IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Adrian Gfeller

21/02/2022

Outline

Executive Summary

- We analyze data from Space X Falcon 9 rocket program. The Space X Falcon 9 is a rocket which aims to reuse it's first stage by landing it back on earth. Looking at past launches we compute the likelihood that such landing is successful and therefore the first stage can be reused. Reusing the first stage hugely reduces the cost of a launch.
- We can predict with an accuracy of around 83.3% if a landing is successful.

Introduction

- We are interested in setting up a new space launch program and want to estimate the cost of a launch. We aim to use a similar launch approach than Space X and reuse the first stage of the rocket.
- In order to estimate the success of a stage one landing we analyze data from Space X.
- We want to find answers to the questions:
  - With what accuracy can we predict if first stage can be reused
  - Are there patterns in launch sites, payloads, orbits or other parameters that make a success more ore less likely.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was collected directly from Space X through an API and web scraped from relevant pages
- Perform data wrangling
  - Data was analyzed by launch site and orbit
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - We use Logistic Regression, Support Vector Machine, Decision Tree, KNN to analyze the data

Data Collection

- Data is collected in two different ways:
  - With the Space X API
  - Through Web Scaping of the Space X Wikipedia page.

Data Collection – SpaceX API

Import library

- Import the required library to make HPPT requests
- **Import requests**

API Address

- Add the URL of the API
- **spacex_url="https://api.spacexdata.com/v4/launches**

Do Request

- Request the data from the API
- **response=requests.get(spacex_url)**

Check Request

- Check the response from the API
- **response.status_code**

Pandas DF

- Turn Data into a Pandas dataframe and then analyse the dataframe
- **data=pd.json_normalize(response.json())**

- The GitHub file with code is on [Capstone/Data Collection API.ipynb at main · agfeller/Capstone (github.com)]

Data Collection - Scraping

Install BeautifulSoup
- Install and import the Web Scraping library
- **!pip3 install beautifulsoup4**
- **From bs4 import BeautifulSoup**

Get Web URL
- Get the Wikipedia Space X page
- static_url="https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922"

Do Request
- Request the data from the webpage
- **response=requests.get(static_url)**

Soup It
- Make a BeautifulSoup Object
- **soup=BeautifulSoup(response.text,'html.parser')**

Format Data
- BeautfulSoup has methods to get text and table data out of its objects

The GitHub file with code is on [Capstone/Web Scraping.ipynb at main · agfeller/Capstone (github.com)]

Data Wrangling

- We perform an explanatory data analysis to find patterns in the data:
- Load data into a Pandas dataframe
- We slice the data in various ways and analyze:
  - Launches by launch site
  - Launches by Orbit
  - Mission outcomes by type

The GitHub file with code is on [Capstone/Data Wrangling.ipynb at main · agfeller/Capstone (github.com)]

EDA with Data Visualization

- We analyze the data and plot:
  - FlightNumber vs. PayloadMass
  - FlightNumber vs. LaunchSite
  - Payload vs. LaunchSite
  - Success rate by Orbit
  - FlightNumber vs. Orbit
  - Orbit vs. PayLoadMass
  - Trend of Success

The GitHub file with code is on [Capstone/EDA with Data Visualization.ipynb at main · agfeller/Capstone (github.com)]

EDA with SQL

- We load the SpaceX data into an IBM db2 on the cloud and analyze it with SQL
- We analyse:
    - Launch sites
    - Payload Mass carried by NASA boosters
    - Average payload by F9 v1.1 boosters
    - First date of successful ground pad landing
    - Booster names in a certain payload range
    - Number of successes and failures
    - Number of boosters with maximum payload
    - Falied drone ship landings in 2015
    - Landing outcomes is a certain time range

The GitHub file with code is on [Capstone/EDA with SQL.ipynb at main · agfeller/Capstone (github.com)]

Build an Interactive Map with Folium

- We created an interactive map with folium.
- On the map we added launch sites with the number of launches and whether the launches were successful of a failure. We also added lines as distance measures.
  - For launch sites we added markers
  - For the amount of launches per site we added marker clusters
  - For the distance of a launch site to the seashore we added lines

The GitHub file with code is on
Capstone/Interactive Visual Analytics.ipynb at main · agfeller/Capstone (github.com)

Build a Dashboard with Plotly Dash

- We built an interactive dashboard with ploty dash where we added:
  - A drop down menu for launch sites
  - A pie chart showing the launches per site and success rates per site
  - A payload range selector
  - Graph showing the correlation between payload and the success rate for the selected launch site and payload
- This allows to interactively analyze data for the selected launch sites

The GitHub file with code is on [Capstone/spacex_dash_app.py.txt at main · agfeller/Capstone (github.com)]

Predictive Analysis (Classification)

- We built a model to predict the success rate of a first stage landing of SpaceX rockets. To do so we used:
  - Logistic Regression
  - Support Vector Machine
  - Decision Tree
  - KNN analysis
- To do the analyses we first normalized the data and then split it up into a training and test set. We analyzed the accuracy using the score function.

The GitHub file with code is on [Capstone/Predictive Analysis.ipynb at main · agfeller/Capstone (github.com)]

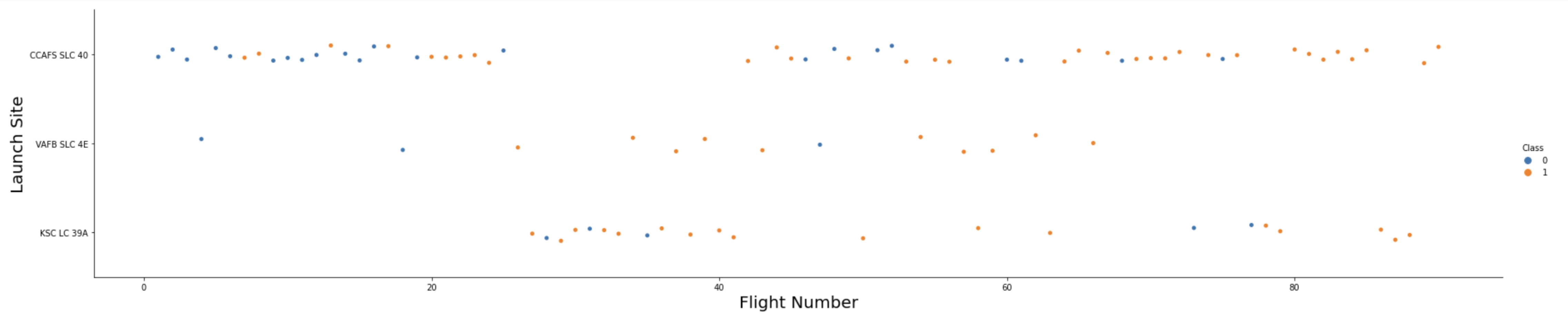
Results

- Exploratory data analysis results
  - Landing success increases over time and seem to be better with heavier payloads
  - Different launch sites are used with different payloads
  - Success seems to depend on the orbit
- Interactive analytics demo in screenshots
  - Success seems to depend on the launch site
  - Success depends on the booster version
- Predictive analysis results
  - Tests to predict the success of a landing have an accuracy of up to 83.3%

Section 2

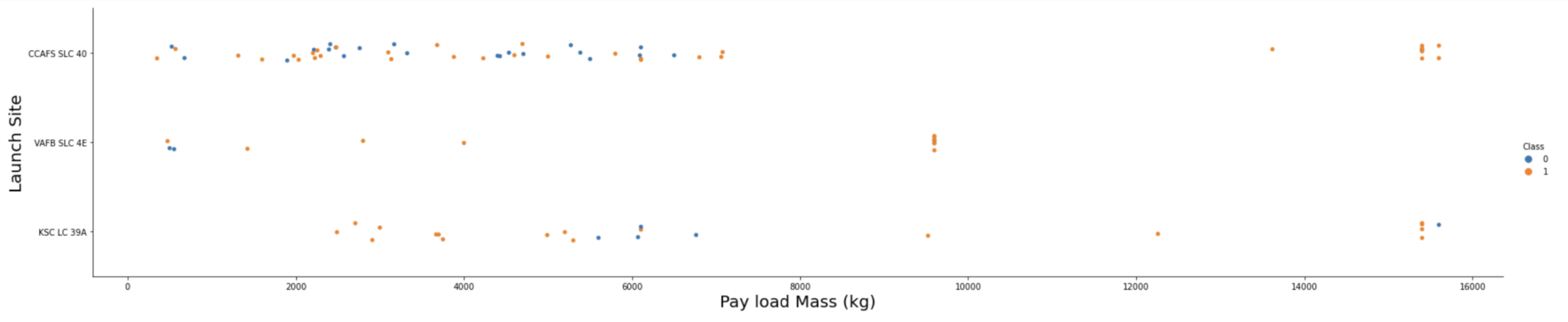
Insights drawn from EDA

Flight Number vs. Launch Site



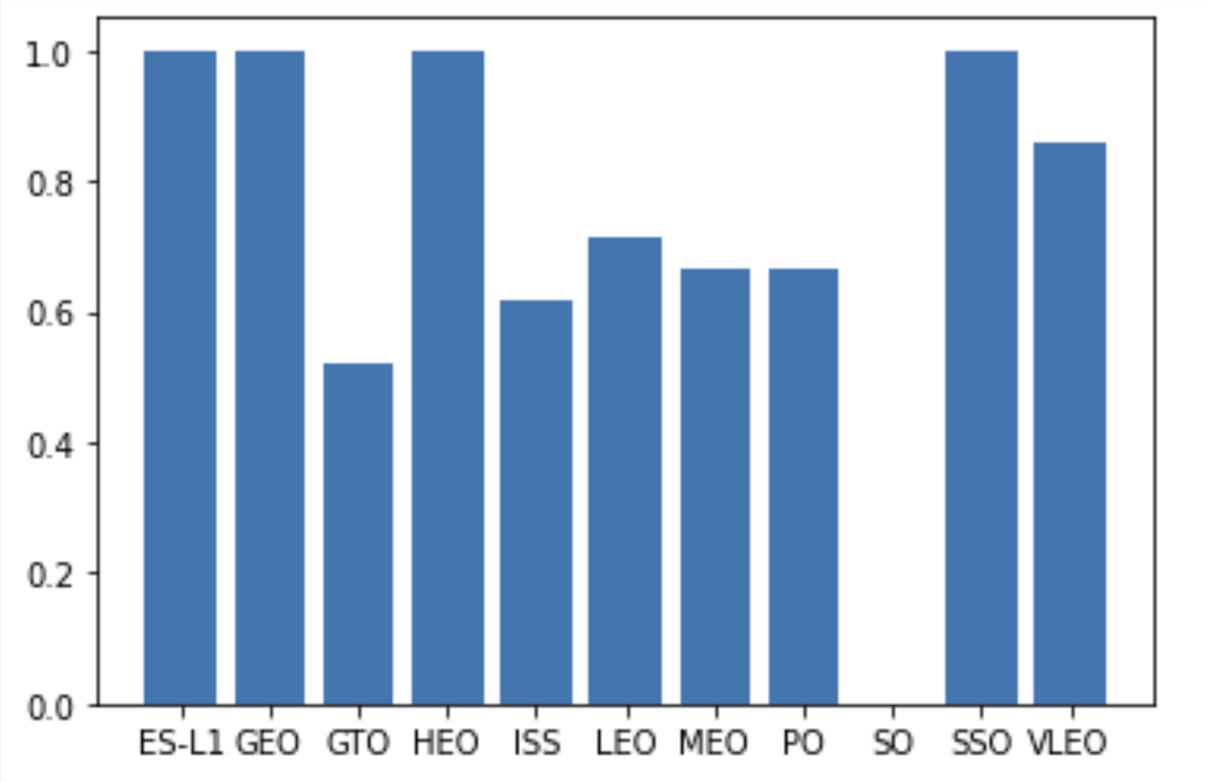
- Class 1 represents a success which clearly increase over time

Payload vs. Launch Site

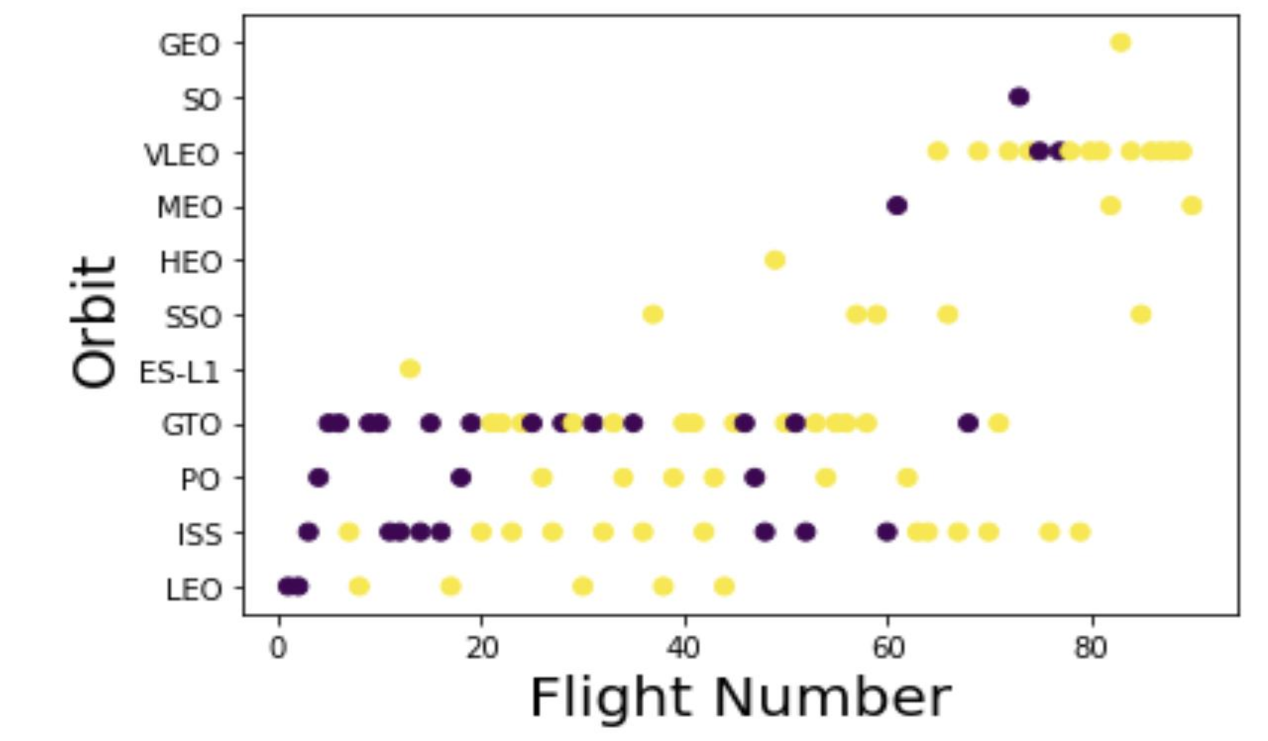


- CCAFS and KSC are used much more frequently than VFAB and are the only ones for very heavy payloads.

Success Rate vs. Orbit Type

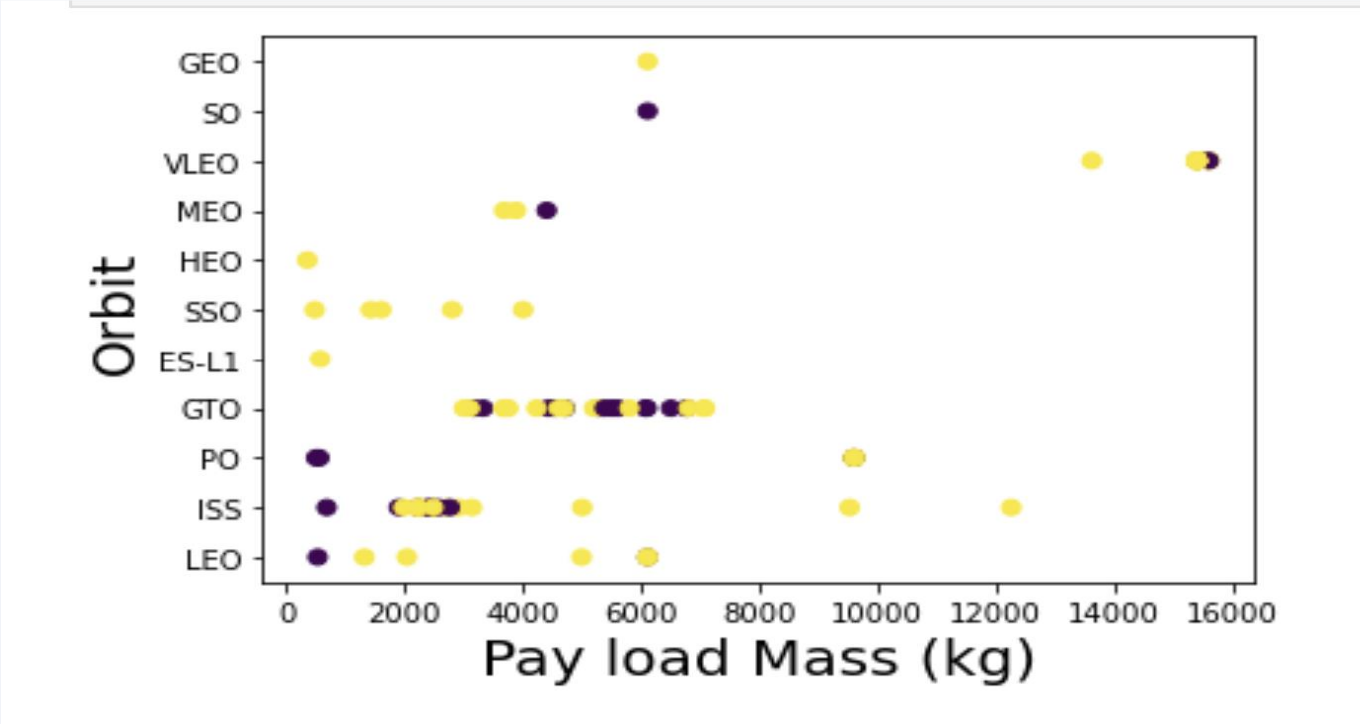


Flight Number vs. Orbit Type



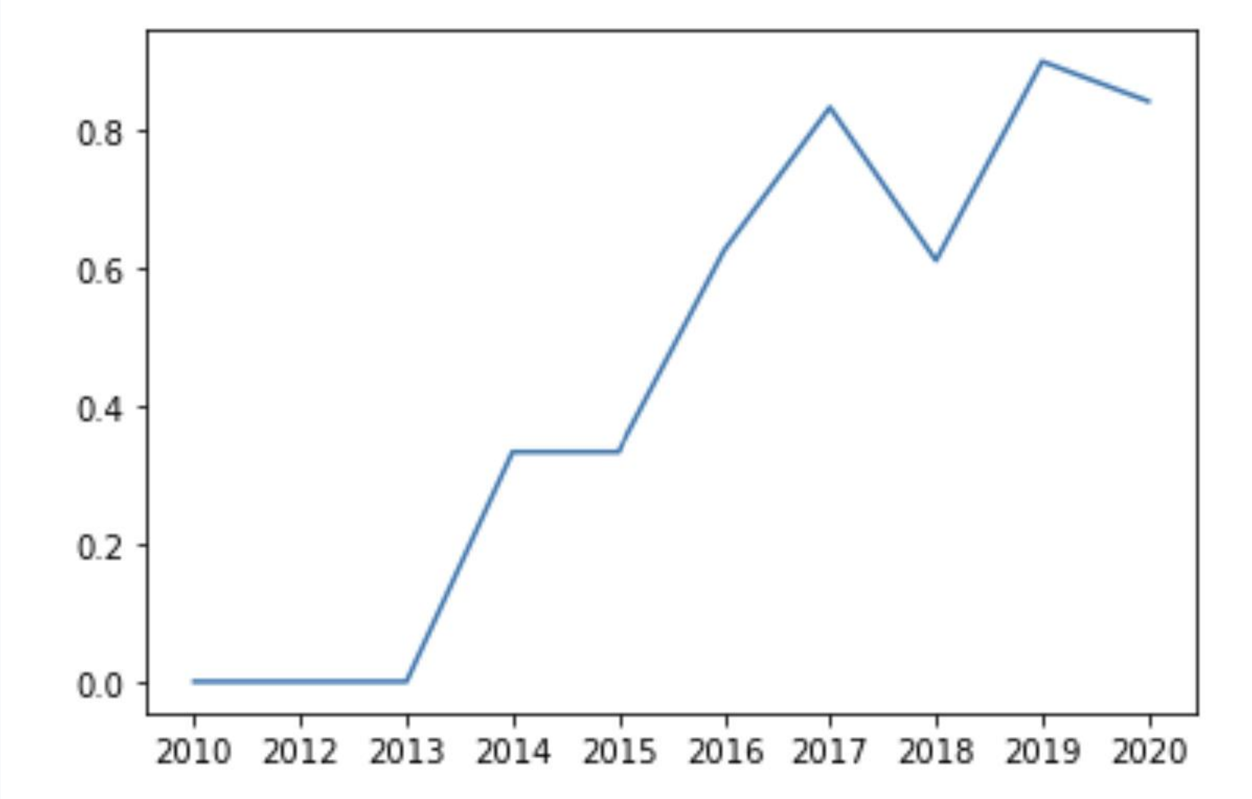
- Some orbits are only done in later launches
- Yellow = success, brown = failure

Payload vs. Orbit Type



- Different orbits have different success rates
- Yellow = success, brown = failure

Launch Success Yearly Trend



- Successes greatly increase over time

All Launch Site Names

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

- Using %sql to write SQL in a Jupiter workbook:
- **%sql select unique launch_site from SPACEXTBL**

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

- LIMIT 5 limits the results to 5 rows:

%sql SELECT * FROM SPACEXTBL WHERE LEFT(launch_site,3) = 'CCA' LIMIT 5

Total Payload Mass

- Total payload carried by boosters from NASA:

45596

```
%sql SELECT SUM(payload_mass__kg_) FROM SPACEXTBL WHERE customer = 'NASA (CRS)'
```

Average Payload Mass by F9 v1.1

- Average payload mass carried by booster version F9 v1.1:

2534

%sql SELECT AVG(payload_mass__kg_) FROM SPACEXTBL WHERE LEFT(booster_version,7) = 'F9 v1.1'

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad:

2015-12-22

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

Successful Drone Ship Landing with Payload between 4000 and 6000

- List the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000:

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

```
%sql SELECT booster_version FROM SPACEXTBL WHERE landing__outcome = 'Success (drone ship)'
AND payload_mass__kg_ >4000 AND payload_mass__kg_<6000
```

Total Number of Successful and Failure Mission Outcomes

Total number of successful and failure mission outcomes:

```
%sql SELECT landing__outcome,COUNT(landing__outcome)
FROM SPACEXTBL GROUP BY landing__outcome
```

landing__outcome	2
Controlled (ocean)	5
Failure	3
Failure (drone ship)	5
Failure (parachute)	2
No attempt	22
Precluded (drone ship)	1
Success	38
Success (drone ship)	14
Success (ground pad)	9
Uncontrolled (ocean)	2

Boosters Carried Maximum Payload

- Names of the booster which have carried the maximum payload mass:

%sql SELECT DISTINCT booster_version FROM SPACEXTBL WHERE payload_mass__kg_ = (SELECT MAX(payload_mass__kg_) FROM SPACEXTBL)

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

2015 Launch Records

Failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015:

DATE	time__utc_	booster_version	launch_site	payload	payload_mass__kg_	orbit	customer	mission_outcome	landing__outcome
2015-01-10	09:47:00	F9 v1.1 B1012	CCAFS LC-40	SpaceX CRS-5	2395	LEO (ISS)	NASA (CRS)	Success	Failure (drone ship)
2015-04-14	20:10:00	F9 v1.1 B1015	CCAFS LC-40	SpaceX CRS-6	1898	LEO (ISS)	NASA (CRS)	Success	Failure (drone ship)

%sql SELECT * FROM SPACEXTBL WHERE LEFT(landing__outcome,7) = 'Failure' AND YEAR(DATE) = '2015'

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

- 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 landingcount
        FROM SPACEXTBL
        WHERE DATE > '2010-06-04'
        AND DATE < '2017-03-20'
        GROUP BY landing__outcome
        ORDER BY landingcount DESC
```

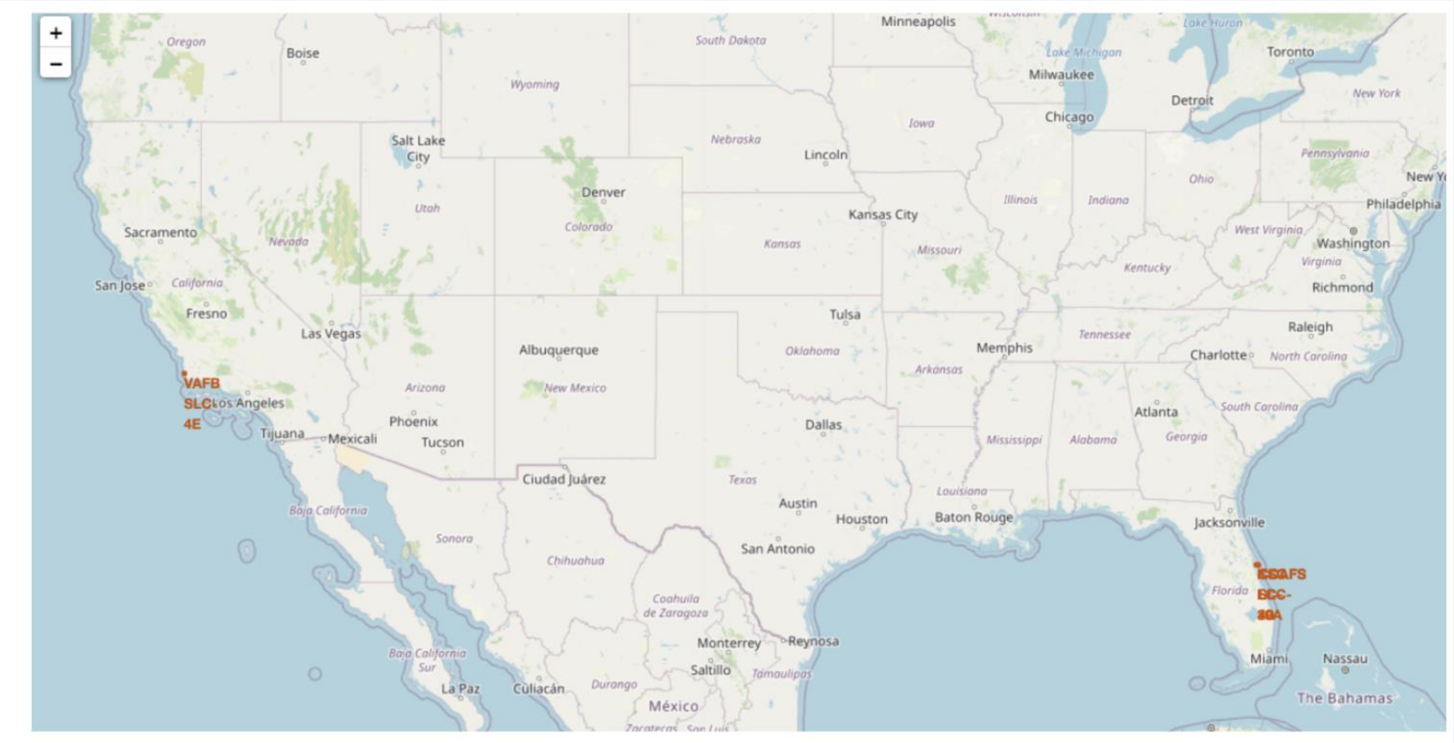
landing__outcome	landingcount
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Uncontrolled (ocean)	2
Failure (parachute)	1
Precluded (drone ship)	1

- Using %%sql for multiple lines of SQL code in a Jupiter workbook

Section 3

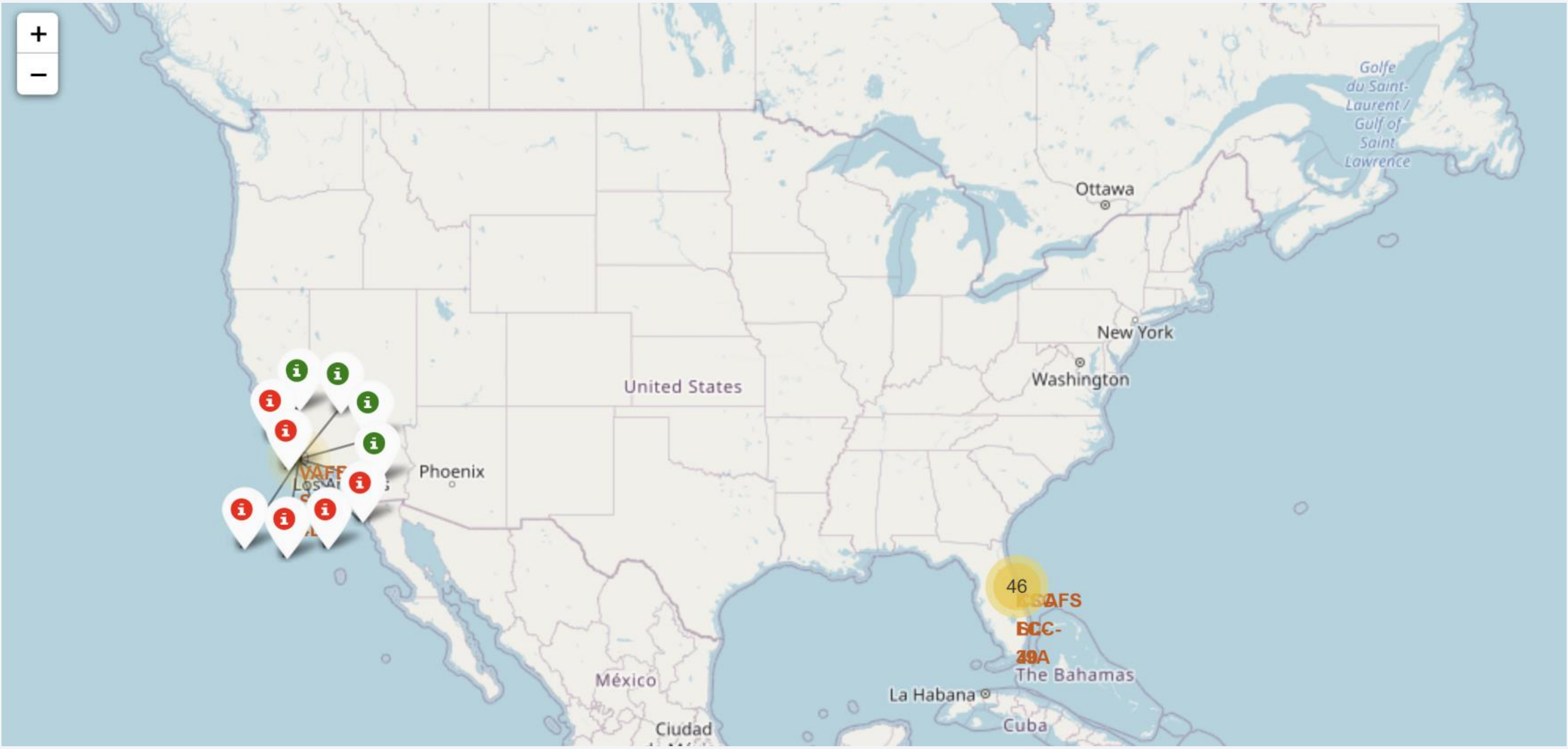
Launch Sites Proximities Analysis

Folium: Launch Sites



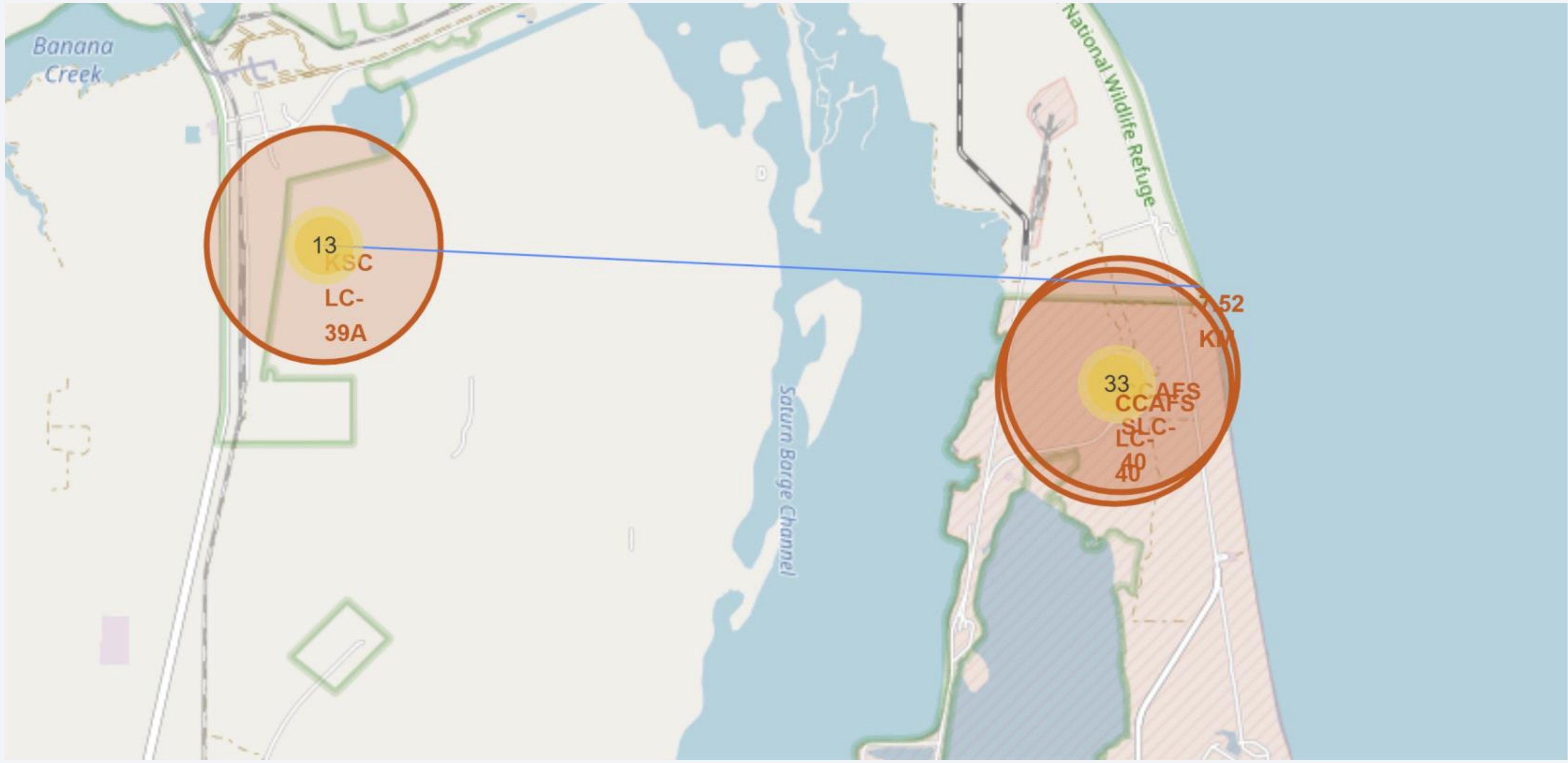
- All launch sites are very close to the sea shore

Folio: Launches by Site



- Launches are grouped with a marked cluster and colour coded: green = success, red = failure

Folio: Launch site proximity to shore

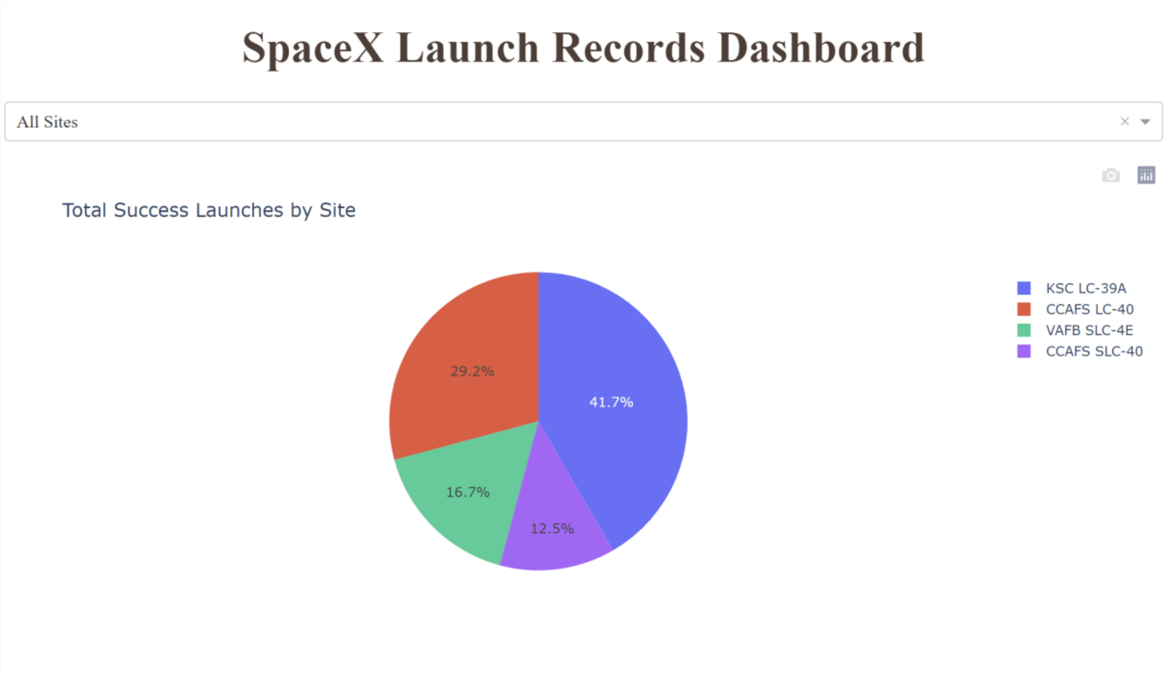


- With Folio lines can be added to the map.
- The line here represents the distance to the shore of 7.52km

Section 4

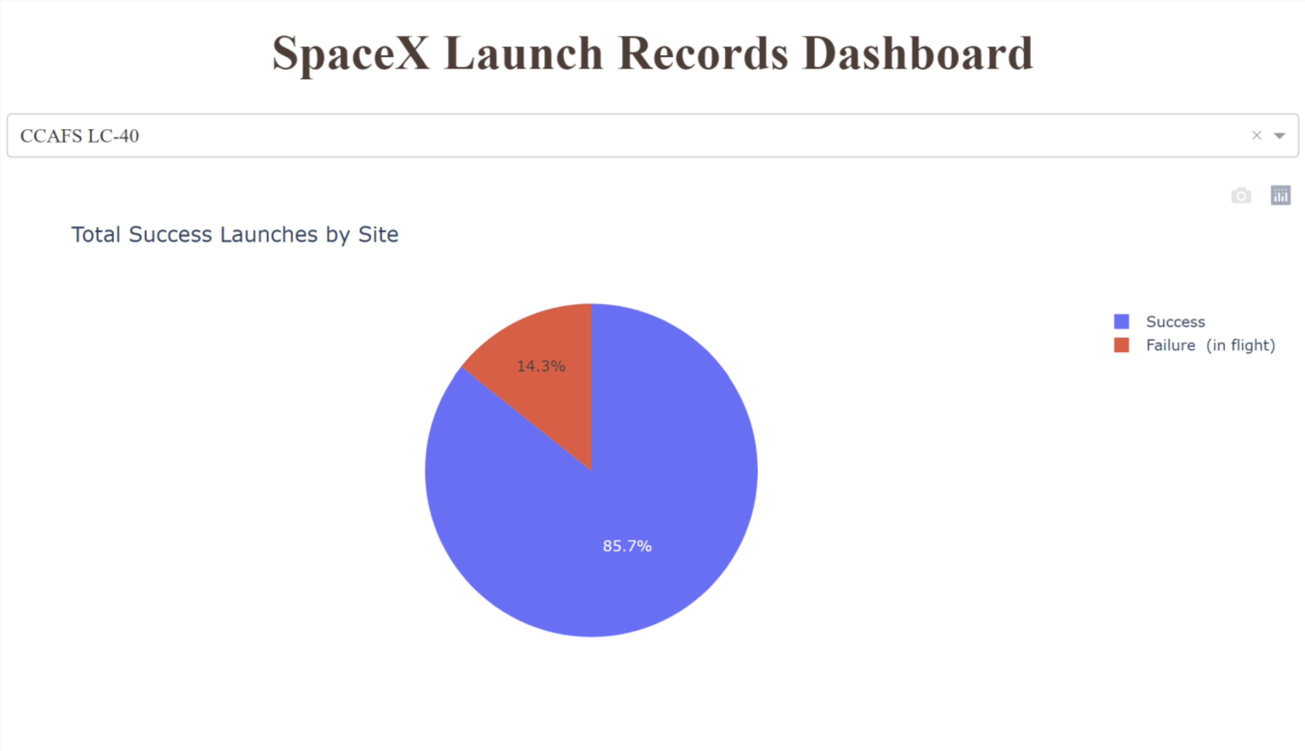
Build a Dashboard with Plotly Dash

Dash: Successful launches by site



- The pie chart shows the percentage of success by launch site for all sites.

Dash: Launch site with highest success ratio



Launch site XXX has a success ratio of 100%

Did not take screen a shot from the 100% success launch site

Dash: Payload vs Success

- The graph shows success (class = 1) and failures (class = 0) by payload. The colors show the various booster versions.

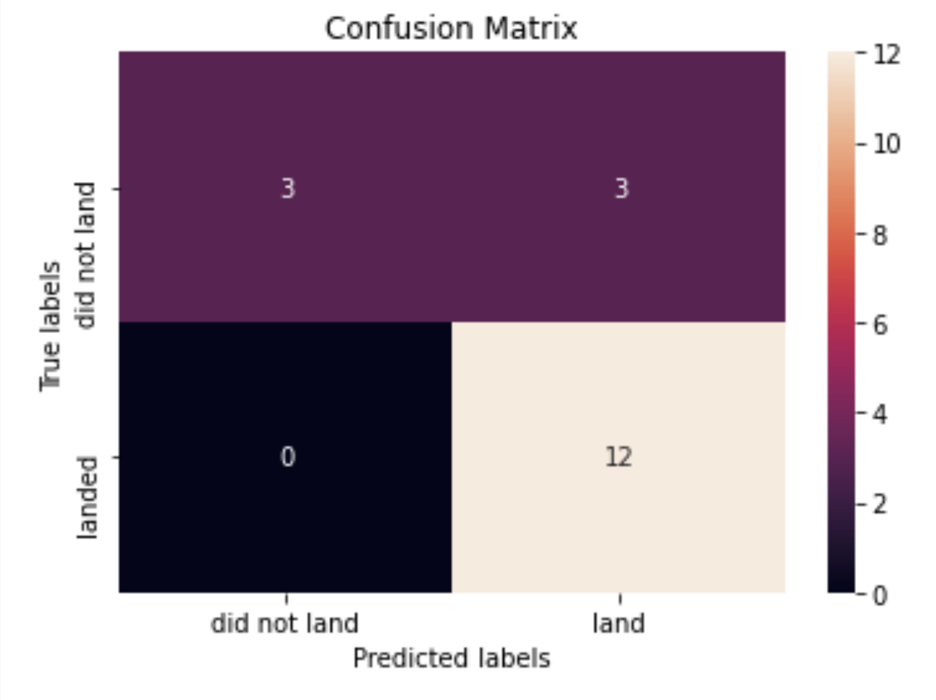
Section 5

Predictive Analysis (Classification)

Classification Accuracy

- Build four models to analyze the accuracy of the model predictions (lognormal regression, support vector analysis, decision tree, KNN)
- All models perform the same with an accuracy of around 83.3%

Confusion Matrix



- The model has some false positives (3 in top right corner)

Conclusions

- Using data from Space X we can analyse the relation of various parameters (launch mass, launch site, orbit, launch number) on the success rate of landing the stage one rocket back on earth.
- Using predictive analysis, we estimate the accuracy of models to predict such a successful landing.
- All the different models all have an accuracy of around 83.3%

Appendix

- All additional data is in the Jupiter workbooks which are linked in the presentation.

Thank you!