



IBM Developer  
SKILLS NETWORK

# Winning Space Race with Data Science

<Lai Zhen Yoong>  
<17/1/2024>



# Outline

---

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

# Executive Summary

---

- Summary of methodologies
- Summary of all results

# Introduction

---

- Commercial space travel is thriving, with companies like SpaceX making it more affordable.
- SpaceX stands out for its cost-effective Falcon 9 rocket launches, priced at \$62 million, a significant saving compared to other providers charging upwards of \$165 million.
- The key to SpaceX's cost efficiency lies in the reusable first stage of the Falcon 9.
- Our goal is to determine the launch price by predicting the first stage's reusability, thereby providing valuable insights for cost estimation.
- This presentation aims to gather pertinent information about SpaceX and create dashboards for our team to analyze and make informed decisions.





Section 1

# Methodology

# Methodology

---

## Executive Summary

- Data collection methodology:
  - Describe how data was collected
- Perform data wrangling
  - Describe how data was processed
- 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

---

- Data collection involved utilizing a RESTful API and web scraping methods.
- The RESTful API was employed to gather structured data, while web scraping helped extract additional relevant information.
- The collected data was converted into a data frame for efficient analysis.
- Subsequent data wrangling techniques were applied to enhance the quality and usability of the dataset.

# Data Collection – SpaceX API

---

- The SpaceX REST API endpoints, or URL, starts with `api.spacexdata.com/v4/`
- Conducting a GET request with the request library to retrieve and utilize launch data
- Our response will be in the form of a list of JSON objects
- The link to the notebook is <http://bit.ly/3HI85KD>

```
url="https://api.spacexdata.com/v4/launches/past"  
response =requests.get(url)  
response.json()
```



# Data Collection - Scraping

- Employ the Python BeautifulSoup package for web scraping HTML containing Falcon 9 launch records.
- Followed by parsing the table data and transforming it into a Pandas data frame.
- The link to the notebook is <http://bit.ly/3O3Ec5i>

2000 (cont.)

In late 2018, SpaceX stated that Falcon 9 had so many as 24 launches for Starlink satellites in 2018, in addition to 14 or 15 non-Starlink launches. In 2019, 13 of which for Starlink satellites, Falcon 9 had its most prolific year, and Falcon rockets were second-most prolific orbital launchers of 2019, only behind China's Long March rocket family.

Flight No.	Date and time (UTC)	Version, Booster	Launch site	Payload	Payload mass	Orbit	Customer	Launch outcome	Booster landing
18	7 January 2019, 00:12:27 <sup>[10]</sup>	F9 101.0, B1046.1	CCAFS, SLC-40	Starlink 2 v1.2 (28 satellites)	15,800 kg (34,830 lb) <sup>[11]</sup>	LEO	SpaceX	Success	Success (core stage)
First stage launch and several operational flights of Starlink constellation. One of the 40 satellites included a test coating to make the satellite less reflective, and thus less likely to interfere with ground-based astronomical observations. <sup>[10]</sup>									
19	18 January 2019, 15:30:04 <sup>[10]</sup>	F9 101.0, B1046.1	WSC, LC-39A	Crew Dragon in-flight abort test <sup>[10]</sup> (Dragon C100.1)	12,800 kg (28,210 lb)	Sub-orbital <sup>[10]</sup>	NASA, NASA <sup>[10]</sup>	Success	No attempt
An atmospheric test of the Dragon-2 abort system after Blue 3. 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 stated to be accomplished with the Crew Dragon Demo-1 capsule <sup>[10]</sup> but that test vehicle experienced a ground test of SuperDraco engines on 20 April 2018. <sup>[10]</sup> The abort test used the capsule originally intended for the first crewed flight. <sup>[10]</sup> As expected, the booster was destroyed by aerodynamic forces after the capsule aborted. <sup>[10]</sup> First flight of a Falcon 9 with only one functional stage – the second stage had a mass simulator in place of its engine.									
20	29 January 2019, 14:27:00 <sup>[10]</sup>	F9 101.0, B1051.1	CCAFS, SLC-40	Starlink 3 v1.2 (28 satellites)	15,800 kg (34,830 lb) <sup>[11]</sup>	LEO	SpaceX	Success	Success (core stage)
First operational and fourth stage launch of Starlink satellites, deployed in a circular 280 km (180 mi) orbit. One of the landing boosters was caught, while the other was refueled out of the ocean. <sup>[10]</sup>									
21	17 February 2019, 15:00:00 <sup>[10]</sup>	F9 101.0, B1055.1	CCAFS, SLC-40	Starlink 4 v1.2 (28 satellites)	15,800 kg (34,830 lb) <sup>[11]</sup>	LEO	SpaceX	Success	Failure (core stage)
Fourth operational and fifth stage launch of Starlink satellites, used a new flight profile which deployed six a 112 km x 300 km (110 mi x 240 mi) elliptical orbit instead of launching into a circular orbit and using the second stage engine burn. The first stage booster failed to land on the drone ship <sup>[10]</sup> due to incorrect wind data. <sup>[10]</sup> This was the first time a flight booster booster failed to land.									
22	1 March 2019, 04:00:00 <sup>[10]</sup>	F9 101.0, B1058.1	CCAFS, SLC-40	Starlink 5 v1.2 (28 satellites)	15,800 kg (34,830 lb) <sup>[11]</sup>	LEO	NASA, NASA	Success	Success (core stage)
Last launch of phase 1 of the CRS contract. Carries Bioneros, an ESA platform for testing external payloads into ISS. <sup>[10]</sup> Originally scheduled to launch on 2 March 2019, the launch date was pushed back due to a second stage engine failure. SpaceX decided to scrap the second stage instead of repairing the faulty part. <sup>[10]</sup> A new SpaceX's 30th anniversary was celebrated on 2 March 2019. The first flight of the Dragon C112 and the last launch of the range Dragon spacecraft.									
23	14 March 2019, 12:18:00 <sup>[10]</sup>	F9 101.0, B1046.1	WSC, LC-39A	Starlink 6 v1.2 (28 satellites)	15,800 kg (34,830 lb) <sup>[11]</sup>	LEO	SpaceX	Success	Failure (core stage)

Web scraping with BeautifulSoup

ion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	I
on 1	20.0	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	Merlin1A	1i
on 1	NaN	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	Merlin2A	1i
on 1	165.0	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	Merlin2C	1i
on 1	200.0	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	Merlin3C	1i
on 9	NaN	LEO	CCAFS SLC 40	None None	1	False	False	False	None	1.0	0	B0003	-2

# Data Wrangling

---

- We would like landing outcomes to be converted into classes  $y$  (either 1 or 0)
  - 0 is a bad outcome i.e., the booster did not land
  - 1 is a good outcome i.e., the booster did land
- Calculated the count of launches at each site and analyzed the frequency of each orbit type and its occurrences.
- The link to the notebook is <http://bit.ly/3O5XWW7>

# EDA with Data Visualization

---

- Some attributes can be used to determine if the first stage can be reused.
- We want to determine what attributes are correlated with successful landings.
- Generated various plots, such as scatter points, bar charts, and line graphs, to examine the impact of one variable on another, such as FlightNumber vs. PayloadMass or LaunchSuccess's Yearly Trend.
- The link to the notebook is <https://bit.ly/3SmgAf1>

# EDA with SQL

---

- Using the SQL queries collectively provide insights into various aspects of space missions such as:
  - The names of unique launch sites in the space mission
  - The total payload mass carried by booster launched by NASA (CRS)
  - The average payload mass carries by booster version F9 v1.1
  - The total number of successful and failure mission outcomes
  - The failed landing outcomes in drone ship, their booster version and launch site names
- The link to the notebook is <https://bit.ly/3RZAET6>

# Build an Interactive Map with Folium

---

- Utilized Folium to create an interactive map to analyze the Launch Site's Geo and Proximities empowering users to make informed decisions through interactive visual analytics.
- Added markers, circles, and lines for launch site locations and proximities.
- Provided answers by utilizing distances between a launch site and its surroundings:
  - Are launch sites near railways, highways and coastlines
  - Do launch sites keep certain distance away from cities
- The link to the notebook is <https://bit.ly/3SmkdAv>



# Build a Dashboard with Plotly Dash

---

- Constructed an interactive dashboard using Plotly Dash.
- Created pie charts to visualize total launches for specific launch sites.
- Generated scatter graphs to illustrate the relationship between outcome and payload mass for various boosters.
- The link to the notebook is <https://bit.ly/3SOaAag>

# Predictive Analysis (Classification)

---

- Loaded data using NumPy and Pandas, transformed it, and then partitioned it into training and testing sets.
- Developed various machine learning models, fine-tuning hyperparameters through GridSearchCV.
- Employed accuracy metrics for model evaluation, identifying the best-performing model. Enhanced model performance through feature engineering and algorithm tuning.
- The link to the notebook is <https://bit.ly/41YzSdz>

# Results

---

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



The background of the slide is an abstract composition. It features a dark blue base color. Overlaid on this are numerous diagonal streaks in shades of red and cyan. A faint, light blue grid pattern is also visible, particularly in the lower-left quadrant. The overall effect is dynamic and technological.

Section 2

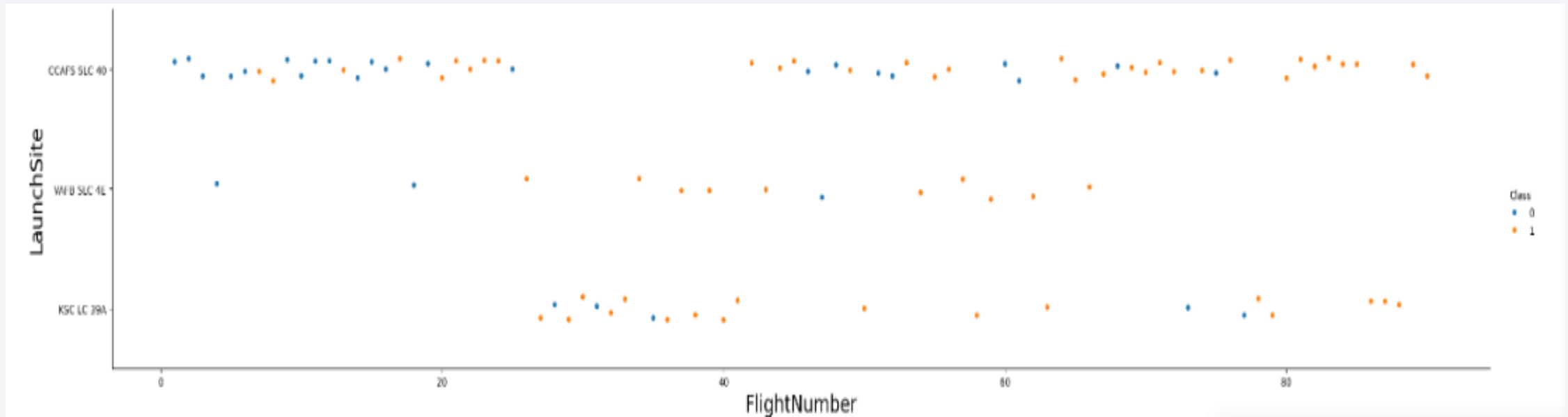
# Insights drawn from EDA



# Flight Number vs. Launch Site



The greater the flight amount at a launch site, the greater the success rate.

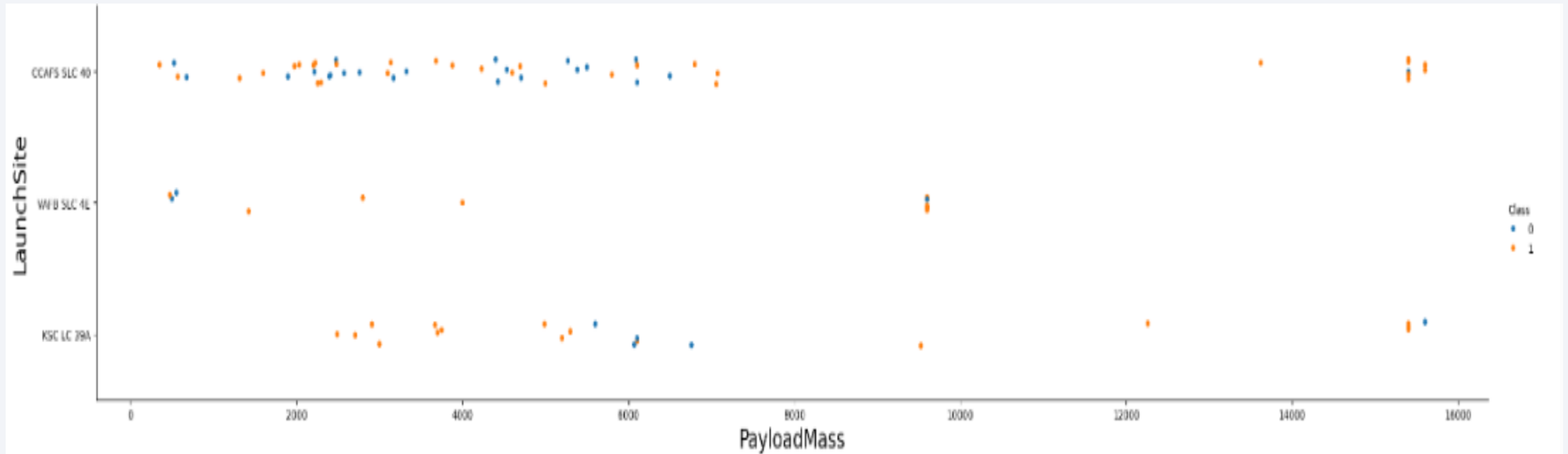




# Payload vs. Launch Site



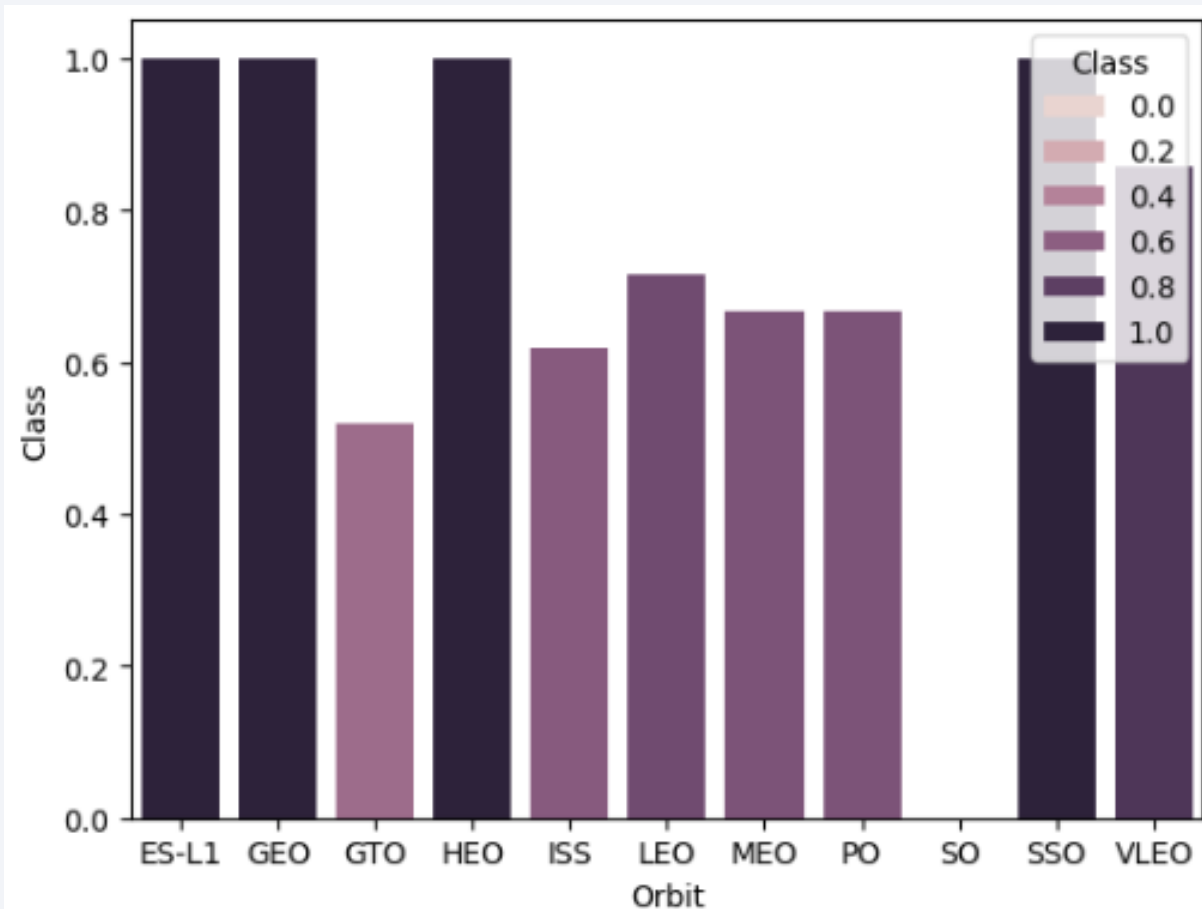
The greater the payload mass,  
the higher the success rate.



# Success Rate vs. Orbit Type



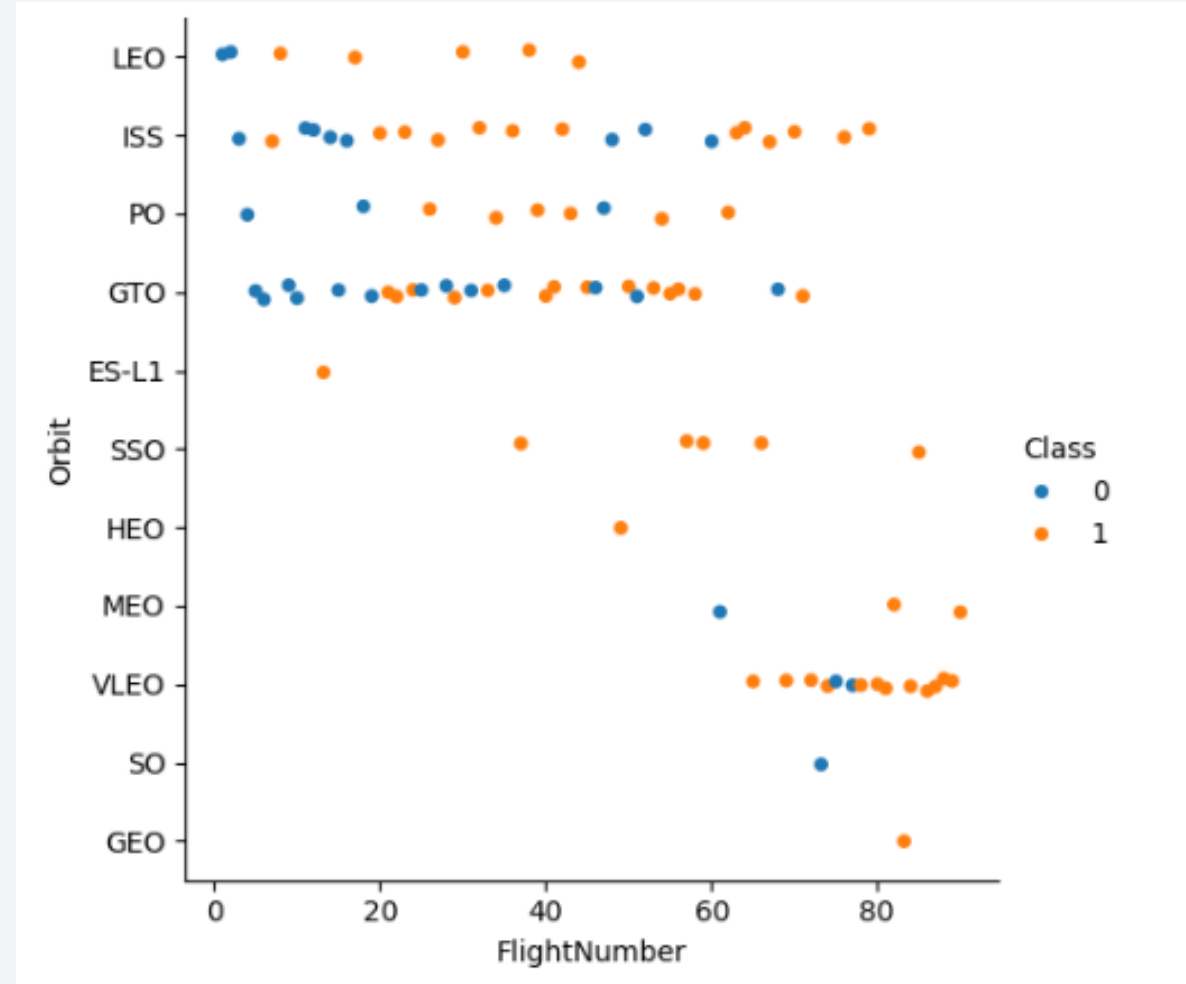
ES-L1, GEO, HEO & SSO achieved the highest success rate.



# Flight Number vs. Orbit Type



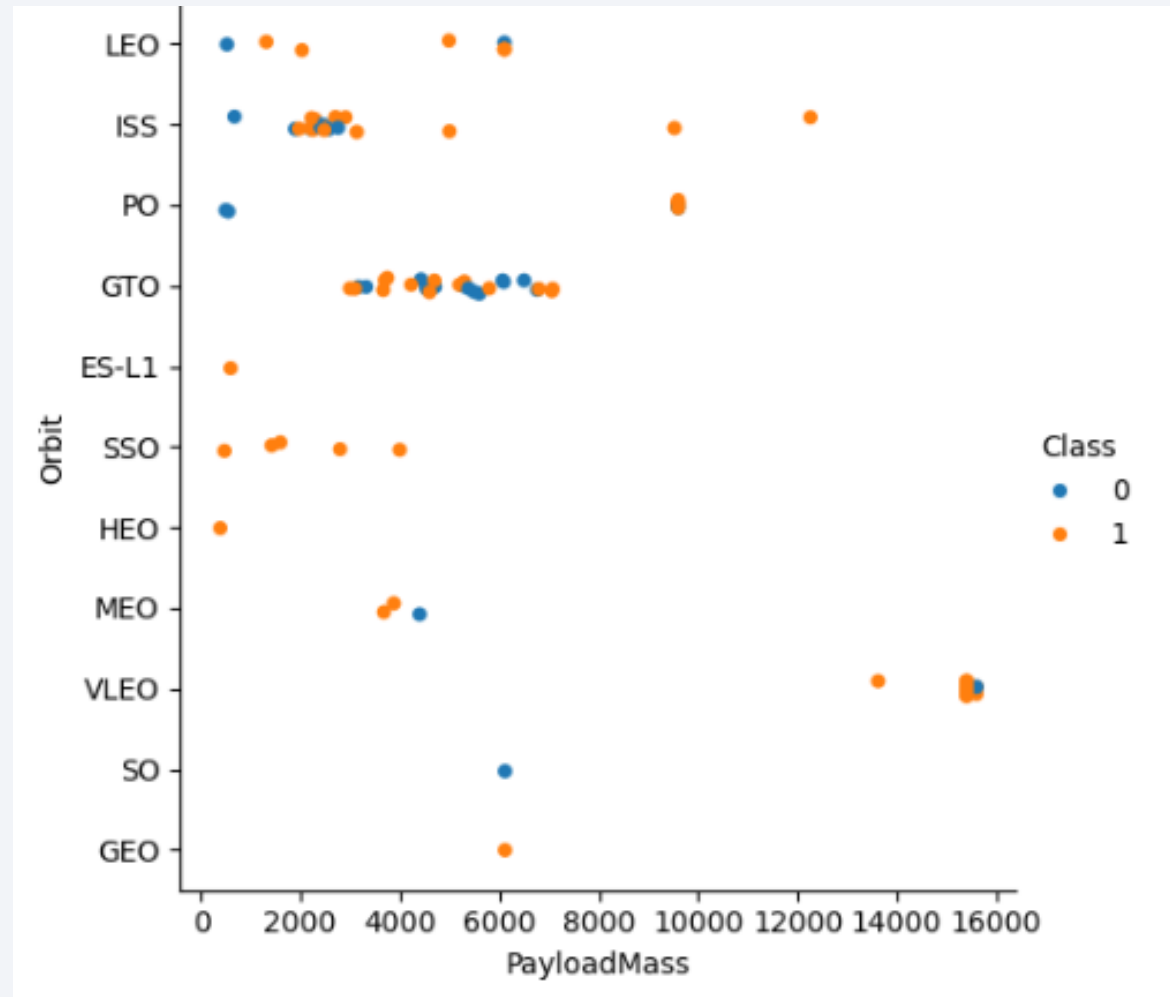
No relationship is found between Flight Number vs Orbit Types.



# Payload vs. Orbit Type



Most payloads are assigned for VLEO, PO and ISS with successful results.

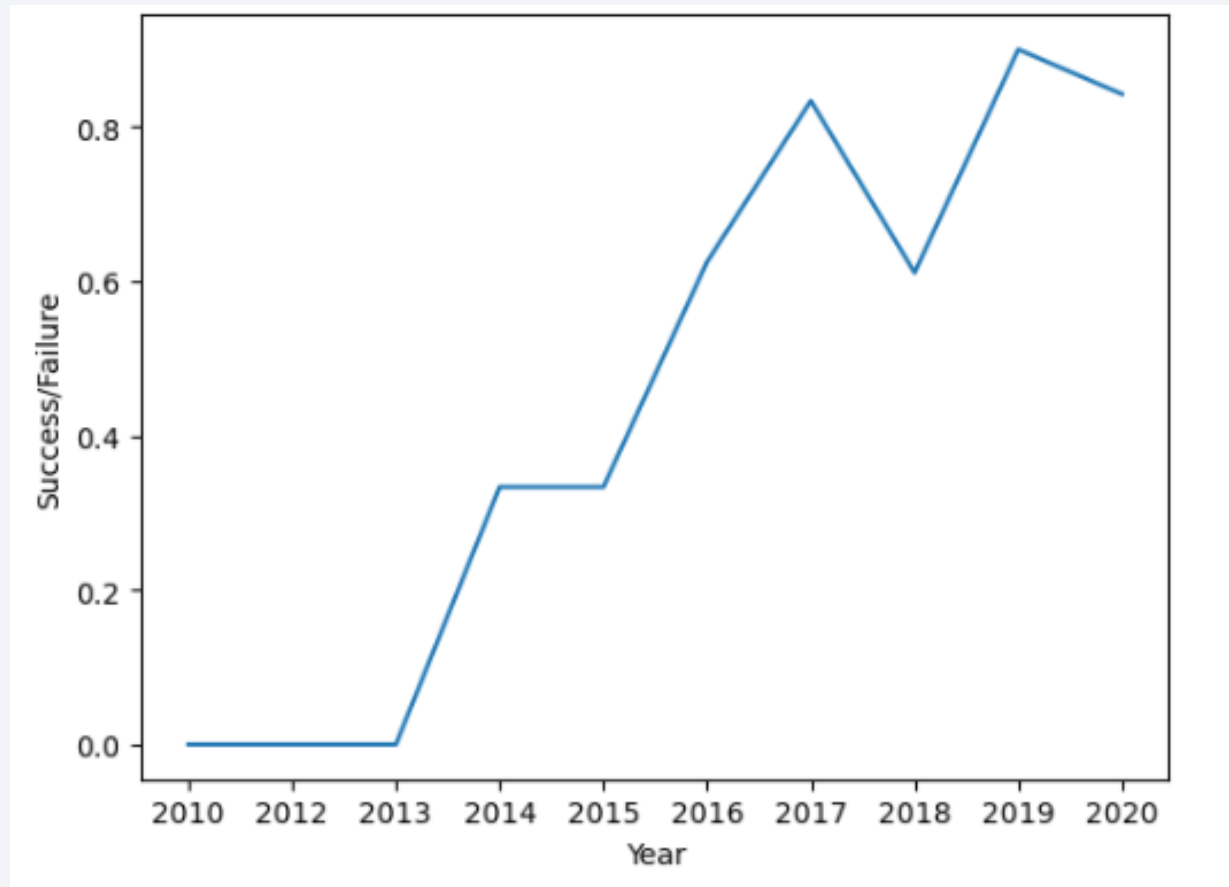


# Launch Success Yearly Trend

---



Launch site success rate increases every year since 2013.





# All Launch Site Names

---

We used the key word  
DISTINCT to show only  
unique launch sites from the  
SpaceX data

```
%sql SELECT DISTINCT "Launch_Site" FROM SPACEXTABLE;
```

```
* 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'

We used the keywords LIKE and LIMIT to filter launch site's beginning with 'CCA' and restricts the number of rows returned

```
%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

# Total Payload Mass

---

Calculated the total payload mass carried by NASA's booster which resulted in 45596 kg in total.

```
%sql SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXTABLE WHERE "Customer" LIKE "NASA (CRS)";
```

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

```
Done.
```

SUM(PAYLOAD_MASS__KG_)
------------------------

45596
-------

# Average Payload Mass by F9 v1.1

---

Calculated the average payload mass carried by booster version F9 v1.1 which resulted in 2928.4 kg in total.

```
%sql SELECT AVG(PAYLOAD_MASS_KG_) FROM SPACEXTABLE WHERE "Booster_Version" LIKE "F9 v1.1";
```

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

```
Done.
```

```
AVG(PAYLOAD_MASS_KG_)
```

---

```
2928.4
```

# First Successful Ground Landing Date

---

The first successful landing outcome on ground pad took place on 22<sup>nd</sup> December 2015.

```
%sql SELECT MIN("Date") FROM SPACEXTABLE WHERE "Landing_Outcome" LIKE "Success (ground pad)";
```

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

```
Done.
```

```
MIN("Date")
```

---

```
2015-12-22
```



# Successful Drone Ship Landing with Payload between 4000 and 6000

---

Using the WHERE clause to filter successful drone ship landing AND payload mass greater than 4000 and less than 6000.

```
%sql SELECT "Booster_Version" FROM SPACEXTABLE WHERE "Landing_Outcome" LIKE "Success (drone ship)" AND "PAYLOAD_MASS_KG_" > 4000 AND "PAYLOAD_MASS_KG_" < 6000
```

```
* 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

---

A GROUP BY clause was used to count mission outcome results.

```
%sql SELECT MISSION_OUTCOME, COUNT(*) 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

A subquery in the WHERE clause selects the maximum payload mass using the MAX function.

```
%%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

A substr() method was used to list out the details for failed launches on drone ship for the year 2015,

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

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

Month	Date	Booster_Version	Launch_Site	Landing_Outcome
01	2015-01-10	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
04	2015-04-14	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

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

We selected Landing outcomes and the COUNT of landing outcomes from the data using a WHERE clause to filter outcomes BETWEEN 2010-06-04 and 2010-03-20 with a GROUPBY clause to group each landing outcome.

```
%%sql
SELECT "Landing_Outcome", COUNT("Landing_Outcome"), "Date"
FROM SPACEXTABLE
WHERE "Date" BETWEEN '2010-06-04' AND '2017-03-20'
GROUP BY "Landing_Outcome"
ORDER BY COUNT("Landing_Outcome") DESC
```

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

Landing_Outcome	COUNT("Landing_Outcome")	Date
No attempt	10	2012-05-22
Success (drone ship)	5	2016-04-08
Failure (drone ship)	5	2015-01-10
Success (ground pad)	3	2015-12-22
Controlled (ocean)	3	2014-04-18
Uncontrolled (ocean)	2	2013-09-29
Failure (parachute)	2	2010-06-04
Precluded (drone ship)	1	2015-06-28

A satellite view of Earth from space, showing the curvature of the planet and city lights at night. The background is a deep blue gradient.

Section 3

# Launch Sites Proximities Analysis

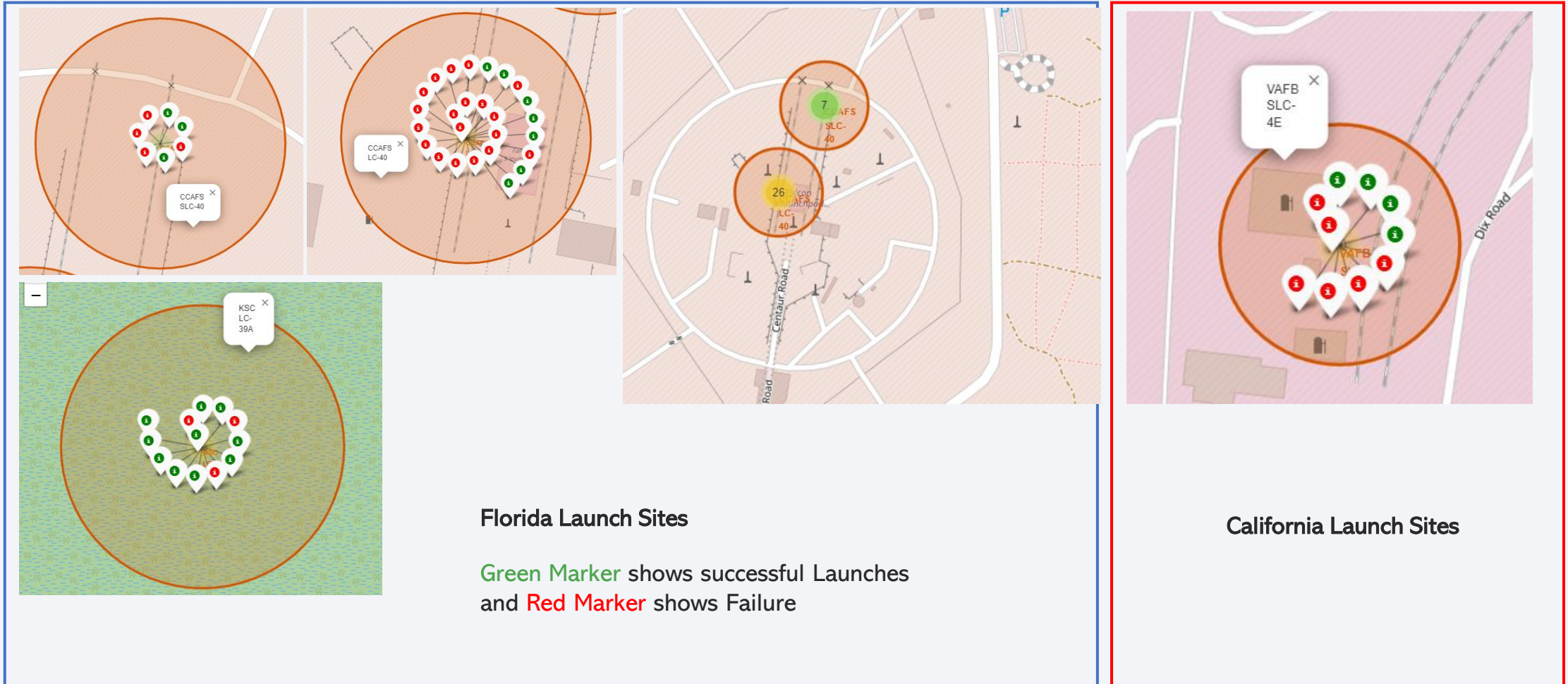
# All Launch Site's Global Location

---



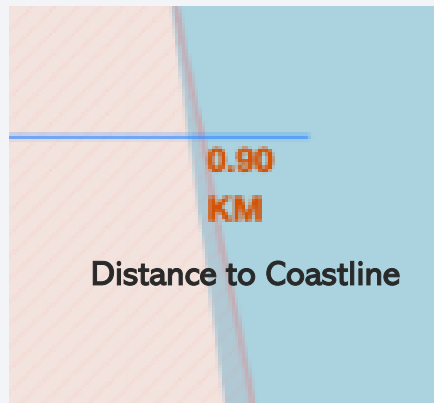
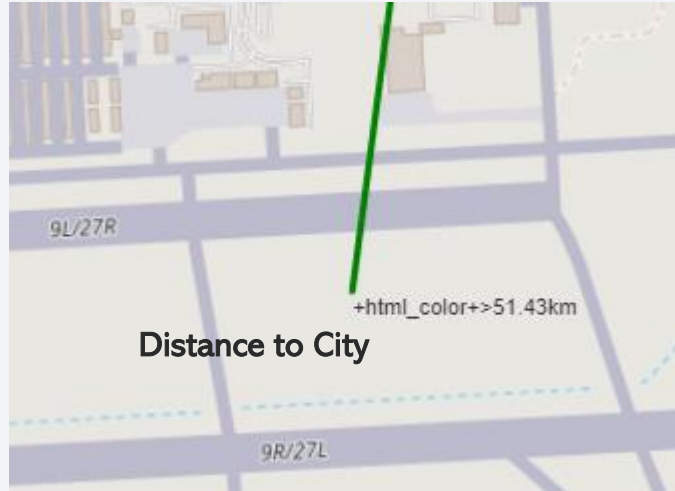


# Markers Showing Launch Site Success/Failure





# Launch Site Distance to Landmark



- \* Are launch sites in close proximity to railways? Yes
- \* Are launch sites in close proximity to highways? Yes
- \* Are launch sites in close proximity to coastline? Yes
- \* Do launch sites keep certain distance away from cities? No



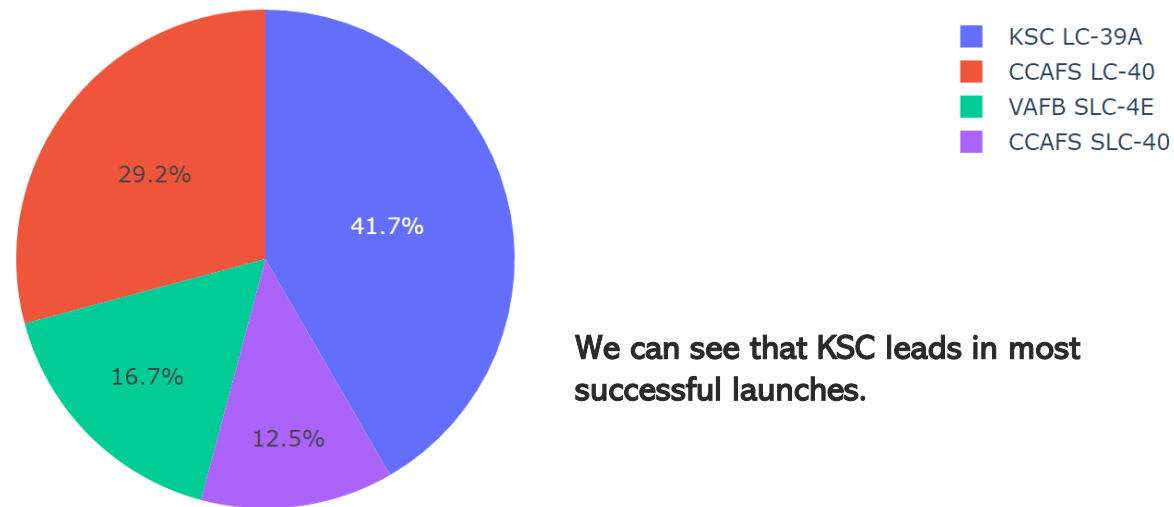
Section 4

# Build a Dashboard with Plotly Dash

# Launch Site's Success using Pie Charts

---

Success Count For All Launch Sites

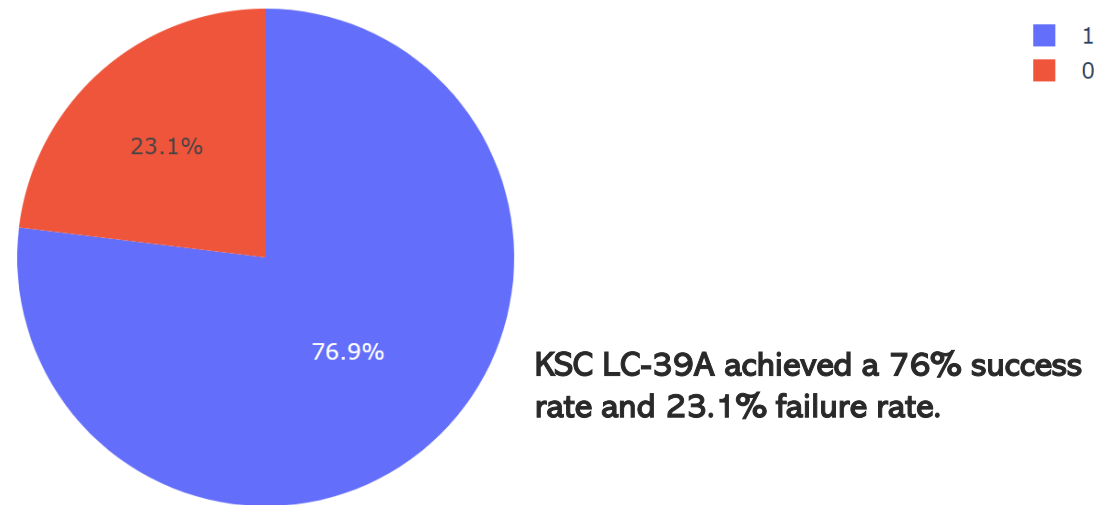


We can see that KSC leads in most successful launches.

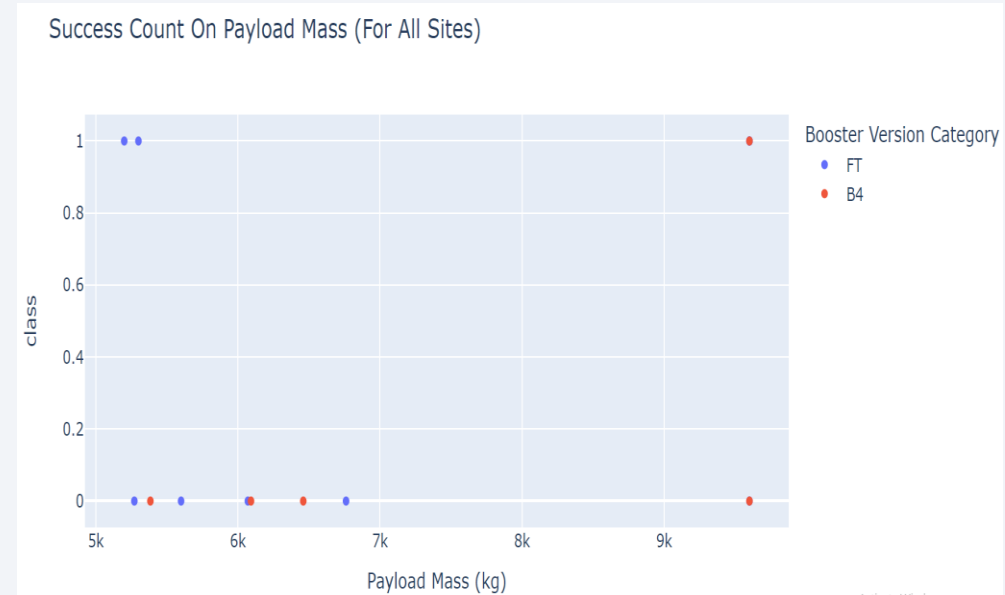
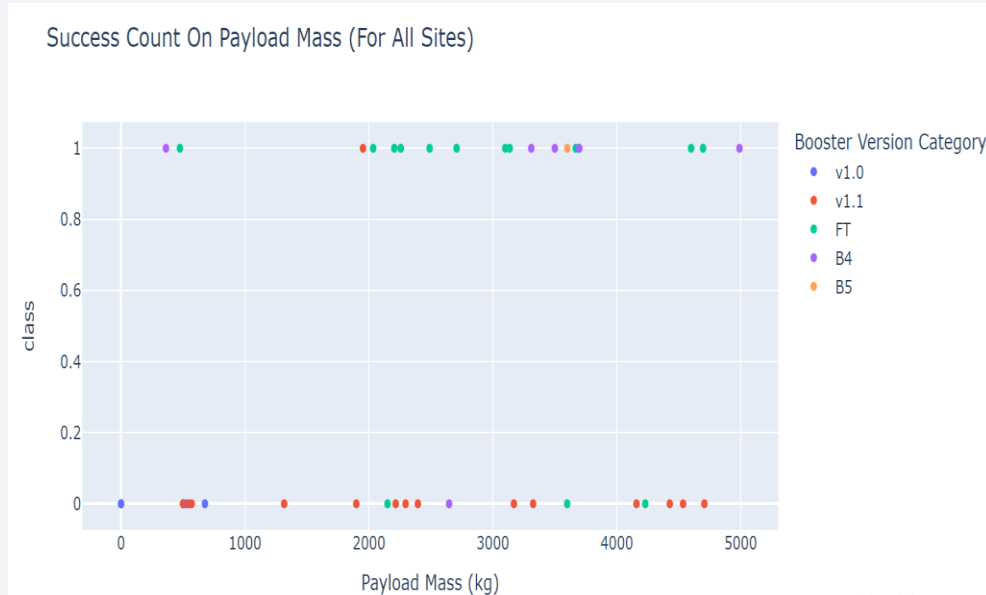
# Highest Launch Success Ratio using Pie Chart

---

Total Success Launches For Site KSC LC-39A



# Payload vs Launch Outcomes using Scatter Plot



**We can see the success rate for lower weighted payloads is higher than the heavier weighted payloads.**

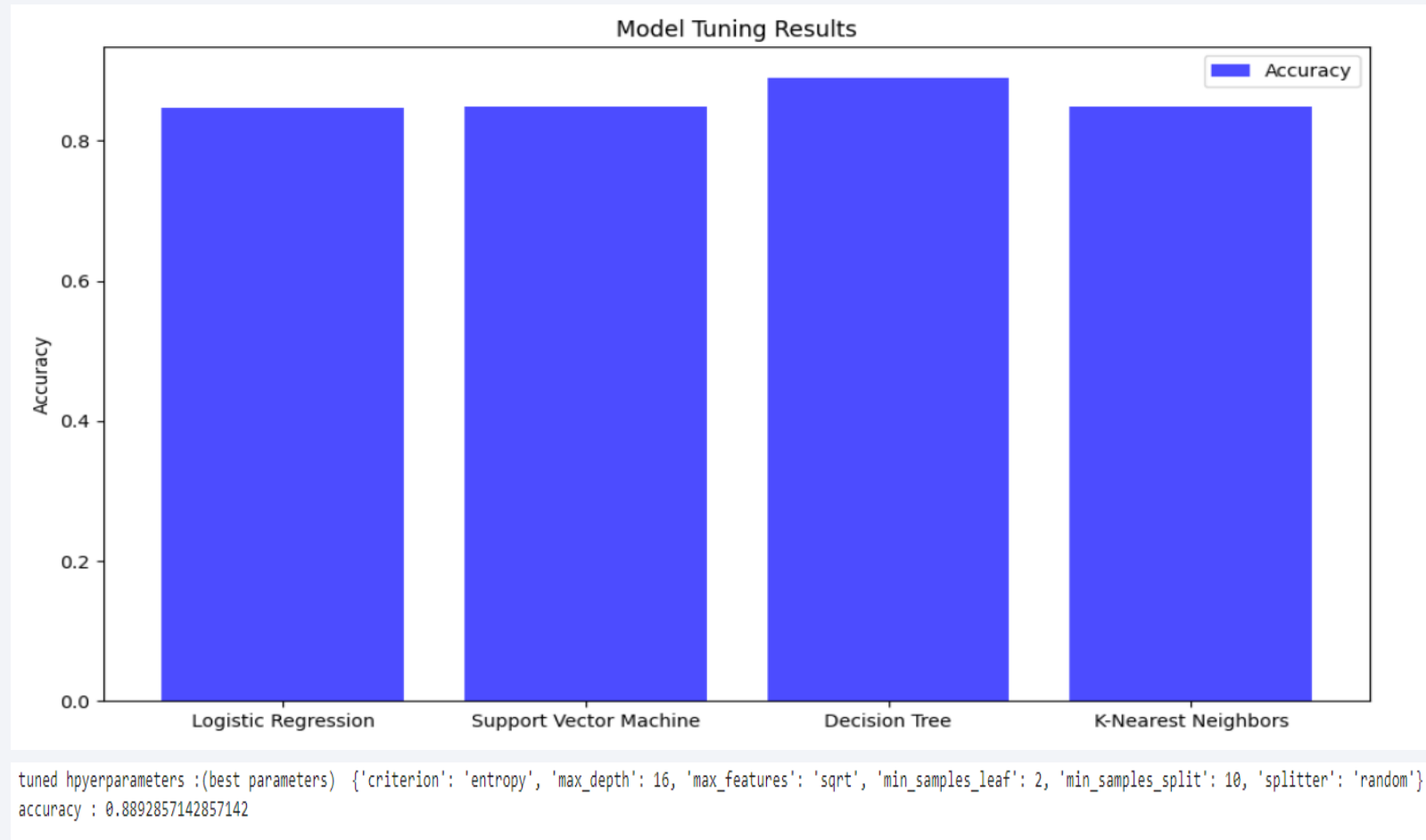


Section 5

# Predictive Analysis (Classification)

# Classification Accuracy

---

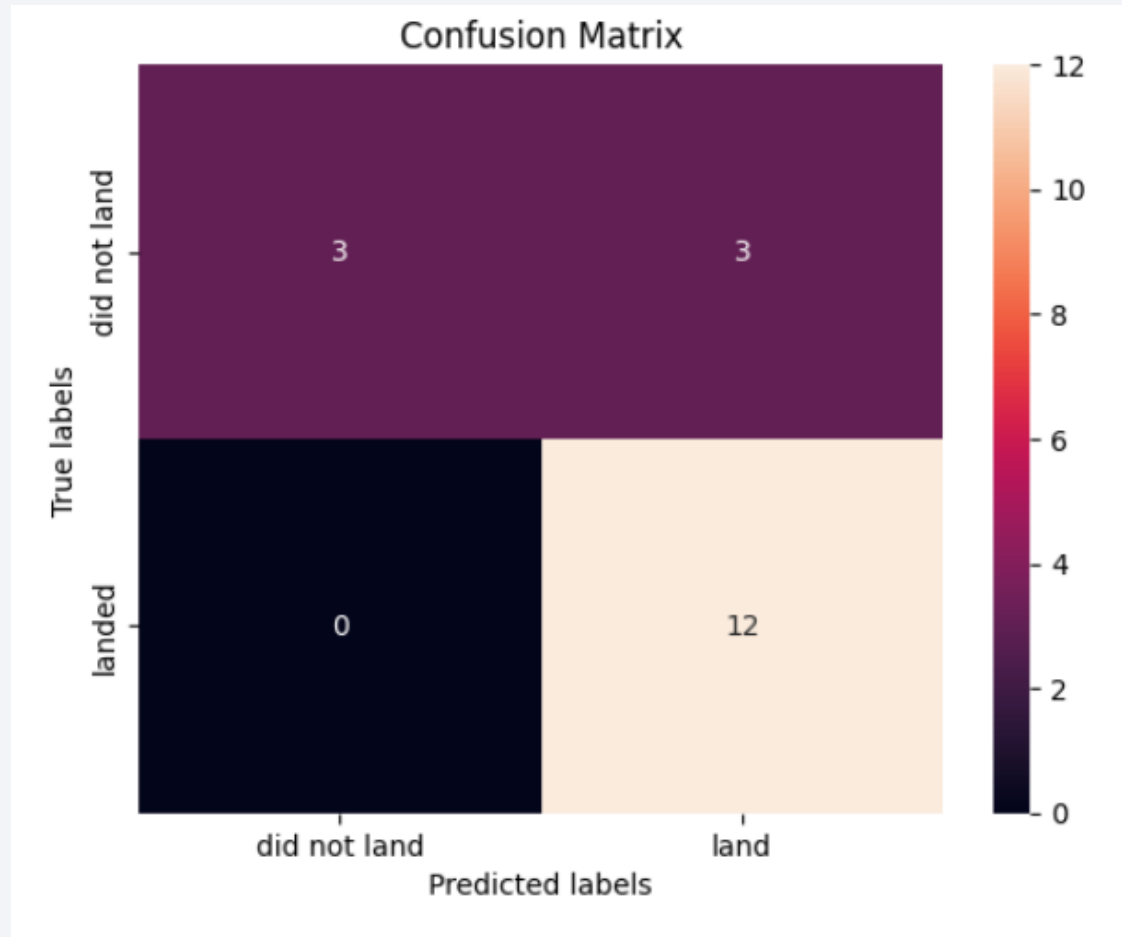


The Decision Tree model resulted to the highest accuracy amongst the 4.

# Confusion Matrix

---

Decision Tree Classifier Confusion Matrix reveals effective class distinction and highlights a False Positives Issue.





# Conclusion

- The greater the flight amount at a launch site, the greater the success rate.
- The greater the payload mass, the higher the success rate.
- ES-L1, GEO, HEO & SSO achieved the highest success rate.
- Launch site success rate increases every year since 2013.
- The Decision Tree model resulted to the highest accuracy





Thank you!

