Launch With Data Science

<Jamal Busby>
<18-5-2023>

Outline

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

Executive Summary

- Summary of methodologies
  - Data collection (API and Web Scraping)
  - Data Wrangling
  - Exploratory Data Analysis (SQL & Data Visualization)
  - Interactive Visual maps with Folium
  - Machine Learning Prediction
- Summary of all results
  - Determine what correlation exist between launch parameters and success or failure of rocket first stage landing
  - Improving success rate over time
  - While accuracy scores were similar between models, decision tree had the highest score.

Introduction

- As humans move into the realm of space equipment and personnel will need to be transported in the most cost effective way.
- Being able to predict a successful booster outcome will help reduce operational costs
- The following steps are used to complete this project:
  - Data collection, wrangling, and formatting
  - Exploratory data analysis
  - Interactive data visualization
  - Machine learning prediction
- The main items we seek to resolve:
  - Determine what correlation exist between launch parameters and success or failure of rocket first stage landing
  - Which ML algorithm will provide the most accurate prediction of a successful landing

Methodology

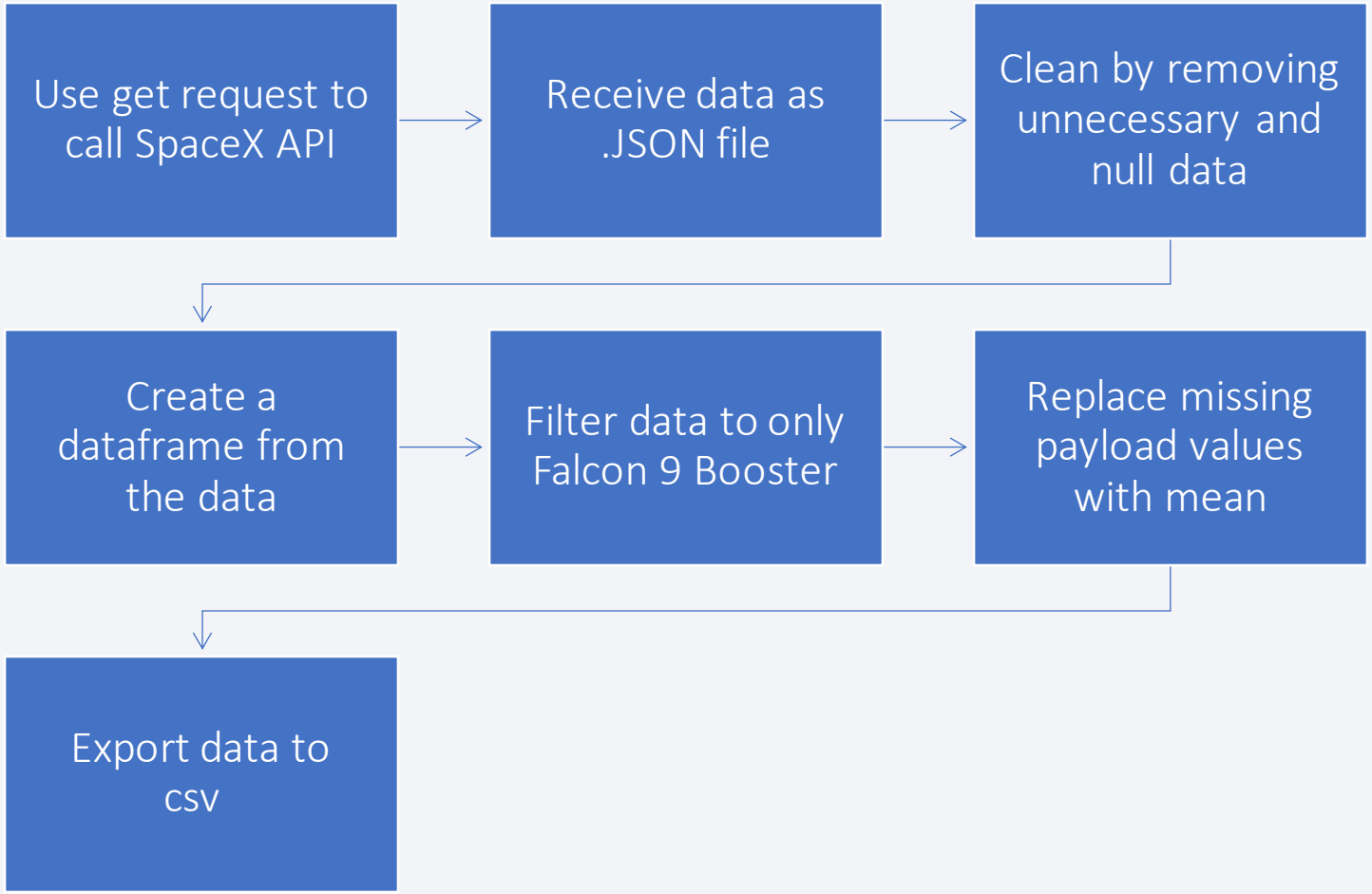
Methodology

- Executive Summary
- Data collection methodology:
  - Data was collected through SpaceX API and Web scraping of Wikipedia data
  - Data from the SpaceX API(https://api.spacexdata.com/v4/launches/past) was placed into a dataframe
  - Beautiful Soup was used to take data from Wikipedia and convert into dataframe
- Perform data wrangling
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

Data Collection

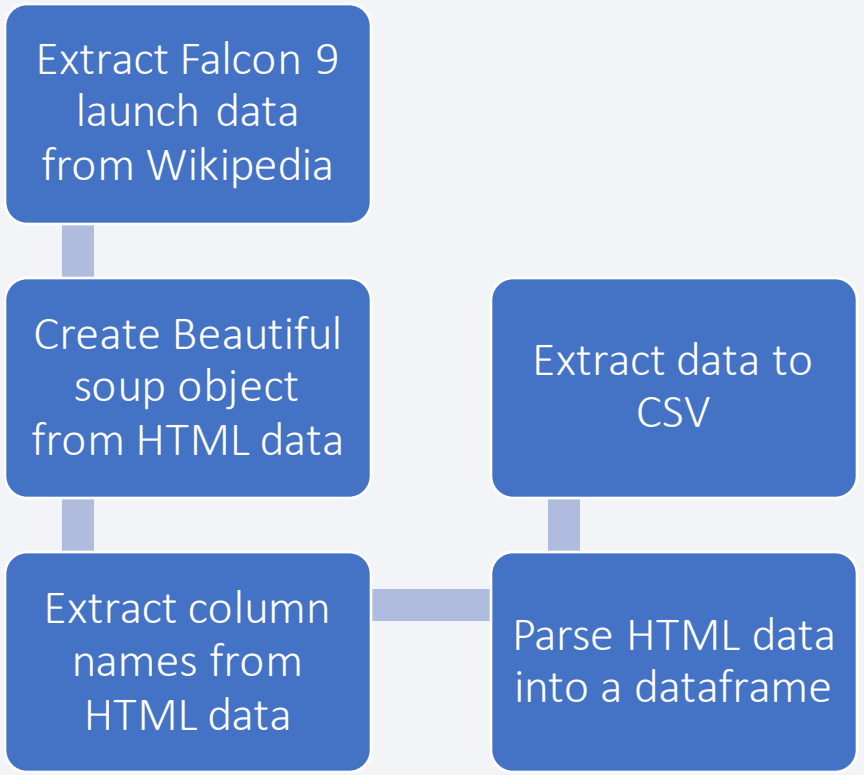
- Data was collected through the SpaceX API, and from scraping SpaceX data from Wikipedia.com
- Once data is retrieved it must be cleaned and made ready to use by removing unnecessary data and resolving missing and null values.
- Once data has been cleaned it is transformed into a dataframe to allow further wrangling of the data

Data Collection – SpaceX API



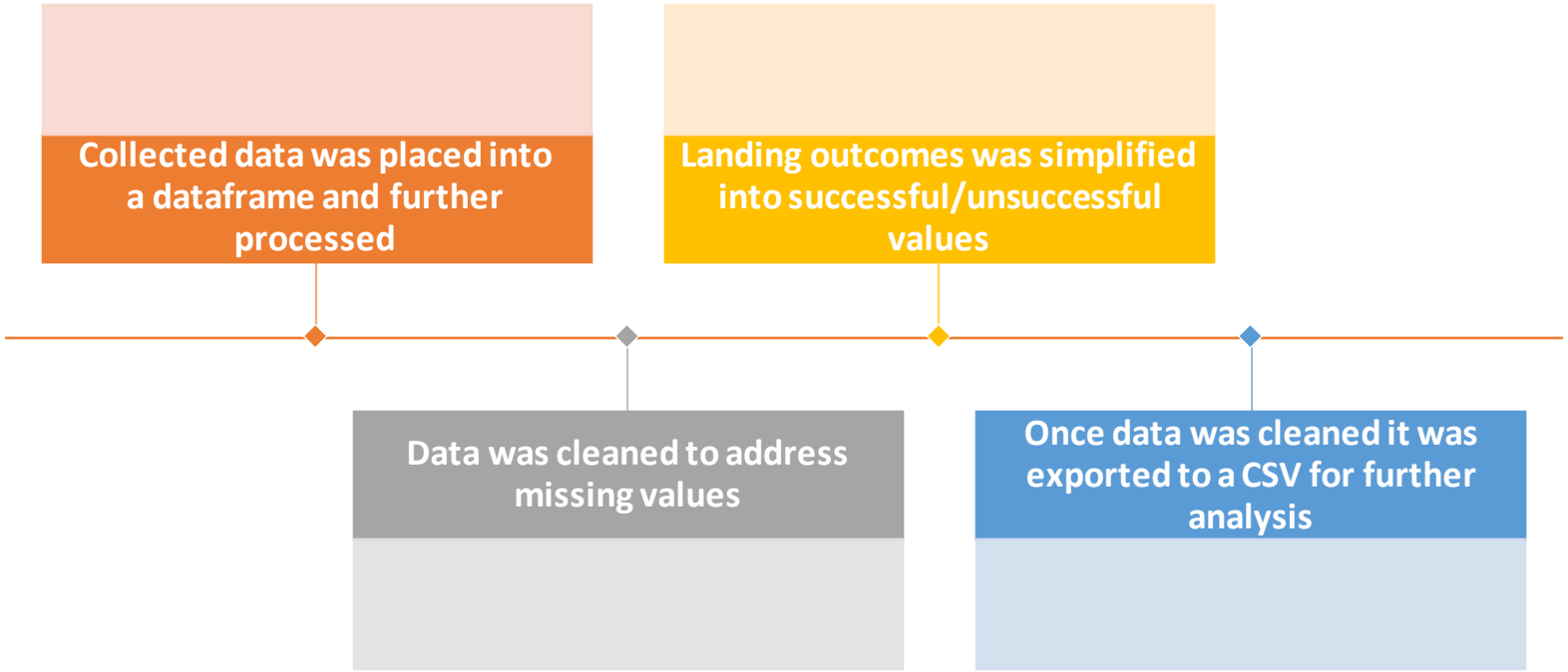
- Data collection Github link

Data Collection - Scraping



- Webscraping Github link

Data Wrangling



- Data Wrangling Github link

EDA with Data Visualization

- Charts plotted included:

Flight number vs. Payload Mass, Flight Number vs. Launch Site, Payload Mass vs. Launch site, Orbit Type vs. Success Rate, Flight Number vs. Orbit Type, Payload Mass vs. Orbit Type, and Success Rate Yearly Trend

- Scatter plots, Bar Charts, and Line Charts are used for the visualization of data

EDA Visualization Github link

EDA with SQL

```
%sql  SELECT DISTINCT LAUNCH_SITE  FROM SPACEXTBL;
```

- Unique Launch sites

```
%sql SELECT * FROM SPACEXTBL WHERE LAUNCH_SITE LIKE 'CCA%' LIMIT 5;
```

- Launch Sites starting with "CCA"

```
%sql SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE CUSTOMER = 'NASA (CRS)';
```

- Total Payload mass Launched by client NASA

```
%sql SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE BOOSTER_VERSION LIKE 'F9 V1.1';
```

- Average Payload Mass by F9v1.1 booster

```
%sql SELECT min(substr(Date,7,4) || substr(Date,4,2) || substr(Date,1,2)) from SPACEXTBL where  "Landing _Outcome" = 'Success (ground pad)';
```

- First successful landing on ground pad

```
%sql SELECT booster_version,PAYLOAD_MASS__KG_ ,"Landing _Outcome" from SPACEXTBL where  "Landing _Outcome" ='Success (drone ship)' and PAYLOAD_MASS__KG_ >4000 and PAYLOAD_MASS__KG_ < 6000
```

- Successful Drone ship landings with payload between 4000 and 6000 kg

```
%sql select count('MISSION_OUTCOME') from SPACEXTBL WHERE MISSION_OUTCOME LIKE '%SUCCESS%' OR  '%FAILURE%';
```

- Total number of successful and failure outcomes

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

- Boosters that have carried the maximum payload

```
%sql SELECT substr(Date, 4, 2) as month,booster_version,"Landing _Outcome" from SPACEXTBL where "Landing _Outcome" ='Failure (drone ship)' and substr(Date,7,4)='2015'
```

- Drone ship failure outcomes in 2015

```
%sql SELECT "Landing _Outcome", COUNT(*) as 'Quantity' FROM SPACEXTBL WHERE DATE BETWEEN '04-06-2010' AND '20-03-2017' GROUP BY "Landing _Outcome" ORDER BY "Quantity"
```

- Descending ordered rank count of landing outcomes from 04-06-2010 thru 20-03-2017

EDA SQL Github link

Build an Interactive Map with Folium

- Folium map was created to show locations of launches, markers were color coded to show successful and unsuccessful launches
- Line was drawn to show distance from launch pad to coast

Folium Map Github link

Build a Dashboard with Plotly Dash

- Summarize what plots/graphs and interactions you have added to a dashboard
- Using Plotly Dash and interactive dashboard was created
- Items included in the dashboard show the relationship between certain parameters such as; Outcomes, Payload Mass, Flight number, and Launch site
- Pie Charts and Scatter plots were used to provide the best visualization of data

Plotly Dash app Github link

Predictive Analysis (Classification)

- Summarize how you built, evaluated, improved, and found the best performing classification model
- Using numpy, pandas, and seaborn data was placed into an array and separated into training and test data sets
- Created SVM(Support Vector Machines), Classification Trees, KNN(K-Nearest Neighbor) and Logistic Regression models to find out which one was the most accurate
- Classification Tree Model obtained the highest accuracy score

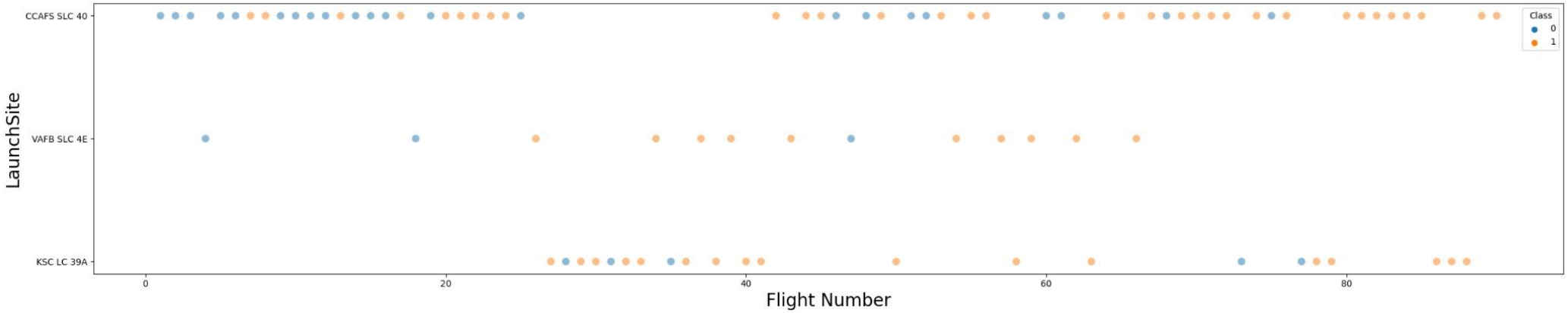
Machine Learning Prediction Github link

Results

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

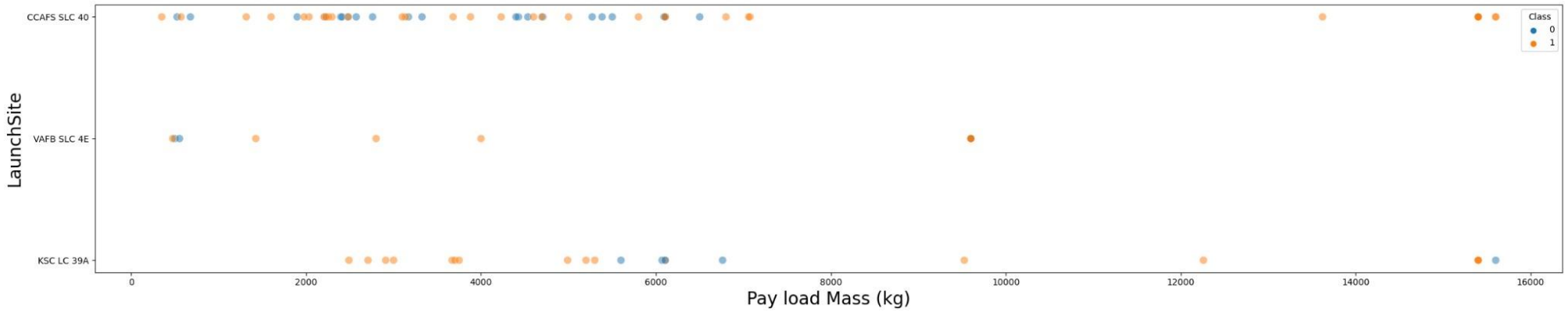
Section 2

Insight from EDA



Flight Number vs. Launch Site

- The greater number of flights has translated to a higher success rate

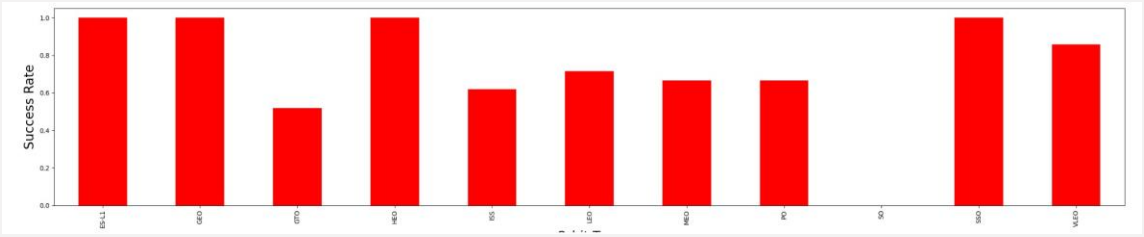


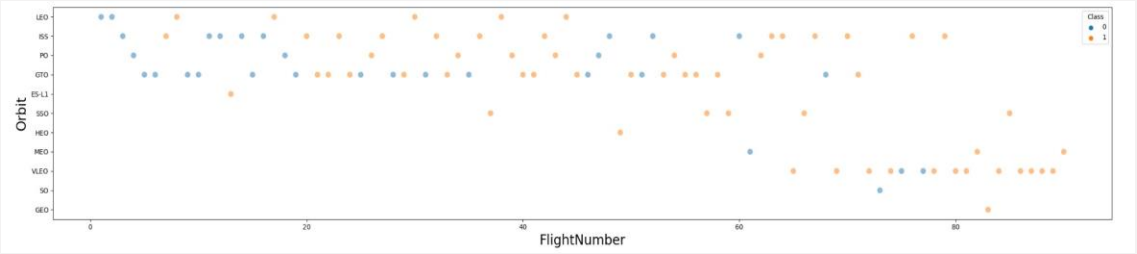
Payload vs. Launch Site

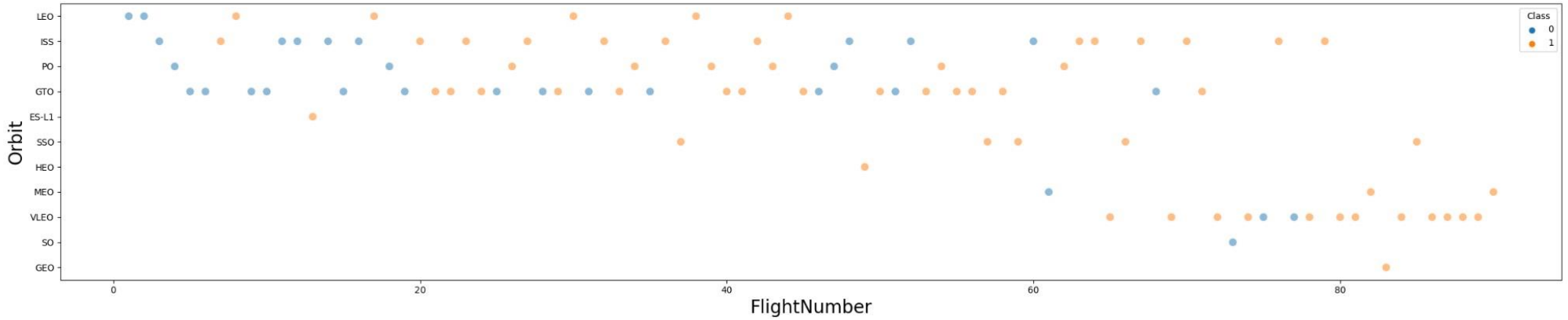
- CCAFS SLC 40 has the greatest number large payload successes

Success Rate vs. Orbit Type

- GEO, HEO, SSO, and ES-L1 have the highest success rates based on Orbit type

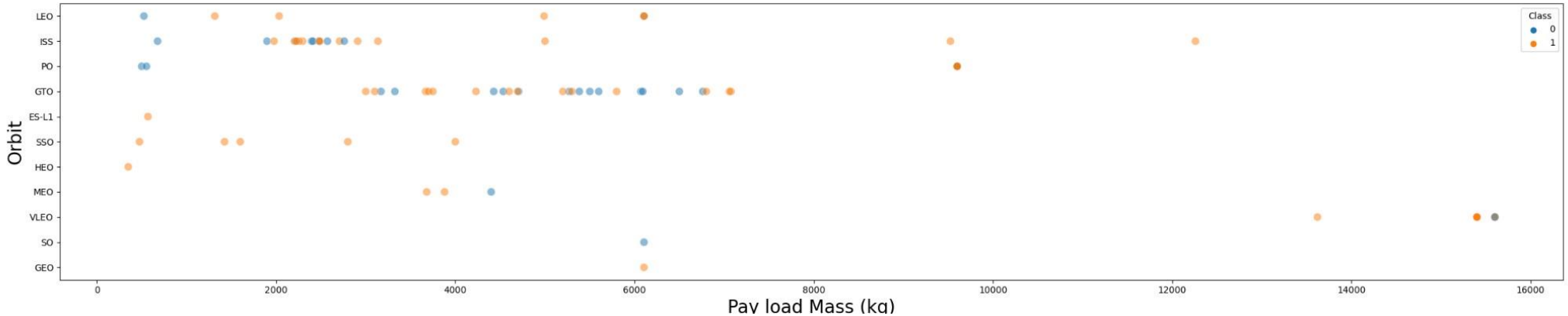






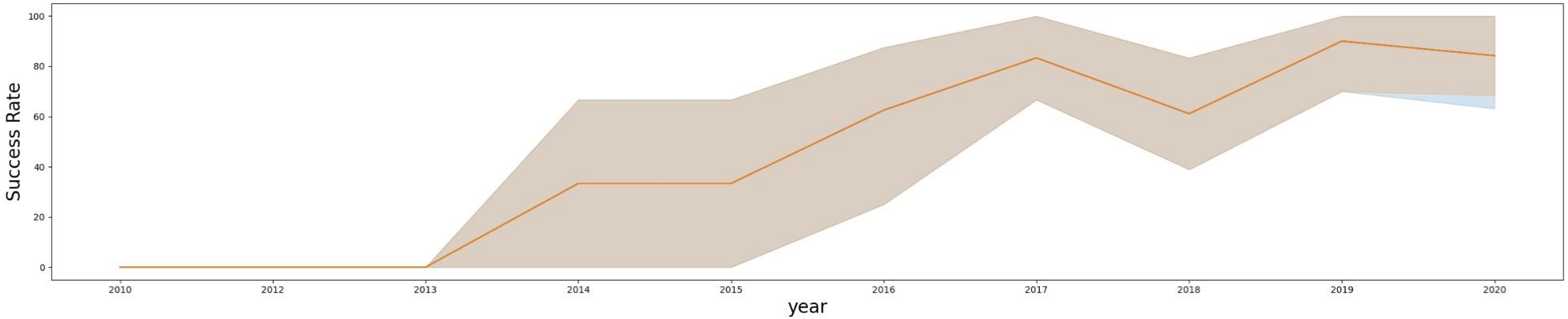
Flight Number vs. Orbit Type

- When viewed based on orbit type the success rate based on the number of flights generally increases



Payload vs. Orbit Type

- scatter point of payload vs. orbit type
- Looking at Success using the parameters Payload and Orbit type there are no clear best as orbits with highest success rate also have a small sample size



Launch Success Yearly Trend

- This line chart show the yearly trend of success
- The average success rate has grown year over year

All Launch Site Names

- There are 4 launch sites being used for launches for SpaceX
- Three sites are located in Fl which are CCAFS LC-40, CCAFS SLC-40, and KSC LC-39A
- One site is located in CA which is VAFB SLC-4E

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

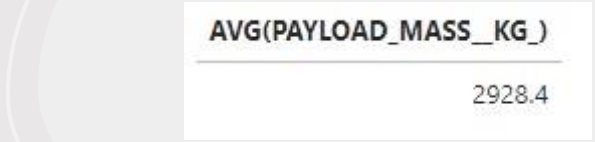
Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome
04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
08-12-2010	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
22-05-2012	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
01-03-2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`
- The records show that although there is a high mission success rate, the landing outcomes failure rate is high

Total Payload Mass

- Total mass of all payloads where NASA is the customer



AVG(PAYLOAD_MASS__KG_)
2928.4

Average Payload Mass by F9 v1.1

- The F9 v1.1 booster carried an average payload of 2928.4 kg

min(substr(Date,7,4) || substr(Date,4,2) || substr(Date,1,2))
20151222

First Successful Ground Landing Date

- Date of the first successful landing outcome on ground pad

Booster_Version	PAYLOAD_MASS__KG_	Landing _Outcome
F9 FT B1022	4696	Success (drone ship)
F9 FT B1026	4600	Success (drone ship)
F9 FT B1021.2	5300	Success (drone ship)
F9 FT B1031.2	5200	Success (drone ship)

Successful Drone Ship Landing with Payload between 4000 and 6000

- List the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

count('MISSION_OUTCOME')
100

Total Number of Successful and Failure Mission Outcomes

- 100 missions of successful or failure mission outcome

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

Boosters Carried Maximum Payload

- The F9 B5 Boosters have carried the Maximum payload

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

2015 Launch Records

- Drone ship failed landing outcomes listing site, Booster version, and month for the year 2015

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

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

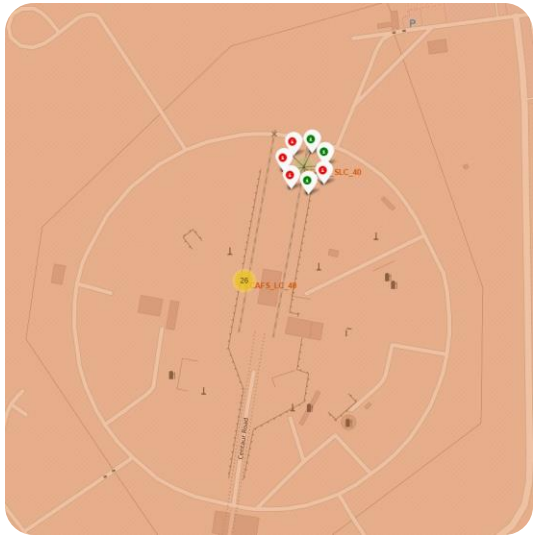
- Ascending order count of all landing outcomes between the above listed range of dates

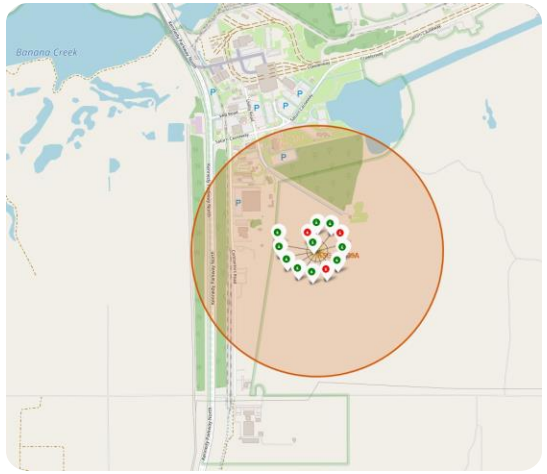
Section 3

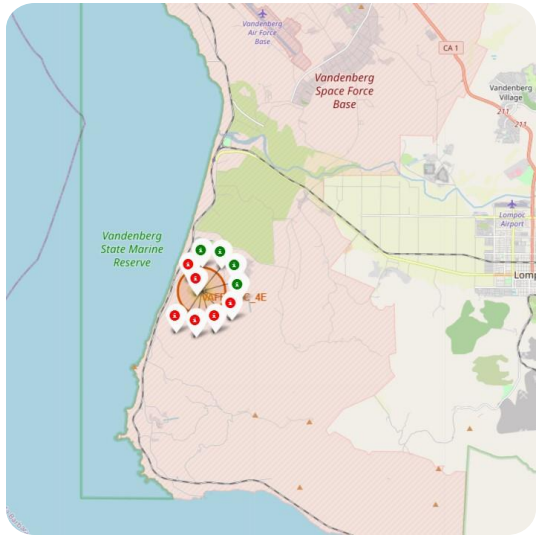
Launch Sites Proximities Analysis

Global Launch site map

- Folium map showing Falcon 9 launches
- 4 launch sites with 3 located in Florida and 1 in California
- Majority of launches and successful landings occurred in Florida





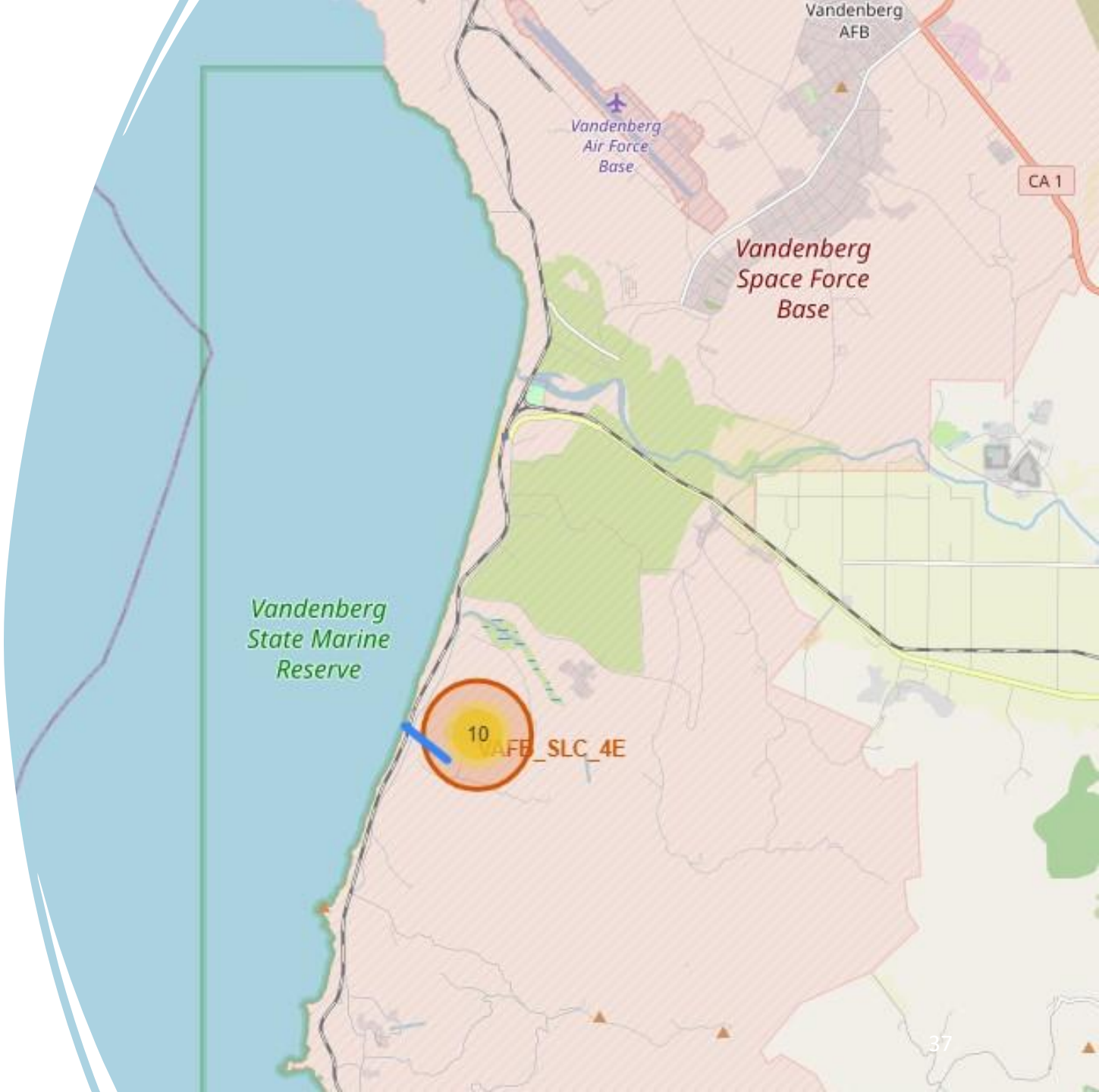


Folium Launch Outcomes

- Color labels show sites success(Green) and failure (Red) launches
- Comparing between the sites KSC LC-39A has the highest success rate

Distance from Coastline

- All launch sites are relatively close to a coastline
- VAFB SLC-4E is located approximately 1 km from coastline
- Launching from sites that are isolated from developments and close to coastline reduces risk by allowing flight path to go over the ocean

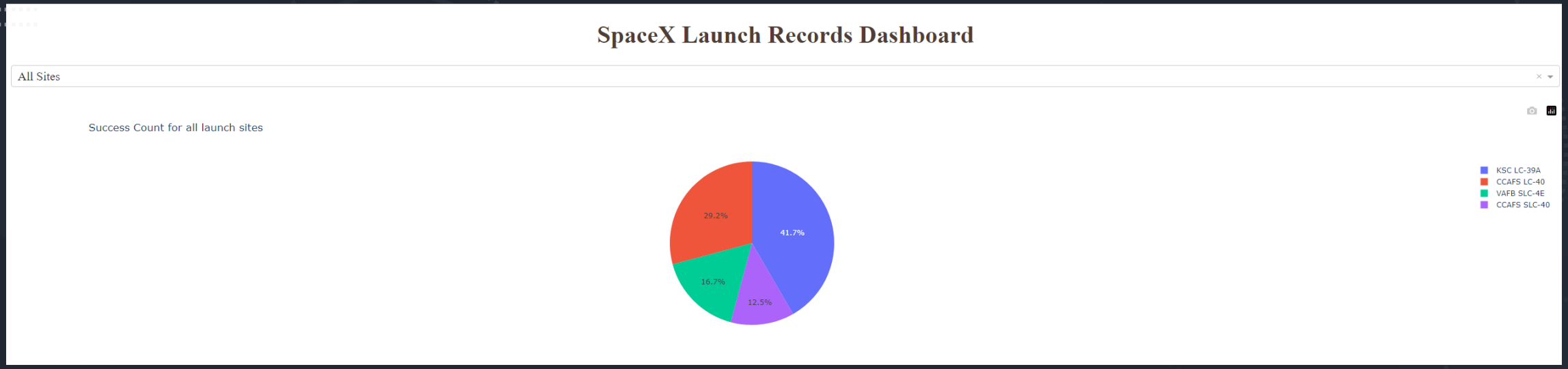


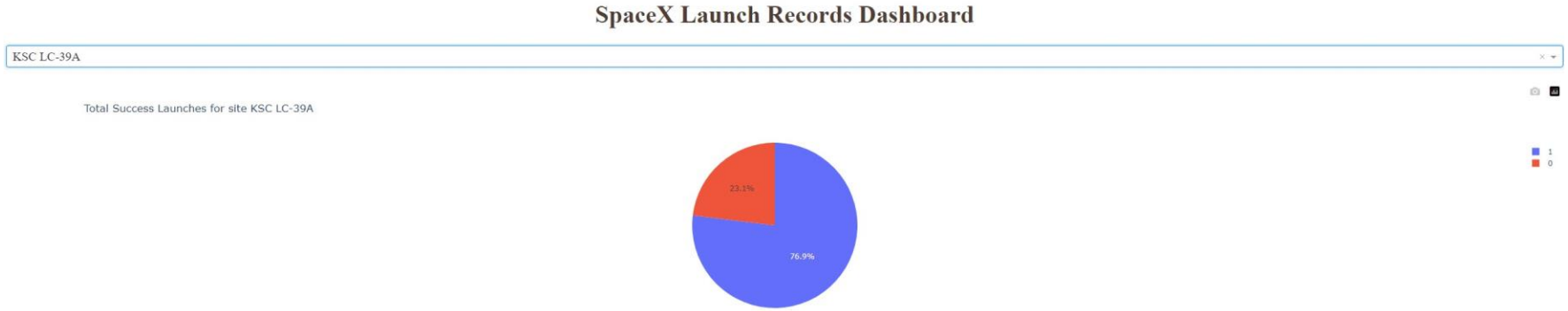
Section 4

Build a Dashboard with Plotly Dash

SpaceX Launch Success Count

- Launch success counts as a percentage of all launches
- Highest number of successful launches occurred at KSC LC-39A with 41.7% of successful launches

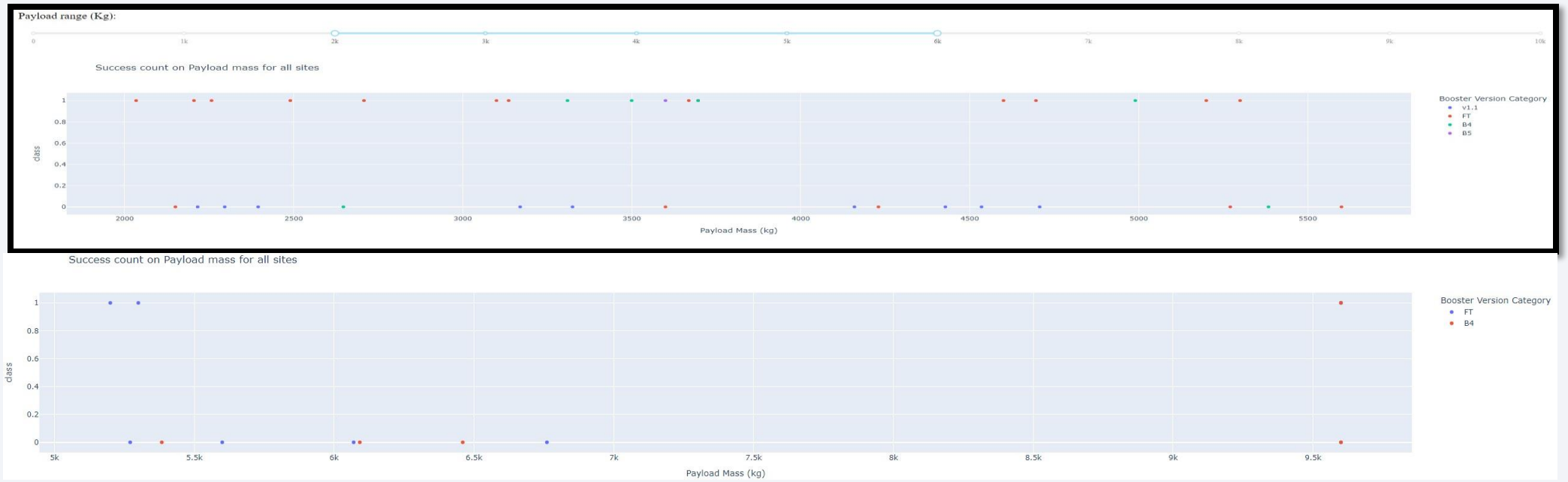




Most Successful Landings

- KSC LC-39A had the highest success ratio with 76.9%
- KSC LC-39A had 10 successes compared to 3 failures

Launch Site Success by Payload Mass



Payload Mass between 2000 and 5500 kg have the highest success rate
Payload Mass between 5000 and 10000 kg have the lowest success rate

Section 5

Predictive Analysis (Classification)

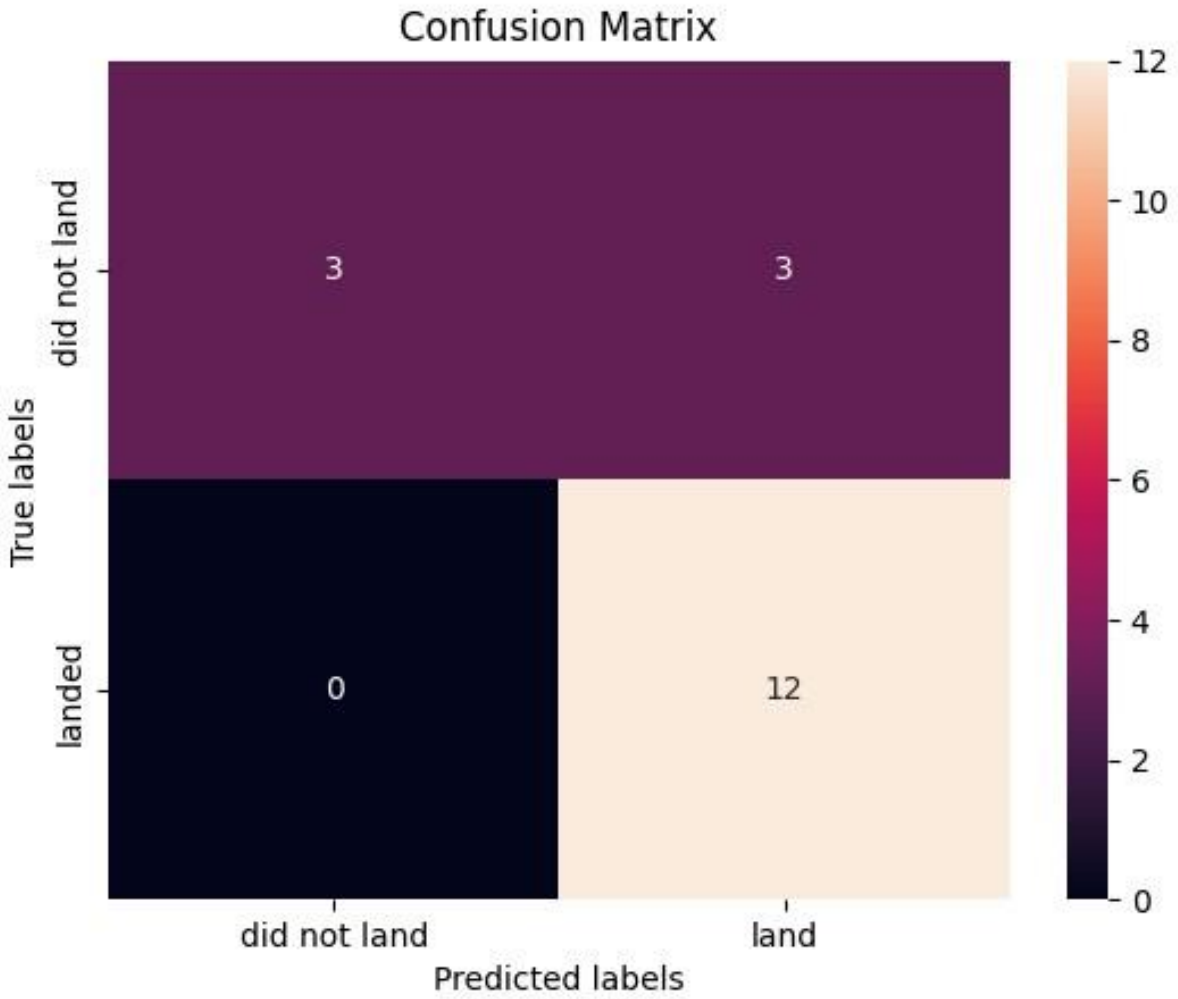
Classification Accuracy

- Decision Tree had the highest Accuracy score

	Best scores
Logistic regresssion	0.846429
SVM	0.848214
Decision tree	0.889286
KNN	0.848214

Confusion Matrix

- While the decision tree had the highest prediction score, all models suffered from a high rate of false positives.



Conclusions

- Decision Tree had the highest accuracy rate
- Greatest success comes with Payloads below 5500 kg
- Launches at KSC LC-39A had the highest success ratio with 76.9%
- Success rate has steadily improved over time
- Because of a high false positive rate additional data would be needed to create a higher accuracy model


Thank You!