IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Gwenn LE GOFF
11.17.2024

Outline

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

Executive Summary

- **Methodologies** :
  - Data Collection
  - Data Wrangling
  - Data Analysis and Visualization
  - Interactive map with Folium
  - Dashboard with Plotly Dash
  - Predictive analysis
- **Results**
  - Data Analysis results
  - Interactive analytics demo
  - Predictive analysis results

Introduction

- **Project Background and Context**
  - SpaceX has revolutionized the commercial space industry by making space travel more affordable and accessible. The company advertises Falcon 9 rocket launches at a price of $62 million per launch, significantly less than the $165 million typically charged by other providers. This cost reduction is largely due to SpaceX's innovative ability to reuse the rocket's first stage. Predicting whether the first stage will successfully land can provide insights into the potential cost-effectiveness of a launch.
  - Using publicly available data and machine learning models, this project aims to predict the likelihood of SpaceX reusing the first stage.
- **Questions to Be Answered:**
  - How do factors like payload mass, launch site, flight count, and orbit type influence the success of the first stage landing?
  - Has the success rate of landings improved over the years?
  - Which algorithm performs best for binary classification in this context?

Section 1

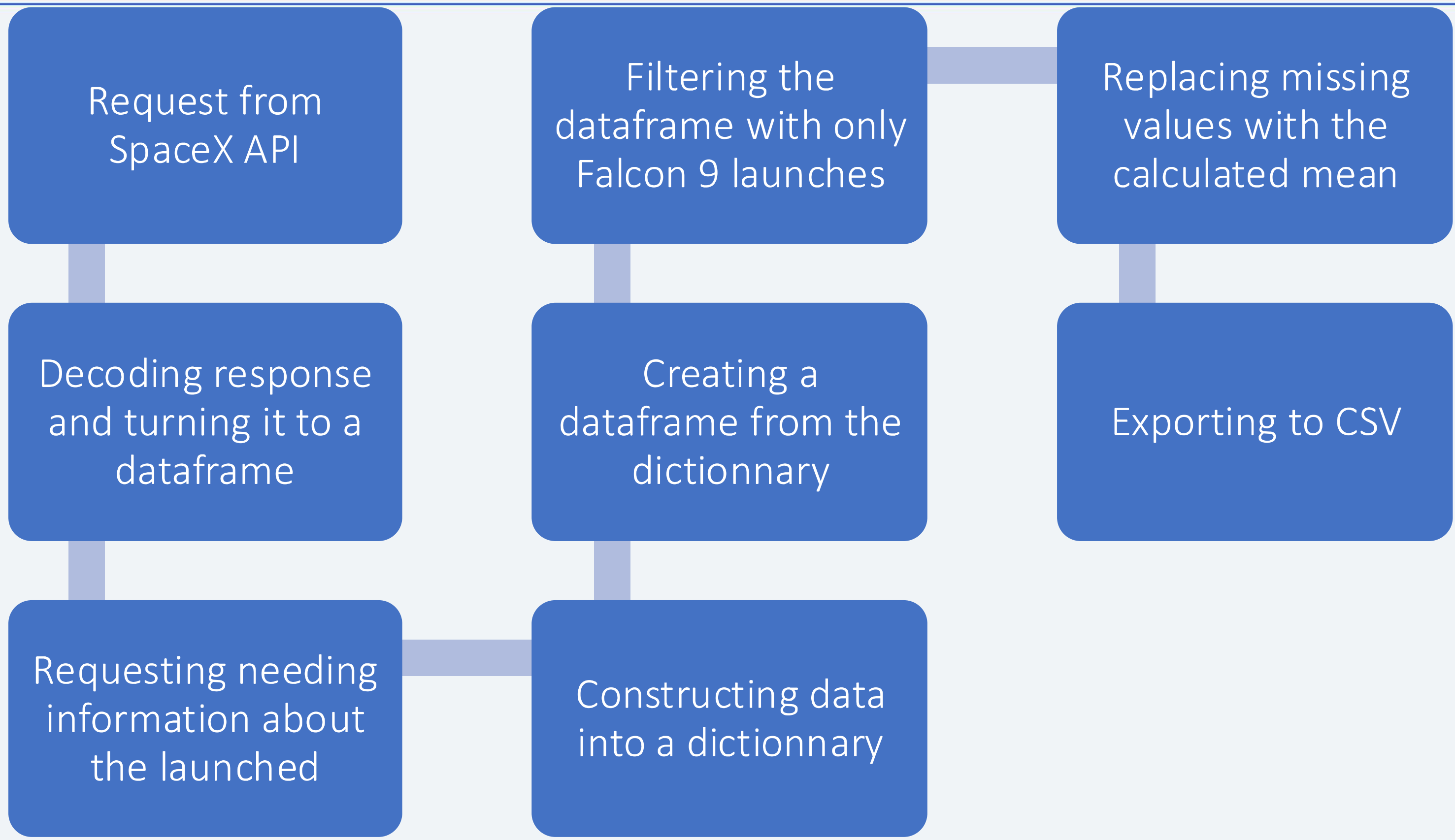
Methodology

Methodology

Executive Summary

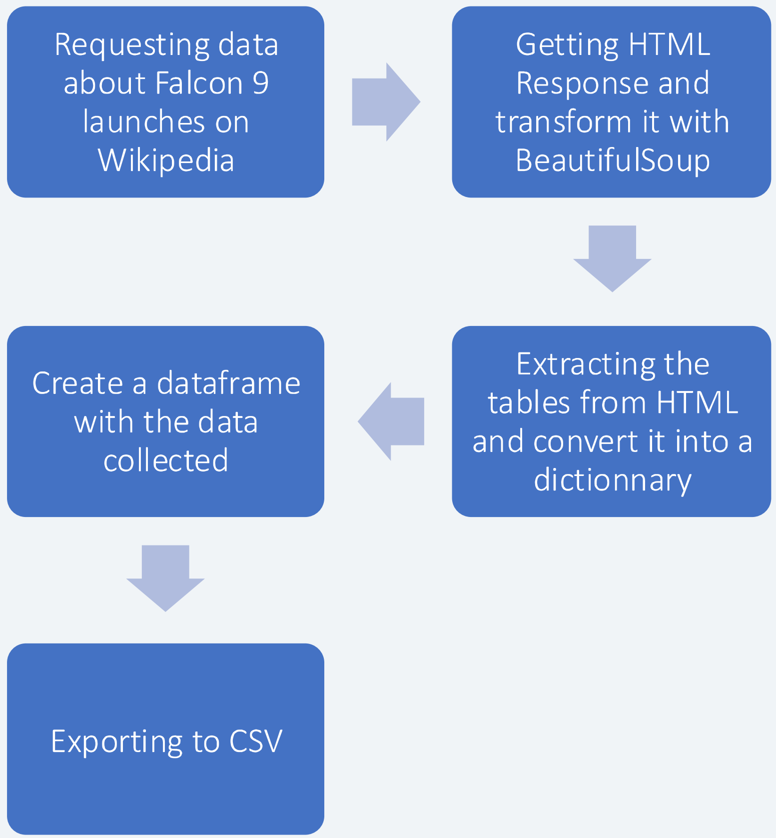
- Data collection methodology:
    - Webscraping and SpaceX API
- Perform data wrangling
    - Convert mission outcomes into training labels using Pandas and Numpy
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
    - Search of the best hyperparameters for SVM, Classification Trees and Logistic Regression, and witch one performs the best

Data Collection – SpaceX API



[GitHub URL]

Data Collection - Scraping



GitHub URL

Data Wrangling

The dataset includes various scenarios where the booster did not land successfully. For instance:

- **True Ocean**: The booster successfully landed in a designated ocean area.
- **False Ocean**: The booster attempted to land in the ocean but failed.
- **True RTLS (Return to Launch Site)**: The booster successfully landed on a ground pad.
- **False RTLS**: The booster failed to land on a ground pad.
- **True ASDS (Autonomous Spaceport Drone Ship)**: The booster successfully landed on a drone ship.
- **False ASDS**: The booster attempted to land on a drone ship but failed.

To simplify these outcomes for machine learning, they are converted into training labels:

- **1** indicates a successful landing.
- **0** represents an unsuccessful landing.

GitHub URL

EDA with Data Visualization

The following charts were plotted to explore the relationships between different variables:

- **Scatter plots**: These charts illustrate the correlation between variables. For example:
    - Flight Number vs. Payload Mass
    - Flight Number vs. Launch Site
    - Payload Mass vs. Launch Site
    - Flight Number vs. Orbit Type
    - Payload Mass vs. Orbit Type

    Scatter plots are useful in identifying patterns or relationships that could inform machine learning models if a significant correlation is found.
- **Bar charts**: These are used to compare discrete categories, such as:
    - Orbit Type vs. Success Rate The aim is to visualize the differences between categories and their associated values.
- **Line charts**: These charts show data trends over time, such as:
    - Success Rate Yearly Trend Line charts are effective in depicting changes over time and identifying trends, making them particularly useful for time series data.

GitHub URL

EDA with SQL

The following SQL queries were executed to analyze the SpaceX dataset:

- Retrieved the names of unique launch sites involved in space missions.
- Extracted five records where the launch site names start with "CCA."
- Calculated the total payload mass carried by NASA's CRS (Commercial Resupply Service) missions.
- Computed the average payload mass transported by boosters of version F9 v1.1.
- Identified the date of the first successful landing outcome on a ground pad.
- Listed boosters with successful drone ship landings and payloads between 4000 and 6000 kg.
- Summarized the total count of successful and failed mission outcomes.
- Identified booster versions that carried the maximum payload mass.
- Analyzed failed drone ship landing outcomes, including booster versions and launch sites for the year 2015.
- Ranked the counts of landing outcomes (e.g., failures on drone ships, successes on ground pads) between June 4, 2010, and March 20, 2017, in descending order.

GitHub URL

Build an Interactive Map with Folium

Markers for Launch Sites:

- **All Launch Sites**: Added markers using latitude and longitude coordinates for all SpaceX launch sites. Each marker includes a circle, a popup label, and a text label to display their geographic locations, along with their proximity to the equator and coastlines.
- **NASA Johnson Space Center**: Plotted a marker with a circle, popup label, and text label at its coordinates to serve as the starting location.

Colored Markers for Launch Outcomes:

- Used colored markers to represent launch outcomes for each site:
    - **Green markers**: Successful launches.
    - **Red markers**: Failed launches.
- Integrated marker clusters to visually group markers and identify which sites demonstrate higher success rates.

Distances from Launch Sites to Proximities:

- Drew colored lines to indicate distances between Launch Site **KSC LC-39A** (as an example) and nearby points of interest, including:
    - Railways.
    - Highways.
    - Coastlines.
    - The closest city.

This approach effectively visualizes the spatial relationships and key metrics for SpaceX launch operations.

GitHub URL

Build a Dashboard with Plotly Dash

Launch Site Dropdown List:

- Implemented a dropdown menu to allow users to select a specific launch site for analysis.

Pie Chart of Launch Success Rates:

- Added a pie chart that visualizes:
    - The total count of successful launches across all sites.
    - The breakdown of success vs. failure counts when a specific launch site is selected.

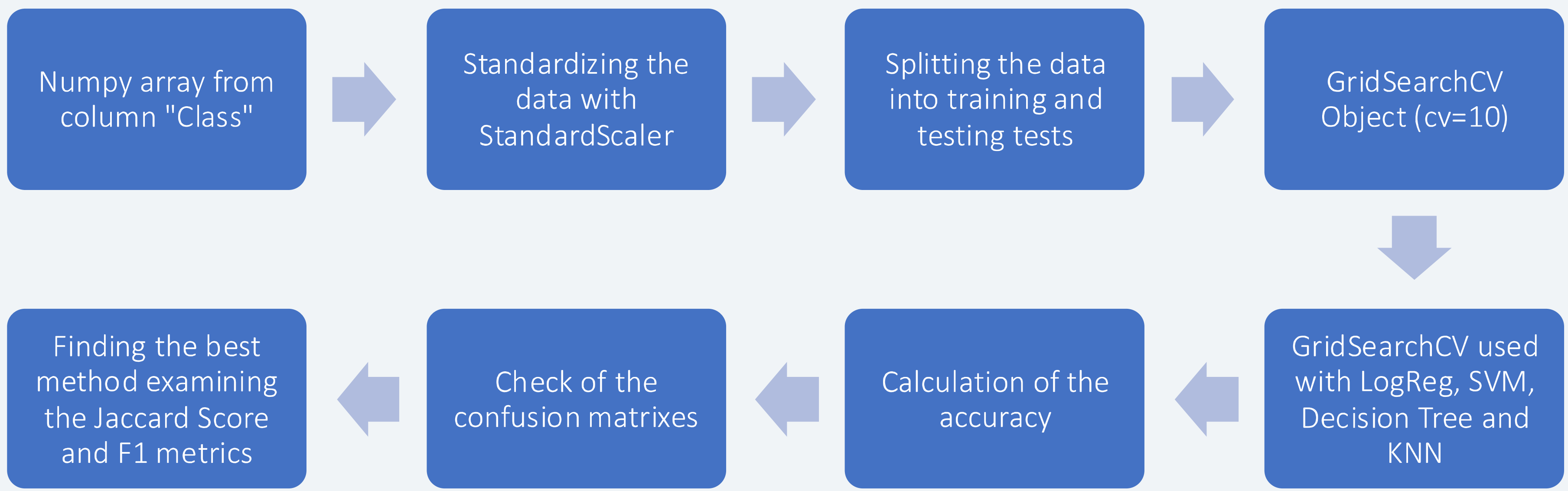
Payload Mass Range Slider:

- Incorporated a slider to let users define a payload mass range for filtering data.

Scatter Plot of Payload vs. Success Rate:

- Created a scatter plot to illustrate the relationship between payload mass and launch success rates for different booster versions, highlighting trends or patterns in performance.

Predictive Analysis (Classification)



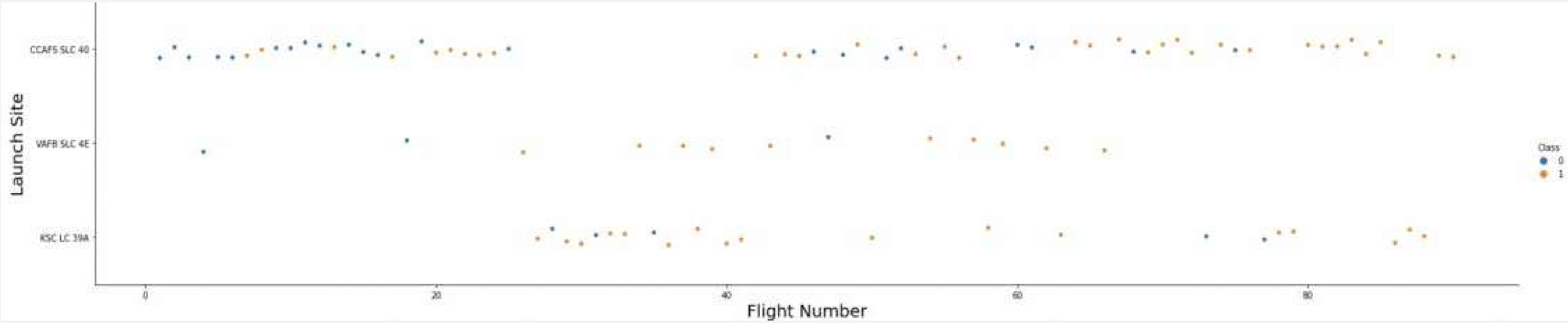
Results

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

Section 2

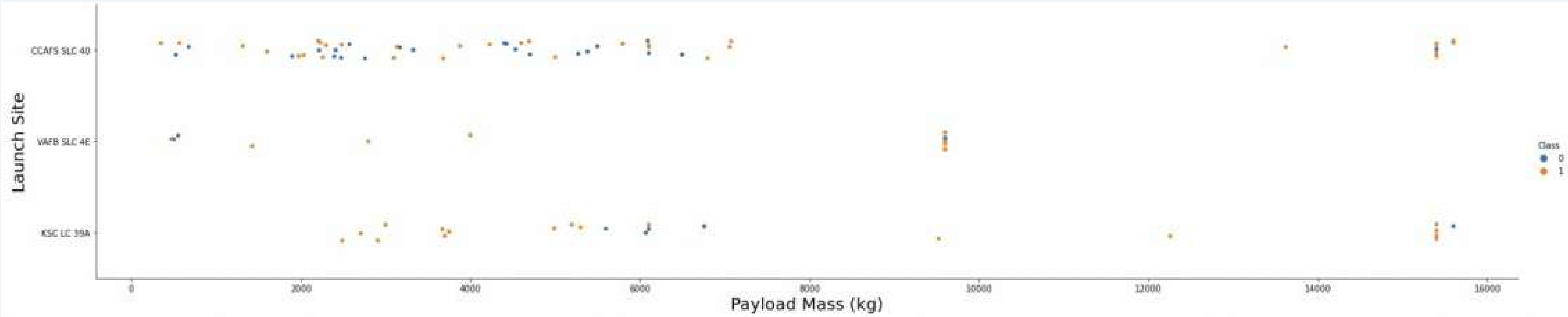
Insights drawn from EDA

Flight Number vs. Launch Site



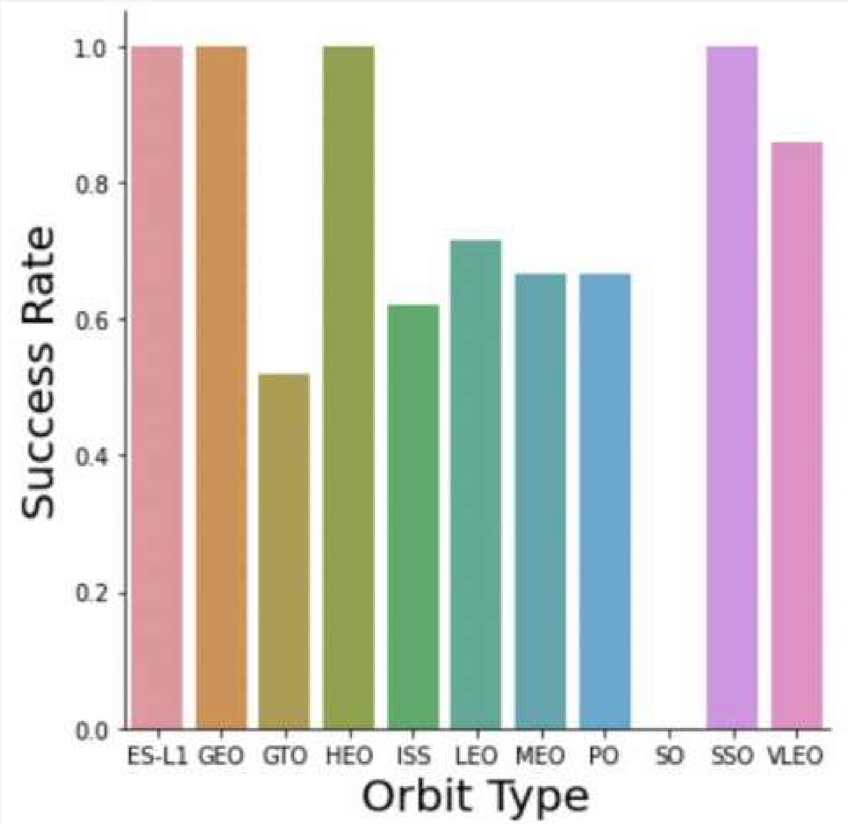
- Ealiest flights all failed
- Latest flights all succeeded
- CCAFS SLC 40 launch site had the half ot the launches

Payload vs. Launch Site



- The higher the payload mass, the higher the success rate
- KSC LC 39A has 100% success rate with payload mass < 5500kg
- Most of the launches > 7000kg were successful

Success Rate vs. Orbit Type



100% success rate on :

- ES-L1, GEO, HEO, SSO

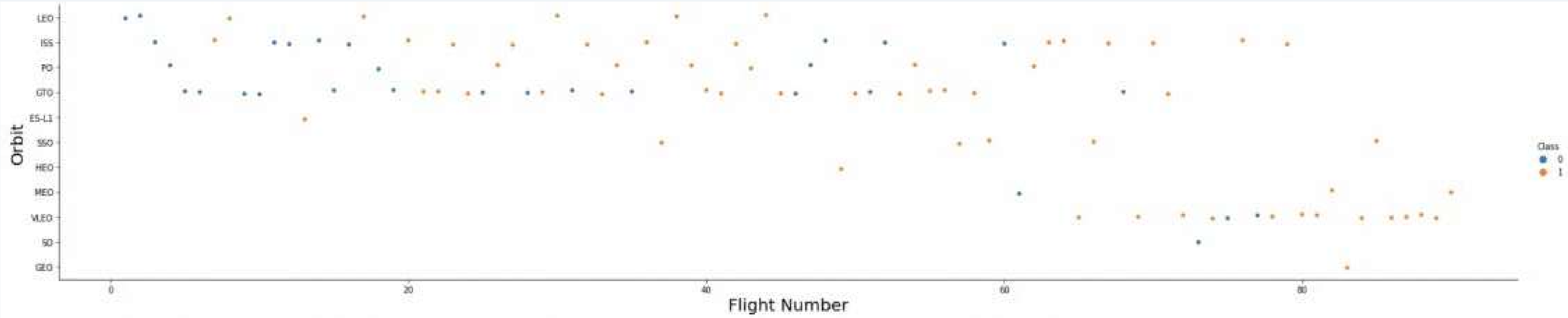
0% success rate on :

- SO

Between 50% and 85% rate :

- GTO, ISS, LEO, MEO, PO

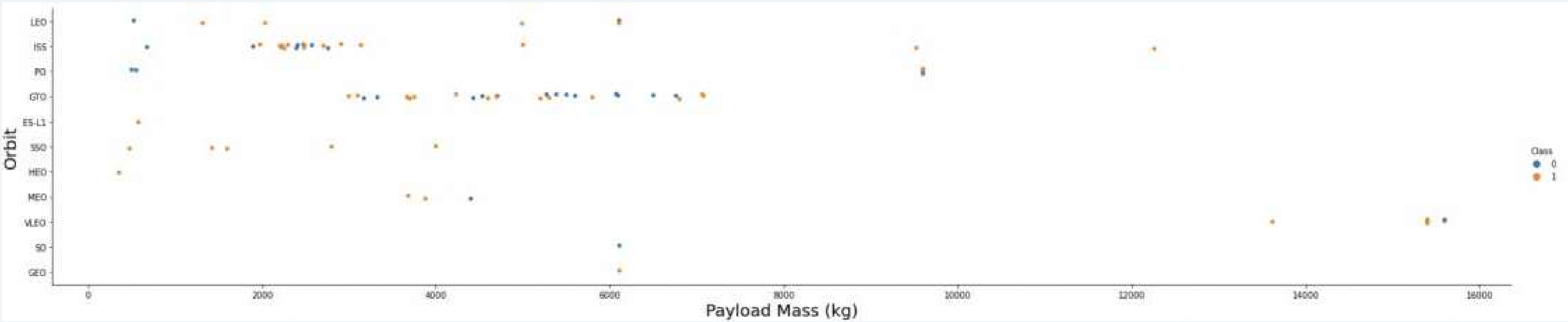
Flight Number vs. Orbit Type



In LEO orbit, the success apperas related to number of flights

No relationship betwee flight number when in GTO orbit

Payload vs. Orbit Type

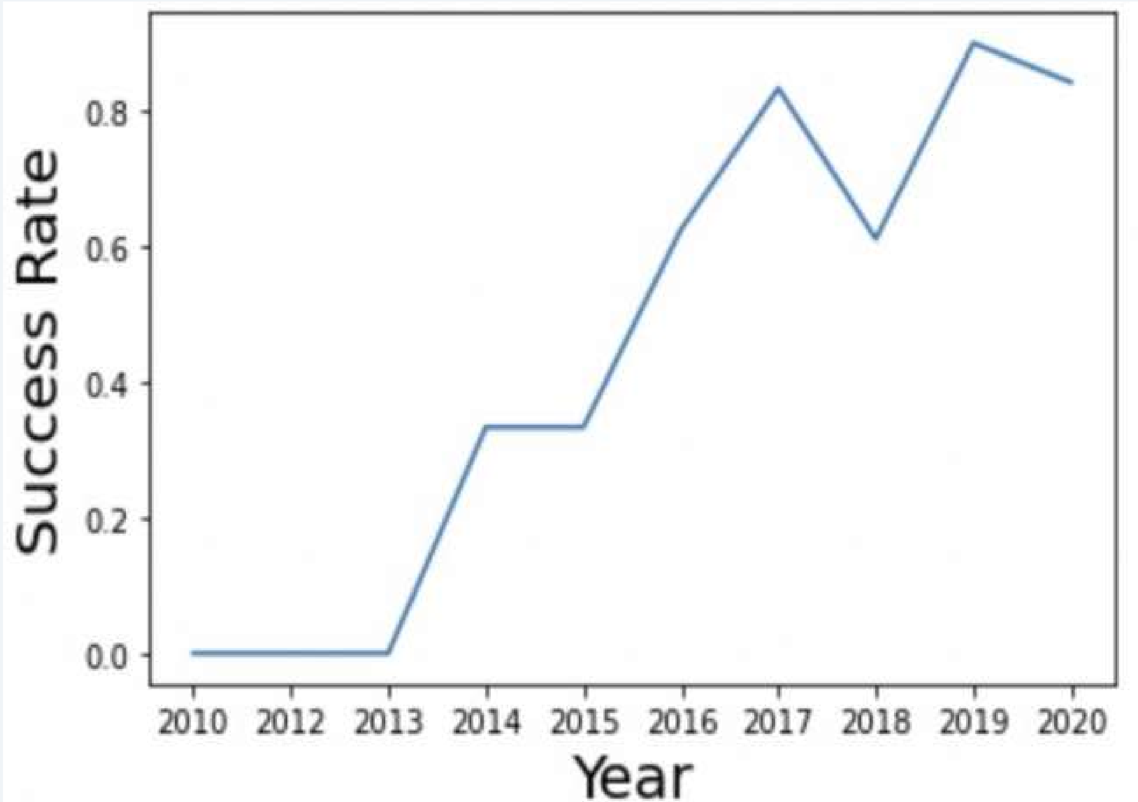


Heavy payloads works less on GTO orbits

They however works better on GTO and ISS orbits

Launch Success Yearly Trend

- The success rate is improving since 2013
- Exception in 2018



All Launch Site Names

```
%sql select distinct launch_site from SPACEXDATASET;
```

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

- SQL request to get the names of the unique launch sites

Launch Site Names Begin with 'CCA'

```
%sql select * from SPACEXDATASET where launch_site like 'CCA%' limit 5;
```

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	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	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

- 5 records with a launch site name beginning with "CCA"

Total Payload Mass

```
%sql select sum(payload_mass__kg_) as total_payload_mass from SPACEXDATASET where customer = 'NASA (CRS)';
```

total_payload_mass
45596

- The total payload Mass is 45,596kg

Average Payload Mass by F9 v1.1

```
%sql select avg(payload_mass__kg_) as average_payload_mass from SPACEXDATASET where booster_version like '%F9 v1.1%';
```

average_payload_mass
2534

- The average Payload Mass is 2,534kg

First Successful Ground Landing Date

```
%sql select min(date) as first_successful_landing from SPACEXDATASET where landing__outcome = 'Success (ground pad)';
```

first_successful_landing
2015-12-22

- The first successful ground landing date wath on 2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%sql select booster_version from SPACEXDATASET where landing__outcome = 'Success (drone ship)' and payload_mass__kg_ between 4
000 and 6000;
```

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

- Displaying the names of the successful Drone Ship Landing with payload betweed 4000 and 6000kg

Total Number of Successful and Failure Mission Outcomes

```
%sql select mission_outcome, count(*) as total_number from SPACEXDATASET group by mission_outcome;
```

mission_outcome	total_number
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

- 1 Failure in flight, 99 successes

Boosters Carried Maximum Payload

```
%sql select booster_version from SPACEXDATASET where payload_mass__kg_ = (select max(payload_mass__kg_) from SPACEXDATASET);
```

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

- Names list of the boosters that carried the maximum Payload

2015 Launch Records

```
%%sql select monthname(date) as month, date, booster_version, launch_site, landing__outcome from SPACEXDATASET
     where landing__outcome = 'Failure (drone ship)' and year(date)=2015;
```

MONTH	DATE	booster_version	launch_site	landing__outcome
January	2015-01-10	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
April	2015-04-14	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

- List of the 2015 Launch Records

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

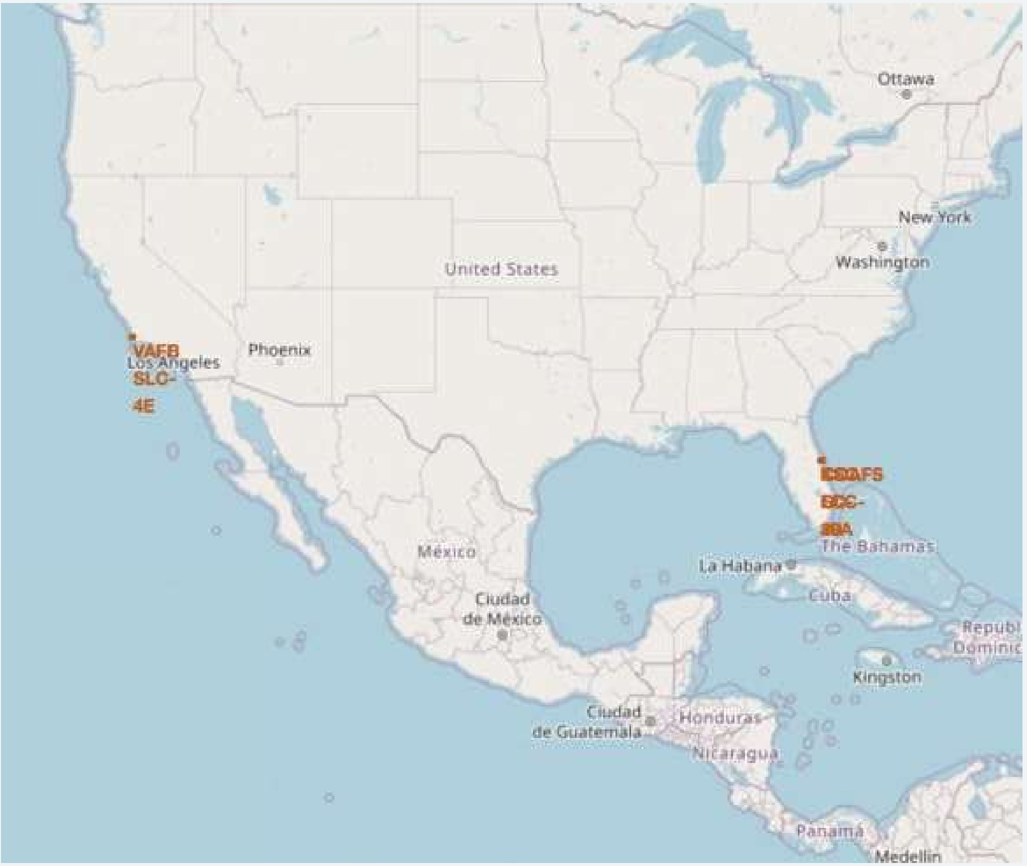
```
%%sql select landing__outcome, count(*) as count_outcomes from SPACEXDATASET
      where date between '2010-06-04' and '2017-03-20'
      group by landing__outcome
      order by count_outcomes desc;
```

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

- Number of landing per outcomes within 2010-06-04 and 2017-03-20

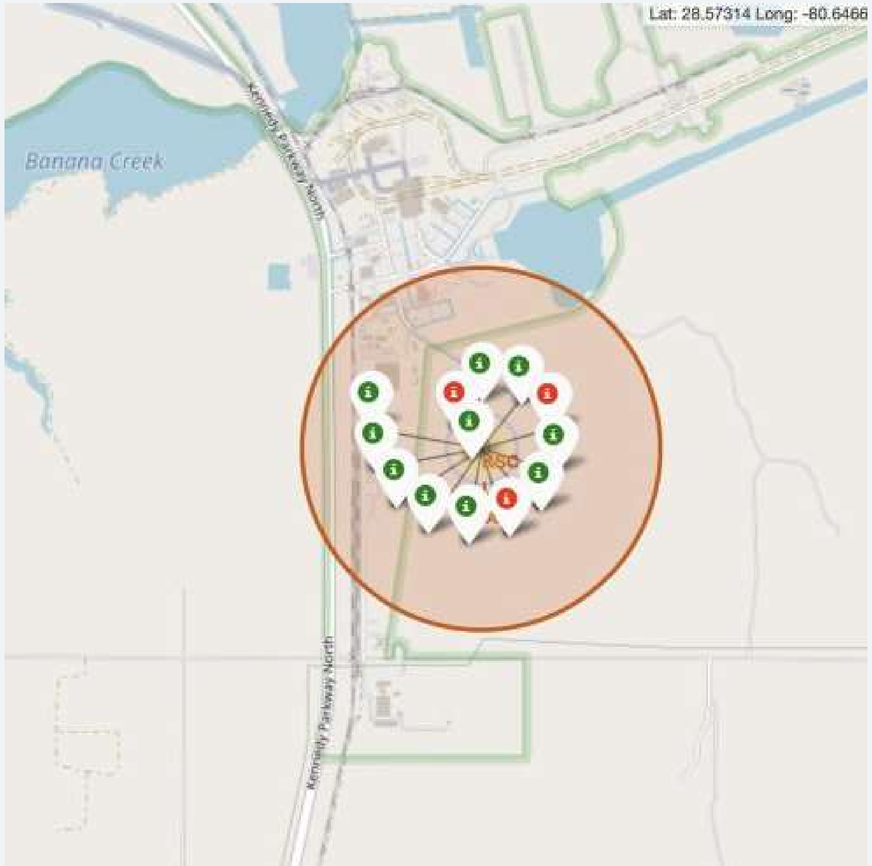

Section 3
Launch Sites
Proximities Analysis

<Folium Map Screenshot 1>



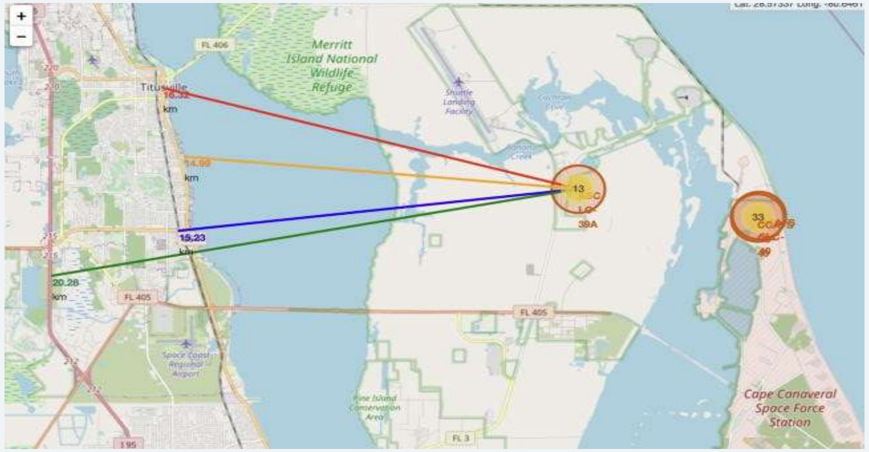
- **Proximity to the Equator**: Most launch sites are located near the equator because Earth's rotational speed is greatest there, at approximately 1670 km/hour. This speed is imparted to rockets at launch, providing an additional velocity boost due to inertia. This advantage helps spacecraft achieve and maintain the speed needed to stay in orbit with reduced fuel consumption.
- **Coastal Proximity**: All launch sites are situated near coastlines to ensure that rockets can be launched over the ocean. This minimizes the risk of debris falling or accidents occurring near populated areas, enhancing safety for nearby communities.

<Folium Map Screenshot 2>



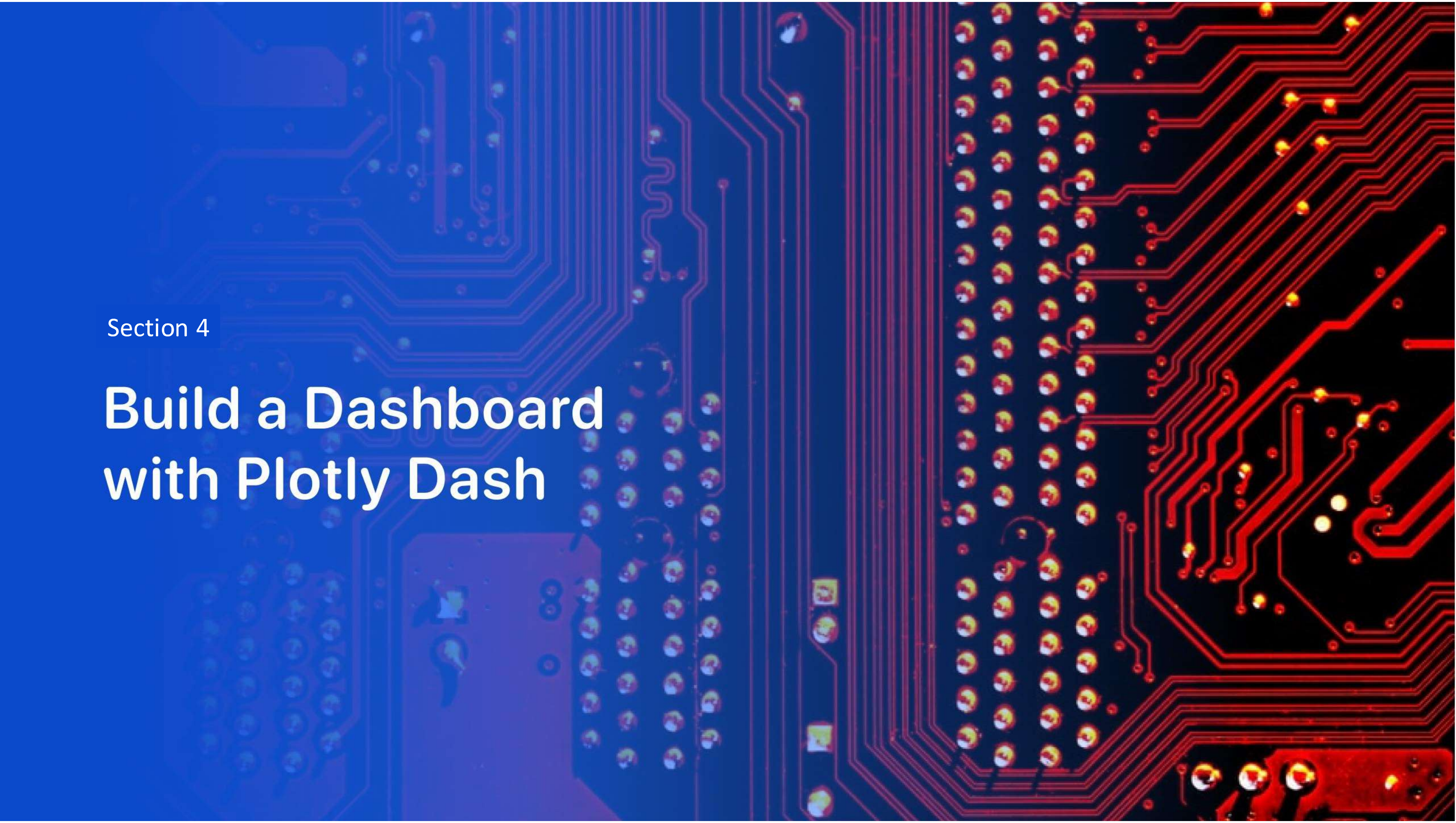
- **Green markers : Successful launches**
- **Red markers : Failed launches**

<Folium Map Screenshot 3>



- Launch Sites are far from the nearest railways, highways and neer to the coastline.

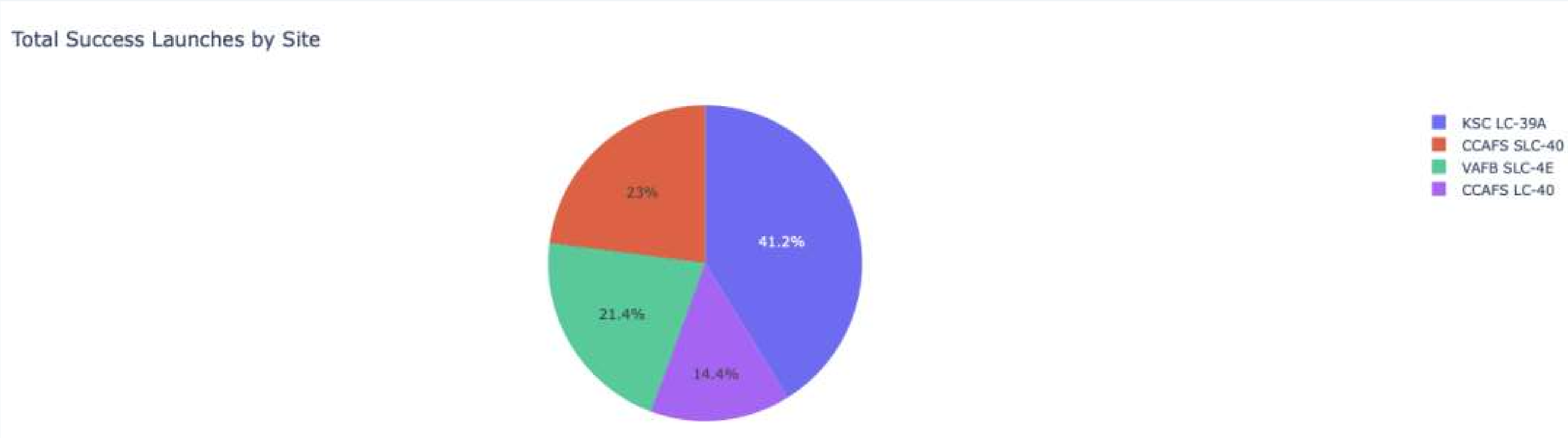
- The aim si to prevent damages on human installations if there is an accident



Section 4

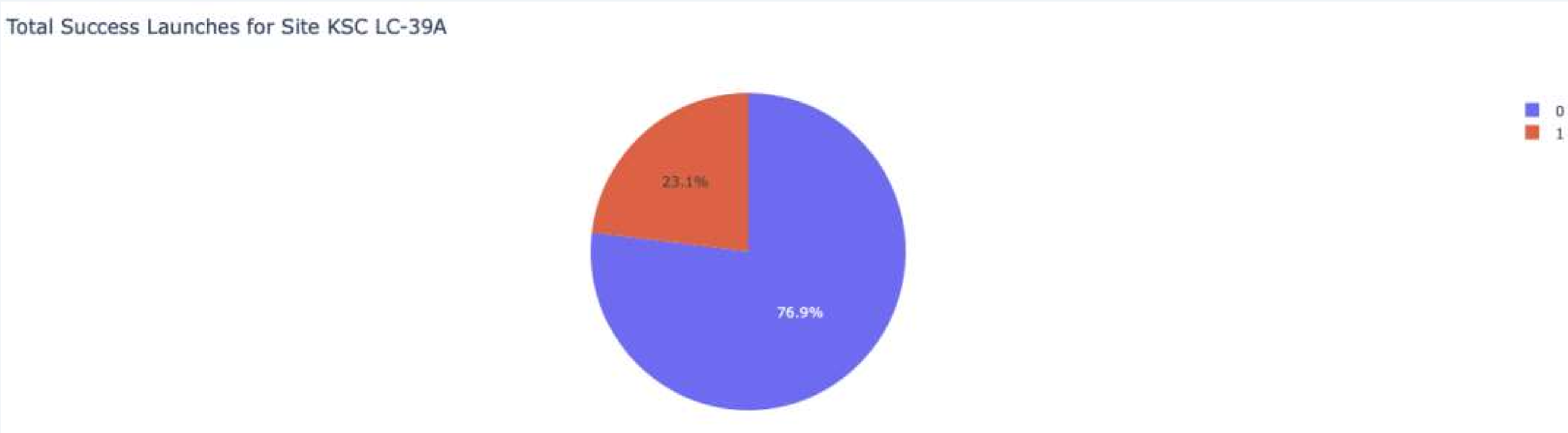
Build a Dashboard with Plotly Dash

Launch success per site



- KSC LC-39A has the most successful launches

Launch site with highest success rate



- KSC LC-39A has the highest success rate with 76.9%

Payload Mass vs Launch Outcome

- Payloads between 2000 and 5500kg have the best success rate

Section 5

Predictive Analysis (Classification)

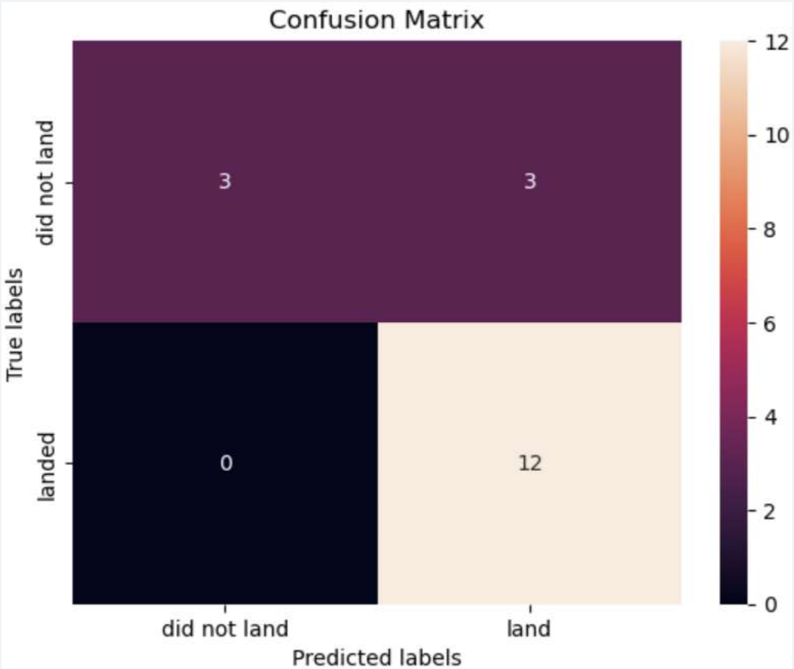
Classification Accuracy

- The best model is the Decision Tree Model with a 91.11% accuracy

```
LogReg accuracy : 0.8222222222222222
SVM accuracy : 0.8482142857142856
KNN accuracy : 0.8482142857142858
Tree Decision accuracy : 0.8607142857142858
```

Confusion Matrix

- Confusion matrix of the Tree Decision Model :



Conclusions

- The Decision Tree Model is the most effective algorithm for this dataset.
- The success rate of launches has improved over the years.
- Launches with a lower payload mass yield better results compared to those with a higher payload mass.
- KSC LC-39A has the highest launch success rate among all sites.
- The majority of launch sites are located near the Equator and are all in close proximity to the coast.
- The orbits ES-L1, GEO, HEO, and SSO have achieved a 100% success rate.


Thank you!