IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<LAURIE ZHAO>
<5/29/2025>

Outline

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

Executive Summary

- Summary of methodologies
  - Data collection with web scraping
  - Exploratory Data Analysis (EDA): data wrangling, data visualization
  - Predictive analysis

- Summary of all results
  - EDA results (best predictive features)
  - Predictive analysis results (best predictive models)

Introduction

- Project background and context
  - The goal of this project is to predict whether or not the Falcon 9 first stage will land successfully. Since SpaceX can reuse the first stage of the rocket, SpaceX is able to launch rockets at a noticeably lower price than other providers. If we can predict the success of the landing, we can estimate the cost of a launch and compare them to other companies.
- Problems you want to find answers
  - What features lead to a successful landing of the Falcon 9 rocket launch given the launch site, orbit type, etc.?

Section 1

Methodology

Methodology

Executive Summary

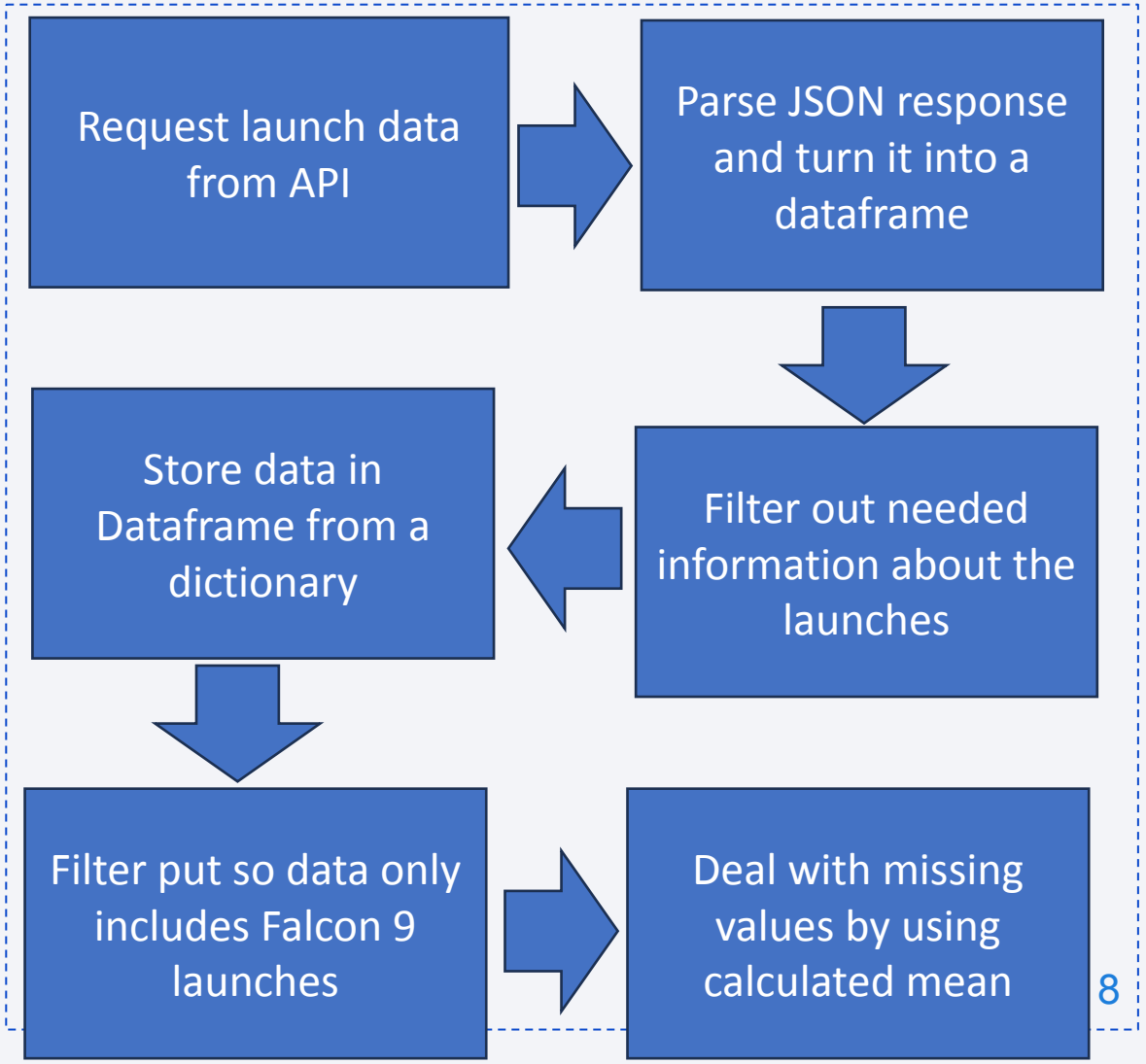
- Data collection methodology:
  - SpaceX API
  - Web Scraping
- Perform data wrangling
  - Filtering through data to find missing values and note key features
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Optimize results and accuracy of each classification model by building, tuning, and evaluation of the models

Data Collection

- Data collection comes from a combination of SpaceX API (https://api.spacexdata.com/v4/rockets/) and web scraping data from a Wikipedia page containing information about Falcon 9 launches (https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922).

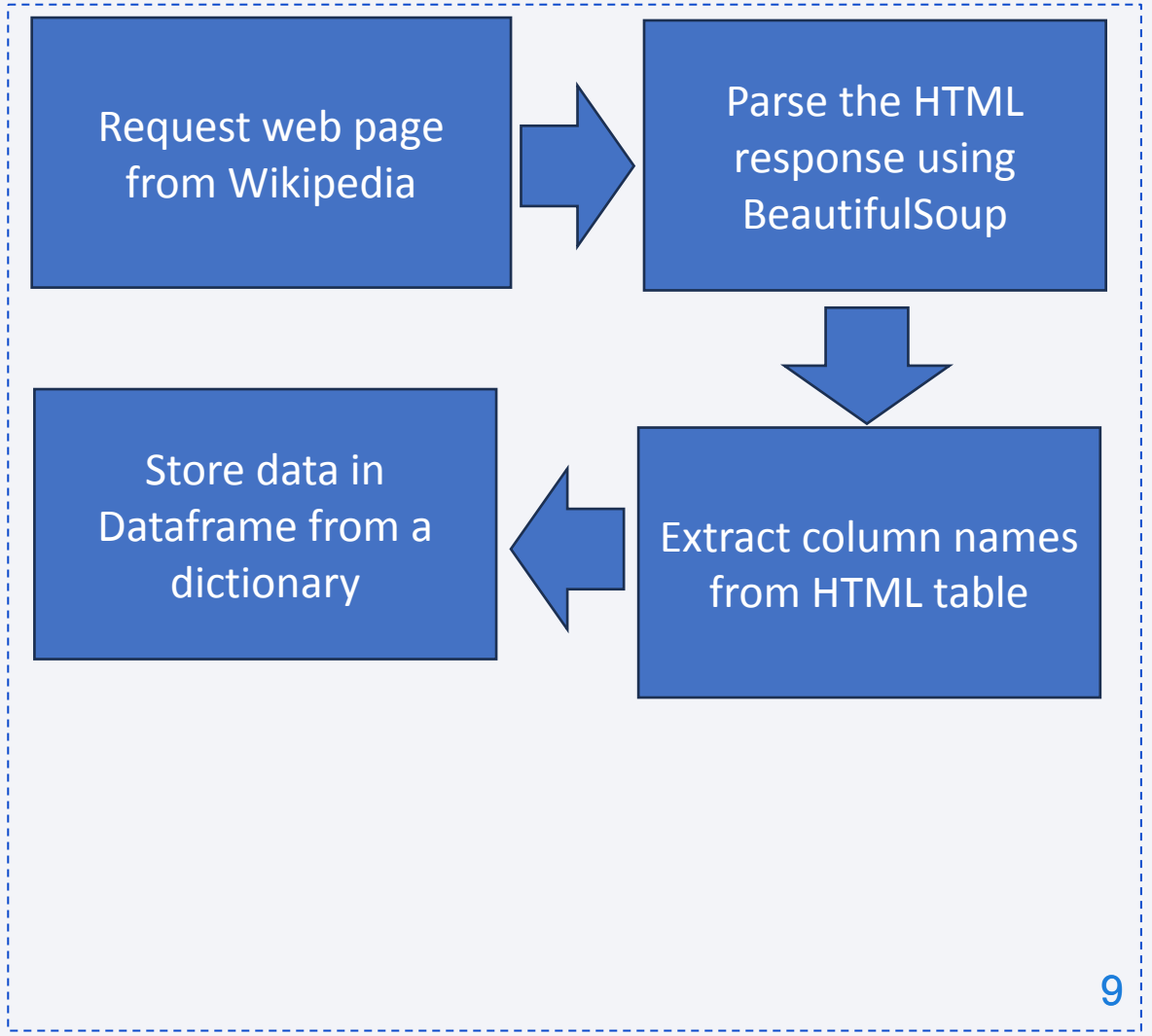
Data Collection – SpaceX API

- https://github.com/lauriezz/Applied-Data-Science-Capstone/blob/main/jupyter-labs-spacex-data-collection-api.ipynb



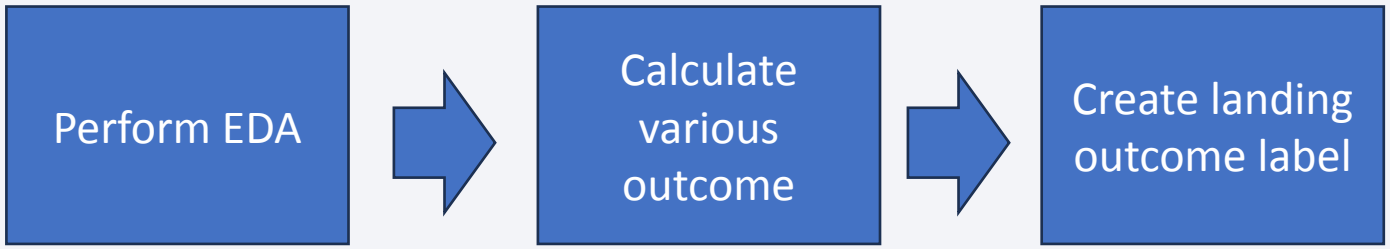
Data Collection - Scraping

- https://github.com/lauriezz/Applied-Data-Science-Capstone/blob/main/jupyter-labs-webscraping.ipynb



Data Wrangling

- https://github.com/lauriezz/Applied-Data-Science-Capstone/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb



- Step 1: perform Exploratory Data Analysis (EDA) on the data
- Step 2: Calculations such as number of launches on each site, occurrence of each orbit, and occurrence of mission outcome per orbit type
- Step 3: creating the landing outcome label from the outcome column and exporting the data to CSV

EDA with Data Visualization

- https://github.com/lauriezz/Applied-Data-Science-Capstone/blob/main/edadataviz.ipynb
- Charts that were plotted to visualize relationship between 2 variables:
  - Flight number vs. Payload Mass, Flight Number vs. Launch site, Payload Mass vs. Launch Site, Orbit Type vs. Success Rate, Flight Number vs. Orbit, Payload Mass vs. Orbit
  - Scatter plots are used to show relationship between 2 variables
  - Bar charts are used to compare data
  - Line charts are used to show trends over time

EDA with SQL

- https://github.com/lauriezz/Applied-Data-Science-Capstone/blob/main/jupyter-labs-eda-sql-coursera_sqllite%20(1).ipynb
- SQL queries that were performed:
  - Names of unique launch sites
  - 5 launch sites whose names began with ‘CCA’
  - Total payload mass carried by boosters launched by NASA
  - Average payload mass carried by booster F9 v1.1
  - Date of first successful landing outcome in ground pad was achieved
  - Names of boosters that have success in drone ship and have payload mass between 4000 and 6000 kg
  - Number of successful and failure missions
  - Names of booster versions which have carried max payload mass
  - Failed landing outcomes in drone ship, their booster versions, and launch site names for the months in year 2015
  - Rank of the count of landing outcomes between the date 2010-06-04 and 2017-03-20 in descending order

Build an Interactive Map with Folium

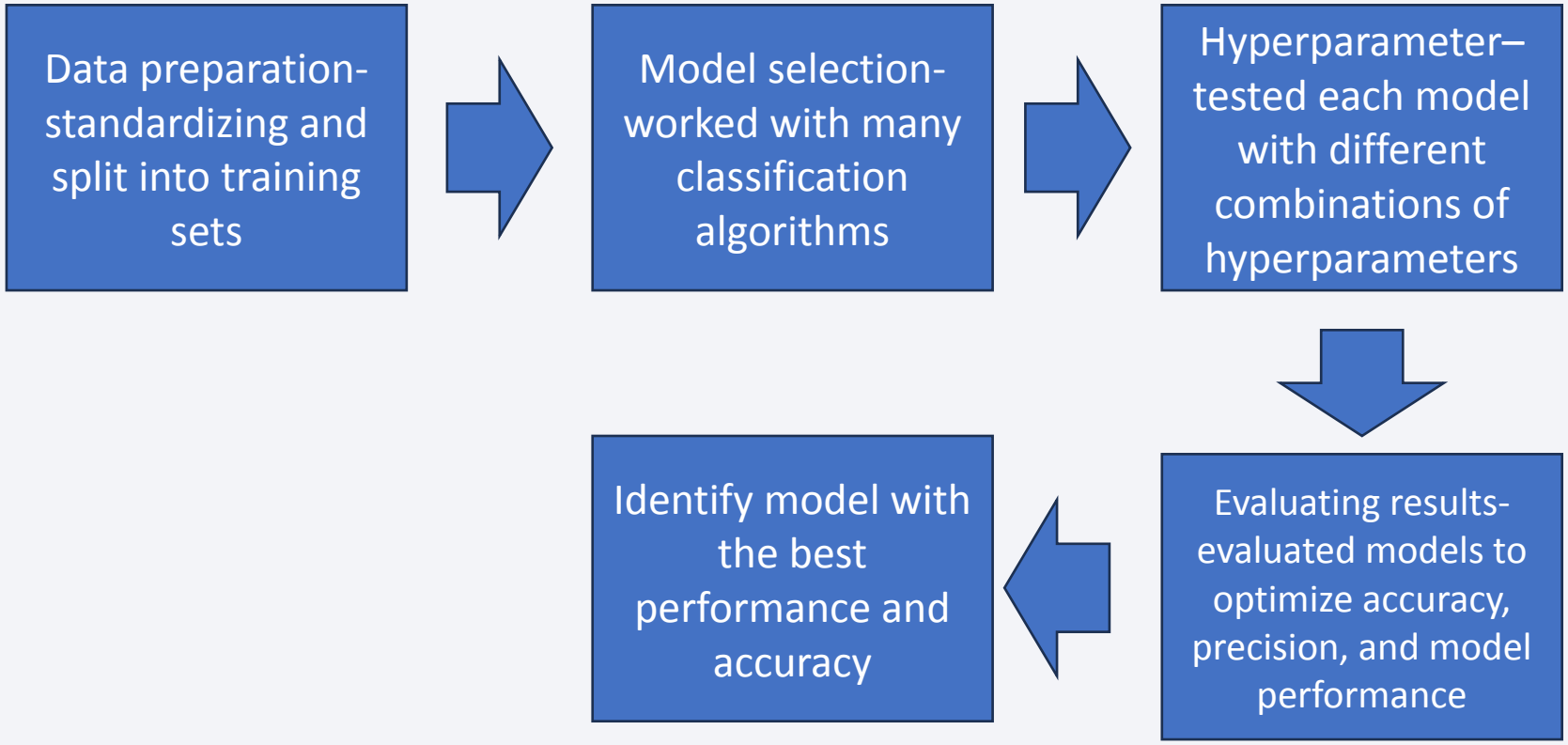
- https://github.com/lauriezz/Applied-Data-Science-Capstone/blob/main/lab_jupyter_launch_site_location.ipynb
- Markers, circles, lines, and marker clusters were used for all launch sites
- Markers- Indicate launch sites on map
- Circles- indicate areas around launch sites
- Lines- connect launch sites with the areas around and other locations
- Marker clusters- indicate events in each coordinate (launches)

Build a Dashboard with Plotly Dash

- Pie chart
  - Showed the percentage of total successful launches vs. failed launches
- Scatterplot of Success Payload
  - Payload mass vs. launch success, shows how payload mass affects outcomes
- Slider for Payload Mass Range
  - Allows for selection of payload range
- Launch site dropdown
  - Allows for selection of launch sites for analysis

Predictive Analysis (Classification)

- https://github.com/lauriezz/Applied-Data-Science-Capstone/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

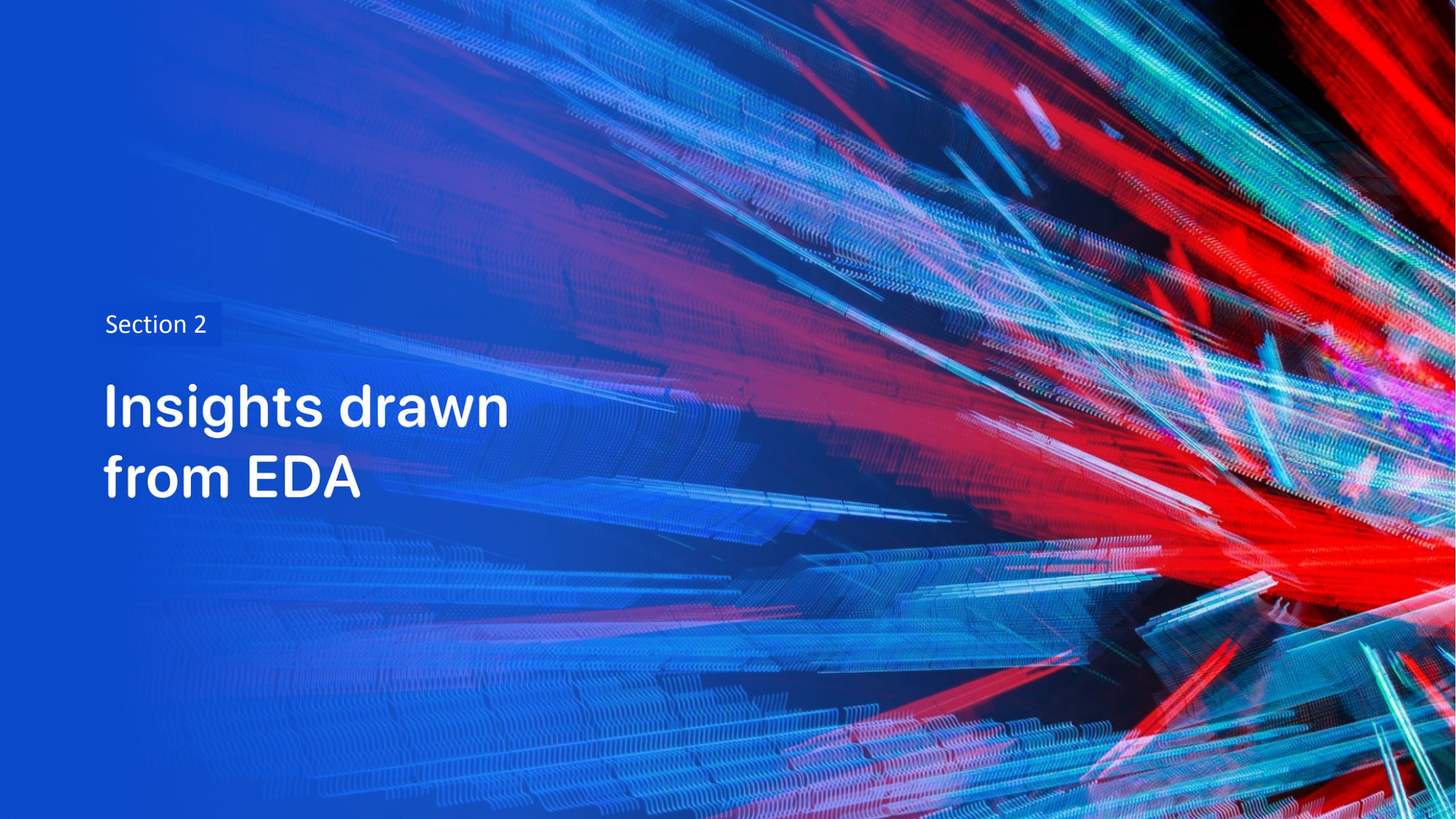


Results

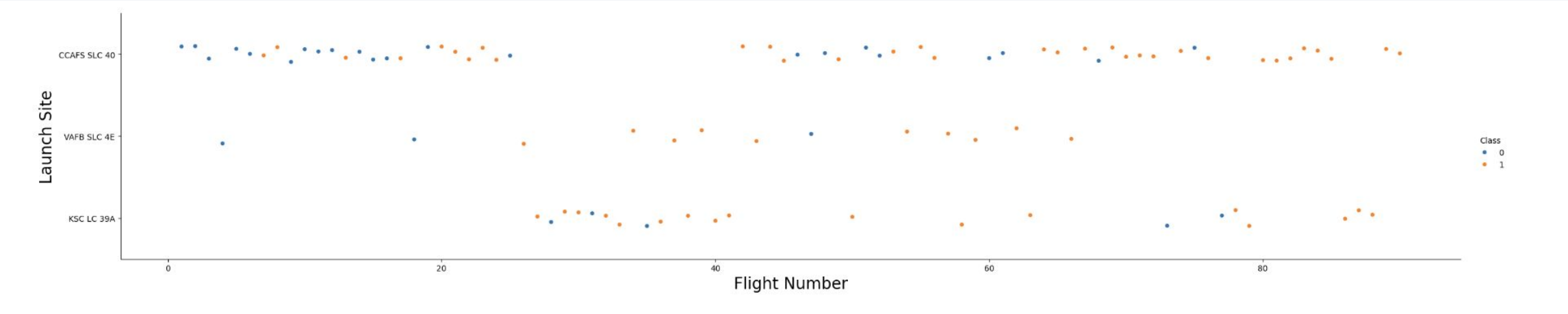
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

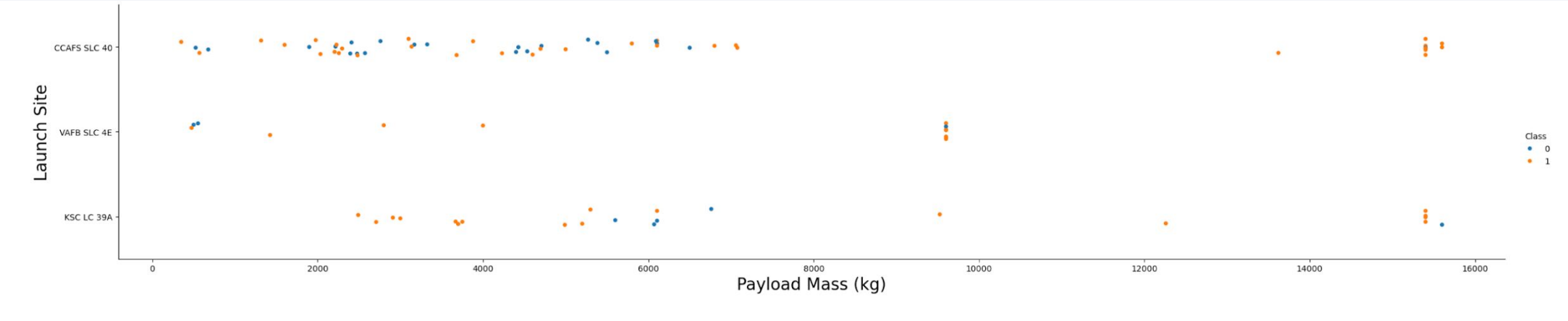


Flight Number vs. Launch Site



- The earlier flights showed mostly failure while the recent ones showed majority success
- CCAFS SLC 40 had most of the launches while VAFB SLC 4E and KSC LC 39A showed more consistent successes

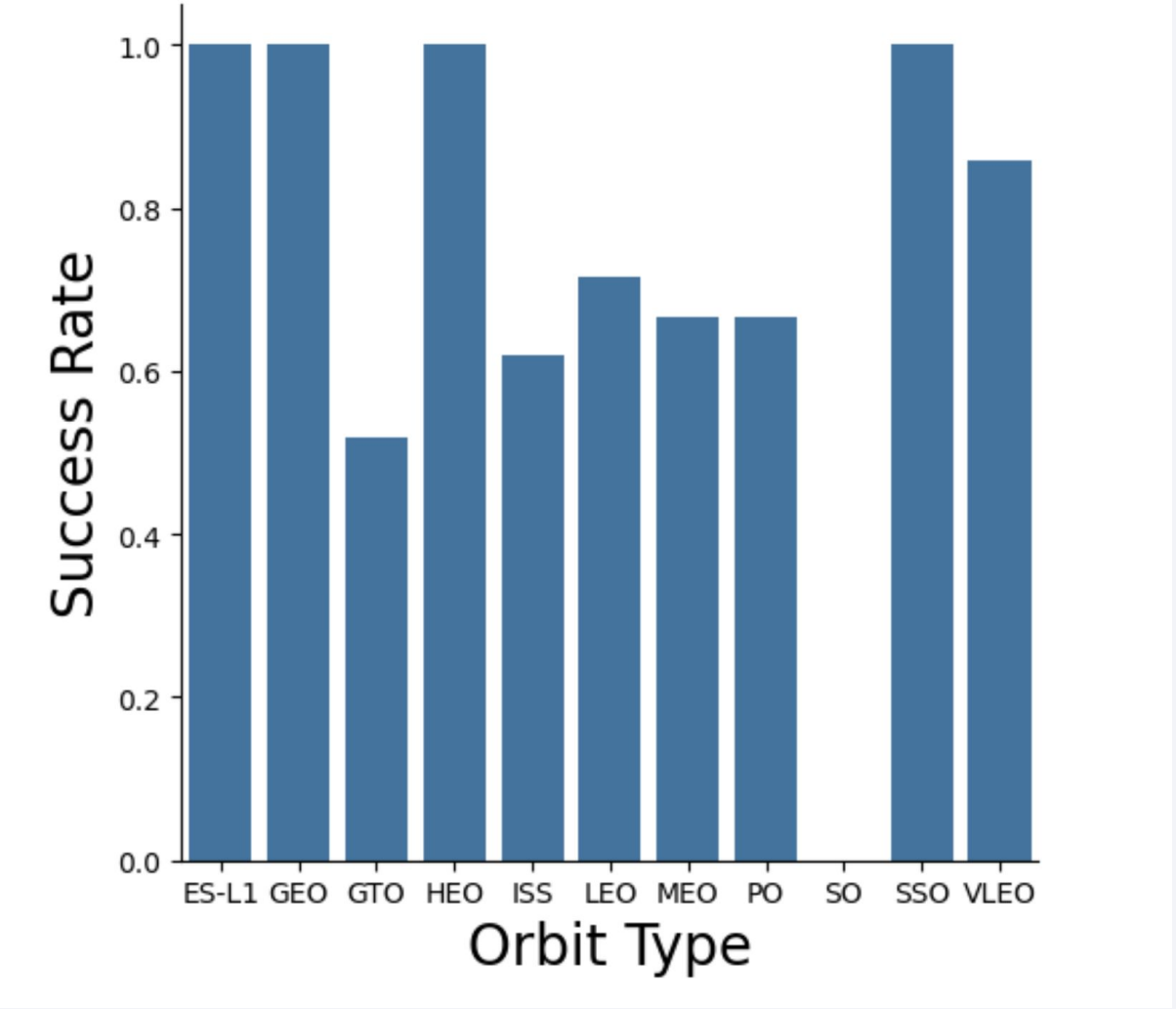
Payload vs. Launch Site



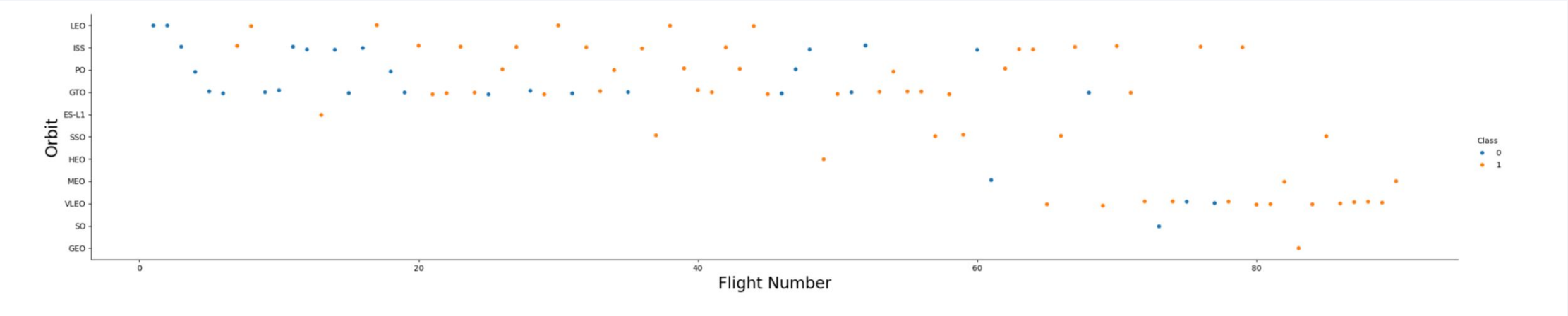
- It seems that as payload mass increases, so does success rate
- Launches with payload mass over about 9000 had very high success rate

Success Rate vs. Orbit Type

- ES-L1, GEO, HEO, and SSO all had a 100% success rate
- Between 80% and 100%:
  - VLEO
- Between 60% and 80%
  - ISS, LEO, MEO, PO
- Between 40% and 60%
  - GTO
- 0% success rate
  - SO

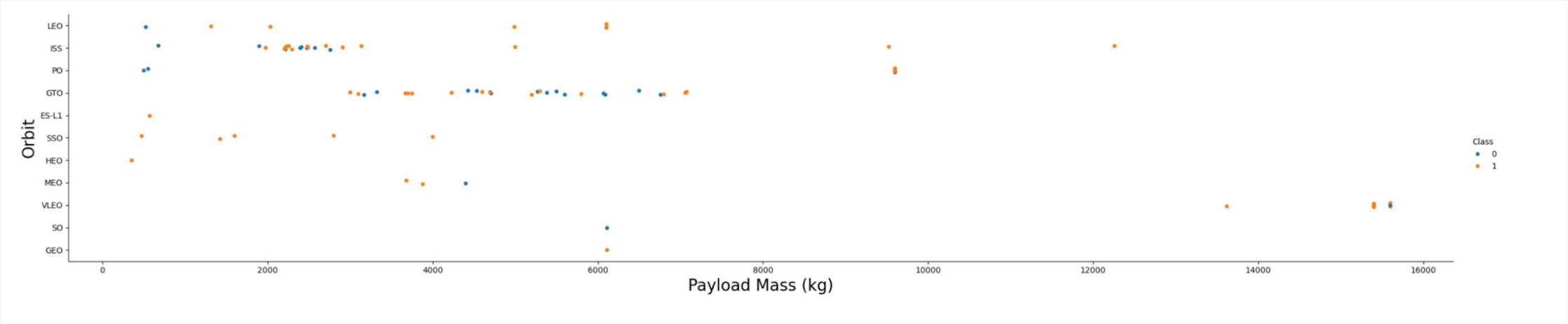


Flight Number vs. Orbit Type



- Generally, it seems that as flight number increased, success increased as well with a few exceptions

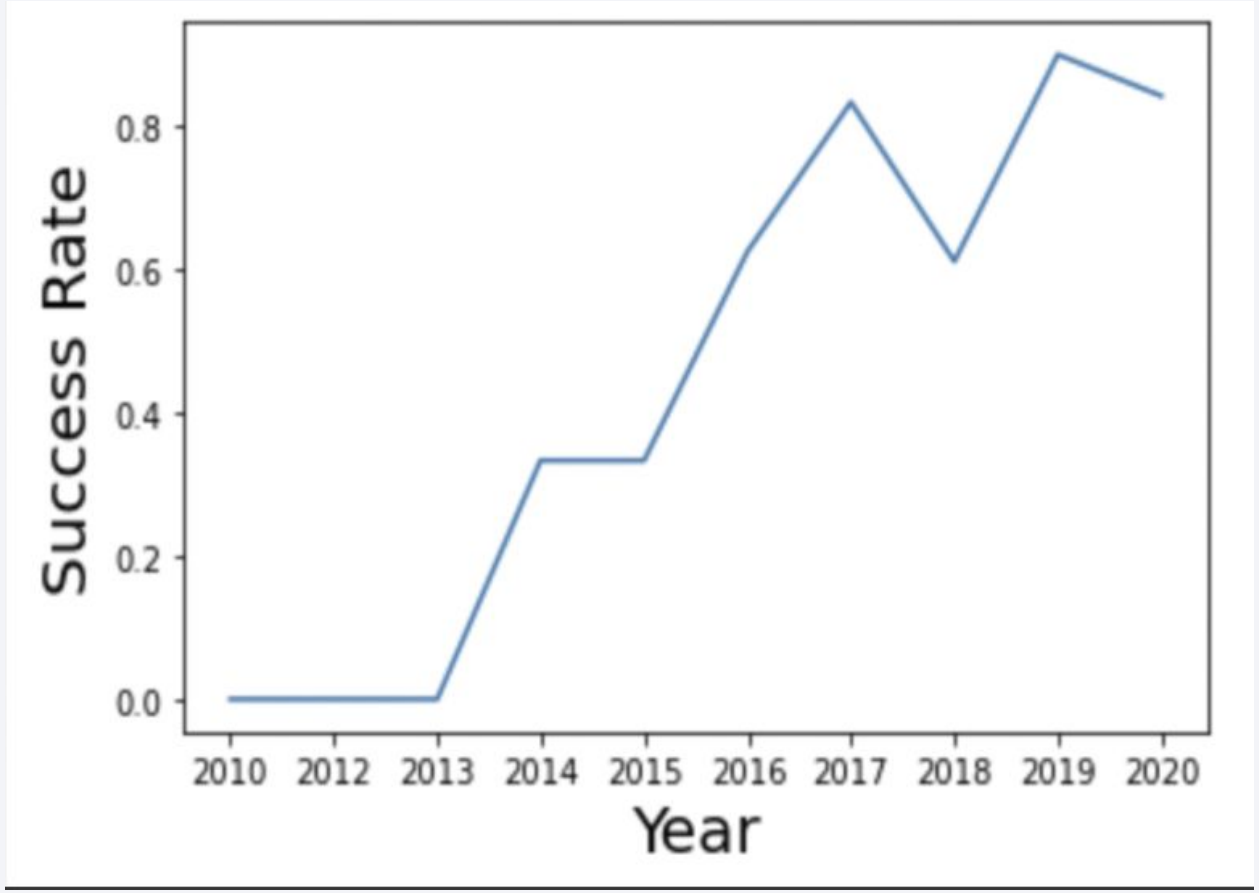
Payload vs. Orbit Type



- There seems to be no correlation between payload mass and success rate for GTO
- Higher payload mass seems to yield higher success rate for ISS

Launch Success Yearly Trend

- The success rate from 2013 to 2020 shows an upward trend



All Launch Site Names

Task 1

Display the names of the unique launch sites in the space mission

```
[11]: %sql SELECT DISTINCT "Launch_Site" FROM SPACEXTABLE;
```

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

[11]:

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

- The name of unique launch sites in the space mission

Launch Site Names Begin with 'CCA'

Task 2

Display 5 records where launch sites begin with the string 'CCA'

[12]:
```
%sql SELECT * FROM SPACEXTABLE WHERE "Launch_Site" LIKE 'CCA%' LIMIT 5;
```

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

[12]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcor
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachu
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 (parachu
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No atten
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No atten
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No atten

- 5 records where launch sites begin with `CCA`

Total Payload Mass

Task 3

Display the total payload mass carried by boosters launched by NASA (CRS)

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

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

SUM("PAYLOAD_MASS__KG_")
45596

- The total payload mass carried by boosters launched by NASA (CRS) is 45596 kgs

Average Payload Mass by F9 v1.1

Task 4

Display average payload mass carried by booster version F9 v1.1

```
[14]: %sql SELECT AVG("PAYLOAD_MASS__KG_") FROM SPACEXTABLE WHERE "Booster_Version" = 'F9 v1.1';
```

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

[14]:

AVG("PAYLOAD_MASS__KG_")
2928.4

- The average payload mass carried by booster version F9 v1.1 is 2928.4

First Successful Ground Landing Date

Task 5

List the date when the first succesful landing outcome in ground pad was acheived.

Hint:Use min function

[15]:
```
%sql SELECT MIN("Date") FROM SPACEXTABLE WHERE "Landing_Outcome" = 'Success (ground pad)';
```

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

[15]:

MIN("Date")
2015-12-22

- The date of the first successful landing outcome in ground pad was 12/22/2015

Successful Drone Ship Landing with Payload between 4000 and 6000

Task 6

List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

```
%sql SELECT DISTINCT "Booster_Version" FROM SPACEXTABLE WHERE "Landing_Outcome" = 'Success (drone ship)' AND "PAYLOAD_
```

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

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

- The names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

Total Number of Successful and Failure Mission Outcomes

Task 7

List the total number of successful and failure mission outcomes

```
19]: %sql SELECT "Mission_Outcome", COUNT(*) AS "Total" FROM SPACEXTABLE WHERE "Mission_Outcome" IN ('Success', 'Failure')

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

19]:

Mission_Outcome	Total
Success	98

- The total number of successful and failure mission outcomes

Boosters Carried Maximum Payload

Task 8

List all the booster_versions that have carried the maximum payload mass. Use a subquery.

```
[20]: %sql SELECT DISTINCT "Booster_Version" FROM SPACEXTABLE WHERE "PAYLOAD_MASS__KG_" = (SELECT MAX("PAYLOAD_MASS__KG_") FR
```

Collapse Output

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

[20]:

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

- The names of the booster which have carried the maximum payload mass

2015 Launch Records

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

[21]:

Month_Name	Mission_Outcome	Booster_Version	Launch_Site
January	Success	F9 v1.1 B1012	CCAFS LC-40
February	Success	F9 v1.1 B1013	CCAFS LC-40
March	Success	F9 v1.1 B1014	CCAFS LC-40
April	Success	F9 v1.1 B1015	CCAFS LC-40
April	Success	F9 v1.1 B1016	CCAFS LC-40
June	Failure (in flight)	F9 v1.1 B1018	CCAFS LC-40
December	Success	F9 FT B1019	CCAFS LC-40

- The failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

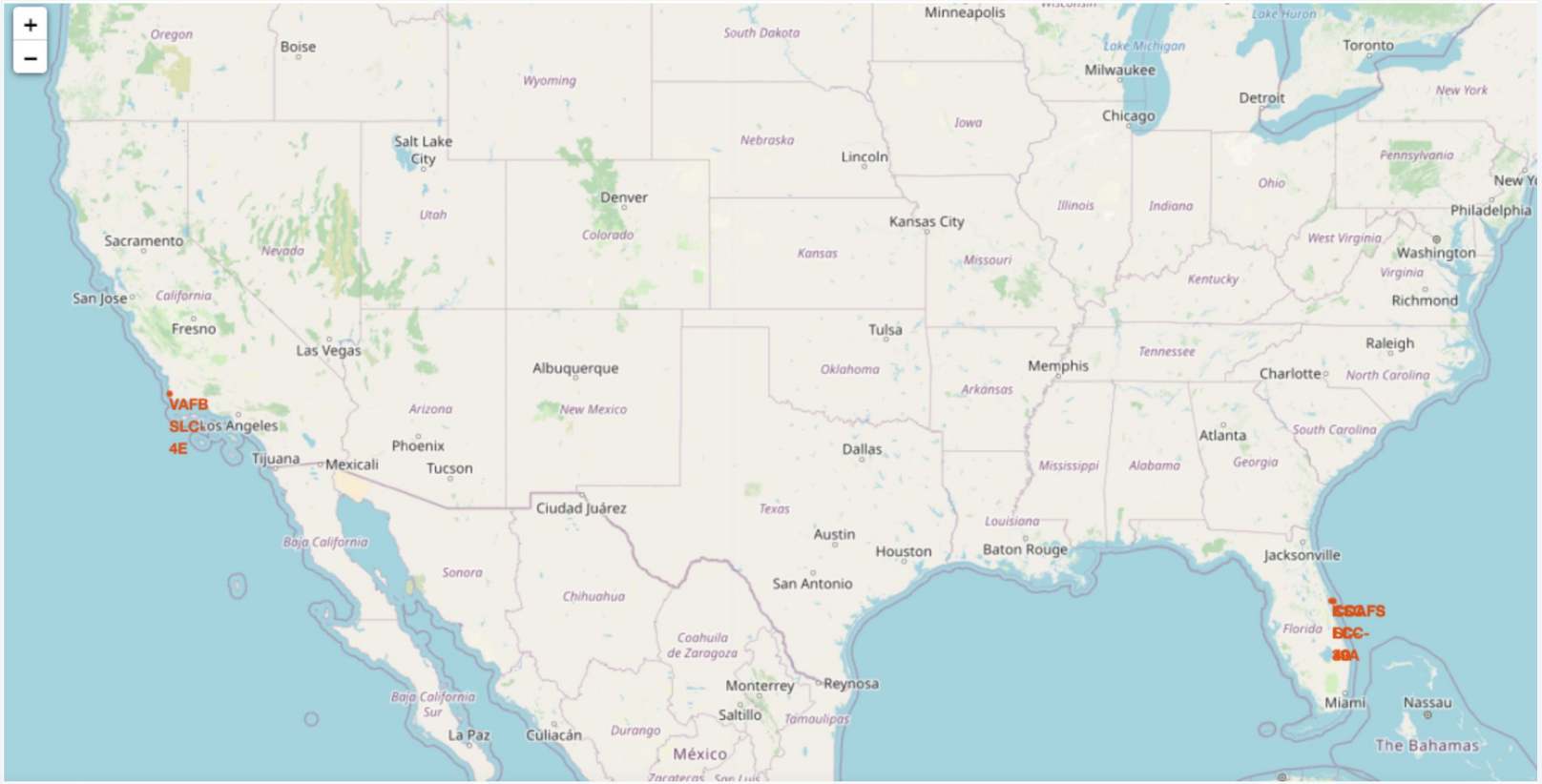
Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

Landing_Outcome	Count
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

- 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

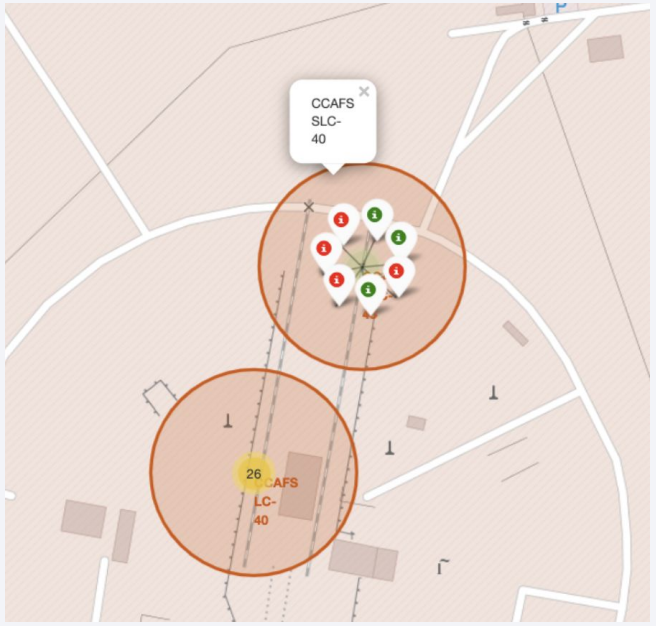

Section 3
Launch Sites
Proximities Analysis

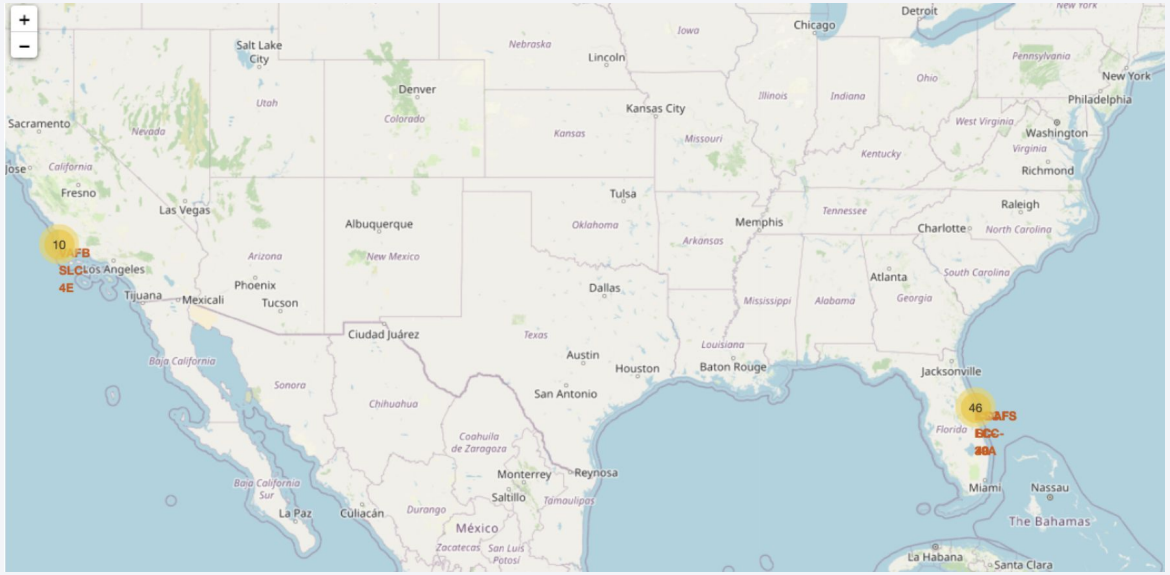
All launch sites on global map



- A few launch sites, all near the coasts, most likely due to safety reasons

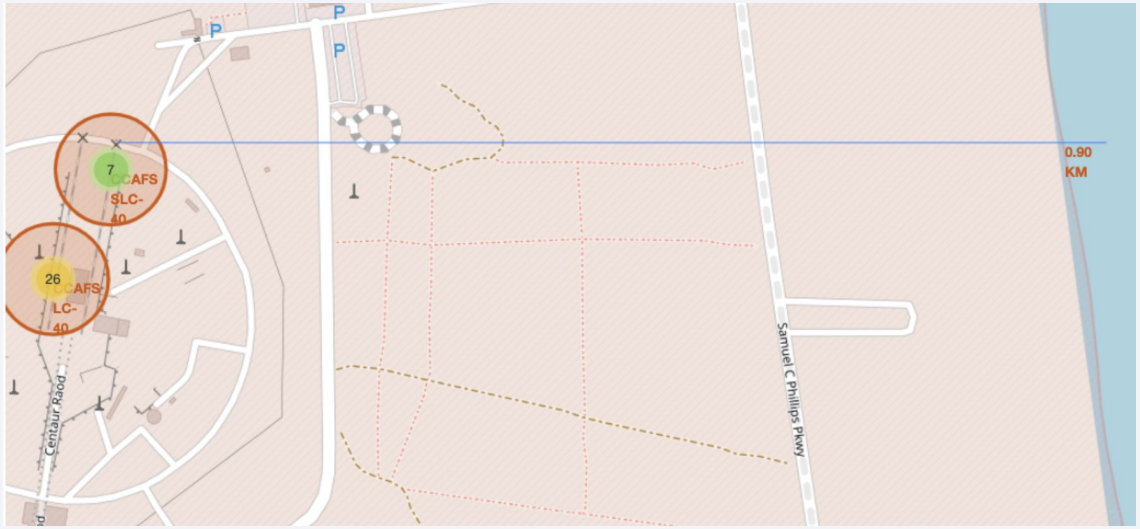
Color coded outcome labels

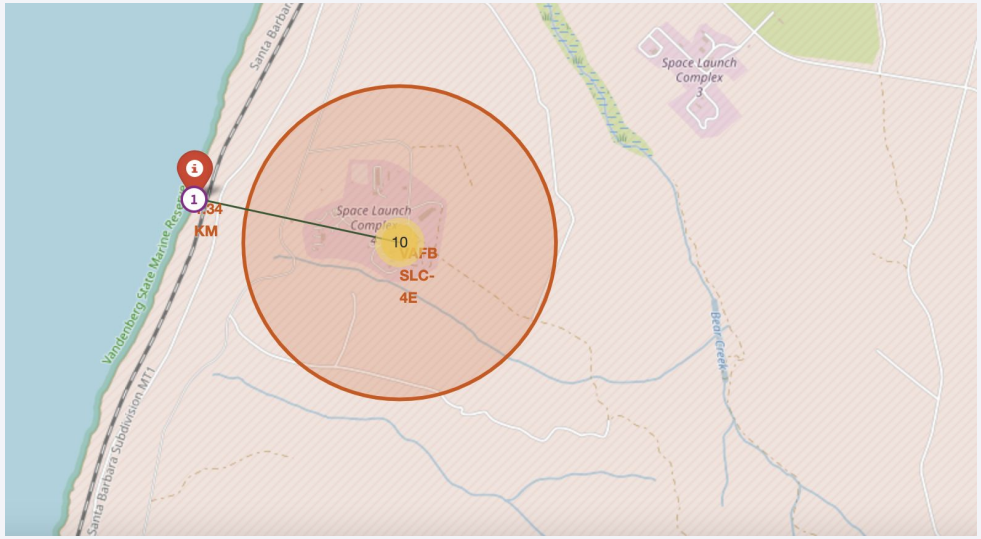




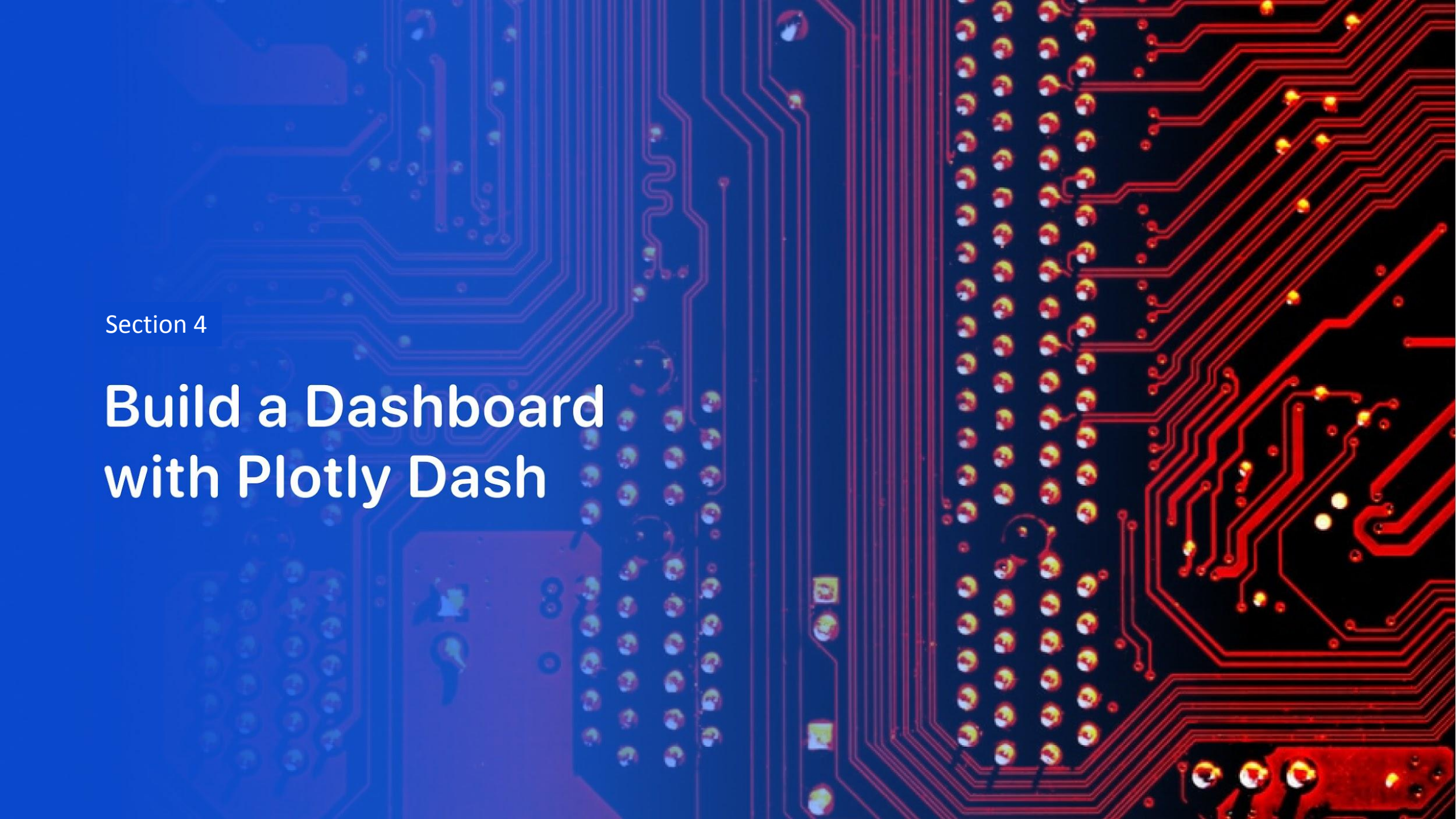
- Shows launch outcomes for each site using red (unsuccessful) and green (successful)

Launch site proximity

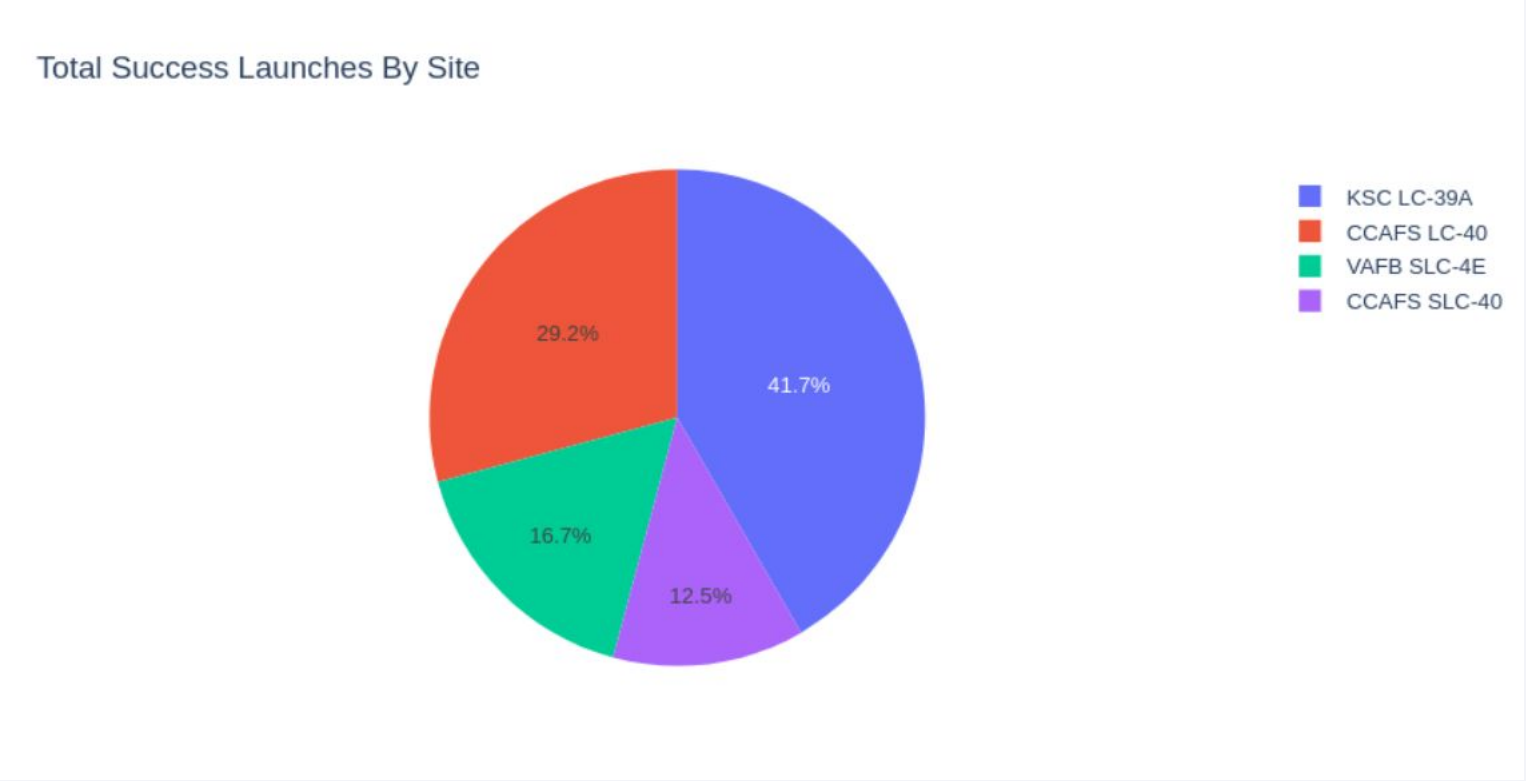




- Launch sites are not in proximity to highways and railways
- Launch sites are in proximity to the coastline
- Launch sites keep distance from cities

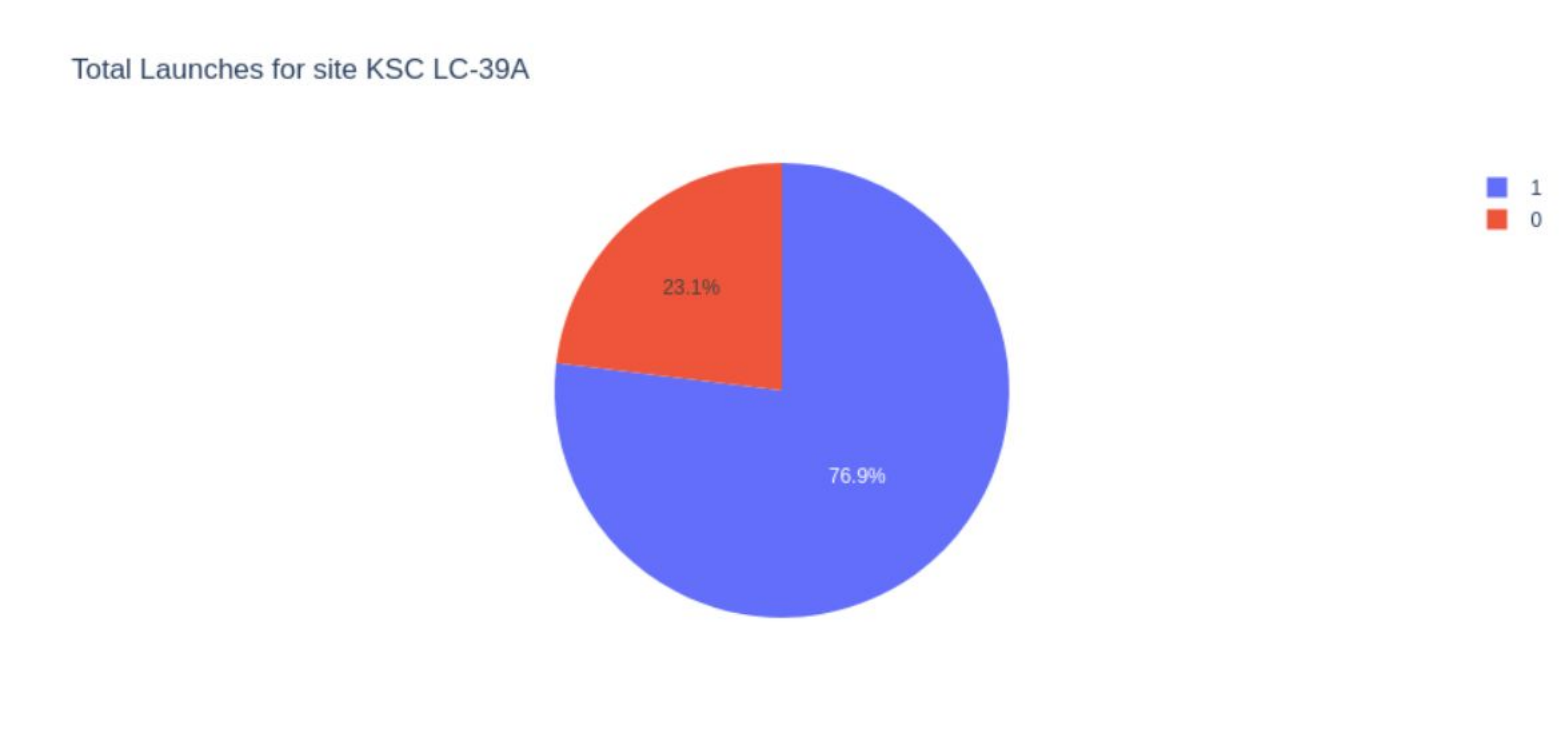

Section 4
Build a Dashboard with Plotly Dash

Success Launches by Site



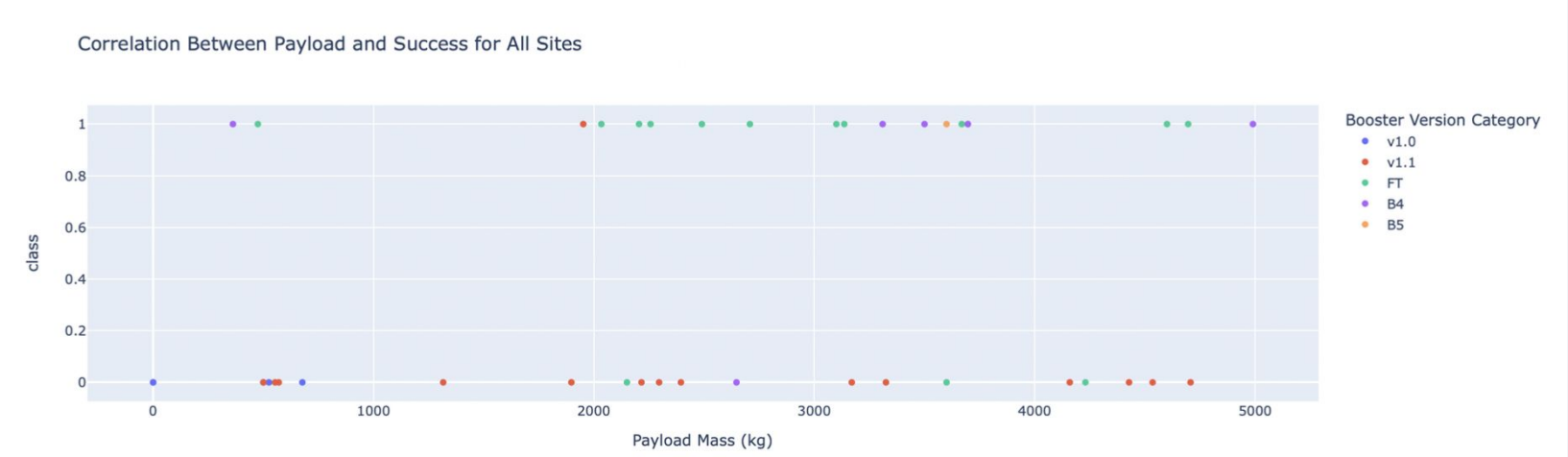
- A pie chart of success launches by site, KSC LC-39A has the most successful launches

Launch site with highest success ratio



- KSC LC-39A has a success rate of 76.9%

Payload Mass vs Launch Outcome



- Payload mass of about 2000 kg to 6000 kg have the highest success rate

Section 5

Predictive Analysis (Classification)

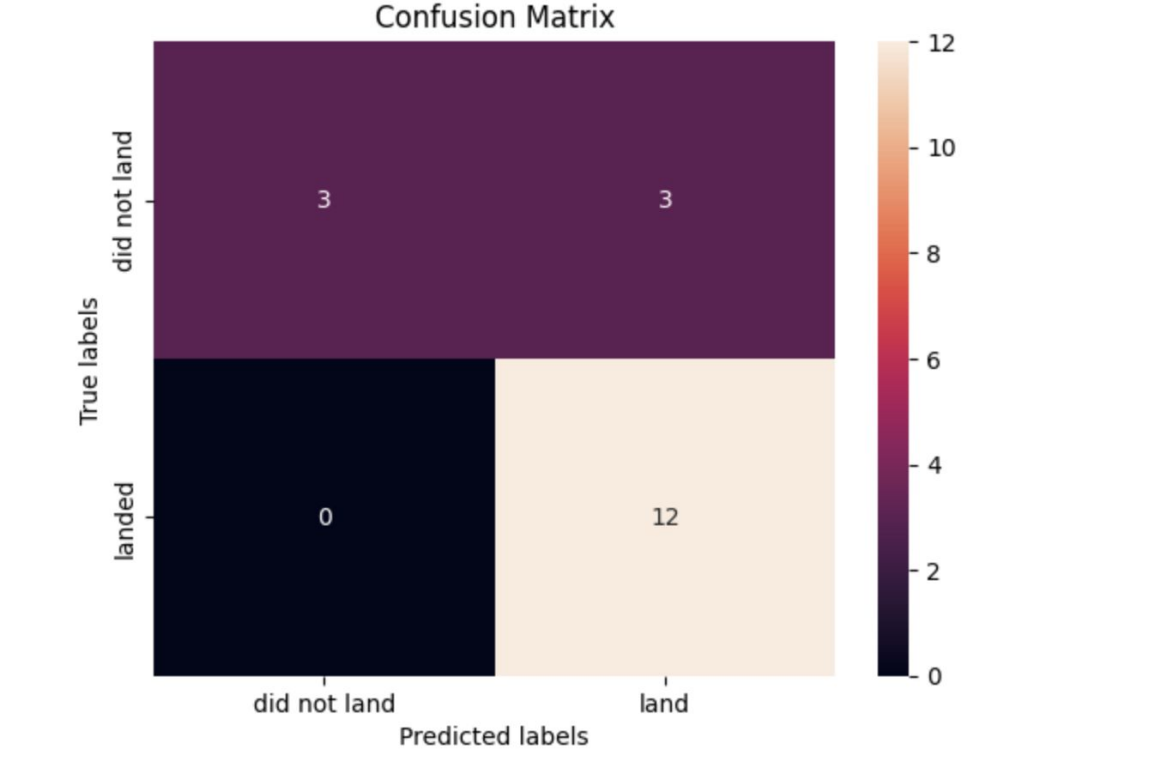
Classification Accuracy

- The model with the highest classification accuracy was the decision tree model with 87% accuracy

[45]:

	LogReg	SVM	Tree	KNN
Jaccard_Score	0.833333	0.845070	0.805556	0.819444
F1_Score	0.909091	0.916031	0.892308	0.900763
Accuracy	0.866667	0.877778	0.844444	0.855556

Confusion Matrix



- The confusion matrix for the decision tree classifier shows that the main issue that is hindering the accuracy are the false positives

Conclusions

- The decision tree was the best model for the dataset
- As flight number increases, the overall success rate increases as well
- Launch success rate increased from 2013 to 2020
- The most successful launch site is KSC LC-39A
- Orbits ES-L1, GEO, HEO, and SSO have the best success rate of 100%
- All launch sites are near the coast and avoid cities, highways, and railways

Thank you!