IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Dominika A. Kalkowska
July 16, 2024

Outline

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

Executive Summary

Research objective:

- ➢ Identify the factors for a successful rocket launch

Used methodologies:

- ✓ Data collection using publicly available data
- ✓ Data processing to extract valuable information
- ✓ Exploratory data analysis
- ✓ Interactive data analysis
- ✓ First stage landing prediction using machine learning

Results:

- ✓ Collected valuable data from public sources
- ✓ Identified features for the best success prediction
- ✓ Identified the best model to predict which characteristics are important

Introduction

Background

- SpaceX is the most successful company of the commercial space age, making affordable space travel a reality.
- SpaceX owes much of its savings' success to reusability of the Falcon 9 rockets during the first stage of the launch.
- Using public data and machine learning models we can predict if the first stage will land successfully, and therefore determine the cost of the launch.

Questions

- How payload mass, launch site, number of flights, and orbits affect the success of the first stage landing?
- Does the rate of successful landings increase over the years?
- What is the best predictive model that can be used for binary classification in this case?

Section 1

Methodology

Methodology

Executive Summary

- **Data collection** using SpaceX REST API and web scraping
- **Data wrangling** by filtering the data, handling missing values and applying one hot encoding to prepare the data for analysis and modeling
- **Exploratory data analysis (EDA)** using visualization and SQL
- **Interactive visual analytics** using Folium and Plotly Dash
- **Predictive analysis** using classification models – tuning and evaluating models to find best model and parameters

Data Collection

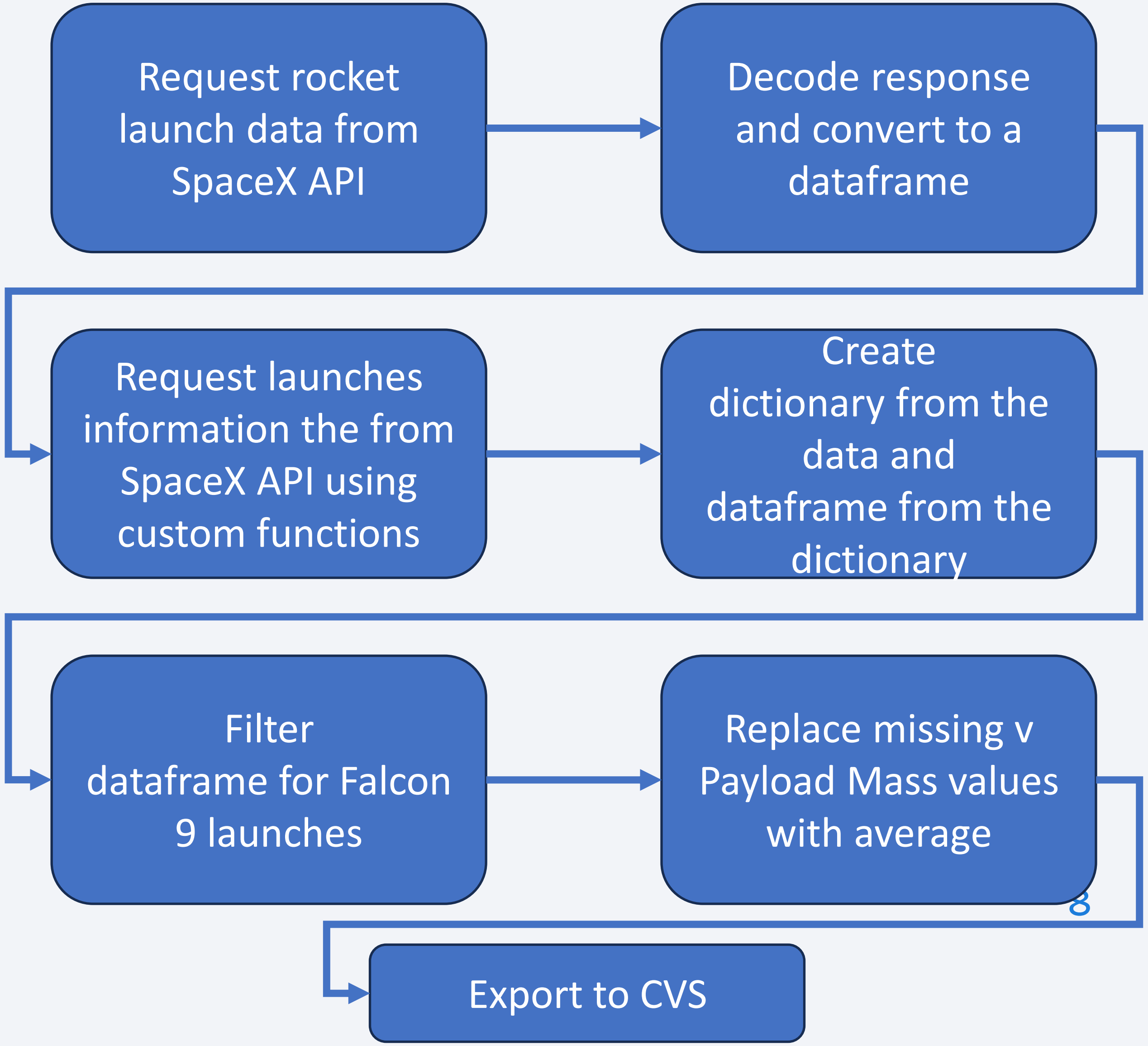
Data sets were collected from

- Space X API (https://api.spacexdata.com/v4/rockets/)
- Wikipedia (https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches)

Data Collection – SpaceX API

- The data was obtained according to the following procedure

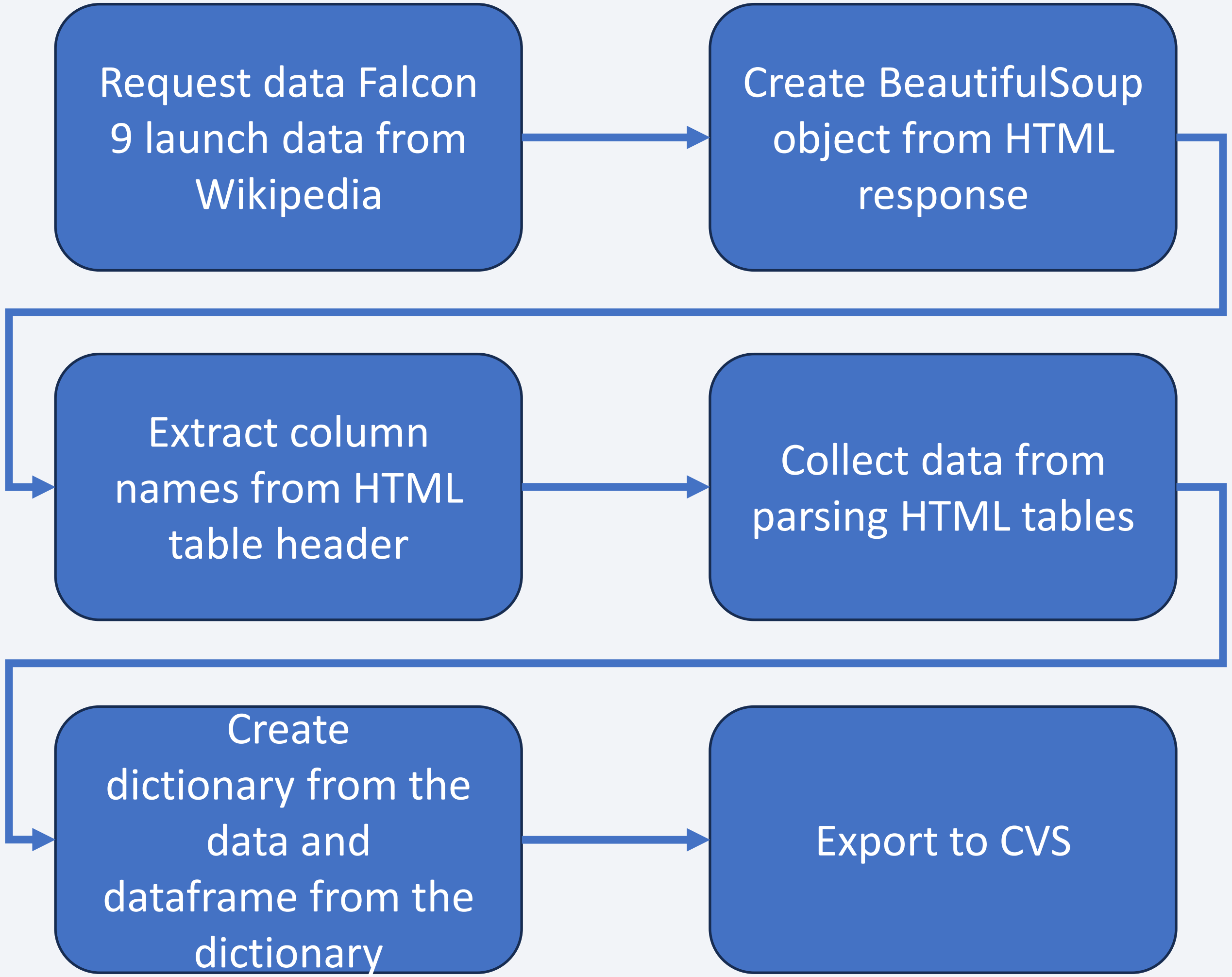
[GitHub URL of the completed SpaceX API calls notebook]



Data Collection - Scraping

- The data was obtained according to the following procedure

[GitHub URL of the completed web scraping notebook]



Data Wrangling

- Steps
  - Perform EDA and determine data labels
  - Calculate:
    - number of launches on each site
    - number and occurrence of orbit
    - number and occurrence of mission outcome per orbit type
  - Create binary landing outcome label (1=success, 0=failure)

GitHub URL of the completed data wrangling notebook

EDA with Data Visualization

Chart

- Flight Number vs. Payload Mass (kg)
- Flight Number vs. Launch Site
- Flight number vs. Orbit
- Payload Mass (kg) vs. Launch Site
- Payload Mass (kg) vs. Orbit
- Orbit vs Success rate

Chart types used

- Scatter plots - show relationship between variables (if relationship exists, could be used in machine learning)
- Bar charts - show comparisons among discrete categories (relationship between the specific categories and measured value)
- Line charts - show trends in data over time

GitHub URL of the completed EDA with data visualization notebook

EDA with SQL

Display:

- Names of the unique launch sites in the space mission
- 5 records where launch sites begin with the string 'CCA’
- Total payload mass carried by boosters launched by NASA (CRS)
- Average payload mass carried by booster version F9 v1.1;

List:

- Date when the first successful landing outcome in ground pad was achieved
- Names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- Total number of successful and failure mission outcomes
- Names of the booster versions which have carried the maximum payload mass
- Failure landing outcomes in drone ship, booster versions, launch site for the months in year 2015
- Count of landing outcomes between June 4, 2010 and March 20, 2017

GitHub URL of the completed EDA with SQL notebook

Build an Interactive Map with Folium

Added markers indicating launch sites

- Marker with Circle, Popup Label and Text Label of NASA Johnson Space Center using latitude and longitude coordinates
- Markers with Circle, Popup Label and Text Label of all Launch Sites using latitude and longitude coordinates

Added colored markers of launch outcomes

- Marker Cluster of success (green) and failed (red) launches at each launch site to show which launch sites have high success rates

Added distances between a launch site to proximities

- Lines indication distance between launch site CCAFS SLC 40 and its proximity to the nearest coastline, railway, highway, and city

[GitHub URL of the completed interactive map with Folium map]

Build a Dashboard with Plotly Dash

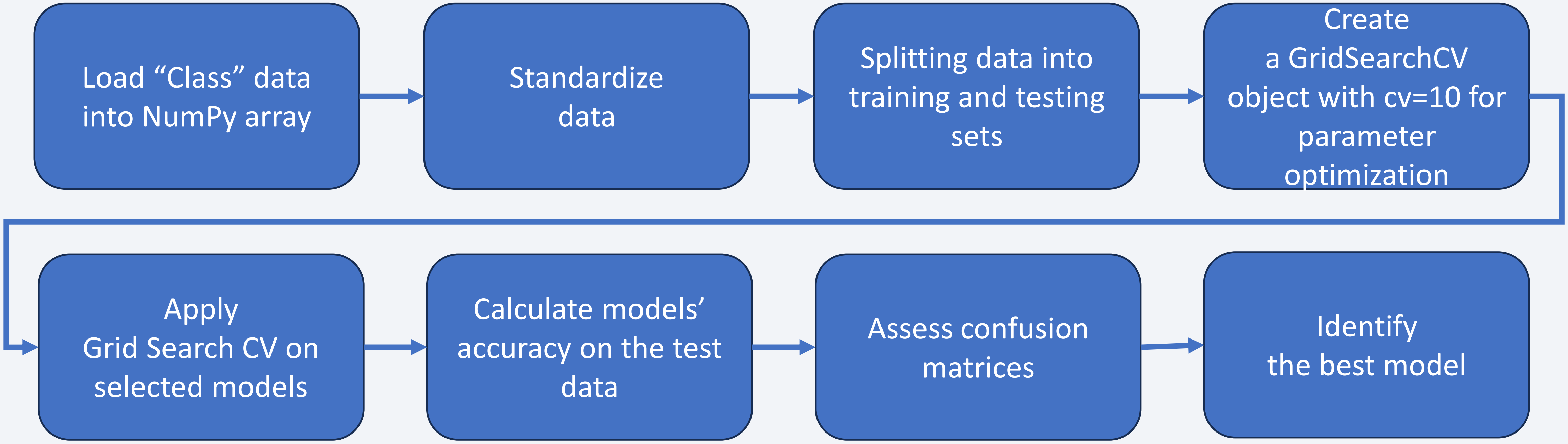
Dashboard elements used:

- Dropdown list with Launch Sites for launch sites selection
- Pie chart showing Successful Launches for successful and failed launches visualization
- Slider of Payload Mass Range for payload mass range selection
- Scatter chart of Payload Mass vs. Success Rate by Booster Version for correlation between Payload and Launch Success

GitHub URL of the completed Plotly Dash lab

Predictive Analysis (Classification)

Four classification models (logistic regression, support vector machine, decision tree and K nearest neighbors) were compared in the following procedure:



[GitHub URL of the completed predictive analysis lab]

15

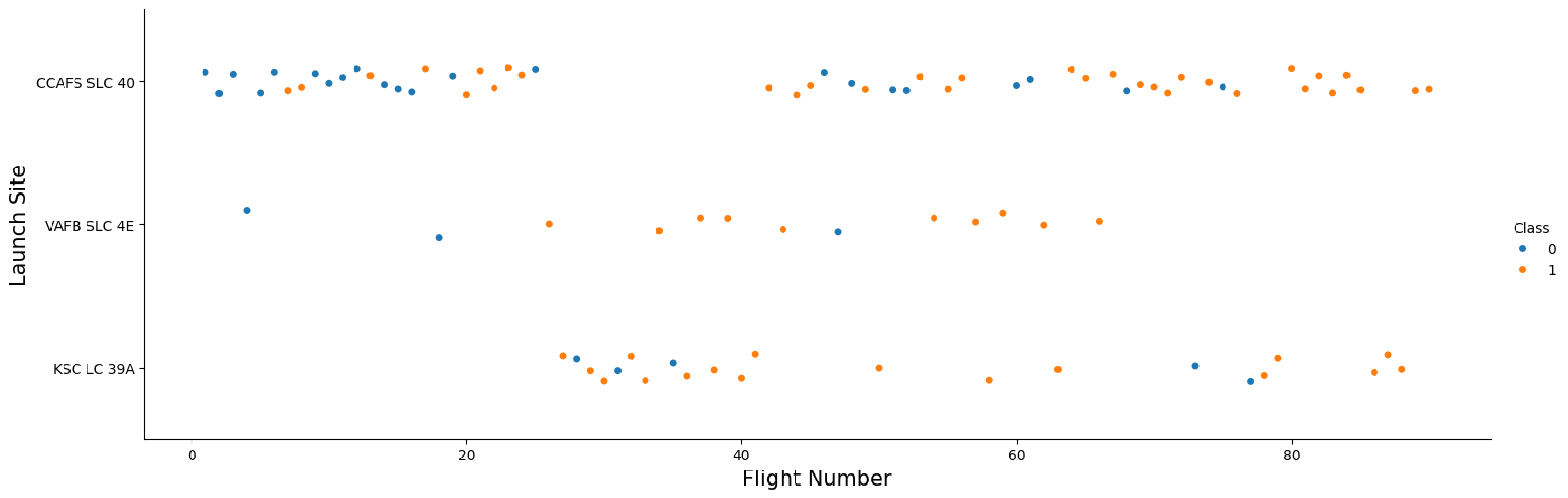
Results

- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

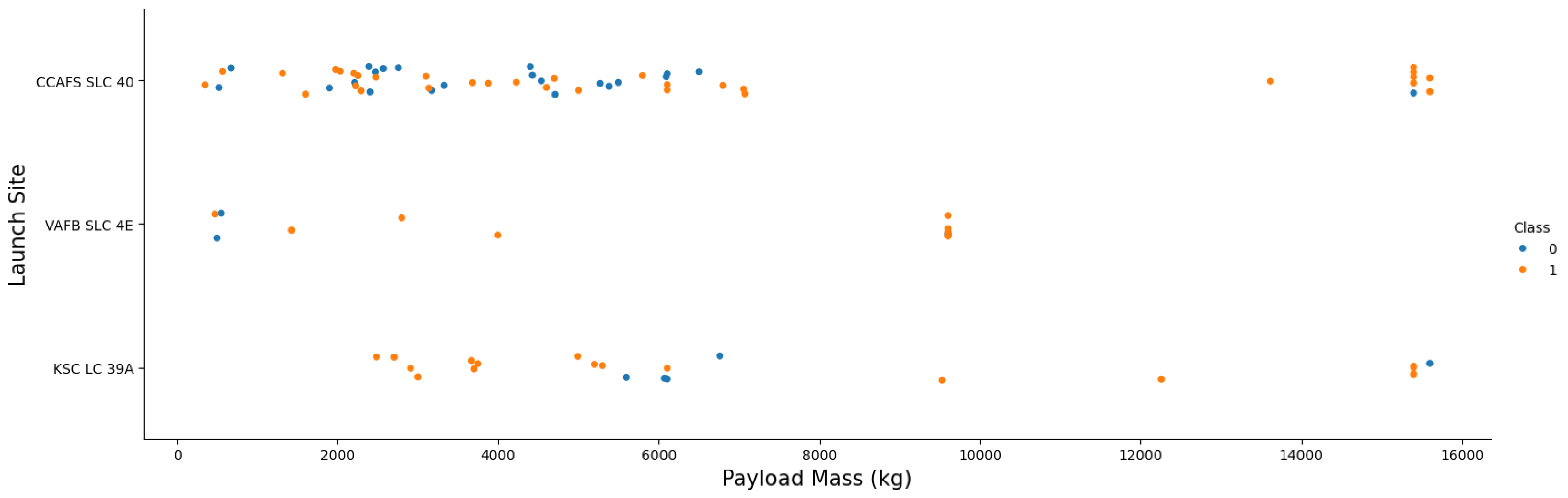
Insights drawn from EDA

Flight Number vs. Launch Site



- The plot verifies that the best launch is CCAF5 SLC 40, where most of recent launches were successful
- New launches have a higher success rate

Payload vs. Launch Site



- Most launces with high payloads (>7,000 kg) were successful
- KSC LC 39A has a 100% success rate for lower payloads (<5,000 kg)
- VAFB SKC 4E has no launches with payloads higher than 10,000 kg

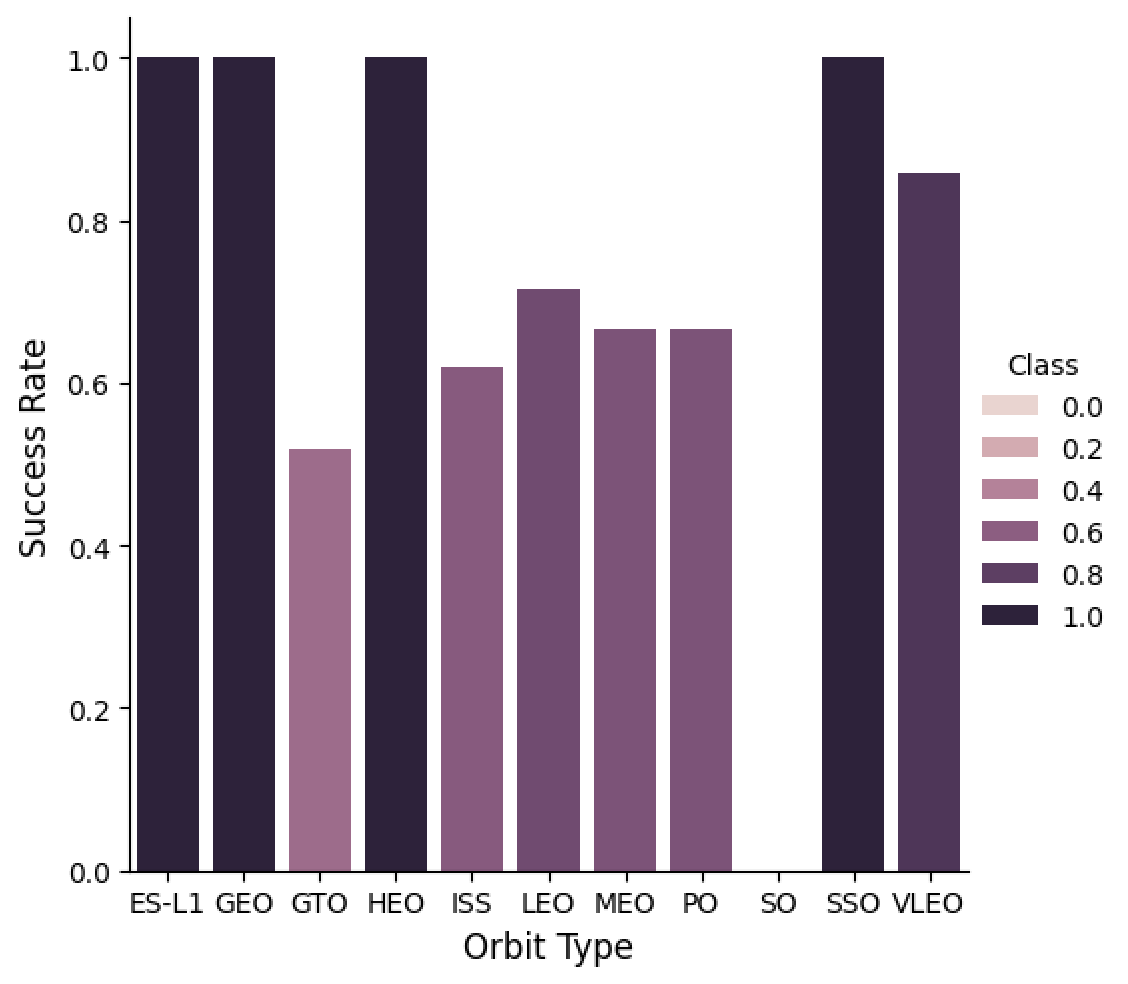
Success Rate vs. Orbit Type

100% success rate orbits:

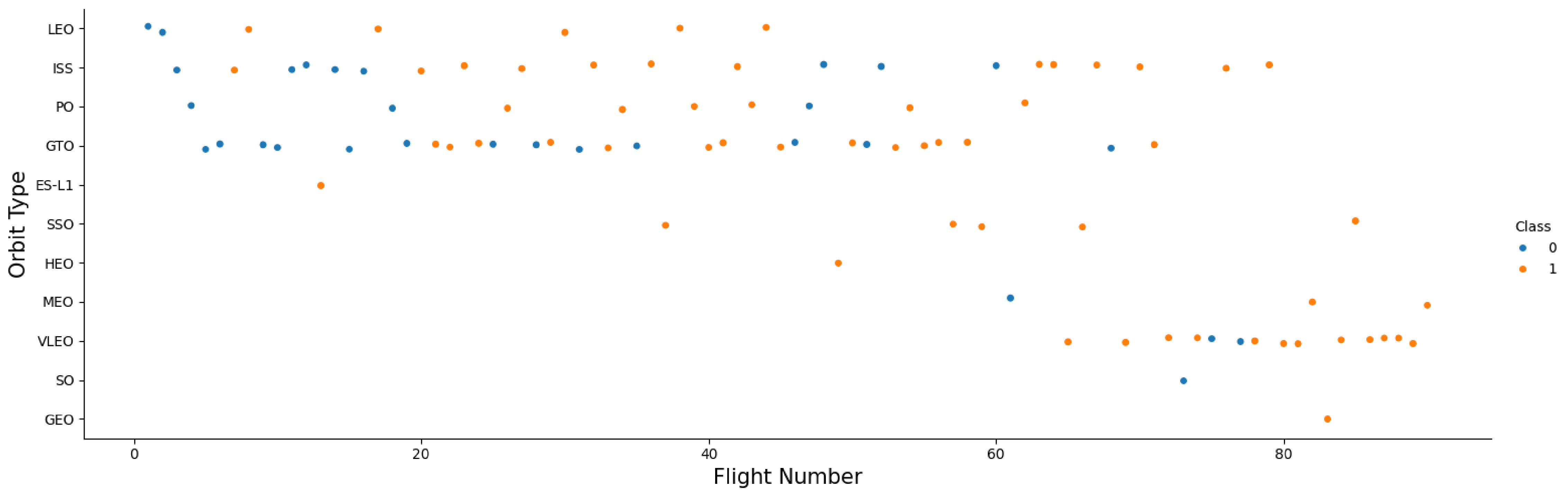
- ES L1
- GEO
- HEO
- SSO

0% success rate orbits:

- SO

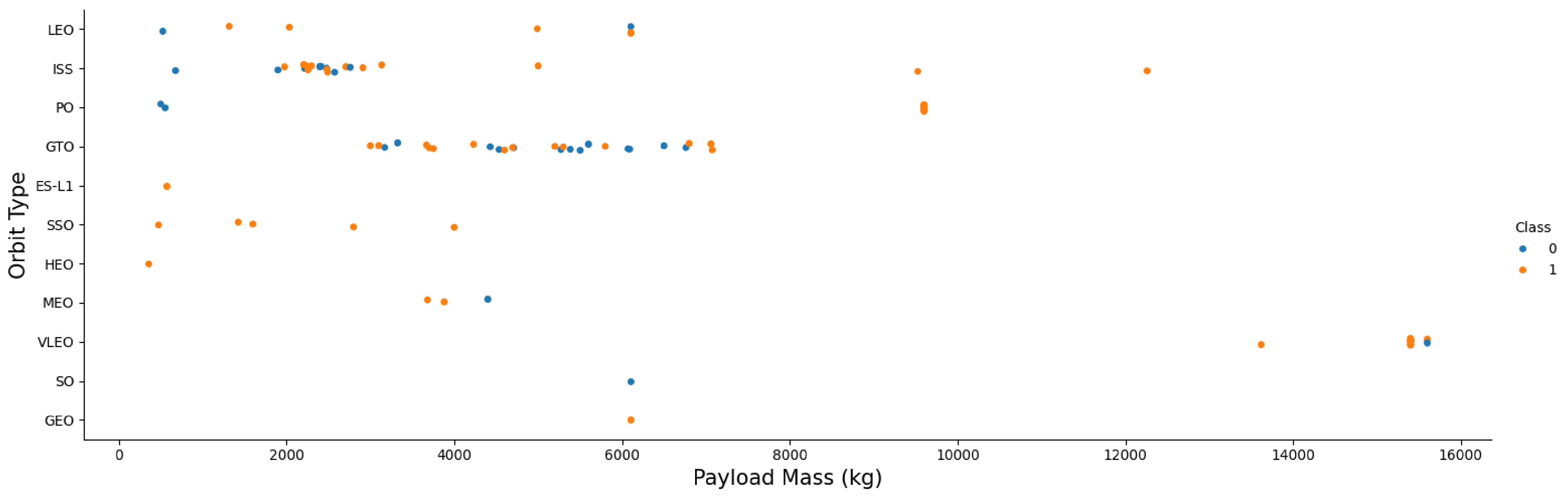


Flight Number vs. Orbit Type



- Generally, success rates improved with the number of flights
- Success of LEO orbit appears related to the number of flights
- No relationship between flight number when in GTO orbit

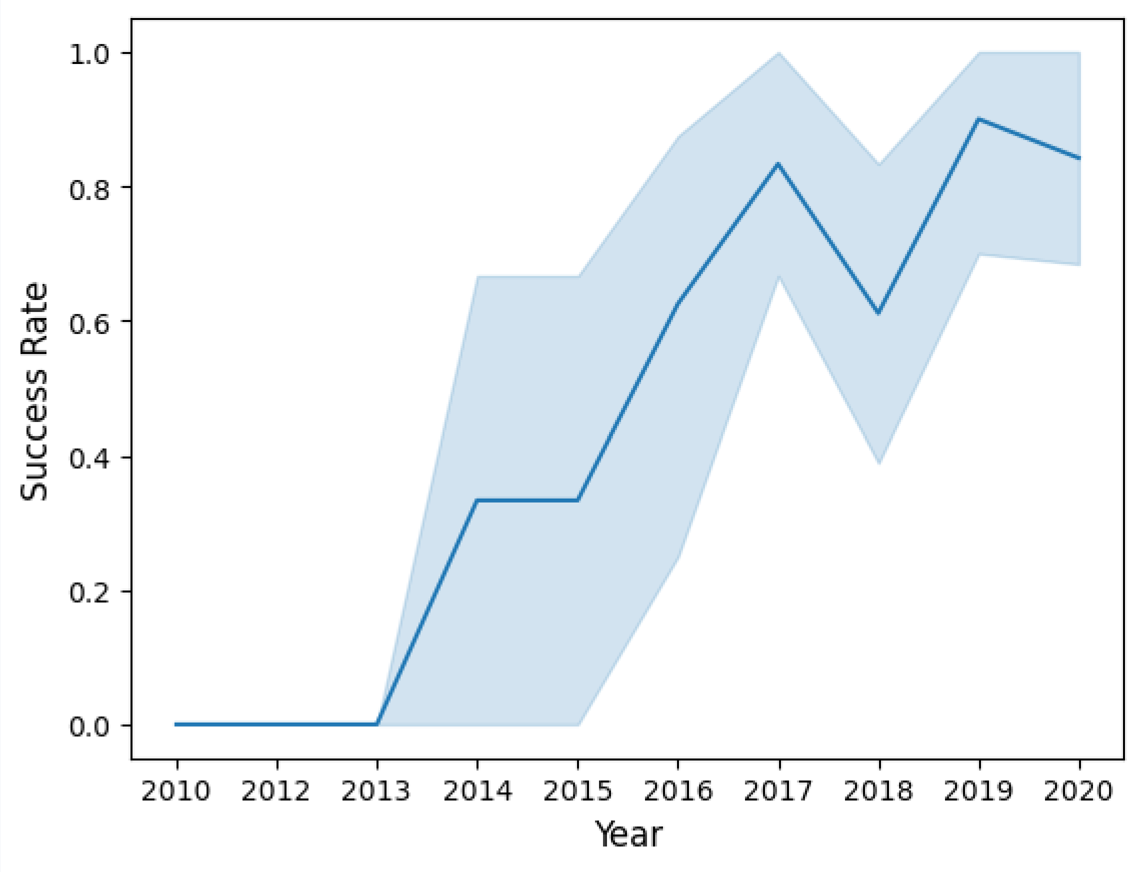
Payload vs. Orbit Type



- LEO, ISS and PO orbits present higher success rate with heavy payloads
- No relation between payload and success rate for GTO orbit

Launch Success Yearly Trend

- Overall, success rate has improved since 2013
- Success rate decreased in 2018



All Launch Site Names

Unique launch site names:

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

Launch Site Names Begin with 'CCA'

5 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	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

Total Payload Mass

Total payload carried by boosters from NASA:

SUM(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

Average payload mass carried by booster version F9 v1.1:

AVG(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

Date of the first successful landing outcome on ground pad:

MIN(DATE)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

Names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000:

Payload
JCSAT-14
JCSAT-16
SES-10
SES-11 / EchoStar 105

Total Number of Successful and Failure Mission Outcomes

Total number of successful and failure mission outcomes:

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

Boosters Carried Maximum Payload

Names of the booster which 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

2015 Launch Records

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

month	Date	Booster_Version	Launch_Site	Landing_Outcome
01	2015-01-10	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
04	2015-04-14	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

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

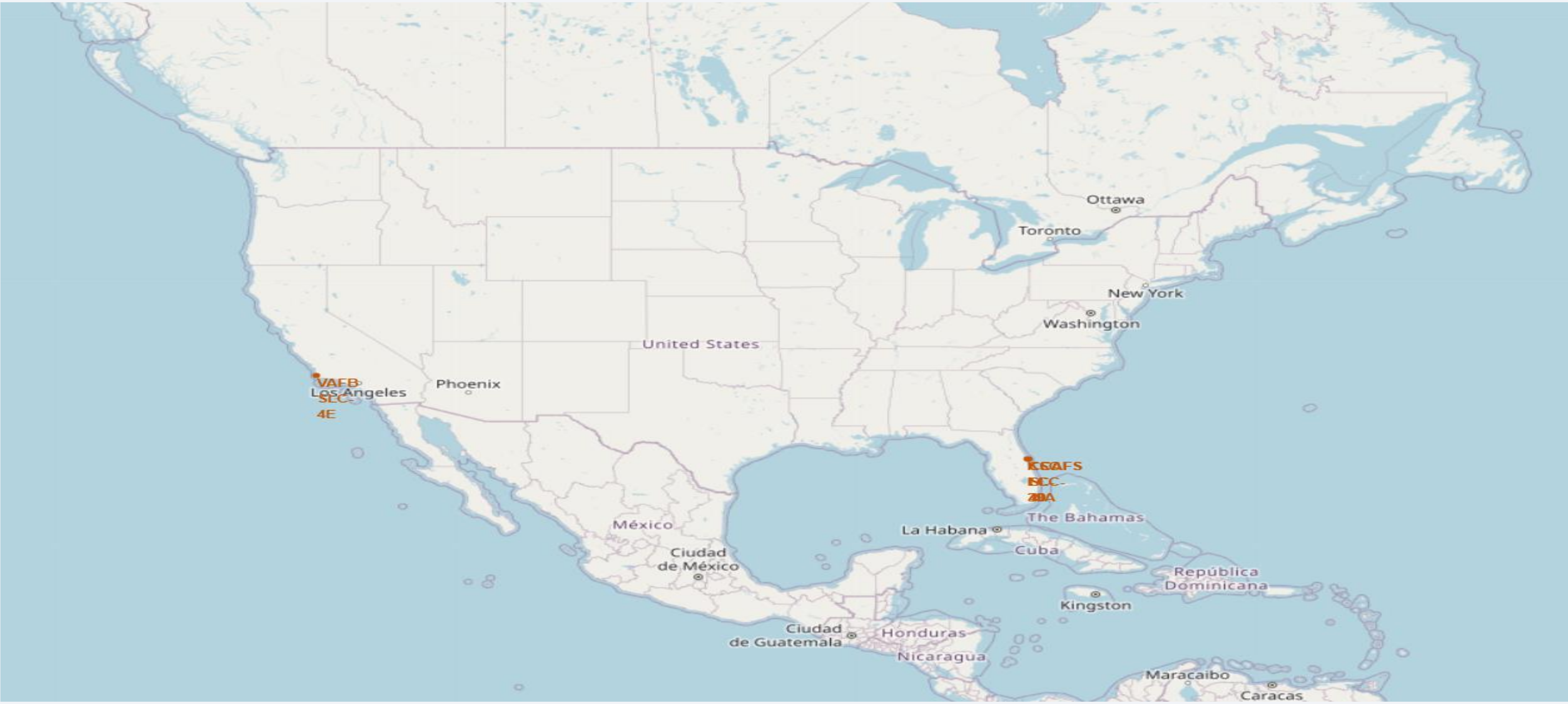
Ranked count of landing between the date 2010-06-04 and 2017-03-20, in descending order:

Landing_Outcome	count_outcomes
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

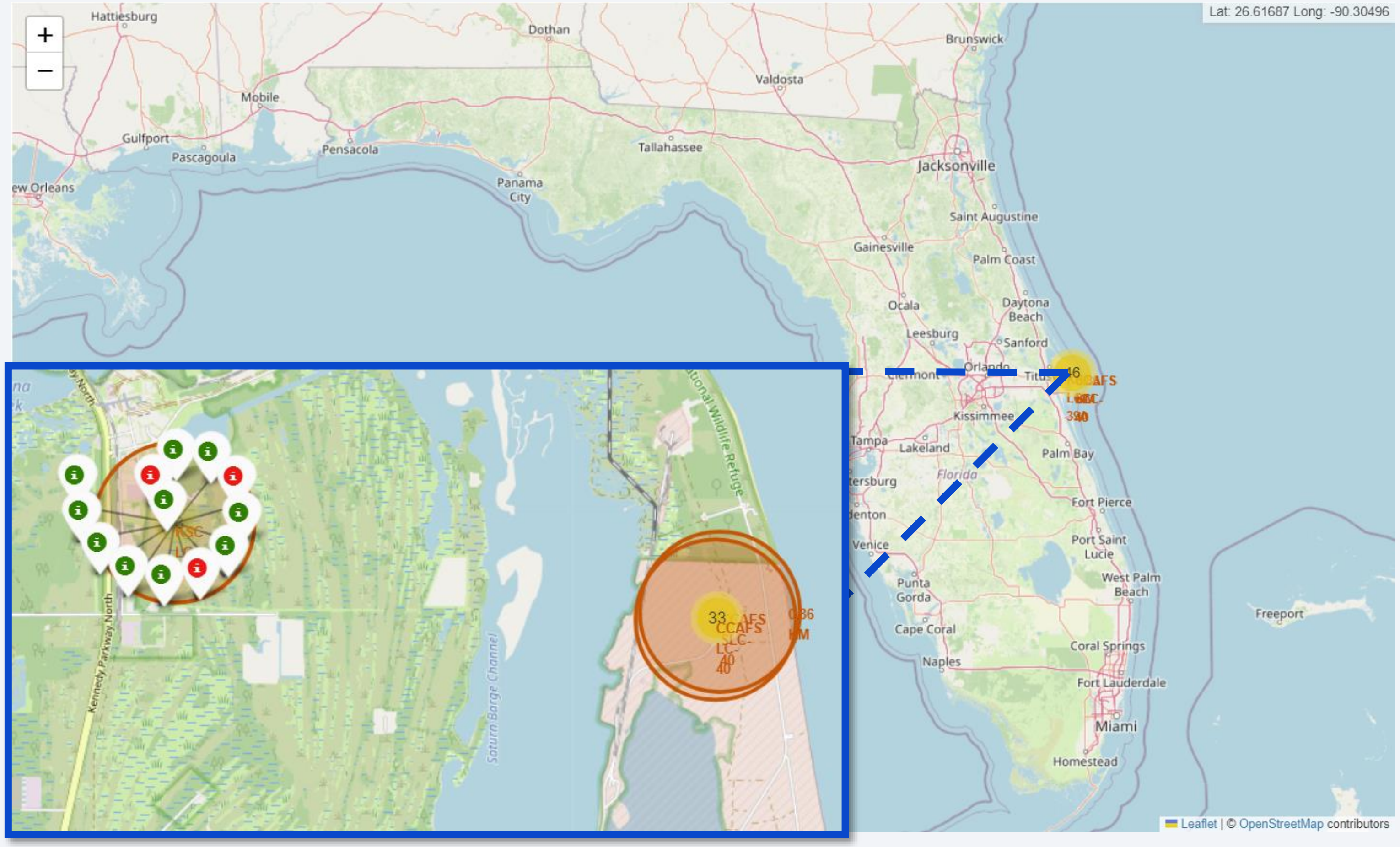
Launch Sites



Launch sites located as close as possible to the equator (within the constraint of access to coast of mainland US), providing natural boost due to the rotational speed of Earth.

Launch Outcomes

- Example launch site (KSS LC 39A) outcomes
- Green markers indicate successful and red ones indicate failure.

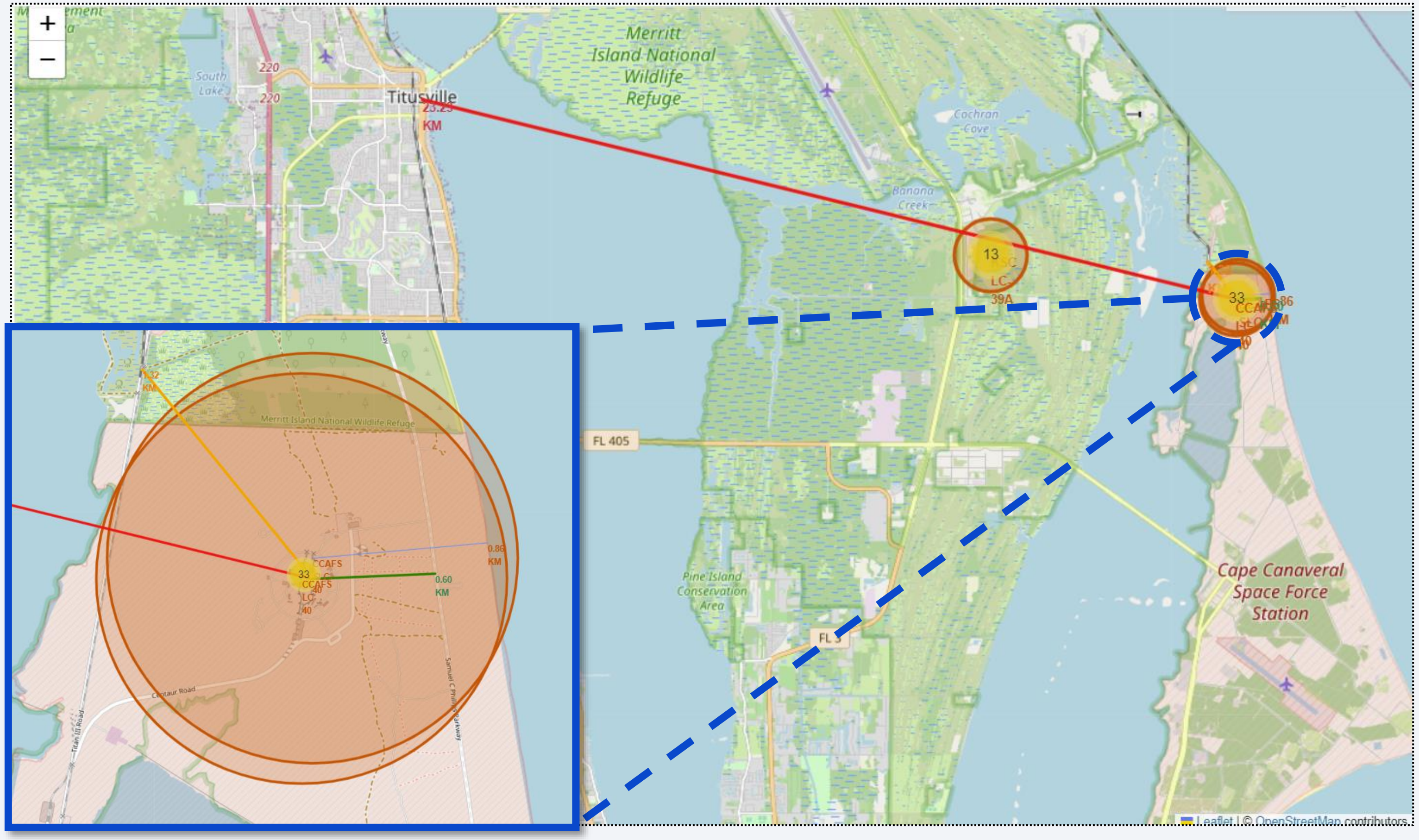


Logistics and Safety

Launch sites are located in close proximity to

- Highways - allowing for easy transport of required personnel and property
- Railways – allowing for easy transport of heavy cargo
- Coastline – allowing for easy access to water landing areas

Launch sites are located in safe distance to cities, which minimizes danger to population dense areas

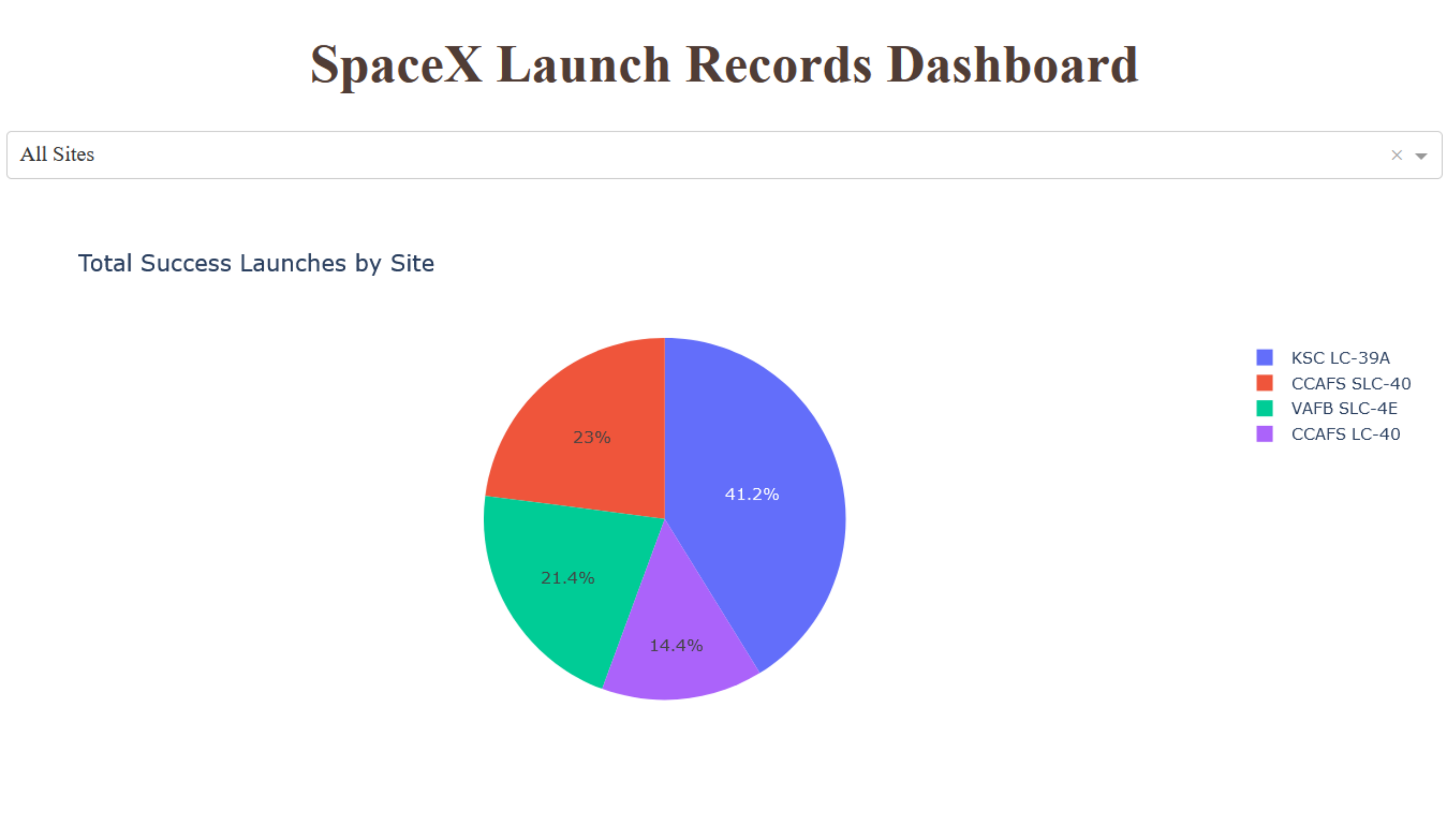


Section 4

Build a Dashboard with Plotly Dash

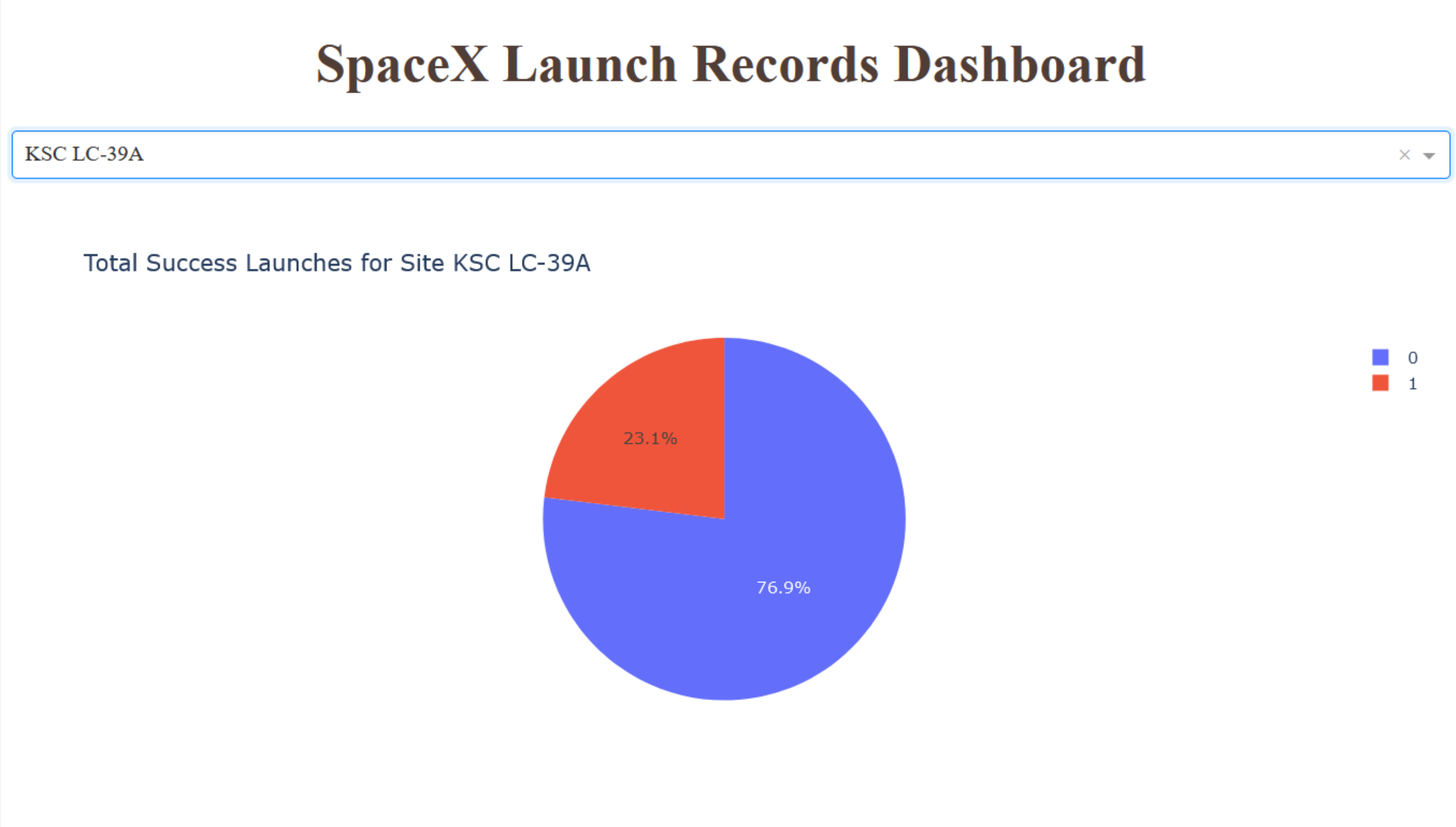
Launch Success

KSC LC 39A has the most successful launches amongst launch sites (41.2%)



Launch Success of KSC LC-39A site

KSC LC 39A has the highest success rate amongst launch sites (76.9%)



Payload Mass and Success

Payloads between 2,000 kg and 6,000 kg have the highest success rate

Section 5

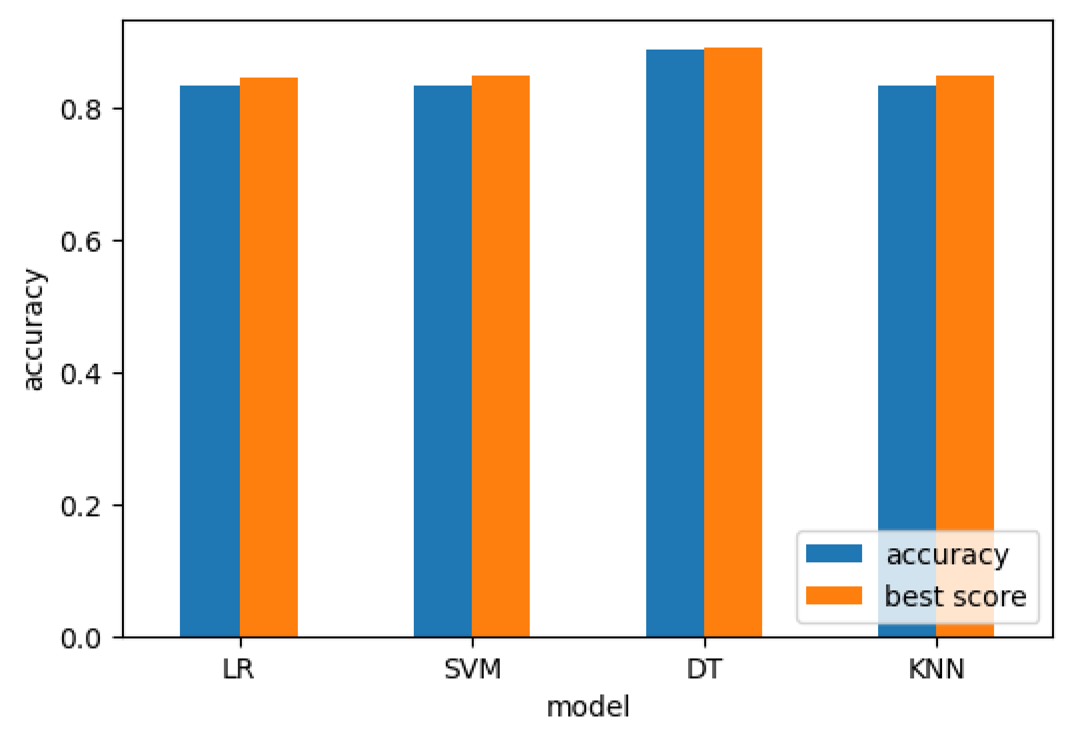
Predictive Analysis (Classification)

Classification Accuracy

Four classification models were tested:

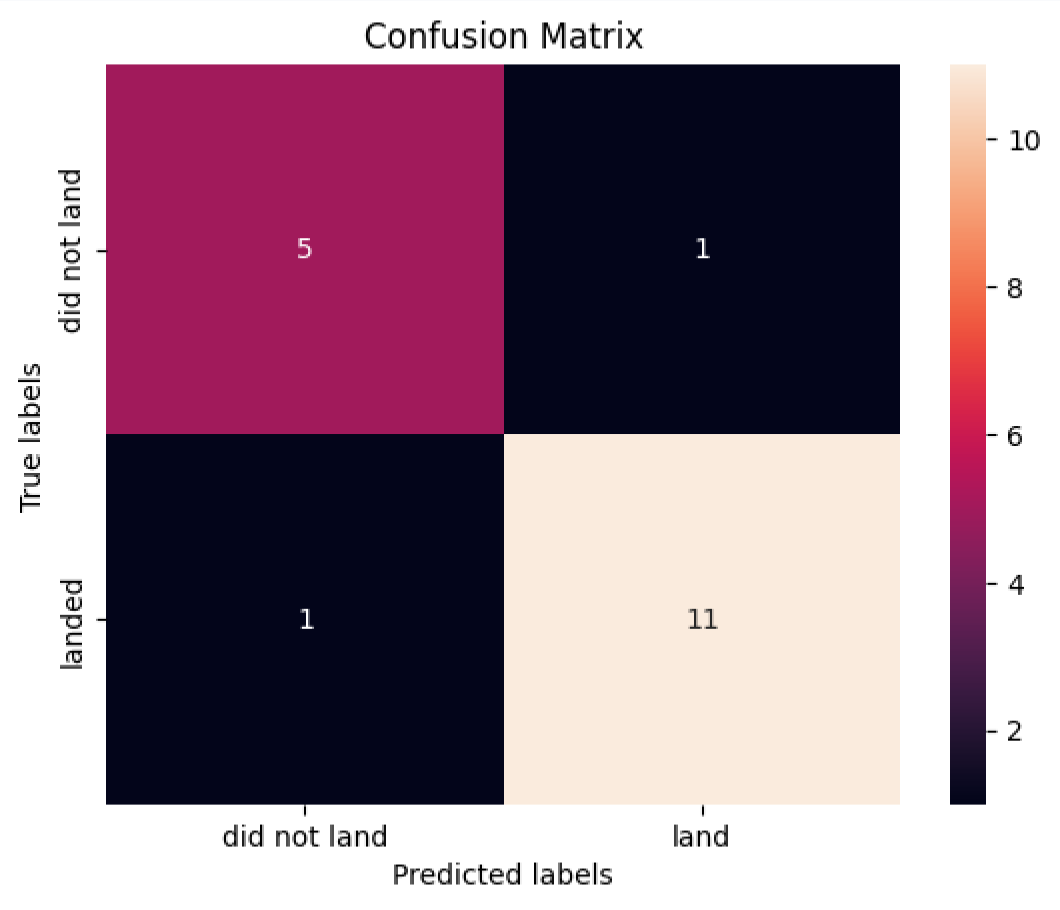
- LR = logistic regression
- SVM = support vector machine,
- DT = decision tree
- KNN = K nearest neighbors

The model with the highest classification accuracy is Decision Tree Classifier which has accuracies over than 87%.



Confusion Matrix

Confusion matrix of Decision Tree Classifier proves its accuracy by showing the big numbers of true positive and true negative compared to the false ones



Conclusions

Data analysis identified

- Best launch site: KSC LC-39A
- Bets orbits: ES L1, GEO, HEO, SSO
- Optimal payload: over 7,000 kg
- Landing success rate continues to improve over time
- Decision Tree Classifier can be used to predict successful landings and increase profits

Thank you!