IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Suresh Tamang
16-08-2024

Outline

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

Executive Summary

- Following Methodologies used to Collect and analyze the Data
  - Web scraping and SpaceX API used to collect data
  - Exploratory Data Analysis(EDA)
  - Machine learning Models for predictions

- Summary of all results
  - Collected all data from public sources
  - Feature selection by EDA
  - Predictive ML model to choose predict the new data.

Introduction

- Project background:
  - Evaluate competebility of a company Space Y against Space X
  - The SpaceY wants to build better Rocket technology using falcon 9 landing Data

- Outcomes:
  - Visual representation of the Falcon 9 Launch Data
  - Visually represent the Data in the map for the precise location of launch station
  - Predicting the success/unsuccess of the new rocket based on features
  - Developing ML models to predict launch outcomes
  - Selecting best / Accurate Models
  - Predicting the success/unsuccess of the new rocket launch
  - Estimating the cost of rocket to launch

Section 1

Methodology

Methodology

- Data collection methodology:
  Data obtained from open sources:
  1. SpaceX API
  2. Web Scraping
- Perform Data Wrangling:
  -Landing outcome labeled success (1) or fai(0)l based on features

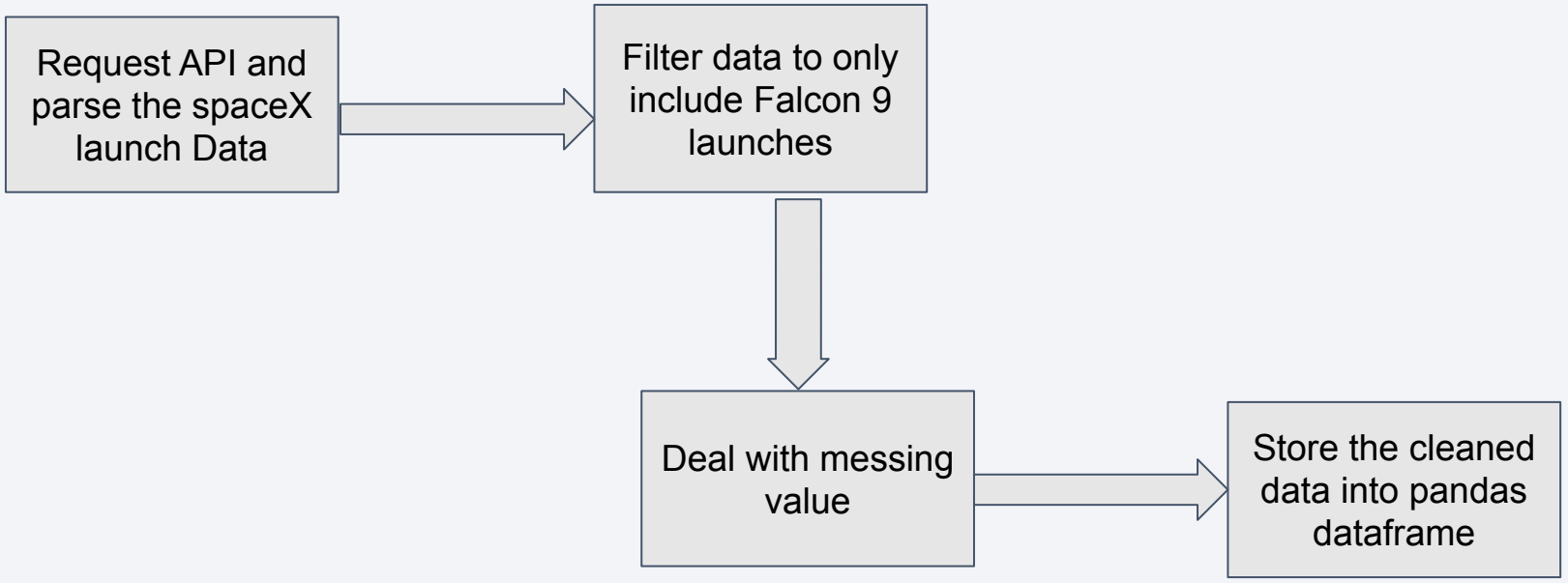
- Perform exploratory data analysis (EDA) using Visualization and SQL

- Perform interactive visual analytics using Folium and Plotly Dash

- Perform predictive analysis using classification models

Data Collection

The Data were collected using SpaceX API by making functions to store it in the pandas dataframe.

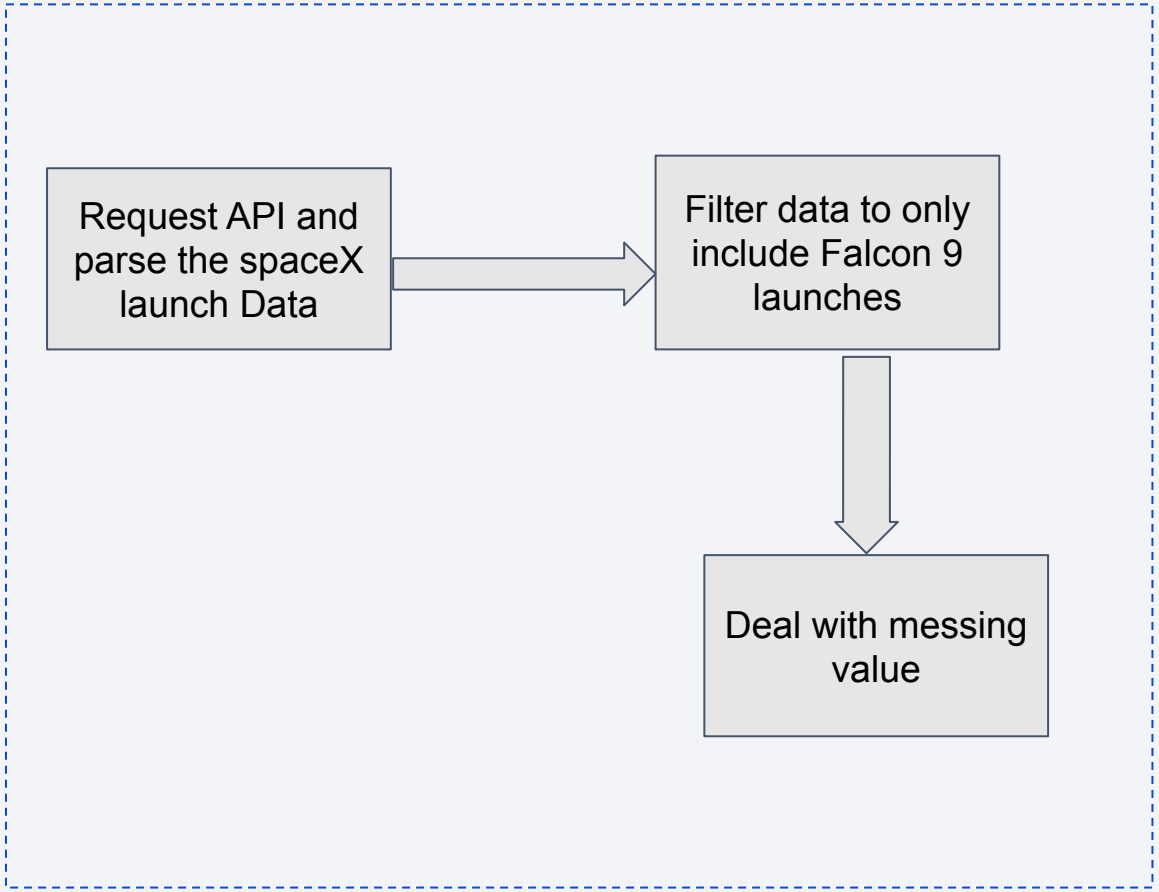


Data Collection – SpaceX API

- Open source SpaceX API

- GitHub URL

https://github.com/Suresh-Tamang/IBM-Data-Science-Capstone/blob/main/1.%20Data%20Collection%20API.ipynb



Data Collection - Scraping

- BeautifulSoup Python Library is used to web scrape the data from wikipedia.

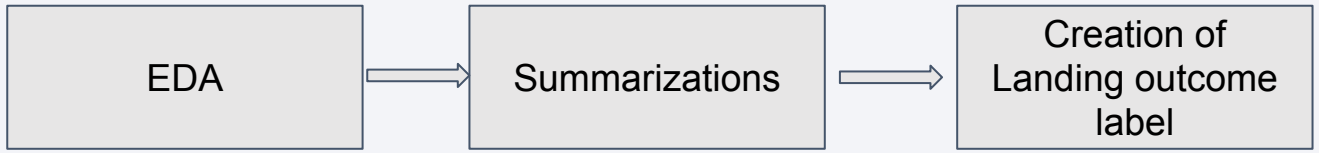
- GitHub URL

https://github.com/Suresh-Tamang/IBM-Data-Science-Capstone/blob/main/2.%20Data%20Webscrapping.ipynb

Data Wrangling

The data is processed so that there are no missing entries and categorical features are encoded using one-hot encoding.
An extra column called ‘Class’ is also added to the dataframe.The column ‘Class’ contains 0 if a given launch is failed and 1 if it is successful.

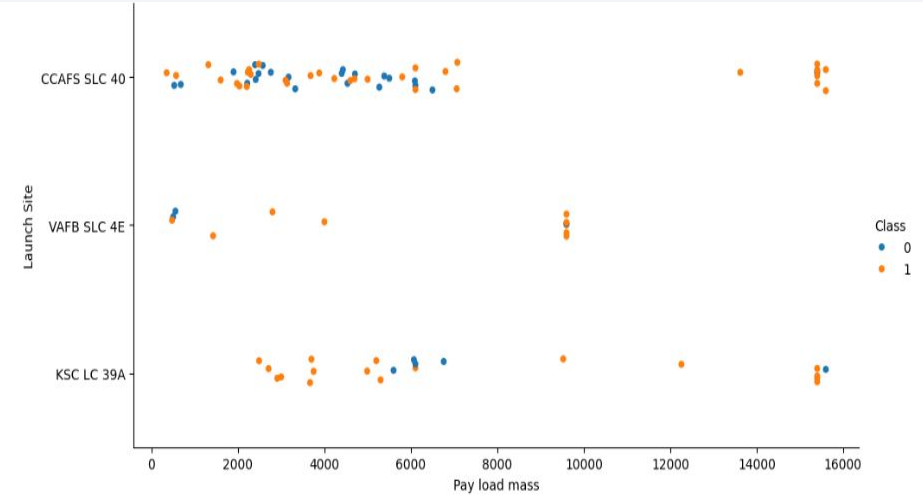
Flowchart:

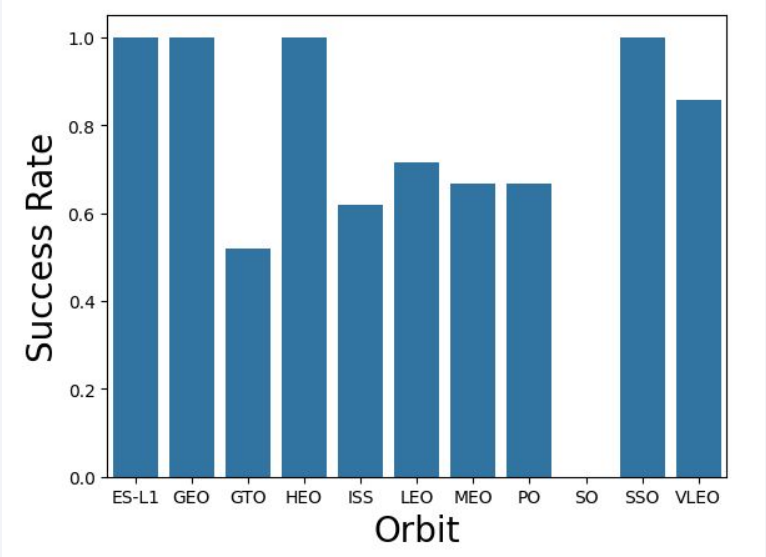


GitHub URL:
https://github.com/Suresh-Tamang/IBM-Data-Science-Capstone/blob/main/3.%20EDA.ipynb

EDA with Data Visualization

- Scatter plots and Barplots were used to visualize the relationship between various features and success rate





GitHub URL:

https://github.com/Suresh-Tamang/IBM-Data-Science-Capstone/blob/main/5.%20Data%20Visualization.ipynb

EDA with SQL

Performed SQL Queries:

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 successful landing outcome in ground pad was achieved
- 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. Use a subquery
- 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:

https://github.com/Suresh-Tamang/IBM-Data-Science-Capstone/blob/main/4.EDA%20With%20SQL.ipynb

Build an Interactive Map with Folium

- Markers, circles , lines and marker clusters were used with Folium Maps
  - The objects are added in the maps to point out the locations of the launch sites.

GitHub URL:

https://github.com/Suresh-Tamang/IBM-Data-Science-Capstone/blob/main/6.%20Interactive%20Visual%20analytics.ipynb

Build a Dashboard with Plotly Dash

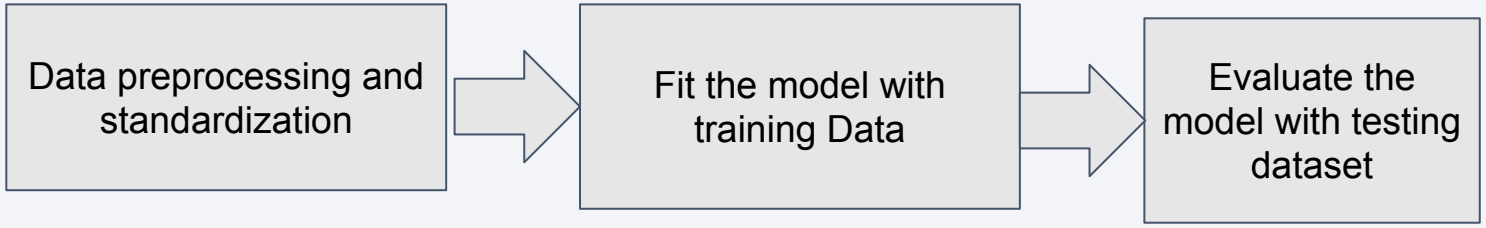
• The following graphs and plots were used to visualize data • Percentage of launches by site
• Payload range
• This combination allowed to quickly analyze the relation between payloads and launch sites, helping to identify where is best place to launch according to payloads.

GitHub URL:
https://github.com/Suresh-Tamang/IBM-Data-Science-Capstone/blob/main/7.%20spaceXDashApp.py

Predictive Analysis (Classification)

Classification Models:

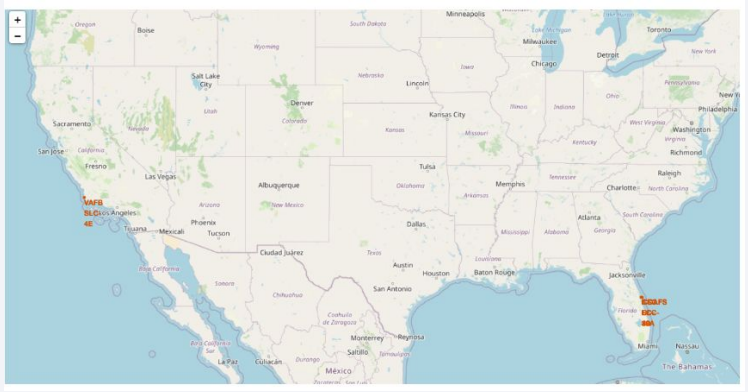
- Logistic regression
- Support Vector Machine
- Decision Tree
- K-Nearest Neighbors



GitHub URL:

https://github.com/Suresh-Tamang/IBM-Data-Science-Capstone/blob/main/8.%20Machine%20Learning%20Prediction.ipynb

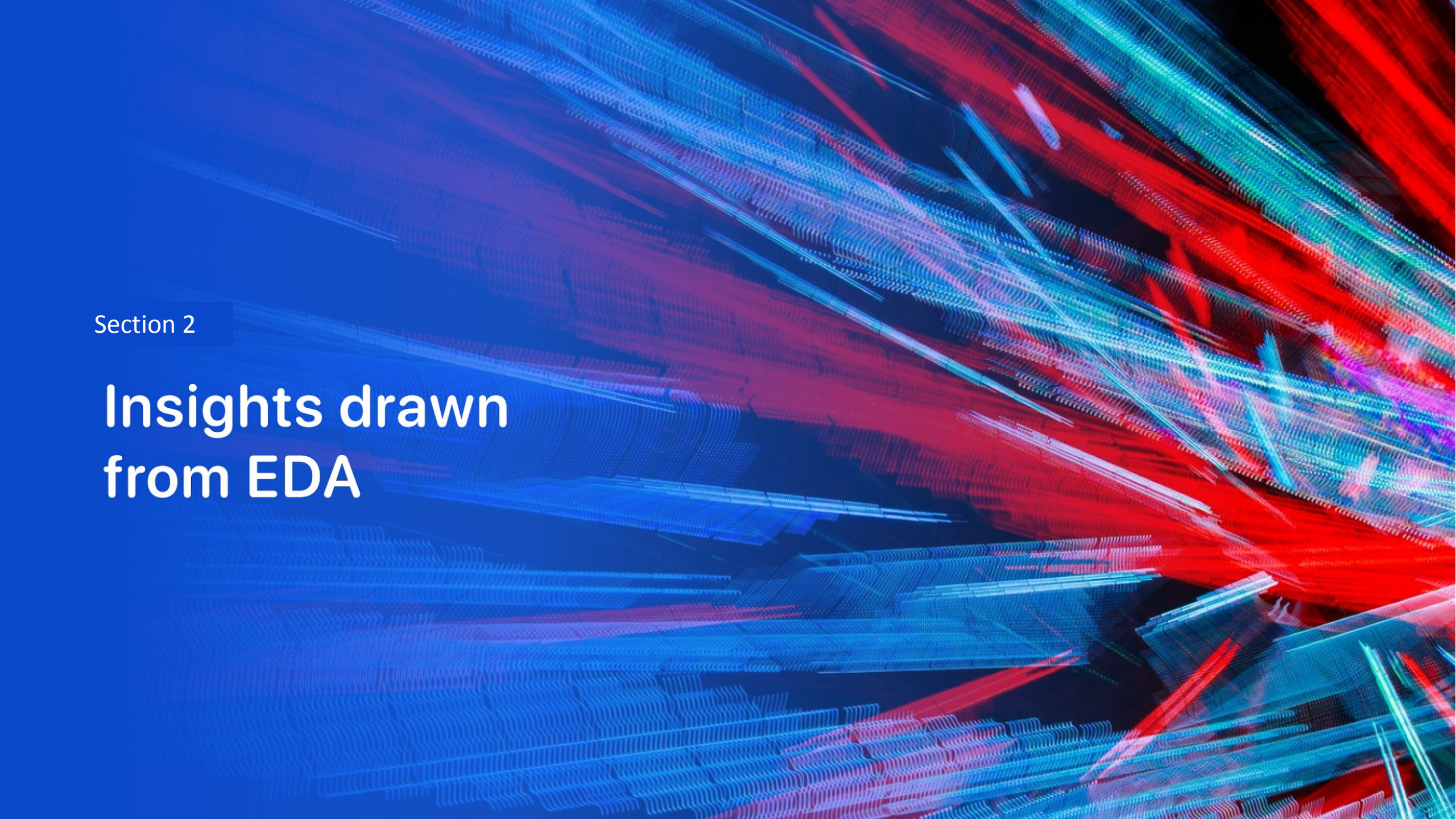
Results

- Exploratory data analysis results
  - There are 4 different launch sites
  - 99 times falcon 9 is launched successfully
  - Two booster versions failed at landing in drone ships in total
  - The number of landing outcomes became better years passed
- Interactive analytics demo in screenshots
- 

- Predictive analysis results

Results

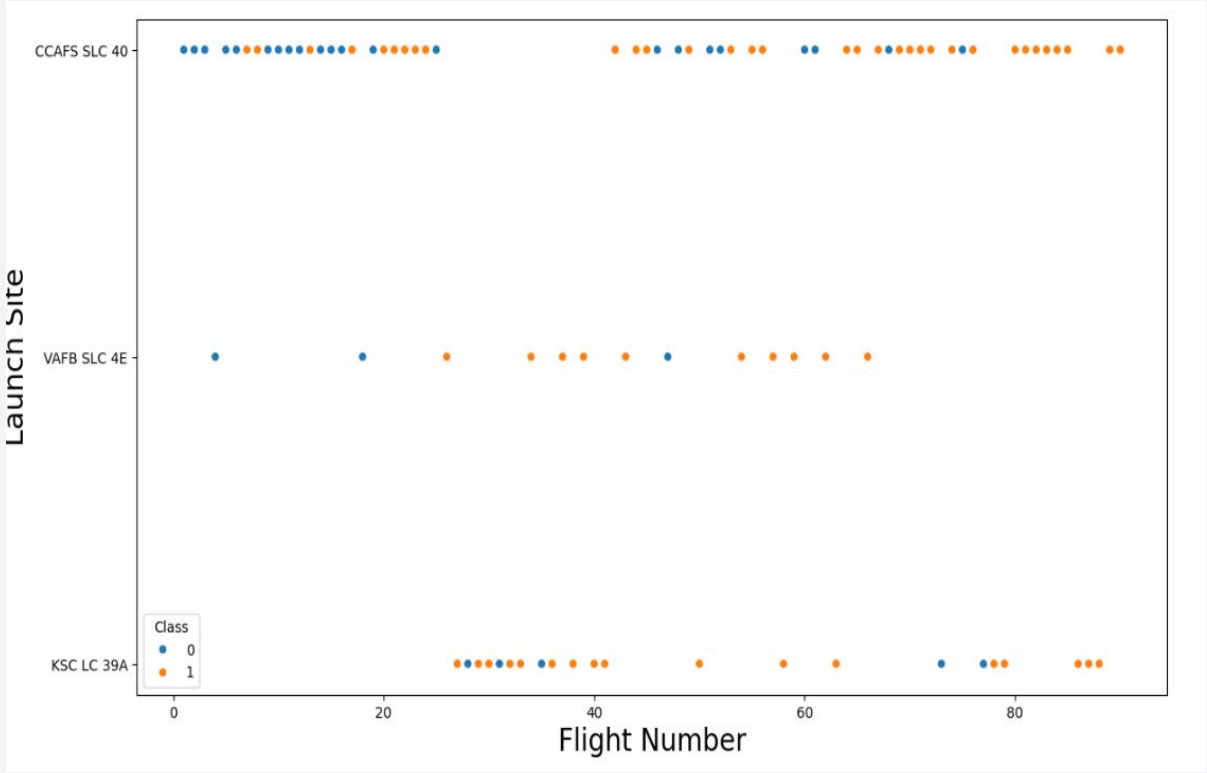
- Predictive analysis results
- Logistic regression (Accuracy 83.33%)
- Support Vector Machine(Accuracy 83.33%)
- Decision Tree (Accuracy 88%)
- K-Nearest Neighbors(Accuracy 85%)


Section 2
Insights drawn from EDA

Flight Number vs. Launch Site

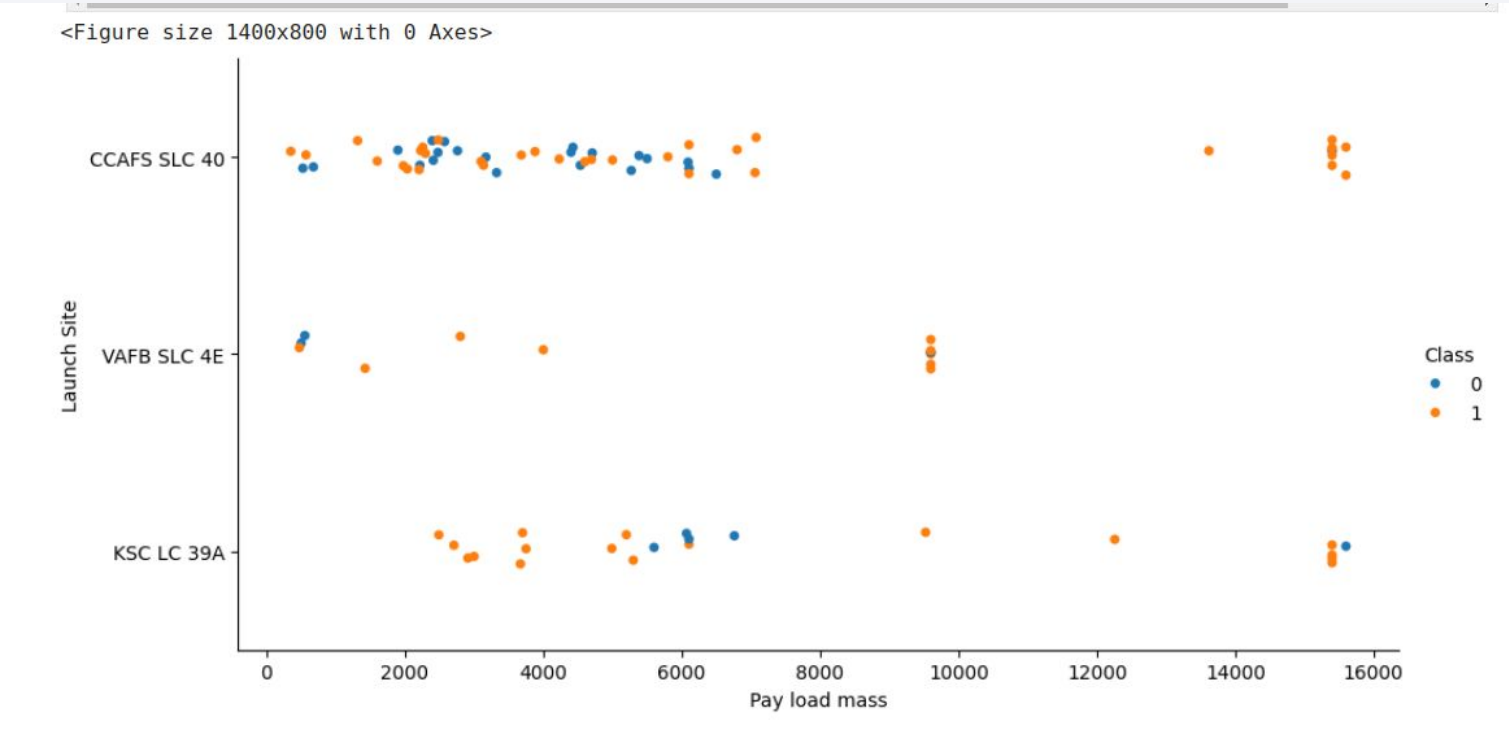
- scatter plot of Flight Number vs. Launch Site

The CCAFS SLC 40 launch site is most used and success rate is high in this site.



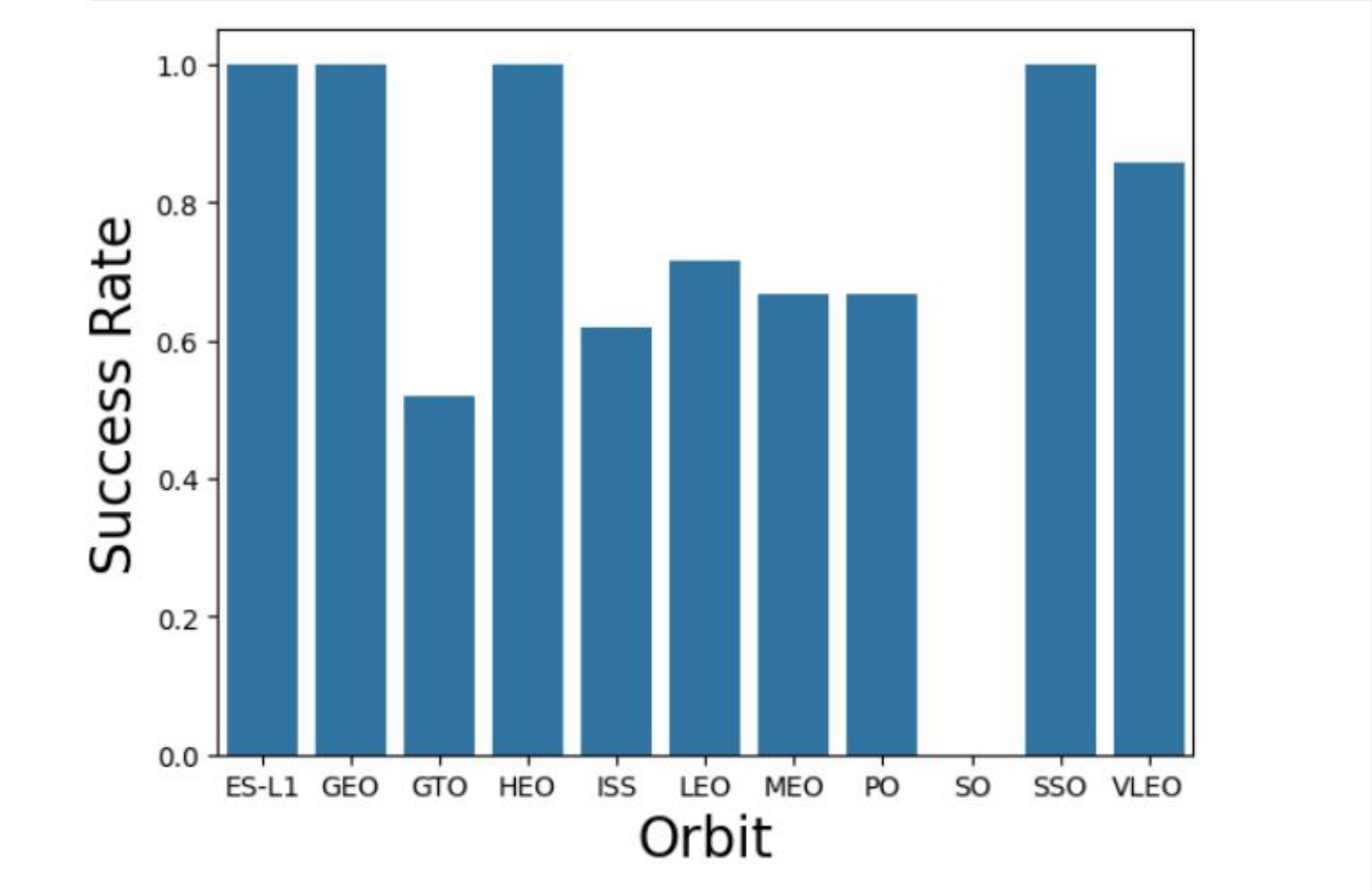
Payload vs. Launch Site

- Scatter plot of Payload vs. Launch Site
- The most Heaviest Rockets are launched from Launch Site CCAFS SLC 40.
- Success rate is high in that launch site



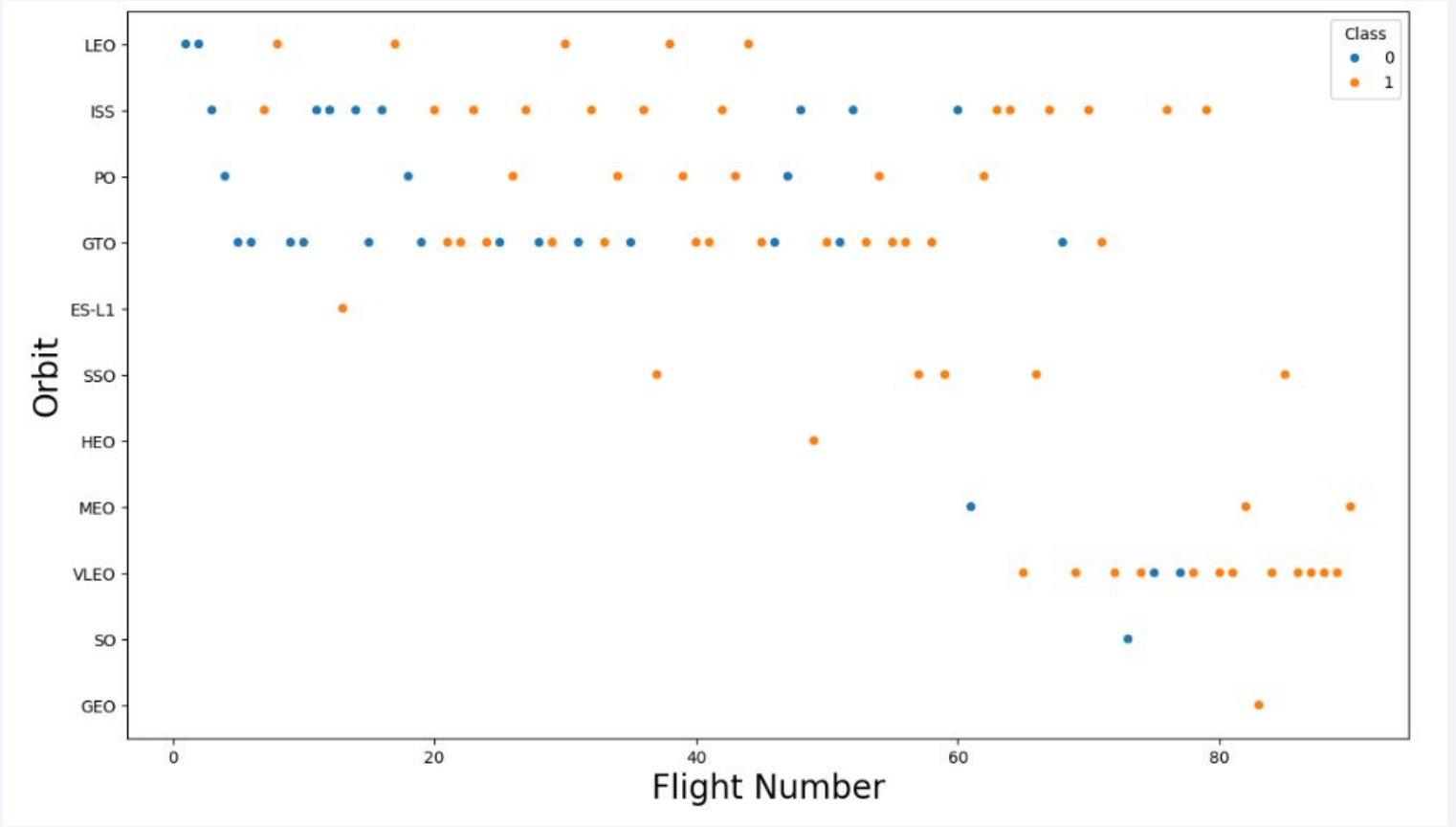
Success Rate vs. Orbit Type

- A bar chart for the success rate of each orbit type
- The four Orbit have 100% success rate. Those orbits are ES-L1,GEO,HEO, SSO



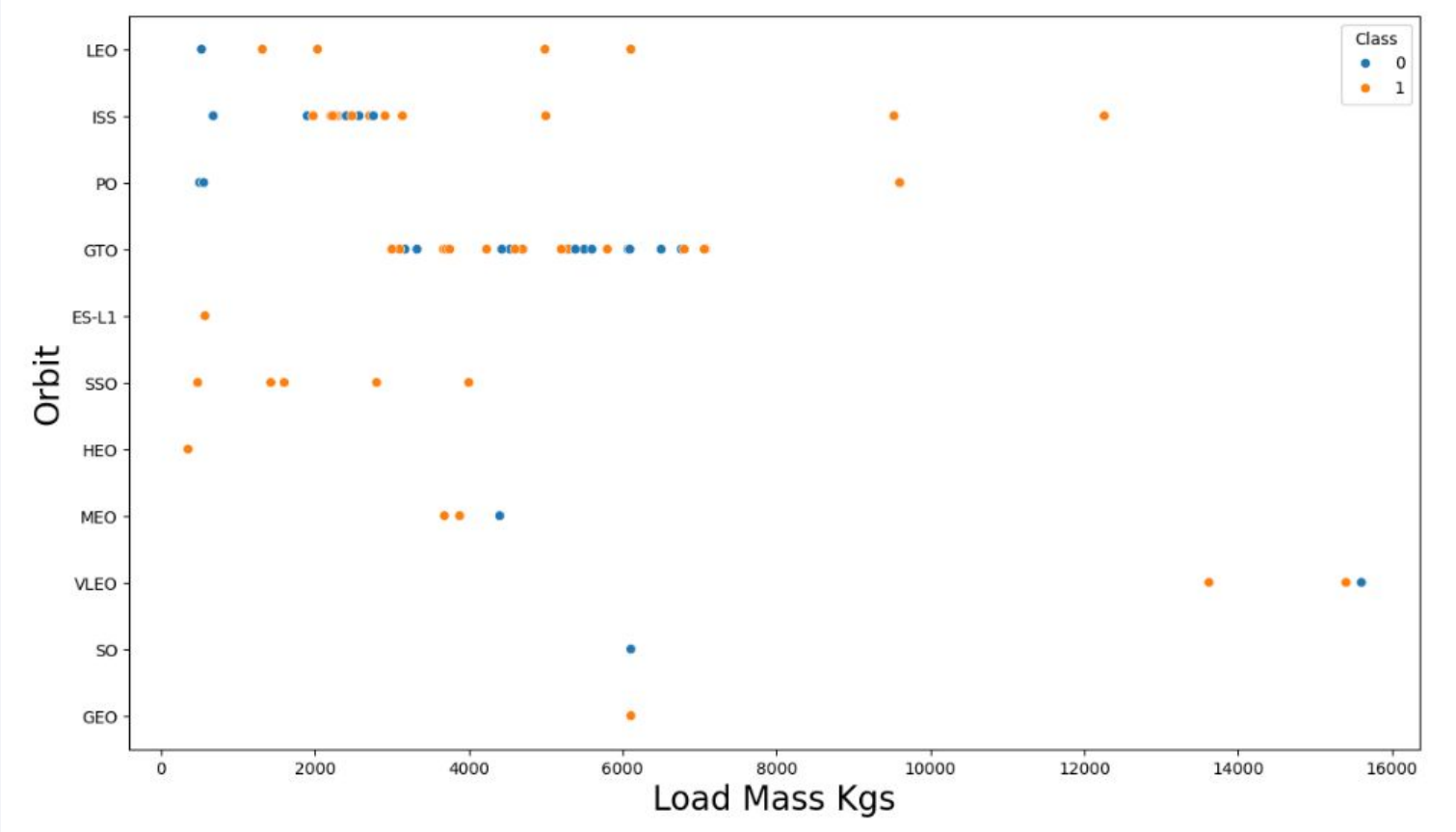
Flight Number vs. Orbit Type

- A scatter point of Flight number vs. Orbit type
- Success rate improved over new launches.



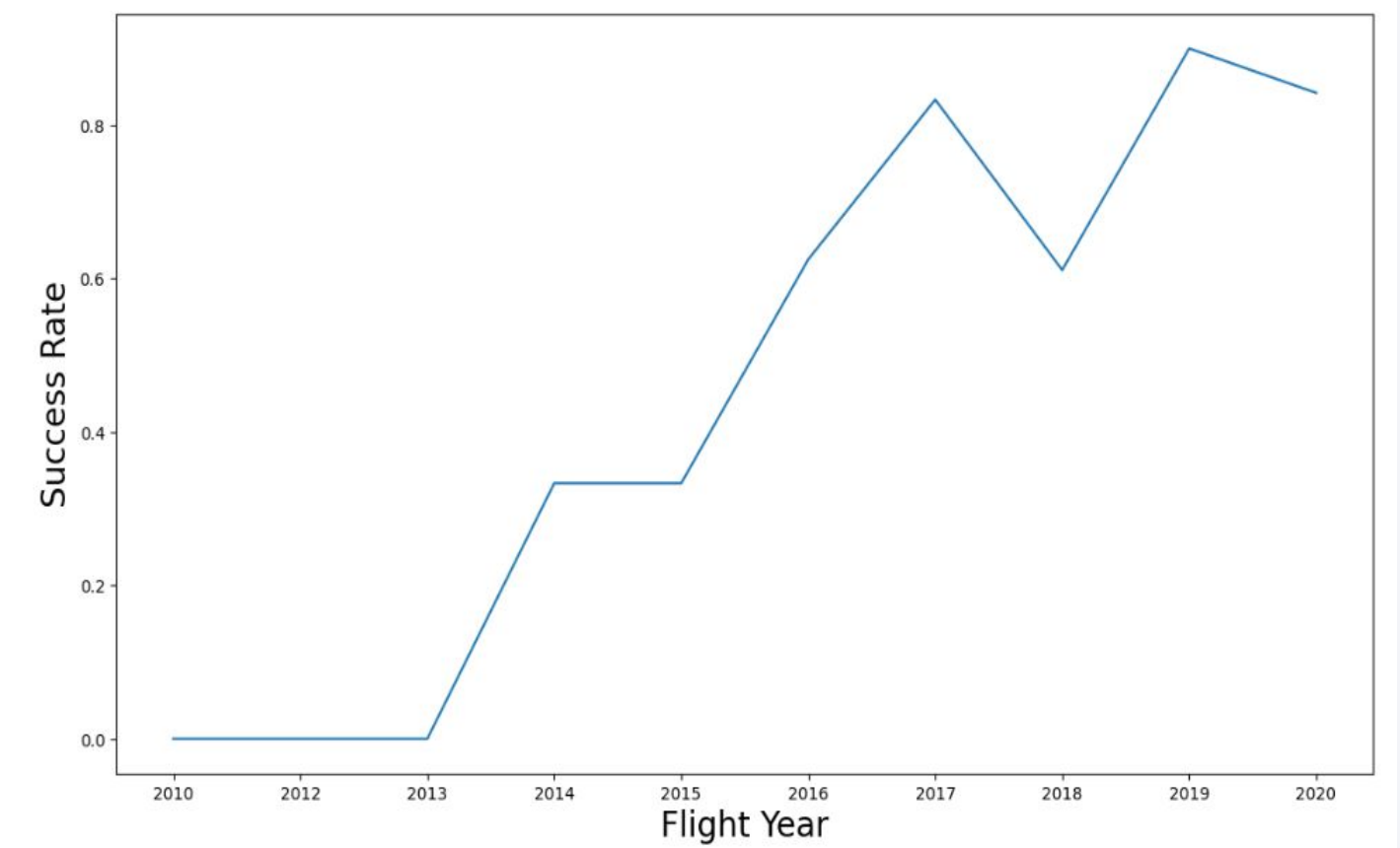
Payload vs. Orbit Type

- A scatter point of payload vs. orbit type
- Most number of rockets are launched in the orbit GTO
- No relation between Payload mass and GTO orbit



Launch Success Yearly Trend

- A line chart of yearly average success rate
- First Three years have 0% success, but from 2014 the success rates are significantly improved.



All Launch Site Names

There are four Unique Launch Sites:

]:

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

Launch Site Names Begin with 'CCA'

5 records where launch sites begin with `CCA`

Those records are from year 2010 to 2013 .

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

Total Payload Mass

Calculate the total payload carried by boosters from NASA

The total payload carried by boosters from NASA is 45596 Kgs

Total Payload Mass(Kgs)	Customer
45596	NASA (CRS)

Average Payload Mass by F9 v1.1

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

Payload Mass Kgs	Customer	Booster_Version
2928.4	SES	F9 v1.1

First Successful Ground Landing Date

The first successful landing outcome on ground pad is 2015-12-22.

min(Date)	Customer
2015-12-22	Orbcomm

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

```
om SPACEXTBL where Landing_Outcome = 'Success (drone ship)' and PAYLOAD_MASS__KG_>4000 and PAYLOAD_MASS__KG_<6000;
```

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

Payload
JCSAT-14
JCSAT-16
SES-10
SES-11 / EchoStar 105

Total Number of Successful and Failure Mission Outcomes

The total number of successful and failure mission outcomes

```
In [26]: %sql select Mission_Outcome,count("Mission_Outcome") as Total from SPACEXTBL group by "Mission_Outcome";
 * sqlite:///my_data1.db
Done.
```

Out[26]:

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

Boosters Carried Maximum Payload

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

Booster_Version	Payload	PAYLOAD_MASS__KG_
F9 B5 B1048.4	Starlink 1 v1.0, SpaceX CRS-19	15600
F9 B5 B1049.4	Starlink 2 v1.0, Crew Dragon in-flight abort test	15600
F9 B5 B1051.3	Starlink 3 v1.0, Starlink 4 v1.0	15600
F9 B5 B1056.4	Starlink 4 v1.0, SpaceX CRS-20	15600
F9 B5 B1048.5	Starlink 5 v1.0, Starlink 6 v1.0	15600
F9 B5 B1051.4	Starlink 6 v1.0, Crew Dragon Demo-2	15600
F9 B5 B1049.5	Starlink 7 v1.0, Starlink 8 v1.0	15600
F9 B5 B1060.2	Starlink 11 v1.0, Starlink 12 v1.0	15600
F9 B5 B1058.3	Starlink 12 v1.0, Starlink 13 v1.0	15600
F9 B5 B1051.6	Starlink 13 v1.0, Starlink 14 v1.0	15600
F9 B5 B1060.3	Starlink 14 v1.0, GPS III-04	15600
F9 B5 B1049.7	Starlink 15 v1.0, SpaceX CRS-21	15600

2015 Launch Records

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

substr(Date,7,4)	substr(Date,4,2)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Mission_Outcome	Landing_Outcome
1-10	5-	F9 v1.1 B1012	CCAFS LC-40	SpaceX CRS-5	2395	Success	Failure (drone ship)
4-14	5-	F9 v1.1 B1015	CCAFS LC-40	SpaceX CRS-6	1898	Success	Failure (drone ship)

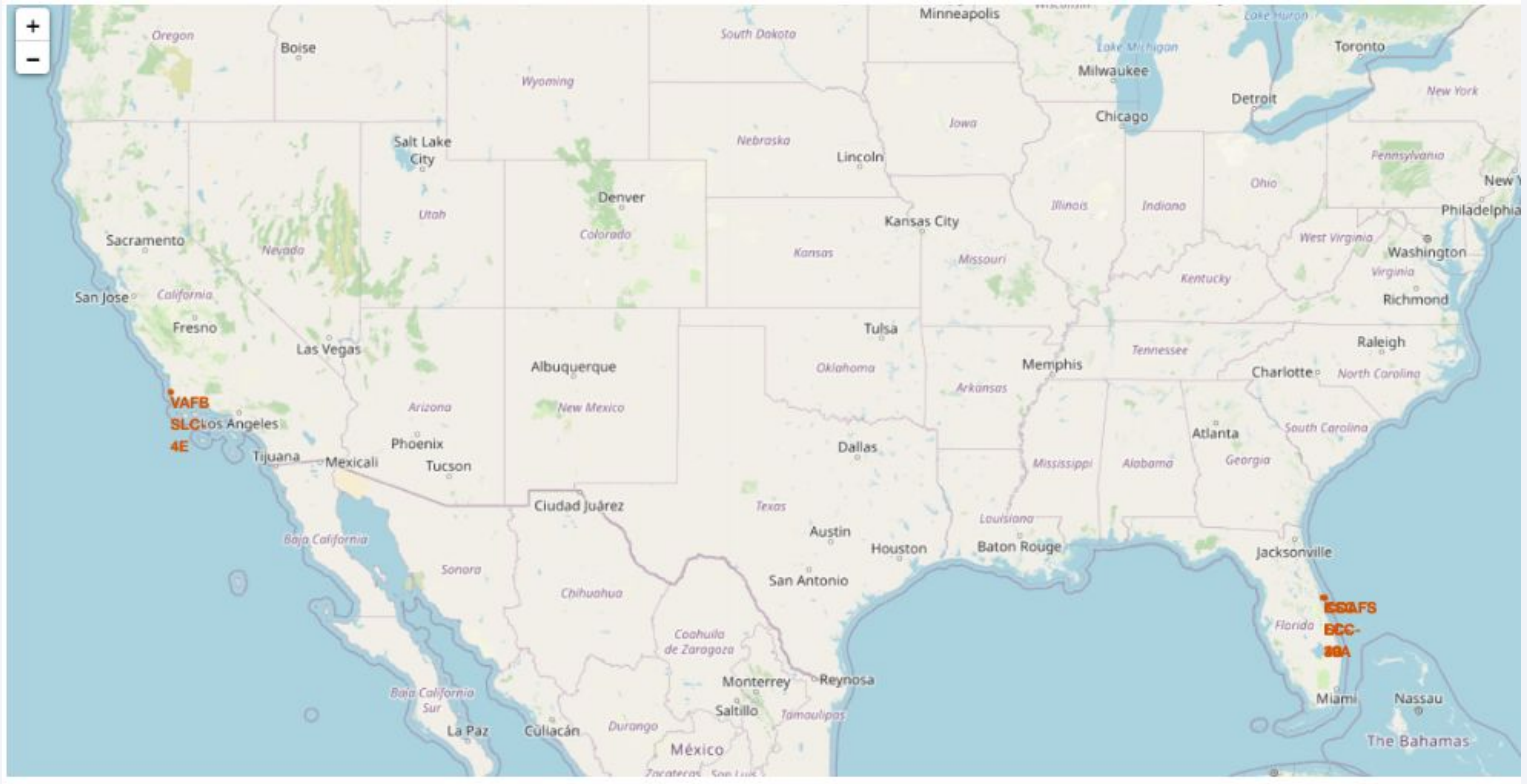
Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

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	total
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

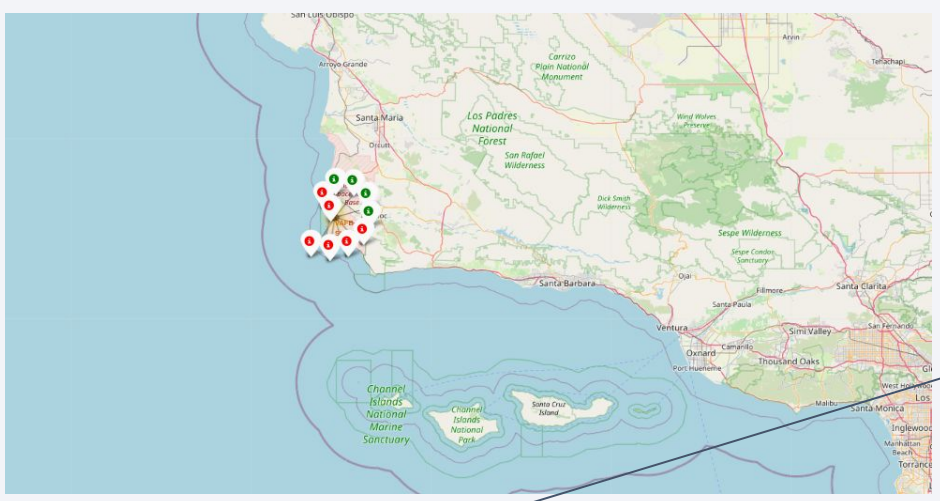

Section 3
Launch Sites
Proximities Analysis

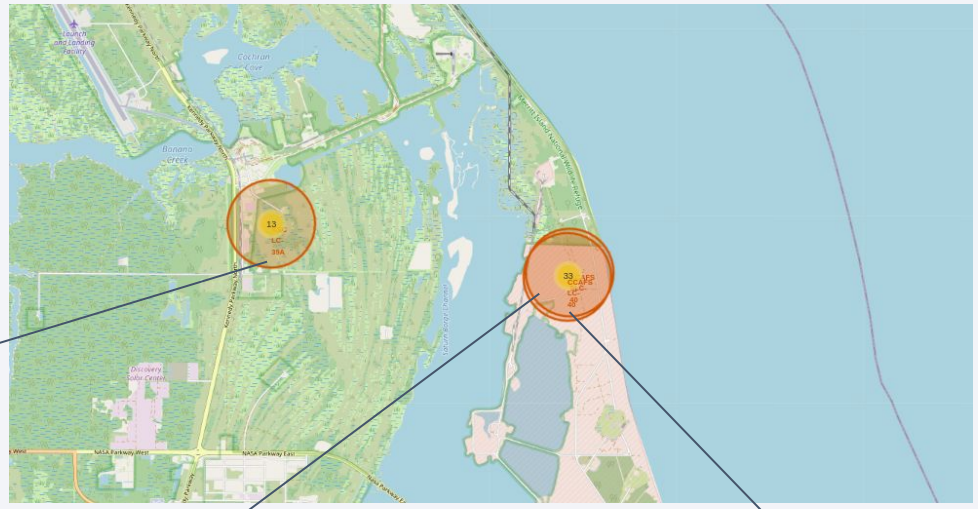
All launch sites

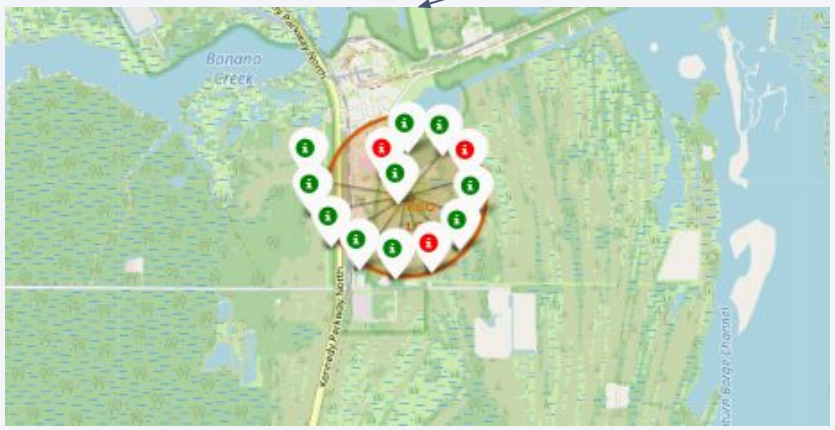


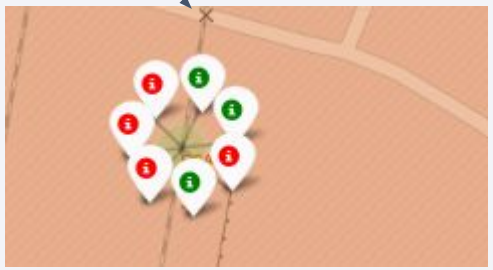
- Launch sites are near sea
- Launch sites are not far from roads and railroads

Success/Failed Launches For Each Site

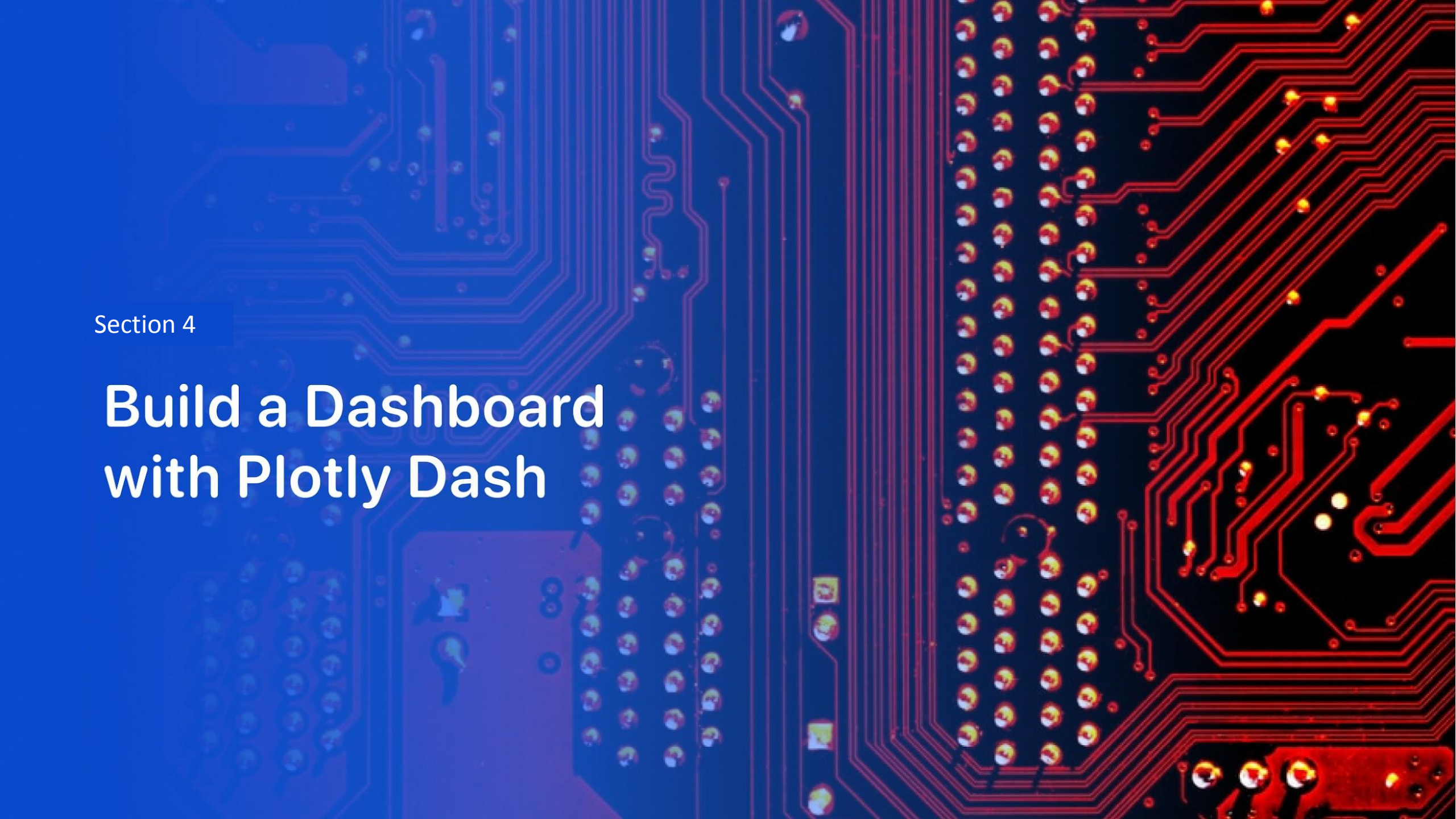




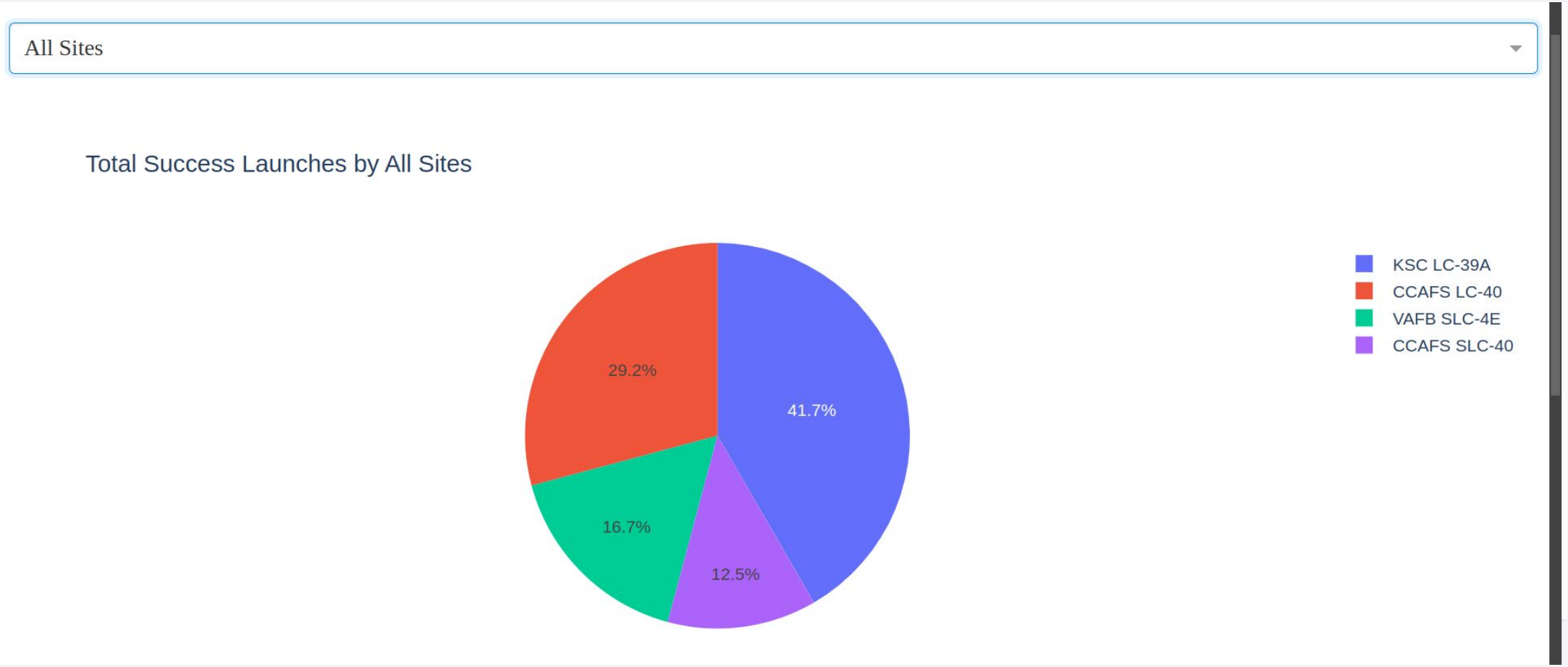




Distances Between a Launch

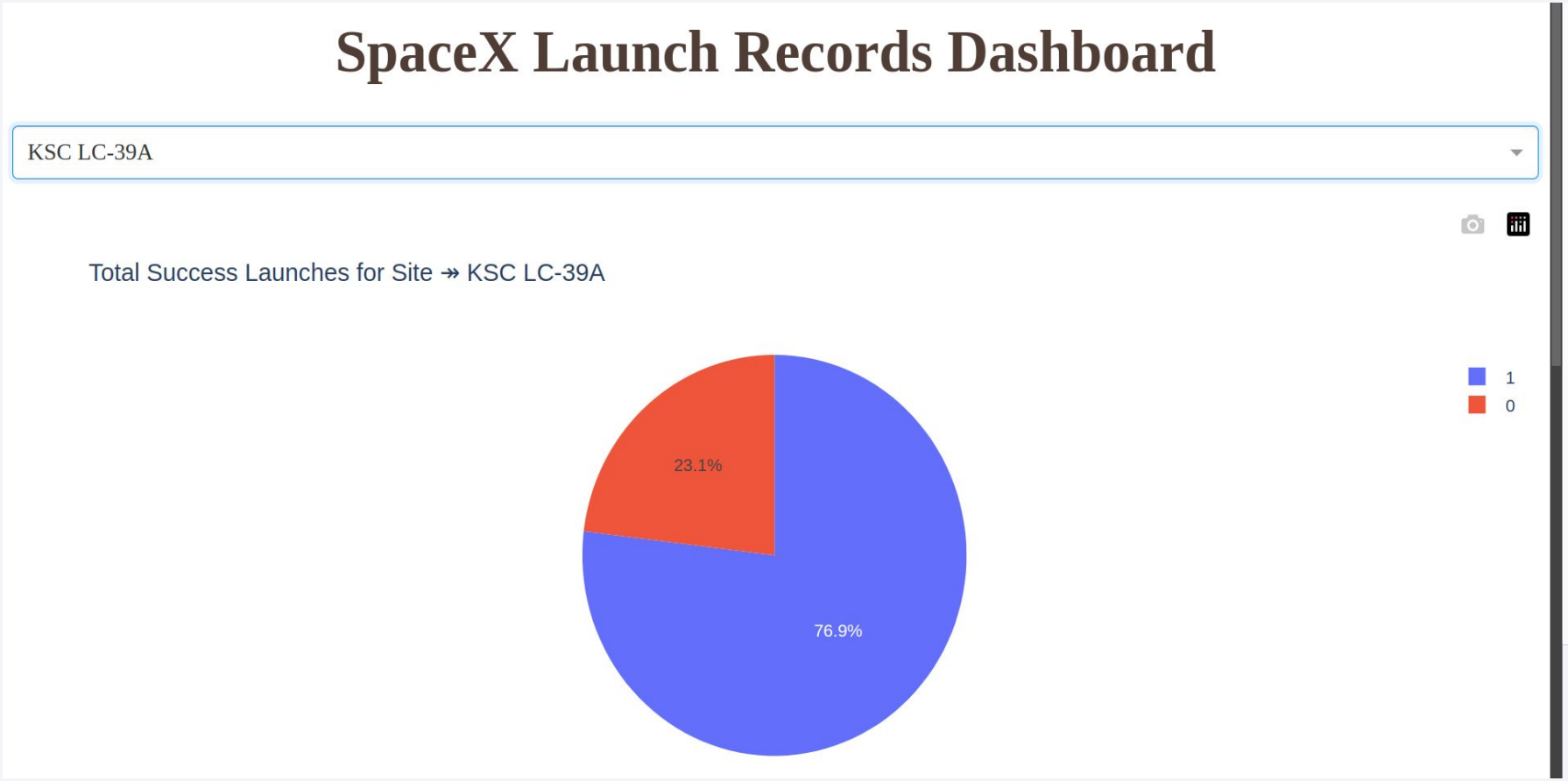

Section 4
Build a Dashboard with Plotly Dash

Successful Launches by Site



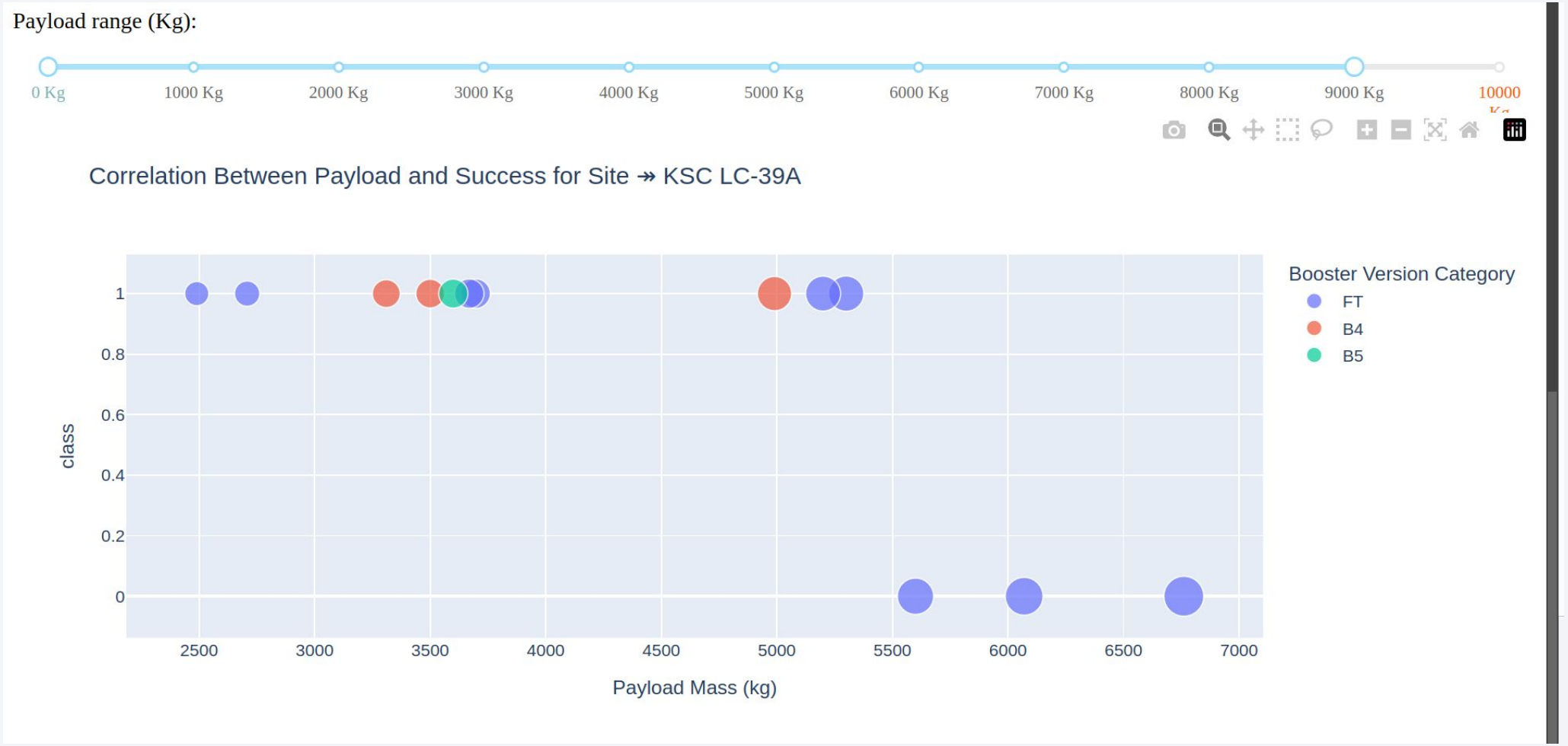
KSC LC-39A has highest rate of success.

Launch Success Ratio for KSC LC-39A



Around 76.9% launches are successful

Payload vs. Launch Outcome



Section 5

Predictive Analysis (Classification)

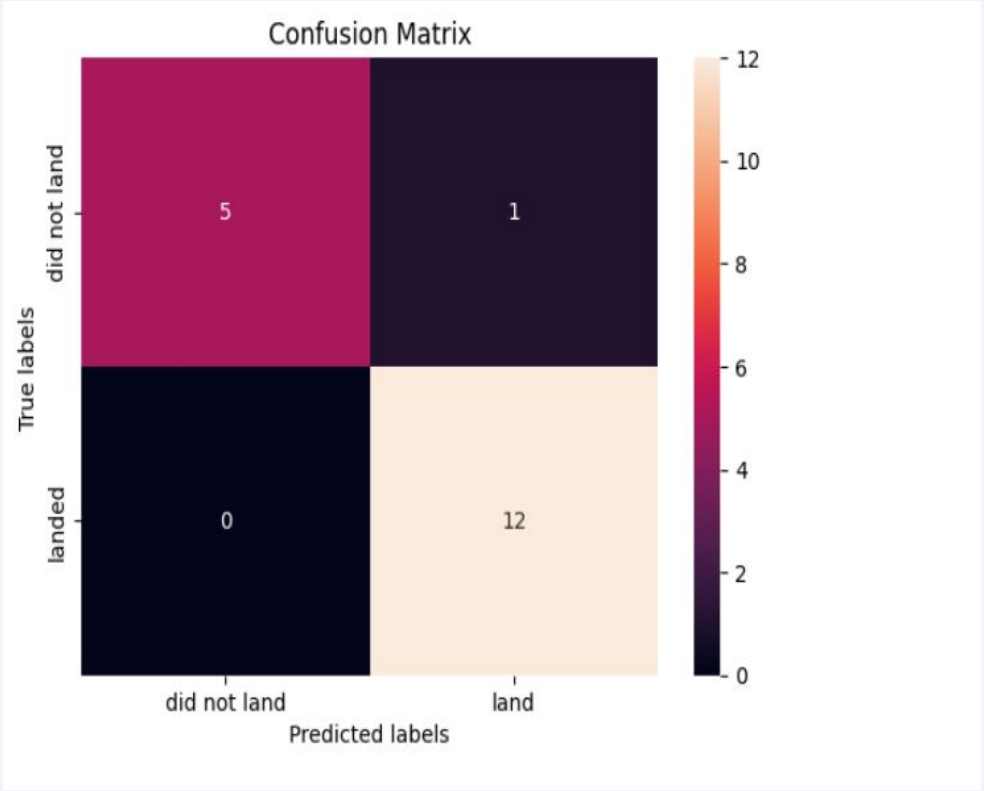
Classification Accuracy

	Best scores
Logistic regresssion	0.846429
SVM	0.848214
Decision tree	0.887500
KNN	0.848214

- The best Classification method is Decision Tree for the model prediction.

Confusion Matrix

True Positives (TP): 12 - The model correctly predicted 12 instances of landing.
True Negatives (TN): 5 - The model correctly predicted 5 instances of not landing.
False Positives (FP): 1 - The model incorrectly predicted 1 instance of landing when it was actually a non-landing case.
False Negatives (FN): 0 - The model did not incorrectly predict any instances of not landing.



Conclusions

- Data collected from open sources : SpaceX API, Wikipedia
- KSC LC-39A is the best launch site among four launch sites
- 0-6000 kg launches are risky
- Success rate improves over time among all launch sites
- Lunch Sites are close to sea, not over populated area
- All ML algorithms seems to work fine, need more fine tuning to choose perfect model

Appendix

- IBM Data Science Professional Certification Course

Thank you!