IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Araik Tamazian
October 16, 2025

Outline

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

Executive Summary

- Methodology
  - Data Collection
  - Data Wrangling
  - Exploratory Data Analysis (EDA)
  - Building an Interactive Map
  - Building a Dashboard
- Results
  - Exploratory Data Analysis Results
  - Interactive Analytics Demo in Screenshots
  - Predictive Analysis Results

Introduction

- The project involves simulating the role of a data scientist at a startup, Space Y, aiming to compete with SpaceX in the rocket launch market.
- The main objective is to predict the price of rocket launches and whether SpaceX will reuse the first stage of its Falcon 9 rocket.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - SpaceX REST API + Web scraping from Wikipedia
- Perform data wrangling
  - Describe how data was processed
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Split the data to train and test set + fit 4 models with usage of grid search to find best parameters + find the best performing model

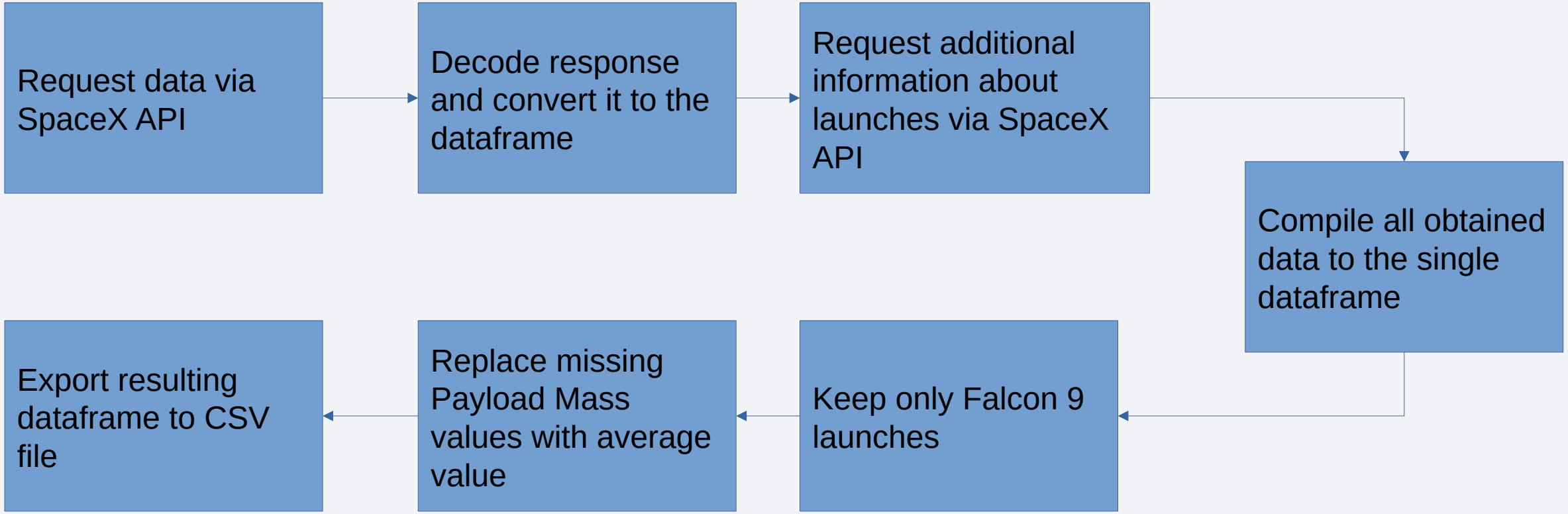
Data Collection

- A combination of SpaceX REST API requests was used to obtain first part of data
- The second part of data was obtained by using web scraping data from SpaceX's Wikipedia article (specific table data was gathered).
- The following data columns were obtained
  - By using SpaceX REST API:

    FlightNumber, Date, BoosterVersion, PayloadMass, Orbit, LaunchSite, Outcome, Flights, GridFins, Reused, Legs, LandingPad, Block, ReusedCount, Serial, Longitude, Latitude
  - By web scraping Wikipedia's SpaceX article:

    Flight No., Launch site, Payload, PayloadMass, Orbit, Customer, Launch outcome, Version Booster, Booster landing, Date, Time

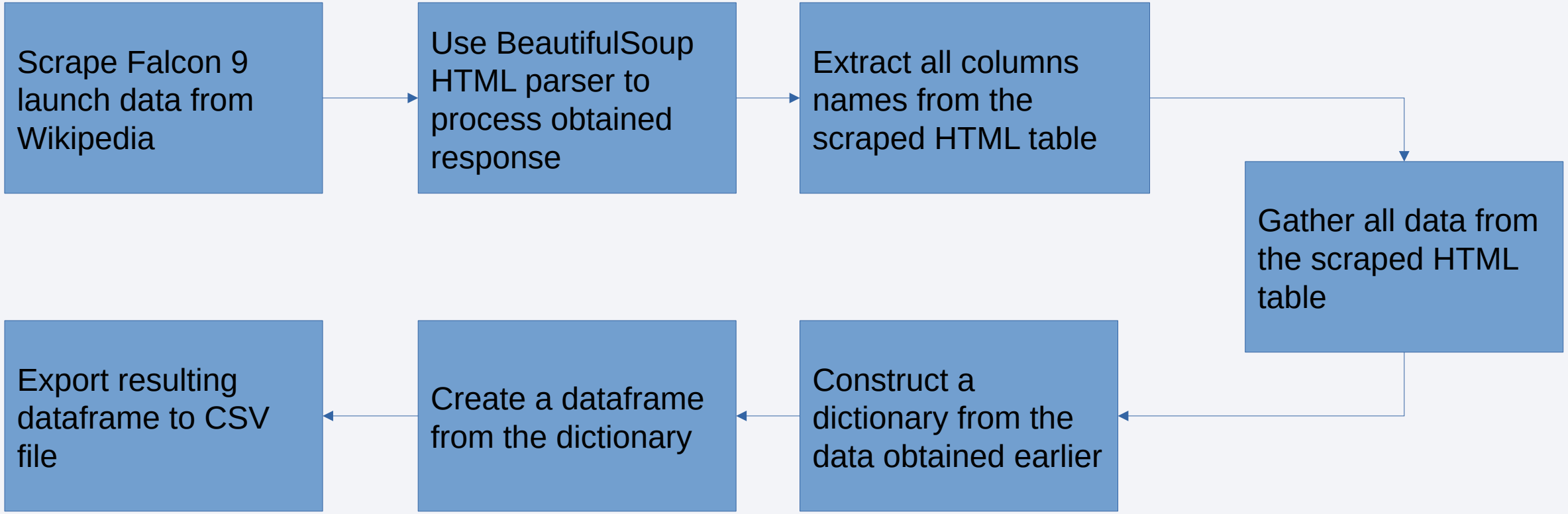
Data Collection – SpaceX API



Link to notebook:

- https://github.com/atamazian/ds-capstone/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

Data Collection - Scraping



Link to notebook:

- https://github.com/atamazian/ds-capstone/blob/main/jupyter-labs-webscraping.ipynb

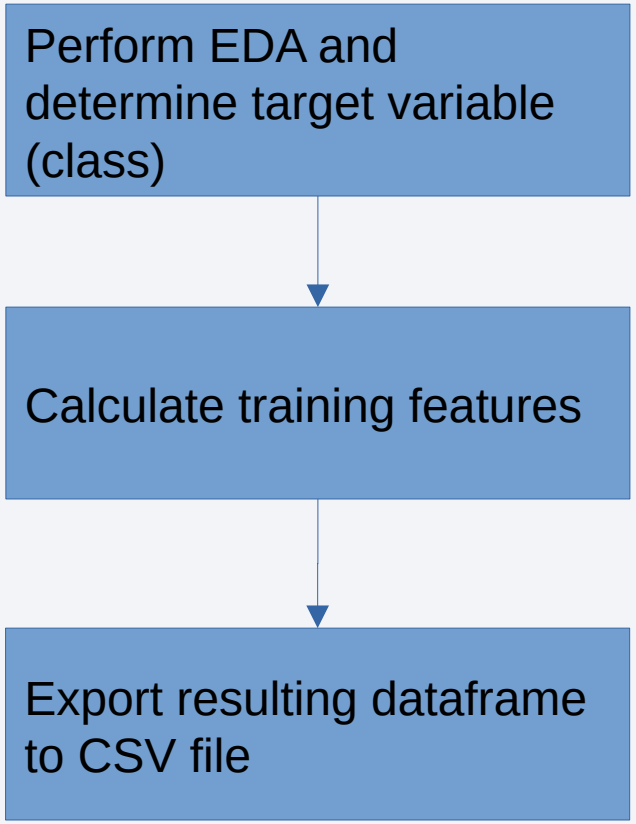
Data Wrangling

- Key Attributes of SpaceX Launch Data

Important attributes include Flight Number, Date, Booster Version, Payload Mass, Orbit, Launch Site, and Outcome, which indicate the status of the first stage of the rocket.

- Launch Outcomes (target variable)

The outcomes are converted into binary classes for analysis, where 0 represents a failed landing and 1 represents a successful landing.



Link to notebook:

- https://github.com/atamazian/ds-capstone/blob/main/labs-jupyter-spacex-Data wrangling.ipynb

EDA with Data Visualization

During the EDA the following plots were produced:

- Flight Number vs. Payload Mass (scatter plot)
- Flight Number vs. Launch Site (scatter plot)
- Payload Mass vs. Launch Site (bar chart)
- Orbit Type vs. Success Rate (scatter plot)
- Flight Number vs. Orbit Type (scatter plot)
- Payload Mass vs Orbit Type (scatter plot)
- Success Rate Yearly Trend (like plot)

Link to notebook:

- https://github.com/atamazian/ds-capstone/blob/main/edadataviz.ipynb

EDA with SQL

Performed SQL queries:

- Print names of the unique launch sites
- Print top 5 records where launch site begin with the string ‘CCA'
- Print the total payload mass carried by boosters launched by NASA (CRS)
- Print average payload mass carried by booster version F9 v1.1
- Print the date when the first successful landing outcome in ground pad was achieved
- Print names of the boosters which have success in drone ship and have payload mass between 4000 and 6000
- Print the total number of successful and failure mission outcomes
- Print names of the booster versions which have carried the maximum payload mass
- Print failed landing outcomes in drone ship, their booster versions and launch site names for all months in 2015
- Print the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20 in descending order

Link to notebook:

- https://github.com/atamazian/ds-capstone/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Added markers of all four launch sites.
- Added marker clusters consisting of colored markers (red for failure, green for success) for all launch sites.
- Plotted a distance from CCAFS LC-40 launch site to the coastal line.

Link to notebook:

- https://github.com/atamazian/ds-capstone/blob/main/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

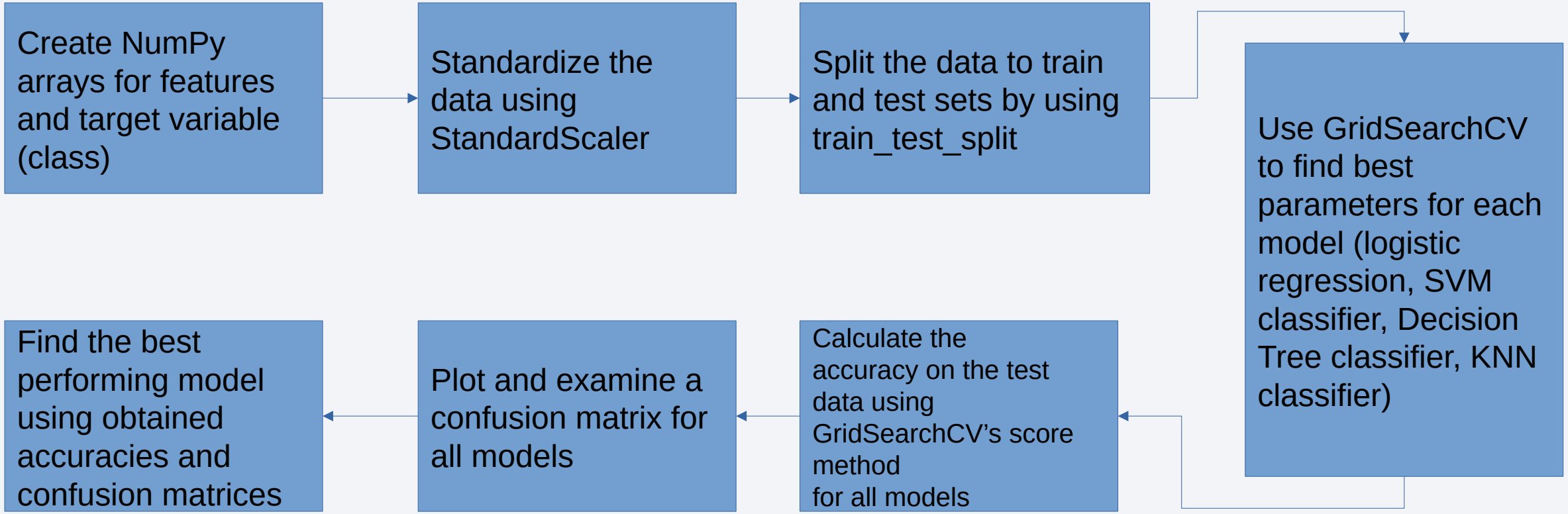
The interactive dashboard shows:

- Pie chart showing Launch Success Rate
  - Dropdown list allows to show either data from all launch sites, or data from the specific launch site.
- Scatter plot of Launch Outcome vs. Payload Mass
  - Dropdown list allows to show either data from all launch sites, or data from the specific launch site.
  - Added a slider to select Payload Mass range.

Link to code:

- https://github.com/atamazian/ds-capstone/blob/main/spacex-dash-app.py

Predictive Analysis (Classification)



Link to code:

- https://github.com/atamazian/ds-capstone/blob/main/SpaceX_Machine Learning Prediction_Part_5.ipynb

Results

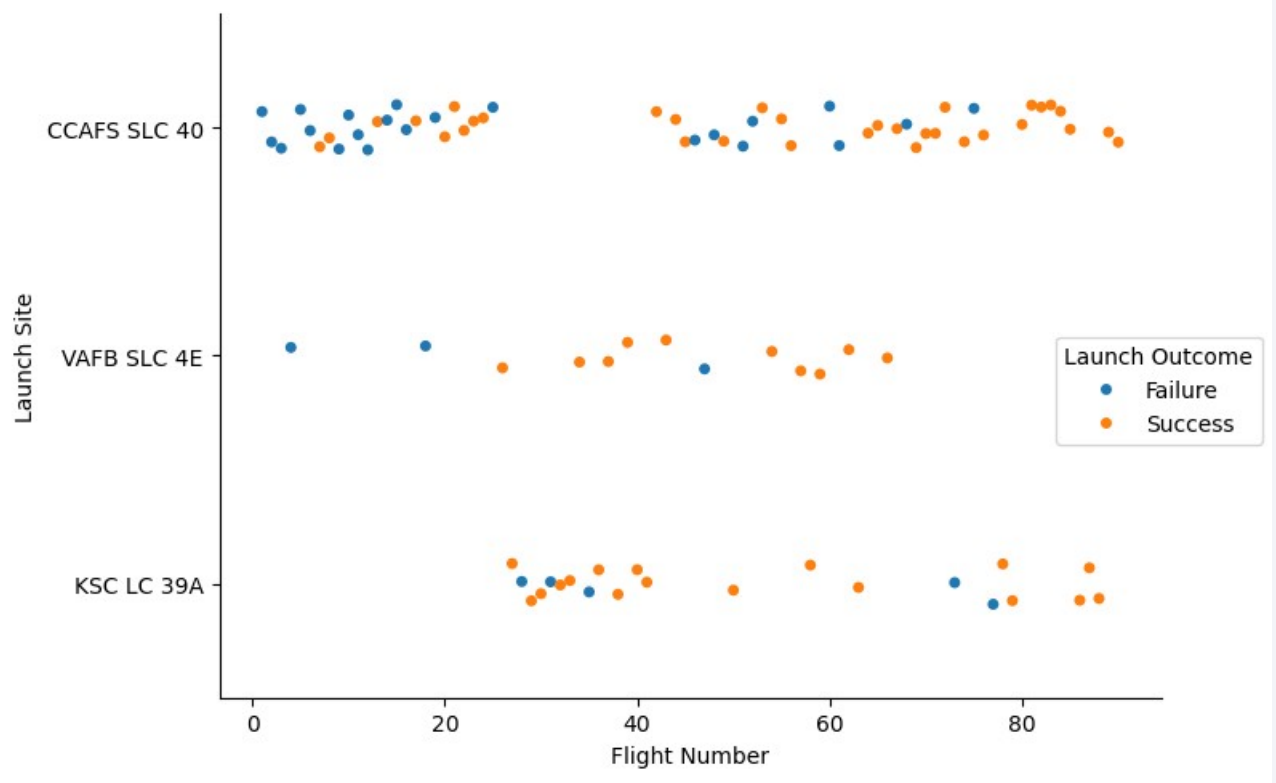
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

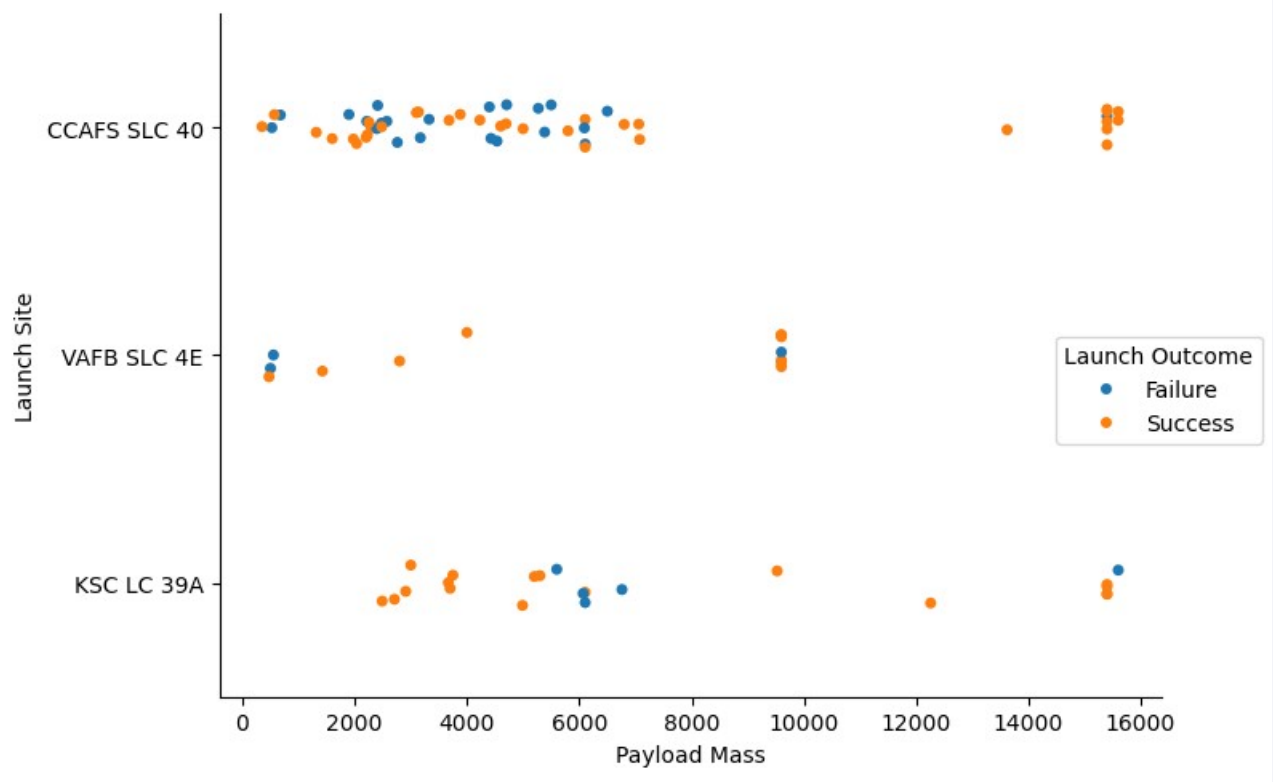
Flight Number vs. Launch Site

- The most early launches failed while the most of later launches succeeded.
- About a half of all launches were done from CCAFS SLC 40 site
- Launches from VAFB SLC 4E and KSC LC 39A sites have higher success rates.



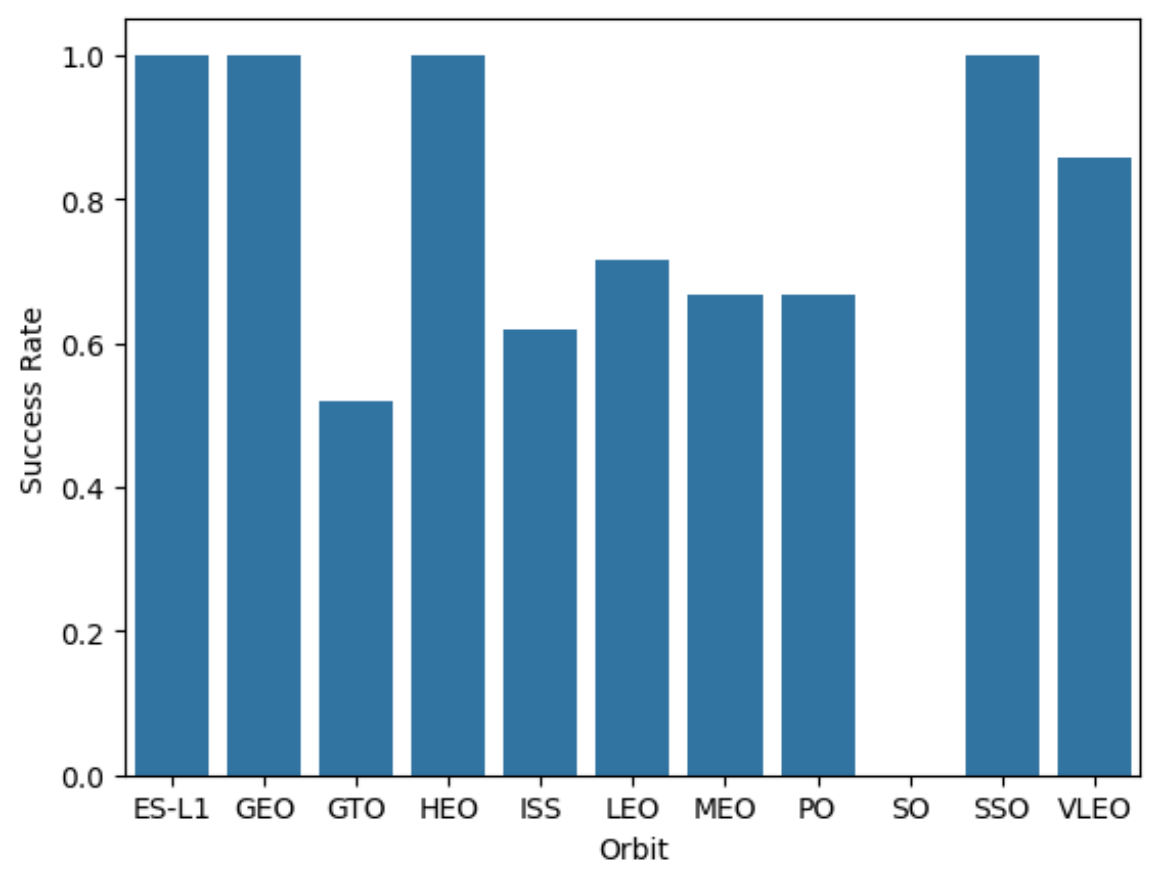
Payload vs. Launch Site

There are no rockets launched for heavy payload mass (greater than 10000) for the VAFB-SLC launch site.



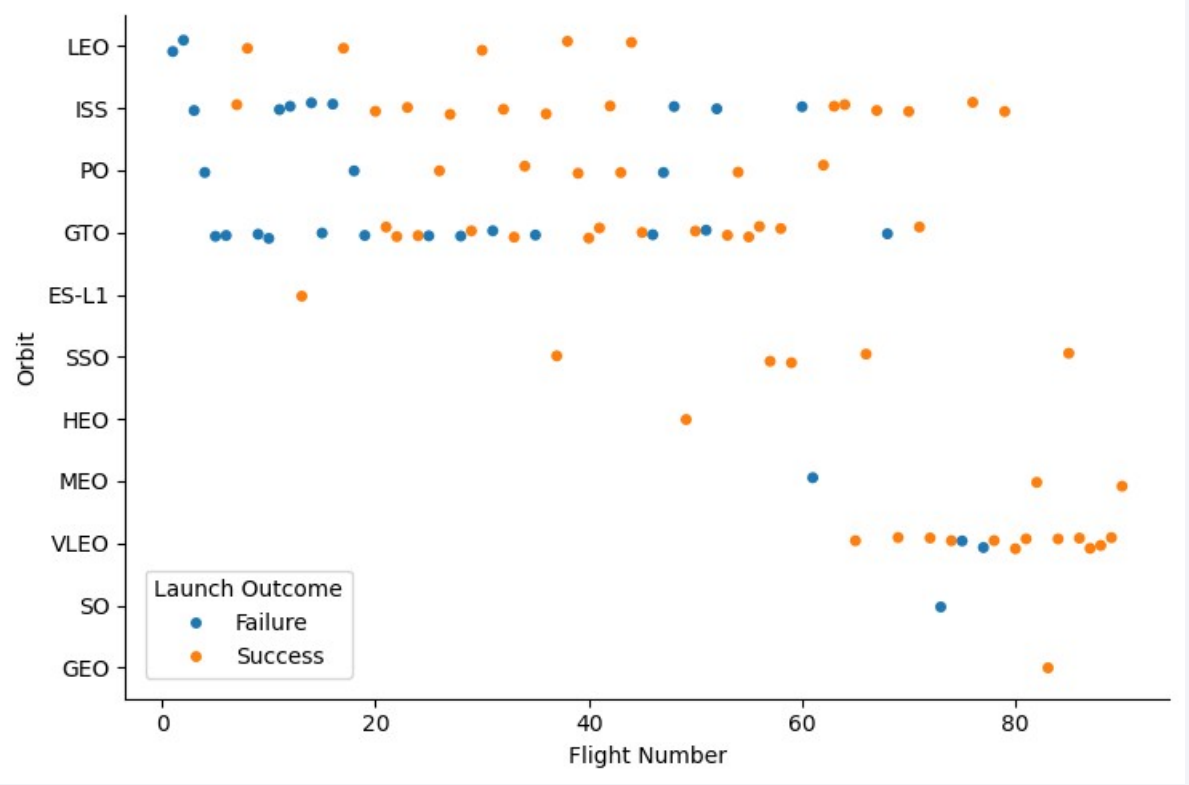
Success Rate vs. Orbit Type

- All launches to ES-L1, GEO, HEO and SSO orbits were succesfull
- All lauches to SO orbit had failed
- Launches to other orbits were partially successfull



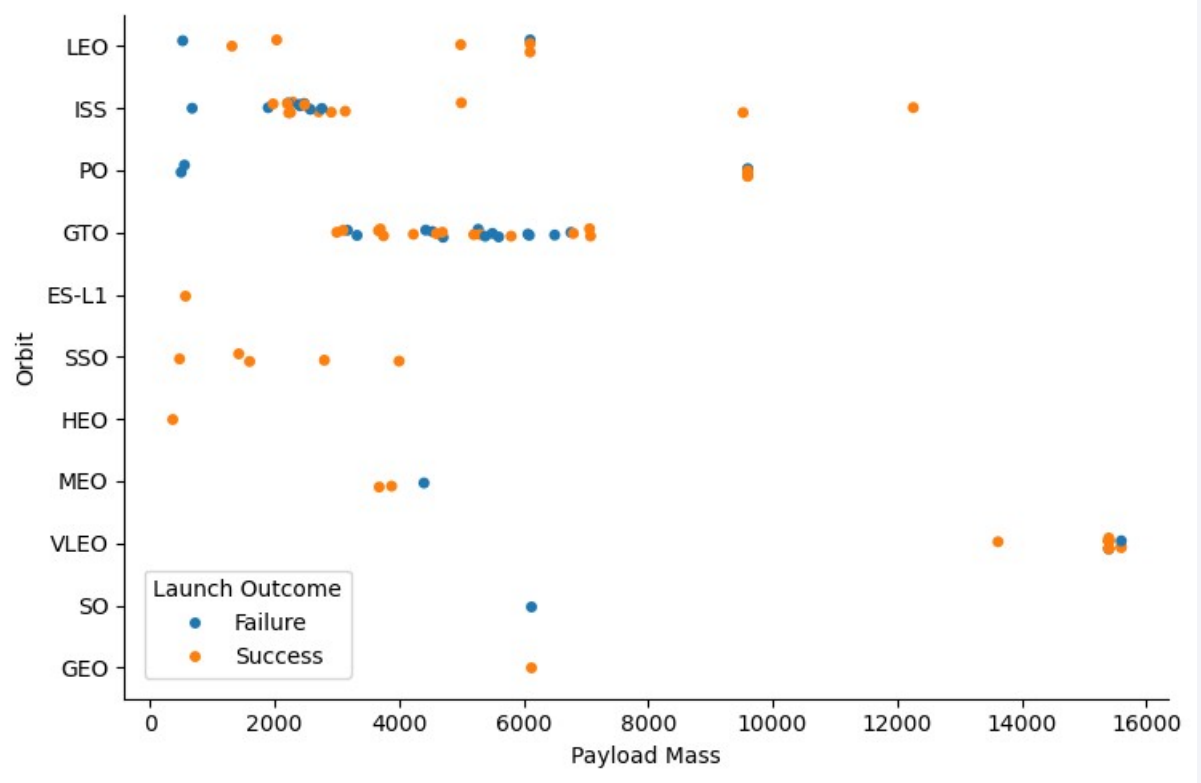
Flight Number vs. Orbit Type

- Success seems to be related to the number of flights in the LEO orbit.
- Conversely, there appears to be no relationship between flight number and success in the GTO orbit.



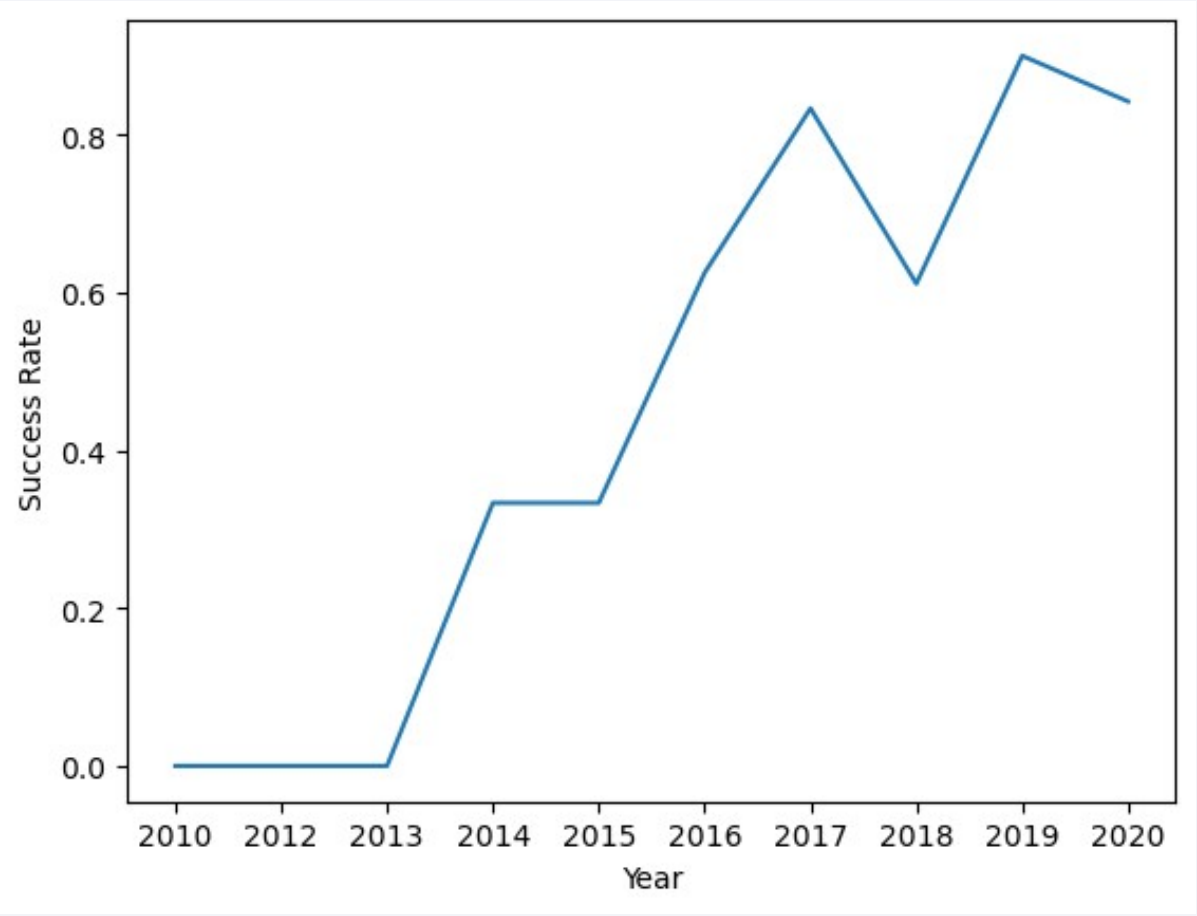
Payload vs. Orbit Type

- Launches carrying heavier payloads tend to have higher success and positive landing rates for Polar, LEO, and ISS launches.



Launch Success Yearly Trend

The launch success rate since 2013 kept increasing till 2020.



All Launch Site Names

Display the names of the unique launch sites in the space mission

```
%sql SELECT DISTINCT "Launch_Site" FROM SPACEXTBL
```

```
 * sqlite:///my_data1.db
Done.
```

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

Launch Site Names Begin with 'CCA'

Display 5 records where launch sites begin with the string 'CCA'

```
%sql SELECT * FROM SPACEXTBL WHERE "Launch_Site" LIKE "CCA%" LIMIT 5
```

```
 * sqlite:///my_data1.db
Done.
```

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_O
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (pa
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 (pa
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No

Total Payload Mass

Display the total payload mass carried by boosters launched by NASA (CRS)

```
%sql SELECT SUM("PAYLOAD_MASS__KG_") FROM SPACEXTBL WHERE Customer LIKE "NASA%"
```

```
 * sqlite:///my_data1.db
Done.
```

SUM("PAYLOAD_MASS__KG_")
99980

Average Payload Mass by F9 v1.1

Display average payload mass carried by booster version F9 v1.1

```
%sql SELECT AVG("PAYLOAD_MASS__KG_") FROM SPACEXTBL WHERE Booster_Version LIKE "F9 v1.1%"
```

```
 * sqlite:///my_data1.db
Done.
```

AVG("PAYLOAD_MASS__KG_")
2534.6666666666665

First Successful Ground Landing Date

List the date when the first succesful landing outcome in ground pad was achieved.

Hint:Use min function

```
%sql SELECT MIN(DATE) FROM SPACEXTBL WHERE Mission_Outcome="Success"
```

```
 * sqlite:///my_data1.db
Done.
```

MIN(DATE)
2010-06-04

Successful Drone Ship Landing with Payload 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

```
%%sql
SELECT Booster_Version
FROM SPACEXTBL
WHERE Landing_Outcome="Success (drone ship)" AND PAYLOAD_MASS__KG_ > 4000 AND PAYLOAD_MASS__KG_ < 6000
```

```
 * sqlite:///my_data1.db
Done.
```

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

Total Number of Successful and Failure Mission Outcomes

```
%sql SELECT Mission_Outcome, COUNT(*) FROM SPACEXTBL GROUP BY Mission_Outcome
```

```
 * sqlite:///my_data1.db
Done.
```

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

Boosters Carried Maximum Payload

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

```
 * sqlite:///my_data1.db
Done.
```

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

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.

Note: SQLLite does not support monthnames. So you need to use substr(Date, 6,2) as month to get the months and substr(Date,0,5)='2015' for year.

```
%%sql
SELECT SUBSTR(Date, 6, 2) AS MONTH, Landing_Outcome, Booster_Version, Launch_Site
FROM SPACEXTBL
WHERE SUBSTR(Date, 0, 5) = '2015' AND Landing_Outcome="Failure (drone ship)"
```

```
 * sqlite:///my_data1.db
Done.
```

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

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

Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order.

```
%%sql
SELECT Landing_Outcome, COUNT(*)
FROM SPACEXTBL
WHERE Date BETWEEN "2010-06-04" AND "2017-03-20"
GROUP BY Landing_Outcome
```

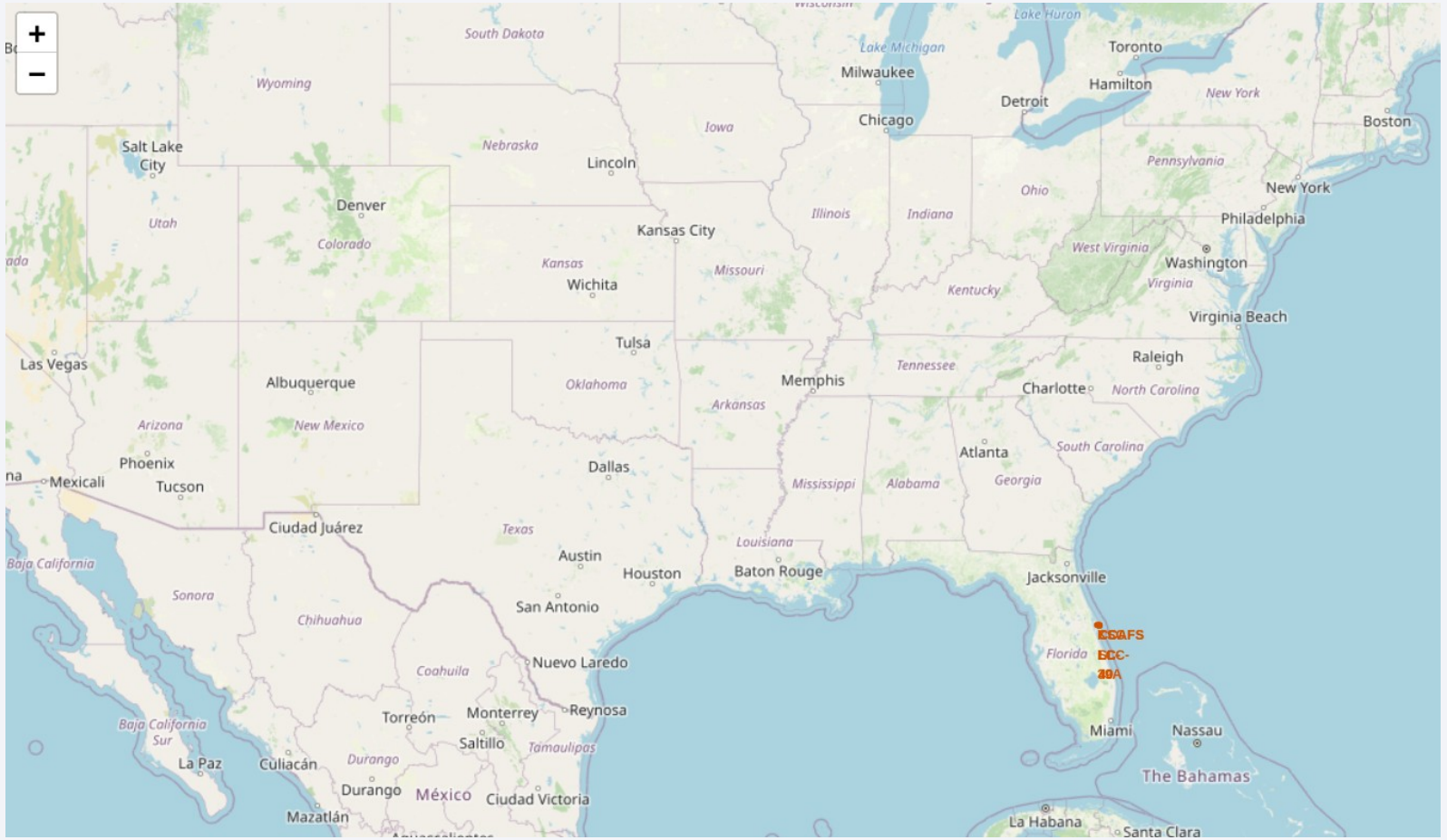
```
 * sqlite:///my_data1.db
Done.
```

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

Section 3

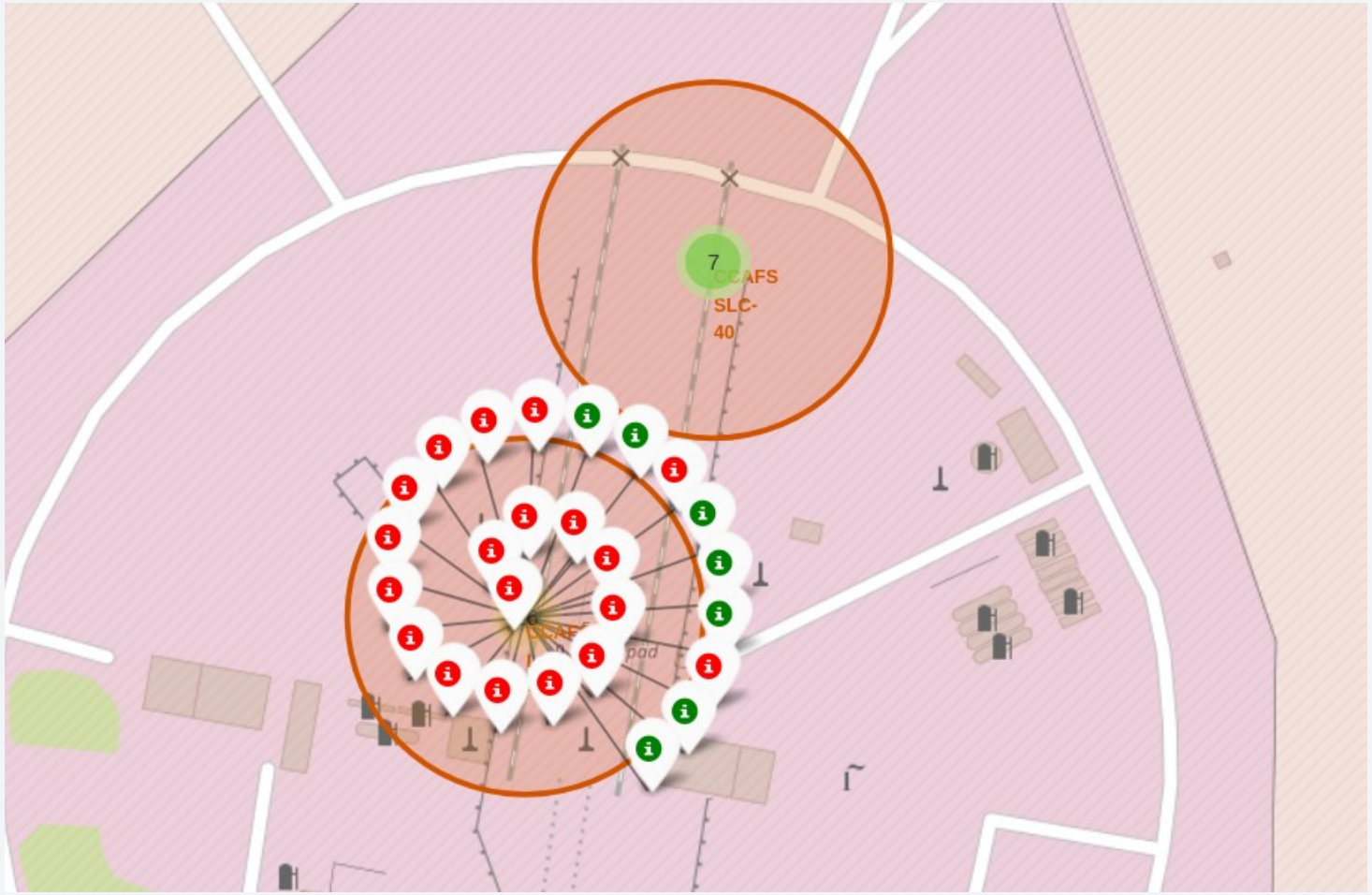
Launch Sites Proximities Analysis

Launch Sites on the Global Map



All launch sites in this study are located near the eastern coast of Florida.

Launch Outcomes for CCAFS LC-40 Launch Site

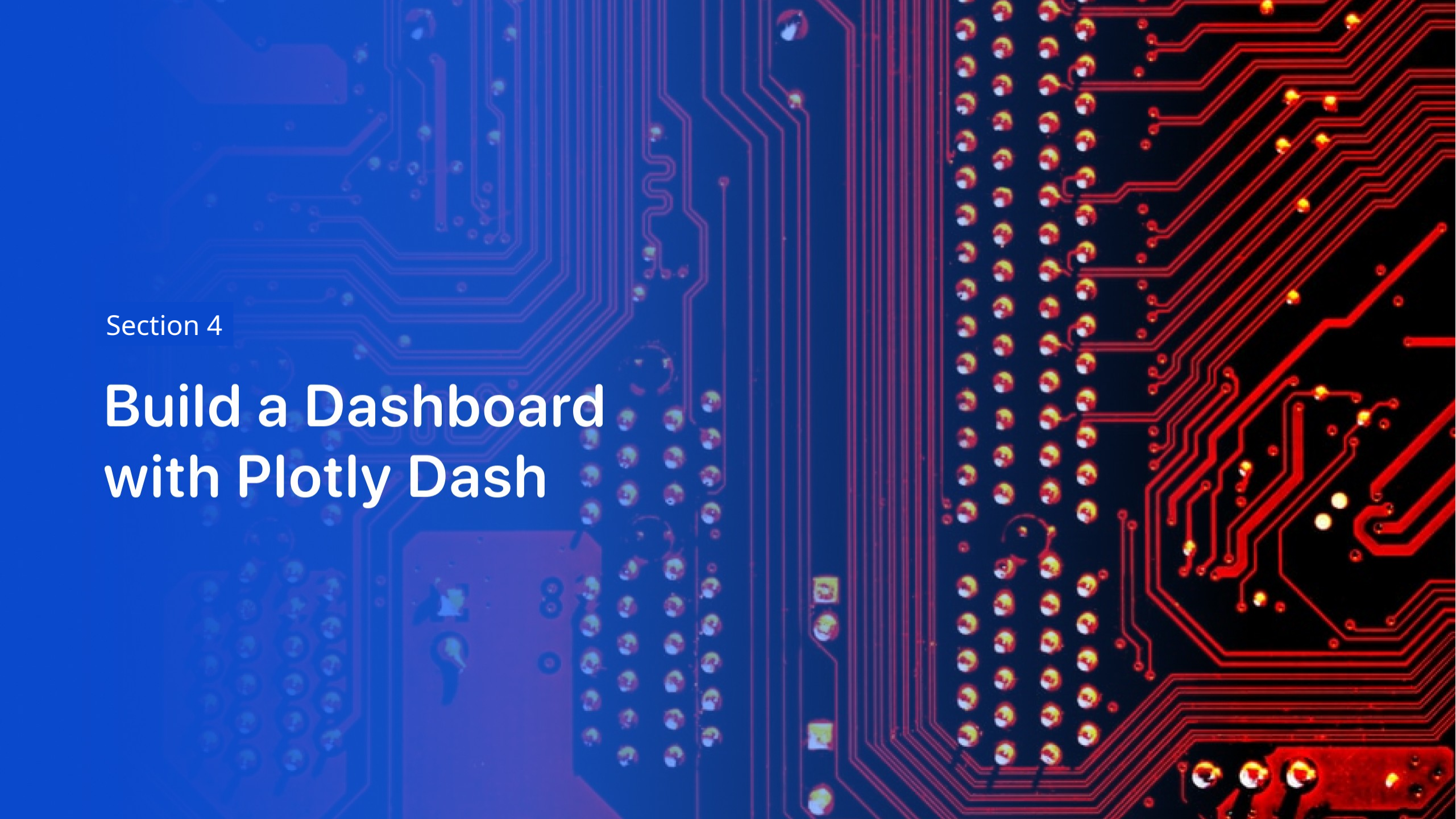


This map shows launch outcomes for CCAFS LC-40 site. There are 25 launches in total.

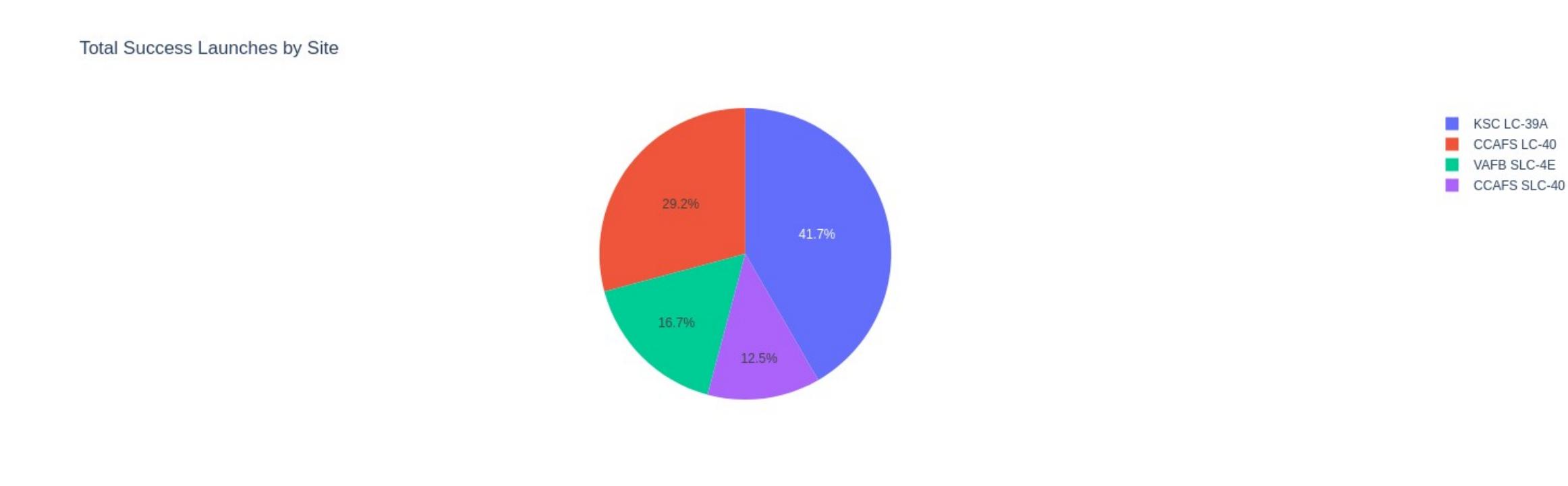
Green markers denote successful launches (7 in total), while red markers – failed launches (18 in total).

CCAFS LC-40 Launch Site and Its Proximity to the Coastline

The distance between CCAFS LC-40 launch site and the coastline is 0.93 km.

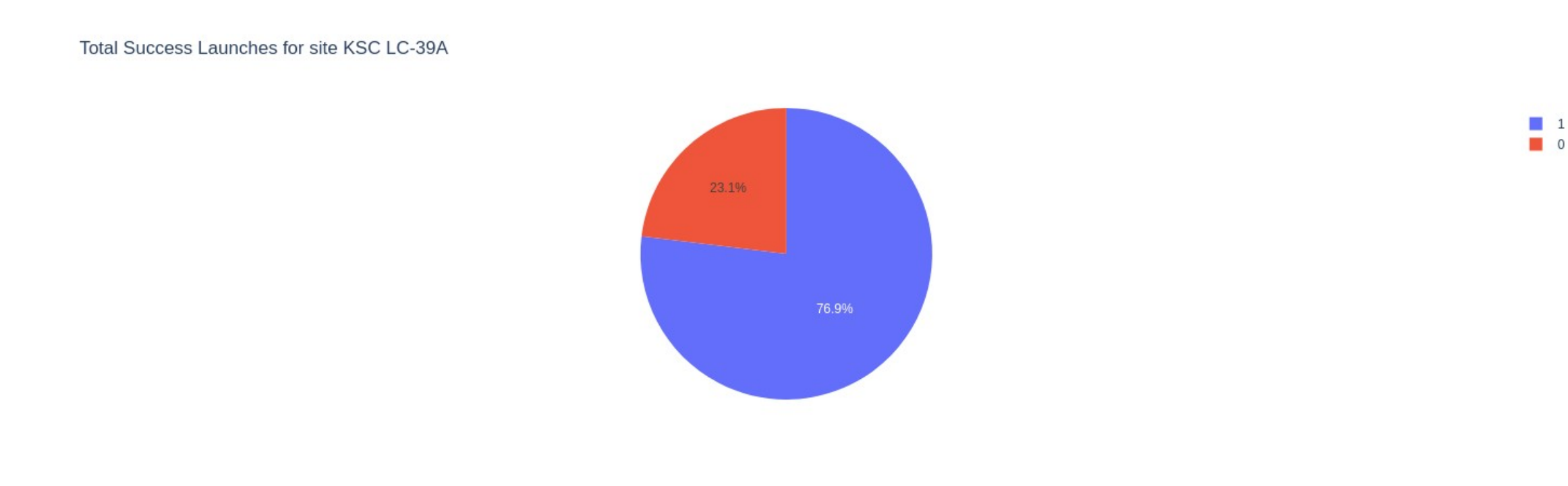

Section 4
Build a Dashboard with Plotly Dash

Total Success Launches by Site



The majority of successful launches were done from KSC LC-39A launch site (41.7%), followed by CCAFS LC-40 (29.2%), VAFB SLC-4E (16.7%) and CCAFS SLC-40 (12.5%).

Total Success Launches for Site KSC LC-39A



Of all launches conducted from KSC LC-39A, 76.9% of them were successful, and 23.1% of them failed.

Correlation between Payload and Success for all Sites

There are less launches in range of 2500 – 7500 kg than in range of 0-5000 kg.

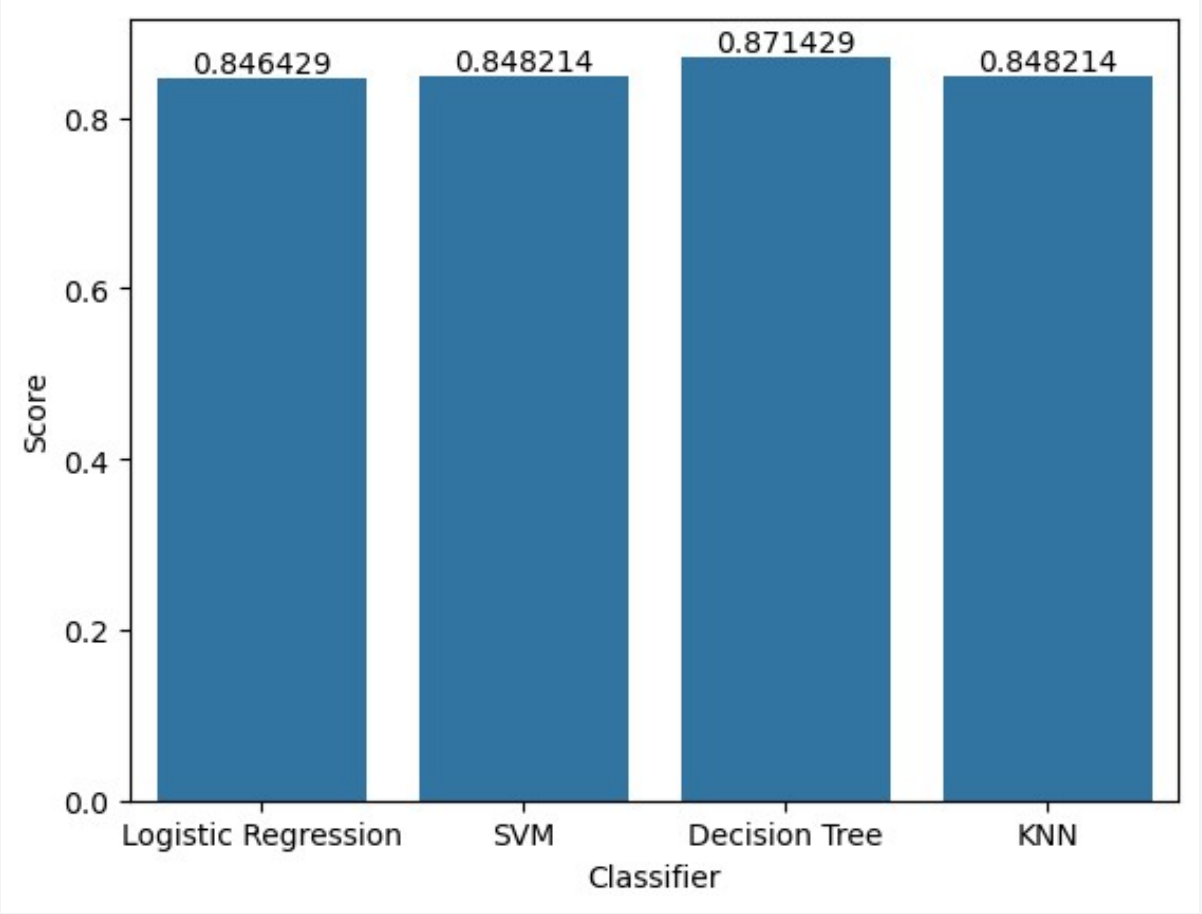
A success rate for launches with payload in range of 0 -5000 kg is higher (52.5%) than in case of payload in range of 2500-7500 kg (44.4%).

Section 5

Predictive Analysis (Classification)

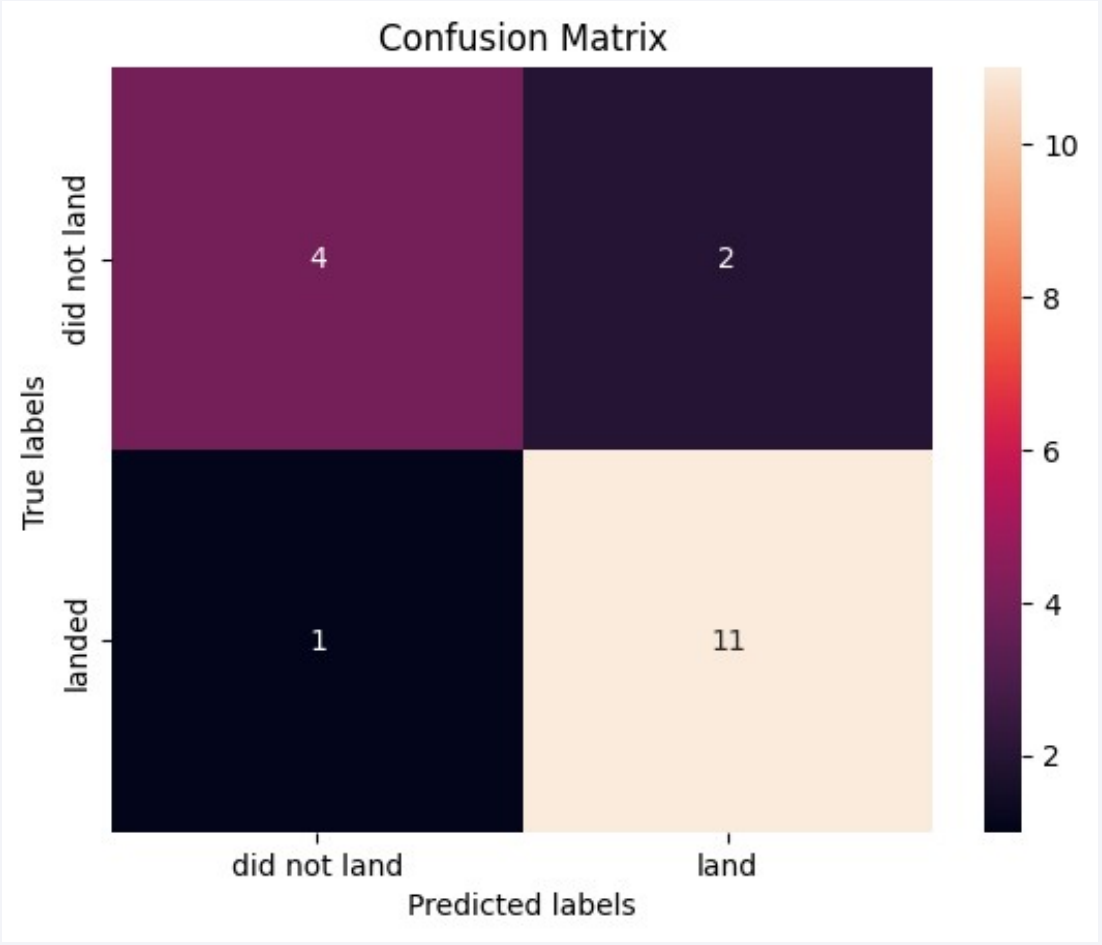
Classification Accuracy

Decision Tree Classifier has the best accuracy (~ 87%), while other three classifiers have nearly the same accuracy (~ 84%)



Confusion Matrix

The Decision Tree Classifier correctly classified 4 out of 6 “did not land” cases and 11 out of 12 “landed” cases.



Conclusions

- DecisionTreeClassifier model had shown the best accuracy among all considered models
- Launches with lower payload mass tend to have more successful outcomes than launches with higher payload mass
- Most launch sites are close to equator and coast line.
- The success rate of launches increased over the years.
- KSC LC-39A site has the highest success rate among considered launch sites.
- Launches aimed at ES-L1, GEO, HEO and SSO orbits have 100% success rate.

Appendix

The GitHub repository containing all Jupyter notebooks used during making this presentation can be accessed here:

https://github.com/atamazian/ds-capstone/tree/main

Thank you!