IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Jose Zacour

01/16/2024

Outline

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

Executive Summary

Summary of methodologies

- Data collection
- Data wrangling
- EDA with data visualization
- EDA with SQL
- Building an interactive map with Folium
- Building a Dashboard with Plotly Dash
- Predictive analysis (Classification)

Summary of all results

- EDA results
- Interactive analytics
- Predictive analysis

Introduction

Unveiling the New Era of Commercial Space Exploration

The dawn of commercial space travel is upon us, with visionary enterprises transforming celestial journeys into attainable adventures. Spearheading this movement, SpaceX offers its Falcon 9 rocket launches at a groundbreaking $62 million – a stark contrast to the typical market rates exceeding $165 million. The cornerstone of these remarkable savings lies in the innovative reuse of the rocket's first stage. Our project aims to delve into the analytics of first-stage reusability to forecast launch expenses with unprecedented precision.

Key Questions Driving Spaceflight Economics

Our investigation seeks to illuminate critical facets of spaceflight cost-efficiency:

- Can the Falcon 9's first-stage landing success be forecasted to inform industry benchmarks?
- Which launch sites have historically yielded the most favorable outcomes?
- What first-stage dimensions correlate with heightened success rates?
- How reliable are our predictions in identifying the most advantageous launch locations?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology
  - SpaceX API
  - Web Scrapping from Wikipedia
- Perform data wrangling
  - We categorized landing outcomes into binary training labels, where '1' signifies a triumphant first-stage landing and '0' denotes a landing attempt that didn't meet our success criteria.
  - Concurrently, we purged null entries and discarded columns that do not contribute to the predictive power of our model.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - LR, KNN, SVM, DT models have been built and evaluated for the best classifier

Data Collection

Datasets were collected using SpaceX API @ https://api.spacexdata.com/v4/

IMPORT LIBRARIES → REQUEST GET API / JSON LOADS / NORMALIZE JSON → CONSTRUCT OUR DATASET USING THE DATA JSON WE HAVE OBTAINED → FILTER THE DATAFRAME TO ONLY INCLUDE FALCON 9 LAUNCHES

Datasets were collected using Wikipedia @https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches

IMPORT LIBRARIES → REQUEST THE FALCON9 LAUNCH WIKI PAGE FROM ITS URL → EXTRACT ALL COLUMN / VARIABLE NAMES FROM THE HTML TABLE HEADER → CREATING A DATAFRAME BY PARSING THE LAUNCH HTML TABLES

→ DATA READY FOR WRANGLING

Data Collection – SpaceX API

1. IMPORT LIBRARIES:

```
# Requests allows us to make HTTP requests which we will use to get data from an API
import requests
# Pandas is a software library written for the Python programming language for data ma
import pandas as pd
# NumPy is a library for the Python programming language, adding support for large, mu
import numpy as np
# Datetime is a library that allows us to represent dates
import datetime
```

2. REQUEST GET API / JSON LOADS/ NORMALIZE JSON:

```
# Use json_normalize meethod to convert the json result into a dataframe
import json
response = requests.get(static_json_url)
data_dict = json.loads(response.content)
data = pd.json_normalize(data_dict)
```

3. CONSTRUCT DATASET USING JSON

```
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}
```

4. FILTER DATAFRAME ONLY INCLUDE FALCON 9 LAUNCH:

```
# Hint data['BoosterVersion']!='Falcon 1'
data_falcon9 = data[data['BoosterVersion'] != 'Falcon 1']
data_falcon9.head()
```

5. DATA READY FOR WRANGLING

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


Files sources: [IBM_DS_Capstone/IMB_DS_Capstone_Lab1A.ipynb at main · jezacour/IBM_DS_Capstone (github.com)]

Data Collection – Scraping

1. IMPORT LIBRARIES:

```
import sys

import requests
from bs4 import BeautifulSoup
import re
import unicodedata
import pandas as pd
pd.set_option('display.max_columns', None)
# Setting this option will print all of the data in a feature
pd.set_option('display.max_colwidth', None)
```

2. REQUEST THE FALCON9 LAUNCH WIKI PAGE FROM ITS URL

```
# use requests.get() method with the
# assign the response to a object
response = requests.get (static_url)
print(response.status_code)
print(response.content [0:100])
```

```
# Use BeautifulSoup() to create a BeautifulSoup object
soup = BeautifulSoup(response.content, "html.parser")
```

3. EXTRACT ALL COLUMNS

```
column_names = []
for row in first_launch_table.find_all('th'):
    name = extract_column_from_header(row)
    if name != None and len(name) > 0:
        column_names.append(name)
```

4. CREATING A DATAFRAME BY PARSING THE LAUNCH HTML TABLES

```
launch_dict= dict.fromkeys(column_names)

# Remove an irrelvant column
del launch_dict['Date and time ( )']

# Let's initial the launch_dict with each v
launch_dict['Flight No.'] = []
launch_dict['Launch site'] = []
launch_dict['Payload'] = []
launch_dict['Payload mass'] = []
launch_dict['Orbit'] = []
launch_dict['Customer'] = []
launch_dict['Launch outcome'] = []
# Added some new columns
launch_dict['Version Booster']=[]
launch_dict['Booster landing']=[]
launch_dict['Date']=[]
launch_dict['Time']=[]
```

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

5. DATA READY FOR WRANGLING

```
df.to_csv('spacex_web_scraped.csv', index=False)
```


Files sources: [IBM_DS_Capstone/IMB_DS_Capstone_Lab1B.ipynb at main · jezacour/IBM_DS_Capstone (github.com)]

Data Wrangling

Check null values → Calculate the number of launches on each site → Calculate the number and occurence of each orbit → Calculate the number and occurence of mission outcome per orbit type → Create a landing outcome label from Outcome comumn → Handle null values

```
# Apply value_counts() on column
df['LaunchSite'].value_counts()
```

```
CCAFS SLC 40    55
KSC LC 39A      22
VAFB SLC 4E     13
```

```
# Apply value_counts on Orbit column
df['Orbit'].value_counts()
```

```
GTO      27
ISS      21
VLEO     14
PO        9
LEO       7
SSO       5
MEO       3
ES-L1     1
HEO       1
SO        1
GEO       1
Name: Orbit, dtype: int64
```

```
landing_outcomes = df['Outcome'].value_counts()
landing_outcomes
```

```
True ASDS      41
None None      19
True RTLS      14
False ASDS      6
True Ocean      5
False Ocean     2
None ASDS       2
False RTLS      1
Name: Outcome, dtype: int64
```

```
0 True ASDS
1 None None
2 True RTLS
3 False ASDS
4 True Ocean
5 False Ocean
6 None ASDS
7 False RTLS
```

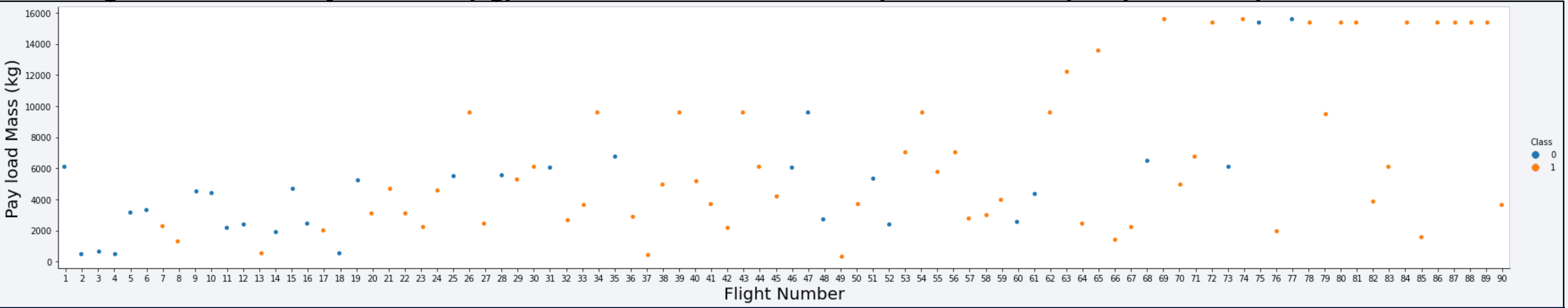
```
df['Class']=landing_class
df[['Class']].head(8)
```

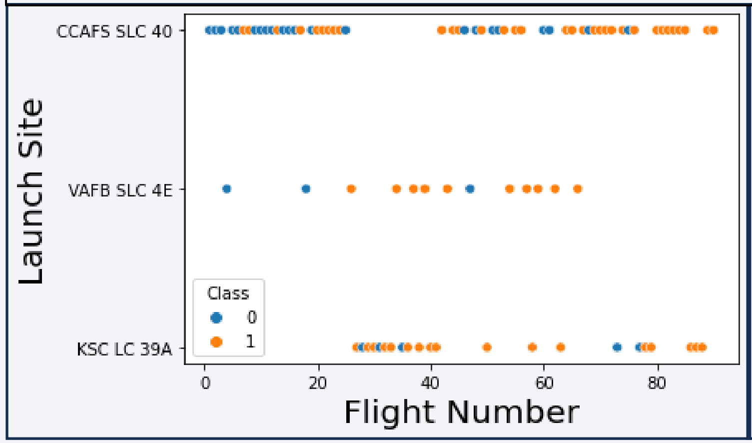
	Class
0	0
1	0
2	0
3	0
4	0
5	0
6	1
7	1

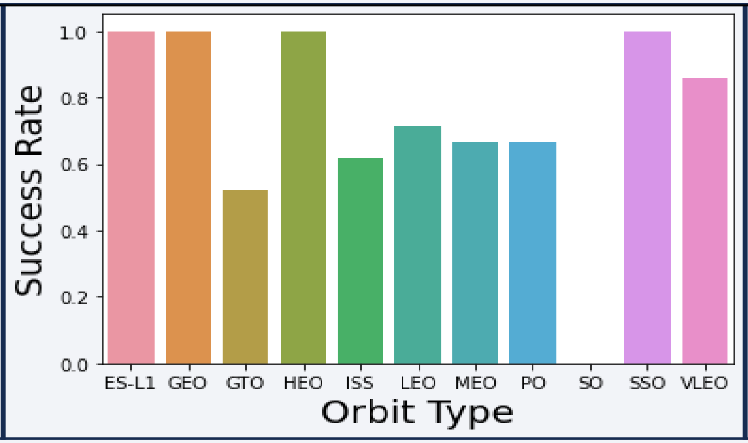
Files sources: [IBM_DS_Capstone/IMB_DS_Capstone_Lab2.ipynb at main · jezacour/IBM_DS_Capstone (github.com)]()

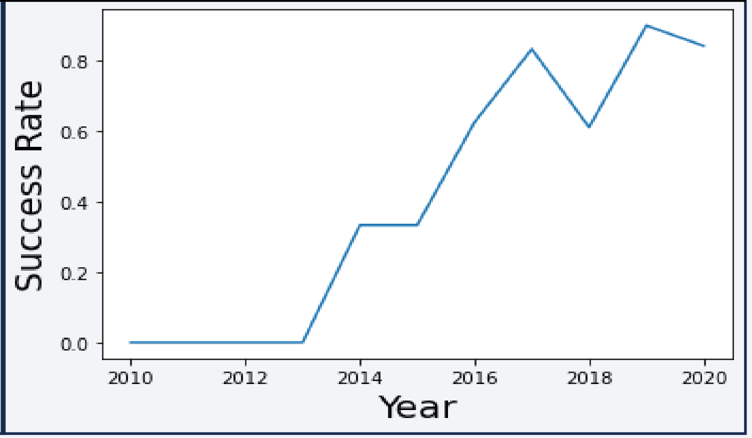
EDA with Data Visualization

Plot FlightNumber vs. PayloadMass (kg) with the launch outcome ● (Not Successful) ● (Successful)









Flight Number vs. Launch Site

Launch Outcomes

- Not Successful
- Successful

Success rate of each orbit type

Launch success yearly trend

Files sources: IBM_DS_Capstone/IBM_DS_Capstone_Vis.ipynb at main · jezacour/IBM_DS_Capstone (github.com)

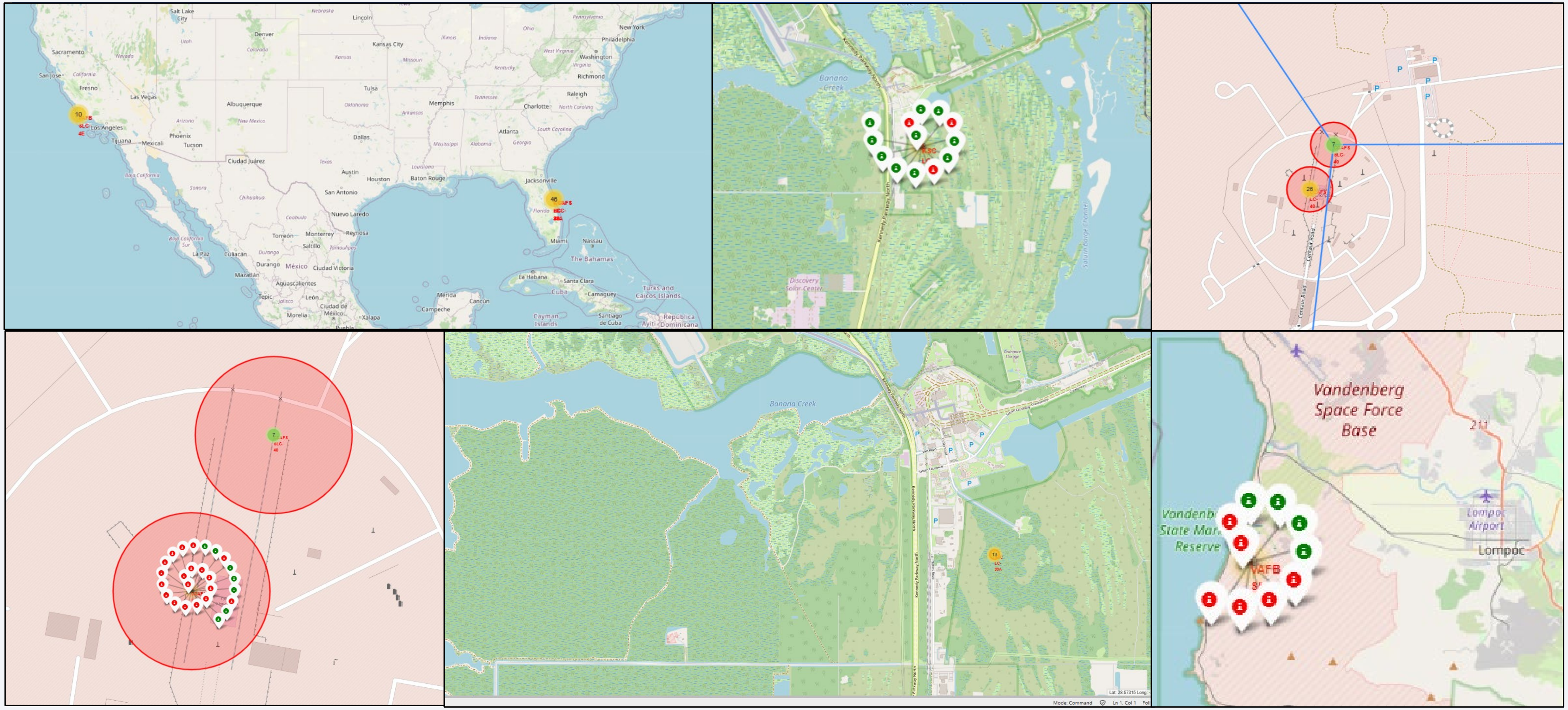
EDA with SQL

SQL queries to solve the assignment tasks:

- 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 achieved
- 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. Use a subquery
- 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

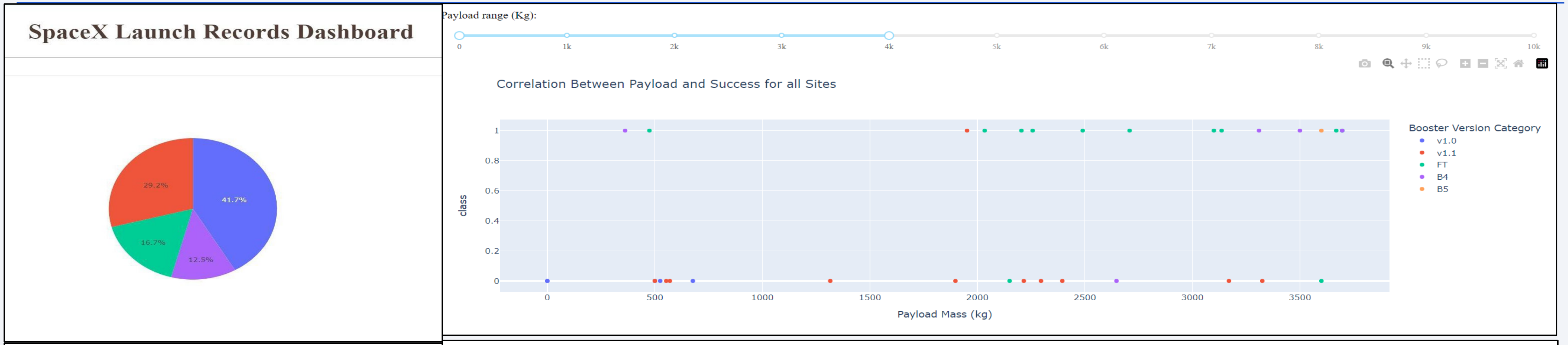
Files sources: [IBM_DS_Capstone/IMB_DS_Capstone_Lab3.ipynb at main · jezacour/IBM_DS_Capstone (github.com)]

Build an Interactive Map with Folium



Files sources: [IBM_DS_Capstone/IMB_DS_Capstone_Folium.ipynb at main · jezacour/IBM_DS_Capstone (github.com)]()

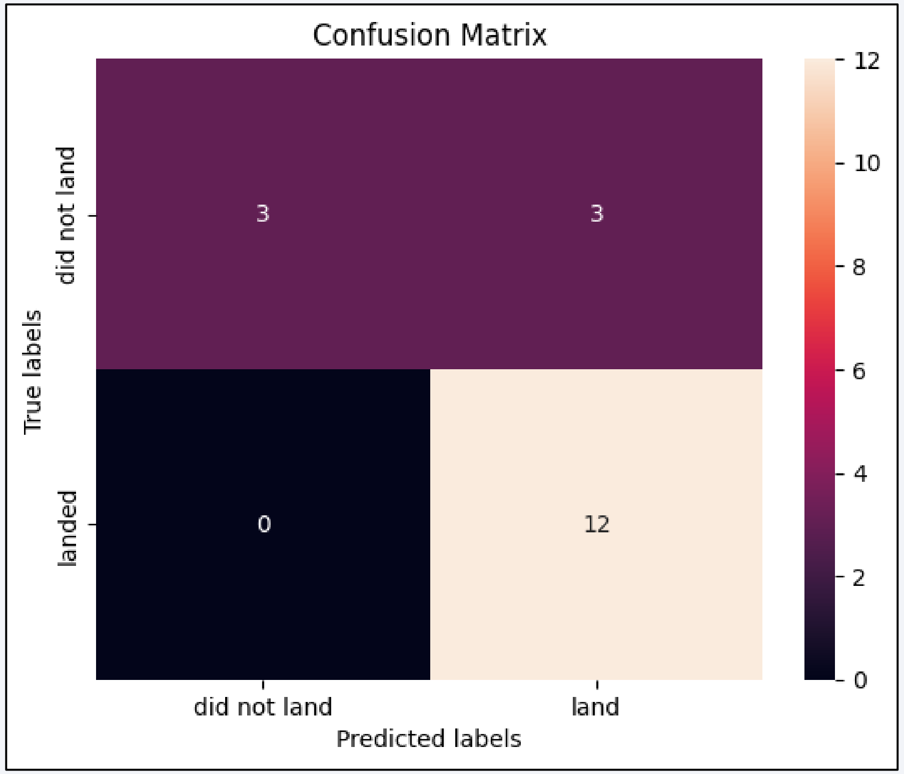
Build a Dashboard with Plotly Dash

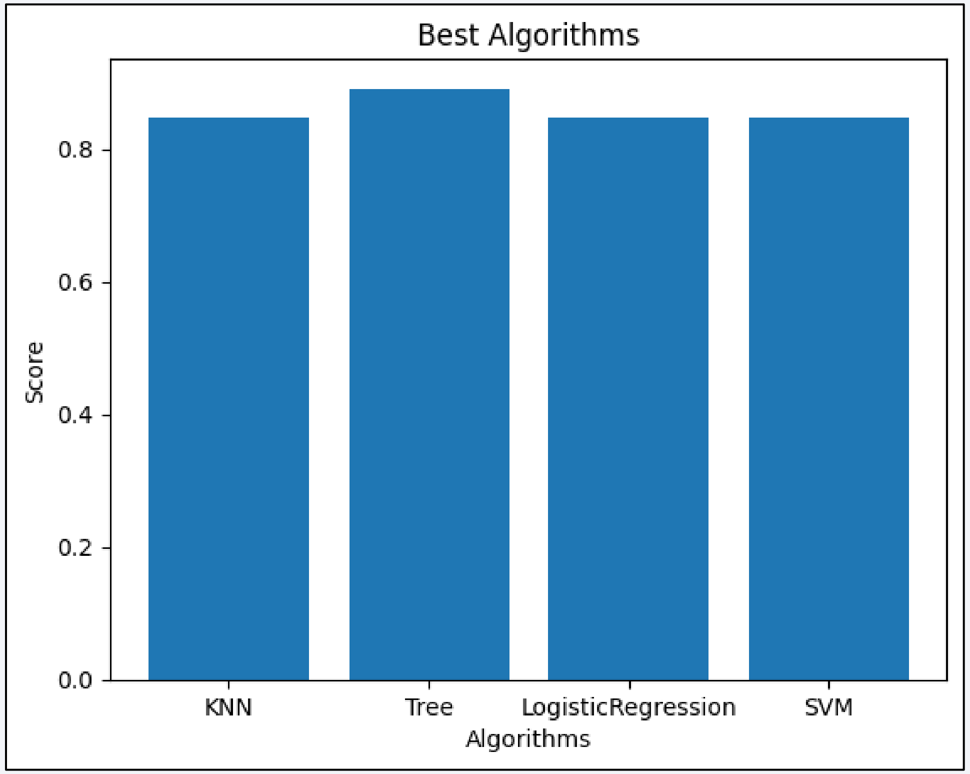


Files sources: [IBM_DS_Capstone/IBM_DS_Capstone_Dash.py at main · jezacour/IBM_DS_Capstone (github.com)]

Predictive Analysis (Classification)

The accuracy of the test data is 83.3%
The best algorithm is Decision Tree with a score of 90.3%





Files sources: IBM_DS_Capstone/IMB_DS_Capstone_ML.ipynb at main · jezacour/IBM_DS_Capstone (github.com)

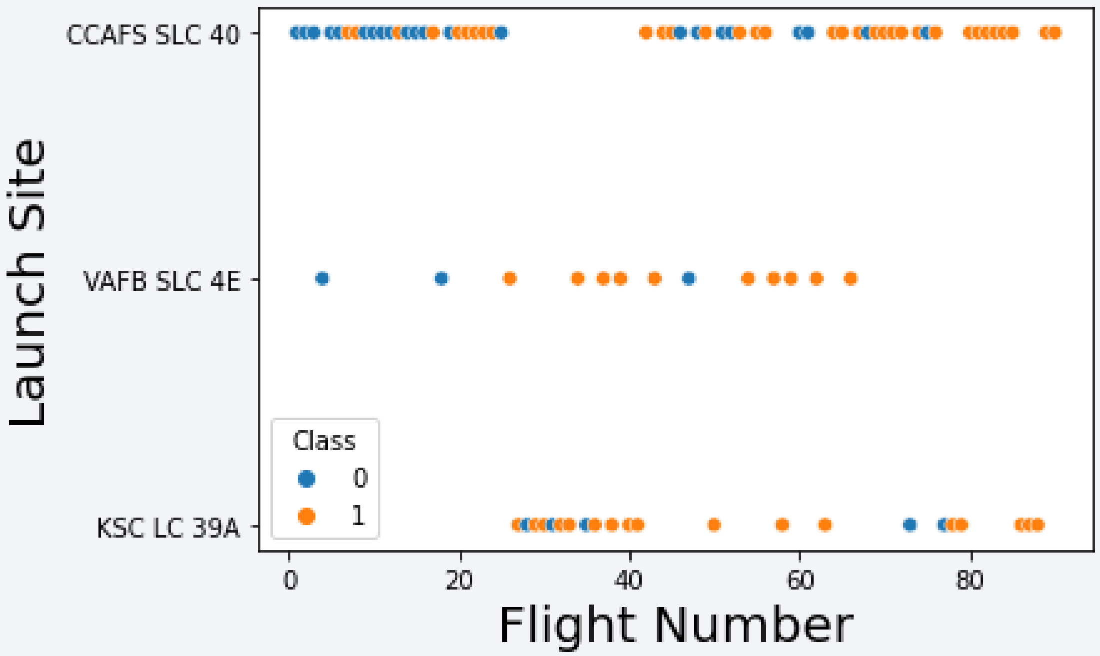
Results

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

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site



Main observations

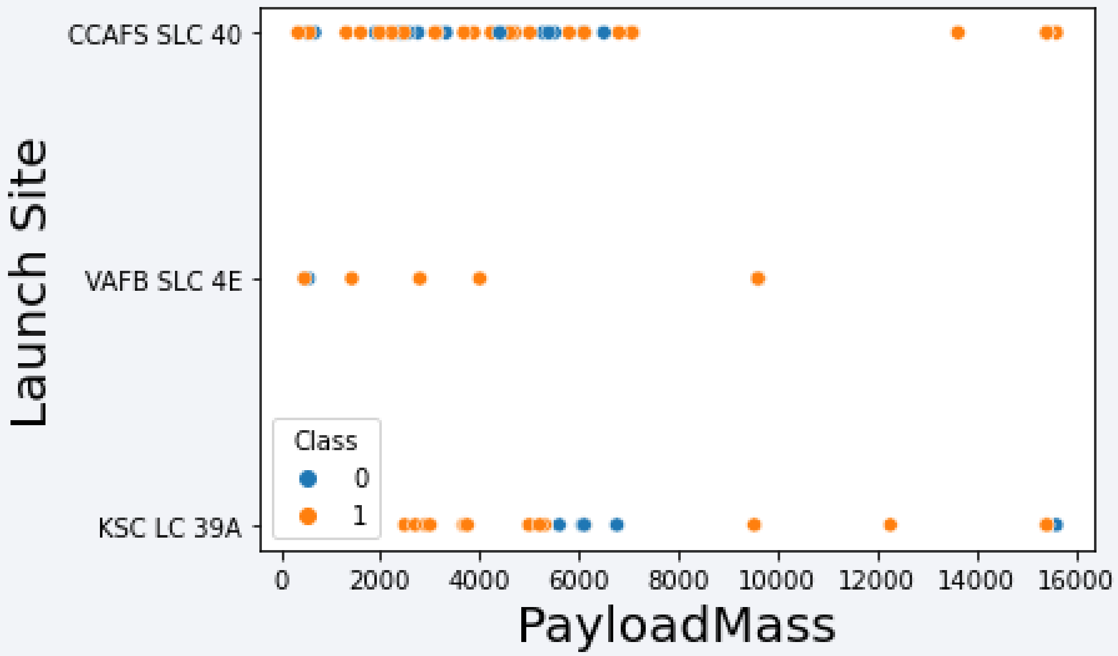
- The initial launches in the series took off from the Cape Canaveral Air Force Station's Space Launch Complex 40 (CCAFS-SLC-40), with subsequent missions launching from Kennedy Space Center's LC-39A (KSC-LC-39A).

- The majority of the missions have embarked from CCAFS-SLC-40, establishing it as the primary launch site.

- A smaller number of launches have originated from the Vandenberg Air Force Base at Space Launch Complex 4E (VAFB SLC 4E), highlighting its role as a secondary launch site.

- With accumulated launch experience at each site, we observe a corresponding ascent in the success rates

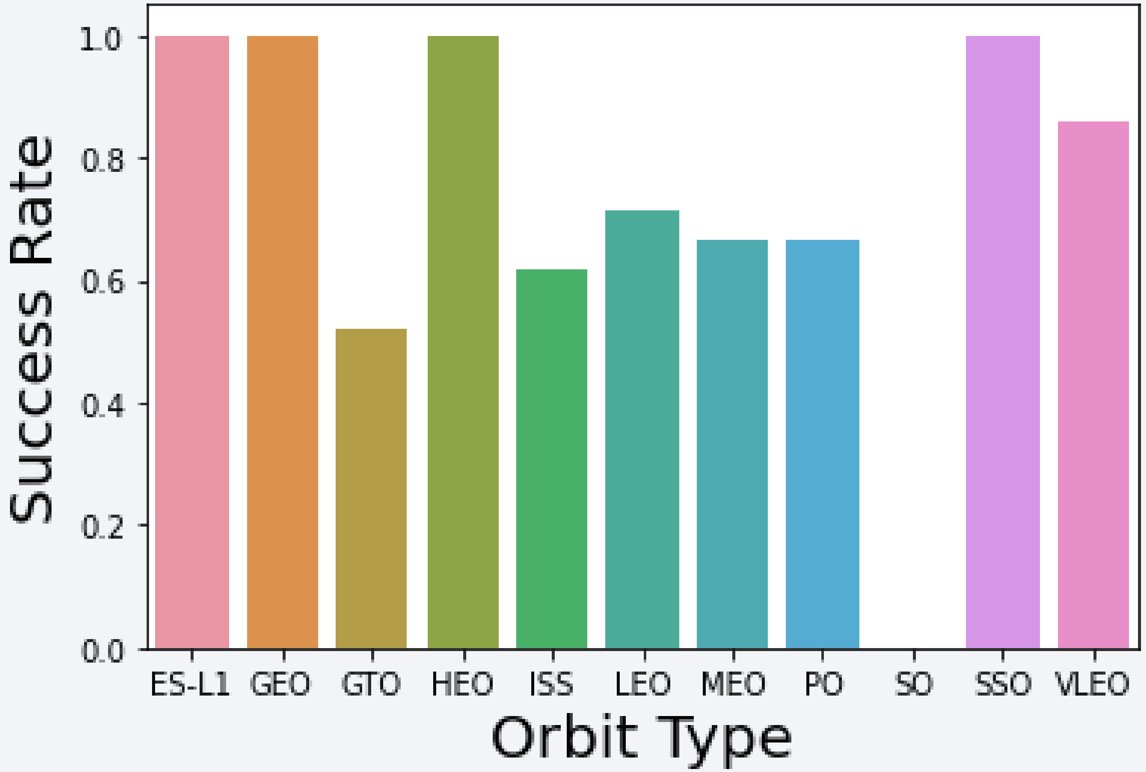
Payload vs. Launch Site



Main observations

- Payloads weighing between 2000-4000kg have achieved the highest success rate in launches.
- Payloads within the 4000-8000kg range have experienced lower success rates comparatively.
- Notably, payloads in the 8000-15000kg bracket have a 100% success rate, with no failed launches recorded.
- Space Launch Complex 40 at Cape Canaveral Air Force Station (CCAFS SLC 40) has been the predominant launch site for payloads on the lighter end of the mass spectrum.

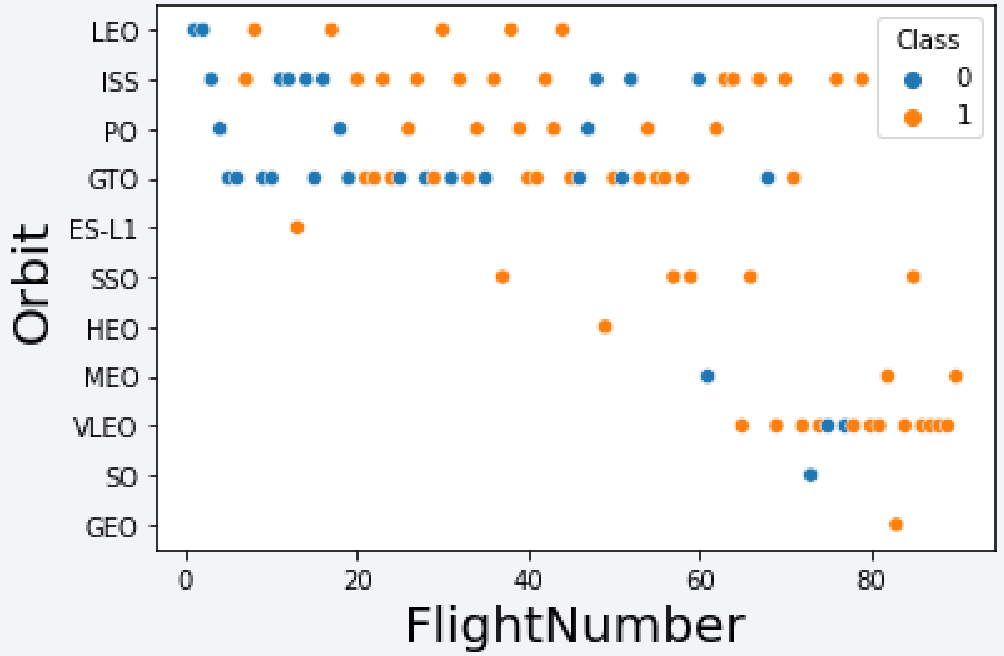
Success Rate vs. Orbit Type



Main observation

- The ES-L1, GEO, HEO, SSO orbit types have a 100% success rates

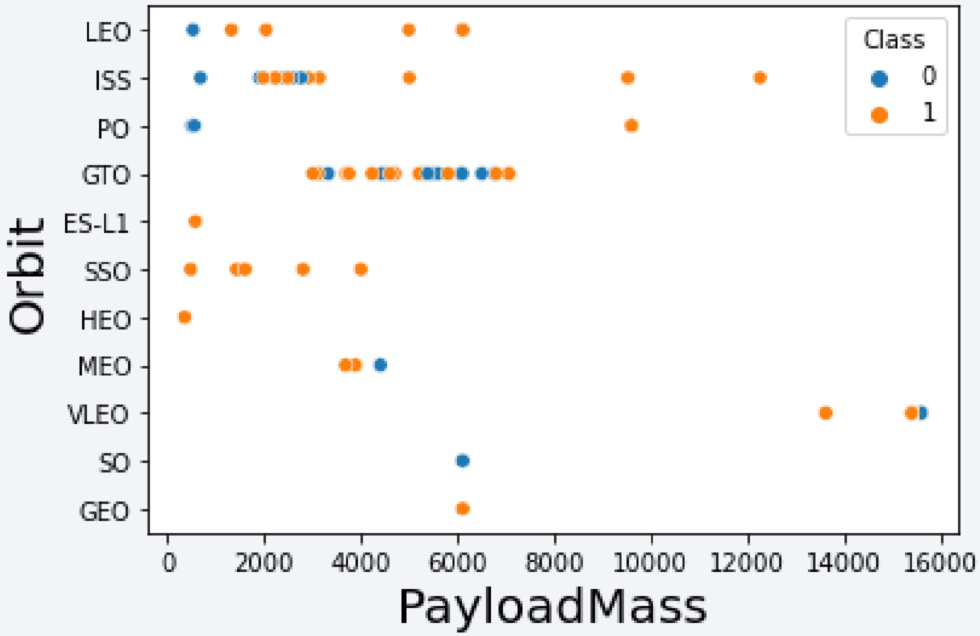
Flight Number vs. Orbit Type



Main observations

- A shift towards VLEO launches is demonstrably rising in recent years due to its close proximity to Earth.
- Experience has fueled an increase in LEO launches.
- GTO's erratic success rate warrants caution until causes are identified.

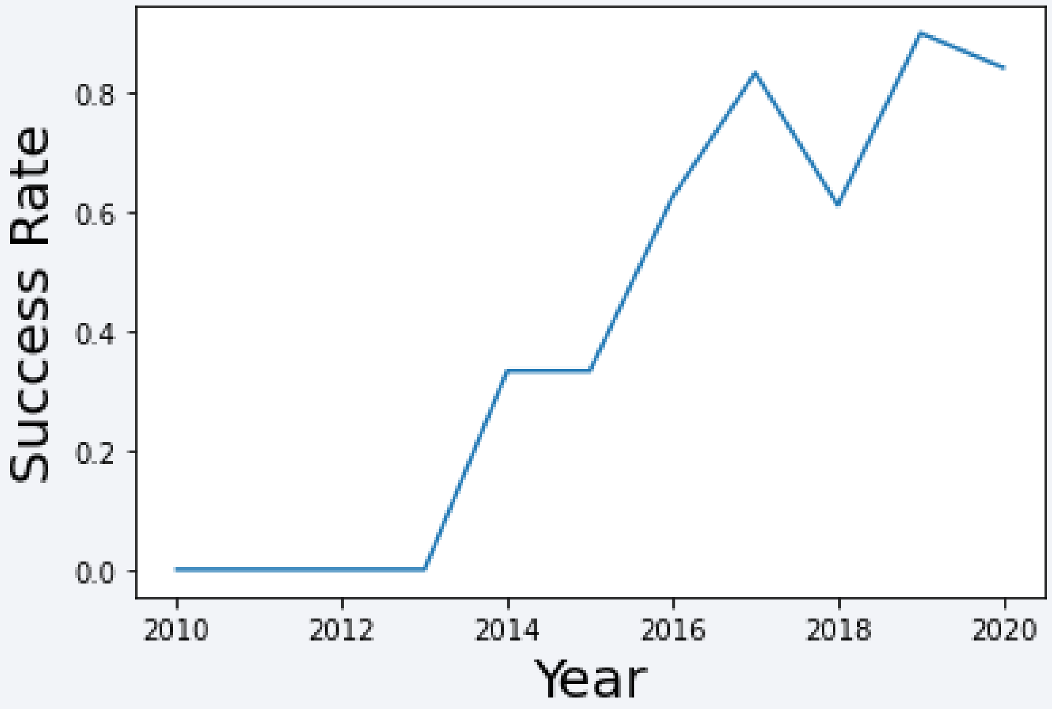
Payload vs. Orbit Type



Main observations

- LEO, ISS, and PO orbits with payloads above 4,000 kg have a 100% success rate.
- SSO orbits have a 100% success rate for payloads up to 4,000 kg.
- GTO orbits have unstable success rates regardless of payload weight.

Launch Success Yearly Trend



Main observation

- Since 2013, we have witnessed a marked enhancement in launch success rates, which have reached a plateau of stability post-2019. This trend likely reflects the cumulative benefits of technological advancements and the assimilation of insights from past experiences.

All Launch Site Names

```
%%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'

```
%%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
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success
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
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success

Total Payload Mass

```
%%sql
SELECT sum(payload_mass__kg_) AS "Total payload mass (NASA (CRS))" FROM SPACEXTBL WHERE customer = 'NASA (CRS)';
```

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

Total payload mass (NASA (CRS))
45596

Average Payload Mass by F9 v1.1

```
%%sql
SELECT AVG(payload_mass__kg_) AS "Average payload mass (booster version F9 v1.1)" FROM SPACEXTBL WHERE booster_version LIKE
```

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

Average payload mass (booster version F9 v1.1)
2534.6666666666665

First Successful Ground Landing Date

```
%%sql
SELECT min(DATE) AS "First successful landing outcome in ground pad" FROM SPACEXTBL WHERE Landing_Outcome = 'Success (ground
```

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

First successful landing outcome in ground pad
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%%sql
SELECT booster_version FROM SPACEXTBL WHERE Landing_Outcome = 'Success (drone ship)' AND payload_mass__kg_ BETWEEN 4000 AND
```

```
 * 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

```
sql SELECT MISSION_OUTCOME, COUNT(MISSION_OUTCOME) FROM SPACEXTBL WHERE MISSION_OUTCOME like 'Success%'
```

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

Mission_Outcome	COUNT(MISSION_OUTCOME)
Success	100

```
sql SELECT MISSION_OUTCOME, COUNT(MISSION_OUTCOME) FROM SPACEXTBL WHERE MISSION_OUTCOME = 'Failure (in flight)'
```

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

Mission_Outcome	COUNT(MISSION_OUTCOME)
Failure (in flight)	1

Boosters Carried Maximum Payload

```
%%sql
SELECT DISTINCT booster_version
FROM SPACEXTBL
WHERE payload_mass__kg_ = (
    SELECT max(payload_mass__kg_)
    FROM SPACEXTBL
)
```

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

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

2015 Launch Records

```
sql SELECT substr(Date, 4, 2) AS Month, substr(Date, 7, 4) as Year,BOOSTER_VERSION, "Landing _Outcome",launch_site FROM SPA(
```

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

Month	Year	Booster_Version	"Landing _Outcome"	Launch_Site

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

```
%%sql

SELECT Landing_Outcome, COUNT(Landing_Outcome) AS "Count"
FROM SPACEXTBL
WHERE (Landing_Outcome like 'Success%') AND DATE BETWEEN '2010-06-04' and '2017-03-20'
GROUP BY Landing_Outcome
ORDER BY Count DESC
;
```

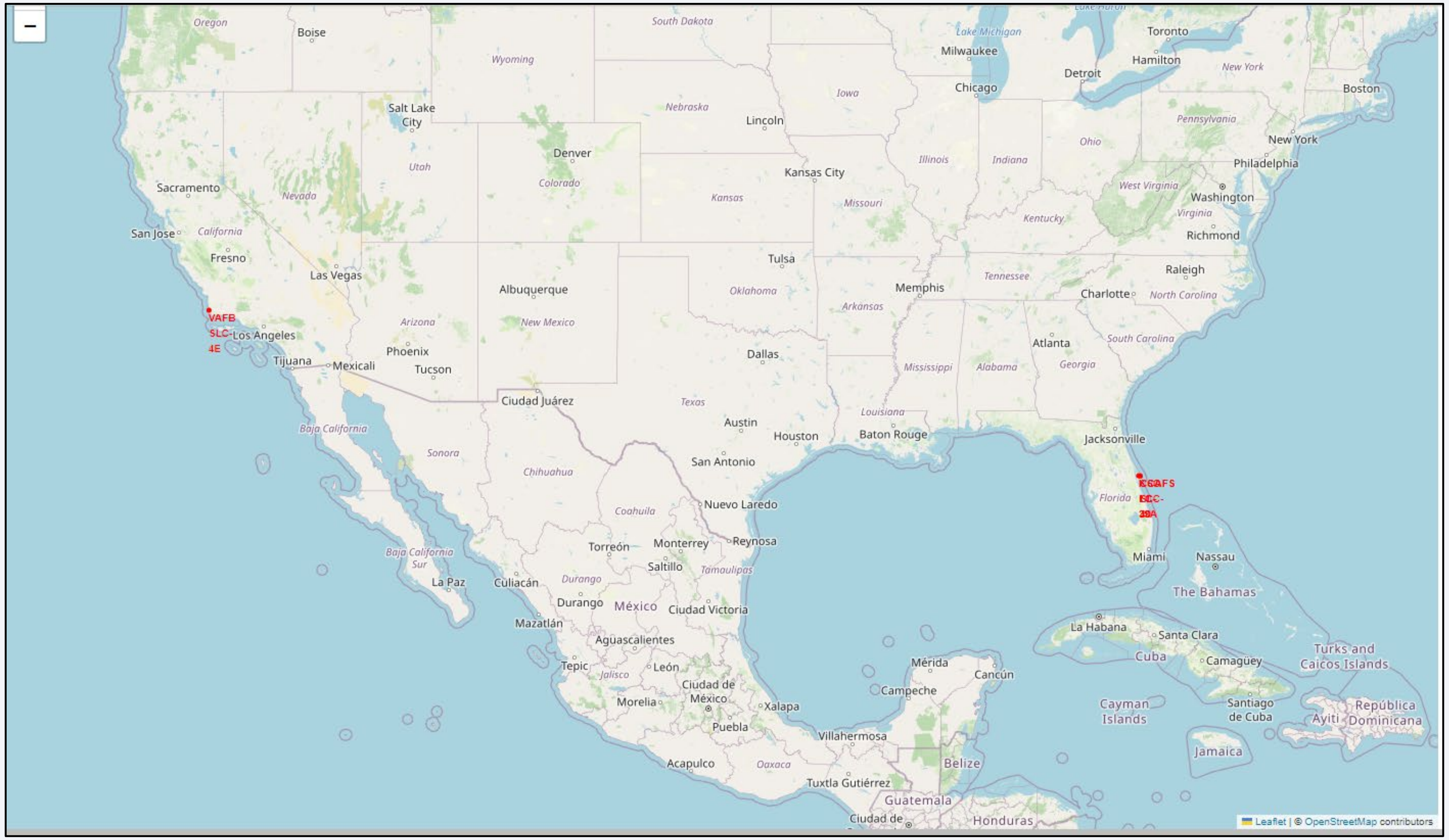
```
 * sqlite:///my_data1.db
Done.
```

Landing_Outcome	Count
Success (drone ship)	5
Success (ground pad)	3

Section 3

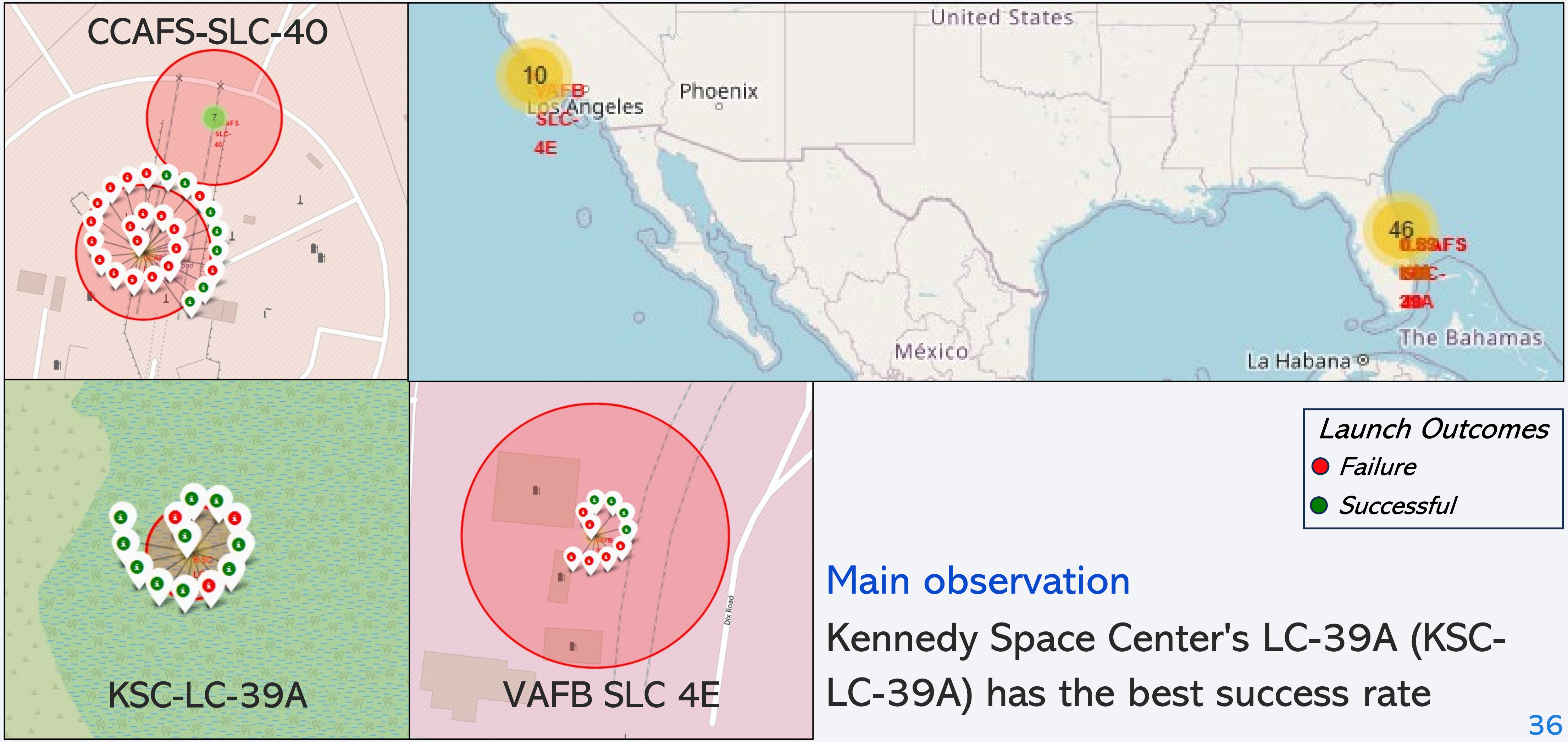
Launch Sites Proximities Analysis

Launch sites' location



All launch sites are located near the ocean

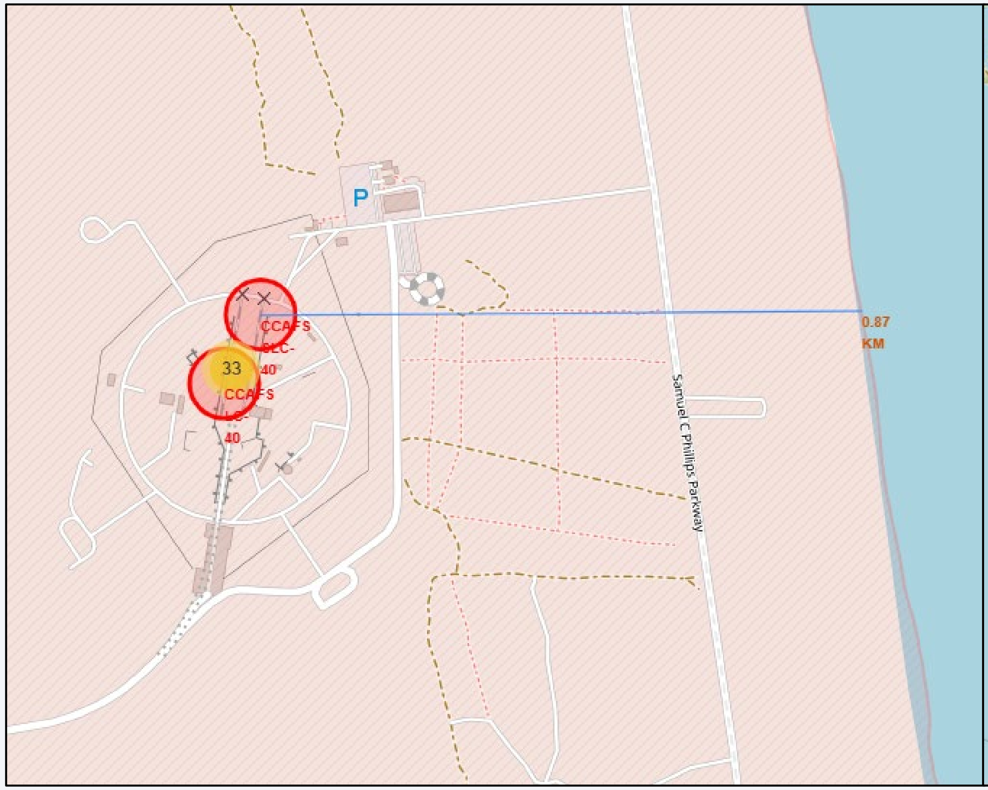
Launch outcomes marked by colors

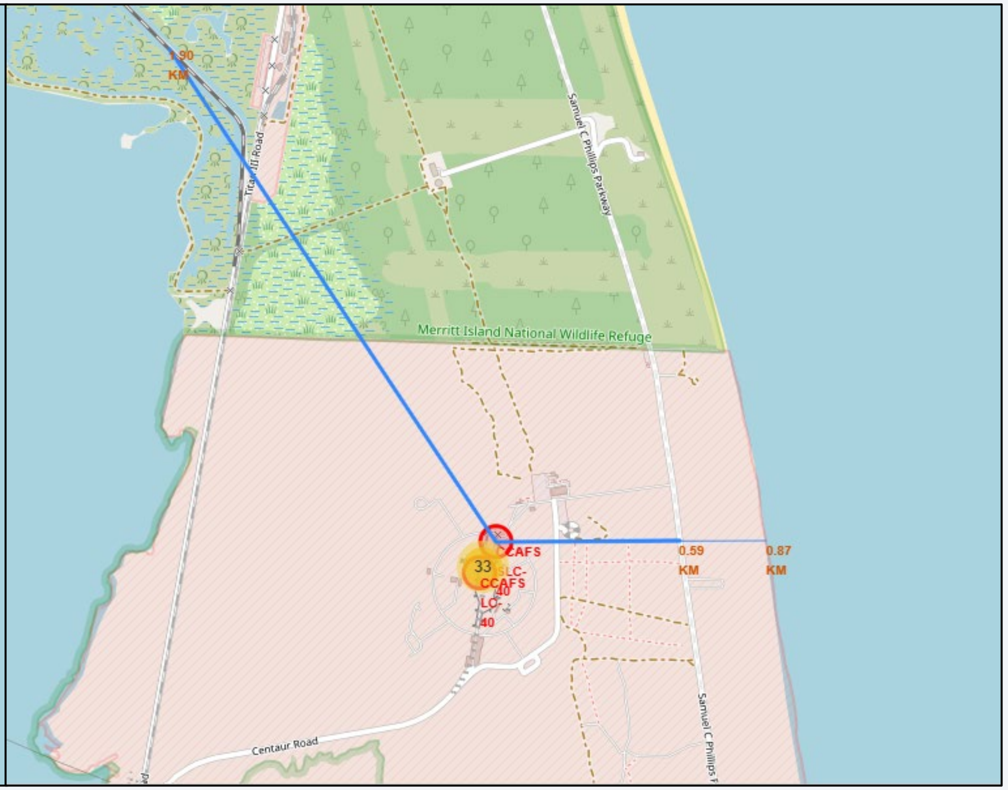


Main observation

Kennedy Space Center's LC-39A (KSC-LC-39A) has the best success rate

Launch sites' surroundings





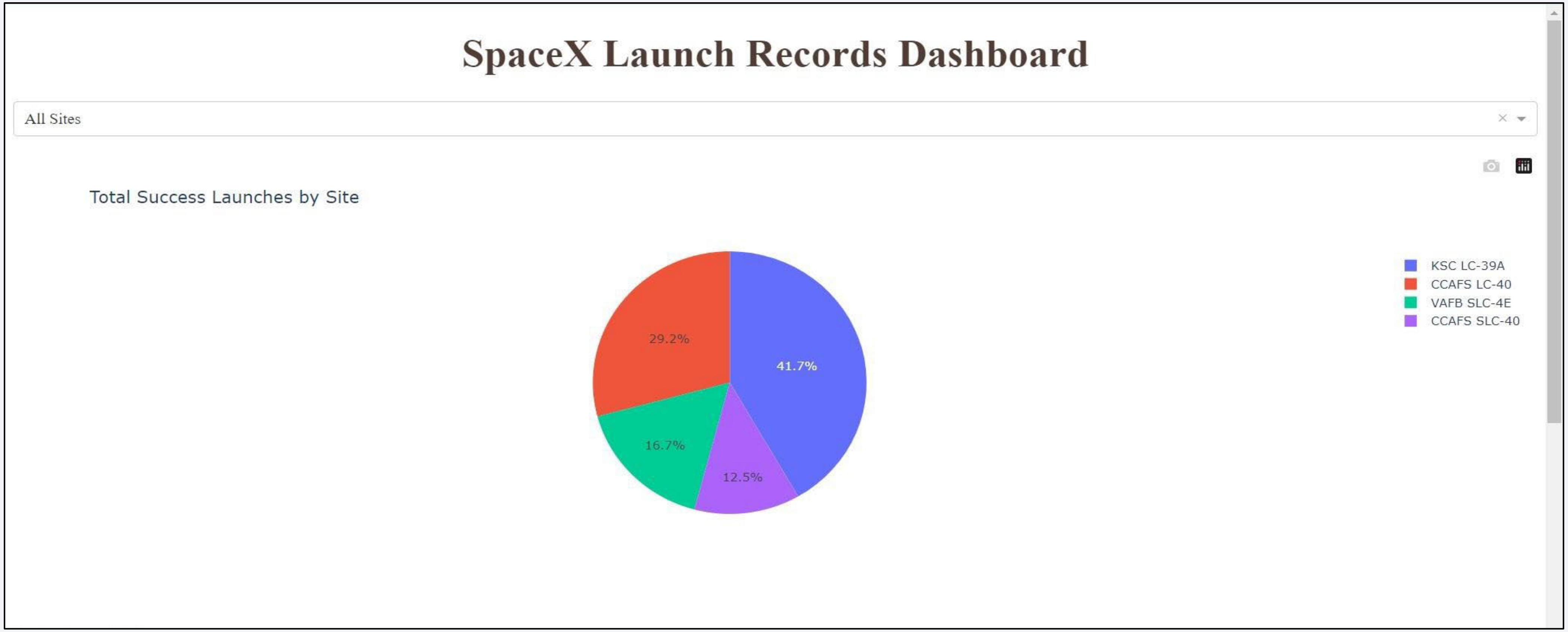
Main observations

Launch sites are strategically positioned within a safe distance from key logistical networks—railways, highways, coastal areas, and major urban centers—facilitating streamlined operations and transport.

Section 4

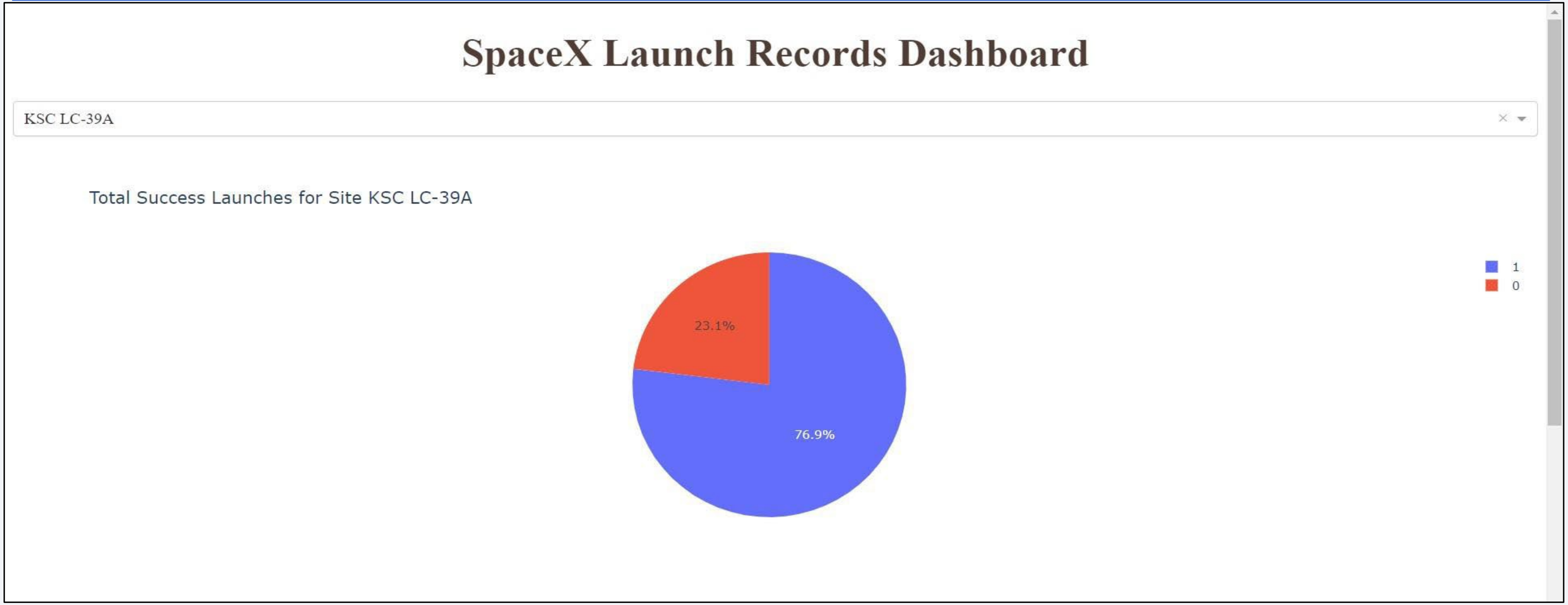
Build a Dashboard with Plotly Dash

Total Success Launches by all Sites



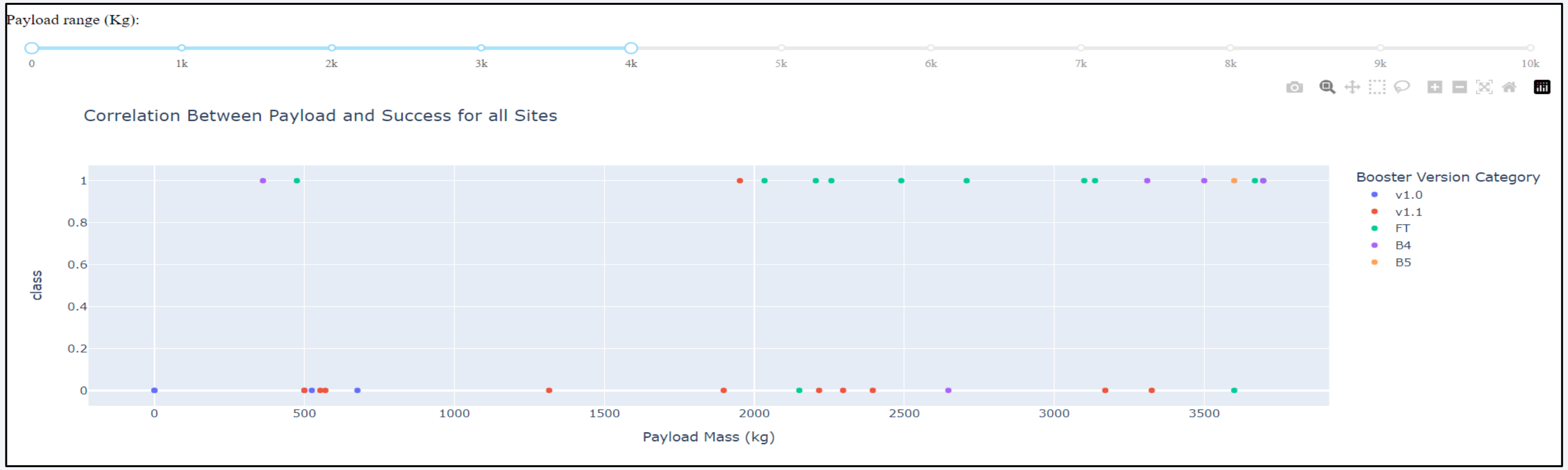
Launch Site KDC LC-39 A has the Greatest Launch Success Rate

Success/Failure Launch Rate by Selected Site

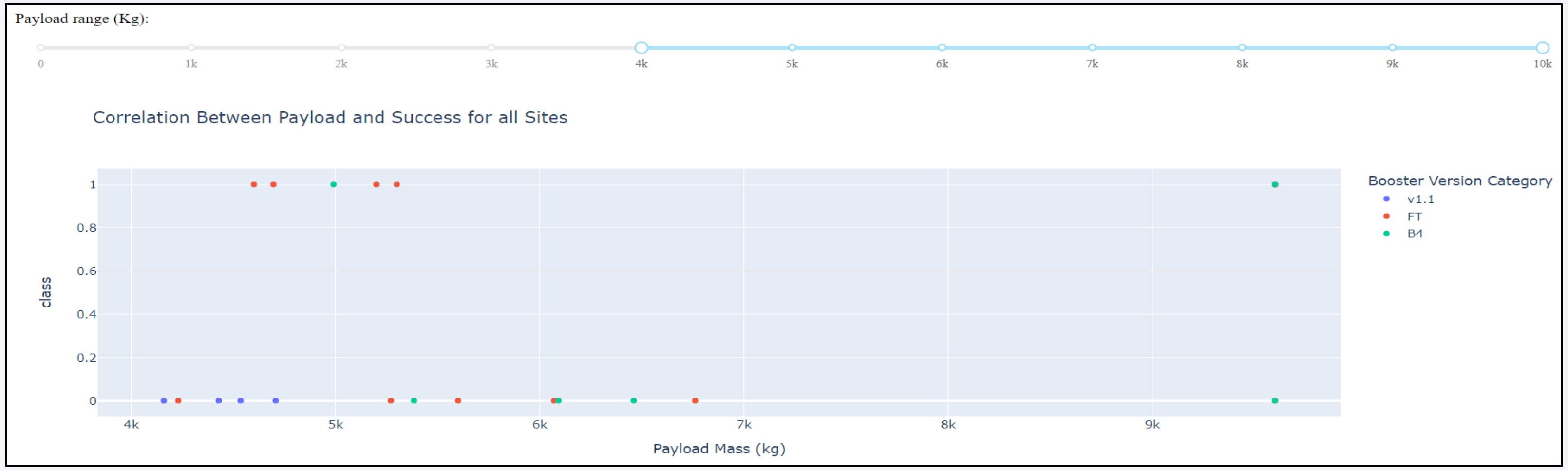


Data shows that Launch Site KDC LC-39 A achieved the highest launch success rate

Payload vs. Launch Outcome



Payload
0 – 4000 kg



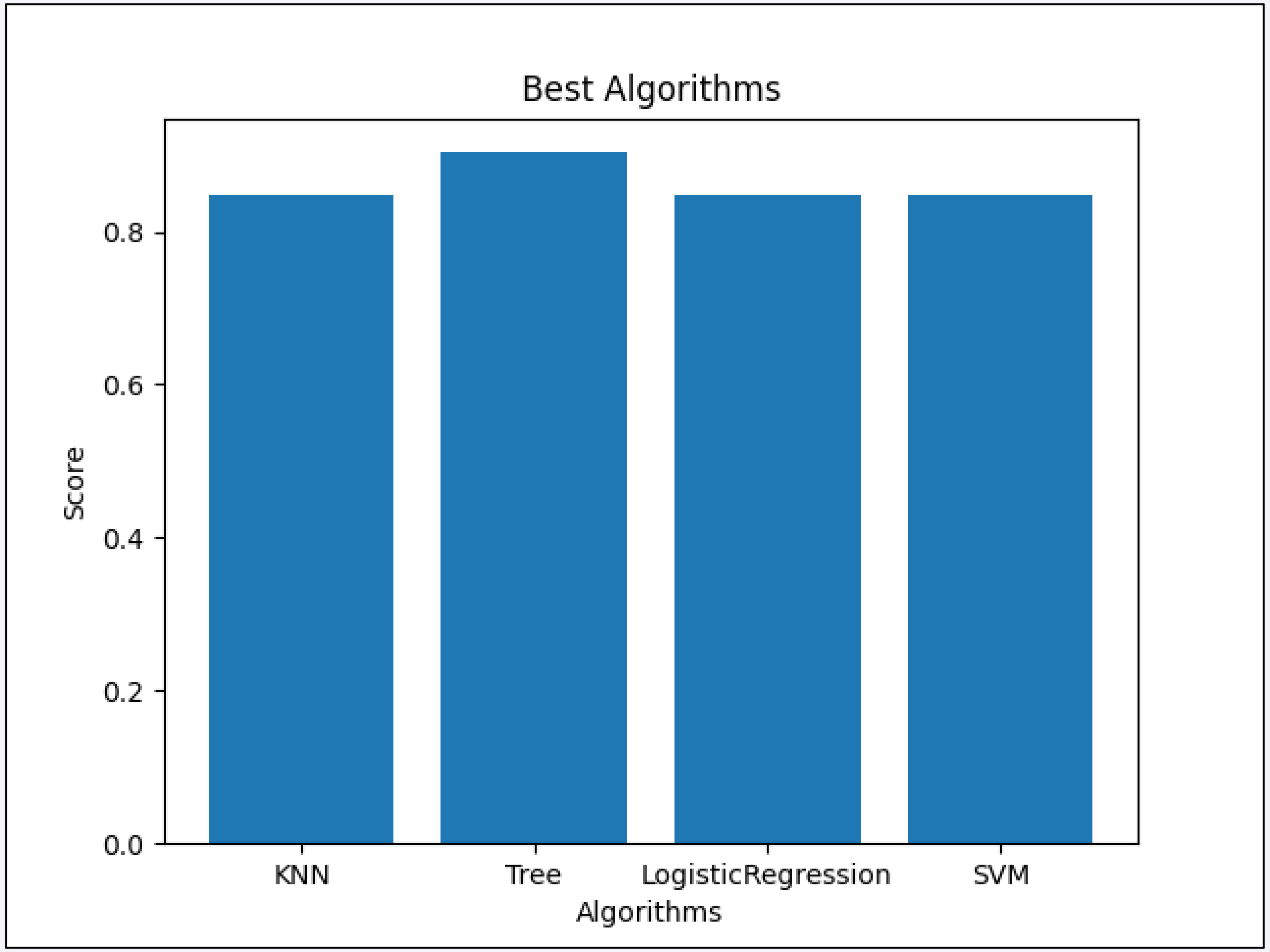
Payload
4000 – 10,000 kg

Lighter payloads have a higher success rate than heavier payloads

Section 5

Predictive Analysis (Classification)

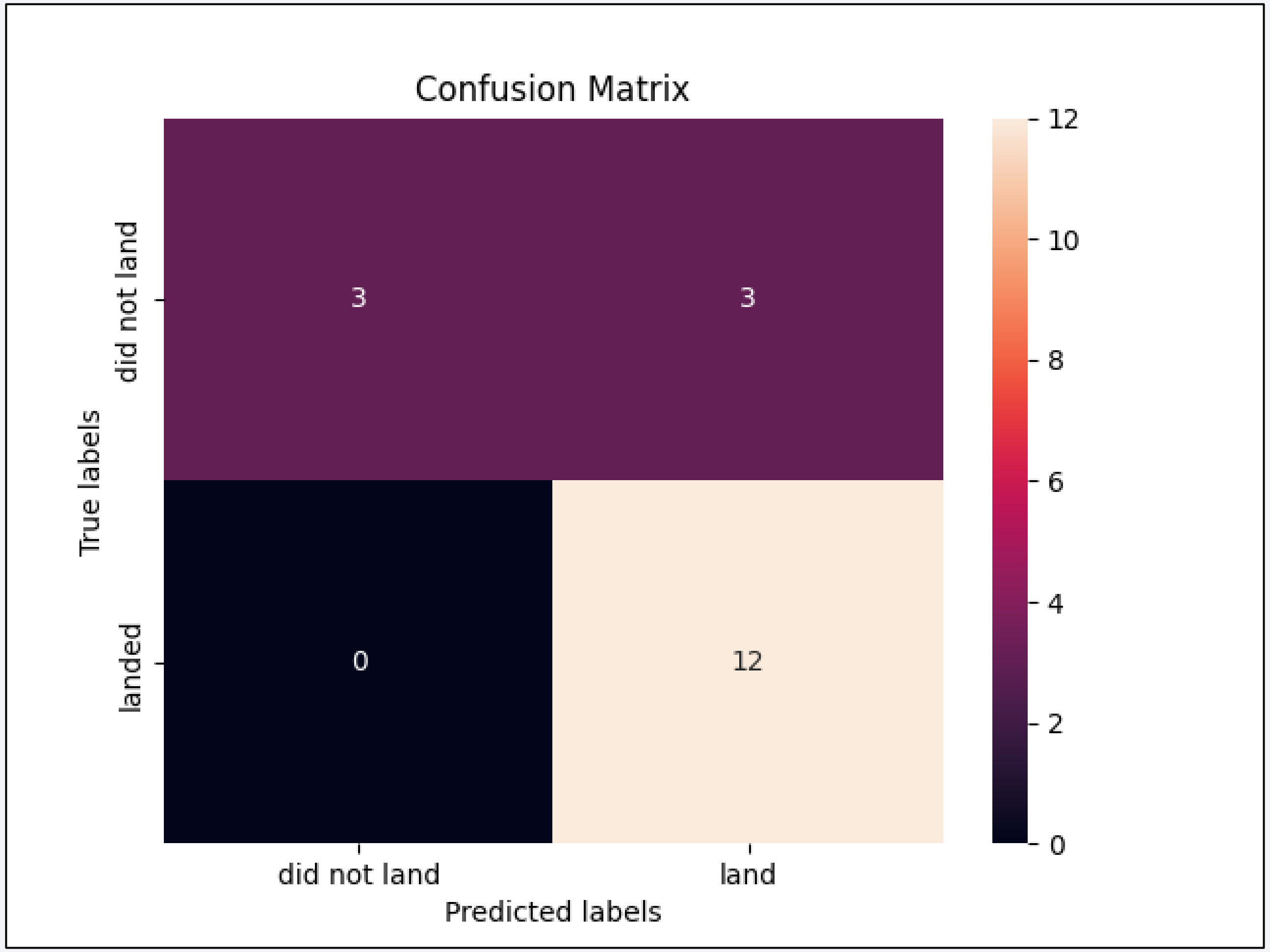
Classification Accuracy



Main finding

- The best algorithm is Decision Tree with a score of 90.3%

Confusion Matrix



Main finding

- Best Model – Decision Tree Classifier
- The confusion matrix for the Decision Tree Classifier reveals a substantial count of false positives, indicating instances where the classifier incorrectly predicts successful rocket landings despite their actual failures

Conclusions

- The success launch rate at a launch site is positively correlated with the the number of flights conducted there
- Orbits such as ES-L1, GEO, HEO, and SSO are associated with the most favorable outcomes in terms of launch success.
- Launches carrying heavier payloads tend to see improved success rates when aimed for VLEO, ISS, and PO orbits.
- From 2013 to 2020, the launch success rate has shown a general uptrend, with only minor variations observed during this interval.
- The Kennedy Space Center's LC-39A (KSC LC-39A) stands out for having the highest rate of successful launches.
- In predicting the likelihood of the SpaceX Falcon 9 rocket's first stage landing success, the Decision Tree Classifier emerges as the most effective machine learning model.

Thank you!