IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Muhammad Zulfikar Muslim
2 January 2024

Outline

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

Executive Summary

- Summary of methodologies
  - Collecting data from public SpaceX API and SpaceX Wikipedia Page.
  - Converting mission outcomes into Training Labels with `1` means the booster successfully landed `0` means it was unsuccessful during data wrangling.
  - Performing some Exploratory Data Analysis (EDA) to find some patterns in data using SQL and Data Visualization.
  - Performing more interactive visual analytics using Folium and creating dashboard.
  - Preprocessing data to create dummy variables to categorical columns, standardize the data in X, and splitting data
  - Using GridSearchCV to find best parameters for machine learning models.
- Summary of all results
  - Four machine learning models were produced: Logistic Regression, Support Vector Machine, Decision Tree Classifier, and K Nearest Neighbors.
  - All produced similar results with accuracy rate about 83.33%.

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
  - Predict if the Falcon 9 first stage will land successfully

Section 1

Methodology

Methodology

Executive Summary

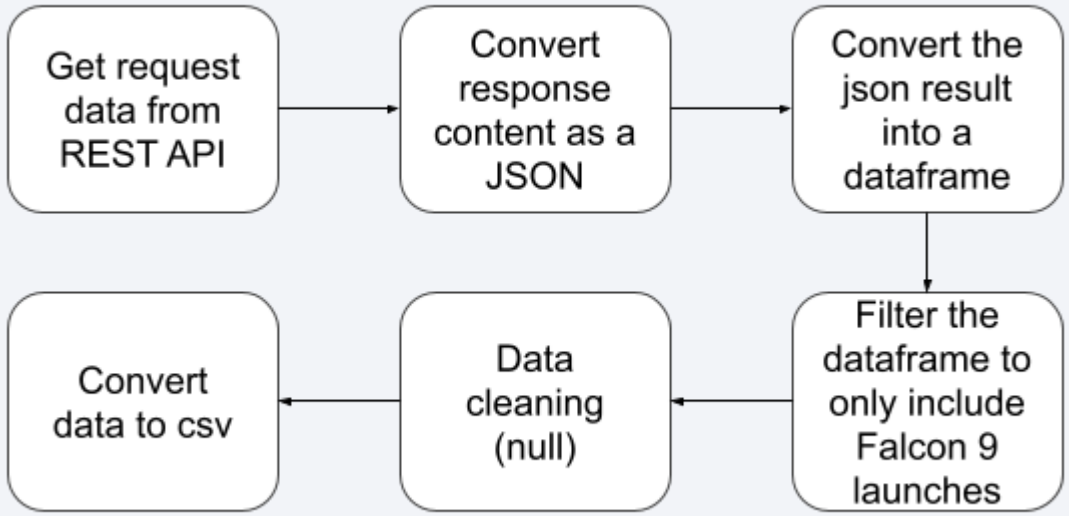
- Data collection methodology:
  - Rest API and Web Scraping
- Perform data wrangling
  - Data are clean of null values and irrelevant columns
  - Categorical columns transformed into one Hot Encoding to be applied to Machine Learning.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Use Logistic Regression, KNN, SVM, and Decision Tree for classification and evaluate its accuracy for the best classifier

Data Collection

- Data collection process involved a combination of API requests from Space X public API and web scraping data from a table in Space X's Wikipedia entry
- API request from https://api.spacexdata.com/v4/
- Web scraping from https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922

Data Collection – SpaceX API

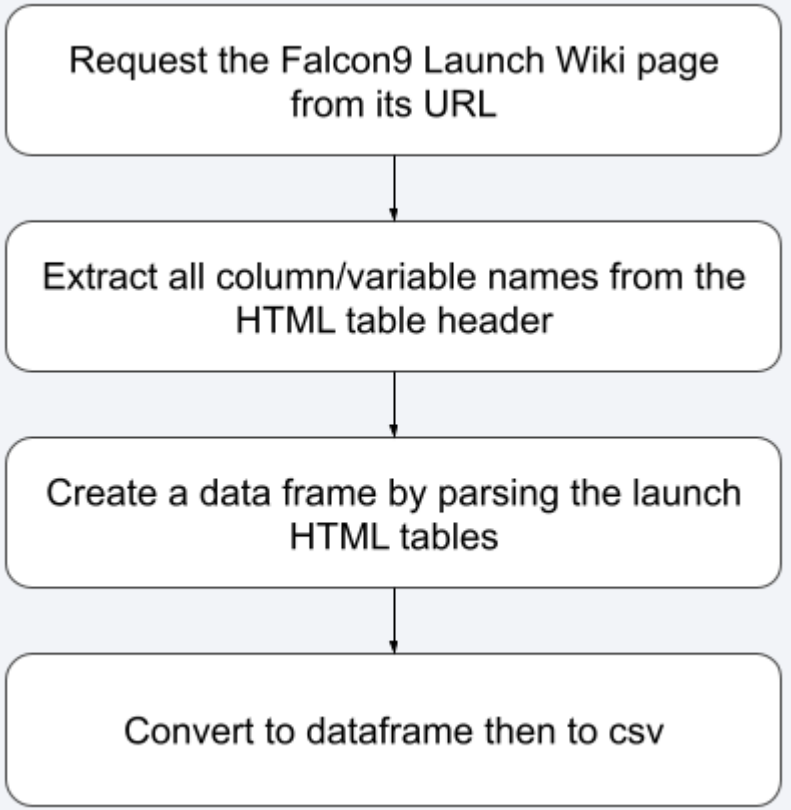
- Launch data is collected from SpaceX REST API
- Github URL: Data Collection - API calls



Data Collection - Scraping

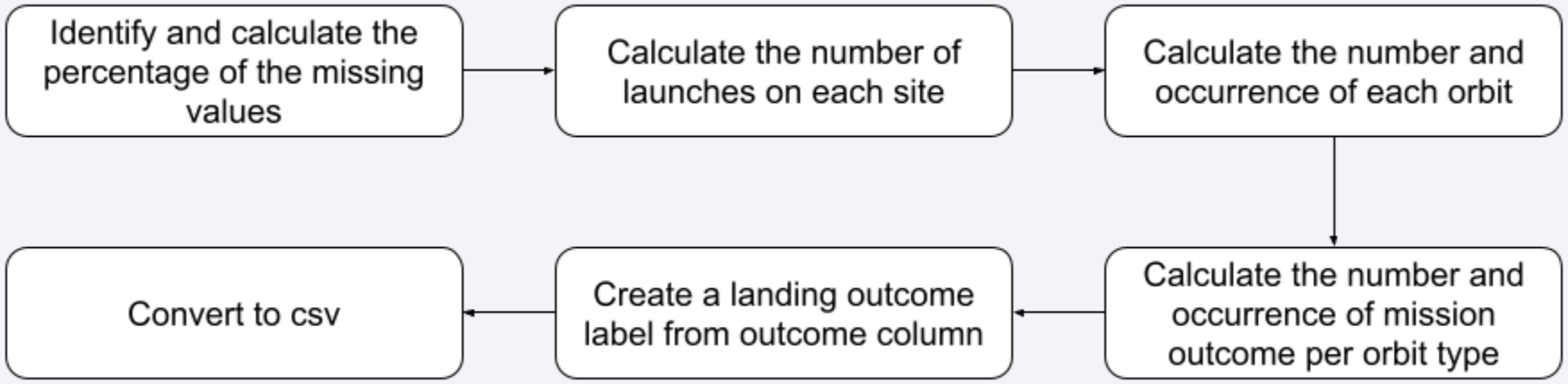
- Web scraping data from a table in Space X's Wikipedia entry

- Github URL: Data Collection – Web scraping



Data Wrangling

- Data were processed to ensure the data are clean of null values and label are ready for analysis and modeling.



- Github URL: Data Wrangling

EDA with Data Visualization

- Plotted charts

Chart	Purpose
Scatter plot Flight Number vs. Launch Site	Visualize the relationship between Flight Number and Launch Site
Scatter plot Payload vs. Launch Site	Visualize the relationship between Payload and Launch Site
Bar chart Success Rate vs. Orbit Type	Visualize the relationship between success rate of each orbit type
Scatter plot Flight Number vs. Orbit Type	Visualize the relationship between Flight Number and Orbit type
Scatter plot Payload vs. Orbit Type	Visualize the relationship between Payload and Orbit type
Line chart Class and Date	Visualize the launch success yearly trend

- Github URL: EDA – Data Vis

EDA with SQL

- SQL queries performed
  - Display the names of the unique launch sites in the space mission
  - Display 5 records where launch sites begin with the string 'CCA'
  - Display the total payload mass carried by boosters launched by NASA (CRS)
  - Display average payload mass carried by booster version F9 v1.1
  - List the date when the first succesful landing outcome in ground pad was acheived.
  - List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
  - List the total number of successful and failure mission outcomes
  - List the names of the booster_versions which have carried the maximum payload mass
  - List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015
  - Rank 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.
- Github URL: EDA – SQL

Build an Interactive Map with Folium

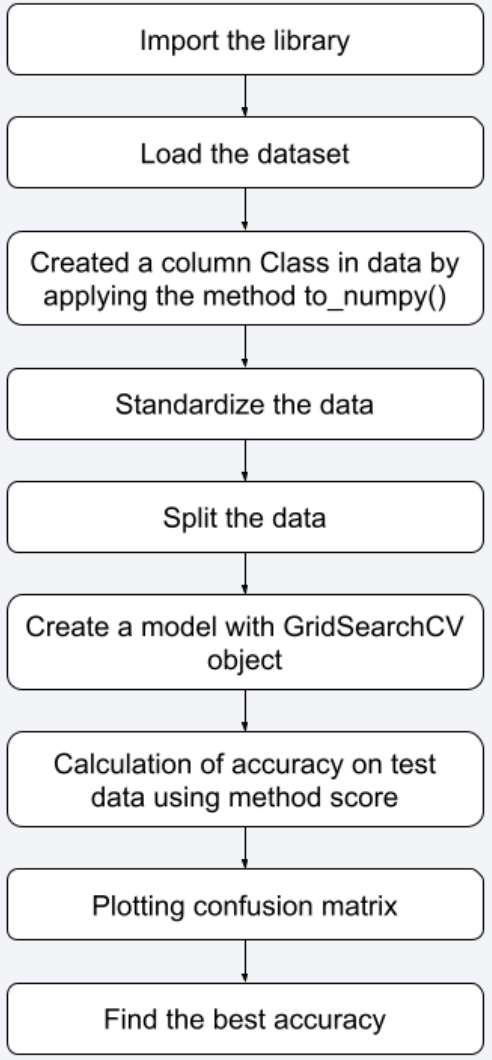
- Folium maps mark
  - Launch Sites,
  - successful and unsuccessful landings, and
  - a proximity example to key locations: Railway, Highway, Coast, and City.
- This allows us to understand why launch sites may be located where they are and also visualizes successful landings relative to location.
- Github URL: Folium

Build a Dashboard with Plotly Dash

- Dashboard includes:
  - Pie chart
    - can be selected to show distribution of successful landings across all launch sites
    - or can be selected to show individual launch site success rates.
  - Scatter plot
    - All sites or individual site
    - payload mass on a slider between 0 and 10000 kg.
- The pie chart is used to visualize launch site success rate while the scatter plot can help us see how success varies across launch sites, payload mass, and booster version category.
- Github URL: Plotly – Dash

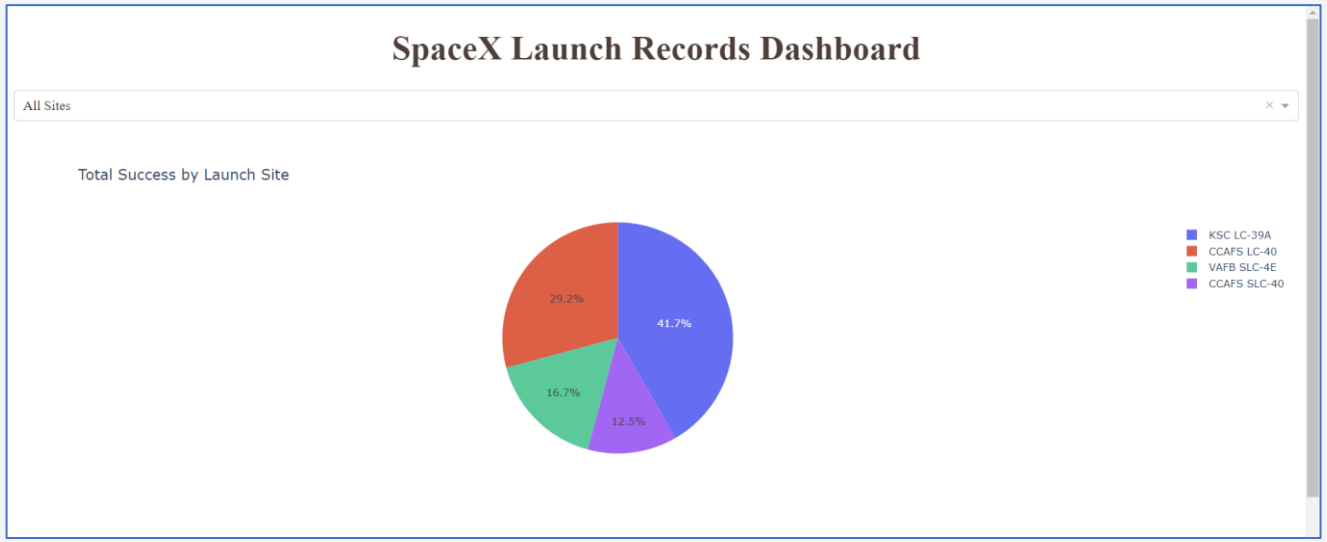
Predictive Analysis (Classification)

- Summary:
  - Standardize: StandardScaler()
  - Test size: 20%
  - Random state: 2
  - Cross validation fold: 10
  - Model: Logistic Regression, SVM, Decision Tree, KNN
  - Metric: Accuracy
- Github URL: Modeling

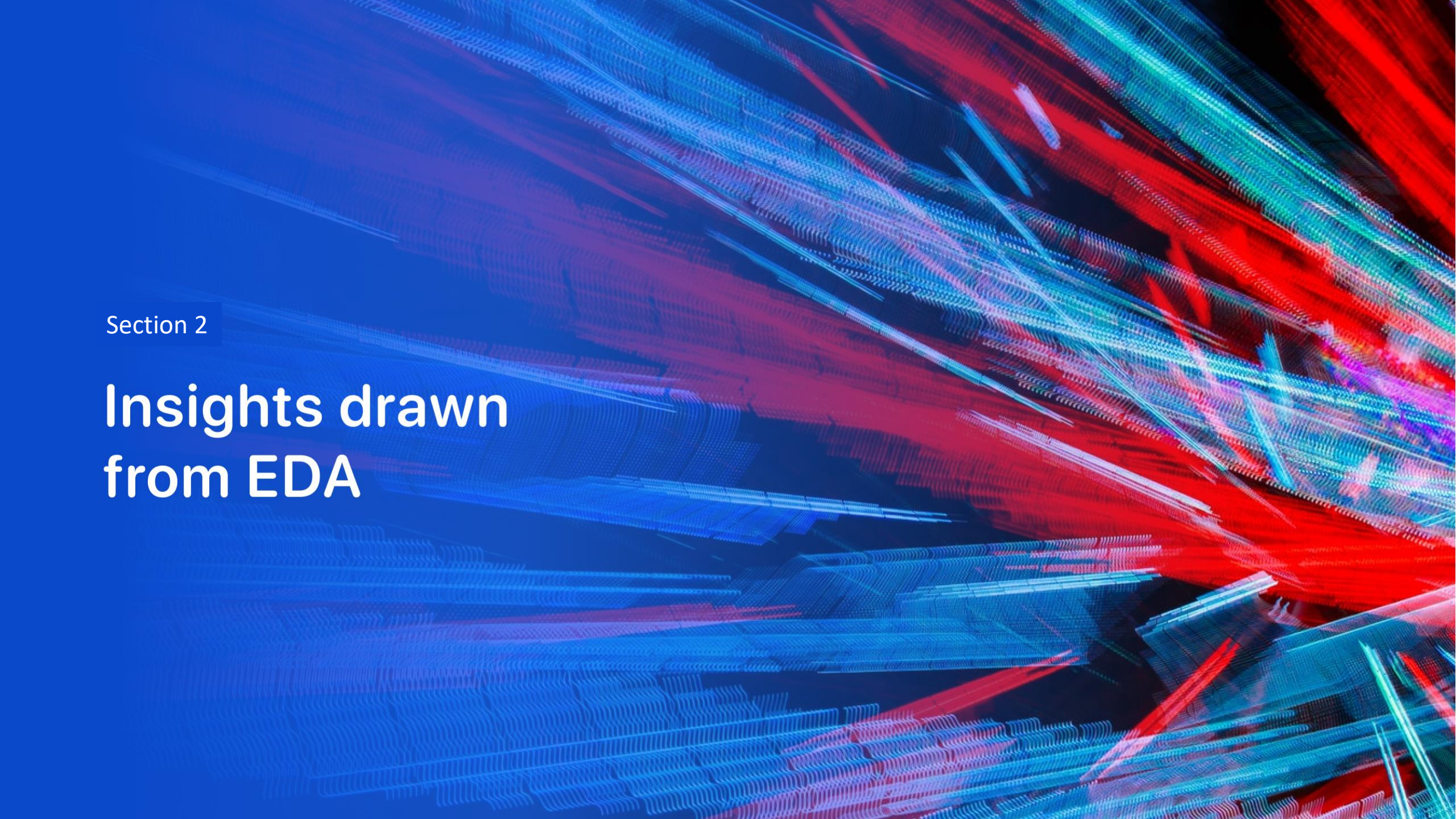


Results

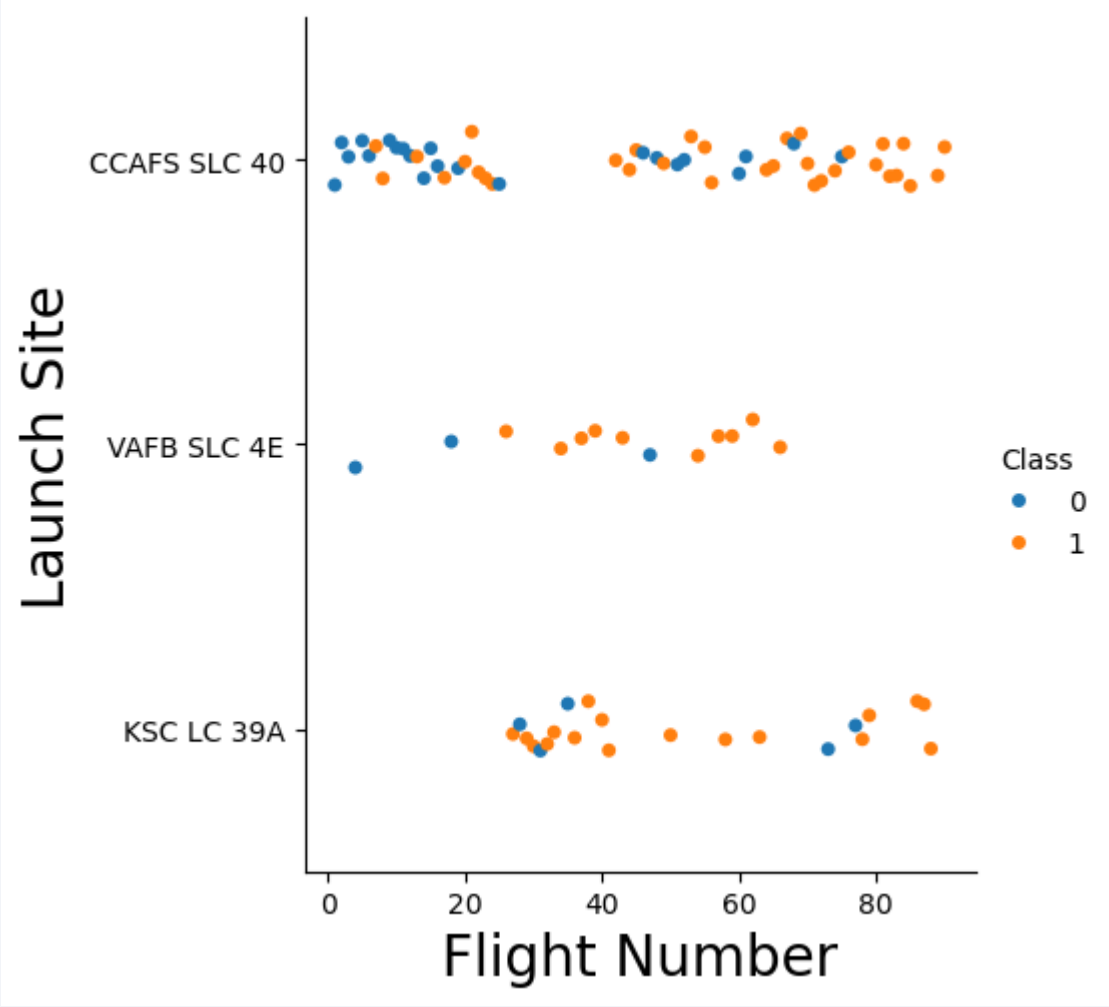
- CCAFS SLC-40 site and KSC LC-39A site has most successful launches from all the sites.



- All of the models perform equally well with the same accuracy score of around 83.33%.

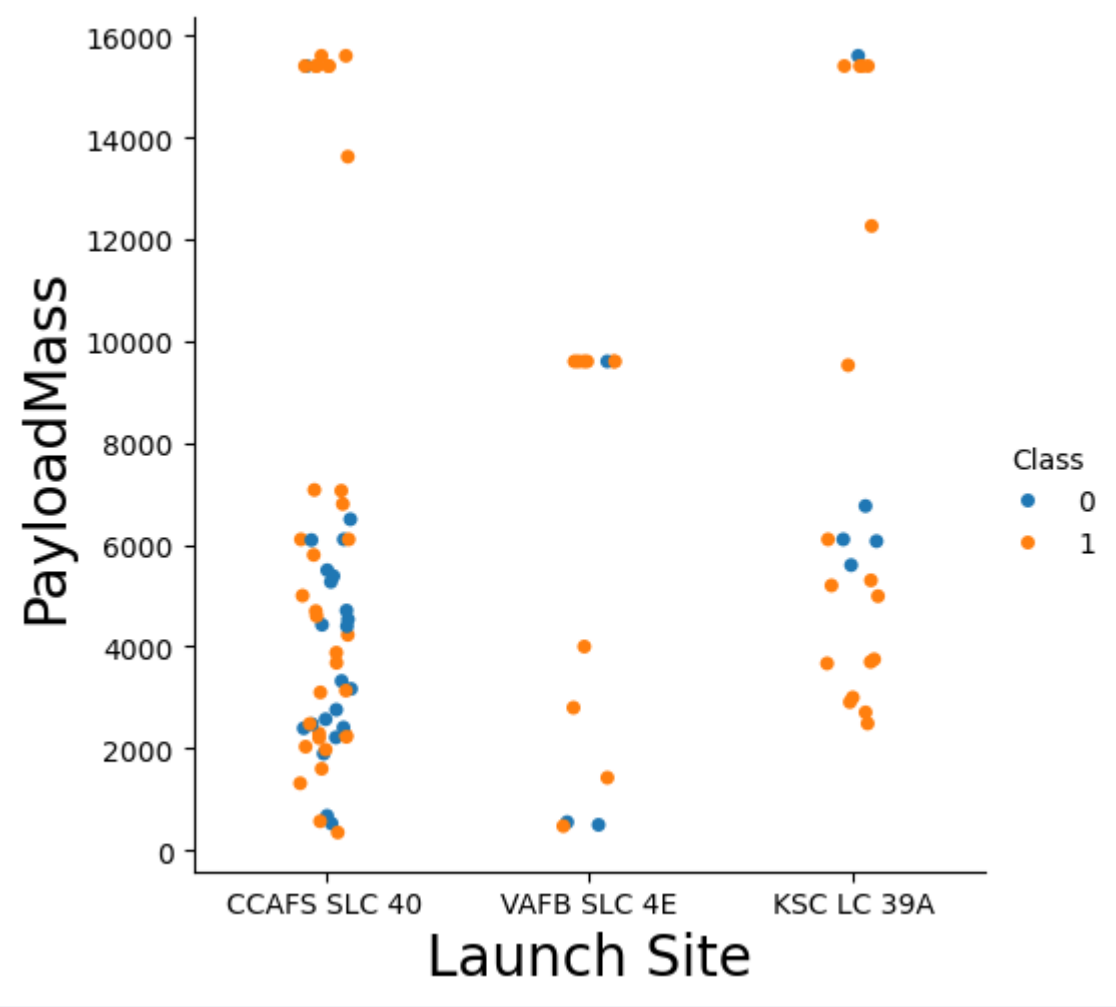

Section 2
Insights drawn from EDA

Flight Number vs. Launch Site



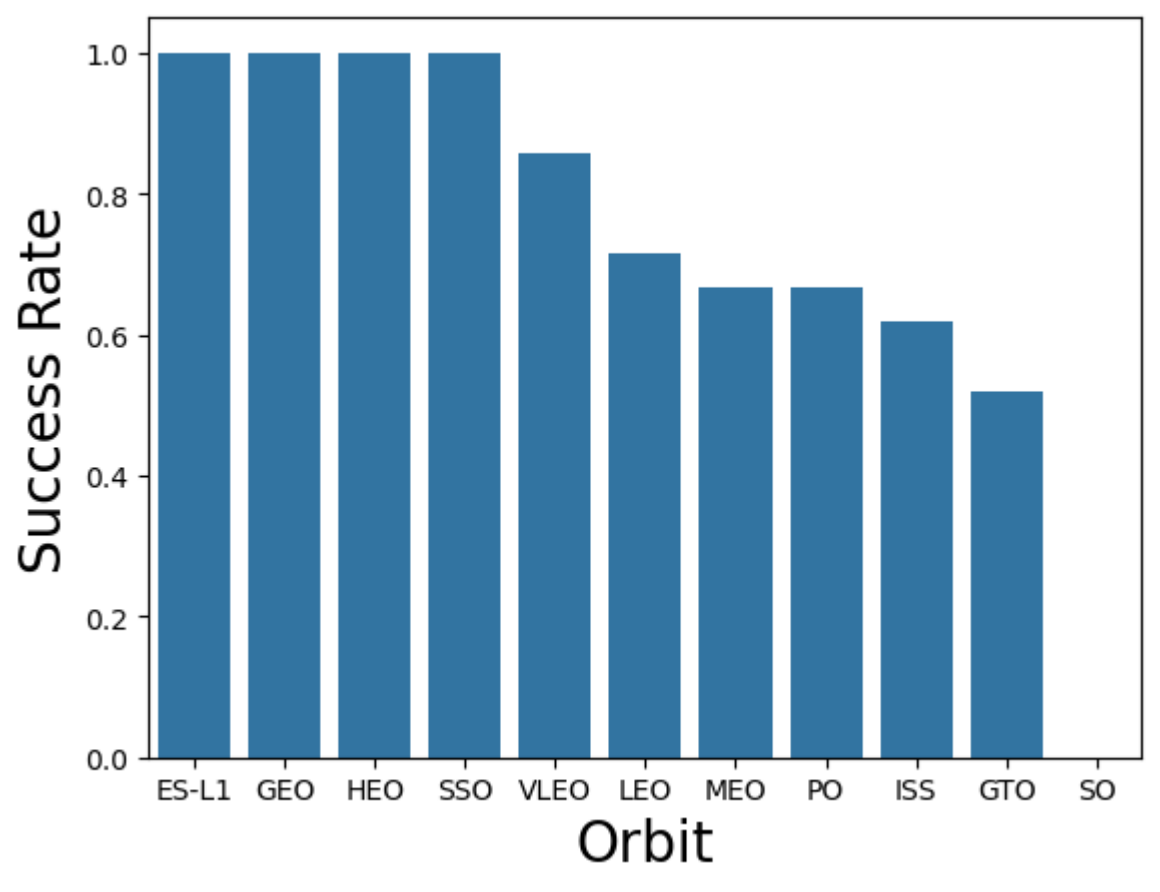
- CCAFS SLC 40 has the highest number of launch.
- Different launch sites have different success rates.
- But as we increase the number of flights the success rate increase.
- VAFB SLC 4E has the smallest number of failed launch

Payload vs. Launch Site



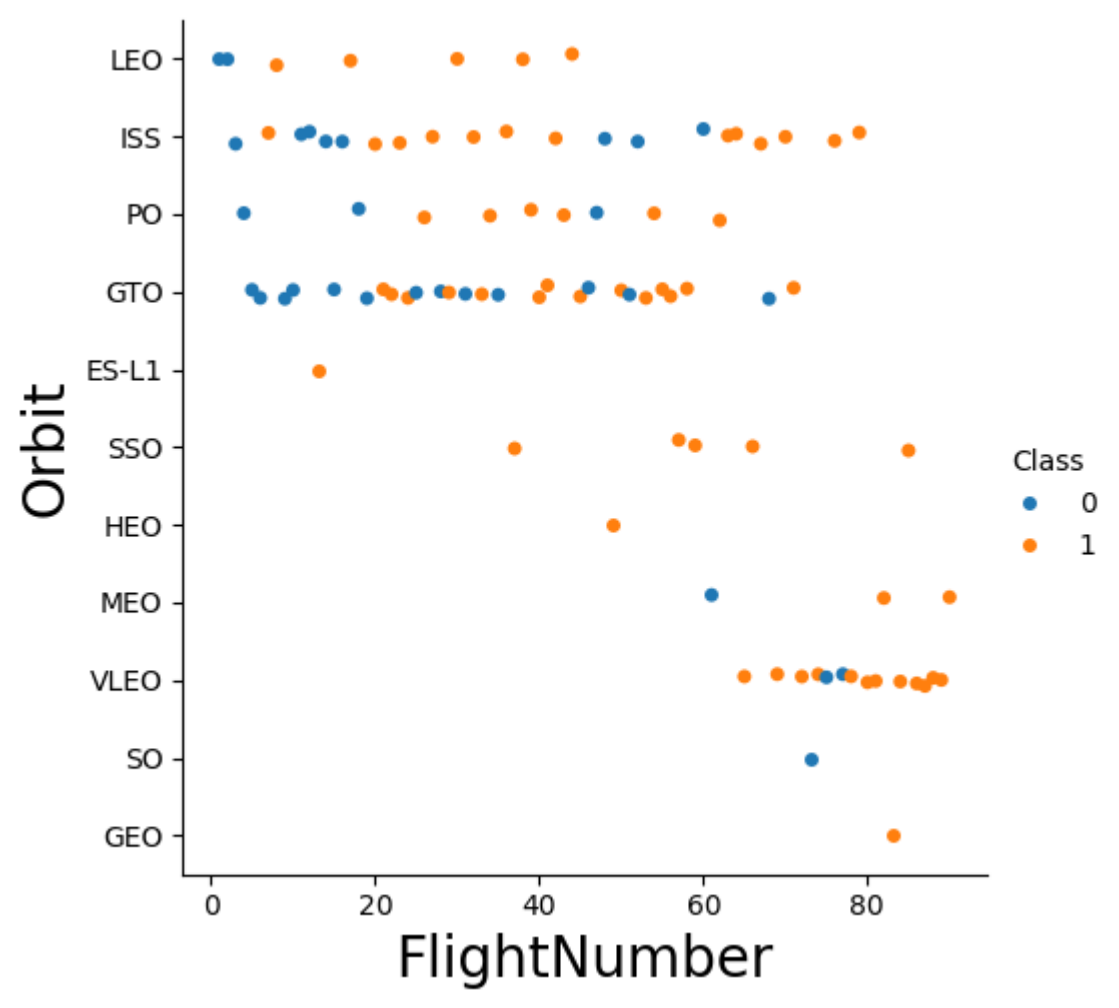
- Rockets launched with heavypayload mass (greater than 10000) tends to success with only 1 failed launch.
- There are no rockets launched with heavypayload mass in the VAFB-SLC launchsite.

Success Rate vs. Orbit Type



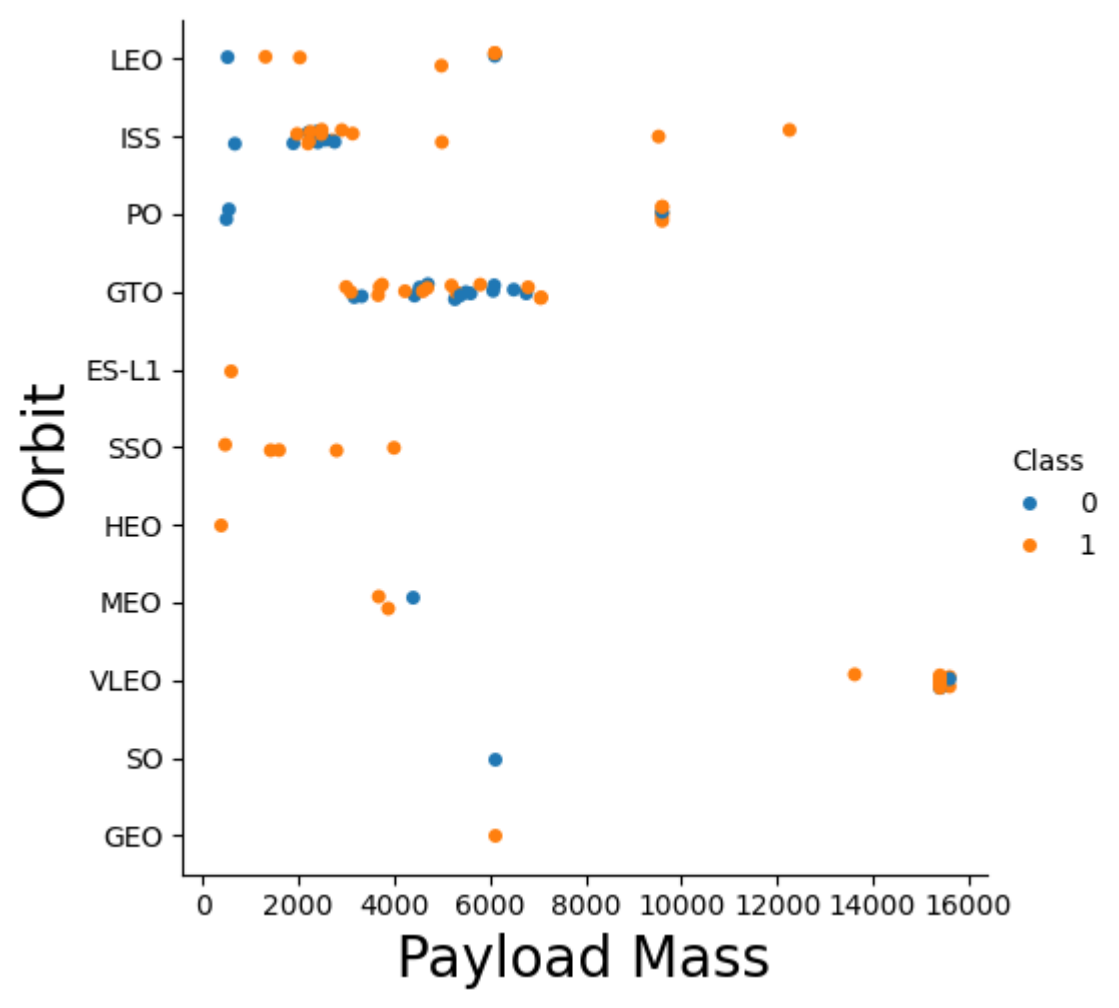
- ES-L1, GEO, HEO, and SSO orbit have 100% success rate
- So has 0% success rate.

Flight Number vs. Orbit Type



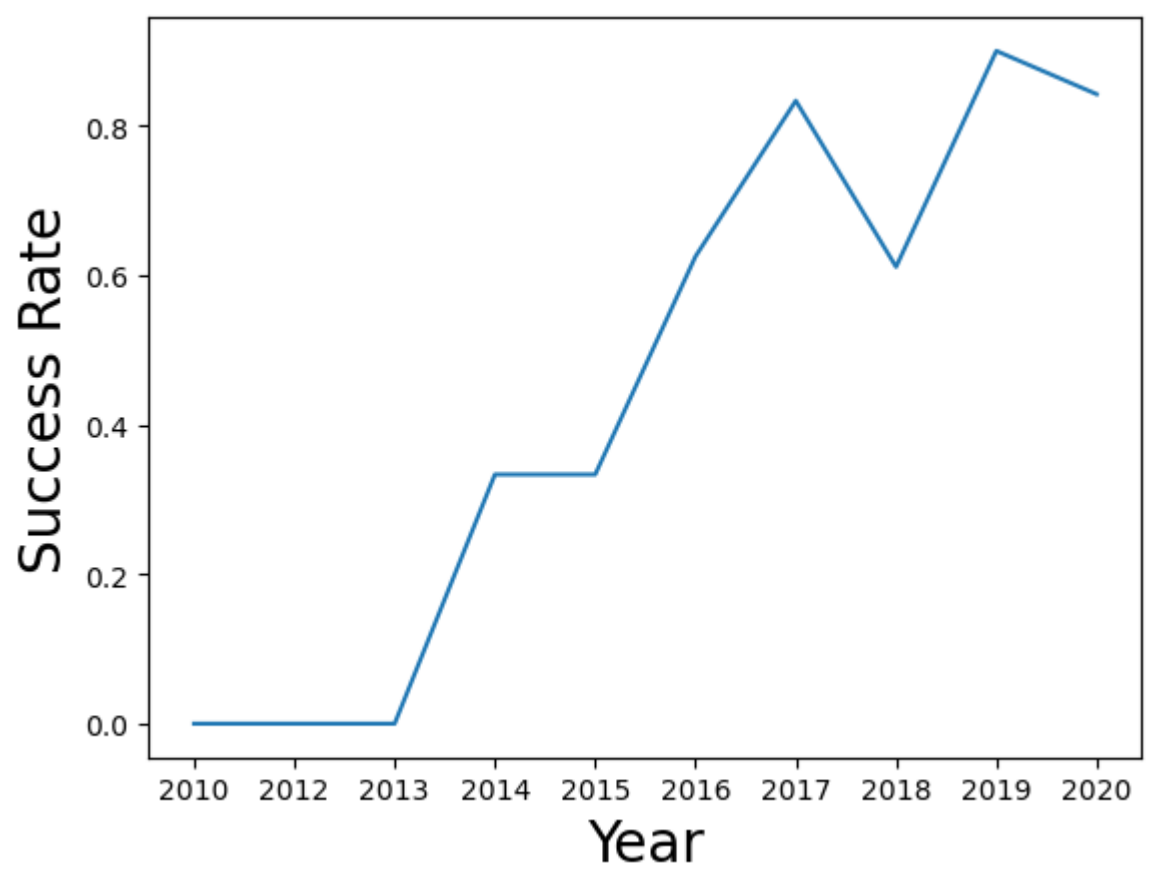
- In the LEO orbit the Success appears related to the number of flights.
- On the other hand, there seems to be no relationship between flight number when in GTO orbit.

Payload vs. Orbit Type



- With heavy payloads the successful landing or positive landing rate are more for Orbit 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



- The success rate since 2013 kept increasing till 2020

All Launch Site Names

- Find the names of the unique launch sites

```
%sql SELECT DISTINCT("Launch_Site") FROM SPACEXTBL;
```

Python

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

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

- There are four unique launch sites : CCAFS LC-40, VAFB SLC-4E, KSC LC-39A, dan CCAFS SLC-40

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`

```
%%sql SELECT * FROM SPACEXTBL
WHERE "Launch_Site" LIKE 'CCA%'
limit 5;
```

Python

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

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

Total Payload Mass

- Calculate the total payload carried by boosters from NASA

```
%%sql SELECT sum("PAYLOAD_MASS__KG_") AS total_payload_mass_by_NASA_CRS FROM SPACEXTBL
WHERE Customer  = 'NASA (CRS)'
```

Python

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

total_payload_mass_by_NASA_CRS
45596

- The total payload carried by boosters from NASA is about 45596 kg.

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1

```
%%sql SELECT AVG("PAYLOAD_MASS__KG_") AS "avg_payload_mass_by_F9_v1.1" FROM SPACEXTBL
WHERE Booster_Version = 'F9 v1.1'
```

Python

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

avg_payload_mass_by_F9_v1.1
2928.4

- The average payload mass carried by boosters version F9 v1.1 is about 2928.4 kg.

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad

```
%%sql SELECT MIN("Date") AS "first_succesful_landing_outcome_in_ground_pad" FROM SPACEXTBL
WHERE Landing_Outcome = "Success (ground pad)";
```

Python

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

first_succesful_landing_outcome_in_ground_pad
2015-12-22

- The first successful landing outcome on ground pad is on 22 December 2015.

Successful Drone Ship Landing with Payload between 4000 and 6000

- List 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_ > 4000 AND PAYLOAD_MASS__KG_ < 6000;
```

Python

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

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

- There are 4 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

- Calculate the total number of successful and failure mission outcomes

```
%sql SELECT COUNT("Mission_Outcome") FROM SPACEXTBL;
```

Python

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

COUNT("Mission_Outcome")
101

- There are 101 mission.

Boosters Carried Maximum Payload

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

```
%%sql SELECT Booster_Version, PAYLOAD_MASS__KG_ FROM SPACEXTBL
WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL);
```

Python

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

Booster_Version	PAYLOAD_MASS__KG_
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

2015 Launch Records

- List 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, substr(Date,0,5) as year, "Landing_Outcome", "Booster_Version", "Launch_Site"
WHERE substr(Date,0,5)='2015' AND Landing_Outcome = "Failure (drone ship)";
```

Python

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

month	year	Landing_Outcome	Booster_Version	Launch_Site
01	2015	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	2015	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

- There are 2 failed landing outcomes in drone ship in year 2015.

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

- Rank 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

```
%%sql SELECT Landing_Outcome, COUNT(*) AS Count_of_Landing_Outcome FROM SPACEXTBL
WHERE Date BETWEEN '2010-06-04' AND '2017-03-20'
GROUP BY Landing_Outcome
ORDER BY 2 DESC
```

Python

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

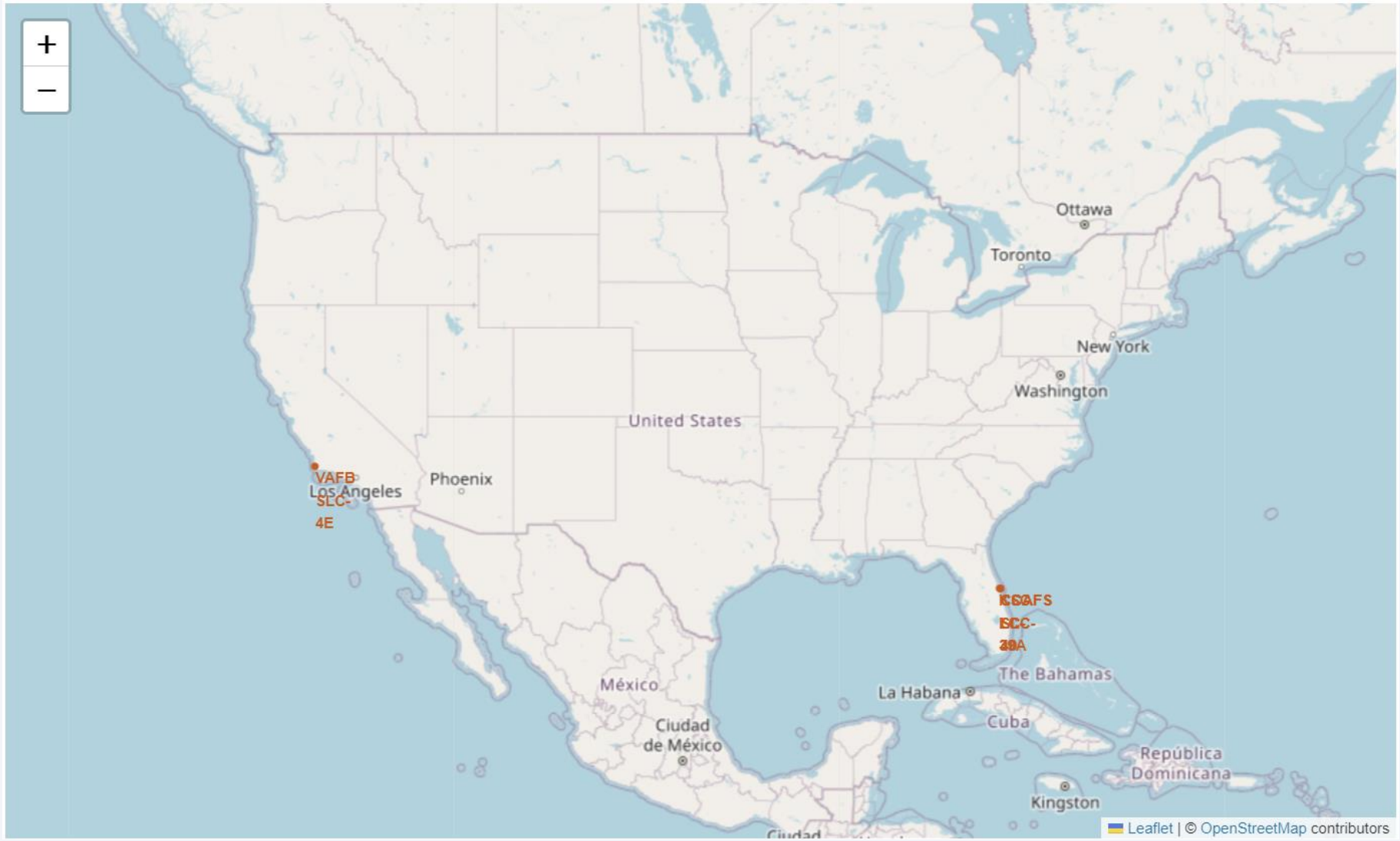
Landing_Outcome	Count_of_Landing_Outcome
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

- Failure (drone ship) and Success (drone ship) is the most landing outcome between the date 2010-06-04 and 2017-03-20.

Section 3

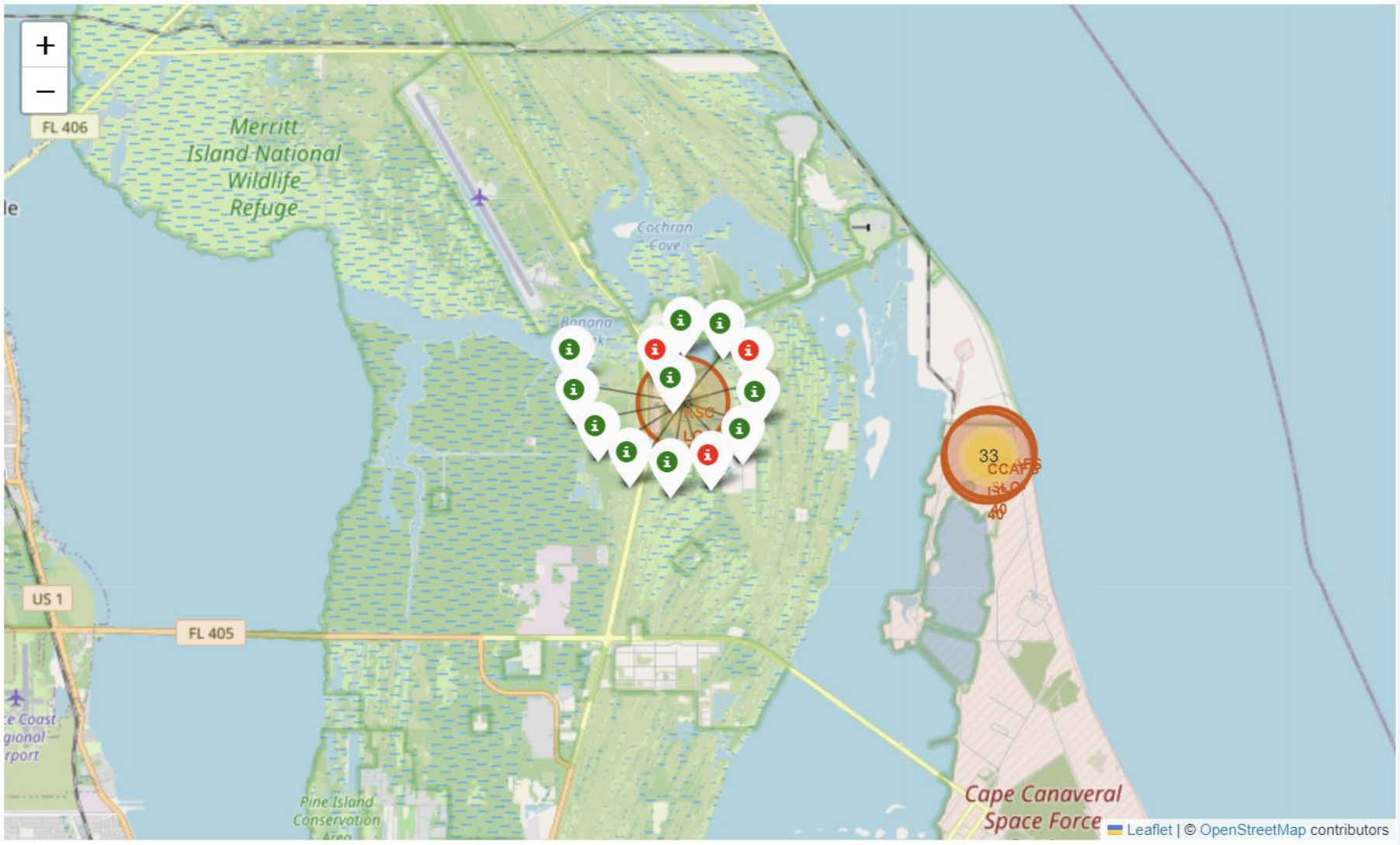
Launch Sites Proximities Analysis

All Launch Sites on a Map



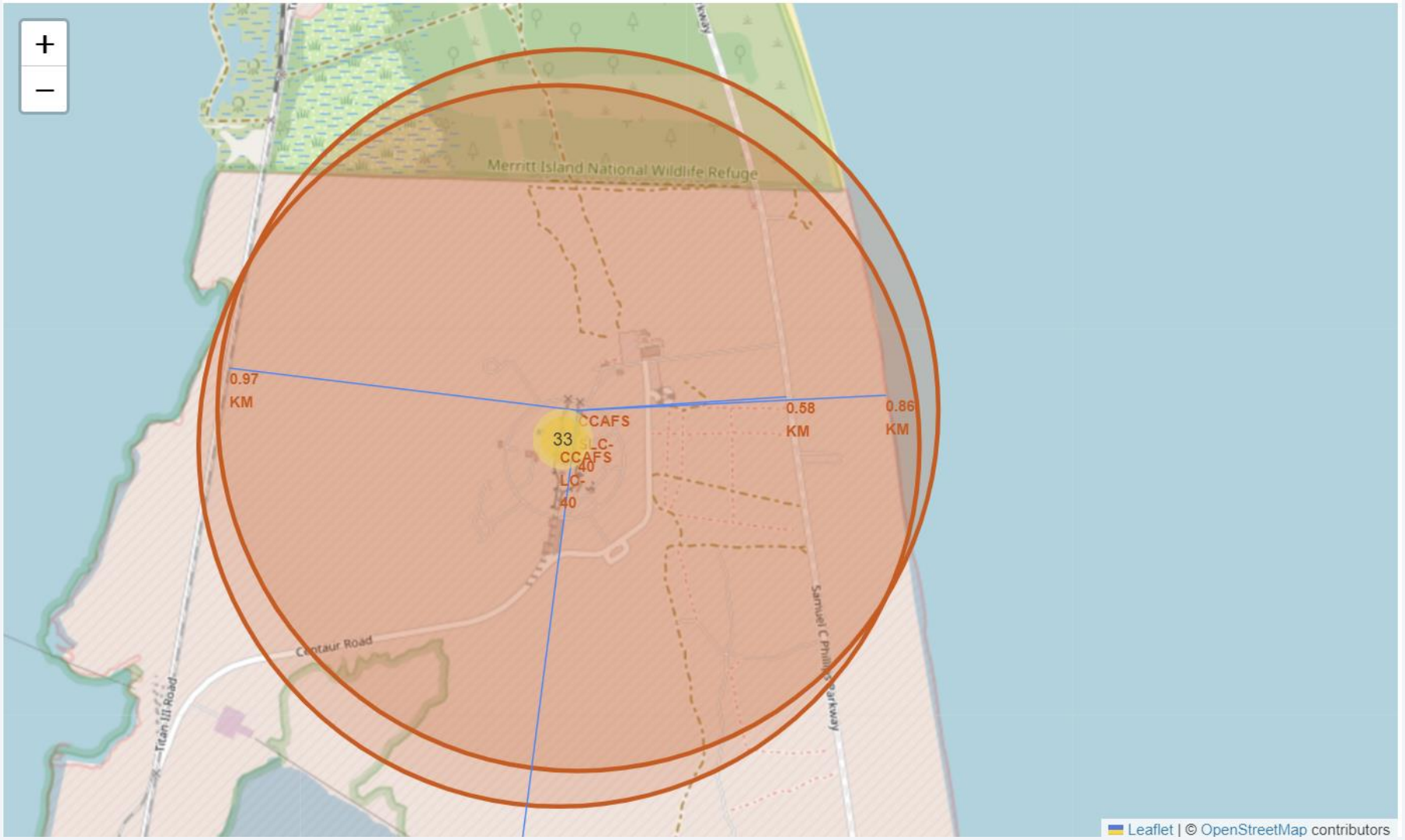
- All launch sites are in proximity to the Equator line. Because if a spacecraft is launched from a site near Earth's equator, it can take optimum advantage of the Earth's substantial rotational speed.
- All launch sites are in very close proximity to the coast. It could provide safety net because the object would have the advantage of flying over the ocean, minimizing the risk of having any debris dropping or exploding near people.

The Success/Failed Launches for Each Site

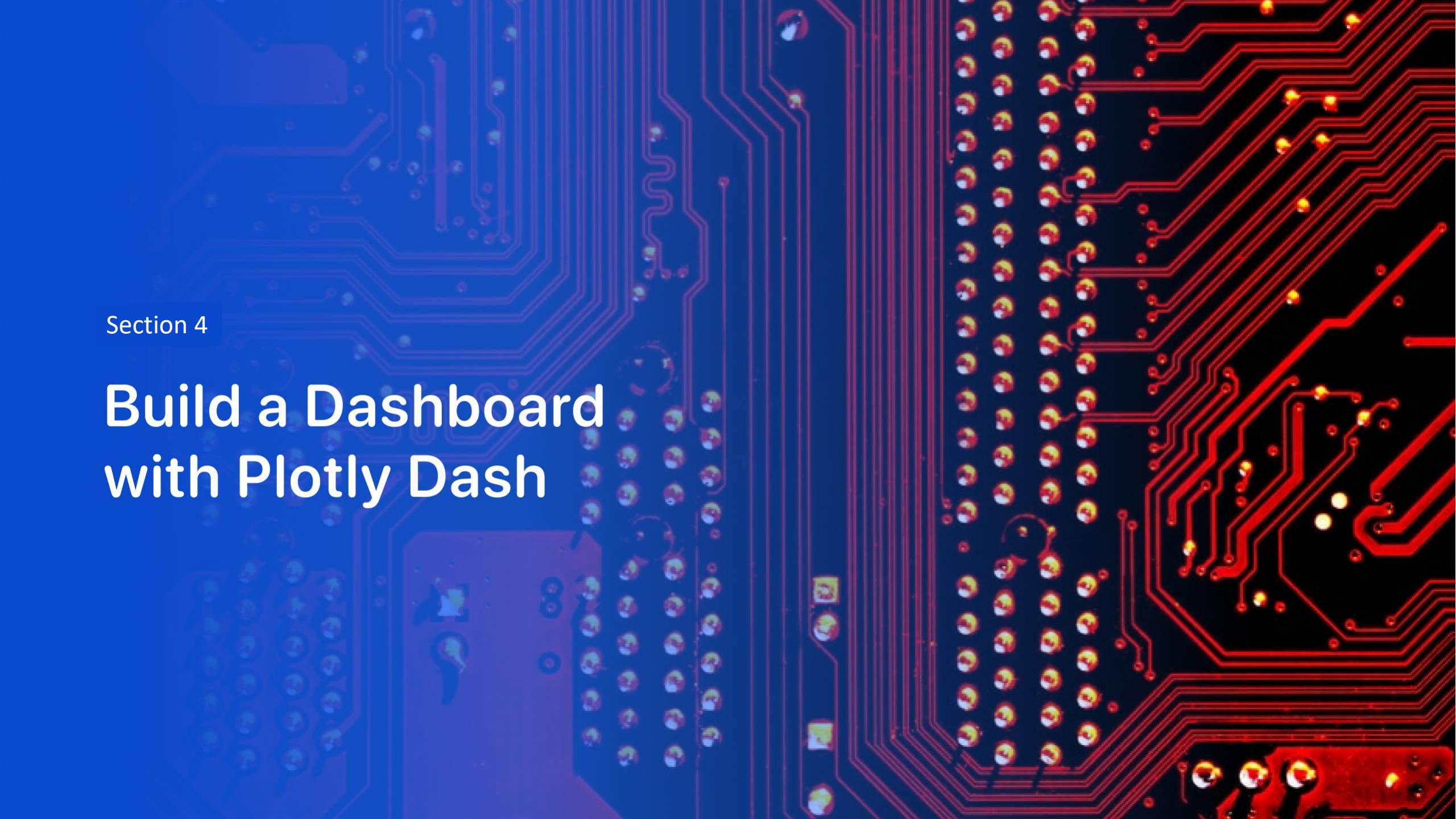


- Clusters on Folium map can be clicked on to display each successful landing (green icon) and failed landing (red icon).
- In this example KSC LC-39A shows 10 successful landings and 3 failed landings.

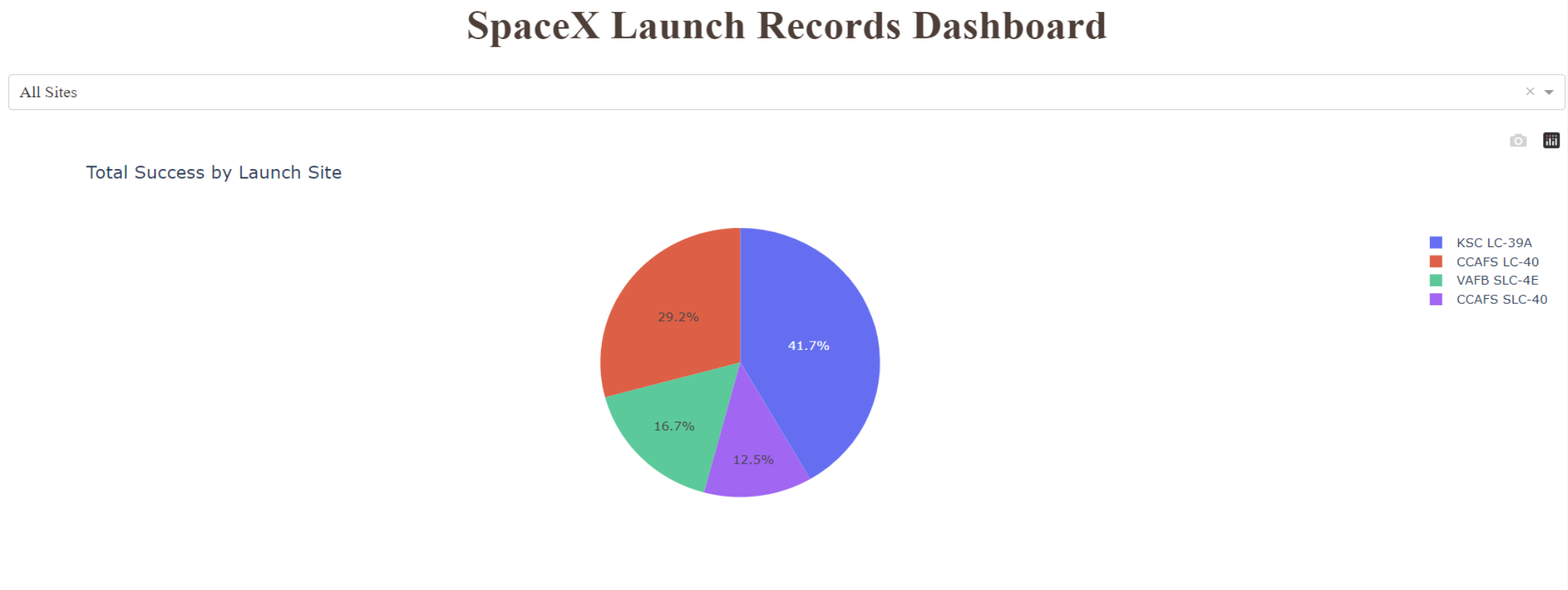
The Distances Between a Launch Site to Its Proximities



- The launch site are in close proximity to railways and highways. Being close to railways and highways make sense as it serves to deliver resources to the launch sites.
- The launch site are in close proximity to coastline and far from the city. It is make sense for minimizing the risk of having any debris dropping or exploding near people.

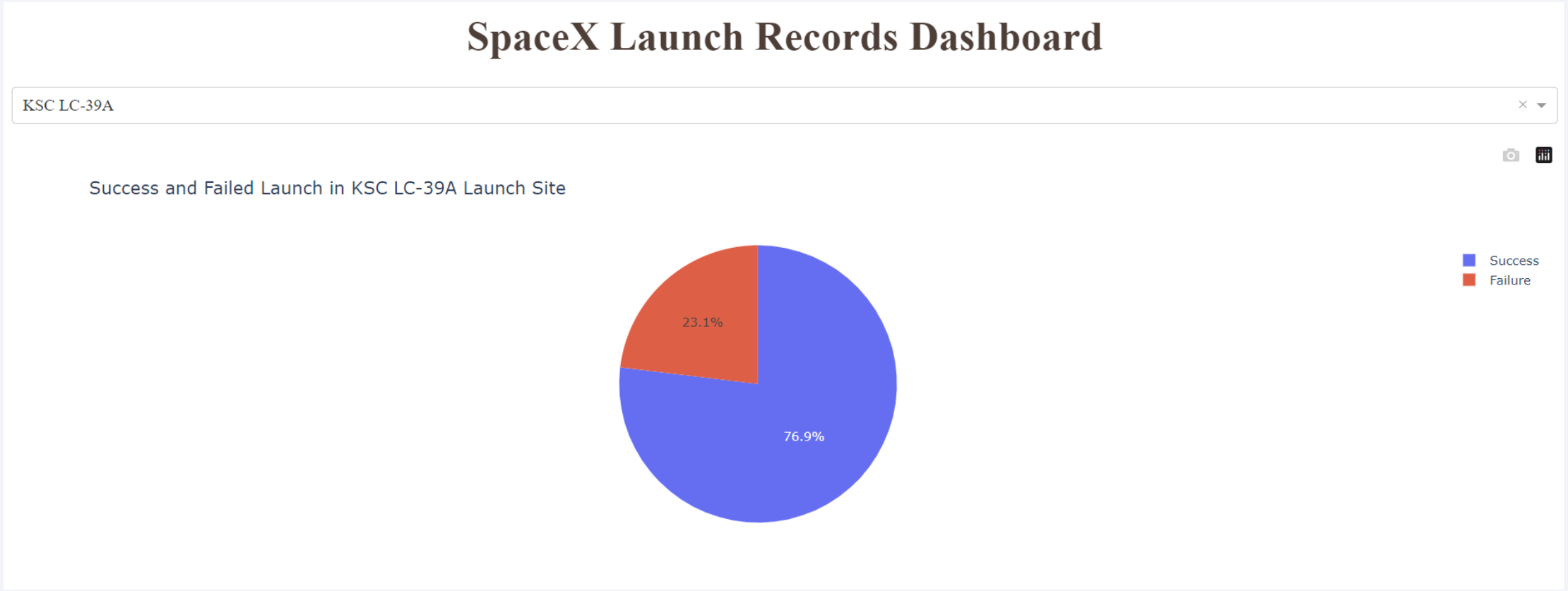

Section 4
Build a Dashboard with Plotly Dash

Successful Launches Across LaunchSites



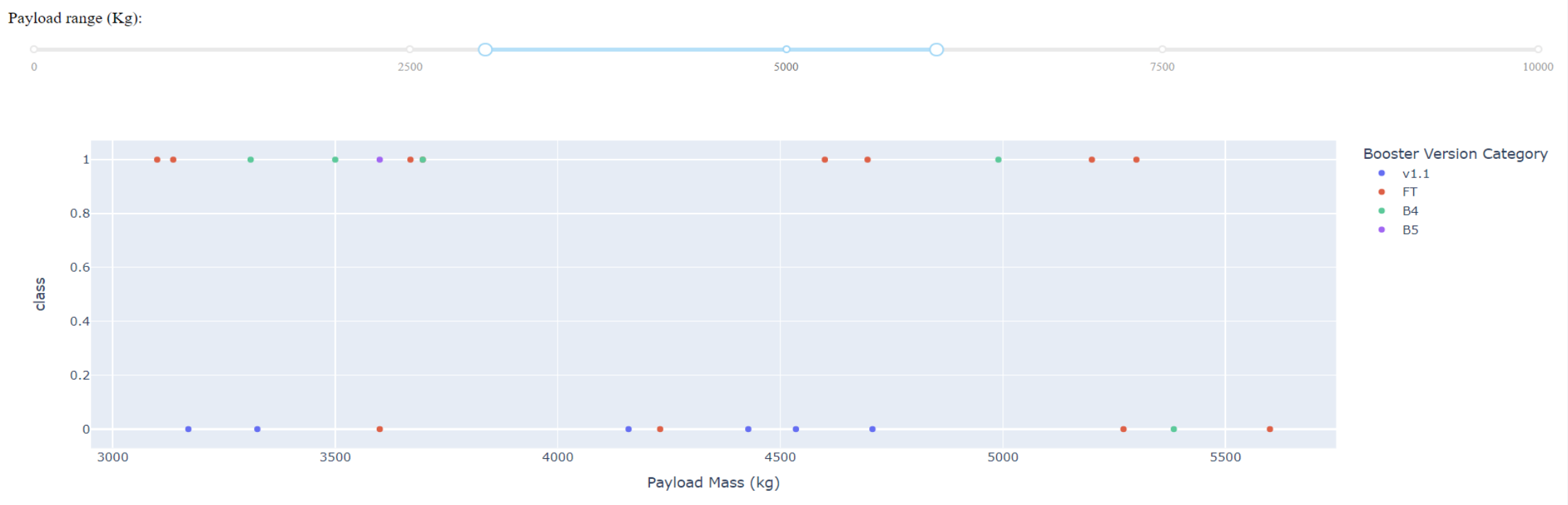
- This is the distribution of successful landings across all launch sites.
- KSC LC-39A has the highest proportion of successful launches accross launch sites.

Highest Success Rate LaunchSite



- KSC LC-39A has the highest success rate with 10 successful landing (76.9%) and 3 failed landing (23.1%)

Relationship between Payload Mass and Landing Outcome

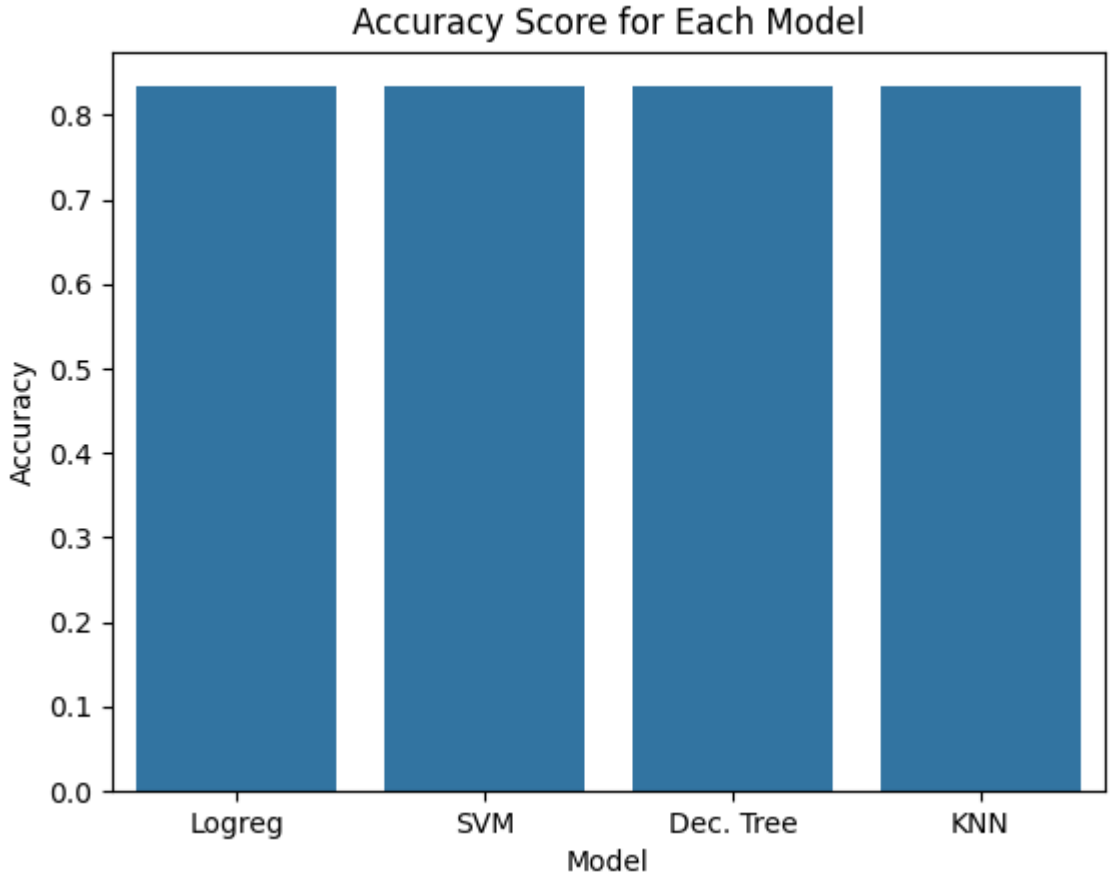


- For payload mass between 3000 – 6000 kg, there are no successful landing for Booster Version Category v1.1 while Booster Version Category B4 has more successful landing.

Section 5

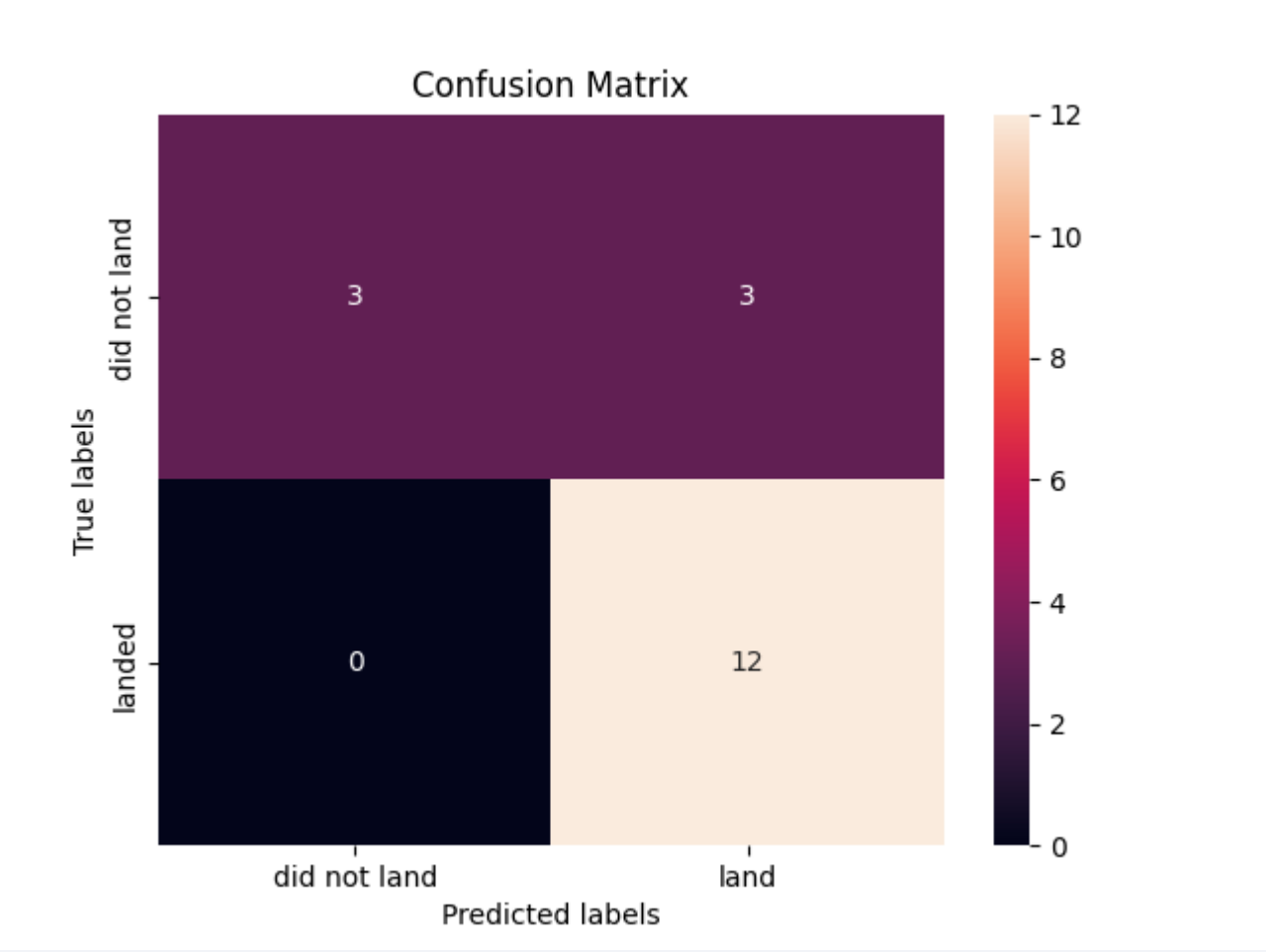
Predictive Analysis (Classification)

Classification Accuracy



- We can see all of the models perform equally well with the same accuracy score of around 83.33%.

Confusion Matrix



- TP = 12, TN = 3, FP = 3, FN = 0
- Recall = TP/(TP + FN) = 100%
- Precision = TP/(TP + FP) = 80%
- Accuracy = 83.33%

- The recall is 100%, then it tells us the model has detected all positive samples as positive and neglects how all negative samples are classified in the model.

Conclusions

- The launch site are in close proximity to railways and highways. Being close to railways and highways make sense as it serves to deliver resources to the launch sites.
- The launch site are in close proximity to coastline and far from the city. It is make sense for minimizing the risk of having any debris dropping or exploding near people.
- Different launch sites have different success rates. But as we increase the number of flights the success rate increase.
- KSC LC-39A has the highest proportion of successful launches accross launch sites.
- KSC LC-39A has the highest success rate with 10 successful landing (76.9%) and 3 failed landing (23.1%)
- Rockets launched with heavypayload mass (greater than 10000) tends to success with only 1 failed launch.
- All models produced similar results with accuracy rate about 83.33%.

Appendix

- Github URL :

https://github.com/mzulfikarmuslim/IBM_FinalAssignment_Course10/tree/main

- Special Thanks to All Instructors

https://www.coursera.org/professional-certificates/ibm-data-science?#instructors

Thank you!