IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Benjamin Jonasson
2023-12-19

Outline

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

Executive Summary

- Methodologies used in this report
  - Data collection
  - Data wrangling
  - Exploratory data analysis
  - Data visualization
  - Predictive analysis
  - Financial forecasting
- Results
  - With the use of machine learning I was able to predict the expected cost of a SpaceX Falcon 9 launch.

Introduction

- SpaceX advertises Falcon 9 rocket launches on its website with a cost of $62 million; other providers cost upward of $165 million each, much of the savings is because SpaceX can reuse the first stage.
- Since SpaceX Falcon 9 launches do not have a 100% success rate. I will use machine learning to predict the expected outcome of any given launch and use that to calculate the real expected cost.

Section 1

Methodology

Methodology

Executive Summary

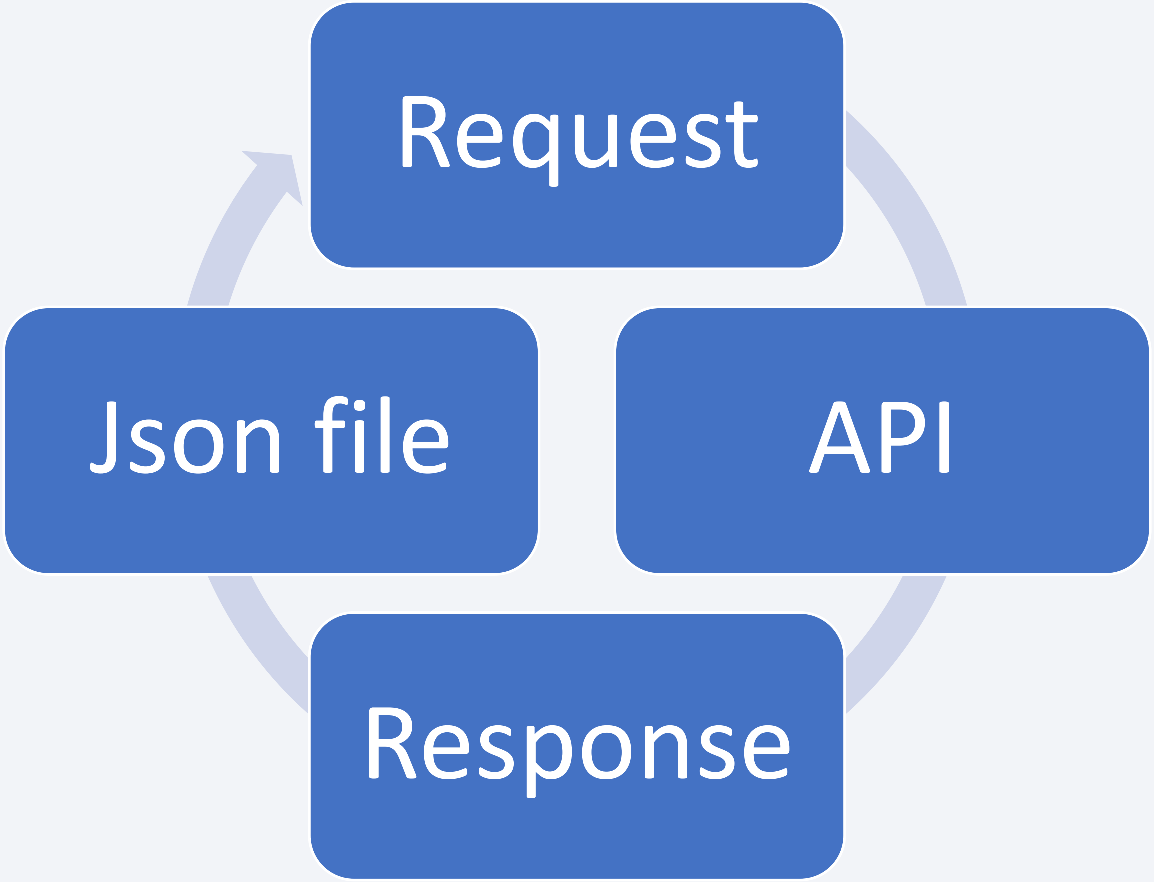
- Data collection methodology:
    - Request and parse the SpaceX launch data using the GET request
    - Filter the dataframe to only include Falcon 9 launches
- Perform data wrangling
    - Dealing with Missing Values
    - Calculate the number of launches on each site
    - Calculate the number and occurrence of each orbit
    - Calculate the number and occurrence of mission outcome of the orbits
    - Create a landing outcome label from Outcome column
- 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

- Request and parse the SpaceX launch data using the GET request
- Filter the dataframe to only include Falcon 9 launches

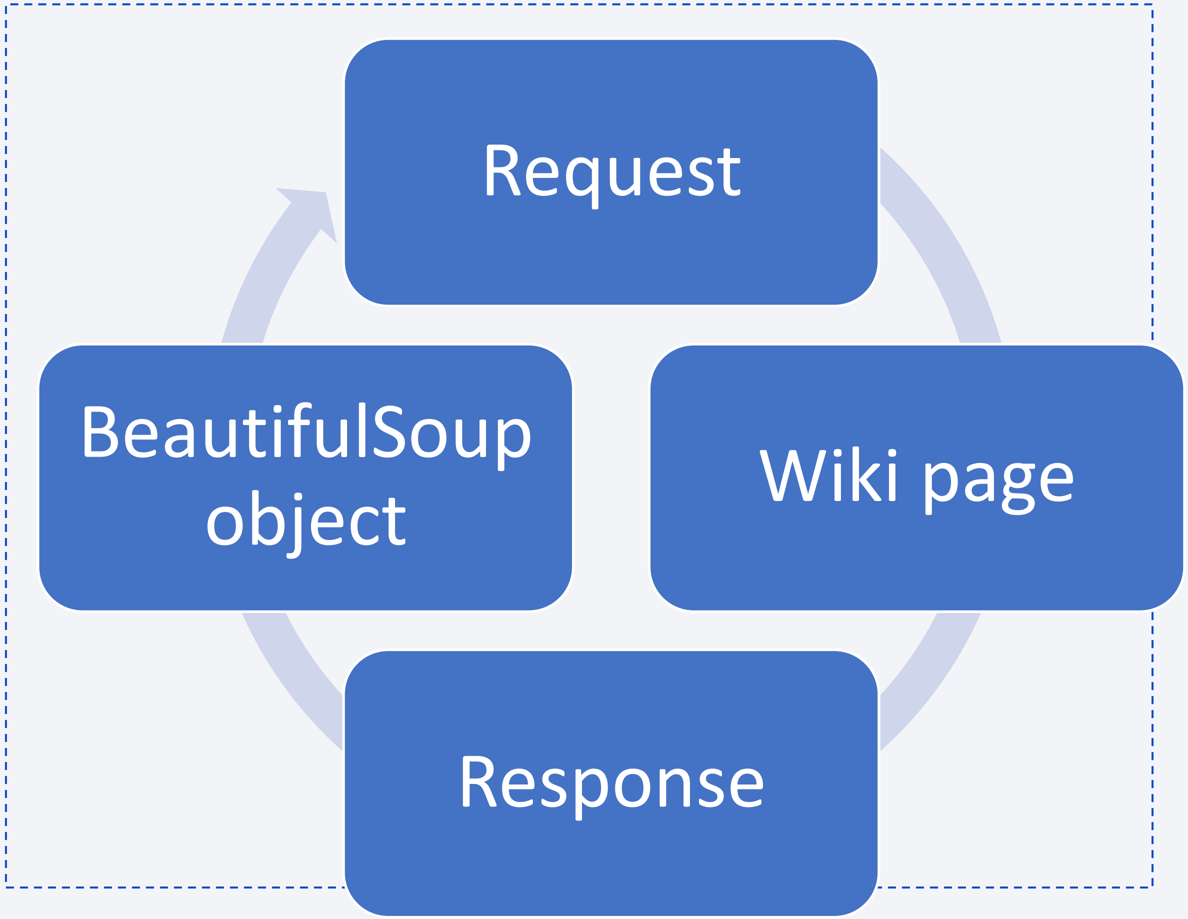
Data Collection – SpaceX API

- requests.get() method was used to make HTTP requests from the SpaceX API.
- Normalize json file and put into dataframe.
- https://github.com/bennxd/space_y/blob/main/jupyter-labs-SpaceX-data-collection-api.ipynb



Data Collection - Scraping

- Request response from wiki web page
- Use BeatifulSoup to parse the html content.
- Extract table from content and put into a dataframe.
- https://github.com/bennxd/space_y/blob/main/jupyter-labs-webscraping.ipynb



Data Wrangling

- Dealing with Missing Values
- 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

- https://github.com/bennxd/space_y/blob/main/labs-jupyter-SpaceX-Data%20wrangling.ipynb

EDA with Data Visualization

- Created scatter plots to see correlation between success rate, flight number, payload mass, launch site, orbit etc. to understand which features correlate with successful launches.
- https://github.com/bennxd/space_y/blob/main/jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb

EDA with SQL

- Distinct launch sites
- Launch sites like ‘CCA’
- Total payload mass for NASA customer
- Average payload mass for booster like F9 v1.1
- First occurrence of a successful launch
- Boosters between 4000-6000 kg
- Count of outcomes
- Boosters that carried the max payload mass
- Specific record in 2015 based on outcome, booster version and launch site
- Count of outcomes ranked in descending order
- https://github.com/bennxd/space_y/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Created markers and circles for different launch sites and their launch outcomes as well as lines between launch sites and their distance to proximities like railway, highway, coastline and city.
- https://github.com/bennxd/space_y/blob/main/dash.PNG

Build a Dashboard with Plotly Dash

- Made an interactive dashboard with a pie chart showing share of successful launches if all sites were selected, and the success rate if a specific site was selected. In addition, I made a scatter plot showing payload mass vs launch success with a payload mass slider.
- https://github.com/bennxd/space_y/blob/main/SpaceX_dash_app.py

Predictive Analysis (Classification)

- Loaded feature data into dataframe
- Created an array for the y values
- Standardized the data
- Split data into train and test set
- Defined parameters for different models (logreg, svm, tree, knn)
- Used GridSearchCV to get the optimal parameter values for the models
- Trained each model on the training data
- Visualized the performance of each model with confusion matrix
- Analyzed the results to evaluate which model was best fit to use.
- https://github.com/bennxd/space_y/blob/main/SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

Results

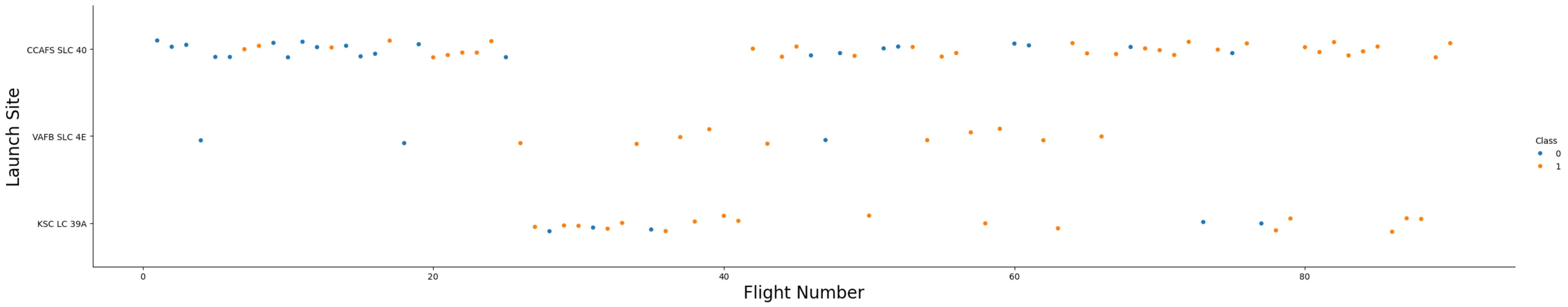
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

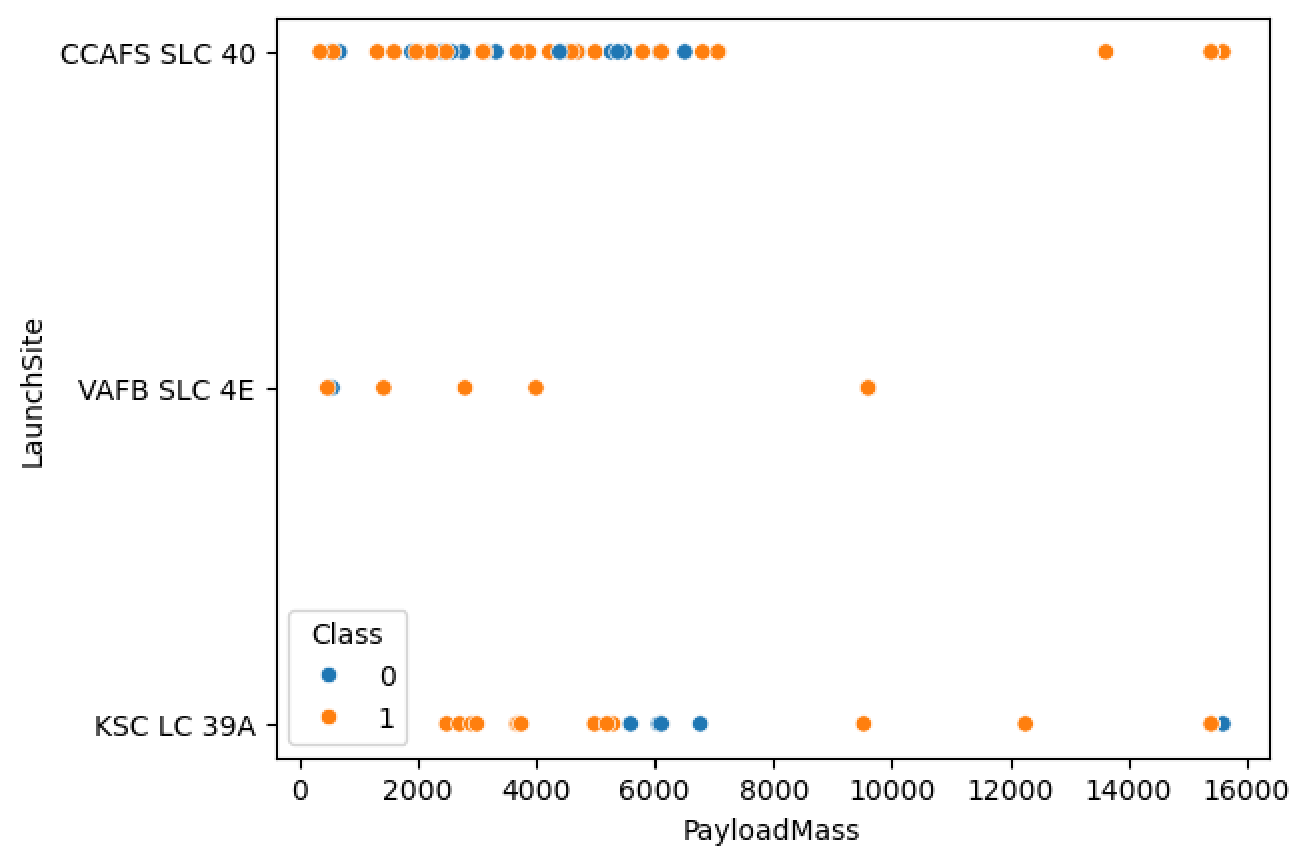
Flight Number vs. Launch Site

- The most used launch site CCAFS SLC 40 has the highest success rate. And its success rate is only increasing.



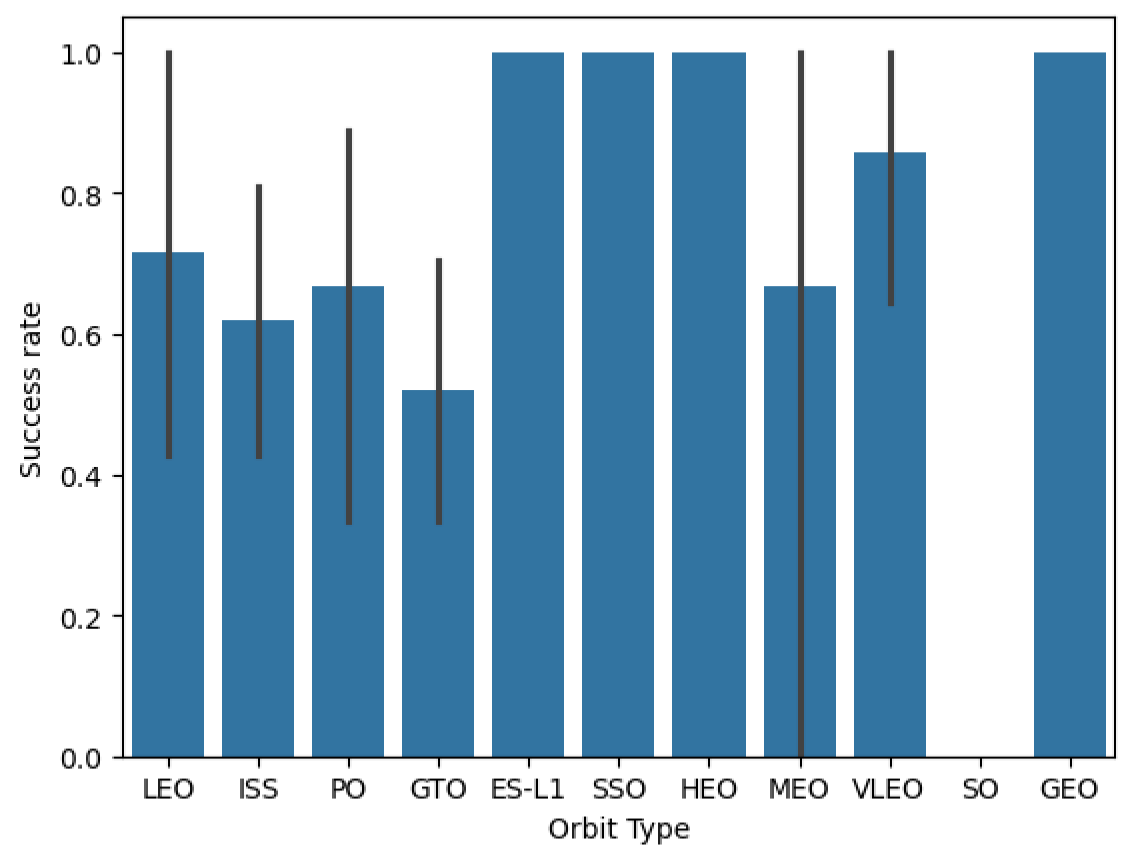
Payload vs. Launch Site

- CCAFS SLC 40 is not used for payloads with a mass in the 8000-13000kg range.
- Payloads over 7000kg have a high success rate regardless of launch site being used.



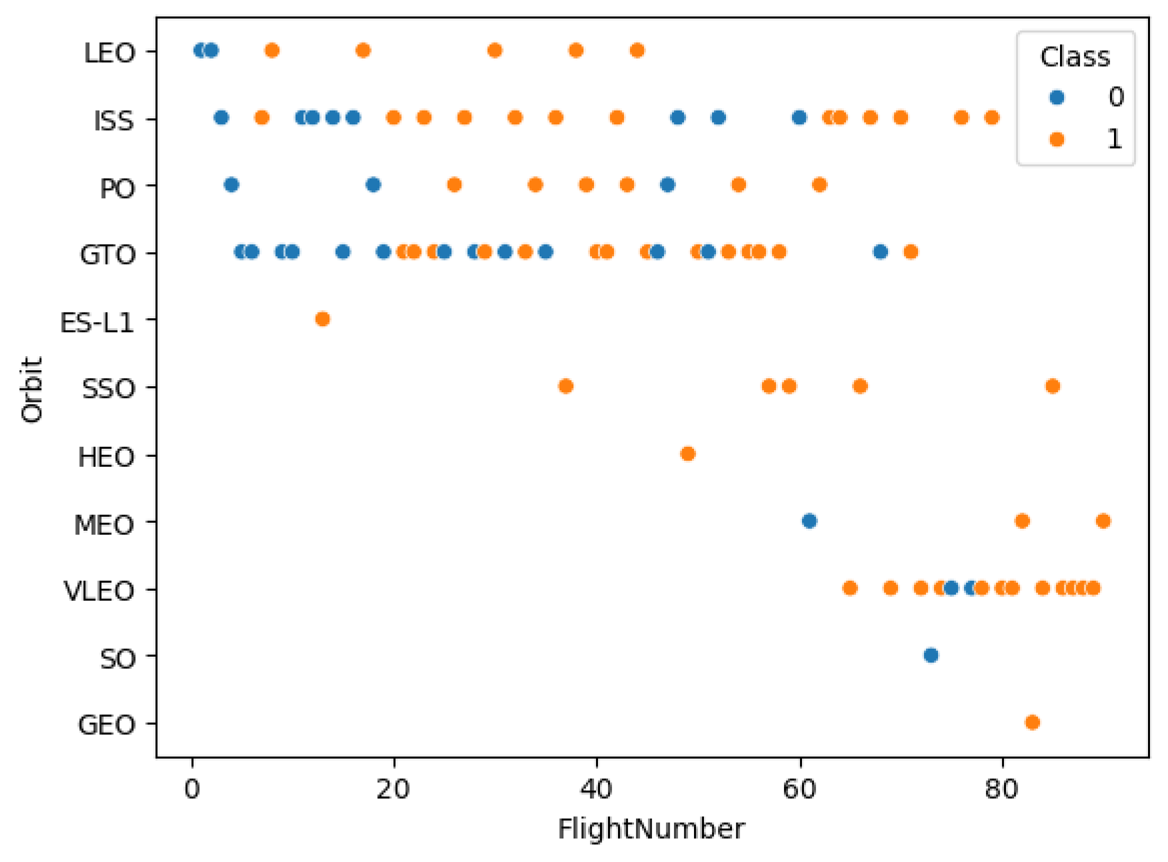
Success Rate vs. Orbit Type

- Orbit types ES-L1, SSO, HEO and GEO have a perfect record.
- SO has never succeeded.
- GTO is a coin toss whether it succeeds or not.
- VLEO has a high success rate.



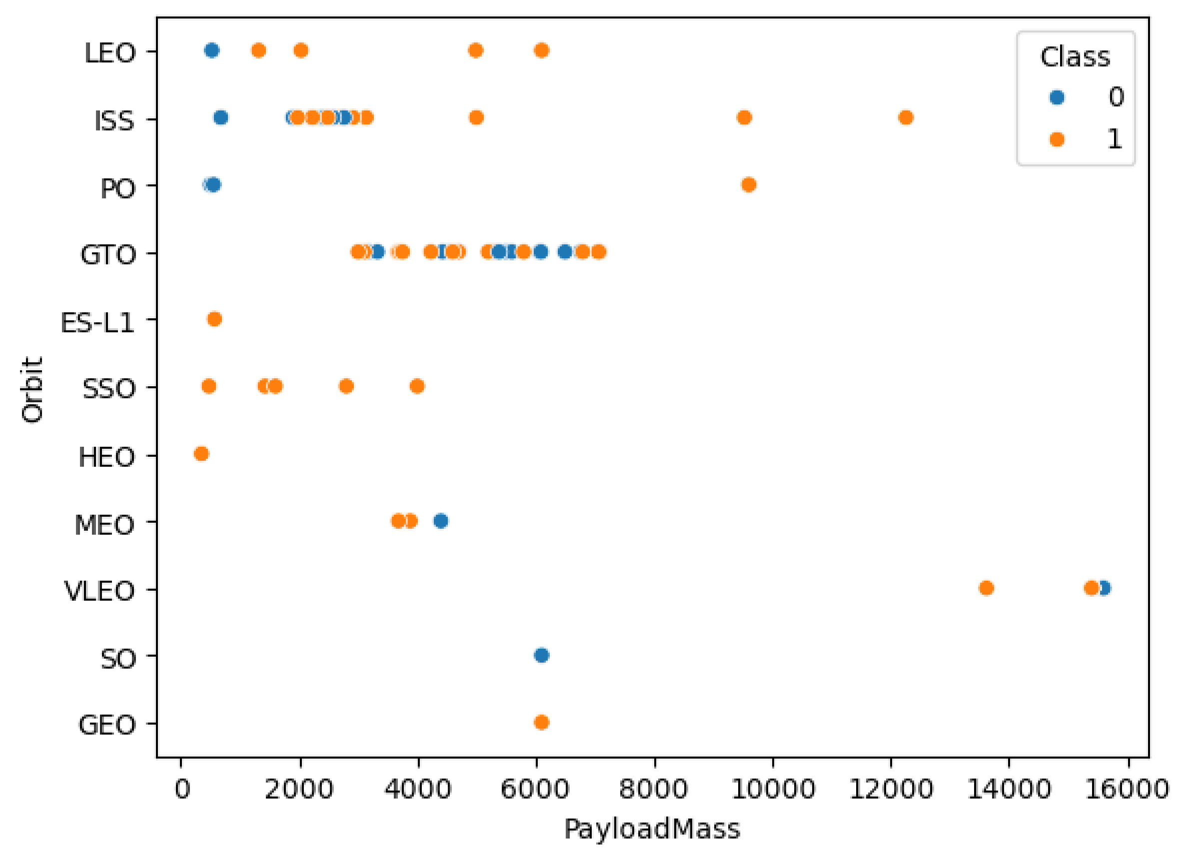
Flight Number vs. Orbit Type

- Success rate is generally low for the early flights, and increasing over time. This could explain why GTO success rate would be lower than the VLEO flights that have only started occuring more recently.



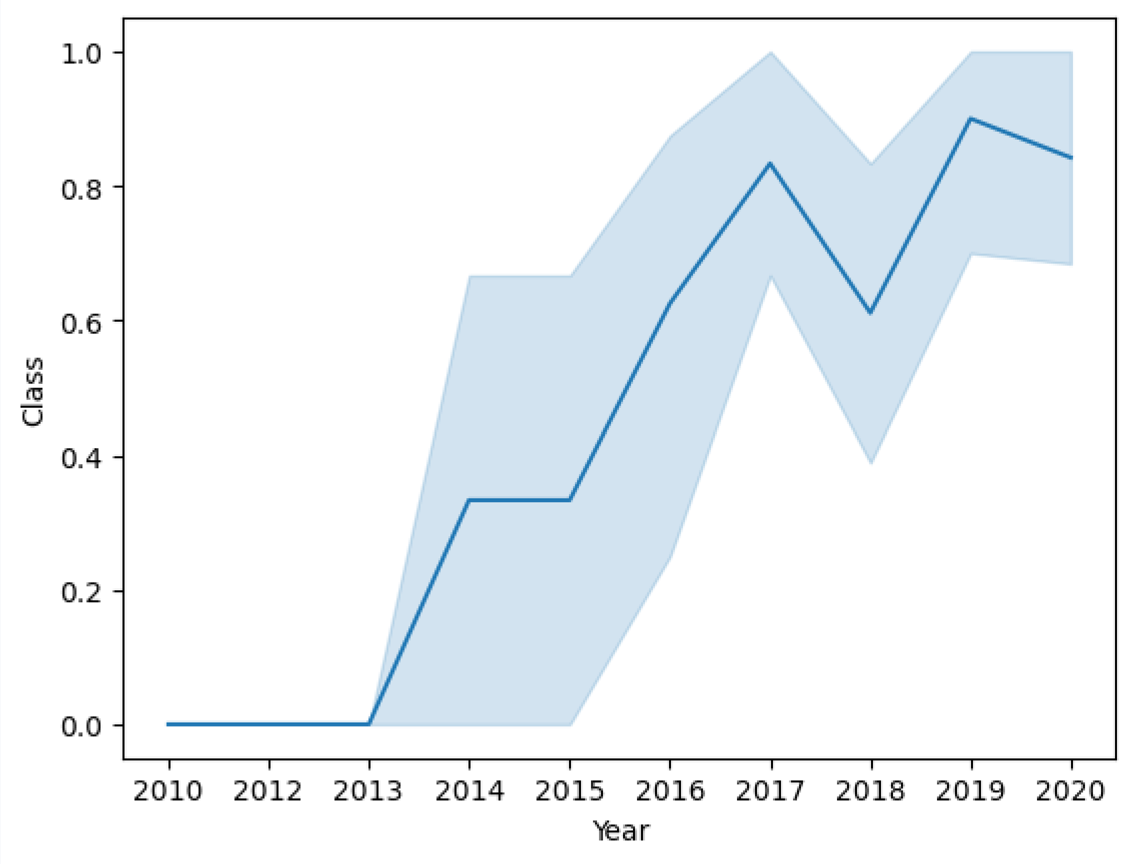
Payload vs. Orbit Type

- Payloads in the 3000-8000 range were mostly sent to GTO.
- Payloads in the 2000-4000 range were mostly sent to ISS.
- The highly successful flights to SSO has been of the lighter type of payload.



Launch Success Yearly Trend

- Here we can see how the success rate increases over the years.



All Launch Site Names

- This table lists all launch sites.

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

Launch Site Names Begin with 'CCA'

- Here are 5 records where the Launch site begins with ‘CCA’. Very interesting…

Launch_Site
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40

Total Payload Mass

- Total payload carried by boosters from NASA:

Customer	total_payload_mass
NASA (CRS)	45596

Average Payload Mass by F9 v1.1

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

Booster_Version	avg_payload_mass
F9 v1.1 B1003	2534.67

First Successful Ground Landing Date

- First successful landing outcome on ground pad.
- This one had all the characteristics of a successful launch.

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2015-12-22	1:29:00	F9 FT B1019	CCAFS LC-40	OG2 Mission 2 11 Orbcomm-OG2 satellites	2034	LEO	Orbcomm	Success	Success (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:

Booster_Version	PAYLOAD_MASS__KG_	Landing_Outcome
F9 FT B1022	4696	Success (drone ship)
F9 FT B1026	4600	Success (drone ship)
F9 FT B1021.2	5300	Success (drone ship)
F9 FT B1031.2	5200	Success (drone ship)

Total Number of Successful and Failure Mission Outcomes

- Total number of launches per outcome:

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

Boosters Carried Maximum Payload

- 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:

year	month	Landing_Outcome	Booster_Version	Launch_Site
2015	01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
2015	04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

- Count per landing outcome between 2010-06-04 and 2017-03-20 in descending order:

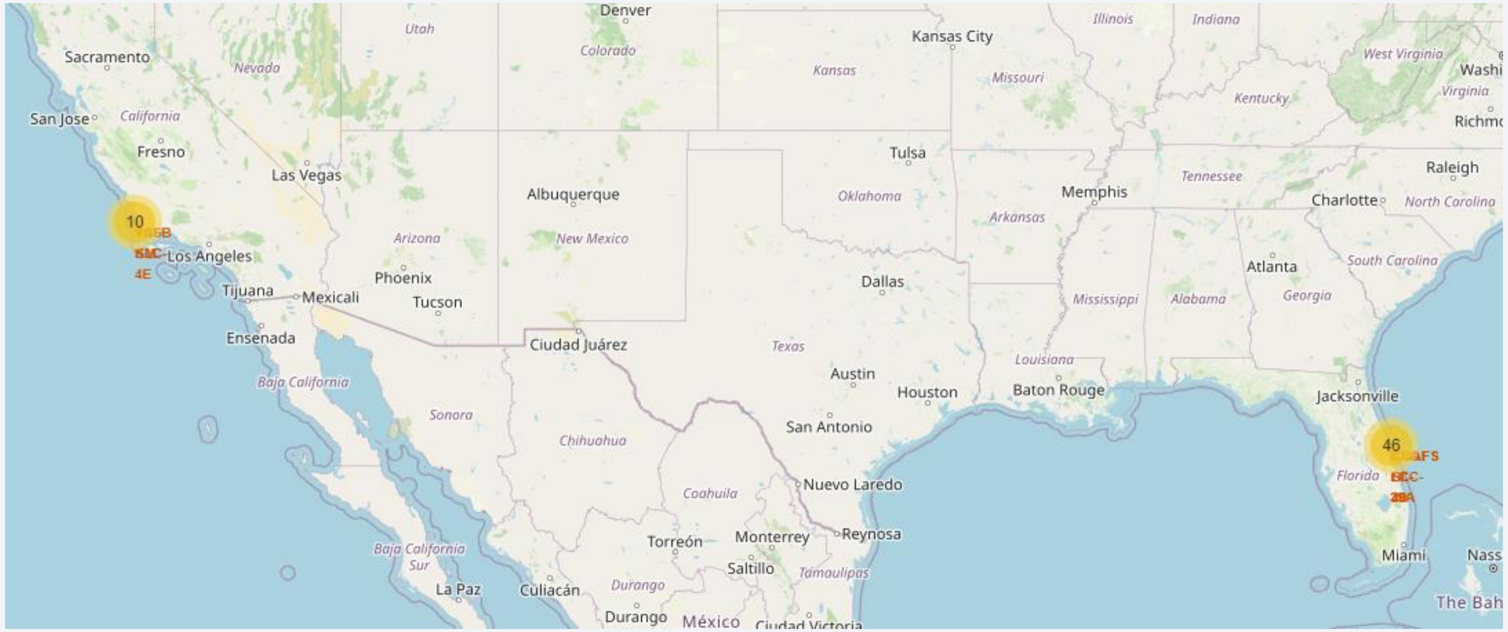
Date	Landing_Outcome	count
2012-05-22	No attempt	10
2016-04-08	Success (drone ship)	5
2015-01-10	Failure (drone ship)	5
2015-12-22	Success (ground pad)	3
2014-04-18	Controlled (ocean)	3
2013-09-29	Uncontrolled (ocean)	2
2010-06-04	Failure (parachute)	2
2015-06-28	Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

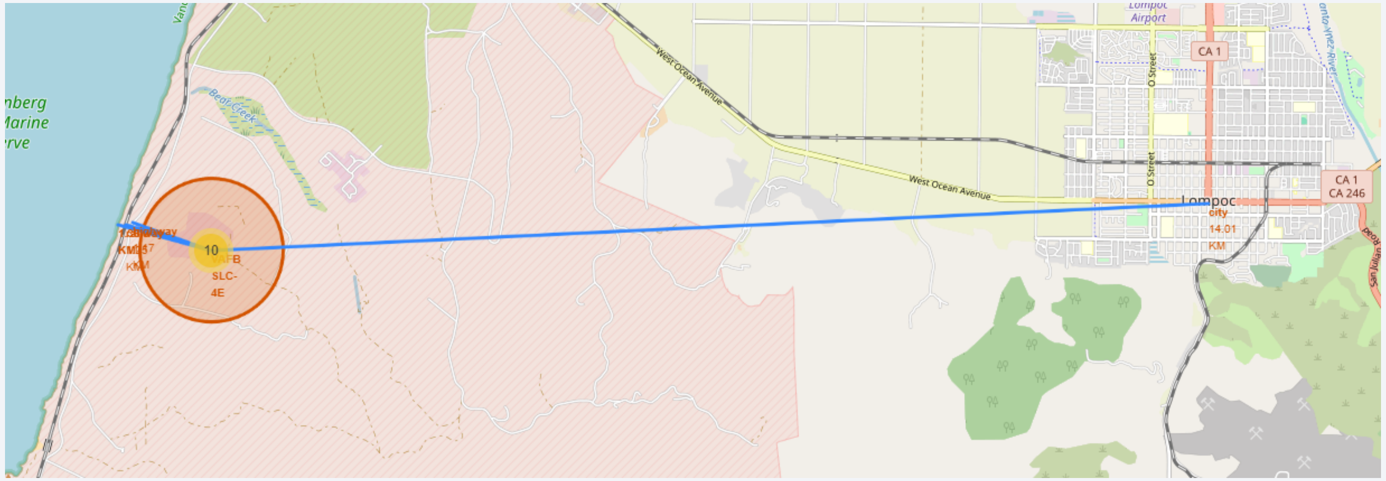
Launch site locations

- Here are the locations of the different launch sites. As can be seen on the map, they are all located close to a coastline.



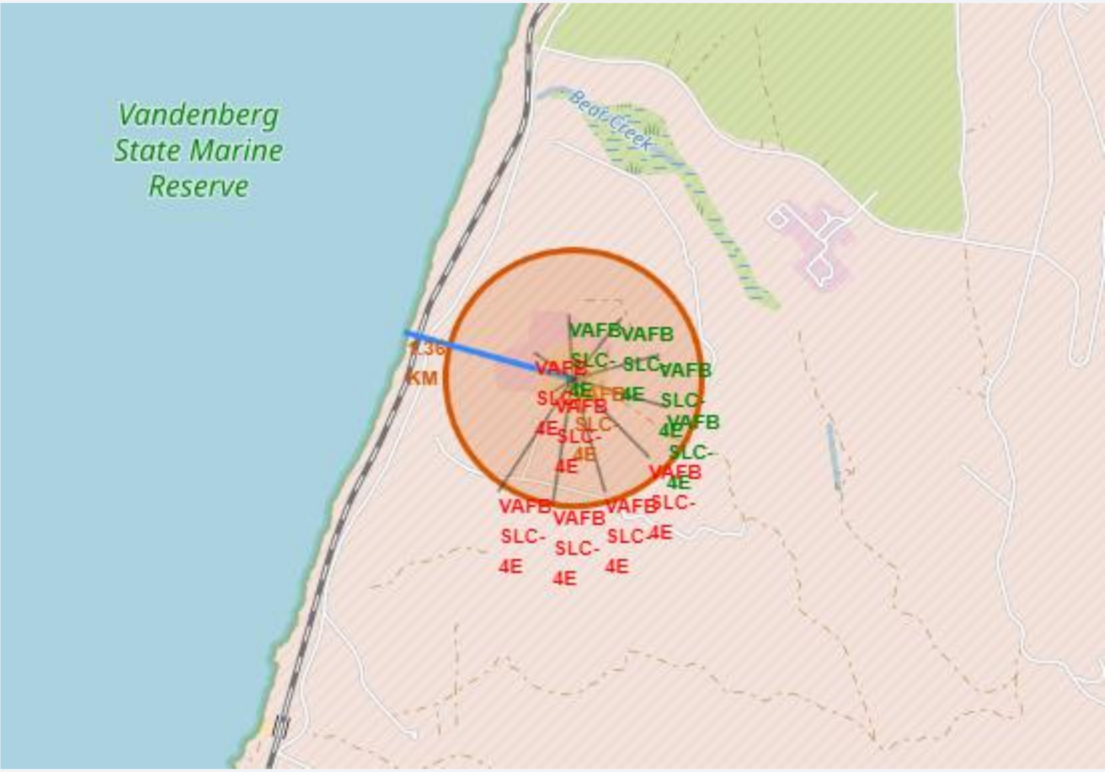
Launch site locations relative to its proximities

- Launch sites are generally located close to a coastline, highway and railway, away from the nearest city, as can be seen in this example of VAFB SLC-4E



Launch outcomes

- Here are the launch outcomes for the VAFB SLC-4E site. Red color indicates a failed launch and green color indicates a successful launch.

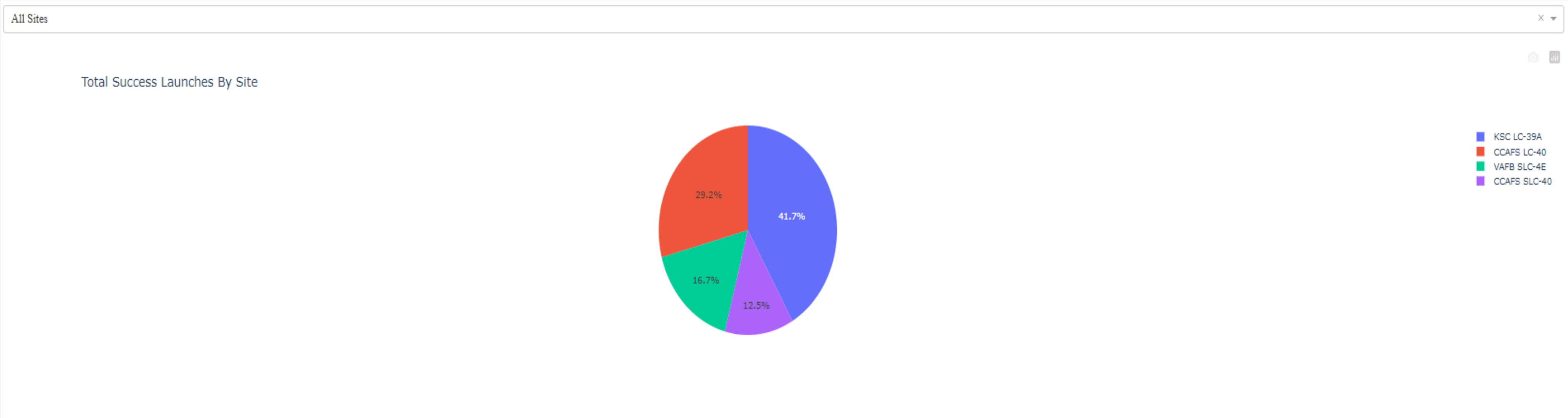


Section 4

Build a Dashboard with Plotly Dash

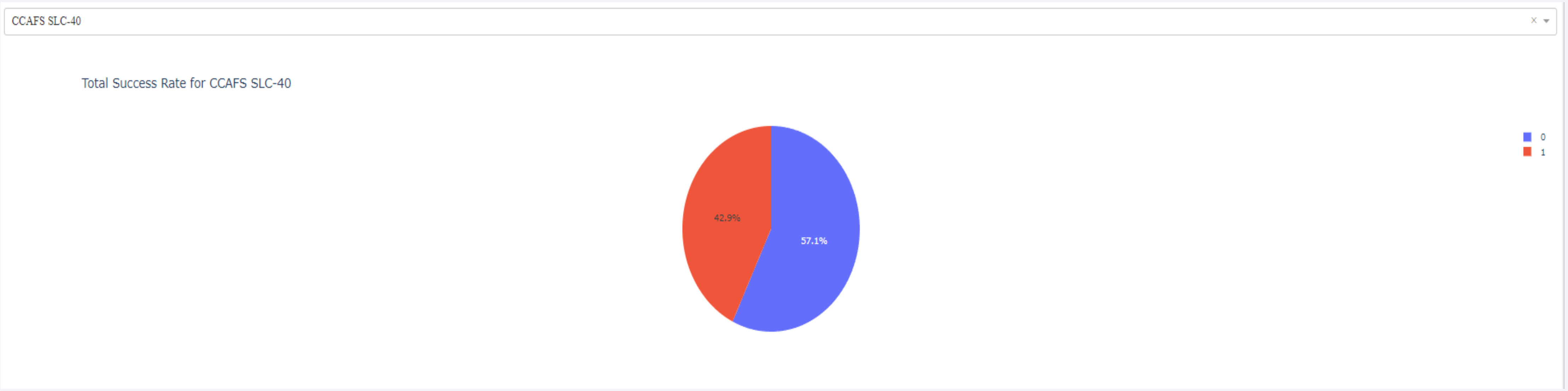
Total successful launches by site

- KSC LC-39A was the launch site with most successful launches. It stood for 41.7% of all successful launches.



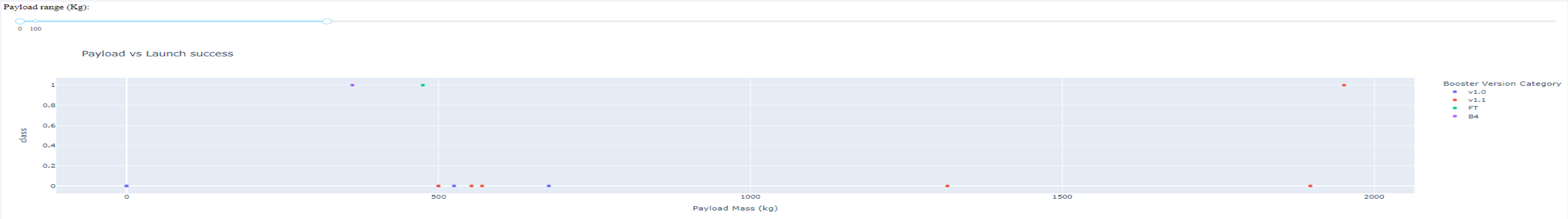
Launch site with highest success rate

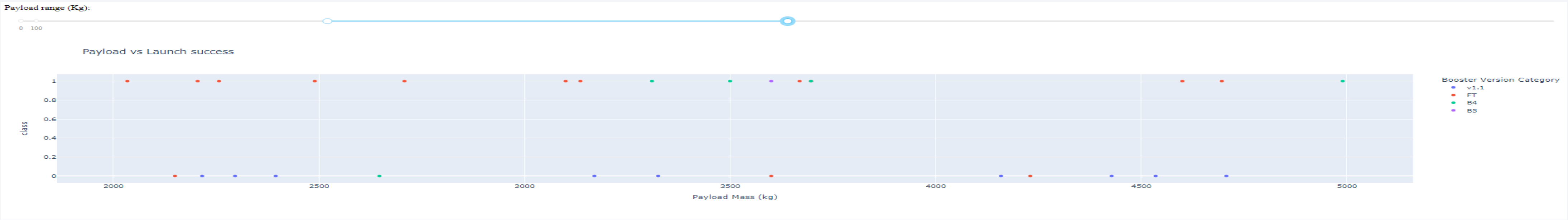
- CCAFS SLC-40 was the launch site with the highest success rate, with it's 42.9% success rate.



Payload mass success rate

- Payload mass in the 0-2000 kg range had a lower success rate than in the 2000-5000 kg range.



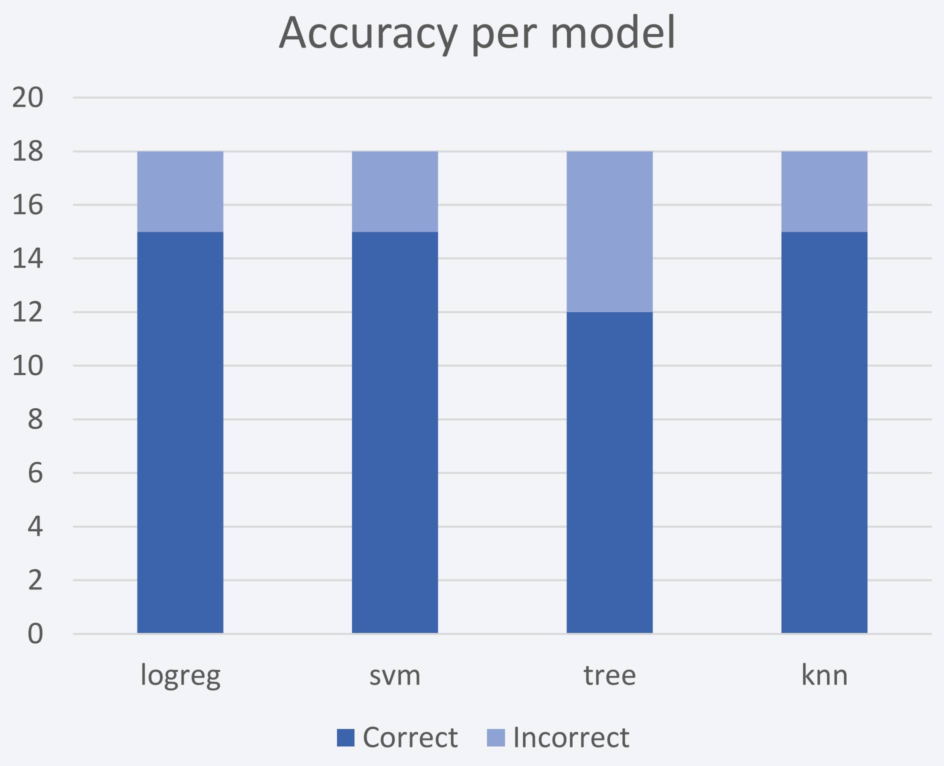


Section 5

Predictive Analysis (Classification)

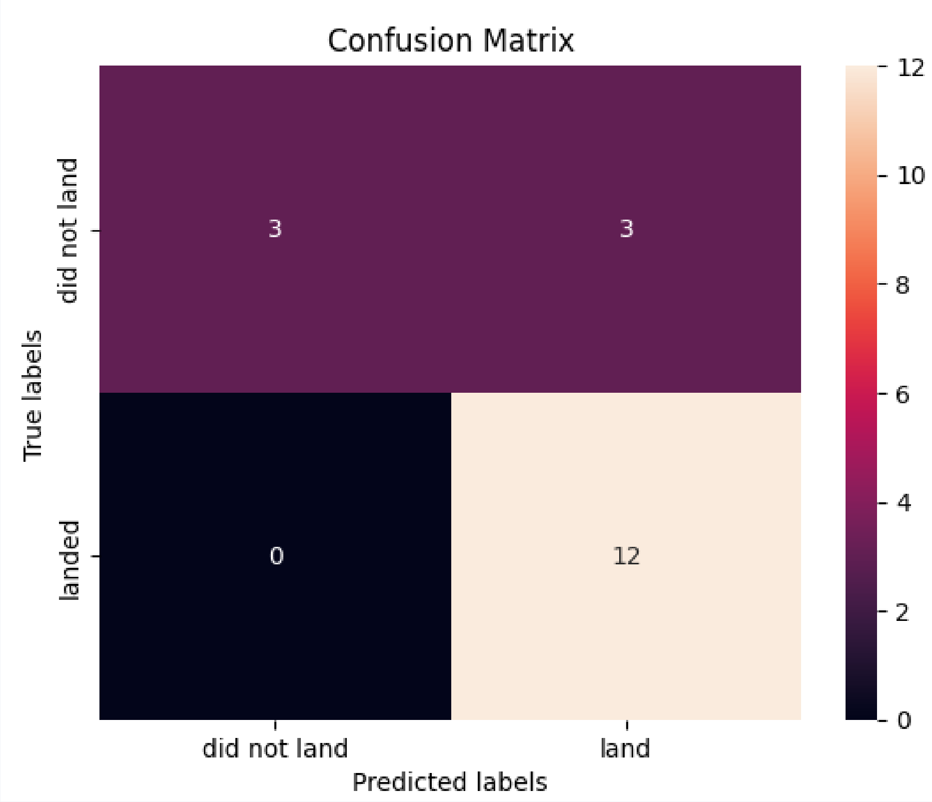
Classification Accuracy

- All models perform similar. The small test size is not enough to decide which model is the best.



Confusion Matrix

- As can be seen in the confusion matrix the best model was accurate except for 3 false positives out of 18 samples.
- Logistic regression , KNN and SVM all showed the same outcome.



Conclusions

- Using the logistic regression model, a given launch is expected to succeed 83.3% of the time.
- Weighting the expected outcome with the costs of $62 million for success and $165 million for failure, the average expected cost is $79 million.
- Further we can adjust for the accuracy of the model and expect it to be within the range of $67 million and $91 million.

Appendix

Cost in millions	Value	
Accuracy of model		84,6%
Expected success		83,3%
Expected failure		16,7%
Cost of success	$	62
Cost of failure	$	165
Weighted cost of success	$	52
Weighted cost of failure	$	28
Total expected cost per launch	$	79
Low	$	67
High	$	91

Thank you!