IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Warodom Phungjununt
March 2024

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion

Executive Summary

Summary of methodologies

Our methodology encompasses several key processes: Initially, we gather data through web scraping and then manipulate this data via an API. This is followed by data sampling and addressing any missing values. We conduct exploratory data analysis and visualize the data for insights. Additionally, we utilize SQL for data querying. For predictive analytics, we employ machine learning algorithms including Support Vector Machines (SVM), Logistic Regression, Decision Trees, and K-Nearest Neighbors (KNN).

Summary of all results

The outcomes will produce machine learning models adept at forecasting the success of a launch, alongside significant data impacting these results. This includes optimal launch locations and whether the payload's mass is associated with the likelihood of success.

Introduction

We have entered the era of commercial space exploration, where companies are democratizing access to space travel. Virgin Galactic offers journeys beyond Earth's atmosphere with its suborbital flights. Rocket Lab focuses on deploying small satellites into orbit. Blue Origin produces both sub-orbital and orbital rockets that can be reused. Leading the pack, SpaceX has made significant strides, including delivering cargo to the International Space Station, launching the Starlink internet constellation to provide global internet coverage, and conducting crewed missions into space. A pivotal aspect of SpaceX's success is its cost-effective approach to space launches. The company promotes its Falcon 9 rocket launches at a price of $62 million on its website, a figure considerably lower than the $165 million or more charged by other providers. This cost efficiency is largely due to SpaceX's innovative reuse of the rocket's first stage.

Our task involves calculating the cost of each launch. To achieve this, we will collect data related to SpaceX and develop dashboards for our team's use. Furthermore, we will assess whether SpaceX intends to reuse the first stage of their rockets. Rather than applying complex rocket science to predict the successful landing of the first stage, we will employ a machine learning model, leveraging publicly available information, to foresee SpaceX's reuse of the first stage.

Section 1

Methodology

Methodology

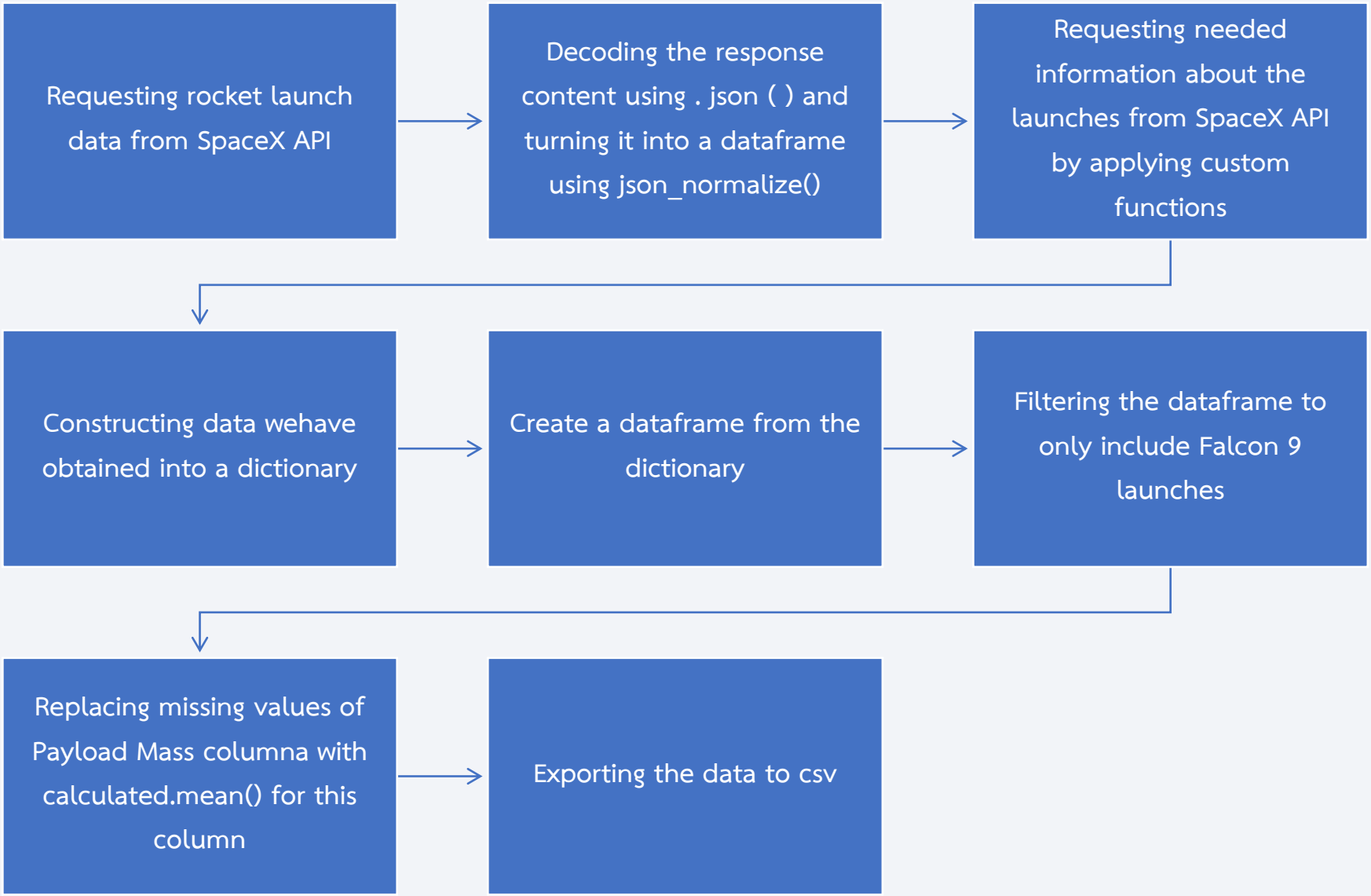
Executive Summary

- Data collection methodology:
  - API was used to collect the data (SpaceX REST API).
  - Data is about launches, payload delivered, launch and landing specifications, and the landing outcome.
- Perform data wrangling
  - Wrangling Data using an API, Sampling Data, and Dealing with Nulls.
- 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

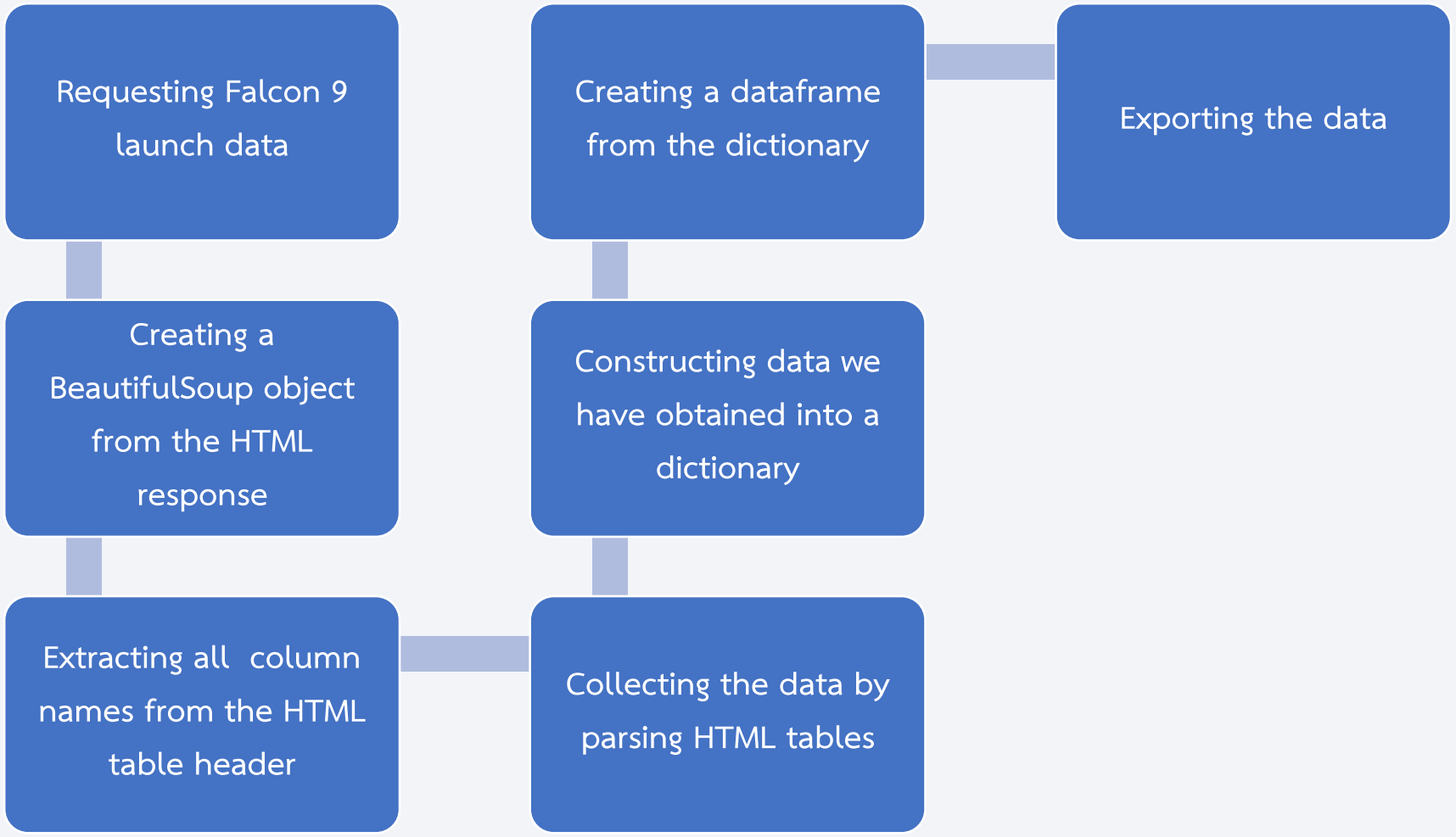
Data Collection

Data gathering was conducted through an API, namely the SpaceX REST API. This particular API will supply us with detailed information on launches, covering aspects such as the rocket utilized, payload details, launch parameters, landing criteria, and the results of the landing.

Data Collection – SpaceX API

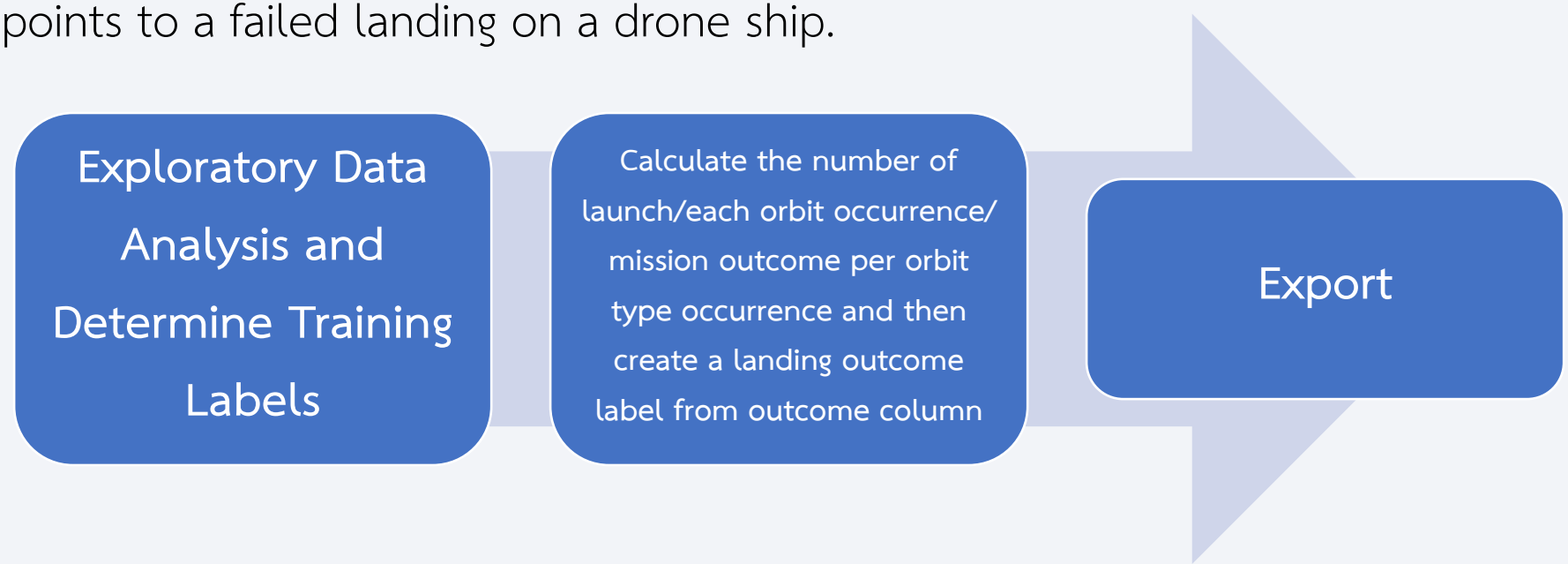


Data Collection - Scraping



Data Wrangling

Within the dataset, various instances exist where the booster failed to land successfully. In some cases, a landing attempt was made but resulted in failure due to mishaps; for instance, "True Ocean" indicates a successful landing in a designated ocean area, whereas "False Ocean" signifies a failed attempt to land in a specified ocean region. Similarly, "True RTLS" denotes a successful return and landing on a ground pad, whereas "False RTLS" indicates an unsuccessful attempt to land on a ground pad. "True ASDS" reflects a successful landing on a drone ship, while "False ASDS" points to a failed landing on a drone ship.



EDA with Data Visualization

Scatter	Bar	Line
• Flight Number vs. Payload Mass • Flight Number vs. Launch Site • Payload Mass vs. Launch Site • Flight Number vs. Orbit Type • Payload Mass vs. Orbit Type	• Orbit Type vs. Success Rate	• Launch Success Rate Yearly Trend

EDA with SQL

- Display the names of the unique launch sites in the space mission
- Display 5 records where launch sites begin with the string 'CCA'
- Display the total payload mass carried by boosters launched by NASA (CRS)
- Display average payload mass carried by booster version F9 v1.1
- List the date when the first succesful landing outcome in ground pad was acheived.
- List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- List the total number of successful and failure mission outcomes
- List the names of the booster_versions which have carried the maximum payload mass.
- 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.
- 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.

Build an Interactive Map with Folium

Markers of all Launch Sites:

• Added Marker with Circle, Popup Label and Text Label of NASA Johnson Space Center using its latitude and longitude coordinates as a start location.

• Added Markers with Circle, Popup Label and Text Label of all Launch Sites using their latitude and longitude coordinates to show their geographical locations and proximity to Equator and coasts.

Coloured Markers of the launch outcomes for each Launch Site:

• Added coloured Markers of success (Green) and failed (Red) launches using Marker Cluster to identify which launch sites have relatively high success rates.

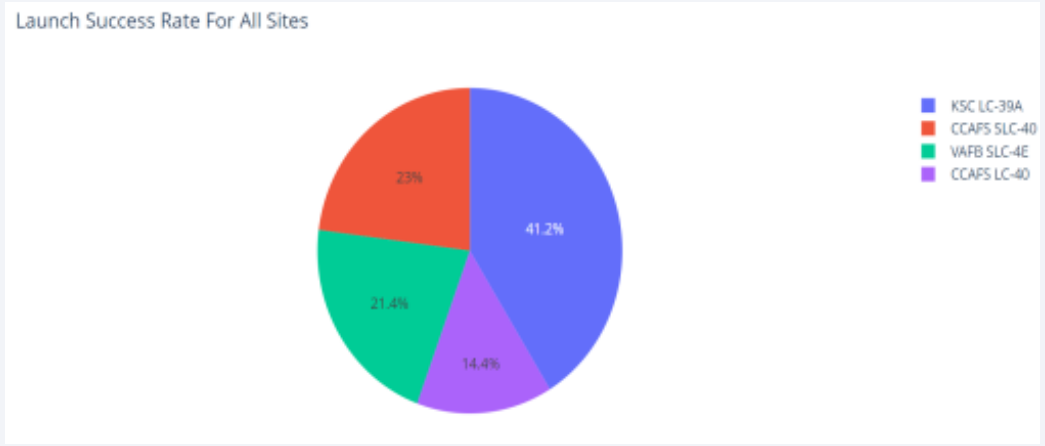
Distances between a Launch Site to its proximities:

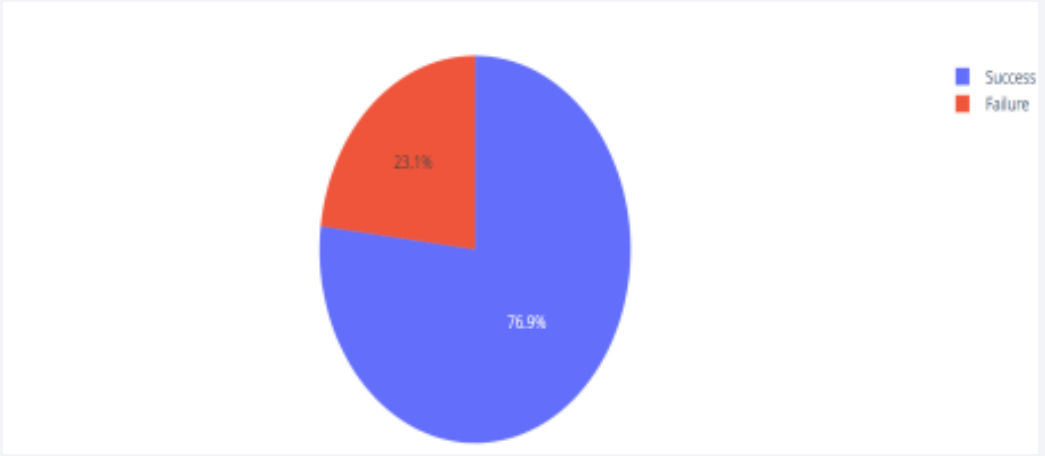
• Added coloured Lines to show distances between the Launch Site KSC LC-39A (as an example) and its proximities like Railway, Highway, Coastline and Closest City.

Build a Dashboard with Plotly Dash

Kennedy Space Center Launch Complex 39A (KSC LC-39A) field was more successful with a success rate of 76.9% and a failure rate of 23.1%.

The Cape Canaveral Launch Complex 40 (CAFS LC-40) obtained a higher failure rate with 73.1 and only success of 26.9%.





Predictive Analysis (Classification)

Load and Prepare Data

Step1: load data

Step2: Extract 'Class' column as Y

Step3: Standardize feature variables as X

Step4: Split data into train and test sets

Model Selection and Evaluation

Step1: Iterate over each machine learning algorithm:

1. Logistic Regression
2. Support Vector Machine (SVM)
3. Decision Tree
4. K Nearest Neighbors (KNN)

Step2: For each algorithm:

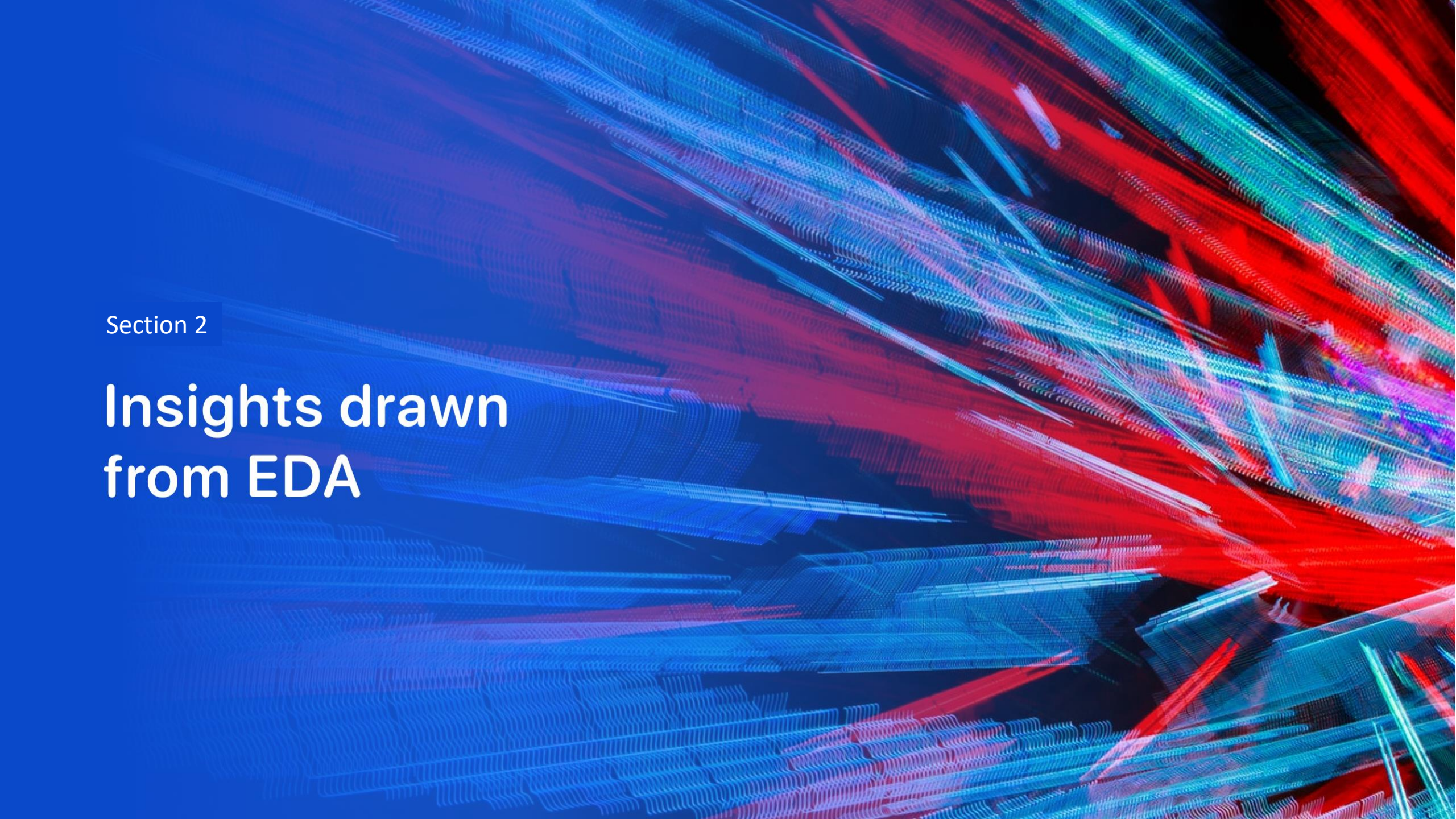
1. GridSearchCV for best parameters
2. Train with best parameters
3. Evaluate accuracy
4. Plot confusion matrix
5. Calculate Jaccard score, F1 score

Step3: Compare model performance metrics

Step4: Determine best-performing model

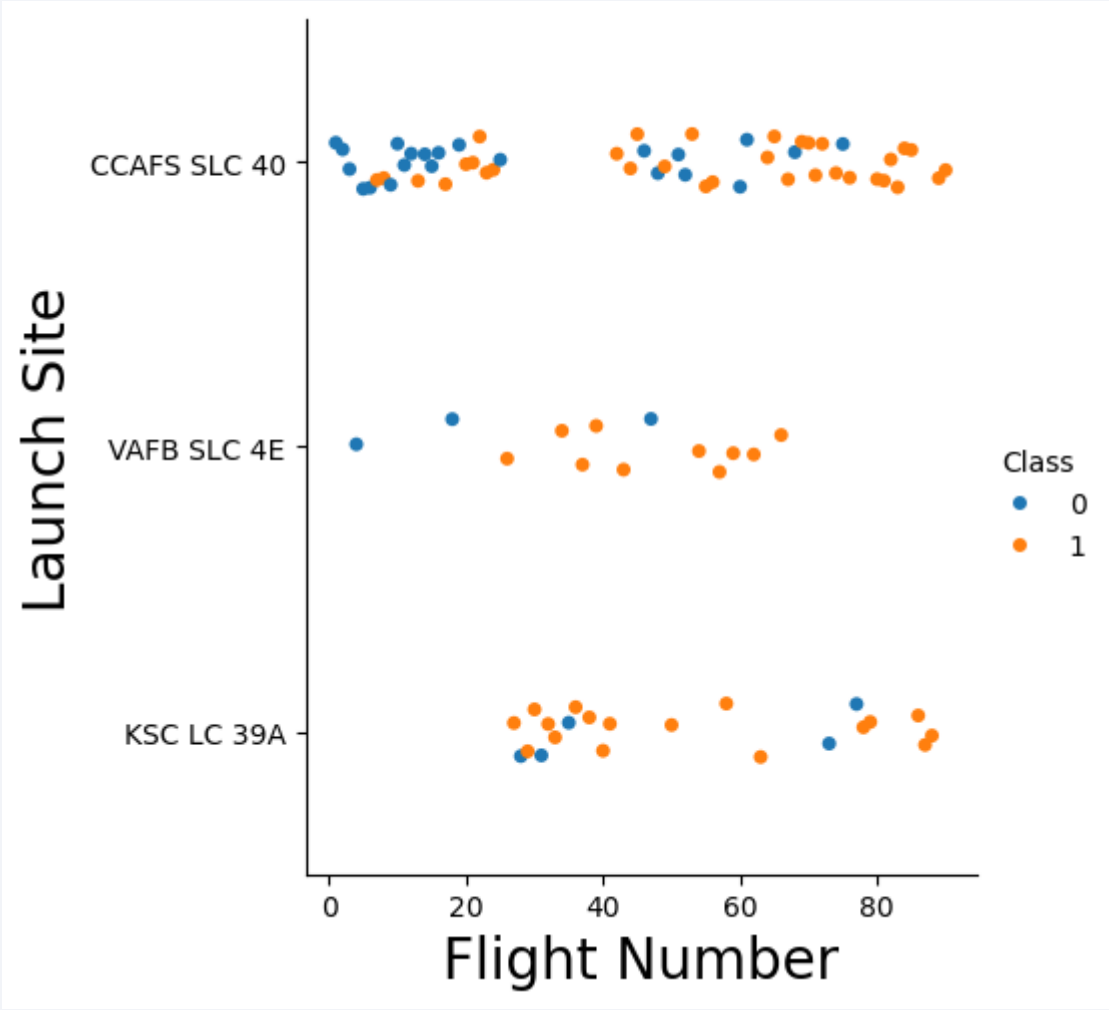
Results

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


Section 2
Insights drawn from EDA

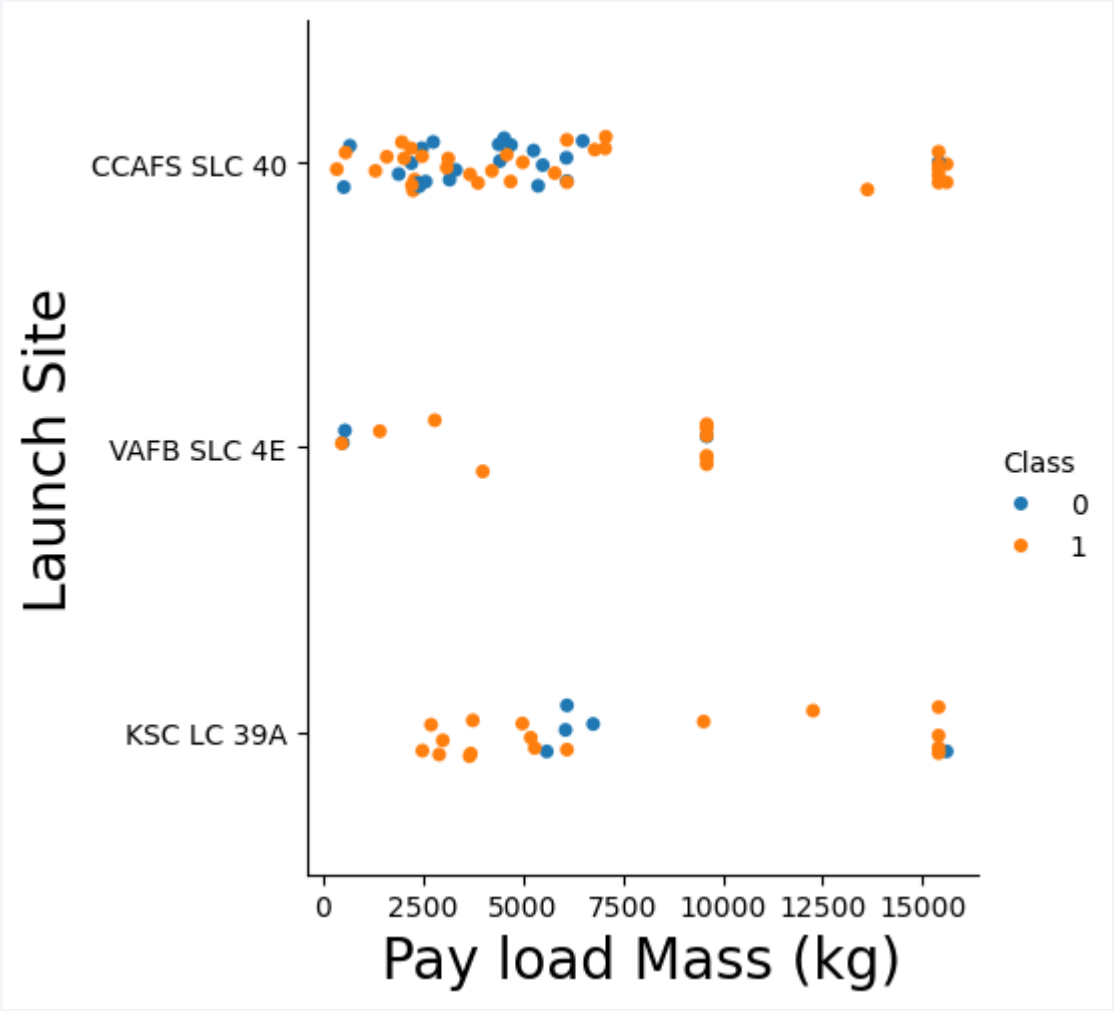
Flight Number vs. Launch Site

- Initially, the CCAFS SLC 40 launch site experiences few successes, but with an increase in flights, the success rate improves.
- A similar trend is observed at the VAFB SLC 4E launch site, where more flights correspond to more successful launches.
- At KSC LC 39A, there appears to be no direct correlation between the number of flights and the success rate.



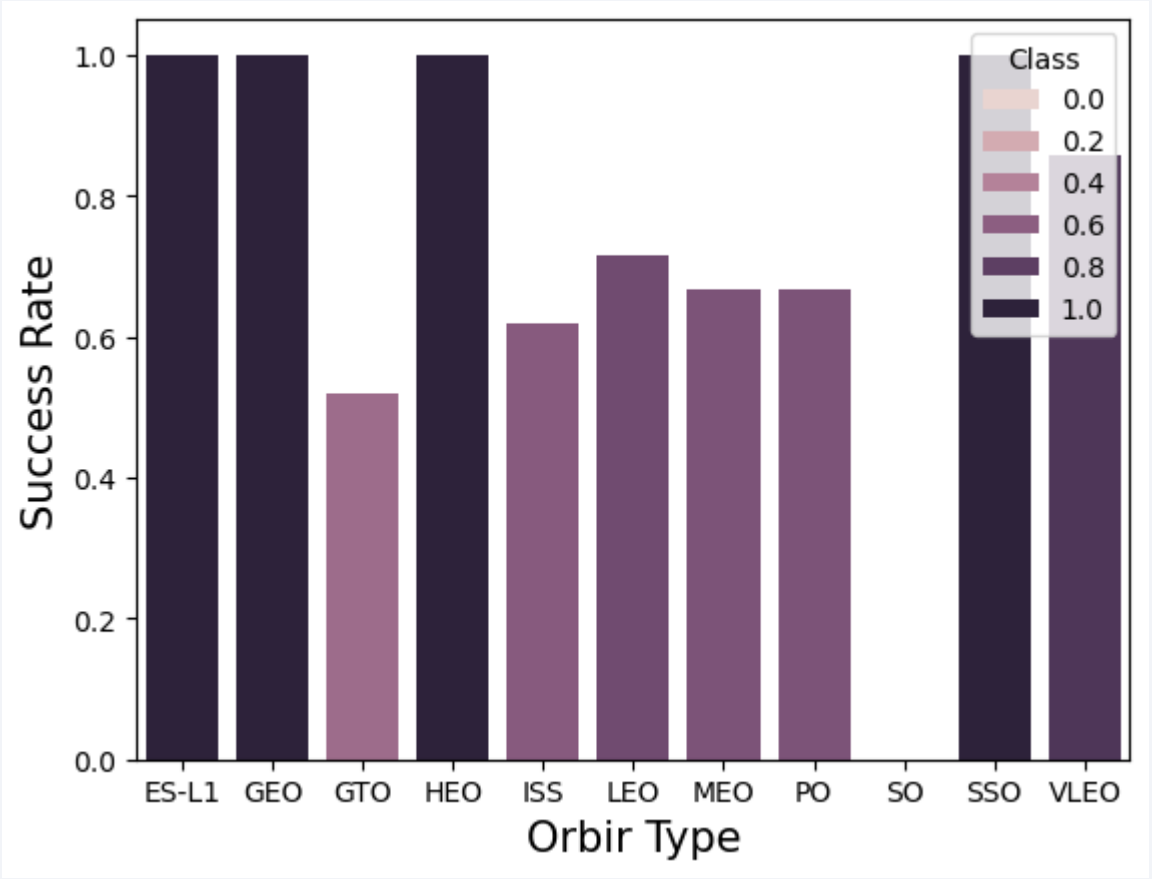
Payload vs. Launch Site

- Observing the Payload Vs. Launch Site scatter plot reveals no launches with heavy payload masses (greater than 10,000) from the VAFB-SLC launch site.
- It appears that an increase in payload mass correlates with an increase in the success rate.



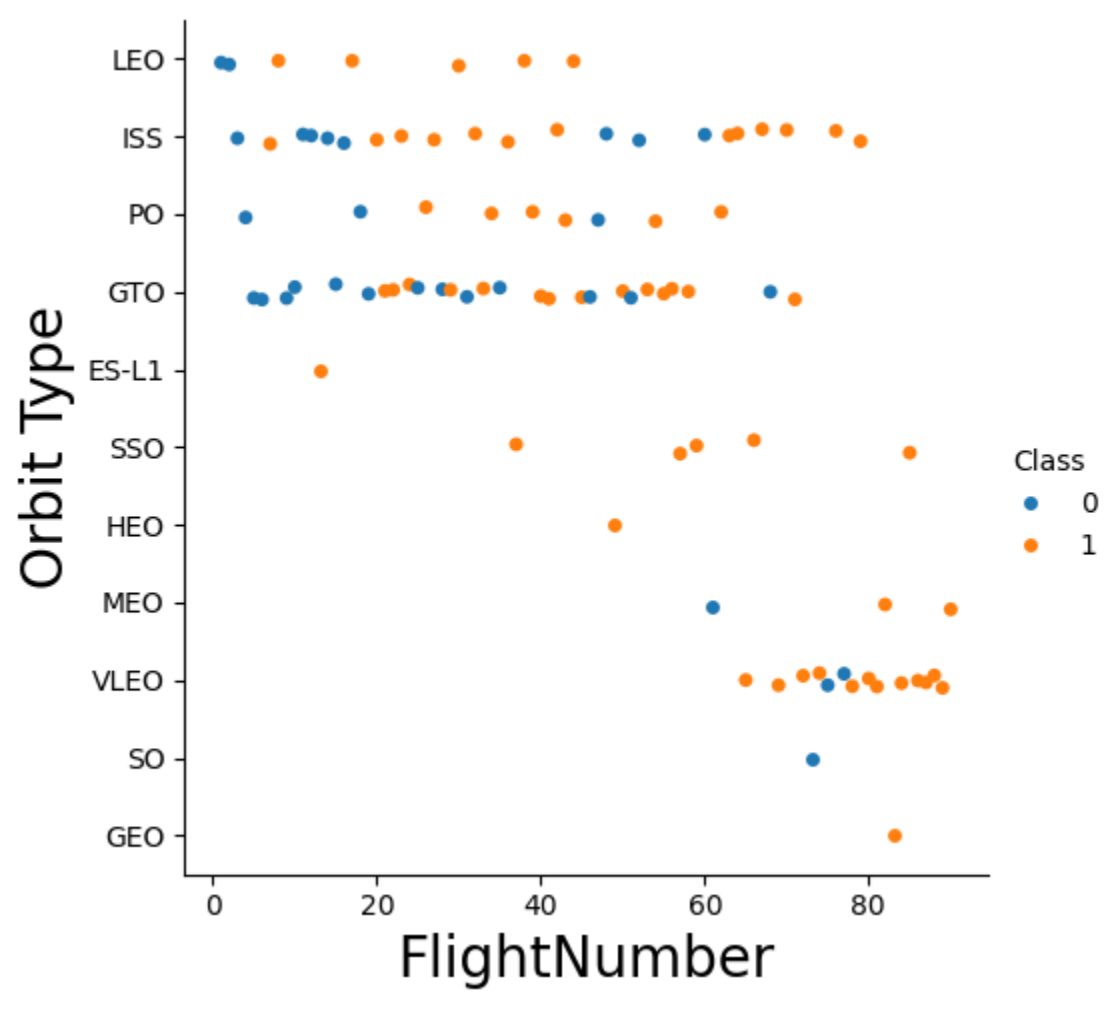
Success Rate vs. Orbit Type

- ES-L1, GEO, ISS, and SSOV orbits have the highest success rates.
- GTO orbit exhibits the lowest success rate, at only 51%.
- Other orbits generally display a respectable success rate.



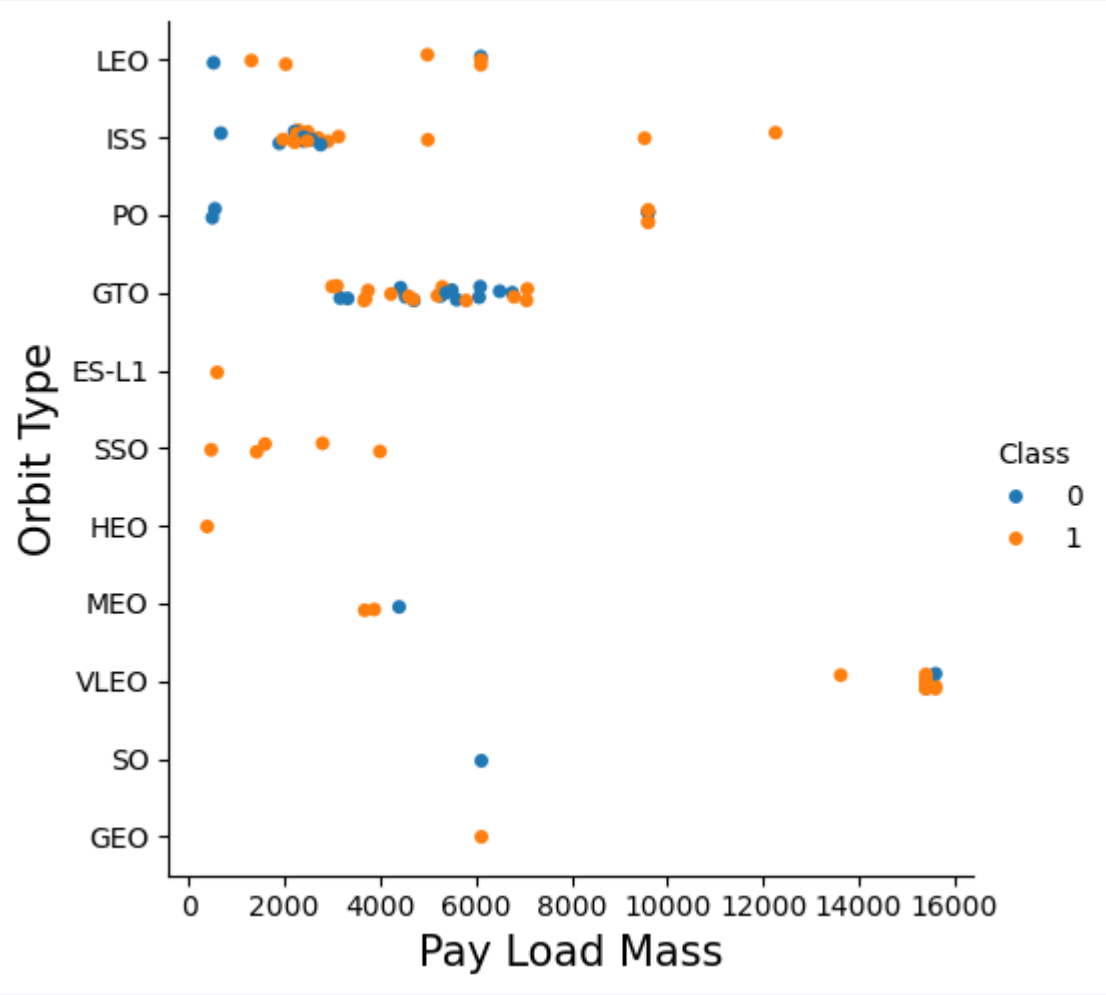
Flight Number vs. Orbit Type

- In the LEO orbit, success rate seems to correlate with the number of flights.
- Conversely, in the GTO orbit, there appears to be no correlation between the number of flights and success rate.



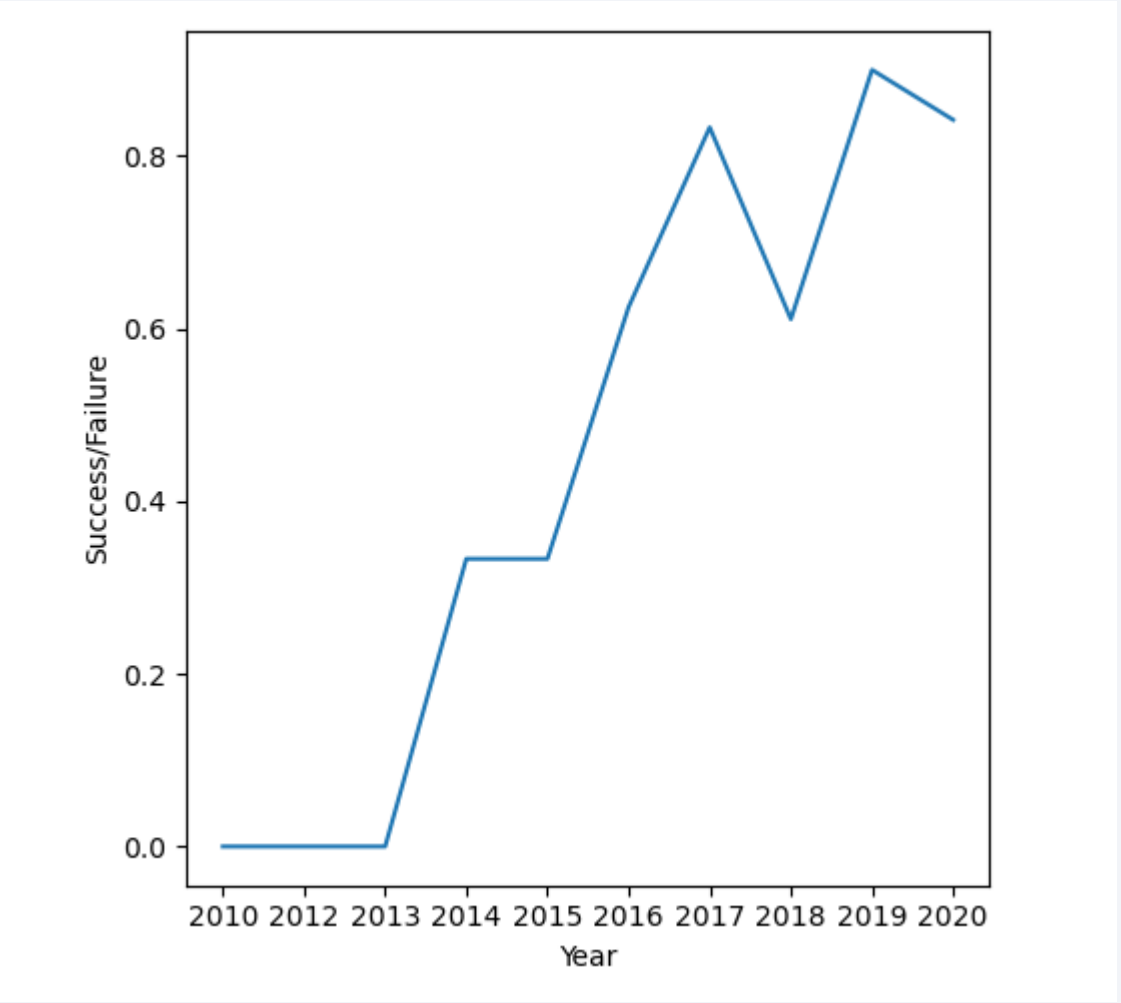
Payload vs. Orbit Type

- For heavy payloads, successful landings or positive landing rates are higher in Polar, LEO, and ISS orbits.
- In the case of GTO orbits, it's difficult to discern a clear pattern as both successful and unsuccessful landings are present.



Launch Success Yearly Trend

The success rate increases from 2013 to 2020.



All Launch Site Names

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

Launch Site Names Begin 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

Total Payload Mass

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

```
%sql select sum("PAYLOAD_MASS__KG_") AS " PAYLOAD MASS IN KILOGRAM" from "SPACEXTBL" where "Customer" = "NASA (CRS)"
```

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

PAYLOAD MASS IN KILOGRAM
45596

Average Payload Mass by F9 v1.1

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

```
%sql select avg("PAYLOAD_MASS__KG_") AS "AVERAGE PAYLOAD MASS IN KILOGRAM" from "SPACEXTBL" WHERE Booster_Version like 'F9 v1.1%'
```

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

AVERAGE PAYLOAD MASS IN KILOGRAM
2534.6666666666665

First Successful Ground Landing Date

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 "Mission_Outcome"= "Success" AND "Landing_Outcome" = "Success"
```

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

min("Date")
2018-07-22

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%sql select DISTINCT Payload from "SPACEXTBL" where "Landing_Outcome" = "Success (drone ship)" and "PAYLOAD_MASS__KG_" > 4000 and "PAYLOAD_MASS__KG_"<6000
```

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

Payload
JCSAT-14
JCSAT-16
SES-10
SES-11 / EchoStar 105

Total Number of Successful and Failure Mission Outcomes

```
%sql SELECT Landing_Outcome, COUNT(*) AS TOTAL From SPACEXTABLE WHERE Landing_Outcome IN ('Success','Failure') GROUP BY Landing_Outcome
```

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

Landing_Outcome	TOTAL
Failure	3
Success	38

Boosters Carried Maximum Payload

```
%sql SELECT Payload FROM SPACEXTABLE WHERE Payload_Mass__KG_ = (SELECT MAX(Payload_Mass__KG_) From SPACEXTABLE)
```

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

Payload
Starlink 1 v1.0, SpaceX CRS-19
Starlink 2 v1.0, Crew Dragon in-flight abort test
Starlink 3 v1.0, Starlink 4 v1.0
Starlink 4 v1.0, SpaceX CRS-20
Starlink 5 v1.0, Starlink 6 v1.0
Starlink 6 v1.0, Crew Dragon Demo-2
Starlink 7 v1.0, Starlink 8 v1.0
Starlink 11 v1.0, Starlink 12 v1.0
Starlink 12 v1.0, Starlink 13 v1.0
Starlink 13 v1.0, Starlink 14 v1.0
Starlink 14 v1.0, GPS III-04
Starlink 15 v1.0, SpaceX CRS-21

2015 Launch Records

```
%%sql
SELECT substr(Date, 6, 2) AS MONTH ,Date,BOOSTER_VERSION,LAUNCH_SITE,LANDING_OUTCOME
FROM SPACEXTABLE
WHERE LANDING_OUTCOME = 'Failure (drone ship)' and substr(Date,0,5)='2015'
```

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

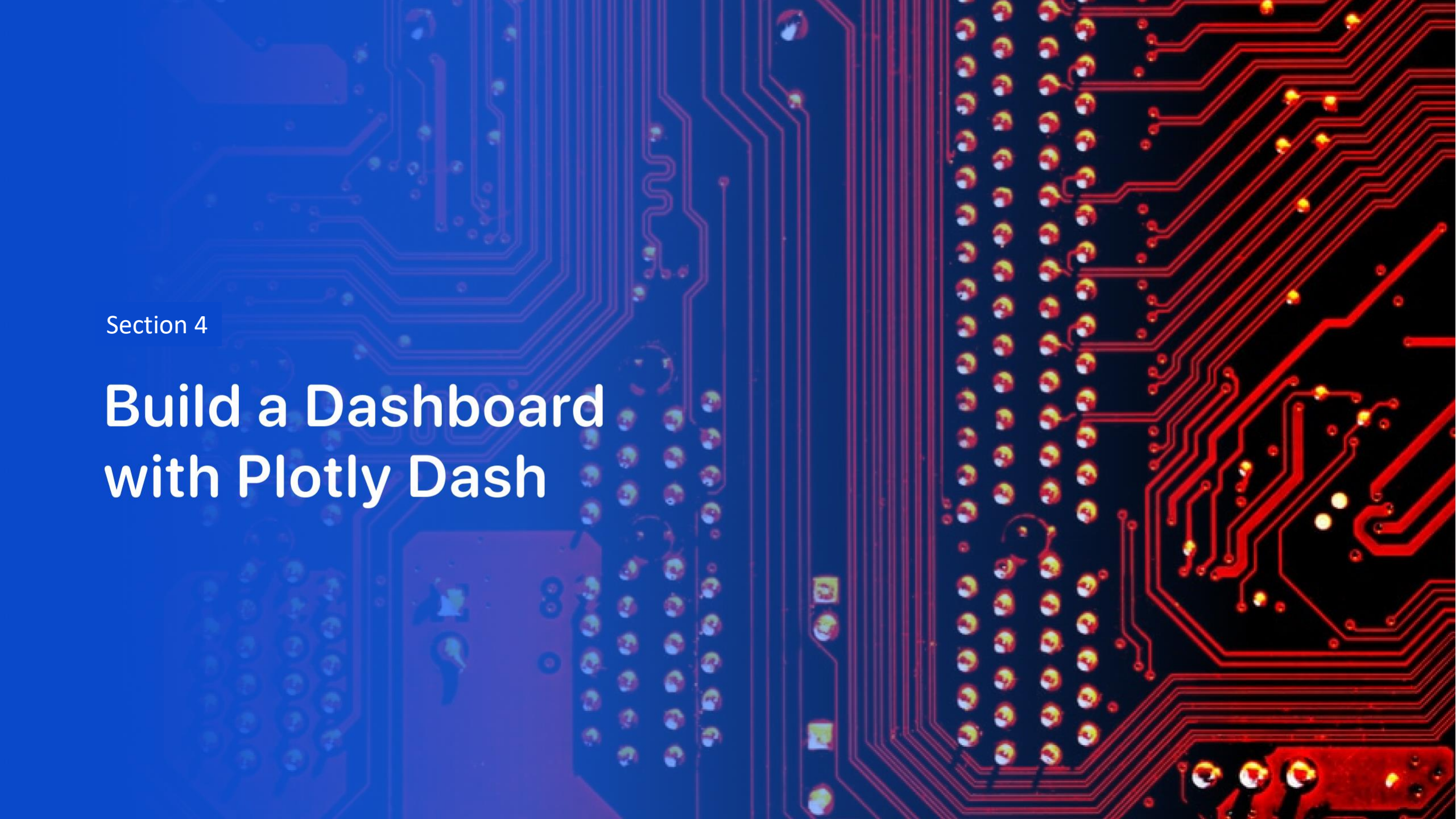
MONTH	Date	Booster_Version	Launch_Site	Landing_Outcome
01	2015-01-10	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
04	2015-04-14	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

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

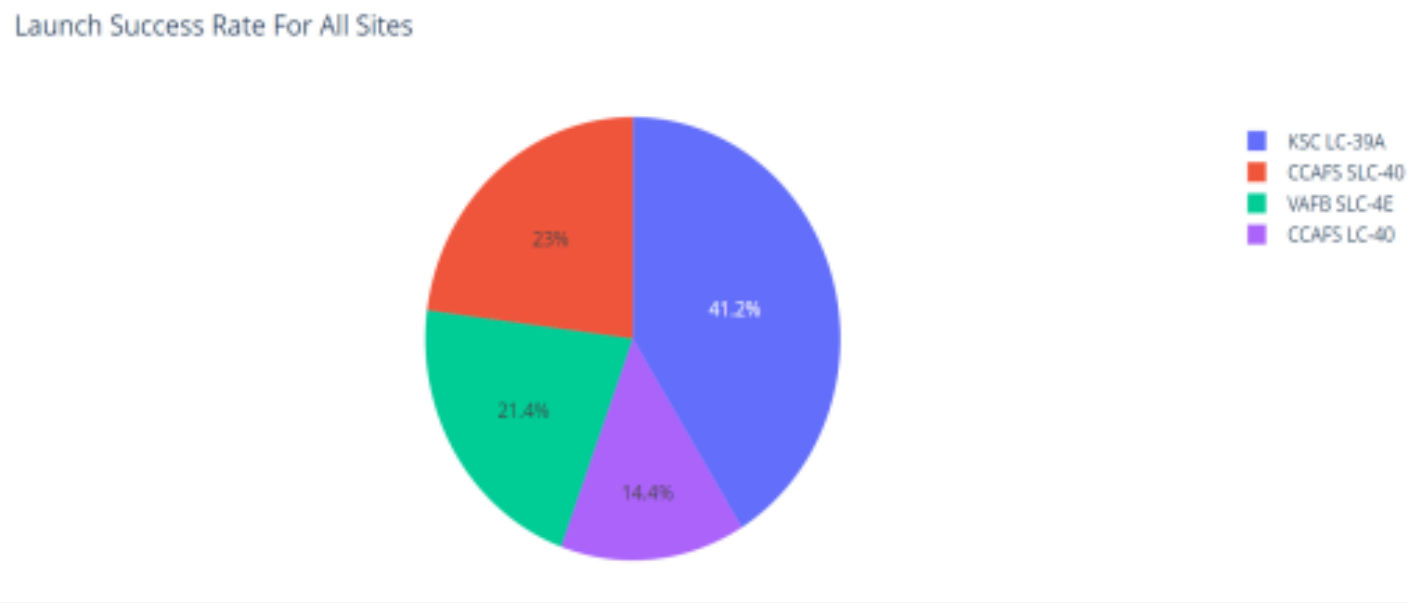
```
%%sql
SELECT LANDING_OUTCOME, count(*) AS count_outcome
FROM SPACEXTABLE
WHERE date between '2010-06-04' and '2017-03-20'
group by Landing_outcome
order by count_outcome desc
```

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

Landing_Outcome	count_outcome
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

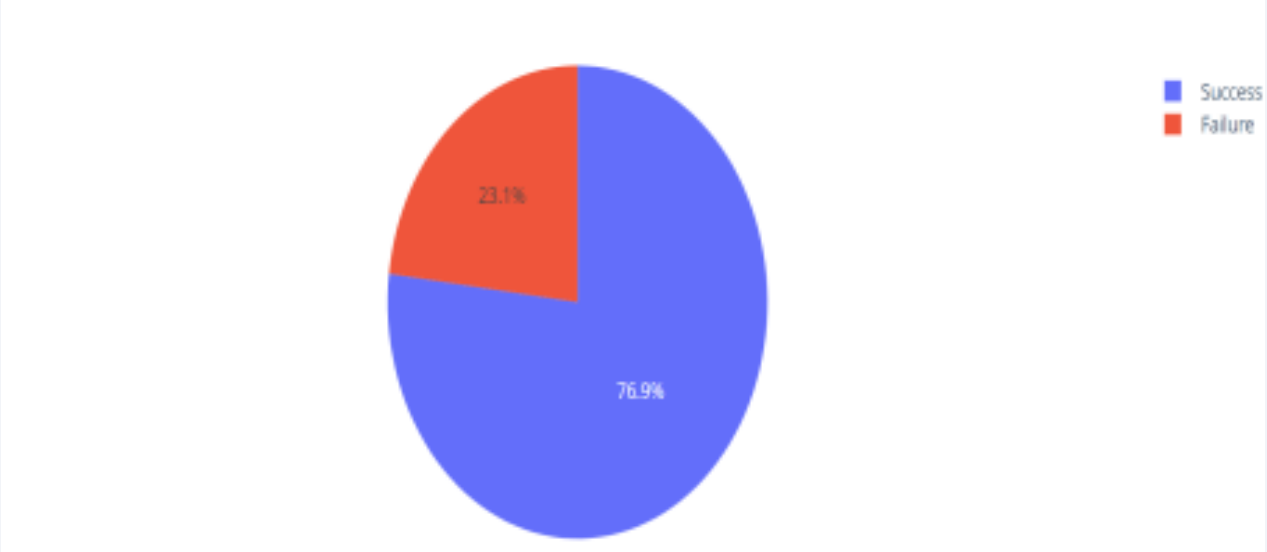

Section 4
Build a Dashboard with Plotly Dash

Launch success count for all site



KSC LC-39A obtained the highest success rate with 41.2% the least was CCAFS-40 with only 14.4%

Piechart for the launch site with highest launch success ratio

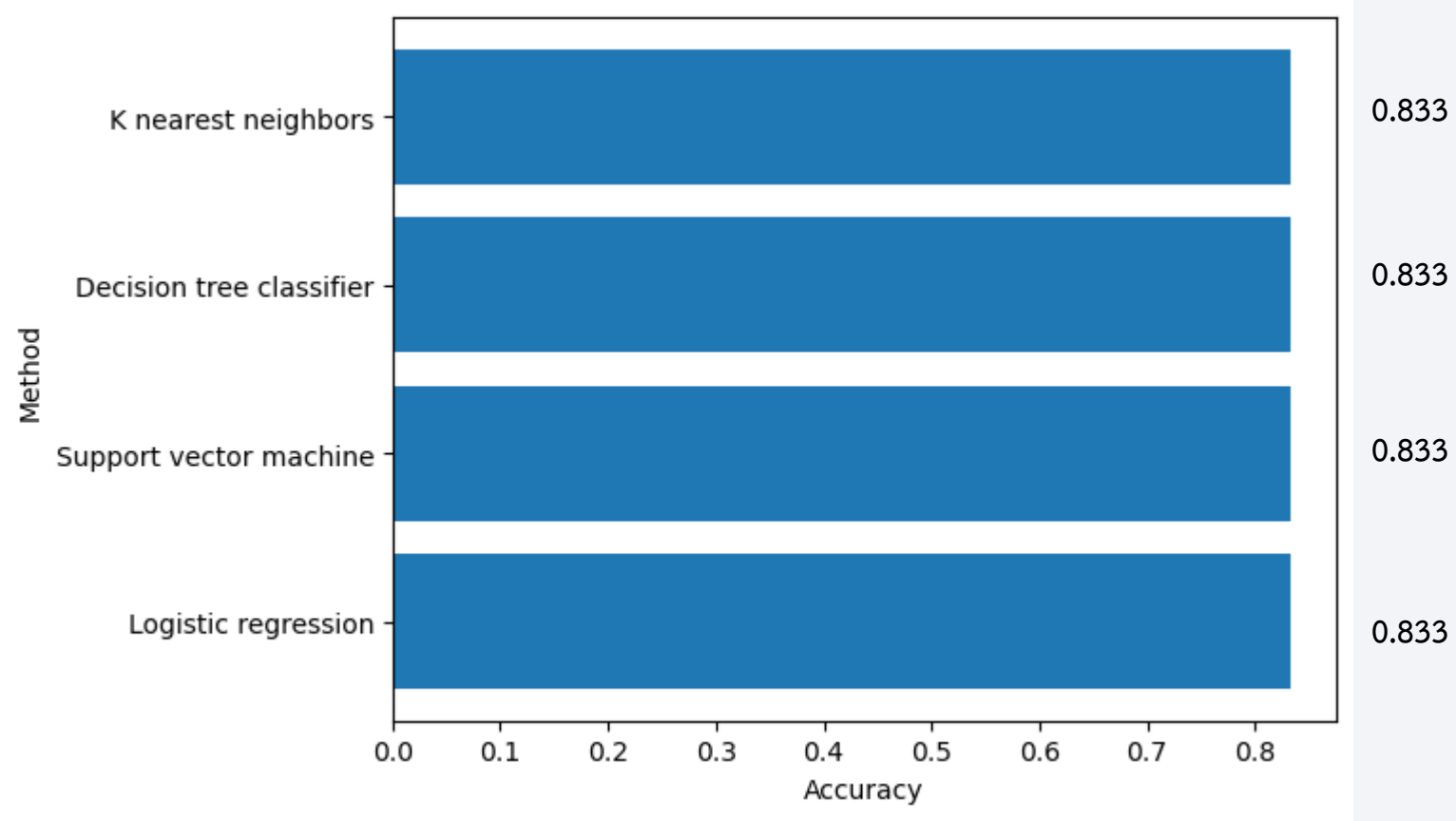


KSC LC-39A is the site that assures us a higher success rate we should choose this place more often for future launches

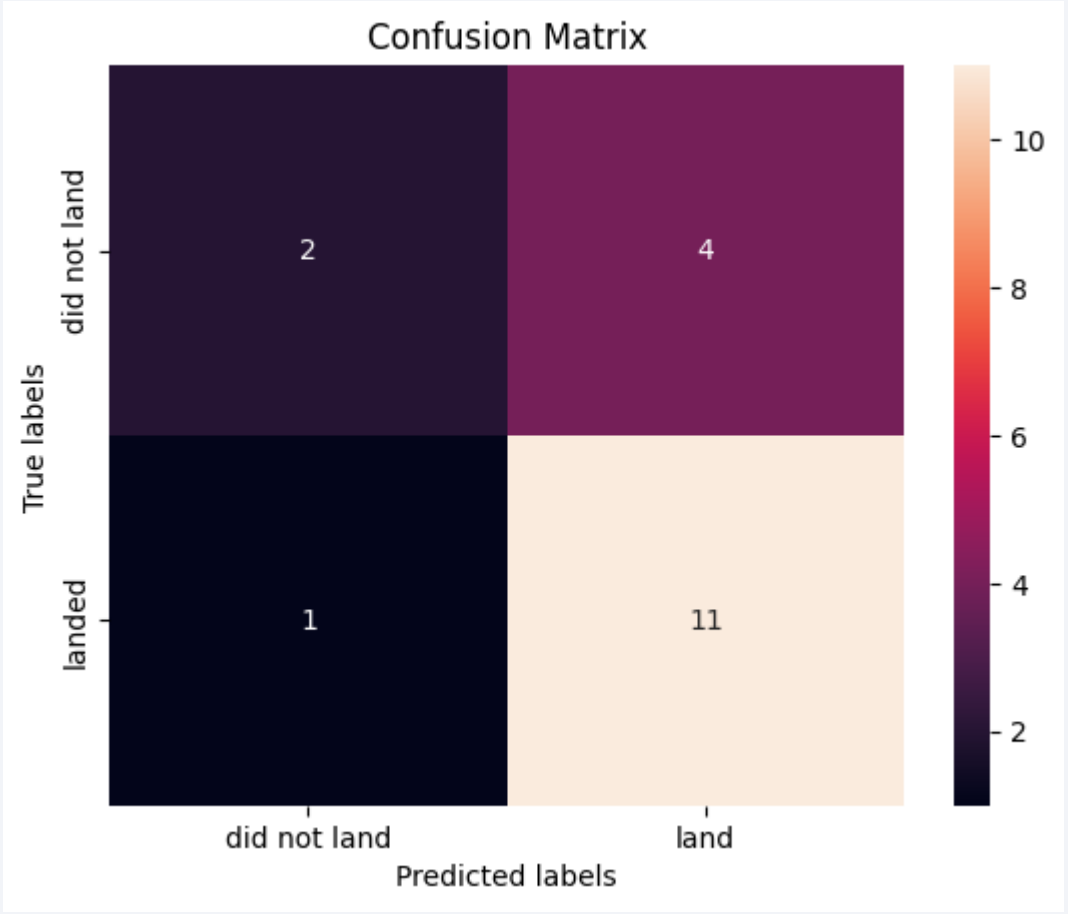
Section 5

Predictive Analysis (Classification)

Classification Accuracy



Confusion Matrix



Conclusions

As the frequency of launches has grown, there has been a notable improvement in the rate of successful missions, recently surpassing the 80% mark. Orbital classifications such as SSO, HEO, GEO, and ES-L1 boast a perfect success rate of 100%. The location of the launch site is strategically chosen for its proximity to railways, highways, and the coastline while maintaining a significant distance from urban areas. Among all the launch sites, KSLC-39A stands out with the highest number of successful launches and the best success rate. Furthermore, launches carrying lighter payloads have demonstrated a higher success rate compared to those with heavier payloads. Within the scope of the dataset, each model has shown an identical accuracy rate of 83.33%. However, the limited size of the dataset suggests a need for more data to accurately identify the most effective model.

Appendix

- GitHub URL
- Coursera Applied Data Science Capstone Course URL

Thank you!