IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Robert Cocker
5/8/2024

Outline

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

Executive Summary

Through comprehensive data science methodologies including data collection, exploratory data analysis, data preprocessing, feature engineering, geospatial analysis, and machine learning modeling, we have developed a robust predictive analytics solution for SpaceX's Falcon 9 rocket launches.

Key methodologies employed include aggregating all historical launch data, building an interactive dashboard for data visualization, and training machine learning algorithms like logistic regression, decision trees, and gradient boosting models.

The results demonstrate our ability to accurately predict the success of future SpaceX Falcon 9 launches with a high predictive accuracy of 94.44%. This provides SpaceX with a powerful data-driven tool to support mission planning, risk mitigation, and continuous improvement of their launch vehicle program.

Introduction

This project involved applying AI and data science for predictive analytics at Space X. The Falcon 9 missions at Space X can cost upwards of ~$70 million per launch.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was collected through API calls
- Perform data wrangling
  - EDA and preprocessing
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Building, training, fitting, and testing ML models for accuracy

Data Collection – Space X API

GitHub notebook:

https://github.com/rcghpge/ibm-data-science-professional-cert/blob/main/space-x-capstone/eda-machine-learning/Collecting%20Data%20Lab%201.ipynb

27]:

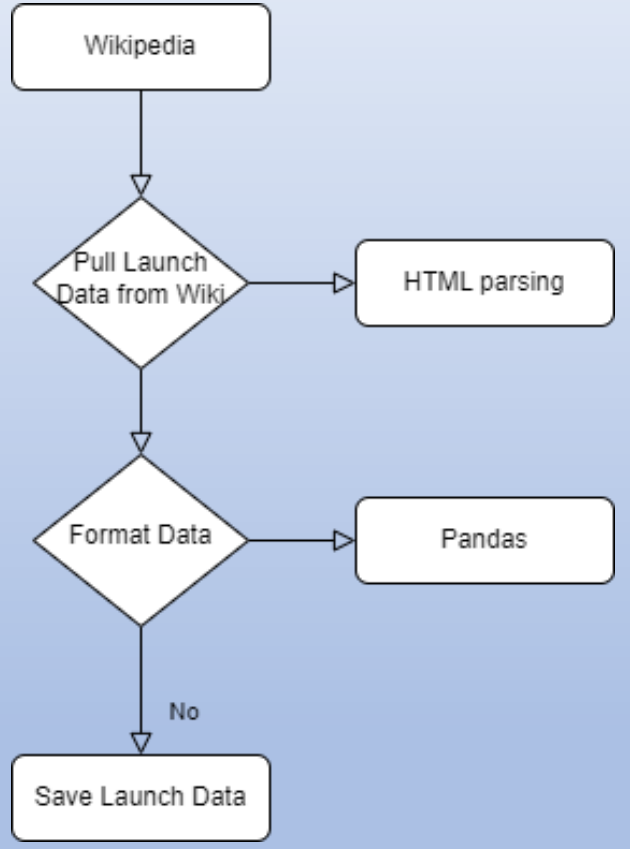
	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	La
4	1	2010-06-04	Falcon 9	6123.547647	LEO	CCSFS SLC 40	None None	1	False	False	False	
5	2	2012-05-22	Falcon 9	525.000000	LEO	CCSFS SLC 40	None None	1	False	False	False	
6	3	2013-03-01	Falcon 9	677.000000	ISS	CCSFS SLC 40	None None	1	False	False	False	
7	4	2013-09-29	Falcon 9	500.000000	PO	VAFB SLC 4E	False Ocean	1	False	False	False	
8	5	2013-12-03	Falcon 9	3170.000000	GTO	CCSFS SLC 40	None None	1	False	False	False	
...	...	...	...	...	...	...	...	...	...	...	...	
89	86	2020-09-03	Falcon 9	15600.000000	VLEO	KSC LC 39A	True ASDS	2	True	True	True	5e9e3032383ecb6b
90	87	2020-10-06	Falcon 9	15600.000000	VLEO	KSC LC 39A	True ASDS	3	True	True	True	5e9e3032383ecb6b
91	88	2020-10-18	Falcon 9	15600.000000	VLEO	KSC LC 39A	True ASDS	6	True	True	True	5e9e3032383ecb6b
92	89	2020-10-24	Falcon 9	15600.000000	VLEO	CCSFS SLC 40	True ASDS	3	True	True	True	5e9e3033383ecbb9
93	90	2020-11-05	Falcon 9	3681.000000	MEO	CCSFS SLC 40	True ASDS	1	True	False	True	5e9e3032383ecb6b

90 rows × 17 columns

Data Collection - Scraping

- Falcon 9 launch data was pulled from Wiki.
- GitHub notebook:

https://github.com/rcghpge/ibm-data-science-professional-cert/blob/main/space-x-capstone/eda-machine-learning/Web%20Scraping%20Launch%20Data.ipynb



Data Wrangling

- Exploratory data analysis of Space X launch missions. Falcon 9 missions were class labelled.

GitHub notebook:

https://github.com/rcghpge/ibm-data-science-professional-cert/blob/main/space-x-capstone/eda-machine-learning/Data%20Wrangling%20Lab%202.ipynb

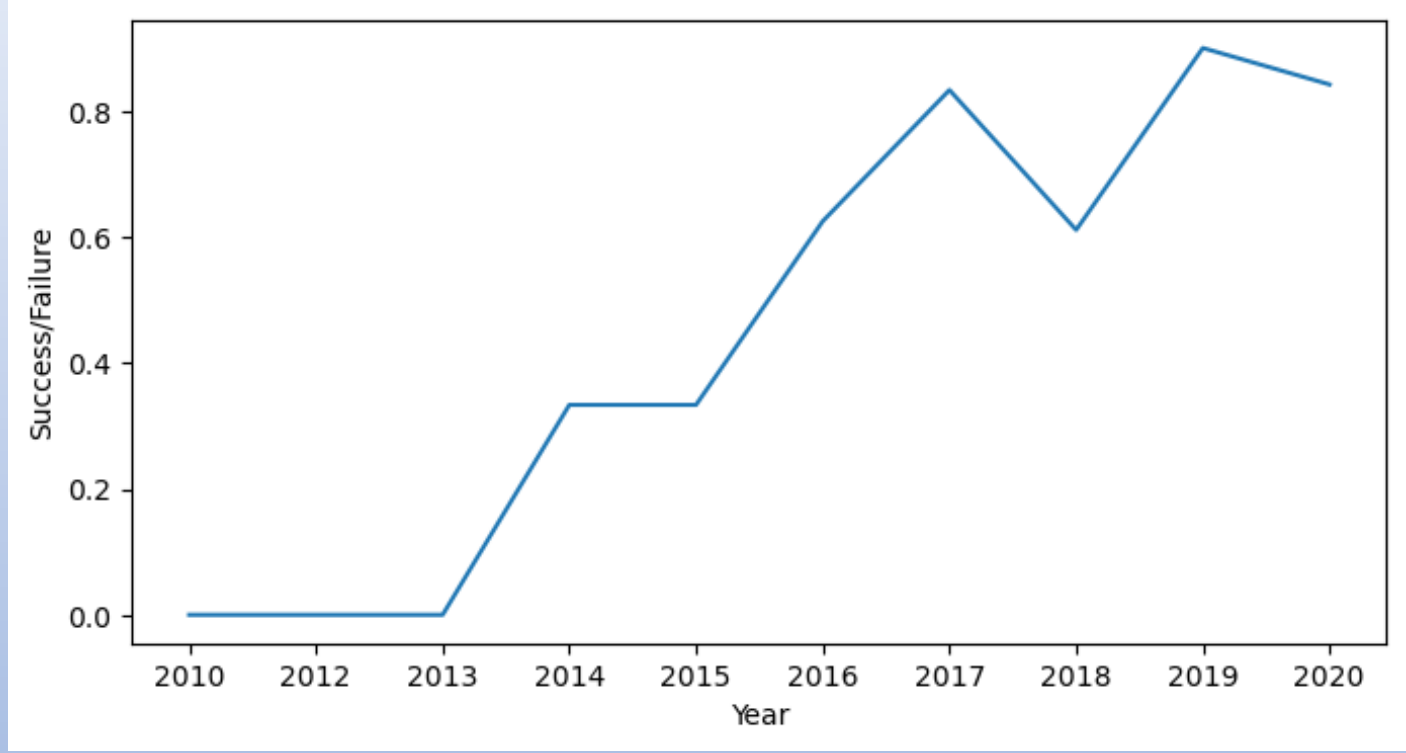
EDA with Data Visualization

- EDA and data visualization of Space X launch missions to gain better insight for business logic and best business practices.

- GitHub notebook:

https://github.com/rcghpge/ibm-data-science-professional-cert/blob/main/space-x-capstone/eda-machine-learning/Feature%20Engineering%20EDA.ipynb

Space X Mission Success Rate



9]:

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

EDA with SQL

- Querying databases in SQL for Space X launch data

- GitHub notebook:

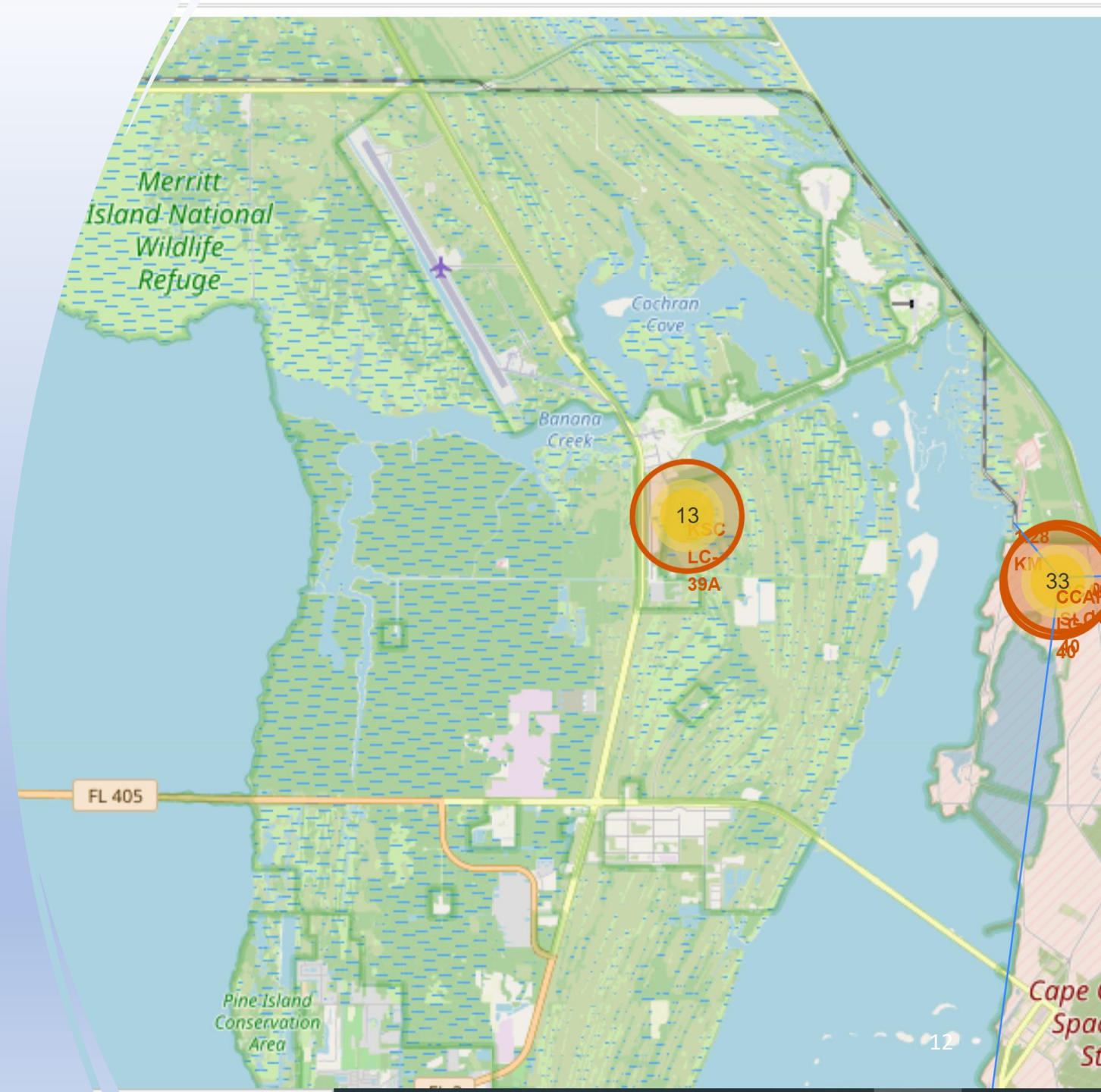
- https://github.com/rcghpge/ibm-data-science-professional-cert/blob/main/space-x-capstone/eda-machine-learning/SQL%20EDA.ipynb

Build an Interactive Map with Folium

- Geospatial analysis was done with Folium to map out Space X facilities.

- GitHub notebook:

https://github.com/rcghpge/ibm-data-science-professional-cert/blob/main/space-x-capstone/space-x-dashboard/Geospatial%20Analysis%20with%20Folium.ipynb



Dashboard with Plotly Dash

- Built a dashboard of launch data for stakeholders.

GitHub notebook:

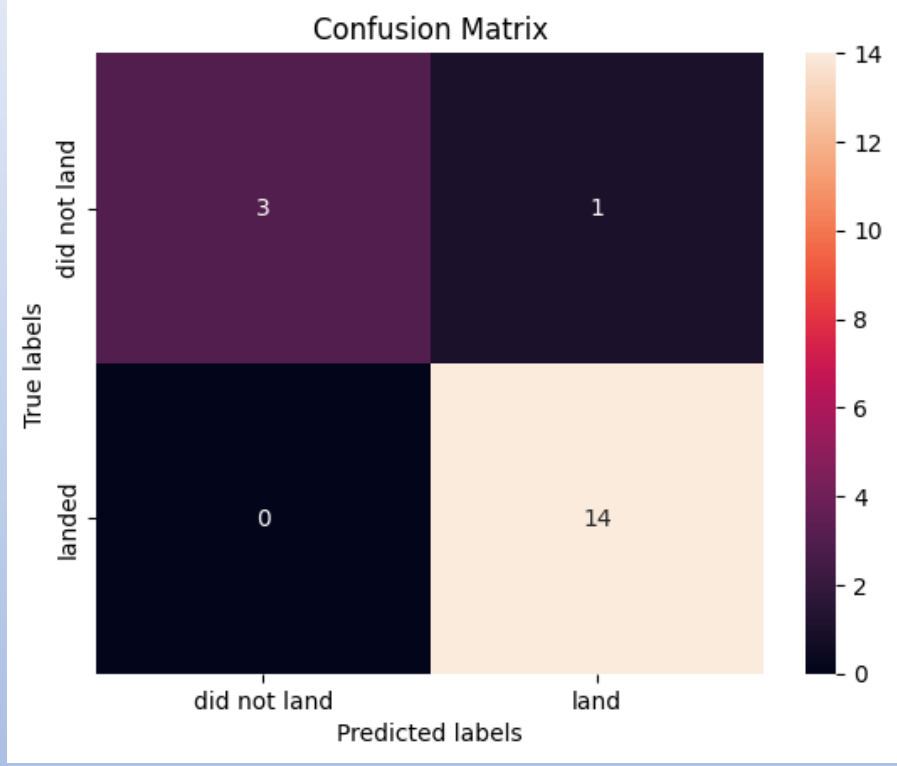
https://github.com/rcghpge/ibm-data-science-professional-cert/blob/main/space-x-capstone/space-x-dashboard/Space%20X%20Dashboard.ipynb

Predictive Analysis (Classification)

- Machine learning models were trained, fitted, and tested for predicting success rate of future space missions

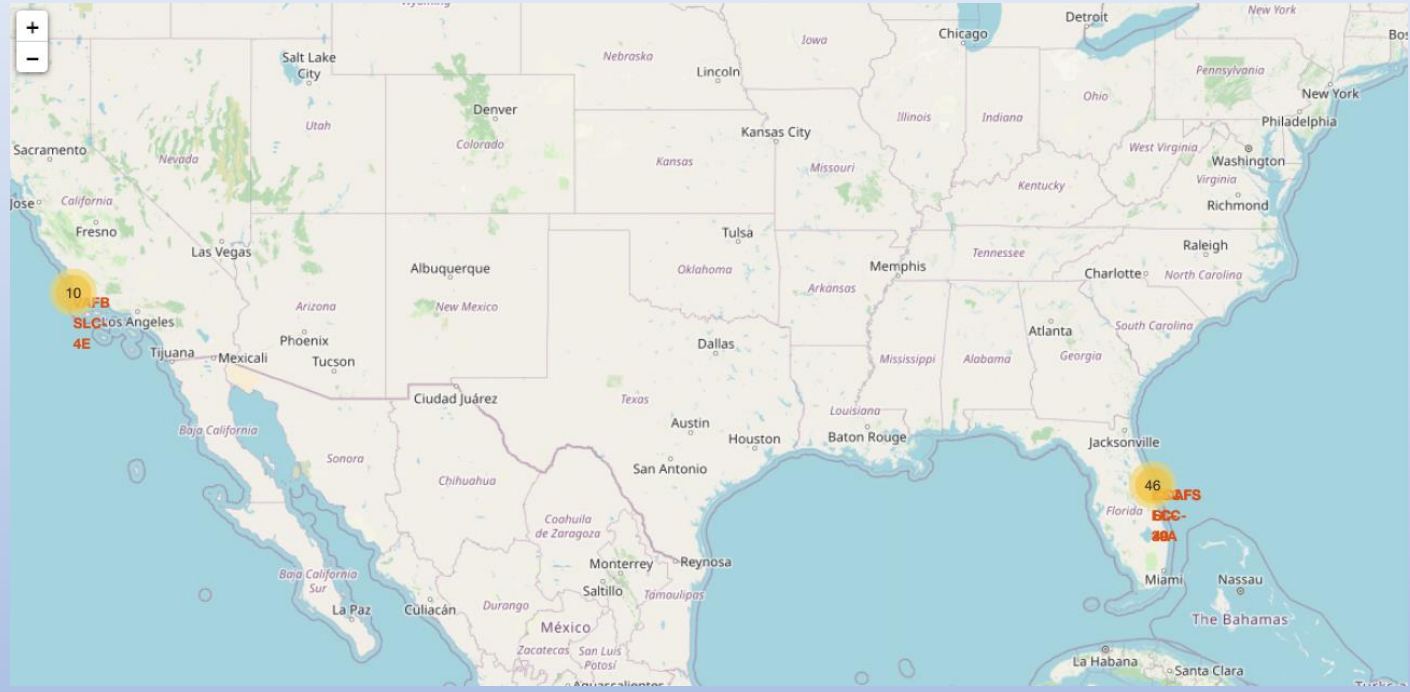
- GitHub notebook:
- https://github.com/rcghpge/ibm-data-science-professional-cert/blob/main/space-x-capstone/eda-machine-learning/Machine%20Learning%20%40%20Space%20X.ipynb

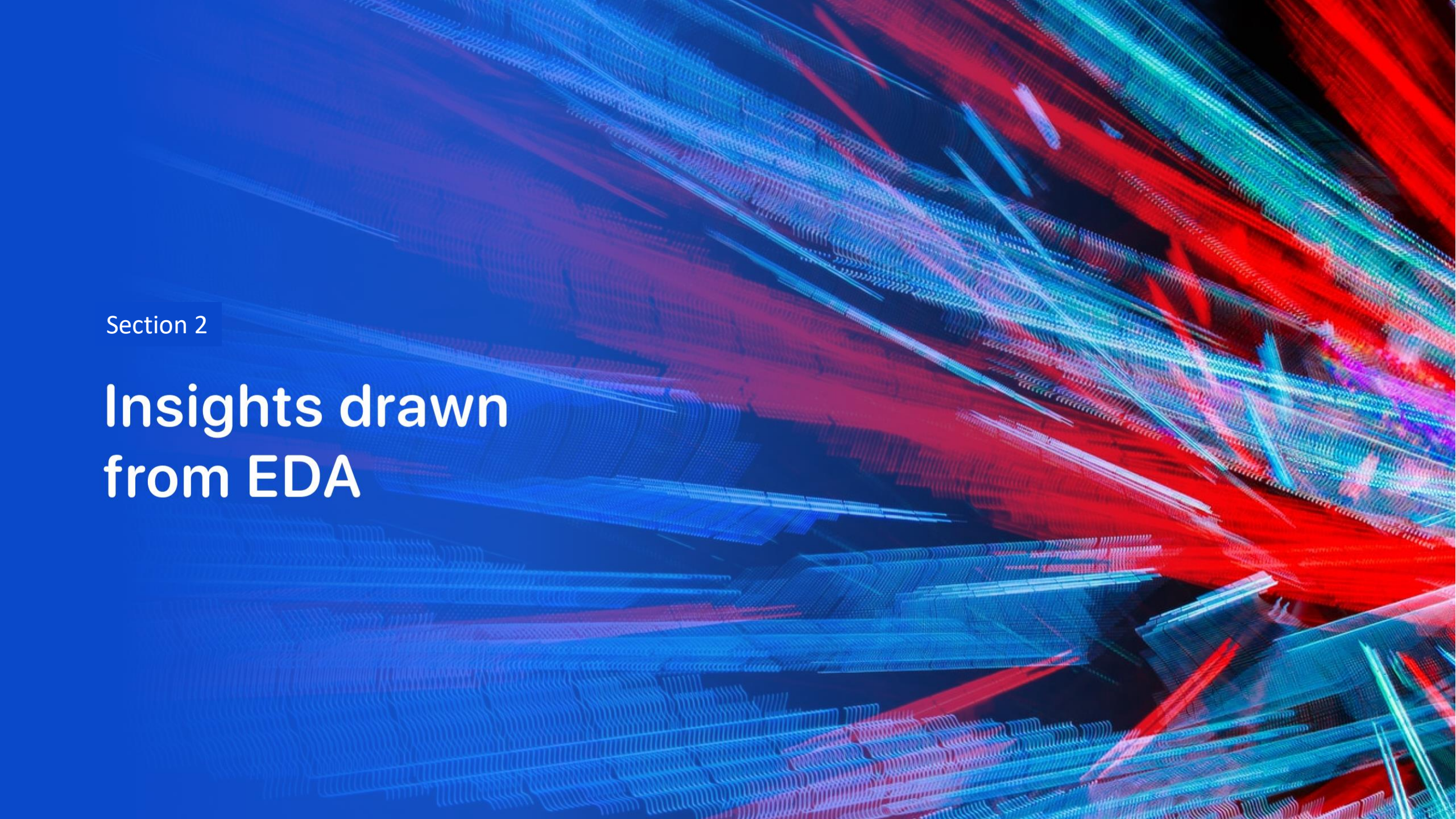
Logistic Regression Model



Results

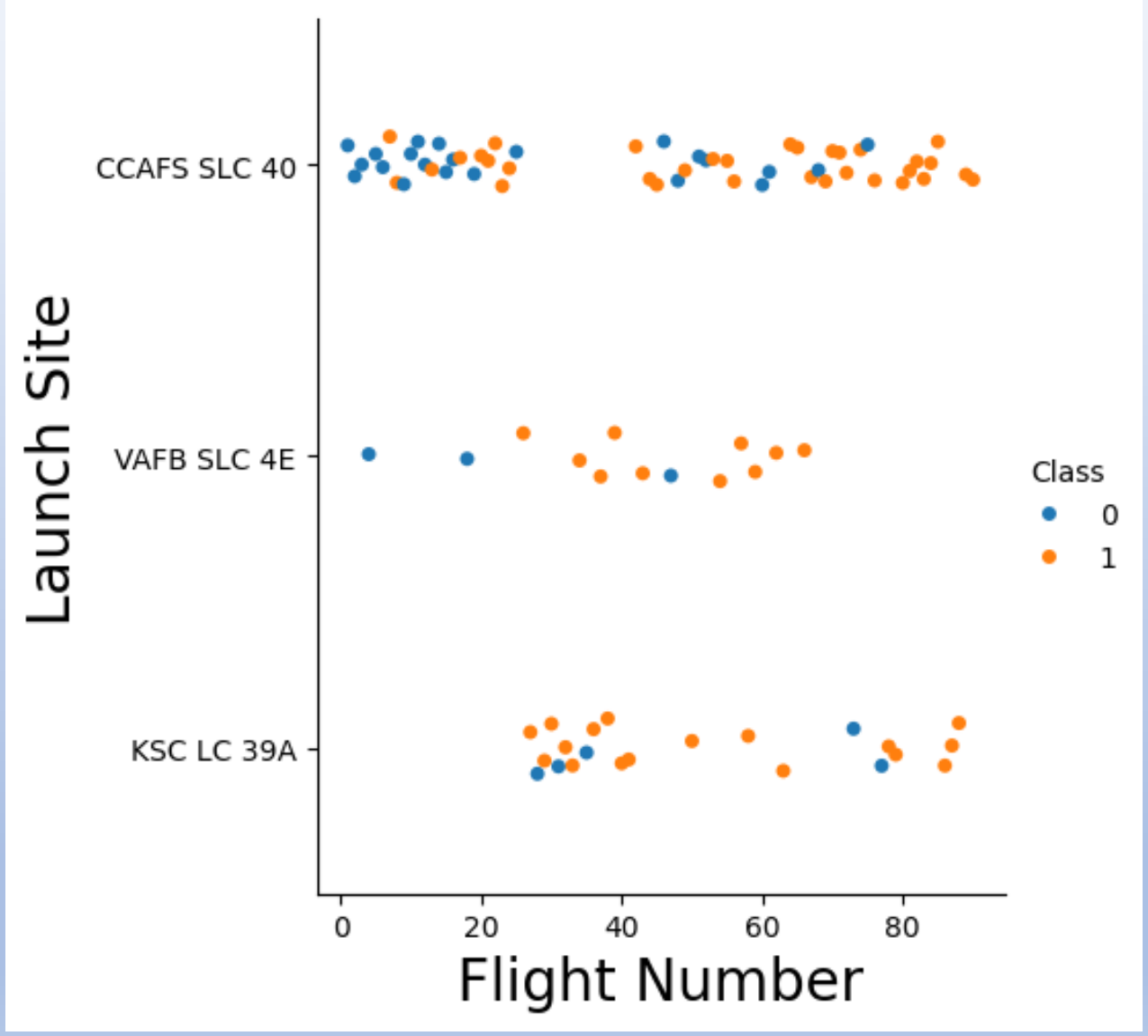
- Business logic and best practices can be analyzed from different missions e.g. Falcon 1 or Falcon 9 missions etc.
- Interactive analytics demo in screenshots
- Pre-trained models for predictive analysis of future space missions




Section 2
Insights drawn from EDA

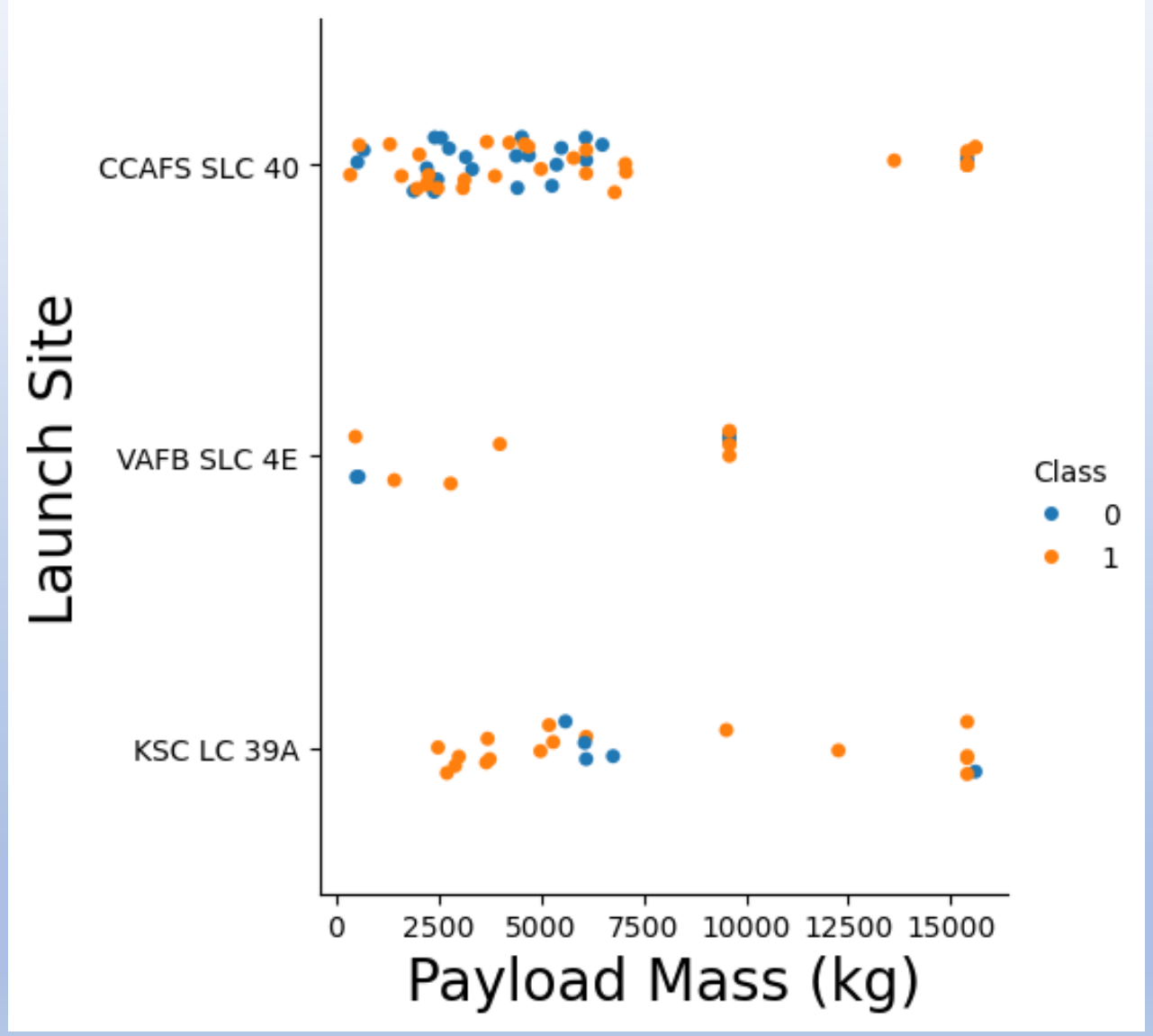
Flight Number vs. Launch Site

A breakdown of previous Space X missions



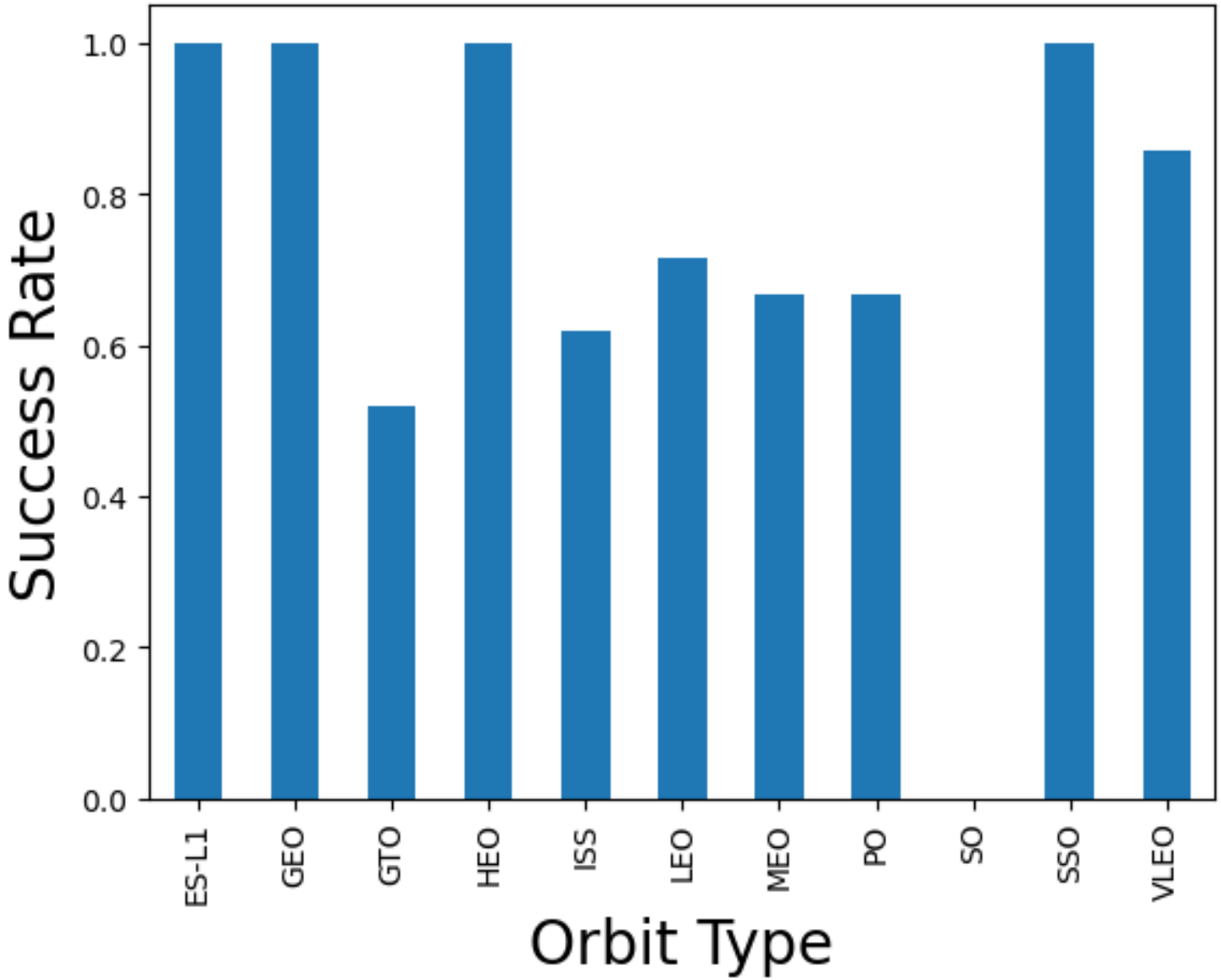
Payload vs. Launch Site

A breakdown of payload and launch site missions.



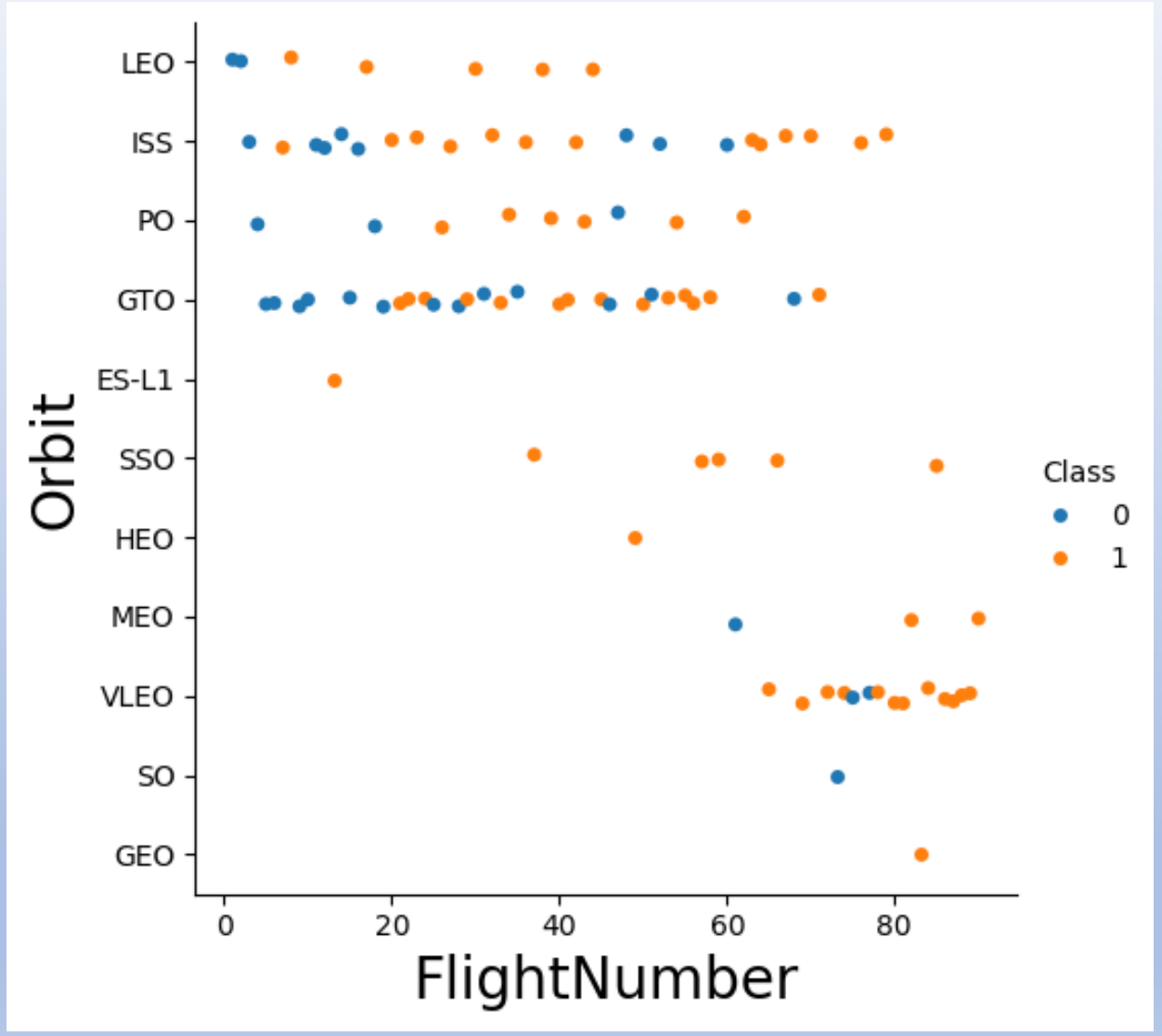
Success rate of missions by orbit type

Success Rate vs. Orbit Type



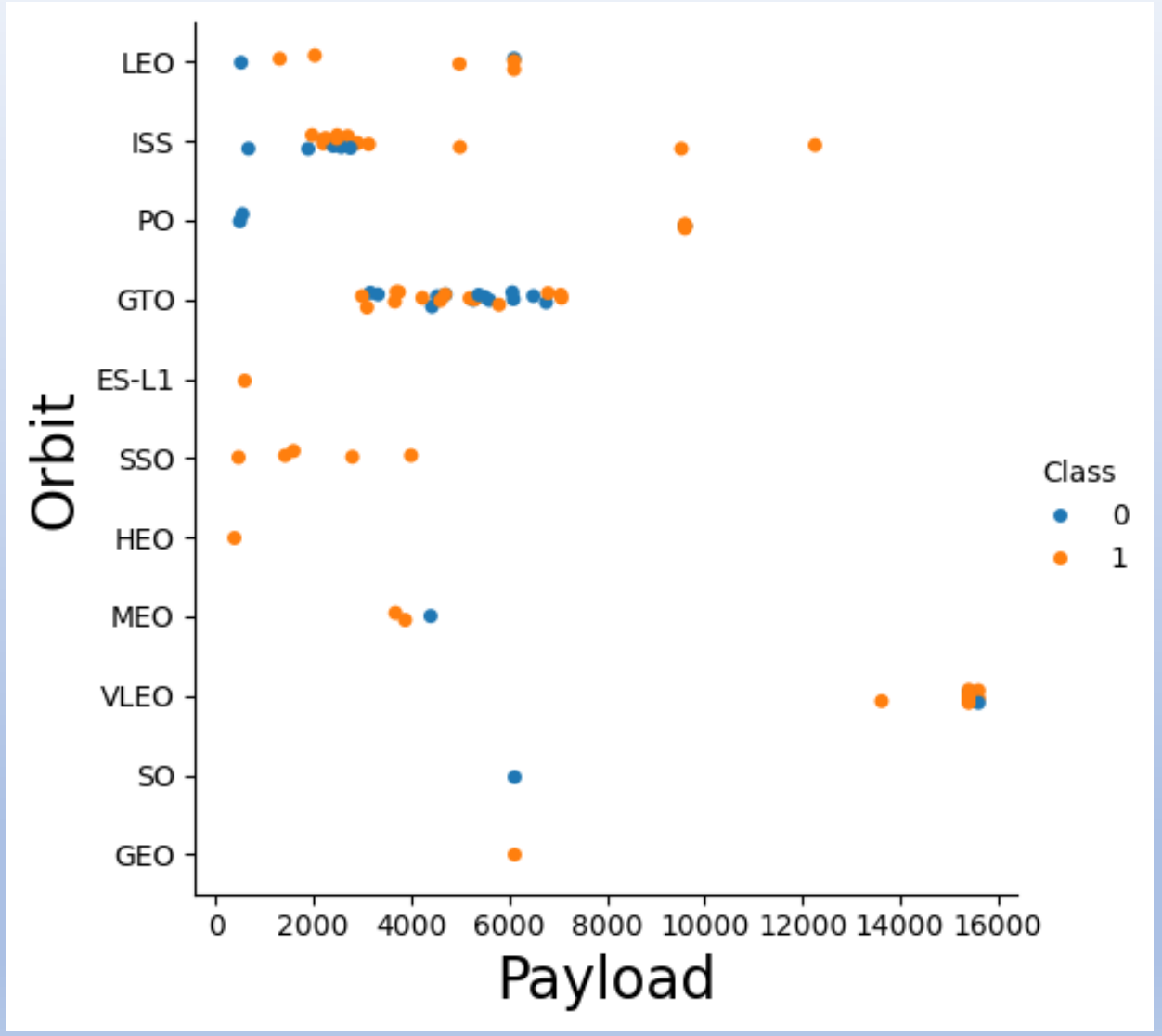
Flight Number vs. Orbit Type

Previous space missions by flight number and orbit type



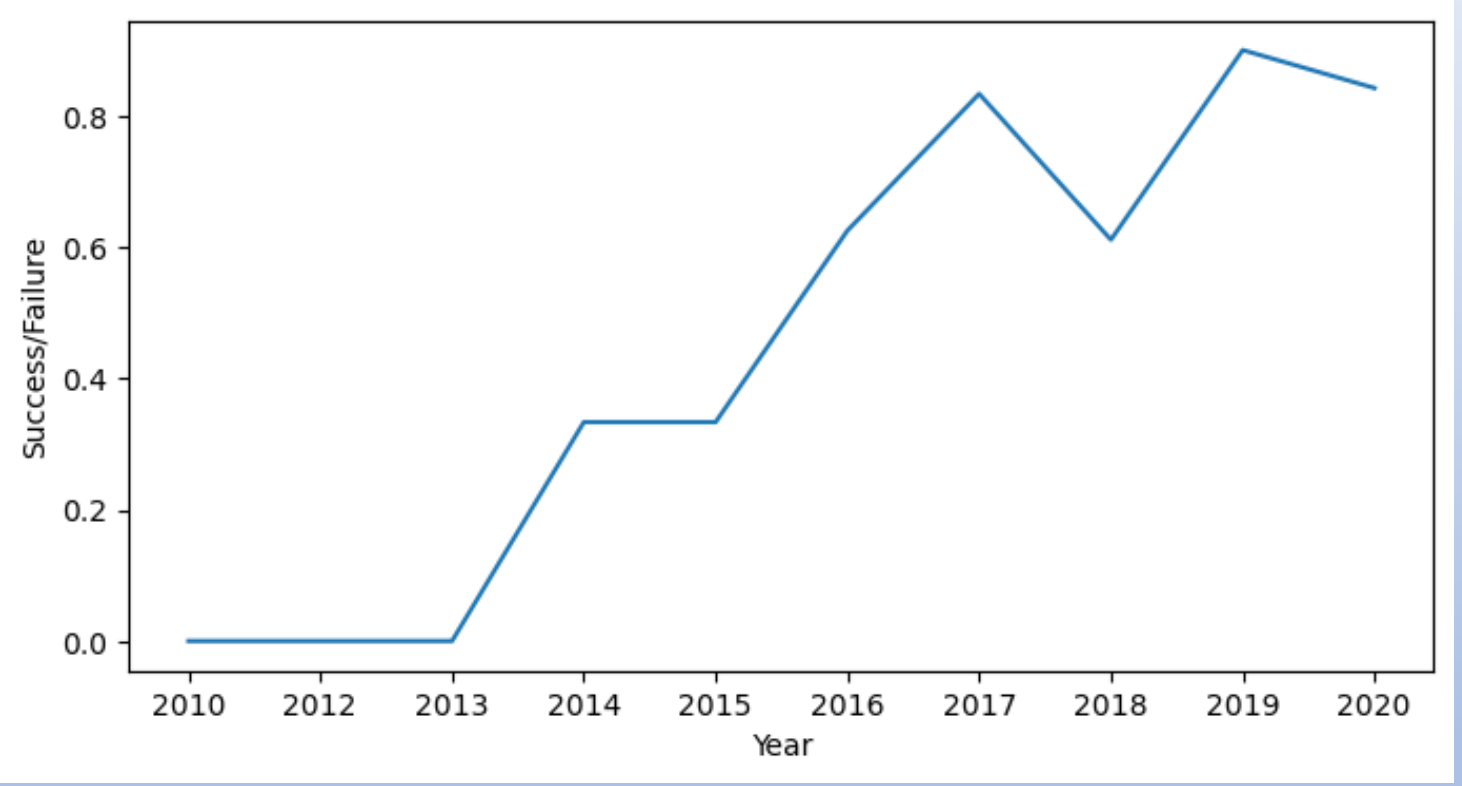
Payload vs. Orbit Type

Payload by orbit type



Launch Success Yearly Trend

- Space X missions have had an average success rate of about 66.7%. With our ML models we can ensure future missions are successful



All Launch Site Names

- Space X has 4 launch sites. Though the 4th, Starbase, first started space missions in 2023 so there isn't much public available data yet. These are the number of missions so far.

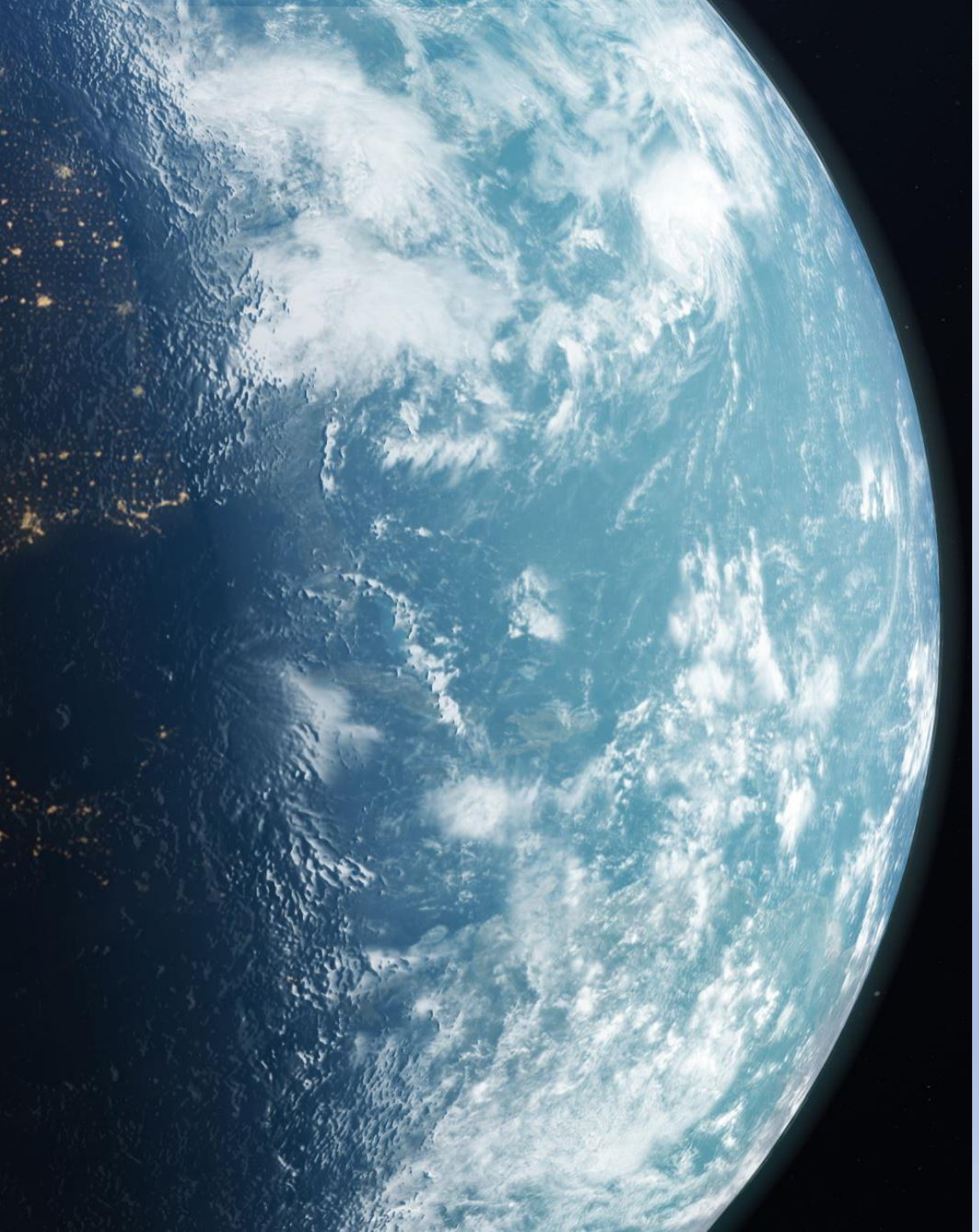
```
Out[6]:  CCAFS SLC 40     55
         KSC LC 39A       22
         VAFB SLC 4E      13
         Name: LaunchSite, dtype: int64
```

Launch Site Names Begin with 'CCA'

SQL EDA was done to pull Space X data. This is the Cape Canaveral launch site in Florida.

Out[10]:

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



Total Payload Mass

- The total payload carried by boosters from NASA was 45,596 kg or 100,522 lbs. or 52.3 tons

Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1 was 340.4 kg
- As we look to expand civilization out into the stars, we look to improving these metrics.

First Successful Ground Landing Date

the first successful landing outcome At space x on a ground pad was December 22nd 2015.

By implementing ai and data science, we can look to further innovate in tech and industry

Successful Drone Ship Landing with Payload between 4000 and 6000

- These are the boosters from successful missions to date.

- Simple SQL queries or general queries can provide valuable insights

Out[15]:

Booster_Version
F9 FT B1021.1
F9 FT B1022
F9 FT B1023.1
F9 FT B1026
F9 FT B1029.1
F9 FT B1021.2
F9 FT B1029.2
F9 FT B1036.1
F9 FT B1038.1
F9 B4 B1041.1
F9 FT B1031.2
F9 B4 B1042.1
F9 B4 B1045.1
F9 B5 B1046.1

Total Number of Successful and Failure Mission Outcomes

TO DATE, THE TOTAL NUMBER OF SUCCESSFUL MISSION OUTCOMES ARE 100 AND 1 FAILURE (IN FLIGHT).

BY EVALUATING METRICS, WE CAN MAKE BETTER BUSINESS DECISIONS

Out[20]:

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

Boosters Carried Maximum Payload

- These are the boosters with maximum payload
- By evaluating these business metrics, we can better iterate on our business decisions.

2015 Launch Records

- In 2015 there were 2 failed missions.
- By evaluating equipment failures, we can better iterate on our products.

35]:

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

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

- These are the landing outcome metrics from 2010 - 2017

37]:

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

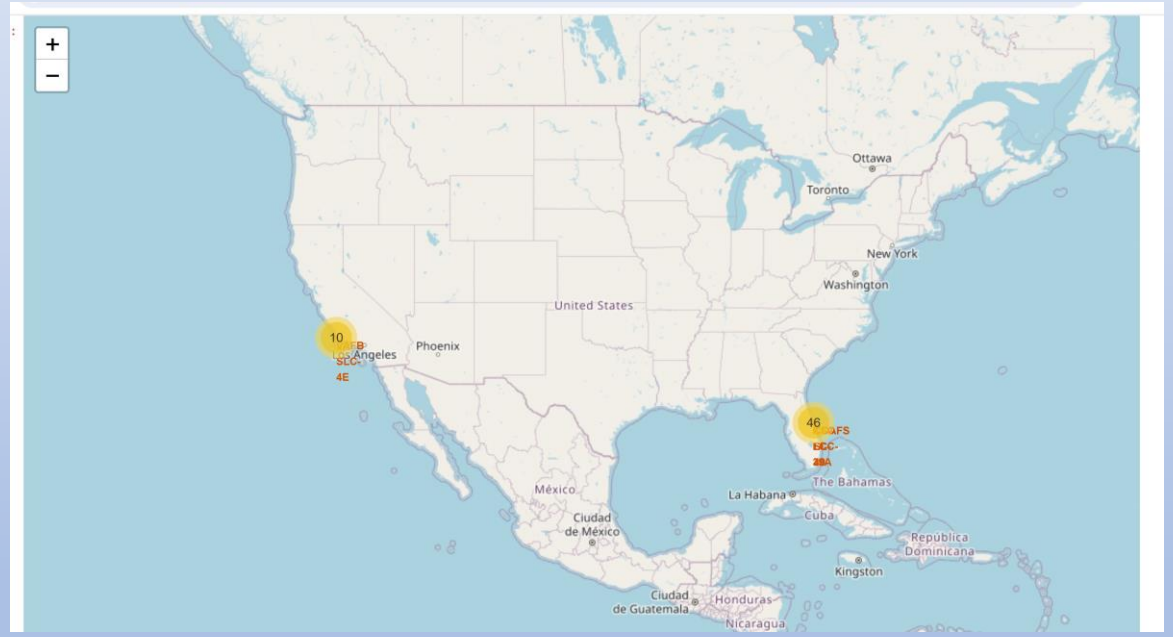

Section 3
Launch Sites
Proximities Analysis

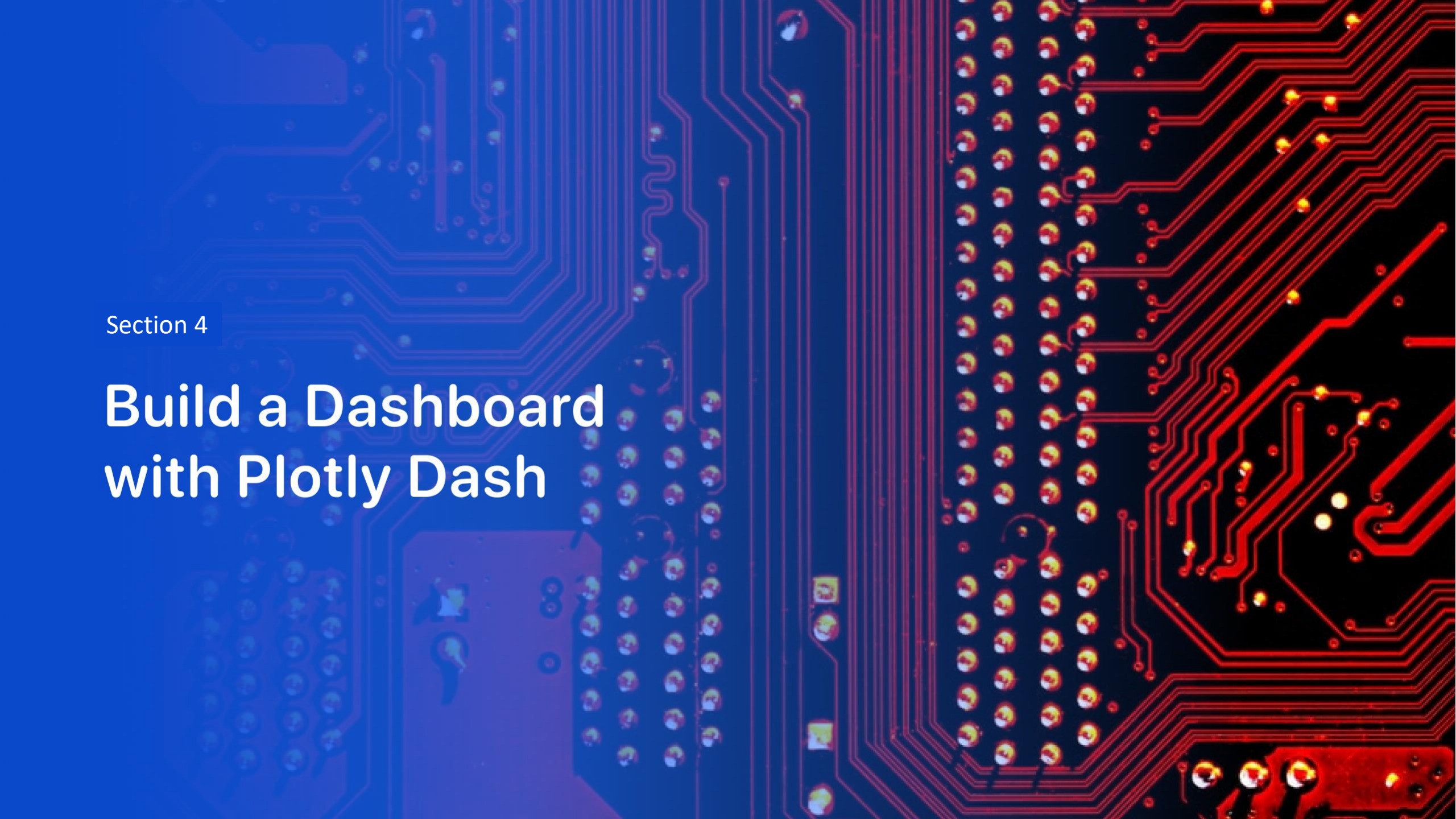
Geospatial analysis with Folium

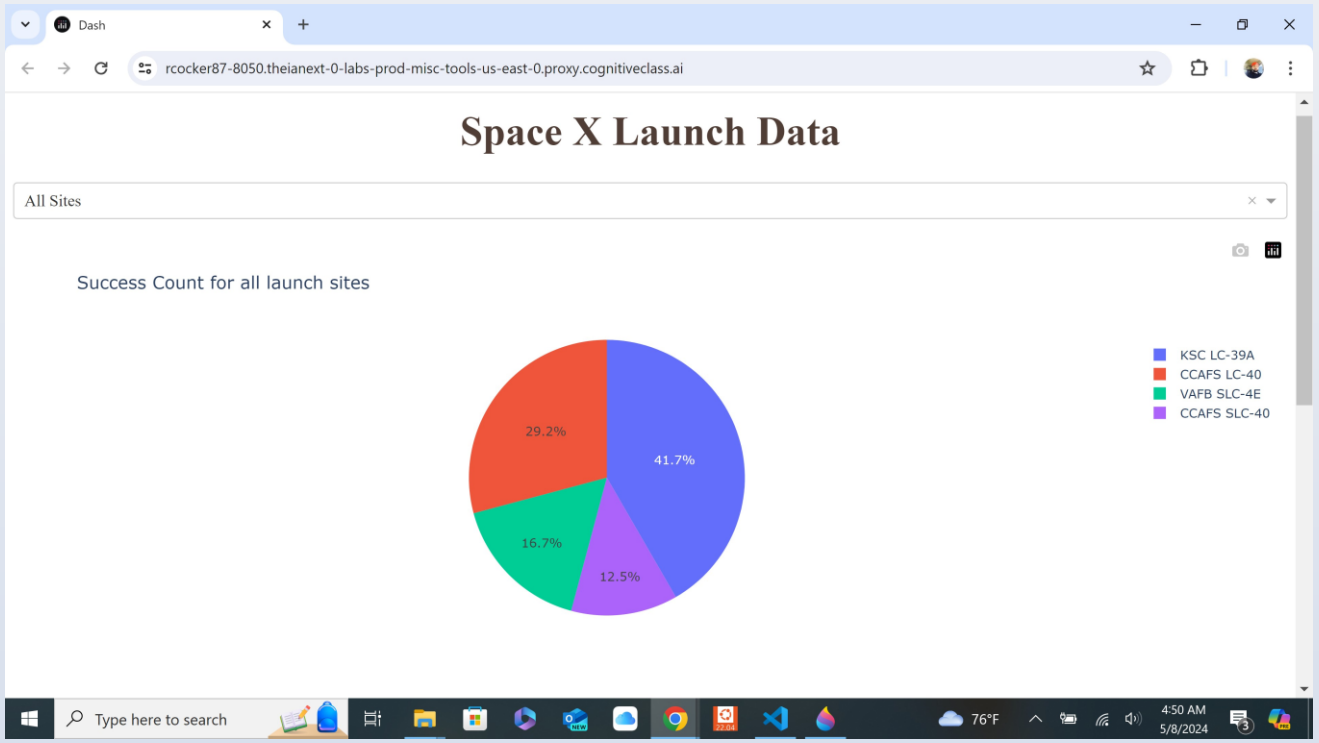
Geospatial analysis for all Space X launch sites including launch data.

By evaluating key metrics, we can understand what has been accomplished, and where we can improve further.

Screenshot of Folium map




Section 4
Build a Dashboard with Plotly Dash



Dashboard of Space X Launch Data

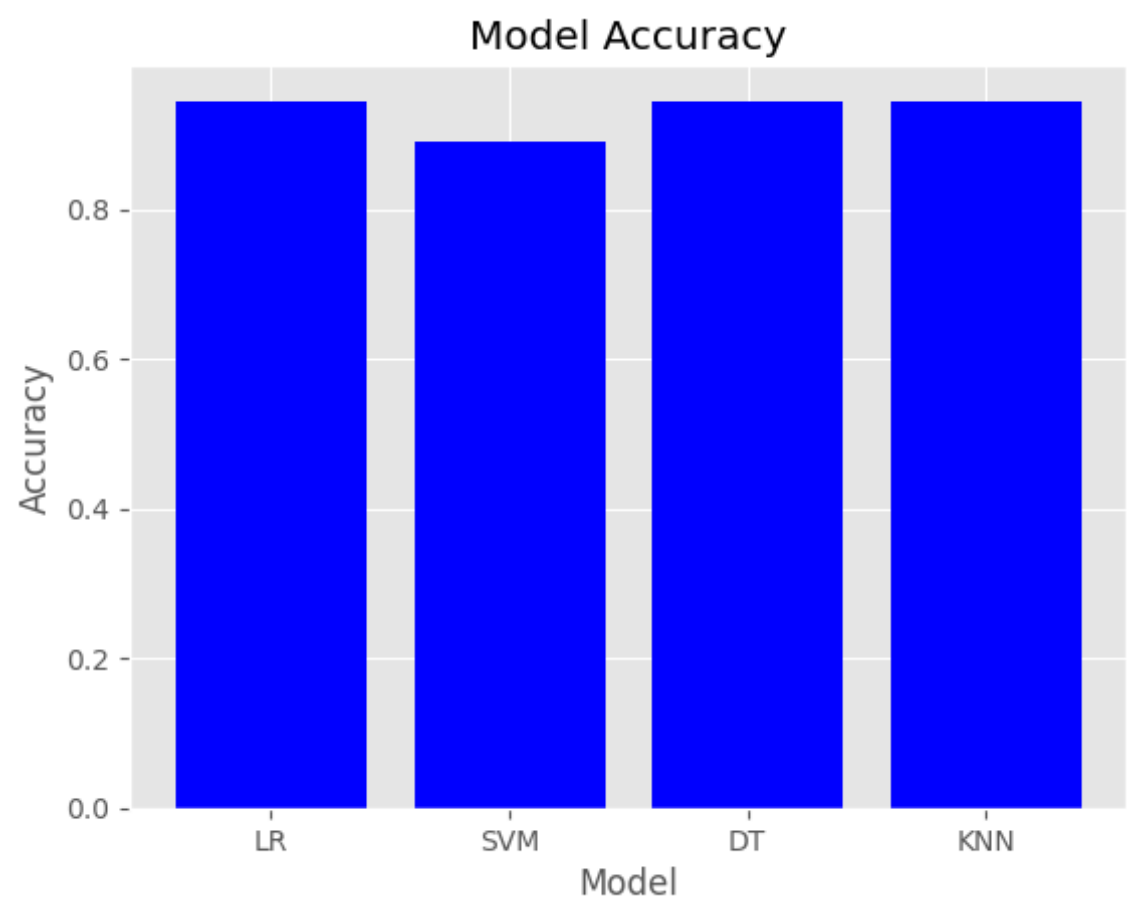
- An interactive dashboard was built of launch data for stakeholders.
- This helps generate data-driven insights of the companies business metrics.

Success Count by Payload

- The dashboard includes metrics such as successful missions to date by payload.
- These metrics can prove to be valuable in making future decisions on future missions.

Section 5

Predictive Analysis (Classification)



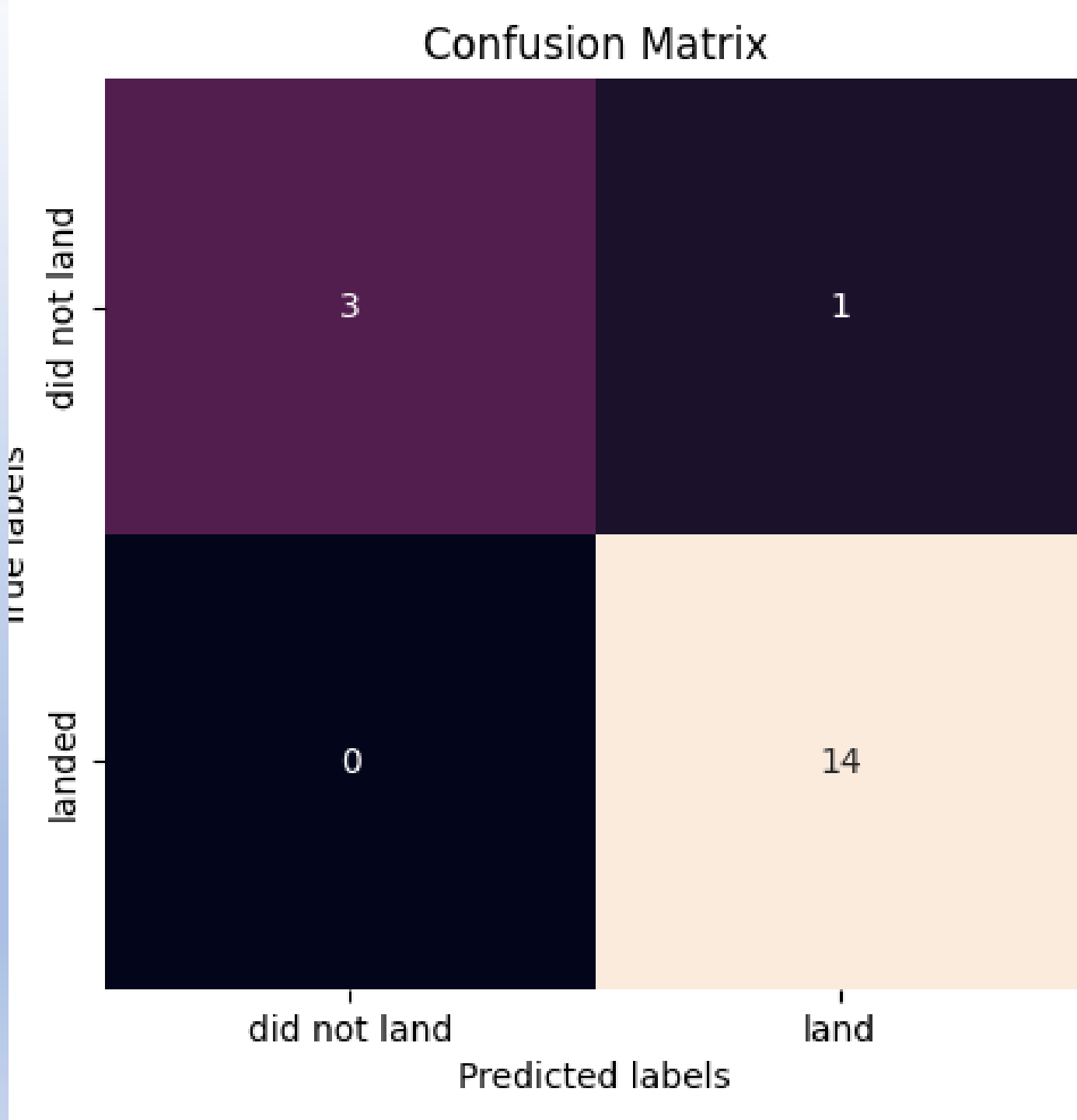
Classification Accuracy

- Logistic Regression, Decision Tree, and K Nearest Neighbors had the best predictive accuracies @ 94.44%

- By utilizing AI and machine learning, we can optimize key company performance metrics.

Confusion Matrix

- Decision Tree Classifier was one of the best models with a 94.44% accuracy for successful future missions.



Conclusions

LEVERAGING AI, MACHINE LEARNING, AND DATA SCIENCE IS ESSENTIAL FOR SPACEX TO OPTIMIZE KEY PERFORMANCE METRICS AND GAIN A COMPETITIVE ADVANTAGE.

MACHINE LEARNING IMPROVES LAUNCH SIMULATIONS FOR BETTER PAYLOAD DELIVERY, WHILE AI DETECTS ANOMALIES TO PREVENT DELAYS.

DATA ANALYTICS STREAMLINES SUPPLY CHAIN OPERATIONS, REDUCING WASTE AND MAXIMIZING COST-EFFECTIVENESS.

THE FUSION OF HUMAN EXPERTISE WITH ADVANCED TECHNOLOGIES EMPOWERS SPACEX TO PIONEER SPACE EXPLORATION AND REALIZE THE DREAM OF BECOMING A MULTI-PLANETARY SPECIES.

Appendix

- GitHub repository: https://github.com/rcghpge/ibm-data-science-professional-cert/tree/main/space-x-capstone

Thank you!