IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Sarah Conley
2 October 2024

Outline

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

Executive Summary

- Collect data with APIs/Web Scraping
- Perform EDA to determine focus of project
- Analyzed what variables could have an effect on mission success probability of success
- Fit models in order to predict a new mission’s outcome
- Found our model performs at an accuracy of 83.33%

Introduction

- What factors impact a missions probability of success
- Can we fit a model to accurately predict a mission’s outcome

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Collected the data by using APIs/Web Scraping
  - Provided by Wikipedia
- Perform data wrangling
  - Cleaned data and put into data frame after scraping
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Including: SVM, KNN, Tree

Data Collection

- Used APIs to gather the data:
  - Acquired historical launch data from an open source REST API for SpaceX
- Web Scraping
  - Acquired historical launch data from Wikipedia
    - List of Falcon 9 Heavy Launches

2020 [edit]

In late 2019, Gwynne Shotwell stated that SpaceX hoped for as many as 24 launches for Starlink satellites in 2020,[490] in addition to 14 or 15 non-Starlink launches. At 26 launches, 13 of which for Starlink satellites, Falcon 9 had its most prolific year, and Falcon rockets were second most prolific rocket family of 2020, only behind China's Long March rocket family.[491]

[hide] Flight No.	Date and time (UTC)	Version, Booster[b]	Launch site	Payload[c]	Payload mass	Orbit	Customer	Launch outcome	Booster landing
78	7 January 2020, 02:19:21[492]	F9 B5 B1049.4	CCAFS, SLC-40	Starlink 2 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Success (drone ship)
	Third large batch and second operational flight of Starlink constellation. One of the 60 satellites included a test coating to make the satellite less reflective, and thus less likely to interfere with ground-based astronomical observations.[493]								
79	19 January 2020, 15:30[494]	F9 B5 B1046.4	KSC, LC-39A	Crew Dragon in-flight abort test[495] (Dragon C205.1)	12,050 kg (26,570 lb)	Sub-orbital[496]	NASA (CTS)[497]	Success	No attempt
	An atmospheric test of the Dragon 2 abort system after Max Q. The capsule fired its SuperDraco engines, reached an apogee of 40 km (25 mi), deployed parachutes after reentry, and splashed down in the ocean 31 km (19 mi) downrange from the launch site. The test was previously slated to be accomplished with the Crew Dragon Demo-1 capsule;[498] but that test article exploded during a ground test of SuperDraco engines on 20 April 2019.[419] The abort test used the capsule originally intended for the first crewed flight.[499] As expected, the booster was destroyed by aerodynamic forces after the capsule aborted.[500] First flight of a Falcon 9 with only one functional stage — the second stage had a mass simulator in place of its engine.								
80	29 January 2020, 14:07[501]	F9 B5 B1051.3	CCAFS, SLC-40	Starlink 3 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Success (drone ship)
	Third operational and fourth large batch of Starlink satellites, deployed in a circular 290 km (180 mi) orbit. One of the fairing halves was caught, while the other was fished out of the ocean.[502]								
81	17 February 2020, 15:05[503]	F9 B5 B1056.4	CCAFS, SLC-40	Starlink 4 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Failure (drone ship)
	Fourth operational and fifth large batch of Starlink satellites. Used a new flight profile which deployed into a 212 km × 386 km (132 mi × 240 mi) elliptical orbit instead of launching into a circular orbit and firing the second stage engine twice. The first stage booster failed to land on the drone ship[504] due to incorrect wind data.[505] This was the first time a flight proven booster failed to land.								
82	7 March 2020, 04:50[506]	F9 B5 B1059.2	CCAFS, SLC-40	SpaceX CRS-20 (Dragon C112.3 ⏃)	1,977 kg (4,359 lb)[507]	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
	Last launch of phase 1 of the CRS contract. Carries *Bartolomeo*, an ESA platform for hosting external payloads onto ISS.[508] Originally scheduled to launch on 2 March 2020, the launch date was pushed back due to a second stage engine failure. SpaceX decided to swap out the second stage instead of replacing the faulty part.[509] It was SpaceX's 50th successful landing of a first stage booster, the third flight of the Dragon C112 and the last launch of the cargo Dragon spacecraft.								
83	18 March 2020, 12:16[510]	F9 B5 B1048.5	KSC, LC-39A	Starlink 5 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Failure (drone ship)
	Fifth operational launch of Starlink satellites. It was the first time a first stage booster flew for a fifth time and the second time the fairings were reused (Starlink flight in May 2019).[511] Towards the end of the first stage burn, the booster suffered premature shut down of an engine, the first of a Merlin 1D variant and first since the CRS-1 mission in October 2012. However, the payload still reached the targeted orbit.[512] This was the second Starlink launch booster landing failure in a row, later revealed to be caused by residual cleaning fluid trapped inside a sensor.[513]								
84	22 April 2020, 19:30[514]	F9 B5 B1051.4	KSC, LC-39A	Starlink 6 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Success (drone ship)

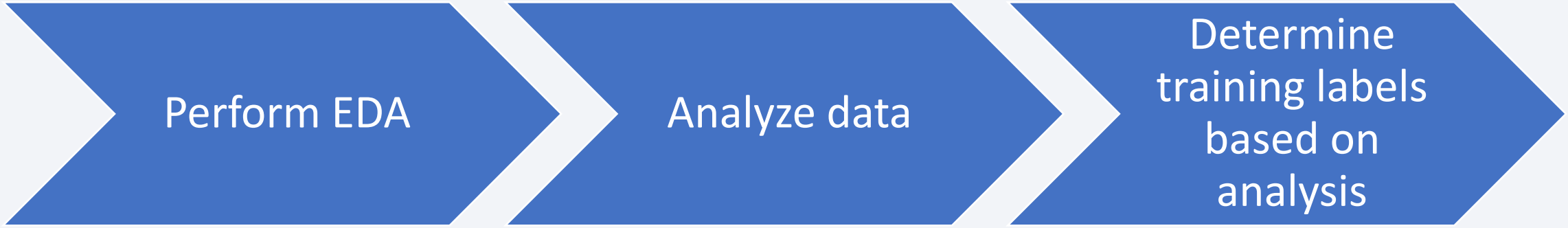
Data Collection – SpaceX API

GitHub URL

Data Collection - Scraping

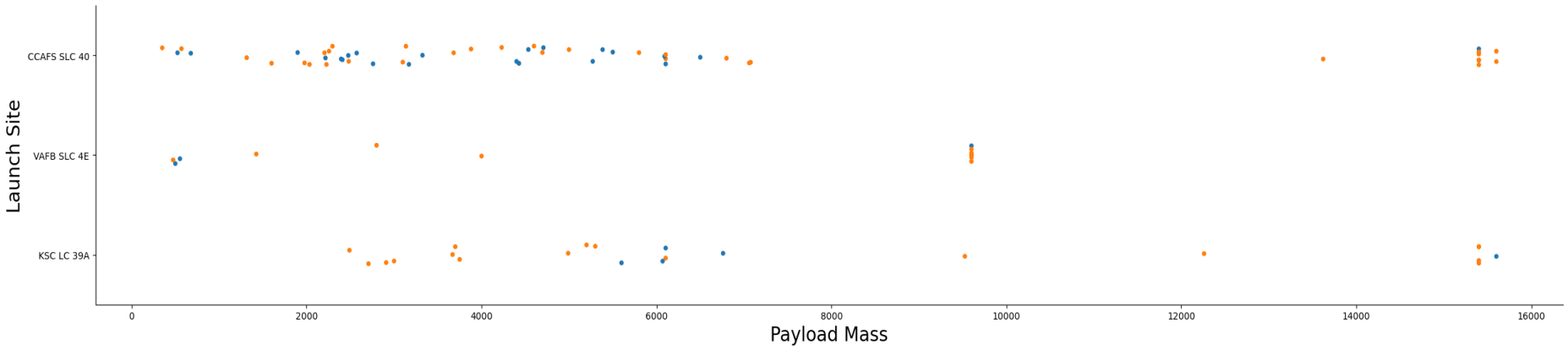
GitHub URL

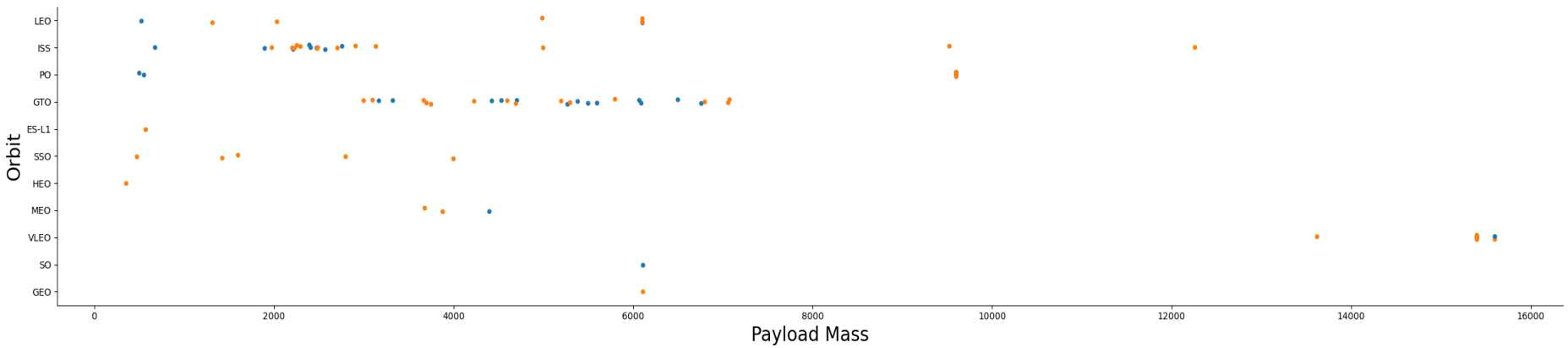
Data Wrangling

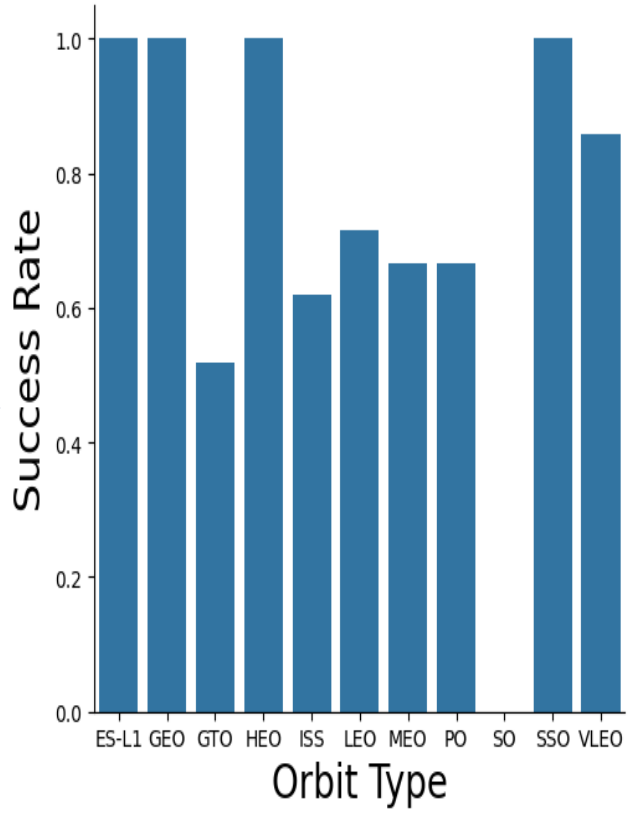


GitHub URL

EDA with Data Visualization







- Payload Mass vs.. Launch Site
- Payload Mass vs. Orbit Type
- Success by Orbit Type

GitHub URL

EDA with SQL

avg(PAYLOAD_MASS__KG_)	Booster_Version
2928.4	F9 v1.1
500.0	F9 v1.1 B1003
2216.0	F9 v1.1 B1010
4428.0	F9 v1.1 B1011
2395.0	F9 v1.1 B1012
570.0	F9 v1.1 B1013
4159.0	F9 v1.1 B1014
1898.0	F9 v1.1 B1015
4707.0	F9 v1.1 B1016
553.0	F9 v1.1 B1017
1952.0	F9 v1.1 B1018

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

Date	Landing_Outcome	count(Landing_Outcome)
2016-04-08	Success (drone ship)	5
2015-12-22	Success (ground pad)	3
2015-06-28	Precluded (drone ship)	1
2015-01-10	Failure (drone ship)	5
2014-04-18	Controlled (ocean)	3
2013-09-29	Uncontrolled (ocean)	2
2012-05-22	No attempt	10
2010-06-04	Failure (parachute)	2

- Can see the average payload mass by Booster version
- Summary of mission outcomes
- Summary of landing outcomes from 06/04/2010 to 04/08/2016

GitHub URL

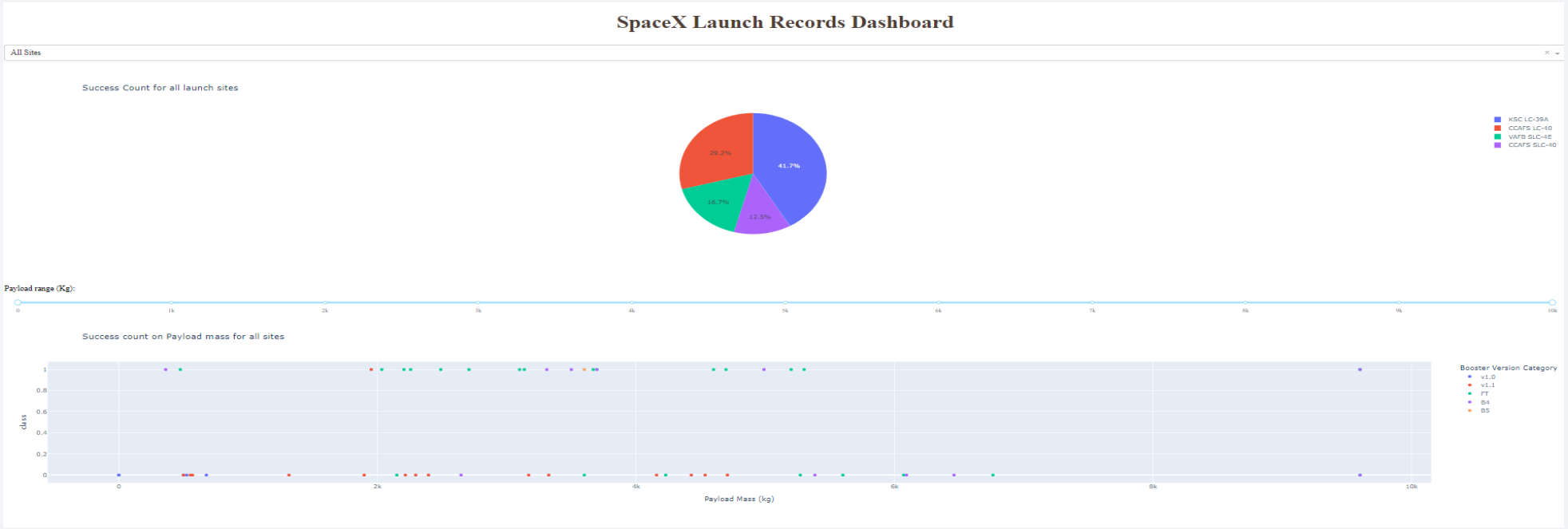
Build an Interactive Map with Folium

- Marked all launch sites on map
- Marked the success/failure launches for ech site on map
- Calculated distances between a launch site to its proximities

GitHub URL

Build a Dashboard with Plotly Dash

- Pie chart of each site, the proportion of launches at that site, and the proportion of successful and unsuccessful launches at those sites
- Scatter plot of success based on payload mass



GitHub URL

Predictive Analysis (Classification)

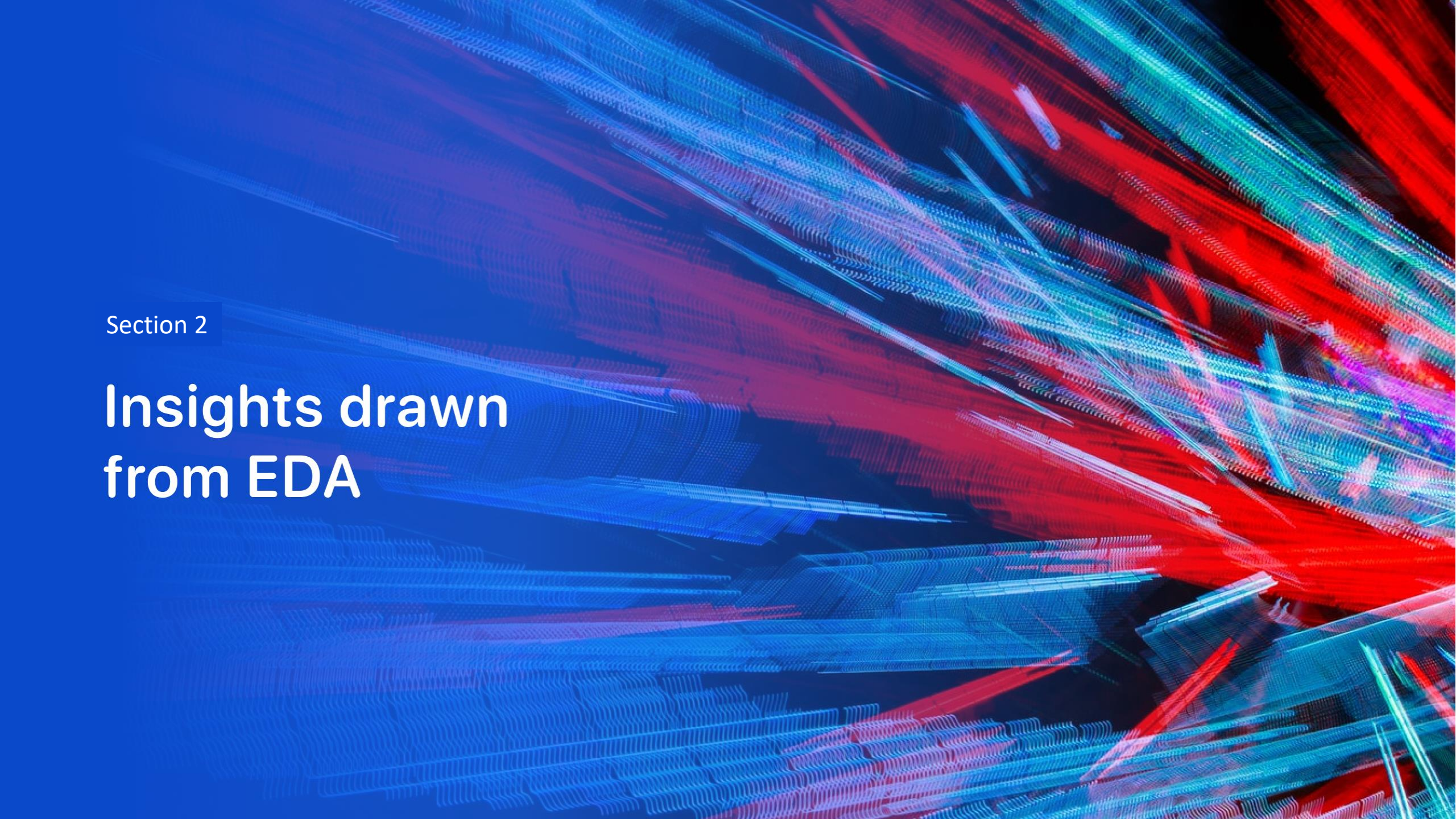
- Create a columns to classify each class
- Standardize the data
- Split into train and test
- Run different ML classification models:
  - Logistic Regression
  - SVM
  - Tree
  - KNN

Results

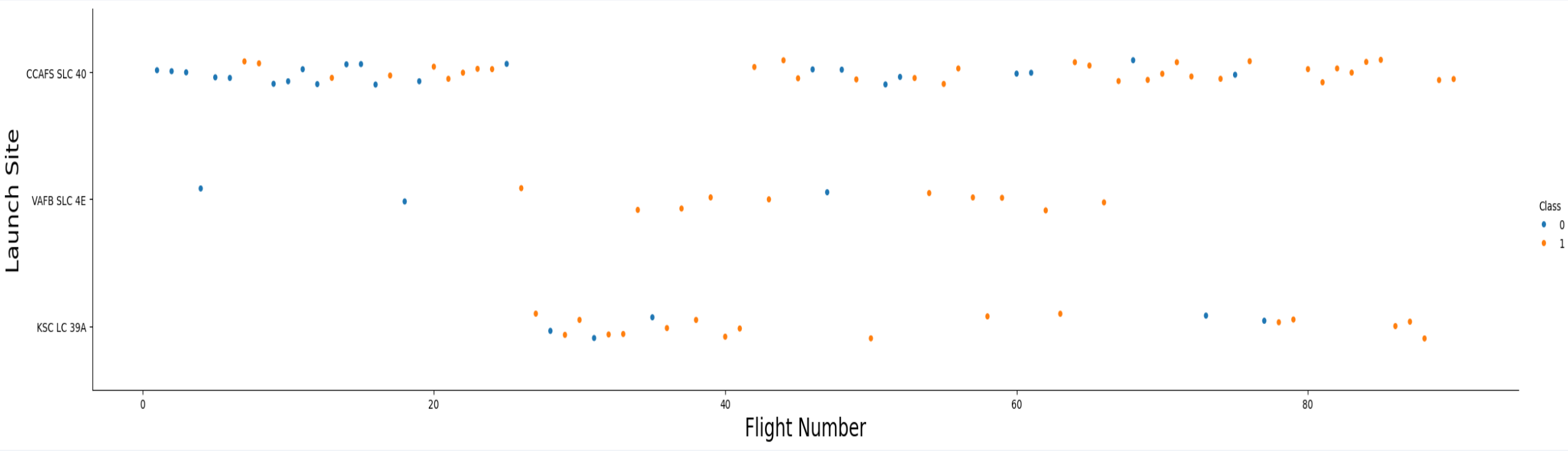
- Train Accuracy:
  - LR: 84.64%
  - SVM: 84.82%
  - Tree: 87.67%
  - KNN: 84.28%
- Test Accuracy:
  - LR: 83.33%
  - SVM: 83.33%
  - Tree: 83.33%
  - KNN: 83.33%

Section 2

Insights drawn from EDA

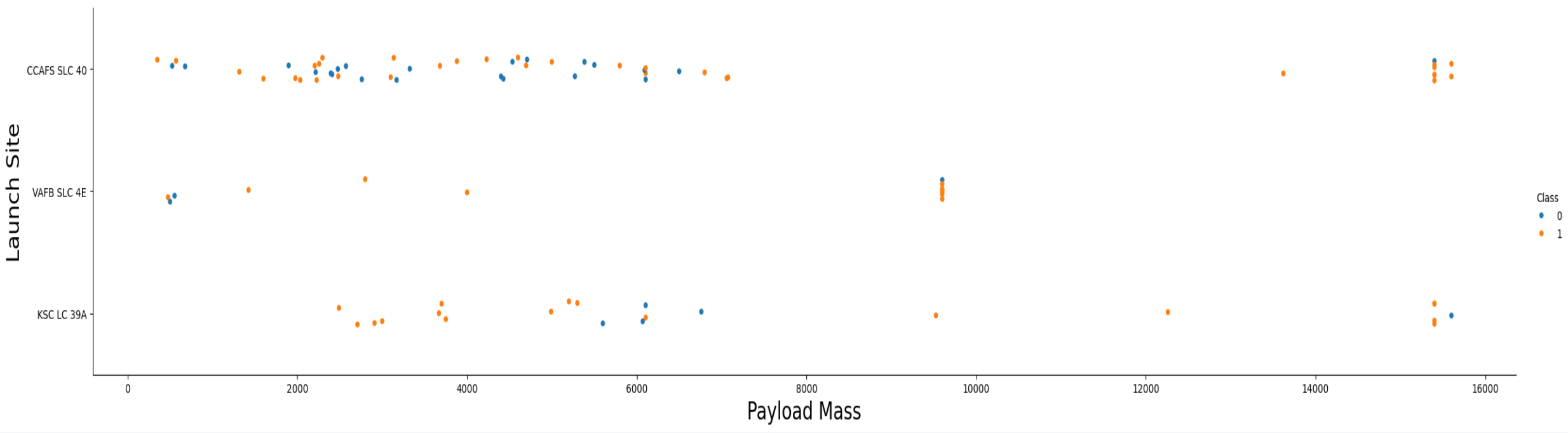


Flight Number vs. Launch Site



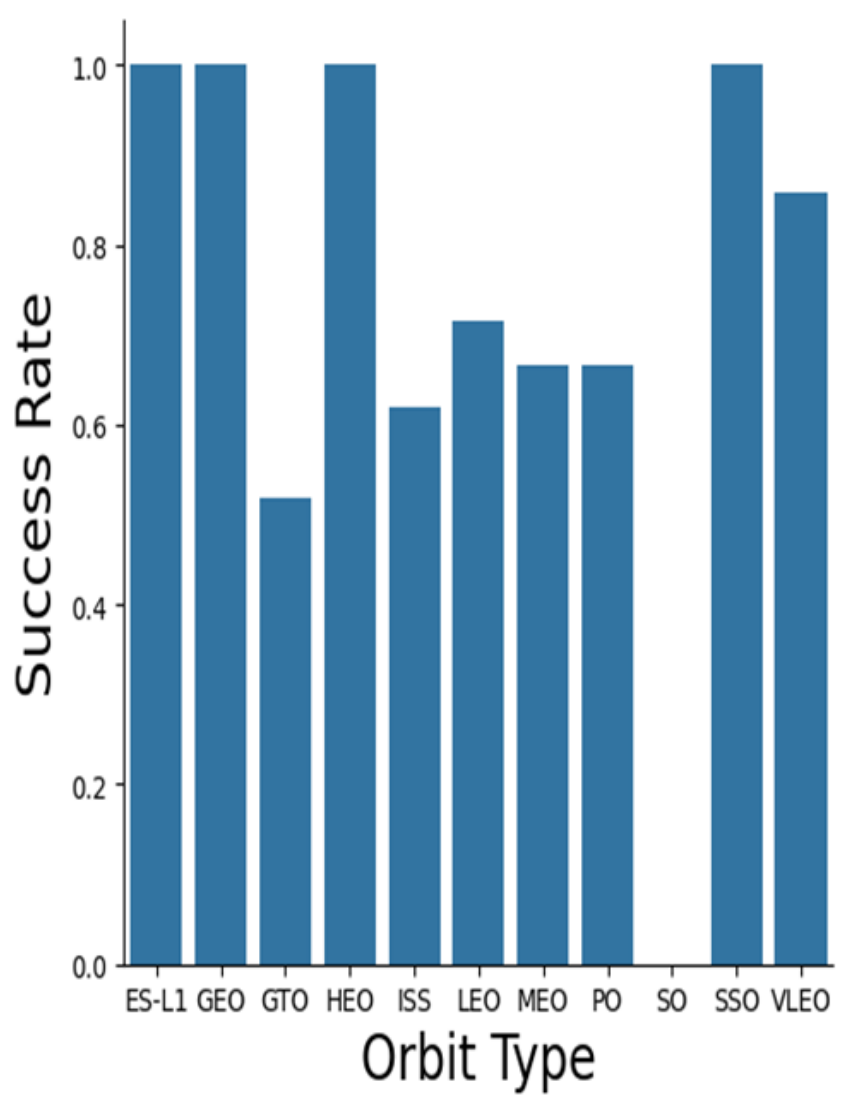
- Blue marks failure, orange success
- Most of these failures appear to be with the CCAFS SLC 40

Payload vs. Launch Site



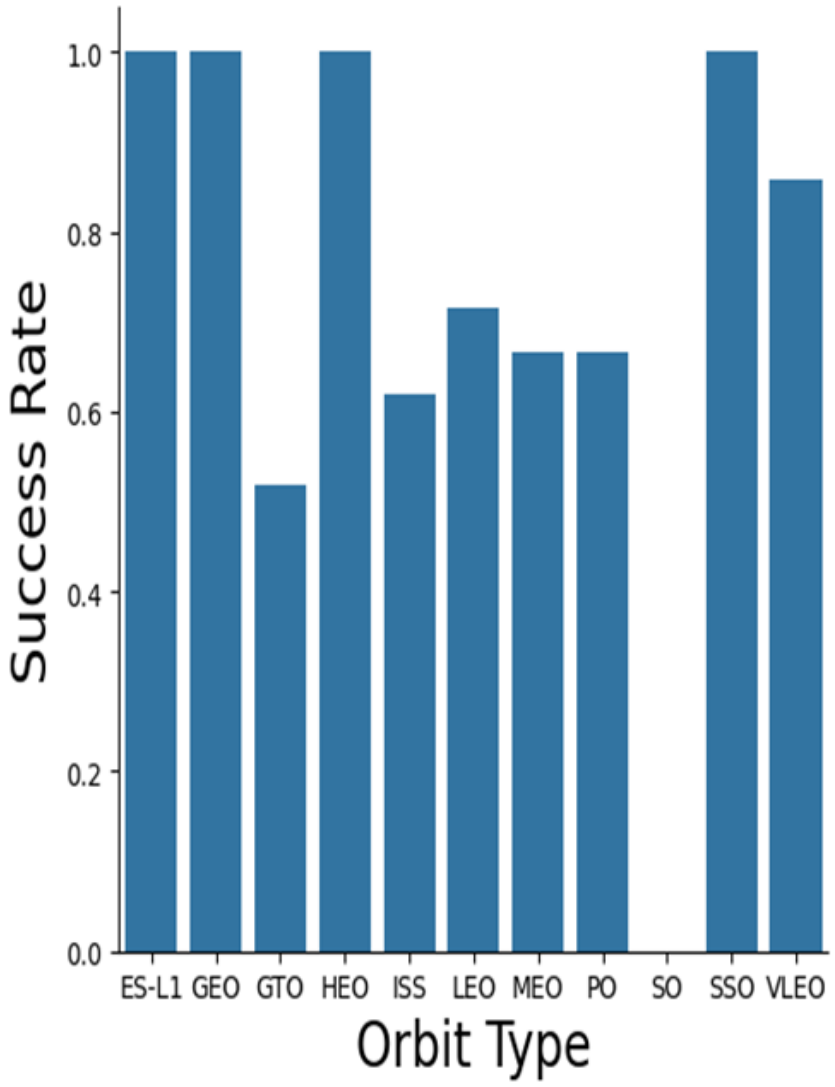
- Most frequent payload mass is between 0-6000
- CCAFS SLC 40, KSC LC 39A used for heavier payloads instead of WFB SLC 4E

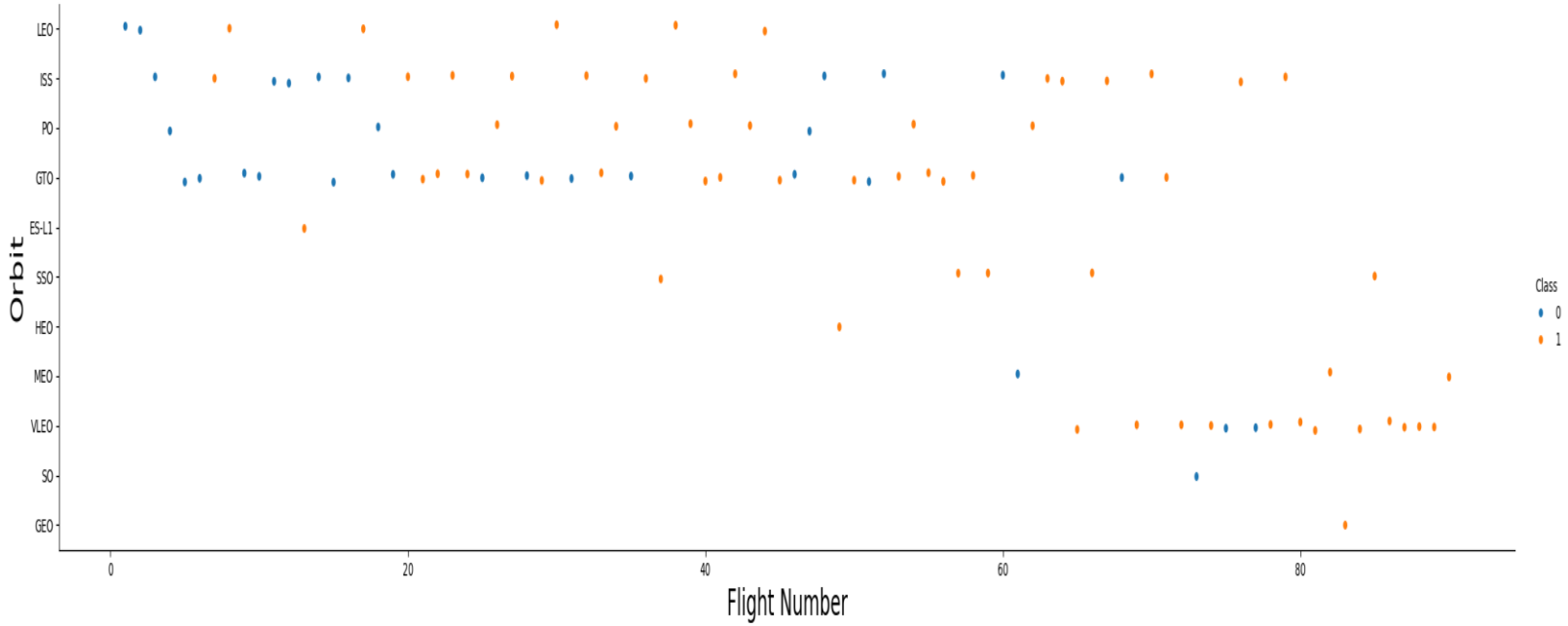
Success Rate vs. Orbit Type



- SO missions have a 0% success rate
- ES-L1, GEO, HEO, SSO all have 100% success rates

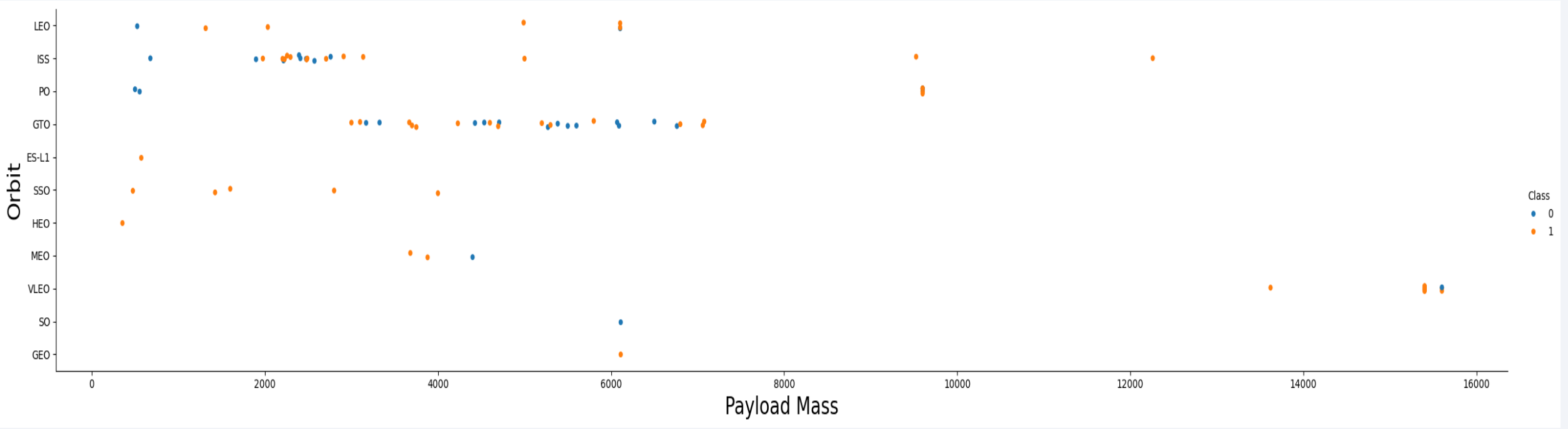
Flight Number vs. Orbit Type





- HEO, MEO, VLEO, SO, GEO tend to only be used by higher flight numbers
- SO missions have a 0% success rate. But there has only been 1
- ES-L1, GEO, HEO, SSO all have 100% success rates, but fewer missions

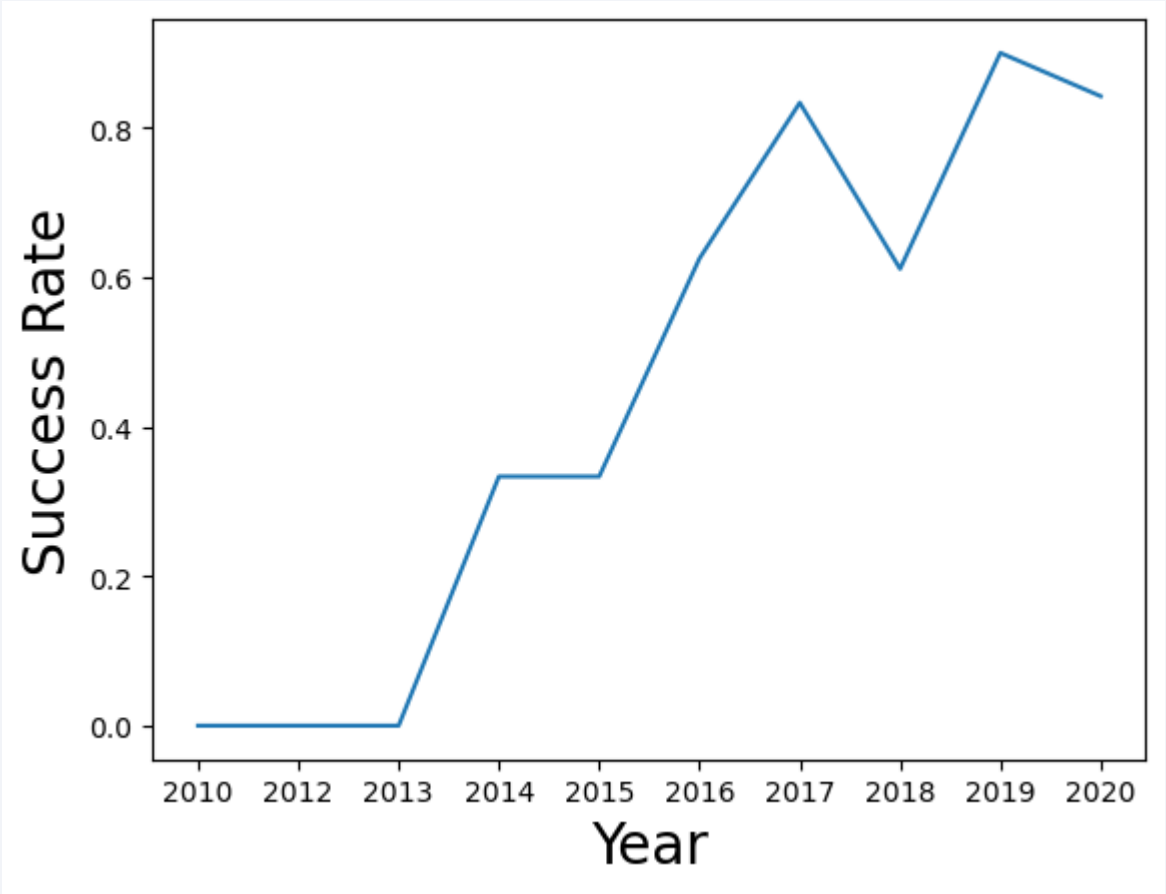
Payload vs. Orbit Type



- VLEO tends to be used by those with the highest payload mass
- Most mission failures are with GTO, or SS orbits

Launch Success Yearly Trend

- As time goes on, we typically see an increase in mission success rate
- Dips in 2018 and 2020



All Launch Site Names

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

- We have 4 distinct launch sites which are shown in the table

Launch Site Names Begin with 'CCA'

```
%sql select * from spacextable where Launch_site like 'CCA%' limit(5)
```

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

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

- Here we have found 5 launches where the launch site begins with CCA

Total Payload Mass

```
#45596
```

```
%sql select sum(PAYLOAD_MASS__KG_), Customer from spacextable group by Customer
```

12055	NASA (CCD)
12530	NASA (CCDev)
12500	NASA (CCP)
525	NASA (COTS)
0	NASA (COTS) NRO
45596	NASA (CRS)
2617	NASA (CRS), Kacific 1
12050	NASA (CTS)
362	NASA (LSP)
553	NASA (LSP) NOAA CNES
1192	NASA / NOAA / ESA / EUMETSAT

- For all NASA related Customers, the total payload mass is 107010

```
%sql select sum(PAYLOAD_MASS__KG_) from spacextable where Customer like '%NASA%'
```

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

sum(PAYLOAD_MASS__KG_)
107010

Average Payload Mass by F9 v1.1

- Average payload mass for the F9 v1.1 is 2928.4 kg

avg(PAYLOAD_MASS__KG_)	Booster_Version
2928.4	F9 v1.1
500.0	F9 v1.1 B1003
2216.0	F9 v1.1 B1010
4428.0	F9 v1.1 B1011
2395.0	F9 v1.1 B1012
570.0	F9 v1.1 B1013
4159.0	F9 v1.1 B1014
1898.0	F9 v1.1 B1015
4707.0	F9 v1.1 B1016
553.0	F9 v1.1 B1017
1952.0	F9 v1.1 B1018

First Successful Ground Landing Date

```
%sql select * from spacextable where date = (select min(date) from spacextable)
```

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

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mi
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	

- First successful landing was June 4, 2010

Successful Drone Ship Landing with Payload between 4000 and 6000

- 24 Booster version have a payload mass between 4000 and 6000 kg

Booster_Version	PAYLOAD_MASS__KG_
F9 B5 B1046.3	4000
F9 v1.1 B1014	4159
F9 B5 B1051.2	4200
F9 FT B1032.2	4230
F9 B5B1060.1	4311
F9 B5B1062.1	4311
F9 B5B1054	4400
F9 v1.1 B1011	4428
F9 v1.1	4535
F9 FT B1026	4600
F9 FT B1022	4696
F9 v1.1 B1016	4707
F9 B5 B1048.3	4850
F9 B4 B1040.1	4990
F9 B4 B1043.1	5000
F9 FT B1031.2	5200
F9 FT B1020	5271
F9 FT B1021.2	5300
F9 FT B1032.1	5300
F9 B5 B1047.2	5300
F9 B4 B1040.2	5384
F9 B5 B1058.2	5500
F9 FT B1030	5600
F9 B5 B1046.2	5800

Total Number of Successful and Failure Mission Outcomes

- Of the 101 attempted missions, only 1 was a failure

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

Boosters Carried Maximum Payload

- Max payload mass carried was 15600 kg
- 12 Booster versions have carried that amount

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

2015 Launch Records

- 6 missions attempted in 2015
- 2 complete failures
  - 01/10/2015
  - 04/14/2015

Date	Booster_Version	Landing_Outcome	Launch_Site
2015-01-10	F9 v1.1 B1012	Failure (drone ship)	CCAFS LC-40
2015-02-11	F9 v1.1 B1013	Controlled (ocean)	CCAFS LC-40
2015-03-02	F9 v1.1 B1014	No attempt	CCAFS LC-40
2015-04-14	F9 v1.1 B1015	Failure (drone ship)	CCAFS LC-40
2015-06-28	F9 v1.1 B1018	Precluded (drone ship)	CCAFS LC-40
2015-12-22	F9 FT B1019	Success (ground pad)	CCAFS LC-40

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

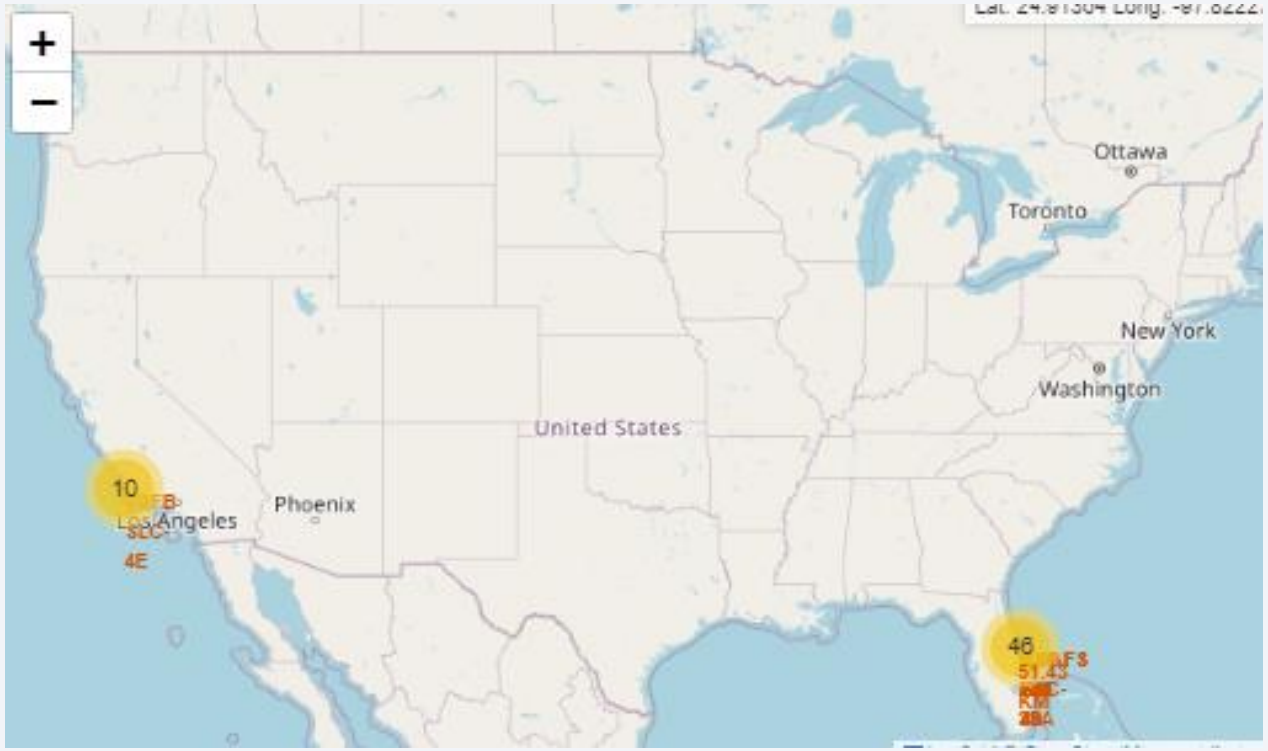
- 31 mssions in the date range 06/04/2010 to 03/20/2017
- Most were not attempted
- Only 8 were a success

Date	Landing_Outcome	count(Landing_Outcome)
2016-04-08	Success (drone ship)	5
2015-12-22	Success (ground pad)	3
2015-06-28	Precluded (drone ship)	1
2015-01-10	Failure (drone ship)	5
2014-04-18	Controlled (ocean)	3
2013-09-29	Uncontrolled (ocean)	2
2012-05-22	No attempt	10
2010-06-04	Failure (parachute)	2


Section 3
Launch Sites
Proximities Analysis

Launch Sites

- We can see that the launch sites are either in CA or FL
- Most of the launches are in FL

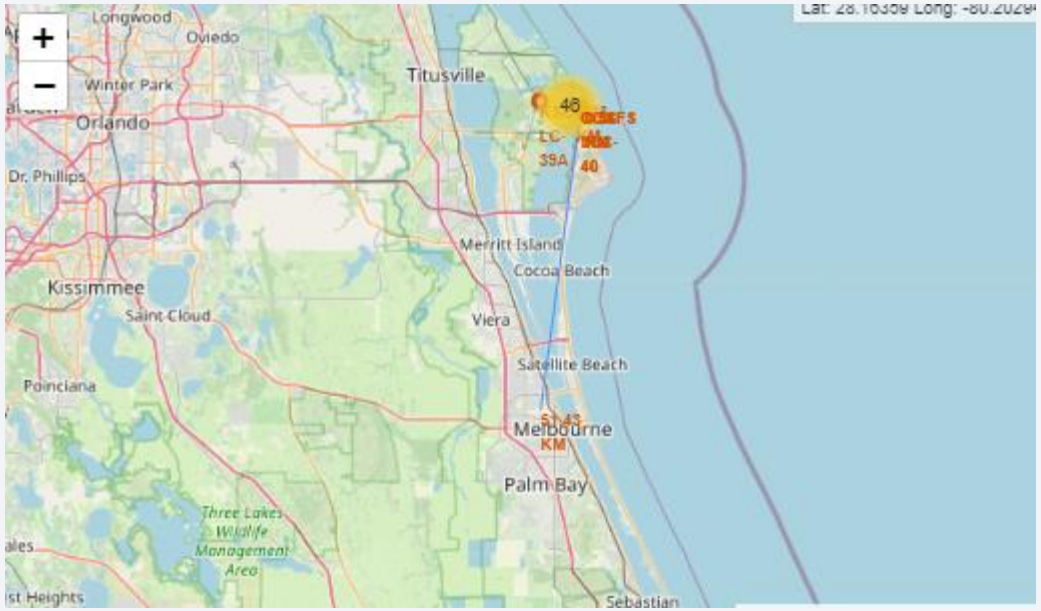


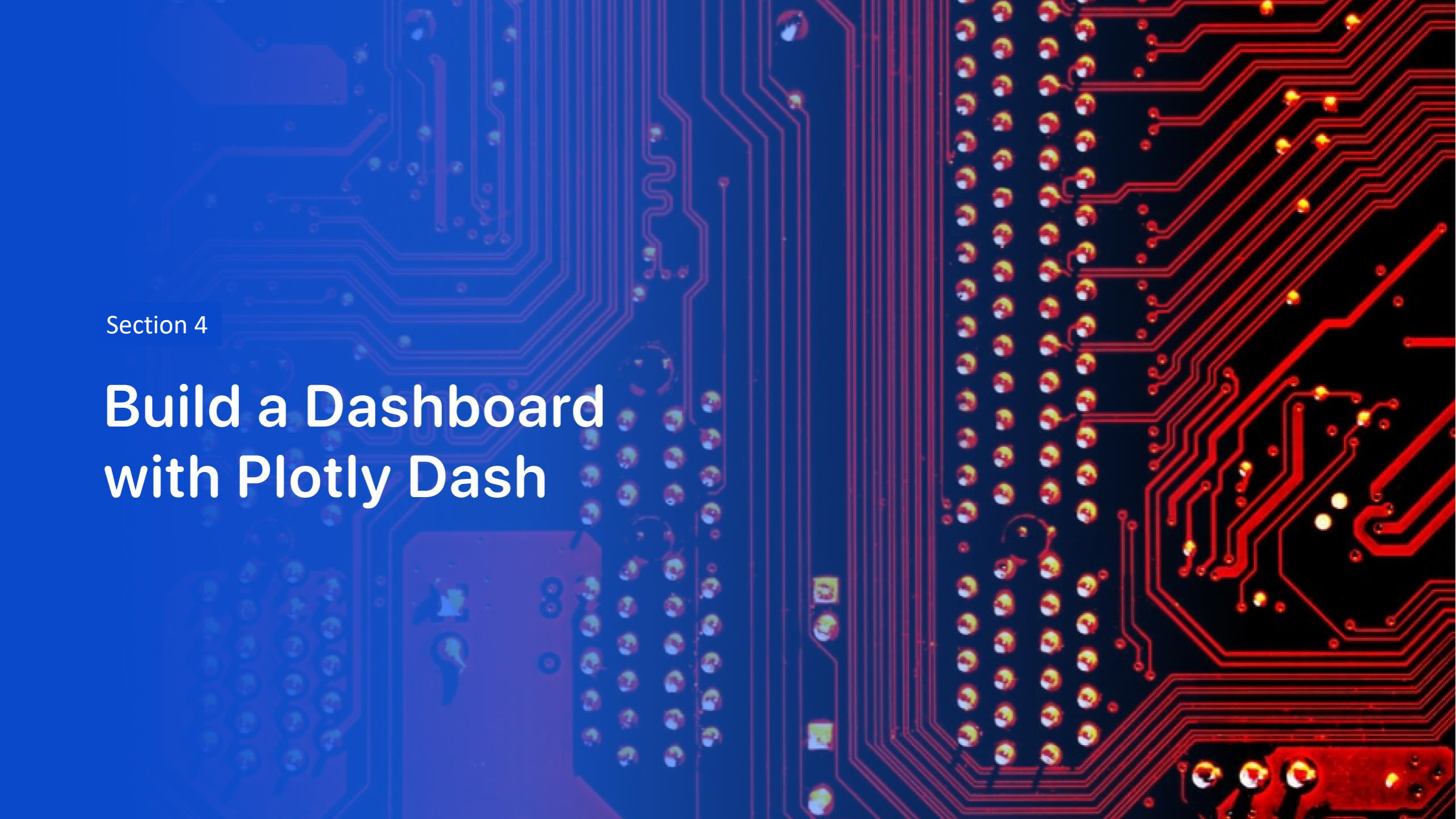
Launch Outcomes

- Labeled as green for success and red for failure/not attempted/etc.

Proximity to an Airport

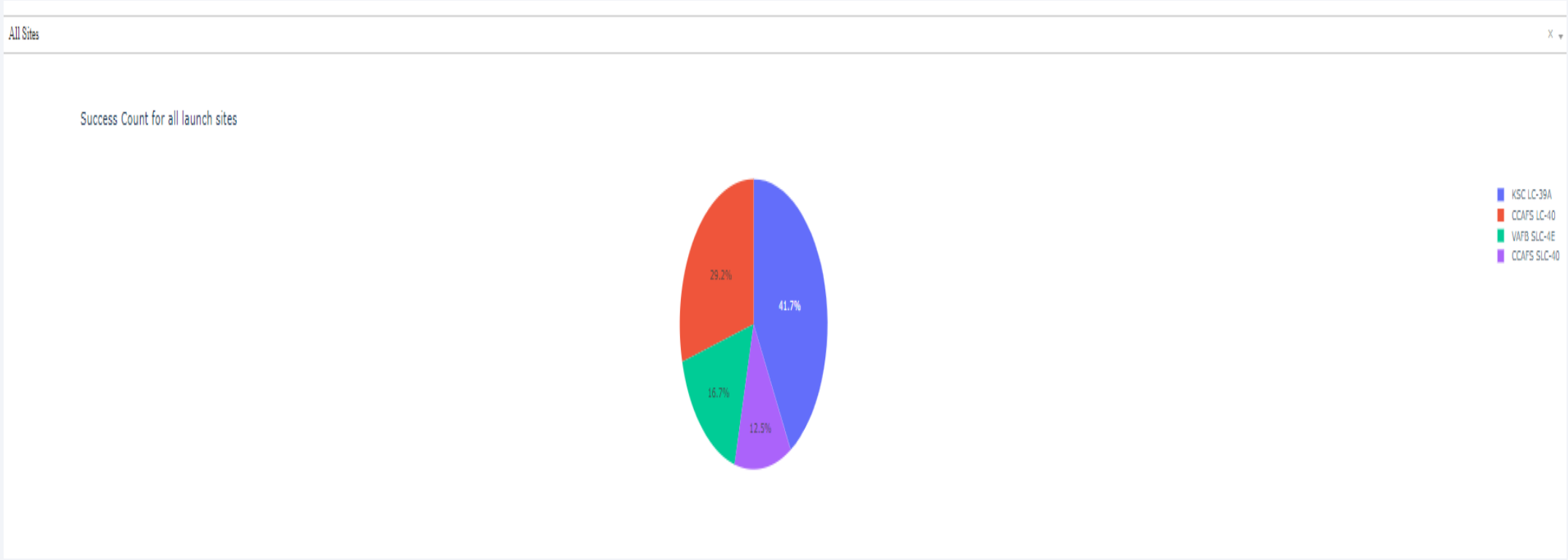
- Shows the distance to the closest airport




Section 4
Build a Dashboard with Plotly Dash

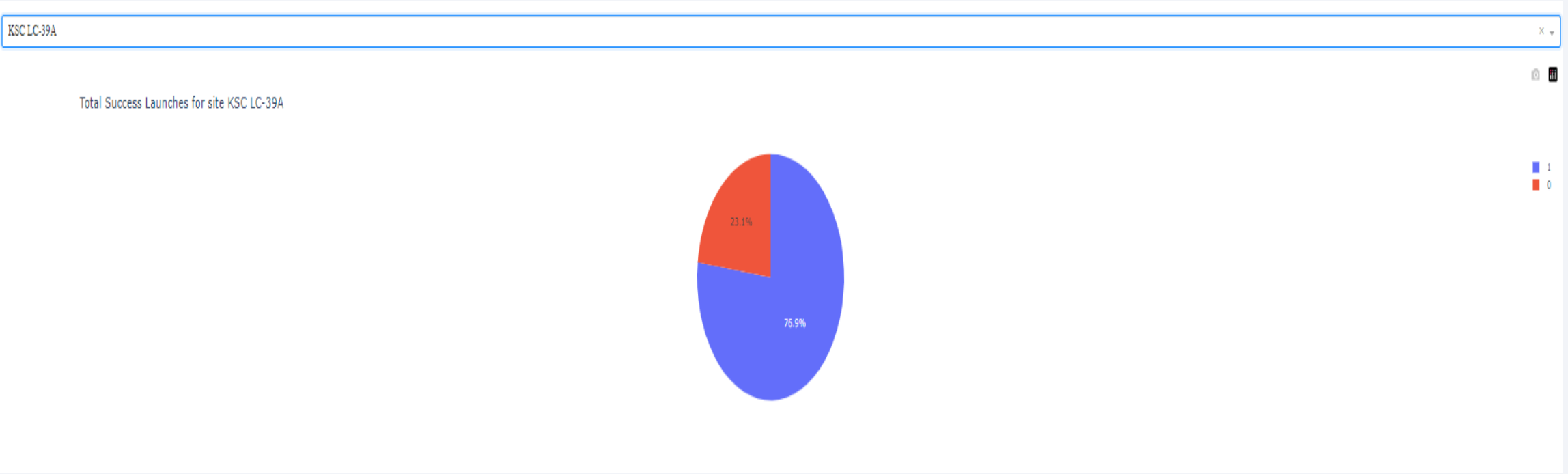
Proportion of Missions by Site

- Majority of launches at KSC LC 39A
- Fewest at CCAFS SLC-40



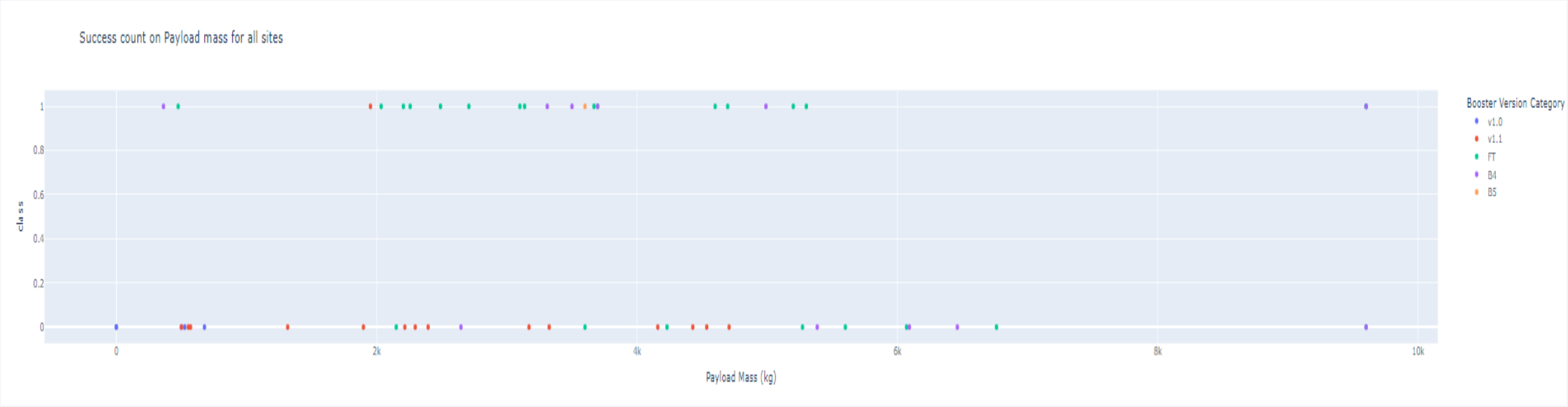
Most Successful Site

- KKSC LC 40 was the most successful site with 76.9% success rate
- This is from 13 observations



Payload vs. Launch Outcome

- Most successes lie in the range of 1952 to 5300 kg of payload mass
- Outliers higher and lower than that range
- Failures also were in that range, but had more outside that range
- Most failures v1.1

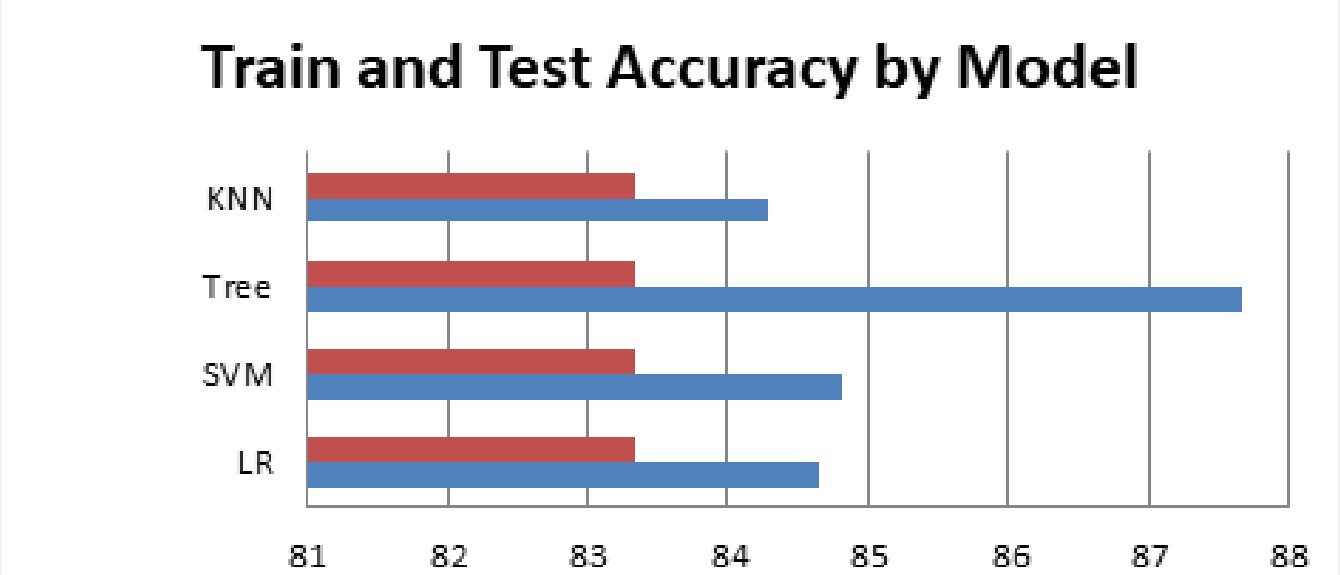


Section 5

Predictive Analysis (Classification)

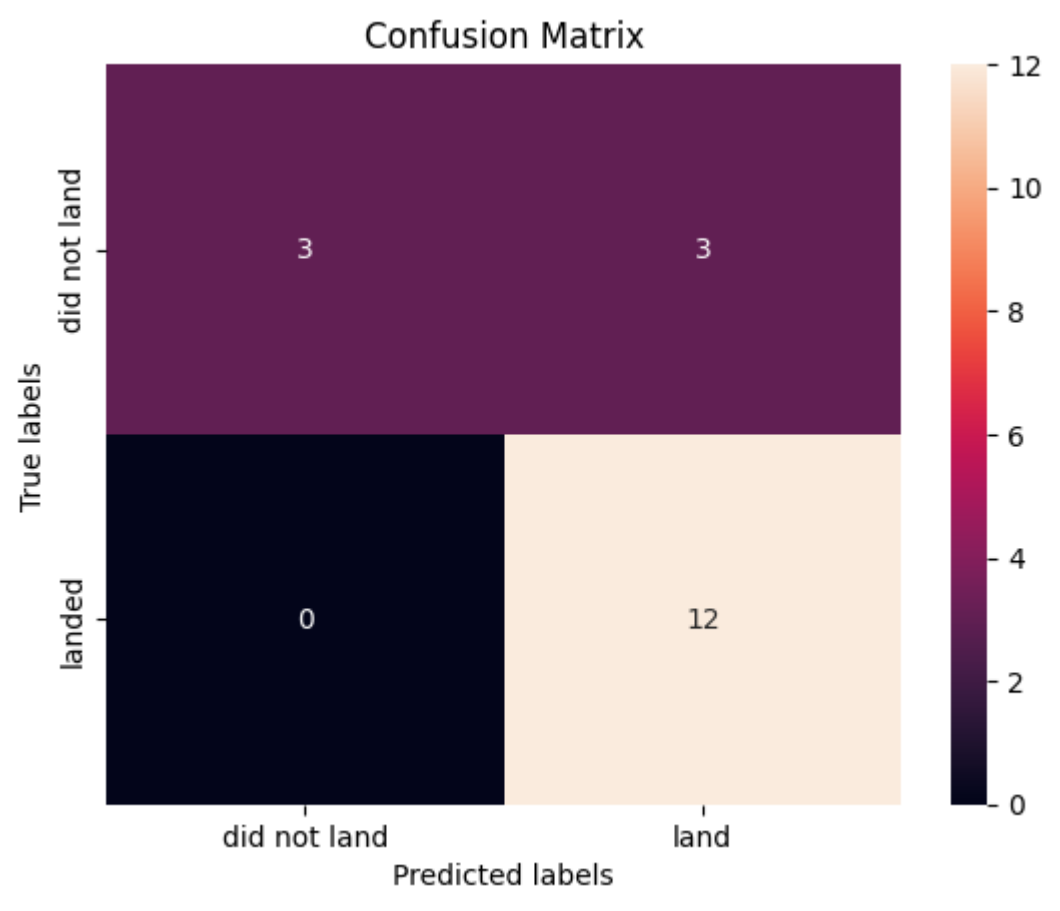
Classification Accuracy

- All the models had the same performance on the test data
- Will say we prefer the KNN model since it performed the lowest on the train data so it likely has the lowest change of having overfitting



	LR	SVM	Tree	KNN
Test Accuracy	83.33	83.33	83.33	83.33
Train Accuracy	84.64	84.82	87.67	84.28

Confusion Matrix



- KNN Confusion Matrix
- Accuracy = (TP + TN) / (TP + TN + FP + FN)
- = (12+3)/(12+3+3+0)
- = 83.33%

Conclusions

- All our models predict with 83.33% accuracy when predicting a failure/success of a launch
- Most failures were of a v1.1
- KSC LC 40 has the most attempts, successes, and success rate of missions
- No perfect booster version, some better than others, but unless it’s only been used once, there isn’t one without any failures

Appendix

GitHub Repository

Thank you!