IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Andrea Neveling
May 2022

Outline

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

Executive Summary

- Summary of methodologies
  - Data was collected by webscraping and using the SpaceX API
  - Data wrangling and Exploratory Data Analysis (EDA) with Jupyter Notebooks
  - Machine Learning prediction of Falcon 9 first stage landings
  - Visualizations with Folium and Plotly Dash
- Summary of all results
  - Data collection, wrangling and exploration was successful
  - EDA and visualizations display certain correlations
  - Models achieved up to 83,33% accuracy

Introduction

- Project background
  - The new player SpaceY wants to compete with SpaceX
  - SpaceX advertises Falcon 9 rocket launches with a cost of 62 million dollars
  - Other providers cost upward of 165 million dollars each
  - SpaceX is cheaper because they can reuse the first stage

- Problems you want to find answers
  - SpaceY wants to bid against SpaceX for a rocket launch
  - Therefor they have to determine the cost of a launch
  - To be able to do this, we have to determine if the first stage will successfully land

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - The data used was gathered from the SpaceX REST API and web scraped from related Wiki pages with the Python module BeautifulSoup
- Perform data wrangling
  - Statistical calculations
  - Five NULL values in the PayloadMass column have been replaced with the mean
  - Additional columns have been added to facilitate the research
- 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

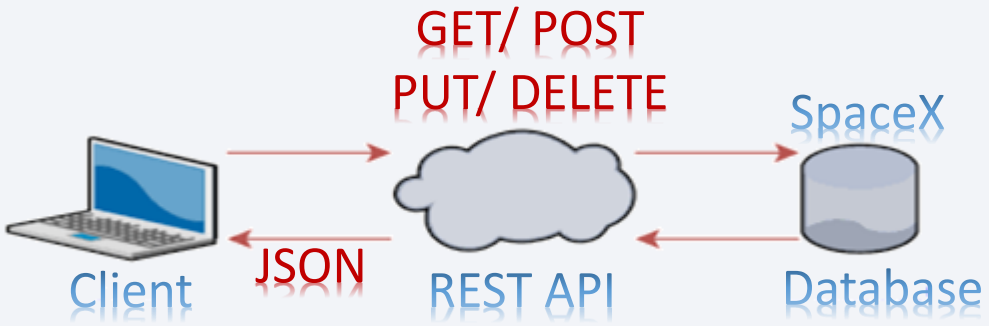
- The data used was gathered from the SpaceX REST API and webscraped from related Wiki pages with the Python module BeautifulSoup
- The gathered data contains information about launches: The rockets used, payload delivered, launch specifications, landing specifications, and landing outcome.
- Data was filtered to Falcon9 launches

Data Collection – SpaceX API

Link: Jupyter Notebook on GitHub

- SpaceX offers a public API:
  spacex_url=https://api.spacexdata.com/v4/launches/past

- The API was used with a GET-request according to the chart below…

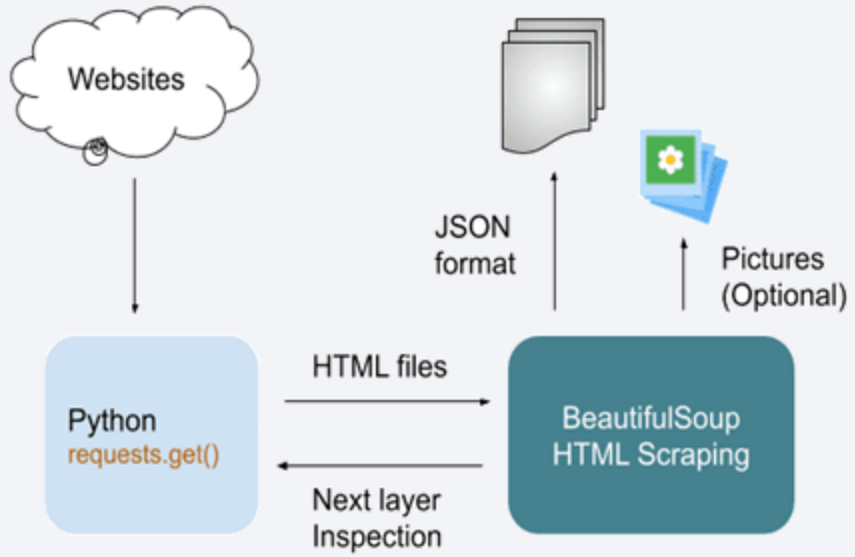


- …then the result was normalized and loaded into a Pandas dataframe:
  data = pd.json_normalize(response.json())

Data Collection - Scraping

Link: Jupyter Notebook on GitHub

- HTML data was downloaded from Wikipedia:
  https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922
- HTML Tables on SpaceX launches were processed through the Python module BeautifulSoup
- Processed data was loaded into a Pandas dataframe



Data Wrangling

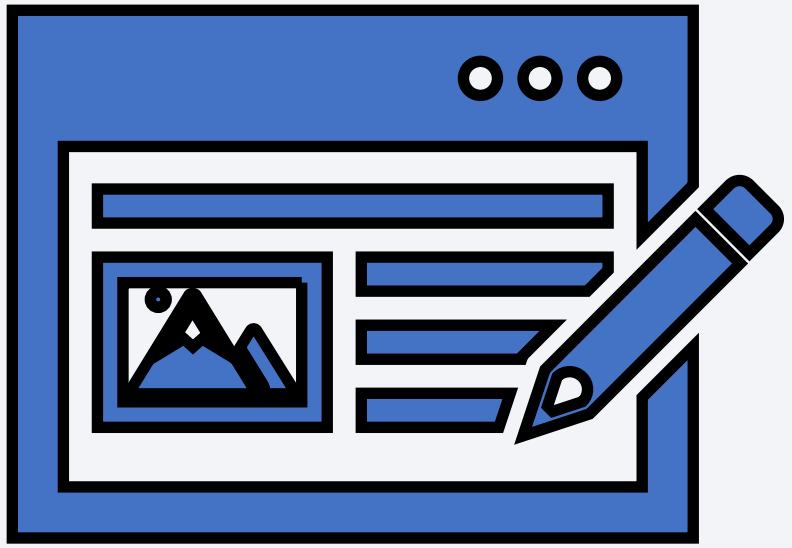
Link: Jupyter Notebook on GitHub

- Identification of data types, missing values etc.
- Calculations of launches per site, mission outcomes etc.
- New column inserted into data frame for labeling different outcomes as landing success (0/ 1)

EDA with Data Visualization

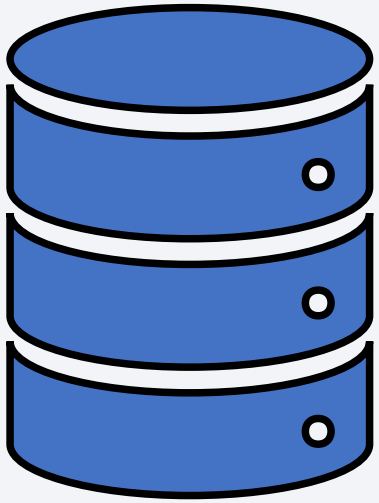
Link: Jupyter Notebook on GitHub

- Python modules Matblotlib and Seaborn were used for visualization
- Scatter plots, bar plots and linear plots were built to find relations of features and mission outcomes:
  - Relationship between Flight Number and Launch Site
  - Relationship between Payload and Launch Site
  - Relationship between success rate of each orbit type
  - Relationship between FlightNumber and Orbit type
  - Relationship between Payload and Orbit type
  - Launch success yearly trend



EDA with SQL

Link: Jupyter Notebook on GitHub



- SQL queries performed
  - Names of the unique launch sites in the space mission
  - Total payload mass carried by boosters launched by NASA (CRS)
  - Average payload mass carried by booster version F9 v1.1
  - Date when the first successful landing outcome in ground pad was achieved
  - Boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
  - Total numbers of successful and failure mission outcomes
  - Booster Versions which have carried the maximum payload mass
  - Failed landing outcomes in drone ship, their booster versions, and launch site names (in year 2015)
  - Ranking of the count of landing outcomes

Build an Interactive Map with Folium

Link: Jupyter Notebook on GitHub

- Finding an optimal location for building a launch site involves many factors
- We built maps with markers to discover some of those factors by analyzing the existing launch site locations (e.g. depending on the outcome)
- We also calculate the distances between a launch site to its proximities

Build a Dashboard with Plotly Dash

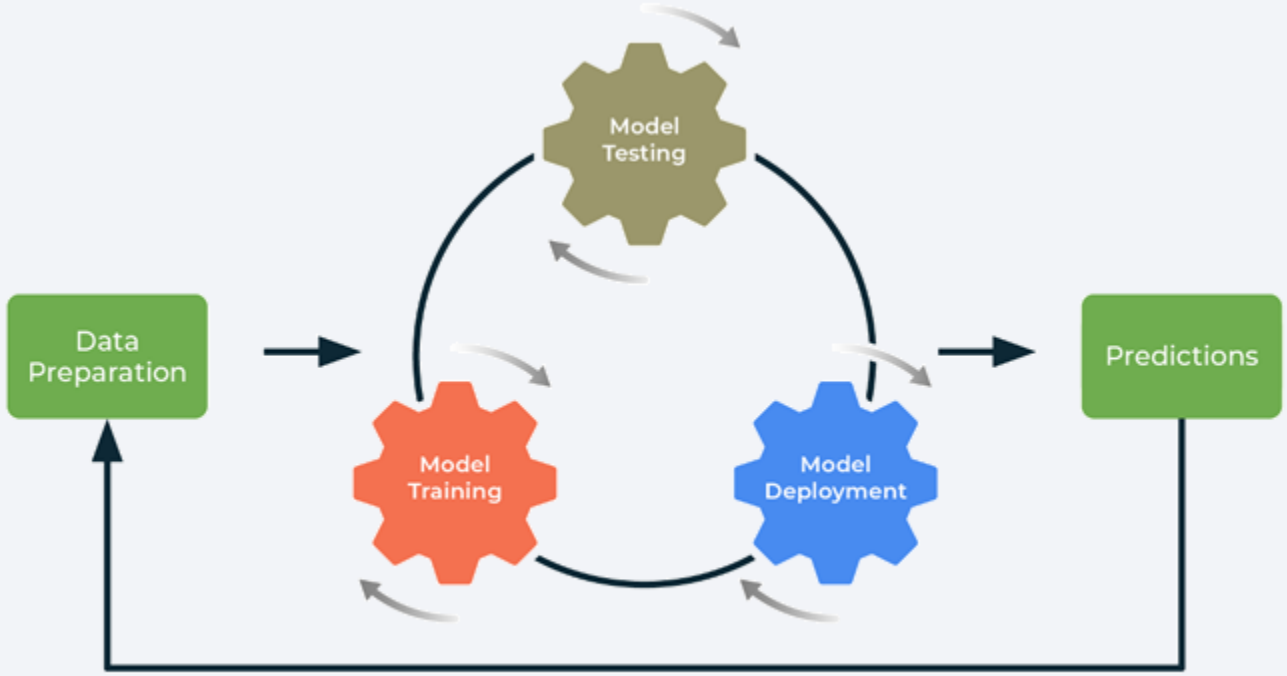
Link: Jupyter Notebook on GitHub

- This interactive dashboard allows the user to choose a site and/ or the payload as filters
- The following plots were used to visualize data
  - Percentage of launches by site
  - Payload range
- A pie chart visualizes the success rate of sites
- A scatter plot visualizes the relationship between payload and outcome

Predictive Analysis (Classification)

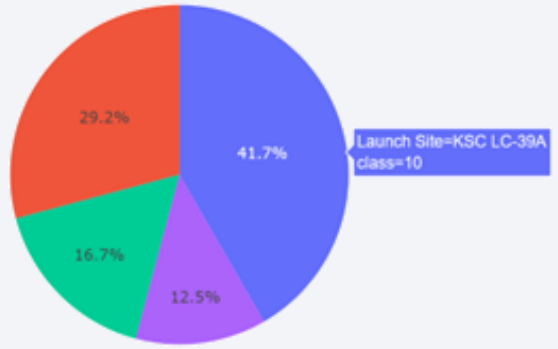
Link: Jupyter Notebook on GitHub

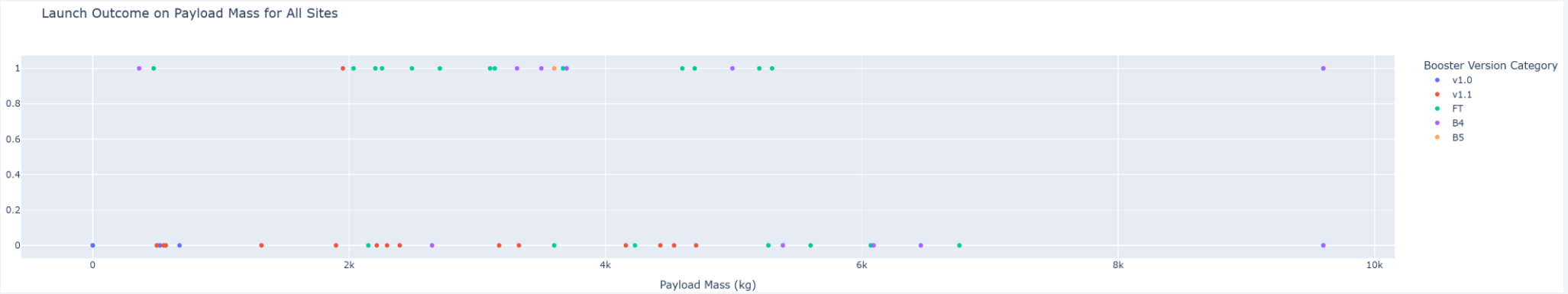
- Comparison of the performance of 4 classification algorithms:
  - Logistic Regression
  - Support Vector Machine
  - Decision Tree
  - K Nearest Neighbors

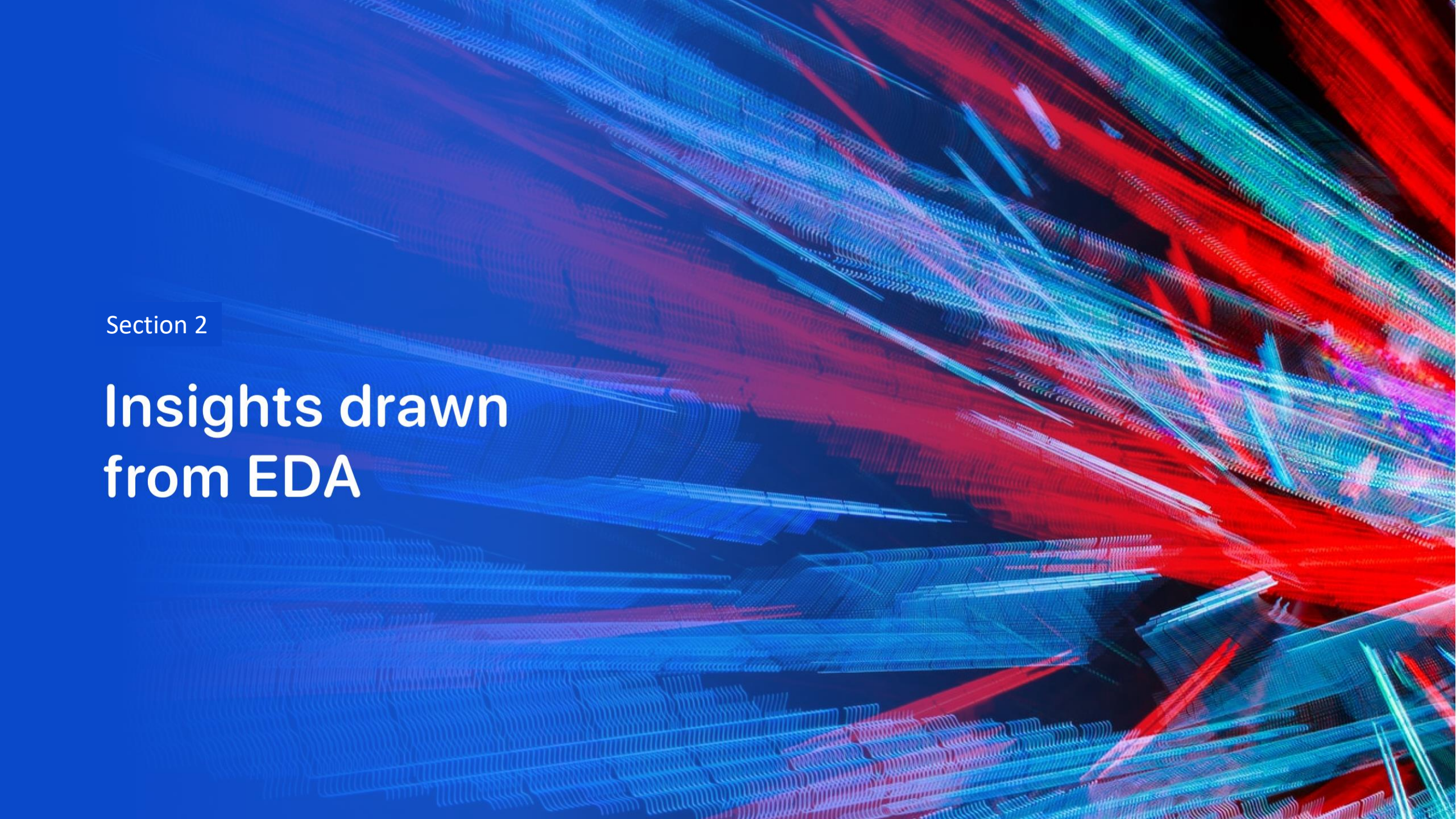


Results

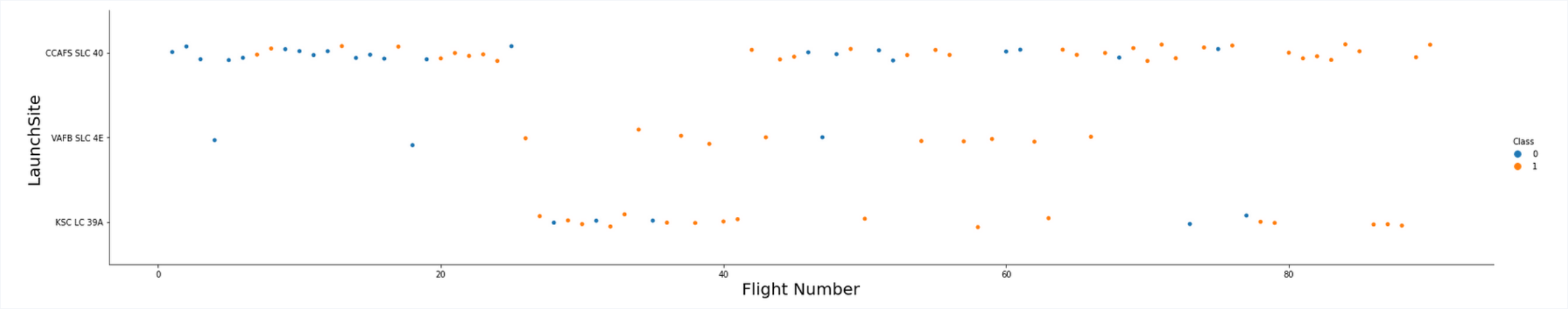
- Space X uses four different launch sites
- Launch site, orbit, payload and flight numbers correlate with the outcome
- Launch site KSC LC-39A had the most successful launches and with 76,9% also the highest launch success rate
- Many Falcon 9 booster versions successfully landed with a payload above the average
- Booster version 1.1 has been only successful below 2000kg payload mass
- The number of successful landing outcomes increases over the years
- The predictive models achieved up to 83,33% accuracy.



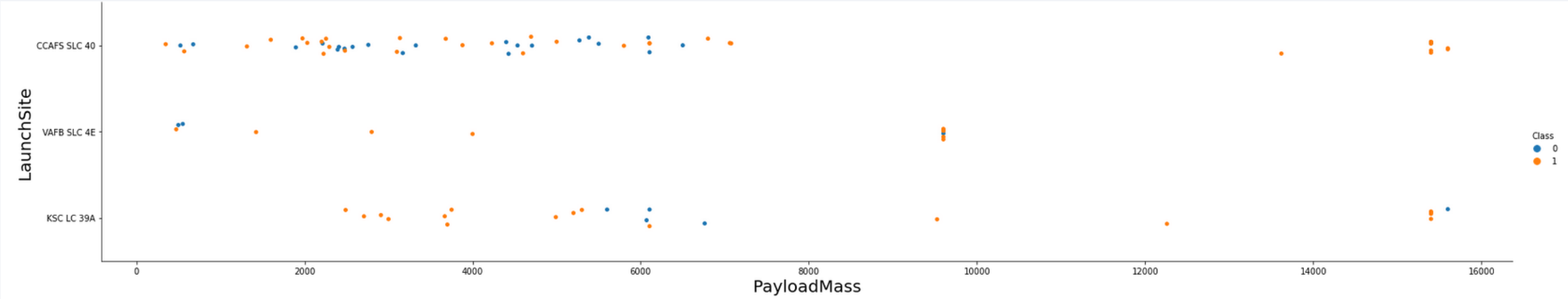



Section 2
Insights drawn from EDA

Flight Number vs. Launch Site



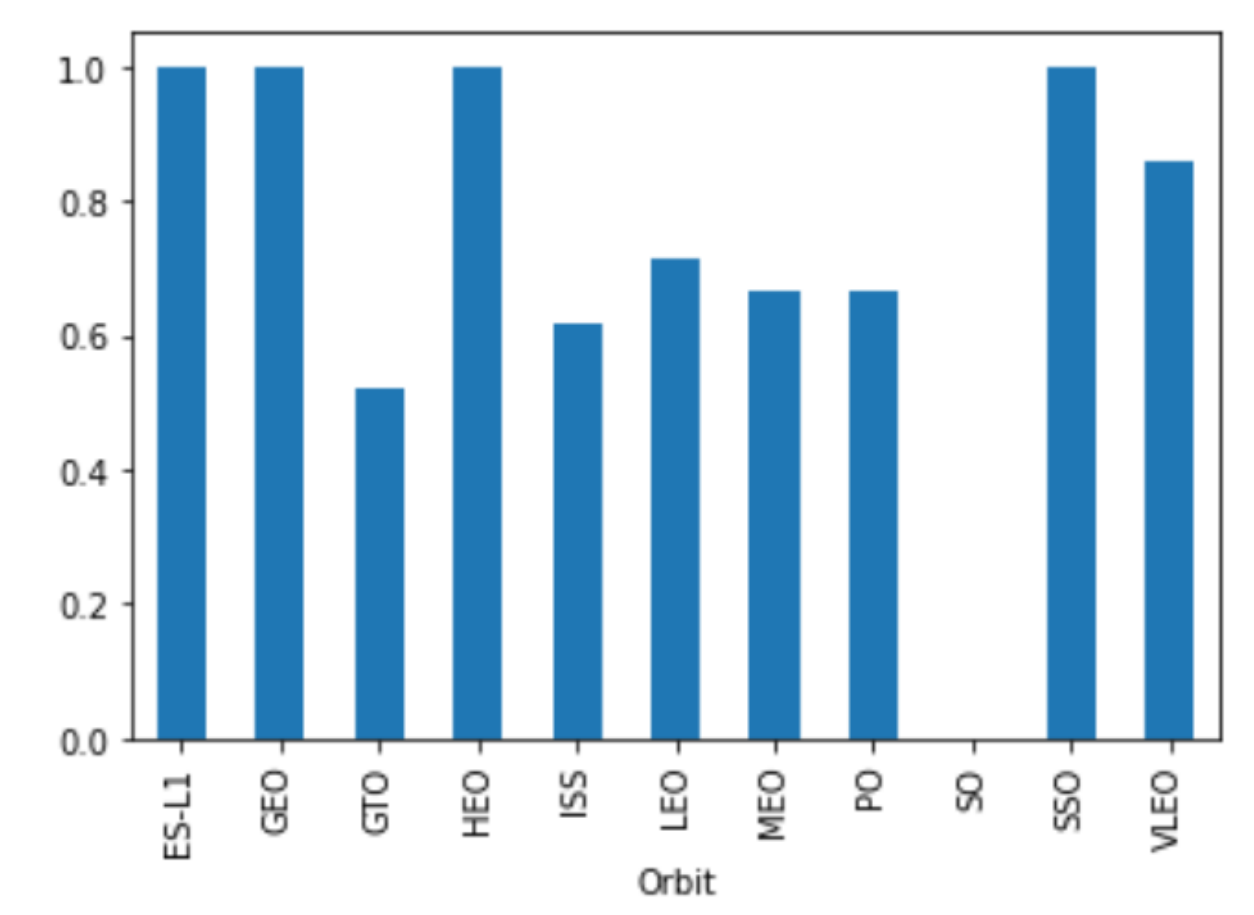
Payload vs. Launch Site



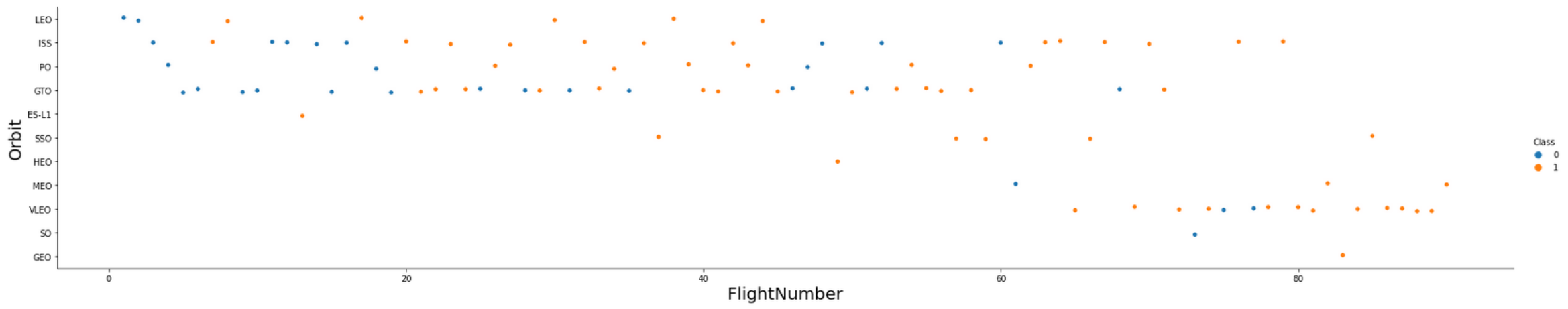
Success Rate vs. Orbit Type

- Show a bar chart for the success rate of each orbit type

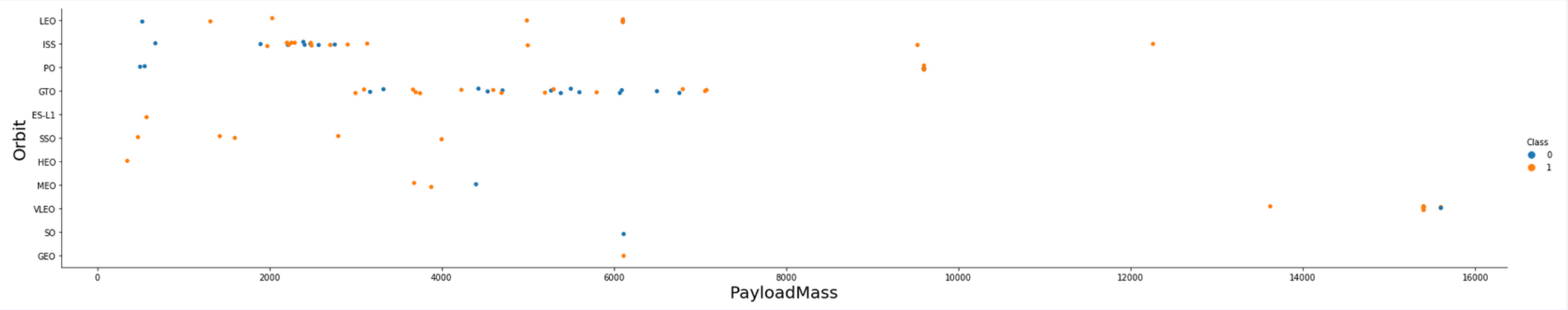
- Show the screenshot of the scatter plot with explanations



Flight Number vs. Orbit Type

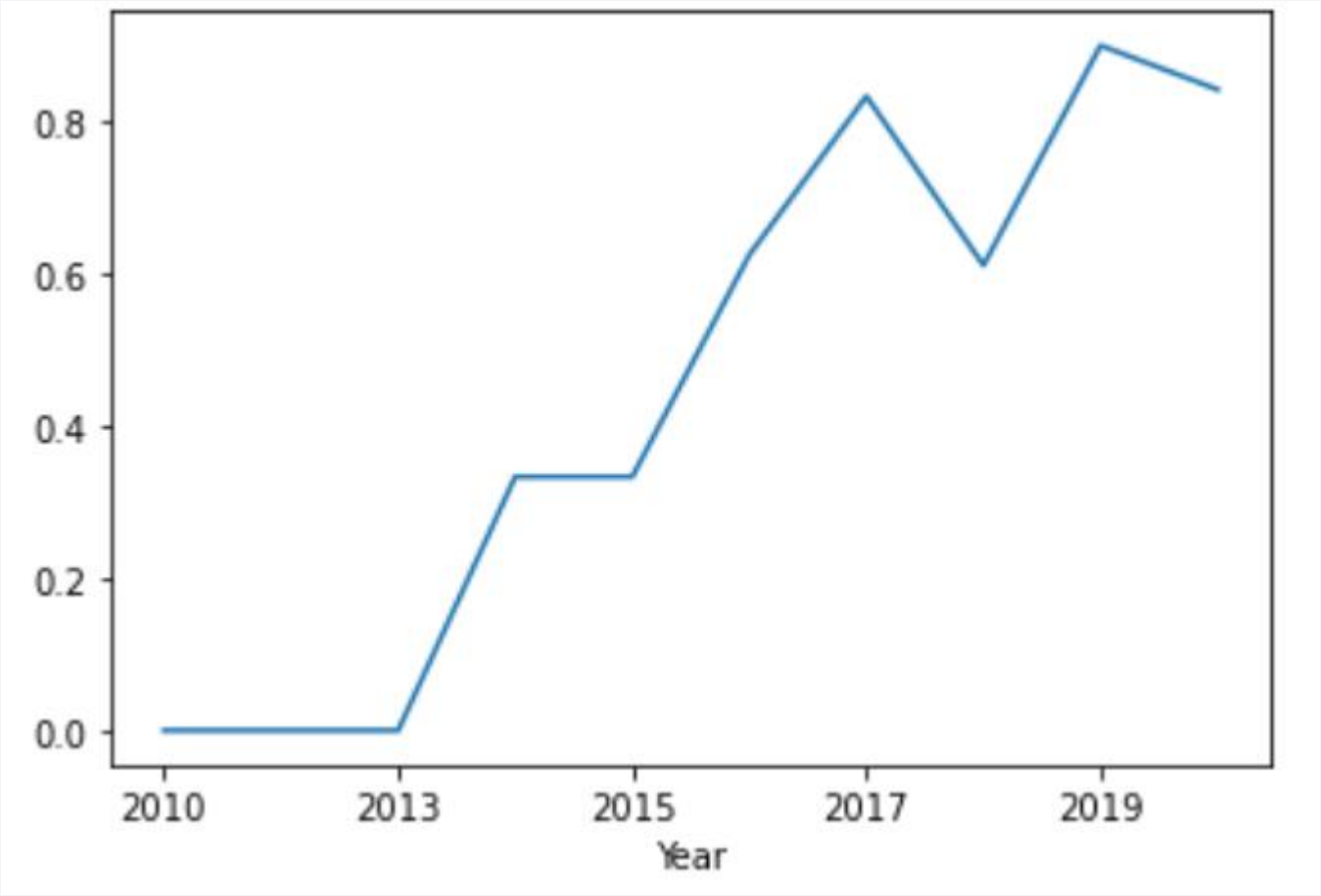


Payload vs. Orbit Type



Launch Success Yearly Trend

- Successful outcomes increase with the years
- This might relate to technological and analytical progress



All Launch Site Names

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

```
%sql SELECT DISTINCT LAUNCH_SITE FROM SPACEXTBL ORDER BY 1
```

Launch Site Names Begin with 'CCA'

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

```
 * ibm_db_sa://zct66184:***@125f9f61-9715-46f9-9399-c8177b21803b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:30426/bludb
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)
2012-08-10	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
2013-03-12	22:41:00	F9 v1.1	CCAFS LC-40	SES-8	3170	GTO	SES	Success	No attempt

Total Payload Mass

The total payload carried by boosters from NASA is 56479 kg

```
%sql SELECT SUM(PAYLOAD_MASS__KG_) AS TOTAL_PAYLOAD FROM SPACEXTBL WHERE PAYLOAD  LIKE '%CRS%'
```

```
 * ibm_db_sa://zct66184:***@125f9f61-9715-46f9-9399-c8177b21803b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:30426/bludb
Done.
```

total_payload
56479

Average Payload Mass by F9 v1.1

The average payload mass carried by booster version F9 v1.1 is 3676 kg

```
%sql SELECT AVG(PAYLOAD_MASS__KG_) AS AVG_PAYLOAD FROM SPACEXTBL WHERE BOOSTER_VERSION = 'F9 v1.1'
```

```
 * ibm_db_sa://zct66184:***@125f9f61-9715-46f9-9399-c8177b21803b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:30426/bludb
Done.
```

avg_payload
3676

First Successful Ground Landing Date

```
%sql SELECT MIN(DATE) AS FIRST_SUCCESS_GP FROM SPACEXTBL WHERE LANDING__OUTCOME = 'Success (ground pad)'
```

```
 * ibm_db_sa://zct66184:***@125f9f61-9715-46f9-9399-c8177b21803b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:30426/bludb
Done.
```

first_success_gp
2017-01-05

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%sql SELECT DISTINCT BOOSTER_VERSION FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000 AND LANDING__OUTCOME = 'Succes
s (drone ship)'
```

```
 * ibm_db_sa://zct66184:***@125f9f61-9715-46f9-9399-c8177b21803b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:30426/bludb
Done.
```

booster_version
F9 FT B1031.2
F9 FT B1022

Total Number of Successful and Failure Mission Outcomes

```
%sql SELECT MISSION_OUTCOME, COUNT(*) AS QTY FROM SPACEXTBL GROUP BY MISSION_OUTCOME ORDER BY MISSION_OUTCOME
```

```
 * ibm_db_sa://zct66184:***@125f9f61-9715-46f9-9399-c8177b21803b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:30426/bludb
Done.
```

mission_outcome	qty
Success	44
Success (payload status unclear)	1

Boosters Carried Maximum Payload

List of boosters which have carried the maximum payload mass:

- F9 B5 B1048.4
- F9 B5 B1049.4
- F9 B5 B1049.5
- F9 B5 B1058.3
- F9 B5 B1060.2

```
%sql SELECT DISTINCT BOOSTER_VERSION FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL) ORDER BY BOOSTER_VERSION
```

```
 * ibm_db_sa://zct66184:***@125f9f61-9715-46f9-9399-c8177b21803b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:30426/bludb
Done.
```

booster_version
F9 B5 B1048.4
F9 B5 B1049.4
F9 B5 B1049.5
F9 B5 B1058.3
F9 B5 B1060.2

2015 Launch Records

```
%sql SELECT BOOSTER_VERSION, LAUNCH_SITE FROM SPACEXTBL WHERE LANDING__OUTCOME = 'Failure (drone ship)' AND DATE_PART('YEAR', DATE) = 2015
```

```
 * ibm_db_sa://zct66184:***@125f9f61-9715-46f9-9399-c8177b21803b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:30426/bludb
Done.
```

booster_version	launch_site
F9 v1.1 B1012	CCAFS LC-40

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

- No attempt: 7
- Failure (drone ship): 2
- Success (drone ship): 2
- Success (ground pad): 2
- Controlled (ocean): 1
- Failure (parachute): 1

```
%sql SELECT LANDING__OUTCOME, COUNT(*) AS QTY FROM SPACEXTBL WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY LANDING__
OUTCOME ORDER BY QTY DESC
```

```
 * ibm_db_sa://zct66184:***@125f9f61-9715-46f9-9399-c8177b21803b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:30426/bludb
Done.
```

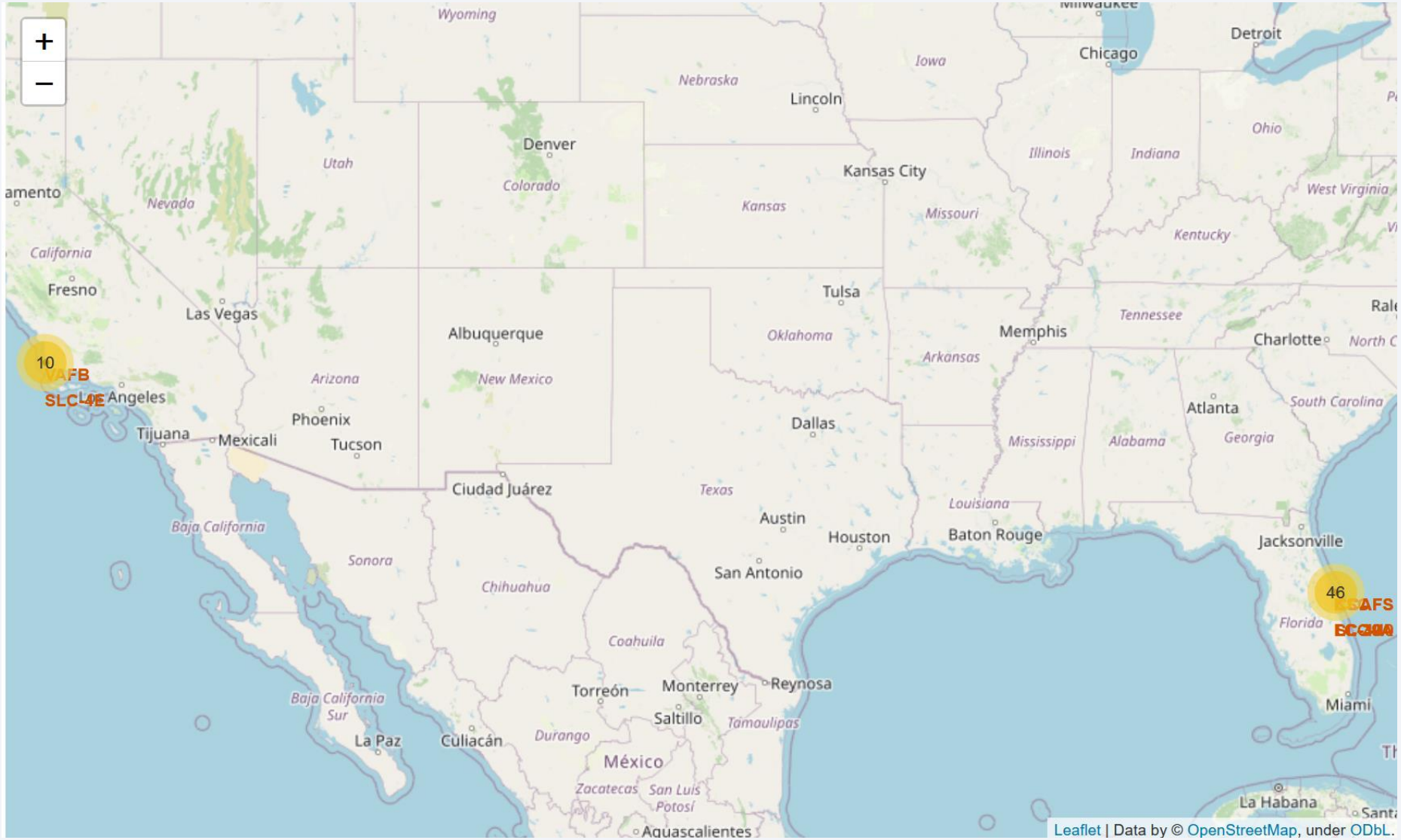
landing__outcome	qty
No attempt	7
Failure (drone ship)	2
Success (drone ship)	2
Success (ground pad)	2
Controlled (ocean)	1
Failure (parachute)	1

Section 3

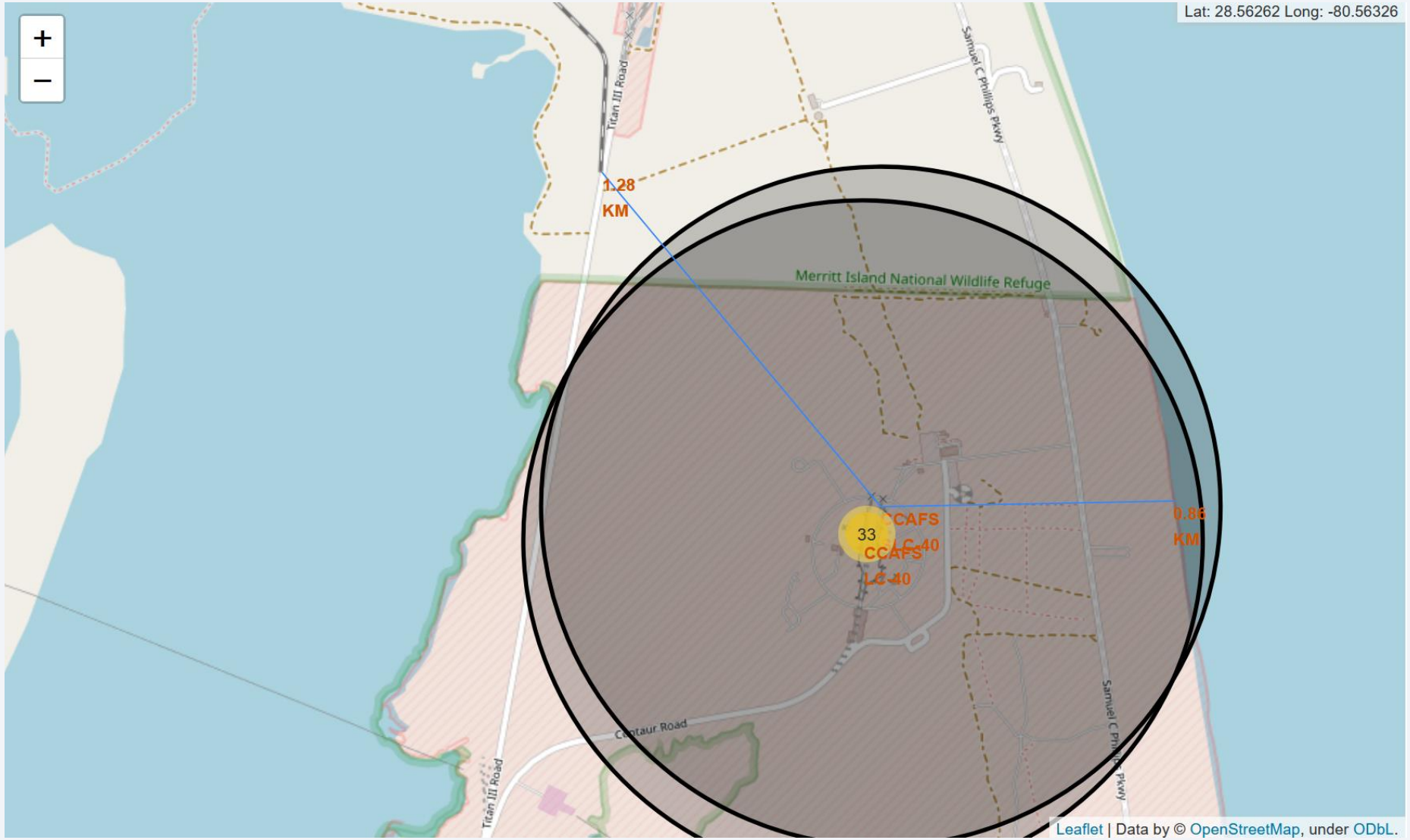
Launch Sites Proximities Analysis

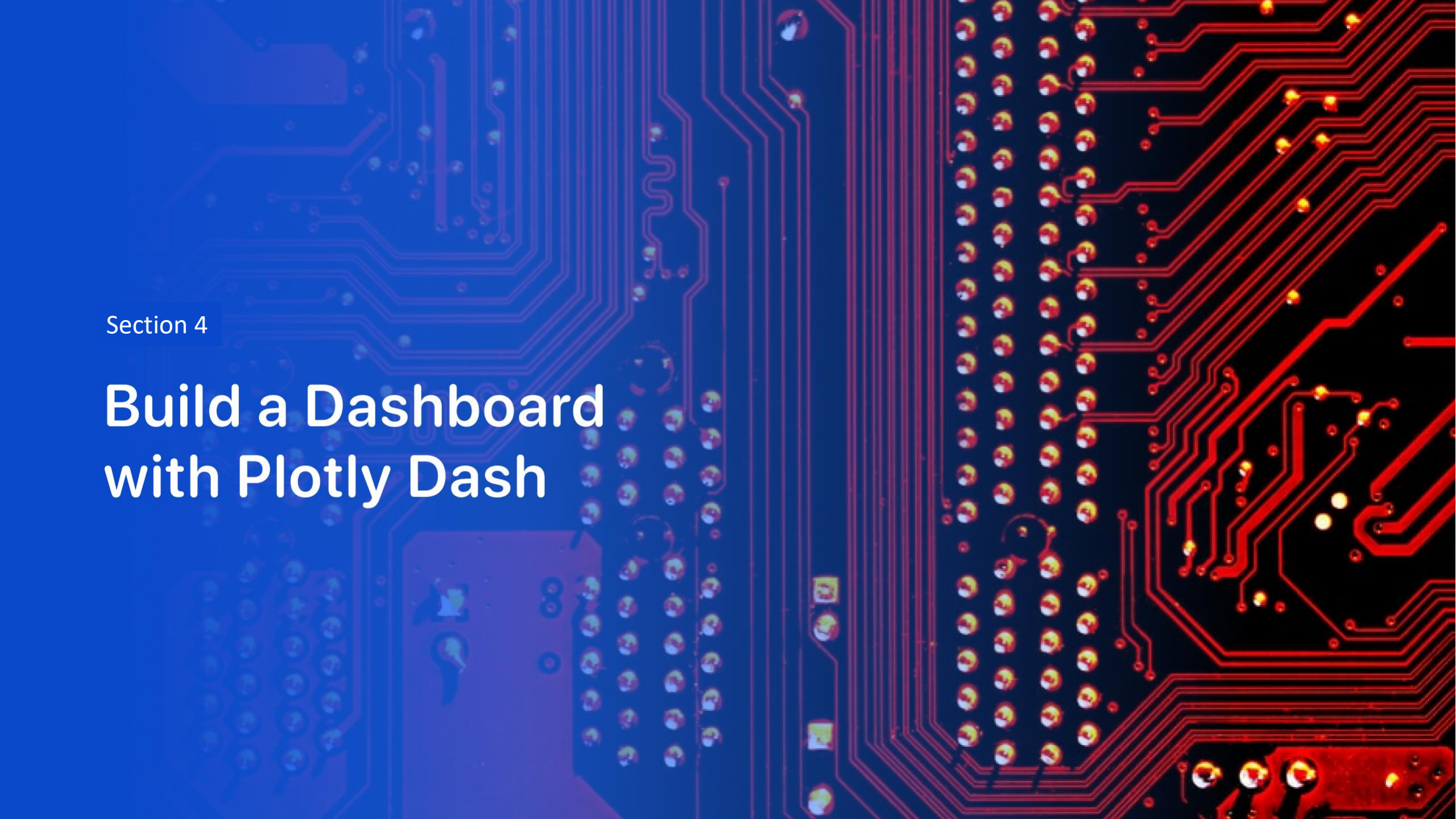
<Folium Map Screenshot 1>

<Folium Map Screenshot 2>

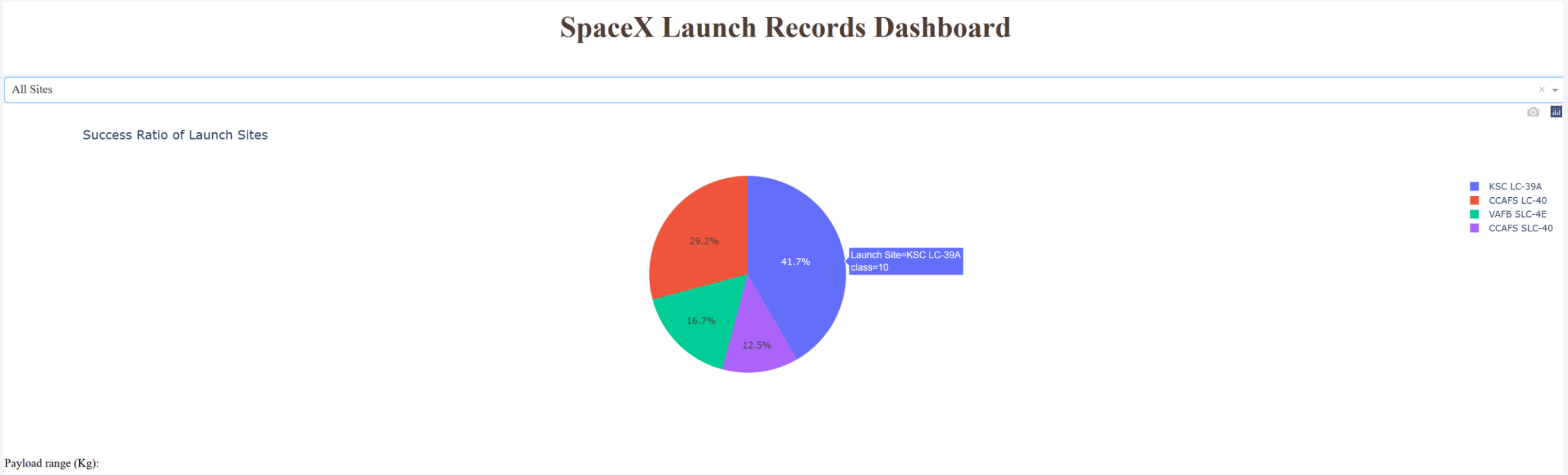


<Folium Map Screenshot 3>



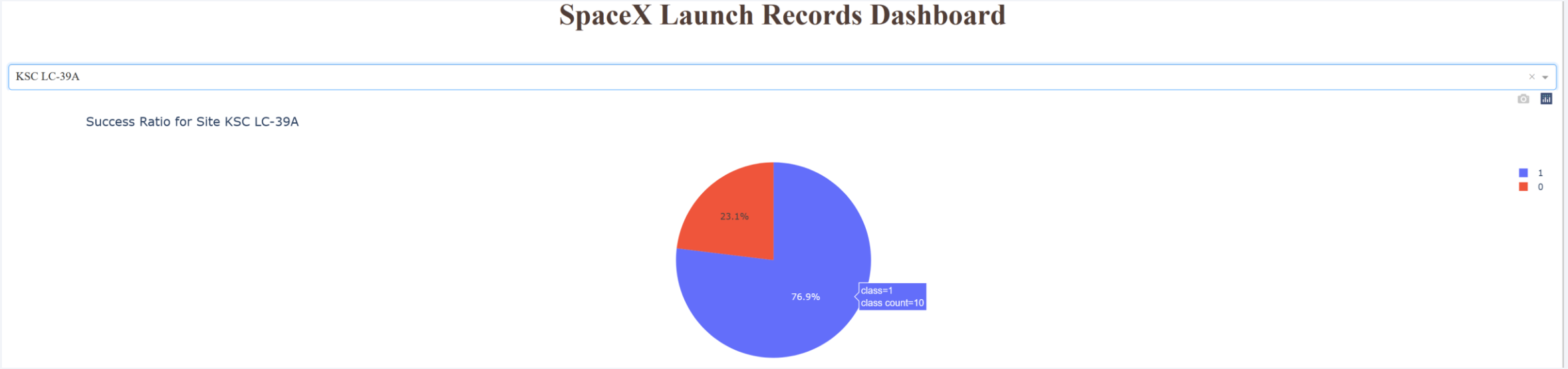

Section 4
Build a Dashboard with Plotly Dash

<Dashboard Screenshot 1>



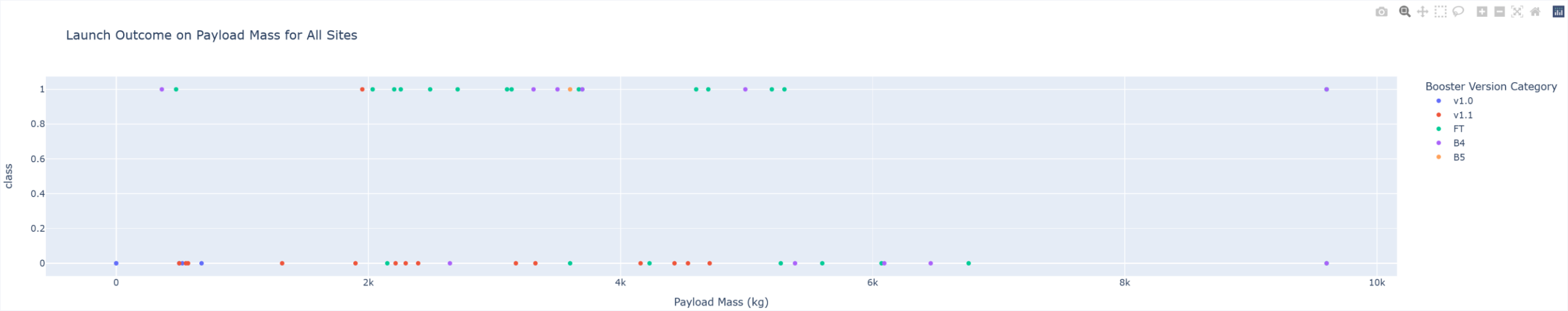
- With 41,7% KSC LC-39A is the launch site with the most successful landing outcomes

<Dashboard Screenshot 2>



- The launch site KSC LC-39A had 10 successful landing outcomes

<Dashboard Screenshot 3>

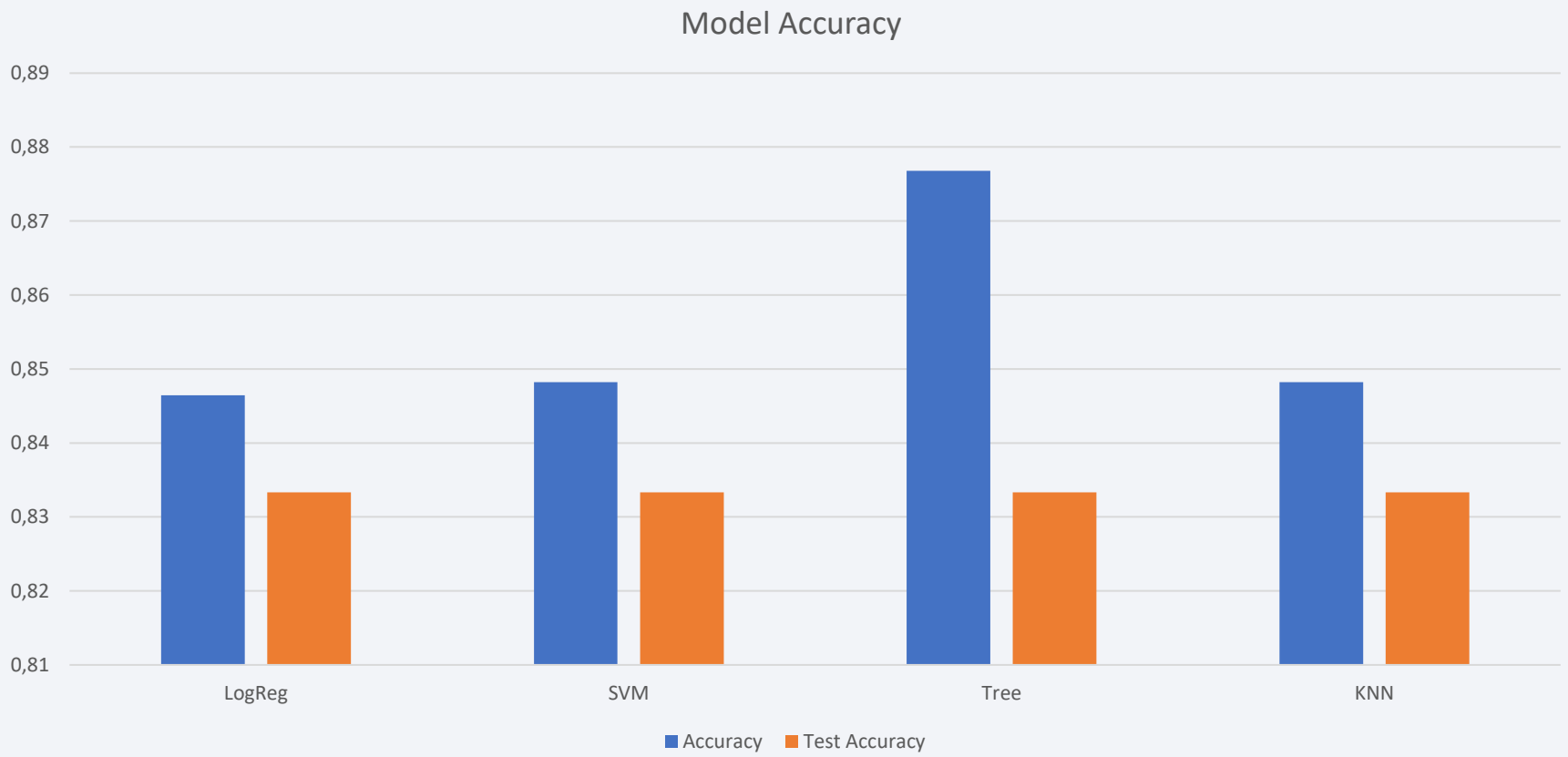


- FT booster version has the highest success rate
- V1.1 booster version has the lowest success rate
- Payload above 6000kg strongly diminishes the success rate

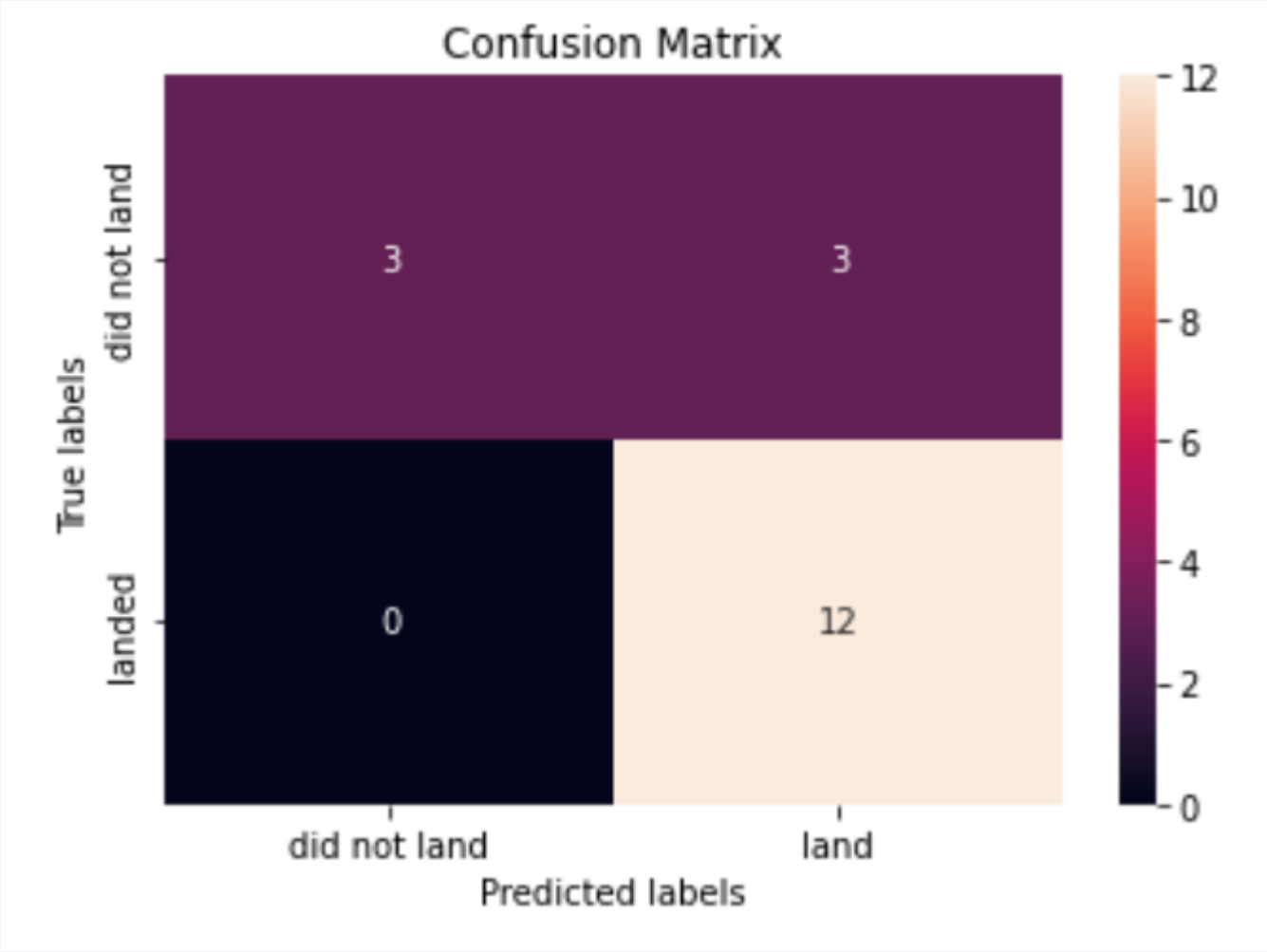
Section 5

Predictive Analysis (Classification)

Classification Accuracy



Confusion Matrix



Conclusions

- The analysis of the data shows the relationships between the mission outcome and features such as payload mass, booster version, launch site
- The most promising launch site is KSC LC 39A
- The number of successful landing outcomes increases over the years, probably to technological and analytical progress
- With the built classification models it is possible to predict the launch outcome with an estimated accuracy of 83,33%.

Thank you!