IBM Developer
SKILLS NETWORK

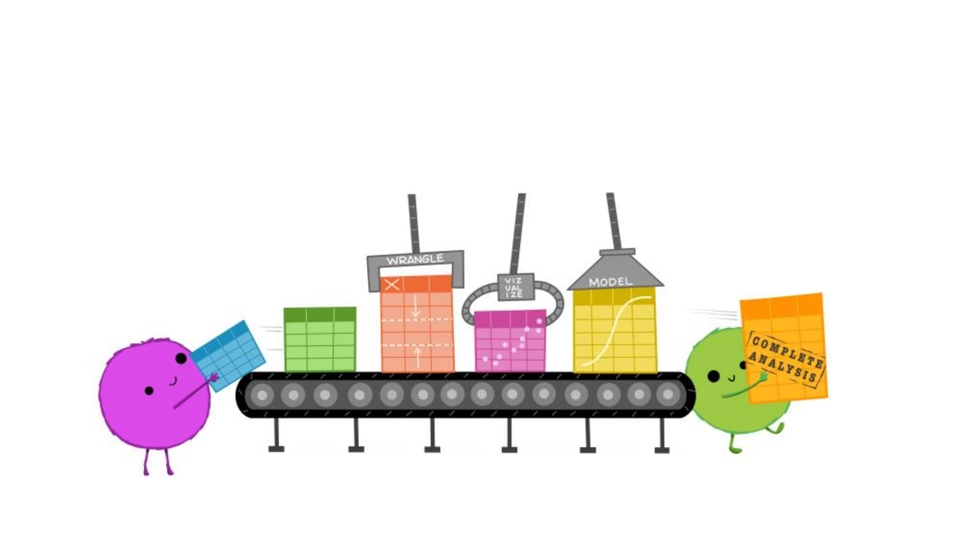
Winning Space Race with Data Science

Tommy Chen
26 April 2024

Outline

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

Executive Summary



- General improvement with launch successes over time.
- Dashboard built to monitor progress.
- Model trained to help predict what the outcome of future launches would be.
- Overall improvements made so that future launches will have the best setup to succeed

Introduction

- Rocket launches are very expensive $50m - $162m
- Tesla has designed reusable rockets that can cut the cost significantly
- Important to be able to understand whether launch will succeed or fail
- Also need to understand the factors contributing to success or failure of a launch

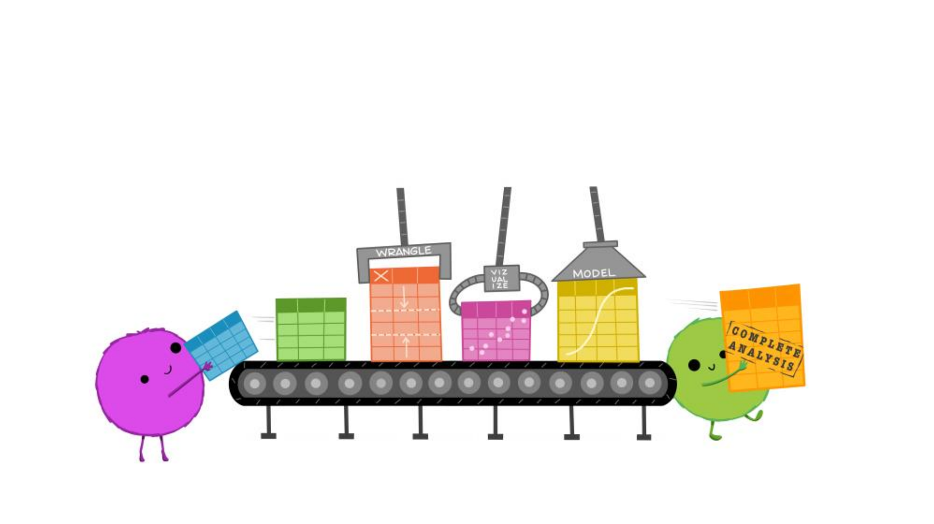
Section 1

Methodology

Methodology

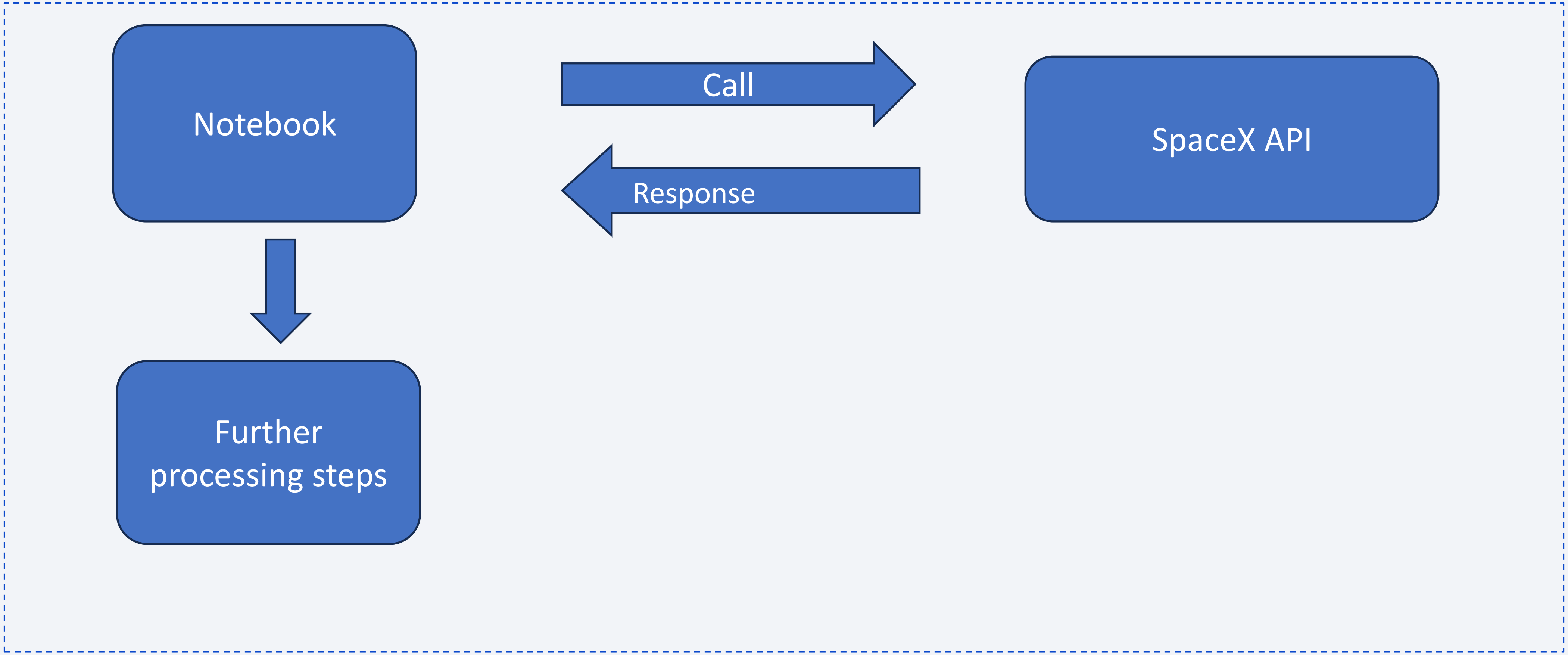
Executive Summary

- Data collection methodology:
  - Data scraped from SpaceX API
- Perform data wrangling
  - Datasets structured and cleaned using pandas
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Grid Search with different models used to determine the best fitting model



Data Collection

Data Collection – SpaceX API



https://github.com/tommychen741/Applied-Data-Science-Capstone/blob/main/1_jupyter-labs-spacex-data-collection-api.ipynb

Data Collection - Scraping

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)

Extraction of column headings

Extraction of data values

https://github.com/tommychen741/Applied-Data-Science-Capstone/blob/main/1_jupyter-labs-spacex-data-collection-api.ipynb

Data Wrangling

- Dataset restricted to falcon 9 data
- Data types made to be consistent
- Non sensical rows dropped
- Missing values were replaced with the average
- Different outcomes normalized to succeed vs fail

https://github.com/tommychen741/Applied-Data-Science-Capstone/blob/main/3_labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- Scatter plot to show relationship between different variables
- Bar plot to show differences by category
- Line plot to show trends over time

https://github.com/tommychen741/Applied-Data-Science-Capstone/blob/main/4_edadataviz.ipynb

EDA with SQL

- SQL queries done:
  - Launch sites
  - Total payload mass
  - Average payload mass for booster version F9 v1.1
  - First successful ground pad landing
  - Different booster versions that were successful on drone with payload of between 4000 and 6000
  - Landing outcomes between 2010 and 2017
  - …and more

https://github.com/tommychen741/Applied-Data-Science-Capstone/blob/main/4_edadataviz.ipynb

Build an Interactive Map with Folium

- Markers and circles used to signify landing sites
- Marker clusters for each mission at landing site created
- Successful launches were shown as green
- Failed launches shown as red

https://github.com/tommychen741/Applied-Data-Science-Capstone/blob/main/6_lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- Dashboard allows users to select launch site and see success rate
- User can also select payload ranges they want to view
- Pie chart shows success rate
- Scatter plot shows success rate across different payloads and booster versions

https://github.com/tommychen741/Applied-Data-Science-Capstone/blob/main/7_Dashboard_code.py

Predictive Analysis (Classification)

- Following models were tested:
  - Logistic Regression
  - SVM
  - Decision Tree
  - K Nearest Neighbours
- Models assessed on:
  - Confusion matrix
  - Accuracy

https://github.com/tommychen741/Applied-Data-Science-Capstone/blob/main/8_SpaceX_Machine%20Learning%20Prediction.ipynb

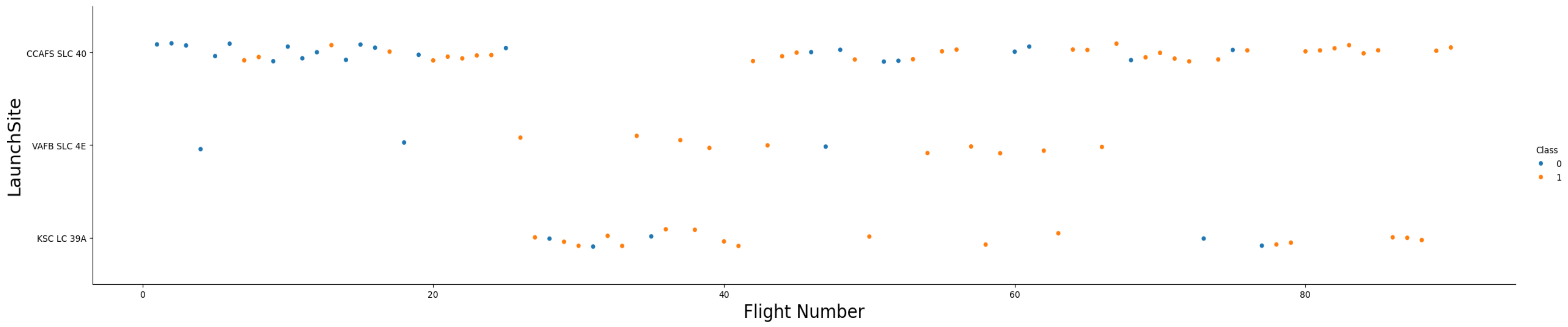
Results

- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

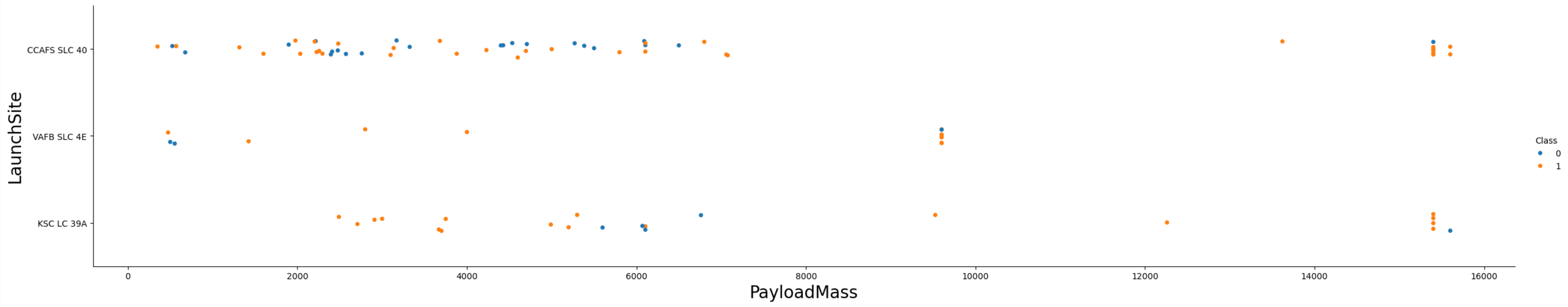
Insights drawn from EDA

Flight Number vs. Launch Site



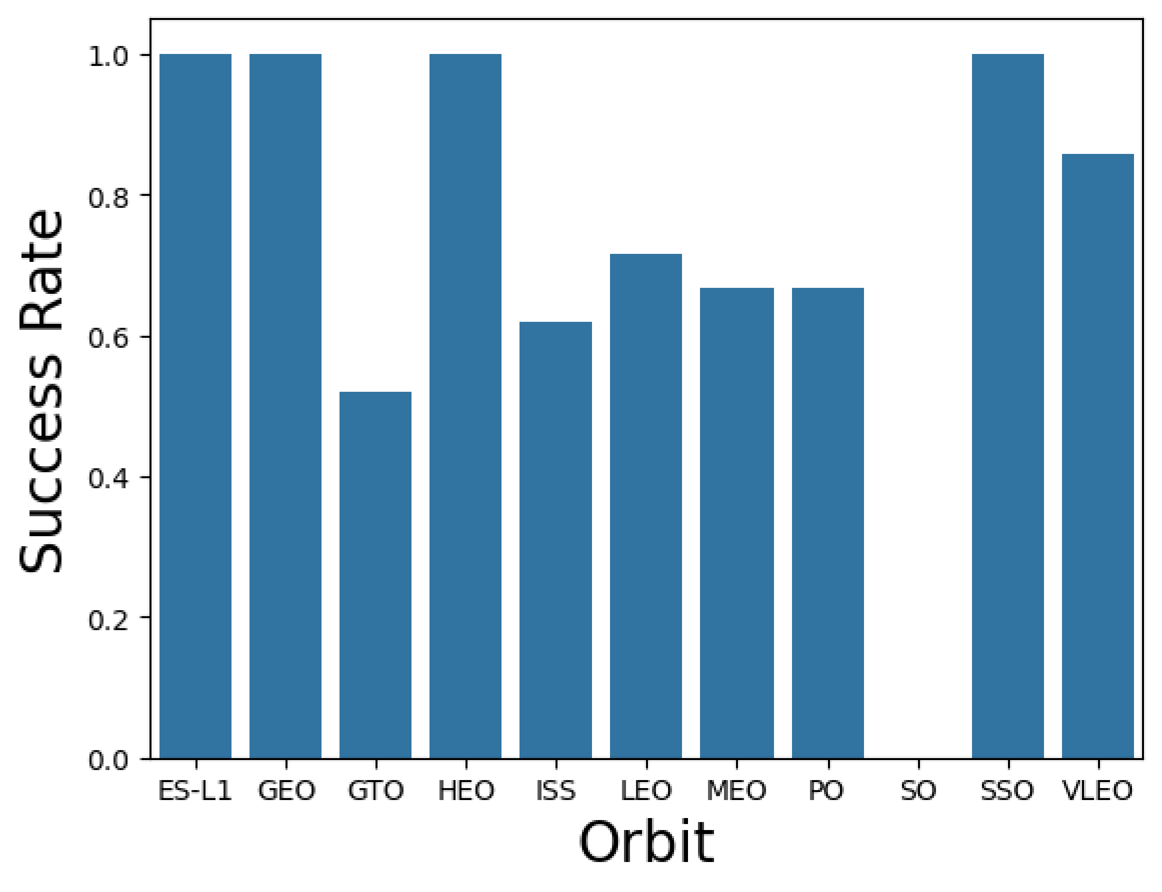
Improvement across all launch sites with more recent launches more likely to succeed

Payload vs. Launch Site



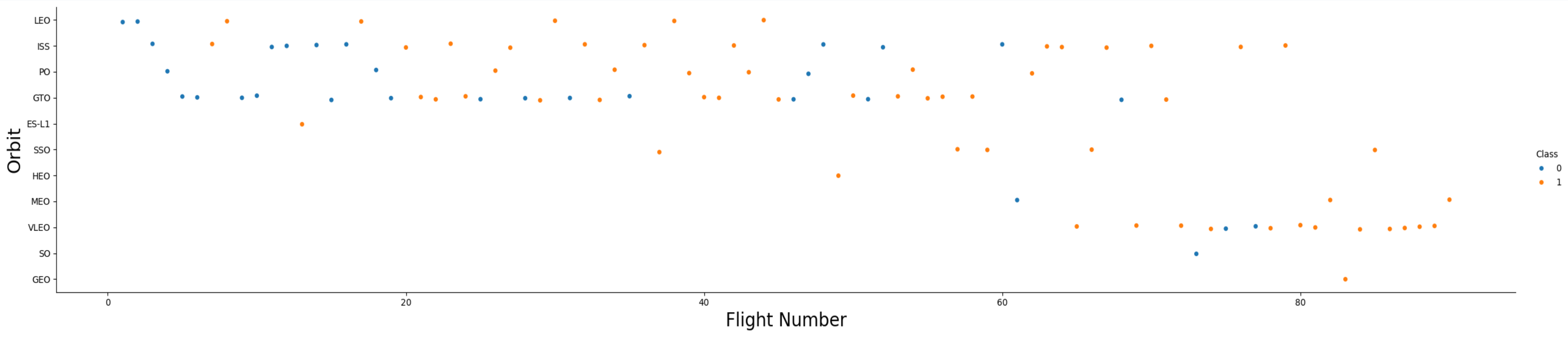
VAFB SLC 4E not doing any launches where payload is greater than 10,000

Success Rate vs. Orbit Type



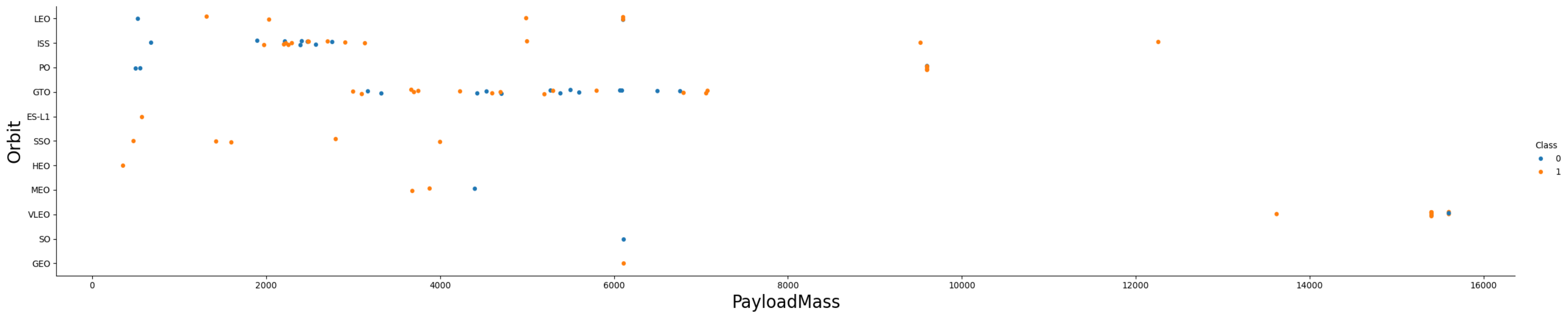
1. ES-L1, GEO, HEO and SSO all have 100% success rates
2. SO site has not succeeded yet

Flight Number vs. Orbit Type



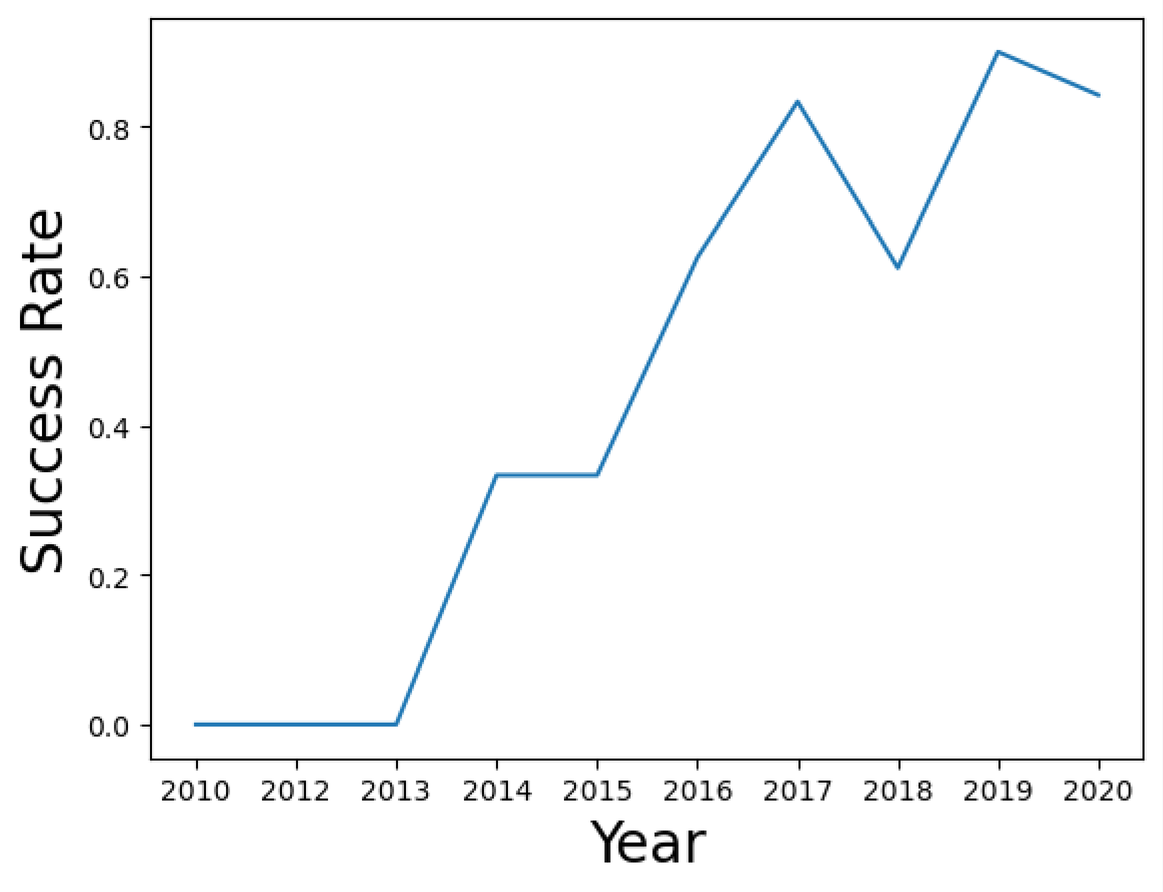
1. Improving success rate across all orbits
2. Most recent launches at VLEO

Payload vs. Orbit Type



1. Heavier payloads done for ISS, PO and VLEO
2. Low payloads were from initial flights

Launch Success Yearly Trend



Increasing success rate over time

All Launch Site Names

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

List of different launch sites in analysis

Launch Site Names Begin with 'CCA'

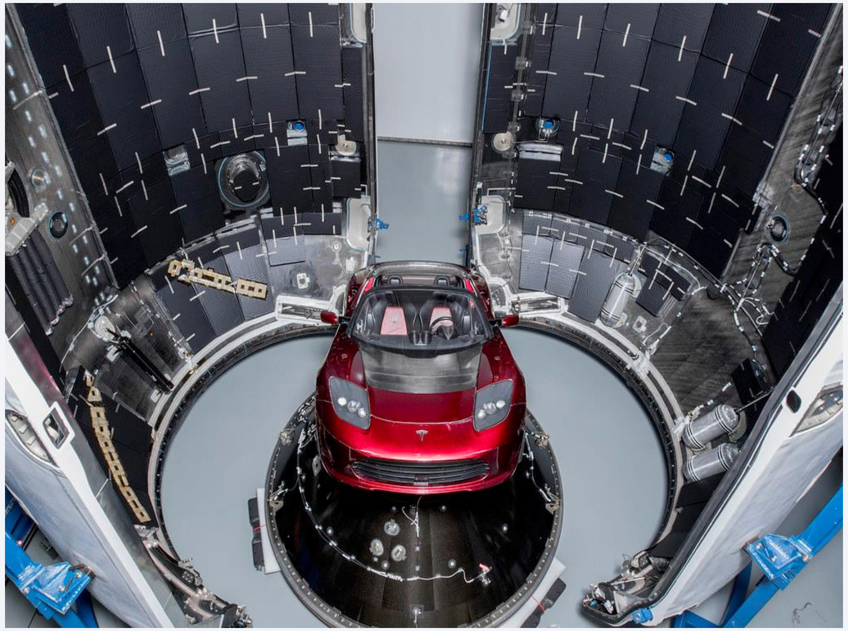
5 records where launch sites begin with `CCA`

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

total payload carried by boosters from NASA

45,596



Average Payload Mass by F9 v1.1

Average payload mass carried
by booster version F9 v1.1:

2,928.4

First Successful Ground Landing Date

Date of the first successful landing outcome on ground pad:

2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

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

Names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

Total Number of Successful and Failure Mission Outcomes

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

Total number of successful and failure mission outcomes

Boosters Carried Maximum Payload

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

Names of the booster which have carried the maximum payload mass

2015 Launch Records

Failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

Month	Landing_Outcome	Booster_Version	Launch_Site
01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

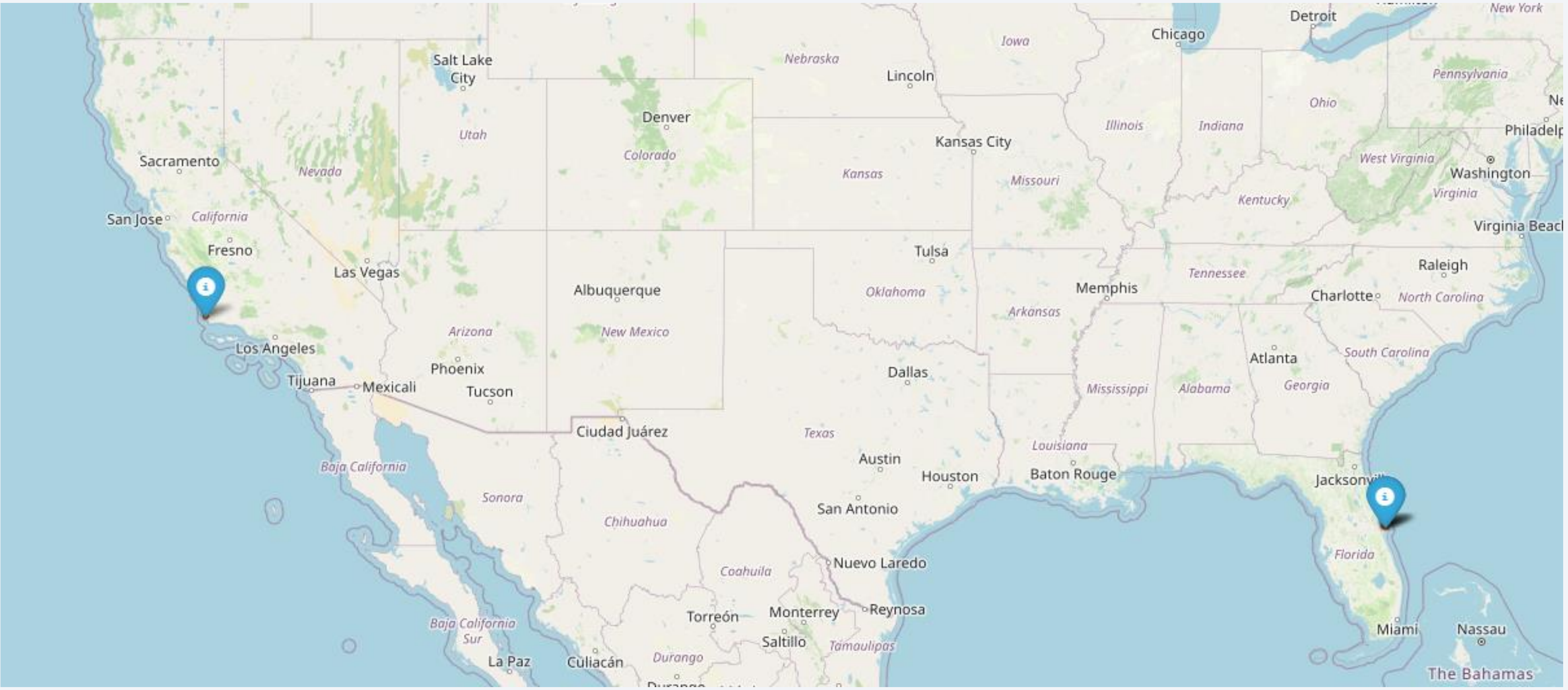
Landing_Outcome	missions
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

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

Section 3

Launch Sites Proximities Analysis

Folium map with launch sites

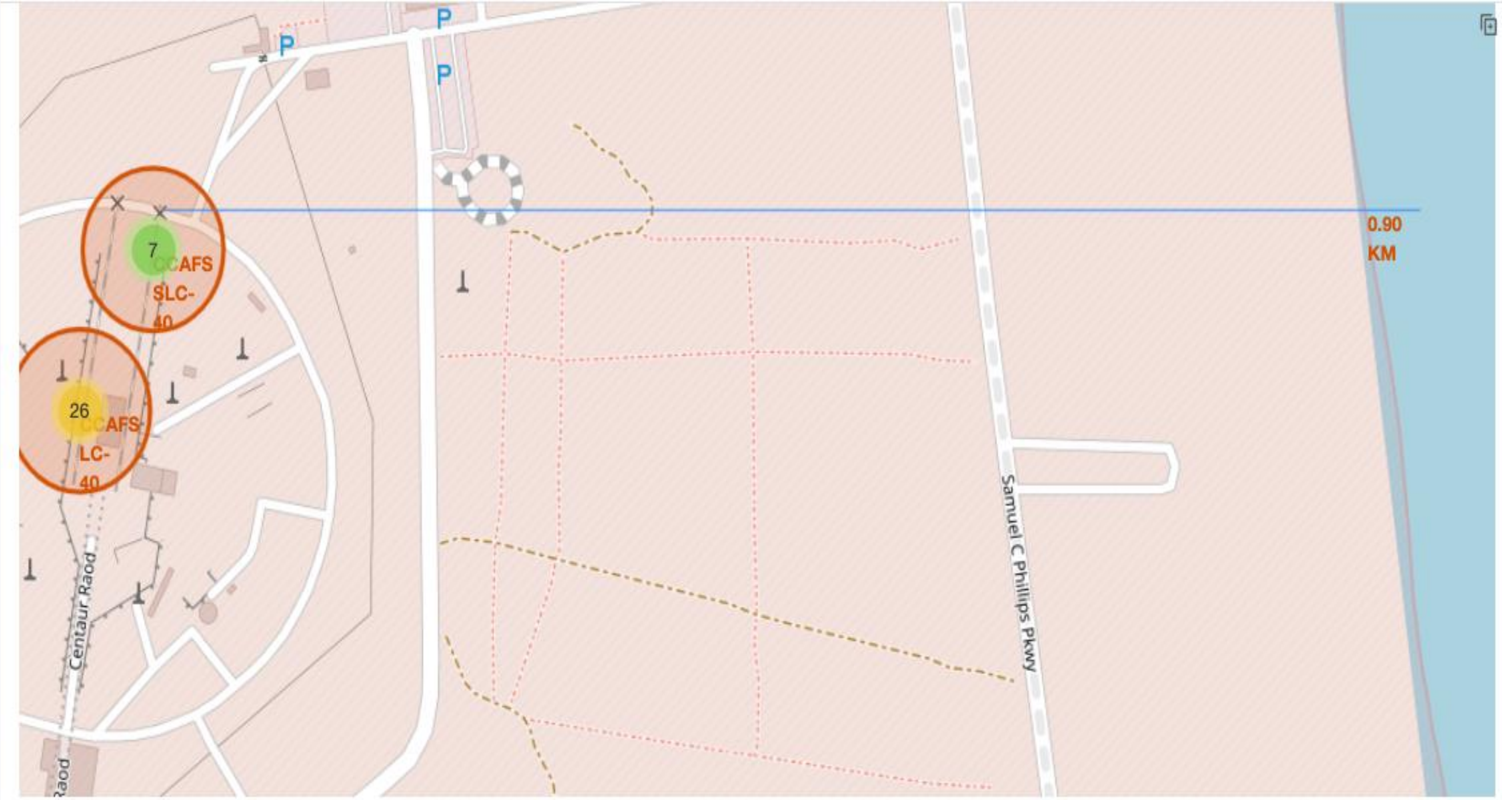


This shows where the different launch sites are distributed

Launch site mission breakdown

Shows successful and failed launches at each site

Distance to other locations

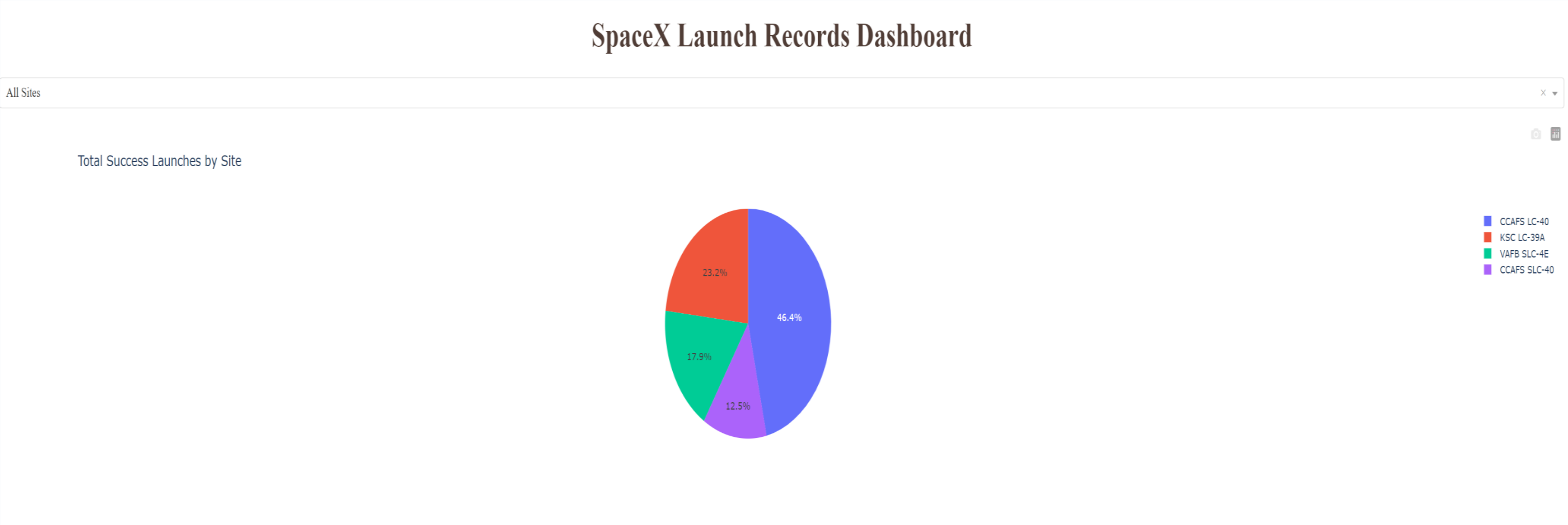


Distance to coastline

Section 4

Build a Dashboard with Plotly Dash

Launch successes across all sites

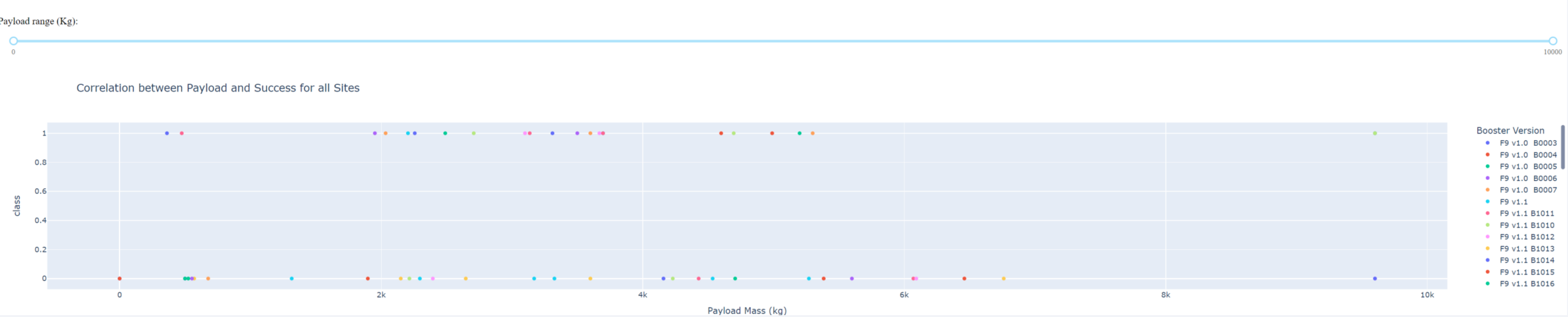


Most launches come from CCAFS LC-40

Most successful site

Most successful launches at KSC LC-39A

Payload vs Launch outcome

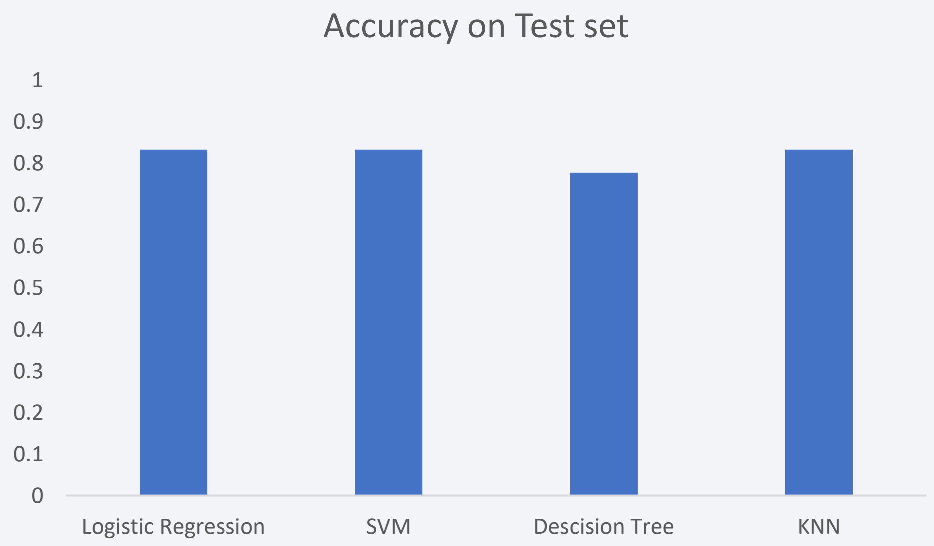


- User able to select different payloads and see success rates for different booster versions

Section 5

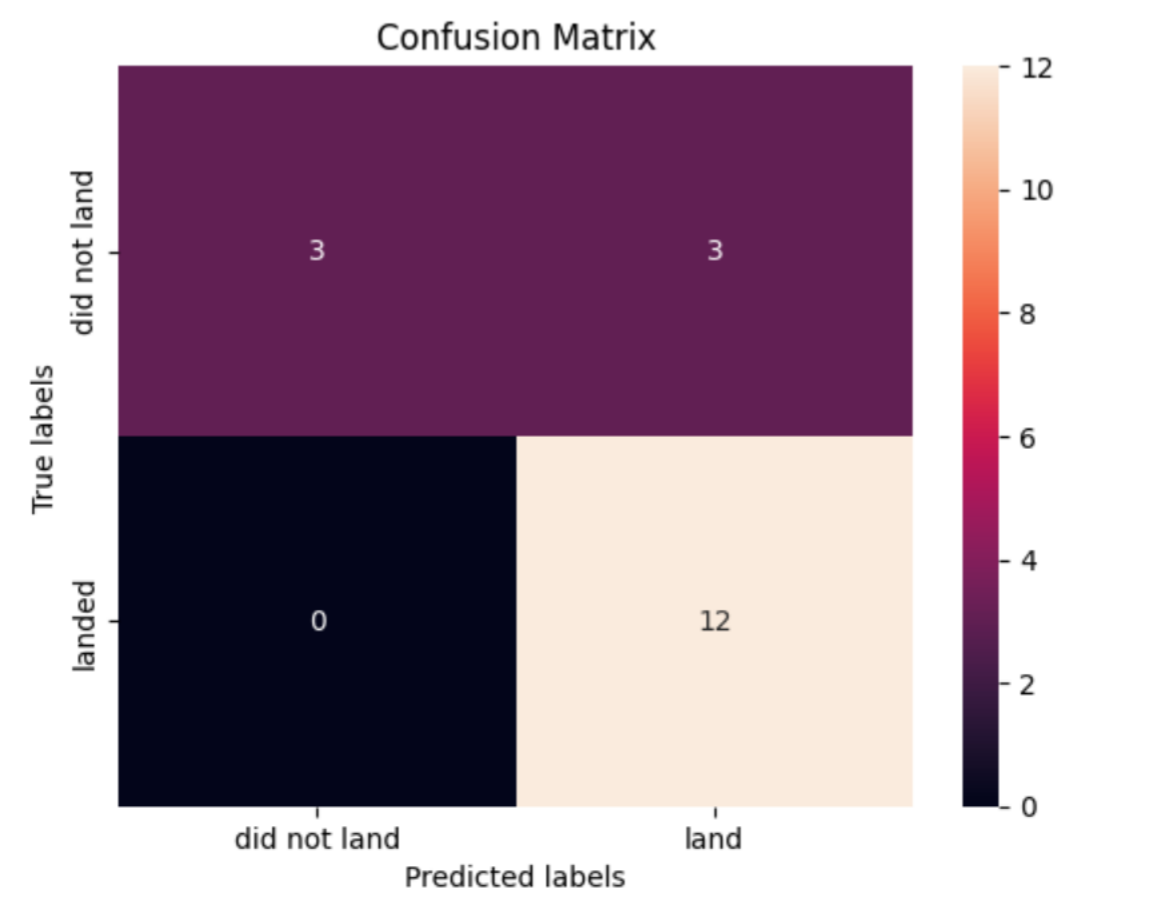
Predictive Analysis (Classification)

Classification Accuracy



All models performed comparably apart from decision tree model

Confusion Matrix



- 3 out of 18 cases with a wrong prediction
- Rest were predicted correctly (5/18)

Conclusions

- General improvement with launch successes over time.
- Dashboard built to monitor progress.
- Model trained to help predict what the outcome of future launches would be.
- Overall improvements made so that future launches will have the best setup to succeed

Thank you!