IBM Developer
SKILLS NETWORK

Winning the Space Race With Data Science

Yochanan Perez
08-Sep-2023

Table of Contents

Executive Summary

Summary of Methodologies

Using trustworthy launch data and various computer models, the optimum parameters of launching a successful competing rocket launching company were determined.

Summary of all results

Optimal Landing Site: Kennedy Space Center Launch Complex 39A **(KSC LC 39A)**

Optimal Pay Load: 2,000 kg – 6,000 kg

Optimal Boosters: F9 v1.1

Optimal Orbit: ISS

Introduction

The primary goal of this research is to develop a predictive model that can accurately determine the likelihood of successfully landing the Falcon 9 first stage of a rocket launch. SpaceX significantly reduce launch costs by reusing the first stage of the Falcon 9 rocket. By reliably predicting the success of this critical phase, the research aims to estimate the overall cost of a launch, a key factor that is crucial for a competitor in the rocket launch market. Real-world launch data are leveraged to build our predictive model.

Section 1

Methodology

Methodology

Collect	Collect raw launch data from high-ranking websites
Preprocess	Clean data by handling missing values and filtering
EDA	Perform exploratory data analysis and SQL
Visualization	Create interactive visual analytics using Folium and Plotly Dash
AI Models	Perform predictive analysis using AI suite in Python

Data Collection – SpaceX API

Access SpaceX API

- Requests library in Python enables HTTP access
- The response is received as a Json

Convert response

- Json converted into dataframe with *json_normalize()*
- Additional data is mapped into final dataframe

Filter for Falcon 9 launches

Handle missing values

GitHub URL of the completed SpaceX API calls notebook

```
data_falcon9.shape
```

```
(90, 17)
```

```
data_falcon9.isnull().sum()
```

```
FlightNumber      0
Date              0
BoosterVersion    0
PayloadMass       0
Orbit             0
LaunchSite        0
Outcome           0
Flights           0
GridFins          0
Reused            0
Legs              0
LandingPad       26
Block             0
ReusedCount       0
Serial            0
Longitude         0
Latitude          0
dtype: int64
```

Data Collection - Scraping

Access Falcon 9 Launch Wiki page

- Requests library in Python enables HTTP access
- The response is received as an HTML

Convert response

- HTML converted into dataframe with BeautifulSoup object
- Additional data is mapped into final dataframe

Parse HTML Tables

GitHub URL of the completed web scraping notebook

```
df=pd.DataFrame(launch_dict)
df.head()
```

	Flight No.	Version,Booster [b]	Launch site	Payload[c]	Payload mass	Orbit	Customer	Launchoutcome	Boosterlanding	1	...	4	5	
0	1	None	CCAFS	None	0	LEO	SpaceX	None	None	None	...	None	None	Non
1	2	None	CCAFS	None	0	LEO	NASA	None	None	None	...	None	None	Non
2	3	None	CCAFS	None	525 kg	LEO	NASA	None	None	None	...	None	None	Non
3	4	None	CCAFS	None	4,700 kg	LEO	NASA	None	None	None	...	None	None	Non
4	5	None	CCAFS	None	4,877 kg	LEO	NASA	None	None	None	...	None	None	Non

5 rows × 22 columns

Data Wrangling

EDA

To determine labels used in *LaunchSite* column	To determine dedicated orbit	To determine mission outcome (success/failure) rate

GitHub URL of completed data wrangling related notebooks

EDA with SQL

SQL Code	Output
%sql SELECT DISTINCT "Launch_Site" FROM SPACEXTBL;	4 unique launch sites listed
%sql SELECT SUM("PAYLOAD_MASS__KG_") AS "Total Payload Mass (kg)" FROM SPACEXTBL WHERE "Customer" LIKE 'NASA%';	Total mass carried by boosters on behalf of NASA: 99980 (kg)
%sql SELECT MIN("Date") AS "First Successful Ground Pad Landing Date" FROM SPACEXTBL WHERE "Mission_Outcome" = 'Success' AND "Landing_Outcome" = 'Success (ground pad)';	Date of first successful ground pad landing: 2015-12-22
%sql SELECT "Booster_Version" FROM SPACEXTBL WHERE "Landing_Outcome" = 'Success (drone ship)' AND "PAYLOAD_MASS__KG_" > 4000 AND "PAYLOAD_MASS__KG_" < 6000;	4 booster versions listed
%sql SELECT "Mission_Outcome", COUNT(*) AS "Total" FROM SPACEXTBL GROUP BY "Mission_Outcome";	Total number of successful and failed missions
%sql SELECT substr('JanFebMarAprMayJunJulAugSepOctNovDec', 1 + 3*strftime('%m', Date), -3) AS "Month", "Launch_Site", "Booster_Version", "Landing_Outcome" FROM SPACEXTBL WHERE substr(Date, 1, 4) = '2015' AND "Landing_Outcome" LIKE 'Failure (drone ship)%';	Failed missions for the months in 2015

GitHub URL of completed *EDA with SQL* notebook

EDA with Data Visualization

Charts plotted and why:

- **FlightNumber vs. PayloadMass** – observe inverse relationship between increasing flight number (increase in positive outcome) and payload mass (increase in negative outcome)
- **Flight Number vs. Launch Site** – highlight the launch site with the most flights
- **PayloadMass vs. LaunchSite** – determine launch site with the least and most massive payloads
- **Success Rate by Orbit Type** – highlight most successful destination orbits
- **FlightNumber vs. Orbit Type** – observe relationship between number of flights and destination orbit
- **PayloadMass vs. Orbit Type** – observe relationship between payload and orbit type
- **Launch Success Rate by Year** – view trend of success rate for years

GitHub URL of completed EDA with data visualization notebook

Build an Interactive Map with Folium

- Added an Orange Circle around NASA Johnson Space Center near Houston, TX.
- Added Marker Clusters for successful (green) and unsuccessful (red) rocket launches.
- Added a line for distances between Launch Site and various landmarks.
- All map objects were added to highlight the common features in geography of launch sites.
- GitHub URL of completed interactive map with Folium map (saved as a Jupyter Notebook and in HTML format).

Build a Dashboard with Plotly Dash

Summary of what plots/graphs and interactions added to dashboard:

Dropdown List added to choose a single or multiple launch sites.

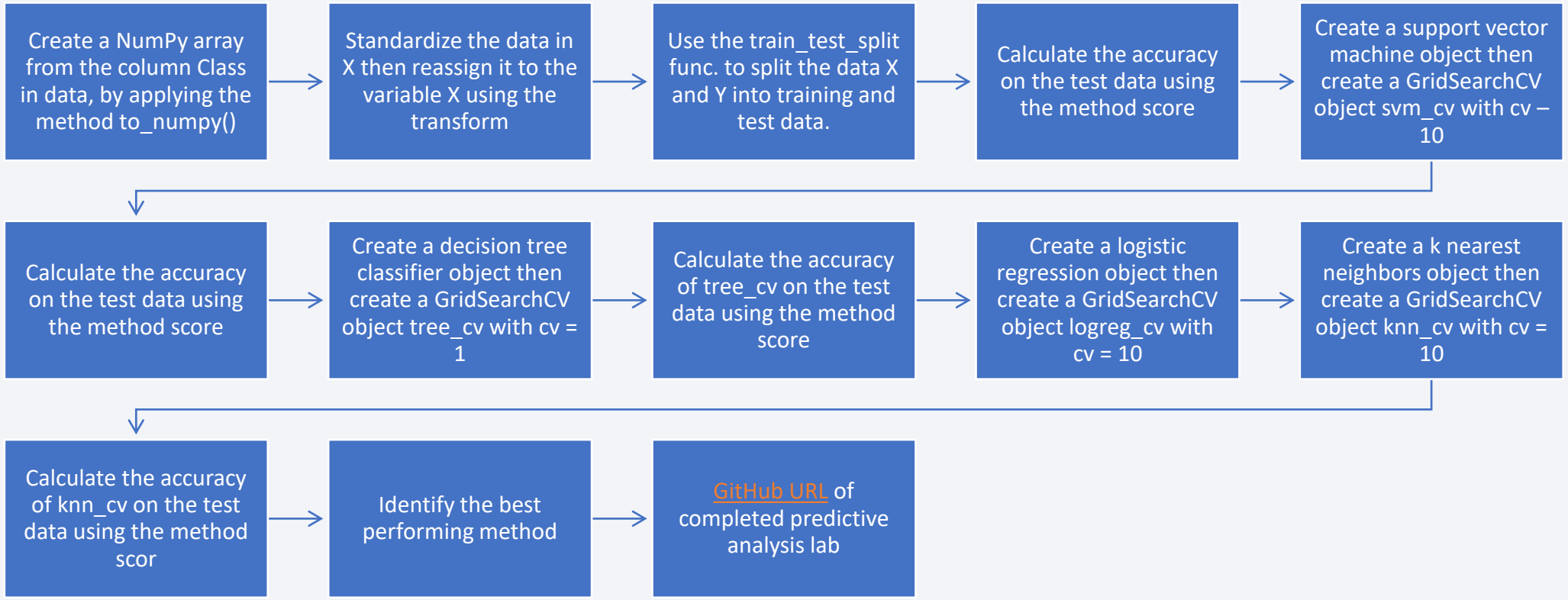
Piechart created to show successful and unsuccessful launches per site.

Dynamic slider to select range of launch payload.

Scatter chart to show the correlation between payload and launch success.

GitHub URL of completed Plotly Dash lab.

Predictive Analysis (Classification)



Results

Exploratory Data Analysis

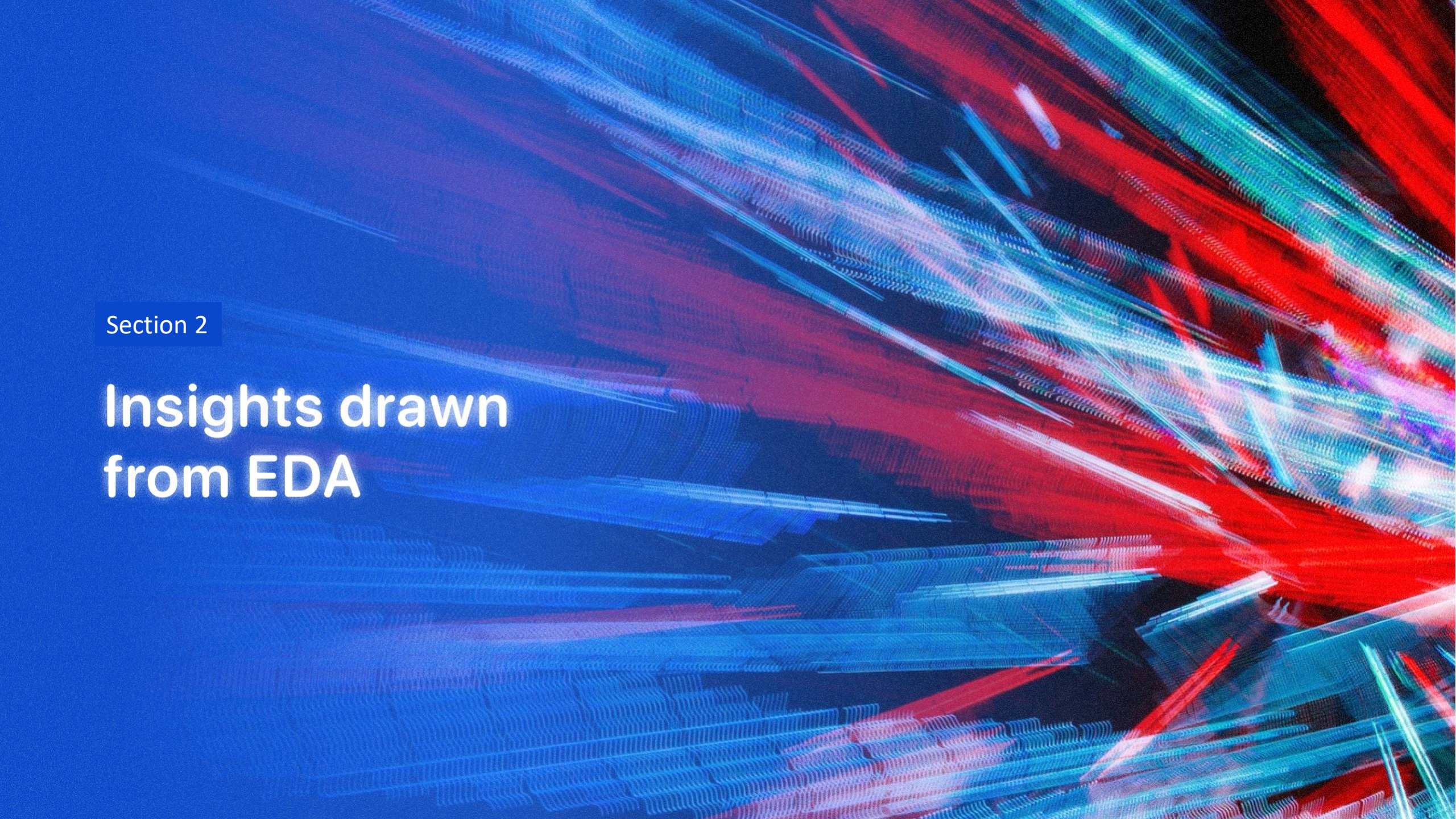
- General improvement in launch success over time
- Orbits with 100% success rate: ES-L1, GEO, HEO and SSO
- KSC LC-39A and VAFB SLC 4E both have a success rate of 77%

Predictive Analysis

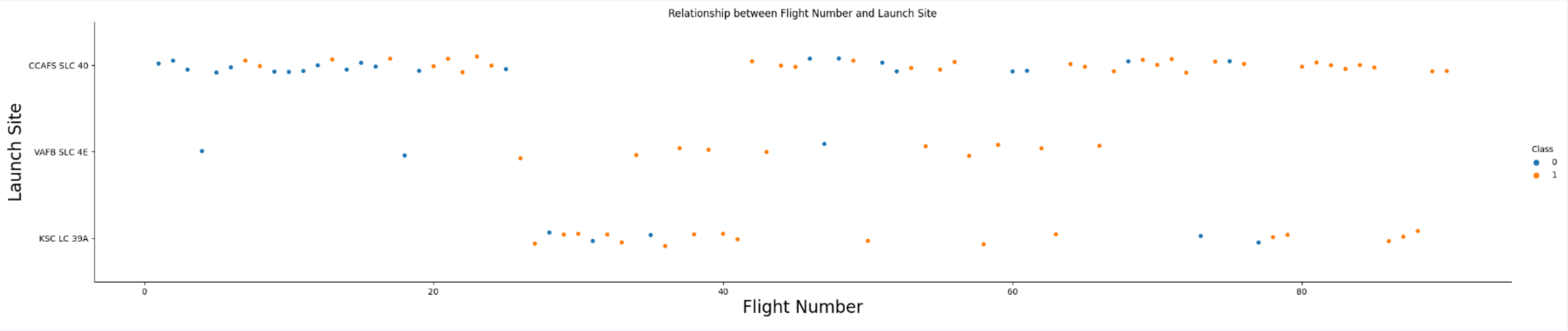
DecisionTreeClassifier() has the best accuracy on validation data

Section 2

Insights drawn from EDA



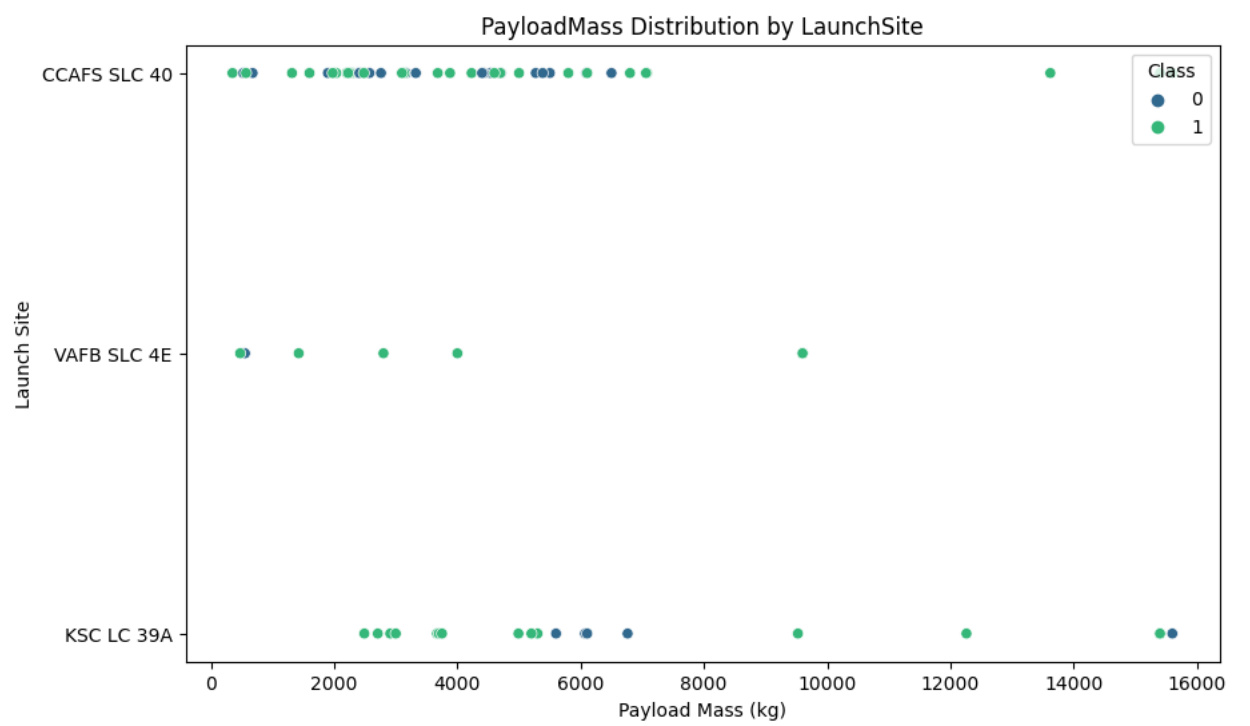
Flight Number vs. Launch Site



- Most unsuccessful landings are planned.
- CCAFS SLC 40 appears to be a main testing site.
- Generally, an increase in continuous launch attempts leads to more successful landings.

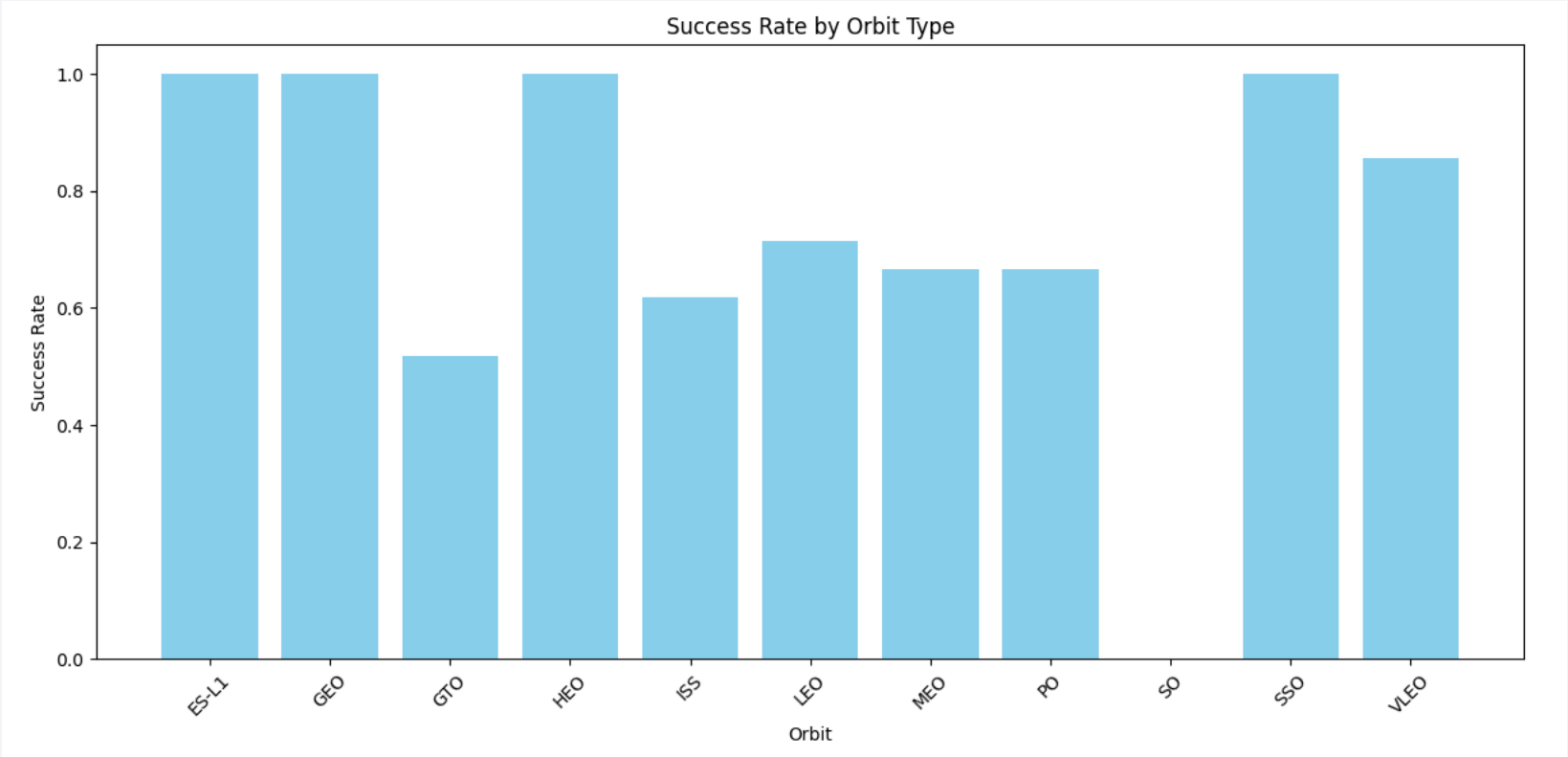
- Flight Number = continuous launch attempts
- Blue dots – Failed launch
- Orange dots – Successful launch

Payload vs. Launch Site



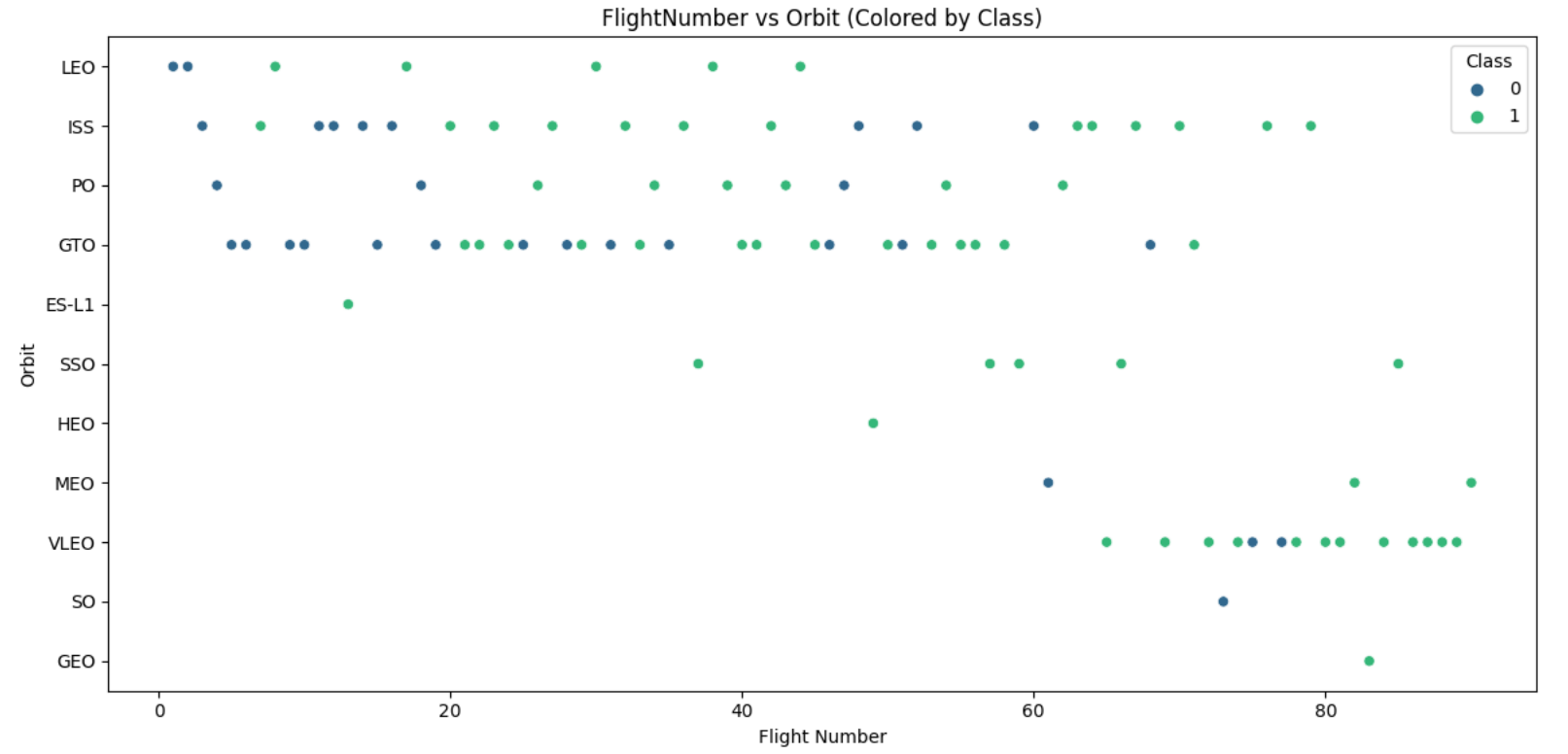
- It appears the more massive the payload, the less likely the first stage will return.
- Blue dot – Failure
- Green dot – Successful launch

Success Rate vs. Orbit Type



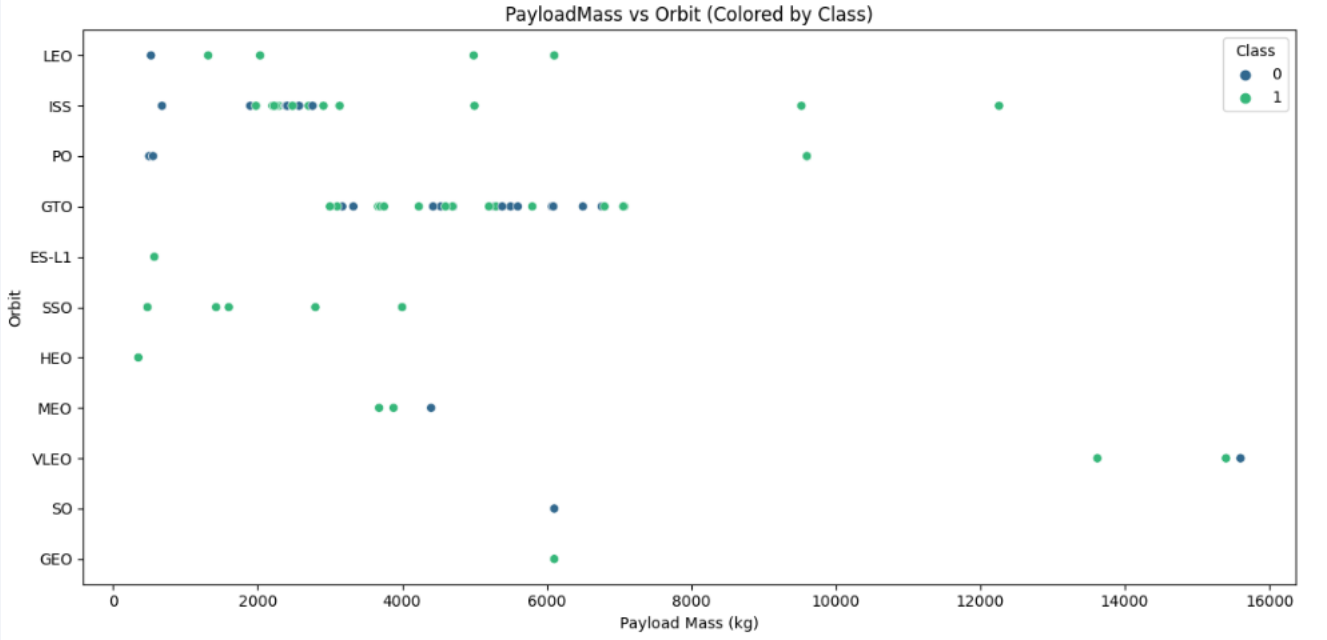
- 4/10 sites have 100% success rate
- Remaining sites with data have between 50% - 80% success rate

Flight Number vs. Orbit Type



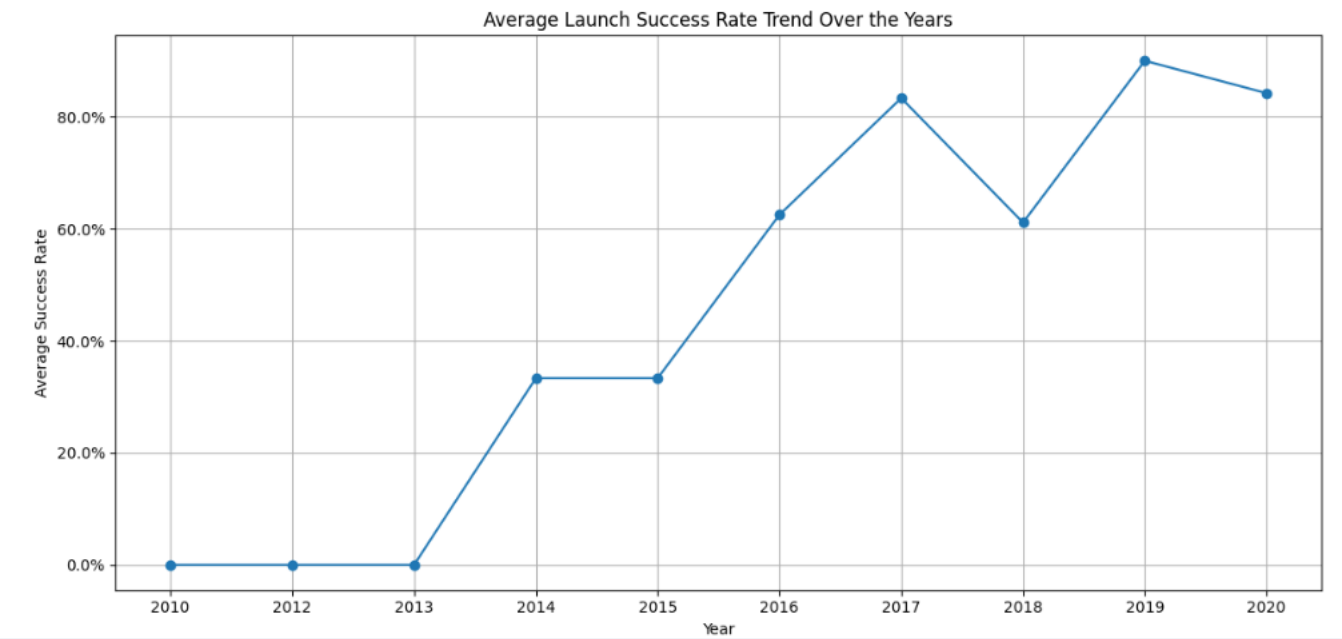
- It appears the more attempted flights, the more likely the first stage will return.
- Blue dot – Failure
- Green dot – Successful launch

Payload vs. Orbit Type



- Heavier payloads are associated with success for Polar, LEO, and ISS orbital destinations.
- Blue dot – Failure
- Green dot – Successful launch

Launch Success Yearly Trend



- In general, the trend is an improved success rate year over year
- There were only two years for regression: 2018 and 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

- Selected “DISTINCT” from *Launch_Site* column
- There are four unique launch sites

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_Outcome
2010-04-06	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-08-12	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	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-08-10	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-01-03	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

- Regex type filter “LIKE” in *Launch_Site* column for launch sites that begin with CCA from SpaceX Table
- Limited to 5 results above

Total Payload Mass

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

```
%sql SELECT SUM("PAYLOAD_MASS__KG_") AS "Total Payload Mass (kg)" FROM SPACEXTBL WHERE "Customer" LIKE 'NASA (CRS)';
```

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

Total Payload Mass (kg)
45596

- The total payload carried by boosters from NASA was 45,596 kg
- Payload mass column was added up for NASA (CRS) customer

Average Payload Mass by F9 v1.1

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

```
# Execute an SQL query to calculate the average payload mass carried by booster version F9 v1.1
%sql SELECT AVG("PAYLOAD_MASS__KG_") AS "Average Payload Mass (kg)" FROM SPACEXTBL WHERE "Booster_Version" = 'F9 v1.1';
```

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

Average Payload Mass (kg)
2928.4

- The average payload mass carried by booster version F9 v1.1 was 2,928.4 kg
- Booster_Version column filtered for F9 v1.1 type

First Successful Ground Landing Date

```
# Execute an SQL query to find the date of the first successful landing on a ground pad
%sql SELECT MIN("Date") AS "First Successful Ground Pad Landing Date" FROM SPACEXTBL WHERE "Mission_Outcome" = 'Success' AND "Landing_Outcome" = 'Success (ground pad)';
```

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

First Successful Ground Pad Landing Date
2015-12-22

The first successful landing outcome on ground pad was December 22, 2015

Successful Drone Ship Landing with Payload between 4000 and 6000

```
# Execute an SQL query to list the names of the boosters meeting the criteria
%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

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

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

Total Number of Successful and Failure Mission Outcomes

```
# Execute an SQL query to count the number of successful and failure mission outcomes
%sql SELECT "Mission_Outcome", COUNT(*) AS "Total" FROM SPACEXTBL GROUP BY "Mission_Outcome";
```

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

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

- 99 Successful flights
- 1 Failure in flight
- 1 Success (payload status unclear)

Boosters Carried Maximum Payload

```
%sql SELECT "Booster_Version","PAYLOAD_MASS__KG_" FROM SPACEXTBL WHERE "PAYLOAD_MASS__KG_" = 15600 GROUP BY "Booster_Version";
```

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

Booster_Version	PAYLOAD_MASS__KG_
F9 B5 B1048.4	15600
F9 B5 B1048.5	15600
F9 B5 B1049.4	15600
F9 B5 B1049.5	15600
F9 B5 B1049.7	15600
F9 B5 B1051.3	15600
F9 B5 B1051.4	15600
F9 B5 B1051.6	15600
F9 B5 B1056.4	15600
F9 B5 B1058.3	15600
F9 B5 B1060.2	15600
F9 B5 B1060.3	15600

The names of the 12 boosters which have carried the maximum payload mass are listed above.

2015 Launch Records

```
# substr('JanFebMarAprMayJunJulAugSepOctNovDec', 1 + 3*strftime('%m', date('now')), -3)
%sql SELECT substr('JanFebMarAprMayJunJulAugSepOctNovDec', 1 + 3*strftime('%m', Date), -3) AS "Month", "Launch_Site", "Booster_Version", "Landing_Outcome"
FROM SPACEXTBL WHERE substr(Date, 1, 4) = '2015' AND "Landing_Outcome" LIKE 'Failure (drone ship)%';
```

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

Month	Launch_Site	Booster_Version	Landing_Outcome
Oct	CCAFS LC-40	F9 v1.1 B1012	Failure (drone ship)
Apr	CCAFS LC-40	F9 v1.1 B1015	Failure (drone ship)

List of the failed *landing_outcomes* in drone ship, their booster versions, and launch site names for in year 2015 above.

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(*) AS "Outcome_Count" FROM SPACEXTBL WHERE Date BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY "Landing_Outcome" ORDER BY "Outcome_Count" DESC;
```

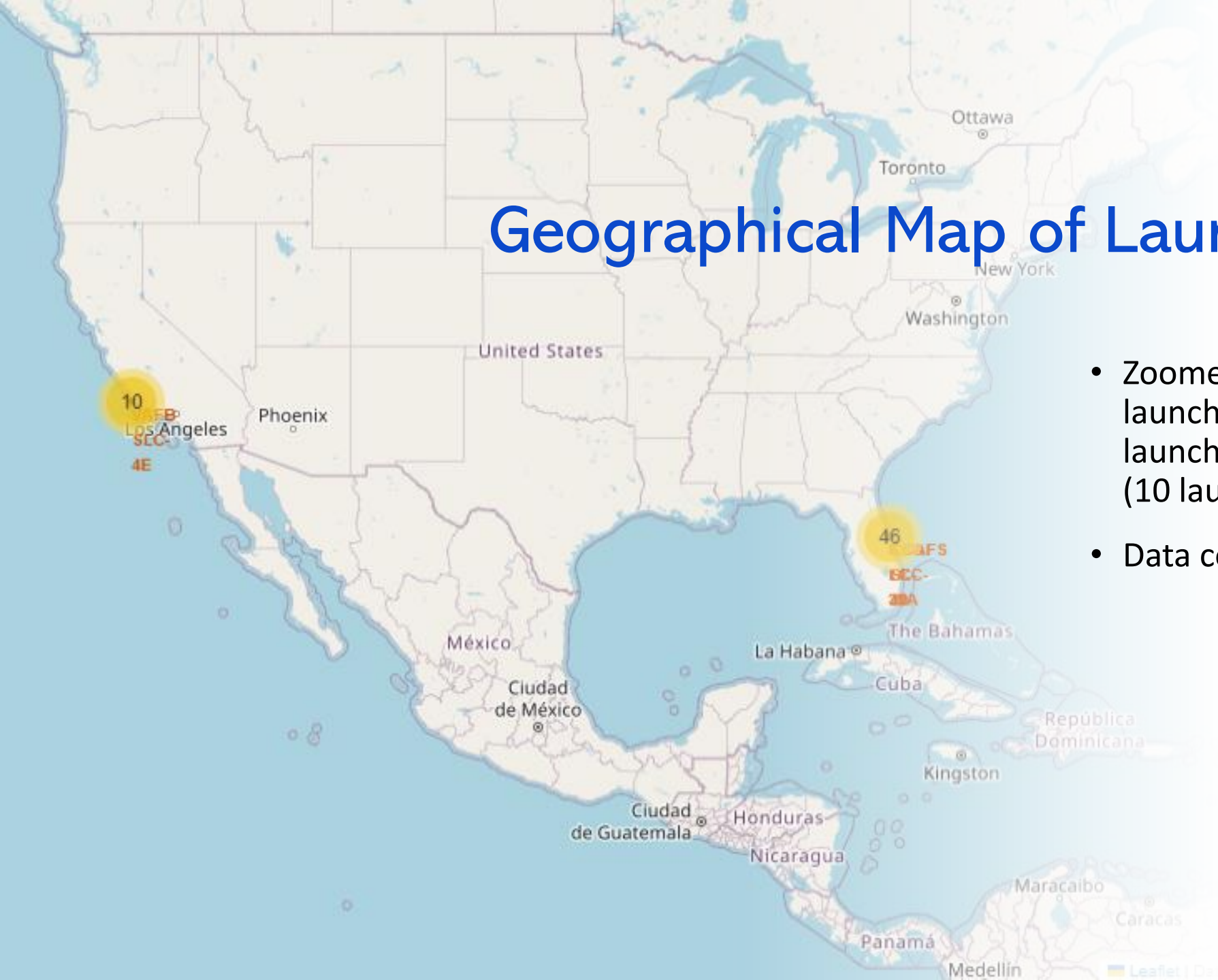
```
 * sqlite:///my_data1.db
Done.
```

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

Rank 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.

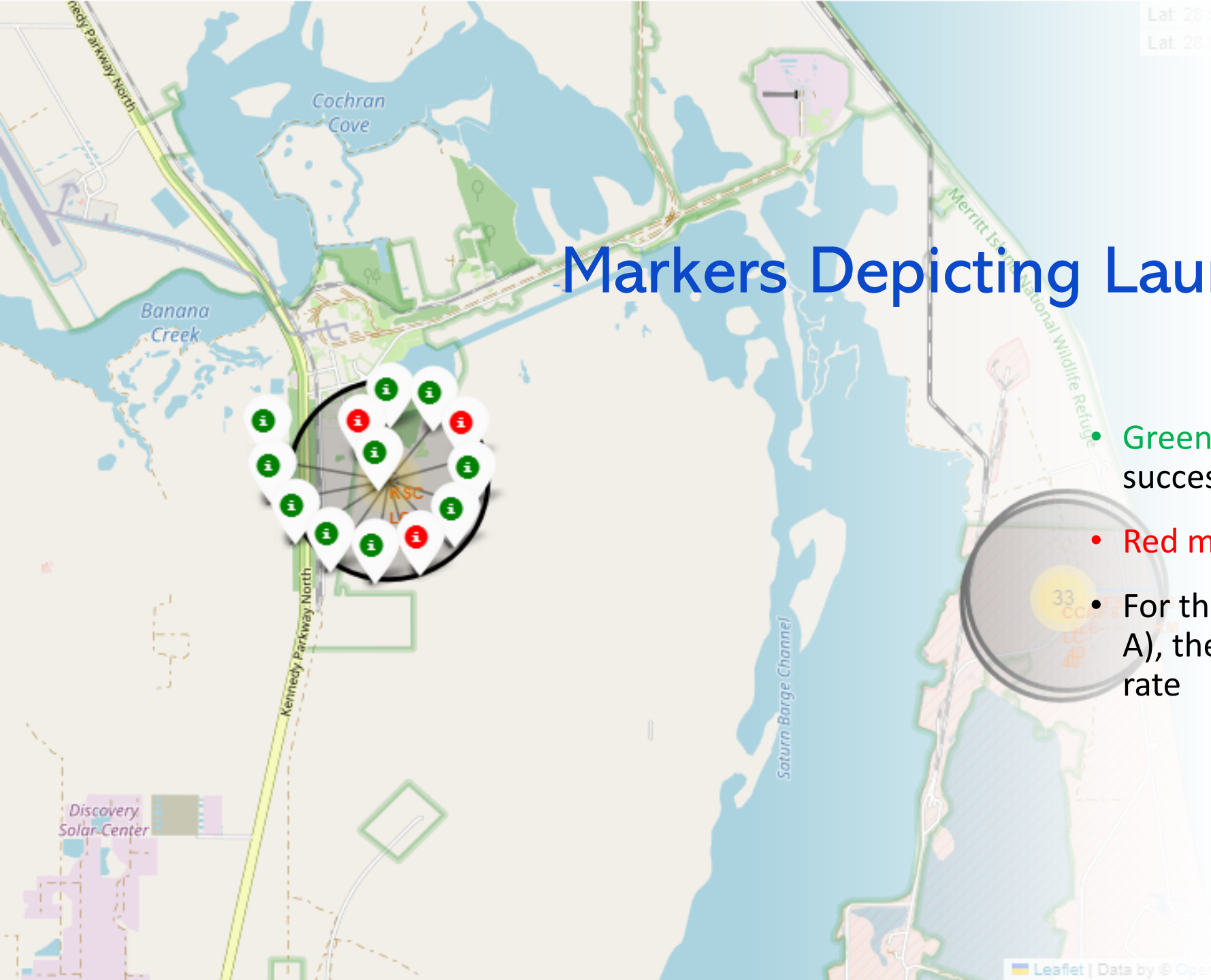

Section 3
Launch Sites
Proximities Analysis

Geographical Map of Launch Sites

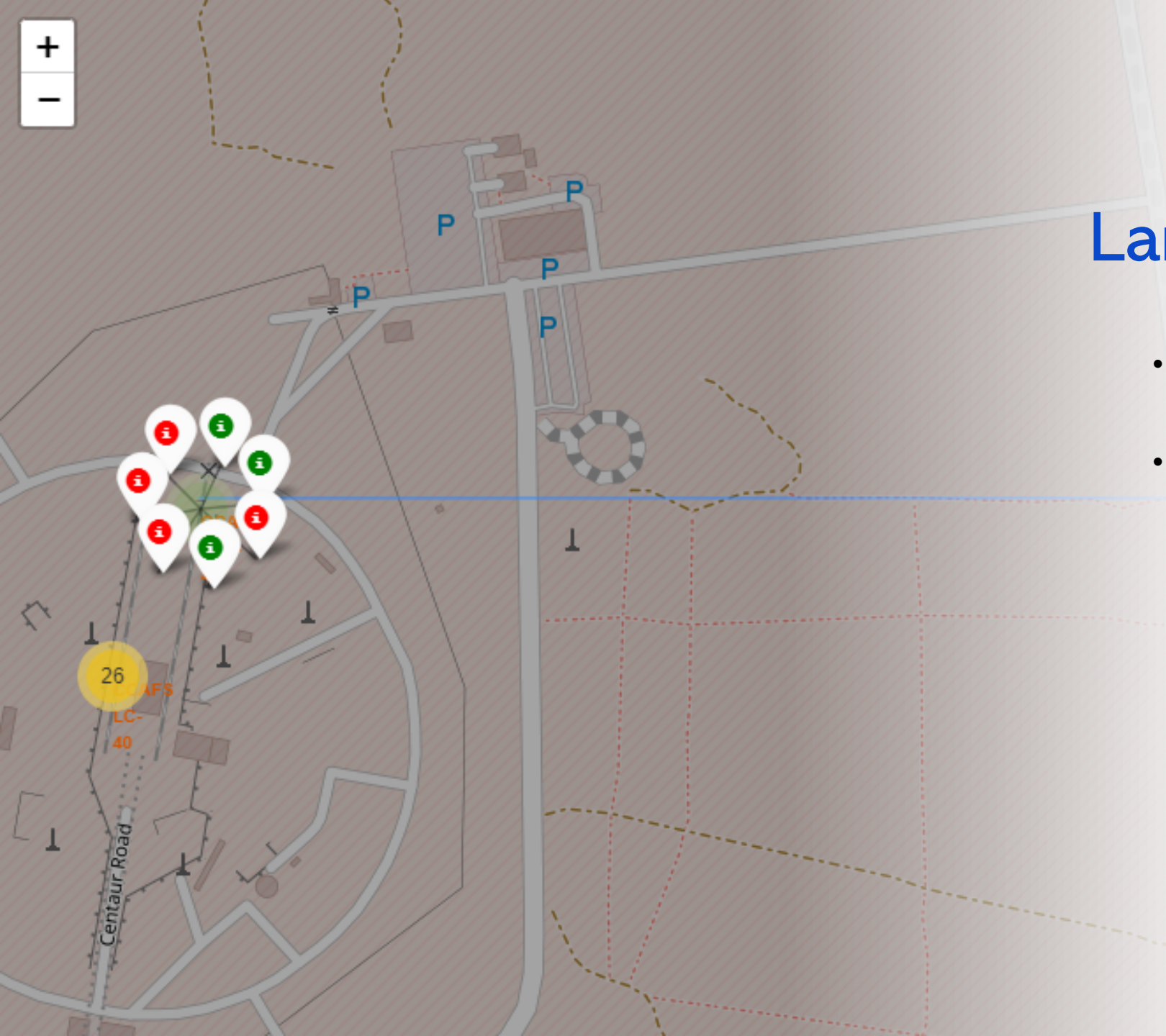


- Zoomed out map shows two launch sites in Florida (46 launches) and one in California (10 launches)
- Data covers 56 total launches

Markers Depicting Launch Outcomes

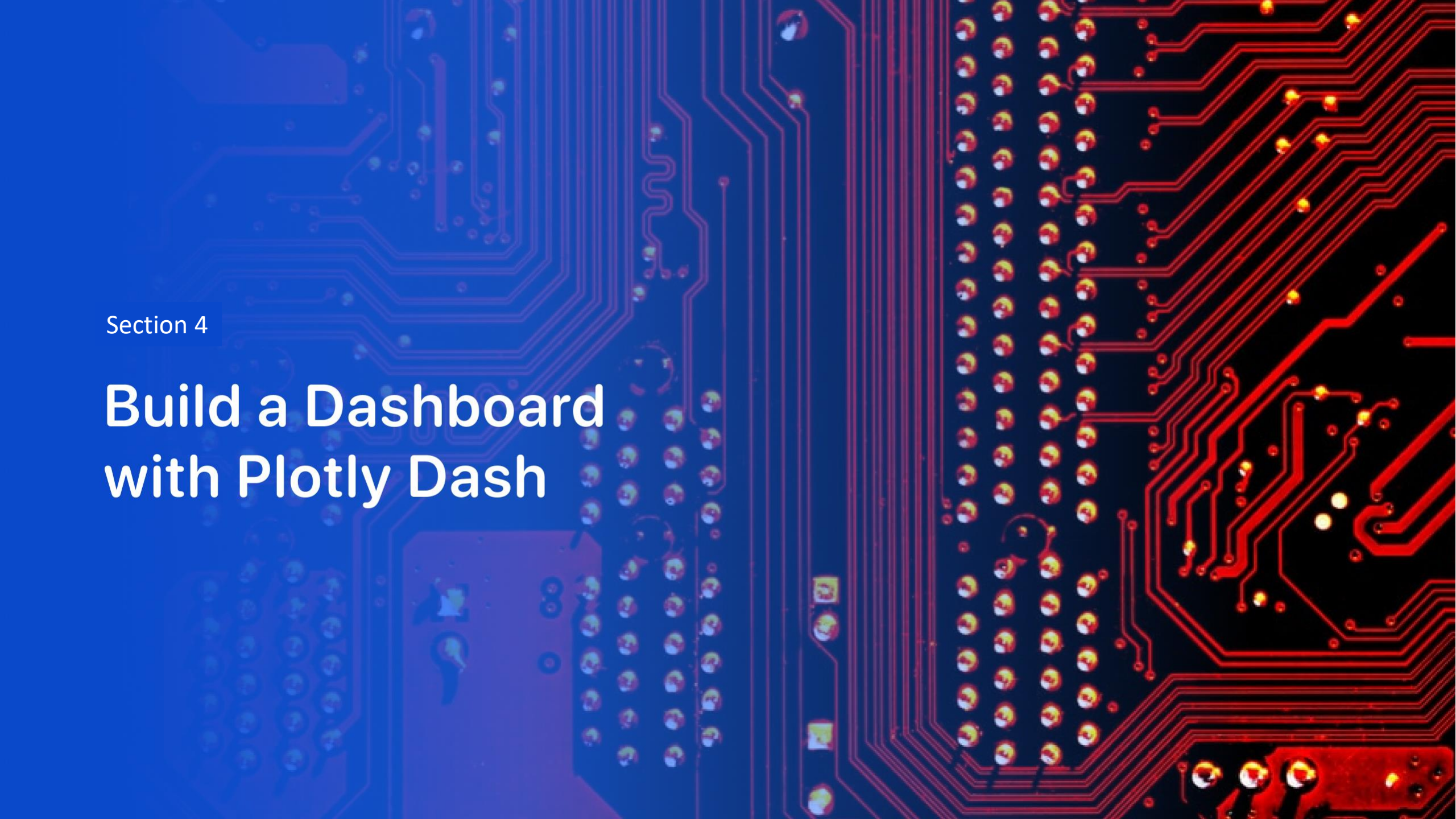


- Green markers denote successful launch
- Red markers denote failure
- For the site depicted (KSC LC 39 A), there is a nearly 77% success rate



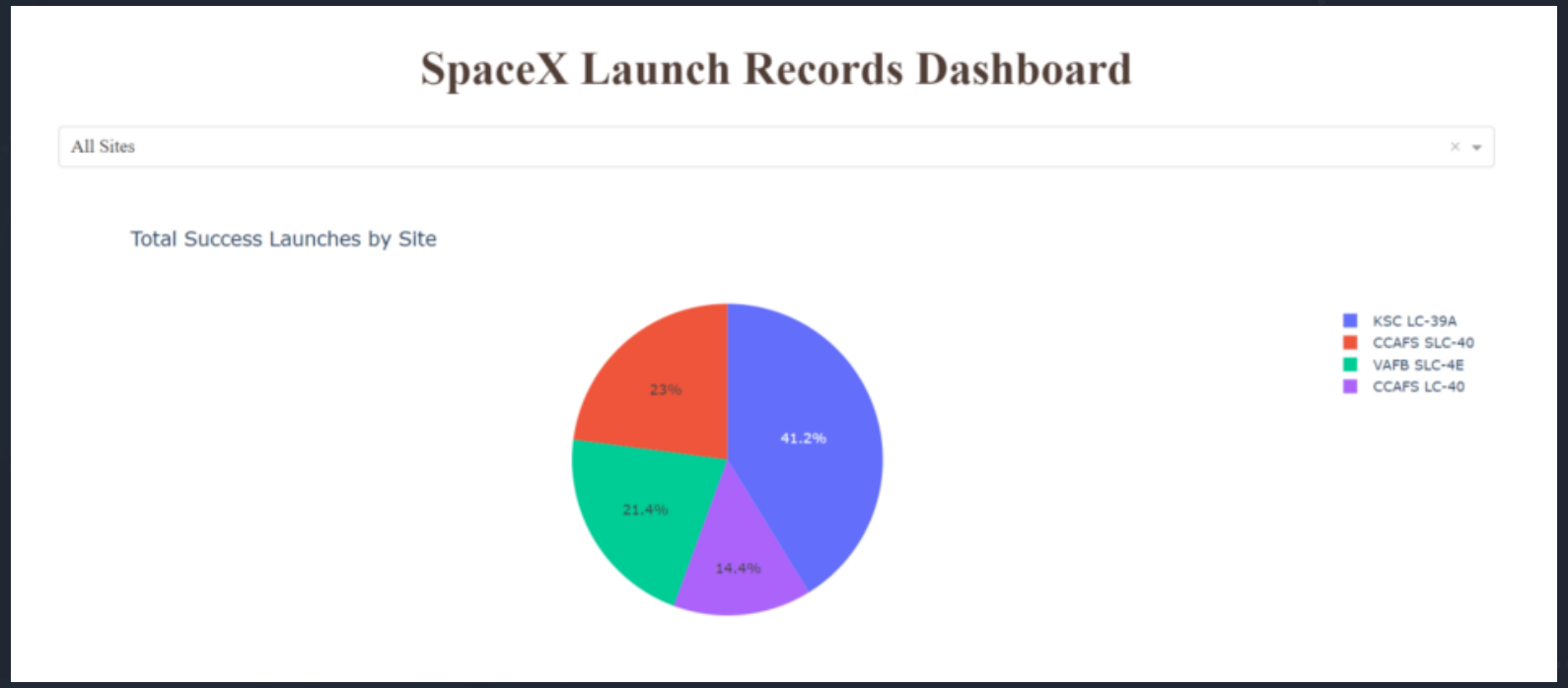
Landmark Proximities

- Blue line shows distance to coastline (approx. 90 km)
- Distance to Houston Launch site is approximately 1435 km (not shown)


Section 4
Build a Dashboard with Plotly Dash

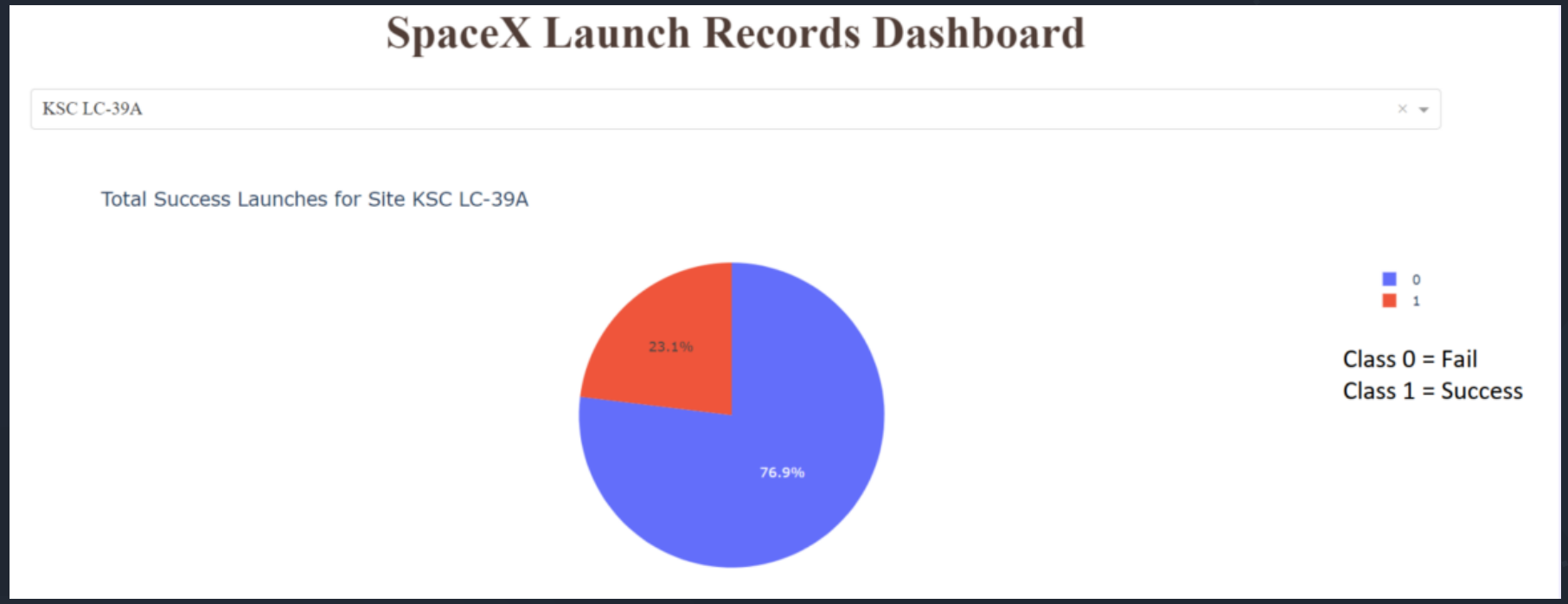
Launch Success Counts: All Sites

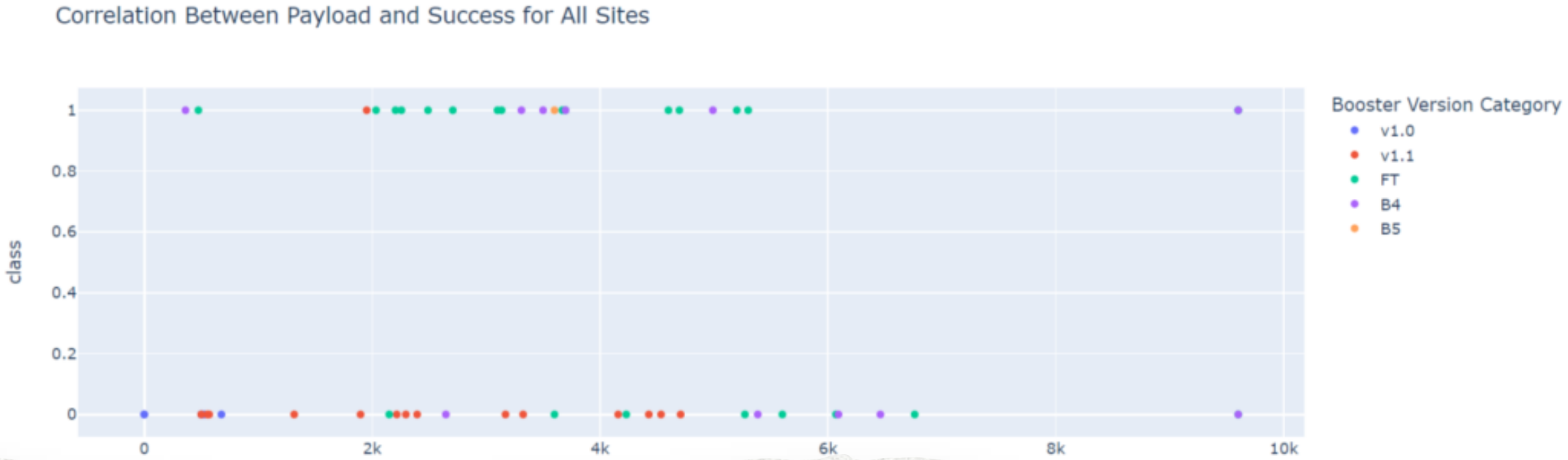
- KSC LC-39A has the highest success percentage of all sites
- CCAFS LC-40 has the lowest



Launch Site with the Highest Success Ratio

- KSC LC-39A has a nearly 77% success rate of total launches
- Failed launches relatively low rate of 23%





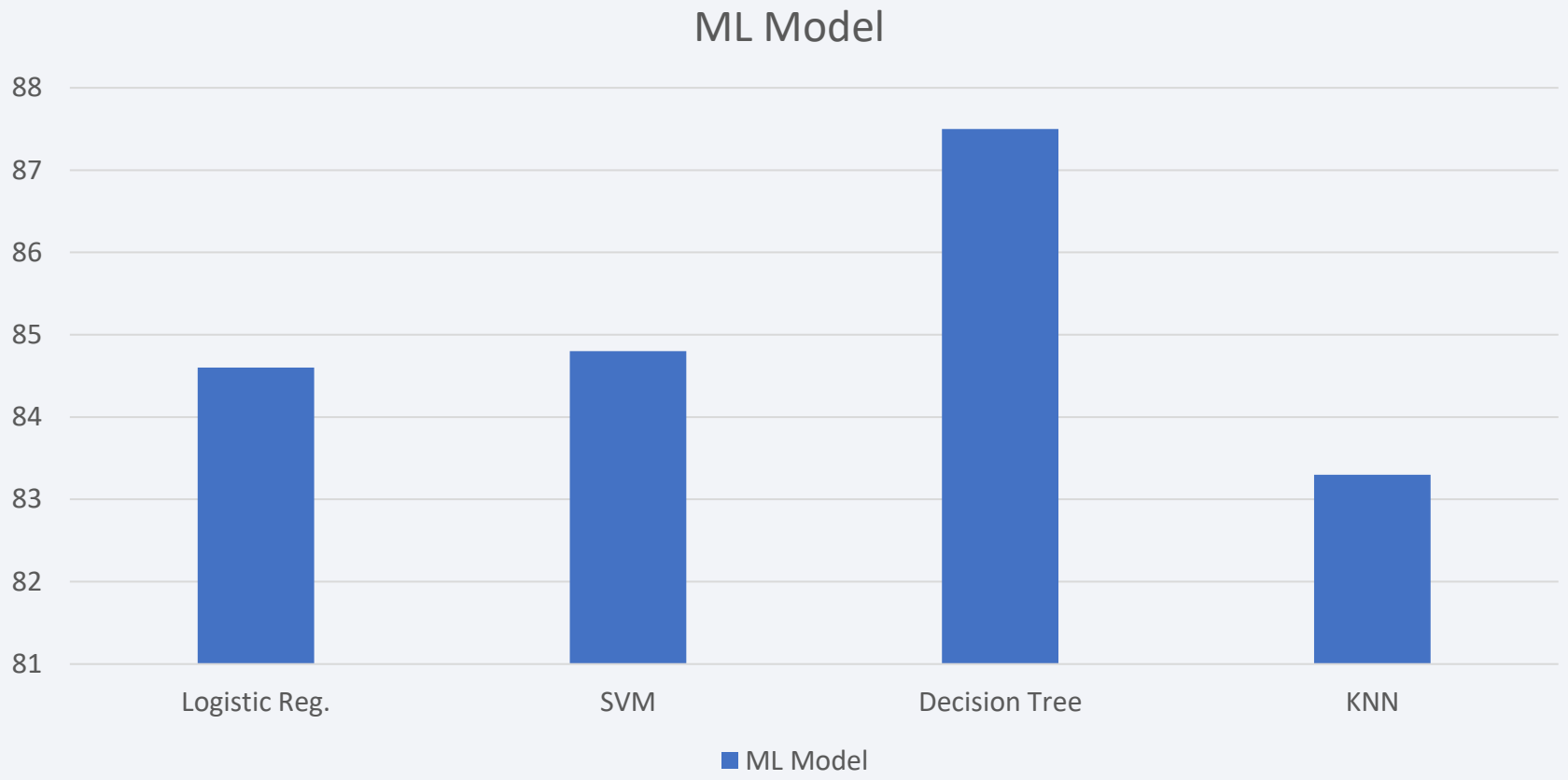
Payload vs. Launch Outcome Scatterplot

- 1 is Successful, 0 is Failure
- Successful launches are clustered between 2,000 kg <= 6,000 kg

Section 5

Predictive Analysis (Classification)

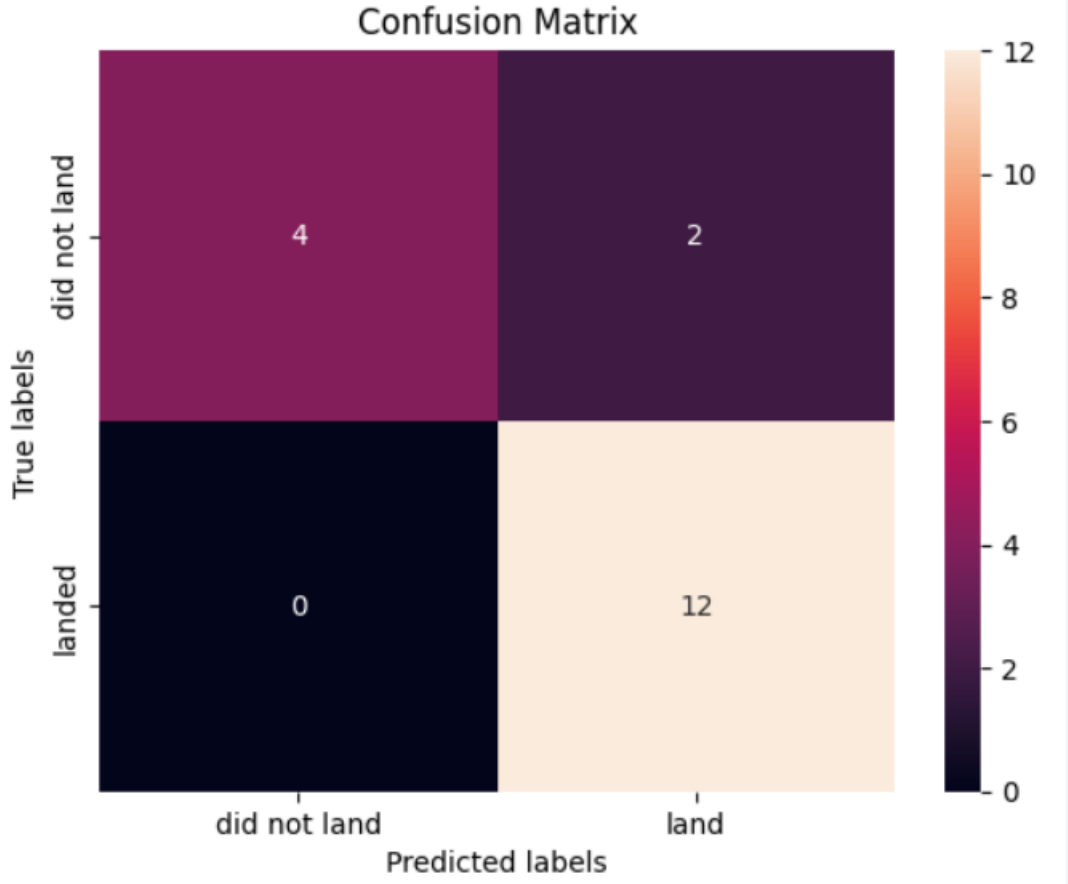
Classification Accuracy



- Bar chart shows the accuracy of various Machine Learning models
- The decision tree performs the best on this limited data

Confusion Matrix

```
yhat = tree_cv.predict(X_test)
plot_confusion_matrix(Y_test,yhat)
```



The confusion matrix of the best performing model is the Decision Tree (*tree_cv*)

Breakdown:

12 True Positives	4 True Negatives	2 False Positives	0 False Negative

Conclusions

Optimal Launch Site: Kennedy Space Center Launch Complex 39A **(KSC LC 39A)**

Optimal Pay Load: 2,000 kg – 5,700 kg

Optimal Booster Type: F9 v1.1

Optimal Destination Orbit: ISS

Optimal Machine Learning Model: Decision Tree

Geographic Location: Coastal

Thank you!