



# SpaceX Launch Stage 1 Outcome Analysis

Josh Flieder  
22 June, 2022

# OUTLINE

---

- Executive Summary
- Introduction
- Methodology
  - Data collection
  - Data wrangling
  - Predictive analysis
- Results
  - Interactive visual analytics
  - Predictive analysis
  - Visual results
  - SQL results
  - Interactive Folium map
  - Plotly Dashboard
  - Predictive analysis



- Discussion
- Findings and Implications
- Conclusion
- Appendix

# EXECUTIVE SUMMARY

---

- Data was collected from a SpaceX API as well as scrapped from a wikipedia page
- Missing data was treated and relevant data was then analyzed
- Plots, queries, and dashboards were made to visualize the data.
- V1.1 boosters have a poor success rate
- KSC launches have the best results with B4 boosters, and with launch payloads under 5.5k kg
- CCAFS LC launches have the best results with FT boosters
- Launches to the ES-L1, GEO, HEO, and SSO orbits have not resulted in failures.

# INTRODUCTION

---

- Reuse of stage 1 saves significant money for SpaceX
- SpaceX requires the knowledge of SpaceX's success in saving stage 1 from launches
- How has SpaceX done in saving stage 1?
- Will future SpaceX launches succeed?
- What factors effect SpaceX launch success?
  - Launch site
  - Payload
  - Booster version
  - Orbit
  - Etc.



# METHODOLOGY: Collection and Wrangling

---

- Data sources for SpaceX launches:
  - <https://api.spacexdata.com/v4/launches/past>
  - [https://en.wikipedia.org/w/index.php?title=List\\_of\\_Falcon\\_9\\_and\\_Falcon\\_Heavy\\_launches&oldid=1027686922](https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922)
- Data wrangling:
  - Functions were used to grab desired columns of data
  - Non-Falcon 9 launches were filtered out
  - Unknown payload masses replaced with average value

```
df = df[ df['BoosterVersion'] == 'Falcon 9']
```

```
# Replace the np.nan values with its mean value
```

```
data_falcon9['PayloadMass'].fillna(data_falcon9['PayloadMass'].mean(), inplace = True)
```

# METHODOLOGY: EDA

---

- Check for success/failure of missions by plotting relationships:
  - Payload mass vs flight number
  - Launch site vs flight number
  - Launch site vs payload mass
  - Orbit vs flight number
  - Orbit vs payload mass
- Look into success percentages by orbit
- Track yearly success overall

# METHODOLOGY: EDA and Viz Analytics

---

- Check success of launches with heaviest payloads
- Compare totals of failures vs successes
- Investigate relationship between booster version and payload mass
- Find the average payload
- Count successes and failures for each launch site
  - Use Folium module and marker clusters

# METHODOLOGY: Predictive analysis

---

- Predictive analysis preparation:
  - Data normalized with `sklearn.preprocessing.StandardScaler()`
  - Data prepared with `train_test_split` w/ test size = 0.2 and random state = 2
  - Test set consisted of 18 samples
- Logistic regression GridSearch conducted
  - "C": [0.01, 0.1, 1]
  - 'penalty': ['l2']
  - 'solver': ['lbfgs']}
  - cv = 10
- SVM GridSearch conducted
  - 'kernel': ('linear', 'rbf', 'poly', 'rbf', 'sigmoid')
  - 'C': `np.logspace(-3, 3, 5)`
  - 'gamma': `np.logspace(-3, 3, 5)`
  - cv = 10



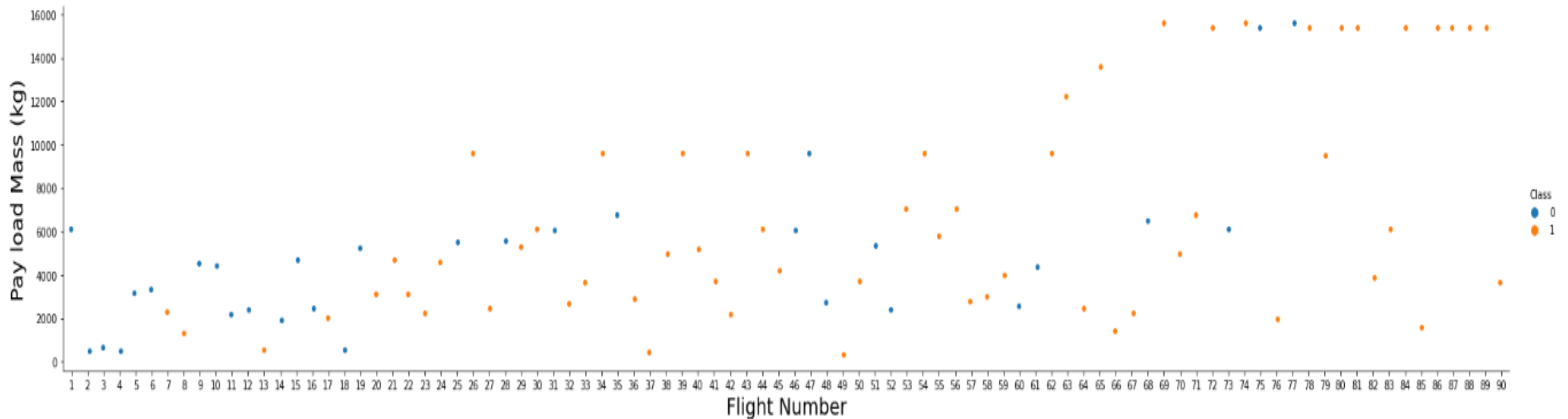
# METHODOLOGY: Predictive analysis

---

- Decision Tree Classifier GridSearch conducted
  - 'criterion': ['gini', 'entropy']
  - 'splitter': ['best', 'random']
  - 'max\_depth': [2\*n for n in range(1,10)]
  - 'max\_features': ['auto', 'sqrt']
  - 'min\_samples\_leaf': [1, 2, 4]
  - 'min\_samples\_split': [2, 5, 10]}
  - cv = 10
- K-Nearest Neighbors GridSearch conducted
  - 'n\_neighbors': [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]
  - 'algorithm': ['auto', 'ball\_tree', 'kd\_tree', 'brute']
  - 'p': [1,2]
  - cv = 10

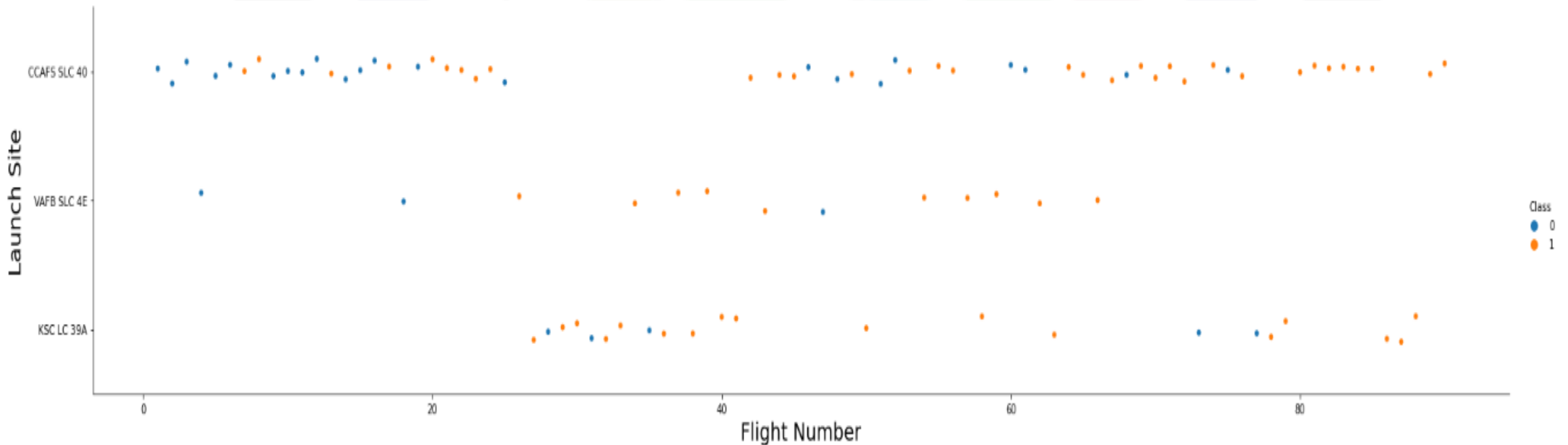
# RESULTS: EDA with Visualization

- Success rate appears to increase as flights progressed, and tended to increase with heavier payload masses



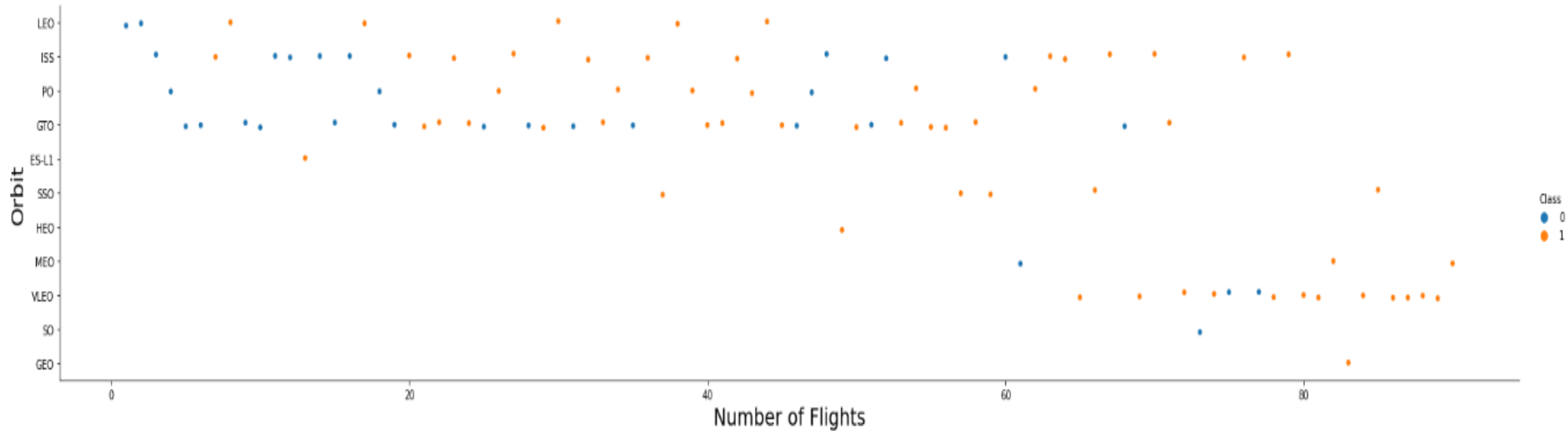
# RESULTS: EDA with Visualization

- CCAFS has the highest number of successes overall



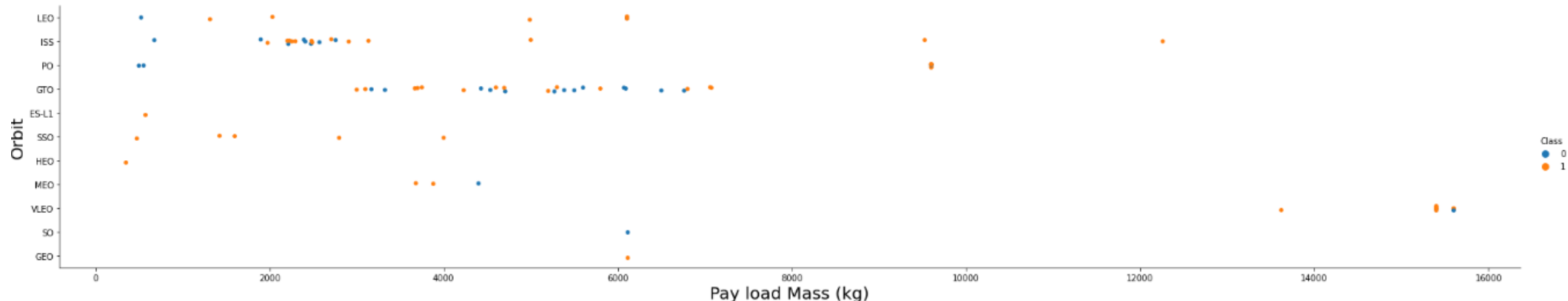
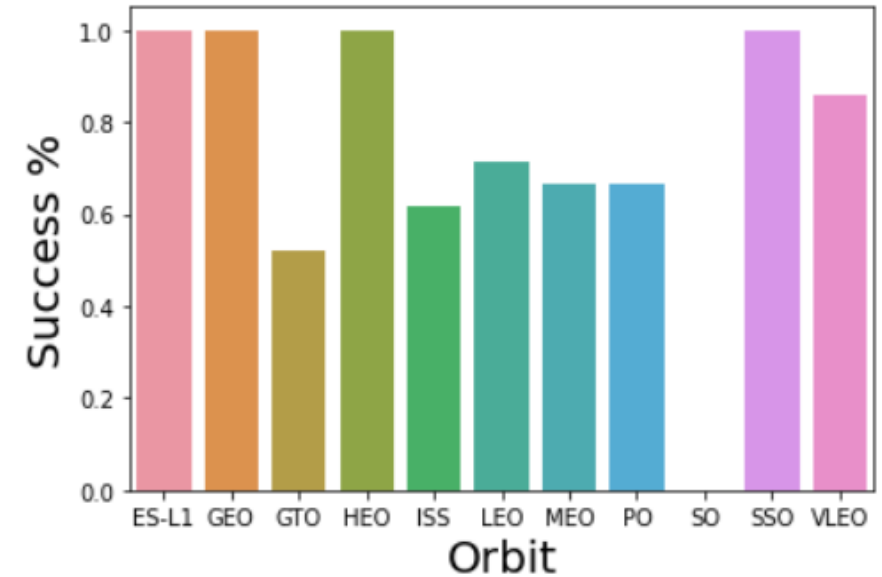
# RESULTS: EDA with Visualization

- Orbit vs number of flights shows nothing in particular



# RESULTS: EDA with Visualization

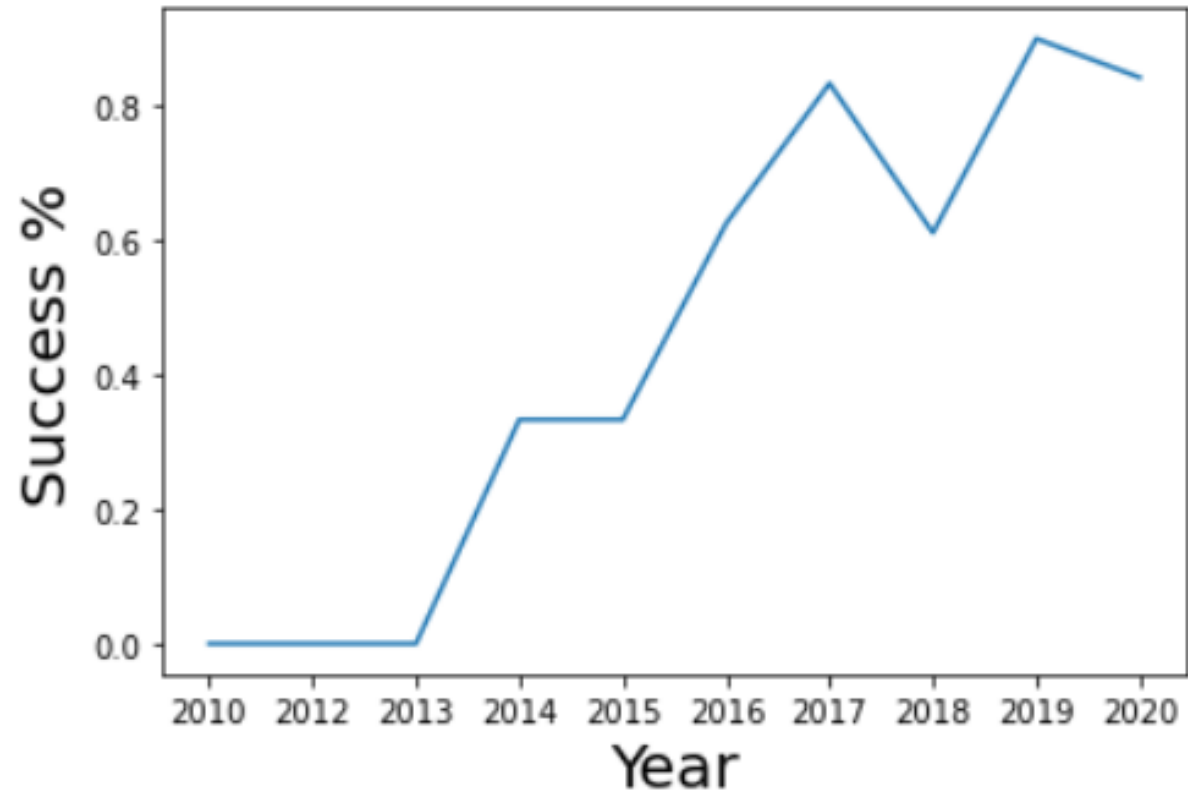
- Launches to the SO orbit have no successes
- Launches to the following orbits have only successes:
  - ES-L1
  - GEO
  - HEO
  - SSO



# RESULTS: EDA with visualization

---

- SpaceX's success rate has generally, but not without exception, trended upwards over time



# RESULTS: EDA with SQL

- Here are the four unique launch sites:



```
%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

- Here are the payload masses of NASA CRS launches:



```
%sql select PAYLOAD_MASS_KG_ From SPACEXTBL Where Customer is 'NASA (CRS)'
```

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

**PAYLOAD\_MASS\_KG\_**

500

677

2296

2216

2395

1898

1952

3136

2257

2490

2708

3310

2205

2647

2697

2500

2495

2268

1977

2972

# RESULTS: EDA with SQL

- Here are the first 5 records from a launch site beginning 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
04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
08-12-2010	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)
22-05-2012	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
01-03-2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt



# RESULTS: EDA with SQL

- The average payload mass in kg:



```
%%sql select AVG(PAYLOAD_MASS_KG_) From SPACEXTBL  
Where Booster_Version like 'F9 v1.1%'
```

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

```
AVG(PAYLOAD_MASS_KG_)
```

```
2534.6666666666665
```

- The date of the first successful ground pad launch:



```
%%sql select MIN(DATE) From SPACEXTBL Where "Landing _Outcome" is 'Success (ground pad)'
```

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

```
MIN(DATE)
```

```
01-05-2017
```

# RESULTS: EDA with SQL

- The boosters between 4k and 6k kg that landed successfully on a drone ship



```
%%sql select Booster_Version From SPACEXTBL
Where "Landing_Outcome" is 'Success (drone ship)
AND PAYLOAD_MASS__KG_ Between 4000 AND 6000
```

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

**Booster\_Version**

```
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2
```

- The count of all launch outcomes



```
%%sql select (Mission_Outcome), count(Mission_Outcome) as Total From SPACEXTBL
Group by Mission_Outcome
```

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

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

# RESULTS: EDA with SQL

- Booster versions which carried the maximum launched payload mass

```
%%sql select Booster_Version, PAYLOAD_MASS_KG_ From SPACEXTBL  
Where PAYLOAD_MASS_KG_ = (select MAX(PAYLOAD_MASS_KG_) From SPACEXTBL)
```

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

Booster_Version	PAYLOAD_MASS_KG_
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

- Details on the failed drone ship landings in 2015



```
%%sql select Date, substr(Date, 4, 2) as Month, Booster_Version, Launch_Site From SPACEXTBL  
Where substr(Date,7,4)='2015'  
AND "Landing_Outcome" = 'Failure (drone ship)'
```

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

Date	Month	Booster_Version	Launch_Site
10-01-2015	01	F9 v1.1 B1012	CCAFS LC-40
14-04-2015	04	F9 v1.1 B1015	CCAFS LC-40

# RESULTS: EDA with SQL

- Count of successful landing\_outcomes between the date 04-06-2010 and 20-03-2017 in descending order

```
%%sql select Date, substr(Date, 4, 2) as Month, substr(Date,7,4) as Year, substr(Date,1,2) as Day,  
Booster_Version, Launch_Site, "Landing_Outcome"  
From SPACEXTBL  
Where "Landing_Outcome" like 'Success%'  
Order by Year desc, Month desc, Day desc
```

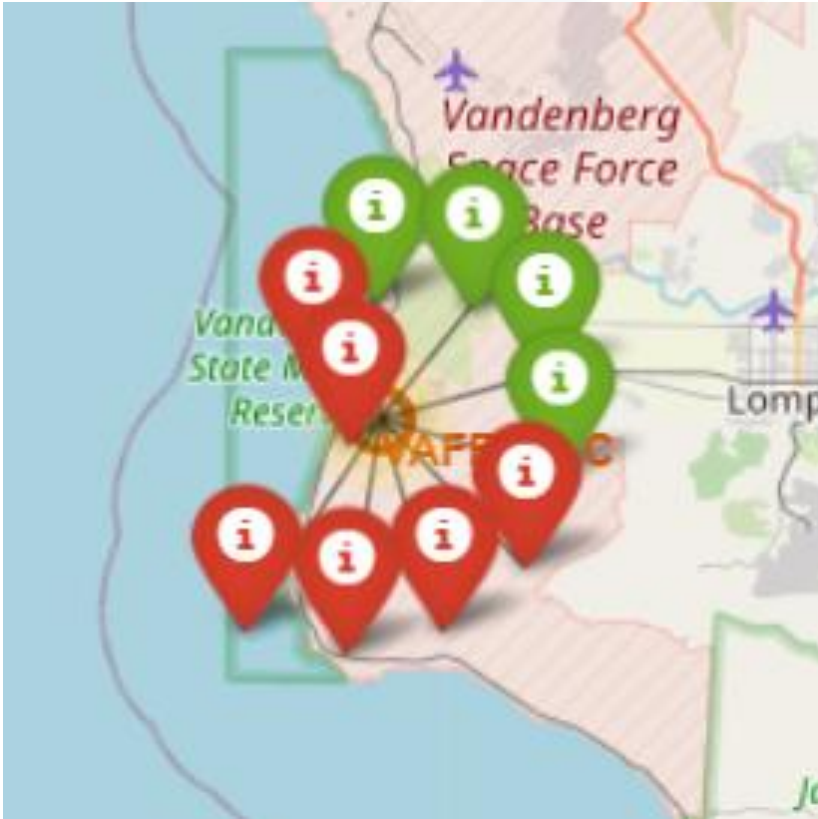
\* sqlite:///my\_data1.db

Done.

Date	Month	Year	Day	Booster_Version	Launch_Site	Landing_Outcome
06-12-2020	12	2020	06	F9 B5 B1058.4	KSC LC-39A	Success
25-11-2020	11	2020	25	F9 B5 B1049.7	CCAFS SLC-40	Success
21-11-2020	11	2020	21	F9 B5B1063.1	VAFB SLC-4E	Success
16-11-2020	11	2020	16	F9 B5B1061.1	KSC LC-39A	Success
05-11-2020	11	2020	05	F9 B5B1062.1	CCAFS SLC-40	Success
24-10-2020	10	2020	24	F9 B5 B1060.3	CCAFS SLC-40	Success
18-10-2020	10	2020	18	F9 B5 B1051.6	KSC LC-39A	Success
06-10-2020	10	2020	06	F9 B5 B1058.3	KSC LC-39A	Success
03-09-2020	09	2020	03	F9 B5 B1060.2	KSC LC-39A	Success

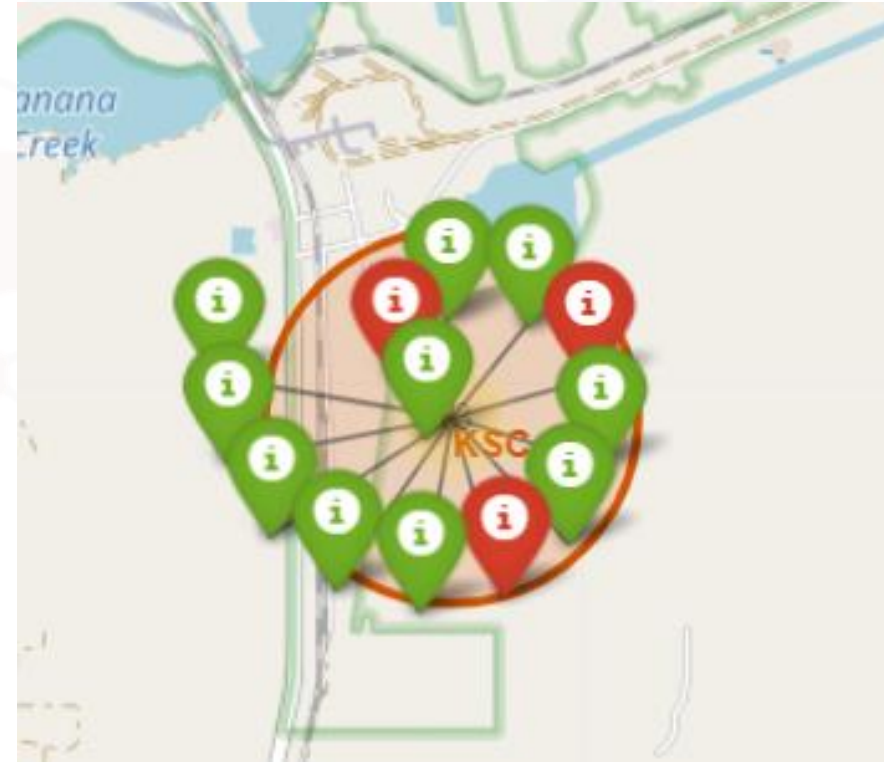
# RESULTS: FoIum

- Here is are the launches from VAFB:



IBM Developer

- Here is are the launches from KSC:



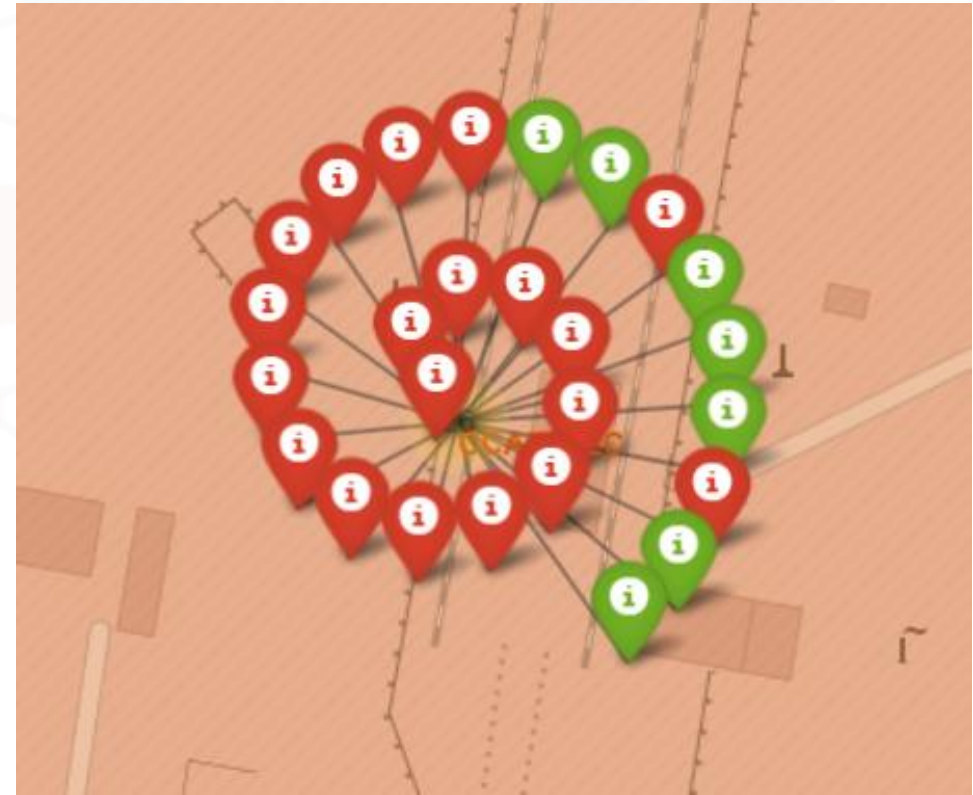
SKILLS NETWORK 

# RESULTS: Fo1ium

- Here is are the launches from CCAFS\_SLC:



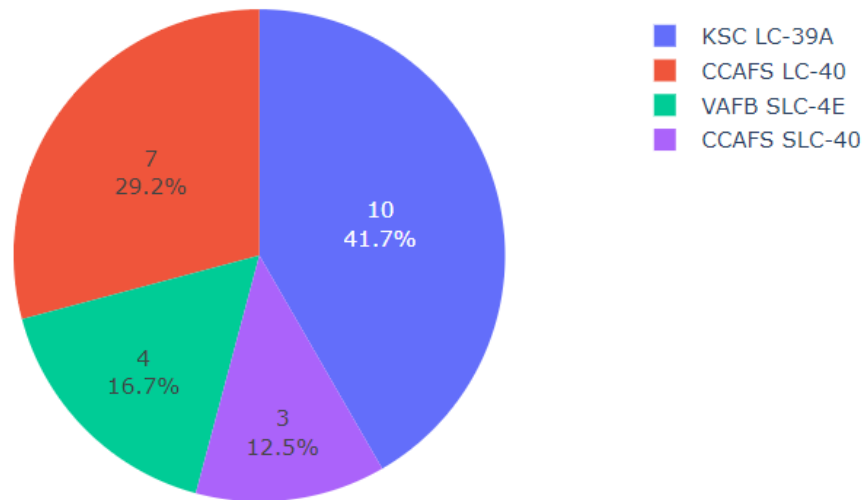
- Here is are the launches from CCAFS\_LC:



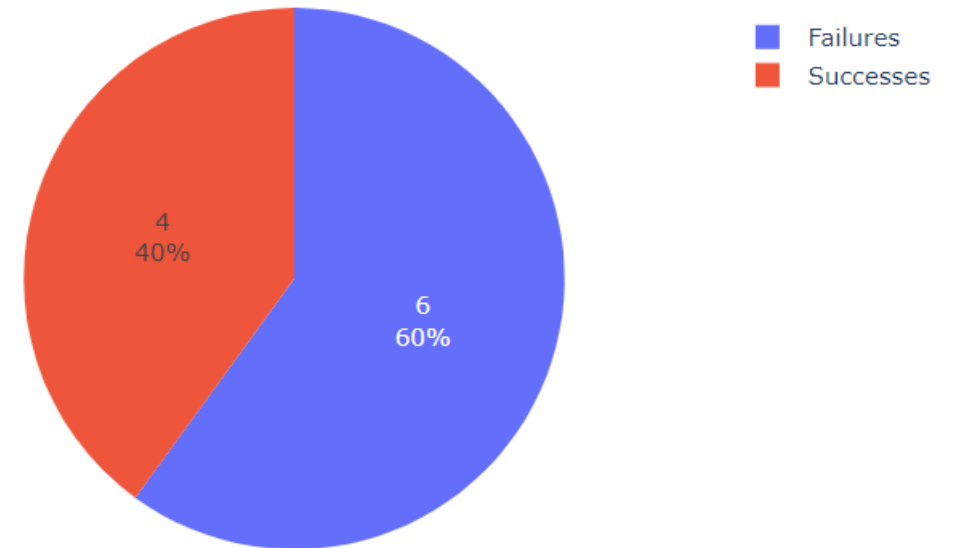
# RESULTS: Plotly Dashboard

- Here is the distribution of successful launches amongst the launch sites:
- Here are the failure and success counts/percentages of launches from VAFB

Total Successful Launches by Launch Site



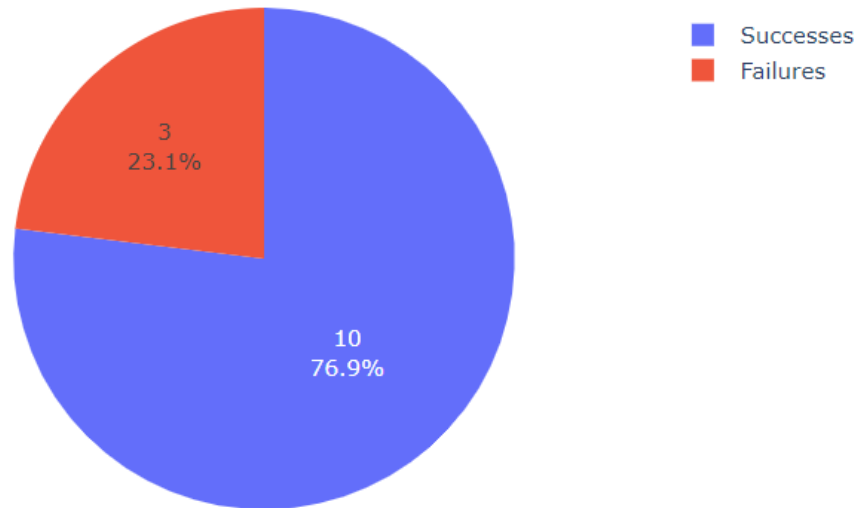
Successful vs Failed Launches for VAFB SLC-4E



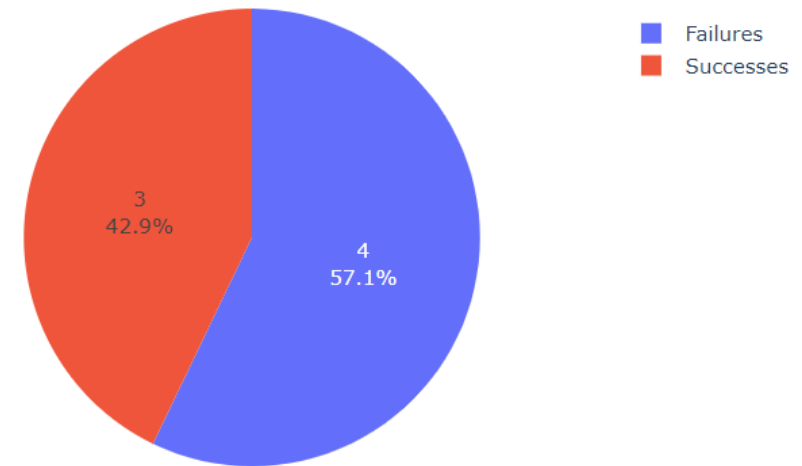
# RESULTS: Plotly Dashboard

- Here are the failure and success counts/percentages of launches from KSC
- Here are the failure and success counts/percentages of launches from CCAFS SLC

Successful vs Failed Launches for KSC LC-39A



Successful vs Failed Launches for CCAFS SLC-40

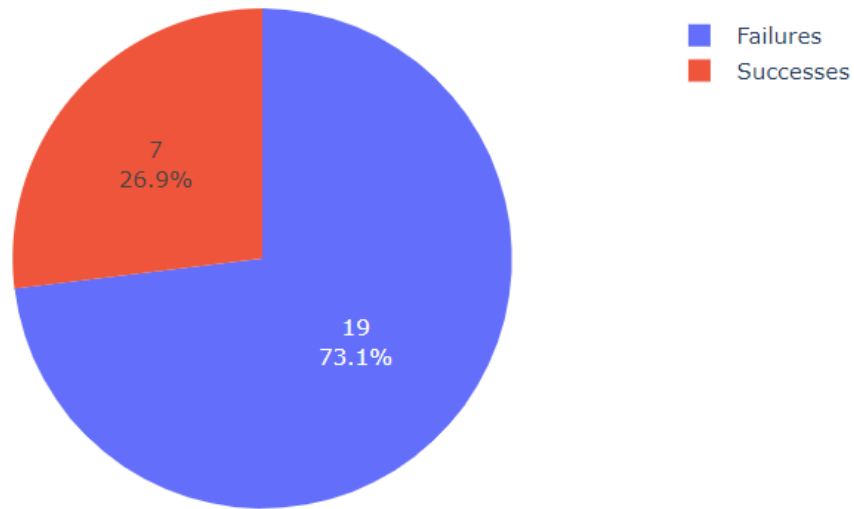




# RESULTS: Plotly Dashboard

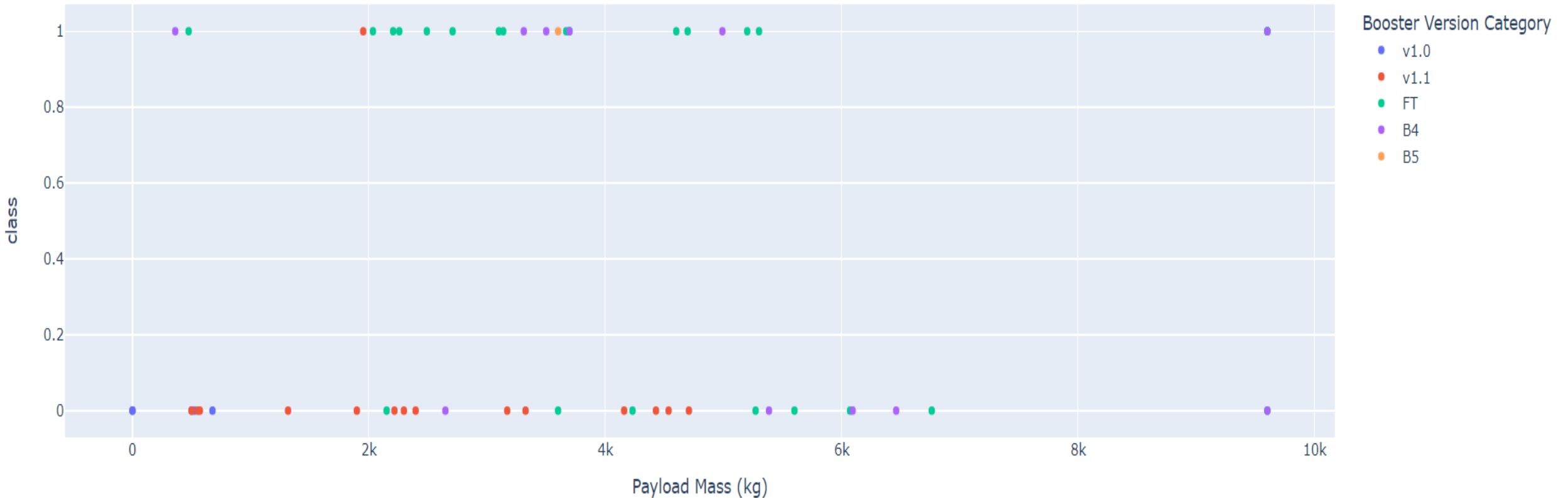
- Here are the failure and success counts/percentages of launches from CCAFS LC

Successful vs Failed Launches for CCAFS LC-40



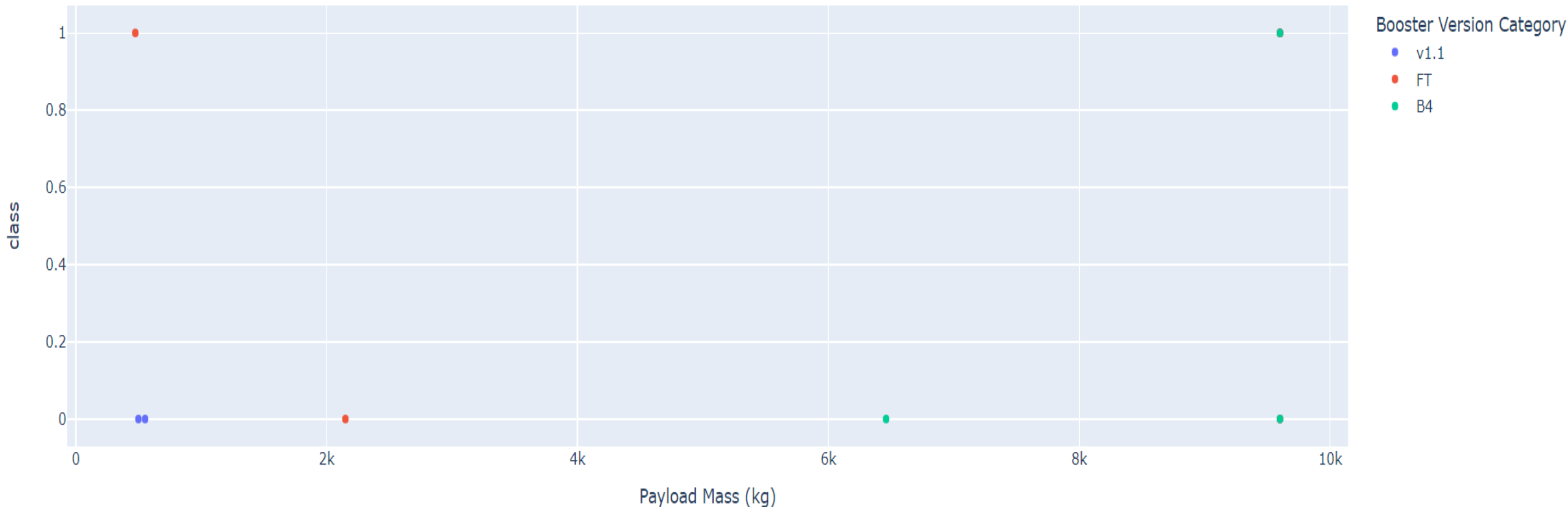
# RESULTS: Plotly Dashboard

- Here is the spread of successes and failures broken down by payload mass and booster version for all launches



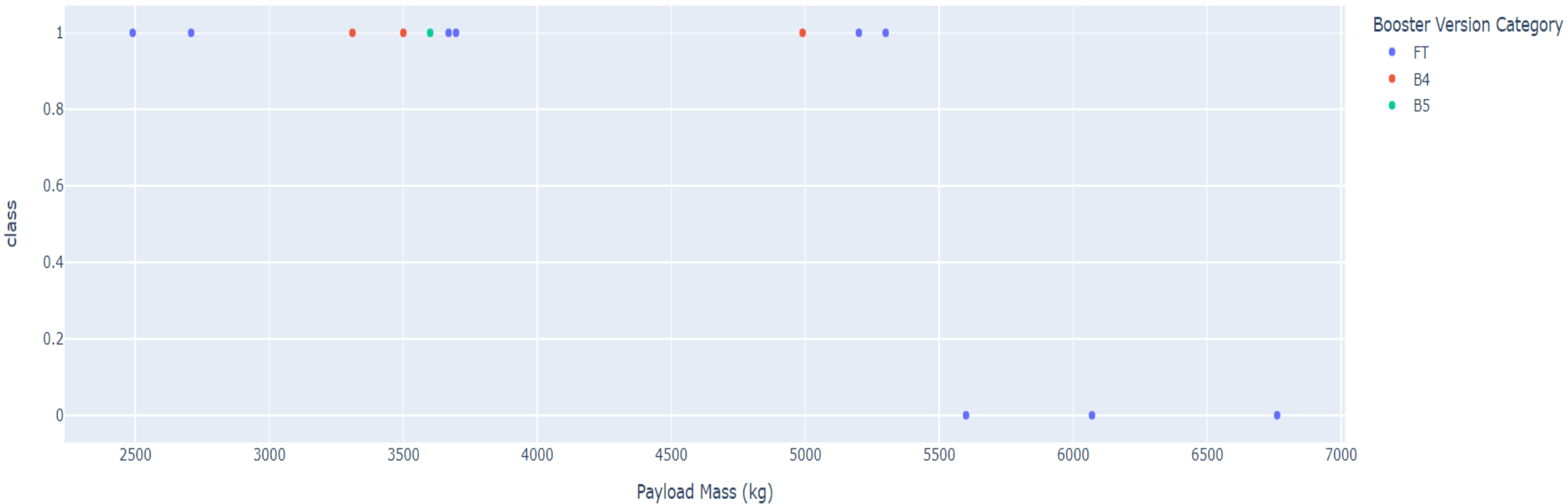
# RESULTS: Plotly Dashboard

- Here is the spread of successes and failures broken down by payload mass and booster version for launches from VAFB



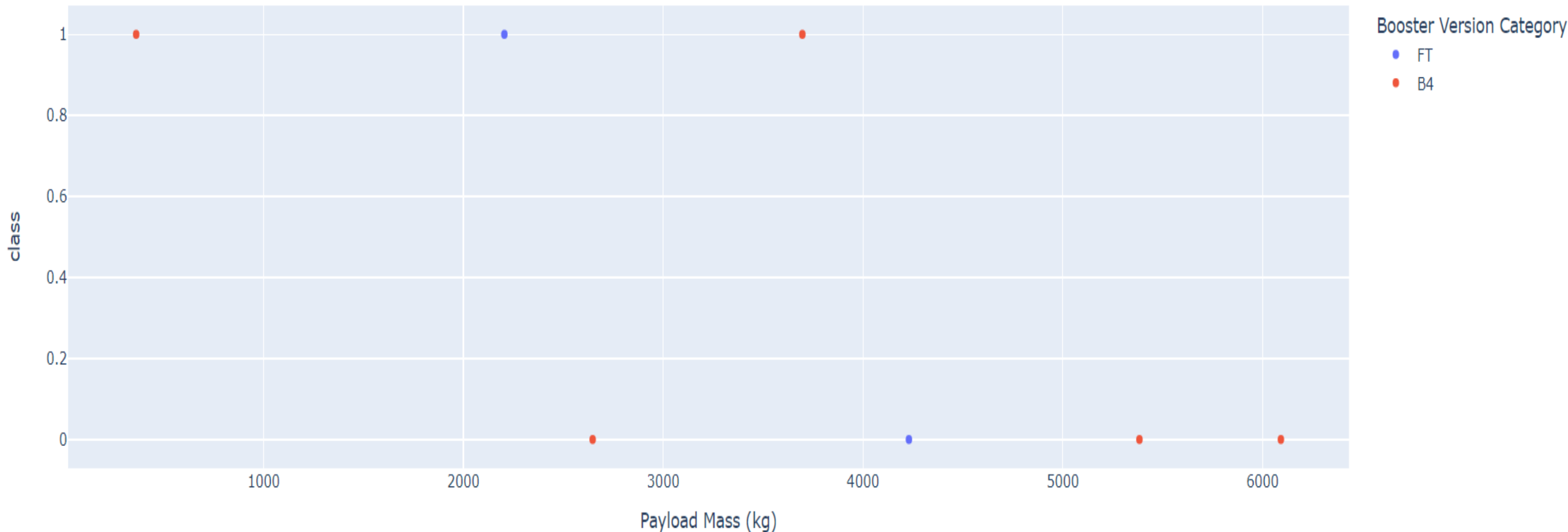
# RESULTS: Plotly Dashboard

- Here is the spread of successes and failures broken down by payload mass and booster version for launches from KSC



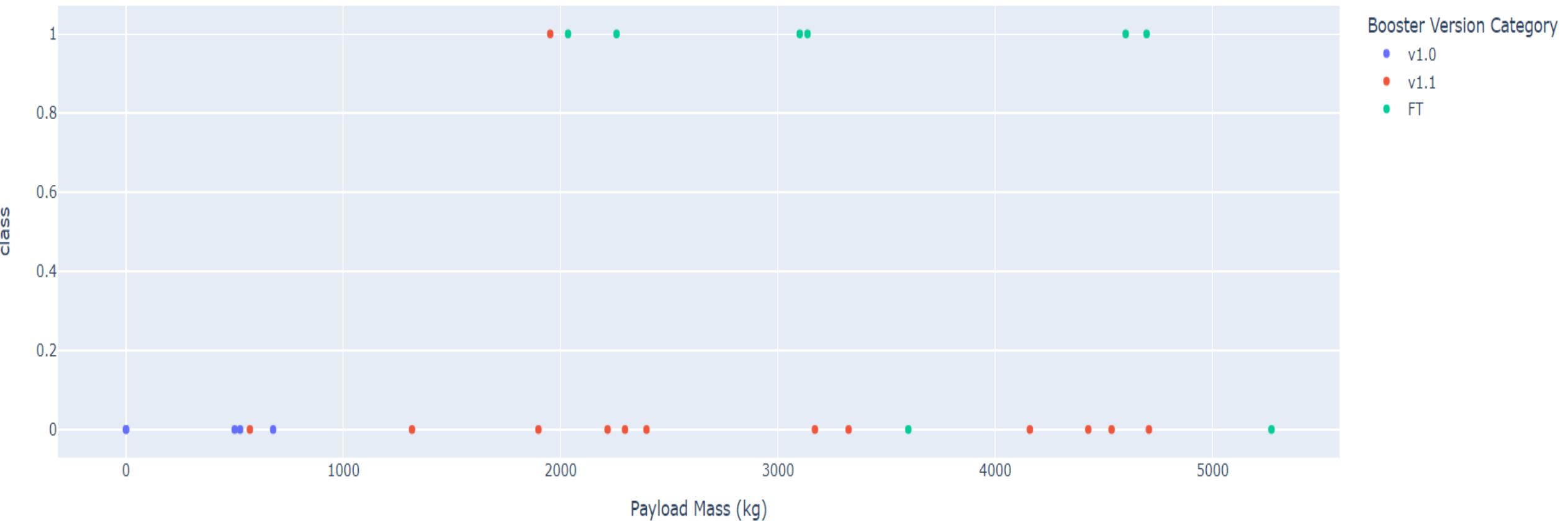
# RESULTS: Plotly Dashboard

- Here is the spread of successes and failures broken down by payload mass and booster version for launches from CCAFS SLC



# RESULTS: Plotly Dashboard

- Here is the spread of successes and failures broken down by payload mass and booster version for launches from CCAFS LC



# RESULTS: Predictive Analysis

---

- The calculated accuracies of the GridSearch models used were:
  - Logistic regression: 82.22%
  - SVM: 82.22%
  - Decision tree classifier: 88.88%
  - K-Nearest Neighbors: 84.44%
- Therefore, the best model to use going forward is the Decision tree classifier.

# CONCLUSION

---

- KSC is the launch site with the best success rate based on the Folium and Plotly results
- FT boosters carrying a payload under 6k kg have a good success rate
- V1.1 boosters have a poor success rate
- KSC launches have the best results with B4 boosters, and with launch payloads under 5.5k kg
- CCAFS LC launches have the best results with FT boosters
- Launches to the ES-L1, GEO, HEO, and SSO orbits have not resulted in failures.