IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Tsuneyoshi Kamoi>
<2024.11.06>

Outline

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

Executive Summary

- Built a machine learning model
  to predict whether SpaceX will reuse the first stage
- A model with a prediction accuracy of about 83%
  was constructed.

Introduction

- The SpaceX Falcon 9 rocket launches with a cost of 62 million dollars. Because SpaceX can reuse the first stage.
- To determine the price of launch, we predict the success rate of the Space X Falcon 9 First Stage Landing.

Section 1

Methodology

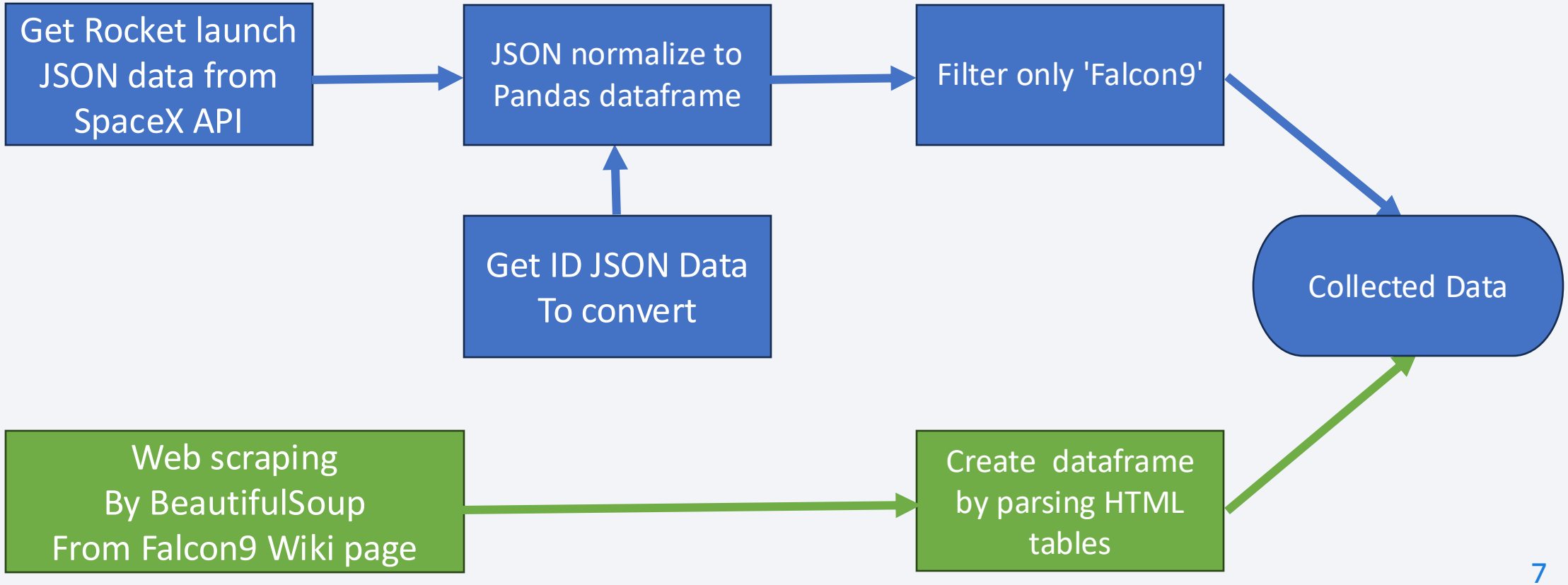
Methodology

Executive Summary

- Data collection methodology:
  - Use SpaceX launch data collected from the SpaceX REST API.
- Perform data wrangling
  - The column 'Outcome' convert to 'Classes' (0 is a bad , 1 is a good)
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Use4 method (Logistic Regression, SVM, Decision Tree, KNN)
    Build by GridSearchCV to evaluate 'accuracy'

Data Collection

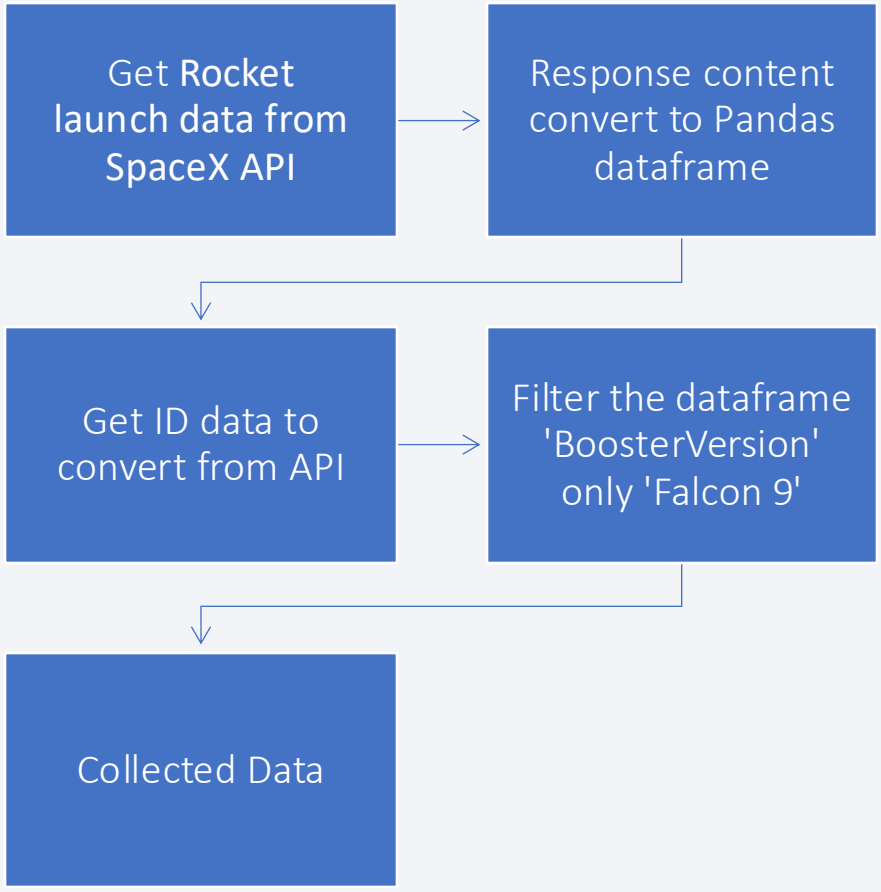
- Get request to the SpaceX API and web scraping from Falcon9 Launch Wiki page



Data Collection – SpaceX API

- Using HTTP GET Requests for Rocket Launch Data from the SpaceX API
- Response content using json() and json_normalize() convert to Pandas dataframe
- Get ID data to convert from API
- Filter the dataframe 'BoosterVersion' only 'Falcon 9'

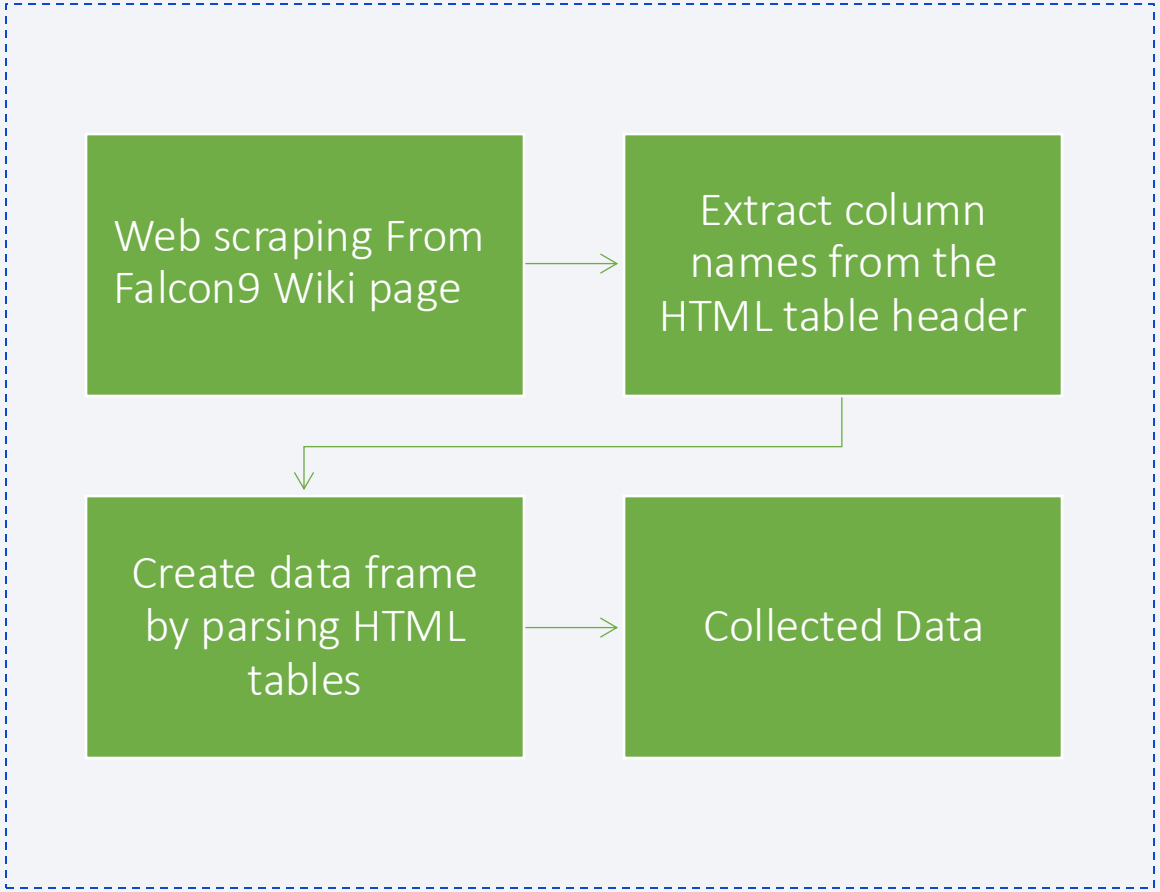
- https://github.com/pat-rap/Applied-Data-Science-Capstone/blob/main/jupyter-labs-spacex-data-collection-api.ipynb



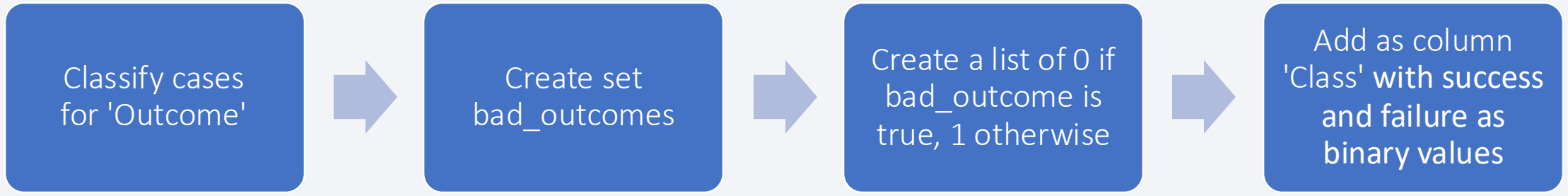
Data Collection - Scraping

- Web scraping By BeautifulSoup
  From Wikipedia 'List of Falcon9 launches'
- Extract all column names from the HTML table header
- Create a data frame by parsing HTML tables

- https://github.com/pat-rap/Applied-Data-Science-Capstone/blob/main/jupyter-labs-webscraping.ipynb

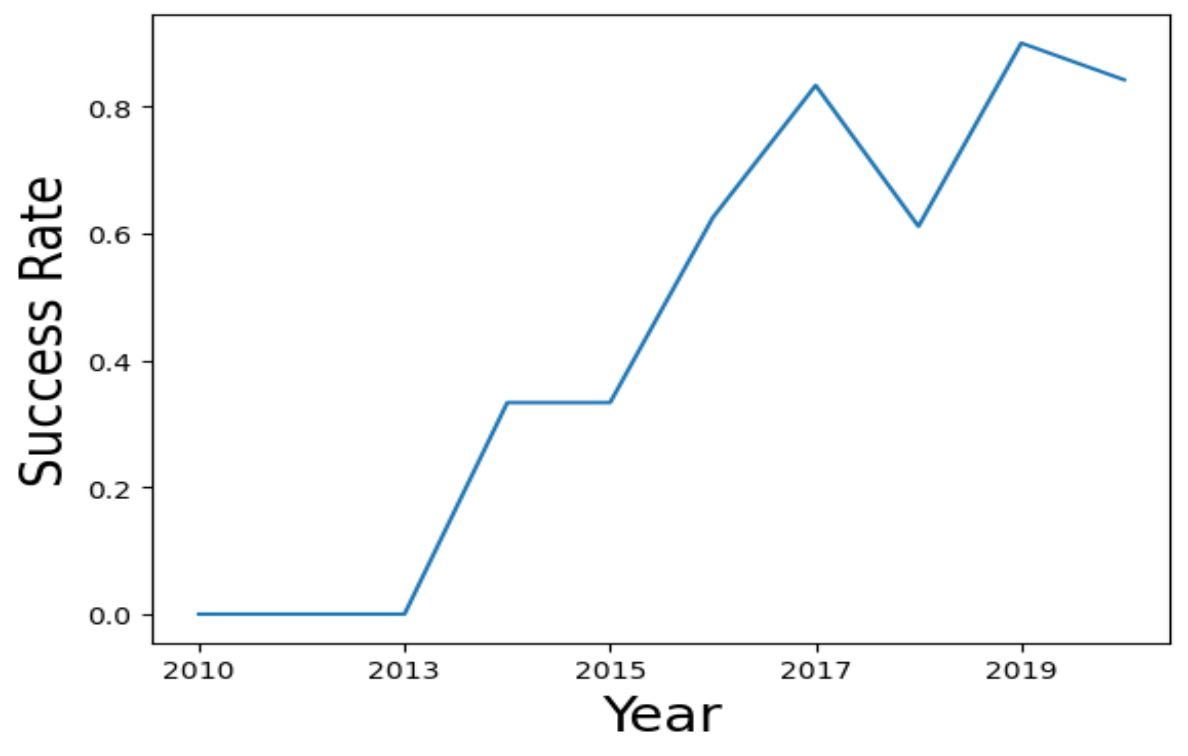


Data Wrangling



- https://github.com/pat-rap/Applied-Data-Science-Capstone/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization



- Sucess rate since 2013 kept increasing till 2020
- https://github.com/pat-rap/Applied-Data-Science-Capstone/blob/main/jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb

EDA with SQL

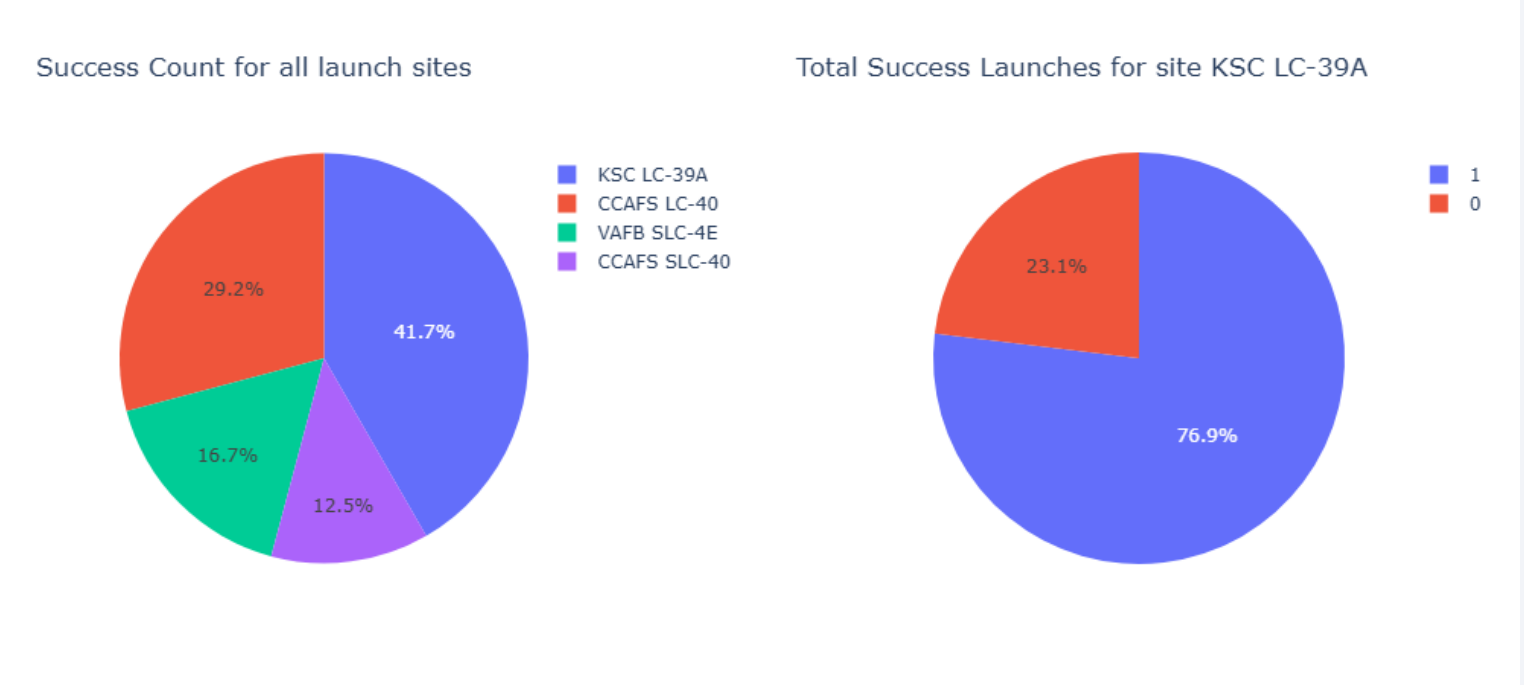
- Names of the unique launch sites in the space mission
- 5 records where launch sites begin with the string 'CCA'
- Total payload mass carried by boosters launched by NASA (CRS)
- Average payload mass carried by booster version 'F9 v1.1'
- First succesful landing outcome in ground pad was achieved
- Names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- Total number of successful and failure mission outcomes
- Names of the booster_versions which have carried the maximum payload mass
- List the records month , failure landing_outcomes in drone ship ,booster versions, launch_site in year 2015
- Count of landing outcomes between the date 2010-06-04 and 2017-03-20

- https://github.com/pat-rap/Applied-Data-Science-Capstone/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

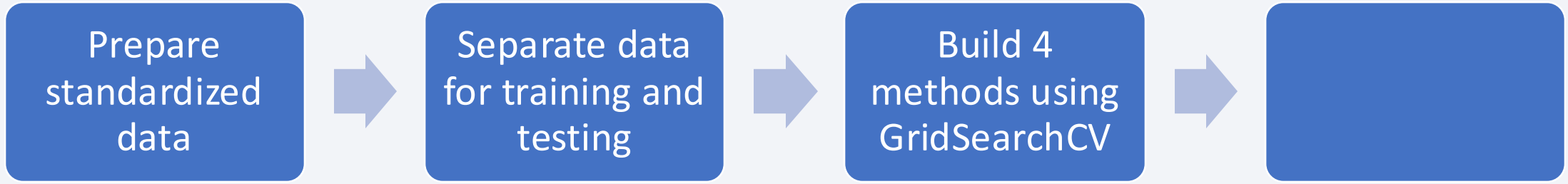
- Three of the four Falcon9 launch sites are concentrated in Florida,. near the equator, which is favorable for rocket launches.
- https://github.com/pat-rap/Applied-Data-Science-Capstone/blob/main/lab_jupyter_launch_site_location.jupyterlite.ipynb

Build a Dashboard with Plotly Dash



- Of the 4 launch sites in Falcon9, LC-39A has a high success rate and a high success rate itself.
- https://github.com/pat-rap/Applied-Data-Science-Capstone/blob/main/spacex_dash_app.py

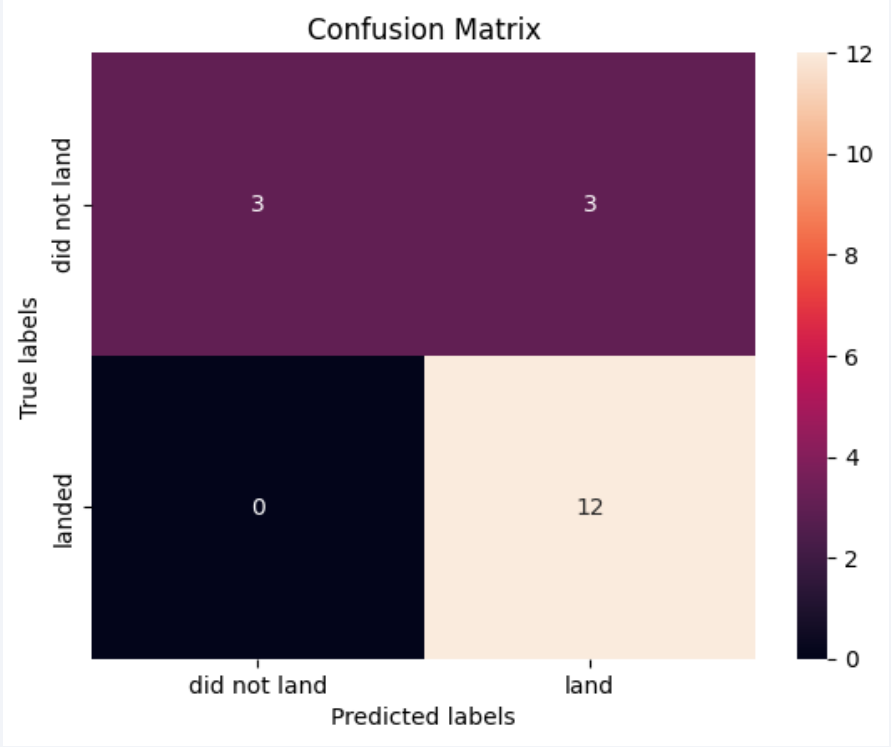
Predictive Analysis (Classification)



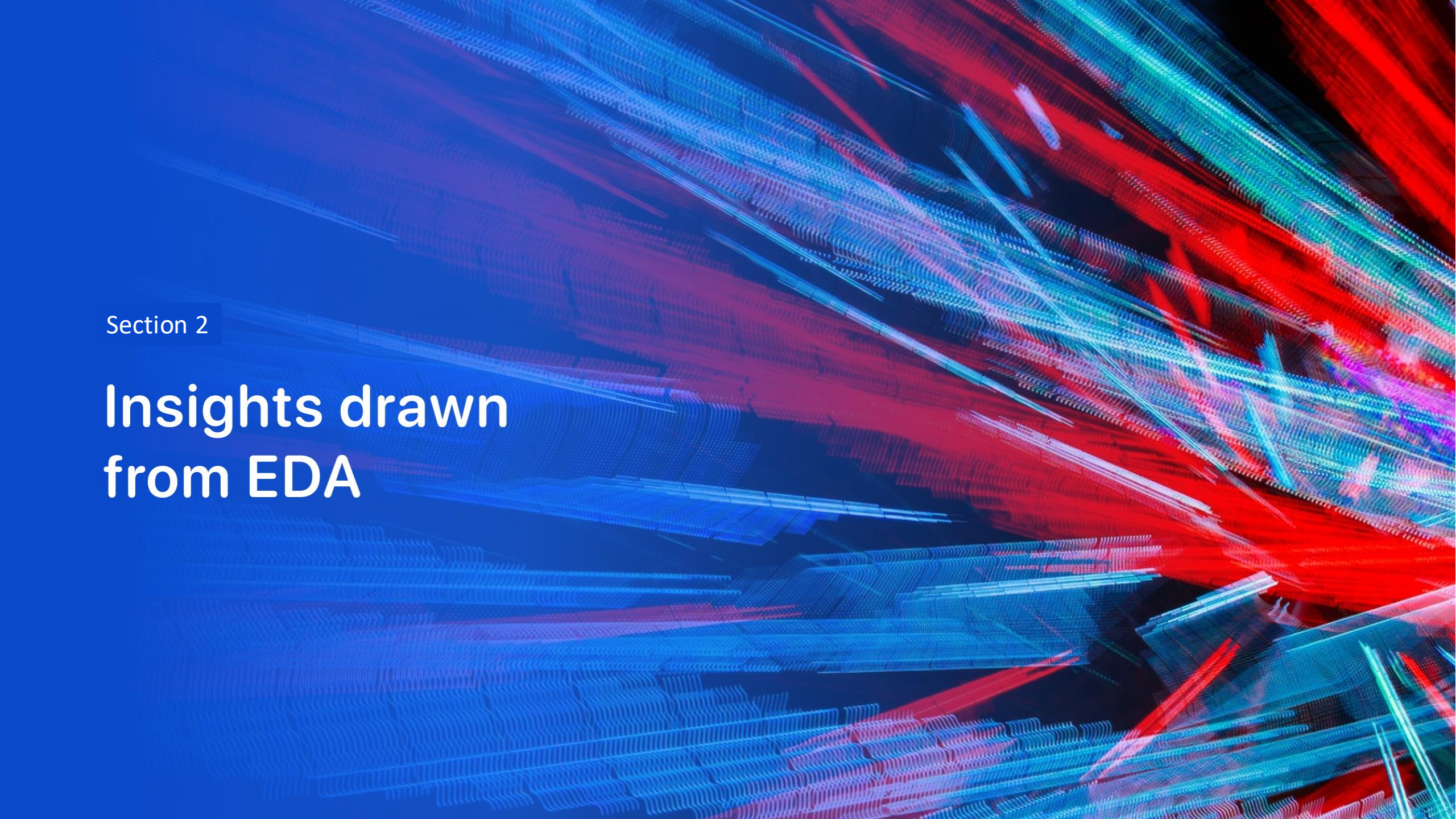
- Build 4 method(Logistics Regression, SVM, Decision Tree, KNN)
- Evaluated 4 methods using Confusion Matrix and Accuracy. all models perform the same.

- https://github.com/pat-rap/Applied-Data-Science-Capstone/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

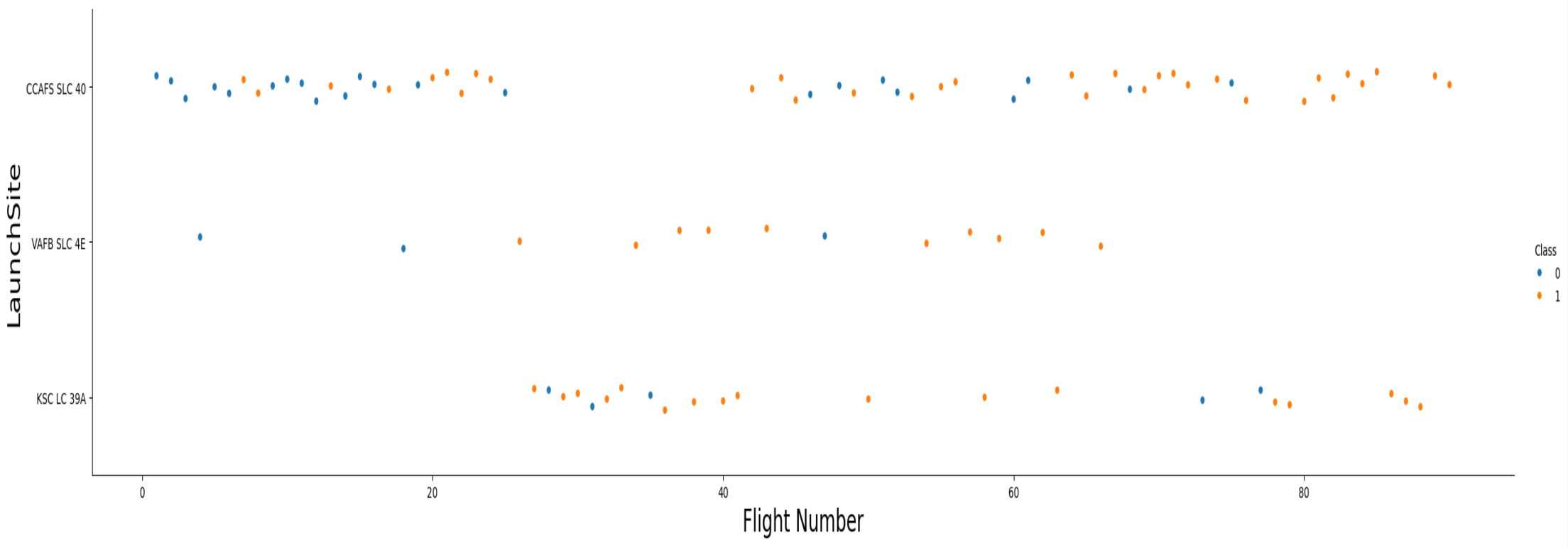
Results



- This is the ConfusionMatrix of our method. Out of 12 successful landing cases in the test data, all were correctly predicted, but 3 out of 6 failures were correctly predicted and 3 were incorrectly predicted as successful, so the prediction accuracy was about 83%.

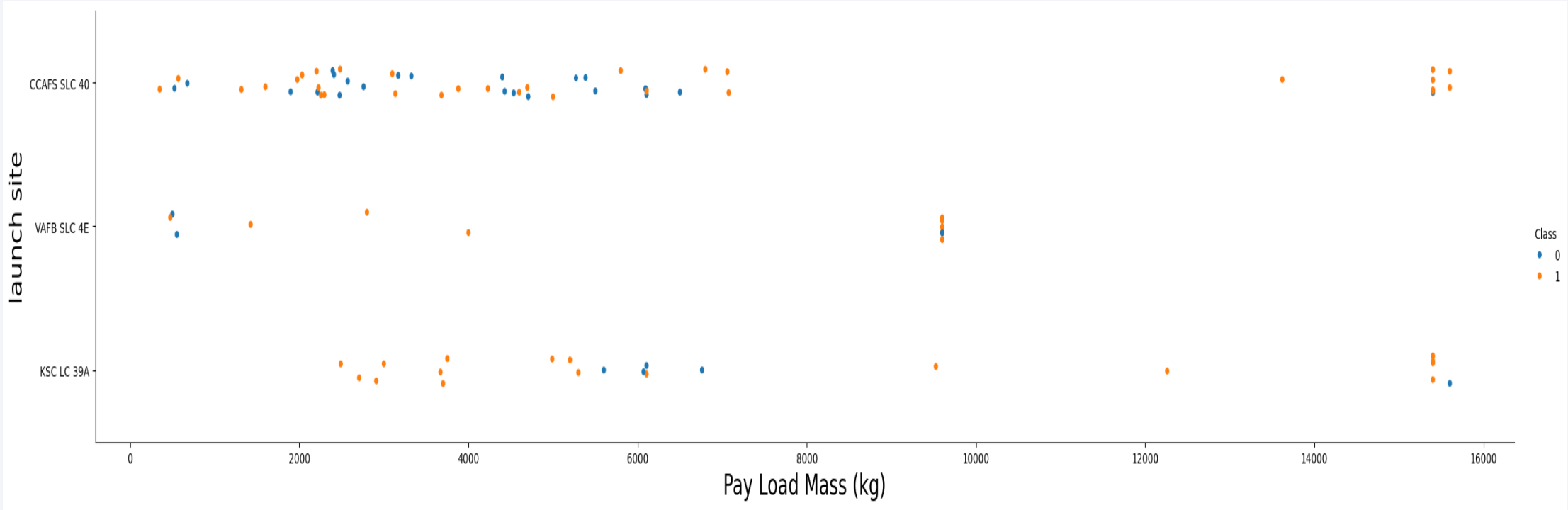

Section 2
Insights drawn from EDA

Flight Number vs. Launch Site



- As a general trend, the blue of failure stands out in the early stage, but the orange of success tends to increase as you gain flight experience later.

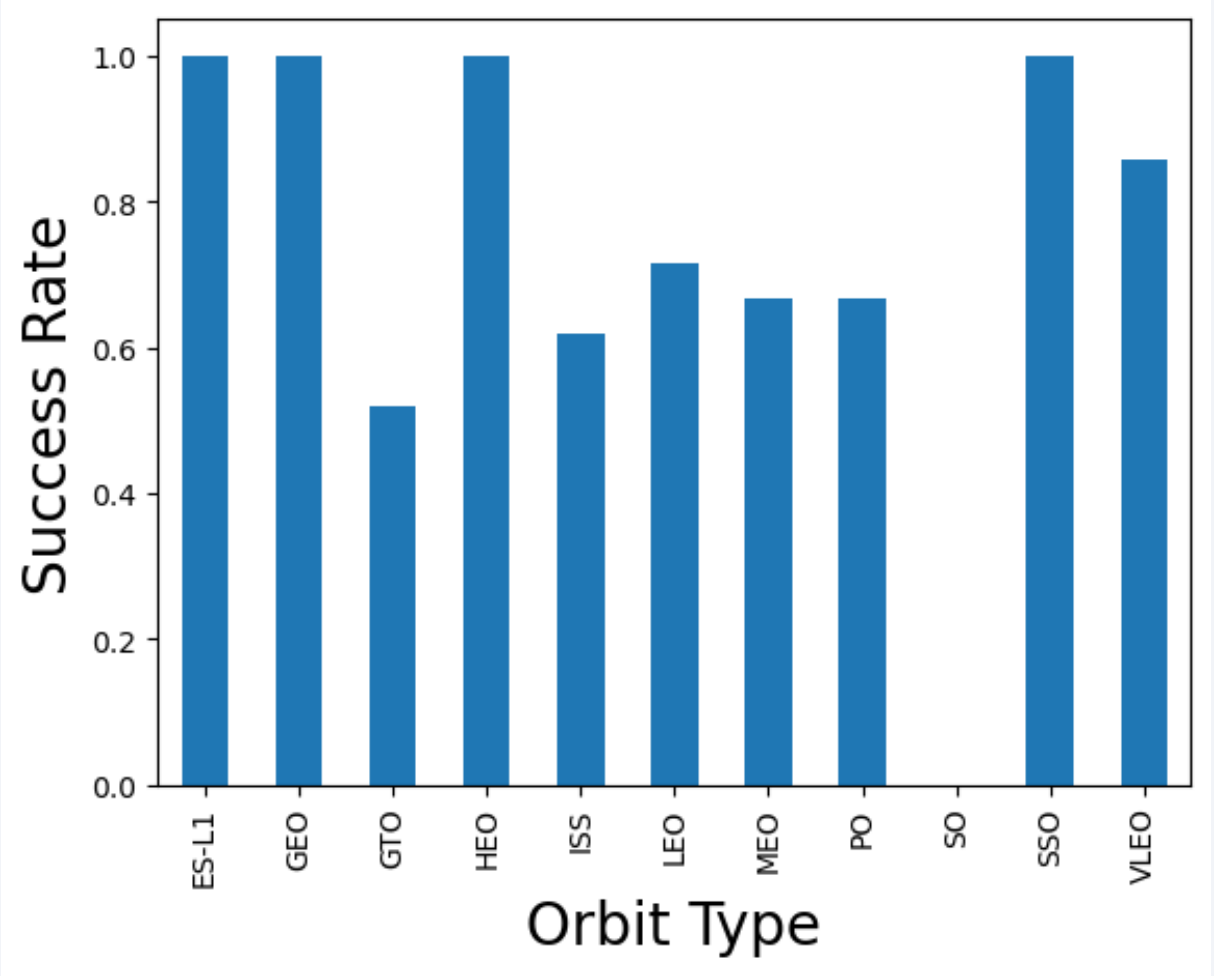
Payload vs. Launch Site



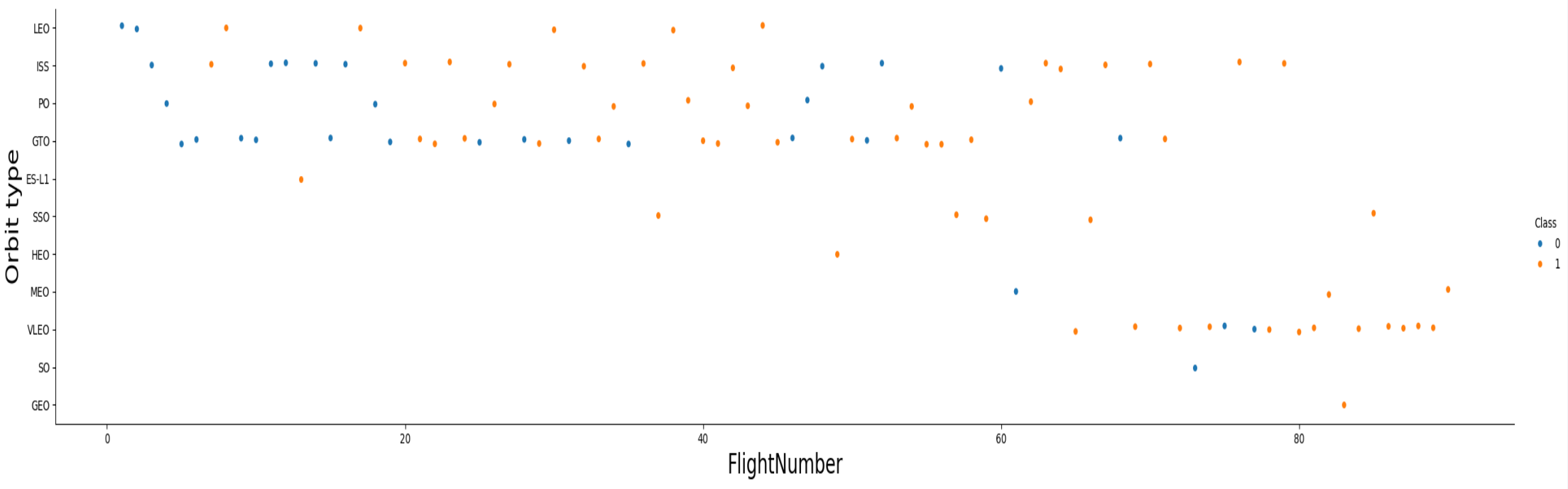
- Observe Payload Mass Vs. Launch Site scatter point chart you will find for the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000).

Success Rate vs. Orbit Type

- Among Orbits, ES-L1 GEO HEO SSO has the highest success rate.
- LEO ISS PO GTO, which was tried earlier, shows a lower success rate.

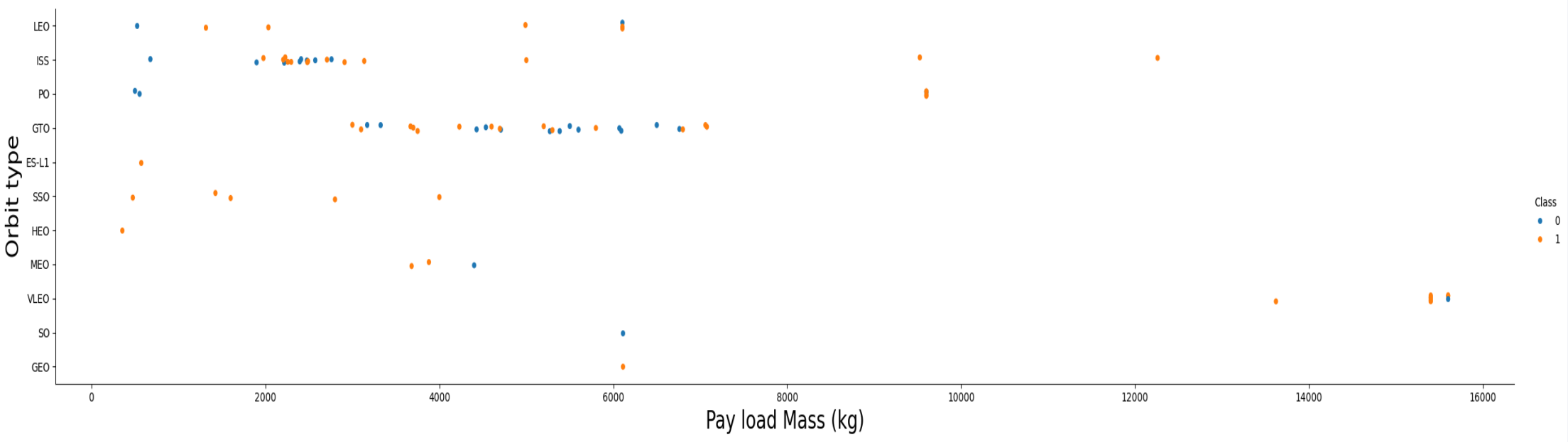


Flight Number vs. Orbit Type



- In LEO orbits, success seems to be related to the number of flights. Conversely, in GTO orbits, there seems to be no relationship between number of flights and success.

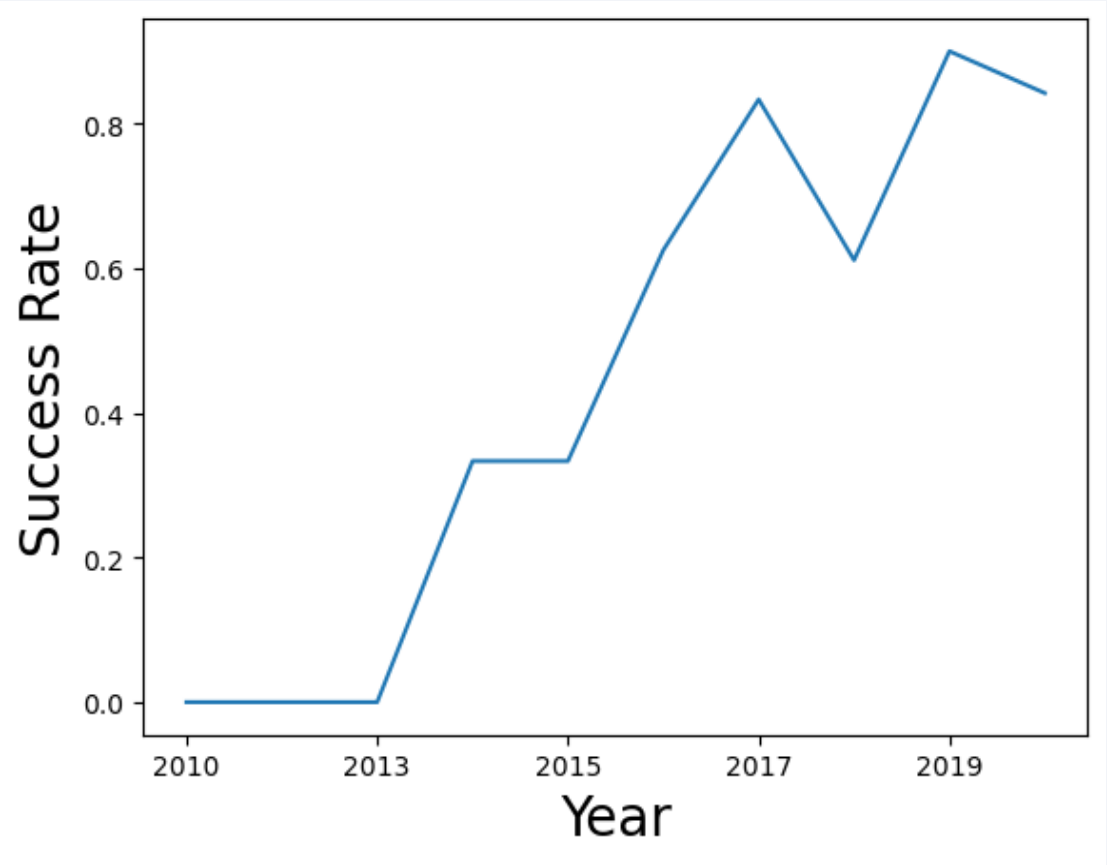
Payload vs. Orbit Type



- With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.
- However, for GTO, it's difficult to distinguish between successful and unsuccessful landings as both outcomes are present.

Launch Success Yearly Trend

- Sucess rate since 2013 kept increasing till 2020



All Launch Site Names

- Use DISTINCT in SQL to get a unique list

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

Launch Site Names Begin with 'CCA'

- Use the wildcard characters% and LIKE in SQL to get a list of launch site data starting at 'CCA'
- Retrieve the first five rows using the LIMIT clause

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

- Use SUM function in SQL to find the sum. Use WHERE to restrict Customer to NASA (CRS).

Total Payload Mass
45596

Average Payload Mass by F9 v1.1

- Use AVG function in SQL to find the average.

booster version F9 v1.1 Average Payload Mass
2928.4

First Successful Ground Landing Date

- Use MIN function in SQL to retrieve the minimum of a value.

first succesful landing outcome in ground pad
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- Use BETWEEN operator in SQL to specify a range.
- restrict the landing to a successful Drone ship landing, and specify PAYLOAD MASS from 4000 to 6000,

Booster_Version	PAYLOAD_MASS__KG_
F9 FT B1022	4696
F9 FT B1026	4600
F9 FT B1021.2	5300
F9 FT B1031.2	5200

Total Number of Successful and Failure Mission Outcomes

- Use COUNT function in SQL to aggregate the number of records.
- To count successes and failures, include a Success or Failure clause in the Mission_outcome column or use the wildcard character% and use the CASE WHEN THEN expression.
- Combine the Success and Failure aggregates with UNION as a single table.

Outcome	count
Failure	1
Success	100

Boosters Carried Maximum Payload

- Use the MAX function in SQL to retrieve the maximum value.
- Use that value as a subquery in the WHERE clause, and DISTINCT non-duplicate list to retrieve the unique booster version that carried the maximum payload.

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

- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

- use Substr function in SQL to retrieve specific parts of a string. The first four characters of the Date column are used as the year in the WHERE clause, and the sixth through two characters are used as the month in the SELECT clause.

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

- 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 GROUP BY clause uniquely lists the Cases in Landing_Outcome. The COUNT function aggregates the number of records, BETWEEN restricts the date range, and ORDER BY DESC sorts from highest to lowest.

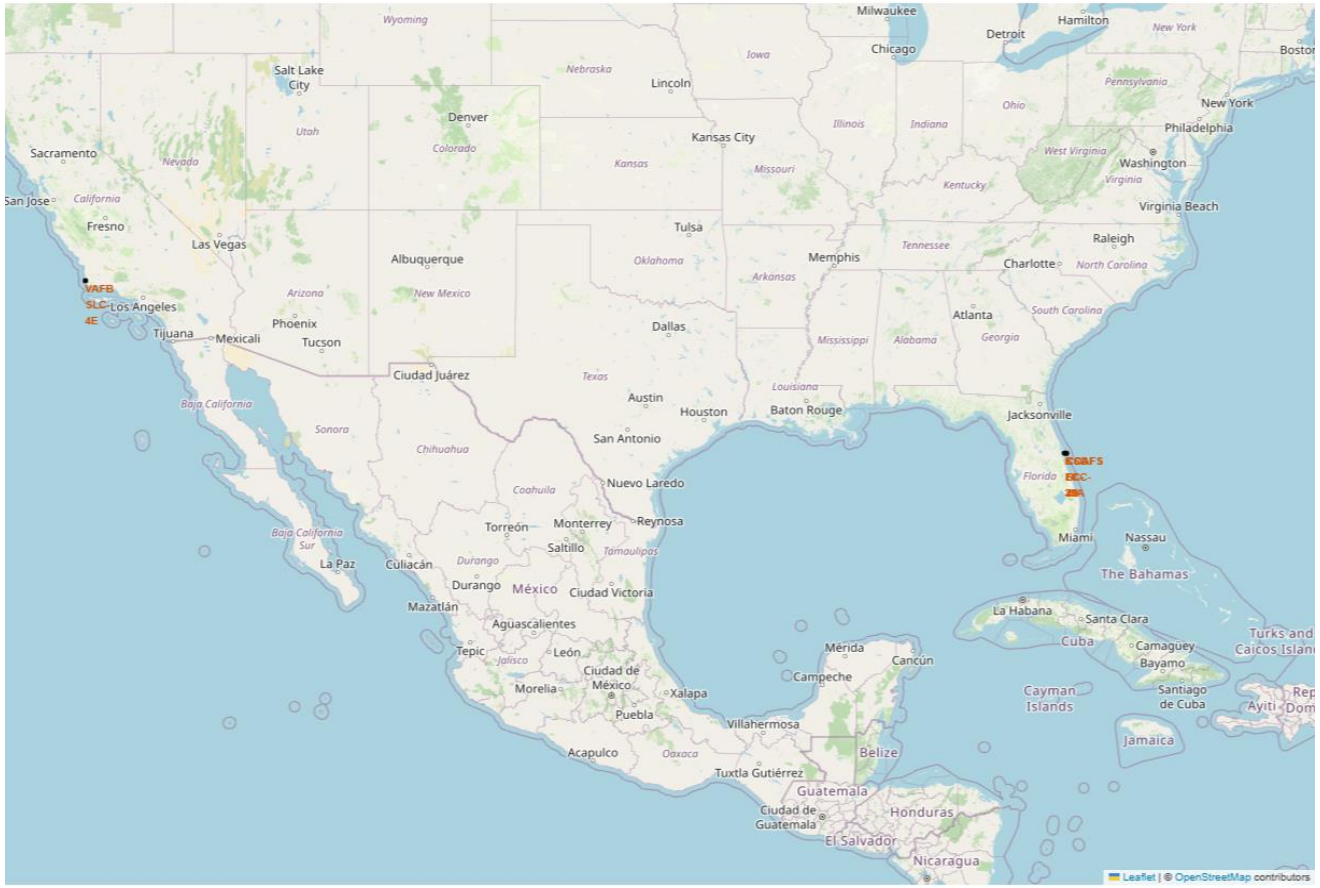
Landing_Outcome	Count
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

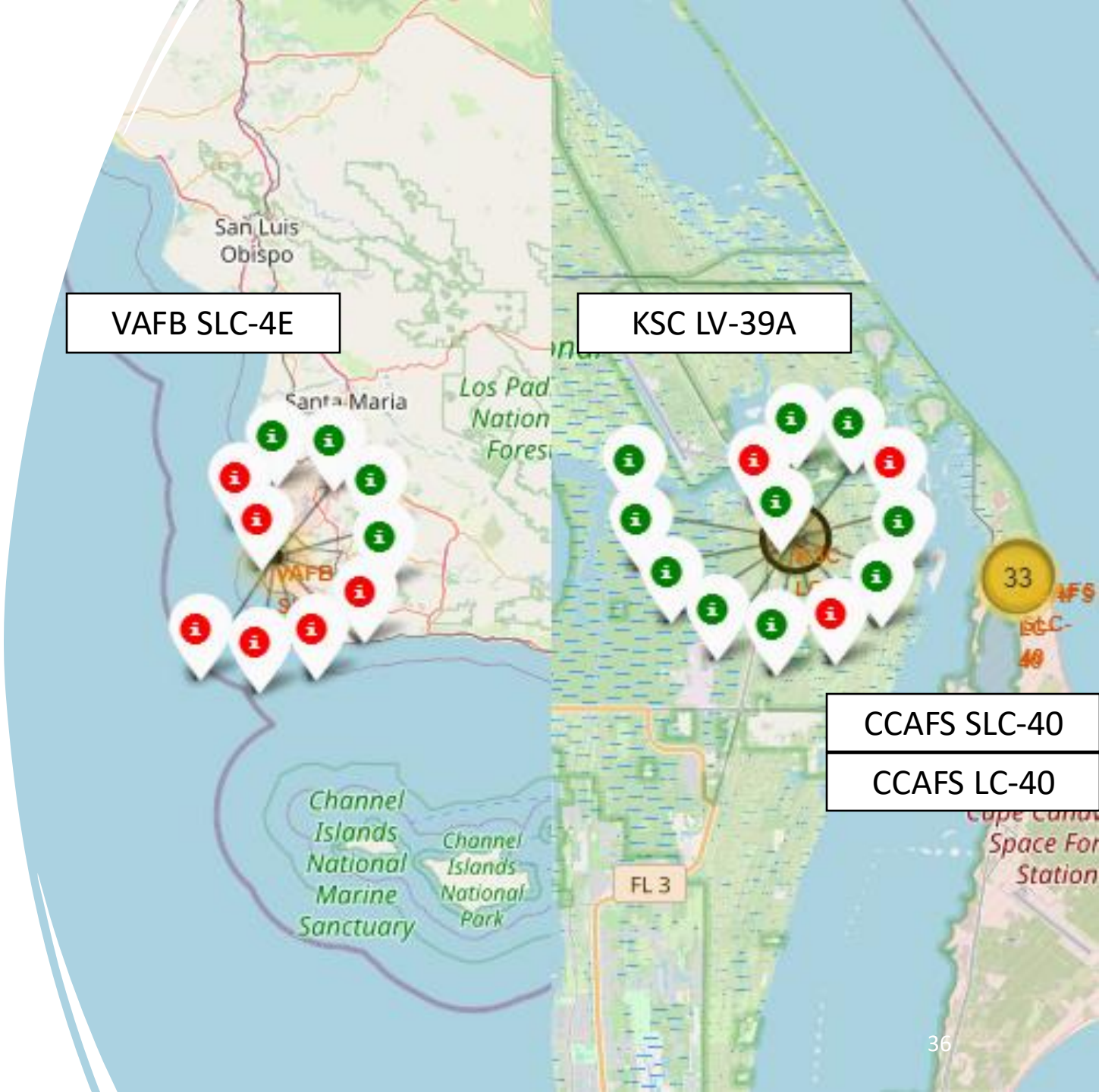
<Launch Site Location Map>

- All launch sites are located near the coast at relatively low latitudes.
- The sea in the direction of polar or geostationary orbit ensures safety, avoiding urban areas.
- The proximity to the equator allows us to take advantage of the Earth's rotation.



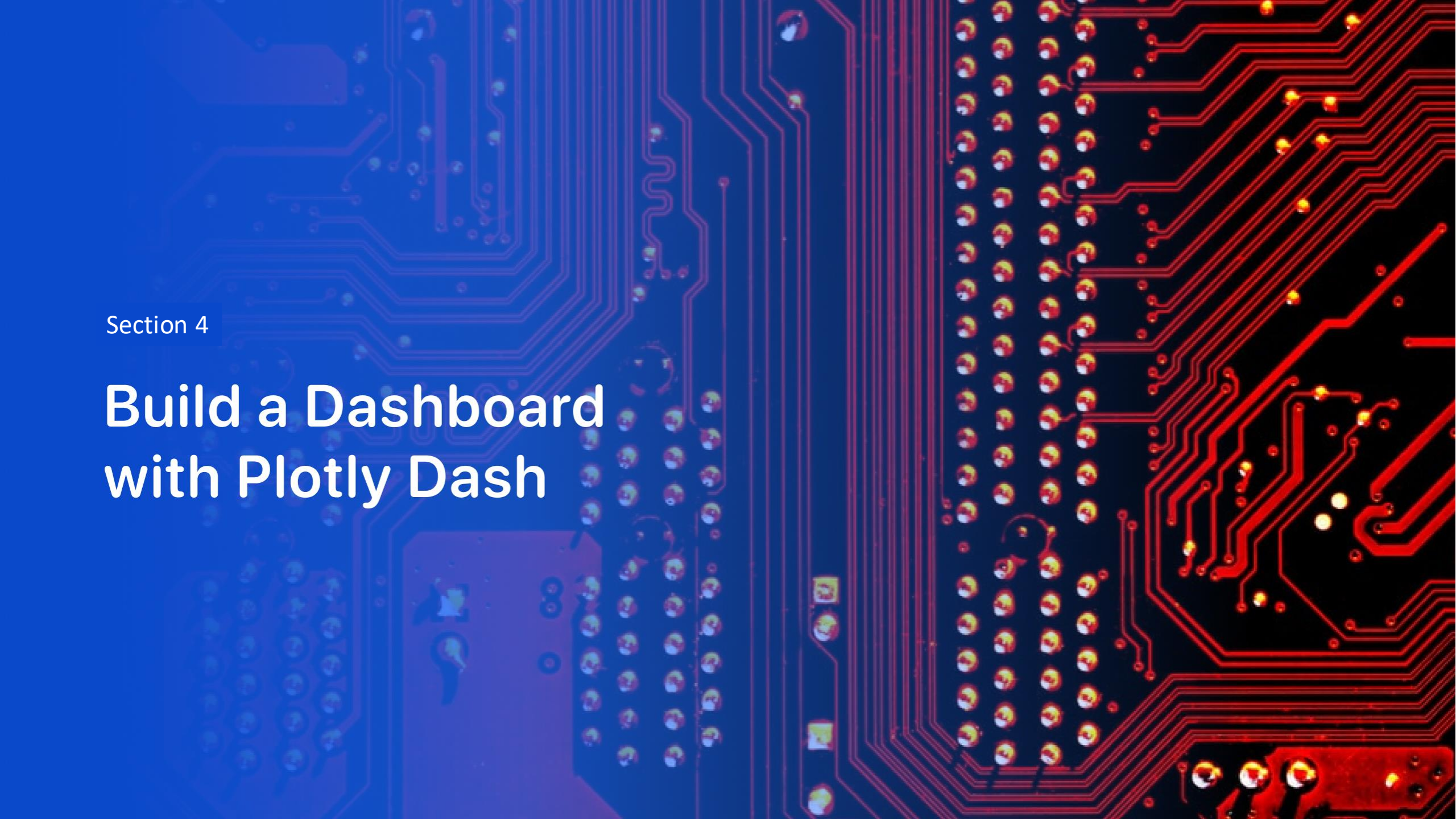
<Mark map of each site launch success/failure>

- The relatively early VAFB SLC-4E shows a lot of failures.
- The KSC LC-39A, which has been used since the middle of the mission, shows a lot of successes.



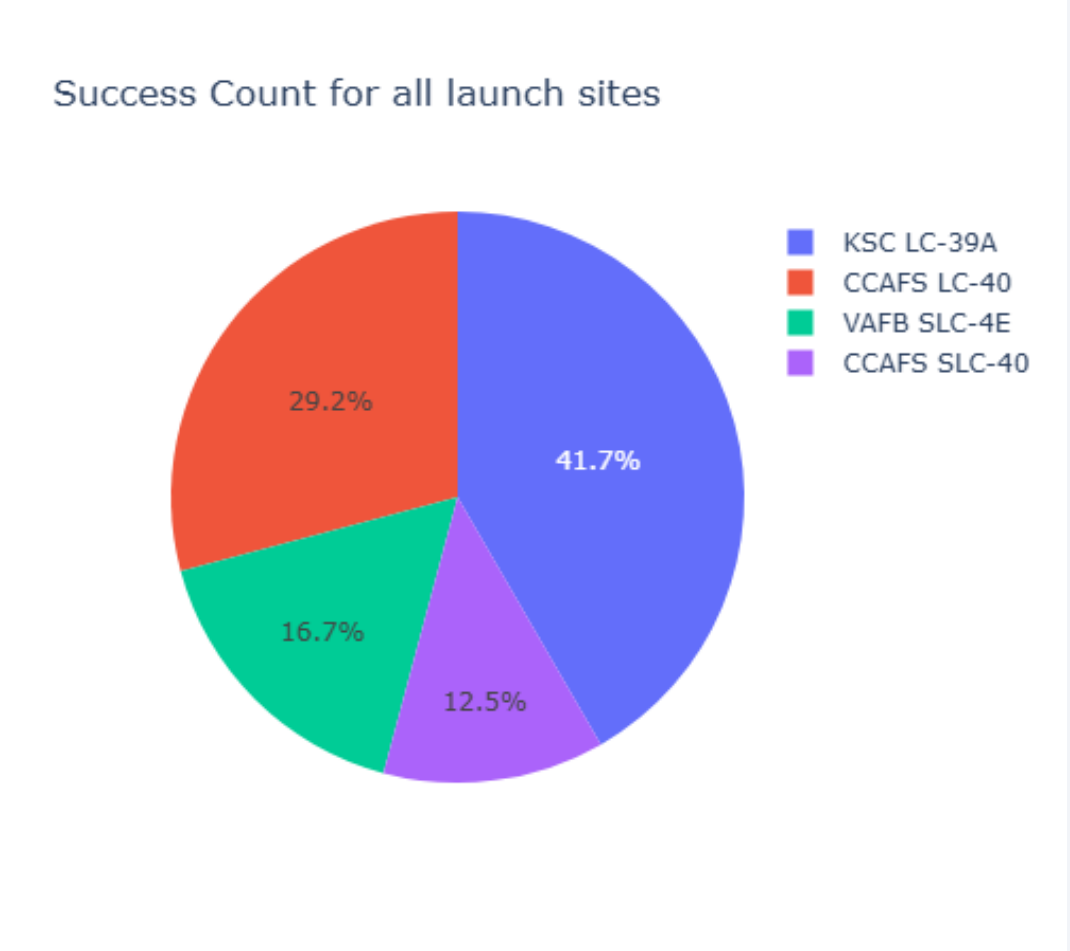
<Map of the distance from the launch site>

- close to the coastline, the railways, the highways, but far from the city


Section 4
Build a Dashboard with Plotly Dash

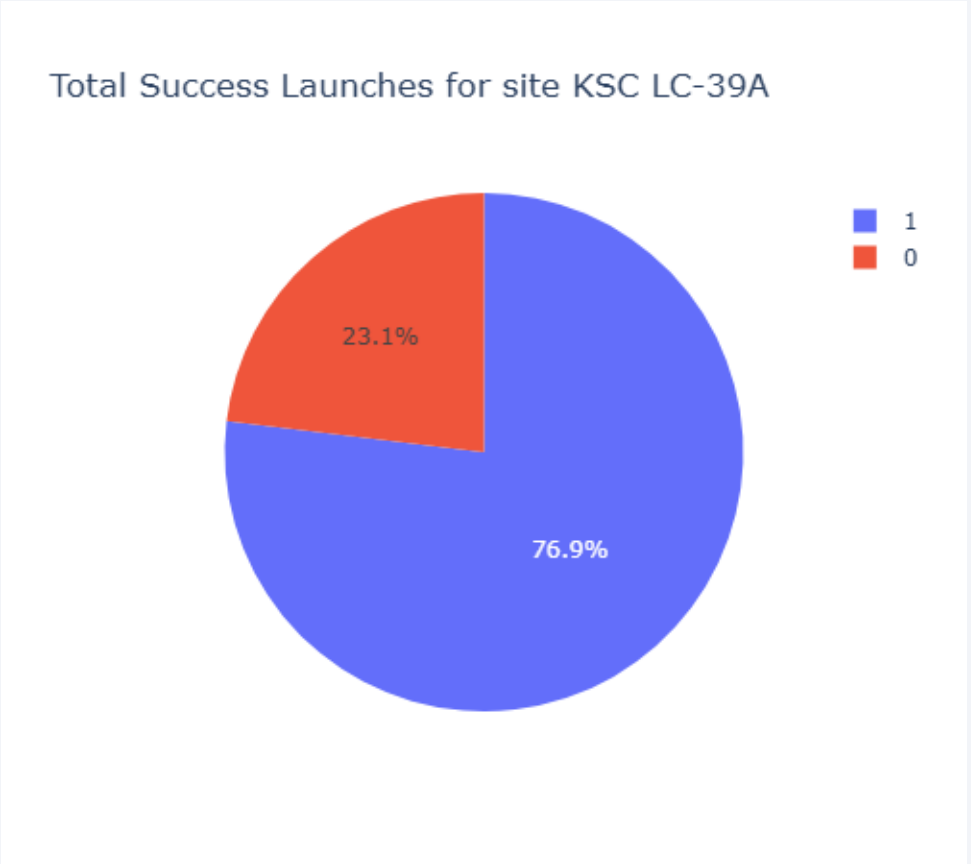
<Success Count for all launch sites>

- Two launch sites, KSC LC-39A and CCAFS LV-40, account for most of the success.



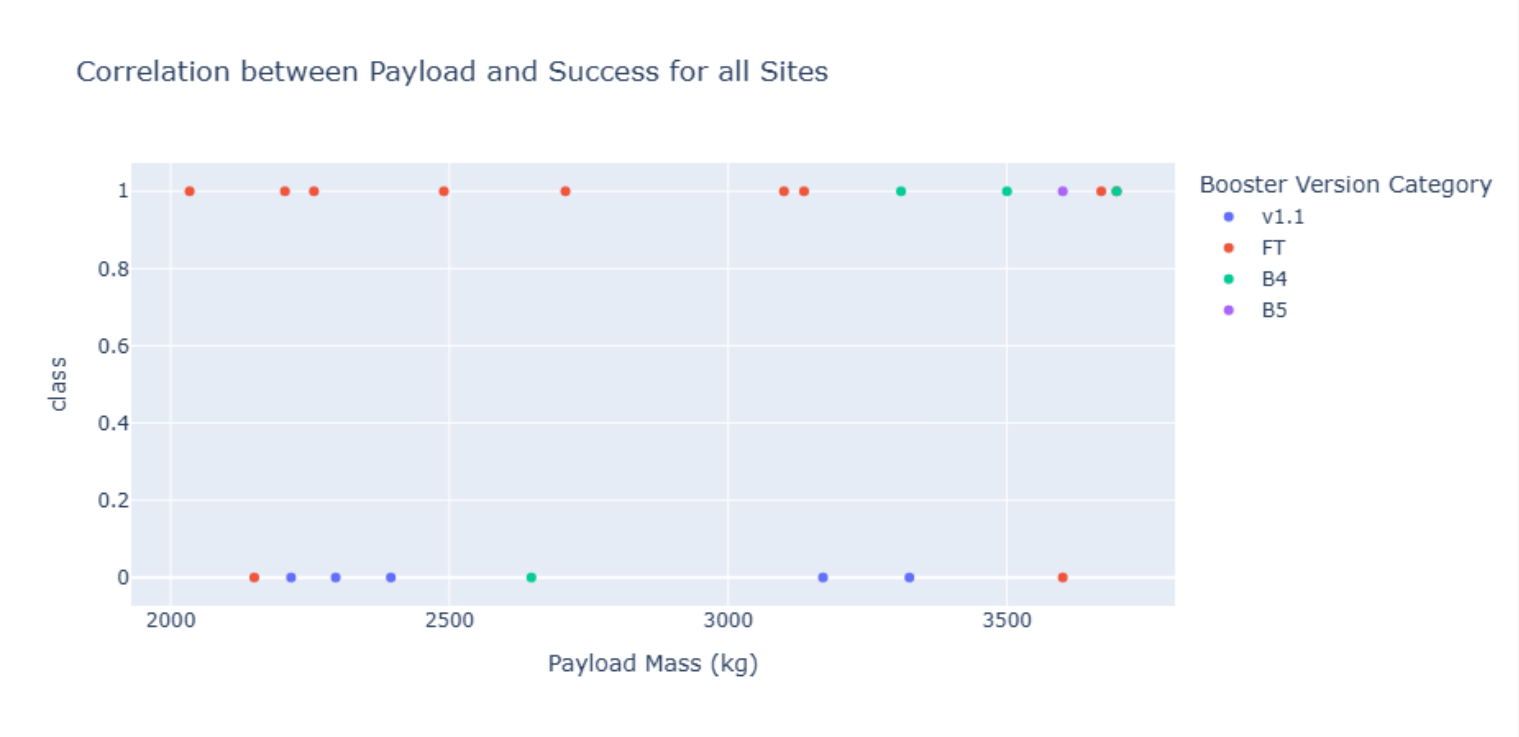
<Total Success Launches for site KSC LC-39A>

- At KSC LC-39A, the launch site with the highest success rate, the success rate is about 77%.



<Dashboard Screenshot 3>

- For payload mass between 2000 and 4000, the booster versions of FT and B4 are more successful.

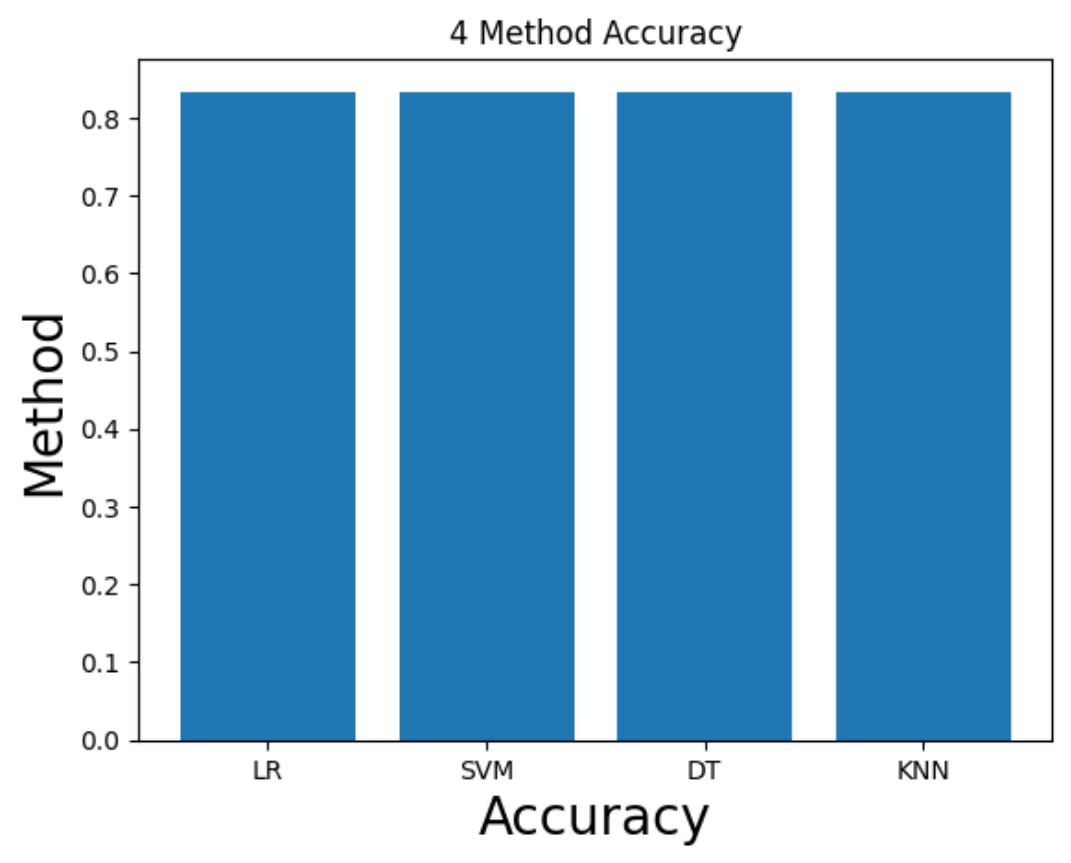


Section 5

Predictive Analysis (Classification)

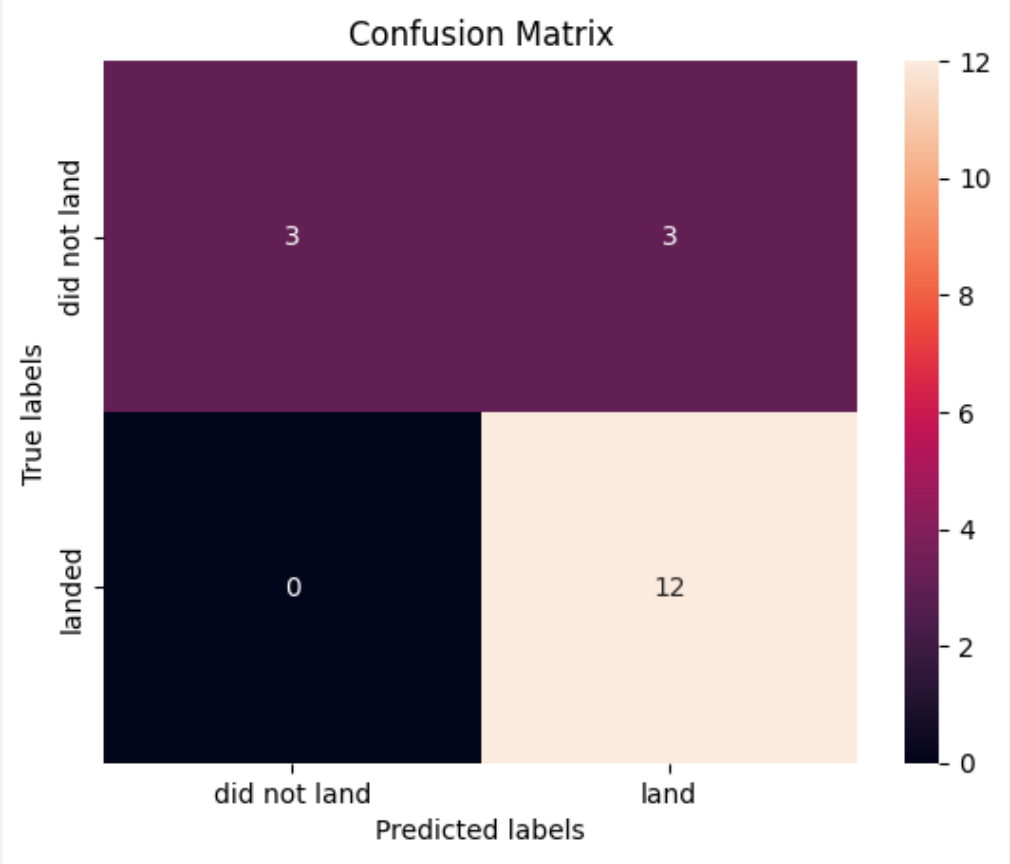
Classification Accuracy

- All four methods examined this time (Logistic regression, Support vector machine, Decision tree, k nearest neighbors) showed the same accuracy.



Confusion Matrix

- This is the ConfusionMatrix of our method. Out of 12 successful landing cases in the test data, all were correctly predicted, but 3 out of 6 failures were correctly predicted and 3 were incorrectly predicted as successful, so the prediction accuracy was about 83%.

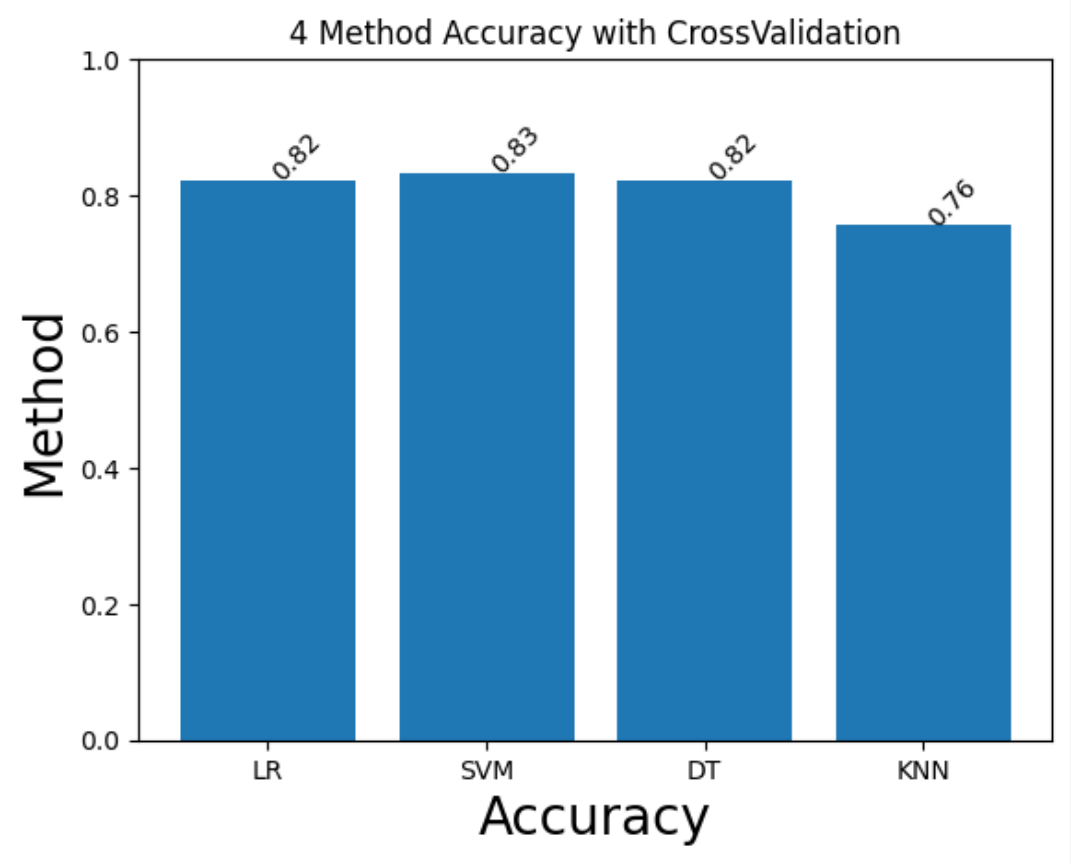


Conclusions

- The prediction model we built predicts success and failure with about 83% accuracy.
- The success rate itself is on the rise every year.
- The success rate varies depending on the launch trajectory.

Appendix

- 4 Method Accuracy with CrossValidation



Thank you!