IBM Data Science Project SpaceX

Carlos Jesus 2025

OUTLINE

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

EXECUTIVE SUMMARY

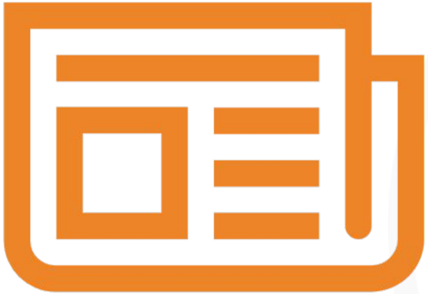
- Summary of Methodologies
  - Data Collection
  - Data Wrangling
  - EDA with Data Visualization
  - EDA with SQL
  - Dashboard with Plotly Dash
  - Predictive Analysis
- Summary of Results
  - Exploratory Results
  - Predictive Analysis Results

INTRODUCTION

- Project Background
  - We wanted to predict if Falcon 9 rocket first stage would land successfully.
  - We also wanted to determine the cost of a launch.

- Roadblocks
  - What influence a successful land?
  - What are the conditions necessary to get the best results?

METHODOLOGY



- Data Collection Methodology
  - SpaceX Rest API
- Exploratory Data Analysis (EDA)
- Visual Analytics with Folium and Plotly Dash
- Predictive Analysis with classification models

RESULTS

- Exploratory Data Analysis
- Visual Analytics
- Predictive Analysis

Data Collection - SpaceX API

Getting response from API

```
spacex_url="https://api.spacexdata.com/v4/launches/past"
```

```
response = requests.get(spacex_url)
```

Convert to .json

```
response = requests.get(static_json_url).json()
data = pd.json_normalize(response)
```

Assign list to dictionary then to data frame

```
launch_dict = {'FlightNumber': list(data['flight_number']),
'Date': list(data['date']),
'BoosterVersion':BoosterVersion,
'PayloadMass':PayloadMass,
'Orbit':Orbit,
'LaunchSite':LaunchSite,
'Outcome':Outcome,
'Flights':Flights,
'GridFins':GridFins,
'Reused':Reused,
'Legs':Legs,
'LandingPad':LandingPad,
'Block':Block,
'ReusedCount':ReusedCount,
'Serial':Serial,
'Longitude': Longitude,
'Latitude': Latitude}
```

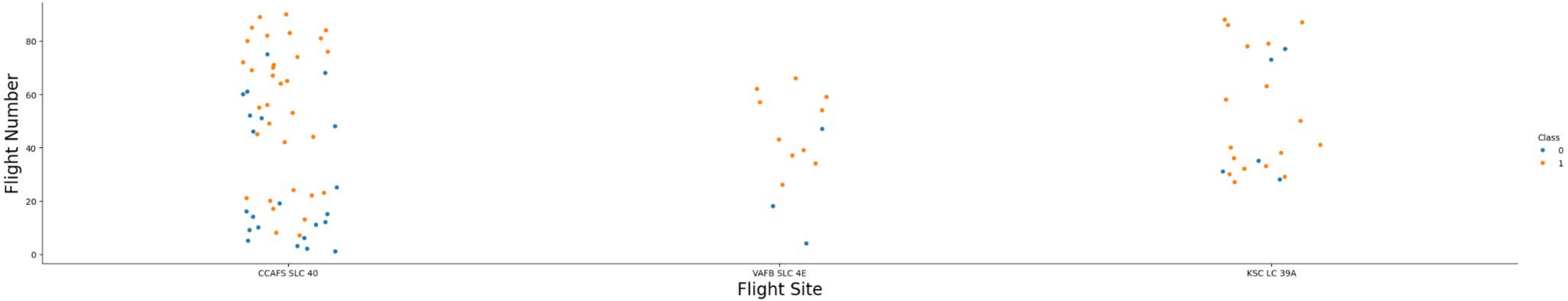
```
df = pd.DataFrame.from_dict(launch_dict)
```

Export to .csv

```
data_falcon9.to_csv('dataset_part_1.csv', index=False)
```

EDA with visualization

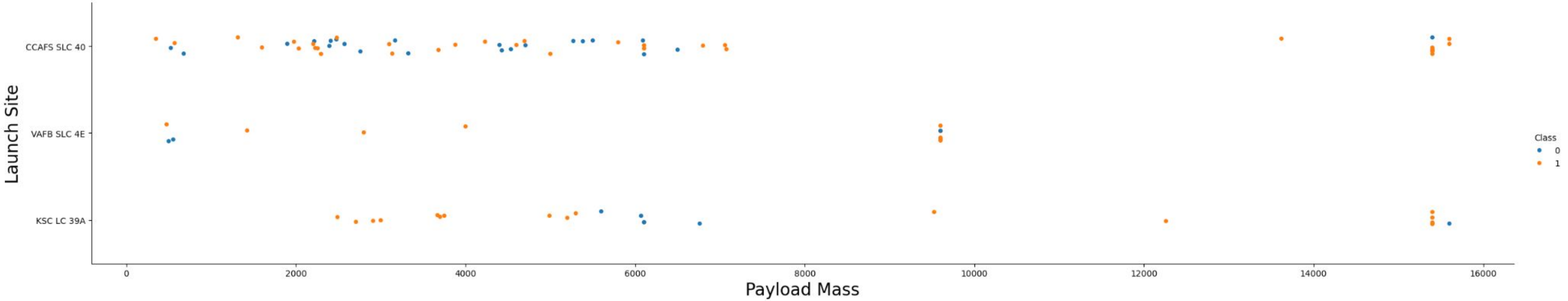
Flight Number vs Flight Site



The more launches exist at a launch site, greater is the success rate at said site.

EDA with visualization

Launch Site vs Payload Mass

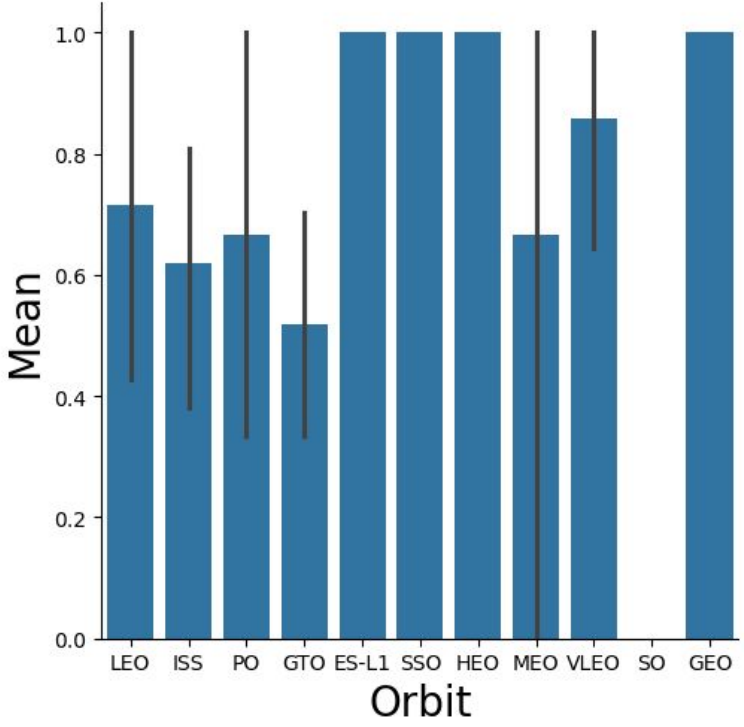


The greater the payload mass for site CCAFS SLC 40 higher is the success rate for the rocket.

EDA with visualization

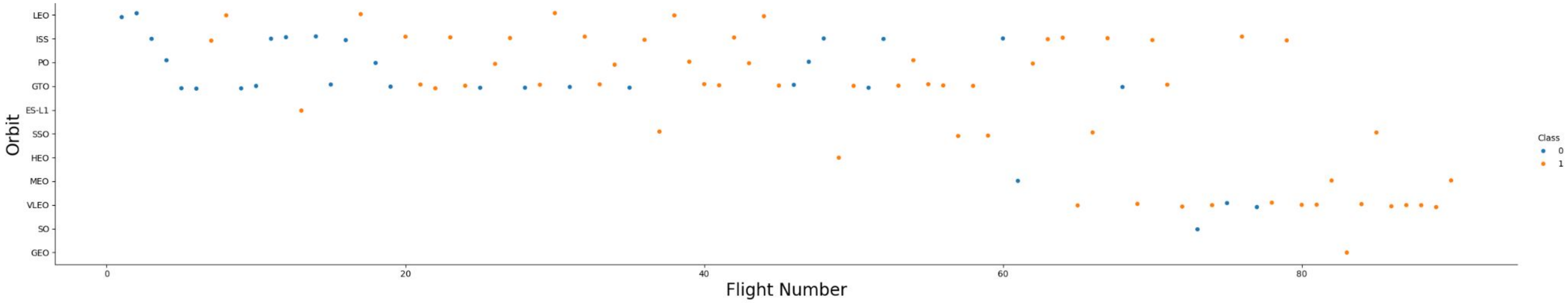
Success Rate vs Orbit Type

Orbits GEO, HEO, SSO and ES L1 have the best Success rate



EDA with visualization

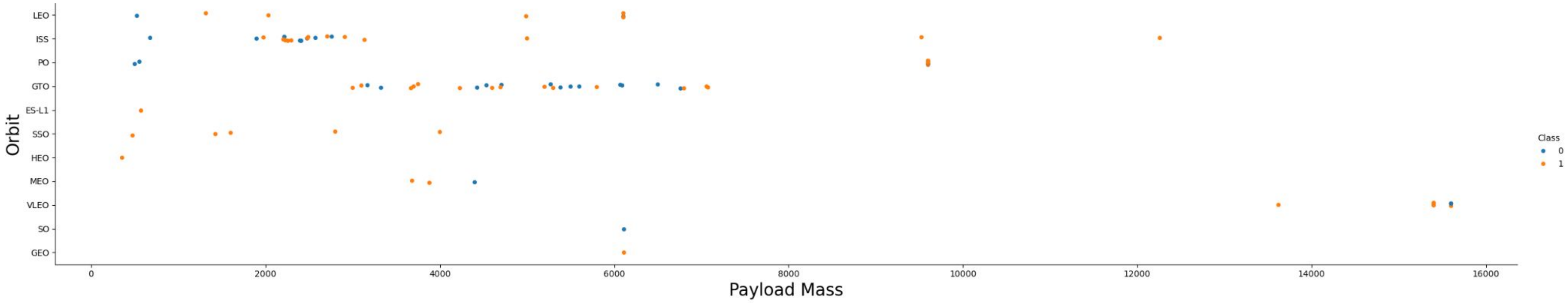
Flight Number vs Orbit Type



LEO orbit success appears to be related to the number of flights. There seems to be no relation on the GTO orbit.

EDA with visualization

Payload Mass vs Orbit Type

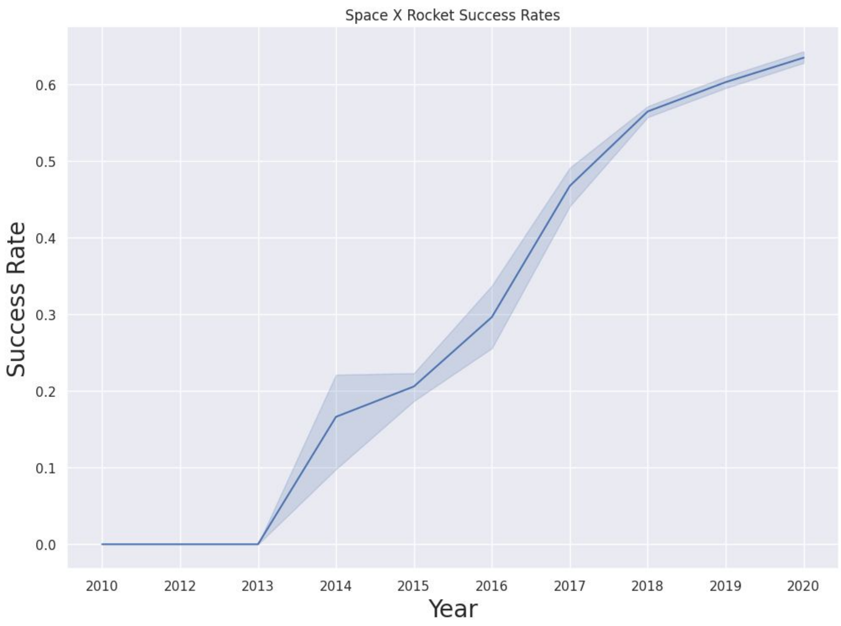


We can observe that heavier payloads has a negative influence on the GTO orbit and positive on the LEO and ISS orbits.

EDA with visualization

Launch Success yearly trend

We can observe that the launch success has been increasing since 2013.



EDA with SQL

SQL query

select DISTINCT Launch_Site from SPACEXTBL

Retrieving unique values for the Launch_Site Column

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

EDA with SQL

SQL query

select * from SPACEXTBL where Launch_Site Like 'CCA%' Limit 5;

Retrieving 5 lines where Launch_Site starts with “CCA”

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

EDA with SQL

SQL query

select SUM(PAYLOAD_MASS__KG_) from SPACEXTBL where Customer = 'NASA (CRS)'

Retrieving the total payload mass carried by boosters launched by “NASA (CRS)”

SUM(PAYLOAD_MASS_KG_)
45596

EDA with SQL

SQL query

select AVG(PAYLOAD_MASS__KG_) from SPACEXTBL where Booster_Version = 'F9 v1.1'

Retrieving the average payload mass by booster version “F9 v1.1”

AVG(PAYLOAD_MASS__KG_)
2928.4

EDA with SQL

SQL query

select MIN(Date) from SPACEXTBL where Landing_Outcome = 'Success (ground pad)'

Retrieving the date when the first successful landing in "ground pade" was achieved

MIN(Date)
2015-12-22

EDA with SQL

SQL query

select Booster_Version from SPACEXTBL where Landing_Outcome = 'Success (drone ship)' AND PAYLOAD_MASS__KG_ > 4000 AND PAYLOAD_MASS__KG_ < 6000

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

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

EDA with SQL

SQL query

select (select count(*) from SPACEXTBL where Mission_Outcome Like 'Success%') as Mission_Success, (select count(*) from SPACEXTBL where Mission_Outcome Like 'Failure%') as Mission_Failure

Retrieving the total number of successes and failures

Mission_Success	Mission_Failure
100	1

EDA with SQL

SQL query

```
select DISTINCT Booster_Version, PAYLOAD_MASS__KG_ FROM SPACEXTBL
WHERE PAYLOAD_MASS__KG_ = (select MAX(PAYLOAD_MASS__KG_) FROM
SPACEXTBL) GROUP BY Booster_Version
```

Retrieving the names of the booster_versions which have carried the maximum payload mass

Booster_Version	PAYLOAD_MASS__KG_
F9 B5 B1048.4	15600
F9 B5 B1048.5	15600
F9 B5 B1049.4	15600
F9 B5 B1049.5	15600
F9 B5 B1049.7	15600
F9 B5 B1051.3	15600
F9 B5 B1051.4	15600
F9 B5 B1051.6	15600
F9 B5 B1056.4	15600
F9 B5 B1058.3	15600
F9 B5 B1060.2	15600
F9 B5 B1060.3	15600

EDA with SQL

SQL query

select substr(Date, 6,2) as Month, Landing_Outcome, Booster_Version, Launch_Site from SPACEXTBL where Landing_Outcome = 'Failure (drone ship)' and substr(Date,0,5)='2015'

Retrieving the records displaying month number in 2015, where the landing outcome on "drone ship" was a failure

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

EDA with SQL

SQL query

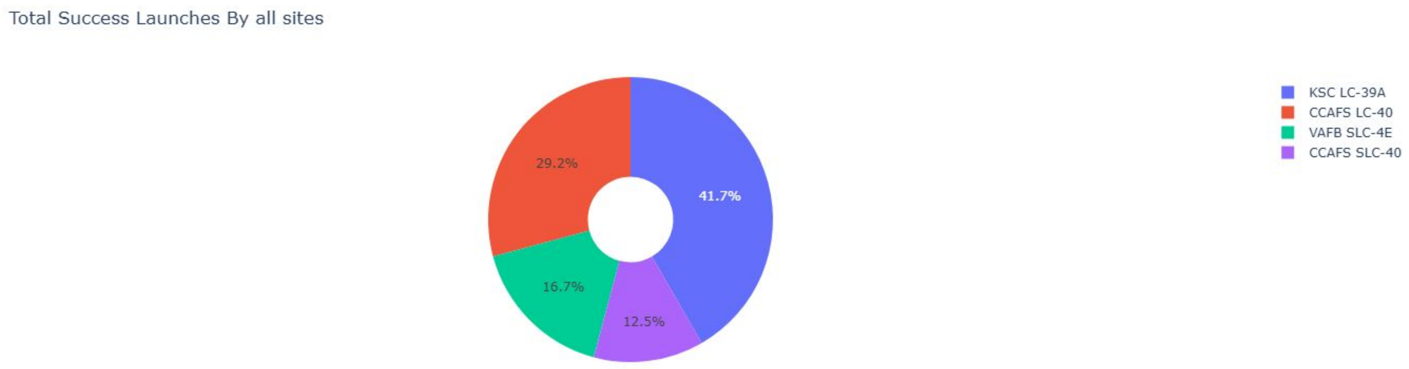
select count(Landing_Outcome) as Count, Landing_Outcome from SPACEXTBL where Date > '2010-06-04' and Date < '2017-03-20' Group by Landing_Outcome Order by Count DESC

Retrieving the rank, by number of attempts, of "Landing_Outcome" between 2010 and 2017

Count	Landing_Outcome
10	No attempt
5	Success (drone ship)
5	Failure (drone ship)
3	Success (ground pad)
3	Controlled (ocean)
2	Uncontrolled (ocean)
1	Precluded (drone ship)
1	Failure (parachute)

Visual Analytics

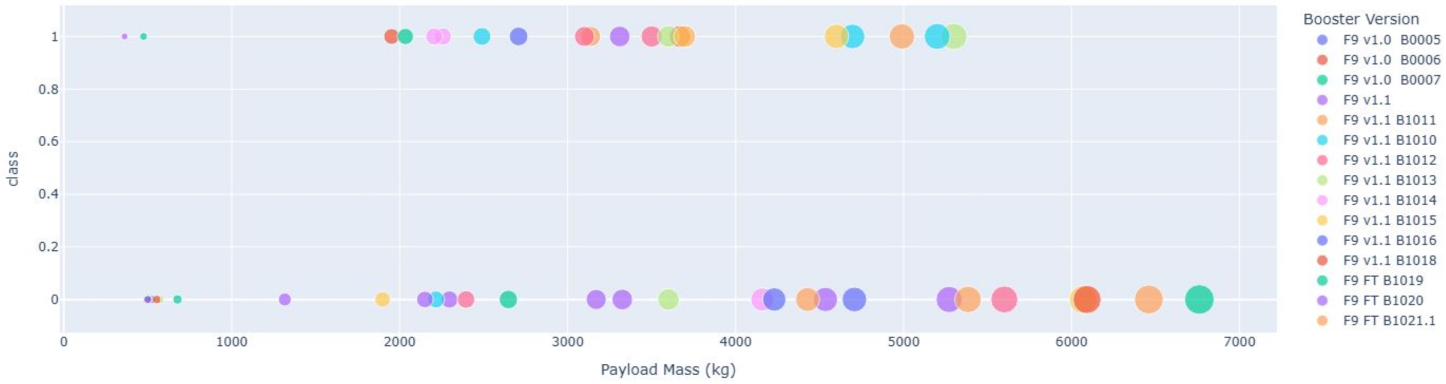
Dashboard - Pie chart showing the success launch percentage by each site



We can observe that KSC LC-39A was the most successful site

Visual Analytics

Dashboard - Scatter plot showing the payload mass vs launch outcome



We can observe that the success rate for lighter payloads is slightly higher than for heavier payloads

Predictive Analysis

Classification accuracy using training data

- The Tree algorithm was the one who performed better

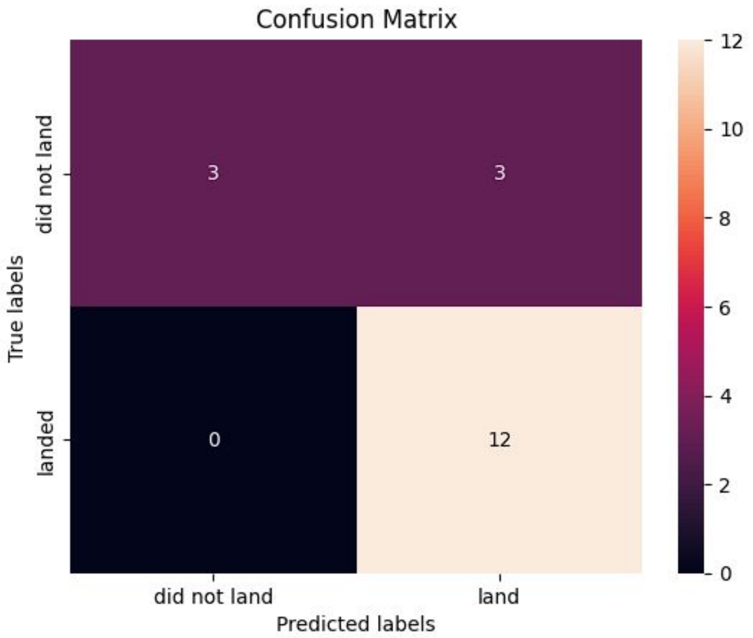
```
algorithms = {'KNN':knn_cv.best_score_,'Tree':tree_cv.best_score_,'LogisticRegression':logreg_cv.best_score_}
bestalgorithm = max(algorithms, key=algorithms.get)
print('Best Algorithm is',bestalgorithm,'with a score of',algorithms[bestalgorithm])
if bestalgorithm == 'Tree':
    print('Best Params is :',tree_cv.best_params_)
if bestalgorithm == 'KNN':
    print('Best Params is :',knn_cv.best_params_)
if bestalgorithm == 'LogisticRegression':
    print('Best Params is :',logreg_cv.best_params_)
```

```
Best Algorithm is Tree with a score of 0.8767857142857143
Best Params is : {'criterion': 'entropy', 'max_depth': 6, 'max_features': 'sqrt', 'min_samples_leaf': 2, 'min_samples_split':
2, 'splitter': 'random'}
```

Predictive Analysis

Confusion matrix for tree

- We can see that the main issue is false negatives



CONCLUSION

- The Tree Classifier Algorithm performed the best for machine learning for this dataset
- Lower weight payloads performed better
- The success rate has been increasing over the years
- Launch site KSC LC-39A had the most successful launches of all sites