SpaceX launch analysis

Chee Ryoutarou

31 Dec 2021

OUTLINE

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

EXECUTIVE SUMMARY

- Competitive analysis of SpaceX revealed that the reusability of first stage is the reason for their cost savings
- Relationship between various factors and success rate was explored
- Machine learning pipeline was built to predict if Space X's launch will land

INTRODUCTION

Problem

- Space X's launch cost is market leading (62M vs 165M+)
- Cost savings is because they can reuse first stage
- By determining if the first stage will land, we can determine cost of launch

Objective

- Build a machine learning pipeline to predict if the first stage will land

METHODOLOGY

- Data collection
  - Data was obtained from SpaceX API and webscraping of Wikipedia
- Exploratory data analysis
  - Data wrangling
  - Interactive visual analytics with Folium & Plotly Dash
- Predictive analysis using machine learning models
  - Logistic regression
  - Support Vector Machine
  - Decision Tree
  - K Nearest neighbours
  - Hyperparameter tuning with GridsearchCV

METHODOLOGY – Data collection & data wrangling

- Data source
  - Space X API (https://api.spacexdata.com/v4/launches/past)
  - Wikipedia
- SpaceX API
  - GET request to request and parse SpaceX launch data
  - Filter data to only include Falcon 9 launches
- Data wrangling
  - Number of launches per site calculated
  - Number and occurence of each orbit calculated
  - Number and occurrence of mission outcome per orbit type calculated
  - Landing outcome label created from outcome column

METHODOLOGY – EDA & interactive visual analytics

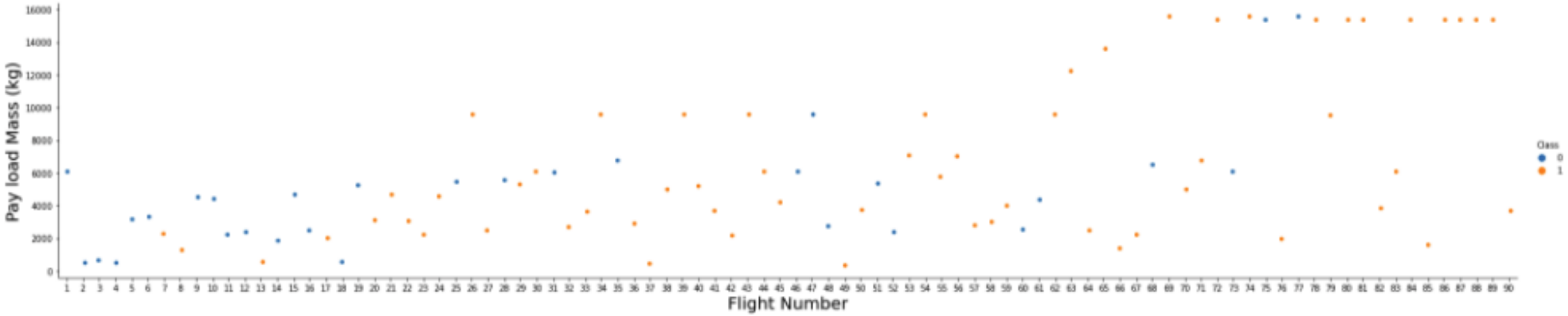
- Exploratory data analysis
  - Scatterplots were plotted to find patterns between variables
    - Load mass vs flight number vs success
    - Flight number vs launch site vs success
    - Payload vs launch site vs success
    - Flight number vs orbit type vs success
    - Payload vs orbit type vs success
  - Bar chart to find success rate by categories
    - Orbit vs success rate
  - Line chart to find trends
    - Success rate vs time
- Interactive visual analytics with Folium
  - All launch sites marked on map
  - Success/failed launched marked on map
  - Distance between launch site to proximites calculated
- Interactive dashboard with Plotly Dash
  - Success rate pie chart rendered based on site
  - Range slider to select payload to show success rate

METHODOLOGY – Predictive analysis

- Predictive analysis using machine learning models
  - Logistic regression
  - Support Vector Machine
  - Decision Tree
  - K Nearest neighbours
- Hyperparameter tuning with GridsearchCV
  - Judge based on score and confusion matrix

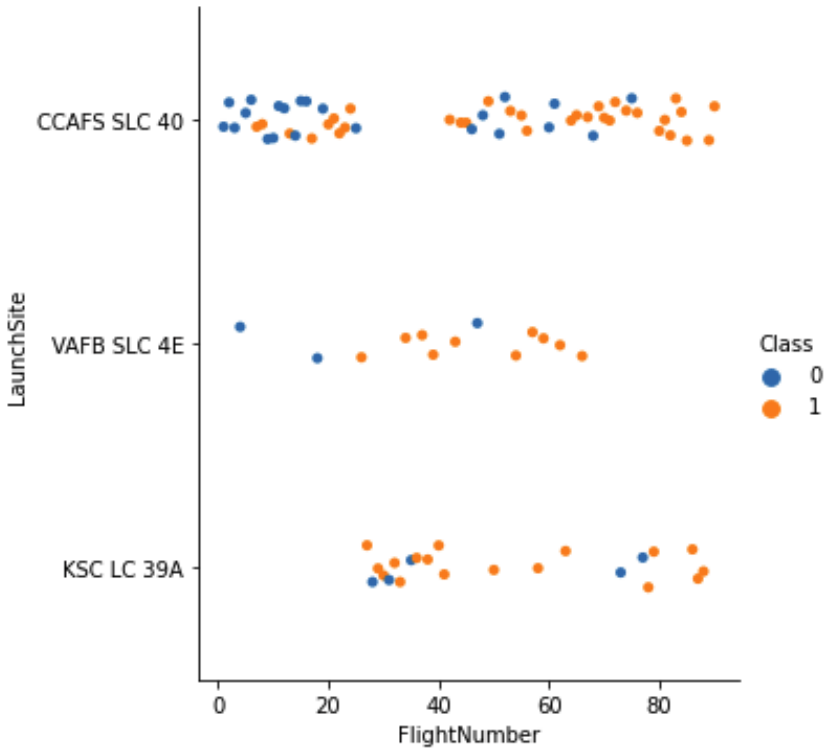
RESULTS – EDA with visualization

- Payload mass vs flight number cs success

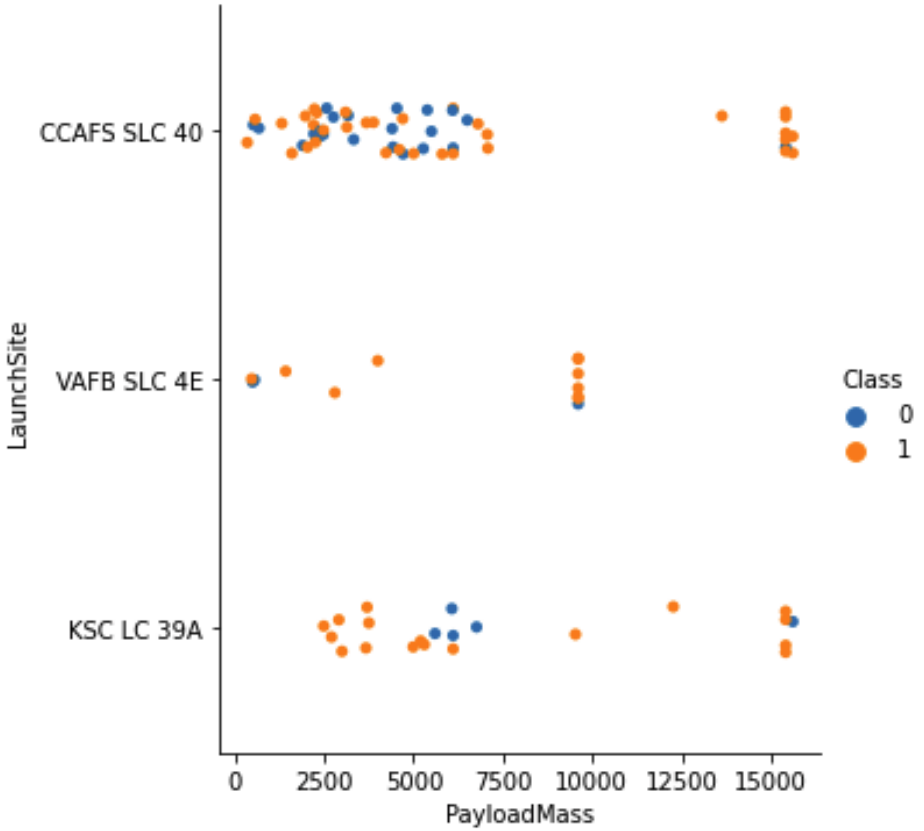


RESULTS – EDA with visualization

- Flight number vs launch site vs success

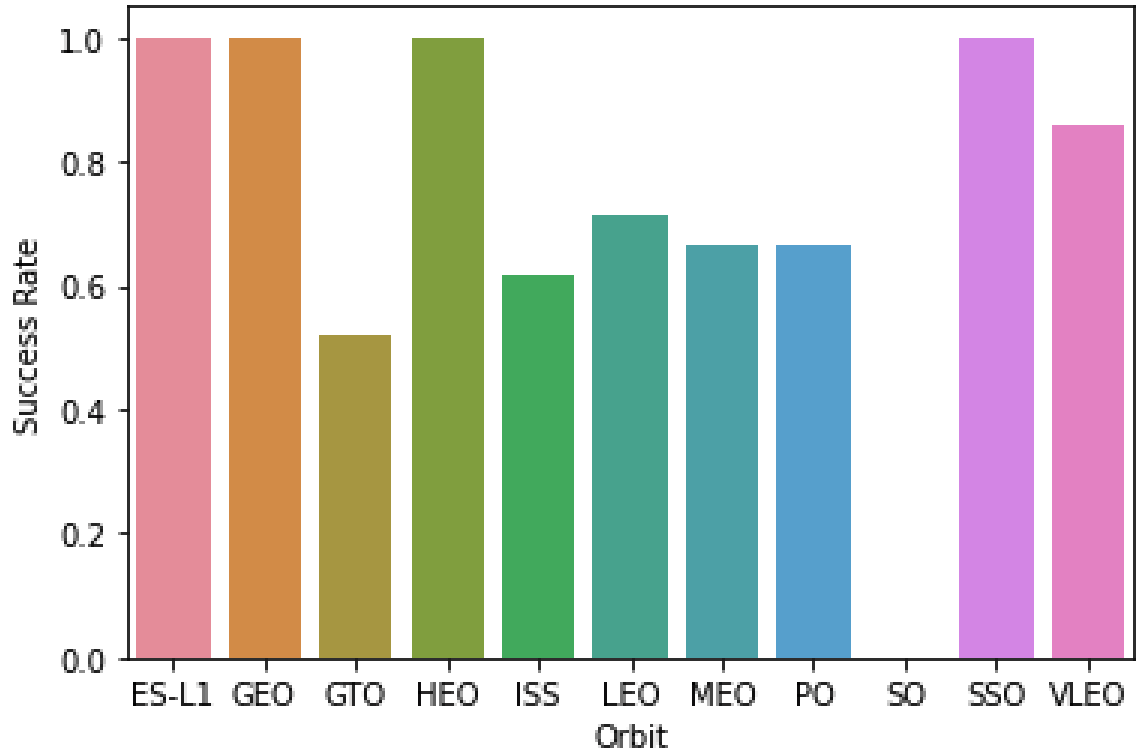


- Payload mass vs launch site vs success



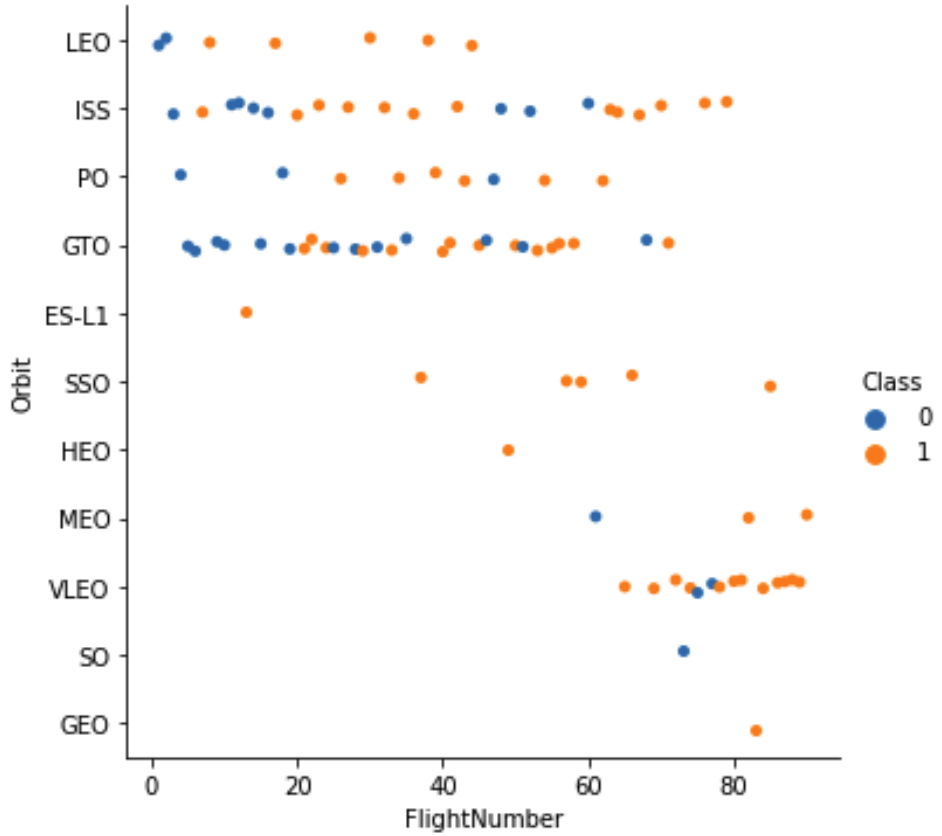
RESULTS – EDA with visualization

- Orbit type vs success rate

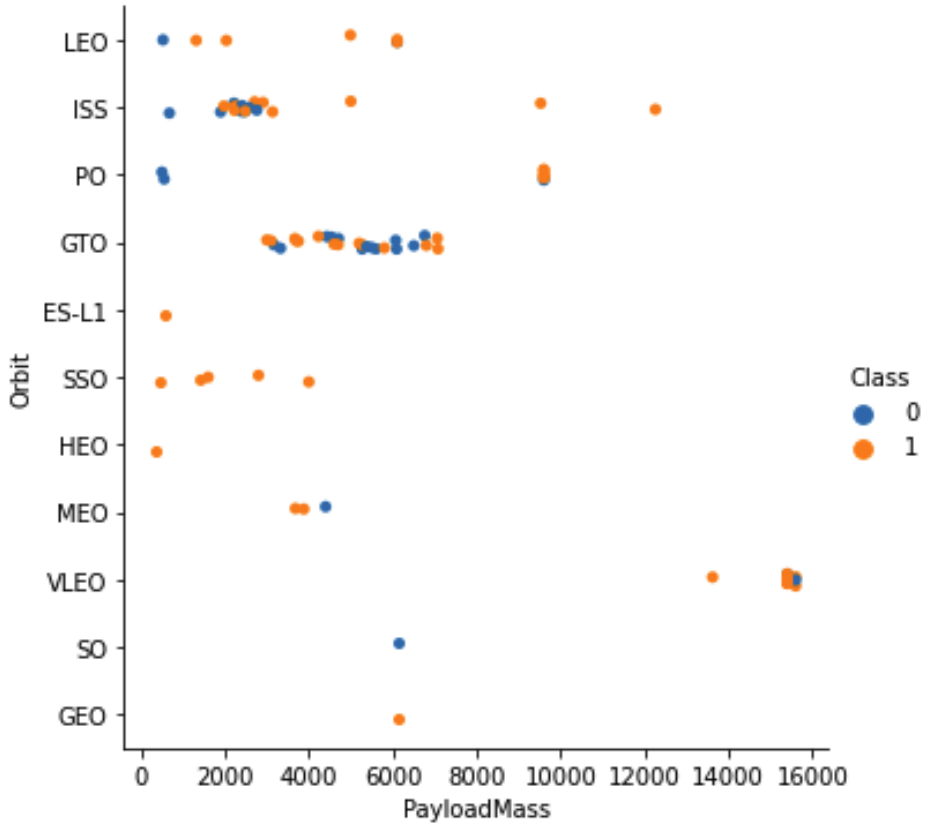


RESULTS – EDA with visualization

- Flight number vs orbit vs success

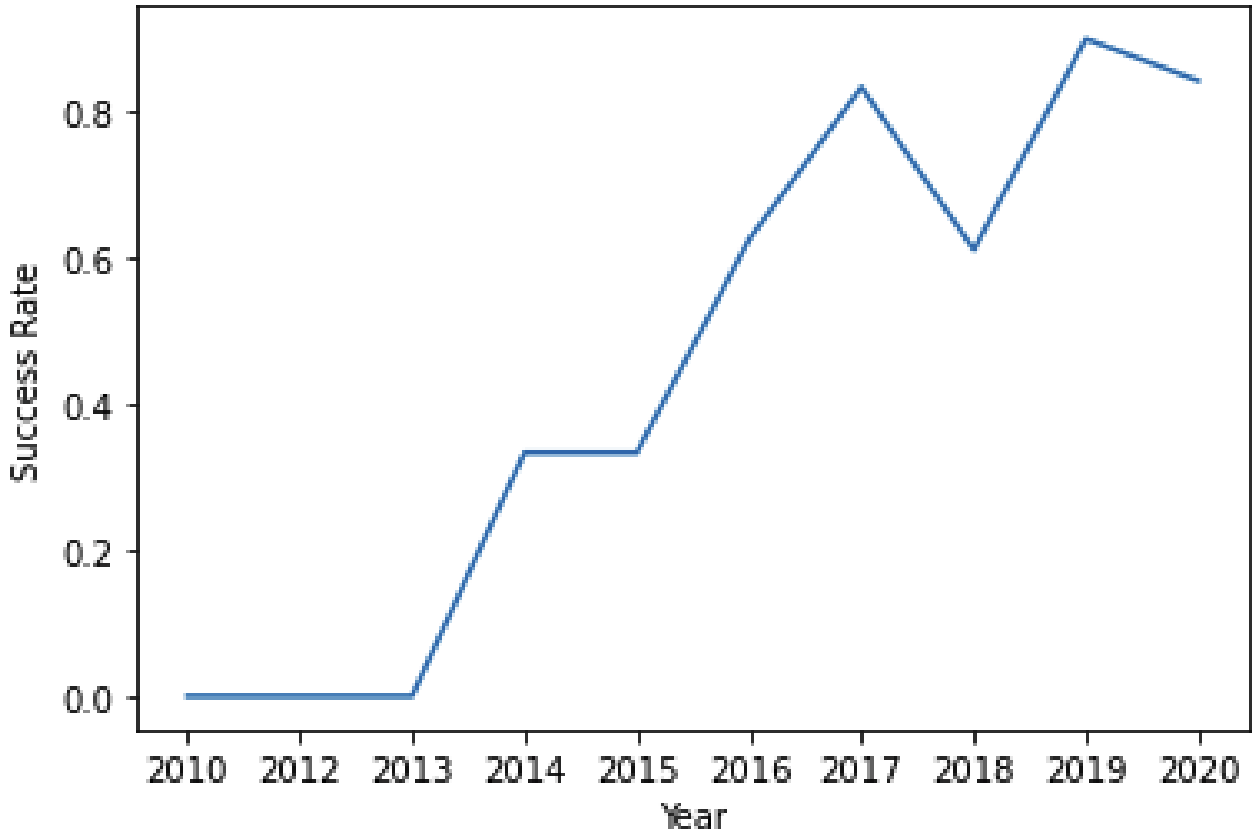


- Payload mass vs orbit vs success



RESULTS – EDA with visualization

- Success rate yearly trend



RESULTS – EDA with SQL

- Unique launch sites

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

RESULTS – EDA with SQL

- Records where launch sites begin 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

- Total payload mass carried by boosters launched by NASA (CRS)

total_payload
45596

- Average payload mass carried by booster ver F9 v1.1

average_payload
2928

- Date when first successful landing outcome in groundpad was achieved

DATE
2015-12-22

RESULTS – EDA with SQL

- Names of boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

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

- Total number of successful and failure mission outcomes

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

- Names of boosters which have carried maximum payload mass

booster_version
F9 B5 B1048.4
F9 B5 B1048.5
F9 B5 B1049.4
F9 B5 B1049.5
F9 B5 B1049.7
F9 B5 B1051.3
F9 B5 B1051.4
F9 B5 B1051.6
F9 B5 B1056.4
F9 B5 B1058.3
F9 B5 B1060.2
F9 B5 B1060.3

RESULTS – EDA with SQL

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

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

- Landing outcomes between 2010-06-04 and 2017-03-20 ranked in descending order

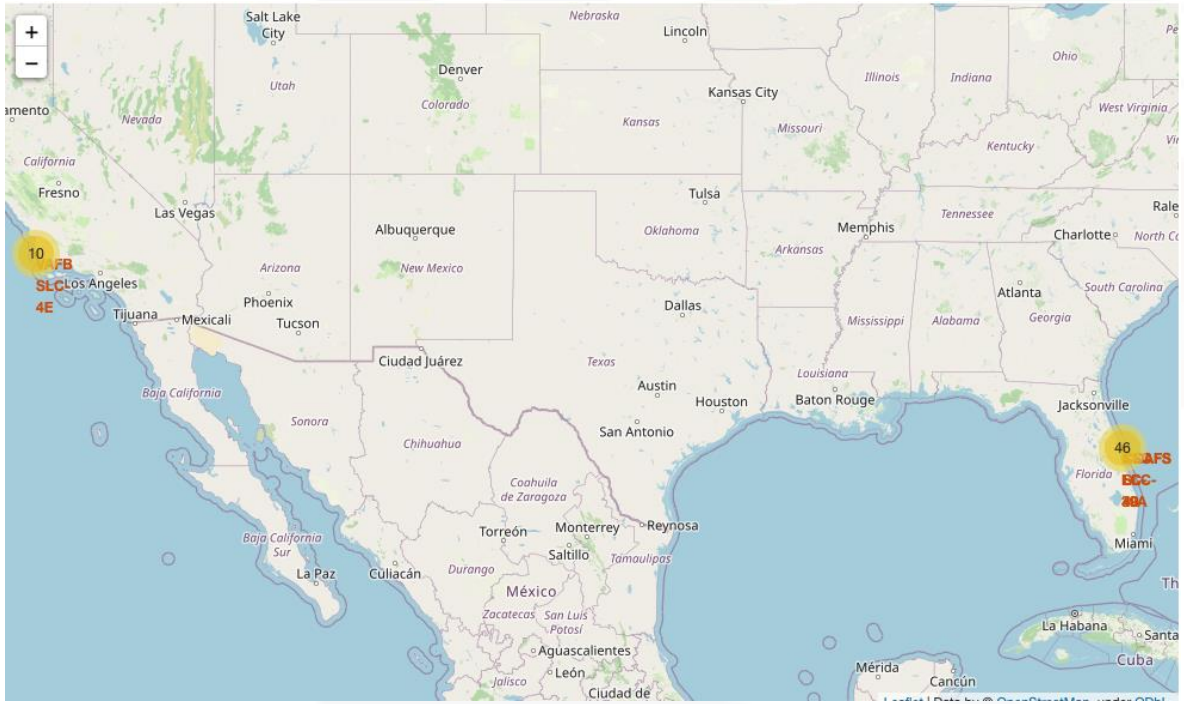
landing__outcome	landing_outcome_count
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

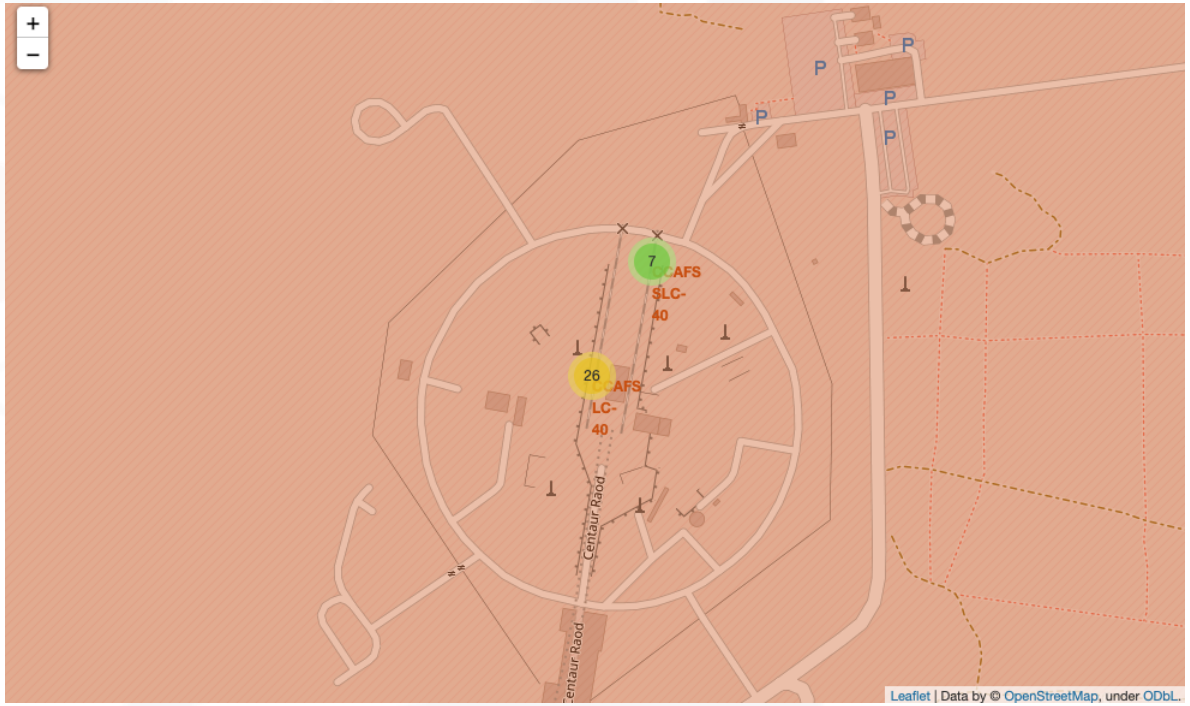
RESULTS – interactive map with Folium

- All launch sites on map

RESULTS – interactive map with Folium

- Success/failed sites on map

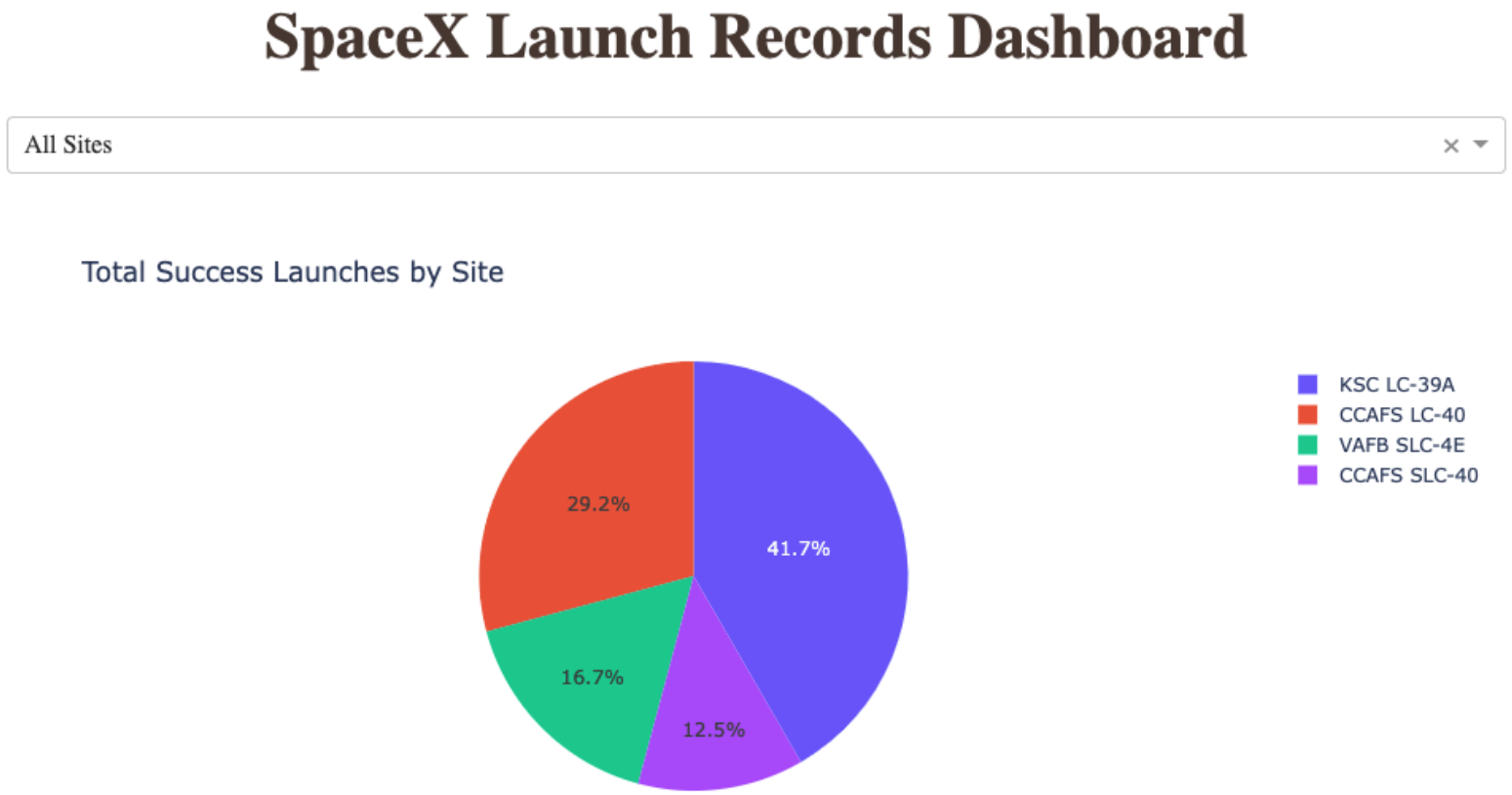




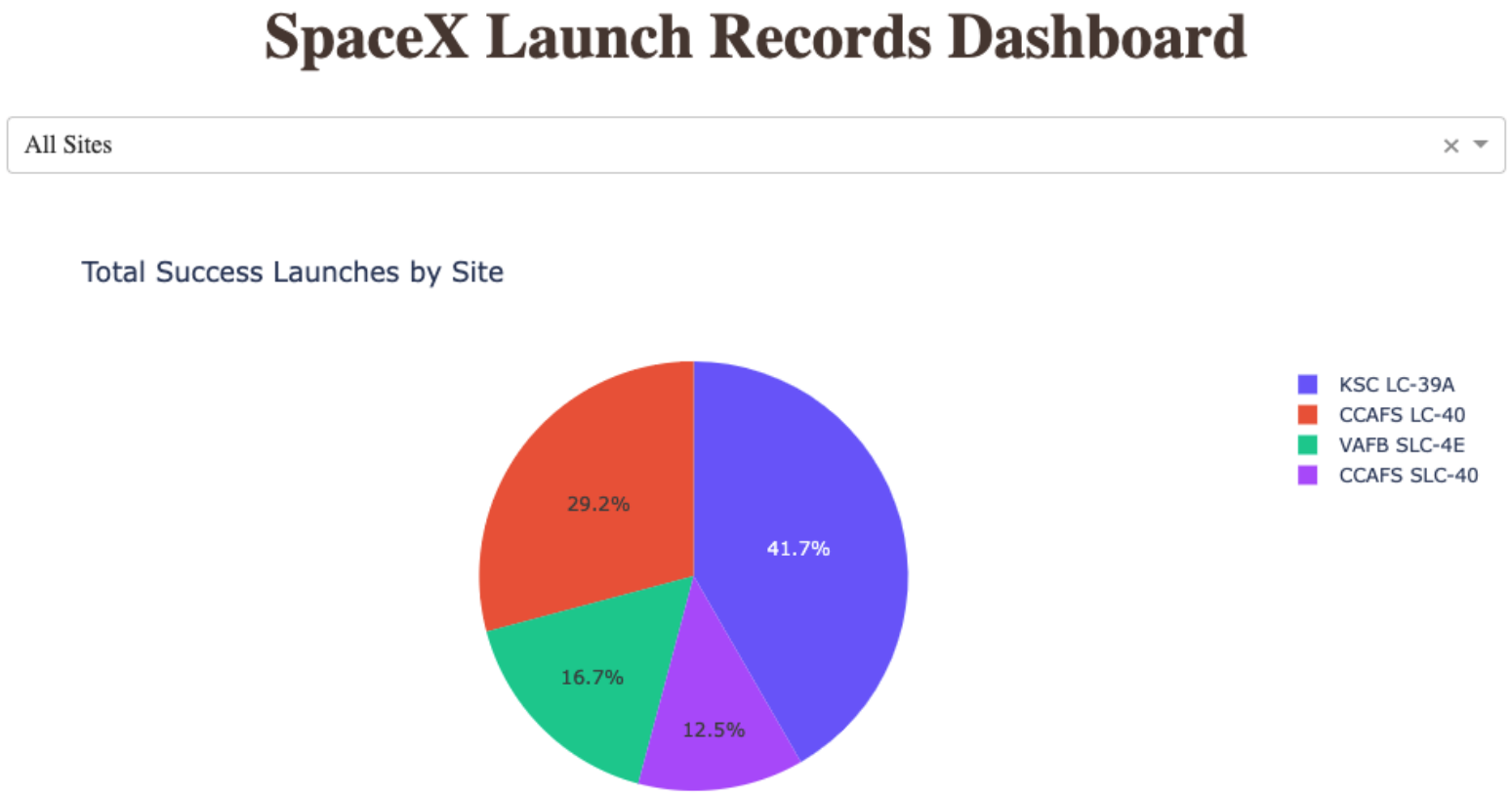
RESULTS – interactive map with Folium

- Launch site to coast line

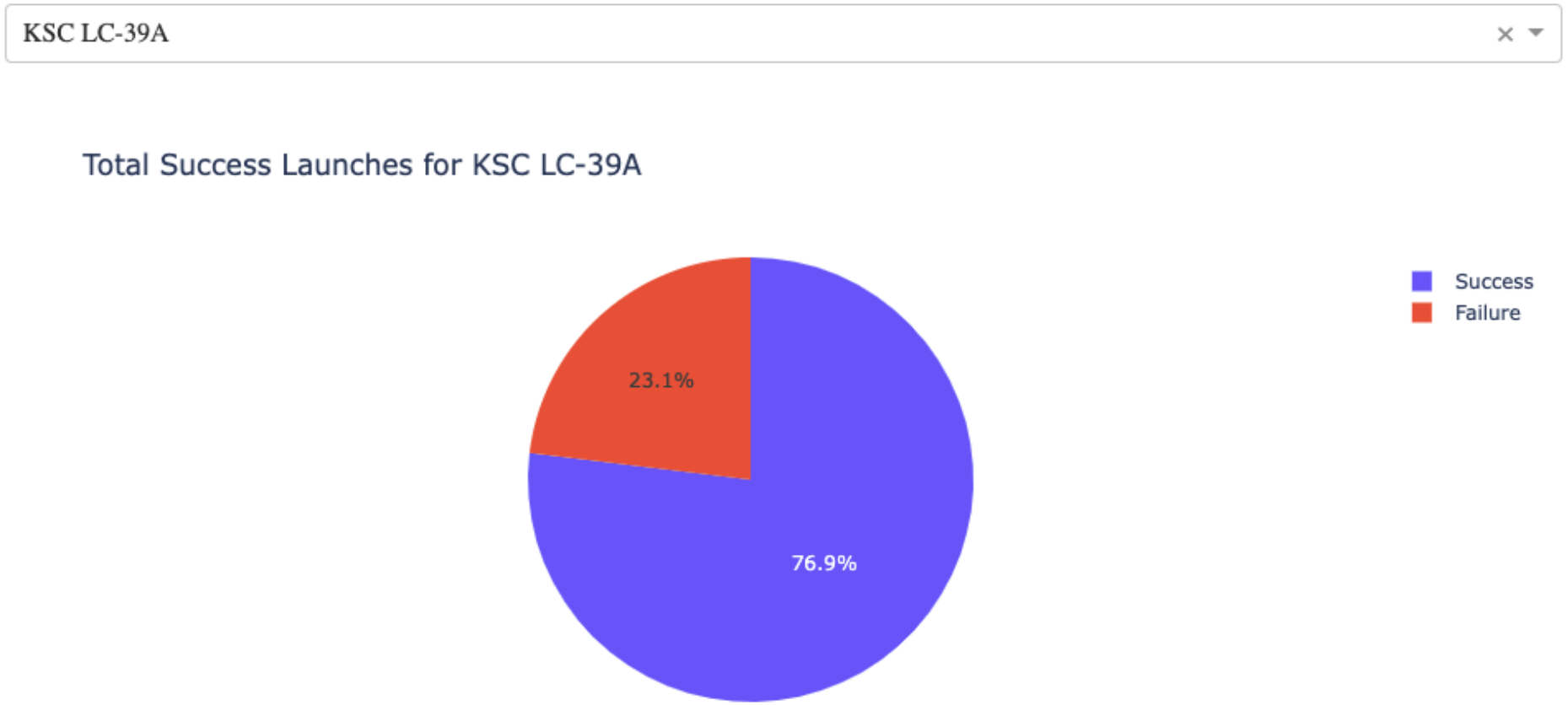
DASHBOARD – Successful launches by site



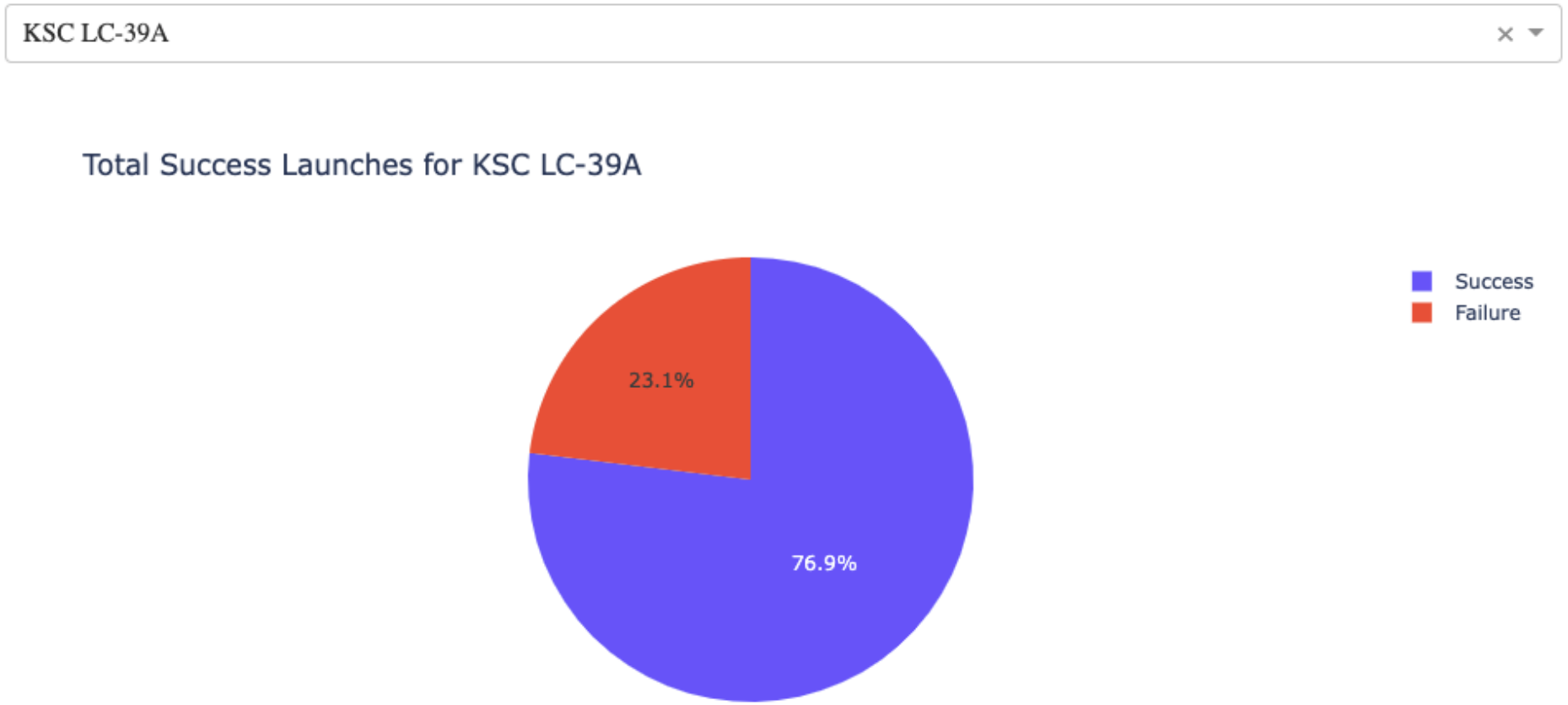
DASHBOARD – Successful launches by site



DASHBOARD – Site with highest launch rate



DASHBOARD – Site with highest launch rate



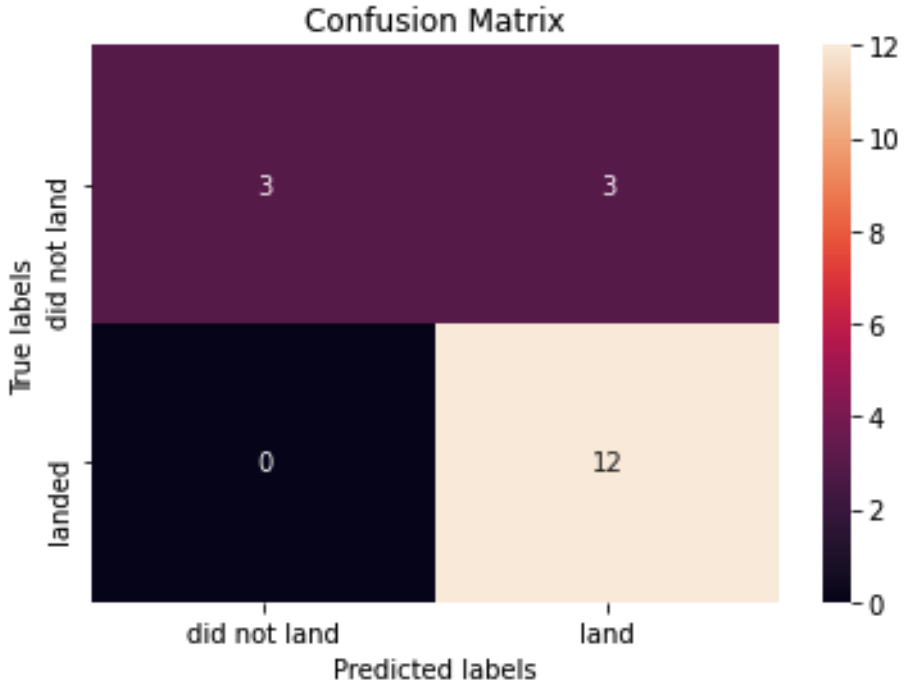
DASHBOARD – Payload mass vs success

RESULTS - Predictive analysis

- Logistic Regression:
  - Tuned hyperparameters & accuracy:

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

  - score 0.8333333333333334

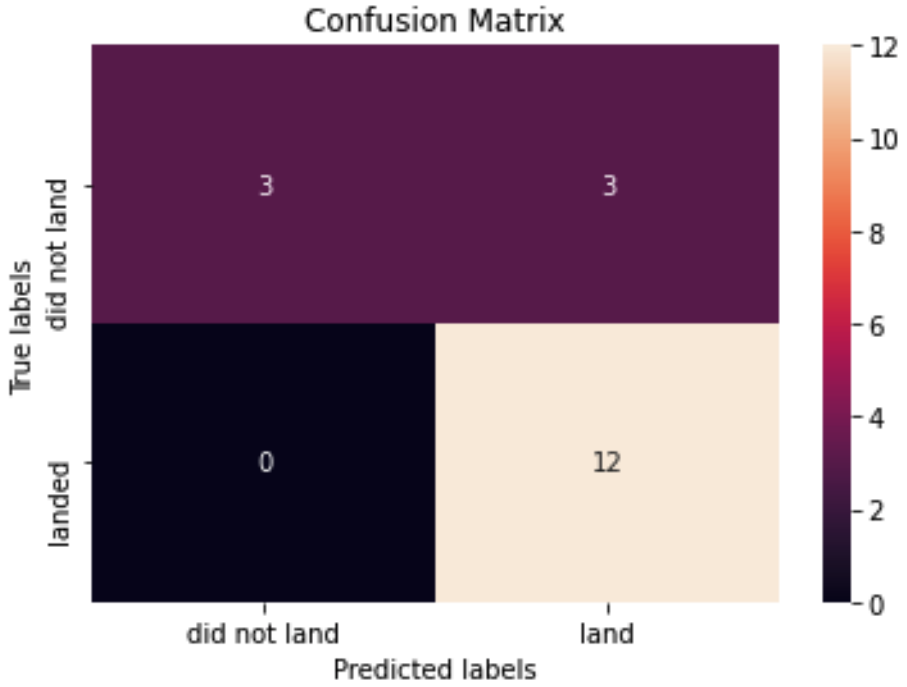


RESULTS - Predictive analysis

- Support vector machine:
  - Tuned hyperparameters & accuracy:

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

  - score `0.8333333333333334`

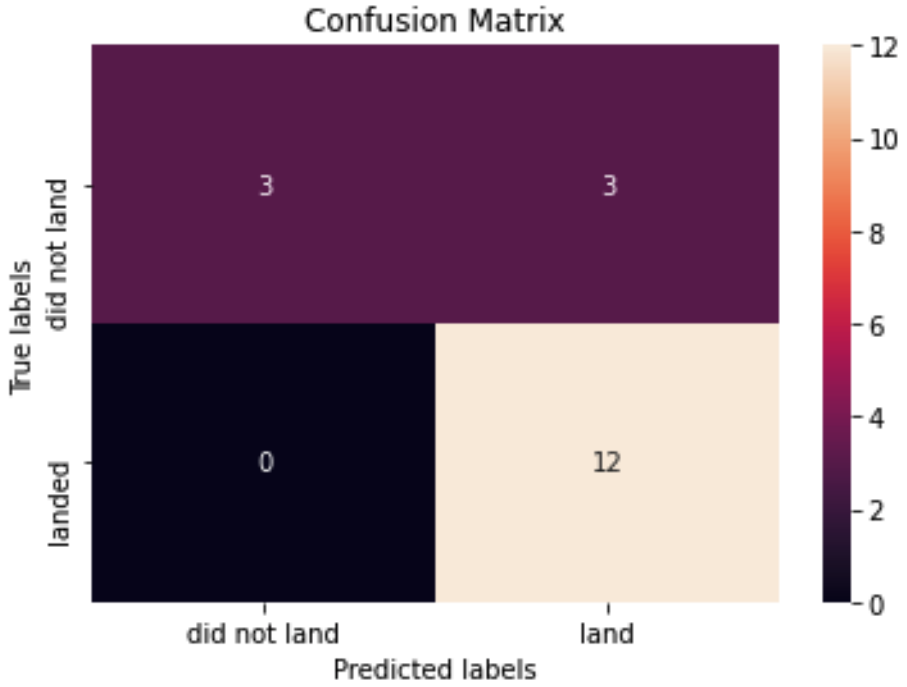


RESULTS - Predictive analysis

- Decision tree:
  - Tuned hyperparameters & accuracy:

```
tuned hpyerparameters :(best parameters)  {'criterion':
'entropy', 'max_depth': 8, 'max_features': 'auto', 'min
_samples_leaf': 1, 'min_samples_split': 10, 'splitter':
'random'}
accuracy : 0.8785714285714284
```

  - score 0.8333333333333334

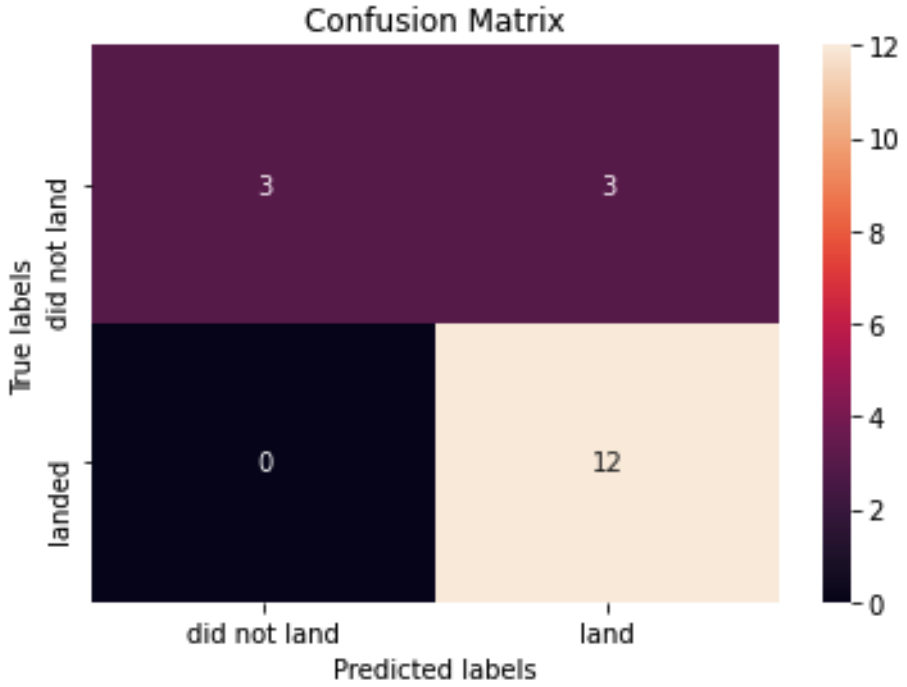


RESULTS - Predictive analysis

- K nearest neighbours:
  - Tuned hyperparameters & accuracy:

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

  - score 0.8333333333333334



CONCLUSION

- Various relationship between factors were found through EDA
- All of the machine learning models trained yielded identical results due to small data sample size. To continue gathering data to improve model.
- In the meanwhile, any of the models can be used to make predictions as accuracy were all around 85% for training data, with scores at 0.83 for test data.

APPENDIX

- Github repository: https://github.com/ryoutarouchee/IBM-Data-Science-Specialization-final-submission