PROPELLING SPACE EXPLORATION WITH DATA SCIENCE

Krishna Chaitanya Kotabhattara

2023-06-25

OUTLINES

EXECUTIVE SUMMARY

- Data sourced from SpaceX API and web scraping.
- Explored, cleaned, and visualized data to identify key patterns.
- Selected relevant features for predicting SpaceX launch success.
- Built Machine Learning model to predict successful booster landings.
- Conclusions empower SpaceX's market dominance through data-driven decisions.

INTRODUCTION

In an industry as revolutionary and fast-paced as aerospace technology, making accurate predictions is key to staying competitive. This project focuses on Space X, the leading aerospace manufacturer and space transportation service provider

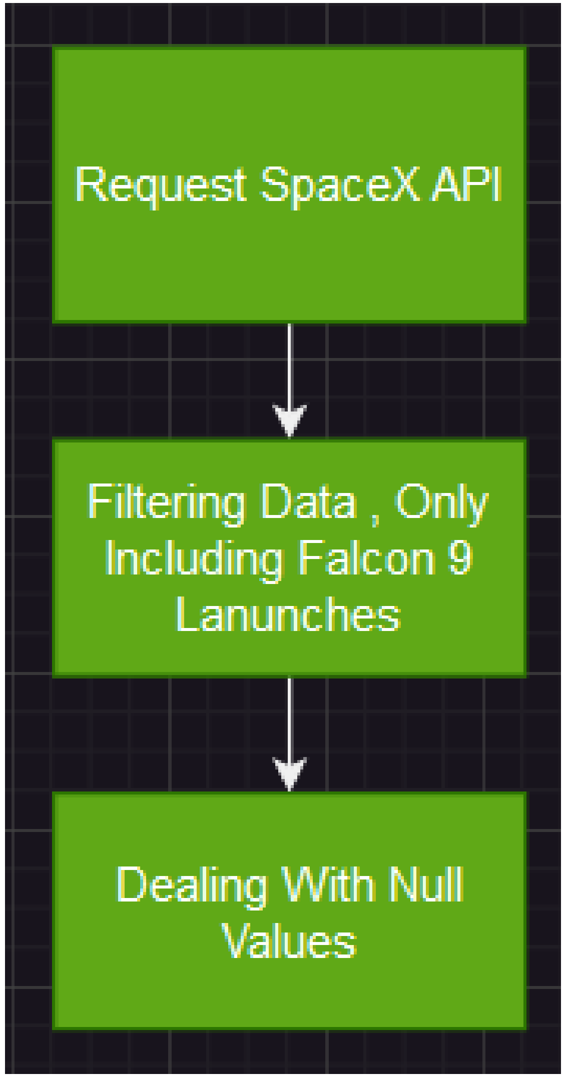
The objective of our analysis is two-fold:

- To predict the success of first stage rocket landings, which play a significant role in cutting the costs of launches.
- To determine the optimal location for these launches to maximize success rates.

METHODOLOGY

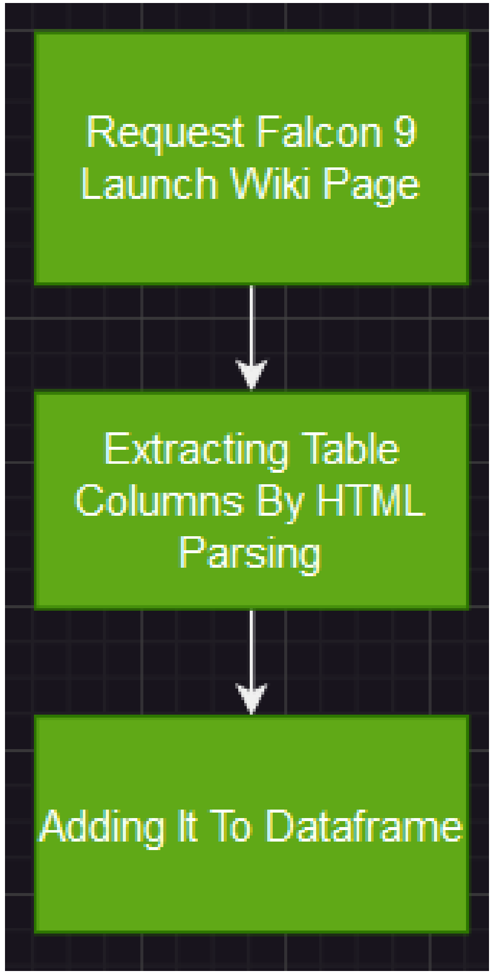
METHODOLOGY OVERVIEW

- Data Collection: Utilized SpaceX API and web scraping (Wikipedia) for primary data collection.
- Data Wrangling: Enriched the data by creating landing outcome labels based on outcome data.
- Exploratory Data Analysis (EDA): Performed EDA using visualization and SQL for initial insights.
- Interactive Visual Analytics: Utilized Folium and Plotly Dash for an interactive representation of our findings.
- Predictive Analysis: Applied classification models on normalized data for landing success prediction. The data was split into training and testing sets. Four different models were evaluated for accuracy with various parameter combinations.



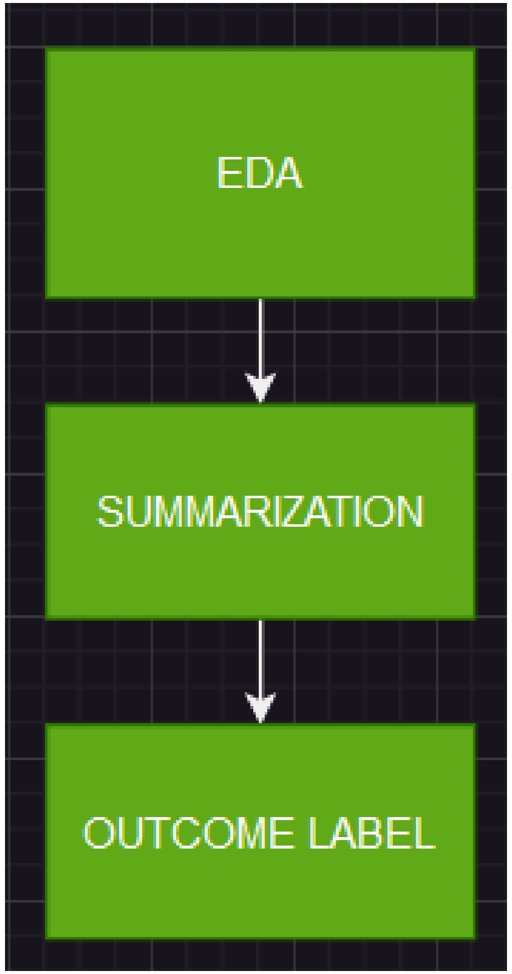
DATA COLLECTION USING SPACEX API

- SpaceX offers a public API
- Data was extracted and stored using the publicly accessible SpaceX API, as outlined in the adjacent flowchart.(https://api.spacexdata.com/v4/rockets/)
- Source Code: https://github.com/DrRuin/IBM-Applied-Data-Science-Capstone/blob/main/Spacex-data-collection-api.ipynb



DATA COLLECTION USING WIKIPEDIA

- Wikipedia is another open source site for information sharing
- Launch data was also sourced from Wikipedia and stored, following the process shown in the flowchart.(https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922)
- Source Code: https://github.com/DrRuin/IBM-Applied-Data-Science-Capstone/blob/main/SpaceX%20Webscraping.ipynb



DATA WRANGLING

- Initial Exploratory Data Analysis (EDA) was executed on the dataset, followed by summarization of launches, orbit occurrences, and mission outcomes. Lastly, a landing outcome label was derived from the Outcome column.
- Source Code: https://github.com/DrRuin/IBM-Applied-Data-Science-Capstone/blob/main/SpaceX%20Data%20Wrangling.ipynb

EDA WITH DATA VISUALIZATION

Scatterplots and barplots visualized relationships between features like:

- Payload Mass & Flight Number, Launch Site & Flight Number, Launch Site & Payload Mass, and Orbit & Flight Number, and Payload & Orbit.

Source Code: https://github.com/DrRuin/IBM-Applied-Data-Science-Capstone/blob/main/EDA_Visualization.ipynb

EDA WITH SQL

1. Identified unique launch sites.
2. Selected top 5 'CCA' sites.
3. Calculated NASA's total payload mass.
4. Computed average payload for F9 v1.1 version.
5. Determined date of first successful ground pad landing.
6. Identified successful drone ship landing boosters carrying 4000-6000 kg payloads.
7. Counted successful and failed mission outcomes.
8. Identified booster versions carrying maximum payload.
9. Investigated 2015's drone ship landing failures with corresponding booster versions and launch sites.
10. Ranked landing outcomes between 2010-06-04 and 2017-03-20.

- Source Code: https://github.com/DrRuin/IBM-Applied-Data-Science-Capstone/blob/main/EDA_SQL.ipynb

INTERACTIVE MAP WITH FOLIUM

- Used markers for points like launch sites.
- Utilized circles for highlighting areas around specific coordinates.
- Implemented marker clusters for group events at each coordinate.
- Employed lines to represent distances between coordinates.

- Source Code: https://github.com/DrRuin/IBM-Applied-Data-Science-Capstone/blob/main/Launch_Site_Location.ipynb

DASHBOARD USING PLOTLY DASH

- Created graphs for visualizing launch percentages by site.
- Constructed plots for payload range.
- Analyzed payload and launch site relations to identify optimal launch locations.

- Source Code: https://github.com/DrRuin/IBM-Applied-Data-Science-Capstone/blob/main/Dash.py

PREDICTIVE ANALYSIS

Four models were compared:

1. Logistic Regression, a statistical model used for predicting the probability of binary outcomes.
2. Support Vector Machine, a type of learning model that's exceptionally efficient in high-dimensional spaces.
3. Decision Tree, a model renowned for its simplicity and usability in both categorical and numerical data.
4. K-Nearest Neighbors, a model rooted in feature similarity, often used in pattern recognition.

Source Code: https://github.com/DrRuin/IBM-Applied-Data-Science-Capstone/blob/main/SpaceX-Final-5.ipynb

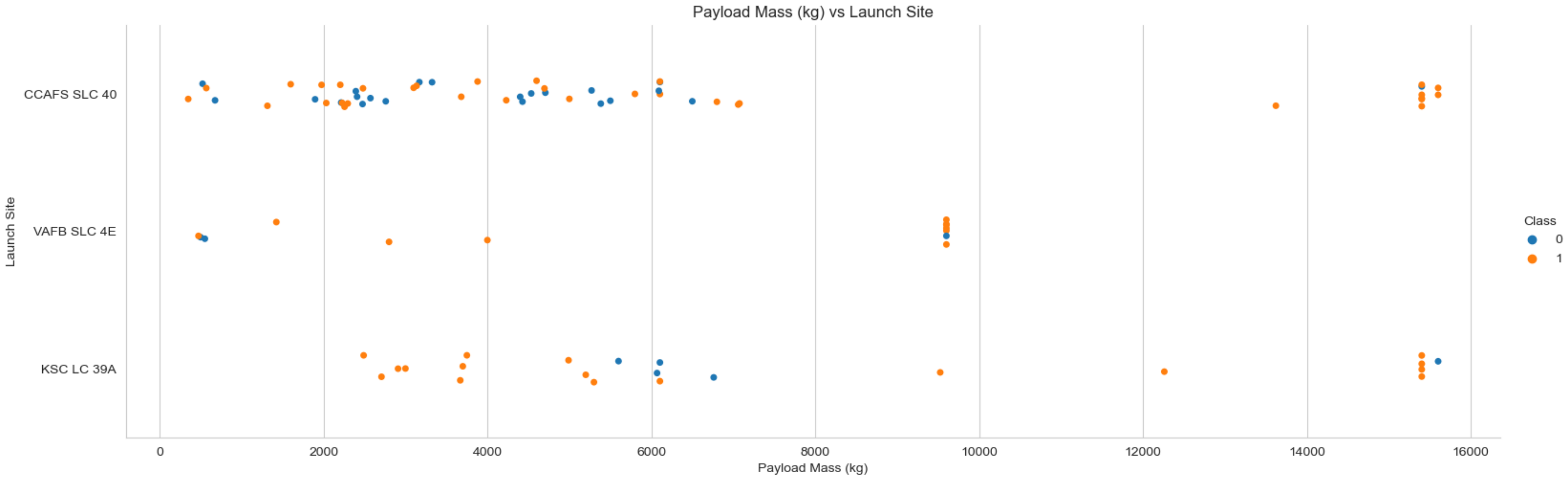
- RESULTS

EXPLORATORY DATA ANALYSIS RESULTS

1. Four launch sites are utilized by SpaceX.
2. SpaceX and NASA were the targets of the initial launches.
3. The average payload for the F9 v1.1 booster is 2,928 kg.
4. The first successful landing happened five years post-initial launch, in 2015.
5. Many Falcon 9 booster versions landed successfully on drone ships, carrying payloads above average.
6. Almost all mission outcomes were successful.
7. Two boosters, F9 v1.1 B1012 and F9 v1.1 B1015, failed to land on drone ships in 2015.
8. Successful landing outcomes increased over the years.
9. Interactive analytics reveal launch sites are typically in safe locations near seas with robust logistics.
10. Most launches occur at east coast launch sites.

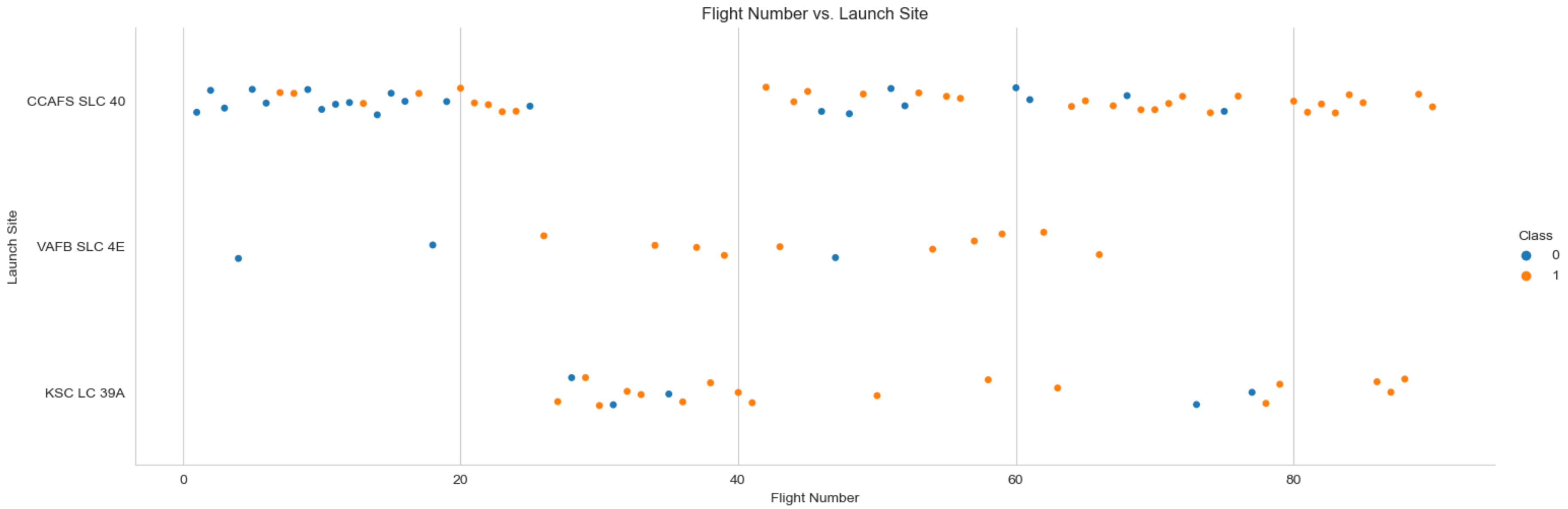
PAYLOAD VS LAUNCH SITE

- Payloads exceeding 9,000kg, roughly equivalent to a school bus, have a high success rate. Payloads surpassing 12,000kg appear feasible only at the CCAFS SLC 40 and KSC LC 39A launch sites.



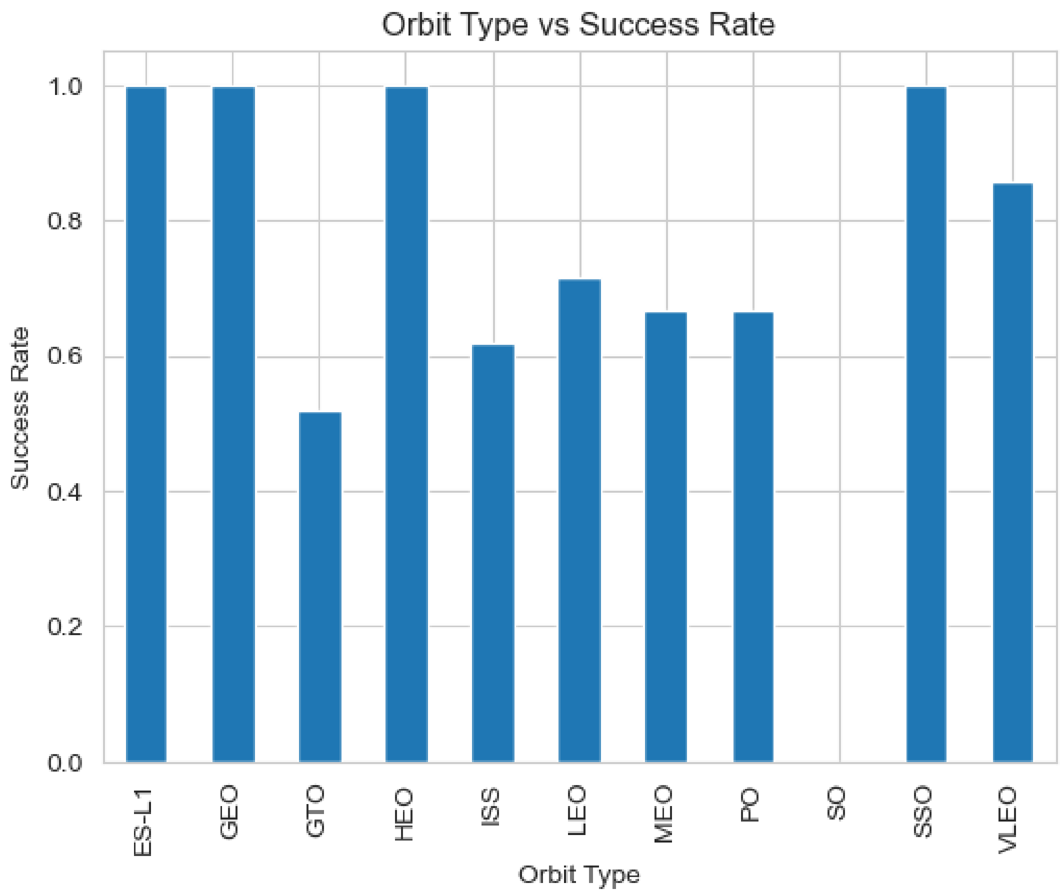
FLIGHT NUMBER VS LAUNCH SITE

- The plot illustrates that the prime launch site is currently CCAFS SLC 40, which boasts the most recent successful launches. VAFB SLC 4E ranks second, followed by KSC LC 39A in third place. A general improvement in success rate over time is also evident.



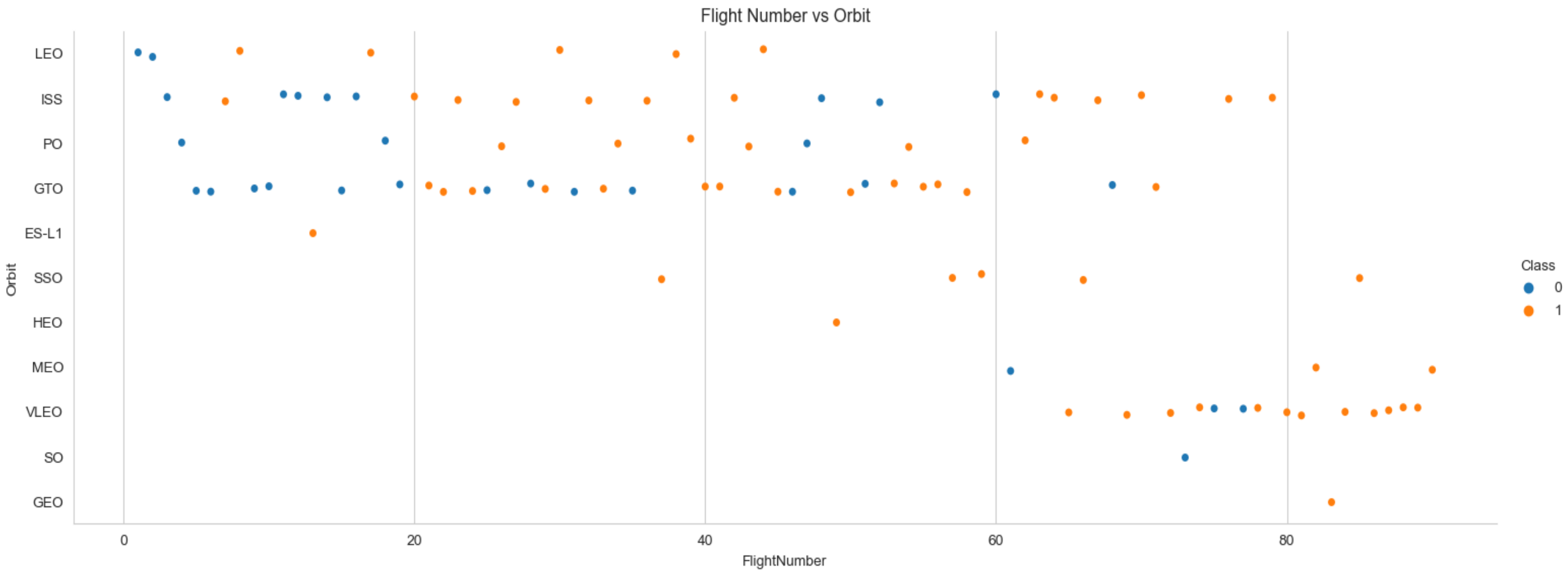
ORBIT TYPE VS SUCCESS RATE

- The highest success rates are associated with ES-L1, GEO, HEO, and SSO orbits. VLEO and LEO orbits follow, with success rates above 80% and 70% respectively.



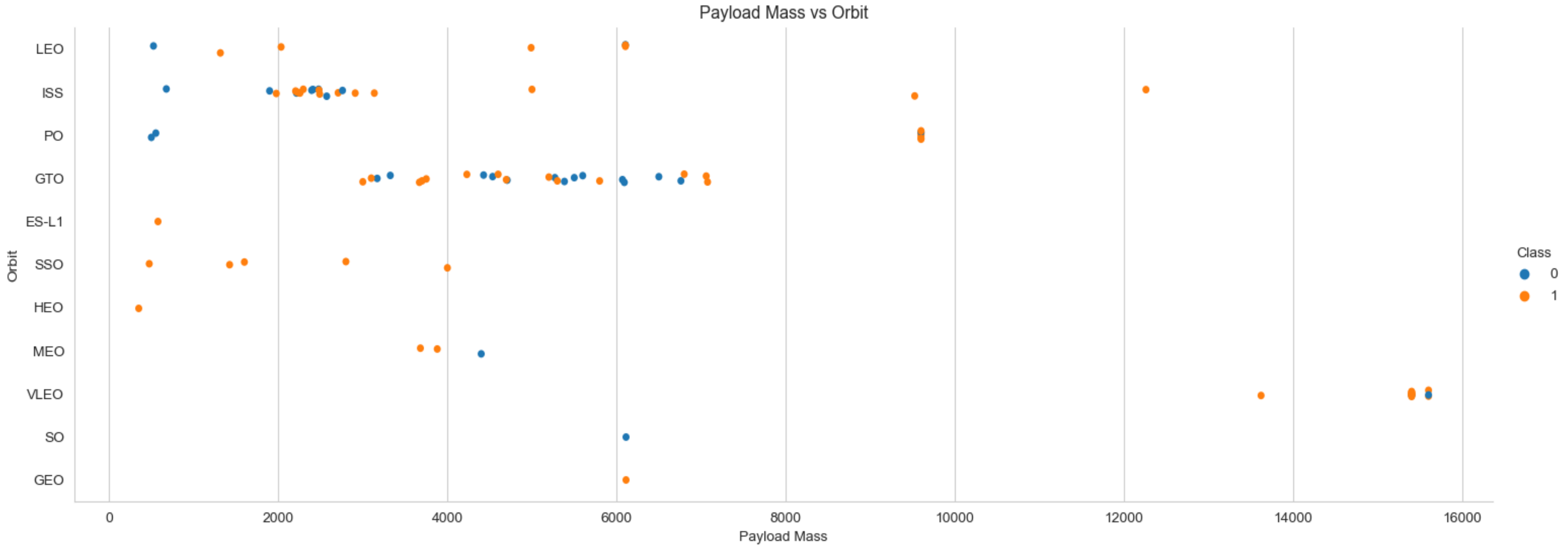
FLIGHT NUMBER VS ORBIT TYPE

- The success rate for all orbits seems to have improved over time. VLEO orbits, in particular, appear to present a new business opportunity due to their recent frequency increase.



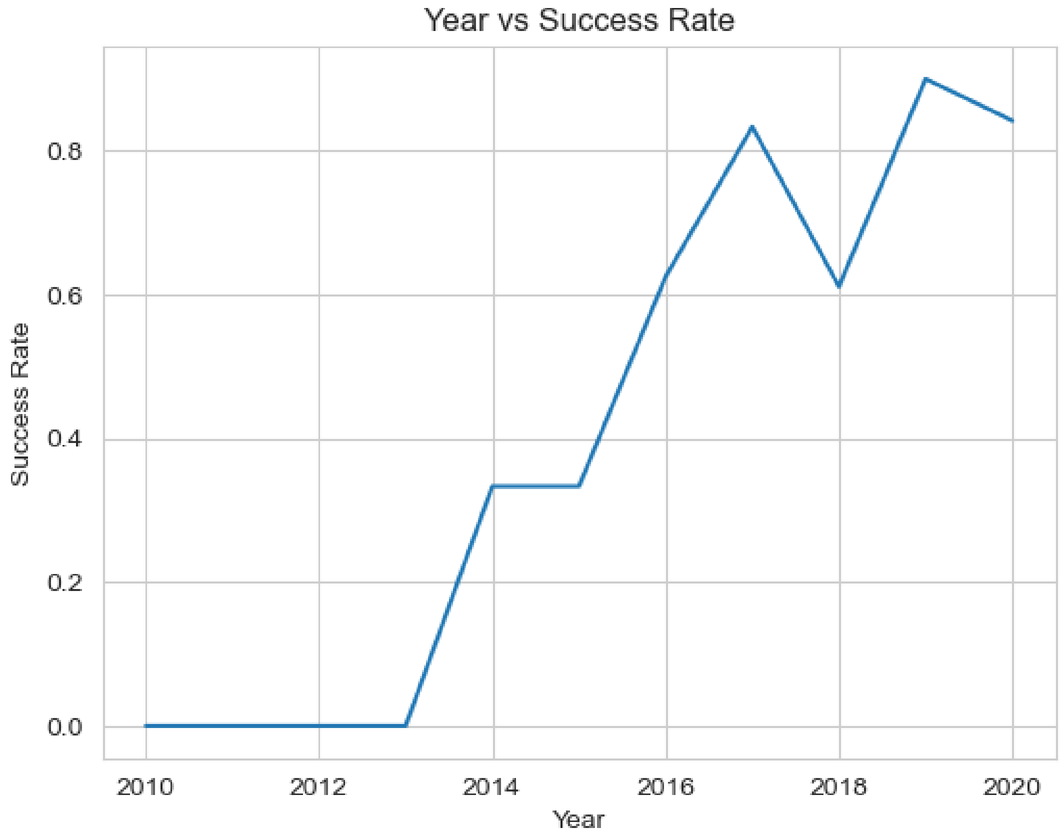
Payload vs Orbit Type

- There seems to be no correlation between payload and success rate for GTO orbit. ISS orbit, with a diverse range of payloads, maintains a strong success rate. Meanwhile, SO and GEO orbits experience relatively fewer launches.



LAUNCH SUCCESS YEARLY TREND

- Success rates began to rise from 2013 and continued until 2020, suggesting that the initial three years served as a critical period for technological adjustments and enhancements.



ALL POSSIBLE LAUNCH SITE NAMES

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

- The data reveals four distinct launch sites, determined by identifying unique "launch_site" values within the dataset.

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	**CCA**FS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-12-08	15:43:00	F9 v1.0 B0004	**CCA**FS 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	**CCA**FS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	F9 v1.0 B0006	**CCA**FS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	**CCA**FS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attemp

TOTAL PAYLOAD MASS BY NASA (CRS) AND AVG PAYLOAD MASS BY F9 v1.1

sum(PAYLOAD_MASS_KG_)
45596.0

avg(PAYLOAD_MASS_KG_)
2928.4

FIRST SUCCESSFUL GROUND LANDING DATE

- First successful landing outcome on ground pad

FirstSuccessfull_landing_date
01/08/2018

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

SUCCESSFUL DRONE SHIP LANDING WITH PAYLOAD BETWEEN 4000 AND 6000

TOTAL NUMBER OF SUCCESSFUL AND FAILURE MISSION OUTCOMES

Mission_Outcome	count
None	0
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

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

2015 LAUNCH RECORDS

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

Date	Booster_Version	Launch_Site	Landing_Outcome
01/10/2015	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
14/04/2015	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

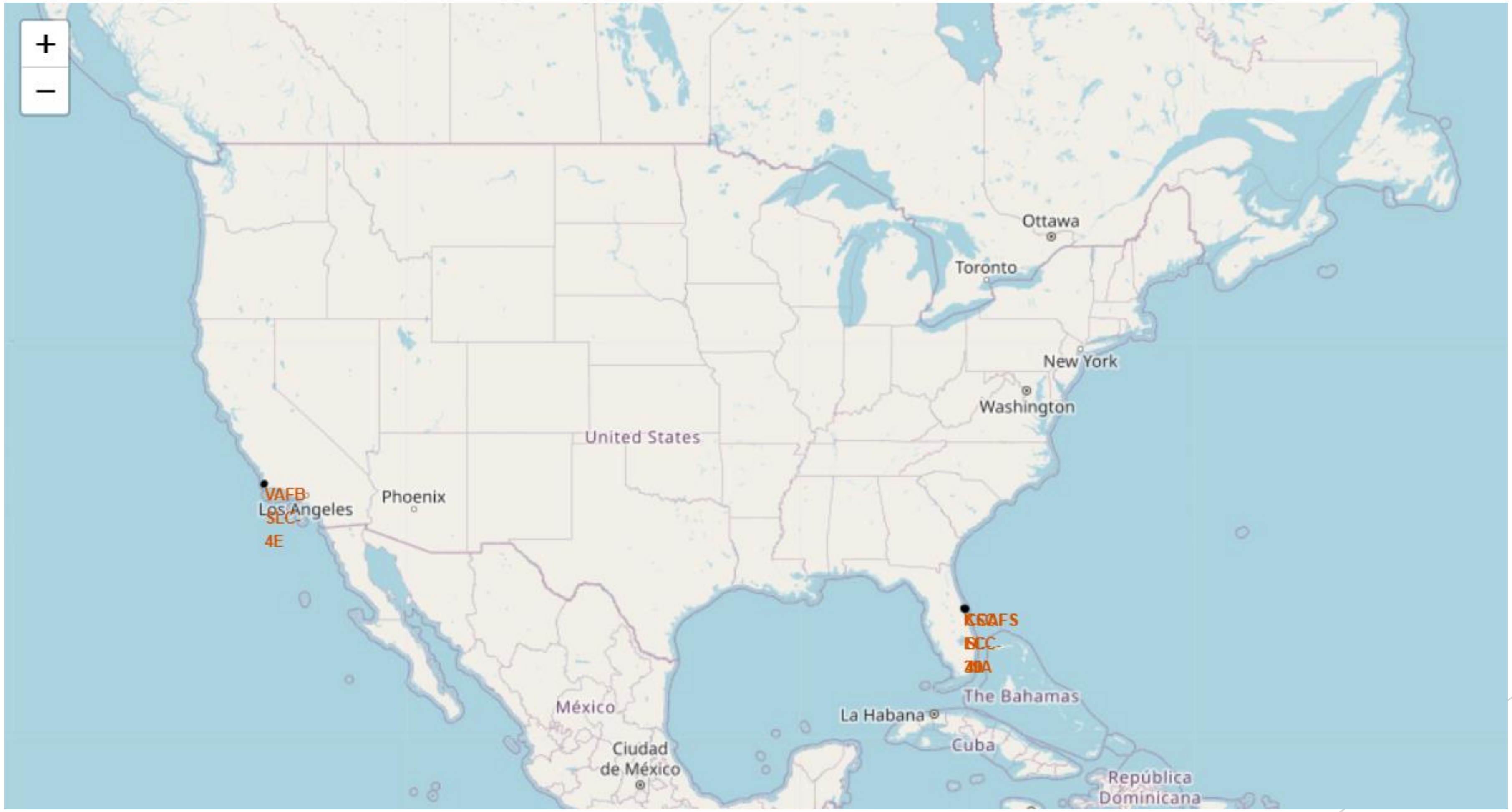
RANK LANDING OUTCOMES BETWEEN 2010-06-04 AND 2017-03-20

Data alerts us that “No attempt” must be taken in account.

	LANDING_OUTCOME	COUNT
3	No attempt	10
1	Failure (drone ship)	5
5	Success (drone ship)	5
6	Success (ground pad)	5
0	Controlled (ocean)	3
7	Uncontrolled (ocean)	2
2	Failure (parachute)	1
4	Precluded (drone ship)	1

LAUNCH SITE PROXIMITY ANALYSIS

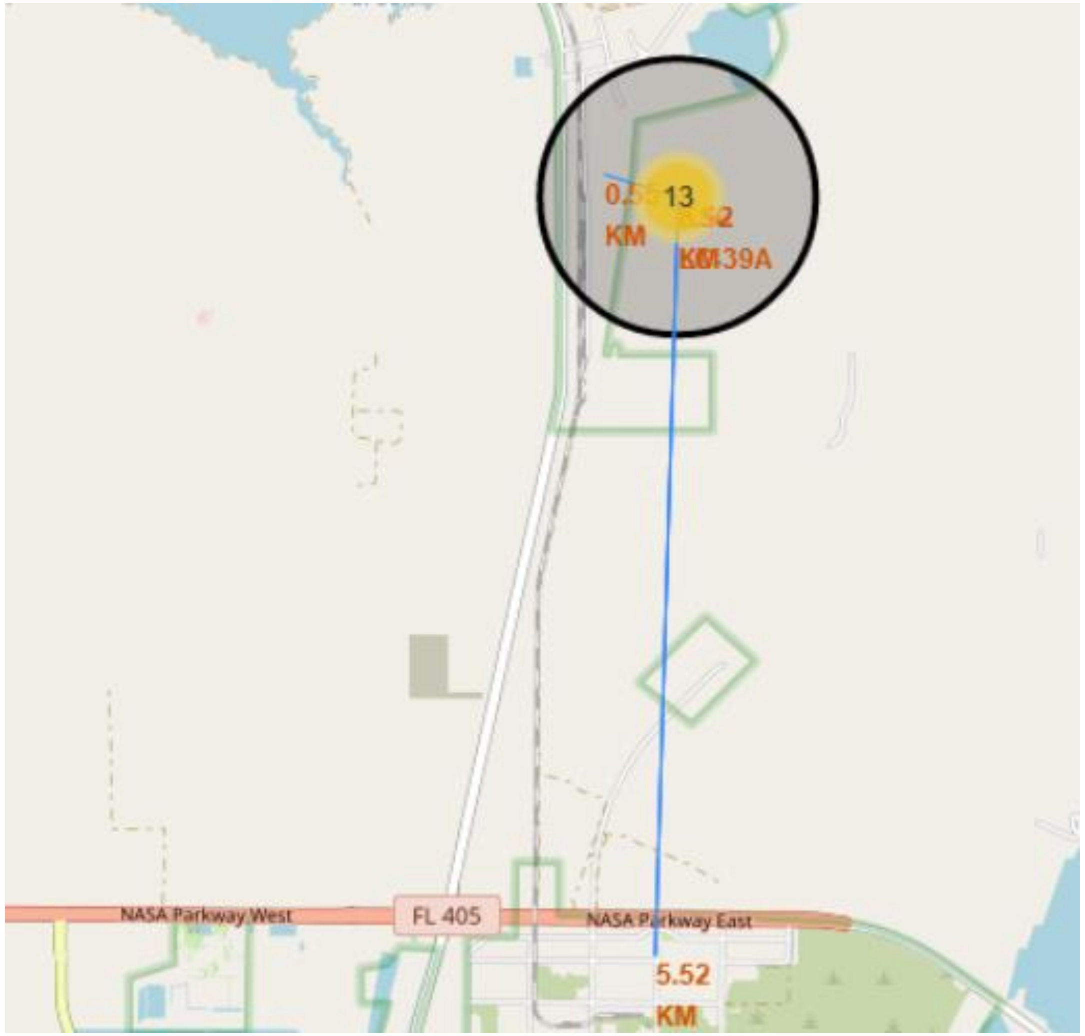
ALL LAUNCH SITES



LAUNCH OUTCOMES BY SITES

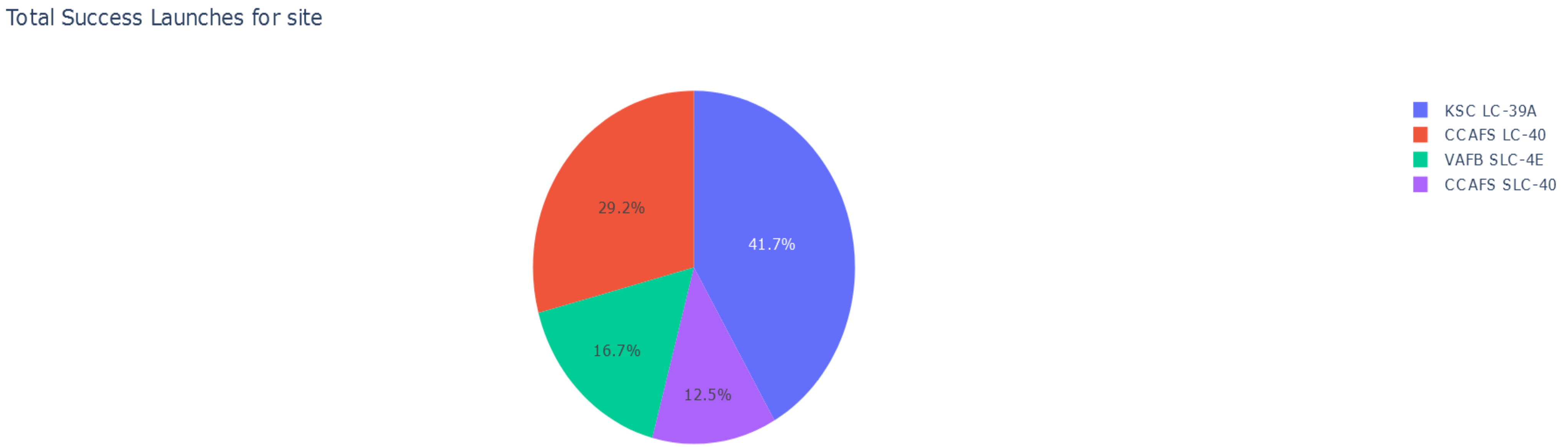
LOGISTIC ASPECT

- Launch site KSC LC-39A has good logistics aspects, being near railroad and road and relatively far from inhabited areas.

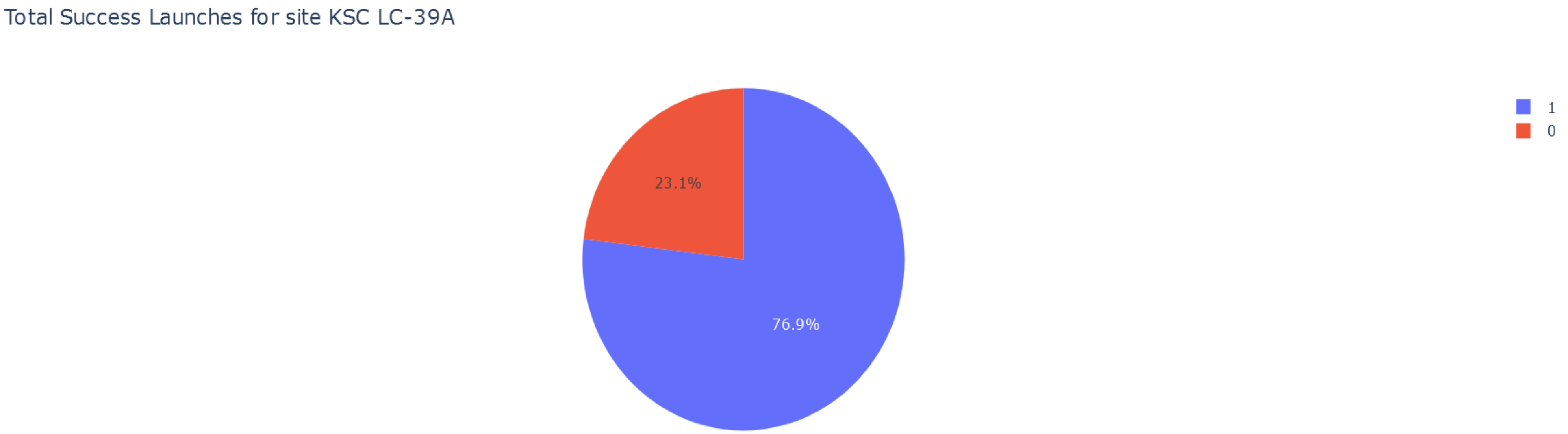


DASHBOARD USING PLOTLY DASH

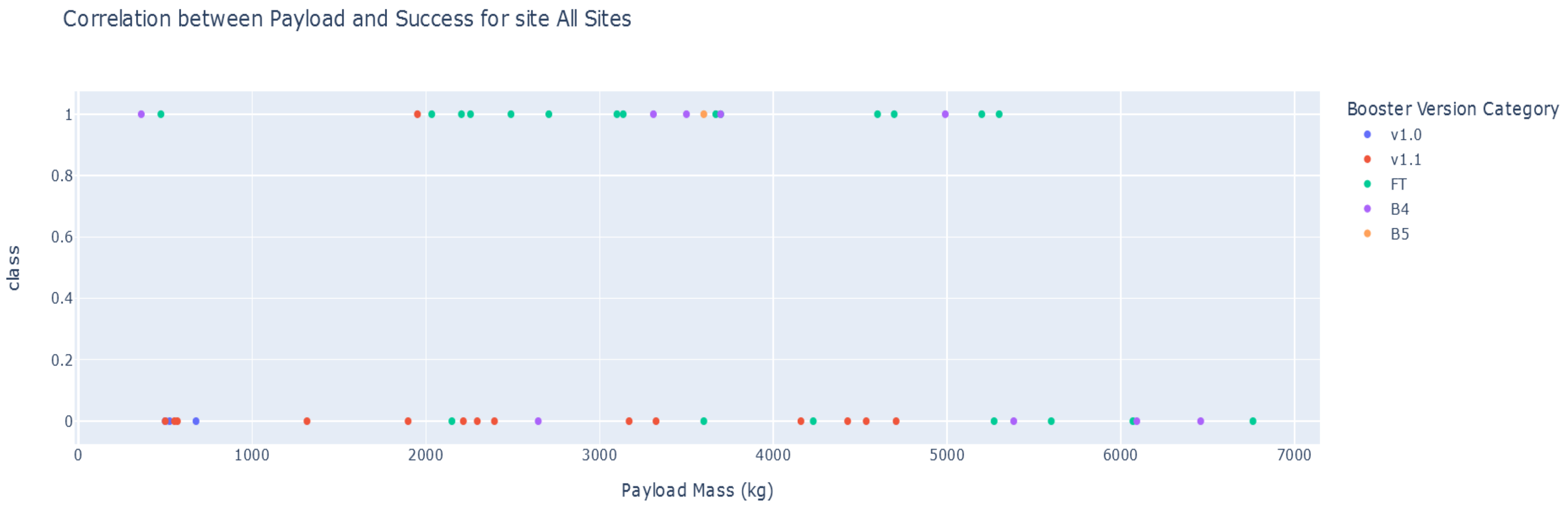
SUCCESSFUL LAUNCHES BY SITE



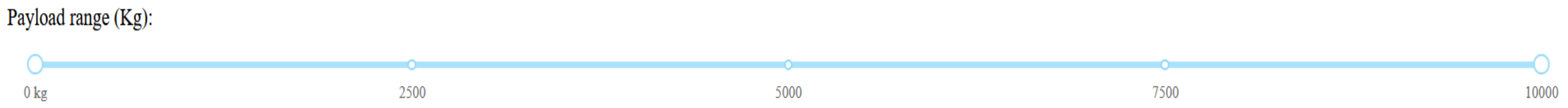
LAUNCH SUCCESS RATIO FOR KSC LC-39A



CORRELATION BETWEEN PAYLOAD AND SUCCESS FOR ALL SITES



PAYLOAD RANGE SEEKBAR



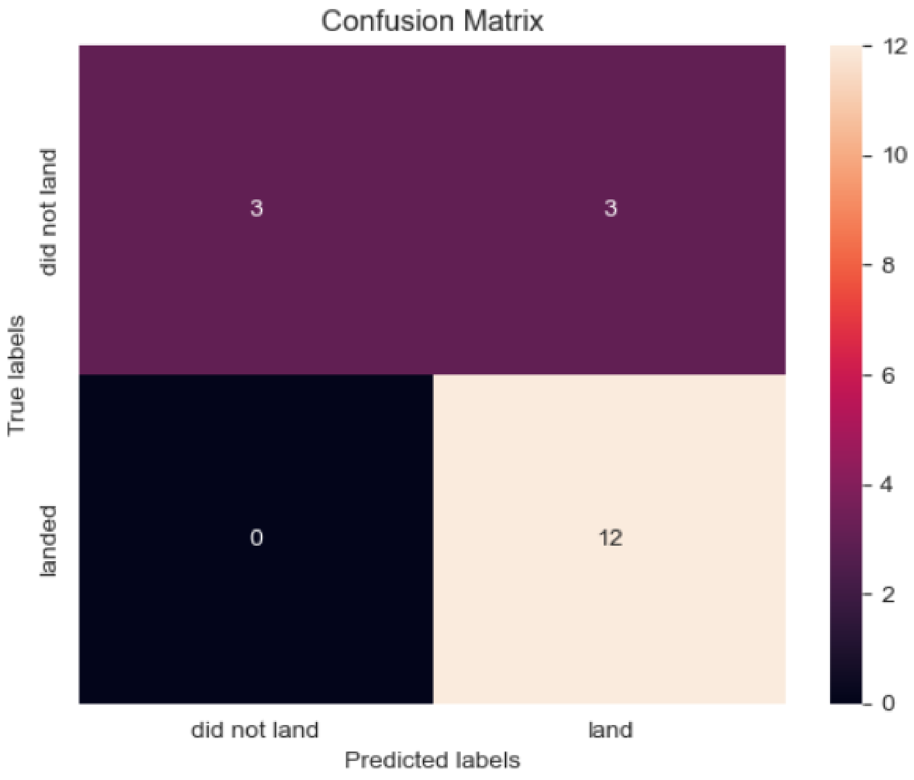
PREDICTIVE ANALYSIS

In predictive analysis we used 4 algorithms :

1. Logreg
2. SVM
3. Tree
4. KNN

It was noticed that all of them had similar accuracy and were pretty much same. Decision Tree was slightly 0.01% better

CONFUSION MATRIX



Find the method performs best:

- Accuracy of all algorithms is above 80%
- Confusion Matrix is consistent
- Results of all algorithms remain consistently effective

Overall summary , All of them perform and give same result

OVERALL PREDICTIVE ANALYSIS RESULT

CONCLUSION

1. Data from multiple sources were scrutinized for insights.
2. KSC LC-39A is identified as the optimal launch site.
3. Launches over 7,000kg exhibit lower risk.
4. Mission success, especially landings, is improving over time.
5. Decision Tree Classifier shows promise in predicting successful landings.

APPENDIX

- Predictive analysis notebook had some validation error , therefore i had to make a new notebook.
- Folium maps didn’t show up on github , so i implemented and posted them on this presentation.
- It was noticed that setting a value for np.random.seed could enhance model testing.

THANK YOU , THE END