SpaceX Falcon 9 First Stage Landing Prediction Model

Tricia Opdahl

July 2024

IBM Developer

SKILLS NETWORK

OUTLINE

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

EXECUTIVE SUMMARY

- SpaceX boasts low launch costs due to reusable first stage
- First stage landing success directly impacts cost
- Landing success has increased to 83%
- Analysis of historical launch data enabled predictive modeling of launch success
- Decision Tree Model predicts landing success with an accuracy of 83%

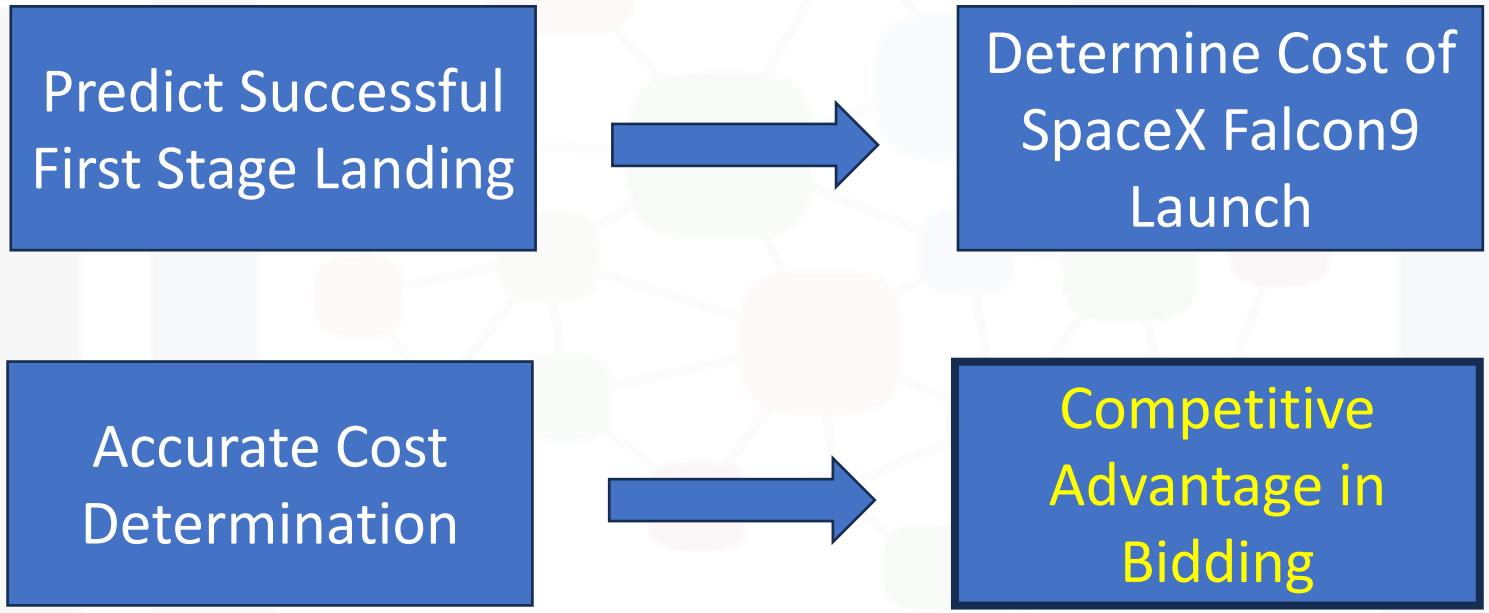
INTRODUCTION

Advertised Rocket Launch Costs

- SpaceX Falcon 9: $62 million
- Other providers: upwards of $165 million
- SpaceX engineered a reusable First Stage

SpaceX quotes $100 million less than other launch providers

Using this information



Can we predict whether the first stage will successfully land?

METHODOLOGY

- Data Collection
- Data Wrangling
- Exploratory Data Analysis
- Data Visualization
- Machine Learning Prediction

METHODOLOGY - DATA COLLECTION

- SpaceX API
  - https://api.spacexdata.com/v4/launches/past
  - Filtered for Falcon 9 data
  - 90 rows each corresponding to a flight
  - 17 columns each corresponding to features (e.g. BoosterVersion, PayloadMass, Orbit, etc)
- Wikipedia
  - https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches
  - Additional data for each flight
  - Features include Launch Outcome, Booster Landing, and Launch Site

METHODOLOGY - DATA WRANGLING

- Missing Values
  - Payload Mass: Replaced with average mass
  - Landing Pad: Retained as is to indicate no landing pad used (28.9% of data)
- Landing Outcome Label
  - Based on Outcome Column
  - 0 if ‘bad outcome’ – Any unsuccessful landing outcome regardless of location
  - 1 if ‘successful outcome’ – Successful landing to ground pad, drone ship, or ocean
- Converted categorical variables using One-Hot-Encoding

METHODOLOGY – EDA

- Loaded data into Db2 database to facilitate initial data characterization
  - Number of launch sites
  - Successful Landing Rates
  - Booster Version

- Exploratory Data visualization using Seaborn, Pandas, and Matplotlib
  - Determine correlation between various features and Landing Outcome

METHODOLOGY – DATA VISUALIZATION

- Launch Sites Locations plotted with Folium
- Interactive Dashboard using Plotly Express and Dash
  - Launch Success statistics for each Launch Site
  - Launch Success vs Payload Mass

METHODOLOGY – ML PREDICTION

- Compared 4 different ML classification models
  - Logistic Regression
  - Support Vector Machine
  - Decision Tree
  - K Nearest Neighbors

METHODOLOGY – ML PREDICTION

Features Used in Model:

- Flight number
- Payload Mass
- Orbit
- Launch Site
- Flights
- GridFins
- Reused
- Legs
- Landing Pad
- Block
- Reused Count
- Serial

METHODOLOGY – ML PREDICTION

- Normalized Data to ensure accurate feature weighting
- Reserved 20% of data for testing using split-test-train: 18 samples
- Performed Grid Search to optimize the models' hyperparameters
- Trained models using all retained features : 83 columns with one-hot-encoding
- Assessed models using Accuracy and Confusion Matrix

RESULTS – EDA with SQL

- 4 Distinct Launch Sites

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

RESULTS – EDA with SQL

- 5 records where launch sites begin with the string '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	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	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

RESULTS – EDA with SQL

Total Mass Launched by NASA (CRS)
45596

Average Mass Carried by BV F9 v1.1
2928.4

First Successful Landing Date
2018-07-22

RESULTS – EDA with SQL

- Booster Version with success in drone ship and payload mass between 4000 and 6000 kg.

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

RESULTS – EDA with SQL

Mission_Outcome	COUNT(*)
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Mission Outcomes

RESULTS – EDA with SQL

Boosters that have carried the maximum payload mass

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

Out[19]:

Month	Landing_Outcome	Booster_Version	Launch_Site
01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
02	Controlled (ocean)	F9 v1.1 B1013	CCAFS LC-40
03	No attempt	F9 v1.1 B1014	CCAFS LC-40
04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40
04	No attempt	F9 v1.1 B1016	CCAFS LC-40
06	Precluded (drone ship)	F9 v1.1 B1018	CCAFS LC-40
12	Success (ground pad)	F9 FT B1019	CCAFS LC-40

- Landing Outcomes in 2015

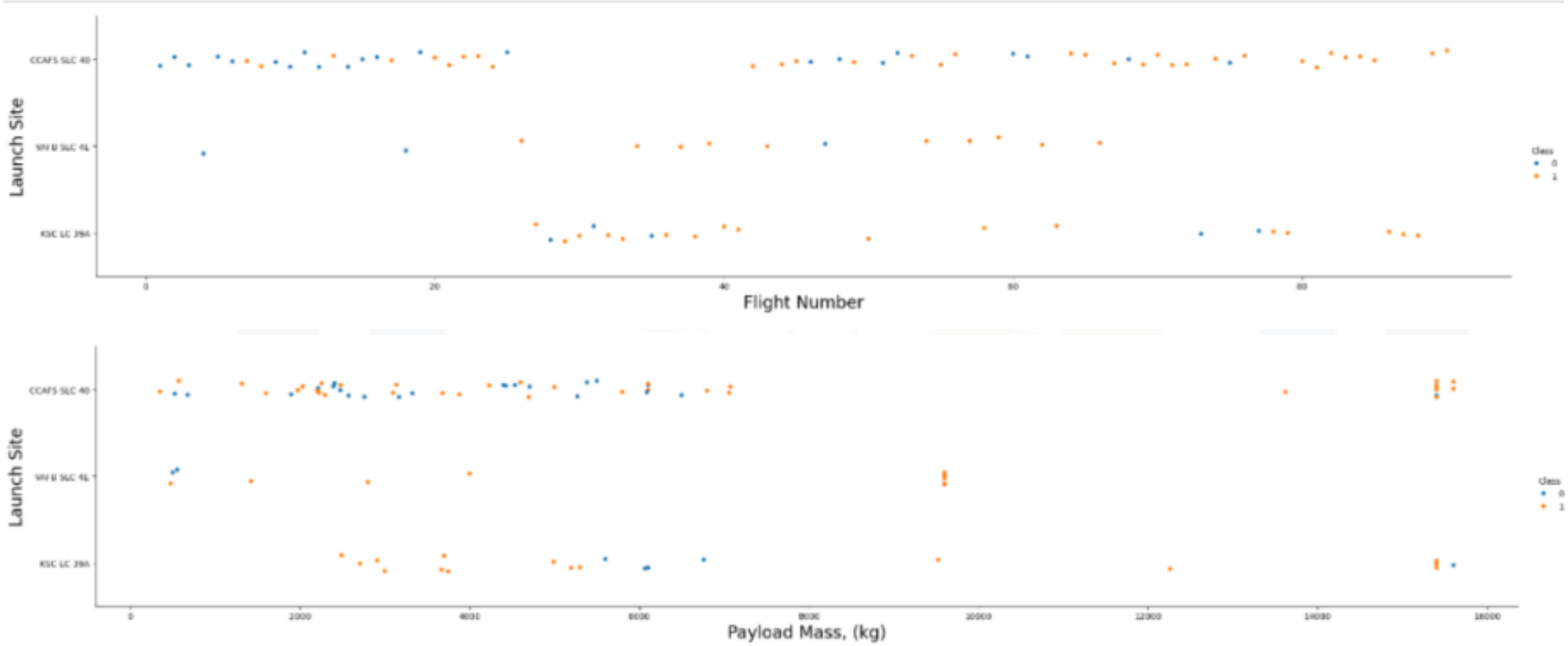
RESULTS – EDA with SQL

- Landing Outcomes from 2010-06-04 to 2017-03-20

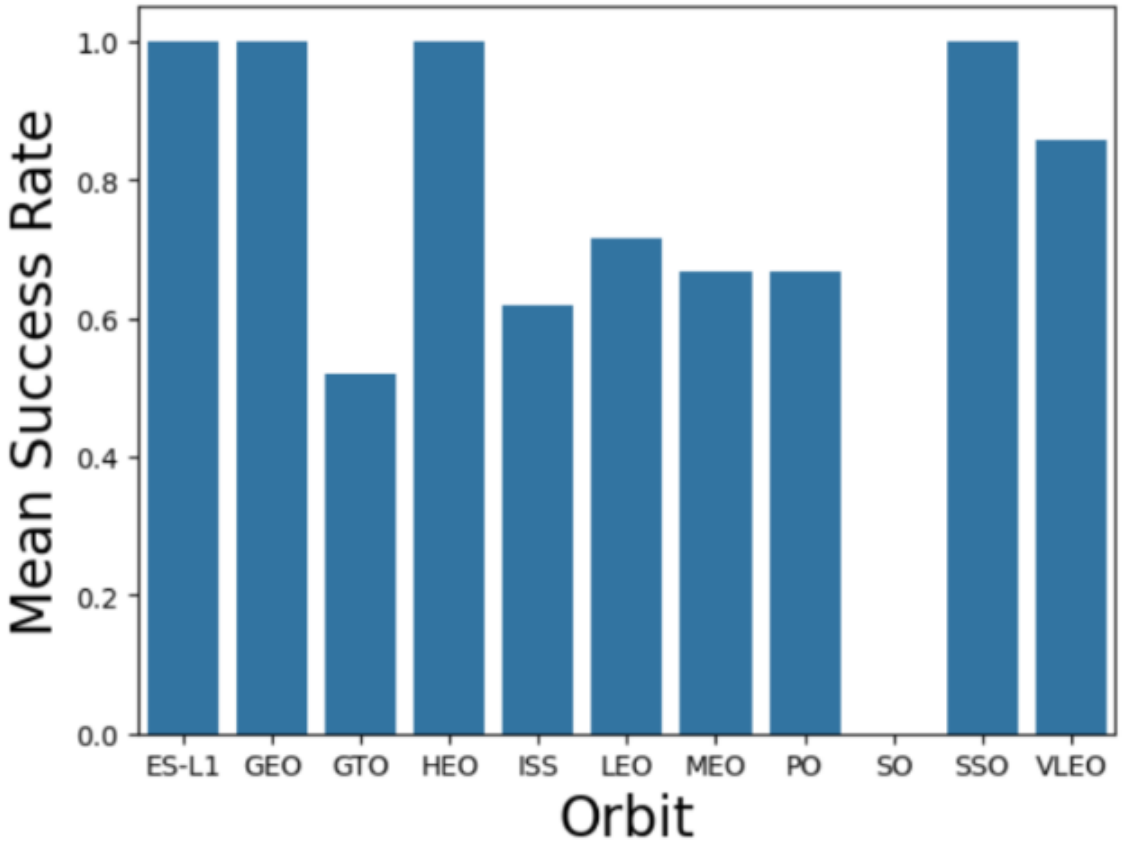
Out[24]:

Landing_Outcome	Outcome Count
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

RESULTS – EDA with Visualization

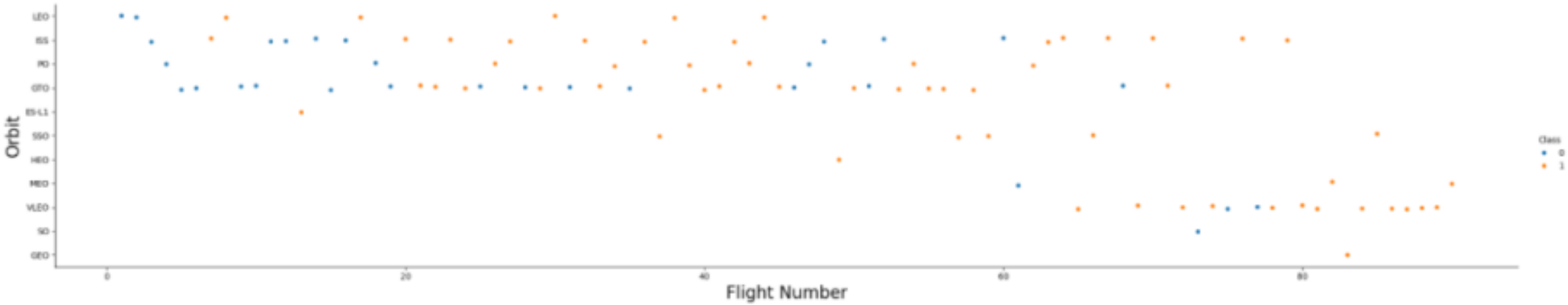


RESULTS – EDA with Visualization



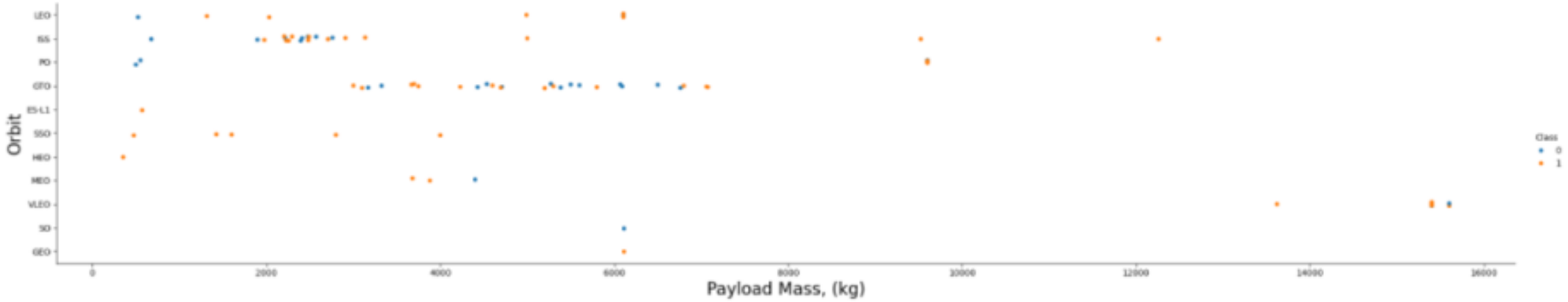
RESULTS – EDA with Visualization

- Orbit vs Flight Number



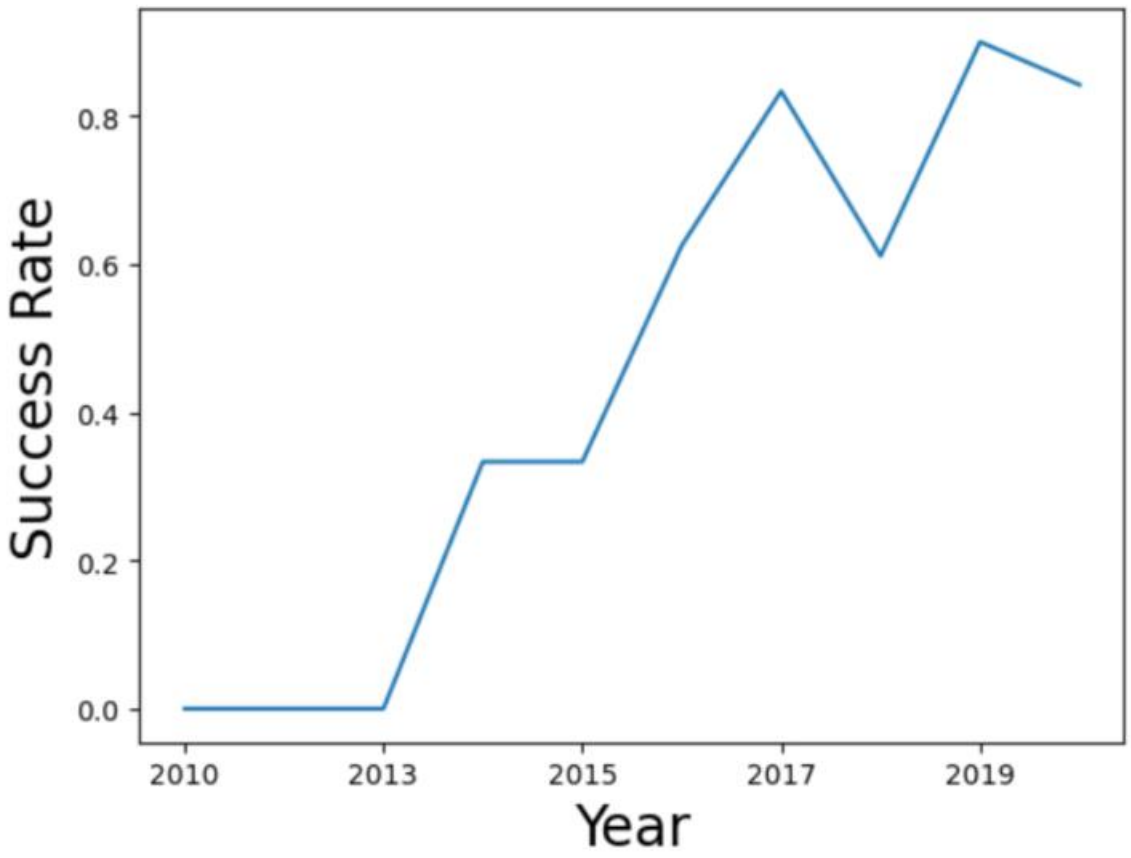
RESULTS – EDA with Visualization

- Orbit vs Payload Mass



RESULTS – EDA with Visualization

- Success Rates increased significantly since 2013



RESULTS – EDA with Visualization

- First 5 records before one-hot-encoding

[14]:

	FlightNumber	PayloadMass	Orbit	LaunchSite	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial
0	1	6104.959412	LEO	CCAFS SLC 40	1	False	False	False	NaN	1.0	0	B0003
1	2	525.000000	LEO	CCAFS SLC 40	1	False	False	False	NaN	1.0	0	B0005
2	3	677.000000	ISS	CCAFS SLC 40	1	False	False	False	NaN	1.0	0	B0007
3	4	500.000000	PO	VAFB SLC 4E	1	False	False	False	NaN	1.0	0	B1003
4	5	3170.000000	GTO	CCAFS SLC 40	1	False	False	False	NaN	1.0	0	B1004

RESULTS – EDA with Visualization

- First 5 records after one-hot-encoding

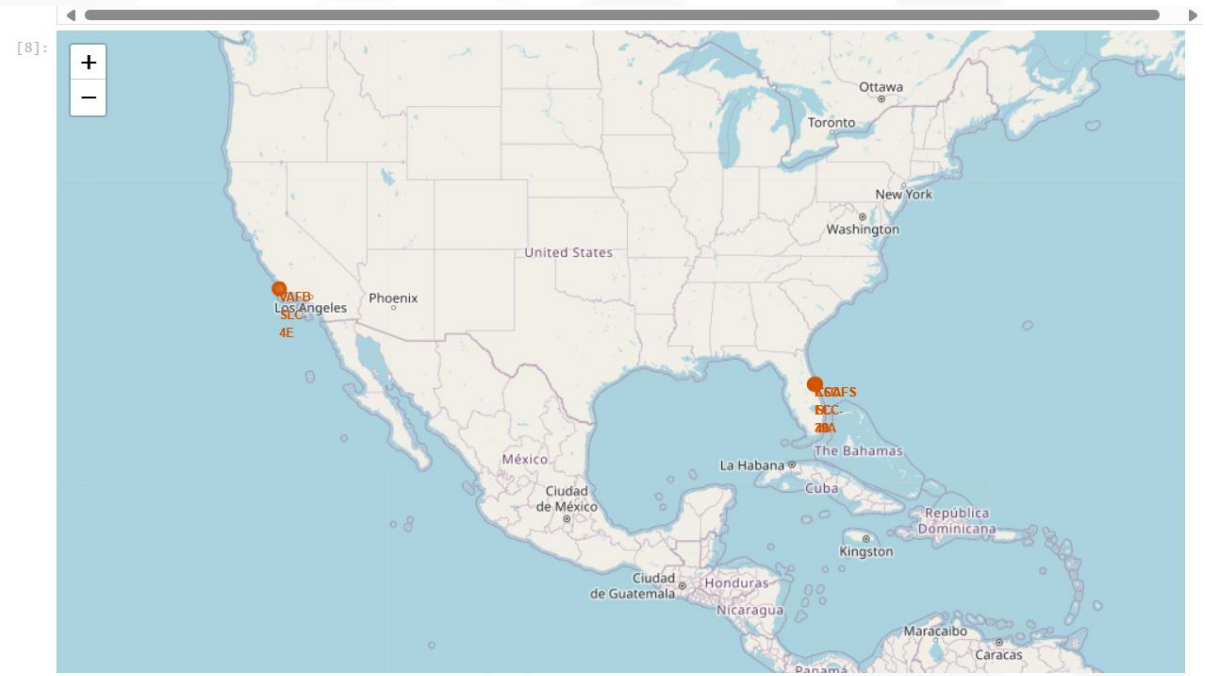
[16]:

	FlightNumber	PayloadMass	Flights	GridFins	Reused	Legs	Block	ReusedCount	Orbit_ES-L1	Orbit_GEO	...	Serial_B1048	Serial_B1049	Serial_I
0	1.0	6104.959412	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	...	0.0	0.0	
1	2.0	525.000000	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	...	0.0	0.0	
2	3.0	677.000000	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	...	0.0	0.0	
3	4.0	500.000000	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	...	0.0	0.0	
4	5.0	3170.000000	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	...	0.0	0.0	

5 rows × 80 columns

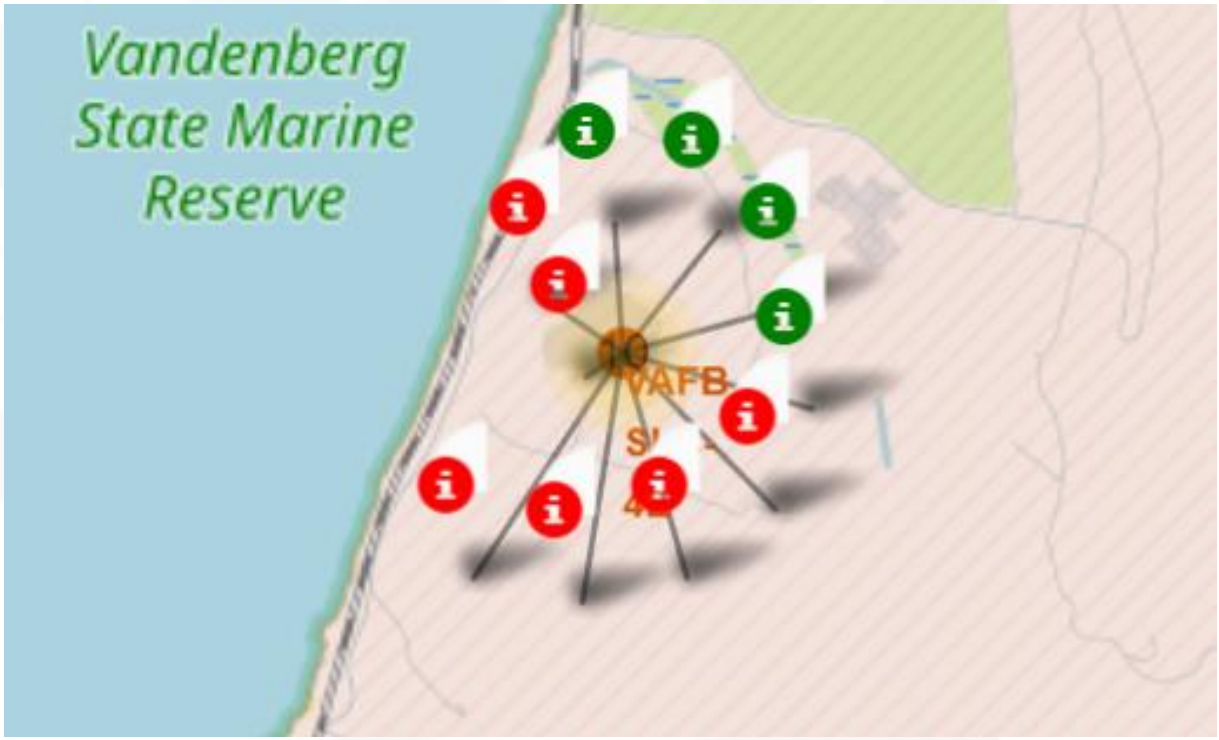
RESULTS – Interactive Map with Folium

- 4 Launch Sites:
  - CCAFS LC-40
  - VAFB SLC-4E
  - KSC LC-39A
  - CCAFS SLC-40
- 1 in California
- 3 in Florida



RESULTS – Interactive Map with Folium

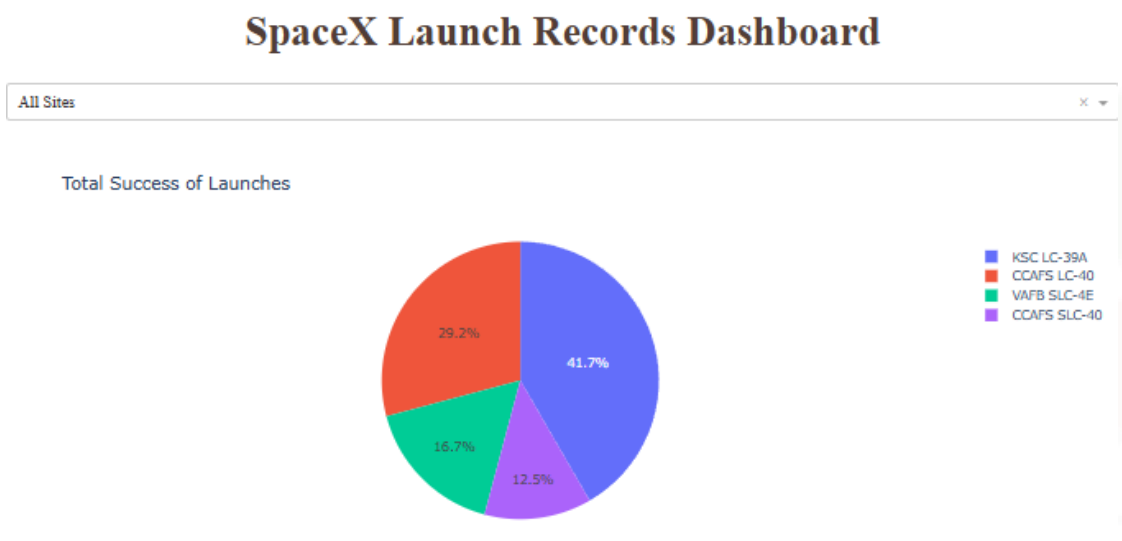
- Successful and Failed outcomes at VAFB

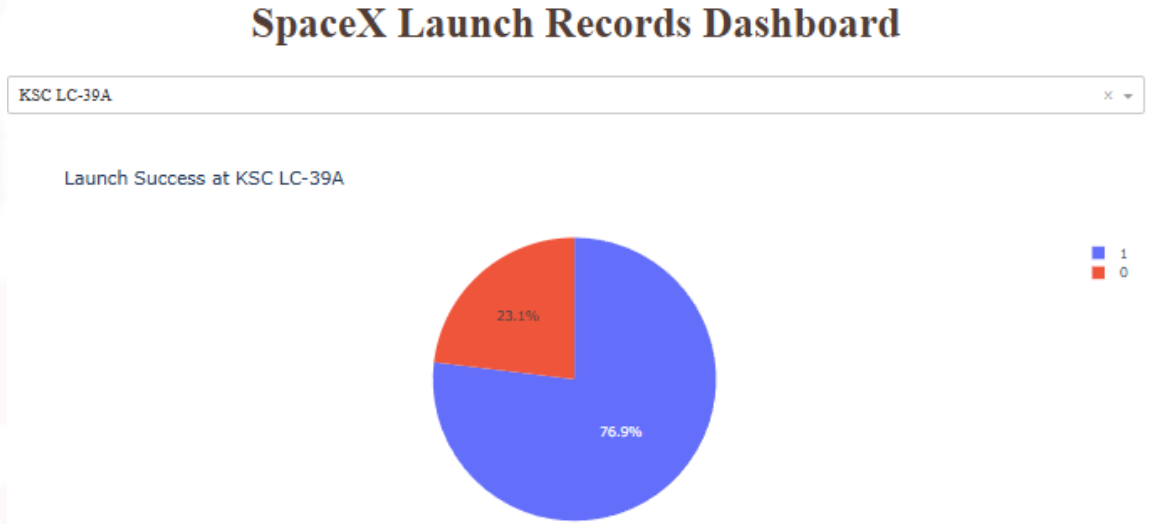


RESULTS – Interactive Map with Folium

- All sites near coastline
- All sites near major modes of transportation

DASHBOARD - Landing Success per Site



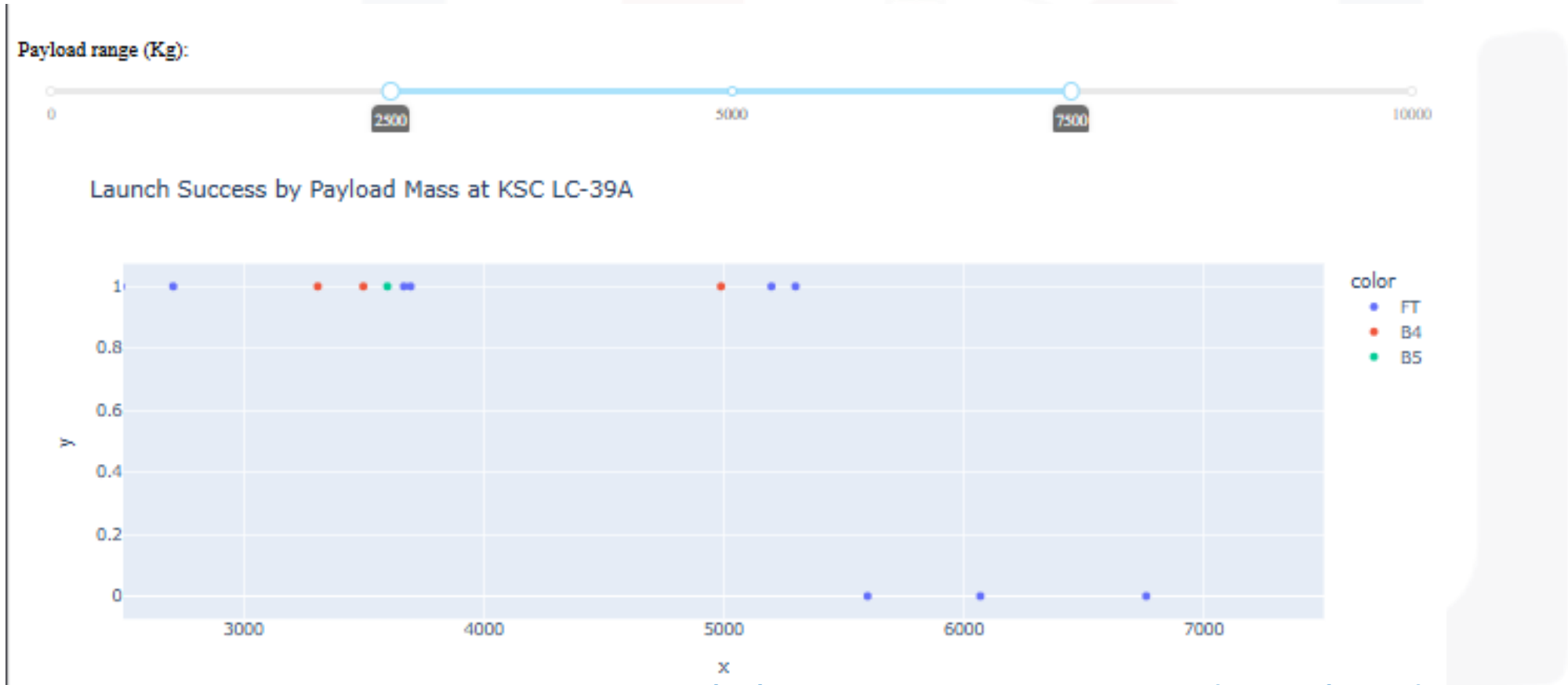


- Success Rates vary across the 4 launch sites

DASHBOARD – Landing Success vs PL Mass, Booster Version, All Sites

- Success Rates vary with booster version and payload mass

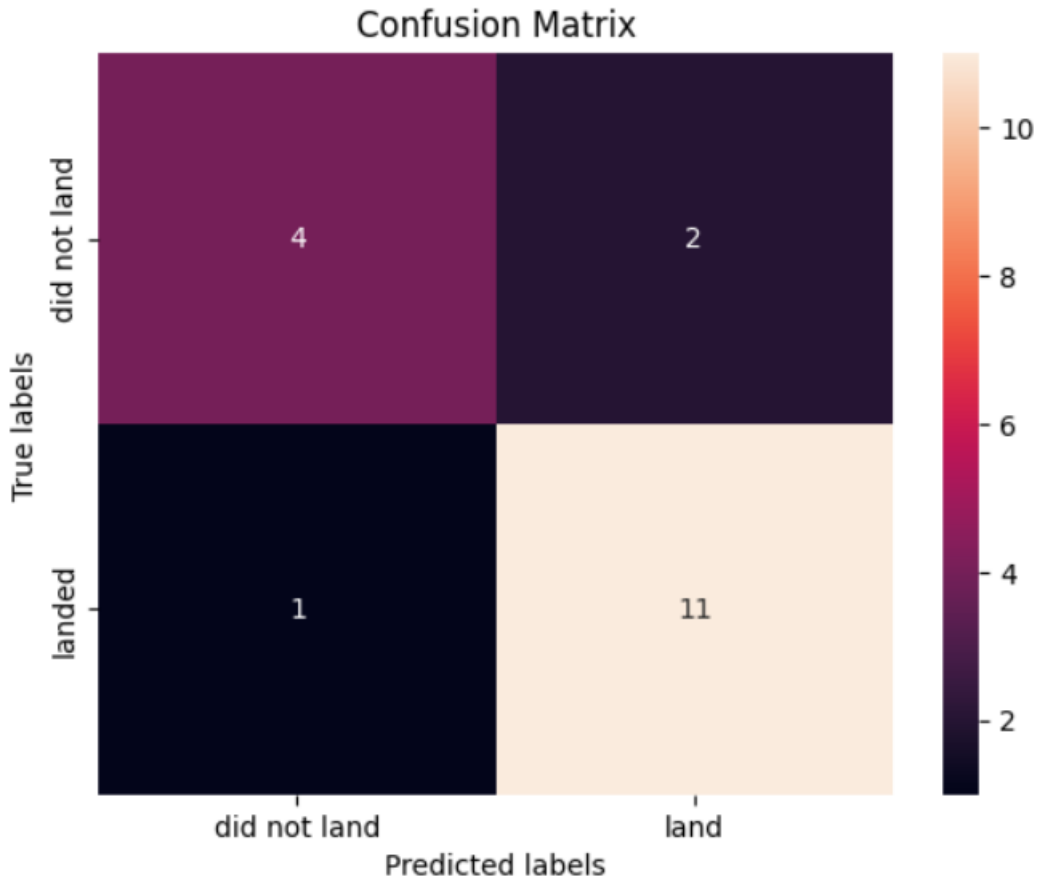
DASHBOARD Landing Success vs PL Mass, Booster Version, One Site



- Success Rates vary with booster version and payload mass

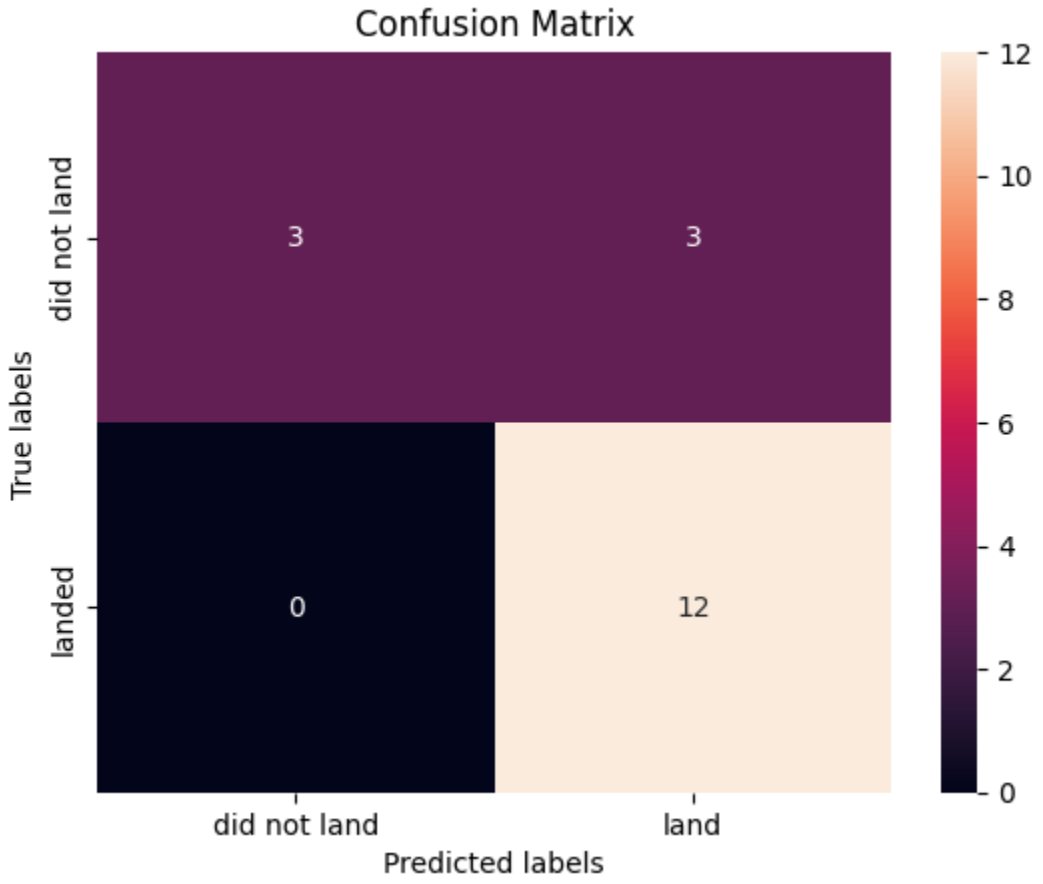
RESULTS – Decision Tree

- All 4 classification models produced an accuracy of 83.3% on test data
- **Decision Tree produced lowest false positive predictions**



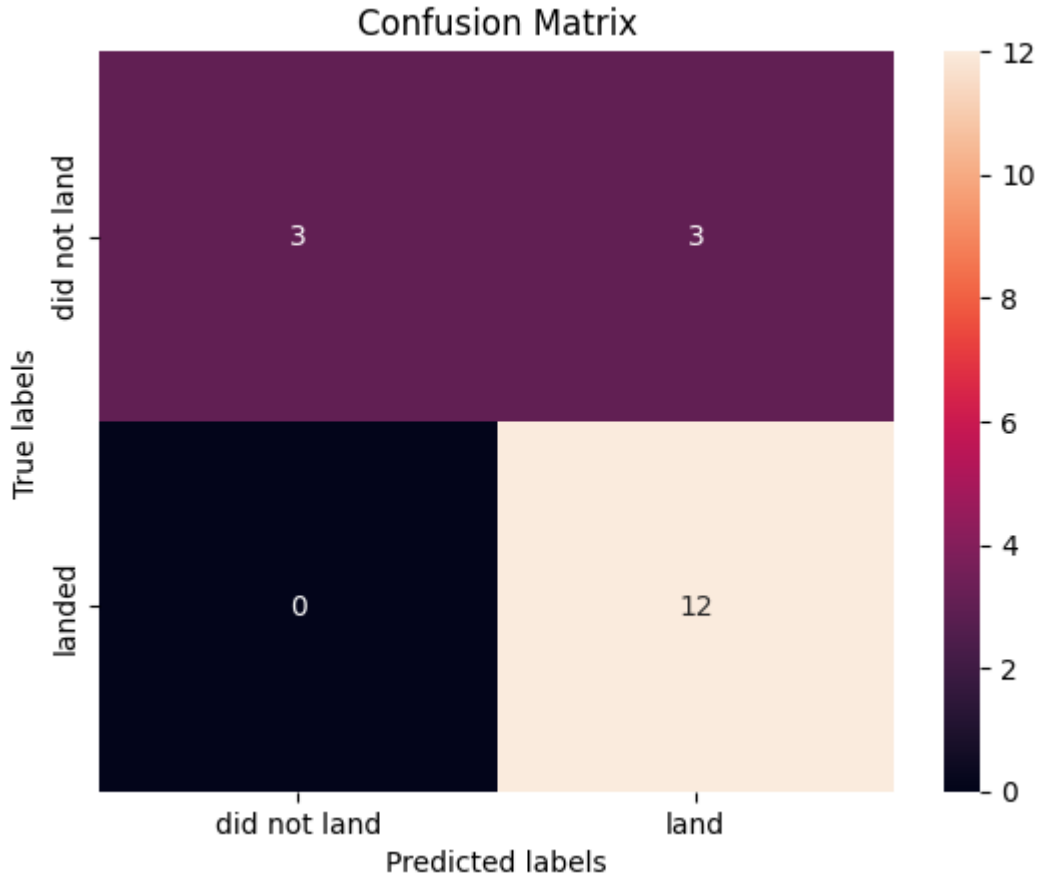
Results – ML Model Data

- Logistic Regression
  - Accuracy on train data: 84.6%
  - Accuracy on test data: 83.3%
  - tuned hyerparameters :(best parameters) {'C': 0.01, 'penalty': 'l2', 'solver': 'lbfgs'}



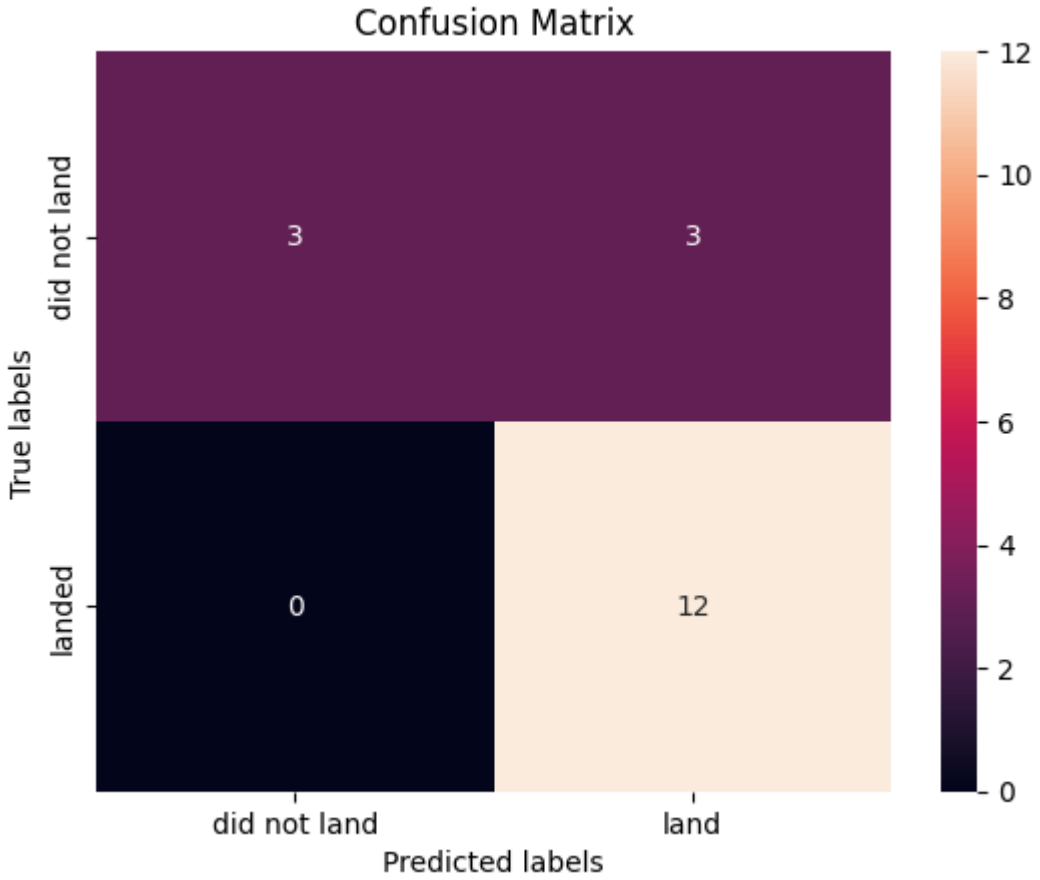
Results – ML Model Data

- Support Vector Machine
  - Accuracy on train data: 84.8%
  - Accuracy on test data: 83.3%
  - tuned hyerparameters :(best parameters) {'C': 1.0, 'gamma': 0.0316227660168379, 'kernel': 'sigmoid'}



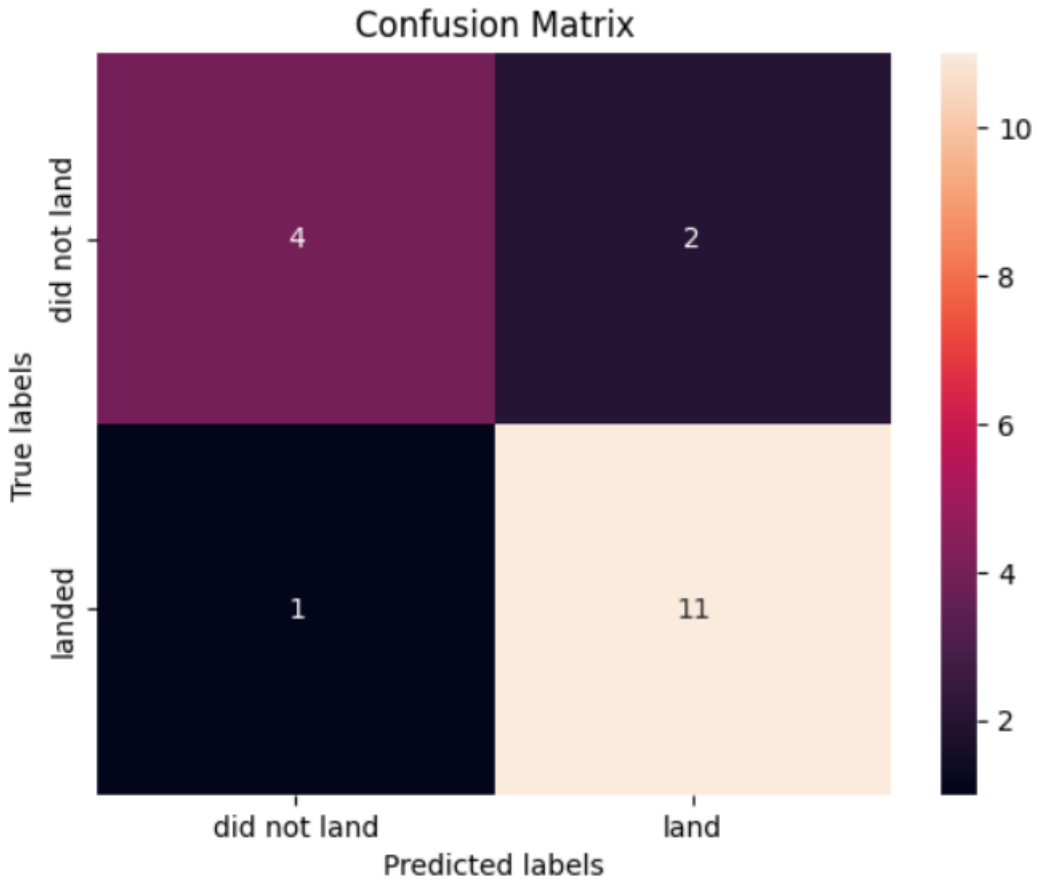
Results – ML Model Data

- K Nearest Neighbors
  - Accuracy on train data: 84.8%
  - Accuracy on test data: 83.3%
  - tuned hyerparameters: (best parameters) {'algorithm': 'auto', 'n_neighbors': 10, 'p': 1}



Results – ML Model Data

- Decision Tree
  - Accuracy on train data: 87.5%
  - Accuracy on test data: 83.3%
  - tuned hyperparameters: (best parameters) {'criterion': 'gini', 'max_depth': 10, 'max_features': 'sqrt', 'min_samples_leaf': 1, 'min_samples_split': 10, 'splitter': 'random'}



DISCUSSION

- Can we predict whether the first stage will successfully land?
- ML model trained with historical landing data.
- ML model utilizes 12 features.
- Decision Tree Model predicts first stage landing success with 83.3% accuracy.
- Further data collection necessary to maintain model as SpaceX technology improves over time.

OVERALL FINDINGS & IMPLICATIONS

Findings

- SpaceX Falcon 9 landing success above 80%
- SpaceX landing success increases over time
- Model predicts landing success with an accuracy of 83%

Implications

- SpaceX Falcon 9 launch costs likely to remain low
- Continued data collection and analysis necessary
- Can predict actual cost of SpaceX Falcon 9 launch

CONCLUSION

- SpaceX continues to advertise low launch costs
- First stage landing success is high, but not flawless
- We can predict landing success with high accuracy
- We can determine actual launch cost with high accuracy

APPENDIX - DASHBOARD

https://github.com/topdahl/Capstone_Project/blob/64e938a5605c1112c366694bcad0ae7ad889b14d/spacex_dash_app.py