IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

K.Sai Nandan Reddy
29-6-25

Outline

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

Executive Summary

Summary of methodologies:

- Dataset is loaded and the identified key features include FlightNumber, Payload Mass, Orbit, Launch Site, Launch pad.
- The categorical features are encoded using One Hot Encoder.
- All the features are standardized for better results.
- Models such as Logistic Regression,SVM,Decision Tree Classifier and KNN are used and compared for best results.
- Identified best tuned hyperparameters that gave us optimal outcomes.

Executive Summary

Summary of results:

- Logisitice regression model gave an accuracy of 84.6% for 10-fold CV.
- Supprt Vector Machine gave an accuracy of 84.8% for 10-fold CV.
- Decision Tree Classifier gave an accuracy of 87.5% for similar foldings.
- KNN model gave an accuracy of 84.8% with 10 foldings

Almost all models gave similar False positives except Decision tree model, it outperformed the other models in predicting with lesser False positives and better accuracy in predicting the successful launches.

Introduction

SpaceX advertises on its websites that the Falcon9 launches with a cost of 62 million dollars with others being 165.The ability to reuse the first stage by SpaceX accounts to its savings.The determination of successful first stage landing can give us an estimation on the cost required to launch.

The main objective of the problem is to predict the successful launch of SpaceX Falcon9 landings using supervised machine learning techniques which ultimately help companies in launch cost estimation.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - From SpaceX api and Wikipedia
- Perform data wrangling
  - Determine the target variable ‘Class’ using landing outcome
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Build, tune, evaluate and compare different classification models

Data Collection

Datasets were collected in two ways i.e., using SpaceX API & Wikipedia.

Used libraries such as requests to get url, Beautiful Soup to parse the html content from webpages and select the required tables using respective html tags.Formed and seperated only useful data from those api information and html content into our datasets and converted them into csv formats.This can now be used as a data frame with python pandas library and ready to perform analysis.

Data Collection – SpaceX API

- The data from api is decoded as json (.json()) and converted into pandas dataframe using .json_normalize()
- Used functions like getBoosterVersion, getLaunchSite etc on data for better construction of the dataset.
- Filtered data to include only Falcon9 version.Removed NULL values.

Github URL:

https://github.com/nandanreddy1014/Final-Project/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

Get data

Decode and normalize

Apply functions

Filter data

Handle missing values

Ready for analysis

Data Collection - Scraping

- Used requests and BeautifulSoup libraries.
- Extracted data using find_all on html tags like table,th.
- Used some functions to extract features from tables and append as a dictonary.
- Converted this to pandas dataframe.

Githhub URL:

https://github.com/nandanreddy1014/Final-Project/blob/main/jupyter-labs-webscraping.ipynb

Get html content

Parse to find tables

Extract features

Append into dictonary

Convert into dataframe

Ready for analysis

Data Wrangling

- Found out different launch sites and launch pads for the rocket launches.
- Assessed different landing outcomes from these launches.
- Seperated them into good outcomes and bad outcomes.
- Added the target column named ‘Class’ has two labels which are ‘1’ for successful launch and ‘0’ for a failure.

GitHub URL:

https://github.com/nandanreddy1014/Final-Project/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

Outcome	
True ASDS	41
None None	19
True RTLS	14
False ASDS	6
True Ocean	5
False Ocean	2
None ASDS	2
False RTLS	1

EDA with Data Visualization

- Visualized relationships between features such as FlightNumber, Payload Mass, Orbit type, Launch Site with scatter plots.
- Plotted yearly success rate using a line plot, Orbit type success rate with a bar chart.
- Found some valuable insights such as Launches with heavy payload are very less but mostly successful.Four Orbit types show 100% success.

GitHub URL:https://github.com/nandanreddy1014/Final-Project/blob/main/edadataviz.ipynb

EDA with SQL

- Found Total Payload Mass by different launch customers.
- Calculated average Payload Mass for Falcon9 booster.
- Found different launch sites.
- Listed out different types of launch outcomes.
- Returned the booster versions that carried maximum payload.
- Ranked all of the different outcomes of rocket launches.

GitHub URL: https://github.com/nandanreddy1014/Final-Project/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

Used Folium,a python library for geospatial visualizatoin that enhances maps with different objects such as markers,circles,lines,colors and give us

- Enhanced visual insights
- Launch and landing rerlationships
- User interactoin
- Geographic contexts
- Data validation
- Differentiation of objects

Build an Interactive Map with Folium

Markers

- Display basic information about launch site via popups

Circles

- Visually emphasize launch site coverage or area of interest

Polylines (lines)

- Visualize trajectories between launch site and recovery

Color coding

- Different marker colors to indicate different information

Build a Dashboard with Plotly Dash

Made a dashboard using Plotly Dash,Python libraries to visulize different components and valuable insights. Dashboard integrated multiple plots & interactions to summarize SpaceX launch activity.

Key elements include:

Dropdown menus

- Allow selection of launch sites for focused analysis

Pie chart

- Success vs Failure distribution
- Provides insights into mission outcome proportions

Build a Dashboard with Plotly Dash

Scatter plot

- Payload Mass vs Launch Outcome
- Highlights how payload weights impact the mission success

Range Slider

- Lets us filter payload payload mass range for scatter plots

These plots help us identify top performing launch sites, offer a quick glance at mission outcomes, targeted payload filtering, reveal patterns across payload ranges, launch site specific information, an interactive and intuitive environment for exploring SpaceX mission data.

Predictive Analysis (Classification)

Everything in the predictive analysis is associated with scikit-learn, a Python library that sees its best usage to build machine learning models.

Data preprocessing was done which include standardizing all features so that no feature show bias and categorical data is converted to numerical for the model to understand relations.

It is important to analyse with different machine learning techniques and build different models to get best results and predictions.

Data is split into training and testing parts where training data is used to fit models and testing data is used to find accuracy of model.

Used a GridSearchCV library from scikit-learn to apply validation by iteratively selecting train-test split data with a cv=10 foldings which gives us the best results.

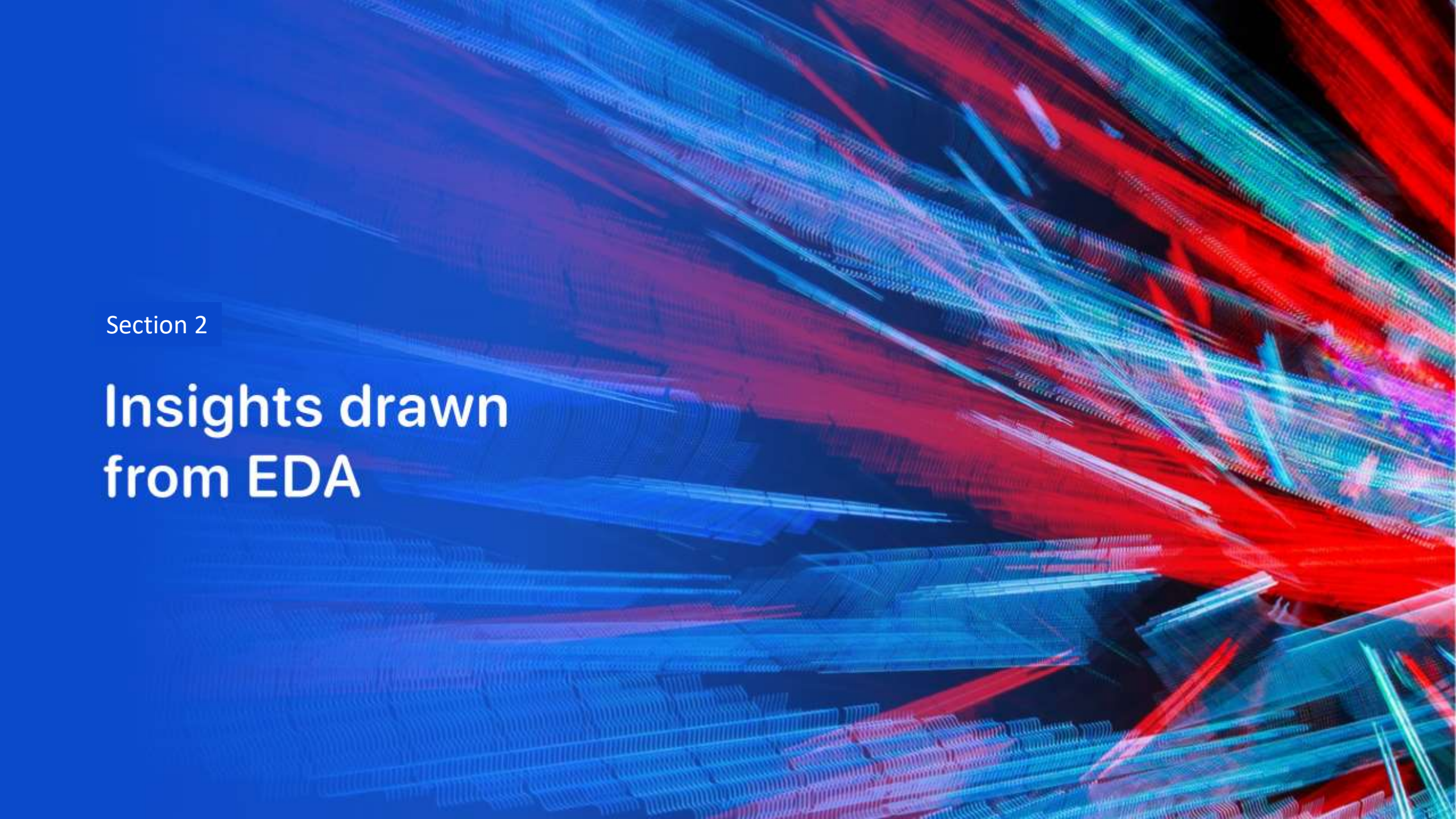
Results

Exploratory data analysis results

- Orbit types such as ES_L1,GEO,HEO,SSO have 100% success rate.
- The payload mass for above orbit types to be low can also be observed.
- The success rate kept increasing since 2013 till 2020.

Predictive analysis results

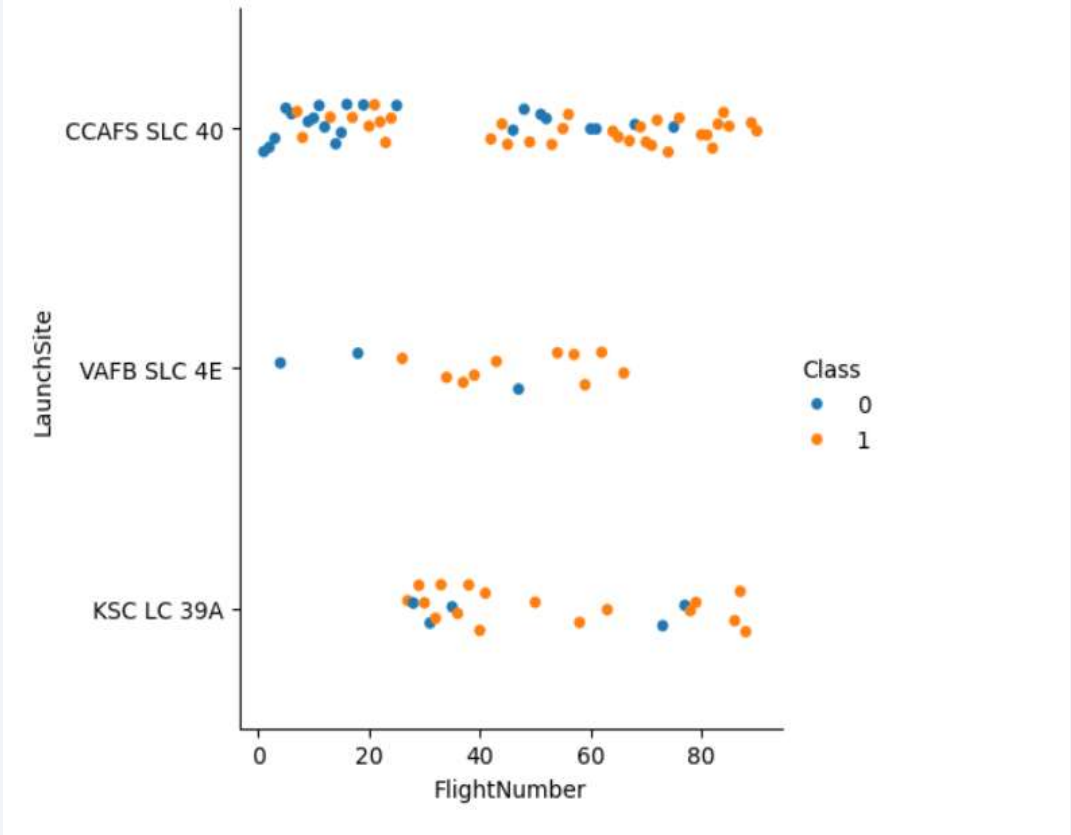
- Logistic regression,SVM,KNN models have performed almost similar.
- Their accuracies are same around 84% and false positive rate is 0.2.
- Decision Tree Classifier outperformed with better accuracy and less false positive rate.


Section 2
Insights drawn
from EDA

Flight Number vs. Launch Site

Scatter plot of Flight Number vs Launch site

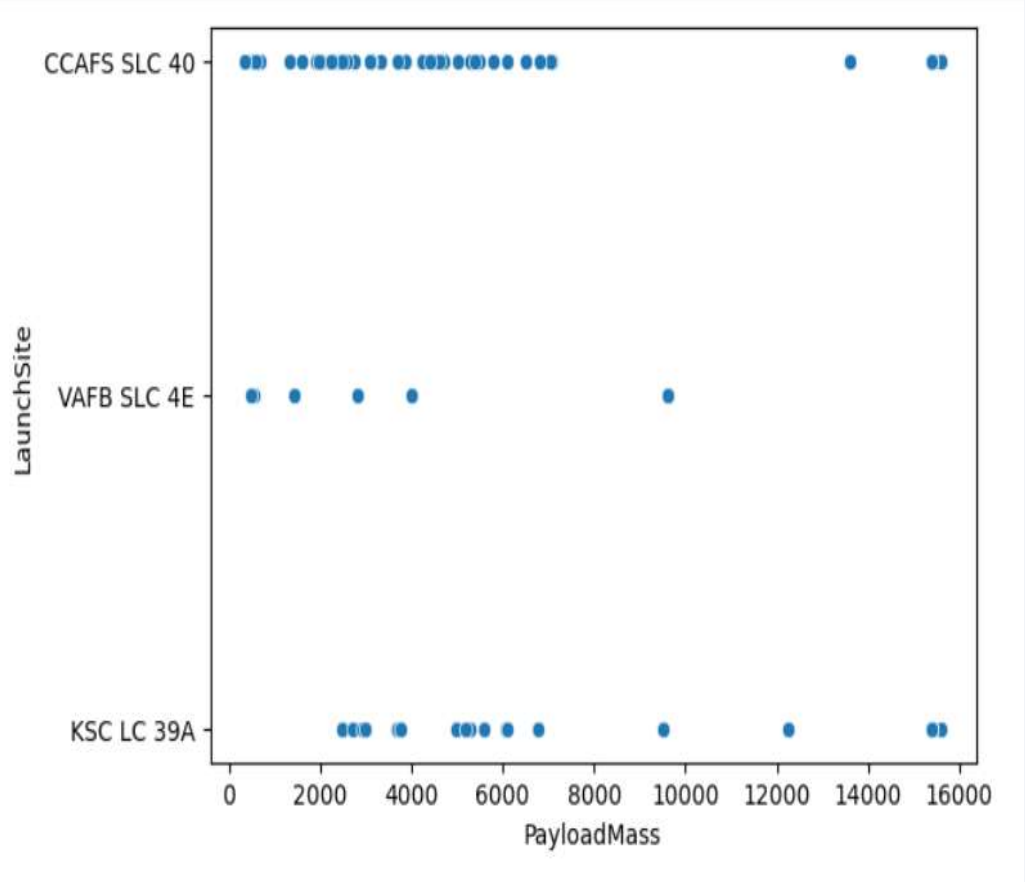
As flight number increased, success rate also increased and CAFS-SLC 40 have launched more rockets than the other combined and KSC LC 39A started their missions late compared to other



Payload vs. Launch Site

Scatter plot of Payload vs. Launch Site

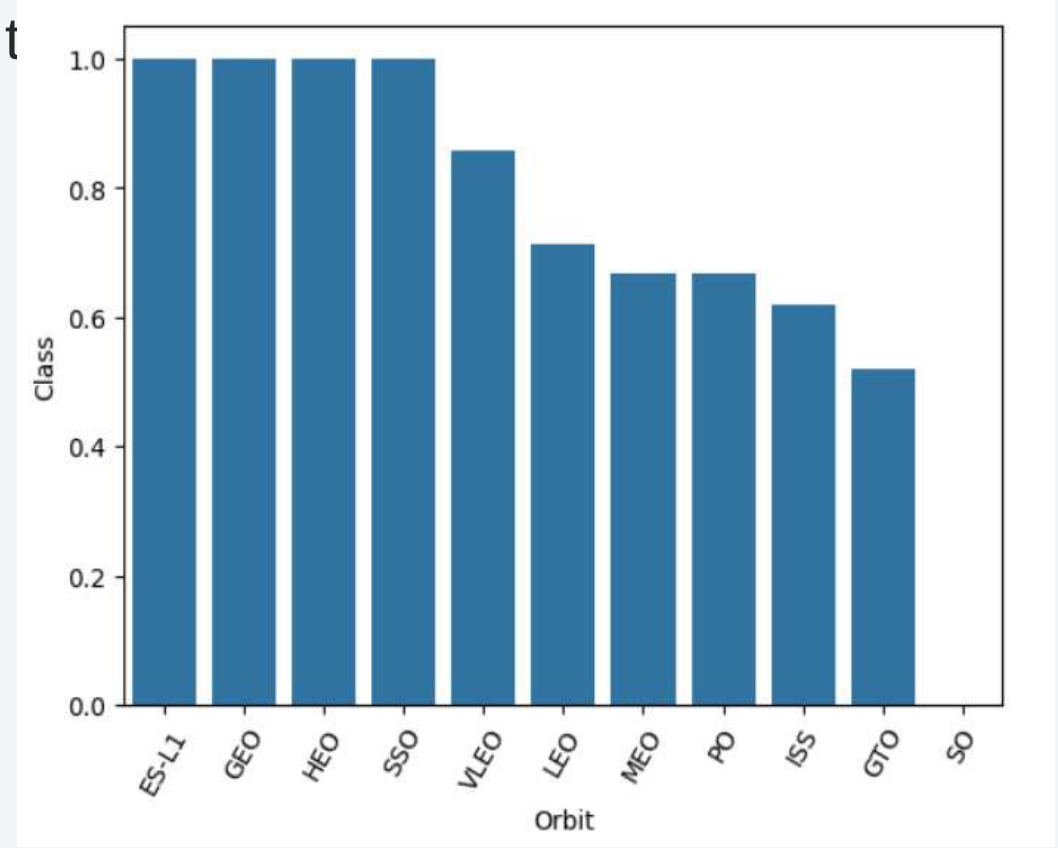
Many rocket carry payload mass less than 8000.Average payload mass for site CCAFS-SLC 40 is around 4000 and site VAFB-SLC 4E does not have any rockets launched for heavy payload mass greater than 10000.



Success Rate vs. Orbit Type

Bar chart for the success rate of each orbit t

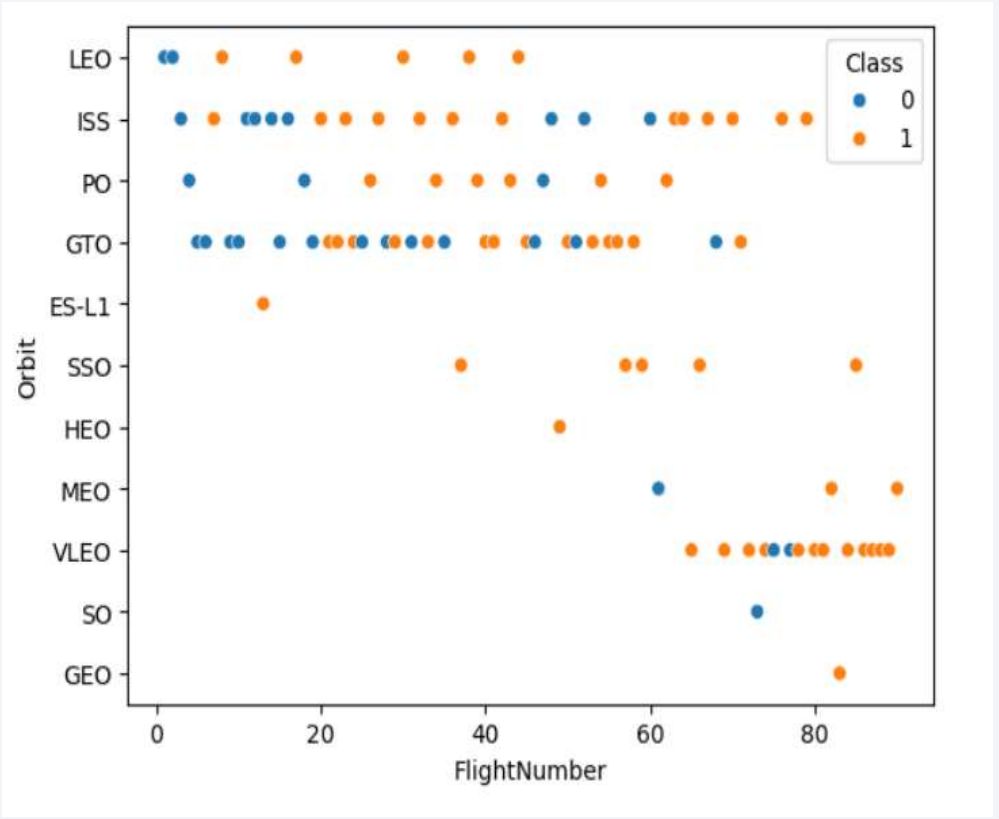
Clearly it is observed that 4 orbits have 100% success rate while one orbit have no success and the average success is around 0.65



Flight Number vs. Orbit Type

Scatter point of Flight number vs. Orbit

Most of the rockets were launched to 4 orbits LEO,ISS,PO,GTO while only one was launched to GEO tells us it is very far from surface of Earth.It can be observed that LEO success rate increased with no.of flights,conversely there is no relationship in GTO

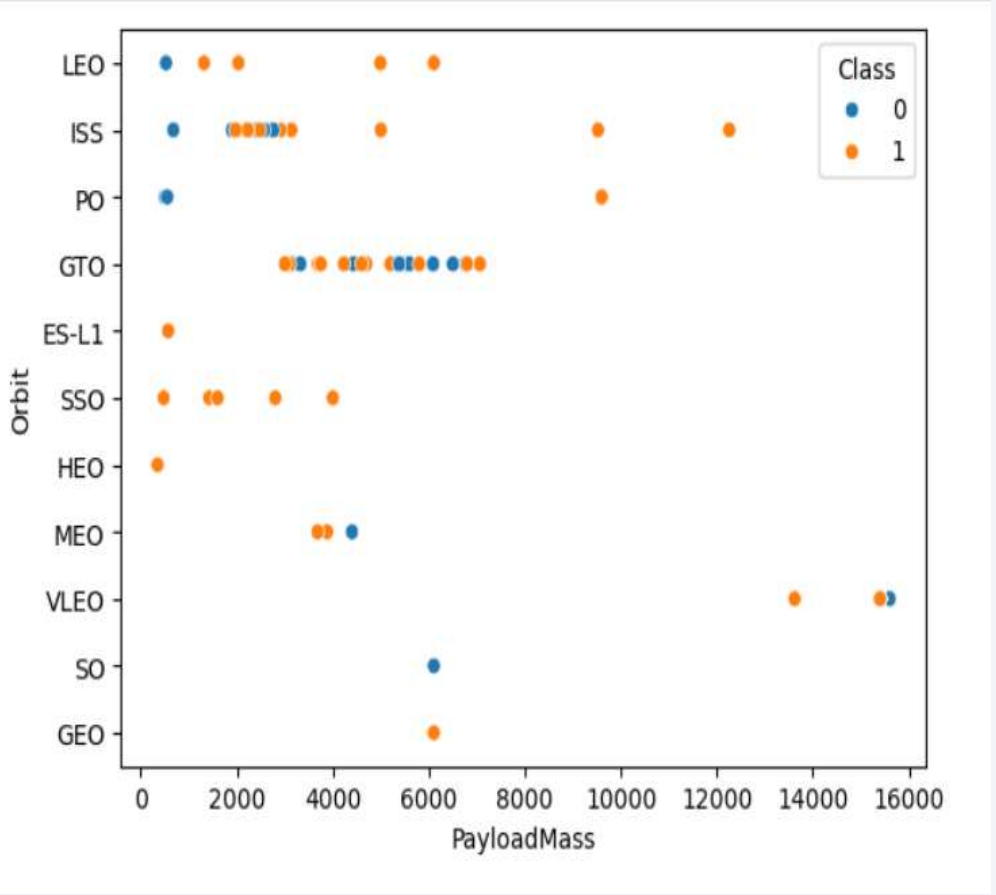


Payload vs. Orbit Type

Scatter point of payload vs. orbit type

With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.

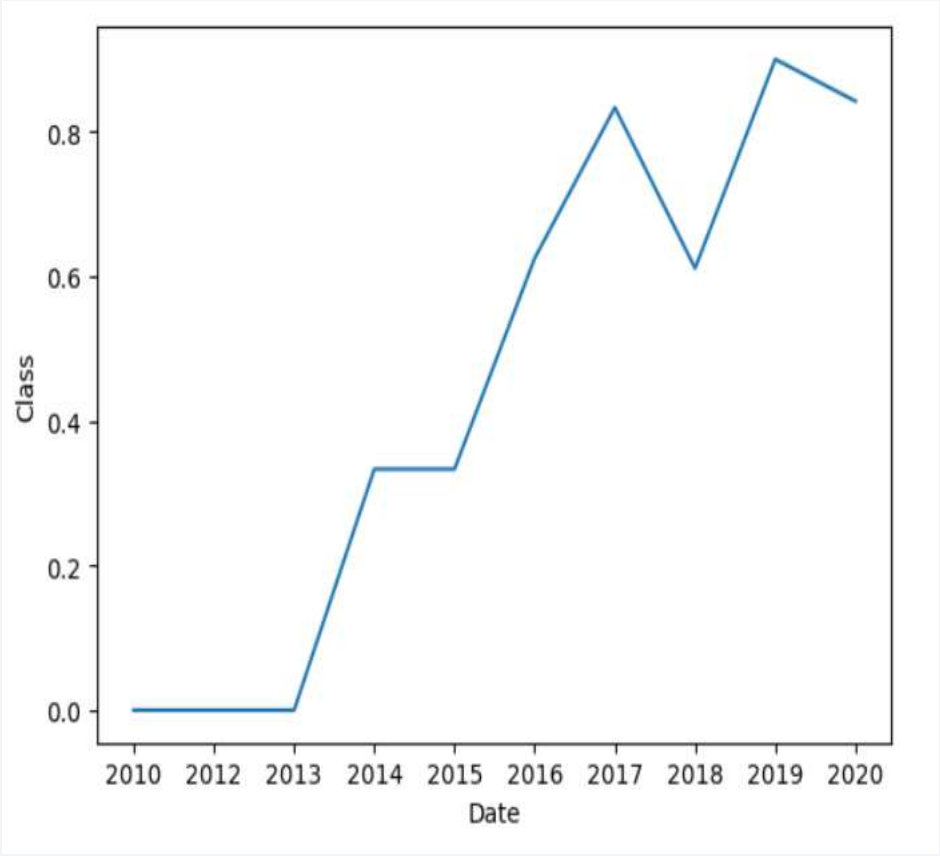
However, for GTO, it's difficult to distinguish between successful and unsuccessful landings as both outcomes are present.



Launch Success Yearly Trend

Line chart of yearly average success rate

A three year period saw no success and then the success rate kept on increasing since 2013 till 2020 with a single dip in 2018.



All Launch Site Names

- Distinct key word is used to find the launch sites from the table.
- It was found that there are 4 different launch sites.
- They are CCAFS LC-40,VAFB SLC-4E,KSC LC-39A, CCAFS SLC-40

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

```
%sql select distinct Launch_Site from SPACEXTBL
```

Launch Site Names Begin with 'CCA'

To retrieve results that have similar pattern ‘like’ key word is used.

```
%sql select * from SPACEXTBL where Launch_site like 'CCA%' limit 5
```

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Out
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (para
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 (para
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No at
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No at
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No at

Total Payload Mass

- The total payload mass carried by boosters launched by NASA are found to be 45596 kgs
- Aggregated function(sum() here) are used for mathematical operations.
- Where clause is used to filter required results.

total
45596

```
%sql select sum(PAYLOAD_MASS__KG_) as total from SPACEXTBL where Customer = 'NASA (CRS)'
```

Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1 was found to be ~ 2500 kgs.
- The aggregated function avg() is used to calculate the average with filtering booster versions like F9 v1.1

average_v1.1
2534.6666666666665

```
%sql select avg(PAYLOAD_MASS__KG_) as 'average_v1.1' from SPACEXTBL where Booster_Version like 'F9 v1.1%'
```

First Successful Ground Landing Date

- The date of the first successful landing outcome on ground pad was 22-12-2015.
- The data was first filtered to ground pad success using where clause and then order by Date to get first successful landing.
- Limit key word was used to retrieve only one result which is the first successful landing.

Date
2015-12-22

```
%sql select Date from SPACEXTBL where Landing_Outcome='Success (ground pad)' order by Date limit 1
```

Successful Drone Ship Landing with Payload between 4000 and 6000

- Results were obtained including the successfully landed on drone ship and had payload mass greater than 4000 but less than 6000 conditions for where clause.

```
%sql select Booster_Version,Landing_Outcome,PAYLOAD_MASS__KG_ from SPACEXTBL where Landing_Outcome = 'Success (drone ship)'
and PAYLOAD_MASS__KG_ > 4000 and PAYLOAD_MASS__KG_ < 6000
```

Booster_Version	Landing_Outcome	PAYLOAD_MASS__KG_
F9 FT B1022	Success (drone ship)	4696
F9 FT B1026	Success (drone ship)	4600
F9 FT B1021.2	Success (drone ship)	5300
F9 FT B1031.2	Success (drone ship)	5200

Total Number of Successful and Failure Mission Outcomes

- The total number of successful and failure mission outcomes were found by grouping all different mission outcomes.
- The aggregated function count() is used to get the total count of each outcome.

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

```
%sql select Mission_Outcome,count(*) as count from SPACEXTBL group by Mission_Outcome
```

Boosters Carried Maximum Payload

- The concept of sub queries is used to query the results.
- First query gets the maximum payload mass.
- Second query gets all booster versions that carry the maximum amount of payload mass.

```
%sql select Booster_Version,PAYLOAD_MASS__KG_ from SPACEXTBL
   from SPACEXTBL where PAYLOAD_MASS__KG_ =
(select max(PAYLOAD_MASS__KG_) from SPACEXTBL)
```

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

2015 Launch Records

- List of the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

```
%sql select substr(Date, 6, 2) AS Month,Landing_Outcome,Booster_Version,Launch_Site FROM SPACEXTBL
        WHERE Landing_Outcome LIKE 'Failure (drone ship)' AND substr(Date, 1, 4) = '2015'
```

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

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

- 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

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

Landing_Outcome	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

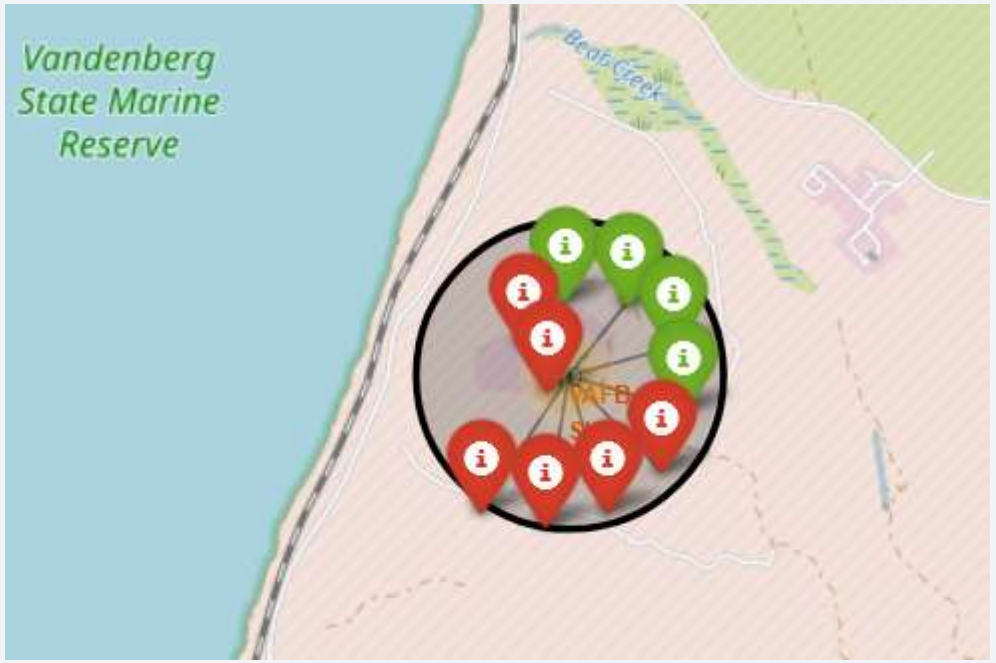
Launch Sites Proximities Analysis

Launch Sites across USA

- One clear finding is that all sites are on the coasts, close to oceans.
- Three locations are clustered in Florida state and one is in California state.
- All launch sites are close to equater.

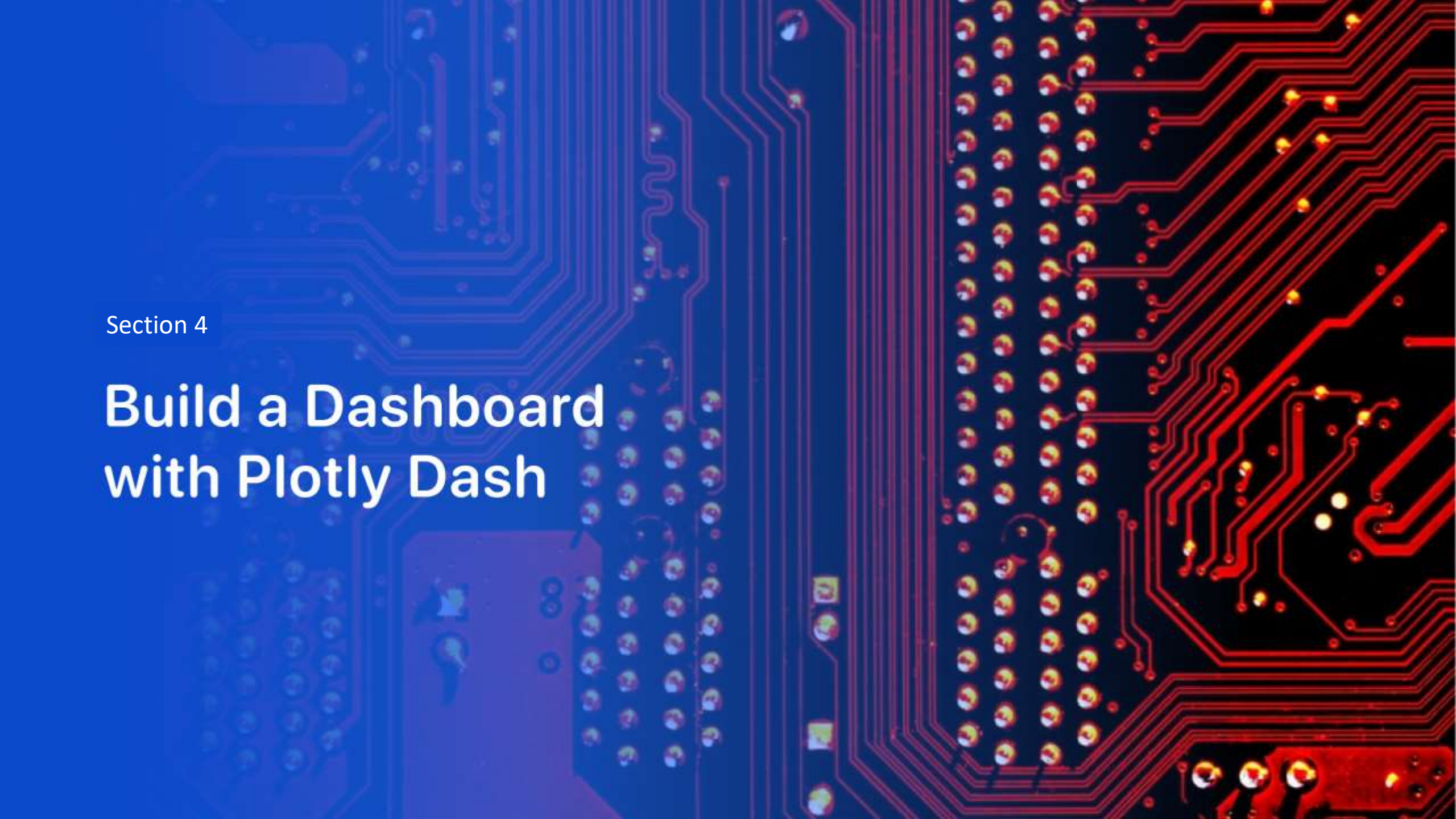
Launch Outcome comparison

- The figure shows launch outcomes at a site on the map.
- Green corresponds successful launch and red corresponds to a failure
- It is clear that the site had more failed launches than successful ones
- Map object color is used to interpret different outcomes



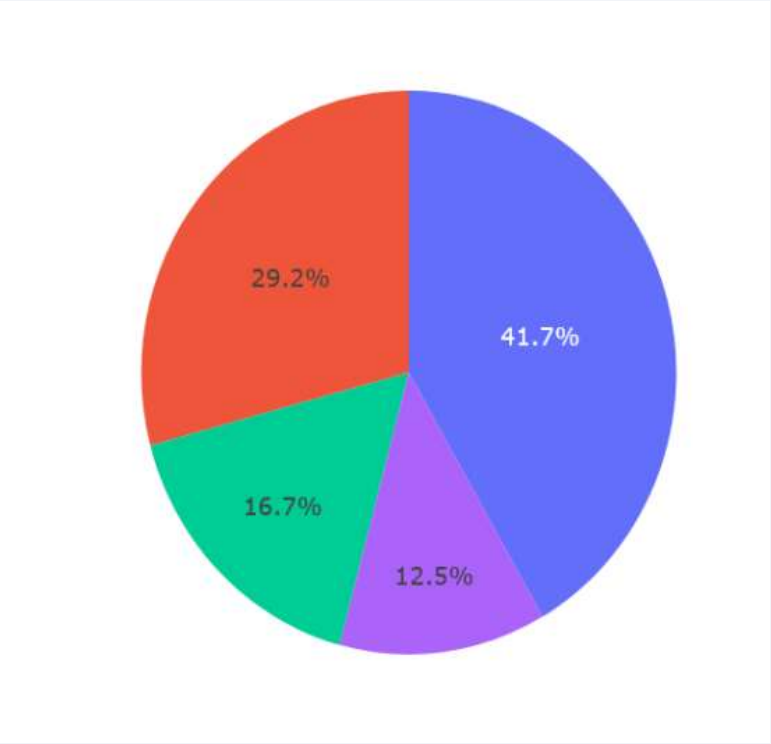
Section 4

Build a Dashboard with Plotly Dash



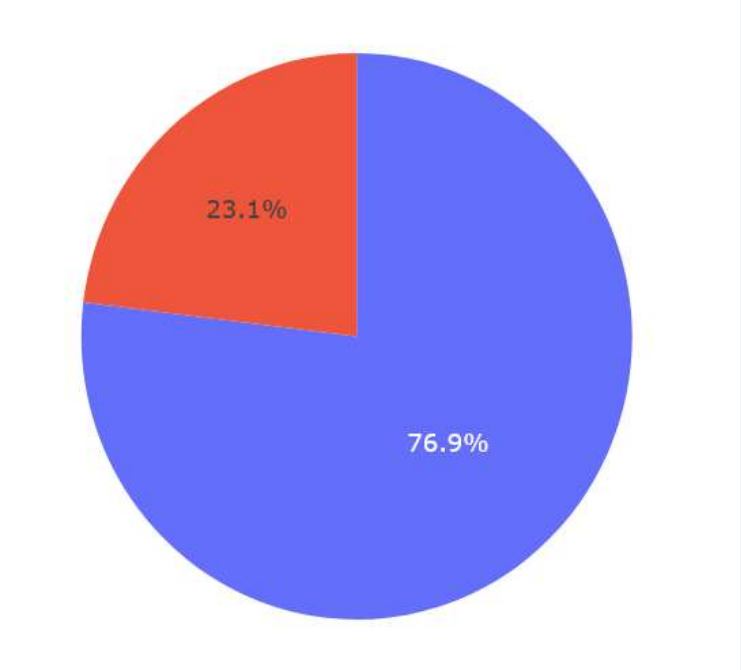
Launch Site Success rate(All Sites)

- Launch site KSC LC-39A has the highest of success rate 41.7% which is depicted in the pie chart with blue color(not individual)
- Launch site CCAFS SLC-40 has the lowest with only 12.5% which is showin using purple color(not individual)

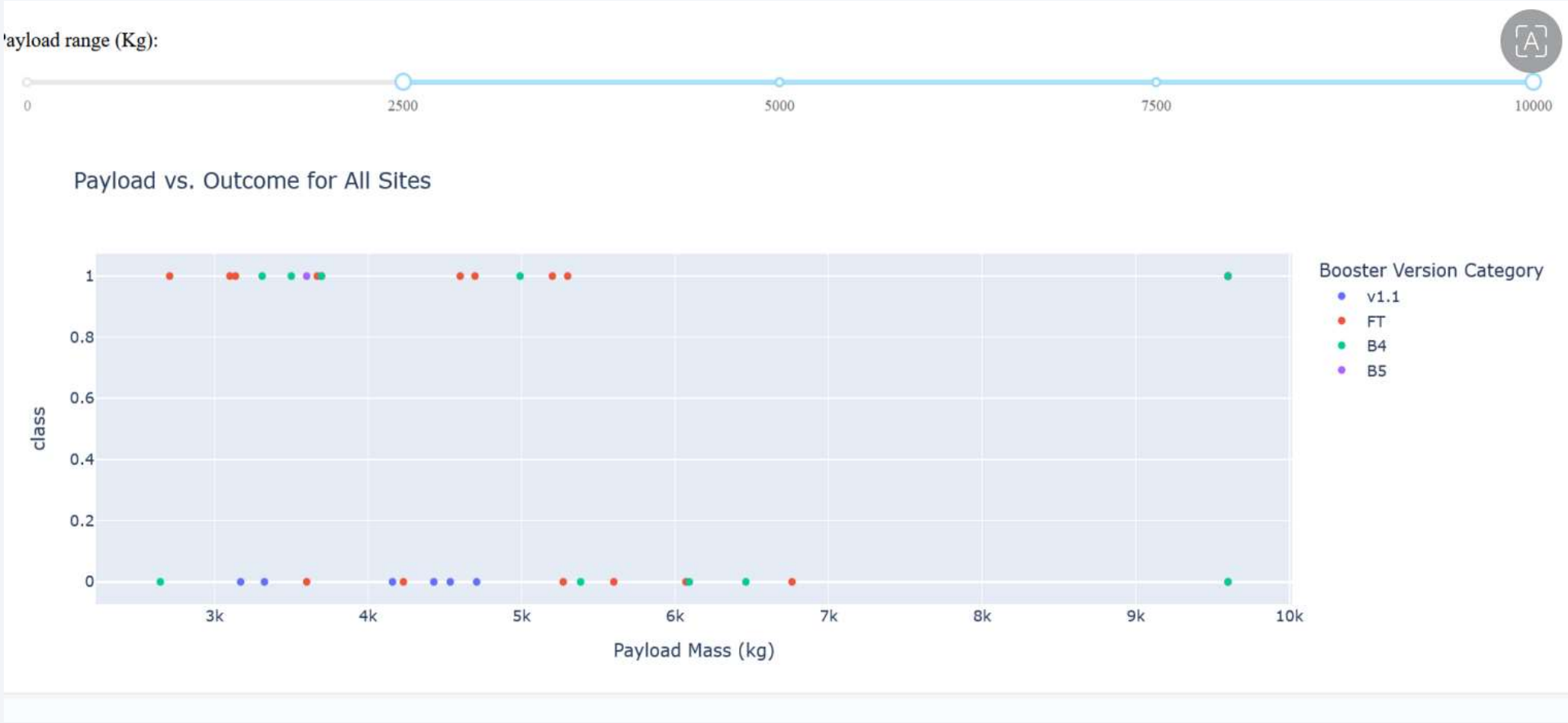


Most successful Launch Site

- Launch site KSC LC-39A has the highest individual success rate of launches with 76.9%.
- Launch site CCAFS LC-40 comes close with the second highest rate of about 73%.



Payload vs. Launch Outcome



Payload vs. Launch Outcome

- Booster Version v1.1 has 0 successful launches with payload heavier than 3000 kgs.
- Booster Version FT has more faliure launches than successful ones with weights heavier than 5000 kgs.
- Booster Version B5 have only one successful launch with payload mass greater than 3000 kgs.
- Booster Version B4 have mixed outcomes upto 5000 kgs and mostly failure above that range.
- There are no payloads in the range 7000-9000 kgs.
- Success rate of all launches drops with heavy payload masses.

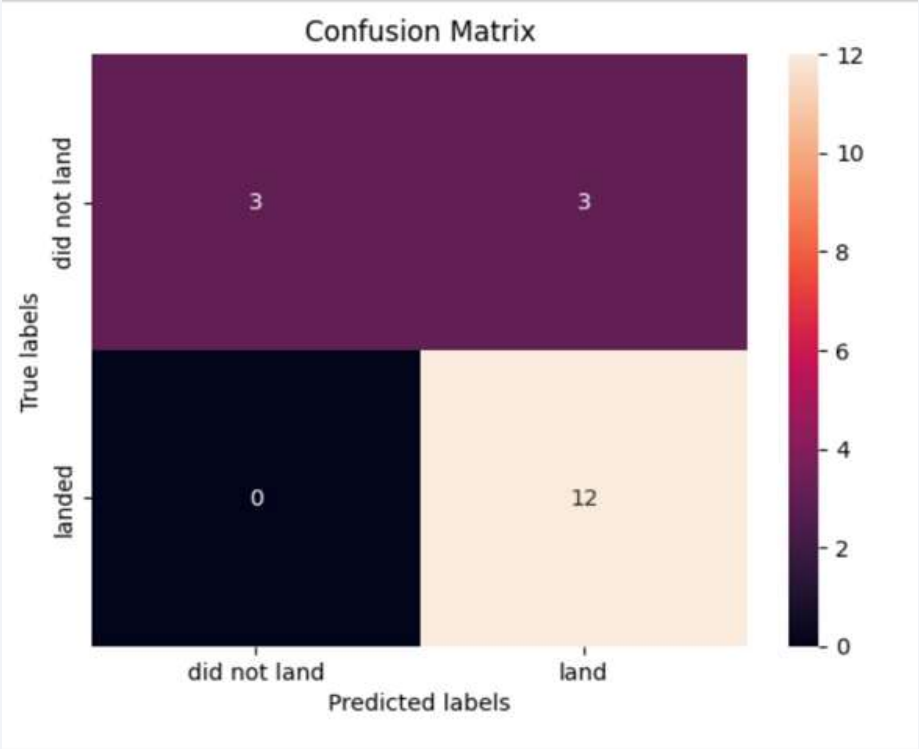
Section 5

Predictive Analysis (Classification)

Classification Accuracy

Logistic Regression

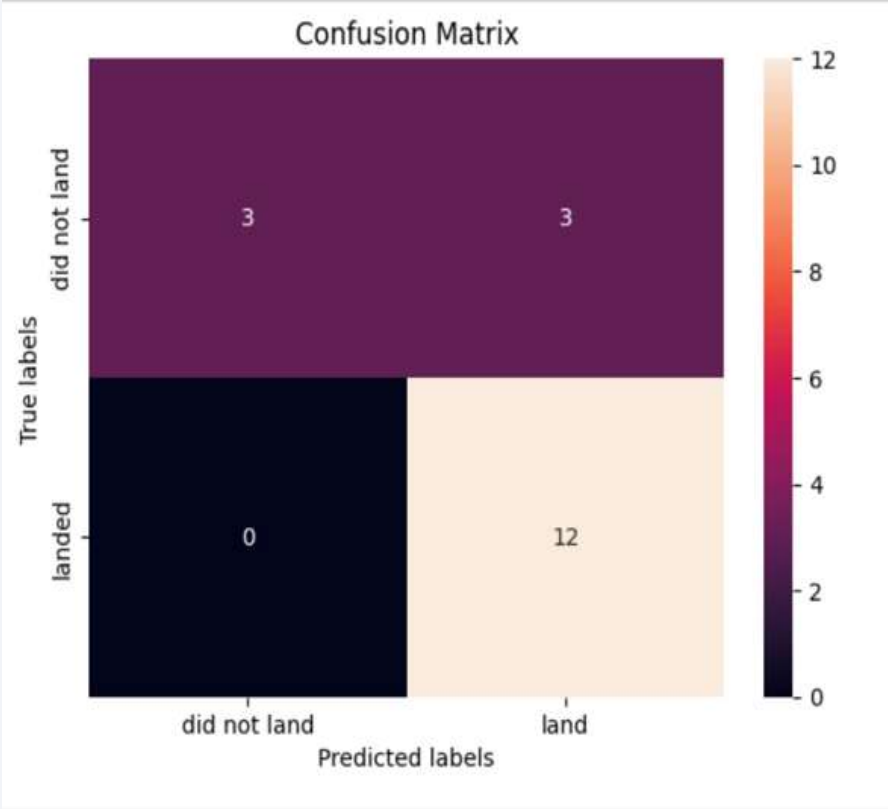
The model predicted with an accuracy of 83% on unseen data and although it differentiated well between classes, there are 3 false positives.



Classification Accuracy

Support Vector Machines

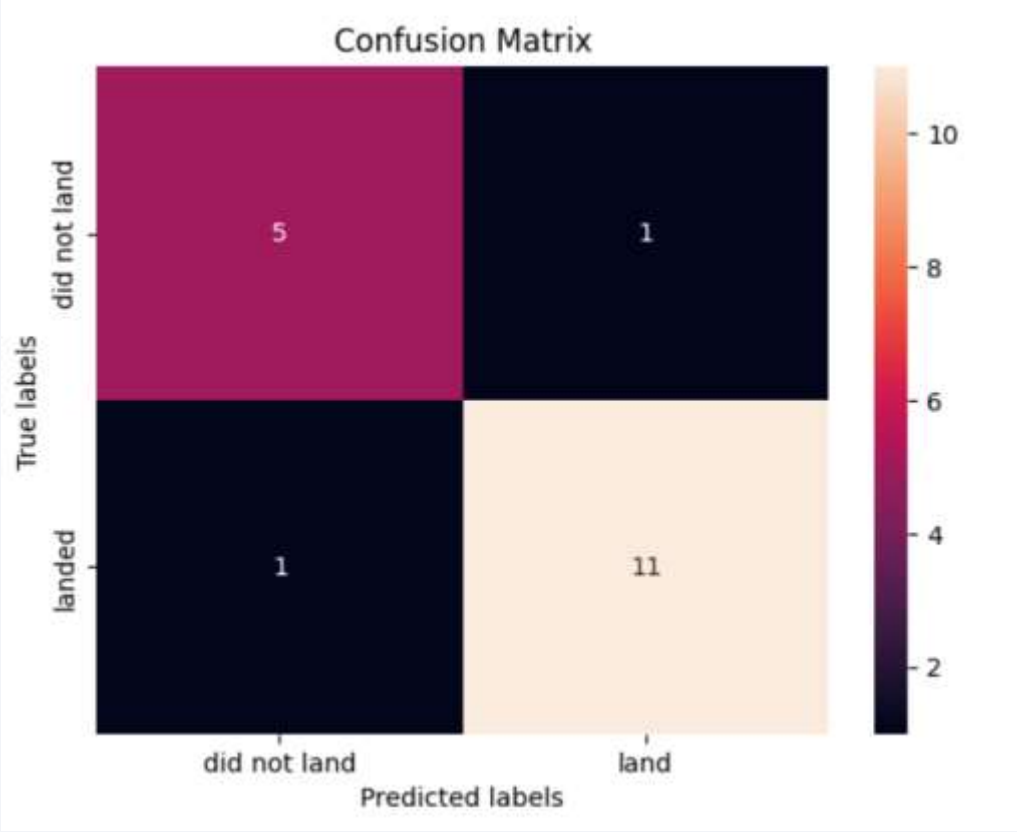
This model also predicted with the same accuracy around 83% on test data and also performed similar in class differentiaion with 3 false positives.



Classification Accuracy

Decision Tree Classifier

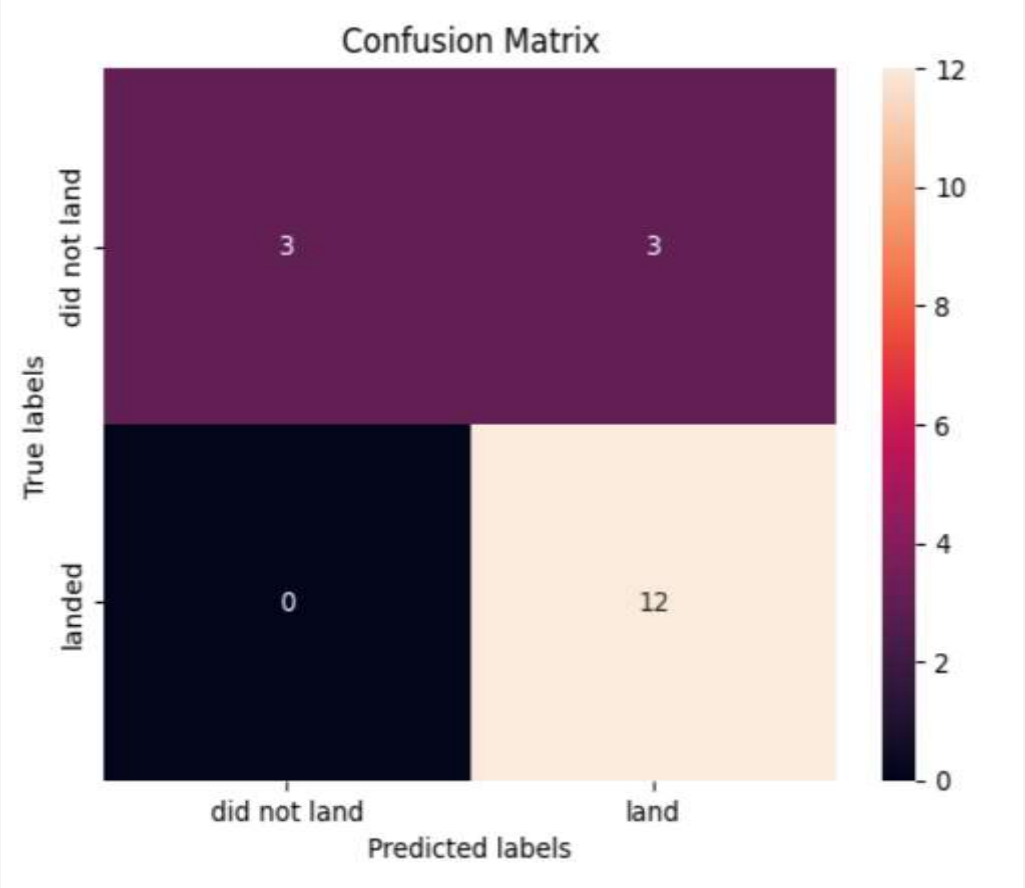
This model predicted better with increased accuracy about 89% on new data and also performed better in differentiating between outcomes with lower false positives.



Classification Accuracy

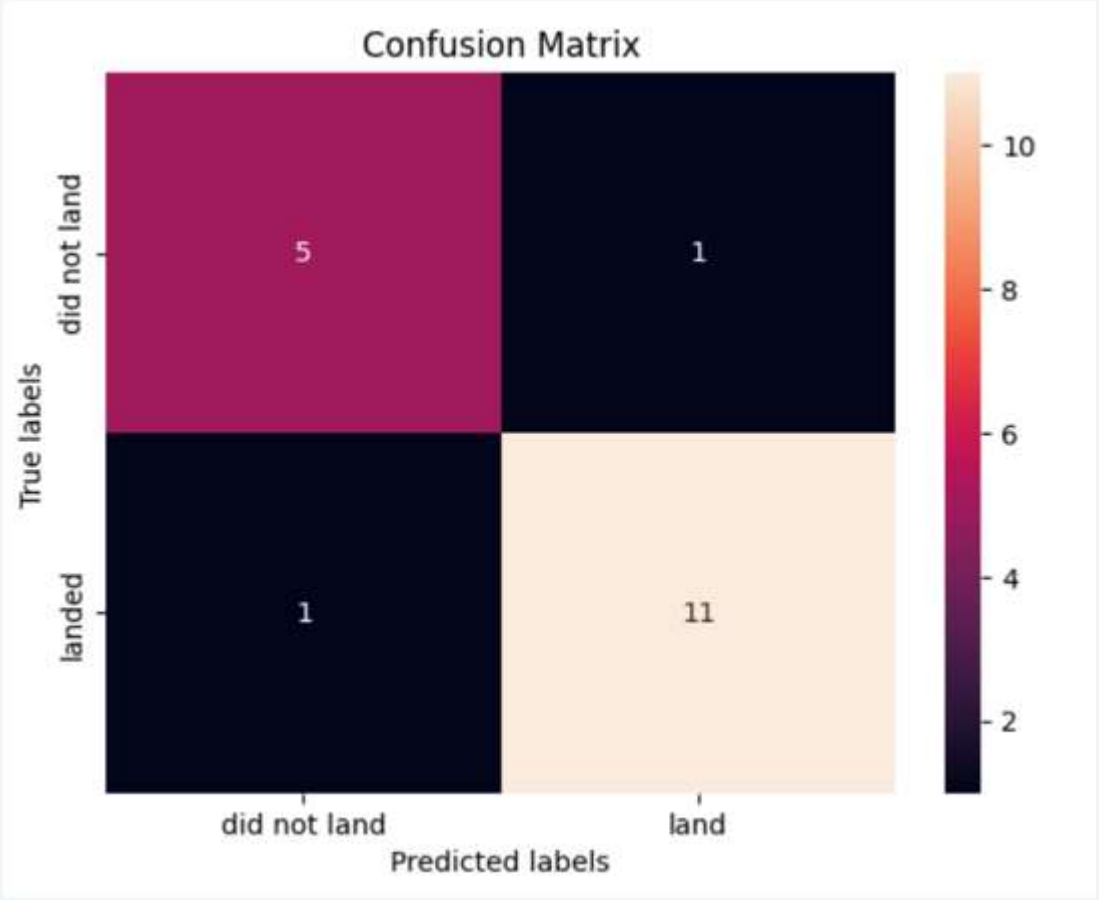
K-Nearest Neighbours

This model also predicted with the same accuracy around 83% on test data and also performed similar in class differentiaion with 3 false positives.



Confusion Matrix

This is the confusion matrix of Decision Tree Classifier that performed better compared to other models with more accuracy and lower false positives.



Conclusions

- Features like Orbit,Payload Mass,Launch Site,Launch Pad are used the best predictors in successful landings.
- Average Payload Mass for successful landing is 2000-5000 kgs.
- The success rate of rocket launch kept increasing and is around 67%.
- Decision Tree Classifier shown best performance in predictions.
- Different tuned hyperparameter are used but overall accuracy is 83%.
- Successful landings in first stage can reduce the cost of rocket launches.

Appendix

Wikipedia page link:

https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches

Dataset used to train the model:

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial
0	1	2010-06-04	Falcon 9	6104.959412	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0003
1	2	2012-05-22	Falcon 9	525.000000	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0005
2	3	2013-03-01	Falcon 9	677.000000	ISS	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0007
3	4	2013-09-29	Falcon 9	500.000000	PO	VAFB SLC 4E	False Ocean	1	False	False	False	NaN	1.0	0	B1003
4	5	2013-12-03	Falcon 9	3170.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B1004

Thank you!