IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Kairav Kadakia
26 Oct 2022

Outline

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

Executive Summary

- Summary of methodologies
  - The project Predicts the landing success of Falcon 9 rocket. Multiple aspects of data science have been used including data wrangling, visualization, analysis and prediction.
- Summary of all results
  - We find that the success rate of the landings is around 66.67%
  - This rate has increased since 2013
  - Predictions show that our models are able to predict this value till 83.33% accuracy

Introduction

- Project background and context
  - We study past data of Falcon 9 landings to predict if the Falcon 9 will land successfully or not.
- Problems you want to find answers
  - If the Falcon 9 First stage will land scuccessfully

Section 1

Methodology

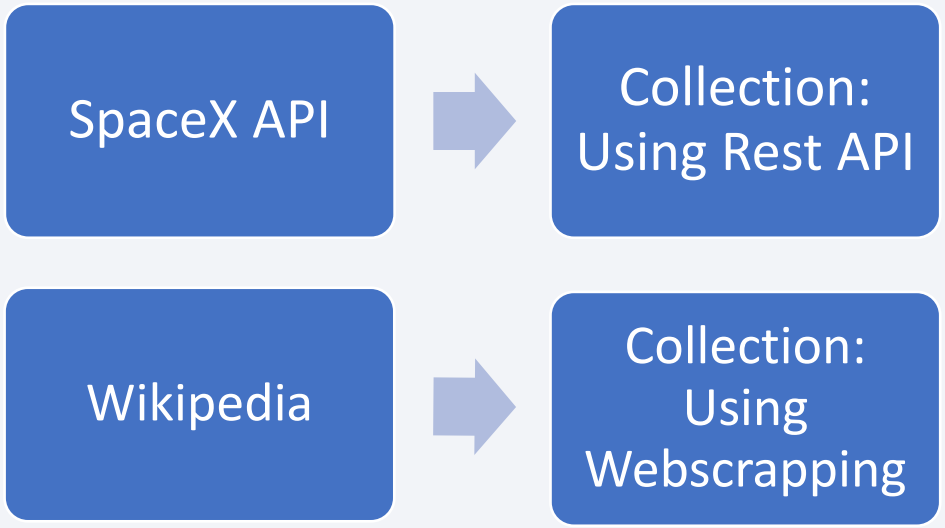
Methodology

Executive Summary

- Data collection methodology:
  - Data was collected from Open Source spaces using web scrapping
- Perform data wrangling
  - Exploratory data analysis was performed
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Different models for prediction were built, their best parameters tested

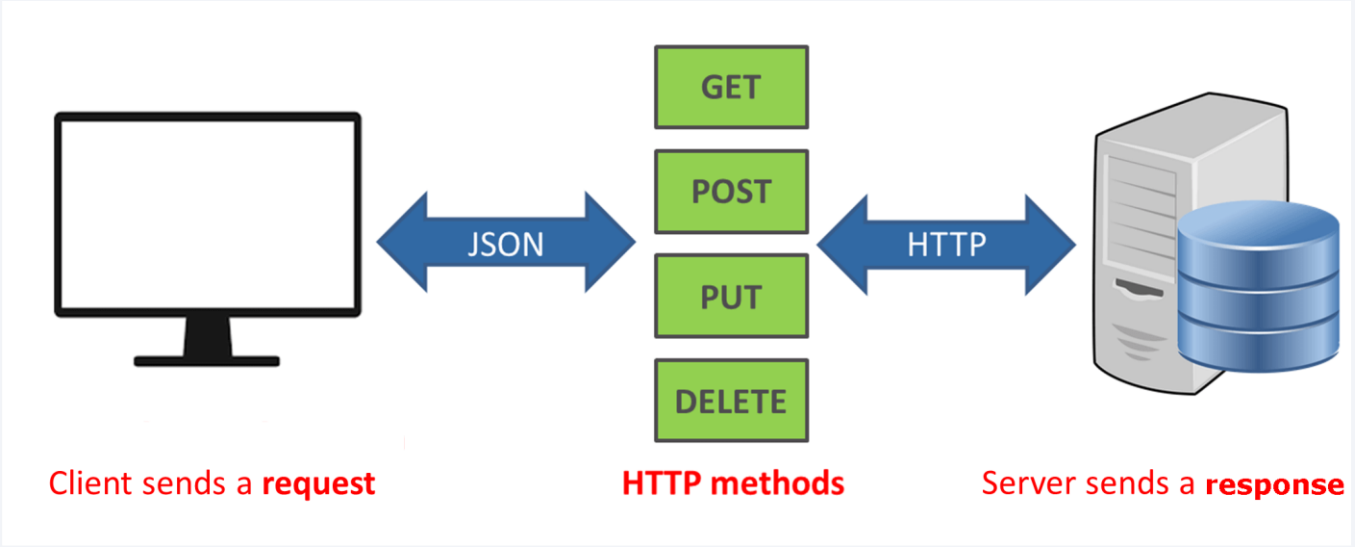
Data Collection

- Data set sources.
  - Space X API and Wikipedia
- Collection



Data Collection – SpaceX API

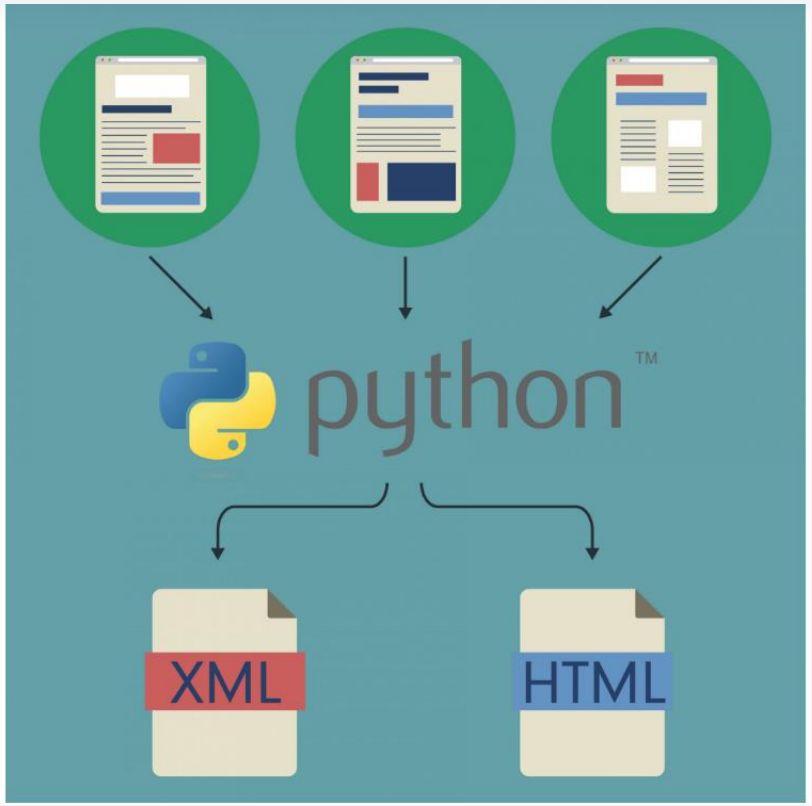
- Data was taken from the SpaceX Rest API using HTTP methods
- Git hub URL



Data Collection - Scraping

- Performed web scrapping using beautiful soup to obtain data from wikipedia

- Git hub URL



Data Wrangling

- Performed data wrangling by structuring and cleanining the data first. Then enriched it by replacing nulls
- Finally checked it for quality by using methods like shape, info and head.
- [Git hub URL]

EDA with Data Visualization

- Charts plotted include scatter plots, bar charts and line charts.
- Scatter plots allow us to see relation between variables
- Bar chart for the various categories of success
- Line chart to observe the yearly trends
- Git hub URL

EDA with SQL

- SQL queries were for the following

  - 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 failed landing_outcomes in drone ship, their booster versions, and launch site names for 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.

- Git hub URL

Build an Interactive Map with Folium

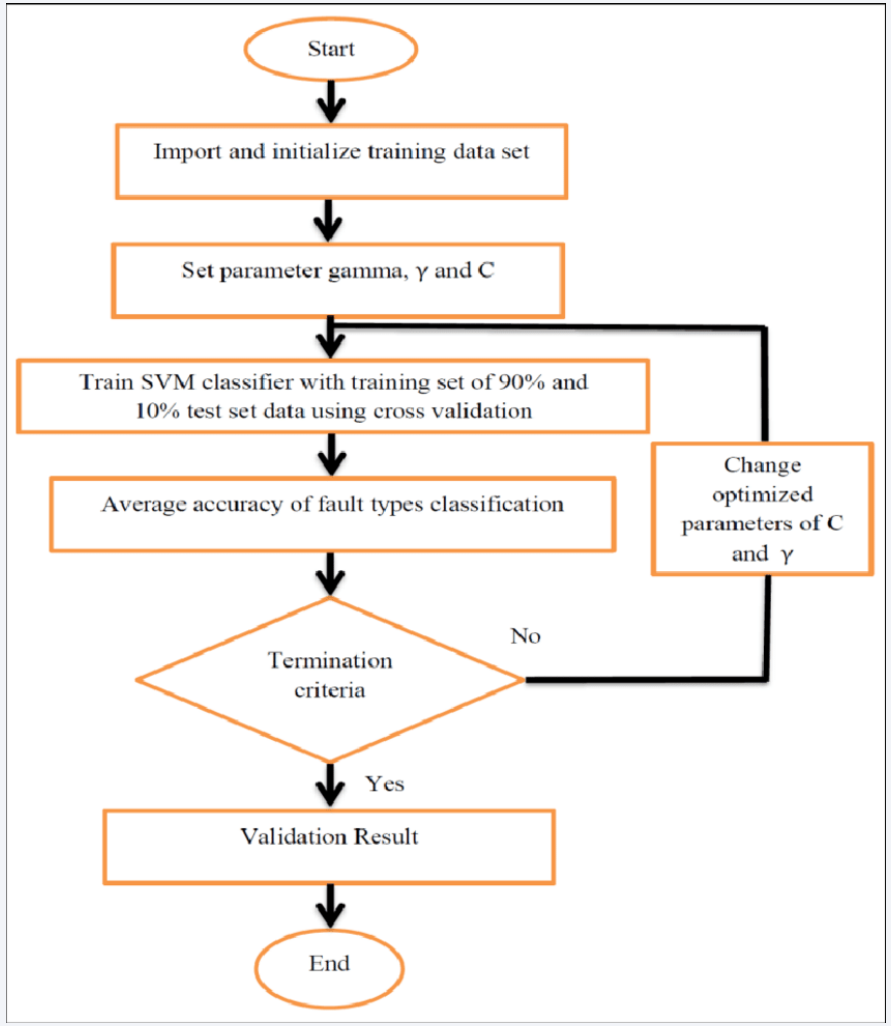
- Folium map used circles, markers, marker clusters, polyline, and mouse position.
- These were used for marking launch sites, their success and failure rates and proximities
- Git hub URL

Build a Dashboard with Plotly Dash

- Plotly dashboard has two main parts: Pie chart and Scatter plot
- Drop down menu helps support the pie chart
- Range slider is present to select the input for the scatter plot
- Git hub URL

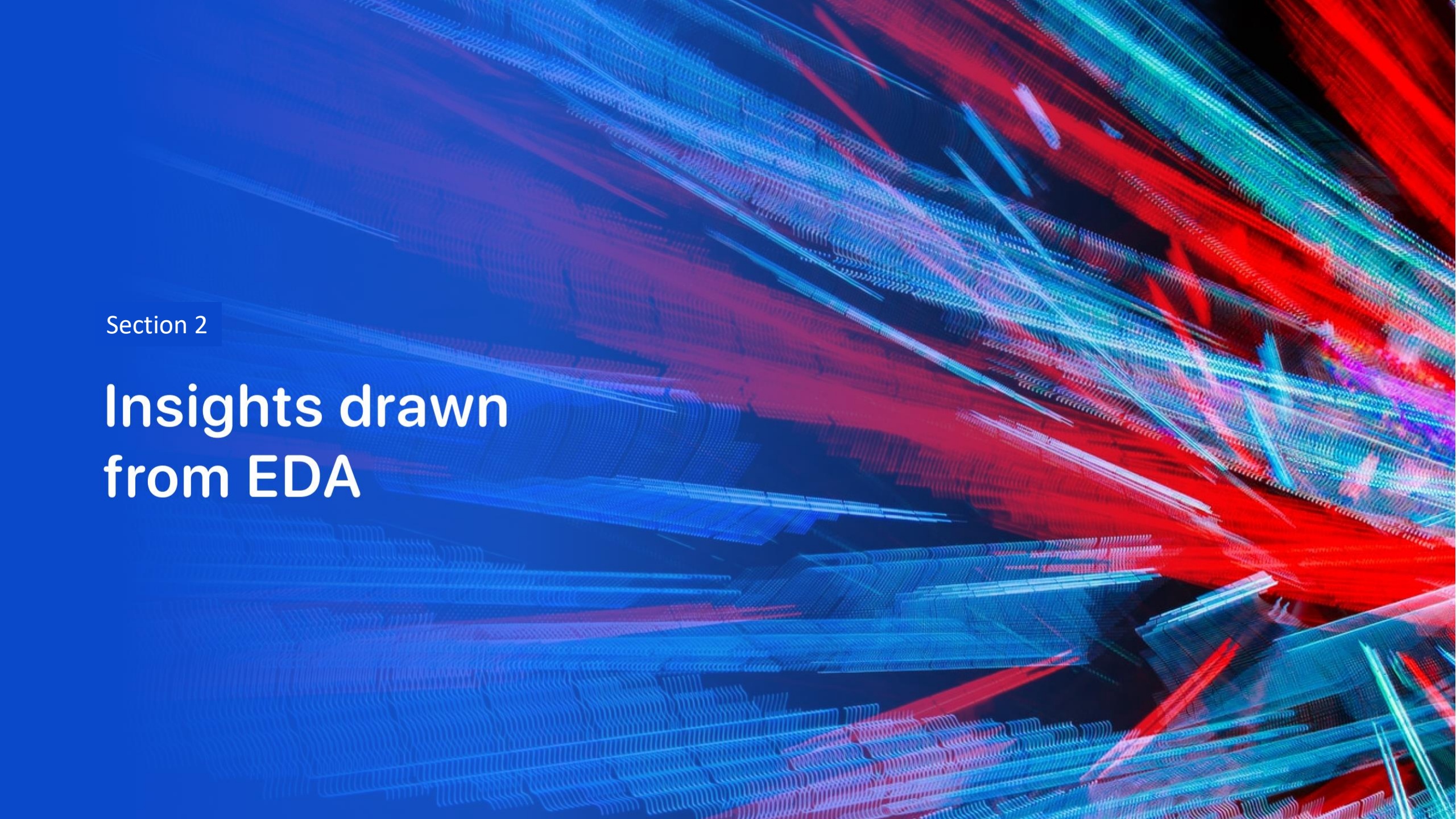
Predictive Analysis (Classification)

- Predictive analysis was done with mulliple models including knn, decision trees.
- Models were trained on training data and their best parameters was found
- Finally they were tested using test data to determine accuracy and performance
- [Git hub URL]



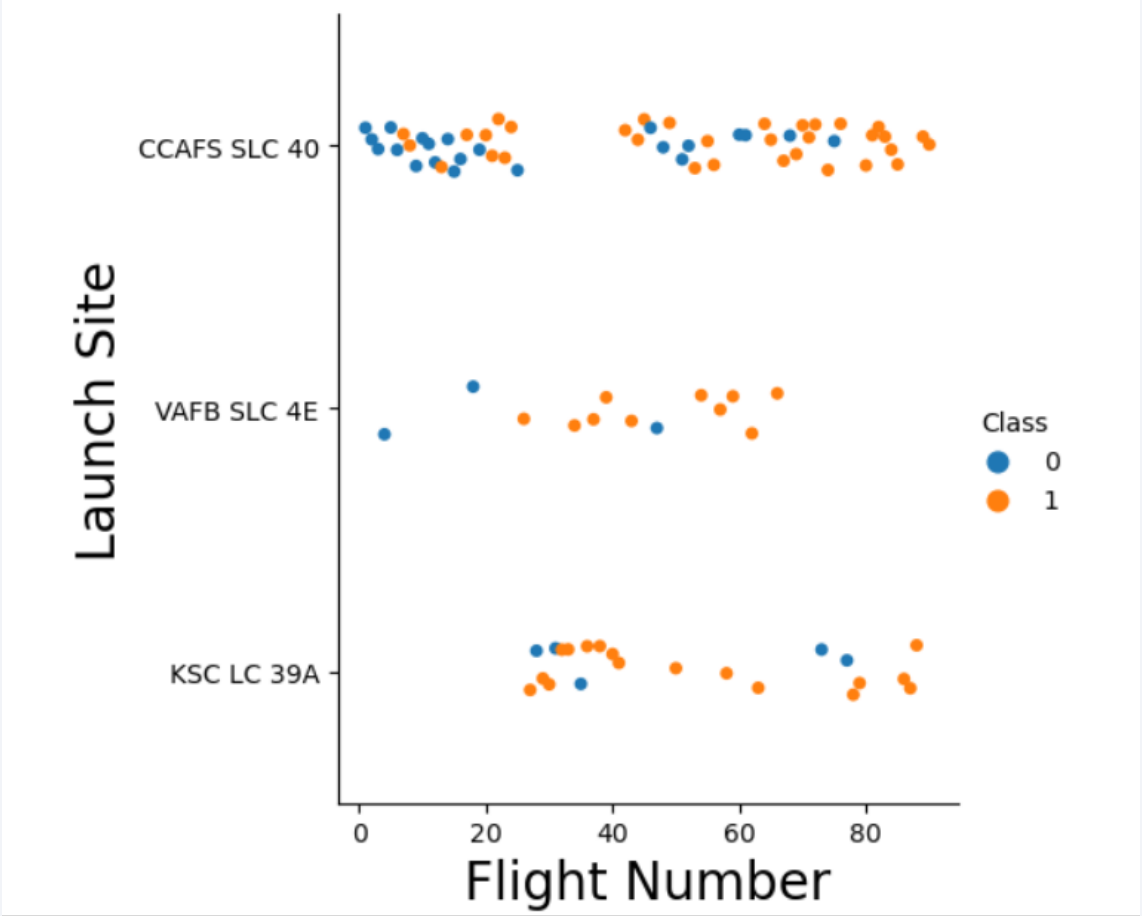
Results

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


Section 2
Insights drawn from EDA

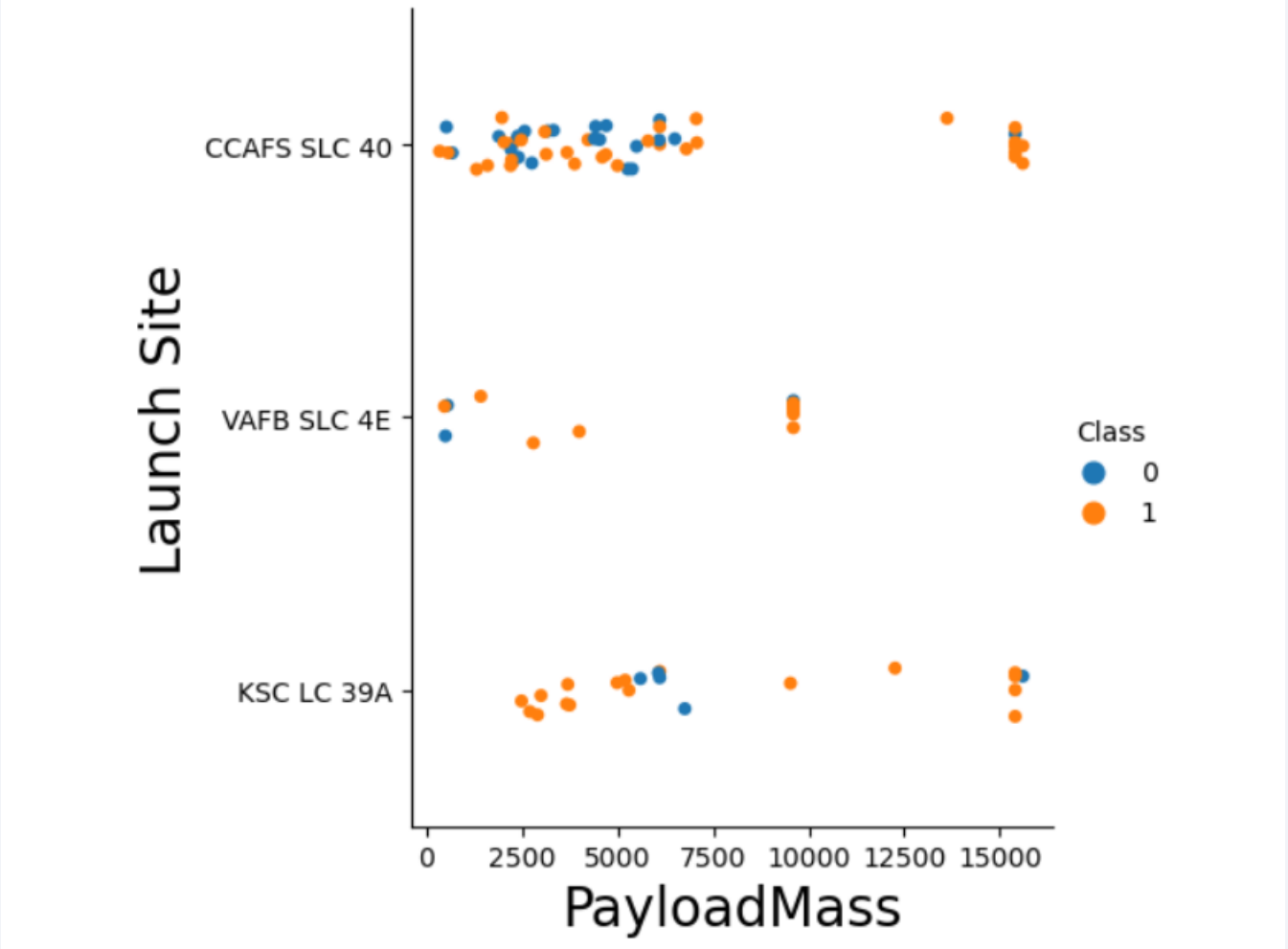
Flight Number vs. Launch Site

- Very few launches from second launch site
- 1st site has maximum launches



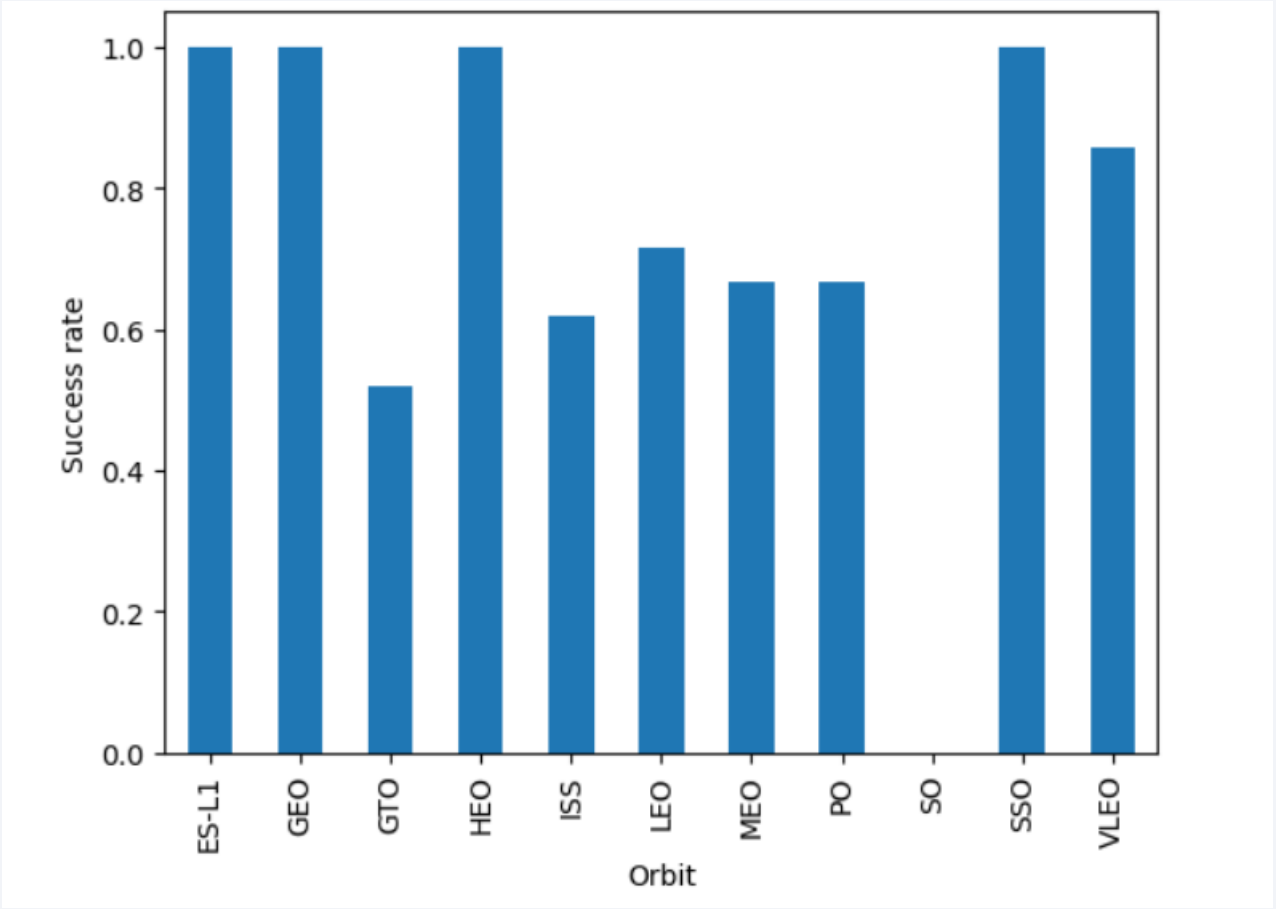
Payload vs. Launch Site

- No heavy payloads from second launch site



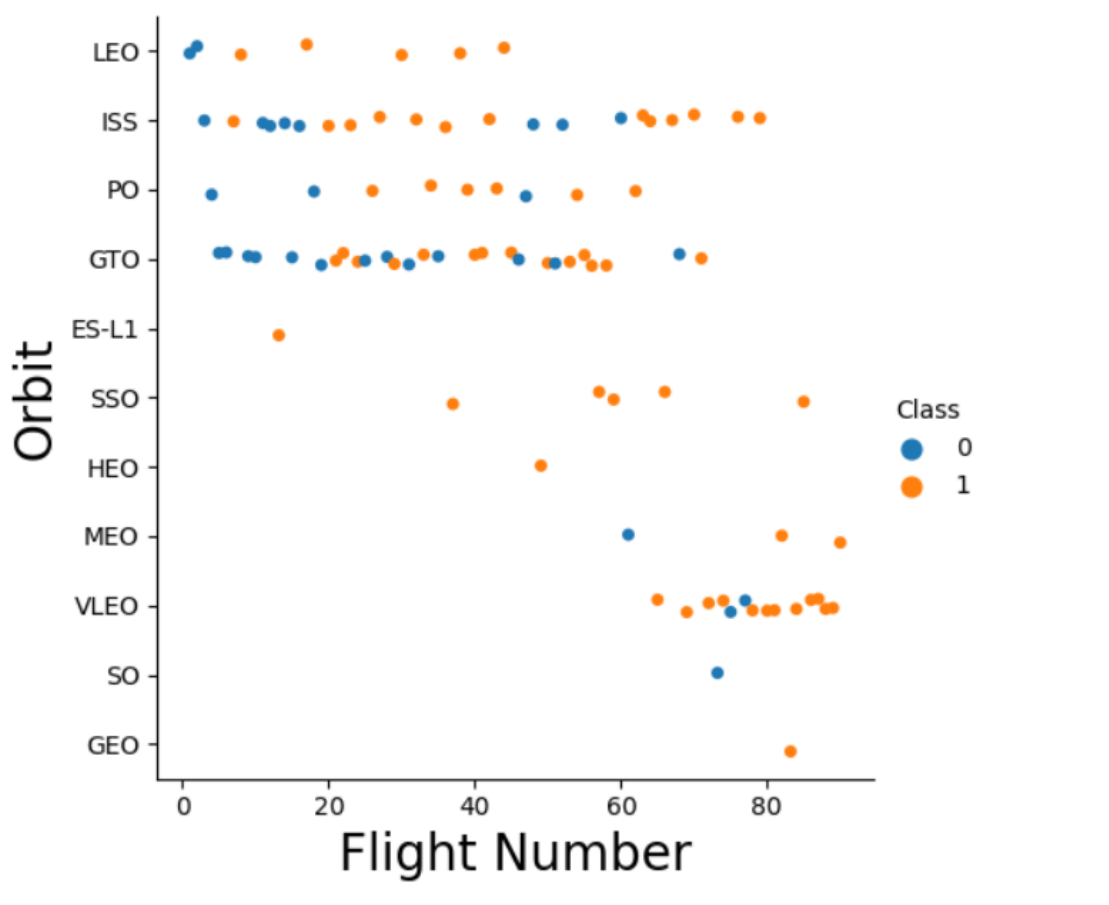
Success Rate vs. Orbit Type

- Most success rate in 4 orbits
- SO has no success



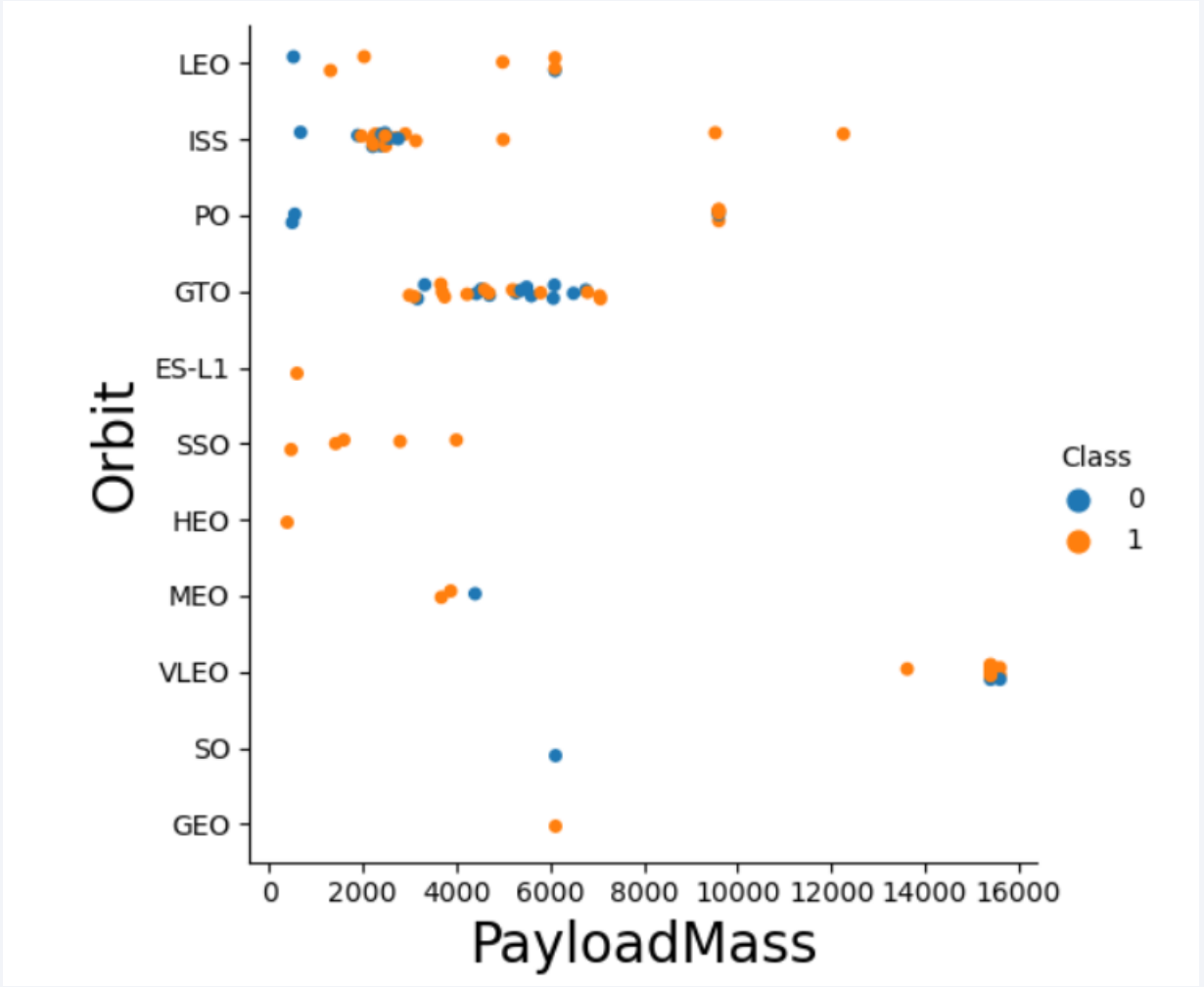
Flight Number vs. Orbit Type

- Leo orbit, success is related to number of flights
- Not the case in GTO orbit



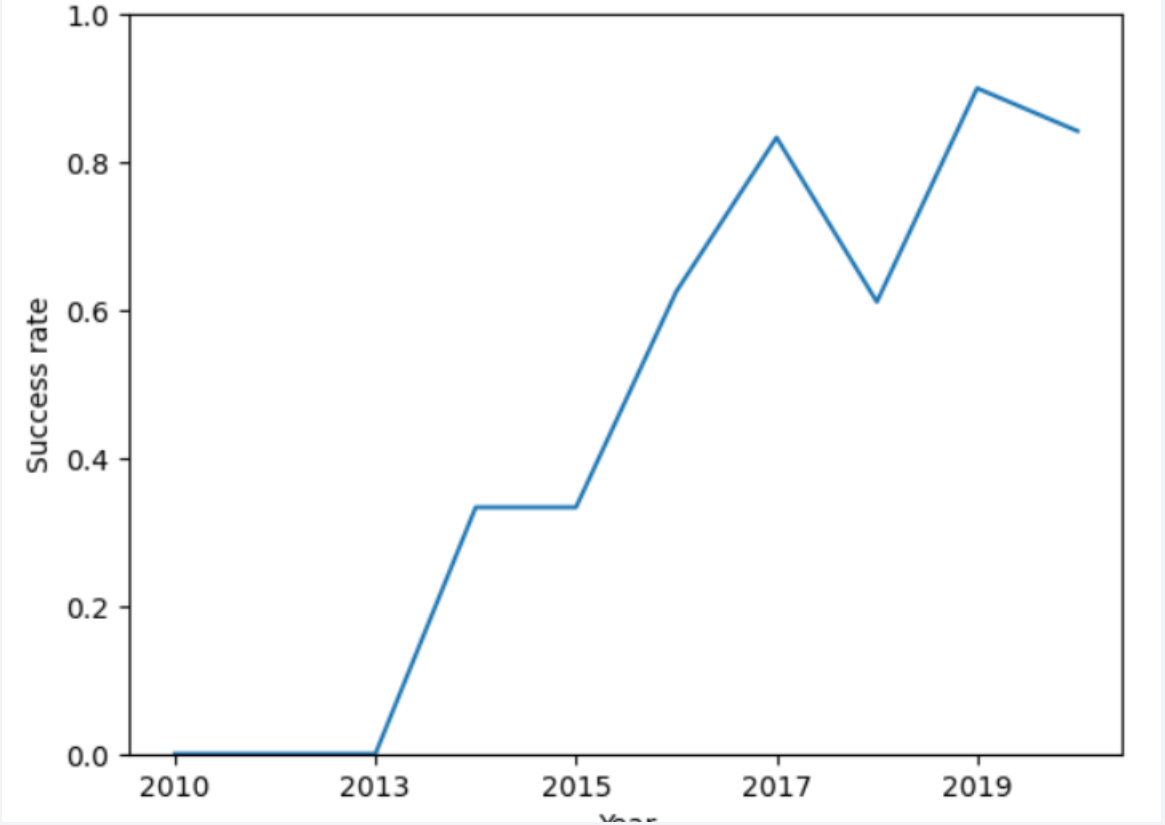
Payload vs. Orbit Type

- For heavy payloads, positive landing rate is more for Polar, Leo and ISS
- No such distension for GTO orbit



Launch Success Yearly Trend

- Launch rate has increased from 2013 to 2020



All Launch Site Names

```
%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'

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

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

Date	Time (UTC)	BoosterVersion	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	LandingOutcome
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

```
%sql SELECT SUM(payload_mass__kg_) AS total_payload_mass FROM SPACEXTBL where customer = 'NASA (CRS)'
```

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

total_payload_mass
45596

Average Payload Mass by F9 v1.1

```
%sql SELECT AVG(payload_mass__kg_) AS average_payload_mass FROM SPACEXTBL WHERE BoosterVersion = 'F9 v1.1'
```

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

average_payload_mass
2928.4

First Successful Ground Landing Date

```
%sql SELECT MIN(DATE) FROM SPACEXtbl WHERE LandingOutcome = 'Success (ground pad)'
```

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

MIN(DATE)
01-05-2017

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%sql select distinct BoosterVersion From Spacextbl WHERE (LandingOutcome = 'Success (drone ship)') AND (payload_mass__kg_ BETWEEN 4001 AND 5999)
```

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

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

Total Number of Successful and Failure Mission Outcomes

```
%sql SELECT mission_outcome, COUNT(mission_outcome) AS total_number FROM SPACEXtbl GROUP BY mission_outcome
```

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

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

Boosters Carried Maximum Payload

```
%sql SELECT BoosterVersion, payload_mass__kg_ FROM SPACEXTBL WHERE payload_mass__kg_ = (SELECT MAX(payload_mass__kg_) FROM SPACEXTBL)
```

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

BoosterVersion	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

```
%sql SELECT substr(Date, 4, 2), LandingOutcome, BoosterVersion, Launch_Site from SPACEXtbl WHERE LandingOutcome = 'Failure (drone ship)' and substr(Da
```

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

substr(Date, 4, 2)	LandingOutcome	BoosterVersion	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

```
%sql SELECT LandingOutcome, count(LandingOutcome) from SPACEXTBL where (DATE BETWEEN '04-06-2010' and '20-03-2017') and (LandingOutcome like '%Success
```

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

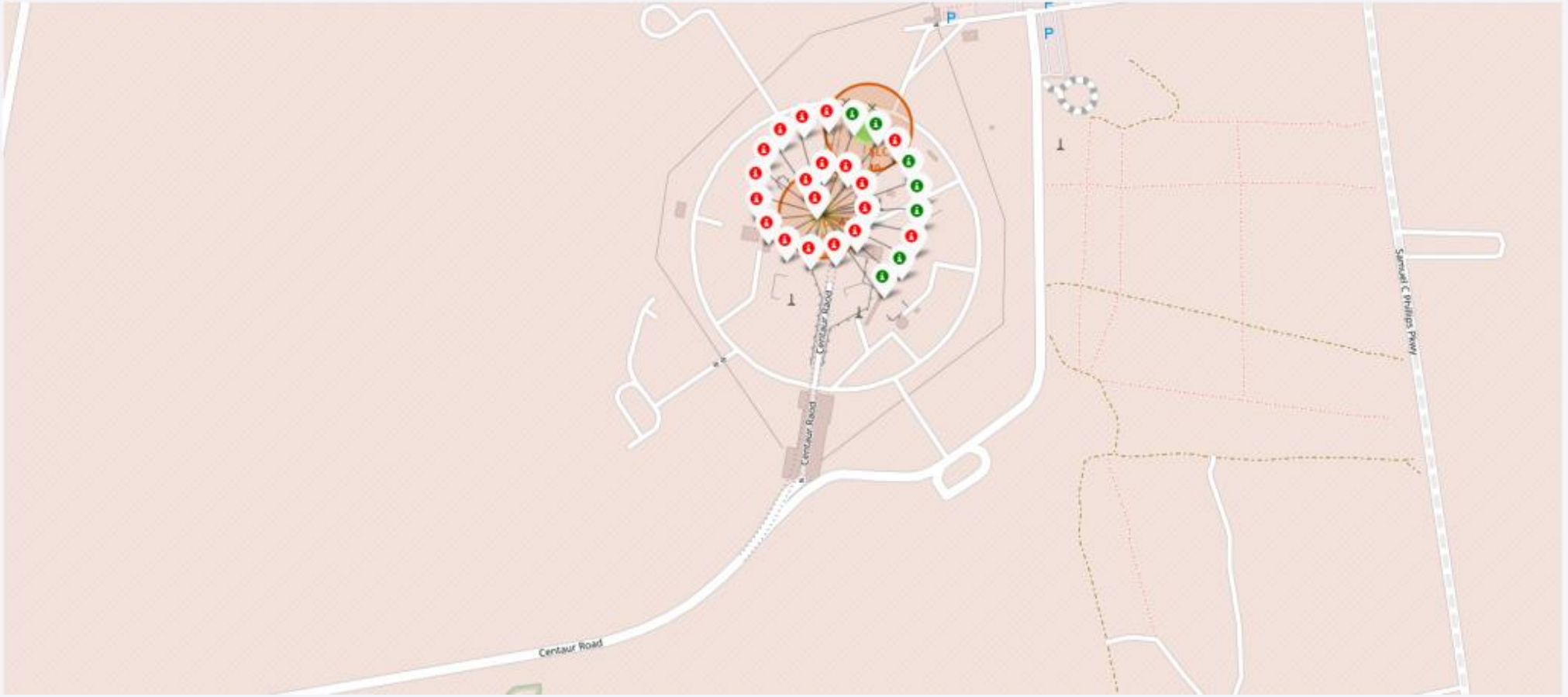
LandingOutcome	count(LandingOutcome)
Success	20
Success (drone ship)	8
Success (ground pad)	6

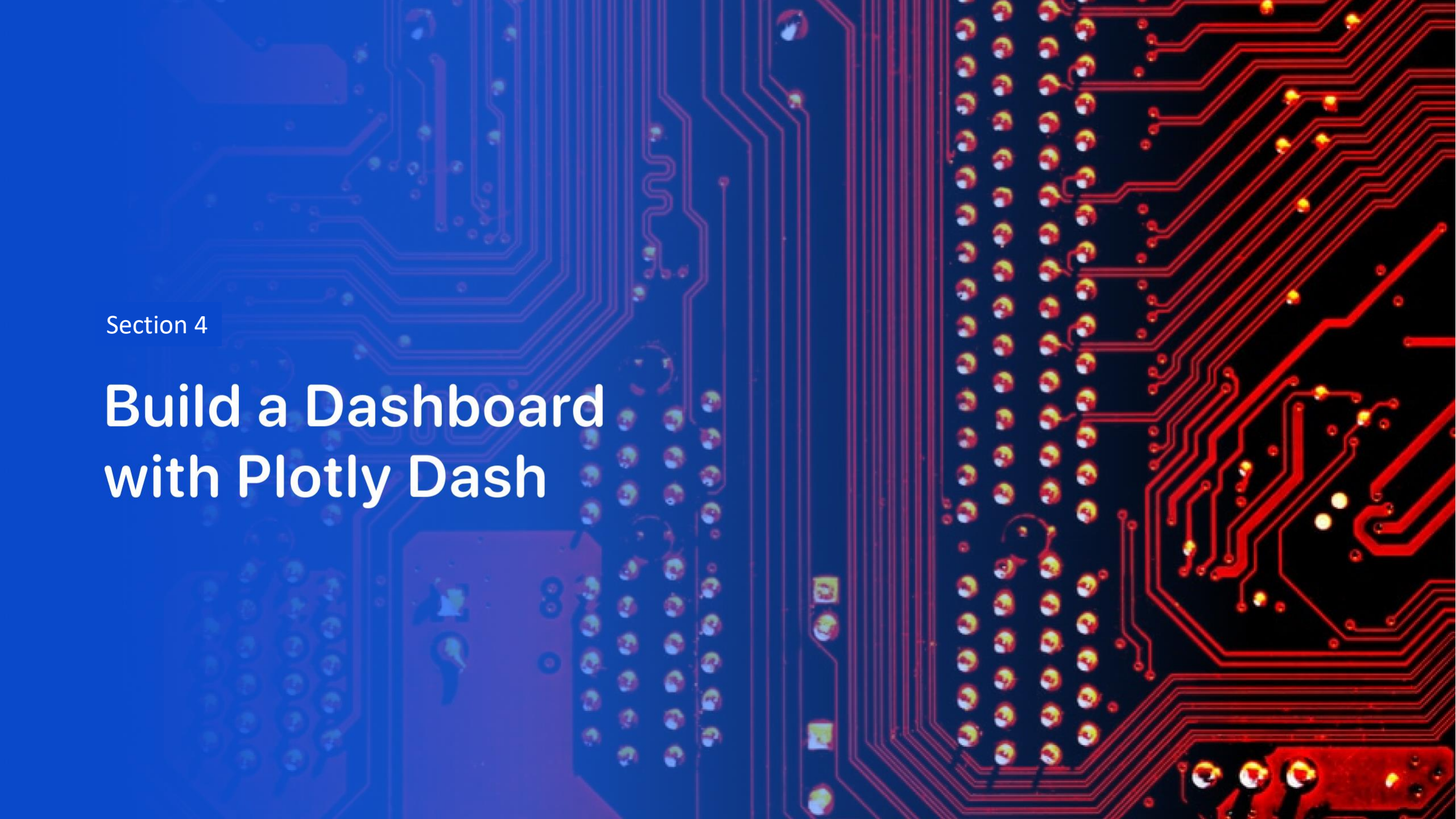
Section 3

Launch Sites Proximities Analysis

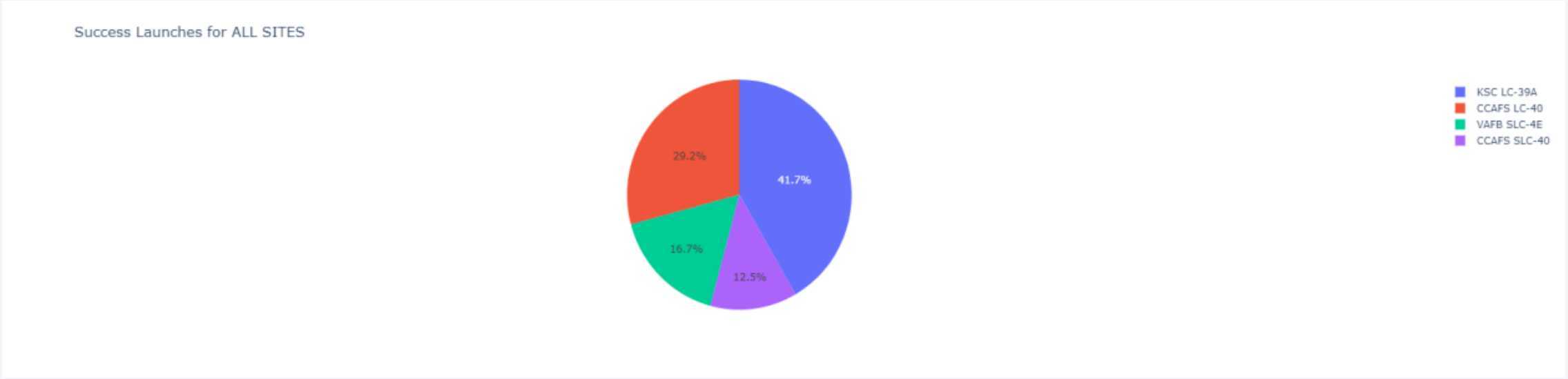
Launch Sites

Successful and failed Launches

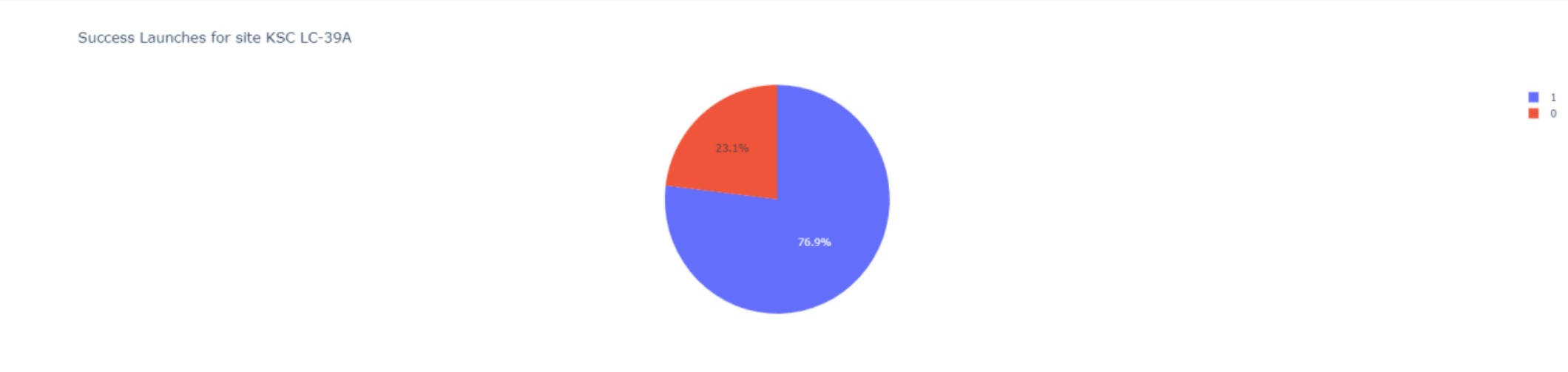



Section 4
Build a Dashboard with Plotly Dash

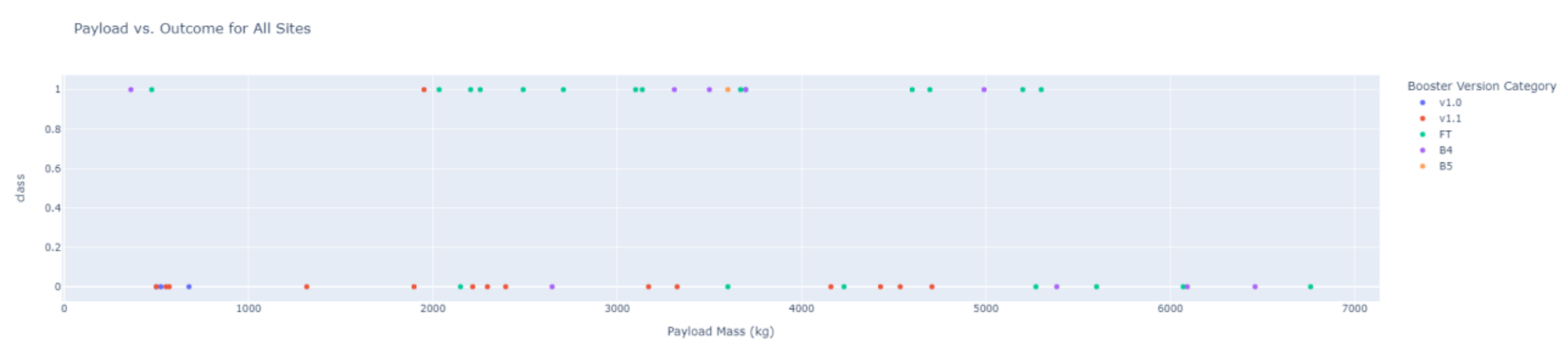
Successful launches



KSC site launch success ratio



Payload vs Launch outcomes



Section 5

Predictive Analysis (Classification)

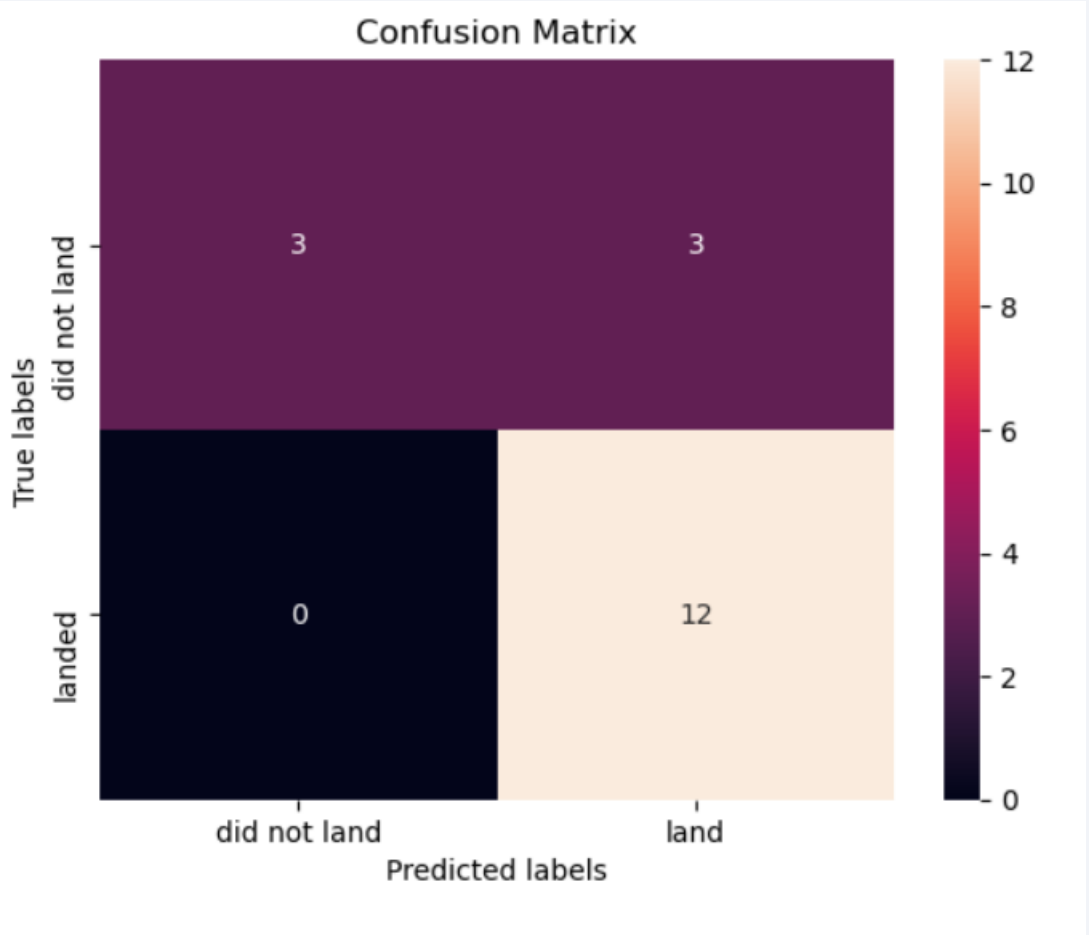
Classification Accuracy

- All models had similar accuracy >80%

model	acc
Logistic Regression	0.833333
SVM	0.833333
Decision Tree	0.833333
KNN	0.833333

Confusion Matrix

- We can see that there are 3 false positives
- All models gave same response



Conclusions

- Project was to predict a successful landing for Flacon 9
- All models were suitable for predicting, they all had an accuracy of >83%
- Successful landings were related to payload. Lighter the payload, better chance for a successful landing
- Landings were increasing in terms of success from 2013 to 2022

Appendix

- NA

Thank you!