

Winning Space Race with Data Science

YU-CHENG LIAO
30 Nov 2022



Outline

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

Executive Summary

- To understand the price of launching a rocket, the purpose of this analysis was to understand the previous launching outcomes from Space X and gave a detailed report about the price prediction and recommendation.
- This analysis was applied with four models for prediction and evaluation. The decision tree had the highest accuracy during the evaluation.
- The recommended site for launching was KSC LC-39A.
- When payload mass was over 8000 kilograms, the rockets had higher rates of landing successfully.

Introduction

Background

- SpaceX has decreased the huge cost down to 62M USD for a rocket launch (other providers cost upwards of 165M USD) due to the ability of recovery from the used first stage of Falcon 9 rocket.
- Space Y wants to compete Space X

Objectives

- To predict the price of each launch in terms of Space X's experiences if the first stage lands successfully .

Section 1

Methodology

Methodology

Executive Summary

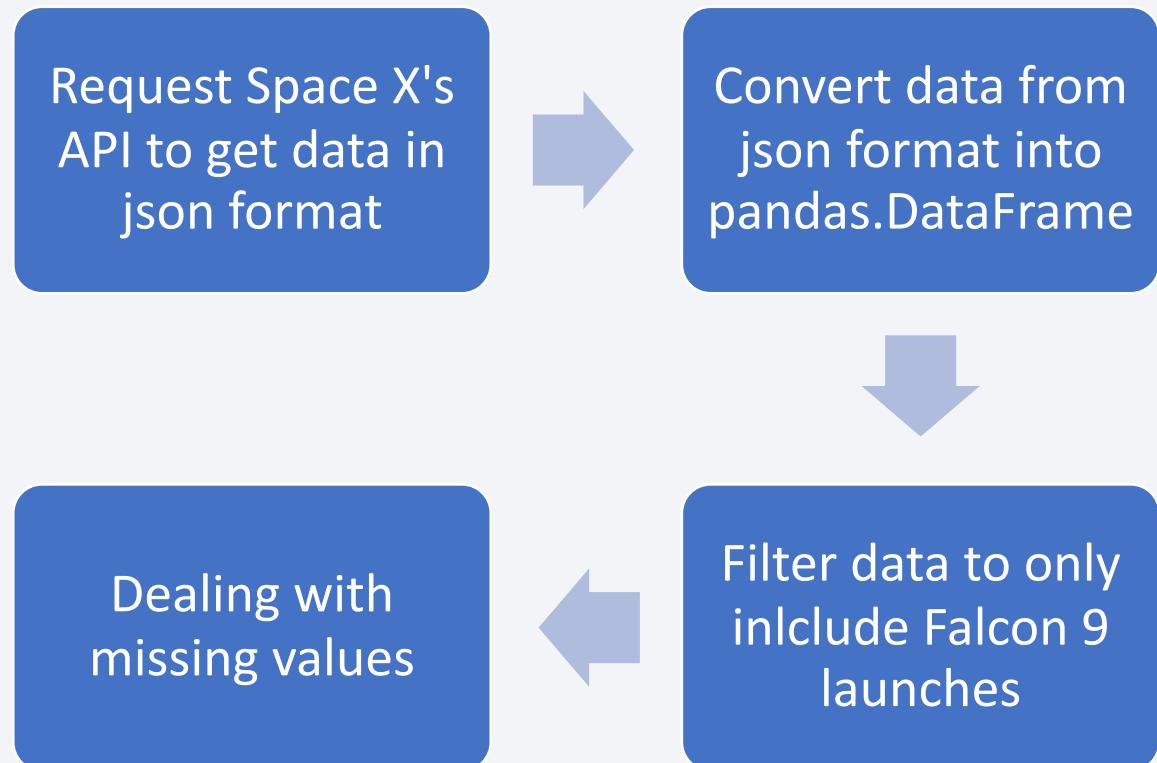
- Data collection methodology:
 - By requesting [Space X's API](#) and the web scraping the records of [Space X's launches](#)
- Perform data wrangling
 - Identified the successful landing outcomes and features
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
 - Used different models to predict, such as Logistic Research, Decision Tree, SVM, KNN, by normalizing data, splitting dataset into training data and testing data, and then cross-validating the accuracy of each model

Data Collection

- Used the request method to select the rockets information and the web scraping method to collect the launches data by rocket information.
 - Rockets information from Space X:
Flight Number, Date, Booster Version, Payload Mass, Orbit, Launch Site, Outcome, Flights, Grid Fins, Reused, Legs, Landing Pad, Block, Reused Count, Serial, Longitude, Latitude
 - Historical launch records records from Wikipedia:
Flight No., Date and time, Launch site, Payload, Payload mass, Orbit, Customer, Launch outcome

Data Collection – SpaceX API

- Space X offers APIs for the public to request and collect information.
- [GitHub Link](#)



Data Collection - Scraping

- The specific rocket data of historical launches could be collect from Wikipedia by web scraping.
- [GitHub Link](#)

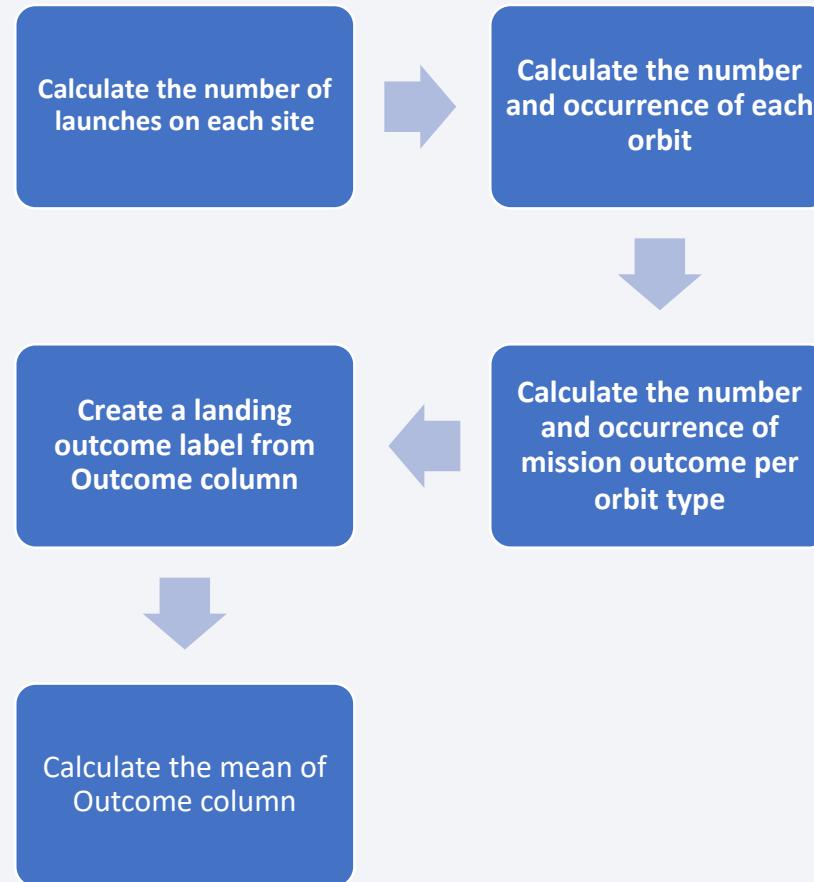
Request the website and scratch all information

Analyse the format of information with BeautifulSoup

Divide the data from html format into DataFrame

Data Wrangling

- Calculate the frequencies and numbers of each column and compute the mean of the Outcome column
- [GitHub Link](#)



EDA with Data Visualization

- To explore the data primitively, a few visualization methods were applied, such as scatter plots, bar plots, and line plots. These plots could be used to investigate the relations of the features among the dataset.
- [GitHub Link](#)

EDA with SQL

- Steps of SQL queries:
 - names of the unique launch sites in the space mission
 - 5 records where launch sites begin with the string 'CCA'
 - the total payload mass carried by boosters launched by NASA (CRS)
 - The average payload mass carried by booster version F9 v1.1
 - Date list 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 ft.
 - The total number of successful and failure mission outcomes
 - The names of the booster versions which have carried the maximum payload mass
 - The records which will display the month names, failure landing outcomes in drone ship ,booster versions, launch site for the months in year 2015.
 - The rank of the count of successful landing outcomes between the date 04-06-2010 and 20-03-2017 in descending order.
- [GitHub Link](#)

Build an Interactive Map with Folium

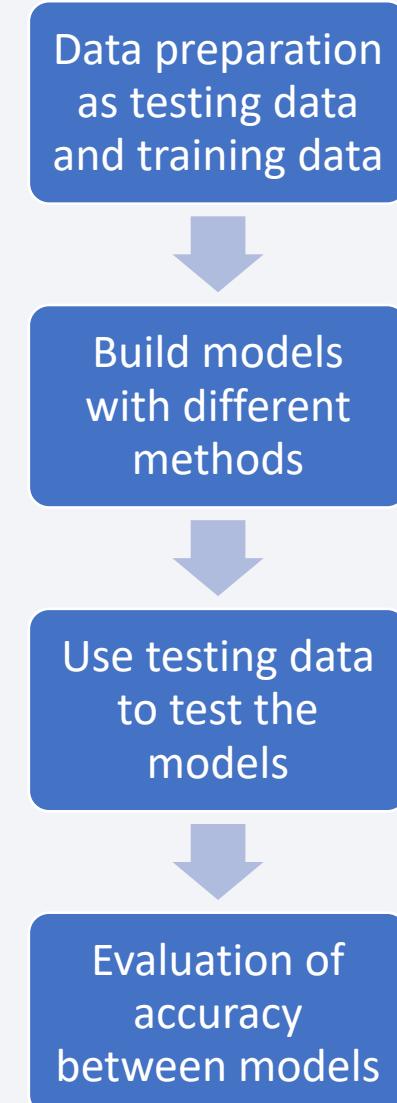
- To build an interactive map, the objects were applied, such as markers, circles, lines, and shown on the US map.
- The successful/failed launches for each site were also marked on the map.
- The circles signified the areas surrounding specific sites, such as NASA.
- The lines indicated the distances from two coordinates.
- [GitHub Link](#)

Build a Dashboard with Plotly Dash

- All successful launch sites were plotted as a pie chart which showed the ratio between four different launch sites. The site 'KSC LC-39A' had the highest ratio.
- An interactive scatter plot was depicted and showed the distributions of payload mass against launch outcomes.
- [GitHub Link](#)

Predictive Analysis (Classification)

- There were four methods applied to prediction, namely Decision Tree, SVM, Logistic Regression, KNN.
- The Decision Tree had the highest accuracy scores in training model but had the lowest scores in testing model.
- [GitHub Link](#)

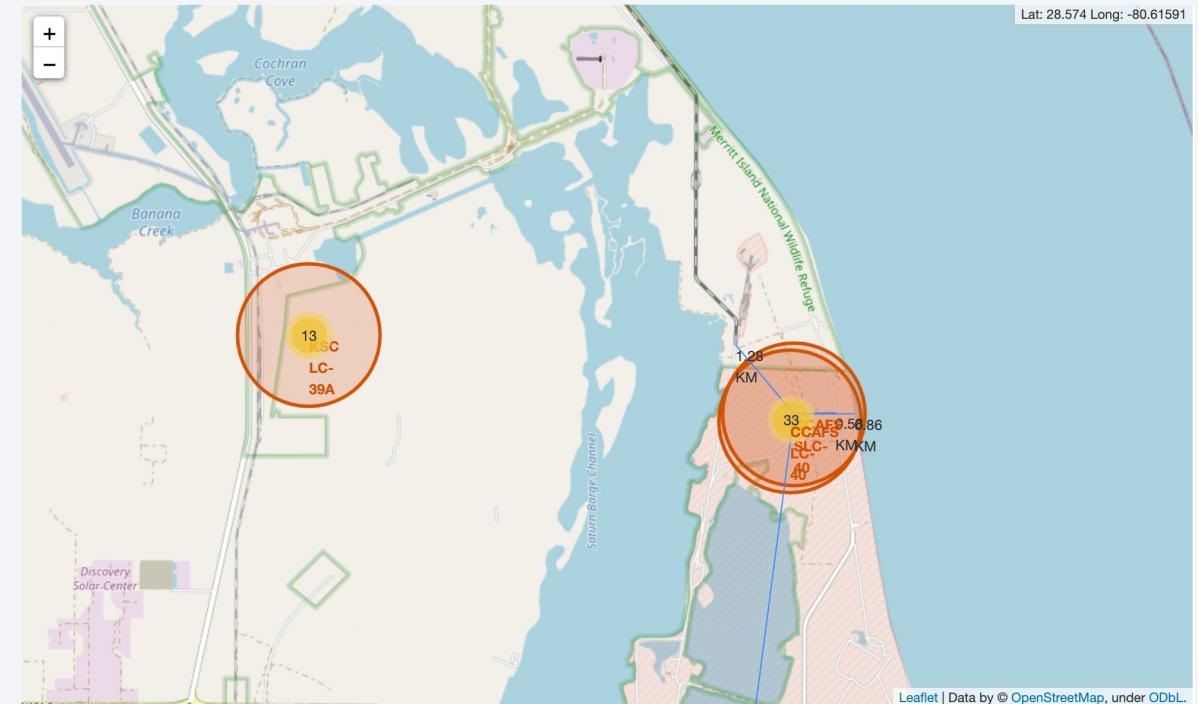
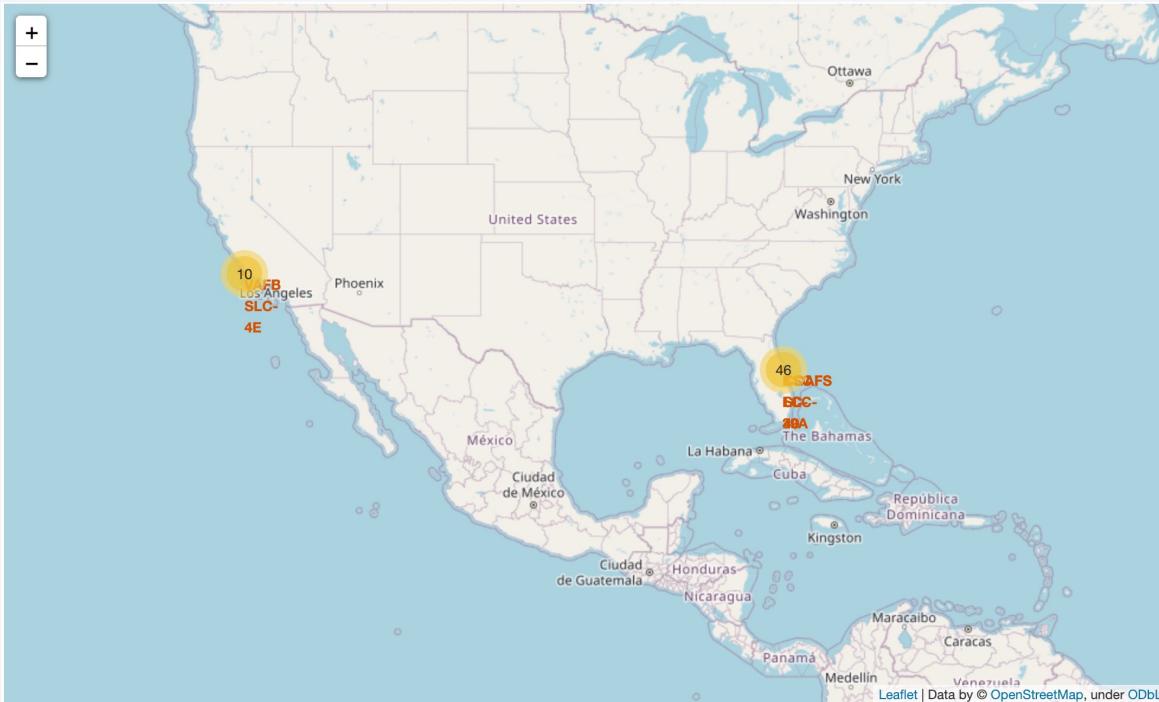


Results

- Exploratory data analysis results
 - Space X had 4 launch site.
 - The first launch date with successful landing outcome was 01-03-2013.
 - The average payload mass for booster version F9 v1.1 was 2534.67 kilograms.
 - Space X was commissioned by NASA and conveyed 45596 kilograms in total.
 - When flight numbers were over 79, the launch outcomes were 100% successful with heavier payload mass over 10000 kilograms.

Results

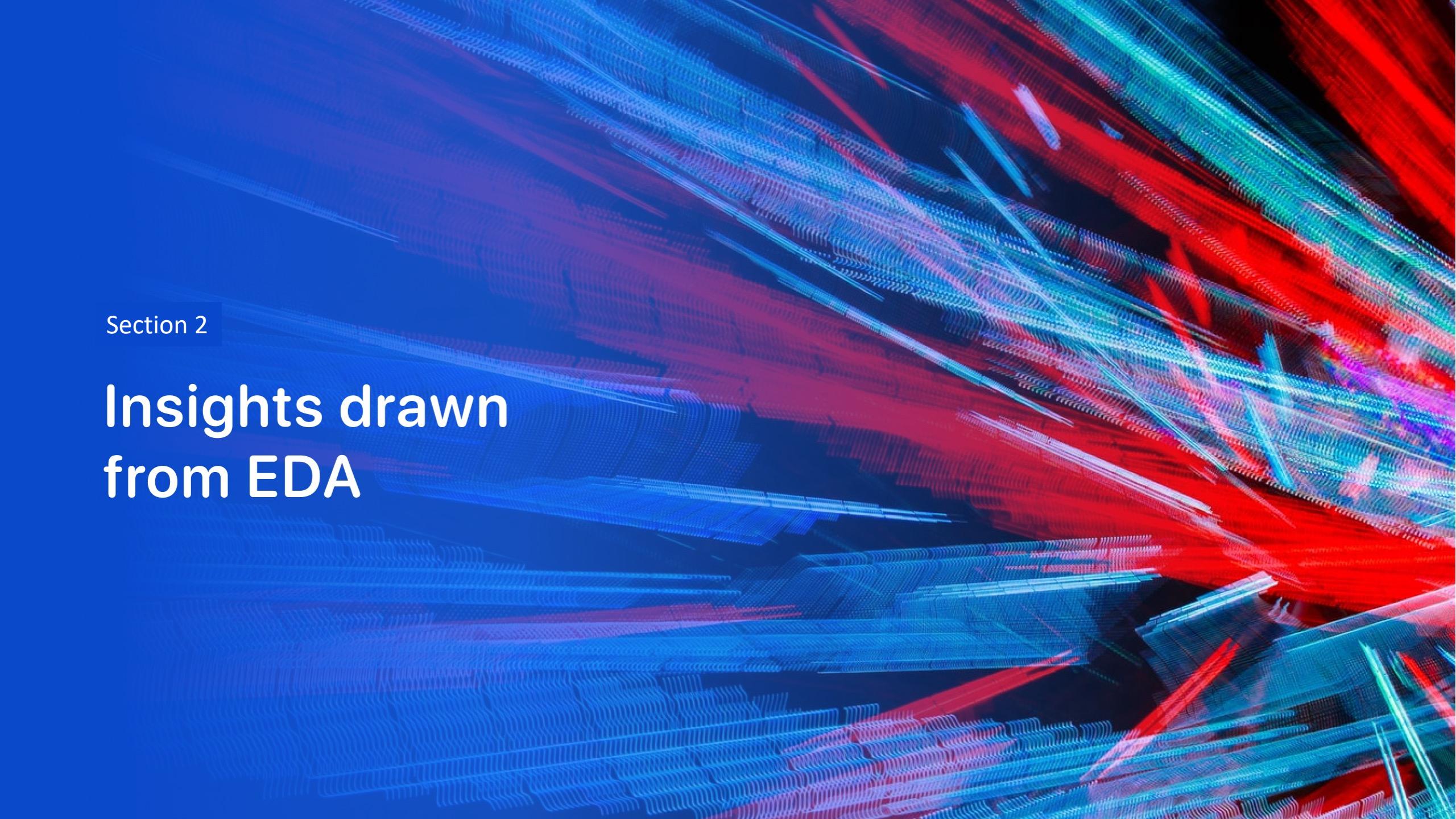
- By interactively analyzing, there were 56 launches for analysis. 10 launches were in west coast and 46 in east coast.
- The map also showed the safety distances from launch sites to important facilities.



Results

- Predictive analysis results by applying four methods showed that the decision tree model had the highest accuracy over 88% in training model but had the least accuracy 72% in testing model.

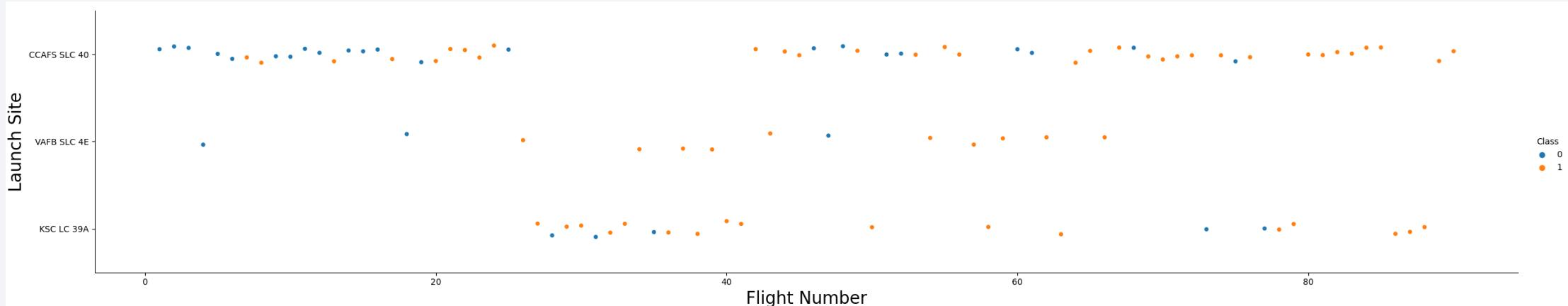
	method	training_accuracy	testing_accuracy
0	logistic regression	0.846429	0.833333
1	SVM	0.848214	0.833333
2	Decition Tree	0.889286	0.722222
3	KNN	0.848214	0.833333

The background of the slide features a complex, abstract digital visualization. It consists of numerous thin, glowing lines that create a sense of depth and motion. The lines are primarily blue and red, with some green and purple highlights. They form a grid-like structure that curves and twists across the frame, resembling a three-dimensional space or a network of data points. The overall effect is futuristic and dynamic.

Section 2

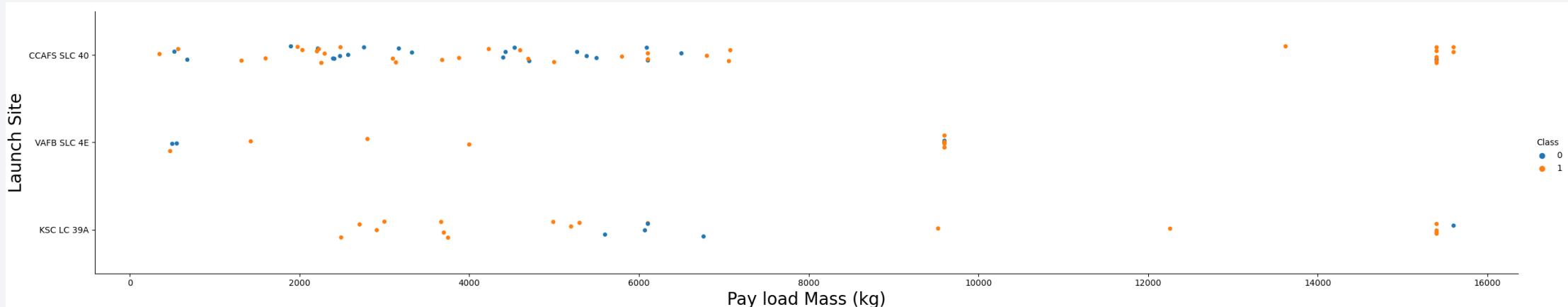
Insights drawn from EDA

Flight Number vs. Launch Site



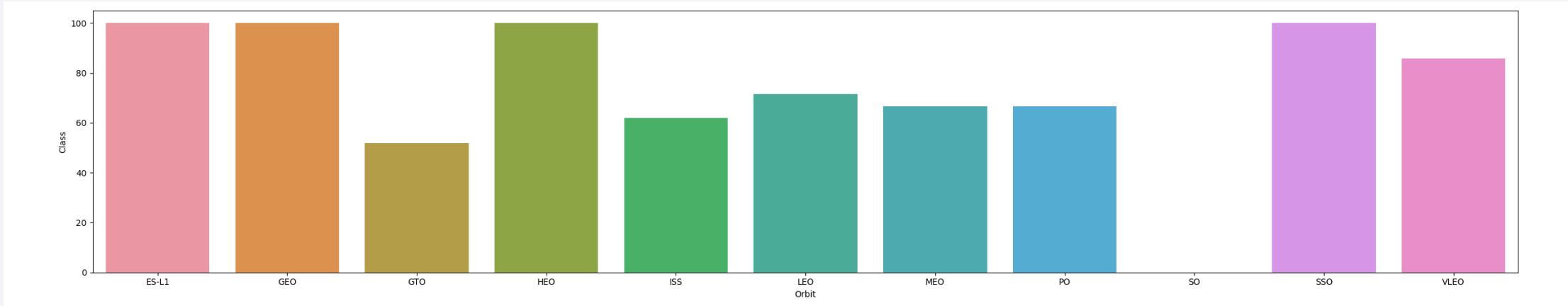
- Flight Number vs. Launch Site
 - CCAFS LC-40 has more times of launches with 60% success rates.
 - KSC LC-39A and VAFB SLC 4E has a success rate of 77%.
- [GitHub Link](#)

Payload vs. Launch Site



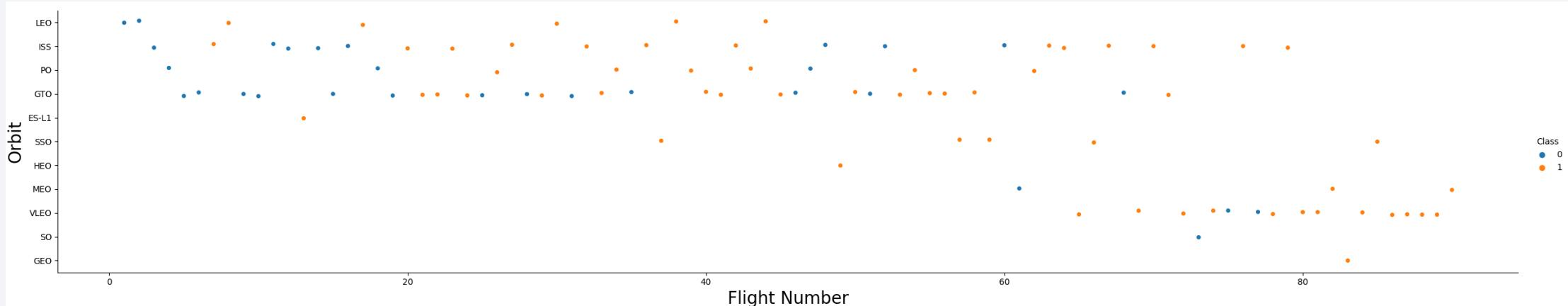
- Play load Mass vs. Launch Site
 - VAFB-SLC 4E launch site has no launches when payload mass is over 10000.
- [GitHub Link](#)

Success Rate vs. Orbit Type



- Orbit vs. Class (successful rate)
 - There is no successful record if a rocket was launched to the SO orbit.
- [GitHub Link](#)

Flight Number vs. Orbit Type



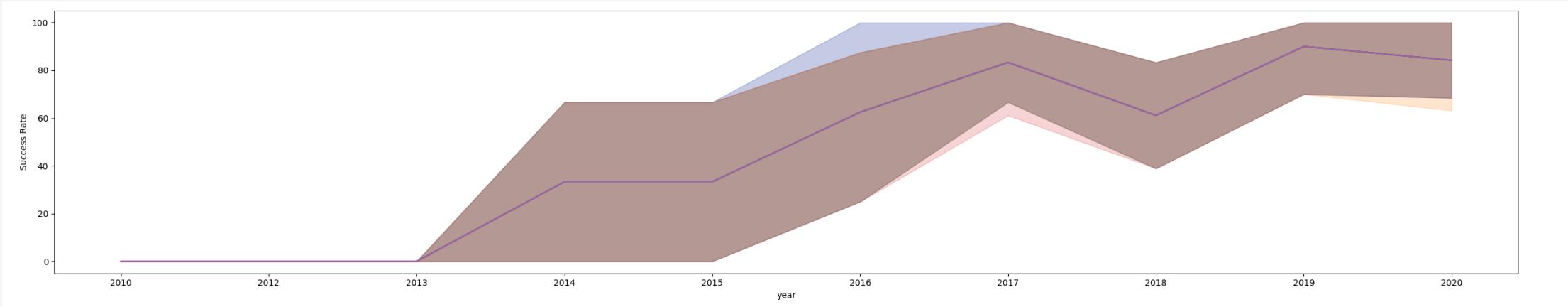
- Flight Number vs. Orbit
 - When flight number is lower than 60, most rockets launch were targeted to Orbits LEO, ISS, PO GTO.
- [GitHub Link](#)

Payload vs. Orbit Type



- Play Load Mass vs. Orbit
 - When play load mass is heavier than 8000, rockets were launched less than under 8000.
- [GitHub Link](#)

Launch Success Yearly Trend



- Year vs. Success Rate
 - After 2017, the success rate could be over 60%.
- [GitHub Link](#)

All Launch Site Names

- Find the names of the unique launch sites

```
%sql select distinct launch_site from SPACEXTBL
```

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

```
Done.
```

Launch_Site

```
CCAFS LC-40
```

```
VAFB SLC-4E
```

```
KSC LC-39A
```

```
CCAFS SLC-40
```

Launch Site Names Begin with 'CCA'

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

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

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

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

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS_KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome
04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
08-12-2010	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)
22-05-2012	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
01-03-2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

- Calculate the total payload carried by boosters from
- The total play load mass is 45596 kilograms.

```
%sql select sum(PAYLOAD_MASS__KG_) as total from spacextbl where customer = 'NASA (CRS)'  
* sqlite:///my_data1.db  
Done.  
total  
45596
```

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1
- The average payload mass is 2534.67 kilograms for booster version F9 v1.1

```
%%sql
select avg(PAYLOAD_MASS__KG_) as average from spacextbl where Booster_Version like 'F9 v1.1%'

* sqlite:///my_data1.db
Done.

average
-----
2534.666666666665
```

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad
- The first successful landing date was 01-03-2013.

```
%%sql
select min(date) as Date from spacextbl where mission_outcome like 'Success'
```

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

```
Done.
```

Date
01-03-2013

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 (mission_outcome like 'Success%')
AND (payload_mass_kg_ BETWEEN 4000 AND 6000) AND ("Landing _Outcome" like 'Success (drone ship)%')

* sqlite:///my_data1.db
Done.

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

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes

```
%sql select distinct Mission_Outcome, count(*) as total from spacextbl \
GROUP by mission_outcome ORDER BY mission_outcome
```

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

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

Boosters Carried Maximum Payload

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

```
%%sql
select distinct Booster_Version from spacextbl where PAYLOAD_MASS__KG_ = (select max(PAYLOAD_MASS__KG_) from spacextbl)

* sqlite:///my_data1.db
Done.

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
```

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", 4, 2) as Month, "Landing _Outcome", booster_version, launch_site
from spacextbl where substr("Date",7,4) = '2015' AND "Landing _Outcome" like 'Failure (drone ship)%'

* sqlite:///my_data1.db
Done.

Month  Landing _Outcome  Booster_Version  Launch_Site
-----  -----  -----  -----
01    Failure (drone ship)  F9 v1.1 B1012  CCAFS LC-40
04    Failure (drone ship)  F9 v1.1 B1015  CCAFS LC-40
```

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("Landing _Outcome") from spacextbl
where Mission_Outcome like "Success%" and "Date" between "04-06-2010" and "20-03-2017"
group by "Landing _Outcome"
order by "Date" DESC
```

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

Landing _Outcome	count("Landing _Outcome")
Success (ground pad)	6
Controlled (ocean)	3
Failure (drone ship)	4
No attempt	10
Success (drone ship)	8
Success	20
No attempt	1
Failure	3
Failure (parachute)	2

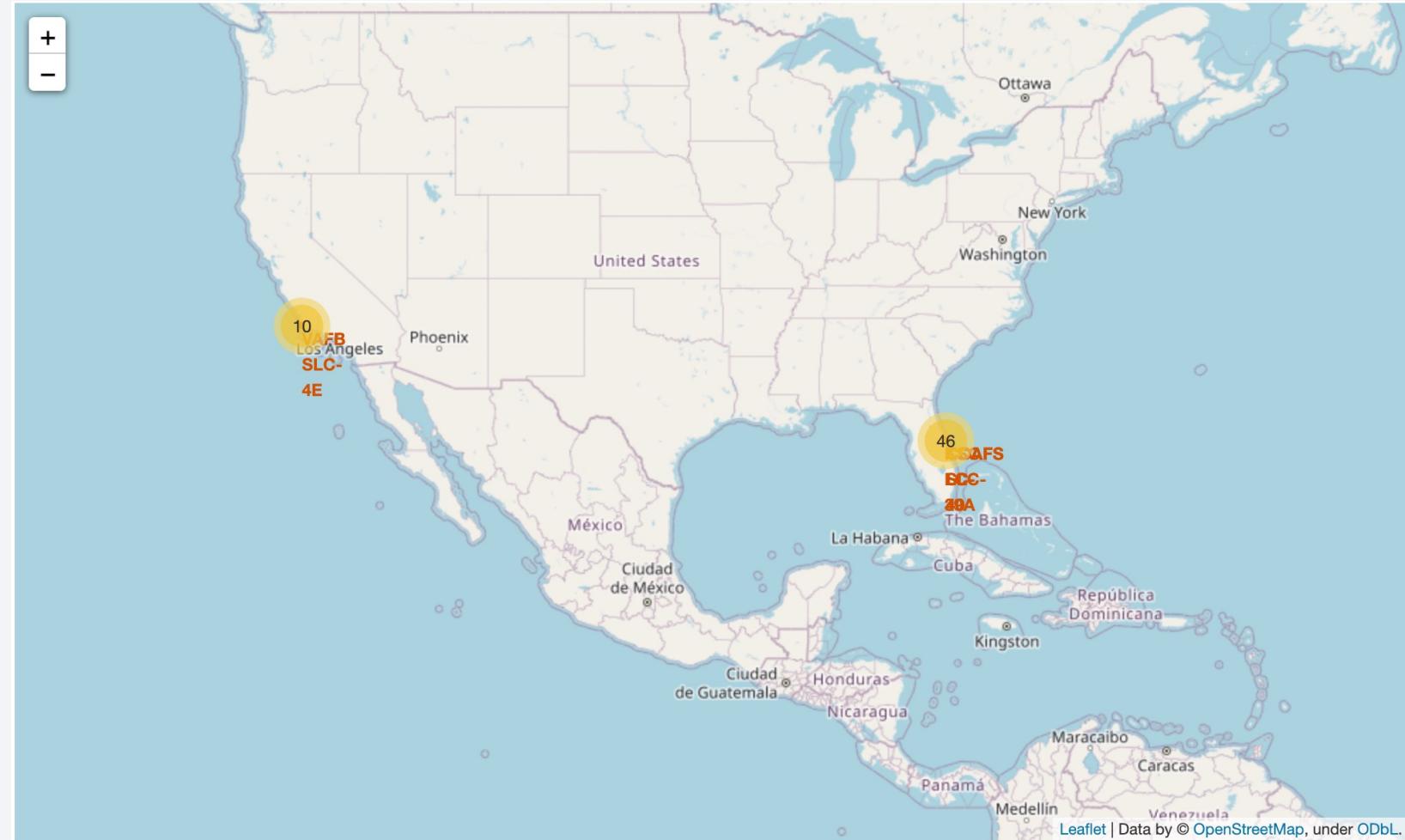
The background of the slide is a photograph taken from space at night. It shows the curvature of the Earth against a dark blue-black void of space. City lights are visible as numerous small white and yellow dots, primarily concentrated in the lower right quadrant where the United States appears. In the upper right, the green and yellow glow of the aurora borealis is visible. The atmosphere of the Earth is thin and hazy, appearing as a light blue band near the horizon.

Section 3

Launch Sites Proximities Analysis

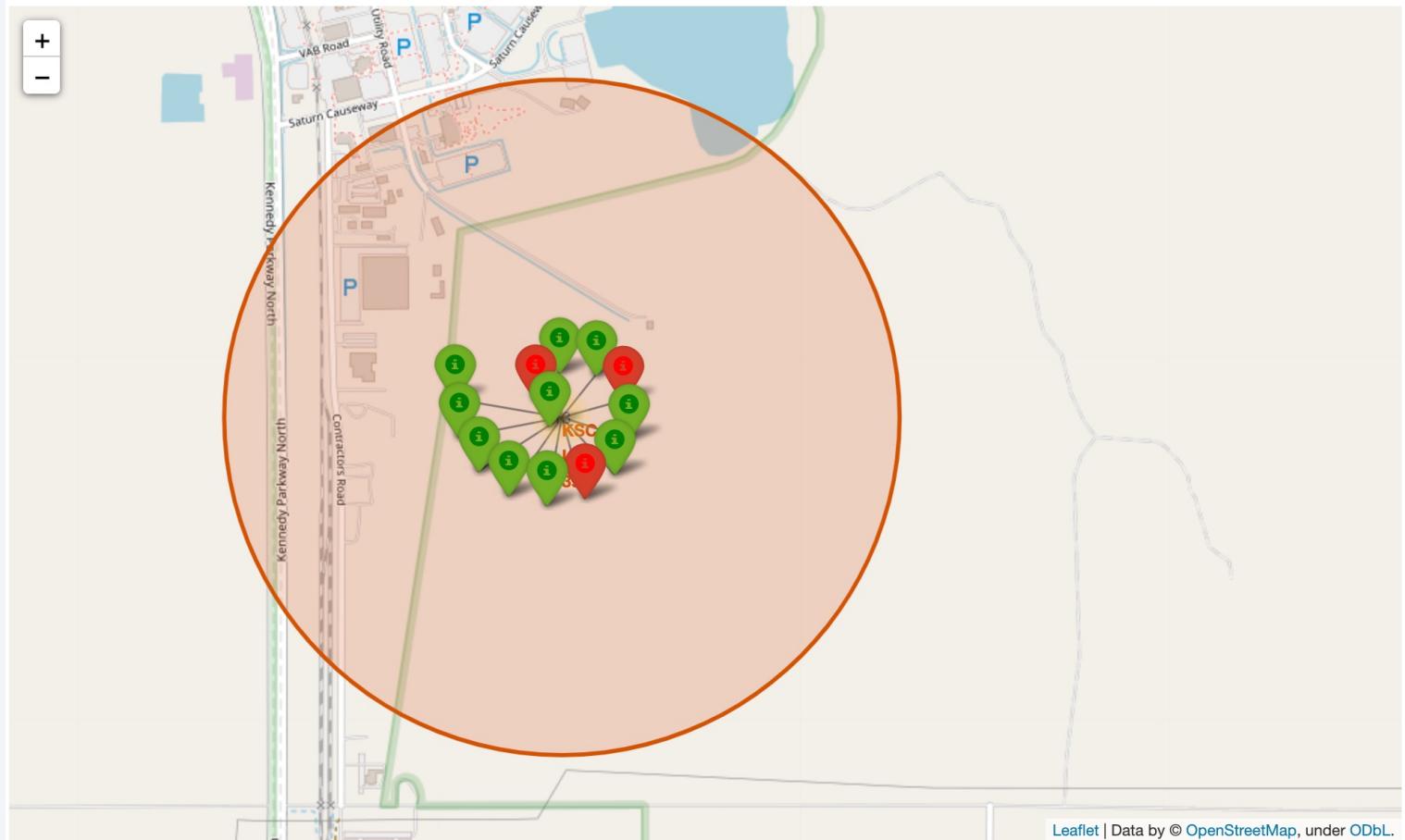
All launch sites

- The launch sites were marked on the US map and there were 46 launches at the eastern coastline and 10 launches at the western coastline.



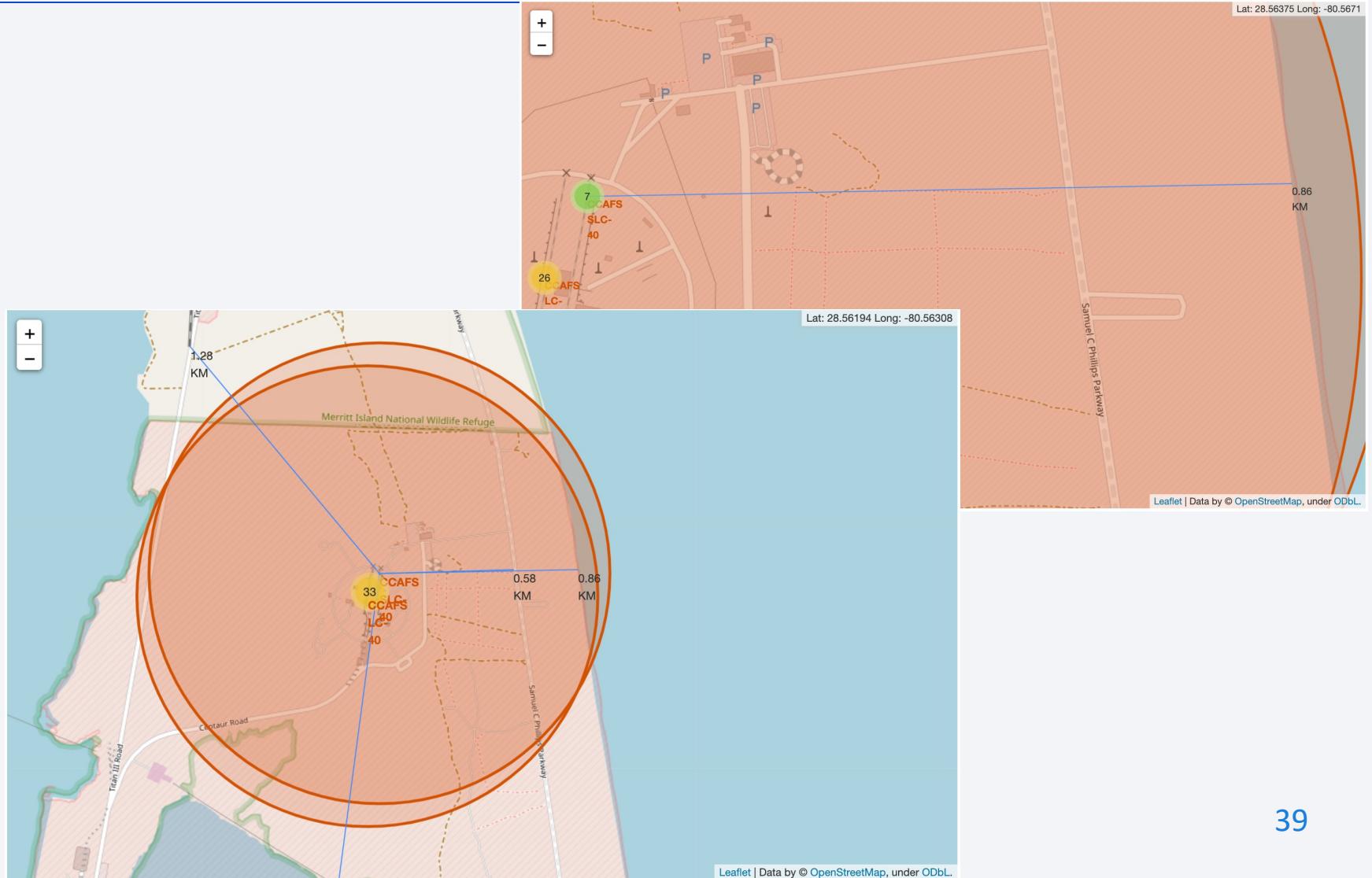
Launch Outcomes by Sites

- The markers indicated the outcomes of launches at this site.
- Green indicated success and red indicate failure.



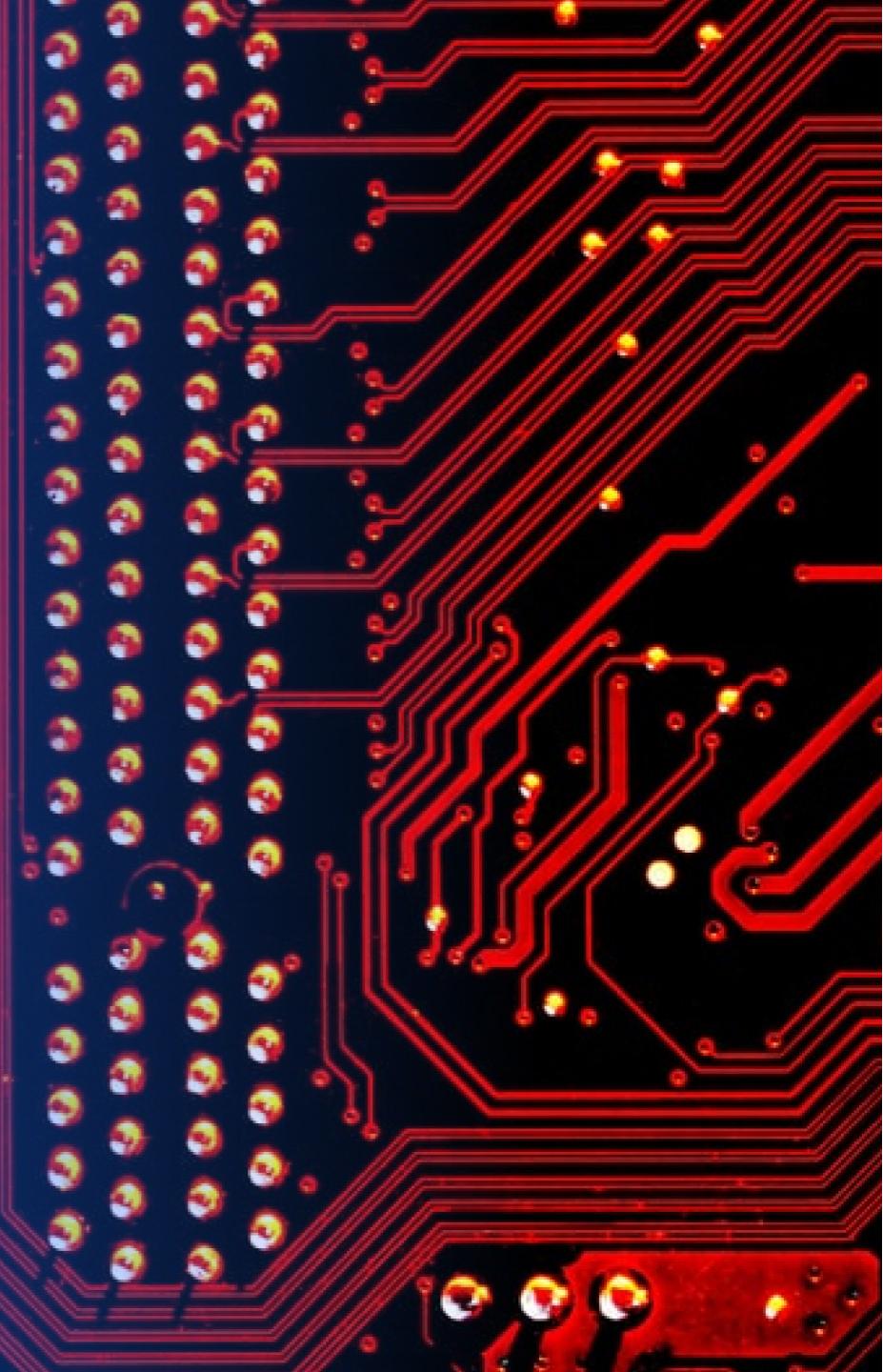
Safe distances from key facilities

- The blue line on the map indicated the distance from important facilities, such as railways, airports.

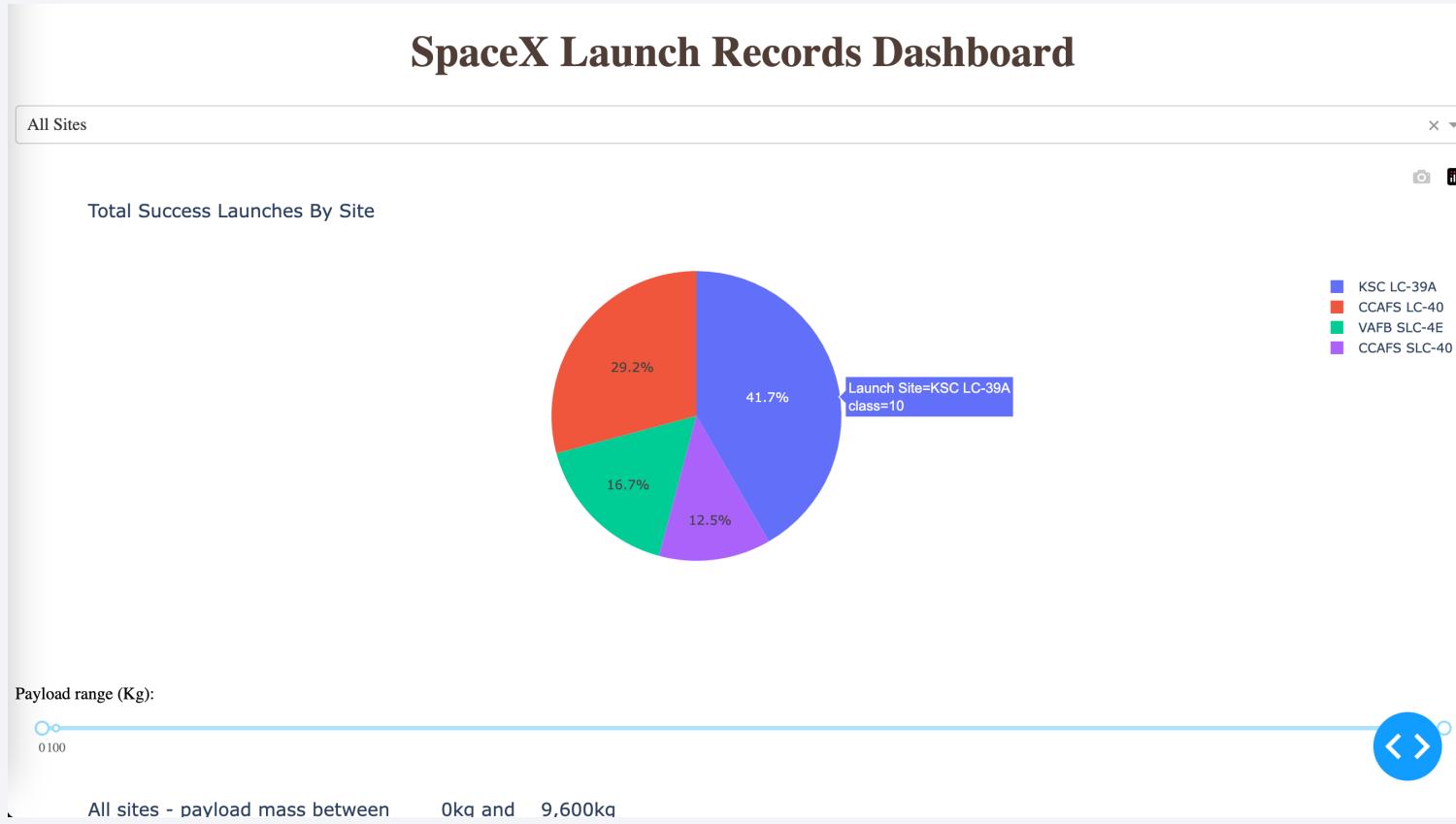


Section 4

Build a Dashboard with Plotly Dash

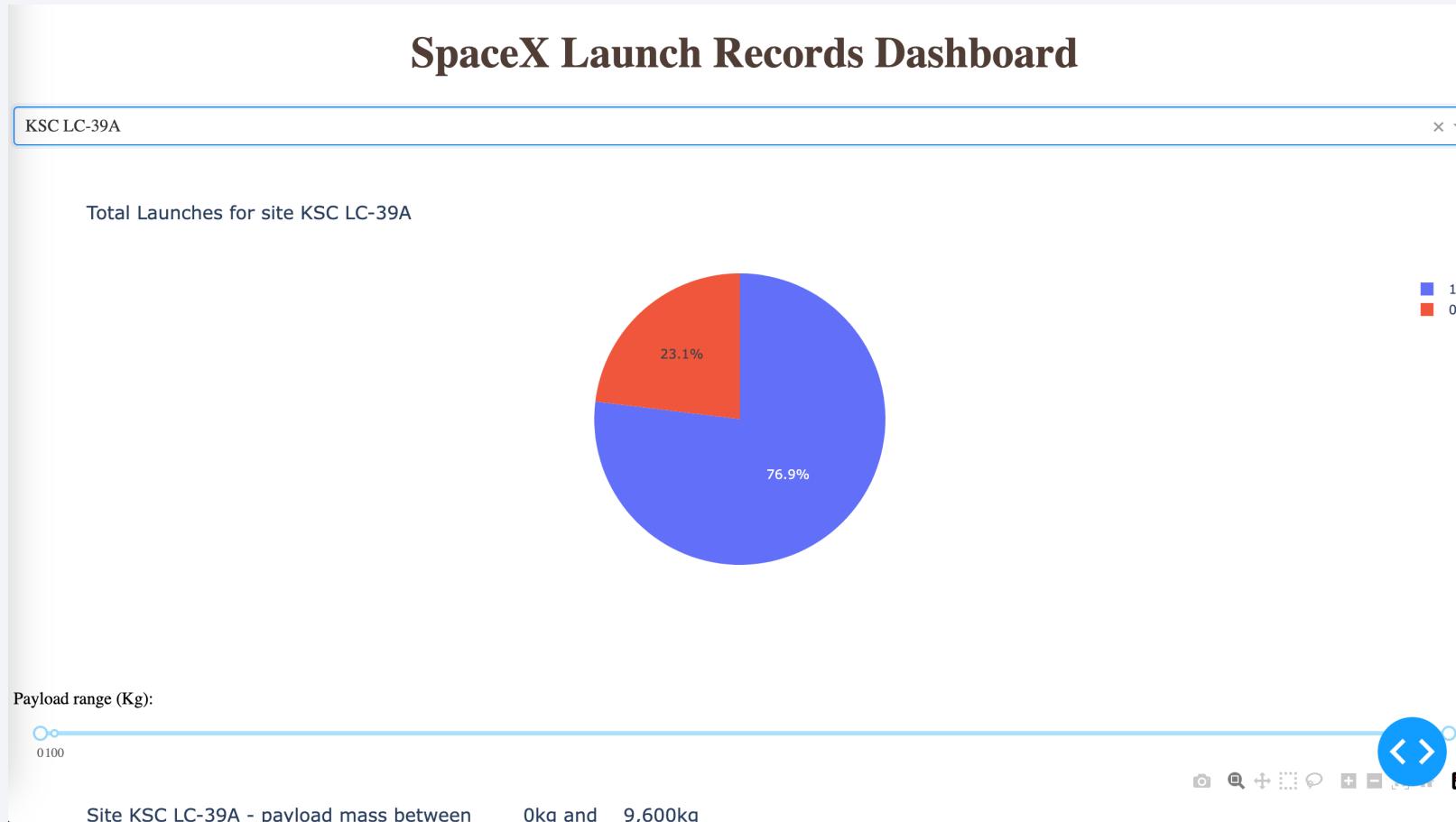


Successful Launches by Sites



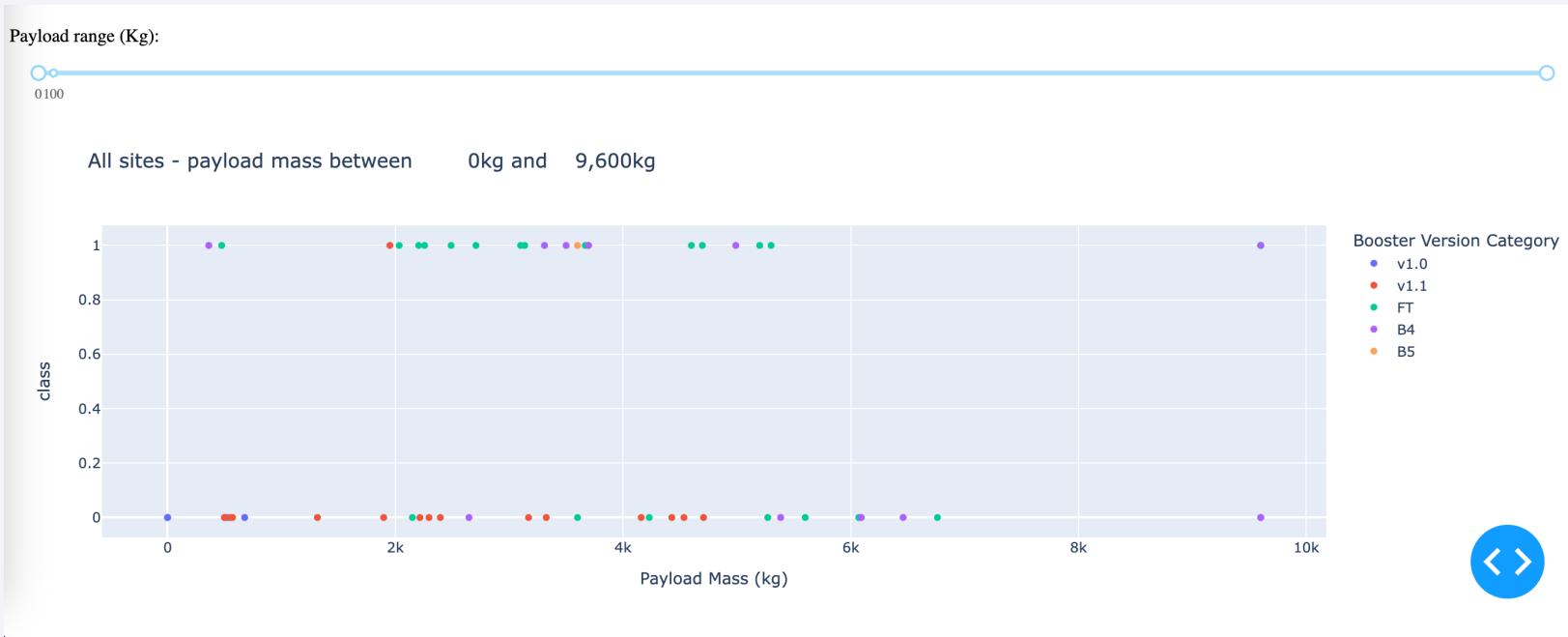
- The successful launches are distributed in four sites as shown in the piechart.

Successful Launches Ratio



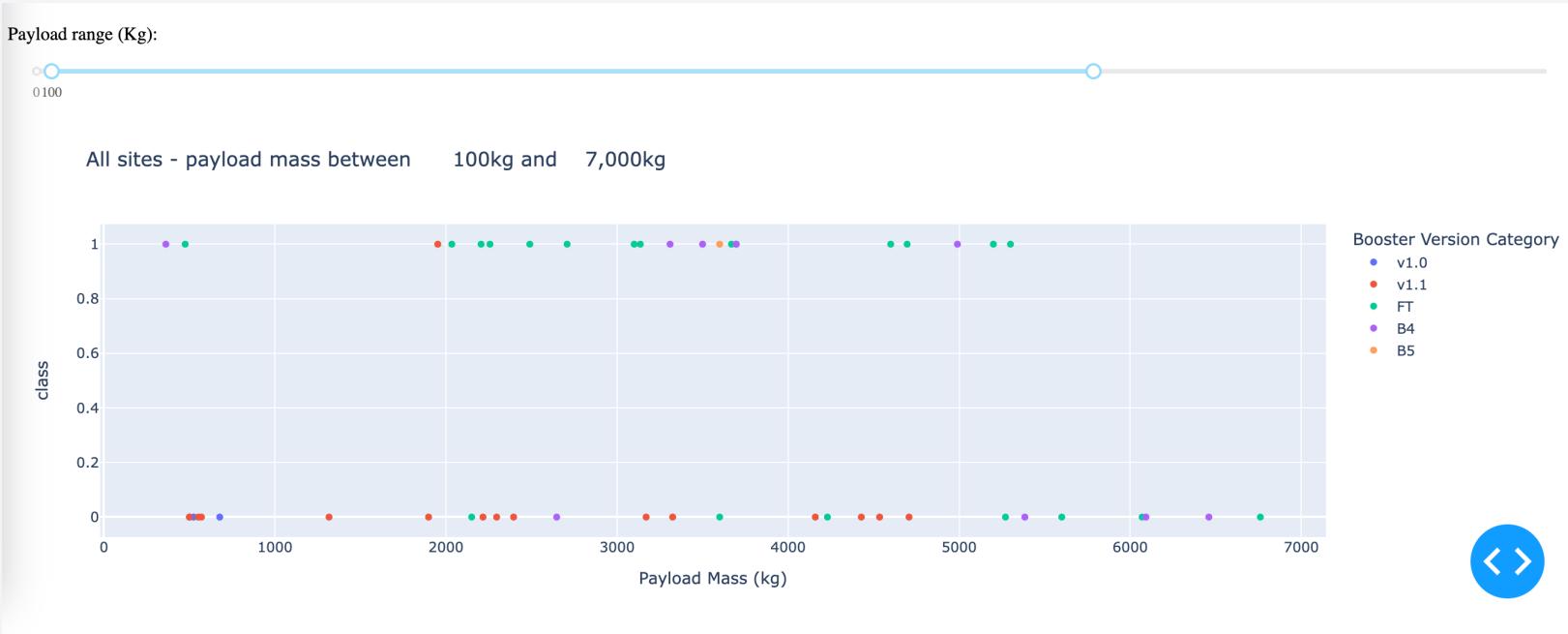
- The launch site 'KSC LC-39A' has the highest ratio of successful launches 76.9%.

Payload Mass vs. Launch Outcomes



- The launch outcomes are distributed across the scatter plot which indicates the distribution of payload mass against launch outcomes.
- Most launches had payload mass less than 8000 kilograms.

Payload Mass vs. Launches Outcomes



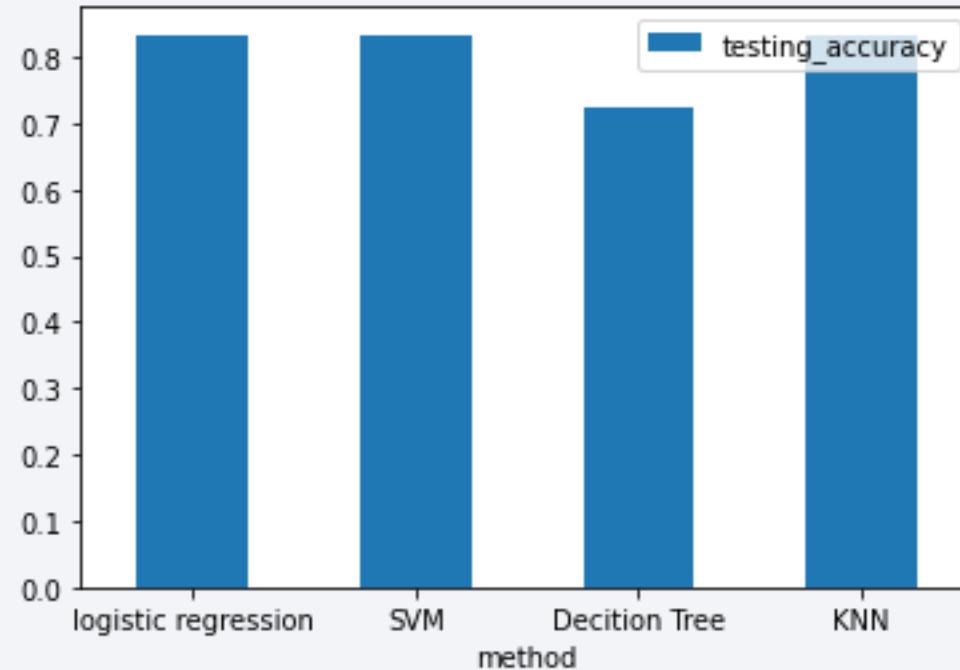
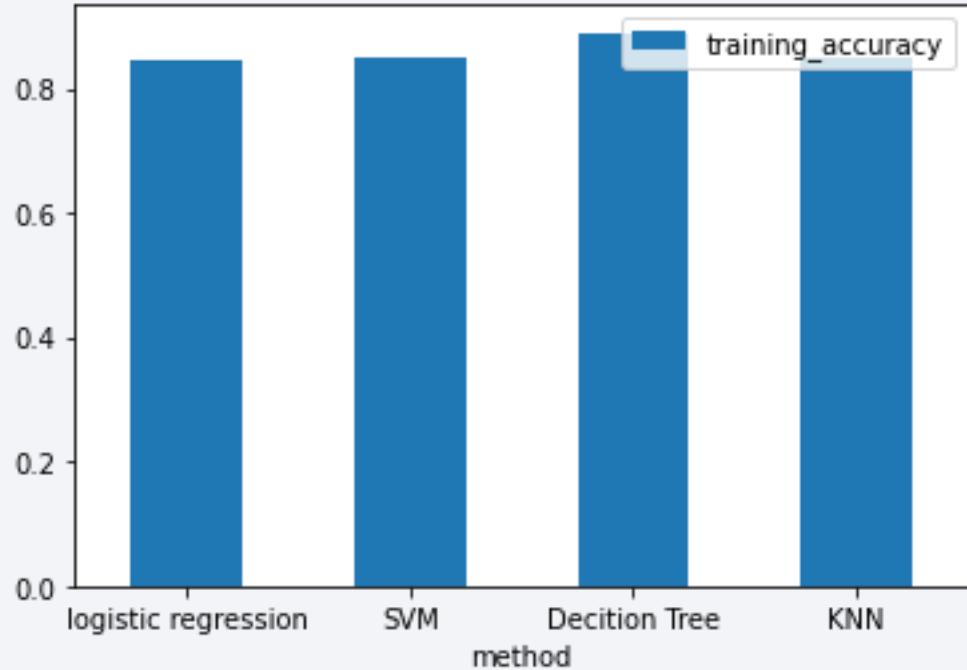
- Most launches had payload mass less than 7000 kilograms.

The background of the slide features a dynamic, abstract design. It consists of several thick, curved lines that transition from a bright yellow at the top right to a deep blue at the bottom left. These lines create a sense of motion and depth, resembling a tunnel or a stylized road. The overall effect is modern and professional.

Section 5

Predictive Analysis (Classification)

Classification Accuracy



- Decision Tree has the highest accuracy for training model. However, its accuracy for testing is the lowest.

	method	training_accuracy	testing_accuracy
0	logistic regression	0.846429	0.833333
1	SVM	0.848214	0.833333
2	Decition Tree	0.889286	0.722222
3	KNN	0.848214	0.833333

Confusion Matrix

- This confusion matrix from the Decision Tree model provides the accuracy by distinguishing true positive, false positive, true negative, and false negative.
- Due to only 18 records, the outcome may not be reliable. We need more data to test.



Conclusions

- The best launch site was KSC LC-39A with highest successful launch ratio.
- Payload Mass over 8000 kilograms has higher rates of successful outcomes.
- Decision Tree model can be used to predict the landing outcomes.
- The launch site VAFB SLC 4E was not suitable for launching with higher payload mas over 10000 kilograms due to no previous records.

Appendix

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

Thank you!

