IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

P. Fields
May 25, 2022

Outline

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

Executive Summary

- The goal of this project was to utilize information pertaining to SpaceX launches to determine whether it was feasible to predict whether a launch would be successful.
- 11 features, including payload mass and launch site among others were examined from 90 launches.
  - Data was collected as publicly available from SpaceX.
- Multiple machine learning approaches were applied included SVM, Logisitic regression, KNN and decision trees. All methods yielded comparable performance and had an 83.3% accuracy on predicting outcome in the test cohort.

Introduction

- SpaceX currently offers launches using their Falcon 9 rocket for 62 million dollars which is considerably cheaper than competitors. This is due to their reusable first stage rocket.
- We are interested in whether we develop a competitor to SpaceX, however it is very important to understand what makes a successful launch.
- The goal of this project is to investigate key attributes of SpaceX launches to determine correlations and whether we can ultimately predict successful launches.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was obtained as publicly available from SpaceX including on wikipedia:

    https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922
- Perform data wrangling
  - Data was analyzed using python v3.9. 11 key attributes were considered.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection

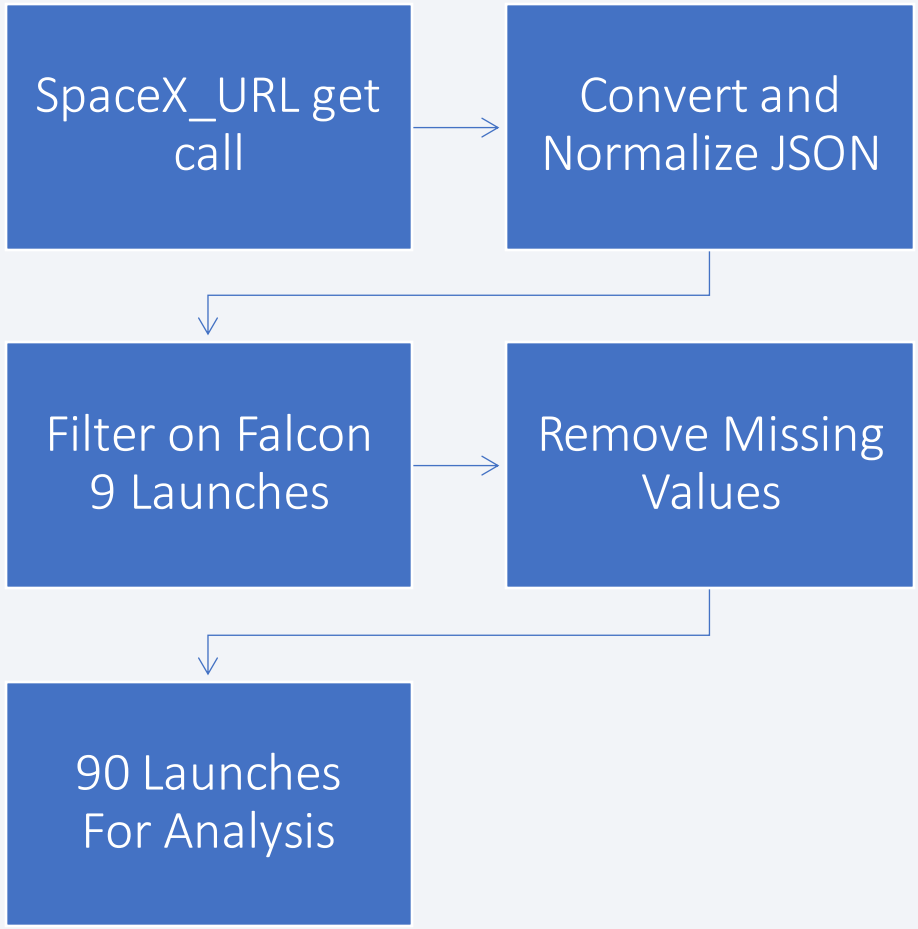
- Data was collected as publicly available on Wikipedia and Space X.
- Data was handled using python notebooks in python v3.9
- Data was stored in csv format as intermediate files to preserve data analysis checkpoints.

Data Collection – SpaceX API

- Key attributes of launches using the Falcon 9 rocket were collected from the SpaceX website using a REST API call and data wrangling.

- Data available at: https://github.com/pfields8/ibm_data_science/blob/main/Data_Collection_NB.ipynb

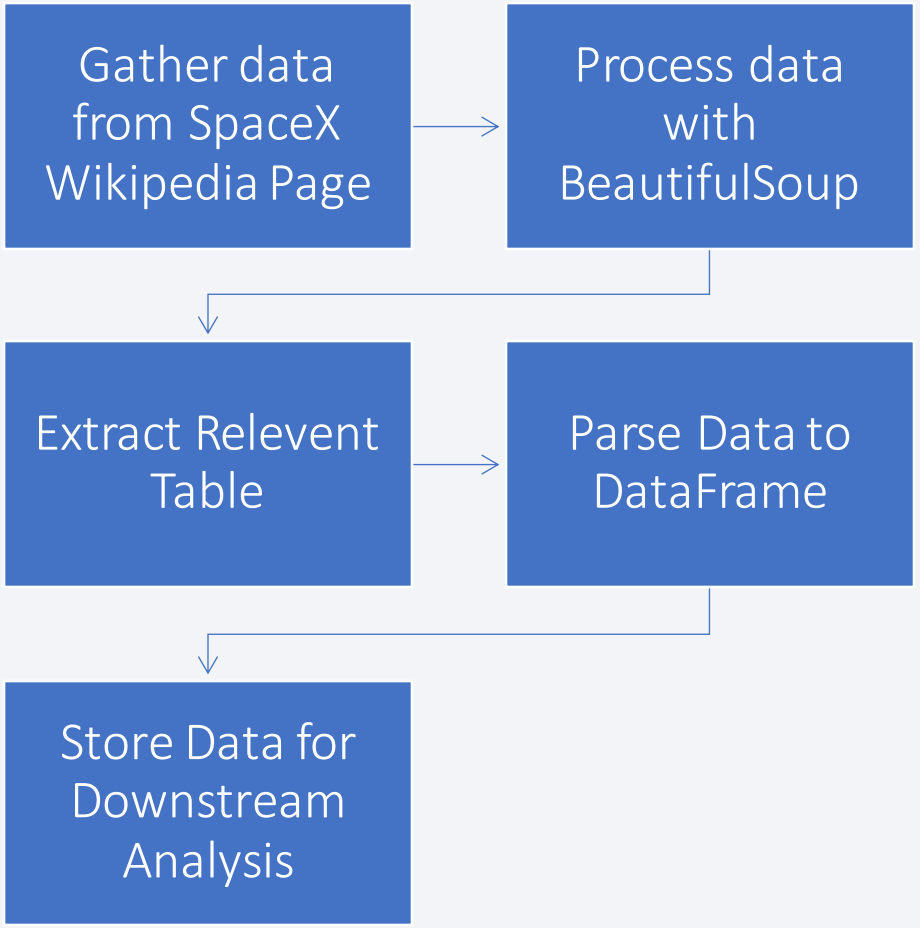
Flow Chart of Collection



Data Collection - Scraping

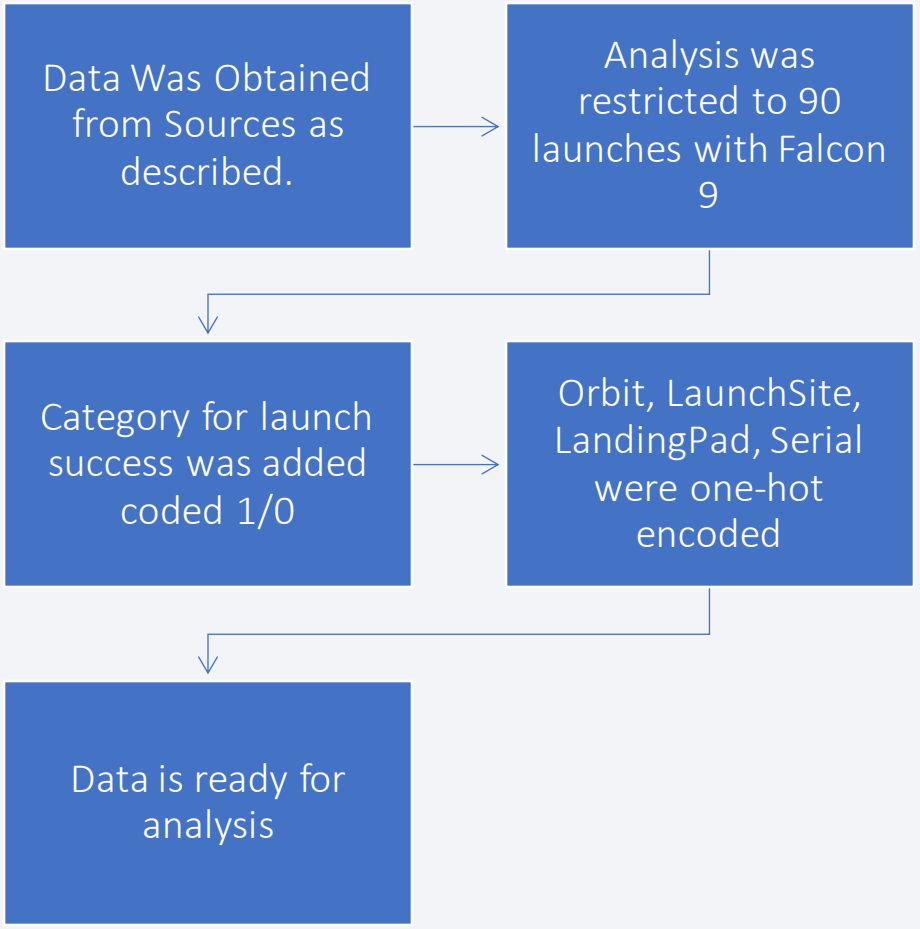
- Additional data was gained by scraping for SpaceX launch wikipedia page.

- Data available at: https://github.com/pfields8/ibm_data_science/blob/main/Web_Scraping_NB.ipynb



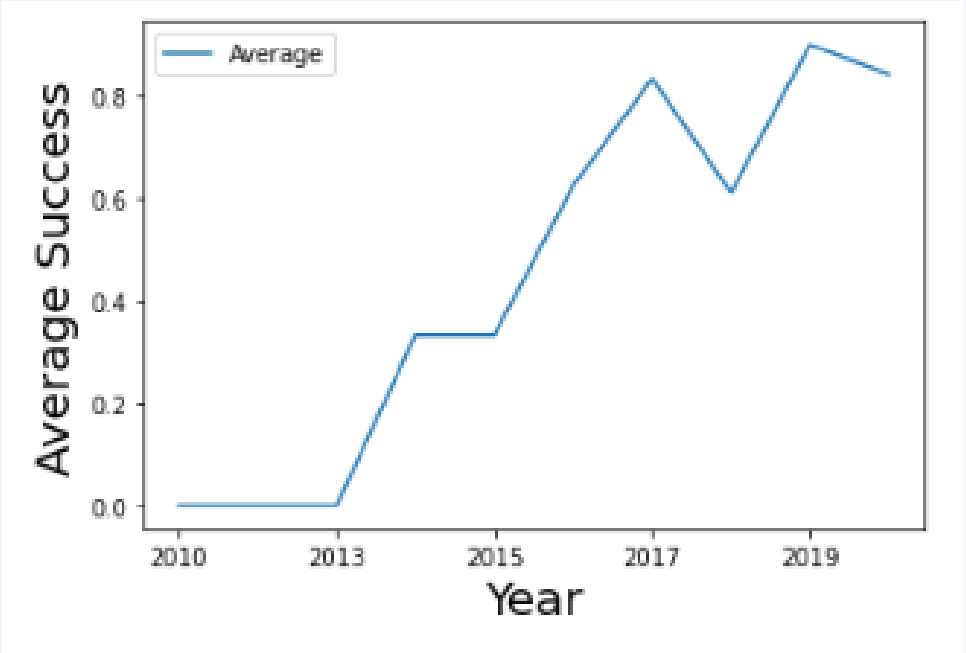
Data Wrangling

- Data was wrangled to optimize attributes and instances to allow for analysis.
- Variables with multiple values such as Orbit were one-hot encoded.
- Data available at: https://github.com/pfields8/ibm_data_science/blob/main/EDA.ipynb



EDA with Data Visualization

- Launch success was analyzed in relation to payload mass, flight number, orbit, and launch site among others.
- Many attributes influence the success rate, such as certain orbits like ES-L1 and GEO have success rates of 100%, while GTO has a success rate <60%.
- One key attribute is time. More recent launches have been much more successful, see figure.
- Data available at: https://github.com/pfields8/ibm_data_science/blob/main/EDA%20with%20Data%20Visualization.ipynb



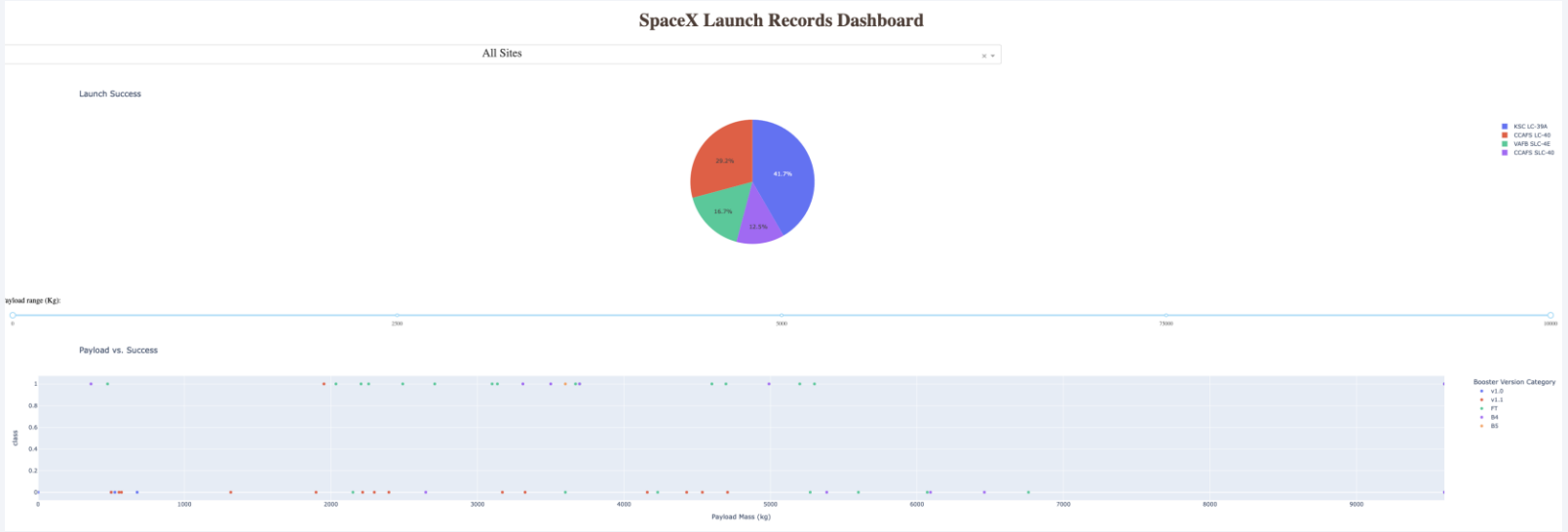
EDA with SQL

- SQL was used as an additional source to query data.
- SpaceX data was stored in a DB2 database which provides a relational database to preserve the data.
- We can identify many attributes in the data, for example the most common landing outcome is No attempt, and the max payload for any launch is 15,600 kg.
- Data available at: https://github.com/pfields8/ibm_data_science/blob/main/EDA%20with%20SQL.ipynb

Build an Interactive Map with Folium

- Folium was utilized to provide an integrated map to assess the location and successes of the various launches.
- 4 launch sites were used by SpaceX. 3 in close proximity in Florida and one in California. All four sites are located in close proximity to the coast.
- Marker clustering allows for the maps to be cleaner while still being able to interactively assess the success for each site.
- Data available at: https://github.com/pfields8/ibm_data_science/blob/main/Interactive%20Visual%20Analytics%20with%20Folium%20lab.ipynb

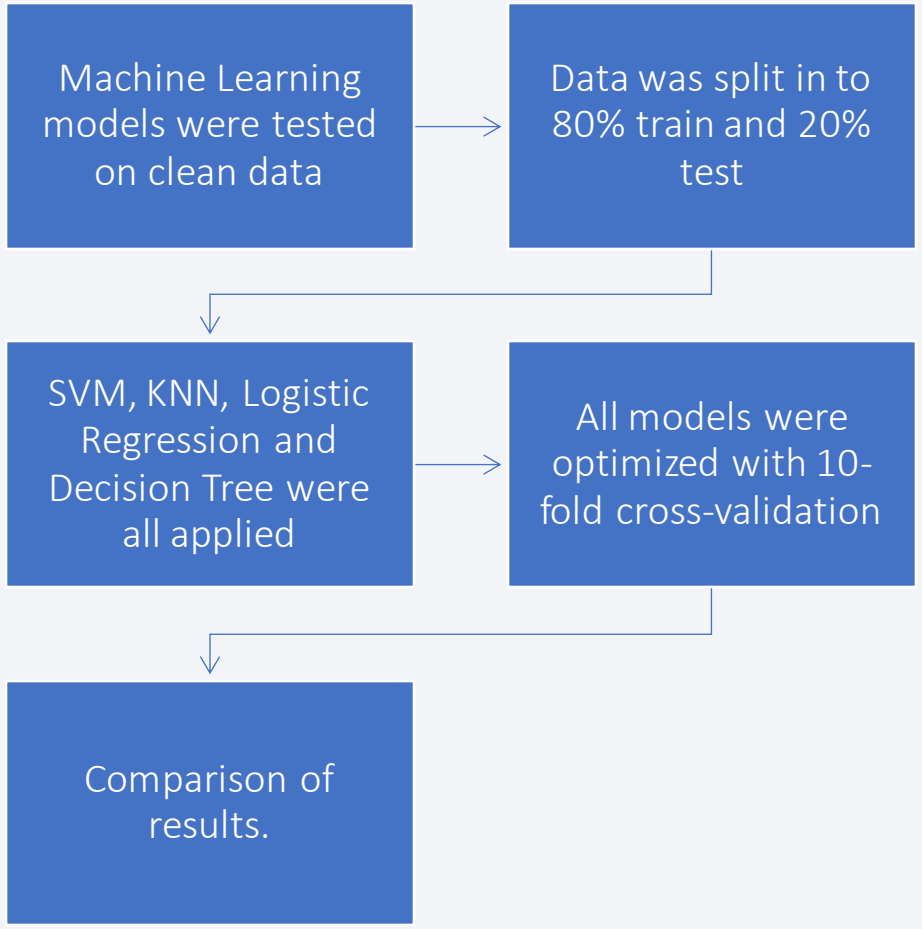
Build a Dashboard with Plotly Dash



- We wanted to provide an interactive assessment of launch success.
- Two key attributes are the launch site and the booster rocket.
- The plotly dashboard allows for location selection to assess success, while also plotting the correlation based on booster rocket
- Code available at: https://github.com/pfields8/ibm_data_science/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)

- 4 different machine learning models were tested to compare the ability to test the ability to predict successful launches.
- While optimization varied within the training, using the optimized hyper parameters all methods performed comparably, with an **83% success rate in the test set.**
- Data available at: https://github.com/pfields8/ibm_data_science/blob/main/Machine%20Learning%20Prediction.ipynb

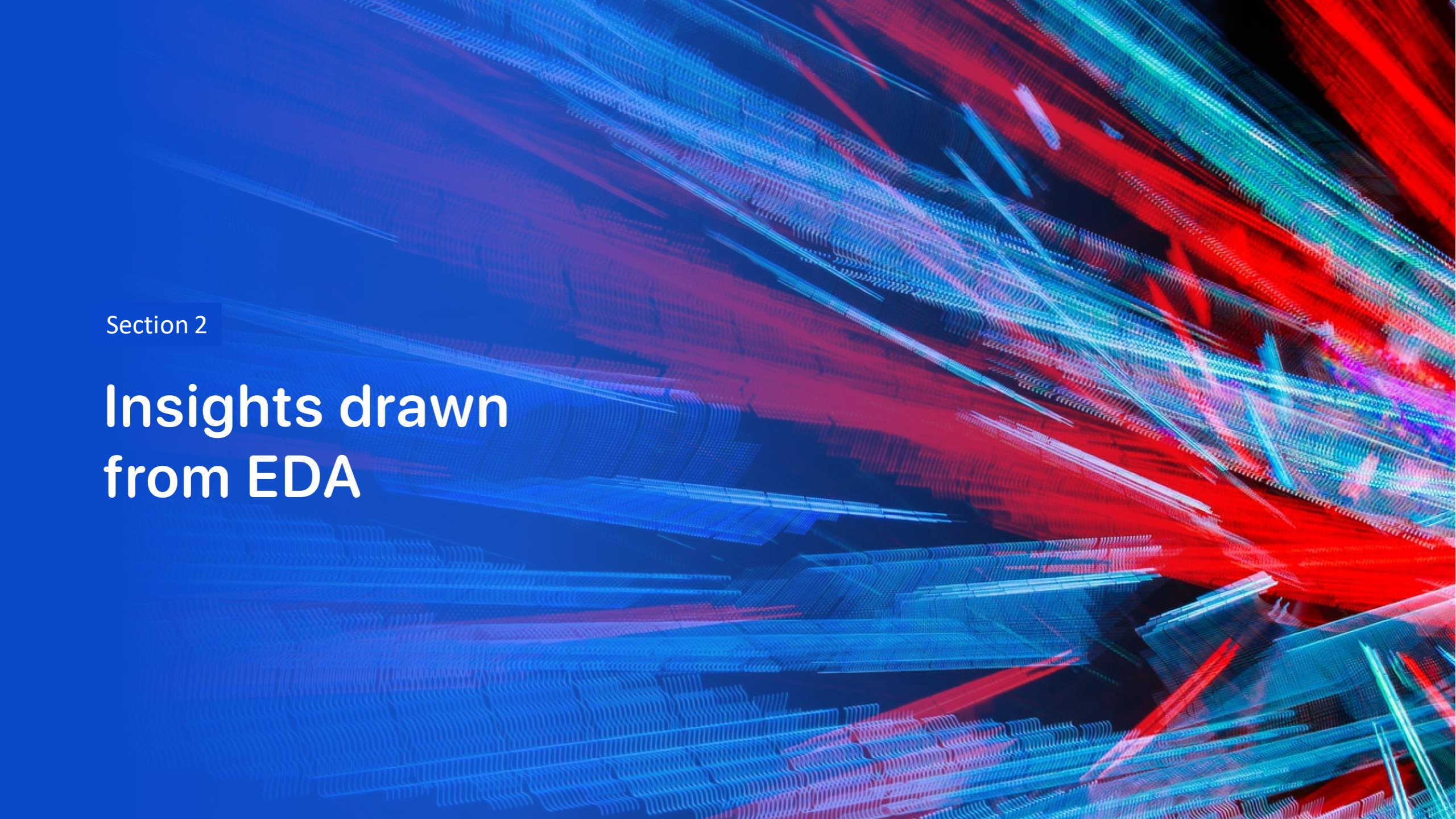


Results

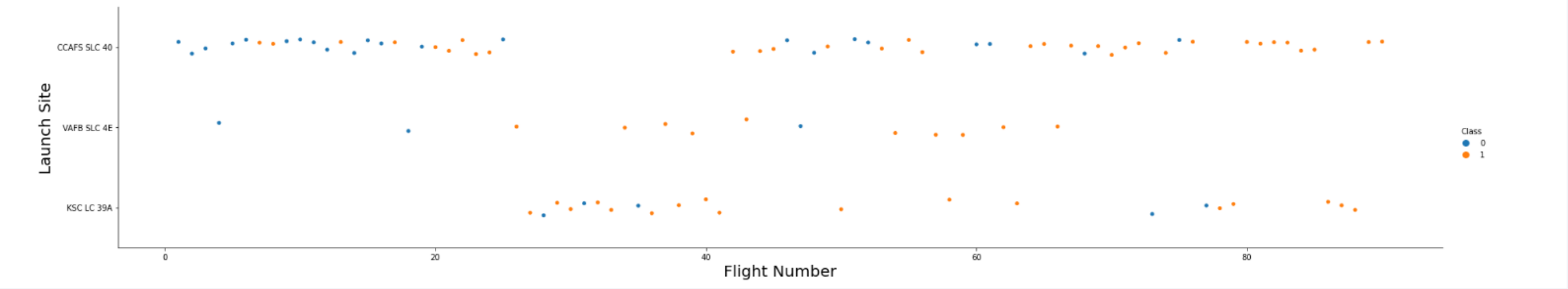
- The key findings of this study are that more recent launches in spite of carrying more weight have a much higher rate of success
- Most launches are from the Florida cape canavaral site.
- ES-L1, GEO, HEO and SSO orbits have the highest success rates.
- All models perform comparably and can predict success at an 83% rate in the test data.

Section 2

Insights drawn from EDA

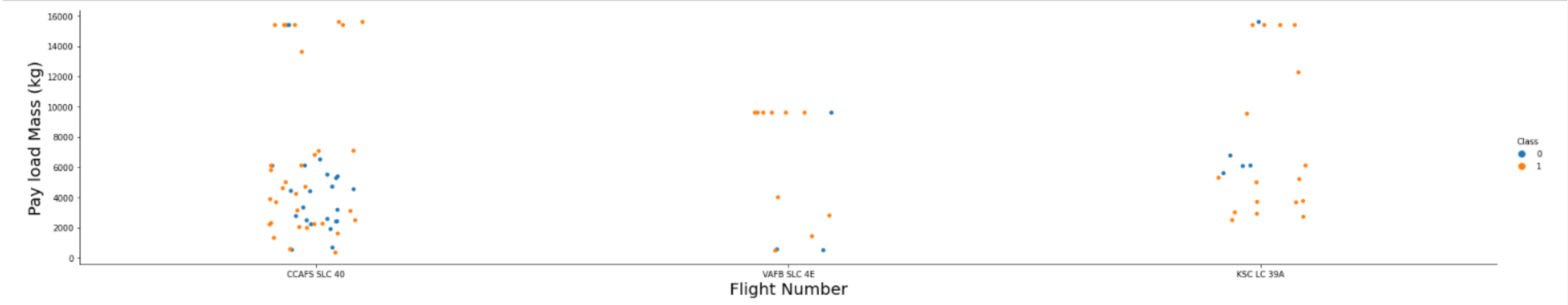


Flight Number vs. Launch Site



- CCAFS SLC 40 is the site of the most launches.
- Most recent launches are successful independent of launch site.
- VAFB SLC 4E has the highest rate of success.

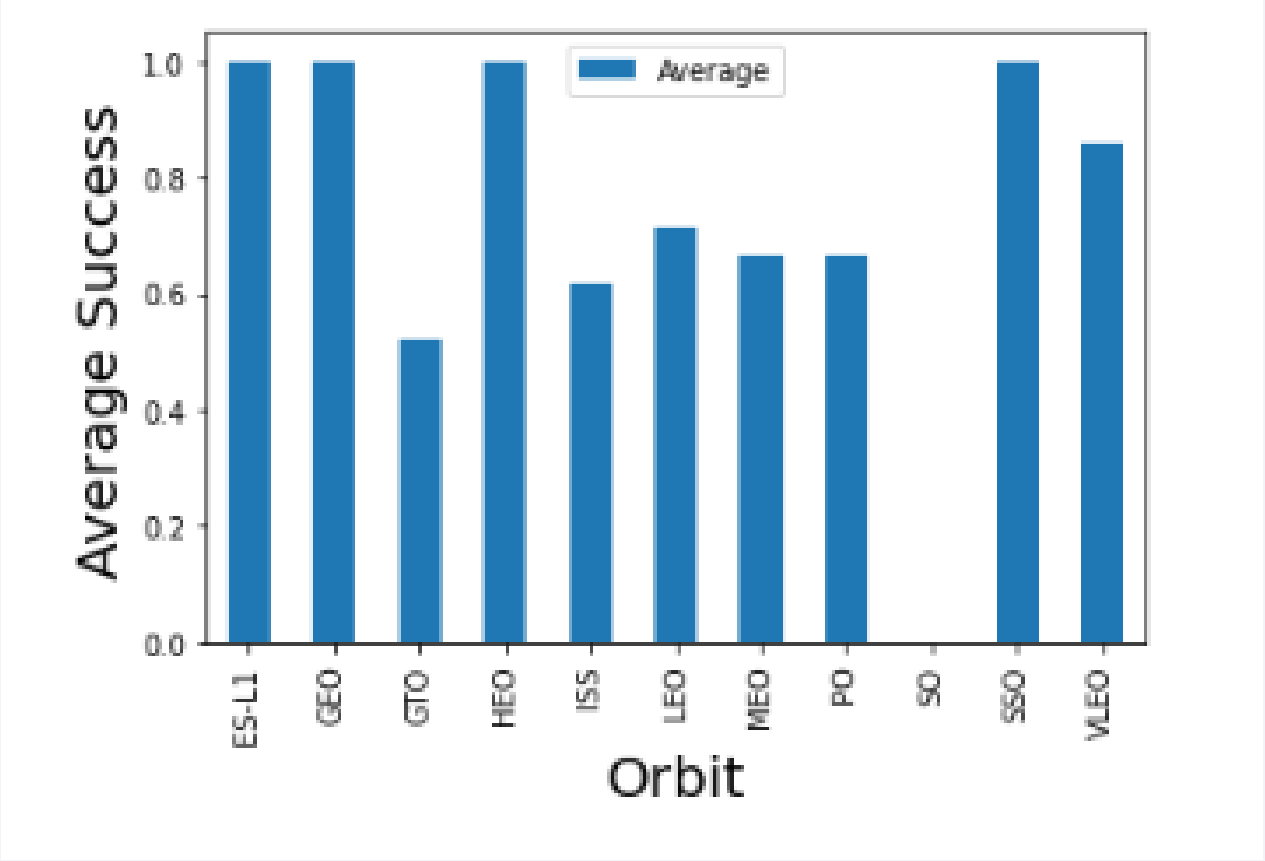
Payload vs. Launch Site



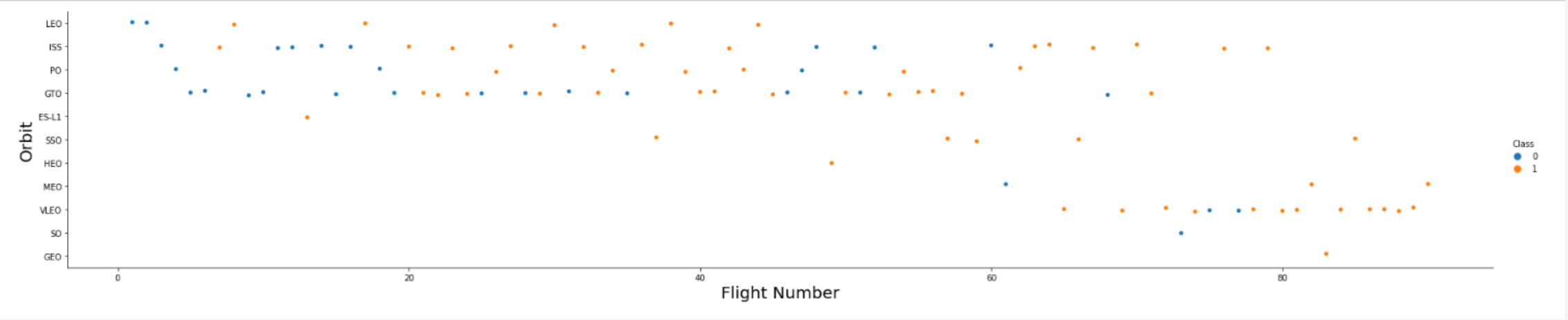
- No rocket launched from VAFB SLC 4E exceeds 10,000 kg, and is most often exactly 10,000 kg.
- Heavy launches have a higher success rate, but this may be confounded by other variables such as launch date.

Success Rate vs. Orbit Type

- ES-L1, GEO, HEO, SSO have high rates of success, approximately 100%
- GTO has one of the lowest at <60%.
- SO has a 0% success rate but only had 1 launch (see next slide).

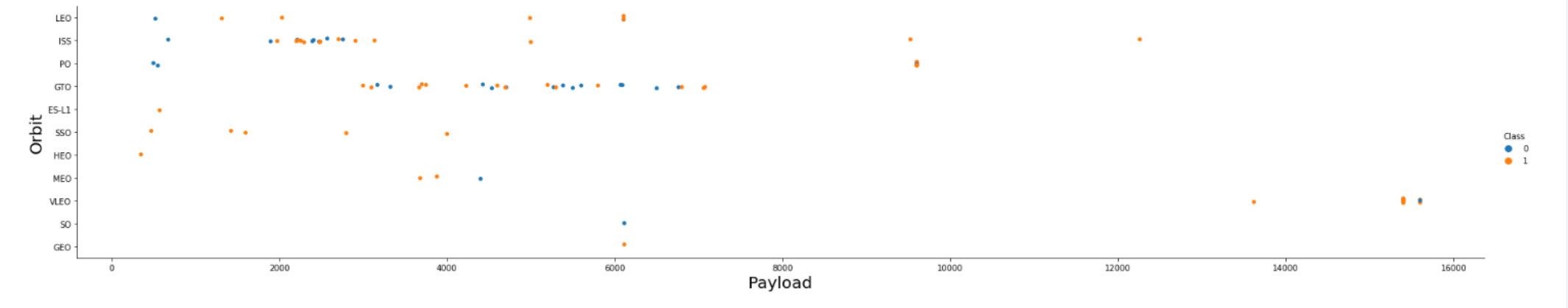


Flight Number vs. Orbit Type



- The prior slide conveys average success, but doesn't capture the time or number.
- Most recent flights have been to VLEO, with an overall high success rate.
- LEO and ISS were most often early launches.
- GTO has the most launches overall.

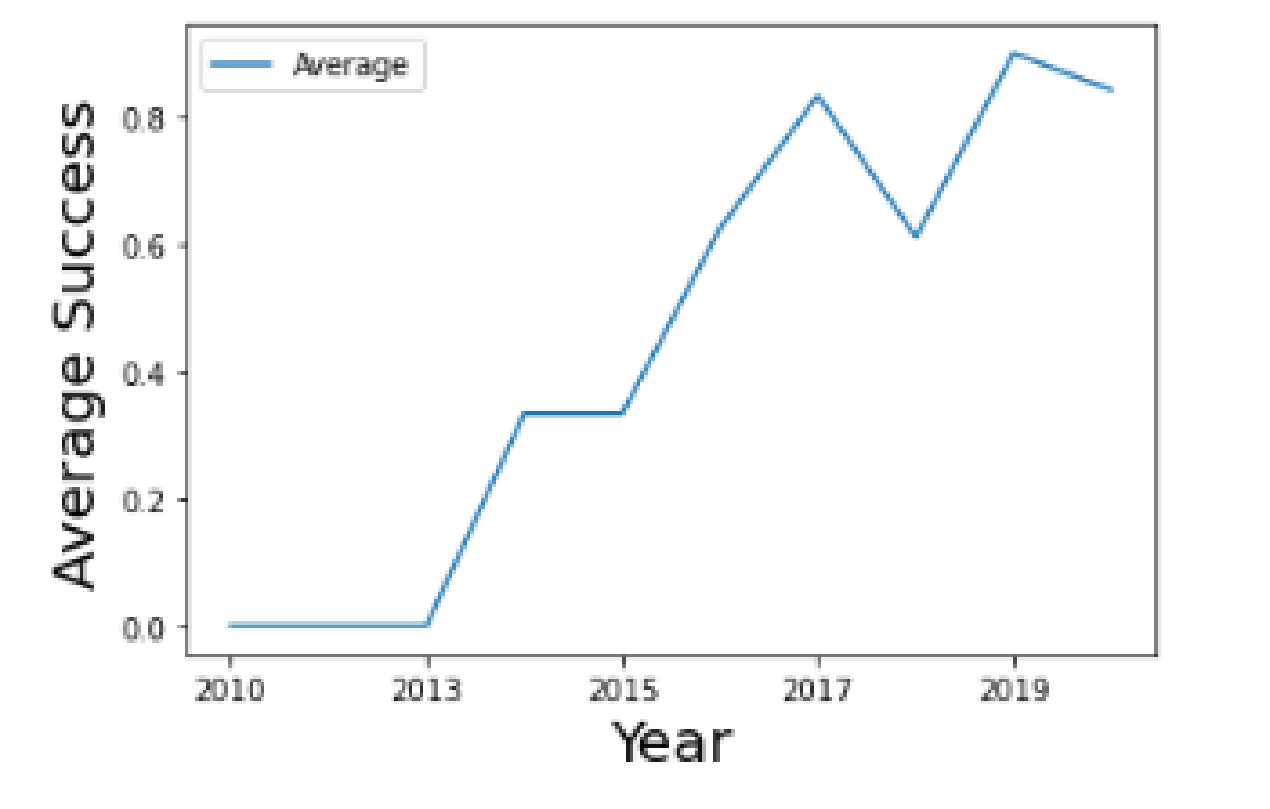
Payload vs. Orbit Type



- All heavy launches are associated with VLEO orbits, while light launches are spread across multiple obrits.
- Launches at 10,000 kg are mostly concentrated at PO orbit type.

Launch Success Yearly Trend

- There has been a steady increase in success rate by time.
- This is important to keep in mind when considering other variables as it may also represent technologic improvements.



All Launch Site Names

- There are 4 unique launch sites within this data.

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

Launch Site Names 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	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	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

- Find 5 records where launch sites begin with `CCA`
- These five launches have variable payload, customer and landing outcomes.

Total Payload Mass

- Calculate the total payload carried by boosters from NASA
- In total rockets have carried 107,010kg of payload for NASA.

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1
- The average mass has been 2,928 kg per launch.

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad
- The first successful ground landing was Dec 22, 2015.

Successful Drone Ship Landing with Payload between 4000 and 6000

- List the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000
- 5 rocket booster versions were utilized on these launches

booster_version
F9 FT B1021.2
F9 FT B1031.2
F9 FT B1020
F9 FT B1022
F9 FT B1026

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes
- Overall 99 launches have a mission outcome have successful outcome, while 1 failure in flight, and 1 had success with payload status unclear.

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass
- Multiple rockets have been involved in max launches.

DATE	time__utc_	booster_version	launch_site	payload	payload_mass__kg_	orbit	customer	mission_outcome	landing__outcome
2019-11-11	14:56:00	F9 B5 B1048.4	CCAFS SLC-40	Starlink 1 v1.0, SpaceX CRS-19	15600	LEO	SpaceX	Success	Success
2020-01-07	02:33:00	F9 B5 B1049.4	CCAFS SLC-40	Starlink 2 v1.0, Crew Dragon in-flight abort test	15600	LEO	SpaceX	Success	Success
2020-01-29	14:07:00	F9 B5 B1051.3	CCAFS SLC-40	Starlink 3 v1.0, Starlink 4 v1.0	15600	LEO	SpaceX	Success	Success
2020-02-17	15:05:00	F9 B5 B1056.4	CCAFS SLC-40	Starlink 4 v1.0, SpaceX CRS-20	15600	LEO	SpaceX	Success	Failure
2020-03-18	12:16:00	F9 B5 B1048.5	KSC LC-39A	Starlink 5 v1.0, Starlink 6 v1.0	15600	LEO	SpaceX	Success	Failure
2020-04-22	19:30:00	F9 B5 B1051.4	KSC LC-39A	Starlink 6 v1.0, Crew Dragon Demo-2	15600	LEO	SpaceX	Success	Success
2020-06-04	01:25:00	F9 B5 B1049.5	CCAFS SLC-40	Starlink 7 v1.0, Starlink 8 v1.0	15600	LEO	SpaceX, Planet Labs	Success	Success
2020-09-03	12:46:14	F9 B5 B1060.2	KSC LC-39A	Starlink 11 v1.0, Starlink 12 v1.0	15600	LEO	SpaceX	Success	Success
2020-10-06	11:29:34	F9 B5 B1058.3	KSC LC-39A	Starlink 12 v1.0, Starlink 13 v1.0	15600	LEO	SpaceX	Success	Success
2020-10-18	12:25:57	F9 B5 B1051.6	KSC LC-39A	Starlink 13 v1.0, Starlink 14 v1.0	15600	LEO	SpaceX	Success	Success
2020-10-24	15:31:34	F9 B5 B1060.3	CCAFS SLC-40	Starlink 14 v1.0, GPS III-04	15600	LEO	SpaceX	Success	Success
2020-11-25	02:13:00	F9 B5 B1049.7	CCAFS SLC-40	Starlink 15 v1.0, SpaceX CRS-21	15600	LEO	SpaceX	Success	Success

2015 Launch Records

booster_version	landing__outcome	launch_site	DATE
F9 v1.1 B1012	Failure (drone ship)	CCAFS LC-40	2015-01-10
F9 v1.1 B1015	Failure (drone ship)	CCAFS LC-40	2015-04-14

- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015
- There have been two failed launches using this criteria.

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

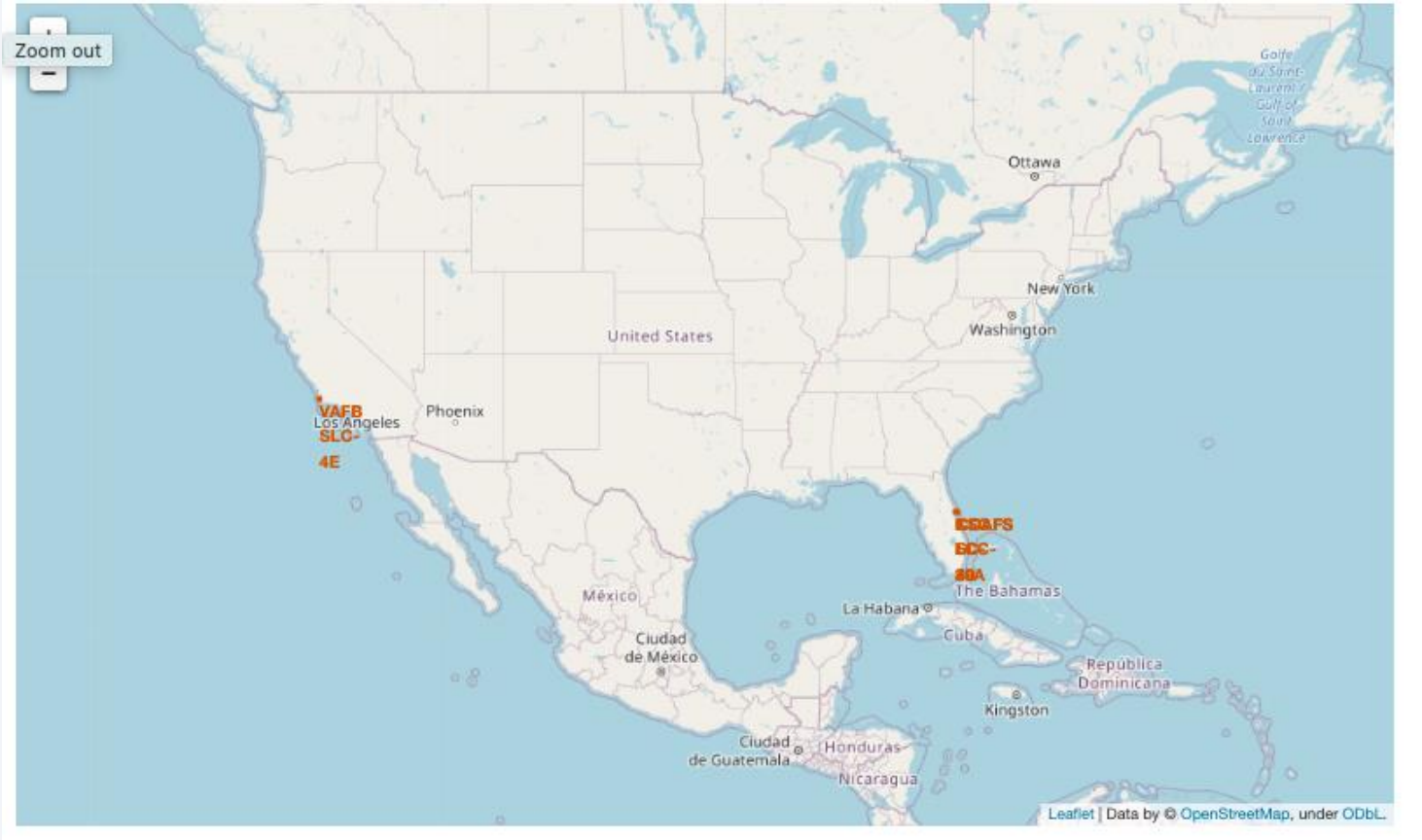
- Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order
- The most common landing outcome result is a no attempt at 10, with drone ship landings at 5 each for success and failure.

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


Section 3
Launch Sites
Proximities Analysis

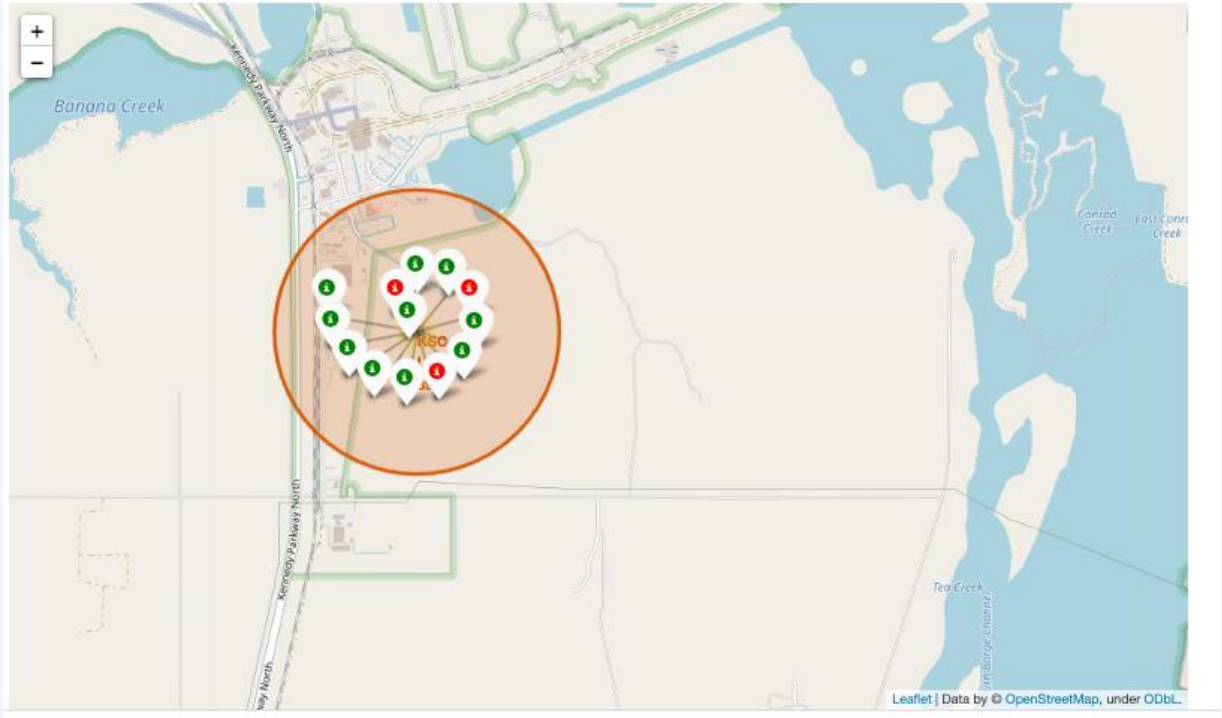
Map of US Launch Sites

- 4 launches sites were utilized for SpaceX launches.
- 3 are concentrated in close proximity in Florida.
- 1 is in California.
- All are near to ocean launches.



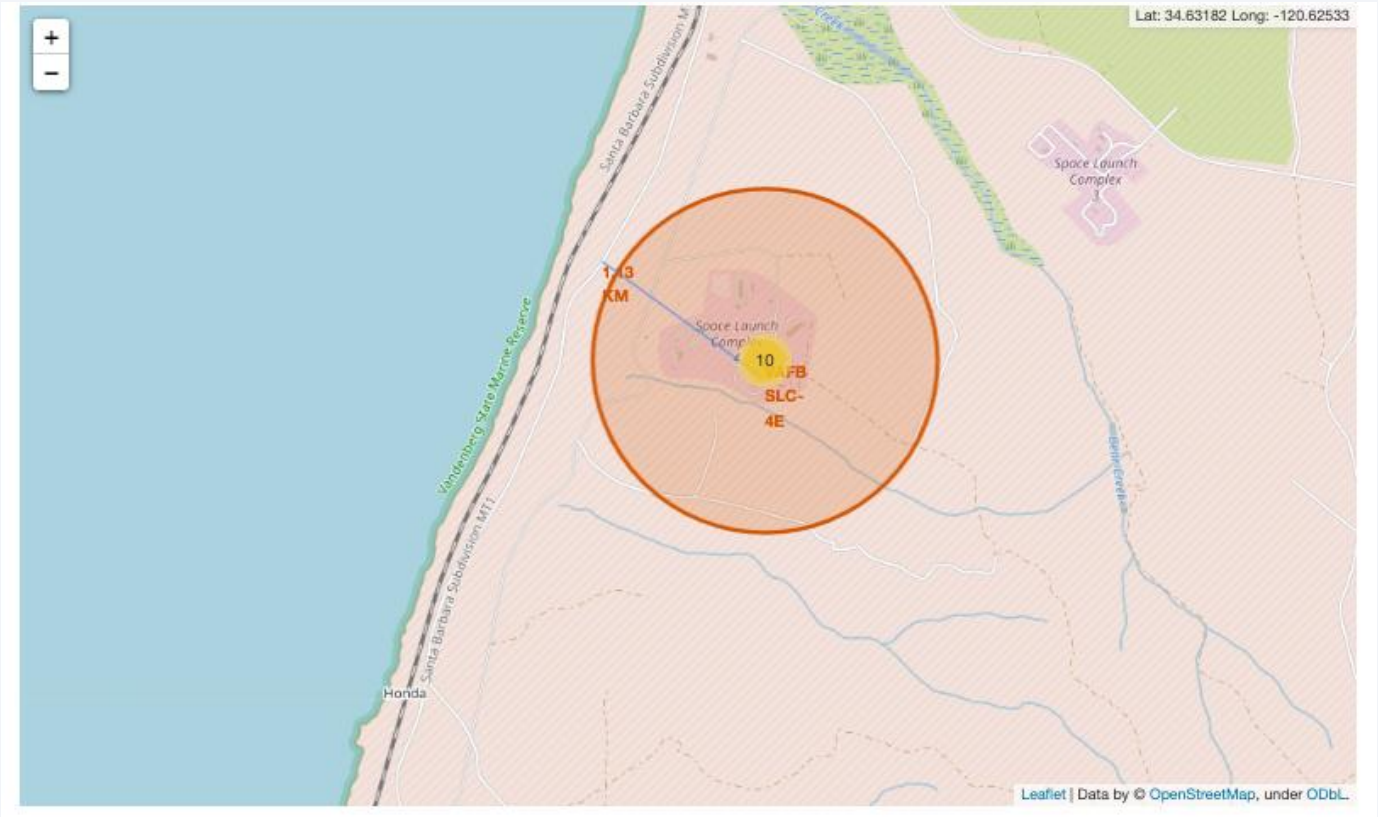
Map of KSC LC-39A Launches

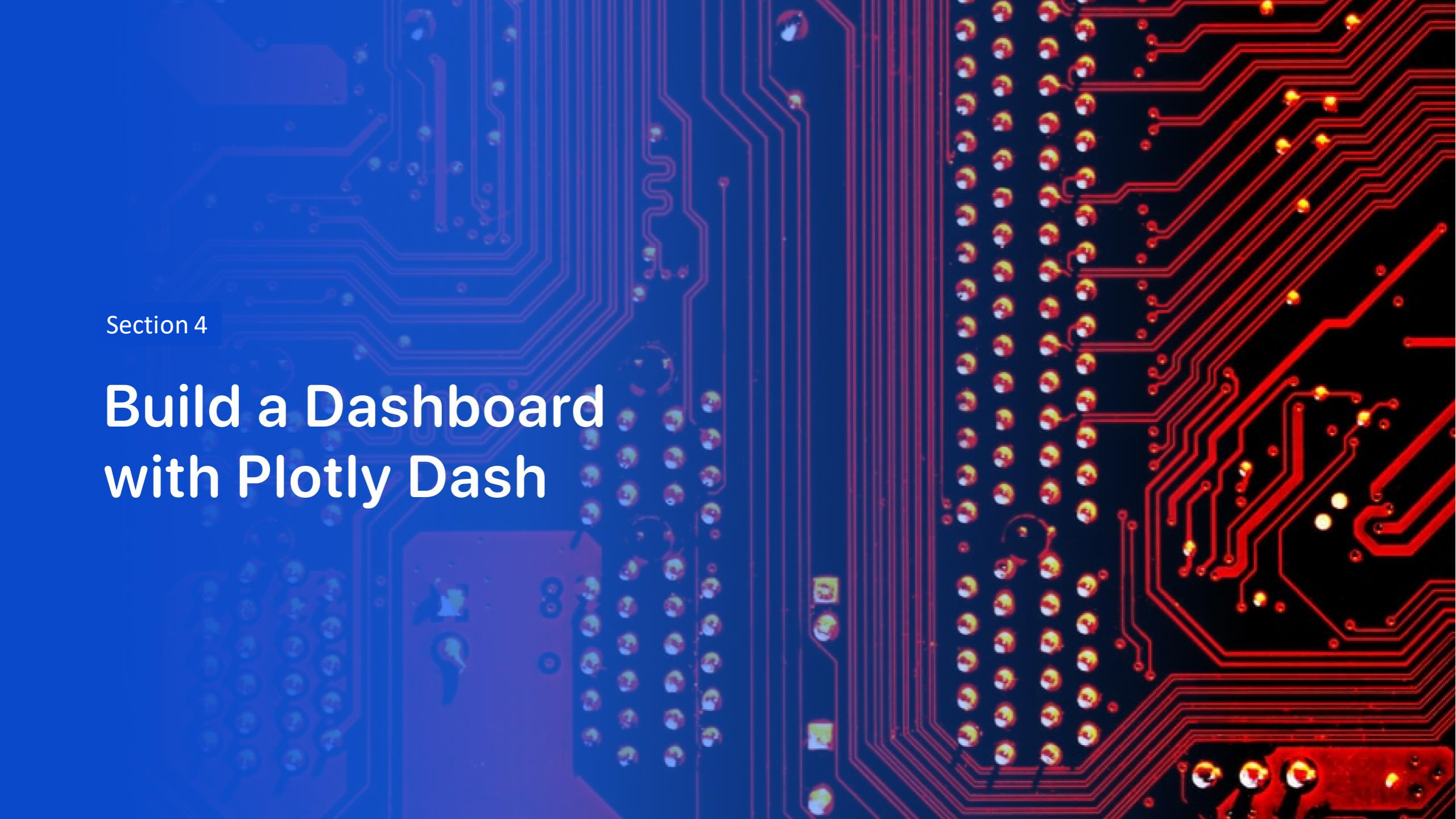
- 13 launches took place as KSC LC-39A in Florida.
- 10 successes in green.
- 3 failures in red.



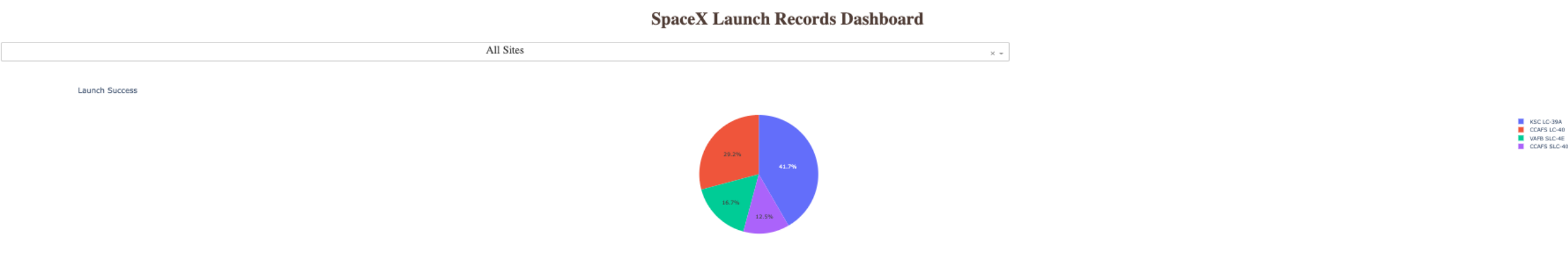
Launch Site Proximity to Ocean

- VAFB Launch site in california is located approximately 1.13 km from the ocean coast.
- Distances can be calculated utilizing interactive Folium plots.



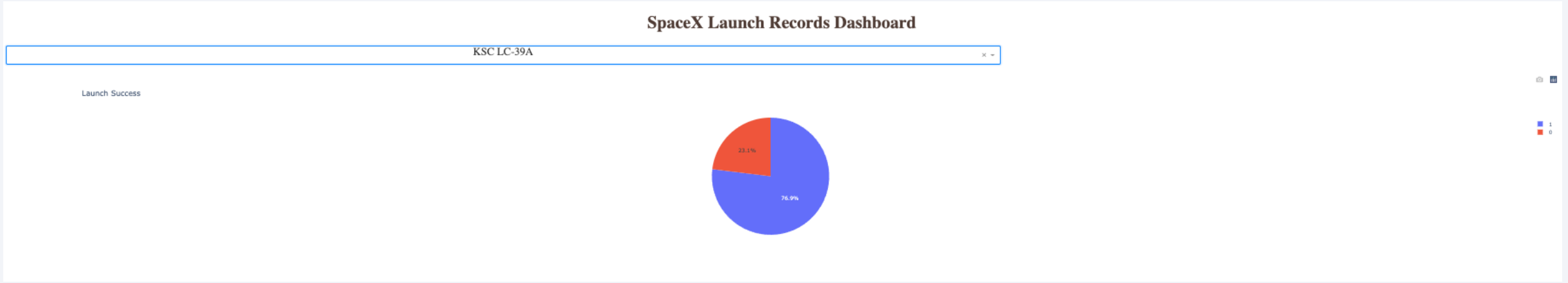

Section 4
Build a Dashboard with Plotly Dash

Success By Launch Site



- Across all launch sites, 41.7% of success occurred form the KSC LC-39A launch site.
- The fewest left from CCAFS SLC-40.
- Note this is not controlled for the number of launches at each site, or the relative distribution of failures.

Launch Success as KSC LC-39A



- 76.9% of launches (total of 13) at KSC LC-39A were a success.

Booster Rocket Version For Payload Between 5000 and 7500 kg

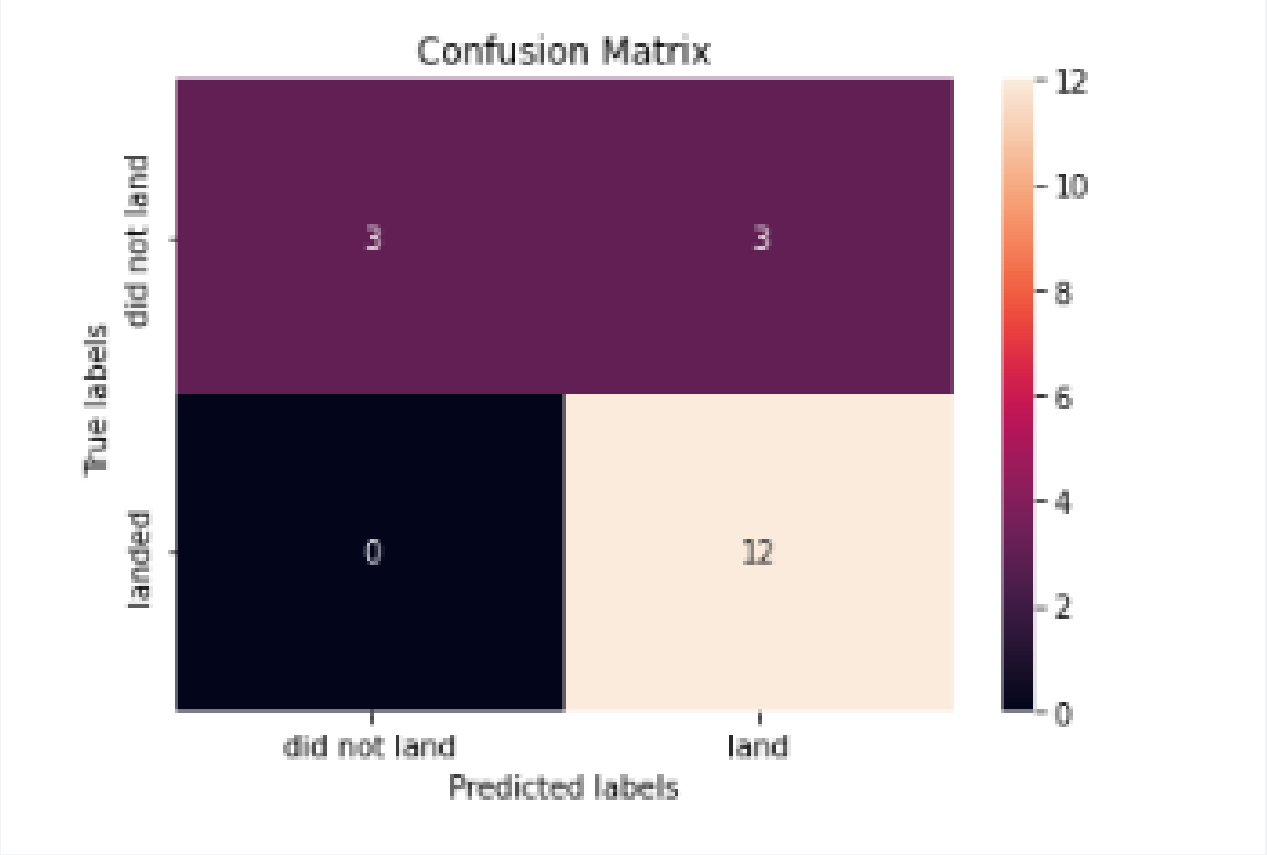
- Between 5000 and 7500 kg most launches were a failure (7/9).
- These are concentrated on FT and B4 versions.

Section 5

Predictive Analysis (Classification)

Confusion Matrix

- All 4 models had comparable accuracy at 83%.
- Each model had a similar confusion matrix with 12 correctly predicted to land, 3 correctly predicted to fail.
- There were 3 incorrect at predicted land that actually failed.



Conclusions

- The key findings of this study are that more recent launches in spite of carrying more weight have a much higher rate of success
- Most launches are from the Florida cape canavaral site.
- ES-L1, GEO, HEO and SSO orbits have the highest success rates.
- All models perform comparably and can predict success at an 83% rate in the test data.

Appendix

- All code is available on github using the links provided throughout the document.
- Python version 3.9

Thank you!