IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Terri Funch
20 September 2021

Outline

This presentation consists of the following sections:

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

Executive Summary

- Summary of methodologies
- Summary of all results

Introduction

This project has been undertaken to predict if the Falcon 9 first stage will land successfully. SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage. Therefore, if we can determine if the first stage will land, we can determine the cost of a launch. This information can be used if an alternate company wants to bid against SpaceX for a rocket launch.

SpaceX Falcon 9 first stage Landing Prediction data analysis, as seen throughout this slide deck, involved:

- Collecting the data
- Web scraping Falcon 9 and Falcon Heavy Launches Records from Wikipedia
- Exploratory Data Analysis to find patterns in the data & determine what would be the label for training supervised models
- Launch Sites Locations Analysis with Folium maps
- Producing an interactive dashboard for Success Launch Sites

Section 1

Methodology

Methodology

Executive Summary

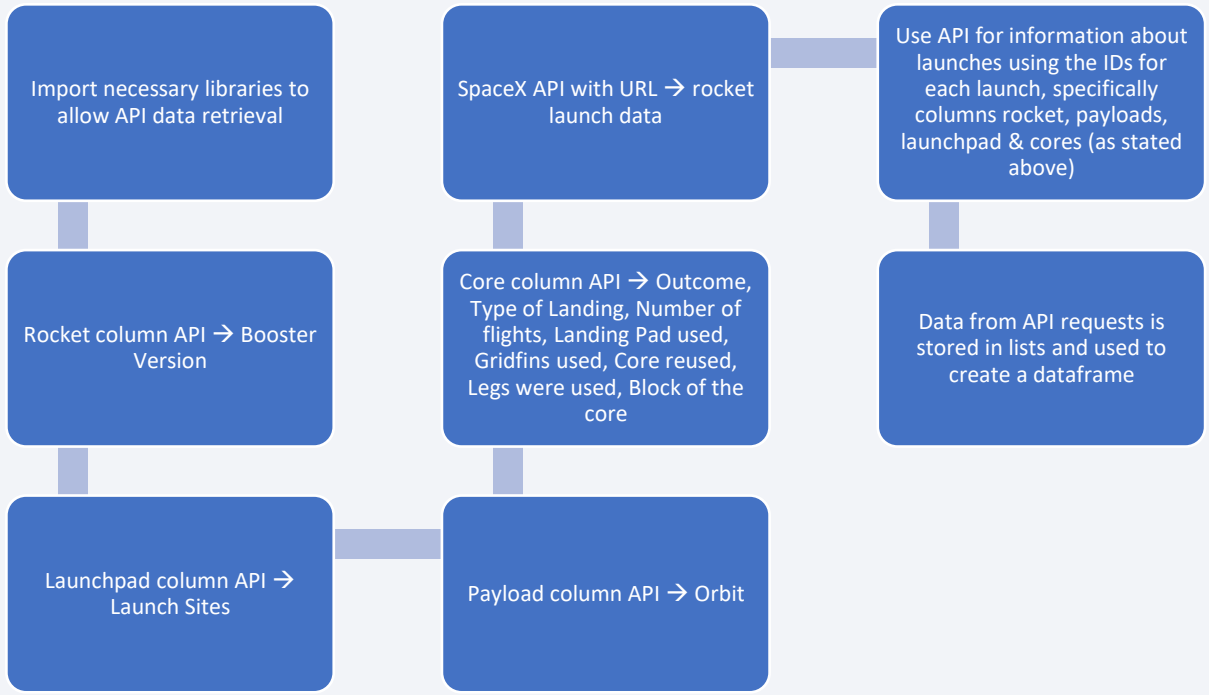
- Data collection methodology
- Perform data wrangling
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

Data Collection

Data set collections and their related process flow charts are found in the following slides.

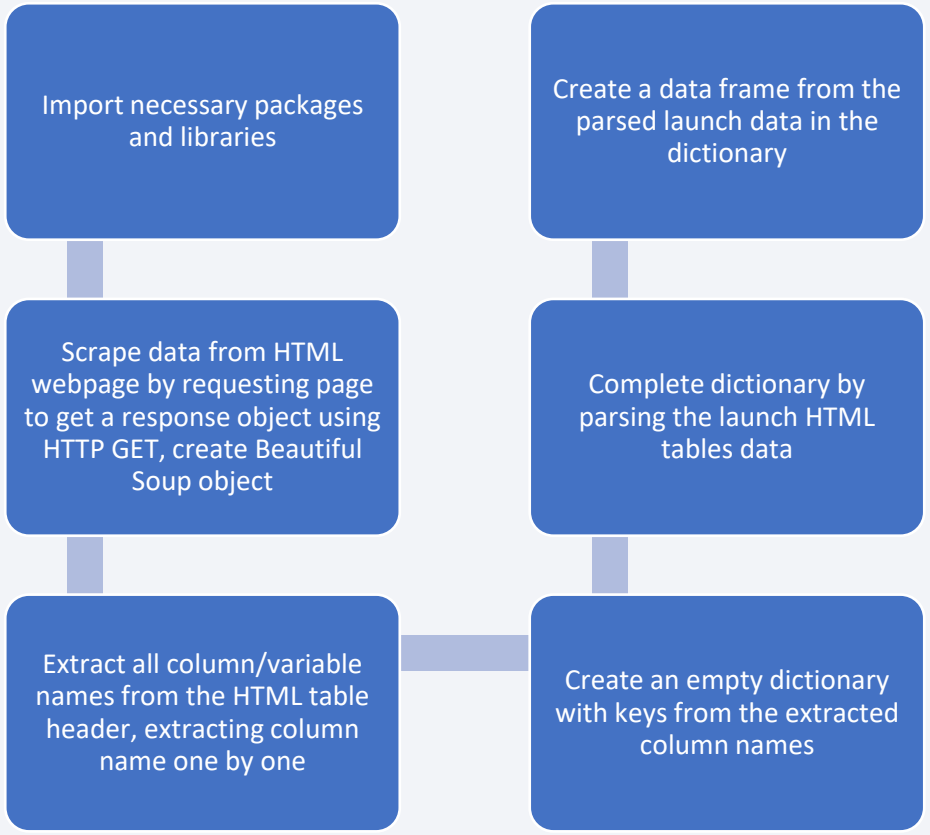
Data Collection – SpaceX API

- GitHub URL of the completed SpaceX API calls notebook: [data_science_capstone/Data Collection API.ipynb at main · terrif2311/data_science_capstone (github.com)]



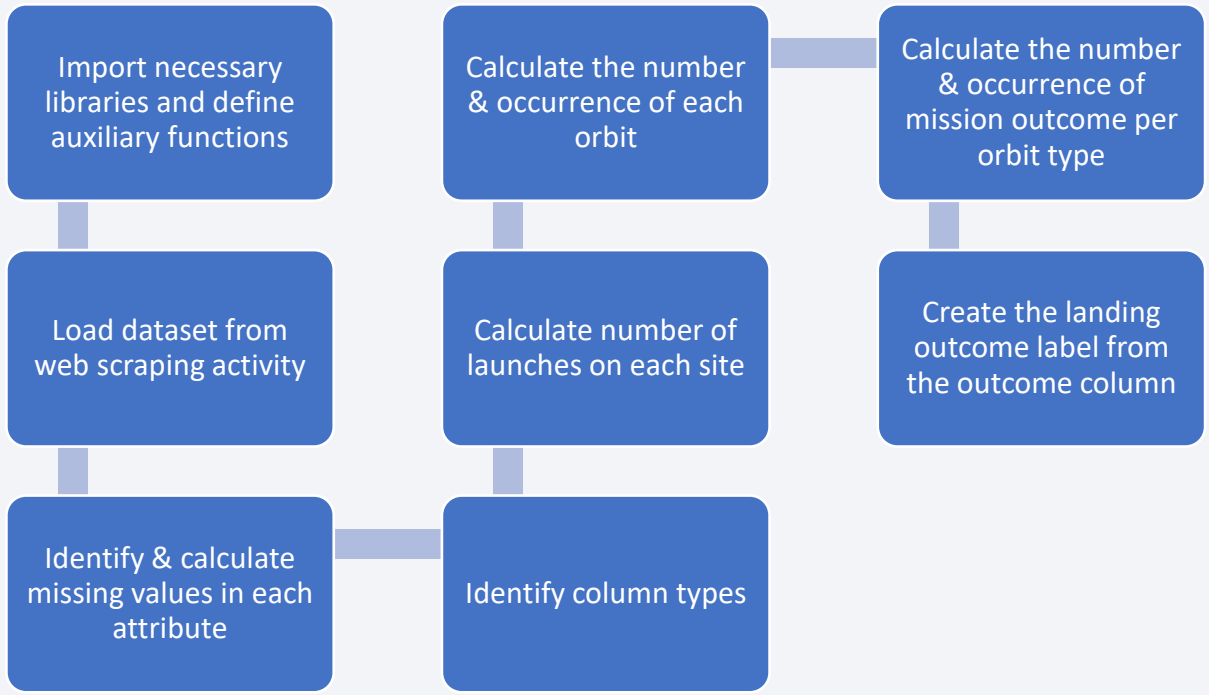
Data Collection - Scraping

- GitHub URL of the completed web scraping notebook
  [data_science_capstone/Data Collection Web Scraping.ipynb at main · terrif2311/data_science_capstone (github.com)]



Data Wrangling

- Exploratory Data Analysis (EDA) is used to find some patterns in the data and determine what would be the label for training supervised models
- Booster landing results:
  - True Ocean means the mission outcome was successfully landed to a specific region of the ocean.
  - False Ocean means the mission outcome was unsuccessfully landed to a specific region of the ocean.
  - True RTLS means the mission outcome was successfully landed to a ground pad.
  - False RTLS means the mission outcome was unsuccessfully landed to a ground pad.
  - True ASDS means the mission outcome was successfully landed on a drone ship.
  - False ASDS means the mission outcome was unsuccessfully landed on a drone ship.
- Outcomes were converted into Training Labels with 1 meant the booster successfully landed 0 meant it was unsuccessful.
- GitHub URL of your completed data wrangling related notebooks [data_science_capstone/EDA.ipynb at main · terrif2311/data_science_capstone (github.com)]



EDA with Data Visualization

- Scatter plots were used to:
  - See how the Flight Number & Payload variables would affect the launch outcome.
  - Visualize the relationship between Flight Number & Launch Site.
  - Visualize the relationship between Payload & Launch Site.
  - Visualize the relationship between Flight Number & Orbit Type.
  - Visualize the relationship between Payload & Orbit Type.
- Bar charts were used to:
  - Visualize the relationship between Success Rate of each Orbit Type.
- Line charts were used to:
  - Visualize the launch success rate yearly trend.

- GitHub URL of completed EDA with data visualization notebook
  data_science_capstone/EDA with Visualisation.ipynb at main · terrif2311/data_science_capstone (github.com)

EDA with SQL

SQL queries performed:

- select DISTINCT LAUNCH_SITE from SPACEXDATASET
- select * from SPACEXDATASET where LAUNCH_SITE like '%CCA%' limit 5
- select customer, sum(payload_mass__kg_) as total_payload from SPACEXDATASET where customer like 'NASA (CRS)' group by customer
- select booster_version, avg(payload_mass__kg_) as average_payload from SPACEXDATASET where booster_version like 'F9 v1.1' group by booster_version
- select min(DATE) as Date_Successful_Landing from SPACEXDATASET where landing__outcome like 'Success (ground pad)'
- select booster_version from SPACEXDATASET where landing__outcome like 'Success (drone ship)' and payload_mass__kg_ >4000 and payload_mass__kg_ < 6000
- select mission_outcome, count(mission_outcome) as total_mission_outcome from SPACEXDATASET group by mission_outcome
- select booster_version, payload_mass__kg_ from SPACEXDATASET where payload_mass__kg_ = (select max(payload_mass__kg_)as max_payload_mass from SPACEXDATASET)
- select landing__outcome, booster_version, launch_site from SPACEXDATASET where date like '%2015%' and landing__outcome like 'Failure (drone ship)'
- select landing__outcome, count(landing__outcome) as total_landing_outcome from SPACEXDATASET where date between '2010-06-04' and '2017-03-20' group by landing__outcome order by total_landing_outcome desc

GitHub URL for completed EDA with SQL notebook [data_science_capstone/EDA with SQL.ipynb at main · terrif2311/data_science_capstone (github.com)]

Build an Interactive Map with Folium

Objects used in the interactive maps include:

- Folium.circle to add a highlighted circle area with a text label on a specific coordinate for the launch sites.
- MarkerCluster to simplify a map containing many markers having the same coordinate.
- MousePosition to easily get coordinates for a mouse over point on the map.
- Folium.marker to show the distance between 2 coordinates.
- Folium.PolyLine to draw a line between the 2 coordinates of the Folium.marker.

GitHub URL of completed interactive map with Folium map data_science_capstone/Interactive Visual Analytics with Folium.ipynb at main · terrif2311/data_science_capstone (github.com)

Build a Dashboard with Plotly Dash

- Interactions found on the dashboard include:
  - Pie chart for summarised All Sites or for individual Launch Sites, showing success landing outcomes rates.
  - Scatter chart for summarised All Sites or for individual Launch Sites, showing the correlation between payload mass & success landing.
  - Range slider to filter on payload mass for the scatter chart, to allow drilling for further data analysis.

GitHub URL of completed Plotly Dash lab data_science_capstone/spacex_dash_app.py at main · terrif2311/data_science_capstone (github.com)

Predictive Analysis (Classification)

Exploratory Data Analysis was undertaken and to determine Training Labels by:

- Creating a column for the class
- Standardizing the data
- Splitting the data into training data and test data

Find the method that performs the best through using the test data and hyperparameters for:

- Support Vector Machine (SVM)
- Classification Trees; and
- Logistic Regression

GitHub URL of completed predictive analysis lab data_science_capstone/Machine Learning Prediction.ipynb at main · terrif2311/data_science_capstone (github.com)

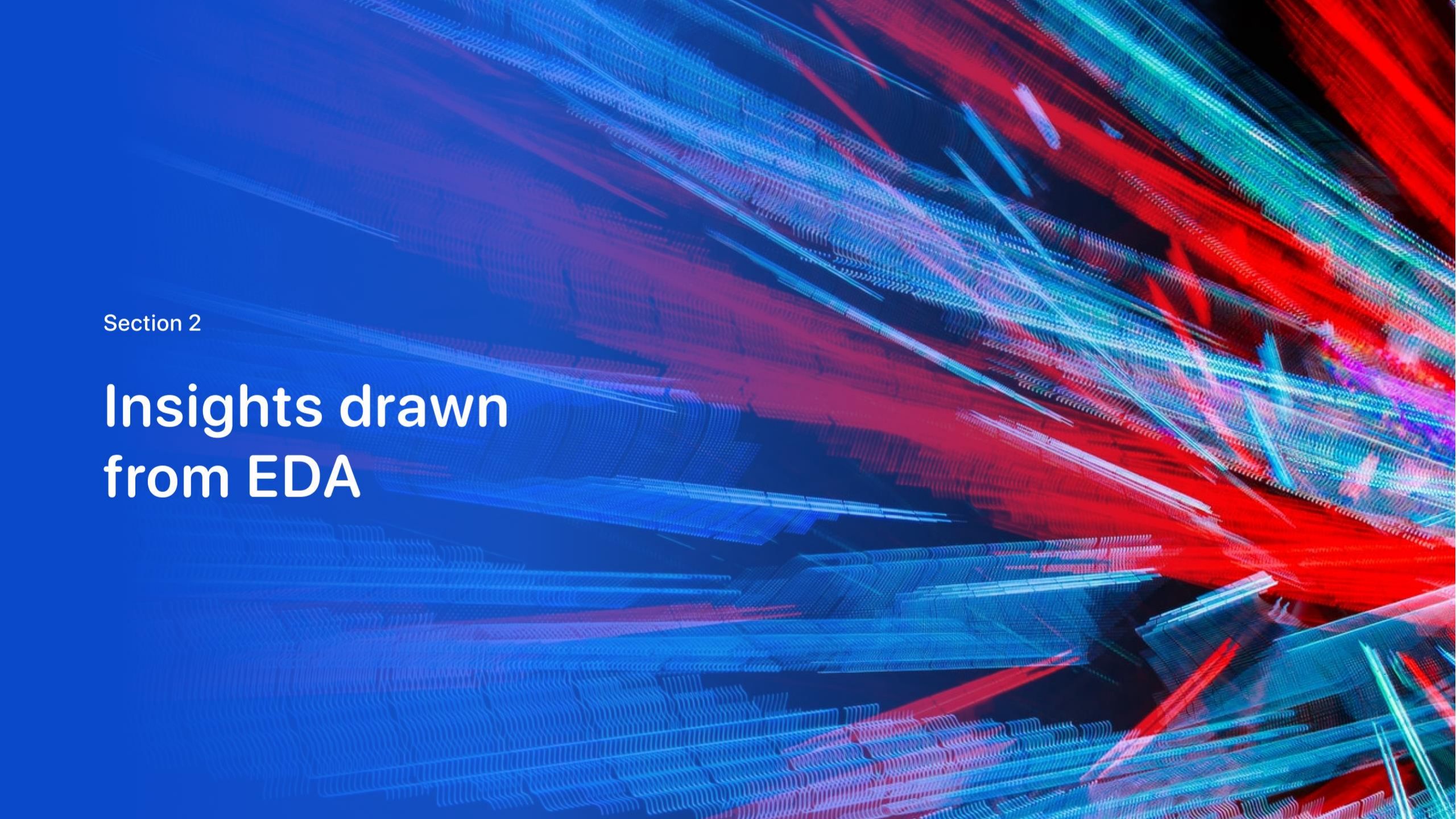
Results

Results of the project analysis are found in:

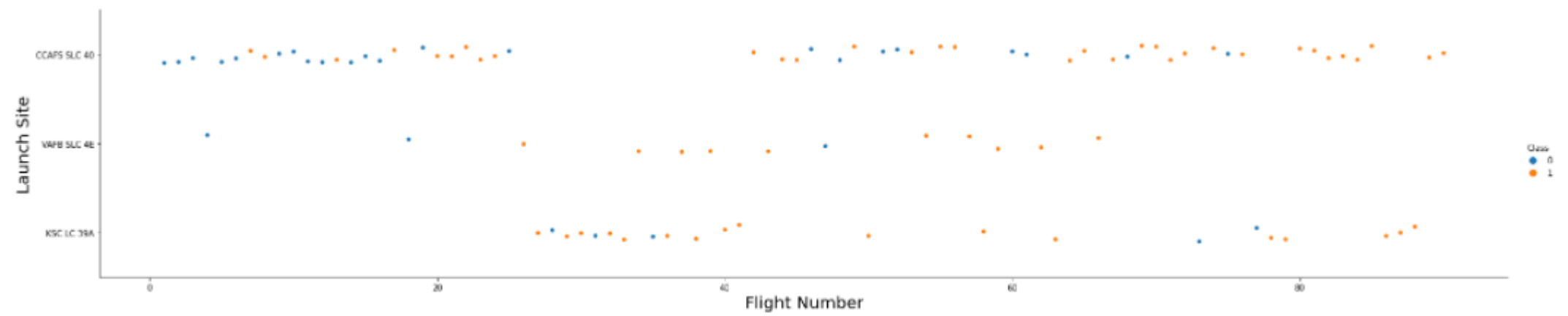
- Exploratory data analysis results
- Interactive analytics demo in screenshots provided
- Predictive analysis results

Section 2

Insights drawn from EDA

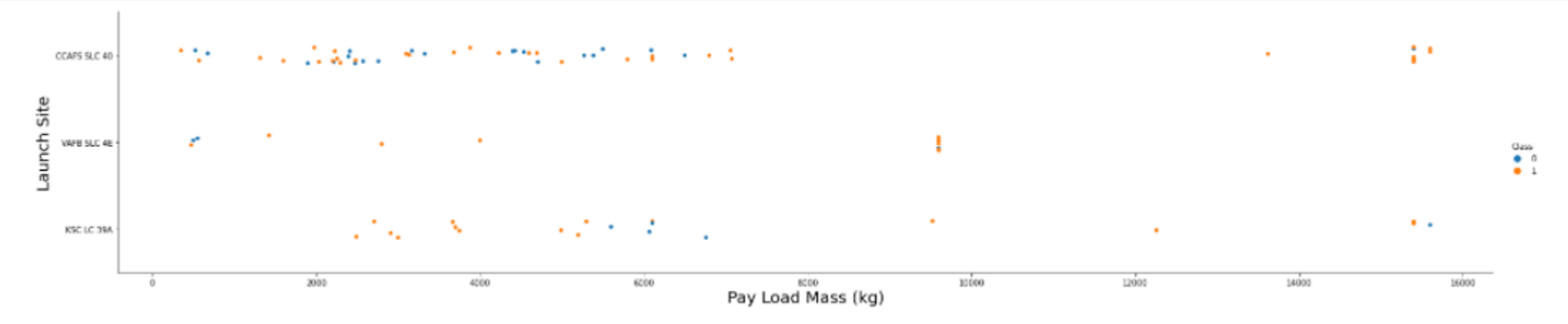


Flight Number vs. Launch Site



There seems to be more success landing outcomes as the flight numbers increase for launch sites.

Payload vs. Launch Site



It seems the lower the payload mass, the more success landing outcomes.

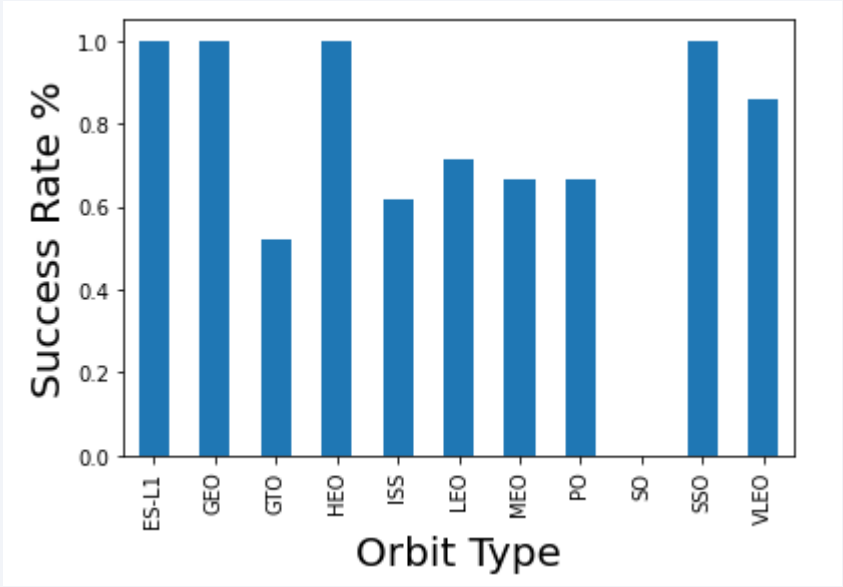
Success Rate vs. Orbit Type

The Orbit Types with the highest success rate are:

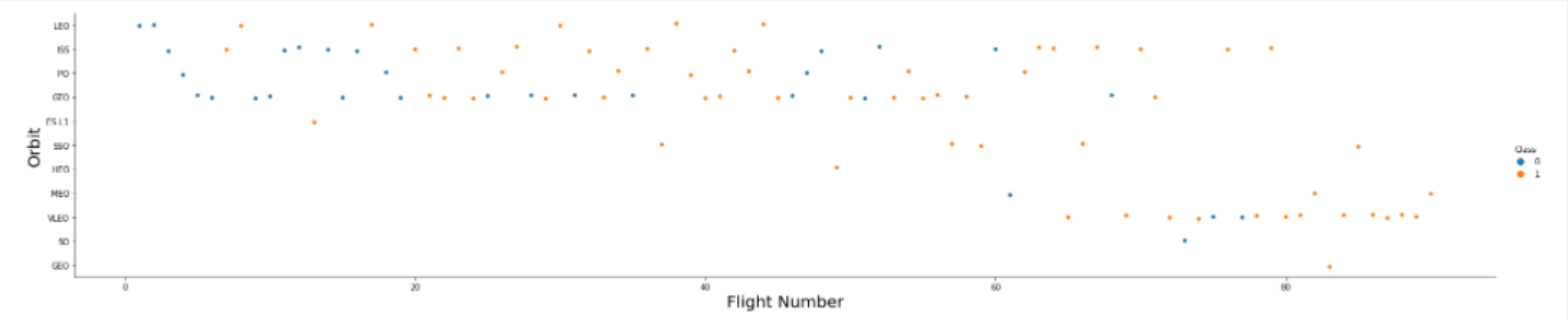
- ES-L1
- GEO
- HEO
- SSO

The Orbit Type with the lowest success rate is:

- SO (no success landing outcomes)

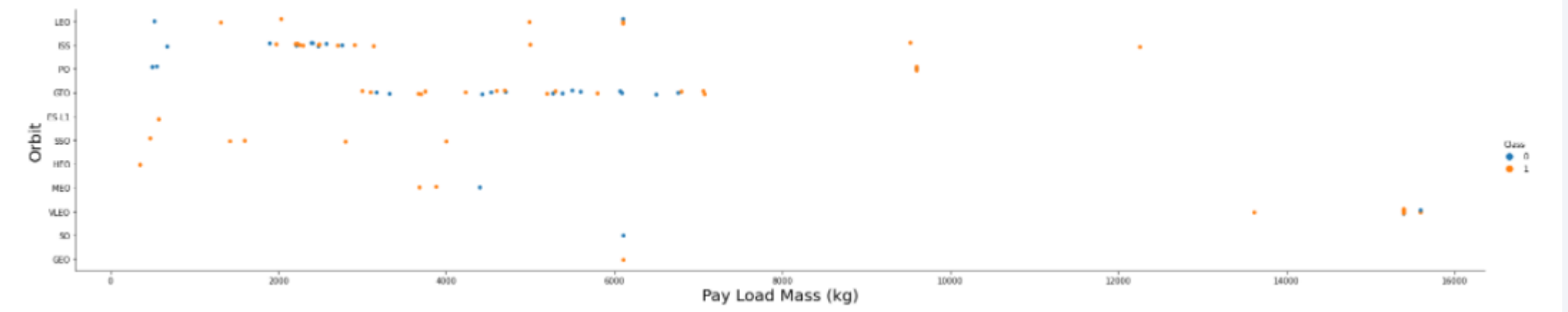


Flight Number vs. Orbit Type



This chart is reflective of the bar chart in the previous slide.

Payload vs. Orbit Type



It seems the lower the payload mass, the more success landing outcomes in multiple Orbits.

Launch Success Yearly Trend

Could not get the line chart of yearly average success rate to work, even after many attempts of help from course instructors

All Launch Site Names

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

```
%sql select DISTINCT LAUNCH_SITE from SPACEXDATASET
```

```
 * ibm_db_sa://rky97748:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32731/BLUDB
Done.
```

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

Display of names of launch sites in the SpaceX dataset

Launch Site Names Begin with 'CCA'

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

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

```
 * ibm_db_sa://rky97748:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32731/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

Display of 5 records in the SpaceX dataset where the launch site begins with CCA

Total Payload Mass

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

```
%sql select customer, sum(payload_mass__kg_) as total_payload from SPACEXDATASET where customer like 'NASA (CRS)' group by custo
mer
```

```
 * ibm_db_sa://rky97748:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32731/BLUDB
Done.
```

customer	total_payload
NASA (CRS)	45596

Display the total (grouped) payload mass carried by boosters in the SpaceX dataset where the customer name is NASA (CRS)

Average Payload Mass by F9 v1.1

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

```
%sql select booster_version, avg(payload_mass__kg_) as average_payload from SPACEXDATASET where booster_version like 'F9 v1.1' g
roup by booster_version
```

```
 * ibm_db_sa://rky97748:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32731/BLUDB
Done.
```

booster_version	average_payload
F9 v1.1	2928

Display the average payload mass carried in the SpaceX dataset where the booster version is F9 v1.1

First Successful Ground Landing Date

List the date when the first successful landing outcome in ground pad was acheived.

Hint:Use min function

```
%sql select min(DATE) as Date_Successful_Landing from SPACEXDATASET where landing__outcome like 'Success (ground pad)'
```

```
 * ibm_db_sa://rky97748:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32731/BLUDB
Done.
```

date_successful_landing
2015-12-22

Display the date of the first successful landing outcome in ground pad in the SpaceX dataset

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 SPACEXDATASET where landing__outcome like 'Success (drone ship)' and payload_mass__kg_ >4000 an
d payload_mass__kg_ < 6000
```

```
 * ibm_db_sa://rky97748:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32731/BLUDB
Done.
```

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

Display the booster versions that have had successful landing outcome in drone ship and the payload mass was greater than 4000 but less than 6000 in the SpaceX dataset

Total Number of Successful and Failure Mission Outcomes

List the total number of successful and failure mission outcomes

```
%sql select mission_outcome, count(mission_outcome) as total_mission_outcome from SPACEXDATASET group by mission_outcome
```

```
 * ibm_db_sa://rky97748:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32731/BLUDB
Done.
```

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

Display the total (grouped) number of mission outcomes by successful and failure in the SpaceX dataset

Boosters Carried Maximum Payload

List the names of the booster_versions which have carried the maximum payload mass. Use a subquery

```
%sql select booster_version, payload_mass__kg_ from SPACEXDATASET where payload_mass__kg_ = (select max(payload_mass__kg_)as max
_payload_mass from SPACEXDATASET)
```

```
 * ibm_db_sa://rky97748:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32731/BLUDB
Done.
```

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

Display the booster versions that carried the maximum payload mass in the SpaceX dataset. A subset was used to find the maximum payload mass value.

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 SPACEXDATASET where date like '%2015%' and landing__outcome like
'Failure (drone ship)'
```

```
 * ibm_db_sa://rky97748:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32731/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

Display the booster versions and launch site where the landing outcome was failure on drone ship during 2015 in the SpaceX dataset.

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(landing__outcome) as total_landing_outcome from SPACEXDATASET where date between '2010-06-0
4' and '2017-03-20' group by landing__outcome order by total_landing_outcome desc
```

```
 * ibm_db_sa://rky97748:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32731/BLUDB
Done.
```

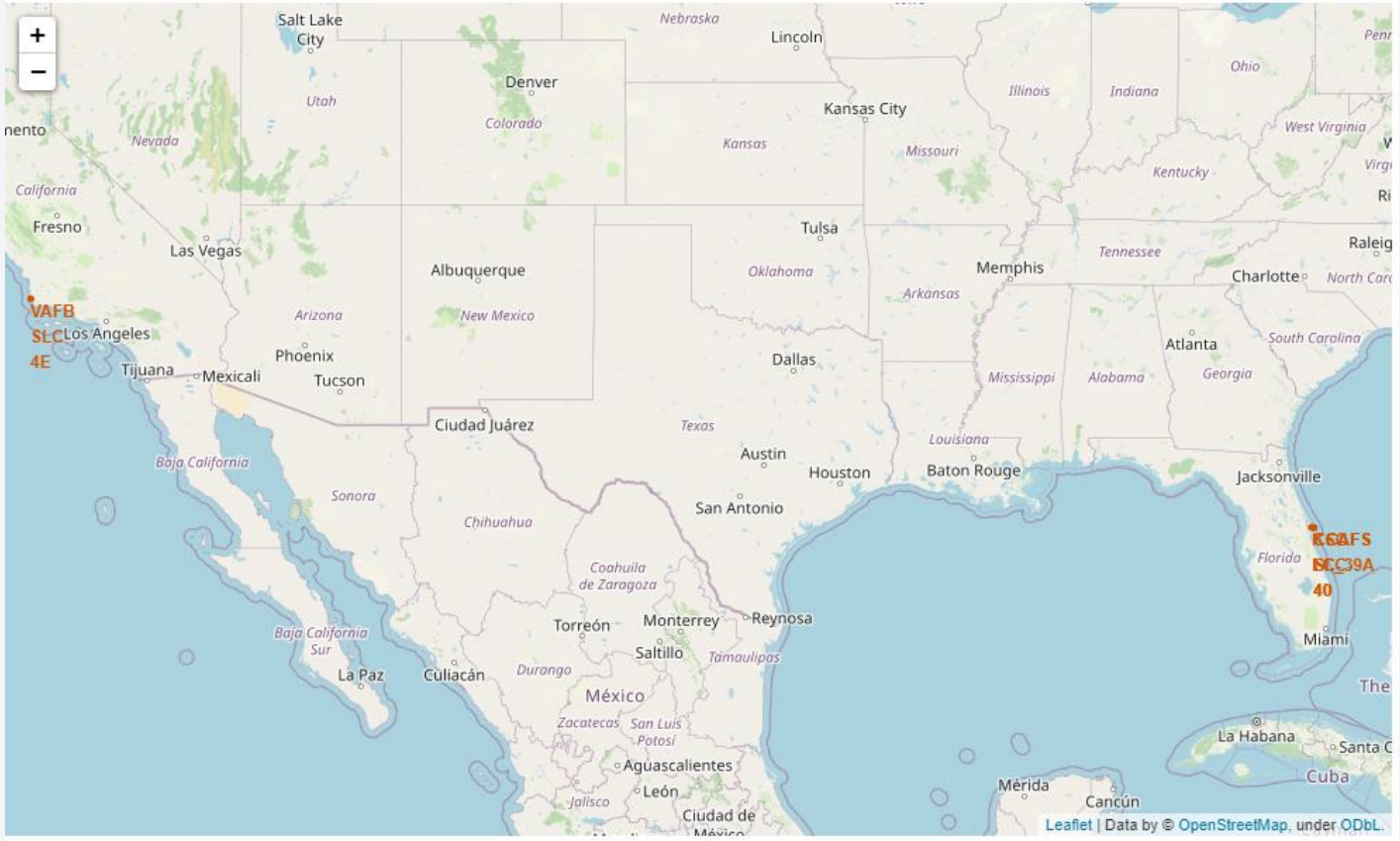
landing__outcome	total_landing_outcome
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

Display the (grouped) landing outcomes, between 4 June 2010 and 20 March 2017, ranked by total landing outcome in descending order in the SpaceX dataset.

Section 4

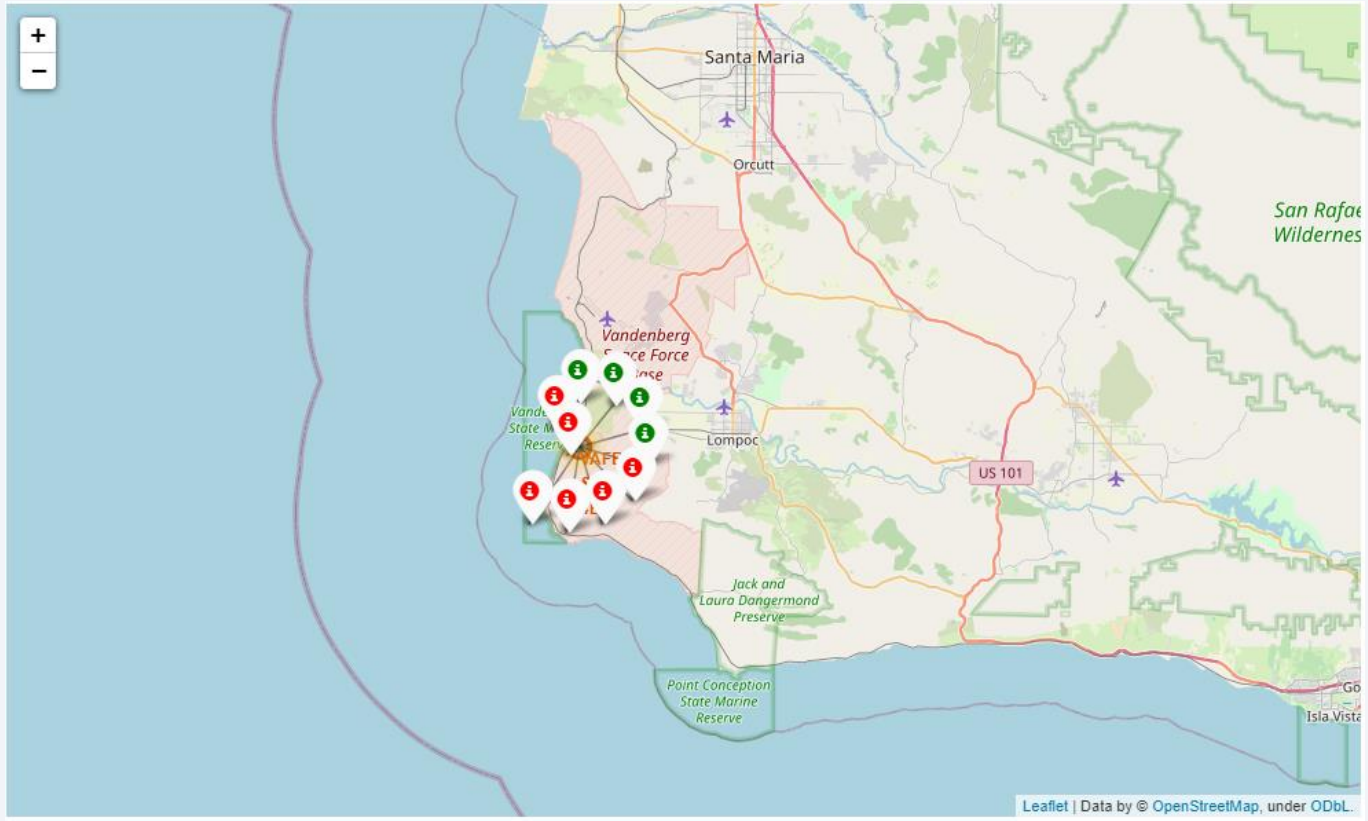
Launch Sites Proximities Analysis

Folium Map: Launch Sites



Map shows the location of the SpaceX launch sites, showing their proximity to the ocean

Folium Map: Launch Site with landing outcomes



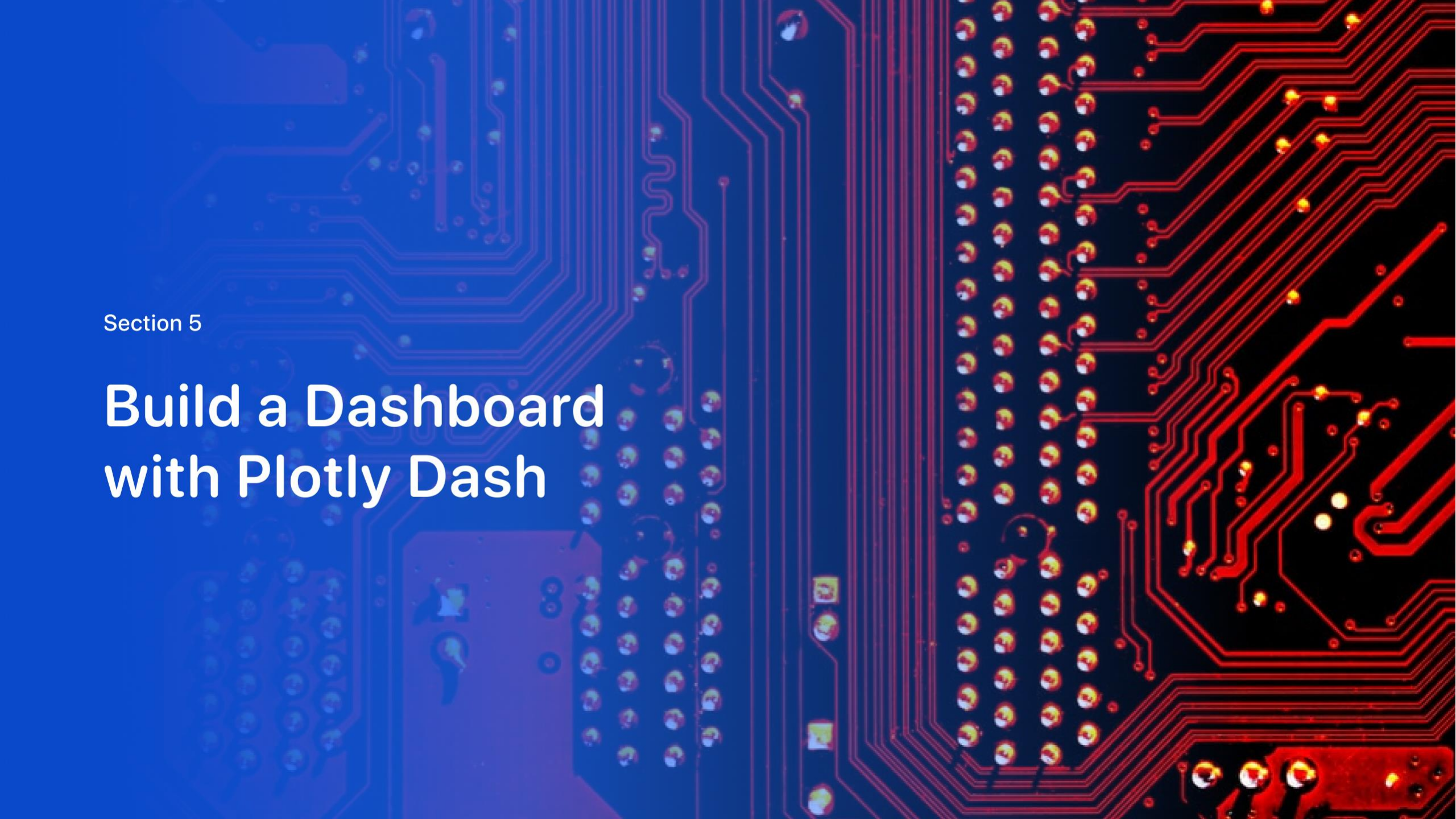
Map is zoomed in on a location of a SpaceX launch site, showing green for successful landings, and red for unsuccessful landing outcomes.

Folium Map: Launch site proximity to coastline

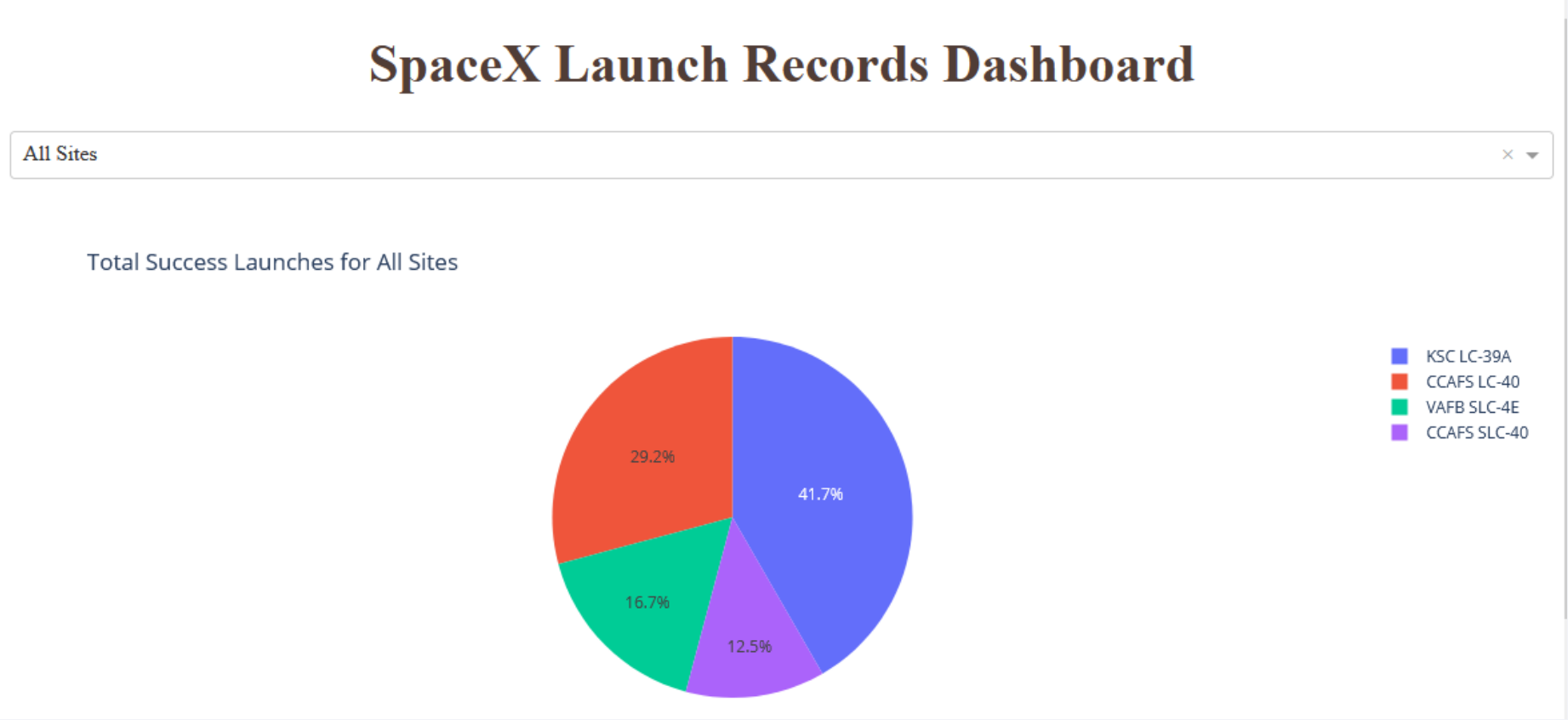
Map is zoomed in on a location of a SpaceX launch site, showing the proximity and distance to the coastline.

Section 5

Build a Dashboard with Plotly Dash

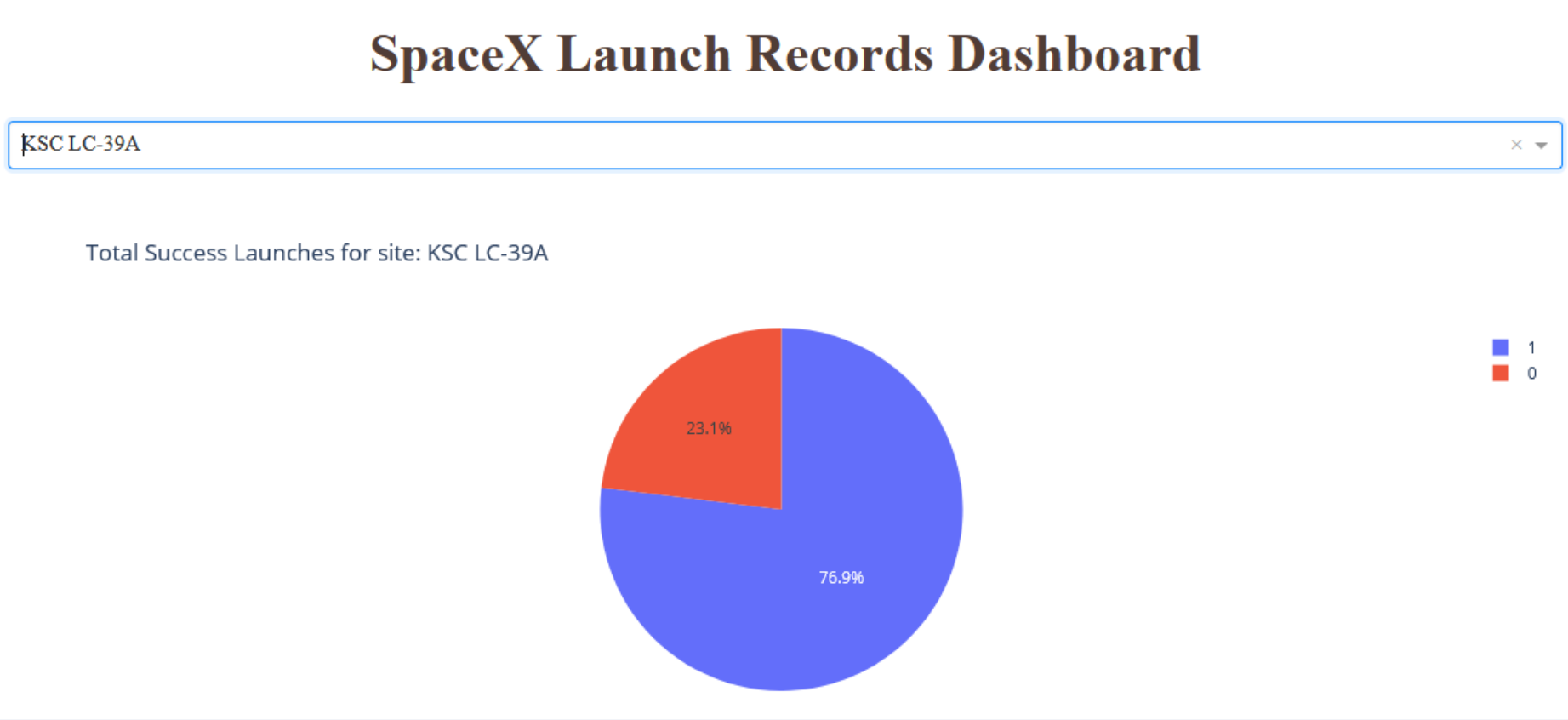


Dashboard: Total Success Launches for All Sites



This pie chart displays the total successful launches for all sites, showing the split by percentages for each launch site.

Dashboard: Launch Site with Highest Launch Success Ratio



This pie chart displays the launch site KSC LC-39A, the site with the highest launch success ratio (76.9% success, 23.1% failure).

Dashboard: All Launch Sites Payload and Success Correlation

This scatter chart displays the correlation between payload and success for all sites, based on a filtered payload range of 2500kgs to 6000kgs, and grouped into booster version categories. In this example, there is no booster version v1.1 that was successful, only failures.

Section 6

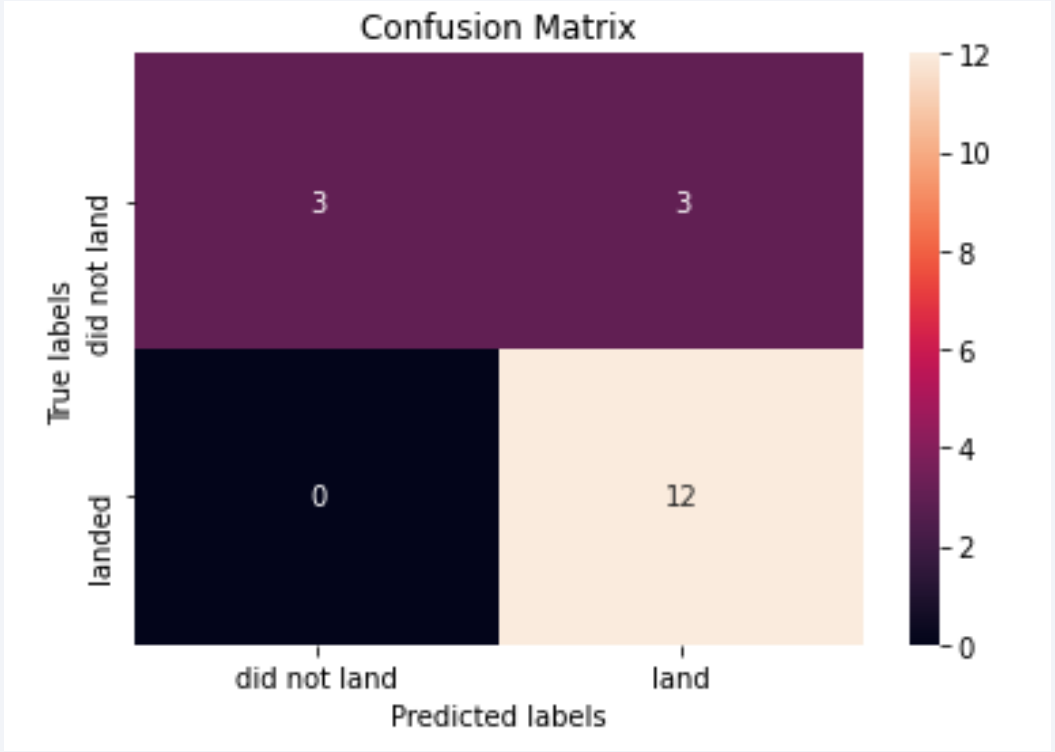
Predictive Analysis (Classification)

Classification Accuracy

Highest classification accuracy are as follows:

Method	Best hyperparameter	Train data accuracy %	Test data accuracy %
Logistic Regression	lbfgs	84.64%	83.33%
SVM	sigmoid	84.82%	83.33%
Classification tree	criterion:gini max depth: 4 max features: sqrt min samples leaf: 1 min samples split: 2 splitter: best	88.92%	83.33%

Confusion Matrix



Above is the confusion matrix result from all of the models produced, ie all returned the same results.

Conclusions

- KSC LC-39A is the most successful launch site that SpaceX has.
- Landing outcome of “no attempt” will always rate the highest rank as SpaceX will incorporate no attempts of some landings.
- There is more success in landing outcomes when the payload mass is not at the maximum kg range.

Appendix

- Falcon 9 historical launch records from Wikipedia
  https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches
- Snapshot of the List of Falcon 9 and Falcon Heavy launches Wikipage, 9 June 2021
  https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922

Thank you!