IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Angela

Outline

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

Executive Summary

- Data collection: used SpaceX API and json response and webscraping with Falcon9 launch html page
- Data wrangling: checked for missing values, filtered rows and columns, calculate new values
- EDA: created graphs, queried with SQL
- Used Folium and Plotly Dash to make interactive visual analytics (maps, charts)
- Built and evaluated different classification models

Introduction

In the commercial space age, many companies are competing to make space travel affordable to for everyone. SpaceX is one of the most successful companies. SpaceX launches Falcon9 rockets for cheap because they are able to reuse the first stage. If we can determine whether the first stage can land, we can determine the cost of a launch.

I am trying to find out the cost of each launch by determining if SpaceX will reuse the first stage.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Describe how data was collected
- Perform data wrangling
  - Describe how data was processed
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection – SpaceX API

Request rocket launch data from SpaceX API → get json response → convert result into a dataframe → clean up data (remove unnecessary information, convert data types) → create dataset using data extracted and cleaned → filter dataset to include only falcon9 launches → deal with missing data → export to csv

Github link for data collection SpaceX API notebook

Data Collection - Scraping

perform an HTTP GET method to request the Falcon9 Launch HTML page → create BeautifulSoup object from the HTML response → Extract all column/variable names from the HTML table header → Create a data frame by parsing the launch HTML tables → export to csv

Github link for webscraping notebook

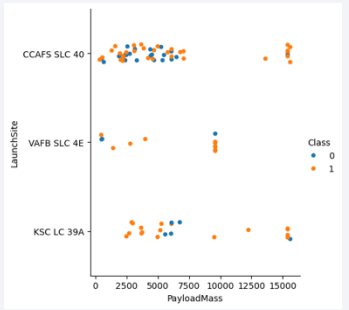
Data Wrangling

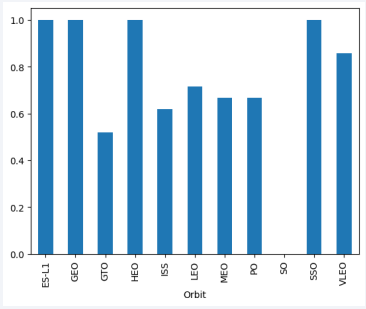
Check for missing values and data types in the dataset → calculate the number of launches on each site → calculate the number and occurrence of each orbit → calculate the number and occurrence of mission outcome of the orbits → create a landing outcome label from Outcome column → export to csv

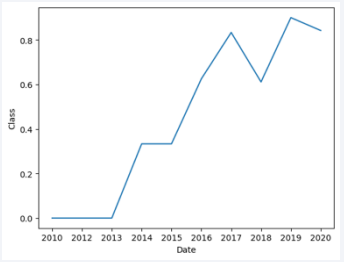
Github link for data wrangling notebook

EDA with Data Visualization

- Scatter plot:
  - Used for Launch Site vs Flight Number and Launch Site vs Payload Mass
  - Can see the relationship between the two variables through the distribution of each data point
- Bar Graph:
  - Success rate of each orbit
  - Shows values for categorical variables (orbit in this case)
- Line plot:
  - Look at success rate over time







[Github link for EDA with Data Visualization notebook]

EDA with SQL

- Display the names of the unique launch sites in the space mission
- Display 5 records where launch sites begin with the string 'CCA'
- Display the total payload mass carried by boosters launched by NASA (CRS)
- Display average payload mass carried by booster version F9 v1.1
- List the date when the first succesful landing outcome in ground pad was acheived.
- 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
- List all the booster_versions that have carried the maximum payload mass, using a subquery with a suitable aggregate function.
- List the 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 landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order.

Github link for EDA with SQL notebook

Build an Interactive Map with Folium

- Created a circle object to add a highlighted circle area with a text label on a specific coordinate
- Used markers to mark the success/failed launches for each site on the map
- Used a line to mark distance from launch site to the nearest coastline

Github link to Folium notebook

Build a Dashboard with Plotly Dash

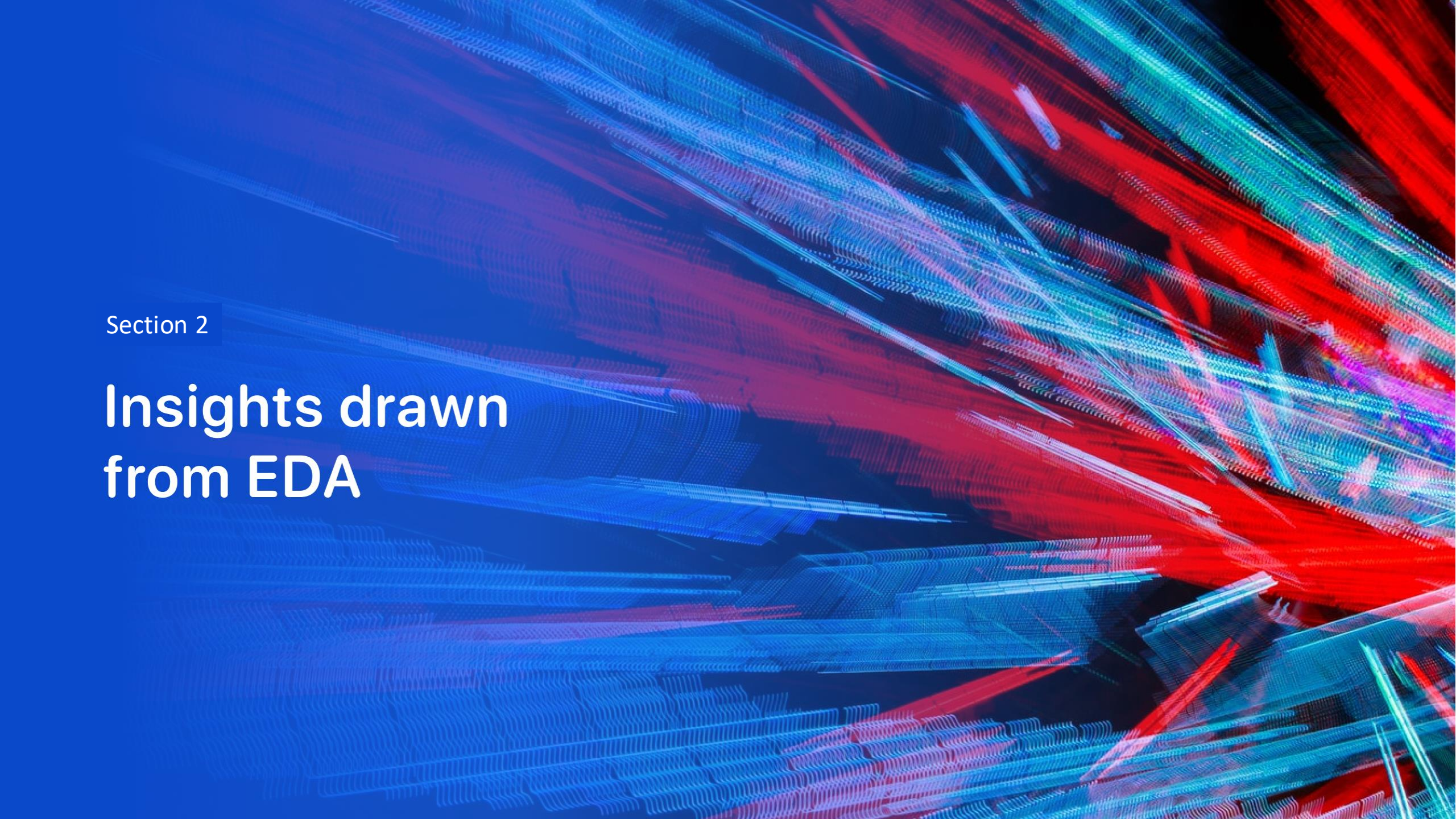
- A dropdown list at the top to filter by sites
- Pie chart showing success rates of all sites or success rate of each site
- Slider to allow user to selected payload range
- Scatter plot showing the correlation between payload and launch success

Github link to Plotly Dash file

Predictive Analysis (Classification)

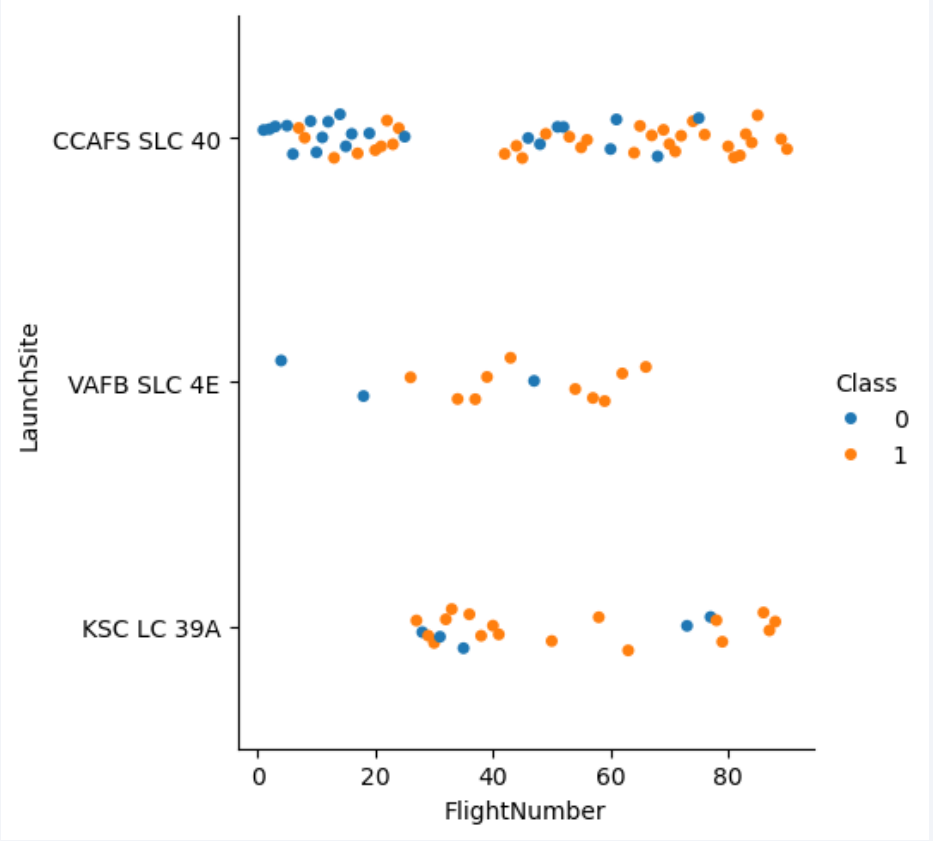
Split the data x and y into training and testing sets → create model object → create a GridSearchCV object → fit object to find best parameters → calculate accuracy using score → look at confusion matrix → repeat with different models and compare results to find best performing model

Github link to machine learning notebook


Section 2
Insights drawn from EDA

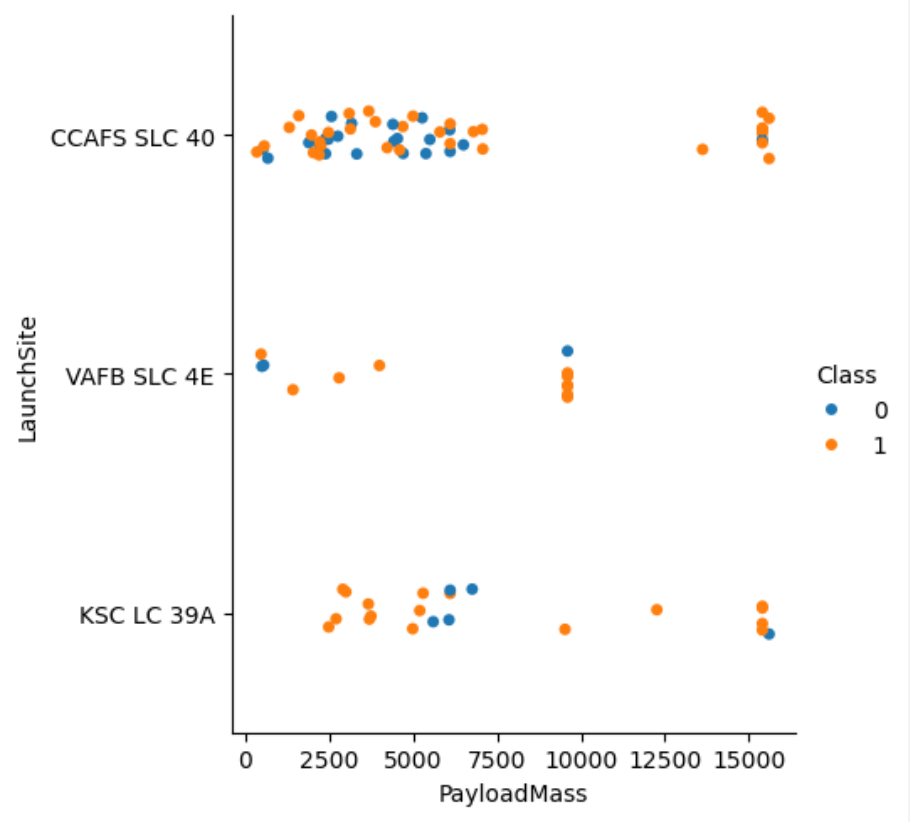
Flight Number vs. Launch Site

- There is not much relationship between flight number and launch site



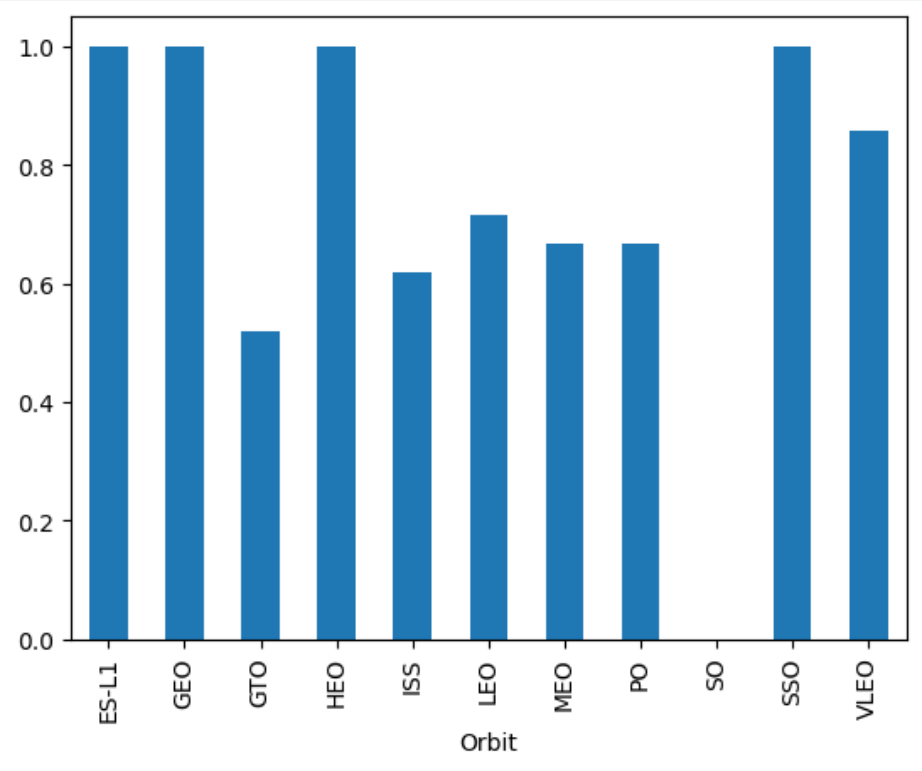
Payload vs. Launch Site

- The VAFB SLC 4E launch site doesn't have any rockets launched for payload mass greater than 10000



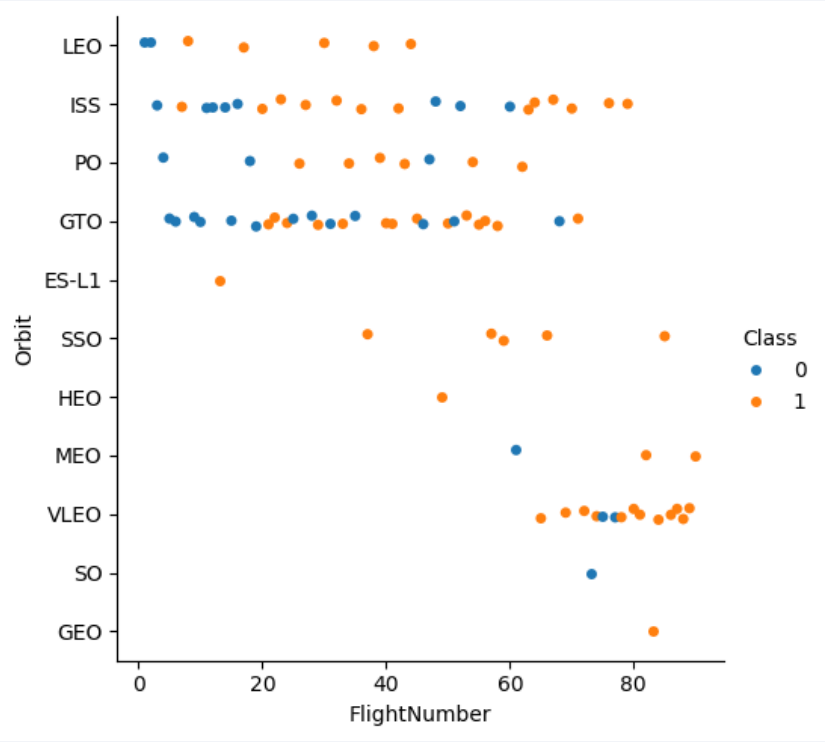
Success Rate vs. Orbit Type

- The orbits with the highest success rate are ES-L1, GEO, HEO, and SSO



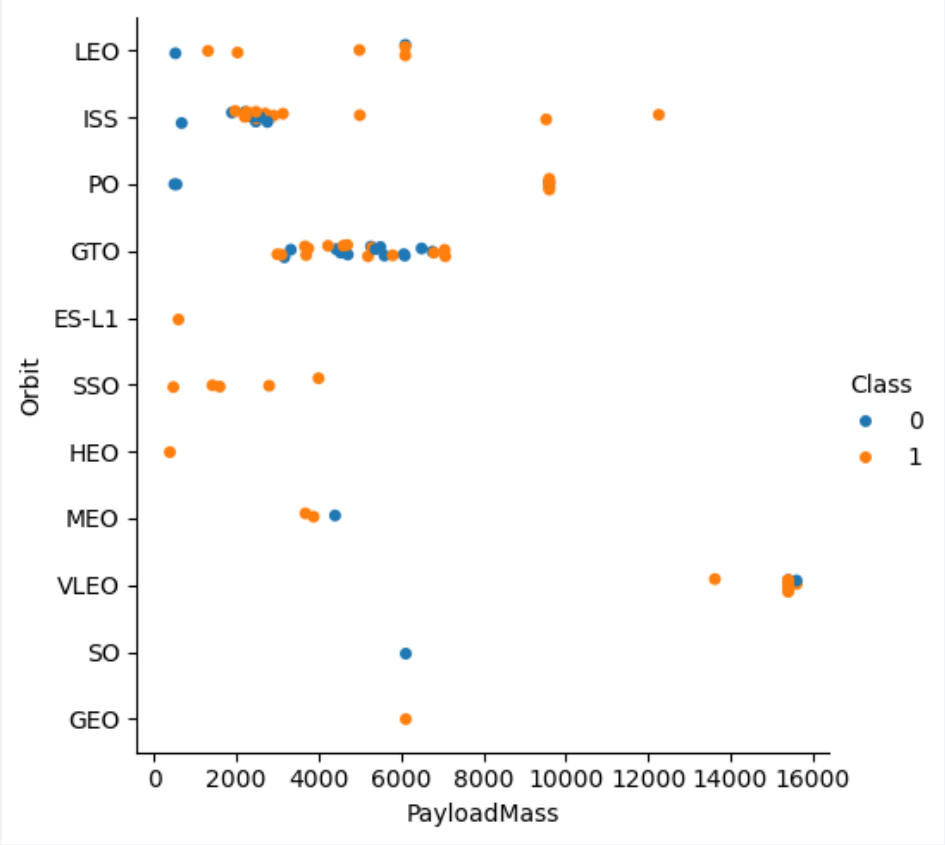
Flight Number vs. Orbit Type

- There is a slight trend of success increasing with more flights



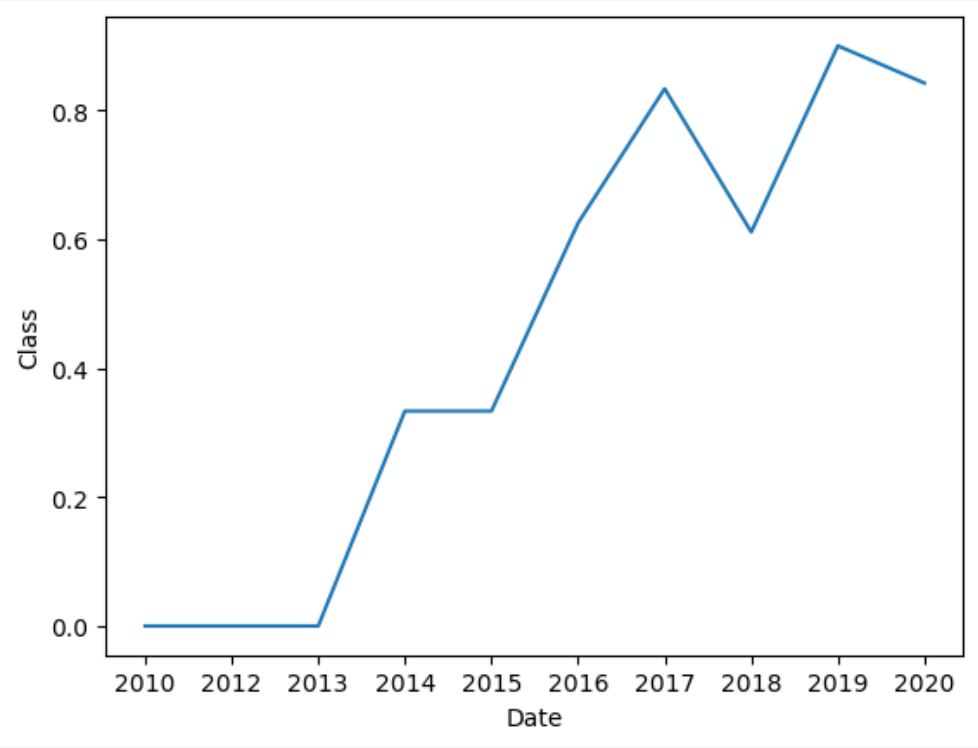
Payload vs. Orbit Type

- For a couple of the orbits, a higher payload mass has more success
- For GTO, the results are all mixed



Launch Success Yearly Trend

- After 2013, there is a continuous overall increase in success rate



All Launch Site Names

- These are the unique names of the launch sites

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

Launch Site Names Begin with 'CCA'

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (p
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 (p
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	N
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	N
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	N

Total Payload Mass

- total payload carried by boosters from NASA: 45596 kg

Average Payload Mass by F9 v1.1

- Average payload mass carried by booster version F9 v1.1: 2928.4 kg

First Successful Ground Landing Date

- Date of the first successful landing outcome on ground pad: 2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- names of boosters which 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

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

Boosters Carried Maximum Payload

- List of names of the boosters which have carried the maximum payload mass

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

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

Month	Landing_Outcome	Booster_Version	Launch_Site
01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

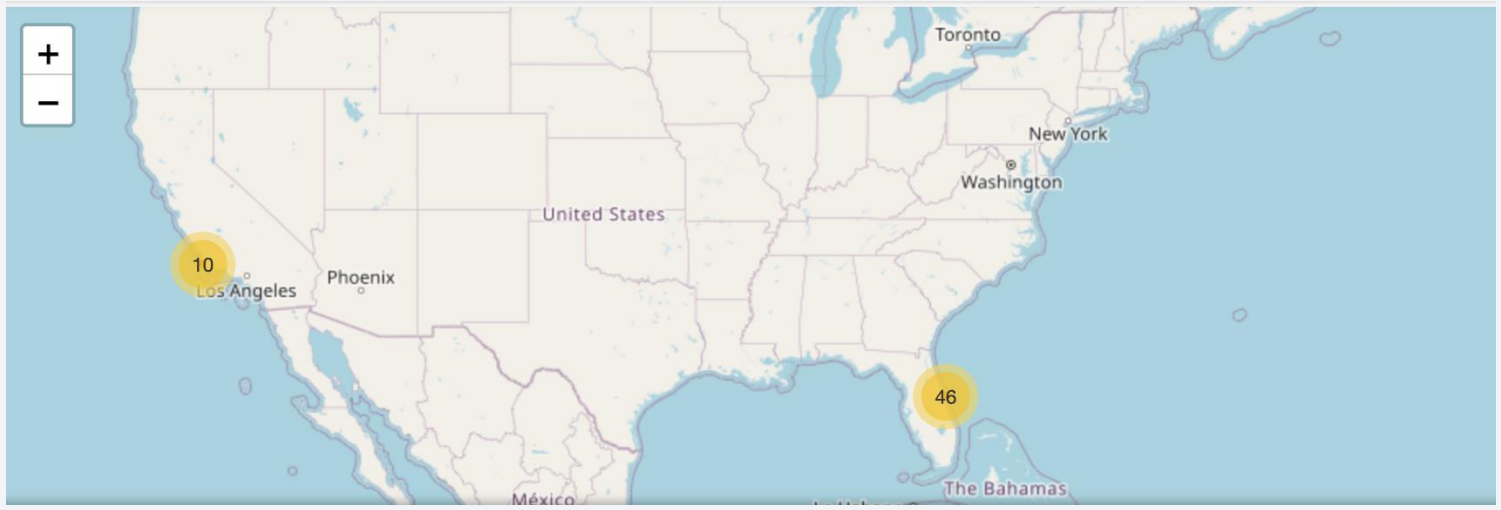
- count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

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

Section 3

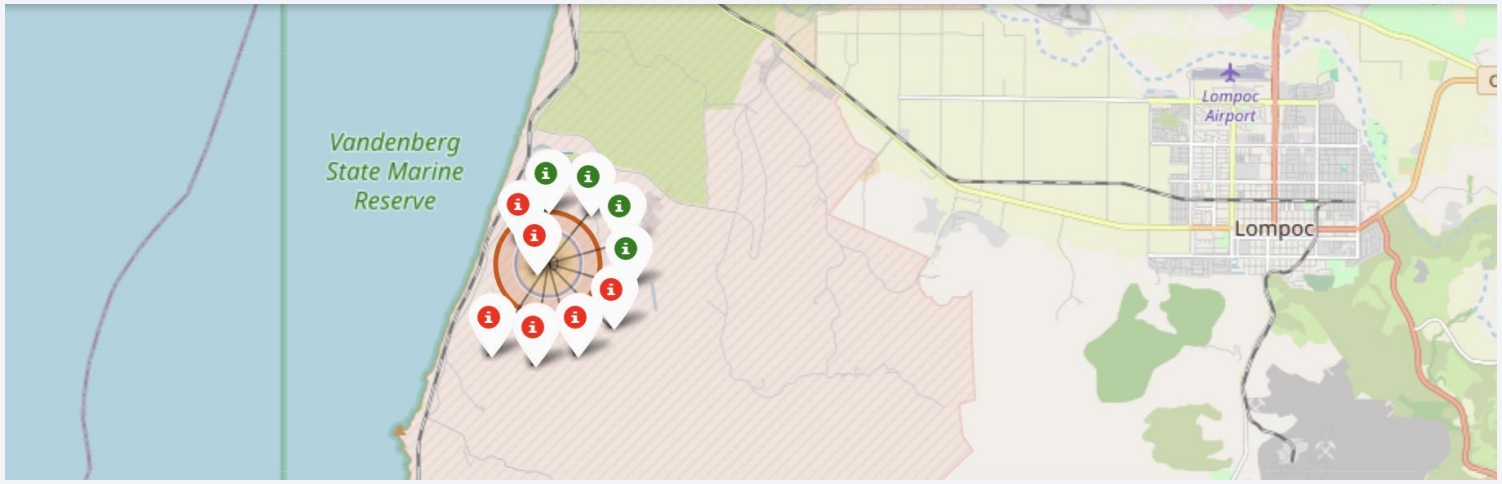
Launch Sites Proximities Analysis

Launch Site Location Markers



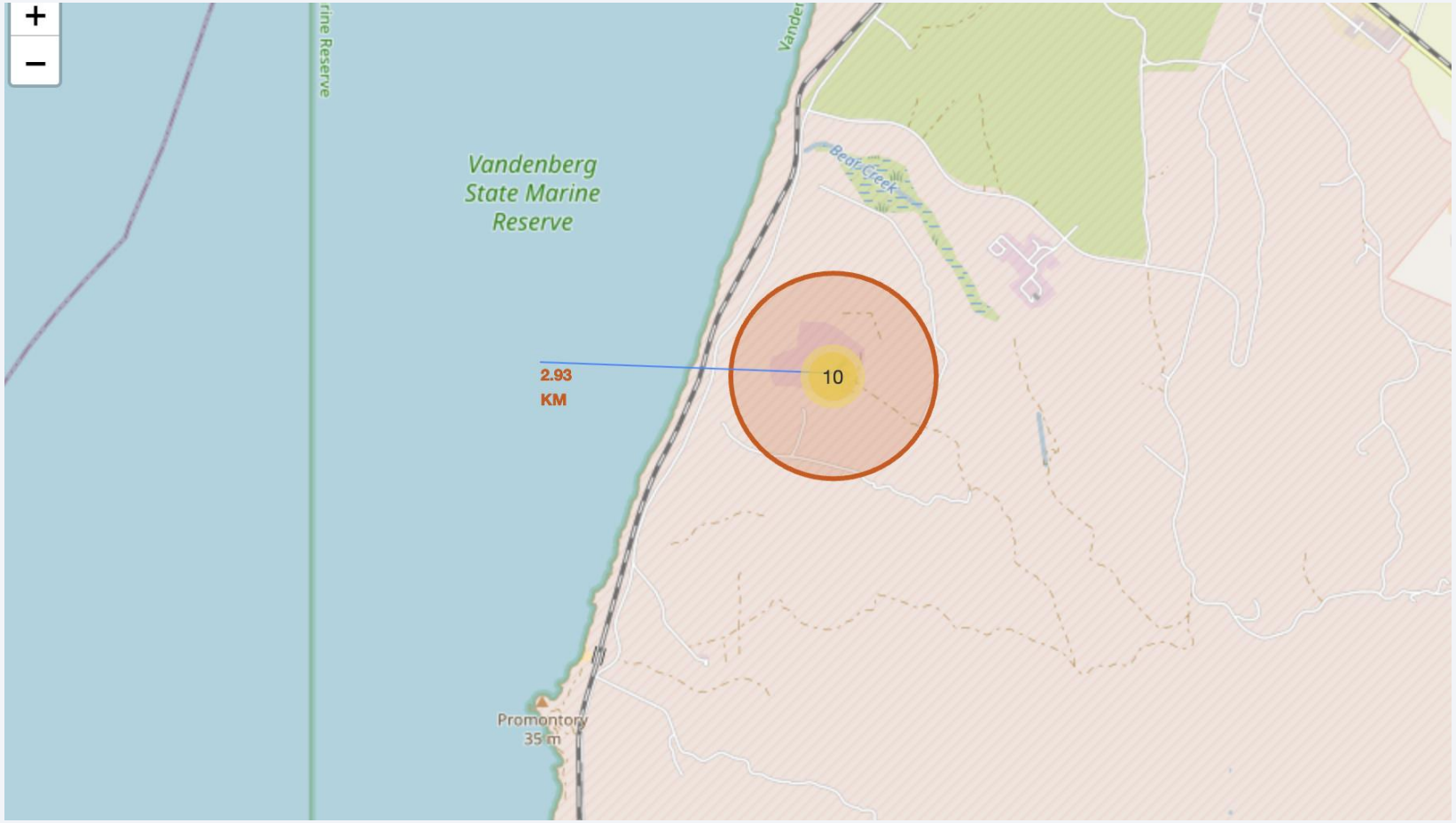
The launch sites are near coastlines

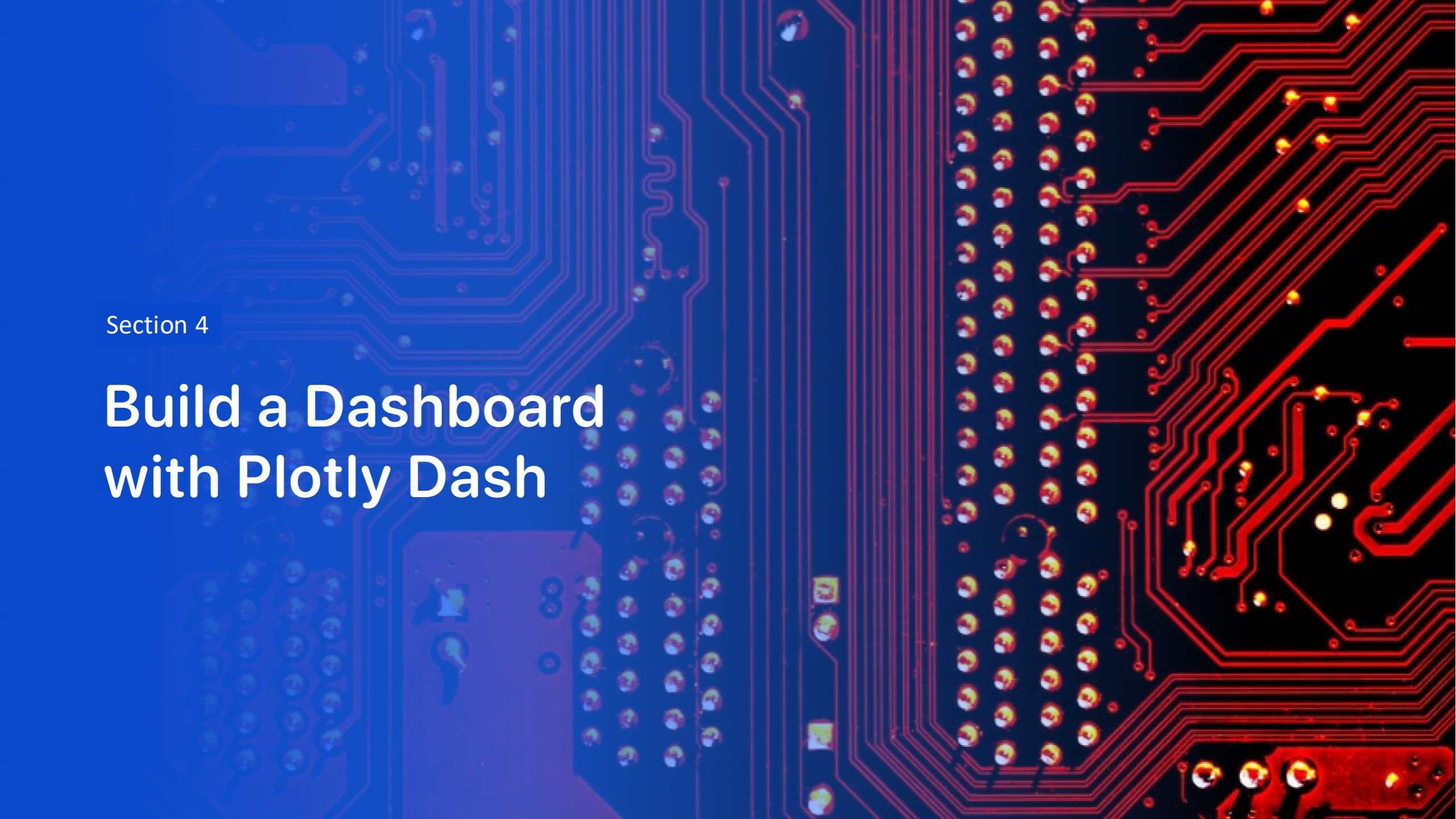
Color Labeled Launch Outcomes



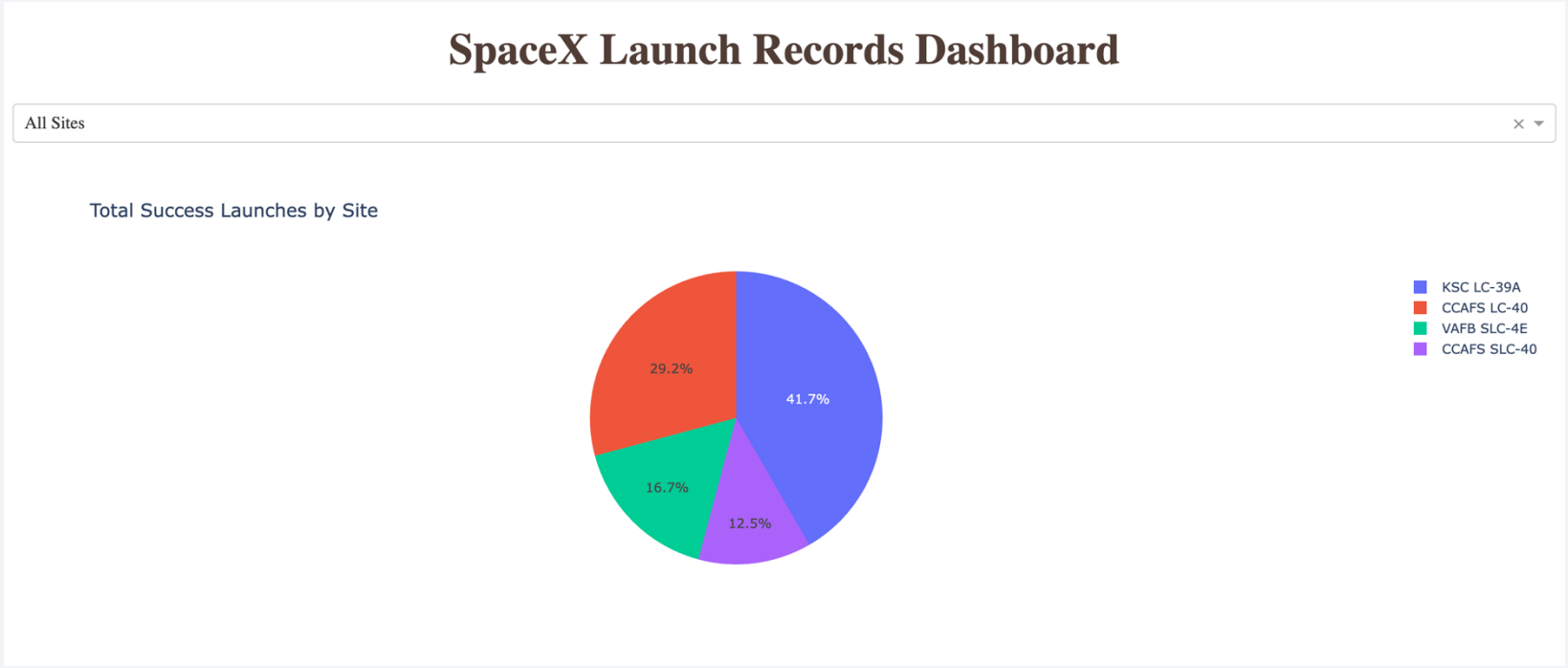
There is a somewhat even mix of failures and successes in this launch site shown.

Distance to Coastline



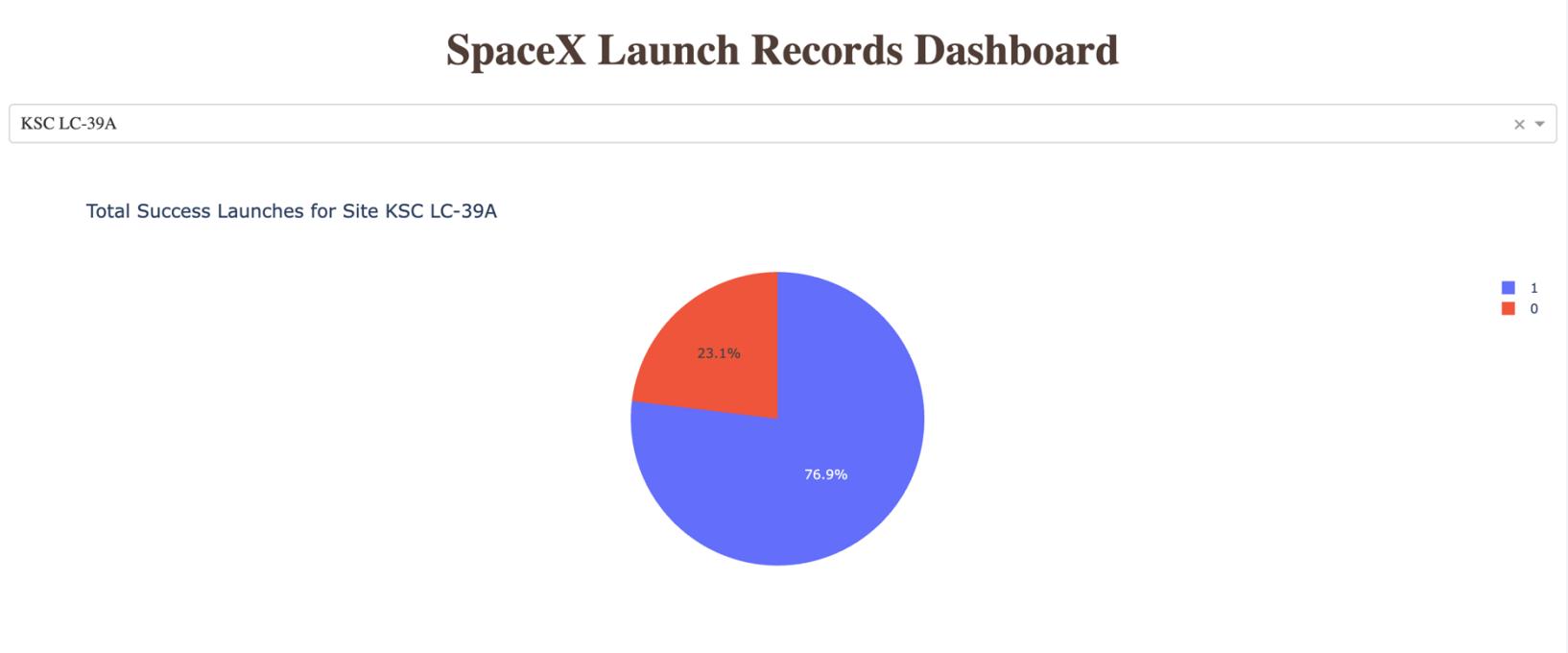

Section 4
Build a Dashboard with Plotly Dash

All Sites Piechart



The most successful launch site is KSC C-39A, with a 41.7% success rate

Piechart for Most Successful Launch Site



76.9% of launches from site KSC LC-39A were successful

Scatterplot for all Sites

Only booster B4 had a payload mass above 7k, and the rest are dispersed in the lower range.

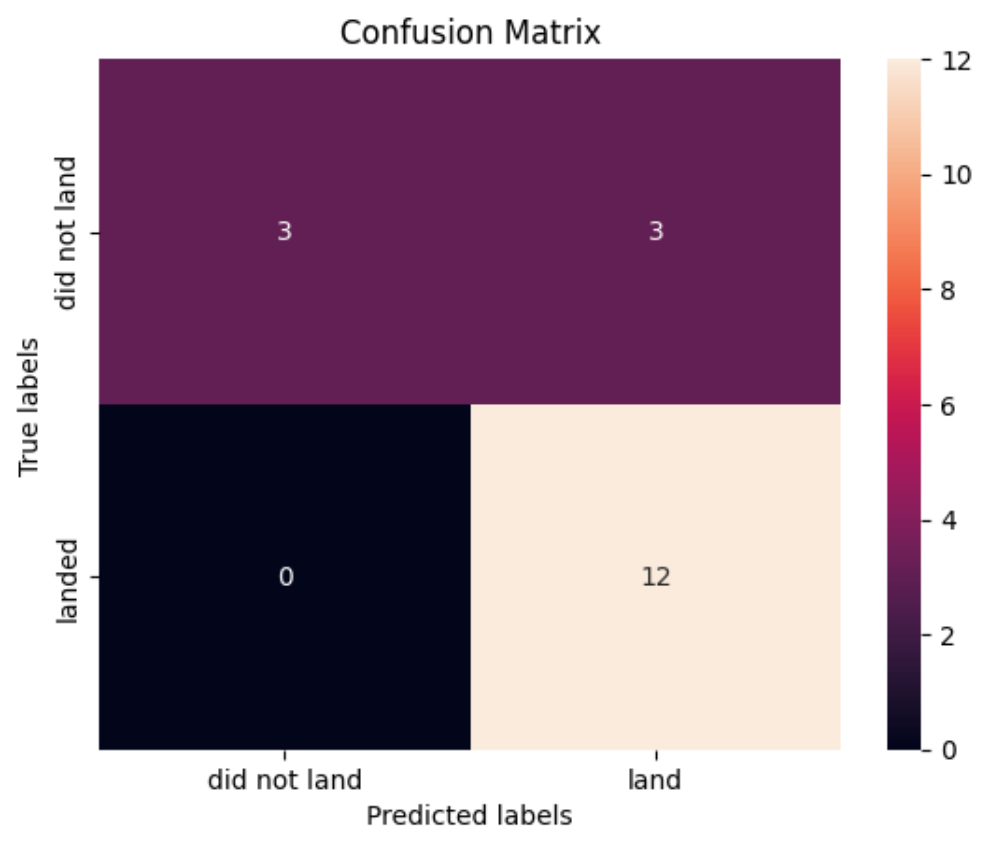
Section 5

Predictive Analysis (Classification)

Classification Accuracy

- The model with the highest accuracy is the support vector machine (svm) model

Confusion Matrix



There are a good amount of true positive, but still a few false positives.

Conclusions

- There is a correlation between the number of flights and success
- Launch success has been continuously improved in the past few years
- Launch sites are close to coastlines
- The most successful launch site is KSC C-39A

Thank you!