SpaceX Falcon 9 Rocket First Stage Launches

THAM NGUYEN

6/1/2022

IBM Developer

SKILLS NETWORK

OUTLINE

- Executive Summary
- Introduction
- Methodology
- Results
  - Visualization – Charts
  - Dashboard
- Discussion
  - Findings & Implications
- Conclusion
- Appendix

EXECUTIVE SUMMARY

- **Summary of the methodologies**
  - Data collection using an API
  - Collect data with Web Scraping
  - Data Wrangling
  - Exploratory Data Analysis with SQL
  - Exploratory Data Analysis with Visualizations
  - Interactives with Folium and Plotly Dash
  - Machine Learning predictions
- **Summary of the results**
  - There are three launch sites located in Florida and one in California
  - Launch site KSC LC-39A is located in Florida and is more successful at landing rockets
  - The success of a rocket launch depends on the payload mass in a certain range
  - Launch sites have to be near railways for transport and far away from cities for safety

INTRODUCTION

- SpaceX 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 SpaceX can reuse the first stage
- If you can accurately predict the likelihood of the first stage rocket landing successfully, you can determine the cost of a launch
- Identify features that will produce successful landing outcomes through correlation techniques
- Find a successful predictive model to help outbid other competitors for a SpaceX rocket launch

METHODOLOGY

- Collect launch data using SpaceX API
- Use Python to web scrape tables of launch records from Wikipedia
- Exploratory Data Analysis with data wrangling
- Load SpaceX dataset into an IBM database and use SQL to extract launch information
- Use Pandas Dataframe to find which features to use to predict landing outcomes
- Create Folium maps and Plotly dashboards to further visualize launch data
- Find the best machine learning method to predict successful landings. The methods include Logistics Regression, Support Vector Machine, Decision Tree, and K Nearest Neighbors

Data Collection with SpaceX API

Request to the SpaceX API

spacex_url="https://api.spacexdata.com/v4/launches/past"

response = requests.get(spacex_url)

This API will help extract SpaceX launch information

This URL contains the rocket launch data

The response will retrieve launch information in JSON format

GitHub URL

Data Collection with Web Scraping

URL to extract a Falcon 9 launch records HTML table from Wikipedia

```
static_url = "https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922"
```

Request the Falcon9 Launch Wiki page from its URL

```
response = requests.get(static_url)
```

Create a BeautifulSoup object from the HTML response

```
soup = BeautifulSoup(response.content, 'lxml')
```

[GitHub URL]

Data Wrangling Methodology

Calculate the number of launches on each site

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

CCAFS SLC 40     55
KSC LC 39A       22
VAFB SLC 4E      13
Name: LaunchSite, dtype: int64
```

Calculate the number and occurrence of each orbit

```
# Apply value_counts on Orbit column
df['Orbit'].value_counts()
GTO        27
ISS        21
VLEO       14
PO          9
LEO         7
SSO         5
MEO         3
HEO         1
GEO         1
ES-L1       1
SO          1
Name: Orbit, dtype: int64
```

Determine the number of landing outcomes

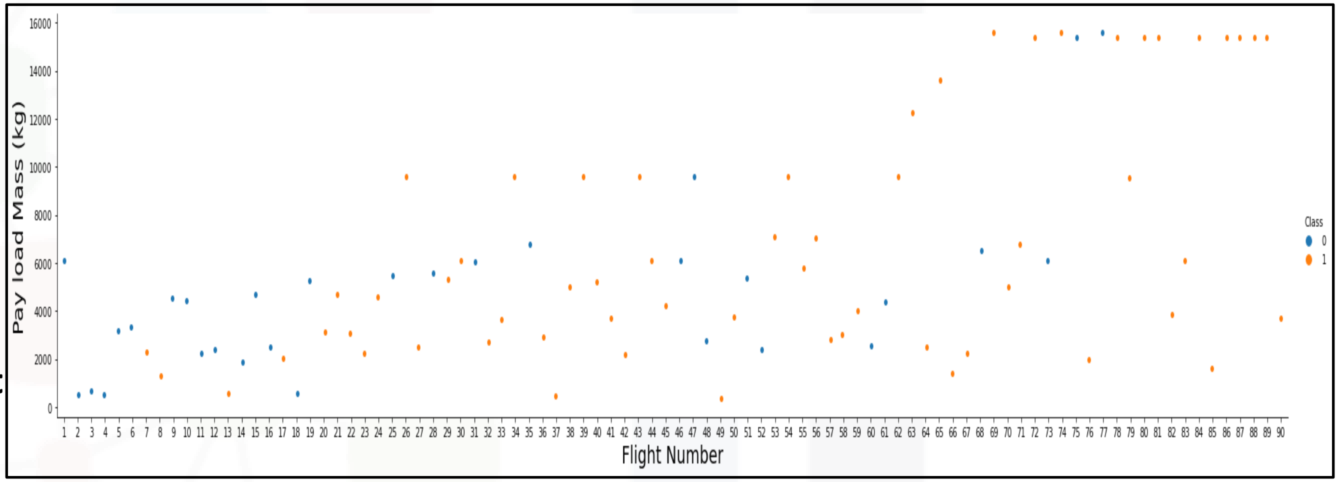
```
# landing_outcomes = values on Outcome column
landing_outcomes = df['Outcome'].value_counts()
landing_outcomes

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
```

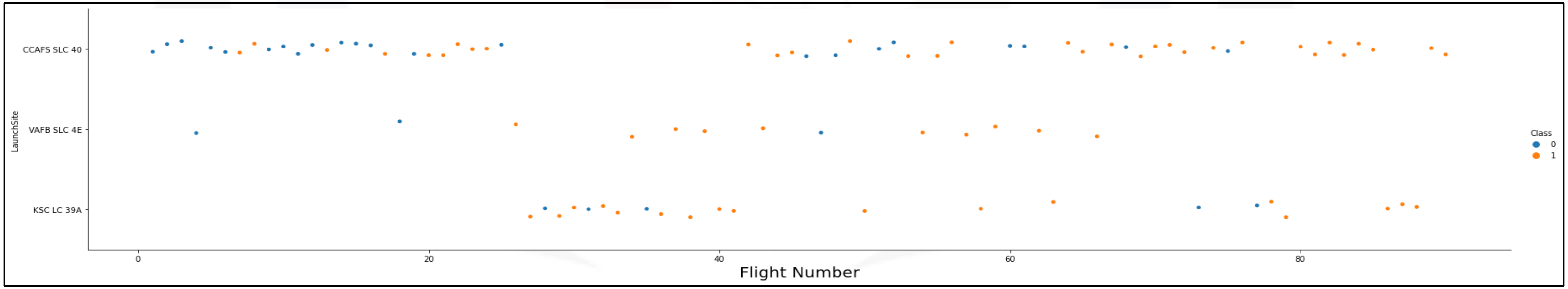
EDA with Data Visualization

These charts show that the flight numbers will affect the outcome of the rocket landing

We see that as the flight number increases, the first stage is more likely to land successfully. Also, the more massive the payload, the less likely it will return.



As the flight number increases, all launch sites have success at landing rockets.



EDA with SQL

- Displayed the names of the unique launch sites in the space mission
- Displayed 5 records where launch sites begin with the string 'CCA'
- Displayed the total payload mass carried by boosters launched by NASA (CRS)
- Displayed average payload mass carried by booster version F9 v1.1
- Listed the date when the first successful landing outcome in ground pad was achieved.
- Listed the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- Listed the total number of successful and failure mission outcomes
- Listed the names of the booster versions which have carried the maximum payload mass
- Listed the failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015
- Ranked 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

GitHub URL

Interactive Visual Analytics with Folium Map and Plotly Dash

- Marked all launch sites on a map by using the sites latitude and longitude coordinate
- Used a Folium circle to highlight the area with a text label on a specific coordinate
- Marked the success/failed launches for each site on the map
- Calculated the distances between a launch site to its proximities, such as coastline, railway, highway, and city
- Created a SpaceX Launch Record Dashboard on Plotly Dash showing a drop down list, pie chart, and scatter chart

GitHub URL

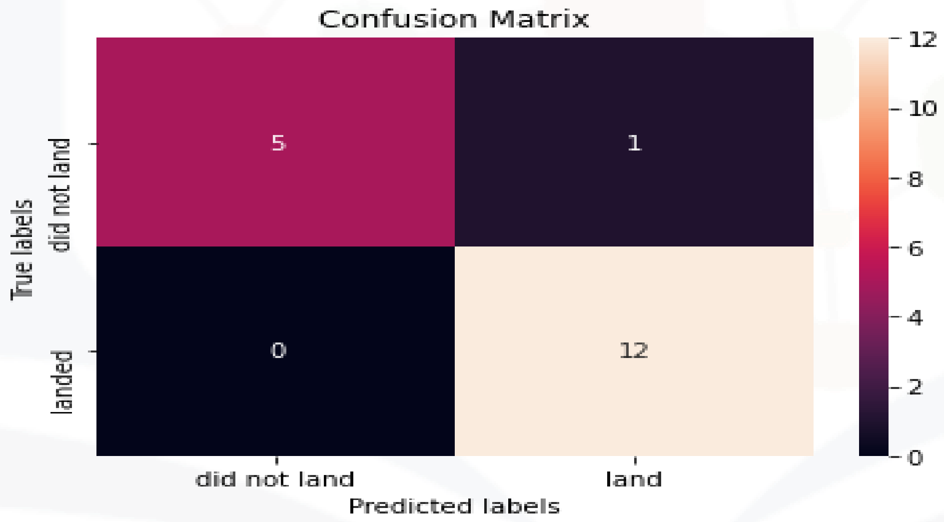
Predictive Analysis Using Classification Models

Create a decision tree classifier object then create a `GridSearchCV` object `tree_cv` with cv = 10

Fit the object to find the best parameters from the dictionary `parameters`

Calculate the accuracy of tree_cv on the test data using the method `score`. `This model has the highest accuracy score compared to other models.`

Confusion matrix of a decision tree classifier

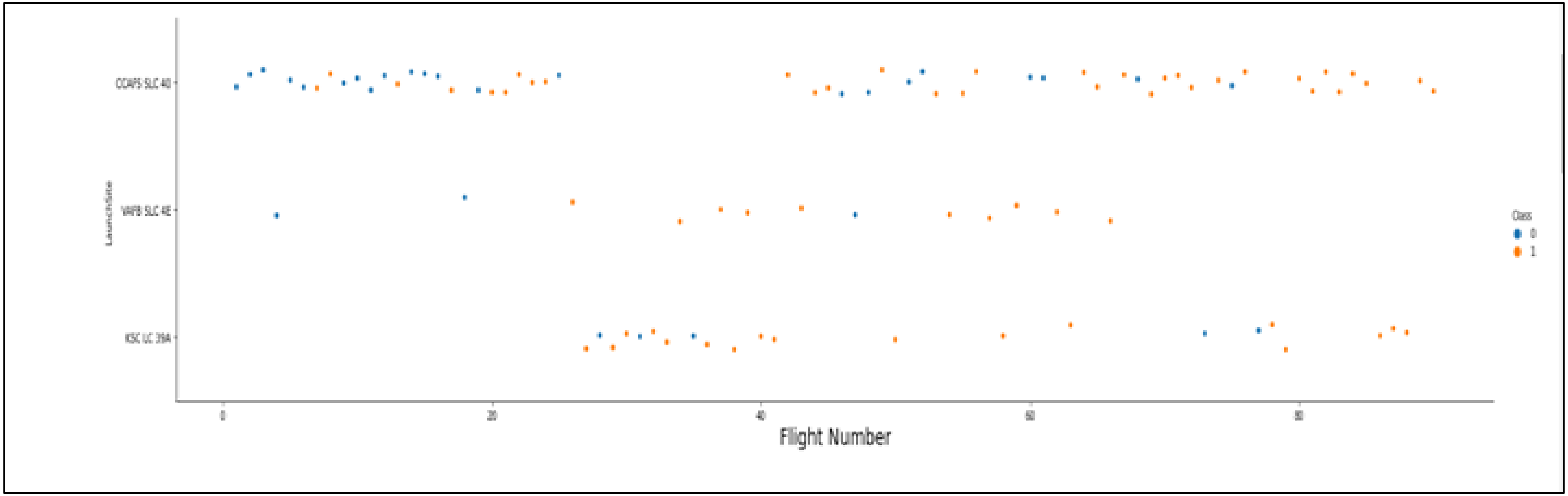



```
tree_cv.score(X_test,Y_test)
0.9444444444444444
```

[GitHub URL]

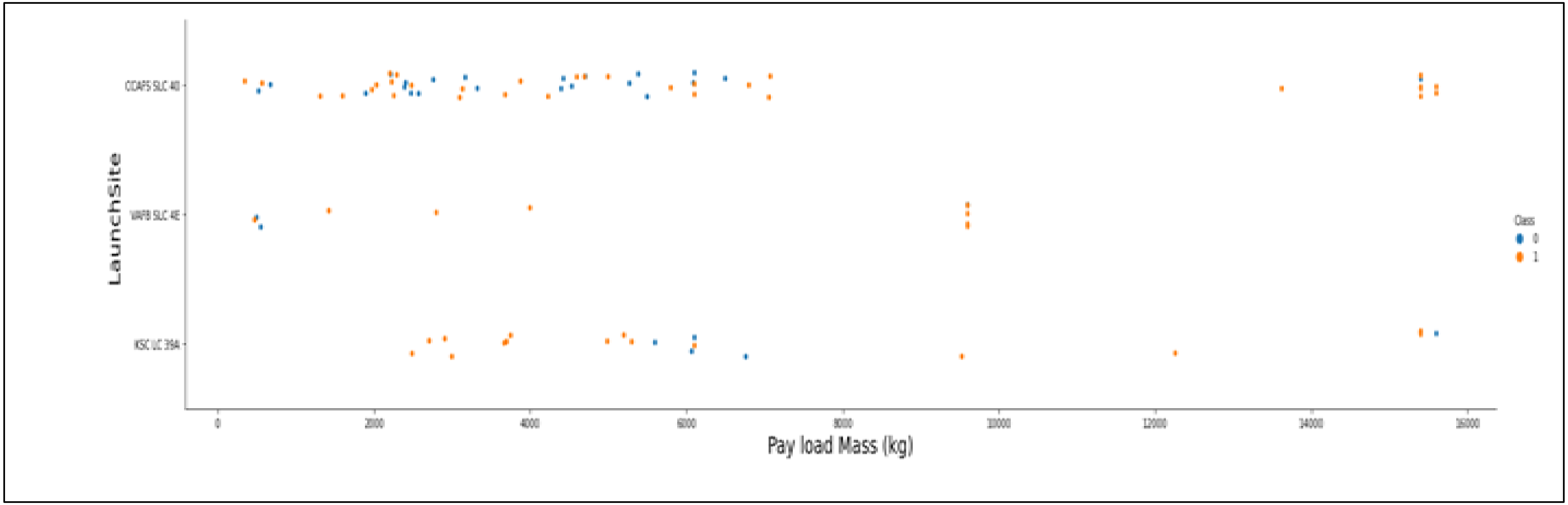
Visualization

Flight Number vs. Launch Site



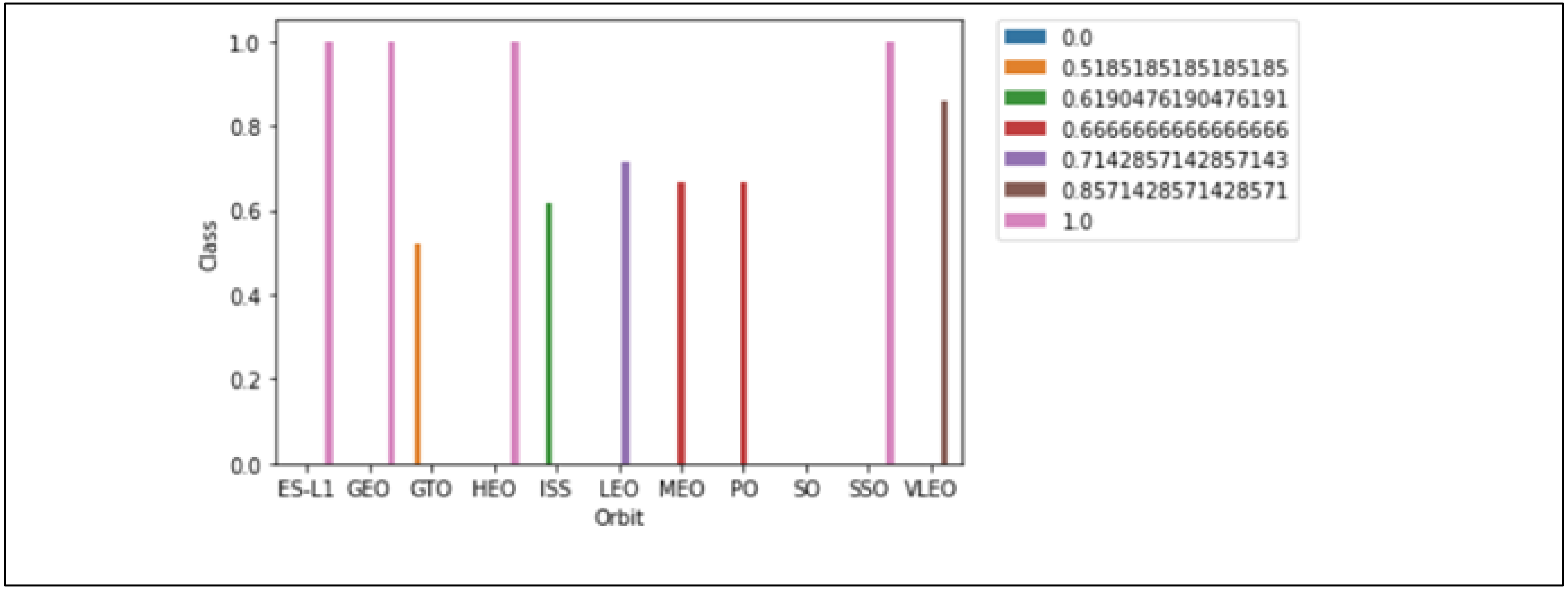
As the Flight Number increased, there were more successful landings at different launch sites

Payload vs. Launch Site



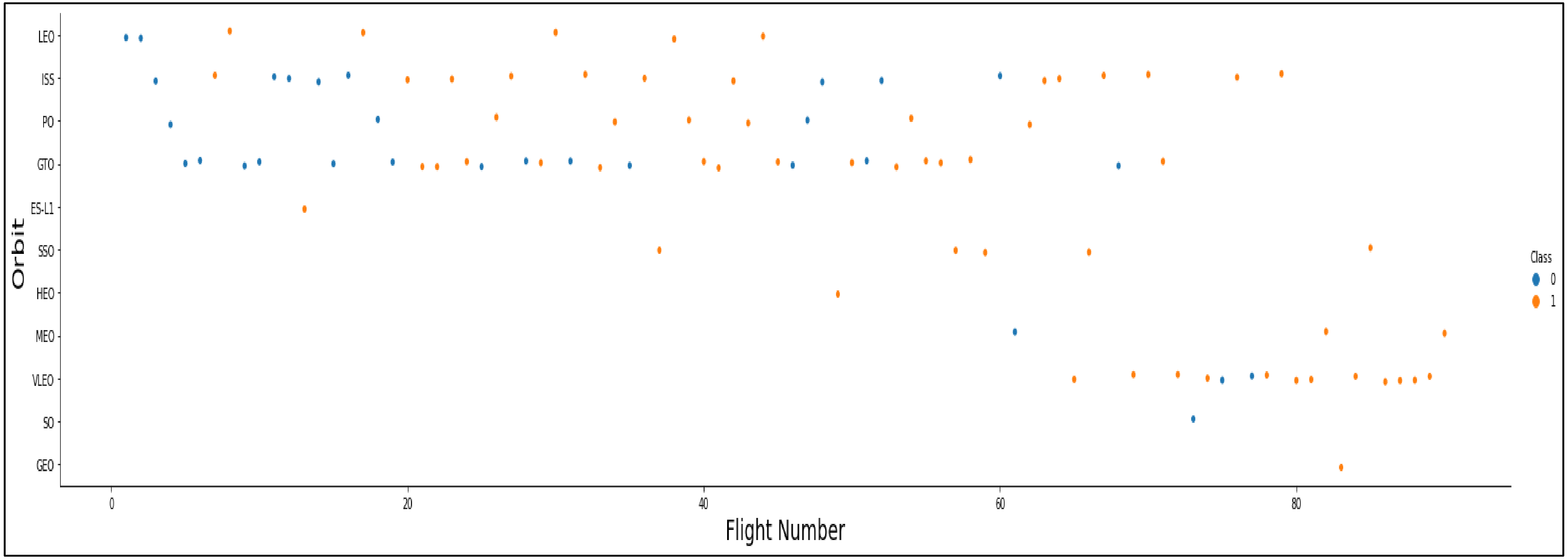
For the VAFB-SLC launch site, there were no rockets launched for heavy payload mass(greater than 10000 kg)

Success Rate vs. Orbit Type



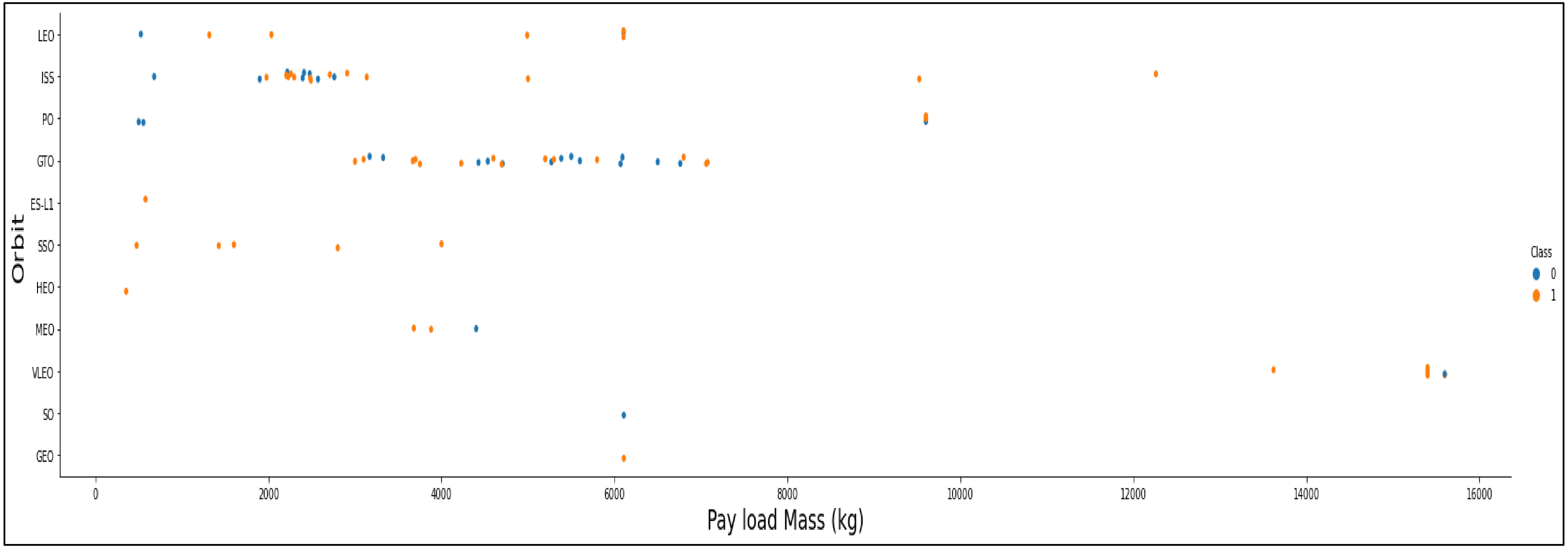
Orbits ES-L1, GEO, HEO, and SSO have high success rates

Flight Number vs. Orbit Type



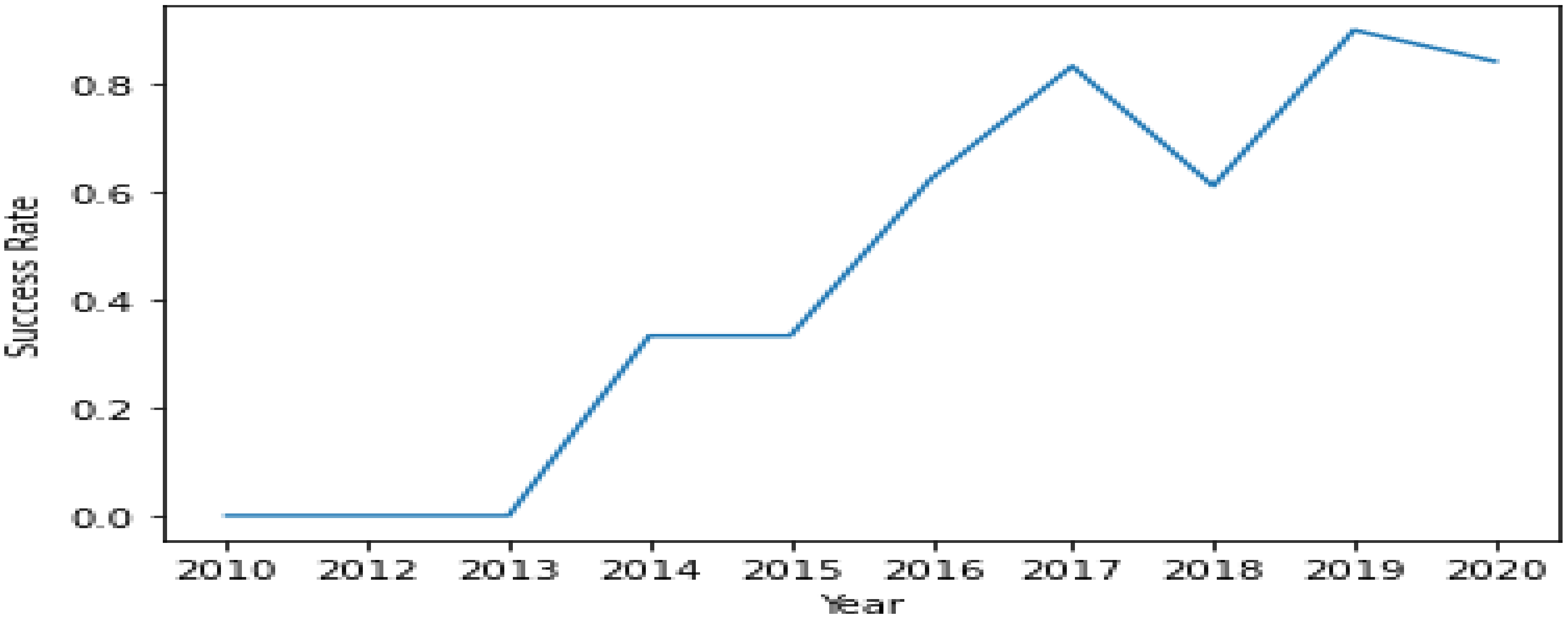
LEO orbit success appears related to the number of flights, however there seems to be no relationship between flight number when in GTO orbit

Payload vs. Orbit Type



With heavy payloads the successful landing rate are more for Polar, LEO and ISS. GTO orbit has mixed results.

Success Yearly Trend Line



The sucess rate since 2013 kept increasing till 2020

SQL Queries

```
%%sql

select DISTINCT(LAUNCH_SITE)
from SPACEX;
```

```
 * ibm_db_sa://wkg27603:***@125f9f61-9715-46f9-9399-c8177b21803b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:30426/bludb
Done.
```

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

Names of the unique launch sites in the space mission

```
%%sql

SELECT *
FROM spacex
WHERE launch_site LIKE 'CCA%'
LIMIT 5;
```

```
 * ibm_db_sa://wkg27603:***@125f9f61-9715-46f9-9399-c8177b21803b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:30426/bludb
Done.
```

DATE	time__utc_	booster_version	launch_site	payload	payload_mass__kg_	orbit	customer	mission_outcome	landing__outcome
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
2014-04-18	19:25:00	F9 v1.1	CCAFS LC-40	SpaceX CRS-3	2296	LEO (ISS)	NASA (CRS)	Success	Controlled (ocean)
2014-07-14	15:15:00	F9 v1.1	CCAFS LC-40	OG2 Mission 1 6 Orbcomm-OG2 satellites	1316	LEO	Orbcomm	Success	Controlled (ocean)
2014-09-21	5:52:00	F9 v1.1 B1010	CCAFS LC-40	SpaceX CRS-4	2216	LEO (ISS)	NASA (CRS)	Success	Uncontrolled (ocean)
2015-04-14	20:10:00	F9 v1.1 B1015	CCAFS LC-40	SpaceX CRS-6	1898	LEO (ISS)	NASA (CRS)	Success	Failure (drone ship)

5 records where launch sites begin with the string 'CCA'

```
%%sql

SELECT SUM(payload_mass__kg_)
FROM spacex
WHERE customer = 'NASA (CRS)';
```

```
 * ibm_db_sa://wkg27603:***@125f9f61-9715-46f9-9399-c8177b21803b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:30426/bludb
Done.
```

1
23589

Total payload mass carried by boosters launched by NASA (CRS)

```
%%sql

SELECT AVG(payload_mass__kg_)
FROM spacex
WHERE booster_version = 'F9 v1.1';
```

```
 * ibm_db_sa://wkg27603:***@125f9f61-9715-46f9-9399-c8177b21803b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:30426/bludb
Done.
```

1
1806

Average payload mass carried by booster version F9 v1.1

```
%%sql

SELECT MIN(DATE)
FROM spacex
WHERE landing__outcome = 'Success (ground pad)';
```

```
 * ibm_db_sa://wkg27603:***@125f9f61-9715-46f9-9399-c8177b21803b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:30426/bludb
Done.
```

1
2015-12-22

The date when the first successful landing outcome in ground pad was achieved

```
%%sql

SELECT Booster_Version
FROM spacex
WHERE landing__outcome = 'Success (drone ship)' AND payload_mass__kg_ > 4000 AND payload_mass__kg_ < 6000;
```

```
 * ibm_db_sa://wkg27603:***@125f9f61-9715-46f9-9399-c8177b21803b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:30426/bludb
Done.
```

booster_version
F9 FT B1026
F9 FT B1021.2

Names of booster versions which have success in drone ship and have payload mass greater than 4000 kg but less than 6000 kg

```
%%sql

SELECT mission_outcome, COUNT(mission_outcome) AS count
FROM spacex
GROUP BY mission_outcome;
```

```
 * ibm_db_sa://wkg27603:***@125f9f61-9715-46f9-9399-c8177b21803b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:30426/bludb
Done.
```

mission_outcome	COUNT
Failure (in flight)	1
Success	55

The total number of successful and failure mission outcomes

```
%%sql

SELECT booster_version
FROM spacex
WHERE payload_mass__kg_ = (SELECT MAX(payload_mass__kg_) FROM spacex);
```

```
 * ibm_db_sa://wkg27603:***@125f9f61-9715-46f9-9399-c8177b21803b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:30426/bludb
Done.
```

booster_version
F9 B5 B1051.3
F9 B5 B1056.4
F9 B5 B1048.5
F9 B5 B1051.4
F9 B5 B1051.6
F9 B5 B1060.3
F9 B5 B1049.7

Names of booster versions which have carried the maximum payload mass

```
%%sql

SELECT booster_version, launch_site
FROM spacex
WHERE landing_outcome = 'Failure (drone ship)' AND date like '2015%'
```

```
 * ibm_db_sa://wkg27603:***@125f9f61-9715-46f9-9399-c8177b21803b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:30426/bludb
Done.
```

booster_version	launch_site
F9 v1.1 B1015	CCAFS LC-40

The failed landing outcome, it's booster version, and launch site name in 2015

```
%%sql

SELECT landing__outcome, COUNT(landing__outcome) AS COUNT
FROM spacex
WHERE date > '2010-06-04' AND date < '2017-03-20'
GROUP BY landing__outcome
ORDER BY COUNT desc;
```

```
 * ibm_db_sa://wkg27603:***@125f9f61-9715-46f9-9399-c8177b21803b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:30426/bludb
Done.
```

landing__outcome	COUNT
Failure (drone ship)	3
No attempt	3
Success (drone ship)	3
Success (ground pad)	3
Controlled (ocean)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

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

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

[38]:
```
%%sql

SELECT Booster_Version
FROM spacex
WHERE landing__outcome = 'Success (drone ship)' AND payload_mass__kg_ > 4000 AND payload_mass__kg_ < 6000;
```

```
 * ibm_db_sa://wkg27603:***@125f9f61-9715-46f9-9399-c8177b21803b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:30426/bludb
Done.
```

[38]:

booster_version
F9 FT B1026
F9 FT B1021.2

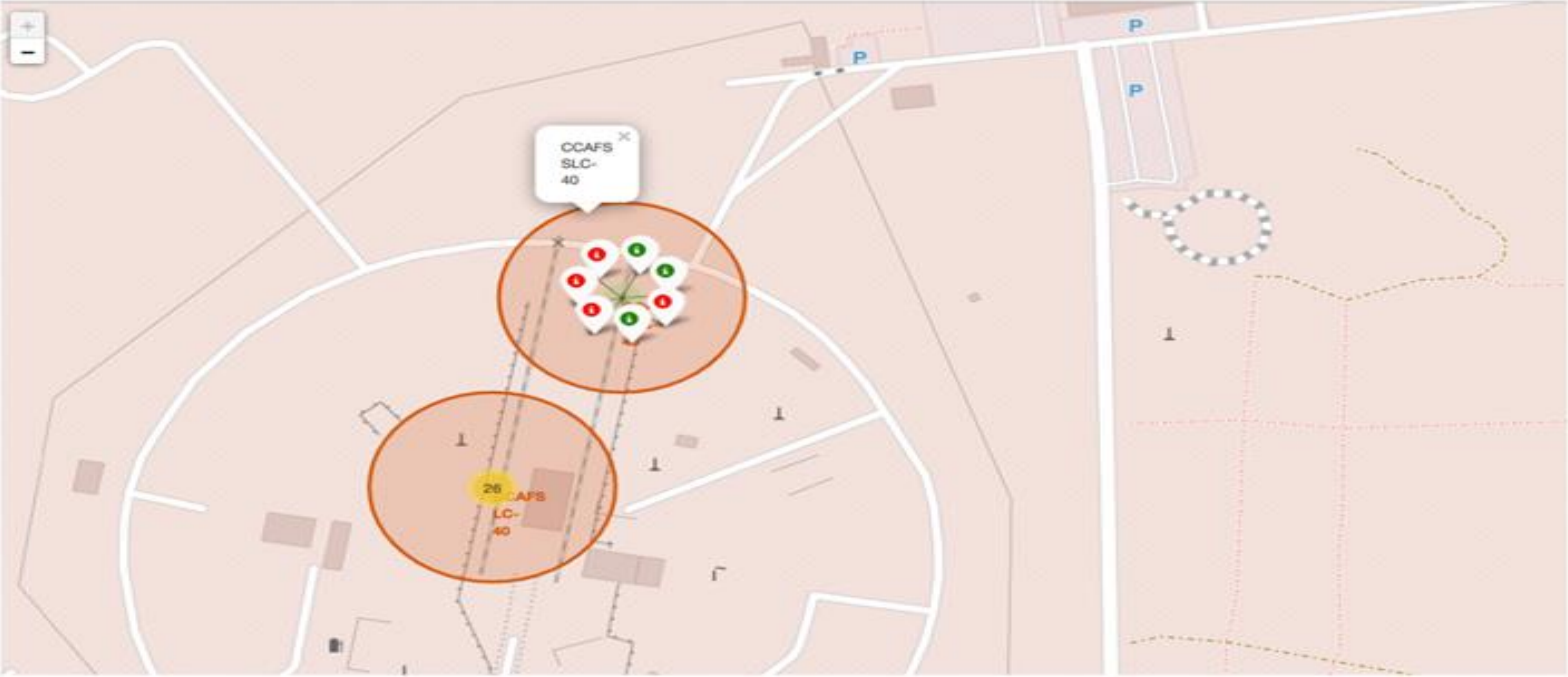
Booster versions F9 FT B1026 and F9 FT B1021.2 have more successful landings in drone ship with payload mass between 4000 kg and 6000 kg

Folium Map

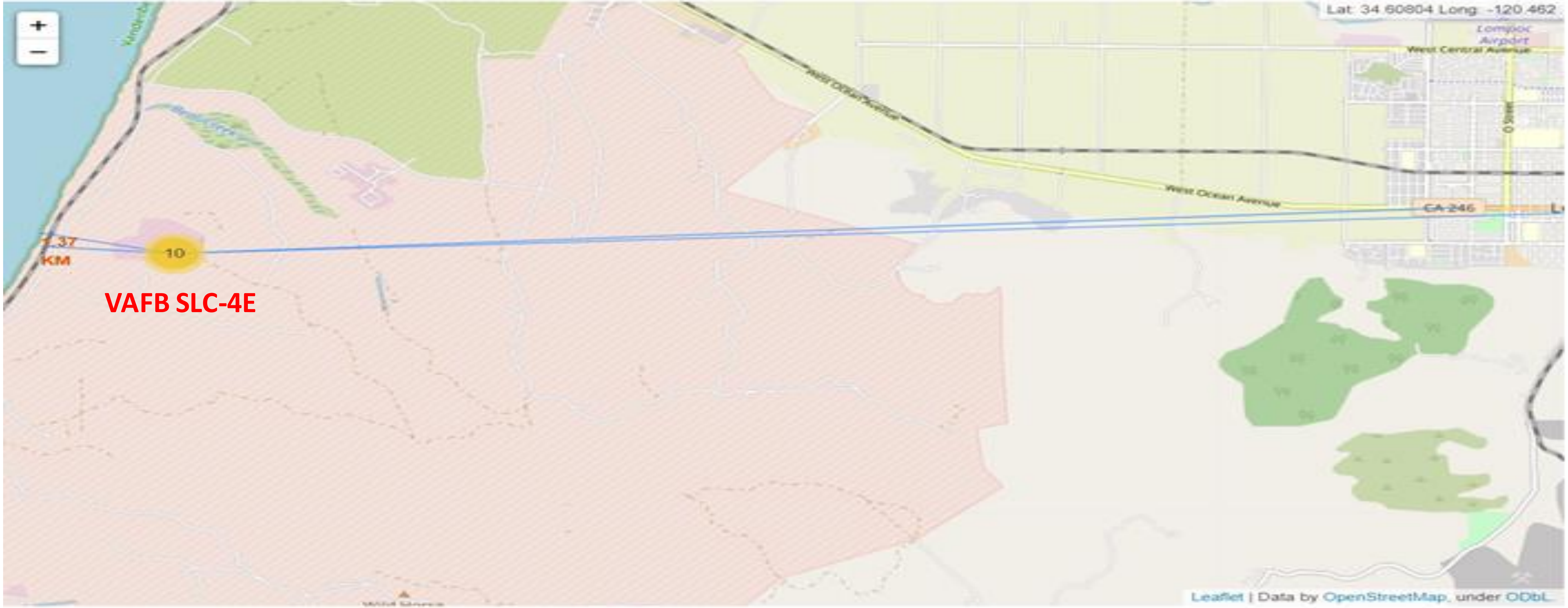
Map of all launch sites marked in red.
Most sites are near the equator and coast.

IBM Developer

SKILLS NETWORK



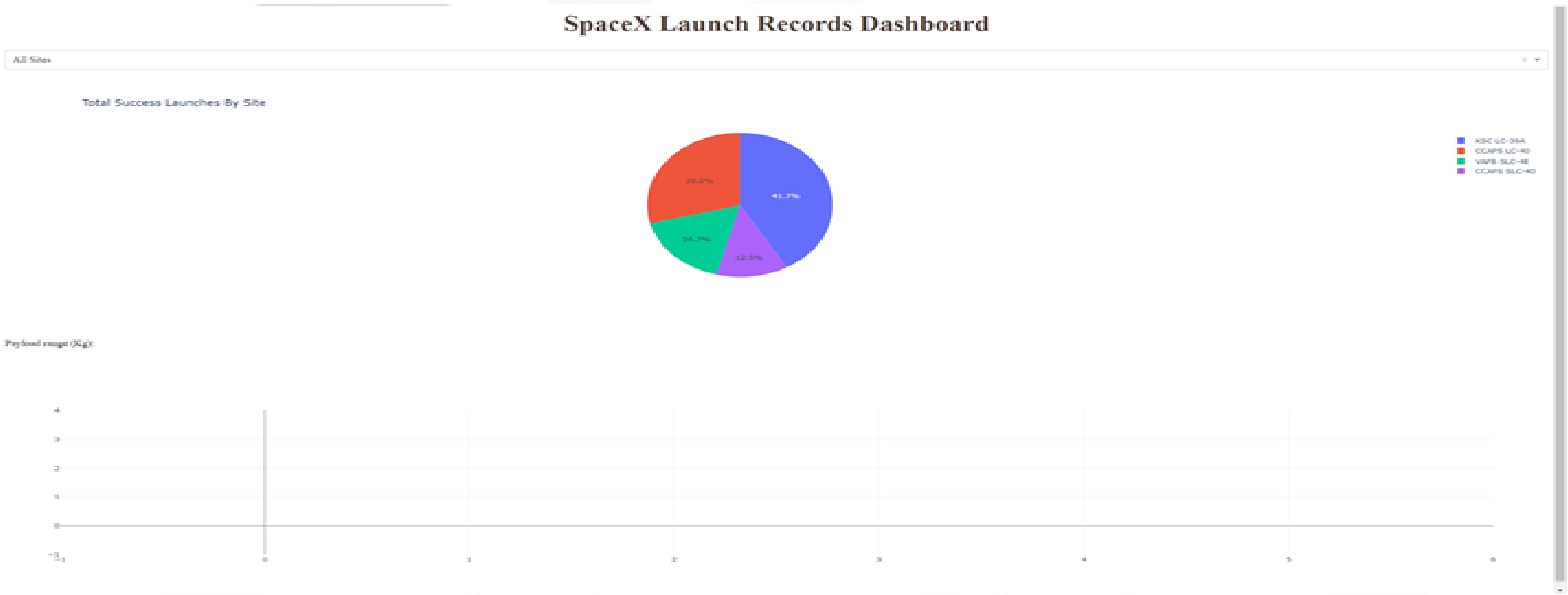
Launch sites with success/fail landing outcomes in green/red



Launch site is close to railway and coastline, but far away from highway and city

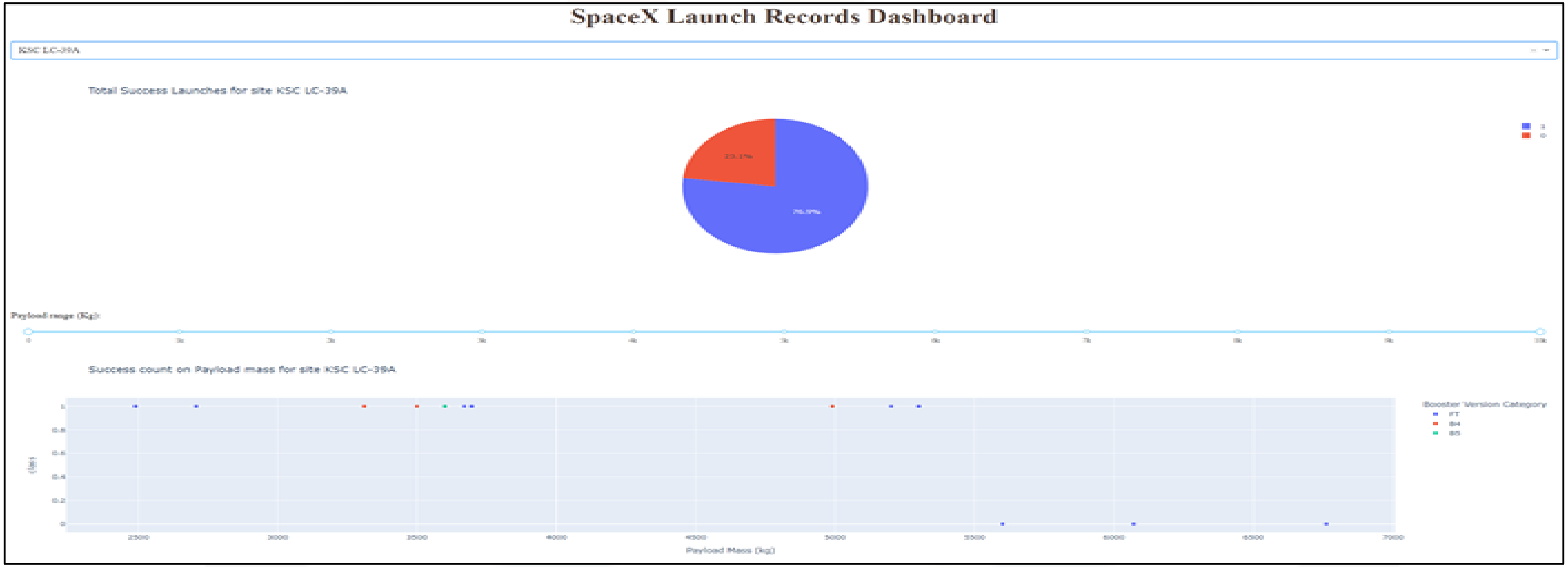
Plotly Dash

DASHBOARD TAB 1



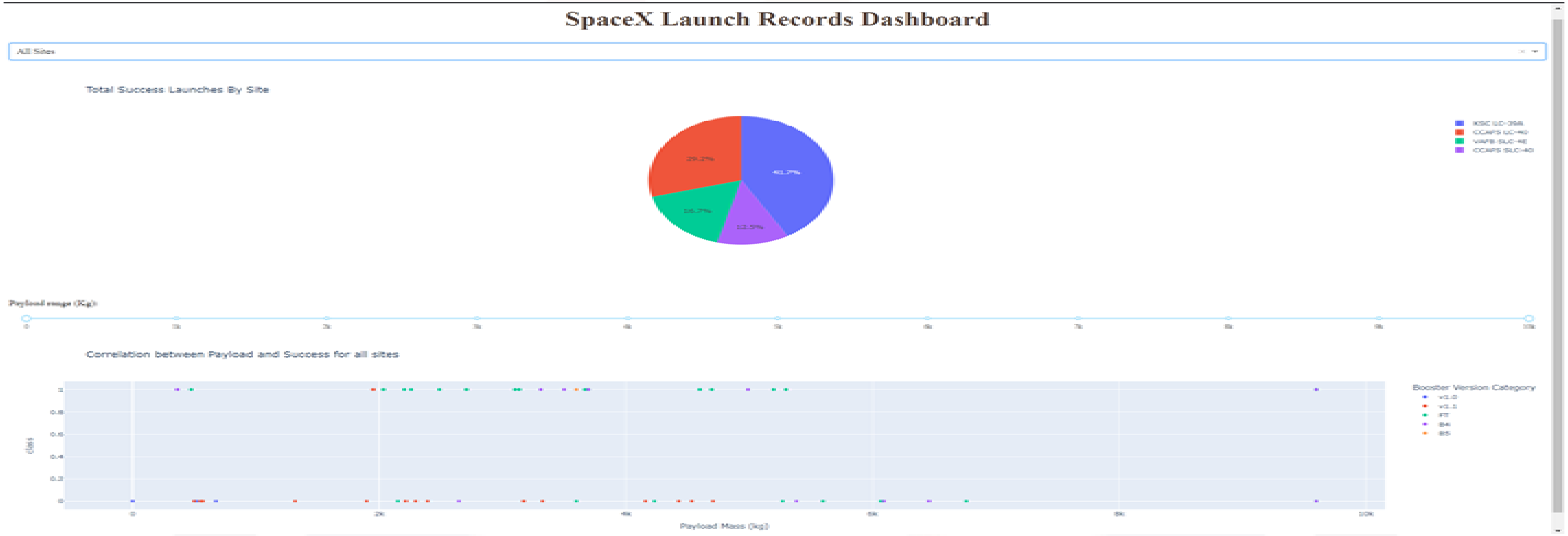
Pie chart with different success ratio of 4 launch sites

DASHBOARD TAB 2



Launch site KSC LC-39A has the highest launch success ratio

DASHBOARD TAB 3



Payload vs launch outcome scatter plot for all sites

Predictive Analysis

KNN model performance

Best parameters to use for KNN model

tuned hpyerparameters :(best parameters) {'algorithm': 'auto', 'n_neighbors': 10, 'p': 1}

accuracy : 0.848

Accuracy of the test data

knn_cv.score(X_test,Y_test) = 0.833

Logistic Regression Model Performance

Best parameters to use for Logistic Regression model

```
tuned hpyerparameters :(best parameters)  {'C':
0.01, 'penalty': 'l2', 'solver': 'lbfgs'}
accuracy : 0.846
```

Accuracy of the test data

logreg_cv.score(X_test,Y_test) = 0.833

Decision Tree Model Performance

Best parameters to use for Decision Tree model

tuned hpyerparameters :(best parameters) {'criterion': 'gini', 'max_depth': 4, 'max_features': 'auto', 'min_samples_leaf': 4, 'min_samples_split': 5, 'splitter': 'best'}

accuracy : 0.875

Accuracy of the test data

tree_cv.score(X_test,Y_test) = 0.944

Support Vector Machine Model Performance

Best parameters to use for Support Vector Machine model

tuned hpyerparameters :(best parameters) {'C': 1.0, 'gamma': 0.03162277660168379, 'kernel': 'sigmoid'}
accuracy : 0.848

Accuracy of the test data

svm_cv.score(X_test,Y_test) = 0.833

Model Evaluation Results

```
Logistics Regression Accuracy: 0.833
Support Vector Machine Accuracy: 0.833
Decision Tree Accuracy: 0.944
K Nearest Neighbors Accuracy: 0.833
```

Decision Tree performs the best with 0.944 accuracy score

Confusion Matrix of Best Model

There is a false positive, where the predicted label has 1 landed, but the true label has 1 not landed

DISCUSSION

After researching and analyzing the SpaceX dataset, It's found that launch site KSC LC-39A is where a rocket should be launched. It has the highest success landing rate, which will increase the chance of reusing the rocket to lower the cost of the launch to 62 million dollars.

CONCLUSION

- As the flight number increased, there were more successful landings
- Launch site KSC LC-39A had the highest success rate at 41.7%
- Payload mass between 2000kg to 6000kg had more successful landings
- Launch site KSC LC-39A had a success rate of 76.9% at payload mass between 2500kg to 5300kg
- Decision Tree model performs the best to predict landing outcomes


PERFECTING
PROPULSIVE
LANDING

APPENDIX

	FlightNumber	PayloadMass	Flights	Block	ReusedCount	Longitude	Latitude
count	94.000000	88.000000	94.000000	90.000000	94.000000	94.000000	94.000000
mean	54.202128	5919.165341	1.755319	3.500000	2.872340	-75.553302	28.581782
std	30.589048	4909.689575	1.197544	1.595288	3.793696	53.391880	4.639981
min	1.000000	20.000000	1.000000	1.000000	0.000000	-120.610829	9.047721
25%	28.250000	2406.250000	1.000000	2.000000	0.000000	-80.603956	28.561857
50%	52.500000	4414.000000	1.000000	4.000000	1.000000	-80.577366	28.561857
75%	81.500000	9543.750000	2.000000	5.000000	4.000000	-80.577366	28.608058
max	106.000000	15600.000000	6.000000	5.000000	11.000000	167.743129	34.632093

Statistics of data used in research

- https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches