IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Thuan Pham
May 22, 2024

Outline

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

Executive Summary

- This presentation will detail findings from analysis of SpaceX Falcon 9 data
- Methodologies used in this report include the following:
  - Data collection via API and Web Scraping
  - Data transformation, wrangling, and formatting
  - Exploratory data analysis using SQL queries and data visuals
  - Built dashboard to display graphs & charts showing data relationships
  - Using machine learning for predictions
- Summary of all results

Introduction

- Background
  - SpaceX's Falcon 9 uses a multi-stage rocket platform
  - The first stage of rocket needs to successfully land allowing it to be reused
  - Reuse allows Falcon 9 more competitive edge, reducing the cost of the platform
- Problem
  - Not all first stage rocket will be successfully land
  - Not all first stage rocket will be reused
  - Need to determine success rate to help determine price

Section 1

Methodology

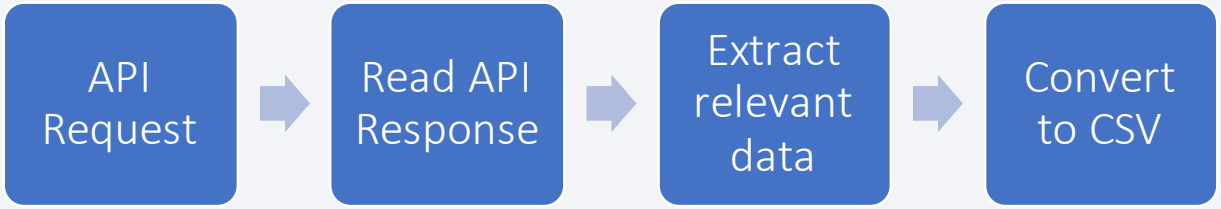
Methodology

Executive Summary

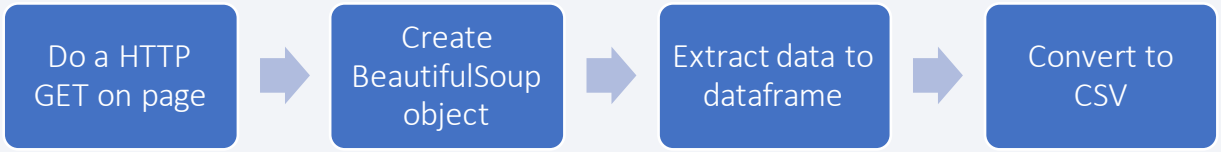
- Data collection methodology:
  - Get data from SpaceX's API
  - Using web scrapping techniques to gather Falcon 9 data from SpaceX wiki page
- Perform data wrangling
  - Updated data into successful and unsuccessful landings
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Logistic regression, Support vector machine(SVM), Decision tree, KNN

Data Collection

- The data collection process included:
  - API GET requests from SpaceX's public API

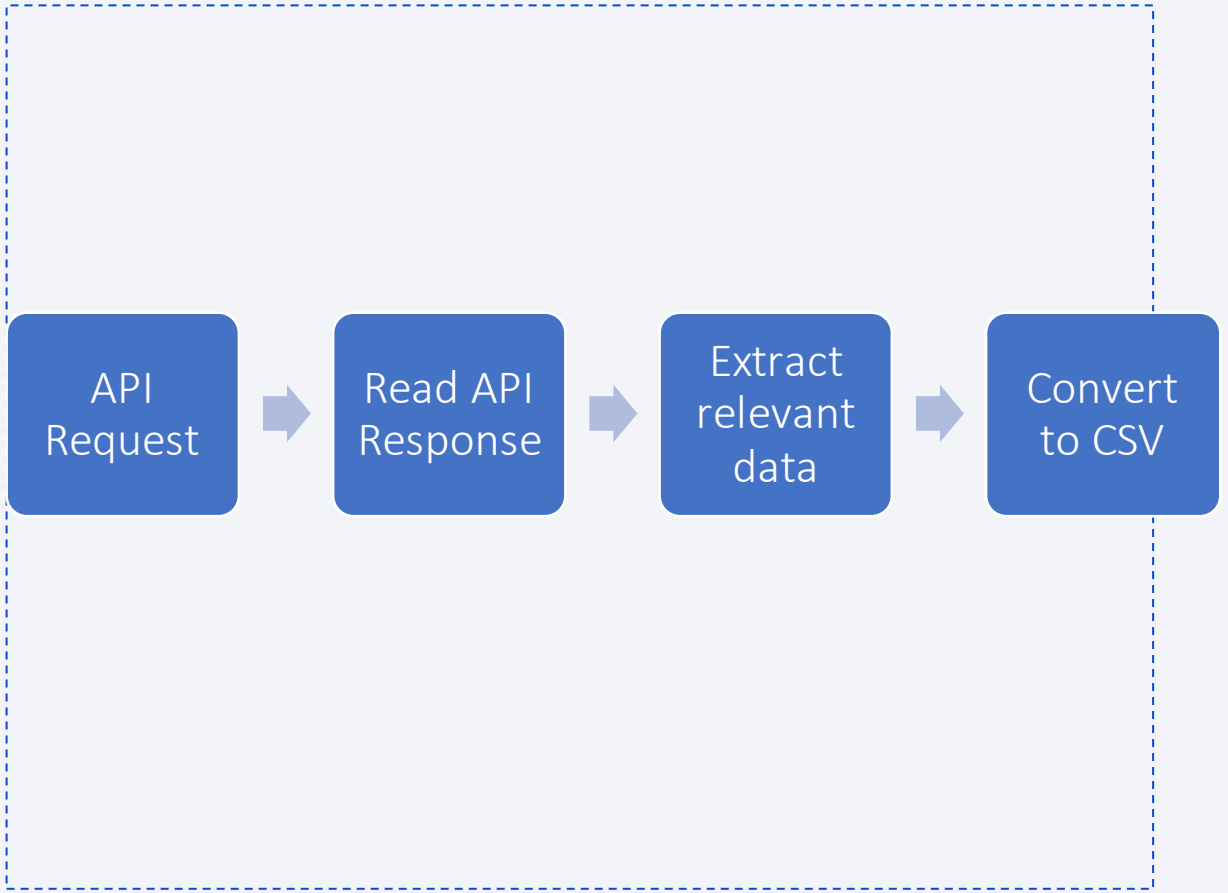


  - Web scraping from SpaceX's page on Wikipedia



Data Collection – SpaceX API

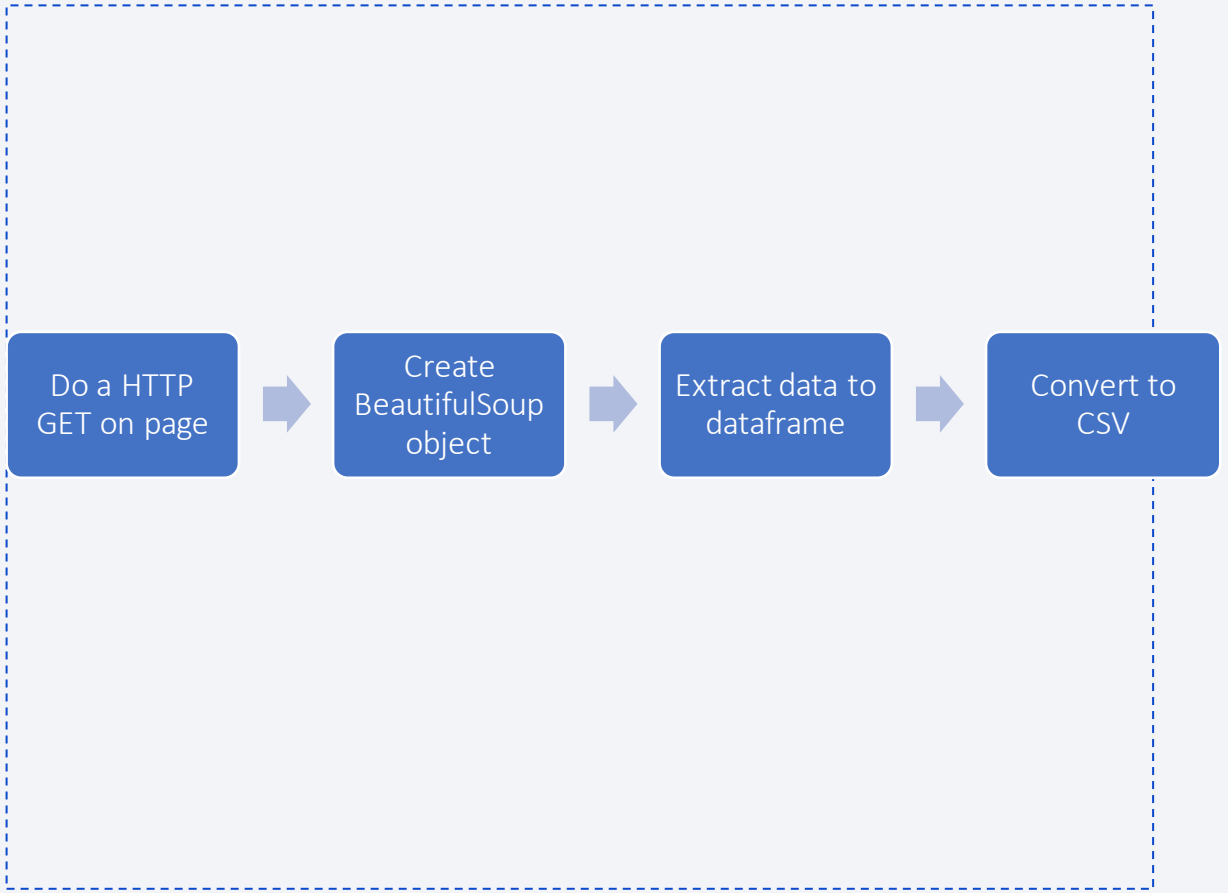
- SpaceX REST calls
  - spacex_url="https://api.spacexdata.com/v4/launches/past"
- Github



Data Collection - Scraping

- Scraping calls:
  - static_url = https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922
  - Html_data = requests.get(static_url)
  - Soup = BeautifulSoup(reponse.content, "html.parser")

- Github



Data Wrangling

- Created training label with landing outcomes
  - 0 = Fail
  - 1 = Success
- Fails include:
  - None None, False ASDS, None ASDS, False Ocean, False RTLS
- Success include:
  - True ASDS, True RTLS, True Ocean
- GitHub URL

EDA with Data Visualization

- In Exploratory Data Analysis, charts and plots were create for insights on the data
  - Scatter plot
  - Bar chart
  - Line chart
- GitHub URL

EDA with SQL

- SQL queries on data performed
  - Unique launch sites
  - Data from specific launch site
  - Total payload from specific customer
  - Average payload from booster version
  - Success from specific landing outcome
  - Distinct booster versions
  - Distinct mission outcomes
- GitHub URL

Build an Interactive Map with Folium

- Used Folium library to create maps with markings of launch sites.
- The maps and markings showed proximity to certain landmarks which could be factors for the Falcon 9
- GitHub URL

Build a Dashboard with Plotly Dash

- Created Dashboard for displaying spaceX data
- Includes pie charts and scatter plots
- Pie chart show success rate
- Scatter plot show launch sites, payload mass, booster version categories

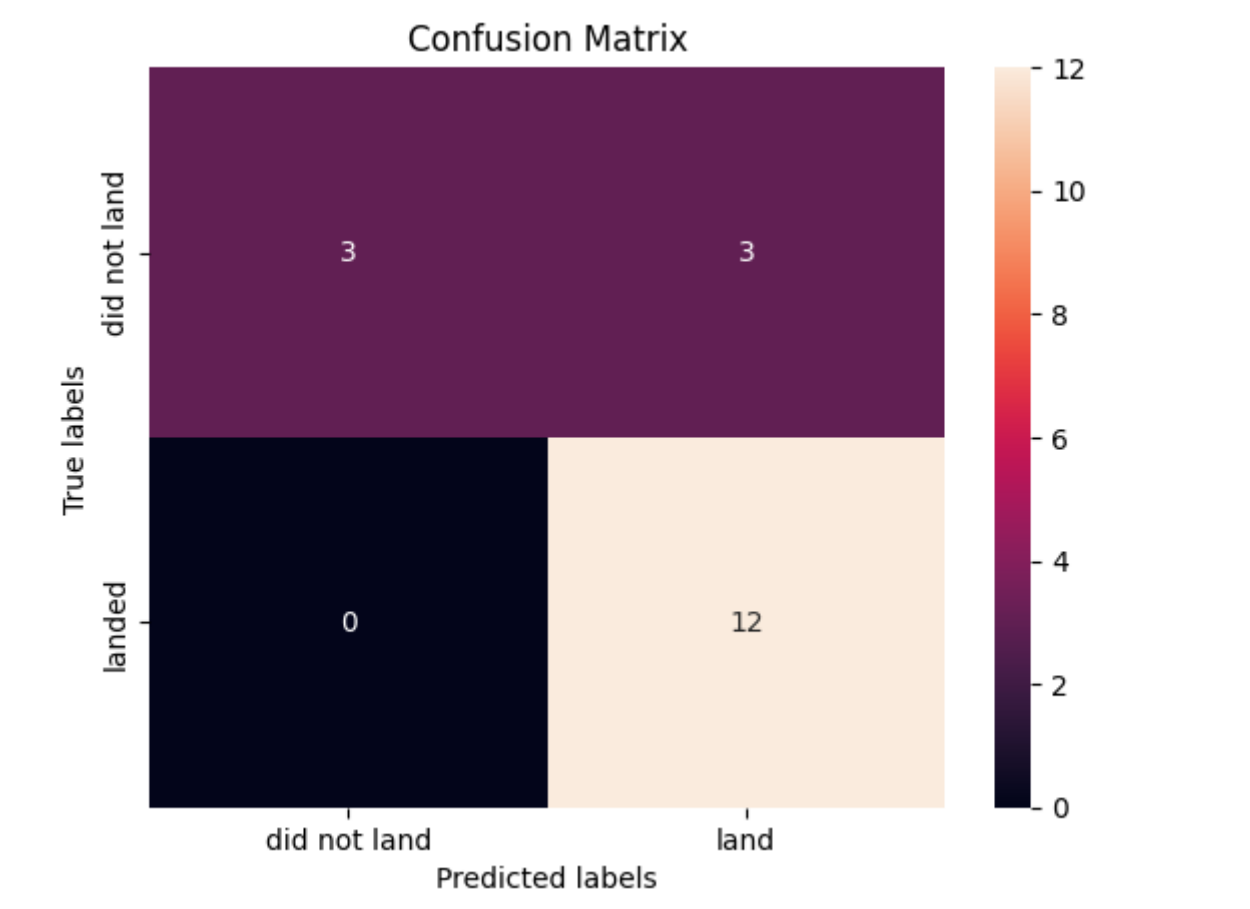
- GitHub URL

Predictive Analysis (Classification)

- Loaded data in dataframe
- Standardized data
- Data was trained test split
- Created models to predict, confusion matrix

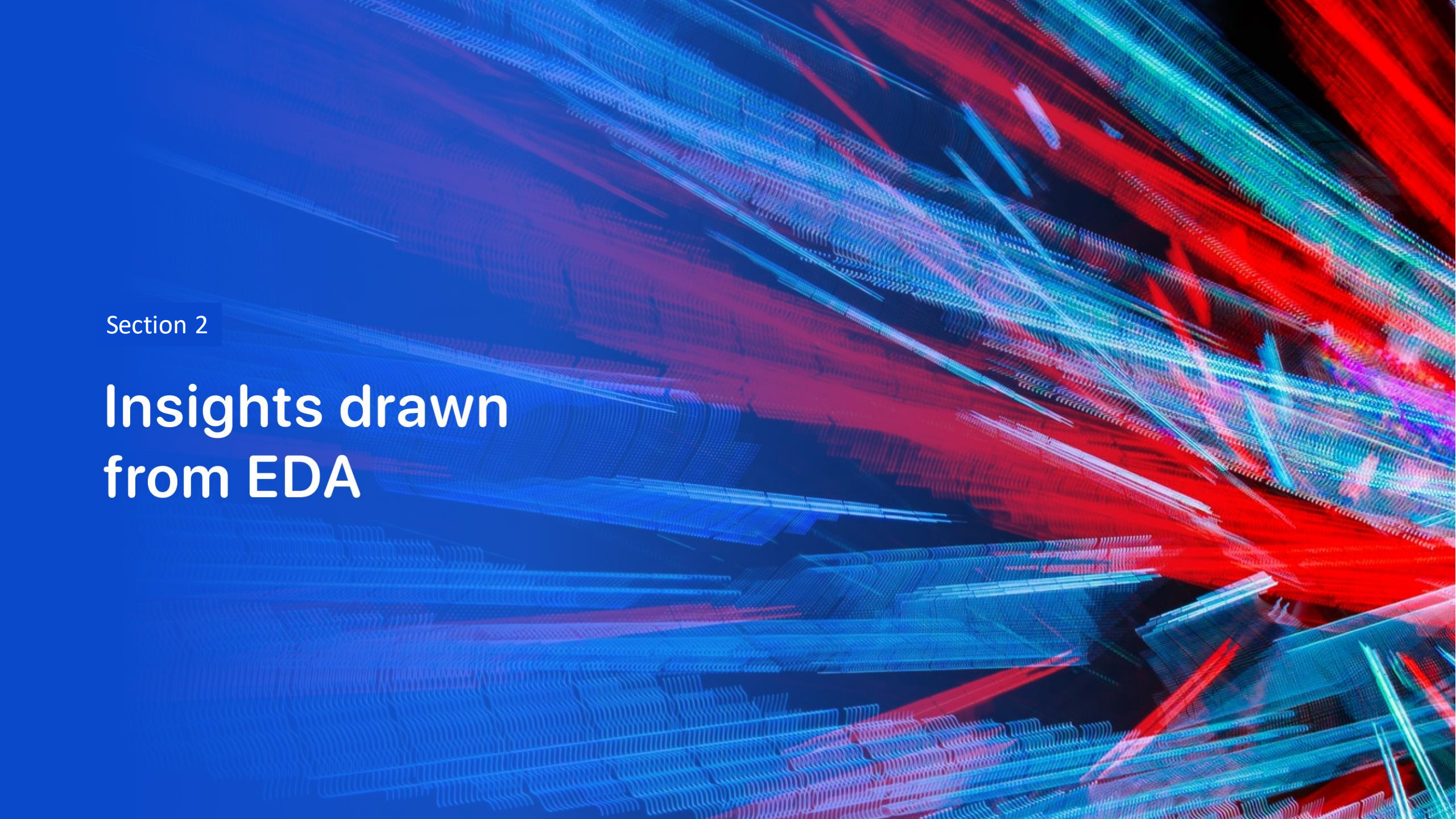
- GitHub URL

Results

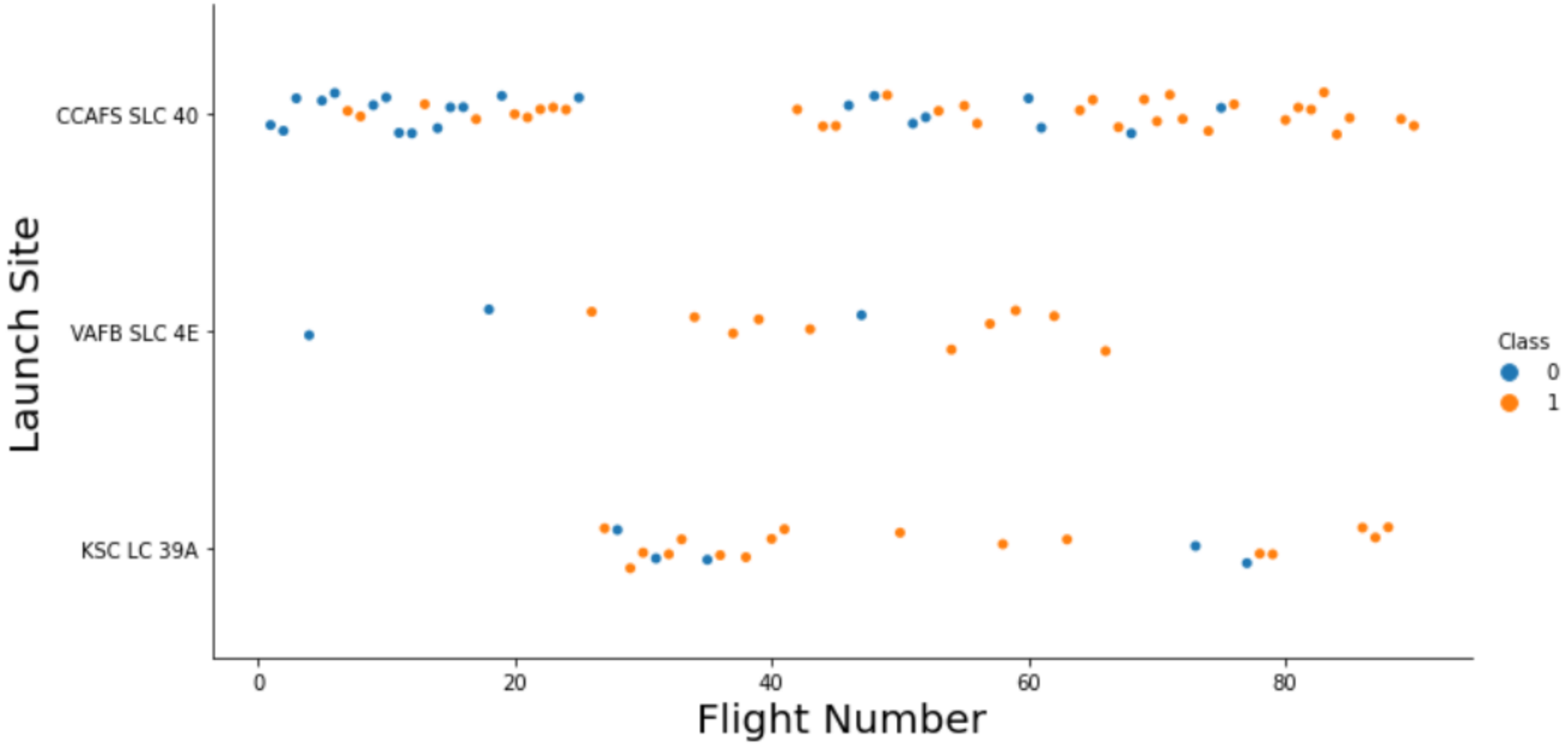

Confusion Matrix
True labels
did not land
landed
3
3
0
12
did not land
land
Predicted labels
12
10
8
6
4
2
0

Section 2

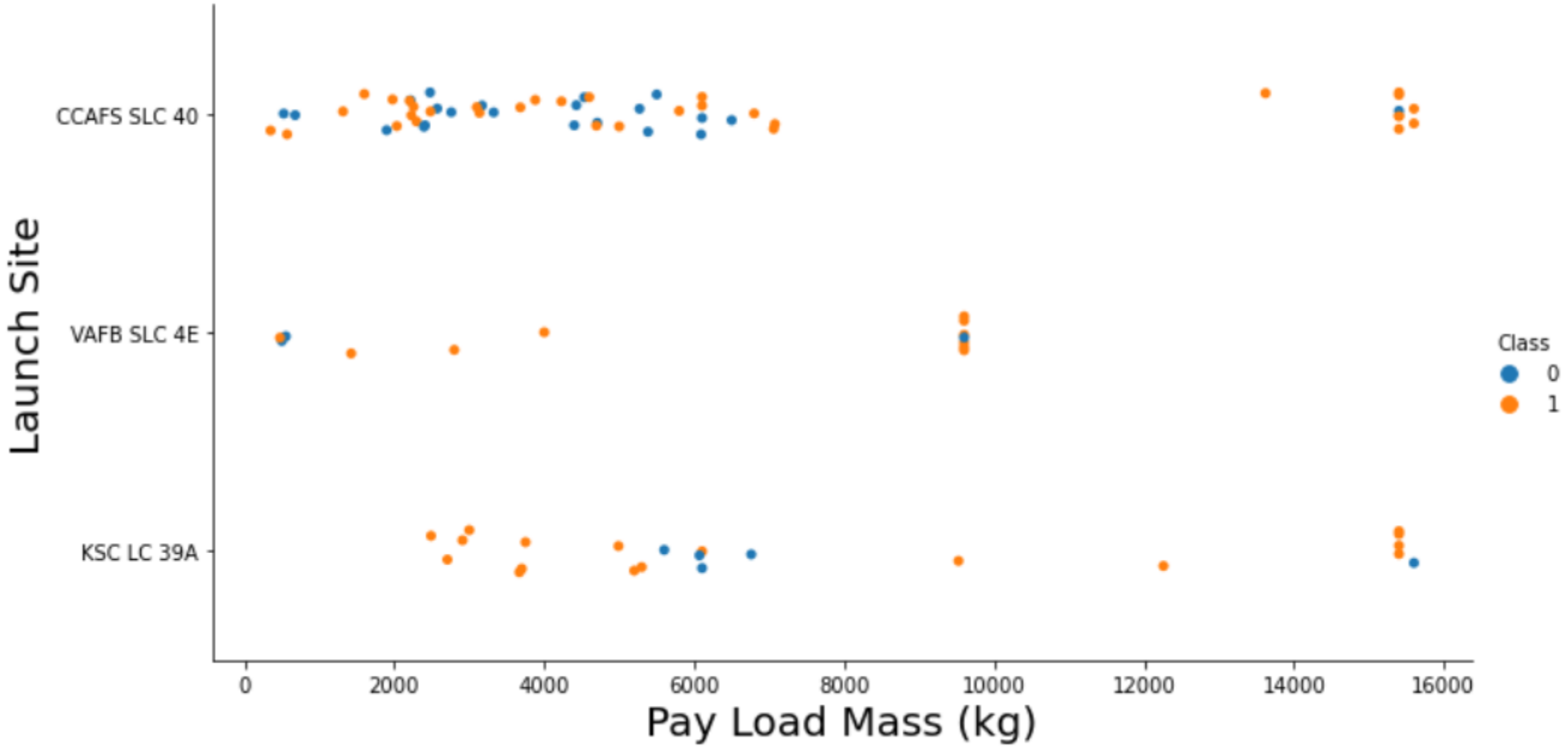
Insights drawn from EDA



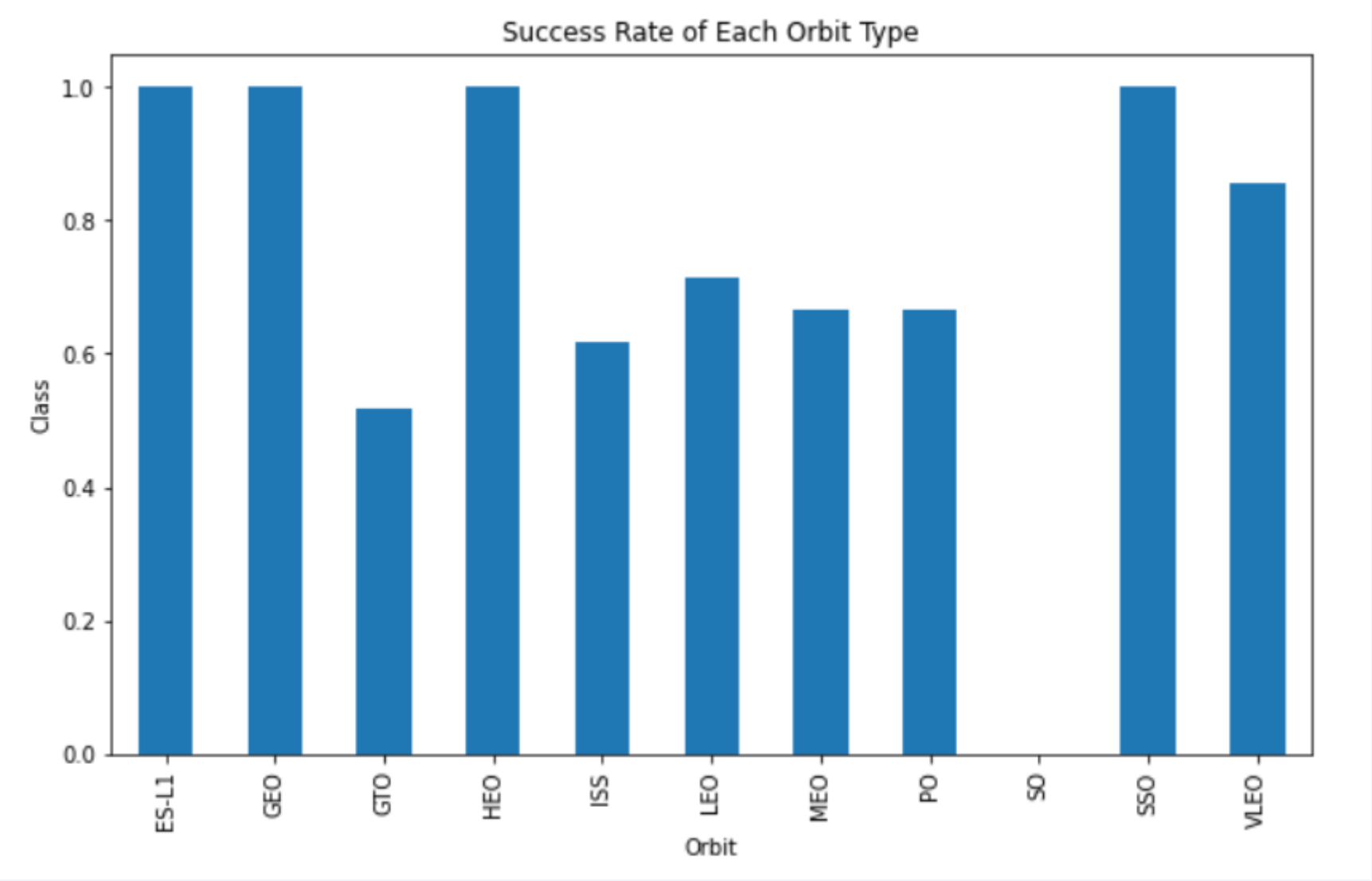
Flight Number vs. Launch Site



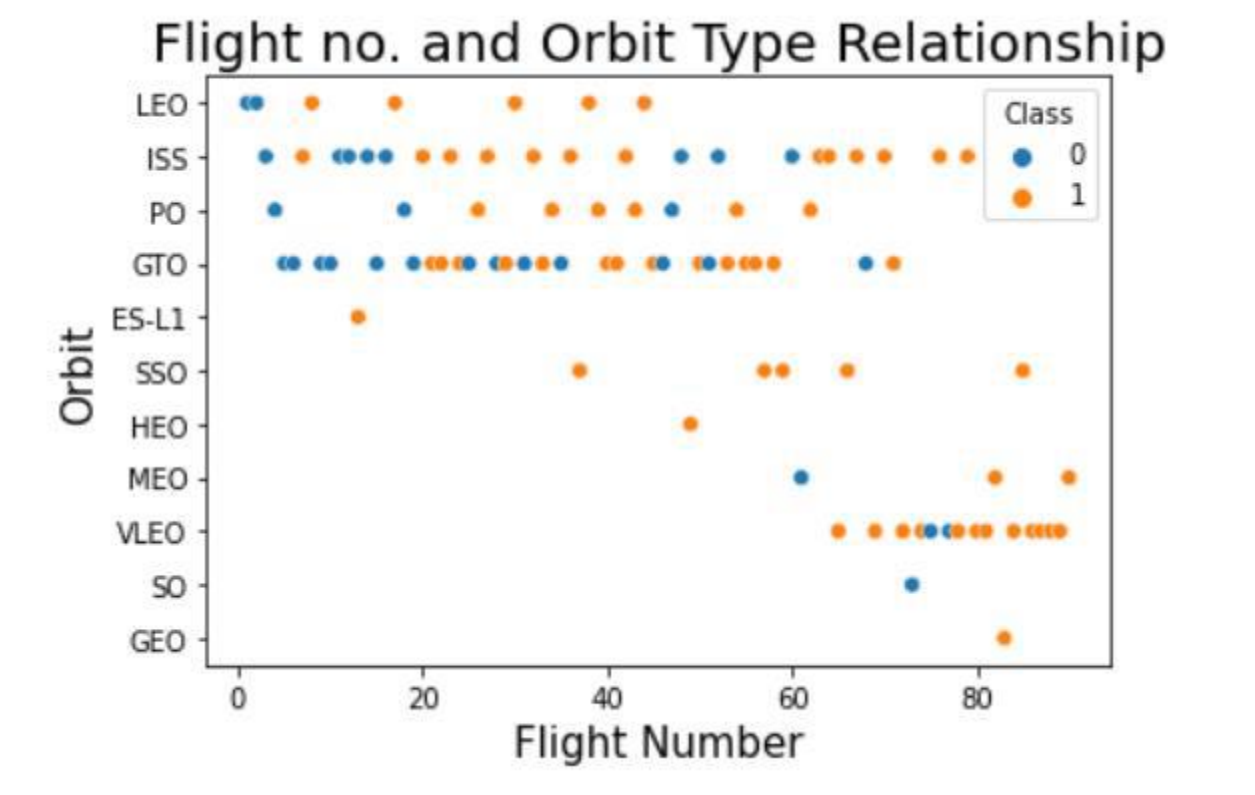
Payload vs. Launch Site



Success Rate vs. Orbit Type

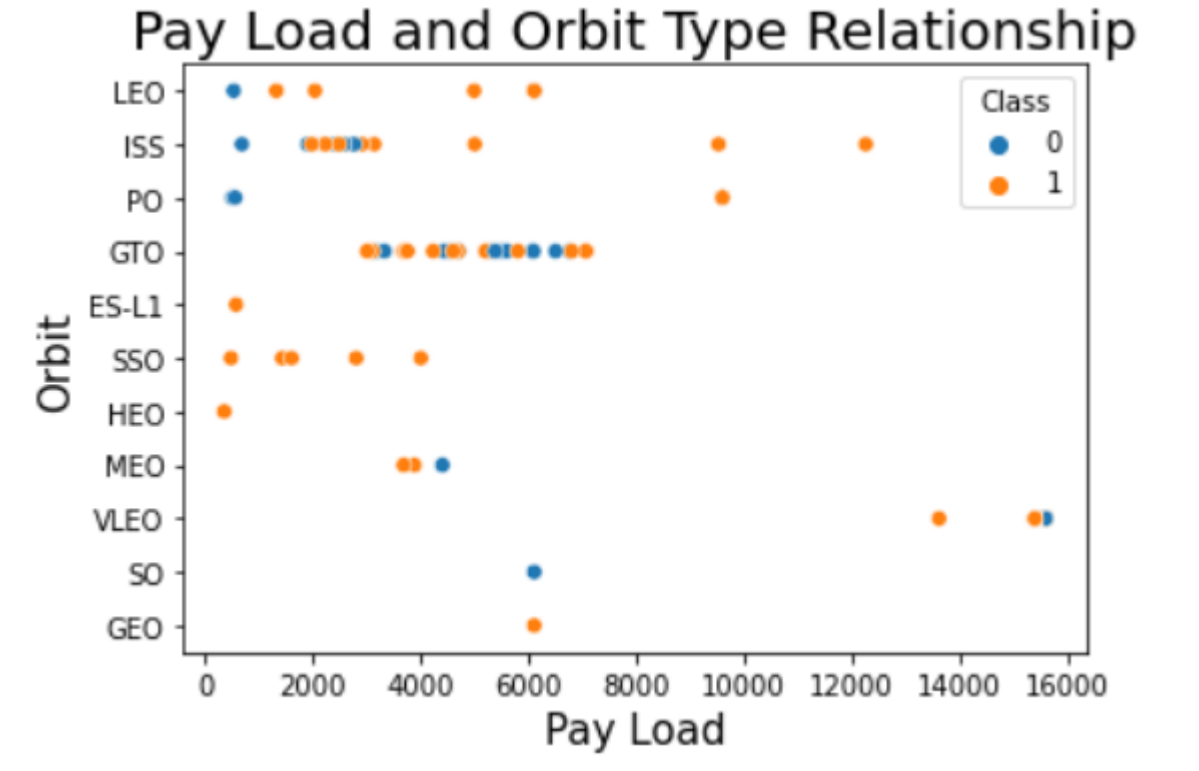


Flight Number vs. Orbit Type



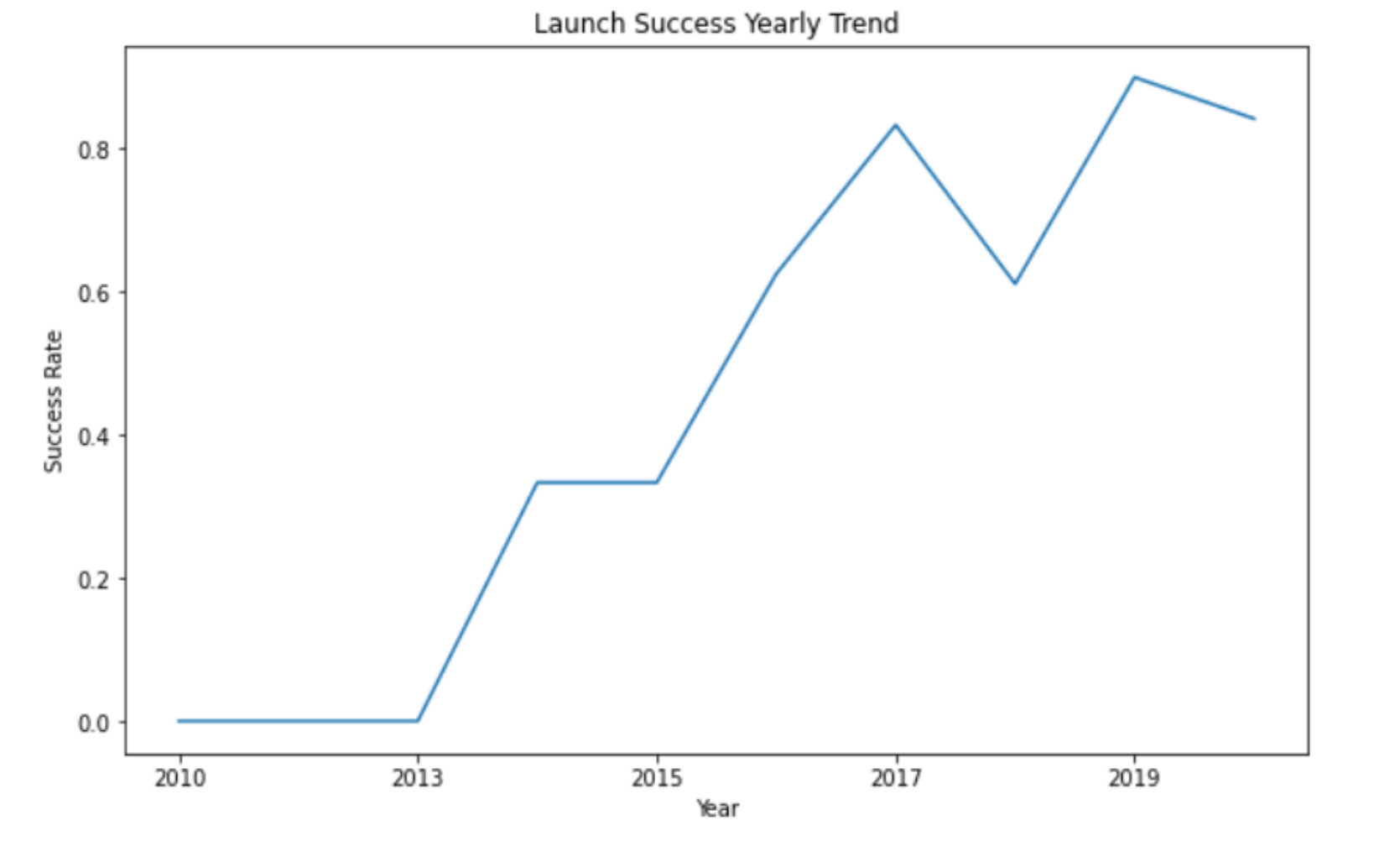
- No relationship

Payload vs. Orbit Type



- Success landing = class 1
- Failed landing = class 0

Launch Success Yearly Trend



All Launch Site Names

- SQL query: %sql select distinct launch_site from spacextable

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

Launch Site Names Begin with 'CCA'

- Query: %sql select * from spacextable where Launch_site like 'CCA%' limit 5

DATE	time__utc_	booster_version	launch_site	payload	payload_mass__kg_	orbit	customer	mission_outcome	landing__outcome
2010-04-06	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-08-12	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
2012-08-10	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-01-03	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

- Query: %sql select sum(PAYLOAD_MASS__KG_) from spacextable where Customer = 'NASA (CRS)'
- Result: 45596

Average Payload Mass by F9 v1.1

- %sql select avg(PAYLOAD_MASS__KG_) from spacextable group by booster_version having booster version = 'F9 v1.1'
- Result: 2928.4

First Successful Ground Landing Date

- %sql select min(Date) from spacextable where Landing_Outcome = 'Success (ground pad)'
- Result: 2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- Query: %sql select distinct booster_version from spacextable where PAYLOAD_MASS__KG_> 4000 and PAYLOAD_MASS__KG_<6000 and landing_outcome = 'Success (drone ship)'
- Results: four items

booster_version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

- Query:
  - `%sql select mission_outcome, count(mission_outcome) from spacextable group by mission_outcome`
- Results: four items; might issue with data

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

Boosters Carried Maximum Payload

- Query
  - `%sql select distinct booster_version from spacextable where PAYLOAD_MASS__KG_ = (select max(PAYLOAD_MASS__KG_) from spacextable)`
- Results:

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:
  - `%sql select Landing_outcome, Booster_Version, Launch_Site from spacextable where landing_outcome = 'Failure (drone ship)' and year(Date) = '2015'`

- Result:

landing__outcome	booster_version	launch_site
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

- Query
  - `%sql select Landing_outcome, count(*) as Landingcounts from spacextable where date between '2010-06-4' and '2017-03-20' group by landing_outcome order by count(*) desc;`

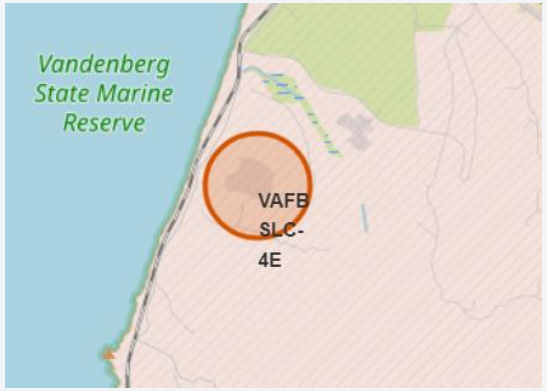
- Result:

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

Section 3

Launch Sites Proximities Analysis

SpaceX Falcon 9 Launch Sites

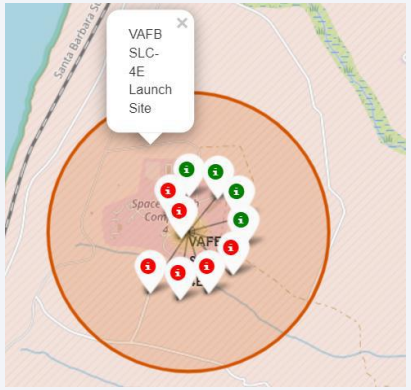


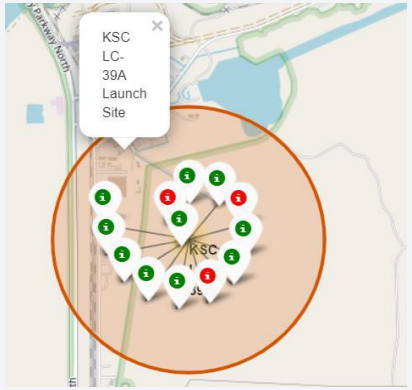
Launch site showing:

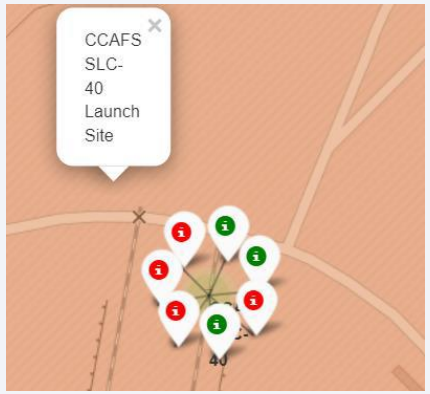
- VAFB SLC-4E (CA)
- CCASFS LC-40 (FL)
- KSC LC-39A (FL)
- CCAFS SLC-40 (FL)

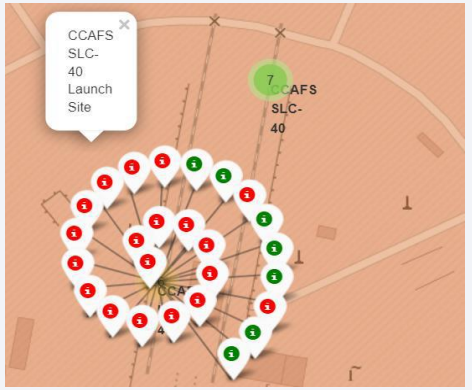
Launch sites with details of success/fails

Details of launch sites and each lauch with their success or failure



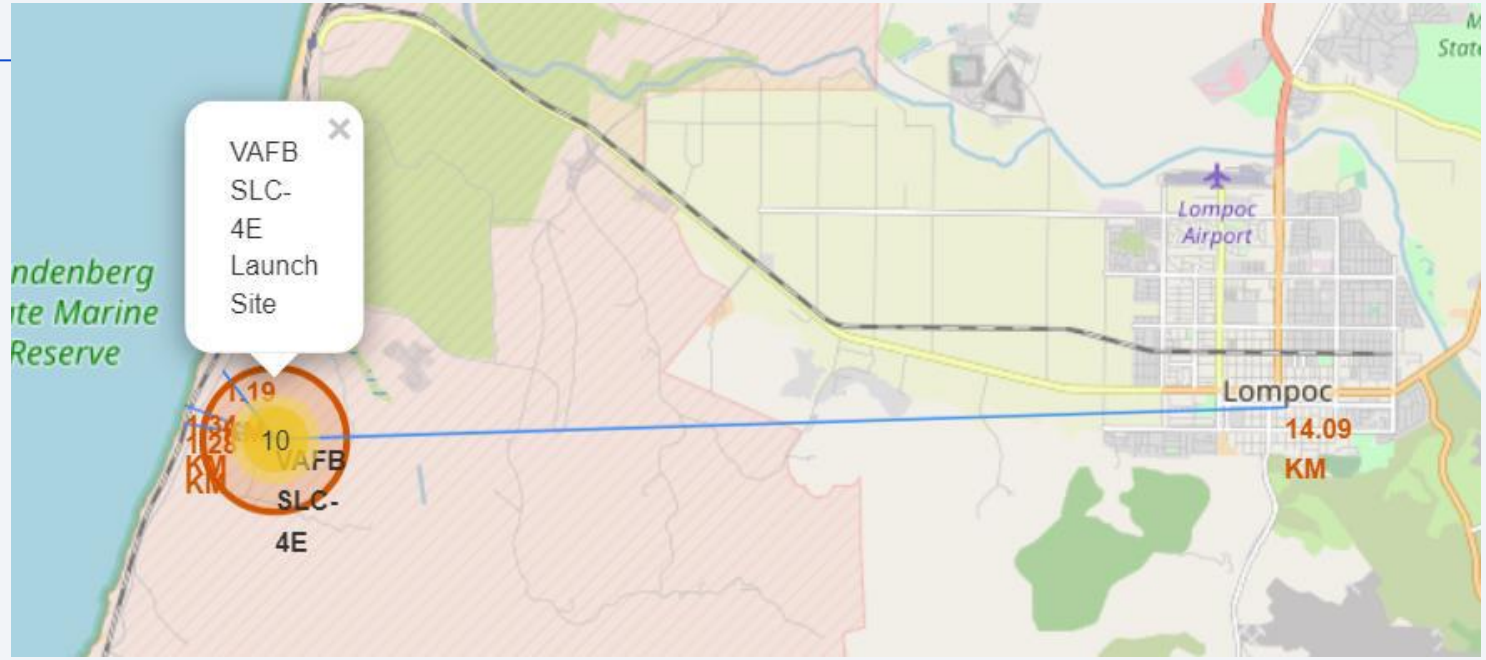


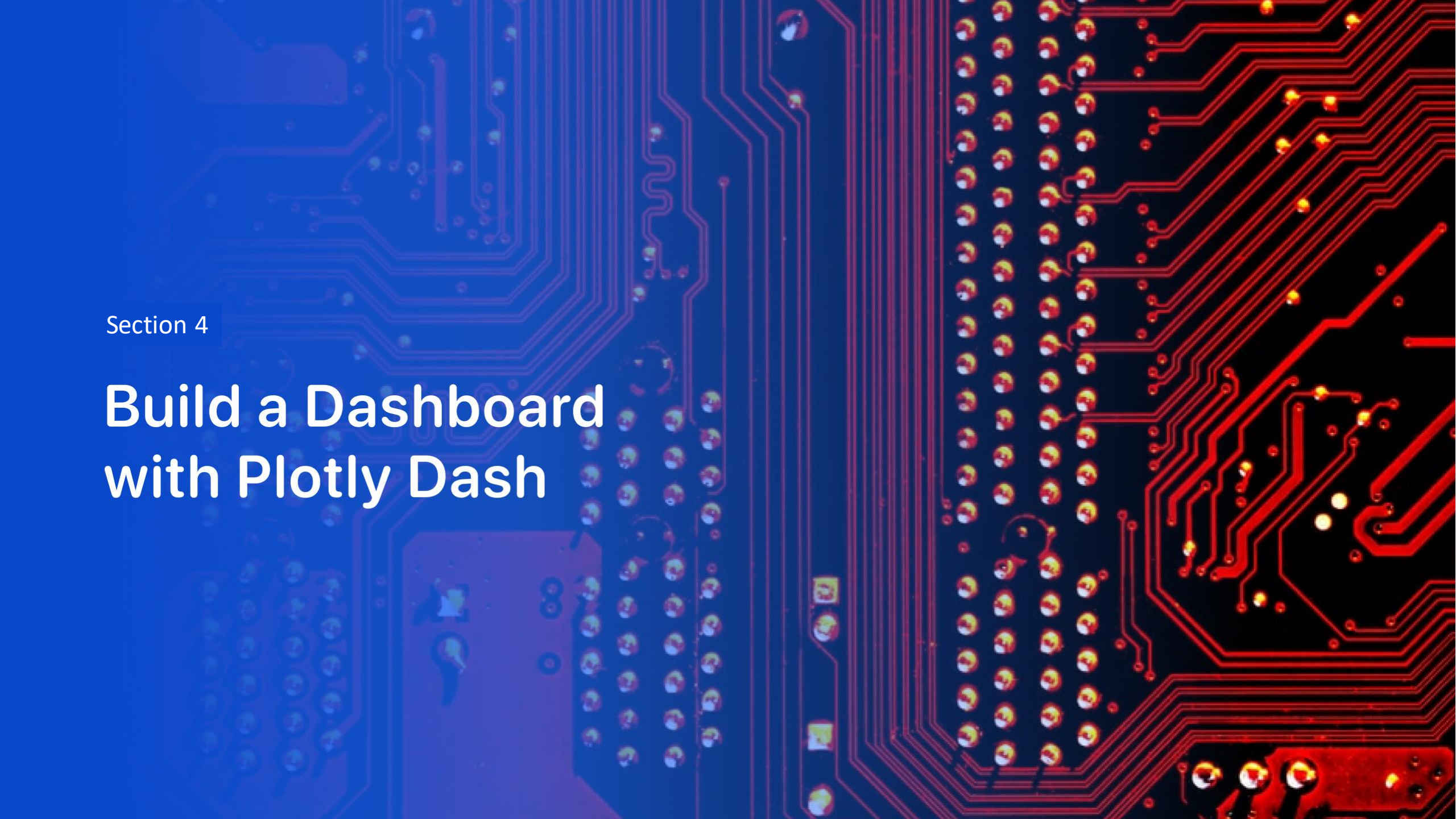




Proximity map of launch site

Detailing proximities from launch site

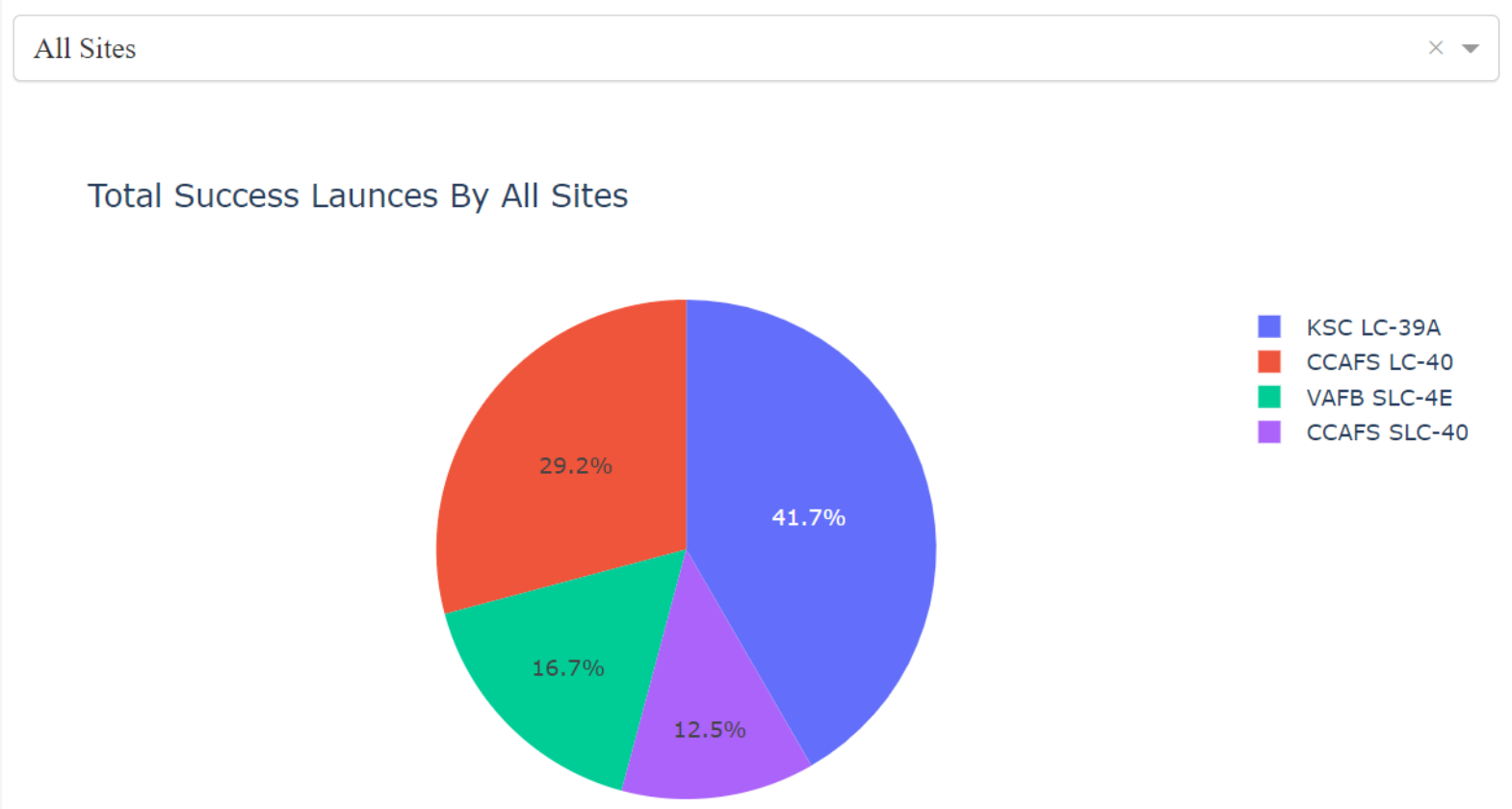



Section 4
Build a Dashboard with Plotly Dash

Launch site success rates

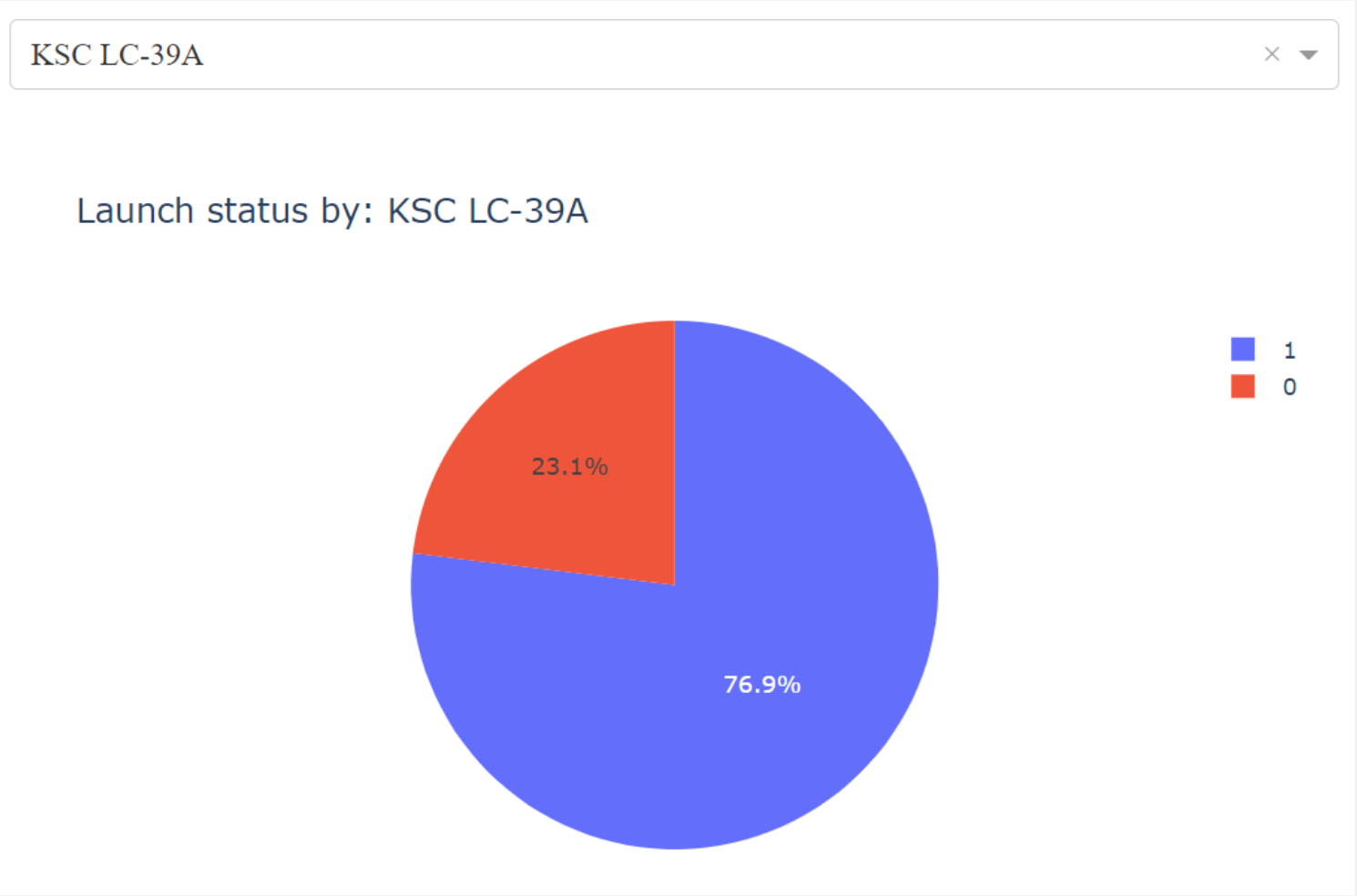
Success Rankings:

1. KSC LC-39A
2. CCAFS LC-40
3. VAFB SLC-4E
4. CCAFS SLC-40

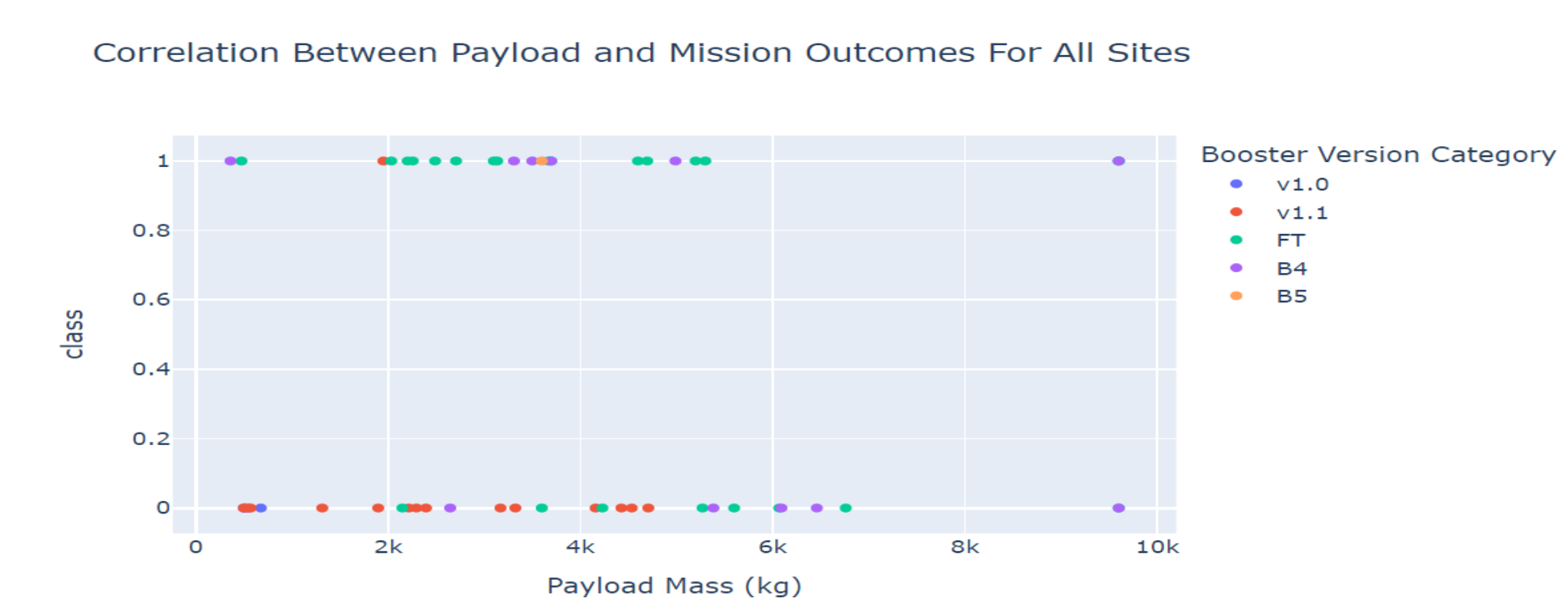


KSC LC-39A Success Ratio

- KSC LC-39A has
  - 76.9 % success rate
  - 23.1% failure rate



Scatter plots with Payload vs Launch outcomes



- Payload vs Launch outcomes
- More success in 2000 to 5500 range

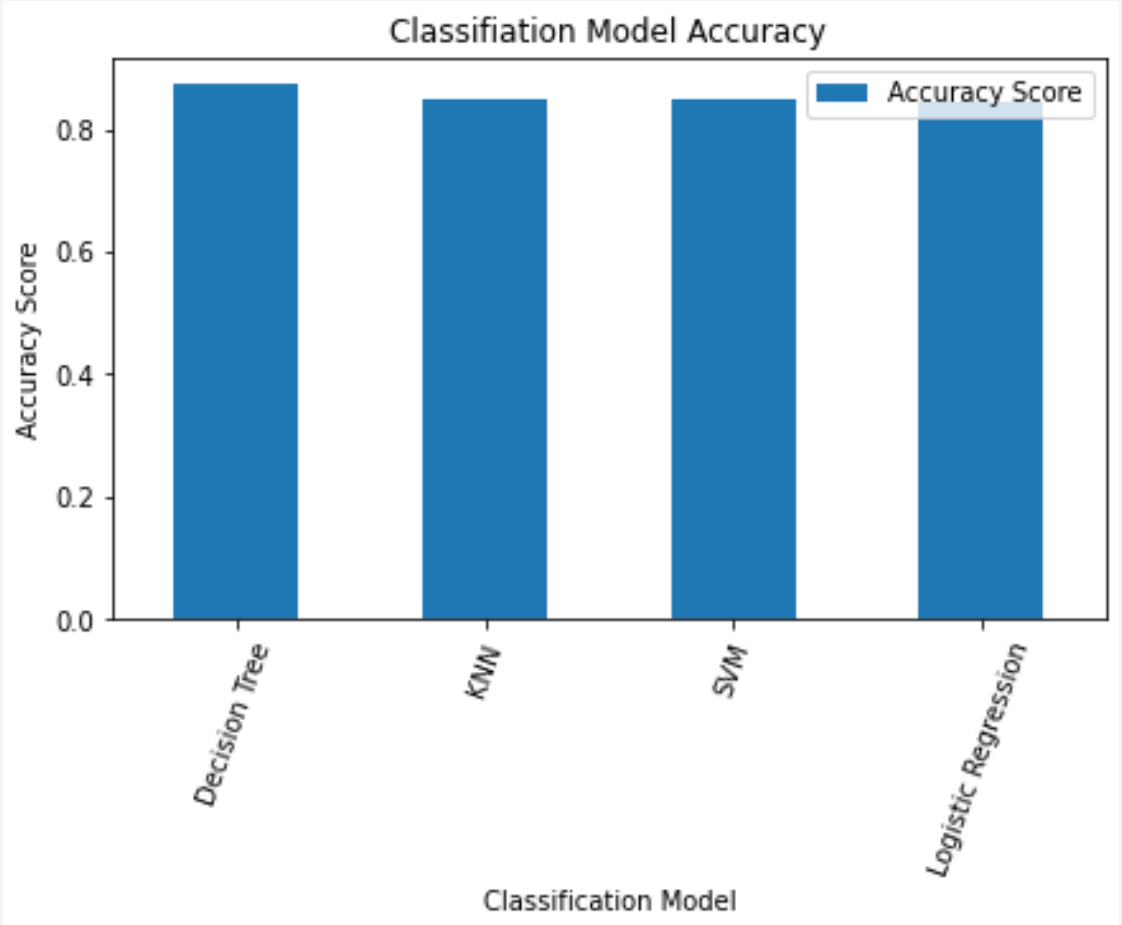
Section 5

Predictive Analysis (Classification)

Classification Accuracy

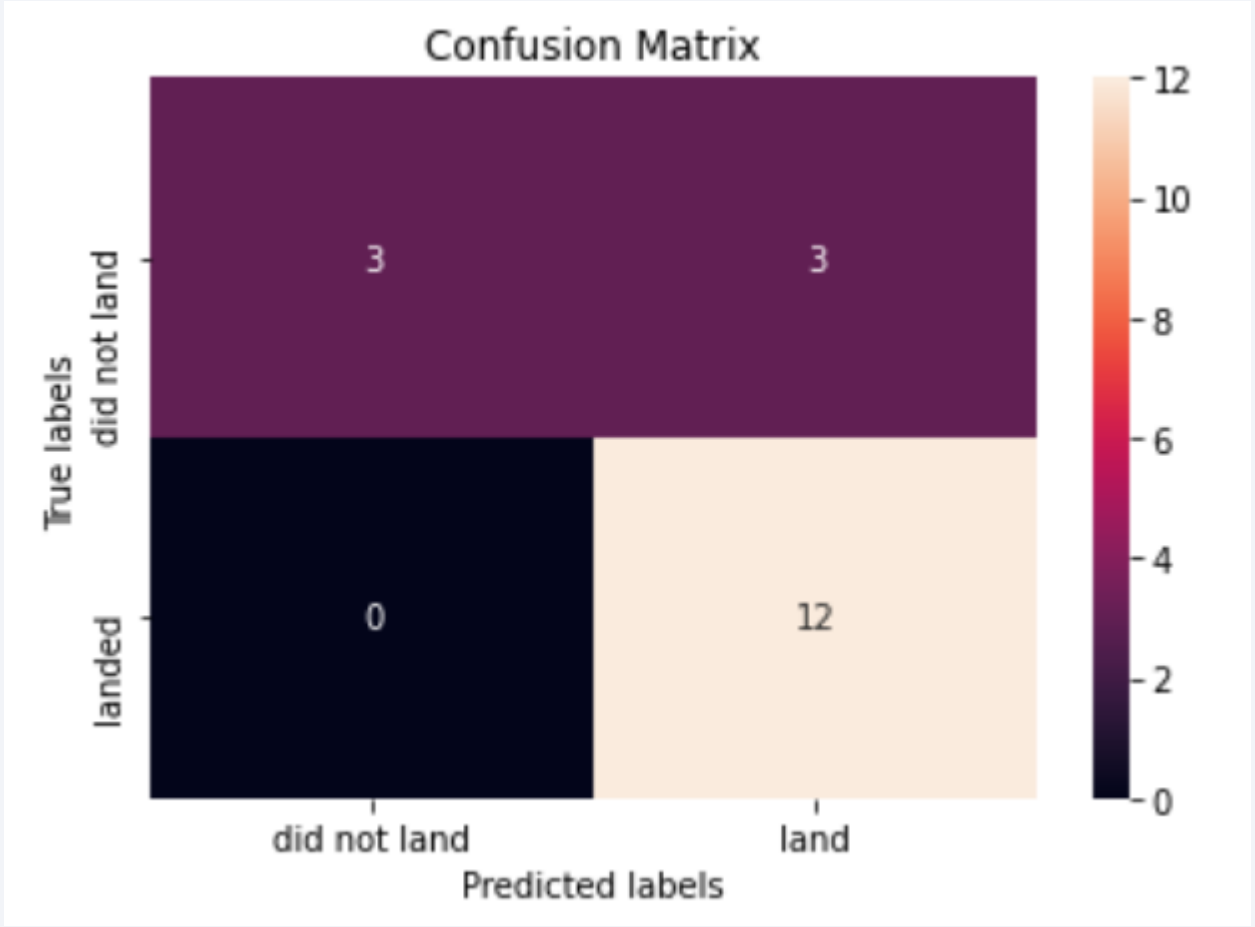
- Accuracy for Decision tree, KNN, SVM, and Logistic Regression

- Decision Tree has highest accuracy



Confusion Matrix

- Correct predictions are in top left and bottom right
- Model predicted 12 successful landings; 3 failed landings; 3 successful landings



Conclusions

- Flights increase, more likely to be sucessful
- Launch success increased from 2013 to 2020
- Model prediction has 83% accuracy

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!