IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Robert Filipp
October 27, 2021

Outline

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

Executive Summary

- Methodology
  - We collected SpaceX launch data using SpaceX API service and Wikipedia
  - Performed data wrangling
  - Performed exploratory data analysis (EDA) using visualization and SQL
  - Performed interactive visual analytics using Folium and Plotly Dash
  - Performed predictive analysis using classification models
- Summary of all results
  - All machine learning models provide the same level of prediction accuracy, at 78% success rate
  - Launch site CCAFS has the highest number of successful launches
  - Successful landing on the ground pad is the most frequent

Introduction

- Multiple space companies are providing flights to space nowadays
- One of them is SpaceX, offering lowest prices, due to their ability to reuse the first stages of the rockets for multiple launches
- As a competing space company, if we can use data science to predict whether the first stage will land or not, we can compete against SpaceX more successfully

Section 1

Methodology

Methodology

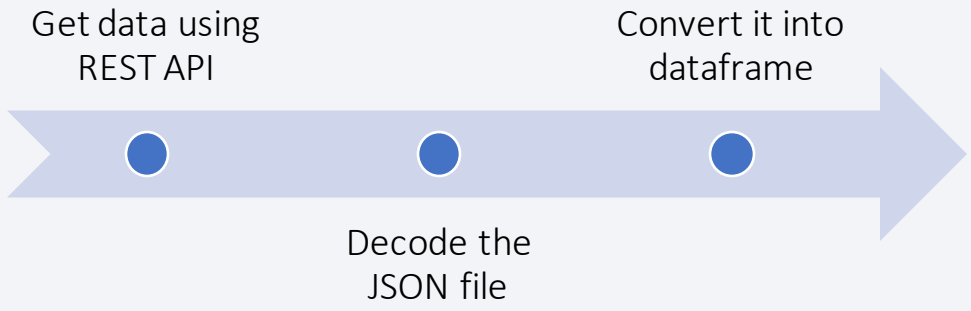
Executive Summary

- Data collection methodology:
  - We collected SpaceX launch data using SpaceX API service and Wikipedia
- Perform data wrangling
  - We processed the data so it can be later used for machine learning and further analysis
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - We built and evaluated various classification models

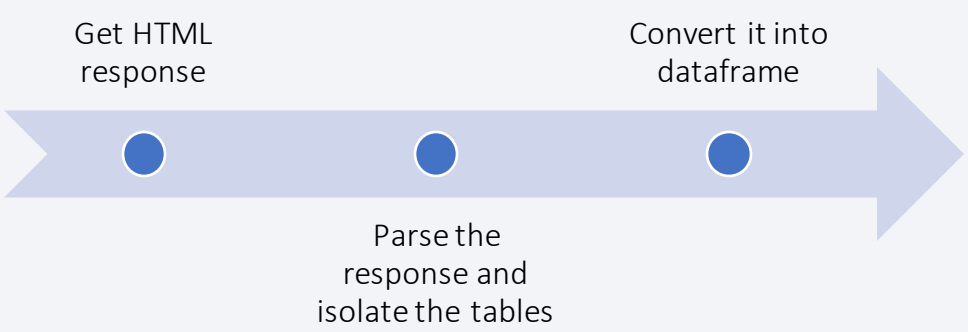
Data Collection

- We collected 2 types of data:
  - SpaceX launch data using the SpaceX API service
  - Additional historical launch data from Wikipedia
- SpaceX launch data was collected using their SpaceX REST API endpoints
- Additional data was collected from Wikipedia using method called scraping

Data Collection Process



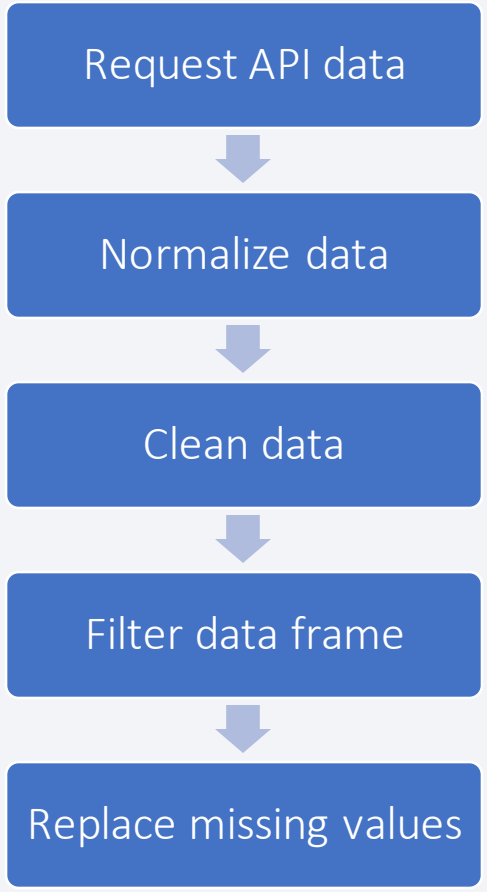
Data Scraping Process



Data Collection – SpaceX API

- Step 1: Call a requests.get function on the API of the url
- Step 2: Normalize data from JSON to data frame
- Step 3: Clean the data frame (select columns, remove unwanted rows, use custom functions)
- Step 4: Filter the data frame for needed variables
- Step 5: Replace the missing values with averages

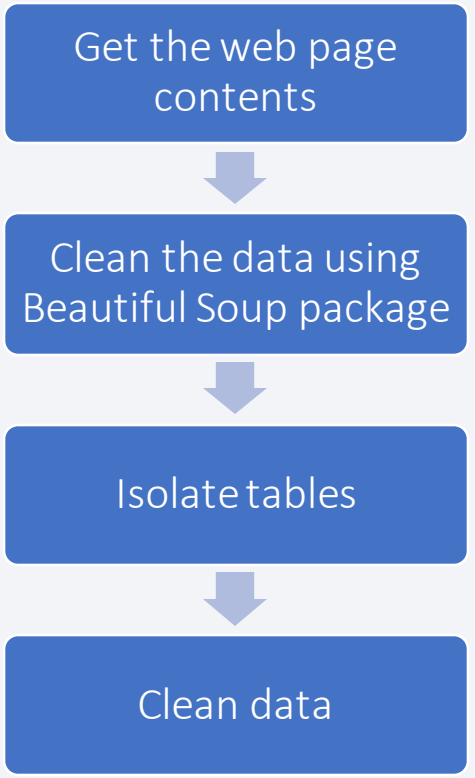
- Link to notebook on GitHub: Link



Data Collection - Scraping

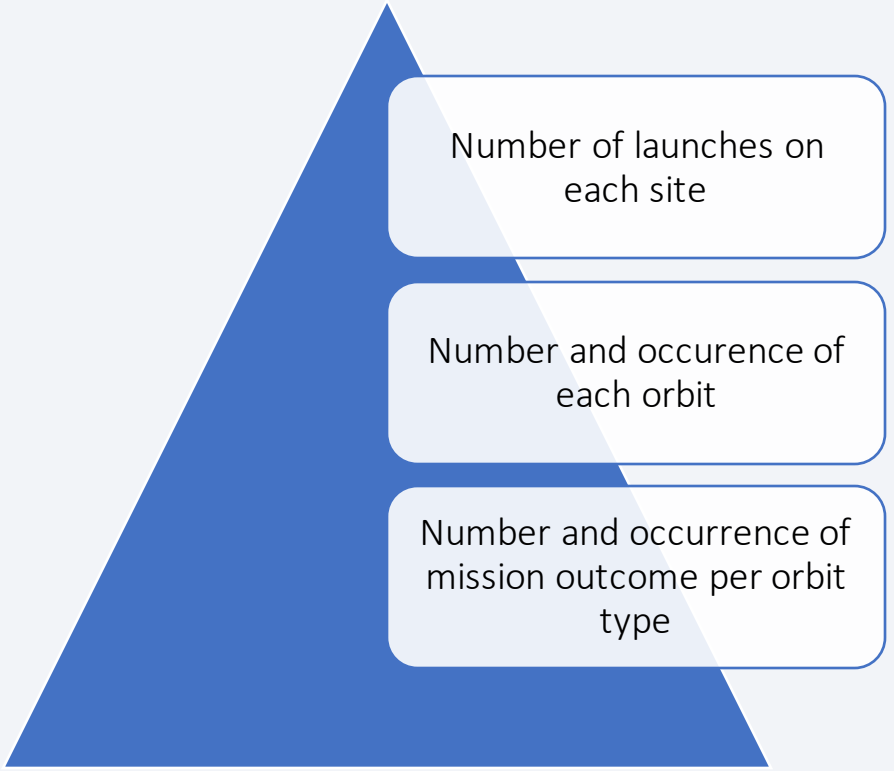
- Step 1: get the web page content using the requests.get method
- Step 2: Clean the response using Beautiful Soup package
- Step 3: Isolate the tables with required data
- Step 4: Clean the data and convert it to a data frame

- Link to notebook on Github: Link

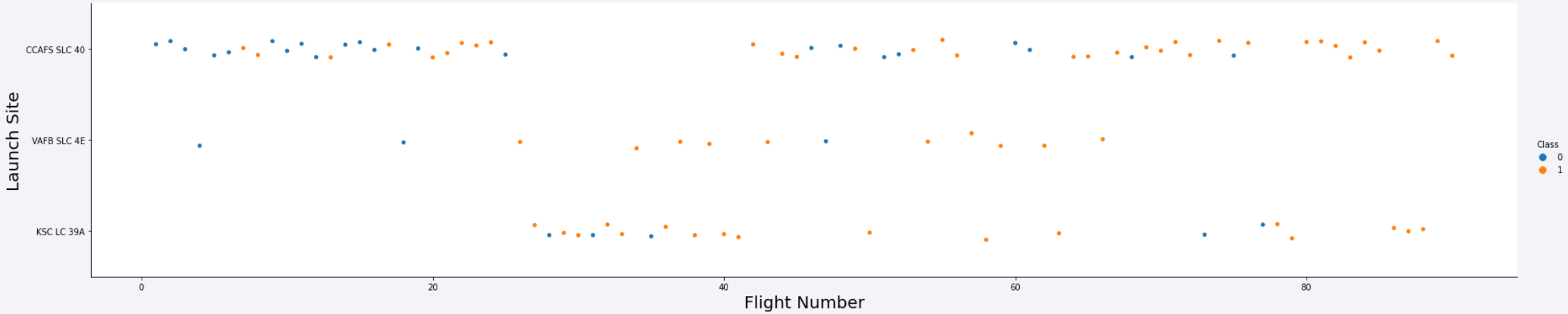


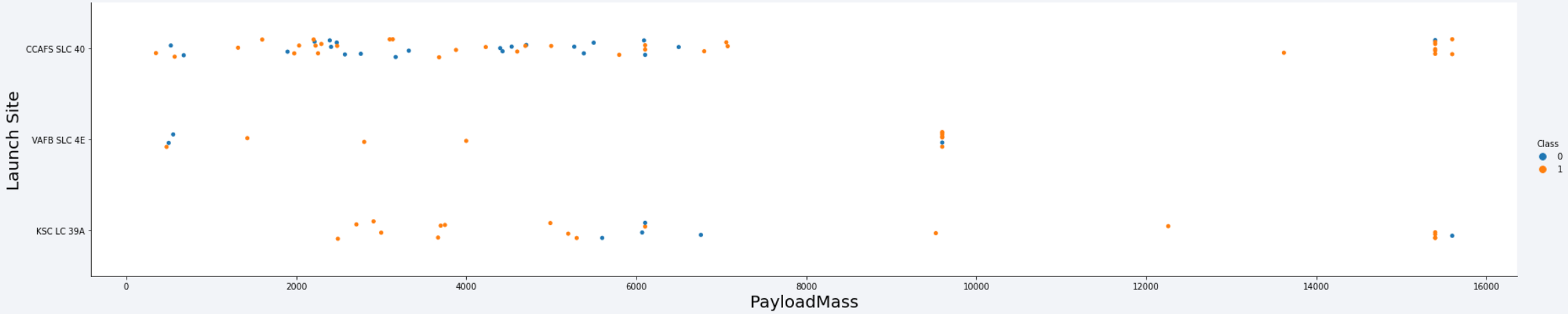
Data Wrangling

- Lean more about data:
  - Calculate the number of launches on each site
  - Calculate the number and occurrence of each orbit
  - Calculate the number and occurrence of mission outcome per orbit type
- Prepare the data frame for further analysis
  - Create a landing outcome label
  - Isolate the landing outcome and insert it into the data frame
- Link to notebook on Github: Link

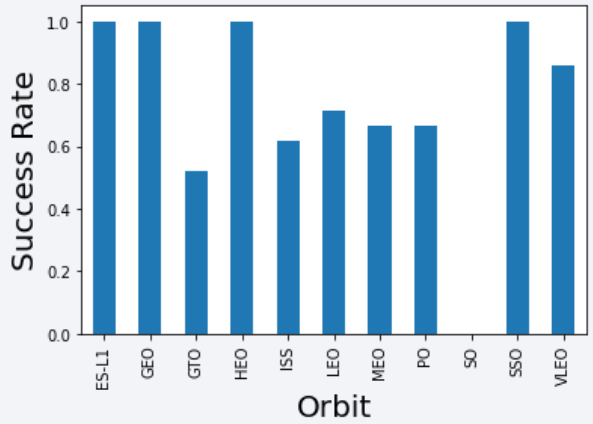


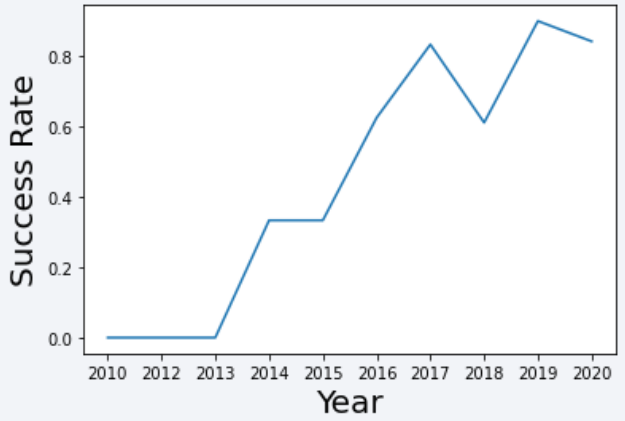
EDA with Data Visualization

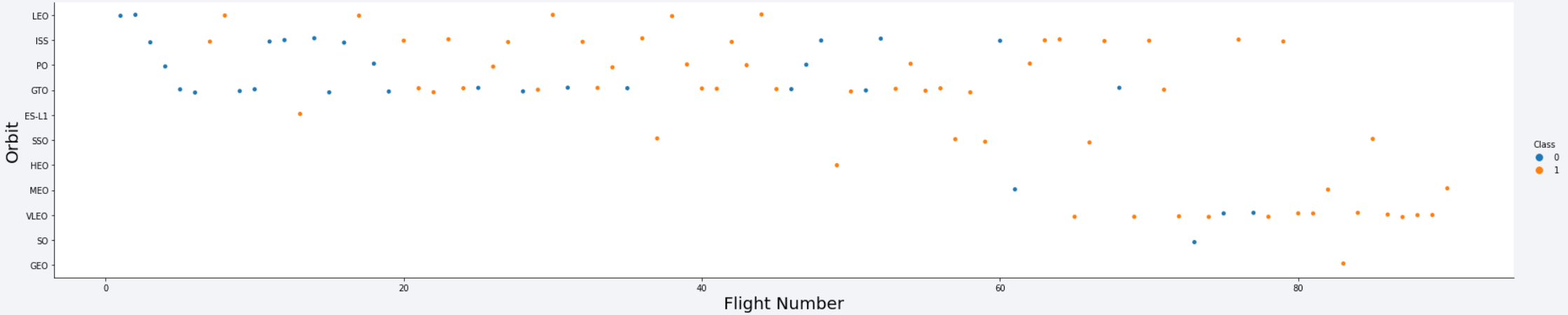




EDA with Data Visualization







Link to notebook on GitHub: [Link]

EDA with SQL

- Some SQL queries performed:
  - Display unique launch sites
  - Show launch sites that begin with 'CCA'
  - Display total payload mass carried by boosters launched by NASA
  - Display average payload mass carried by booster version F9 v1.1
  - List the first successful landing outcome in ground pad date
  - List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
  - List the total number of successful and failure mission outcomes
- Link to notebook on GitHub: Link

Build an Interactive Map with Folium

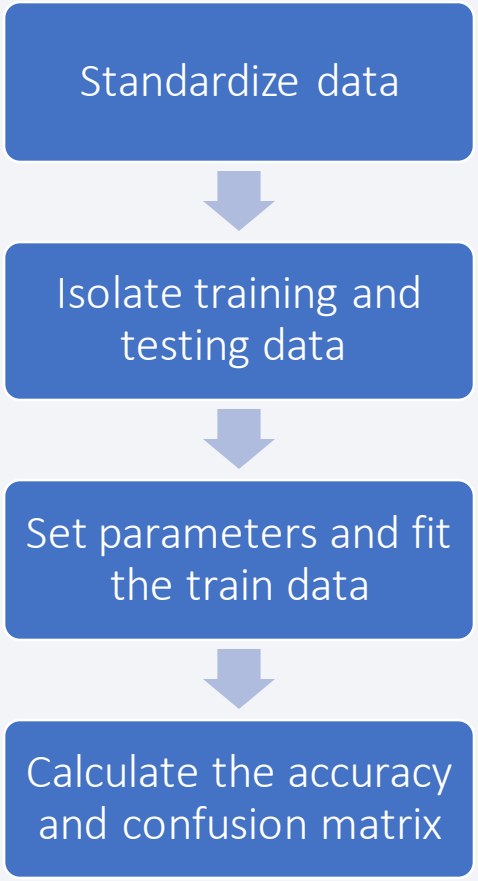
- To better visualize and understand the data, we created the following elements on interactive maps:
  - All launch sites as circles and labels
  - Successful and failed launches as markers
  - Distance from a launch site to nearby objects (beach, naval base) as circles and lines
- Link to notebook on GitHub: Link

Build a Dashboard with Plotly Dash

- To allow users to interact with data and find new insights, we created an interactive dashboard with the following plots and graphs:
  - Total Success Launches for All Sites and Individual Sites
  - Payload vs. Outcome for All Sites and Individual Sites
- Link to notebook on Github: Link

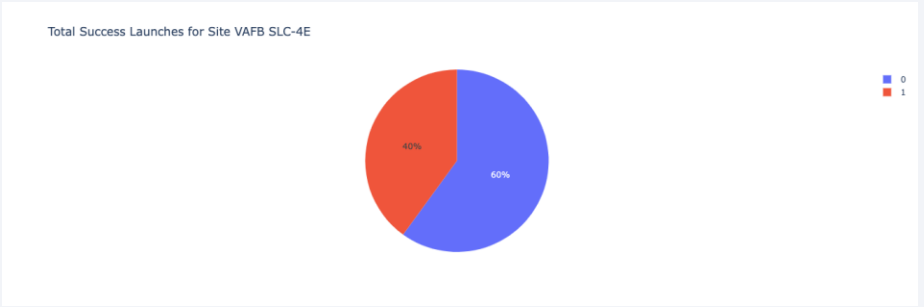
Predictive Analysis (Classification)

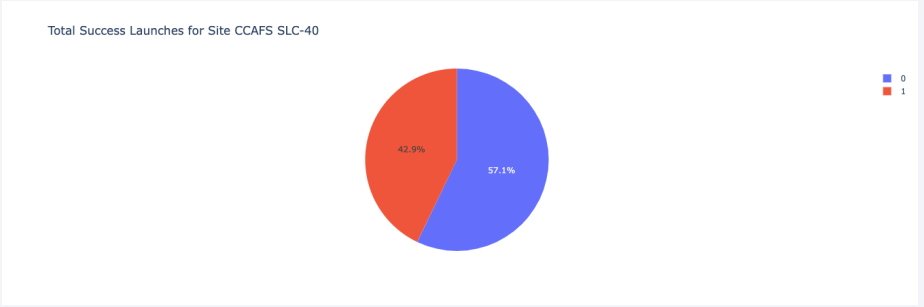
- In order to be able to predict wheather the launch and the land would be successful, we need to find a predictive model with highest percentage of success
- We explored multiple models:
  - Logistic regression
  - Support Vector Machine
  - Decision Tree Classifier
  - K Nearest Neighbors
- We found that based on our data, all models were performing equally well
- Link to notebook on GitHub: Link

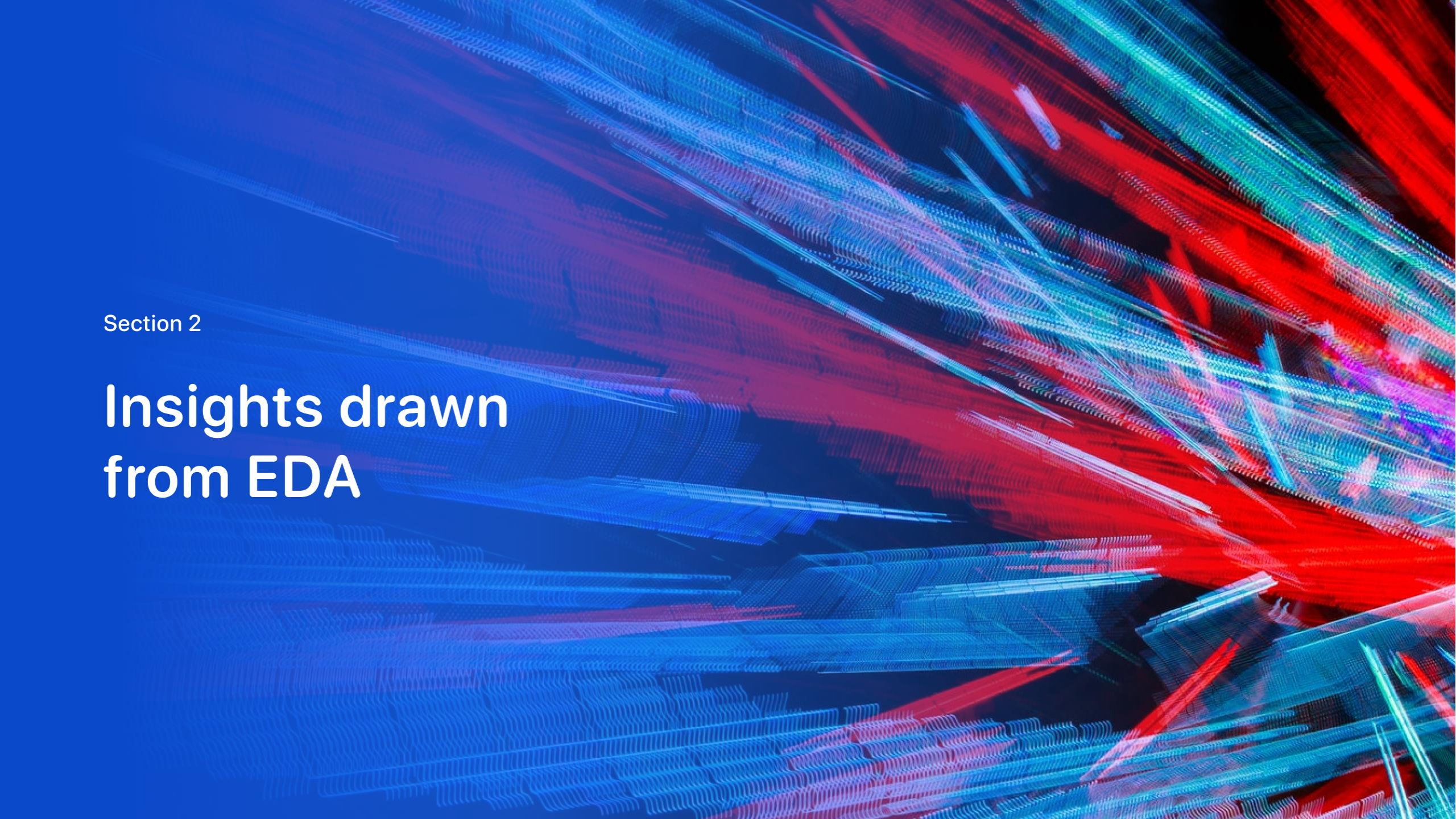


Results

- Exploratory data analysis results
  - Site CCAFS had the most launches
  - GTO is the most frequent orbit
  - Successful landing on the ground pad is the most frequent
- Interactive analytics demo in screenshots
- Predictive analysis results
  - All models have equal success rates
  - We achieved 78% success rate in predicting outcomes



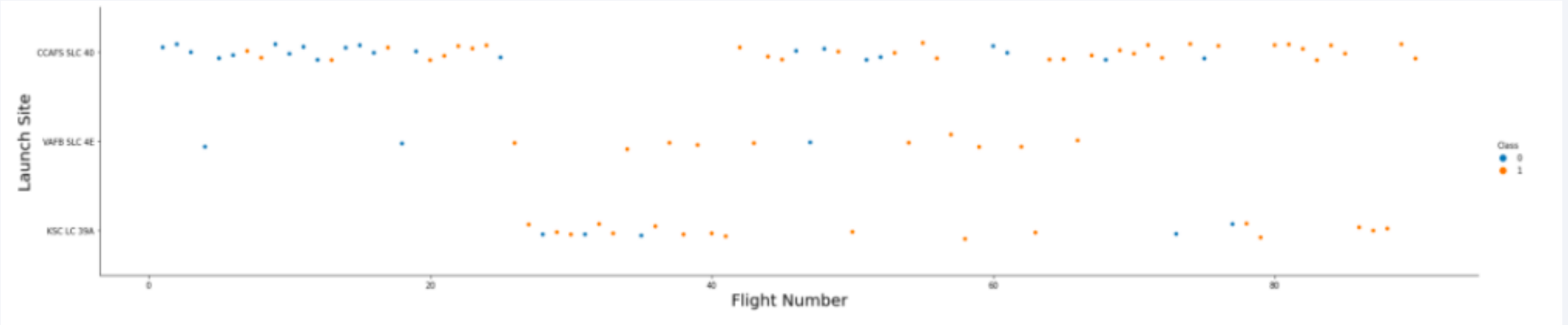




Section 2

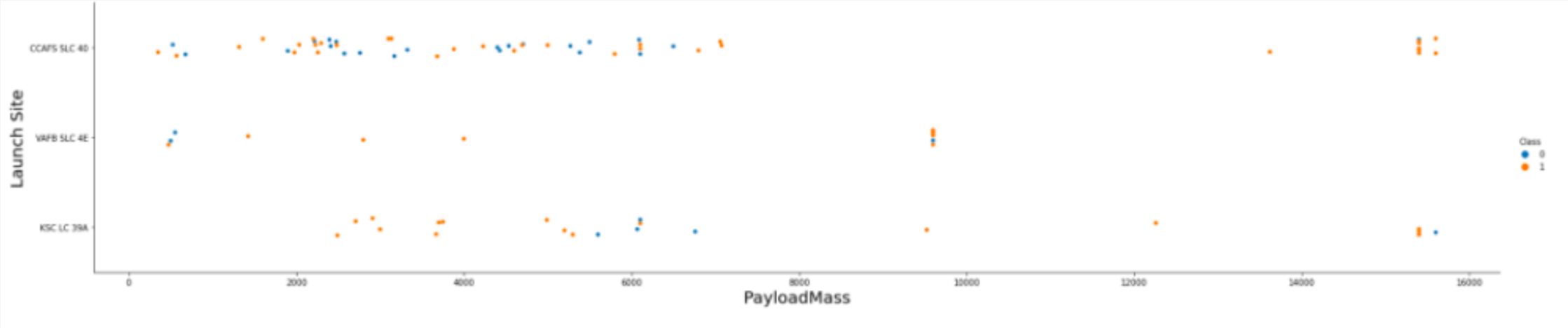
Insights drawn from EDA

Flight Number vs. Launch Site



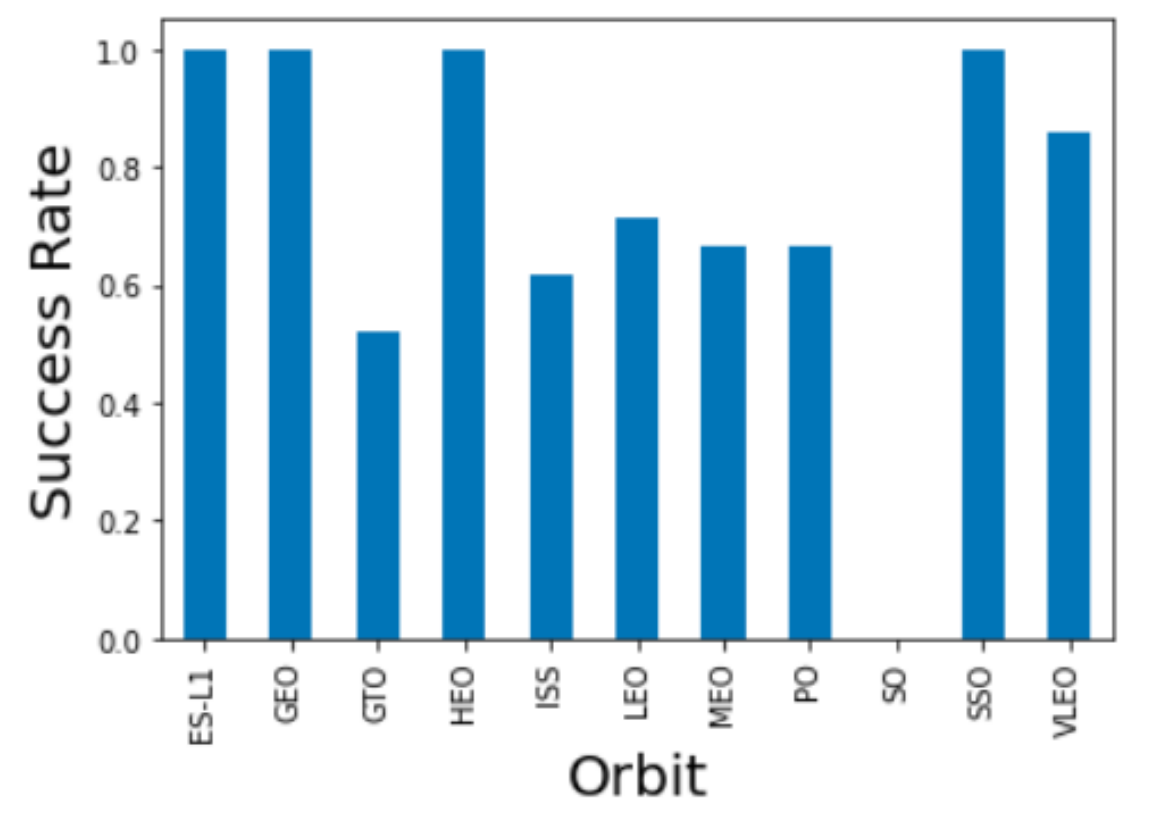
- The highest number of flights were launched from CCAFS
- Almost all most recent launches from CCAFS were successful
- None of the last 25 flights were launched from VAFB

Payload vs. Launch Site



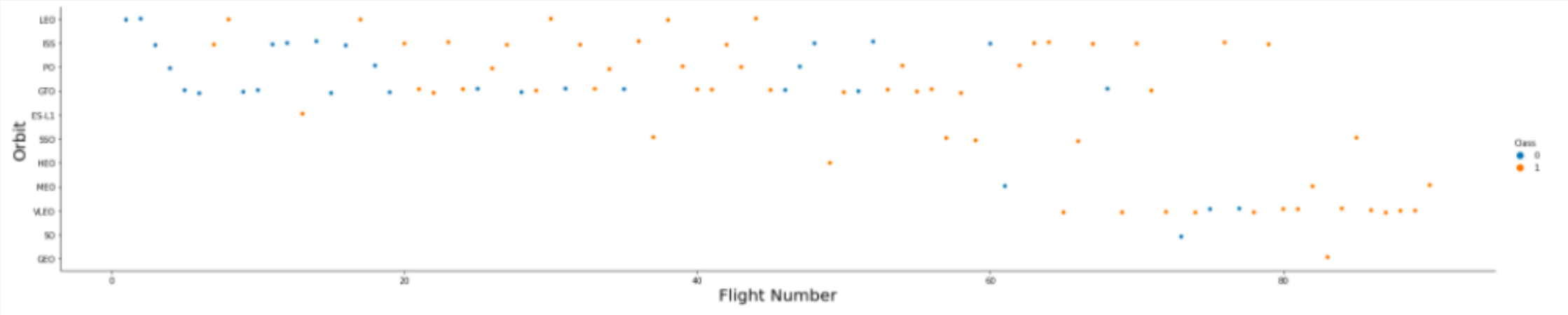
- The highest number of payloads of less than 8,000 lb was launched from CCAFS
- Almost all flights with payloads of more than 8,000 lb were successful
- Almost all flights with payloads of less than 8,000 lb from KSC were successful

Success Rate vs. Orbit Type



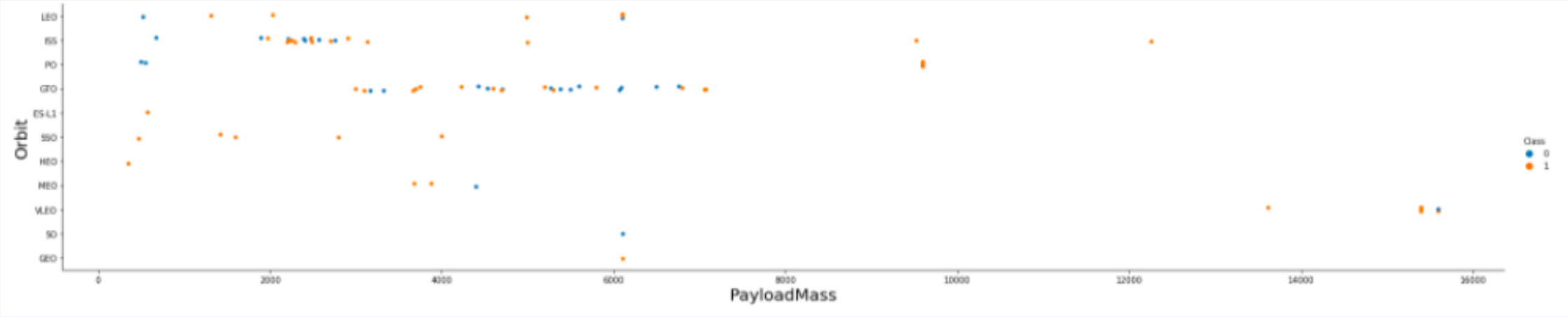
- 100% of ES-L1, GEO, HEO, and SSO were successful
- Flights to GTO were least successful (50%)
- Flights to SO were all unsuccessful

Flight Number vs. Orbit Type



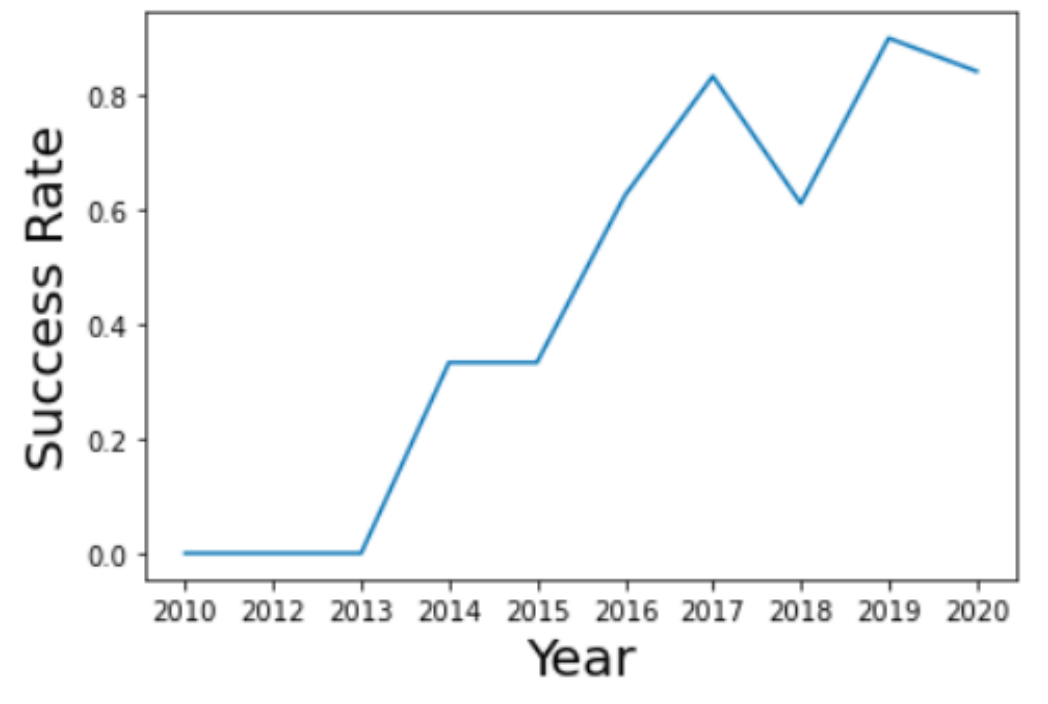
- Flights to VLEO is a rather new type of missions
- Flights to ISS have happened throughout the time period
- Whle there were only a few flights to SSO, all were successful

Payload vs. Orbit Type



- Flights to GTO were all under 8,000 lbs
- Flights to VLEO were all above 12,000 lbs
- Flights with payload of more than 8,000 lbs were almost all successful

Launch Success Yearly Trend



- Since 2016, annual success rate has been always more than 50%
- The success rate has been improving steadily over the last 11 years
- The success rate in last 4 years has been almost 100%

All Launch Site Names

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

- These are the names of all launch sites from the data used

Launch Site Names Begin with 'CCA'

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	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	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

- These are the first 5 launch records which launch site names begin with 'CCA'

Total Payload Mass

1
45596

- The total payload carried by booster from NASA has been more than 45,000 lbs

Average Payload Mass by F9 v1.1

1
2928

- The average payload mass carried by booster version F9 v1.1 was almost 3,000 lbs

First Successful Ground Landing Date

1
2015-12-22

- The first successful landing outcome on ground pad happened on 12/22/2015

Successful Drone Ship Landing with Payload between 4000 and 6000

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

- These are the names of boosters which have successfully landed on a drone ship and had payload mass greater than 4,000 but less than 6,000 lbs

Total Number of Successful and Failure Mission Outcomes

mission_outcome	2
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

- There has been a total of 99 successful mission outcomes

Boosters Carried Maximum Payload

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

- These are the names of the boosters which carried the maximum payload mass

2015 Launch Records

DATE	time__utc_	booster_version	launch_site	payload	payload_mass__kg_	orbit	customer	mission_outcome	landing__outcome
2015-01-10	09:47:00	F9 v1.1 B1012	CCAFS LC-40	SpaceX CRS-5	2395	LEO (ISS)	NASA (CRS)	Success	Failure (drone ship)
2015-04-14	20:10:00	F9 v1.1 B1015	CCAFS LC-40	SpaceX CRS-6	1898	LEO (ISS)	NASA (CRS)	Success	Failure (drone ship)

- These are the failed landings on a drone ship in 2015

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

landing__outcome	2
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

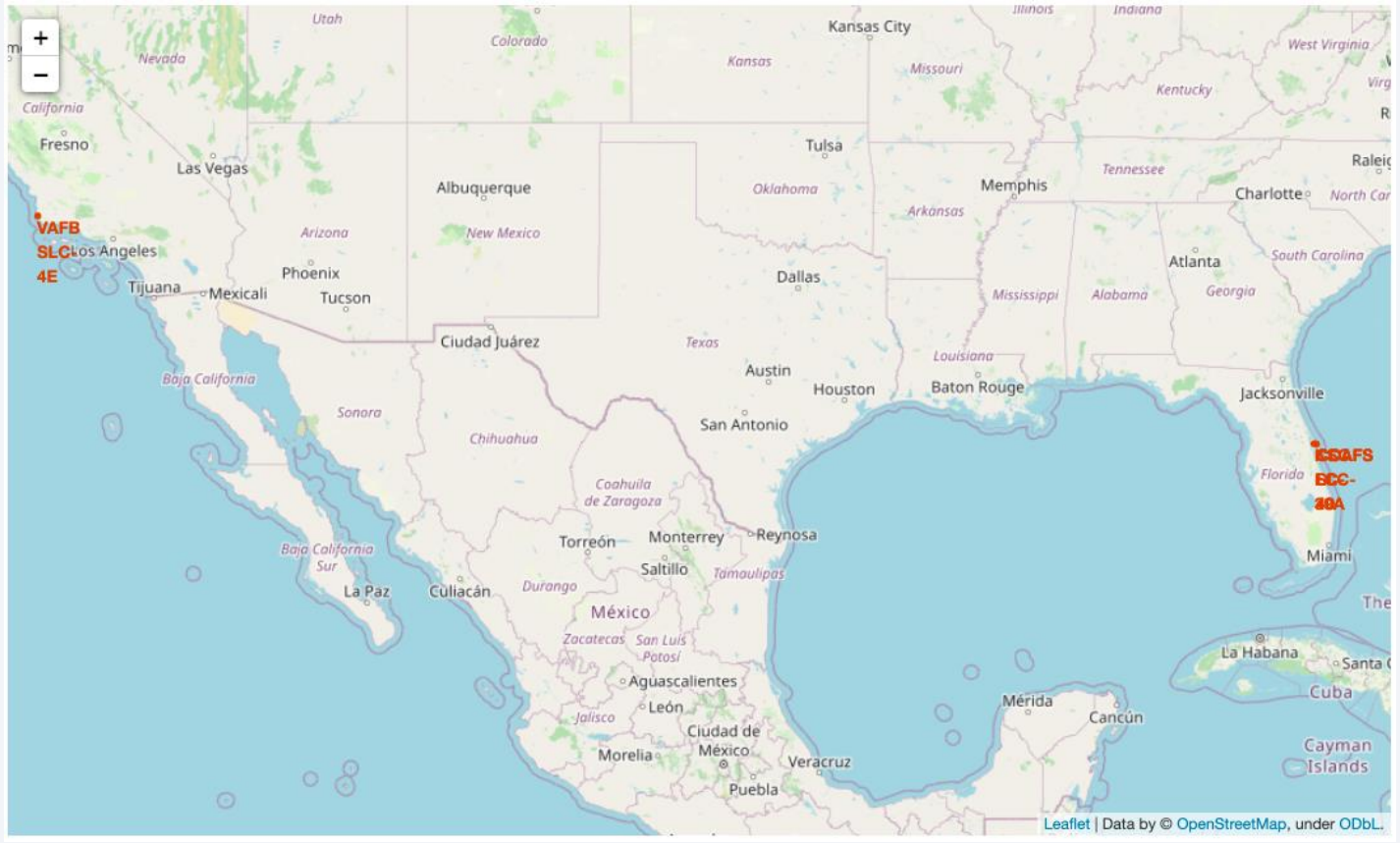
- This is the list of successful and failed landing outcomes between 4/6/2010 and 20/3/2017
- The most frequent outcome is that there has not been even an attempt

Section 4

Launch Sites Proximities Analysis

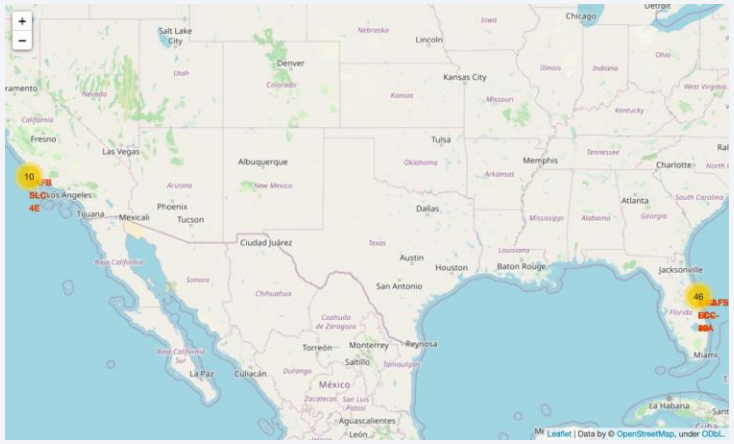
Map: All Launch Site Locations

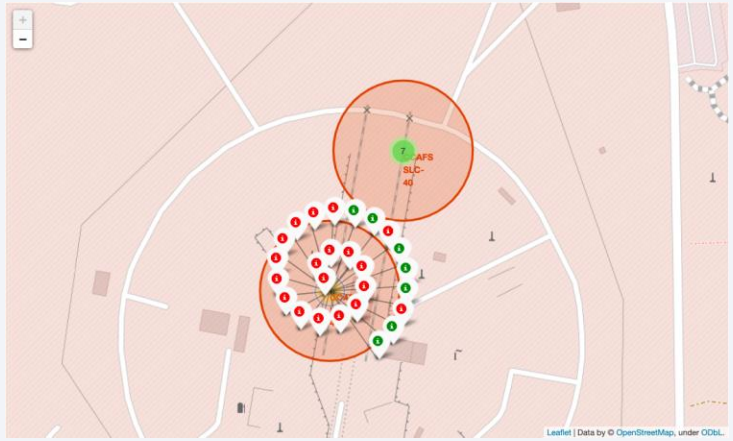
- The screenshot shows all the launch sites in the US
- One site is on the West Coast in California, and multiple ones are on the East Coast in Florida



Map: Mission Outcomes by Location

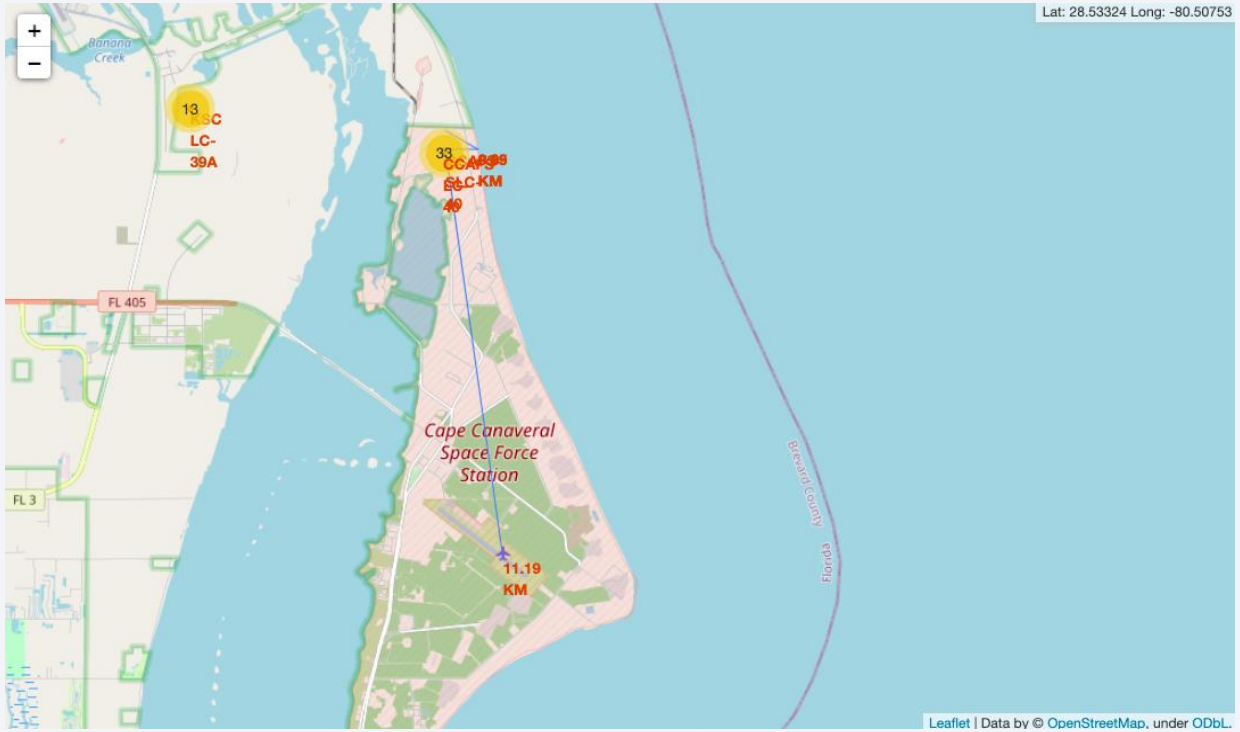
- These maps show the mission outcomes (color coded) by location
- With proper interactive zoom, you are able to see the details of each launch, and whether the landing was successful or not

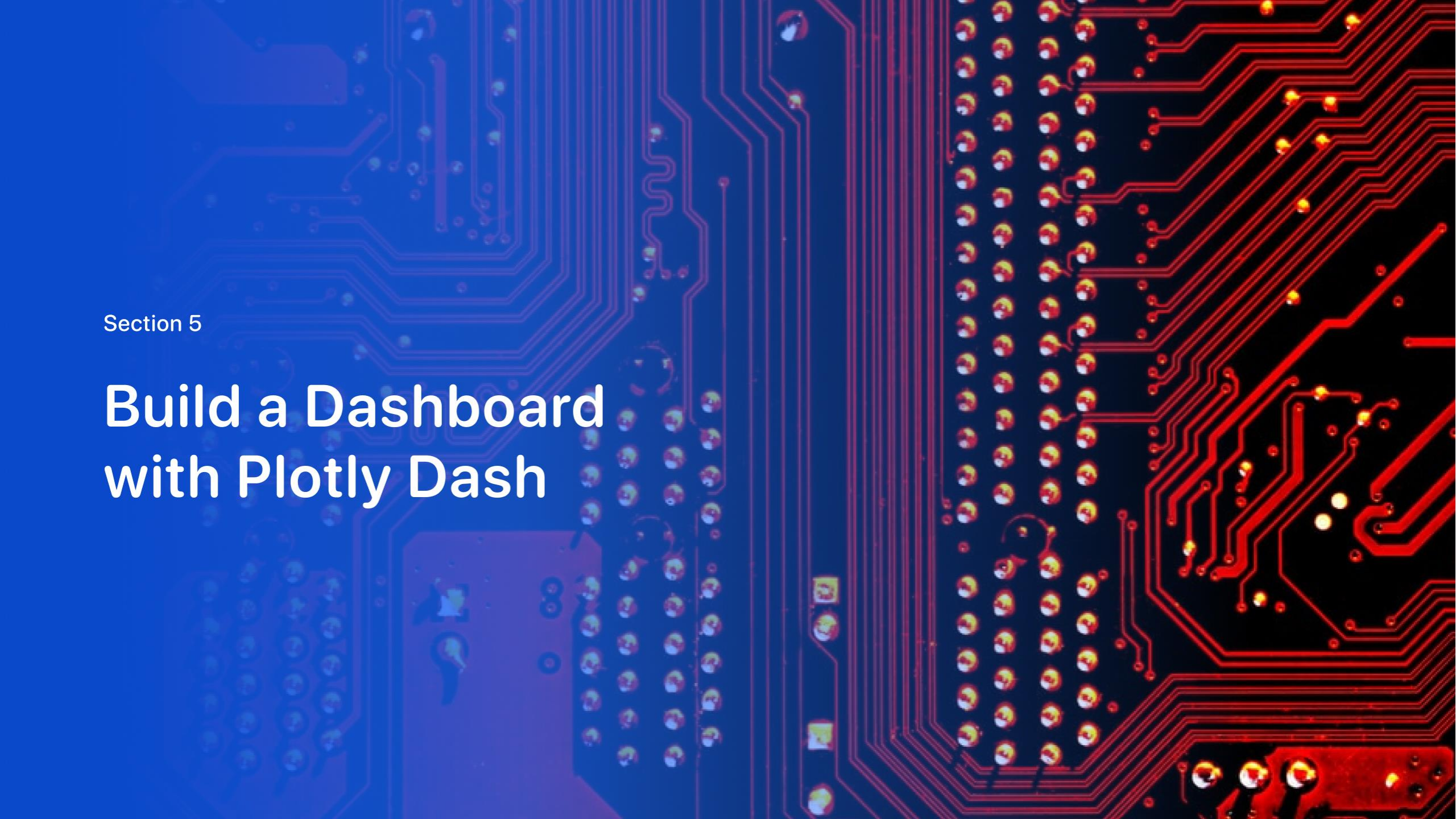




Map: Launch Location Proximity to Landmarks

- This map shows one of the launch sites on East Coast, and its distance from the coast, and the naval base

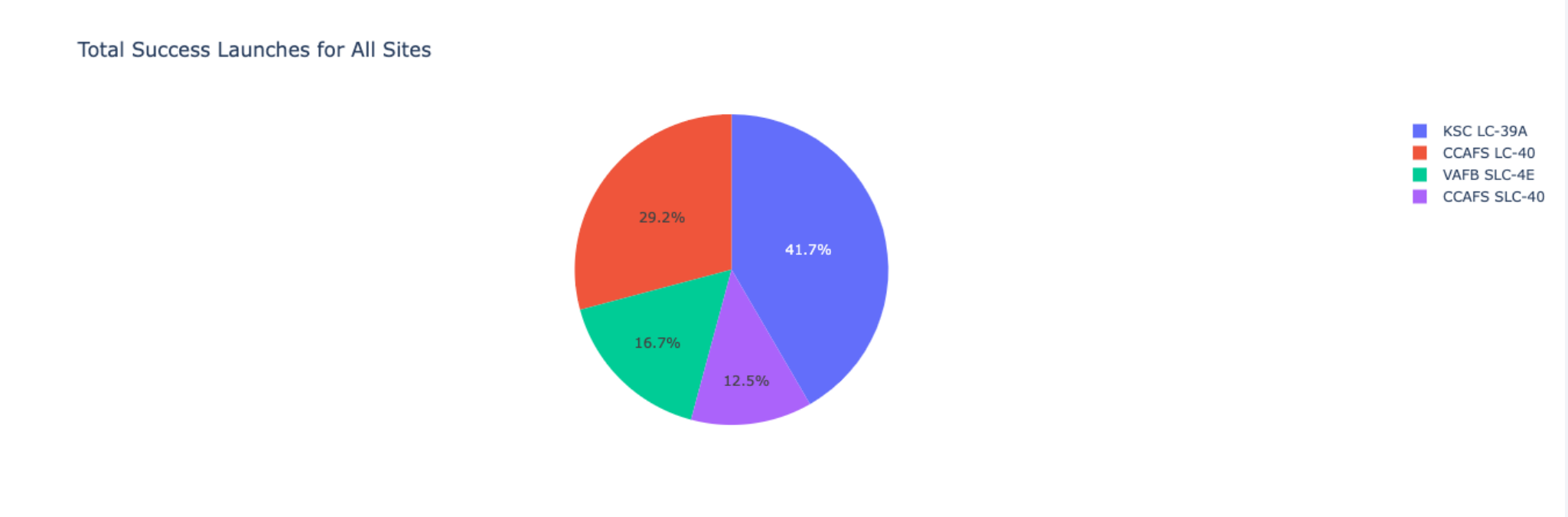




Section 5

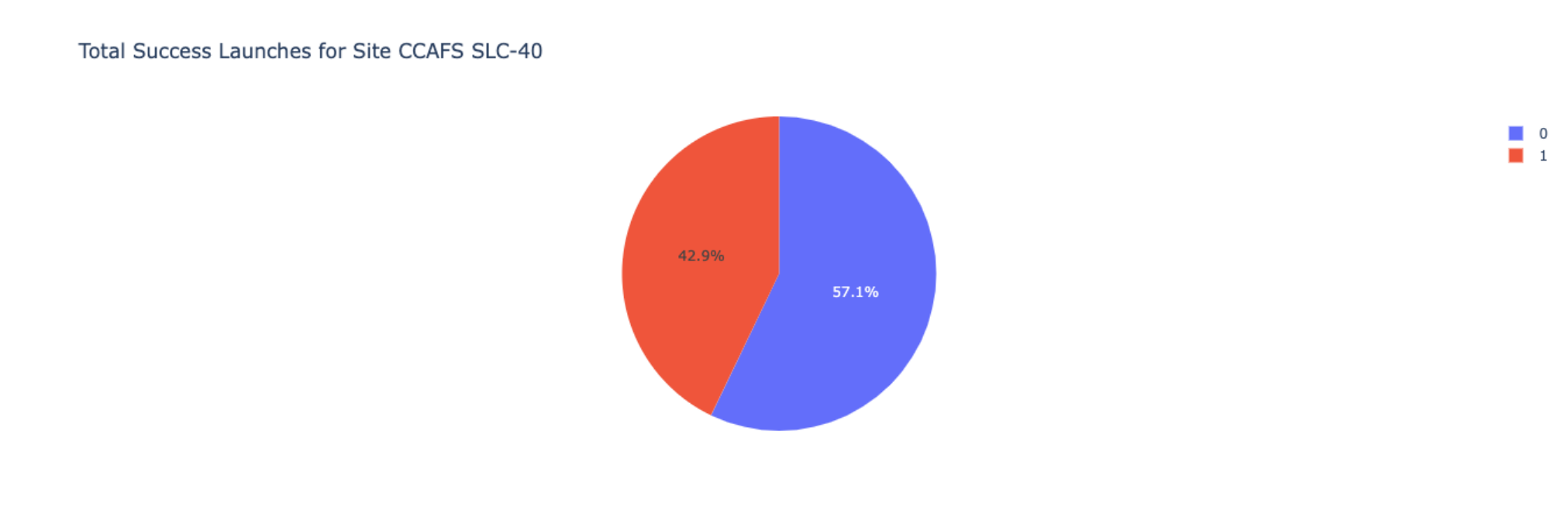
Build a Dashboard with Plotly Dash

Dashboard: Launch Success Count for All Sites



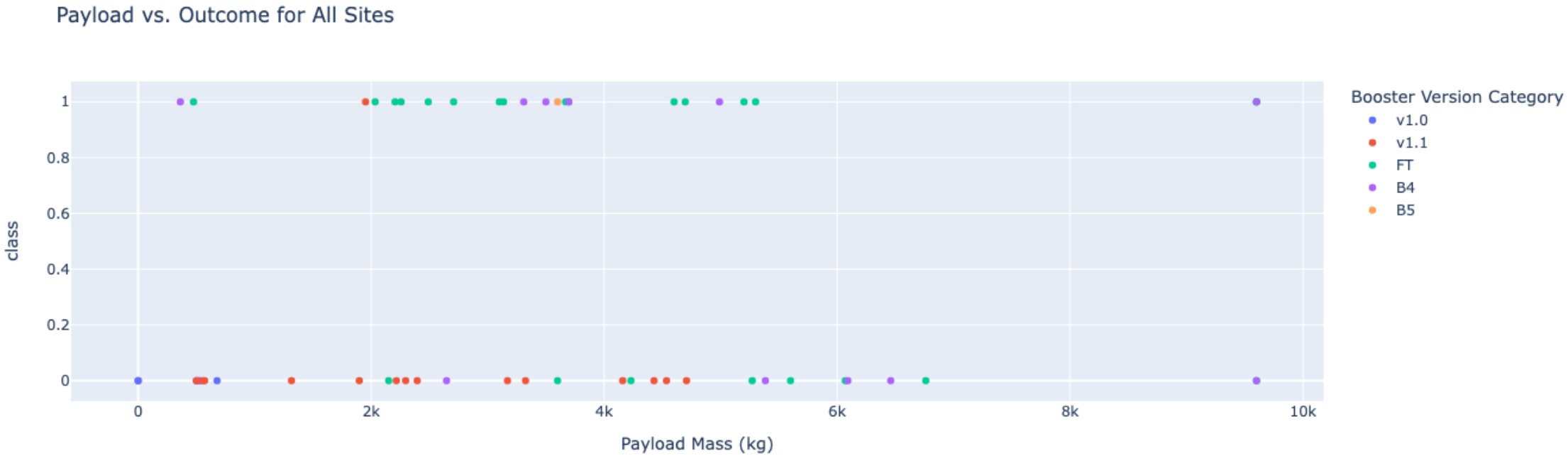
- This interactive dashboard pie chart shows the count or successful landings by location
- The highest number of successful landing has CCAFS launch site

Dashboard: CCAFS success rates



- This interactive pie chart shows the success rate of landing for the missions launched from CCAFS launch site
- This pie chart shows that 43% of launches landed successfully

Dashboard: Payload vs. Launch Outcome for All Sites



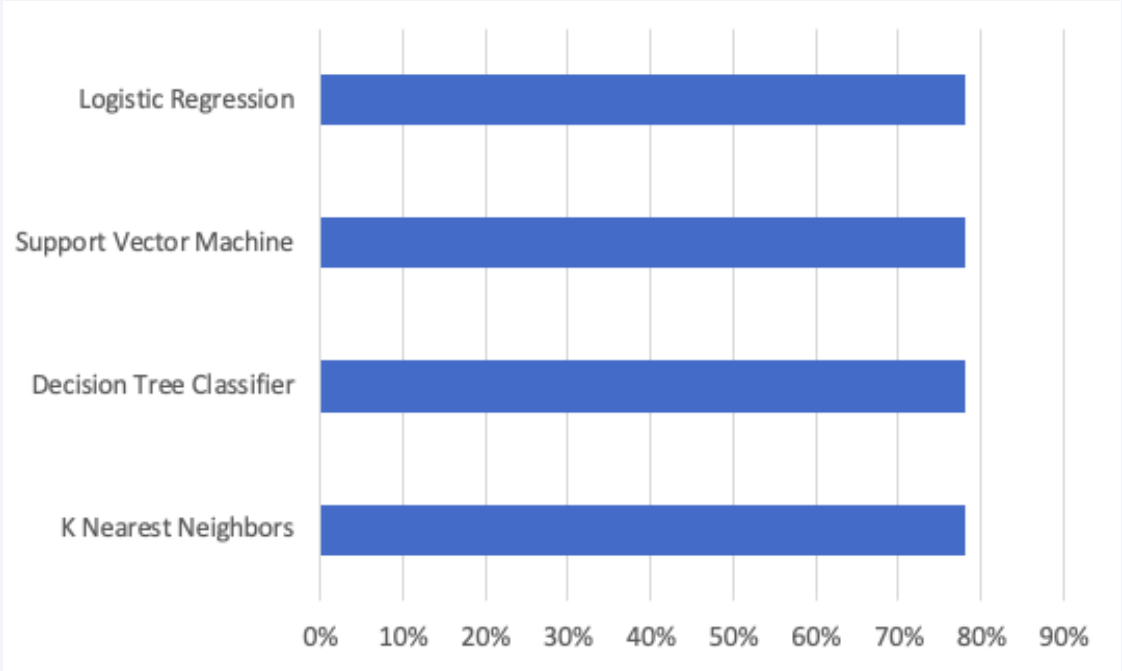
- This scatter plot shows the payloads, launch outcomes and booster version for all sites
- This chart shows that launches with payloads of 6,000 lbs or more were unsuccessful

Section 6

Predictive Analysis (Classification)

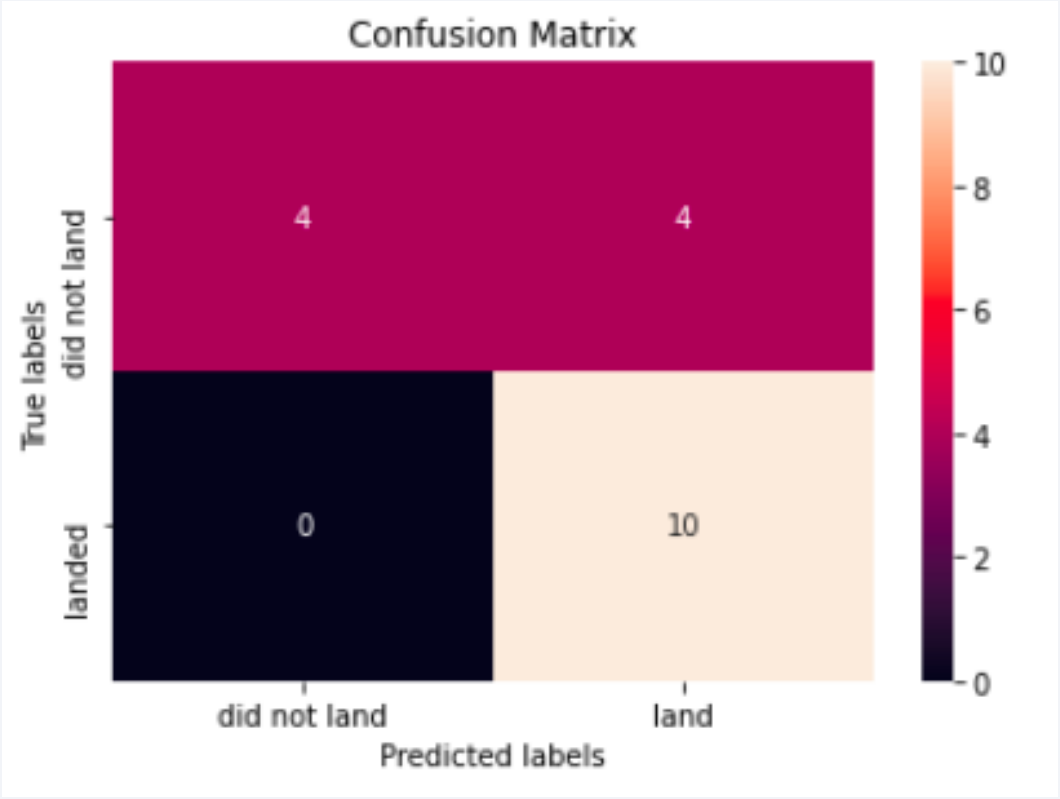
Classification Accuracy

- We explored multiple models:
  - Logistic regression
  - Support Vector Machine
  - Decision Tree Classifier
  - K Nearest Neighbors
- We found that based on our data, all models were performing equally well, at 78% accuracy



Confusion Matrix

- The following confusion matrix explains the prediction outcomes:
  - 10 correctly predicted successful landings
  - 4 correctly predicted failed landings
  - 4 incorrectly predicted successful landings



Conclusions

- All machine learning models provide the same level of prediction accuracy, at 78% success rate
- Launch site CCAFS has the highest number of successful launches
- Successful landing on the ground pad is the most frequent

Thank you!