IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

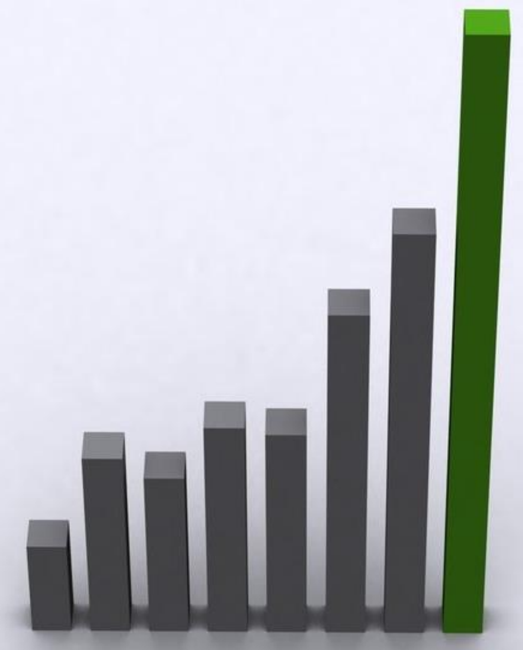
Gary R. Corder
26 April 2023

Outline

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

Executive Summary

- **Methodology**
  - Webscrape data from url: https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches
  - Use relevant JSON type data to develop cost prediction model in python
  - Clean data/present findings
- **Results Summary**
  - Accuracy on decision tree test data: 83.3%
  - SVM able to predict cost well



Introduction

- **Project background**
  - SpaceX rocket launches are relatively inexpensive
  - Costs of launch dependent on if first stage is reused
- **The Underlying Problem**
  - Is there a way to ultimately determine the cost of launch?
  - If the first stage is reused, how much does the launch cost lower?

Section 1

Methodology

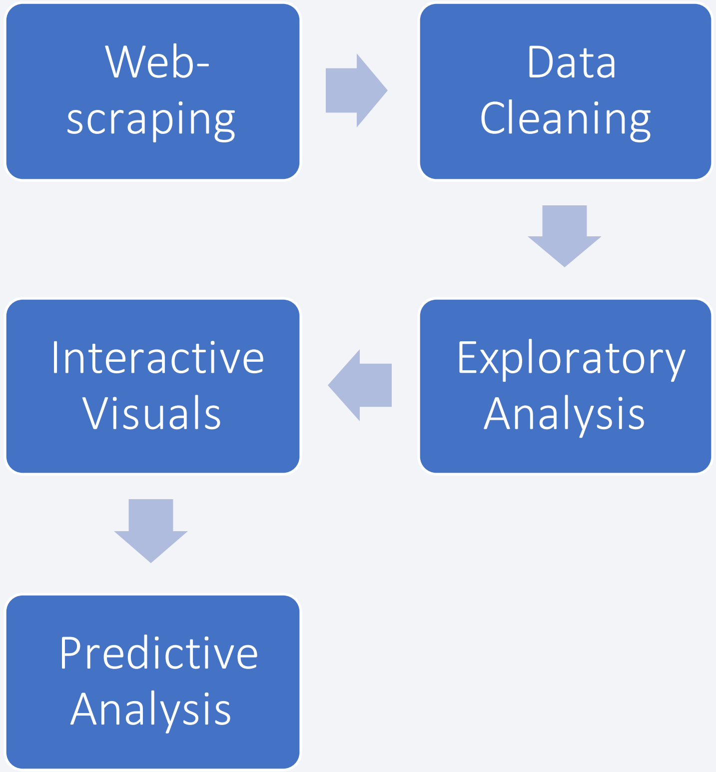
Methodology

Executive Summary

- Data collection methodology:
  - Data collected using web scraping techniques
- Data wrangling
  - Data processed by pandas software, converted to a dataframe through JSON, calculations performed using python
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Support Vector Machine model used for predictive analysis

Data Collection

- Data sets
  - Data converted from raw html code into a JSON file
  - Collected by using pandas software within python to store and clean data
  - Key Data: booster name from rocket, mass of payload, name of launch site and location
- You need to present your data collection process use key phrases and flowcharts



Data Collection – SpaceX API

- CALLS
  - response = requests.get(spacex_url)
  - print(response.content)

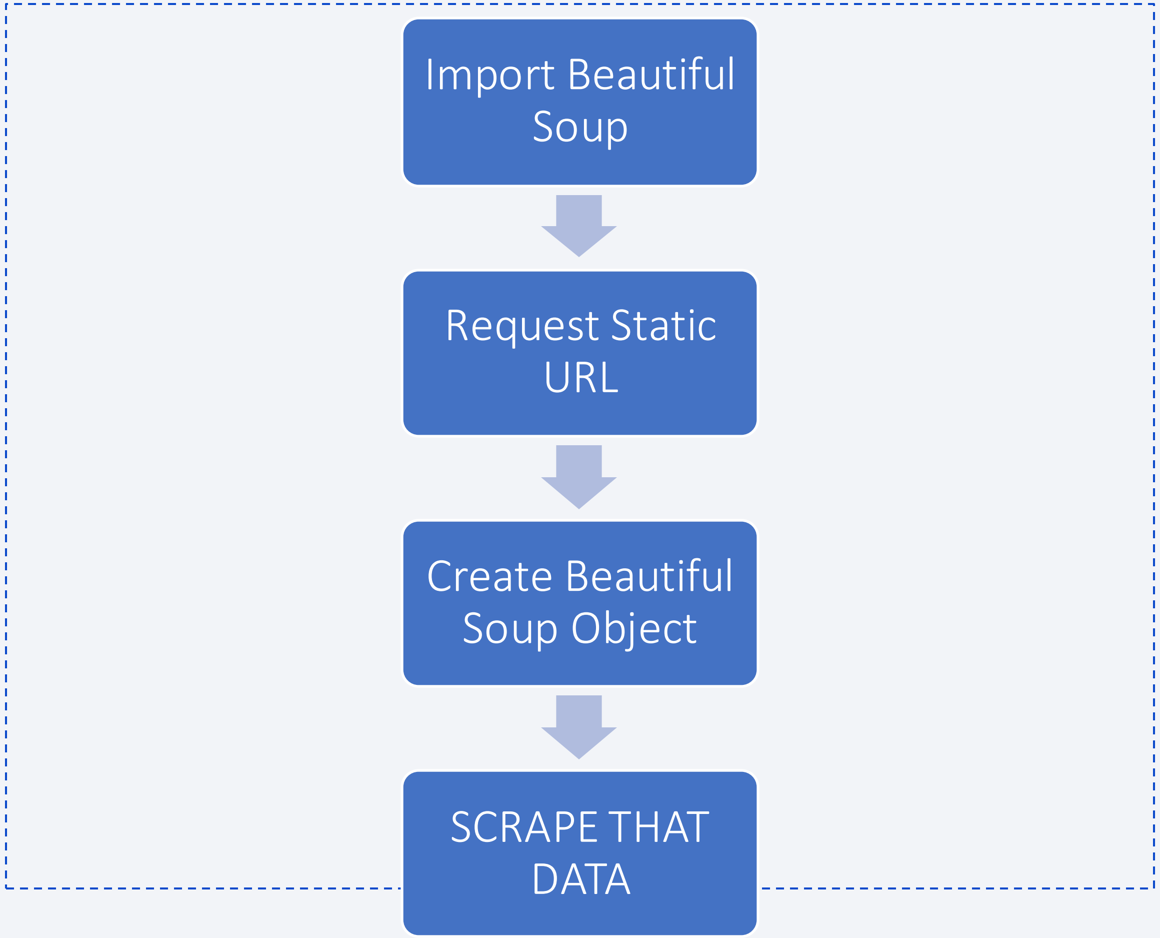
- GitHub URL: https://github.com/cordergr/FinalProjectIBM/blob/main/datacollection.ipynb

Request	Request to join REST API
Convert	Convert data into a JSON file and normalize
Save	Save data into a pandas dataframe to be used for future work

Data Collection - Scraping

- Web Scraping Flowchart

- GitHub URL: https://github.com/cordergr/FinalProjectIBM/blob/main/Webscraping.ipynb



Data Wrangling

USE "DATA_FALCON9.ISNULL().SUM()" TO FIND NULL VALUES IN DATA

DROP THE NULL VALUES BY USING MEAN OF DATA VALUES, OR REPLACING WITH ZERO.

THAT DATA IS WRANGLED

GITHUB
URL: HTTPS://GITHUB.COM/CORDERGR/FINALPROJECTIBM/BLOB/MAIN/DATACOLLECTION.IPYNB

EDA with Data Visualization

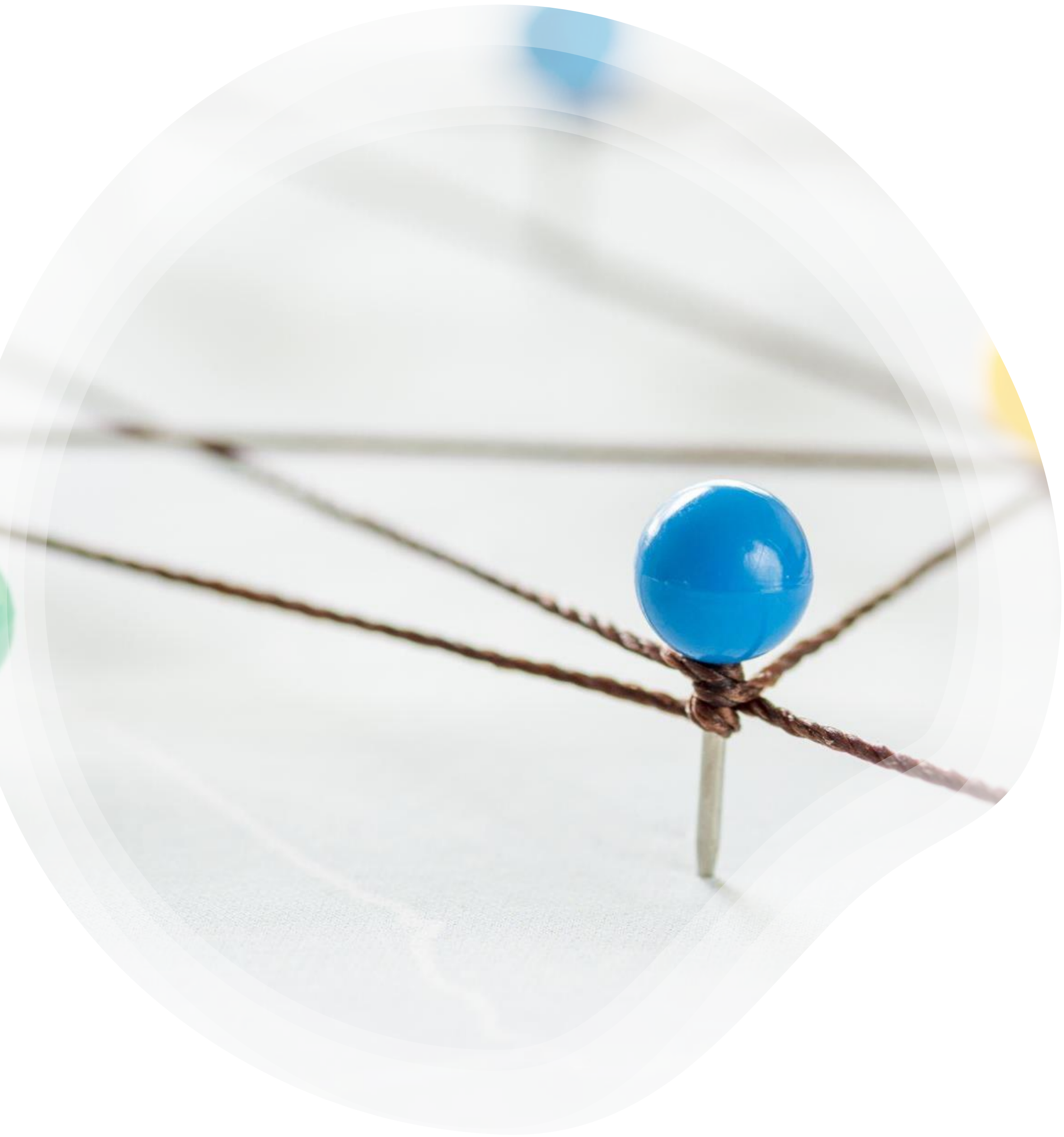
Scatterplots for holistic visualization

Data frames for easy data manipulation

GitHub URL: https://github.com/cordergr/FinalProjectIBM/blob/main/EDA2.ipynb

EDA with SQL

- SELECT DISTINCT
- SELECT*FROM
- SELECT SUM
- SELECT AVG
- SELECT MIN
- AND MORE! In order to explore all important data.
- GitHub URL: https://github.com/cordergr/FinalProjectIBM/blob/main/EDA.ipynb



Build an Interactive Map with Folium

- Map objects such as map marker, map circle, marker cluster, mouse position, and distance marker were added to the interactive map with Folium.
- All objects added to increase visualization and interaction with the data for enhanced understanding.
- GitHub URL: https://github.com/cordergr/FinalProjectIBM/blob/main/LaunchSitesLocationAnalysis.ipynb

Build a Dashboard with Plotly Dash

Interactive pie charts, scatterplots, and line graphs were chosen to display the relevant information.

Each plot and interaction helps to visualize the data and tell a story behind each SpaceX launch. Time, success, and accuracy are all of extreme importance.

GitHub URL for python dash app: https://github.com/cordergr/FinalProjectIBM/blob/main/dash.py

Predictive Analysis (Classification)

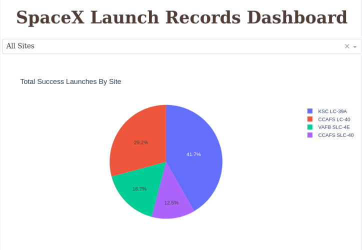
The Classification Models were built after cleaning/normalizing data. Accuracy tests were performed to choose the best model.

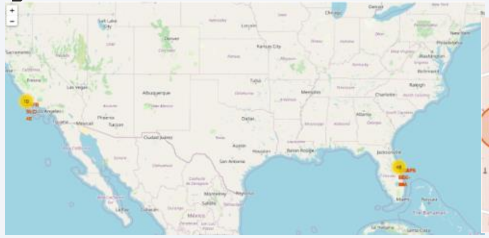
The best model for predictive analysis of a success is the Decision Tree, by use of accuracy.

Github Predictive Analysis
URL: https://github.com/cordergr/FinalProjectIBM/blob/main/MachineLearningPredictionModel.ipynb

Results

- EDA results: launch success rate generally rises each year.
    - Total mission outcome success is positive! (query on right)
- Interactive analytics demo in screenshots



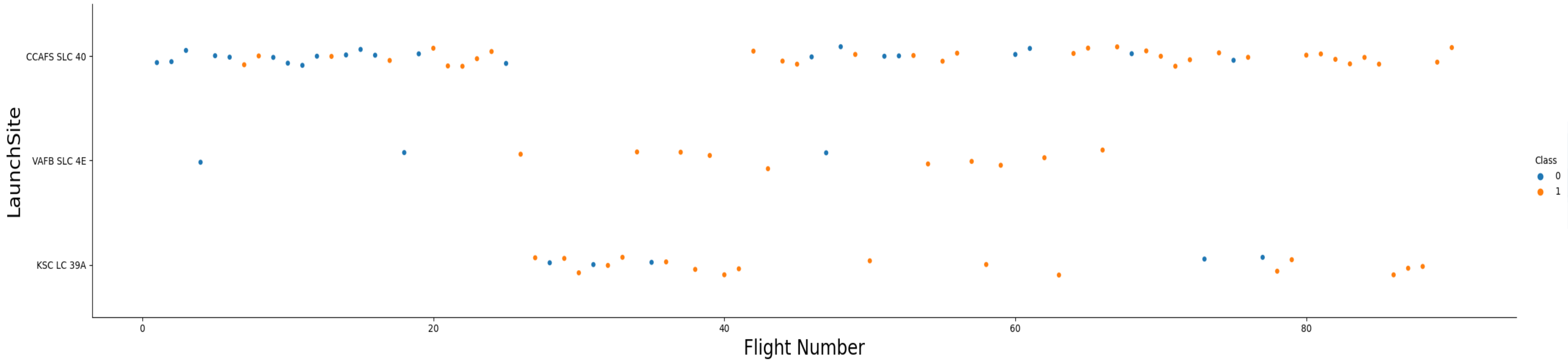


- Predictive analysis results
    - Decision Tree chosen as best model for predictive analysis because it has the highest accuracy (**87.7%**)

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

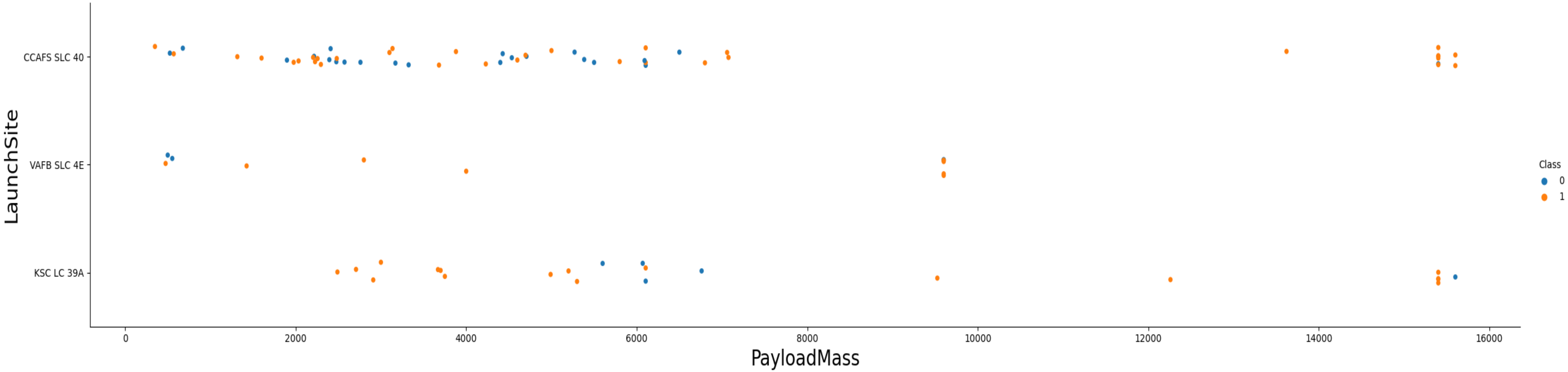
Section 2

Insights drawn from EDA



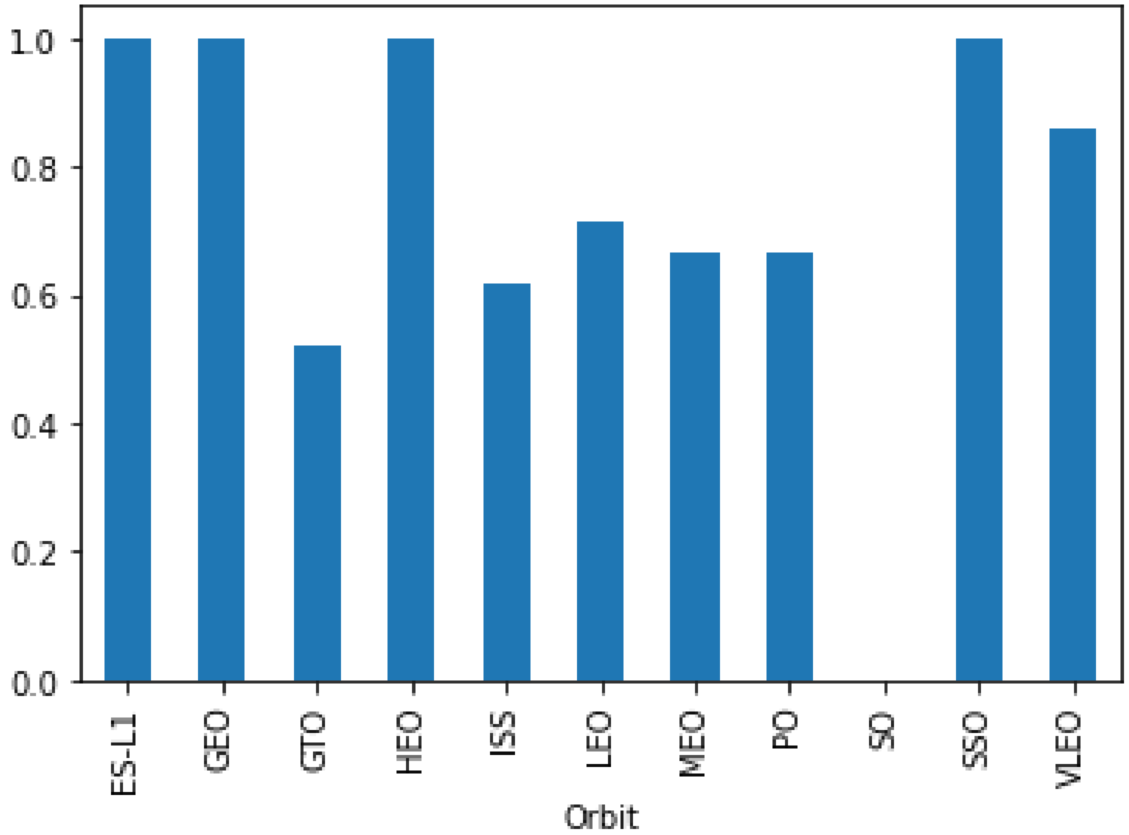
Flight Number vs. Launch Site

- The graph above shows the flight number vs. Launch site. The darker points are successful launches, the lighter points are unsuccessful launches. The graph makes it easier to see if there is any correlation between the launch site and successful launches.



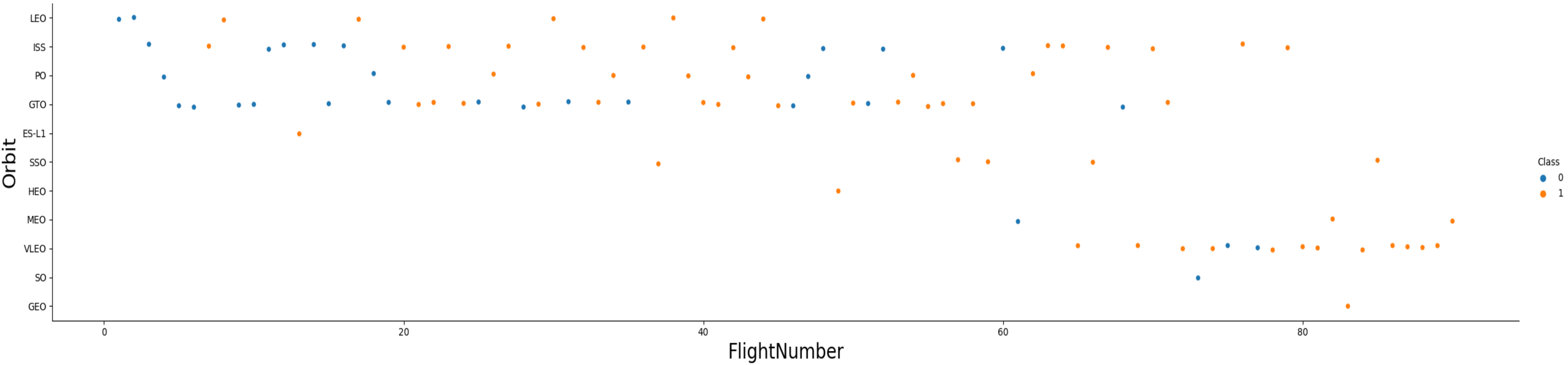
Payload vs. Launch Site

- The graph above shows the payload vs. Launch site. The darker points are successful launches, the lighter points are unsuccessful launches. The graph makes it easier to see if there is any correlation between the rocket payload and successful launches.



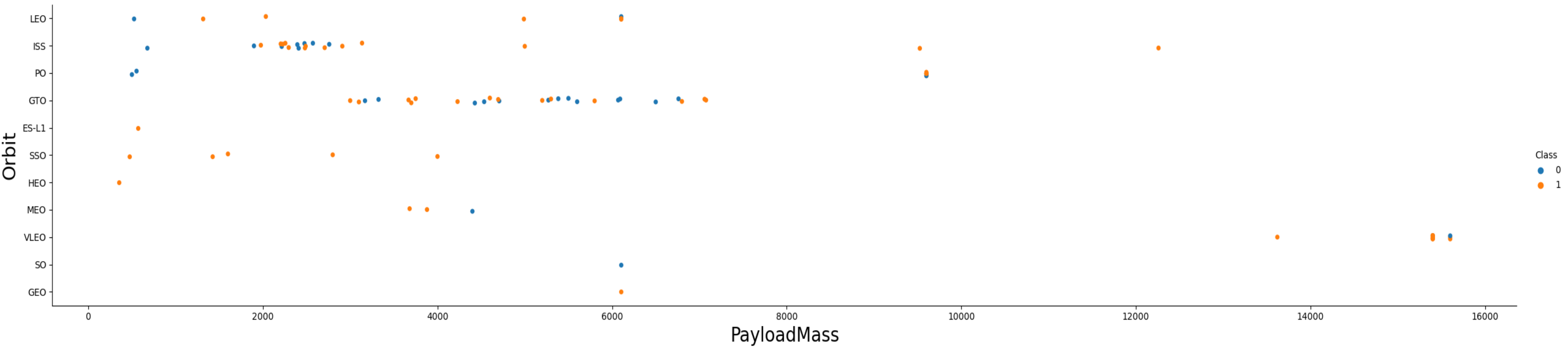
Success Rate vs. Orbit Type

- The bar chart shows how successful (from 0% success to 100% success) each orbit type is with launch.



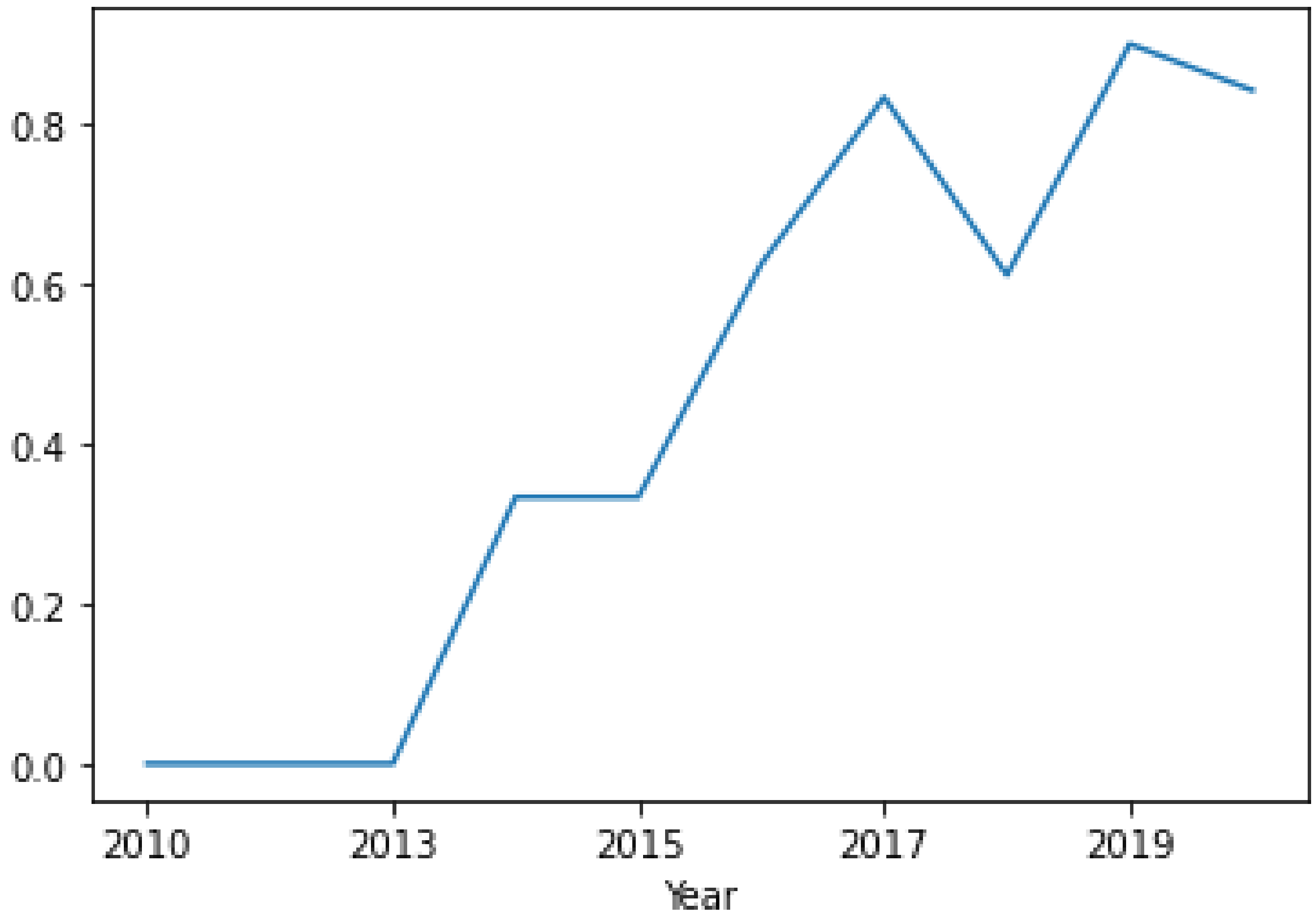
Flight Number vs. Orbit Type

- The plot above shows that there might be a correlation between flight numbers and LEO orbit type. It seems that the lower flight numbers have had higher success.



Payload vs. Orbit Type

- Heavy payloads seem to have a negative influence on GTO orbits and positive on GTO and Polar LEO (ISS) orbits.



Launch Success Yearly Trend

It is easy to tell from the graph that the success rate increases overall with the year.

All Launch Site Names

[7]:

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

- All unique launch site names are found in the query to the left. The names include a numerical value followed by a number/letter code.

[8]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome
04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
08-12-2010	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)
22-05-2012	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
01-03-2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Launch Site Names Begin with 'CCA'

- The launch site names that begin with 'CCA' are in the query above. Also listed is the date, time, booster version, payload, payload mass, orbit type, customer, mission outcome, and landing outcome

[9]:

TOTAL_PAYLOAD
111268

Total Payload Mass

- The total payload mass is displayed in the query above, reaching 111,268 kilograms. That's a lot of mass!

Average Payload Mass by F9 v1.1

- The average payload by mass (in kilograms) for F9 v1.1 is displayed in the query on the right.

AVG_PAYLOAD
2928.4

First Successful Ground Landing Date

- The date of the first successful landing outcome on the ground pad is displayed in the query to the right, with 'gp' meaning 'ground pad'.

first_success_gp
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- The names of boosters which have successfully landed on a drone ship and had payload mass between 4000 and 6000 kilograms are displayed in the query to the right.

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

Total Number of Successful and Failure Mission Outcomes

- The total number of successful and failure mission outcomes are in the query below.
- Overall, the mission outcome success rate was high.

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

Boosters Carried Maximum Payload

- The boosters which carried the maximum payload mass are displayed in the query to the right.
- There are a total of 15 boosters which have carried the 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

2015 Launch Records

- The failed landing outcomes in drone ship, their booster versions, and launch site names are listed in the query for 2015.
- The booster version is listed next to their corresponding launch site.

booster_version	launch_site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

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

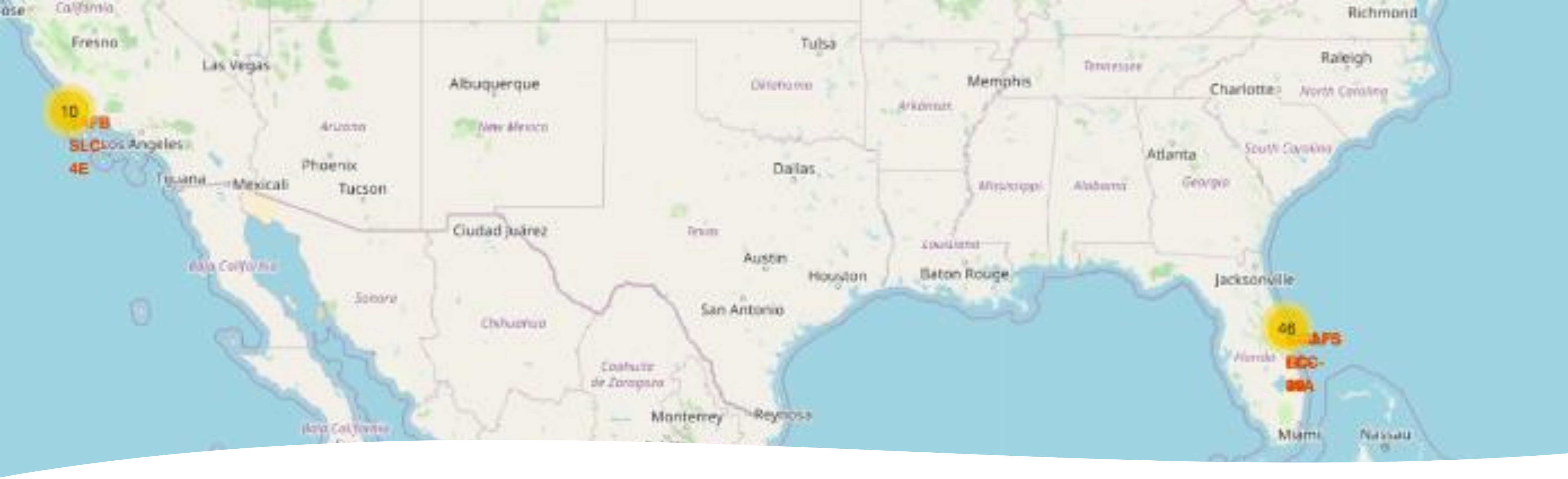
- Landing outcomes are ranked in descending order from June 4, 2010 to March 20, 2017.

- There is a brief description of each landing outcome beside the frequency.

landing_outcome	qty
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

Section 3

Launch Sites Proximities Analysis

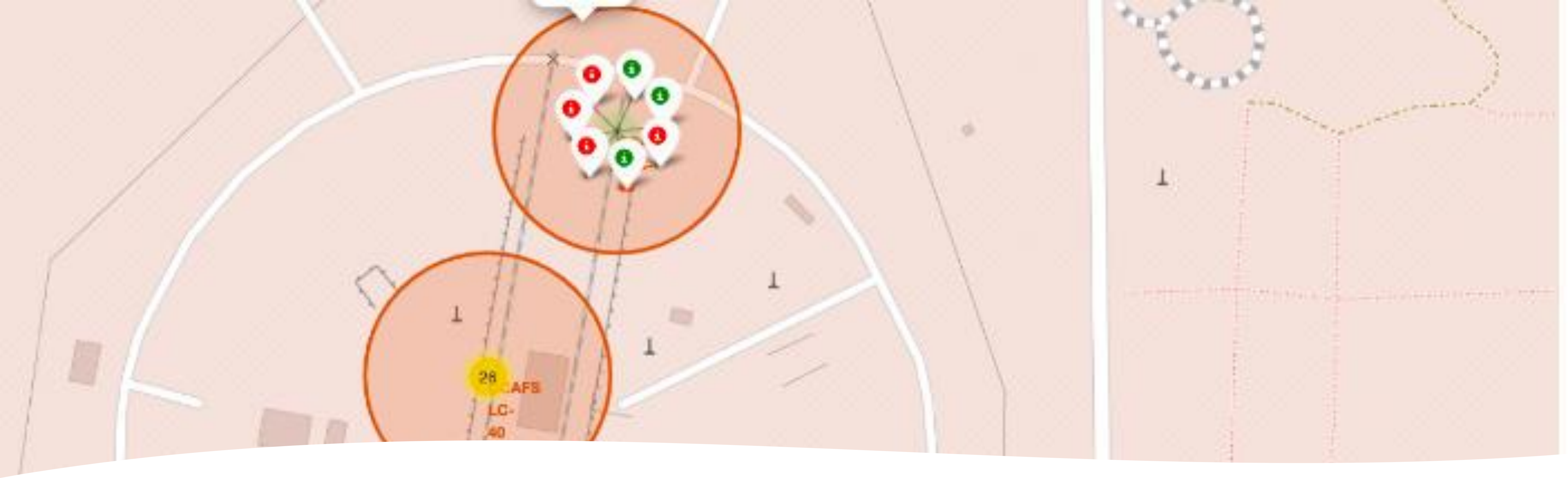


Folium Map (zoomed out)

- Above shows an interactive map made in folium, zoomed out. The map shows the dfferent launch sites and how many launches occurred.

Folium Map, zoomed in

- Above shows the folium map zoomed in. The map shows how far away some of the launch sites are from the closest coastline.



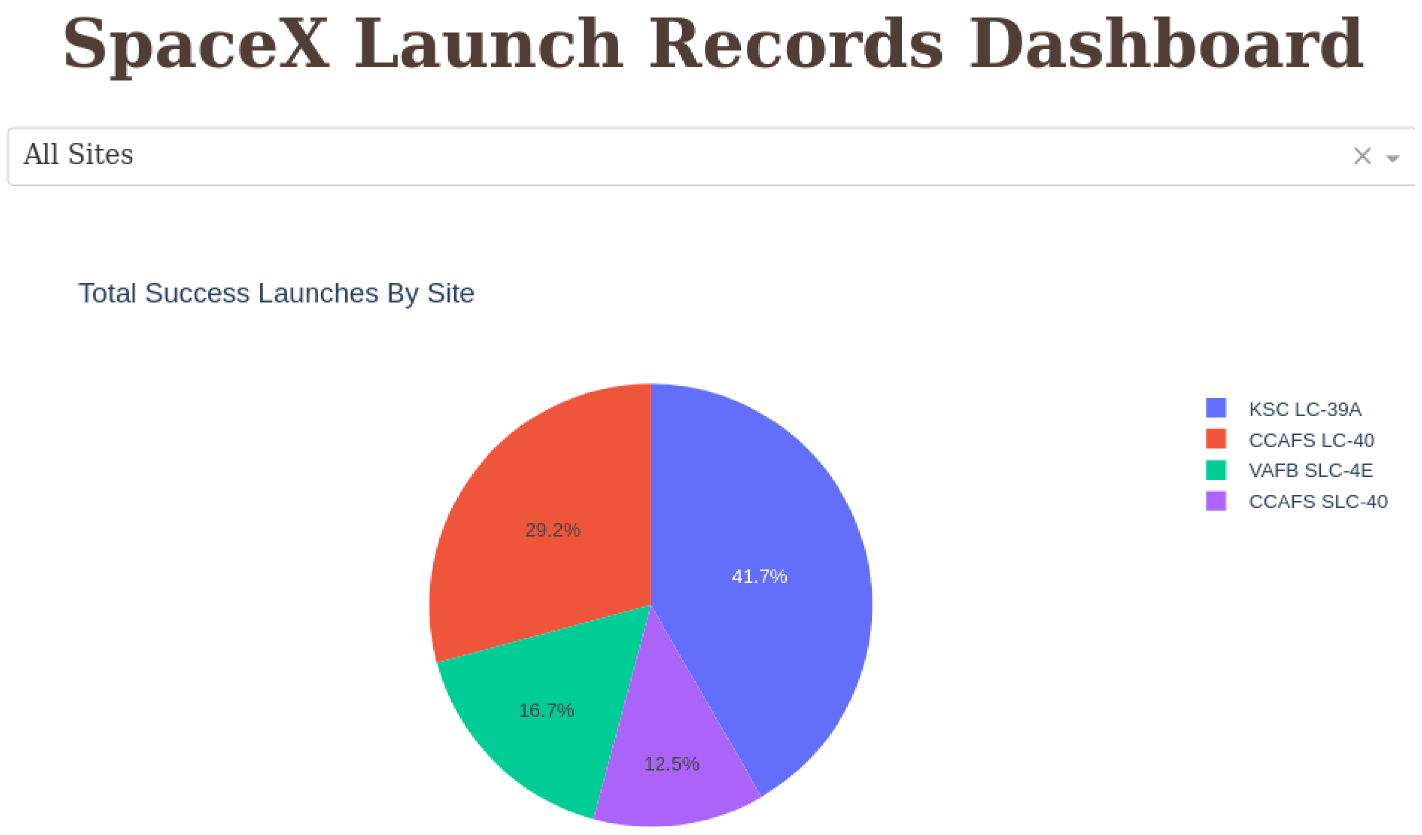
Folium Map, zoomed in (cont)

- Above, the folium map shows if each launch was successful. The green markers indicate a successful launch, while the red ones mean an unsuccessful launch.

Section 4

Build a Dashboard with Plotly Dash

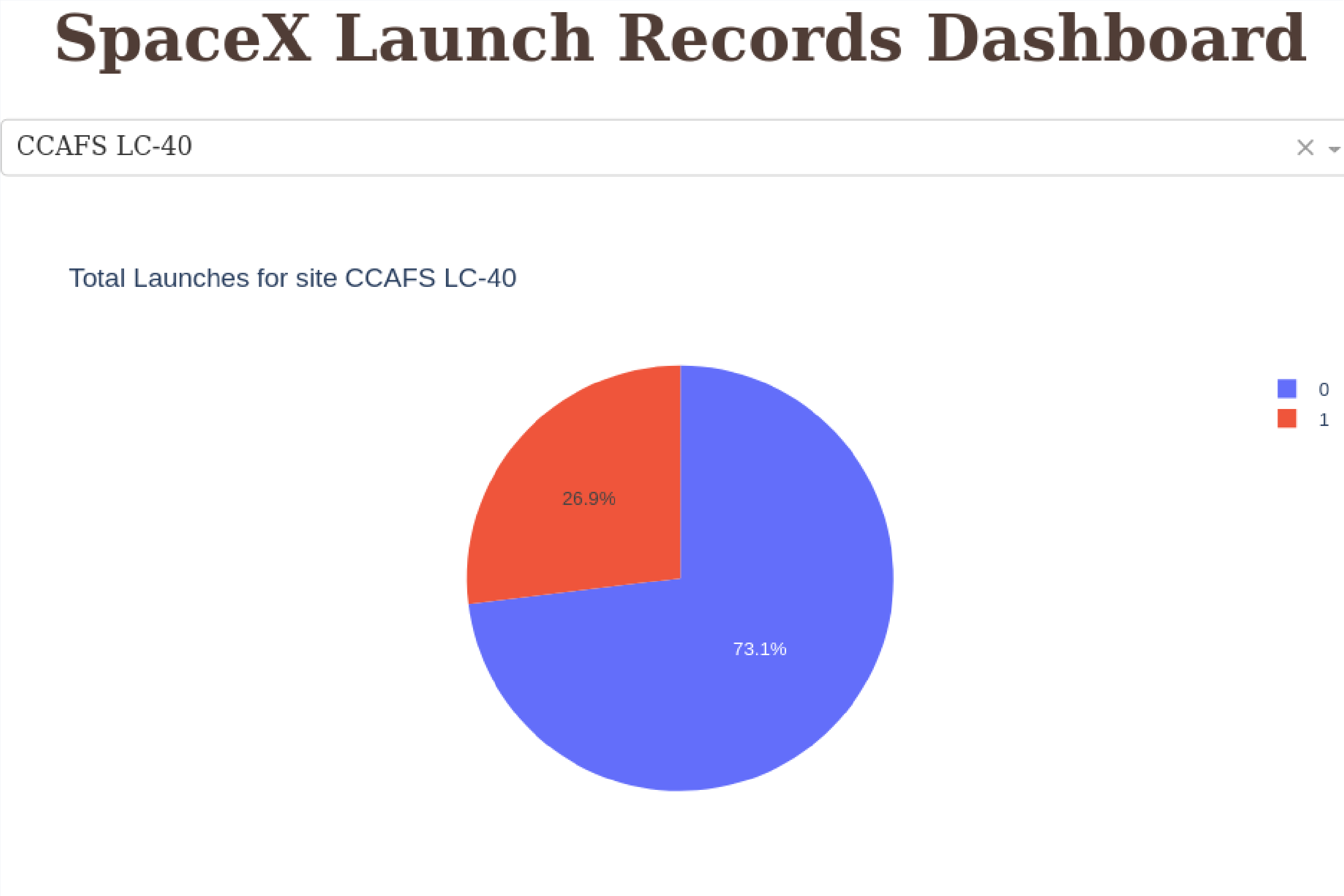
Launch Success Count for all sites



- The Dashboard shows the total launch success count for all launch sites. The site with the highest number successful launches is KSC LC-39A, while CCAFS SLC-40 has the least number of successful launches.

Highest Launch Success Ratio

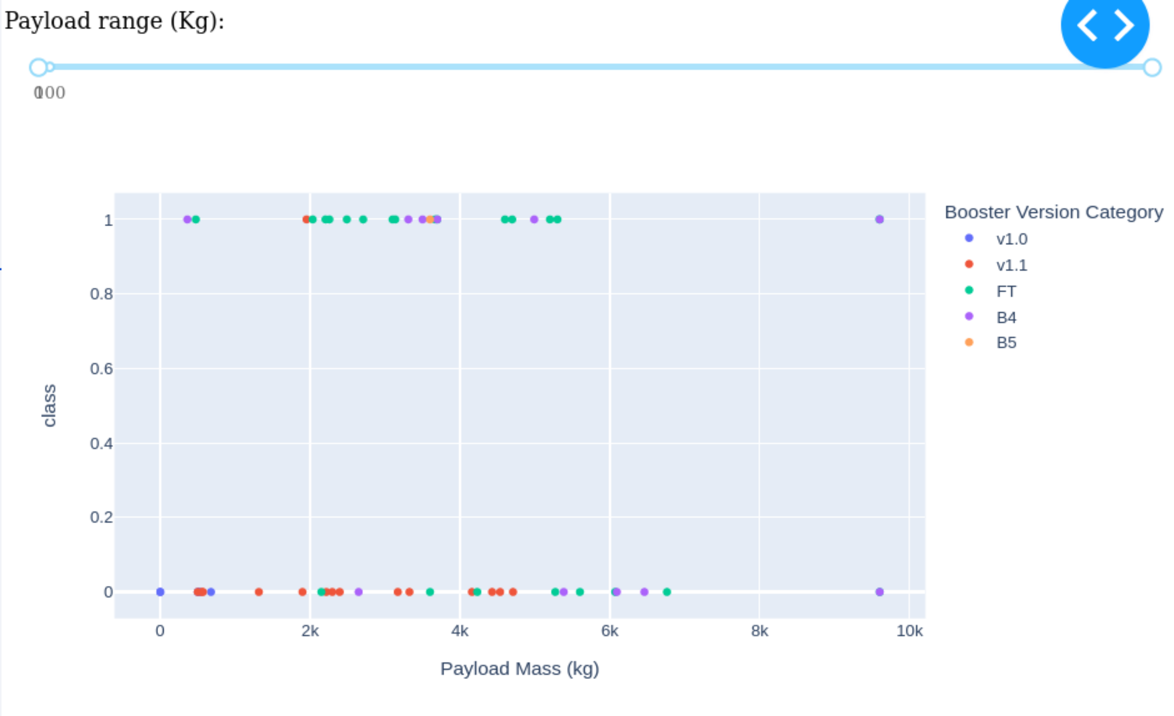
- Show the screenshot of the piechart for the launch site with highest launch success ratio
- CCAFS LC-40 was the launch site with the highest launch success ratio, with **73.1%.**

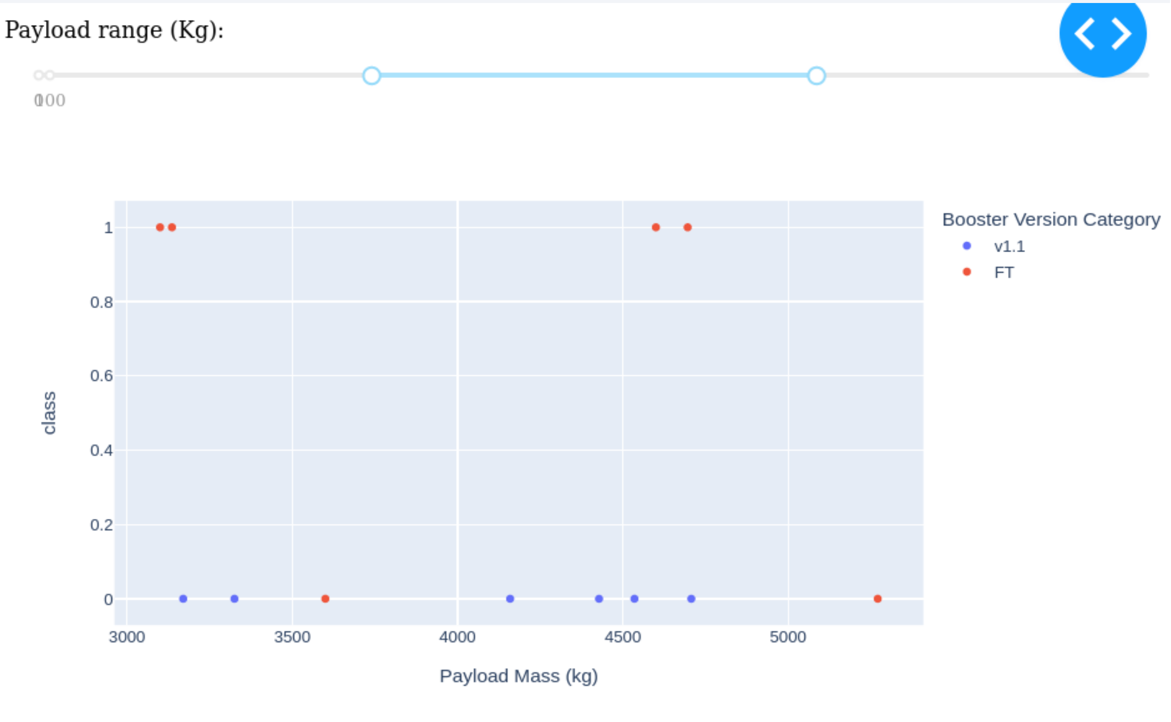


Payload range V. Booster

Every booster version and their success rate is shown along with their full payload range in kgs (top graph).

Booster version FT has a better success rate than v1.1 for payload range 3000-5500 kgs (bottom graph).





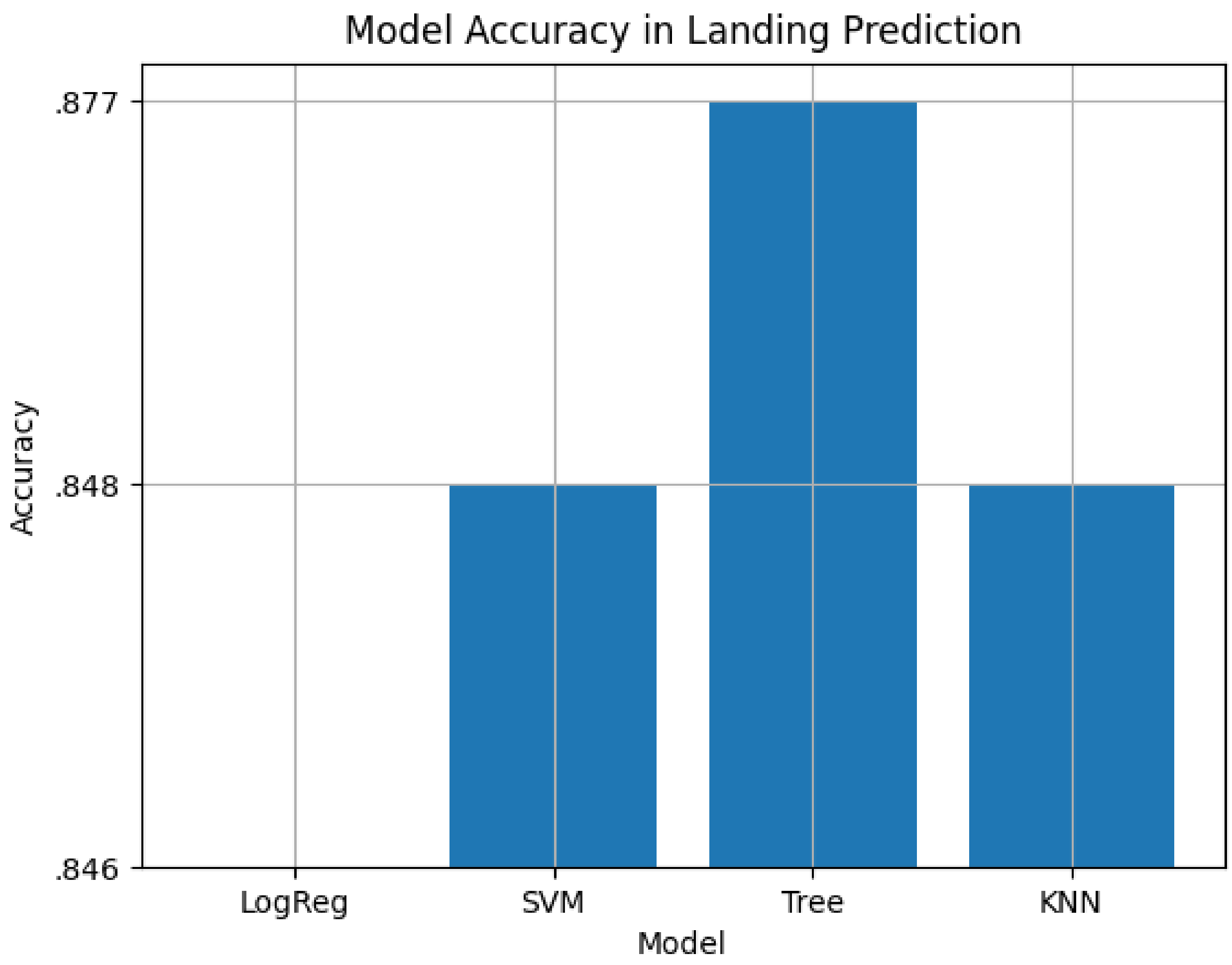
Section 5

Predictive Analysis (Classification)

Classification Accuracy

- LogReg = Logistic Regression
- SVM = Support Vector Machines
- Tree = Decision Tree
- KNN = K-nearest neighbors

- **Decision Tree model** has highest accuracy



Confusion Matrix

- Confusion Matrix for Decision Tree Model
  - Accuracy of **87.7%**
  - Test Accuracy of **88.9%**
  - Predicted all 12 true landings correctly
  - Predicted 5/6 failures to land

Conclusions

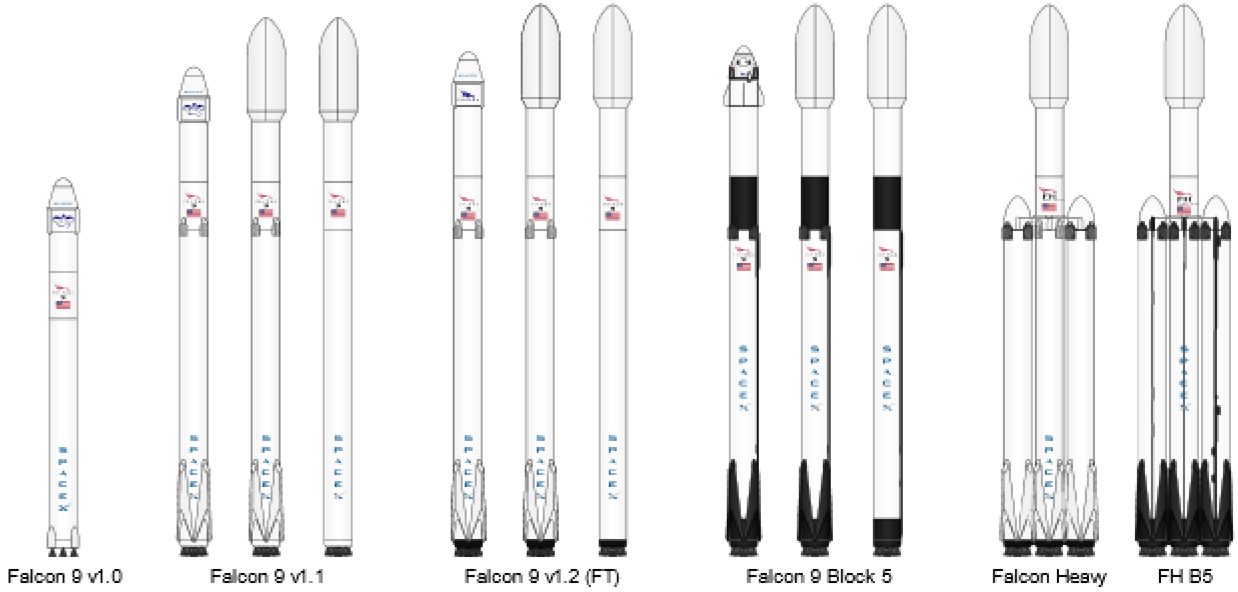
- Decision Tree model is the best model to predict the outcome of SpaceX launch cost reduction using predictive analysis
  - Compared to SVM, KNN, logistic regression
  - Accuracy shown for each model on the right

- Web scraping using the Beautiful Soup package is an ideal way to gain data insight from HTML data

Model	Accuracy	TestAccuracy
LogReg	0.84643	0.83333
SVM	0.84821	0.83333
Tree	0.87679	0.88889
KNN	0.84821	0.83333

Appendix

- Useful code:
- soup = BeautifulSoup(data, 'html5lib')
- df=pd.DataFrame(launch_dict)
- tree_cv = GridSearchCV(estimator=tree, cv=10, param_grid=parameters)
- tree_cv.fit(X_train, Y_train)
- Many thanks to IBM for creating tasks and helpful code.



Thank you!