Winning Space Race with Data Science

Zane van Campfort
30/5/2024

Outline

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

Executive Summary

- Data Preparation and Wrangling to organise and filter relevant data such that the feature and target variable relationships can be investigated.
- Exploratory Data Analysis (EDA) using Pandas, Matplotlib and SQL, to gauge the importance of each feature in predicting the target outcome.
- Data Visualisation, with Folium and Plotly Dash, to investigate trends and relative effects of changing each feature variable.
- Model Evaluation to determine predictive power of feature variables.

- Future SpaceX launches have a high probability of successfully landing the first stage, particularly for heavy payloads, those sent into VLEO orbit, and those launched at KSC LC 39A.

Introduction

- The company executives at SpaceCorp wish to make a **data-driven** decision on whether they should bet against the next SpaceX launch by analysing the cost effectiveness of SpaceX payload launches.

- The most significant factor determining cost of a launch is whether the first stage of the launch spacecraft will **remain intact** after launch, as first stages may be reused at no additional cost for multiple launches.

Section 1

Methodology

Methodology

Executive Summary

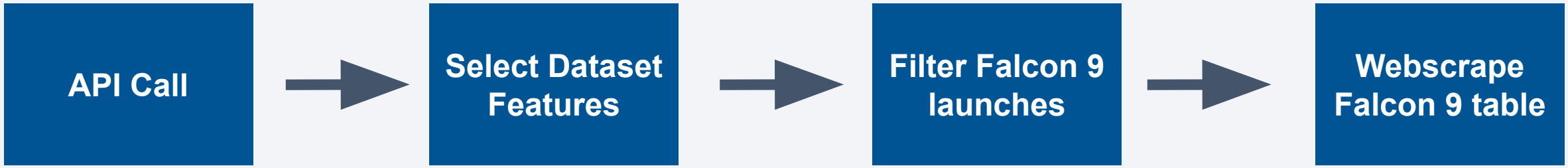
- Data collection methodology:
  - Describe how data was collected
- Perform data wrangling
  - Describe how data was processed
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection

Datasets Used:

- SpaceX Launch Historical Data - https://api.spacexdata.com/v4/launches/past
- Booster Falcon 9 Flights Data table -

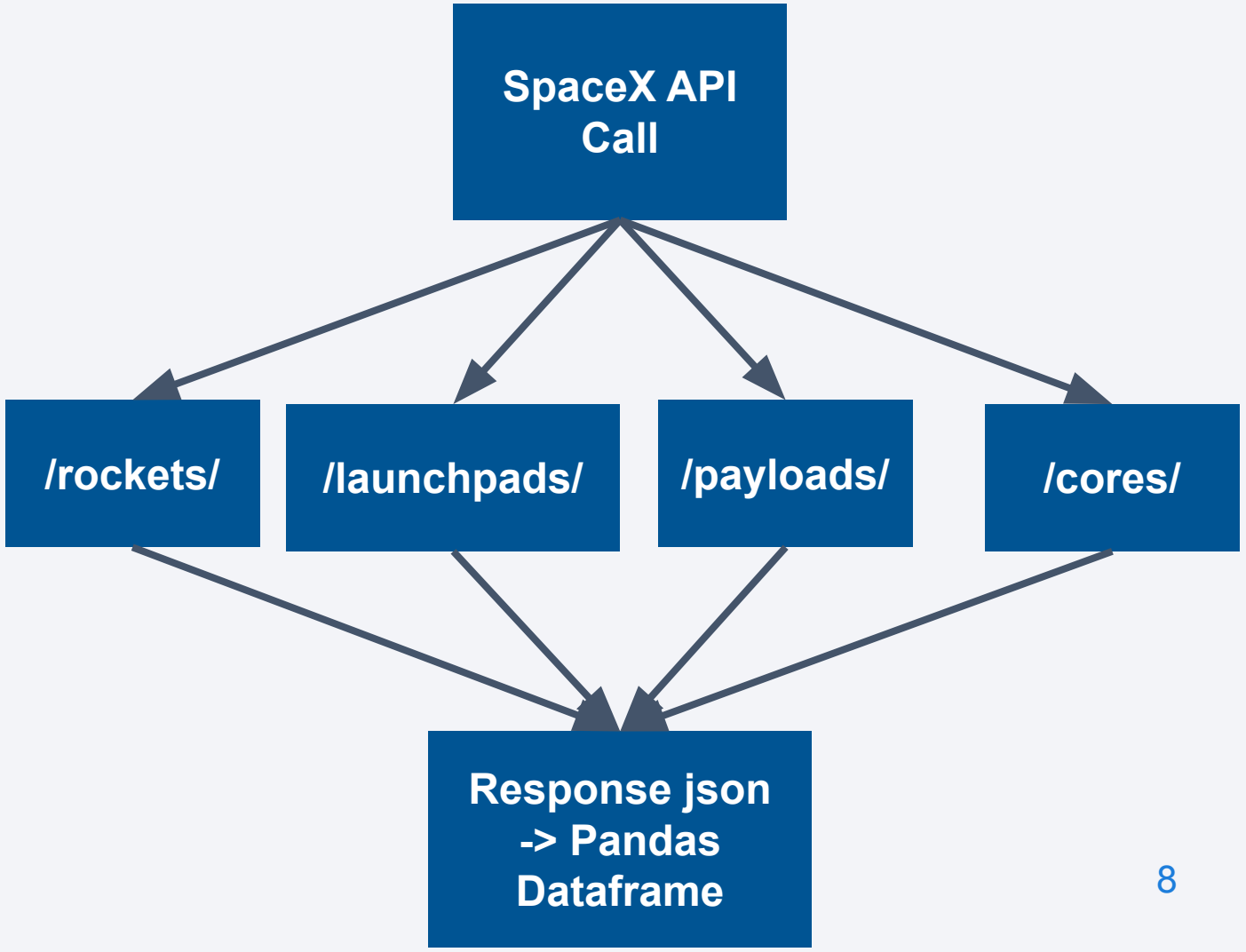
Collection Process



Data Collection – SpaceX API

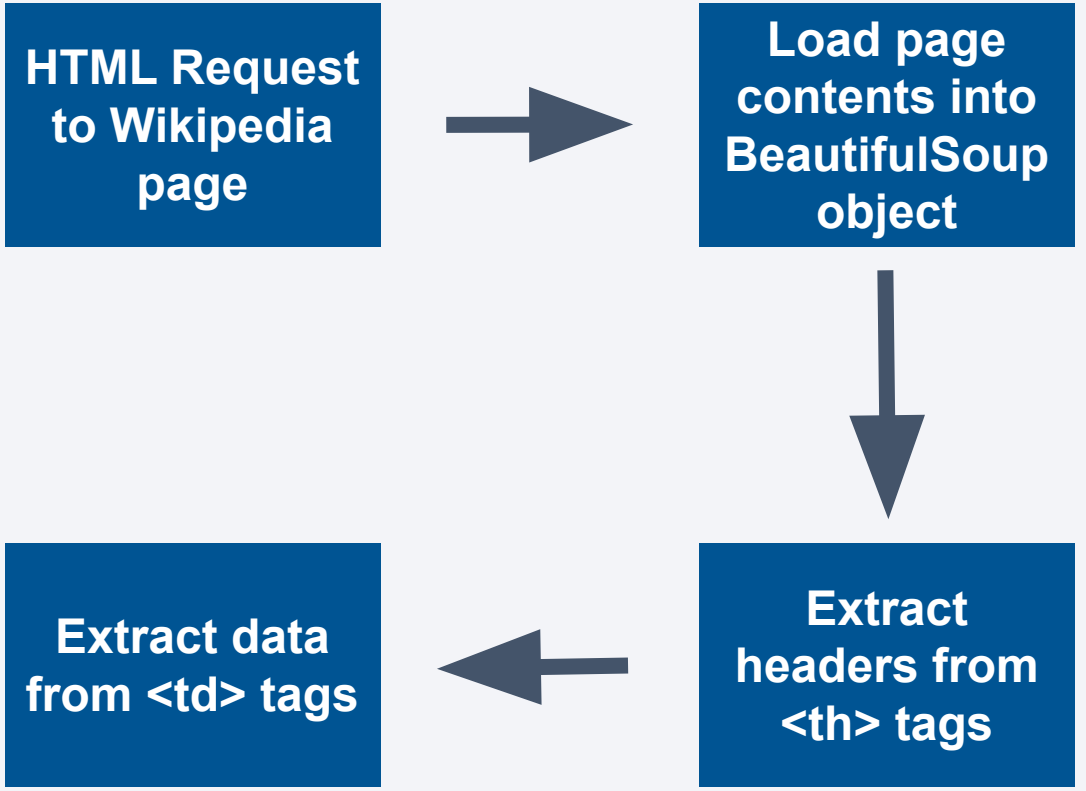
- Get JSON response to API calls and store as Pandas Dataframe

- https://github.com/zanevc5555/IBM_DataScience_Capstone/blob/main/jupyter-labs-spacex-data-collection-api.ipynb



Data Collection - Scraping

- Extract a Falcon 9 launch records HTML table from Wikipedia
- Parse the table and convert it into a Pandas data frame



- https://github.com/zanevc5555/IBM_DataScience_Capstone/blob/main/jupyter-labs-webscraping.ipynb

Data Wrangling

Problem:

- Dataset is not properly formatted for modelling
- Target variable too specific; only need True and False outcomes

Solution:

- Transform categorical target data to binary numerical data

https://github.com/zanevc5555/IBM_DataScience_Capstone/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

Chart Types:

- Scatter Plots (Flight Number vs. Feature) - Observe overall trend of feature variables.
- Scatter Plots (Payload Mass vs. Feature) - Characterise SpaceX launches by payload mass.
- Bar Plot (Orbit vs. Class) - Investigate relationship between launch orbit and success rate.

- https://github.com/zanevc5555/IBM_DataScience_Capstone/blob/main/eda dataviz.ipynb

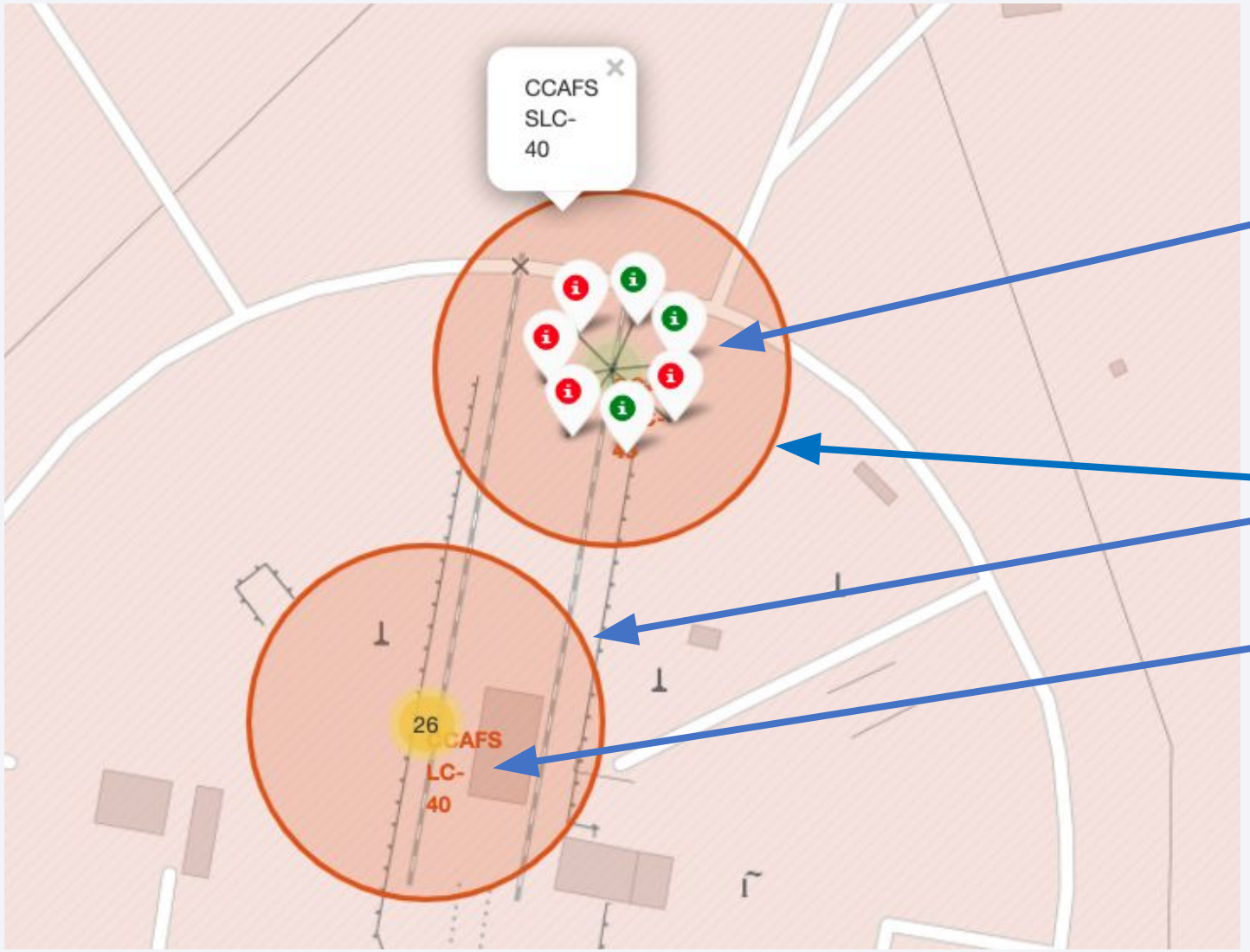
EDA with SQL

SQL Queries:

- Display the names of the unique launch sites in the space mission
- Display the total payload mass carried by boosters launched by NASA (CRS)
- List the total number of successful and failure mission outcomes
- List the date when the first successful landing outcome in ground pad was achieved.
- 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.

- https://github.com/zanevc5555/IBM_DataScience_Capstone/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium



Folium Map Features:

- MarkerCluster object groups icons by coordinates, green for launch success and red for launch failure.
- Circle object shows coloured zone and pop up label at selected coordinate.
- Individual markers provide text label for each site of interest.

Build a Dashboard with Plotly Dash

Dashboard Features:

- Dropdown widget to choose which launch site to analyse.
- First plot - Pie chart to visualise distribution of launch successes by launch site, or visualise proportion of successes/failure for a selected site.
- Range slider to filter second plot results by a given range of payload masses.
- Second plot - Scatter plot of launch success/failure ordered by payload mass, grouped by booster category.

https://github.com/zanevc5555/IBM_DataScience_Capstone/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)

- Standardise feature variables
- Split data into training and testing sets
- Optimise model parameters with GridSearchCV
- Use metrics such as R-squared to evaluate model accuracy.

Models:

- **Logistic Regression**
- **Support Vector Machine (SVM)**
- **Decision Tree**
- **K-Nearest NeighboZZrs (KNN)**

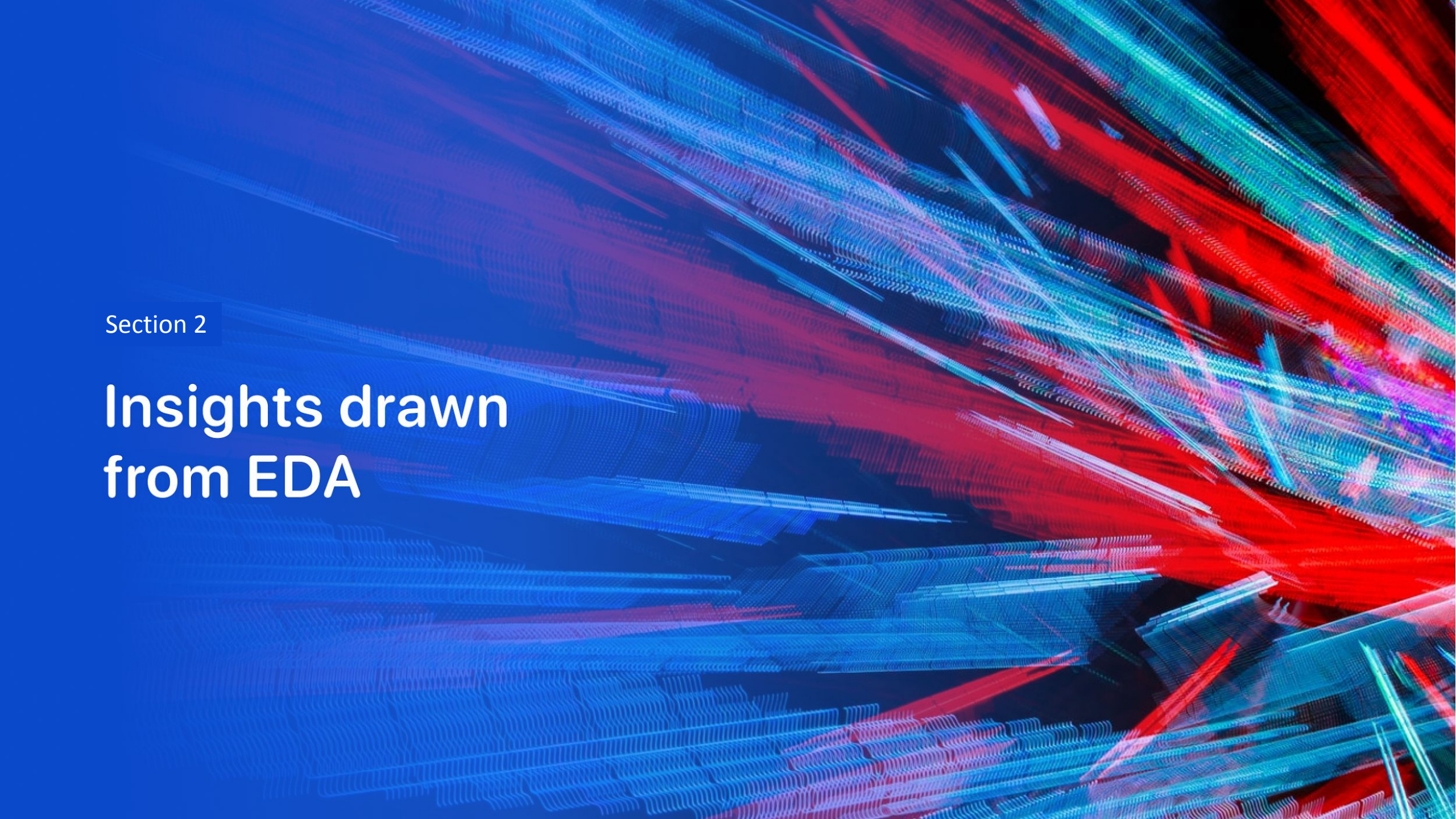
https://github.com/zanevc5555/IBM_DataScience_Capstone/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Results

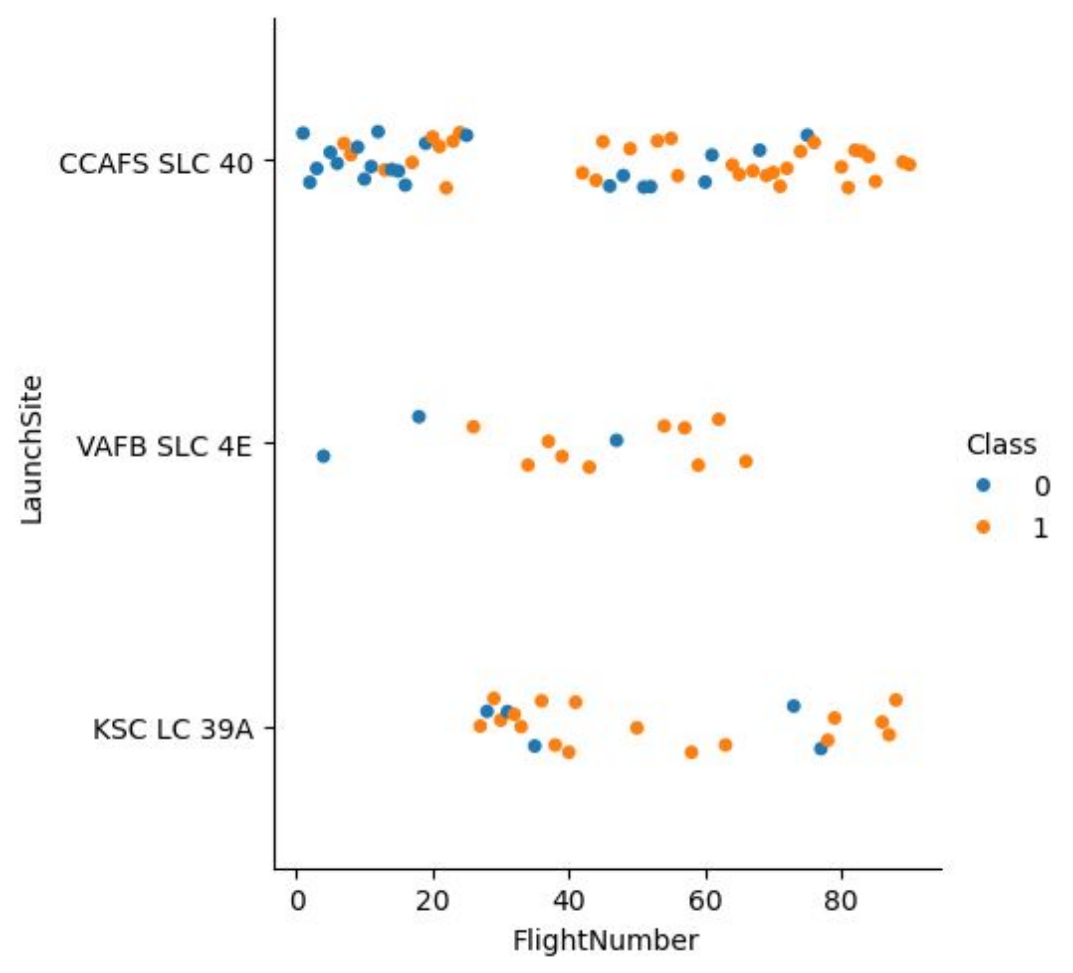
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA



Flight Number vs. Launch Site



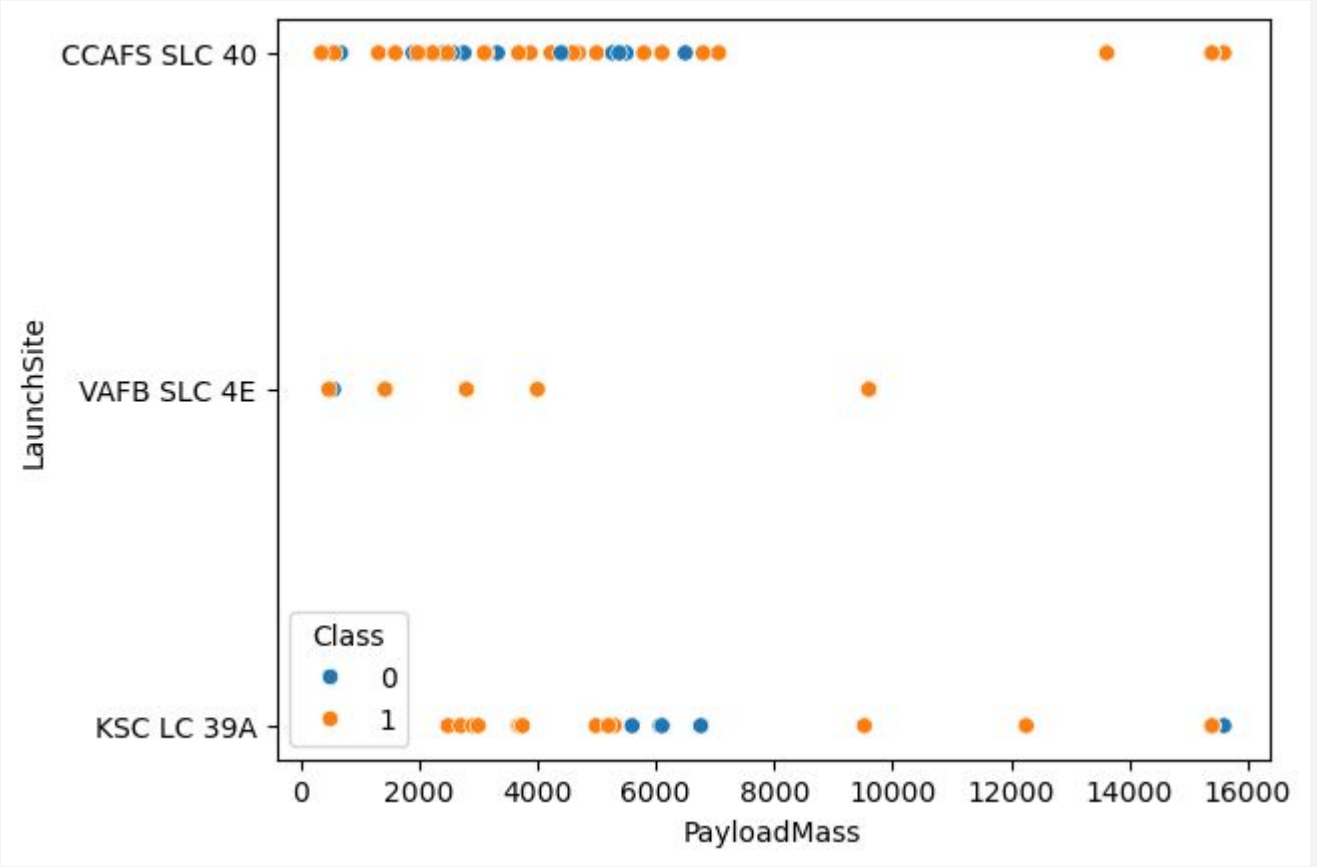
This scatter plot orders all SpaceX launches by launch site and denotes each first stage landing outcome as dot colour. Blue dots are failures, while red dots are successes.

The launch site category **CCAFS SLC 40** features both the highest number of launches and the highest proportion of landing failures.

Payload vs. Launch Site

This plot shows the variation of payload masses corresponding to each launch site category, alongside landing outcome for each launch.

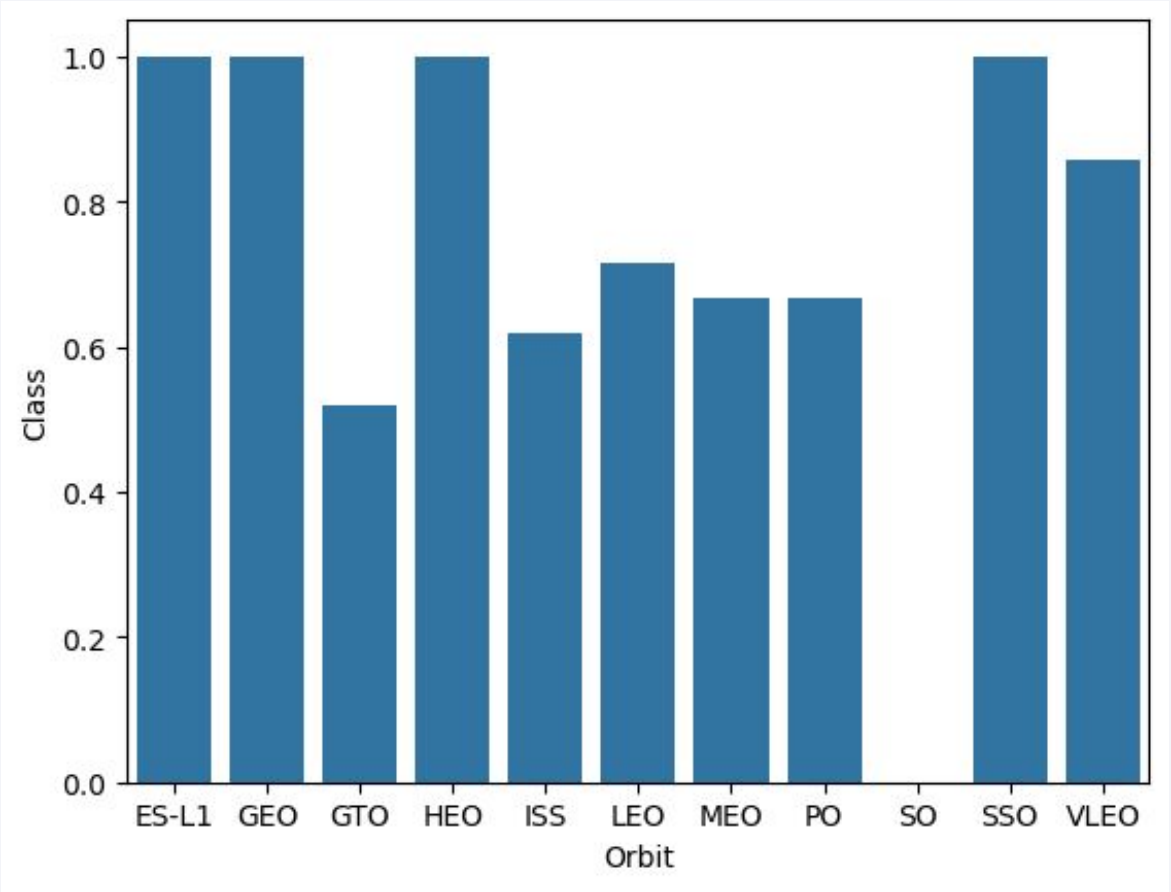
The majority of payloads at sites **CCFAS SLC 40** & **KSC LC 39A** are evenly distributed across the 500-7000 kg range, while payloads launched at **VAFB SLC 4E** cluster around the 500 kg mark with some right skew.



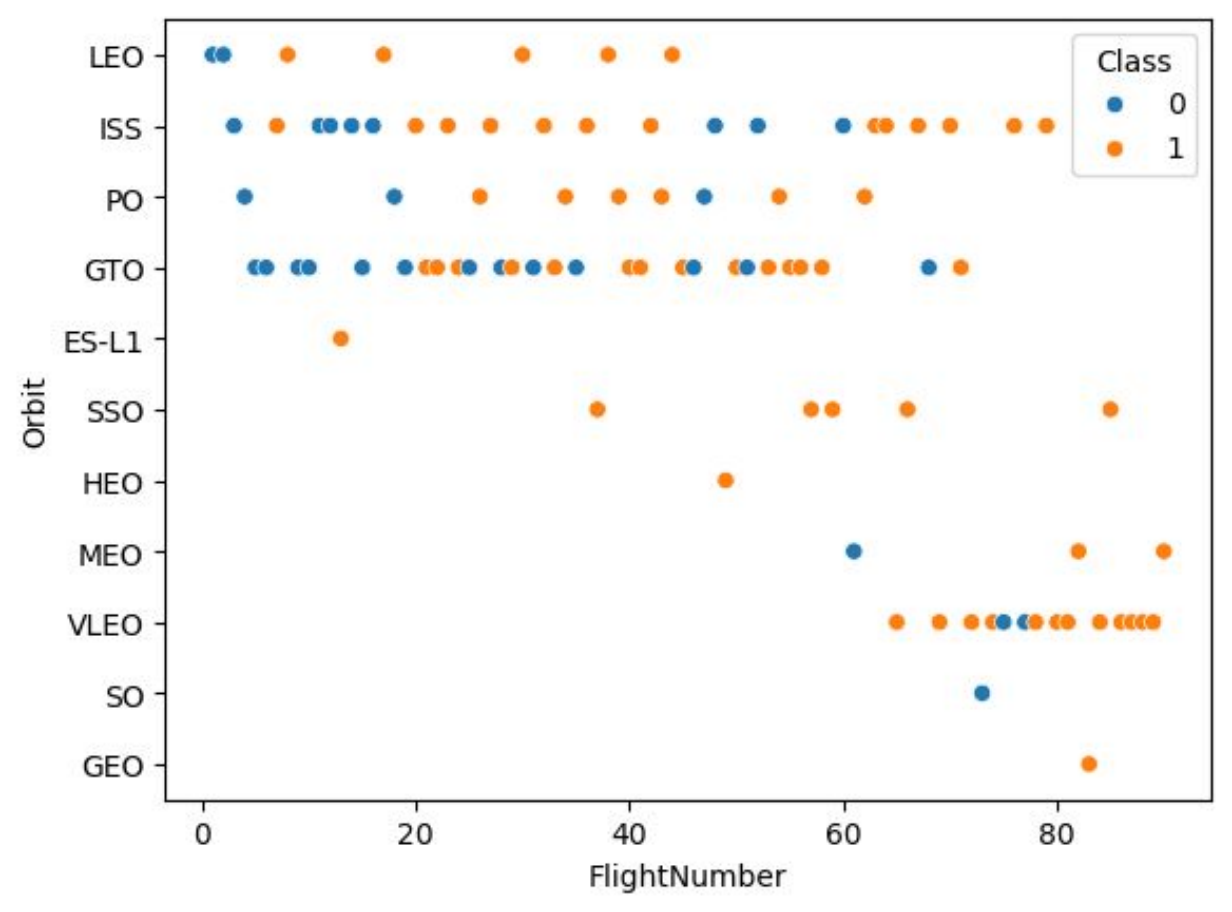
Success Rate vs. Orbit Type

Bar chart of the proportion of first stage landing successes where launches are grouped by the destination payload orbit.

Payloads sent to certain orbits like **GEO**, **SSO** and **VLEO**, correspond to a high landing success rate. Orbits like **ISS**, **LEO** and **MEO**, correspond to ~60% success rates.



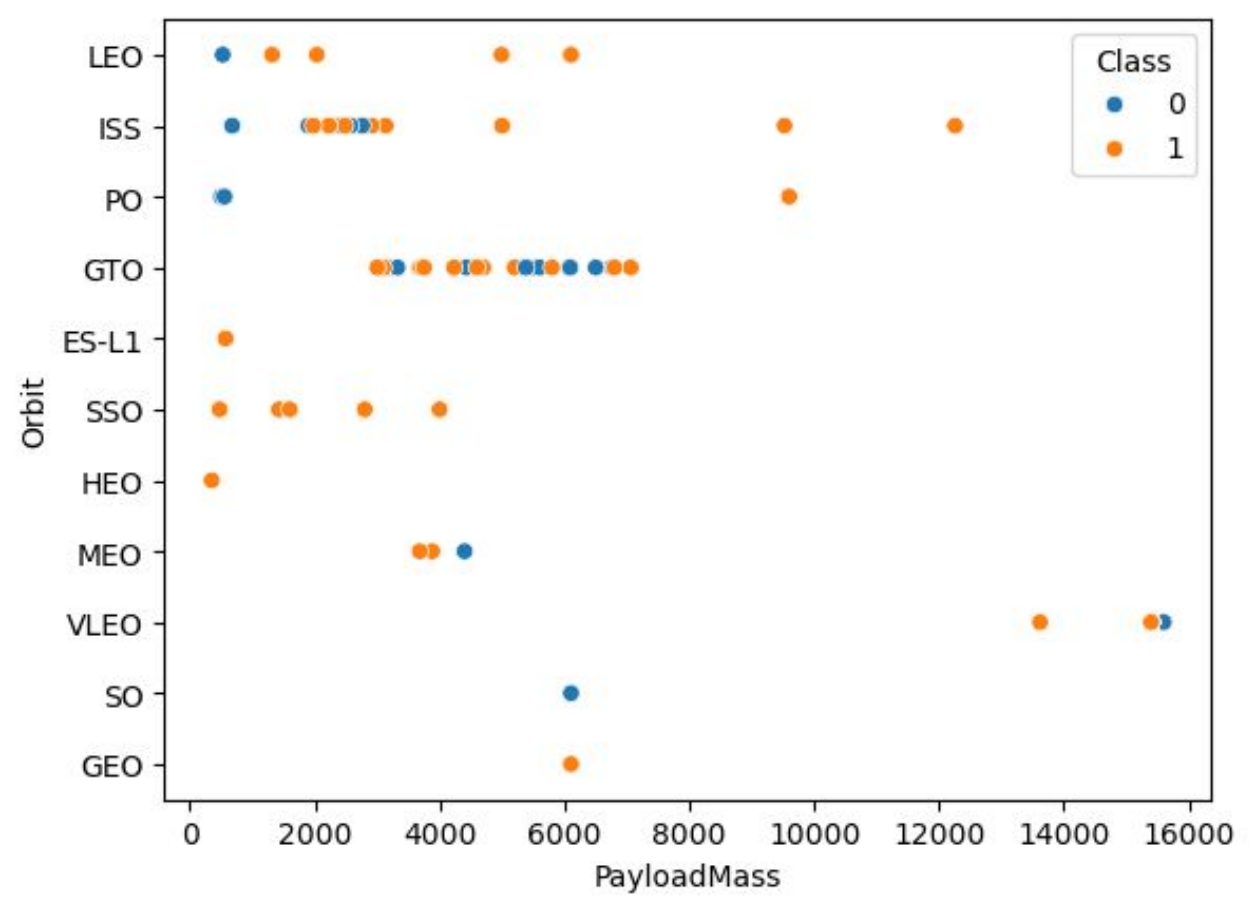
Flight Number vs. Orbit Type



Scatter plot showing all launches grouped by the payload orbit, with landing outcome represented in colour.

Many payloads have been sent into **GTO** and **VLEO** orbits, while very few payloads have been sent into **HEO** and **MEO** orbits.

Payload vs. Orbit Type



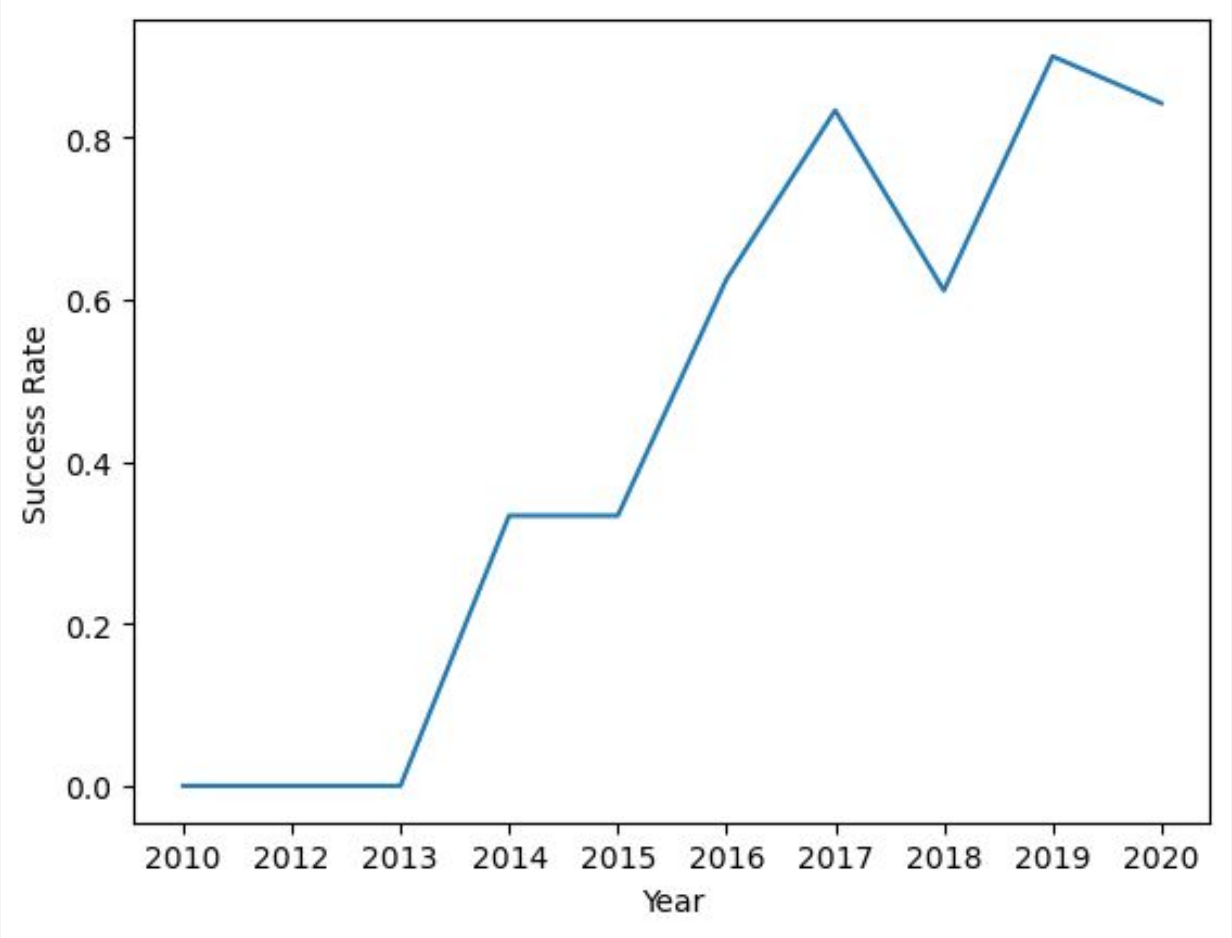
Scatter plot of all launches grouped by payload orbit and ordered by payload mass, with landing outcome by dot colour.

Most orbits groups feature launches with low payload masses, however all launches sent into **VLEO** orbit had payload masses exceeding 13000 kg.

Landing Success Yearly Trend

Line chart showing trends in annual landing success rate across the years from 2010 to 2020.

First stage landing success rates drastically increased from zero percent to nearly ninety percent starting from 2013.



All Launch Site Names

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

All four SpaceX payload launch sites.

Launch Site Names Begin with 'CCA'

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYL
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	
2010-12-08	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	

First five entries in dataset with a launch site name starting with **'CCA'**, where all five entries were launched from **CCAFS LC-40.**

Total Payload Mass

sum(PAYLOAD_MASS__KG_)
48213

Sum of payload masses from NASA (CRS) payloads launched by SpaceX.

Average Payload Mass by F9 v1.1

avg(PAYLOAD_MASS__KG_)
2928.4

Payloads with the booster **F9 v1.1** have an average payload mass of 2928.4 kilograms.

First Successful Ground Landing Date

min(Date)
2015-12-22

The first successful first stage landing was during the 12th of December 2015

Successful Drone Ship Landing with Payload between 4000 and 6000

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_
2016-05-06	5:21:00	F9 FT B1022	CCAFS LC-40	JCSAT-14	4696
2016-08-14	5:26:00	F9 FT B1026	CCAFS LC-40	JCSAT-16	4600
2017-03-30	22:27:00	F9 FT B1021.2	KSC LC-39A	SES-10	5300
2017-10-11	22:53:00	F9 FT B1031.2	KSC LC-39A	SES-11 / EchoStar 105	5200

There are four launches with a successful drone ship landing with a payload between 4000 and 6000 kilograms.

Total Number of Successful and Failure Mission Outcomes

Successes

61

```
%%sql
select count(*) as Fa
where Landing_Outcome
```

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

Failures

10

Total landing successes: 61

Total landing failures: 10

Boosters Carried Maximum Payload

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

All payloads with the maximum payload mass were launched using one of these boosters.

2015 Launch Records

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

There are two launches that resulted in drone ship failure in the year 2015. One was during January with te booster version F9 v1.1 B1012, and the other was during April with the booster version F9 v1.1 B1015.

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

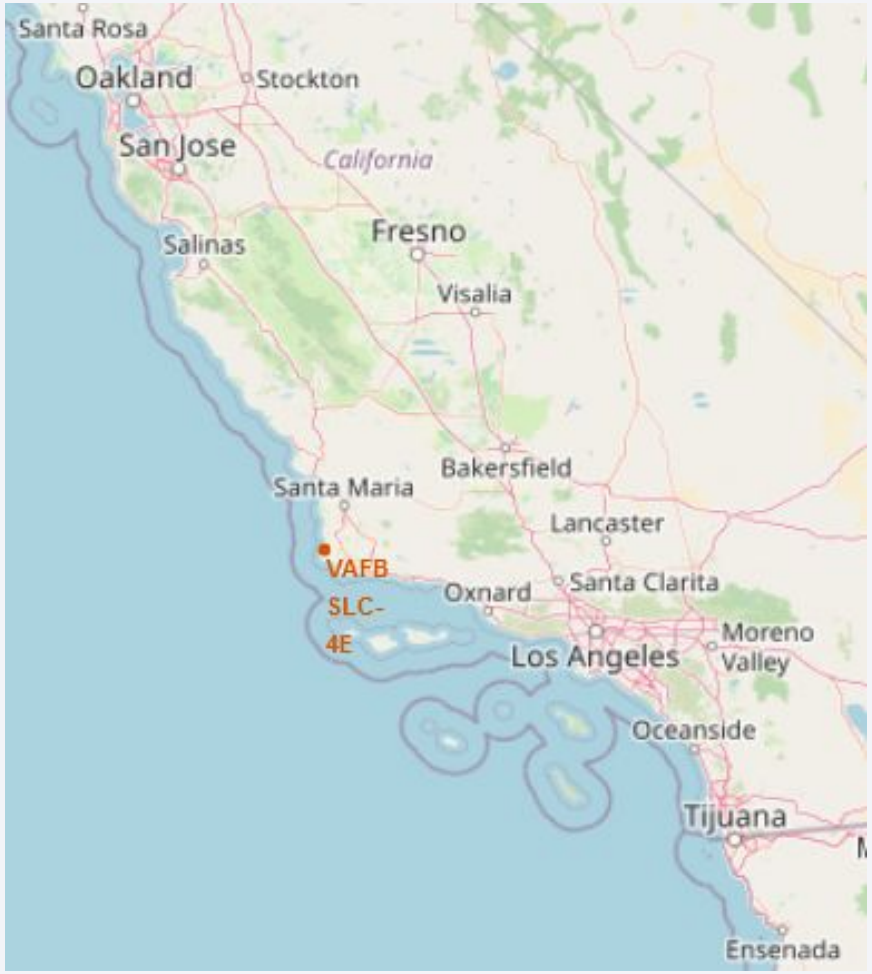
Landing_Outcome	Tally
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

A ranking of landing outcomes for launches between 4th June 2010 and 20th March 2017.

The most common outcome was not attempt to recover first stage, and the least common outcome was a precluded landing by drone ship.

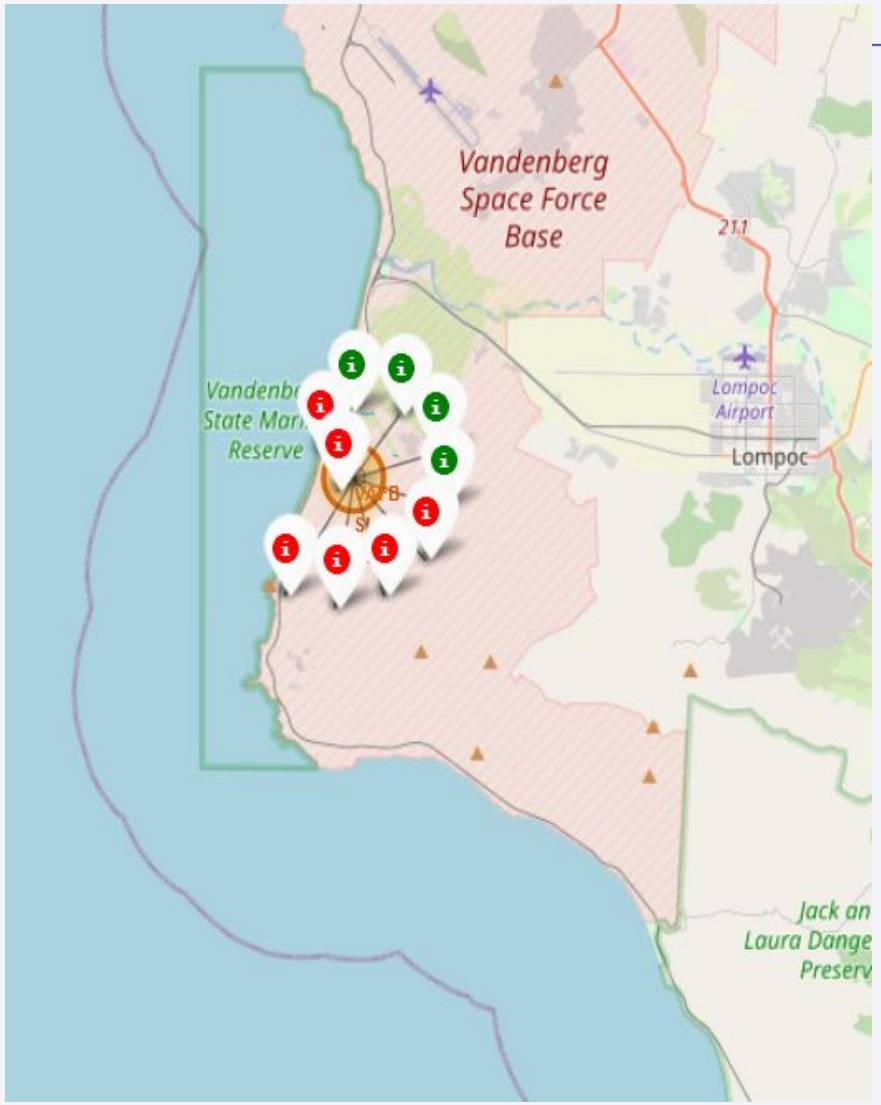

Section 3
Launch Sites
Proximities Analysis

Global SpaceX Launch Sites



SpaceX launch sites were located either at Vandenberg Space Force base near Los Angeles or next to John F. Kennedy Space Center in Florida.

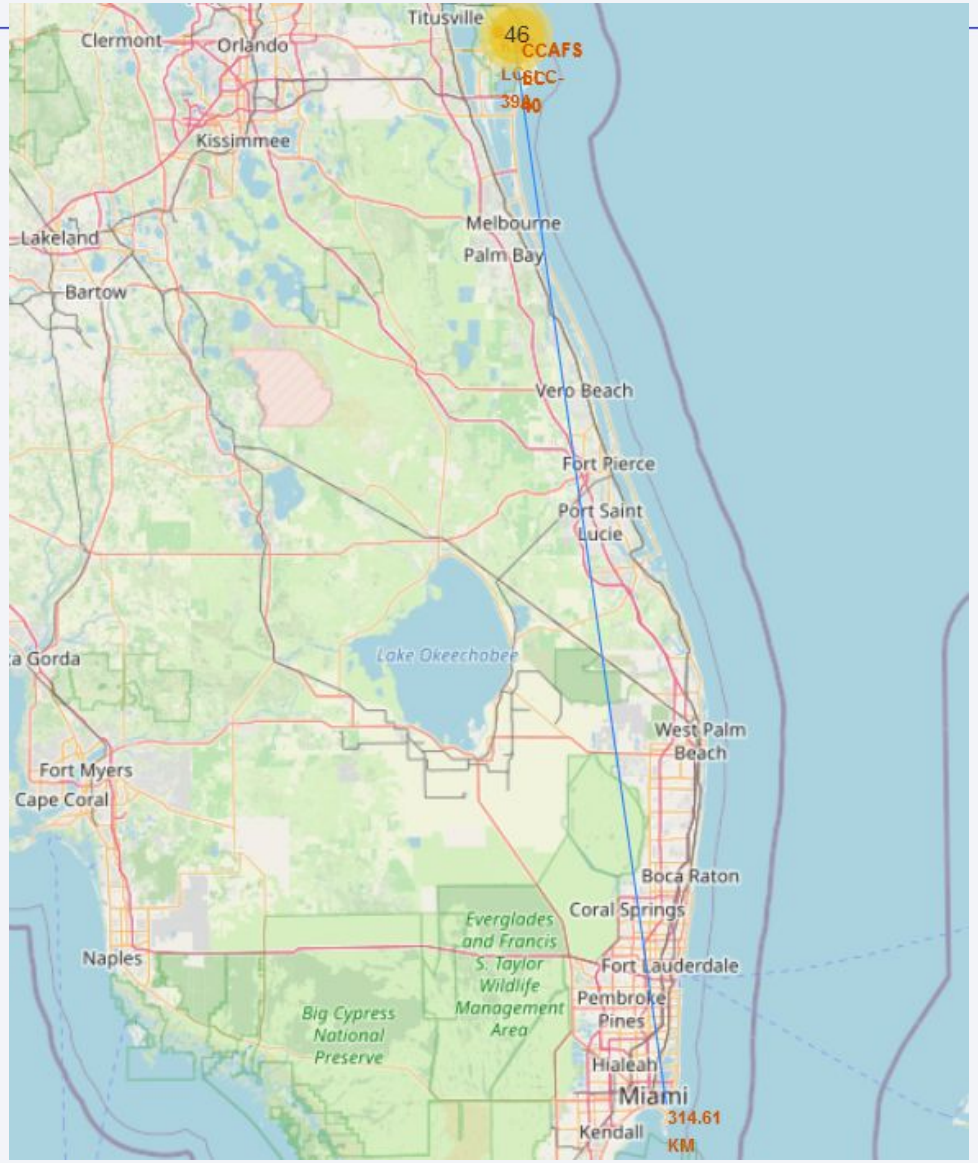
Launch Locations with Outcome



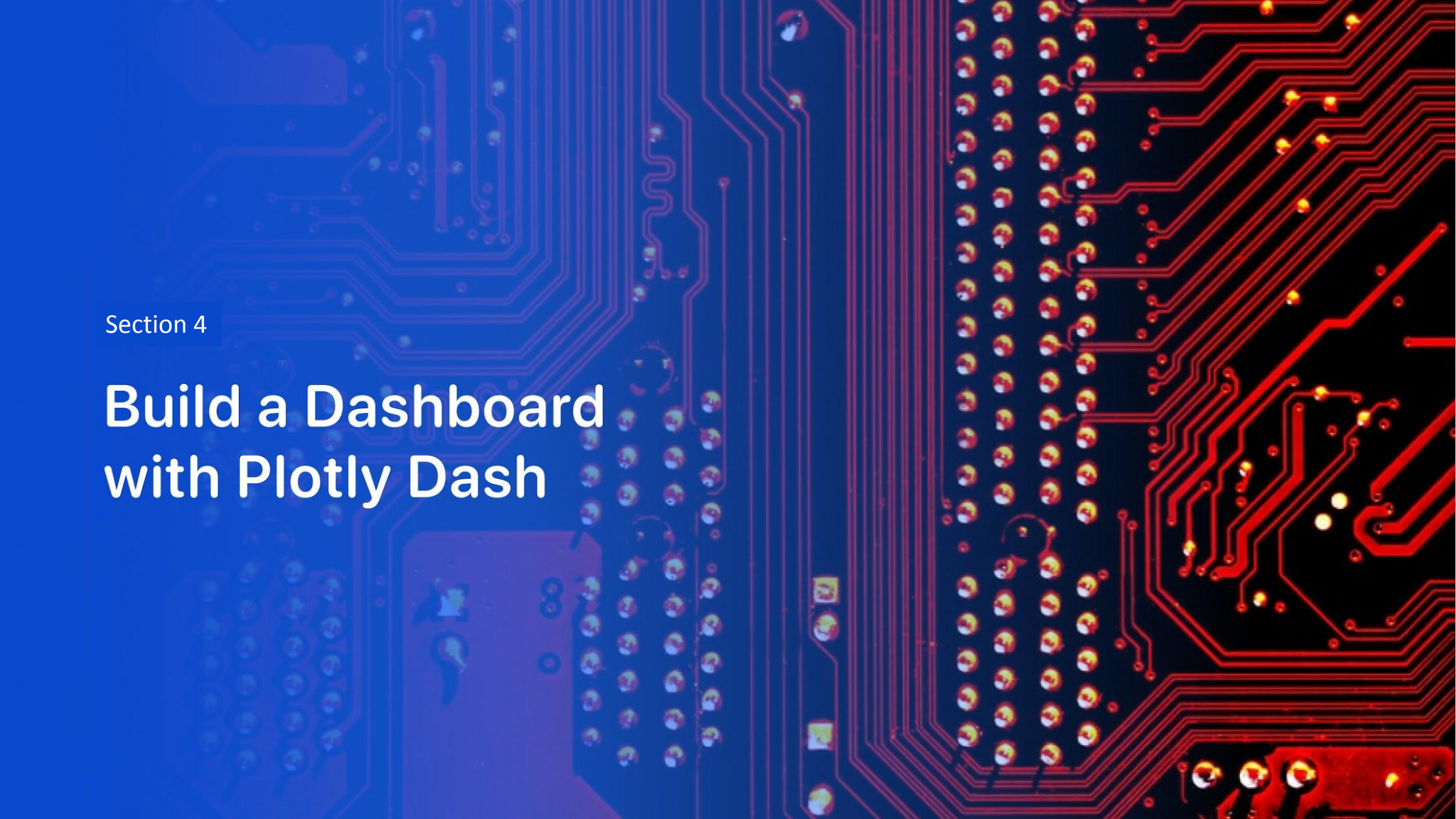
There were eleven launches, with four successful first stage landings, at Vandenberg Space Force base.

There were forty-six launches, with more than ten successful first stage landings, next to John F. Kennedy Space Center.

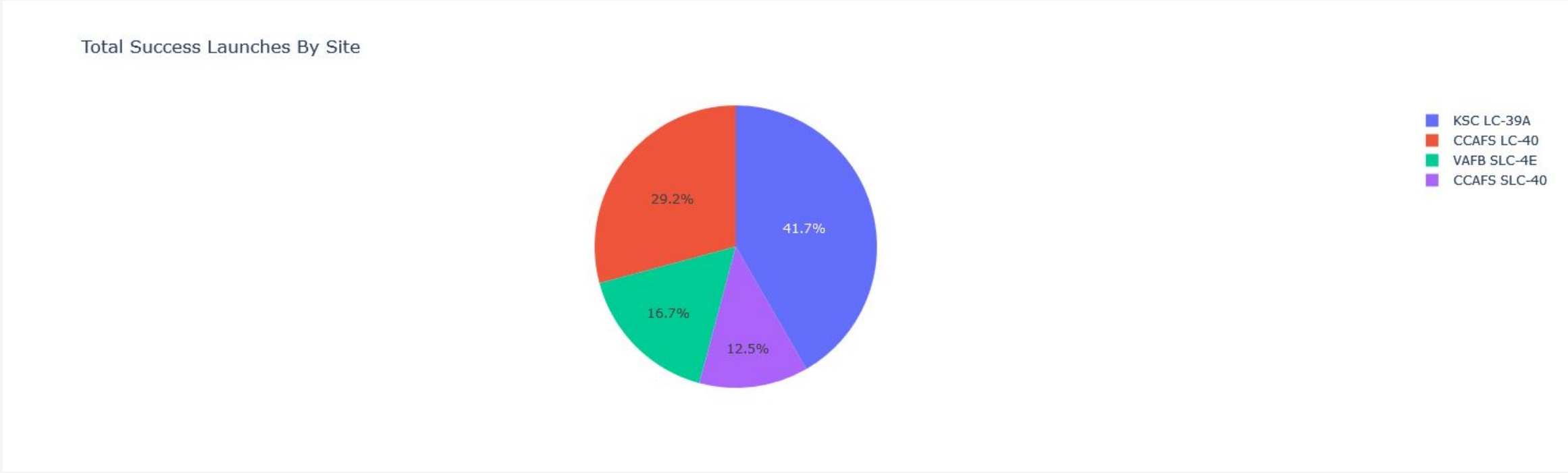
Distance between Launch Site and Major Cities



There is a distance of 314.61 km between the JFK SC launch sites and the city of Miami, heart of the largest metropolitan area in Florida.

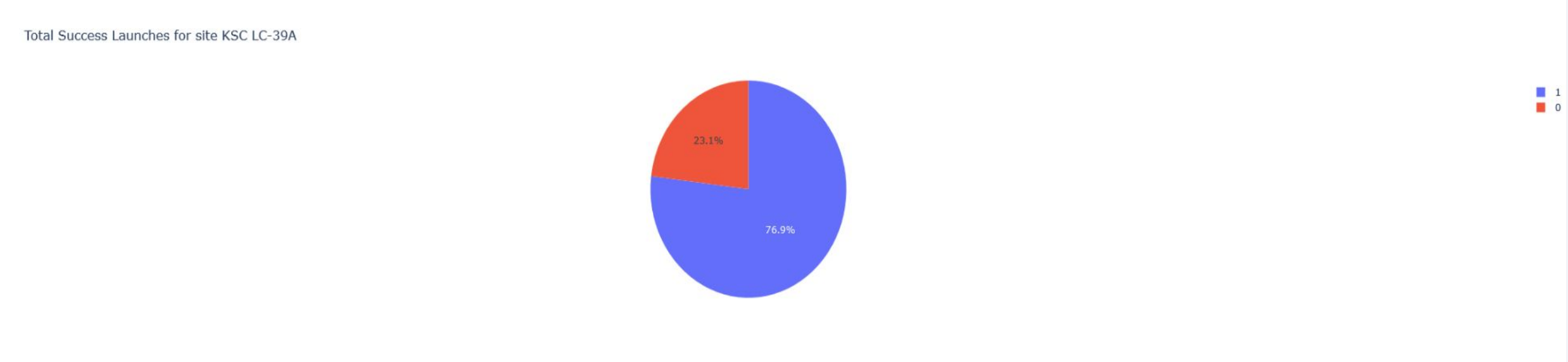

Section 4
Build a Dashboard with Plotly Dash

Launch Successes by Launch Site



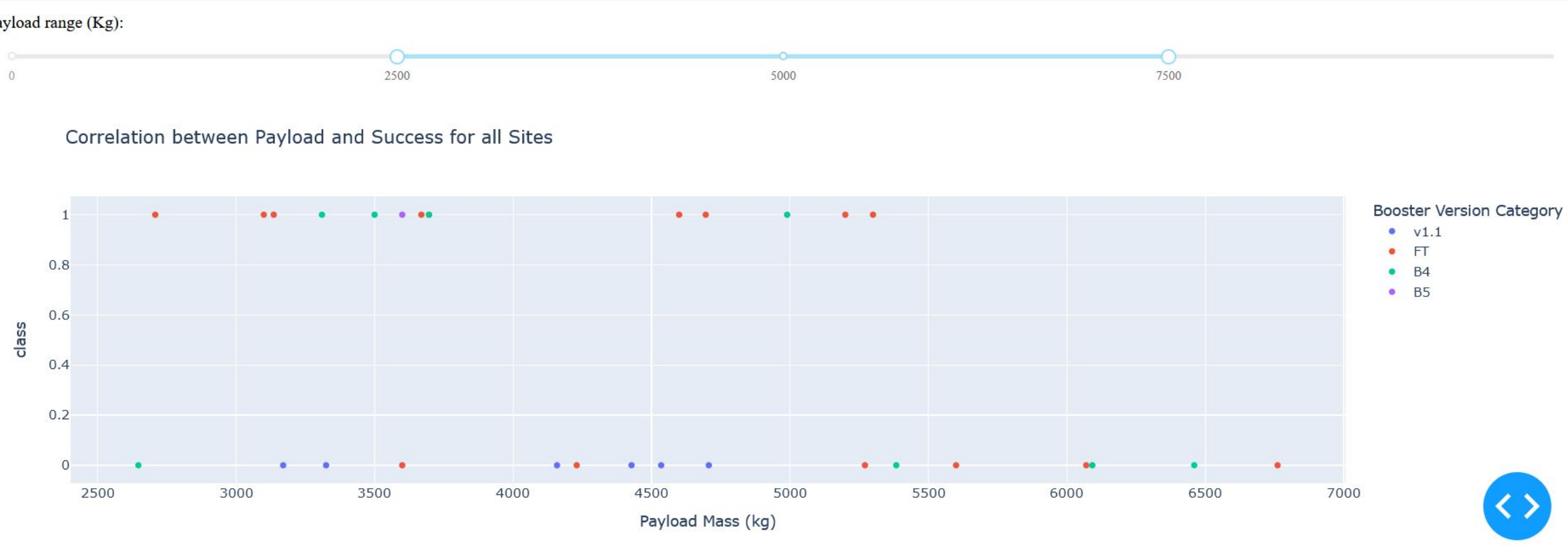
The highest proportion of successes come from site **KSC LC-39A**.
The lowest proportion of successes come from site **CCAFS SLC-40**.

Successes/Failures for KSC LC-39A



The launch site with the highest proportion of successes, KSC LC-39A, has a 76.9% success rate.

Payload Mass vs. Launch Outcome for all Sites



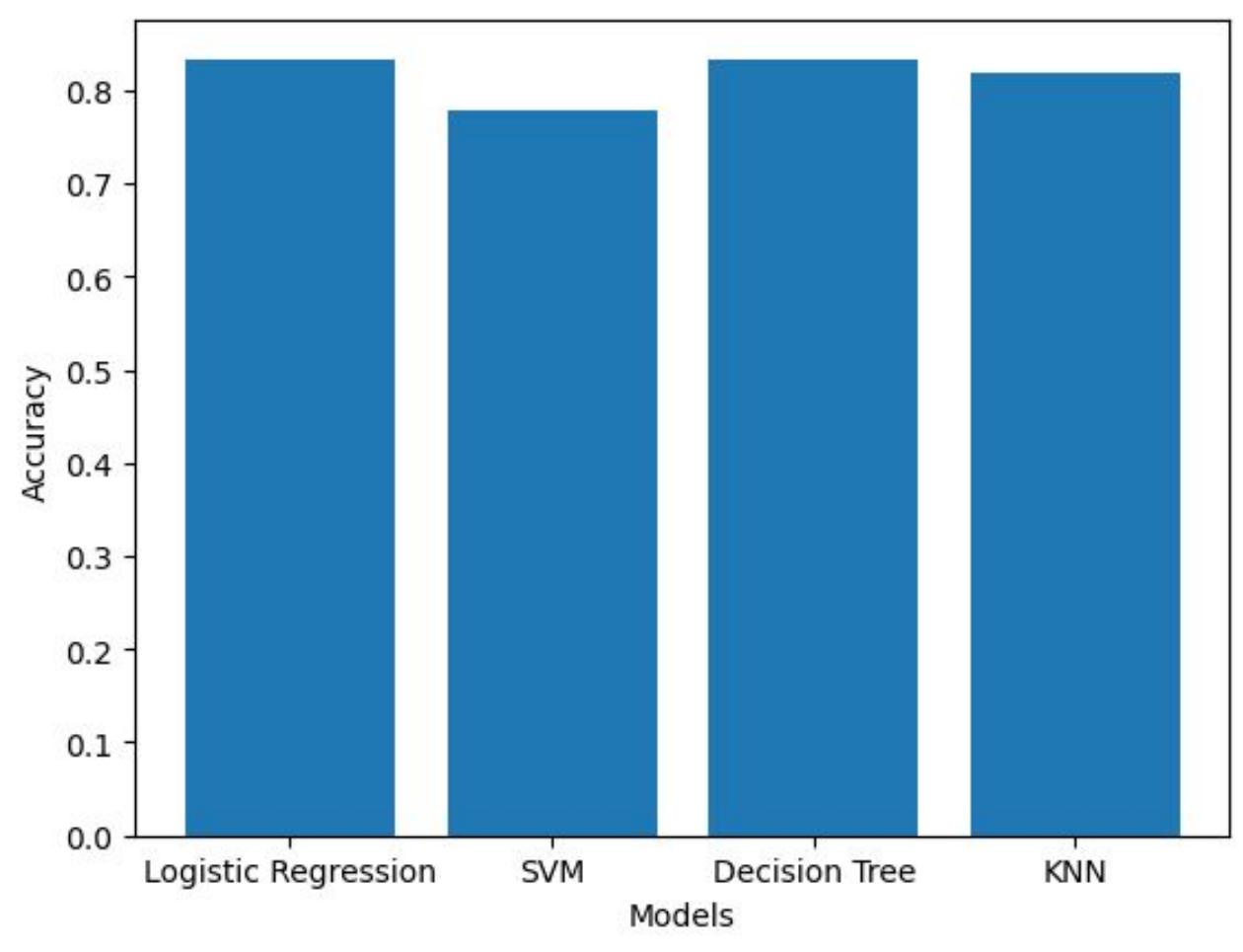
The average payload mass of payloads with booster versions **FT** and **B4** are higher than the average payload mass of payloads with booster versions **v1.1** and **B5**.

Section 5

Predictive Analysis (Classification)

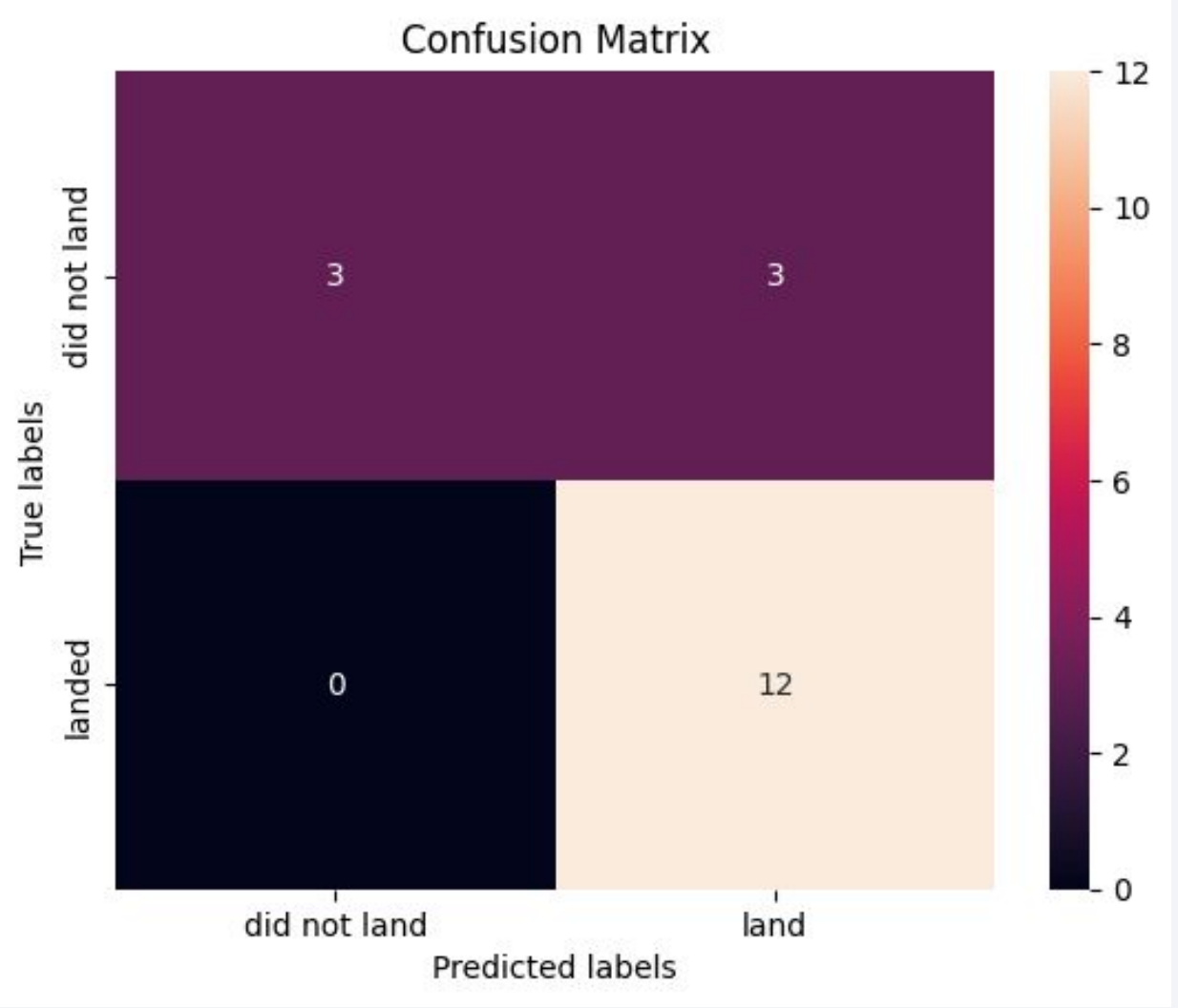
Classification Accuracy

- The logistic regression and decision tree models most accurately classify the landing outcome given the feature variables
- Mutual accuracy score of **0.833**



Confusion Matrix

- 100% precision, 50% recall for failed outcomes
- 80% precision, 100% recall for successful outcomes

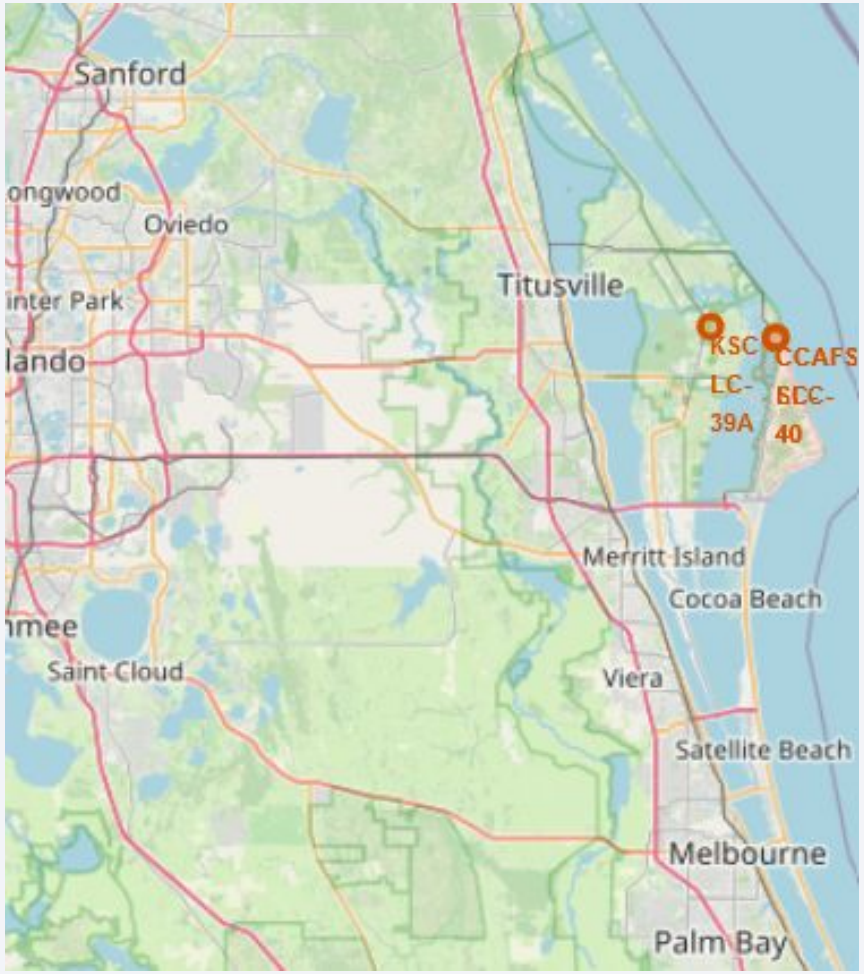


Conclusions

- The overall trend of landing success rate over time has changed from certain failure during 2010 to **very likely success** (~90%) during 2020.
- Landing success rate does tend to increase for **heavy payloads** over light payloads, however, the trend is not conclusive as there are comparatively few launches with heavy (>10000 kg) payloads.
- Most orbit types that correlate with high success rate have less than five launch data points, except for **VLEO** orbit, with high success rate (~90%) and sufficient data points (14 launches).
- The launch site **KSC LC 39A** has the highest overall success rate per site (76.9%) with sufficient data to support this conclusion (22 launches).

Appendix

Global Sites Second Screenshot



Launch Outcome Icons Second Screenshot

Thank you!