IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Charlie Radke
10/5/2023
charlesradke58@gmail.com

Outline

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

Executive Summary

This presentation will develop into a story of a data science journey for this project, and is compelling and easy to understand. The following slides describe the methods used to collect the data, wrangle the data, transform the data and present a visual of the analysis the data enabled.

Summary of all results:

The resulting charts, data frames, maps, dashboards, and machine learning model exploration enable further refinement of the presented analysis visuals as needed by the customer.

Introduction

Space X advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because Space X can reuse the first stage. Therefore if we can determine if the first stage will land, we can determine the cost of a launch. This information can be used if an alternate company wants to bid against space X for a rocket launch.

Section 1

Methodology

Methodology

Executive Summary

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

Data Collection

- The following slides describe how data sets were collected.

Data Collection – SpaceX API

Data collection results are represented here:

FlightNumber	Date LaunchSite Reused ReusedCount	BoosterVersion Outcome Legs Serial	PayloadMass Flights LandingPad Longitude	Orbit GridFins Block Latitude
4	1 LEO False 1.0 28.561857	2010-06-04 CCSFS SLC 40 False 0	Falcon 9 None None False B0003	NaN 1 None -80.577366
5	2 LEO False 1.0 28.561857	2012-05-22 CCSFS SLC 40 False 0	Falcon 9 None None False B0005	525.0 1 None -80.577366
6	3 ISS False 1.0 28.561857	2013-03-01 CCSFS SLC 40 False 0	Falcon 9 None None False B0007	677.0 1 None -80.577366

- GitHub URL of the completed SpaceX API calls notebook :

https://github.com/cradke58/This_Should_be_Good/blob/main/jupyter-labs-spacex-data-collection-api%20(1)%20(1).ipynb

To complete Data Collection using an API:

- Requested to the SpaceX API. Requesting rocket launch data from SpaceX API with the following URL: spacex_url = https://api.spacexdata.com/v4/launches/past with this code snippet - response = requests.get(spacex_url)
- Cleaned the requested data
- Completed some Data Wrangling
- Filtered the dataframe to only include Falcon 9 launches
- Resulting data is represented in left pane.

Data Collection – Web Scraping

After Web Scraping Data frame

112	113	114	115	116	117	118	119	120	121
112	113	114	115	116	117	118	119	120	121
24 March 2021	7 April 2021	23 April 2021	29 April 2021	4 May 2021	9 May 2021	15 May 2021	26 May 2021	3 June 2021	6 June 2021
08:28	16:34	9:49	03:44	19:01	06:42	22:56	18:59	17:29	04:26
F9 5B1060.6	F9 B5 ♺	F9 B5B1061.2	F9 B5B1060.7	F9 B5B1049.9	F9 B5B1051.10	F9 B5B1058.8	F9 B5B1063.2	F9 B5B1067.1	F9 B5
CCSFS	CCSFS	KSC	CCSFS	KSC	CCSFS	KSC	CCSFS	KSC	CCSFS

- GitHub URL of the completed web scraping notebook here :https://github.com/cradke58/This_Should_be_Good/blob/main/Web_Scraping_NB%20(1).ipynb

To complete Data Collection using Web Scraping:

- Extracted a Falcon 9 launch records HTML table from Wikipedia at this URL: static_url = https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922
- Requested the Falcon9 Launch Wiki page from its URL with this code: response = requests.get(static_url)
- Extracted all column/variable names from the HTML table header for use in constructing data frame
- Parsed the table and converted it into a Pandas data frame
- Data frame represented by graphic at the left

Data Wrangling

- First, completed Exploratory Data Analysis (EDA) using this URL = 'https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/dataset_part_1.csv' and this code snippet:

  ```
  resp = await fetch(URL)
  dataset_part_1_csv = io.BytesIO((await resp.arrayBuffer()).to_py())
  ```

- Then, Calculated the number of launches on each site, number and occurrence of each orbit, and number and occurrences of mission outcome per orbit type
- Next, created a landing outcome label from Outcome column
- Finally, determine the success rate: 0.6666666666666666 based on the "Class" column
- GitHub URL of completed data wrangling related notebooks is at this URL: https://github.com/cradke58/This_Should_be_Good/blob/main/module_1_L3_labs-jupyter-spacex-data_wrangling_jupyterlite.jupyterlite.ipynb

EDA (Exploratory Data Analysis) with Data Visualization

- We used Matplotlib, a plotting library for python and pyplot to give us a MatLab like plotting framework.
- We use Seaborn, a Python data visualization library based on matplotlib for drawing informative statistical graphics
- We plotted charts to visualize relationships described below:
  - Plotted out the FlightNumber vs. PayloadMass and overlay the outcome of the launch showing that different launch sites have different success rates
  - Plotted FlightNumber vs LaunchSite to visualize the relationships to success
  - A Payload Vs. Launch Site scatter point chart displayed the VAFB-SLC launchsite hadno rockets launched for heavypayload mass(greater than 10000)
  - Used a bar chart to find which orbits have high success rates
  - Used a scatter plot to success relationships between FlightNumber vs. Orbit types and Payload vs. Orbit
  - Used a line chart with x axis to be Year and y axis to be average success rate, to get the average launch success trend
  - GitHub URL of completed EDA with data visualization notebook: https://github.com/cradke58/This_Should_be_Good/blob/main/jupyter-labs-eda-dataviz.ipynb

EDA (Exploratory Data Analysis) with SQL

SQL queries performed:

- Displayed the names of the launch sites using the “Select” command
- Displayed 5 records where launch sites begin with 'CCA’ using - Limit 5
- Displayed the total payload mass launched by NASA (CRS) using - Sum()
- Displayed average payload mass carried by F9 v1.1 using – Avg()
- Listed the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000 using - Select, Where
- Listed the total number of successful and failure mission outcomes using – Count, Group By
- Listed the names of the booster versions which have carried the maximum payload mass using a subquery
- Listed the records displaying the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015
- Ranked the count of landing outcomes between the date 2010-06-04 and 2017-03-20, in descending order using – Select, Order By
- GitHub URL of your completed EDA with SQL notebook: https://github.com/cradke58/This_Should_be_Good/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

Built an interactive map :

- Created a folium Map object and added the following
- folium.Circle and folium.Marker for each launch site to create a visual of locations
- Marker clusters for success/fail to simplify the map containing many markers having the same coordinates and easily identify which launch sites have relatively high success rates
- MousePosition on the map to get coordinate for a mouse over a point on the map
- PolyLine with distance between a launch site to the selected points to show proximity
- GitHub URL of completed interactive map with Folium map: https://github.com/cradke58/This_Should_be_Good/blob/main/lab_jupyter_launch_site_locations.ipynb

Build a Dashboard with Plotly Dash

Summary of plots/graphs and interactions added to a dashboard:

- Created a dropdown list to enable Launch Site selection
- Placed a pie chart to show the total successful launches count for all sites. If a specific launch site is selected, shows the Success vs. Failed counts for the selected site
- Inserted a slider to select payload range
- Developed a scatter chart showing correlation between payload and launch success
- GitHub URL of completed Plotly Dash Dashboard Build: https://github.com/cradke58/This_Should_be_Good/blob/main/Plotly_Dashboard.ipynb

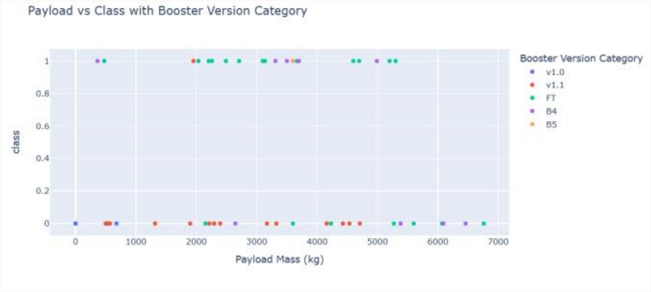
Predictive Analysis (Classification)

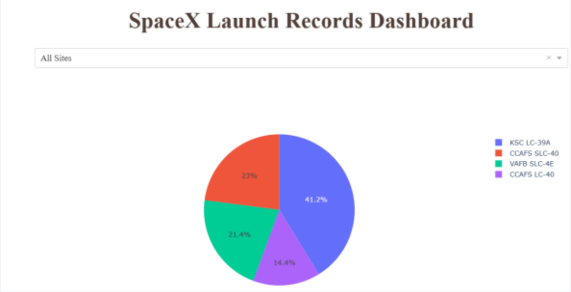
- Summarize how you built, evaluated, improved, and found the best performing classification model
- Flow Steps: Performed exploratory Data Analysis and determined Training Labels -> Created a column for “class” -> Standardized the data -> Split into training data and test data -> Found best Hyperparameter for SVM, Classification Trees, K Nearest Neighbor and Logistic Regression
- Calculated the accuracy on the test data using the method score ()
- Logistic Regression performs the best
- Find the method performs best using test data GitHub URL of completed predictive analysis: https://github.com/cradke58/This_Should_be_Good/blob/main/Predictions_Macine_Learning.ipynb

Results

- Exploratory data analysis results: can observe that the success rate since 2013 kept increasing until 2020.

- Interactive analytics dashboard show success rates for payloads and launch sites:





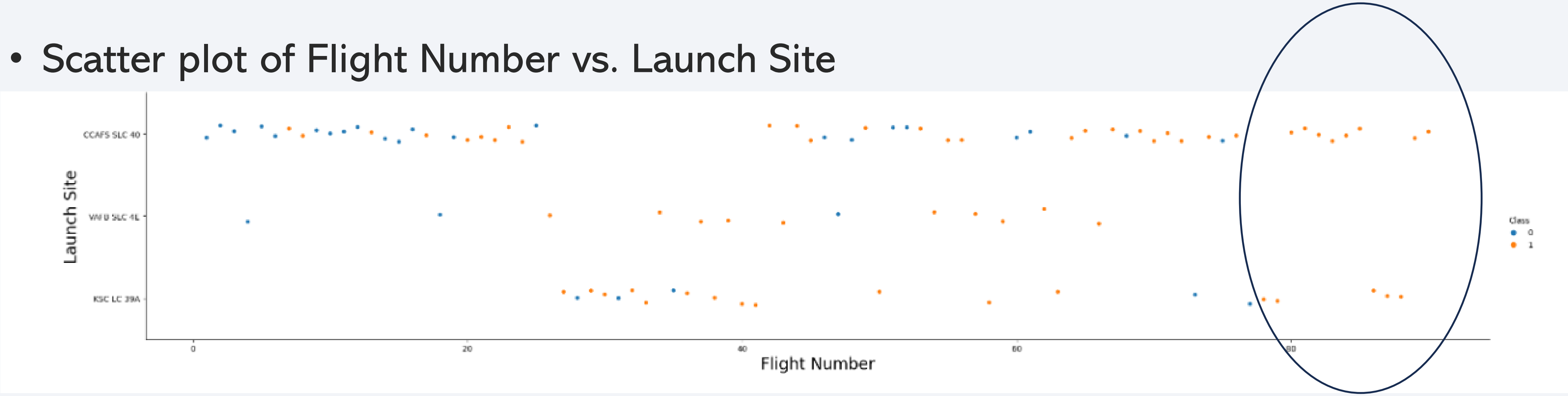
- Predictive analysis results : Logistic Regression performs the best

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

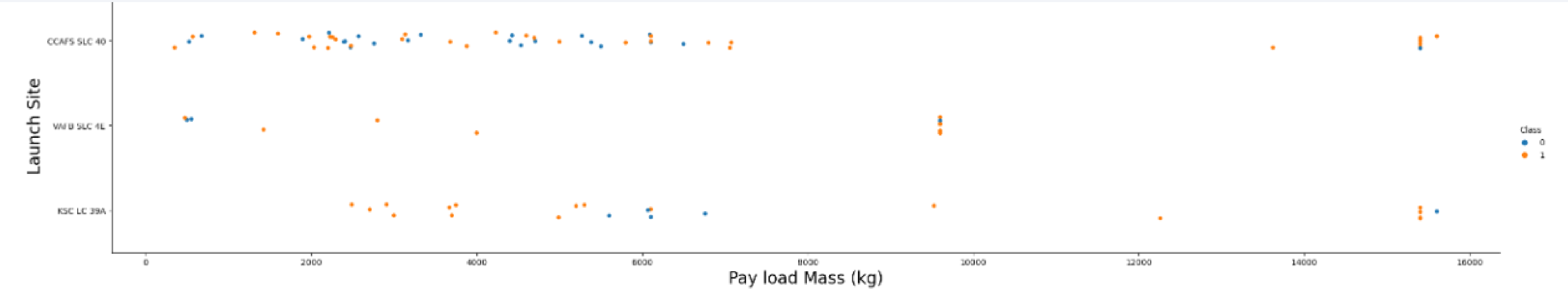
- Scatter plot of Flight Number vs. Launch Site



This scatter plot show that as the number of the flight increases at all launch sites the success/fail become all success

Payload vs. Launch Site

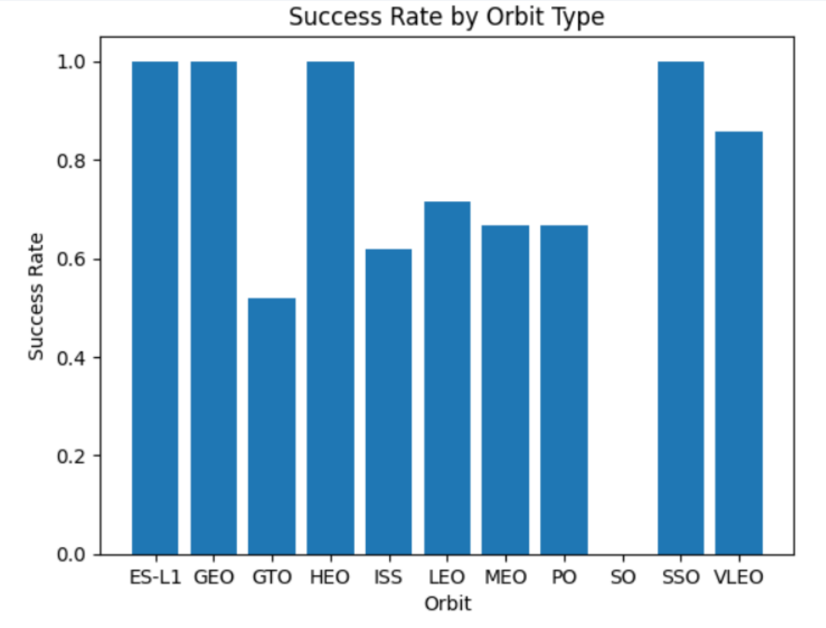
- Scatter plot of Payload vs. Launch Site



Scatter plot shows Launch Site KSC LC-39A launched payloads greater than 2000 kg and the other launch sites launched payloads bot greater than and less than 2000 kg

Success Rate vs. Orbit Type

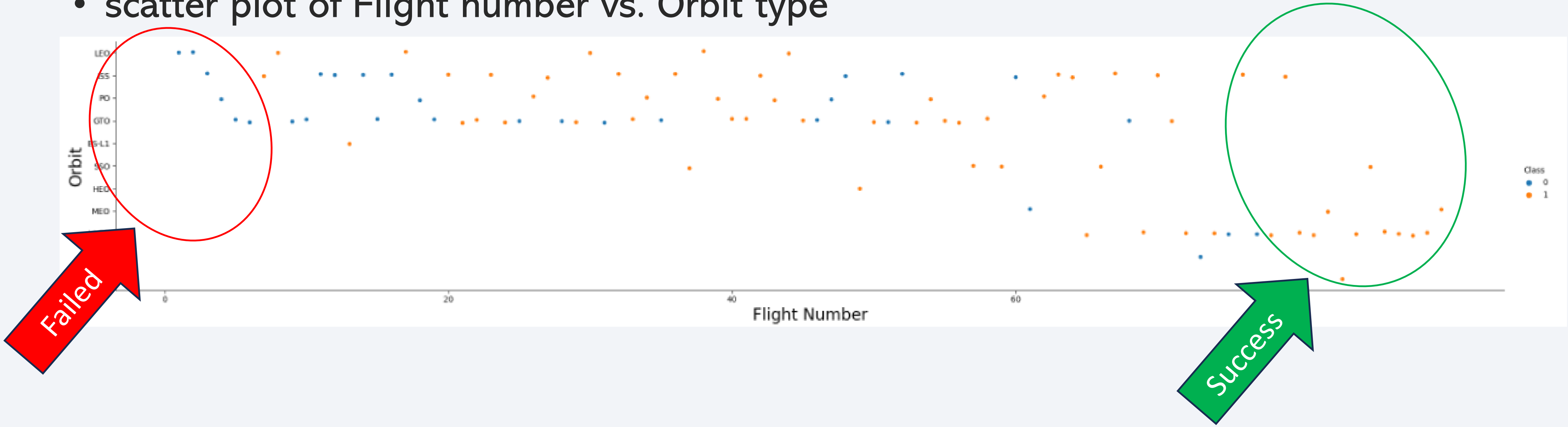
- bar chart for the success rate of each orbit type



This bar chart shows that orbit types: ES-L1, GEO,HEO, and SSO had 100 percent success rates

Flight Number vs. Orbit Type

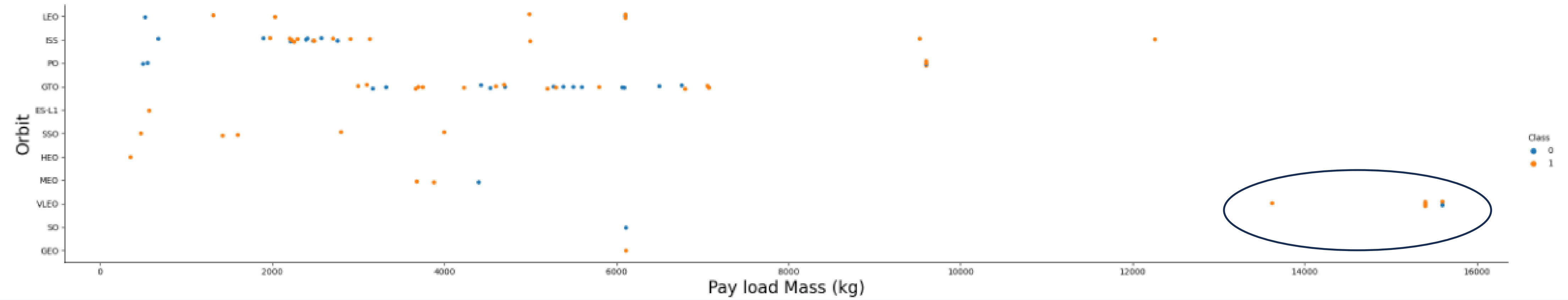
- scatter plot of Flight number vs. Orbit type



This scatter plot shows early flights failed (red oval) and later flights achieved 100 percent success (green oval) independent of the orbit type.

Payload vs. Orbit Type

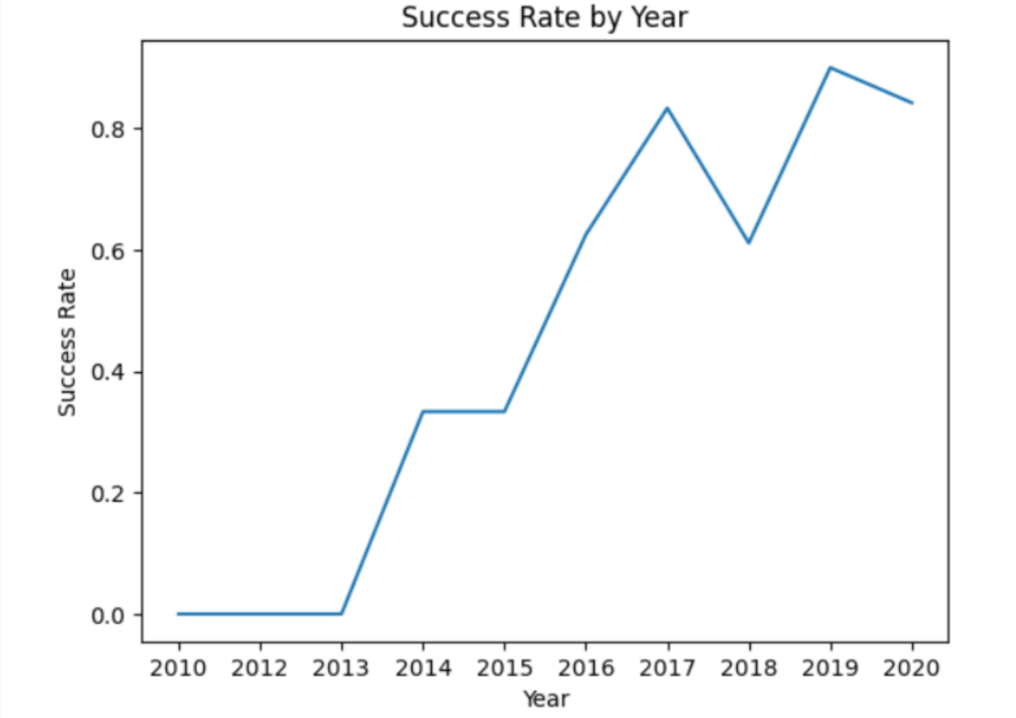
- scatter point of payload vs. orbit type



This scatter plot show payloads greater than 13000 kg were at lower orbits and achieved an increased success rate.

Launch Success Yearly Trend

- line chart of yearly average success rate



This line chart show as the years moved into later times the success rates trended toward and increase

All Launch Site Names

- names of the unique launch sites

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

This is a table of the launch site names

Launch Site Names Begin with 'CCA'

- 5 records where launch sites begin with `CCA`

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

This is a table of 5 launch sites whose launch site begins with ‘CCA’

Total Payload Mass

- calculated total payload carried by boosters from NASA (CRS)

Total_Payload_Mass
45596

This is the calculated total payload carried by boosters from NASA (CRS)

Average Payload Mass by F9 v1.1

- calculated average payload mass carried by booster version F9 v1.1

Average_Payload_Mass
2928.4

This is the calculated average payload mass carried by booster version F9 v1.1

First Successful Ground Landing Date

- date of the first successful landing outcome on ground pad

First_Successful_Landing
2010-04-06

This is the date of the first successful landing outcome on ground pad

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

Landing_Outcome	Booster_Version
Success (drone ship)	F9 FT B1022
Success (drone ship)	F9 FT B1026
Success (drone ship)	F9 FT B1021.2
Success (drone ship)	F9 FT B1031.2

These are the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

Total Number of Successful and Failure Mission Outcomes

- total number of successful and failure mission outcomes

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

This is a table of the total number of successful and failure mission outcomes

Boosters Carried Maximum Payload

- names of the booster which have carried the maximum payload mass

PAYLOAD_MASS__KG_	Booster_Version
15600	F9 B5 B1048.4
15600	F9 B5 B1049.4
15600	F9 B5 B1051.3
15600	F9 B5 B1056.4
15600	F9 B5 B1048.5
15600	F9 B5 B1051.4
15600	F9 B5 B1049.5
15600	F9 B5 B1060.2
15600	F9 B5 B1058.3
15600	F9 B5 B1051.6
15600	F9 B5 B1060.3
15600	F9 B5 B1049.7

This is a list of the names of the booster which have carried the maximum payload mass

2015 Launch Records

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

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

This is a list of failed landing outcomes in drone ship, their booster versions, and launch site names for the year 2015

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

- ranked 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 (ground pad)	5
Success (drone ship)	5
Failure (drone ship)	5
Controlled (ocean)	3
Uncontrolled (ocean)	2
Precluded (drone ship)	1
Failure (parachute)	1

This is a ranked 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

Section 3

Launch Sites Proximities Analysis

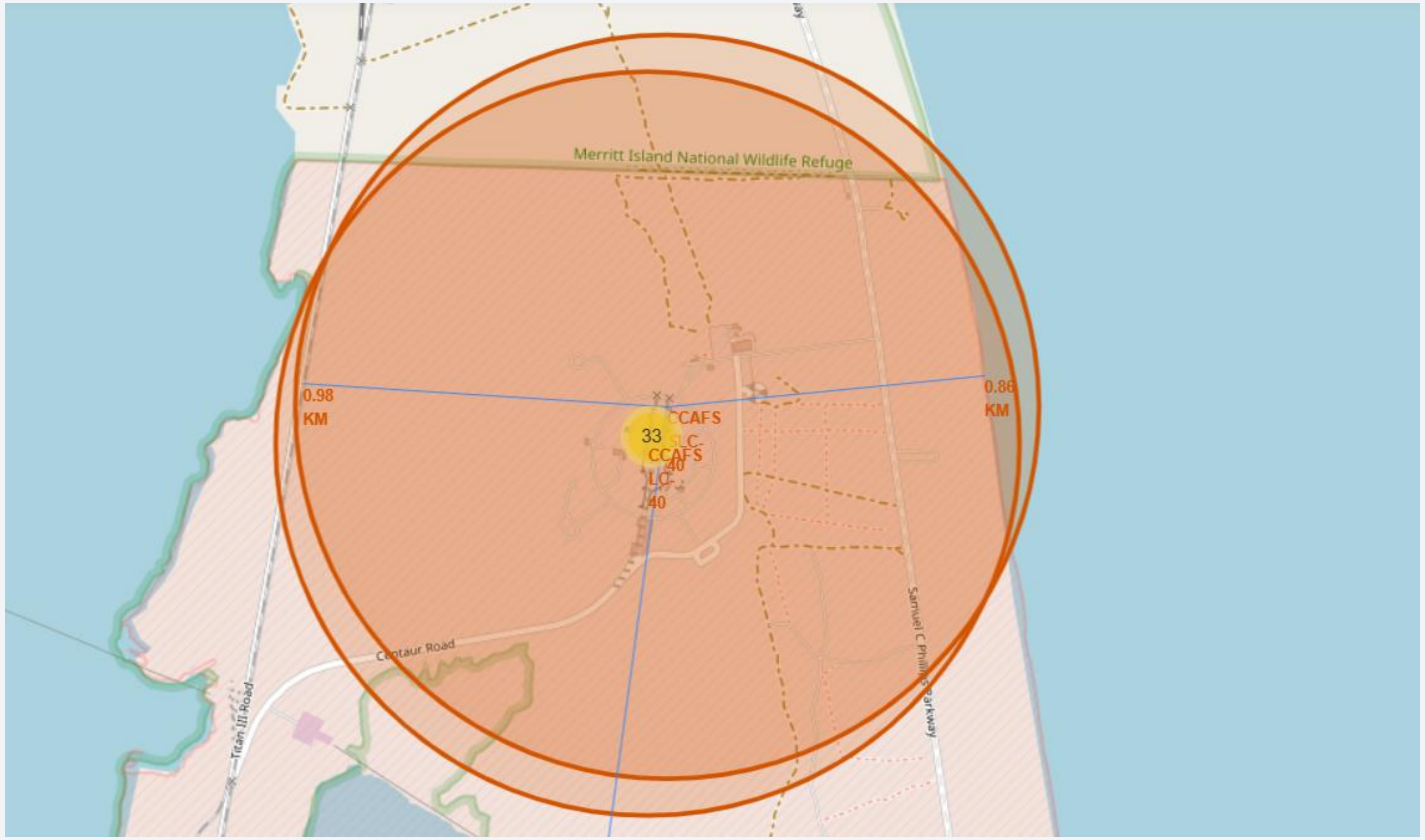
Launch Site Locations

This map shows the labelled launch site locations which are near the East and West ocean coasts

Launch Site with Color Coded Success Markers

There is a capability to drill down on launch sites and determine success rates based on color coded markers. Fewer green markers indicate low success rate.

Launch Site Proximity to Features (Coast, Rail, City)

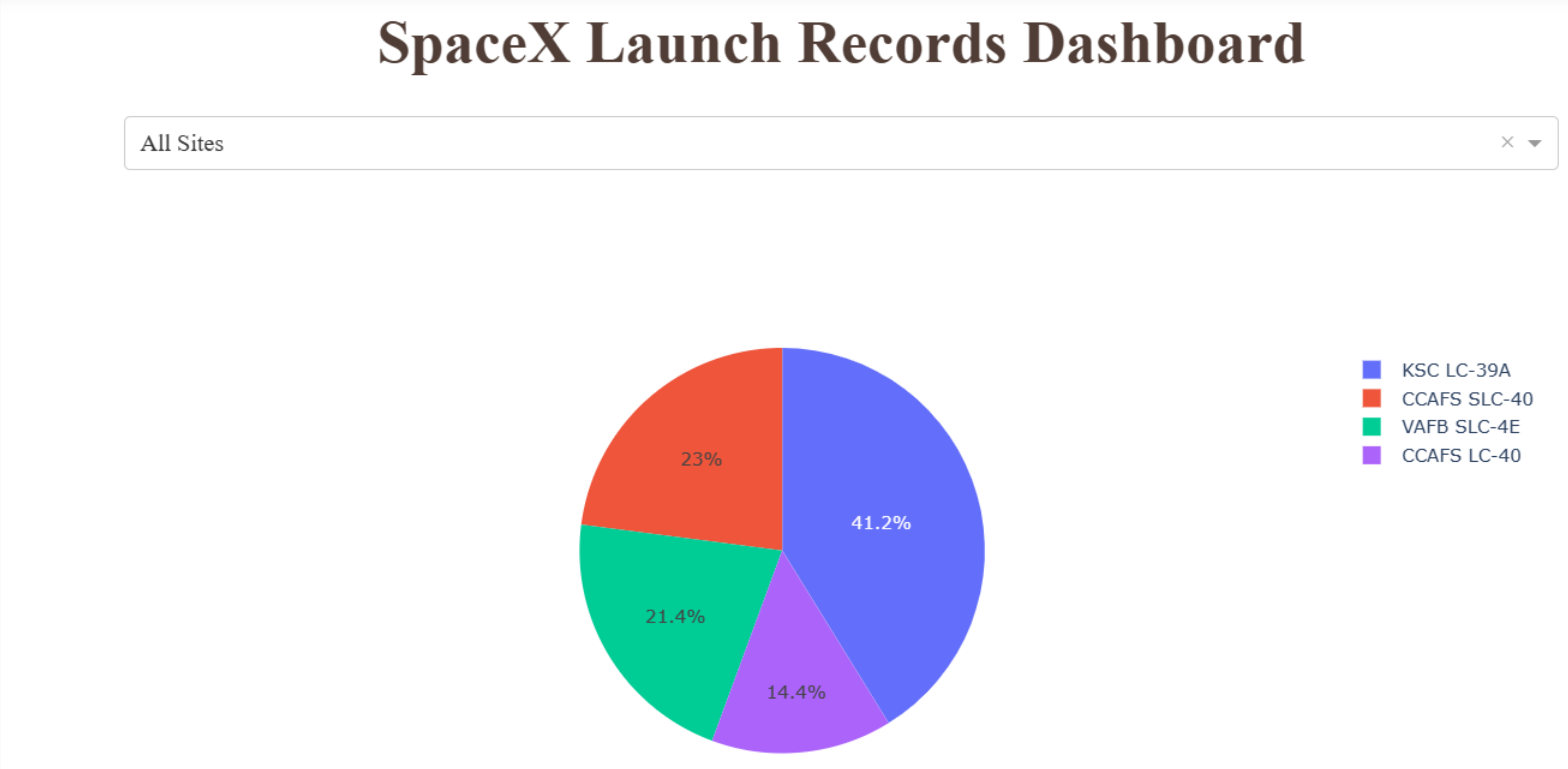


This map view shows the proximity the Launch Site has to the coast, railroad, and city.

Section 4

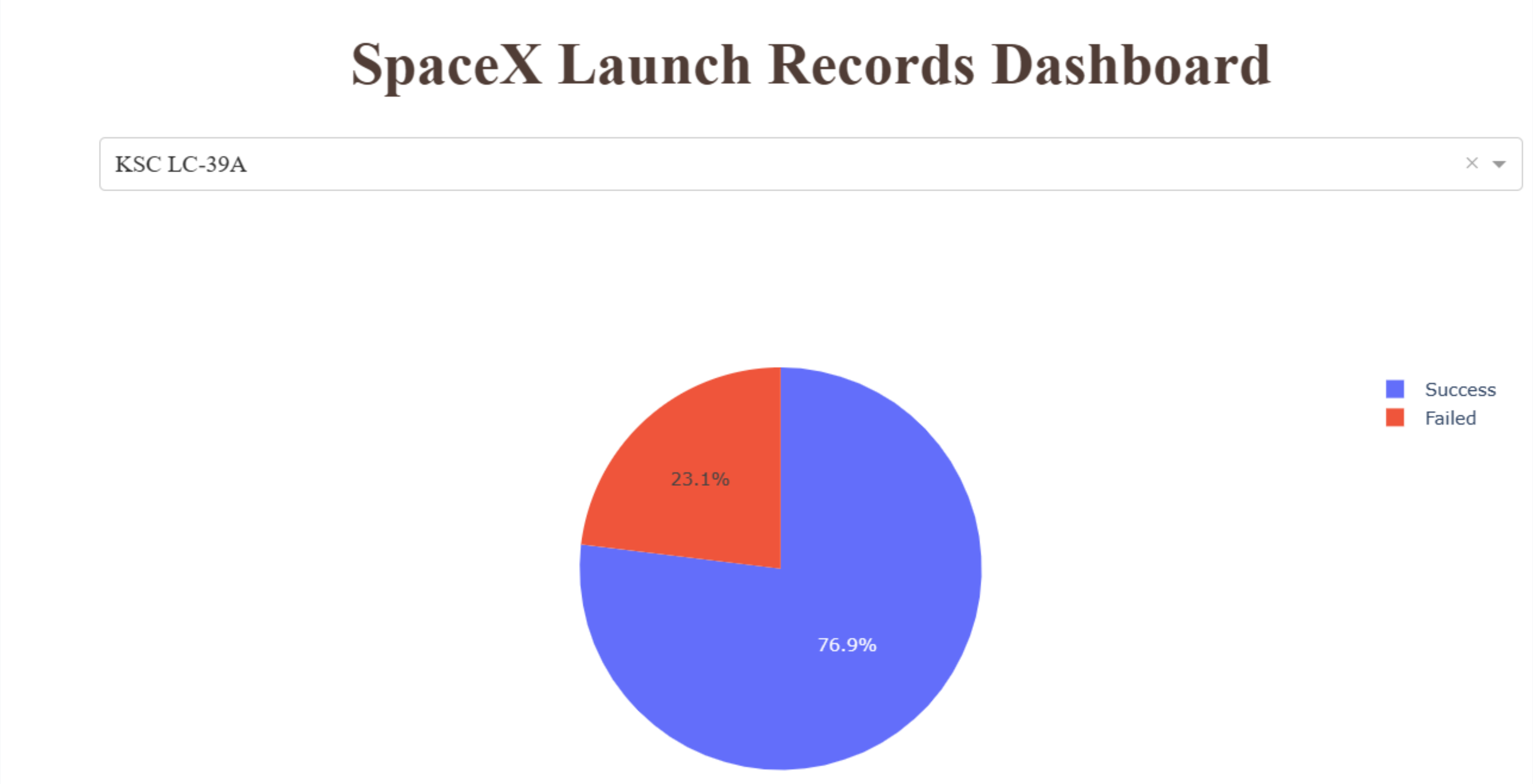
Build a Dashboard with Plotly Dash

Dashboard with ALL Launch Site Success Rates %



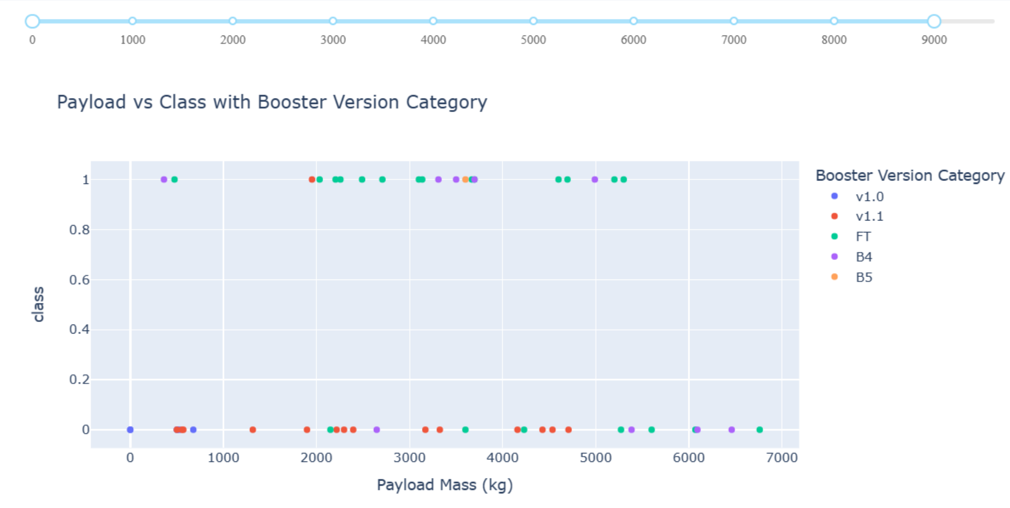
Dashboard view of all launch site success rates. Blue slice of KSC LC-39A has the highest (41.2 %) portion of success rates.

Launch Site with Highest Success Rate



Pie chart shows the launch site (KSC LC-39A) with the highest success rate of 76.9%

Slider Helps Visualize Payload and Booster Success



Two graphics, side by side, show all success and failures on left graphic. Graphic on the right clearly shows two booster versions with payloads less than 5000 kg have high or low success rates.

Section 5

Predictive Analysis (Classification)

Classification Accuracy

Find the method performs best:

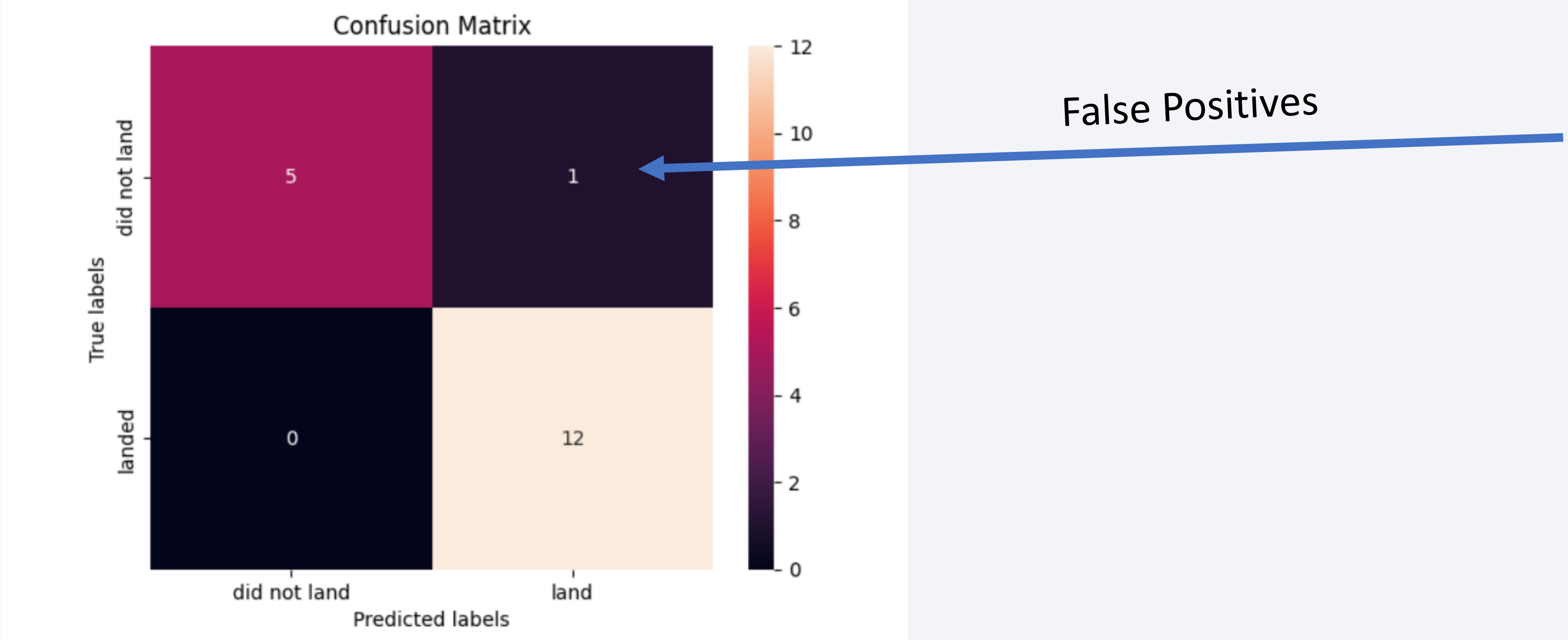
```python
# Select the method with the highest accuracy as the best-performing method
best_method = max(svm_accuracy, trees_accuracy, logreg_accuracy, knn_accuracy)

# Print the best-performing method
if best_method == svm_accuracy:
    print("SVM performs the best.")
elif best_method == trees_accuracy:
    print("Classification Trees performs the best.")
elif best_method == logreg_accuracy:
    print("Logistic Regression performs the best.")
else:
    print("K Nearest Neighbors performs the best.")
```

```
Logistic Regression performs the best.
```

The Logistics Regression Model performs the best

Confusion Matrix

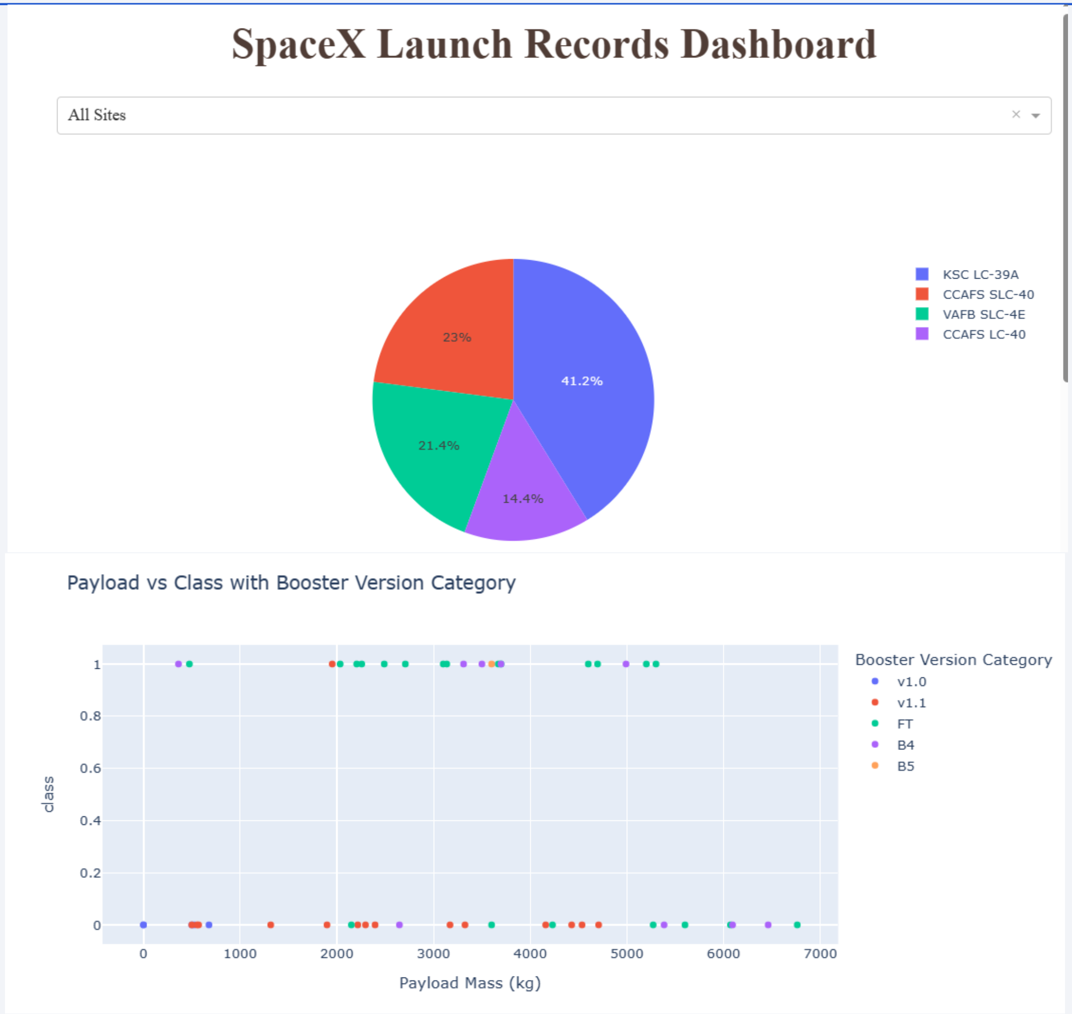
Logistic Regression Confusion Matrix



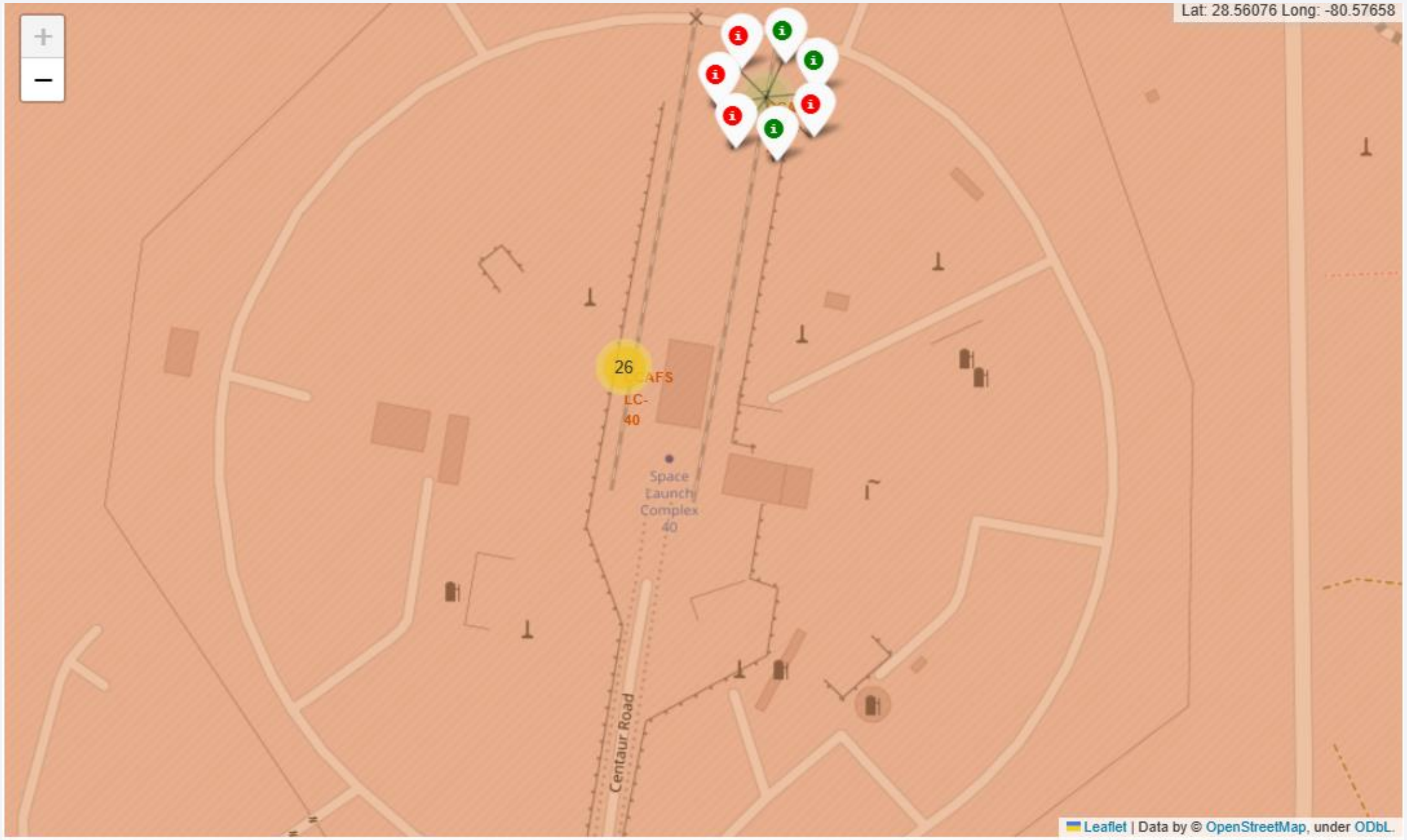
Examining the confusion matrix, we see that logistic regression model performed the best and can distinguish between the different classes. We see that the major problem is false positives.

Conclusions

- The Extract, Transform, Load sequence on SpaceX launch data resulted in a Dashboard that facilitates further analysis. By manipulating the controls on the dashboard launch site success rates can be explored and compared. Payloads with different booster versions can be explored with success visuals to the finest detail. Much can be determined using this dashboard of SpaceX launches.



Appendix



When presenting the analysis of SpaceX success and failure, geographic visuals are at hand to provide a spacial representation of previous launches. This data can be presented to a machine learning model and predictions made of future successes based on launch site, payload, booster types, and captured parameters.

Thank you!