
IBM Developer
SKILLS NETWORK
Winning Space Race with Data Science
Robert Mo
12-24-2025
Robert Mo
2025

Outline

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

Executive Summary

- Scraped publicly available data from Space X API and Wikipedia
- Visualized data using different plots
- Used different machine learning algorithms to predict successful launches (best of 87.5% accuracy)

Introduction

- Space X has reusable first stage; cuts costs of rocket launches
- Success rate of first stages' landings
- Can we predict future successful landings of first stages?

Section 1

Methodology

Methodology

Executive Summary

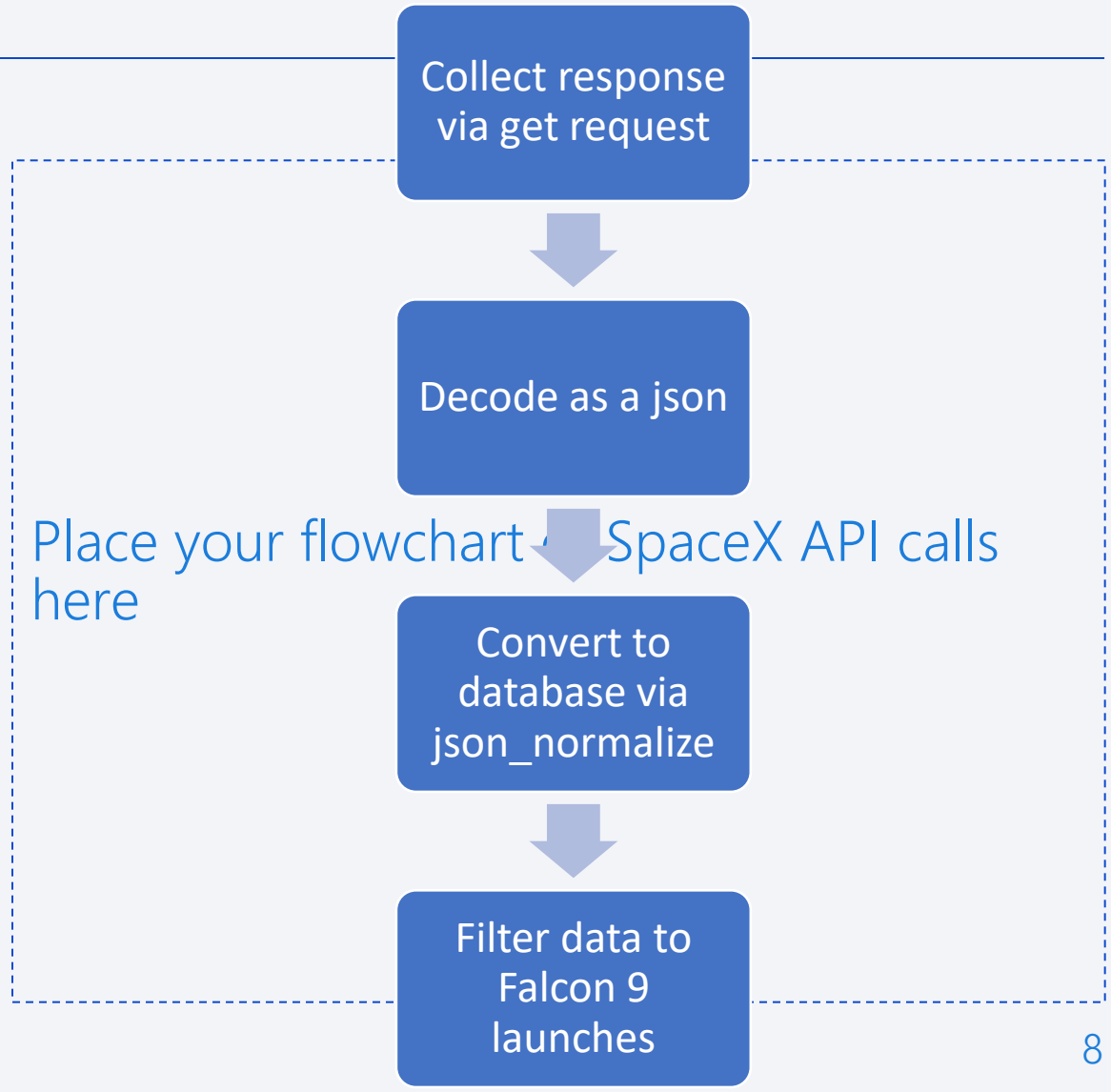
- Obtained Space X API data via get request
- Obtained Wikipedia data via webscraping (Beautifulsoup)
- Wrangled data based on columns (ie. launch site, orbit)
- Visualized data using scatter/bar/line plots
- Made Folium maps and Plotly Dash dashboard
- Predicted via logistic regression, decision tree, KNN, and support vector machine (best of ~83% accuracy)

Data Collection

- Data is collected from publicly available sources (Space X API, Wikipedia)

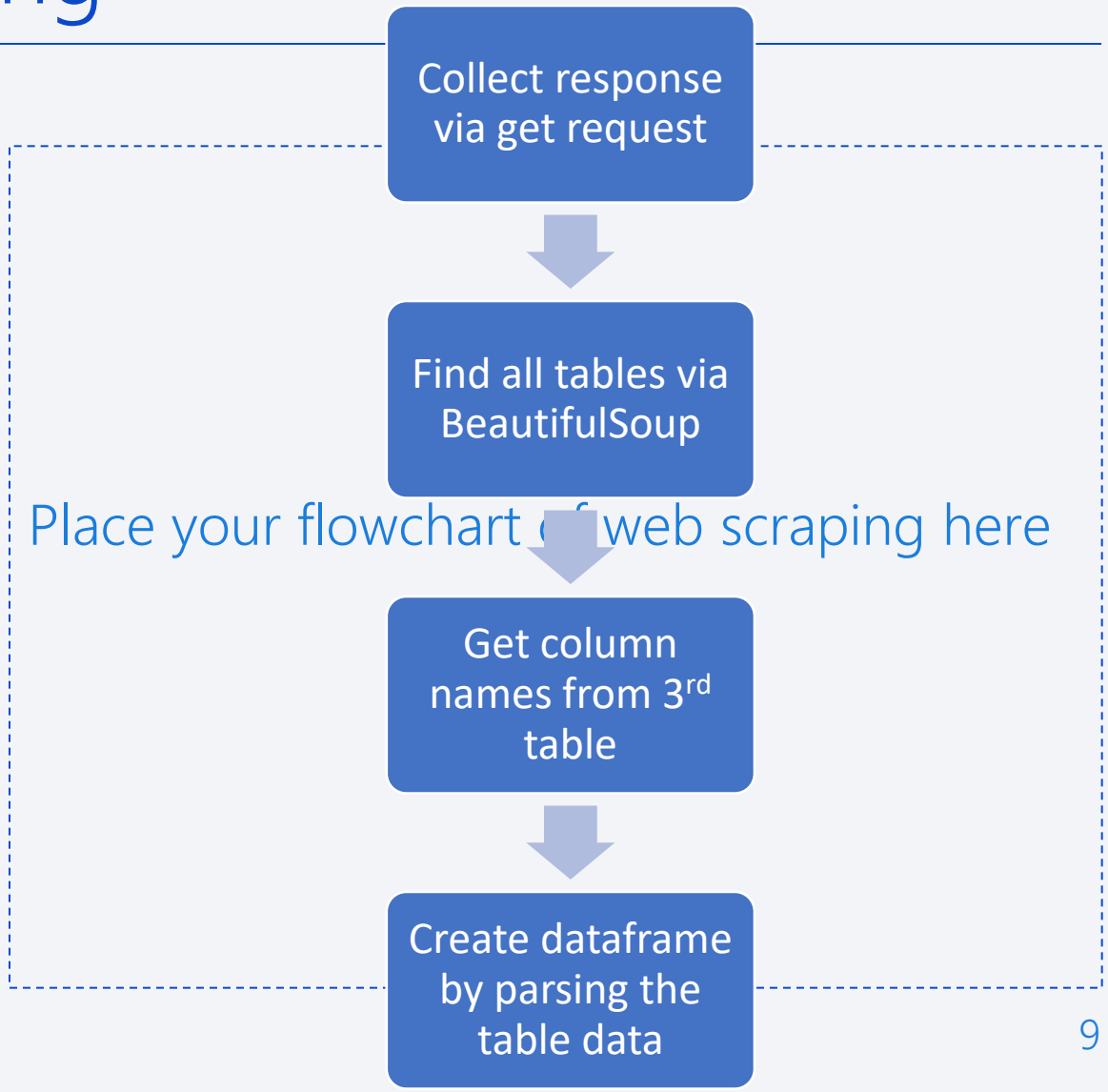
Data Collection – SpaceX API

- Github: [Data-Science-Capstone-Project-Space-X-Data-/jupyter-labs-spacex-data-collection-api.ipynb at main · rwmo2008/Data-Science-Capstone-Project-Space-X-Data-]



Data Collection - Scraping

- Github: [Data-Science-Capstone-Project-Space-X-Data-/jupyter-labs-webscraping.ipynb at main · rwmo2008/Data-Science-Capstone-Project-Space-X-Data-]



Data Wrangling

- Github: [Data-Science-Capstone-Project-Space-X-Data-/labs-jupyter-spacex-Data wrangling.ipynb at main · rwmo2008/Data-Science-Capstone-Project-Space-X-Data-]

Count up column entries via value_count

Create landing class column to sort successful and unsuccessful landings

EDA with Data Visualization

- Flight number vs. payload mass (scatter plot; hue=class): shows that even with increasing payload mass, there were more successful returns with more flight experience
- Flight number vs. launch site (scatter plot; hue=class): KSC LC 39A has a higher success rate in later flights
- Payload mass vs. launch site (scatter plot; hue=class): CCAFS SLC 40 handles the most heavy payloads
- Orbit type success rate (bar plot): ES-L1, GEO, HEO, SSO have 100% success rate; GTO has 50% success rate
- Flight number vs. orbit type (scatter plot; hue=class):
- Github: Data-Science-Capstone-Project-Space-X-Data-/edadataviz.ipynb at main · rwmo2008/Data-Science-Capstone-Project-Space-X-Data-

EDA with SQL

- Used SQL to get
  - Distinct names of launch sites
  - Total payload mass
  - Average payload size
  - Total number of successes and failures in mission outcomes
  - All booster versions that carried max payload size
- Github: Data-Science-Capstone-Project-Space-X-Data-/jupyter-labs-eda-sql-coursera_sqllite.ipynb at main · rwmo2008/Data-Science-Capstone-Project-Space-X-Data-

Build an Interactive Map with Folium

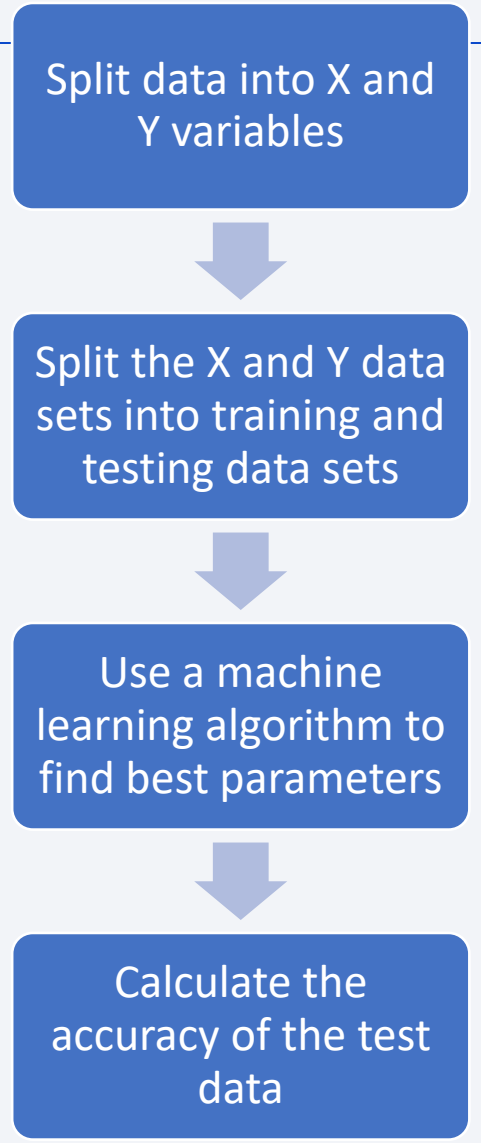
- Markers are used to point out locations of launch sites
- Marker colors are used to differentiate launch success or failure
- Lines are used to display the distance between launch sites to certain places such as towns or roads.
- Github: Data-Science-Capstone-Project-Space-X-Data-/lab_jupyter_launch_site_location.ipynb at main · rwmo2008/Data-Science-Capstone-Project-Space-X-Data-

Build a Dashboard with Plotly Dash

- Circle graphs used to show success rates of launches from each launch site
- Scatter plot used to show correlation of payload mass and mission success (while also showing booster versions)
- Github: Data-Science-Capstone-Project-Space-X-Data-/spacex-dash-app.py at main · rwmo2008/Data-Science-Capstone-Project-Space-X-Data-

Predictive Analysis (Classification)

- Github: [Data-Science-Capstone-Project-Space-X-Data-/SpaceX_Machine Learning Prediction_Part_5.ipynb at main · rwmo2008/Data-Science-Capstone-Project-Space-X-Data-]



Results

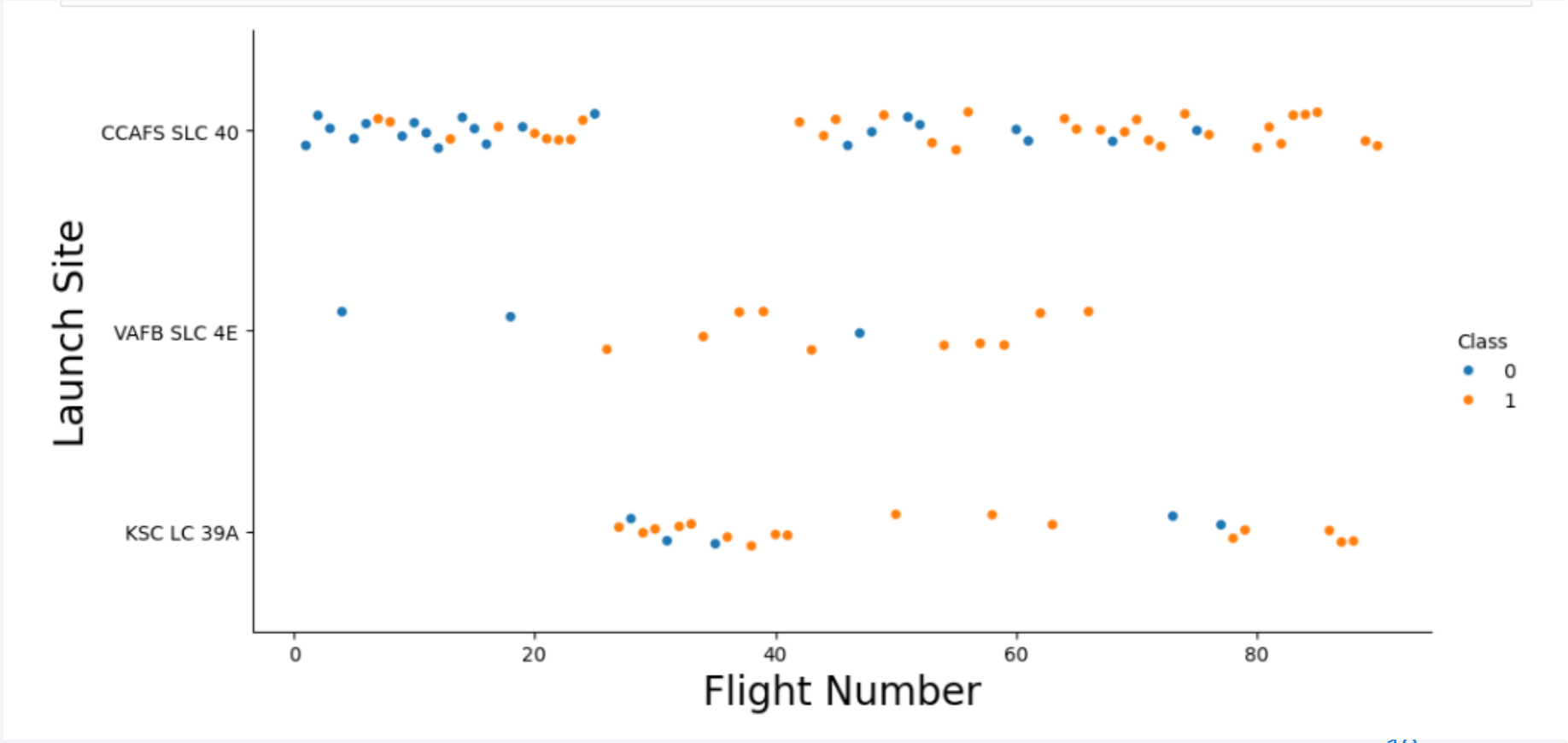
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

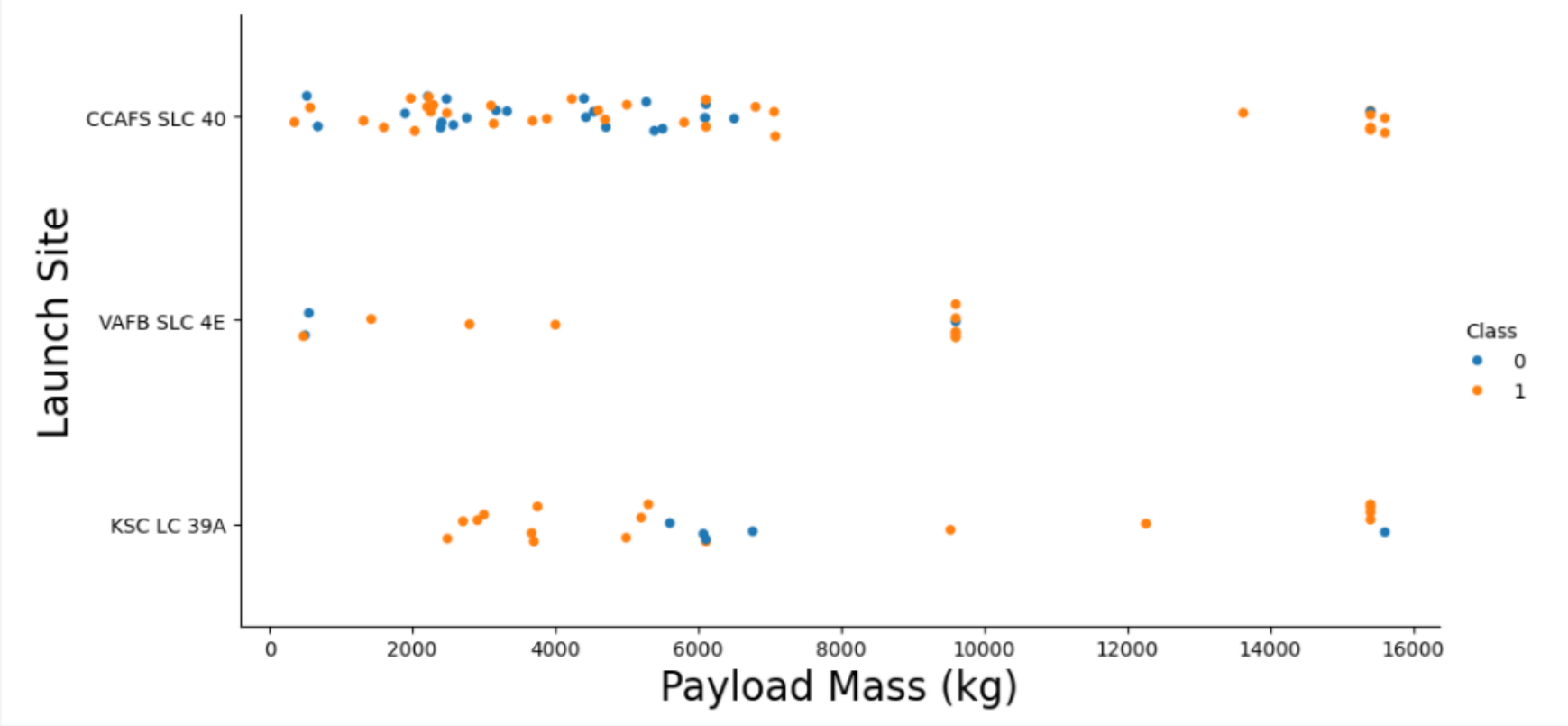
Flight Number vs. Launch Site

- The scatter plot shows that there were more successes as the number of flights increase thanks to the increased experience.



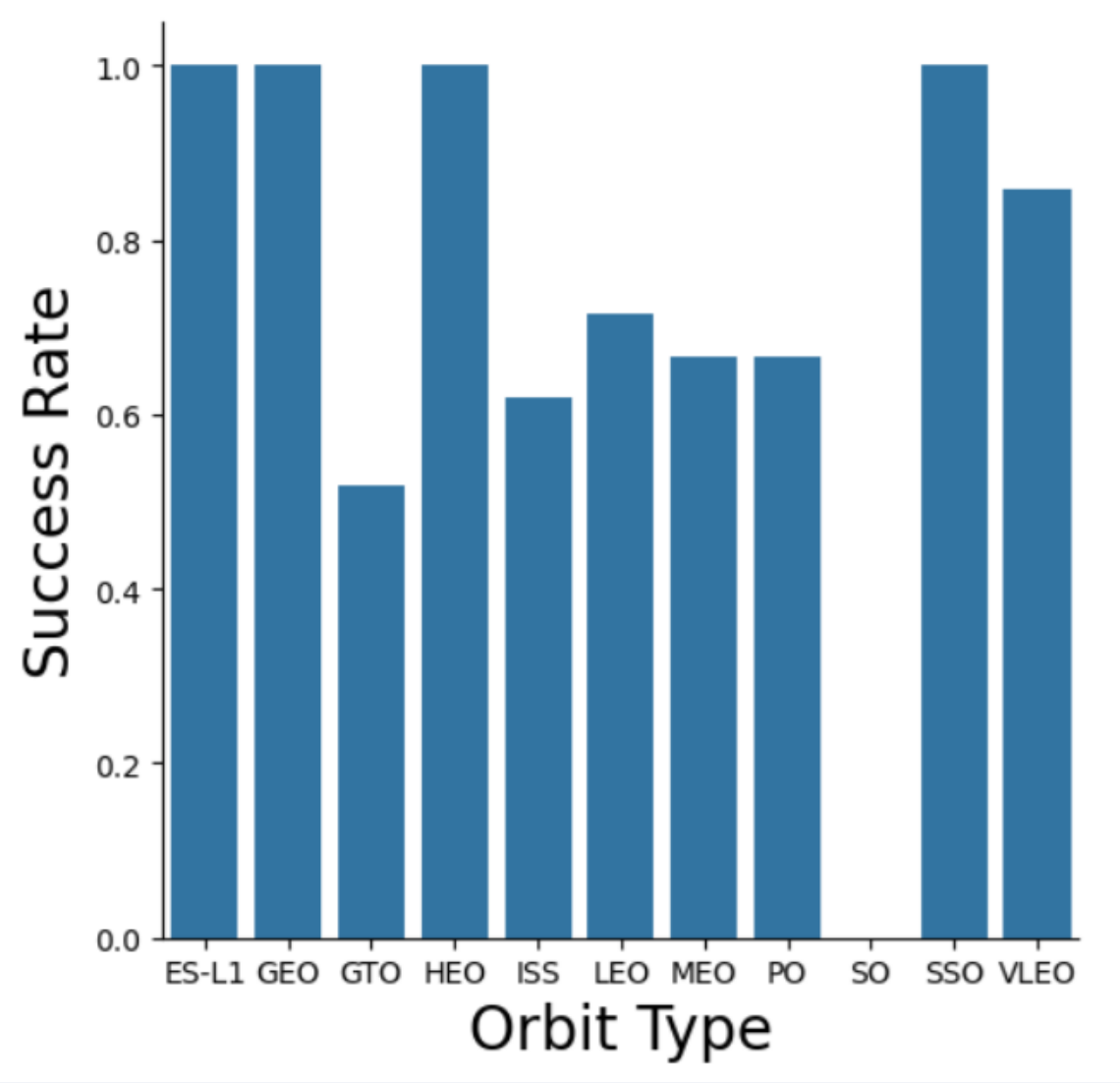
Payload vs. Launch Site

- CCAFS SLC 40 carries most of the heaviest payloads



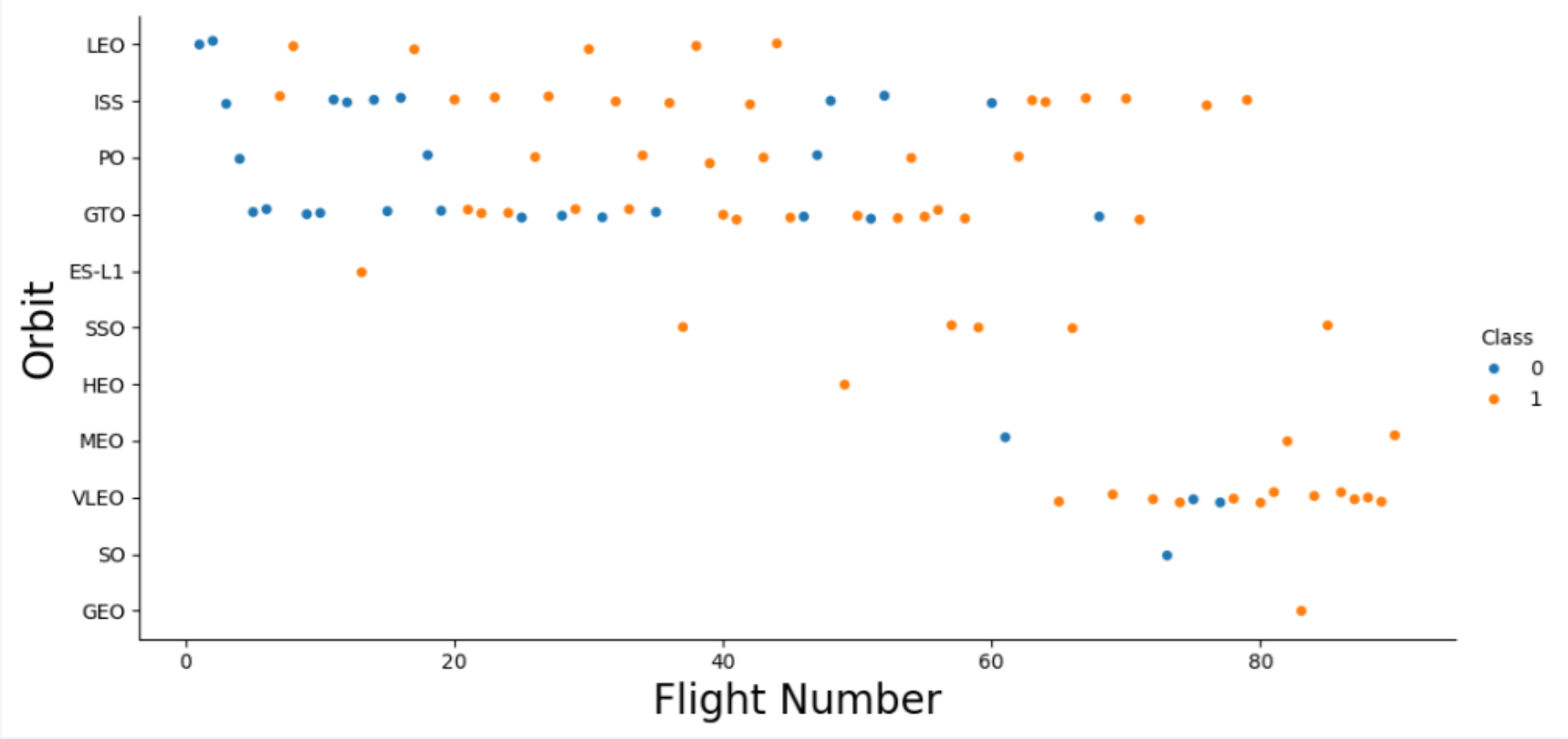
Success Rate vs. Orbit Type

- Missions involving ES-L1, GEO, HEO, and SSO orbits have a 100% success rate
- Missions involving GTO orbits have a 50% success rate



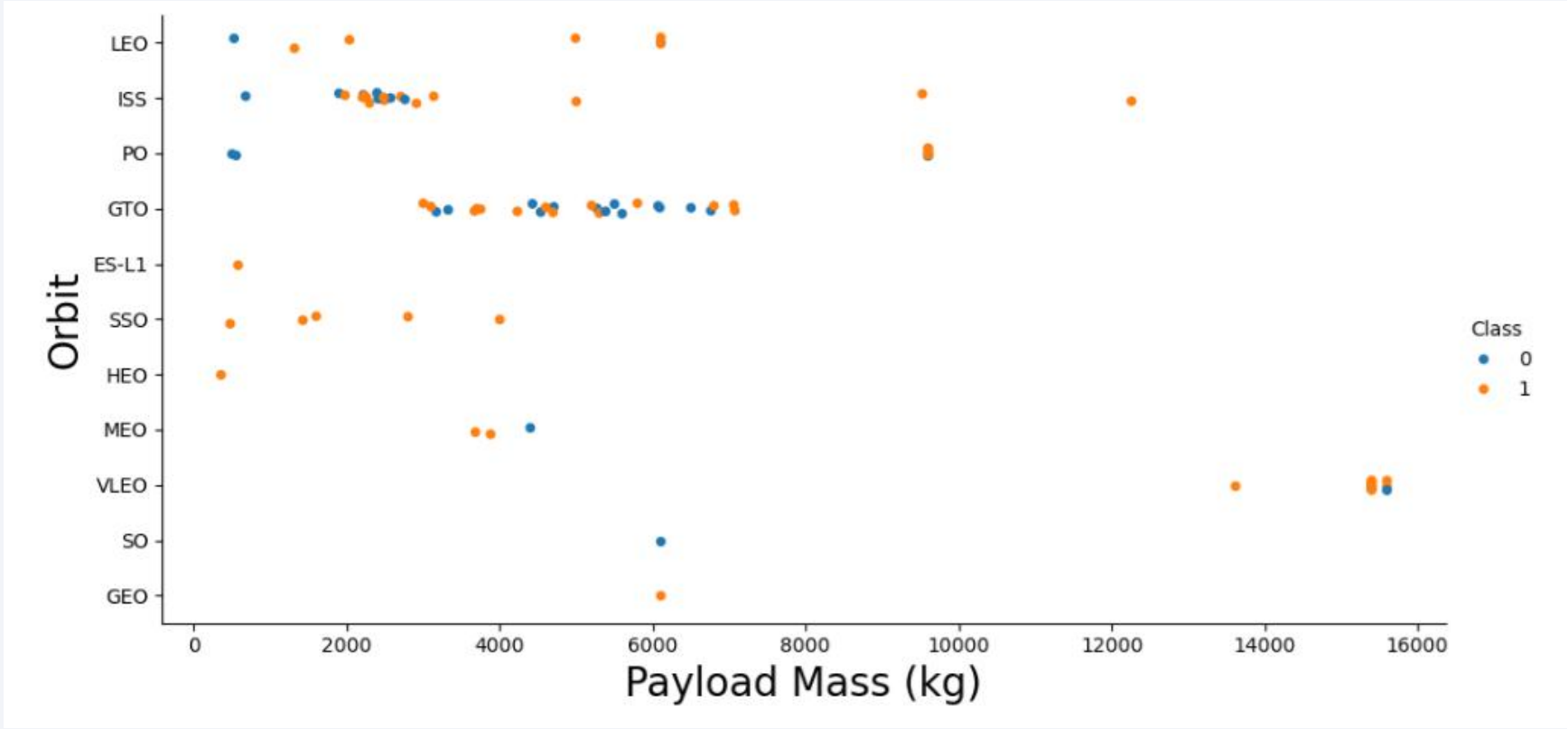
Flight Number vs. Orbit Type

- There were more VLEO orbits as flight numbers increase.



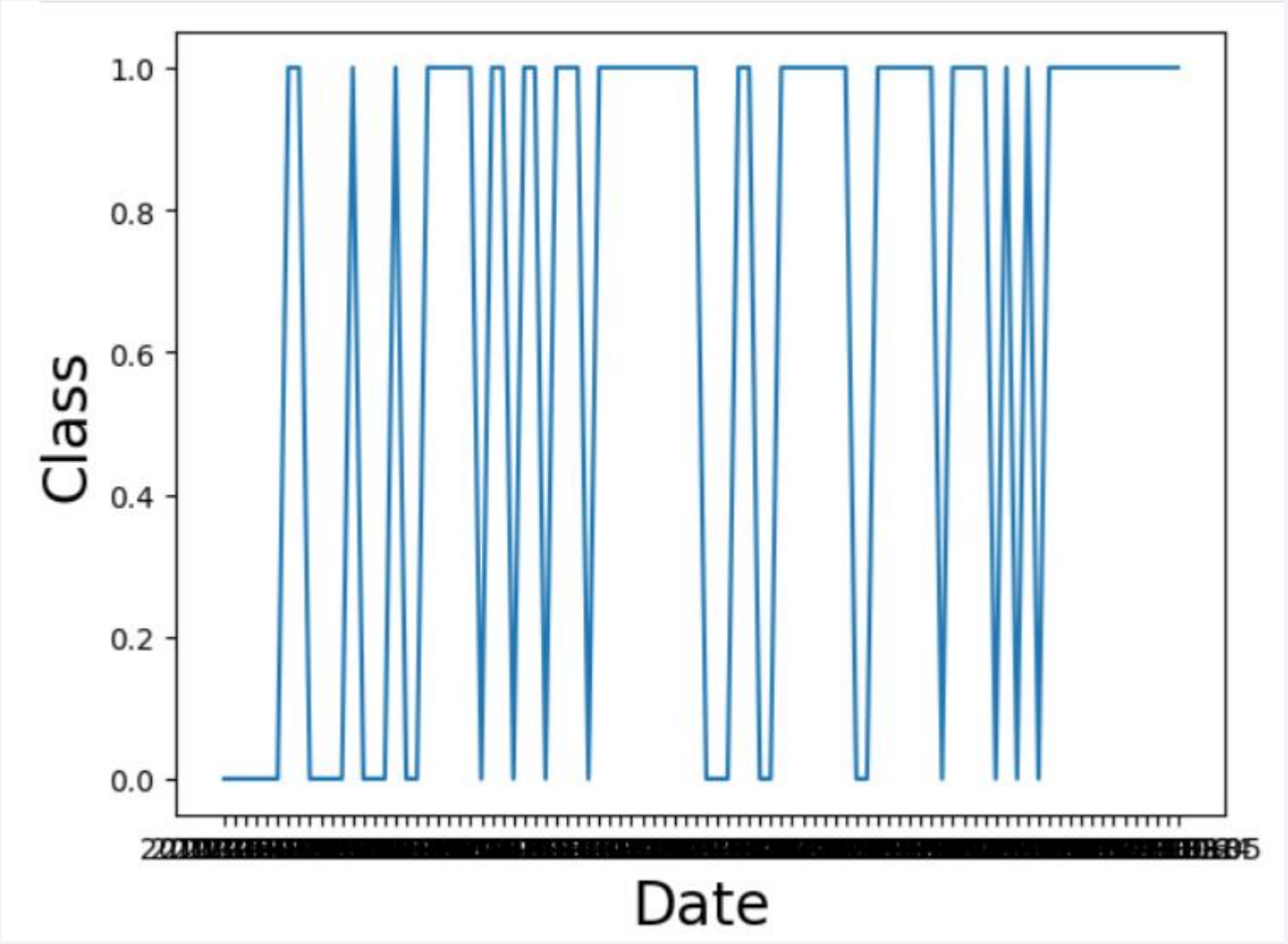
Payload vs. Orbit Type

- The heaviest payloads are seen being carried for the VLEO orbits.



Launch Success Yearly Trend

- Failures decrease in rate as time went on.



All Launch Site Names

- Found using the Select DISTINCT query

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

Launch Site Names Begin with 'CCA'

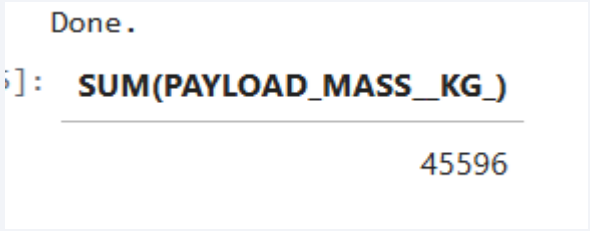
- Used a query that included a WHERE clause with “LIKE ‘CCA%’”, limited to 5 results

]:

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

Total Payload Mass

- Used the select sum query with a Where clause of “CUSTOMER = ‘NASA (CRS)’”



Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1
- Present your query result with a short explanation here

First Successful Ground Landing Date

- MIN (Date) query with a Where clause of “Landing_Outcome = ‘Success (ground pad)’”

First_Successful_Landing_Date
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- Query with drone ship landing outcome and payload mass of 4000-6000 kg

Boosters_in_successful_drone_landings
JCSAT-14
JCSAT-16
SES-10
SES-11 / EchoStar 105

Total Number of Successful and Failure Mission Outcomes

- Query of mission outcome column and Count of successes and failures

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

Boosters Carried Maximum Payload

- Query of booster versions with a where clause with a subquery of max payload mass

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

- Query selecting multiple columns with Where clause set to find failure landing outcomes of a specific time frame

month	Date	Booster_Version	Launch_Site	Landing_Outcome
01	2015-01-10	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
04	2015-04-14	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

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

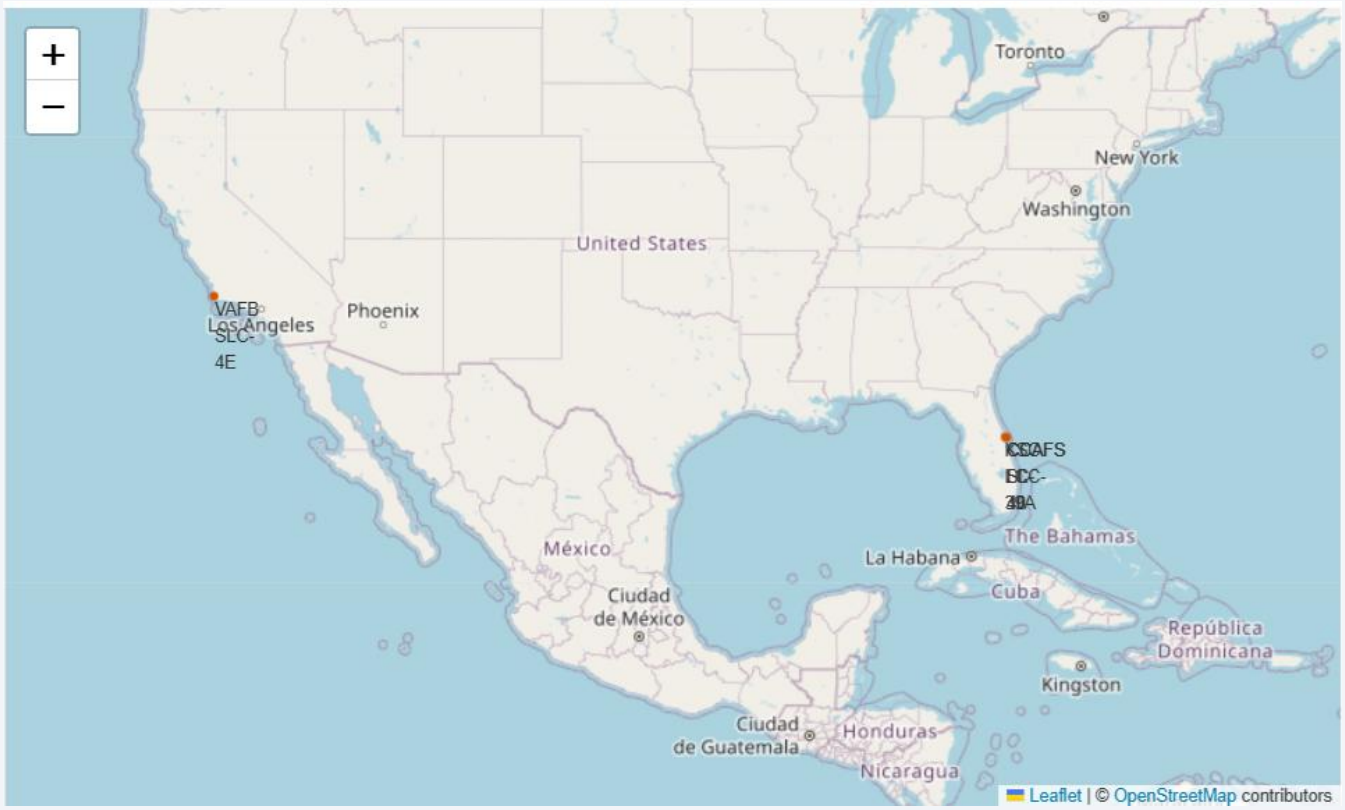
- Query of landing outcomes counted up and ranked in descending order

Landing_Outcome	count_outcomes
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


Section 3
Launch Sites
Proximities Analysis

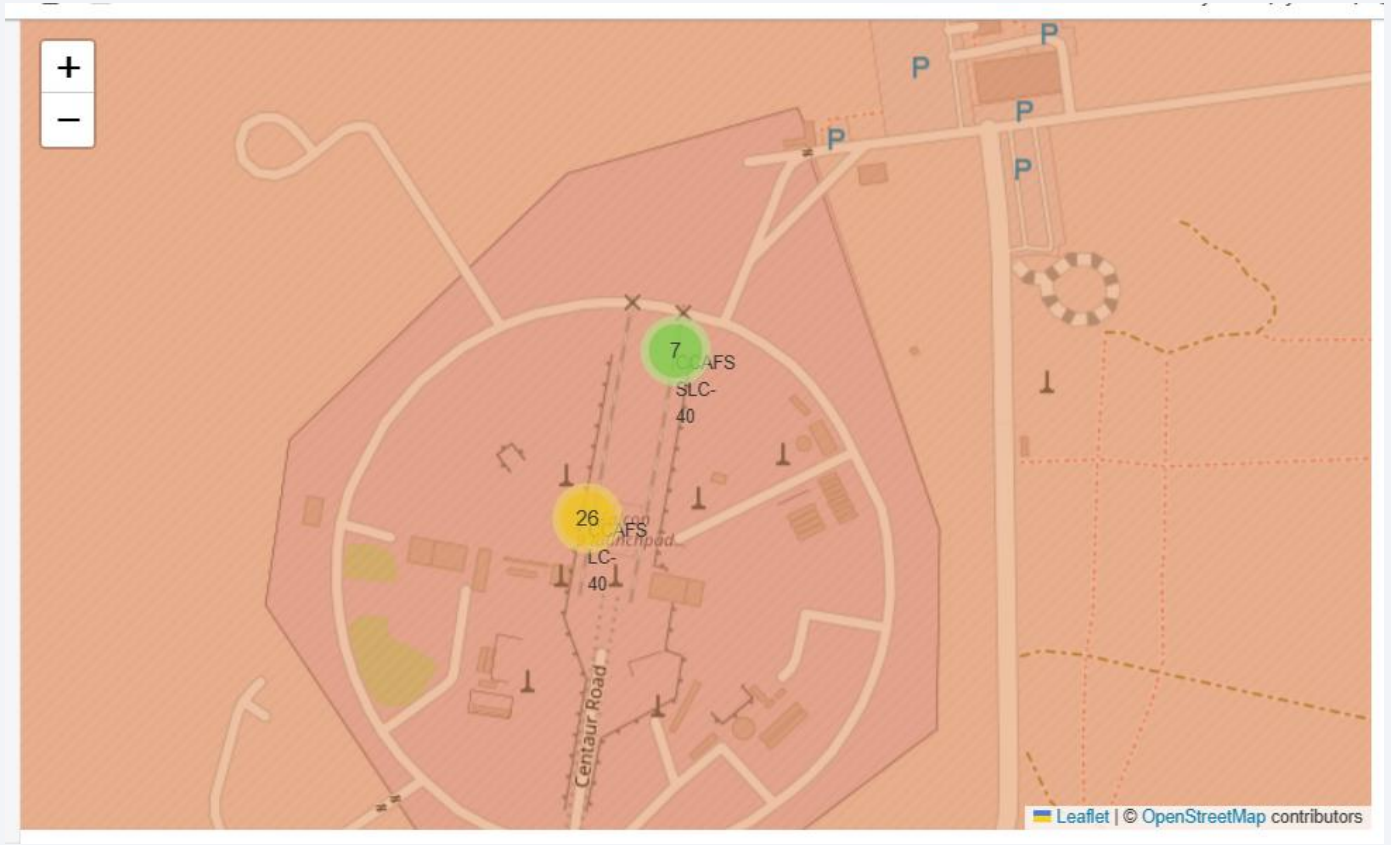
Map of launch locations

- All launch locations are located in either Florida or California (USA)



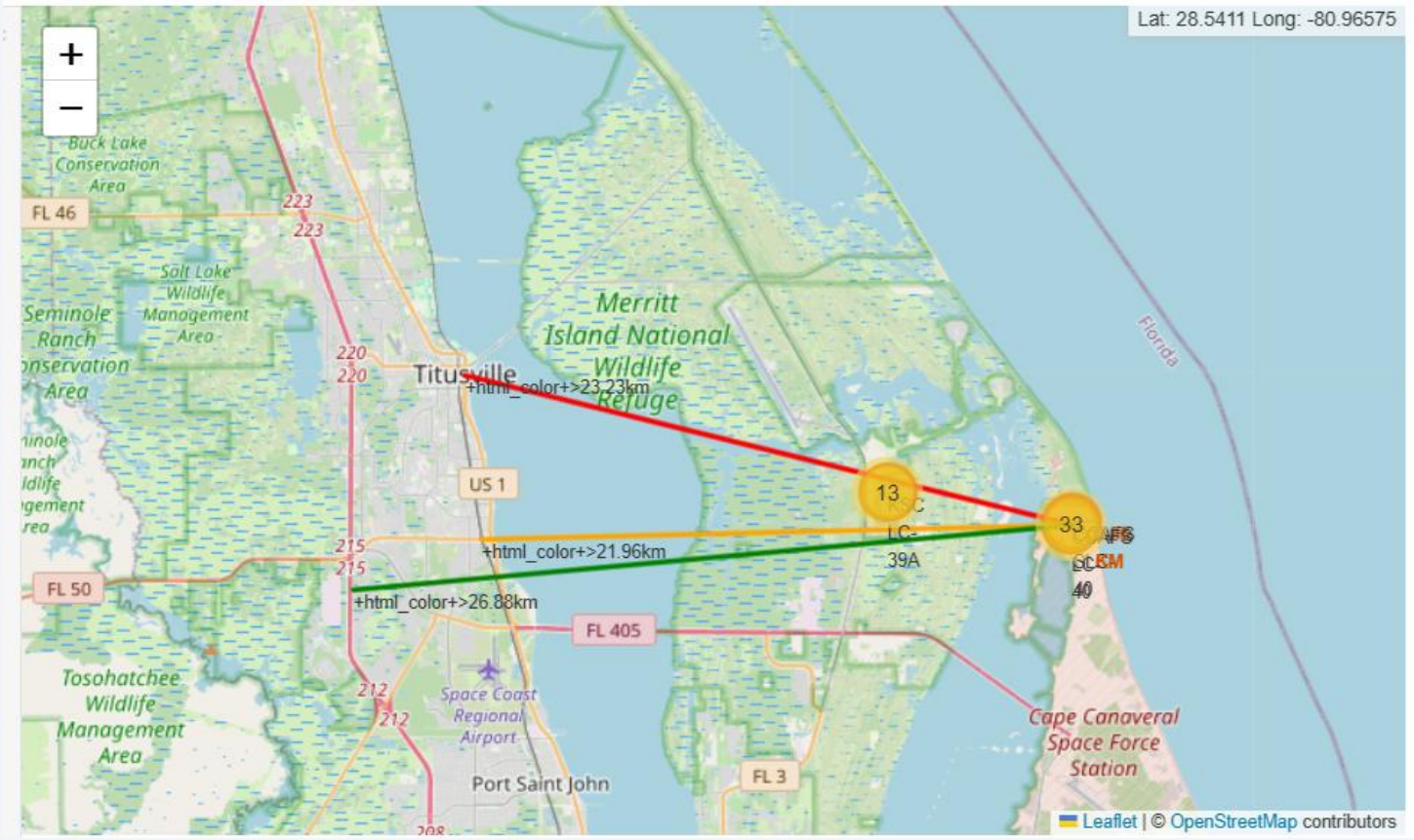
Some successes and failures

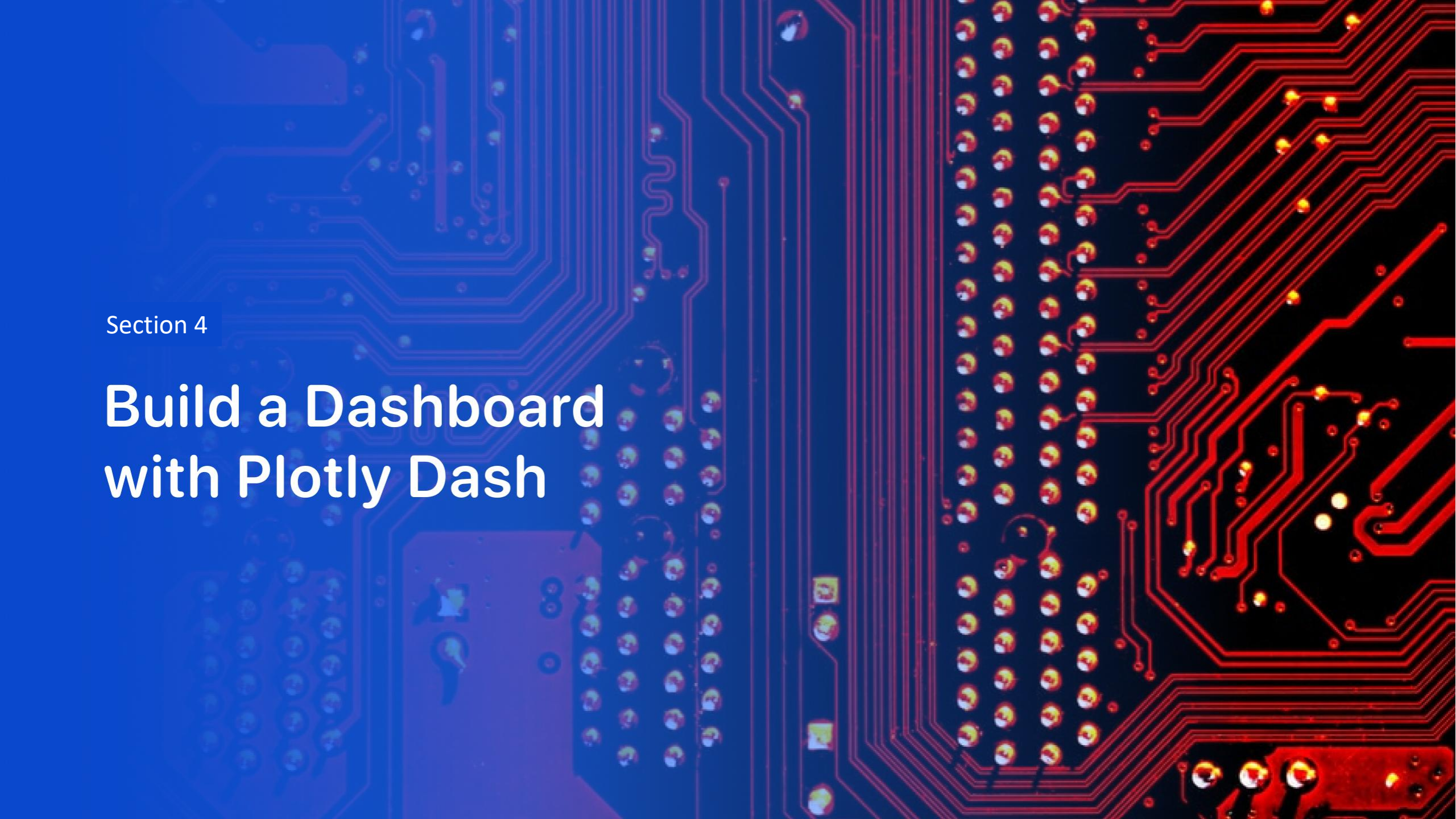
- We can see that FL has a lot of launches with site CCAFS SLC-40 having a high success rate.



Distances from Infrastructure

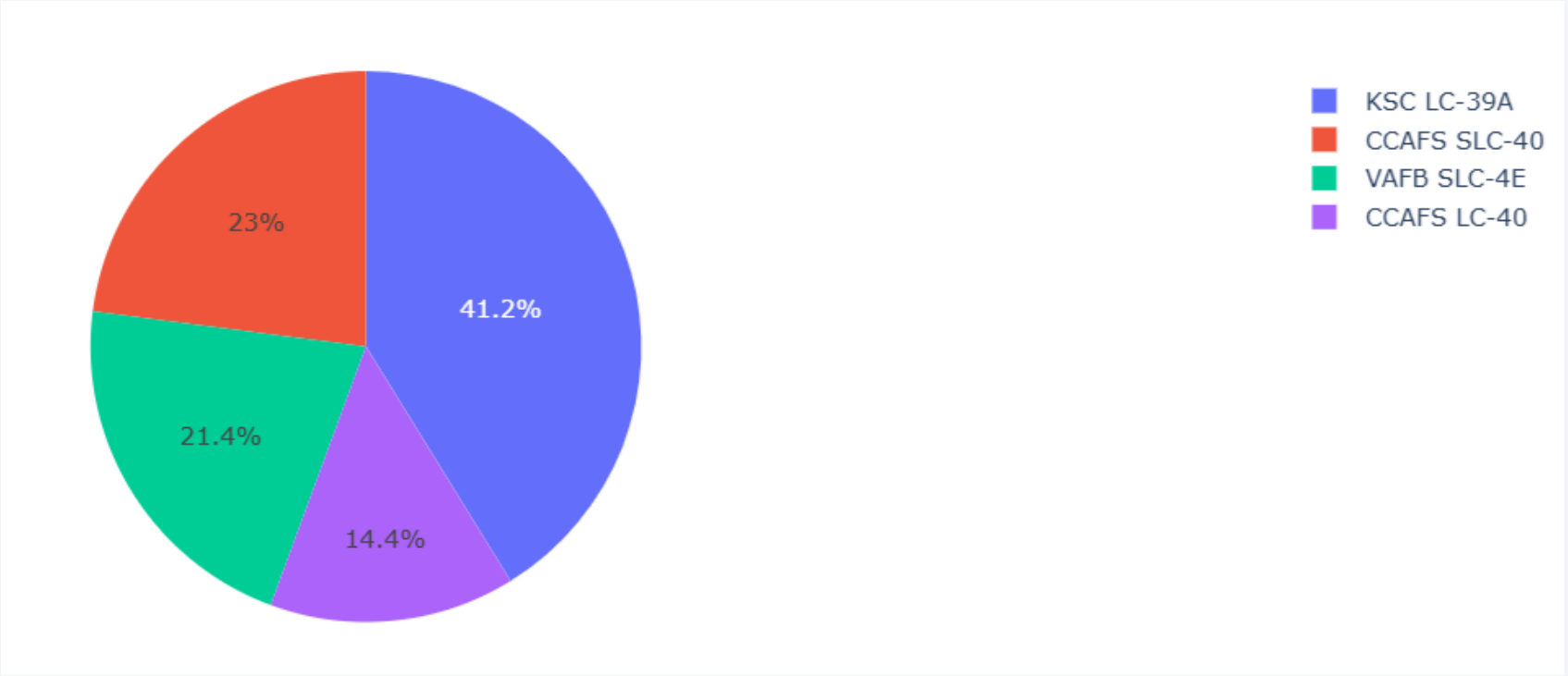
- The launch sites are a fair distance away from roads, railroad lines, and cities.




Section 4
Build a Dashboard with Plotly Dash

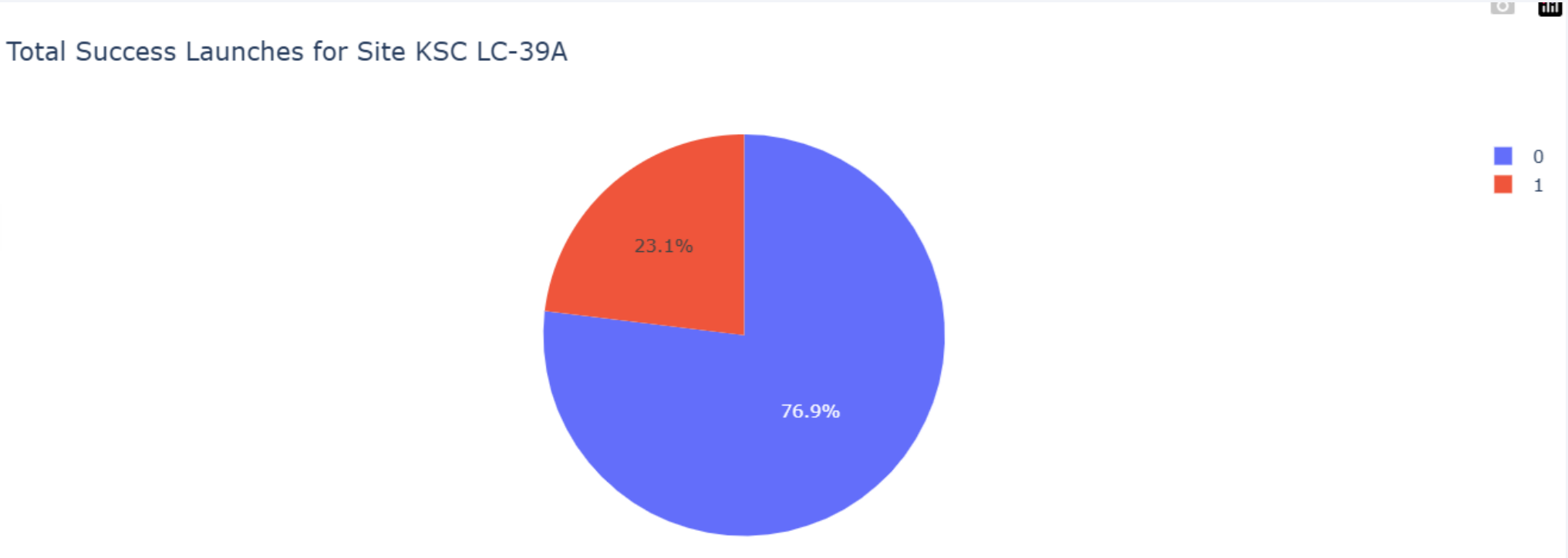
Total Successful Launches by Site

- KSC LC-39A has the largest proportion of successful launches
- CCAFS LC-40 has the smallest proportion of successful launches



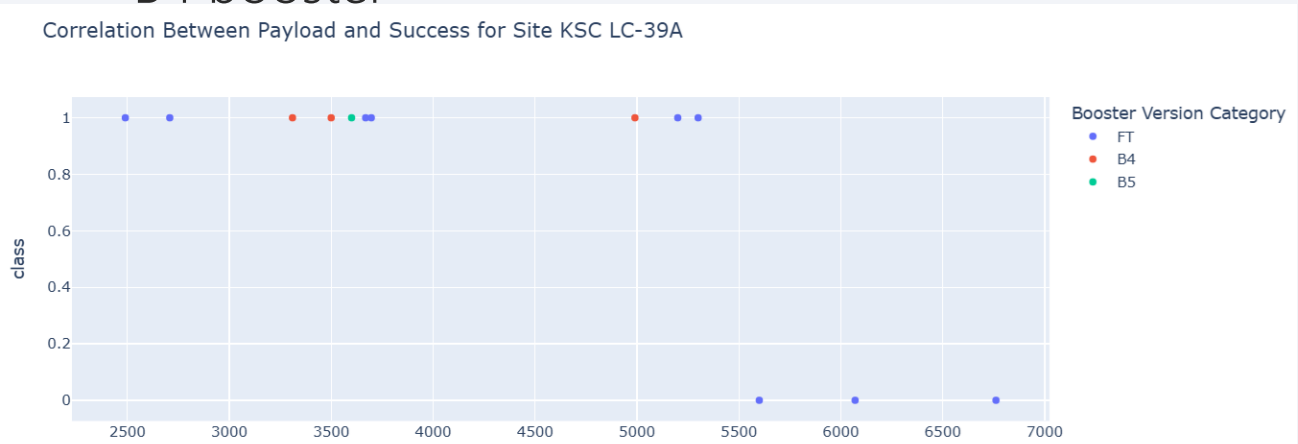
Launch site with highest success ratio

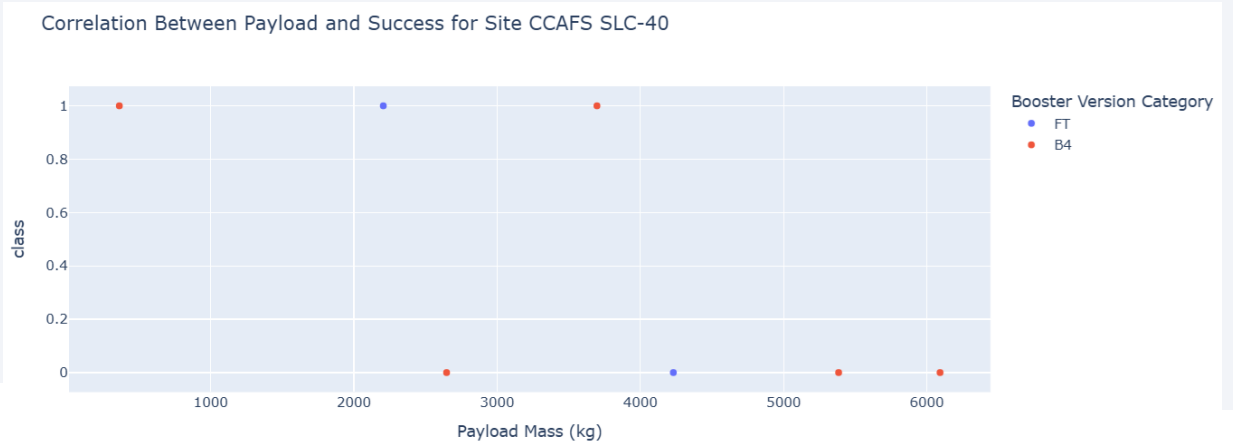
- Site KSC LC-39A has the highest success ratio, with a 76.9% success rate.

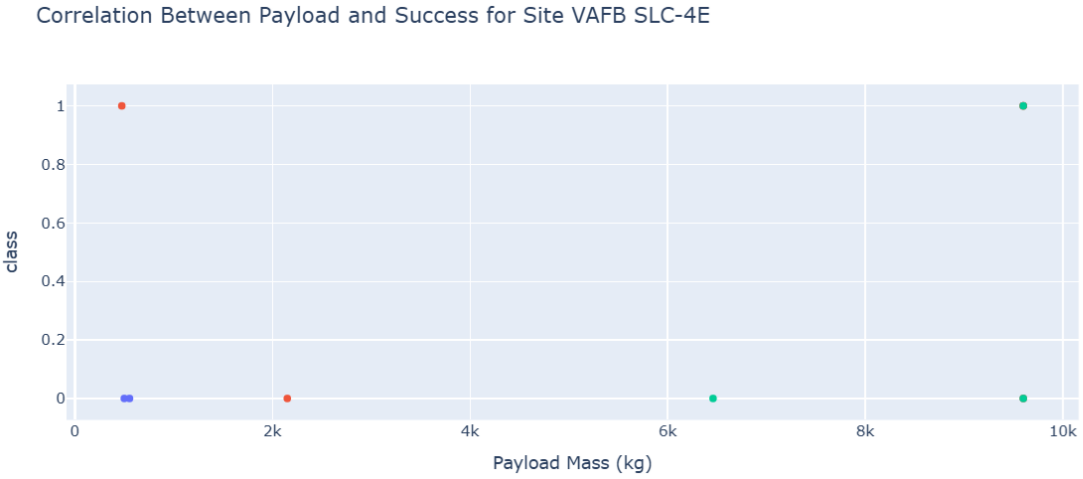


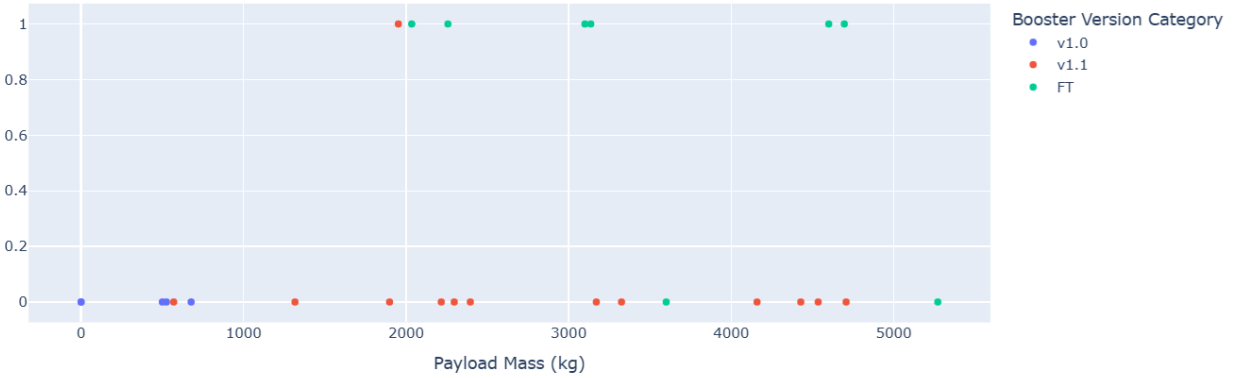
Payload vs. Launch at all Sites

- Site KSC LC-39A's successful launches have payloads of <5500kg, with most using FT boosters
- Site VAFB SLC-4E has the largest payload for a successful launch, at over 9000kg, using a B4 booster







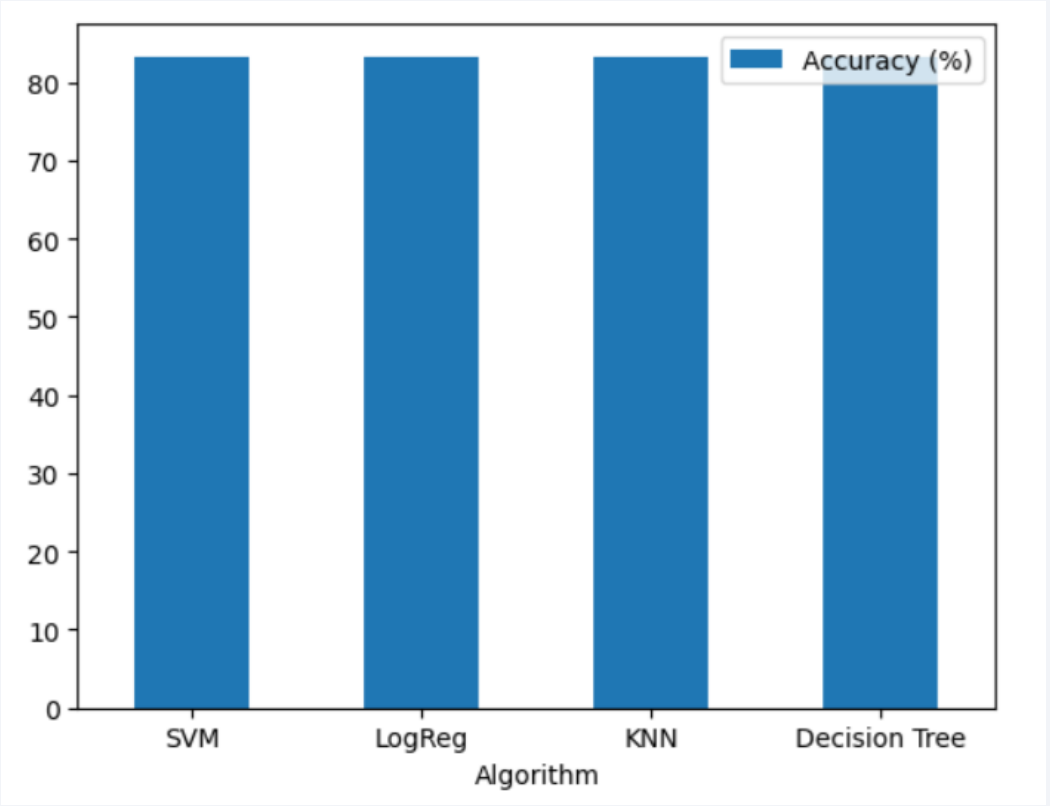


Section 5

Predictive Analysis (Classification)

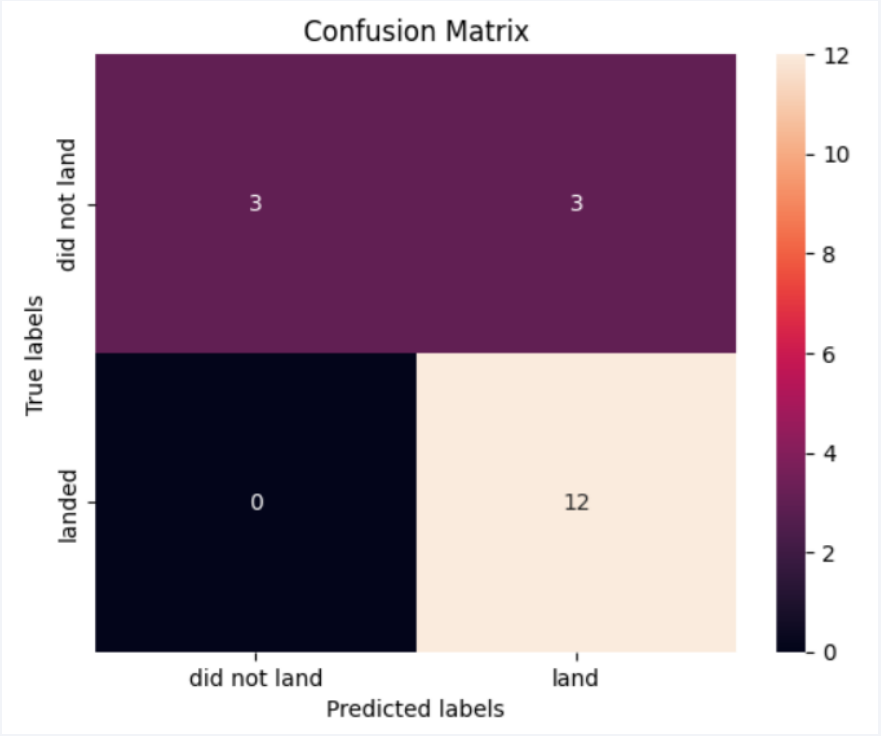
Classification Accuracy

- The models have about equal accuracy.



Confusion Matrix

- The models were able to correctly predict most successful landings.



Conclusions

- All models are useful
- During training, decision trees have highest accuracy among the models

Appendix

- Used:
- Python, numpy, seaborn, pandas, Dash, sklearn, matplotlib, Jupyter Notebook, JavaScript, SQL, BeautifulSoup, PrettyTable, Folium
- Space X API, Wikipedia

Thank you!