IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Adriana Melges Quintanilha Weingart
June 2024

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion

Executive Summary

- Data collection: data was collected from SpaceX API and Wikipedia launch table, for launches prior to 2020.
- Data wrangling / preparation: data was cleaned in preparation for EDA and Feature Engineering.
- EDA (Exploratory Data Analysis): visualizations via plots and SQL plots were used to explore and know the data.
- Additionally, interactive visual analysis was performed using Folium and Plotly Dash.
- Finally, after splitting the dataset into train and test sets, predictive analysis using classification models was performed. Models used were: Logistic Regression, SVM, Decision Trees and K-NN.
- Confusion Matrix and Accuracy (method score) were used to evaluate the models performance.
- Considering the test set, all models performed similarly. But, the accuracy of the train set was better for Decision Tree model.

Introduction

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.

Questions of interest:

- Does the success or failure depend on the launch site? Does it depend on the landing target?
- Is it possible to predict a successful first stage landing of Falcon9?
- Which machine learning model would work best (highest accuracy) to predict the outcome of Falcon9 first stage landing in a future launch?

Section 1

Methodology

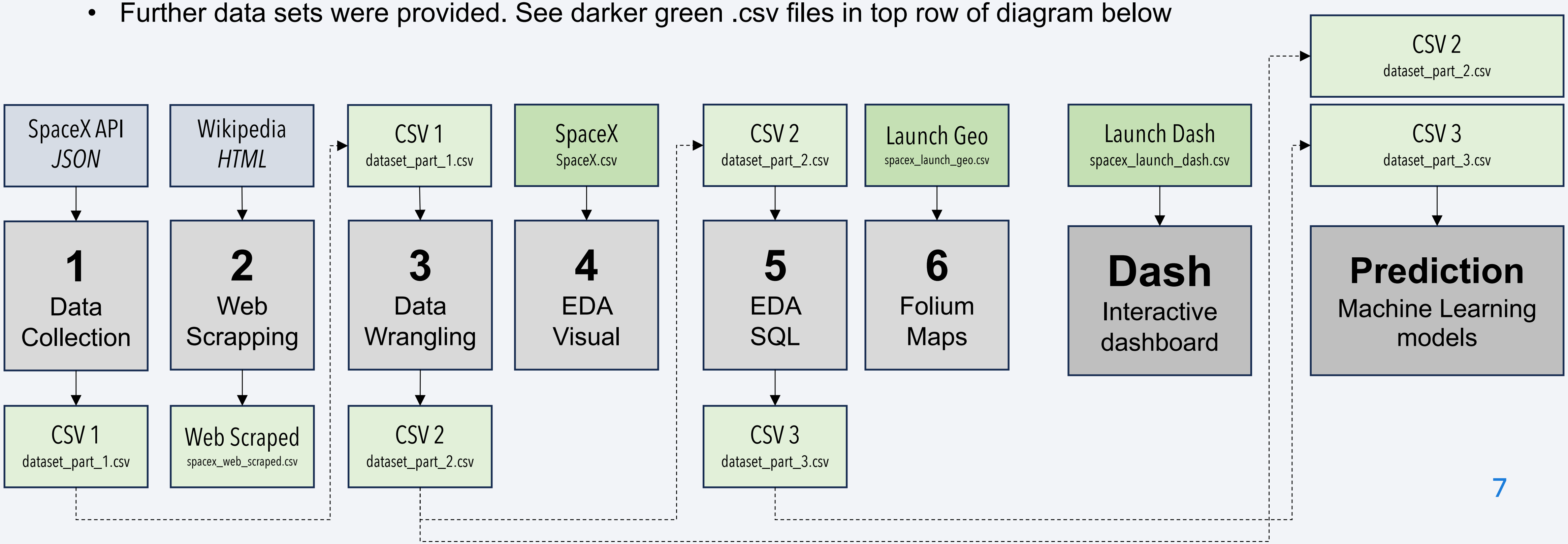
Methodology

Executive Summary

- Data collection: data was collected from SpaceX API and Wikipedia launch table, for launches prior to 2020.
- Data wrangling / preparation: data was cleaned in preparation for EDA and Feature Engineering.
- EDA (Exploratory Data Analysis): visualizations via plots and SQL plots were used to explore and know the data.
- Additionally, interactive visual analysis was performed using Folium and Plotly Dash.
- Finally, after splitting the dataset into train and test sets, predictive analysis using classification models was performed. Models used were: Logistic Regression, SVM, Decision Trees and K-NN.
- Confusion Matrix and Accuracy (method score) were used to evaluate the models performance.

Data Collection

- Source of datasets:
  - An IBM copy of a call to the publicly accessible SpaceX API with launch data in JSON format.
  - A permanently linked Wikipedia page with launch data in HTML tables (9 June 2021 revision).
  - Further data sets were provided. See darker green .csv files in top row of diagram below



Data Collection – SpaceX API

- SpaceX launch data is available via REST API Endpoints.
- The flowchart presents the structure of the API Endpoints, and those used in this notebook.
- Data was requested using Python method .get() and made into a json format, which was normalized and made into a Pandas dataframe.

To make the requested JSON results more consistent, we will use the following static response object for this project:

```
static_json_url = 'https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/API_call_spacex_api.json'
```

We should see that the request was successfull with the 200 status response code

```
response.status_code
```

```
200
```

Now we decode the response content as a Json using `.json()` and turn it into a Pandas dataframe using `.json_normalize()`

```
# Use json_normalize meethod to convert the json result into a dataframe
data = pd.json_normalize(response.json())
```

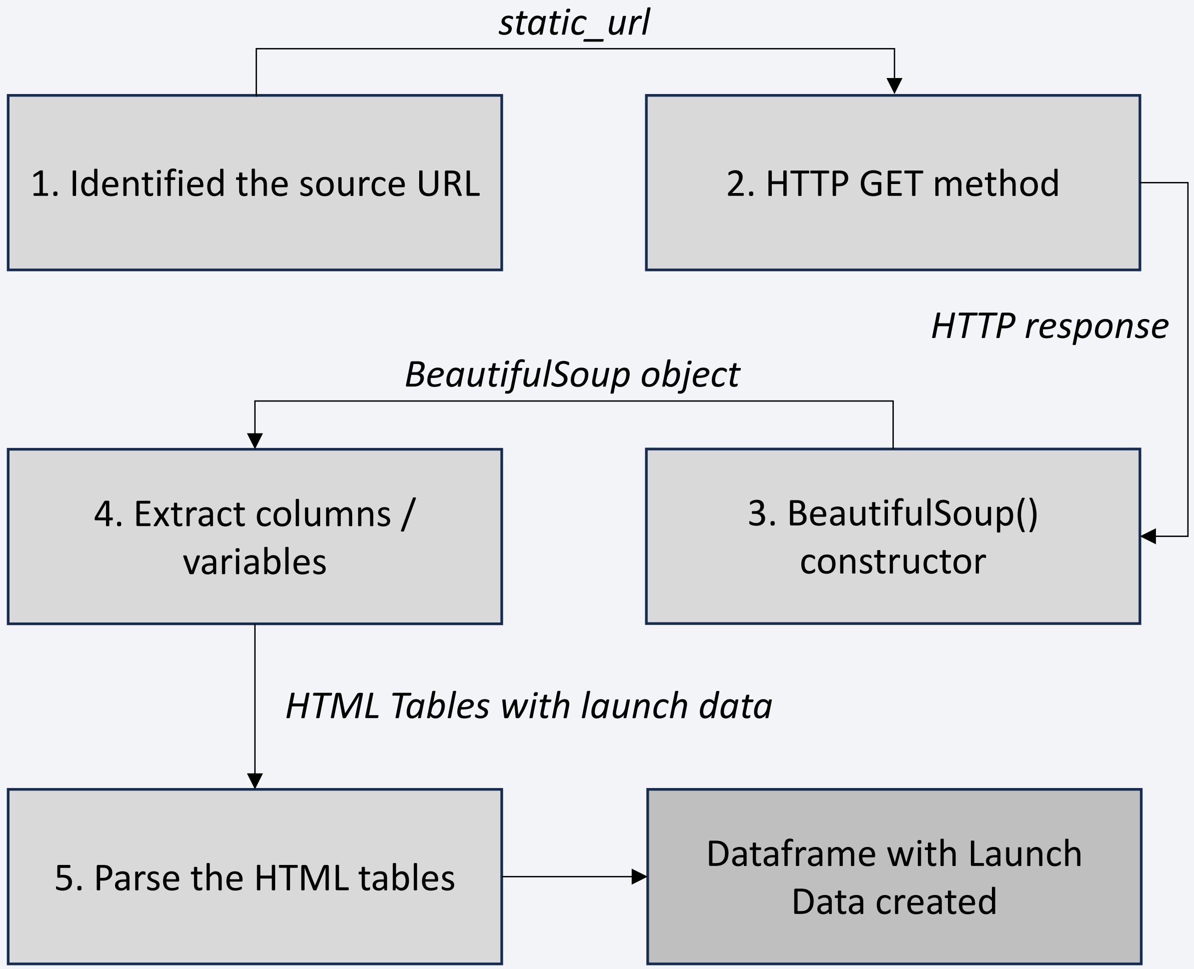
- The notebook is available at https://bit.ly/AW_SpaceXFalcon9_Lab1

SpaceX REST API Endpoints
https://api.spacexdata.com/v4/...

- ***Rockets launch data:*** *https://api.spacexdata.com/v4/launches/past*
- ***Booster Version data:*** *https://api.spacexdata.com/v4/rockets*
- ***Launch Site data:*** *https://api.spacexdata.com/v4/launchpads*
- ***Payload data:*** *https://api.spacexdata.com/v4/payloads*
- ***Core data:*** *https://api.spacexdata.com/v4/cores*

Data Collection - Scraping

- Wikipedia documents data of rockets from the Falcon9 family that have been launched since June 2010 at https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches
- Python Method requests.get() was used to get the data from this wiki, which was extracted using BeautifulSoup() constructors, and parsed to build the dataframe with the launch data for analysis. The diagram on the right illustrates this process.
- The notebook is available at https://bit.ly/AW_SpaceXFalcon9_Lab2



Data Wrangling

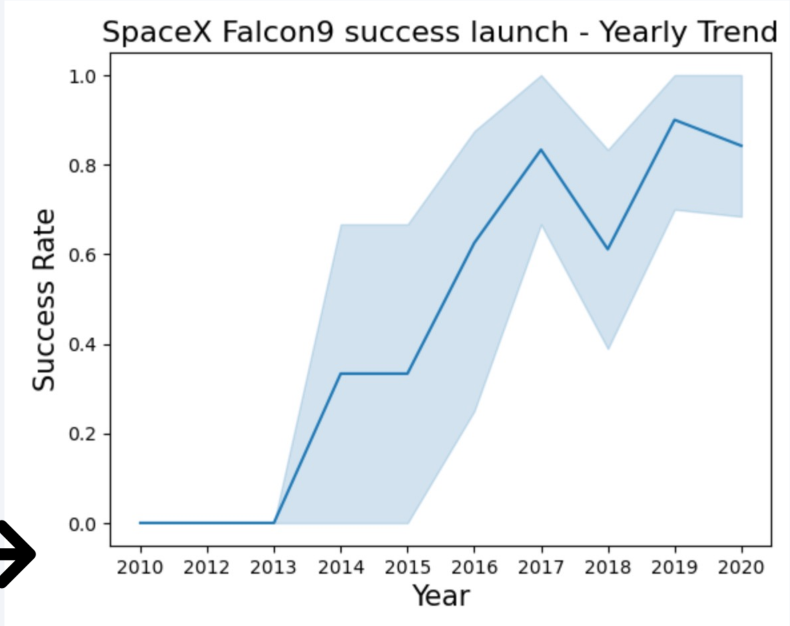
1. EDA:
    1. Check Missing Data.
    2. Check number of launches on each site.
    3. Check number and occurrence of each orbit (as each launch aims to a dedicated orbit, within 11 different types).
    4. Check number and occurrence of mission outcome of the orbits.
2. Feature Engineering:
    1. Create a landing outcome label from *outcome* column, where 1 indicates a successful landing for the Falcon9 first stage and 0 means failure, as shown in the table.

- The notebook is available at https://bit.ly/AW_SpaceXFalcon9_Lab3

Mission Outcome	Explanation	Outcome label
True Ocean	Successful land to an ocean region	1
False Ocean	Unsuccessful land to an ocean region	0
True RTLS	Successful land to a ground pad	1
False RTLS	Unsuccessful land to a ground pad	0
True ASDS	Successful land on a drone ship	1
False ASDS	Unsuccessful land on a drone ship	0

EDA with Data Visualization

Trends	Chart Type	Python command	Mission Outcome relationship between
Launch Site Trends	Scatterplot	sns.catplot	Launch Site vs Flight Number
	Scatterplot	sns.catplot	Launch Site vs Payload
Orbit Types trends	Bar chart	sns.catplot / kind = 'bar'	Success rate of each orbit type
	Scatterplot	sns.catplot	Orbit Type vs Flight Number
	Scatterplot	sns.catplot	Orbit Type vs Payload
Trends based on time	Line plot	sns.lineplot	Trend by Year



- Different types of charts were used to visually explore and understand the data and its distribution and relationships. The table above describes the types of charts used and their objective.
- In evidence, lineplot chart for SpaceX Falcon9 success launch yearly trend.

- The notebook is available at https://bit.ly/AW_SpaceXFalcon9_Lab4

EDA with SQL

- EDA was executed with SQL to extract information about:
  - Launch Sites
  - Payload masses
  - Dates
  - Booster Types
  - Mission outcomes
- The notebook is available at https://bit.ly/AW_SpaceXFalcon9_Lab5
- As there were issues in the lab saving and downloading the notebook, the code and results ref. this activity are in the .pdf file available at https://bit.ly/AW_SpaceXFalcon9_Lab5_pdf .
- Issue documented in the Discussion Forum, in thread https://www.coursera.org/learn/applied-data-science-capstone/discussions/forums/cS7b6vOeEeqgoAqp1bXgEw/threads/rQTtyi8TEe--cRJ6pAVjWw

Build an Interactive Map with Folium

- An interactive map was built with Folium, with the markers:
    - Launch sites
    - Success and failed launches for each site
    - Distances between a launch site to elements (coastline, city, highway and railway) in its proximity
- An interactive map can give a quick overview and understanding of the data in the dataset, helping either for real-time information, or for data preparation for future analysis. The items marked in the map are related to launch site, which may affect the mission outcome results.
- The notebook is available at https://bit.ly/AW_SpaceXFalcon9_Lab6

Build a Dashboard with Plotly Dash

- An interactive dashboard was created with Plotly Dash, allowing:
    - Initial view of Success launches by site (using a Pie Chart).
    - Drop Down menu to select individual Launch Site, which in turn, will present a Pie Chart with Successful and Failed launches for the selected site.
    - Correlation scatterplot chart between Payload and Success (per site, or for all sites, depending on previous selection), indicating Booster Version Cathegory, also controlled by a range bar, to select payload mass
- These charts allow a quick and dynamic view of the data – success/failure rate, and overview for all the launch sites. While this does not contribute directly for a prediction analysis, it allows a quick overview and understanding of the data.
- The notebook is available at https://bit.ly/AW_SpaceXFalcon9_Lab7

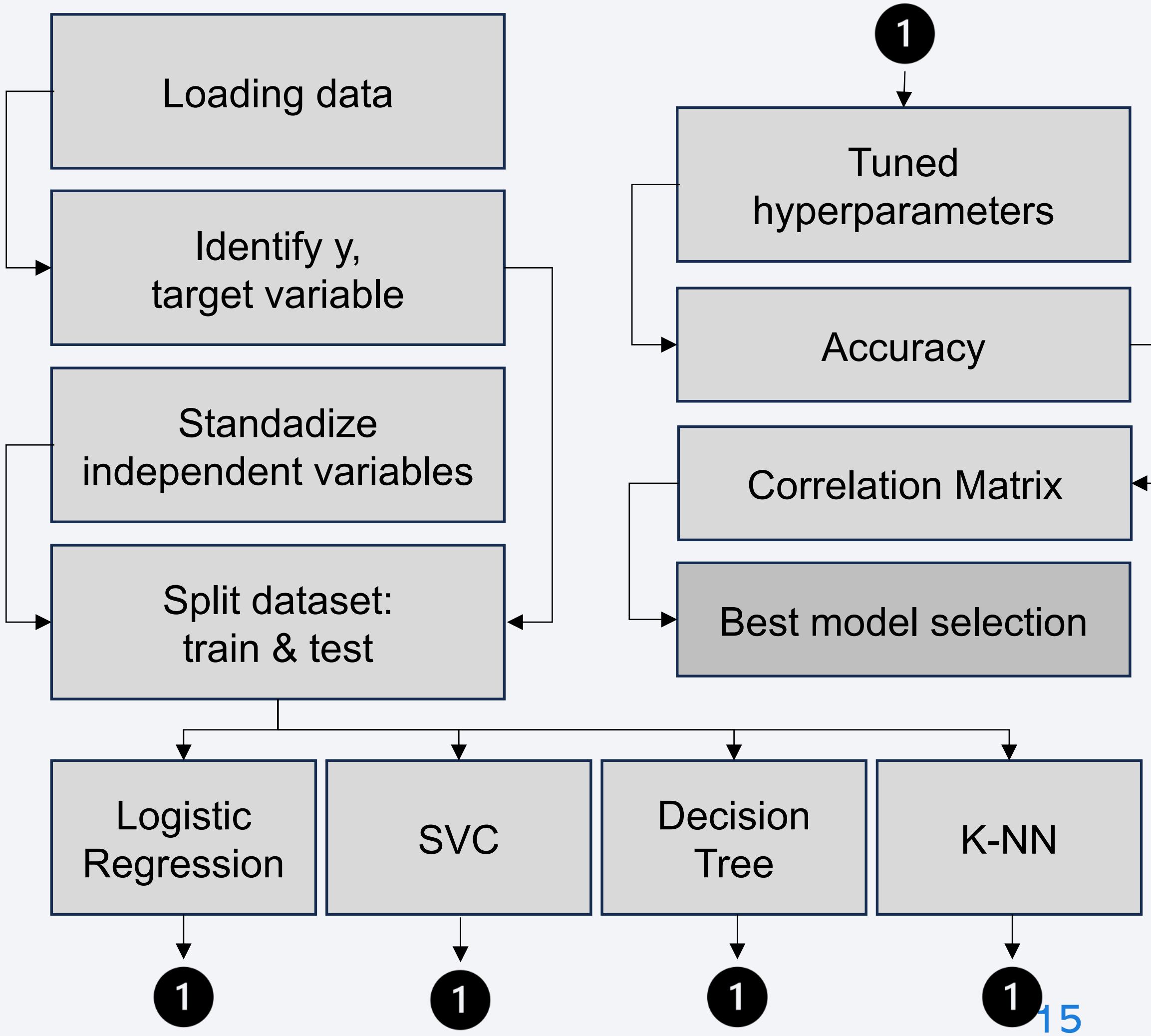
Predictive Analysis (Classification)

- **Preparing the data:**
    1. Loading the data
    2. Standardize the data
    3. Split the data between train (80%) and test (20%)

- **Models:**
    - Logistic Regression
    - SVC (Support Vector Classifier)
    - Decision Tree
    - K-NN (K-Nearest Neighbors)

- **Actions:**
    - Model train
    - Hyperparameters selection (using GridSearchCV)
    - Accuracy
    - Confusion Matrix

- The notebook is available at https://bit.ly/AW_SpaceXFalcon9_Lab8



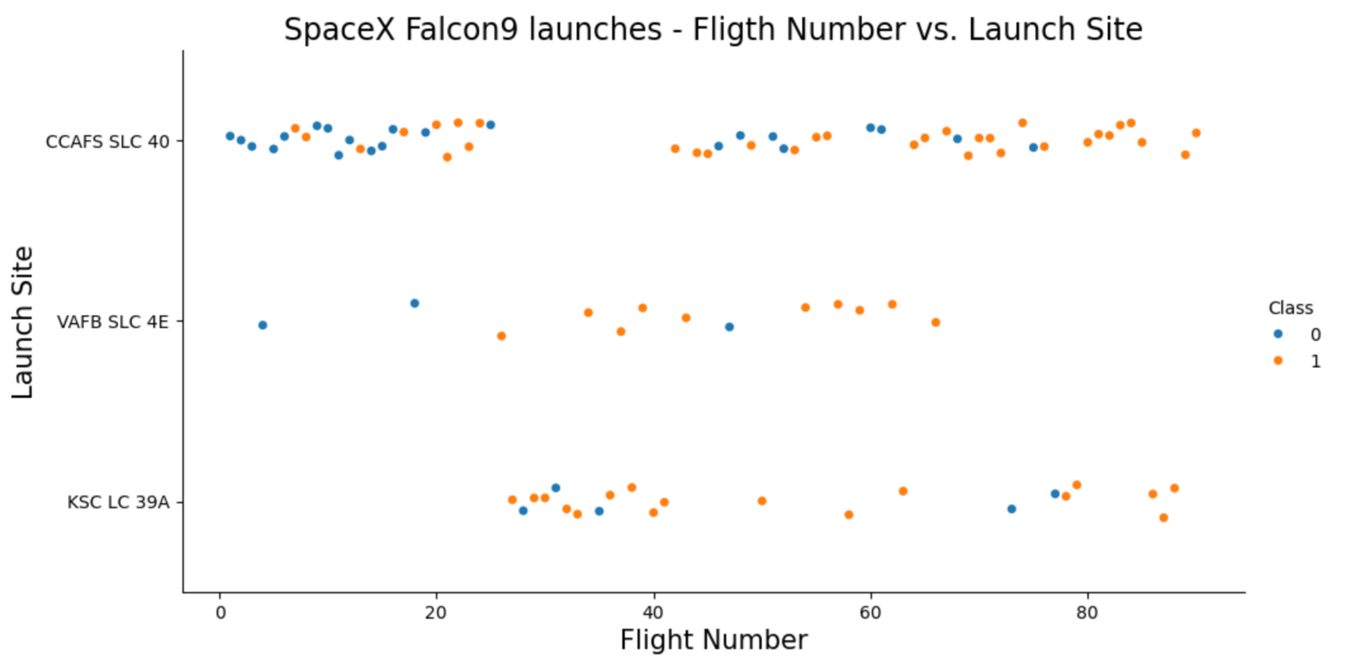
Results

- Exploratory data analysis results → Section 2
- Interactive analytics demo in screenshots → Sections 3 and 4
- Predictive analysis results → Section 5

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

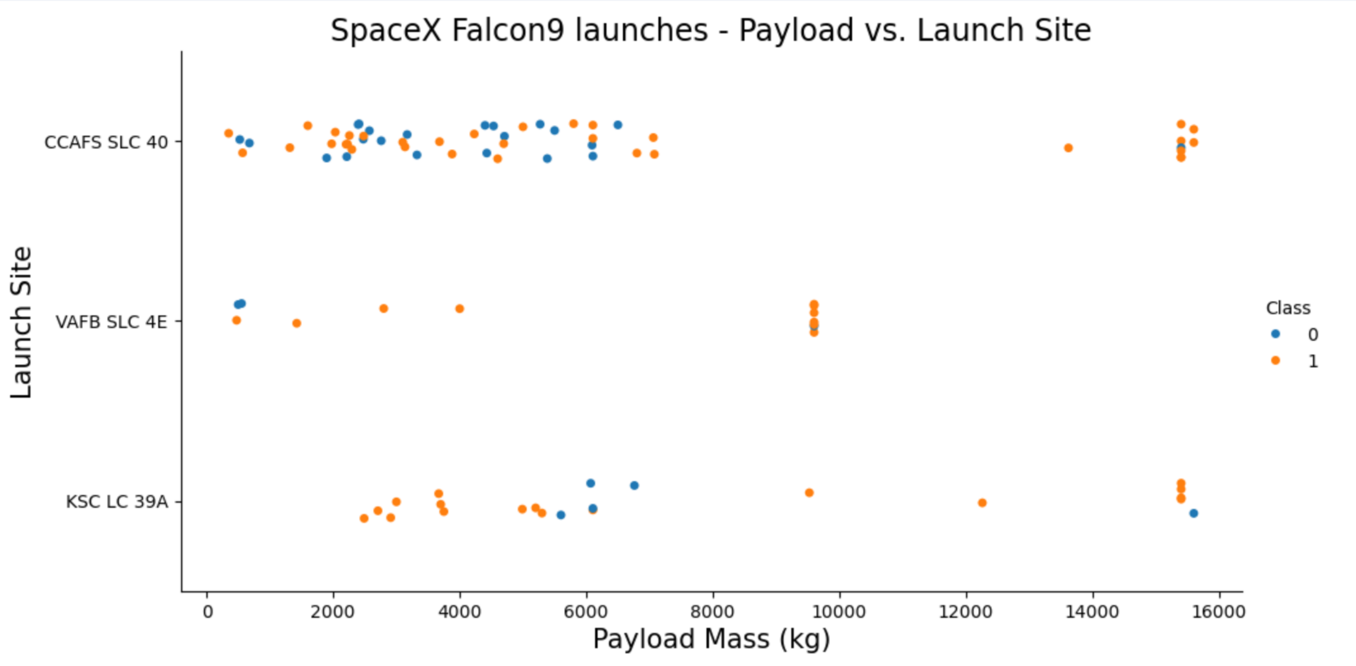


Class	Legend marker	Outcome
Class 0	● blue marker	Failure
Class 1	● orange marker	Success

Findings:

- Success rate vaires with launch site.
- As Flight Number increases, the Success rate also increases for all 3 sites.
- After a specific Flight Number (50, for VAFB SLC 4E site, and 80 for the other 2 sites), success rate was 100%.

Payload vs. Launch Site

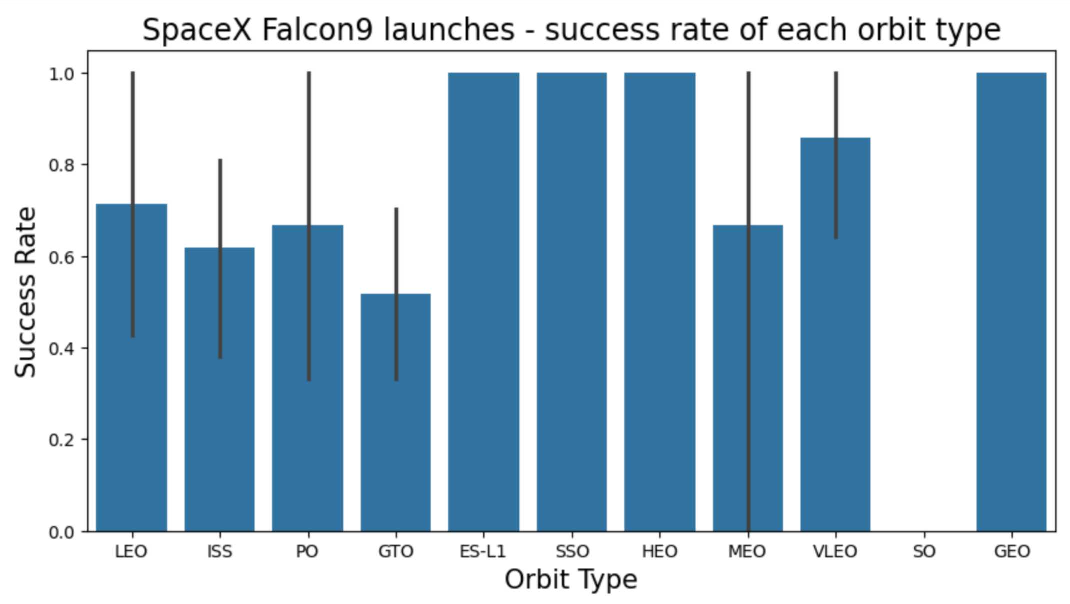


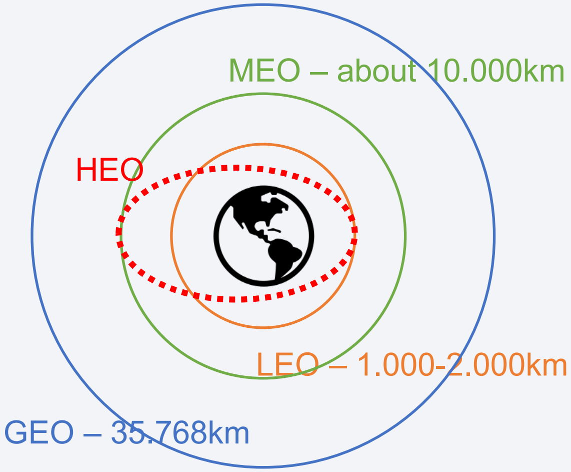
Class	Legend marker	Outcome
Class 0	● blue marker	Failure
Class 1	● orange marker	Success

Findings:

- For VAFB-SLC Launch Site, there era no rockets launched for heavy payload mass (no payload mass greater than 10.000kg).
- For Launch Site CCAF5 SLC 40, for Payload Mass smaller than 8.000kg, the success and failures are distributed. But for heavy payload mass (greater than 12.000kg), the success outcome are majority.
- Visually, it looks like the majority of failures for Launch Site KSC LC 39A happen with Payload Mass around 6.000kg.

Success Rate vs. Orbit Type



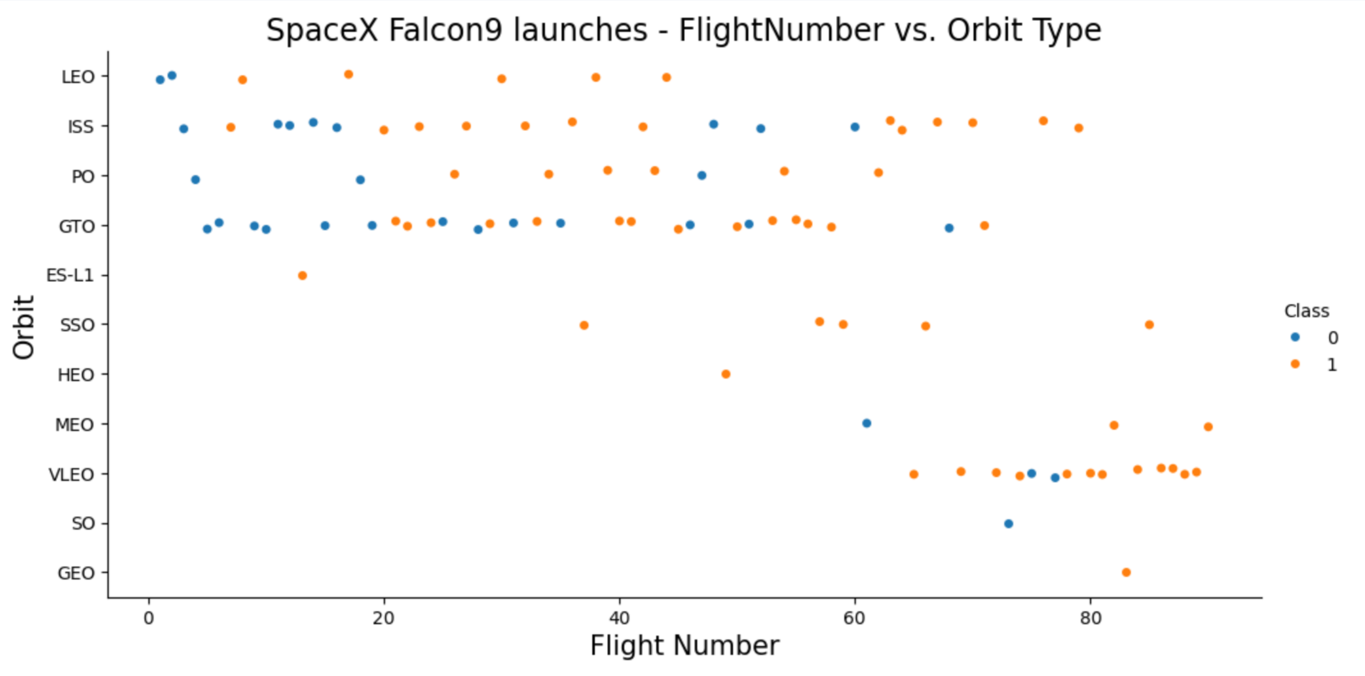


Findings:

- SO (Sub-Orbitital) has no successful first state landing.
- Orbits ES-L1, SSO, HEO and GEO had successful landing for all launches.

Orbit	Name	Outcome
LEO	Low Earth Orbit	Earth-centred orbit with altitude of 2.000km
VLEO	Very Low Earth Orbit	Mean altitude belo 450km
GTO	Geosynchronous Orbit	High Earth orbit that allows satellites to match Earth's rotation (located at 35.786km, above Earth's equator).
SSO	Sun-synchronous Orbit	Also called Heliosynchronous orbit, nearly polar orbit around the planet.
ES-L1	Lagrange points	At the Lagrange points the gravitational forces of the two large bodies cancel out in such a way that a small object placed in orbit there is in equilibrium relative to the center of mass of the large bodies.
HEO	Highly Elliptical Orbit	Elliptic orbit with high eccentricity
ISS	Modular Space Station in LEO	Multinational collaborative Project – International Space Station
MEO	Geocentric Orbit	Altitude from 2.000km to just below GTO. Also known as Intermediate Circular Orbit
GEO	Circular GTO	Above Equator, following direction of Earth's rotation
PO	Type of satellites	Type of satellite that passes above or nealy above both poles.
SO	Sub-Orbital	It reaches outer space, but its trajectory intersects the surface of the gravitating body.

Flight Number vs. Orbit Type

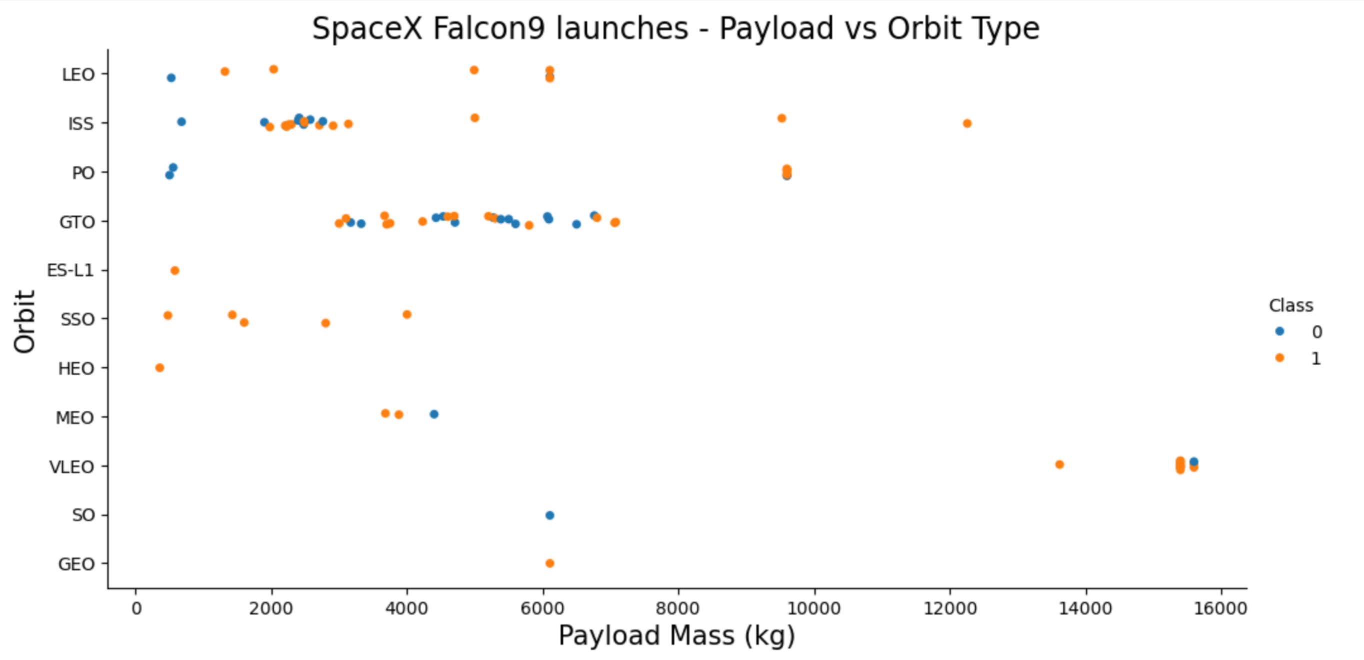


Class	Legend marker	Outcome
Class 0	● blue marker	Failure
Class 1	● orange marker	Success

Findings:

- Visually, can be observed that:
    - For LEO Orbit, the success seems to be related to the Number of Flights.
    - For GTO Orbit, it seems that there is no relationship between the number of flights and the orbit.
- It looks like there is a correlation between Fligth Number and Success Rate with larger flights number being associated with higher success rates.

Payload vs. Orbit Type

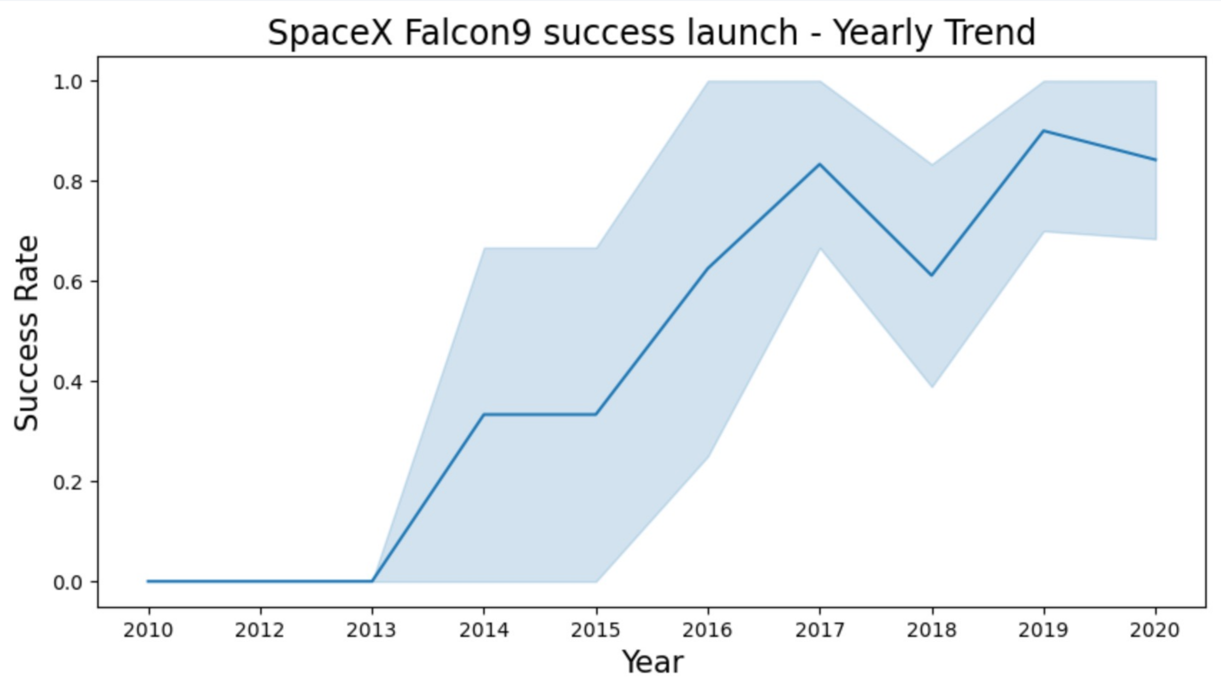


Class	Legend marker	Outcome
Class 0	● blue marker	Failure
Class 1	● orange marker	Success

Findings:

- For Orbits PO, LEO and ISS, successful landing (or positive landing) looks related to the heavier payloads.
- For GTO Orbit, it is not possible to identify impacts of payload (positive or negative).
- Considering all orbit types, it is not possible to establish if there is a correlation between Success Rate and Payload Mass

Launch Success Yearly Trend



Findings:

- Sucess Rate has increased significantly over the Years, starting on 2013.
- There was a small decrease in 2018. Unknown causes without deeper analysis.

All Launch Site Names

Launch Sites:

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

Query:

%sql SELECT DISTINCT LAUNCH_SITE FROM SPACEXTBL;

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

Findings:

- Understanding data, number of unique launch sites.

Launch Site Names Begin with 'CCA'

Launch Sites, with name starting with ‘CCA’:

Query:

```
%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

Findings:

- Undestanding data: quick view of data related to a single Launch_site, to obtain a better understanding it.

Total Payload Mass

Total Payload carreried by boosters from NASA:

Query:

```
%sql SELECT SUM(PAYLOAD_MASS__KG_) \
    AS 'Total Payload Mass (kg) – NASA' \
    FROM SPACEXTBL \
    WHERE CUSTOMER LIKE '%NASA (CRS)% ';
```

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

Total Payload Mass (kg) - NASA
48213

```
%sql SELECT SUM(PAYLOAD_MASS__KG_) \
    FROM SPACEXTBL \
    WHERE CUSTOMER = 'NASA (CRS)';
```

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

SUM(PAYLOAD_MASS__KG_)
45596

Auxiliary Query:

```
%sql SELECT DISTINCT CUSTOMER \
    FROM SPACEXTBL \
    WHERE CUSTOMER LIKE '%NASA (CRS)%';
```

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

Customer
NASA (CRS)
NASA (CRS), Kacific 1

Findings:

- Quick veiw of total payload carried by boosters from NASA – 48.213kg.
- There are 2 diferent entries containing "NASA" for Customer, and depending on how the SQL query is written, it may or may not capture all the information. This also leads to the question: is the "NASA (CRS), Kacific 1" an incorrect typing, or a diferente customer to be considered.

Average Payload Mass by F9 v1.1

Average Payload mass carried by booster version F9 v1.1:

Query:

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

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

Average Payload Mass carried by booster F9 v1.1
2928.4

```
%sql SELECT COUNT(DISTINCT BOOSTER_VERSION) \
     FROM SPACEXTBL;
```

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

COUNT(DISTINCT BOOSTER_VERSION)
97

Findings:

- Average payload mass carried by booster version F9 v1.1 is 2.928,4 kg.
- There are 97 Booster Versions in the dataset. Knowing the average payload mass carried by each booster type may help analysis if this information affects the launches and successful/failure outcome.

First Successful Ground Landing Date

Dates When the first successful landing outcome in ground pad was achieved:

Query:

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

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

MIN(DATE)
2015-12-22

Findings:

- The first successful landing outcome on ground pad happend on December 22nd, 2015.

Successful Drone Ship Landing with Payload between 4000 and 6000

Names of booster which have successfully landed on drone ship and has payload mass between 4000 and 6000kg

Query:

```
%sql SELECT BOOSTER_VERSION \
    FROM SPACEXTBL \
    WHERE LANDING_OUTCOME = 'Success (drone ship)' \
    AND PAYLOAD_MASS__KG_ > 4000 AND PAYLOAD_MASS__KG_ < 6000;
```

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

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

Findings:

- 4 (four) booster versions have successfully landed on drone ship, having payload mass between 4.000kg and 6.000kg.

Total Number of Successful and Failure Mission Outcomes

Number of total successful and failure mission outcomes:

Query:

```
%sql SELECT MISSION_OUTCOME, COUNT(*) as total_number \
    FROM SPACEXTBL \
    GROUP BY MISSION_OUTCOME;
```

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

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

Findings:

- There are 4 distinct mission outcome status – success outcome has 3 possibilities, which should lead to a deeper understanding if there is a typo and/or if they can be grouped together in a single "success" status.

```
%sql SELECT DISTINCT MISSION_OUTCOME \
    FROM SPACEXTBL
```

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

Mission_Outcome
Success
Failure (in flight)
Success (payload status unclear)
Success

- 100 launches had successful outcome, against 1 failure.

Boosters Carried Maximum Payload

Booster Version that carried the maximum payload mass:

Findings:

- 12 diferente Booster Versions carried the maximum payload mass of 15.600kg.

```
%sql SELECT COUNT(DISTINCT BOOSTER_VERSION) \
     FROM SPACEXTBL \
     WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL);
```

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

COUNT(DISTINCT BOOSTER_VERSION)
12

Query:

```
%sql SELECT DISTINCT BOOSTER_VERSION, PAYLOAD_MASS__KG_ \
    FROM SPACEXTBL \
    WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL););
```

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

Booster_Version	PAYLOAD_MASS__KG_
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

2015 Launch Records

Failed Landing_Outcomes in drone ship, including booster versions and launch site names in 2015:

Query:

```
%sql SELECT substr(Date, 6, 2) as month, DATE, BOOSTER_VERSION, LAUNCH_SITE, Landing_Outcome \
    FROM SPACEXTBL \
    WHERE Landing_Outcome = 'Failure (drone ship)' AND SUBSTR(Date, 0, 5) = '2015';
```

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

month	Date	Booster_Version	Launch_Site	Landing_Outcome
01	2015-01-10	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
04	2015-04-14	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

Findings:

- There were 2 failed landing outcomes with a drone ship in 2015, both launched from Site CCAFS LC-40, one in January, and the other in April.

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

Ranking of the landing outcomes Count between 04-Jun-2010 and 20-Mar-2017, in descending order:

Query:

```
%sql SELECT Landing_Outcome, COUNT(*) as COUNT_OUTCOMES \
    FROM SPACEXTBL \
    WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20' \
    GROUP BY Landing_Outcome \
    ORDER BY COUNT_OUTCOMES DESC;
```

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

Landing_Outcome	COUNT_OUTCOMES
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

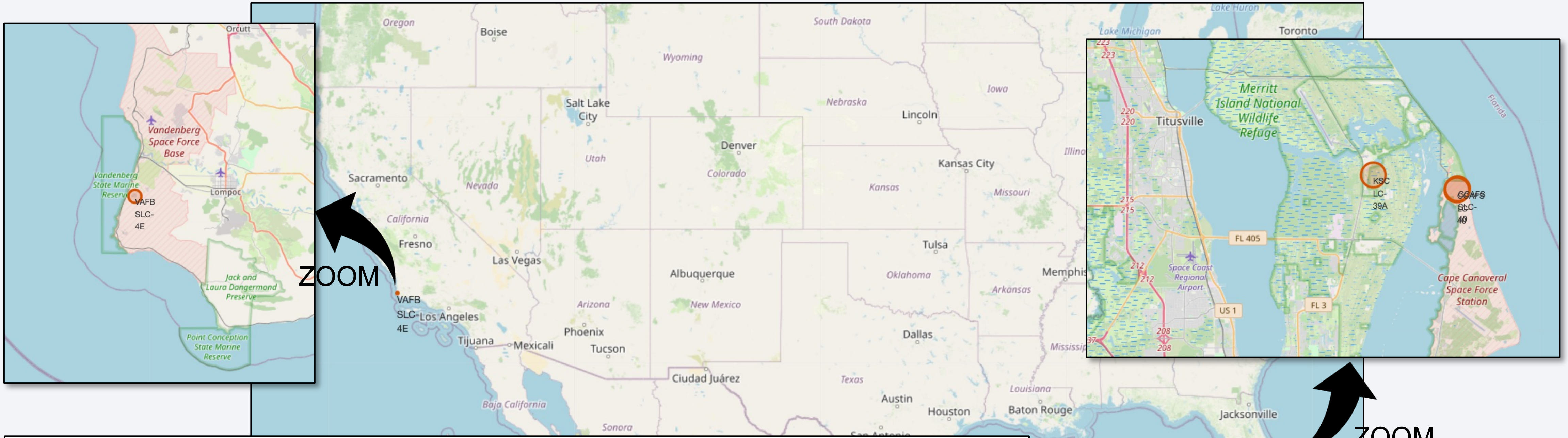
Findings:

- The most common landing outcome was "No Attempt.

Section 3

Launch Sites Proximities Analysis

Falcon9 Launch Site locations



Findings:

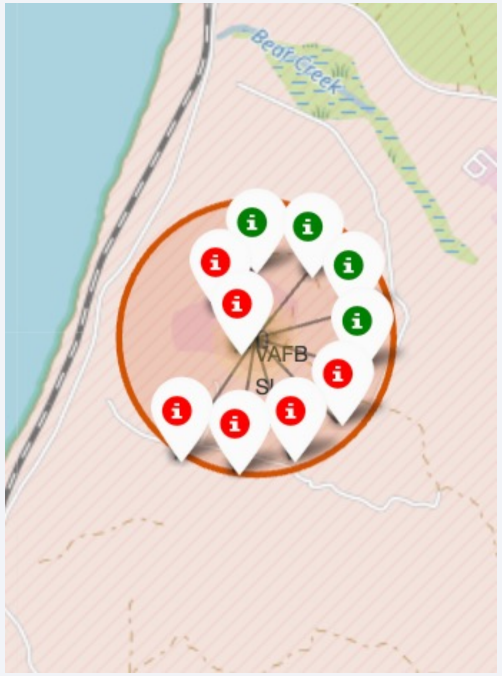
- All launch sites are in low latitude, closest to Equator. This may be explained that rockets can most easily reash satellite orbits if launched near the Equator in an easterly direction (as this maximizes use of Earth's rotational speed) – this also provides a desirable orientation for arriving at GEO (source: https://en.wikipedia.org/wiki/Spaceport#:~:text=vertical%20satellite%20launch.-,Location,arriving%20at%20a%20geostationary%20orbit.) So, the closer to the Equator, the greater boost it gets from rotation of the Earth.
- Now, the site location close to the coast can be explained be security: if anything goes wrong, there is the whole ocean to crash in.

	Launch Site	Lat	Long
0	CCAFS LC-40	28.562302	-80.577356
1	CCAFS SLC-40	28.563197	-80.576820
2	KSC LC-39A	28.573255	-80.646895
3	VAFB SLC-4E	34.632834	-120.610745

Success and Failed launches in each site

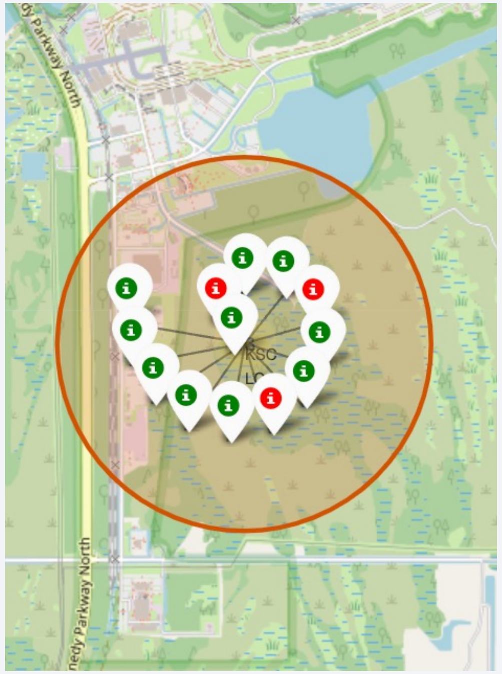
WEST COAST

VAFB SLC-4E

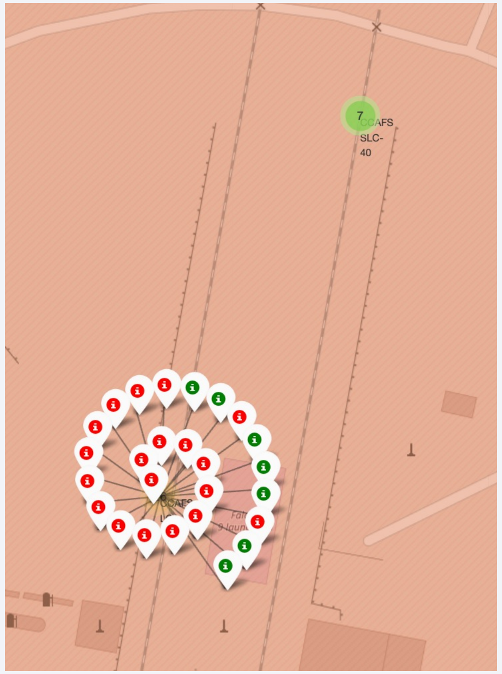


EAST COAST

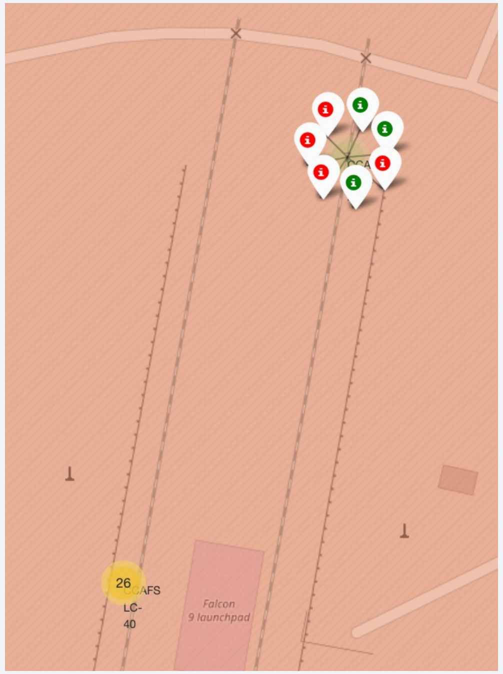
KSC LC-39A



CCAFS LC-40



CCAFS SLC-40



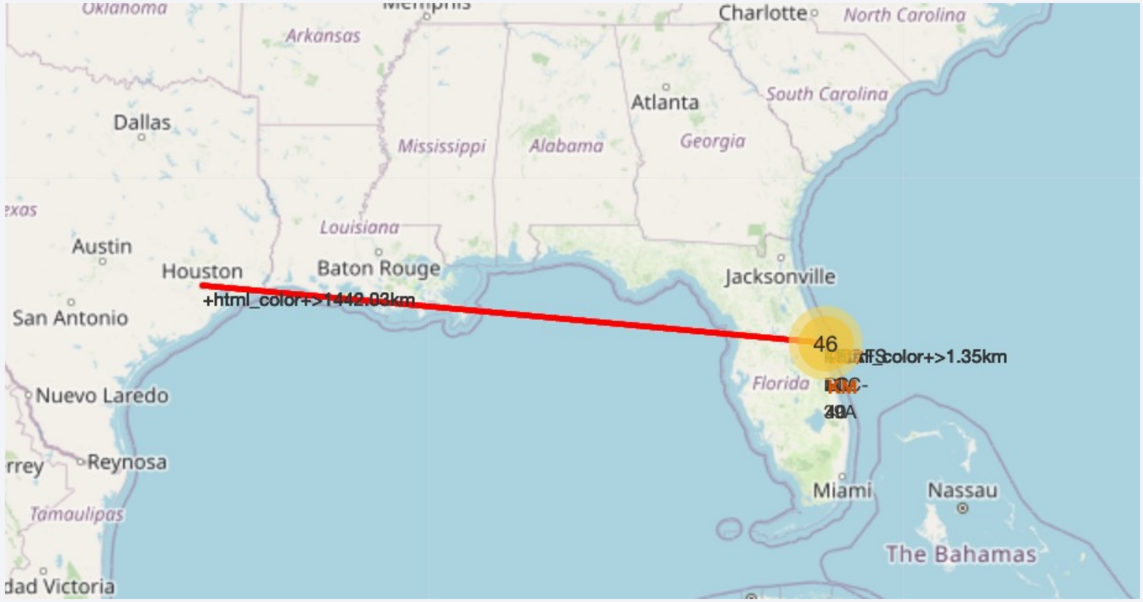
Green Marks: Successful launches

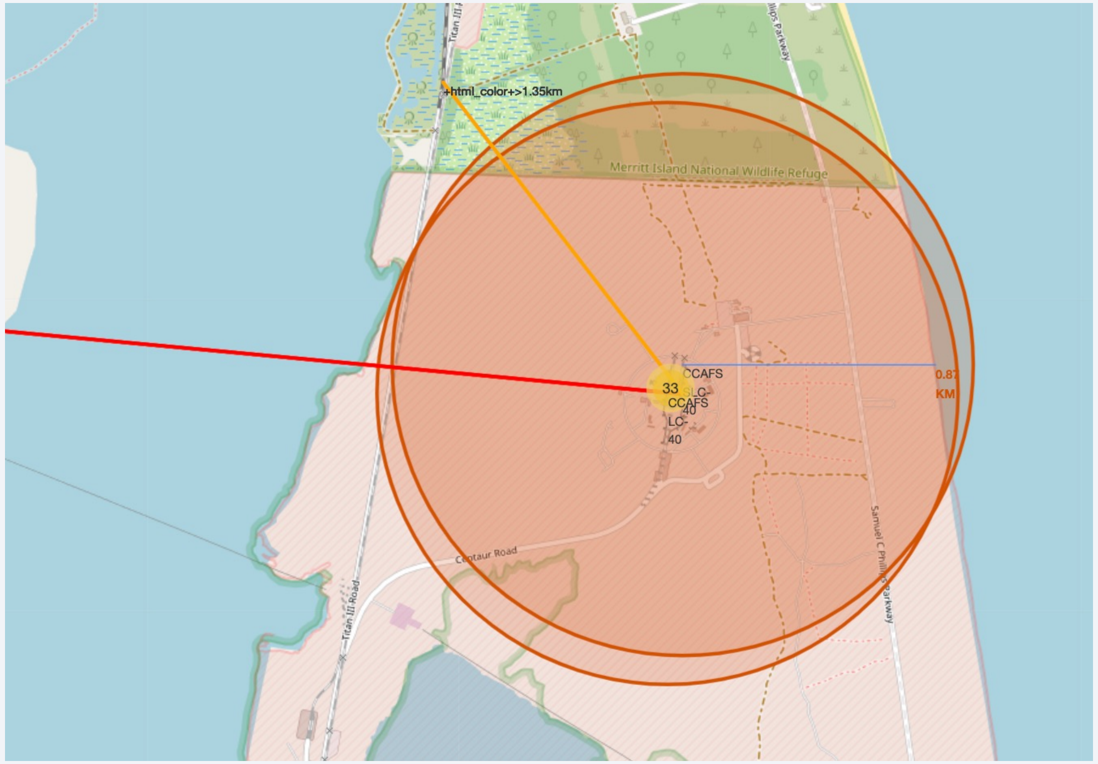
Red Marks: Failed launches

Findings:

- These views and analysis allow a quick understanding of success rate for each site When launching the rockets. While it does not explain the why some sites have more failures than others, it allows this understanding and provide the basis for deeper furture checks.

Distance from Launch Site to other geographical marks



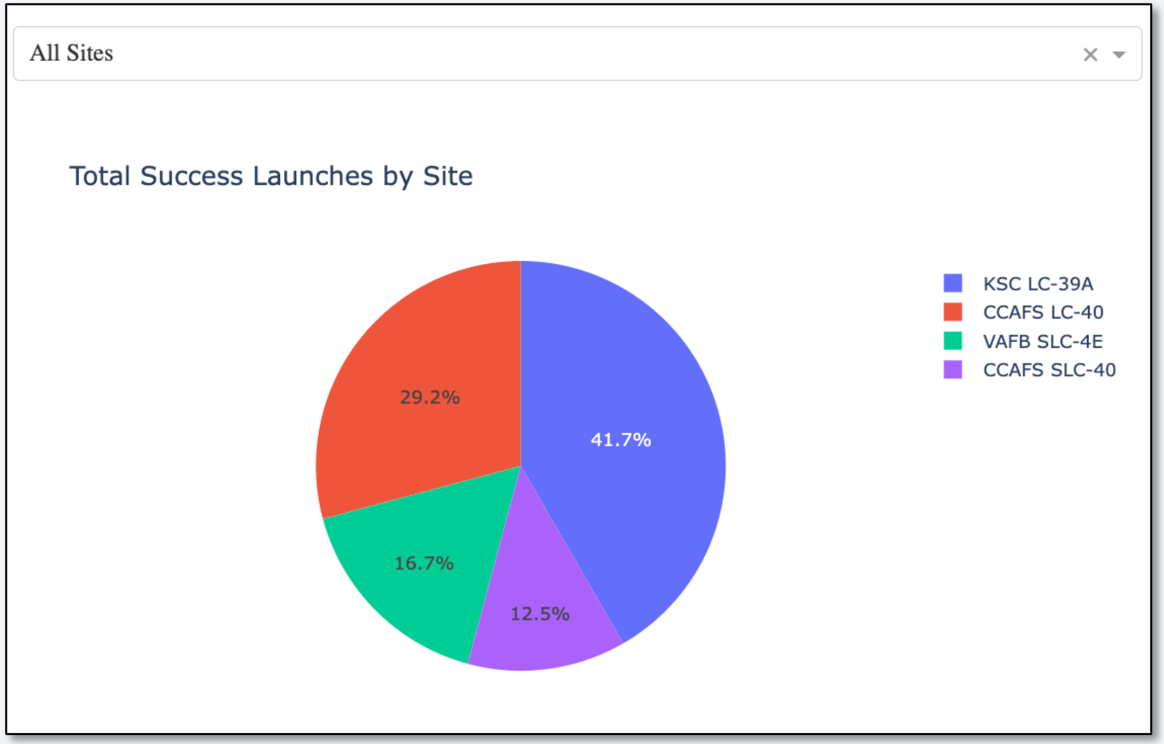


- A Folium (dynamic) map also allows some measurement checks and marks between geographical locations.
- From Launch Site CCAFS SLC-40 to the coastline (Atlantic), there is 0.87km.
- From Launch Site CCAFS LC-40 to the point where Titan III Road meets the NASA Railroad, there is 1,35km. Still, from this site (in Florida) to Houston (TX), the distance is about 1.442km.

Section 4

Build a Dashboard with Plotly Dash

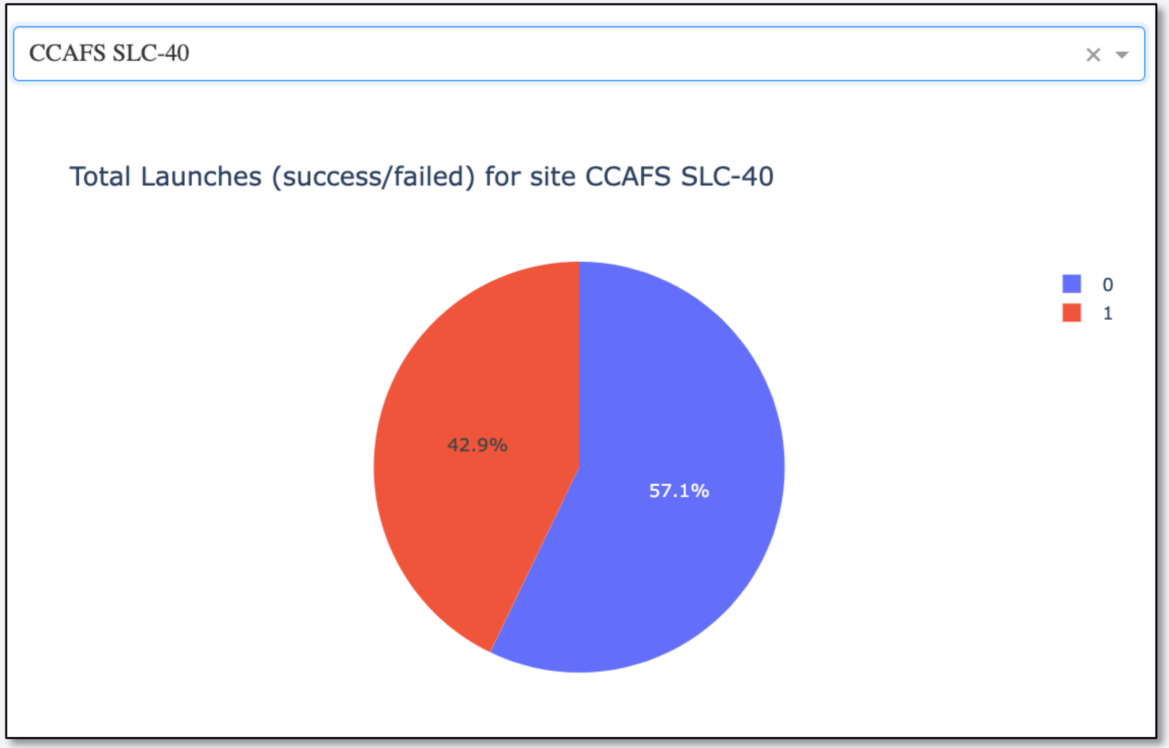
Success Launches per site, for all sites



FINDINGS & COMMENTS:

- A dropdown menu, by default, showing “All Site”, select the data for the Pie Chart below.
- With this selection, the pie chart displays the distribution of successful launces per site.
- It allows a quick view that:
  - Greatest success rate is for Site KSC LC-39A, with 41,7%;
  - Lowest success rate is for Site CCAFS SLC-40, with 12,5%.

Launch site with highest launch success ratio

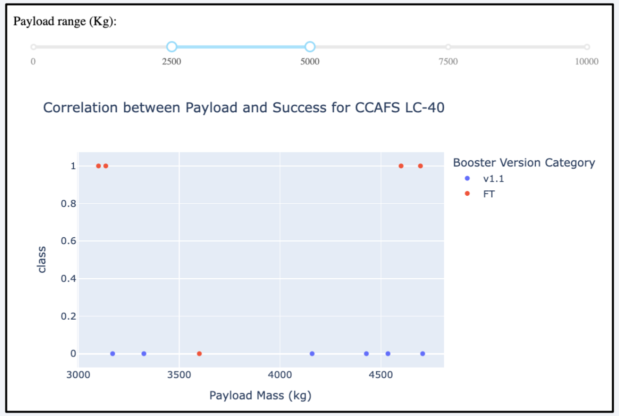


FINDINGS & COMMENTS:

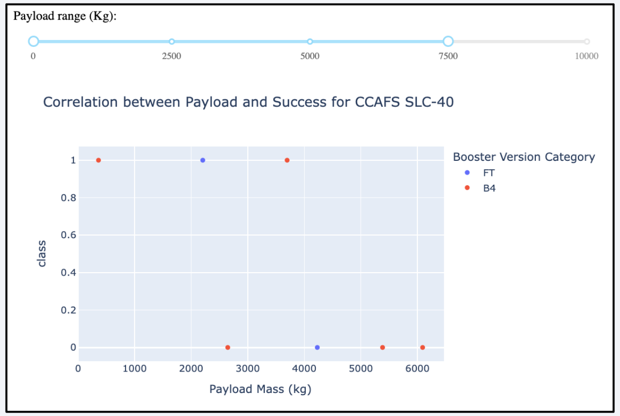
- The site with highest launch success ratio is CCAFS SLC-40. 42,9% launches were successful (**in red**), against 57,1% failed (**in blue**).

Payload vs Launch Outcome

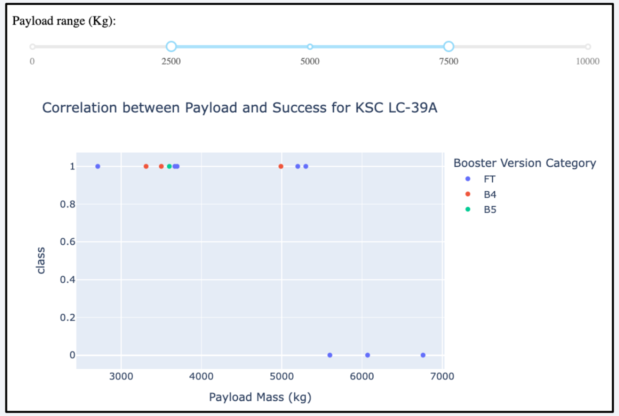
CCAFS LC-40



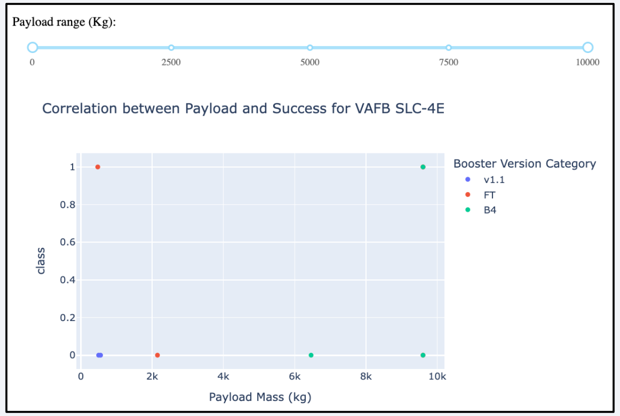
CCAFS SLC-40



KSC LC-39A



VAFB SLC-4E



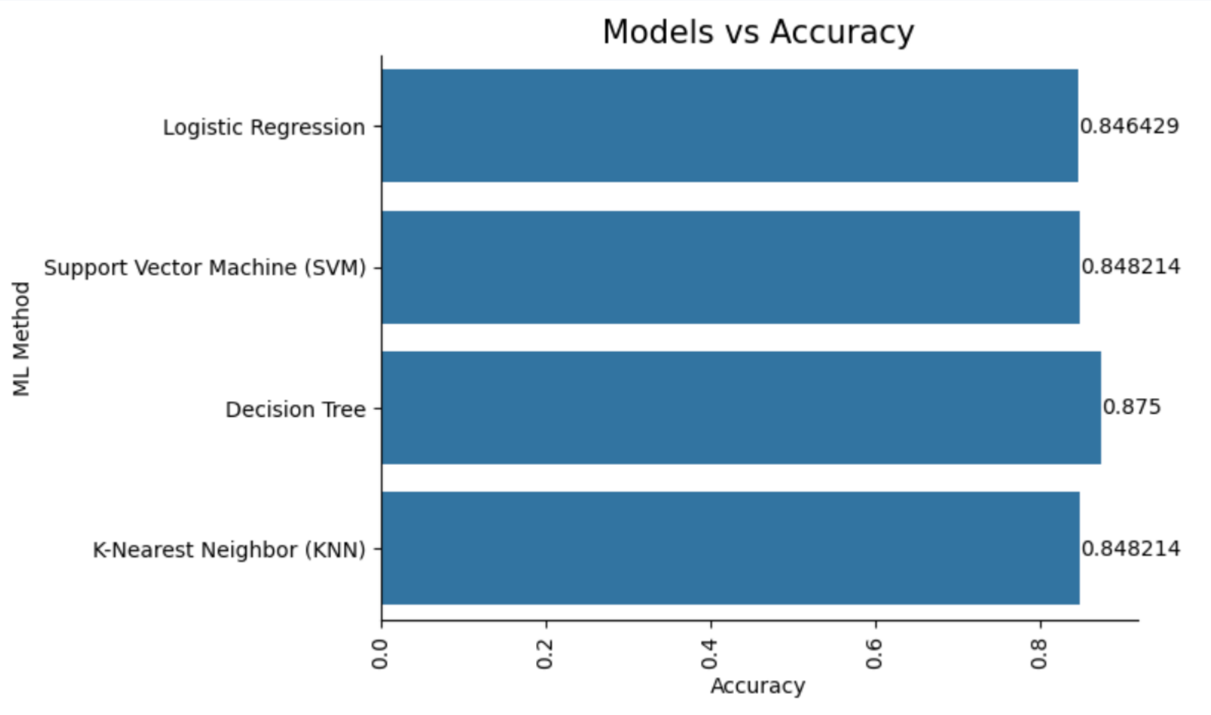
FINDINGS & COMMENTS:

- Screenshots of Payload vs Launch outcome, for different payload selection.
- Payload range between aprox. 2.000kg and 5.000kg has the largest success rate.
- Booster Version FT has the largest success rate.

Section 5

Predictive Analysis (Classification)

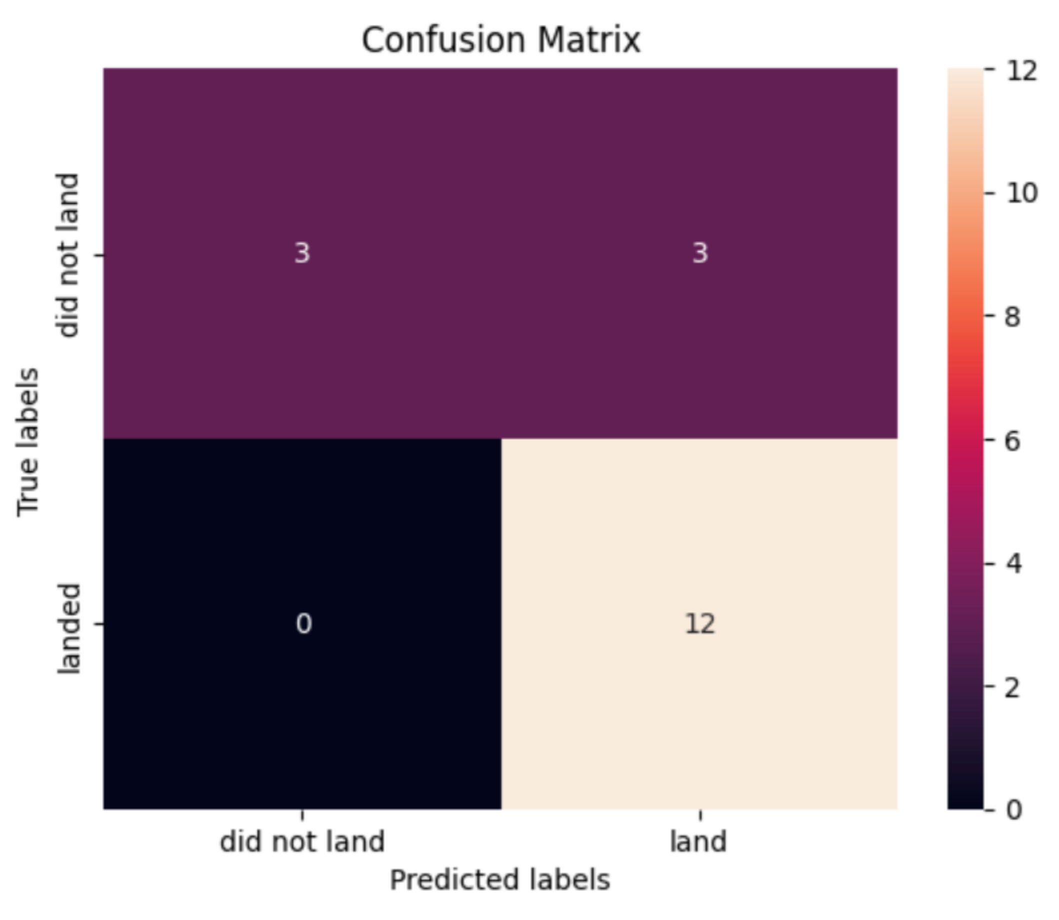
Classification Accuracy



- All methods performed equally (84,64%), but Decision Tree- that had Accuracy of 87,5% on train data.
- On the test data, all methods presented similar accuracy, using method score: 0,83333.
- Conclusion: all the algorithms tested give the same result, same accuracy for this dataset.

	ML Method	Accuracy	Score	Parameters
0	Logistic Regression	0.846429	0.833333	{'C': 0.01, 'penalty': 'l2', 'solver': 'lbfgs'}
1	Support Vector Machine (SVM)	0.848214	0.833333	{'C': 1.0, 'gamma': 0.03162277660168379, 'kernel': 'sigmoid'}
2	Decision Tree	0.875000	0.833333	{'criterion': 'gini', 'max_depth': 6, 'max_features': 'sqrt', 'min_samples_leaf': 4, 'min_samples_split': 5, 'splitter': 'random'}
3	K-Nearest Neighbor (KNN)	0.848214	0.833333	{'algorithm': 'auto', 'n_neighbors': 10, 'p': 1}

Confusion Matrix



- Confusion Matrix for the Decision Tree model, which is the one that had best accuracy on train data: 0,875.
- Confusion Matrix details:

True Negative	False Positive
False Negative	True Positive

  - 12 True Positives and 3 True Negatives;
  - 3 False Positives and 0 False Negatives.
- The 3 False Positive mapped is the major problem.

Conclusions

- The trend shows that as the flight number increases, so does the success rate
- Different launch sites have different success rates.
- Orbits ES-L1, SSO, HEO and GEO have the highest success hate, at 100%, while SO orbit sucess rate was the lower, at 0%.
- With heavy payloads, the successful landing are for orbits Polar, LEO and ISS. But, for GTO orbit, it is not possible to distinguished clearly, as both positive and negative landing are both there equally distributed.
- Success rate is increasing since 2013 (tilll 2020, last year of the data analyzed).
- Prediction models studied (Logistic Regression, SVM, Decision Trees and k-NN) have the same Accuracy (method Score) for test set. For the train set, the Decision Tree model presented better Accuracy.
- The Correlation Matrix for all methods presented similar results, with 3 False Positives, which presentes a point of concern and possible improvements for the models.

Thank you!