IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Stuart Clayton
Dec 2, 2021

Outline

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

Executive Summary

- Summary of methodologies

  Collection of data from unofficial SpaceX API

  Collection of data from web scraping

- Summary of all results

  - Launch site KSC LC-39A has the highest success rate
  - More launches occur from Florida launch sites than from California.
  - The success of launches increases continuously and considerably from 2013 up to 2020.
  - All launch sites are located on coasts.
  - The Decision Tree model was the best performing.

Introduction

- Project background and context
  - SpaceX is arguably the most successful private rocket company today
  - They were the first to reuse a orbital booster
  - Reusable boosters is a key pillar of their business model and success
- Problems you want to find answers
  - What makes their launches successful?
  - How many successful launches do they have?
  - How many successful landings do they have?
  - What is the true cost for launches?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was collected via unofficial SpaceX API and web scraping
- Perform data wrangling
  - Data was processed 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
  - K-nearest Neighbors, Logistic Regression, Decision Tree, Support Vector Machine

Data Collection

- Data was collected from https://api.spacexdata.com/v4/launches/past
- Additional data was collected from web scraping Wiki pages

Data pulled from API using requests library.

↓

Data frame created and columns of interest isolated.

↓

Data cleaned using Pandas.

Data scraped from wiki pages using beautiful soup.

↓

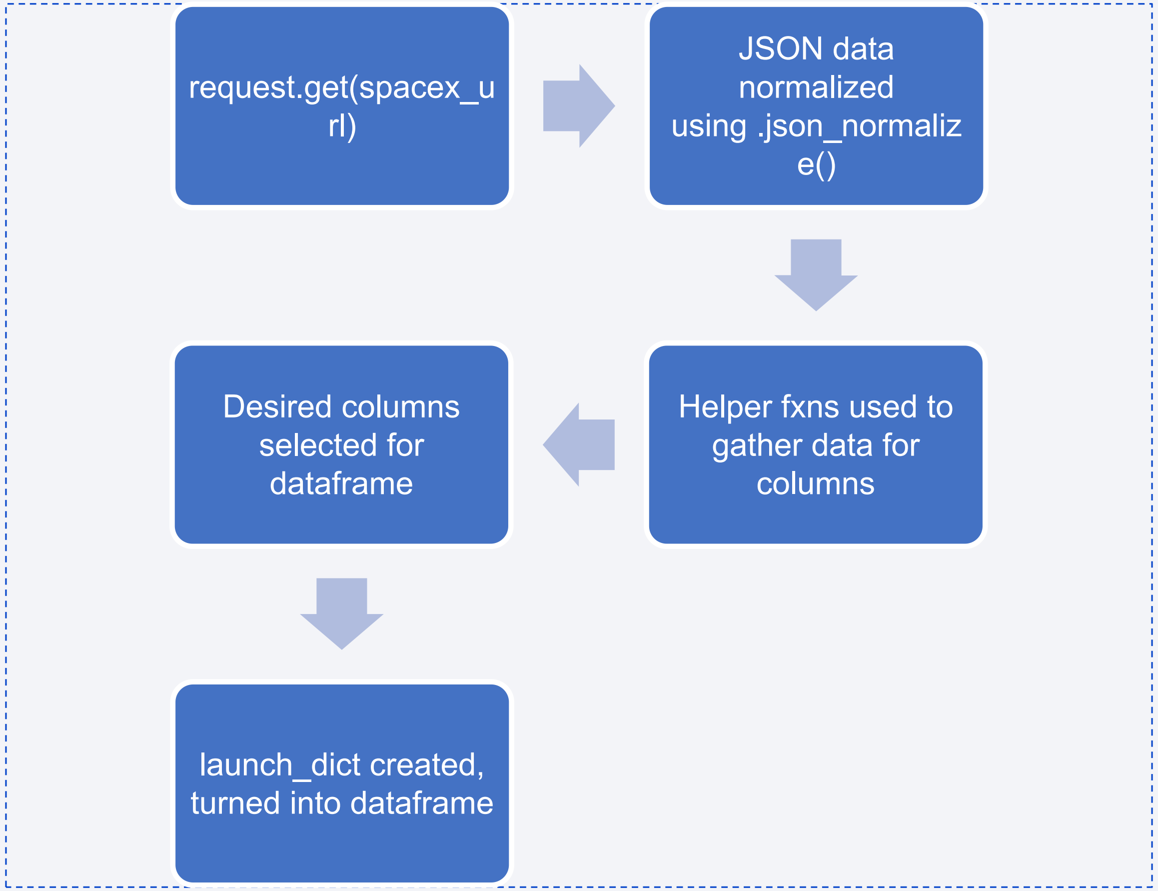
Data frame created from HTML.

↓

Data frame cleaned by filtering through tables.

Data Collection – SpaceX API

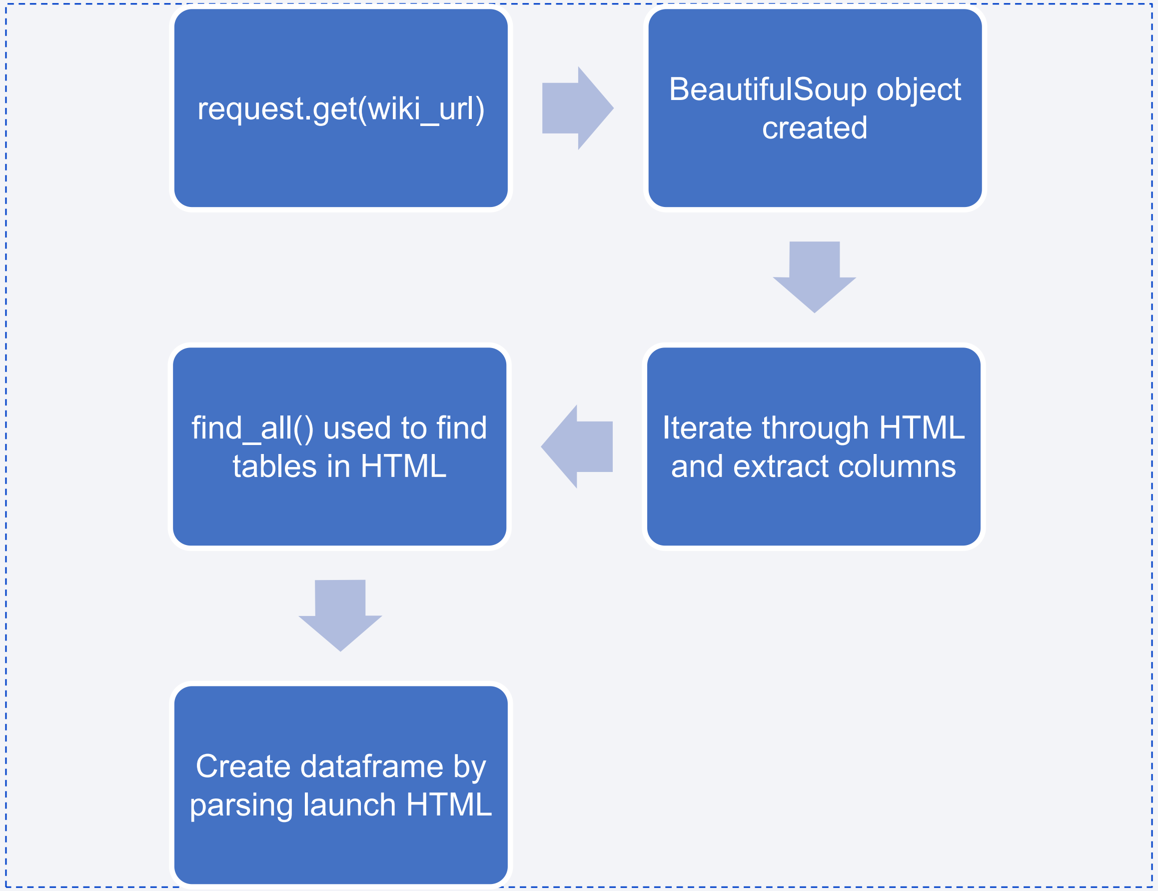
- SpaceX launch data extracted from API using helper functions.
- Cleaning was done using Pandas.
- [Github link to SpaceX API call notebook.]()



Data Collection – Scraping

- Additional data was gathered by web scraping Wikipedia.
- BeautifulSoup was used to help process the raw HTML.

- [Github link to Web Scraping notebook.]



Data Wrangling

- Data analysis primarily acheived using Pandas
  - Identify missing values and data types for each column
  - Use .value_counts() to calculate number of launches from each site.
  - Use .value_counts() to explore the number of launches for different orbital targets.
  - Identify the different landing outcomes and the number of each.
- [Github link to Data Wrangling notebook.]()

EDA with Data Visualization

- The primary charts used were scatter plot, bar and line.
  - Scatter plots were useful for:
    - Comparing Flight Number to Payload Mass
    - Comparing Flight Number to Launch Site
    - Comparing Payload Mass to Launch Site
  - Bar and line charts were used to compare Success Rate to Year
- GitHub link to EDA with Visualization notebook.

EDA with SQL

- Identify unique launch sites
- Total payload mass launched by NASA
- First successful landing outcome on ground pad
- Total succesful and failed mission outcomes
- Booster version which have carried maximum payload
- Failed landing outcome on drone ship
- Landing outcomes, between the date 2010-06-04 and 2017-03-20, in descending order
- GitHub link to EDA with SQL notebook.

Build an Interactive Map with Folium

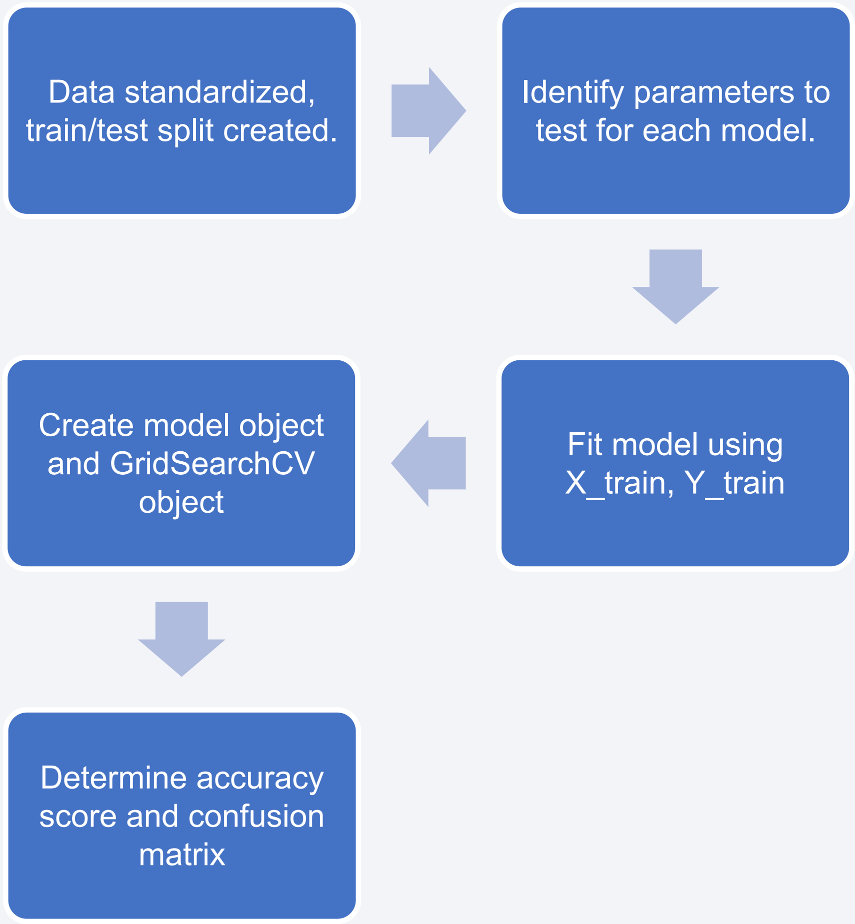
- Add a Circle to better visualize on the map where the location is. Include popup with detailed description.
- Add a Marker to better pinpoint the exact location.
- Add MarkerCluster for areas with a lot of markers for better clarity.
- Add MousePosition to extract latitude and longitude from map.
- Add PolyLine to visualize distances between points.
- GitHub link to Interactive Map with Folium notebook.

Build a Dashboard with Plotly Dash

- The dashboard provides two main graphical views.
- The pie chart displays the percentage of succssful or failed launches at each site.
- The scatter plot displays the Payload Mass and Class for each booster.
- GitHub link to Dashboard with Plotly Dash python file.

Predictive Analysis（Classification）

- Models were all created and developed using sklearn.
- Data standardized using StandardScaler()
- Train/test splits created from dataset
- Different models tests (KNN, decision tree, logistic regression, SVM)
- [GitHub link to Predictive Analysis notebook.]

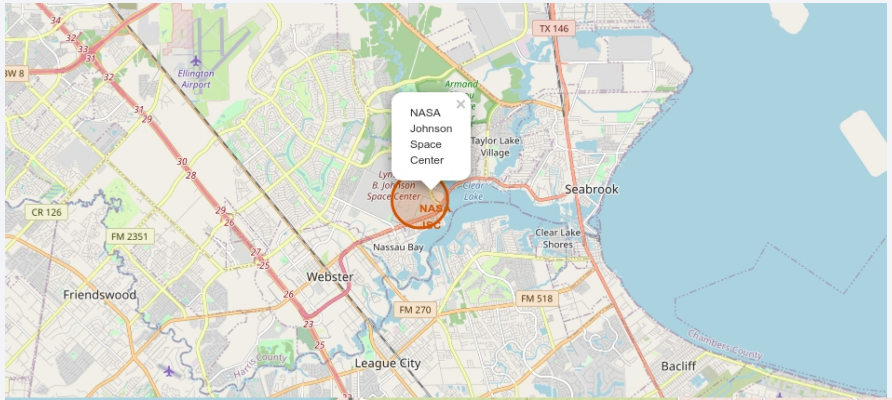


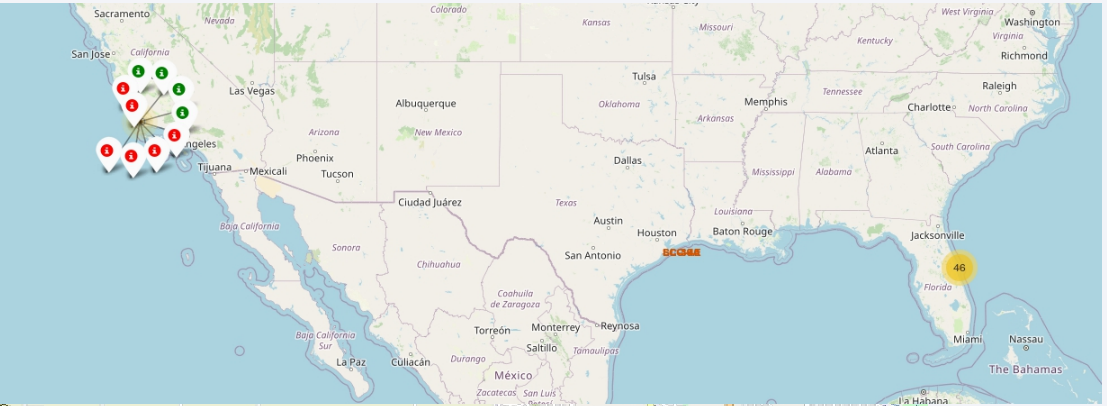
Results

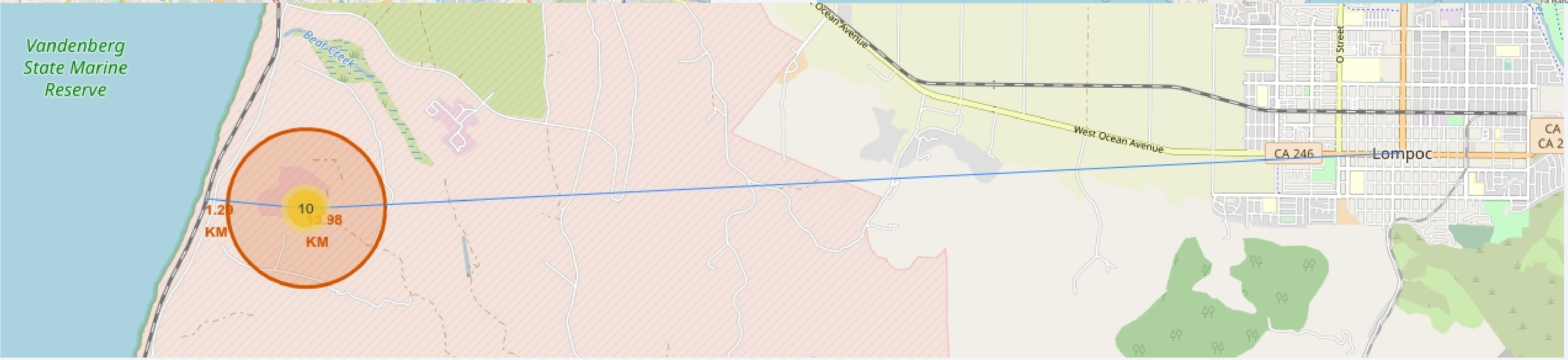
- Exploratory data analysis results
  - Launch sites: CCAFS LC-40, CCAFS SLC-40, KSC LC-39A, VAFB SLC-4E
  - Payload mass launched at each site varies
  - ES-L1, GEO, SSO, and HEO orbits have 100% success rate
  - Launch success trend increases from 2013 until 2020 (end of data)
- Predictive analysis results
  - The best performing algorithm was Decision Tree.

Results

- Interactive analytics





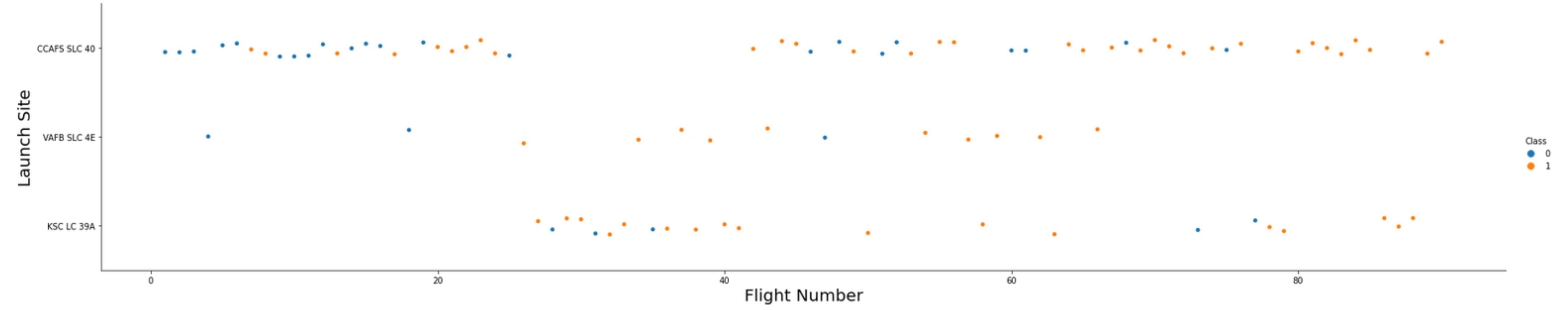


Section 2

Insights drawn from EDA

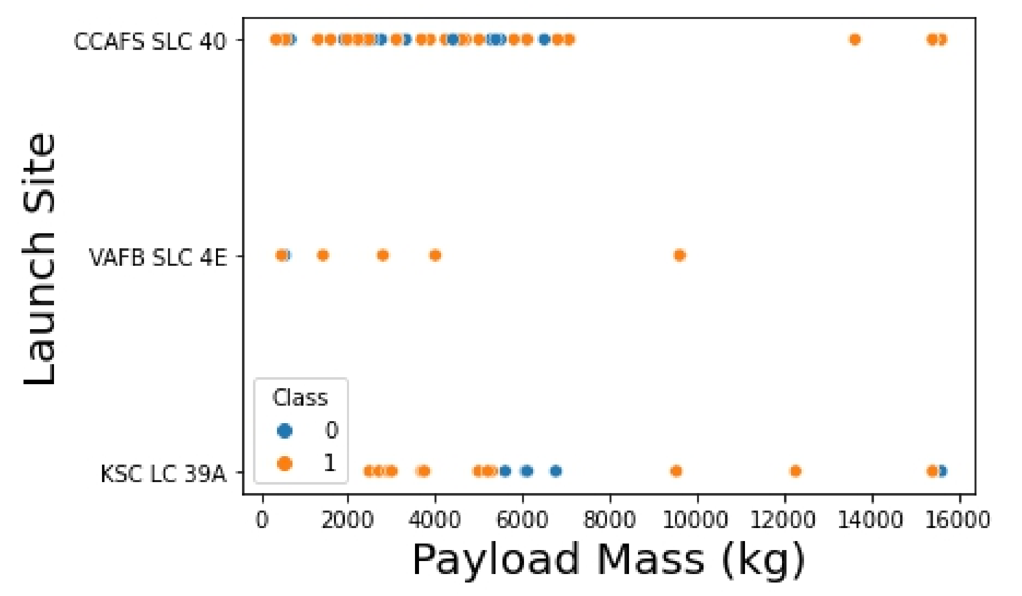
Flight Number vs. Launch Site

Launch site VAFB SLC-4E appears to have the least launches. Launches from KSC LC-39A began occuring later into the companies history.



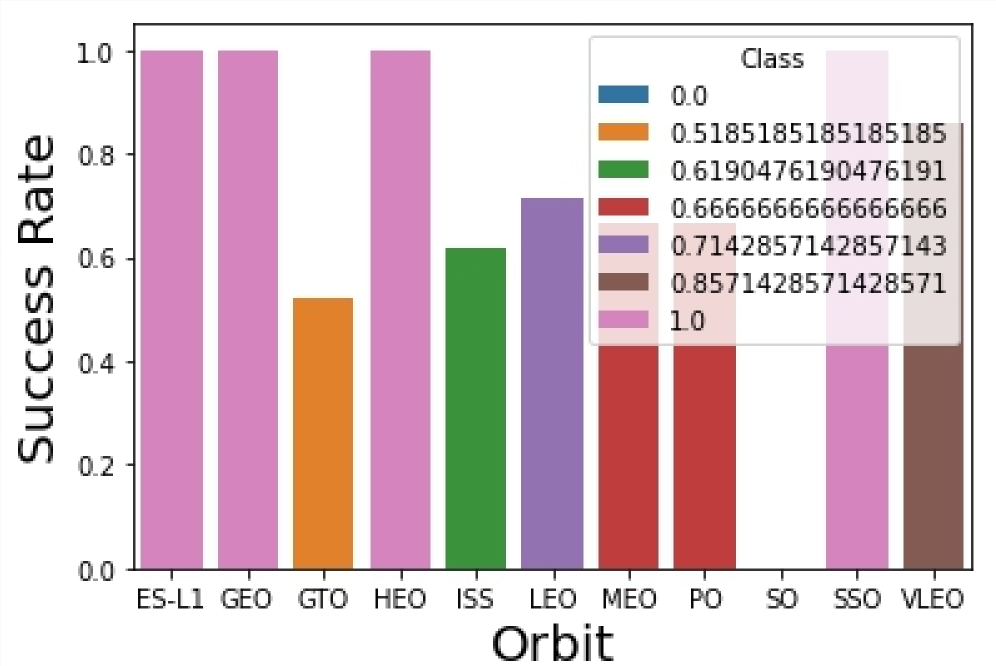
Payload vs. Launch Site

- VAFB SLC-4E has no launches with a payload mass over 10000 kg.
- Both CCAFS SLC-40 and KSC LC-39A support heavy launches up to around 16000 kg.



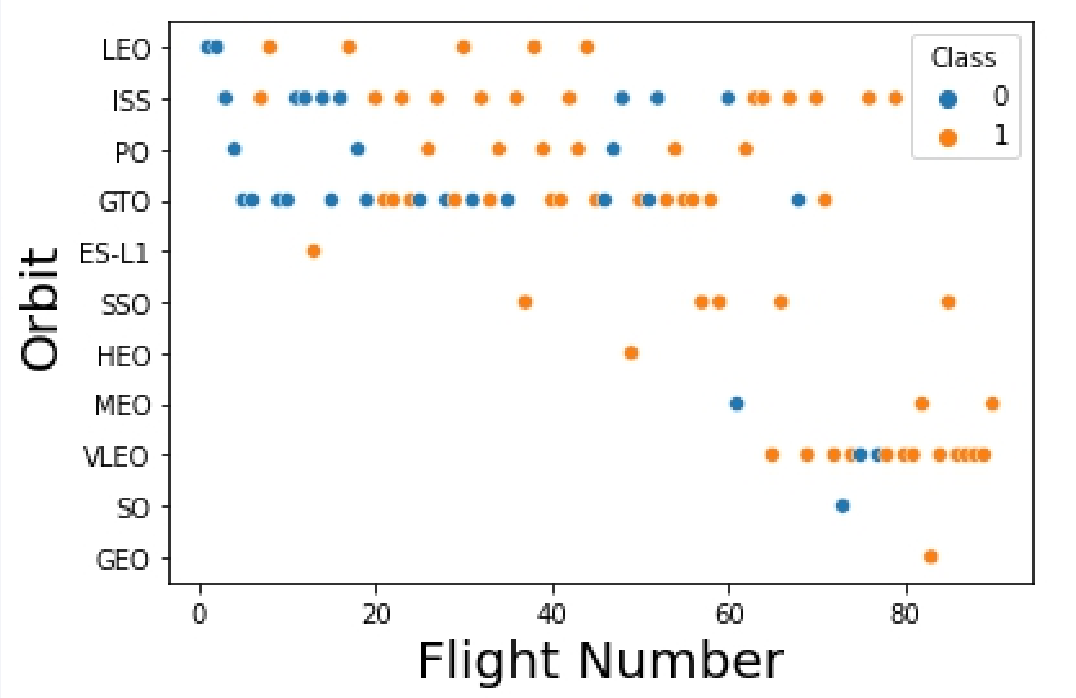
Success Rate vs. Orbit Type

- Orbits ES-L1, GEO, HEO, SSO have a 100% success rate.
- GTO and ISS have a lower success rate at around 50% to 60%.



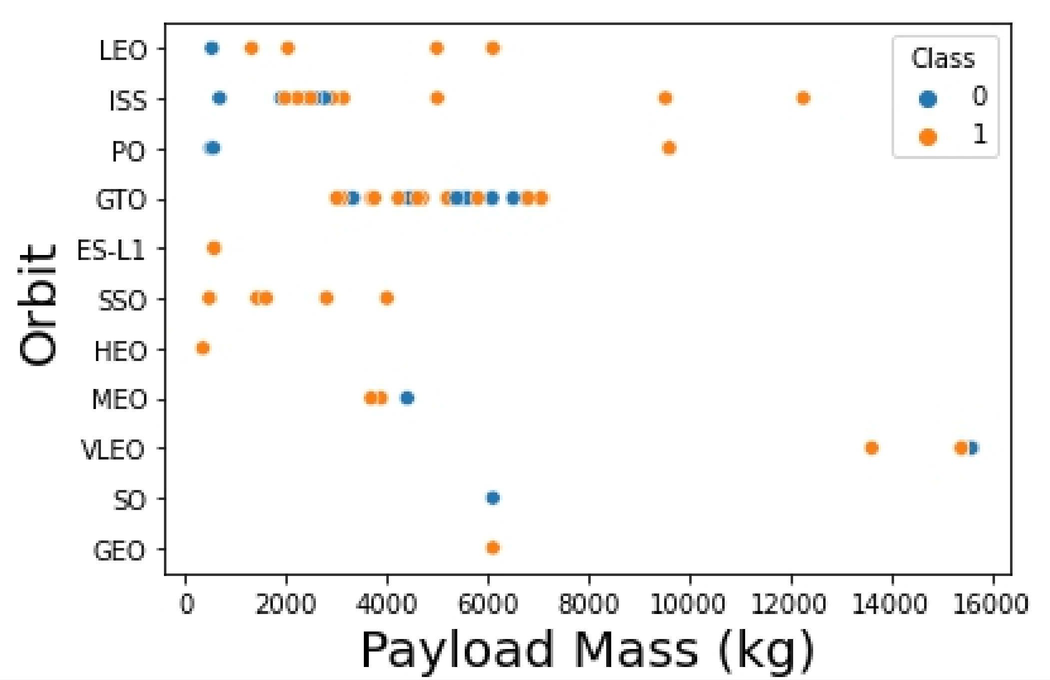
Flight Number vs. Orbit Type

- For LTO, the success appears to be related to the flight number.
- GTO appears to have no correlation, as do most other orbit types.



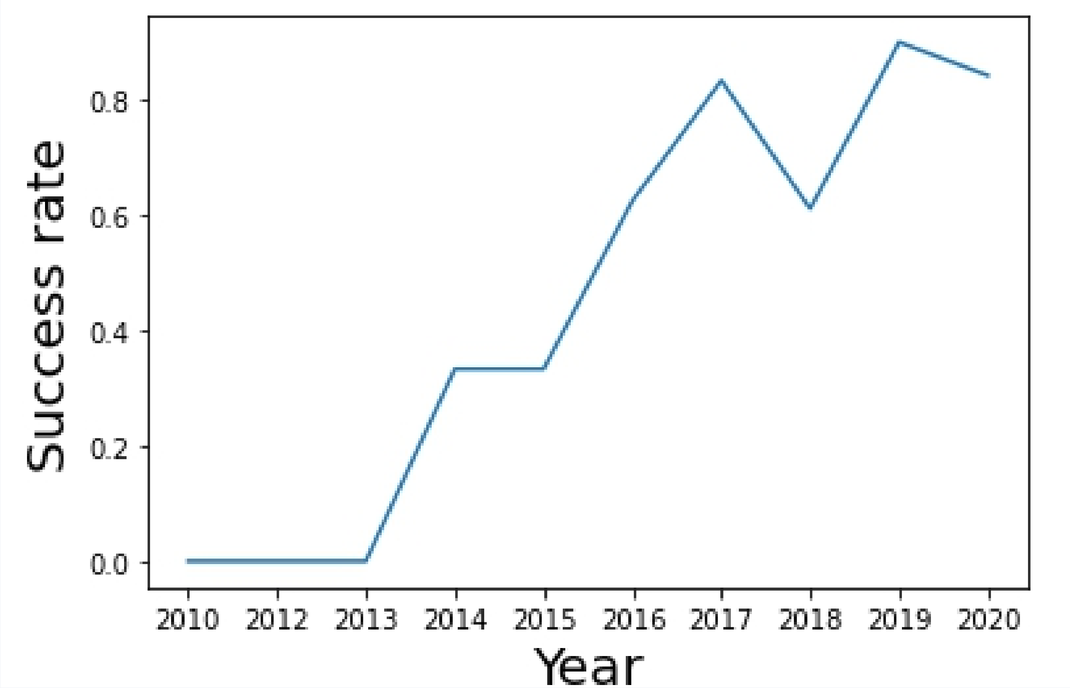
Payload vs. Orbit Type

- PO, LEO, and ISS have higher success rates with a heavier payload.
- GTO, GEO, SO and others have no apparent correlation.



Launch Success Yearly Trend

- Success rate increases starting in 2013 up untill 2020, when the data collected ends.



All Launch Site Names

There are 4 unique launch sites.

```
%sql SELECT UNIQUE(launch_site) FROM SPACEXTBL;
```

```
 * ibm_db_sa://kxs69266:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:31321/BLUDB
Done.
```

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

Launch Site Names Begin with 'CCA'

Find 5 records where launch sites begin with `CCA`

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

```
 * ibm_db_sa://kxs69266:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:31321/BLUDB
Done.
```

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

Total Payload Mass

Calculate the total payload carried by boosters from NASA. Number given is in kg.

```
%sql SELECT SUM(payload_mass__kg_) FROM SPACEXTBL WHERE customer LIKE 'NASA (CRS)';
```

```
 * ibm_db_sa://kxs69266:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3sd0tgtu0lqde00.dat
Done.
```

1
45596

Average Payload Mass by F9 v1.1

Calculate the 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';
```

```
 * ibm_db_sa://kxs69266:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:31321/BLUDB
Done.
```

1
2928

First Successful Ground Landing Date

Find the dates of the first successful landing outcome on ground pad

```
%sql SELECT MIN(date) FROM SPACEXTBL WHERE landing__outcome LIKE 'Success (ground pad)';
```

```
 * ibm_db_sa://kxs69266:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:31321/BLUDB
Done.
```

1
2015-12-22

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

```
%sql SELECT booster_version FROM SPACEXTBL WHERE landing__outcome LIKE 'Success (drone ship)' AND payload_mass__kg_ BETWEEN 4001 AND 5999;
```

```
 * ibm_db_sa://kxs69266:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:31321/BLUDB
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

Calculate the total number of successful and failure mission outcomes

```
%sql SELECT count(*) FROM SPACEXTBL WHERE mission_outcome LIKE '%Success%';
```

```
 * ibm_db_sa://kxs69266:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:31321/BLUDB
Done.
```

1
100

```
%sql SELECT count(*) FROM SPACEXTBL WHERE mission_outcome LIKE '%Failure%';
```

```
 * ibm_db_sa://kxs69266:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:31321/BLUDB
Done.
```

1
1

Boosters Carried Maximum Payload

List the names of the booster which have carried the maximum payload mass

```
%sql SELECT booster_version FROM SPACEXTBL WHERE payload_mass__kg_ = (SELECT MAX(payload_mass__kg_) FROM SPACEXTBL);
```

```
 * ibm_db_sa://kxs69266:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:31321/BLUDB
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 failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

```
%sql SELECT landing__outcome, booster_version, launch_site FROM SPACEXTBL WHERE landing__outcome LIKE '%Fail%' AND date LIKE '%2015%';
```

```
 * ibm_db_sa://kxs69266:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:31321/BLUDB
Done.
```

landing__outcome	booster_version	launch_site
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

No attempt had the most landings outcomes, Success and failure drone ship both had 5. All others had 3 or fewer attempts.

```
%sql SELECT landing__outcome, COUNT(*) FROM SPACEXTBL WHERE date BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY landing__outcome;
```

```
 * ibm_db_sa://kxs69266:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:31321/BLUDB
Done.
```

landing__outcome	2
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 4

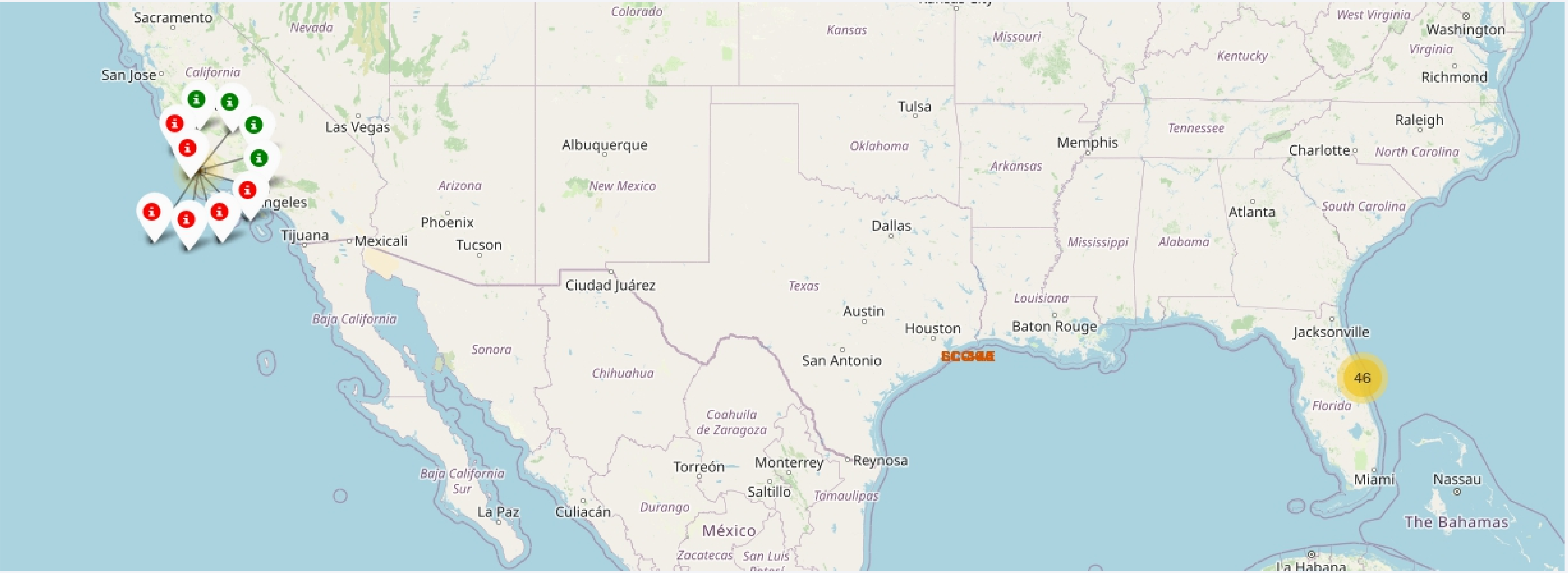
Launch Sites Proximities Analysis

Launch sites US map

Displays the different launch sites in the US

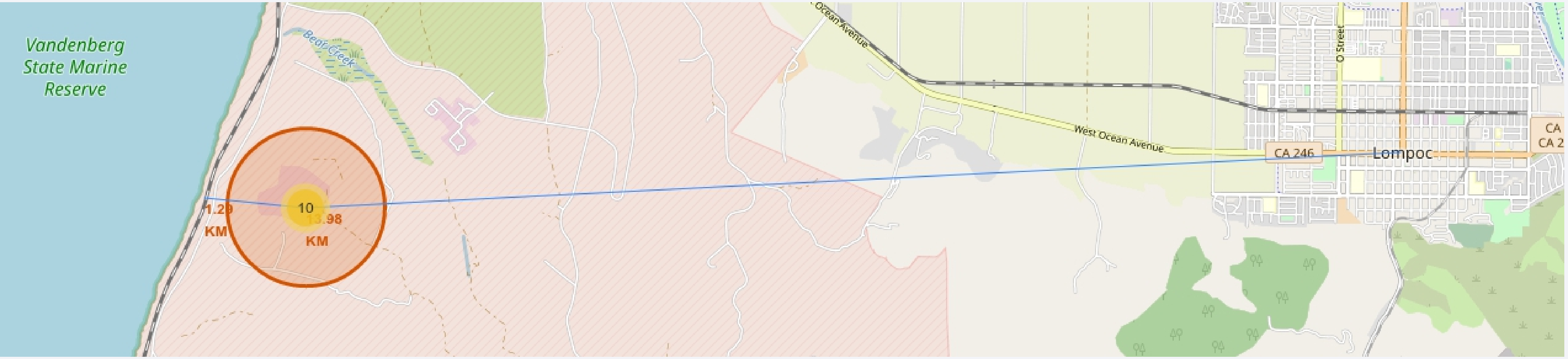
Success Vs Failed launches

The markers indicate if the launch was successful (green) or failed (red)



Launch site proximity map

Shows the proximity of VAFB to the nearest road and to the nearest city

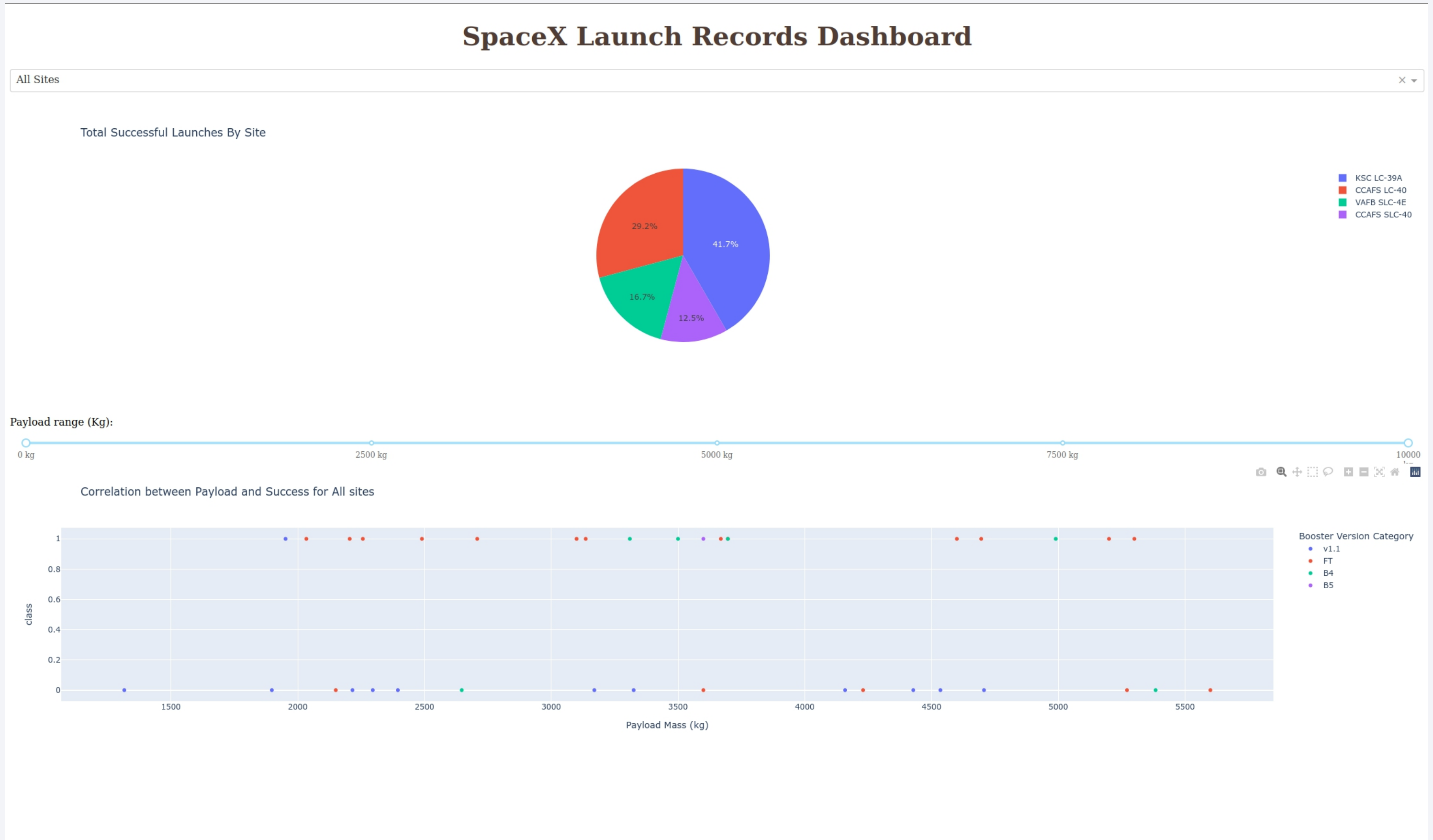


Section 5

Build a Dashboard with Plotly Dash

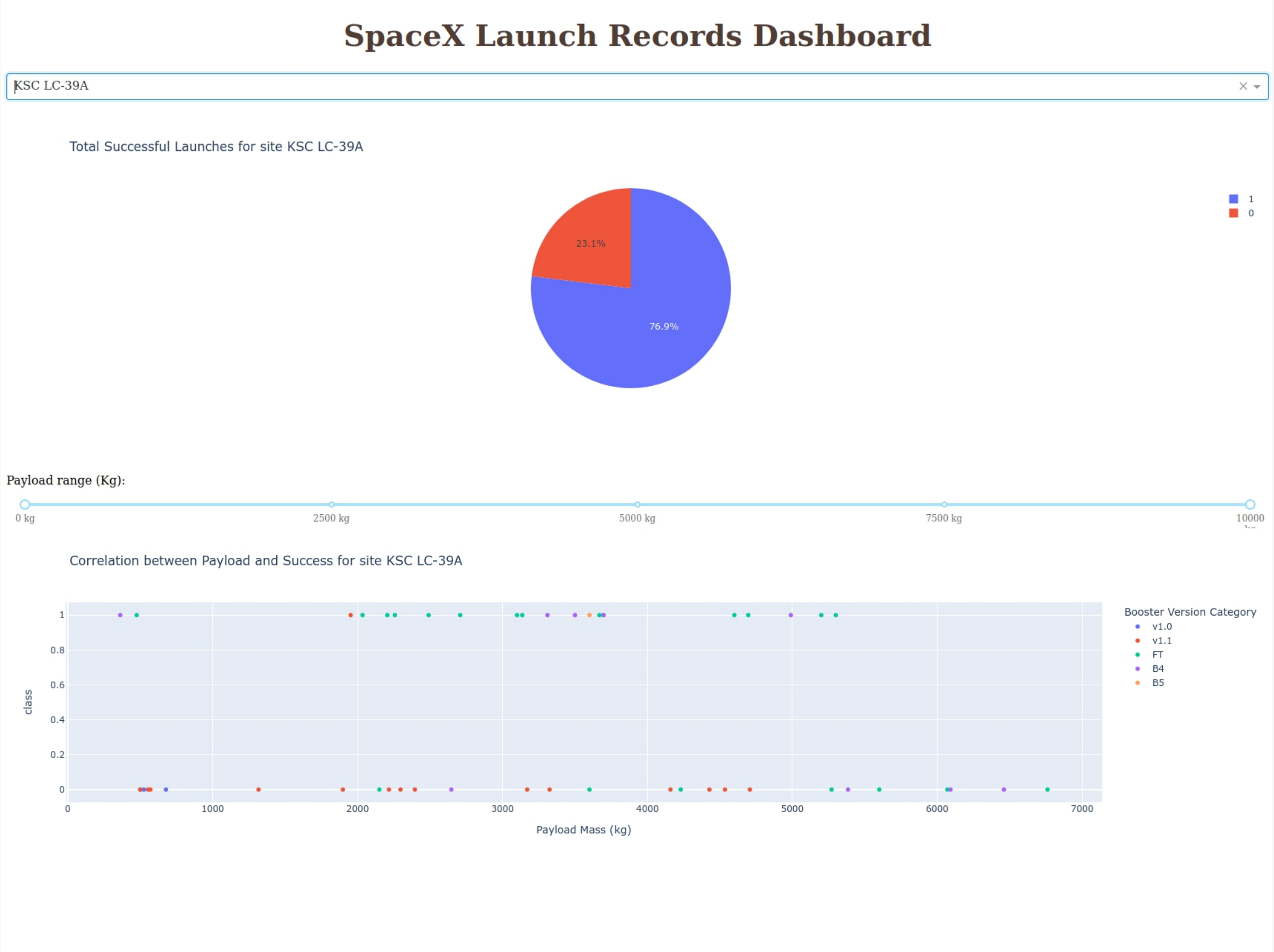
Dashboard Overview

Over 40% of successful launches are from KSC LC-39A.



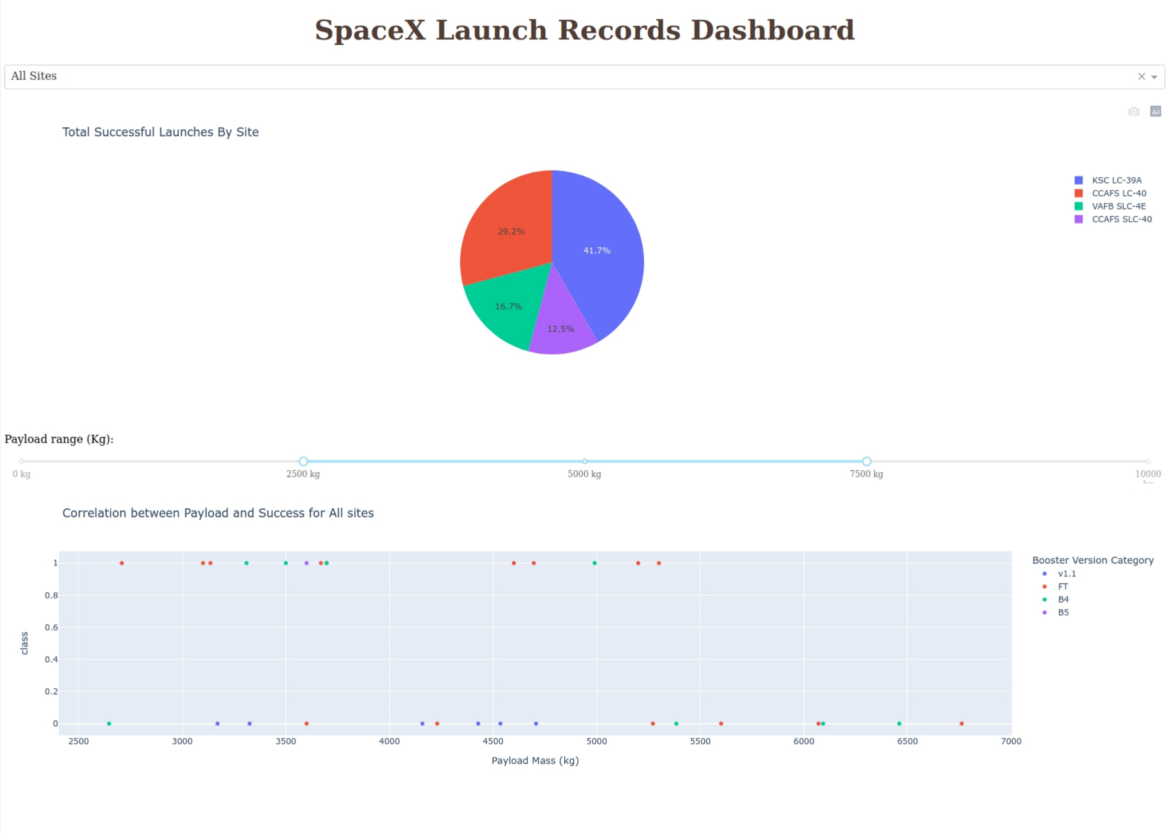
Launch site with highest success rate

Launch site KSC LC-39A has the highest success rate.



Payload vs. Launch Outcome

Payload range between 2000 and 6000 have the highest success rate. Booster version FT has highest success rate.

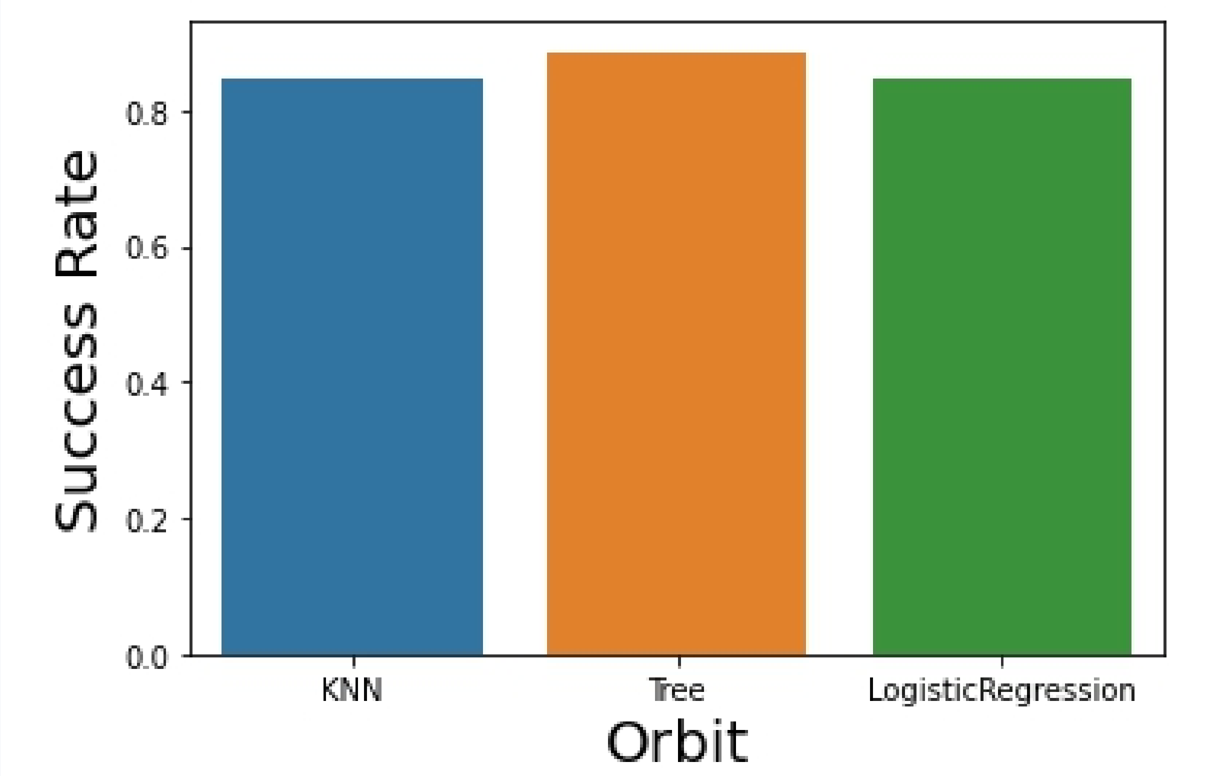


Section 6

Predictive Analysis (Classification)

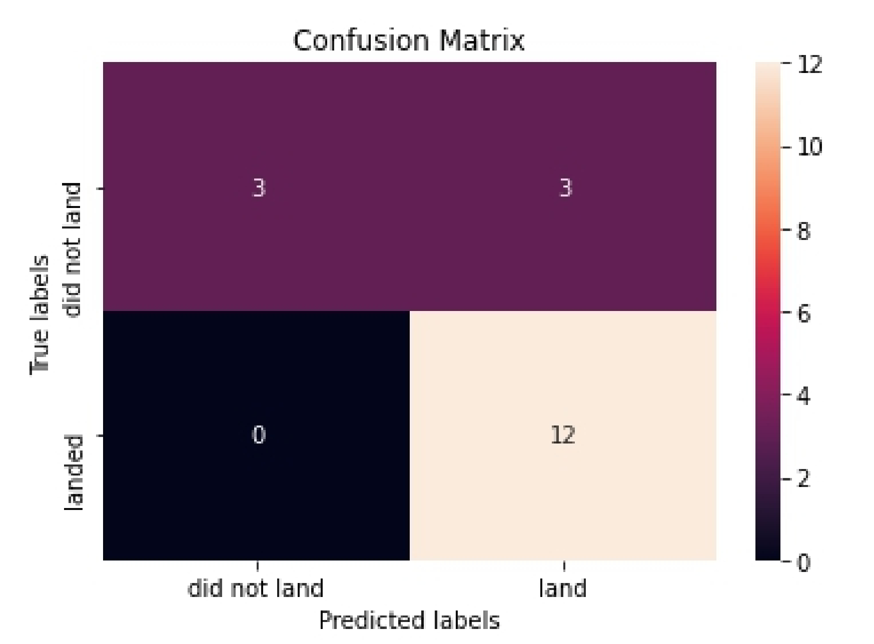
Classification Accuracy

- Decision Tree has the highest accuracy



Confusion Matrix

- The Decision Tree model was the best performing. It has the highest accuracy.



Conclusions

- Launch site KSC LC-39A has the highest success rate
- More launches occur from Florida launch sites than from California.
- The success of launches increases continuously and considerably from 2013 up to 2020.
- All launch sites are located on coasts.
- The Decision Tree model was the best performing.

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.

Link to screenshots used in this report.

Thank you!