IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Sascha Höche
06.06.2023

Outline

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

Executive Summary

- In this project, the following methodologies were used
  - For data collection, web scraping as well as APIs were used
  - An exploratory data analysis (EDA) was performed, that included data wrangling and visualization
  - Predictive analysis using machine learning tools was performed
- Summary of all results
  - All public sources delivered workable data
  - The features that allowed the best prediction of successful launches have been determined during EDA
  - The best model to predict the important characteristics of a successful launch have been determined via the machine learning phase.

Introduction

- Project context
  - Determine the viability of a hypothetical company “Space Y” to compete with the existing company “Space X”
- Research Questions
  - What are the best predictors for a successful launch, thereby minimizing the costs?
  - What launch location is the most desirable?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Space X launch data was obtained from two sources:
    - Space X official API (https://api.spacexdata.com/v4/rockets/)
    - Web Scraping Falcon 9 Wiki Page (https://en.wikipedia.org/wiki/Falcon/_9/_and_Falcon_Heavy_launches/)
- Perform data wrangling
  - Summarizing and analyzing features as well as creating a landing outcome label as a result
- Perform exploratory data analysis (EDA) using visualization and SQL

Methodology

Executive Summary

- Perform interactive visual analytics using Folium and Plotly Dash:
- Perform predictive analysis using classification models
  - Normalizing data
  - Performing a test/train 20/80 split
  - Evaluating four classification

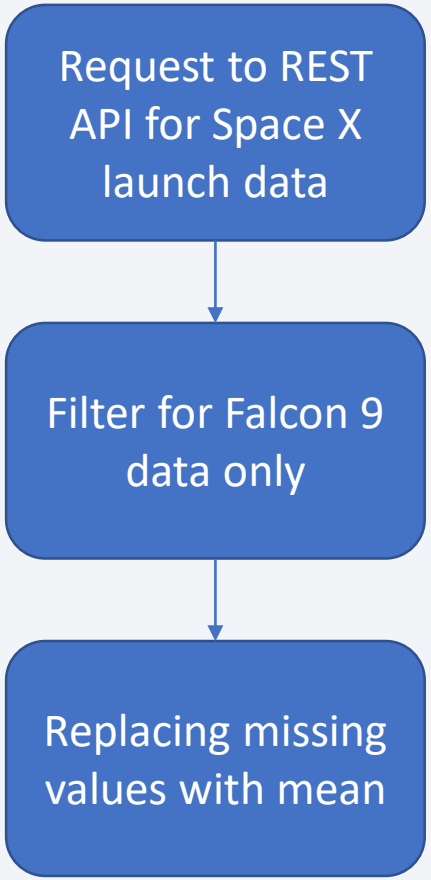
Data Collection

- Data was collected via an API call to https://api.spacexdata.com/v4/rockets/
- Secondary data was collected via WebScraping with BeautifulSoup4 on https://en.wikipedia.org/wiki/Falcon/_9/_and_Falcon_Heavy_launches/

Data Collection – SpaceX API

- Using the official Space X Rest API

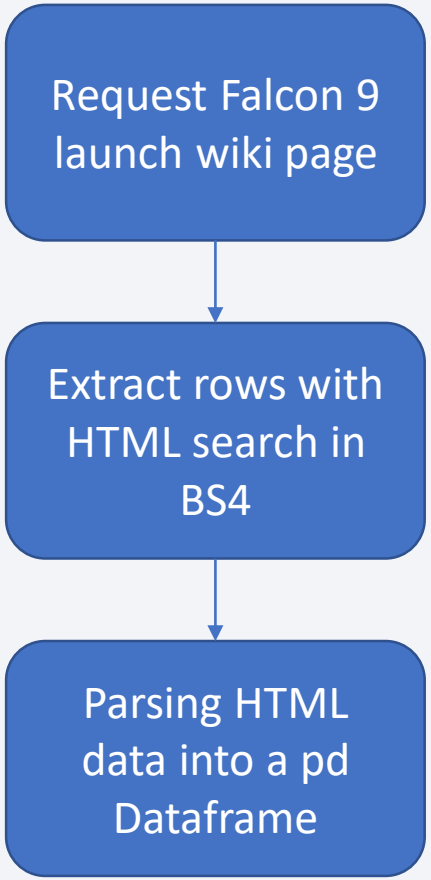
- Github URL: DS_Capstone_Project/notebook_SpaceX_Data_Collection_a7gpUg3Rl.ipynb at main · saschahoeche/DS_Capstone_Project · GitHub



Data Collection - Scraping

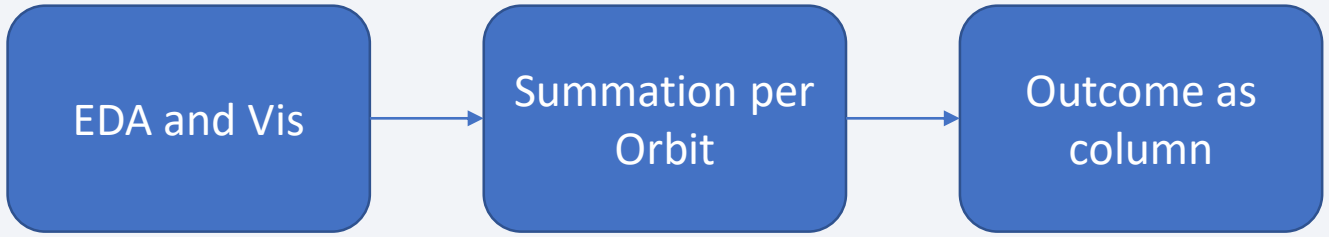
- WebScraping the falcon 9 wiki page with Beautiful Soup module

- Github Url: DS_Capstone_Project/notebook_SpaceX_Webscraping_9tL4IJnwd.ipynb at main · saschahoeche/DS_Capstone_Project · GitHub



Data Wrangling

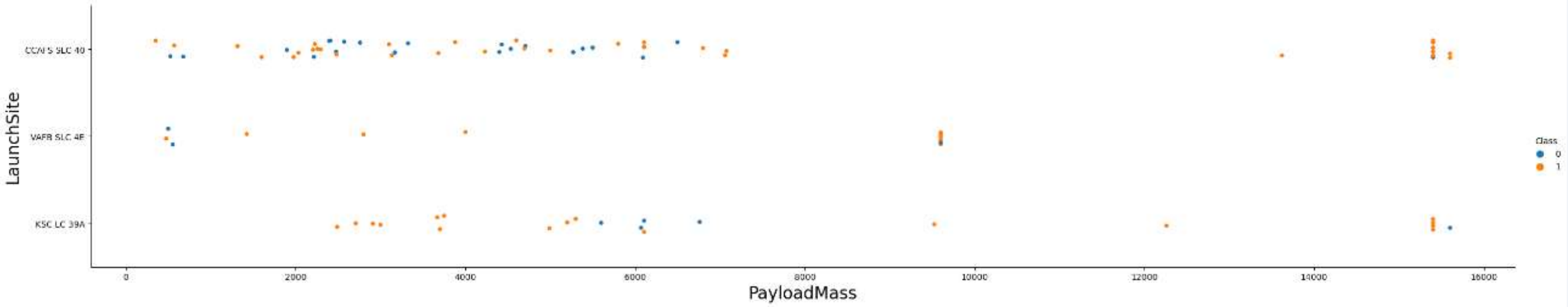
- EDA including visualizations were performed
- The positive outcomes were calculated for each type of orbit
- The outcomes were translated into a Outcome column



- Github Url: DS_Capstone_Project/notebook_SpaceX_Data_Wrangling_GcebiAJvN.ipynb at main · saschahoeche/DS_Capstone_Project · GitHub

EDA with Data Visualization

- To visualize relationships between pairs of features, scatterplots and barplots were used



- Github Url: [DS_Capstone_Project/notebook_SpaceX_EDL_with_Vis_dOEPS91SK.ipynb at main · saschahoeche/DS_Capstone_Project · GitHub]

EDA with SQL

- SQL Queries performed:
  - Select unique launch sites
  - Top 5 launch sites beginning with ‘CCA’
  - Total payload mass carried
  - Average payload mass carried by booster F9 v1.1
  - Date of first successful landing pad landing
  - Booster names with successful drone ship landing and 4k to 6k payload
  - Number of successful and not successful missions
  - Count of landing outcomes between 04.06.2010 and 20.03.2017
- Github Url: DS_Capstone_Project/notebook_SpaceX_EDL_with_SQL_0KH9YtJM3.ipynb at main · saschahoeche/DS_Capstone_Project · GitHub

Build an Interactive Map with Folium

- Markers, circles, lines and clusters were created
  - Markers to indicate launch sites
  - Circles to highlight areas around those launch sites
  - Clusters of markers to group events by coordinates
  - Lines as distance measure between coordinates

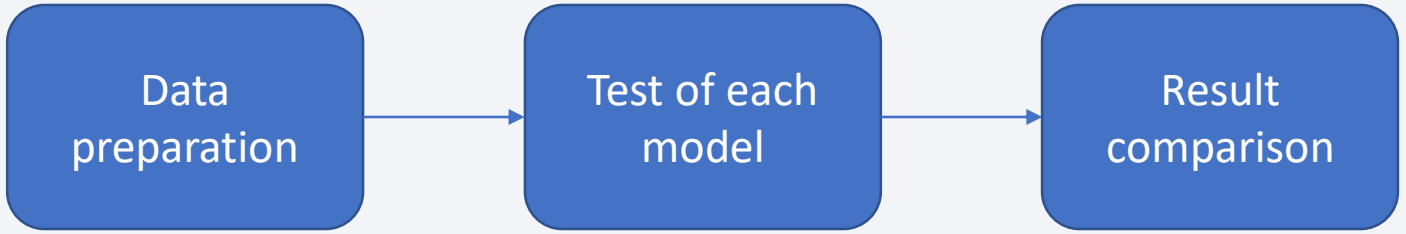
- Github Url: DS_Capstone_Project/notebook_SpaceX_Interactive_Dashboard_L5vVtXuSg.ipynb at main · saschahoeche/DS_Capstone_Project · GitHub

Build a Dashboard with Plotly Dash

- Visualized data with following features
  - Percentage of launches per site
  - Payload range
- This allowed a quick overview of the relations between launch sites and payloads and helped identifying where best to launch depending on the payload
- Github Url: DS_Capstone_Project/notebook_SpaceX_Interactive_Dashboard_L5vVtXuSg.ipynb at main · saschahoeche/DS_Capstone_Project · GitHub

Predictive Analysis (Classification)

- Four classification models were compared:
  - Logistic Regression
  - Support Vector Machine
  - Decision Tree
  - K – nearest neighbour



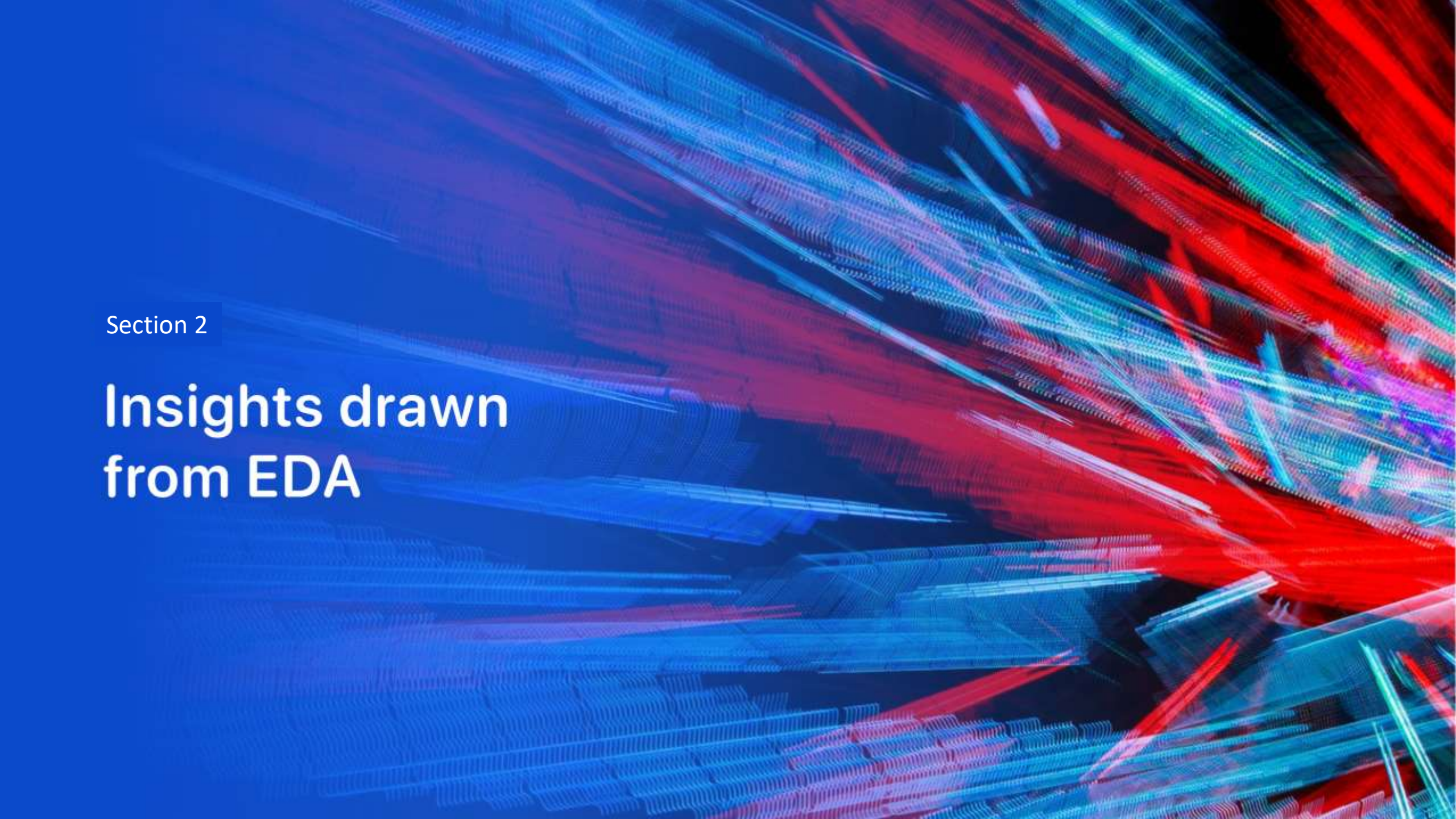
- Github Url: DS_Capstone_Project/notebook_Machine_Learning_Prediction_Part_5.ipynb at main · saschahoeche/DS_Capstone_Project · GitHub

Results

- Exploratory data analysis results
  - Four different launch sites are used
  - The average payload of Falcon 9 v1.1 is 2928kg
  - First successful landing was in 2015
  - Many above average payload Falcon 9 landed ob drone ships
  - Landing outcomes improved over the years significantly
  - Most launches happened on the east coast of the US

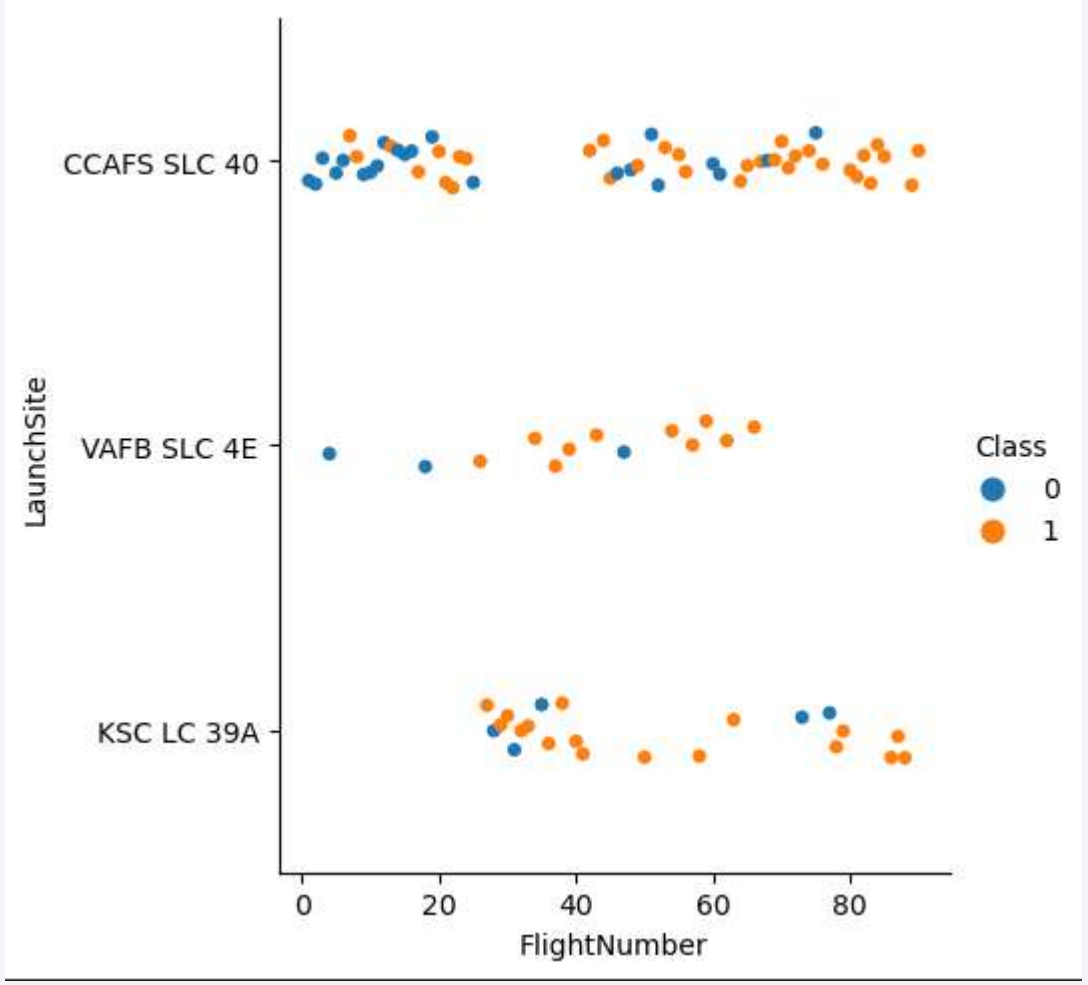
Results

- The thorough testing of different classification algorhythms showed that the Decision Tree Classifier was the most efficient to predict successful landings with an accuracy of 87%

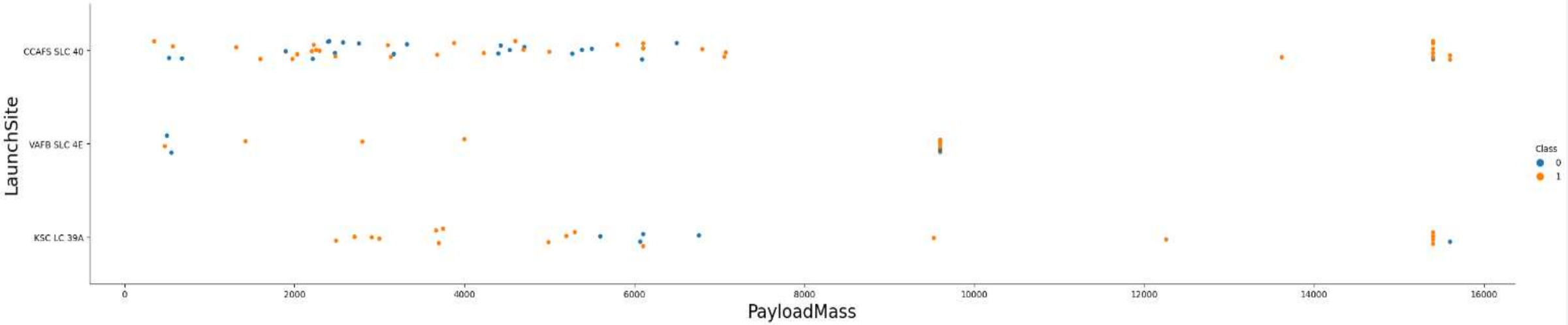

Section 2
Insights drawn
from EDA

Flight Number vs. Launch Site

- The best launch site for recent launches is CCAFS SLC 40
- The number of successful launches increased over time



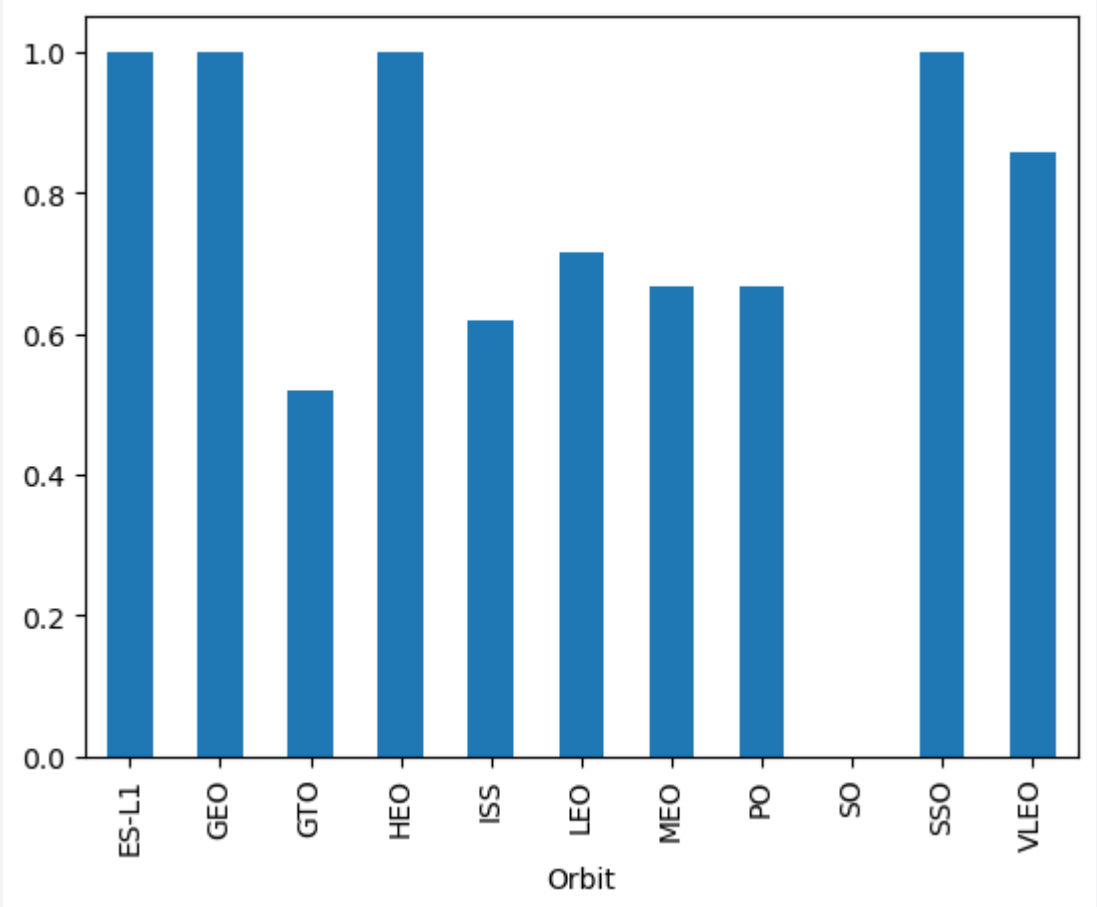
Payload vs. Launch Site



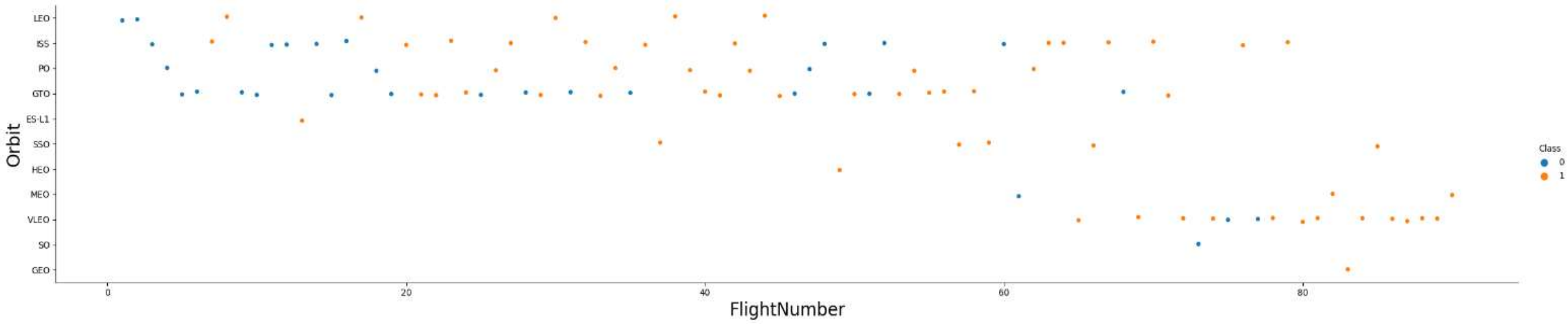
- Payloads over 9000kg have the highest success rate
- Payloads over 12000kg have only be performed on CCAFS SLC 40 and KSC LC 39A

Success Rate vs. Orbit Type

- ES-L1, GEO, HEO and SSO have been launched without failure

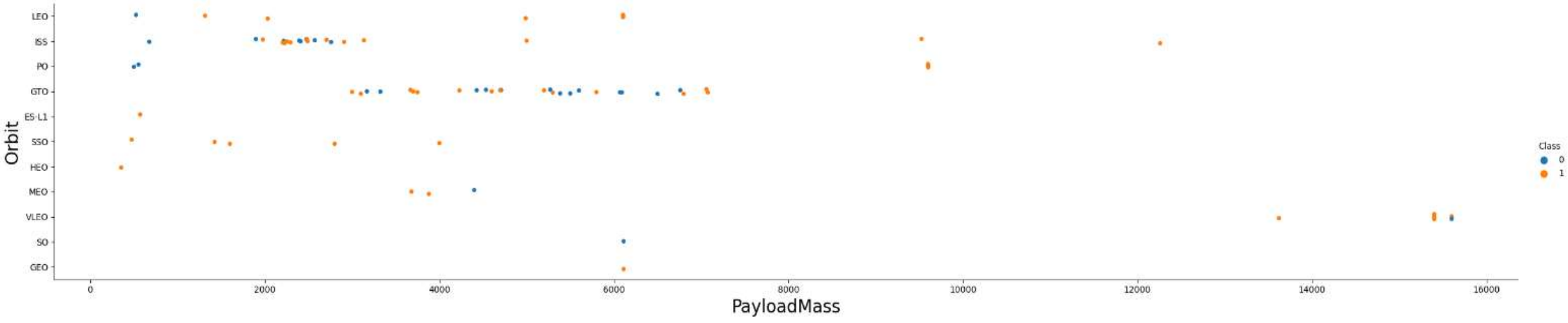


Flight Number vs. Orbit Type



- Success rate improved over time with all Orbit types
- Launches into VLEO orbit have increased recently

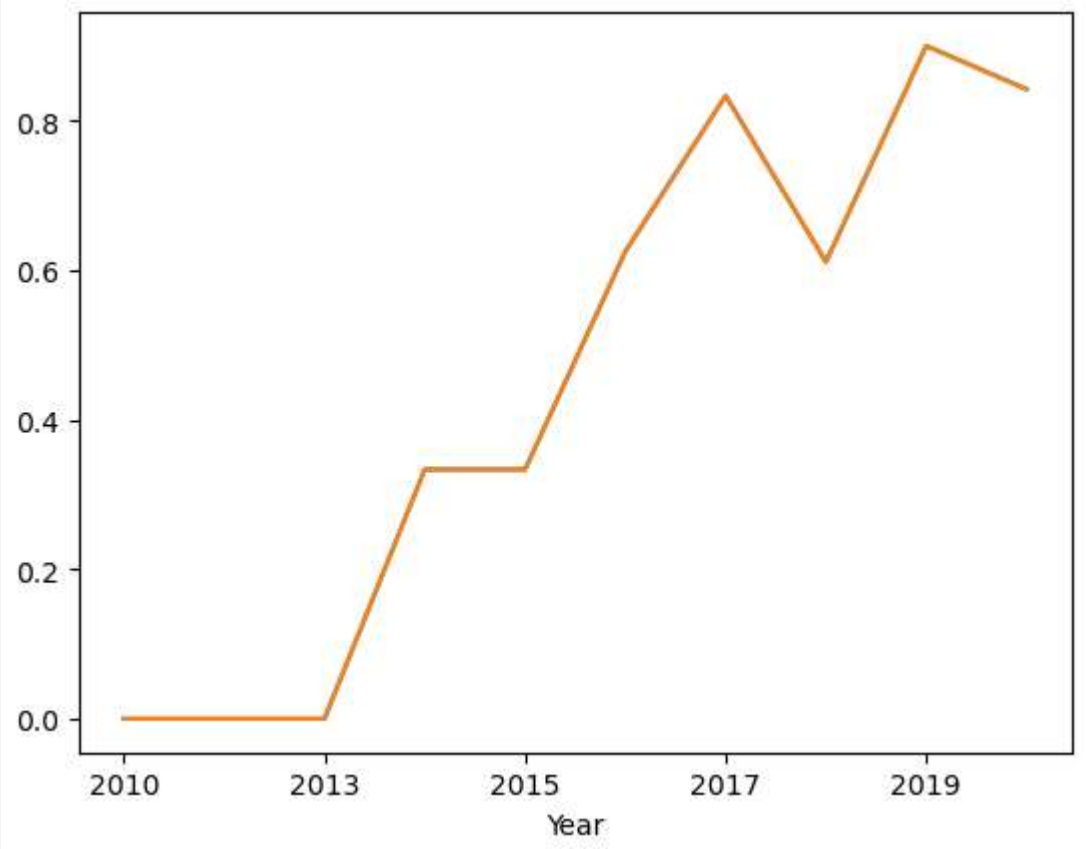
Payload vs. Orbit Type



- No relationship between payload and GTO orbit success rate
- Launches to ISS orbit have the highest range of payloads with reasonably high success rates
- SO and GEO orbits are used the least

Launch Success Yearly Trend

- Steadily increasing launch success rate
- No successes during the first three years



All Launch Site Names

- There are four launch sites:

Launch Sites
CCAFS LC 40
CCAFS SLC 40
KSC LC 39A
VAFB SLC 4E

- These have been extracted from the “launch_sites” value in the dataset

Launch Site Names Begin with 'CCA'

- Five records of launch sites that begin with `CCA`

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
06/04/2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0.0	LEO	SpaceX	Success	Failure (parachute)
12/08/2010	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0.0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
22/05/2012	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525.0	LEO (ISS)	NASA (COTS)	Success	No attempt
10/08/2012	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500.0	LEO (ISS)	NASA (CRS)	Success	No attempt
03/01/2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677.0	LEO (ISS)	NASA (CRS)	Success	No attempt

- CCA stands for Cape Canaveral

Total Payload Mass

- Tpayload carried by boosters from NASA

Total Payload (kg)
111268

- The sum of all payloads containing the code ‘CRS’

Average Payload Mass by F9 v1.1

- Average payload mass carried by booster version F9 v1.1

Average Payload (kg)
2928

- Data filtered by booster version and then averaged

First Successful Ground Landing Date

- First successful landing outcome on ground pad

Min Date
2015-12-22

- Filtering the data for the first successful landing on a ground pad returned above date

Successful Drone Ship Landing with Payload between 4000 and 6000

- Boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

Booster Version
F9 FT B1021.2
F9 FT B1031.2
F9 FT B1022
F9 FT B1026

- Filtering booster data according to above limitations

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes

Mission Outcome	Occurrences
Success	99
Success (payload status unclear)	1
Failure (in flight)	1

- Grouping mission outcomes by counts for each group

Boosters Carried Maximum Payload

- Booster which have carried the maximum payload mass
- The listed boosters have carried the maximum payload amount in the dataset

Booster Version
F9 B5 B1048.4
F9 B5 B1048.5
F9 B5 B1049.4
F9 B5 B1049.5
F9 B5 B1049.7
F9 B5 B1051.3
F9 B5 B1051.4
F9 B5 B1051.6
F9 B5 B1056.4
F9 B5 B1058.3
F9 B5 B1060.2
F9 B5 B1060.3

2015 Launch Records

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

Booster Version	Launch Site
F9 v1.1 B1012	CCAFS LC 40
F9 v1.1 B1015	CCAFS LC 40

- Only two drone ship landings failed in 2015

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

- Rank all outcomes between 2010-06-04 and 2017-03-20:

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

- Data shows that most attempts have been aborted in the above time frame

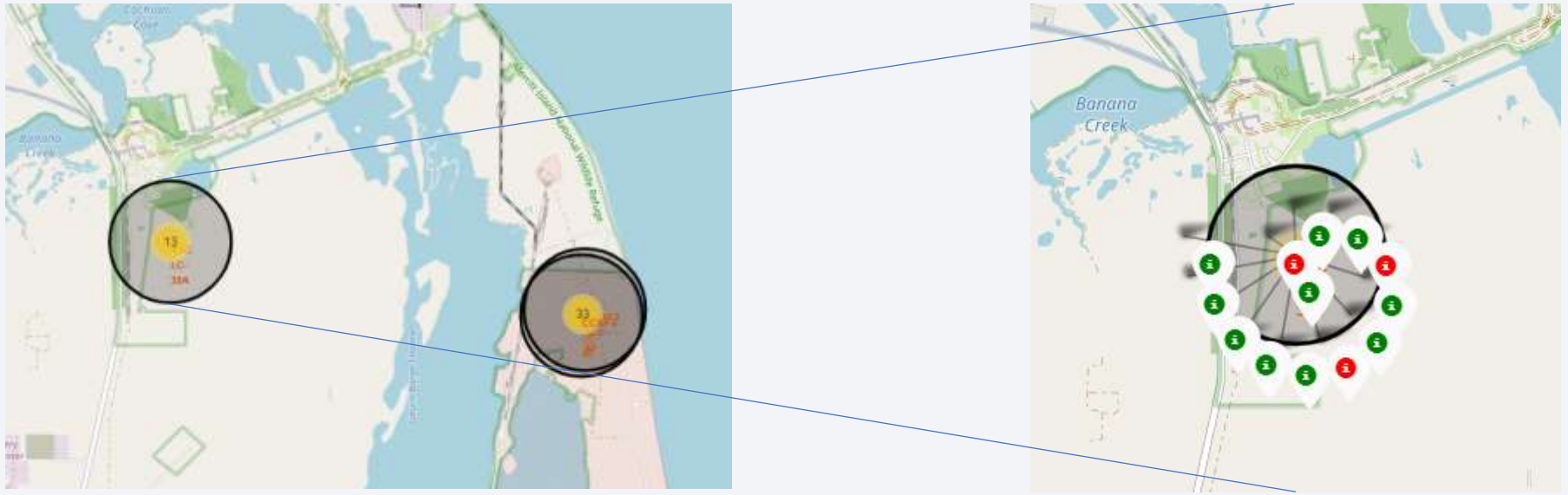
Section 3

Launch Sites Proximities Analysis

Launch Sites

- Launch Sites are usually located near the ocean for safety reasons

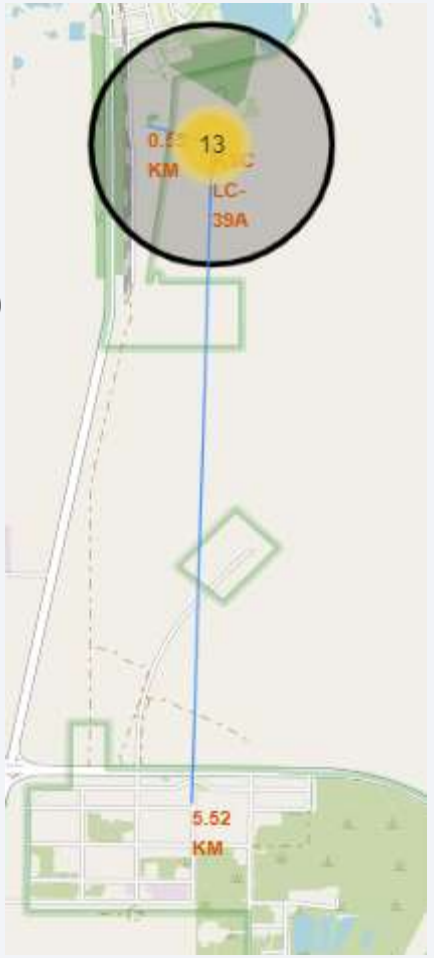
Launch Outcomes



- Example launch outcomes for KSC LC 39A, red = failure, green = success

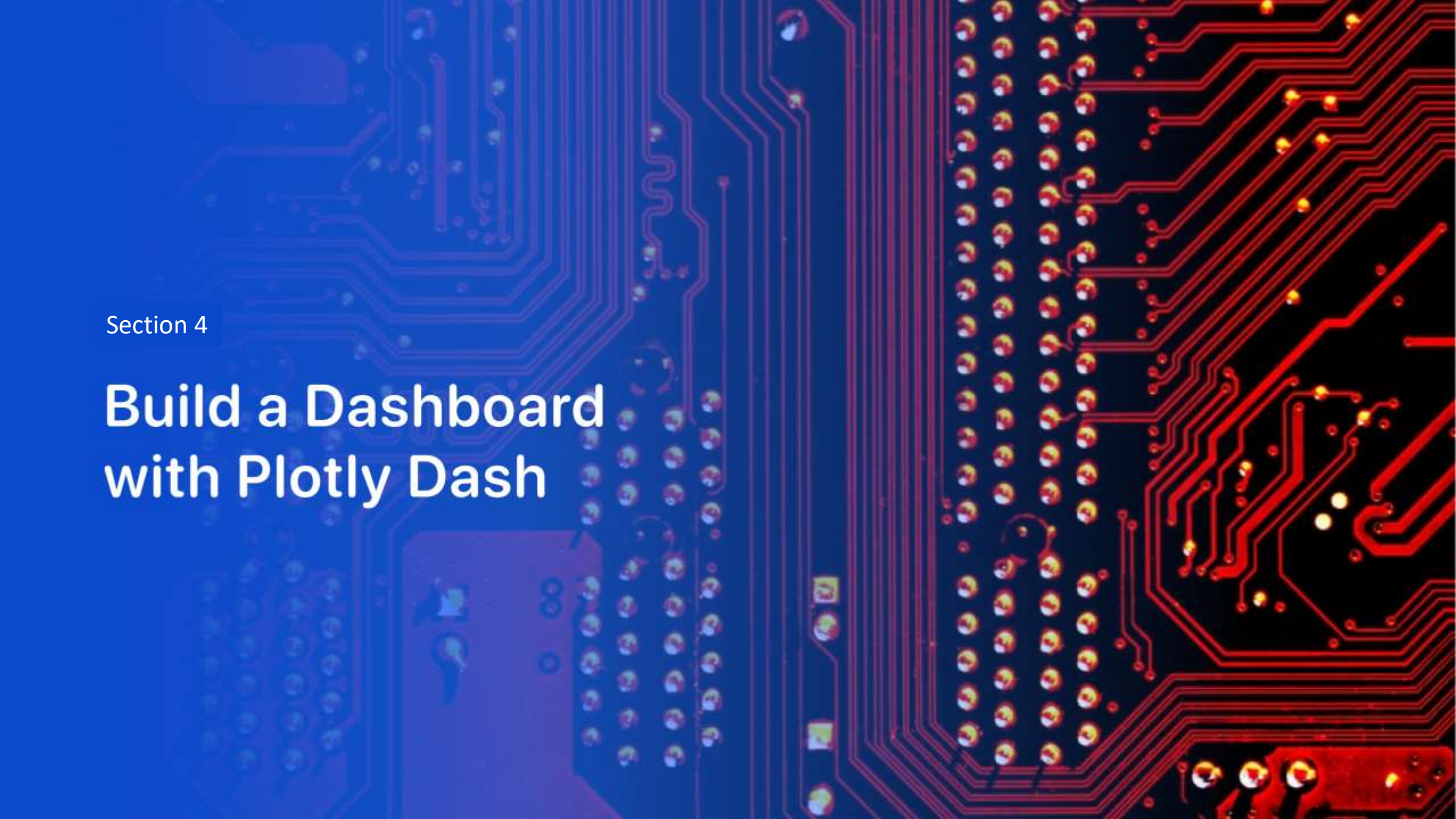
Logistics and Safety

- As displayed KSC LC 39A is 0.55km close to the next rail connection for easy transport
- And it is 5.52km away from the next urban area fo safety and noise reasons

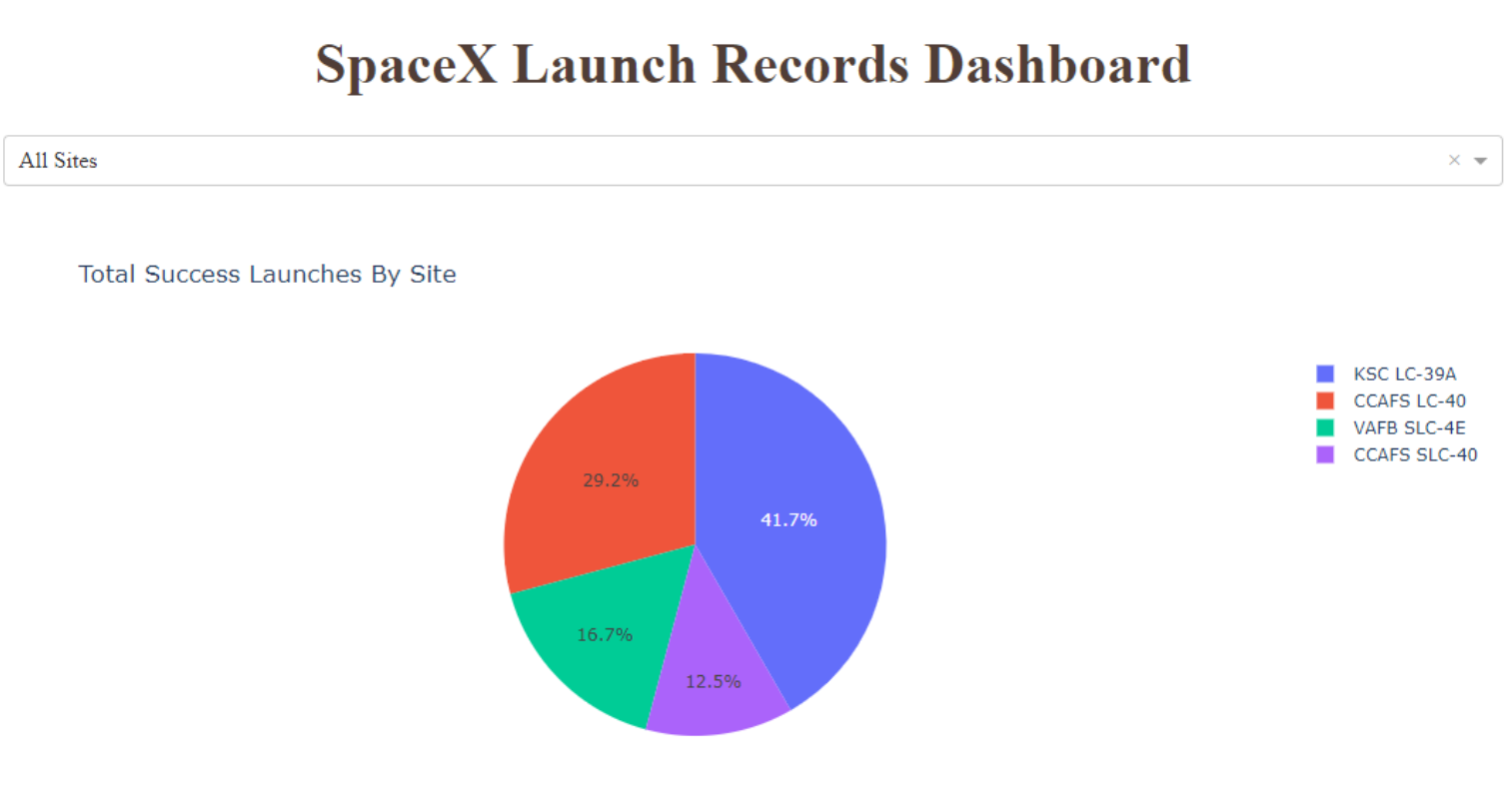


Section 4

Build a Dashboard with Plotly Dash

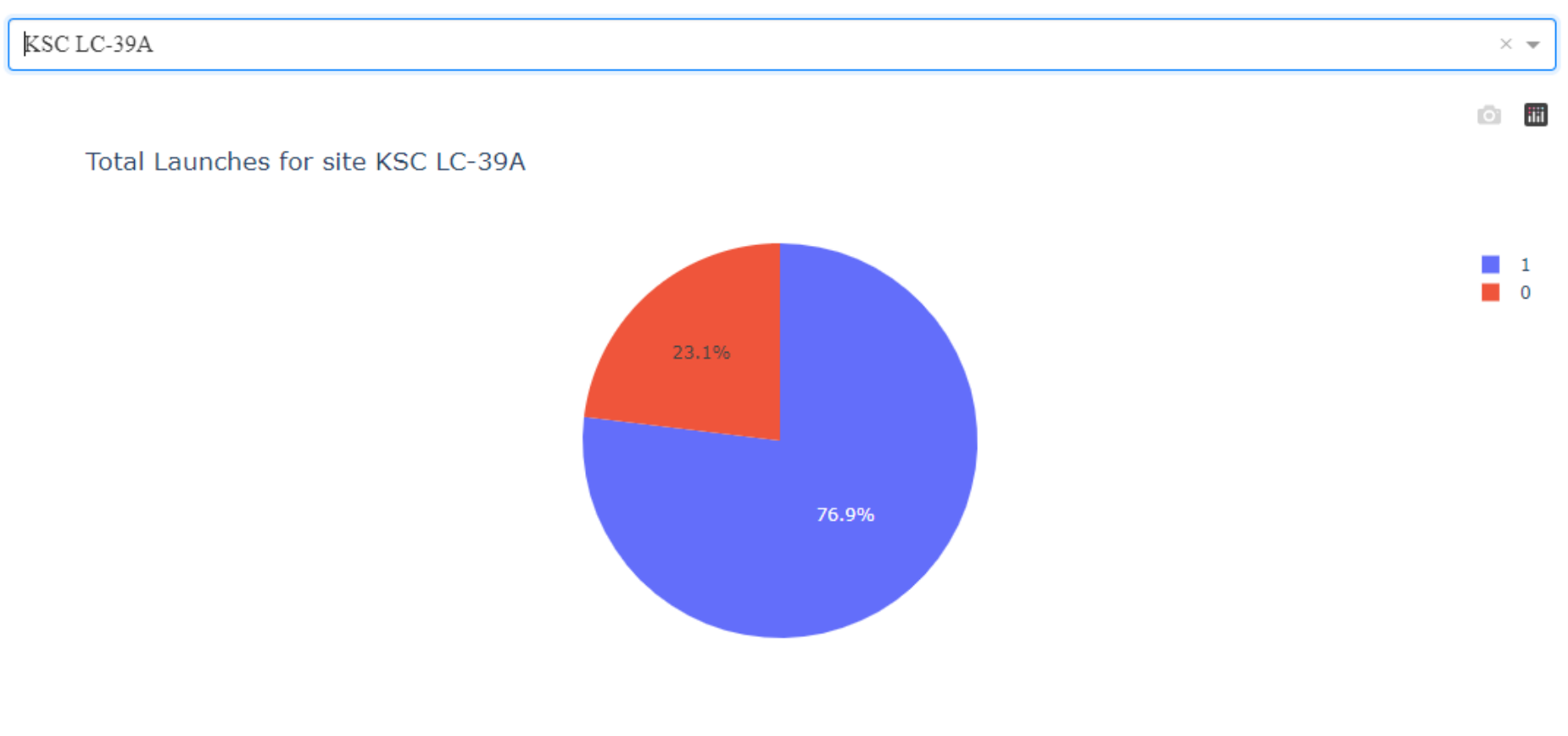


Successful Launches by Site



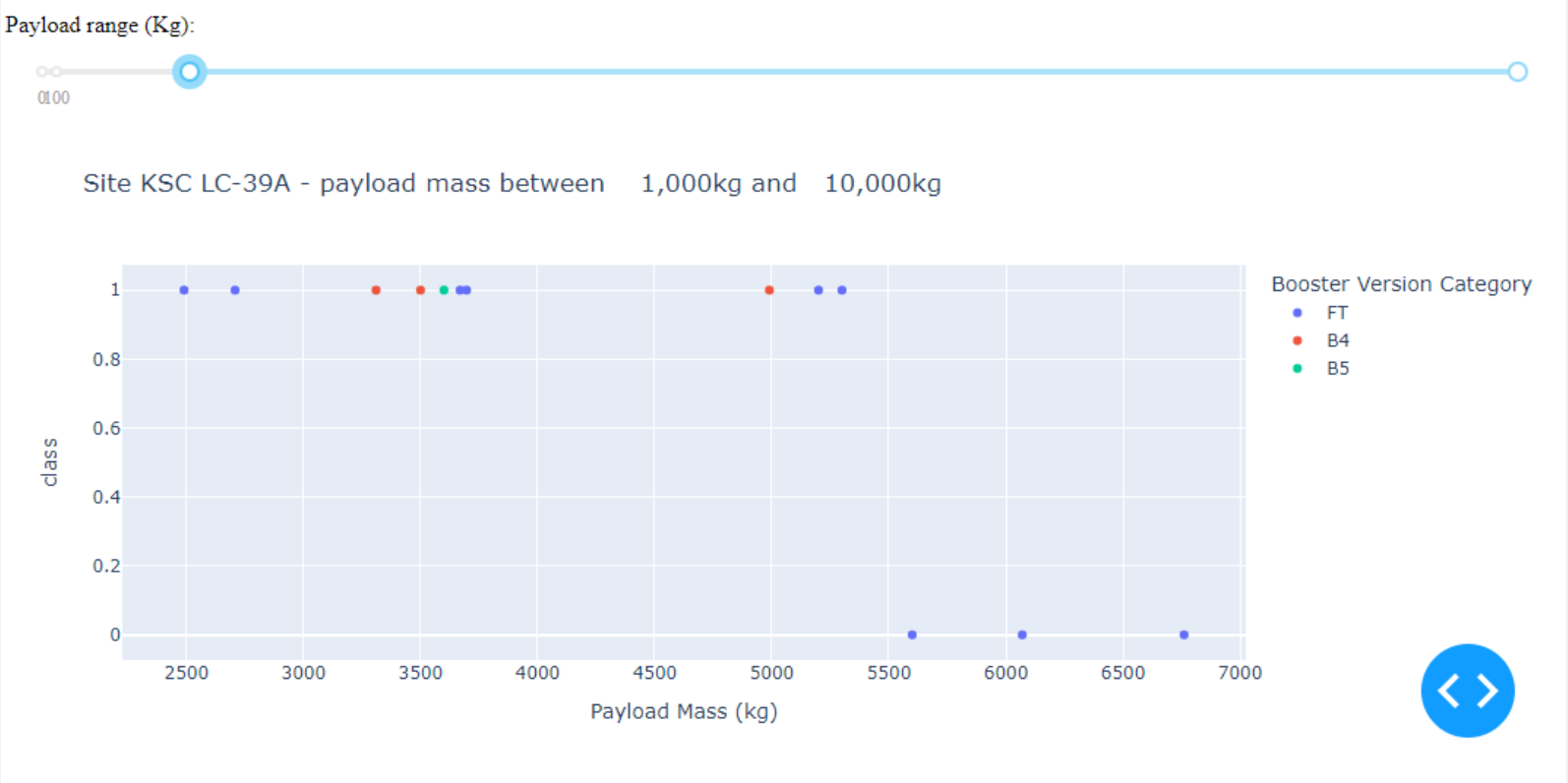
- Identifying the most successful launch sites

<Dashboard Screenshot 2>



- Three quarters of the launches at KSC LC 39A are successful

Payload vs. Launch Outcome



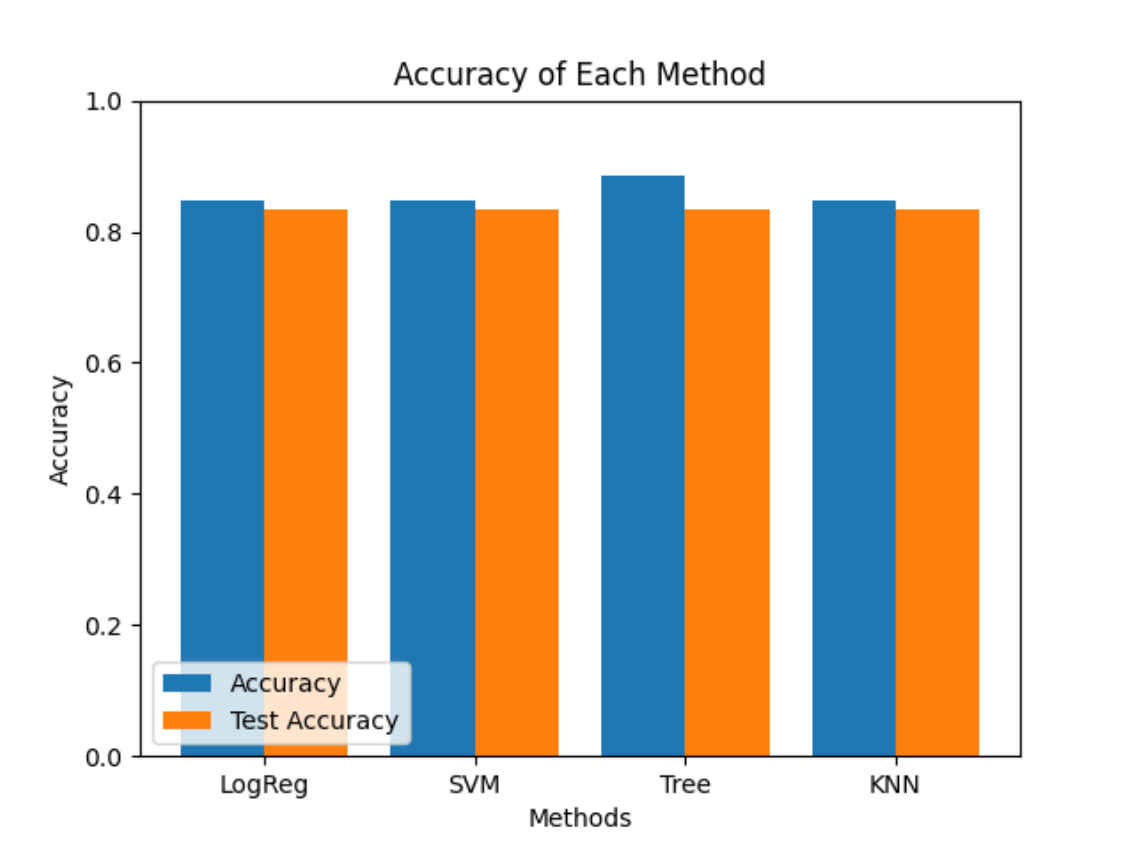
- Payloads under 6000kg ant FT boosters are the most successful

Section 5

Predictive Analysis (Classification)

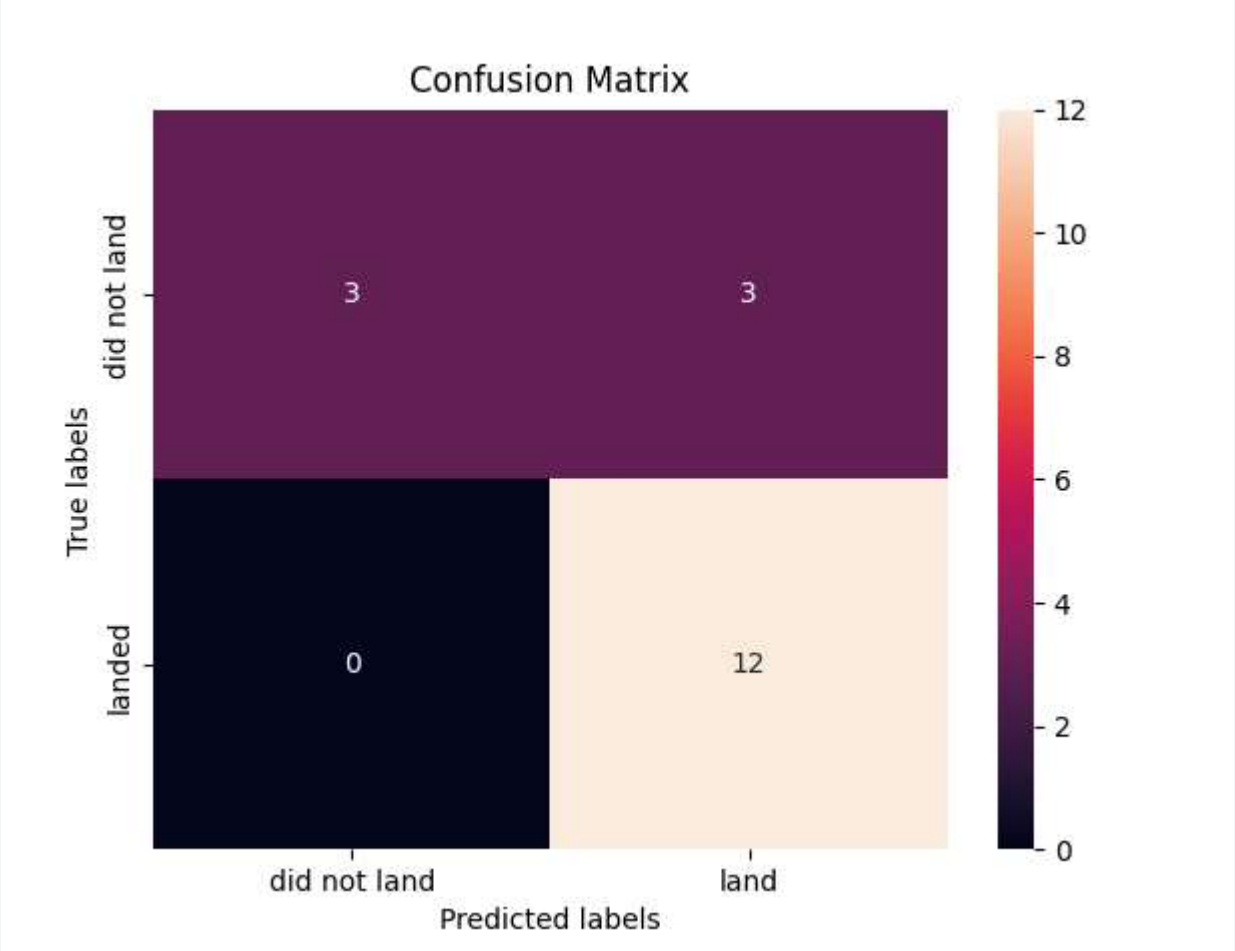
Classification Accuracy

- Four classification models were tested
- Decision tree classifier has the highest accuracy with 87%



Confusion Matrix Decision Tree

- Decision Tree Classifier shows the highest number of true positives



Conclusions

- Two data sources were analyzed, both public
- The best launch site is KSC LC 39A
- Launches over 7000kg have less chance of failure
- The majority of missions are successful
- The percentage of successful landings increased over time due to improvements in rocket design
- Decision Tree classifier can predict a successful landing with 87% accuracy

Thank you!