IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Brian Croom
23 Jan 2025

Outline

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

Executive Summary

- Summary of methodologies
- Summary of all results

Introduction

- Project background and context
- Problems you want to find answers

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was scraped from SpaceX’s Wikipedia
- Perform data wrangling
  - Data was processed using pandas and BeautifulSoup
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - k nearest neighbors, decision trees, and logistic regression

Data Collection

Data was collected using the Python pandas and BeautifulSoup libraries. The

Data itself was scraped from SpaceX’s Wikipedia page containing information for each launch.

Data Collection – SpaceX API

- Collected data via REST calls to retrieve SpaceX data into pandas DataFrame

- **GitHub:** https://github.com/brianpccroom/datasciencecapstone/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

Data Collection - Scraping

- Requests library was used to retrieve the data, while BeautifulSoup was used to parse the HTML data into a pandas DataFrame.

- https://github.com/brianpccroom/datasciencecapstone/blob/main/jupyter-labs-webscraping.ipynb

Data Wrangling

The data wrangling process involved replacing missing numerical values with the mean average of the column, replacing NaN values with 0, and filtering out Falcon 1 launches to ensure only the most relevant data is considered.

See the following GitHub files for examples on the data collection/scraping/wrangling process:

https://github.com/brianpccroom/datasciencecapstone/blob/main/jupyter-labs-webscraping.ipynb

https://github.com/brianpccroom/datasciencecapstone/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

EDA with Data Visualization

Scatter plots were mostly used in this project since most of the comparisons involved numerical data against categorical data. For example, many of the analyses considered relationships between payload mass and launch site as well as flight number and launch site.

See the following notebook for EDA examples:

https://github.com/brianpccroom/datasciencecapstone/blob/main/edadataviz.ipynb

EDA with SQL

- Selecting distinct launch sites in the space mission.
- Listing when the first succesful landing outcome in ground pad was acheived.
- Listing the total number of successful and failure mission outcomes.

https://github.com/brianpccroom/datasciencecapstone/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

We used circles, text markers, and lines when analyzing the data with Folium.

These specific markers were used to facilitate the ease of reading the map as well as to draw our attention to important areas and distances from amenities such as railroads, roads, and coastlines.

https://github.com/brianpccroom/datasciencecapstone/blob/main/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

We used pie charts and slider charts to our Dashboards by using plotly. These charts were configured via callbacks and dropdown menus to facilitate interactivity.

The pie charts were immensely useful in visibly conveying percentages of launch successes for each launch site.

https://github.com/brianpccroom/datasciencecapstone/blob/main/plotly_dash.py

Predictive Analysis (Classification)

The best model was found by comparing the accuracy score of each model.

The development process involved creating a GridSearchCV object for each method and then checking its score with the score method

https://github.com/brianpccroom/datasciencecapstone/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Results

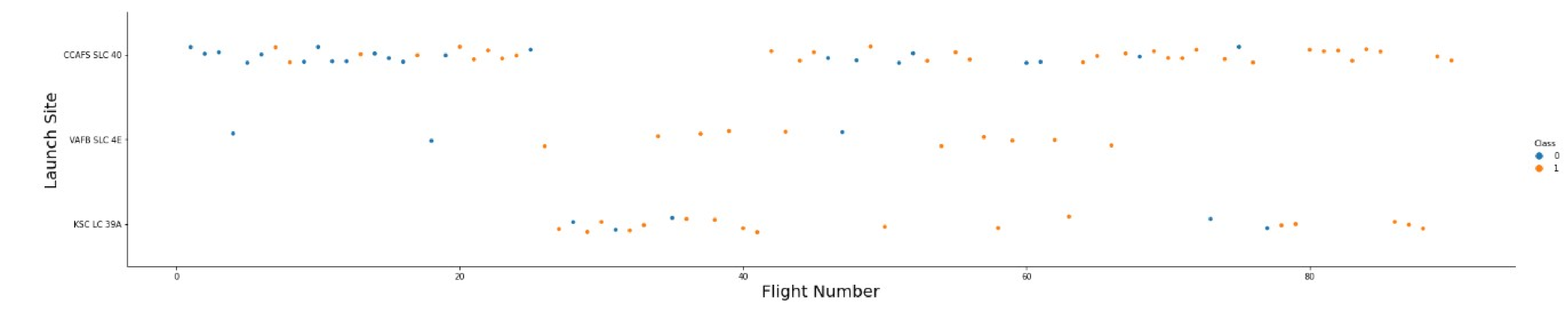
- All machine learning algorithms performed about the same

	ML Method	Accuracy Score (%)
0	Support Vector Machine	83.333333
1	Logistic Regression	83.333333
2	K Nearest Neighbour	83.333333
3	Decision Tree	83.333333

Section 2

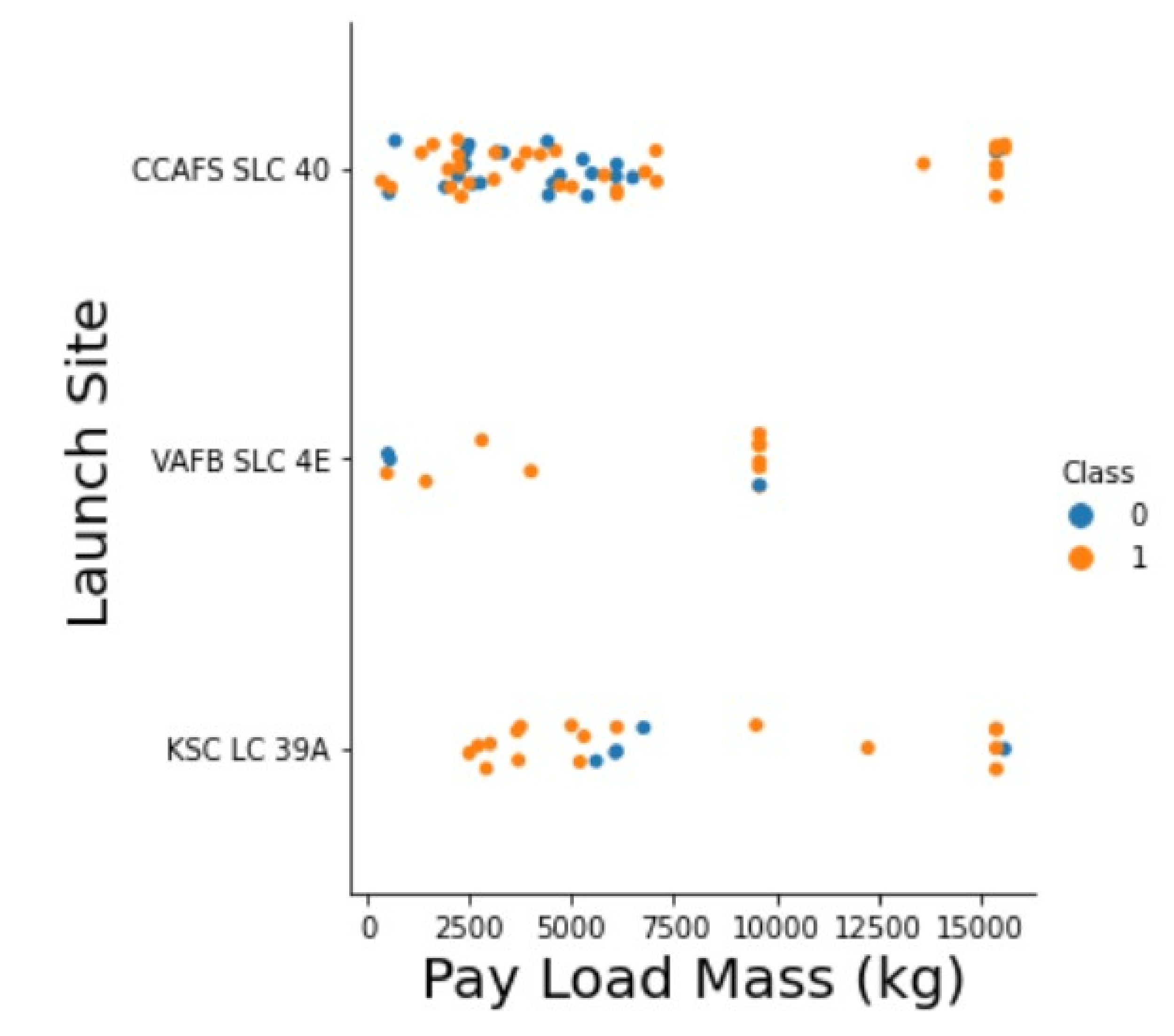
Insights drawn from EDA

Flight Number vs. Launch Site



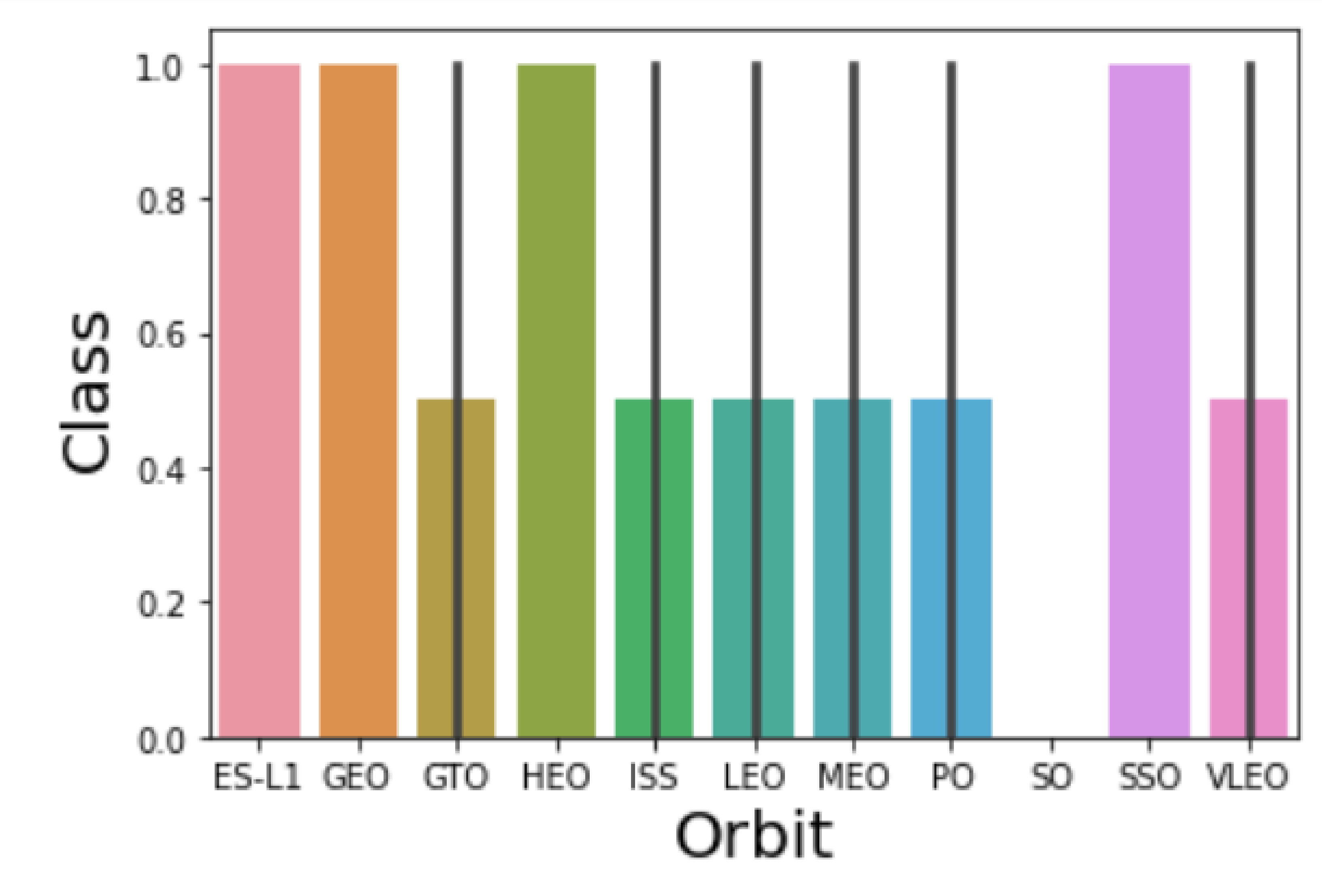
- VAFB SLC 4E had no failures after FL 50
- There were generally less failures as the FL increased

Payload vs. Launch Site

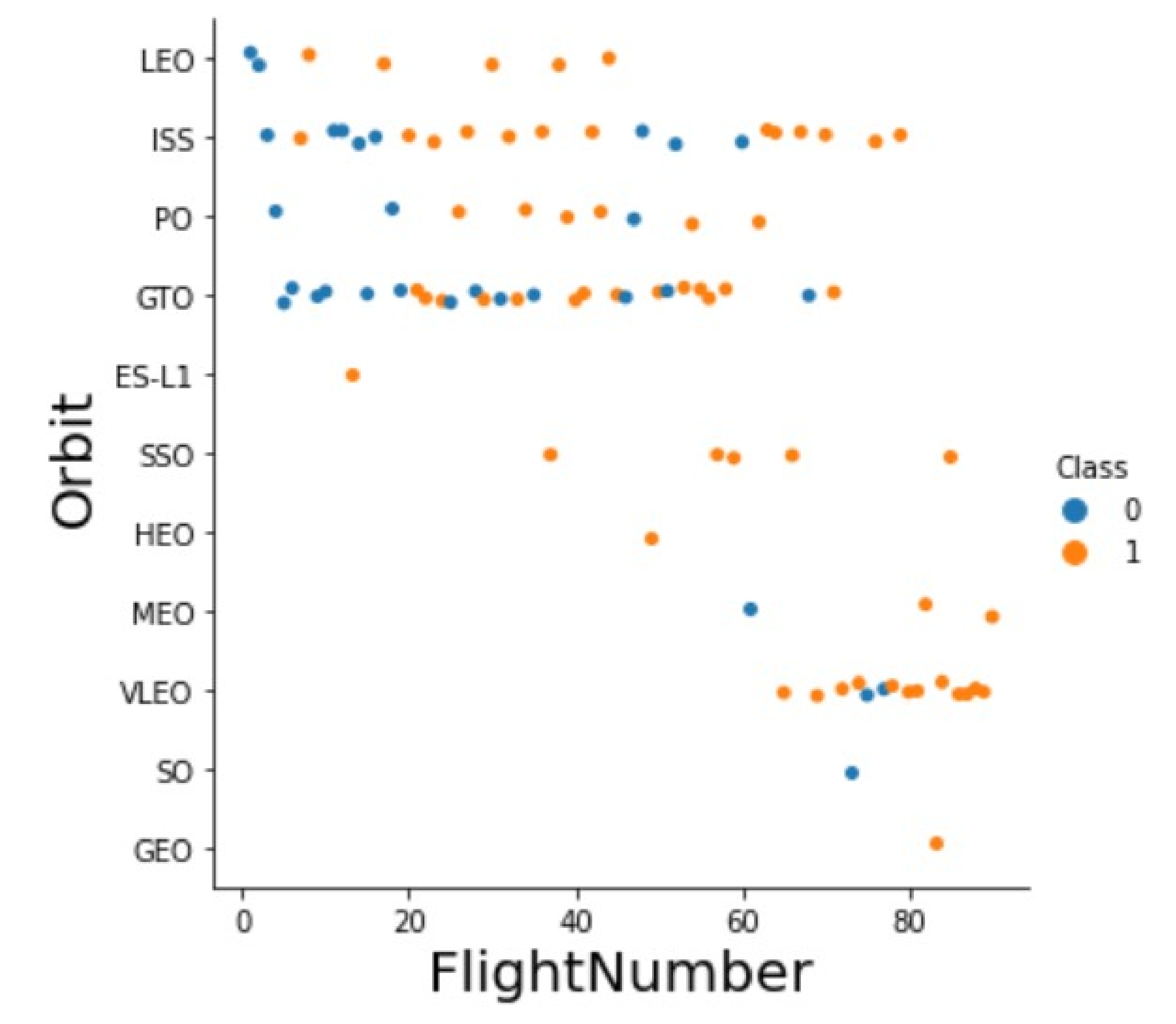


- CCAFS SLC 40 had the least failures with heavy payloads
- VAFB SLC 4E had a good success rate with heavier payloads, but did not launch with loads heavier than 10000kg

Success Rate vs. Orbit Type

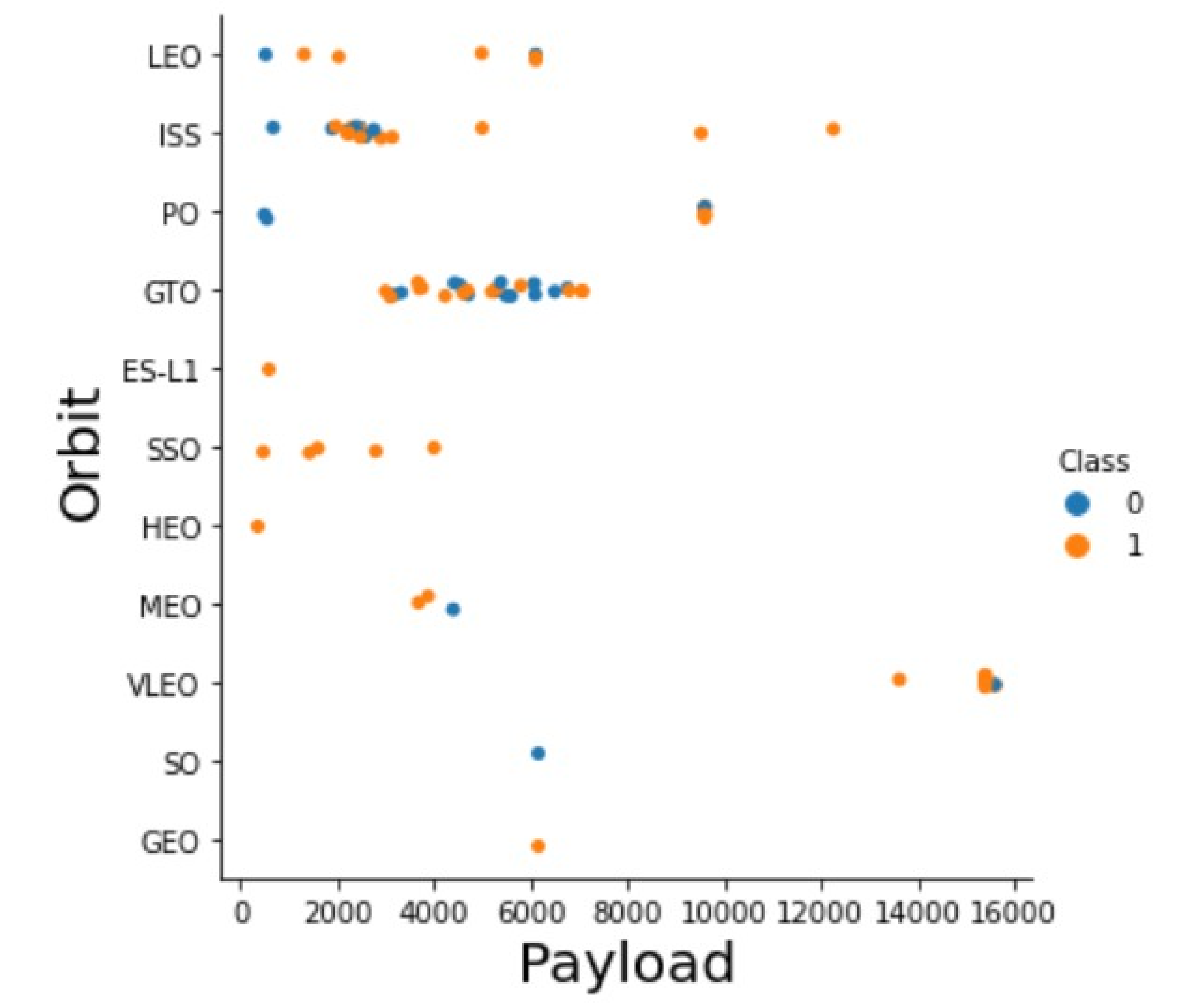


Flight Number vs. Orbit Type

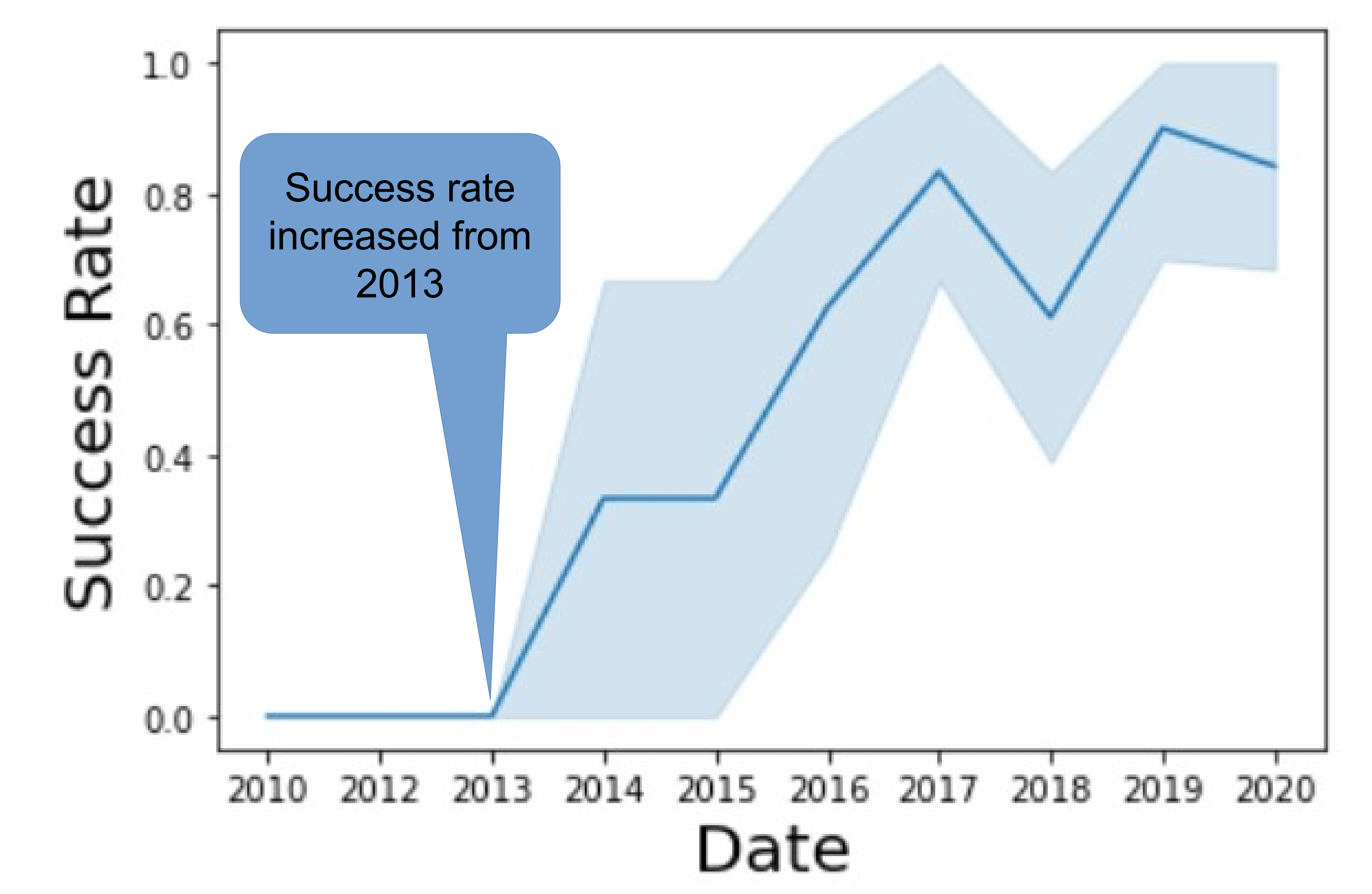


- The LEO success rate is linked with the flight number
- SSO orbits have no failures
- VLEO orbits have a very high success rate

Payload vs. Orbit Type



Launch Success Yearly Trend



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	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

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

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

SUM(PAYLOAD_MASS__KG_)
None

Average Payload Mass by F9 v1.1

```
%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

```
%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

```
%sql SELECT BOOSTER_VERSION from SPACEXTBL WHERE LANDING_OUTCOME = 'Success (drone ship)' and  PAYLOAD_MASS__KG_  >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 count(MISSION_OUTCOME) from SPACEXTBL where MISSION_OUTCOME = 'Success' or MISSION_OUTCOME = 'Failure (in flight
```

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

count(MISSION_OUTCOME)
99

Boosters Carried Maximum Payload

```
%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

```
%sql SELECT BOOSTER_VERSION,LAUNCH_SITE,LANDING_OUTCOME FROM SPACEXTBL WHERE LANDING_OUTCOME = 'Failure (drone ship)' and DA
```

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

Booster_Version	Launch_Site	Landing_Outcome
F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)
F9 v1.1 B1017	VAFB SLC-4E	Failure (drone ship)
F9 FT B1020	CCAFS LC-40	Failure (drone ship)
F9 FT B1024	CCAFS LC-40	Failure (drone ship)

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

```
%sql select * from SPACEXTBL where Landing_Outcome = 'Success (ground pad)' or (DATE between '2010-06-04' and '2017-03-20')
```

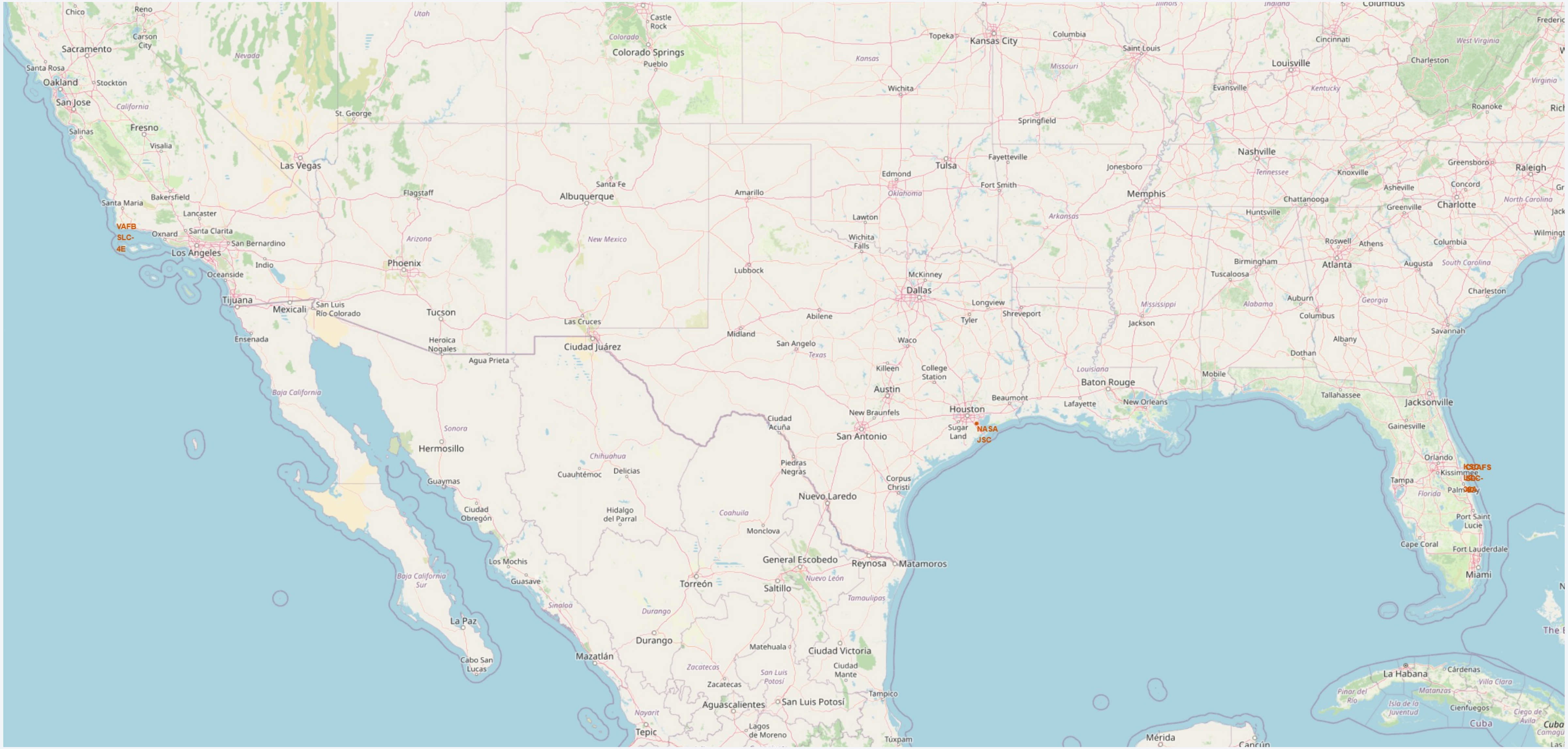
```
 * sqlite:///my_data1.db
Done.
```

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Out
2018-01-08	1:00:00	F9 B4 B1043.1	CCAFS SLC-40	Zuma	5000	LEO	Northrop Grumman	Success (payload status unclear)	Success (gr
2017-12-15	15:36:00	F9 FT B1035.2	CCAFS SLC-40	SpaceX CRS-13	2205	LEO (ISS)	NASA (CRS)	Success	Success (gr
2017-09-07	14:00:00	F9 B4 B1040.1	KSC LC-39A	Boeing X-37B OTV-5	4990	LEO	U.S. Air Force	Success	Success (gr
2017-08-14	16:31:00	F9 B4 B1039.1	KSC LC-39A	SpaceX CRS-12	3310	LEO (ISS)	NASA (CRS)	Success	Success (gr
2017-06-03	21:07:00	F9 FT B1035.1	KSC LC-39A	SpaceX CRS-11	2708	LEO (ISS)	NASA (CRS)	Success	Success (gr
2017-05-01	11:15:00	F9 FT B1032.1	KSC LC-39A	NROL-76	5300	LEO	NRO	Success	Success (gr
2017-03-16	6:00:00	F9 FT B1030	KSC LC-39A	EchoStar 23	5600	GTO	EchoStar	Success	No att
2017-	14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX	2490	LEO	NASA (CRS)	Success	Success (gr

Section 3

Launch Sites Proximities Analysis

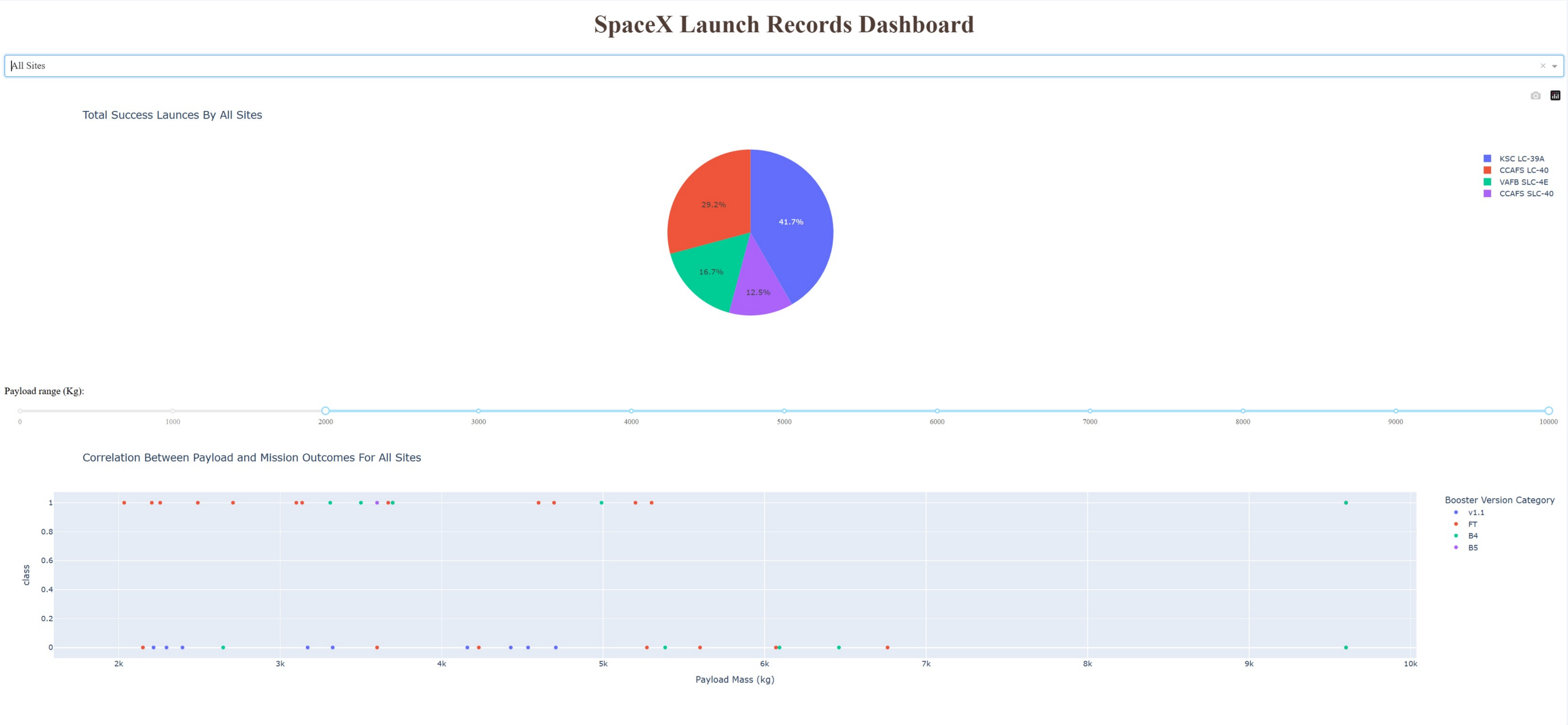
Launch Site Locations



Section 4

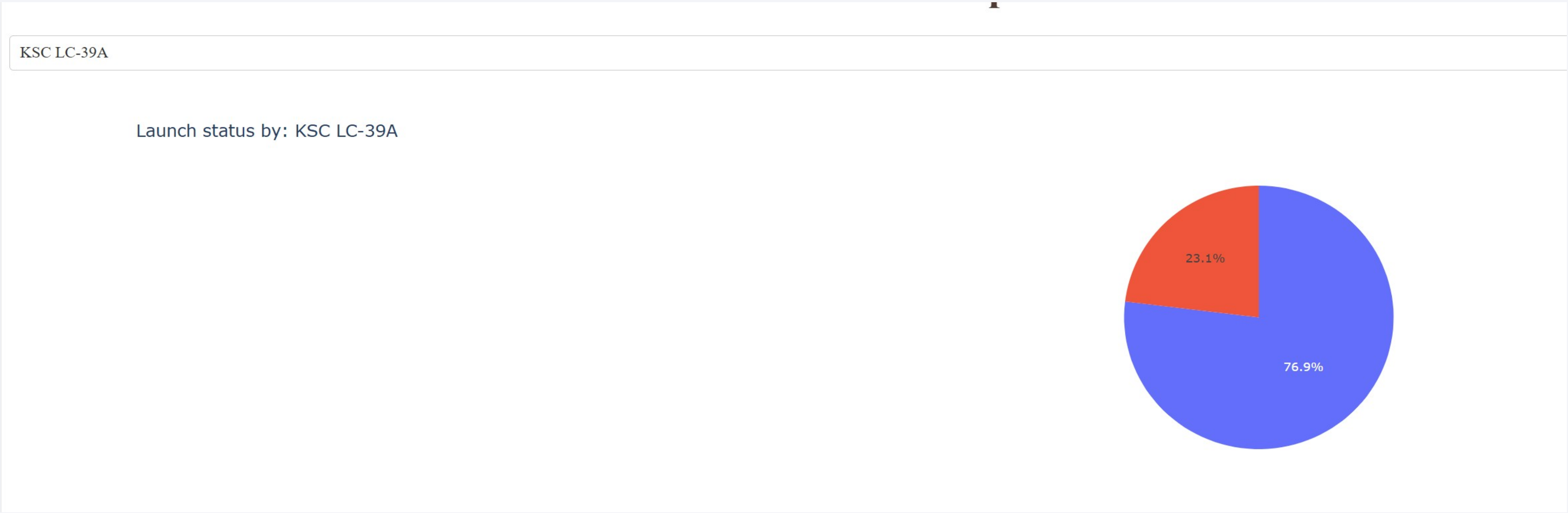
Build a Dashboard with Plotly Dash

Dashboard



SpaceX Launch Records Dashboard
All Sites
Total Success Launces By All Sites
KSC LC-39A
CCAFS LC-40
VAFB SLC-4E
CCAFS SLC-40
41.7%
29.2%
16.7%
12.5%
Payload range (Kg):
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000
Correlation Between Payload and Mission Outcomes For All Sites
class
Payload Mass (kg)
Booster Version Category
v1.1
FT
B4
B5

Launch Site Outcome



We can use Dash to interactively view launch success rates for each launch site

<Dashboard Screenshot 3>

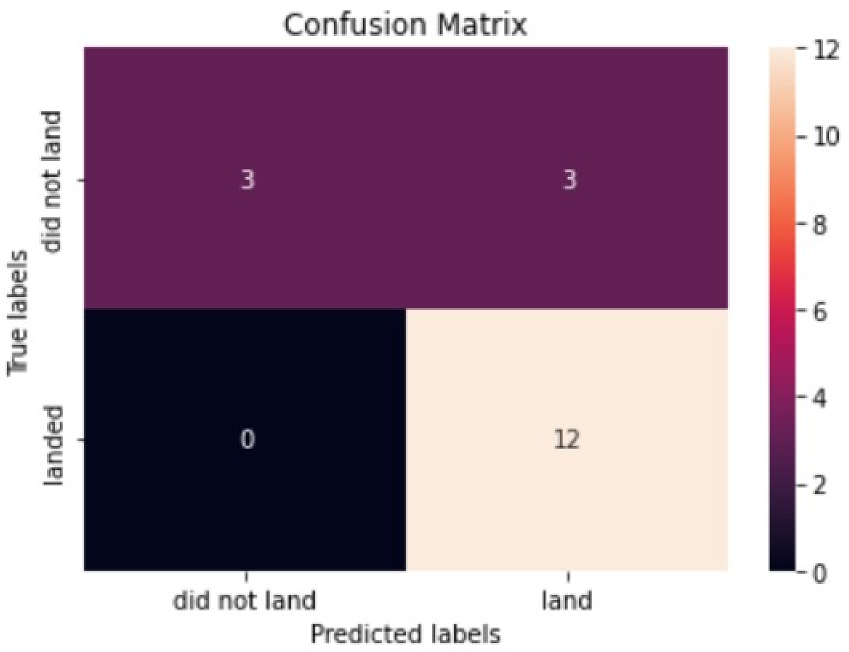
Booster FT has the highest success rate!

Section 5

Predictive Analysis (Classification)

Confusion Matrix

```
yhat=logreg_cv.predict(X_test)
plot_confusion_matrix(Y_test,yhat)
```



Conclusions

- KSC LC 39A was the most successful launch site overall.
- The rate of success is still increasing for all launch sites and payload configuration; a trend which began in 2013 and has been steadily increasing since.
- GTO has the least successful returns, while GEO, HEO, and SSO orbits have the highest rates of success.
- There is no clear indication that the payload mass has an effect on the landing outcome.

Appendix

Thank you for reviewing my capstone project!

Thank you!