DATA SCIENCE CAPSTONE PROJECT

Bodala Sri Varshitha
23rd Dec, 2022

Outline

- Executive Summary
- Introduction
- Methodology
- Results
  - Visualization – Charts
  - Dashboard
- Discussion
  - Findings & Implications
- Conclusion
- Appendix

Executive Summary

Executive Summary

Methodologies used to analyze the data:

- Data collection
- Data wrangling
- Exploratory data analysis with Data Visualization, SQL
- Building an interactive map and dashboard with Folium & Plotly Dash
- Predictive analysis using classification technique

Summary of results:

- Exploratory data analysis results
- Predictive analysis results

Introduction

Introduction

The commercial space age is here, companies are making space travel affordable for everyone. SpaceX is one of these companies, with relatively inexpensive rocket launches. SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upwards of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage. Therefore, if we can determine if the first stage will land, we can determine the cost of a launch. Spaces X's Falcon 9 launch like regular rockets. The main objective is to determine the price of each launch by gathering the information from SpaceX for the company SpaceY, which would like to compete with SpaceX. Also, a machine learning model will be trained to determine whether SpaceX will reuse the first stage.

Methodology

Methodology

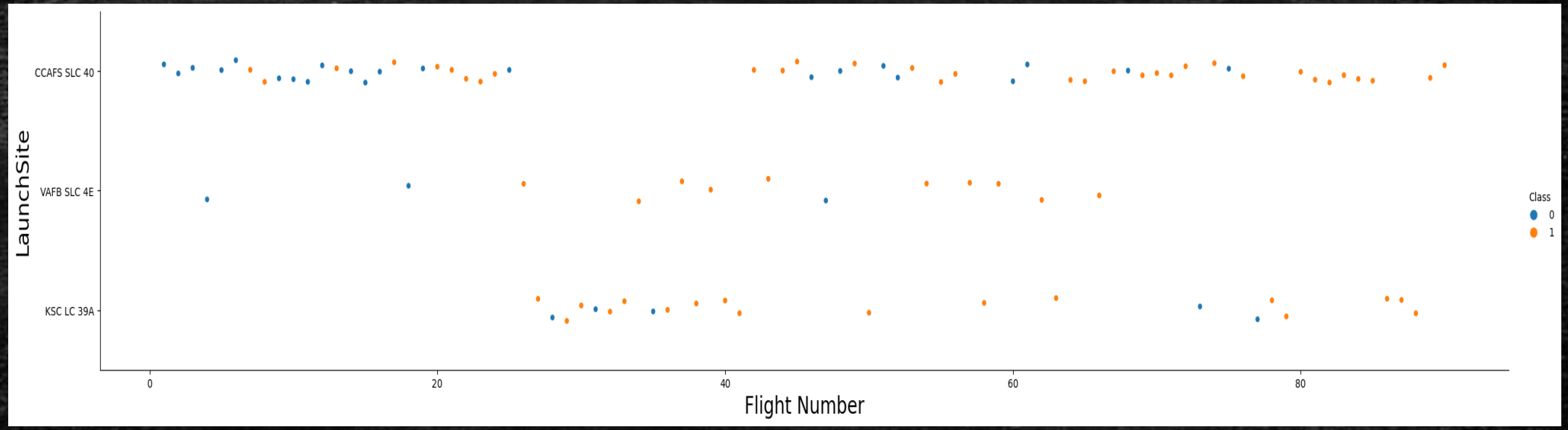
- Data collection using SpaceX REST API and web scrapping from Wikipedia
- Data wrangling (Cleaning the data, replacing or dropping missing values, using one-hot encoding for categorical data)
- Exploratory data analysis using visualization and SQL
- Building an interactive dashboard and map using Folium and Plotly Dash for data analysis.
- Predictive analysis using classification techniques (Logistic regression, k nearest neighbor, decision trees, support vector machine - These models were built and were evaluated to choose the best one.

Results

EDA With Visualization Results

Results

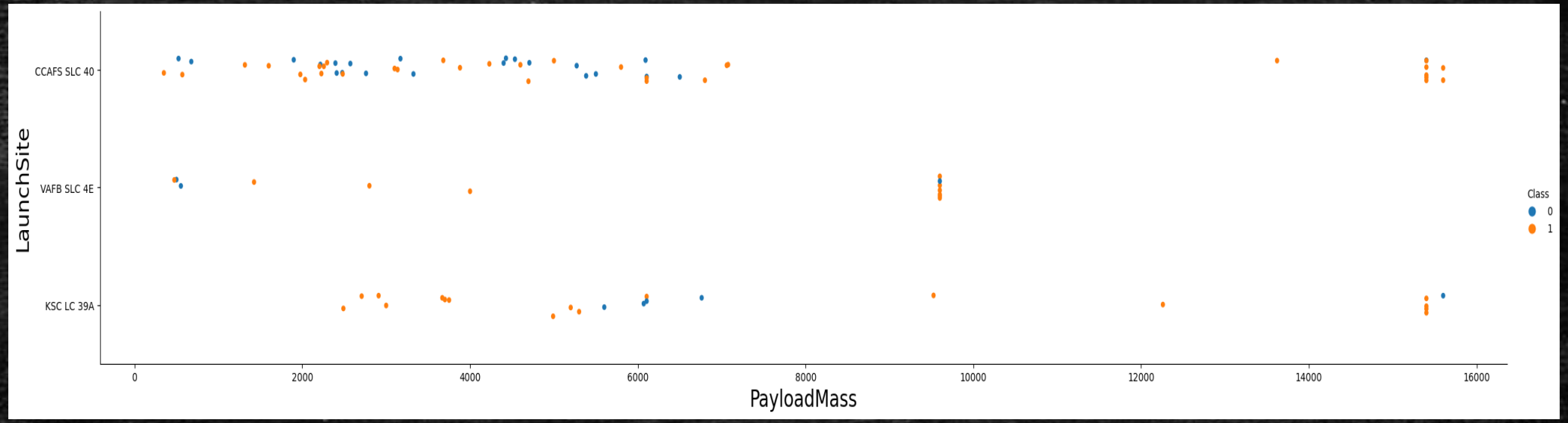
Flight Number Vs Launch Site



It can be observed that most of the earliest flights failed whereas the latest ones succeeded. VAFB SLC 4E and KSC LC 39A have high success rates. CCAFS SLC 40 also high success rate for the latest flights.

Results

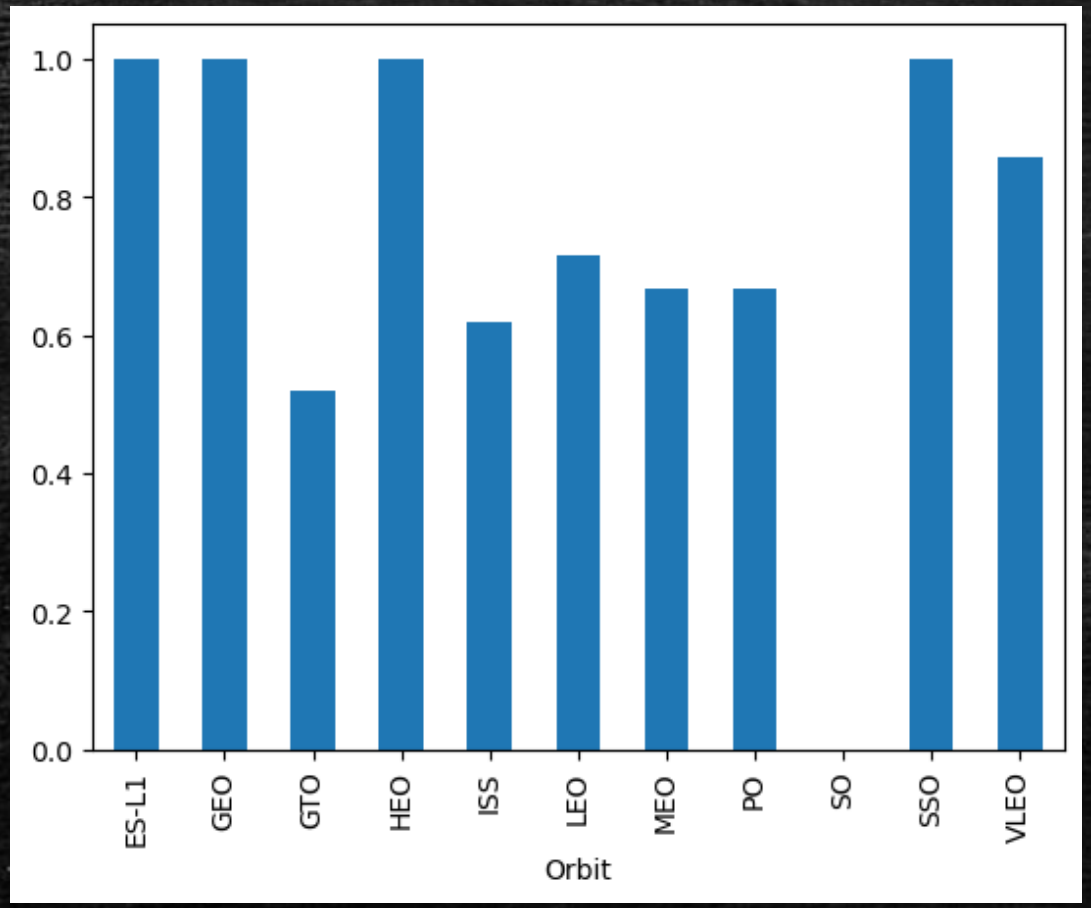
Payload Mass vs Launch Site



Payloads with masses less than 5000kg have 100 percent success rate with KSC LC 39A.
Payloads with masses over 13000kg have 100 percent rate with the CCAFS SLC 40 launch site.

Results

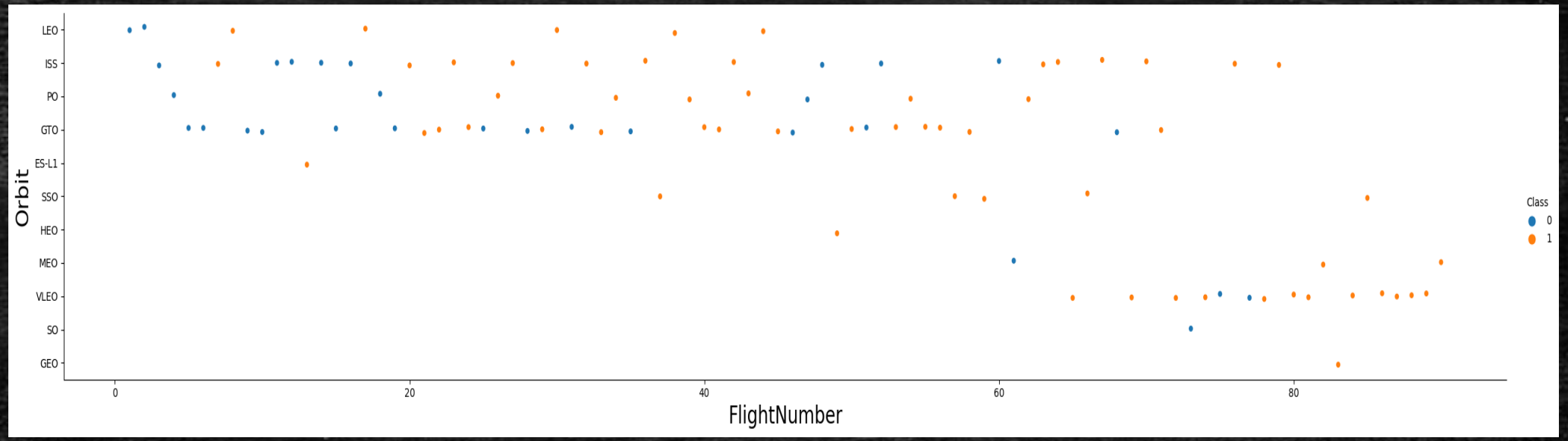
Orbit Type vs Success Rate



- The following orbits have 100 percent success rate:
  - ES-L1
  - GEO
  - HEO
  - SSO
- The SO orbit has zero success rate. The VLEO orbit has a success rate above 80 percent.
- The orbits GTO, ISS, LEO, MEO and PO have a success rate above 50 percent.

Results

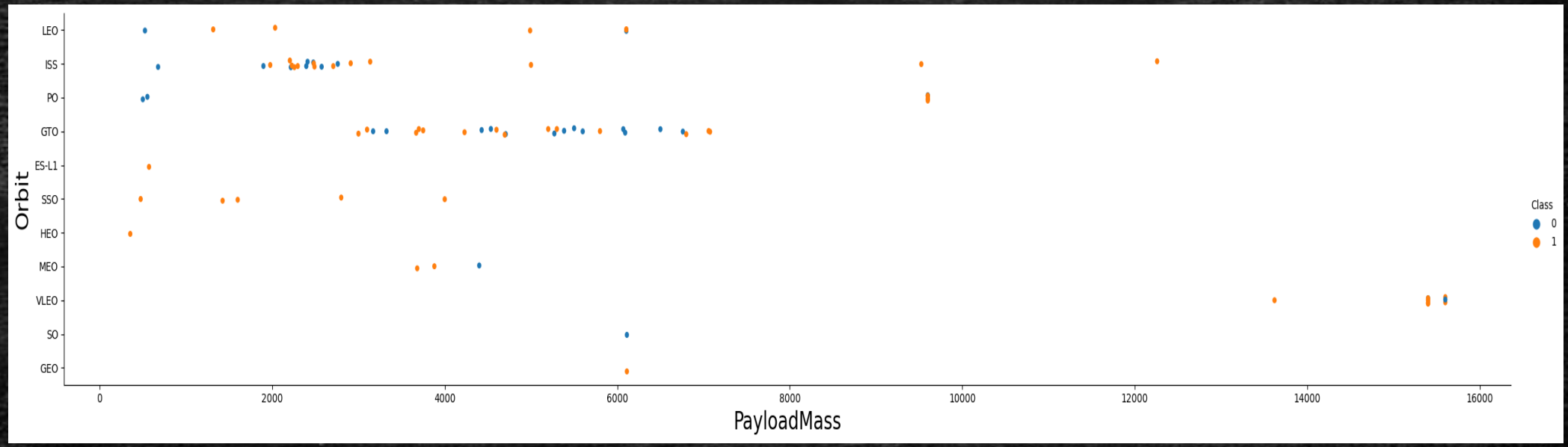
Flight Number vs Orbit



It is observed that most of the recent flights have been successful. That is, success rate increased over time.

Results

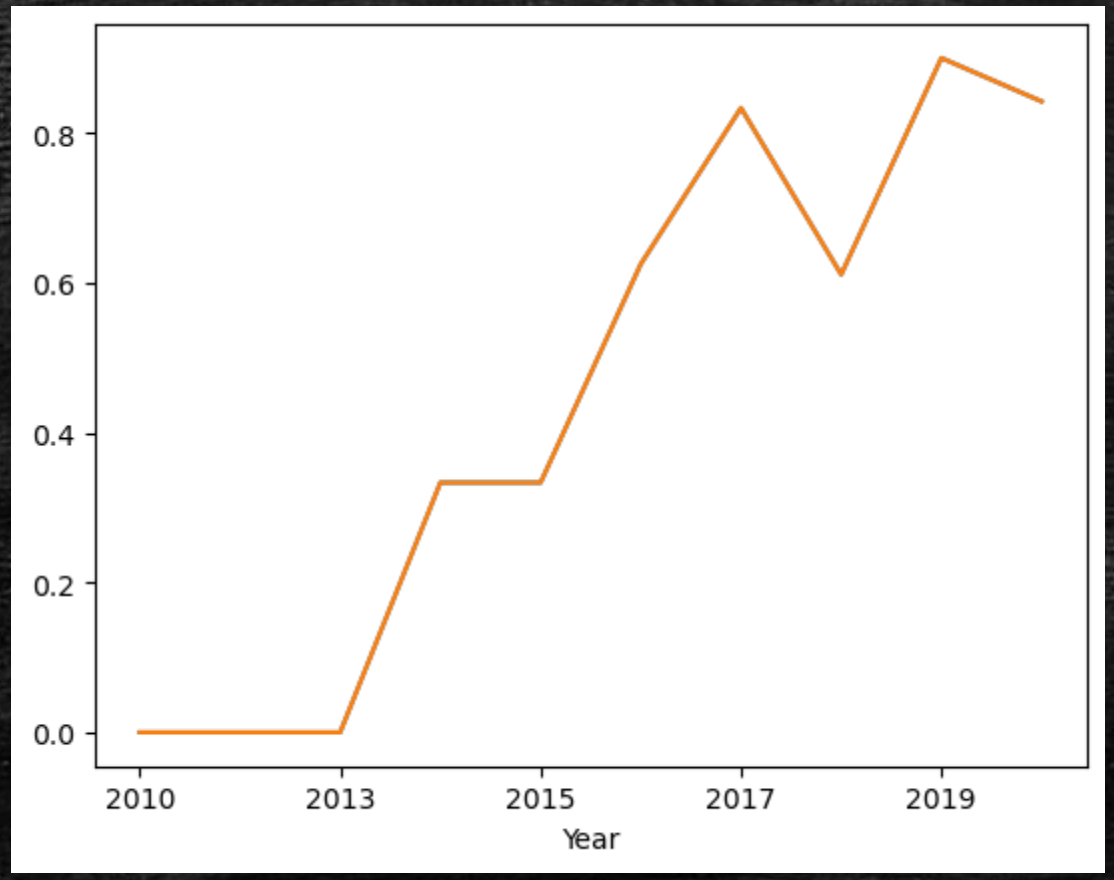
Payload Mass vs Orbit



VLEO orbit has had success with the highest payload masses.
ISS orbit has had success with a varying range of payload masses.
GTO orbit seems to be better for payloads with relatively lower masses.

Results

Launch Success – Yearly Trend



The success rate has increased over the years (Though it fell a little between 2017 and 2019).

EDA With SQL Results

Results

All Launch Site Names:

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

Results

Launch Site Names Beginning With 'CCA':

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

5 launch site with names beginning with 'CCA' are shown.

Results

Total Payload Mass

total_payload_mass
45596

Total payload mass carried by boosters launched by NASA(CRS) is displayed.

Results

Average Payload Mass Carried by Booster Version F9 v1.1

average_payload_mass
2534

Average payload mass carried by booster version F9 v1.1 is as shown.

Results

First Successful Ground Landing Date

first_successful_landing
2015-12-22

The date on which the first successful landing outcome in ground pad was achieved is shown above.

Results

Successful drone ship landing with payload between 4000 and 6000

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

The names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000 are listed.

Results

Total number of successful and failed mission outcomes

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

The total number of successful and failure mission outcomes are listed above.

Results

Boosters carried maximum payload

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

The names of the booster_versions which have carried the maximum payload mass are displayed in this table.

Results

2015 Failed Launch Records

MONTH	DATE	booster_version	launch_site	landing__outcome
January	2015-01-10	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
April	2015-04-14	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

The failed landings in drone ship, their booster versions, and launch site names for the year 2015 are shown.

Results

Rank Landing Outcome Count Between 2010-06-04 and 2017-03-20

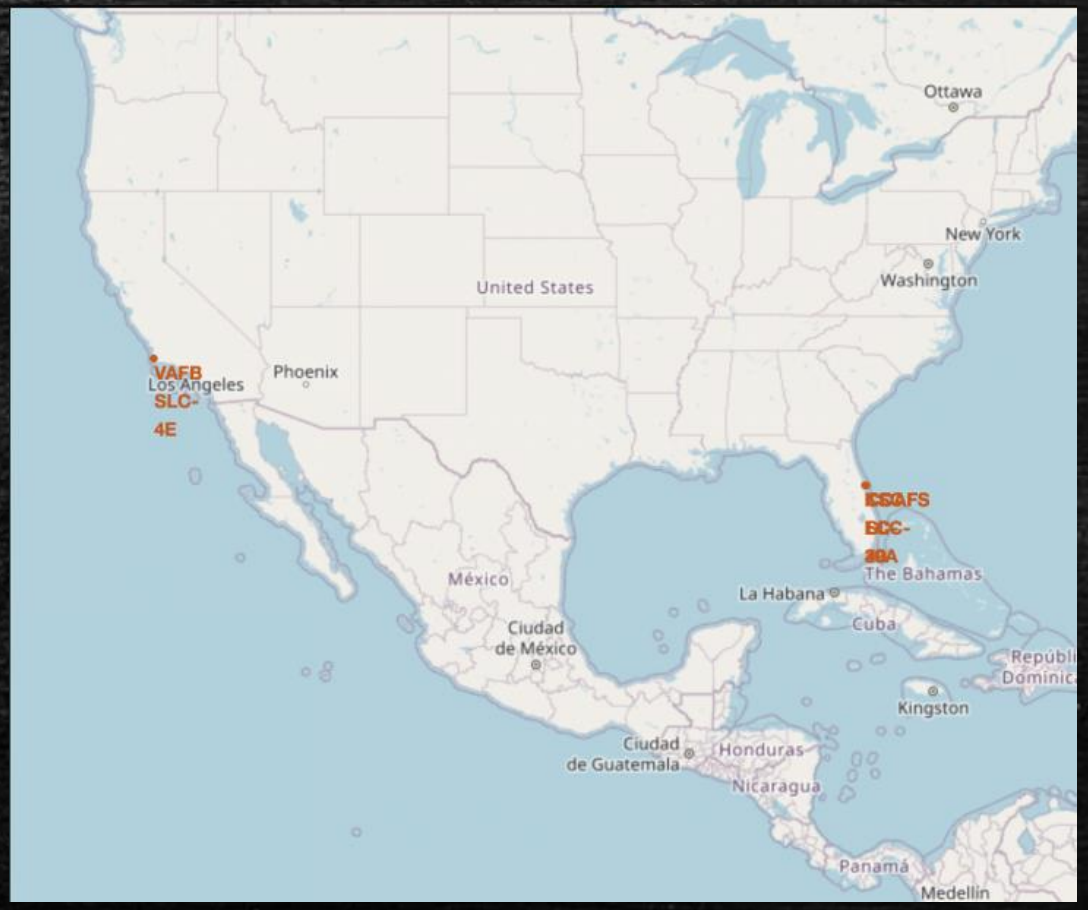
landing__outcome	count_outcomes
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

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 is shown

Interactive Map With Folium Results

Results

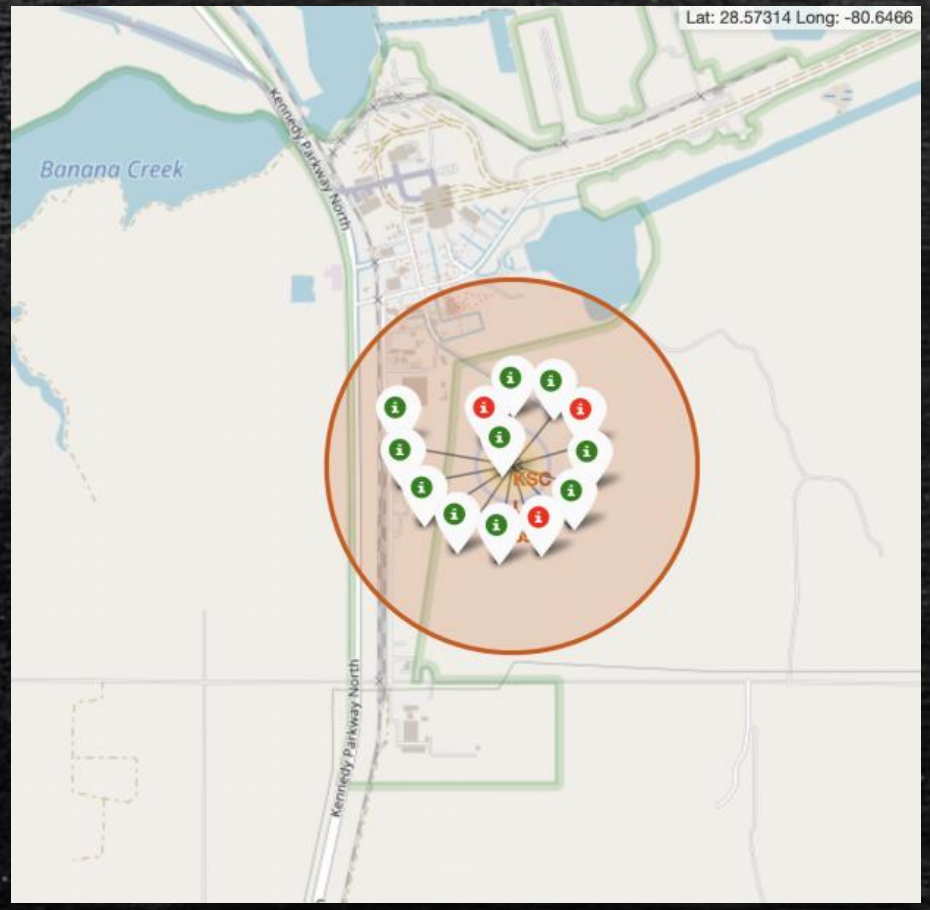
Marking All Launch Sites on a Map



It can be observed that the launch sites are closer to the sea.

Results

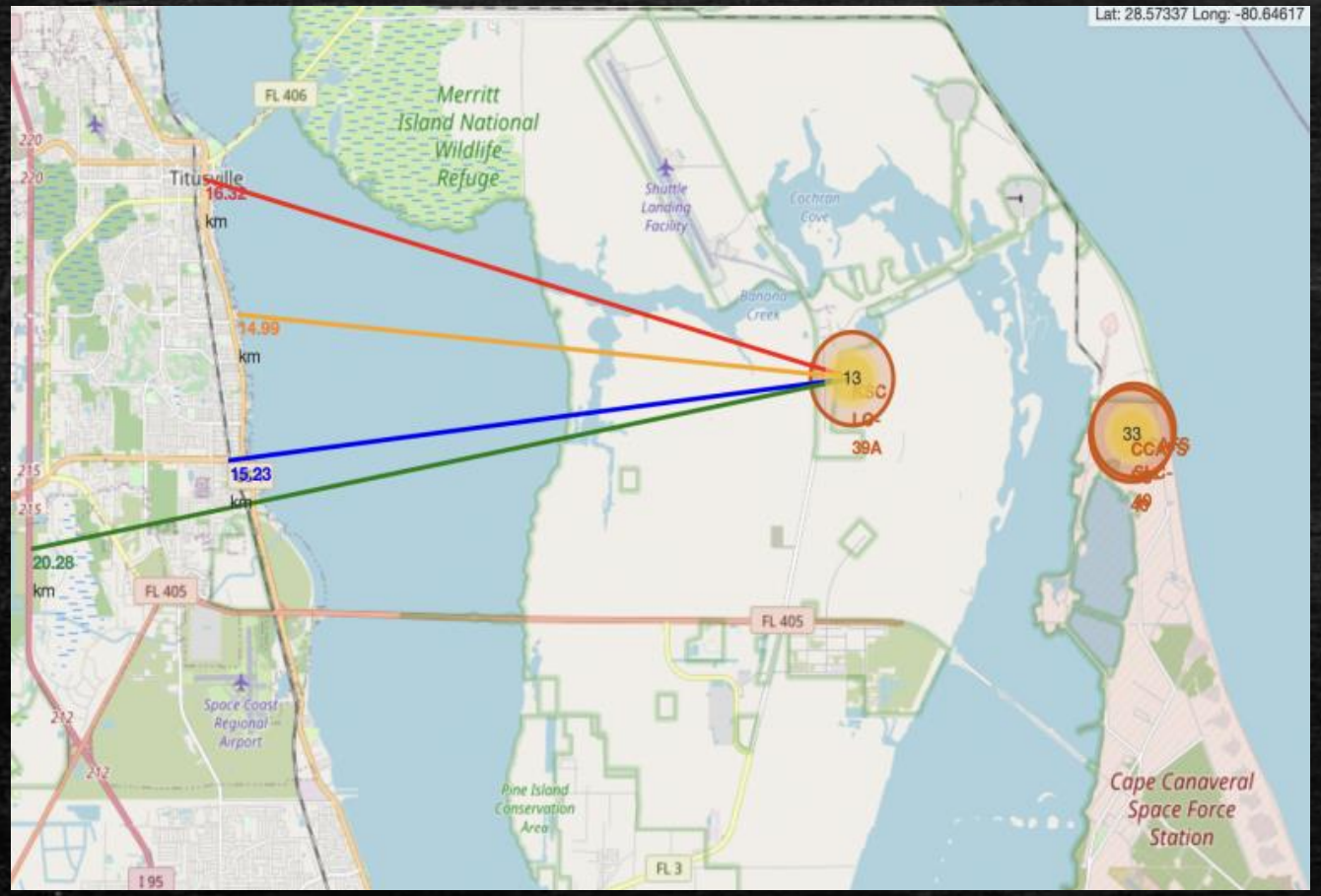
Successful/failed launches for each site on the map



The launches for the launch site KSC LC-39A are shown.
The green markers indicate successful launches.
The red markers indicate failed launches

Results

Distances between a launch site to its proximities

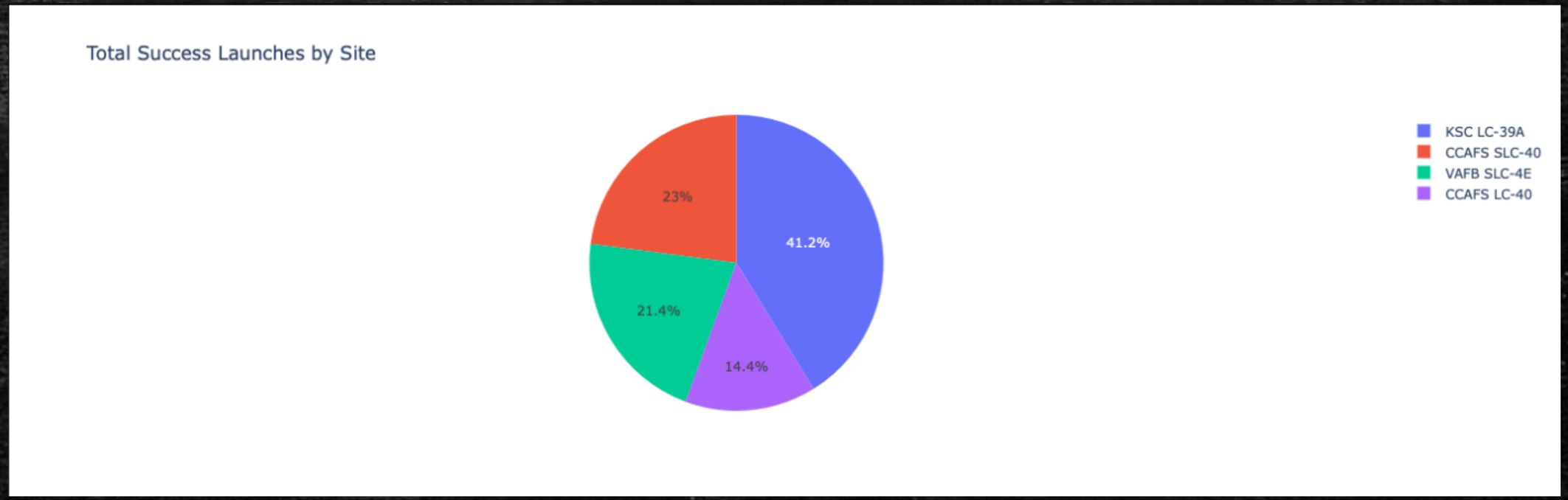


Distances of the launch site e KSC LC-39A are shown. The site is close to railways, highway and the coastline. It's closest city is Titusville.

Plotly Dash Dashboard Results

Results

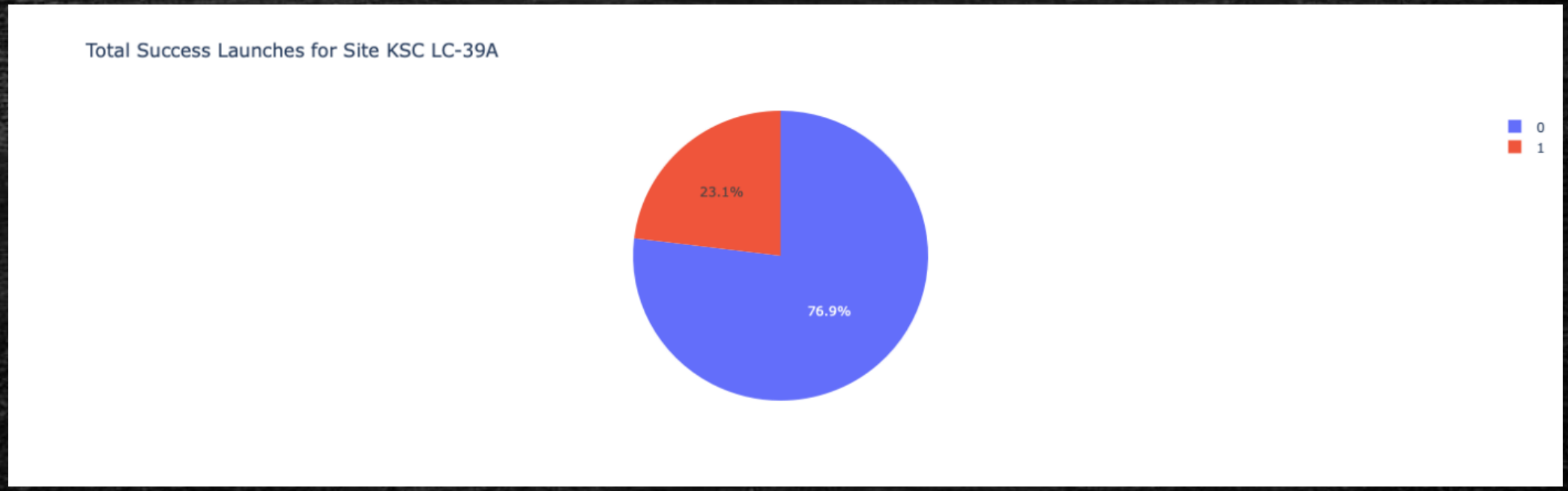
Successful Launches by Site



The launch site KSC LC-39A has the highest success rate among all sites.

Results

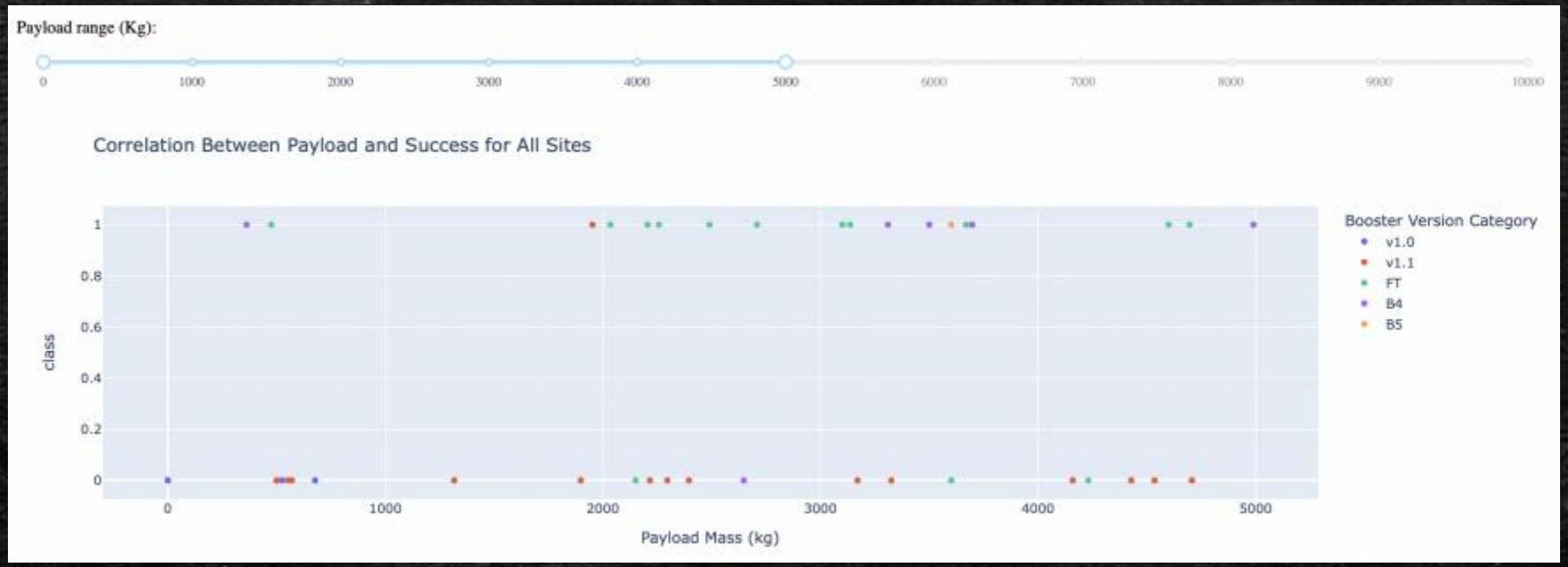
Total Success Launches for KSC LC-39A



KSC LC-39A has a success rate of 76.9%

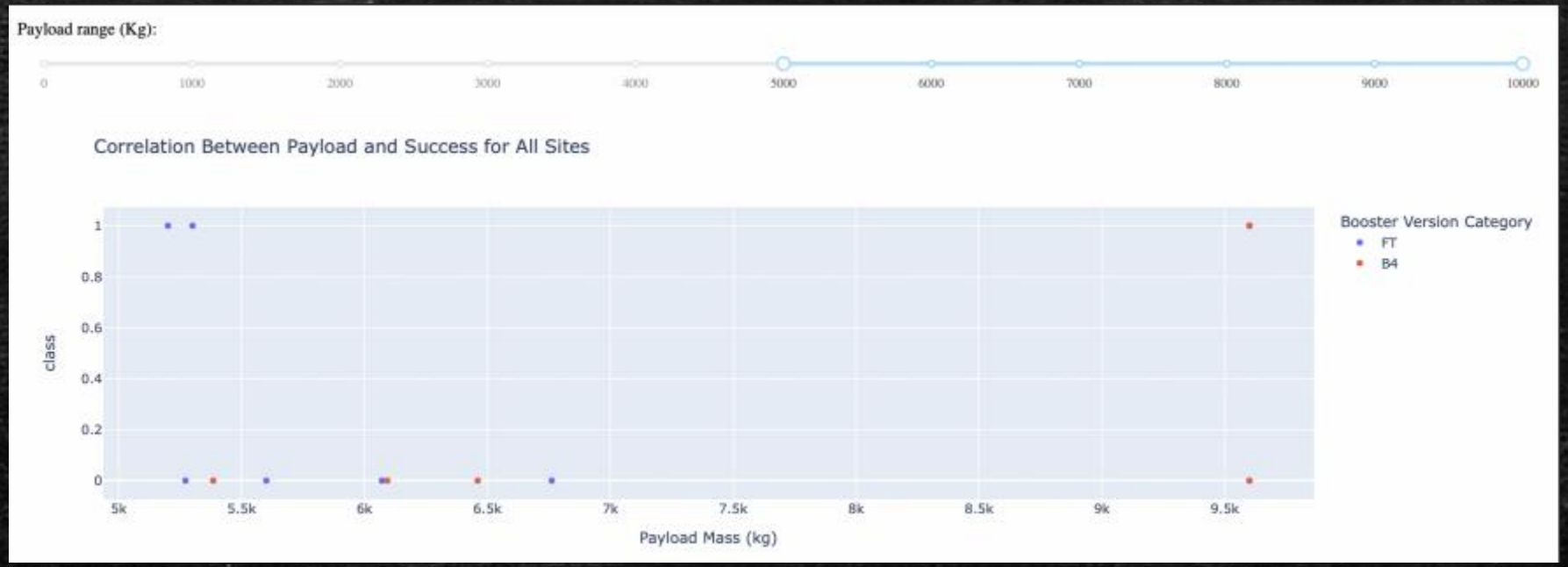
Results

Payload vs. Launch Outcome



Results

Payload vs. Launch Outcome



Predictive Analysis

Results

Classification Accuracy
Accuracy and score for different models:

	LogReg	SVM	Tree	KNN
Jaccard_Score	0.800000	0.800000	0.800000	0.800000
F1_Score	0.888889	0.888889	0.888889	0.888889
Accuracy	0.833333	0.833333	0.833333	0.833333

Four models were trained and tested – Logistic Regression, Support Vector Machine, Decision Tree, KNN.

Results

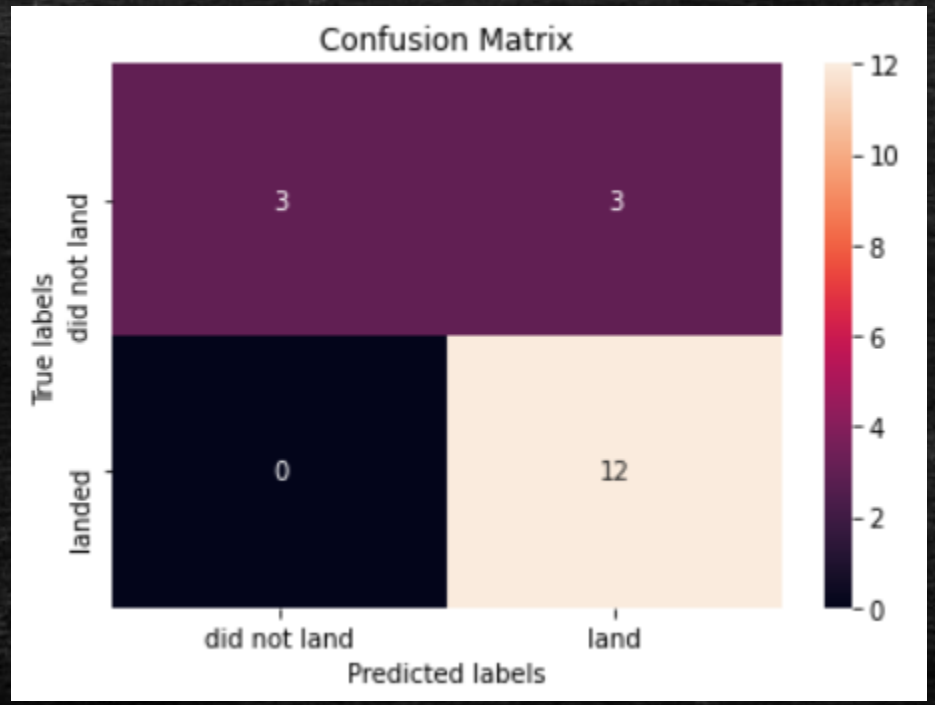
Classification Accuracy

	LogReg	SVM	Tree	KNN
Jaccard_Score	0.833333	0.845070	0.882353	0.819444
F1_Score	0.909091	0.916031	0.937500	0.900763
Accuracy	0.866667	0.877778	0.911111	0.855556

We can see that the decision tree model gives the best accuracy out of all models.

Results

Confusion Matrix:



Conclusion

Conclusion

- The launch site KSC LC-39A has a high success rate for launches
- The success rate has increased over the years probably due to the improvements in the technology.
- Almost all of the launch sites are closer to the coastline.
- Decision tree model turned out to be the best model to predict the success or failure of a launch.

Thank you :)