IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Gabe Gagliardo
June 12, 2025

Outline

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

Executive Summary

- The goal of this project is to predict, using machine learning, if the SpaceX Falcon 9 first stage will land successfully.
- The steps of this evaluation include:
    - Data collection, wrangling, and formatting
    - Exploratory data analysis
    - Interactive data visualization
    - Machine learning prediction
- Our graphs indicate that some features of the launches have a correlation with the outcome of the launches
- It is concluded that decision tree may be the best machine learning algorithm to predict successful landing of the Falcon 9 first stage.

Introduction

- The purpose of this task is to predict if the Falcon 9 first stage will land successfully. SpaceX states the cost of $62 M for a Falcon 9 rocket launch.
  - Competitors can cost up to $165 M per launch.,
- The bulk of the savings is attributed to reuse the first stage. If it can determined the first stage will land, we can determine the cost of a launch. If the successful landing of the first stage, after use, can be predicted, we can predict a cost of the mission. This information can be used for budget planning and bidding on future contracts.
- The question to answer…..
  - Do features of a rocket launch (Payload mass, Orbit, etc) correlate to successful landing of the first stage?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - The SpaceX API was used to collect data for this effort
- Perform data wrangling
  - Web scraping was used to wrangle the 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
  - Logistic regression
  - Support vector machine (SVM)
  - Decision tree
  - K-nearest neighbors (KNN)

Data Collection

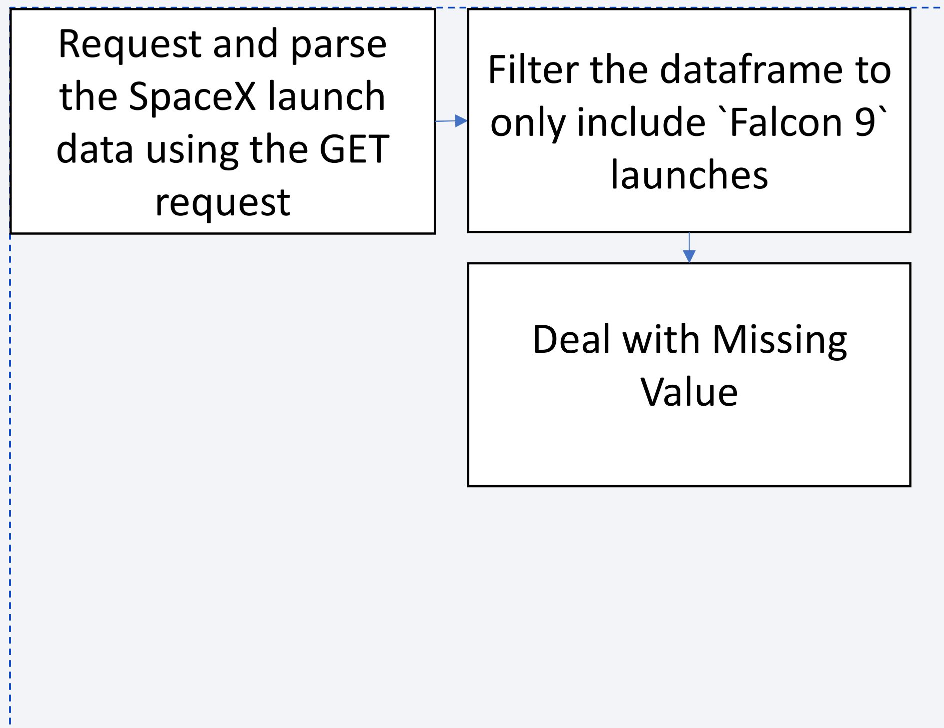
- SpaceX API
  - API URL = https://api.spacexdata.com/v4/launches/past
  - The API provides SpaceX launch data. It will be filtered for Falcon 9 launches.
  - Missing values in the data set are replaced with the mean in the relevant column
  - The data set is comprised of 90 rows or instances and 17 columns/features.

Filter the dataframe to only include Falcon 9 launches
Filter the dataframe to only include Falcon 9 launches

Data Collection – SpaceX API

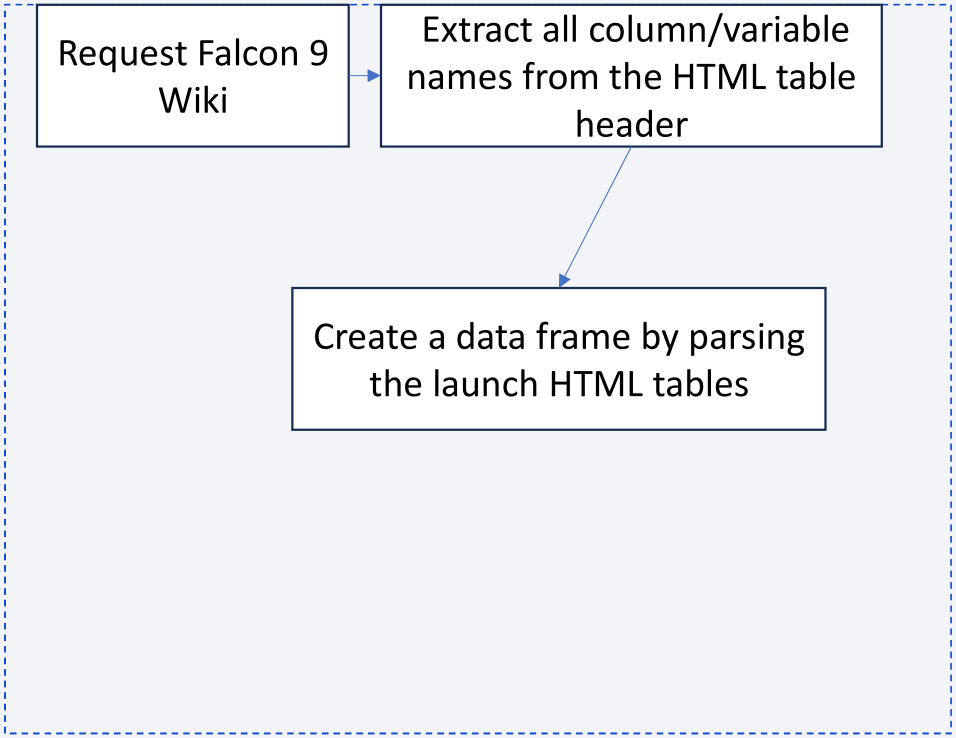
- Data collection with SpaceX REST calls

- GitHub URL for Data Collection Notebook
  - https://github.com/Gabe-Gagliardo/IBM-Data-Science-Capstone



Data Collection - Scraping

- GitHub URL for WebScraping Notebook
  - https://github.com/Gabe-Gagliardo/IBM-Data-Science-Capstone



Data Wrangling

- The flow chart below illustrates how the data was processed.

Calculate the number of launches on each site → Calculate the number and occurrence of each orbit → Calculate the number and occurrence of mission outcome of the orbits → Create a landing outcome label from Outcome column

- GitHub URL for Data Wrangling Notebook
  - https://github.com/Gabe-Gagliardo/IBM-Data-Science-Capstone

EDA with Data Visualization

- Charts used
  - Scatter plots to visualize the relationship between Flight Number and Launch Site
  - Scatter plot to visualize the relationship between Payload Mass and Launch Site
  - Bar chart to visualize the relationship between success rate of each orbit type
  - Scatter plot to visualize the relationship between FlightNumber and Orbit type
  - Scatter plot to visualize the relationship between Payload Mass and Orbit type
  - Line chart to visualize the launch success yearly trend
- GitHub URL for Data Visulatization Notebook
  - https://github.com/Gabe-Gagliardo/IBM-Data-Science-Capstone

EDA with SQL

- SQL Query Summary
  - %sql SELECT DISTINCT LAUNCH_SITE as "Launch_Sites" FROM SPACEXTBL;
  - %sql SELECT * FROM 'SPACEXTBL' WHERE Launch_Site LIKE 'CCA%' LIMIT 5;
  - %sql SELECT SUM(PAYLOAD_MASS__KG_) as PM_KG_TOTAL, Customer FROM 'SPACEXTBL' WHERE Customer = 'NASA (CRS)'
  - %sql SELECT AVG(PAYLOAD_MASS__KG_) as PM_KG_AVG FROM 'SPACEXTBL' WHERE Booster_Version LIKE 'F9 v1.1%'
  - %sql select BOOSTER_VERSION from SPACEXTBL where LANDING_OUTCOME='Success (drone ship)' and PAYLOAD_MASS__KG_ BETWEEN 4000 and 6000;
  - %sql select count(MISSION_OUTCOME) as missionoutcomes from SPACEXTBL GROUP BY MISSION_OUTCOME;
  - %sql select BOOSTER_VERSION as boosterversion from SPACEXTBL where PAYLOAD_MASS__KG_=(select max(PAYLOAD_MASS__KG_) from SPACEXTBL);
  - %sql SELECT MONTH(DATE),MISSION_OUTCOME,BOOSTER_VERSION,LAUNCH_SITE FROM SPACEXTBL where EXTRACT(YEAR FROM DATE)='2015';
  - %sql SELECT LANDING_OUTCOME FROM SPACEXTBL WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20' ORDER BY DATE DESC;
- GitHub URL for EDA SQL Notebook
  - https://github.com/Gabe-Gagliardo/IBM-Data-Science-Capstone

Build an Interactive Map with Folium

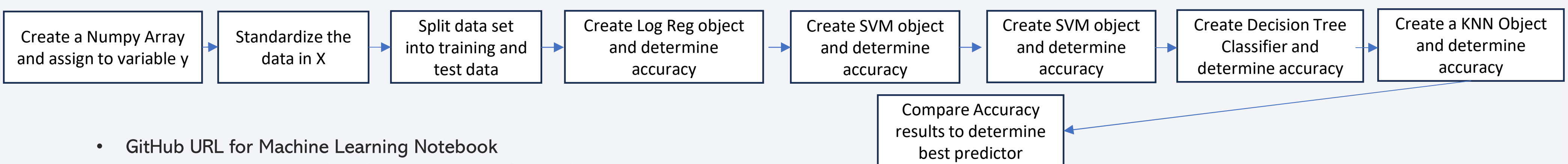
- Launch Site markers were added to easily locate launch locations. Additionally, for each site, Marker clusters were added for each launch at given location. A Ploy line was added to indicate distance from a launch site to the nearest coast line.
- These objects were added to understand if launch sites are determined by proximity to key features such as coast lines and large cities.
- GitHub URL for Folium Map Notebook
  - https://github.com/Gabe-Gagliardo/IBM-Data-Science-Capstone

Build a Dashboard with Plotly Dash

- Summarize what plots/graphs and interactions you have added to a dashboard
    - Launch site Drop-down Component
    - Callback function to render based on selected site Dropdown
    - Range Slider to Select Payload
    - Callback function to render the scatter plot
- These interactions are needed to filter and visualize data dynamically
- Add the GitHub URL of your completed Plotly Dash lab, as an external reference and peer-review purpose
    - Can not provide a Ploty dash link of the completed dashboard. IT polices on my work computer prevent me from launching the app in this lab

Predictive Analysis (Classification)

- Summarize how you built, evaluated, improved, and found the best performing classification model
- Functions from the Scikit-learn library are used to create our machine learning models.
- The machine learning prediction phase include the following steps:
  - Standardizing the data
  - Splitting the data into training and test data
  - Creating machine learning models, which include:
    - Logistic regression
    - Support vector machine (SVM)
    - Decision tree
    - K nearest neighbors (KNN)
  - Fit the models on the training set
  - Find the best combination of hyperparameters for each model
  - Evaluate the models based on their accuracy scores and confusion matrix
  - It was determined that Dtree was the best model for this project



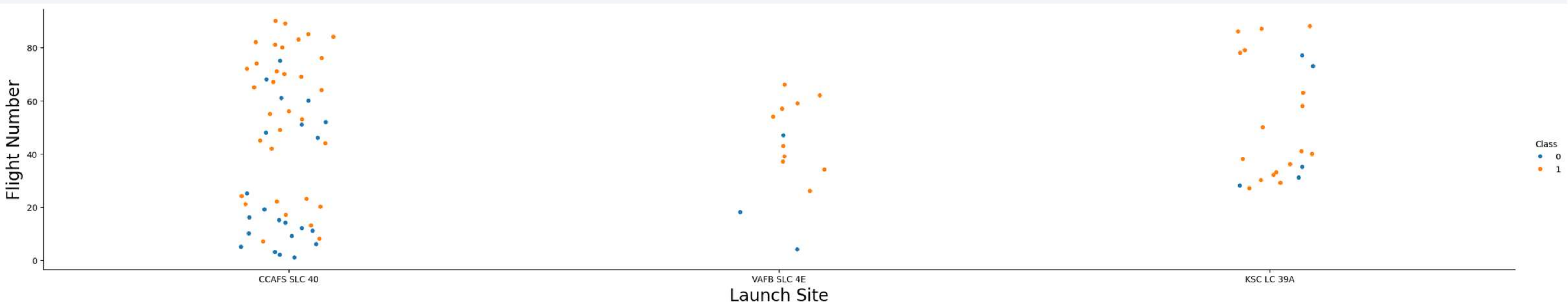
- GitHub URL for Machine Learning Notebook
  - https://github.com/Gabe-Gagliardo/IBM-Data-Science-Capstone

Section 2

Insights drawn from EDA

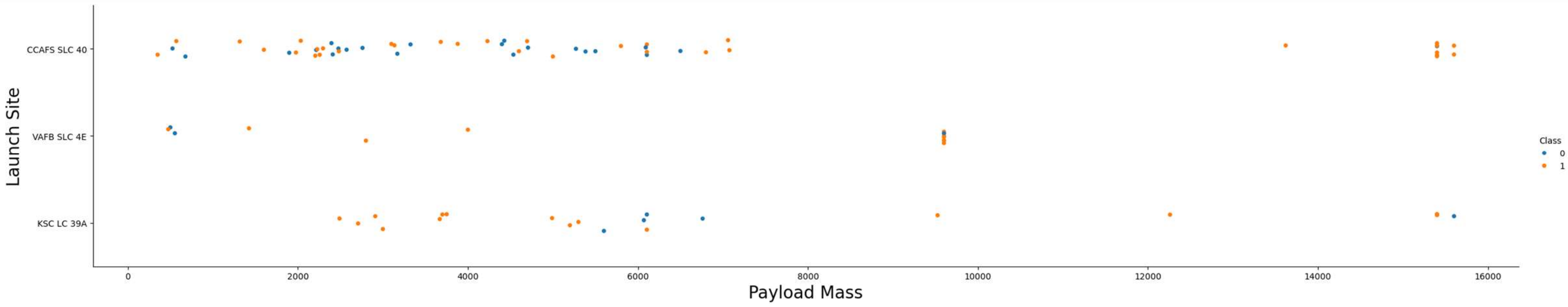
Flight Number vs. Launch Site

- Show a scatter plot of Flight Number vs. Launch Site
- The Scatter plot indicates the relationship between Launch Sites and Flight number
- CCAFS SLC 40 has the most flights



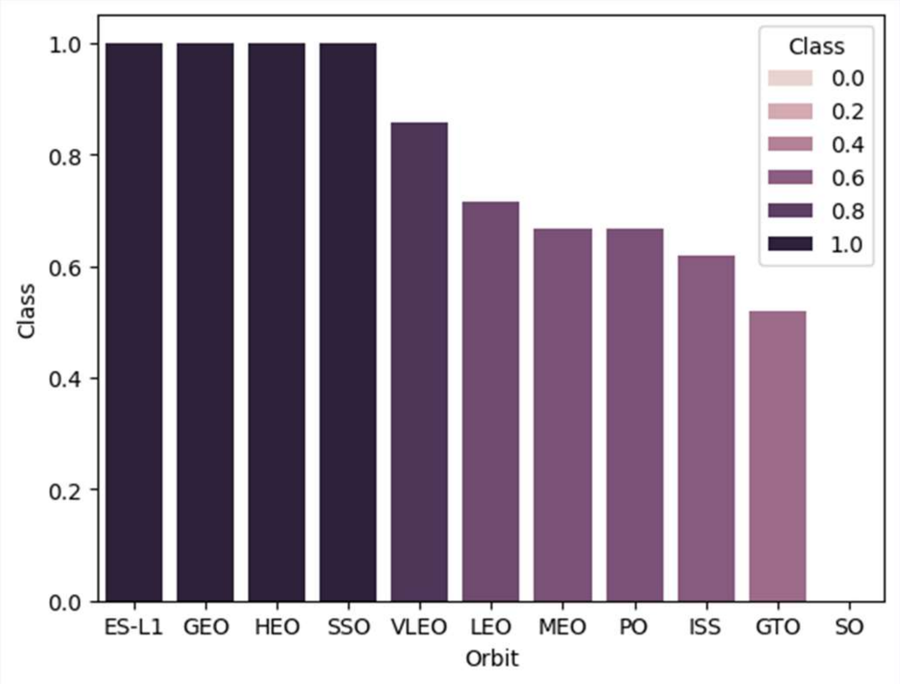
Payload vs. Launch Site

- Show a scatter plot of Payload vs. Launch Site
- This scatter plot shows the relationship between launch site and payload mass. At the VAFB-SLC launch site there are no rockets launched with a heavy payload mass(greater than 10000).



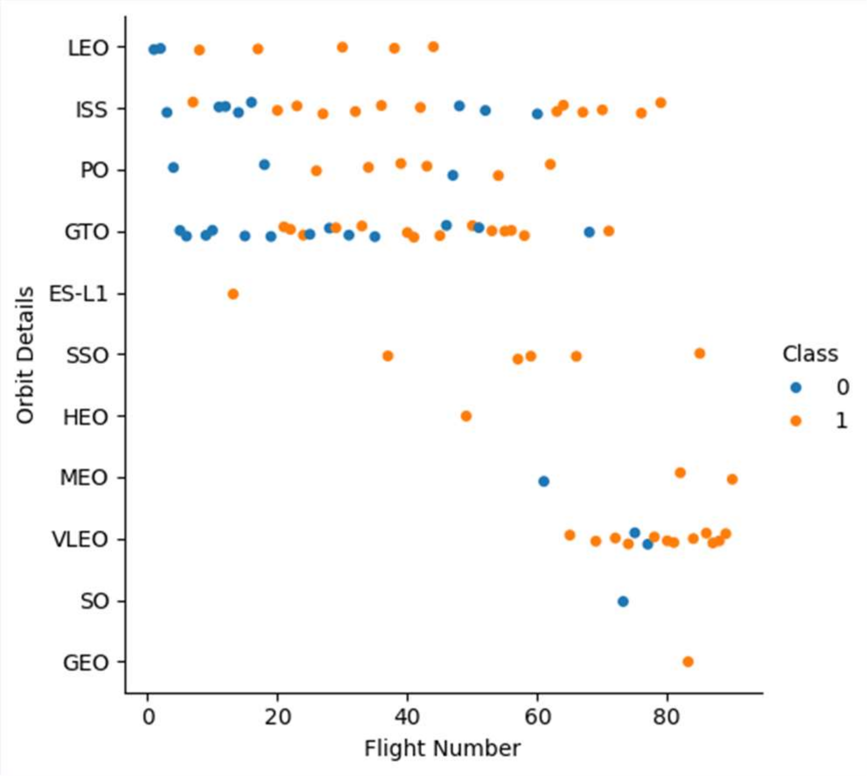
Success Rate vs. Orbit Type

- Show a bar chart for the success rate of each orbit type
- ES-L1, GEO, HEO and SSO are the mission orbits with the highest success rate.



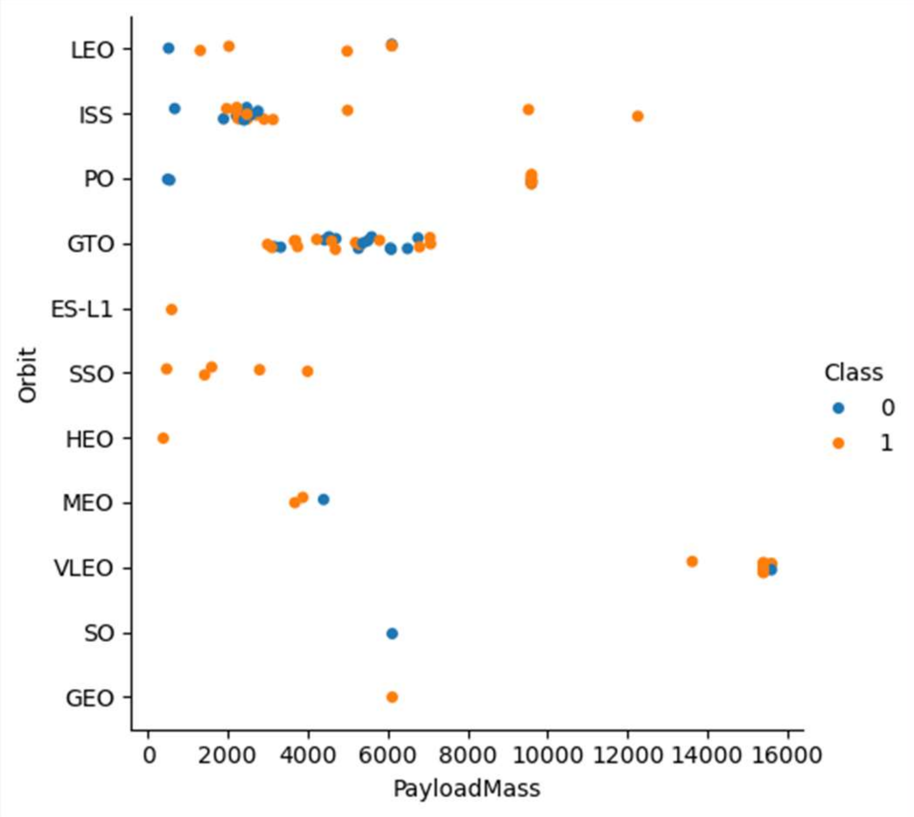
Flight Number vs. Orbit Type

- Show a scatter point of Flight number vs. Orbit type
- The orbit type with the highest flight numbers is VLEO. This suggests that most flight towards the end of this data-set timeline are VLEO because flight number is related to flight date.



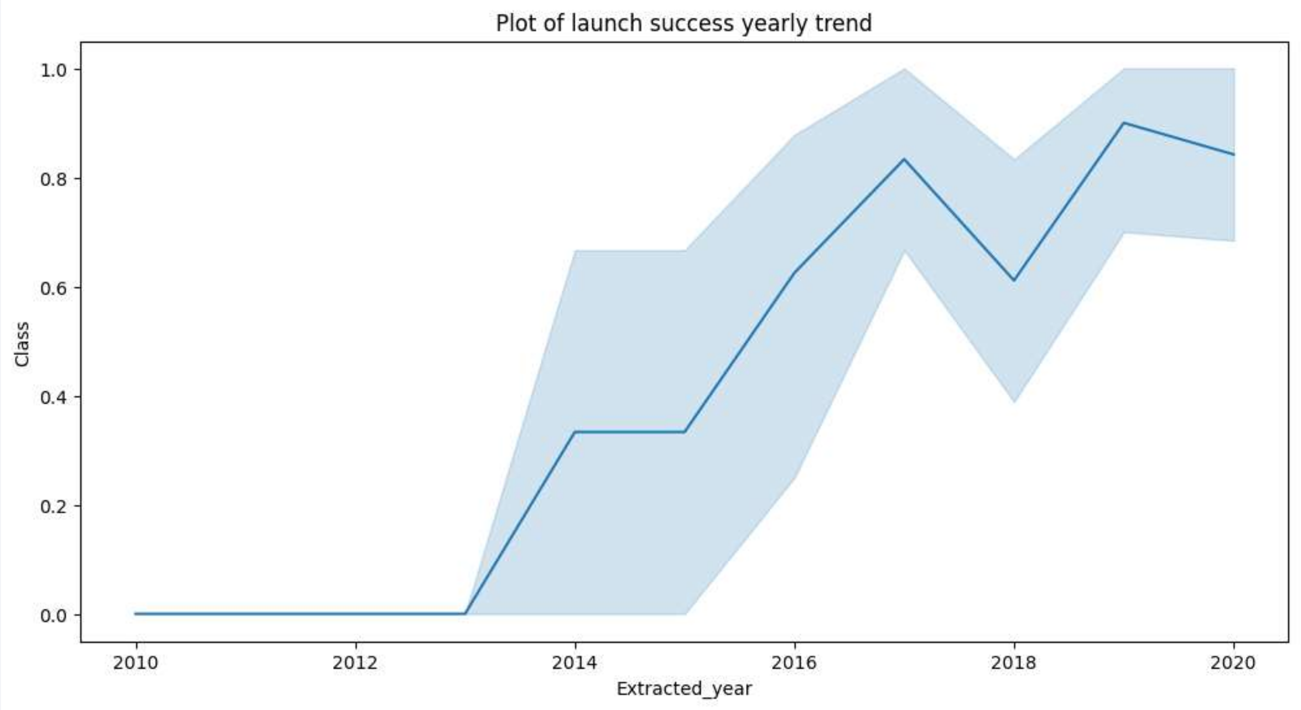
Payload vs. Orbit Type

- Show a scatter point of payload vs. orbit type
- There is a direct correlation of heavier payloads and lower orbits. Lighter payloads correlate to a wider ranges of orbit types.



Launch Success Yearly Trend

- Show a line chart of yearly average success rate
- As time goes on, there is a higher success rate. This is most likely due to lessons learned from failed launches and advancements in technology over time



All Launch Site Names

- Find the names of the unique launch sites
- Present your query result with a short explanation here
  - There are 4 launch sites used
    - 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`
- Present your query result with a short explanation here

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
- Total payload mass for NASA (CRS) IS 45596 KG

Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1 is 2534.7 KG

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad
- The first successful landing outcome on a ground pad was December 12, 2015

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
- Booster_Version
  - 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
- There is a total of 100 successful mission outcomes and 1 failure

Boosters Carried Maximum Payload

- List the names of the booster which 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

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

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

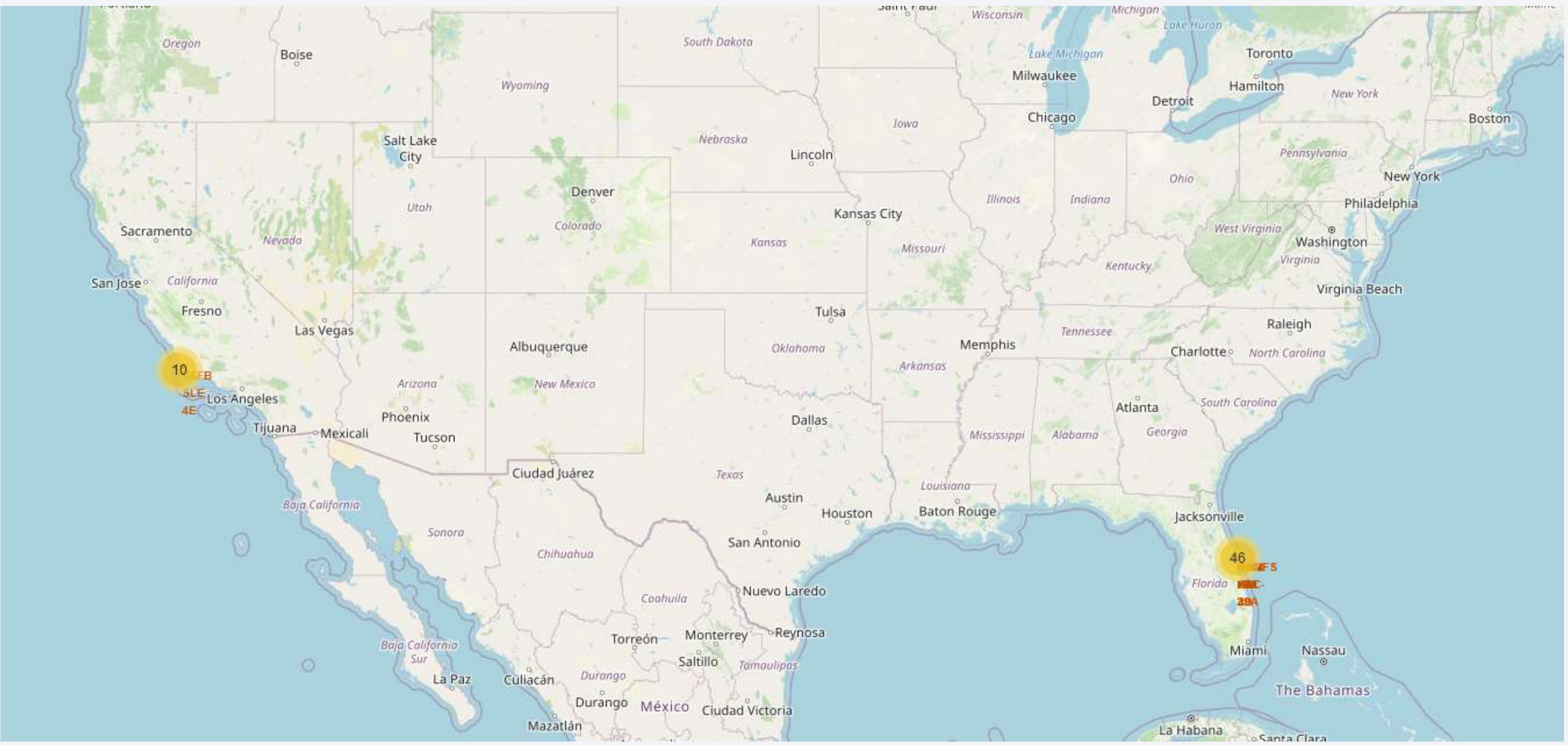
- Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

Landing_Outcome	count("Landing_Outcome")
Success (drone ship)	12
No attempt	12
Success (ground pad)	8
Failure (drone ship)	5
Controlled (ocean)	4
Uncontrolled (ocean)	2
Precluded (drone ship)	1

Section 3

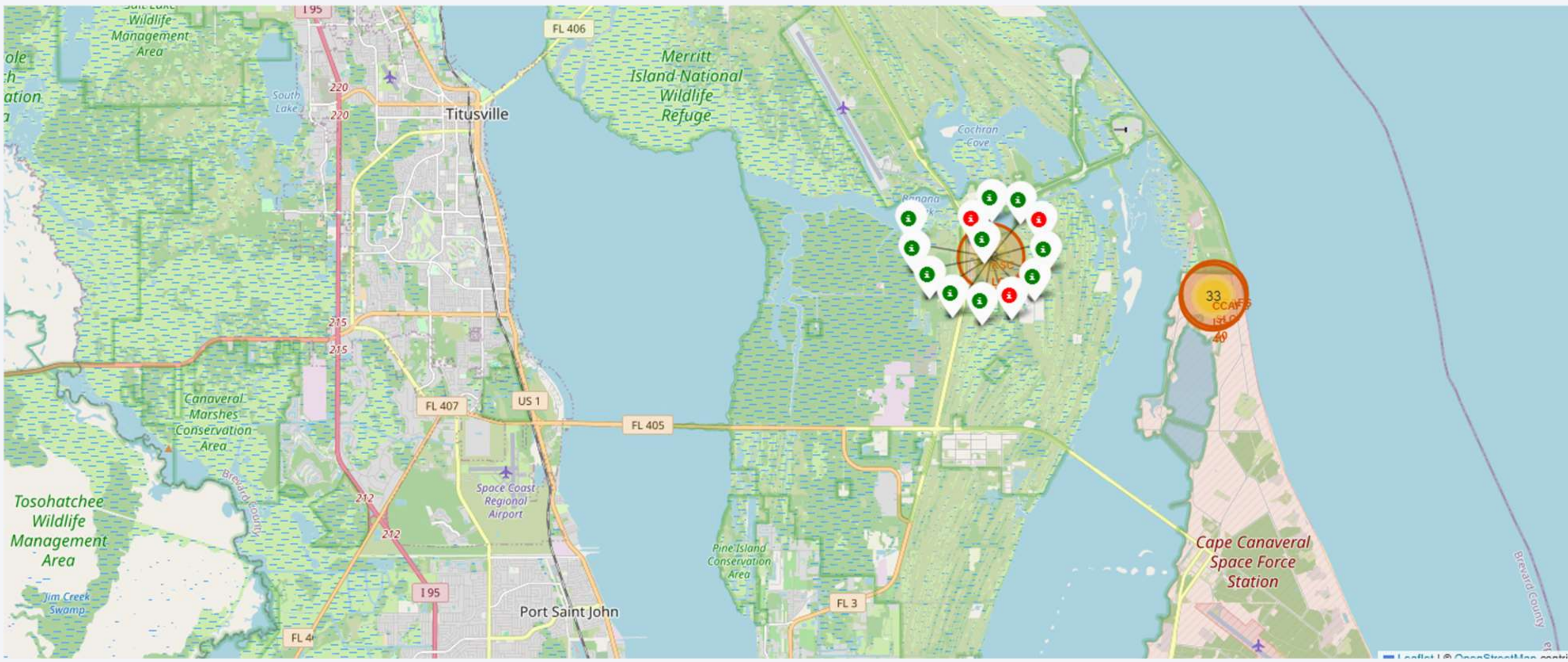
Launch Sites Proximities Analysis

Launch Locations on Map



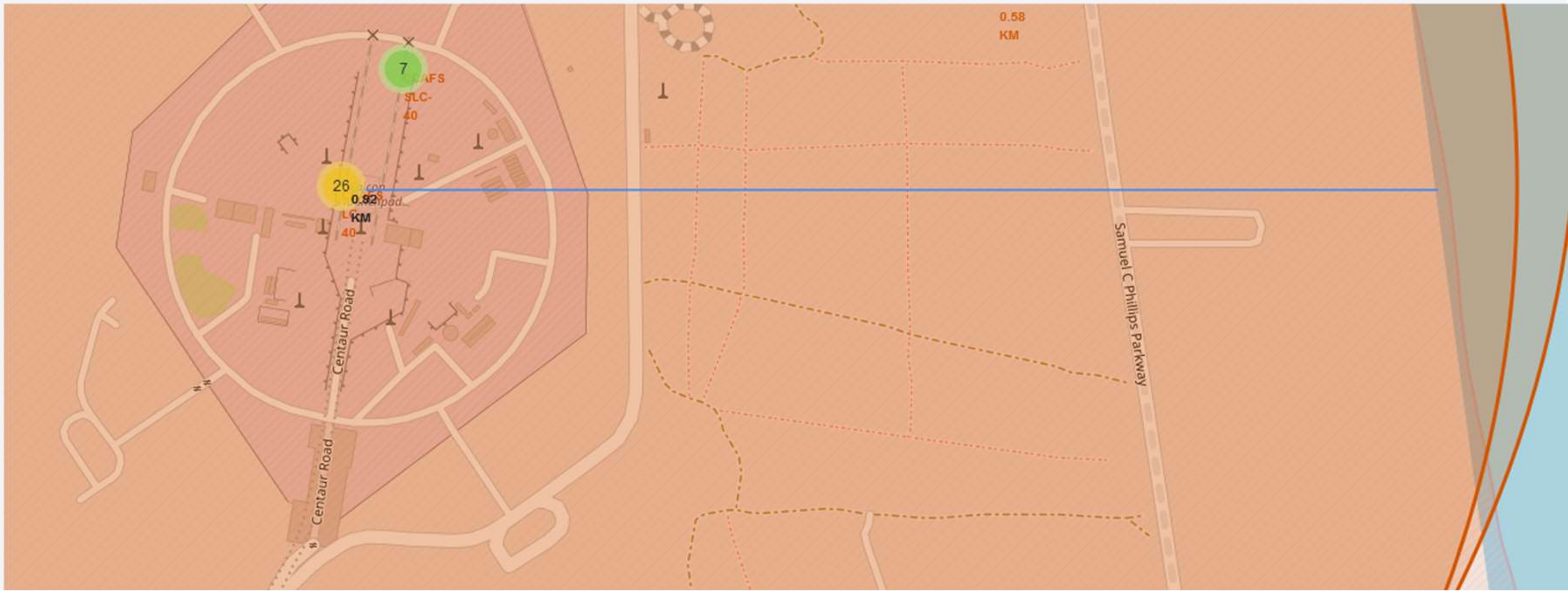
Mission Outcome Marker Example

- Green indicates successful launch at site. Red marker indictaes failed launch.



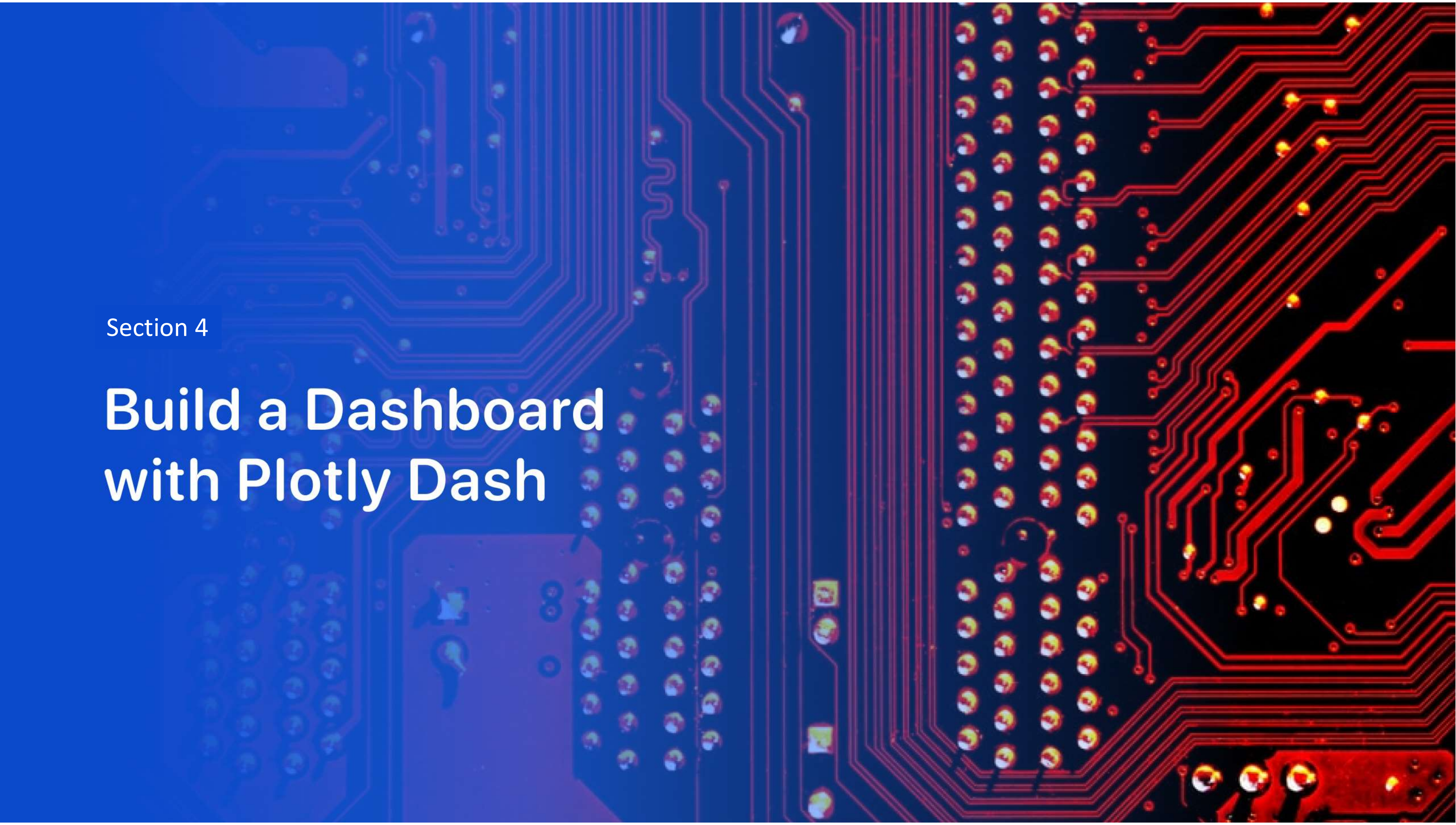
Launch Site in Relation to Distance to Coastline

- Launch site in relation to distance from coastline



Section 4

Build a Dashboard with Plotly Dash



<Dashboard Screenshot 1>

- Replace <Dashboard screenshot 1> title with an appropriate title
- Show the screenshot of launch success count for all sites, in a piechart
- Explain the important elements and findings on the screenshot

<Dashboard Screenshot 2>

- Replace <Dashboard screenshot 2> title with an appropriate title
- Show the screenshot of the piechart for the launch site with highest launch success ratio
- Explain the important elements and findings on the screenshot

<Dashboard Screenshot 3>

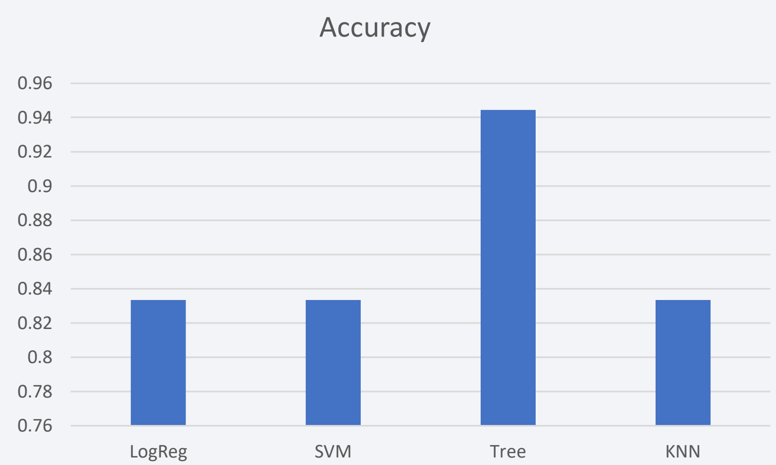
- Replace <Dashboard screenshot 3> title with an appropriate title
- Show screenshots of Payload vs. Launch Outcome scatter plot for all sites, with different payload selected in the range slider
- Explain the important elements and findings on the screenshot, such as which payload range or booster version have the largest success rate, etc.

Section 5

Predictive Analysis (Classification)

Classification Accuracy

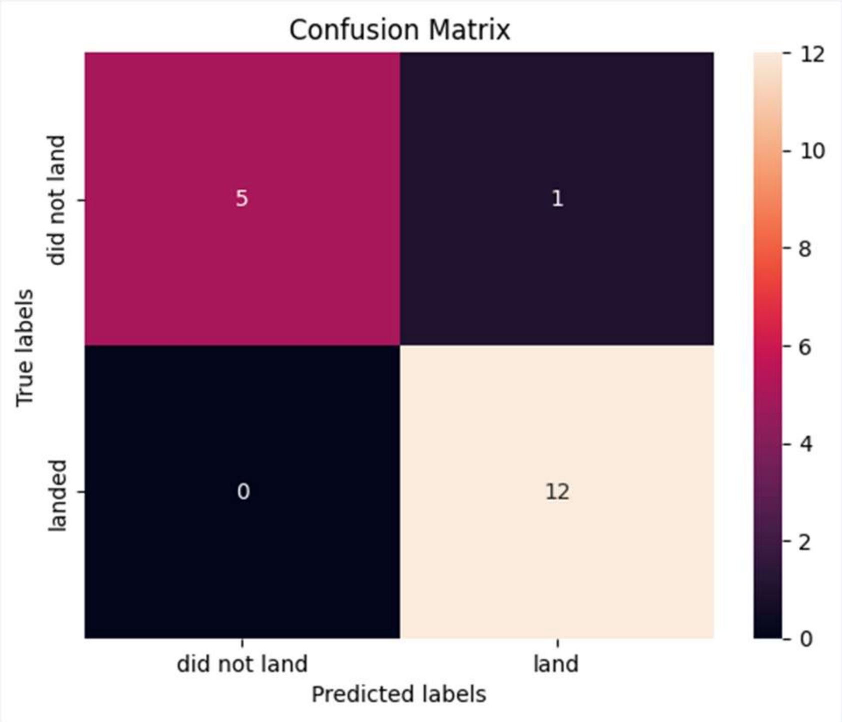
- Visualize the built model accuracy for all built classification models, in a bar chart
- According the testing, Dtree has the highest accuracy



	LogReg	SVM	Tree	KNN
Jaccard_Score	0.800000	0.800000	0.923077	0.800000
F1_Score	0.888889	0.888889	0.960000	0.888889
Accuracy	0.833333	0.833333	0.944444	0.833333
GridSearch CV	.846	.848	.88	.848

Confusion Matrix

- Show the confusion matrix of the best performing model with an explanation
- The Tree confusion matrix is shown here. It indicates the most accurate true and false predictions among the evaluated models.



Conclusions

- The data suggests there is a slightly higher success rate for drone ship landing than ground, which is counterintuitive considering the variables a body of water introduces to the task.
- Data also indicates there is a very high success rate overall, which is encouraging considering the associated cost savings with the SpaceX solution.
- Success rates have also increased over time, which is indicative of solid engineering practices and learning from failures
- Based on the accuracy of all models, Tree appears to be the most reliable method to predict launch/mission Success.


Thank you!