IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Carmen

Oct 2025

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection API
  - Data Collection with Webscraping
  - Data Wrangling
  - Exploratory Analysis Using SQL
  - Exploratory Data Analysis for Data Visualization
  - Interactive Visual Analytics with Folium lab
  - Interactive Visual Analytics and Dashboard
  - Predictive Analysis
- Summary of all results
  - Dashboard
  - Predictive Analysis

Introduction

- Project background and context
  - Companies are making space travel affordable for everyone and Space Y would like to compete with SpaceX founded by Billionaire industrialist Allon Musk
- Problems you want to find answers
  - Predict if the Falcon 9 first stage will land successfully by training a machine learning model
  - Determine the cost of the launch

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - SpaceX API and webscrapping
- Perform data wrangling
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

Data Collection – SpaceX API

Steps to Process SpaceX Launch Data

1. Send GET request to retrieve launch data
2. Filter data for Falcon 9 launches only
3. Handle missing values in the dataset
4. URL: https://github.com/carmenlwm526-pixel/testrepo/blob/5c6d3bd530e6ed242c324bc51860508cdaadea1c/jupyter-labs-spacex-data-collection-api.ipynb

Data Collection - Scraping

Steps to Process Falcon9 Launch Data

1. Retrieve Falcon9 Launch Wiki page URL
2. Extract column and variable names from table header
3. Parse HTML tables to create data frame
4. URL: https://github.com/carmenlwm526-pixel/testrepo/blob/5c6d3bd530e6ed242c324bc51860508cdaadea1c/jupyter-labs-webscraping.ipynb

Data Wrangling

Data Analysis Steps

1. Calculate launches per site
2. Determine number and frequency of each orbit
3. Analyze mission outcomes by orbit
4. Generate landing outcome label from Outcome column
5. URL: https://github.com/carmenlwm526-pixel/testrepo/blob/5c6d3bd530e6ed242c324bc51860508cdaadea1c/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

1. Visualize below relationships with scatter chart
   - Flight Number & Launch Site
   - Flight Number & Pay Load Mass
   - Flight Number & Orbit Type
   - Pay Load Mass & Orbit Type
2. Visualize the relationship between success rate of each orbit type with bar chart
3. Visualize the launch success yearly trend with line chart
4. URL: https://github.com/carmenlwm526-pixel/testrepo/blob/f3f76f63bd3f364725bf75a6414724a8f77e9150/edadataviz%20(3).ipynb

EDA with SQL

SQL performed

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

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

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

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

5. List the date when the first succesful landing outcome in ground pad was acheived.

6. List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

7. List the total number of successful and failure mission outcomes

8. List all the booster_versions that have carried the maximum payload mass, using a subquery with a suitable aggregate function.

9. 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

10. 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

11. URL: https://github.com/carmenlwm526-pixel/testrepo/blob/6bb51a2b1d57d21f40daa87bd5c6254c4230d0c3/jupyter-labs-eda-sql-coursera_sqllite_2.ipynb

Build an Interactive Map with Folium

- Added map objects such as markers, circles, lines, etc. to a folium map
- To indicate the successful / failure launch and coastal line, highway etc.
- URL: https://github.com/carmenlwm526-pixel/testrepo/blob/45fc3b4558fe7967d9cd0bc65eeab1507d61fb10/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- Visualize the success count for all sites in pie chart
- Visualize the relationship of outcome and pay load mass in scatter chart
- URL: https://github.com/carmenlwm526-pixel/testrepo/blob/93f525e41ca93fc4d0b6bf58cfc7acb66595172f/project%20(1).py

Predictive Analysis (Classification)

Steps for Model Training and Evaluation

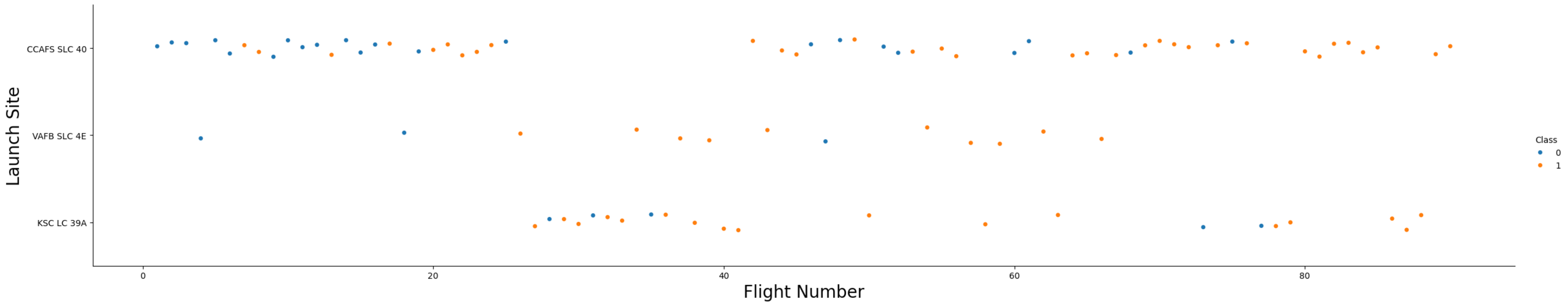
1. Divide data into training and test sets
2. Build logistic regression model and assess accuracy
3. Build support vector machine model and assess accuracy
4. Build decision tree classifier and assess accuracy
5. Build k nearest neighbors model and assess accuracy
6. Determine best performing model
7. URL: https://github.com/carmenlwm526-pixel/testrepo/blob/39ad35444ae0eeff39a56a9bacfe0a81844c4d90/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Section 2

Insights drawn from EDA

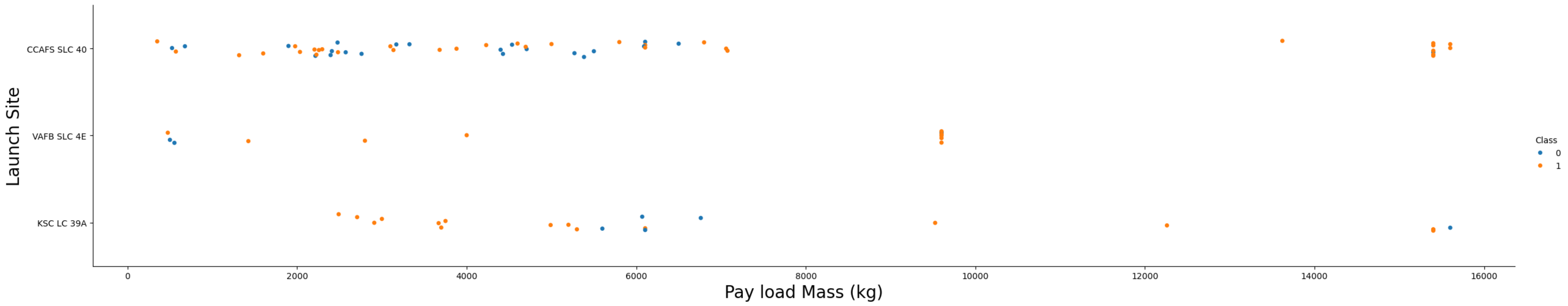
Flight Number vs. Launch Site

- Higher flight number, higher launch rate



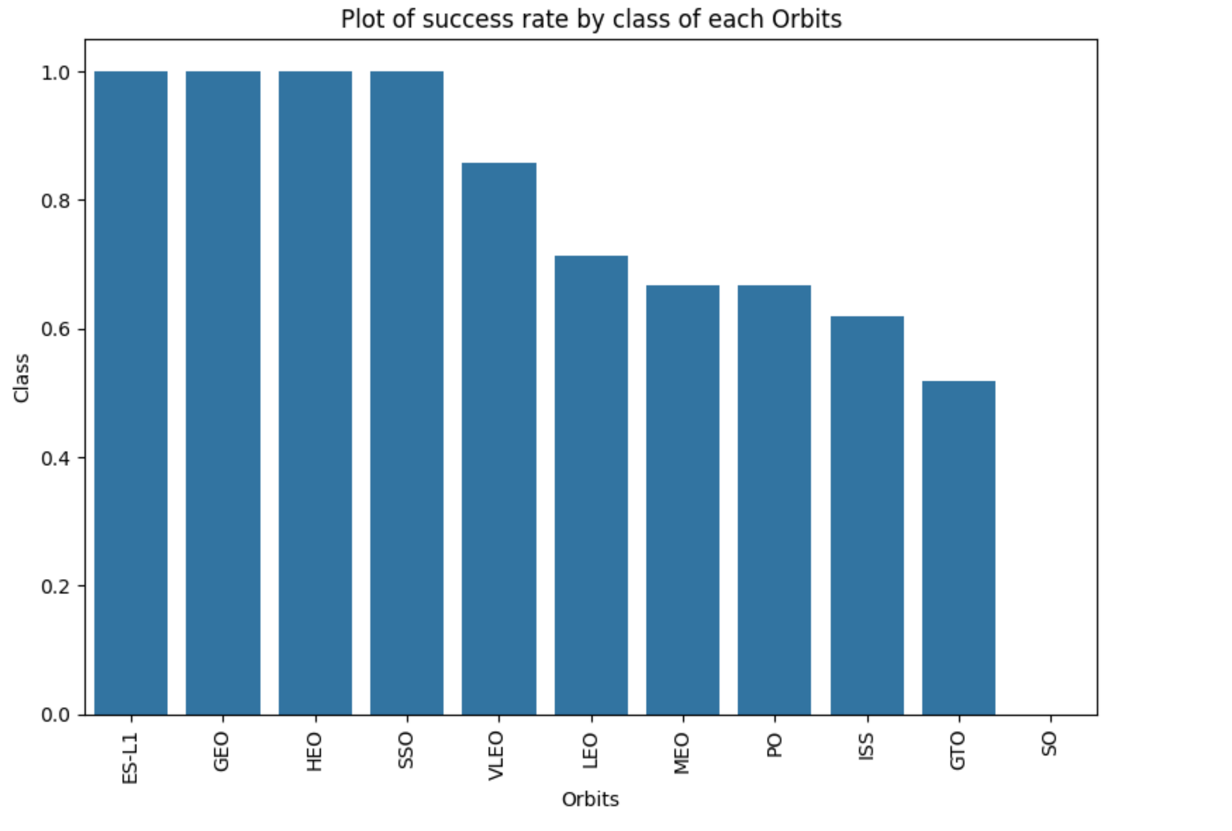
Payload vs. Launch Site

- Higher launch rate for CCAFS and KSC, especially on higher pay load mass



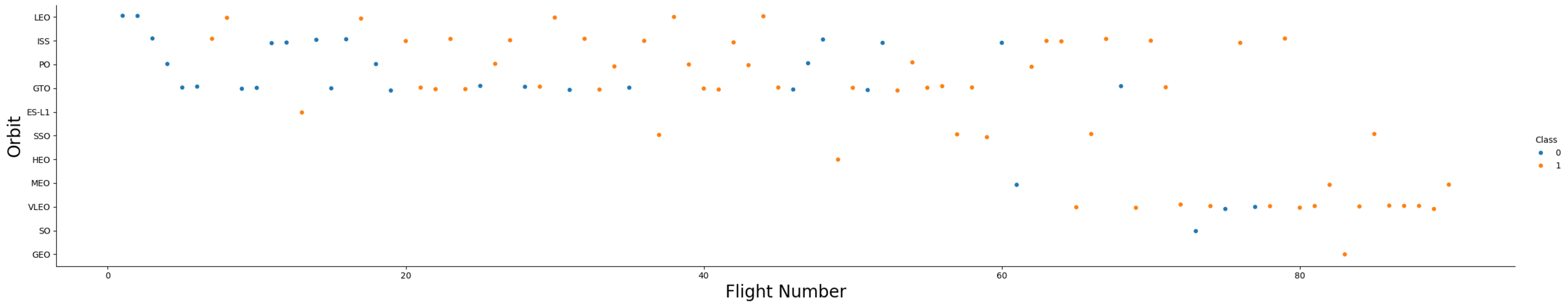
Success Rate vs. Orbit Type

- ES-L1, GEO, HEO, SSO have highest success rate



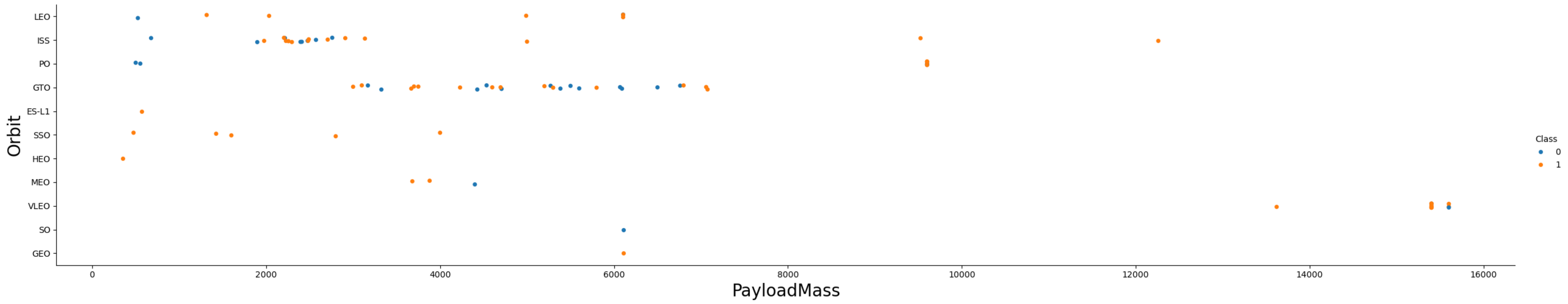
Flight Number vs. Orbit Type

- LEO orbit, success seems to be related to the number of flights. Conversely, in the GTO orbit, there appears to be no relationship between flight number and success.



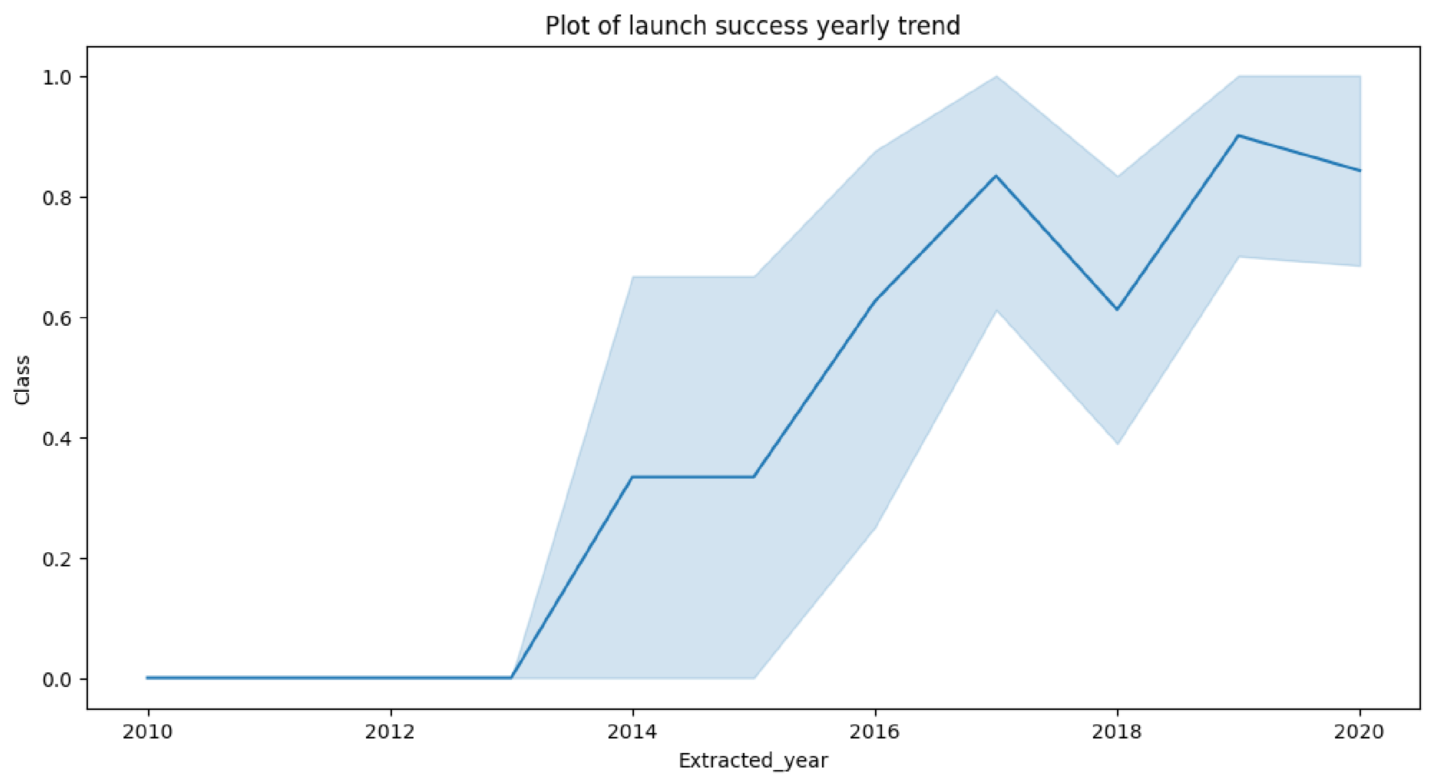
Payload vs. Orbit Type

- With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.
- However, for GTO, it's difficult to distinguish between successful and unsuccessful landings as both outcomes are present.



Launch Success Yearly Trend

- The success rate since 2013 kept increasing till 2020



All Launch Site Names

- Find the names of the unique launch sites

Task 1

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

```
%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'

- Find 5 records where launch sites begin with `CCA`

Task 2

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

```
%sql SELECT * from SPACEXTBL 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

- Calculate the total payload carried by boosters from NASA

Task 3

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

```
%sql select sum(PAYLOAD_MASS__KG_) as payloadmass from SPACEXTBL WHERE Customer = 'NASA (CRS)';
```

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

payloadmass
45596

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1

Task 4

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

```
%sql select avg(PAYLOAD_MASS__KG_) as payloadmass from SPACEXTBL WHERE Booster_Version = 'F9 v1.1';
```

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

payloadmass
2928.4

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad

Task 5

List the date when the first succesful landing outcome in ground pad was acheived.

Hint:Use min function

```
%sql select min(DATE) from SPACEXTBL 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

- List the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

Task 6

List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

```
lect Booster_Version from SPACEXTBL where Landing_Outcome='Success (drone ship)' and PAYLOAD_MASS__KG_ BETWEEN 4000 and 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

- Calculate the total number of successful and failure mission outcomes

Task 7

List the total number of successful and failure mission outcomes

```
%sql select Mission_Outcome, count(MISSION_OUTCOME) as missionoutcomes from SPACEXTBL GROUP BY MISSION_OUTCOME;
```

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

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

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass

Task 8

List all the booster_versions that have carried the maximum payload mass, using a subquery with a suitable aggregate function.

```
lect BOOSTER_VERSION as boosterversion from SPACEXTBL where PAYLOAD_MASS__KG_=(select max(PAYLOAD_MASS__KG_) from SPACEXTBL);
```

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

boosterversion
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 failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

Task 9

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.

```
substr(Date, 6,2) as month ,BOOSTER_VERSION,LAUNCH_SITE FROM SPACEXTBL where Landing_Outcome LIKE 'Failure (drone ship)' AND
```

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

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

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

Task 10

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.

```
ME, count(*) FROM SPACEXTBL WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20' 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
Failure (parachute)	2
Precluded (drone ship)	1

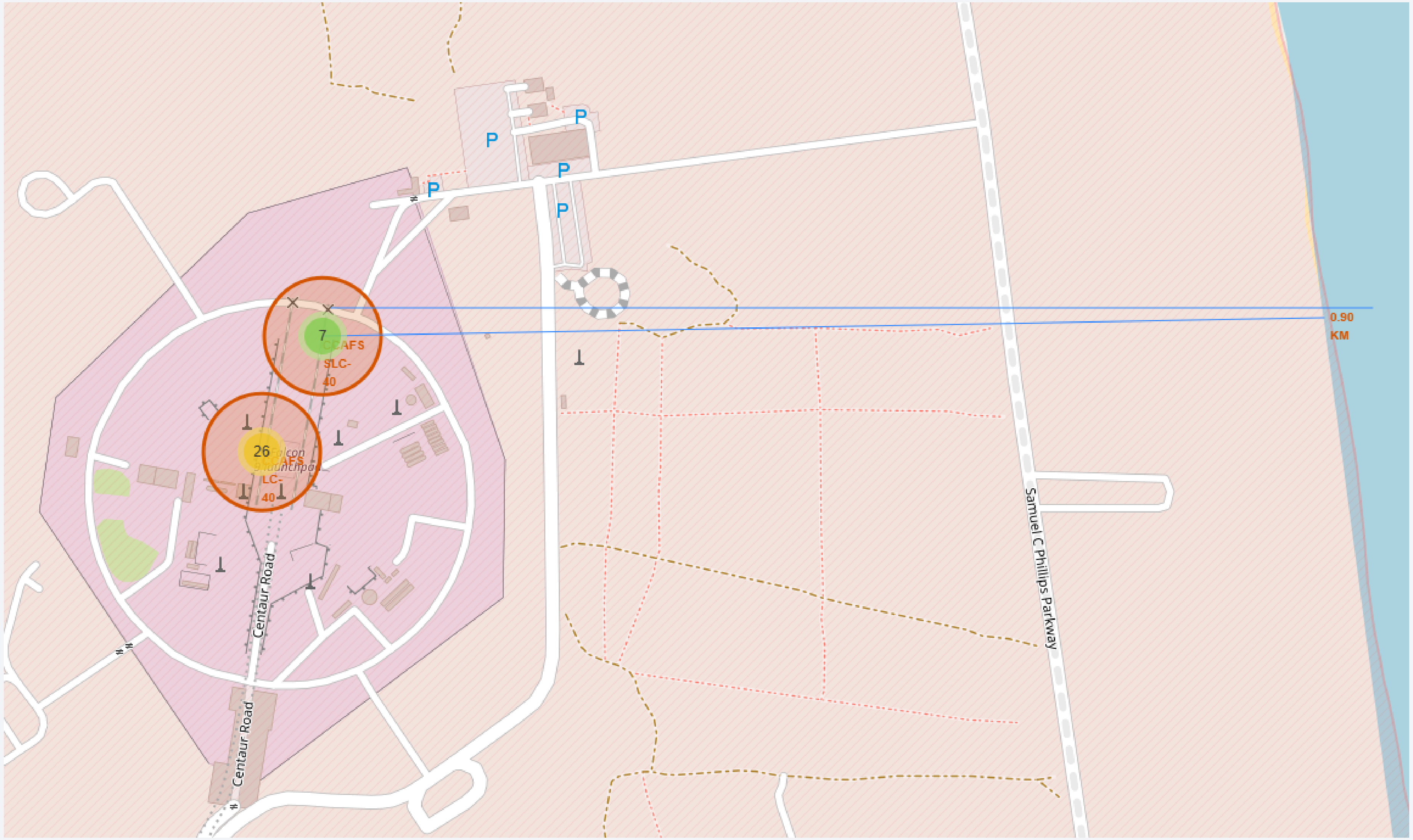
Section 3

Launch Sites Proximities Analysis

Mark all launch sites on a map

Mark the success/failed launches for each site on the map

Selected launch site with railway, highway, coastline and distance

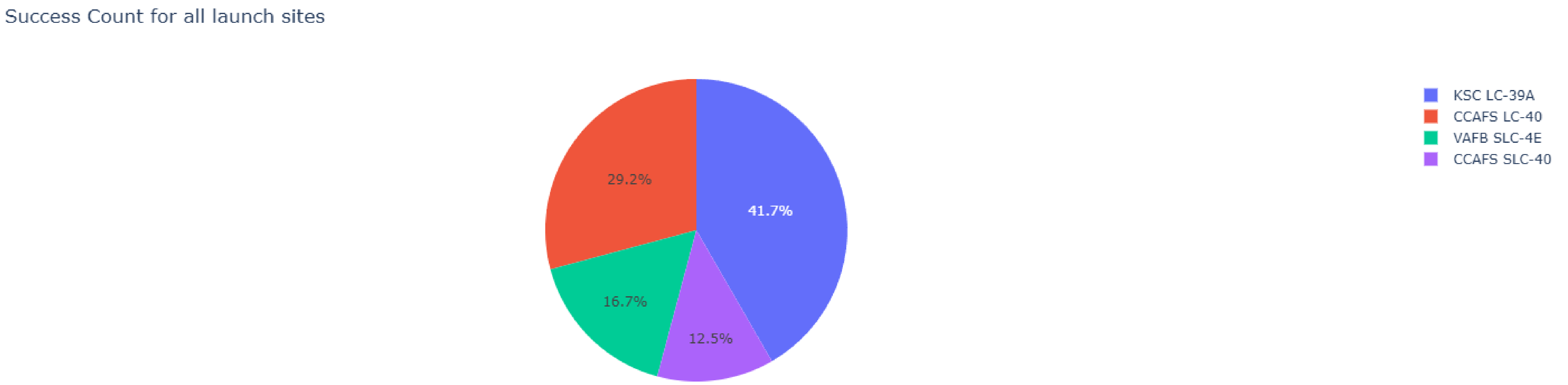


Section 4

Build a Dashboard with Plotly Dash

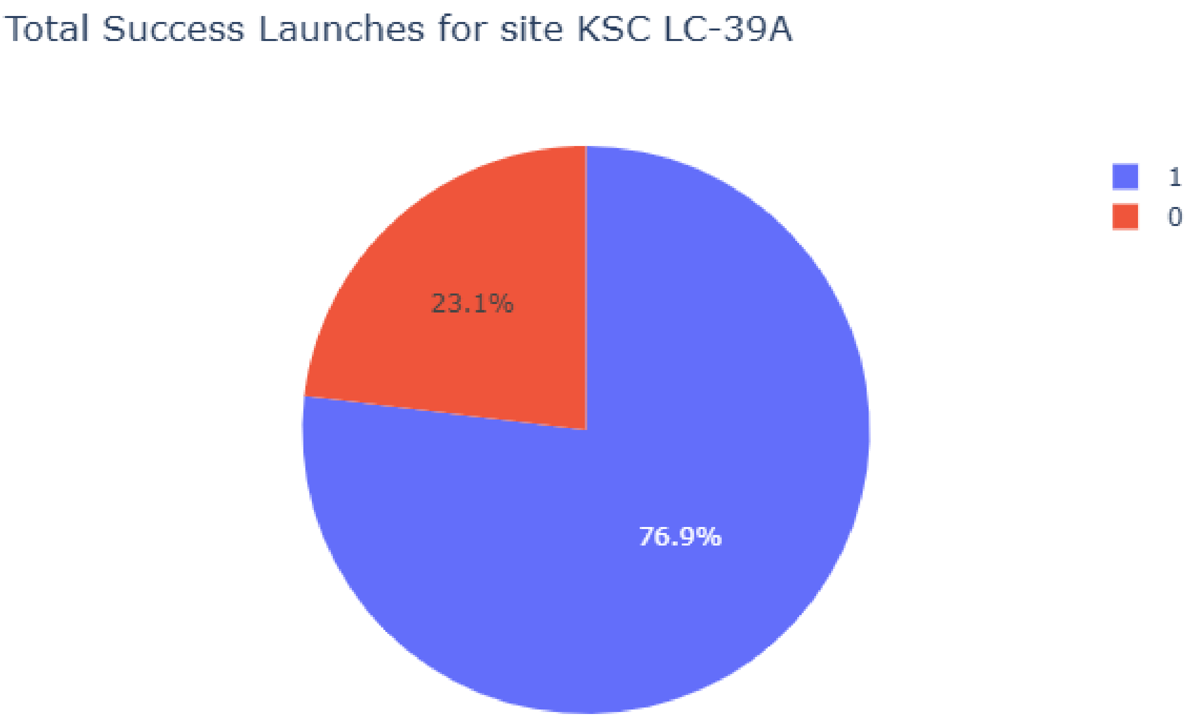
Success Count for all launch sites

- KSC has the highest success count



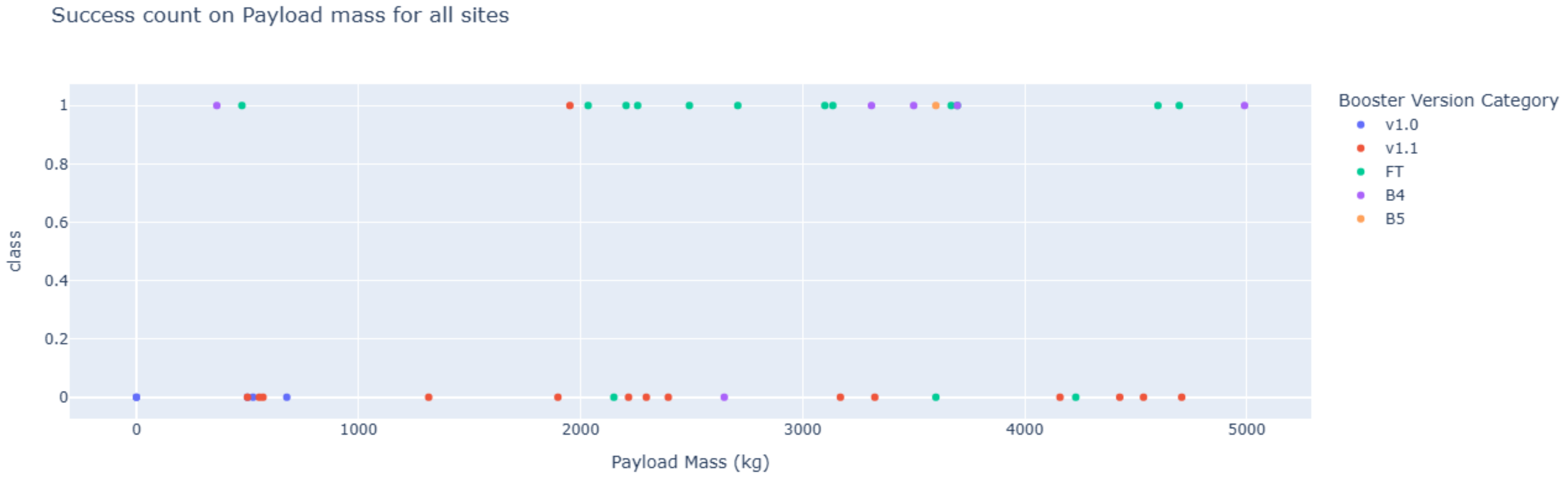
Total Success Launch for each site

- KSC has the highest success rate



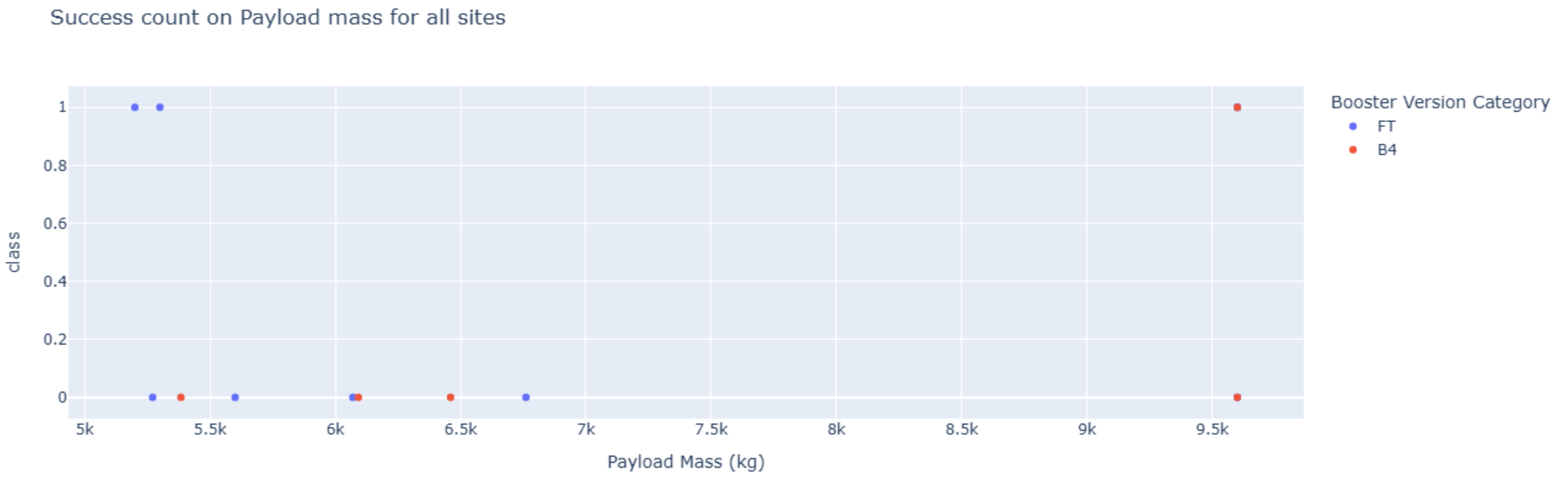
Success rate on payload mass for different classes

- 0-5000 kg



- Lower payload mass has higher success rate

- 5000-10000 kg

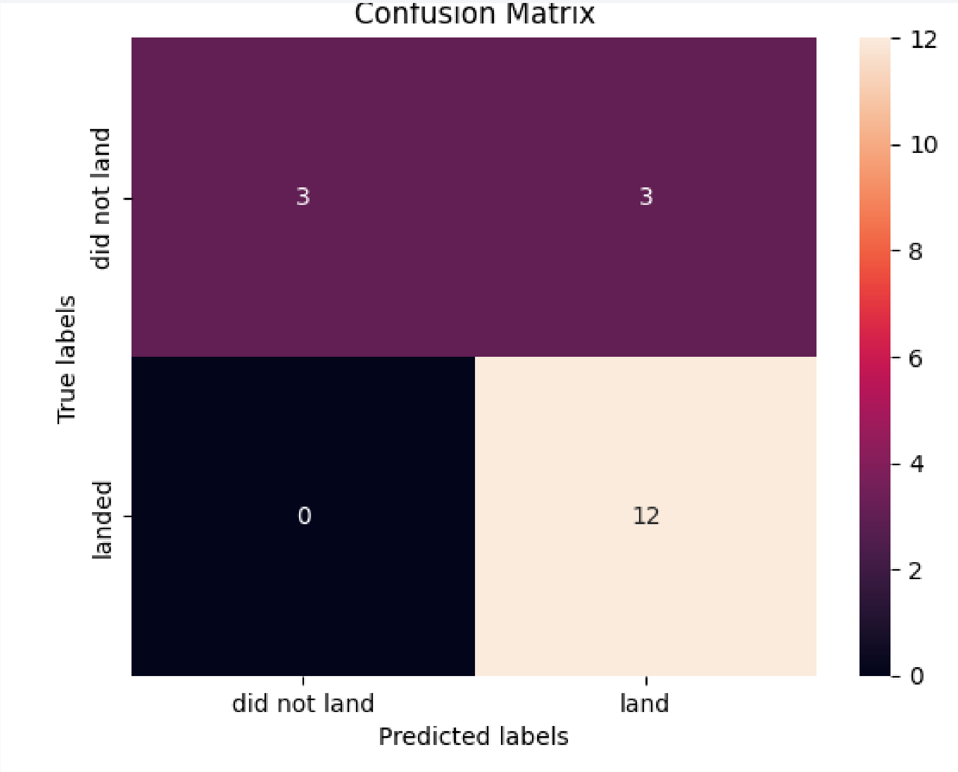


Section 5

Predictive Analysis (Classification)

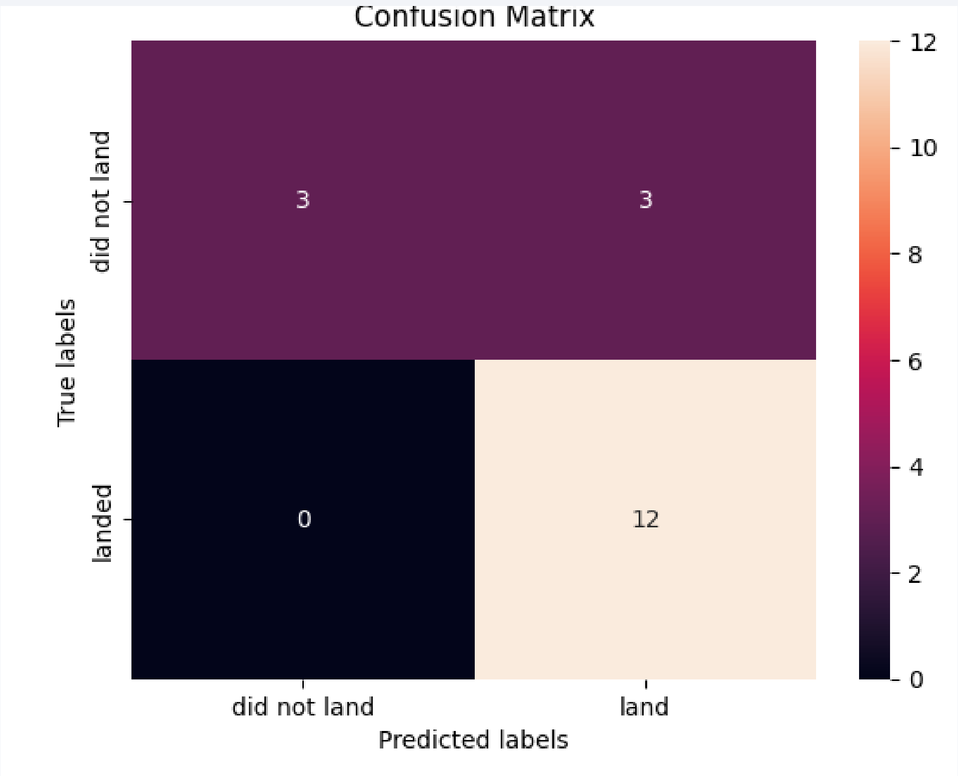
Classification Accuracy

- Best model is DecisionTree with a score of 0.87



Confusion Matrix of decision tree

- 12 correct predictions for successful landings (True Positives).
- 3 correct predictions for failed landings (True Negatives).
- No False Negatives — the model didn't miss any actual landings
- 3 False Positives — the model predicted a landing when it didn't actually happen.



Conclusions

- Higher flight number, higher launch rate
- The success rate since 2013 kept increasing till 2020
- KSC has the highest success count
- ES-L1, GEO, HEO, SSO have highest success rate
- Lower payload mass has higher success rate
- Launch sites are also located near highways and railways
- Decision tree performs best among 4 machine training models

Thank you!