IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Neelesh Anand
30-Dec-22

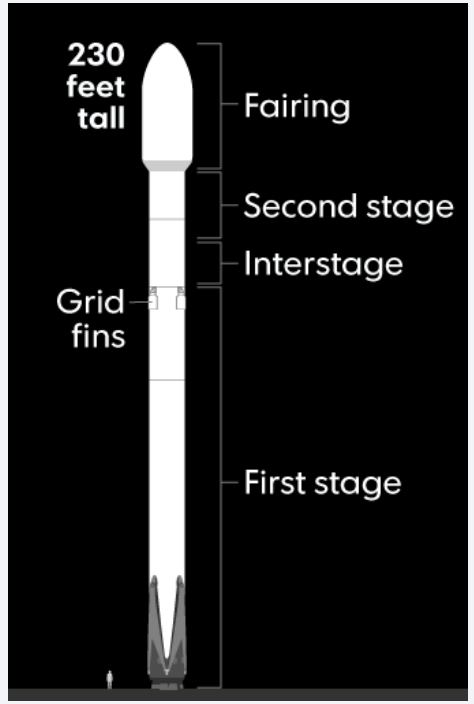
Outline

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

Executive Summary

The goal of the project is to develop a model that can accurately predict the likelihood of successful landing for a rocket's first stage, based on a variety of input factors.

Data Source: SpaceX launch information (public data)



- Summary of methodologies
  - Exploratory data analysis (EDA) of Falcon 9 first-stage landings
  - Launch sites proximity analysis
  - Predictive classification analysis
- Summary of all results
  - Exploratory data analysis (EDA) of Falcon 9 first-stage landings
  - Classification model can predict the successful landing likelihood by accuracy of 83% .

Introduction

- Project background and context
  - Setup a new company in Space Exploration field
  - Launch cost can be reduced by recovering Stage1 of the rocket (~70% less)
  - SpaceX Falcon9 can recover first stage, but sometimes the first stage does not land due to crash or mission parameters.

- Problems you want to find answers
  - Whether first stage recovery is recommended or not for project.
  - Develop a model based on SpaceX data to predict probability of successfully landing the first stage, given various mission parameters.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data Collection with SpaceX API (rocket launch data)
  - Data Collection with Web Scraping (wikipedia Table)
- Perform data wrangling
  - Determine Training Labels (input and outcome features)
  - Missing data correction
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Comparison of various classification algorithms and parameter fine tuning

Data Collection

Data Source and Methods:

1. Data Collection with SpaceX API (rocket launch data)
2. Data Collection with Web Scraping (wikipedia Table)

1.

SpaceX API

⬇

Json Result

⬇

Json Normalize Method

⬇

DataFrame *(filtering)*

⬇

Falcon 9 First Stage Landing Dataset

2.

Wikipedia page

⬇

HTML Table

⬇

Parse the table

⬇

Pandas Data Frame

⬇

Falcon 9 historical launch records

Data Collection – SpaceX API

SpaceX API
⬇
Json Result
⬇
Json Normalize Method
⬇
DataFrame *(filtering)*
⬇
Falcon 9 First Stage
Landing Dataset

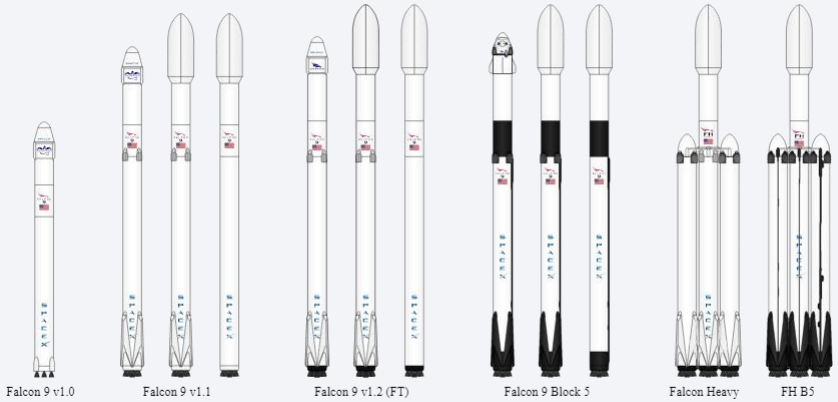
Objectives

In this lab, you will make a get request to the SpaceX API. You will also do some basic data wrangling and formatting.

- Request to the SpaceX API
- Clean the requested data

GitHub URL of the completed SpaceX API calls notebook : Link

Data Collection - Scraping



Wikipedia page

⬇

HTML Table

⬇

Parse the table

⬇

Pandas Data Frame

⬇

Falcon 9 historical launch records

Objectives

Web scrap Falcon 9 launch records with BeautifulSoup:

- Extract a Falcon 9 launch records HTML table from Wikipedia
- Parse the table and convert it into a Pandas data frame

GitHub URL of the completed SpaceX webscraping notebook : Link

Data Wrangling

Objectives

Perform exploratory Data Analysis and determine Training Labels

- Exploratory Data Analysis
- Determine Training Labels

```
Index(['FlightNumber', 'Date', 'BoosterVersion', 'PayloadMass', 'Orbit',
       'LaunchSite', 'Outcome', 'Flights', 'GridFins', 'Reused', 'Legs',
       'LandingPad', 'Block', 'ReusedCount', 'Serial', 'Longitude',
       'Latitude'],
```

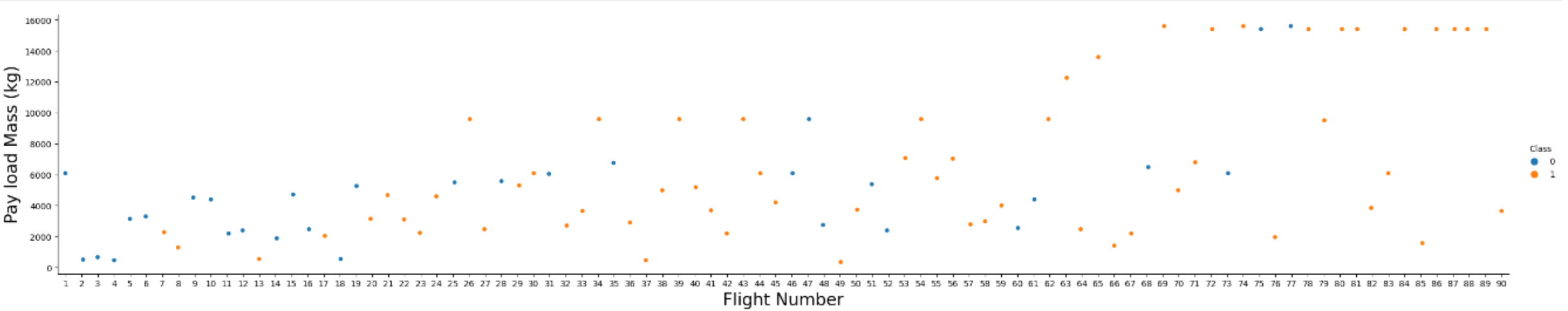
```
Index(['FlightNumber', 'Date', 'BoosterVersion', 'PayloadMass', 'Orbit',
       'LaunchSite', 'Outcome', 'Flights', 'GridFins', 'Reused', 'Legs',
       'LandingPad', 'Block', 'ReusedCount', 'Serial', 'Longitude', 'Latitude',
       'Class'],
```

GitHub URL of your completed data wrangling : Link

EDA with Data Visualization

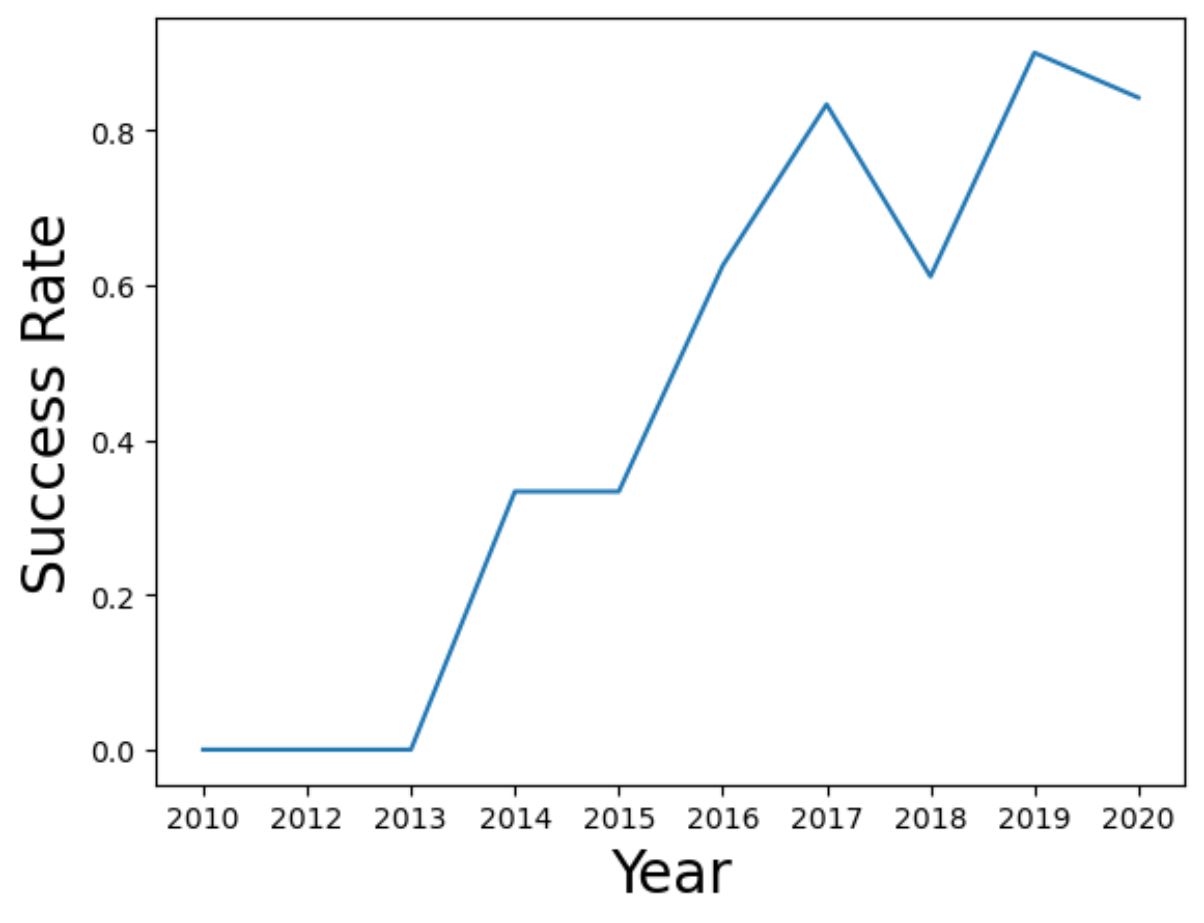
• Scatterplots and barplots were used to visualize the relationship between pair of features:

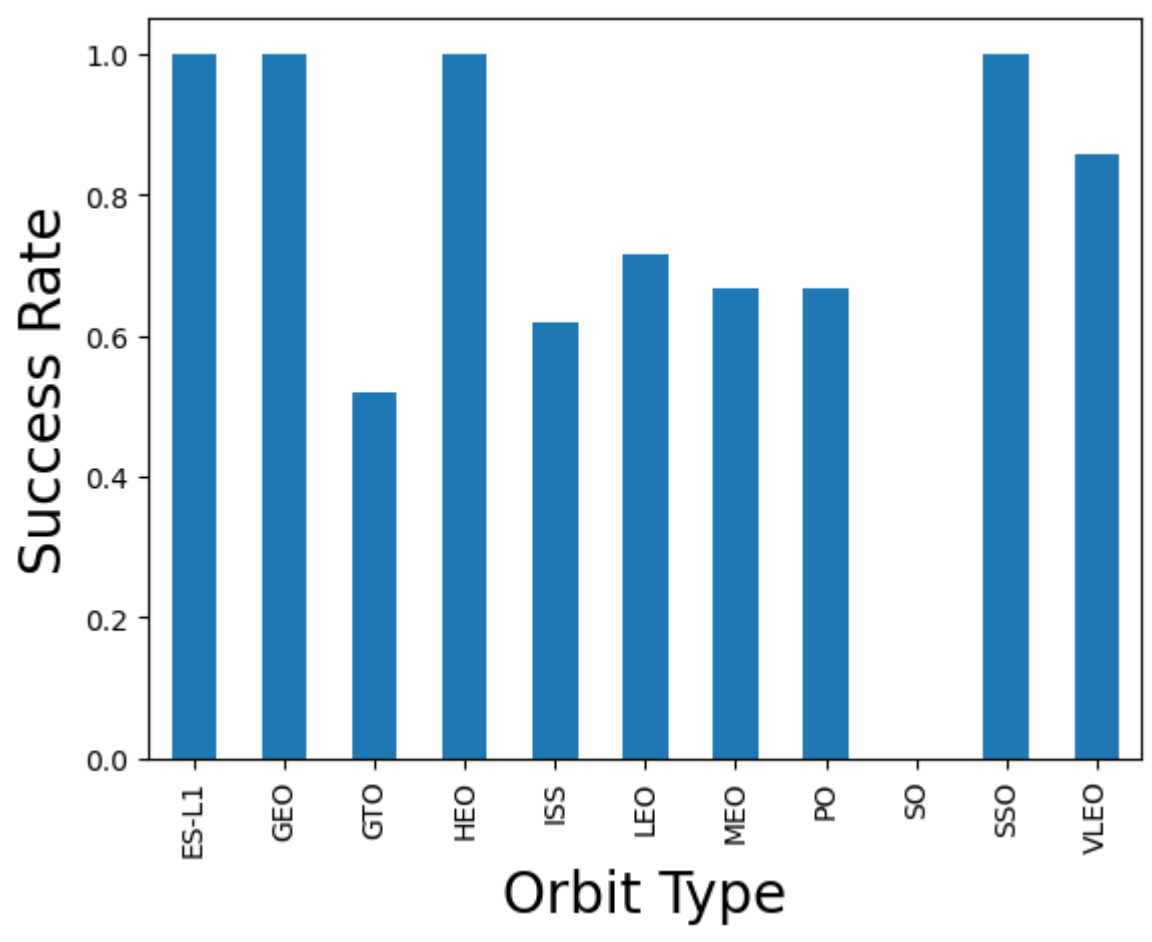
Payload Mass X Flight Number, Launch Site X Flight Number, Launch Site X Payload Mass, Orbit and Flight Number, Payload and Orbit



GitHub URL : Link

EDA with Data Visualization





GitHub URL : [Link]

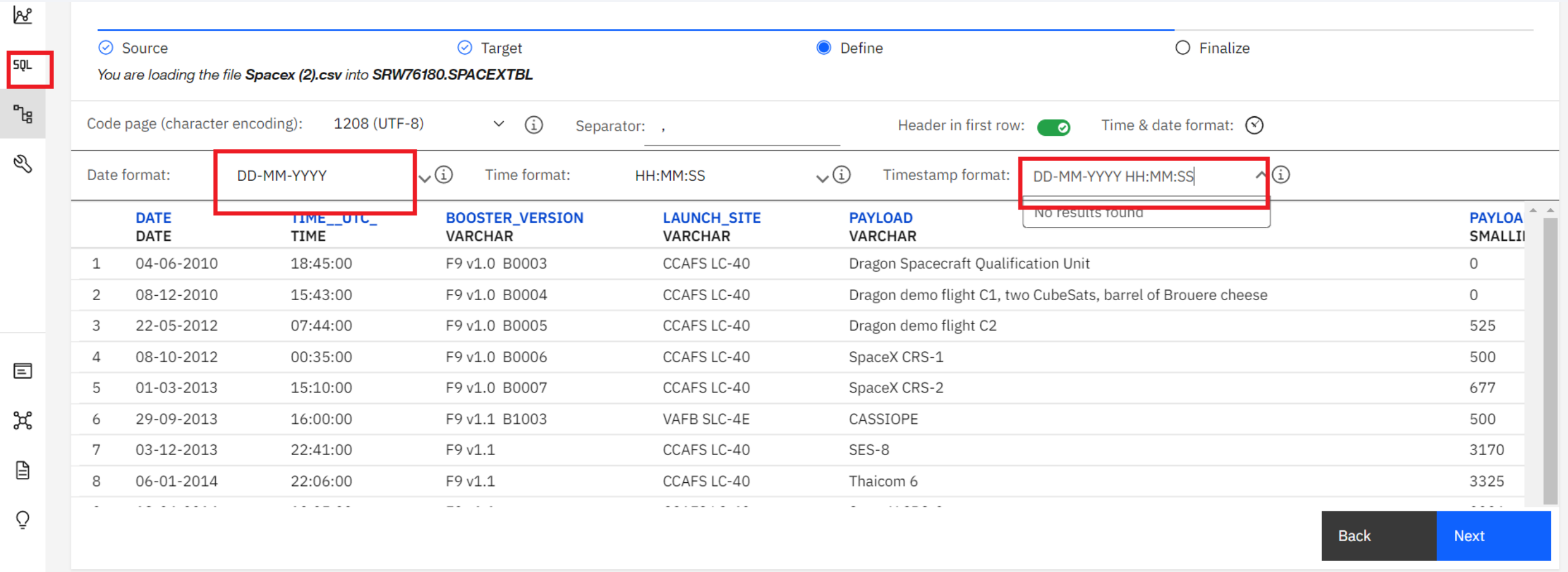
EDA with SQL

SQL queries:

- 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.

GitHub URL : Link

EDA with SQL



	DATE DATE	TIME__UTC_ TIME	BOOSTER_VERSION VARCHAR	LAUNCH_SITE VARCHAR	PAYLOAD VARCHAR	PAYLOA SMALLI
1	04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0
2	08-12-2010	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0
3	22-05-2012	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525
4	08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500
5	01-03-2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677
6	29-09-2013	16:00:00	F9 v1.1 B1003	VAFB SLC-4E	CASSIOPE	500
7	03-12-2013	22:41:00	F9 v1.1	CCAFS LC-40	SES-8	3170
8	06-01-2014	22:06:00	F9 v1.1	CCAFS LC-40	Thaicom 6	3325

GitHub URL : Link

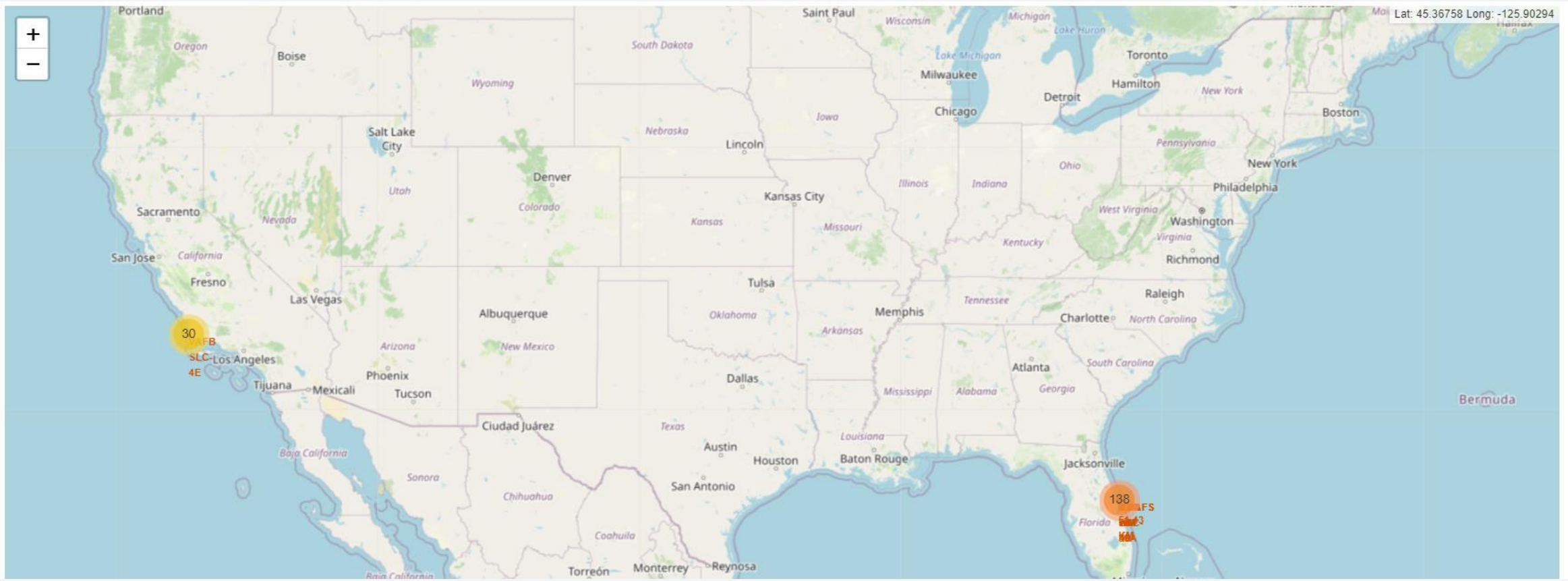
Build an Interactive Map with Folium

Folium Map : Markers, circles, lines and marker clusters

- Markers : launch sites.
- Circles : highlighted areas around specific coordinates
- Marker clusters : groups of events in each coordinate
- Lines : indicate distances between two coordinates.

GitHub URL : Link

Build an Interactive Map with Folium



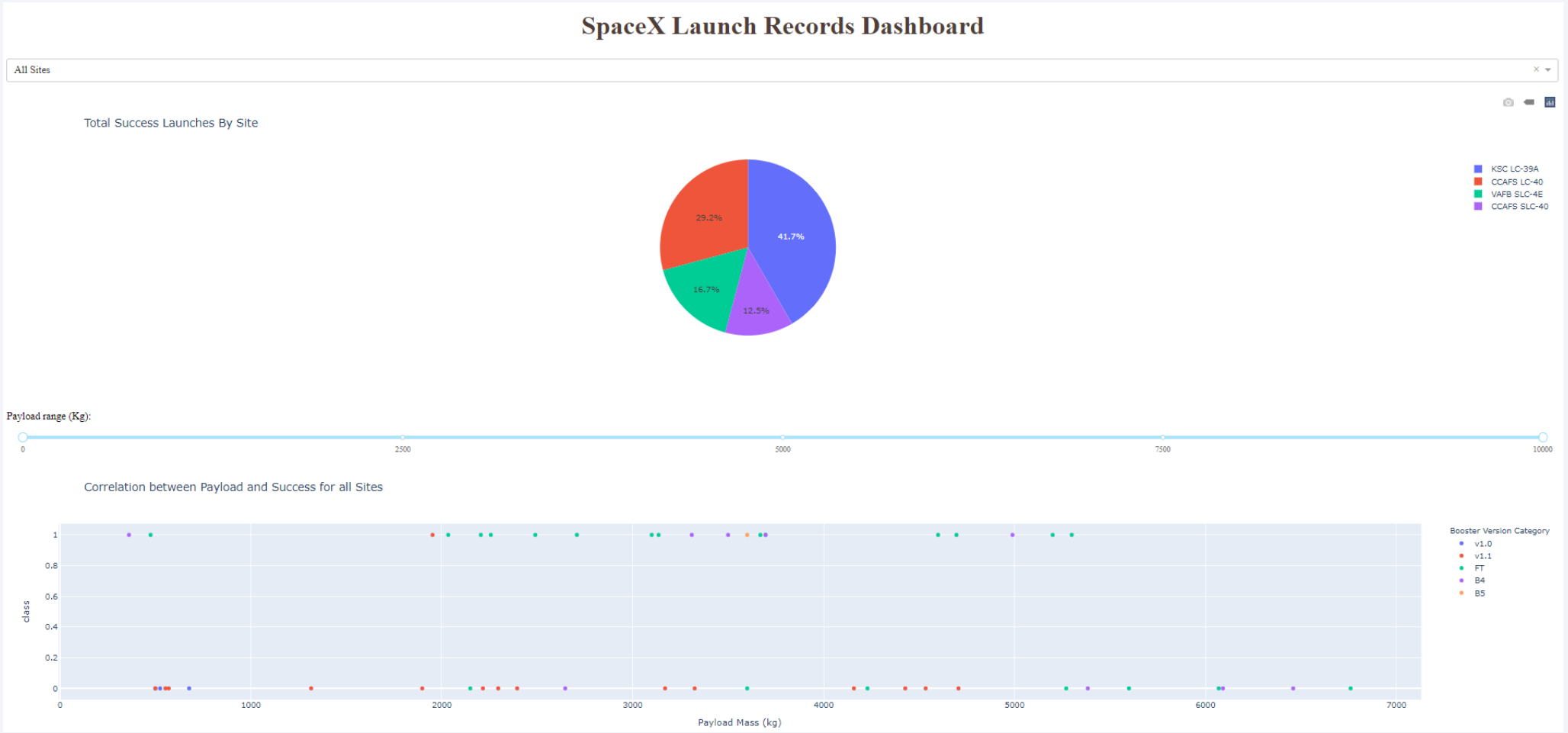
GitHub URL : [Link]

Build a Dashboard with Plotly Dash

- The following graphs and plots were used to visualize data
  - Percentage of launches by site
  - Payload range
- Analyze the relation between payloads and launch sites, helping to identify where is best place to launch according to payloads.

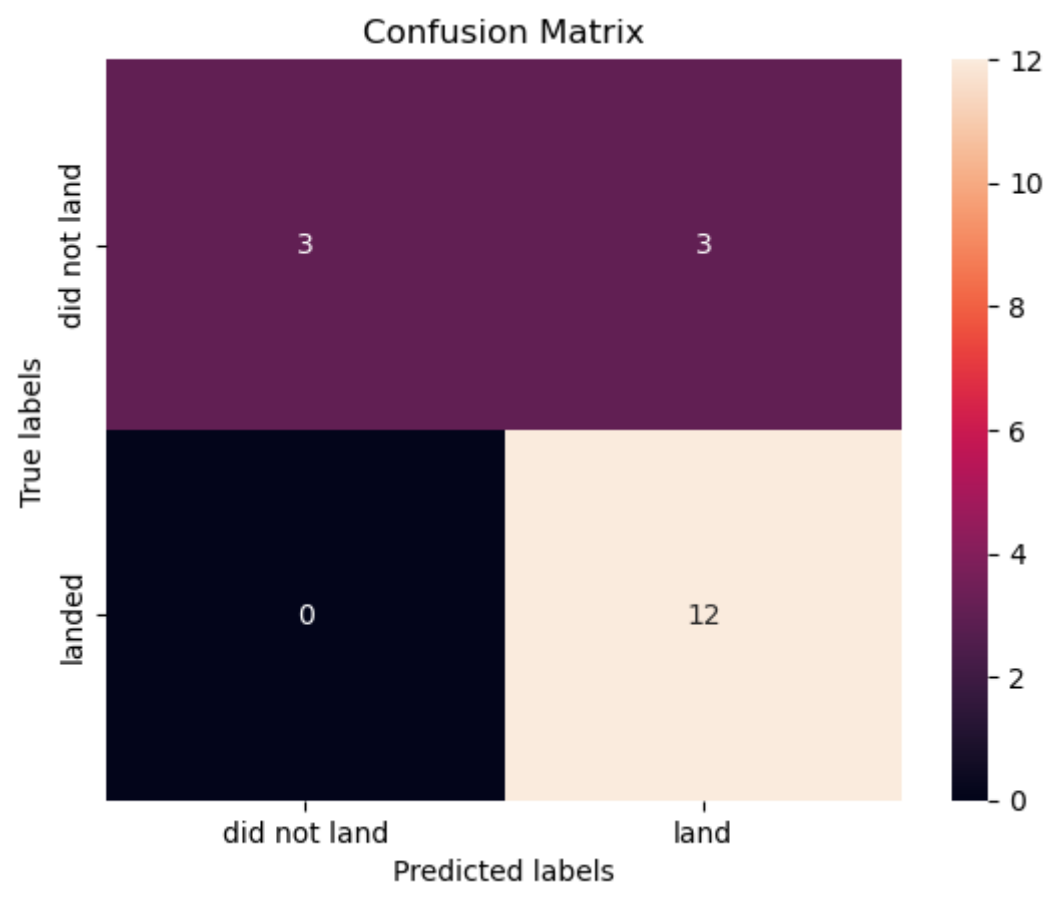
GitHub URL : Link

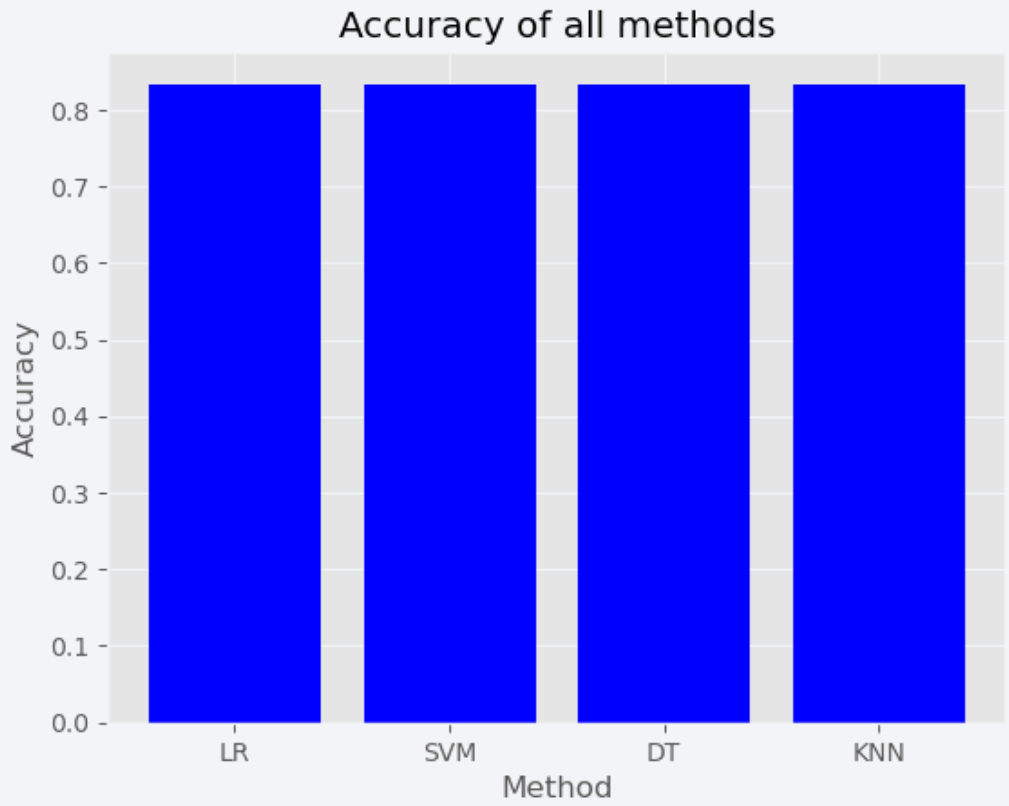
Build a Dashboard with Plotly Dash



GitHub URL : [Link]

Predictive Analysis (Classification)

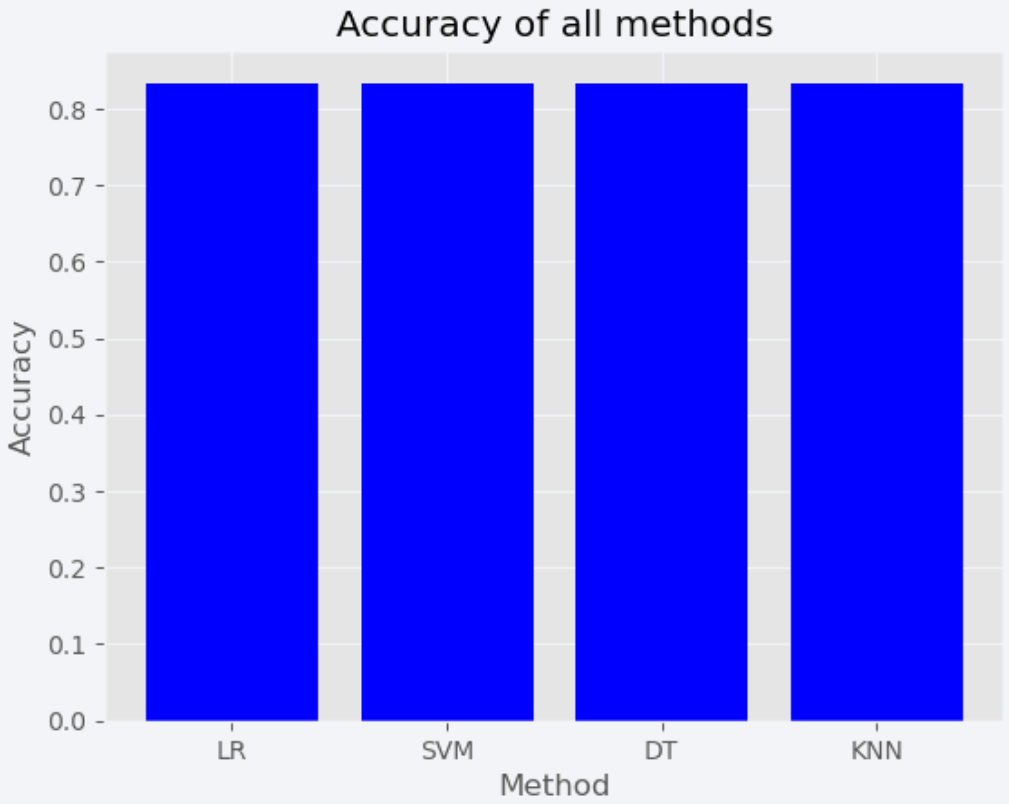


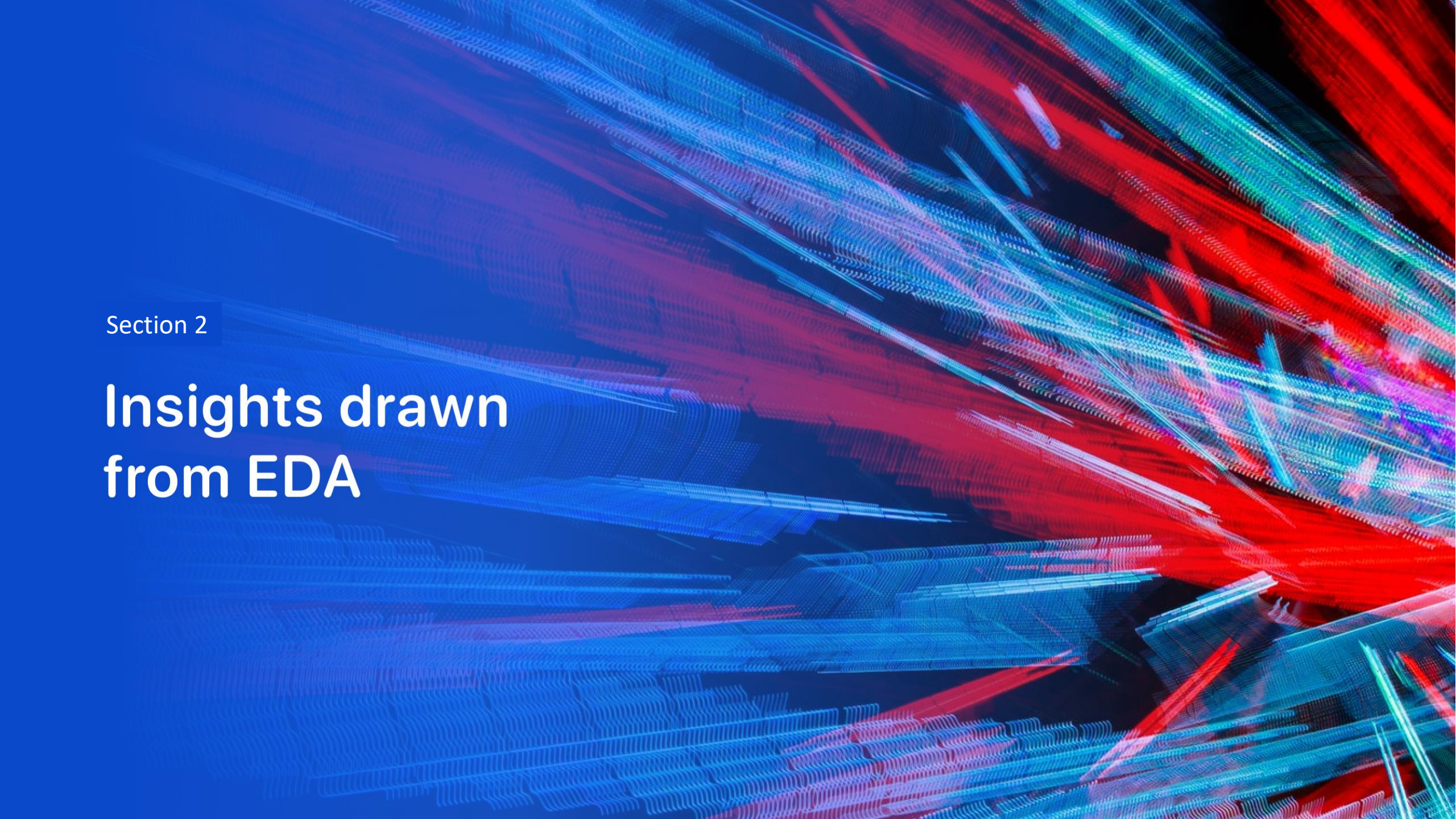


GitHub URL : [Link]

Results

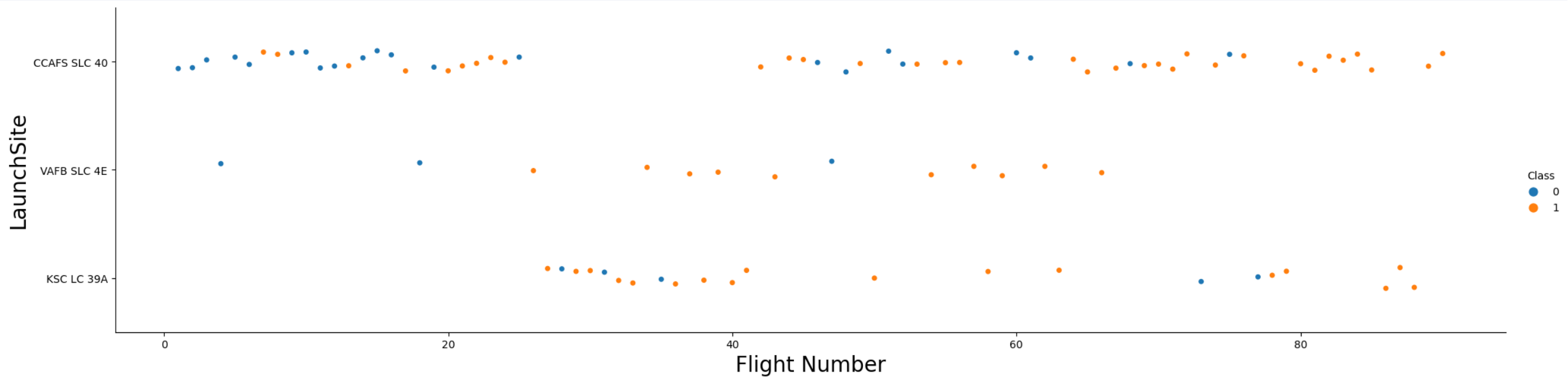
- Predictive analysis shows that the accuracy of the trained model is over 83% on test data.




Section 2
Insights drawn from EDA

Flight Number vs. Launch Site

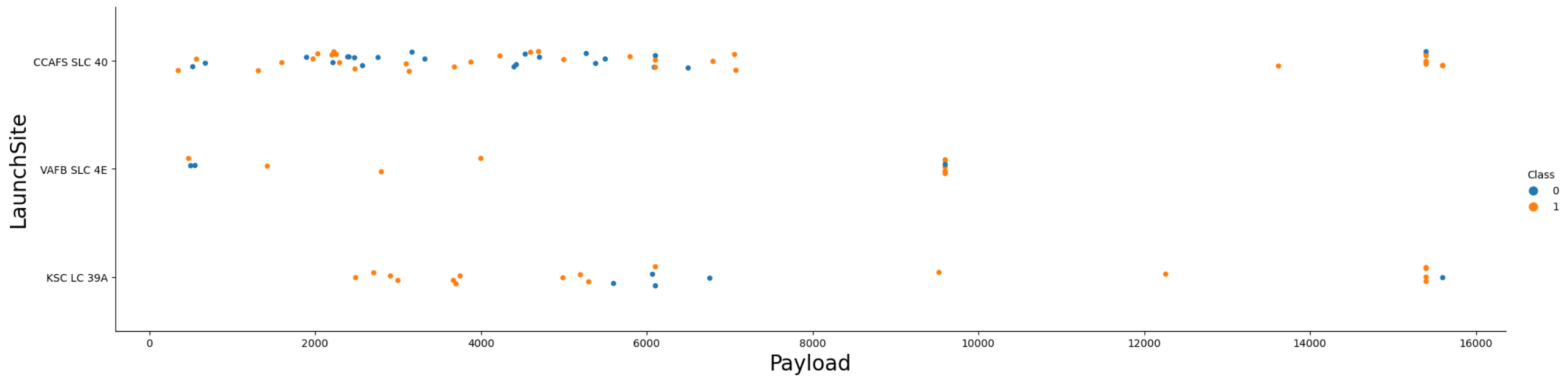
Scatter plot of Flight Number vs. Launch Site



- Best launch site option: CCAFS SLC 40
- General success rate is increasing over time.

Payload vs. Launch Site

Scatter plot of Payload vs. Launch Site

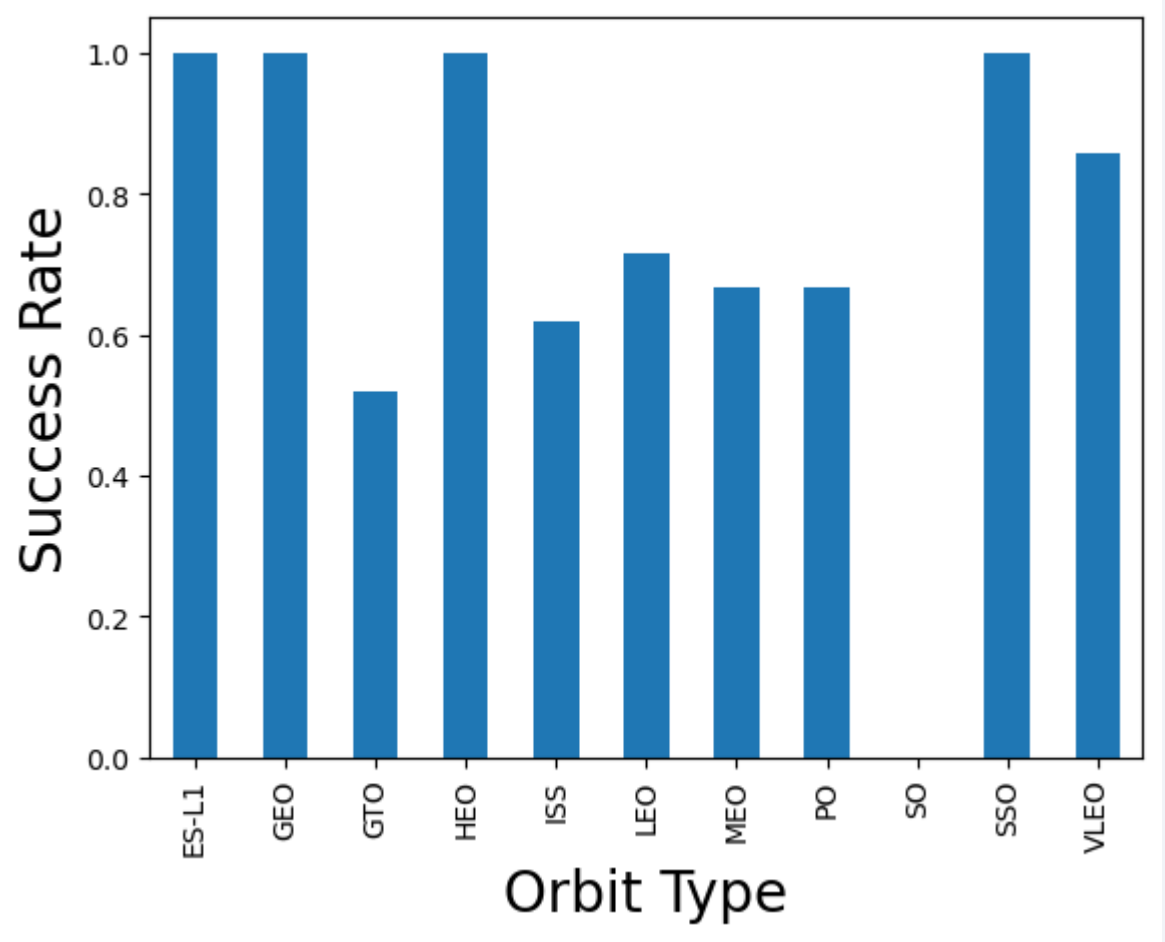


- Higher payload have better success rate
- KSC LC 39A have excellent success rate for payload of under 5000kg

Success Rate vs. Orbit Type

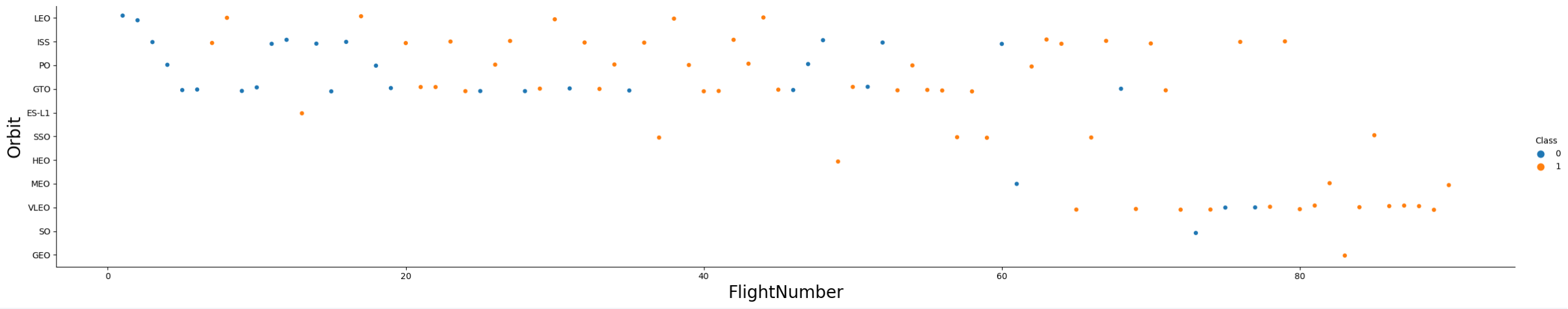
Bar chart for the success rate of each orbit type

- Best orbit to launch
  - ES-L1
  - GEO
  - HEO
  - SSO



Flight Number vs. Orbit Type

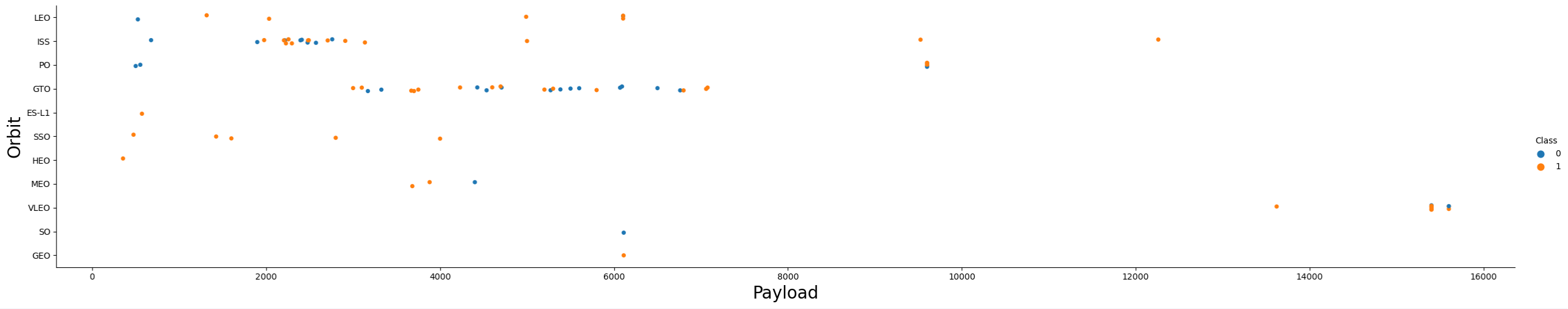
Scatter point of Flight number vs. Orbit type



- VLEO orbit have excellent success rate in last 10 launces

Payload vs. Orbit Type

Scatter point of payload vs. orbit type

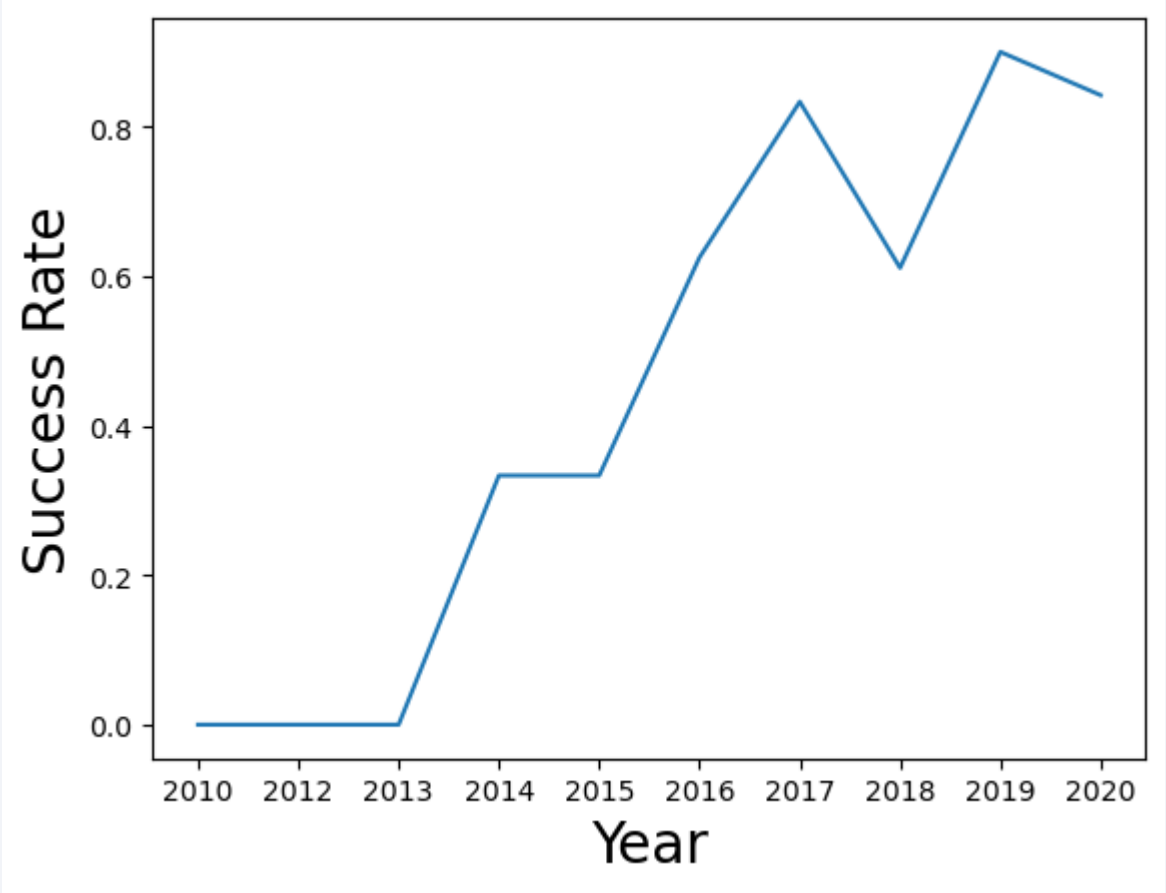


- VLEO orbit is preferred for heavy payload category
- SSO orbit have 100% success rate for low payload category

Launch Success Yearly Trend

Line chart of yearly average success rate

- Over the years, success rate has been increased.
- This may be caused by technological advances in recent years.



All Launch Site Names

- Names of the unique launch sites
  - CCAFS LC-40
  - CCAFS SLC-40
  - KSC LC-39A
  - VAFB SLC-4E

- Select unique occurrences of “ launch_site ” values from the dataset.

Launch Site Names Begin with 'CCA'

- 5 records where launch sites begin with `CCA`

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 attemp

Total Payload Mass

- Total payload carried by boosters from NASA : **111268 kg**

- Total payload calculated above, by summing all payloads whose codes contain ‘CRS’, which corresponds to NASA.

Average Payload Mass by F9 v1.1

- Average payload mass carried by booster version F9 v1.1 : **2928kg**
- Filtering data by the booster version above and calculating the average payload mass we obtained the value of 2928 kg.

First Successful Ground Landing Date

- Date of the first successful landing outcome on ground pad : 22-Dec-2015
- By filtering data by successful landing outcome on ground pad and getting the minimum value for date it's possible to identify the first occurrence, that happened on 12/22/2015.

Successful Drone Ship Landing with Payload between 4000 and 6000

- Boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

Booster Version
F9 FT B1021.2
F9 FT B1031.2
F9 FT B1022
F9 FT B1026

- Select distinct booster versions according to the filters above, these 4 are the result.

Total Number of Successful and Failure Mission Outcomes

- Total number of successful and failure mission outcomes

Mission Outcome	Occurrences
Success	99
Success (payload status unclear)	1
Failure (in flight)	1

- Grouping mission outcomes and counting records for each group led us to the summary above.

Boosters Carried Maximum Payload

- Names of the booster which have carried the maximum payload mass

Booster Version (...)
F9 B5 B1048.4
F9 B5 B1048.5
F9 B5 B1049.4
F9 B5 B1049.5
F9 B5 B1049.7
F9 B5 B1051.3

Booster Version
F9 B5 B1051.4
F9 B5 B1051.6
F9 B5 B1056.4
F9 B5 B1058.3
F9 B5 B1060.2
F9 B5 B1060.3

- These are the boosters which have carried the maximum payload mass registered in the dataset.

2015 Launch Records

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

Booster Version	Launch Site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

- The list above has the only two occurrences.

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

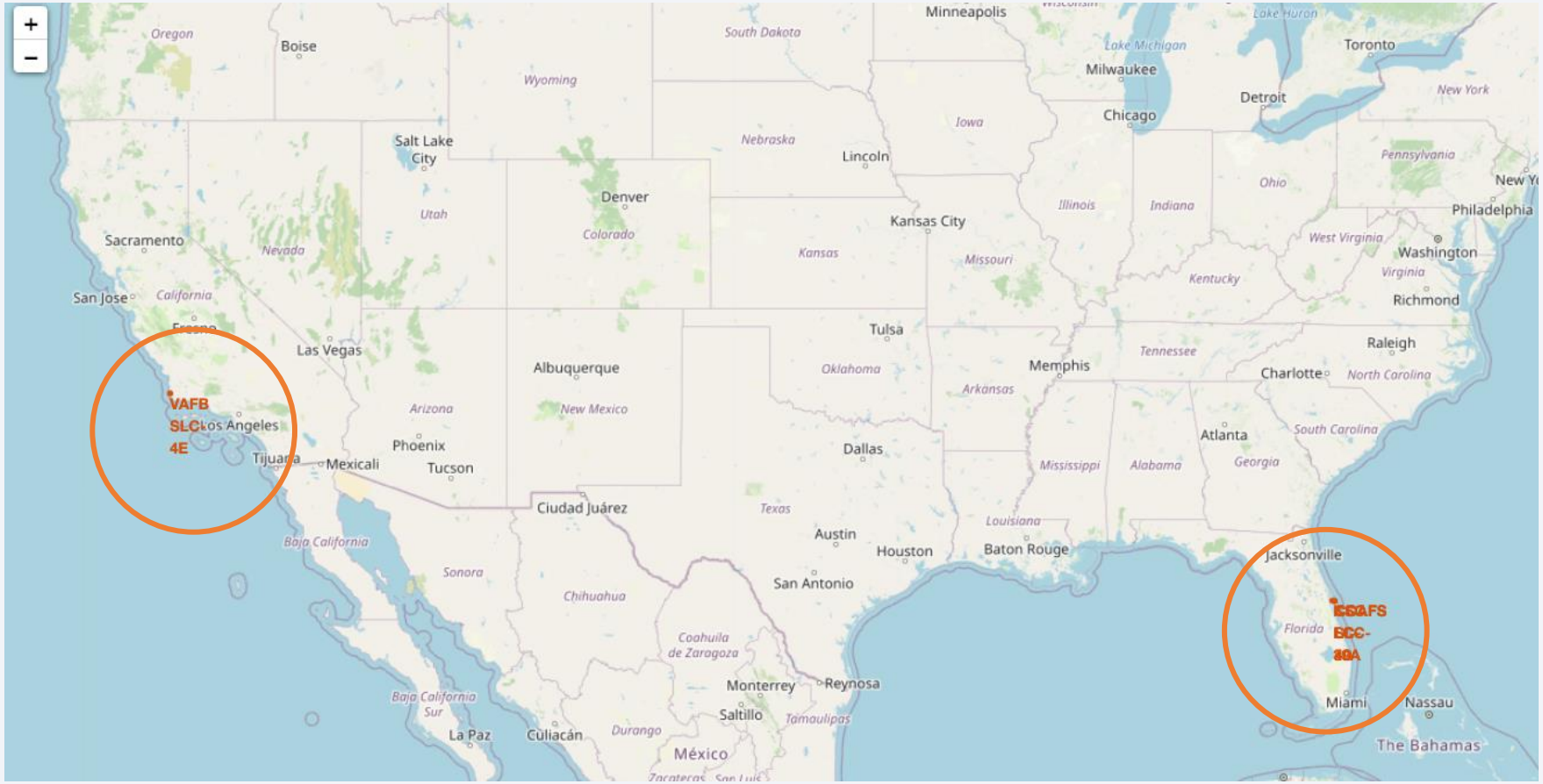
Landing Outcome	Occurrences
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

- This view of data alerts us that “No attempt” must be taken in account.

Section 3

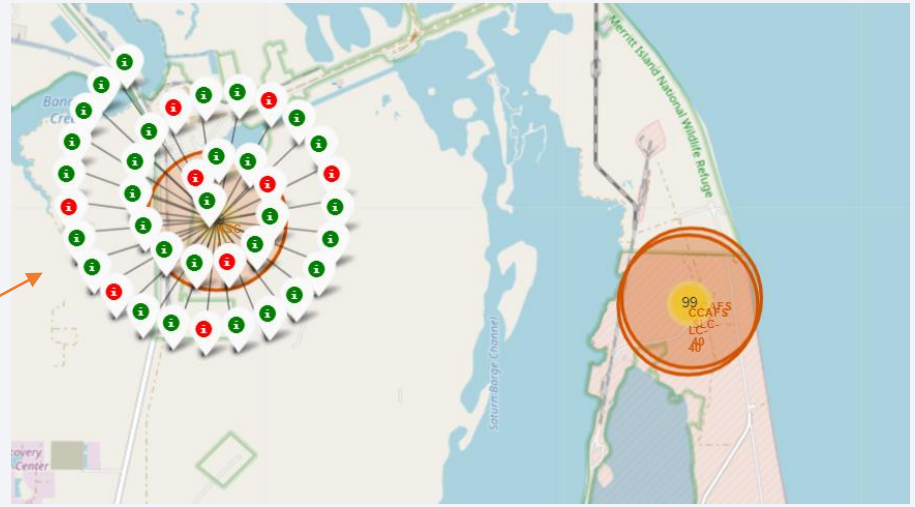
Launch Sites Proximities Analysis

Folium Map : Location of Launch Sites



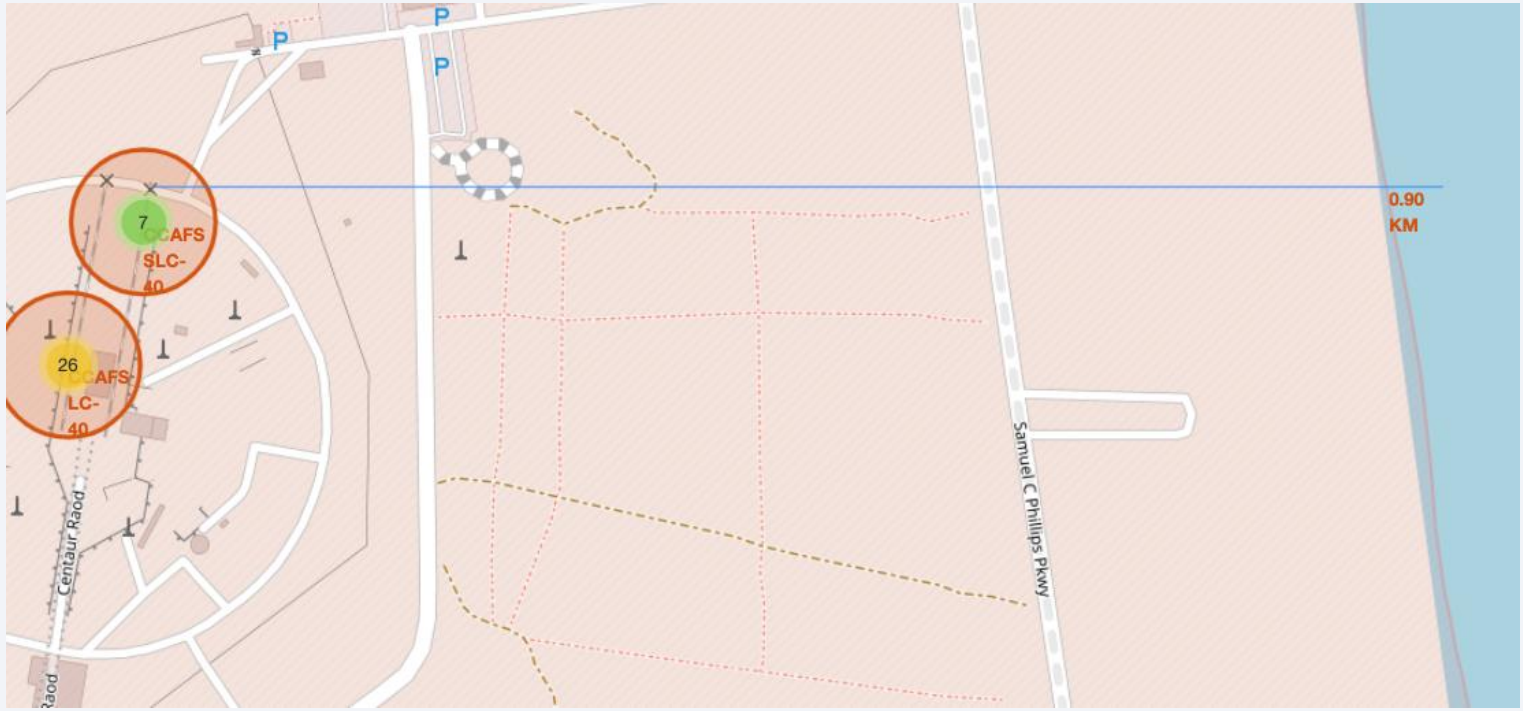
All launch sites in very close proximity to the coast

Folium Map : Launch Site Frequency

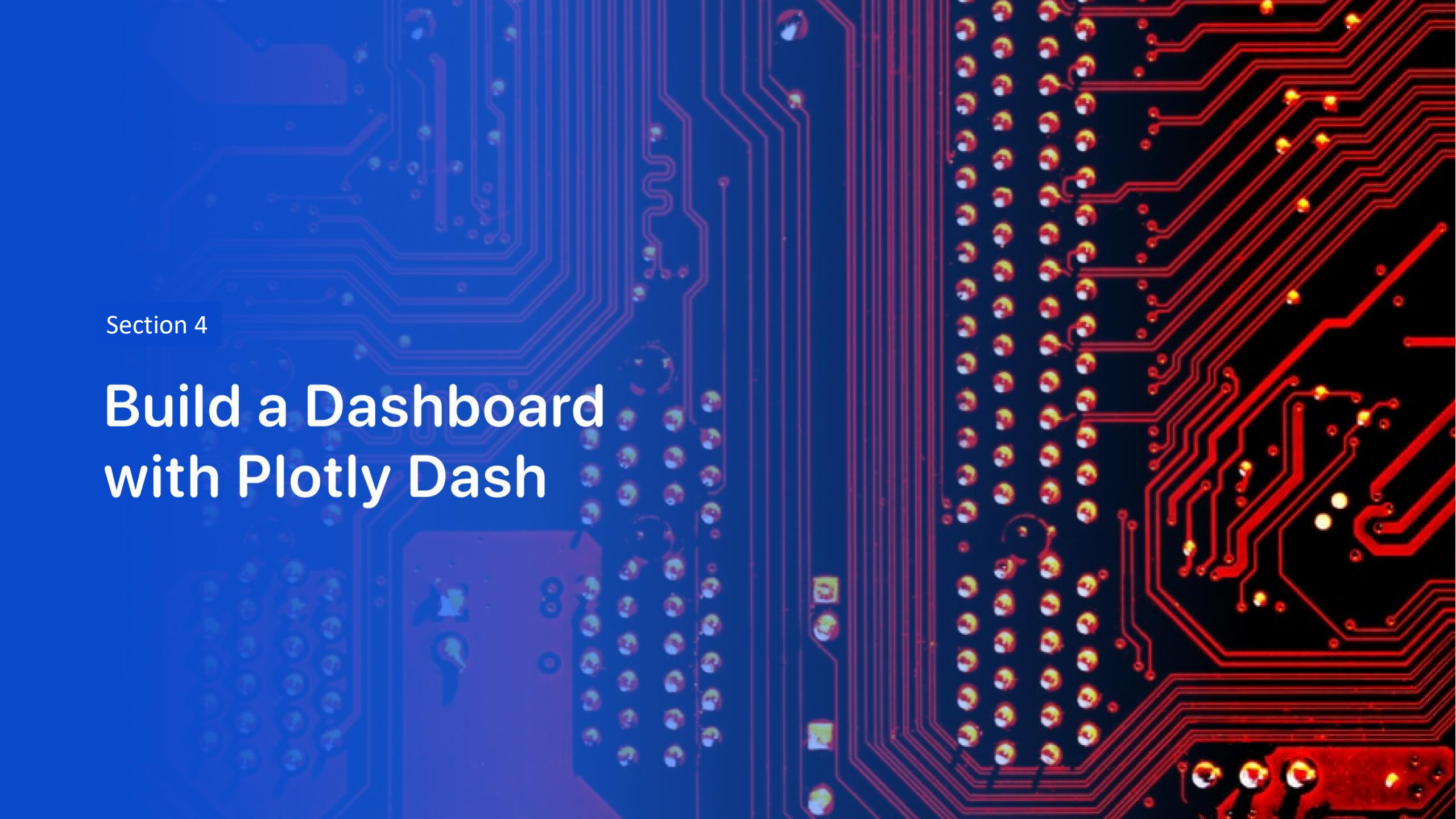


Site KSC LC-39A have high landing success rate

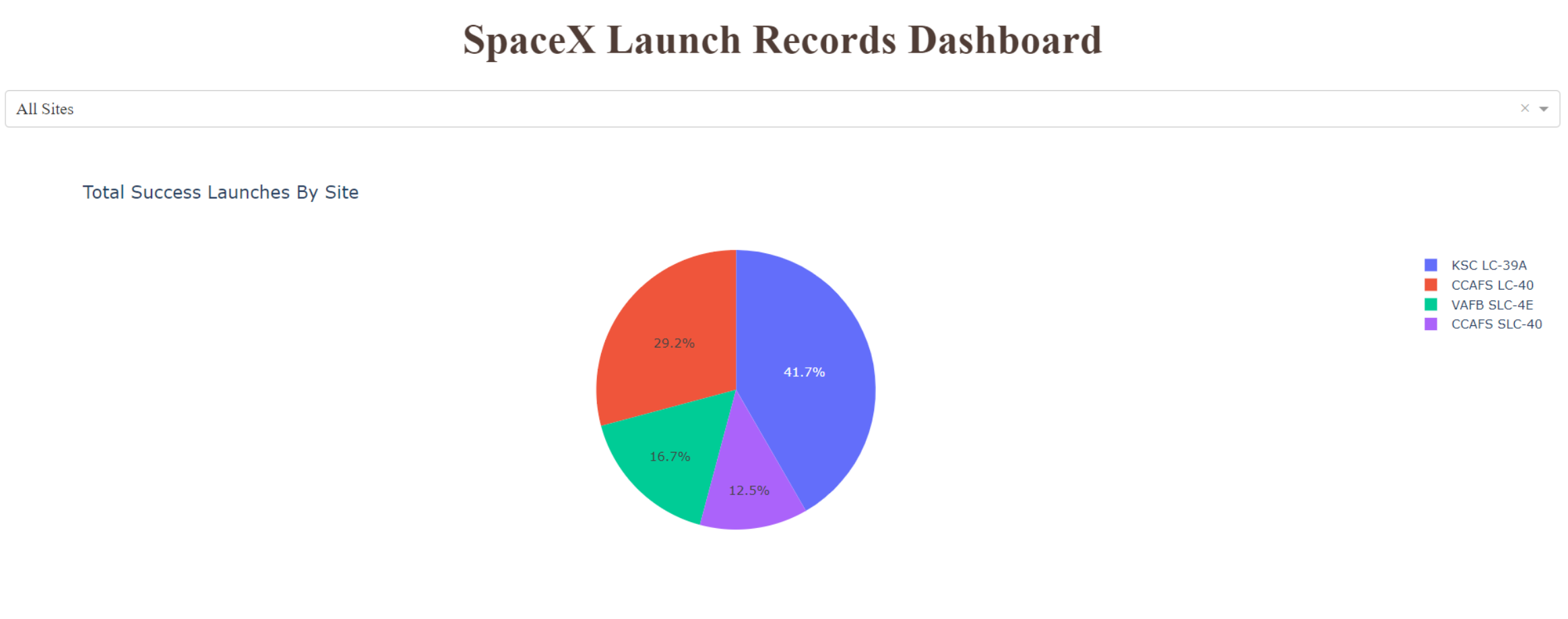
Folium Map : Launch site proximities



Launch site is close to coastal area (nearest distance : 0.8km)

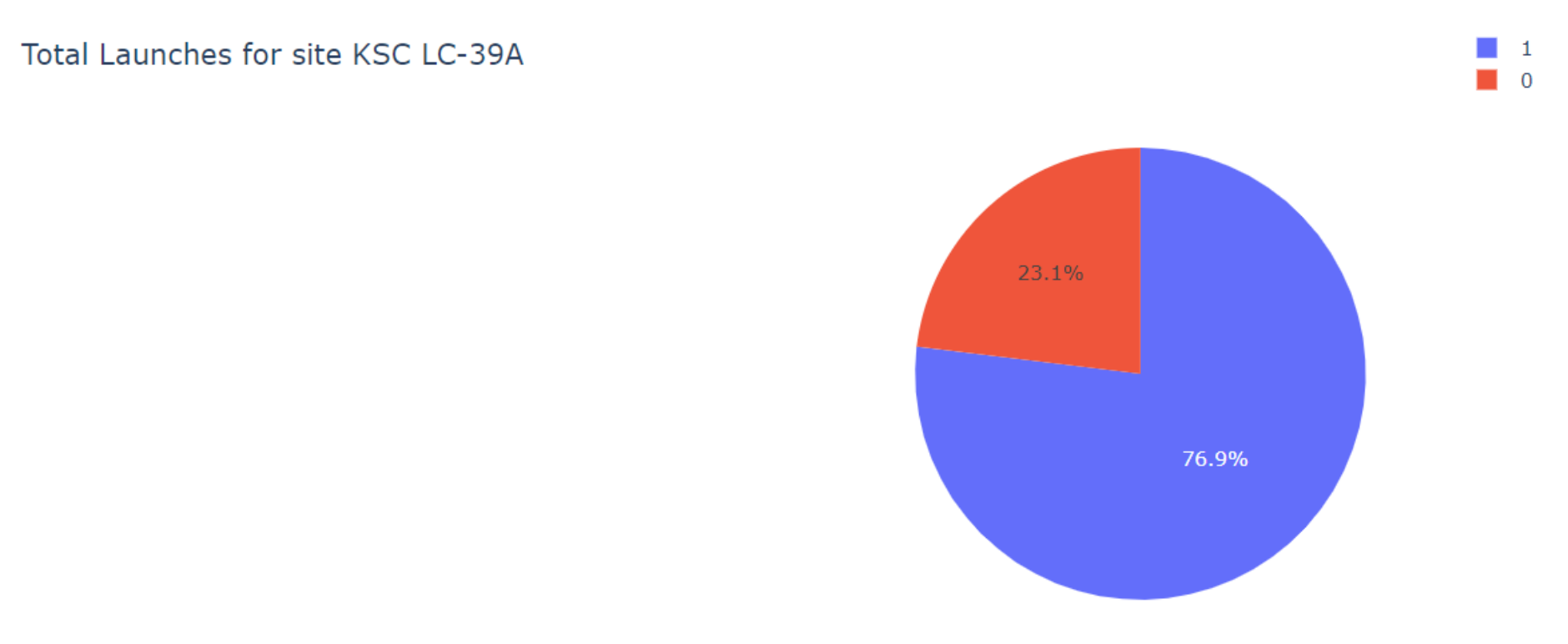

Section 4
Build a Dashboard with Plotly Dash

Dashboard Screenshot : Successful Launches by Site



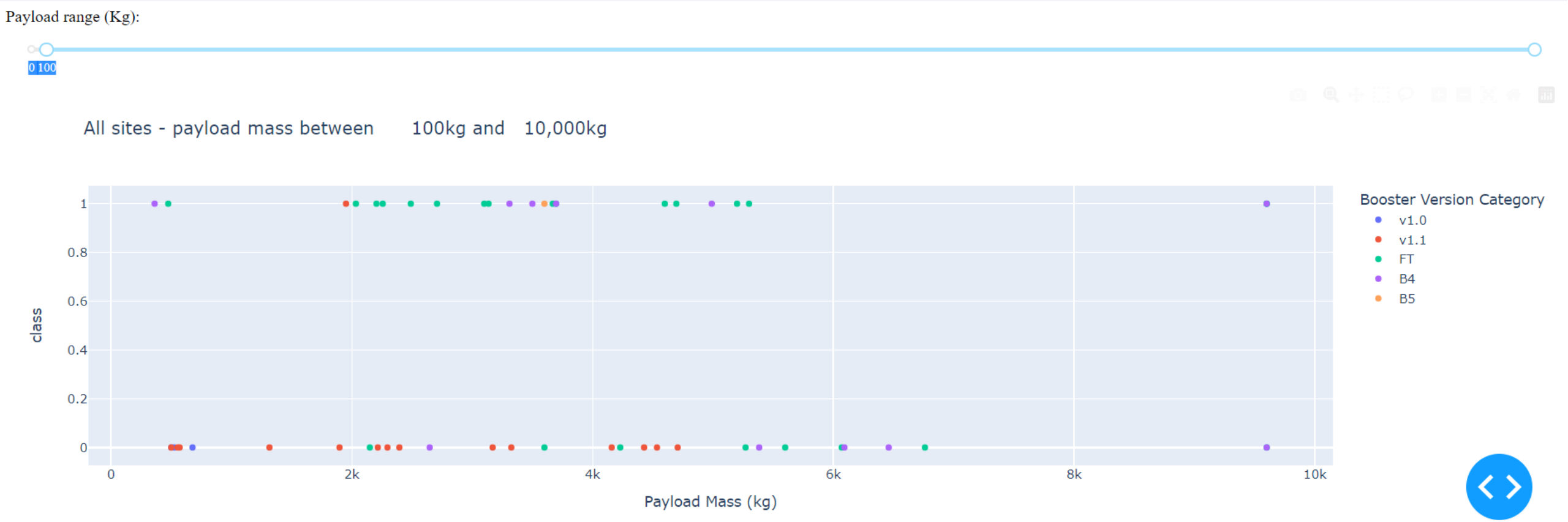
KSC LC-39A have highest success rate

Highest Successful Launch Site



- KSC LC-39A is the highest successful launch site with success rate of 76.9%

Payload vs. Launch Outcome



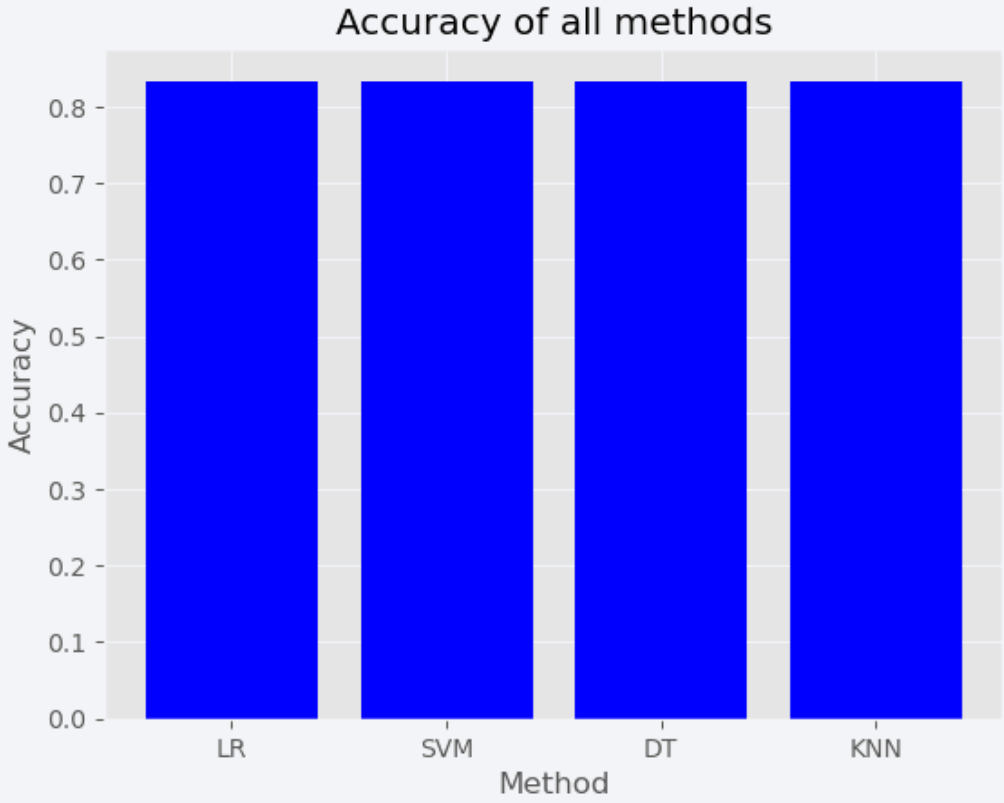
- Best combination for successful launch : payload <6k, booster version - FT

Section 5

Predictive Analysis (Classification)

Classification Accuracy

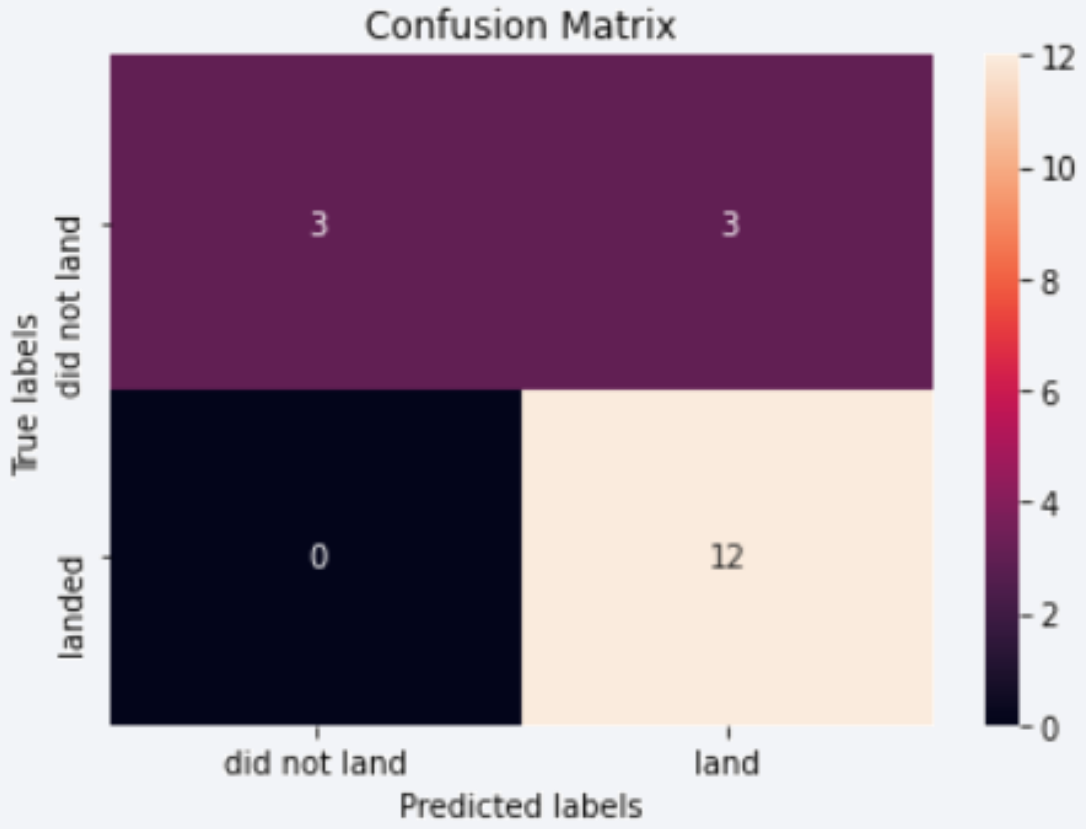
Model accuracy for all built classification models visualization:



All classification models have same accuracy i.e. **83.3%**

Confusion Matrix

Confusion Matrix for SVM model:



There is chance of false positive:

→ Out of 15 observations 3 have false positive.

Conclusions

- Trained classification model can predict the probability of landing first stage of a rocket with an accuracy of **83.3%.**
- Site KSC LC-39A have high landing success rate.
- Payload over 7000kg have high success rate.

Appendix

- Data from Space X was obtained from 2 sources:
  - https://api.spacexdata.com/v4/rockets/
  - https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches/

Thank you!