SpaceY: Success for the Launch future

Vignesh Sivakumar

October 30, 2022

IBM Developer

SKILLS NETWORK

OUTLINE

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

EXECUTIVE SUMMARY

- Data Collection with API and Web Scraping
- Exploratory Data Analysis
  - Data Wrangling
  - Interactive Visual Analytics
  - Data Visualization
- Interactive Launch Maps with Folium
- Launch Records Dashboard using Plotly
- Machine Learning Prediction Analysis

INTRODUCTION

- Evaluate the ability for the new SpaceY company to launch rocket technology, in perspective of rocket costs, variables, and geographic viability.

➢ Prediction of successful launches and landings

What rocket variables affect these outcomes?

How can SpaceY compete with SpaceX using predictive models and evaluation of prior launches?

METHODOLOGY

- Data Collection Methodology
  - Data was sourced from the following:
    - SpaceX API: https://api.spacexdata.com/v4/rockets/
    - Wikipedia Web Scraping

- Data Wrangling
  - One Hot Encoding was performed
  - Landing Outcomes added based on feature data
  - Removal of unnecessary columns

METHODOLOGY

- Exploratory Data Analysis
  - SQL and Visualization tools
  - Exploratory Analysis with Rocket Database
  - Category plotting with launch variables
- Interactive Visual Analytics
  - Folium Map Visual
  - Plotly Launch Dashboard Visual
- Predictive Analysis
  - Classification Models
  - Accuracy testing

Data Collection: API and Web Scraping

- SpaceX API
- Variables: Payload mass, Rocket launch, and Rocket name
  1. Rest API called for data parsing
  2. JSON file into DataFrame
  3. Data prepared for export
- SpaceX Web Scraping
  1. Falcon 9 Wikipedia HTML response obtained
  2. BeautifulSoup Method employed for data extraction
  3. DataFrame creation from launch tables
  4. Data prepared for export

Data Wrangling

- EDA using SQL
- The following conditions were queried:
  - Unique launch sites in SpaceX’s repertoire
  - Launch sites beginning with “CCA” limited to 5 query results
  - Total payload mass by NASA (CRS) boosters
  - Average payload mass by booster version F9 v.1.1
  - Date of first successful landing using ground pad
  - Booster names of payload mass between 4000-6000 kg
  - Total number of successful and failed launches
  - Booster versions with the maximum payload mass
  - Records of months, landing outcomes, booster versions, launch site, and months in 2015
  - Count of successful landing outcomes between June 4th 2010 and March 20th 2017

Data Wrangling

- EDA with Data Visualization
- Relationships between different launch variables
  - Payload Mass, Launch Site, Flight Number, Payload, and Orbit
  - Scatter Plot, Category Plot, and Success Rate
- Preparing Data Feature Engineering
  - Success rate with feature selection using “get_dummies” method

Interactive Visual Analytics

- Map Interactivity using Folium
- Launch Site Location Analysis
  - Identify the launch sites within the United States
  - Successful and Failed Launches in each site
  - Distances between launch site and geographic sites
- NASA Johnson Space Center (Houston, Texas)
  - Markers, Clusters, and Circles employed for visual coordinate grouping

Interactive Visual Analytics

- Plotly Launch Dashboard
  - Launch Site Records and the Success Rate
  - Interactive dashboard with pie chart, scatter chart, slider, and dropdowns
  - Launch site selection compared with payload mass range adjustor and relationship between payload mass and success rate

- Identification of relationship between launch site, payload mass, booster version category, and respective success rates

Predictive Analysis

- Classification using Machine Learning
  - Standardize data and Split into Training/Test data
  - Best parameters for the following classifiers:
    - Support Vector Machines
    - Logistic Regressions
    - Decision Tree
    - K-Nearest Neighbors
  - Model evaluation and accuracy comparison
    - Plot Confusion Matrix

RESULTS: EDA with SQL

- The unique launch site names are presented:
  - SQL code + Resulting Output

```
%%sql
SELECT DISTINCT "LAUNCH_SITE" FROM SPACEXTBL
```

```
 * sqlite:///my_data1.db
Done.
```

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

RESULTS: EDA with SQL

- The launch sites with the name "CCA"
  - SQL code + Resulting Output

```
%%sql
SELECT * FROM SPACEXTBL
WHERE "LAUNCH_SITE" LIKE '%CCA%' LIMIT 5
```

```
 * sqlite:///my_data1.db
Done.
```

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

RESULTS: EDA with SQL

- The total payload mass from NASA (CRS)
  - SQL code + Resulting Output

```
%%sql
SELECT SUM("PAYLOAD_MASS__KG_") FROM SPACEXTBL
WHERE "CUSTOMER" = 'NASA (CRS)'
```

```
 * sqlite:///my_data1.db
Done.
```

SUM("PAYLOAD_MASS__KG_")
45596

RESULTS: EDA with SQL

- The average payload mass from booster version F9 v1.1
  - SQL code + Resulting Output

```
%%sql
SELECT AVG("PAYLOAD_MASS__KG_") FROM SPACEXTBL
WHERE "BOOSTER_VERSION" LIKE '%F9 v1.1%'
```

```
 * sqlite:///my_data1.db
Done.
```

AVG("PAYLOAD_MASS__KG_")
2534.6666666666665

RESULTS: EDA with SQL

- The first successful ground landing date
  - SQL code + Resulting Output

```
%%sql
SELECT MIN("DATE") FROM SPACEXTBL
WHERE "Landing _Outcome" LIKE '%Success%'
```

```
 * sqlite:///my_data1.db
Done.
```

MIN("DATE")
01-05-2017

RESULTS: EDA with SQL

- The successful drone ship landing (between 4000 kg-6000kg)
  - SQL code + Resulting Output

```
%%sql
SELECT "BOOSTER_VERSION" FROM SPACEXTBL
WHERE "LANDING _OUTCOME" = 'Success (drone ship)'
AND "PAYLOAD_MASS__KG_" > 4000 AND "PAYLOAD_MASS__KG_" < 6000;
```

```
 * sqlite:///my_data1.db
Done.
```

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

RESULTS: EDA with SQL

- The number of successful and failed missions
  - SQL code + Resulting Output

```
%%sql
SELECT (SELECT COUNT("MISSION_OUTCOME") FROM SPACEXTBL
        WHERE "MISSION_OUTCOME" LIKE '%Success%') AS SUCCESS,
(SELECT COUNT("MISSION_OUTCOME") FROM SPACEXTBL WHERE "MISSION_OUTCOME" LIKE '%Failure%') AS FAILURE
```

```
 * sqlite:///my_data1.db
Done.
```

SUCCESS	FAILURE
100	1

RESULTS: EDA with SQL

- The boosters carrying maximum payload
  - SQL code + Resulting Output

```
%%sql
SELECT DISTINCT "BOOSTER_VERSION" FROM SPACEXTBL
WHERE "PAYLOAD_MASS__KG_" = (SELECT max("PAYLOAD_MASS__KG_") FROM SPACEXTBL)
```

```
 * sqlite:///my_data1.db
Done.
```

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

RESULTS: EDA with SQL

- The launch records for 2015 (Failure outcome, booster version, and launch site
  - SQL code + Resulting Output

```
%%sql
SELECT substr("DATE", 4, 2) AS MONTH, "BOOSTER_VERSION", "LAUNCH_SITE" FROM SPACEXTBL
WHERE "LANDING _OUTCOME" = 'Failure (drone ship)' and substr("DATE",7,4) = '2015'
```

```
 * sqlite:///my_data1.db
Done.
```

MONTH	Booster_Version	Launch_Site
01	F9 v1.1 B1012	CCAFS LC-40
04	F9 v1.1 B1015	CCAFS LC-40

RESULTS: EDA with SQL

- The rank of successful landing outcomes (between 04-06-2010 and 20-03-2017)
  - SQL code + Resulting Output

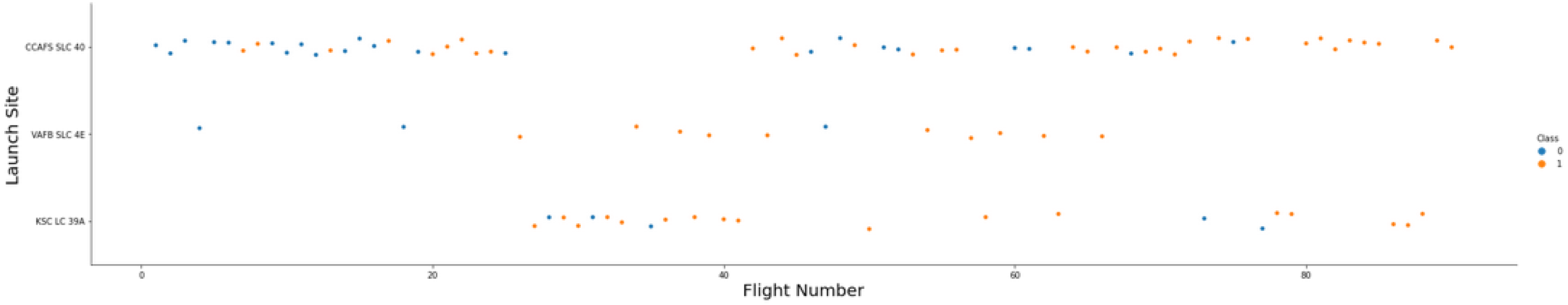
```
%%sql SELECT "LANDING _OUTCOME", COUNT("LANDING _OUTCOME") FROM SPACEXTBL
WHERE "DATE" >= '04-06-2010' and "DATE" <= '20-03-2017' and "LANDING _OUTCOME" LIKE '%Success%'
GROUP BY "LANDING _OUTCOME"
ORDER BY COUNT("LANDING _OUTCOME") DESC;
```

```
 * sqlite:///my_data1.db
Done.
```

Landing _Outcome	COUNT("LANDING _OUTCOME")
Success	20
Success (drone ship)	8
Success (ground pad)	6

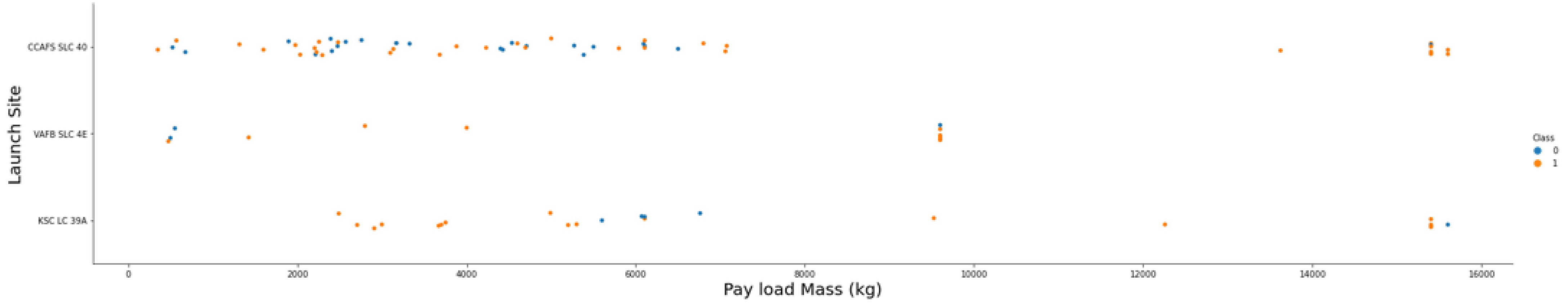
RESULTS: EDA with Visualization

- Flight Number vs. Launch Site
  - CCAF5 SLC 40 ranks the most successful launches
  - VAFB SLC 4E ranks higher than KSC LC 39A
  - General success rates have increased over the number of flights



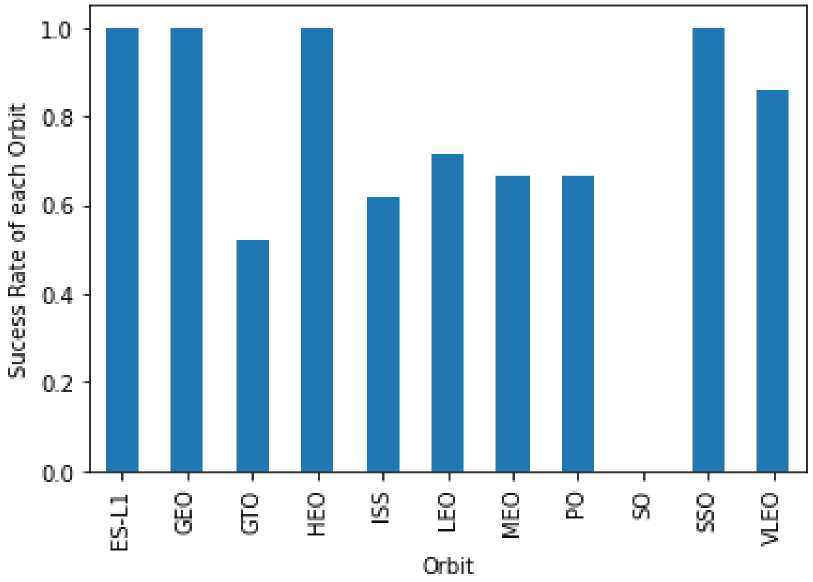
RESULTS: EDA with Visualization

- Payload vs. Launch Site
  - Payload masses below 8000 kg are variable in their success rates
  - Payload masses above 8000 kg are generally successful
  - VAFB SLC 4E appears to not be viable with payload masses above 14000 kg



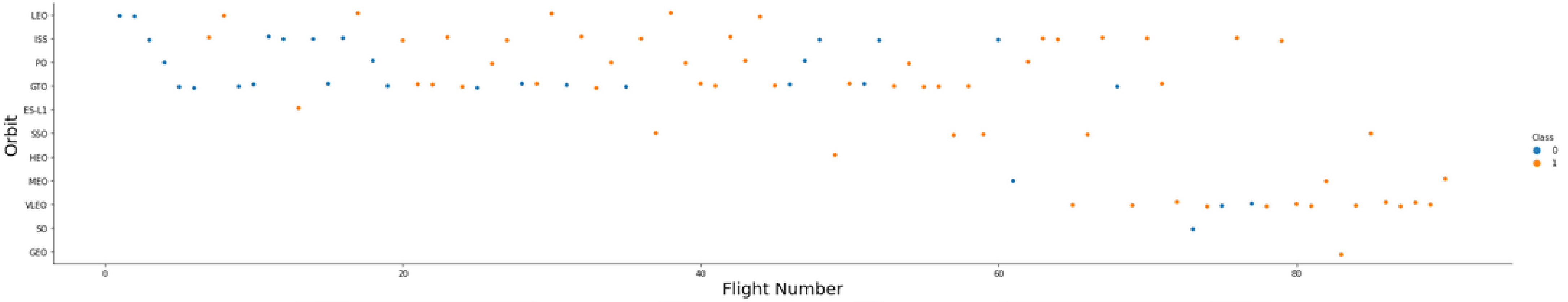
RESULTS: EDA with Visualization

- Success Rate vs. Orbit Type
  - ES-L1, GEO, HEO, and SSO have the highest success rates
  - VLEO is considerably close to these rates, with GTO falling short of <0.6 success rate
  - SO data is unverified



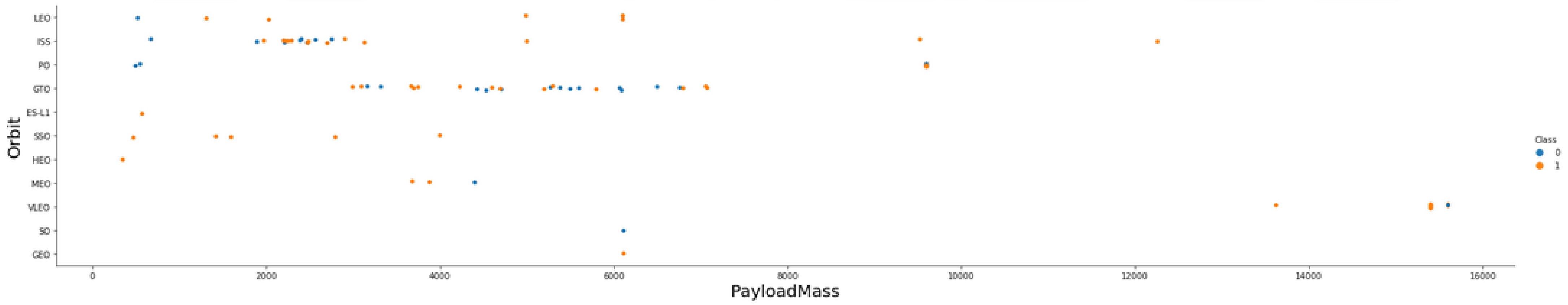
RESULTS: EDA with Visualization

- Flight Number vs. Orbit Type
  - VLEO is considerably new in the flight number success frequencies
  - Success rates are relatively consistent among the top contenders previously mentioned



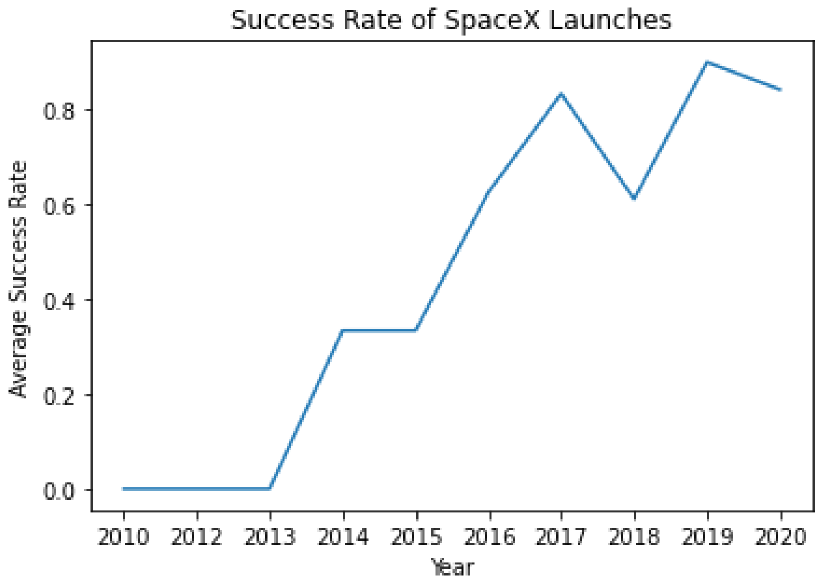
RESULTS: EDA with Visualization

- Payload vs. Orbit Type
  - SO and GEO do not have enough launch information relative to GTO
  - GTO contains a diverse payload and success rate trends
    - No correlation between both variables for GTO
  - ISS orbit has a payload mass range from >300 kg to <13000 kg



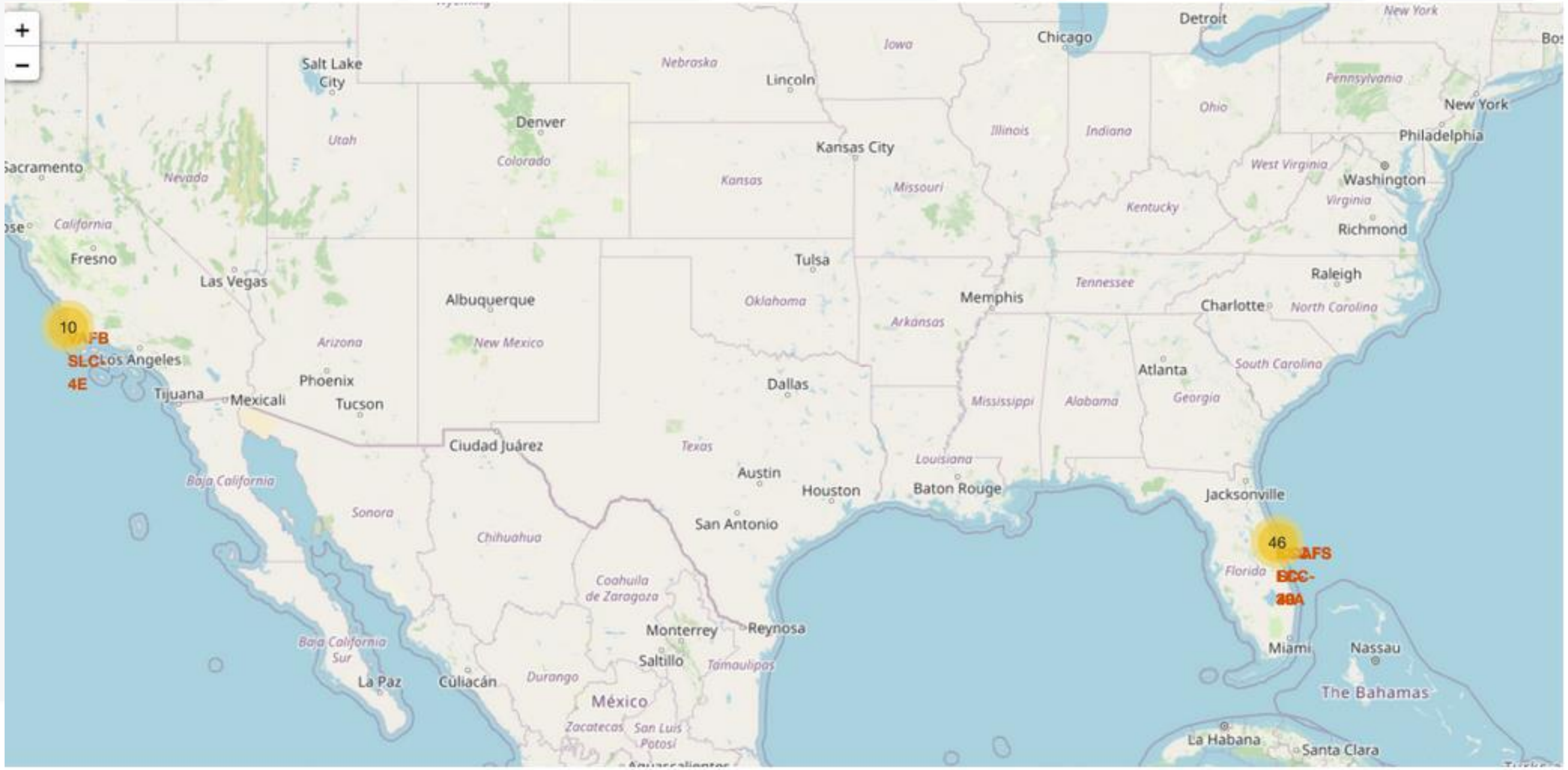
RESULTS: EDA with Visualization

- Launch Success Yearly Trend
  - In 2015 the success rate began rapidly climbing until 2017
  - From 2017 and 2018, certain launch variables impacted the success rate
  - In 2020 the average success rate of 0.8 remains slightly lower than that of 2019



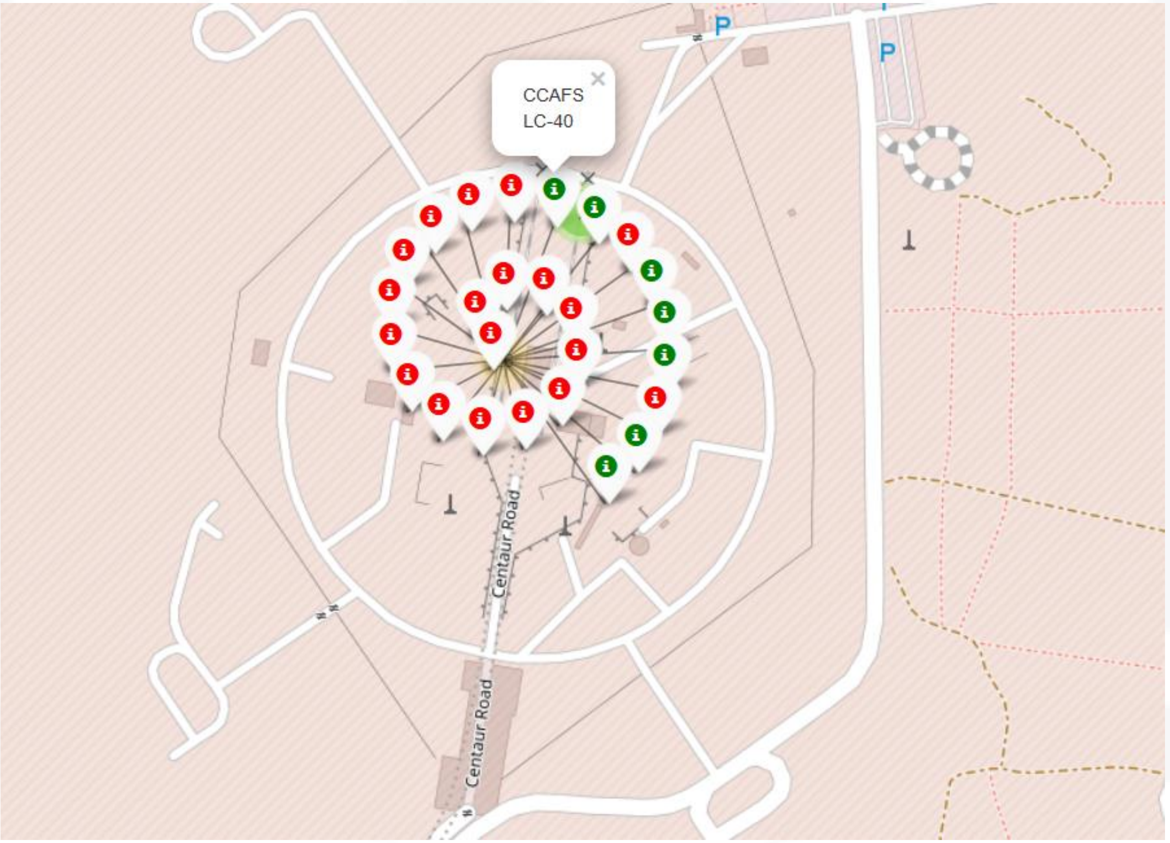
RESULTS: Folium Visualization

- The launch sites are situated in California and Florida, with proximity to the coastlines



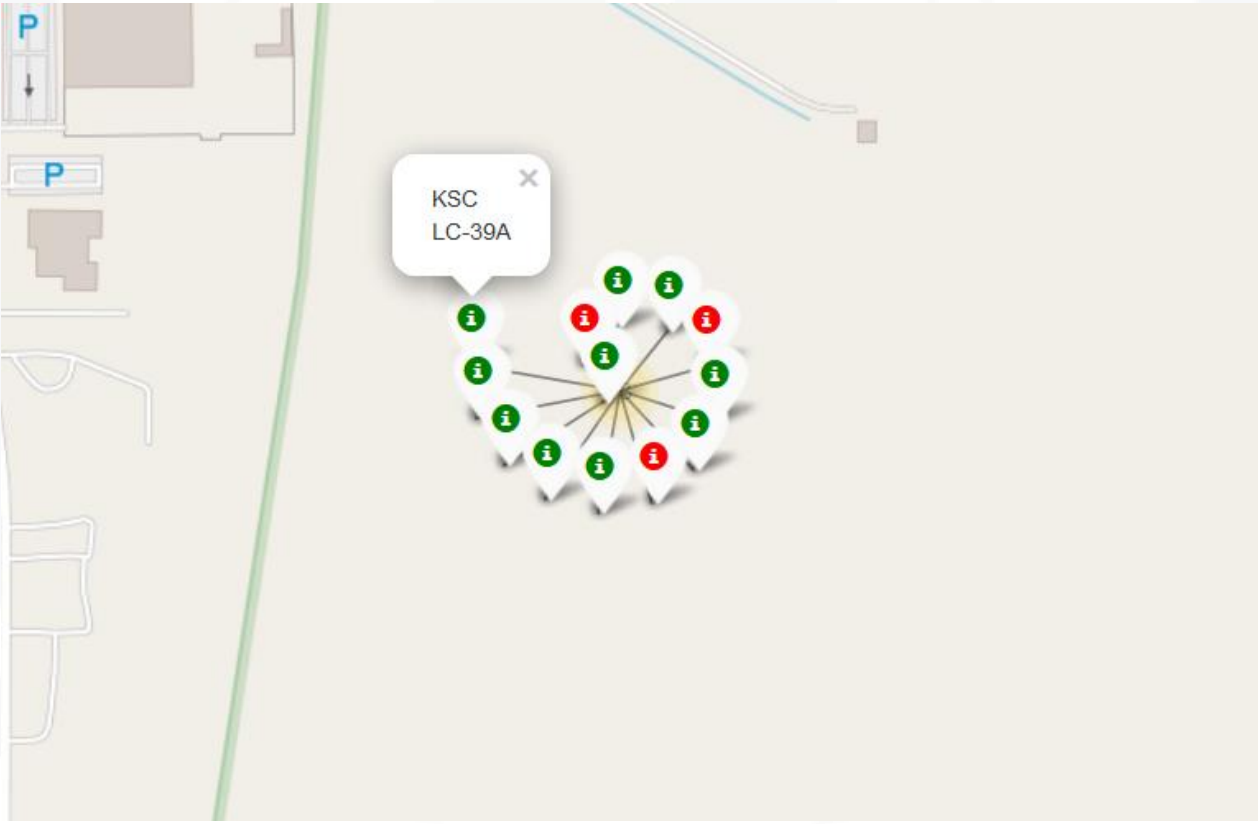
RESULTS: Folium Visualization

- The CCAFS LC-40 launches (highlighted green), indicates a successful launch near the coastline



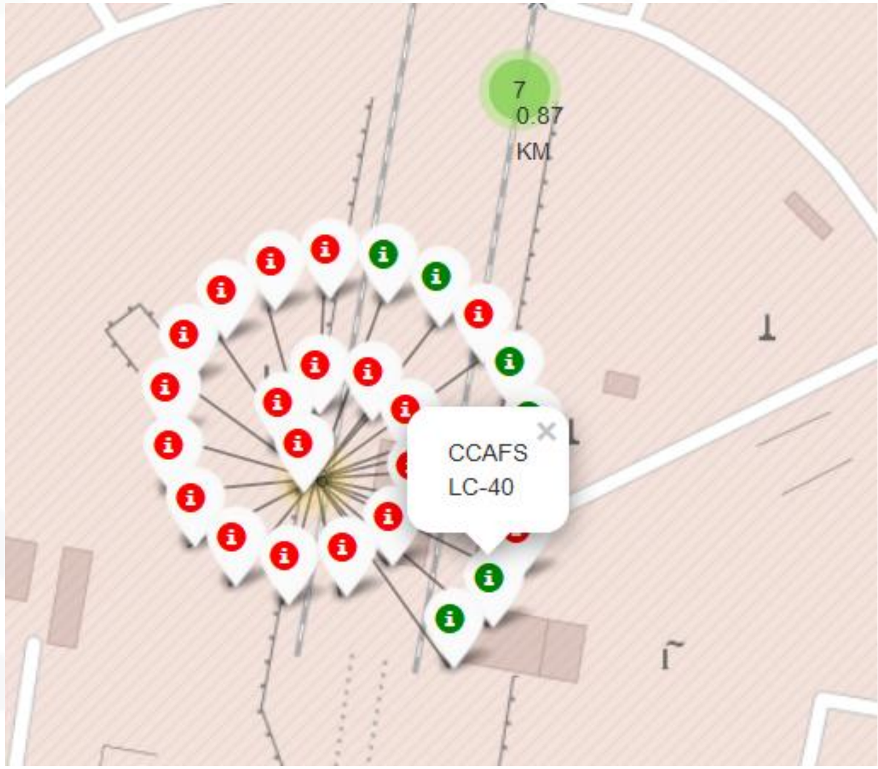
RESULTS: Folium Visualization

- The LC-39 A launches (highlighted green), contrasts higher number of launches further inland



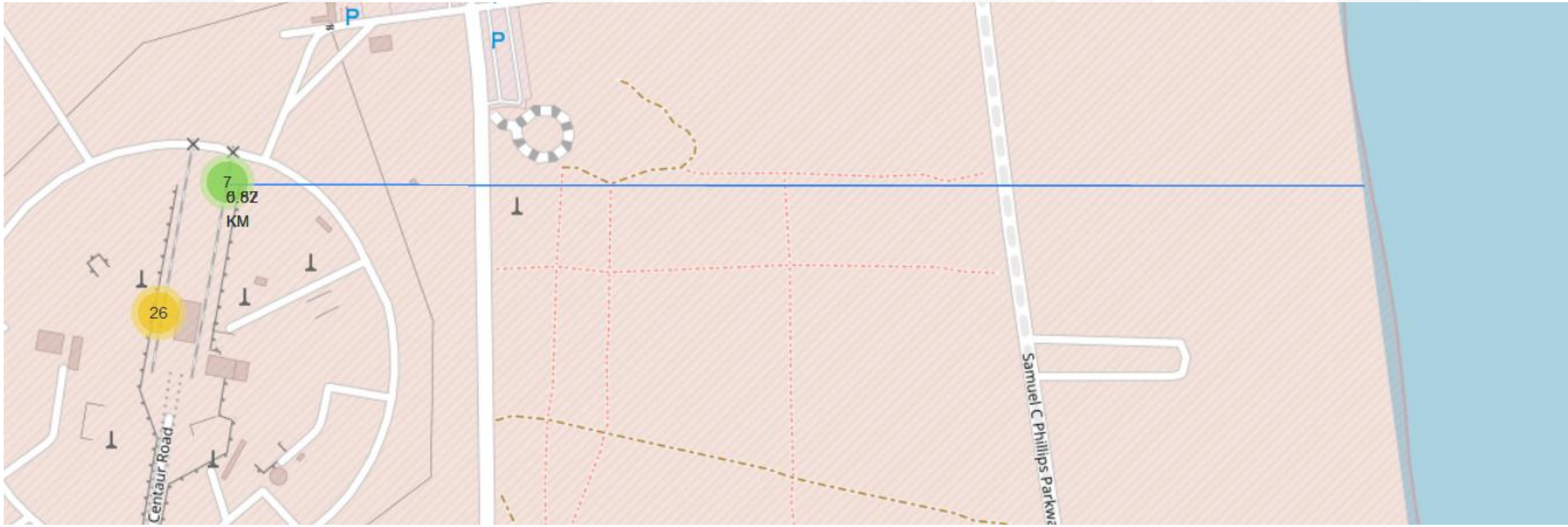
RESULTS: Folium Visualization

- Compare the kilometer distance between a relatively successful launch site region and a relatively failure-prone region (0.87 km)



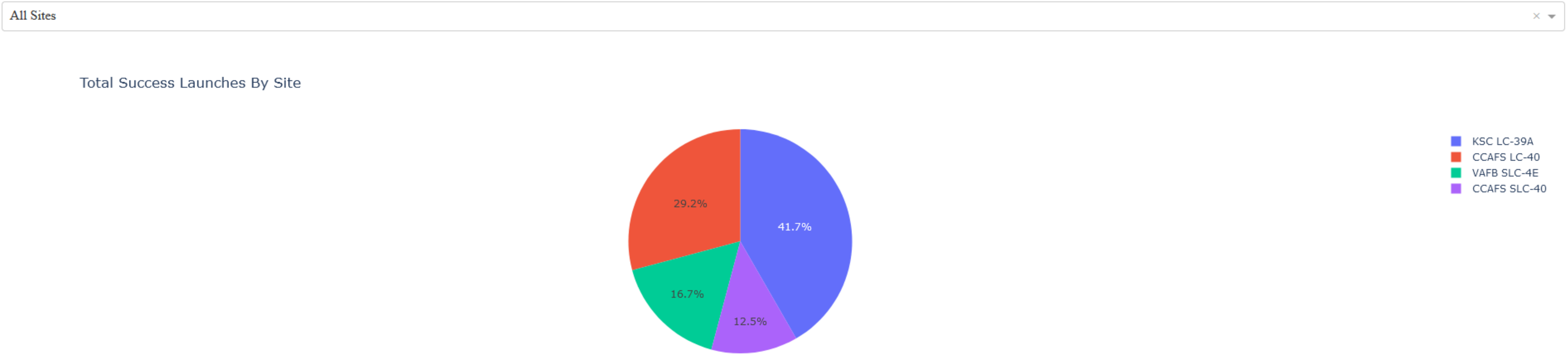
RESULTS: Folium Visualization

- Kilometer distance from the coastline (6.82 km) to the relatively successful launch site can be visualized
- The proximity to other geographic entities can be interacted with using Folium



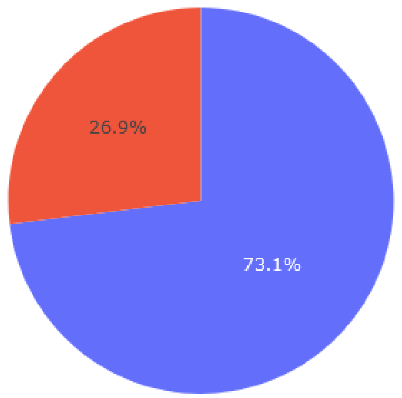
Results: Plotly Dashboard

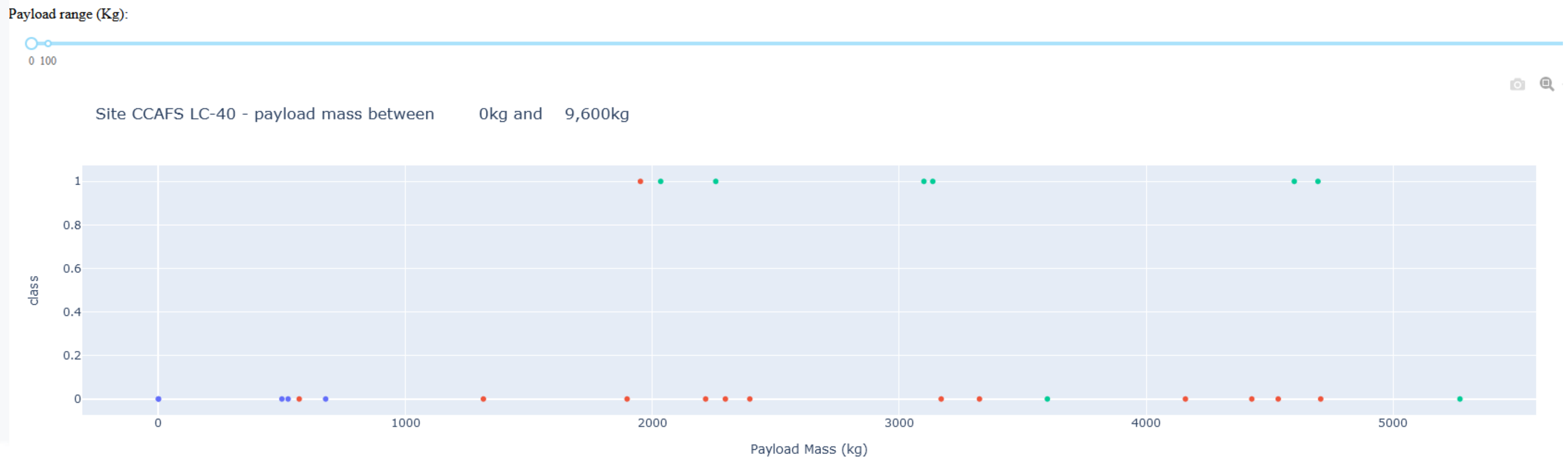
- All Sites for SpaceX launch records
- KSC LC 39-A represent significant proportion of successful launches



Results: Plotly Dashboard

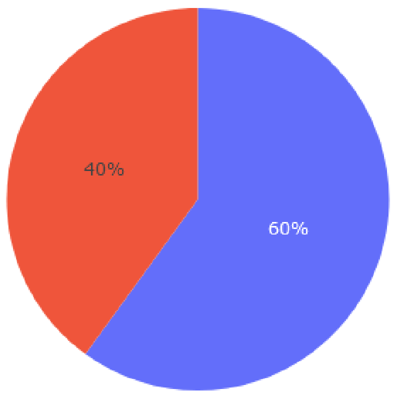
- CCAFS LC-40 for SpaceX launch records
- Booster v1.0 significantly larger proportion, yet lower payload mass

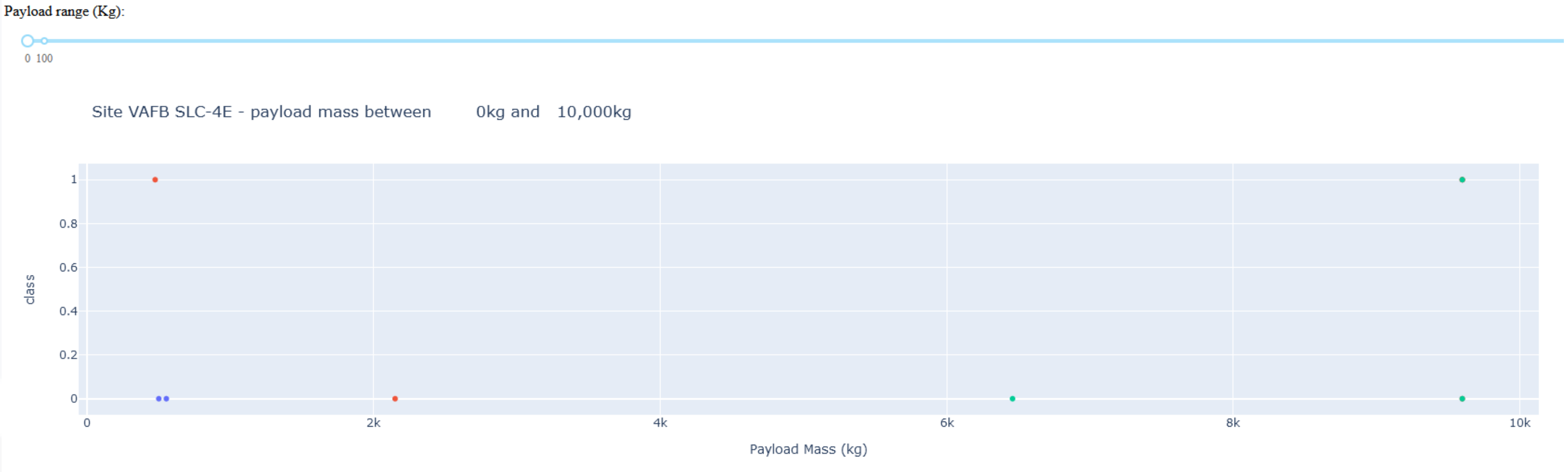




Results: Plotly Dashboard

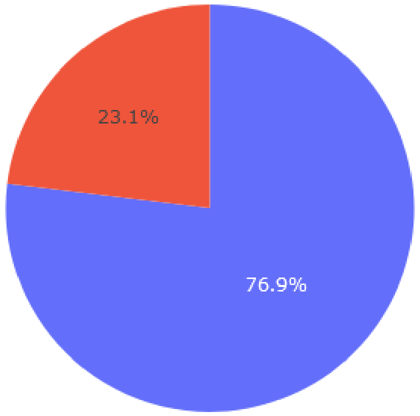
- VAFB SLC-4E for SpaceX launch records
- Booster v1.1 significantly larger proportion
- Payload masses at <2000 kg between FT and v1.1 are negligible

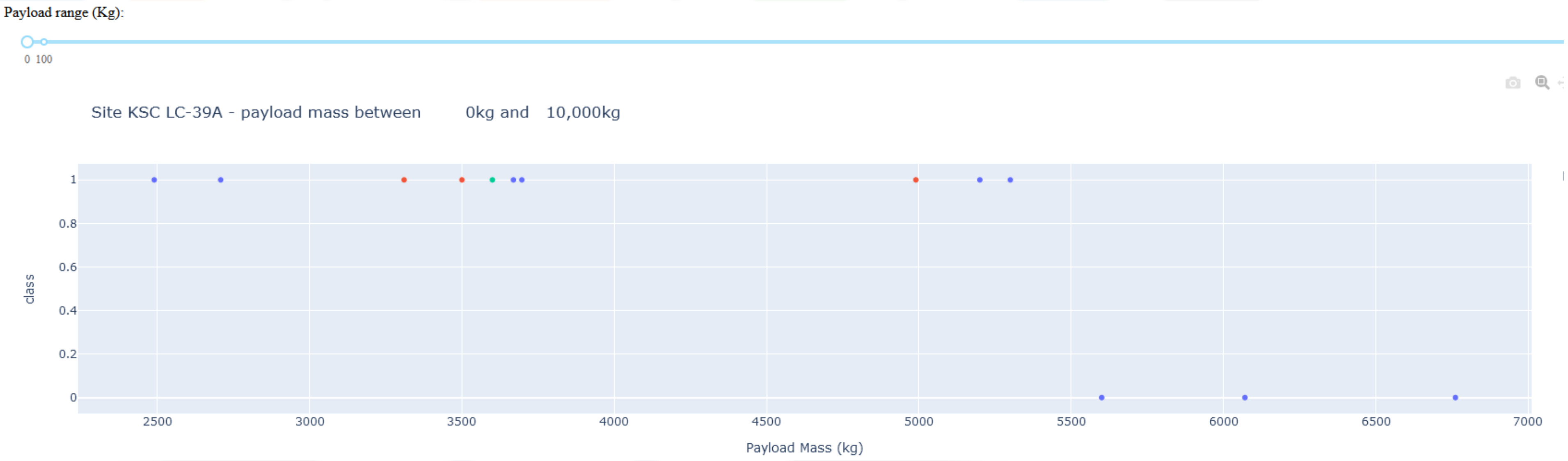




Results: Plotly Dashboard

- KSC LC-39A for SpaceX launch records
- Booster FT significantly larger proportion
- Payload mass capacities are unsuccessful at masses above 5600 kg



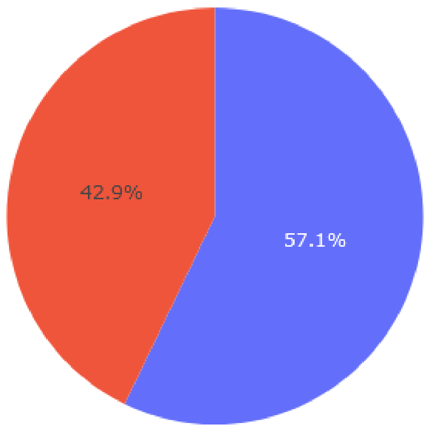


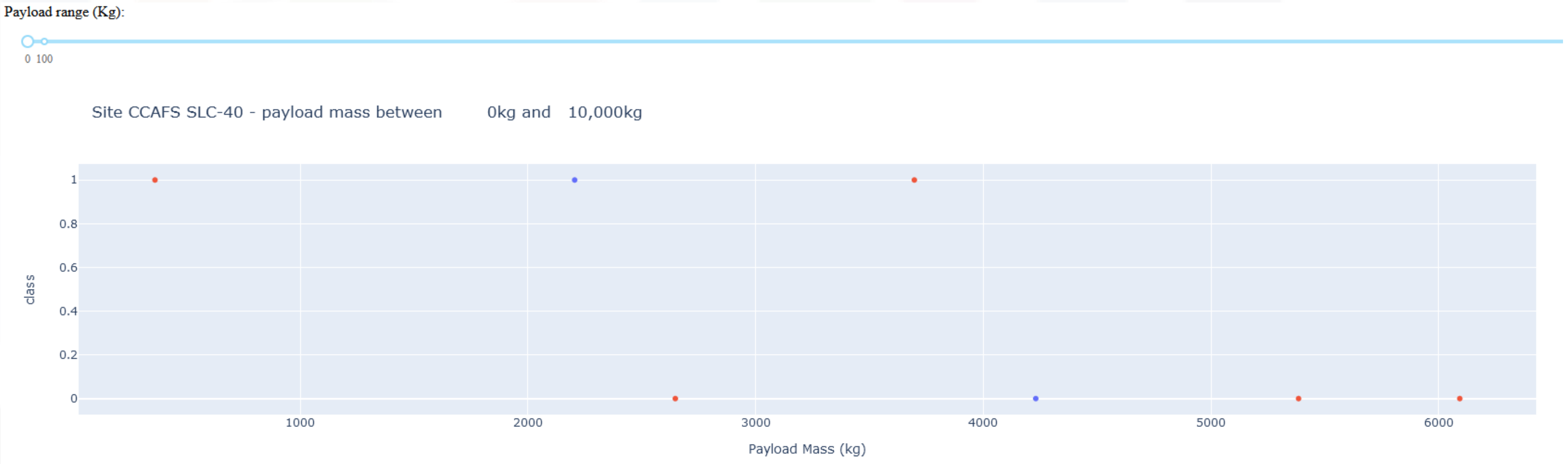
IBM Developer

SKILLS NETWORK

Results: Plotly Dashboard

- CCAFS SLC-40 for SpaceX launch records
- B4 category of booster performs better at lower payload masses
- Both booster categories are negligible in payload to success rate evaluation





IBM Developer

SKILLS NETWORK

Results: Predictive Analysis

- Logistic Regression Classifier

```
tuned hpyerparameters :(best parameters)  {'C': 0.01, 'penalty': 'l2', 'solver': 'lbfgs'}
accuracy : 0.8464285714285713
```

- Support Vector Machine Classifier

```
tuned hpyerparameters :(best parameters)  {'C': 1.0, 'gamma': 0.03162277660168379, 'kernel': 'sigmoid'}
accuracy : 0.8482142857142856
```

- Decision Tree Classifier

```
tuned hpyerparameters :(best parameters)  {'criterion': 'entropy', 'max_depth': 14, 'max_features': 'sqrt', 'min_samples_leaf': 2, 'min_samples_split': 1
0, 'splitter': 'random'}
accuracy : 0.8892857142857145
```

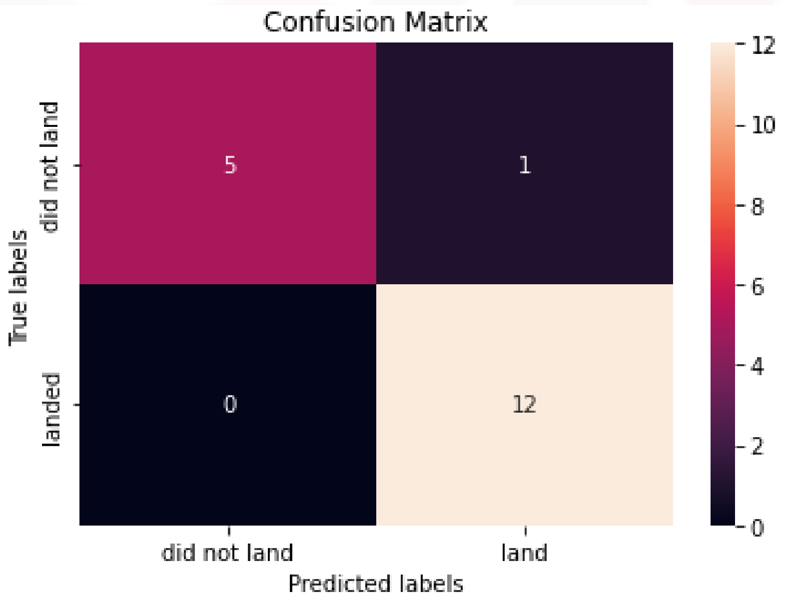
- k-Nearest neighbors Classifier

```
tuned hpyerparameters :(best parameters)  {'algorithm': 'auto', 'n_neighbors': 10, 'p': 1}
accuracy : 0.8482142857142858
```

➢ The Decisions tree classifier, with the best parameters achieves the highest accuracy

Results: Predictive Analysis

- Decision Tree classifier
  - Based on the prior accuracy comparisons between all four classifiers, the DT model demonstrates the best accuracy at 0.8892
  - DT Confusion Matrix: The number of true positives and true negatives outweighs false positives and false negatives



Conclusions

- Variable prior launch factors can influences the success rates of previous and future launch missions
- For SpaceY, the focus of launch site **KSC LC-39A** is recommended
- The orbits of **ES-L1**, **GEO**, **HEO**, and **SSO** achieve the highest success rates
- The successful landings appear to improve over time for all missions, with improvements to booster version and variability
- A payload mass of above **8,000 kg** appear to achieve a lower failure rate
- The **Decision Tree classifier** can be employed to predict future successful launches and landing