IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Edmund Lee
2023.10.26

Outline

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

Executive Summary

- Summary of methodologies
    a. Data collection with SpaceX API
    b. Data collection with web scrapping
    c. Data wrangling
    d. Exploratory data analysis with SQL
    e. Exploratory data analysis with data visualization
    f. Interactive visual analytics with Folium
    g. Machine learning prediction
- Summary of all results
    a. Exploratory data analysis
    b. Interactive analytics visualization
    c. Predictive analysis

Introduction

Source: *Popular Mechanics*

Project Background

Space X claimed that the Falcon 9 rocket launches cost 62 million dollars whereas other provides cost 165 million dollars each launch. This significant different is due to the reusable technology by Space X that land back Falcon 9 at first stage. Hence, this study aim to build a predictive model on the success of a rocket landing.

Research Questions:

1. What factors influence the landing of the rocket?
2. What are the features for the model?
3. What is the success rate of landing?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Collect data from Space X's API and web scrapping from Wikipedia.
- Perform data wrangling
  - Apply one-hot encoding method upon categorical features
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Apply several models – SVM, decision trees, KNN, logistics regression
  - Compare the accuracy results

Data Collection

1. Data collection with Space X's API
    a. Set up the GET request to the API URL
    b. Decode the response and turn it into Pandas data frame using .json_normalize()
    c. Perform data wrangling, clean the missing values

2. Data collection with web scrapping from Wikipedia
    a. Set up the GET request to the Wikipedia page URL
    b. Extract all column from the HTML table header
    c. Create a data frame by parsing the launch HTML tables

Both methods serve as a way to obtain the data of Falcon 9 launch records.

Data Collection - Scraping

Data Wrangling

EDA with Data Visualization

EDA with SQL

Build an Interactive Map with Folium

Build a Dashboard with Plotly Dash

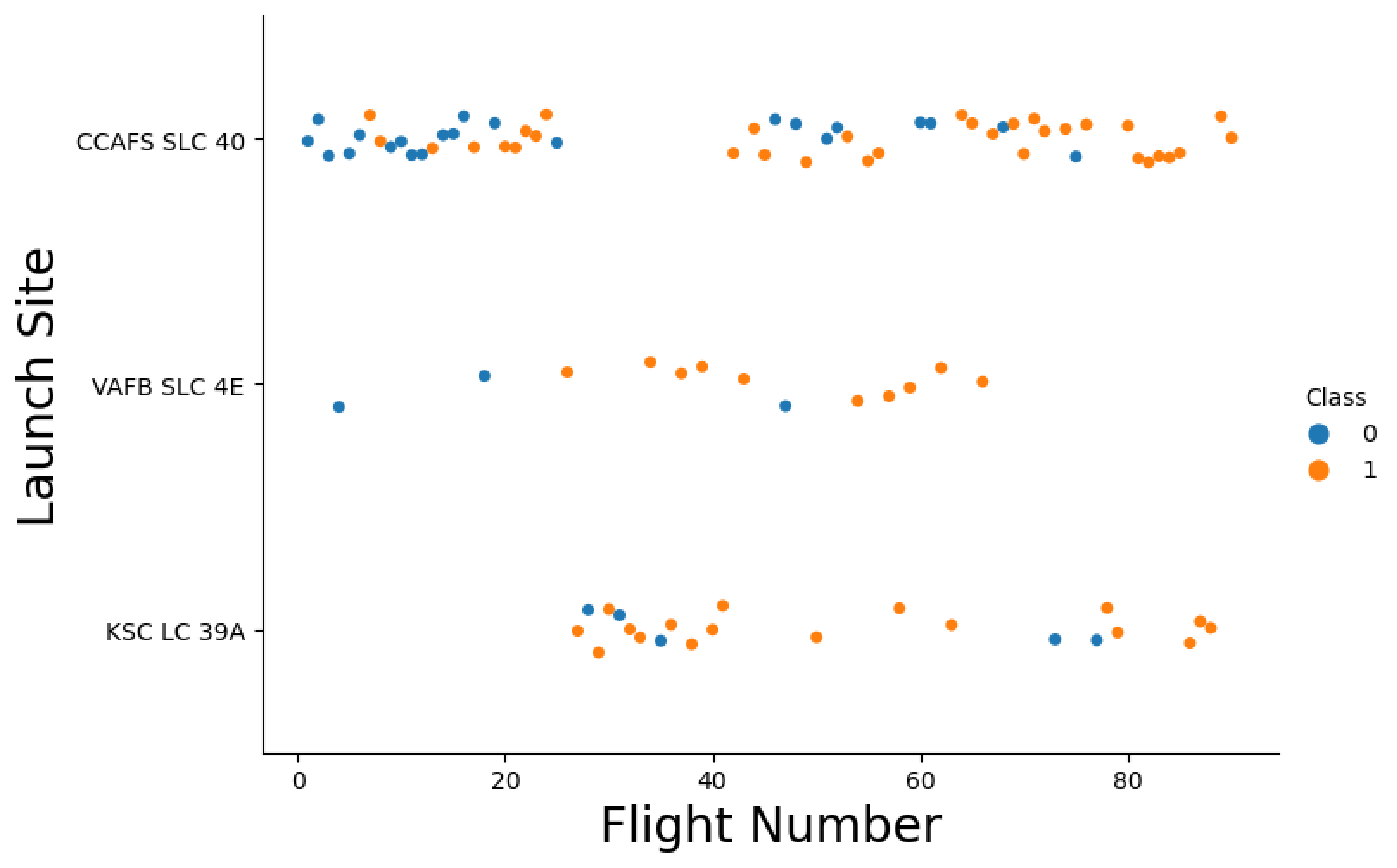
Predictive Analysis (Classification)

Section 2

Insights drawn from EDA

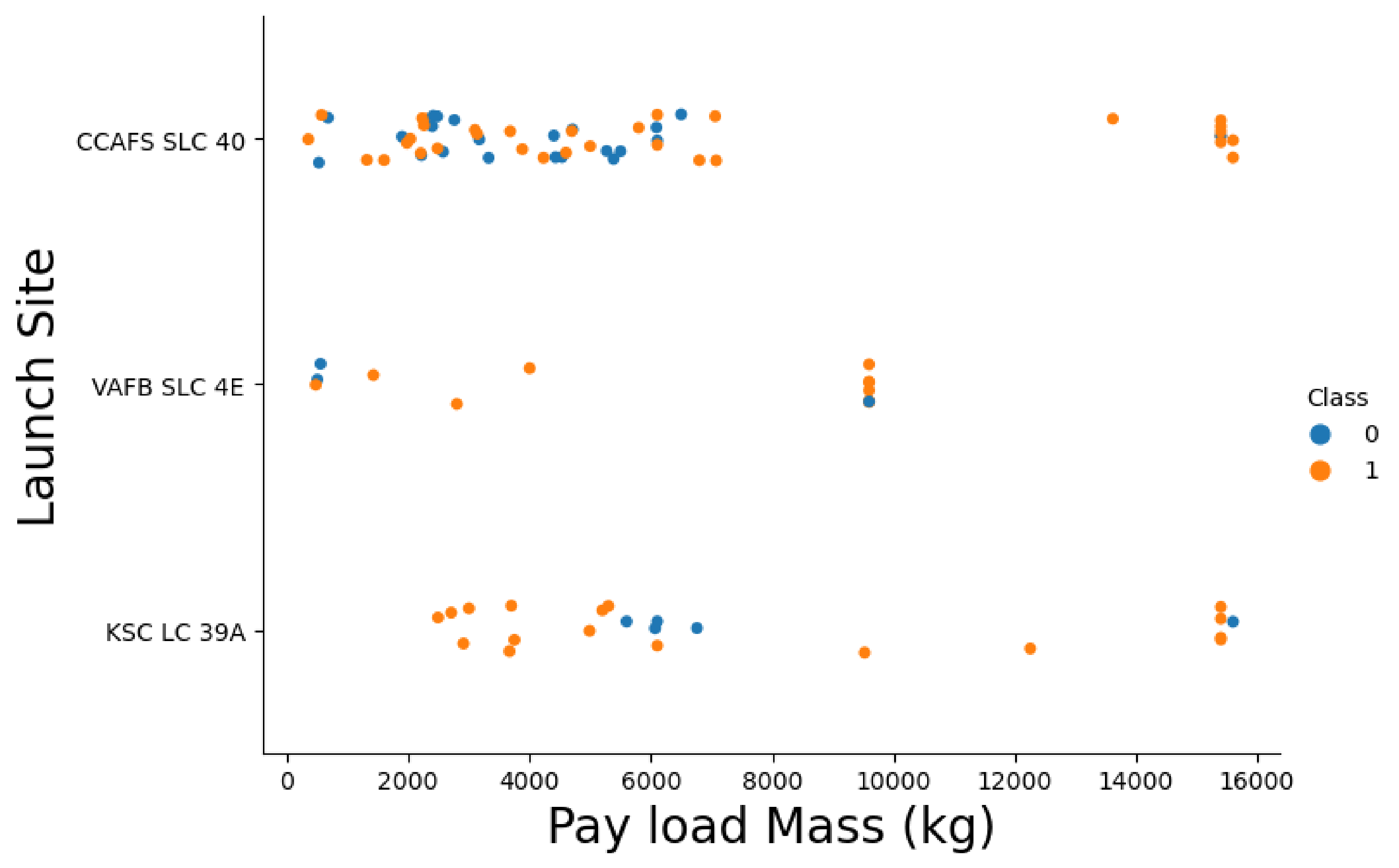
Flight Number vs. Launch Site

- It shows that the larger the flight number the higher the success rate at a launch site



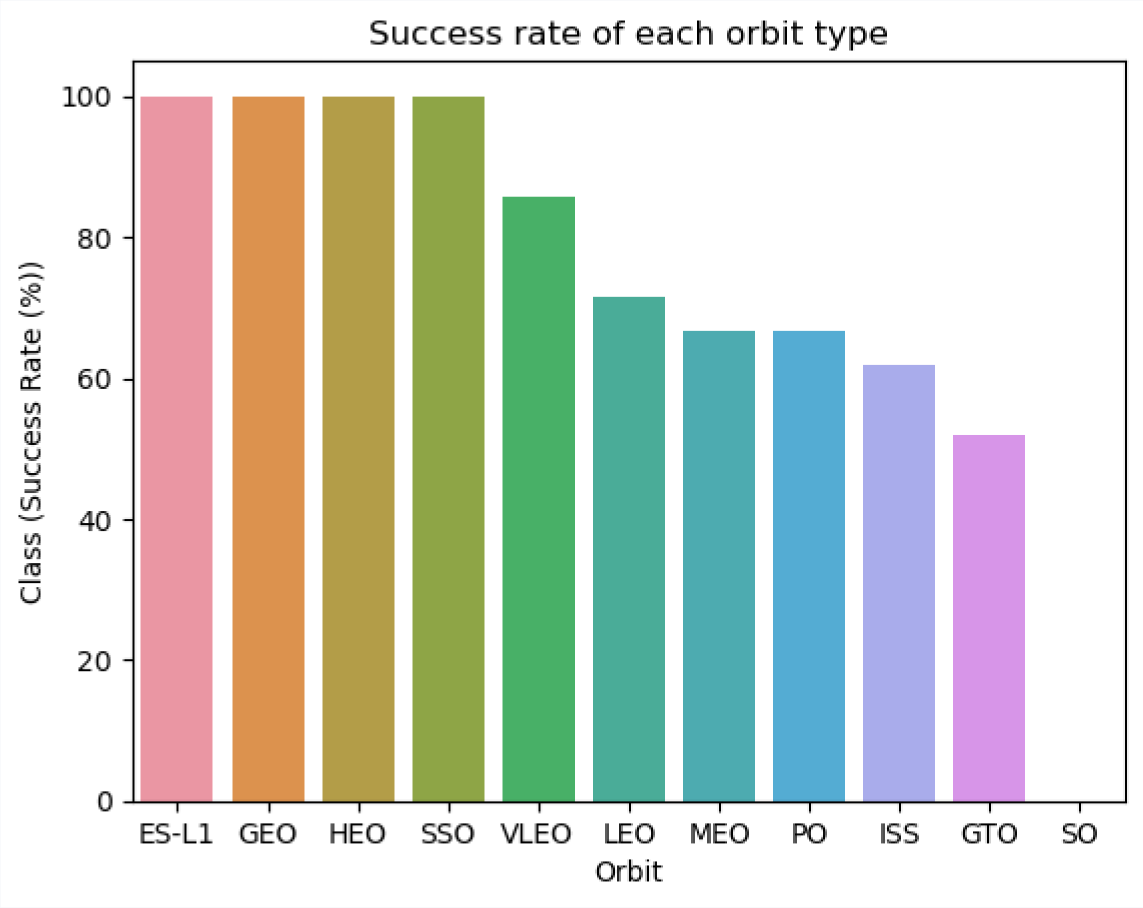
Payload vs. Launch Site

- It shows that the greater the payload mass, the higher the success rate.



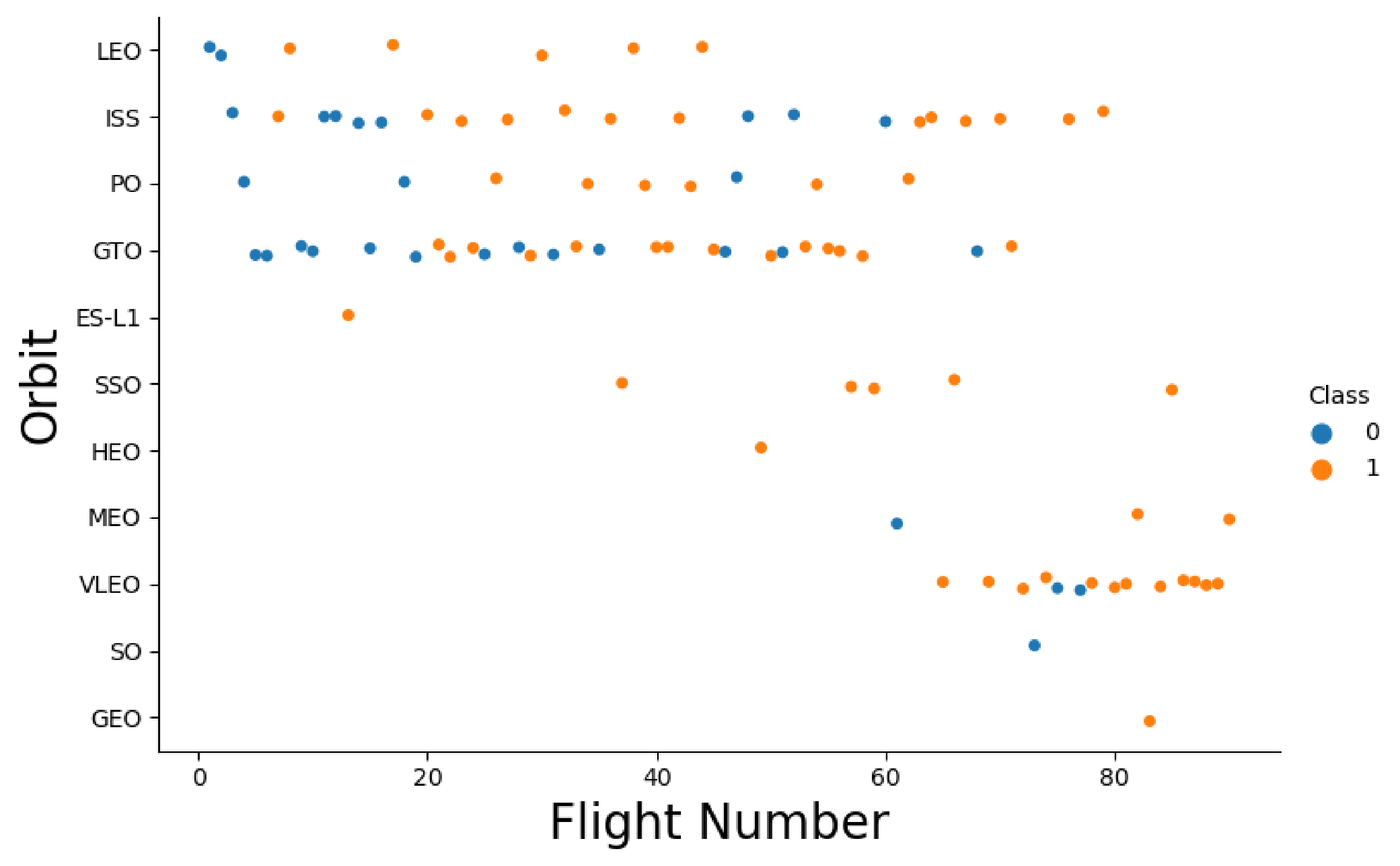
Success Rate vs. Orbit Type

- It shows that ES-L1, GEO, HEO, SSO, VLEO have the highest success rate



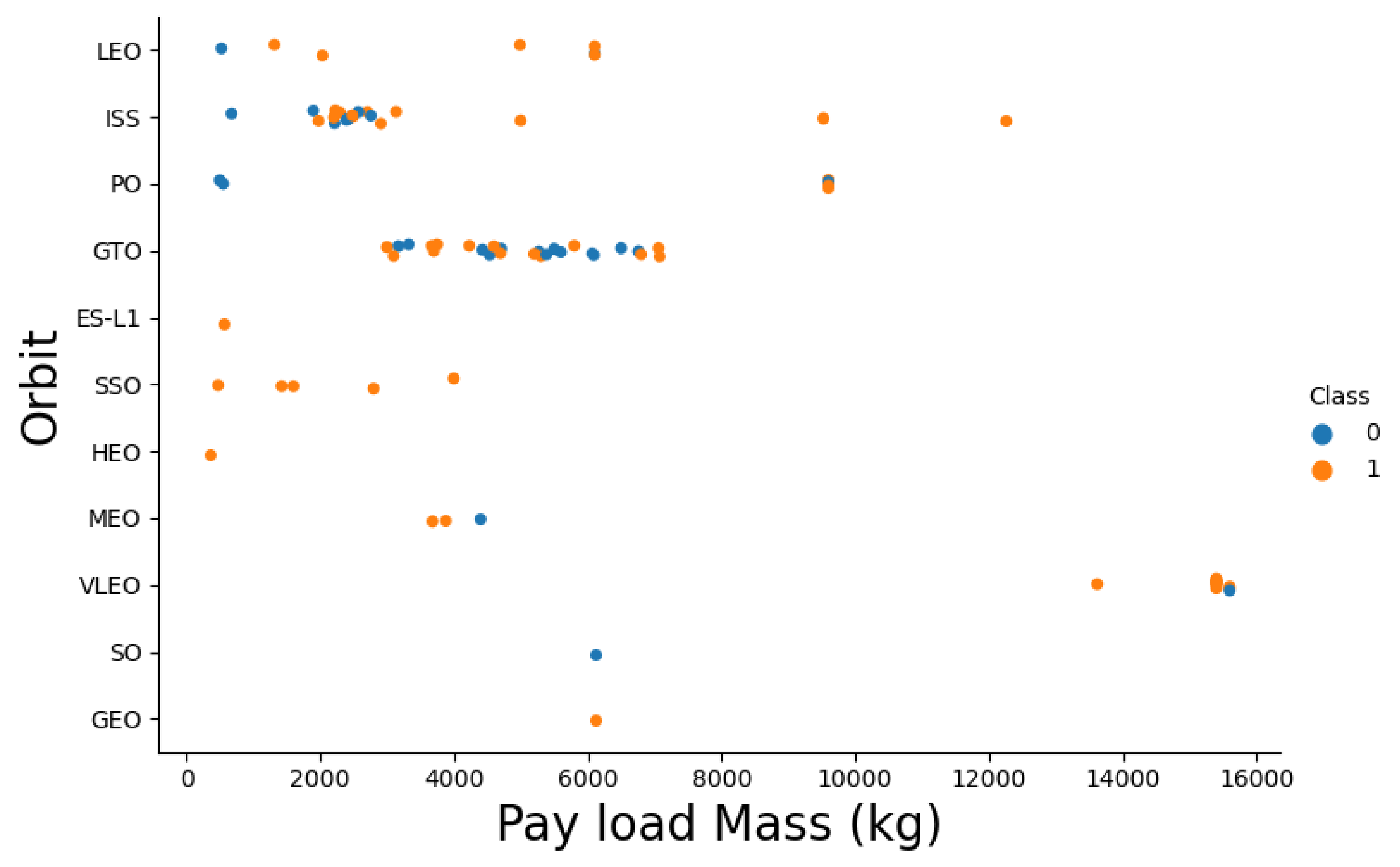
Flight Number vs. Orbit Type

- It shows the higher the flight number, the higher the success rate



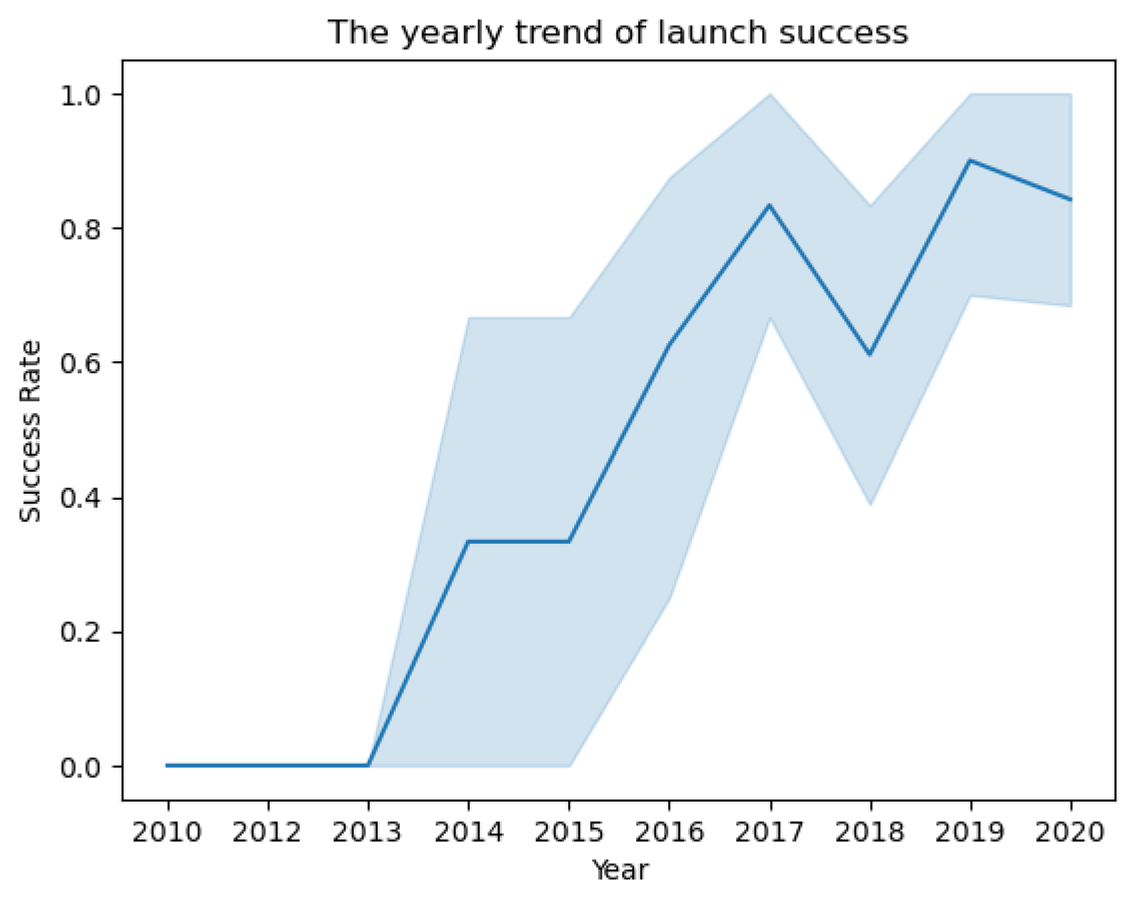
Payload vs. Orbit Type

- It shows that higher payload return high success for LEO, ISS.



Launch Success Yearly Trend

- It shows that the launch success is increasing from 2010 to 2020 generally, albeit with significant decline in 2018.



All Launch Site Names

```
In [9]:  %%sql
         SELECT DISTINCT LAUNCH_SITE as "Launch_Sites"
         FROM SPACEXTBL

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

Out[9]:

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

Launch Site Names Begin with 'CCA'

In [10]:

```
%%sql
SELECT *
FROM 'SPACEXTBL'
WHERE Launch_Site LIKE 'CCA%'
LIMIT 5;
```

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

Out[10]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orb
2010-04-06	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LE
2010-08-12	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0	LE (IS
2012-05-22	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LE (IS
2012-08-10	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LE (IS
2013-01-03	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LE (IS

Total Payload Mass

In [11]:

```
%%sql
SELECT SUM(PAYLOAD_MASS__KG_) as "Total Payload Mass(Kgs)", Customer
FROM 'SPACEXTBL'
WHERE Customer = 'NASA (CRS)'
```

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

Out[11]:

Total Payload Mass(Kgs)	Customer
45596	NASA (CRS)

Average Payload Mass by F9 v1.1

In [12]:

```
%%sql
SELECT AVG(PAYLOAD_MASS__KG_) as "Payload Mass Kgs", Customer, Booster_Versic
FROM 'SPACEXTBL'
WHERE Booster_Version LIKE 'F9 v1.1%';
```

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

Out[12]:

Payload Mass Kgs	Customer	Booster_Version
2534.6666666666665	MDA	F9 v1.1 B1003

First Successful Ground Landing Date

In [13]:

```
%%sql
SELECT MIN(DATE)
FROM SPACEXTBL
WHERE "Landing_Outcome" = "Success (ground pad)"
```

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

Out[13]:

MIN(DATE)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

In [14]:
```
%%sql
SELECT DISTINCT Booster_Version, Payload
FROM SPACEXTBL
WHERE Landing_Outcome = "Success (drone ship)"
  AND PAYLOAD_MASS__KG_ > 4000
  AND PAYLOAD_MASS__KG_ < 6000;
```

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

Out[14]:

Booster_Version	Payload
F9 FT B1022	JCSAT-14
F9 FT B1026	JCSAT-16
F9 FT B1021.2	SES-10
F9 FT B1031.2	SES-11 / EchoStar 105

Total Number of Successful and Failure Mission Outcomes

In [15]:
```
%%sql
SELECT Mission_Outcome, COUNT(Mission_Outcome) as Total
FROM SPACEXTBL
GROUP BY Mission_Outcome
```

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

Out[15]:

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

Boosters Carried Maximum Payload

In [16]:

```
%%sql
SELECT Booster_Version, Payload, PAYLOAD_MASS__KG_
FROM SPACEXTBL
WHERE "PAYLOAD_MASS__KG_" = (SELECT MAX("PAYLOAD_MASS__KG_") FROM SPACEXTBL);
```

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

Out[16]:

Booster_Version	Payload	PAYLOAD_MASS__KG_
F9 B5 B1048.4	Starlink 1 v1.0, SpaceX CRS-19	15600
F9 B5 B1049.4	Starlink 2 v1.0, Crew Dragon in-flight abort test	15600
F9 B5 B1051.3	Starlink 3 v1.0, Starlink 4 v1.0	15600
F9 B5 B1056.4	Starlink 4 v1.0, SpaceX CRS-20	15600
F9 B5 B1048.5	Starlink 5 v1.0, Starlink 6 v1.0	15600
F9 B5 B1051.4	Starlink 6 v1.0, Crew Dragon Demo-2	15600
F9 B5 B1049.5	Starlink 7 v1.0, Starlink 8 v1.0	15600
F9 B5 B1060.2	Starlink 11 v1.0, Starlink 12 v1.0	15600
F9 B5 B1058.3	Starlink 12 v1.0, Starlink 13 v1.0	15600
F9 B5 B1051.6	Starlink 13 v1.0, Starlink 14 v1.0	15600
F9 B5 B1060.3	Starlink 14 v1.0, GPS III-04	15600
F9 B5 B1049.7	Starlink 15 v1.0, SpaceX CRS-21	15600

2015 Launch Records

In [17]:
```
%%sql
SELECT substr(Date,0,5) AS 'Year', substr(Date, 6, 2) AS 'Month', Booster_Ver
FROM SPACEXTBL
WHERE Year='2015' AND Landing_Outcome = 'Failure (drone ship)'
```

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

Out[17]:

Year	Month	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Mission_(
2015	10	F9 v1.1 B1012	CCAFS LC-40	SpaceX CRS-5	2395	
2015	04	F9 v1.1 B1015	CCAFS LC-40	SpaceX CRS-6	1898	

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

In [18]:

```
%%sql
SELECT *
FROM SPACEXTBL
WHERE Landing_Outcome LIKE 'Success%' AND (Date BETWEEN '2010-06-04' AND '201
```

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

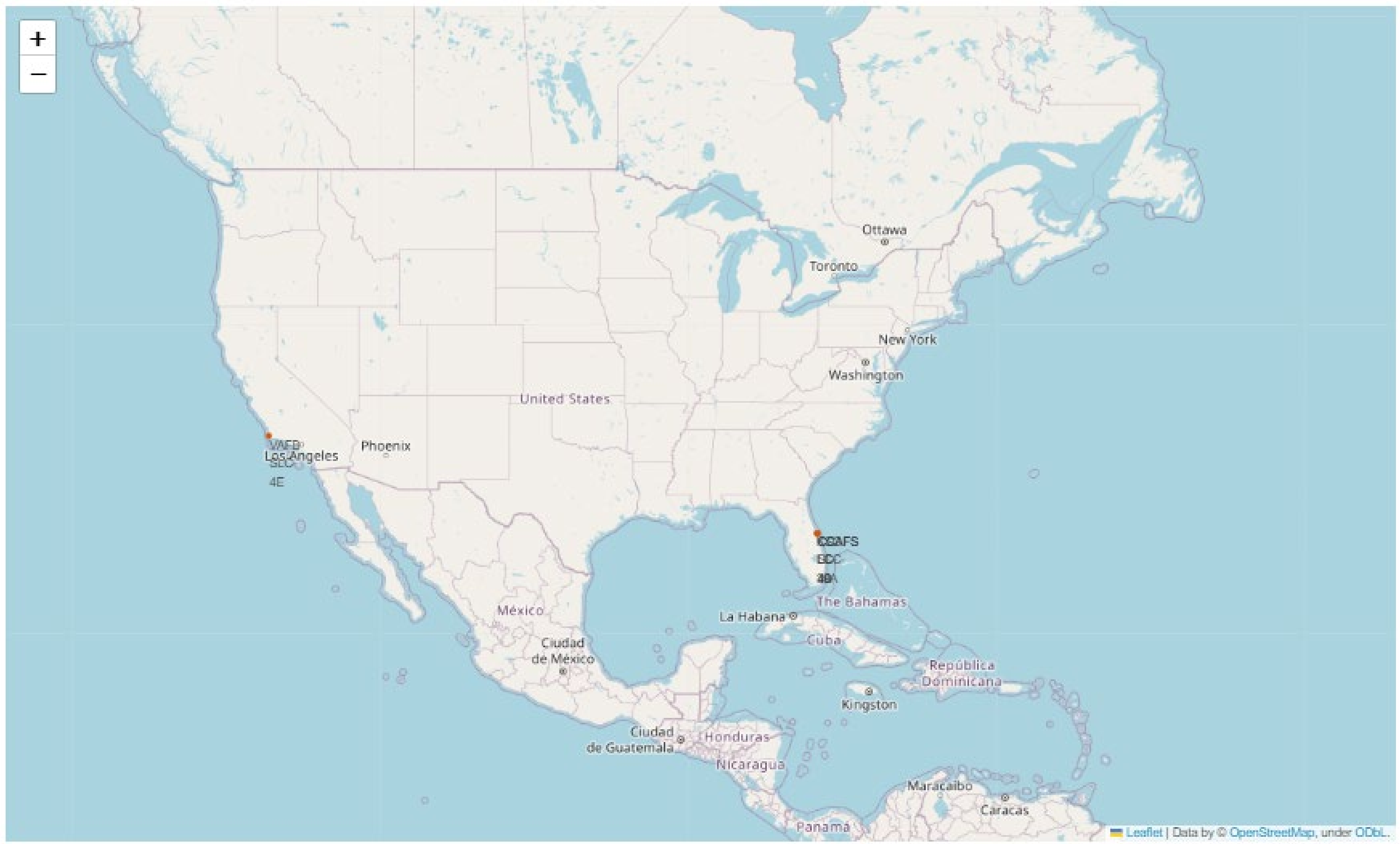
Out[18]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbi
2017-03-06	21:07:00	F9 FT B1035.1	KSC LC-39A	SpaceX CRS-11	2708	LEC (ISS
2017-02-19	14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490	LEC (ISS
2017-01-14	17:54:00	F9 FT B1029.1	VAFB SLC-4E	Iridium NEXT 1	9600	Pola LEC
2017-01-05	11:15:00	F9 FT B1032.1	KSC LC-39A	NROL-76	5300	LEC
2016-08-14	05:26:00	F9 FT B1026	CCAFS LC-40	JCSAT-16	4600	GTC
2016-08-04	20:43:00	F9 FT B1021.1	CCAFS LC-40	SpaceX CRS-8	3136	LEC (ISS
2016-07-18	04:45:00	F9 FT B1025.1	CCAFS LC-40	SpaceX CRS-9	2257	LEC (ISS
2016-06-05	05:21:00	F9 FT B1022	CCAFS LC-40	JCSAT-14	4696	GTC
2016-05-27	21:39:00	F9 FT B1023.1	CCAFS LC-40	Thaicom 8	3100	GTC
2015-12-22	01:29:00	F9 FT B1019	CCAFS LC-40	OG2 Mission 2 11 Orbcomm-OG2 satellites	2034	LEC

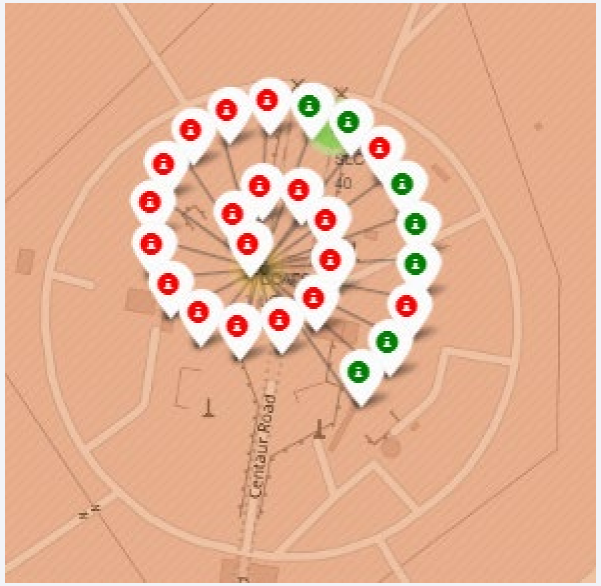
Section 3

Launch Sites Proximities Analysis

All launch sites in the United States



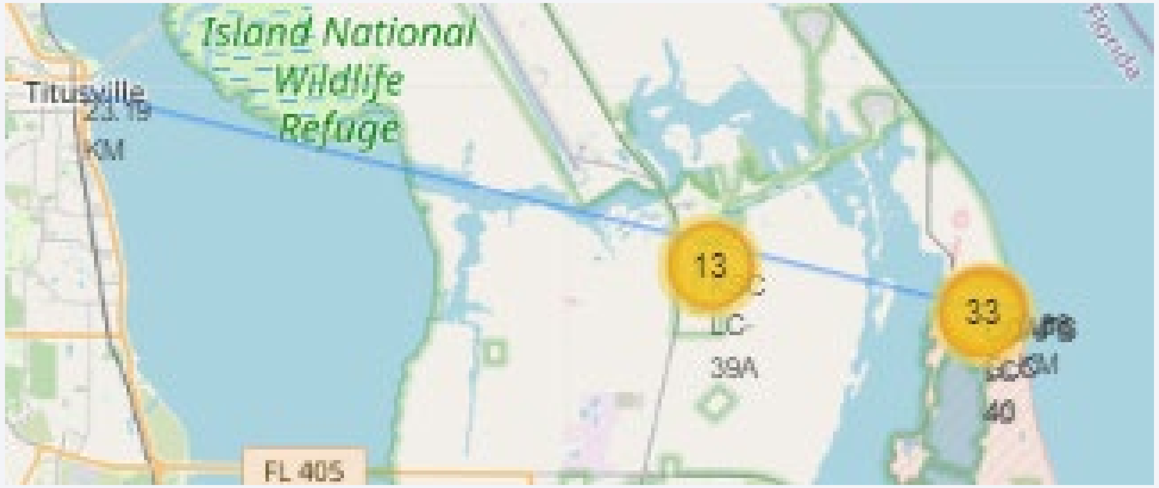
Markers showing launch sites with color labels



Green marker shows successful launches
Red marker shows failure launches

Launch Site Distance to Landmarks

0.86KM distance to coast line



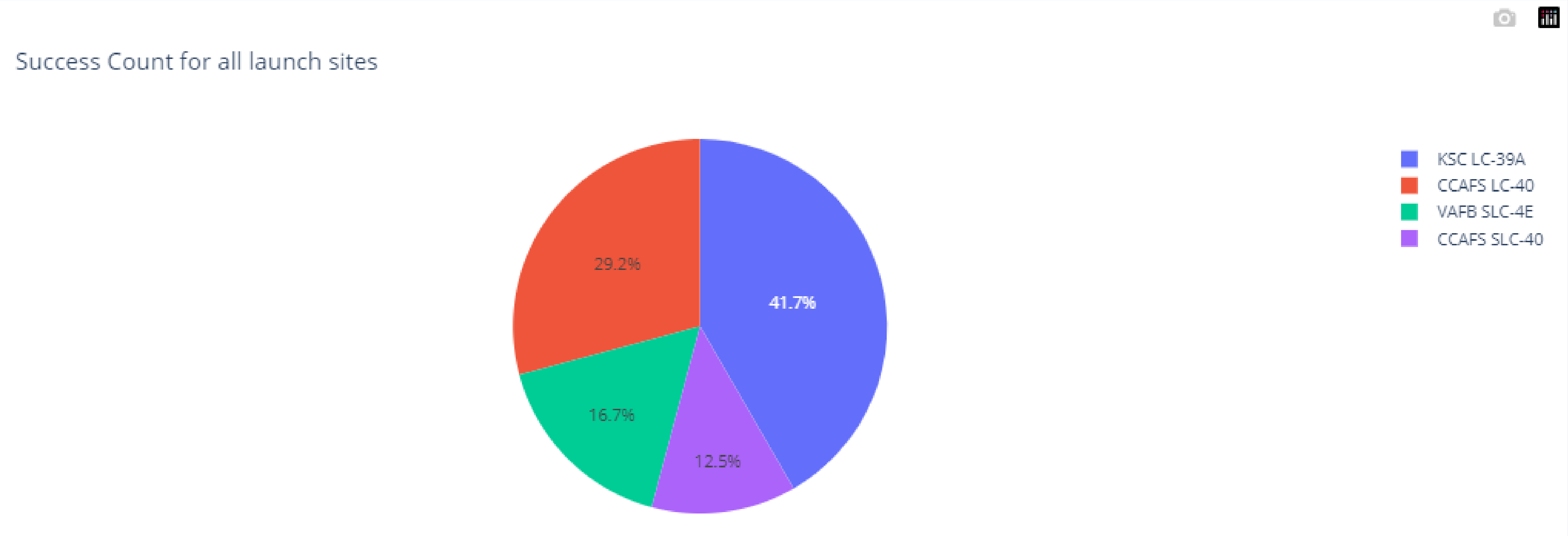
23.19KM distance to city

Section 4

Build a Dashboard with Plotly Dash

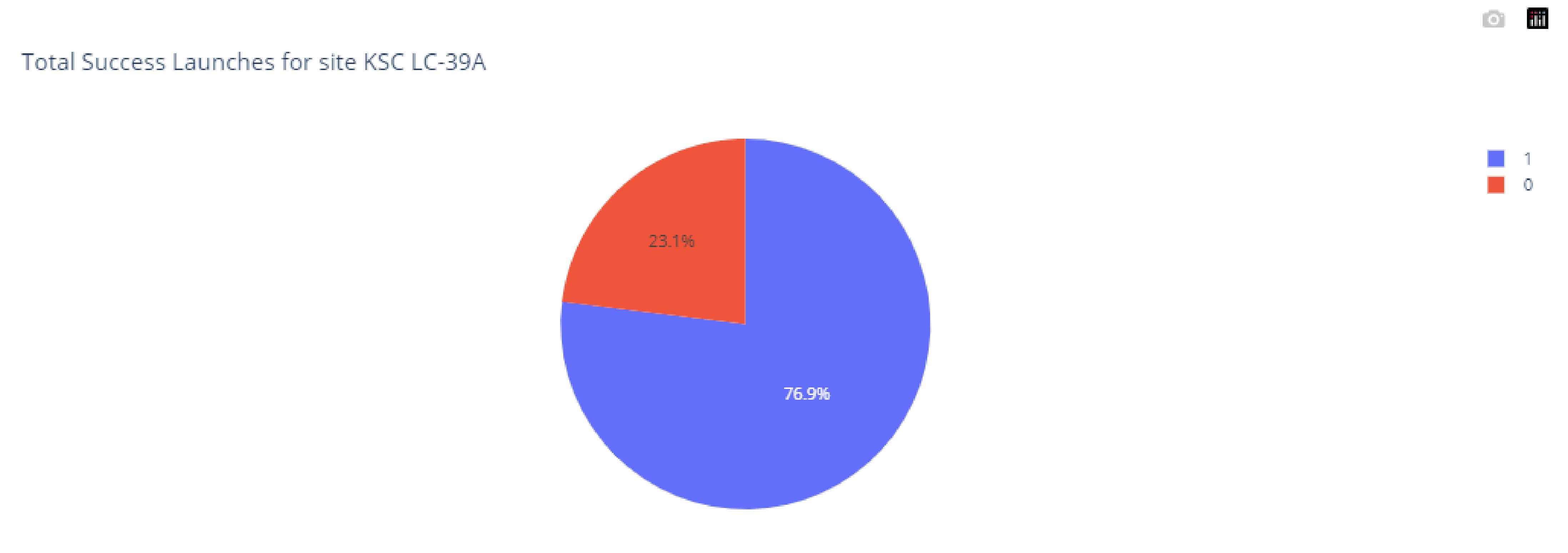
Success Count for All Launch Sites

The pie chart shows that KSC LC-39A had the highest success launch.



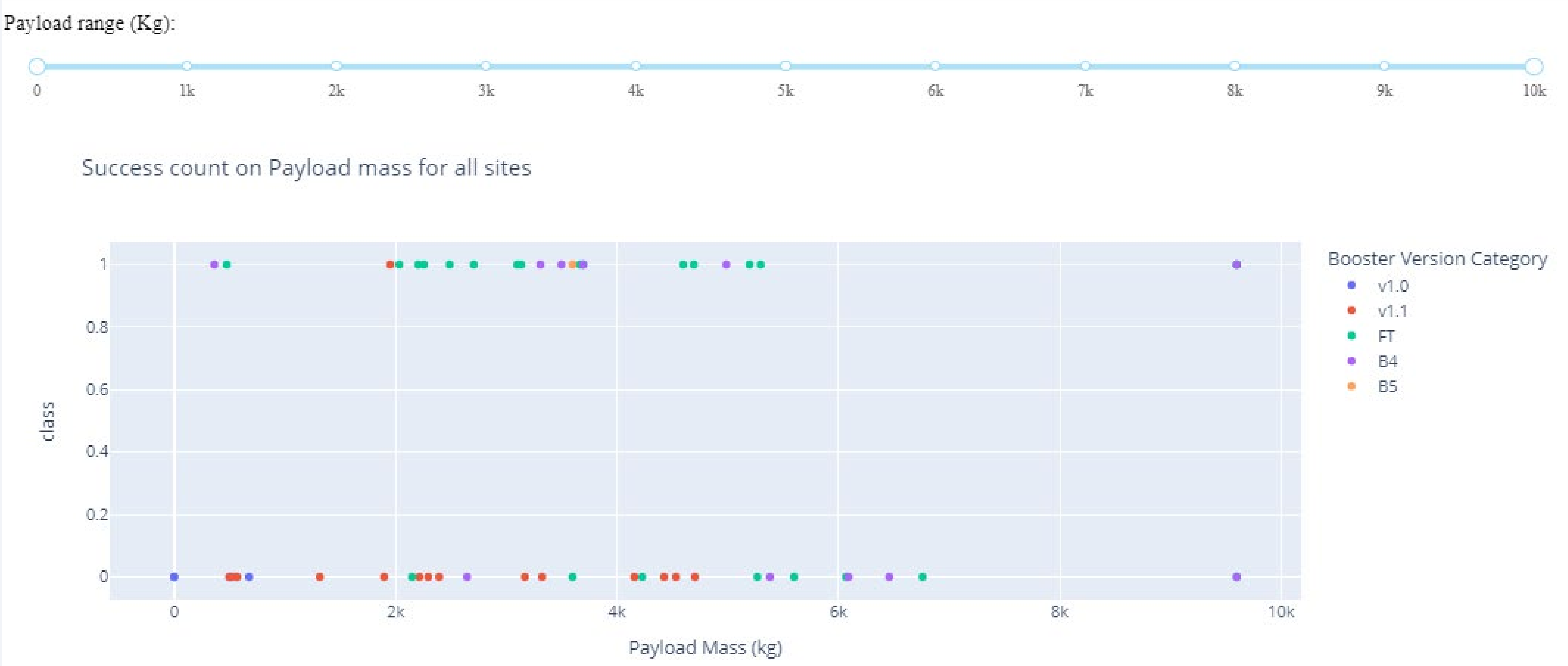
The Launch Site with the Highest Launch Success Ratio

The distribution for KSC LC-39A.



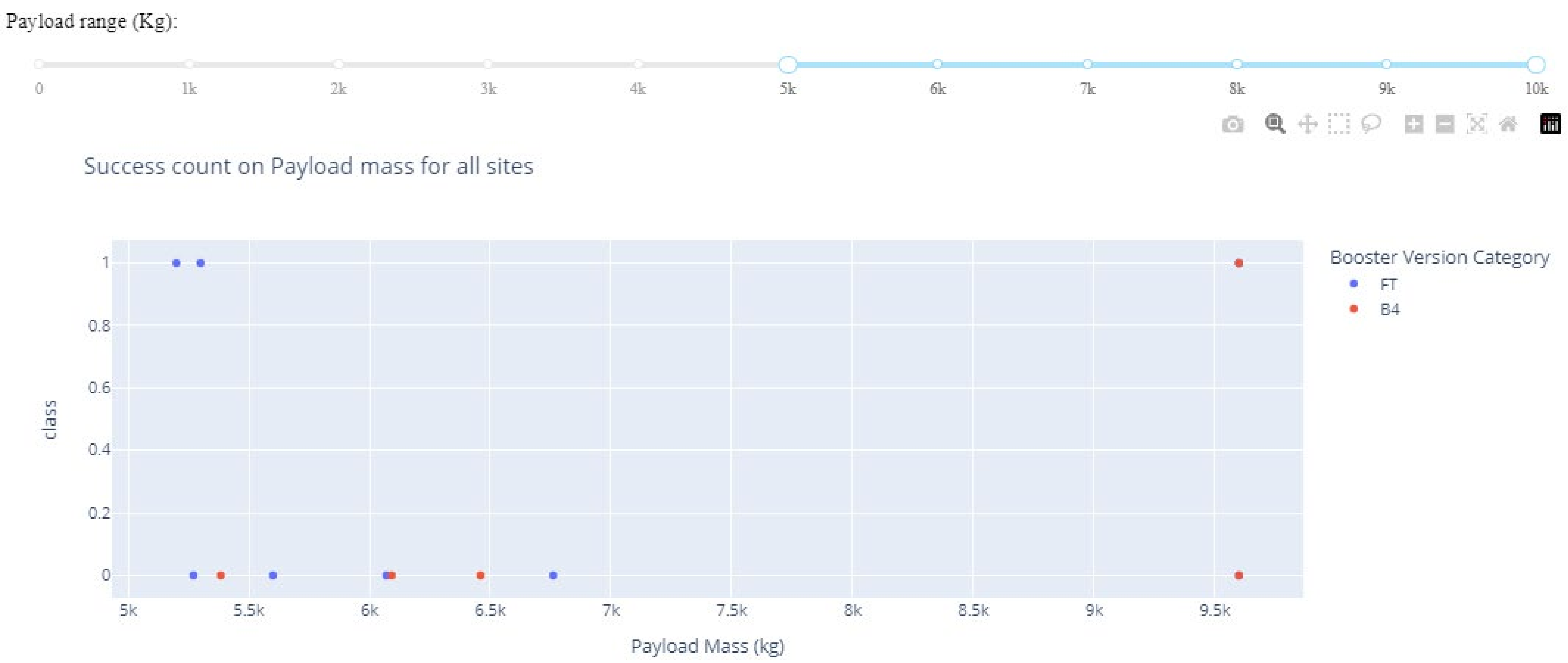
Scatter Plot of Playload mass for All Sites

Payload 0 KG to 10k KG



Scatter Plot of Playload mass for All Sites

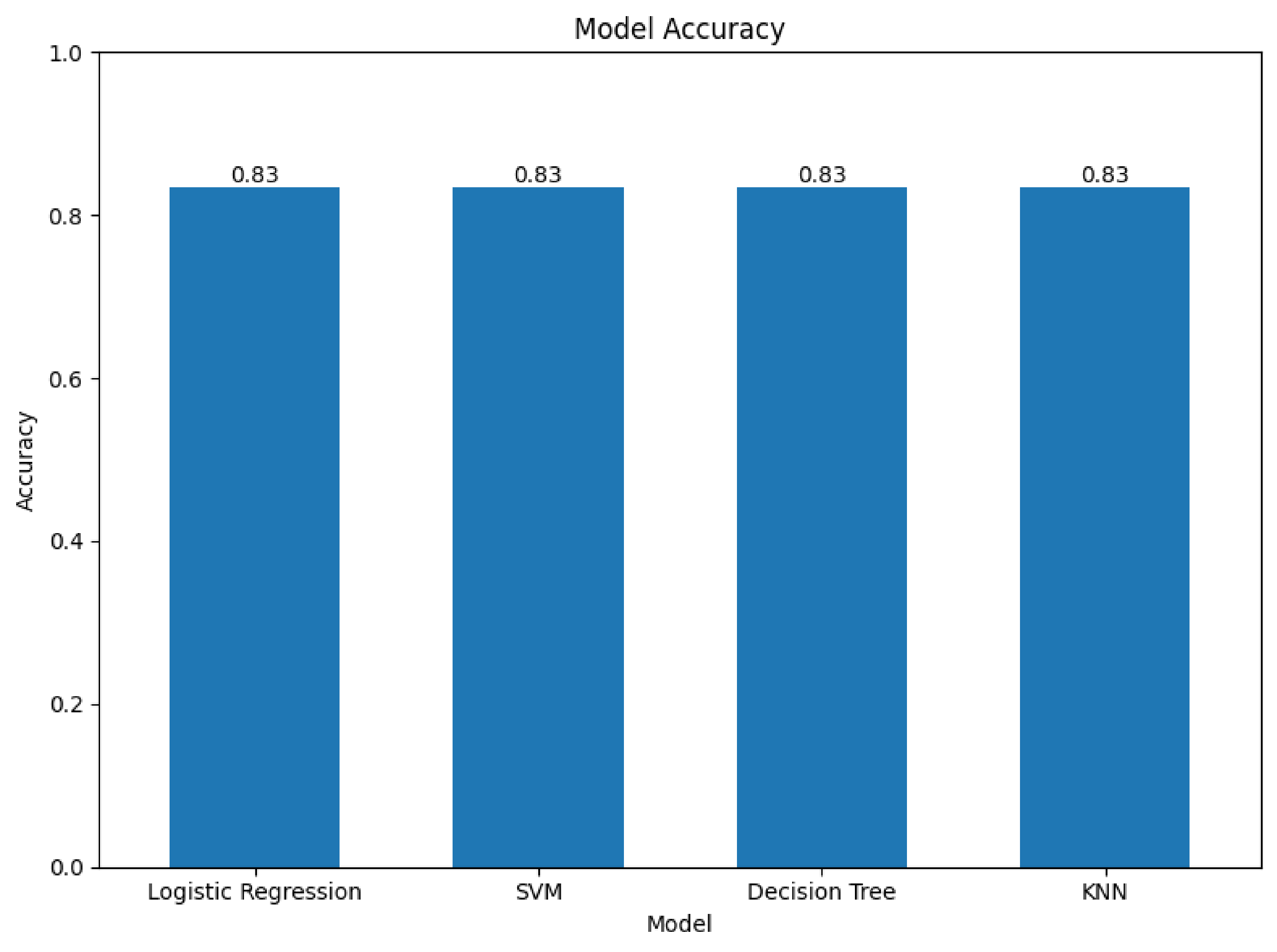
Payload 5 KG to 10k KG



Section 5

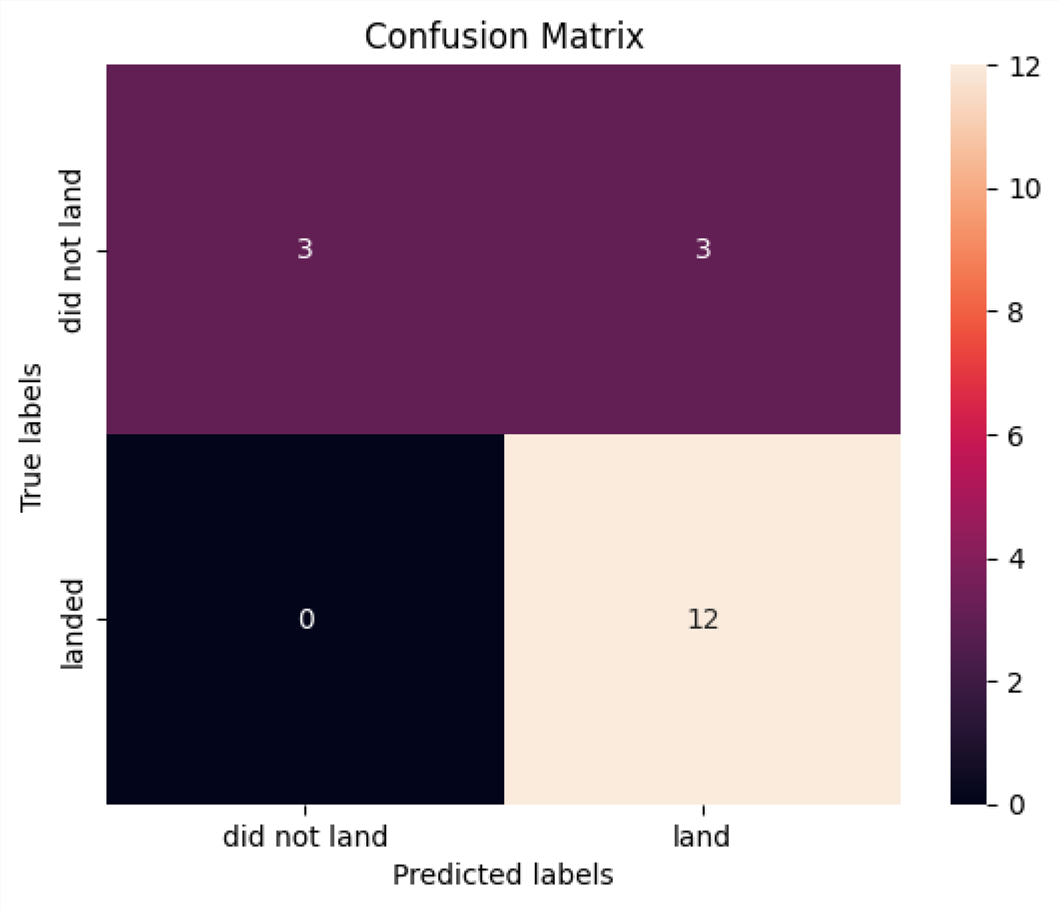
Predictive Analysis (Classification)

Classification Accuracy



Apparently, all four model had the same accuracy on the test dataset. Hence, they are equally good.

Confusion Matrix



Confusion matrix for KNN model.
The model has problem for False Positive meaning incorrectly predicting launch as landing successful but turned out it was a failure.

Conclusions

- In conclusion, this project shows a complete machine learning predictive workflow, starts from data collection, data wrangling, exploratory data analysis, data visualization, model prediction and reporting.
- The launch rate had been increasing from 2010 to 2020.
- The prediction on 4 models had been the same with 83% accuracy, but the model has problem in false positive.

Thank you!