IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Hubert Kalaus

5.11.2024

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection through API
  - Data Collection with Web Scraping
  - Data Wrangling
  - Exploratory Data Analysis with SQL
  - Exploratory Data Analysis with Data Visualization
  - Interactive Visual Analytics with Folium & Dash
  - Machine Learning Prediction
- Summary of all results
  - Exploratory Data Analysis results
  - Interactive analytics screenshots
  - Predictive Analytics result

Introduction

- Project background and context
  - Commercial space age: space travel becomes accessible
  - Through: cost-effective SpaceX launches with reusable Falcon 9 rockets.
    - Reusability of Falcon 9's first stage significantly reduces launch costs.
- Problems you want to find answers
  - Explore launch data, identify key contributors to success
  - Project goal: Develop a model to predict SpaceX first stage reusability for Space Y, a new competitor.

Section 1

Methodology

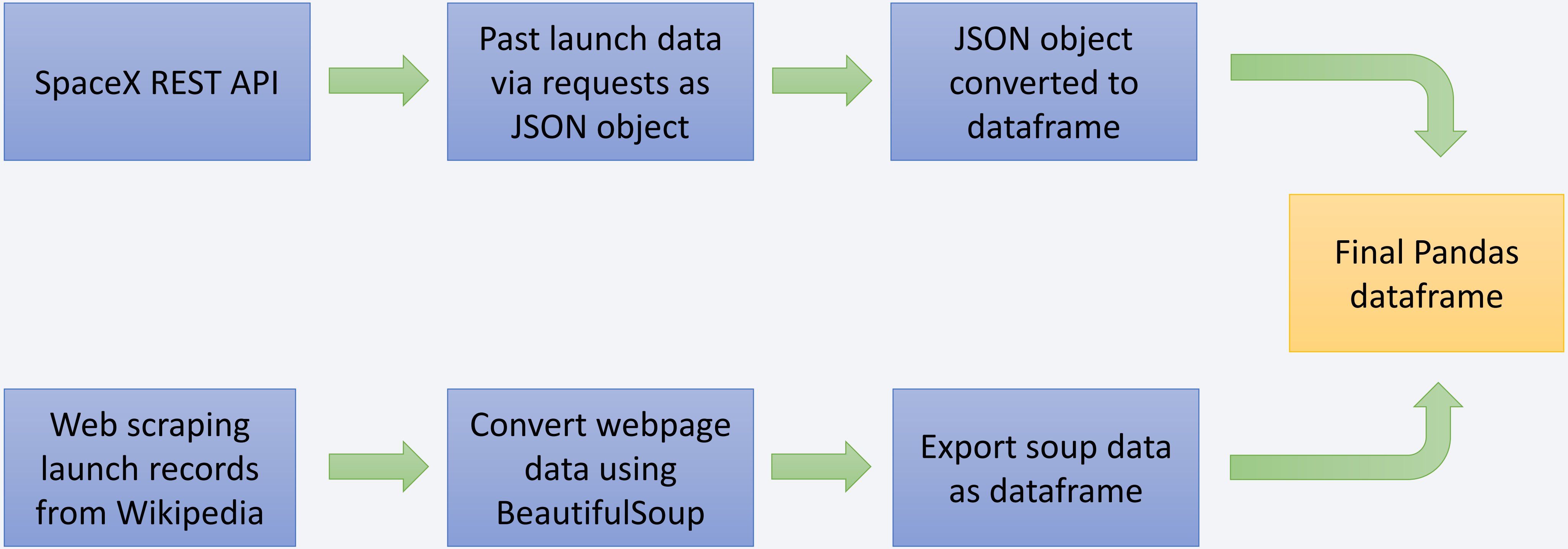
Methodology

Executive Summary

- Data collection methodology:
  - Flight data from the SpaceX Rest API
  - additional information from wikipedia
- Perform data wrangling
  - JSON objects and html tables were converted to Pandas dataframes
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Evaluation of machine learning models to predict Falcon 9 lauch success

Data Collection

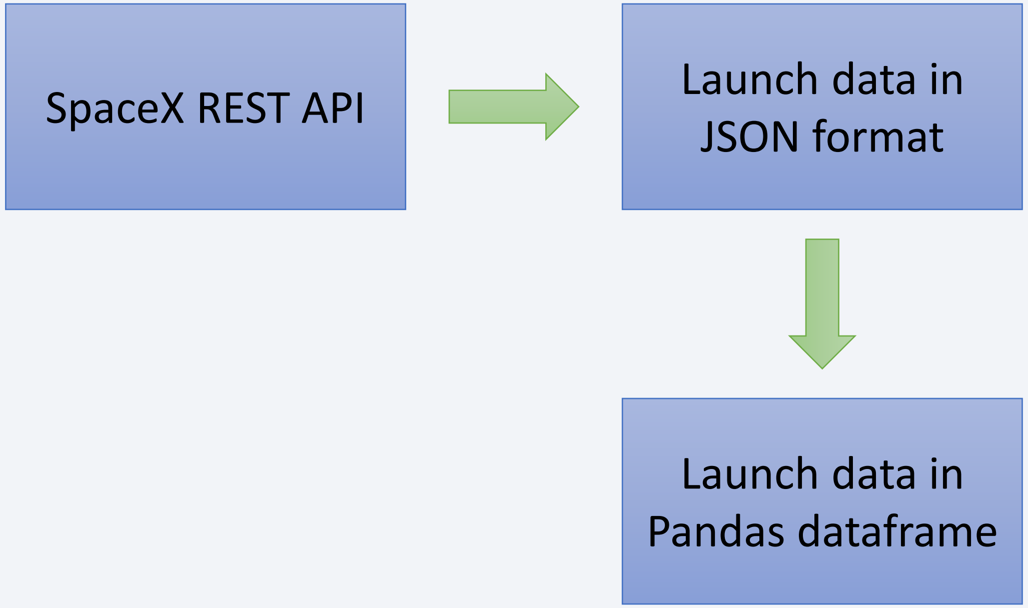
- Flight data was collected from the SpaceX Rest API and Wikipedia pages



Data Collection – SpaceX API

- Launch data obtained from SpaceX REST API
- Converted to Pandas dataframe via JSON object

- GitHub Link: https://github.com/Javert15/Final-Data-Science-Coursera-project/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

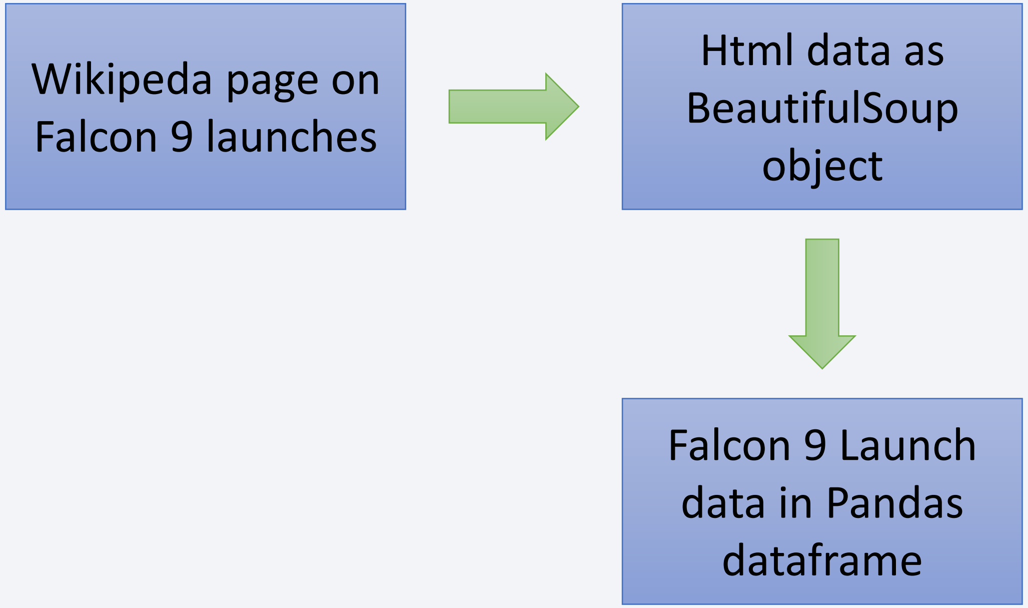


	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	...
0	1	2006-03-24	Falcon 1	20.0	LEO	Kwajalein Atoll	None None	1	
1	2	2007-03-21	Falcon 1	NaN	LEO	Kwajalein Atoll	None None	1	
2	4	2008-09-28	Falcon 1	165.0	LEO	Kwajalein Atoll	None None	1	
3	5	2009-07-13	Falcon 1	200.0	LEO	Kwajalein Atoll	None None	1	

...

Data Collection - Scraping

- Launch data for Falcon 9 flights available on Wikipedia
- Web scraped using BeautifulSoup for further processing
- GitHub Link: https://github.com/Javert15/Final-Data-Science-Coursera-project/blob/main/jupyter-labs-webscraping.ipynb



Data Wrangling

- First, the missing values were identified

Identification of missing values → Control of correct data types → Determined launches per site, targeted orbit and mission outcome → Introduced landing outcome label „Class“

- GitHub Link: https://github.com/Javert15/Final-Data-Science-Coursera-project/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- Summarize what charts were plotted and why you used those charts
  - Categorial plot of Flight Number for each Lanch Site
  - Categorial plot of Payload Mass for each Lanch Site
  - Bar chart of Success Rate for each targeted Orbit type
  - Scatterplot of targeted Orbit for each Flight number, with indication of success
  - Scatterplot of targeted Orbit for each Payload Mass, with indication of success
  - Lineplot of Success Rate per year
- GitHub Link: https://github.com/Javert15/Final-Data-Science-Coursera-project/blob/main/edadataviz.ipynb

EDA with SQL

- Summary of SQL queries performed:
  - Selected distinct Launch Sites
  - Investigated the initial 5 launches from CCAFS LC-40
  - Calculated total masscarried by boosters launched by NASA and F9
  - Identified the initial successful landing
  - Identified the Boosters that achieved successful landing in a drone ship with a payload mass between 4000 and 6000 kg
  - Identified the Booster versions, that carried the maximum payload
  - Listed records for months in 2015
  - Counted the Successes and Failures between 2010-06-04 and 2017-03-20

- GitHub Link: https://github.com/Javert15/Final-Data-Science-Coursera-project/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Markers were added to the locations of launch sites
- Later, additional markers were added, to indicate the success of the missions from the respective launch sites
- In turn, lines were drawn to indicate the distance to the nearest coast, railway etc. to show the launch sites' proximity to these locations
- GitHub Link: https://github.com/Javert15/Final-Data-Science-Coursera-project/blob/main/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

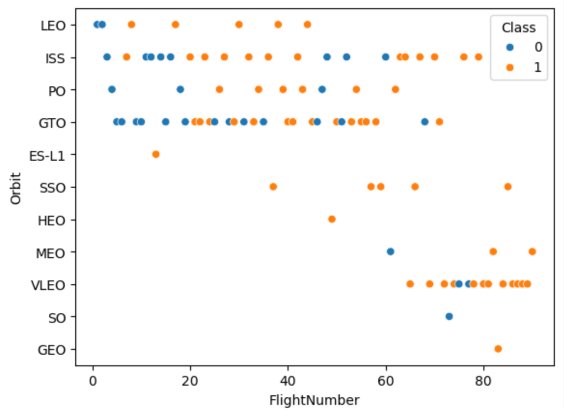
- A Dashboard was used to visualize the success rate from different launch sites and with different payloads
- The success rate is visualized using a pie chart, where the overall success rates and the success rates for individual sites can be compared
- The payload scatter plot allows to select a range for the payload of interest and indicates whether these missions were successful or not
- GitHub Link: https://github.com/Javert15/Final-Data-Science-Coursera-project/blob/main/spacex_dash_app.py

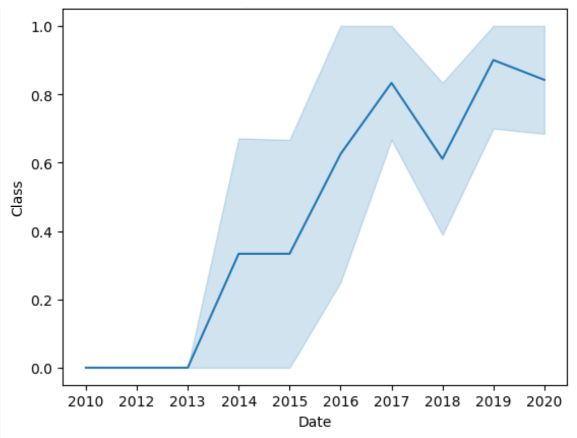
Predictive Analysis (Classification)

- Summarize how you built, evaluated, improved, and found the best performing classification model
- You need present your model development process using key phrases and flowchart
- Add the GitHub URL of your completed predictive analysis lab, as an external reference and peer-review purpose

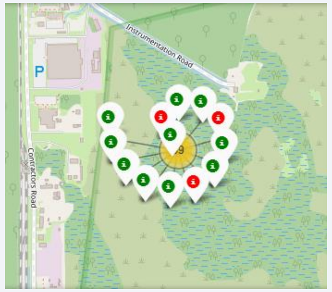
Results

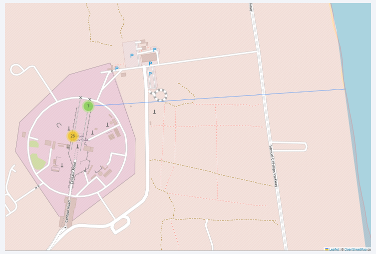
- Exploratory data analysis results



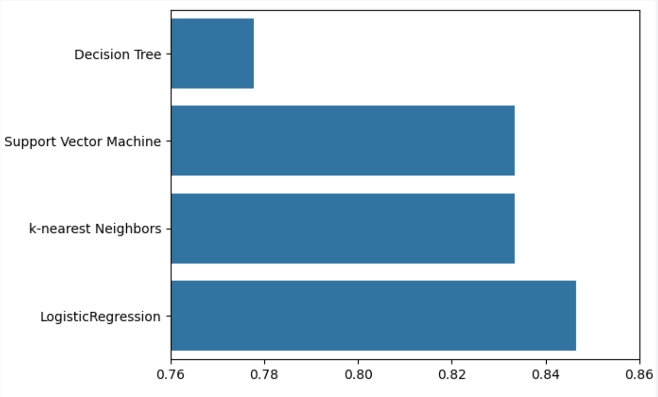


- Interactive analytics demo in screenshots





- Predictive analysis results: Best results with logistic regression

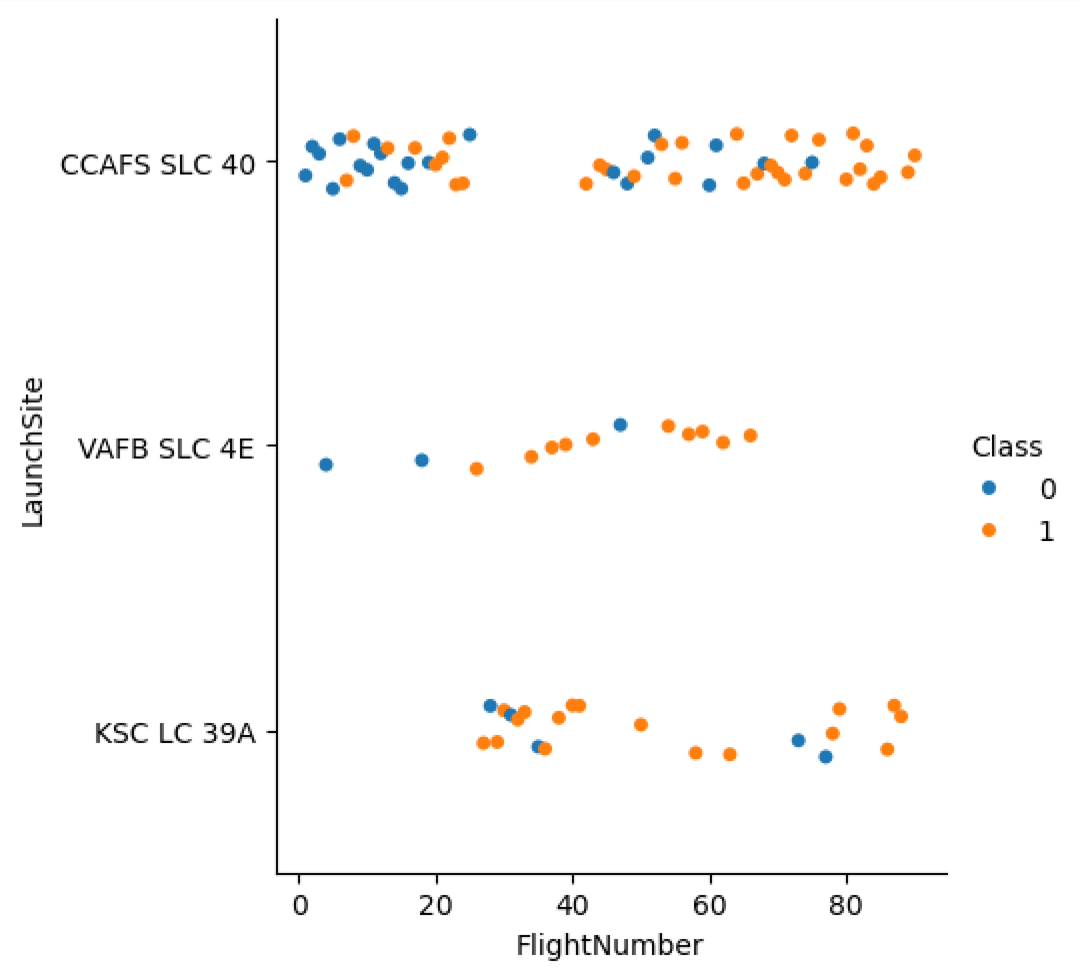


Section 2

Insights drawn from EDA

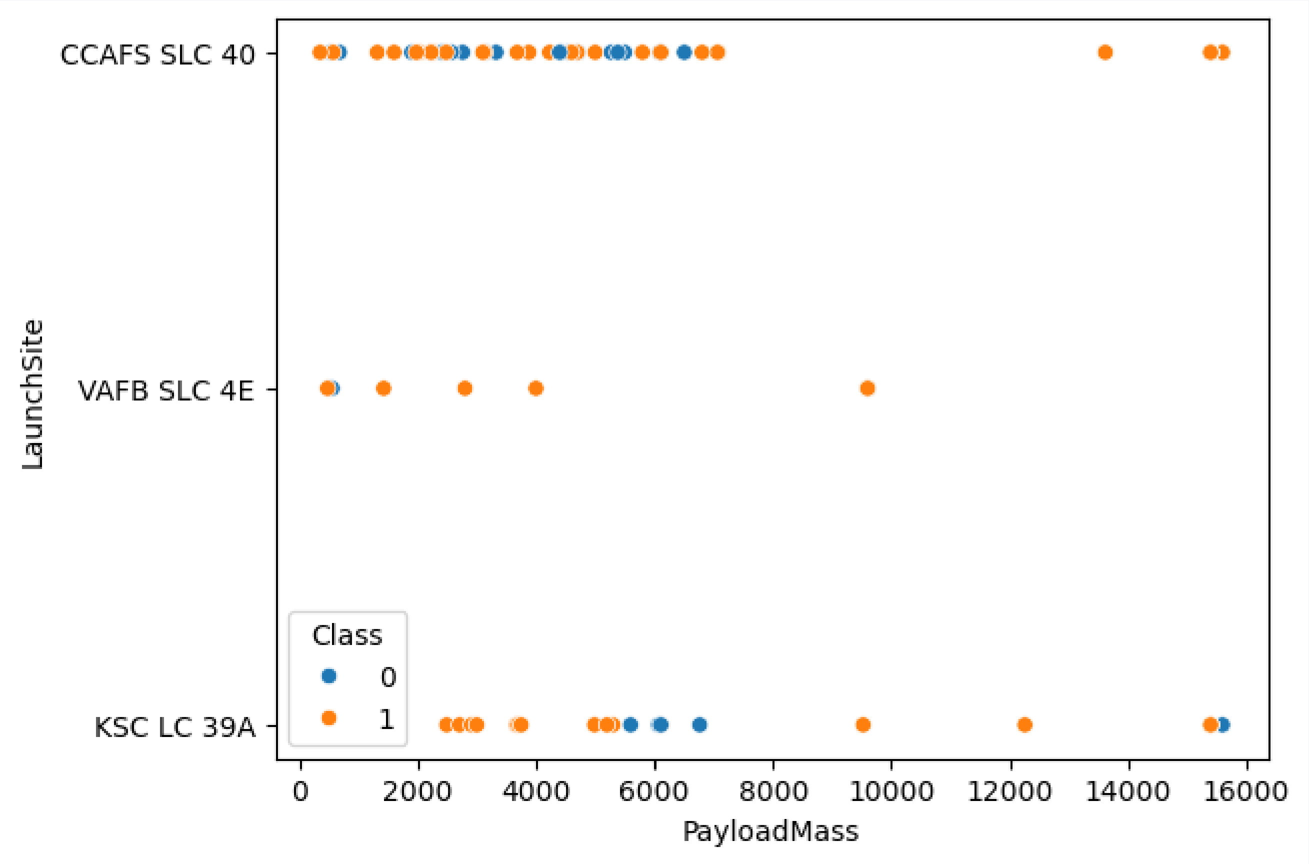
Flight Number vs. Launch Site

- Flights at CCAFS SLC 40 were carried out continuously, with a short break at Flight Number ~ 40, where KSC was primarily used
- VAFB seems to be discontinued, as no more flights after Flight Nr 70 are recorded
- KSC started later, with no flights before Nr 20 are recorded



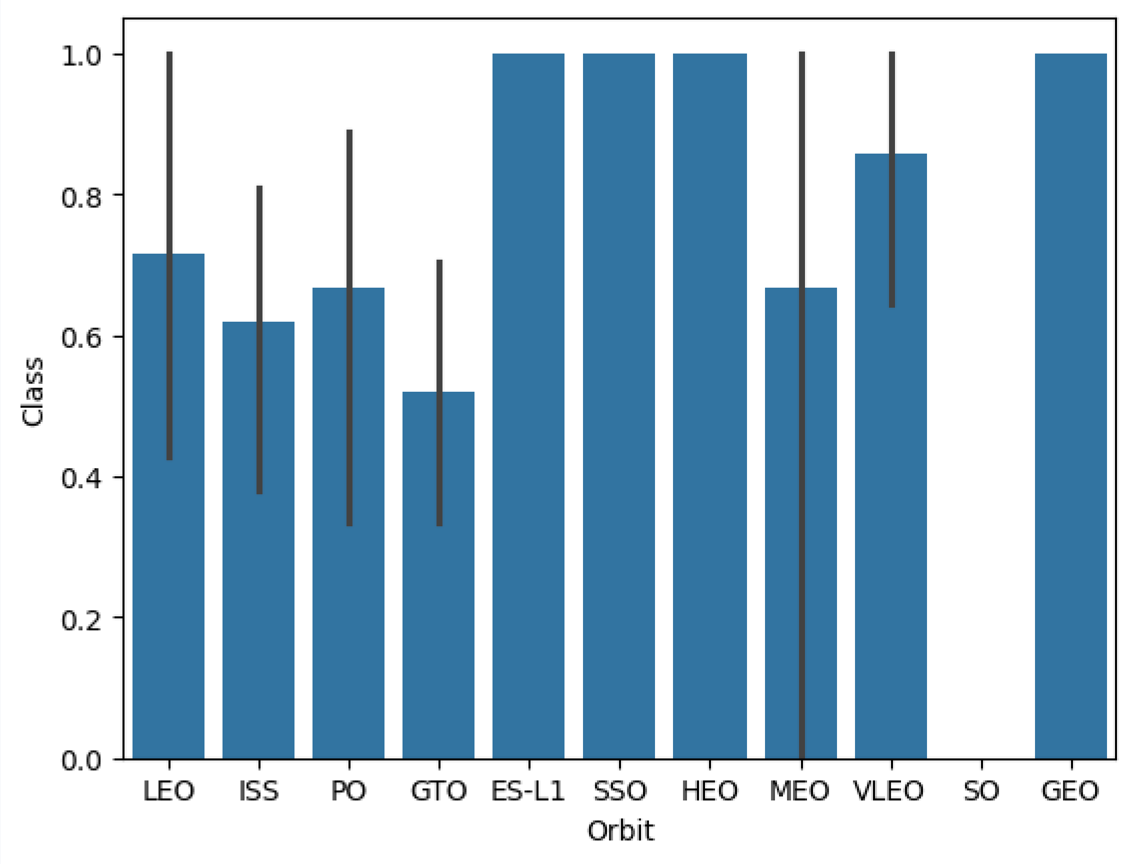
Payload vs. Launch Site

- VAFB was used for less and usually smaller payloads
- Most flights from CCAFS were below 8000 kg, with only 3 exceptions
- KSC has a continuous range of payload mass



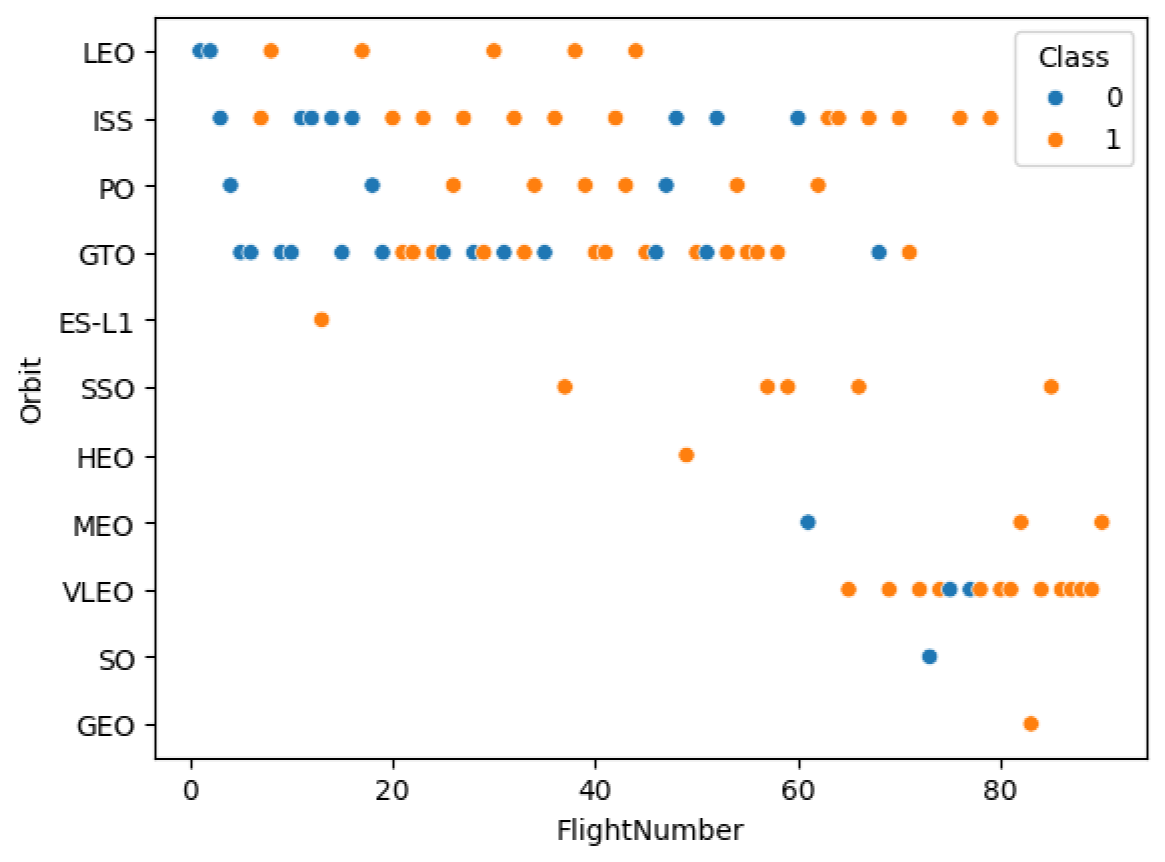
Success Rate vs. Orbit Type

- The bar chart shows the success rates for different Orbit types



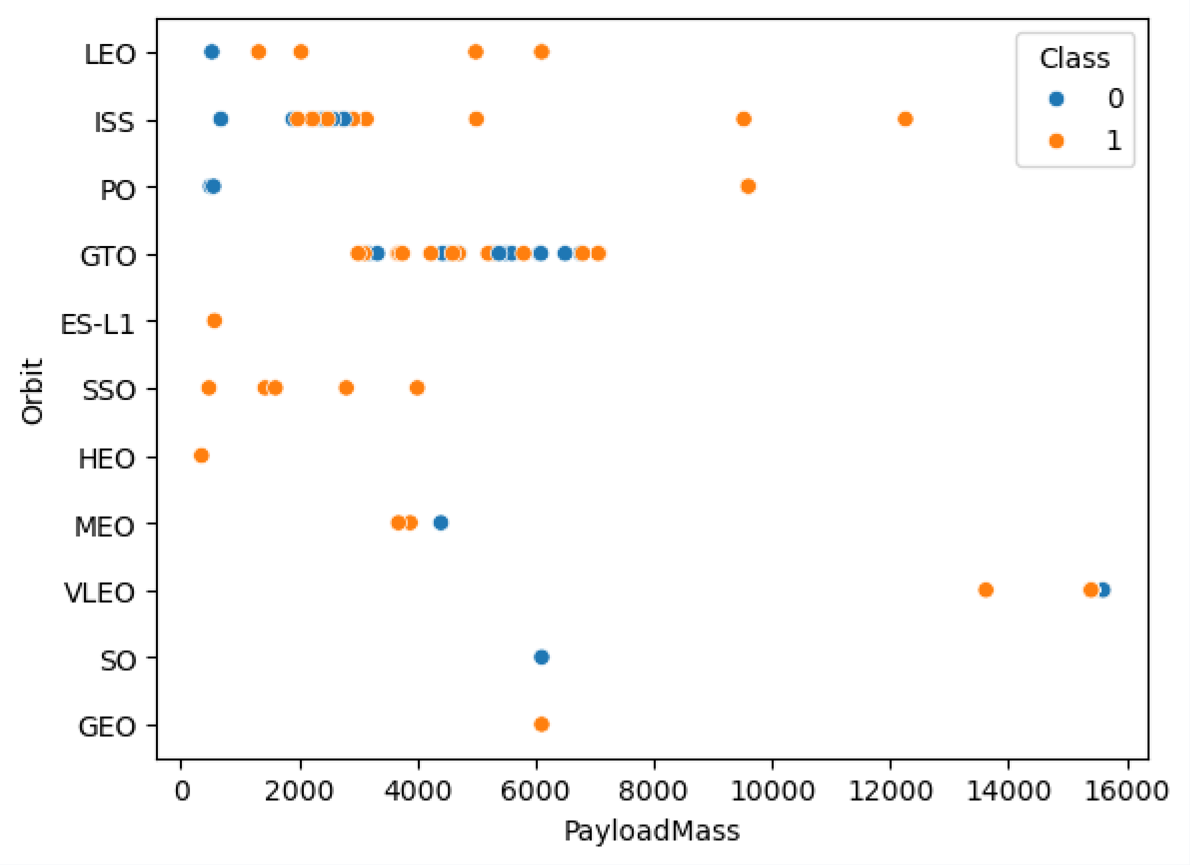
Flight Number vs. Orbit Type

- The scatter plot shows, that initially only LEO, ISS, PO and GTO were targeted
- Later, SSO HEO, and MEO were targeted, before focusing on VLEO
- One failed SO and one successful GEO flight are also recorder



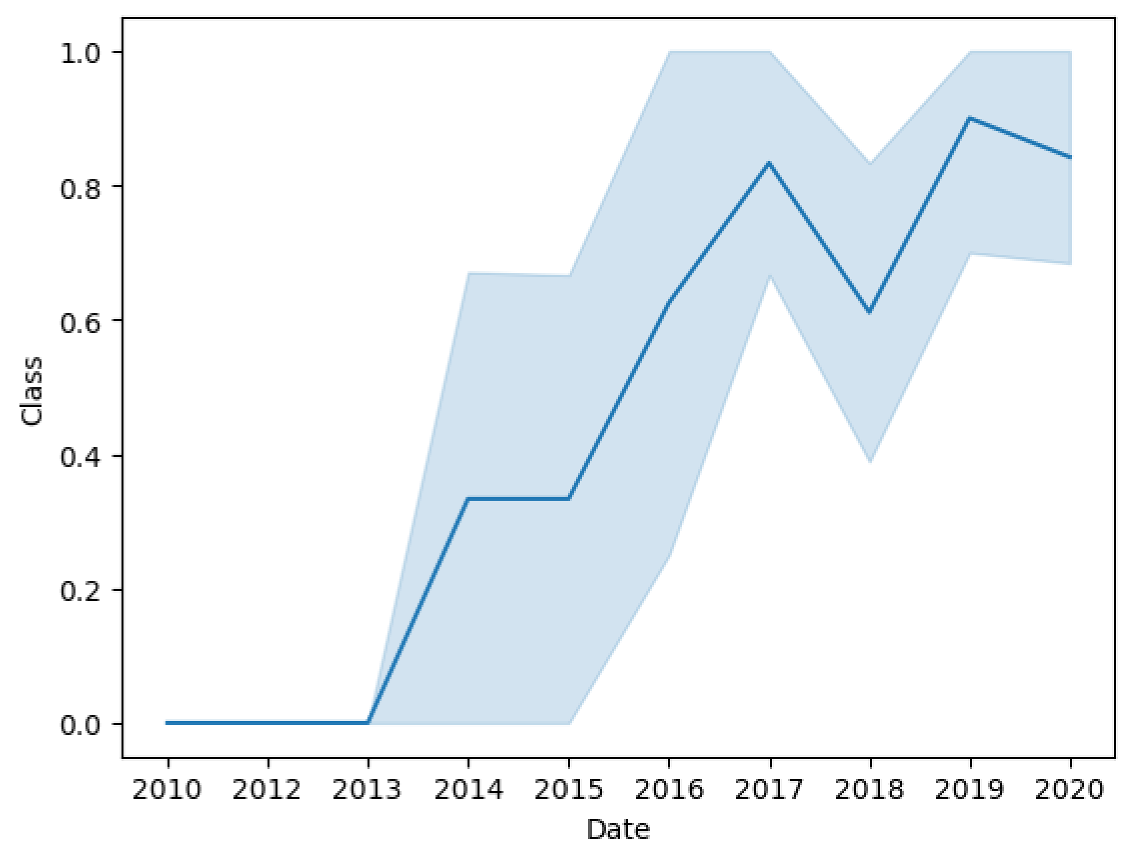
Payload vs. Orbit Type

- The mission success of different Payloads for different Orbit Types are depicted in the scatterplot in the right



Launch Success Yearly Trend

- First successes happened in 2014, that increased until 2017
- 2018 showed some setbacks
- But 2019 and 2020 could return to success rates slightly better than 2017



All Launch Site Names

- Find the names of the unique launch sites
- SELECT DISTINCT(LAUNCH_SITE) FROM SPACEXTBL
  - CCAFS LC-40
  - CCAFS SLC-40
  - KSC LC-39A
  - VAFB SLC-4E

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`
- SELECT * FROM SPACEXTBL WHERE LAUNCH_SITE LIKE "CCA%" LIMIT 5

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

- Calculate the total payload carried by boosters from NASA
- SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE CUSTOMER="NASA (CRS)“
  - 45.596 kg

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1
- SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE Booster_Version LIKE "F9 V1.1%“
    - 2534.66 kg

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad
- SELECT MIN(DATE) FROM SPACEXTBL WHERE LANDING_OUTCOME="Success";
    - 2018-07-22

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
- SELECT Booster_Version FROM SPACEXTBL WHERE (PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000) AND (Landing_Outcome="Success (drone ship)")
    - F9 FT B1022
    - F9 FT B1026
    - F9 FT B1021.2
    - F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes
- SELECT Mission_Outcome, COUNT(Mission_Outcome) FROM SPACEXTBL GROUP BY Mission_Outcome;

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

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass
- SELECT Booster_Version FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_=(SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL)
  - 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

- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015
- SELECT substr(Date,6,2) as month, DATE,BOOSTER_VERSION, LAUNCH_SITE, Landing_Outcome FROM SPACEXTBL WHERE Landing_Outcome='Failure (drone ship)' AND substr(Date,0,5)='2015'

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

- 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
- SELECT Landing_Outcome, COUNT(*) FROM SPACEXTBL WHERE DATE BETWEEN "2010-06-04" AND "2017-03-20" GROUP BY Landing_Outcome HAVING Landing_Outcome="Success (ground pad)" OR Landing_Outcome="Failure (drone ship)" ORDER BY Landing_Outcome DESC

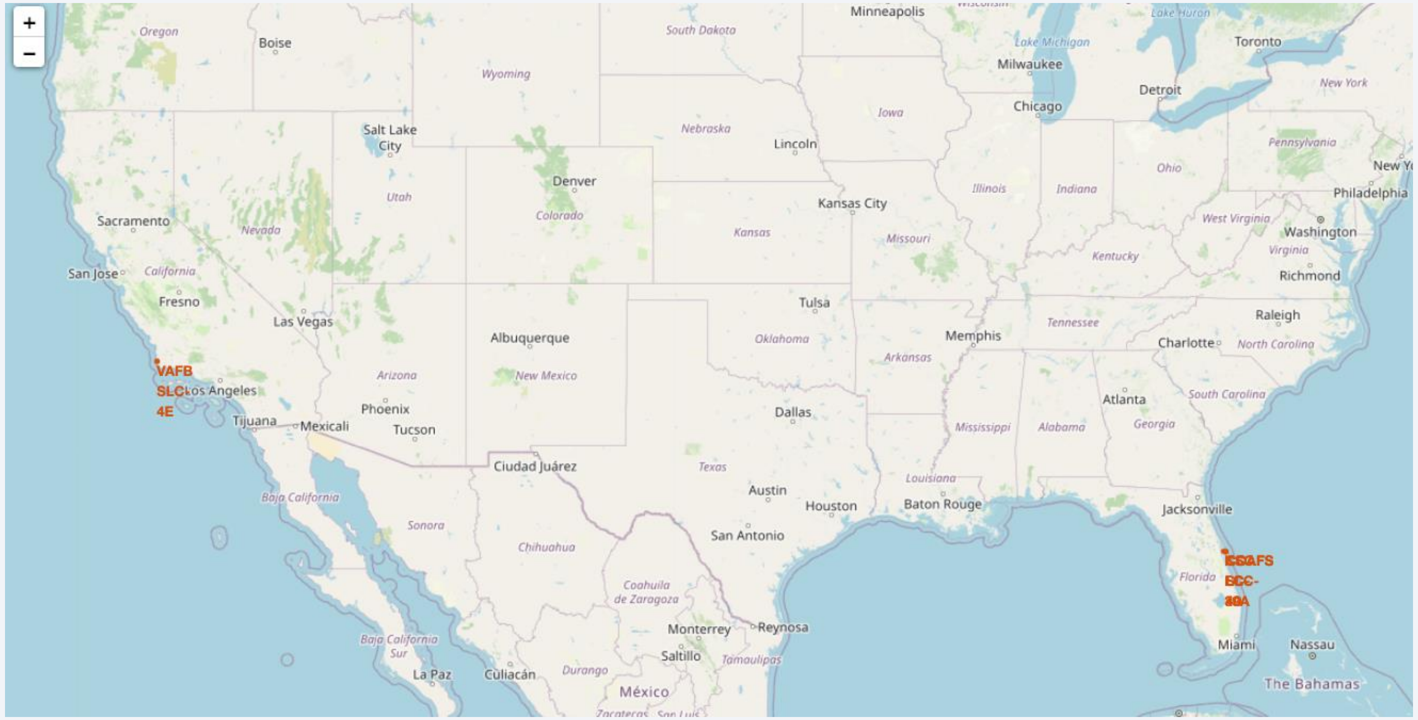
Landing_Outcome	COUNT(*)
Success (ground pad)	3
Failure (drone ship)	5

Section 3

Launch Sites Proximities Analysis

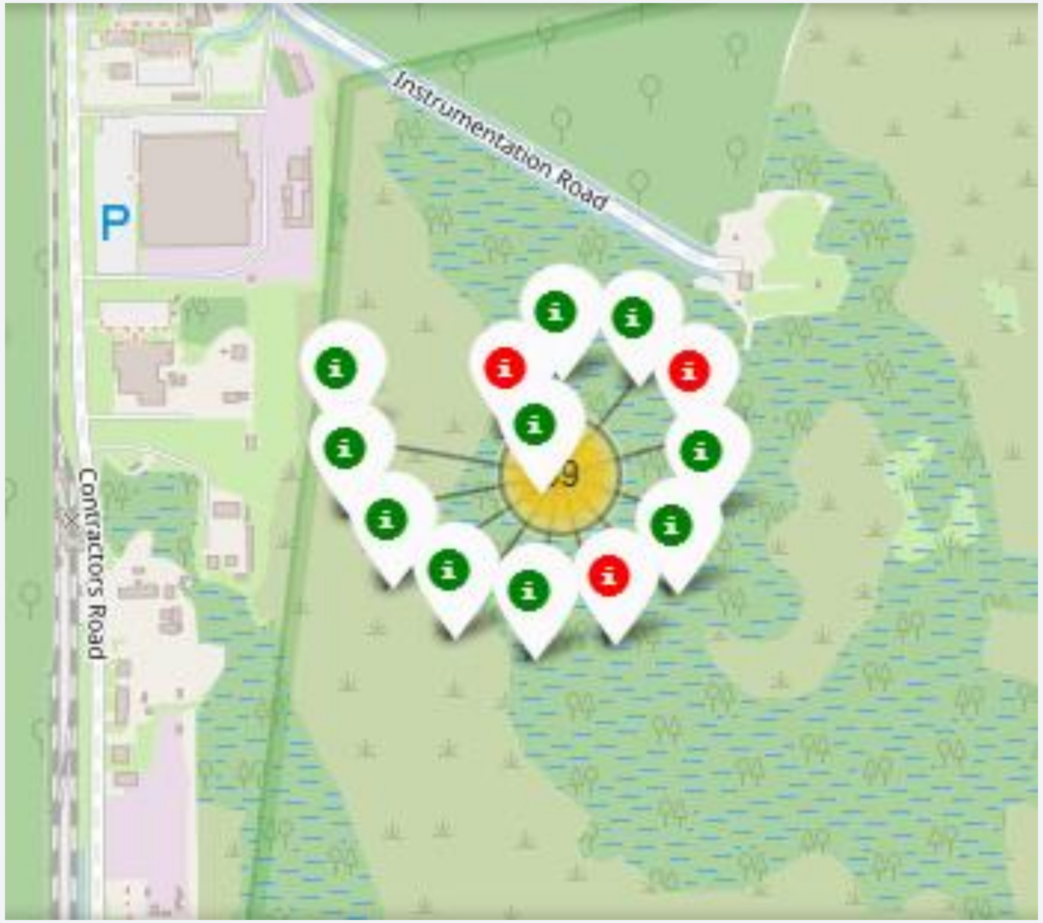
Launch Sites' location

- One launch site is located close to Los Angeles
  - VAFB SLC-4E
- The other three are close to one another in Florida:
  - KSC LC-39A
  - CCAFS LC-40
  - CCAFS SLC-40



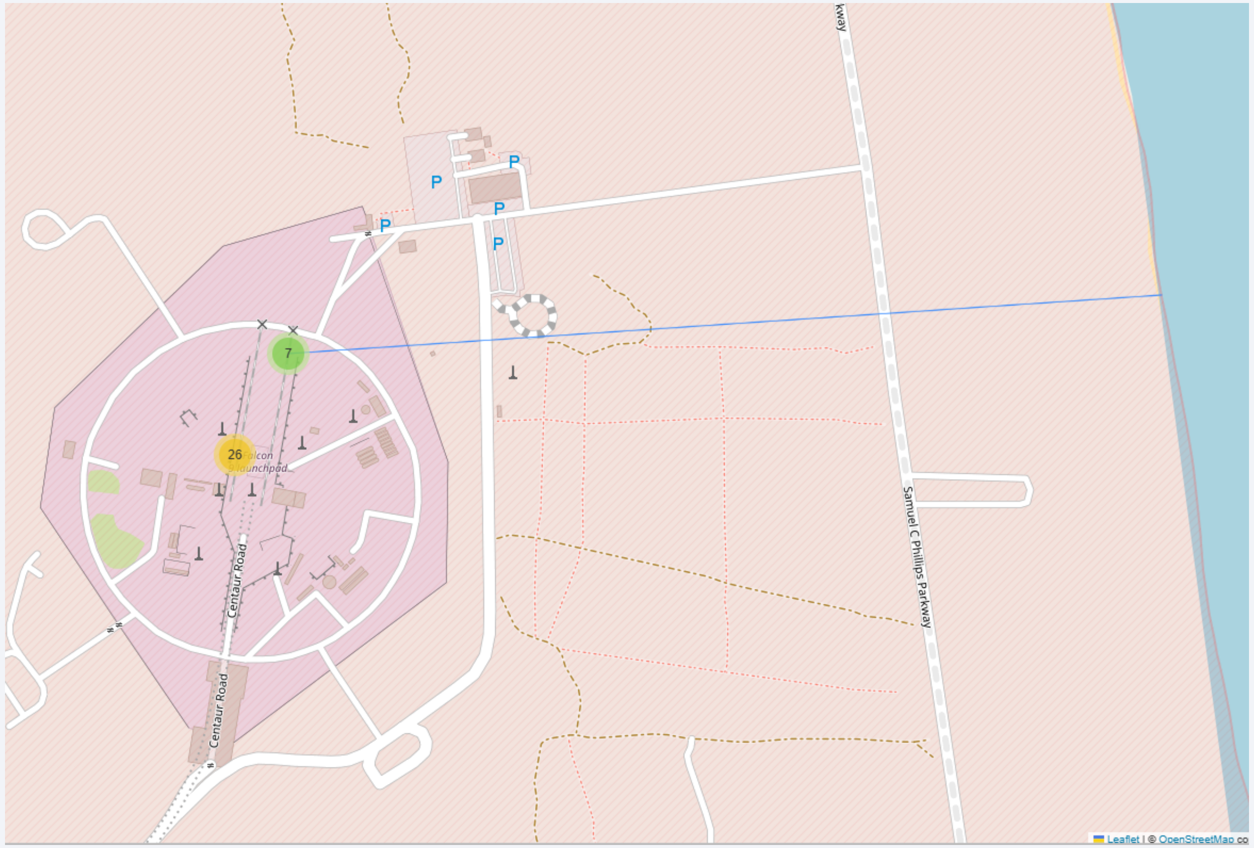
Markers showing success of missions on map

- The success of a mission is marked at each launch site
  - With a green marker for successful missions
  - With a red marker for failed missions



Distance from Launch Site to Ocean

- Distance from Launch Site CCAFS SLC-40 to the ocean indicated by blue line

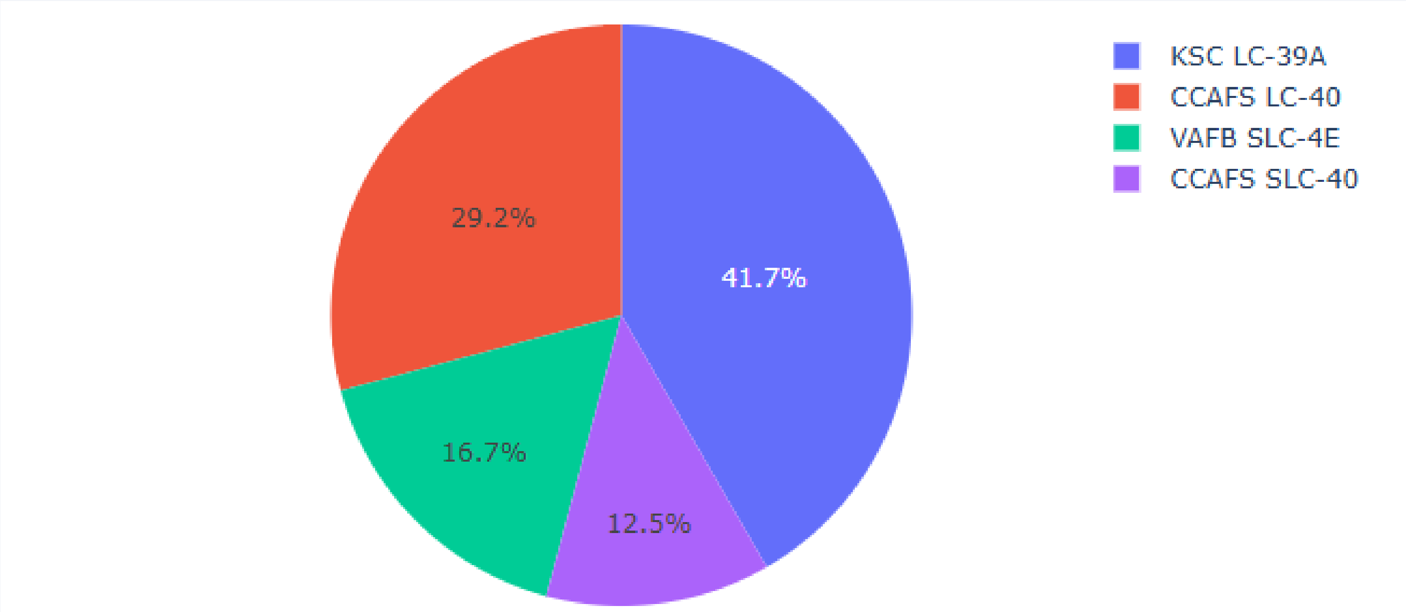


Section 4

Build a Dashboard with Plotly Dash

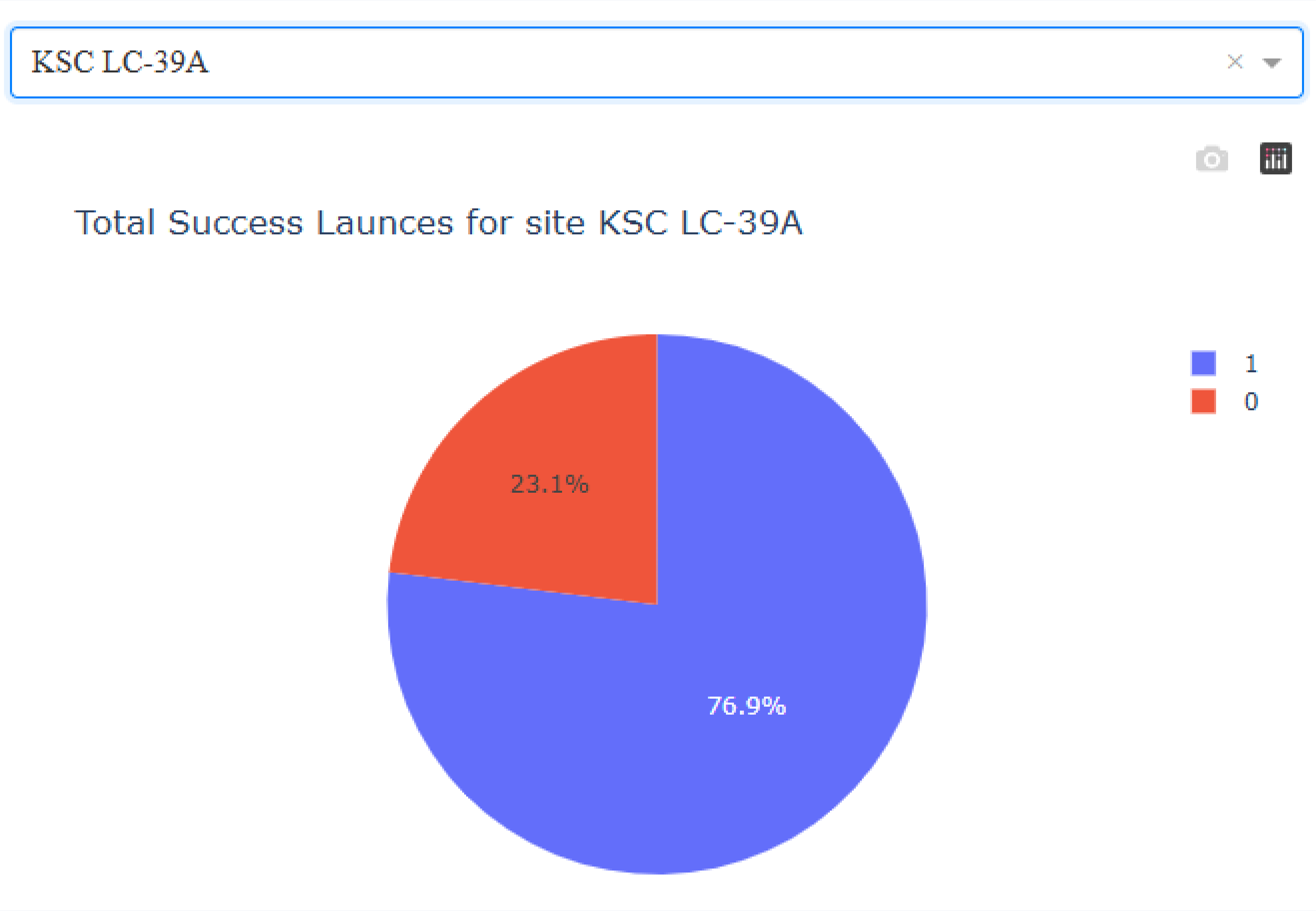
Total Successful Launches By Site

- Most successful launches happen from KSC LC-39A (42%)
- 29% launch from CCAFS LC-40
- VAFB SLC-4E (located on the west coast) attributes only to 17% of the successful launches
- The remaining 13% lauch from CCAFS SLC-40



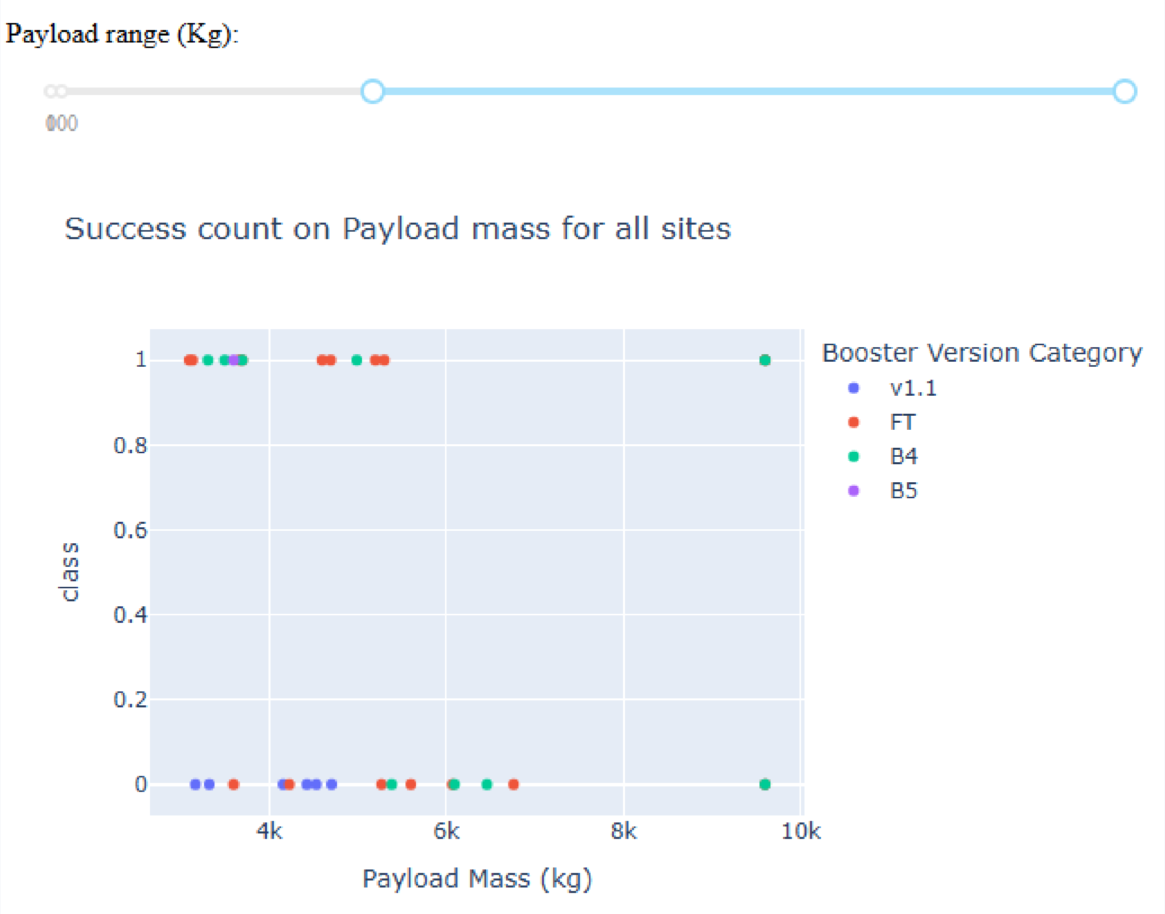
Best success rate: launches from KSC LC-39A

- 77% of all launches from KSC LC-39A were successful, making it the best performing launch site



Successes of Payload-Booster combinations

- Payloads between 3.000 kg and 10.000 kg are displayed
- The color indicates the employed booster version
- Class corresponds to success or failure of that mission

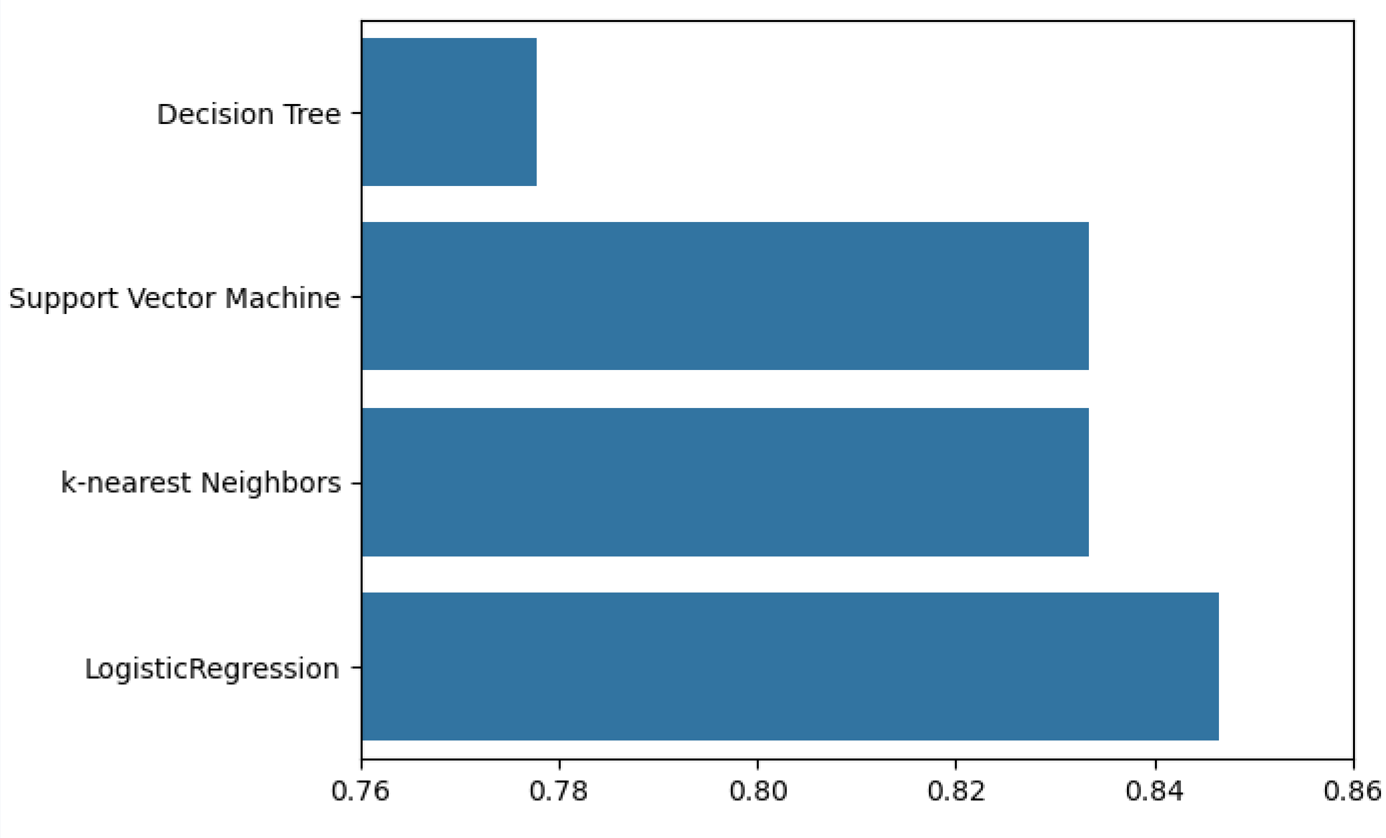


Section 5

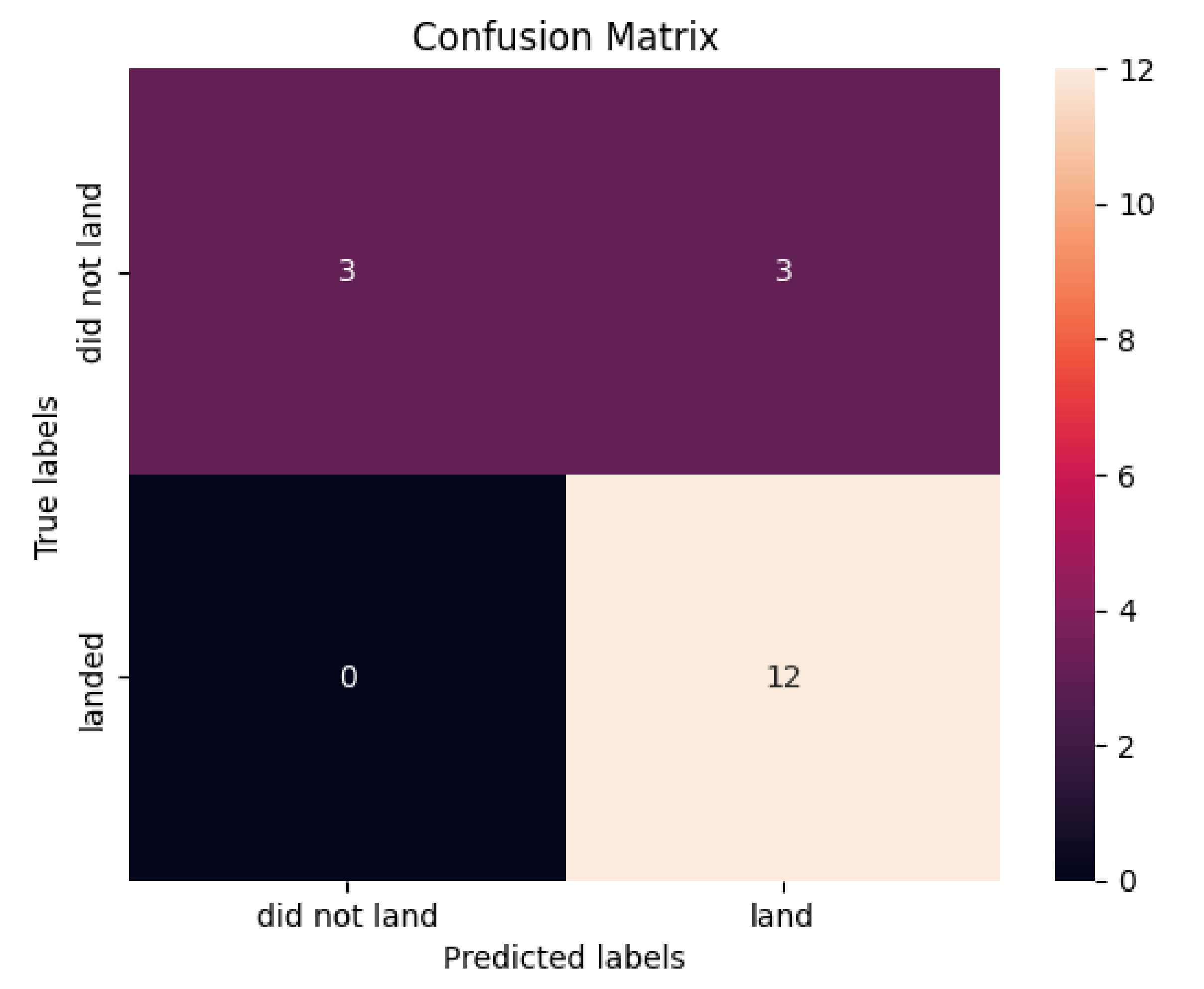
Predictive Analysis (Classification)

Classification Accuracy

- Among the 4 investigated models, the Logistic Regression gave the best result
  - Accuracy = 0.846



Confusion Matrix of Logistic Regression



Conclusions

- More flights indicate a higher success rate at a launch site
- Launch success rate increased from 2013 to 2020 (with a setback in 2018)
- Orbits LEO, ISS, PO and GTO were targeted initially, followed by SSO, HEO, and MEO, before focusing on VLEO. Only one flight each for SO and GEO
- Logistic regression gave the best results in predicting launch success

Thank you!