IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Gabriel Herz
13/04/2023

Outline

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

Executive Summary

Summary of methodologies

- Data about Falcon 9 launches was retrieved from the Space X API and from Wikipedia.
- Various machine learning algorithms were used to identify the ideal conditions for a successful rocket mission
- Summary of all results
- ...

Introduction

Project background and context

- Commercial satellite launches have been pioneered by Space X.
- We want to identify the drivers for successful launches from these early missions to enable ideal conditions for future launches.

Problems you want to find answers

- Where is an ideal rocket starting position located?
- Which orbit is best suited for what kind of payload?

Section 1

Methodology

Methodology

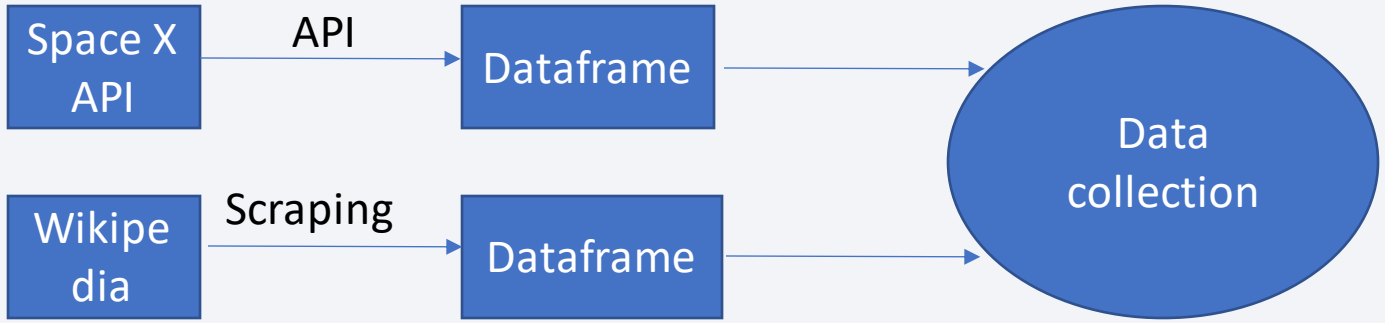
Executive Summary

- Data collection methodology:
  - Data was collected using the Space X API and from the Wikipedia homepage through web scraping.
- Perform data wrangling
  - The API response was filtered for Falcon 9 starts. Missing values in Payload Mass were replaced with averages. Various minor data manipulation steps were performed on the Wikipedia data.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - A multitude of classification models was used concurrently to identify the best fit.

Data Collection

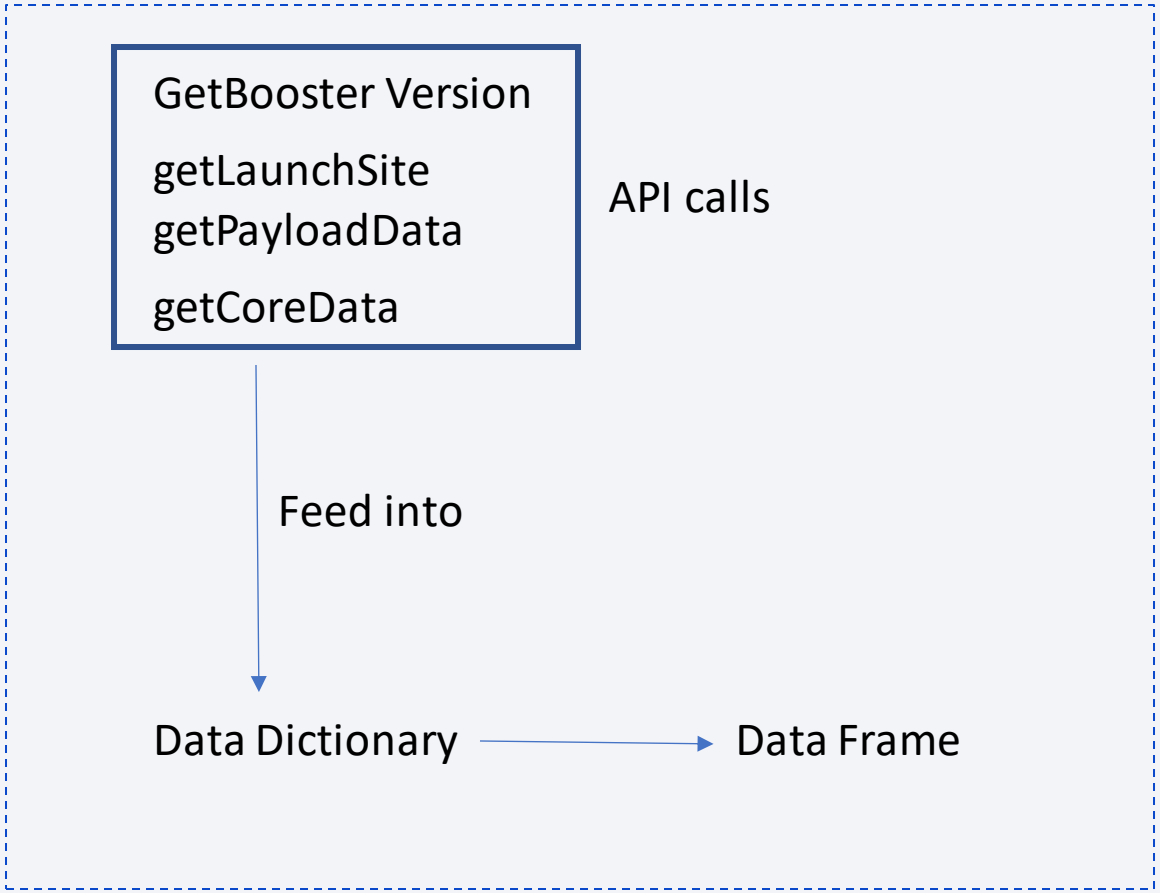
- Datasets were collected from two sources

1. The Space X API
2. The Wikipedia page *https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches*



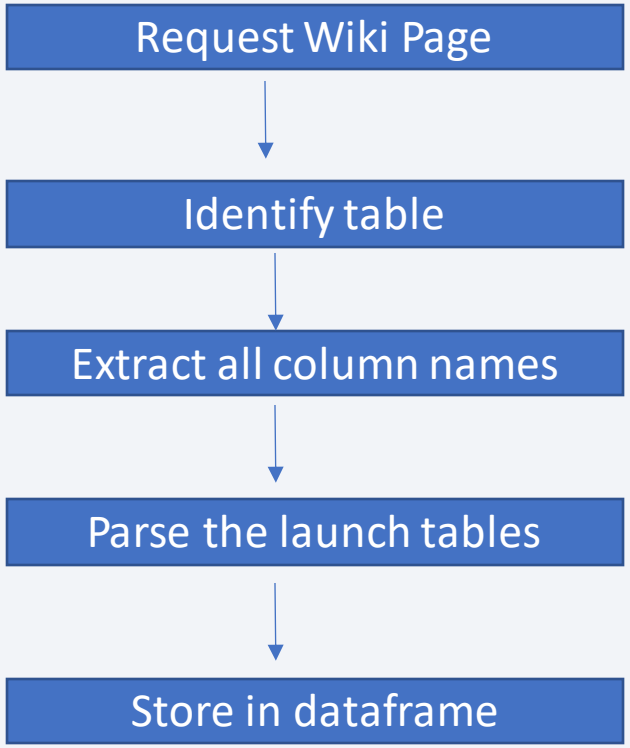
Data Collection – SpaceX API

- The following API call was used:
  spacex_url="https://api.spacexdata.com/v4/..."
  response = requests.get(spacex_url)
- *The specific calls are shown on the right hand side.*
- GitHub URL of workbook:
  https://github.com/GabrielFHerz/IBMDatascienceCourse/blob/2ddb0648fc1d4769b90c442c357a7de1e031f863/Falcon%20Lab1.ipynb



Data Collection - Scraping

- Webscraping with BeautifulSoup
- static_url = "https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922"
- soup = BeautifulSoup(launcherwiki.content, "html.parser")
- GitHUB URL of workbook: https://github.com/GabrielFHerz/IBMDatascienceCourse/blob/2ddb0648fc1d4769b90c442c357a7de1e031f863/Falcon%20M1Lab2.ipynb



Data Wrangling

- The following steps were performed:

1. Load data
2. Calculate share of missing values and identify data types
3. Calculate the number of launches on each site
4. Calculate the number and occurrence of each orbit¶
5. Calculate the number and occurrence of mission outcome per orbit type¶
6. Create a landing outcome label from Outcome column
7. Calculate average landing success rate

- GitHUB: https://github.com/GabrielFHerz/IBMDatascienceCourse/blob/2ddb0648fc1d4769b90c442c357a7de1e031f863/Falcon%20M1Lab3.ipynb

EDA with Data Visualization

The following charts were plotted to identify correlation between variables and identify change of success rate over time.

- Scatter plot FlightNumber vs. PayloadMass
- Scatter plot FlightNumber vs LaunchSite
- Scatter plot Payload and Launch Site
- Bar chart success rate and orbit type
- Scatter plot FlightNumber and Orbit type
- Scatter plot Payload vs. Orbit
- Line graph year vs success rate
- Add the GitHub URL of your completed EDA with data visualization notebook, as an external reference and peer-review purpose: https://github.com/GabrielFHerz/IBMDatascienceCourse/blob/2ddb0648fc1d4769b90c442c357a7de1e031f863/Falcon%20M2Lab1.ipynb

EDA with SQL

- **The following sql queries were performed**
- Names of launch sites
- Launch sites beginning with CCA
- Total payload mass carried by NASA (CRS)
- Average payload mass carried by F9 v1.1
- date when the first successful 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
- records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015
- Rank the count of successful landing_outcomes between the date 04-06-2010 and 20-03-2017 in descending order

- https://github.com/GabrielFHerz/IBMDatascienceCourse/blob/2ddb0648fc1d4769b90c442c357a7de1e031f863/Falcon%20M2Lab2.ipynb

Build an Interactive Map with Folium

- Created maps with folium to show launch sites with markers, numbers of successful starts with clusters and the distance to the sea with lines.
- These objects were added to the map to illustrate the location and important features of the launch sites.
- GitHub: https://github.com/GabrielFHerz/IBMDatascienceCourse/blob/2ddb0648fc1d4769b90c442c357a7de1e031f863/Falcon%20M3Lab1.ipynb

Build a Dashboard with Plotly Dash

- A dashboard containing a drop-down menu, a range slider, a pie-chart and a scatter plot was created.
- This enables the user to interactively explore how launch sites and payloads effect the success rate of launches.
- GitHub: https://github.com/GabrielFHerz/IBMDatascienceCourse/blob/1f2abaed92298a686d4ddc37b13e1f661257595b/spacex_dash_app.py

Predictive Analysis (Classification)

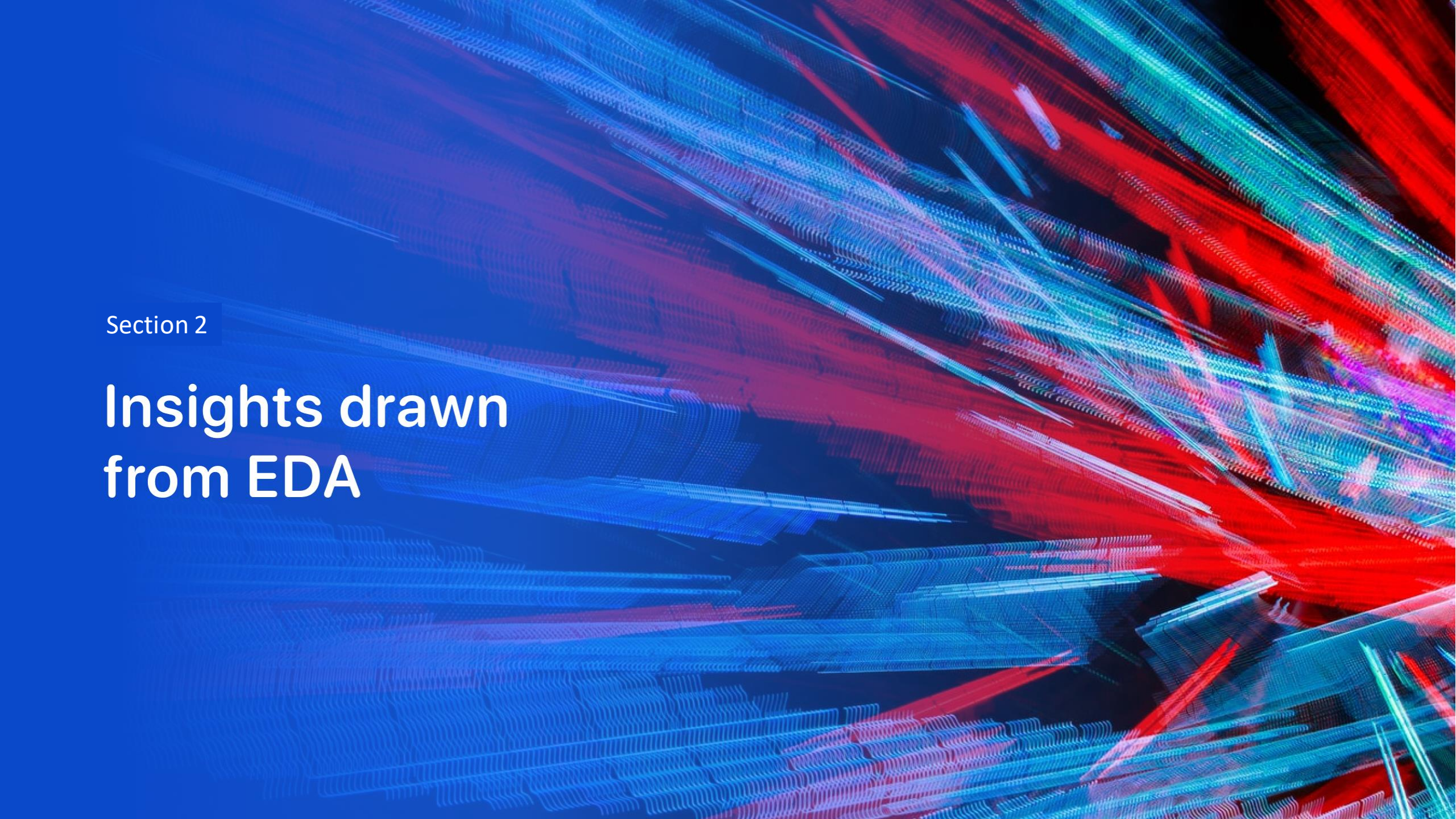
- The data
- Summarize how you built, evaluated, improved, and found the best performing classification model
- You need present your model development process using key phrases and flowchart
- Add the GitHub URL of your completed predictive analysis lab, as an external reference and peer-review purpose

Results

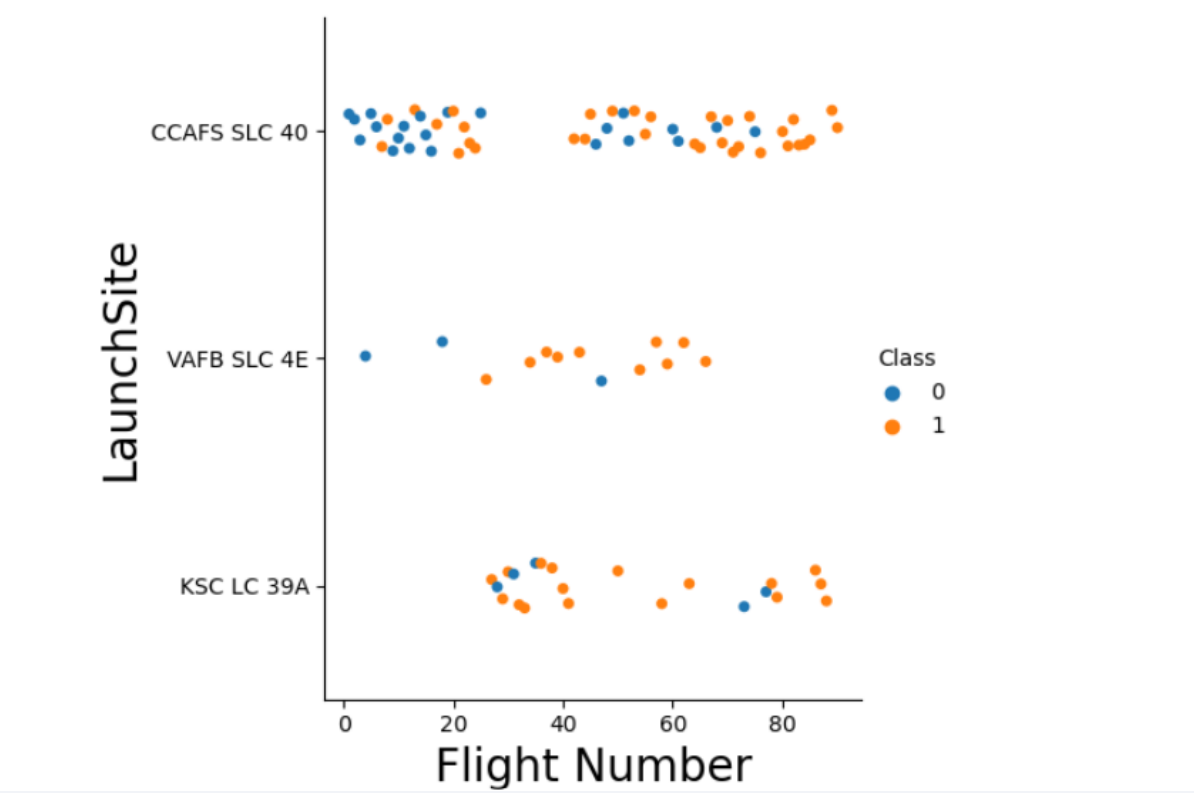
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

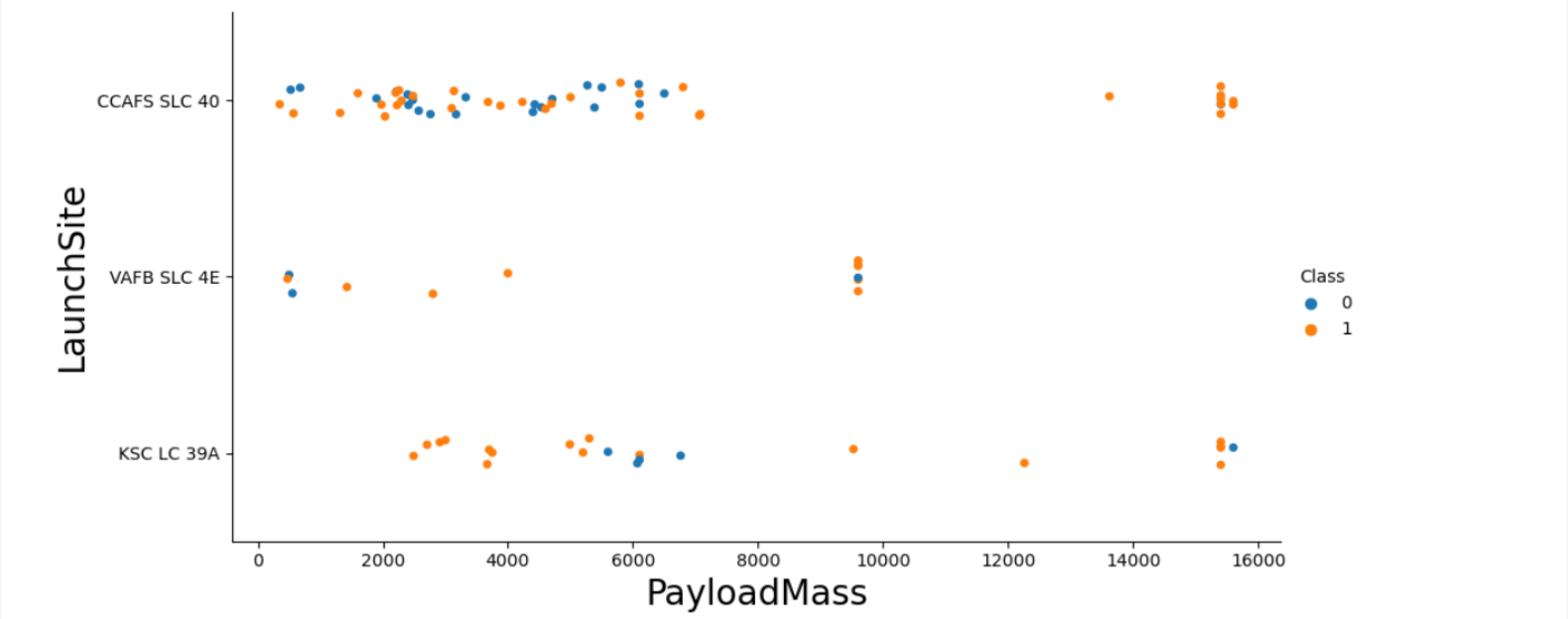


Flight Number vs. Launch Site



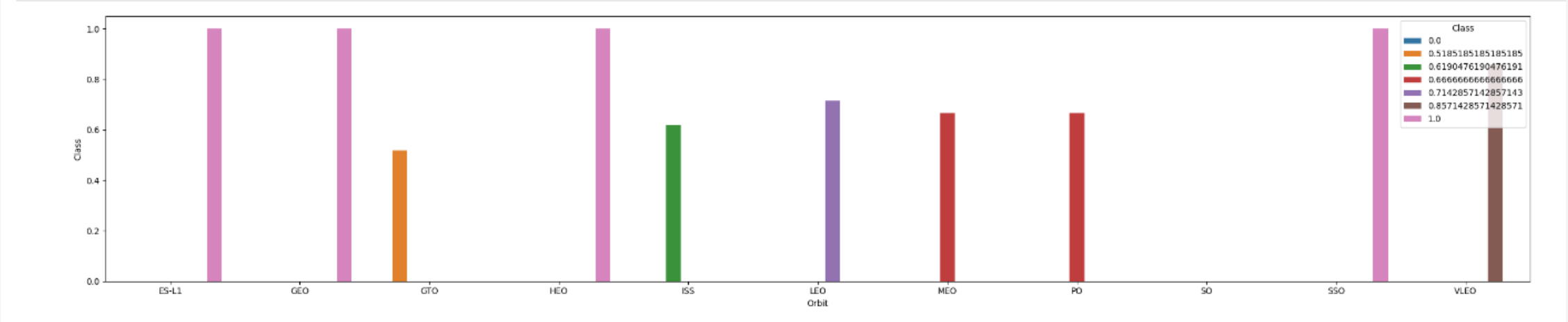
- The plot shows that launches became more sucessful with higher flight numbers. It also shows how launch sites have changed with flight numbers and how sucessful different launch sites are.

Payload vs. Launch Site



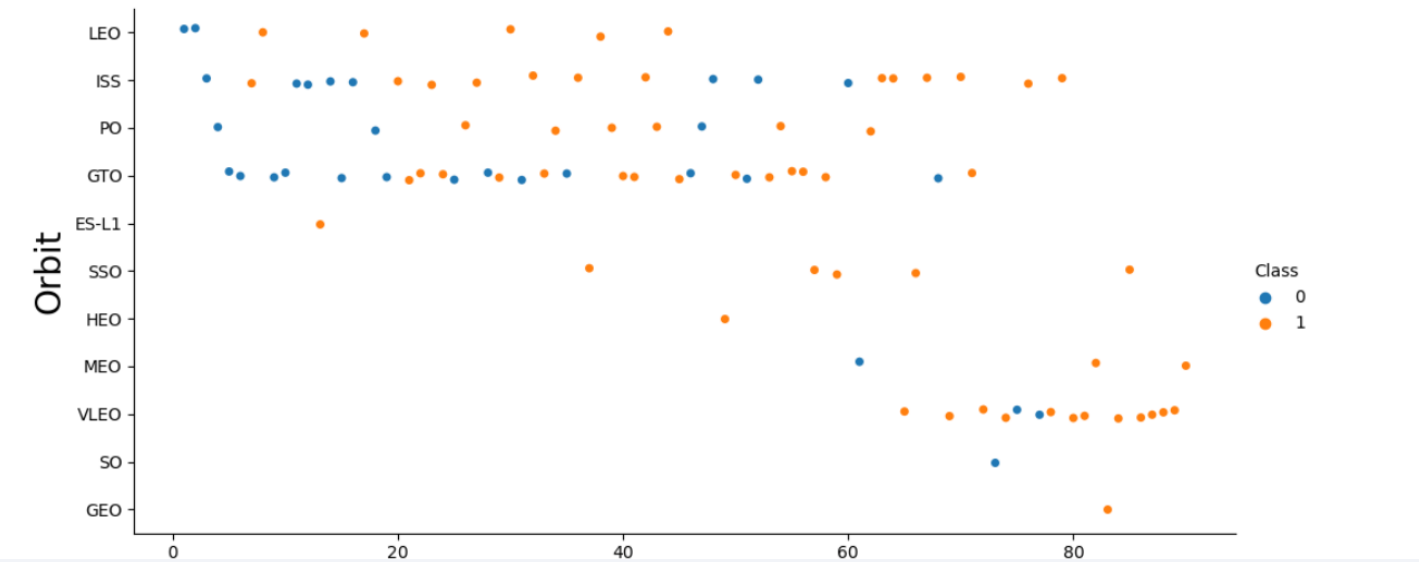
- The plot shows that very high payload are more sucessful fropm CCAFS.
- Lower payloads were most common from CCAFS.
- Low to medium payloads were most sucesfull from KSC.

Success Rate vs. Orbit Type



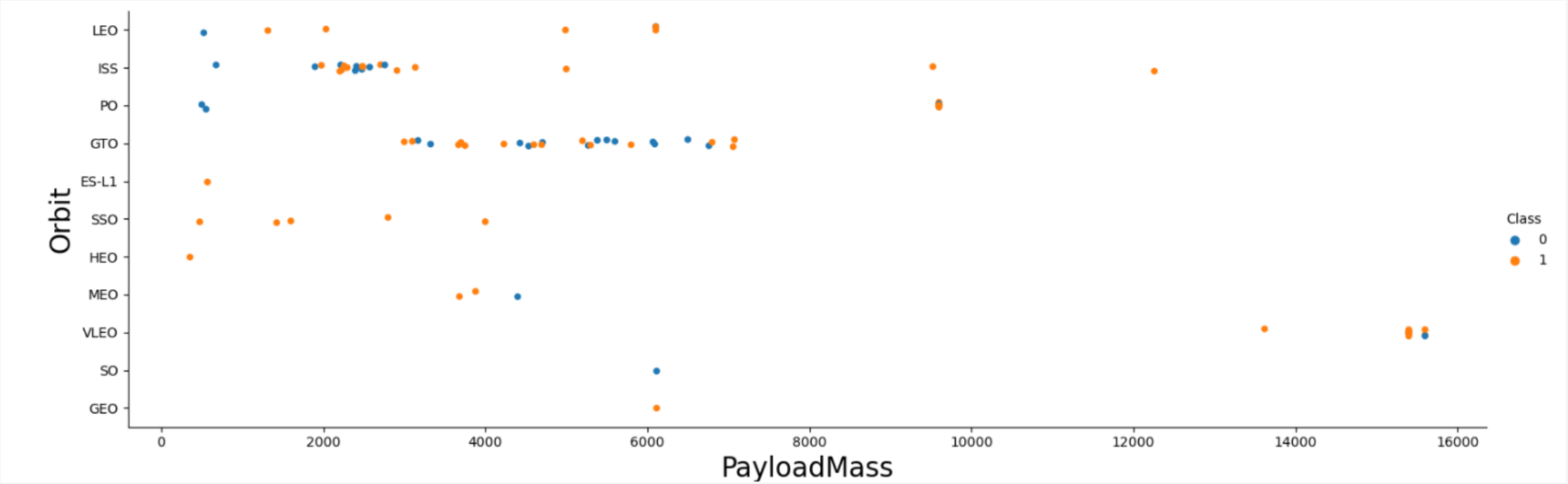
- There is a clear difference in success rate depending on the orbit.
- The orbits shown in purple are the most successful. The orange orbit is the least sucesful.

Flight Number vs. Orbit Type



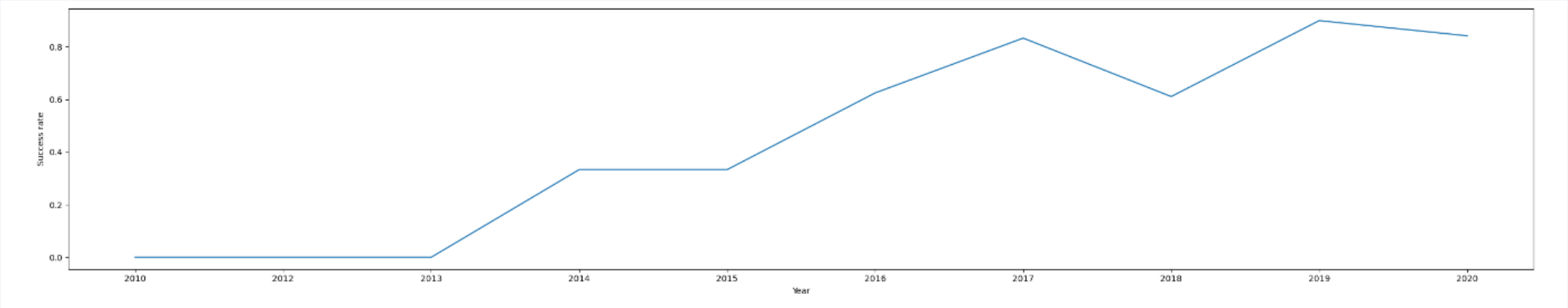
- Orbits changes drastically with higher flight numbers.
- The initially used orbits LEO, PO and GTO were not used in the latest flights but VLEO was used consistently from flight number 60-70 onwards.

Payload vs. Orbit Type



- Very high payloads are only launched from orbit VLEO
- Successful low payloads have only been launched from ES-L1, SSO and HEO.
- Specific orbits are more suited for a specific payload mass.

Launch Success Yearly Trend



- The success rate increased over time with the exception of 2017/18 and 2019/2020.
- The latest success rate is at about 80%.

All Launch Site Names

- The unique launch sites are:
- CCAFS LC-40
- VAFB SLC-4E
- KSC LC-39A
- CCAFS SLC-40

Launch Site Names Begin with 'CCA'

- Find 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
- 04-06-2010 18:45:00 F9 v1.0 B0003 CCAFS LC-40 Dragon Spacecraft Qualification Unit 0 LEO SpaceX Success Failure (parachute)
- 08-12-2010 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)
- 22-05-2012 07:44:00 F9 v1.0 B0005 CCAFS LC-40 Dragon demo flight C2 525 LEO (ISS) NASA (COTS) Success No attempt
- 08-10-2012 00:35:00 F9 v1.0 B0006 CCAFS LC-40 SpaceX CRS-1 500 LEO (ISS) NASA (CRS) Success No attempt
- 01-03-2013 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
- The total payload carried by boosters from NASA is 45596 kg

Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1 is 2534.67 kg.

First Successful Ground Landing Date

- The first successful landing outcome on ground pad was achieved on 01-05-2017.

Successful Drone Ship Landing with Payload between 4000 and 6000

- The following boosters have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000:
- F9 FT B1022
- F9 FT B1026
- F9 FT B1021.2
- F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

- Summary of successful and failure mission outcomes
- Success: 99
- Success (payload status unclear): 1
- Failure (in flight): 1

Boosters Carried Maximum Payload

- The following booster have carried the maximum payload mass:
- 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

- Failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015
- JANUARY Failure (drone ship) F9 v1.1 B1012 CCAFS LC-40
- APRIL Failure (drone ship) F9 v1.1 B1015 CCAFS LC-40

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

- Landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

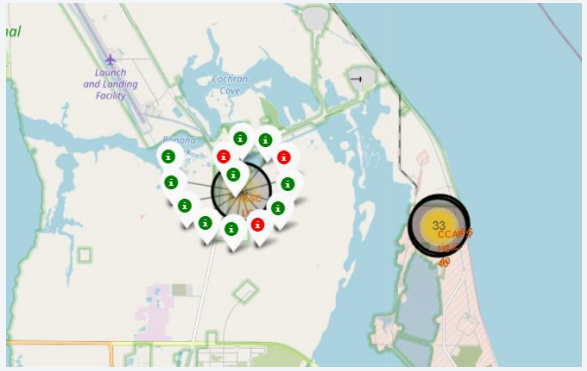
- Landing_Outcome COUNT
- Success 38
- No attempt 21
- Success (drone ship) 14
- Success (ground pad) 9
- Failure (drone ship) 5
- Controlled (ocean) 5
- Failure 3
- Uncontrolled (ocean) 2
- Failure (parachute) 2
- Precluded (drone ship) 1
- No attempt 1

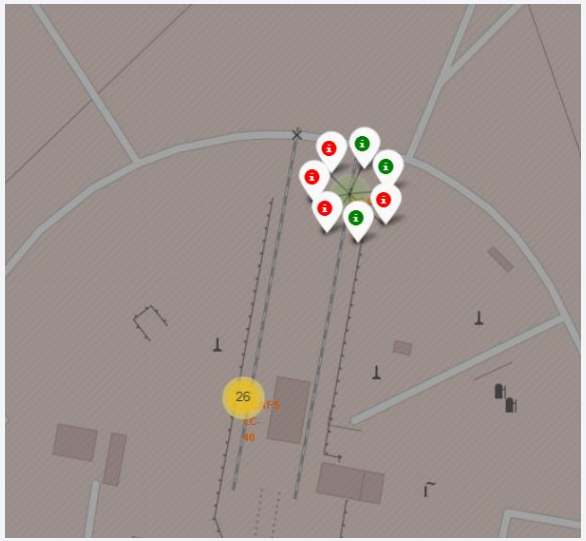

Section 3
Launch Sites
Proximities Analysis

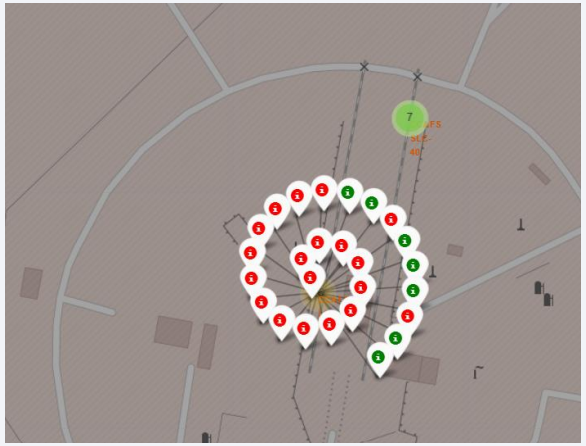
Locations of launch sites

- Launch sites are located at the western and eastern US coastlines.
- Launch sites on each coast line are closely together.

Launch outcomes

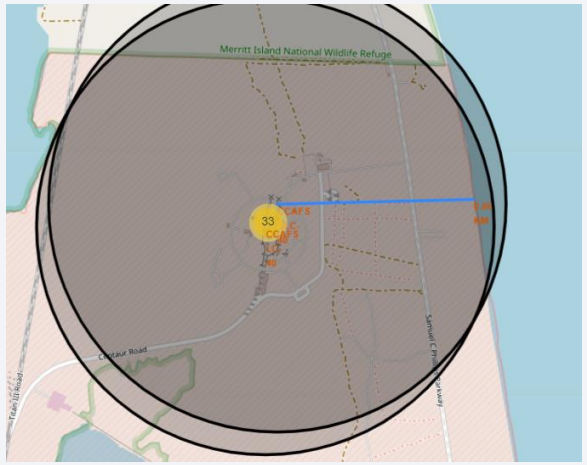




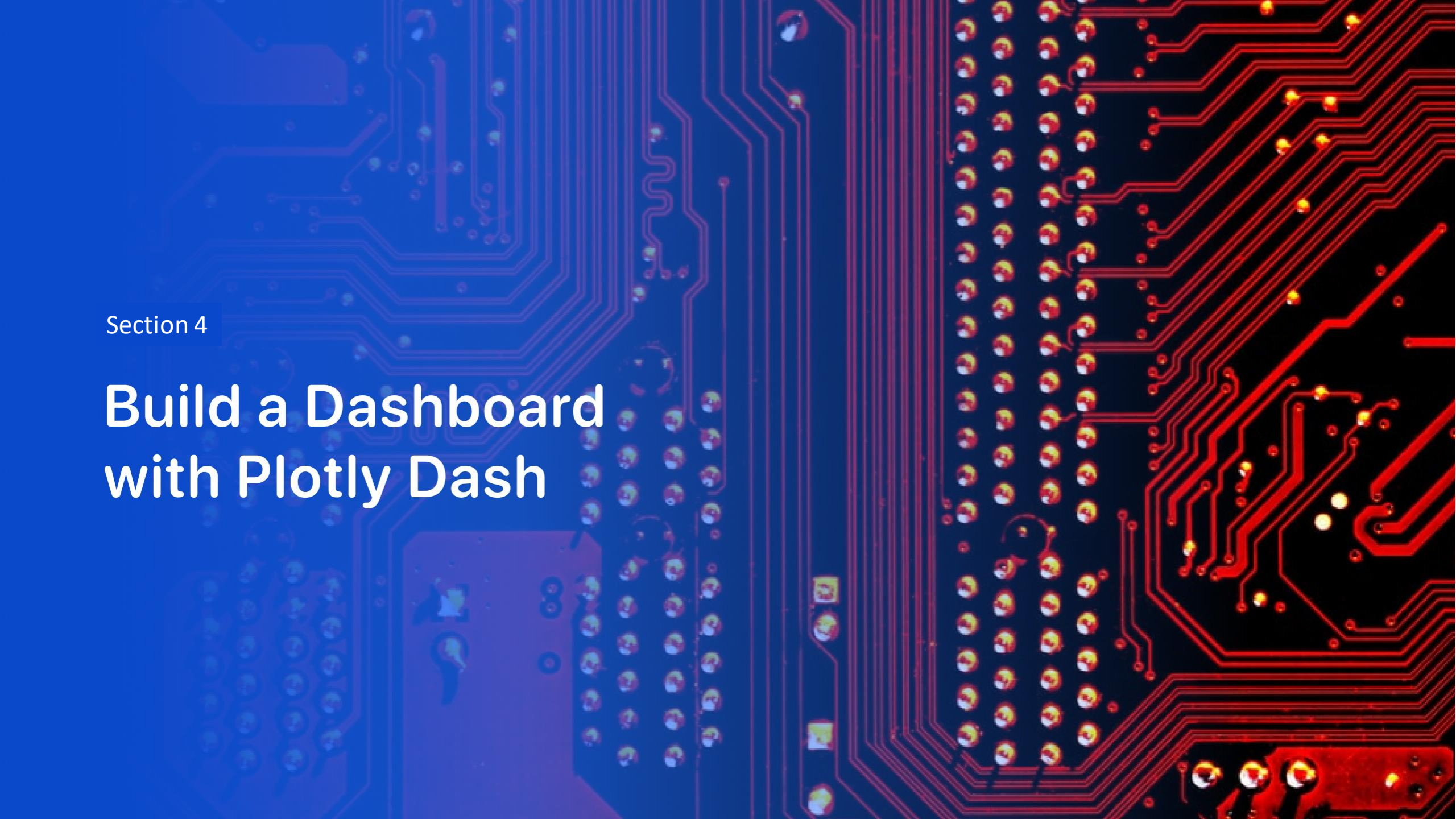


- Above screenshots show the launch outcomes per site.

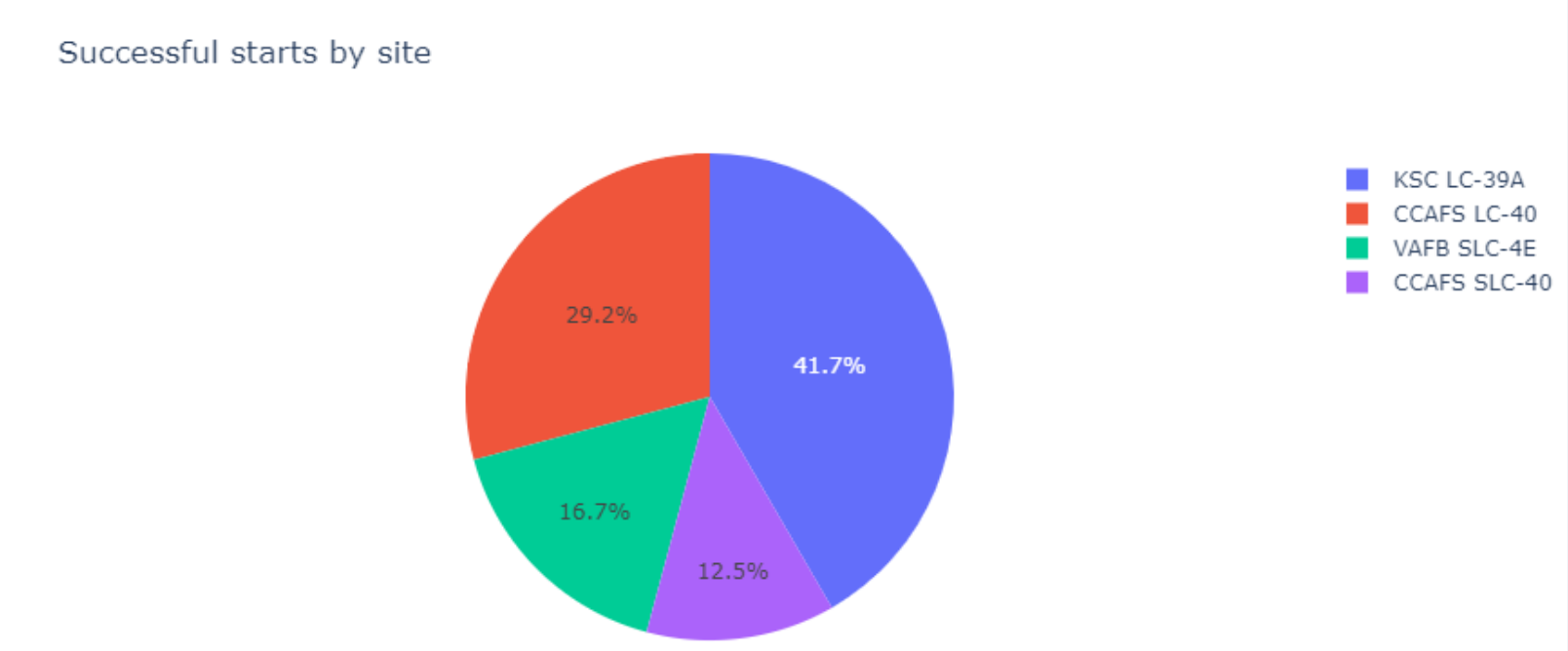
Surroundings of launch sites



- Launch sites are close to the sea.
- The blue line indicates that the launch site is only 860m away from the coastline.
- The launch site is also close to a railway.

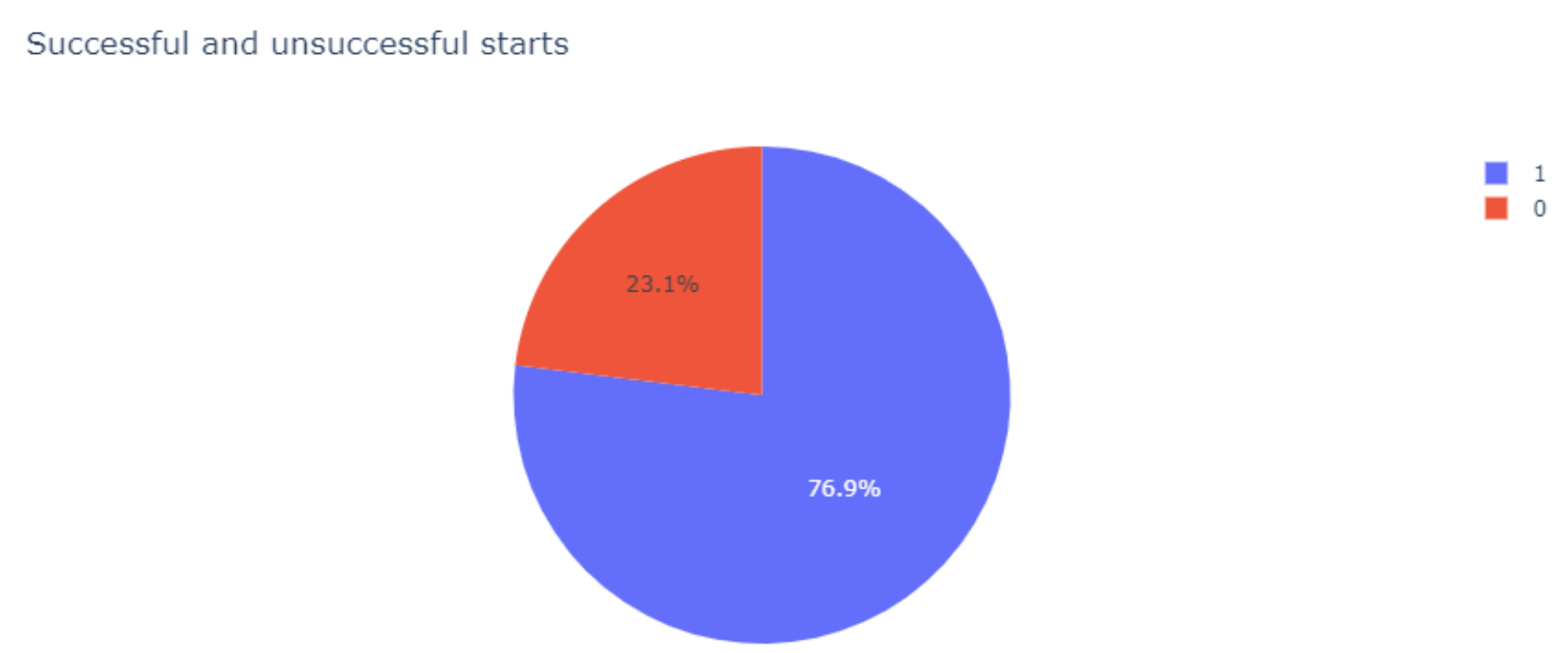

Section 4
Build a Dashboard with Plotly Dash

Launch Success Rates for all sites



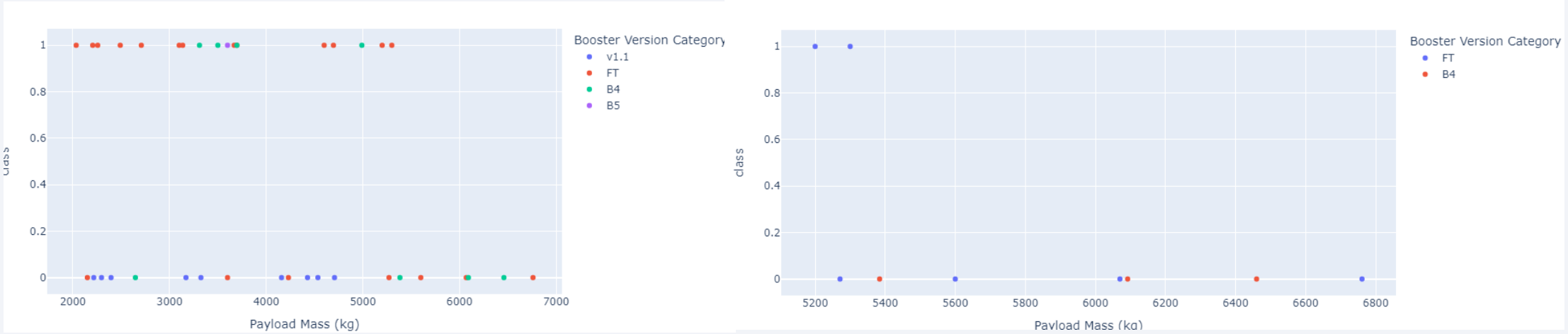
- KSC LC-39A is the most promising launch site.
- CCAFS SLC-40 is the least promising launch site.

Launches on KSC LC-39A



- More than three quarters of launches at KSC LC-39A are successful. This makes it the most successful launch sites relative to launches and in absolute terms.

Payload vs. Launch Outcome

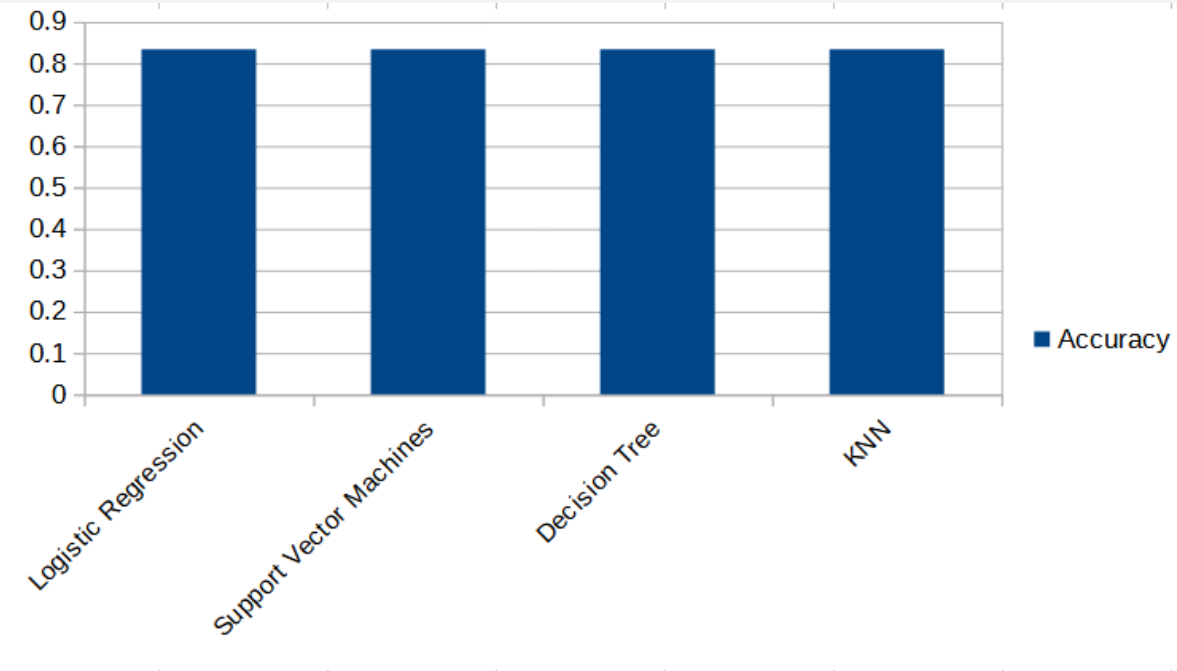


- It is clear from the screenshots that the chance of a successful launch is very low with very high payloads.
- For low payloads, booster version FT seems to be most successful.

Section 5

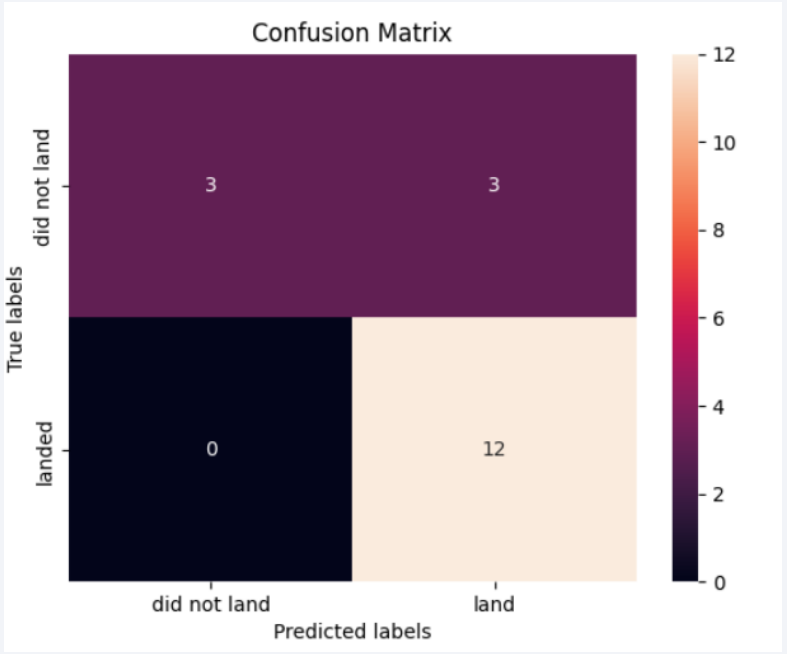
Predictive Analysis (Classification)

Classification Accuracy



- All models have the same accuracy on the test set.

Confusion Matrix



- All models have the same confusion matrix.
- In 3 out of 15 (20%) landing predictions, the actual landing was unsuccessful.
- Unsuccessful landings were predicted with high accuracy.

Conclusions

- Launch sites are close to the seashore and to train lines.
- Launch success increased with flight number. This indicates that valuable experience was gained during early launches.
- Launch success is higher if payload mass and orbit type match.
- Launch success for very heavy payload (more than 6,000kg) is low.
- ML models can identify unsuccessful launches with high accuracy.
- ML models can identify successful launches with good accuracy but have a failure rate of 20%: 20% of predicted successful launches were indeed unsuccessful.

Appendix

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

Thank you!