IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

José Ramírez Cabello
December - 2023

Outline

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

Executive Summary

Determining the price of each launch, predicting the reuse of the first phase of each rocket, and communicating the results obtained require the execution of various steps and processes

- Summary of methodologies
  - Data collection from 2 sources available on the web
  - Exploration and analysis (via SQL)
  - Statistical analysis, data visualization, and categorization of variables for subsequent processes
  - Construction and evaluation of predictive models
  - Prepare report and final presentation of the project
- Summary of all results
  - Organized and structured data to meet requirements
  - Dashboard for review and interactive crossing of variables
  - Analysis and review of predictive model results
  - Favorable applicability of a model to solve the proposed problem

Introduction

Project background and context

- The **Space Y** company is planning to join the commercial space services market.
- The costs of launching rockets into space vary from USD$62M to more than USD$165M.
- The most successful company on the market (Space X) has a two-stage rocket (Falcon 9). The first of them is reusable, with a significant cost reduction.
- Space X can sacrifice the first stage rockets depending on some mission parameters (payload, orbit, client). On some occasions due to failures it has not been possible to recover these stage.

Problems you want to find answers

- The task entrusted to data scientists is to be able **to determine prices for rocket launches** by **predicting the reuse of the first stage of rockets** in future launches.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Base information obtained from two different sources:
    - Obtain rocket and launch information in .json format from Space X API REST (IBM Developer version)
    - https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/API_call_spacex_api.json
    - Collect historical records of Falcon 9 launches from a Wikipedia page titled "List of Falcon 9 and Falcon Heavy launches“, updated on 9th June 2021
    - https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922
- Perform data wrangling
  - Rescue, cleaning and filtering of valuable information available on the web, in html format tables, from falcon 9 rockets, using BeautifulSoup library tools.

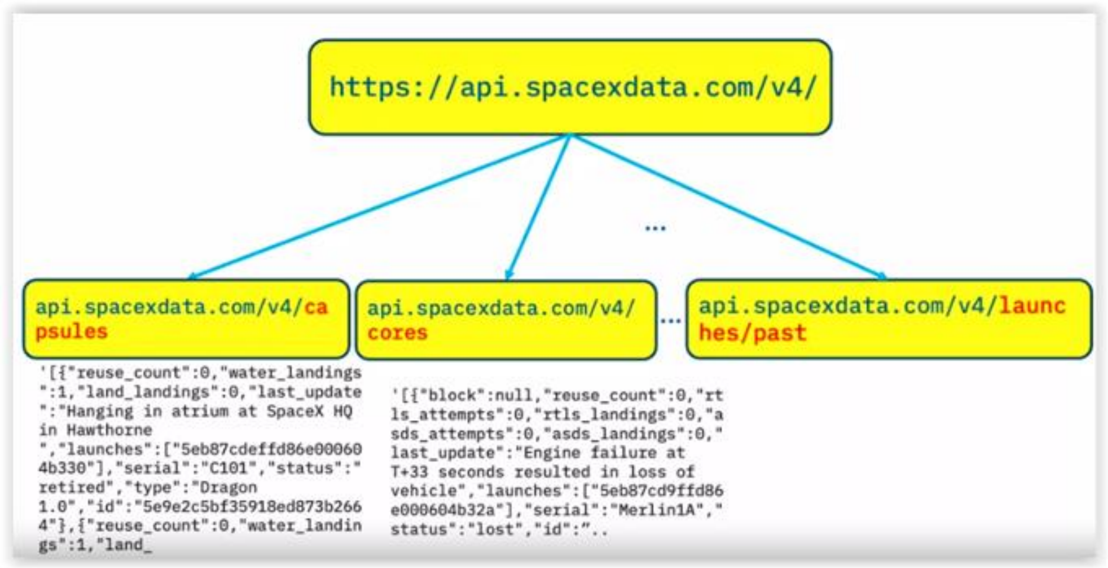
Methodology

Executive Summary

- Perform exploratory data analysis (EDA) using visualization and SQL
  - It involves loading .csv data file into database for queries via SQL.
  - Exploratory analysis of information to evaluate automatic determination of the effective landing rocket first stage.
- Perform interactive visual analytics using Folium and Plotly Dash
  - Detect patterns and perform proximity analysis to choose the optimal launch site.
- Perform predictive analysis using classification models
  - Perform exploratory analysis and structure training data to create a machine learning system to predict whether the first phase of the Falcon 9 rocket will have a successful landing.
  - Used models: Logistic Regression (**lr**), support vector machine (**svm**), Regression tree (**tree**), K-Nearest Neighbors (**KNN**)

Data Collection

- Two sources of input information:
  - Space X REST API with various rocket launch information, obtained in .json format and converted to a data frame for analysis.



  - Processing via scraping the Wiki page "**Falcon 9 and Falcon Heavy launch list**" updated on "June 9, 2021", to create data structure from tables on website.

[hide] Flight No.	Date and time (UTC)	Version, Booster[b]	Launch site	Payload[c]	Payload mass	Orbit	Customer	Launch outcome	Booster landing
78	7 January 2020, 02:19:21[492]	F9 B5 ♺ B1049.4	CCAFS, SLC-40	Starlink 2 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Success (drone ship)
	Third large batch and second operational flight of Starlink constellation. One of the 60 satellites included a test coating to make the satellite less reflective, and thus less likely to interfere with ground-based astronomical observations.[493]								
79	19 January 2020, 15:30[494]	F9 B5 ♺ B1046.4	KSC, LC-39A	Crew Dragon in-flight abort test[495] (Dragon C205.1)	12,050 kg (26,570 lb)	Sub-orbital[496]	NASA (CTS)[497]	Success	No attempt
	An atmospheric test of the Dragon 2 abort system after Max Q. The capsule fired its SuperDraco engines, reached an apogee of 40 km (25 mi), deployed parachutes after reentry, and splashed down in the ocean 31 km (19 mi) downrange from the launch site. The test was previously slated to be accomplished with the Crew Dragon Demo-1 capsule;[498] but that test article exploded during a ground test of SuperDraco engines on 20 April 2019.[419] The abort test used the capsule originally intended for the first crewed flight.[499] As expected, the booster was destroyed by aerodynamic forces after the capsule aborted.[500] First flight of a Falcon 9 with only one functional stage — the second stage had a mass simulator in place of its engine.								
80	29 January 2020, 14:07[501]	F9 B5 ♺ B1051.3	CCAFS, SLC-40	Starlink 3 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Success (drone ship)
	Third operational and fourth large batch of Starlink satellites, deployed in a circular 290 km (180 mi) orbit. One of the fairing halves was caught, while the other was fished out of the ocean.[502]								
81	17 February 2020, 15:05[503]	F9 B5 ♺ B1056.4	CCAFS, SLC-40	Starlink 4 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Failure (drone ship)
	Fourth operational and fifth large batch of Starlink satellites. Used a new flight profile which deployed into a 212 km × 386 km (132 mi × 240 mi) elliptical orbit instead of launching into a circular orbit and firing the second stage engine twice. The first stage booster failed to land on the drone ship[504] due to incorrect wind data.[505] This was the first time a flight proven booster failed to land.								
82	7 March 2020, 04:50[506]	F9 B5 ♺ B1059.2	CCAFS, SLC-40	SpaceX CRS-20 (Dragon C112.3 ♺)	1,977 kg (4,359 lb)[507]	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
	Last launch of phase 1 of the CRS contract. Carries *Bartolomeo*, an ESA platform for hosting external payloads onto ISS.[508] Originally scheduled to launch on 2 March 2020, the launch date was pushed back due to a second stage engine failure. SpaceX decided to swap out the second stage instead of replacing the faulty part.[509] It was SpaceX's 50th successful landing of a first stage booster, the third flight of the Dragon C112 and the last launch of the cargo Dragon spacecraft.								
83	18 March 2020, 12:16[510]	F9 B5 ♺ B1048.5	KSC, LC-39A	Starlink 5 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Failure (drone ship)
	Fifth operational launch of Starlink satellites. It was the first time a first stage booster flew for a fifth time and the second time the fairings were reused (Starlink flight in May 2019).[511] Towards the end of the first stage burn, the booster suffered premature shut down of an engine, the first of a Merlin 1D variant and first since the CRS-1 mission in October 2012. However, the payload still reached the targeted orbit.[512] This was the second Starlink launch booster landing failure in a row, later revealed to be caused by residual cleaning fluid trapped inside a sensor.[513]								
84	22 April 2020, 19:30[514]	F9 B5 ♺ B1051.4	KSC, LC-39A	Starlink 6 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Success (drone ship)

Data Collection – SpaceX API

SpaceX API REST launch data from launches/past

- Request: using .json format
- Parse: Create DataFrame using json_normalize

Extract information via GET Request from /cores

- Create a new dataframe adding data on rockets used, landings made, landing locations and conditions.
- API Get functions: getBoosterVersion, getLaunchSite, getPayloadData, getCoreData

Filtering process

- We exclude records of the Falcon 1 rocket
- data_falcon9 = data2[data2['BoosterVersion']!='Falcon 1']

Fixing some null values

- We replace null values of PayloadMass with the mean of that attribute
- mean = data_falcon9["PayloadMass"].mean()
- data_falcon9["PayloadMass"] = data_falcon9["PayloadMass"].replace(np.nan, mean)

- Source code: https://github.com/agrofacilchile/SpaceX_capstone_project_jrrc/blob/main/jupyter-labs-spacex-data-collection-api_lab_jrrc.ipynb

Data Collection – Scraping

Extract a Falcon 9 launch records HTML table from Wikipedia

- Source: https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922
- Python library: BeautifulSoap.

Parse the table and convert it into a Pandas data frame

- Request the Falcon9 Launch Wiki page from previous URL
- Extract all column/variable names from the HTML table header.
- Create a data frame by parsing the launch HTML tables

Save DataFrame into .cvs file

- Export DataFrame to file in .cvs format, for use in subsequent processes.

- Source code: https://github.com/agrofacilchile/SpaceX_capstone_project_jrrc/blob/main/jupyter-labs-webscraping_lab_jrrc.ipynb

Data Wrangling

Perform exploratory Data Analysis and determine Training Labels	• Source: https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/dataset_part_1.csv • Python library: Pandas and Numpy
Data Analysis	• Calculate the 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.
Save DataFrame into .cvs file	• Export DataFrame to file in .cvs format, for use in subsequent processes.

- Source code: https://github.com/agrofacilchile/SpaceX_capstone_project_jrrc/blob/main/labs-jupyter-spacex-Data%20wrangling_lab_jrrc.ipynb

EDA with Data Visualization

Type Chart	Variables	Use descriptions
Scatter point	FlightNumber vs Payload	See how the FlightNumber and Payload variables would affect the launch outcome.
Scatter point	FlightNumber vs LaunchSite	It allows you to relate the number of launches per site and the success rate.
Scatter point	LaunchSite vs Payload	We can observe if there is any relationship between launch sites and their payload mass.
Bar chart	Orbit vs Class	We can visually check if there are any relationship between success rate and orbit type.
Scatter point	Orbit vs FlightNumber	For each orbit, we want to see if there is any relationship between FlightNumber and Orbit type.
Scatter point	Orbit vs Payload	It allows identifying the relationship that may exist between payload and type of orbit.
Line chart	Year vs Average success rate	Allows you to obtain the average launch success trend, considering year and average launch success rate.

- Source code: https://github.com/agrofacilchile/SpaceX_capstone_project_jrrc/blob/main/jupyter-labs-eda-dataviz_lab_jrrc.ipynb.jupyterlite.ipynb

EDA with SQL

- ✓ **Understand the Spacex DataSet**
  - ✓ Dataset includes a record for each payload carried during a SpaceX mission into outer space.
- ✓ **Load the dataset into the corresponding table in a Db2 database**.
  - ✓ Connection with database using sqlite3 and SQL extension
- ✓ **Execute SQL queries to answer different questions:**
  - ✓ 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
  - ✓ Names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
  - ✓ Total number of successful and failure mission outcomes
  - ✓ Names of the booster_versions which have carried the maximum payload mass.
  - ✓ List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for months in year 2015.
  - ✓ Ranking the count of landing outcomes (such as Failure or Success) between the date 2010-06-04 and 2017-03-20.

- Source code: https://github.com/agrofacilchile/SpaceX_capstone_project_jrrc/blob/main/jupyter-labs-eda-sql-coursera_sqllite_lab_jrrc.ipynb

Build an Interactive Map with Folium

Geographical location of the launch point is important for several factors: **Ground rotation speed** (best at the equator), **Tangential takeoff speed** (favorable on the east coast of the USA) and the **Safety of the population** (near the coast with flight towards the ocean). The launch sites are analyzed considering these elements as success factors:

- ✓ **Mark all launch sites on a map**.
  - ✓ Identify the geographical position and determine its distance from the Equator.
  - ✓ Dots on the map identify the position of the launch pads.
- ✓ **Mark the success/failed launches for each site on the map**.
  - ✓ Circles and texts identifying the number of launches made on each platform..
  - ✓ Specific symbols identifying success/failure of each launch and platforms.
- ✓ **Calculate the distances between a launch site to its proximities**.
  - ✓ Measurement of distances on the map, for critical points identified by geographic coordinates:
    - ✓ Are launch sites in close proximity to railways?
    - ✓ Are launch sites in close proximity to highways?
    - ✓ Are launch sites in close proximity to coastline?
    - ✓ Do launch sites keep certain distance away from cities?

- Source code: https://github.com/agrofacilchile/SpaceX_capstone_project_jrrc/blob/main/lab_jupyter_launch_site_location.jupyterlite_lab_jrrc.ipynb

Build a Dashboard with Plotly Dash

Tool that in real time through **input components**, **drop-down lists** and **sliders** allows you to perform interactive visual analyzes that allow you to determine:

- ✓ Which site has the most successful releases?
- ✓ Which site has the highest launch success rate?
- ✓ Which payload range(s) has the highest launch success rate?
- ✓ Which payload range(s) has the lowest launch success rate?
- ✓ Which version of F9 Booster has the highest launch success rate?

Features incorporated into the dashboard:.

- ✓ launch site entry (selection/deployment) component
- ✓ representation via pie chart of launch success (according to drop-down menu of the selected site)
- ✓ slider using ranges to select payload
- ✓ successful payload scatterplot

- Source code: https://github.com/agrofacilchile/SpaceX_capstone_project_jrrc/blob/main/spacex_dash_app_jrrc.py

Predictive Analysis (Classification)

Process to standardize data, divide it into training and test groups, to train the models and through hyperparameters determine which ones work best.

The models are: **linear regression** (lr), **support vector machine** (SVM), **decision tree** (tree) and **k nearest neighbor** (KNN).

Process flowchart:

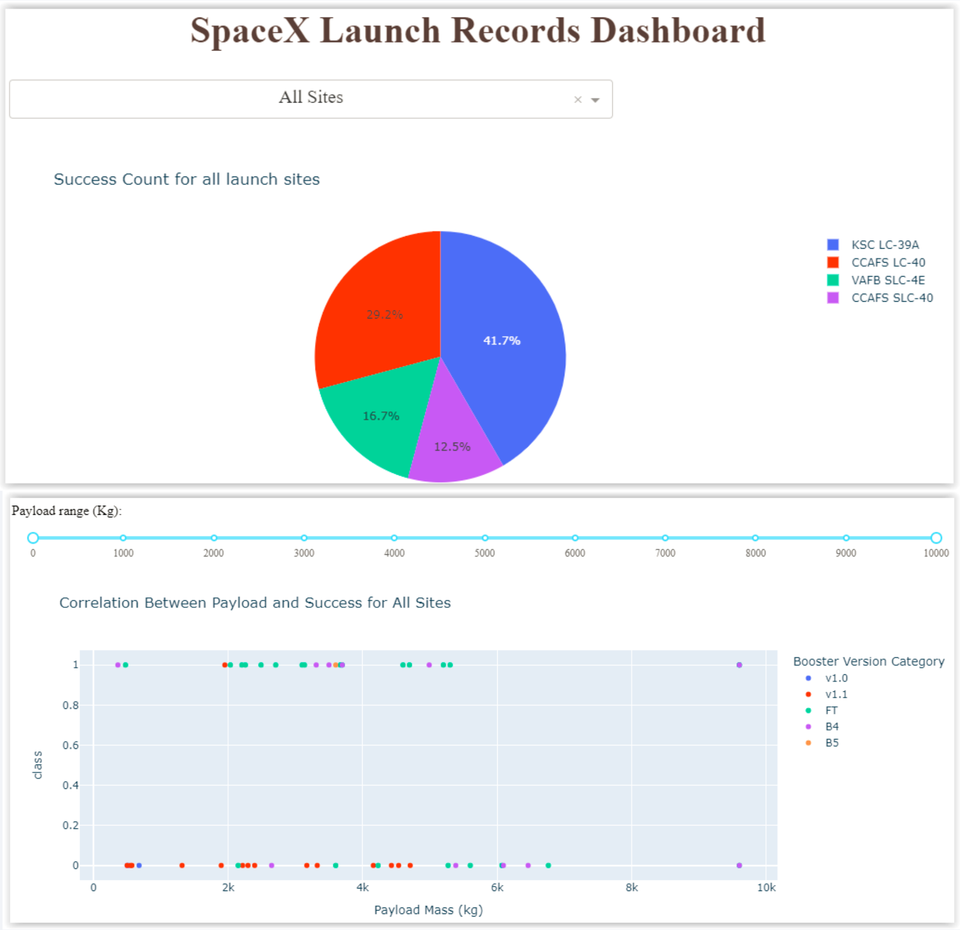
Load dataframe	• Two data file: dataset_part_2.csv - dataset_part_3.csv. • Create a NumPy array (**Class** column).
Standardize data	• Function to change vectors into a better representation for later estimators. • preprocessing.StandardScaler().fit(X).transform(X).
Split: training data and test data	• Split data: training data (80%) and test data (20%). • Function used: **train_test_split**, with test_size=0.2, and random_state=2).
Find the method performs best, using test data	• Evaluated method **byGridSearchCV** : lr, SVM, tree and KNN. • Criteria: validation accuracy (**best_score_**), accuracy on test data**(.score**), and plot confusion matrix(**plot_confusion_matrix**).

- Source code: https://github.com/agrofacilchile/SpaceX_capstone_project_jrrc/blob/main/SpaceX_Machine_Learning_Prediction_Part_5_jrrc_v3.jupyterlite.ipynb

Results

- **Exploratory data analysis results**
  - The success rate of rocket launches is improving over time.
  - By increasing the flight number, the landing success rate of the first stage also improves.
  - The KSC LC-39A platform shows the highest success rate among landing sites.
  - A good ratio of successful landings of the first stage is observed, with payload masses around 7000 kg.
  - As experience in rocket launches increases, the success rate of rocket launches also increases when planning to recover the first stage.
- **Predictive analysis results**
  - The **decision tree model (tree)** shows better quality parameters as a predictive model for the data set.

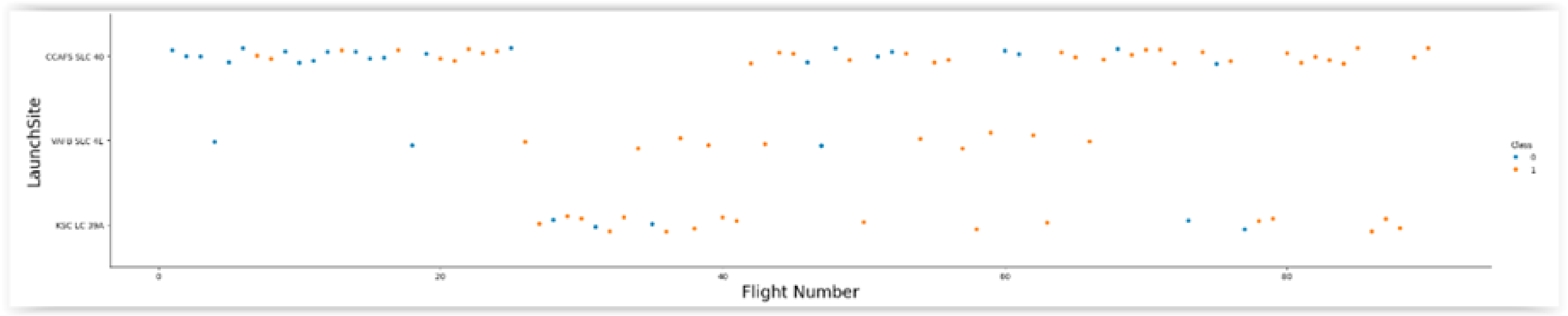
Interactive analytics demo in screenshots



Section 2

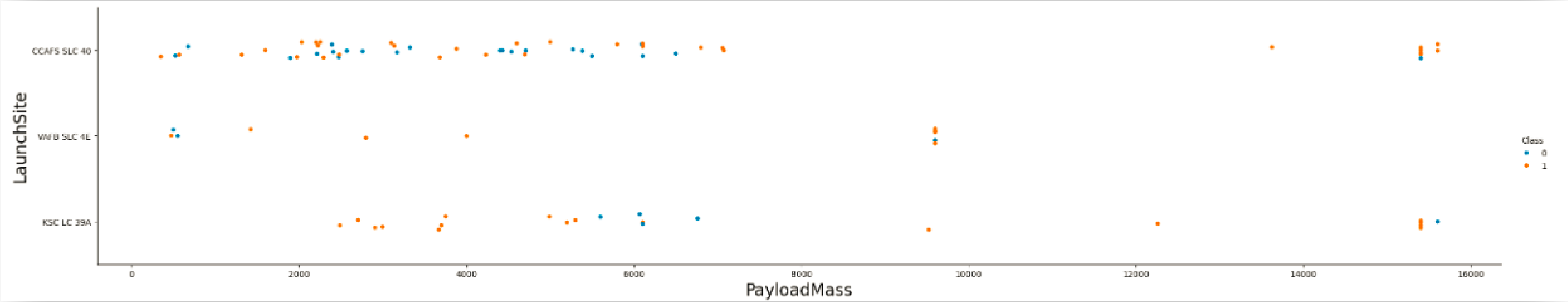
Insights drawn from EDA

Flight Number vs. Launch Site



- There is a clear positive progression in the success rate of launches, as the number of flights carried out increases.
- There is a significant concentration of flight launches from the CCAFS SCL-40 platform.
- The VAFB SLC-4E launch pad does not record flights in the last recorded period.

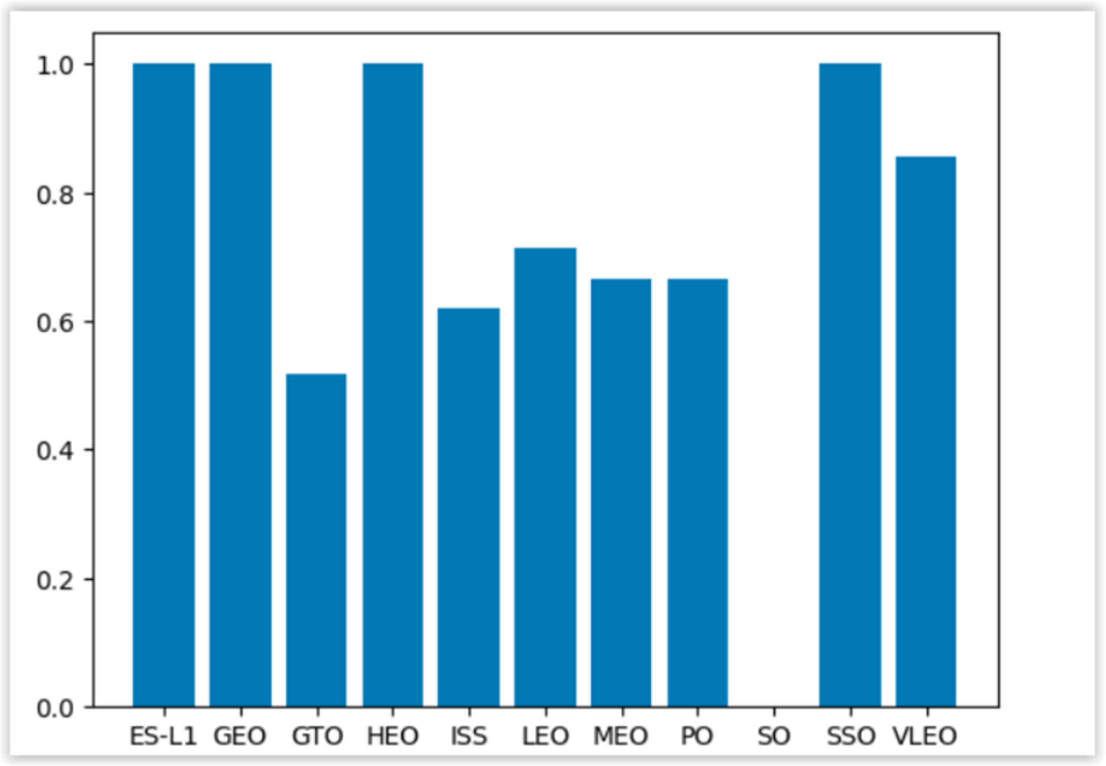
Payload vs. Launch Site

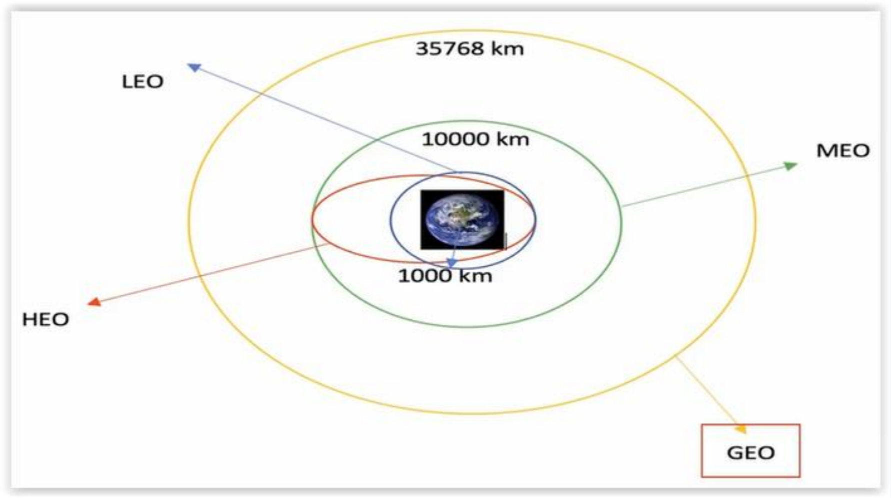


- Above 7000 kg load, the success rate is high.
- About 20000 kg. there are only launches from the CCAFS SCL-40 and KSC CL-39A platforms. Each one records only one unsuccessful case.
- The CCAFS SCL-40 platform is the one that concentrates the greatest number of launches under 7000 kg. In this range, a significant rate of failed recovery is observed.

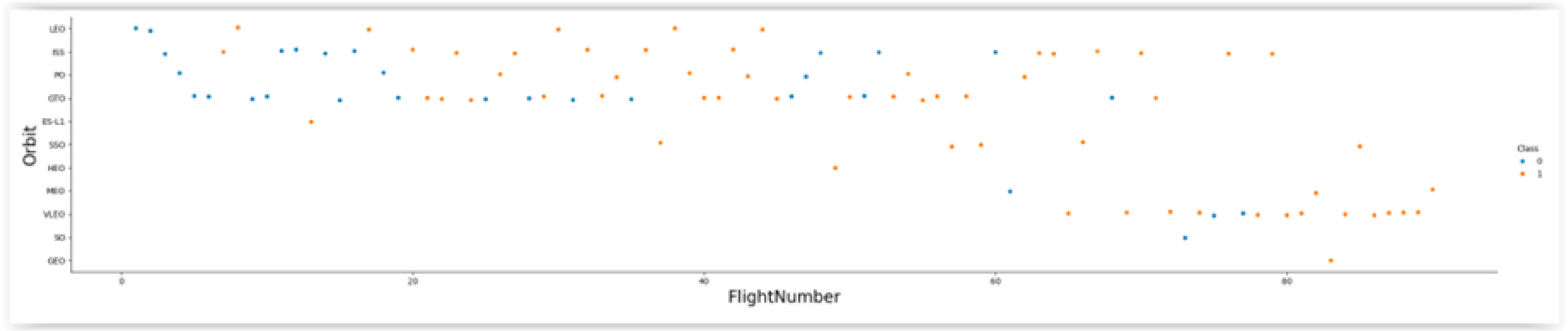
Success Rate vs. Orbit Type

- The orbits that show the maximum success rate are:
  - Stationary satellite orbit at Lagrange point 1 (ES-L1)
  - Geostationary orbit (GEO)
  - Highly elliptical orbit (HEO)
  - Helisynchronous orbit (SSO)



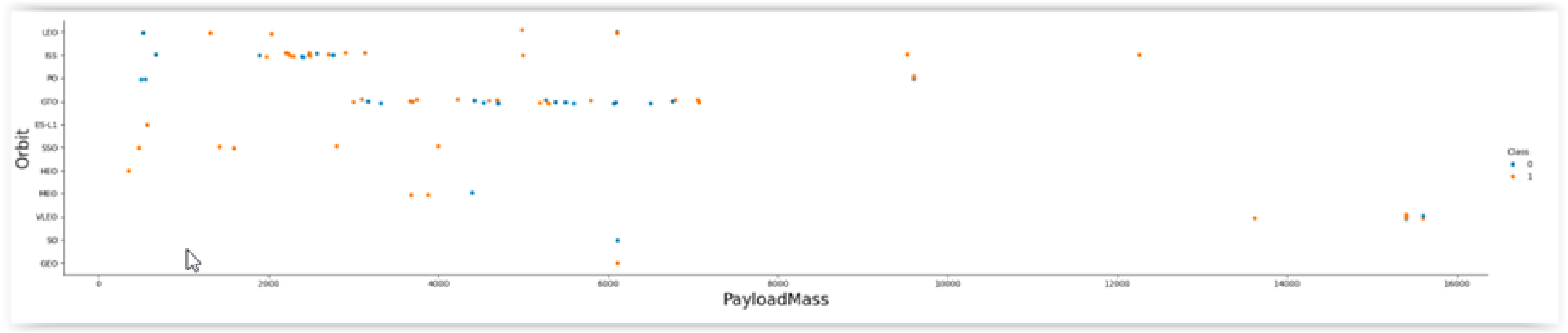


Flight Number vs. Orbit Type



- The FlightNumber carried out are concentrated in 5 orbits:
  - Low Earth Orbit (LEO)
  - International Space Station (ISS)
  - Polar orbit (PO)
  - Geostationary Transfer Orbit (GTO)
  - Very Low Earth Orbit (VLEO)
- Success rate increases noticeably when the flight number is greater.

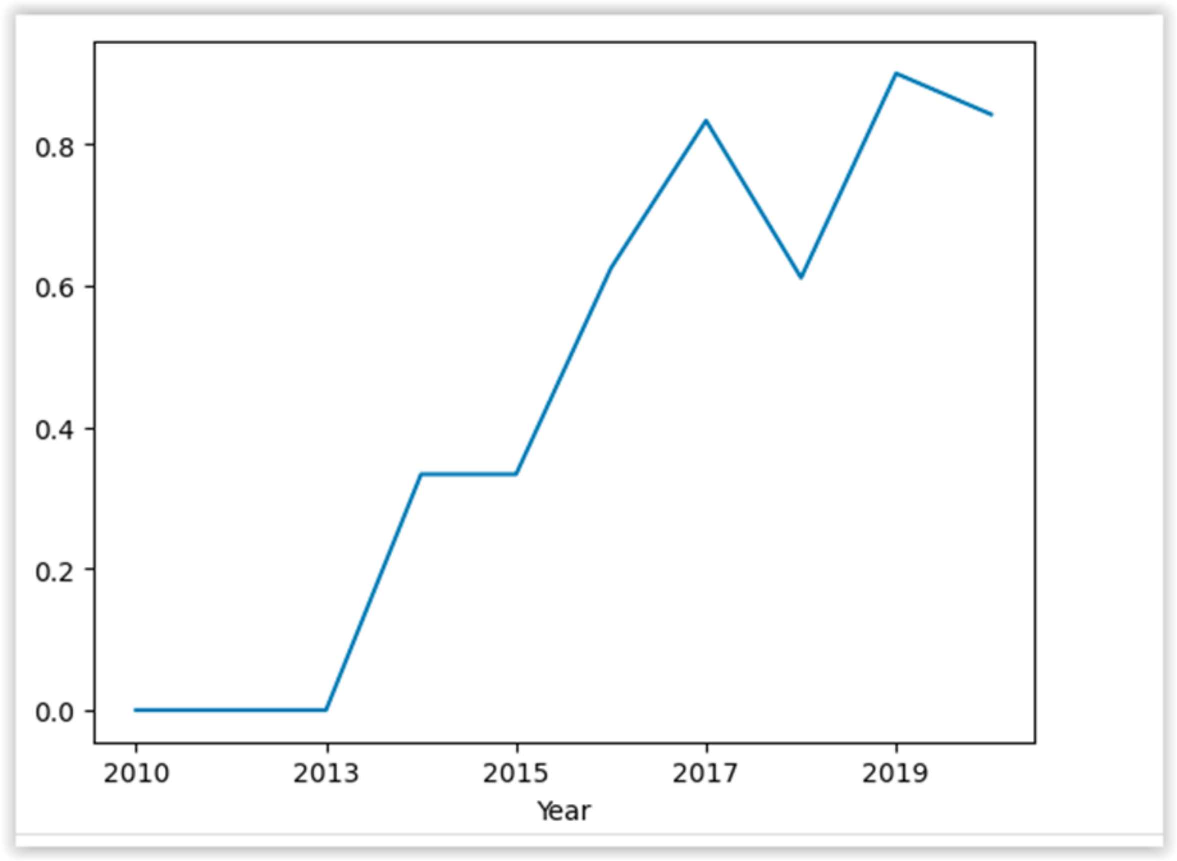
Payload vs. Orbit Type



- Launches with payload exceeding 13,000 kg., they are associated with very low Earth orbit (VLEO), and their success rate is high.
- Most of the payload is below 8,000 kg, and is distributed as follows:
  - In the orbit of the International Space Station (ISS), loads between 2,000 and 3,000 kg are preferably placed, with a success rate that exceeds 70%.
  - n the Geostationary Transfer Orbit (GTO), rockets with a payload between 3,000 and 7,000 kg are sent, presenting an exit rate slightly higher than 50%.

Launch Success Yearly Trend

- The annual success rate shows a significant progression from 2010 (rate = 0.0) to 2020 (rate over 0.8).
- The main features are:
  - Rate stagnant at 0.0 in the period between 2010 and 2013.
  - Sustained rate in 2014 and 2015 (rate close to 0.35).
  - Progressive increase in the rate between 2015 and 2019 (rate above 0.8), with two decreases in the progression in 2018 and 2020, but the rate remains above 0.8.



All Launch Site Names

- SQL query to retrieve unique names of rocket launch sites.
- Key: Use the **distinct** clause to get unique values for the name of launch sites.
- Allows you to identify 4 different launch sites:
    - CCAFS LC-40
    - VAFB SCL-4E
    - KSC LC-39A
    - CCAFS SLC-40

Task 1

Display the names of the unique launch sites in the space mission

```
[8]: %sql select distinct ("Launch_Site") from SPACEXTBL
```

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

[8]:

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

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`

```
%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

- Key: Use of the **limit 5** clause to rescue the first 5 records from database.

Total Payload Mass

- Calculate the total payload carried by boosters from NASA

Task 3

Display the total payload mass carried by boosters launched by NASA (CRS)

```
%sql select sum(PAYLOAD_MASS__KG_) from SPACEXTBL where Customer =  'NASA (CRS)'
```

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

sum(PAYLOAD_MASS__KG_)
45596

- Keys: Use of **sum()** function to add payload and the **where** clause to select only records from NASA (CRS) customer.
- Result: 45596 kg.

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1

Task 4

Display average payload mass carried by booster version F9 v1.1

```
%sql select avg(PAYLOAD_MASS__KG_) from SPACEXTBL where Booster_Version =  'F9 v1.1'
```

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

avg(PAYLOAD_MASS__KG_)
2928.4

- Keys: Use of **avg()** function to calculate the average weight of payload, and the **where** clause to select only records from booster version F9 v1.1
- Result: 2928,4 kg.

First Successful Ground Landing Date

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

Task 5

List the date when the first succesful landing outcome in ground pad was acheived.

Hint:Use min function

```
%sql select min(date) from SPACEXTBL where "Landing_Outcome" =  'Success (ground pad)'
```

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

min(date)
2015-12-22

- Keys: Use of **min()** function to obtain the smallest value of date attribute, and **where** clause to select only records of successful landings on the terrestrial platform.

- Result: date is 2015-12-22.

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

Task 6

List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

```
%sql select booster_version,PAYLOAD_MASS__KG_,Landing_Outcome 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	PAYLOAD_MASS__KG_	Landing_Outcome
F9 FT B1022	4696	Success (drone ship)
F9 FT B1026	4600	Success (drone ship)
F9 FT B1021.2	5300	Success (drone ship)
F9 FT B1031.2	5200	Success (drone ship)

- Keys: Use of the **select** clauses for selection and **where** for filtering. Where clause uses **and** conditions, equal (=), greater than (>) and less than (<).

- Result: 4 boosters met the indicated specifications.

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes

Task 7

List the total number of successful and failure mission outcomes

```
%sql select Mission_Outcome, count(*) as total from SPACEXTBL Group by Mission_Outcome Order by Mission_Outcome
```

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

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

- Keys: Combines several actions on the Mission_Outcome attribute: **select** clauses for selection, **count()** function to determine occurrences, **group by** clause for grouping and **order by** clause for record ordering.

- Result: 101 records grouped into 4 classes according to values of Mission_outcome attribute.

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass
- Keys: The information is obtained through **query** and **subquery**, combining **select**, **where** and **order by** clauses, integrated with **max()** function.
- Result: Shows a list with 12 boosters name that have transported the maximum payload registered in the database (15600 kg).

Task 8

List the names of the booster_versions which have carried the maximum payload mass. Use a subquery

```
%sql select Booster_Version, PAYLOAD_MASS__KG_ from SPACEXTBL where
PAYLOAD_MASS__KG_ = (select max(PAYLOAD_MASS__KG_)
                     from SPACEXTBL) order by Booster_Version
```

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

Booster_Version	PAYLOAD_MASS__KG_
F9 B5 B1048.4	15600
F9 B5 B1048.5	15600
F9 B5 B1049.4	15600
F9 B5 B1049.5	15600
F9 B5 B1049.7	15600
F9 B5 B1051.3	15600
F9 B5 B1051.4	15600
F9 B5 B1051.6	15600
F9 B5 B1056.4	15600
F9 B5 B1058.3	15600
F9 B5 B1060.2	15600
F9 B5 B1060.3	15600

2015 Launch Records

- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

Task 9

List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.

Note: SQLLite does not support monthnames. So you need to use substr(Date, 6,2) as month to get the months and substr(Date,0,5)='2015' for year.

```
%sql select "Booster_Version", "Launch_Site" from SPACEXTBL where
"Landing_Outcome" = 'Failure (drone ship)' and substr(Date,1,4) = '2015'
```

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

Booster_Version	Launch_Site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

- Query uses **select** and **where** clauses combined with **substr()** function, to obtain the requested information.
- Result: Shows failed drone landings from 2015, identifying 2 boosters at the CCAFS LC-40 launch site.

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

- Rank the 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

Task 10

Rank the 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(*) as 'COUNT' from SPACEXTBL
where substring(Date,1,4) || substring(Date,6,2) || substring(Date,9,2)
between '20100604' and '20170320' 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

- Complex query that combines clauses (**select, where, between, group by, order by** and **DESC**), with functions (**count()** and **substring()**), plus logical connector **and**.
- Result: Displays a list with 8 types of landing results and 31 observed cases for time period indicated in the query.

Section 3

Launch Sites Proximities Analysis

Folium Map: Launch Sites Locations

- There are facilities near the Pacific and Atlantic coasts.
- The Atlantic shelves are located closer to the Equator.
- Launches from the Atlantic coast allow us to meet better safety standards.

Folium Map: Launch outcomes for each site

- Total launches pads analyzed: 56
  - Successful 24 (green symbol)
  - Unsuccessful 32 (red symbol)
- West Coast: one launch pads
  - VAFB SCL-4E: 10 launches, 4 successful
- East Coast: 3 launch pads, 46 launches, 20 successful
  - KSC LC-39A: 13 launches, 10 successful
  - CCAFS LC-40: 26 launches, 7 successful
  - CCAFS SCL-40: 7 launches, 3 successful

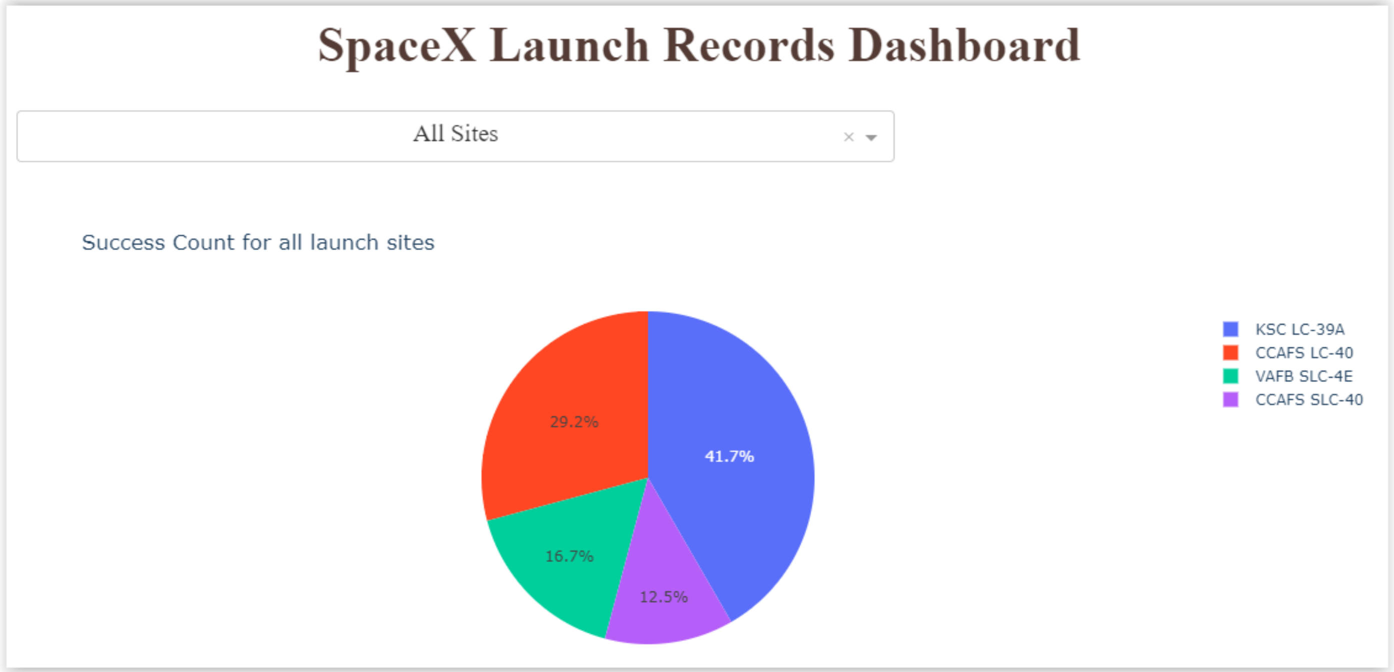
Folium Map: Security and mobility around a launch pad

- The territorial proximity to transportation infrastructure (railways, highways) or to geographic locations due to security criteria (coastline, cities) are important factors in the analysis of the work carried out.
- As an example, the distance from the CCAFS SCL-40 launch platform to 4 specific geographical points was calculated:
    - Coastline: 0.88 km. (blue line on the map)
    - Nearest city (Titusville): 23.73 km
    - Most railway line (Railway NASA): 1.06 km
    - Nearest highway (Highway FL405): 7.50 km

Section 4

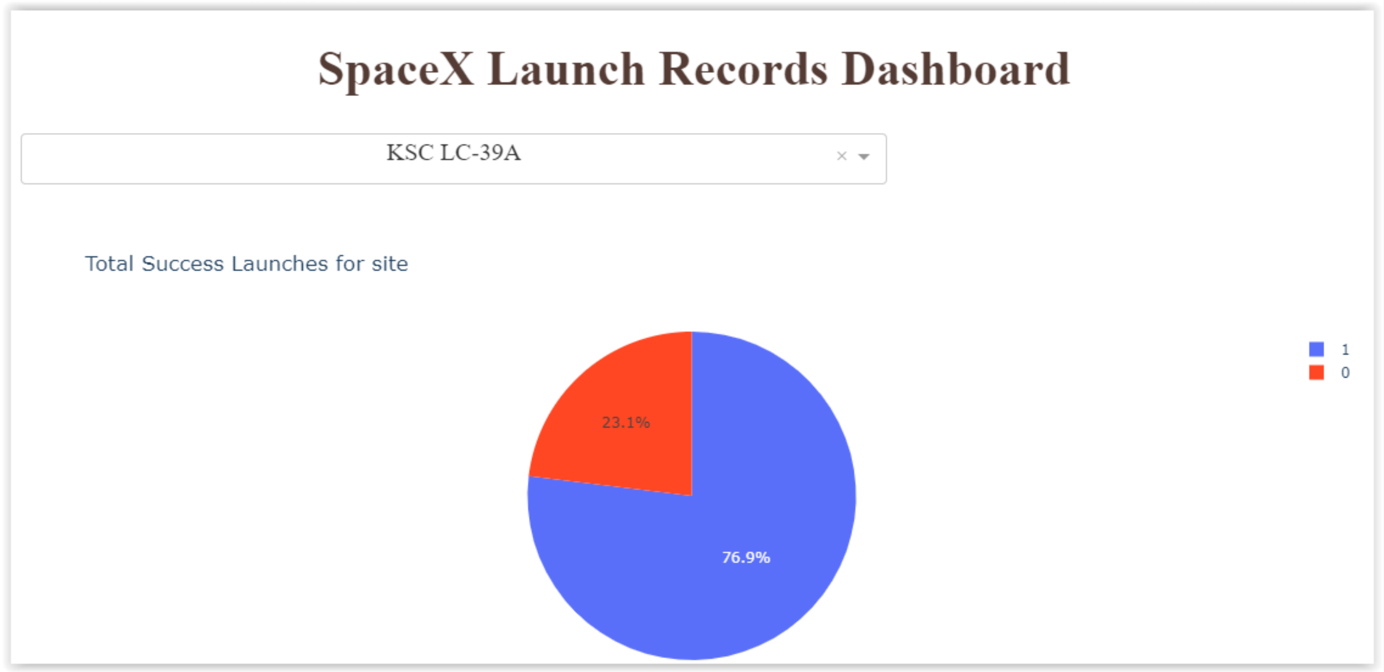
Build a Dashboard with Plotly Dash

Dashboard: Dropdown list panel - global information



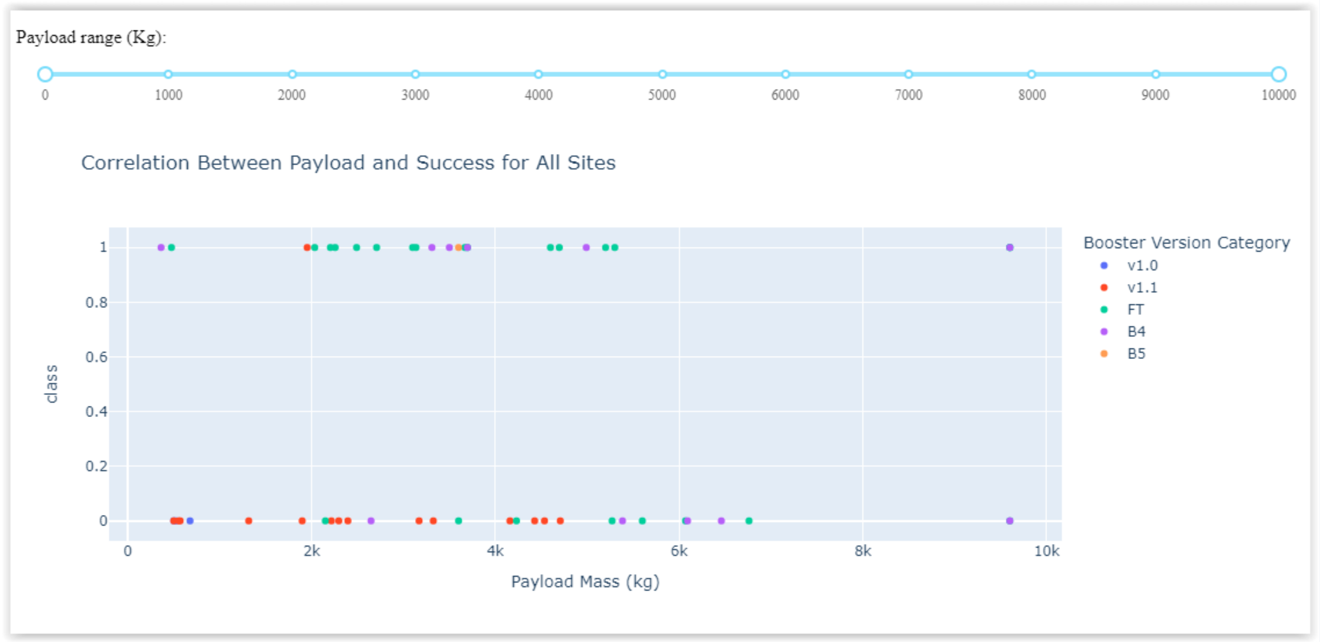
- The drop-down panel allows you to analyze the information for each launch site, or all of them together.
- When the global information (all sites) is reviewed in the drop-down panel, pie chart displays the percentage share of each launch site in the total successful landings.
- The pie chart shows that KSC LC-39A platform has the best rate, with 41.7% of successful landings, while CCAFS LC-40 platform only accounts for 12.5% of successful landings.

Dashboard: Dropdown list panel - launch platform information



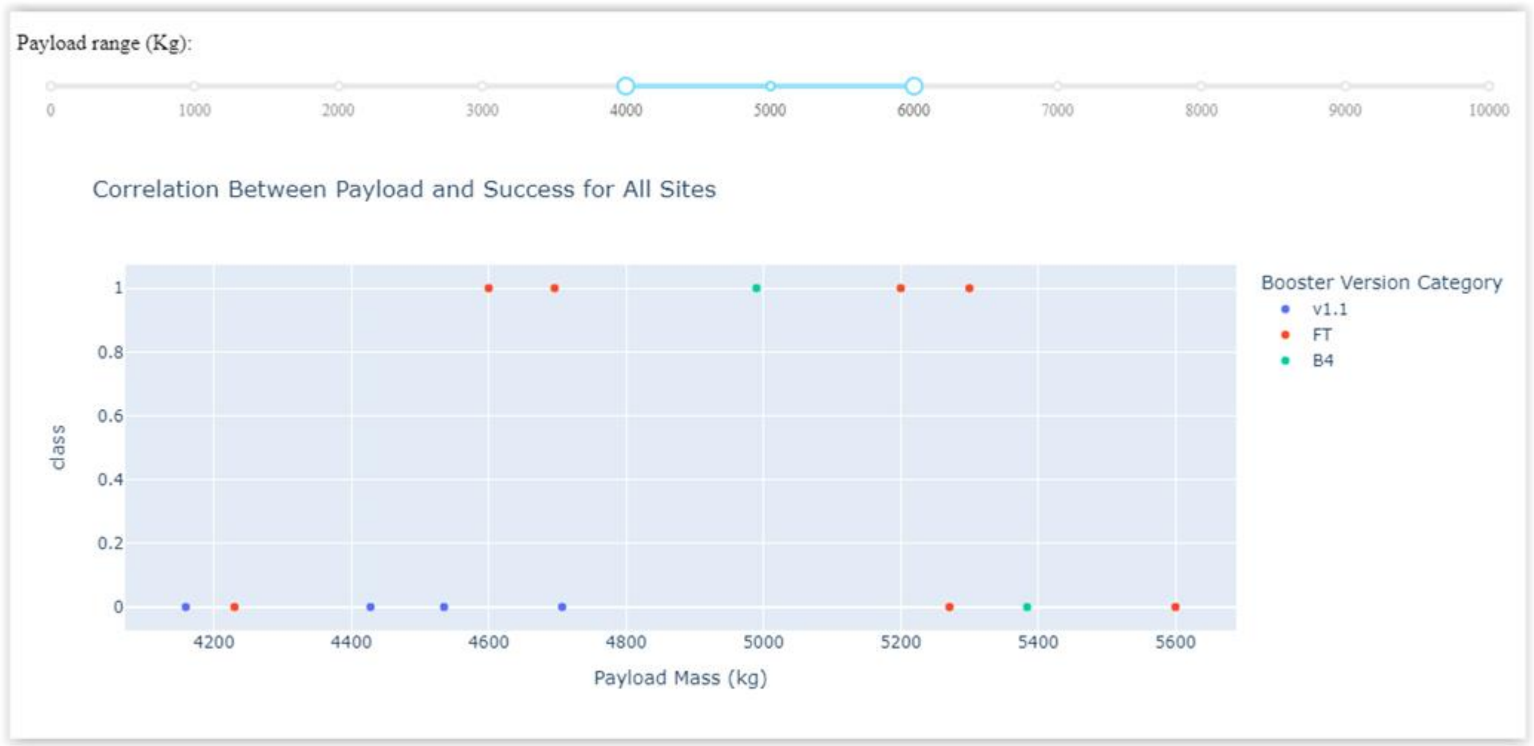
- When a launch pad is identified in the drop-down panel, the pie chart displays the successful landing rate (blue color) and unsuccessful landing rate (red color).
- The pie chart shows that on the KSC LC-39A platform, 76.9% of landings have been successful, and only 23.1% of them have been unsuccessful.

Dashboard: Range slider



- A range slider is used to define search criteria by payload and relate it to drive model used.
- The dot scatter plot shows a successful (class=1) or failed (class=0) point using a different color for each booster category (5), located into space according to payload transported.
- For example, the boosters model FT has 12 successful landings with loads between 0 and 6000 kg, and also records 6 failed landings with loads between 2000 and 7000 kg.

Dashboard: Range slider



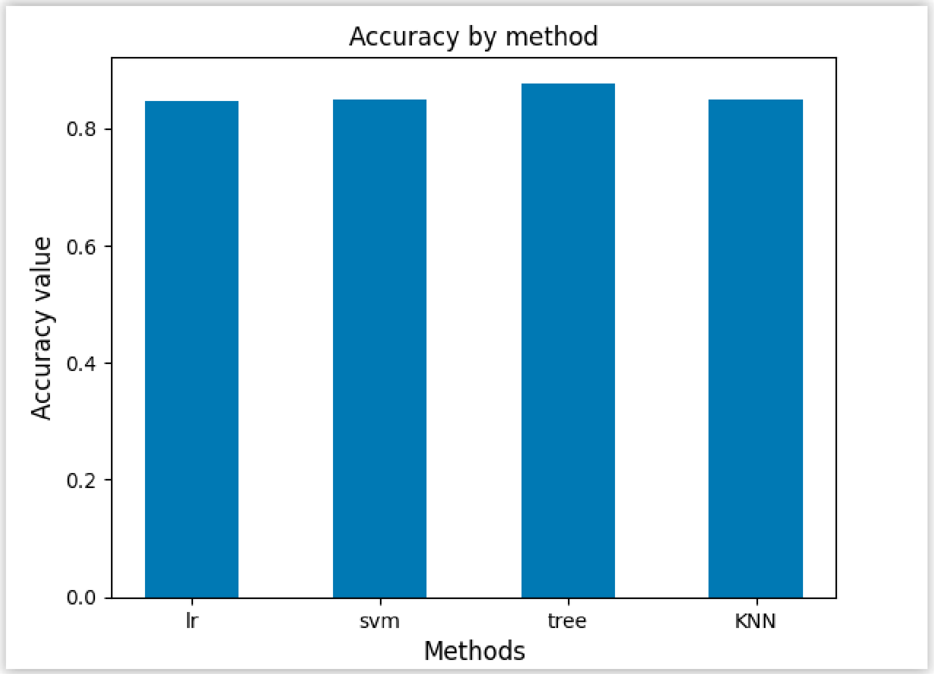
- This example shows the distribution when adjusting the payload range between 4000 and 6000 kg.
- It is shown that only 3 booster models were used for this range of payloads.
- Only two booster models record successful landings: FT and B4.

Section 5

Predictive Analysis (Classification)

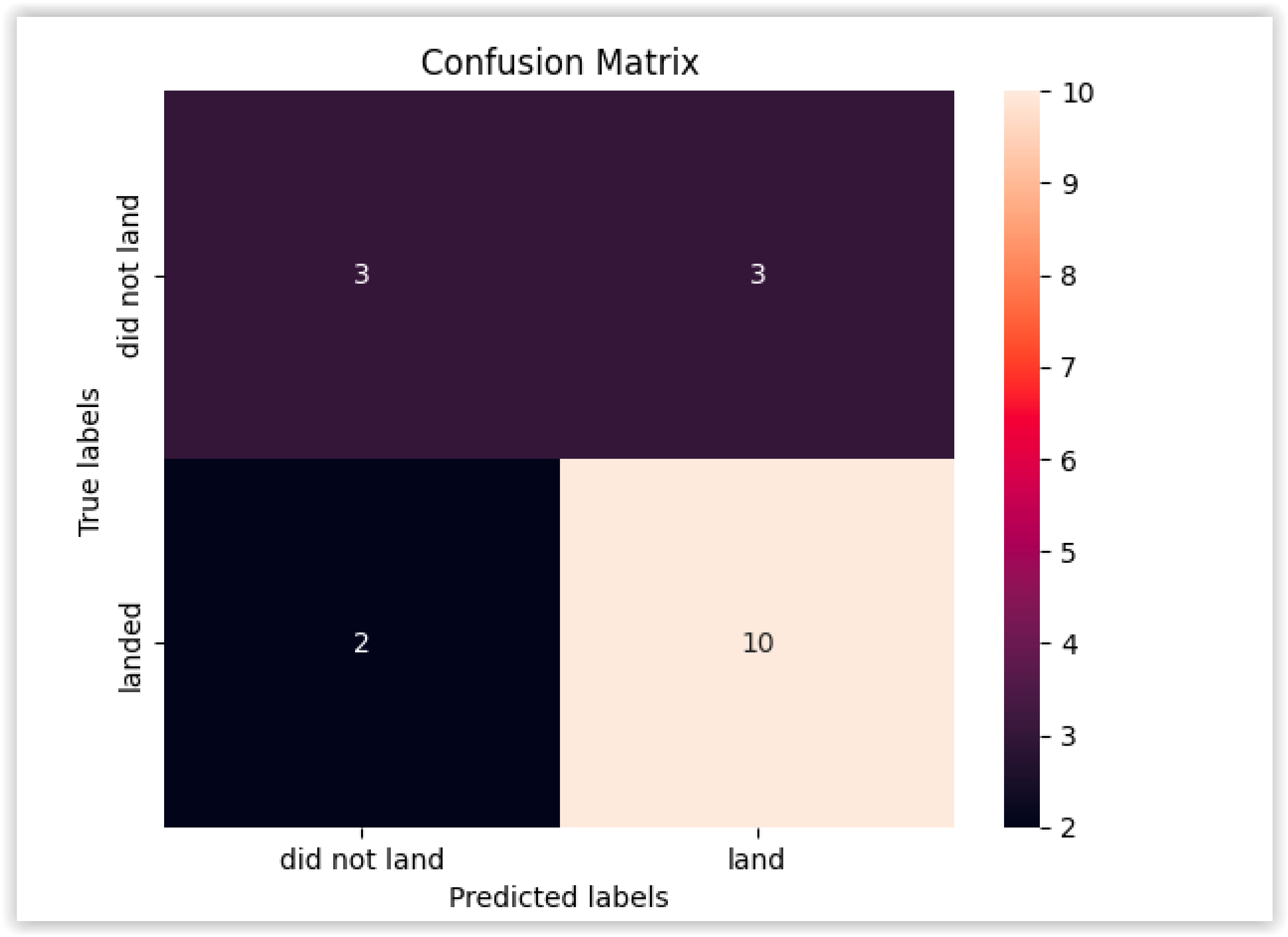
Classification Accuracy

- The percentage of observations correctly classified with respect to the total predictions shows an accuracy greater than 0.84 for each model evaluated:
  - Logistic Regression (lr)
  - support vector machine (svm)
  - Regression tree (tree)
  - K-Nearest Neighbors (KNN)
- The model that obtains the best results is **Tree**, with an accuracy of 0.8767



Confusion Matrix

- The confusion matrix for the prediction made with the **xtree** regression model, on 18 observations, shows the following results:
  - 10 true positives (well-classified successful landings)
  - 3 true negatives (well-classified failed landings)
  - 3 false positives (failed landings but classified as successful)
  - 2 false negatives (successful landings but classified as failed)



Conclusions

- Successful landings present significant and progressive improvement over time.
- A range is identified (2000 to 6000 kg) where successful landings have a greater positive rate.
- The prediction methods used show good levels of accuracy, and their confusion matrices yield similar results.
- The **decision tree method (tree)** is the one that obtains the best accuracy results, reaching 0.8767 in the training test data.
- It is possible to **recommend the use of this model (tree) to Space Y** to decide its entry into the space travel market.

Appendix

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

Thank you!