IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

< Shiny5228>
<2024-11-26>

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection through API and Web Scraping
  - Data Wrangling
  - Exploratory Data Analysis (EDA) with SQL and Data Visualization
  - Interactive Visual Analysis with Folium und Dash
  - Machine Learning (ML) Predictions
- Summary of all results
  - EDA results, screenshots and ML results

Introduction

- Space Exploration Technologies Corp., commonly referred to as SpaceX, is an American space technology company headquartered at the SpaceX Starbase near Brownsville, Texas. Since its founding in 2001, the company has made numerous advancements in rocket propulsion, reusable launch vehicle, human spaceflight and satellite constellation technology. By the late 2010s, SpaceX had become the world's dominant space launch provider.
- Problems you want to find answers
  - Identifying all factors that influence the landing outcome.
  - The relationship between each variables and how it is affecting the outcome.
  - The best condition needed to increase the probability of successful landing.

Section 1

Methodology

Methodology

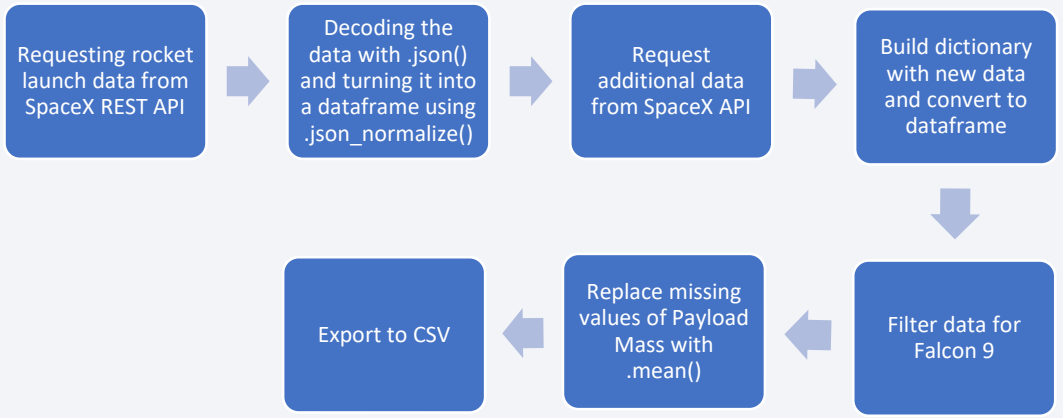
Executive Summary

- Data collection methodology:
  - Data collection using SpaceX REST API and web scraping
- Perform data wrangling
  - Convert landing outcomes in training labels (1 and 0)
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection

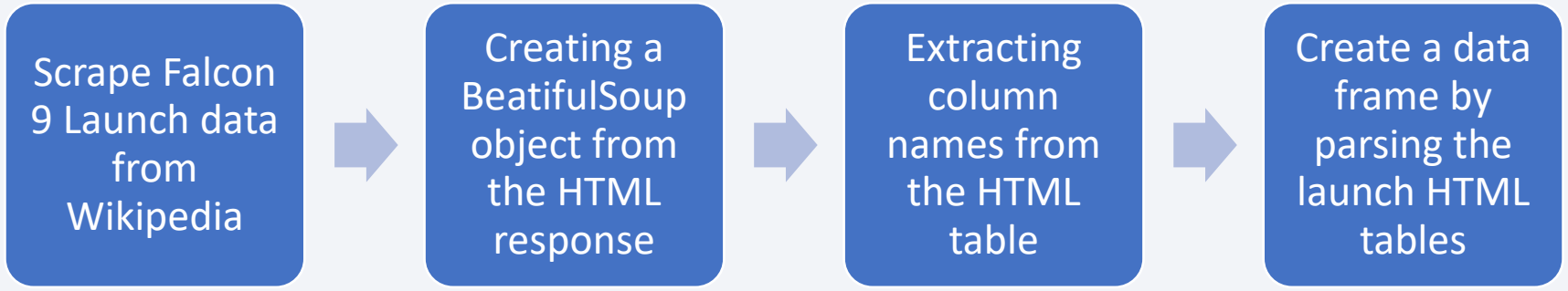
- Describe how data sets were collected.
- You need to present your data collection process use key phrases and flowcharts

Data Collection SpaceX API



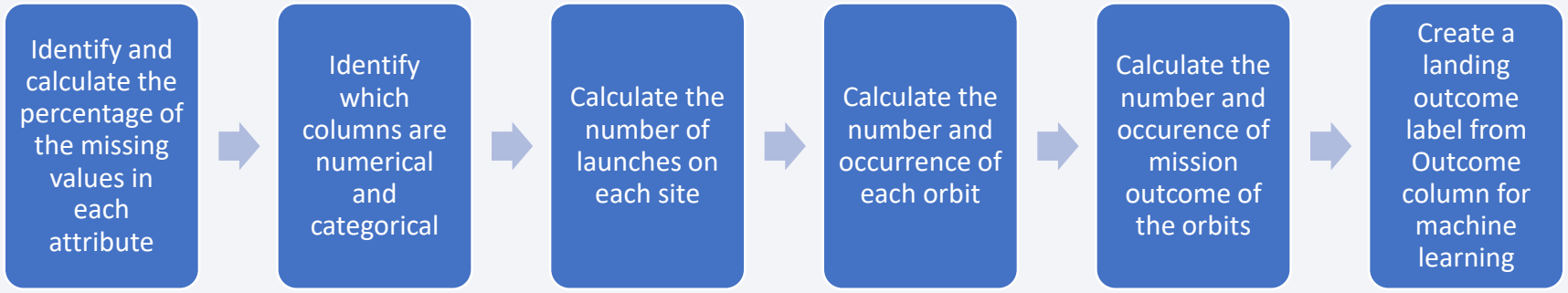
Github

Data Collection – Scraping



Github

Data Wrangling



Github

EDA with Data Visualization

- Scatter plots show the relationship between variables. Scatter point charts for:
  - Flight Number and Launch Site
  - Payload Mass and Launch Site
  - FlightNumber and Orbit type
  - Payload Mass and Orbit type
- Bar charts show comparisons among discrete categories
  - success rate of each orbit type
- Line charts show trends in data over time
  - launch success yearly trend

Github

EDA with SQL

- Displaying the names of the unique launch sites.
- Displaying 5 records where launch sites begin with the string ‘CCA’.
- Displaying the total payload mass carried by booster launched by NASA (CRS).
- Displaying the average payload mass carried by booster version F9 v1.1.
- Listing the date when the first successful landing outcome in ground pad was achieved.
- Listing the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000.
- Listing the total number of successful and failure mission outcomes.
- Listing the names of the booster_versions which have carried the maximum payload mass.
- Listing the failed landing_outcomes in drone ship, their booster versions, and launch sites names for in year 2015.
- Rank the count of landing outcomes or success between the date 2010-06-04 and 2017-03-20, in descending order.

Github

Build an Interactive Map with Folium

- Marked all launch sites on a map
- Marked the success/failed launches for each site on the map
- Calculated the distances between a launch site to its proximities

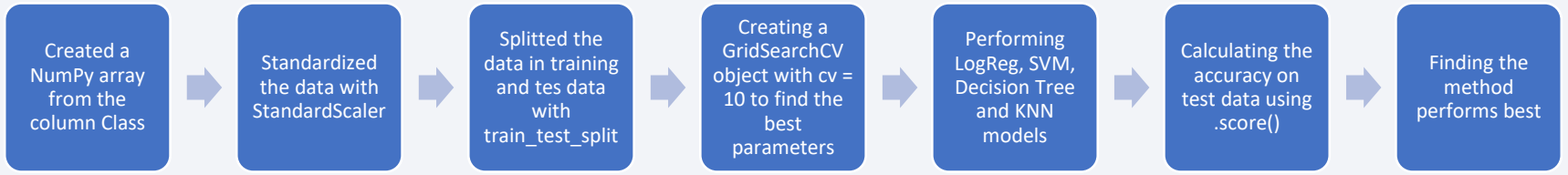
Github

Build a Dashboard with Plotly Dash

- Dropdown list for launch site selection
- pie chart to show the total successful launches count for all sites and the success vs. failed counts for the site, if a specific launch site was selected
- Slider for payload mass range
- Scatter Chart of payload mass vs. success rate for different booster versions

Github

Predictive Analysis (Classification)



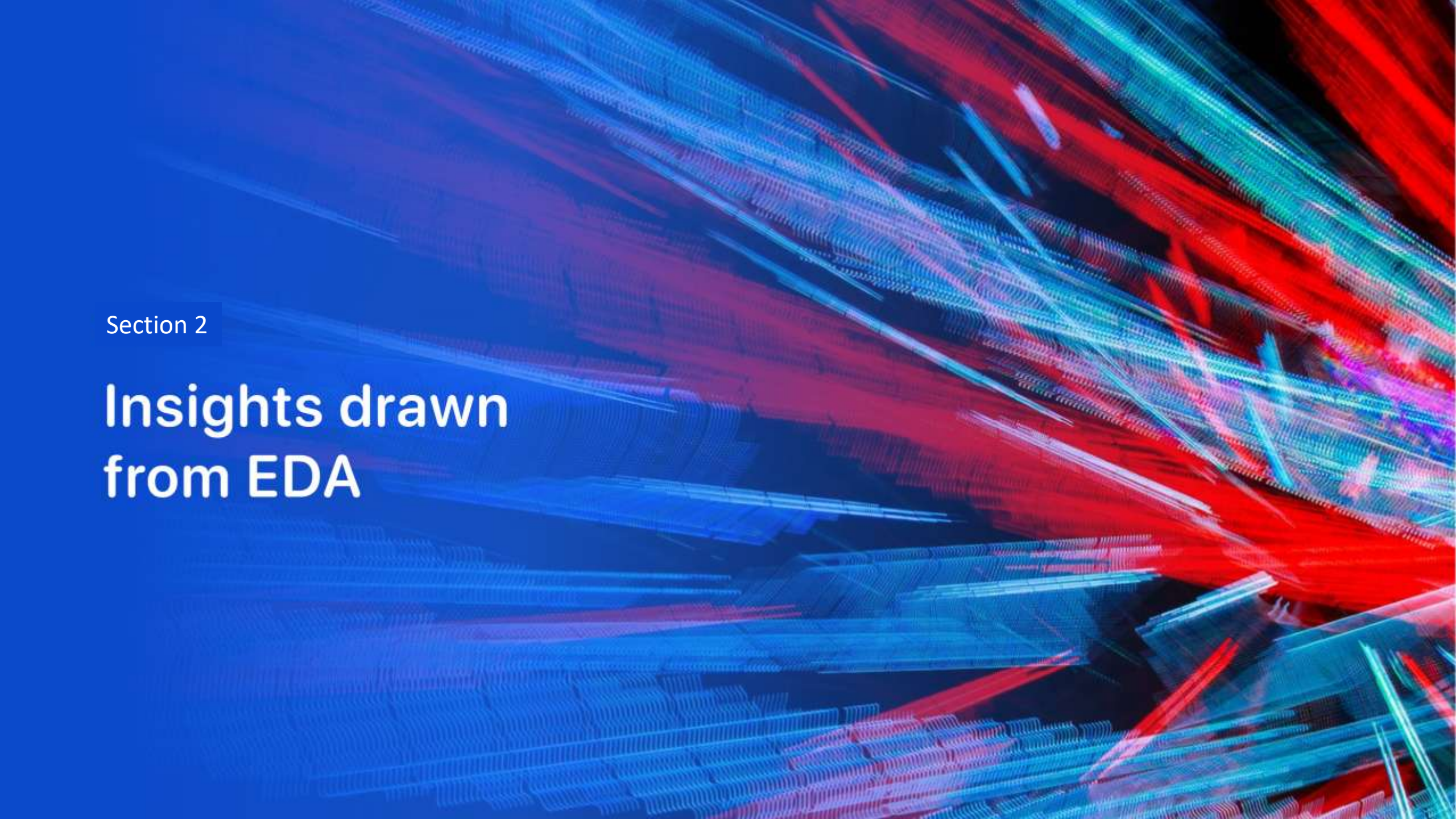
Github

Results

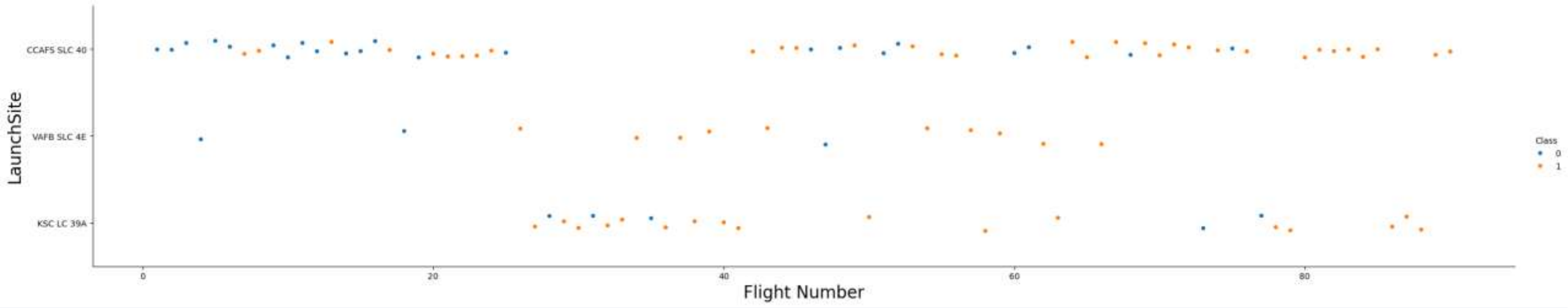
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

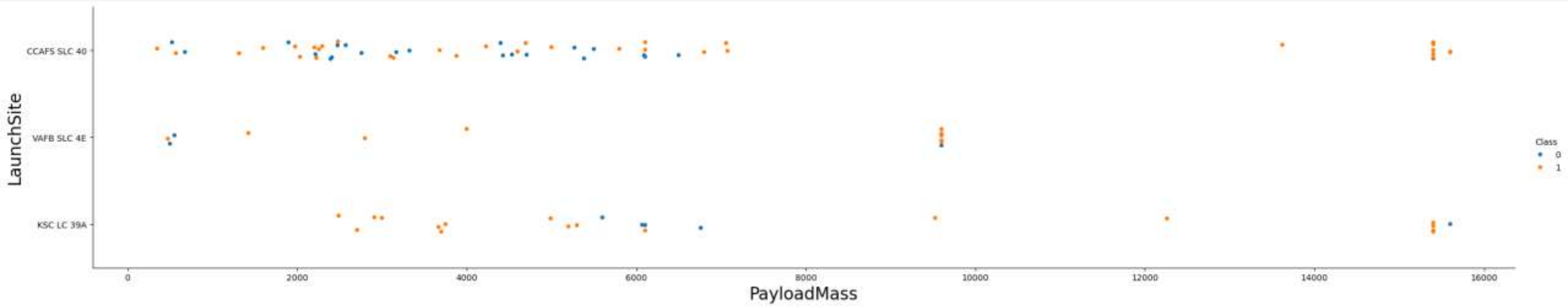


Flight Number vs. Launch Site



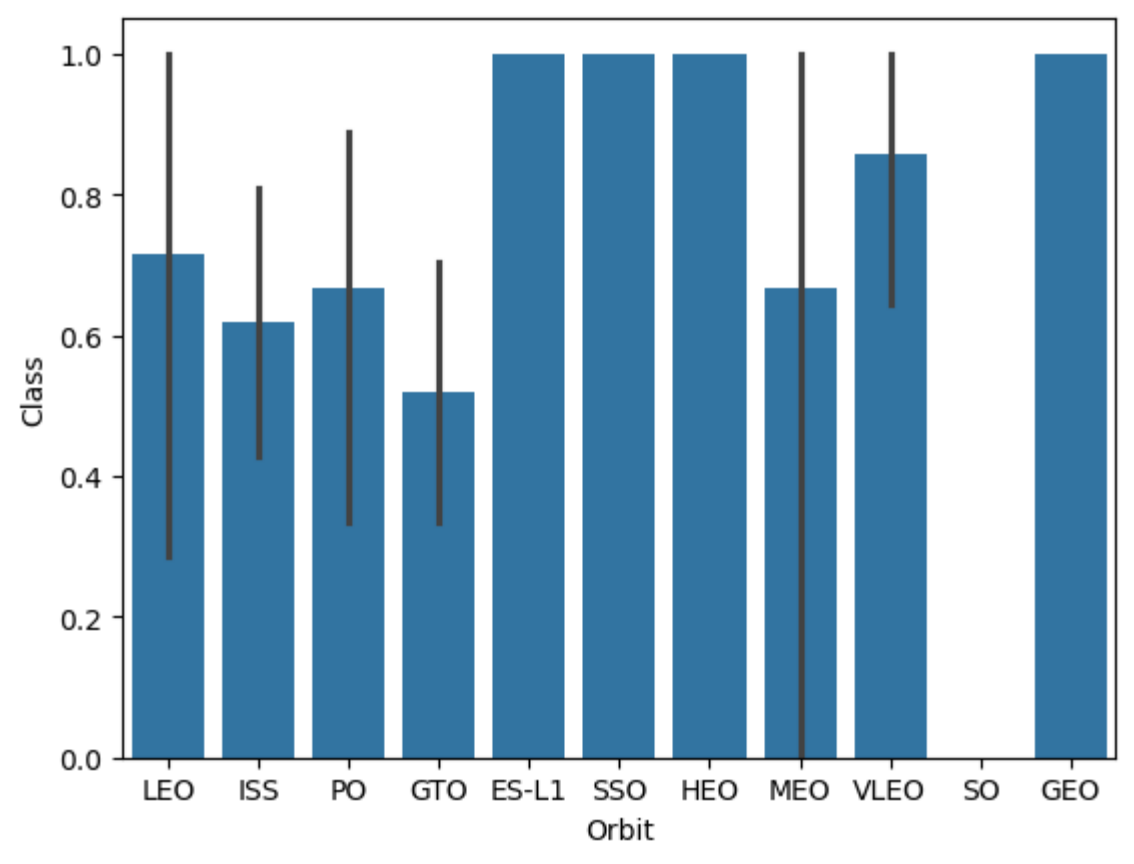
- Later flights were more successful
- Most flights were on CCAFS SLC 40
- VAFB SLC 4E and KSC LC 39A have higher success rates

Payload vs. Launch Site



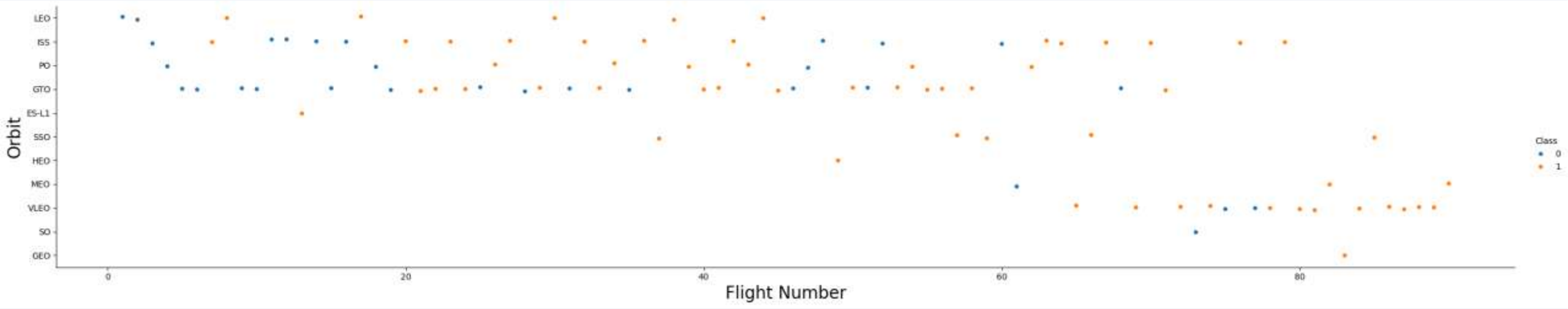
- VAFB SLC 4E only has payloadmass < 10000 kg
- Higher payloadmass were more successful

Success Rate vs. Orbit Type



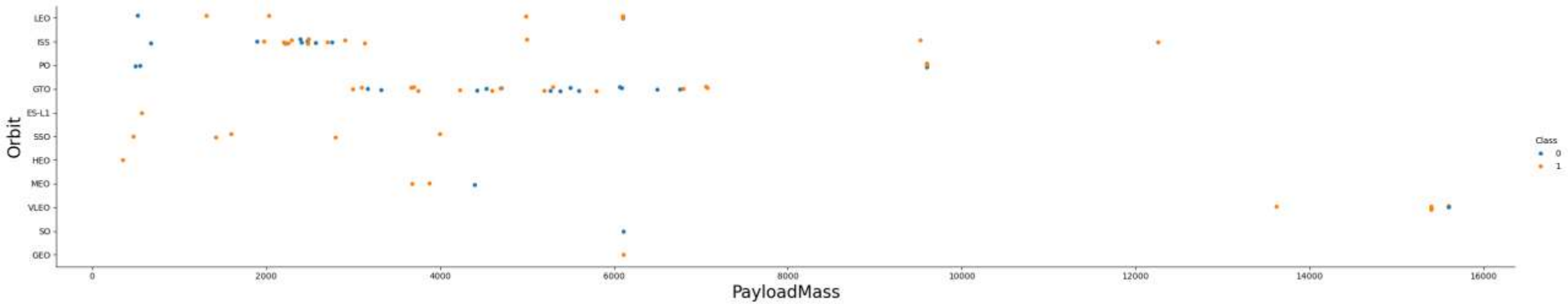
- ES-L1, SSO, HEO, GEO 100% success rate
- SO 0% success rate

Flight Number vs. Orbit Type



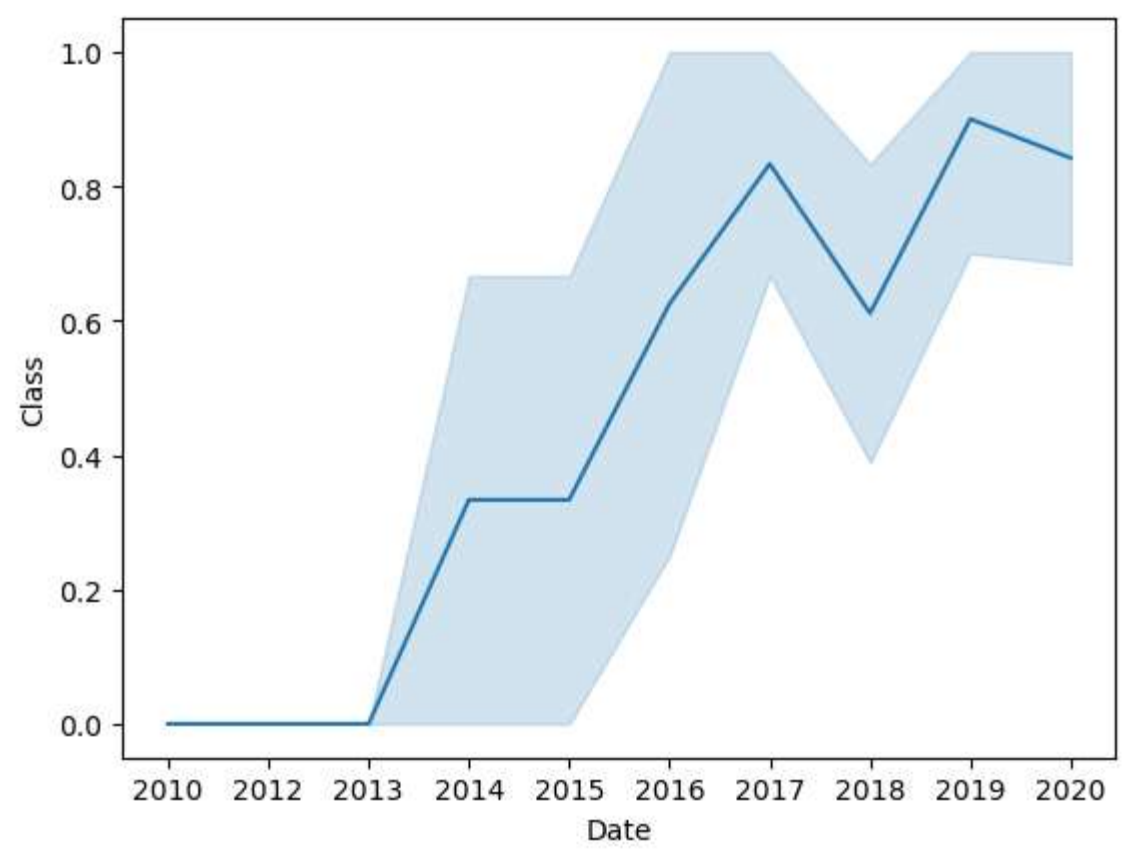
- Later flights in LEO were more successful.
- Recent successful flight were to ISS and VLEO

Payload vs. Orbit Type



- Payloads > 13000kg were all in VLEO
- Heavy payloads in GTO were less successful

Launch Success Yearly Trend



- sucess rate since 2013 kept increasing till 2020

All Launch Site Names

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

```
%sql SELECT distinct(Launch_Site) FROM SPACEXTABLE
```

[11]

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

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

- Names of the unique launch sites

Launch Site Names Begin with 'CCA'

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

```
%sql SELECT * FROM SPACEXTABLE 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
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	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	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 attempt

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

Total Payload Mass

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

Average Payload Mass by F9 v1.1

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

```
%sql SELECT avg(PAYLOAD_MASS__KG_) FROM SPACEXTABLE WHERE Booster_Version = 'F9 v1.1'
```

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

avg(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

```
[14]  %sql SELECT min(Date) FROM SPACEXTABLE where Landing_Outcome = "Success (ground pad)"

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

...   min(Date)
      2015-12-22
```

Successful Drone Ship Landing with Payload between 4000 and 6000

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 Booster_Version FROM SPACEXTABLE where Landing_Outcome = 'Success (drone ship)' and PAYLOAD_MASS__KG_ > 4000 and PAYLOAD_MASS__KG_ < 6000
```

[19]

```
 * 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

List the total number of successful and failure mission outcomes

```
%sql SELECT mission_outcome, count(*) FROM SPACEXTABLE GROUP BY mission_outcome
```

15]

* sqlite:///my_data1.db
Done.

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

Boosters Carried Maximum Payload

List the names of the booster_versions which have carried the maximum payload mass. Use a subquery

```
%sql SELECT Booster_Version FROM SPACEXTABLE where PAYLOAD_MASS__KG_ = (SELECT max(PAYLOAD_MASS__KG_) FROM SPACEXTABLE)
```

```
 * 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 records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.

Note: SQLLite does not support monthnames. So you need to use substr(Date, 6,2) as month to get the months and substr(Date,0,5)='2015' for year.

```
%%sql
SELECT substr(Date, 6,2), Booster_Version, Launch_Site, Landing_Outcome
FROM SPACEXTABLE
WHERE substr(Date,0,5)='2015'
AND Landing_Outcome = 'Failure (drone ship)'
```

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

substr(Date, 6,2)	Booster_Version	Launch_Site	Landing_Outcome
01	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
04	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

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
    FROM SPACEXTABLE
    WHERE date < '2017-03-20'
    and date > '2010-06-04'
    GROUP BY Landing_Outcome
    ORDER BY count desc
```

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

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


Section 3
Launch Sites
Proximities Analysis

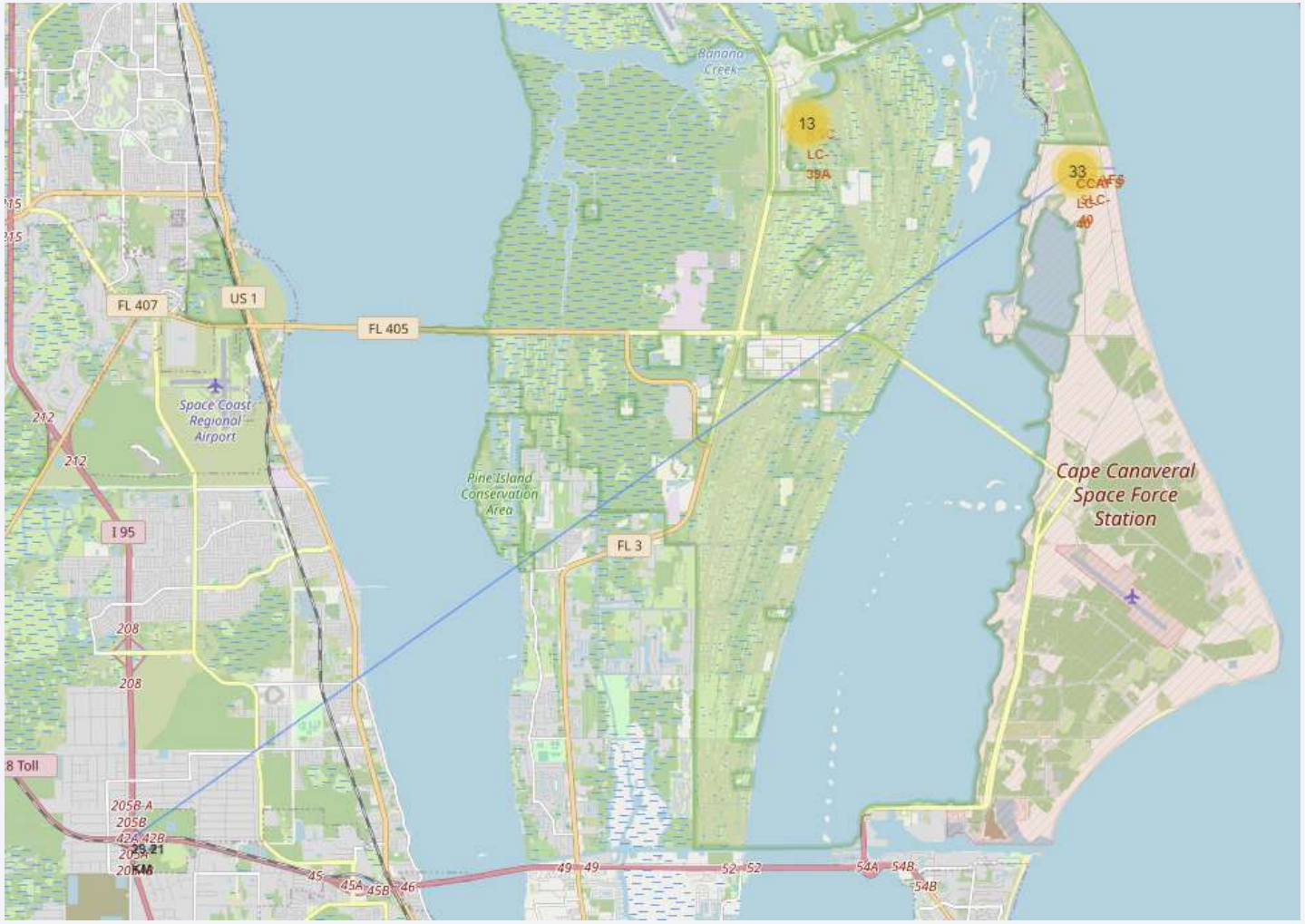
All launch sites' location markers on a global map

- All launch sites were in the US.
- All sites were close to the ocean.
- All sites were close to the equator.

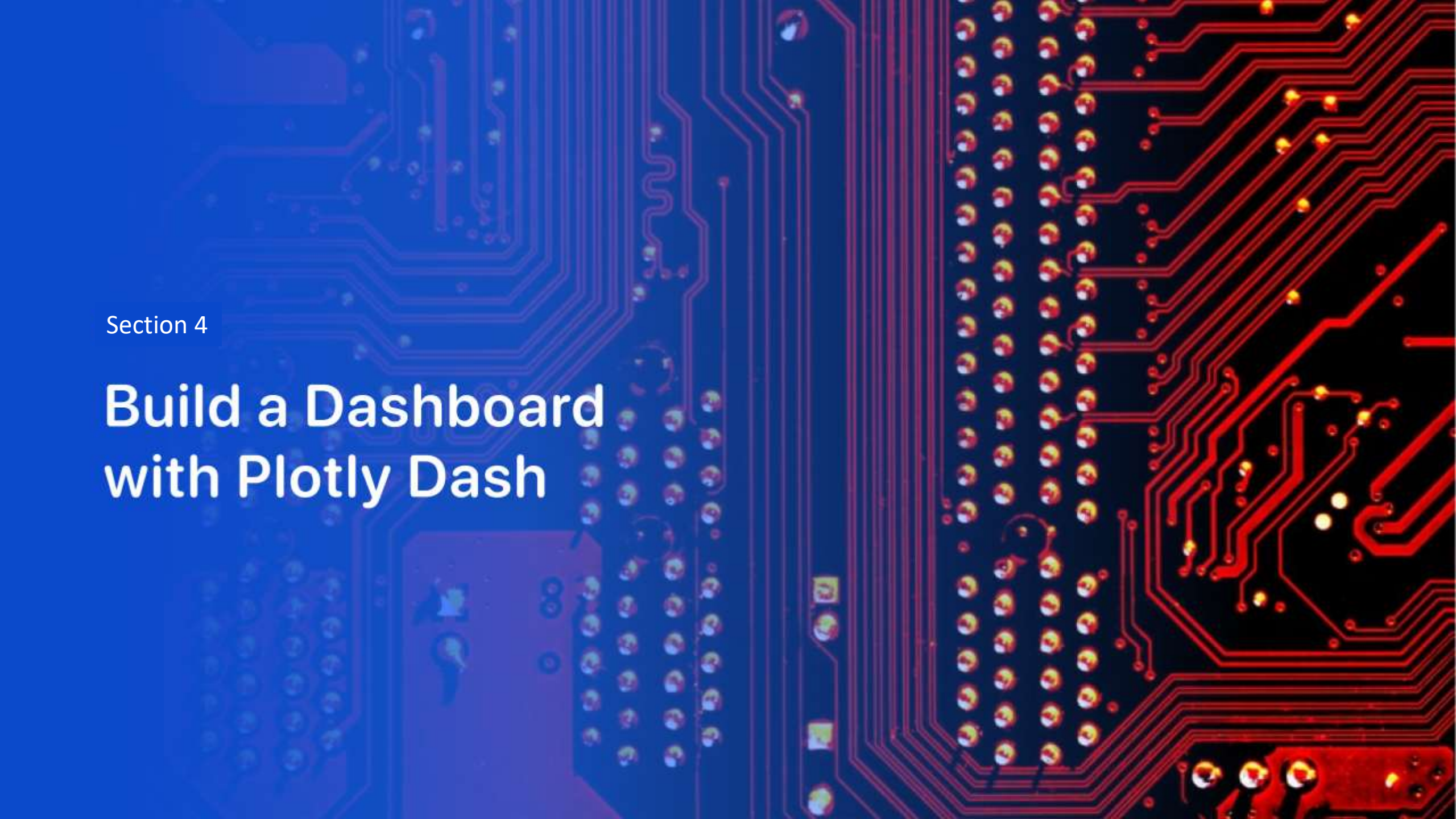
Colour-labeled launch records on the map

- Launches on each launch site were color coded:
  - Green = Successful launch
  - Red = Failed launch

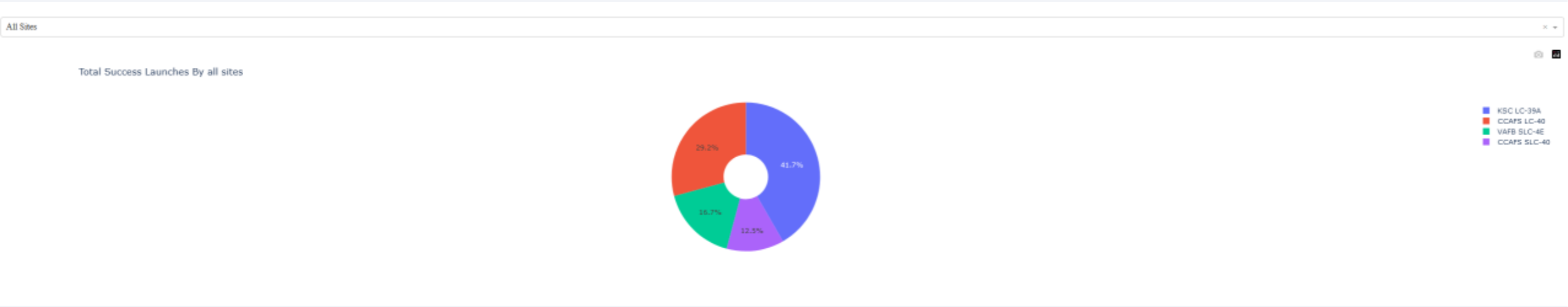
Distance from the launch site CCAFS SLC-40 to its proximities



- 29,21 km to the nearest highway
- 0,9km to nearest ocean

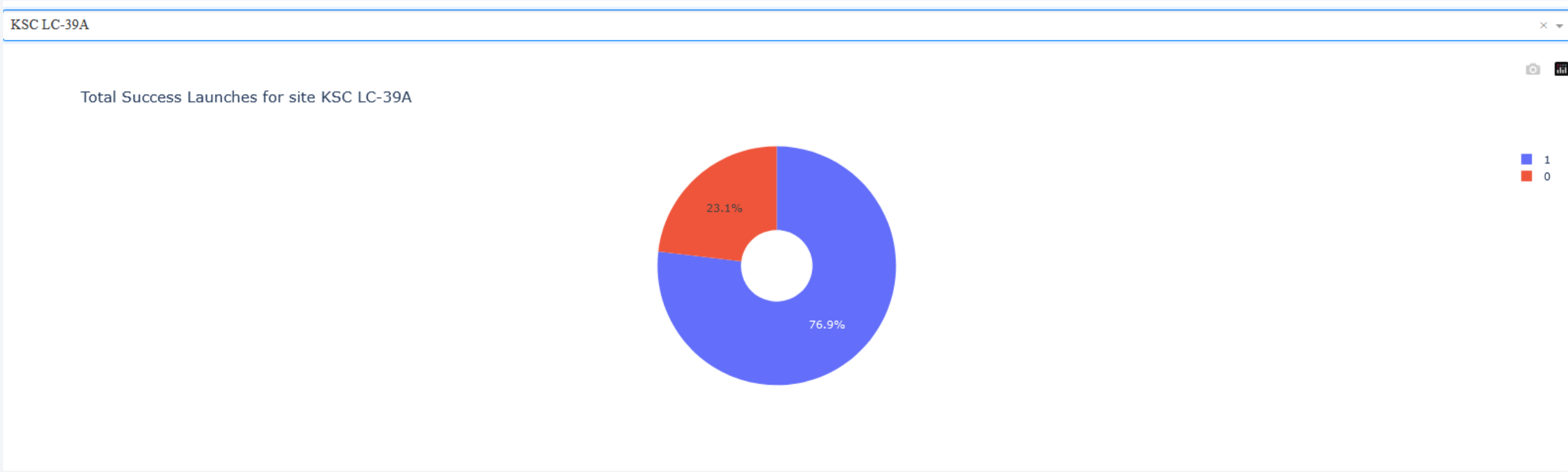

Section 4
Build a Dashboard with Plotly Dash

Launch success count for all sites



- The chart clearly shows that from all the sites, KSC LC-39A has the most successful launches

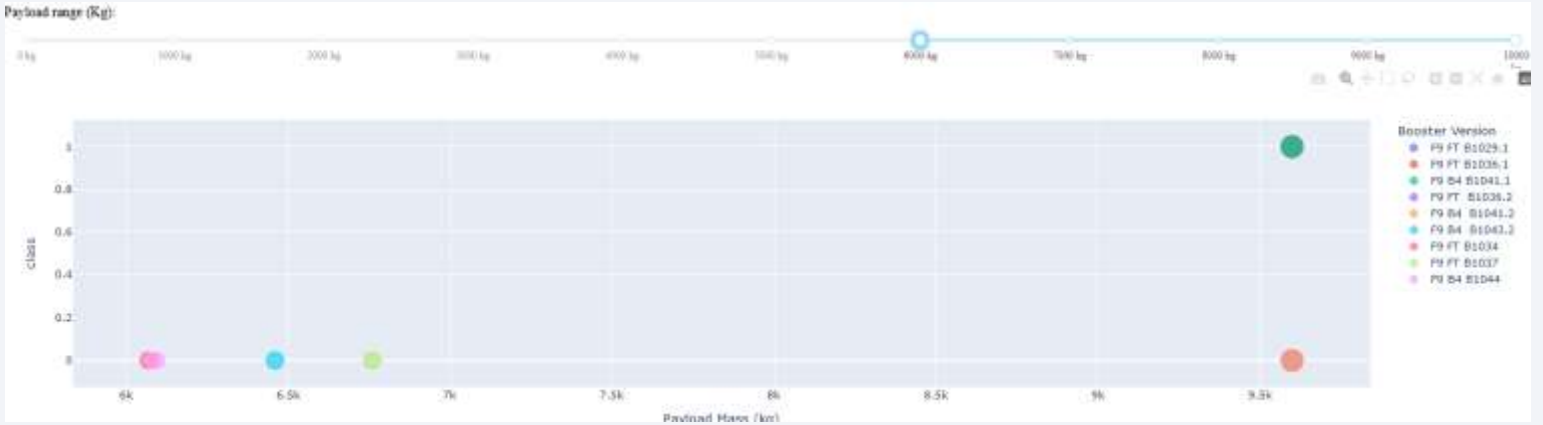
Launch site with highest launch success ratio



- The chart clearly shows that KSC LC-39A has the most successful launches with 77%

Payload Mass vs. Launch Outcome for all sites

- We can see that all the success rate for low weighted payload is higher than heavy weighted payload



Section 5

Predictive Analysis (Classification)

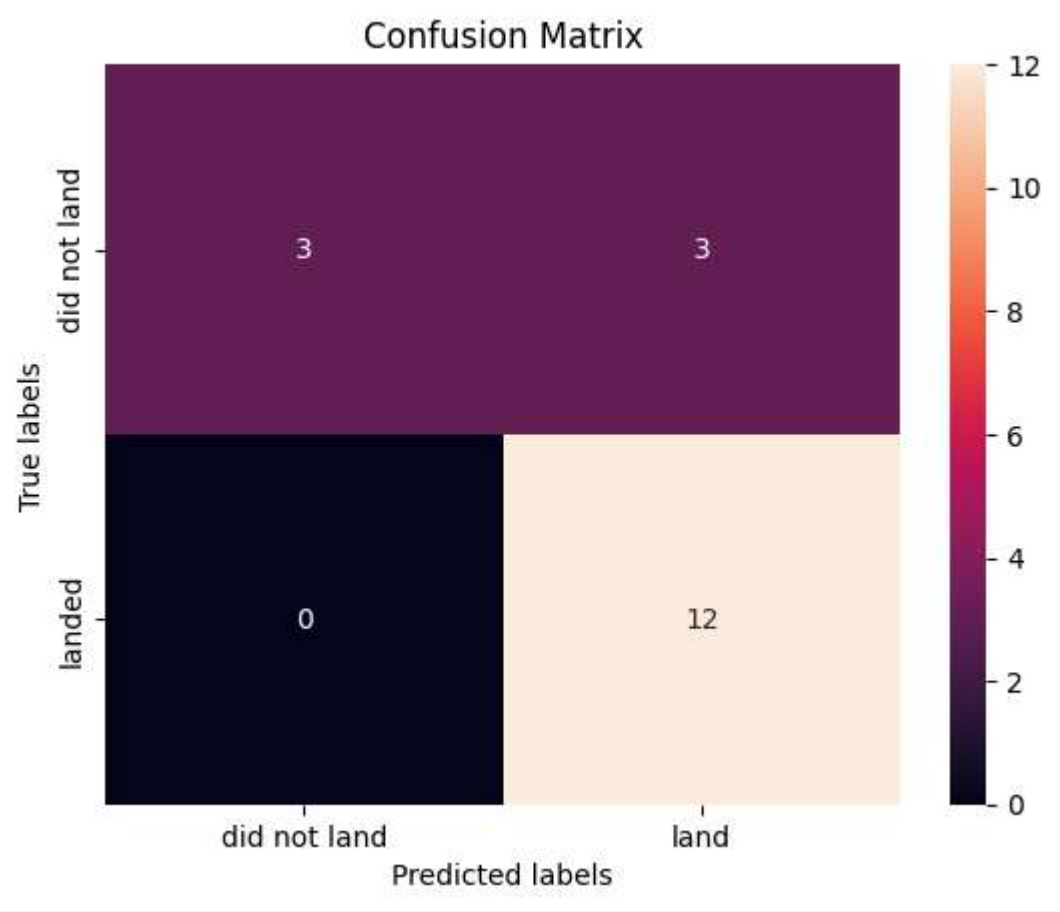
Classification Accuracy

- All models performed the same

	Model	Score
0	Logistic Regression	0.833333
1	SVM	0.833333
2	Decision Tree	0.833333
3	KNN	0.833333

Confusion Matrix

- Examining the confusion matrix, we see that logistic regression can distinguish between the different classes. We see that the major problem is false positives.



Conclusions

- All machine learning models performed the same.
- The low weighted payloads (which define as 4000kg and below) performed better than the heavy weighted payloads.
- Starting from the year 2013, the success rate for SpaceX launches is increased until 2020.
- KSC LC-39A have the most successful launches of any sites; 76.9%
- SSO orbit have the most success rate; 100% and more than 1 occurrence.

Thank you!