100% Made with Linux and LibreOffice

IBM Professional Certificate Data Science

Winning Space Race with Data Science

Omer Fatih Kaya
15 May 2023
https://github.com/aeromars/data_science_capstone/

Capstone Project

Outline

Executive Summary

- Major success came through in 2015 indicating technological advances. These rockets are in development phase and success rate expected to increase over time as flight numbers increase.
- Launch site distribution are CCAFS SLC 40 (61.11%), KSC LC 39A (24.44%), and VAFB SLC 4E (14.44%), but VAFB SLC 4E discontinued. Since these sites are in similar ranges to equator, launch sites are almost indifferent for space projectile motion as shown in data.
- Orbit distribution are GTO (27), ISS (21), VLEO (14), PO (9), LEO (7), SSO (5), MEO (3), ES-L1 (1), GEO (1), HEO (1), SO (1).
- Rockets have version numbers indicating technological advances. As technological advances increase, other factors such as orbit types, launch sites are expected to be less importance. In another words, as flight numbers increase, other factors mentioned are less importance.

Executive Summary

- Prediction model for rocket first stage success has accuracy of 83.33%. Remarkably, all prediction models (Logistic Regression, Support Vector Machine, Decision Tree, K-Nearest Neighbors) have same accuracy of 83.33%.

Introduction

- Analyzing SpaceX, specifically rocket first stage, makes SpaceX competitive for space cargo. Rocket first stage includes rocket reusability and rocket landing success rate. Market prices for space cargo range from $62 million offered by SpaceX to $165 million by other companies.
- We will analyze SpaceX rocket first stage to conclude whether feasible to compete by incumbents or prospective companies. Measurements are rocket first stage success rate, payload mass, orbit types, launch sites.
- Space launch sites are more important compared to air ports because space projectile motion are different from air projectile motion. Since launch sites are near equator, launch sites are insignificant.

Flowchart Diagram/Steps Methodology

- Flowchart Diagram/Steps proceeding from 1 to 10.

```
[A] Methodology: Data Collection & Wrangling
        [1] Data Collection
        [2] Data Wrangling
[B] Methodology: EDA & Interactive Visualizations
        [3] EDA
        [4] Features Engineering
[C] Methodology: Predictive Analysis (Classification)
        [5] Standardize Data
        [6] Split Train and Test Datasets
        [7] Machine Learning Algorithms
        [8] Train Models
        [9] Test Models
        [10] Feature Selection
```

Data Collection

- Two processes selected to collect data
  - SpaceX API (Unofficial)
  - Web Scraping (Wikipedia)

Data Collection API
Methodology: Data Collection & Wrangling
Data Collection

- SpaceX API (Unofficial)
  - https://api.spacexdata.com/v4/launches/past
- IBM Skills Network Dataset
  - https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/API_call_spacex_api.json
- Dataset contains identification number as references. For example, rocket column has no information about rocket but only an identification number. These identification numbers must be used to extract more information from different API endpoints. Specifically rocket, payloads, launchpad, and cores columns are extracted.
- Source
  - aeromars Capstone Data Collection API

Web Scraping
Methodology: Data Collection & Wrangling
Data Collection

- requests, pandas, re, unicodedata, BeautifulSoup packages used for web scraping.
- Extract Falcon 9 launch records HTML table from Wikipedia (List of Falcon 9 and Falcon Heavy launches on 9 June 2019).
  - https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922
- Parse table and convert it to Pandas dataframe.
- Source
  - aeromars Capstone Data Collection via Web Scraping

Data Wrangling
Methodology: Data Collection & Wrangling
Data Wrangling

- Pandas and Numpy used for performing basic statistics.
- Replace missing values with mean.
- Create binary (Success = 1, Fail = 0) "Class" column from "Outcome" category for training supervised models label. In another words create a landing outcome label from "Outcome" column.
- Filter only Falcon 9 rockets from dataset.

EDA Basic Statistics
Methodology: EDA & Interactive Visualizations
EDA

- Aggregate Site Launches
- Aggregate Orbit Types
- Aggregate Mission Outcome per Orbit Type
- Aggregate Success Rate

EDA with Data Visualization
Methodology: EDA & Interactive Visualizations
EDA

- Examining relationships:
  - Flight Number and Launch Sites
  - Payload and Launch Sites
  - Success Rate and Orbit Types
  - Flight Number and Orbit Types
  - Payload and Orbit Types
  - Average Success Rate (Year and Average Success Rate)

EDA with SQL
Methodology: EDA & Interactive Visualizations
EDA

- Task 1
  - %sql SELECT DISTINCT(Launch_Site) FROM SPACEXTBL;
- Task 2
  - %sql SELECT * FROM SPACEXTBL WHERE Launch_Site LIKE 'CCA%' LIMIT 5;
- Task 3
  - %sql SELECT SUM("PAYLOAD_MASS__KG_") FROM SPACEXTBL WHERE Customer = 'NASA (CRS)';
- Task 4
  - %sql SELECT AVG("PAYLOAD_MASS__KG_") FROM SPACEXTBL WHERE "Booster_Version" = 'F9 v1.1';

EDA with SQL
Methodology: EDA & Interactive Visualizations
EDA

- Task 5
  - %sql SELECT MIN(Date) FROM SPACEXTBL WHERE "Landing _Outcome" = 'Success (ground pad)';
- Task 6
  - %sql SELECT "Booster_Version" FROM SPACEXTBL WHERE "Landing _Outcome" = "Success (drone ship)" AND ("PAYLOAD_MASS__KG_" BETWEEN 4000 AND 6000);
- Task 7
  - %sql SELECT "Mission_Outcome", COUNT("Mission_Outcome") FROM SPACEXTBL GROUP BY "Mission_Outcome";

EDA with SQL
Methodology: EDA & Interactive Visualizations
EDA

- Task 8
  - %sql SELECT Booster_Version FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL);
- Task 9
  - Skills Network

    %sql SELECT substr(Date, 4, 2) AS Dates, "Booster_Version", "Launch_Site", "Landing _Outcome" FROM SPACEXTBL WHERE "Landing _Outcome" = 'Failure (drone ship)' AND substr(Date, 7, 4) = '2015';
  - Watson

    %sql SELECT "Date", "Booster_Version", "Launch_Site", "Landing _Outcome" FROM SPACEXTBL WHERE "Landing _Outcome" = 'Failure (drone ship)' AND YEAR("Date") = 2015;

EDA with SQL
Methodology: EDA & Interactive Visualizations
EDA

- Task 10
  - Skills Network

    %sql SELECT "Landing _Outcome", COUNT(*) AS Counter FROM SPACEXTBL WHERE substr(Date, 4, 2) BETWEEN '04062010' AND '20032017' GROUP BY "Landing _Outcome" ORDER BY Counter DESC;

  - Watson

    %sql SELECT "Landing _Outcome", COUNT(*) AS Counter FROM SPACEXTBL WHERE "Date" BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY "Landing _Outcome" ORDER BY Counter DESC;

Interactive Map with Folium
Methodology: EDA & Interactive Visualizations
EDA

- Mark all launch site locations on map. All launch sites located near equator.
- Launch outcomes shown on map for easy statistic on the fly. In our case we only have success or failure outcomes, but we can optionally display other features on map.
- Explore map for possible collateral near launch site locations.

Dashboard with Plotly Dash
Methodology: EDA & Interactive Visualizations
EDA

- Launch sites and success rate graphs are compared for best locations.
- Payload and success rate graph examined for best rocket selections.

Features Engineering

- Use dummy variables on categorical columns.
- Cast all numeric columns to float64 type.

Standardize Data

Methodology: Predictive Analysis (Classification)

Standardize Data

- pandas, numpy, matplotlib, seaborn, preprocessing (sklearn), train_test_split, GridSearchCV, LogisticRegression, SVC, DecisionTreeClassifier, KNeighborsClassifier packages used.
- Synchronize two datasets
- Dataset links
  - https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/dataset_part_2csv
  - https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/dataset_part_3.csv

Standardize Data

- Standardize data using sklearn

```
standardization_prep = preprocessing.StandardScaler().fit(X)
X_standard = standardization_prep.transform(X)
X = X_standard
```

Split Train and Test Datasets

- Split dataset into train and test datasets using sklearn's train_test_split function with parameters test_size = 0.2 and random_state = 2.

  X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size = 0.2, random_state = 2)

Machine Learning Algorithms

- Four algorithms used:
  - Logistic Regression
  - Support Vector Machines
  - Decision Tree
  - K-Nearest Neighbors

Train Models

- Best hyperparameters via sklearn's GridSearchCV with parameters
  - Logistic Regression parameters

    parameters ={"C":[0.01,0.1,1],'penalty':['l2'], 'solver':['lbfgs']}# l1 lasso l2 ridge
  - Support Vector Machines parameters

    parameters = {'kernel':('linear', 'rbf','poly','rbf', 'sigmoid'), 'C': np.logspace(-3, 3, 5), 'gamma':np.logspace(-3, 3, 5)}
  - Decision Tree

    parameters = {'criterion': ['gini', 'entropy'], 'splitter': ['best', 'random'], 'max_depth': [2*n for n in range(1,10)], 'max_features': ['auto', 'sqrt'], 'min_samples_leaf': [1, 2, 4], 'min_samples_split': [2, 5, 10]}

Train Models

- K-Nearest Neighbors

  parameters = {'n_neighbors': [1, 2, 3, 4, 5, 6, 7, 8, 9, 10], 'algorithm': ['auto', 'ball_tree', 'kd_tree', 'brute'], 'p': [1,2]}

- Fit the models using training data
- Training set accuracy via best_score_ function.

Methodology: Predictive Analysis (Classification)
Test Models

- Training set accuracy via best_score_ function.
- Test models accuracy using test data.
- Compare all models' accuracy.
- Compare all models' confusion matrices.

Feature Selection

- Evaluate which features for optimizing future models. From analyzing data, rocket versions, payload mass, and launches suitable for building future models while remaining features are insignificant.

EDA Basic Statistics
Results: EDA & Interactive Visualizations
EDA

- Aggregate Success Rate: 0.6666
- Source:
  - aeromars Capstone Data Wrangling

```
Aggregate Site Launches
        CCAFS SLC 40: 55
        KSC LC 39A: 22
        VAFB SLC 4E: 13
```

```
Aggregate Orbit Types
        GTO: 27
        ISS: 21
        VLEO: 14
        PO: 9
        LEO: 7
        SSO: 5
        MEO: 3
        ES-L1: 1
        GEO: 1
        HEO: 1
        SO: 1
```

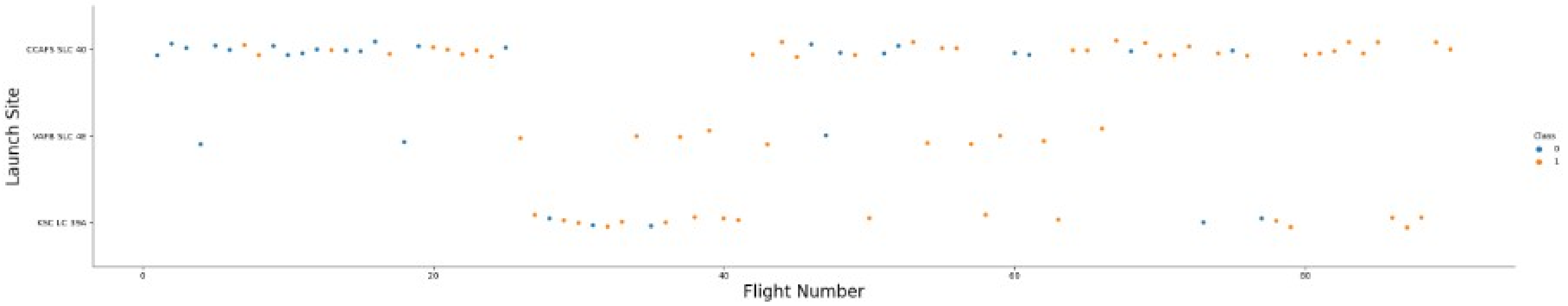
```
Aggregate Mission Outcome per Orbit Types
        True ASDS: 41
        None None: 19
        True RTLS: 14
        False ASDS: 6
        True Ocean: 5
        False Ocean: 2
        None ASDS: 2
        False RTLS: 1
```

Flight Numbers and Launch Sites
Results: EDA & Interactive Visualizations
EDA

- We see that different launch sites have different success rates. CCAFS LC-40 has a success rate of 60% while KSC LC-39A and VAFB SLC-4E has a success rate of 77%.
- Initial launches from CCAFS SLC are less successful and continuously success rate increase where flight number over 80 are all success. CCAFS SLC site is also highest number of launches occurred.
- VAFB SLC site is the least used site with very high success rate.

Flight Numbers and Launch Sites
Results: EDA & Interactive Visualizations
EDA

- Source:
  - aeromars Capstone EDA via Visualizations

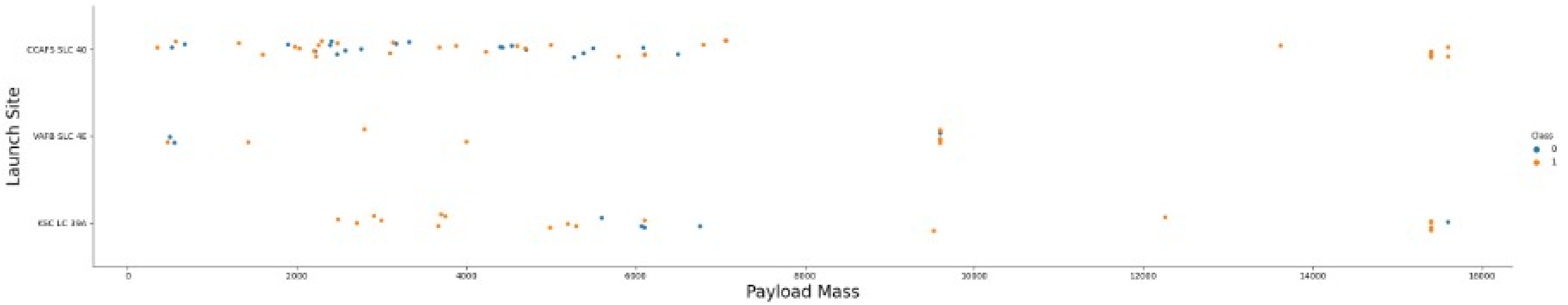


Payload and Launch Sites
Results: EDA & Interactive Visualizations
EDA

- VAFB-SLC launch site there are no rockets launched for heavy payload mass greater than 10000.
- CCAFS SLC has more failures compared to other sites and also most used site.
- There aren't many rocket launches for payload mass greater than 7000.
- Hardly any failures for payload mass over 9000.

Payload and Launch Sites
Results: EDA & Interactive Visualizations
EDA

- Source:
  - aeromars Capstone EDA via Visualizations

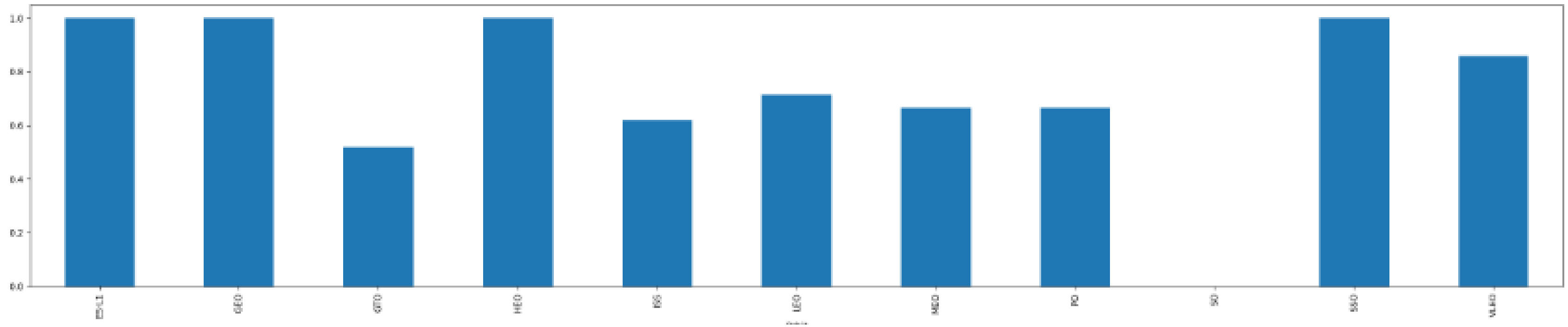


Success Rate and Orbit Types
Results: EDA & Interactive Visualizations
EDA

- ES-L1, GEO, HEO, SSO orbit types have perfect success rate. VLEO orbit type is the next success rate. SO orbit type has no success rate. GTO orbit type have least success rate with over 55%.
- Aggregate success rate for orbit types seems to be over 80%.

Success Rate and Orbit Types
Results: EDA & Interactive Visualizations
EDA

- Source:
  - aeromars Capstone EDA via Visualizations

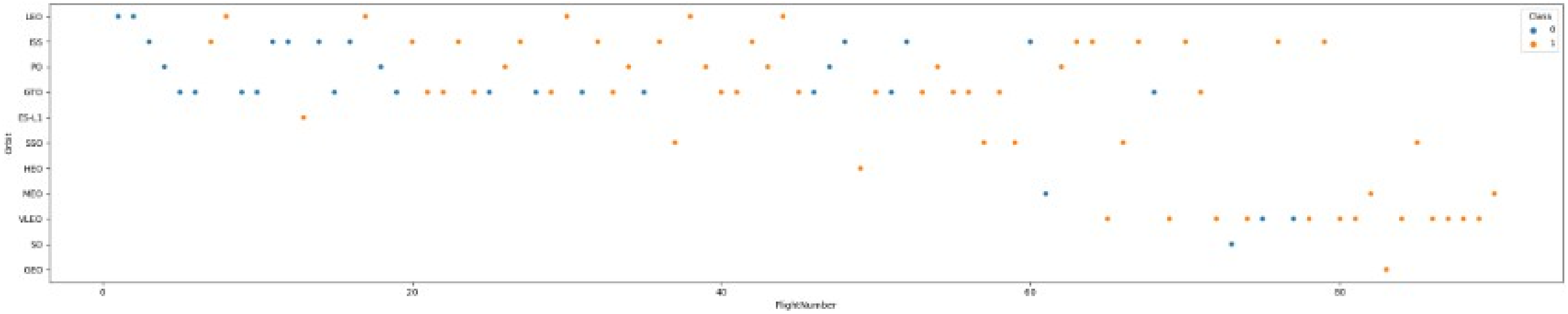


Flight Number and Orbit Types
Results: EDA & Interactive Visualizations
EDA

- LEO orbit type have success appears related to number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.

Flight Number and Orbit Types
Results: EDA & Interactive Visualizations
EDA

- Source:
  - aeromars Capstone EDA via Visualizations

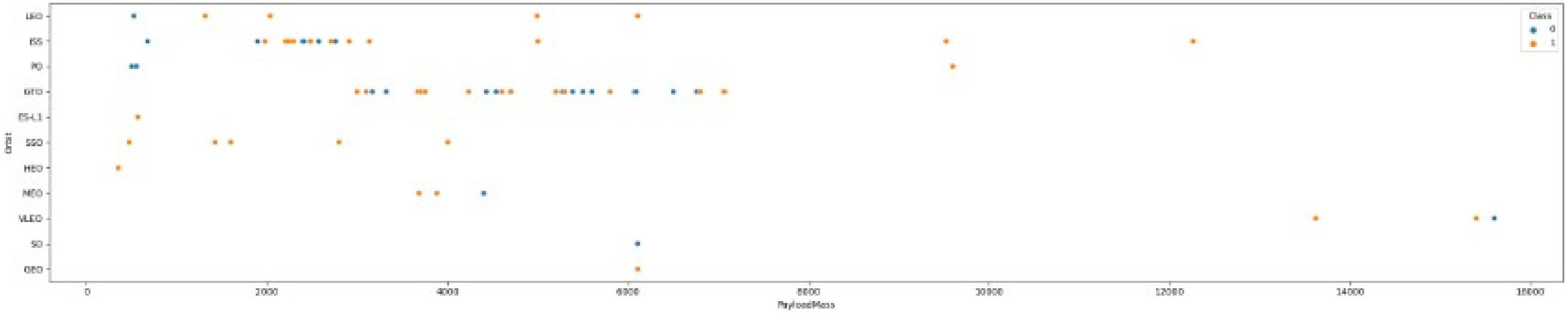


Payload and Orbit Types
Results: EDA & Interactive Visualizations
EDA

- With heavy payloads the successful landing or positive landing rate are more for Polar, LEO, ISS orbit types. However for GTO orbit type we can't distinguish this well as both positive landing rate and negative landing (unsuccessful mission) are both there.
- GTO orbit type have most frequent number of successes and failures followed by ISS orbit type.

Payload and Orbit Types
Results: EDA & Interactive Visualizations
EDA

- Source:
  - aeromars Capstone EDA via Visualizations



Average Success Rate (Year and Average Success Rate)

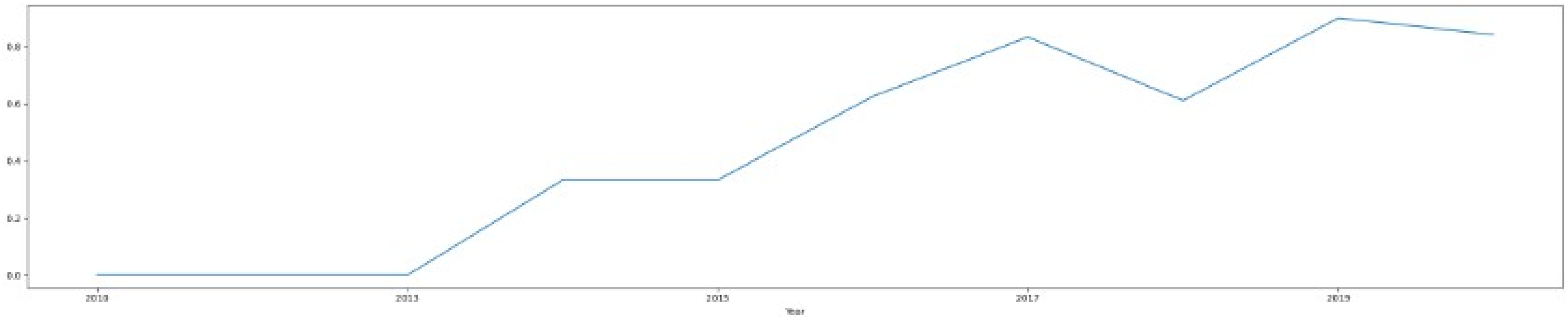
Results: EDA & Interactive Visualizations

EDA

- You can observe that success rate since 2013 increasing until 2020.
- Average success rate increased reaching max in year 2019 with over 80% value.

Average Success Rate (Year and Average Success Rate) Results: EDA & Interactive Visualizations EDA

- Source:
  - aeromars Capstone EDA via Visualizations



EDA with SQL
Results: EDA & Interactive Visualizations
EDA

TASK 1

Description:
Display the names of the unique launch sites in the space mission.
Query:
```
%sql SELECT DISTINCT(Launch_Site) FROM SPACEXTBL;
```
Result:
CCAFFS LC-40, VAFB SLC-4E, KSC LC-39A, CCAFS SLC-40

TASK 2

Description:
Display 5 records where launch sites begin with the string 'CCA'.
Query:
```
%sql SELECT * FROM SPACEXTBL WHERE Launch_Site LIKE 'CCA%' LIMIT 5;
```
Result:

EDA with SQL
Results: EDA & Interactive Visualizations
EDA

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

EDA with SQL
Results: EDA & Interactive Visualizations
EDA

```
TASK 3

Description:
Display the total payload mass carried by boosters launched by NASA (CRS).
Query:
%sql SELECT SUM("PAYLOAD_MASS__KG_") FROM SPACEXTBL WHERE Customer = 'NASA (CRS)';
Result:
45596
```

```
TASK 4

Description:
Display average payload mass carried by booster version F9 v1.1.
Query:
%sql SELECT AVG("PAYLOAD_MASS__KG_") FROM SPACEXTBL WHERE "Booster_Version" = 'F9 v1.1';
Result:
2928
```

EDA with SQL
Results: EDA & Interactive Visualizations
EDA

TASK 5

Description:
List the date when the first successful landing outcome in ground pad was achieved.
Query:
```
%sql SELECT MIN(Date) FROM SPACEXTBL WHERE "Landing _Outcome" = 'Success (ground pad)';
```
Result:
01-05-2017

TASK 6

Description:
List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000.
Query:
```
%sql SELECT "Booster_Version" FROM SPACEXTBL WHERE "Landing _Outcome" = "Success (drone ship)" AND
("PAYLOAD_MASS__KG_" BETWEEN 4000 AND 6000);
```
Result:
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

EDA with SQL
Results: EDA & Interactive Visualizations
EDA

TASK 7

Description:
List the total number of successful and failure mission outcomes.
Query:

```
%sql SELECT "Mission_Outcome", COUNT("Mission_Outcome") FROM SPACEXTBL GROUP BY "Mission_Outcome";
```

Result:
Failure (in flight): 1
Success: 98
Success: 1
Success (payload status unclear): 1

EDA with SQL
Results: EDA & Interactive Visualizations
EDA

```
TASK 8

Description:
List the names of the booster_versions which have carried the maximum payload mass. Use a subquery.
Query:
%sql SELECT Booster_Version FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL);
Result:
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
```

EDA with SQL
Results: EDA & Interactive Visualizations
EDA

```
TASK 9

Description:
List the records which will display the month names, failure landing outcomes in drone ship, booster versions,
launch site for the months in year 2015.
Query:
# Skills Network solution
%sql SELECT substr(Date, 4, 2) AS Dates, "Booster_Version", "Launch_Site", "Landing _Outcome" FROM SPACEXTBL WHERE
"Landing _Outcome" = 'Failure (drone ship)' AND substr(Date, 7, 4) = '2015';

# Watson solution
# %sql SELECT "Date", "Booster_Version", "Launch_Site", "Landing _Outcome" FROM SPACEXTBL WHERE "Landing _Outcome"
= 'Failure (drone ship)' AND YEAR("Date") = 2015;
Result:
Dates   Booster Version Launch Site     Landing Outcome
01      F9 v1.1 B1012   CCAFS LC-40     Failure (drone ship)
04      F9 v1.1 B1015   CCAFS LC-40     Failure (drone ship)
```

EDA with SQL
Results: EDA & Interactive Visualizations
EDA

TASK 10

Description:
Rank the count of successful landing outcomes between the date 04-06-2010 and 20-03-2017 in decending order.
Query:

```
# Skills Network solution
%sql SELECT "Landing _Outcome", COUNT(*) AS Counter FROM SPACEXTBL WHERE substr(Date, 4, 2) BETWEEN '04062010' AND
'20032017' GROUP BY "Landing _Outcome" ORDER BY Counter DESC;

# Watson solution
# %sql SELECT "Landing _Outcome", COUNT(*) AS Counter FROM SPACEXTBL WHERE "Date" BETWEEN '2010-06-04' AND
'2017-03-20' GROUP BY "Landing _Outcome" ORDER BY Counter DESC;
```

Result:
Success: 31
No attempt: 11
Success (drone ship): 10
Success (ground pad): 7
Uncontrolled (ocean): 2
Failure (parachute): 2
Controlled (ocean): 2
Precluded (drone ship): 1
No attempt: 1
Failure (drone ship): 1
Failure: 1

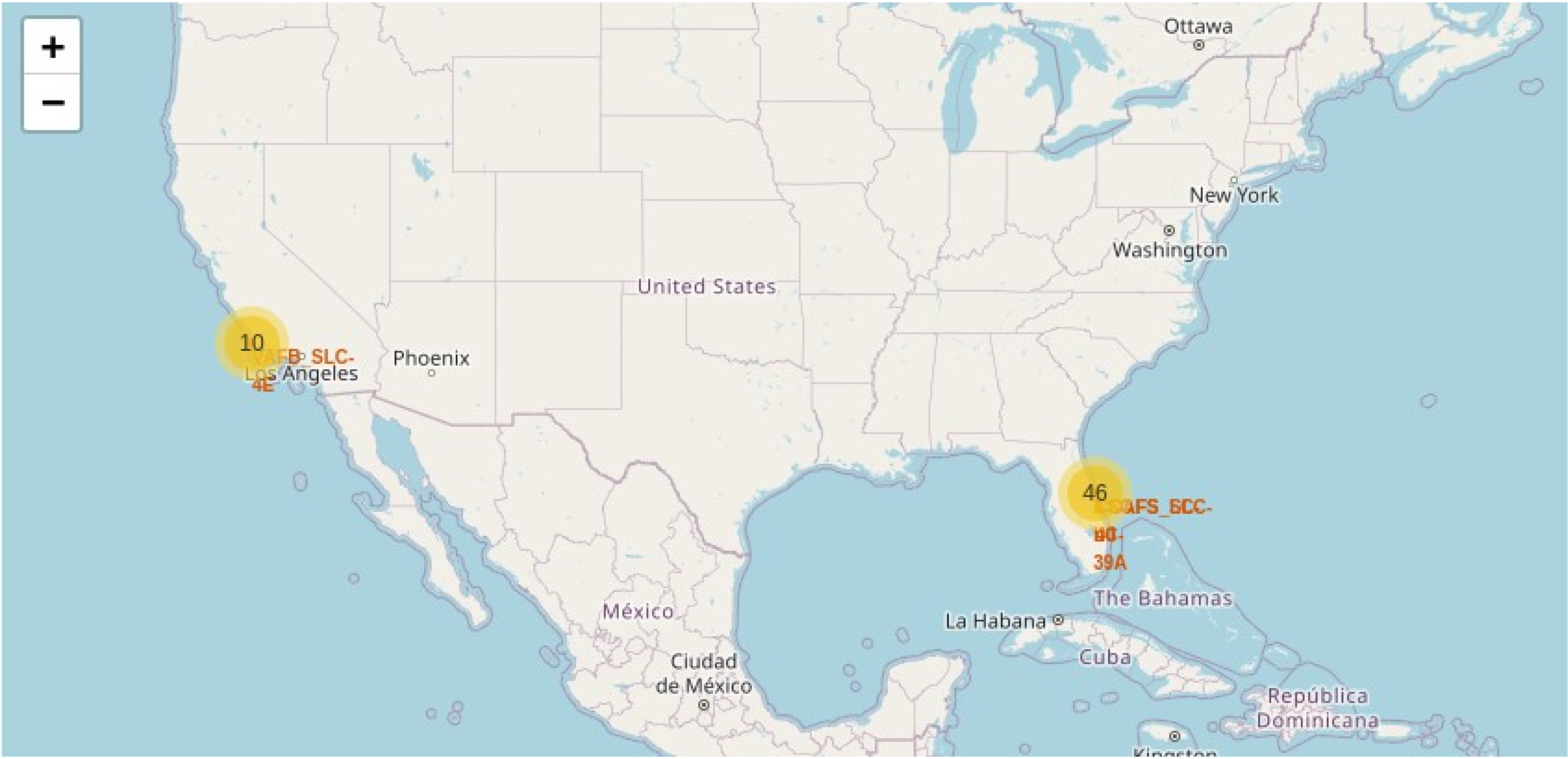
EDA with SQL
Results: EDA & Interactive Visualizations
EDA

- Source:
  - aeromars Capstone EDA via SQL

Interactive Map with Folium
Results: EDA & Interactive Visualizations
EDA

- All launch sites located in southern states due to proximity to equator. For example, there are no launch sites in Chicago.
- Source:
  - aeromars Capstone Data Visualization via Folium

Interactive Map with Folium
Results: EDA & Interactive Visualizations
EDA

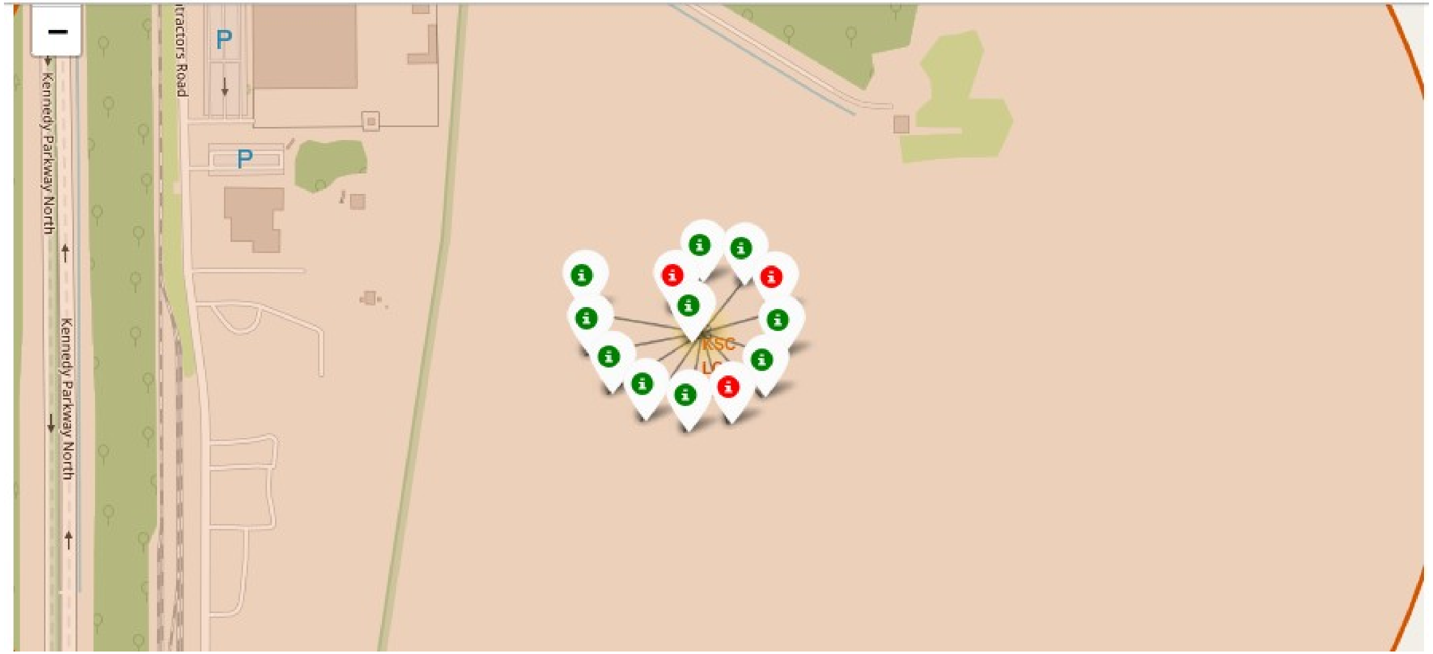


Interactive Map with Folium
Results: EDA & Interactive Visualizations
EDA

- Easy to compare statistics on the fly with map.
- Source:
  - aeromars Capstone Data Visualization via Folium

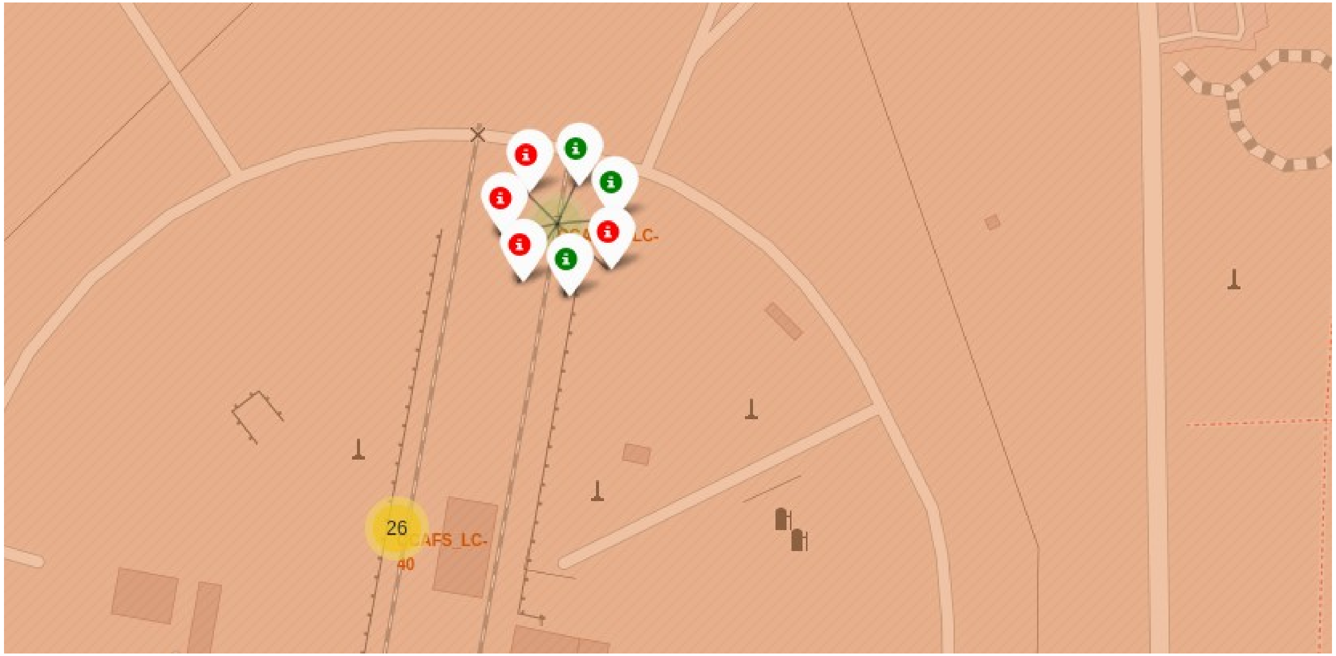
Interactive Map with Folium
Results: EDA & Interactive Visualizations
EDA

- KSC LC



Interactive Map with Folium
Results: EDA & Interactive Visualizations
EDA

- CCAFS SLC

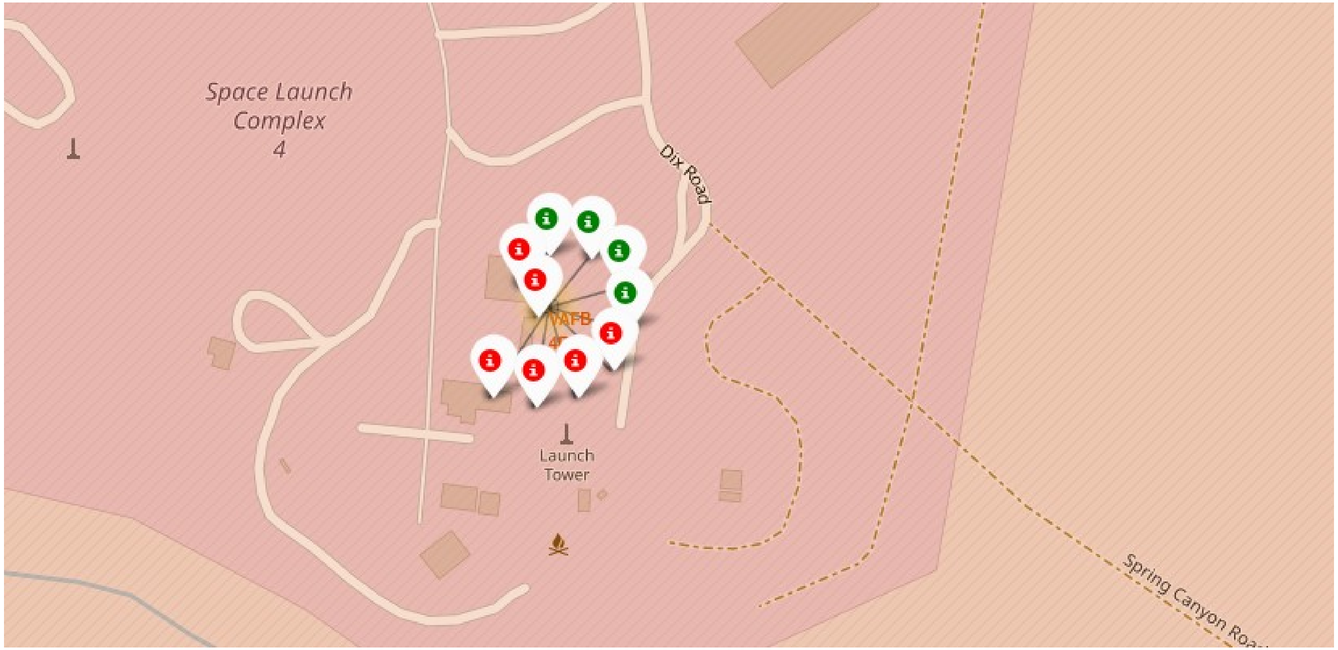


Interactive Map with Folium
Results: EDA & Interactive Visualizations
EDA

- CCAFS LC

Interactive Map with Folium
Results: EDA & Interactive Visualizations
EDA

- VAFB SLC



Interactive Map with Folium
Results: EDA & Interactive Visualizations
EDA

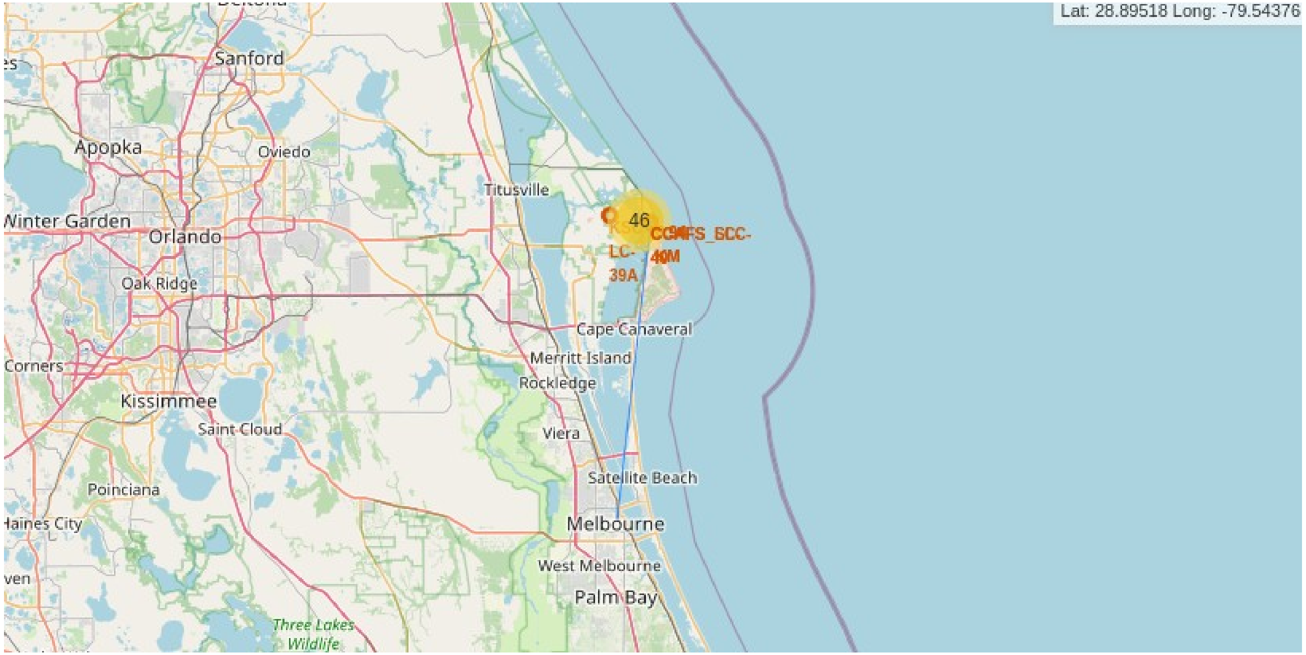
- For safety reasons, all launch sites are in reserved areas far from population. Coastlines seem ideal for rockets in test phrase as oceans capture failed rocket debris.
- Source:
  - aeromars Capstone Data Visualization via Folium

Interactive Map with Folium
Results: EDA & Interactive Visualizations
EDA

- Launch sites close to coastlines

Interactive Map with Folium
Results: EDA & Interactive Visualizations
EDA

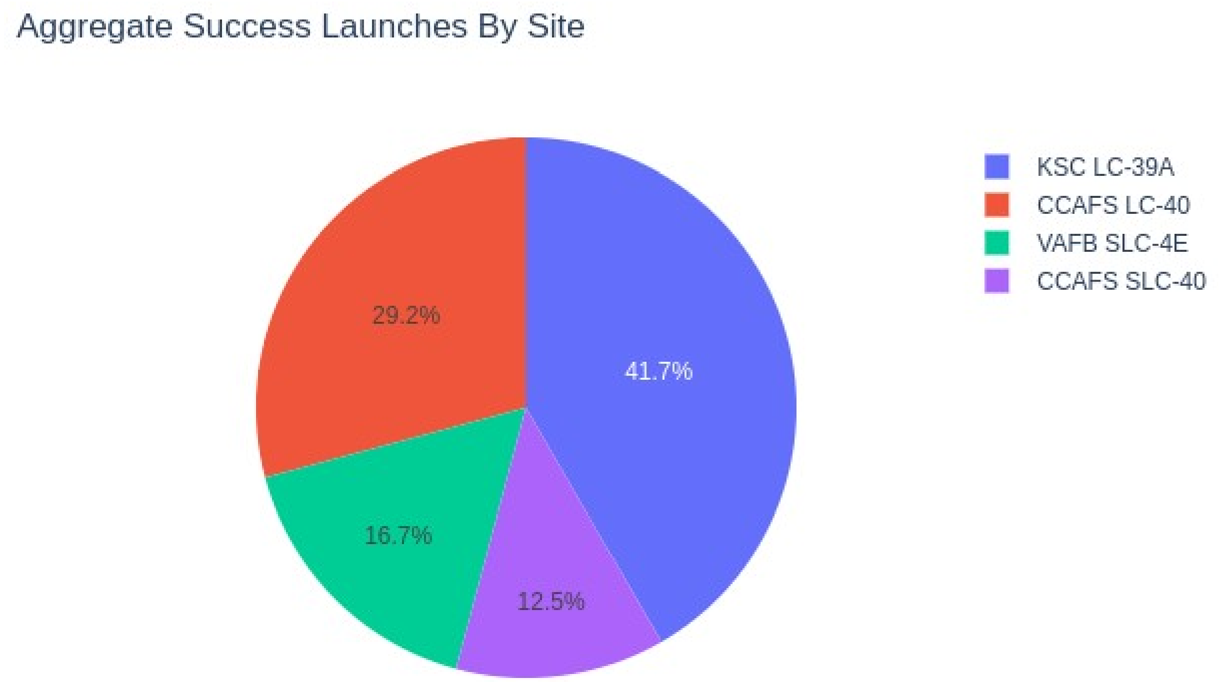
- Launch sites far from cities



Dashboard with Plotly Dash
Results: EDA & Interactive Visualizations
EDA

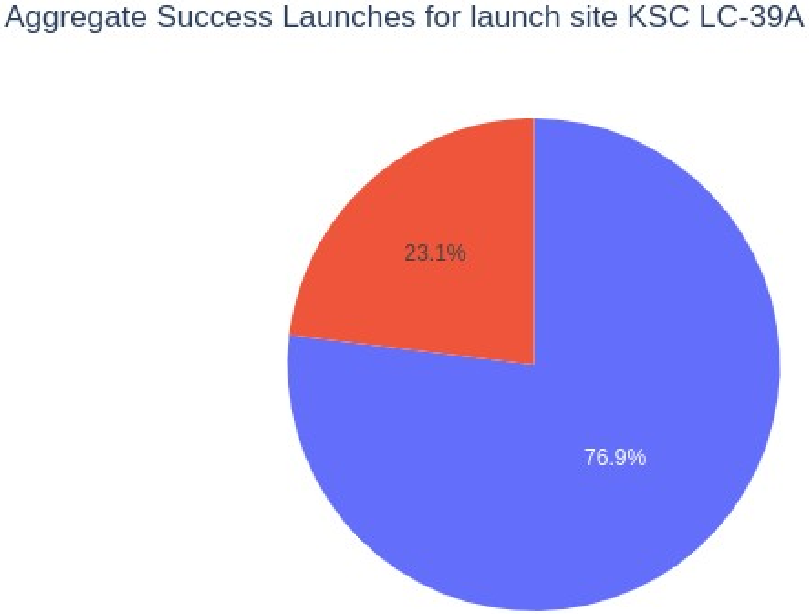
- Aggregate launch success for all launch sites
- Aggregate launch success for specific sites.
- Source:
  - aeromars Capstone Data Visualization via Dash

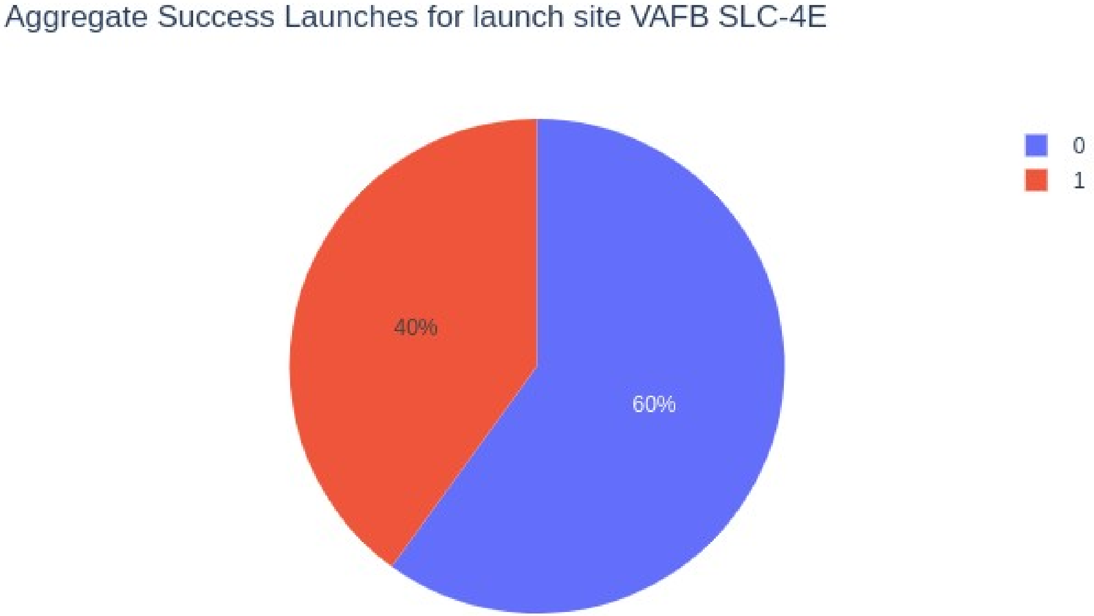
Dashboard with Plotly Dash
Results: EDA & Interactive Visualizations
EDA



Dashboard with Plotly Dash
Results: EDA & Interactive Visualizations
EDA

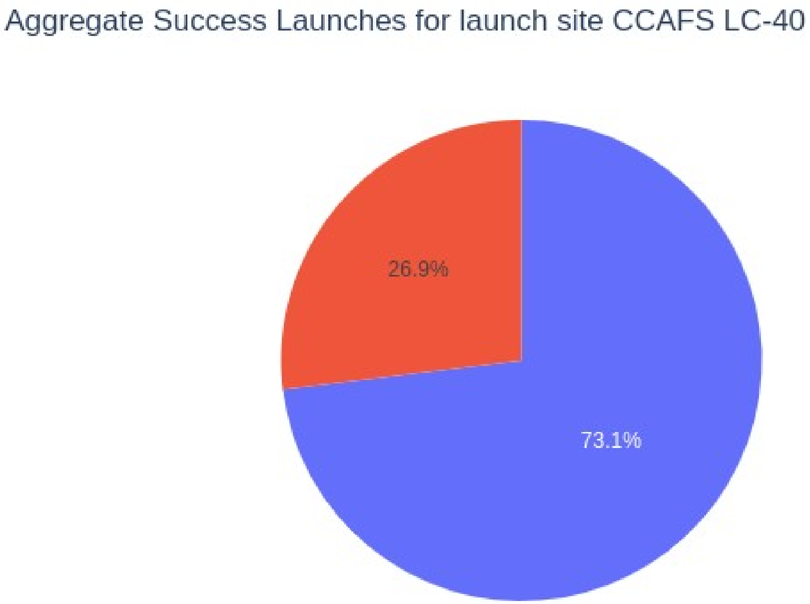
- KSC LC-39A has highest success ratio.
- 1 = Success, 0 = Fail

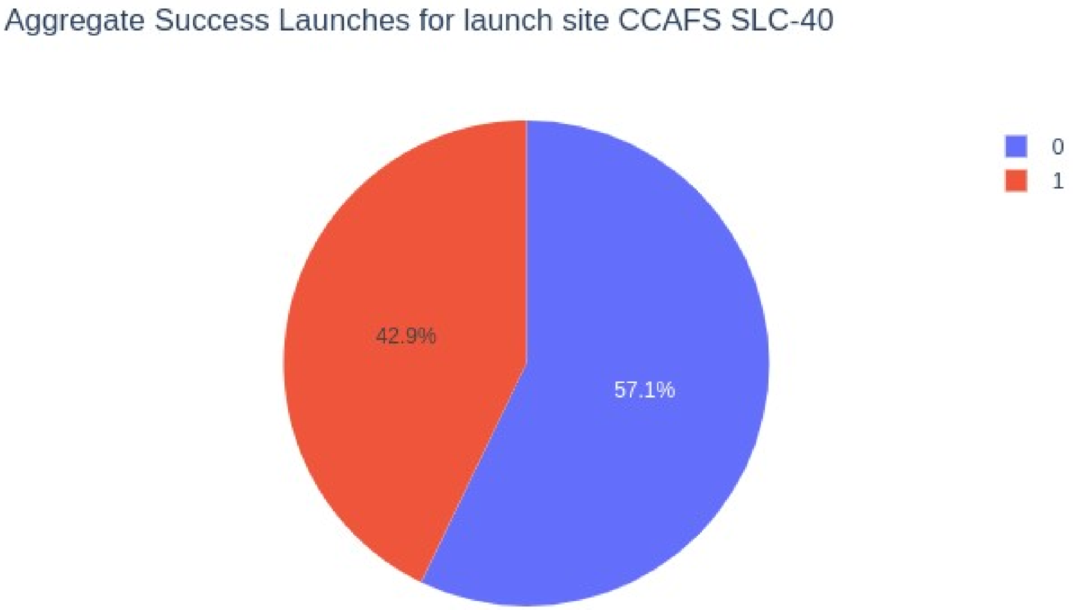




Dashboard with Plotly Dash
Results: EDA & Interactive Visualizations
EDA

- 1 = Success, 0 = Fail



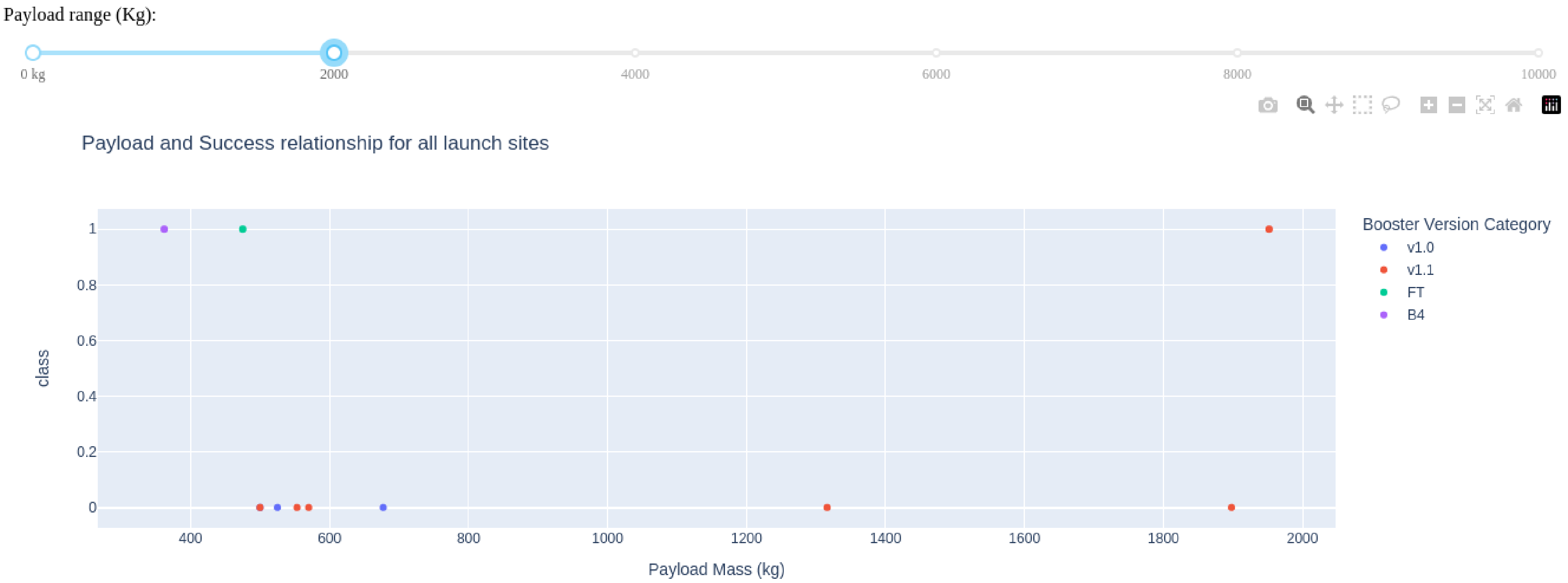


Dashboard with Plotly Dash
Results: EDA & Interactive Visualizations
EDA

Payload and Launch Outcome scatter plot for all launch sites with different payload selected in the range slider.

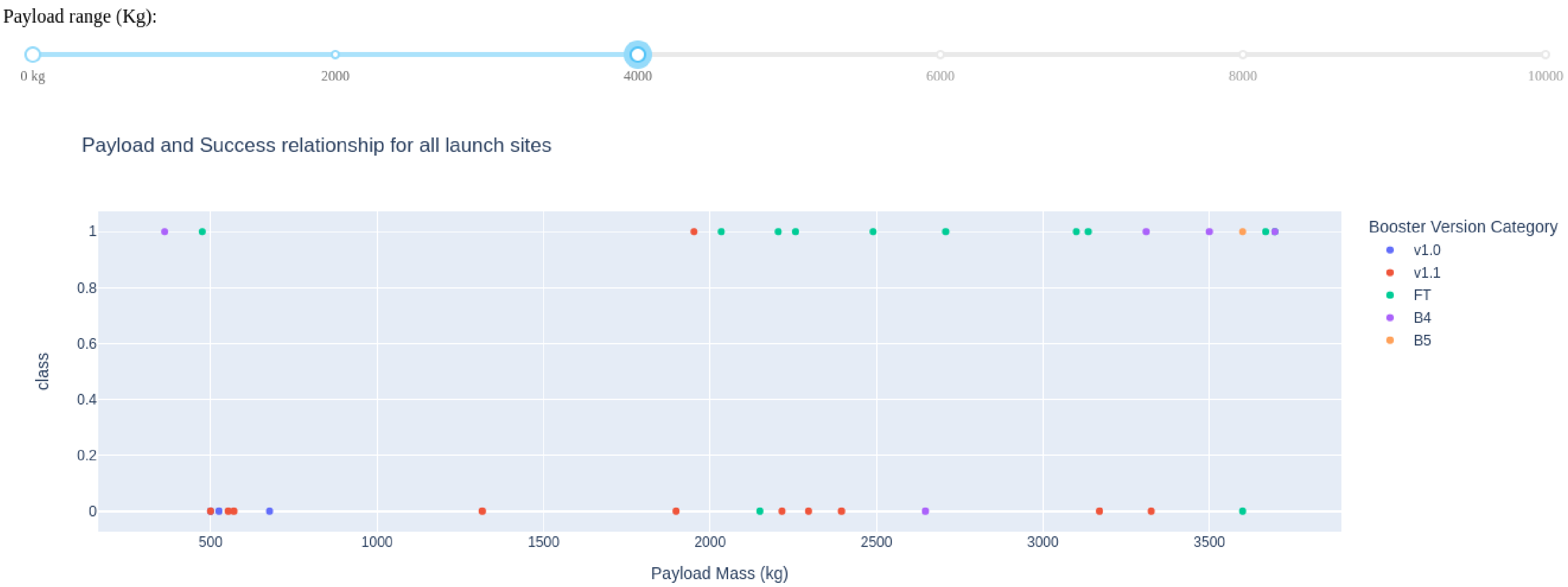
- Booster Version FT has best overall success rating from light payload to heavy payload compared to other Booster Versions. Booster Version v1.1 has worst overall success rating from light payload to heavy payload.
- Source:
  - aeromars Capstone Data Visualization via Dash

Dashboard with Plotly Dash Results: EDA & Interactive Visualizations
EDA



Dashboard with Plotly Dash Results: EDA & Interactive Visualizations

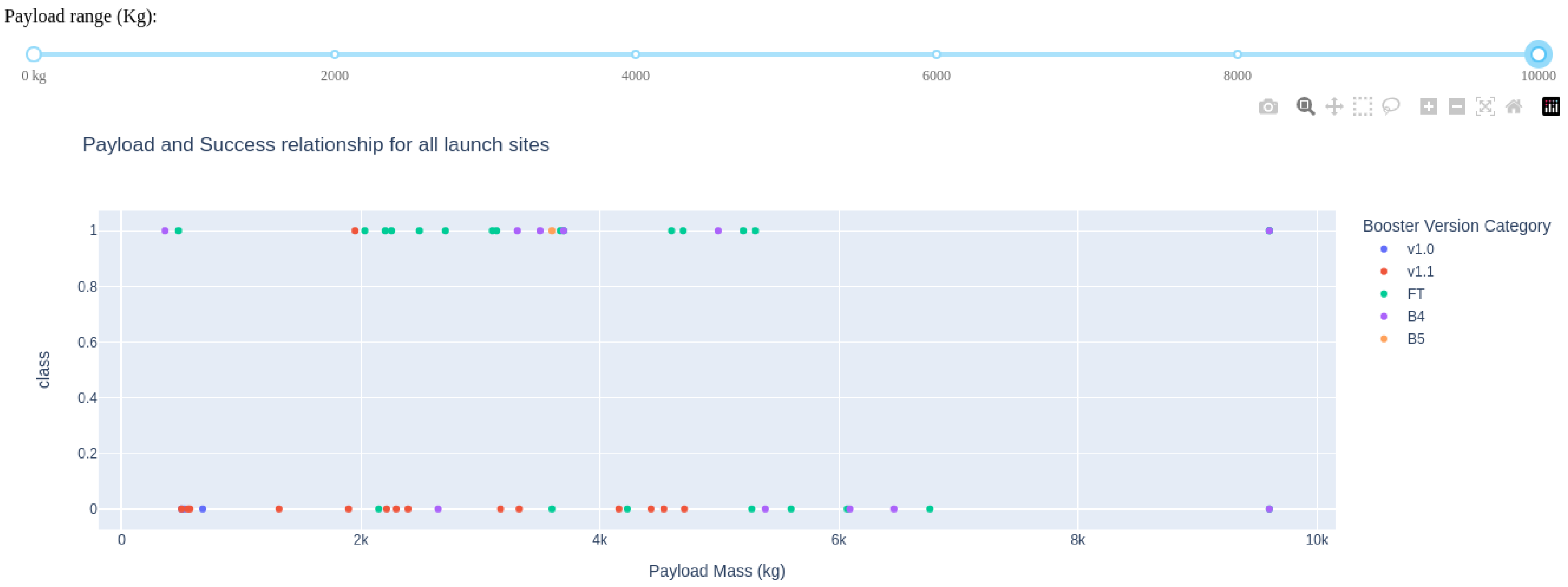
EDA



Dashboard with Plotly Dash
Results: EDA & Interactive Visualizations
EDA

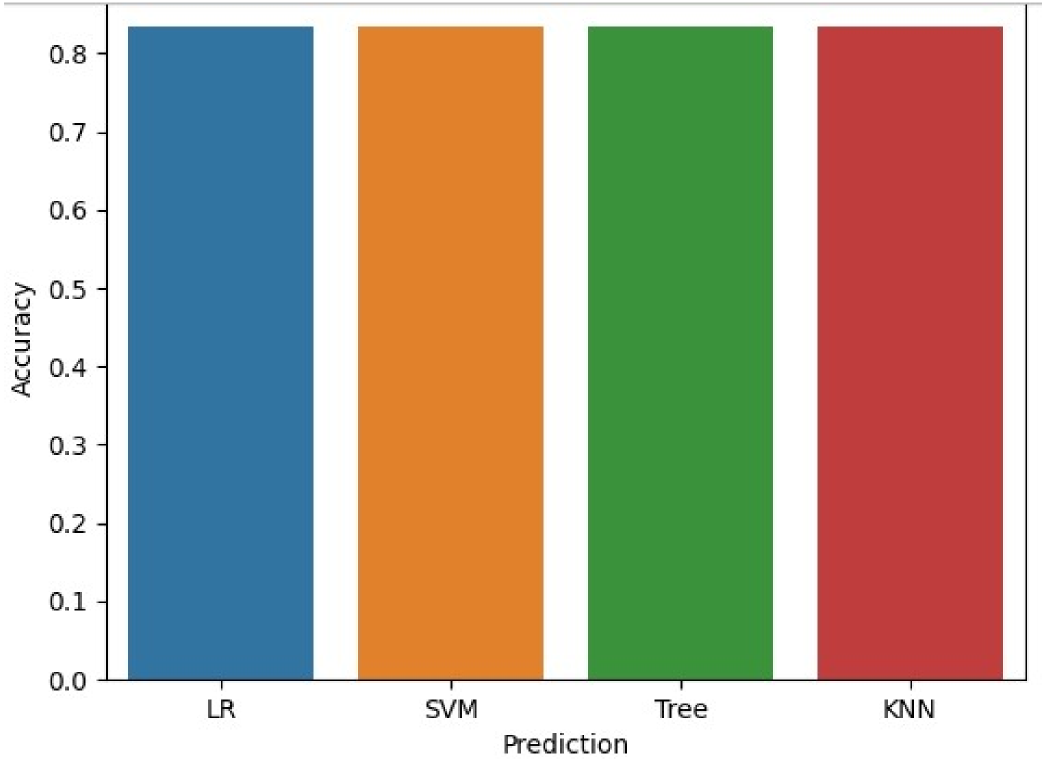
Dashboard with Plotly Dash
Results: EDA & Interactive Visualizations
EDA

Dashboard with Plotly Dash
Results: EDA & Interactive Visualizations
EDA



- All models have accuracy of 0.83333.
- Accuracy of 0.83333 is acceptable but a better model (besides mentioned four models) highly recommended.
- Source:
  - aeromars Capstone Predictions

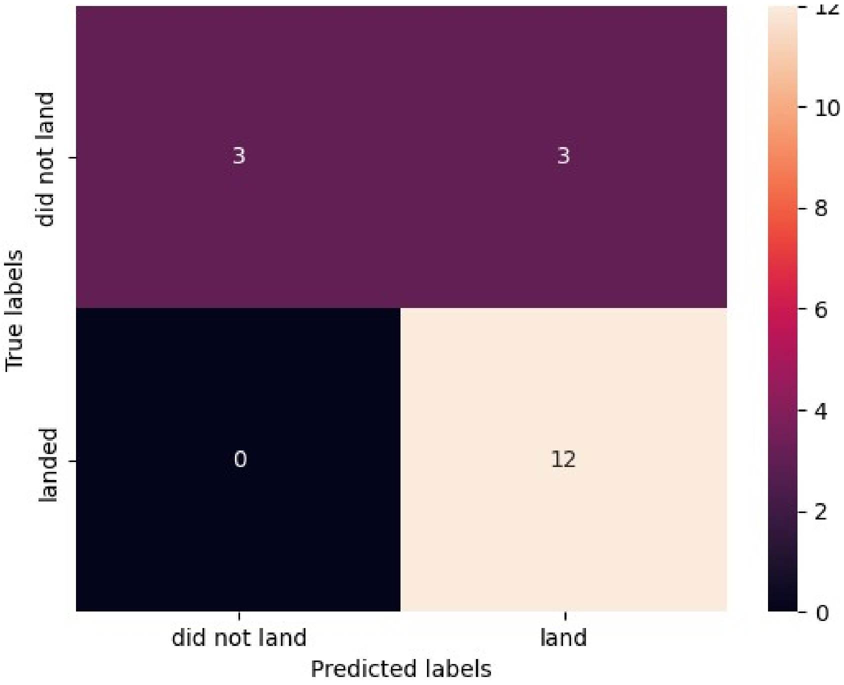
Classification Accuracy
Results: Predictive Analysis (Classification)
Predictive Analysis



Confusion Matrix
Results: Predictive Analysis (Classification)
Predictive Analysis

- All models have same confusion matrix.
- Confusion matrix show prediction model “did not land” as “landed” about 20% of the time.
- Source:
  - aeromars Capstone Predictions

Confusion Matrix
Results: Predictive Analysis (Classification)
Predictive Analysis



Results: Predictive Analysis (Classification)

Feature Selection

- All prediction models (Linear Regression, Support Vector Machines, Decision Tree, K-Nearest Neighbor) result in same accuracy of 0.8333 and confusion matrix. 0.8333 accuracy is acceptable, but a better model highly exists using different machine learning algorithms.
- Most important features are rocket versions, flight numbers, and payload mass. Remaining features are important but may be excluded to test new models for better model accuracy.

Conclusion

- Obvious that success rate mostly depends on product development denoted by rocket versions but also payload mass, flight numbers. As rocket versions increase, payload mass also increased, orbit types and launch sites become irrelevant.
- Rocket versions denote technological advances in particular rocket first stage. Later rocket versions have higher success rate.
- All prediction models (Linear Regression, Support Vector Machines, Decision Tree, K-Nearest Neighbor) result in same accuracy of 0.8333.

Appendix

- Capstone Project
  - https://github.com/aeromars/data_science_capstone/
- Capstone Project Data Collection API
  - https://github.com/aeromars/data_science_capstone/blob/main/Capstone_01-Data_Collection_API.ipynb
- Capstone Project Data Web Scraping
  - https://github.com/aeromars/data_science_capstone/blob/main/Capstone_02-Data_Collection_via_Web_Scraping.ipynb
- Capstone Project Data Wrangling
  - https://github.com/aeromars/data_science_capstone/blob/main/Capstone_03-Data_Wrangling.ipynb

Appendix

- Capstone Project EDA via SQL
  - https://github.com/aeromars/data_science_capstone/blob/main/Capstone_04-EDA_via_SQL.ipynb
- Capstone Project EDA via Visualization
  - https://github.com/aeromars/data_science_capstone/blob/main/Capstone_05-EDA_via_Visualizations.ipynb
- Capstone Project Data Visualization via Folium
  - https://github.com/aeromars/data_science_capstone/blob/main/Capstone_06-Data_Visualization_via_Folium.ipynb
- Capstone Project Data Visualization via Dash
  - https://github.com/aeromars/data_science_capstone/blob/main/Capstone_07-Data_Visualization_via_Dash.py

Appendix

- Capstone Project Predictions
  - https://github.com/aeromars/data_science_capstone/blob/main/Capstone_08-Predictions.ipynb
- Capstone Project Presentation