IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Animal Lady
September 20, 2024

Outline

- Executive Summary
- Introduction
- Methodology
- Results
  - Section 2
  - Section 3
  - Section 4
  - Section 5
- Conclusion

Executive Summary

1. **Summary of methodologies**
   - Data collection
   - Data wrangling
   - Data exploration
   - Data visualization
   - Model building
   - Model evaluation

2. **Summary of Results**
   - The best model is the Decision Tree.
   - Total number of successful mission outcomes: 99 flights.
   - Total number of failed mission outcomes: 1 flight.
   - Most launch sites are located near the coastal areas.
   - ES-L1, GEO, and SSO had a 100% success rate.

Introduction

1. **Project background and context**

This assignment predicts if the Falcon 9 first stage will land successfully. 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.

2. **Problems**

- How payload mass, launch site, number of flights, and orbits affect first-stage landing success.
- Rate of successful launch.
- Best predictive models for successful launch.

Section 1

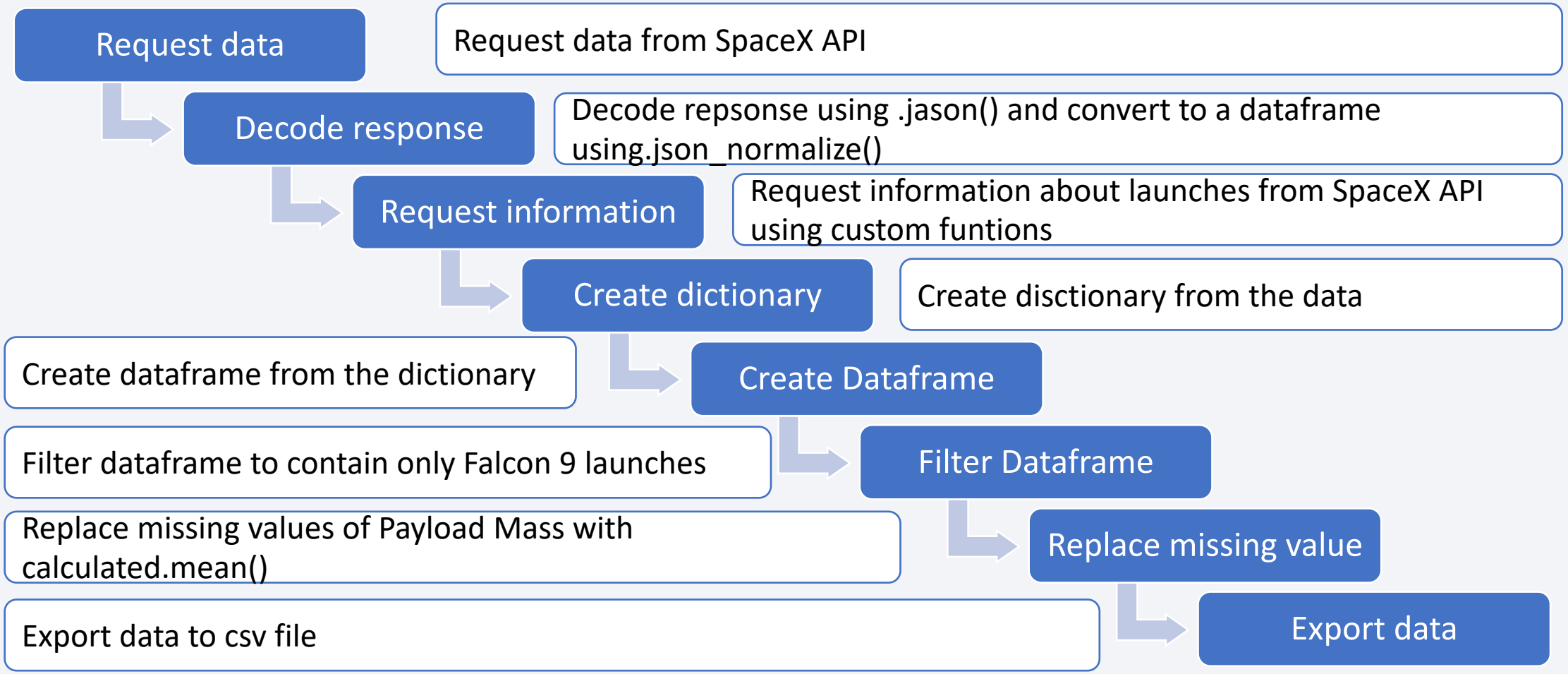
Methodology

Methodology

Executive Summary

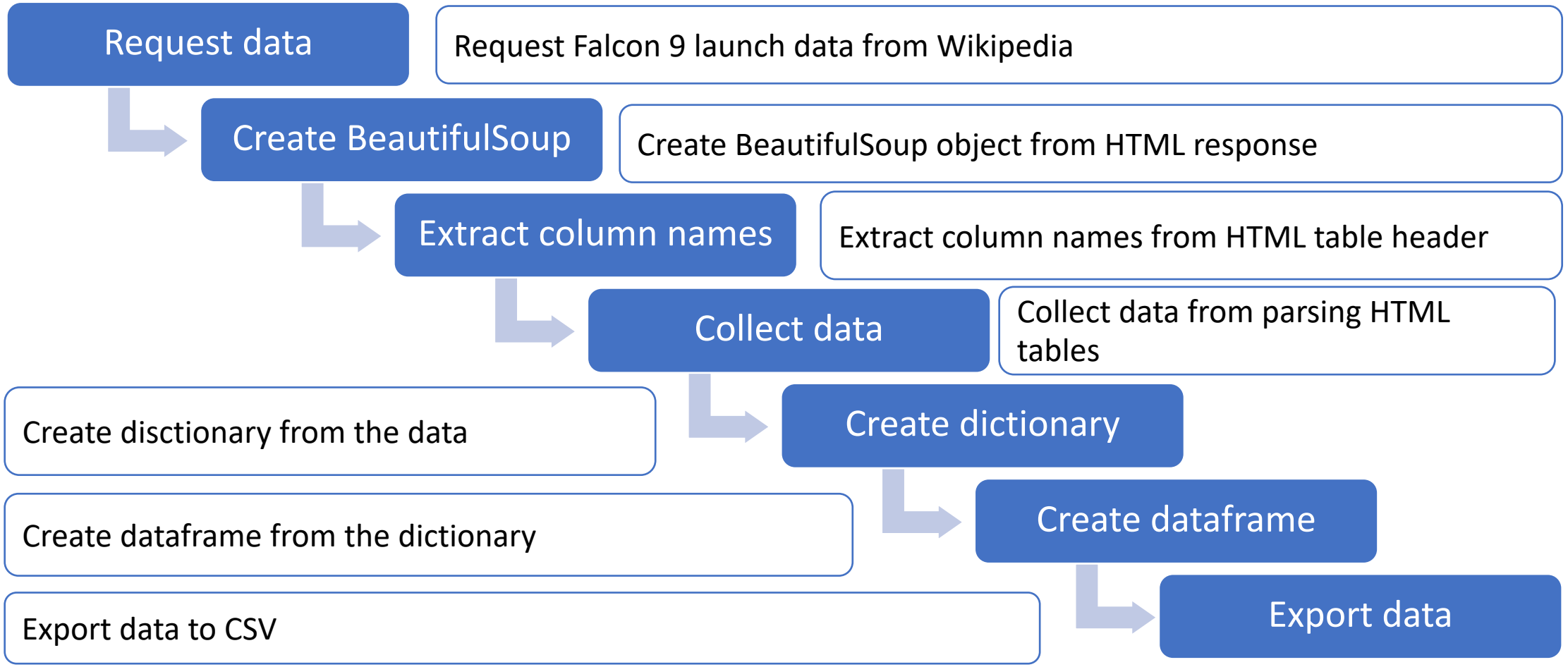
- Data collection methodology: This assignment used SpaceX Rest API data and web scraping with BeautifulSoup methods.
- Perform data wrangling: Data wrangling used API method, filtering data, handling missing values, and data sampling.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models: Models were built using GridSearchCV on different algorisms: logistic regression, support vector machine, decision tree, K-Nearest Neighbor. The predictive models were evaluate using .score() for all models. Confusion matrix, Jaccard, and F1_Score were employed to identify best models.

Data Collection – SpaceX API



GitHub URL: https://github.com/animallady/tesrepo/blob/main/Lab%201-API.ipynb

Data Collection – Web Scraping



GitHub URL: https://github.com/animallady/tesrepo/blob/main/Lab%202-Webscrapig.ipynb

Data Wrangling

```
df['Orbit'].value_counts()
```

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

- Import data and handle missing values
- Check data types
- Calculate the number of launches on each site
- Calculate the number and occurrence of each orbit
- Calculate the number and occurrence of each orbit
- Calculate the number and occurrence of mission outcome of the orbits
- Create outcome columns

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

```
Outcome
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: count, dtype: int64
```

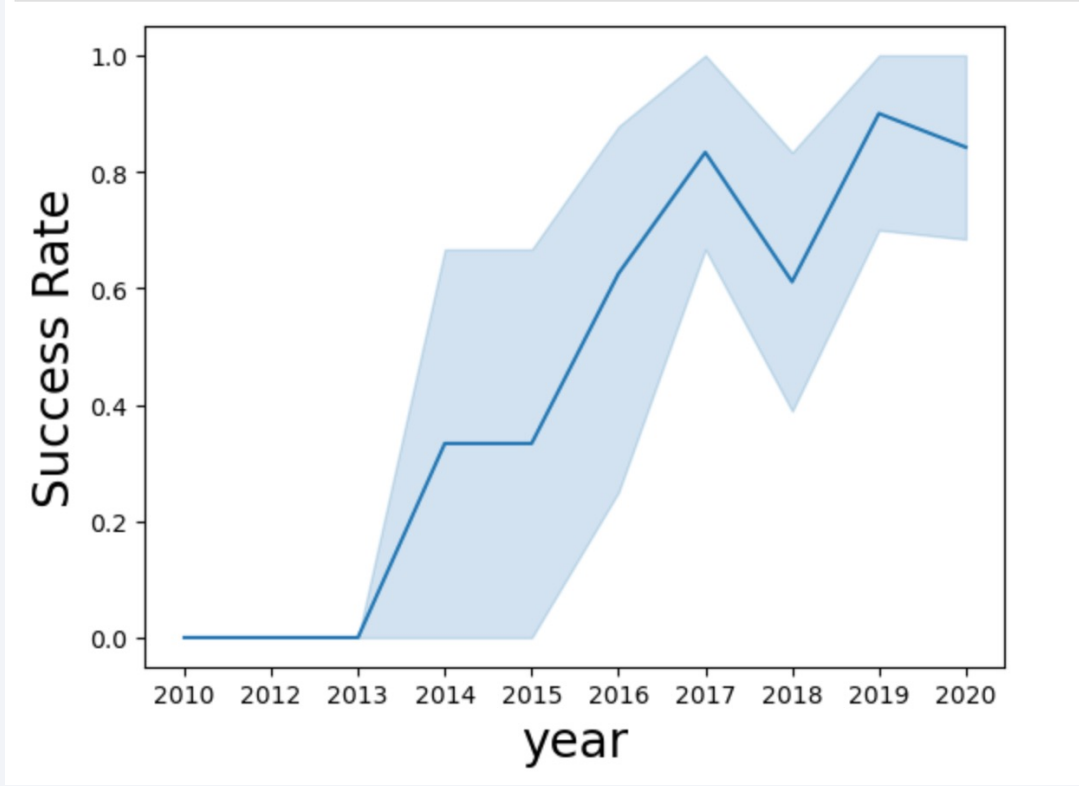
GitHub URL: https://github.com/animallady/tesrepo/blob/main/Lab%203-Data%20wranglig.ipynb

EDA with Data Visualization

Scatterplots, line plots, and bar charts were used to visualize the data.

- Scatterplots were used to visualize relationships between two or more variables
  - Flight Number vs. Payload Mass
  - Flight Number Vs. Launch Site
  - Launch Site vs. Payload Mass
  - Orbit vs. Flight Number
  - Orbit vs. Payload Mass
- A line plot was used to observe the success rate between 2010 and 2020.
- A bar chart was used to visualize the relationship between the success rate of each orbit type.

```
sns.lineplot(data=df, x="Date", y="Class")
plt.xlabel("year",fontsize=20)
plt.ylabel("Success Rate",fontsize=20)
plt.show()
```



GitHub URL: https://github.com/animallady/tesrepo/blob/main/Lab5-EDA%20with%20Visualization%20.ipynb

EDA with SQL

1. Display the names of the unique launch sites in the space mission
2. Display 5 records where launch sites begin with the string 'CCA'
3. Display the total payload mass carried by boosters launched by NASA (CRS)
4. Display average payload mass carried by booster version F9 v1.1
5. List the date when the first successful landing outcome in the ground pad was achieved
6. List the names of the boosters which have success in drone ships and have payload mass greater than 4000 but less than 6000
7. List the total number of successful and failed mission outcomes
8. List the names of the booster versions that have carried the maximum payload mass.
9. List the records that will display the month names, failed landing outcomes in drone ships, booster versions, and launch sites for the months of 2015.
10. Landing outcomes between 2010-06-04 and 2017-03-20

Landing_Outcome	count_outcomes
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

GitHub URL: https://github.com/animallady/tesrepo/blob/main/Lab%204-EDA%20with%20SQL.ipynb

Build an Interactive Map with Folium

- All launch sites were marked ‘Black.’
- Colors for Launch Outcomes
  - ‘Green’ means ‘successful .’
  - ‘Red’ means ‘Failure’
- Distance from proximities
  - ‘Red’ refers to the distance from the ‘nearest city .’
  - ‘Orange’ refers to the distance from the ‘nearest railway .’
  - ‘Green’ refers to the distance from the ‘nearest highway .’

GitHub URL: https://github.com/animallady/tesrepo/blob/main/Lab%206-Fulium.ipynb

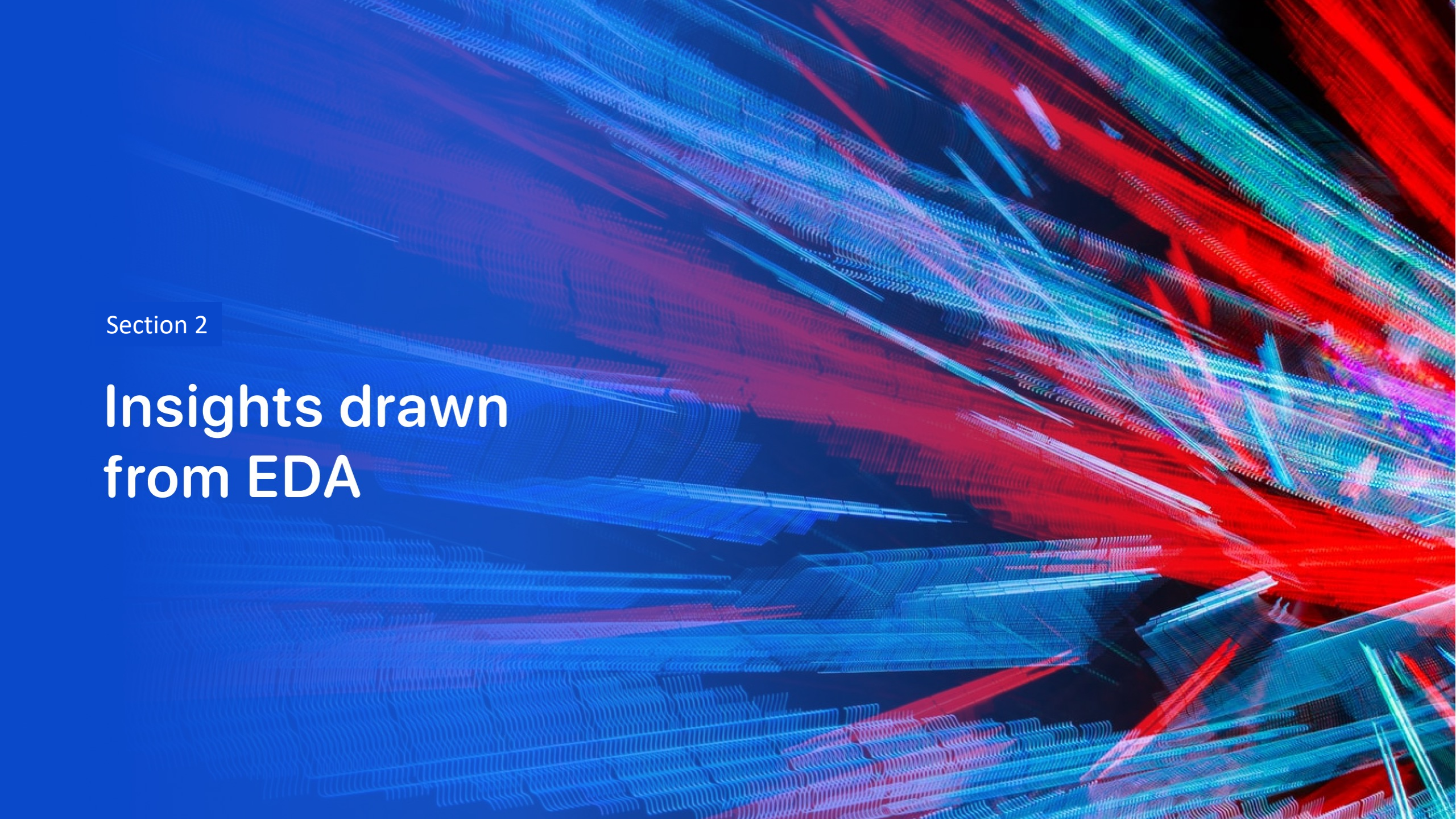
Build a Dashboard with Plotly Dash

- Create a dropdown list with launch sites that allow users to view all launch sites or select a specific site of their interest.
- Create a pie chart based on a dropdown list to show total success launches for each site.
- Add a range slider to select the payload mass.
- Create Scatterplots to show the relationship between the payload mass and launching success rate.

GitHub URL: https://github.com/animallady/tesrepo/blob/main/Lab-7-Dash%20Plotly.ipynb

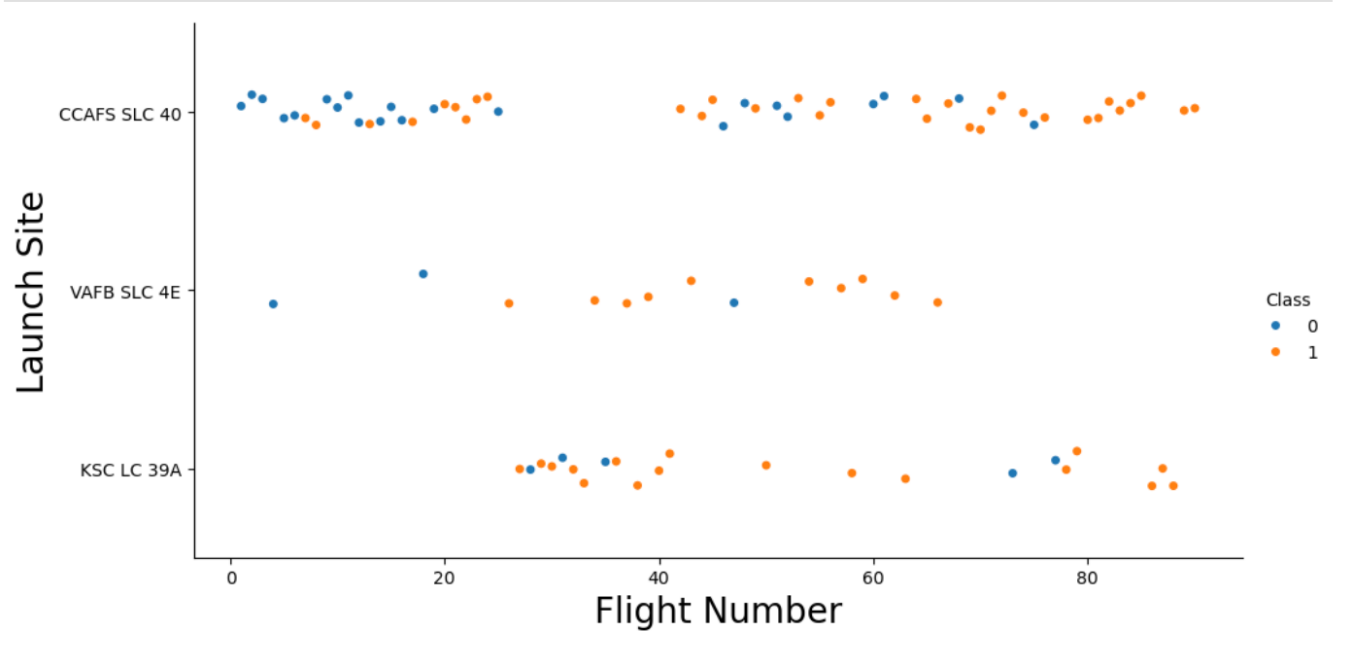
Predictive Analysis (Classification)

- Loading data
- Create NumPy array
- Standardize the data using the StandardSclaer function.
- Fit and transform data.
- Split the data using train_test_split
- Create a GridSearchCV object with cv = 10
- Apply GridSearch CV on:
  - Logistic Regression
  - Support Vector Machine
  - Decision Tree
  - K-nearest Neighbor
- Evaluate models by calculating accuracy, Jaccard_Score, and F1_Score.


Section 2
Insights drawn from EDA

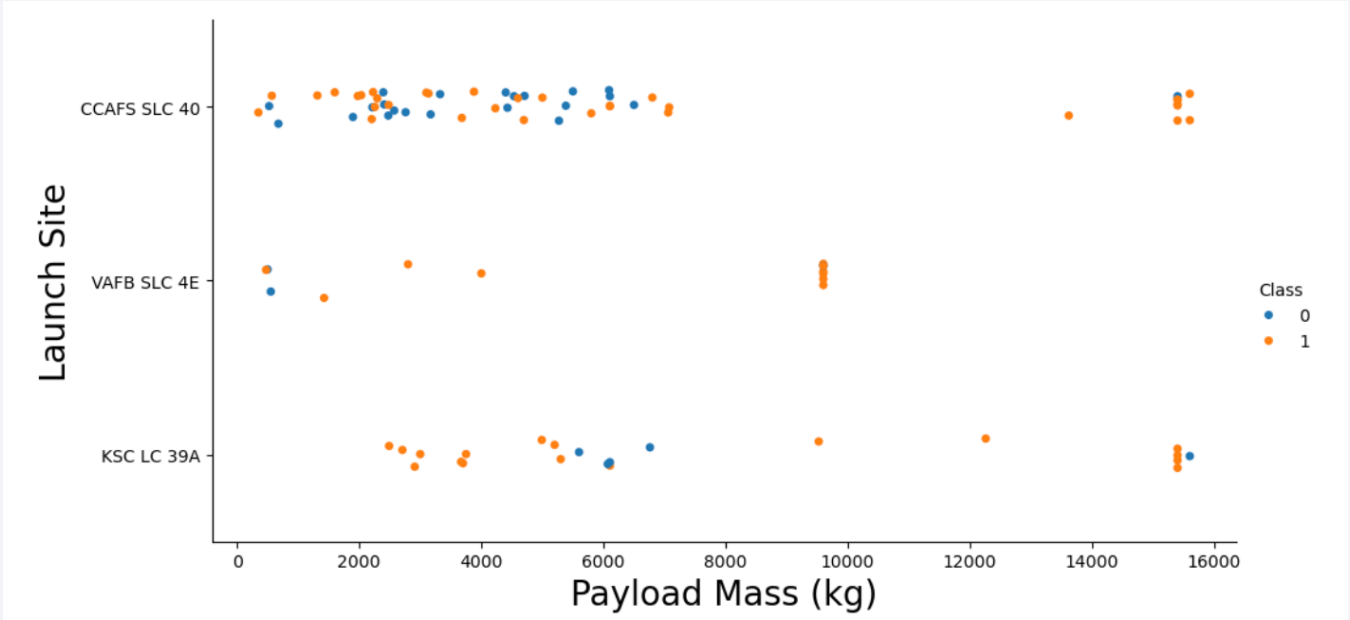
Flight Number vs. Launch Site

- Across all launch sites, earlier flights had lower success rates than later flights.
- VAFB SLC 4E and KSC LC 39A have launched fewer flights than CCAFS SLC 40 and have higher success rates.



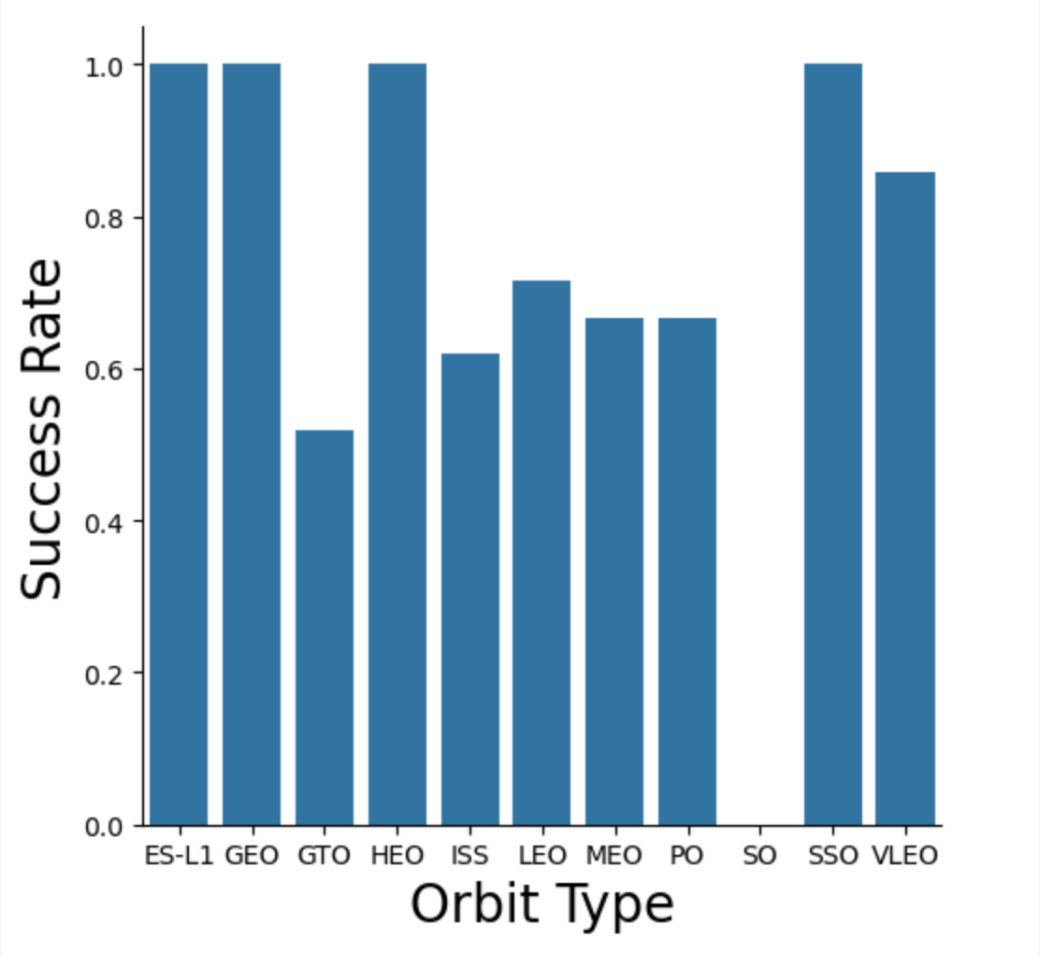
Payload vs. Launch Site

- Most flights across all launch sites had a payload mass of less than 8,000 Kg.
- Across all launch sites, payload mass higher than 8,000 Kg had very high success rates.
- VAFB SLC 4E had launched flights with payload masses of less than 10,000 Kg, yet they had a high success rate.
- Most flights from CCAFS SLC 40 had payload masses of less than 8,000 Kg and had a low success rate.
- All flights from KSC LC 39 A, which had payload masses lower than 5,500 Kg, had an extremely high success rate.



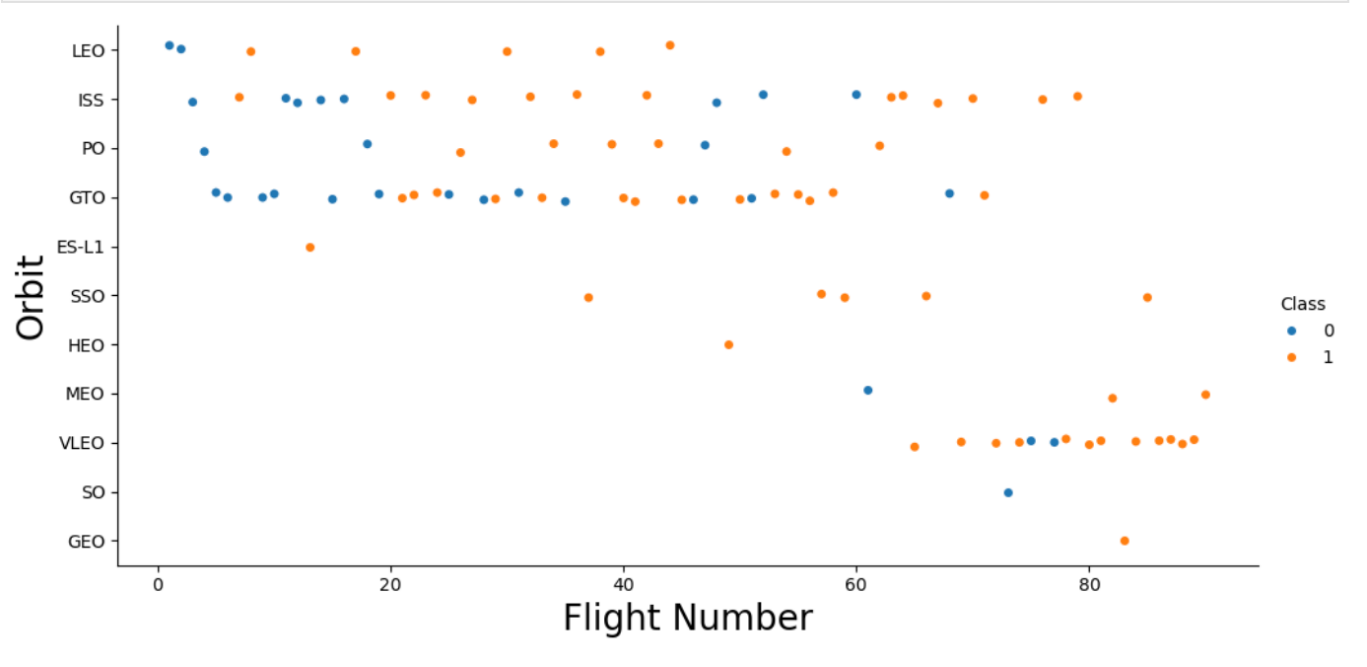
Success Rate vs. Orbit Type

- ES-L1, GEO, and SSO had a 100% success rate.
- ISS, LEO, MEO, and PO had between 60% and 70%.
- GTO had a success rate of around 50%.
- No flight was successful in SO Orbit.



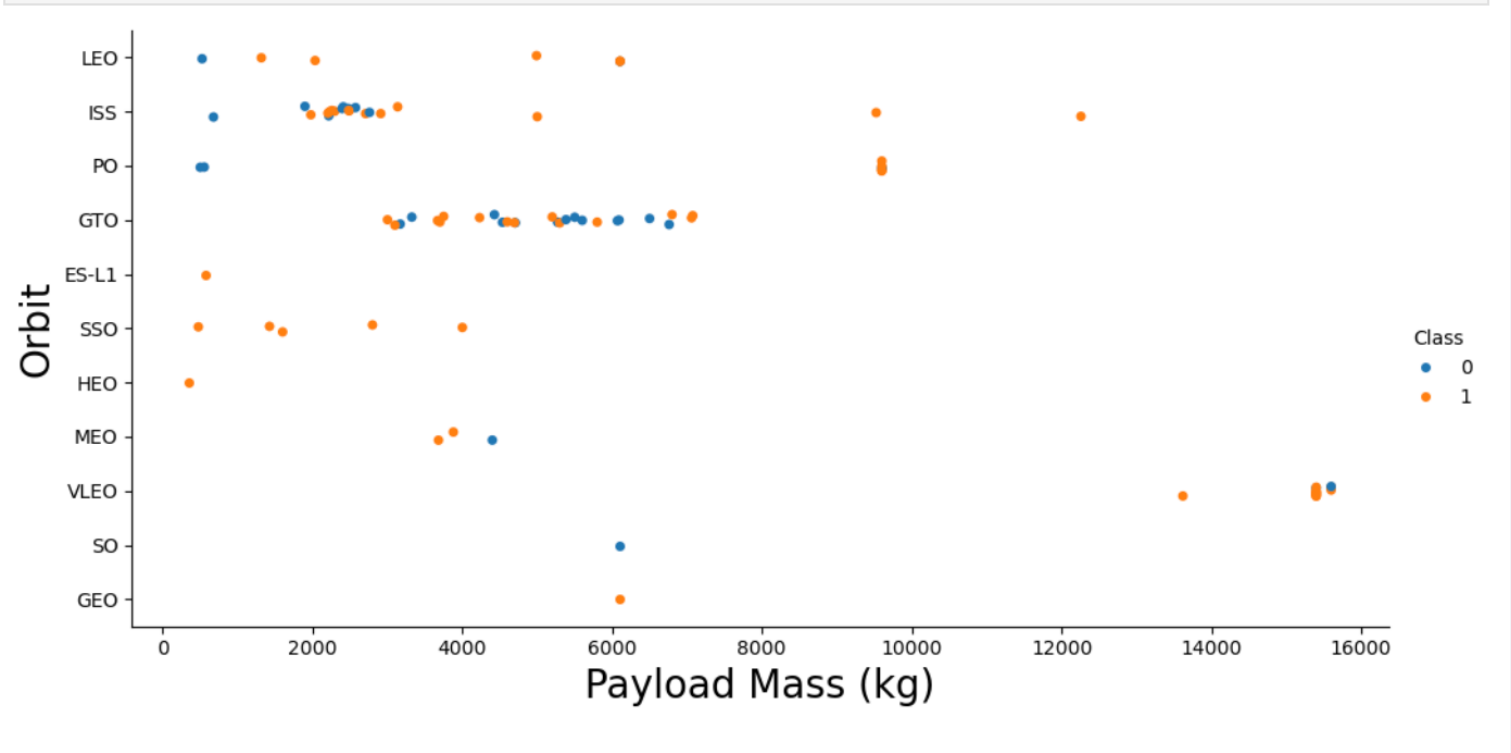
Flight Number vs. Orbit Type

- In most orbits, the higher the number of flights, the higher the success rate.
- The GTO orbit displayed a different trend from the rest of the orbits.



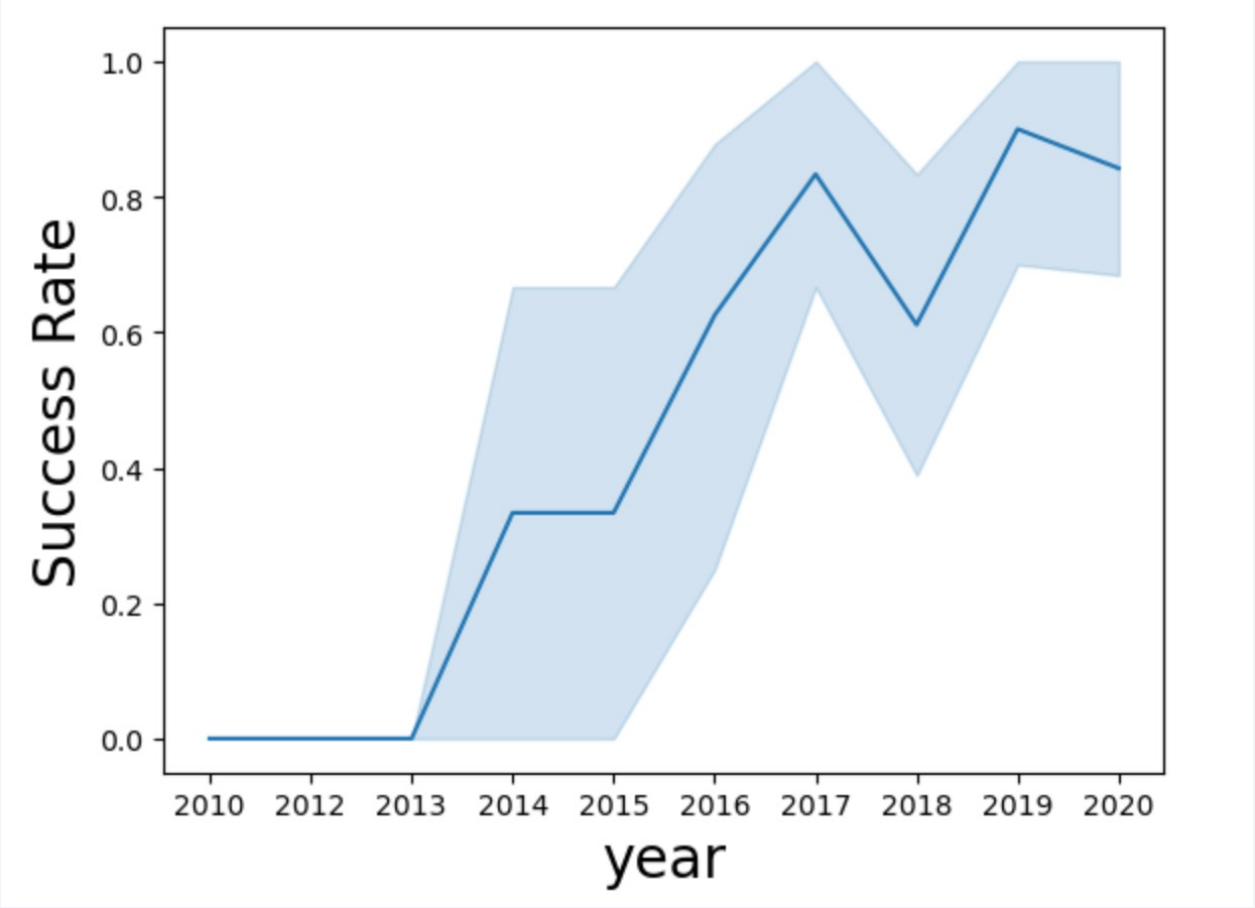
Payload vs. Orbit Type

- Flights with payload masses less than 4,000 Kg were successfully launched into ES-L1, SSO, HEO, and MEO orbits.
- Flights with payload masses greater than 10,000 Kg had a greater success rate in ISS, PO, and VLEO orbits.
- Flights with payload masses between 4,000 Kg and 8,000 Kg were launched into GTO orbit and displayed mixed results.



Launch Success Yearly Trend

- Overall, the success rate has improved since 2013.
- There were no successful launches before 2013.
- The success rate reached its peak in 2019 between the period of 2010 and 2020.



All Launch Site Names

There were three launch sites:

- CCAFS SLC 40 has launched 55 flights.
- KSC LC 39 A has launched 22 flights.
- VAFB SLC 4E has launched 13 flights.

```
df['LaunchSite'].value_counts()
```

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

Launch Site Names Begin with 'CCA'

- All five flights launched at CCAFS LC 40 entered LEO orbit.
- All five flights were successfully launched.
- Four flights were contracted by NASA.
- One flight was launched by SpaceX.

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

Total Payload Mass

The total payload mass carried by boosters launched by NASA was 45,596 Kg.

```
%sql SELECT SUM(PAYLOAD_MASS__KG_) \
    FROM SPACEXTBL \
    WHERE Customer = 'NASA (CRS)';
```

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

SUM(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

The average payload mass carried by booster version F9 v1.1 was 2,928.4 Kg.

```
%sql SELECT AVG(PAYLOAD_MASS__KG_) \
    FROM SPACEXTBL \
    WHERE Booster_Version = 'F9 v1.1';
```

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

AVG(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

The first successful ground landing was on December 22, 2015.

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

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

MIN(DATE)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

List the names of boosters that have successfully landed on drone ships and had payload mass greater than 4000 but less than 6000:

- JCSAT-14
- JCSAT-16
- SES-10
- SES-11/EchoStar 105

```
%sql SELECT PAYLOAD \
FROM SPACEXTBL \
WHERE Landing_Outcome = 'Success (drone ship)' \
AND PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000;
```

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

Payload
JCSAT-14
JCSAT-16
SES-10
SES-11 / EchoStar 105

Total Number of Successful and Failure Mission Outcomes

- Total number of successful mission outcomes: 99 flights
- Total number of failed mission outcomes: 1 flight.
- One flight was successfully launched, yet the payload mass was unclear.

```
%sql SELECT Mission_Outcome, COUNT(*) as total_number \
FROM SPACEXTBL \
GROUP BY Mission_Outcome;
```

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

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

Boosters Carried Maximum Payload

- 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

```
%sql SELECT 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

List the failed landing outcomes in drone ships, their booster versions, and launch site names for the year 2015:

- There were two flights that failed to land in drone ships in 2015.
- Both flights were launched at CCAFS LC 40.

```
%sql SELECT substr(Date,6,2) as month, Date,Booster_Version, Launch_Site, [Landing_Outcome] \
FROM SPACEXTBL \
where [Landing_Outcome] = 'Failure (drone ship)' and substr(Date,0,5)='2015';
```

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

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

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

Rank the count of landing outcomes between 2010-06-04 and 2017-03-20:

- No attempt: 10 flights
- Failure: 7 flights
- Success: 8 flights

```
%sql SELECT [Landing_Outcome], count(*) as count_outcomes \
FROM SPACEXTBL \
WHERE DATE between '2010-06-04' and '2017-03-20' group by [Landing_Outcome] order by count_outcomes DESC;
```

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

Landing_Outcome	count_outcomes
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


Section 3
Launch Sites
Proximities Analysis

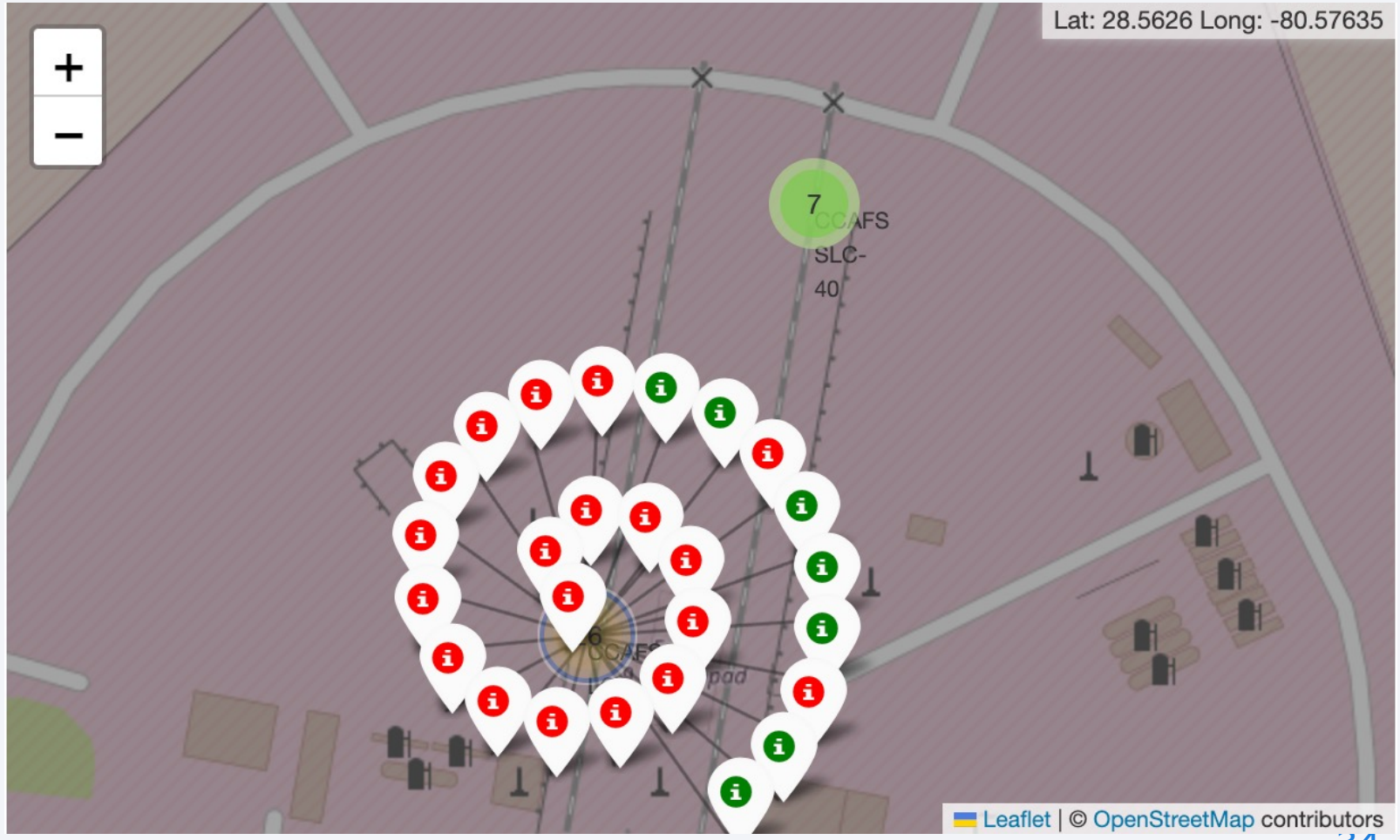
Launch Sites

Launch Sites are located in the coastal areas of the USA.

- VAFB SLC 4E is on the West Coast (California).
- KSC LC-39A, CCAFS SLC-40, and CCAFS LC-40 are located in Florida.

Launch Outcomes

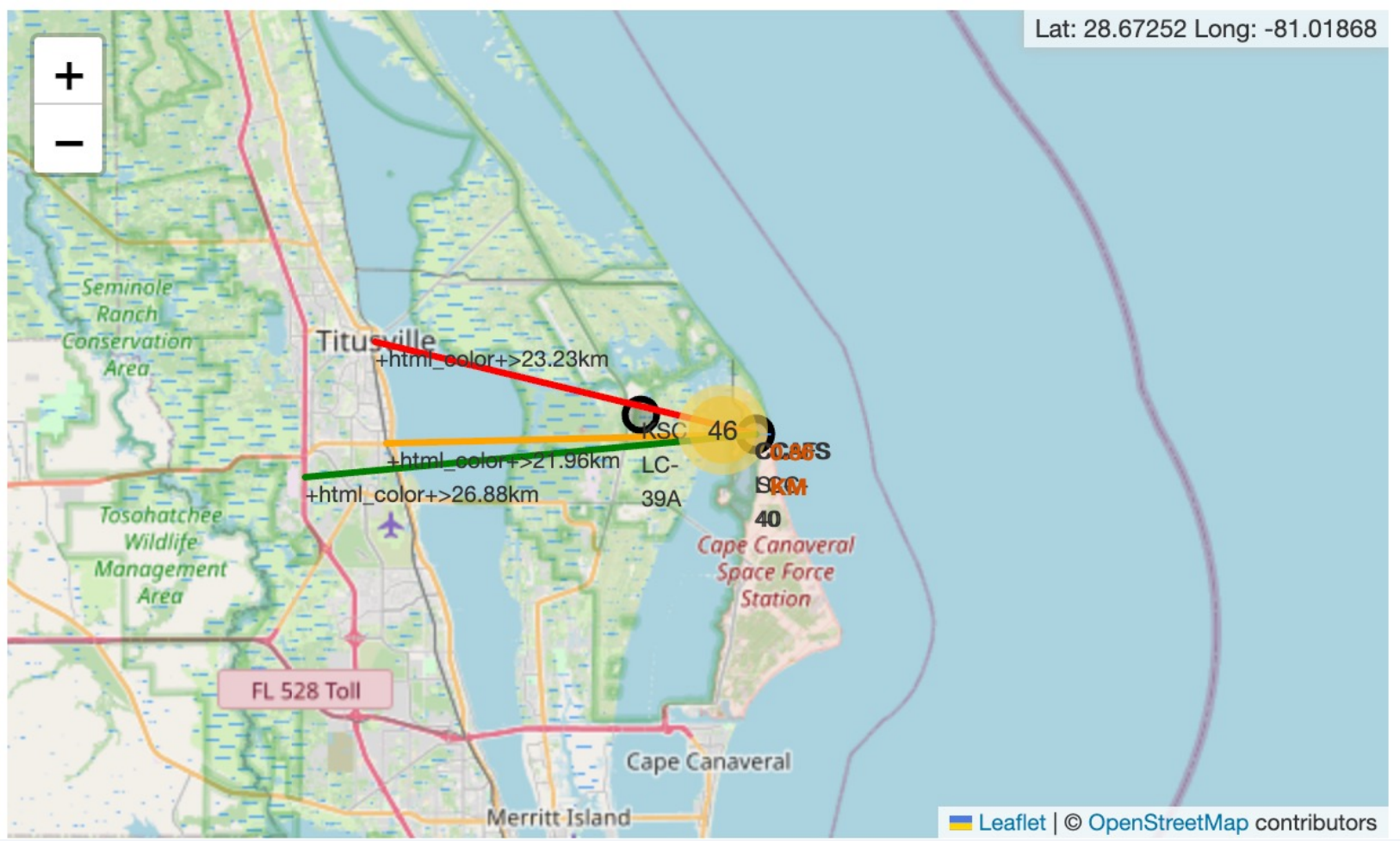
- ‘Green’ marks a successful launch.
- ‘Red’ marks a failed launch.

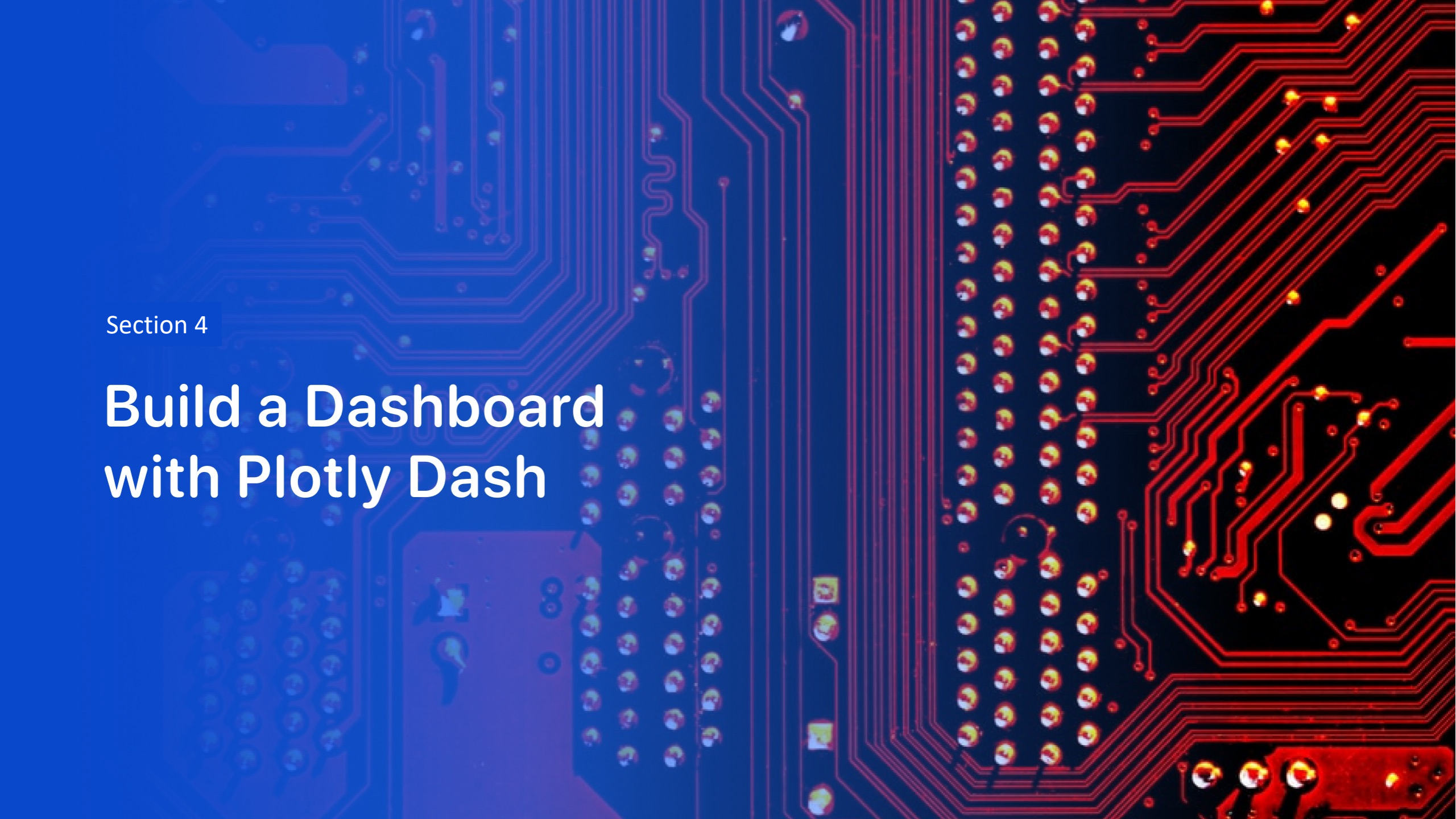


Distance Lines to Prosimities

CCAFS LC-40

- 23.23 Km from Titusville
- 21.96 Km from the Railway (nearest).
- 26.88 Km from the Highway (nearest).

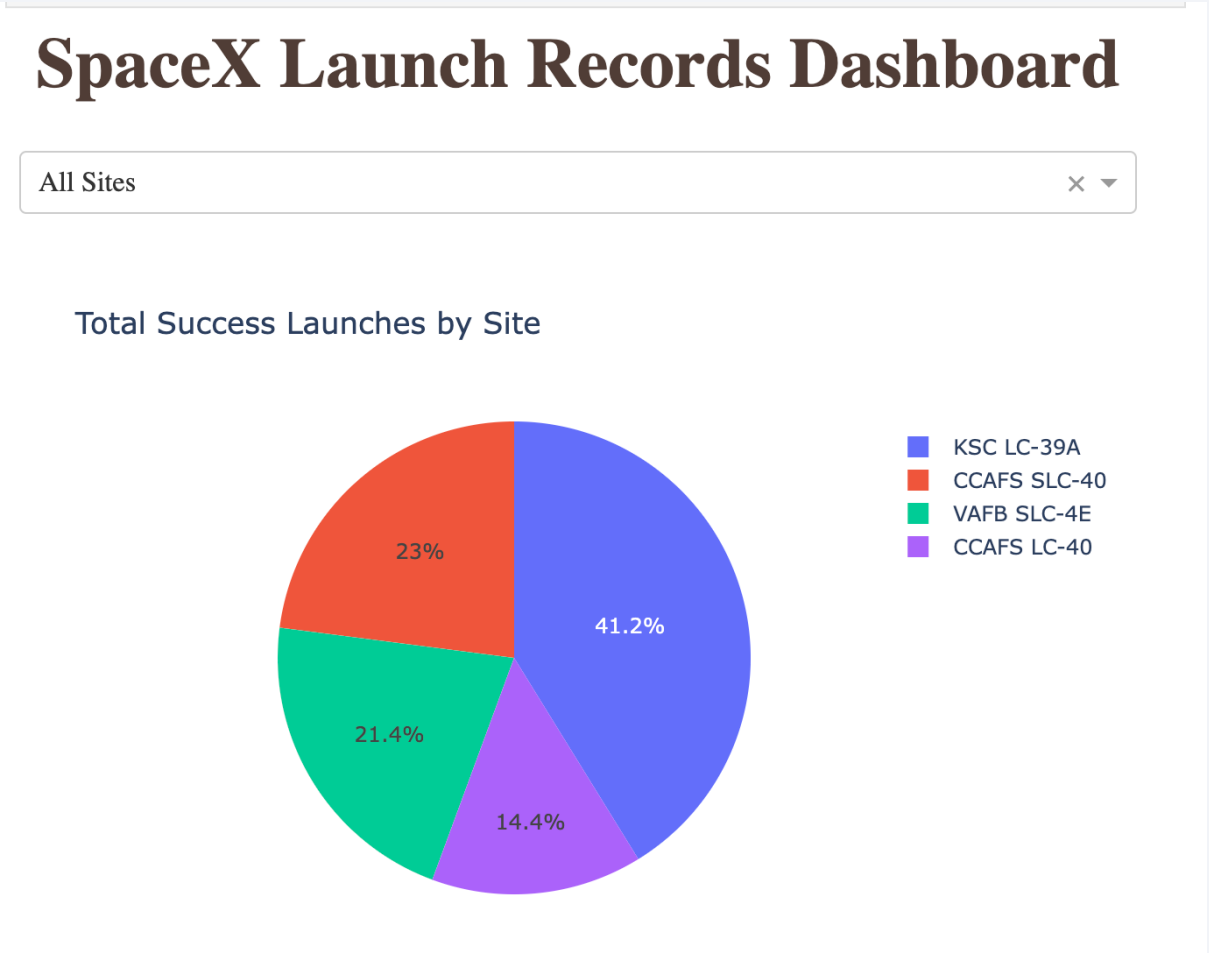



Section 4
Build a Dashboard with Plotly Dash

Total Success Launches by Site

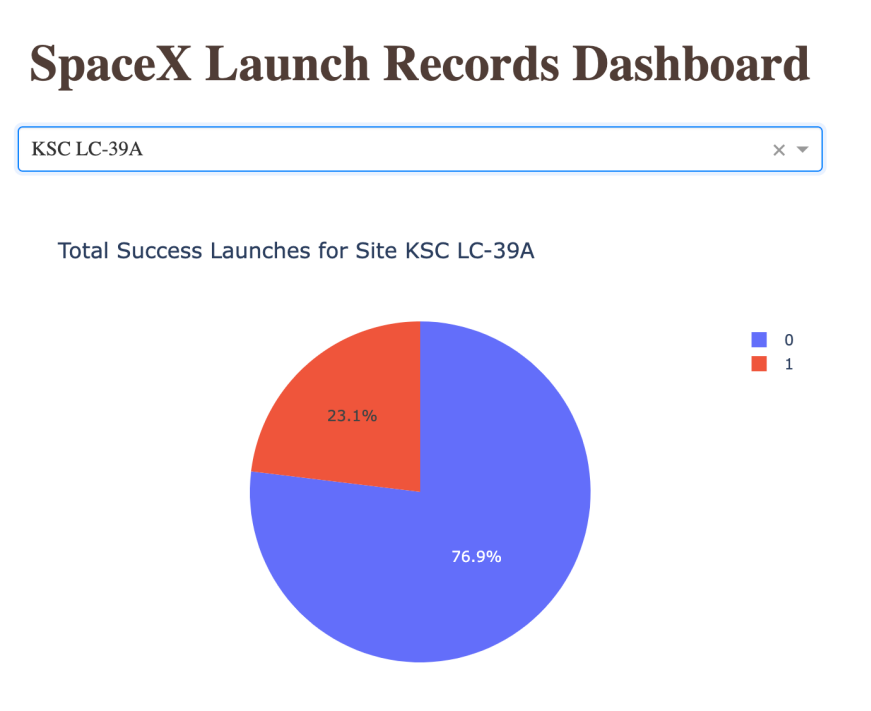
There are four launch sites, accounted for 100% of success rate in this pie chart:

- KSC LC-39A shared a 41.2% success rate.
- VAFB SLC-4E shared a 21.4% success rate.
- CCAFS SLC-40 shared a 21% success rate.
- CCAFS LC-40 shared a 14.4% success rate.

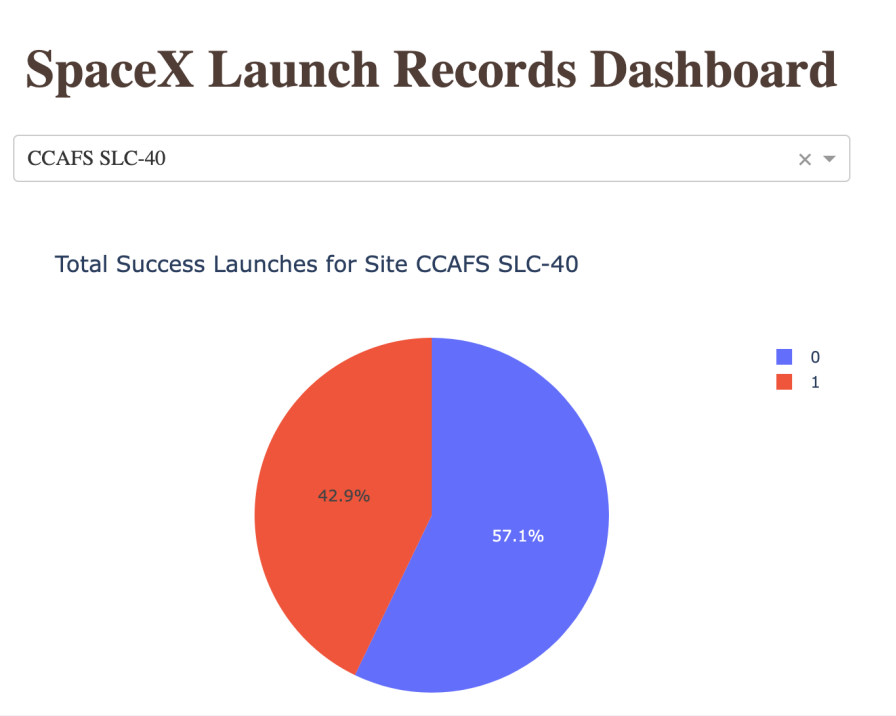


Total Success Launches for Site KSC LC-39A

- KSC LC-39A shared the most significant percentage of success rate among the four sites.
- However, only 23.1% of flights from KSC LC-39A were successful.



- CCAFS SLC-40 shared the lowest percentage of success rate among the four sites.
- However, 42.9% of flights from CCAFS SLC-40 were successful.



Section 5

Predictive Analysis (Classification)

Classification Accuracy

	LogReg	SVM	Tree	KNN
Jaccard_Score	0.800000	0.800000	0.769231	0.800000
F1_Score	0.888889	0.888889	0.869565	0.888889
Accuracy	0.833333	0.833333	0.833333	0.833333

```
models = {'KNeighbors':knn_cv.best_score_,
              'DecisionTree':tree_cv.best_score_,
              'LogisticRegression':logreg_cv.best_score_,
              'SupportVector': svm_cv.best_score_}

bestalgorithm = max(models, key=models.get)
print('Best model is', bestalgorithm,'with a score of', models[bestalgorithm])
if bestalgorithm == 'DecisionTree':
    print('Best params is :', tree_cv.best_params_)
if bestalgorithm == 'KNeighbors':
    print('Best params is :', knn_cv.best_params_)
if bestalgorithm == 'LogisticRegression':
    print('Best params is :', logreg_cv.best_params_)
if bestalgorithm == 'SupportVector':
    print('Best params is :', svm_cv.best_params_)
```

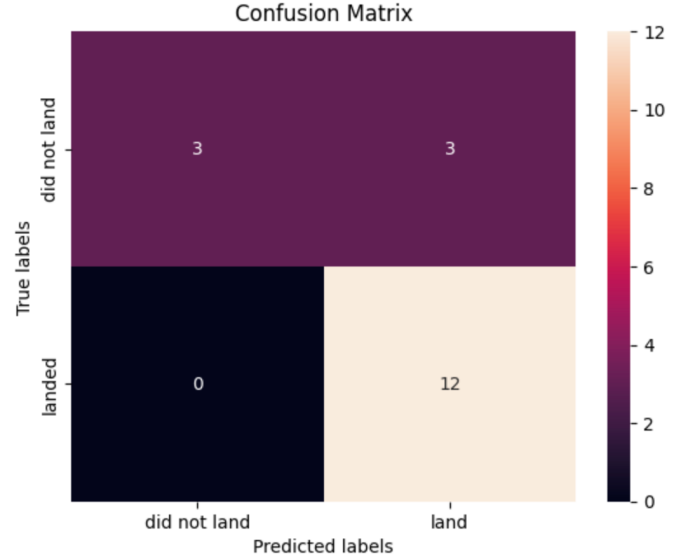
```
Best model is DecisionTree with a score of 0.8732142857142856
Best params is : {'criterion': 'entropy', 'max_depth': 6, 'max_features': 'sqr
t', 'min_samples_leaf': 2, 'min_samples_split': 2, 'splitter': 'random'}
```

- All models have the same accuracy score (0.833).
- Jaccard and F1 scores for LogReg, SVM, and KNN were slightly higher than the Decision Tree's scores.
- However, the Decision Tree outperformed other models when computing to identify the best model.

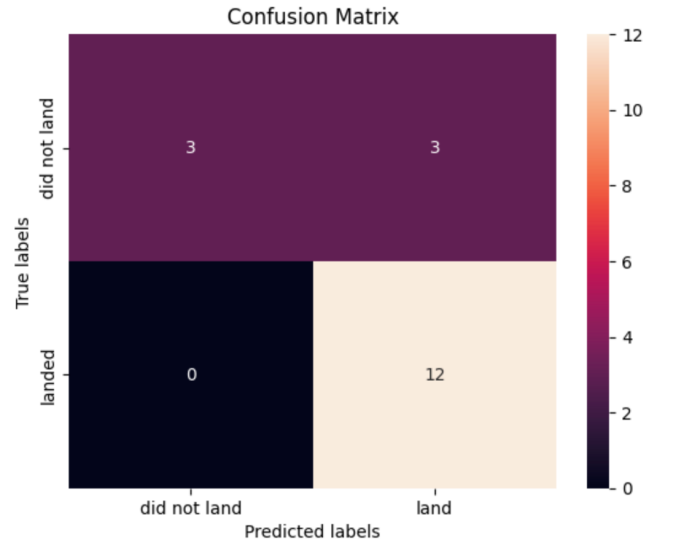
Confusion Matrix

- The outputs of confusion matrixes of SVM, Decision Tree, and K-nearest Neighbor were identical: 12 True Positive, 3 True Negative, and 3 False Positive.

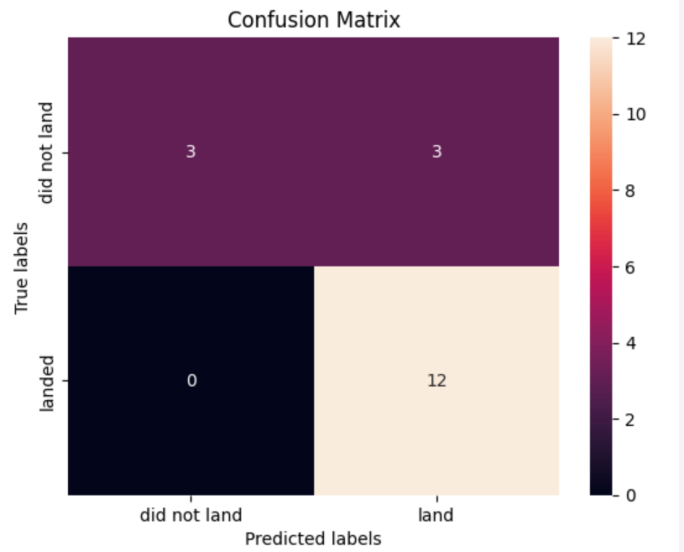
```
svm_yhat=svm_cv.predict(X_test)
plot_confusion_matrix(Y_test,yhat)
```




```
tree_yhat = tree_cv.predict(X_test)
plot_confusion_matrix(Y_test,yhat)
```




```
knn_yhat = knn_cv.predict(X_test)
plot_confusion_matrix(Y_test,yhat)
```



Conclusions

- The success rate of launching has increased since 2013.
- ES-L1, GEO, and SSO had a 100% success rate.
- All launch sites are located near the coastal areas.
- KSC LC-39A shared the most significant percentage of success rate among the four sites (41.2%).
- Logistic Regression, Support Vector Machines, Decision Tree, and K-nearest Neighbor generated the same accuracy scores. However, the Decision Tree outperformed other models when computing to define the best model.

Thank you!