IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Tomas Kondela
18.06.2024

Outline

Executive Summary

- Summary of methodologies
  - Data Collection through API
  - Data Collection with Web Scraping
  - Data Wrangling
  - Exploratory Data Analysis with SQL
  - Exploratory Data Analysis with Data Visualization
  - Interactive Visual Analytics with Folium
  - Machine Learning Prediction
- Summary of all results
  - Exploratory Data Analysis result
  - Interactive analytics in screenshots
  - Predictive Analytics result

Introduction

- Project background and context

  Space X advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because Space X can reuse the first stage. Therefore if we can determine if the first stage will land, we can determine the cost of a launch. This information can be used if an alternate company wants to bid against space X for a rocket launch. This project is focusing to predict success of first stage landing with machine learning approach

- Problems you want to find answers
  - What are the factors for successful landing
  - Statistics of landings

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - API (https://api.spacexdata.com/v4/rockets/)
  - Web Scarping (https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches)
- Perform data wrangling
  - Added labels into Training data with 1 means the booster successfully landed 0 means it was unsuccessful.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash

Methodology

Executive Summary

- Perform predictive analysis using classification models
  - Data were normalized, divided in training and test data sets and evaluated by four different classification models: Logistic regression, SVM, Decision tree, KNN and evaluated

Data Collection

- Datasets were collected from SpaceX API (https://api.spacexdata.com/v4/rockets/)

- and from Wikipedia (https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches), using web scraping technics.

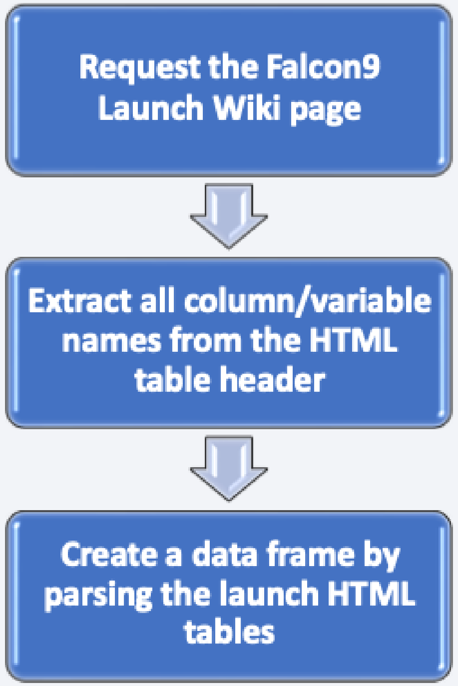
Data Collection – SpaceX API

- spacex_url=https://api.spacexdata.com/v4/launches/past
- response = requests.get(spacex_url)
- Link to GitHub: https://github.com/tomasa22/Coursera_IBM_Capstone/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

Data Collection - Scraping

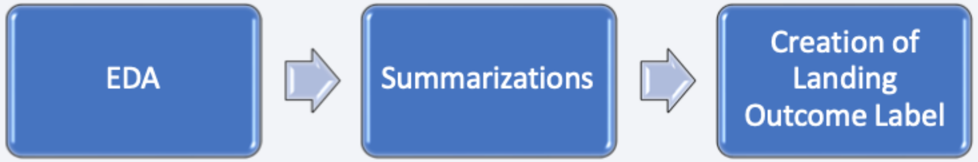
- Data obtained from: https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922

- Link to GitHub: https://github.com/tomasa22/Coursera_IBM_Capstone/blob/main/jupyter-labs-webscraping.ipynb



Data Wrangling

- Calculated the number of launches on each site
- Calculated the number and occurrence of each orbit
- Calculated the number and occurence of mission outcome of the orbits
- Created a landing outcome label



- Link to GutHuB: https://github.com/tomasa22/Coursera_IBM_Capstone/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- Following graphs were created to better understand data
  - relationship between Flight Number and Launch Site
  - relationship between Payload and Launch Site
  - relationship between success rate of each orbit type
  - relationship between FlightNumber and Orbit type
  - relationship between Payload and Orbit type
  - launch success yearly trend

- Link to GitHub: https://github.com/tomasa22/Coursera_IBM_Capstone/blob/main/edadataviz.ipynb

EDA with SQL

- SQL queries performed
  - Names of the unique launch sites in the space mission
  - 5 records where launch sites begin with the string 'CCA‘
  - The total payload mass carried by boosters launched by NASA (CRS)
  - Average payload mass carried by booster version F9 v1.1
  - The date when the first succesful landing outcome in ground pad was acheived.
  - The names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
  - The total number of successful and failure mission outcomes
  - The names of the booster_versions which have carried the maximum payload mass
  - 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.
- Link to GitHub: https://github.com/tomasa22/Coursera_IBM_Capstone/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Markers, circles, lines and marker clusters were used with Folium Maps
- Markers indicate points like launch sites;
- Circles indicate highlighted areas around specific coordinates, like NASA Johnson Space Center;
- Marker clusters indicates groups of events in each coordinate, like launches in a launch site; and
- Lines are used to indicate distances between two coordinates.
- Link to GitHub: https://github.com/tomasa22/Coursera_IBM_Capstone/blob/main/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- Created pie chart visualizing launch success counts
- Add a Range Slider to Select Payload
- Plot a scatter plot with the x axis to be the payload and the y axis to be the launch outcome
- Link To GitHub: https://github.com/tomasa22/Coursera_IBM_Capstone/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)

- LR, SVM, Decision Tree and KNN object were created and fit with provided GridSearchCV object to find best parameters.
- Accuracy of models were evaluated:

Find the method performs best:

```
print('LR accuracy: ',logreg_cv.score(X_test, Y_test))
print('SVM accuracy: ',svm_cv.score(X_test, Y_test))
print('Decision Tree accuracy: ',tree_cv.score(X_test,Y_test))
print('KNN accuracy: ',knn_cv.score(X_test,Y_test))
```

✓ 0.0s

```
LR accuracy:  0.8333333333333334
SVM accuracy:  0.8333333333333334
Decision Tree accuracy:  0.6666666666666666
KNN accuracy:  0.8333333333333334
```

- Link to GitHub: https://github.com/tomasa22/Coursera_IBM_Capstone/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

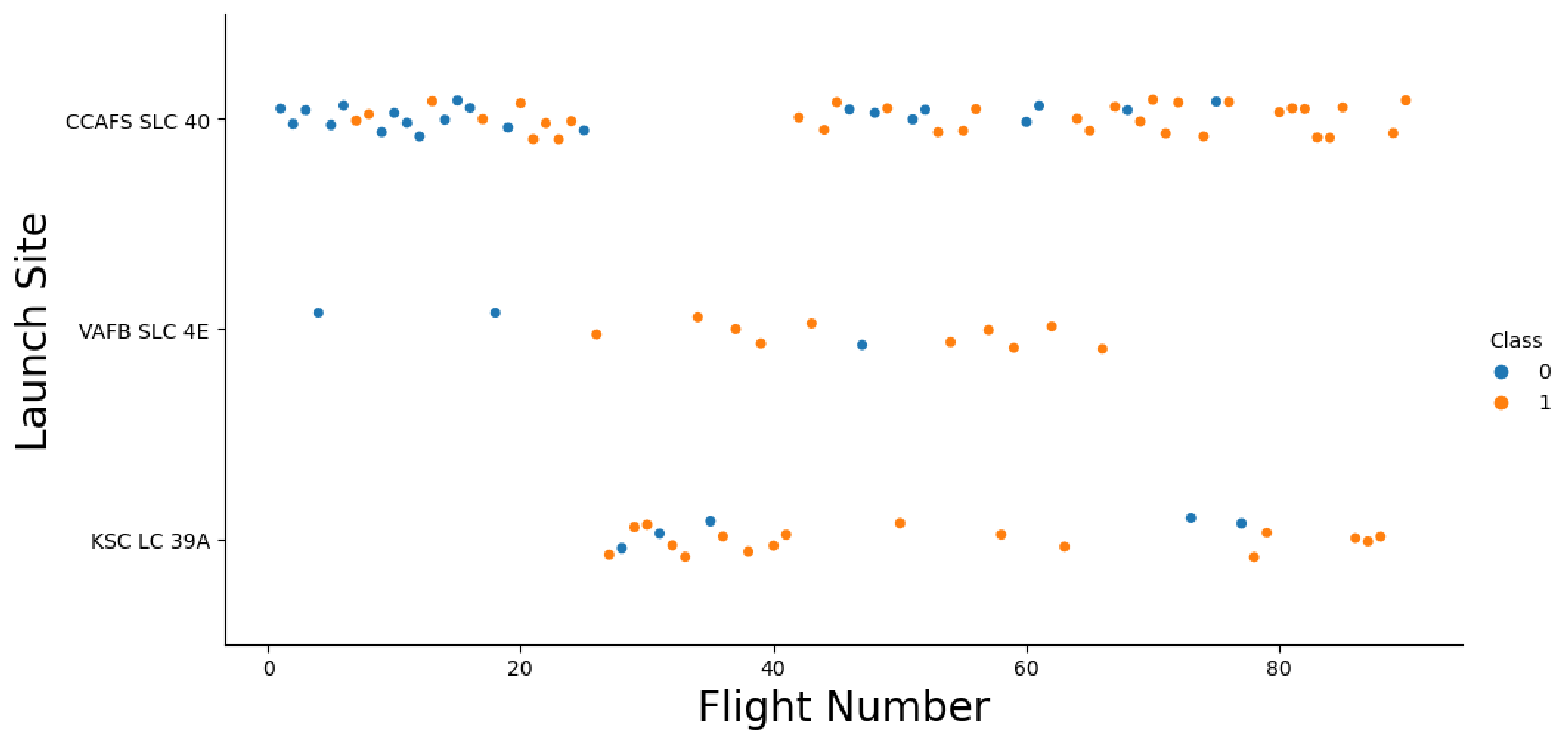
Results

- Success of launches for SpaceX increased with number of years of experience
- Best accuracy model are LR, SVM and KNN
  - 83.3% accuracy for all 3 models
  - Why are results so similar ?
- Most success launch sites is KSC LC-39A

Section 2

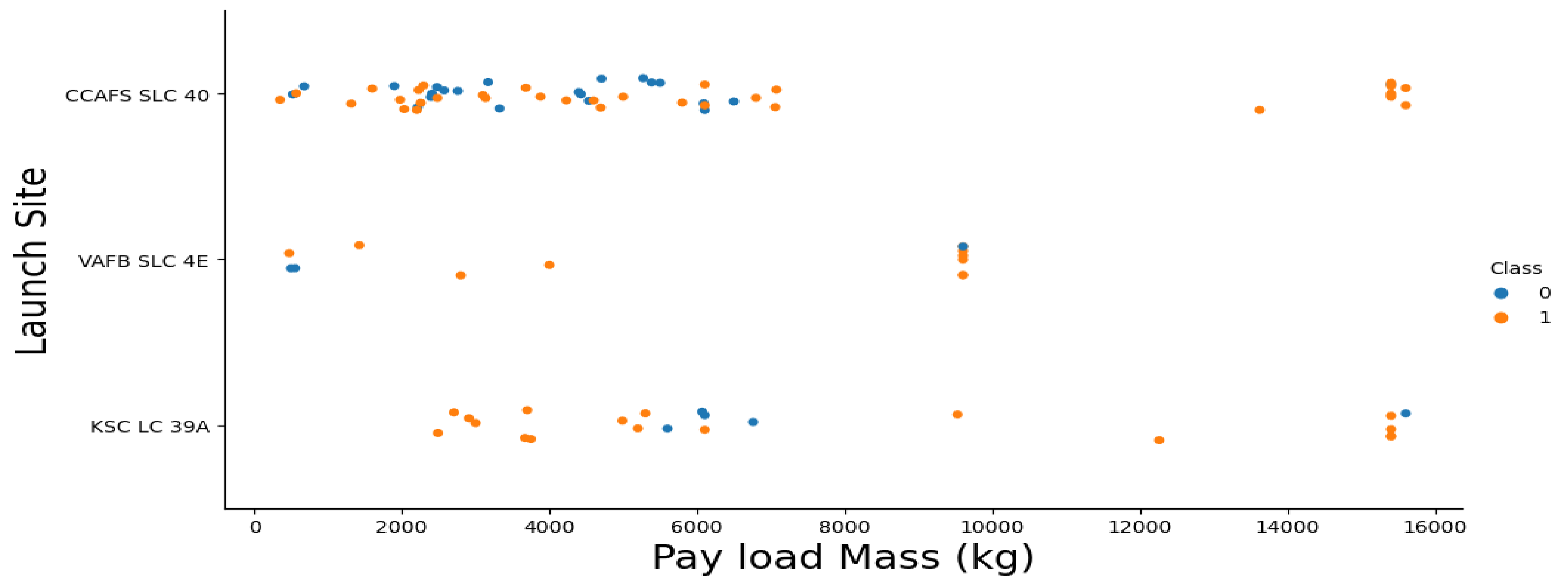
Insights drawn from EDA

Flight Number vs. Launch Site



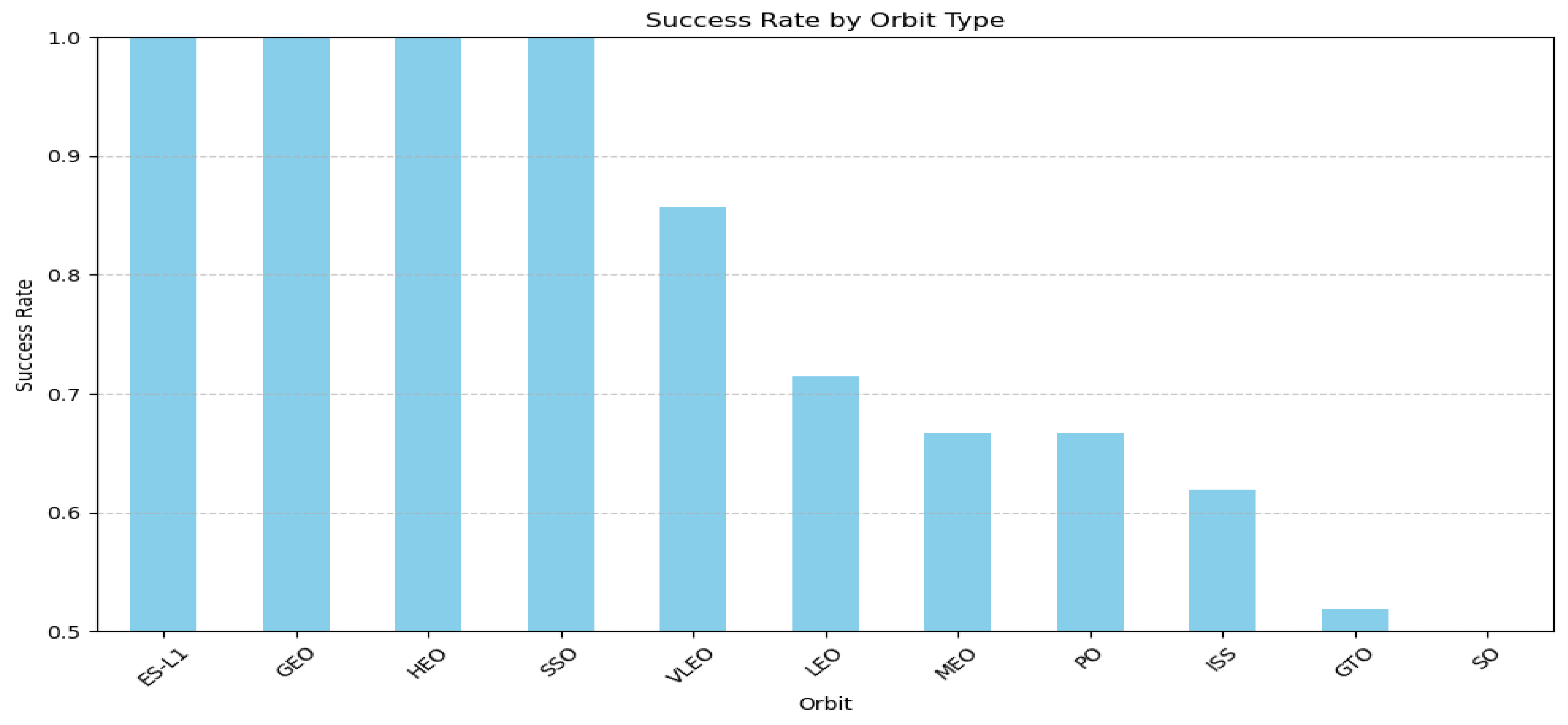
Most launches are from site CCAFS SLC 40

Payload vs. Launch Site



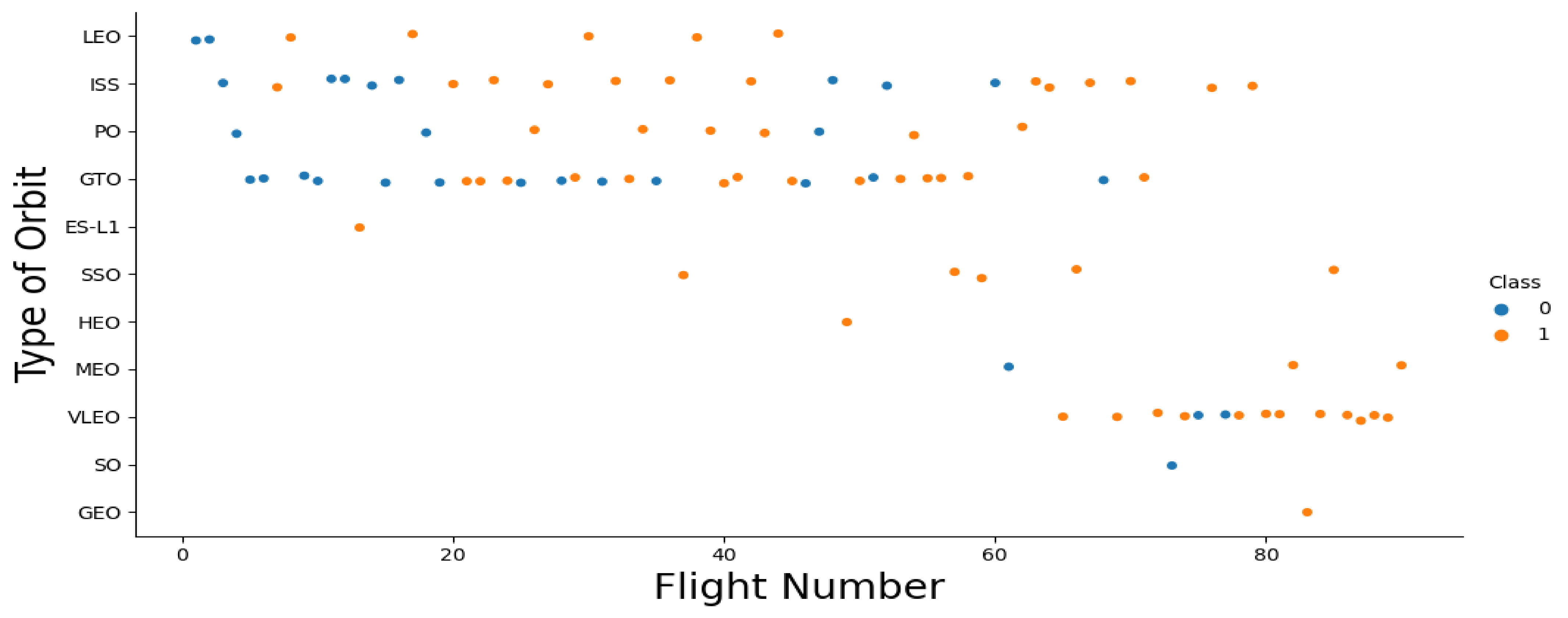
Most common launches are with payload up to 8000 kg

Success Rate vs. Orbit Type



Highest success rate are orbits : ES-L1, GEO, HEO, SSO

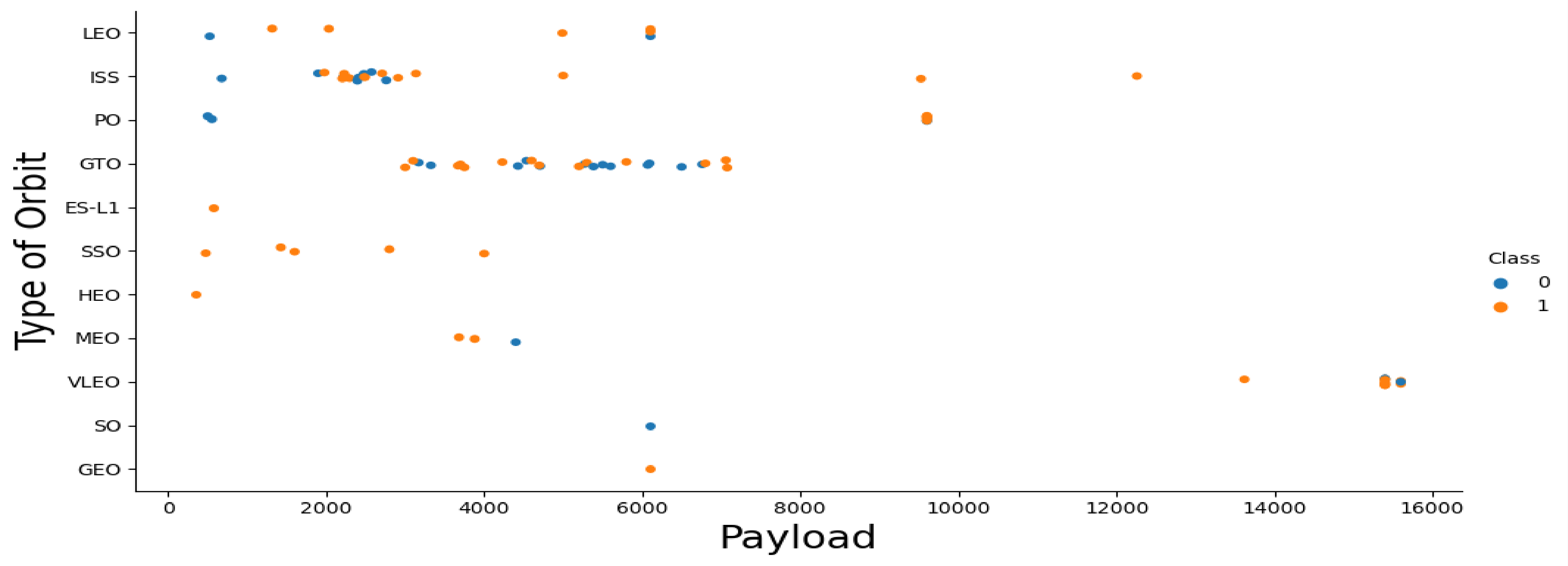
Flight Number vs. Orbit Type



With increasing flight numbers (years) more prominent type of orbits are reached.

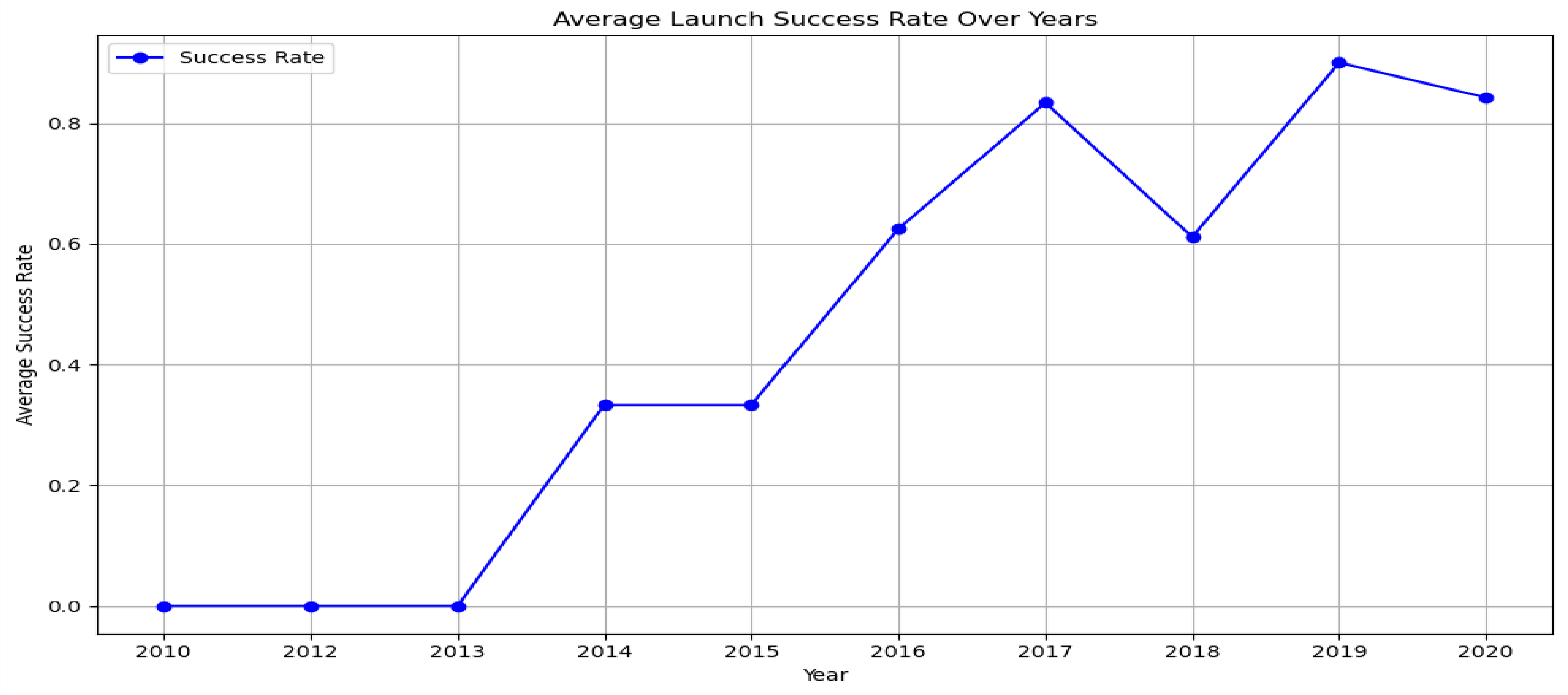
Showing increasing experience of SpaceX delivery system

Payload vs. Orbit Type



Heavy payloads have more successful rate

Launch Success Yearly Trend



After 2013 success of of launched increased systematically

All Launch Site Names

- CCAFS LC-40
- VAFB SLC-4E
- KCS LC-39A
- CCAFS SLC-40

In [6]:
```
%sql select distinct Launch_Site from SPACEXTABLE
```

Running query in 'sqlite:///my_data1.db'

Out[6]:

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

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`

In [8]:
```
%sql SELECT * FROM SPACEXTABLE WHERE Launch_Site LIKE 'CCA%' LIMIT 5;
```

Running query in 'sqlite:///my_data1.db'

Out[8]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	L
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

- Total payload carried by booster launched by NASA = 45596 kg

Task 3

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

[12]:

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

Running query in 'sqlite:///my_data1.db'

[12]:

TOTAL_PAYLOAD_MASS
45596

Average Payload Mass by F9 v1.1

- Average payload mass carried by booster version F9 v1.1 = 2928.4 kg

Task 4

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

```
%sql select avg(PAYLOAD_MASS__KG_) as AVERAGE_PAYLOAD_MASS from SPACEXTABLE where Booster_Version  = 'F9 v1.1'
```

Running query in 'sqlite:///my_data1.db'

AVERAGE_PAYLOAD_MASS
2928.4

First Successful Ground Landing Date

SQL query to find dates of the first successful landing outcome on ground pad

In [18]:
```
%sql select * from SPACEXTABLE where Landing_Outcome = 'Success (ground pad)' order by Date ASC limit 1
```

Running query in 'sqlite:///my_data1.db'

Out[18]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2015-12-22	1:29:00	F9 FT B1019	CCAFS LC-40	OG2 Mission 2 11 Orbcomm-OG2 satellites	2034	LEO	Orbcomm	Success	Success (ground pad)

Successful Drone Ship Landing with Payload between 4000 and 6000

- List the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

%sql select distinct Booster_Version from SPACEXTABLE where Landing_Outcome = 'Success (drone ship)' and PAYLOAD_MASS__KG_ between 4000 and 6000

Task 6

List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

```
%sql select distinct Booster_Version from SPACEXTABLE where Landing_Outcome = 'Success (drone ship)' and PAYLOAD_MASS__KG_ b
```

unning query in 'sqlite:///my_data1.db'

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 query

Task 7

List the total number of successful and failure mission outcomes

```
%sql SELECT Mission_Outcome, COUNT(*) AS Count FROM SPACEXTABLE GROUP BY Mission_Outcome
```

:unning query in 'sqlite:///my_data1.db'

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

Boosters Carried Maximum Payload

- SQL query to list of the booster version which have carried maximum payload

Task 8

List the names of the booster_versions which have carried the maximum payload mass. Use a subquery

```
]: %%sql SELECT Booster_Version
FROM SPACEXTABLE
WHERE PAYLOAD_MASS__KG_ = (
     SELECT MAX(PAYLOAD_MASS__KG_)
     FROM SPACEXTABLE
)
```

Running query in 'sqlite:///my_data1.db'

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

2015 Launch Records

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

In [41]:

```
%%sql SELECT
    CASE substr(Date, 6, 2)
        WHEN '01' THEN 'January'
        WHEN '02' THEN 'February'
        WHEN '03' THEN 'March'
        WHEN '04' THEN 'April'
        WHEN '05' THEN 'May'
        WHEN '06' THEN 'June'
        WHEN '07' THEN 'July'
        WHEN '08' THEN 'August'
        WHEN '09' THEN 'September'
        WHEN '10' THEN 'October'
        WHEN '11' THEN 'November'
        WHEN '12' THEN 'December'
    END AS Month_Name,  -- Extract month (2 digits) and renema it to names of months
  Landing_Outcome,
  Booster_Version,
  Launch_Site,
  Date
FROM SPACEXTABLE
WHERE Landing_Outcome = 'Failure (drone ship)'
AND substr(Date, 1, 4) = '2015'   -- Filter for year 2015
```

Running query in 'sqlite:///my_data1.db'

Out[41]:

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

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

- 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

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.

```
%%sql SELECT Landing_Outcome, COUNT(*) AS Count
FROM SPACEXTABLE
WHERE Landing_Outcome IN ('Success (ground pad)', 'Failure (drone ship)', 'Failure (ocean)', 'Success (drone ship)')
  AND DATE BETWEEN '2010-06-04' AND '2017-03-20'
GROUP BY Landing_Outcome
ORDER BY Count DESC;
```

Running query in 'sqlite:///my_data1.db'

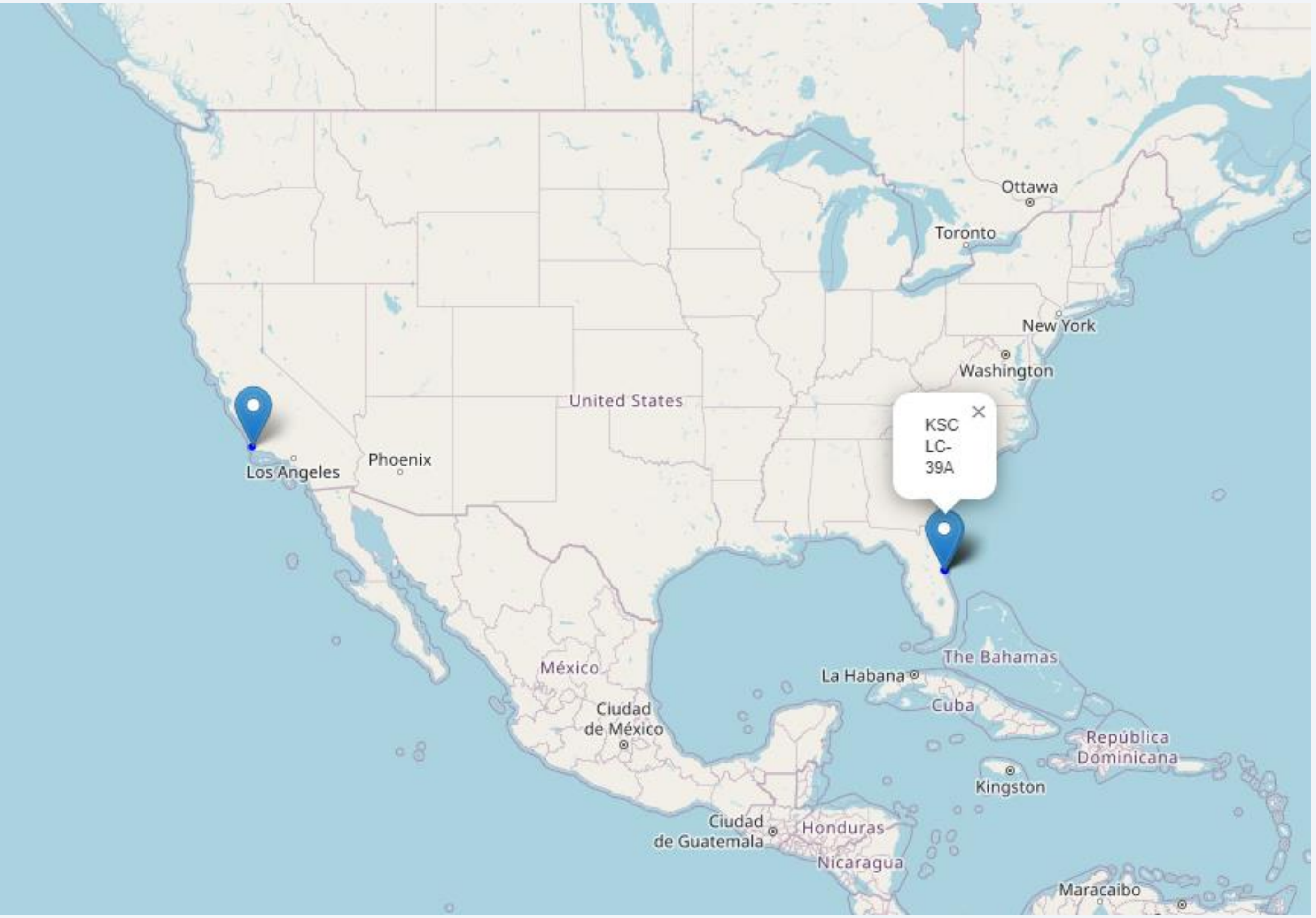
Landing_Outcome	Count
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3

Section 3

Launch Sites Proximities Analysis

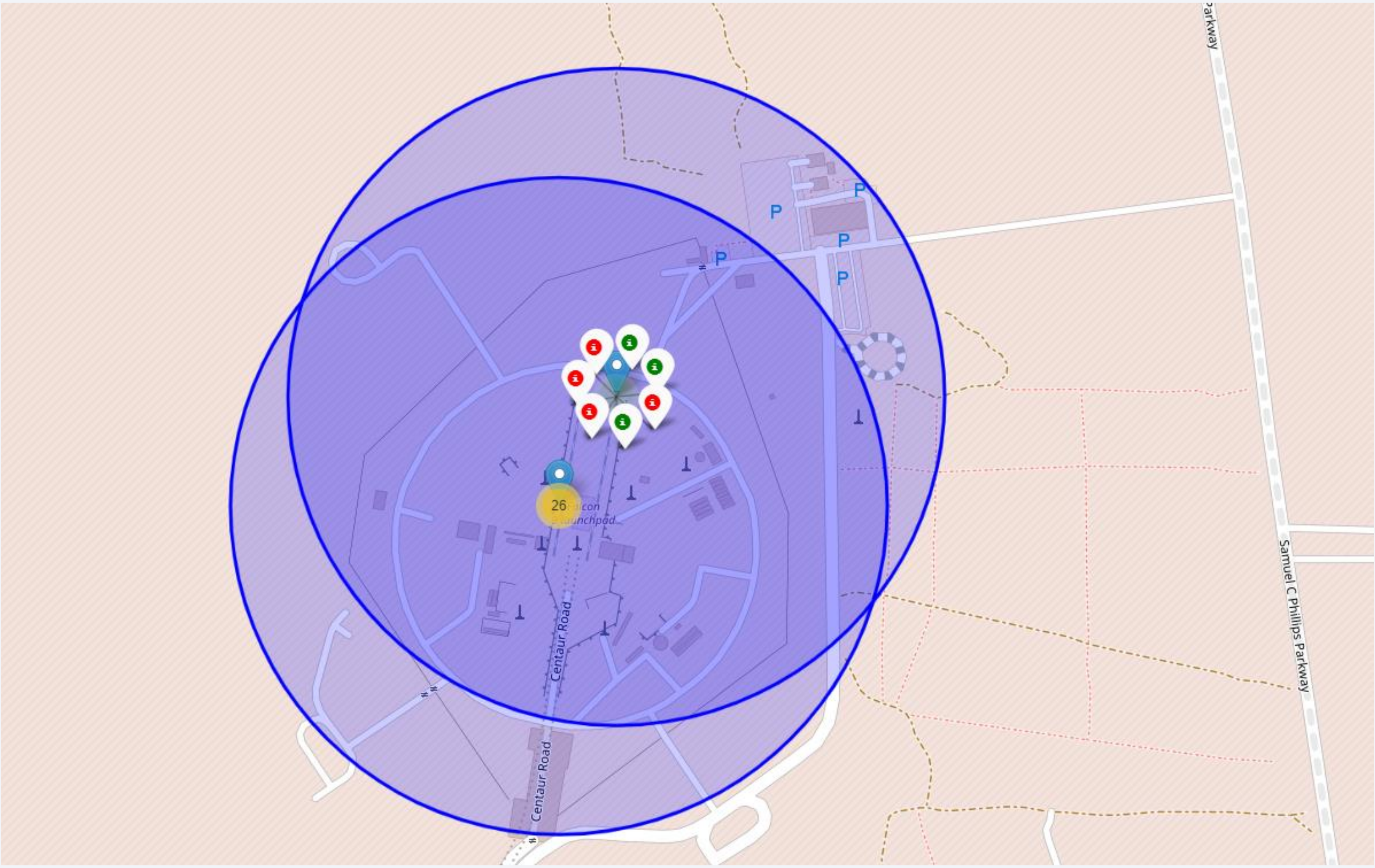
Position of launch sites on map

- Position of launch sites on map with markers



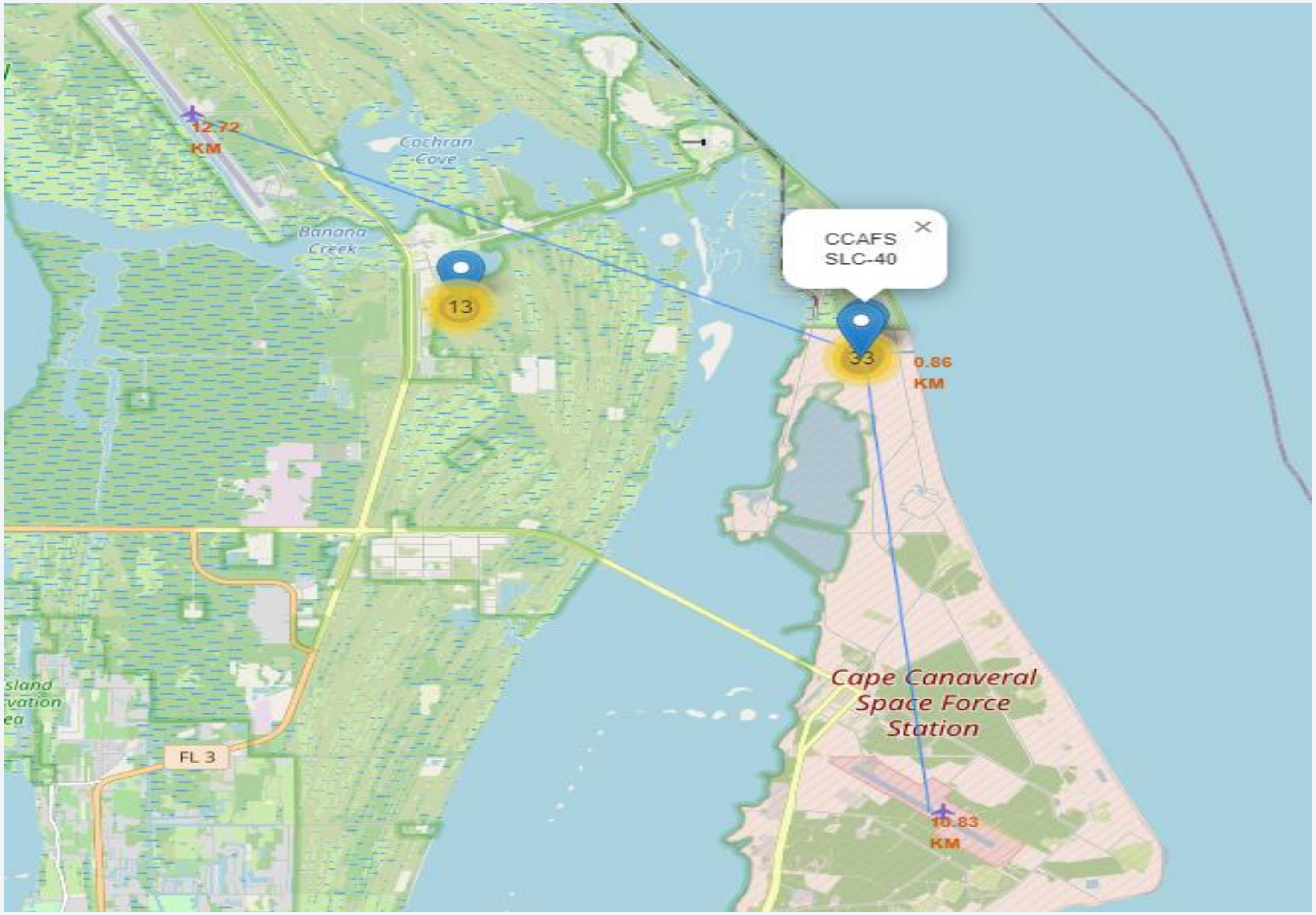
Success/Failure launches for each launch site

- Launches are grouped to clusters are labeled by green marker for success and red for failure



Nearest landing strips and shoreline

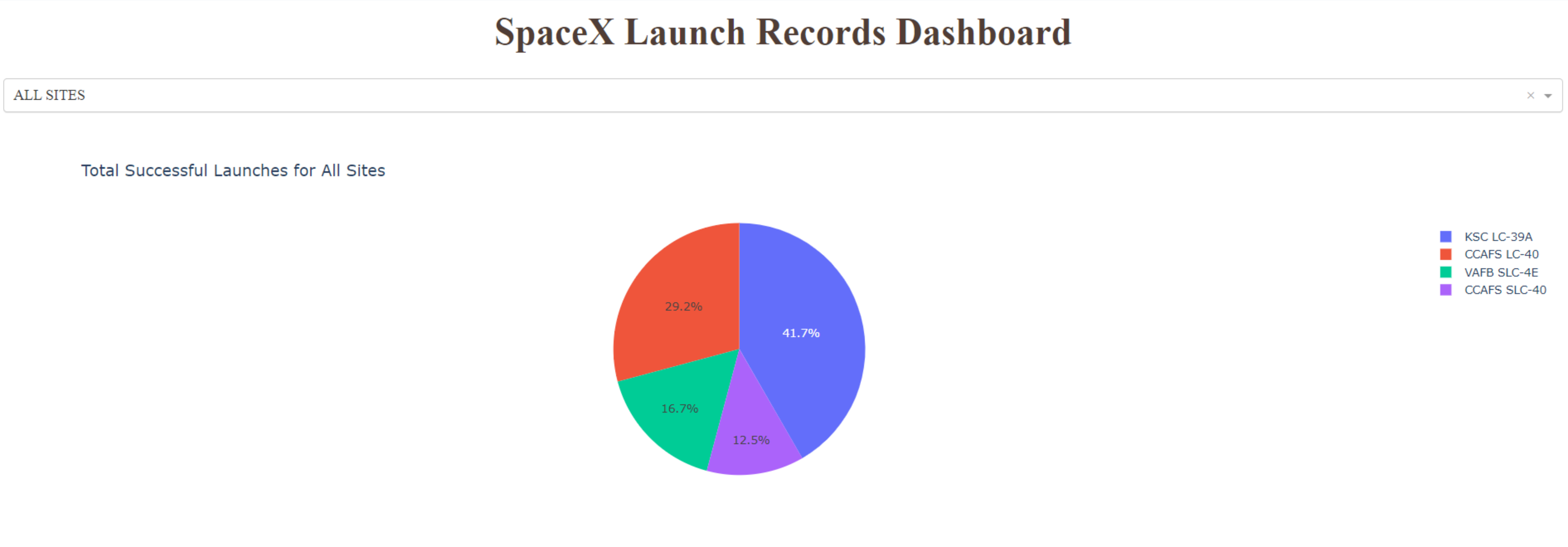
- Nearest landing strips (in km) and shorelines for launch site CCAFS SLC-40



Section 4

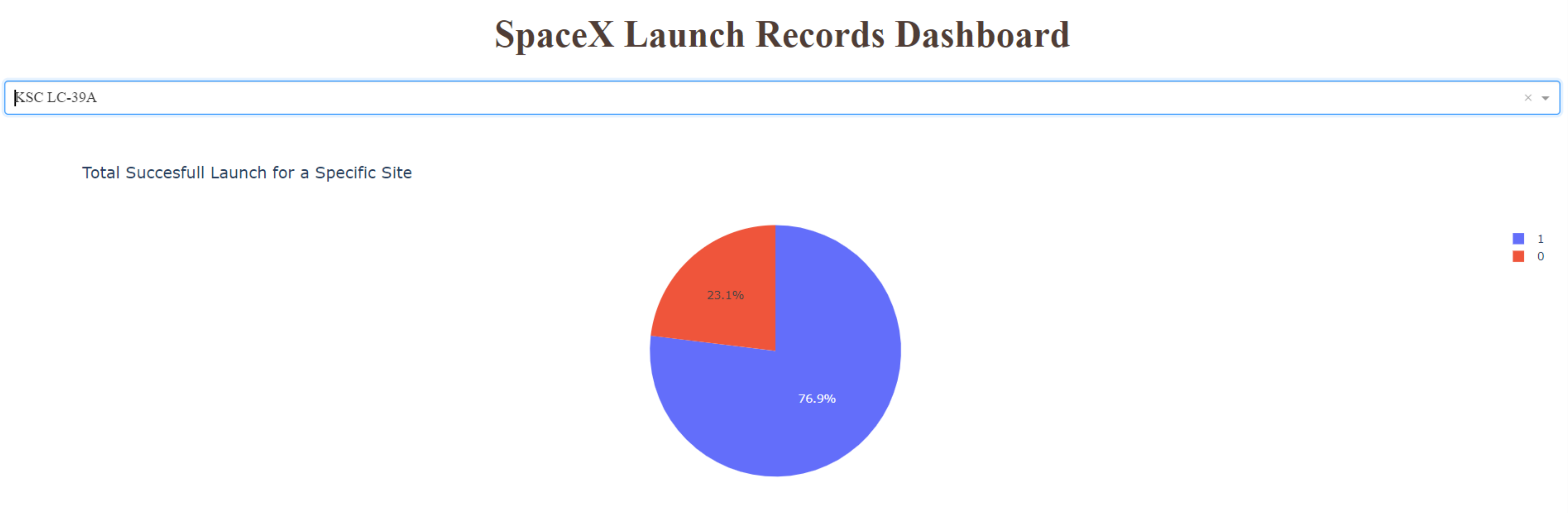
Build a Dashboard with Plotly Dash

Total Launches for All Sites



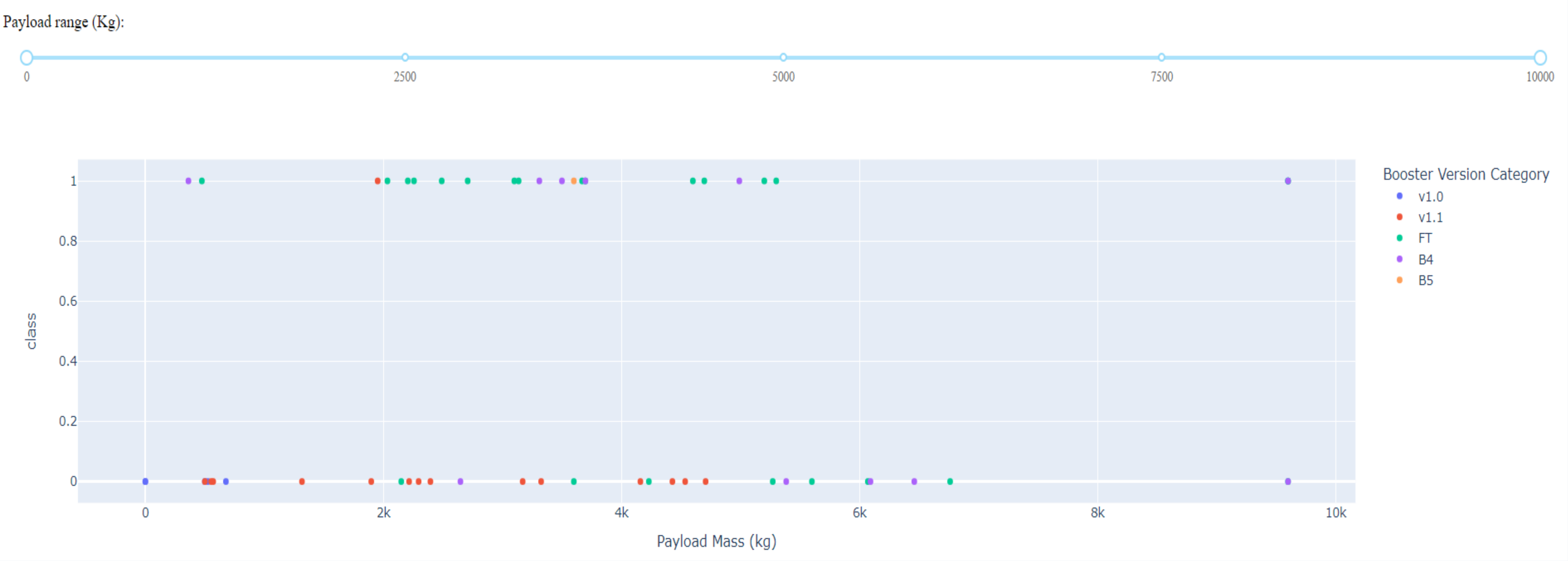
Launch site KCS LC-39A has the most successful rate with 41.7%

Total Successful Launch for Specific Site



KSC LC-39A has 76.9% success rate and failure rate 23.1%

Payload vs launch outcome

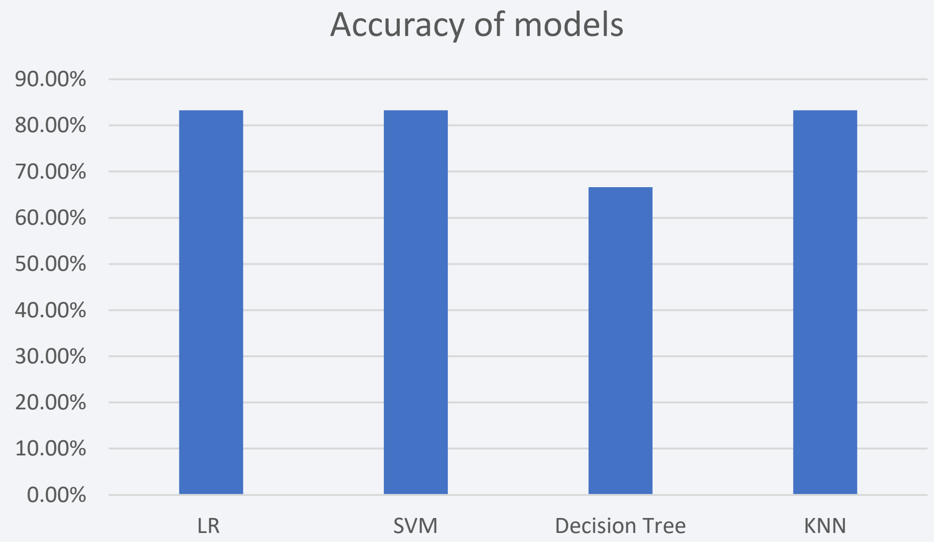


Here is show success/failure (class) for each booster version and how much payload they were carrying

Section 5

Predictive Analysis (Classification)

Classification Accuracy



Find the method performs best:

```
print('LR accuracy: ',logreg_cv.score(X_test, Y_test))
print('SVM accuracy: ',svm_cv.score(X_test, Y_test))
print('Decision Tree accuracy: ',tree_cv.score(X_test,Y_test))
print('KNN accuracy: ',knn_cv.score(X_test,Y_test))
```

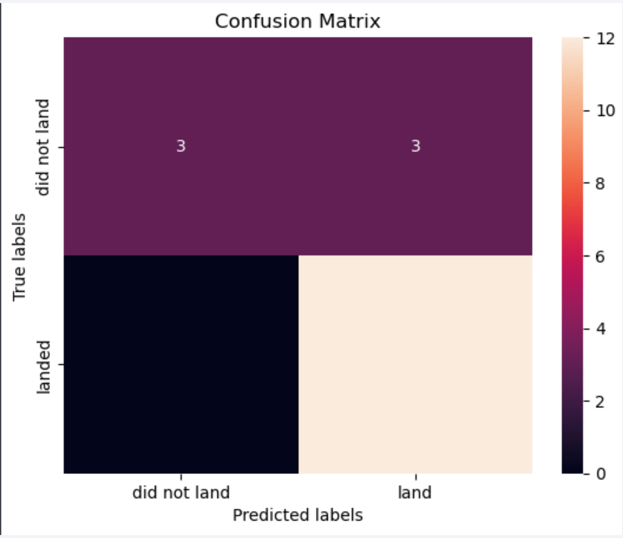
✓ 0.0s

```
LR accuracy:  0.8333333333333334
SVM accuracy:  0.8333333333333334
Decision Tree accuracy:  0.6666666666666666
KNN accuracy:  0.8333333333333334
```

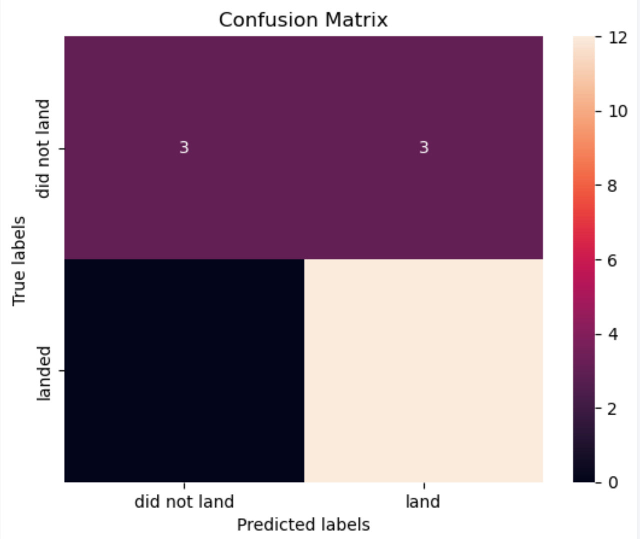
- Three best performing models are LR, SVM and KNN

Confusion Matrix

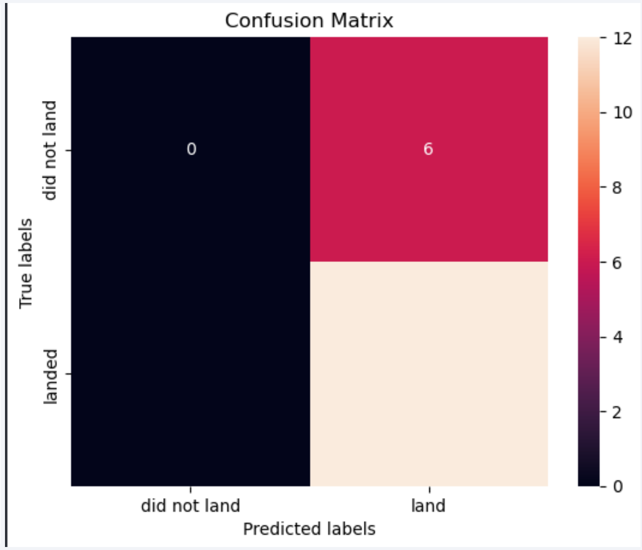
- Here is my problem I have 3 best models not only one and final results are confusing: no number for did not land , land in Landed idk why ☺



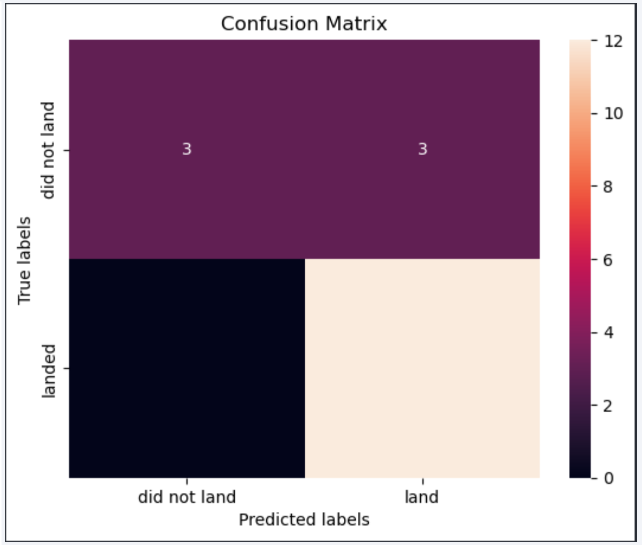
LR



SVM



Decision Tree



KNN

Conclusions

- The larger the flight amount at a launch site, the greater the success rate at a launch site.
- Launch success rate started to increase in 2013 till 2020.
- Orbits ES-L1, GEO, HEO, SSO, VLEO had the most success rate.
- KSC LC-39A had the most successful launches of any sites

Thank you!