IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Name>
<Date>

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection through API
  - Data Collection with Web Scraping
  - Data Wrangling
  - Exploratory Data Analysis with SQL
  - Exploratory Data Analysis with Data Visualization
  - Interactive Visual Analytics with Folium
  - Machine Learning Prediction
- Summary of all results
  - Exploratory Data Analysis result
  - Interactive analytics in screenshots
  - Predictive Analytics result

Introduction

- **Project name:** SpaceX Analysis
- **Background and context:**

  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.

- **Problems to be answered:**
  - What factors are influencing the outcome?
  - What are the relationships between the dimensions?
  - How can we predict if the first stage will land?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data is collected via SpaceX API and web scraping from an HTML page
- Perform data wrangling
  - One-hot encoding for categorical features was used
- 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

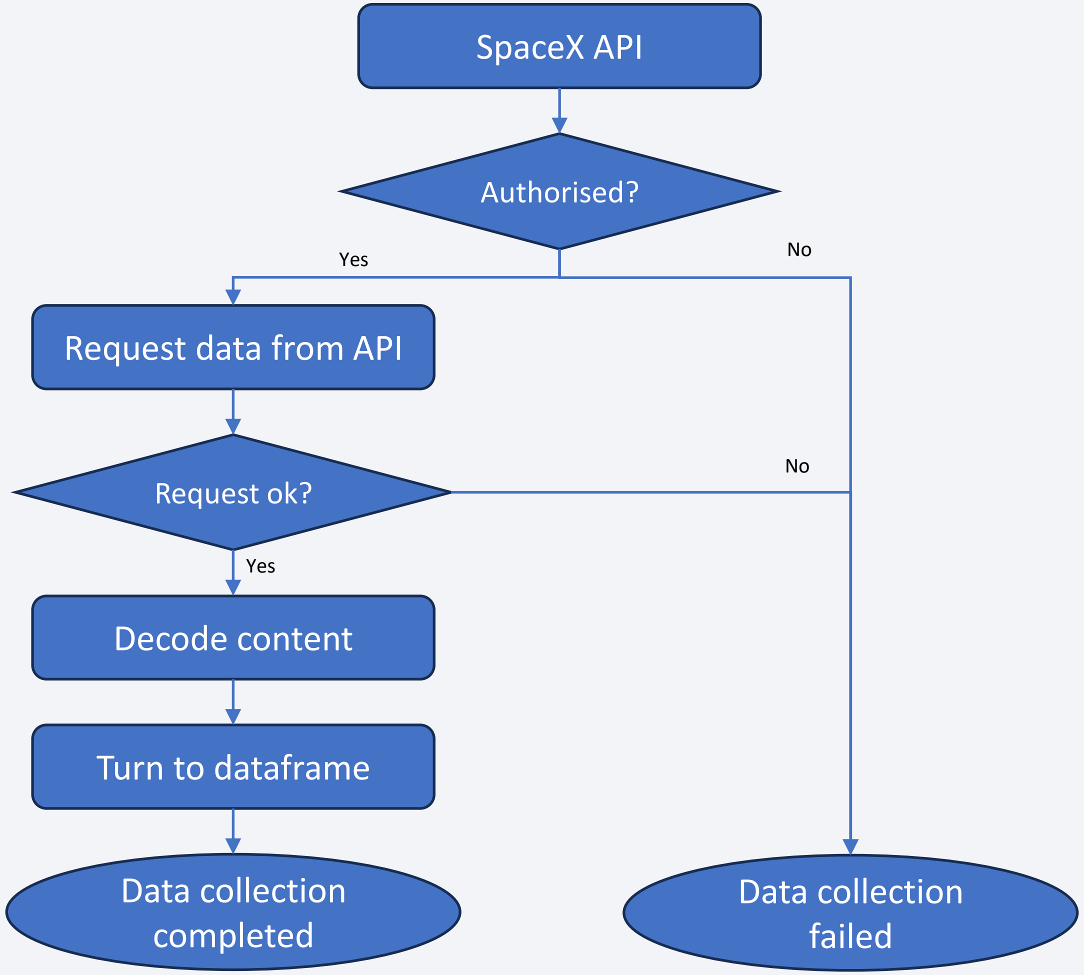
Data Collection

- Data Source: SpaceX API
- Steps:
  - Import needed libraries like numpy and pandas
  - Request rocket launch data from SpaceX API
  - decode the response content as a Json
  - Clean data
  - Turn it into a dataframe

Data Collection – SpaceX API

- spacex_url=https://api.spacexdata.com/v4/launches/past
- response = requests.get(spacex_url)
- df_json = response.json()
- data = pd.json_normalize(df_json)

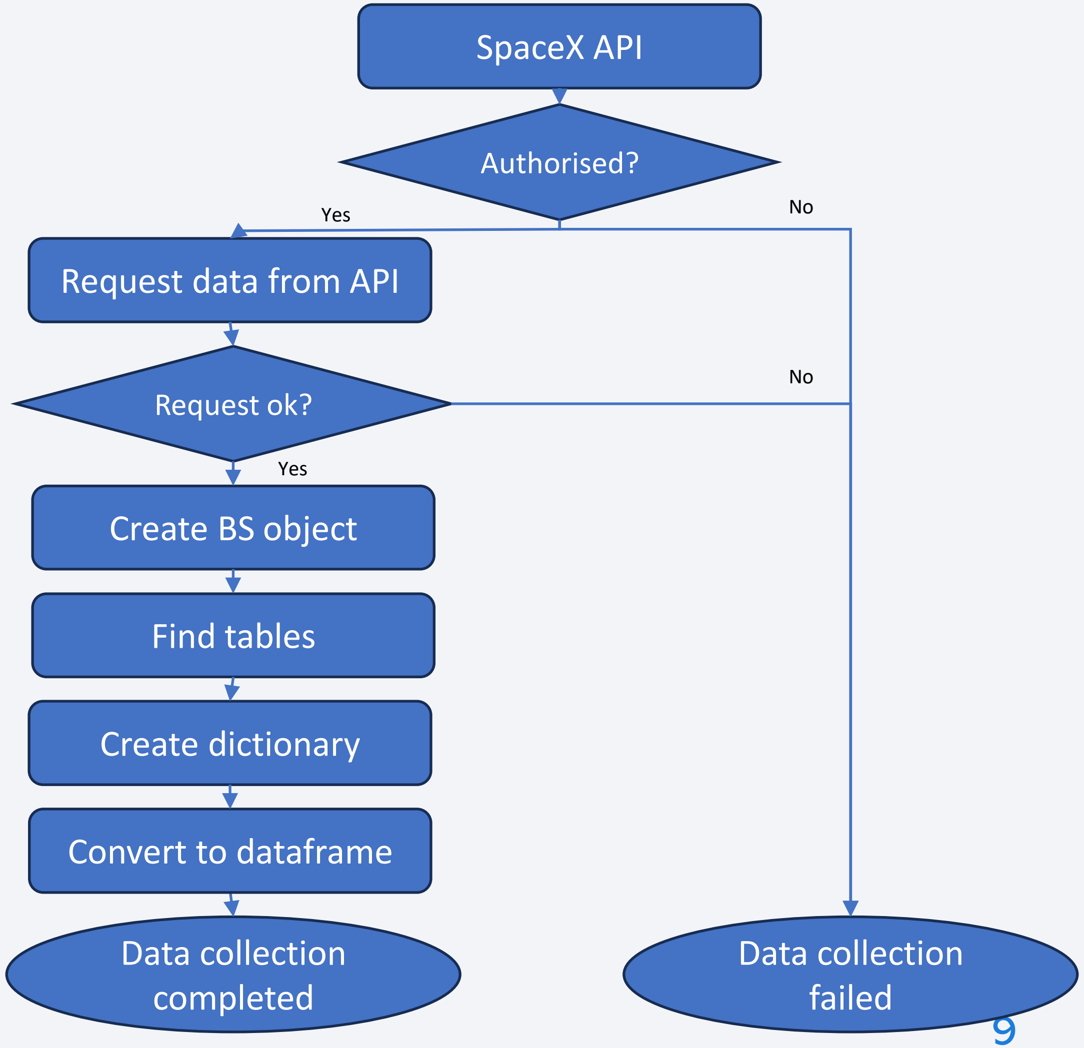
GitHub URL: capstone/jupyter-labs-spacex-data-collection-api.ipynb



Data Collection - Scraping

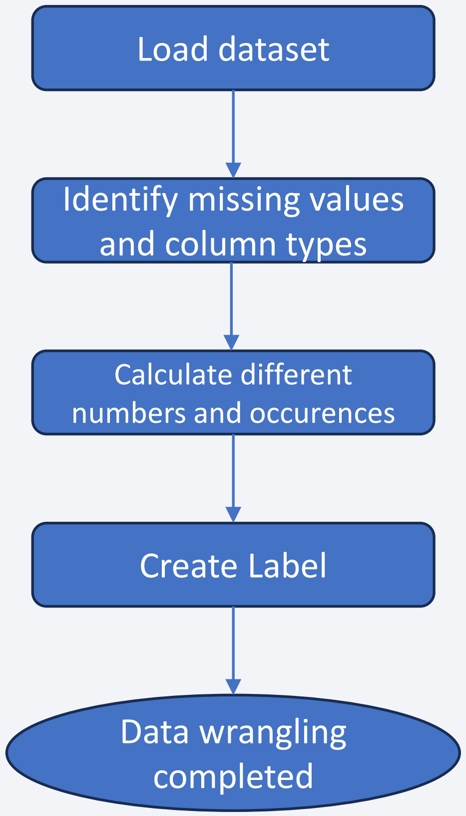
- Request data as HTTP response
- Create a BeautifulSoup object
- find all tables on the wiki page
- create an empty dictionary and convert it to dataframe

GitHub URL: capstone/jupyter-labs-webscraping.ipynb



Data Wrangling

- Load dataset
- Identify missing values in each attribute
- Identify column data types
- Calculate number of launches on each site
- Calculate the number and occurrence of each orbit
- Calculate the number and occurence of mission outcome of the orbits
- Create a landing outcome label from Outcome column



GitHub URL: capstone/labs-jupyter-spacex-Data wrangling.ipynb

10

EDA with Data Visualization

- Scatterplot with Class, Flight Number, Orbit, PayloadMass and Launch Site to identify if correlations are existent
- BarCharts to compare different dimensions
- Linechart to identify trends

GitHub URL: capstone/jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb

EDA with SQL

- Find average, highest or lowest values for e.g. first successful landing or highest payloadmass or average success
- Group data e.g. to count number of successful landings

GitHub URL: capstone/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- circles for launch sites to identify location
- Markers for launch site names and success information to show where successful locations are and how many launches are made where
- Lines for distances to see how far the launch site is away from the cost

GitHub URL:
capstone/lab_jupyter_launch_site_location.jupyterlite.ipynb

Build a Dashboard with Plotly Dash

- Pie chart to analyze which launch site was most successful based on the number of successes (dropdown possible to select a specific site)
- Scatterplot to identify correlations between payload and success and to find the most successful Booster version or payload

GitHub URL: capstone/spacex_dash_app.py

Predictive Analysis (Classification)

- Load and transform data
- Split into test and train data set
- Building different models like decision tree, knn, support vector machine
- Calculate accuracy
- Find best performing model

GitHub URL:
capstone/SpaceX_Machine_Learning_Prediction.ipynb

Results

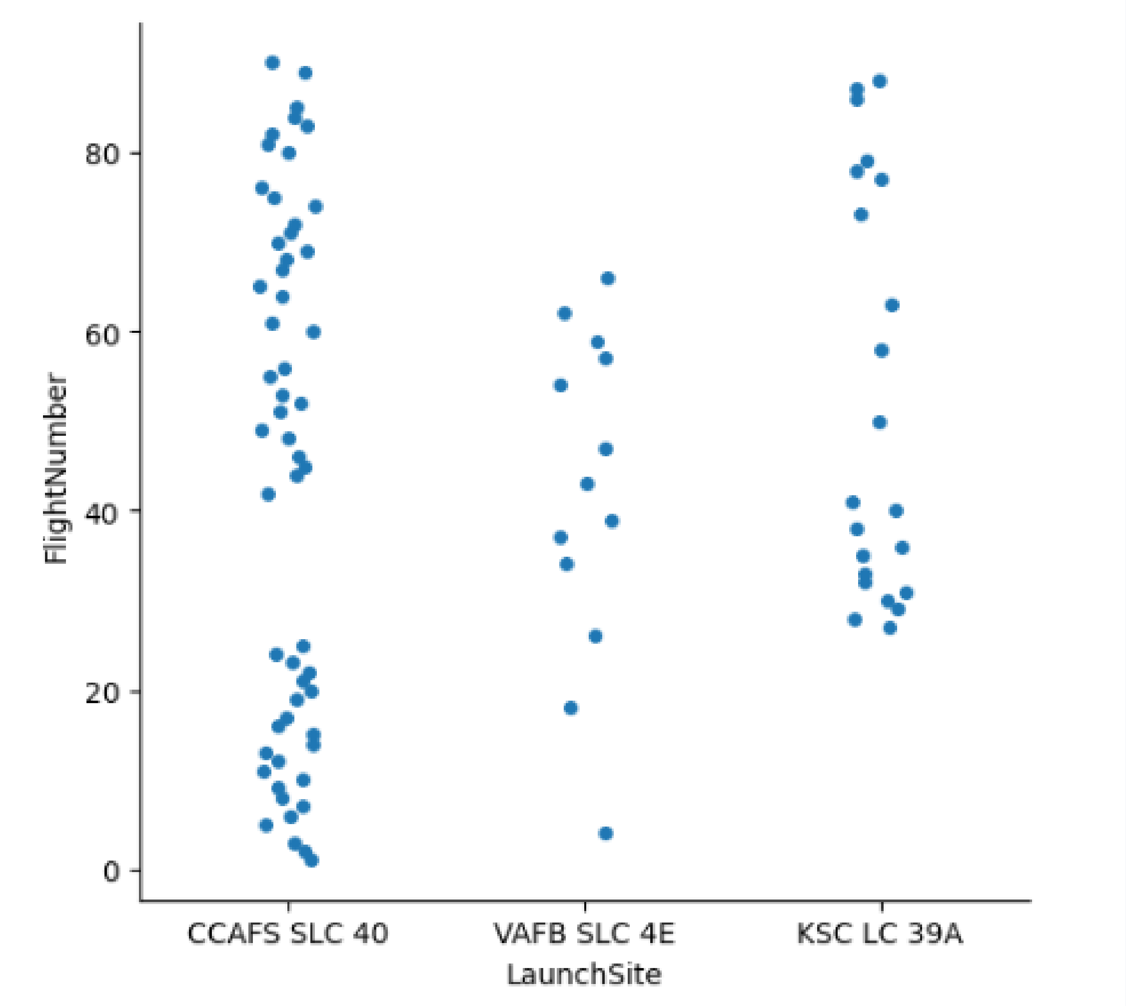
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

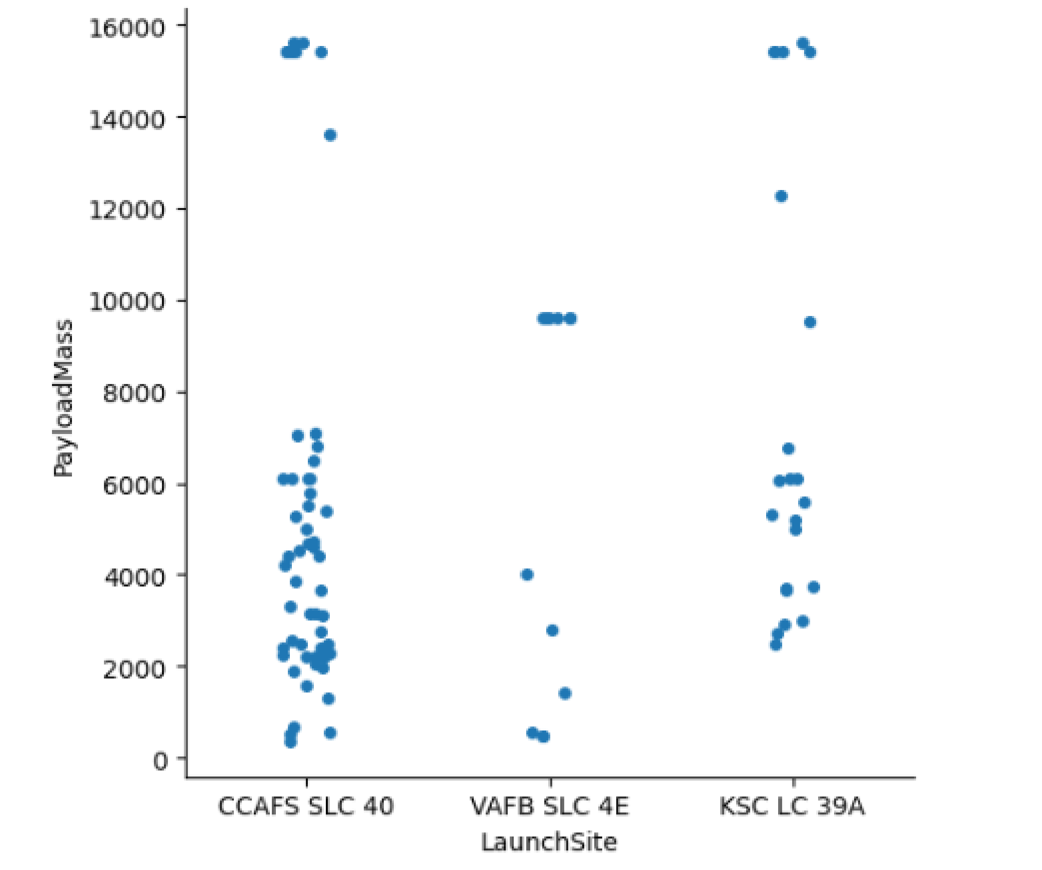
Flight Number vs. Launch Site

- Most of the flights were launched at the site CCAF
- The least were launched at VAFB
- KSC has not launched any of the flights before Flight Number 20



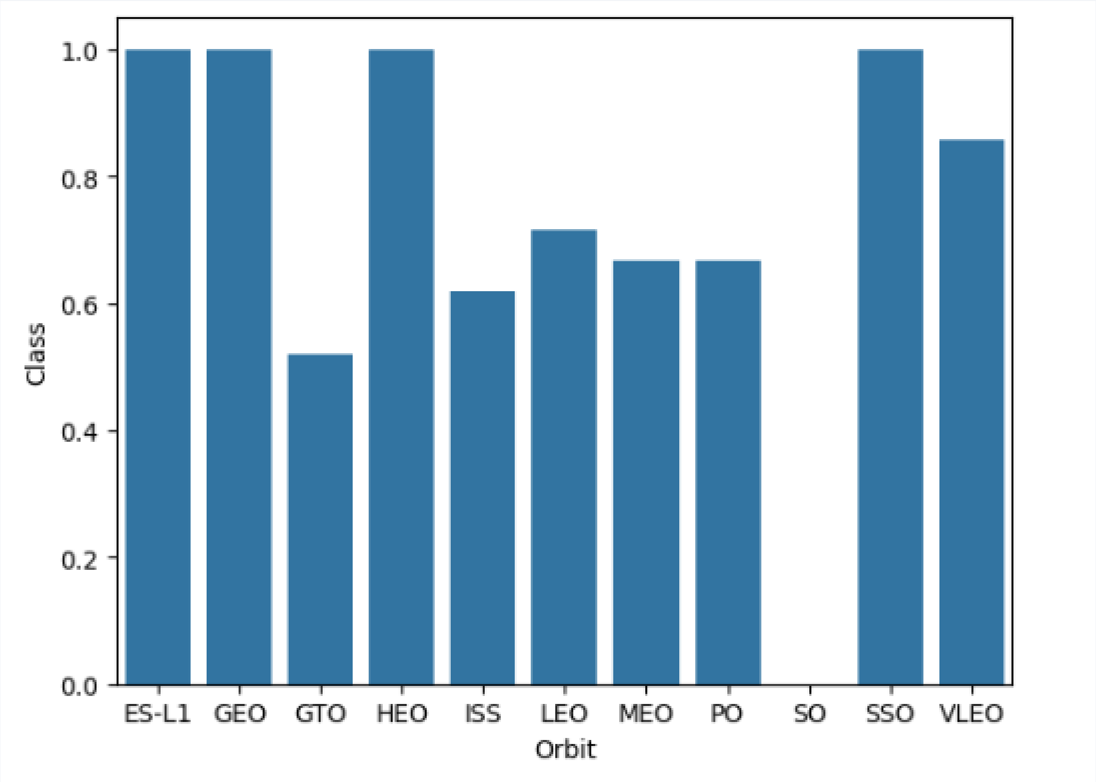
Payload vs. Launch Site

- Most of the flights at launch site CCAFS had a payloadmass below 8000
- Only CCAFS and KSC had payloadmasses above 10000
- KSC had no payloadmasses below 2000



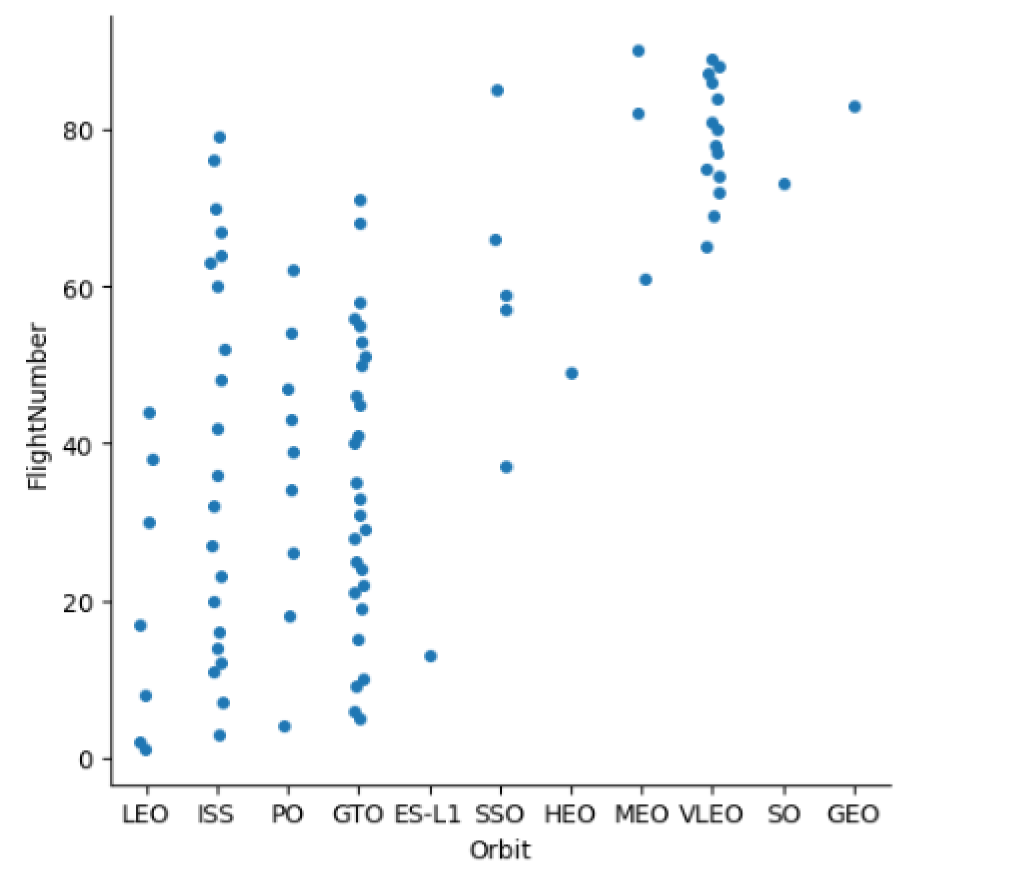
Success Rate vs. Orbit Type

- SO was never succesful
- The most succesful orbits are: ES-L1, GEO, HEO and SSO
- 6 of 11 orbits have an average succes rate below 80%



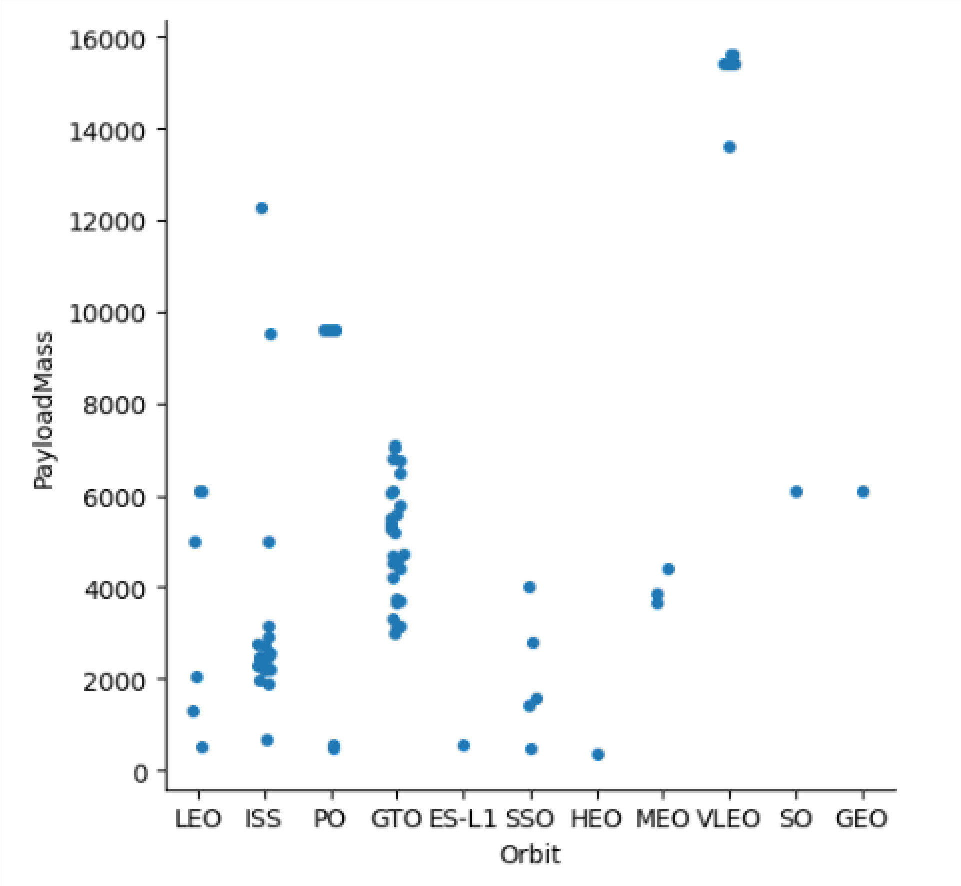
Flight Number vs. Orbit Type

- GEO, SO, HEO and ES-L1 only consist of 1 flight
- 4 orbits only have flight numbers higher than 60
- Orbits GTO and ISS have the most flights



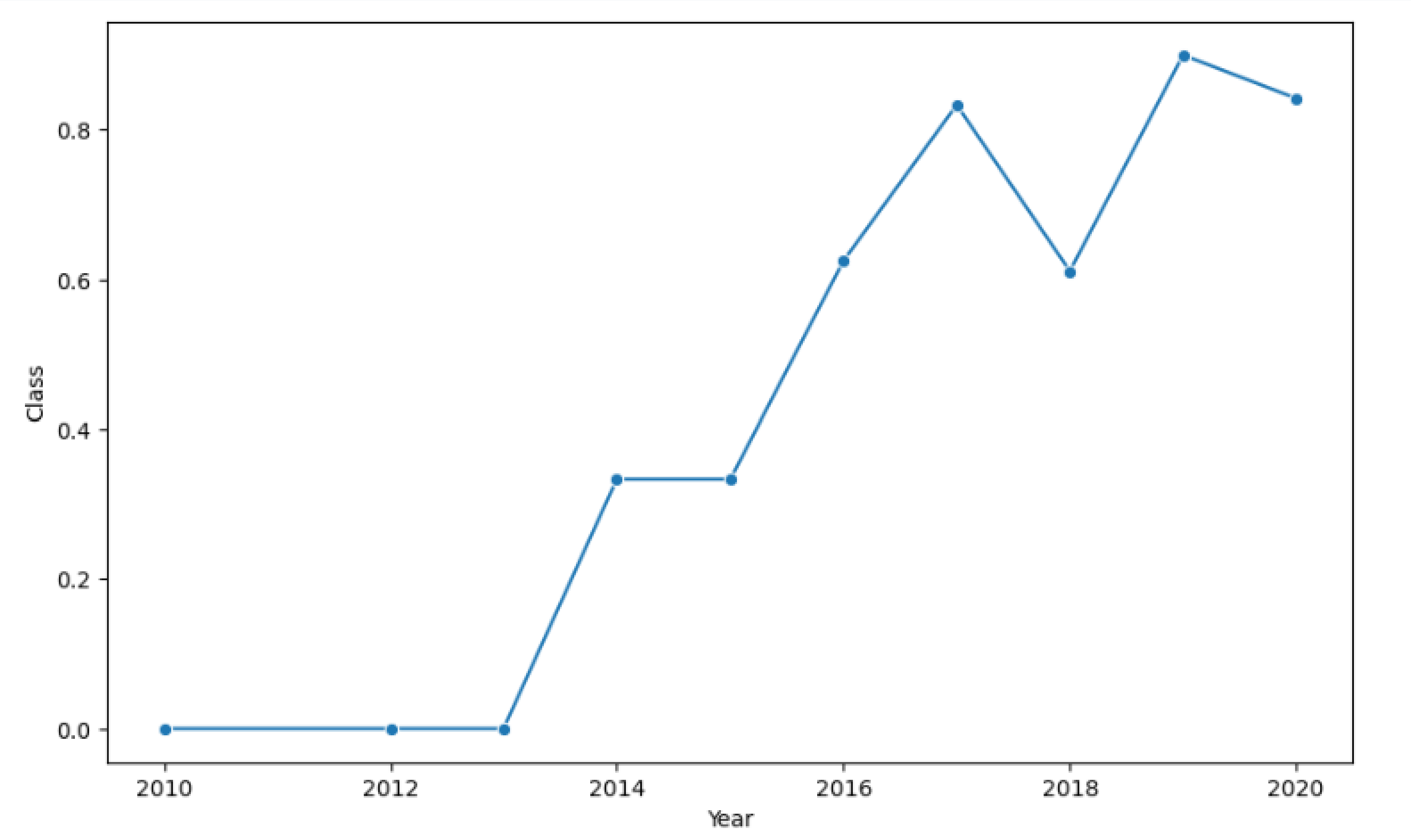
Payload vs. Orbit Type

- Most of the orbits do not have an payloadmass higher than 8000
- VLEO has the highest payloadmasses
- GTO has payloadmasses only between 3000 and 8000



Launch Success Yearly Trend

- The success rate was 0% until 2013
- In 2014 the first successful flight was launched
- The highest success rate was achieved in 2019



All Launch Site Names

There are 4 unique launch sites listed in the SPACEXTBL:

```
%sql Select Distinct Launch_Site from SPACEXTBL
```

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

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

Launch Site Names Begin with 'CCA'

- The mission outcome of the first 5 records with the launch site like 'CCA%' were "Success"

```
%sql Select * from SPACEXTBL where Launch_Site like 'CCA%' Limit 5
```

```
 * sqlite:///my_data1.db
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	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	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

- The total payload was 45.596 kg for NASA launches

```
%sql Select customer, sum(PAYLOAD_MASS__KG_) from SPACEXTBL where Customer = 'NASA (CRS)'
```

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

Customer	sum(PAYLOAD_MASS__KG_)
NASA (CRS)	45596

Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1 was 2.928,4 kg

```
%sql Select Booster_Version, AVG(PAYLOAD_MASS__KG_) from SPACEXTBL where Booster_Version = 'F9 v1.1'
```

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

Booster_Version	AVG(PAYLOAD_MASS__KG_)
F9 v1.1	2928.4

First Successful Ground Landing Date

- The date of the first successful landing outcome on ground pad was the 2015-12-22

```
%sql Select * from SPACEXTBL where Date = (Select min(Date) from SPACEXTBL where Landing_Outcome = 'Success (ground pad)')
```

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

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2015-12-22	1:29:00	F9 FT B1019	CCAFS LC-40	OG2 Mission 2 11 Orbcomm-OG2 satellites	2034	LEO	Orbcomm	Success	Success (ground pad)

Successful Drone Ship Landing with Payload between 4000 and 6000

- These are 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 =  'Success (drone ship)' and PAYLOAD_MASS__KG_ between 4000 and 6000
```

```
 * sqlite:///my_data1.db
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

- List with the total number of successful and failure mission outcomes:

```
%sql Select distinct Mission_Outcome, count (*) from SPACEXTBL group by Mission_Outcome
```

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

Mission_Outcome	count (*)
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- List 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)

 * sqlite:///my_data1.db
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 of failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015:

```
%sql Select substr(Date, 6,2), Landing_Outcome, Booster_Version, Launch_Site from SPACEXTBL where Landing_Outcome = 'Failure (drone ship)' and substr(Date,0,5)='2015'
```

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

substr(Date, 6,2)	Landing_Outcome	Booster_Version	Launch_Site
01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

- Rank 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 Distinct Landing_Outcome, count(*) from SPACEXTBL where Date between '2010-06-04' and '2017-03-20' group by Landing_Outcome order by count(*) desc
```

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

Landing_Outcome	count(*)
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

Section 3

Launch Sites Proximities Analysis

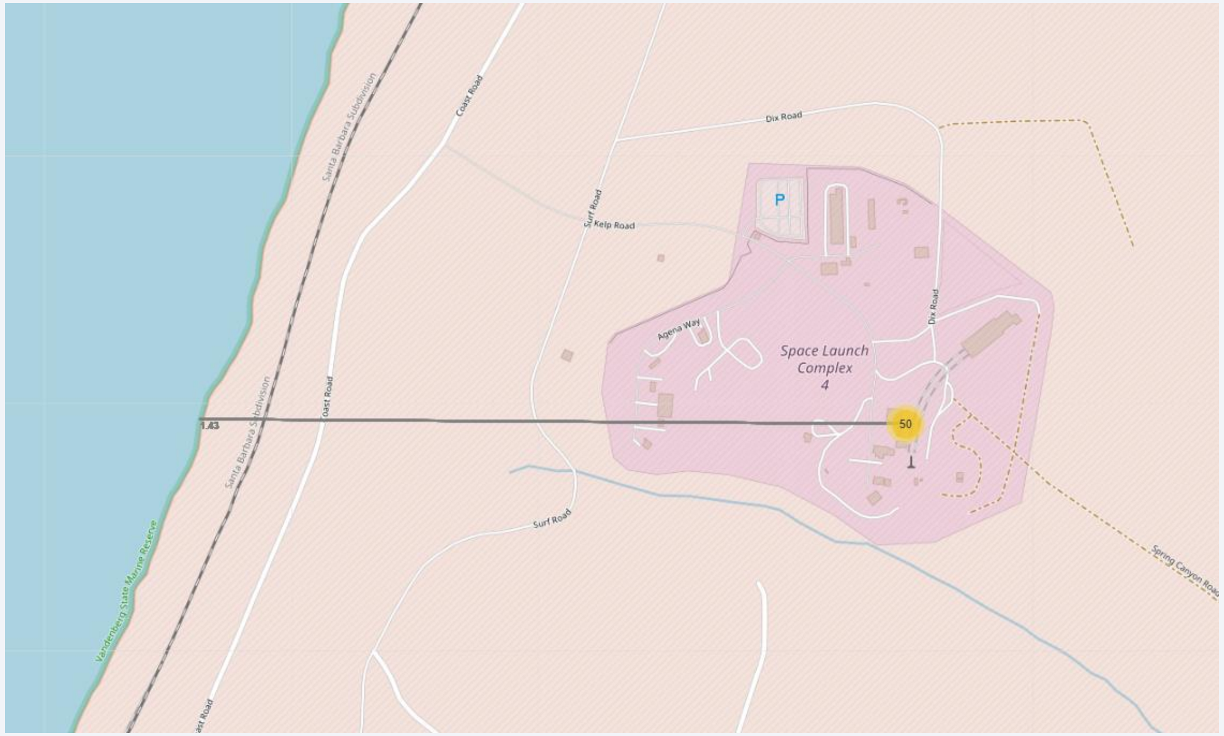
Launch Sites Locations

- Circle and markers show the location of the launch sites
- 3 of 4 launch sites are in Florida (East USA)

Launch success for each site

- Markers show where the locations are
- By clicking on them colored markers appear which represent if a launch was successful or not

Distances between a launch site to its proximities

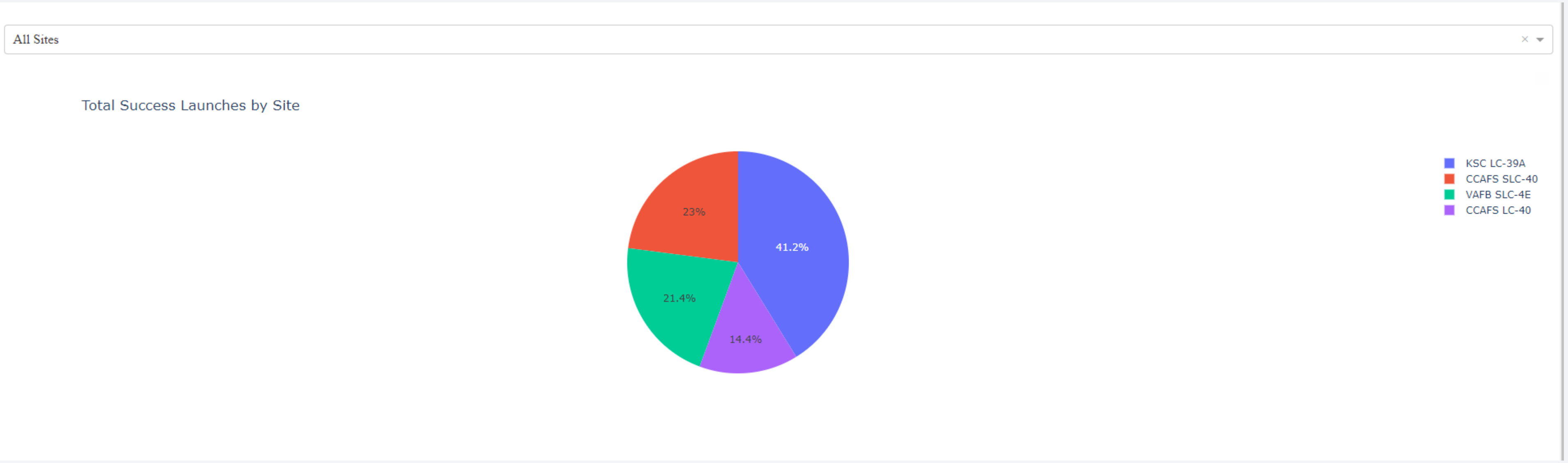


- Marker for Location with number of launches
- Grey line for distance to coast line with displayed distance in km

Section 4

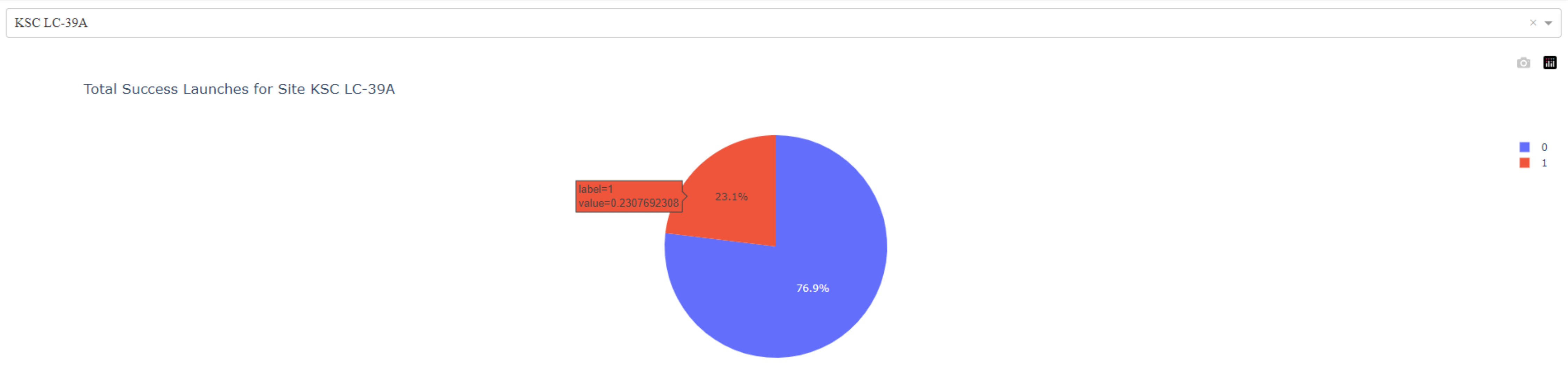
Build a Dashboard with Plotly Dash

Total Success Launches by Side



- Dropdown to select all or a specific launch site
- Piechart to show the success per site in % of total

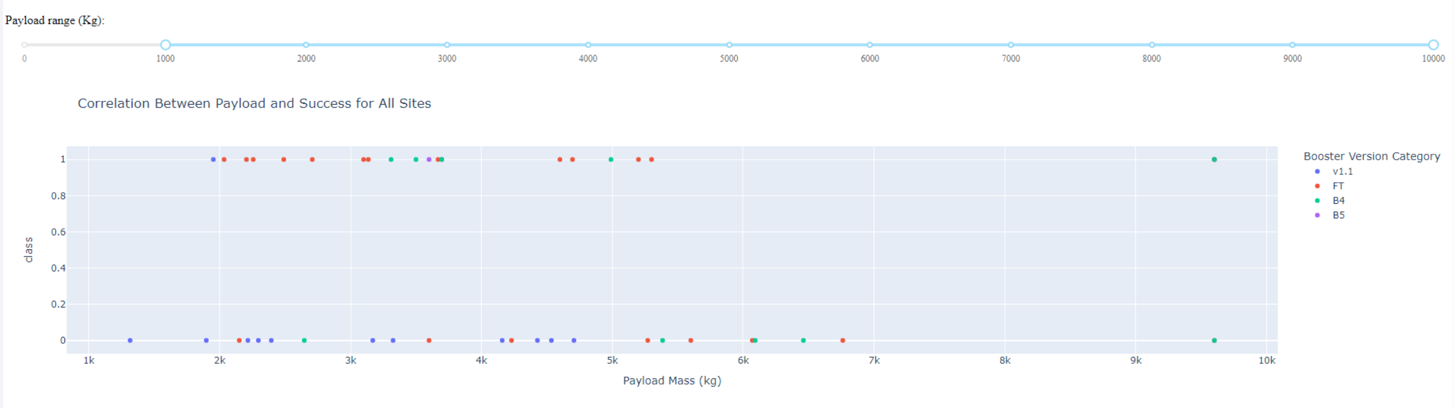
Total Success for KSC LC-39A

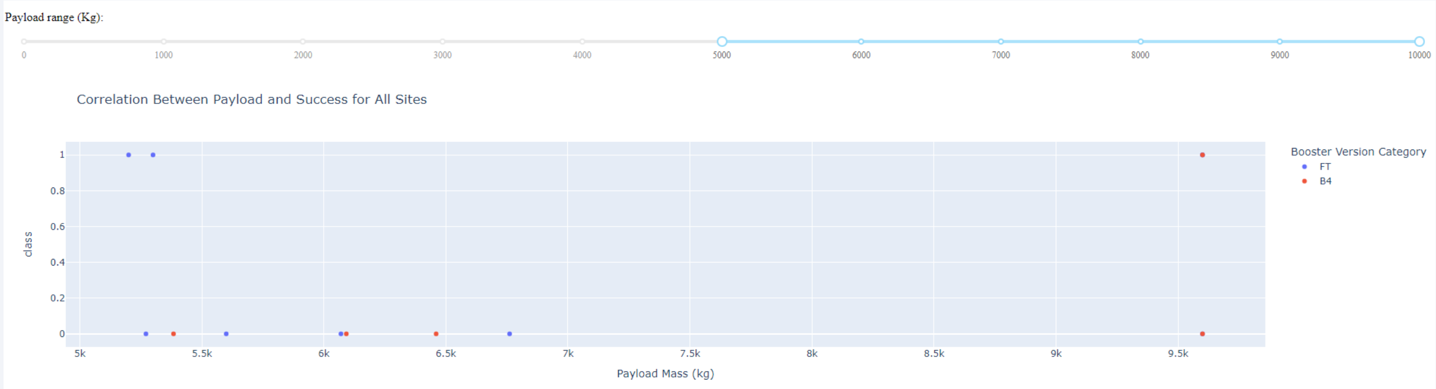


- 23,1% auf the launches for site KSC LC-39A failed
- The other 76,9% were a success

Correlation between payload and success

- Booster version FT has the most successes
- Most successes occurred with a payload between 2000 and 6000



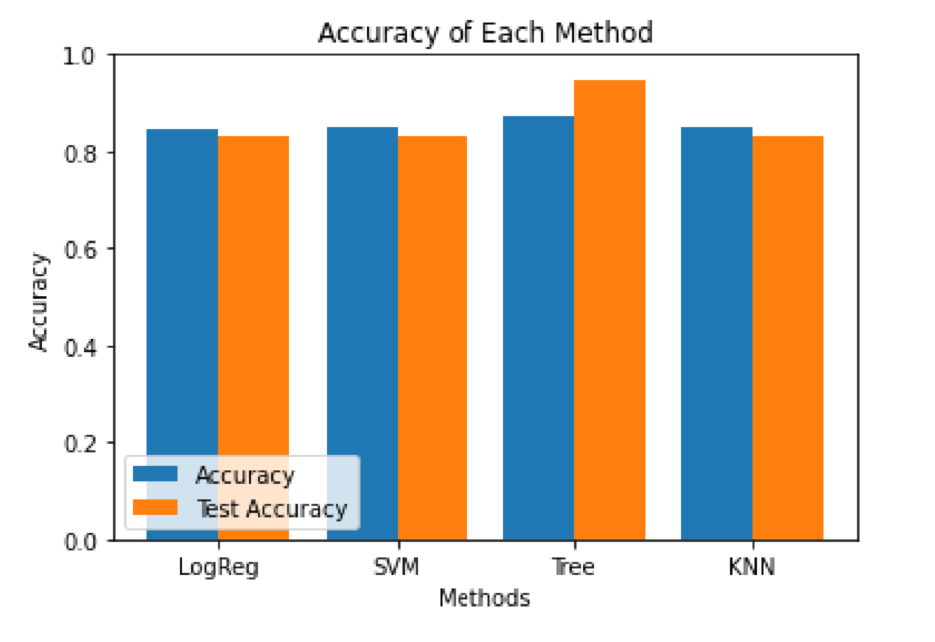


Section 5

Predictive Analysis (Classification)

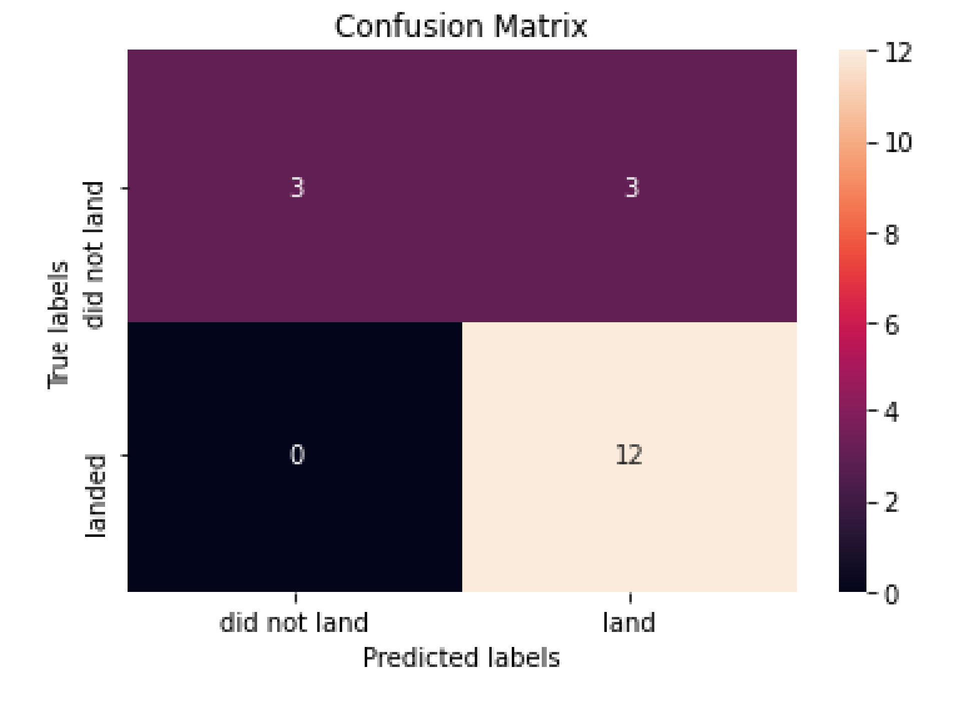
Classification Accuracy

- Best model = decision tree



Confusion Matrix

- Can distinguish between different classes
- False positive are biggest problem



Conclusions

- KSC LC-39A is the most successful launch site
- The higher the payload the more risky
- Success improved over time period
- Best model: decision tree classifier

Thank you!