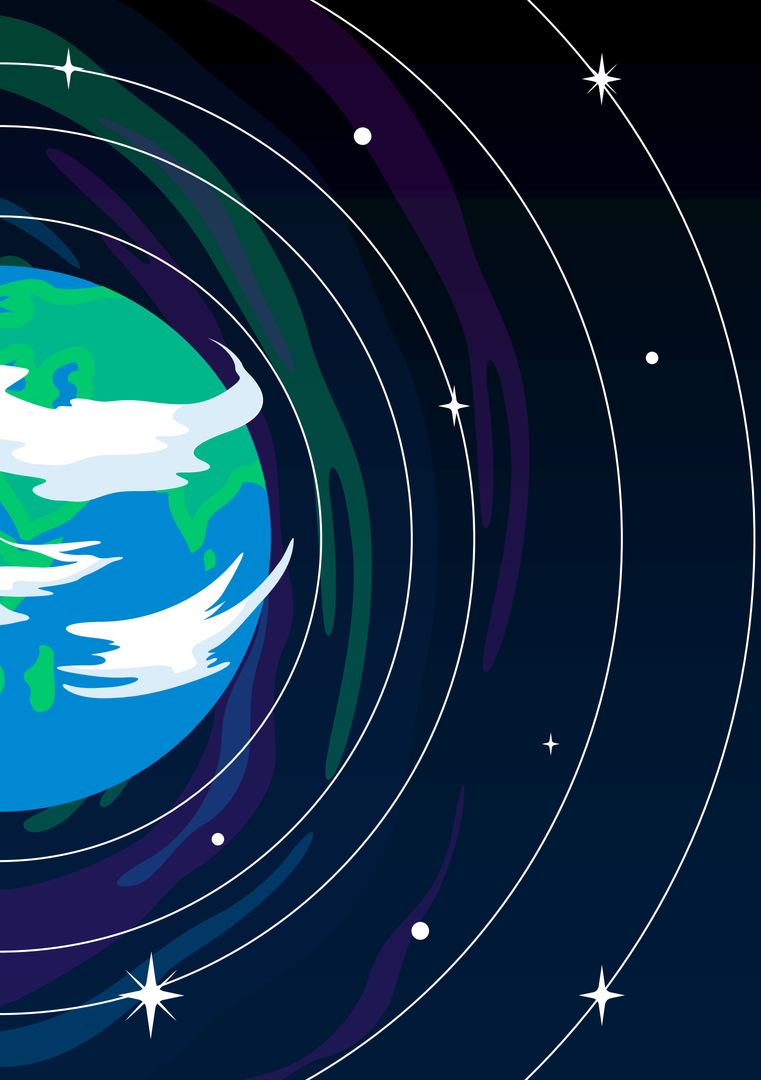


Data Science Capstone Project

Elia Alhanach
8th of May 2024

TABLE OF CONTENTS

01 Executive Summary

02 INTRODUCTION

03 Methodology

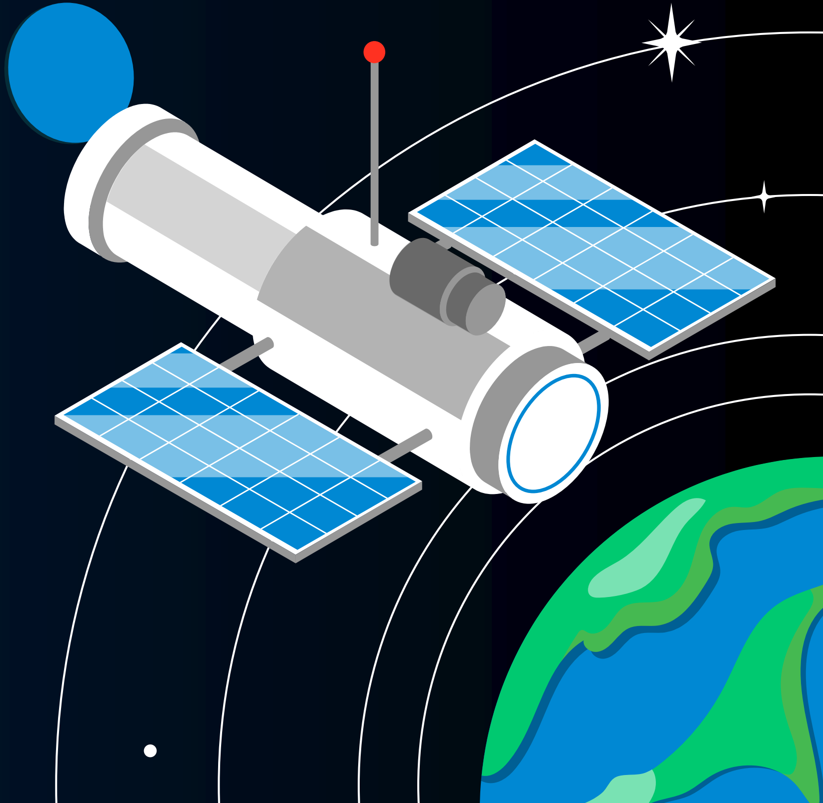
04 Results

05 Conclusion

06 Acknowledgement

01

Executive Summary



Executive Summary

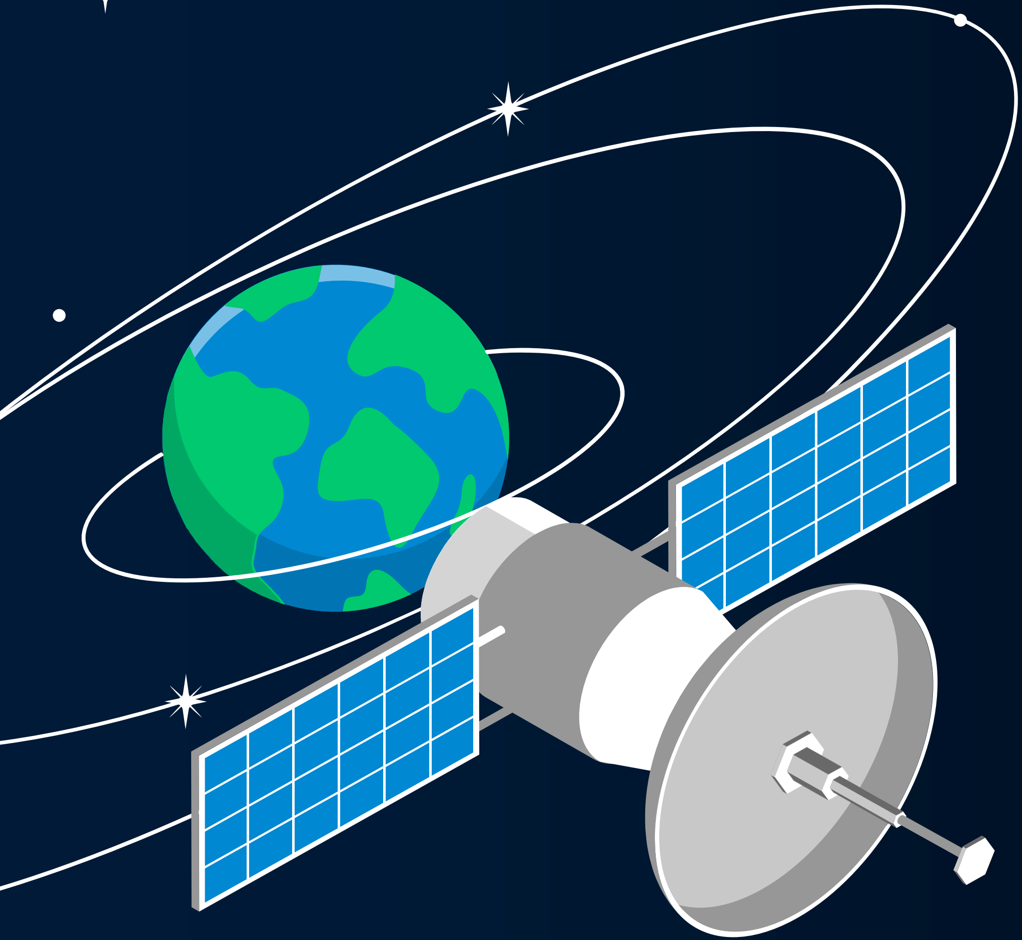
The following methodologies were used to analyze data:

1. Data Collection
2. Data Wrangling
3. Exploratory Data Analysis with Data Visualization
4. Exploratory Data Analysis with SQL
5. Building an interactive map with Folium
6. Building a dashboard with Ploty Dash
7. Predictive Analysis and Classification

Executive Summary

Summary of all results:

1. Exploratory Data Analysis
2. Interactive Analytics
3. Predictive analysis resutls



02

INTRODUCTION

INTRODUCTION

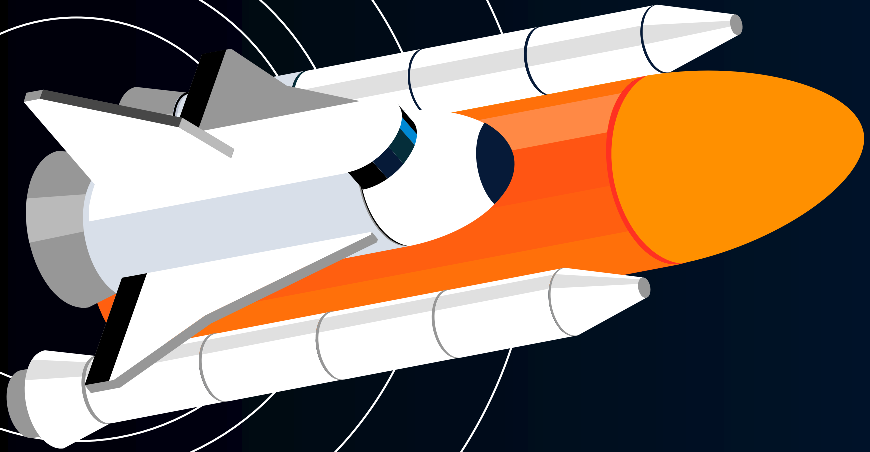
SpaceX advertises Falcon 9 rocket launches on its website, with a cost of 62 million dollars; much of the savings is because SpaceX can reuse the first stage.

Therefore, if we can determine if the first stage will land, we can determine the cost of a launch. Based on public information and machine learning models, we are going to predict if SpaceX will reuse the first stage.

Questions to be answered:

- How do the variables affect the success of the first stage landing ?
- Does the rate of successful landings increase over the years ?
- What is the best algorithm that can be used for binary classification in this case?

INTRODUCTION



03

Methodology

Data Collection

We had to use a process involving a combination of API requests from SpaceX REST API and Web Scraping data from a table in SpaceX Wikipedia entry, to get the complete information about the launching for a more detailed analysis.

Data Collection

Using SpaceX REST API we can obtain the Flight Number, Dates, Booster Version, Payloads, Orbit, Launch Sites, Outcome, Flights, Landing Pads,…..

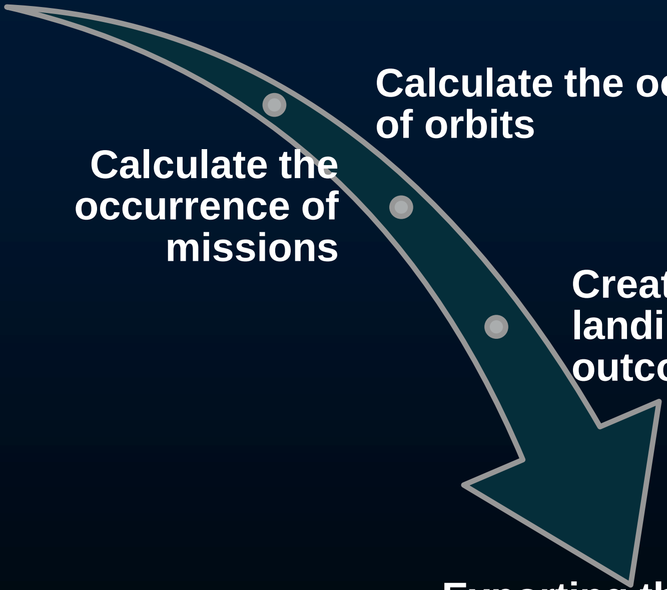
Data Collection

Using Wikipedia Web scraping we can obtain Filght Bumbers, Payloads, Customer, Time, Launch Outcome, …

Data Wrangling

True RTLS means the mission outcome was successfully landed to a ground pad False RTLS means the mission outcome was unsuccessfully landed to a ground pad.True ASDS means the mission outcome was successfully landed on a drone ship False ASDS means the mission outcome was unsuccessfully landed on a drone ship. We mainly convert those outcomes into Training Labels with “1” means the booster successfully landed, “0” means it was unsuccessful.

Data Wrangling



Calculate the number of launches

Calculate the occurrence of orbits

Calculate the occurrence of missions

Create a landing outcome

Exporting the data to CSV

EDA and Data Visualization

A scatter plot shows the relationship between variables: it can be then used in the machine learning model.

A bar chart show a comparison between categories: the relationship between categories is measured.

A line chart shows the trend of the data over time.

EDA and Data Visualization

We plotted multiple charts including:

Flight Number vs. Payload Mass, Flight Number vs. Launch Site, Payload Mass vs. Launch Site, Orbit Type vs. Success Rate, Flight Number vs. Orbit Type, Payload Mass vs Orbit Type and Success Rate Yearly Trend

Predictive Analysis

1- Create a NumPy array from the “Class” in data

2- Standardize the data with Standard Scaler as well as fit and transform it.

3- Split the data into training and testing sets.

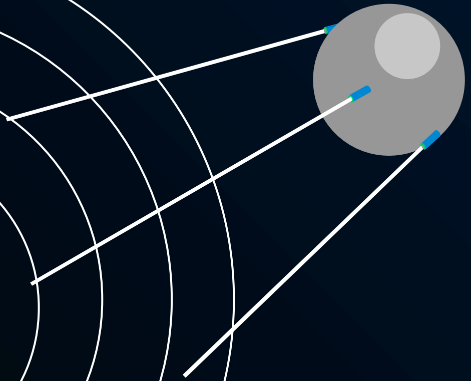
4- Create a GridSearchCV object to find the parameter.

5- Apply it on LogReg, SVM, Decision Tree, KNN.

6- Calculate the Accuracy on the test data.

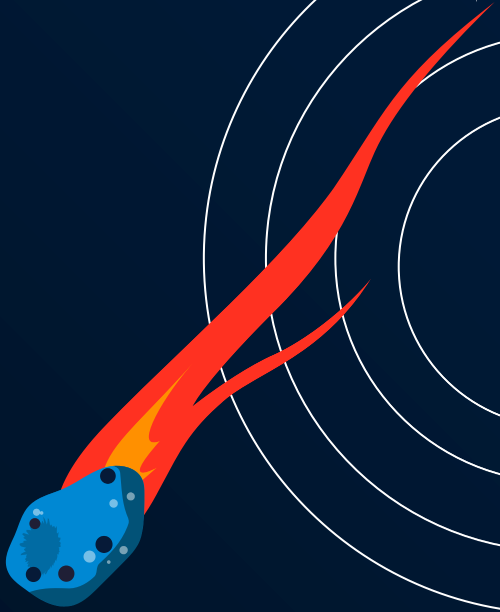
7- Examine the confusion matrix for these models.

8- Find the method that performed the best.



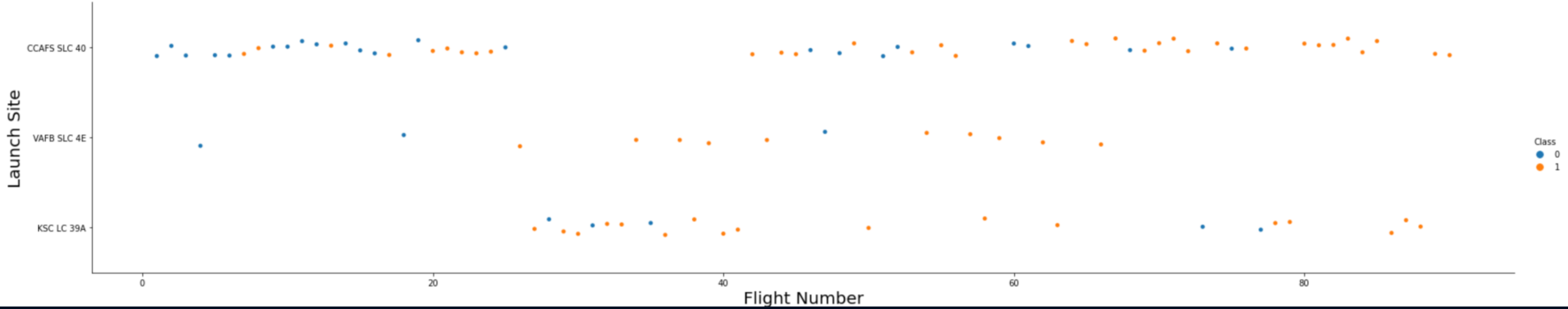
04

Results



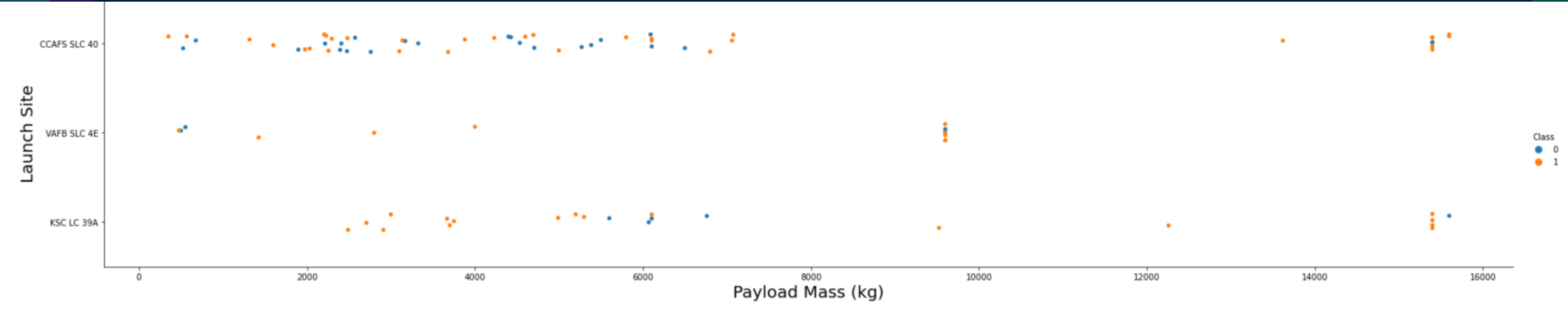
EDA with Visualization Results

Flight Number vs Launch Site



The earliest flights all failed while the latest flights all succeeded.
The CCAFS SLC 40 launch site has about a half of all launches.
VAFB SLC 4E and KSC LC 39A have higher success rates.
It can be assumed that each new launch has a higher rate of success.

Payload vs Launch Site



For every launch site the higher the payload mass, the higher the success rate.

Most of the launches with payload mass over 7000 kg were successful.

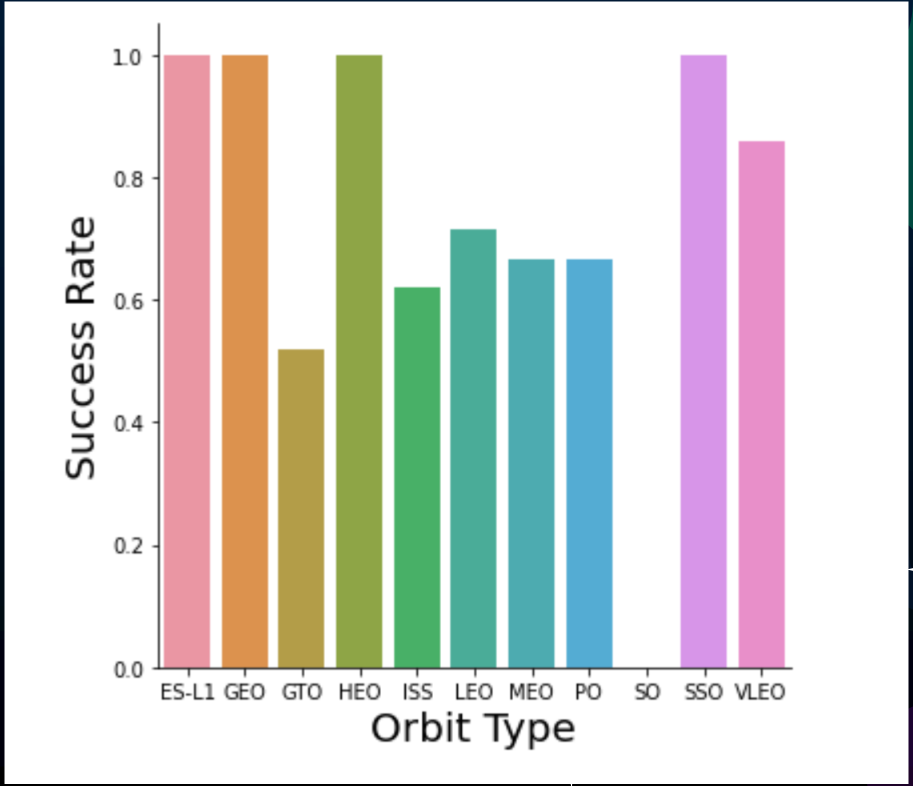
KSC LC 39A has a 100% success rate for payload mass under 5500 kg too.

Success Rate vs Orbit Type

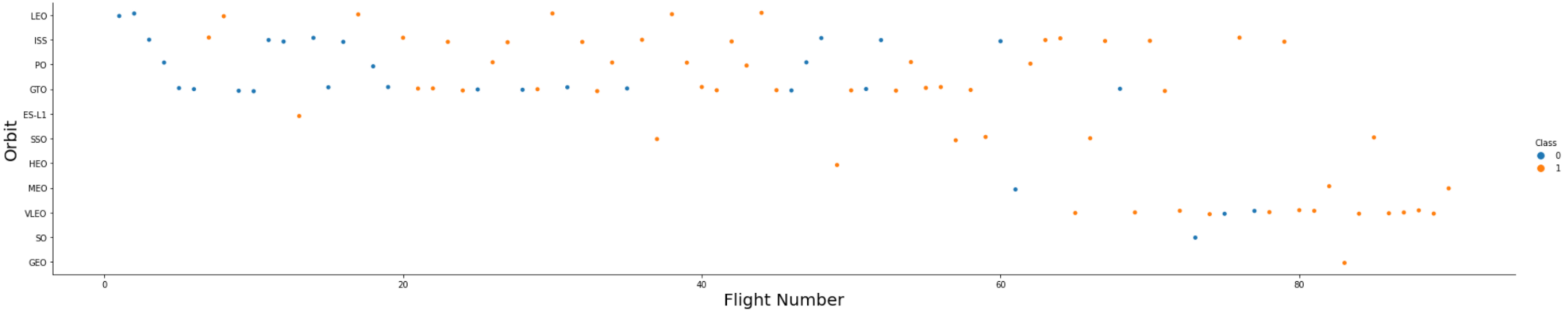
Orbits with 100% success rate are: ES-L1, GEO, HEO, SSO

Orbits with 0% success rate are: SO

- Orbits with success rate between 50% and 85%: GTO, ISS, LEO, MEO, PO

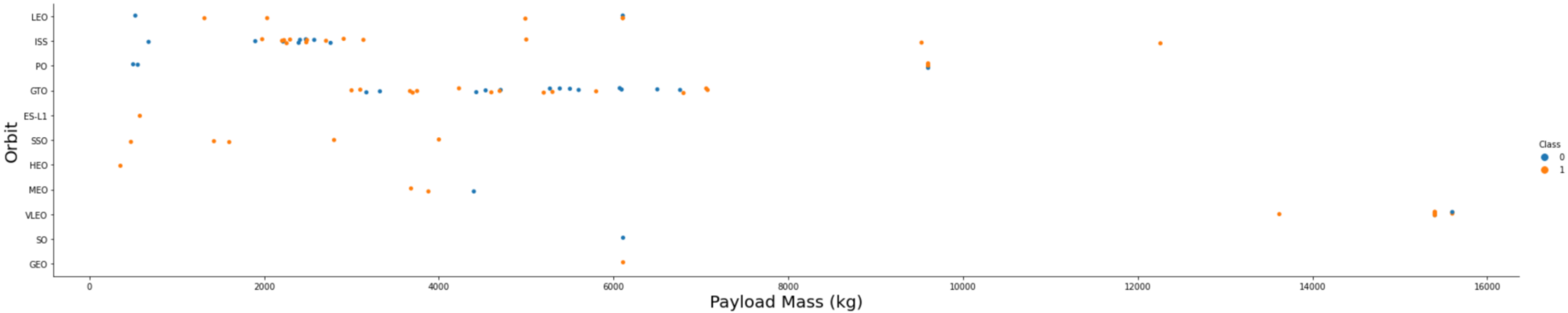


Flight Number vs Orbit Type



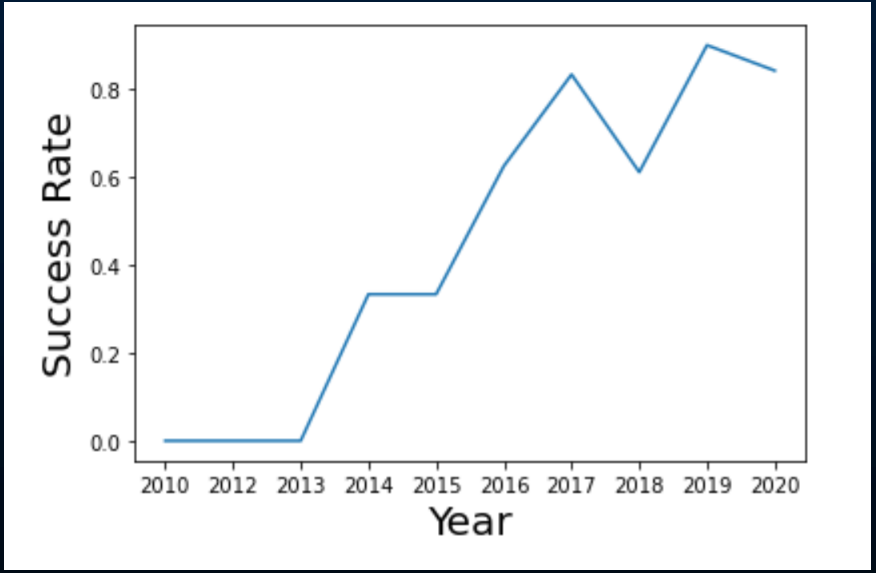
In the LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.

Payload Mass vs Orbit Type



Heavy payloads have a negative influence on GTO orbits and positive on GTO and Polar LEO (ISS) orbits.

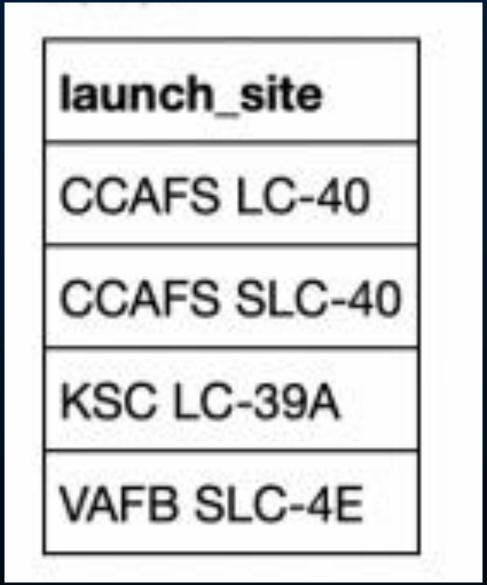
Success Rate vs Years



The sucess rate since 2013 kept increasing till 2020

EDA with SQL results

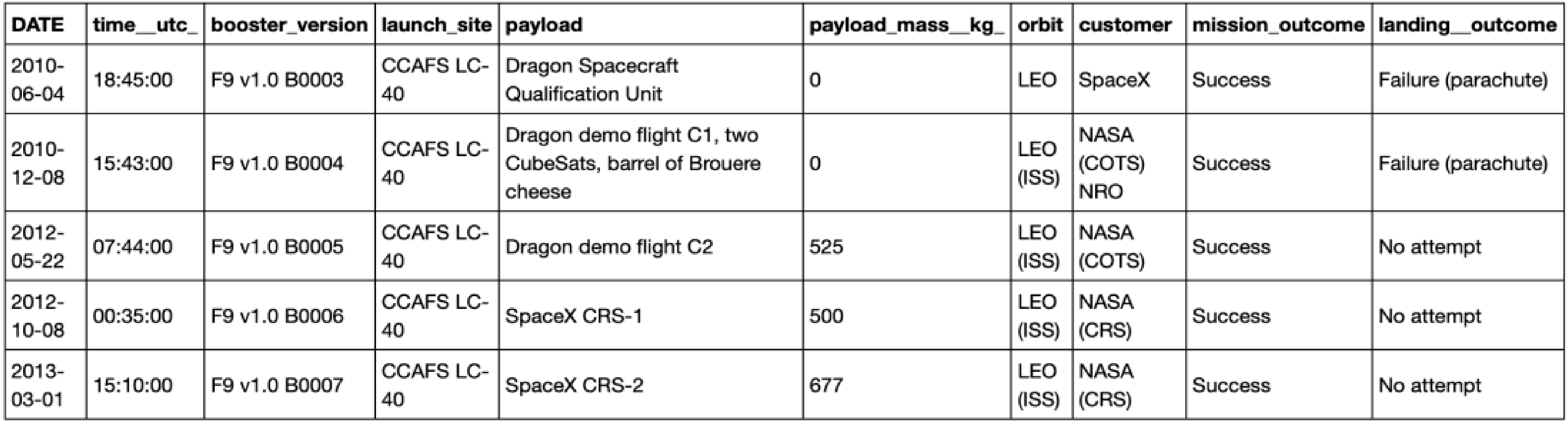
Launch Site Names



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

The Names of the Launch Sites for this mission.

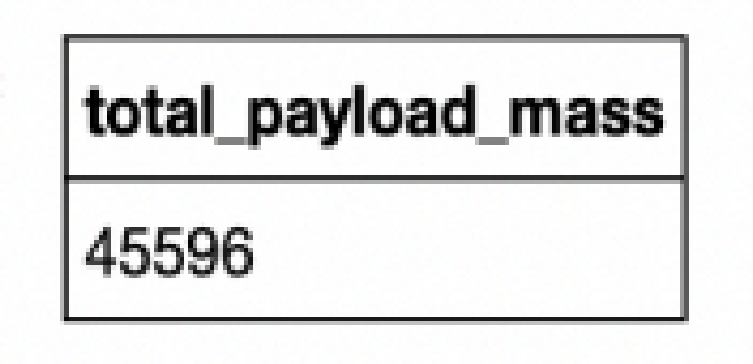
Launch Sites with CCA in their name



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

The records shown are all for launch sites with their name biggening with CCA.

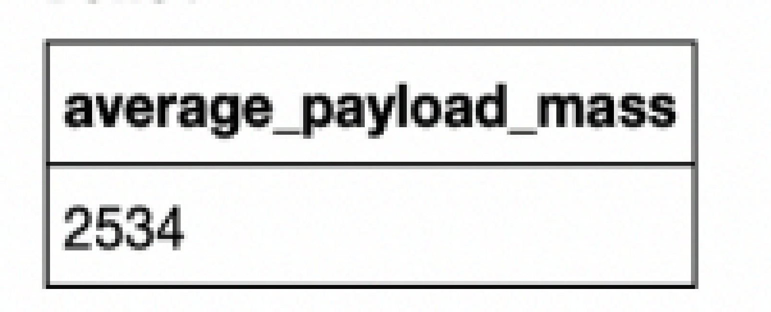
Total Payload Mass



total_payload_mass
45596

The total Payload Mass carried by boosters launched by NASA.

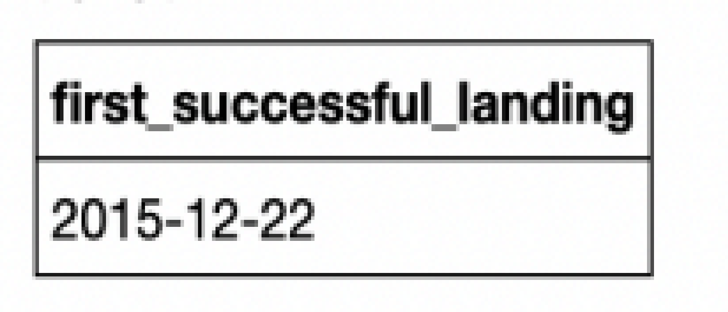
Average Payload Mass by F9 v1.1



average_payload_mass
2534

The Average Payload Mass carried by the booster F9 v1.1

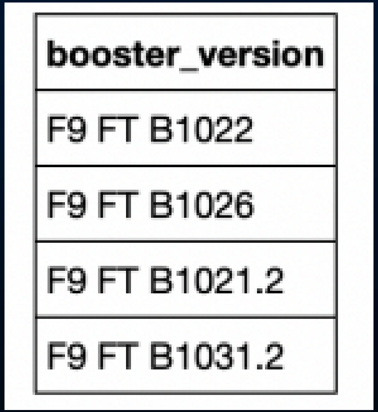
First Successful Ground Landing Date



first_successful_landing
2015-12-22

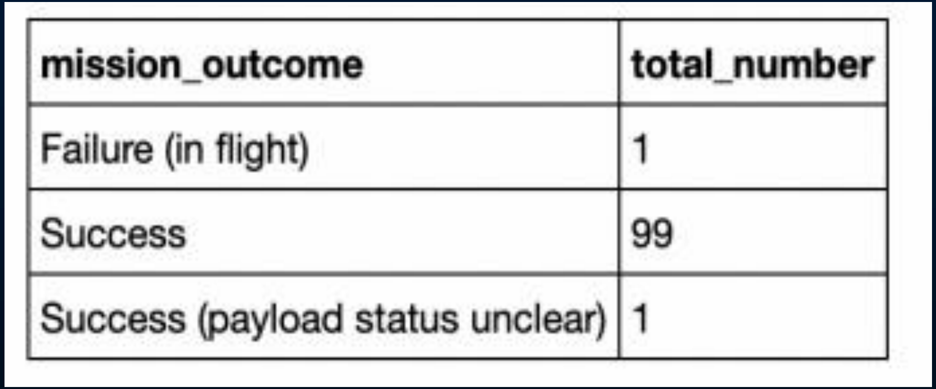
The date when the first successful landing outcome in ground pad was achieved.

Successful Drone Landing with Payload between 4000 and 6000



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

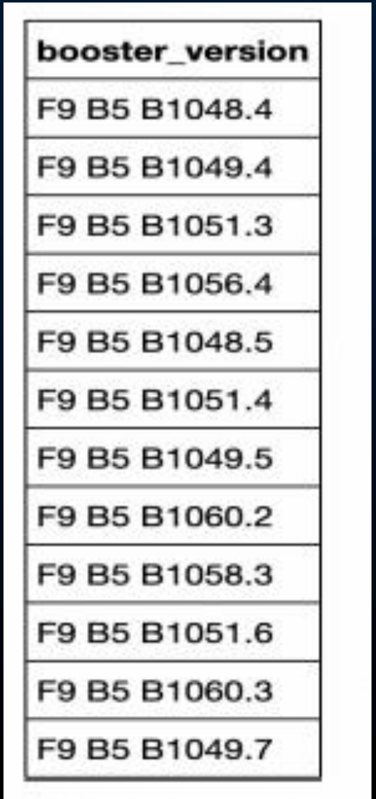
Mission Outcomes



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

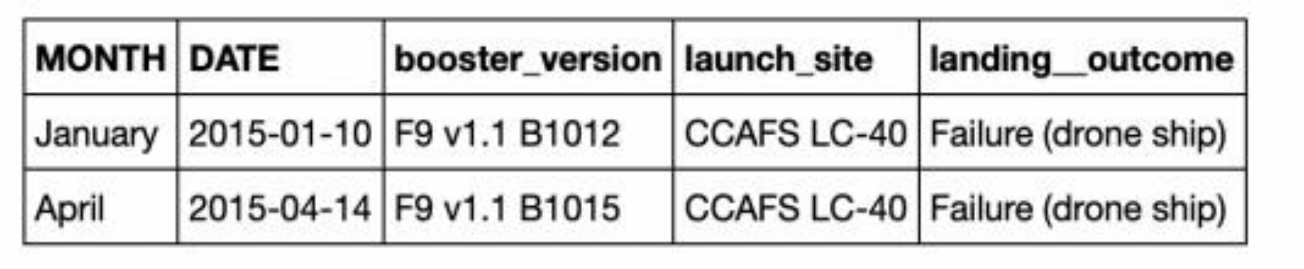
The total number of successful and failed missions.

Boosters that carried the maximum Payload



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

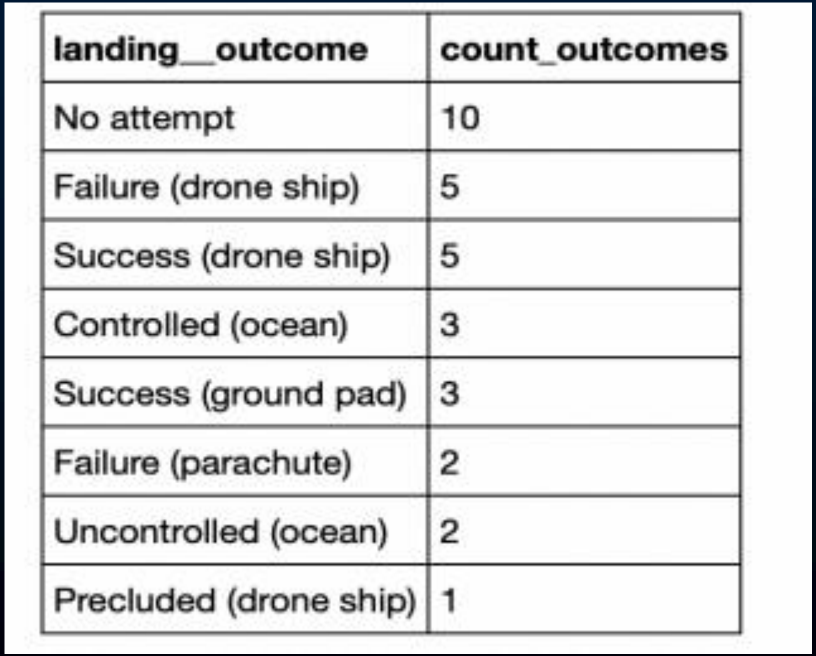
Launch Records of 2015



MONTH	DATE	booster_version	launch_site	landing__outcome
January	2015-01-10	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
April	2015-04-14	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

The failed landing outcomes in drone ship, their booster versions and launch site names for 2015.

Success Count between 4/6/2010 and 20/3/2017

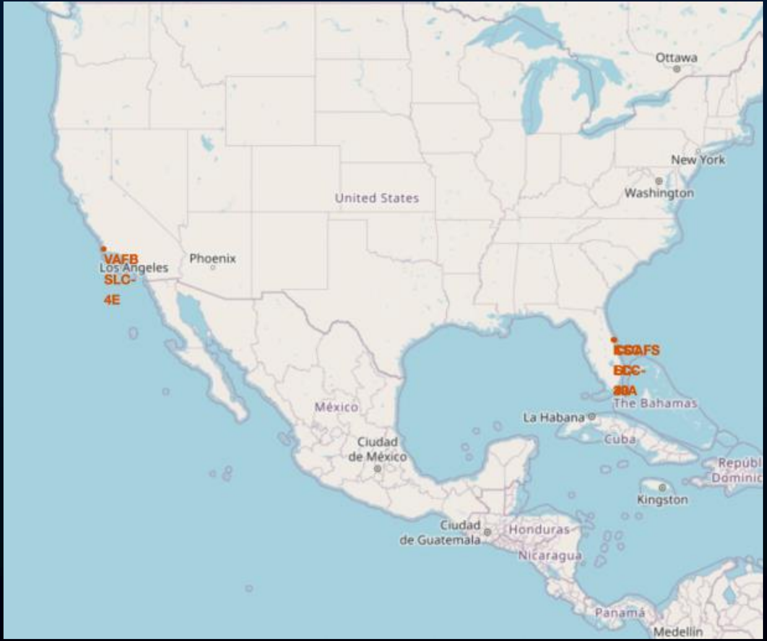


landing__outcome	count_outcomes
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

Interactive map with Folium

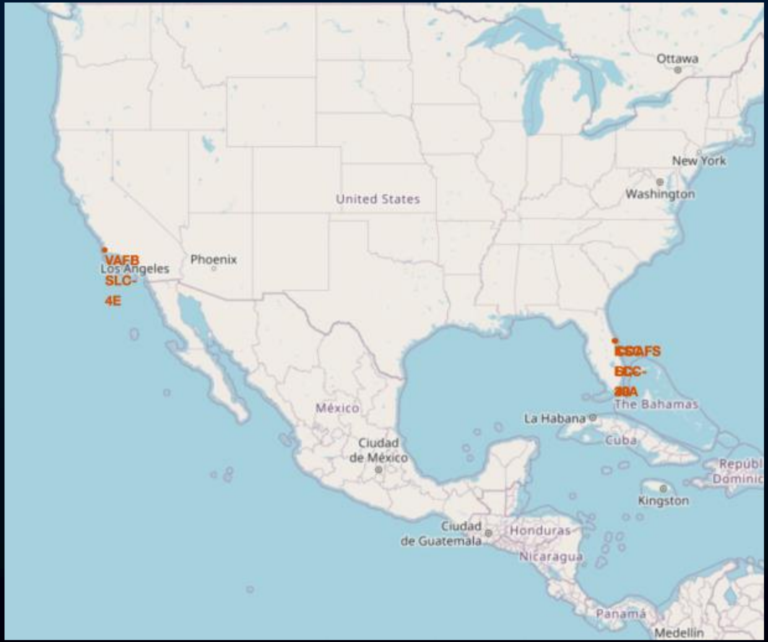
Launch Sites Locations on the Global Map

Most of Launch sites considered in this project are in proximity to the Equator line. Launch sites are made at the closest point possible to Equator line, because anything on the surface of the Earth at the equator is already moving at the maximum speed (1670 kilometers per hour). For example launching from the equator makes the spacecraft move almost 500 km/hour faster once it is launched compared half way to north pole.



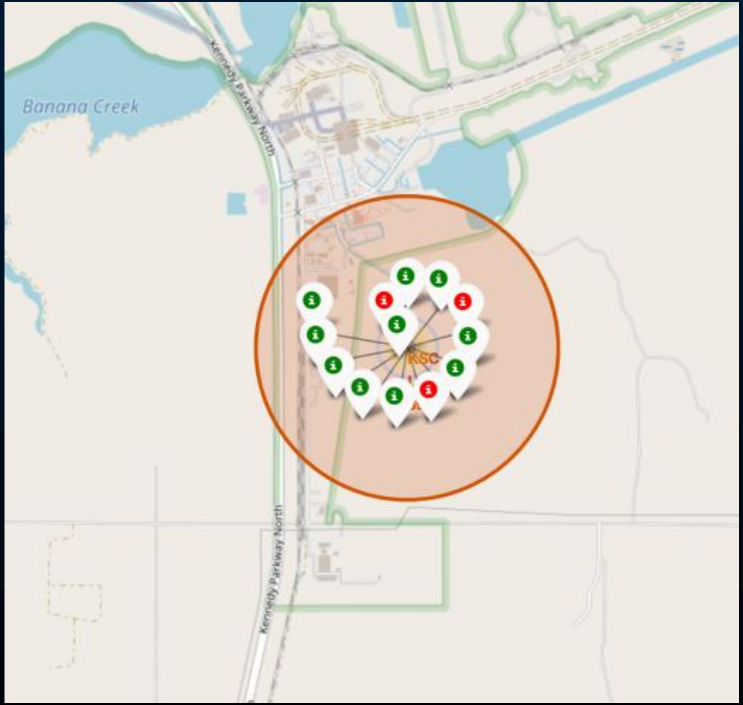
Launch Sites Locations on the Global Map

All launch sites considered in this project are in very close proximity to the coast While starting rockets towards the ocean we minimize the risk of having any debris dropping or exploding near people.



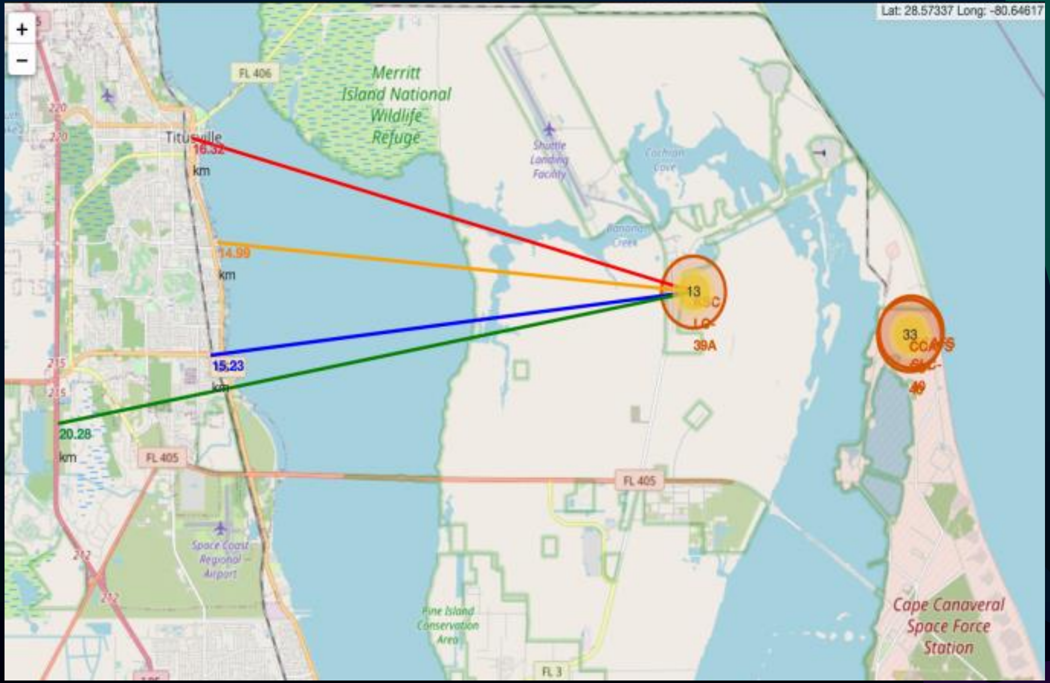
Labels for Launch Records on the map

From the color-labeled markers in marker clusters, you should be able to easily identify which launch sites have relatively high success rates.



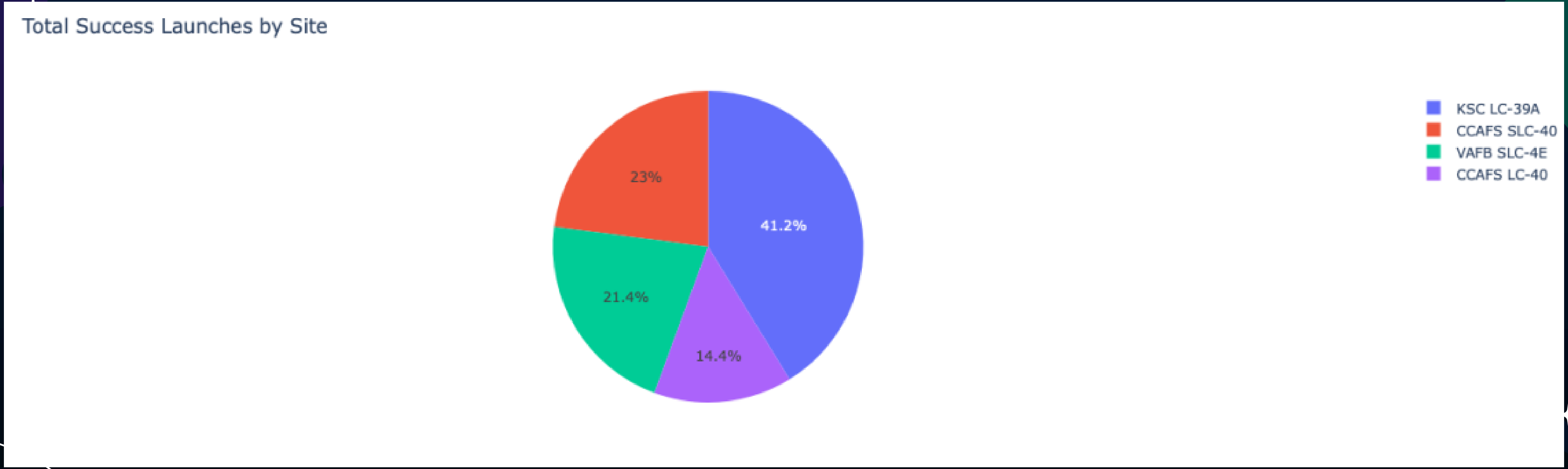
Distance from the Launch Site KSC LC-39A

- From the visual analysis of the launch site KSC LC-39A we can clearly see that it is:
    - relative close to railway (15.23 km)
    - relative close to highway (20.28 km)
    - relative close to coastline (14.99 km)
- Also the launch site KSC LC-39A is relative close to its closest city Titusville (16.32 km).
- Failed rocket with its high speed can cover distances like 15-20 km in few seconds. It could be potentially dangerous to populated areas.

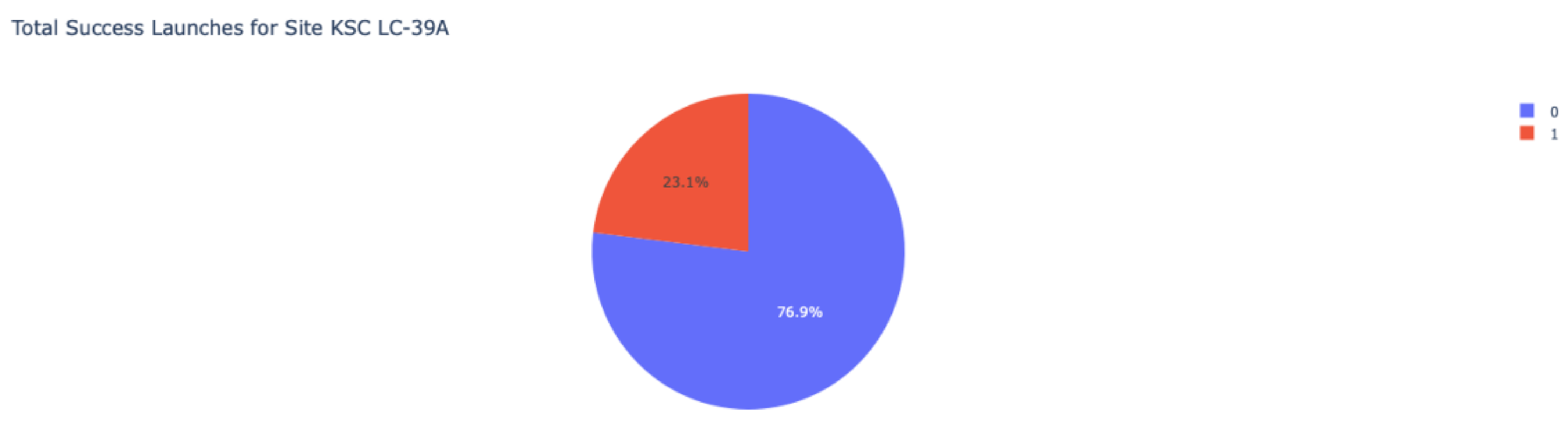

Lat: 28.57337 Long: -80.64617
+
−
Merritt Island National Wildlife Refuge
Titusville
16.32 km
14.99 km
15.23 km
20.28 km
13
KSC LC-39A
33
CCAFS SLC-40
Shuttle Landing Facility
FL 406
FL 405
FL 3
Space Coast Regional Airport
Pine Island Conservation Area
Cape Canaveral Space Force Station

Dashboard with Ploty Dash

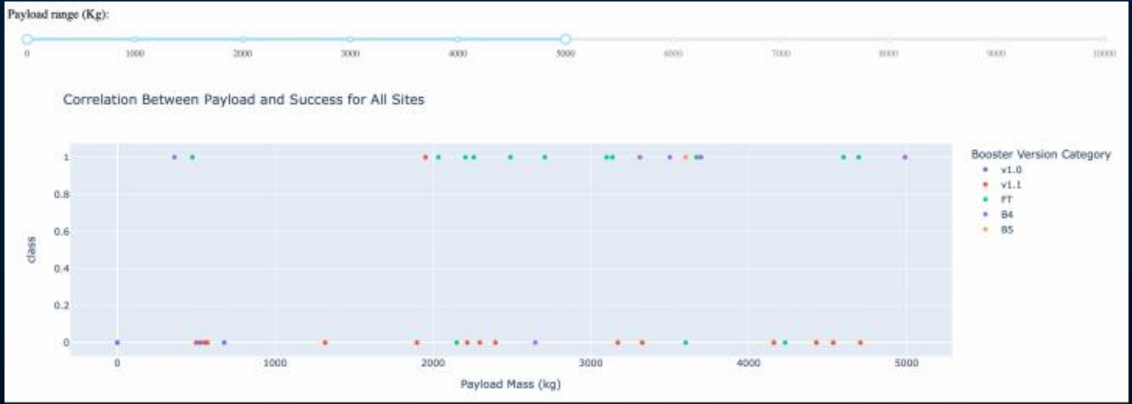
Launch Success Count for all sites



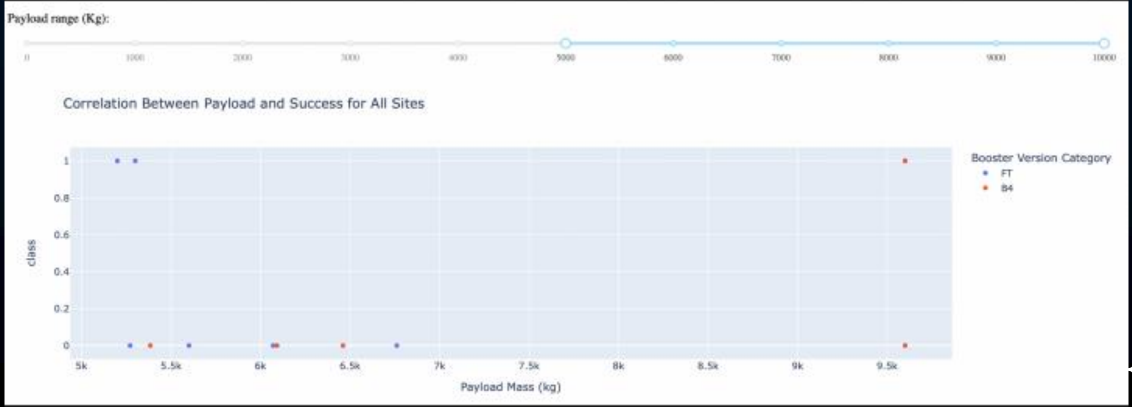
KSC LC-39A the site with the highest launch success ratio



Payload Mass vs Launch Outcomes for all sites



Payload range (Kg):



Payload range (Kg):

Predictive Analysis

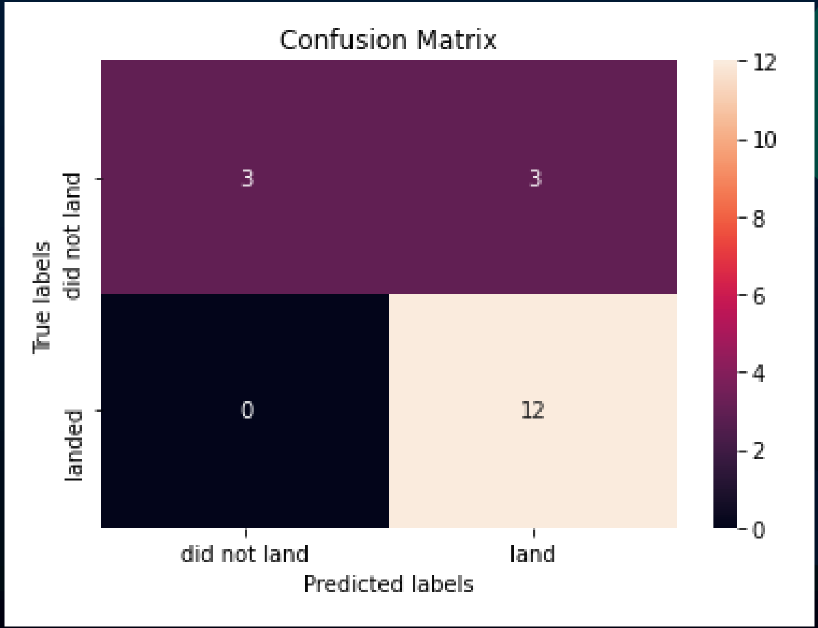
Classification Accuracy

Based on the scores of the Test Set, we can not confirm which method performs best.

- Same Test Set scores may be due to the small test sample size (18 samples). Therefore, we tested all methods based on the whole Dataset.
- The scores of the whole Dataset confirm that the best model is the Decision Tree Model. This model has not only higher scores, but also the highest accuracy.

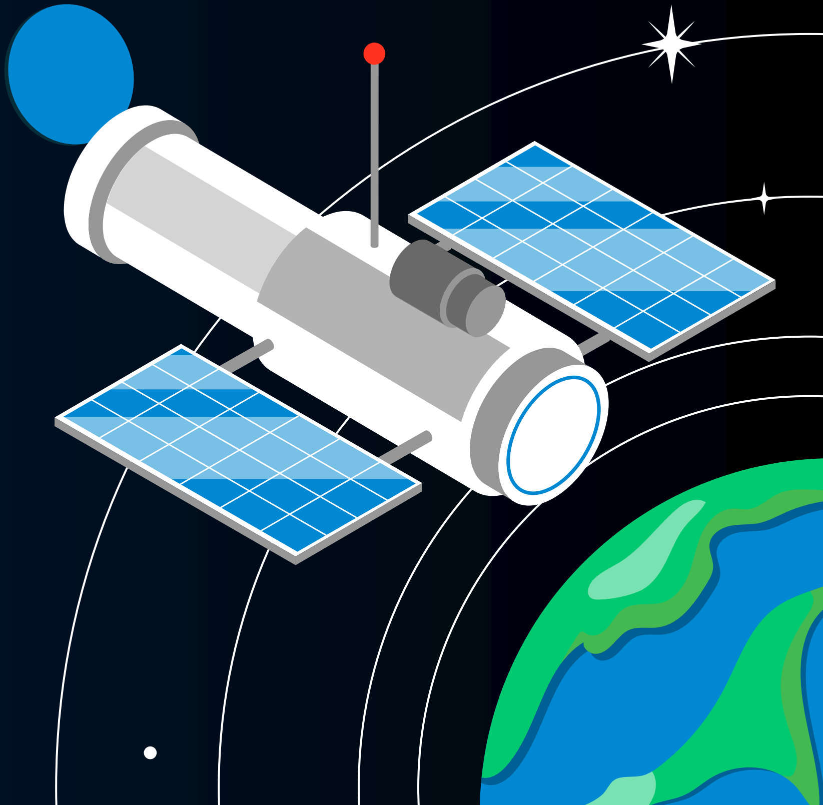
Confusion Matrix

Examining the confusion matrix, we see that logistic regression can distinguish between the different classes. We see that the major problem is false positives.



05

Conclusion



Conclusion

1- The decision Tree Model is the best algorithm for this dataset

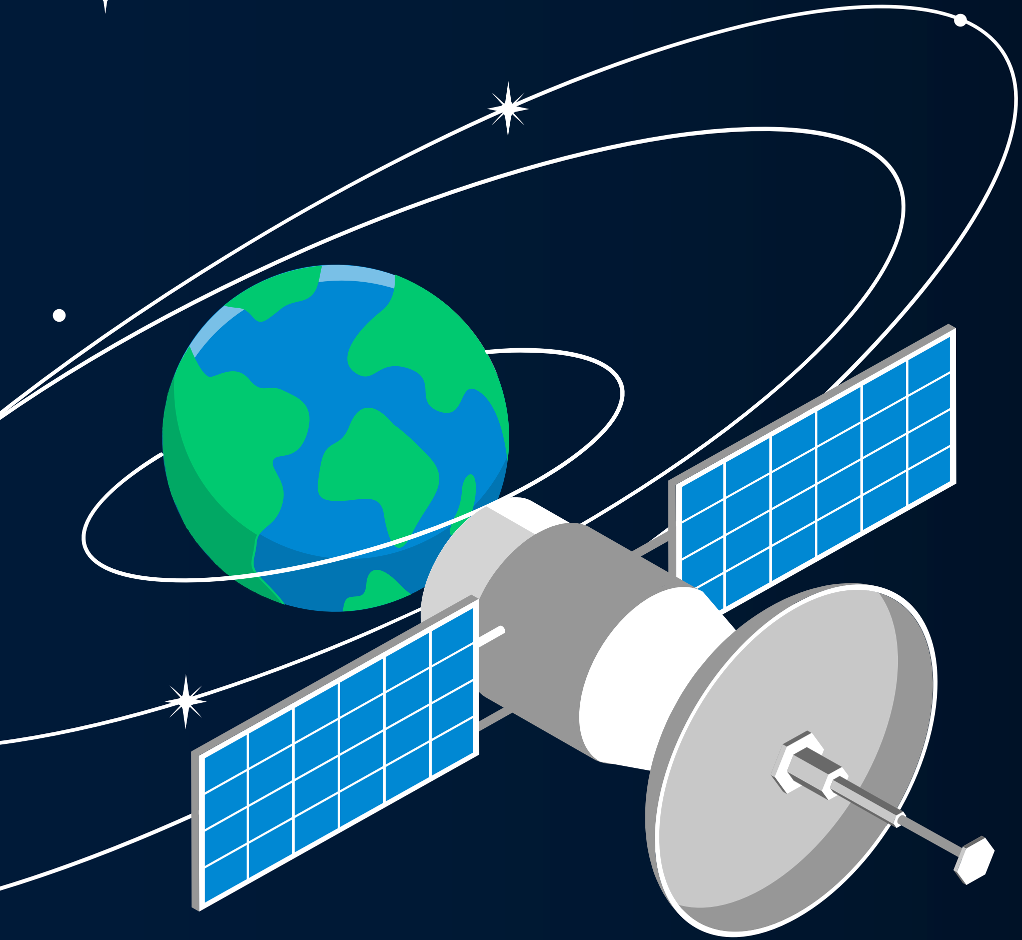
2- Most of the launch sites are in proximity to the Equator Line and all the sites are in very close proximity to the coast.

Conclusion

3- The success rate increased over the years.

4- KSC LC-39A has the highest success rate.

5- Orbits ES-L1, GEO, HEO, and SSO have 100% success rate.

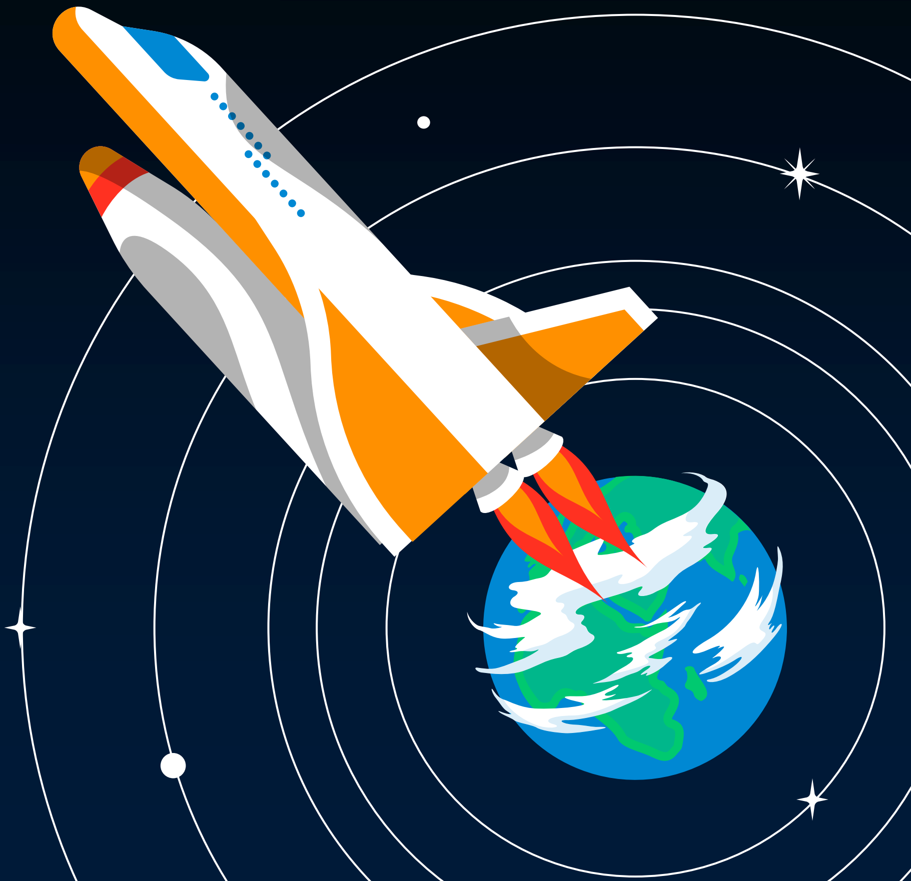


06

Acknowledgement

Acknowledgement

IBM
Coursera
Slidesgo



Thank You

Elia Alhanach
8th of May 2024