IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Akash Nigale
22-Nov-2023

Outline

Executive Summary

- Summary of methodologies

The purpose of this research is to analyze what factors contribute to a successful rocket landing, using the following methodologies:

1. Collect
2. Wrangle
3. Explore
4. Analyze
5. Visualize
6. Model

- Summary of all results

Introduction

- BACKGROUND

As a leader in the aerospace industry, SpaceX aims to create affordable and accessible space travel and has distinguished itself from competitors by successfully developing a method by which the first stage of the Falcon 9 Rocket is reused thus resulting in costs cut by more than half.

- PURPOSE

Analyze which of the following factors affect the success of first-stage landing:

- Payload Mass
- Launch Site
- Flight Quantity
- Orbit Trajectory

Section 1

Methodology

Methodology

Data Collection

Two methodologies: web scraping and SpaceX's REST API

Data Wrangling

Clean data via missing value extraction, one hot encoding, and filtering

Exploratory Data Analysis

Exploration via EDA and SQL

Visual Analysis

Create plots and charts with Plotly Dash

Predictive Analysis

Insert variables into python models to predicts future landing outcomes

Data Collection

API

Resource from SpaceX allowing you to retrieve launch data – requires basic data wrangling and formatting

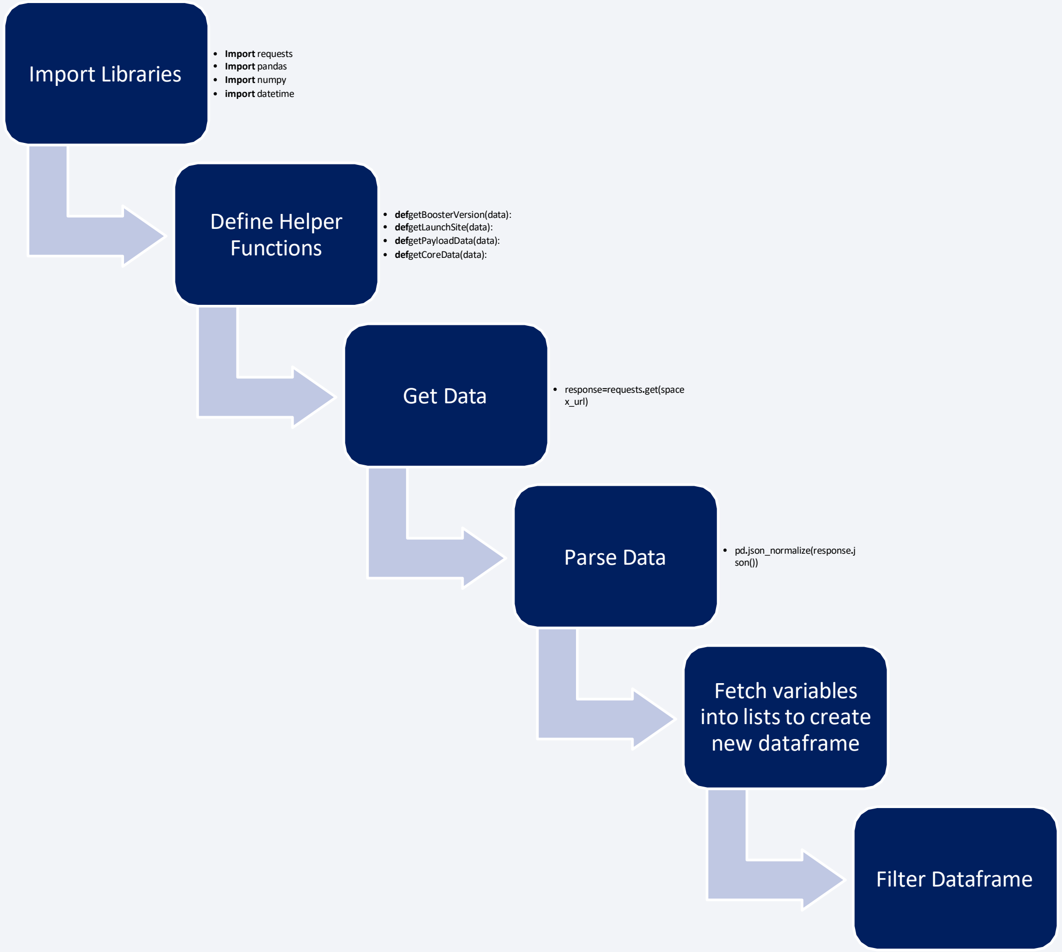
WEB SCRAPING

Extraction of data from Wikipedia page titled *List of Falcon 9 and Falcon Heavy Launches*

SPACEX API

1. *Import required libraries into Python*
2. *Define helper functions*
3. *Request data from SpaceX API (rocket launch data)*
4. *Decode response using .json() and convert to a dataframe using .json_normalize()*
5. *Extract information with functions*
6. *Create dictionary from the data then create dataframe from the dictionary*
7. *Filter dataframe to contain only Falcon 9 launches*
8. *Replace missing values to only include Falcon 9 Launches*

https://github.com/OLSI7046/COURSERAUSB/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

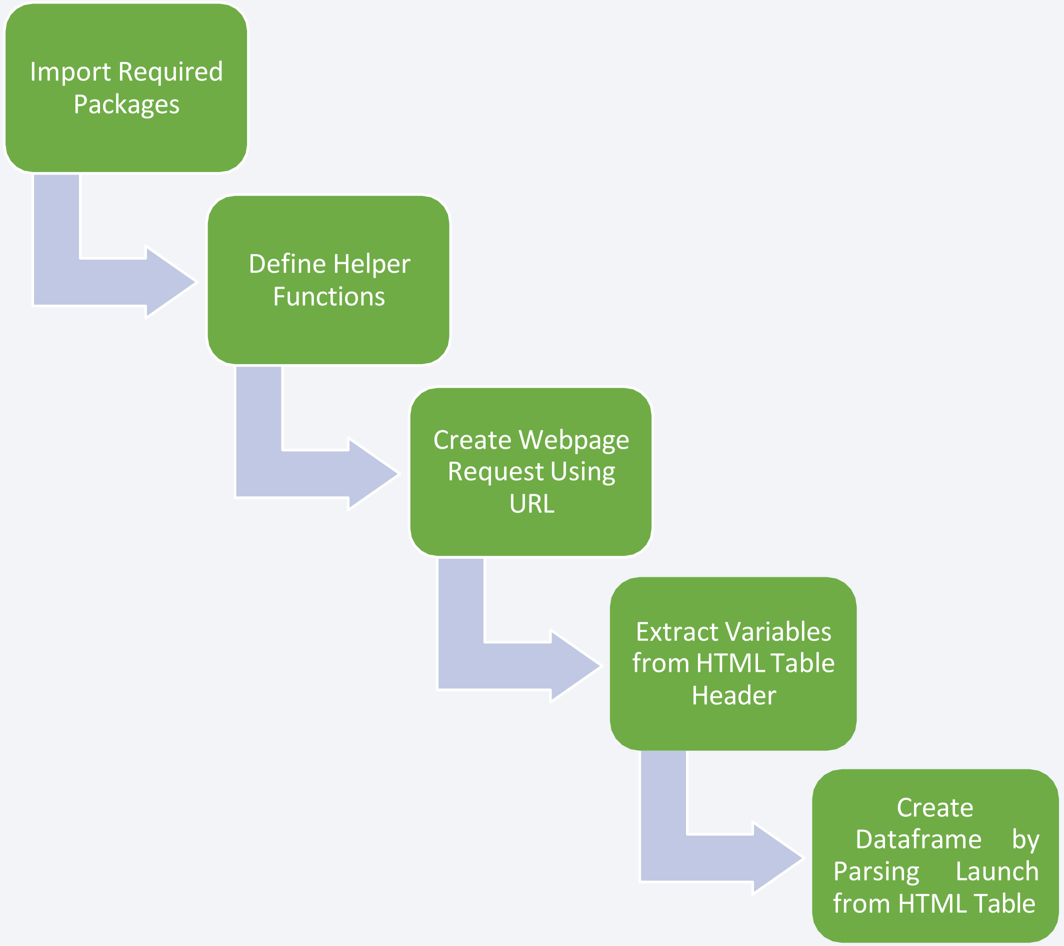


Data Collection

WEB SCRAPING

1. *Request data (Falcon 9 launch data) from Wikipedia*
2. *Create BeautifulSoup object from HTML response*
3. *Extract column names from HTML table header*
4. *Collect data from parsing HTML tables*
5. *Create dictionary from the data*
6. *Create dataframe from the dictionary*
7. *Export data to csv file*

https://github.com/OLSI7046/COURSERAUSB/blob/main/jupyter-labs-webscraping.ipynb



Data Wrangling

- Perform EDA and determine data labels
- Calculate:
  - # of launches for each site
  - # and occurrence of orbit
  - # and occurrence of mission outcome per orbit type]
- Create binary landing outcome column (dependent variable)

CHARTS PLOTTED:

- Flight Number vs. Payload
- Flight Number vs. Launch Site
- Payload Mass (kg) vs. Launch Site
- Payload Mass (kg) vs. Orbit type

EDA with SQL

SQL QUERIES PERFORMED:

Query launch site data

a.SELECT Unique(LAUNCH_SITE) FROM SPACEXTBL;
- Display the names of the unique launch sites in the space mission

- SELECT * FROM SPACEXTBL WHERE LAUNCH_SITE LIKE'CCA%' LIMIT 5;
  - Display 5 records where launch sites begin with the string 'CCA’
- SELECT MIN(DATE) FROM SPACEXTBL WHERE LANDING__OUTCOME = 'Success (ground pad)’
  - List the date when the first successful landing outcome in ground pad was achieved.

Query payload mass data

a.SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE CUSTOMER = 'NASA (CRS)’;
- Display the total payload mass carried by boosters launched by NASA (CRS)

- SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE BOOSTER_VERSION = 'F9 v1.1’;
  - Display average payload mass carried by booster version F9 v1.1
- SELECT PAYLOAD FROM SPACEXTBL WHERE LANDING__OUTCOME = 'Success (drone ship)' AND PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000;
  - List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- SELECT BOOSTER_VERSION FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL);
  - List the names of the booster versions which have carried the maximum payload mass. Use a subquery

Query successful launches

a.SELECT MISSION_OUTCOME, COUNT(*) as total_number FROM SPACEXTBL GROUP BY MISSION_OUTCOME;
- List the total number of successful and failure mission outcomes
- SELECT substr(Date,4,2) as month, DATE,BOOSTER_VERSION, LAUNCH_SITE, [Landing _Outcome] FROM SPACEXTBL where [Landing _Outcome] = 'Failure (drone ship)' and substr(Date,7,4)='2015’;
- 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.
- SELECT [Landing _Outcome], count(*) as count_outcomes FROM SPACEXTBL WHERE DATE between '04-06-2010' and '20-03-2017' group by [Landing _Outcome] order by count_outcomes DESC;
  - Rank the count of successful landing_outcomes between the date 04-06-2010 and 20-03-2017 in descending order.

Interactive Map with Folium

Launch Site Markers

- Blue Circle at NASA Johnson Space Center's coordinate with a popup label showing its name using its latitude and longitude coordinates
- Red circles at all launch sites coordinates with a popup label showing its name using its name using its latitude and longitude coordinates

Launch Outcomes

- Green marker for successful launches
- Red marker for unsuccessful launches

Launch Site Proximities

- Colored lines to show distance between launch site CCAFS SLC40 and its proximity to the nearest coastline, railway, highway, and city

Dashboard with Plotly Dash

Dropdown List Launch Sites

- Allow user to select all launch sites or a certain launch site

Pie Chart Successful Launches

- Allow user to see successful and unsuccessful launches as a percent of the total

Slider Payload Mass Range

- Allow user to select payload mass range

Scatter Chart Payload Mass vs. Success Rate

- Allow user to see the correlation between Payload and Launch Success

- Summarize how you built, evaluated, improved, and found the best performing classification model
- You need present your model development process using key phrases and flowchart
- Add the GitHub URL of your completed predictive analysis lab, as an external reference and peer-review purpose

Results

Exploratory Data Analysis

- Launch success has improved over time
- KSC LC-39A has the highest success rate among landing sites
- Orbits ES-L1, GEO, HEO and SSO have a 100% success rate Results Summary

Visual Analytics

- Most launch sites are near the equator, and all are close to the coast
- Launch sites are far enough away from anything a failed launch can damage (city, highway, railway), while still close enough to bring people and material to support launch activities

Predictive Analytics

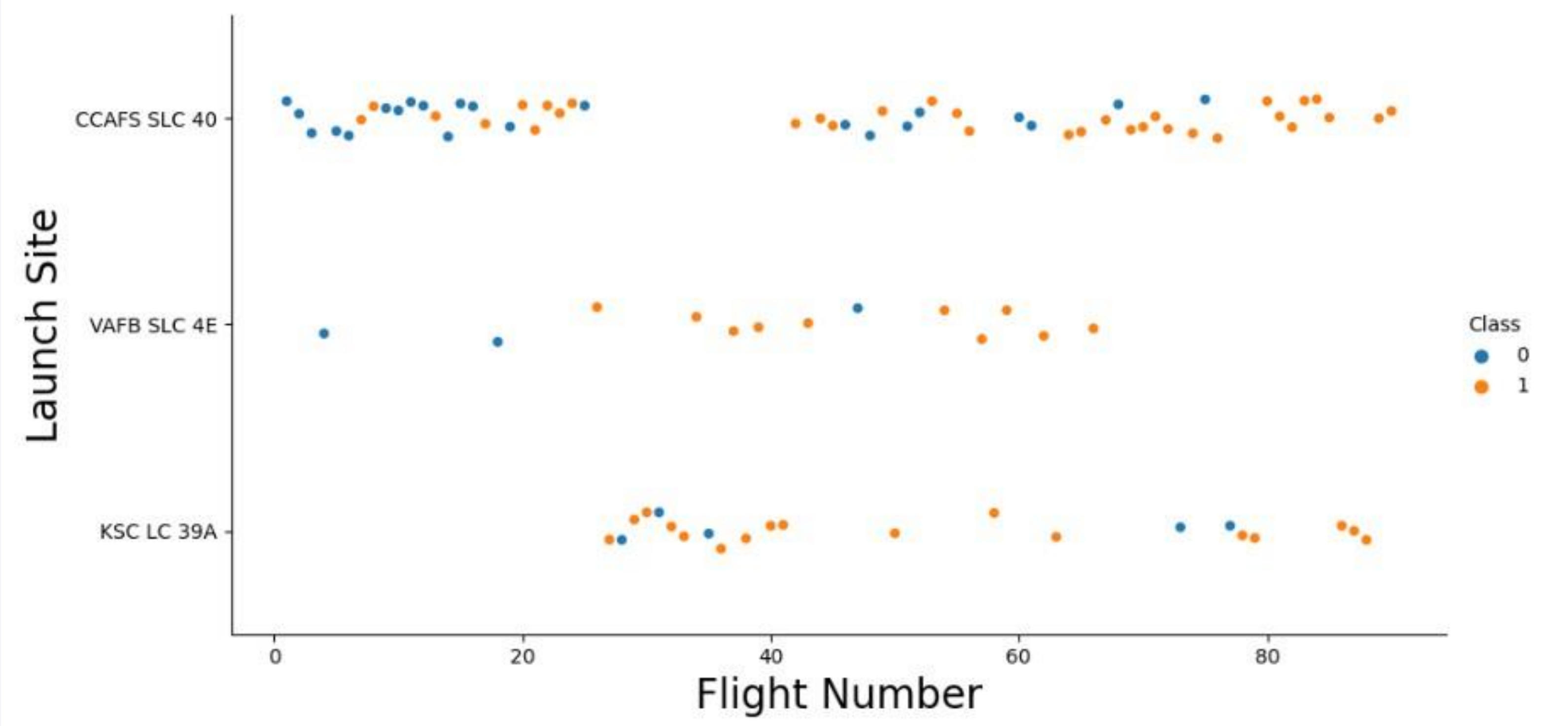
- Decision Tree model is the best predictive model for the dataset

Section 2

Insights drawn from EDA

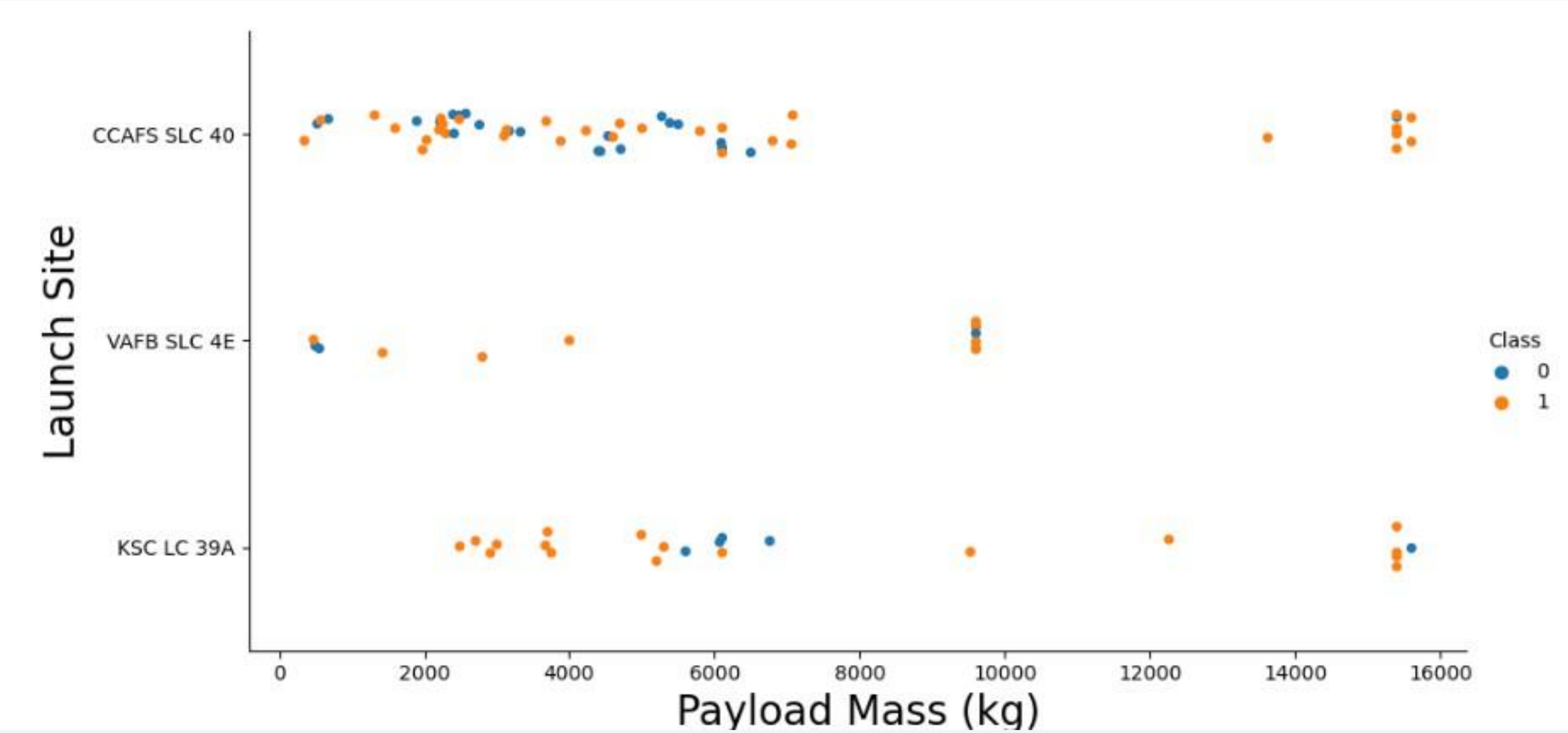
Flight Number vs. Launch Site

- Scatter plot indicates number of continuous launch attempts at each of three launch sites
- Color indicates “class” or success



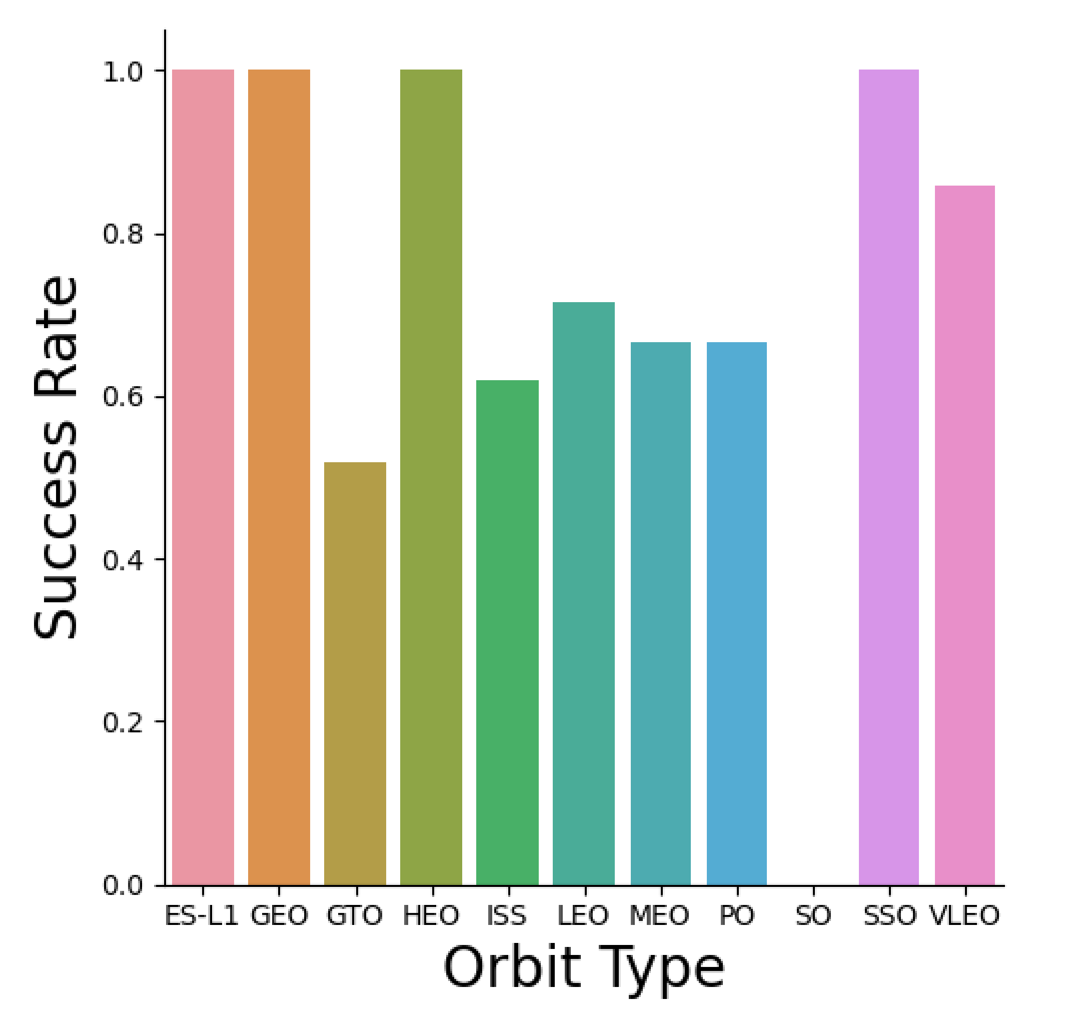
Payload vs. Launch Site

- Scatter plot indicates payload mass of each launch at each of three launch sites
- Color indicates "class" or success
- Most launches occur at site CCAFS SLC 40



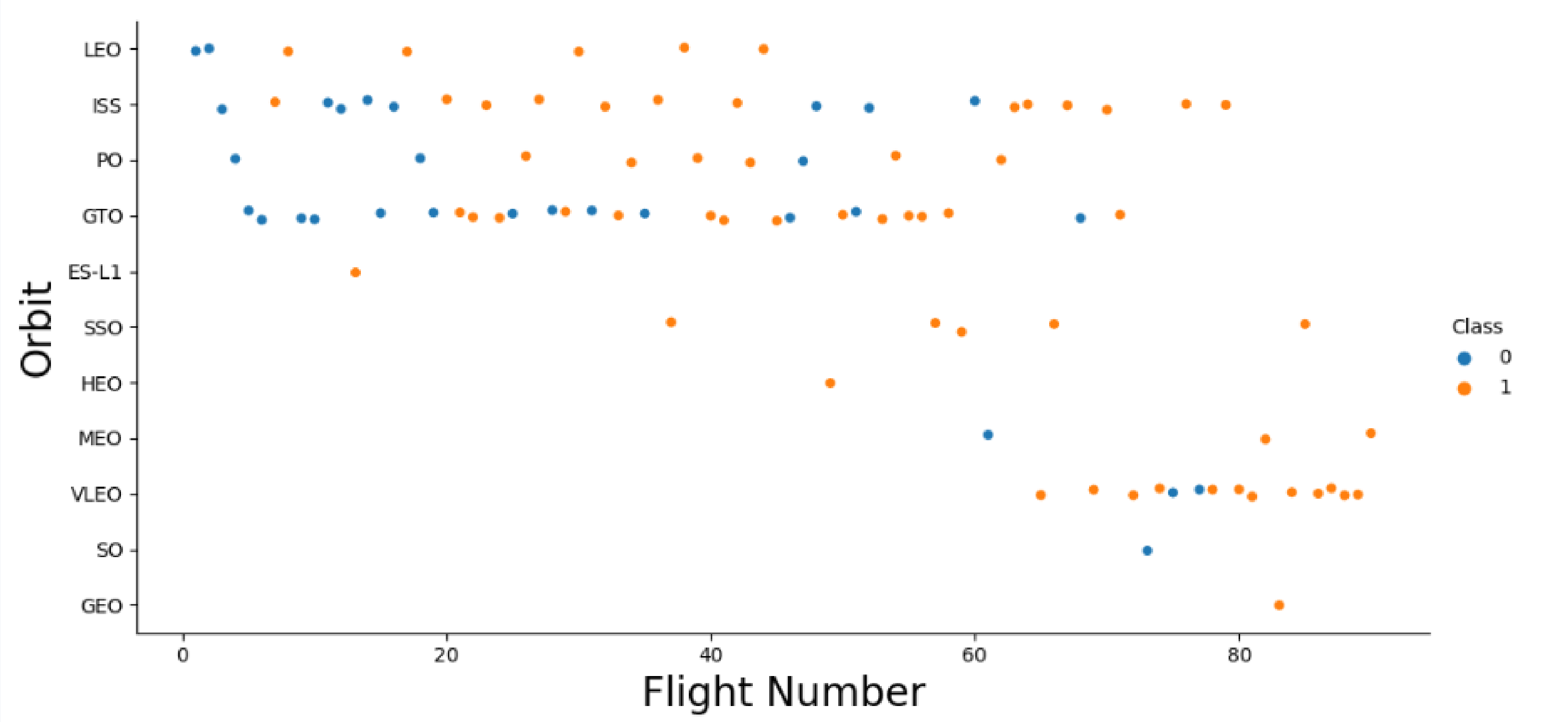
Success Rate vs. Orbit Type

- Each orbit type is accentuated with a varying hue
- Outliers include orbit type SO with 0% success rate
- Most successful orbit types with 100% success rate include:
  - ES-L1
  - GEO
  - HEO
  - SSO



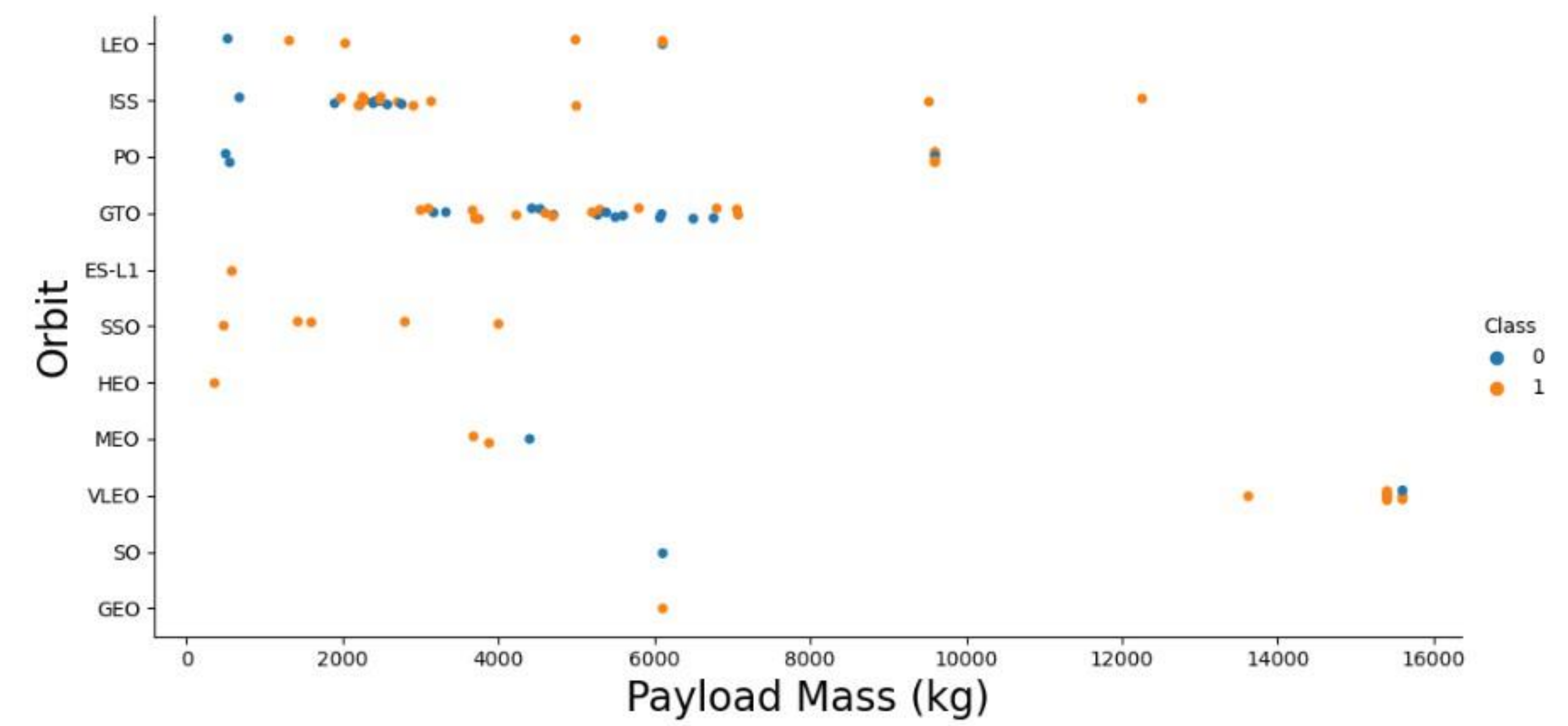
Flight Number vs. Orbit Type

- Scatter plot indicates number of continuous launch attempts per orbit type
- Orbit types SSO to GEO have higher flight numbers whereas orbit type LEO to ES-L1 have more variation



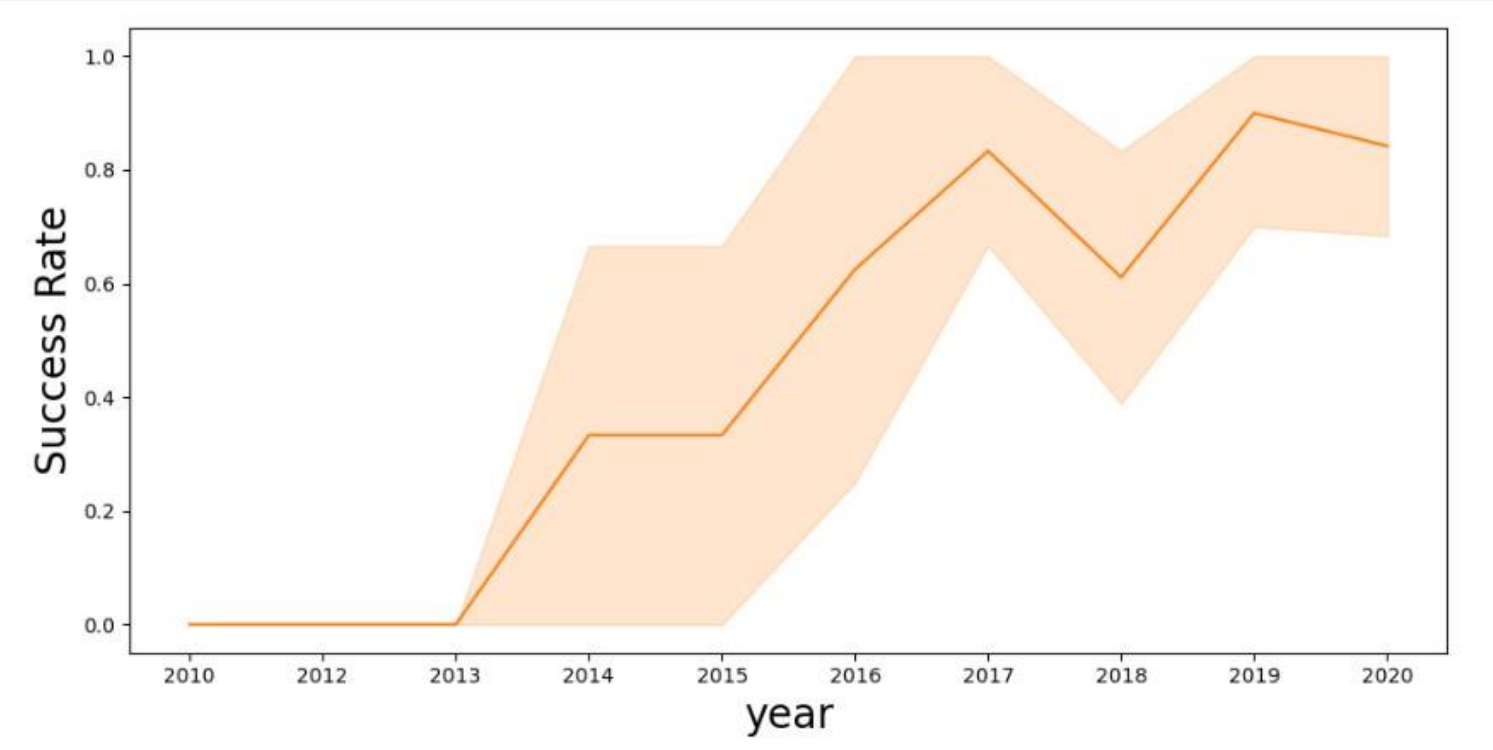
Payload vs. Orbit Type

- Scatter plot indicates payload mass of each launch per orbit type
- Color indicates “class” or success
- Average Payload Mass <8000kg



Launch Success Yearly Trend

- Line graph indicates average success rate of launches on a yearly basis
- Strong positive trend (up and to the right) indicates launches have improved over last decade



All Launch Site Names

```
SELECT Unique(LAUNCH_SITE)
FROM SPACEXTBL
```

Launch Site Names Begin with 'CCA'

```
SELECT *

FROM SPACEXTBL

WHERE LAUNCH_SITE LIKE 'CCA%'
LIMIT 5;
```

Dragon Spacecraft Qualification Unit	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	Dragon demo flight C2	SpaceX CRS-1	SpaceX CRS-2

Total Payload Mass

45,596
kg

Total Payload mass carried by NASA boosters equals roughly 45 thousand kilograms.
Query includes simple summation filtering to the NASA customer:

```
SELECT SUM(PAYLOAD_MASS_KG_)
FROM SPACEXTBL
WHERE CUSTOMER = 'NASA (CRS)’;
```

Average Payload Mass by F9 v1.1

2,928.4

kg

Average Payload mass carried by NASA boosters equals almost 3 thousand kilograms.
Query includes simple summation filtering to the ‘F9 v1.1’ booster version:

```
SELECT AVG(PAYLOAD_MASS__KG_)
FROM SPACEXTBL
WHERE BOOSTER_VERSION = 'F9 v1.1';
```

First Successful Ground Landing Date

date

Average Payload mass carried by NASA boosters equals almost 3 thousand kilograms.
Query includes simple summation filtering to the ‘F9 v1.1’ booster version:

```
SELECT MIN(Date)
FROM SPACEXTBL
WHERE LANDING_OUTCOME = 'Success (ground pad)';
```

Successful Drone Ship Landing with Payload between 4000 and 6000

Query locates booster names by filtering two factors: landing outcome and payload mass:

```
SELECT PAYLOAD
FROM SPACEXTBL
WHERE LANDING_OUTCOME = 'Success (drone ship)'
AND PAYLOAD_MASS_KG_ BETWEEN 4000 AND 6000;
```

name

Total Number of Successful and Failure Mission Outcomes

1 Failure (in flight)	**99** Success	**1** Success (Payload Status Unclear)

Query results indicate >99% success rate with only one failure

```
SELECT MISSION_OUTCOME, COUNT(*) as 'Total Quantity'
FROM SPACEXTBL
GROUP BY MISSION_OUTCOME
```

Boosters Carried Maximum Payload

- 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

Boosters that carried the maximum payload mass are widely of the F9 kind

```
SELECT BOOSTER_VERSION
FROM SPACEXTBL
WHERE PAYLOAD_MASS__KG_ =
      (SELECT MAX(PAYLOAD_MASS__KG_)
      FROM SPACEXTBL);
```

2015 Launch Records

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

Querying failed landing outcomes for 2015 results in only two records – one in January and one in April – both from the same launch site

```
SELECT substr(Date,4,2) as ‘Month’, DATE, BOOSTER_VERSION, LAUNCH_SITE, [Landing _Outcome]
FROM SPACEXTBL
WHERE [Landing _Outcome] = 'Failure (drone ship)’ AND substr(Date,7,4) = '2015';
```

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

Landing _Outcome	Outcome_Count
Success	20
No attempt	10
Success (drone ship)	8
Success (ground pad)	6
Failure (drone ship)	4
Failure	3
Controlled (ocean)	3
Failure (parachute)	2
No attempt	1

Majority of Landing Outcomes between 2010 and 2017 were successful closely followed by no attempts.

Total number of failures = 9, constituting 15.79% of all launches in this timeframe

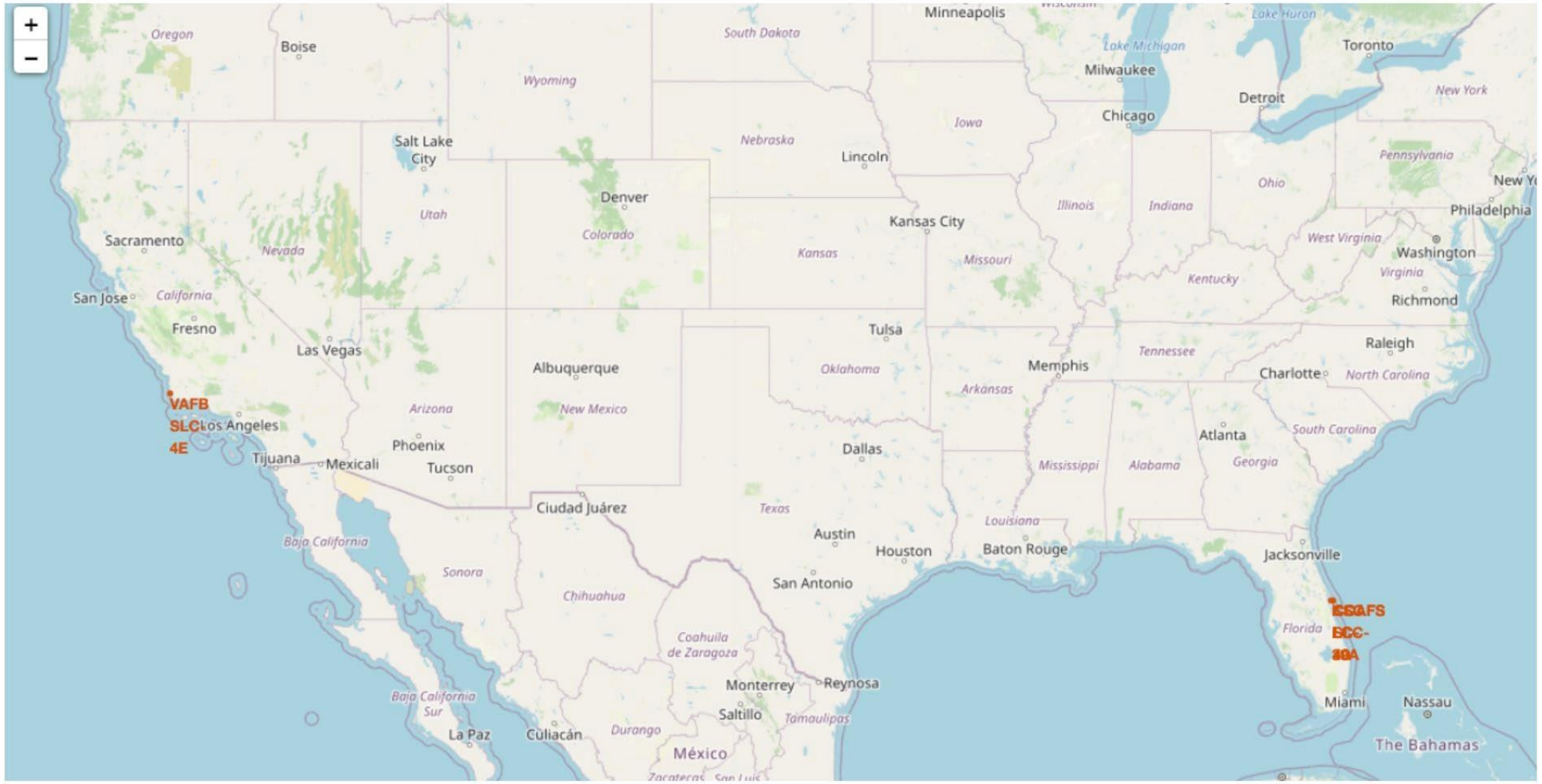
Section 3

Launch Sites Proximities Analysis

Launch Sites Map

Launch sites are situated close to the equator

- Easier to launch to equatorial orbit
- Rotation of the earth assists in launch
- Less fuel needed in boosters = saved costs



Launch Outcomes Map

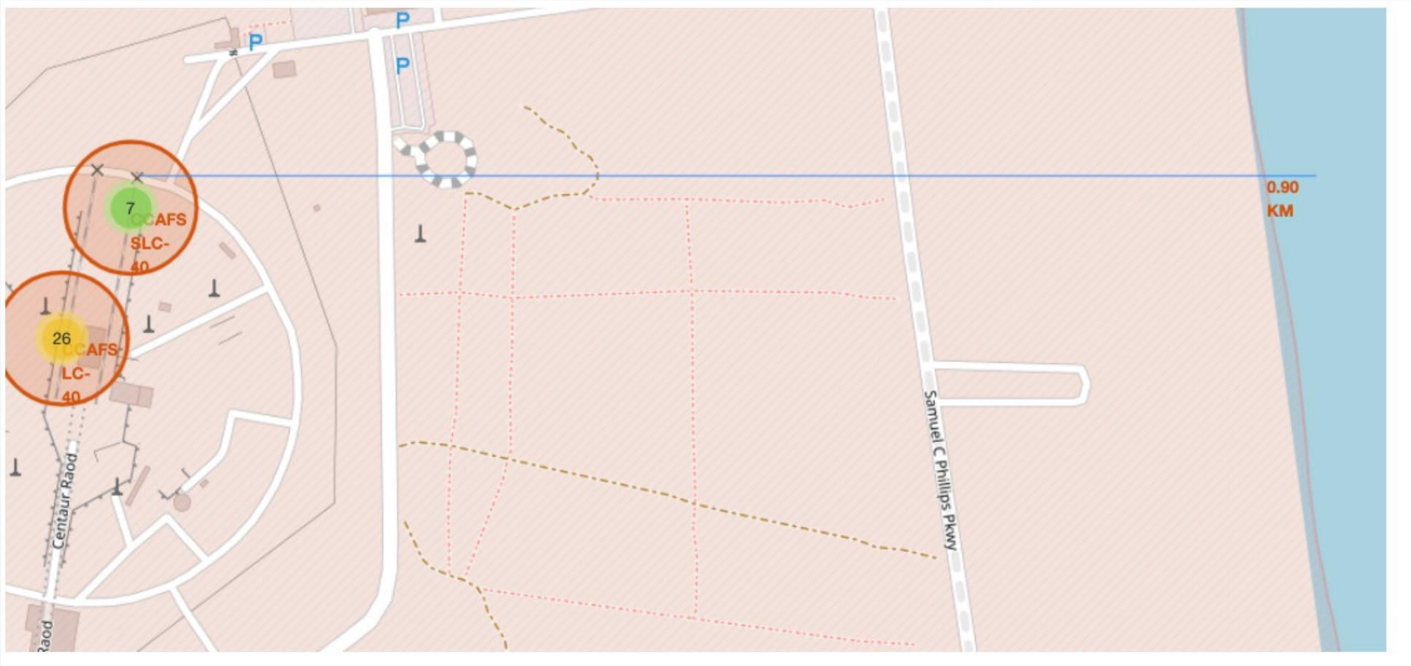
CCAFS SLC-40 Success Rate = 42.9%

- Successful Launches: 3
- Unsuccessful Launches: 4

Landmark Proximities

CCAFS SLC-40

- .86 km from nearest coastline:
  - *Stages dropping along launch path will not hit pedestrian property*
- 21.96 km from nearest railway
  - *Material transportation costs reduced with direct access to railway*
- 23.23 km from nearest city
  - *Accessible by employees*
- 26.88 km from nearest highway
  - *Supported infrastructure for employees and material transportation*



Section 4

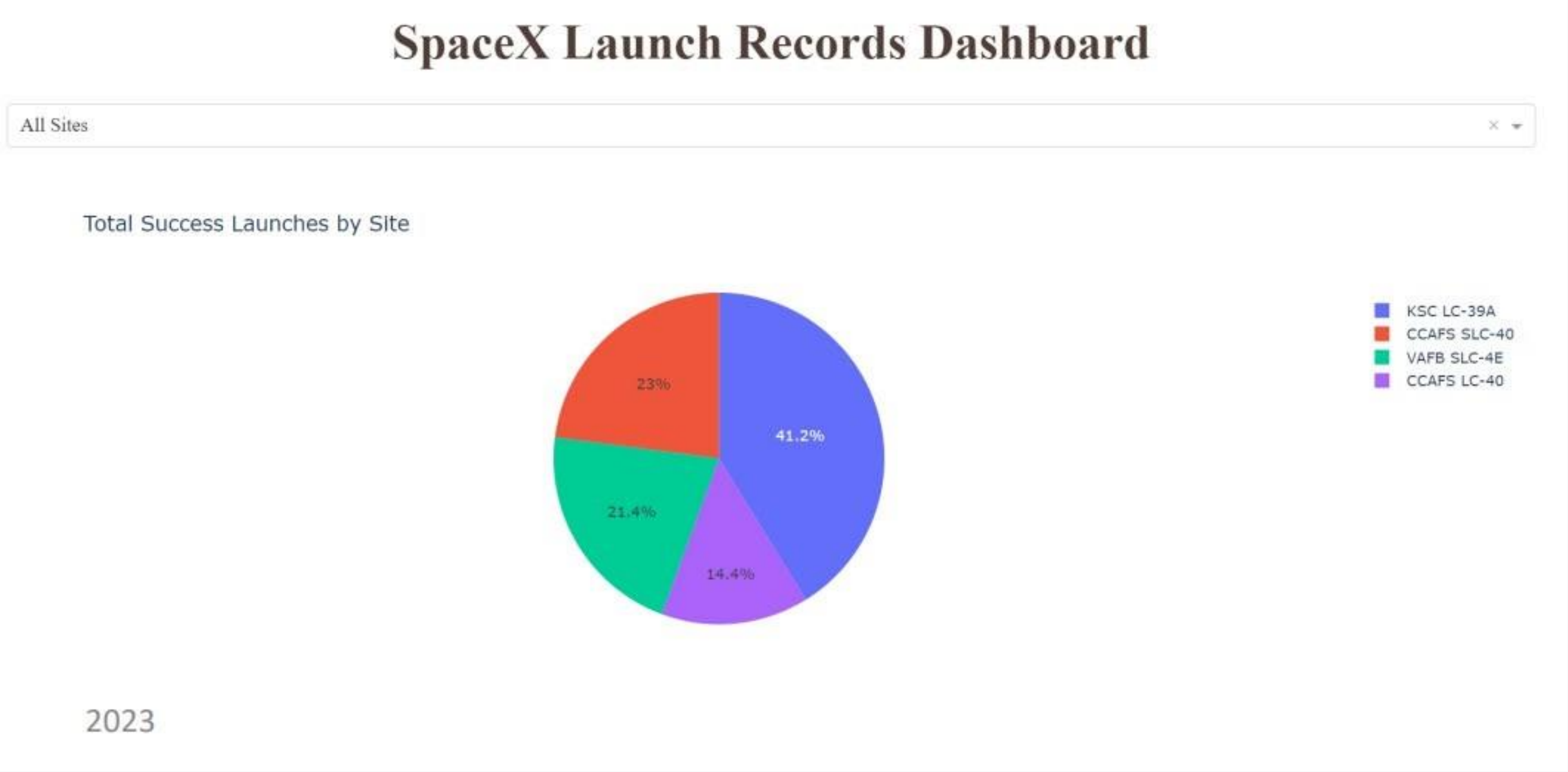
Build a Dashboard with Plotly Dash

Total Successful Launches by Site

Most successful site:

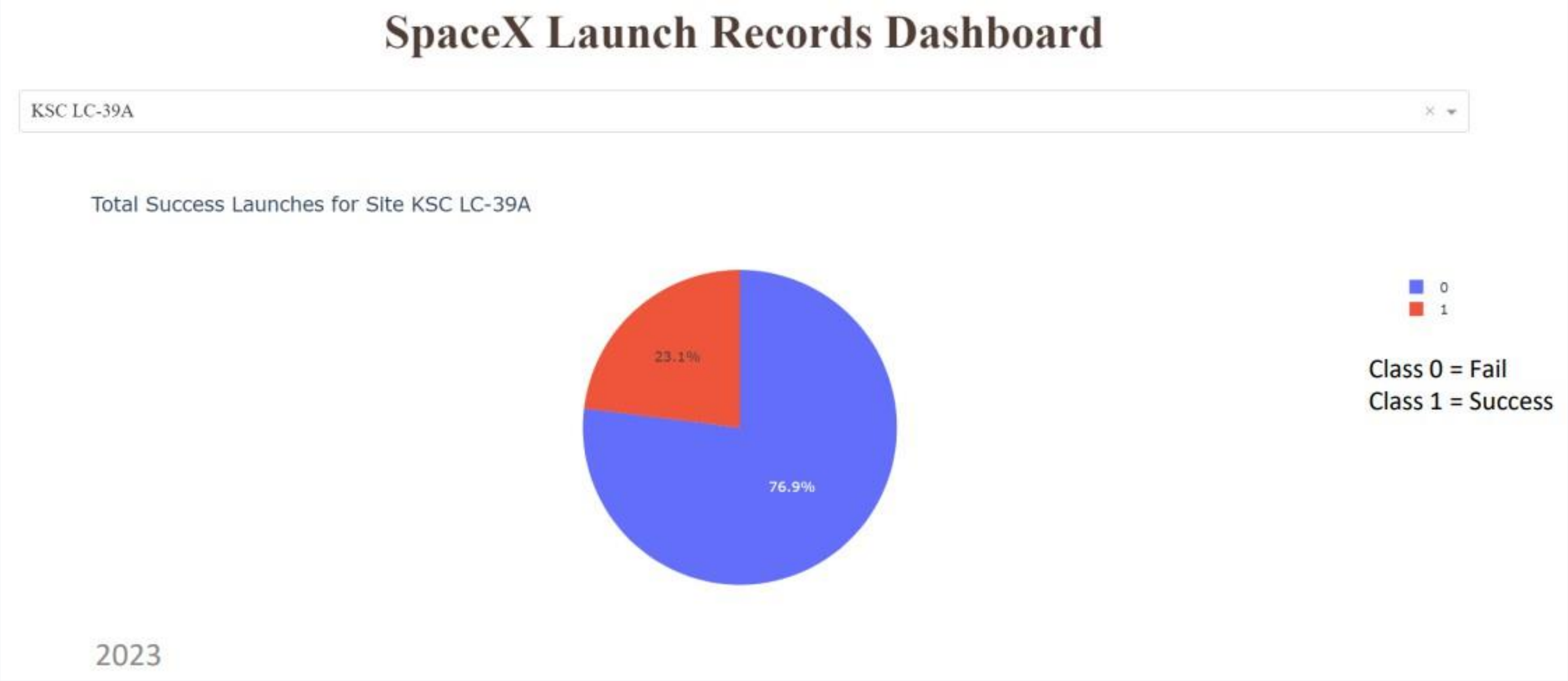
KSC LC-39A

41.2% of all successful launches



Total Successful Launches for Site KSC LC-39A

10 Successful Launches
3 Failed Launches
76.9% Success Rate



Payload vs Launch Success Correlation

2000-5000kg : range of payloads with highest success
FT, B4 : Majority of Successful Booster Version

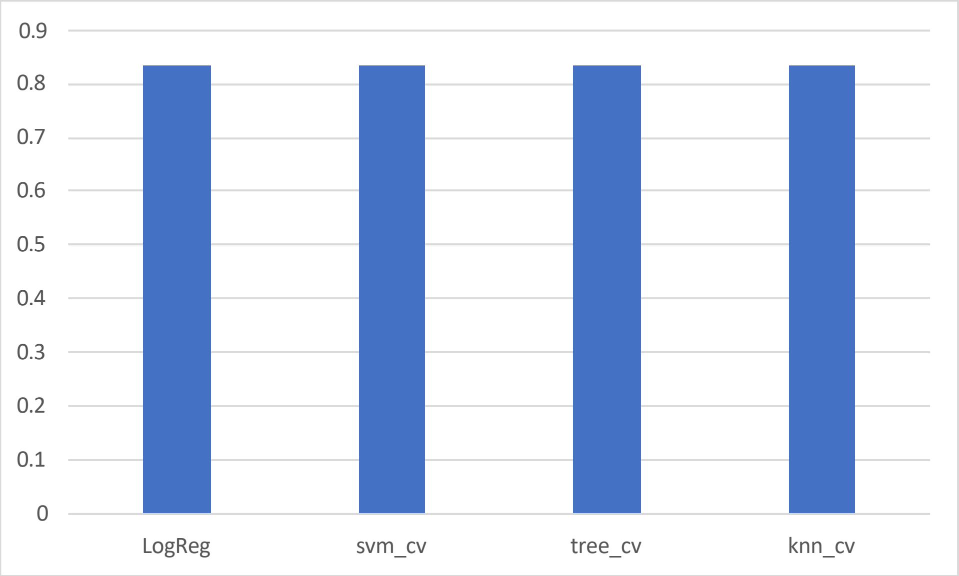
Section 5

Predictive Analysis (Classification)

Classification Accuracy

NO CLEAR MODEL OUTPERFORMED THE OTHERS

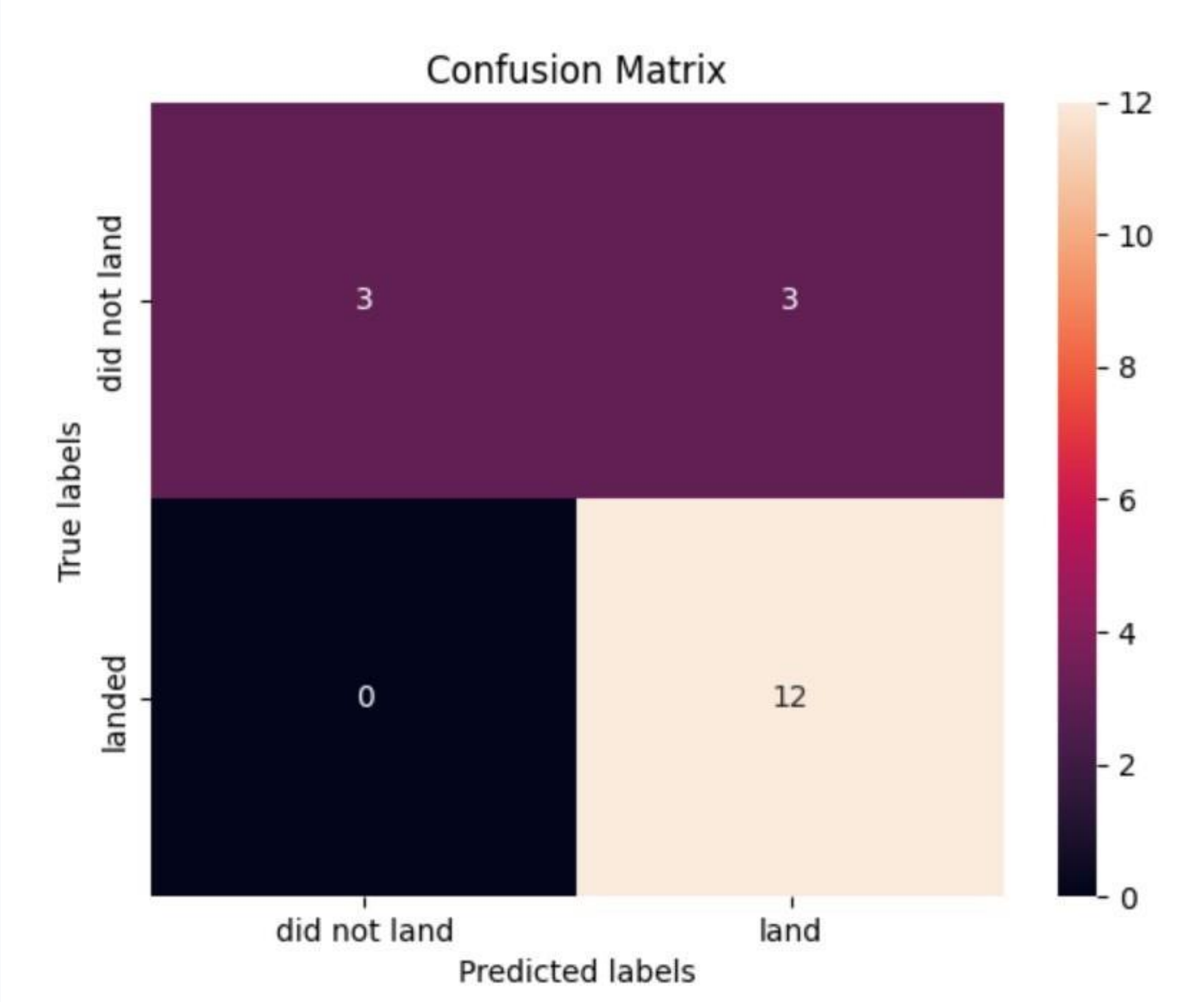
- Most likely due to small sample size (18)



	ML Method	Accuracy Score (%)
0	Support Vector Machine	83.333333
1	Logistic Regression	83.333333
2	K Nearest Neighbour	83.333333
3	Decision Tree	83.333333

Confusion Matrix

- All the confusion matrices were identical
- False positives present in matrix
  - 12 True positive
  - 3 True negative
  - 3 False positive
  - 0 False Negative



Conclusions

Model Performance: The models performed similarly on the test set with the decision tree model slightly outperforming

Equator: Most of the launch sites are near the equator for an additional natural boost - due to the rotational speed of earth - which helps save the cost of putting in extra fuel and boosters

Coast: All the launch sites are close to the coast

Launch Success: Increases over time

KSC LC-39A: Has the highest success rate among launch sites. Has a 100% success rate for launches less than 5,500 kg

Orbits: ES-L1, GEO, HEO, and SSO have a 100% success rate

Payload Mass: Across all launch sites, the higher the payload mass (kg), the higher the success rate

Appendix

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

Thank you!