IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Almaz Sadenov
26 Oct 2021

Outline

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

Executive Summary

- During the capstone real work pipeline has been simulated, this experience going to be helpful in any future data project.
- Pipeline steps included:
  - Data collection (REST API / web scraping)
  - Data preparation and cleaning
  - Data Wrangling using Pandas
  - Data Visualization using plot/seaborn/folium/plotly libraries
  - Machine learning alghorithm performing and choosing the best
- These capstone project tasks lead you to improve your data science skills

Introduction

Project Background

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.

In this capstone, we will predict if the Falcon 9 first stage will land successfully.

Section 1

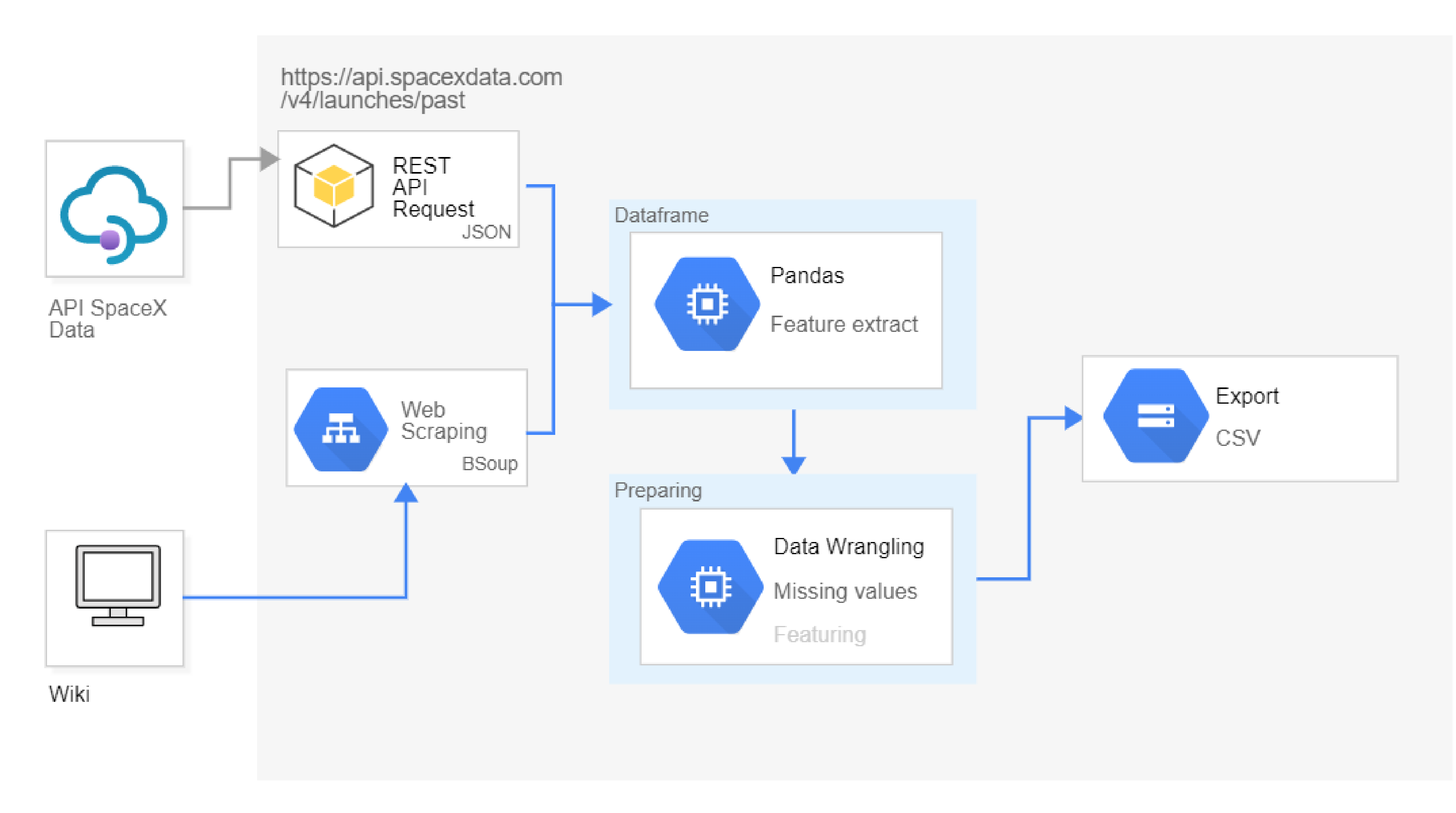
Methodology

Methodology

Executive Summary

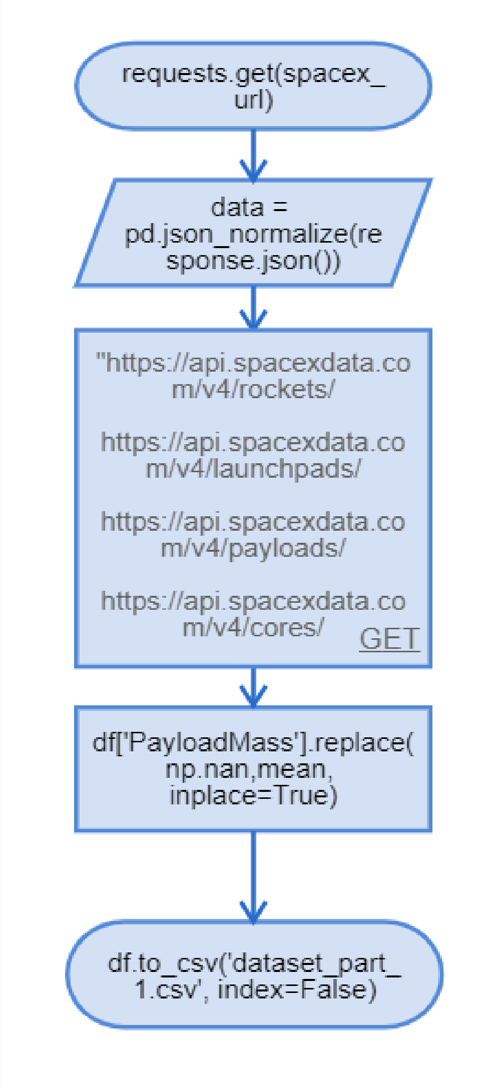
- Data collection methodology:
  - Data was collected and checked that is in the correct format from an API.
  - Also launch data extracted from HTML table from Wikipedia
- Data wrangling
  - Features created, missing values checked, visualized through pandas dataframe
- Exploratory data analysis (EDA) using visualization and SQL
- Interactive visual analytics using Folium and Plotly Dash
- Predictive analysis using classification models

Data Collection



Data Collection – SpaceX API

- Data collection with SpaceX REST calls



Decode the response content as a Json using json() and turn it into a Pandas dataframe using.json_normalize()

We will now use the API again to get information about the launches using the IDs given for each launch. Specifically we will be using columns

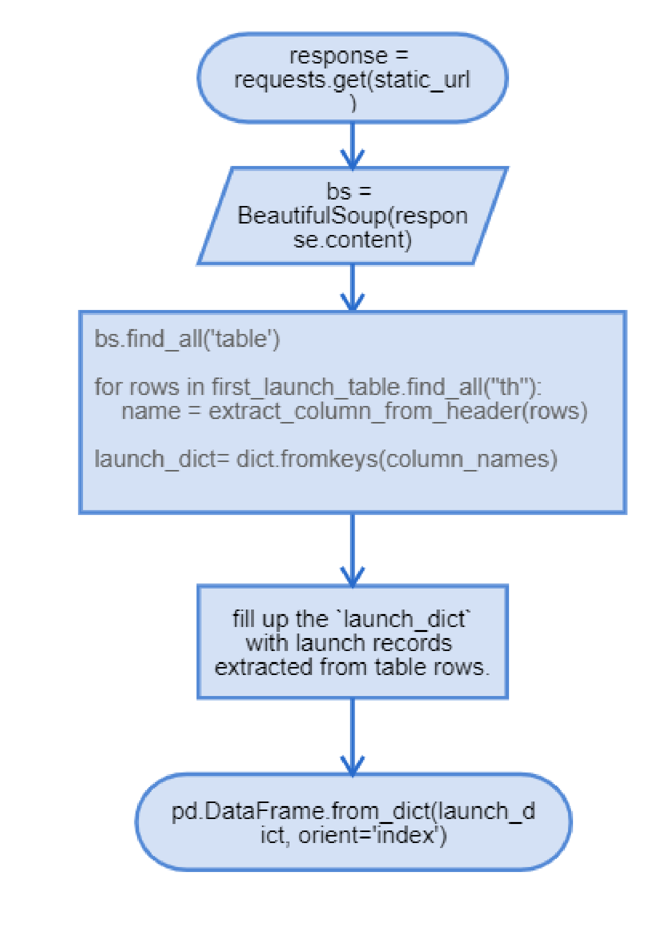
- Rocket
- Payloads,
- Launchpad,
- Cores

The mean and the .replace() function to replace `np.nan` values in the data with the mean you calculated.

- GitHub URL of the completed SpaceX API calls notebook
- https://github.com/alm4z/ibm-ds-capstone/blob/8cdab60ff7a87ed2bcf4b462053d98607dbf94fe/Week%201/Data%20Collection%20API.ipynb

Data Collection – Scraping

- Data collection with Web Scraping



Request the HTML page from the above URL and get a `response` object

Use BeautifulSoup() to create a BeautifulSoup object from a response text content

Iterate each required html element and apply the provided extract_data() to get a value

Create dataframe and export it to a CSV.

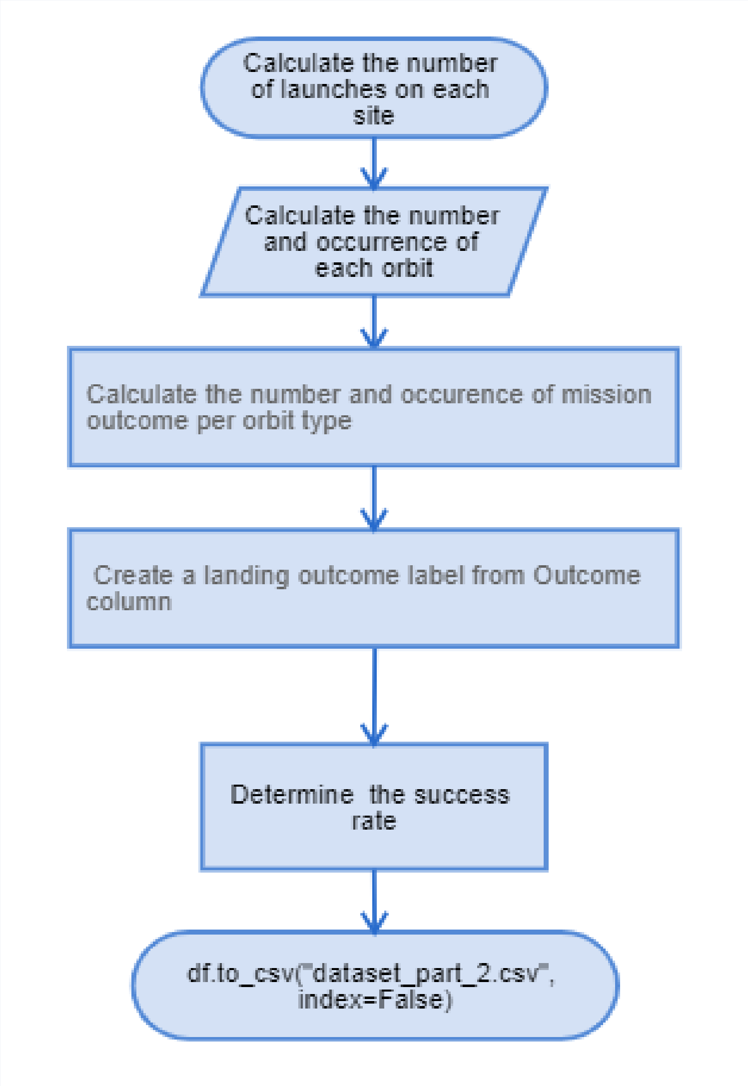
- GitHub URL of the completed Web Scraping notebook
- https://github.com/alm4z/ibm-ds-capstone/blob/8cdab60ff7a87ed2bcf4b462053d98607dbf94fe/Week%201/Data%20Collection%20with%20Web%20Scraping.ipynb

Data Wrangling

- Data was processed using pandas dataframe features

- Data wrangling process contains 5 steps for data understanding

- GitHub URL of completed data wrangling related notebook
- https://github.com/alm4z/ibm-ds-capstone/blob/8cdab60ff7a87ed2bcf4b462053d98607dbf94fe/Week%202/EDA.ipynb



EDA with Data Visualization

- **Scatter Point Charts:**
  - The relationship between Flight Number and Launch Site
  - The relationship between Payload and Launch Site
  - The relationship between Payload and Orbit type
- **Bar Chart**
  - The relationship between success rate of each orbit type
- **Line Chart**
  - The launch success yearly trend¶

https://github.com/alm4z/ibm-ds-capstone/blob/8cdab60ff7a87ed2bcf4b462053d98607dbf94fe/Week%202/EDA%20with%20Data%20Visualization.ipynb

EDA with SQL

- **Performed SQL queries:**
  - The names of the unique launch sites in the space mission
  - 5 records where launch sites begin with the string 'CCA'
  - Average payload mass carried by booster version F9 v1.1
  - List of the first successful landing outcome in ground pad was acheived.
  - List of the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
  - List of the total number of successful and failure mission outcomes
  - List of the names of the booster_versions which have carried the maximum payload mass. Use a subquery
  - List of the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015
  - Ranked 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/alm4z/ibm-ds-capstone/blob/8cdab60ff7a87ed2bcf4b462053d98607dbf94fe/Week%202/EDA%20with%20SQL.ipynb

Build an Interactive Map with Folium

- **Next Folium maps were generated:**
  - All launch sites were marked on a map¶
  - The success/failed launches were marked for each site on the map
  - The distances between a launch site to its proximities¶

- These maps helped to understand distance relationship between launches

https://github.com/alm4z/ibm-ds-capstone/blob/8cdab60ff7a87ed2bcf4b462053d98607dbf94fe/Week%203/Interactive%20Visual%20Analytics%20with%20Folium%20lab.ipynb

Build a Dashboard with Plotly Dash

- Next elements and functions were coded:
  - Dropdown list to enable Launch Site selection
  - Pie chart to show the total successful launches count for all sites
  - Slider to select payload range
  - Scatter chart for success payload
  - Callback function for `site-dropdown` as input, `success-pie-chart` as output
  - Callback function for `site-dropdown` and `payload-slider` as inputs, `success-payload-scatter-chart` as output

- Interactive dash helps to understand relationship between payload and success launches count in different circumstances

https://github.com/alm4z/ibm-ds-capstone/blob/8cdab60ff7a87ed2bcf4b462053d98607dbf94fe/Week%203/spacex_dash_app.py

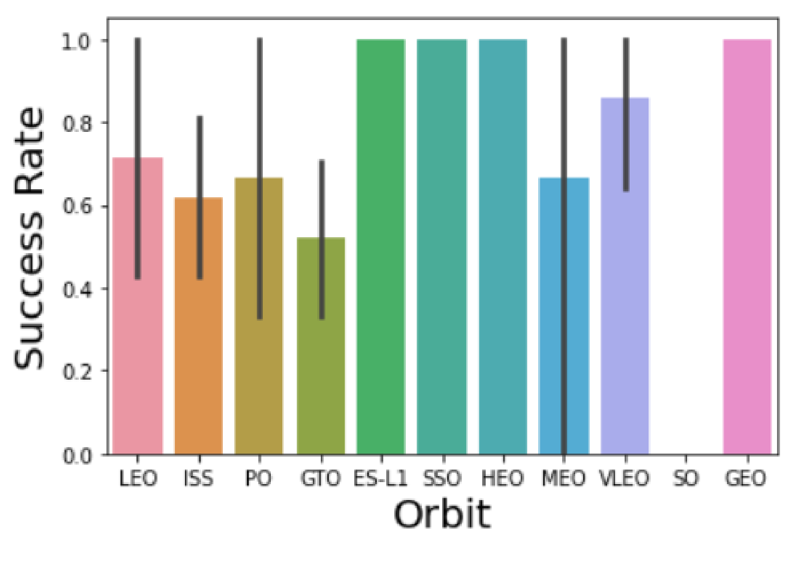
Predictive Analysis (Classification)

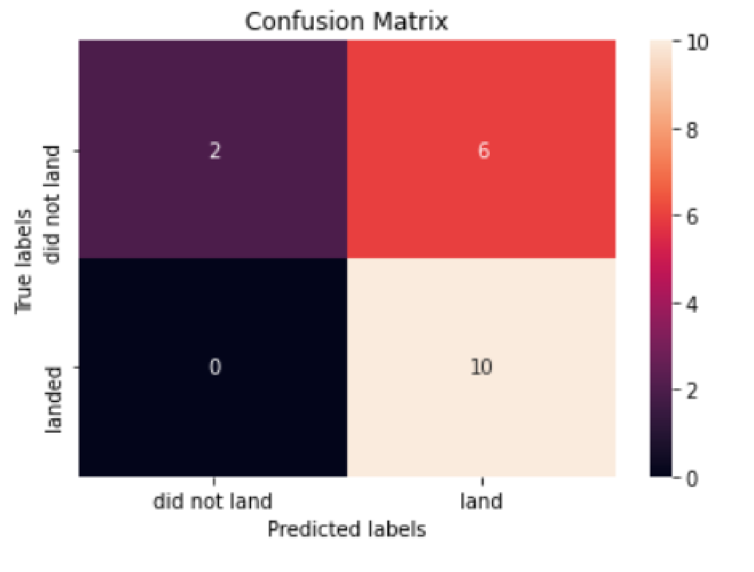
- Performed exploratory Data Analysis and determined Training Labels
- Created a column for the class
- Standardized the data
- Splited into training data and test data
- Finded best Hyperparameter for SVM, Classification Trees and Logistic Regression
- Selected the method performs best using test data

https://github.com/alm4z/ibm-ds-capstone/blob/8cdab60ff7a87ed2bcf4b462053d98607dbf94fe/Week%204/Machine%20Learning%20Prediction.ipynb

Results

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedC
0	1	2010-06-04	Falcon 9	6104.959412	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0
1	2	2012-05-22	Falcon 9	525.000000	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0
2	3	2013-03-01	Falcon 9	677.000000	ISS	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0
3	4	2013-	Falcon 9	500.000000	PO	VAFB SLC	False [illegible]	1	False	False	False	NaN	1.0	0
4	5	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1.0	0

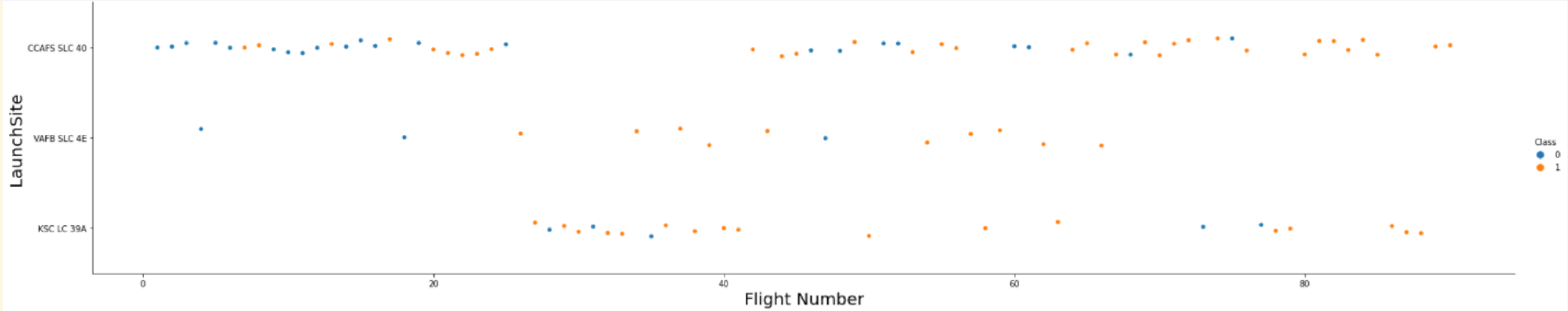




Section 2

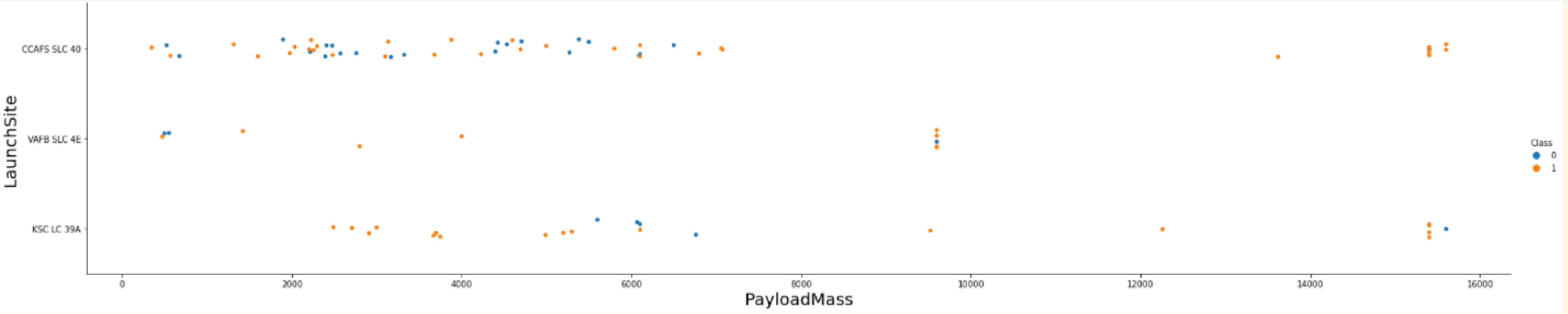
Insights drawn from EDA

Flight Number vs. Launch Site



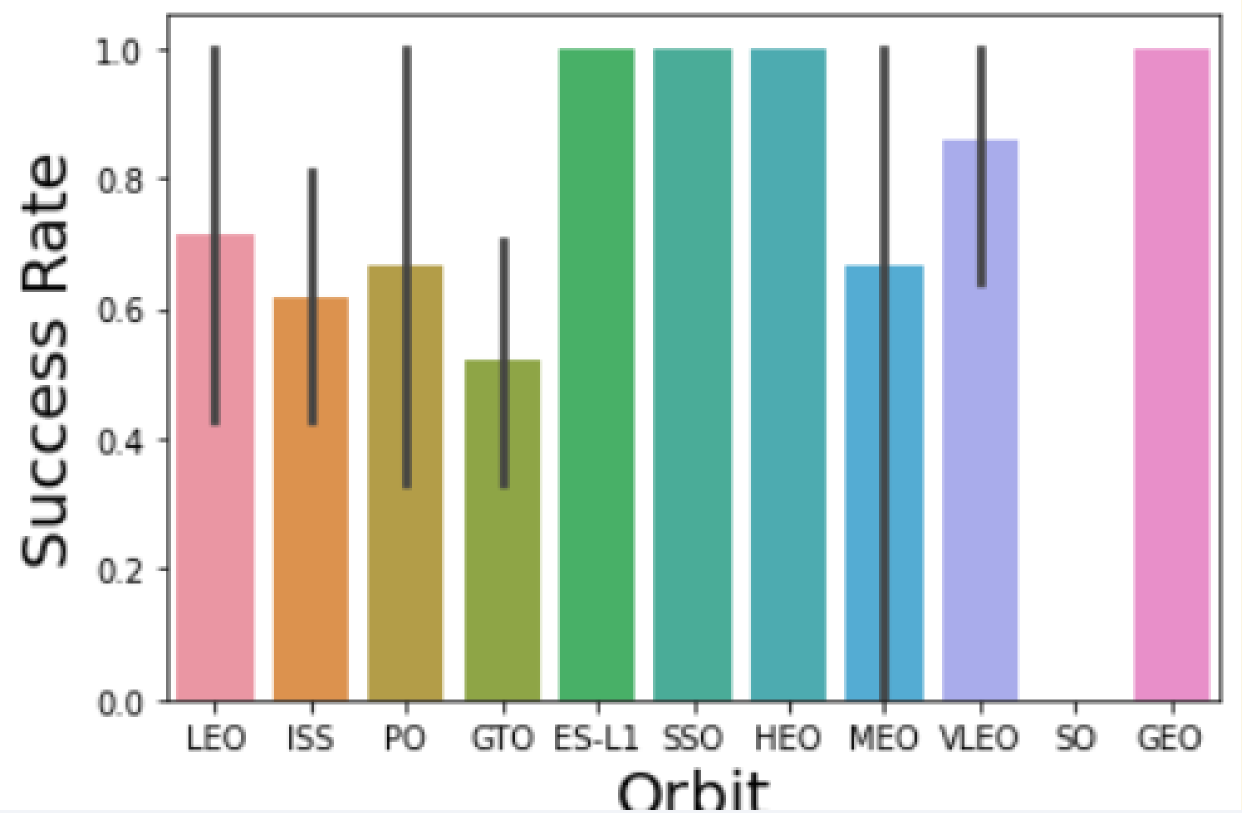
A significant part of launches started from CCAFS SLC 40

Payload vs. Launch Site



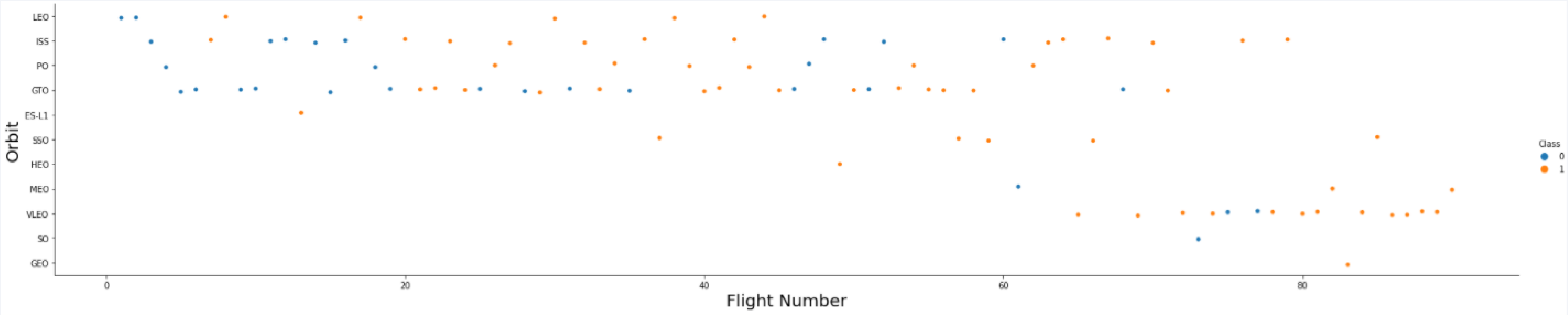
Median payload mass related to CCAFS SLC 40

Success Rate vs. Orbit Type



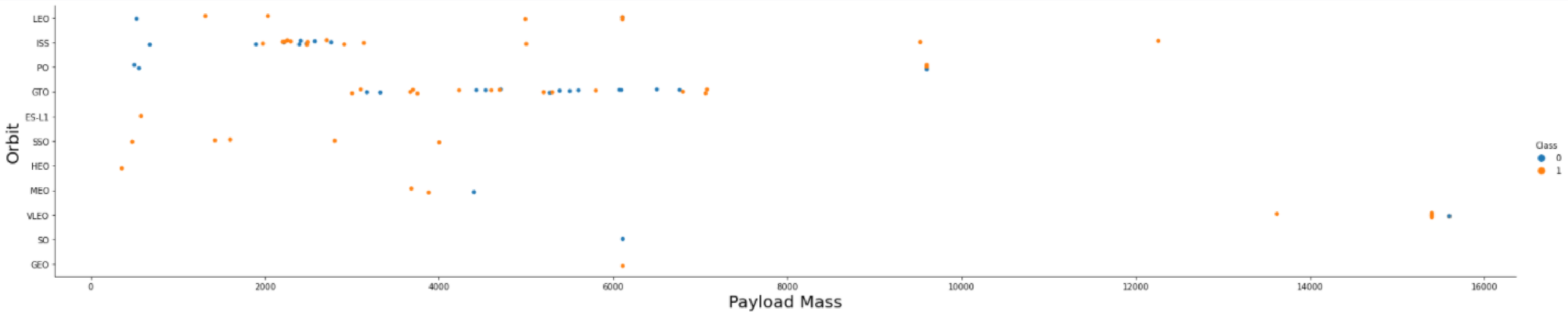
100% success rate observed in ES-L1/SSO/HEP and GEO orbit types

Flight Number vs. Orbit Type



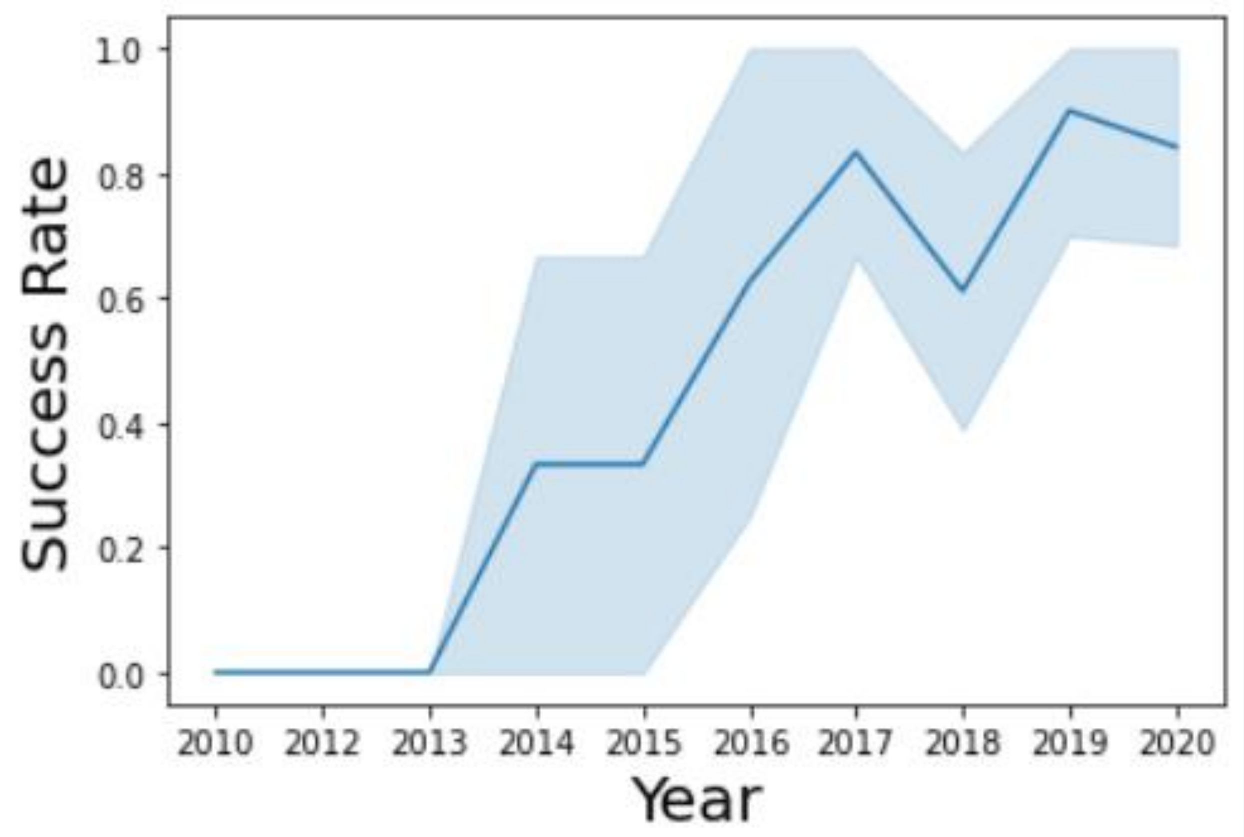
First 60 flights were launched to LEO/ISS/PO/GEO/ES-l1 orbit types, in the end of observed period the most frequent orbits are MEO/VLEO/SO/GEO

Payload vs. Orbit Type



GEO is the most frequently used orbit for any payload mass

Launch Success Yearly Trend



Success rate showed an incredible rise during the all period of time

All Launch Site Names

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

SQL query, launch_site selected using grouping :

```
select LAUNCH_SITE from SPACEXTBL group by LAUNCH_SITE
```

Launch Site Names Begin with 'CCA'

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

SQL query

```
select * from SPACEXTBL
where LAUNCH_SITE LIKE
'CCA%' LIMIT 5
```

Total Payload Mass

45 596

SQL query:

```
SELECT sum(payload_mass__kg_) from spacextbl
where customer='NASA (CRS)'
```

Average Payload Mass by F9 v1.1

2 928

SQL query:

```
SELECT AVG(payload_mass__kg_) from spacextbl where
booster_version='F9 v1.1'
```

First Successful Ground Landing Date

2015-12-22

SQL query:

```
select min(DATE) from spacextbl where
landing__outcome = 'Success (ground pad)'
```

Successful Drone Ship Landing with Payload between 4000 and 6000

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

SQL query:

```
select booster_version from spacextbl where
landing__outcome = 'Success (drone ship)' and
payload_mass__kg_ between 4000 and 6000
```

Total Number of Successful and Failure Mission Outcomes

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

SQL query:

```
select mission_outcome, count(*) from spacextbl
group by mission_outcome
```

Boosters Carried Maximum Payload

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

SQL query:

```
select
booster_version,payload_mass__
kg_ from spacextbl where
payload_mass__kg_ in (select
max(payload_mass__kg_) from
spacextbl)
```

2015 Launch Records

booster_version	launch_site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

SQL query:

```
select booster_version,launch_site from spacextbl where
landing__outcome='Failure (drone ship)' and year(date)=2015
```

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

Failure (drone ship)	5
Success (ground pad)	3

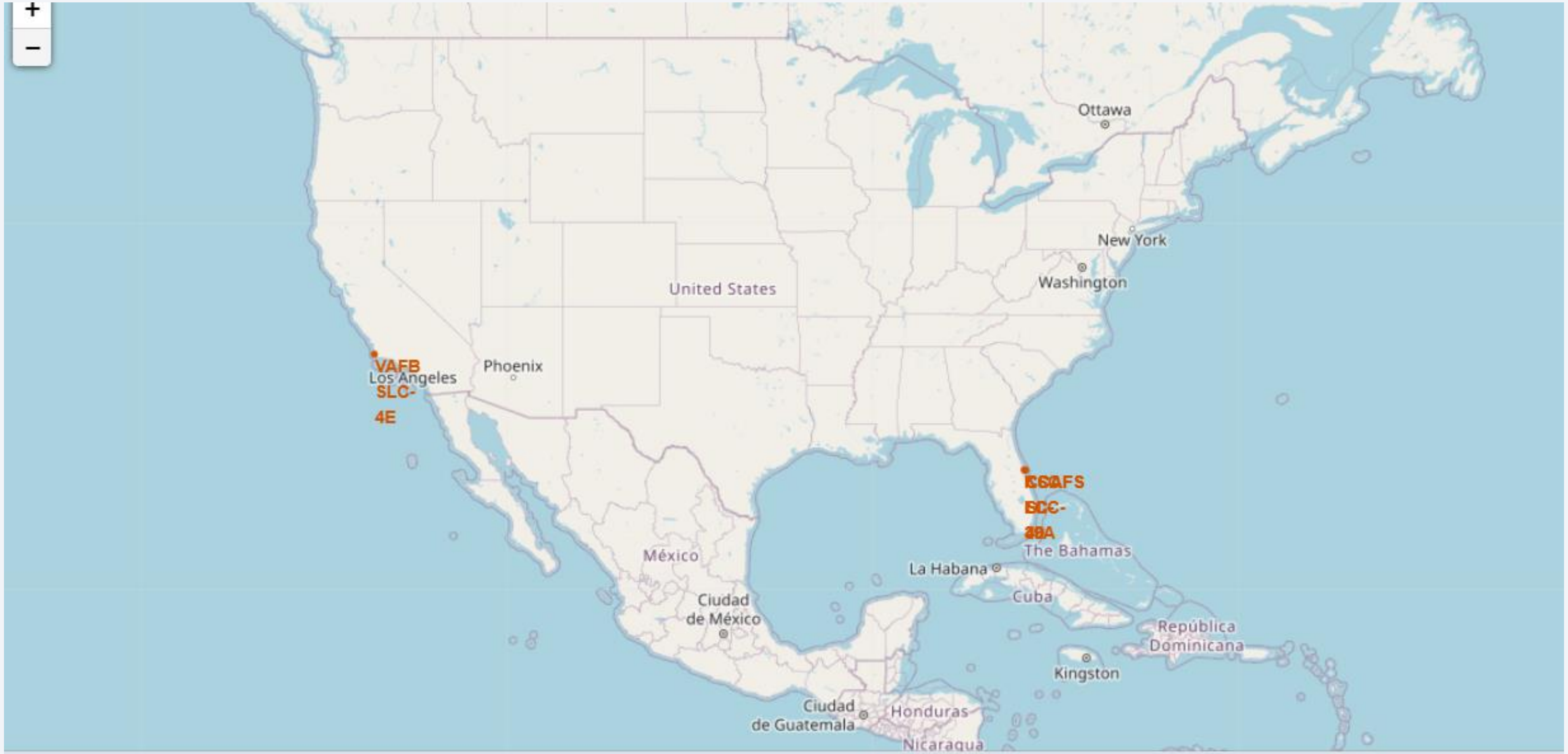
SQL query:

```
select landing__outcome,count(*) from spacextbl where landing__outcome
in ('Failure (drone ship)','Success (ground pad)') and date between
'2010-06-04' and '2017-03-20' group by landing__outcome
```

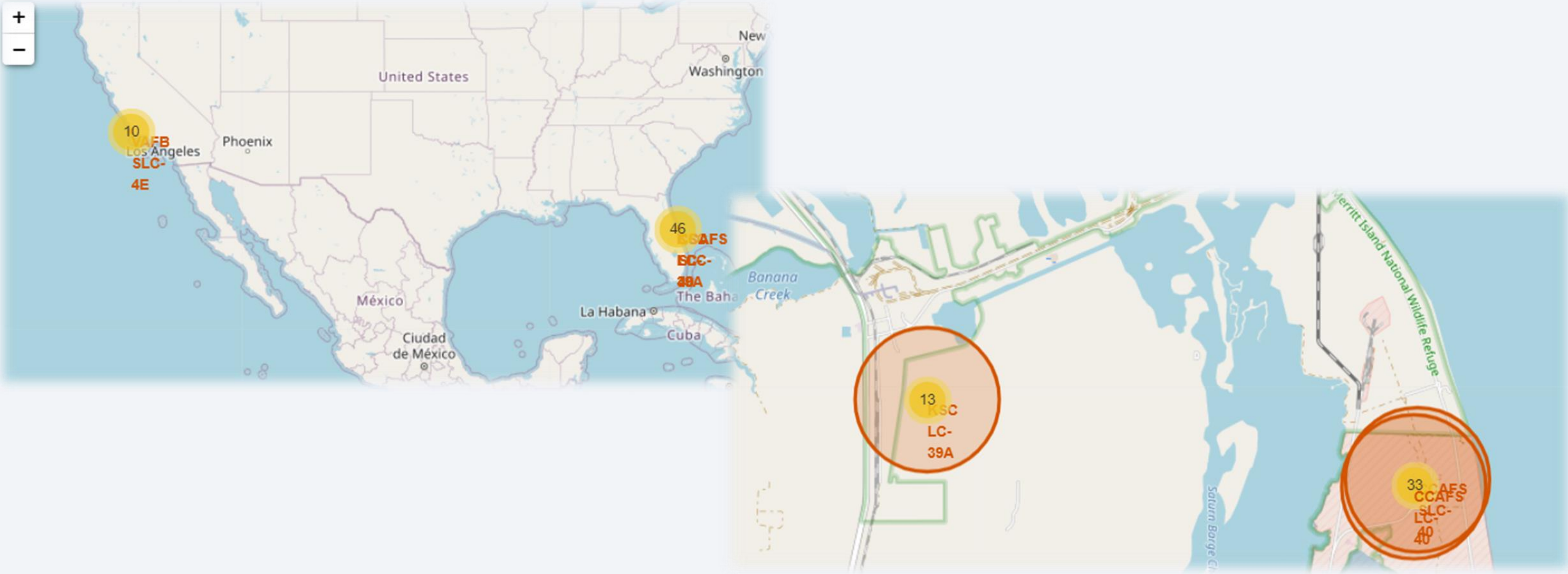
Section 4

Launch Sites Proximities Analysis

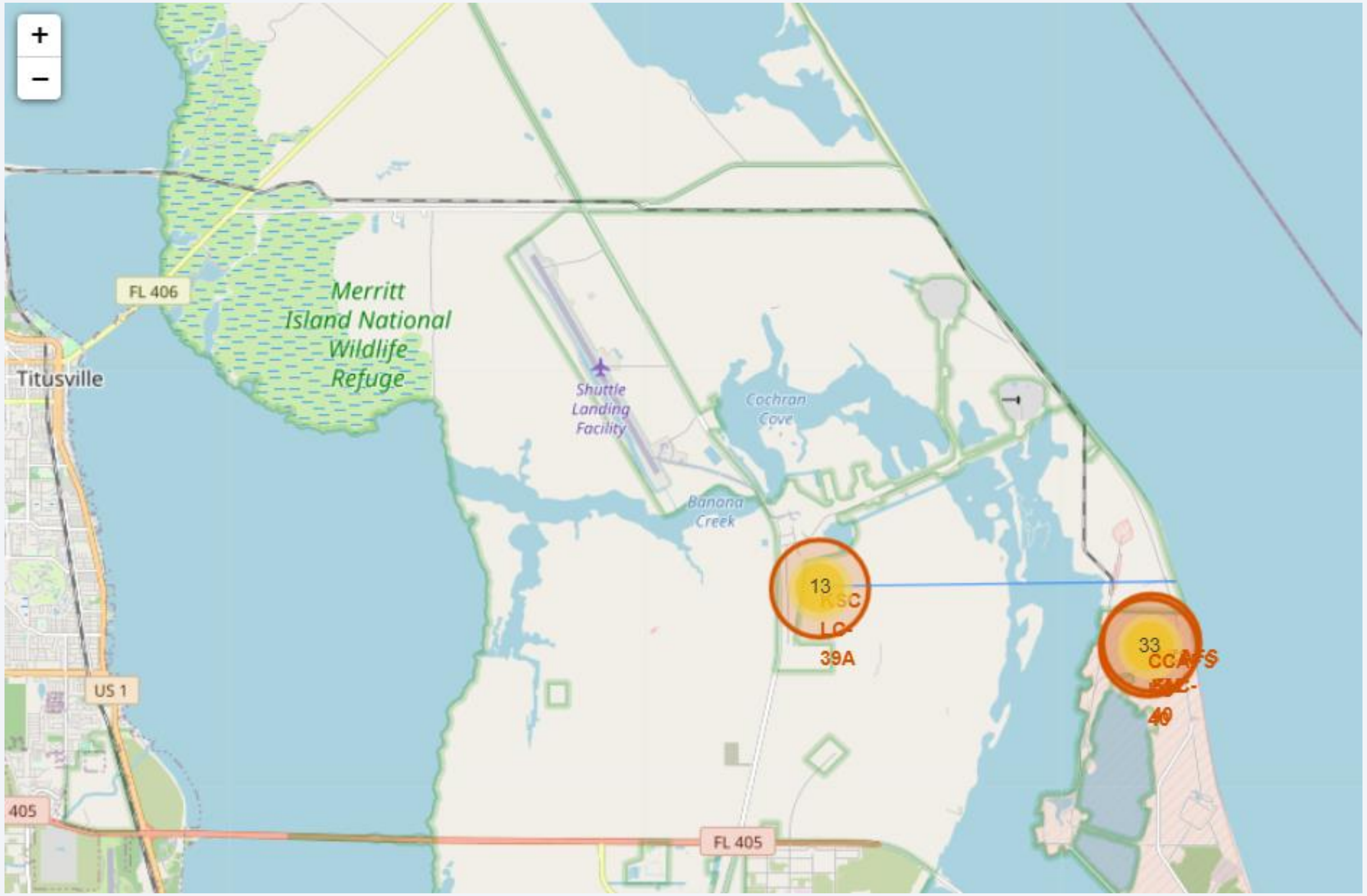
<Folium Map Screenshot 1>



<Folium Map Screenshot 2>



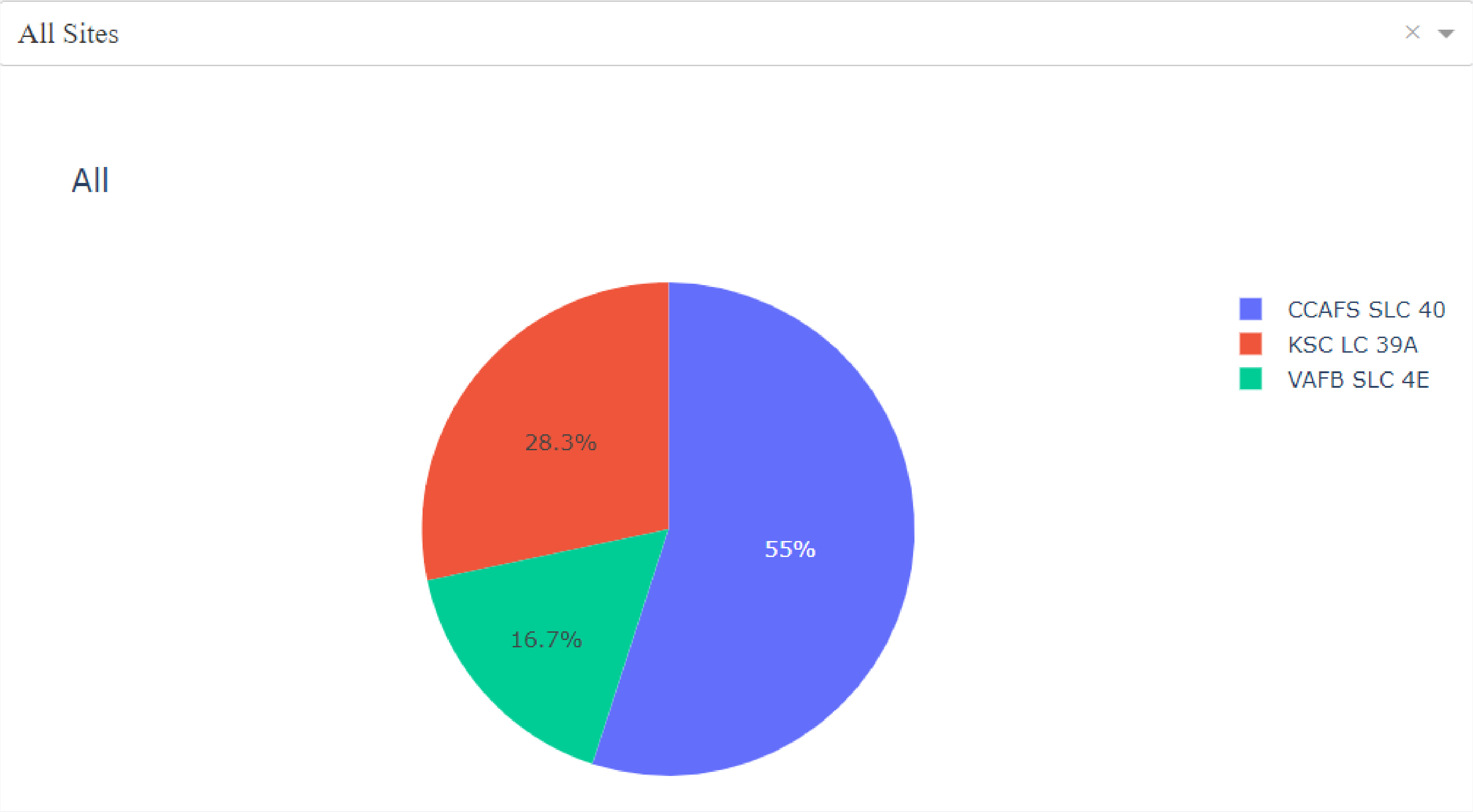
<Folium Map Screenshot 3>



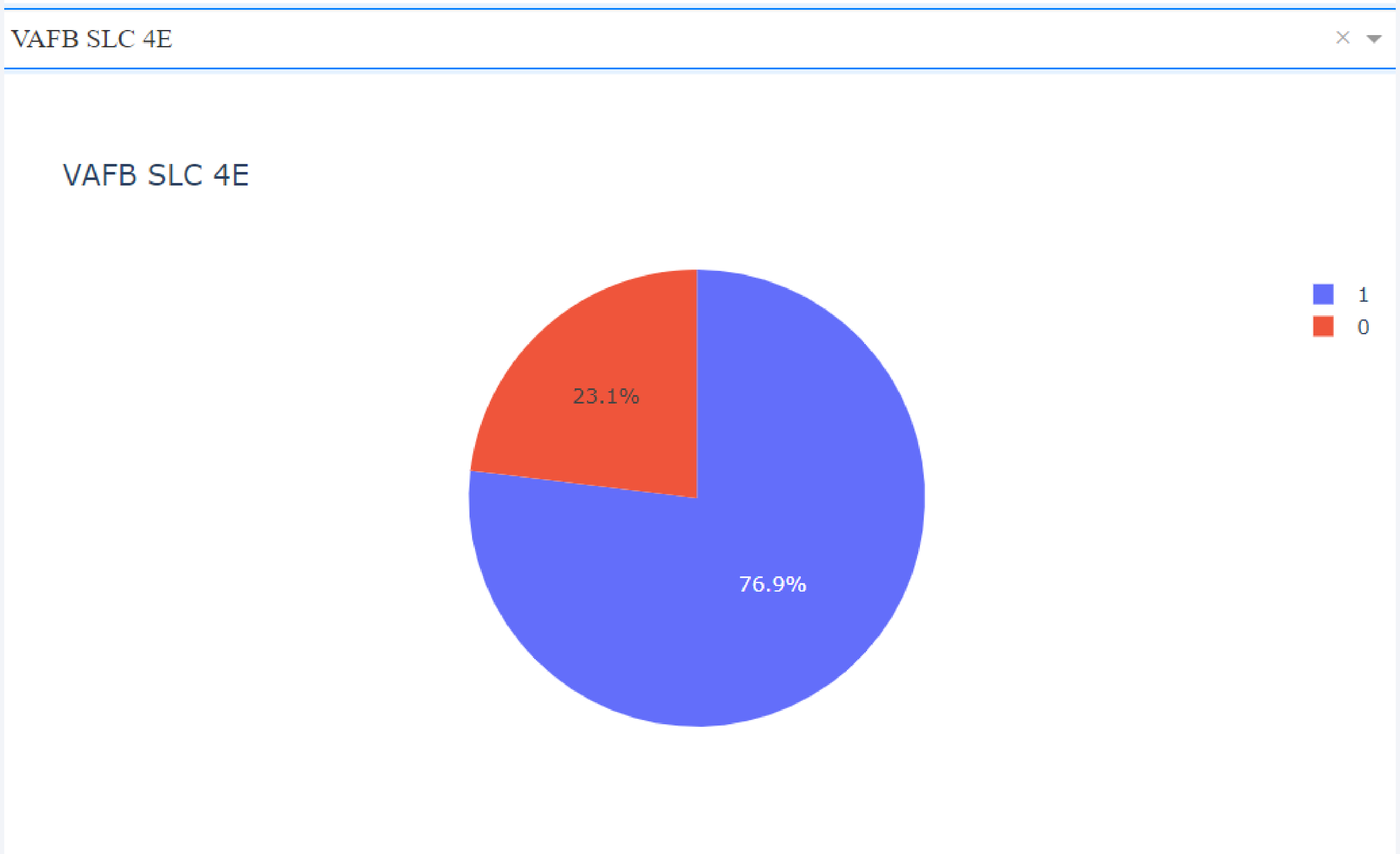
Section 5

Build a Dashboard with Plotly Dash

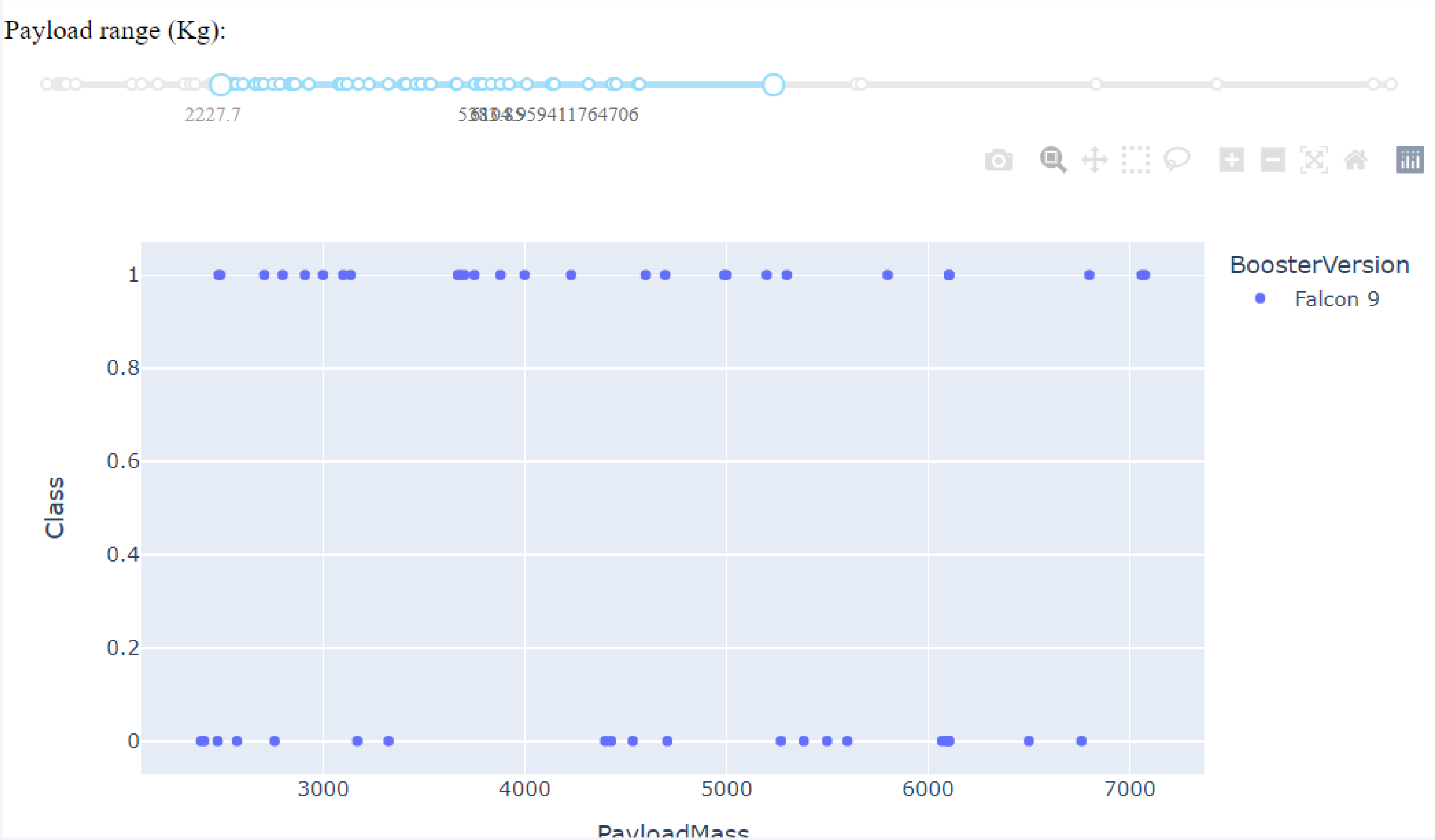
<Dashboard Screenshot 1>



<Dashboard Screenshot 2>



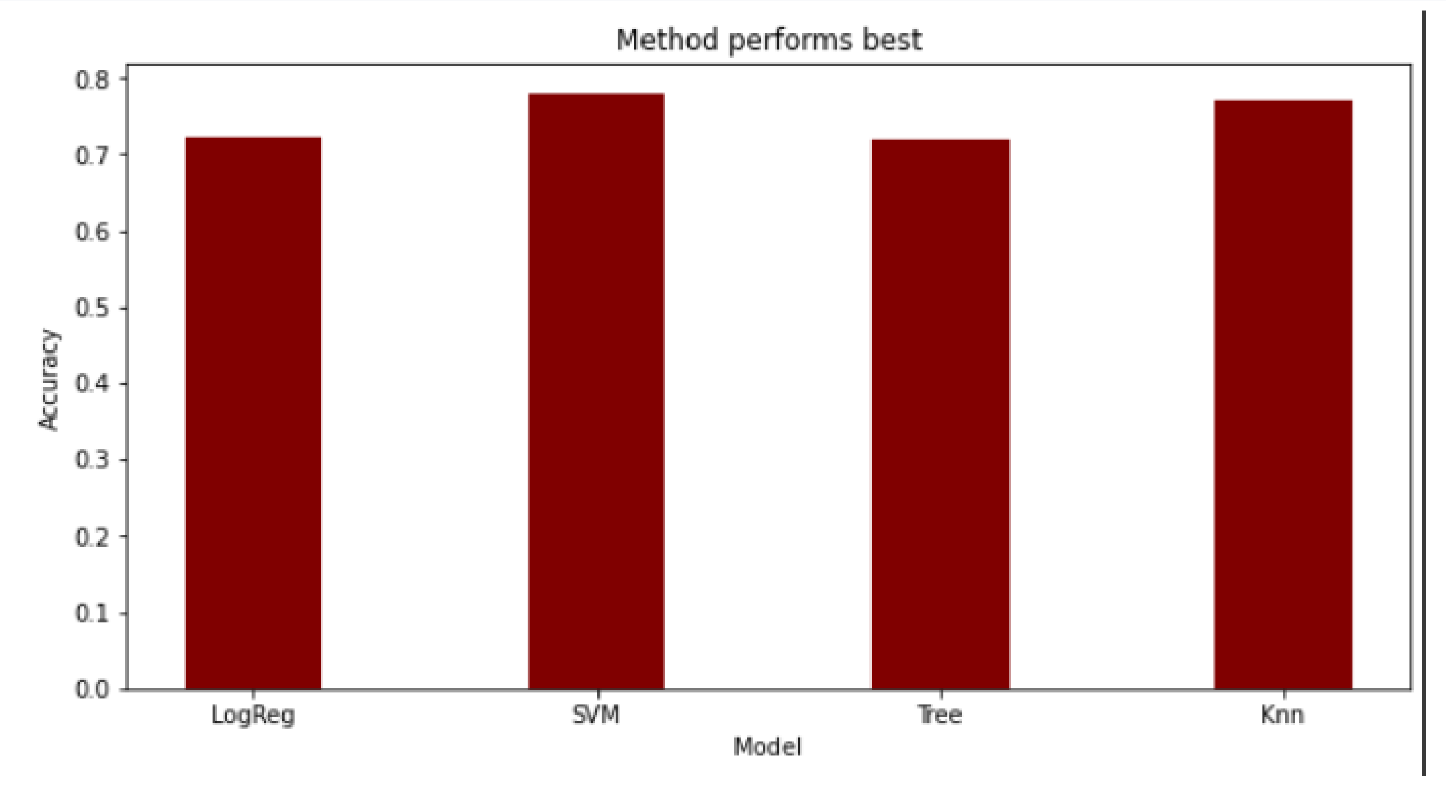
<Dashboard Screenshot 3>



Section 6

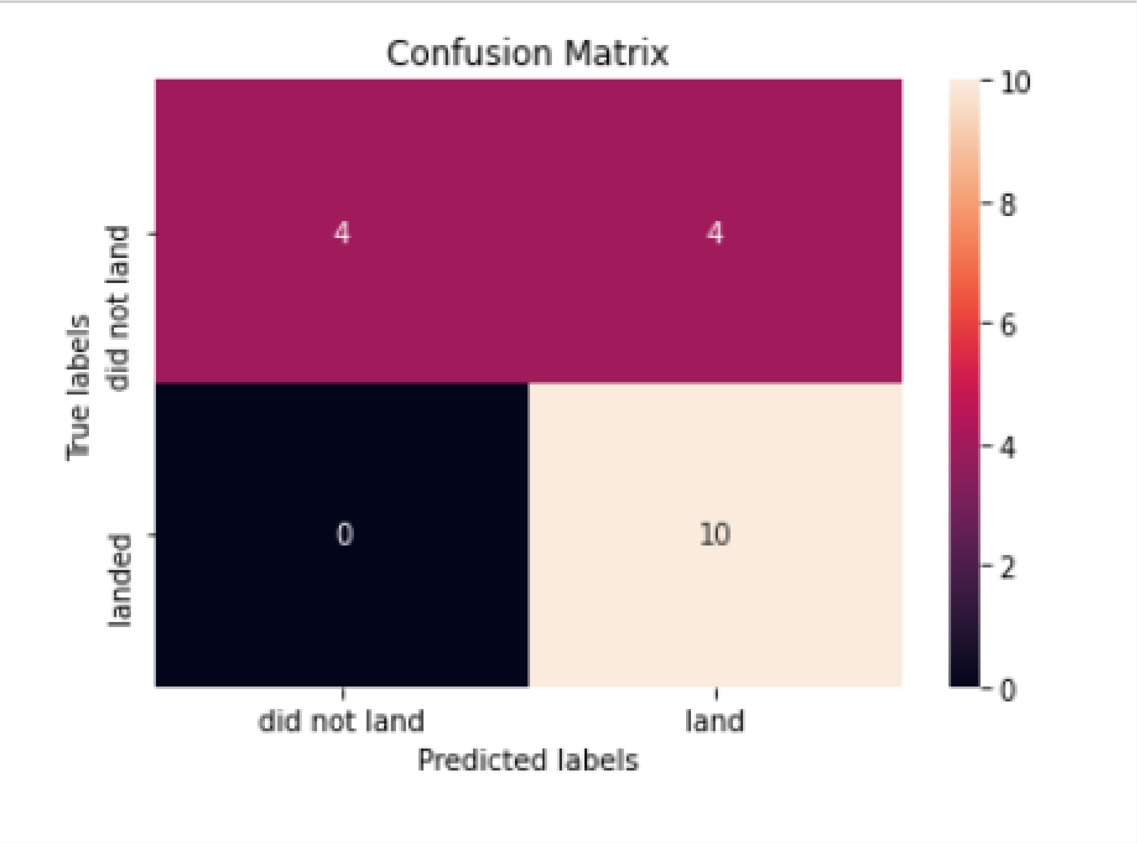
Predictive Analysis (Classification)

Classification Accuracy



Best accuracy on test data - SVM

Confusion Matrix



Conclusions

Space X advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars;

Therefore now 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 space X for a rocket launch.

In this lab, we have been created a machine learning pipeline to predict if the first stage will land given the data from the preceding labs.

Appendix

How to autofill dropdown with categories

```
dcc.Dropdown(id='site-dropdown',
             options=[{'label': 'All Sites', 'value': 'ALL'}] +
                     [ {'label': i, 'value': i} for i in
                       spacex_df['LaunchSite'].unique().tolist() ],
             value='ALL',
             placeholder="Select a Launch Site here",
             searchable=True)
```

Thank you!