SpaceX, will it land?

Joohwan Lee

7/27/2024

OUTLINE

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

EXECUTIVE SUMMARY

- Methodologies
  - Data collection using SpaceX API
  - Data cleaning using Python
  - Exploratory data analysis using SQL and Python
  - Preliminary results using data visualization
  - Predictive analysis using Machine learning algorithms
- Summary
  - Results of Exploratory Data Analysis
  - Inference from data visualization
  - Determining the best classification model

INTRODUCTION

- SpaceX is an American spacecraft manufacturer whose mission is "Making Humanity Multiplanetary"
- Falcon 9 is a reusable, two stage rocket that SpaceX designed and manufactured for the reliable and safe transportation of people and payloads into Earth orbit and beyond
- For the Falcon 9 rocket to be reused (first-stage), the rockets must make a safe landing.
- A safe landing (successful outcome) results in rockets that can be reused and lower the cost of launches.
- Question SpaceX has to answer
  - What factors lead to successful landings?
- In this presentation we will try to answer this question

Methodology

Data Collection

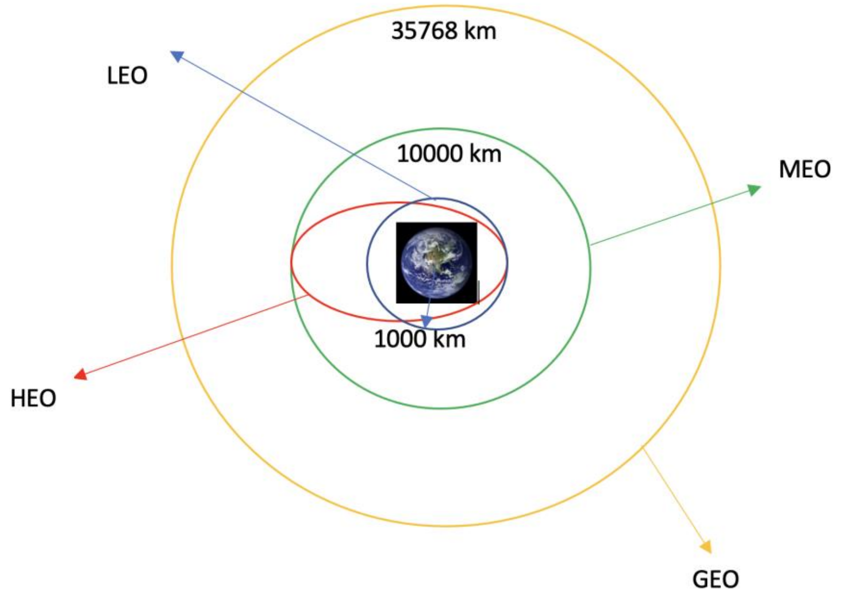
- Using SpaceX API to collect data.
- .Json() function to decode the content and turn into a pandas dataframe
- Cleaned the data looking for nulls (missing values)
- Nulls found in the PayloadMass attribute was replaced with the mean PayloadMass
- Filtered out Falcon 1 (older) launches.
- Here is the link to the notebook1

Data Collection-Webscraping

- Using webscraping technique (BeautifulSoup) to collect launch records from Wikipedia
- Parsed the table and converted into a pandas dataframe
- Here is the link to the notebook2

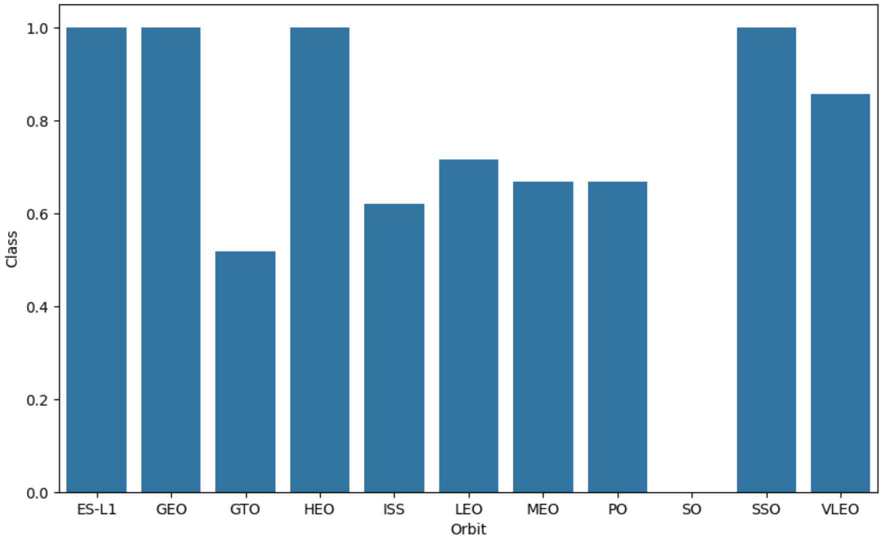
Exploratory Data Analysis

- Performed exploratory data analysis (EDA)
- Identified training labels ‘Orbit’ and ‘LaunchSite’
- Classification of outcome label (1=success, 0=failure)
- Here is the link to the notebook3

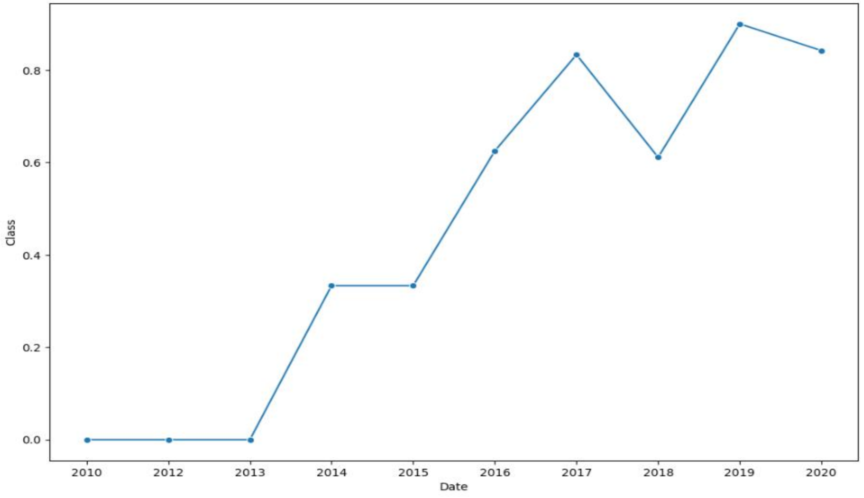


EDA with Data Visualization

- Here is a bar graph between proportion of successful landing (y-axis) and different 'Orbits' (x-axis)
- Different 'Orbits' can determine success/failure landing



- Here is a line graph that shows the proportion of successful landing (y-axis) over time (x-axis)
- The trend is the proportions of successful landing are increasing over time
- Here is the link to the notebook4



EDA with SQL(Magic)

- Loaded the SpaceX dataset into a PostgreSQL database
- Applied EDA with SQL to get insights from the data. Wrote queries to:
  - The names of unique launch sites
  - The total payload mass carried by boosters
  - The average payload mass carried by boosters
  - The total number of successful and failures landings
- The link to the notebook5

Interactive Map with Folium

- Marked all launch sites. Added map objects such as markers to indicate success or failure for each map site
- Assigned the attribute launch outcomes (1 = success, 0 = failure).
- Identified launch site success rates using color-labels
- Calculate distances from launch sites to certain important map features
- Here is a link to the Folium Map

Dashboard with Plotly Dash

- Interactive dashboard with Plotly dash
- Pie chart showing total launches relative to launch sites
- Scatterplot showing relationships between Outcome and PayloadMass(Kg)
- Here is the link to the notebook6

Predictive Analysis

- Use one-hot coding to transform categorical attributes into attributes for use in multi-variable logistic regression
- Split the data into training and testing data.
- Applied different machine learning classification models tuned to different hyperparameters using GridSearchCV
- Applied different accuracy metric to determine best performing classification model
- Here is the link to the notebook7

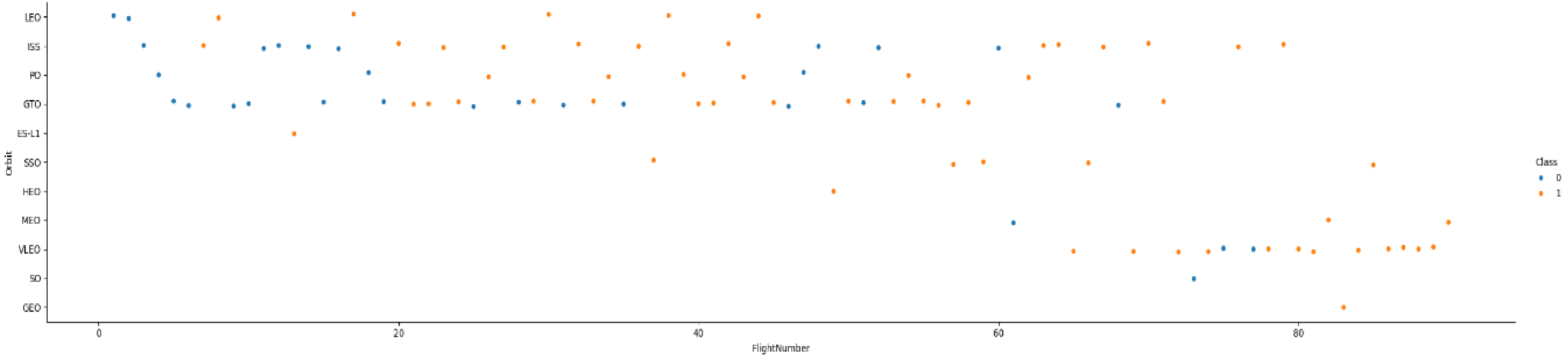
Results

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

Insights from EDA

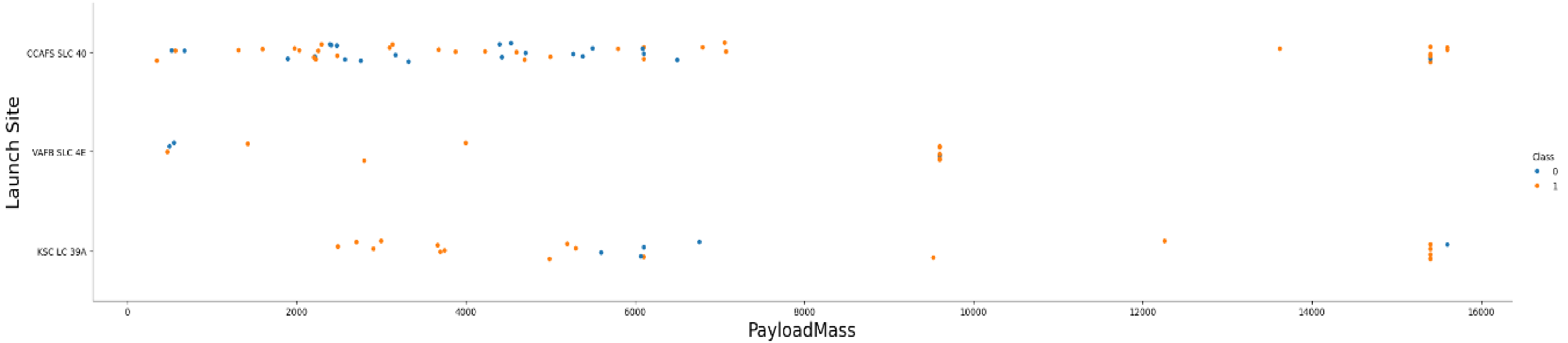
Flight Number vs Launch Site

- From the plot, we found that the larger the flight number, the greater the success rate at a launch sites



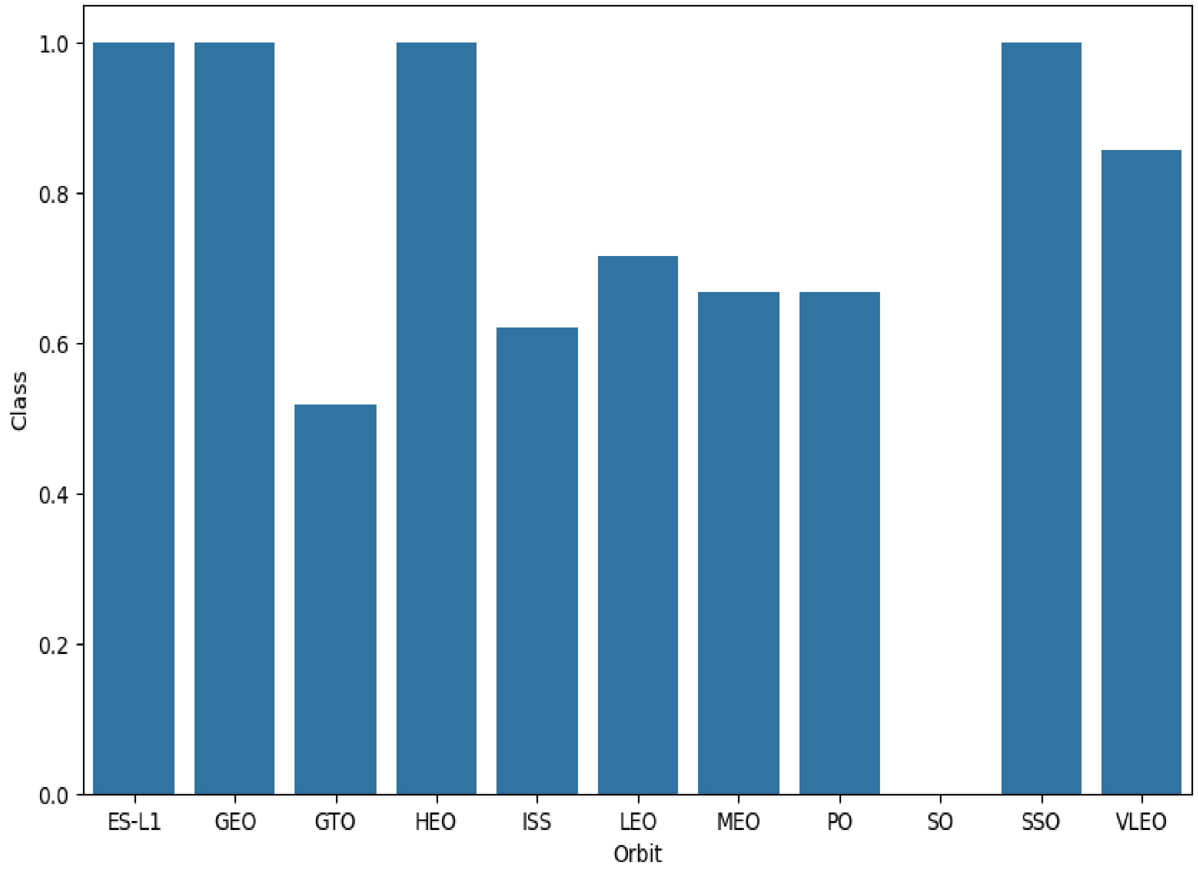
Payload vs Launch Site

- Greater the payload mass for launch site CCAFS SLC 40 the higher the success rate for the rocket



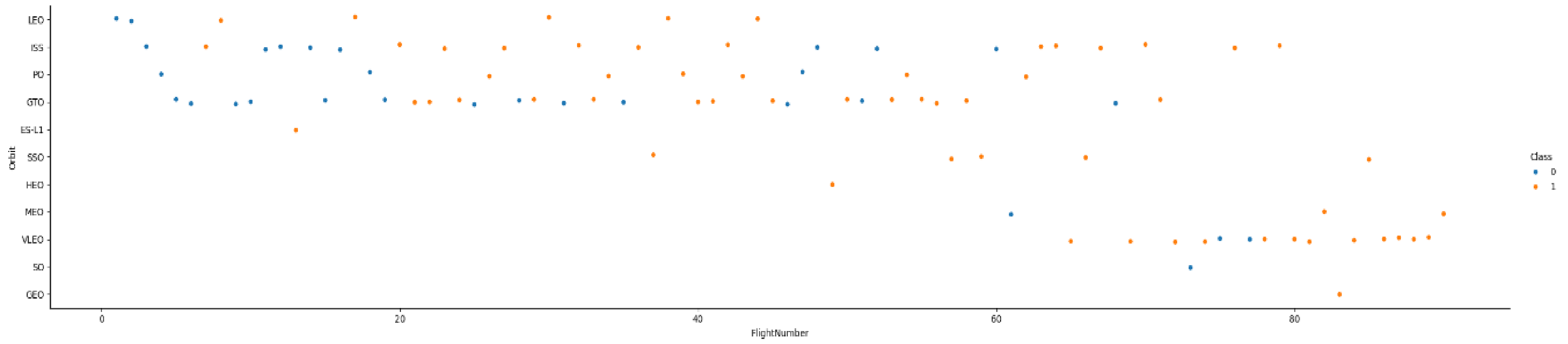
Success Rate vs Orbit Type

- From the bar graph, we can see that ES-L1, GEO, HEO, SSO, VLEO orbits had the most success rate
- The types of orbits will be defined in the Appendix1



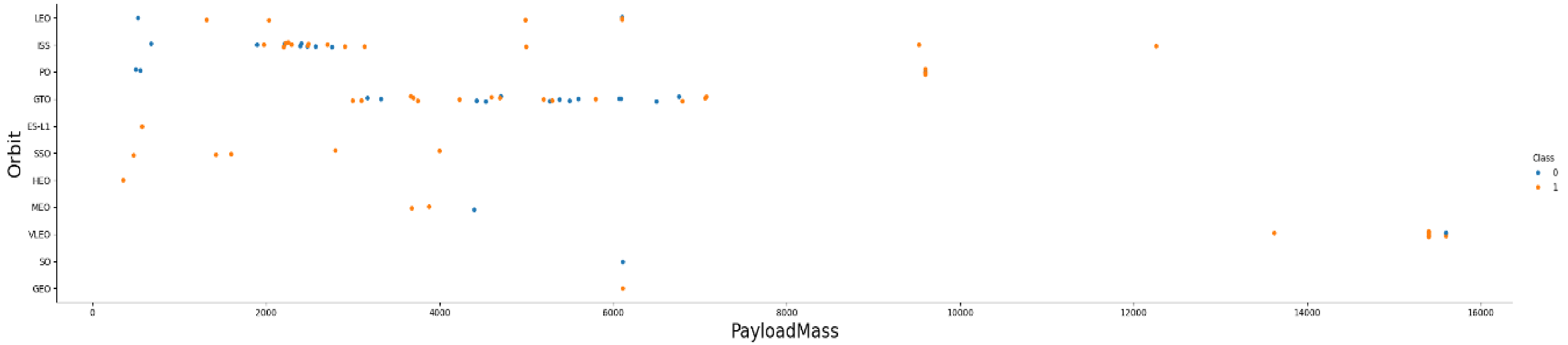
Flight Number vs Orbit Type

- The plot shows the flight number vs Orbit type. There are some vague correlations but nothing to indicate a relationship



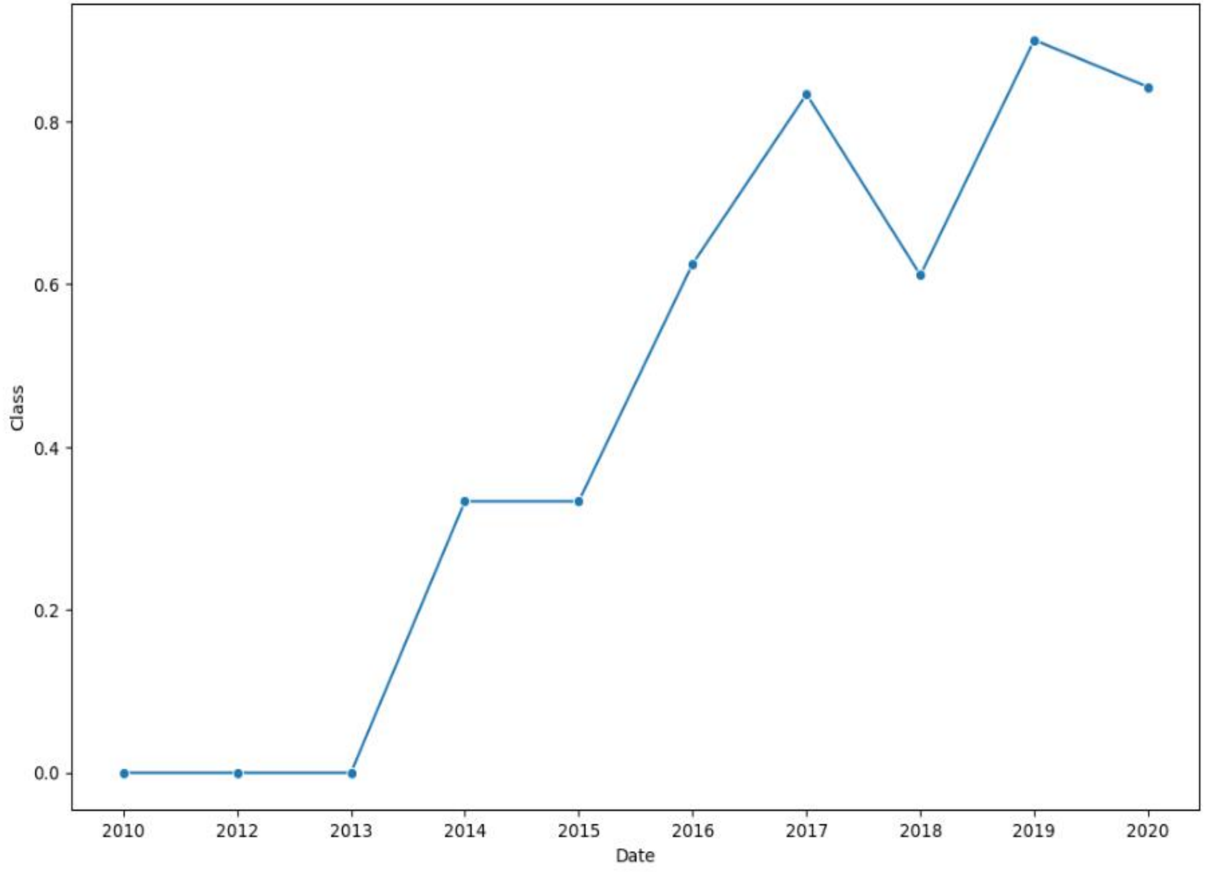
Payload vs Orbit Type

- Observe that with heavy payloads, there are more successful landings for PO, LEO and ISS orbits



Launch Success Yearly Trend

- From the line graph we can observe there is an increasing trend of proportions of successful landings



All Launch Site Names

- We used keyword DISTINCT to show only unique launch sites from the SpaceX data.
  - CCAFS LC-40 (Cape Canaveral Space Launch Complex 40)
  - VAFB SLC-4E (Vandenberg Space Launch Complex 4)
  - KSC LC-39A (Kennedy Space Center Launch Complex 39A)
  - CCAFS SLC-40 (Cape Canaveral Launch Complex 40)

```
%%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 query to display 5 records where the launch site begins with ‘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

Total Payload Mass

- Calculated the total payload carried by the boosters from NASA as 45,596 Kg.

```
%%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

- Calculated the average payload carried by the booster version F9 v1.1 as 2,534.7 Kg.

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

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

AVG("PAYLOAD_MASS__KG_")
2534.6666666666665

First Successful Ground Landing Date

- The date of the first successful landing outcome on ground pad was 22nd December 2015.

```
%%sql
SELECT MIN("Date") AS First_Successful_Landing_Date
FROM SPACEXTABLE
WHERE "Landing_Outcome" = 'Success (ground pad)';
```

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

First_Successful_Landing_Date
2015-12-22

Successful Drone Ship Landing with Payload between 4000 to 6000 kg

- Found the booster versions of successful outcomes with quantifies of payload mass between 4000 to 6000 Kg

```
%%sql
SELECT DISTINCT "Booster_Version"
FROM SPACEXTABLE
WHERE "Landing_Outcome" = '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

- Mission outcomes

```
%%sql
SELECT "Mission_Outcome", COUNT(*)
FROM SPACEXTABLE
GROUP BY "Mission_Outcome";
```

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

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

Booster Version That Carried the Maximum Payload

- The booter version that was able to carry the maximum payload of 15,600 kg.

```
%%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 display the month names, failure landing_outcomes in drone shipes, booster version, launch_site for the months in year 2015

```
%%sql
SELECT
    CASE
        WHEN substr("Date", 6, 2) = '01' THEN 'January'
        WHEN substr("Date", 6, 2) = '02' THEN 'February'
        WHEN substr("Date", 6, 2) = '03' THEN 'March'
        WHEN substr("Date", 6, 2) = '04' THEN 'April'
        WHEN substr("Date", 6, 2) = '05' THEN 'May'
        WHEN substr("Date", 6, 2) = '06' THEN 'June'
        WHEN substr("Date", 6, 2) = '07' THEN 'July'
        WHEN substr("Date", 6, 2) = '08' THEN 'August'
        WHEN substr("Date", 6, 2) = '09' THEN 'September'
        WHEN substr("Date", 6, 2) = '10' THEN 'October'
        WHEN substr("Date", 6, 2) = '11' THEN 'November'
        WHEN substr("Date", 6, 2) = '12' THEN 'December'
    END AS Month_Name,
    "Booster_Version",
    "Launch_Site"
FROM SPACEXTABLE
WHERE substr("Date", 1, 4) = '2015'
AND "Landing_Outcome" = 'Failure (drone ship)';
```

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

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

2015 Launch Records

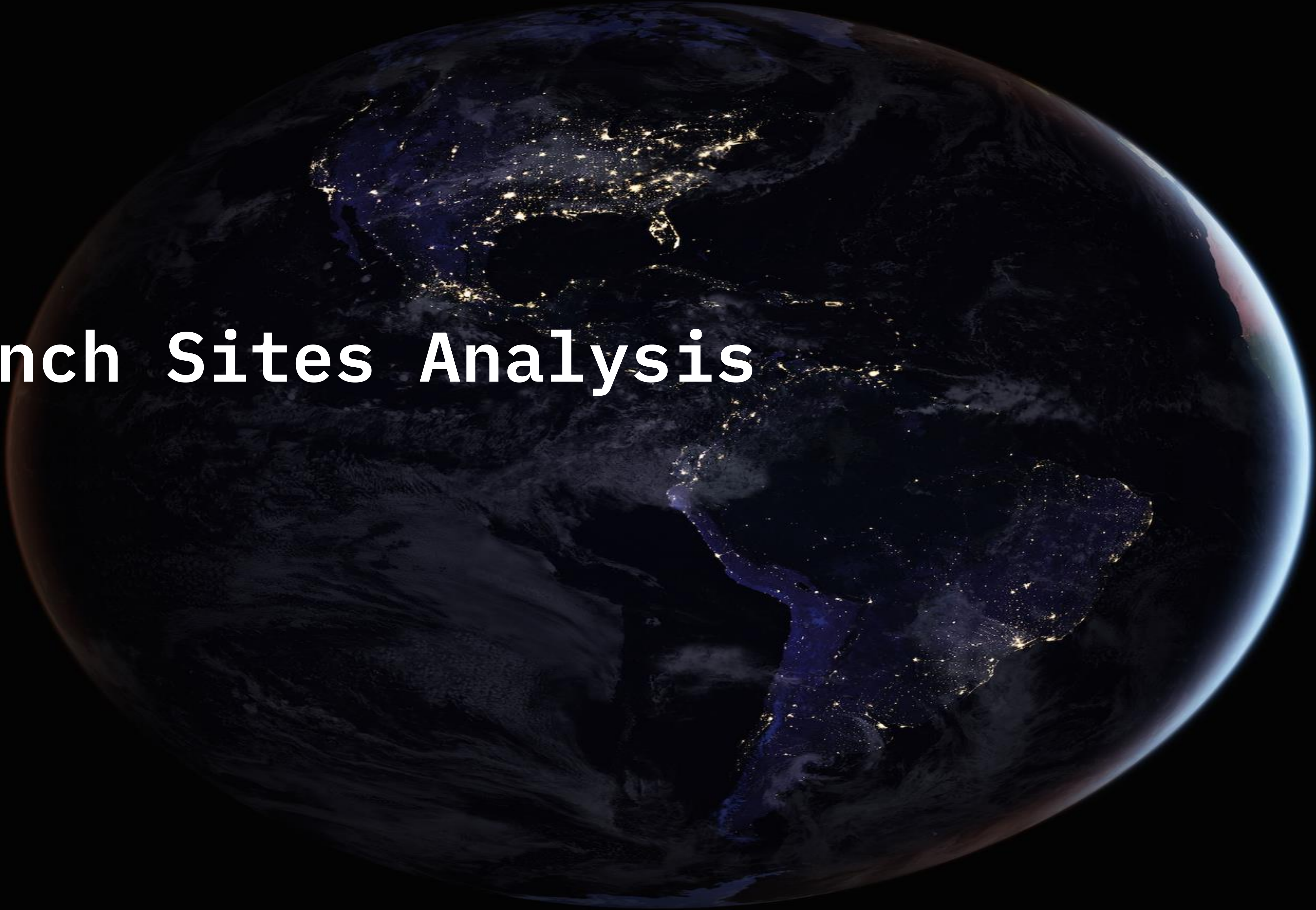
- List the records which display the month names, failure landing_outcomes in drone shipes, booster version, launch_site for the months in year 2015

```
%%sql
SELECT
    CASE
        WHEN substr("Date", 6, 2) = '01' THEN 'January'
        WHEN substr("Date", 6, 2) = '02' THEN 'February'
        WHEN substr("Date", 6, 2) = '03' THEN 'March'
        WHEN substr("Date", 6, 2) = '04' THEN 'April'
        WHEN substr("Date", 6, 2) = '05' THEN 'May'
        WHEN substr("Date", 6, 2) = '06' THEN 'June'
        WHEN substr("Date", 6, 2) = '07' THEN 'July'
        WHEN substr("Date", 6, 2) = '08' THEN 'August'
        WHEN substr("Date", 6, 2) = '09' THEN 'September'
        WHEN substr("Date", 6, 2) = '10' THEN 'October'
        WHEN substr("Date", 6, 2) = '11' THEN 'November'
        WHEN substr("Date", 6, 2) = '12' THEN 'December'
    END AS Month_Name,
    "Booster_Version",
    "Launch_Site"
FROM SPACEXTABLE
WHERE substr("Date", 1, 4) = '2015'
AND "Landing_Outcome" = 'Failure (drone ship)';
```

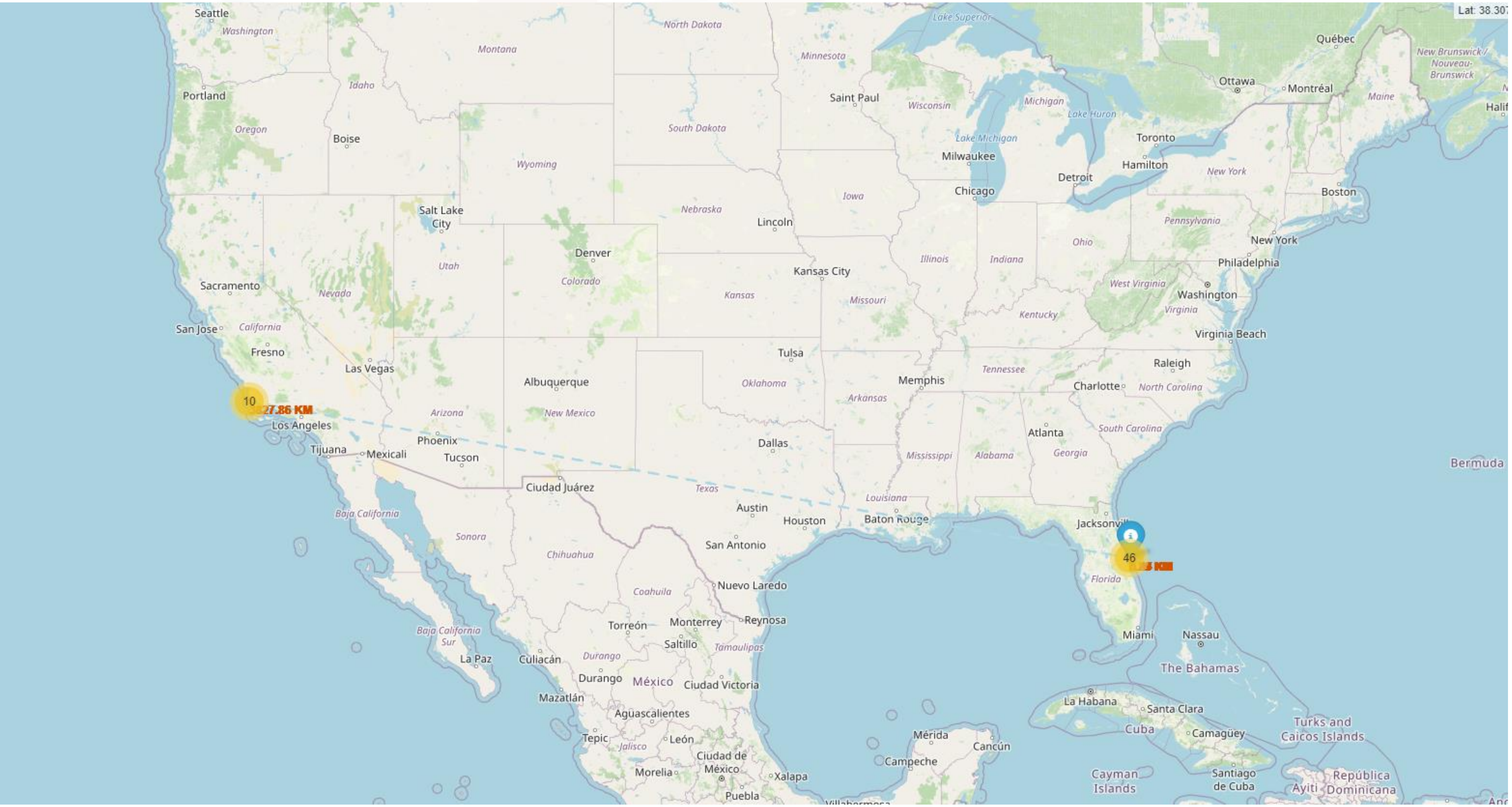
```
 * sqlite:///my_data1.db
Done.
```

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

Launch Sites Analysis



Launch Sites

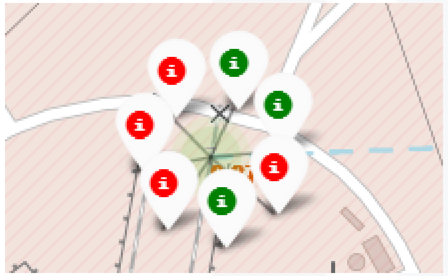


IBM Developer

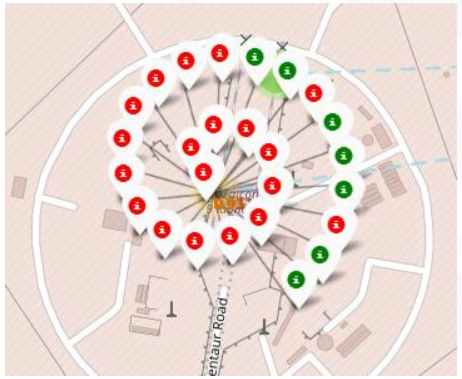
SKILLS NETWORK

Launch Outcomes at Various Launch Sites

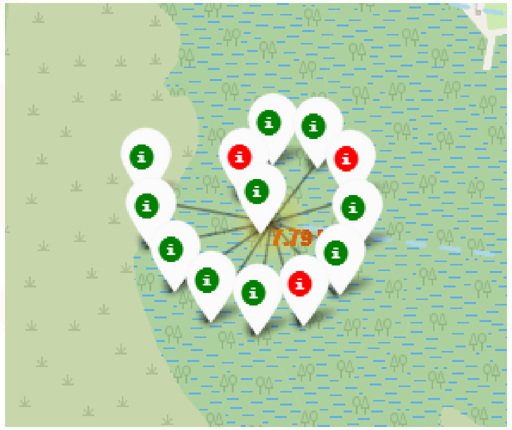
- CCAFS SLC-40 (Florida)



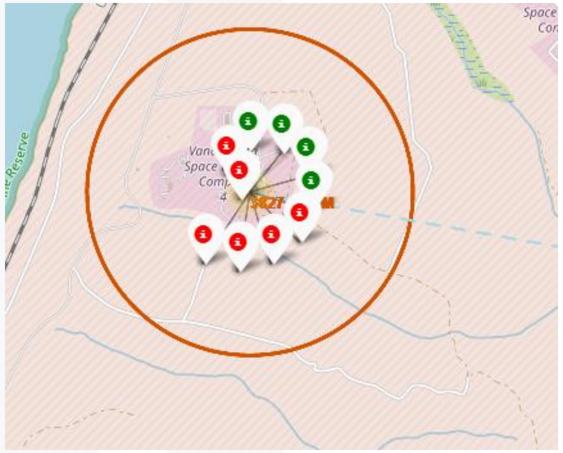
- CCAFS LC-40 (Florida)



- KSC LC-39A (Florida)



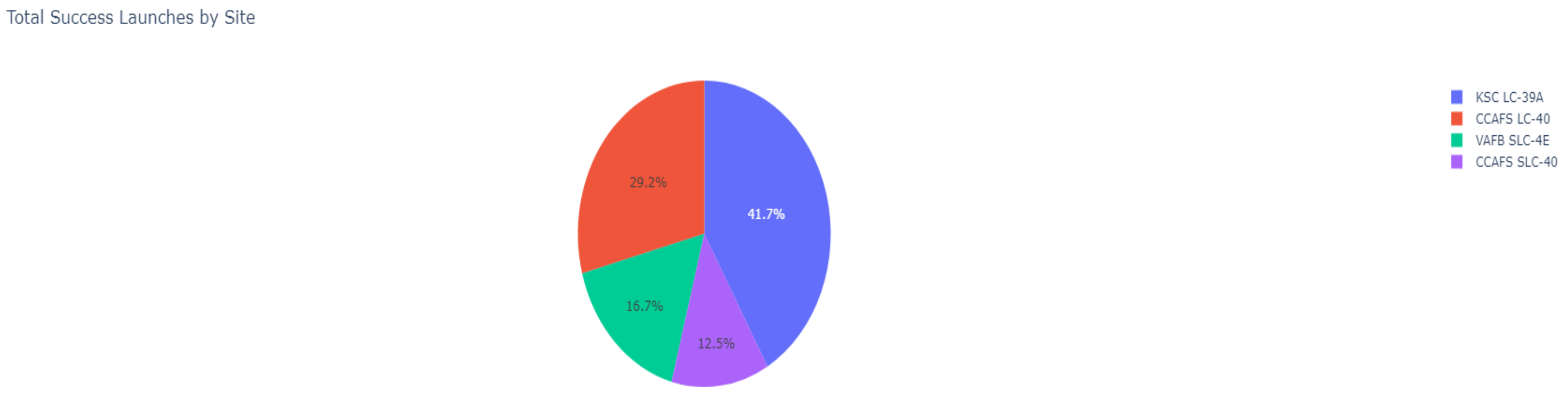
- VAFB SLC-4E (California)



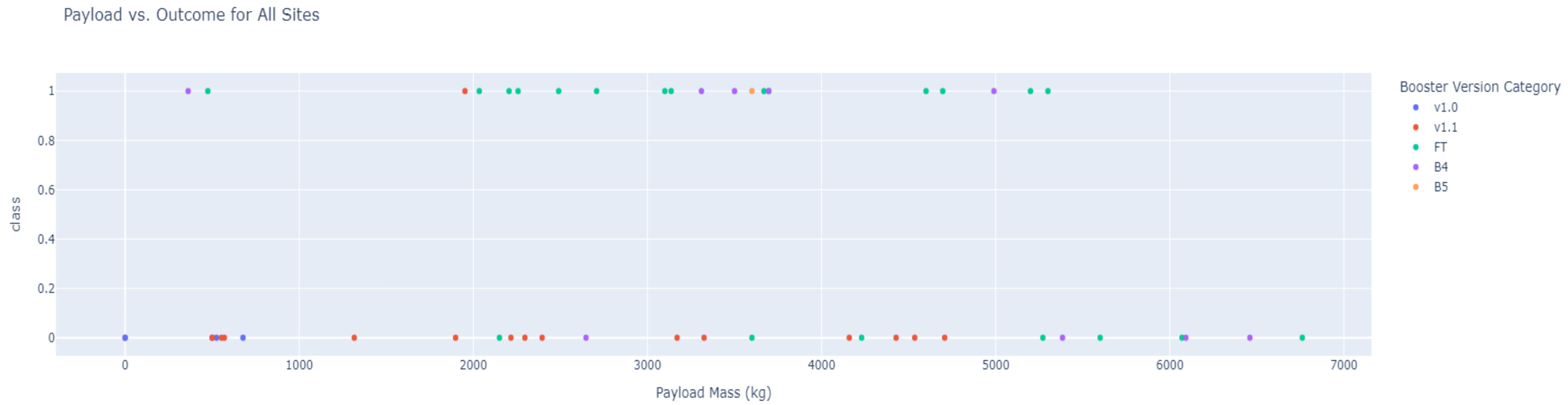
- Green = Success, Red = Failure

Dashboard with Plotly Dash

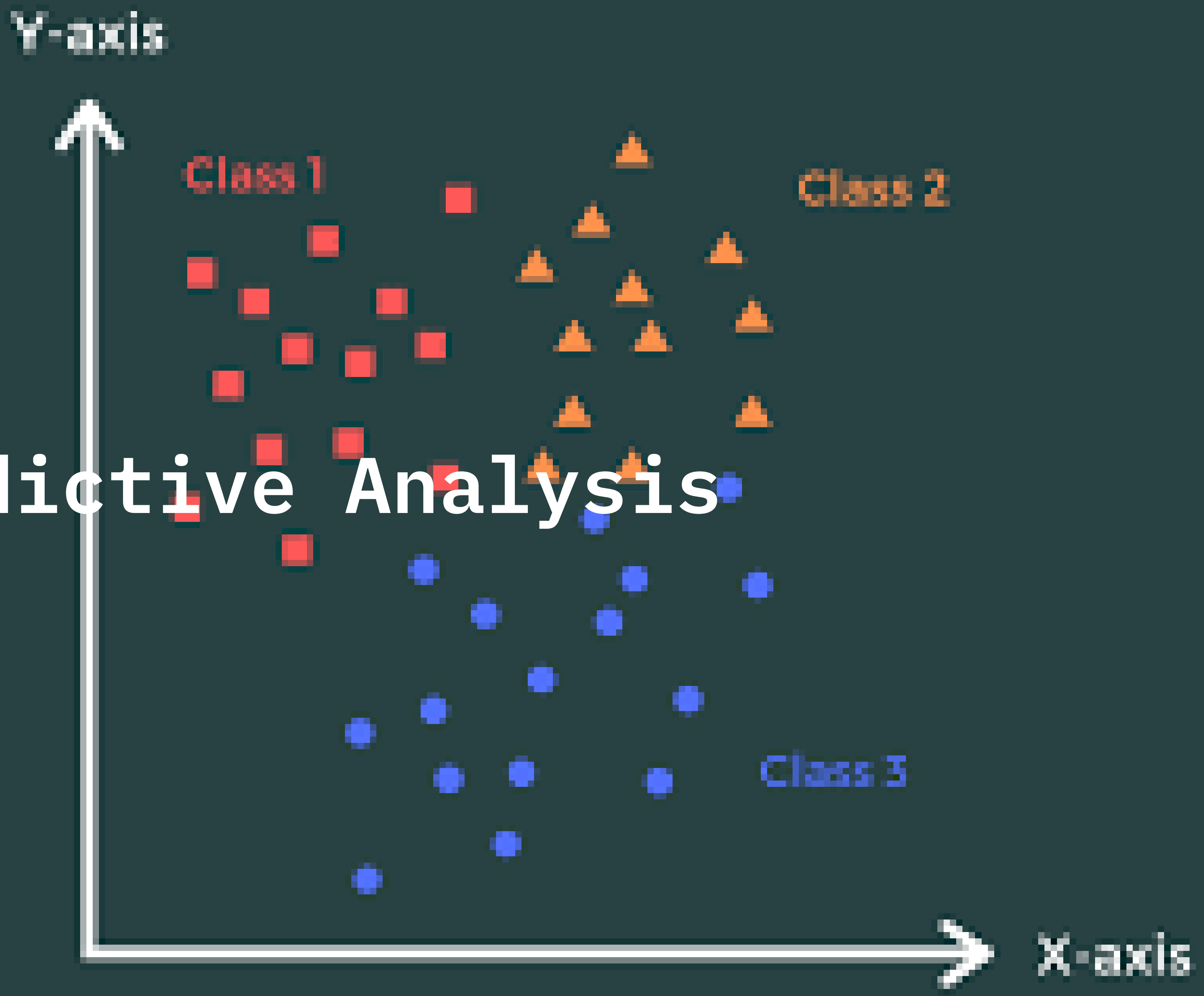
Pie Chart Showing Success Rate by Launch Site



Scatterplot Payload vs Launch Outcomes

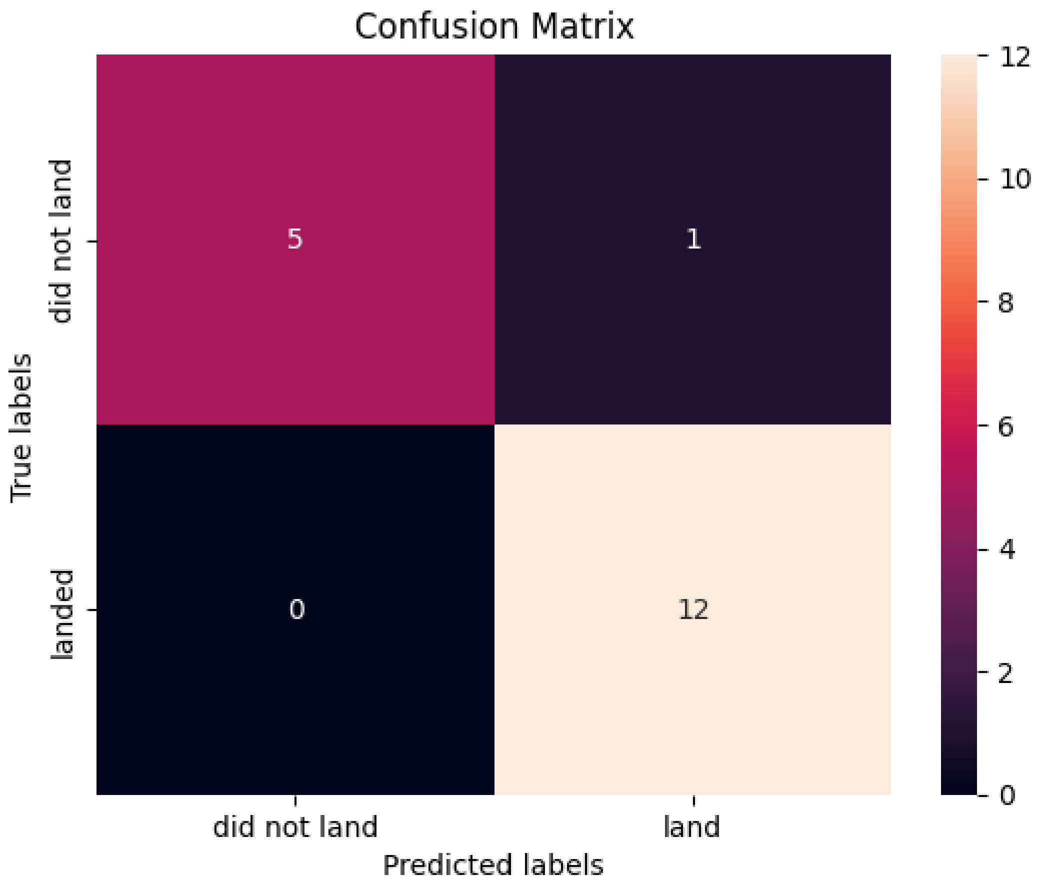


Predictive Analysis



Classification

- The decision tree classifier is the model with the highest accuracy score



Conclusions

- The larger flight number at a launch site, the greater the rate of success.
- Launch success rate increased from 2013 to 2020.
- Orbits ES-L1, GEO, HEO, SSO, VLEO had the most success rate.
- KSC LC-39A had the most successful launches of any sites.
- The Decision Tree classifier is the best machine learning algorithm to determine the task of “will it land?”.


Thank You

APPENDIX1

- Different orbit types
- **LEO**: Low Earth orbit (LEO)is an Earth-centred orbit with an altitude of 2,000 km (1,200 mi) or less (approximately one-third of the radius of Earth),[1] or with at least 11.25 periods per day (an orbital period of 128 minutes or less) and an eccentricity less than 0.25.[2] Most of the manmade objects in outer space are in LEO [1].
- **VLEO**: Very Low Earth Orbits (VLEO) can be defined as the orbits with a mean altitude below 450 km. Operating in these orbits can provide a number of benefits to Earth observation spacecraft as the spacecraft operates closer to the observation[2].
- **GTO** A geosynchronous orbit is a high Earth orbit that allows satellites to match Earth's rotation. Located at 22,236 miles (35,786 kilometers) above Earth's equator, this position is a valuable spot for monitoring weather, communications and surveillance. Because the satellite orbits at the same speed that the Earth is turning, the satellite seems to stay in place over a single longitude, though it may drift north to south," NASA wrote on its Earth Observatory website [3] .
- **SSO (or SO)**: It is a Sun-synchronous orbit also called a heliosynchronous orbit is a nearly polar orbit around a planet, in which the satellite passes over any given point of the planet's surface at the same local mean solar time [4] .
- **ES-L1** :At the Lagrange points the gravitational forces of the two large bodies cancel out in such a way that a small object placed in orbit there is in equilibrium relative to the center of mass of the large bodies. L1 is one such point between the sun and the earth [5] .
- **HEO** A highly elliptical orbit, is an elliptic orbit with high eccentricity, usually referring to one around Earth [6].
- **ISS** A modular space station (habitable artificial satellite) in low Earth orbit. It is a multinational collaborative project between five participating space agencies: NASA (United States), Roscosmos (Russia), JAXA (Japan), ESA (Europe), and CSA (Canada) [7]
- **MEO** Geocentric orbits ranging in altitude from 2,000 km (1,200 mi) to just below geosynchronous orbit at 35,786 kilometers (22,236 mi). Also known as an intermediate circular orbit. These are "most commonly at 20,200 kilometers (12,600 mi), or 20,650 kilometers (12,830 mi), with an orbital period of 12 hours [8]
- **HEO** Geocentric orbits above the altitude of geosynchronous orbit (35,786 km or 22,236 mi) [9]
- **GEO** It is a circular geosynchronous orbit 35,786 kilometres (22,236 miles) above Earth's equator and following the direction of Earth's rotation [10]
- **PO** It is one type of satellites in which a satellite passes above or nearly above both poles of the body being orbited (usually a planet such as the Earth [11]