IBM Developer
SKILLS NETWORK

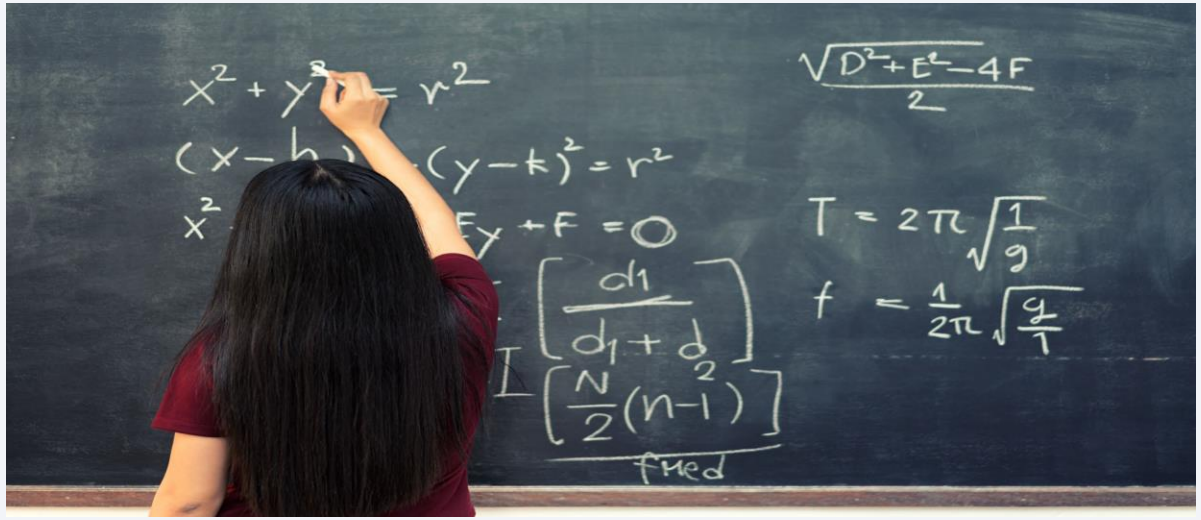
Winning Space Race with Data Science

Jason F
06/29/2023

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion

Executive Summary



- Space X claims to have lower cost on their rocket launch because they are able to reuse parts of previous launches.

- Data Shows there is an 83.3% chance of the First Stage Landing.

Introduction

- The Data we set out to collect is to help us SpaceY determine the probability of SpaceX reusing the First Stage.

- The Answers we seek are the chance of successful launch and reuse along with other factors such as launch site.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Utilized BeautifulSoup to Webscrape historical launch data for SpaceX
- Perform data wrangling
  - Processed Data into a Dataframe to determine which launches had a Success.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Utilized Python libraries for machine learning to choose the best predictive route.

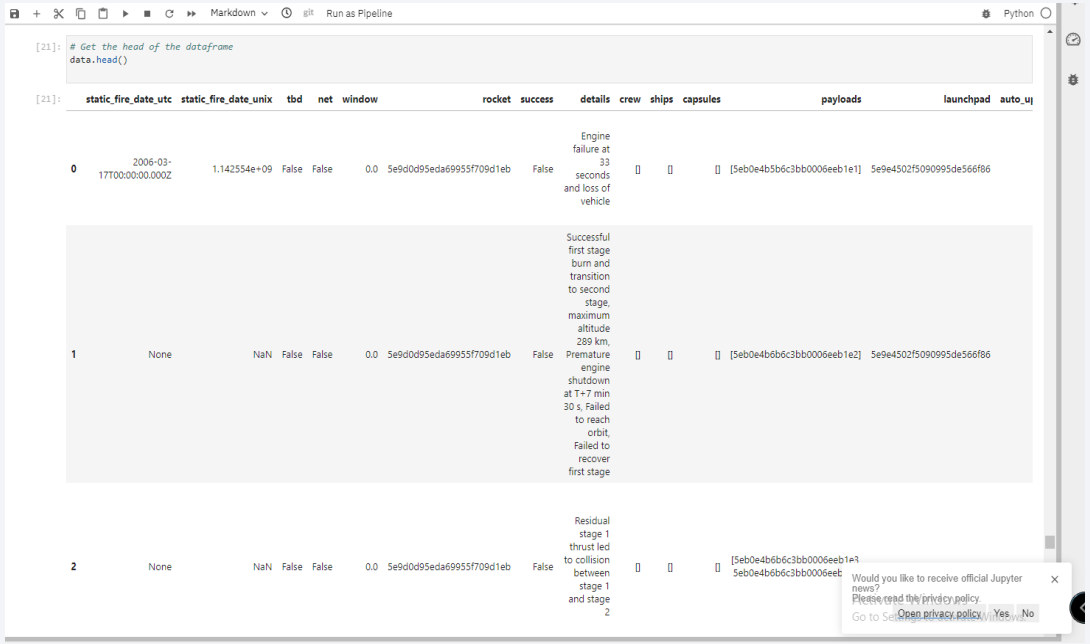
Data Collection

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

Data Collection – SpaceX API

We do a request and parse utilizing GET request and the SpaceXAPI.

BEFORE:



```
[21]: # Get the head of the dataframe
      data.head()
```

[21]:

	static_fire_date_utc	static_fire_date_unix	tbd	net	window	rocket	success	details	crew	ships	capsules	payloads	launchpad	auto_u
0	2006-03-17T00:00:00.000Z	1.142554e+09	False	False	0.0	5e9d0d95eda69955f709d1eb	False	Engine failure at 33 seconds and loss of vehicle	[]	[]	[]	[5eb0e4b5b6c3bb0006eeb1e1]	5e9e4502f5090995de566f86	
1	None	NaN	False	False	0.0	5e9d0d95eda69955f709d1eb	False	Successful first stage burn and transition to second stage, maximum altitude 289 km, Premature engine shutdown at T+7 min 30 s, Failed to reach orbit, Failed to recover first stage	[]	[]	[]	[5eb0e4b6b6c3bb0006eeb1e2]	5e9e4502f5090995de566f86	
2	None	NaN	False	False	0.0	5e9d0d95eda69955f709d1eb	False	Residual stage 1 thrust led to collision between stage 1 and stage 2	[]	[]	[]	[5eb0e4b6b6c3bb0006eeb1e3, 5eb0e4b6b6c3bb0006eeb		

AFTER:

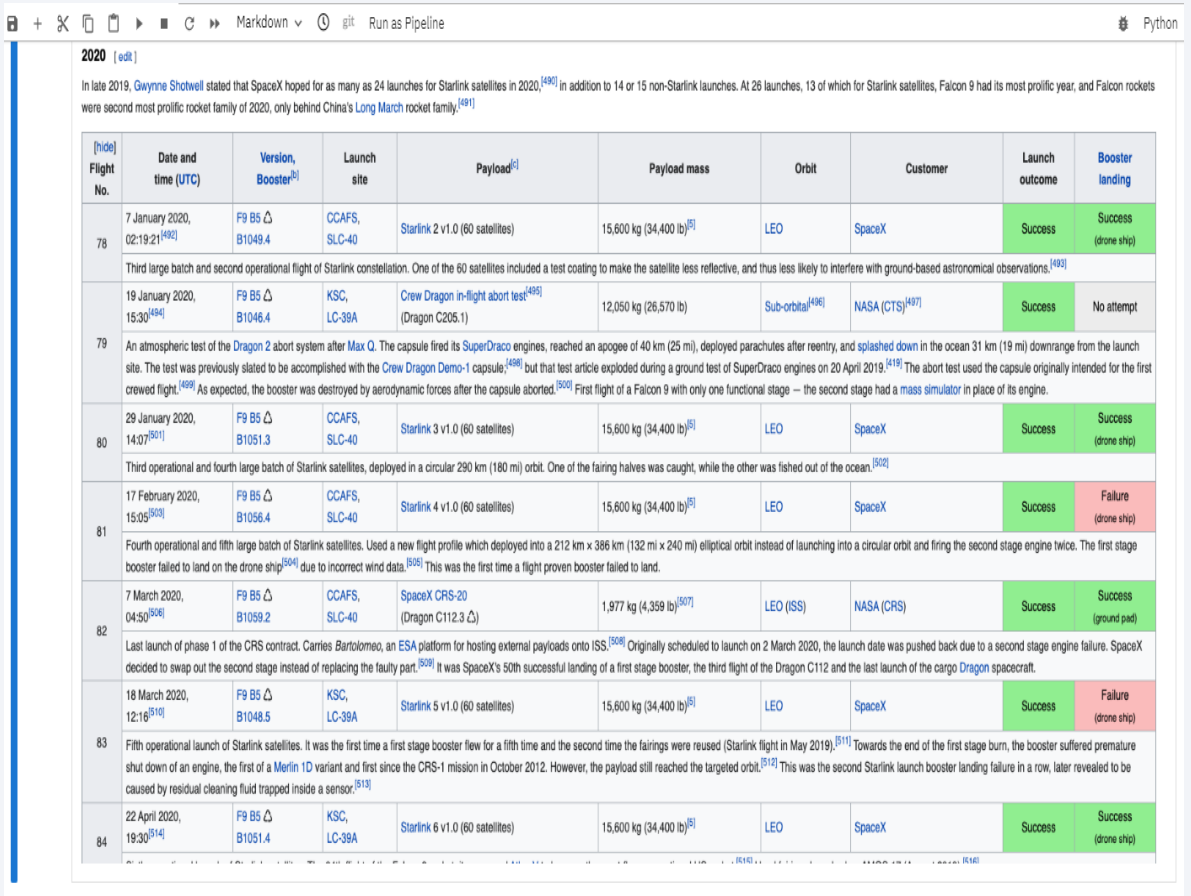
[35]:

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude
0	1	2006-03-24	Falcon 1	20.0	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	Merlin1A	167.743129	9.047721
1	2	2007-03-21	Falcon 1	NaN	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	Merlin2A	167.743129	9.047721
2	4	2008-09-28	Falcon 1	165.0	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	Merlin2C	167.743129	9.047721
3	5	2009-07-13	Falcon 1	200.0	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	Merlin3C	167.743129	9.047721
4	6	2010-06-04	Falcon 9	NaN	LEO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0003	-80.577366	28.561857

Data Collection - Scraping

- Used BeautifulSoup to take in a web page and convert the HTML into JSON so we could parse information into a readable data frame.

- To the right is an HTML table we had to convert into a Dataframe.



Markdown git Run as Pipeline Python

2020 [edit]

In late 2019, Gwynne Shotwell stated that SpaceX hoped for as many as 24 launches for Starlink satellites in 2020,[490] in addition to 14 or 15 non-Starlink launches. At 26 launches, 13 of which for Starlink satellites, Falcon 9 had its most prolific year, and Falcon rockets were second most prolific rocket family of 2020, only behind China's Long March rocket family.[491]

[hide] Flight No.	Date and time (UTC)	Version, Booster[b]	Launch site	Payload[c]	Payload mass	Orbit	Customer	Launch outcome	Booster landing
78	7 January 2020, 02:19:21[492]	F9 B5 △ B1049.4	CCAFS, SLC-40	Starlink 2 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Success (drone ship)
	Third large batch and second operational flight of Starlink constellation. One of the 60 satellites included a test coating to make the satellite less reflective, and thus less likely to interfere with ground-based astronomical observations.[493]								
79	19 January 2020, 15:30[494]	F9 B5 △ B1046.4	KSC, LC-39A	Crew Dragon in-flight abort test[495] (Dragon C205.1)	12,050 kg (26,570 lb)	Sub-orbital[496]	NASA (CTS)[497]	Success	No attempt
	An atmospheric test of the Dragon 2 abort system after Max Q. The capsule fired its SuperDraco engines, reached an apogee of 40 km (25 mi), deployed parachutes after reentry, and splashed down in the ocean 31 km (19 mi) downrange from the launch site. The test was previously slated to be accomplished with the Crew Dragon Demo-1 capsule,[498] but that test article exploded during a ground test of SuperDraco engines on 20 April 2019.[419] The abort test used the capsule originally intended for the first crewed flight.[499] As expected, the booster was destroyed by aerodynamic forces after the capsule aborted.[500] First flight of a Falcon 9 with only one functional stage – the second stage had a mass simulator in place of its engine.								
80	29 January 2020, 14:07[501]	F9 B5 △ B1051.3	CCAFS, SLC-40	Starlink 3 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Success (drone ship)
	Third operational and fourth large batch of Starlink satellites, deployed in a circular 290 km (180 mi) orbit. One of the fairing halves was caught, while the other was fished out of the ocean.[502]								
81	17 February 2020, 15:05[503]	F9 B5 △ B1056.4	CCAFS, SLC-40	Starlink 4 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Failure (drone ship)
	Fourth operational and fifth large batch of Starlink satellites. Used a new flight profile which deployed into a 212 km × 386 km (132 mi × 240 mi) elliptical orbit instead of launching into a circular orbit and firing the second stage engine twice. The first stage booster failed to land on the drone ship[504] due to incorrect wind data.[505] This was the first time a flight proven booster failed to land.								
82	7 March 2020, 04:50[506]	F9 B5 △ B1059.2	CCAFS, SLC-40	SpaceX CRS-20 (Dragon C112.3 △)	1,977 kg (4,359 lb)[507]	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
	Last launch of phase 1 of the CRS contract. Carries *Bartolomeo*, an ESA platform for hosting external payloads onto ISS.[508] Originally scheduled to launch on 2 March 2020, the launch date was pushed back due to a second stage engine failure. SpaceX decided to swap out the second stage instead of replacing the faulty part.[509] It was SpaceX's 50th successful landing of a first stage booster, the third flight of the Dragon C112 and the last launch of the cargo Dragon spacecraft.								
83	18 March 2020, 12:16[510]	F9 B5 △ B1048.5	KSC, LC-39A	Starlink 5 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Failure (drone ship)
	Fifth operational launch of Starlink satellites. It was the first time a first stage booster flew for a fifth time and the second time the fairings were reused (Starlink flight in May 2019).[511] Towards the end of the first stage burn, the booster suffered premature shut down of an engine, the first of a Merlin 1D variant and first since the CRS-1 mission in October 2012. However, the payload still reached the targeted orbit.[512] This was the second Starlink launch booster landing failure in a row, later revealed to be caused by residual cleaning fluid trapped inside a sensor.[513]								
84	22 April 2020, 19:30[514]	F9 B5 △ B1051.4	KSC, LC-39A	Starlink 6 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Success (drone ship)

Data Wrangling

We calculate the number of launches from each site.

- Calculate occurrence of each orbit
- Determine the mission outcome per orbit type
- Convert outcomes into a number based on outcome on a new column
- Zero for Failed
- One for Success

[30]:

	Class
0	0
1	0
2	0
3	0
4	0
5	0
6	1
7	1

EDA with Data Visualization

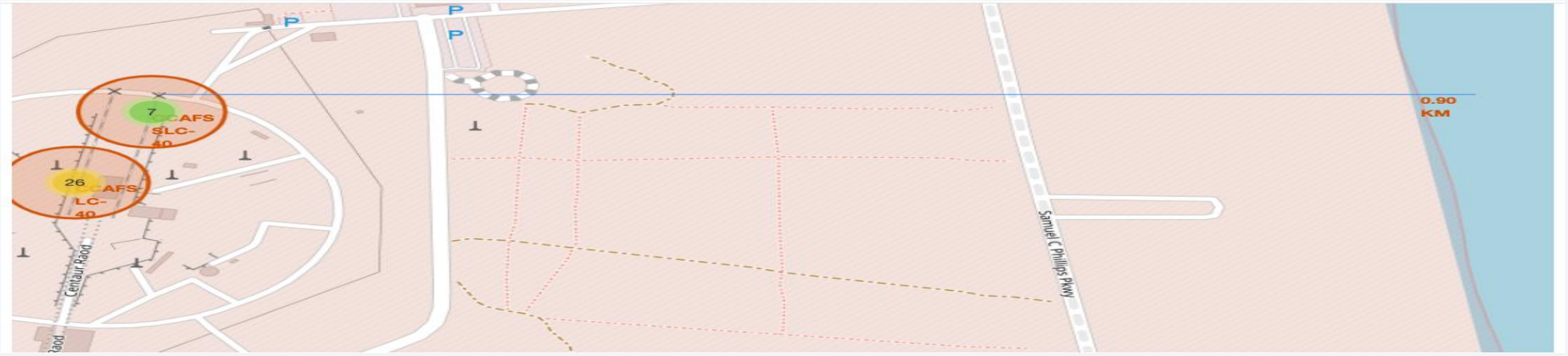
- Summarize what charts were plotted and why you used those charts
- Used a ScatterPlot to visualize the relationship between Flightnumbers and payload mass.
- Followed by Launch Site information compared with Payload mass.
- A Bar Graph was choose to visualize the success rate of each oribit.
- Scatter Plotted a Payload vs Orbit to reveal any relationship between the two.
- A line graph was used to pair success rate and year to show a steady incline of success rate increase over the years.

EDA with SQL

- Using SELECT and DISTINCT I pull information on Launch Sites
- Used SUM on the PAYLOAD column to find the total amount of payload in kg.
- Used BETWEEN to find successful launches with a payload mass between 4000 and 6000kg.
- Used COUNT(*),GROUP BY, and ORDER BY to list total amount of successful and failed launches.
- Utilized '%2015%' to search for a specfic date.

Build an Interactive Map with Folium

- Summarize what map objects such as markers, circles, lines, etc. you created and added to a folium map
- Designed an interactive map to mark launch sites and find their proximity to coastlines, railroad tracks, highways and cities.
- I discovered launch sites have a close proximity to coastlines , average proximity to highways but have decent distance from railroads and cities Understandably.



Build a Dashboard with Plotly Dash

- Created a drop down to represent showing a pie chart.
- Also created a Scatter Plot with a slider to choose payload size.
- This allowed users to scale payload and visualize a change if any.
- The pie chart selector allowed users to view launch sites as a whole or see specific success or failure rates from each launch location.

Predictive Analysis (Classification)

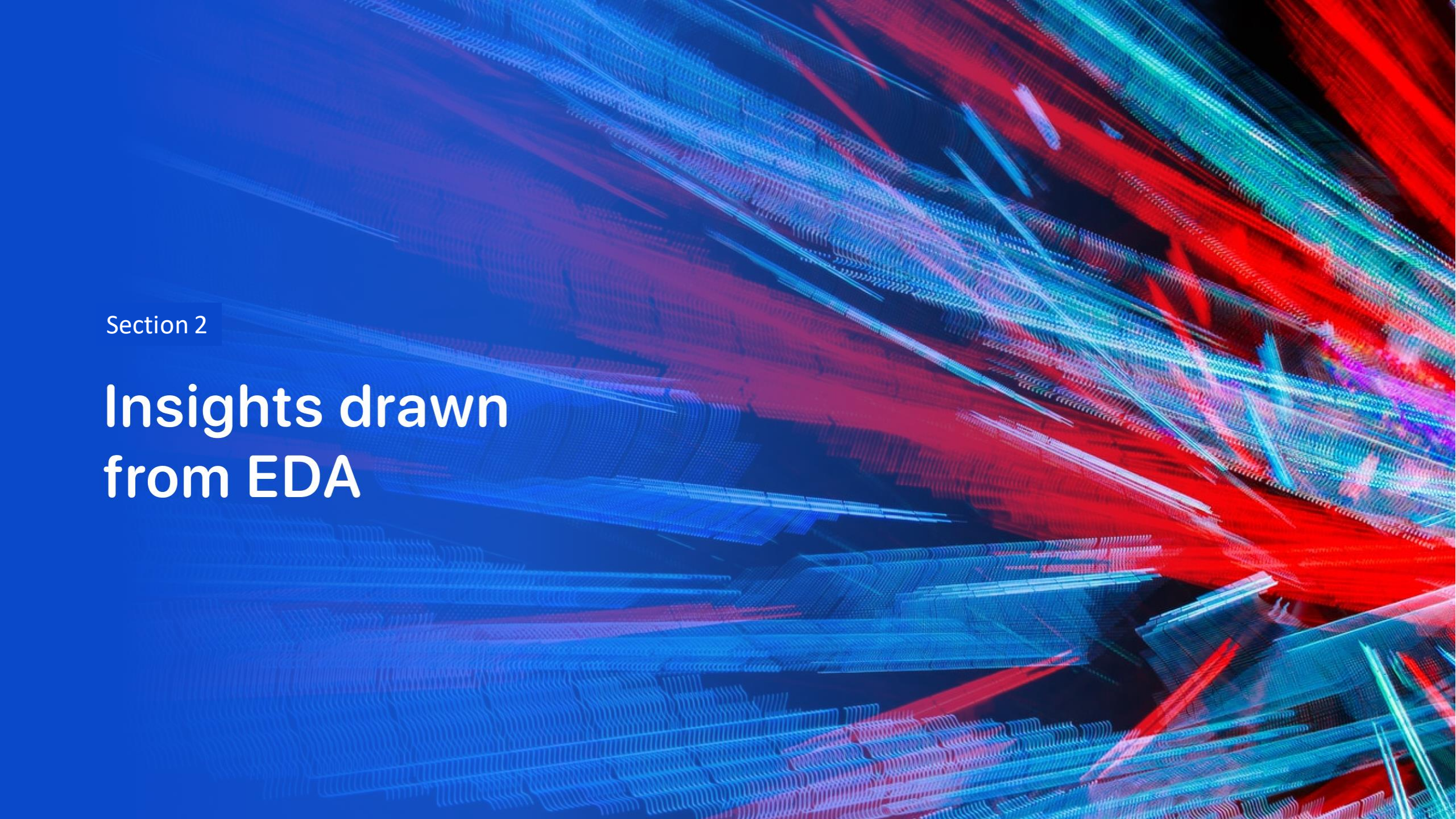
- Summarize how you built, evaluated, improved, and found the best performing classification model
- Created a NumPy array from the Dataframe and Transfomed it.
- Ran the Transformed Data into a train_test_split function
- Then trained a logistic regression model and tested its accuracy.
- Tested the accuracy of a SVC(Support-Vector), a Decision Tree, and K-nearest-neighbor.
- I favored the Decision Tree in the end.

Results

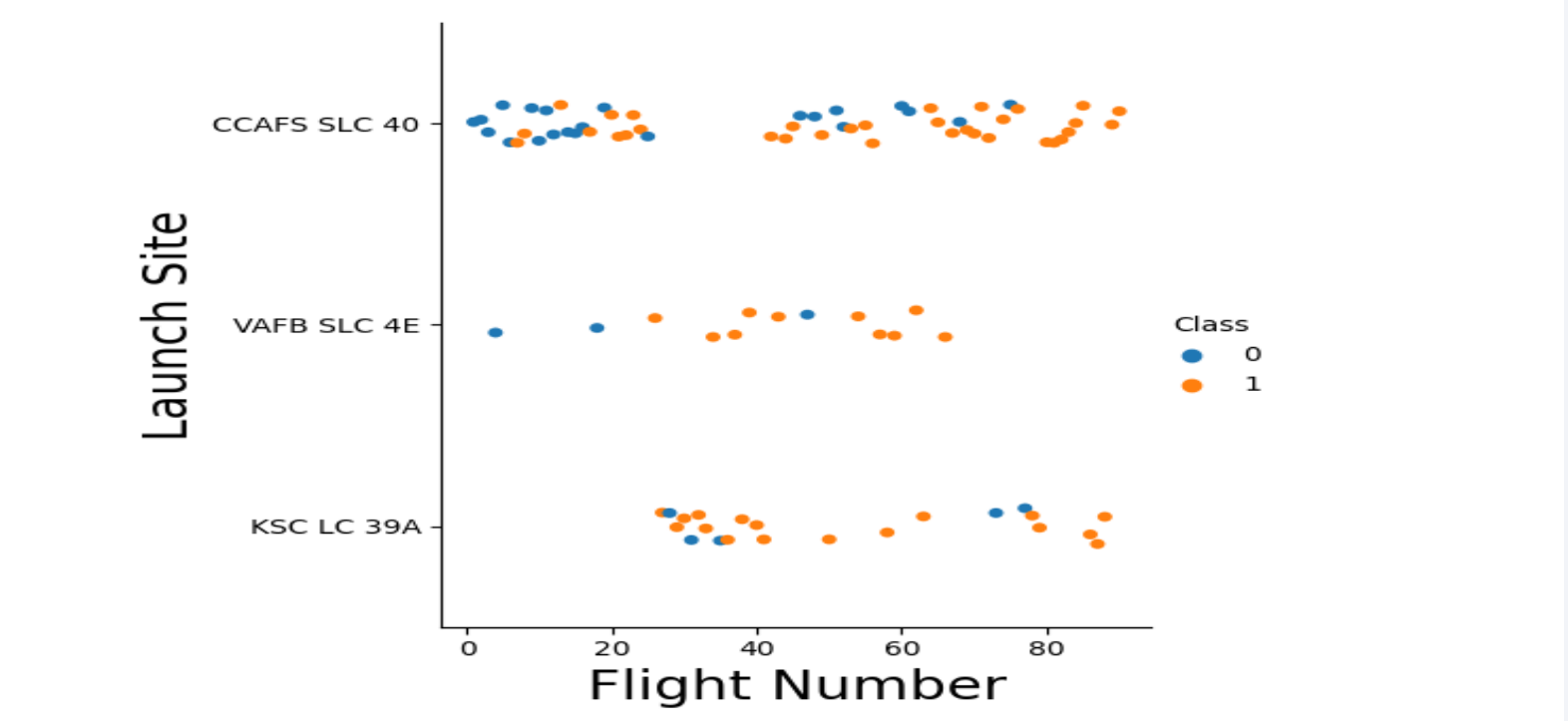
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

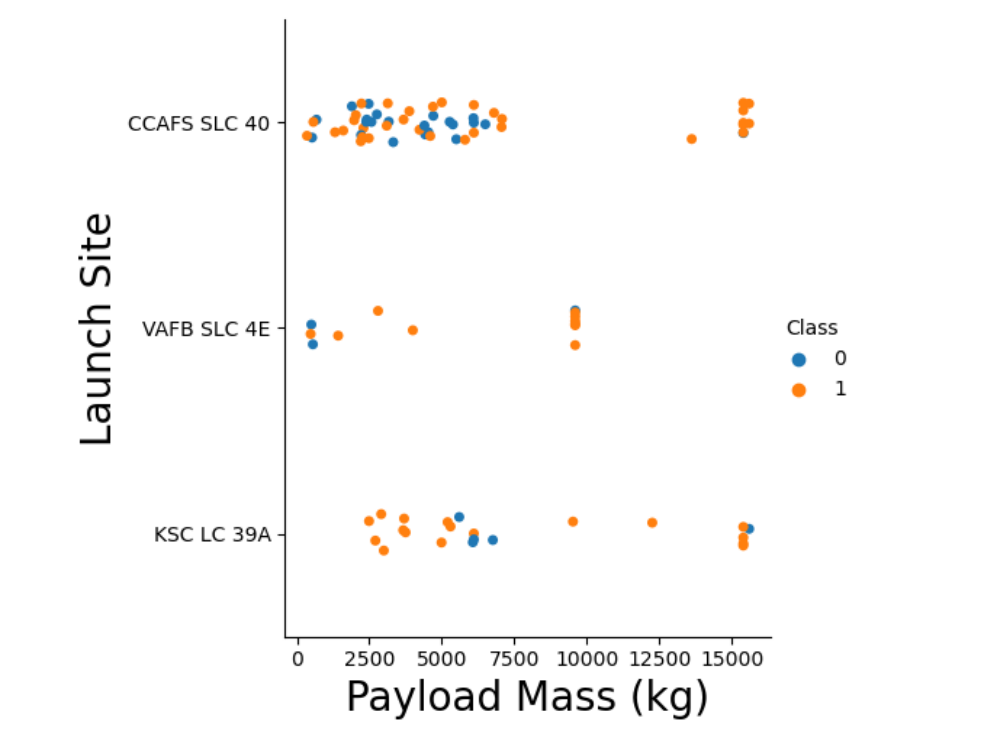


Flight Number vs. Launch Site



- There is no extreme correlation in Flight Number vs Launch Site

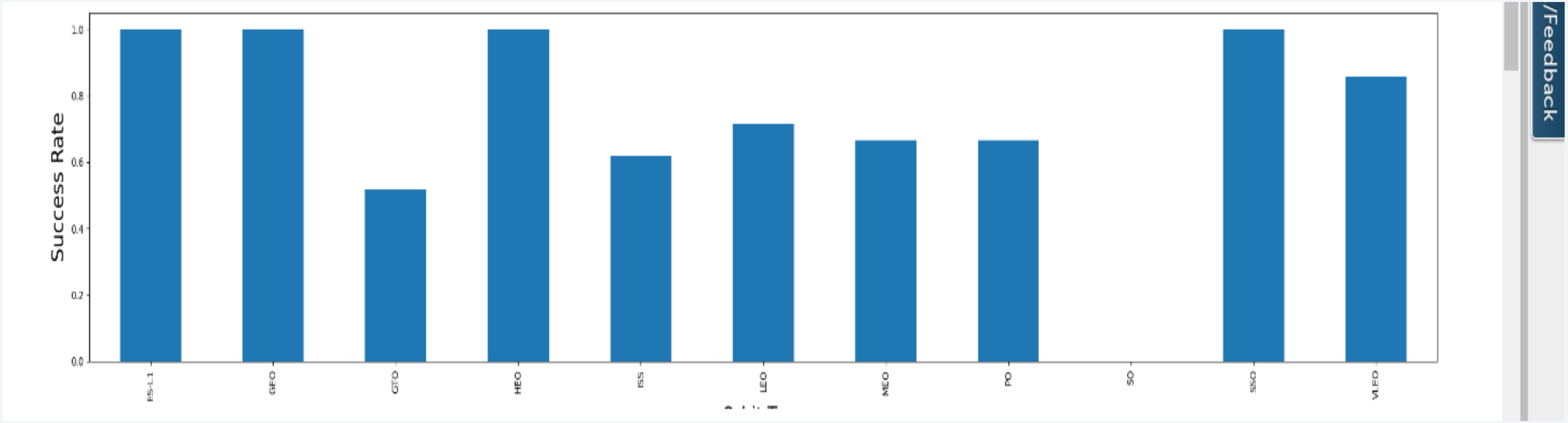
Payload vs. Launch Site



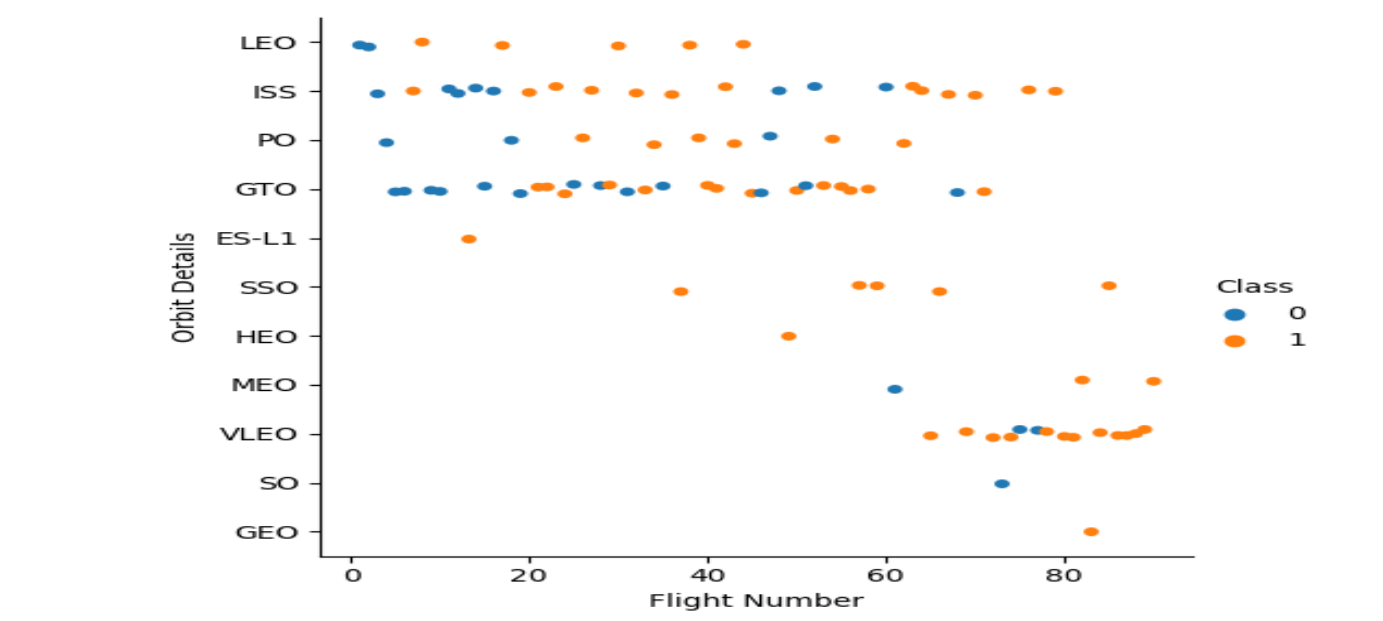
- VAFB launch site payload mass does not exceed 10000kg

Success Rate vs. Orbit Type

- Success Rate at certain Orbit types are higher but overall over medium



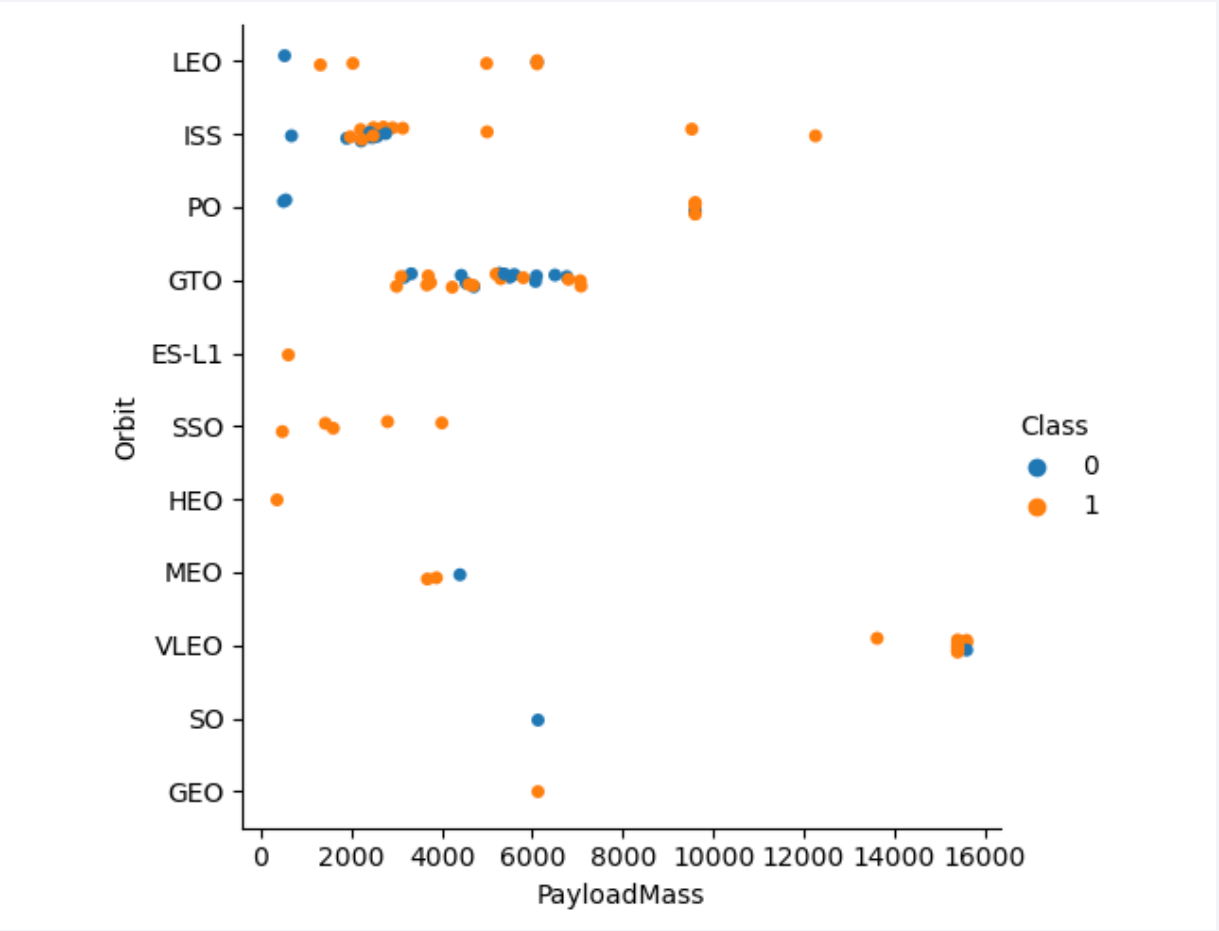
Flight Number vs. Orbit Type



- Higher Flight numbers at orbit of VLEO
- Steady pace for Orbit LEO, ISS, PO, and GTO

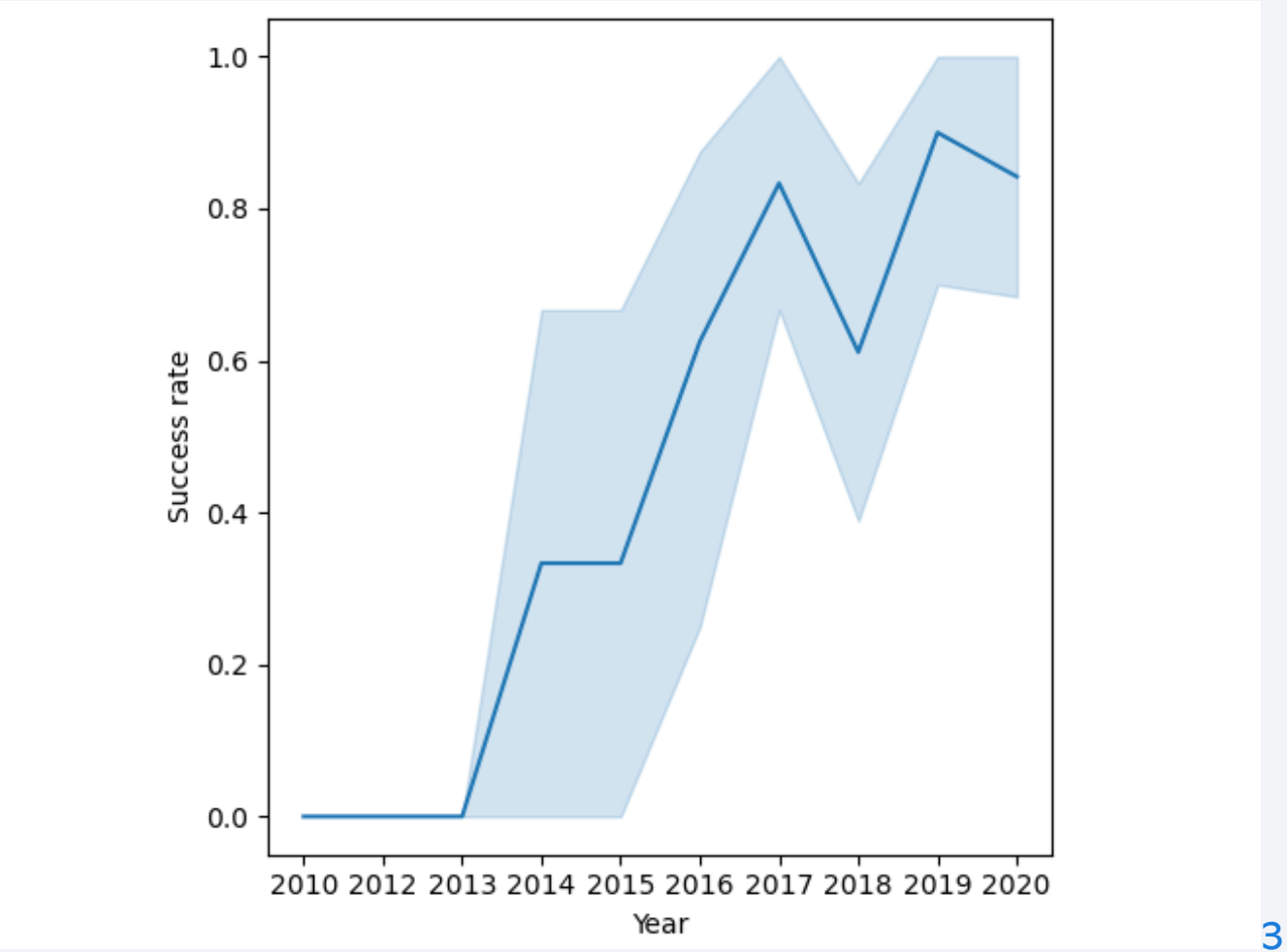
Payload vs. Orbit Type

- Lower Orbits have a low to mid Payload Mass.



Launch Success Yearly Trend

- Success Increase each year from 2013 with a small decrease in 2018.



All Launch Site Names

[7]:

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

Through a SELECT SQL query we find the names of each launch site.

Launch Site Names 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;
```

^ Our SQL query

Our results ->

Launch_Site
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40

Total Payload Mass

Display the total payload mass carried by boosters launched by NASA (CRS)

```
%sql SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE PAYLOAD  LIKE '%CRS%';
```

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

SUM(PAYLOAD_MASS__KG_)
111268.0

Average Payload Mass by F9 v1.1

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

```
%sql SELECT AVG(PAYLOAD_MASS__KG_) AS AVG_PAYLOAD FROM SPACEXTBL WHERE BOOSTER_VERSION = 'F9 v1.1';
```

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

AVG_PAYLOAD
2928.4

First Successful Ground Landing Date

```
%sql SELECT Date FROM SPACEXTBL WHERE Landing_Outcome = 'Success (ground pad)' limit 1;
```

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

Date
22/12/2015

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 SPACEXTBL WHERE PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000 AND Landing_Outcome = '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

```
%sql SELECT Mission_Outcome, COUNT(*) FROM SPACEXTBL GROUP BY Mission_Outcome ORDER BY Mission_Outcome;
```

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

Mission_Outcome	COUNT(*)
None	898
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 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

Would you like to receive official Jupyter

2015 Launch Records

```
%sql SELECT Booster_Version, Launch_Site FROM SPACEXTBL WHERE Landing_Outcome = 'Failure (drone ship)' AND Date like '%2015%';
```

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

Booster_Version	Launch_Site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

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

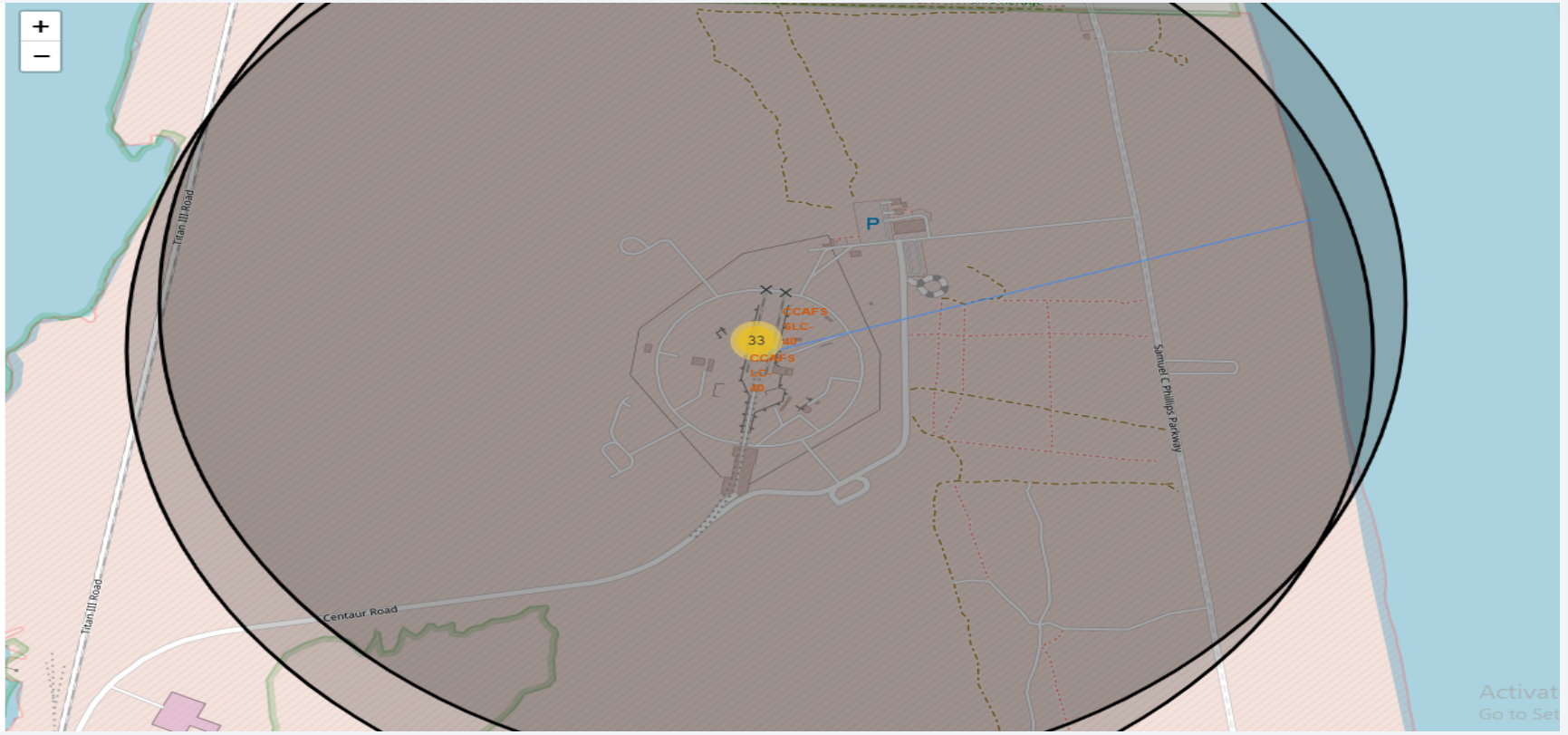
```
%sql SELECT Landing_Outcome, COUNT(Landing_Outcome) AS TOTAL FROM SPACEXTBL WHERE DATE BETWEEN '04/06/2010' AND '20/03/2017' GROUP BY Landing_Outcome ORDER BY TOTAL DESC
```

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

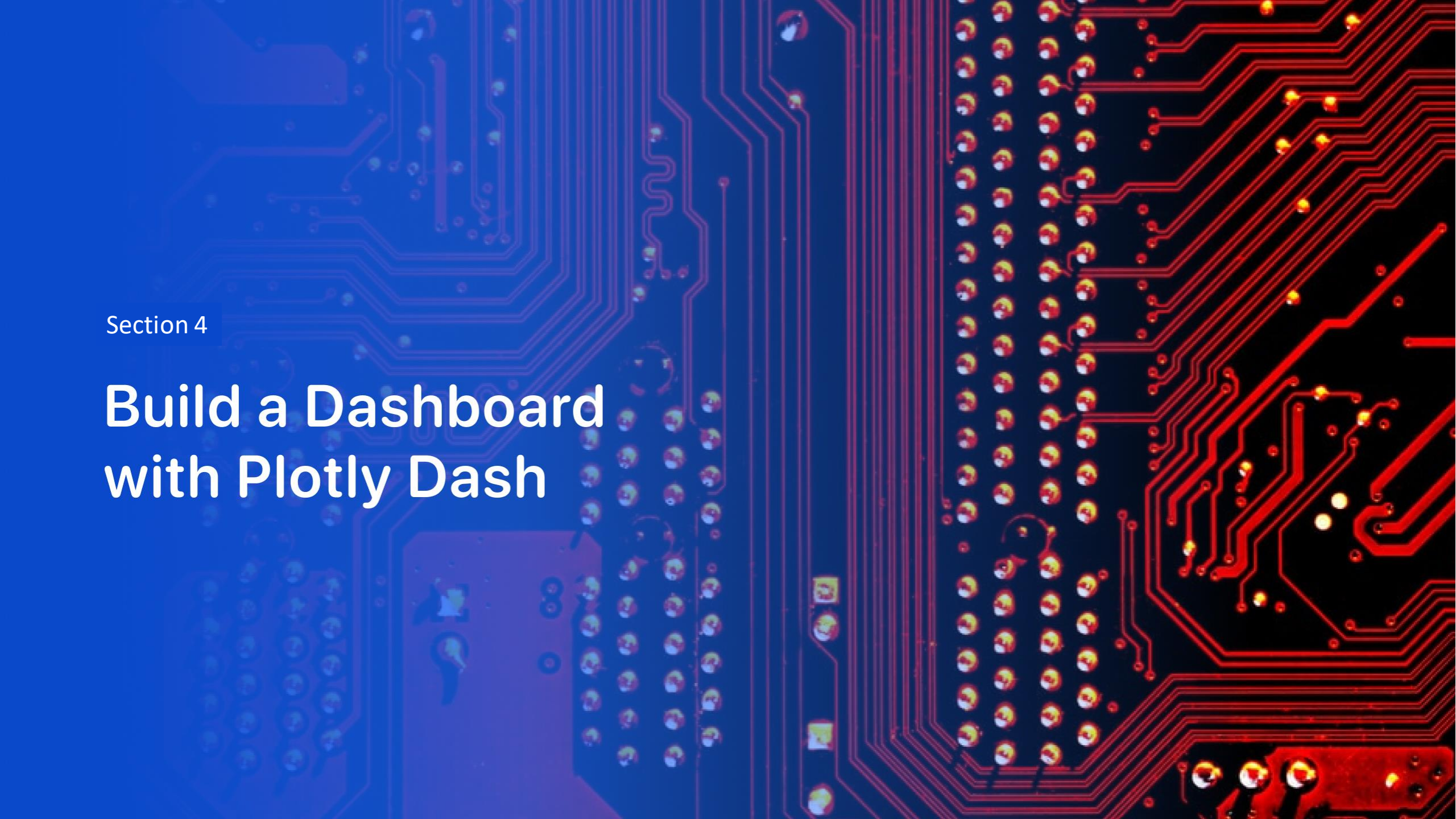
Landing_Outcome	TOTAL
Success	20
No attempt	9
Success (drone ship)	8
Success (ground pad)	7
Failure (drone ship)	3
Failure	3
Failure (parachute)	2
Controlled (ocean)	2
No attempt	1


Section 3
Launch Sites
Proximities Analysis

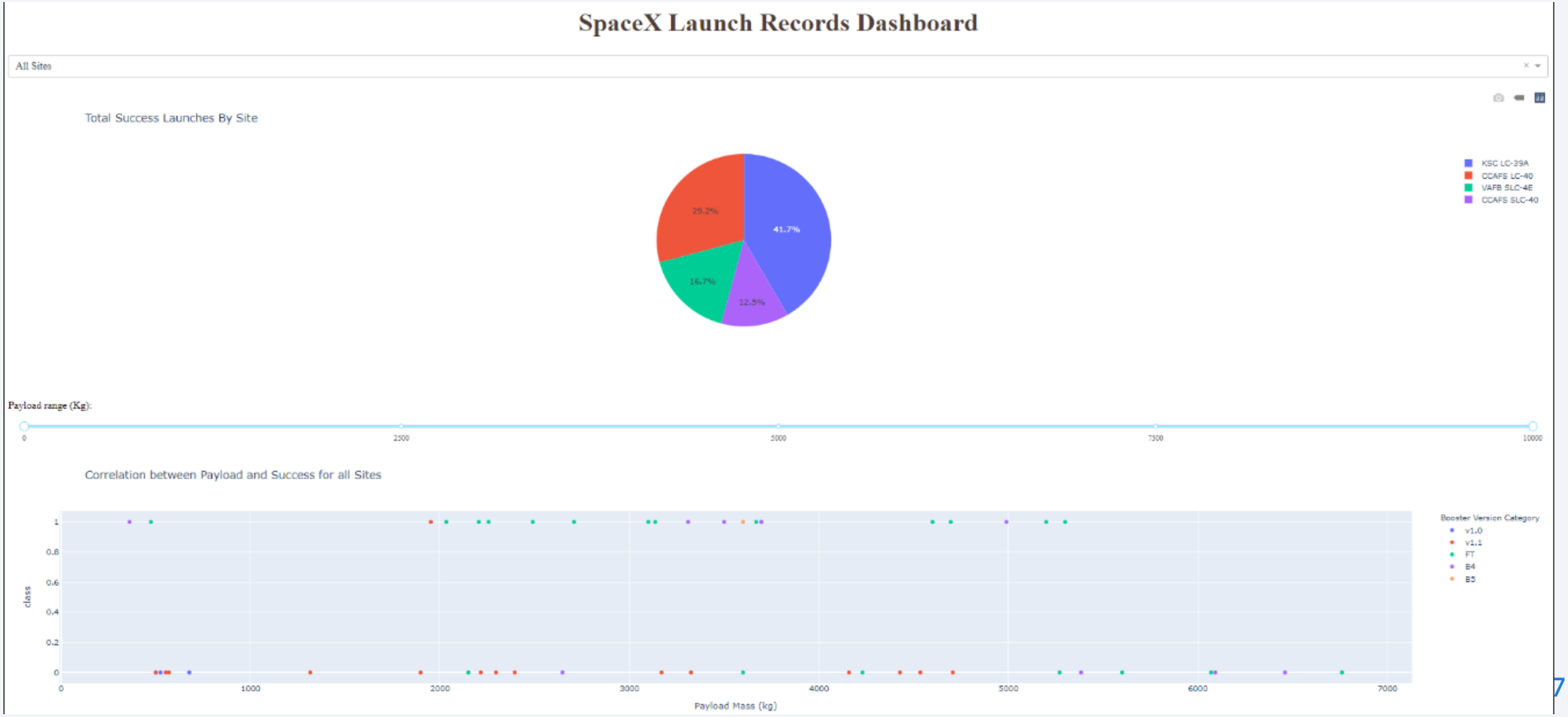
<Folium Map Screenshot 1>



Discovered the Proximity of Launch sites to other locations such as coastline.


Section 4
Build a Dashboard with Plotly Dash

Dashboard with Plotly Dash

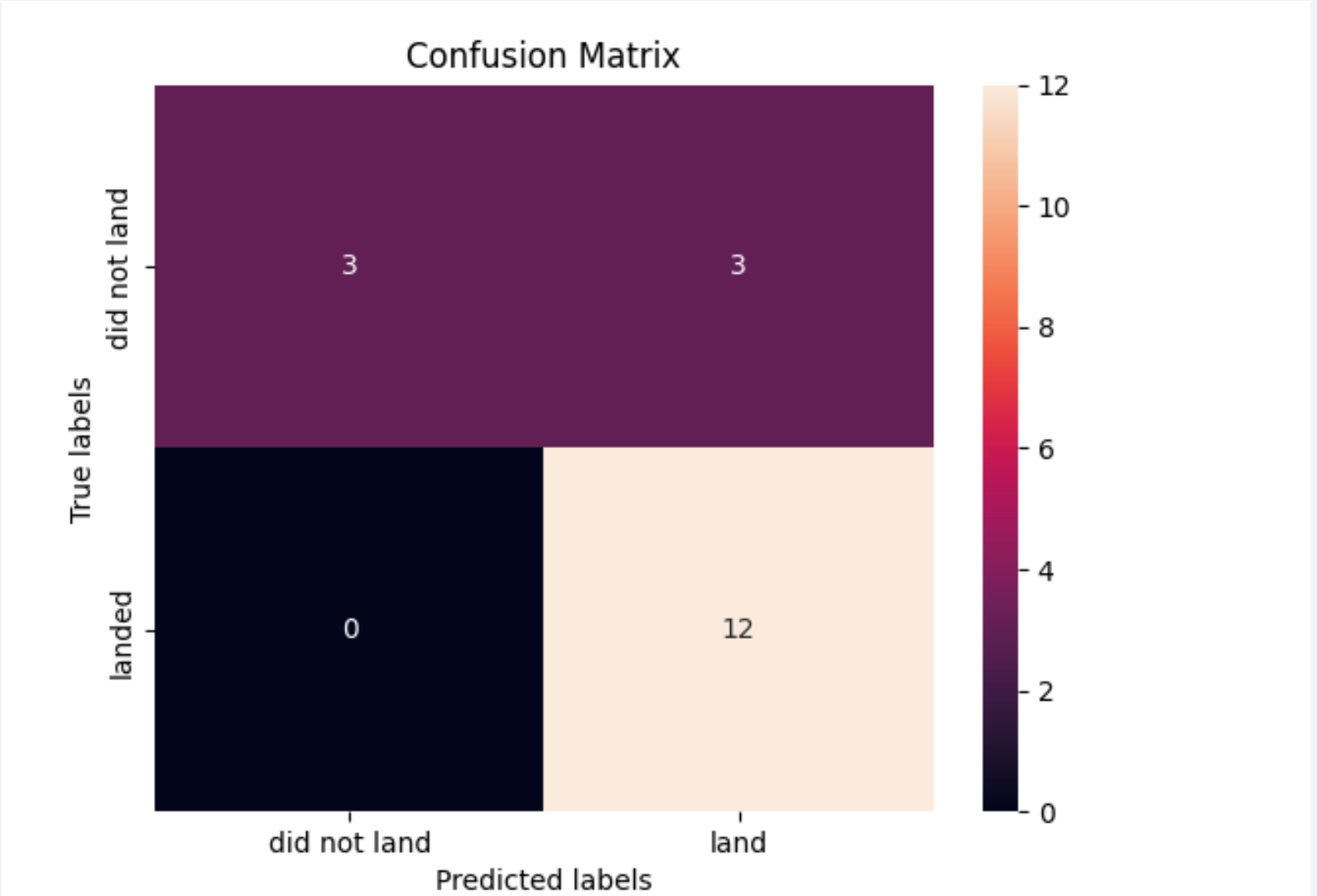


Section 5

Predictive Analysis (Classification)

Confusion Matrix

- Best performer



Conclusions

- SpaceX has a confirmed 83.3% succesful launch
- This in return saves money on reusability
- Their probability of success has increased steadily over the past 8 years and I think it will continue.

Thank you!