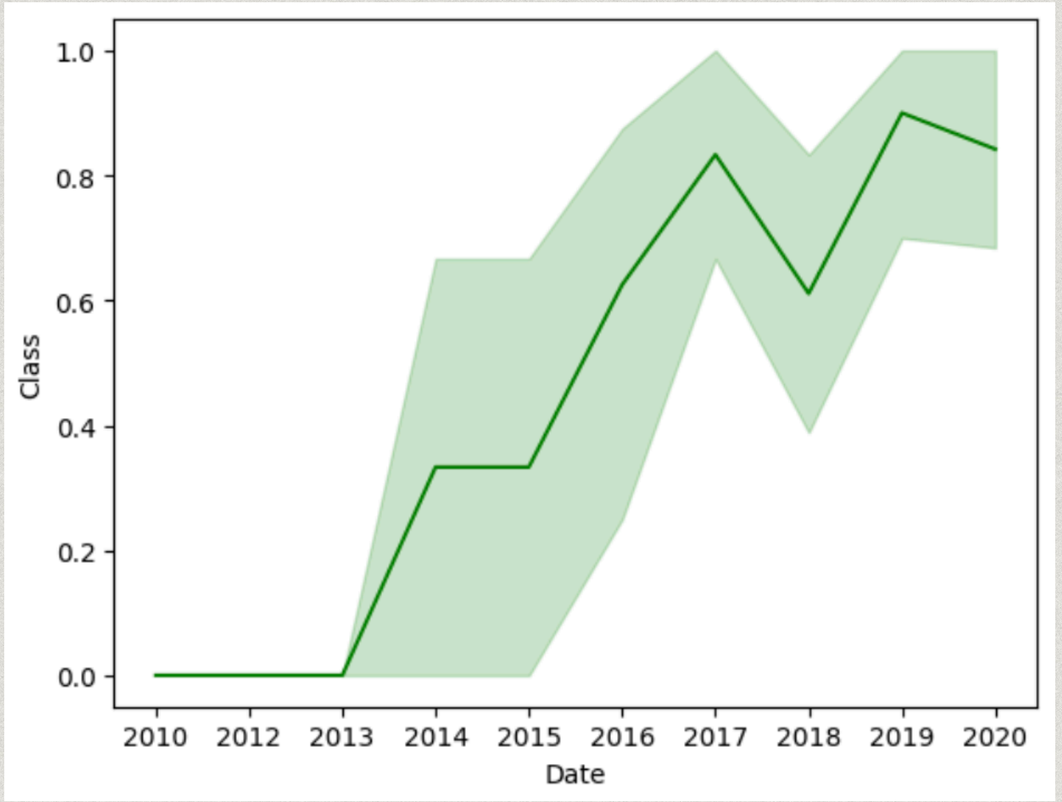
Payload vs Orbit Type

- Graph shows Payload Mass launched to different Orbits. Where Blue means Failed Launch and Orange means Successful Launch.
- Overwhelming majority of Launches where made with Payload lower than 8000 kg.
- VLEO has the highest Payload Mass average around 14500 kg.
- All 100% Success Rate Orbits have Payload Mass lower than 6000 kg.



Launch Success Yearly Trend

- Graph shows Launch Success over Years. Where Line represents mean and Area around it - min\max values.
- For only 4 years (2013-2017) Success Rate increased from 0% to 83%.
- The most successful year was 2019 with 90% Rate mean.



All Launch Site Names

- Image shows SQL query and it’s result for finding the names of the unique launch sites.

```
%%sql
select distinct LAUNCH_SITE from SPACEXTBL;
```

Running query in 'sqlite:///my_data1.db'

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

Launch Site Names: ‘CCA%’

- Image shows SQL query and it’s result for finding 5 records where launch sites begin with ‘CCA’.

```
%%sql
select * from SPACEXTBL where LAUNCH_SITE like 'CCA%' limit 5;
```

Running query in 'sqlite:///my_data1.db'

Date	Time (UTC)	Booster_Version	Launch_Site	
2010-04-06	18:45:00	F9 v1.0 B0003	CCAFS LC-40	
2010-08-12	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flig
2012-05-22	07:44:00	F9 v1.0 B0005	CCAFS LC-40	
2012-08-10	00:35:00	F9 v1.0 B0006	CCAFS LC-40	
2013-01-03	15:10:00	F9 v1.0 B0007	CCAFS LC-40	

Total Payload Mass

- Image shows SQL query and it’s result for calculating the total payload mass carried by boosters from NASA.

```
%%sql
select sum(PAYLOAD_MASS__KG_) as Total_Payload_Mass
from SPACEXTBL where CUSTOMER like 'NASA (CRS)';
```

✓ 0.4s

Running query in 'sqlite:///my_data1.db'

Total_Payload_Mass
45596

Average Payload Mass: F9 v1.1

- Image shows SQL query and it’s result for calculating the average payload mass carried by booster version F9 v1.1.

```
%%sql
select avg(PAYLOAD_MASS__KG_) as Average_Payload_Mass
from SPACEXTBL where BOOSTER_VERSION = 'F9 v1.1';
```

Running query in 'sqlite:///my_data1.db'

Average_Payload_Mass
2928.4

First Successful Ground Landing

- Image shows SQL query and it’s result for finding the dates of the first successful landing outcome on ground pad.

```
%%sql
select min(Date) as First_Successful_Landing from SPACEXTBL
where LANDING_OUTCOME like 'Success (ground pad)';
```

Running query in 'sqlite:///my_data1.db'

First_Successful_Landing
2015-12-22

Successful Drone Ship Landing

- Image shows SQL query and it’s result for listing the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000.

```
%%sql
select BOOSTER_VERSION from SPACEXTBL where
LANDING_OUTCOME = 'Success (drone ship)'
and (PAYLOAD_MASS__KG_ between 4000 and 6000);
```

Running query in 'sqlite:///my_data1.db'

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

Total Score of Mission Outcomes

- Image shows SQL query and it's result for calculating the total number of successful and failure mission outcomes.
- Total Successful missions: 100
- Total Failed missions: 1

```
%%sql
select MISSION_OUTCOME, COUNT(*) as Total_Number
from SPACEXTBL group by MISSION_OUTCOME;
```

✓ 0.4s

Running query in 'sqlite:///my_data1.db'

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

Boosters Maximum Payload

- Image shows SQL query and it's result for listing the names of the booster which have carried the maximum payload mass.
- Note: the list was truncated so had 2 values hidden.

```
%%sql
select BOOSTER_VERSION from  SPACEXTBL where
PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_)
from SPACEXTBL);
```

Running query in 'sqlite:///my_data1.db'

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

Launch Records: 2015

- Image shows SQL query and it’s result for listing the failed landing outcomes in drone ship, their booster versions and launch site names for in year 2015.

```
%%sql
select substr(Date,6,2) as Month, BOOSTER_VERSION, LAUNCH_SITE, LANDING_OUTCOME
from SPACEXTBL where LANDING_OUTCOME  = 'Failure (drone ship)'
and substr(Date,1,4) = '2015';
```

Running query in 'sqlite:///my_data1.db'

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

Landing Outcomes Rank

- Image shows SQL query and it's result for ranking the count of landing outcomes between the date 2010-06-04 and 2017-03-20, in descending order.

```
%%sql
select LANDING_OUTCOME, count(*) as Count_Landing_Outcomes
from SPACEXTBL where DATE between '2010-06-04' and '2017-03-20'
group by LANDING_OUTCOME order by Count_Landing_Outcomes desc;
```

Running query in 'sqlite:///my_data1.db'

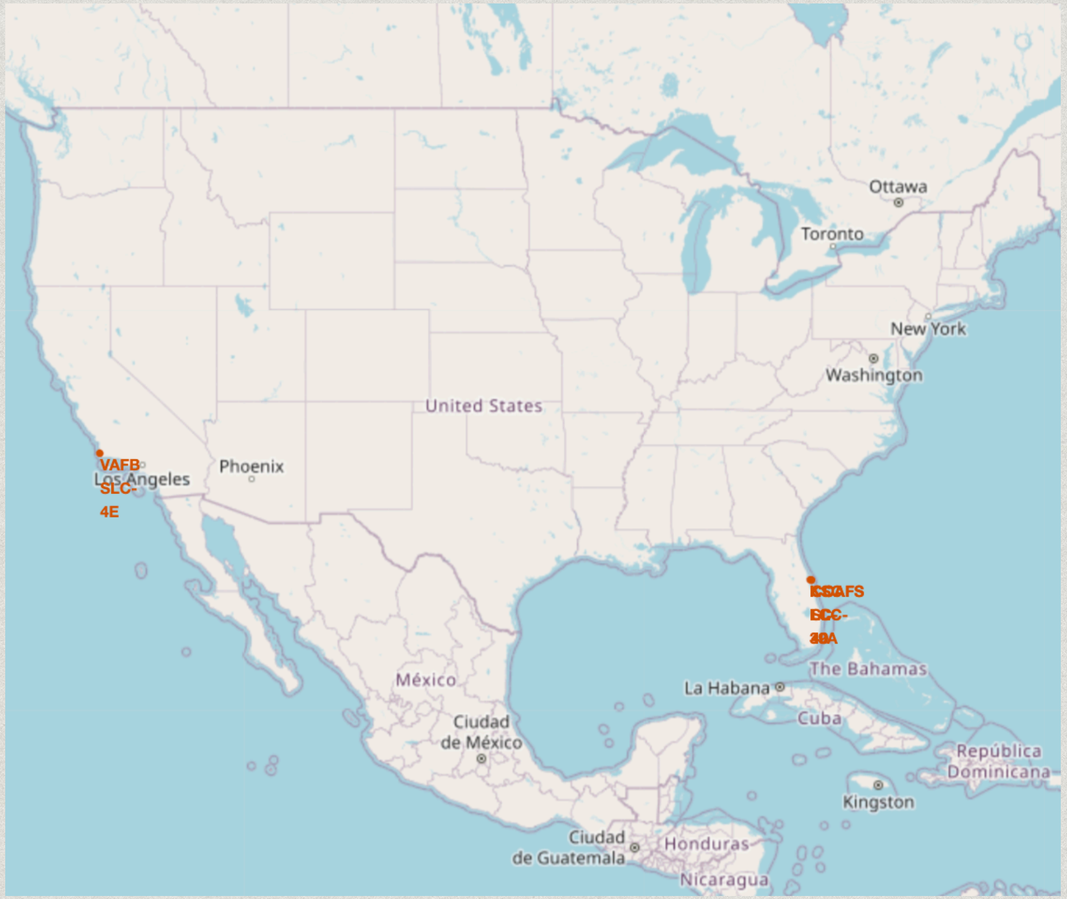
Landing_Outcome	Count_Landing_Outcomes
No attempt	10
Success (ground pad)	5
Success (drone ship)	5
Failure (drone ship)	5
Controlled (ocean)	3
Uncontrolled (ocean)	2
Precluded (drone ship)	1
Failure (parachute)	1

LAUNCH SITES PROXIMITIES ANALYSIS

SECTION №3

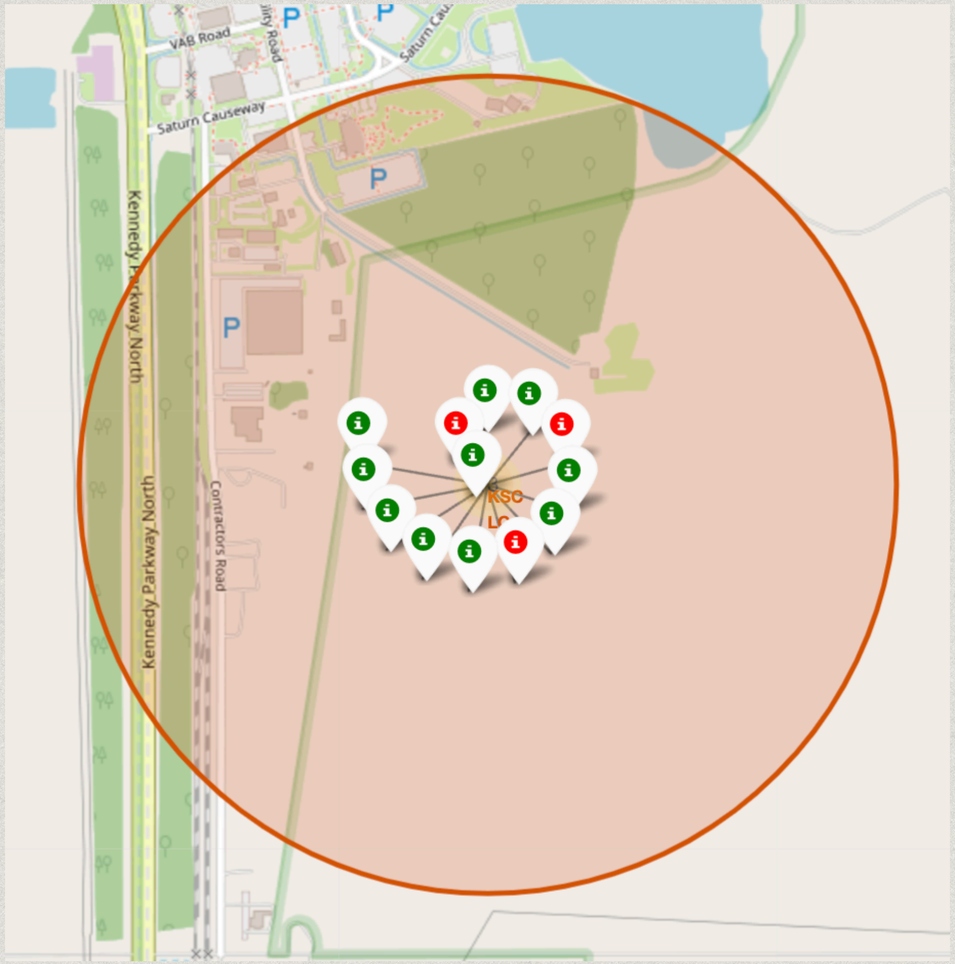
Launch Sites Locations

- Map shows Locations of Launch Sites.
- All of them are in the US: California and Florida States. Built on the south coasts, near Equator.



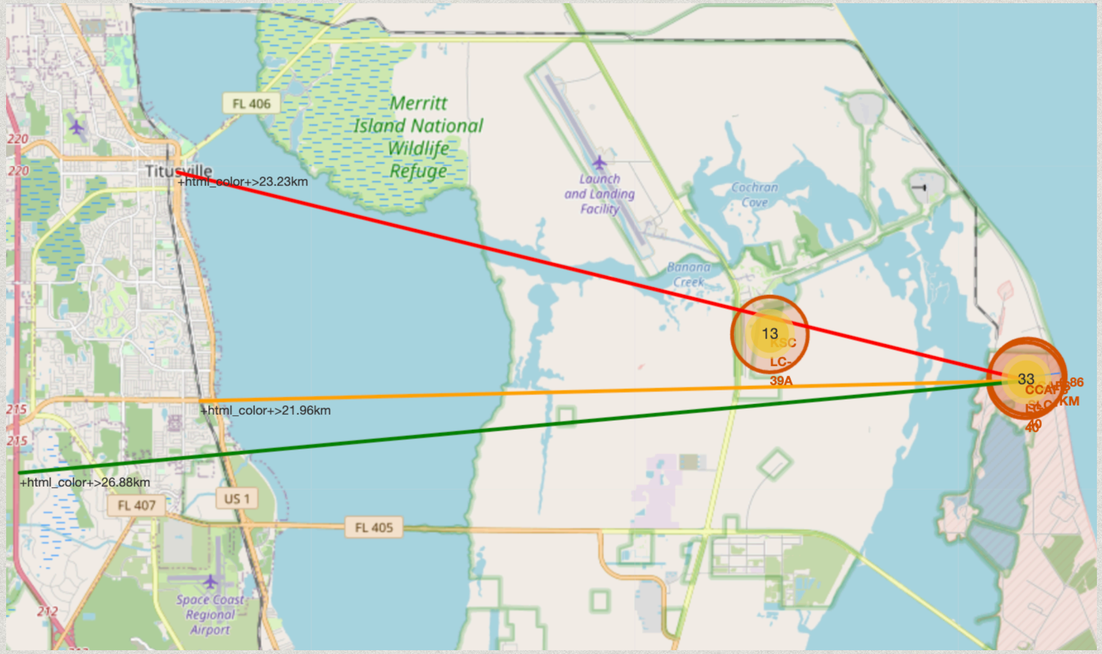
Launch Outcome Markers

- Map shows Launch Outcome Markers from KSC LC 39A site. Where Red means Failed Launch and Green means Successful Launch.



Launch Sites Proximities

- Map shows Launch Site Proximities from CCAFS SLC 40 site.
- Red line: distance to the nearest town is 23 km.
- Orange line: distance to the nearest railway is 22 km.
- Green line: distance to the nearest highway is 27 km.
- Blue line: distance to the nearest coast is 0.86 km.

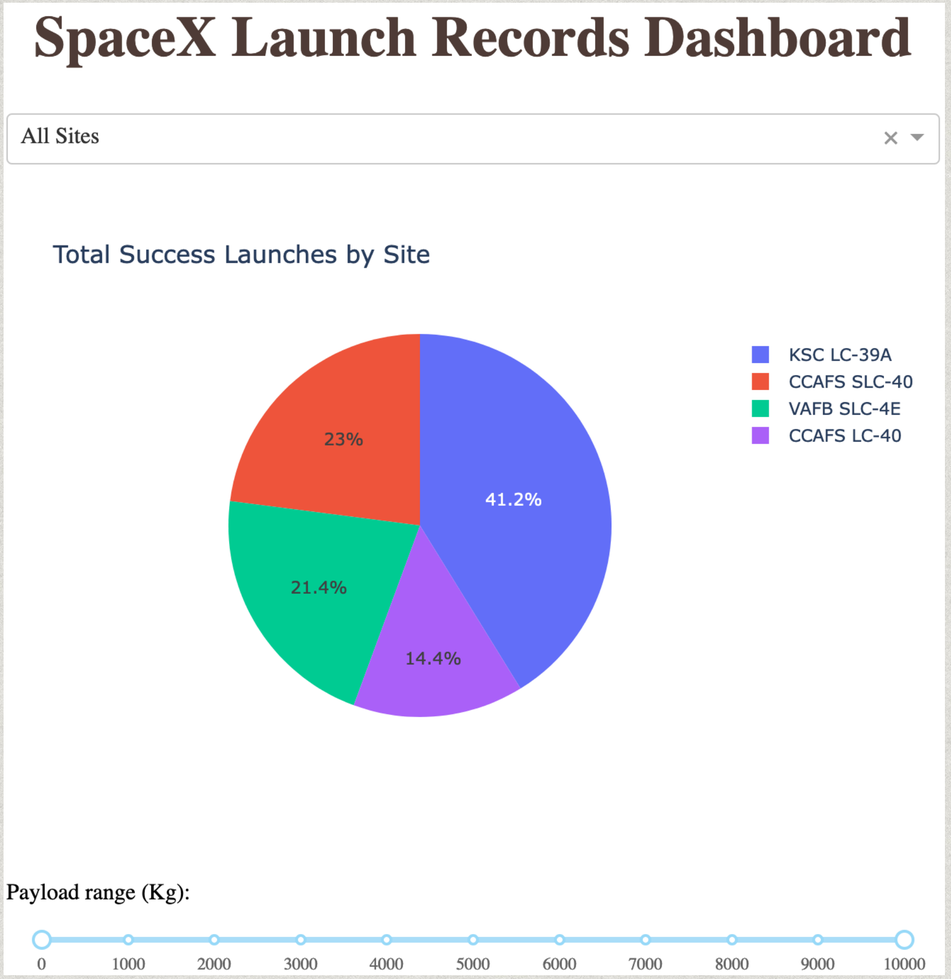


DASHBOARD WITH PLOTLY DASH

SECTION №4

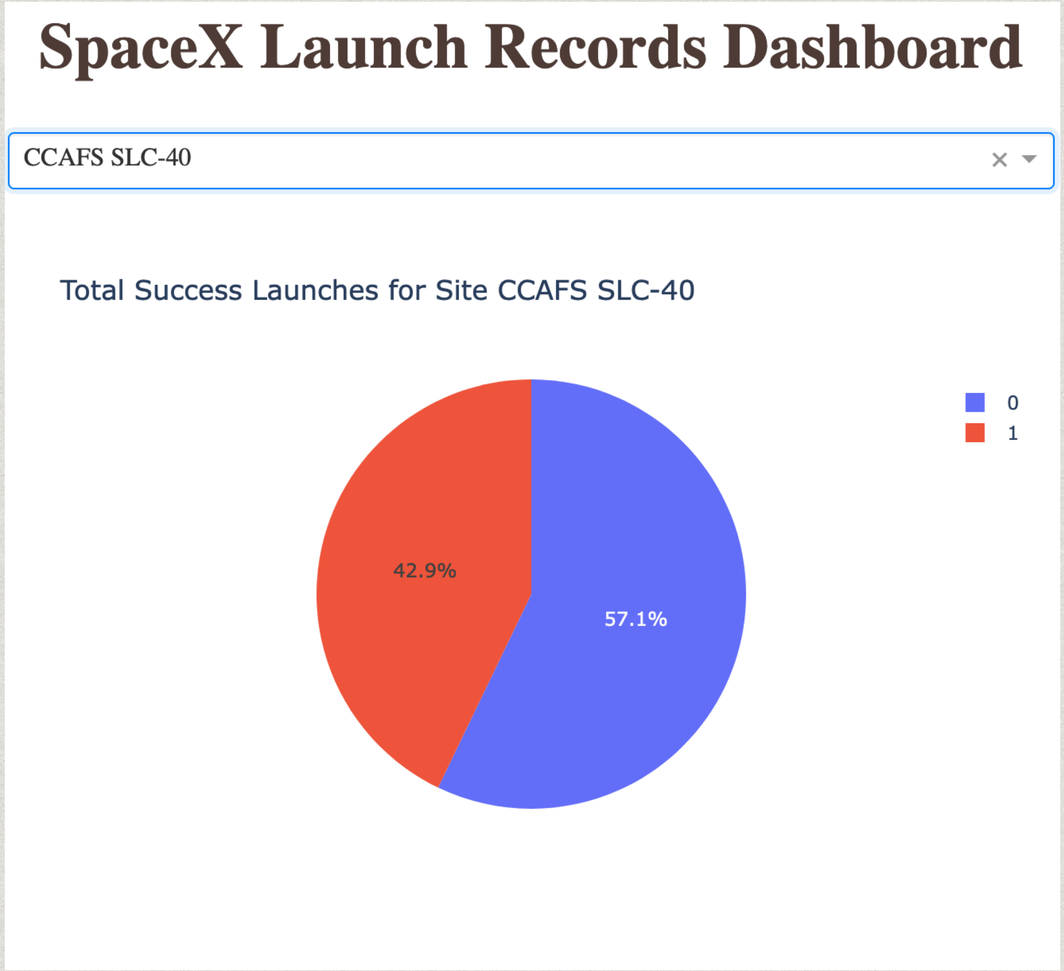
Success Launches by Site

- Chart shows Success Rate from different Launch Sites.
- KSC LC-39A has the highest absolute number of Success Launches.



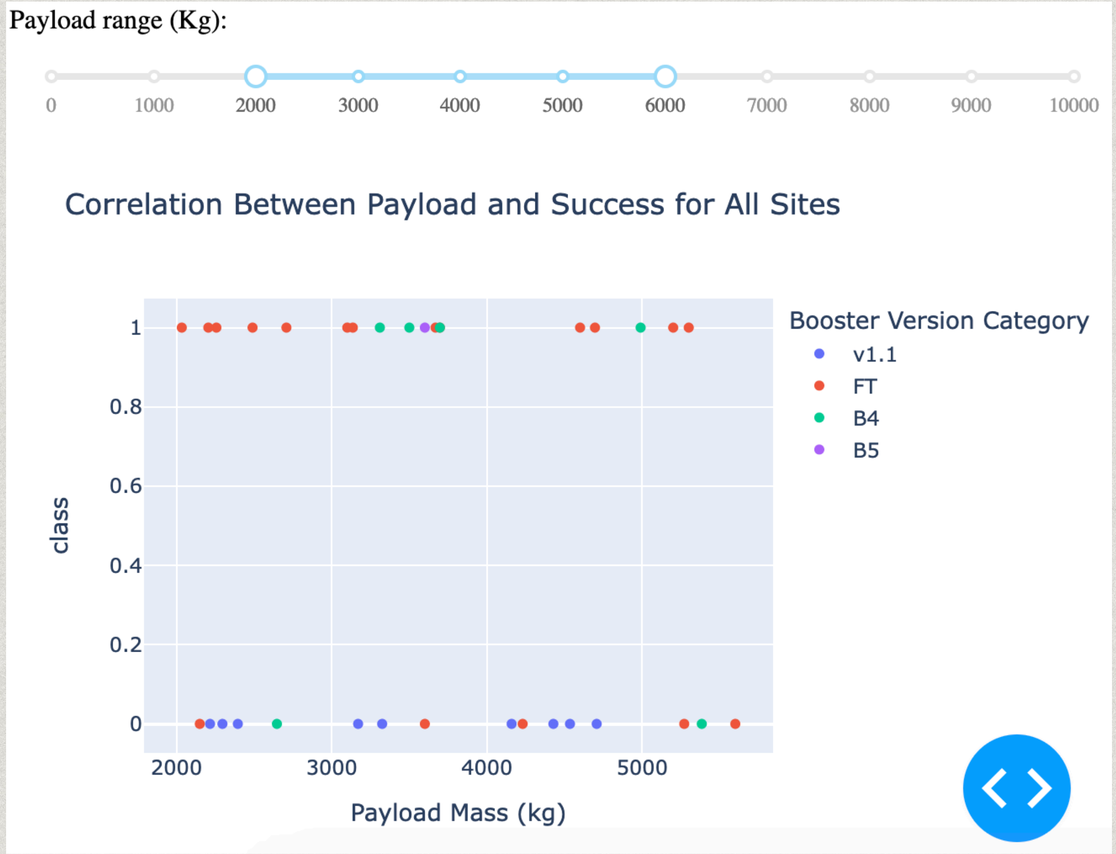
Best Success Ratio Launch Site

- Chart shows Success Rate from CCAFS SLC-40 Launch Site. Where Blue means Failed Launch and Red means Successful Launch.
- CCAFS SLC-40 has the highest rational number of Success Launches equal to 42.9%.



Best Success Ratio Payload Mass

- Graph shows Launch Outcome for different Payload Mass. Where 1 is Success, 0 is Failure and colors represent Booster Version Category.
- Payload mass between 2000 and 6000 kg has the highest Success Rate.

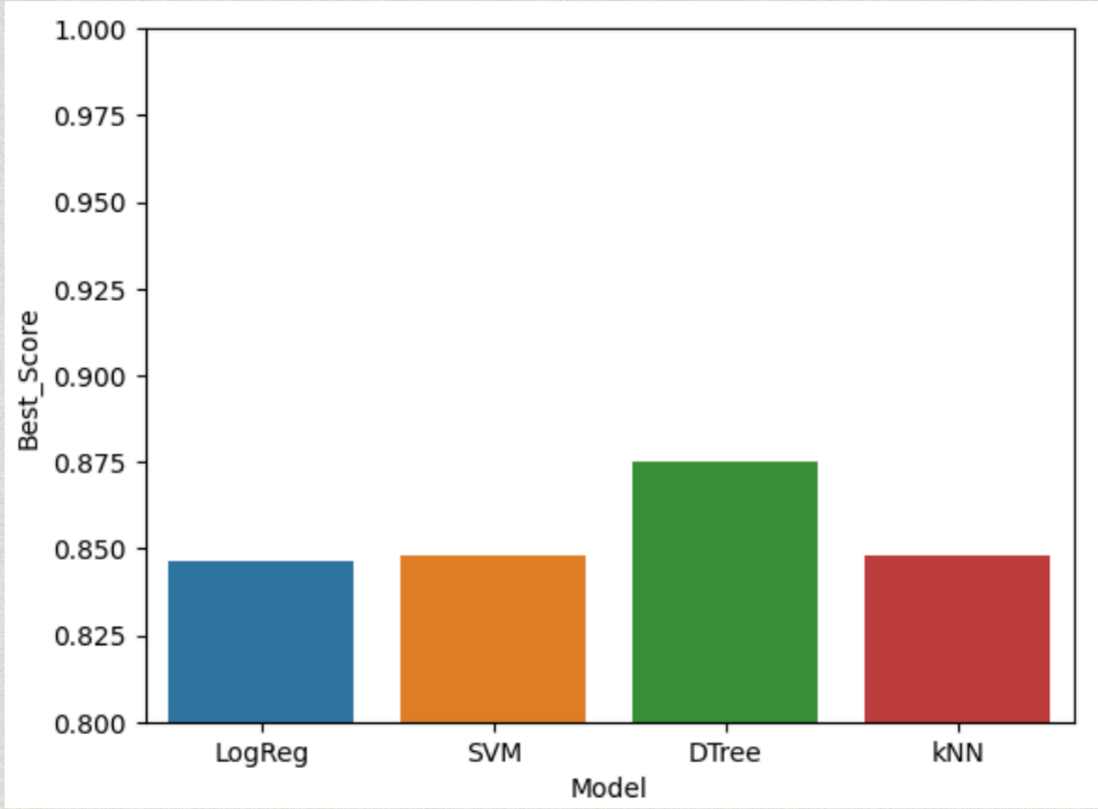


PREDICTIVE ANALYSIS: CLASSIFICATION

SECTION №5

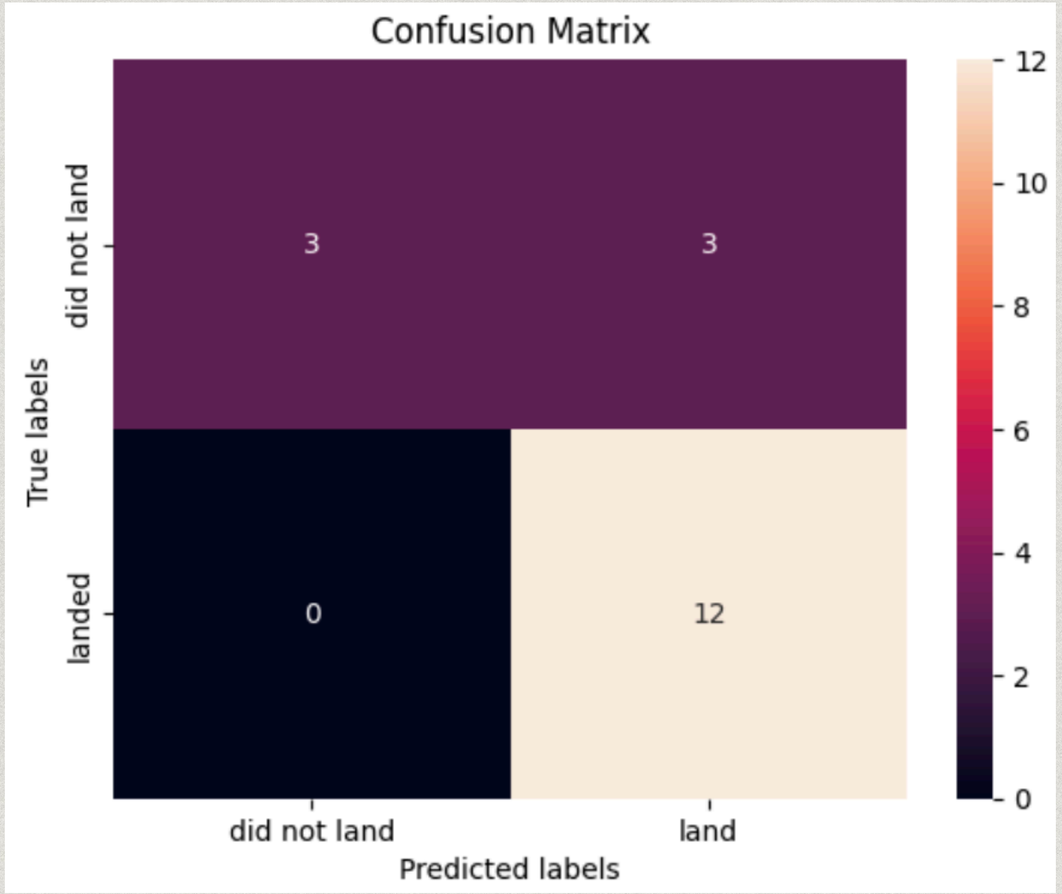
Classification Accuracy

- Graph shows .best_score_ values for different models.
- All models have the same accuracy and score. The .best_score_ value was the highest for Decision Tree model.
- Values are:
  LogReg: 0.84643
  SVM: 0.84821
  DTree: 0.87500
  kNN: 0.84821
  (*ylim was set to 0.8 for better expression)*



Confusion Matrix

- Image shows Confusion Matrix for any of the models. All models have the same values.
- There are no False Negatives, but 3 False Positives in every model.



Conclusion

- To maximize mission success rate:
  launch from CCAFS SLC-40 to SSO orbit with payload mass lower than 5000 kg.
- To maximize prediction of launch outcome:
  use decision tree classifier.
- To maximize model performance:
  consider larger dataset and test other models.
- And always remember…


DON'T
PANIC!