IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Tasha Fahira
March 2024

Outline

Executive Summary

SpaceX revolutionized the space exploration industry by making the first-stage booster rockets to land back so it can be reused. We used public data from the internet to analyze SpaceX Falcon 9 launches and predict landing success rate. Here are highlight of this report:

- **Launch Success Trends**: An upward trend in success rates from 2013 to 2020, with Kennedy Space Center's LC-39A site achieving over a 75% success rate, highlights technological and operational advancements.
- **Payload Analysis**: There are no clear correlation between payload mass and success rate, especially with the Falcon 9 (FT) booster, indicates the nuanced factors influencing launch outcomes.
- **Solution through predictive analysis**: We found Logistic Regression to be both simple and effective, underscoring the potential of predictive analytics to enhance decision-making and optimize launch success.

In conclusion, our analysis underscores the value of predictive analytics. By leveraging data-driven insights and efficient algorithms, we can better navigate the complexities of space launches, ultimately enhancing the reliability and efficiency of space exploration missions.

Back to Outline

Introduction

- **The Challenge**
  - Rocket launches are expensive (upwards of $165M)
- **The Innovation**
  - SpaceX cuts the cost of rocket launch missions by roughly half because they're able to reuse their first-stage booster rockets by making them land
- **The Role of Data**
  - Analysis of launch data – mission and landing outcomes, payload masses, launch site performance, etc. + empowered by predictive analysis is central to make strategic decision
- **Objective:**
  - This presentation explores key trends in space launches, including success rates, payload impacts, and the effectiveness of predictive models, aiming to highlight factors behind successful missions

Section 1

Methodology

Back to Outline

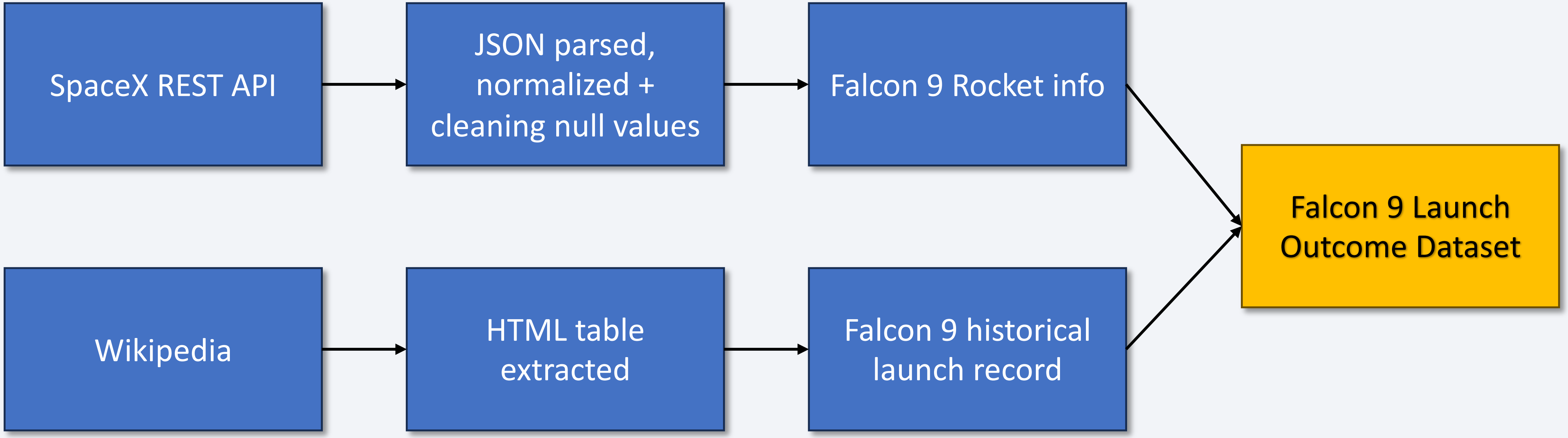
Methodology

Executive Summary

- Data collection methodology:
  - SpaceX REST API
  - Web scraping from Wikipedia
- Data wrangling
  - Created target label: “Class” column – represents landing outcome of the rocket
- Exploratory data analysis (EDA) using visualization (Matplotlib) and SQL
- Interactive visual analytics using Folium and Plotly Dash
- Predictive Analysis using 4 machine learning algorithms: logistic regression, SVM, Decision Tree, KNN

Data Collection

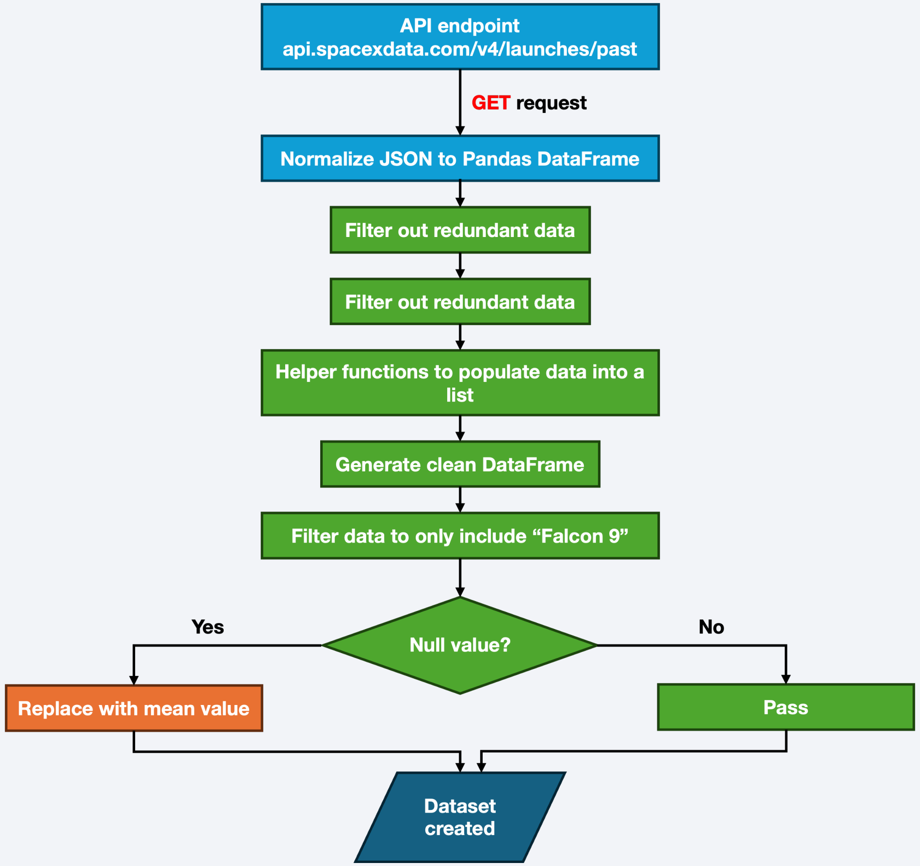
- Data collected using two methods: REST API call and Web scraping
- We collected information about the rocket used, payload delivered, launch specifications, landing specifications, and landing outcome



Data Collection – SpaceX API

1. Request and parse data using GET request
2. Filter DataFrame to only include Falcon 9 launches
3. Replace null values with mean values

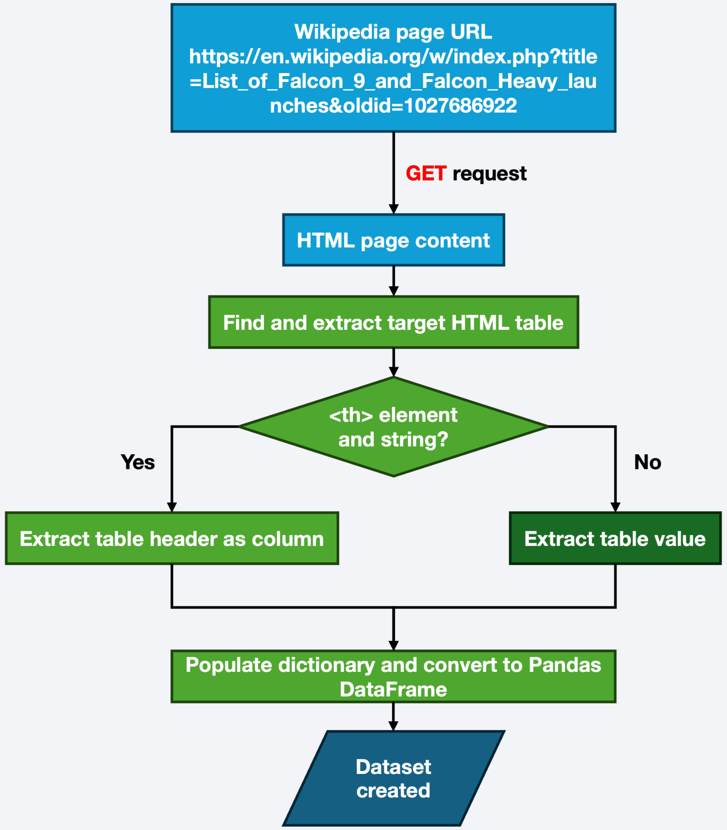
[GitHub URL here](#)



Data Collection - Scraping

1. Request the Falcon 9 Launch Wiki page from URL
2. Extract all column/variable names from the HTML table header
3. Create DataFrame by parsing HTML table

GitHub URL here



Back to Outline

Data Wrangling

1. Calculate the number of launches per site

Launch site	Number of launches
CCAFS SLC 40	55
KSC LC 39A	22
VAFB SLC 4E	13

2. Calculate the number and occurrence of each orbit

Orbit name	Number of occurence
GTO	27
ISS	21
…	
GEO	1

3. Differentiate between booster landing outcomes

Landing outcome	Terms included
Successful	True ASDS, True RTLS, True Ocean
Failed	None None, False ASDS, False Ocean, None ASDS, False RTLS

4. Created "Class" column

Landing outcome	Class label
Successful	1
Failed	0

- EDA was done to find out why landing outcome is successful or failed and then created a target label column called "Class"

GitHub URL here

Back to Outline

EDA with Data Visualization

- We did exploratory data analysis (EDA) on the Falcon 9 Launch Outcome Dataset, analyzing the relationship of the following features:
  - Flight number
  - Launch site
  - Payload mass
  - Launch success rate
  - Launch orbit type
- See the results slide for more details

GitHub URL here

Back to Outline

EDA with SQL

We did EDA on Falcon 9 Launch Outcome dataset with SQL. Here are some of our primary findings:

- There are 4 unique launch sites
- Average payload mass carried by F9 is around 2,500 kg
- The first successful ground landing attempt was on Dec 22nd, 2015
- Etc.

Click here to jump to the actual query results

GitHub URL here

Back to Outline

Build an Interactive Map with Folium

- We created various Folium objects such as circles, markers, lines, etc. to analyze the following aspects of the launch sites:
  - Location of the Launch Sites
  - Landing Outcome in Each Site
  - Proximity Between Site and Important Landmarks incl. highway, coastline, airport, railway, and city center

GitHub URL here

[Back to Outline](#)

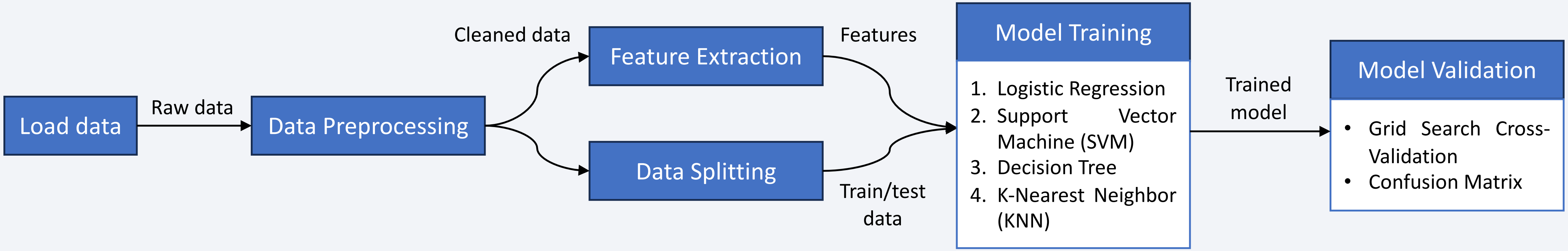
Build a Dashboard with Plotly Dash

- We plotted two kinds of graph:
  - [Launch Success Rate of All Sites](#)
  - [Correlation between Payload Mass and Success Rate](#)
- We want to see which launch site has the greatest number of successful launches
- We can analyze whether there is any correlation between payload mass and the launch success rate

[GitHub URL here](#)

Predictive Analysis (Classification)

We build a machine learning pipeline to train and predict if Falcon 9 launch will succeed or fail



GitHub URL here

Results

Click the blocks to jump to result of each section

Exploratory Data Analysis Results

Predictive analysis result

Interactive Proximities Analytics

Interactive Dashboard Screenshots

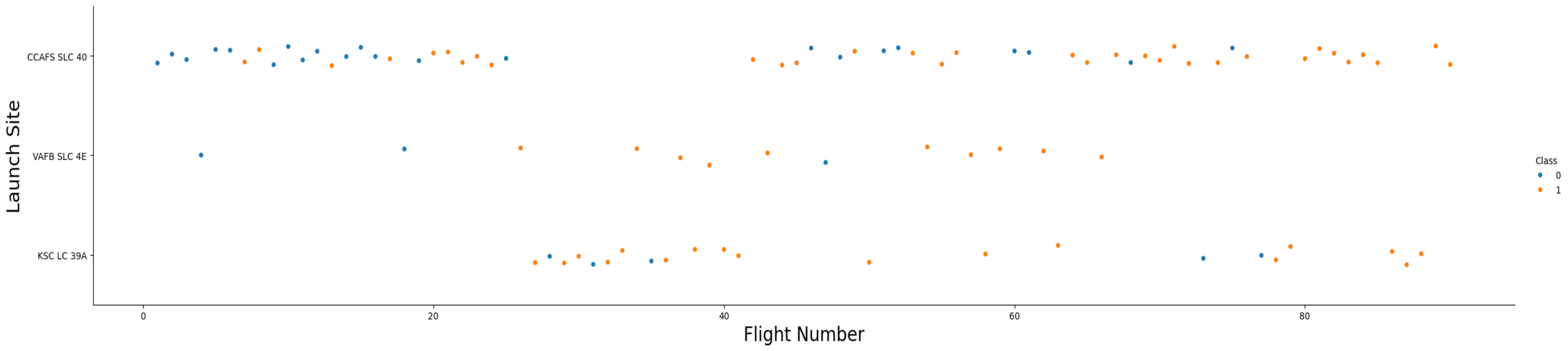
Section 2

Insights drawn from EDA

Back to Outline

Back to Results

Flight Number vs. Launch Site



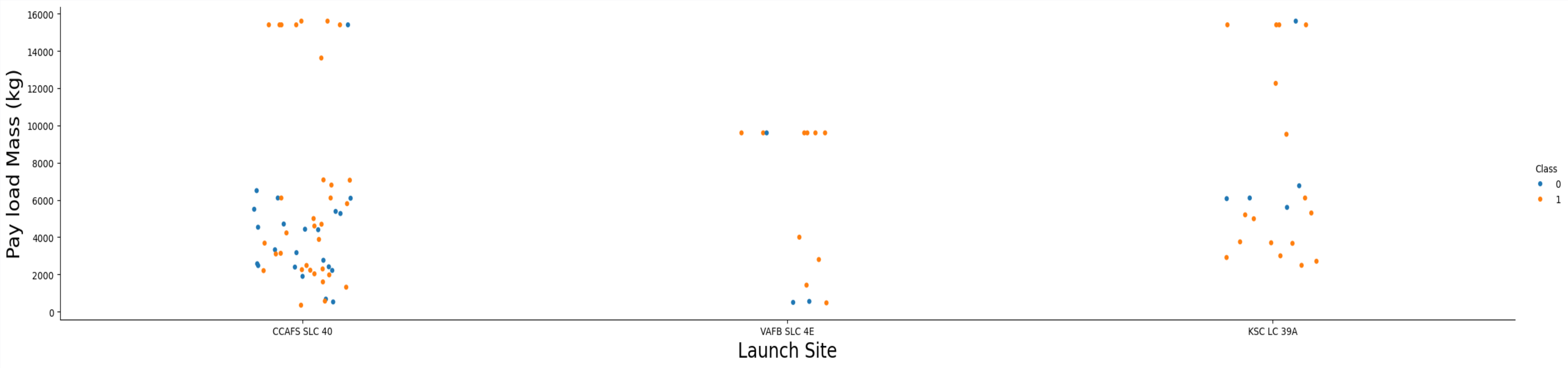
We see that different launch sites have different success rates.

“CCAFS LC-40” has a success rate of 60%, while “KSC LC-39A” and “VAFB SLC 4E” has a success rate of 77%

Back to Outline

Back to Results

Payload vs. Launch Site



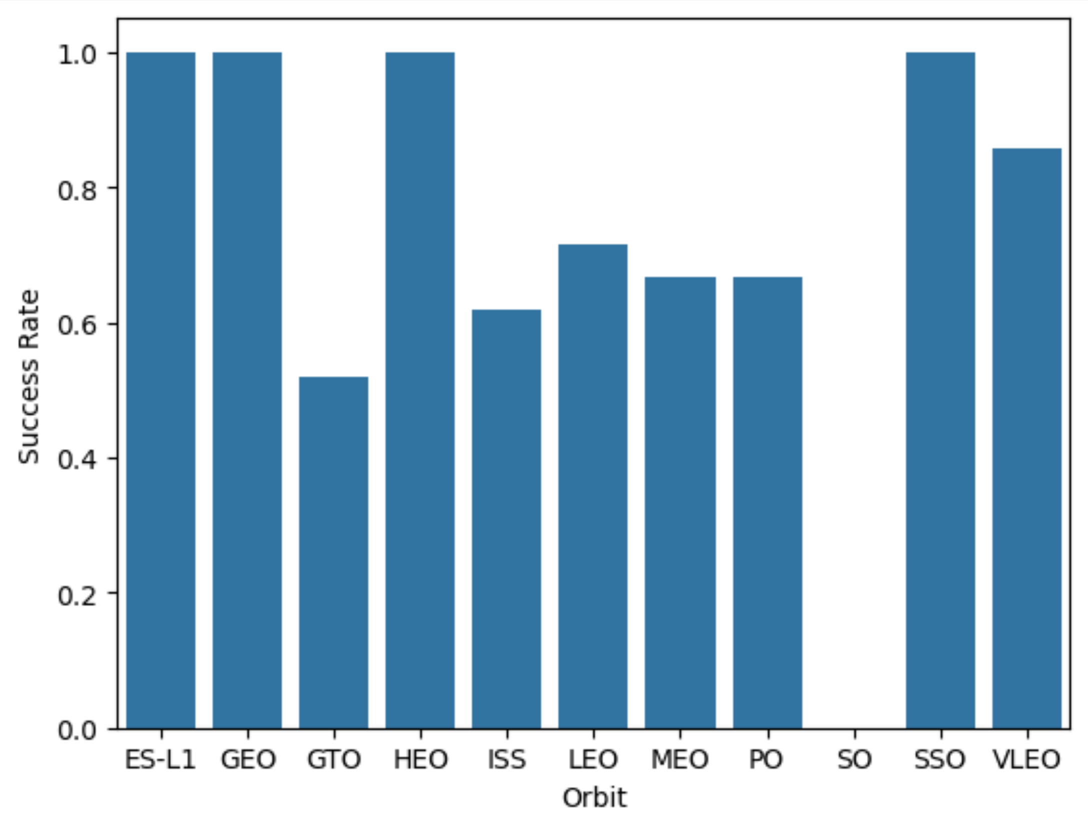
We can observe that there are no rockets launched with payload mass greater than 10,000 kg in VAFB SLC 4E

[Back to Outline](#)

[Back to Results](#)

Success Rate vs. Orbit Type

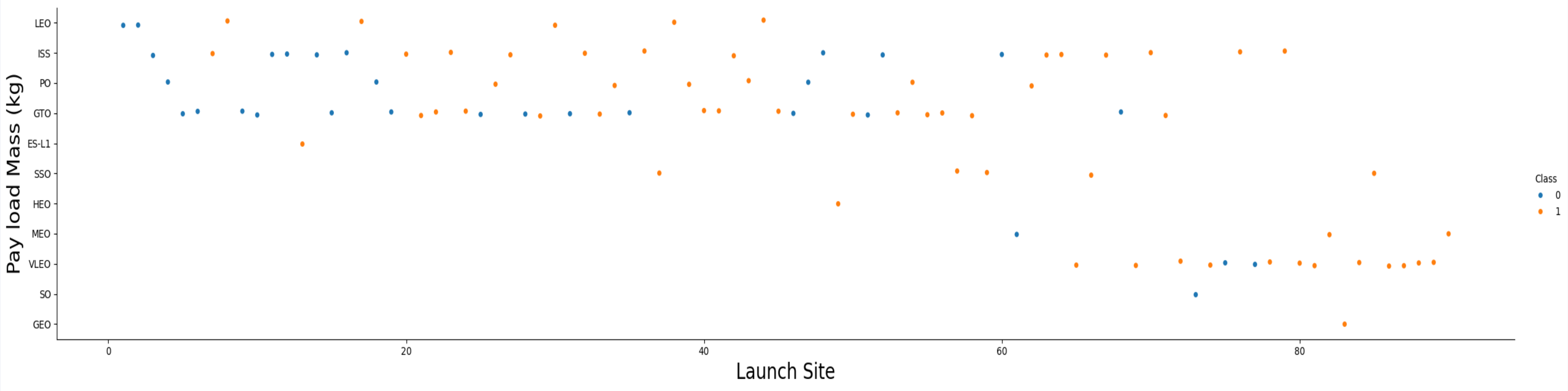
The destination orbits with high success rate (> 80%) are ES-L1, GEO, HEO, SSO, and VLEO



Back to Outline

Back to Results

Flight Number vs. Orbit Type

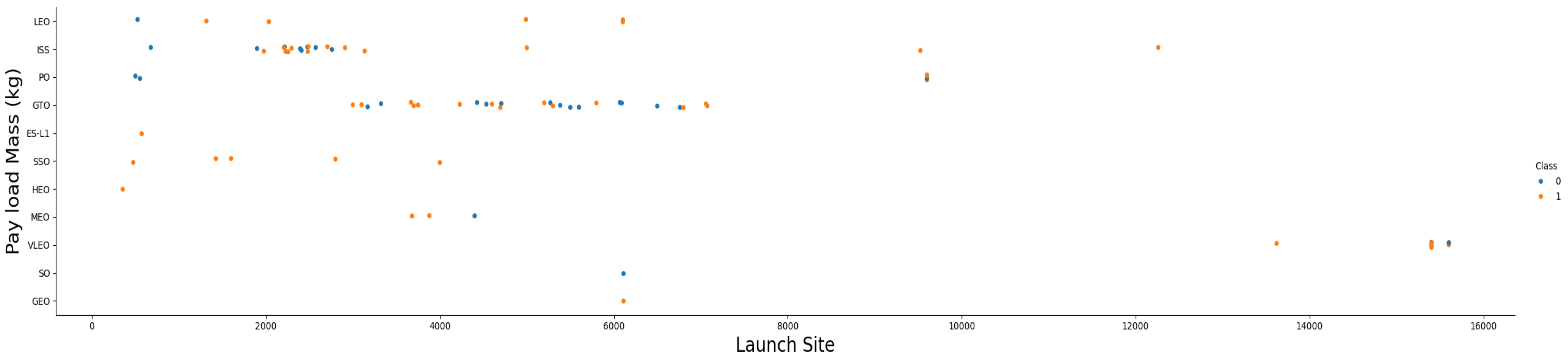


- In the LEO orbit, the success rate appears related to the number of flights
- There seems to be no relationship between flight number when in GTO orbit

[Back to Outline](#)

[Back to Results](#)

Payload vs. Orbit Type



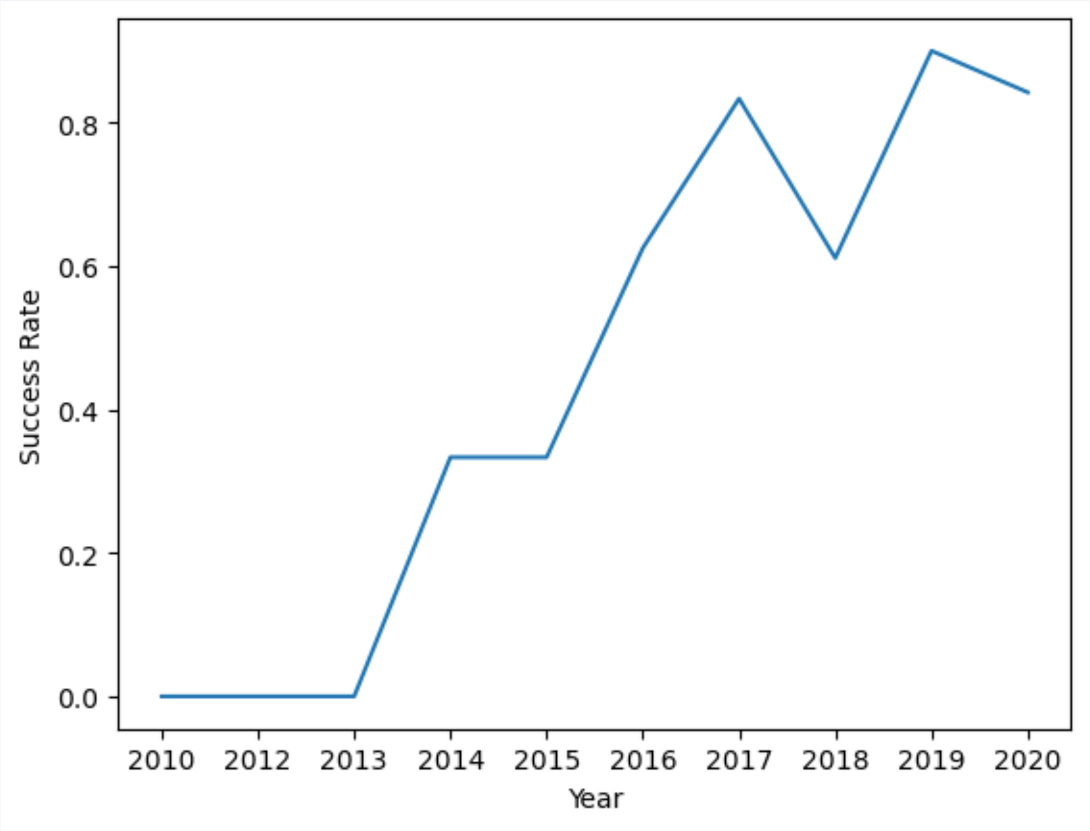
- For heavy payloads (> 10,000 kg), successful landing can be observed for Polar, LEO, and ISS
- GTO orbit has the most launch attempts, but no obvious landing outcome patterns can be observed

Back to Outline

Back to Results

Launch Success Yearly Trend

The success rate kept increasing from 2013 to 2020



Back to Outline

Back to Results

All Launch Site Names

- Display list of distinct launch sites:

```
select distinct(Launch_Site) from SPACEXTABLE;
```

- Query result:

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

- Explanation:
  Four launch sites can be observed, but “CCAFS LC-40” is in the same location as “CCAFS SLC-40” -> **Three site locations**

Back to Outline

Back to Results

Launch Site Names Begin with 'CCA'

- Display 5 records where launch sites begin with ”CCA”:

```
select * from SPACEXTABLE WHERE Launch_Site like 'CCA%' LIMIT 5;
```

- Query result:

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

- Explanation:
  The destination of the first 5 launches at CCAFS LC-40 is the LEO orbit type. While the mission was successful, the landing outcome was a failure or no attempt was made

Back to Outline

Back to Results

Total Payload Mass

- Calculate the total payload mass carried by boosters launched by NASA:

```
select SUM(PAYLOAD_MASS__KG_) from SPACEXTABLE WHERE Customer LIKE '%NASA%';
```

- Query result:

SUM(PAYLOAD_MASS__KG_)
107010

- Explanation:
  Total payload mass carried by boosters launched by NASA is 107,000 kg

Back to Outline

Back to Results

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1:
```
select AVG(PAYLOAD_MASS__KG_) from SPACEXTABLE WHERE Booster_Version LIKE '%F9 v1.1%’;
```

- Query result:

AVG(PAYLOAD_MASS__KG_)
2534.6666666666665

- Explanation:
  Average payload mass carried by booster version F9 v1.1 is 2,534.6 kg

Back to Outline

Back to Results

First Successful Ground Landing Date

- Find the date of first successful landing in ground pad was achieved:
```
select min(Date) from SPACEXTABLE WHERE Landing_Outcome = 'Success (ground pad)';
```
- Query result:

min(Date)
2015-12-22

- Explanation:
  The first successful landing attempt in ground pad was in Dec 22nd, 2015

Back to Outline

Back to Results

Successful Drone Ship Landing with Payload between 4000 and 6000

- List booster versions which have success in drone ship and have 6000 > payload mass > 4000:

```
select Booster_Version from SPACEXTABLE WHERE Landing_Outcome = 'Success (drone ship)' and PAYLOAD_MASS__KG_ between 4000 and 6000;
```

- Query result:

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

- Explanation:
F9 FT are the only successful booster version which has successfully landed in a drone ship and has the payload between 4,000 – 6,000 kg

Back to Outline

Back to Results

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failed landing outcomes:
```
SELECT (SELECT COUNT(*) FROM SPACEXTABLE WHERE Mission_Outcome LIKE '%Success%') AS Successful_Missions, (SELECT COUNT(*) FROM SPACEXTABLE WHERE Mission_Outcome LIKE '%Failure%') AS Failure_Missions;
```

- Query result:

Successful_Missions	Failure_Missions
100	1

- Explanation:
  There was only 1 out of 101 missions that ended in failure due to presumed second stage issue.
  Find out more here

Back to Outline

Back to Results

Boosters Carried Maximum Payload

- List the booster names that have carried the max. payload mass:
```
SELECT Booster_Version FROM SPACEXTABLE WHERE PAYLOAD_MASS__KG_ = (SELECT
MAX(PAYLOAD_MASS__KG_) FROM SPACEXTABLE);
```

- Query result:

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

- Explanation:
Booster F9 B5 were used to carry the maximum payload mass (which was 15,600 kg)

Back to Outline

Back to Results

2015 Launch Records

- Display the month, booster version, and launch site for failed landing in drone ship during year 2015:

```
SELECT substr(Date, 6,2) as month, Landing_Outcome, Booster_Version, Launch_Site 
FROM SPACEXTABLE WHERE substr(Date,0,5)='2015' AND Landing_Outcome='Failure (drone 
ship)';
```

- Query result:

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

- Explanation:
  The failed landing in 2015 was in January which was the first attempt at landing on a floating barge while in April was the third attempt to do so

[Back to Outline](#)

[Back to Results](#)

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

- Rank the count of landing outcomes between the date 2010-06-04 and 2017-03-20, in descending order:

```
SELECT Landing_Outcome, COUNT(*) as count FROM SPACEXTABLE WHERE Date BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY Landing_Outcome ORDER BY count DESC;
```

- Query result:

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

- Explanation:
  There are less successful attempted landing outcomes than has been attempted (8 out of 31) from 2010 to 2017

Section 3

Launch Sites Proximities Analysis

[Back to Outline](#)

[Back to Results](#)

Location of Launch Sites

- All launch sites are near the equatorial line and located near coasts
- Launching rockets near the equator utilizes the Earth's rotation for speed boost, which saves fuel and increase payload capacity
- Launching near the coast minimizes potential risk of rocket debris impacting populated areas

[Back to Outline]

[Back to Results]

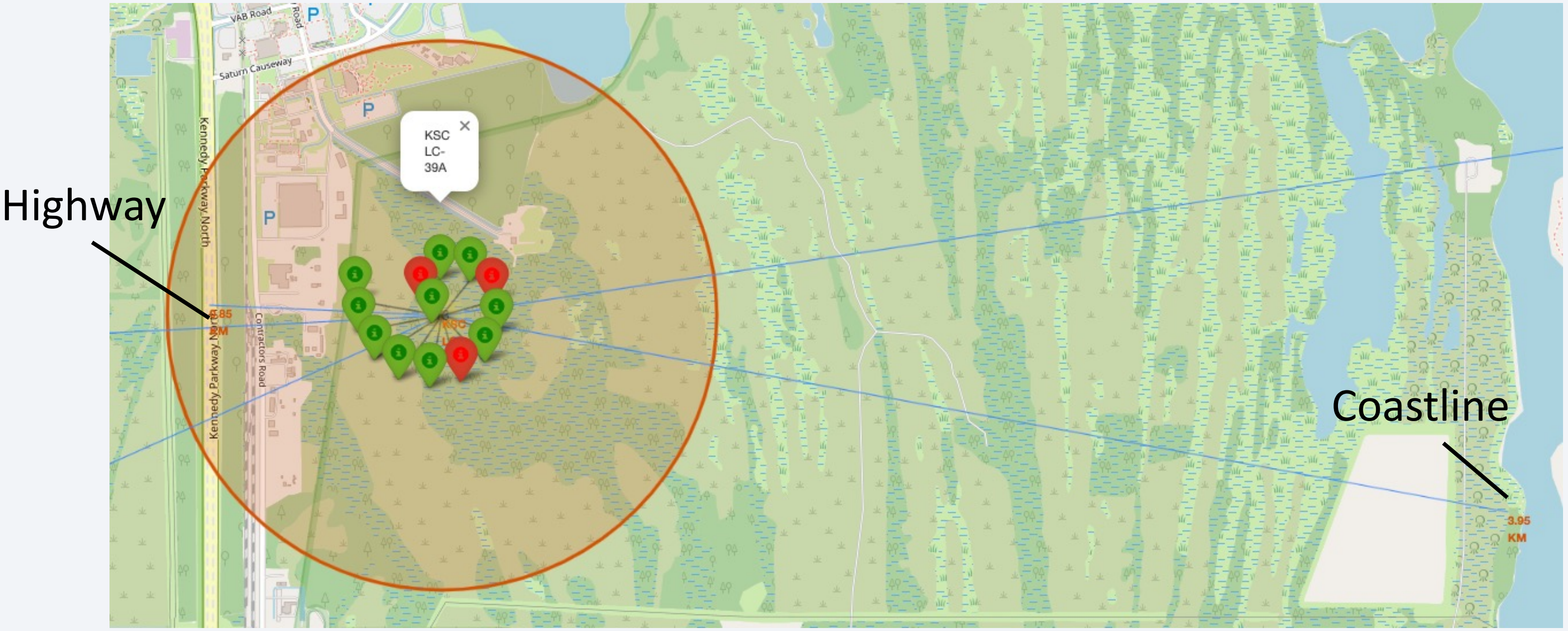
Landing Outcome of Launch Sites

- Different landing outcomes visualized:
  - Green: Success
  - Red: Fail
- Most successful landing in KSC LC-29A

Back to Outline

Back to Results

Proximity of Landmarks to KSC LC-39A

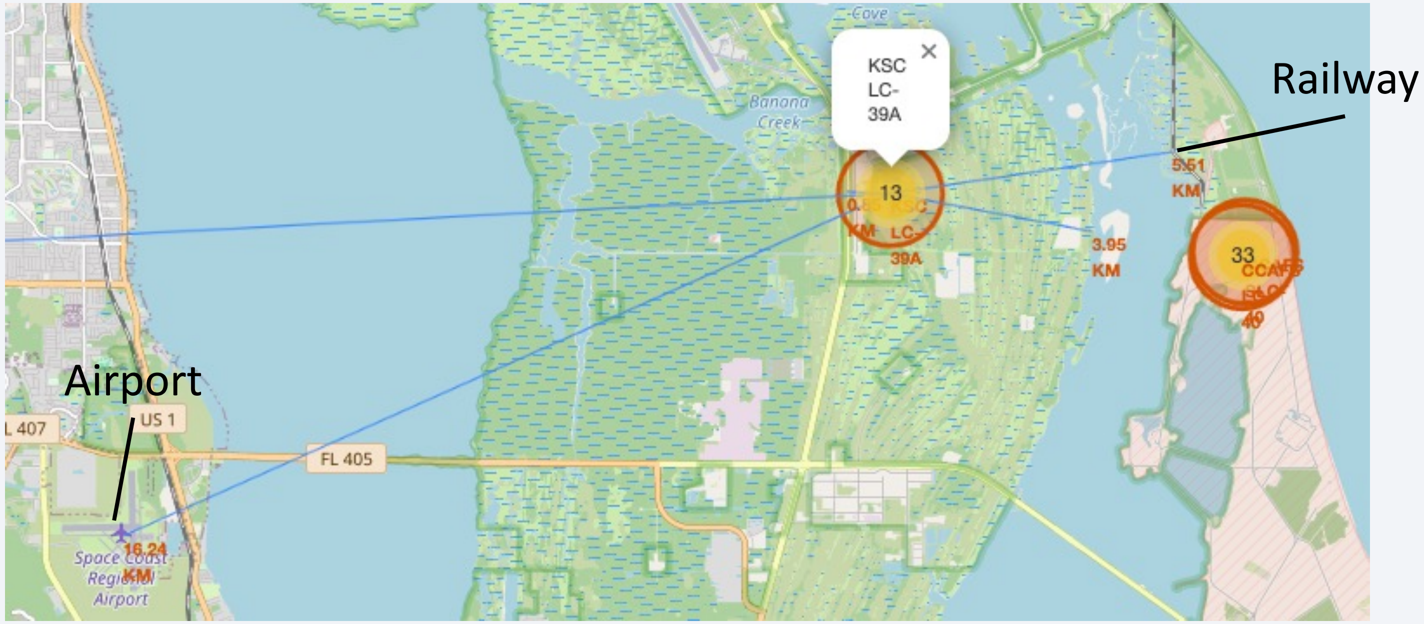


- Close to highway (0.85 km) could be be beneficial for easy accessibility to the launch site
- Close to coastline (3.95 km) provide a clear path over the sea and avoiding risk to populated areas

[Back to Outline](#)

[Back to Results](#)

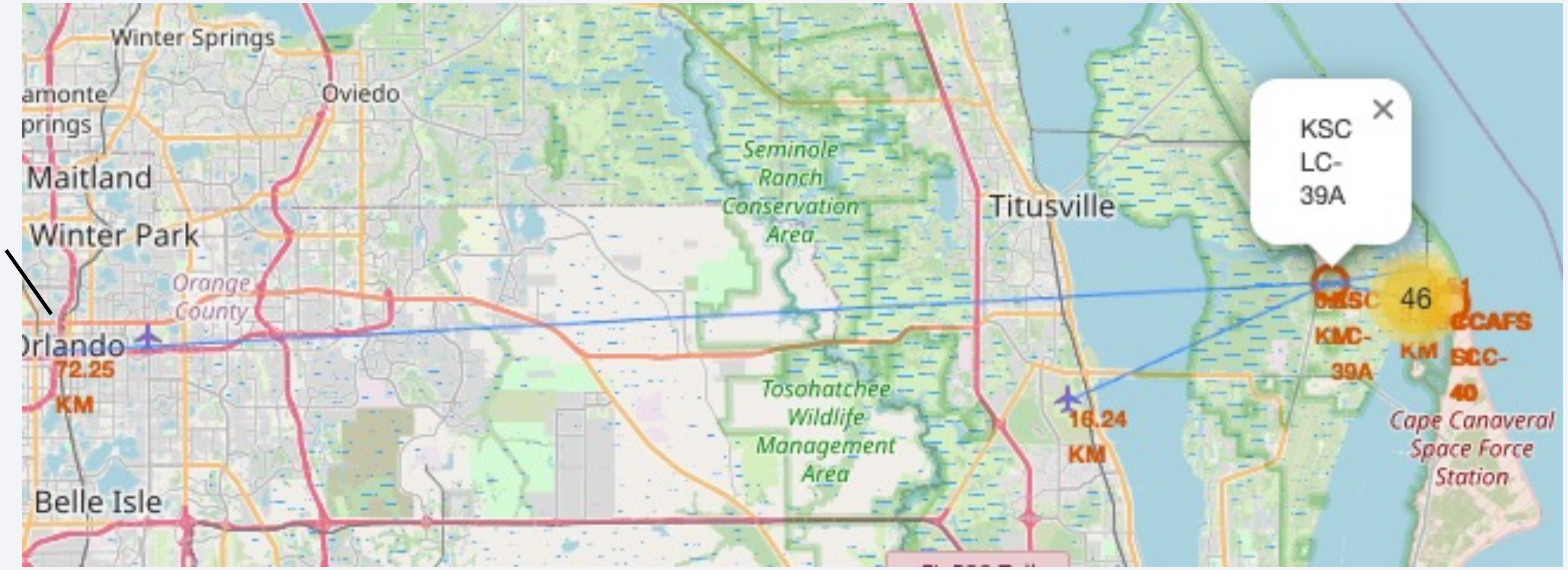
Proximity of Landmarks to KSC LC-39A



- Proximity to railway (5.61 km) could be useful for transporting heavy equipment and materials (CCAFS SLC-40)
- Far from airport (16.24 km) avoiding risk to populated areas and preventing interference with commercial air flights

Proximity of Landmarks to KSC LC-39A

City of
Orlando, FL

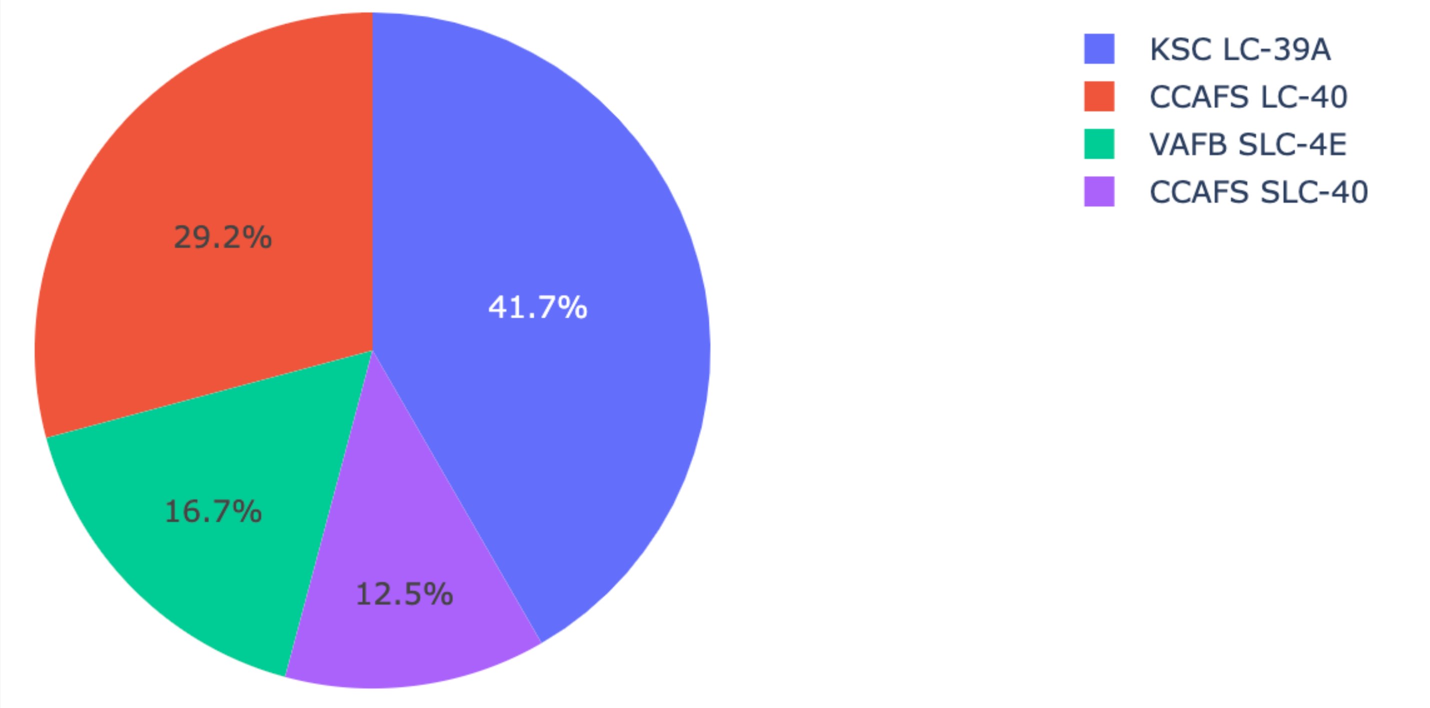


- Far from the city centers (72.25 km) to minimize potential damage or casualties in the event of a failed launch. It could also help in reducing noise pollution impact on populated areas

Section 4

Build a Dashboard with Plotly Dash

Launch Success Rate by Site

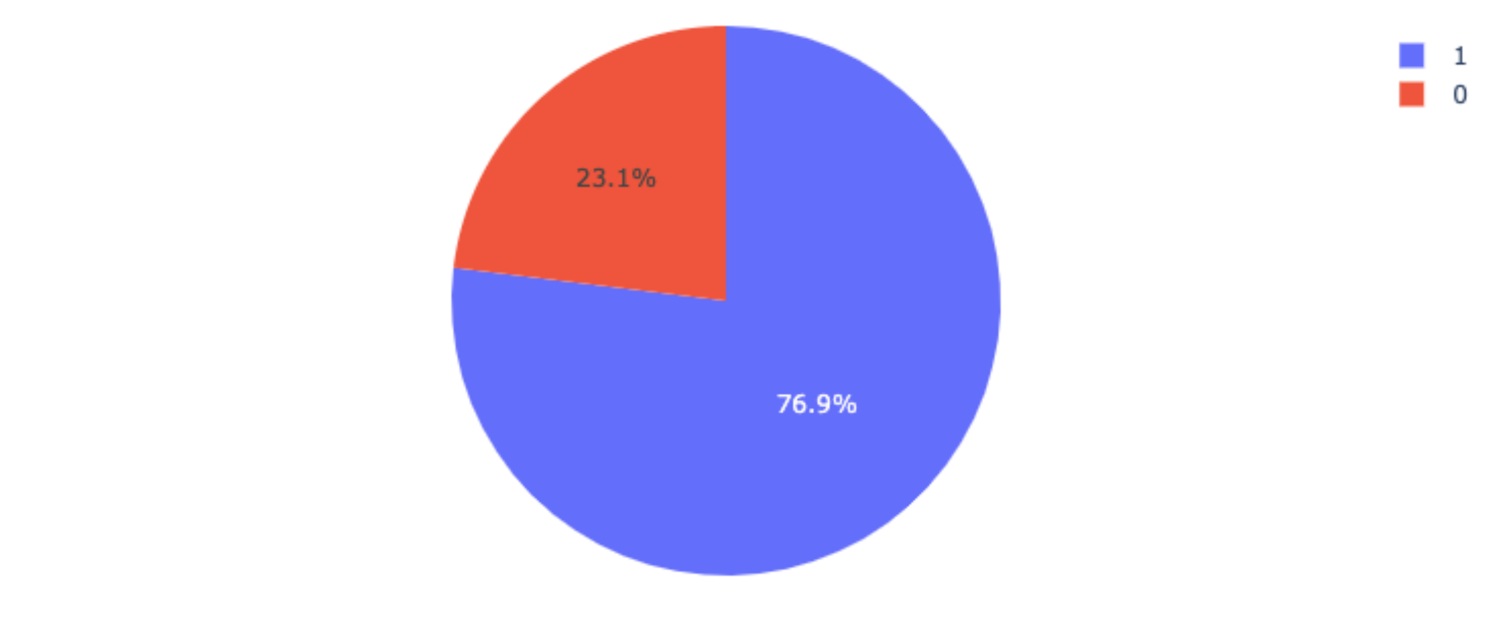


- KSC LC-39A site has the highest launch success rate among all sites

[Back to Outline](#)

[Back to Results](#)

Launch Success Rate of KSC LC-39A



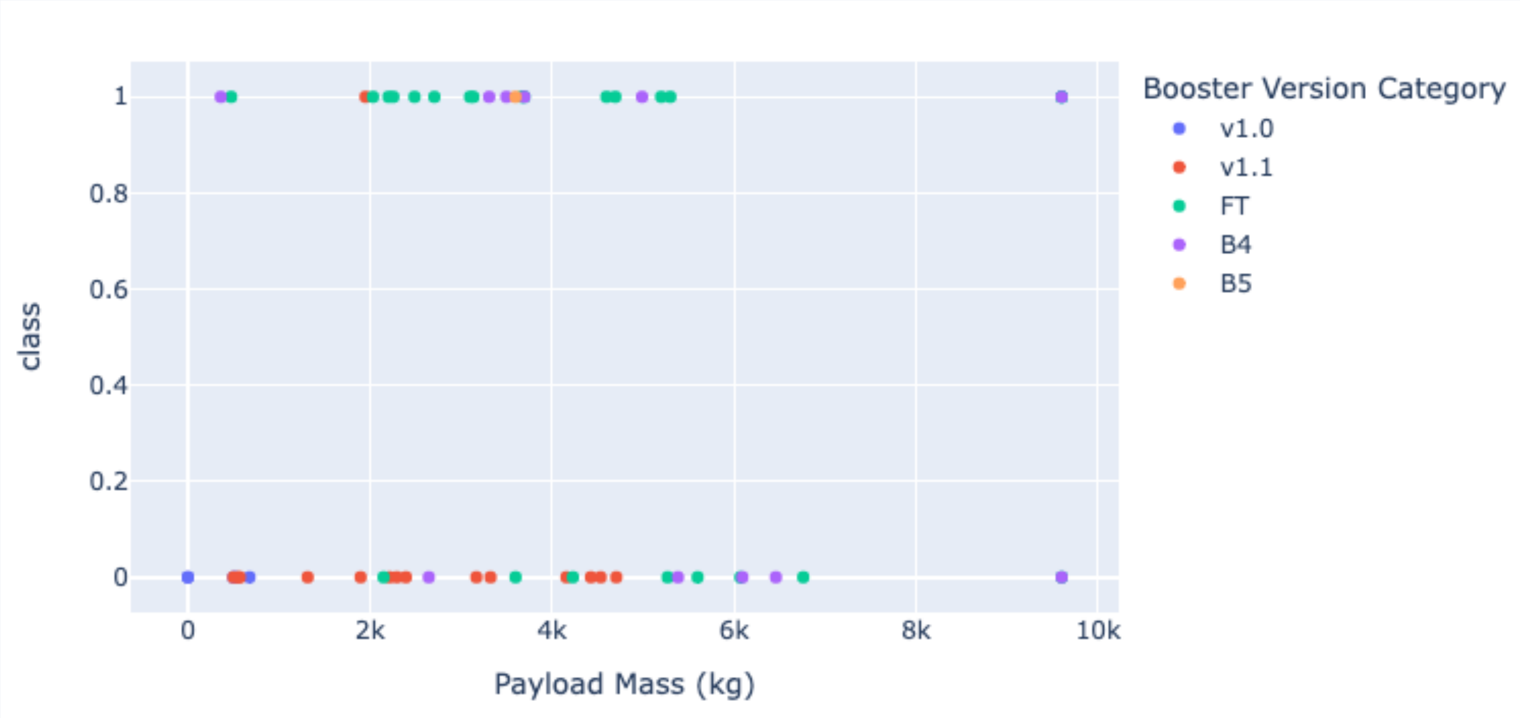
- Over 75% launch success rate

Plot of other launch sites in Appendix

[Back to Outline](#)

[Back to Results](#)

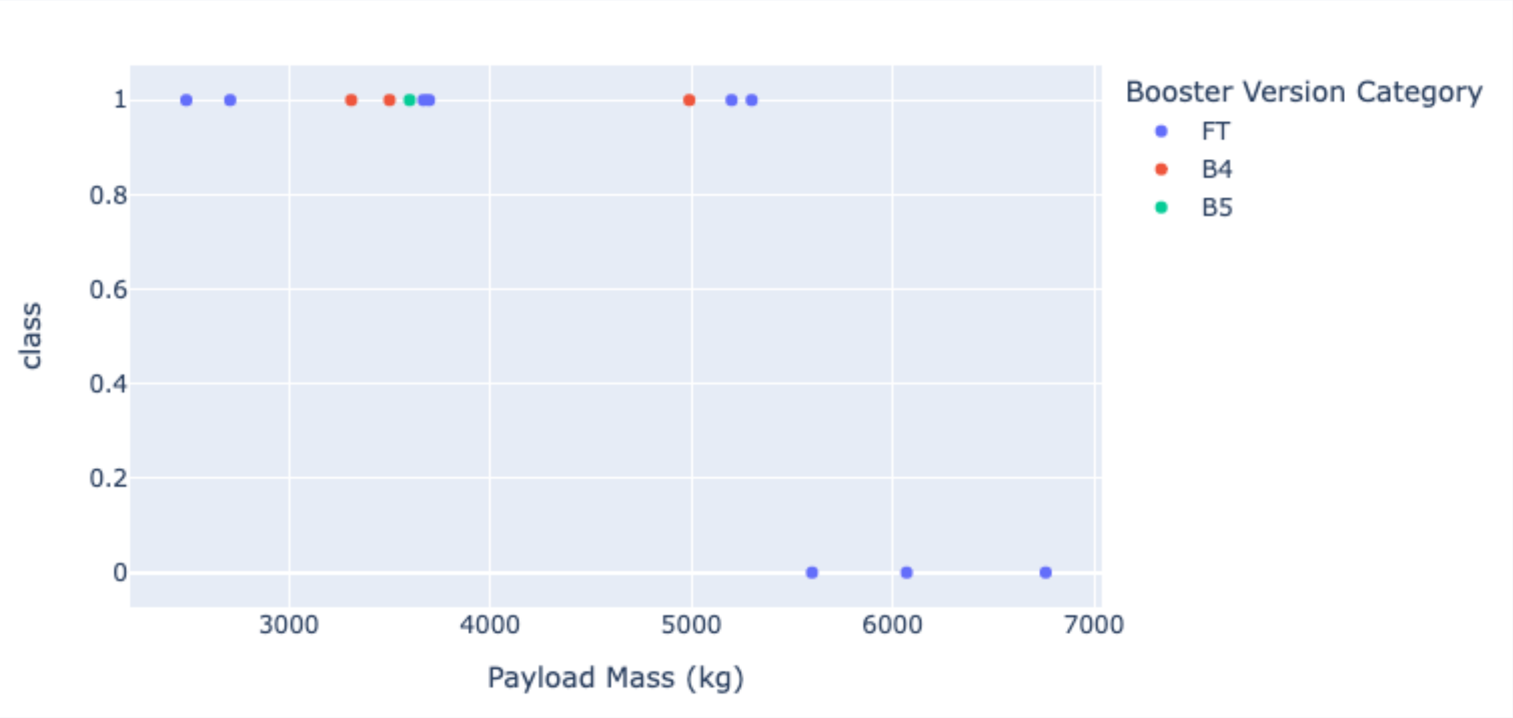
Correlation between Payload and Success Rate



- There does not seem to be any obvious correlation between payload mass with success rate
- FT is the most used booster version amongst successful launches, with most of the payload mass ranging somewhere between 2,000 – 5,000 kg

Back to Outline
Back to Results

Payload vs Success Rate in KSC LC-39A



- Higher payload mass (> 5,500 kg) are less likely to result in successful launch
- FT is the most used booster version

Section 5

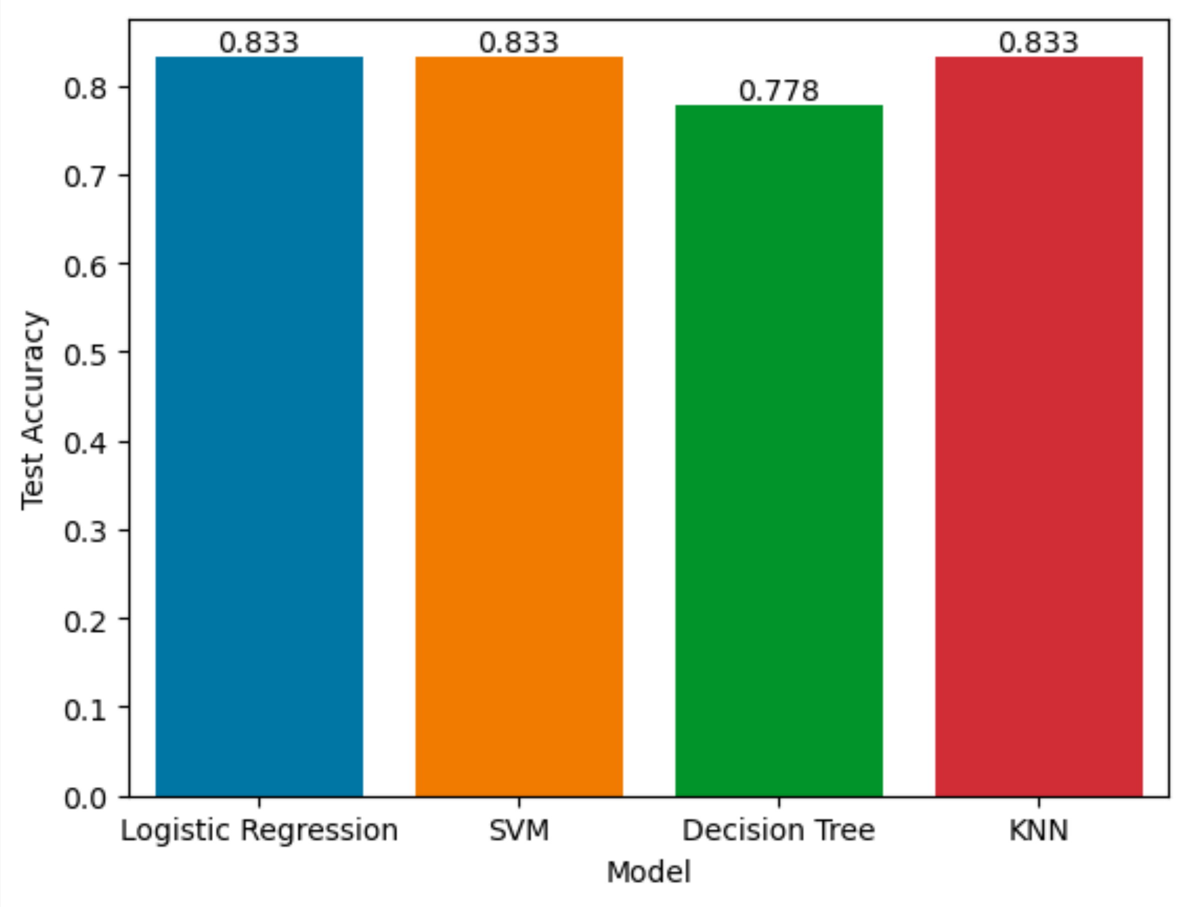
Predictive Analysis (Classification)

[Back to Outline](#)

[Back to Results](#)

Classification Accuracy

- Overall, the performance of the algorithms are competitive
- Decision Tree yields the lowest accuracy and taking the longest to train due to many numbers of making it a poor choice
- **Logistic Regression** is the simplest and yields same results as other more complex algorithms

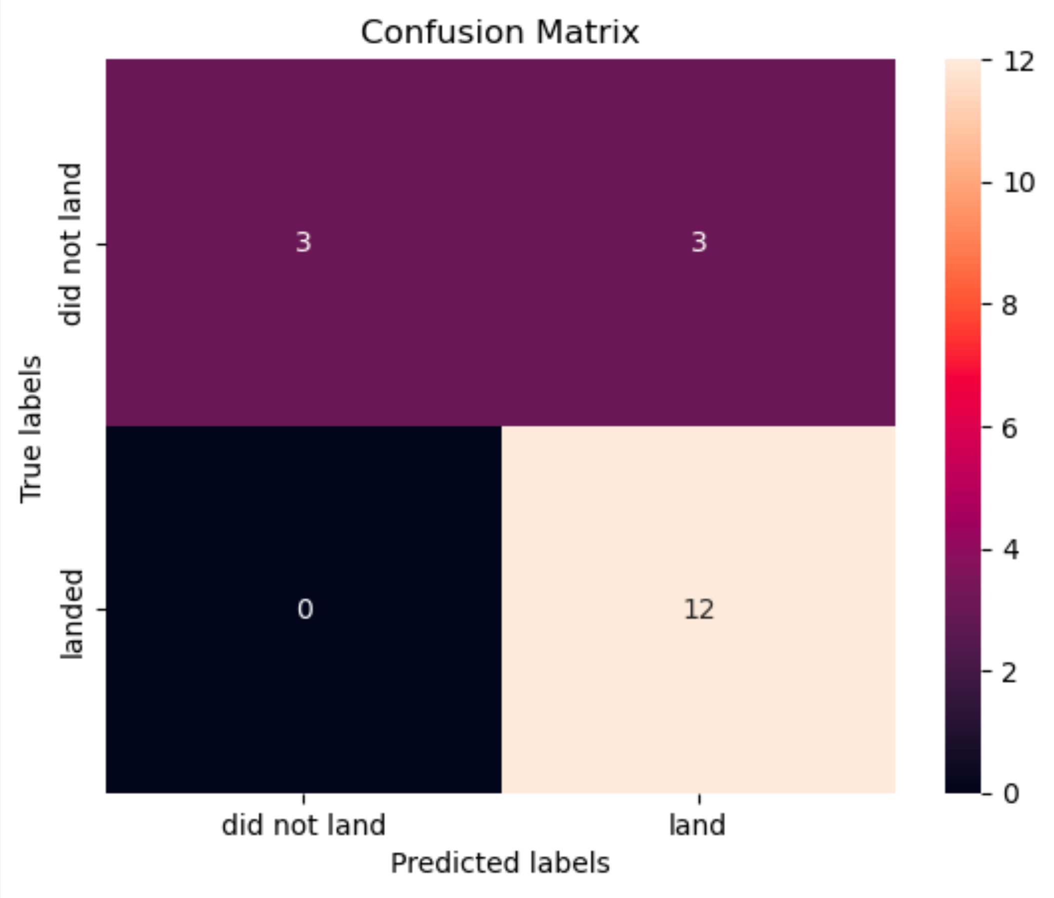


Back to Outline

Back to Results

Confusion Matrix

- Confusion matrix of Logistic Regression
  - (SVM and KNN also yield the same results)
- This algorithm (and all others) has a major **problem with false positives**



Conclusions

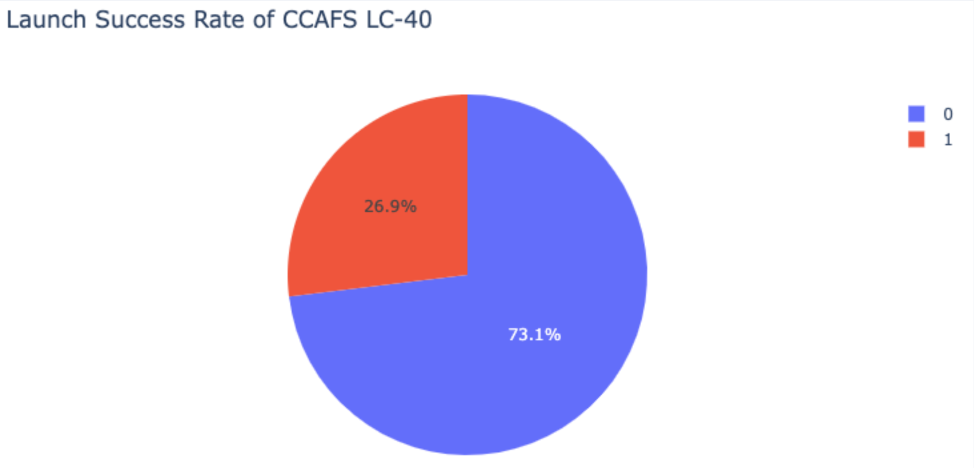
Key insights of SpaceX launch analysis

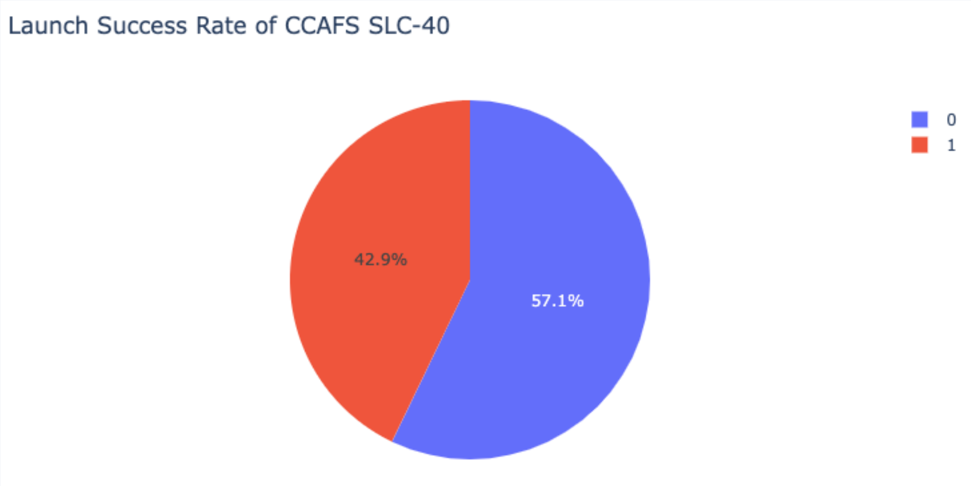
- **Technological Progress**: Launch success rates have risen from 2013 to 2020, underscoring advancements in technology and operations within the space sector.
- **Launch Site Importance**: Kennedy Space Center's LC-39A, with a success rate over 75%, highlights the critical role of selecting optimal launch sites.
- **Payload Complexity**: No direct link between payload mass and launch success was found, pointing to the intricate factors affecting launch dynamics.
- **Predictive Analytics**: Logistic Regression proved effective, achieving 83.3% accuracy, and underscored the value of predictive analytics in improving launch decisions.
- **Looking Ahead**: The future of space exploration will lean heavily on innovation and data analytics to boost mission success and efficiency.

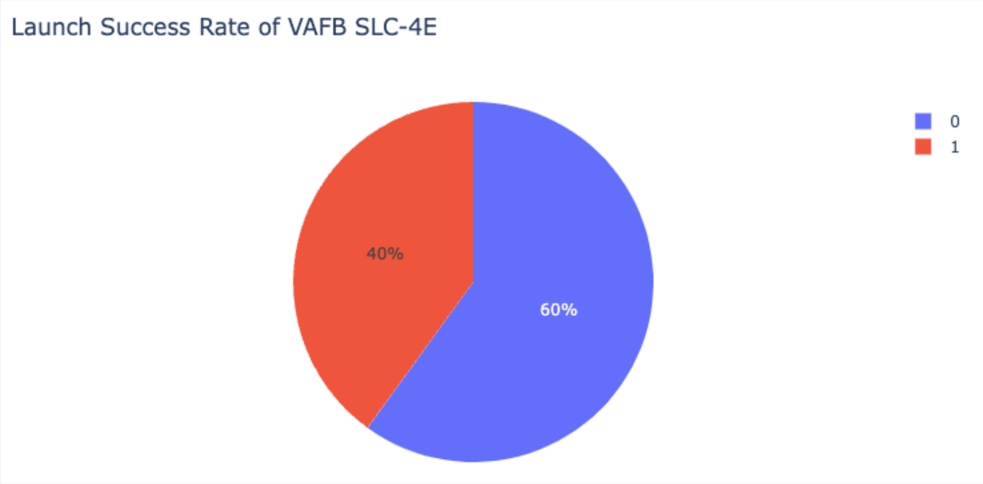
Appendix

Back to Outline

Launch Success Rate of Other Sites

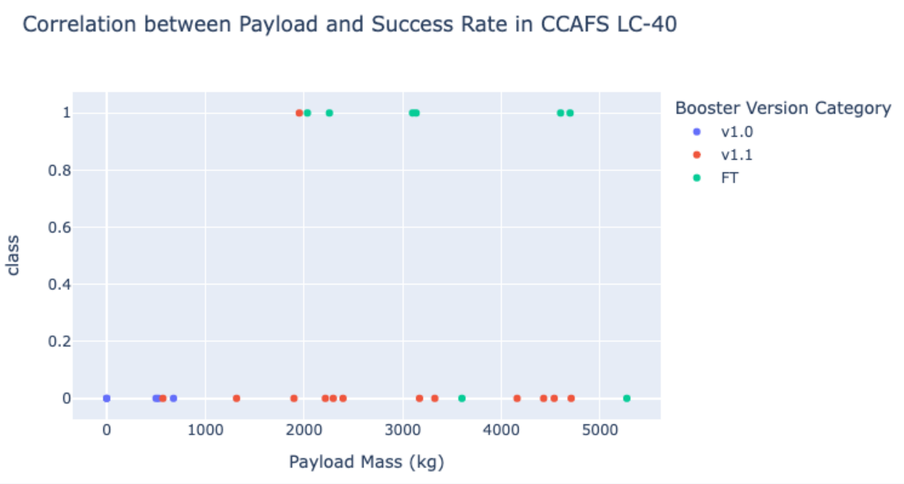


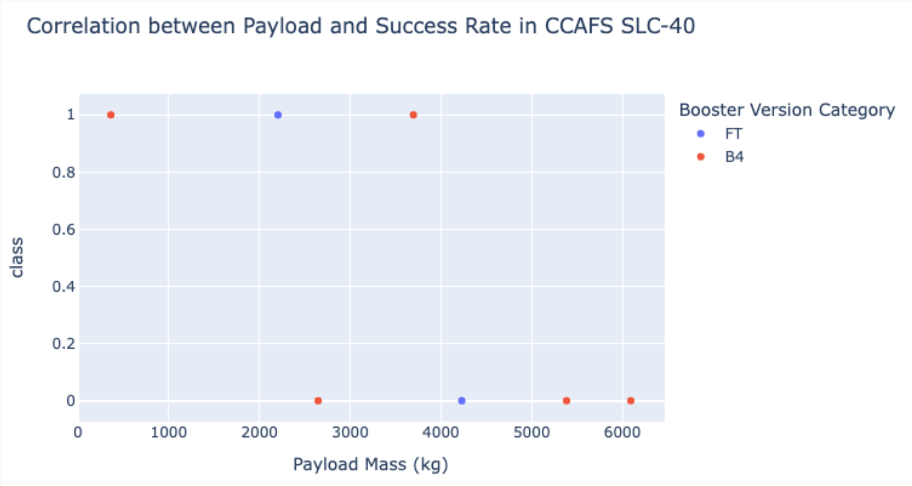


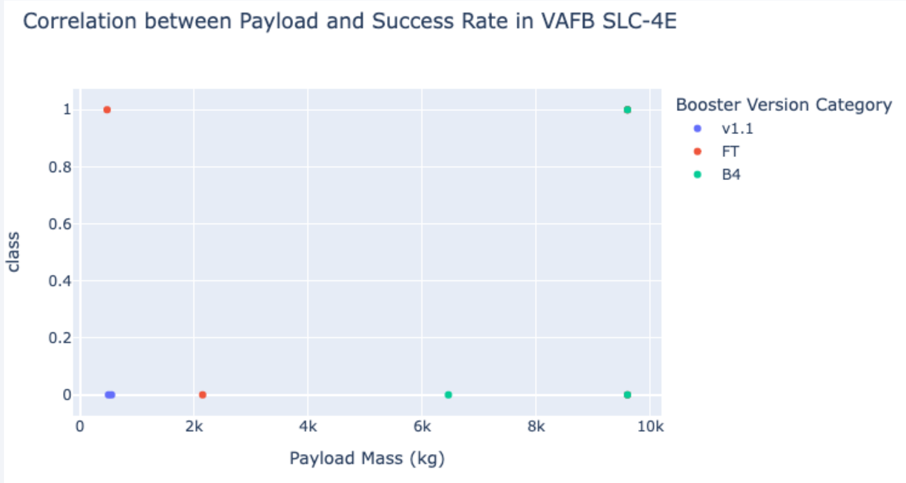


Back to Outline

Payload Mass vs Success Rate in Other Sites



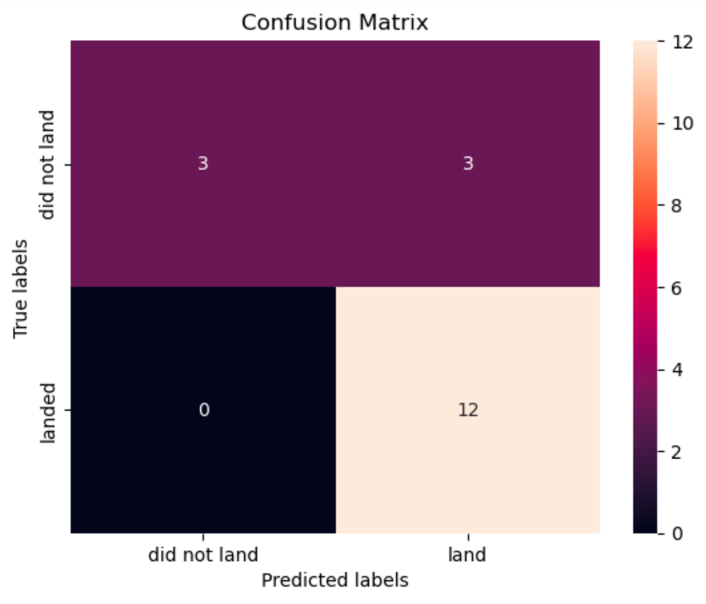




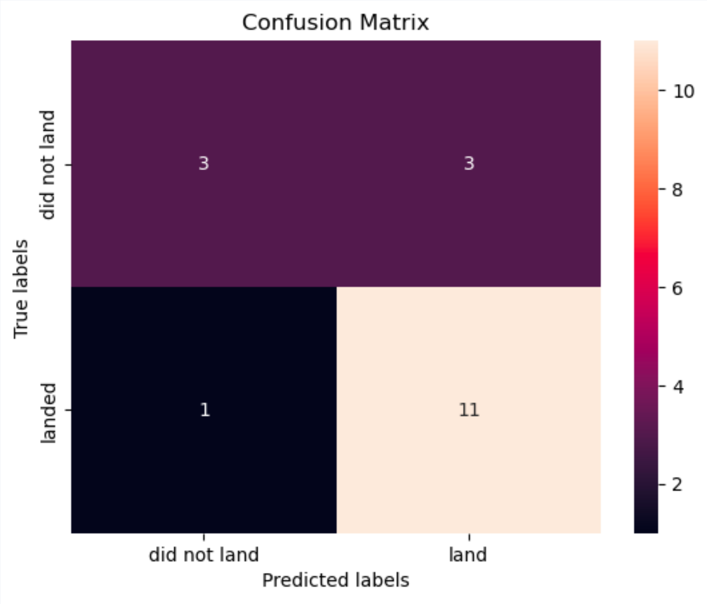
Results of Predictive Analysis

Model	Best Hyperparameters	Training Accuracy	Test Accuracy
Logistic Regression	{'C': 0.01, 'penalty': 'l2', 'solver': 'lbfgs'}	0.846	0.833
SVM	{'C': 1.0, 'gamma': 0.03162277660168379, 'kernel': 'sigmoid'}	0.848	0.833
Decision Tree	{'criterion': 'gini', 'max_depth': 6, 'max_features': 'sqrt', 'min_samples_leaf': 1, 'min_samples_split': 2, 'splitter': 'random'}	0.886	0.778
KNN	{'algorithm': 'auto', 'n_neighbors': 10, 'p': 1}	0.848	0.833

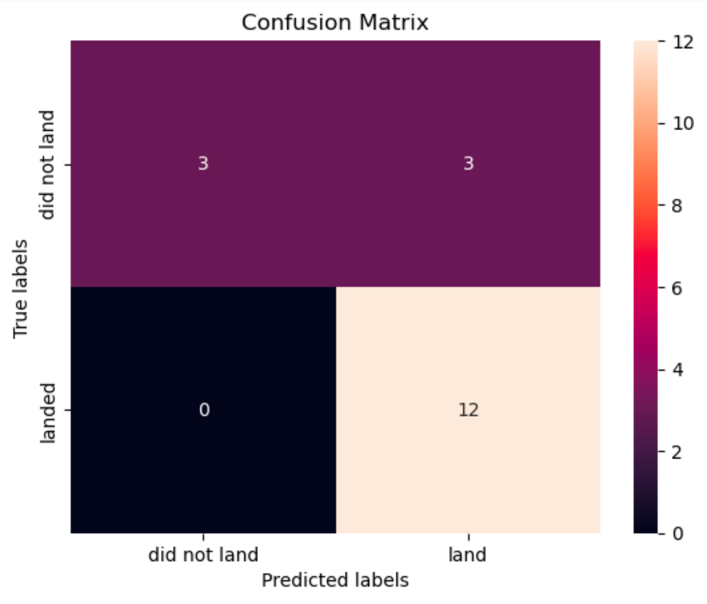
Confusion Matrix of Classifiers



SVM



Decision Tree



KNN

Thank you!