IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Maurizio PAPI
January 17th 2025

Outline

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

Executive Summary

- The commercial space transportation sector is significantly impacted by the high costs associated with each individual launch. SpaceX has disrupted the industry by introducing an innovative approach that involves reusing the first stage of the Falcon 9 rocket, drastically reducing launch costs.
- This project aims to predict the success of a Falcon 9 launch. To achieve this, data collected online will undergo preparation (data wrangling) and exploratory data analysis. The processed data will then be used to train a machine learning algorithm, resulting in a predictive model capable of estimating the likelihood of a successful launch.

Introduction

- SpaceX, founded by Elon Musk in 2002, has revolutionized space travel and commercial space transportation. The company's primary goal is to reduce space transportation costs and enable the colonization of Mars. SpaceX's innovative approach includes the development of reusable rockets, such as the Falcon 9, which has significantly lowered the cost of satellite launches. This cost-effectiveness has made space more accessible to a wider range of customers, including commercial, government, and scientific entities.
- In this project, we will predict if the Falcon 9 first stage will land successfully. 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.

Section 1

Methodology

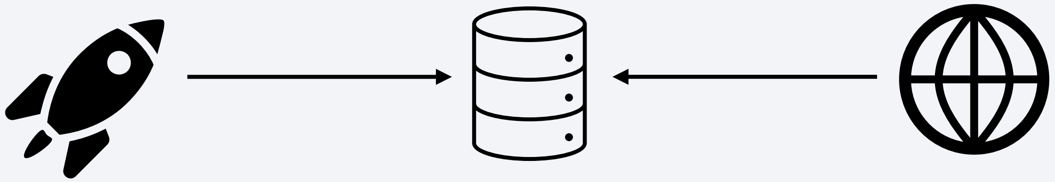
Methodology

Executive Summary

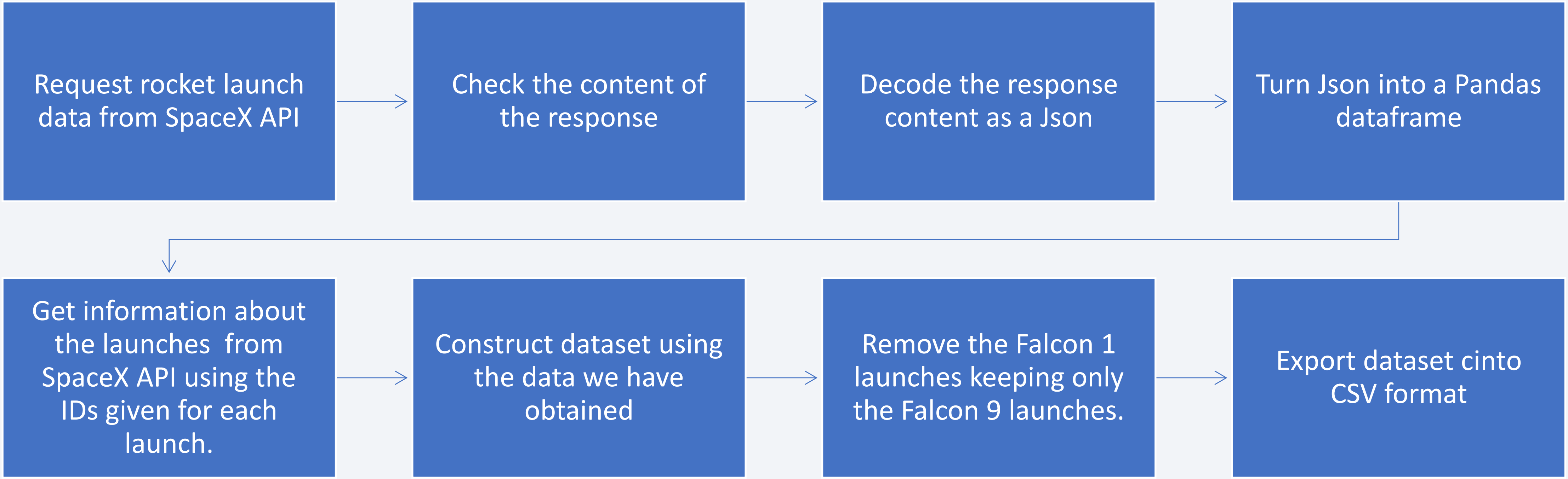
- Data collection methodology:
  - Describe how data was collected
- Perform data wrangling
  - Describe how data was processed
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection

- we will be working with SpaceX launch data that is gathered from an API, specifically the SpaceX REST API. This API will give us data about launches, including information about the rocket used, payload delivered, launch specifications, landing specifications, and landing outcome: https://github.com/r-spacex/SpaceX-API/tree/master
- Another popular data source for obtaining Falcon 9 Launch data is web scraping related Wiki pages: https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches
- These two data sources have the advantage of being free but at the same time having a decent accuracy.



Data Collection – SpaceX API



See the GitHub URL of the completed SpaceX API calls notebook:

https://github.com/MaurizioPapi/IBM_Data_Science/blob/main/01-jupyter-labs-spacex-data-collection-api.ipynb

Data Collection – Scraping

Perform an HTTP GET method to request the Falcon9 Launch HTML page

Print the page title to verify if the object was created properly

find all tables on the wiki page: third table is our target table

extract column name.

create an empty dictionary with keys from the extracted column names in the previous task.

fill up the dictionary with launch records extracted from table rows

Create a DataFrame from the dictionary.

Export dataset cinto CSV format

See the GitHub URL of the completed SpaceX web scraping notebook:

https://github.com/MaurizioPapi/IBM_Data_Science/blob/main/02-jupyter-labs-webscraping.ipynb

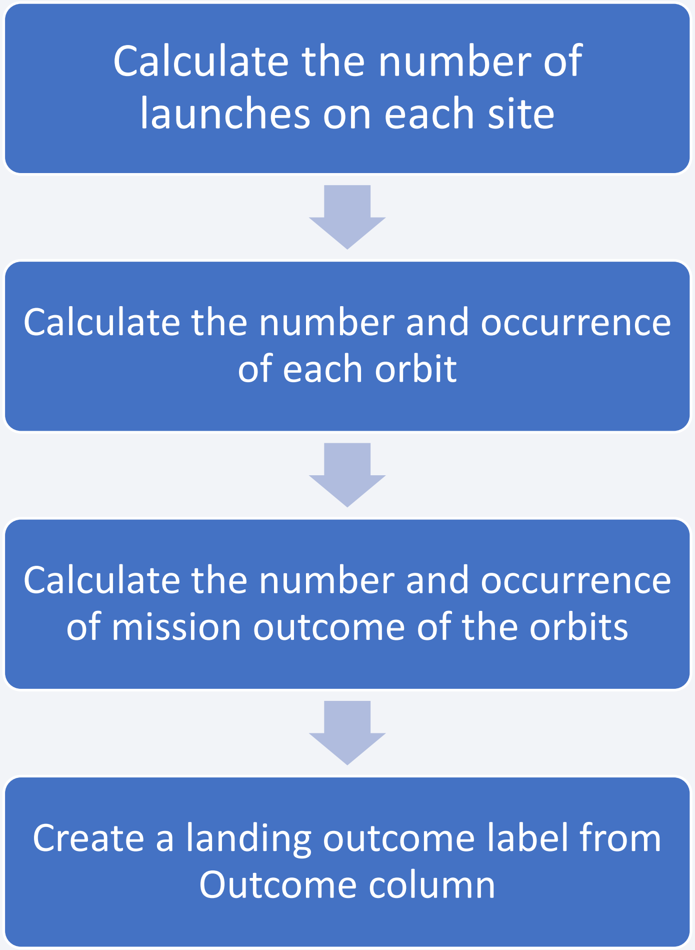
Data Wrangling

- One of the most critical phases of a Data Science project is data wrangling. Our database needs to:
    - Identify and calculate the percentage of missing values in each attribute.
    - Determine the training labels.

The diagram on the right shows the data wrangling phases adopted in this project.

See the GitHub URL of the completed data wrangling notebook:

https://github.com/MaurizioPapi/IBM_Data_Science/blob/main/03-labs-jupyter-spacex-Data%20wrangling.ipynb



EDA with Data Visualization

Data visualization through charts and figures is of great help for Exploratory Data Analysis and Feature Engineering.

See the GitHub URL of the completed data visualization notebook:

https://github.com/MaurizioPapi/IBM_Data_Science/blob/main/05-jupyter_labs_eda_dataviz_v2.ipynb

EDA with SQL

After the data cleaning phase, the dataset is saved in an SQL database.

Below, we list the main tasks performed on our database:

- Display the names of the unique launch sites in the space mission
- Display average payload mass carried by booster version F9 v1.1
- List the date when the first successful landing outcome in ground pad was achieved.
- List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000.
- List the names of the booster_versions which have carried the maximum payload mass.

See the GitHub URL of the completed EDA with SQL notebook:

https://github.com/MaurizioPapi/IBM_Data_Science/blob/main/04-jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

Here's a summary of the map objects I created and added to a Folium map, along with the reasons for adding them:

- Circle area with a text label, to represent all launch sites on the map.
- Markers of successful / failed launches for each site on the map, to have a visual success rate.
- Lines to connect launch sites to their proximities, to explore railways, highways, coastlines and cities.

See the GitHub URL of the completed interactive map notebook:

https://github.com/MaurizioPapi/IBM_Data_Science/blob/main/06-lab-jupyter-launch-site-location-v2.ipynb

Build a Dashboard with Plotly Dash

An interactive dashboard is highly useful for analyzing and exploring the data in our dataset. In our case, the dashboard features:

- a dropdown menu that allows the selection of the launch site
- a pie chart of the rate of successes
- a scatter plot for payload mass and rate of success.

See the GitHub URL of the completed Plotly Dash notebook:

https://github.com/MaurizioPapi/IBM_Data_Science/blob/main/07_spacex_dash_app.ipynb

Predictive Analysis (Classification)

The final step is training models capable of predicting the success or failure of launches.

After splitting the dataset into training and test sets, we tested the following algorithms: Logistic Regression; SVM; Decision Tree; KNN. Then, we compared the results on the test set.

Optimization was performed to find the best parameters for each model.

See the GitHub URL of the completed predictive analysis notebook:

https://github.com/MaurizioPapi/IBM_Data_Science/blob/main/08_SpaceX_Machine_Learning_Prediction_Part_5_v1.ipynb

- create a column for the class
- Standardize the data
- Split into training data and test data
- Find best Hyperparameter for SVM, Classification Trees and Logistic Regression
- Find the method performs best using test data

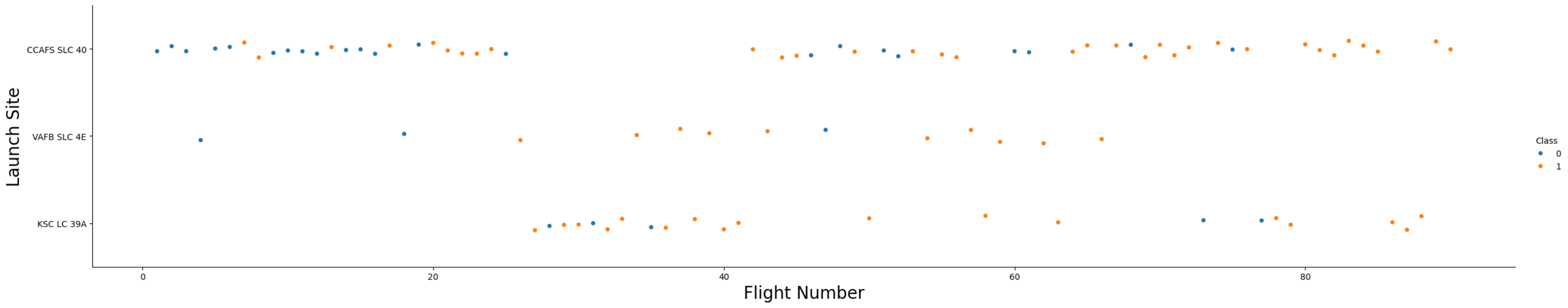
Results

- From the data exploration, we can assert that the success of missions significantly improved from 2014 to 2015, thanks to SpaceX's increased experience in recovering first stages.
- The ability to launch from multiple sites and the variety of missions, along with the capability to place payloads into many orbits, demonstrate SpaceX's flexibility in meeting its clients' needs.
- The predictive models are able to forecast mission success with a good degree of accuracy (around 83%).

Section 2

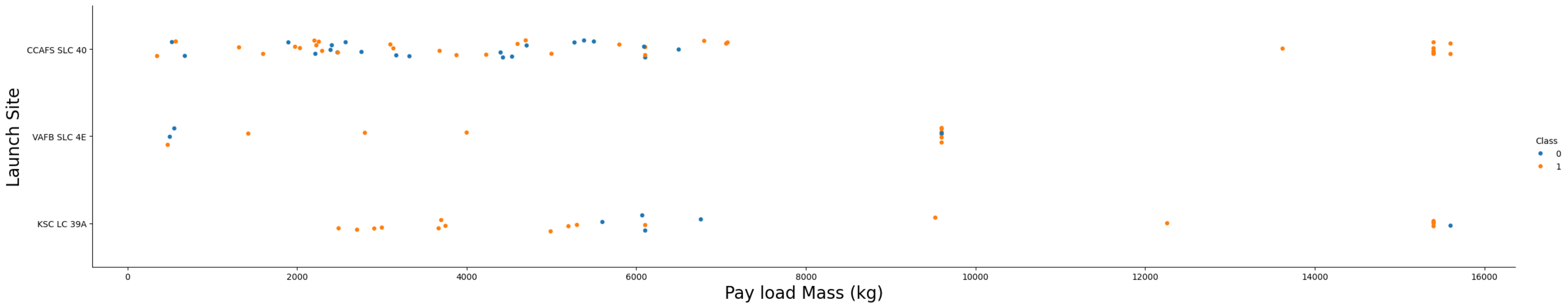
Insights drawn from EDA

Flight Number vs. Launch Site



- The scatter plot shows the launch sites for each flight number.
- The majority of launches occurred at SLC 40.
- The use of LC 39A is limited, and the Vandenberg site is used on few occasions.

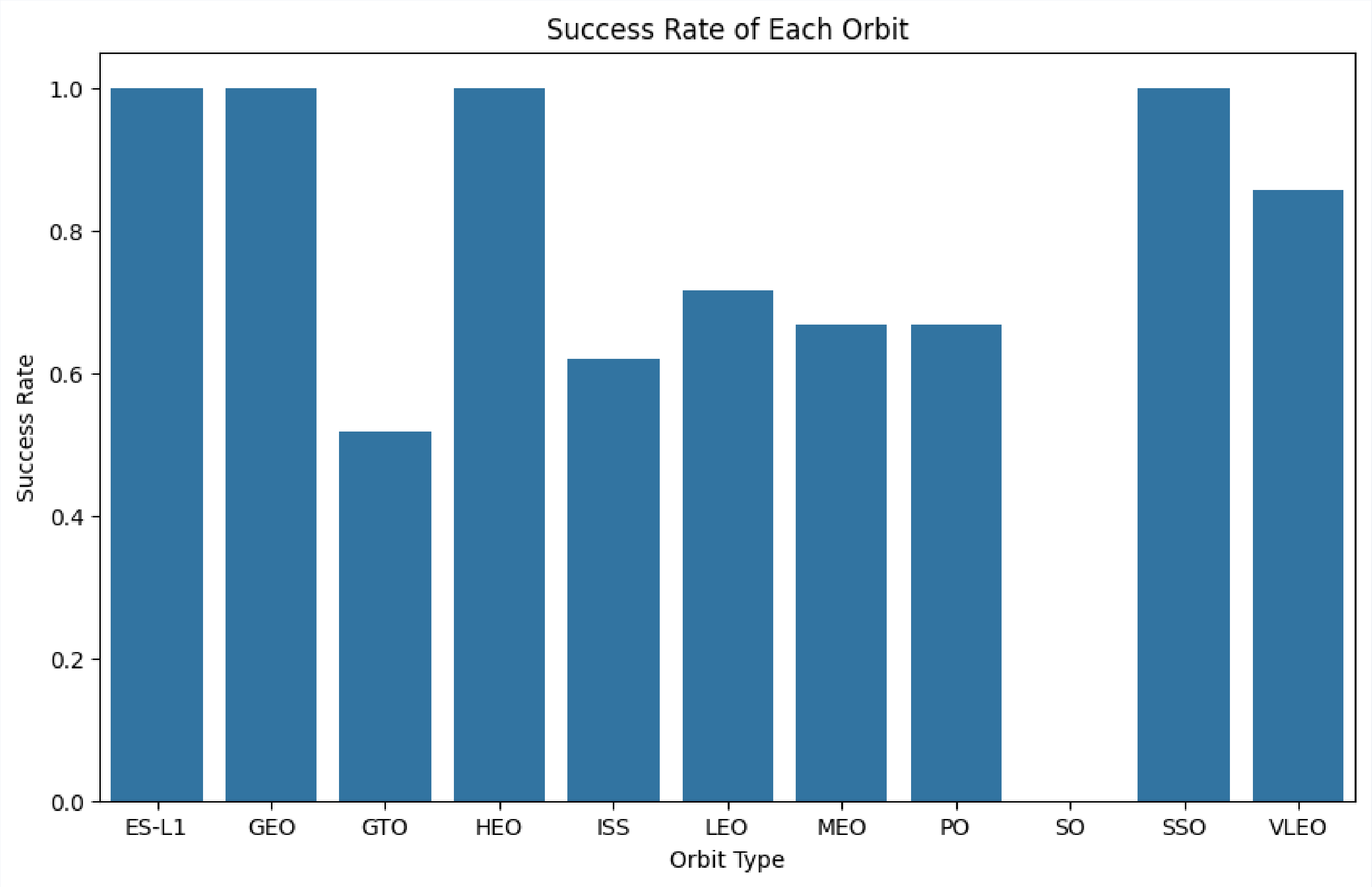
Payload vs. Launch Site



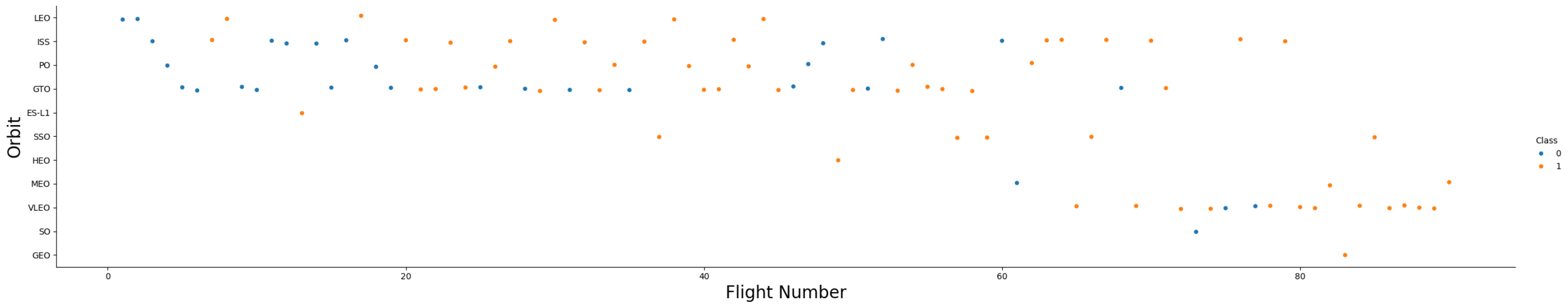
- The scatter plot shows the payload mass for each launch site.
- Payload mass higher than 10°000Kg have been launched only from Cape Canaveral.

Success Rate vs. Orbit Type

- The chart bar shows the success rate for each orbit type.
- Highest success rat are for ES-L1, GEO, HEO an SSO.



Flight Number vs. Orbit Type



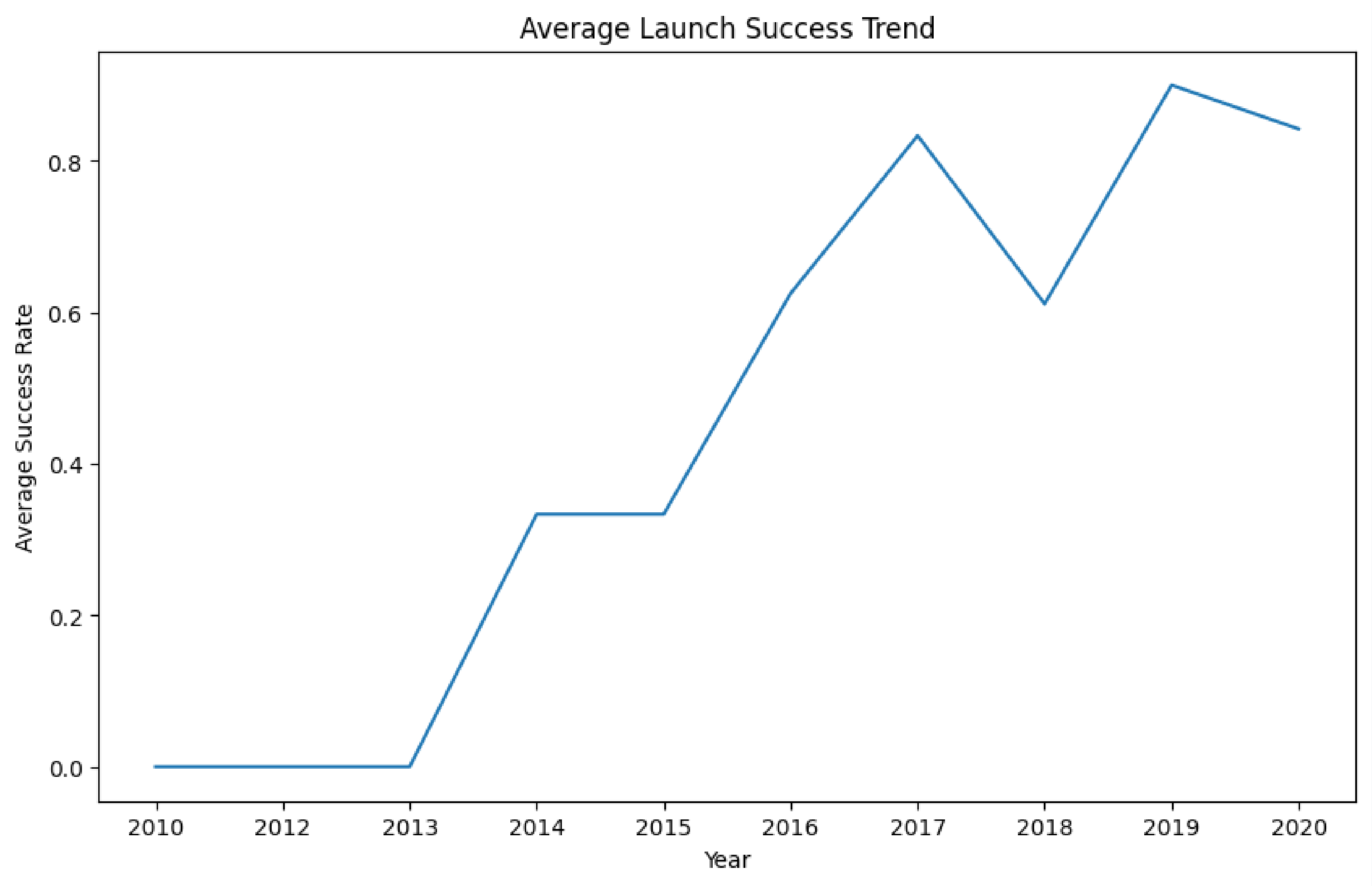
- The scatter plot shows the Orbit Type for each Flight Number.
- Most successful orbit types have few flights: their high success rate is not statistically significant.

Payload vs. Orbit Type

- The scatter plot shows the Orbit Type for Payload Mass.
- There is no clear correlation between the success rate and the payload mass.

Launch Success Yearly Trend

- The line chart shows the average launch success trend for each years.
- Since the 2014-2015 period, the level of reliability has significantly increased, confirming SpaceX's growing expertise in recovering first stages.



All Launch Site Names

- The names of the unique launch sites are:
  - CCAFS LC-40
  - VAFB SLC-4E
  - KSC LC-39A
  - CCAFS SLC-40

Launch Site Names Begin with 'CCA'

- 5 records where launch sites begin with `CCA` are:

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

- They are all related to LEO missions to ISS.

Total Payload Mass

- The total payload carried by boosters from NASA is:

619°967 Kg

Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1 is:

2°535 Kg

First Successful Ground Landing Date

- The dates of the first successful landing outcome on ground pad is December 22nd 2015
- Here below the flight mission information:

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

Successful Drone Ship Landing with Payload between 4000 and 6000

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

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2016-05-06	5:21:00	F9 FT B1022	CCAFS LC-40	JCSAT-14	4696	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
2016-08-14	5:26:00	F9 FT B1026	CCAFS LC-40	JCSAT-16	4600	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
2017-03-30	22:27:00	F9 FT B1021.2	KSC LC-39A	SES-10	5300	GTO	SES	Success	Success (drone ship)
2017-10-11	22:53:00	F9 FT B1031.2	KSC LC-39A	SES-11 / EchoStar 105	5200	GTO	SES EchoStar	Success	Success (drone ship)

- These missions were all carried out during the period with a good success rate (after 2015).

Total Number of Successful and Failure Mission Outcomes

- The total number of successful and failure mission outcomes are:
  - 100 Success missions
  - 1 Failure missions
- Success mission outcomes are very high, with a failure rate <1%.

Boosters Carried Maximum Payload

- The names of the booster which have carried the maximum payload mass are:
  - 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
- All them all carried 15°600 Kg of mass payload.

2015 Launch Records

- The failed landing_outcomes in drone ship, with their booster versions and launch site names in year 2015, are:

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

- These missions took place in the first part of the year.

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:

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

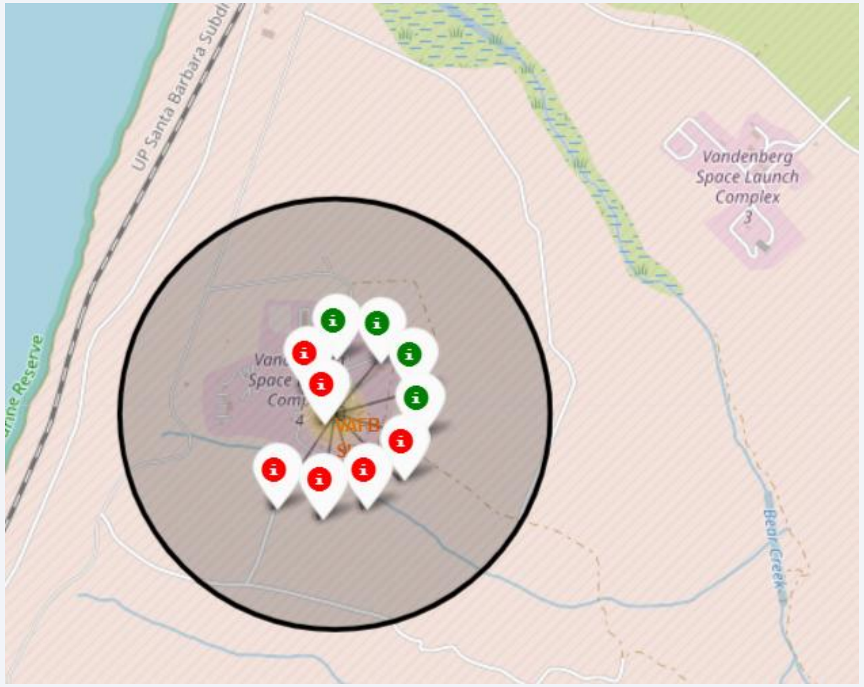
Section 3

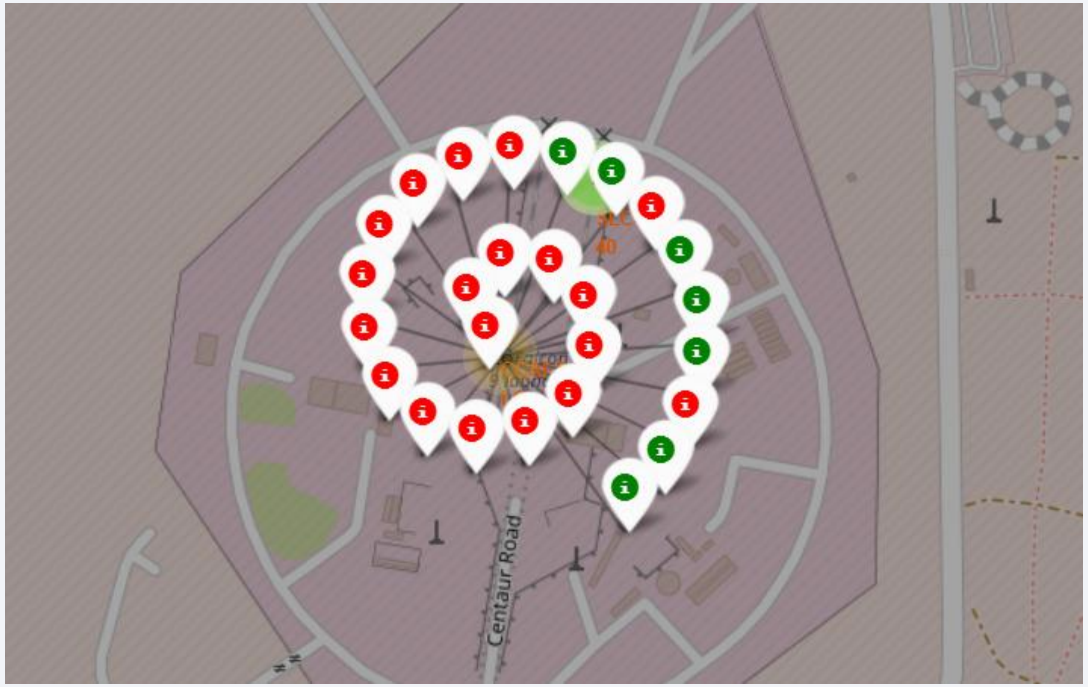
Launch Sites Proximities Analysis

Space X Falcon 9 - Launch Sites

- 4 launch sites, but there are two places: Cape Kennedy in Florida and Vandenberg in California.

Space X Falcon 9 - Launch Outcomes

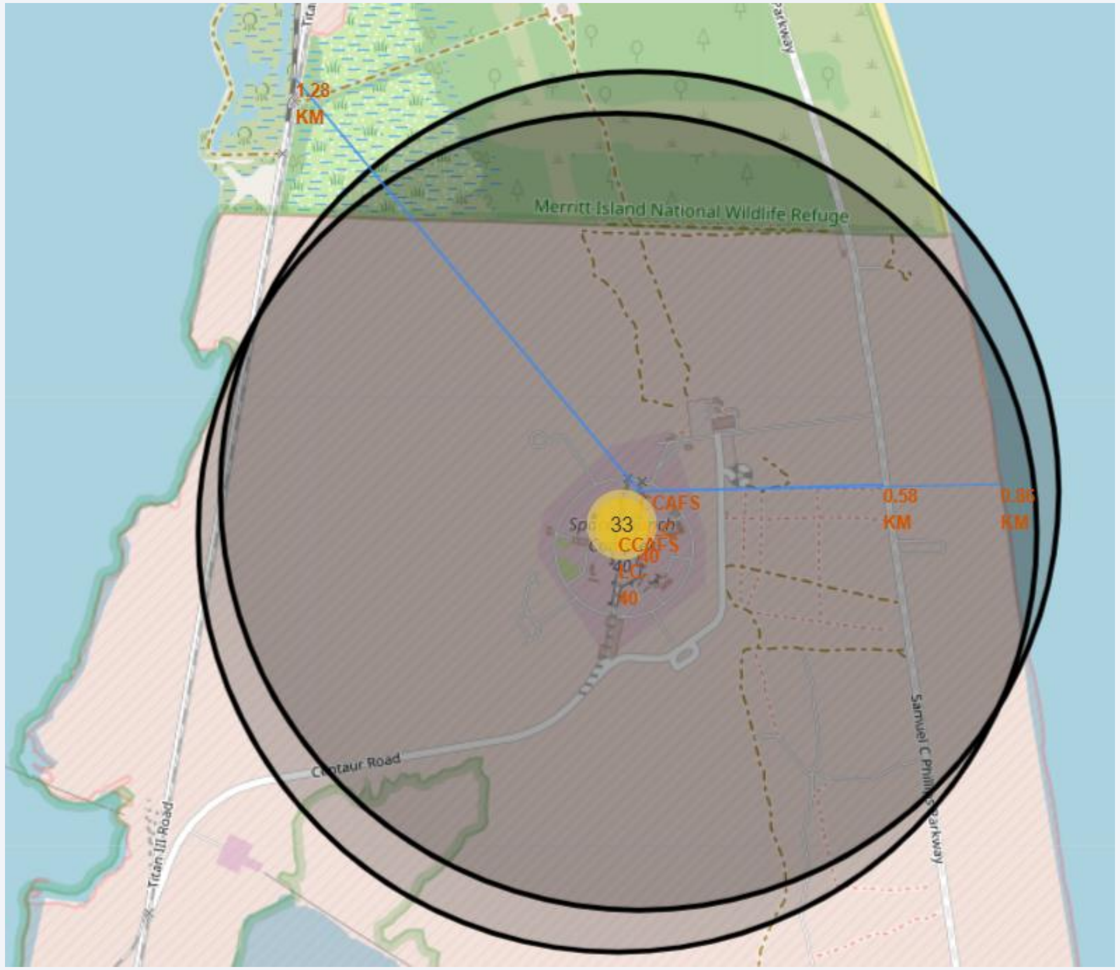




- The figures show the launch outcomes for the 2 main launch sites: Vandenberg (left) and SLC-40 at Cape Kennedy.

Space X Falcon 9 - Proximities

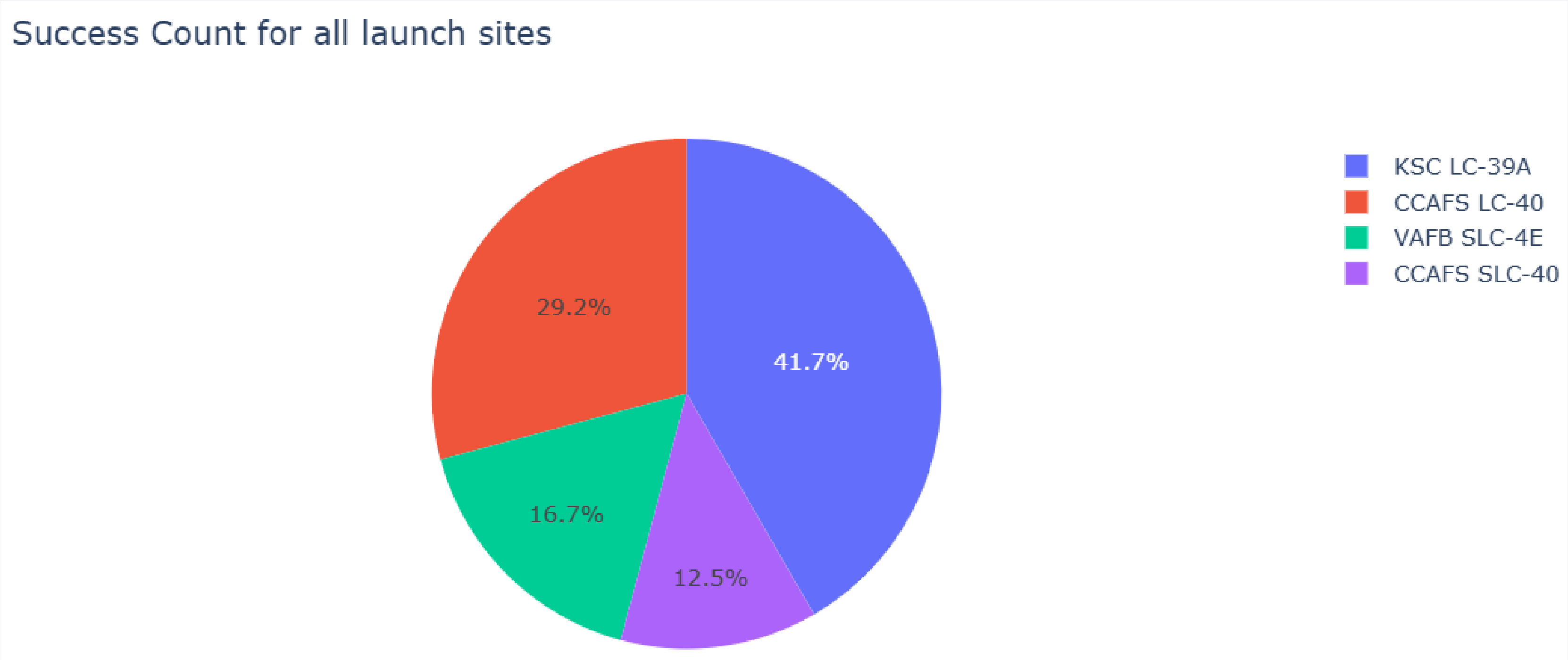
- Launch sites are located far from crowded places and strategic for infrastructure
- In the image on the right, you can see how the Cape Kennedy launch complex is located some distance from the coast and the nearest train station.



Section 4

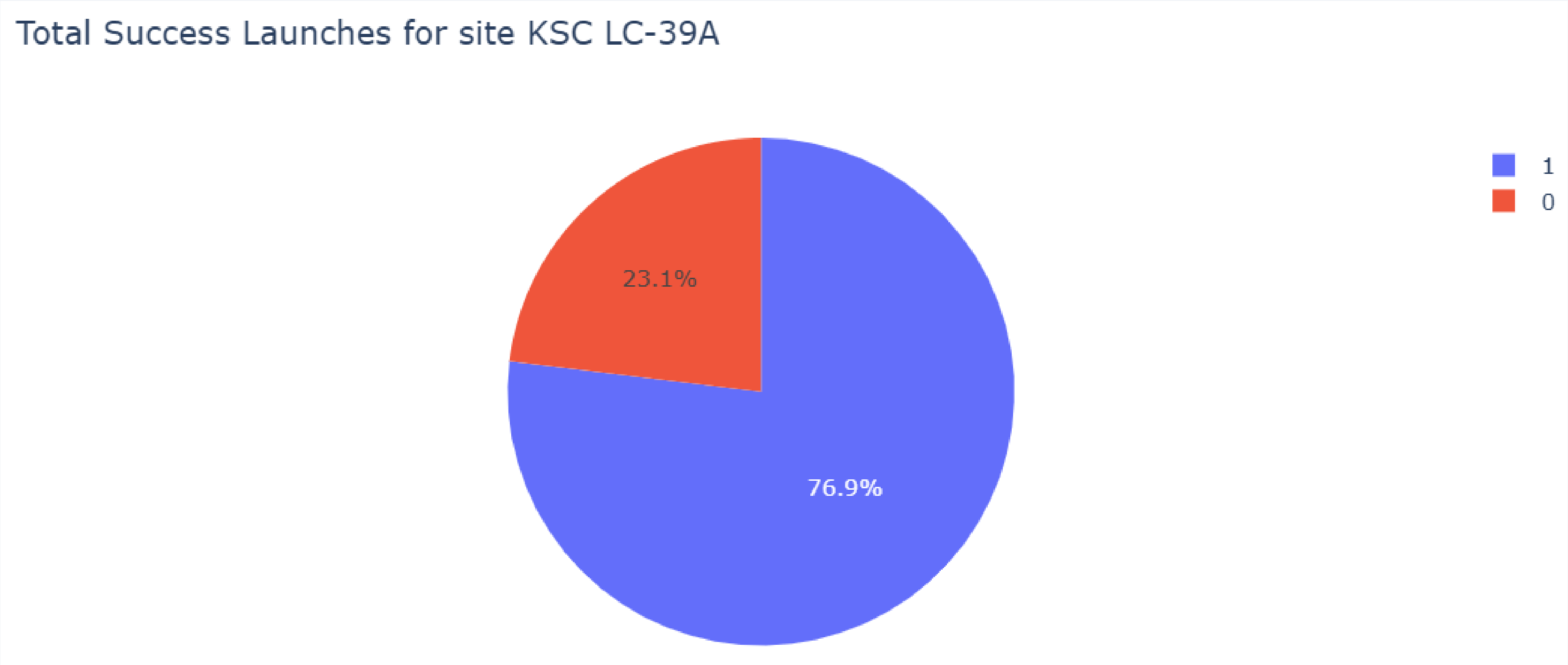
Build a Dashboard with Plotly Dash

Success Count - Launch Sites



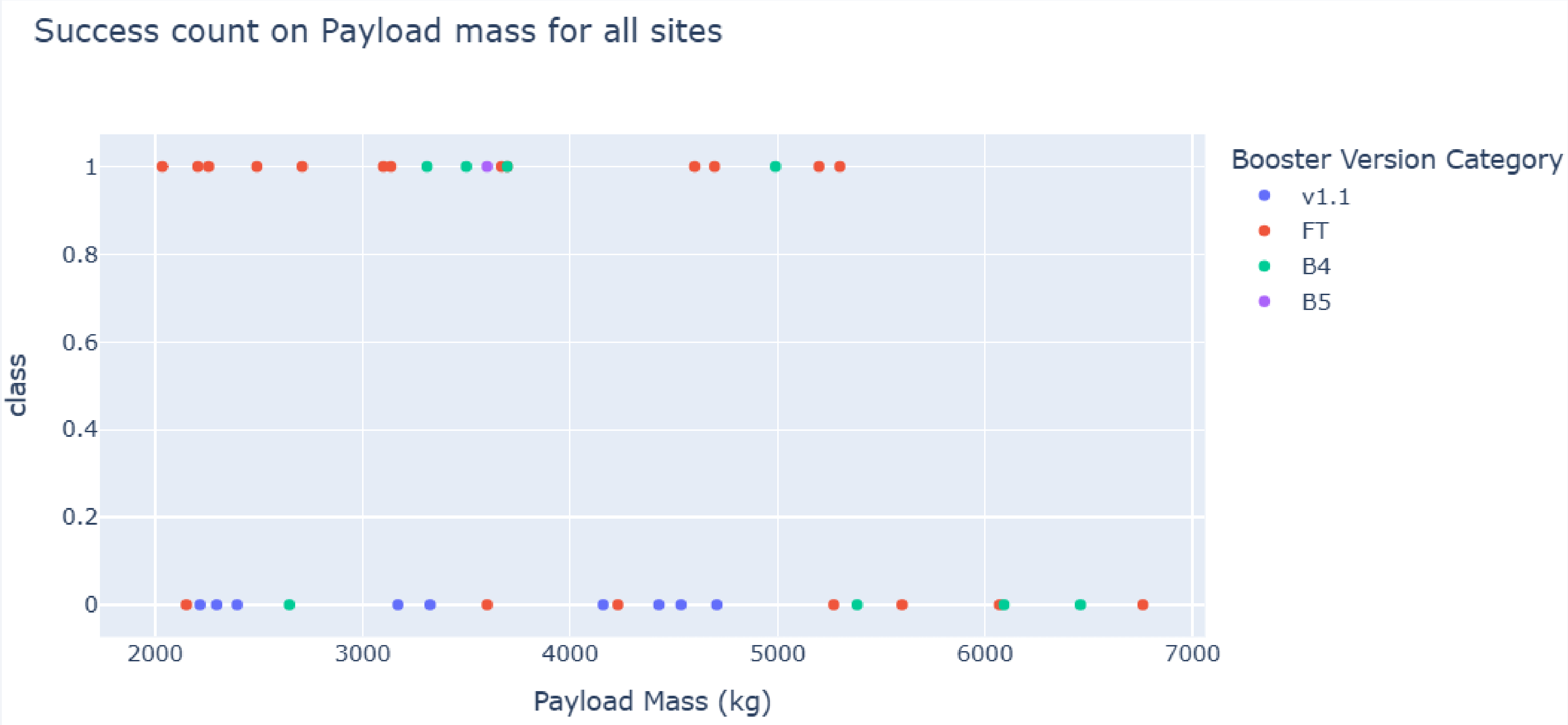
- Most of the successful launches have been made at Cape Kennedy.

Success Count - KSC LC-39A



- The success rate for the KSC LC-39A site is high: the reason is that the site was not used in the first years of service of the Falcon 9 – years with the lowest reliability.

Success Count on Payload Mass - All Sites



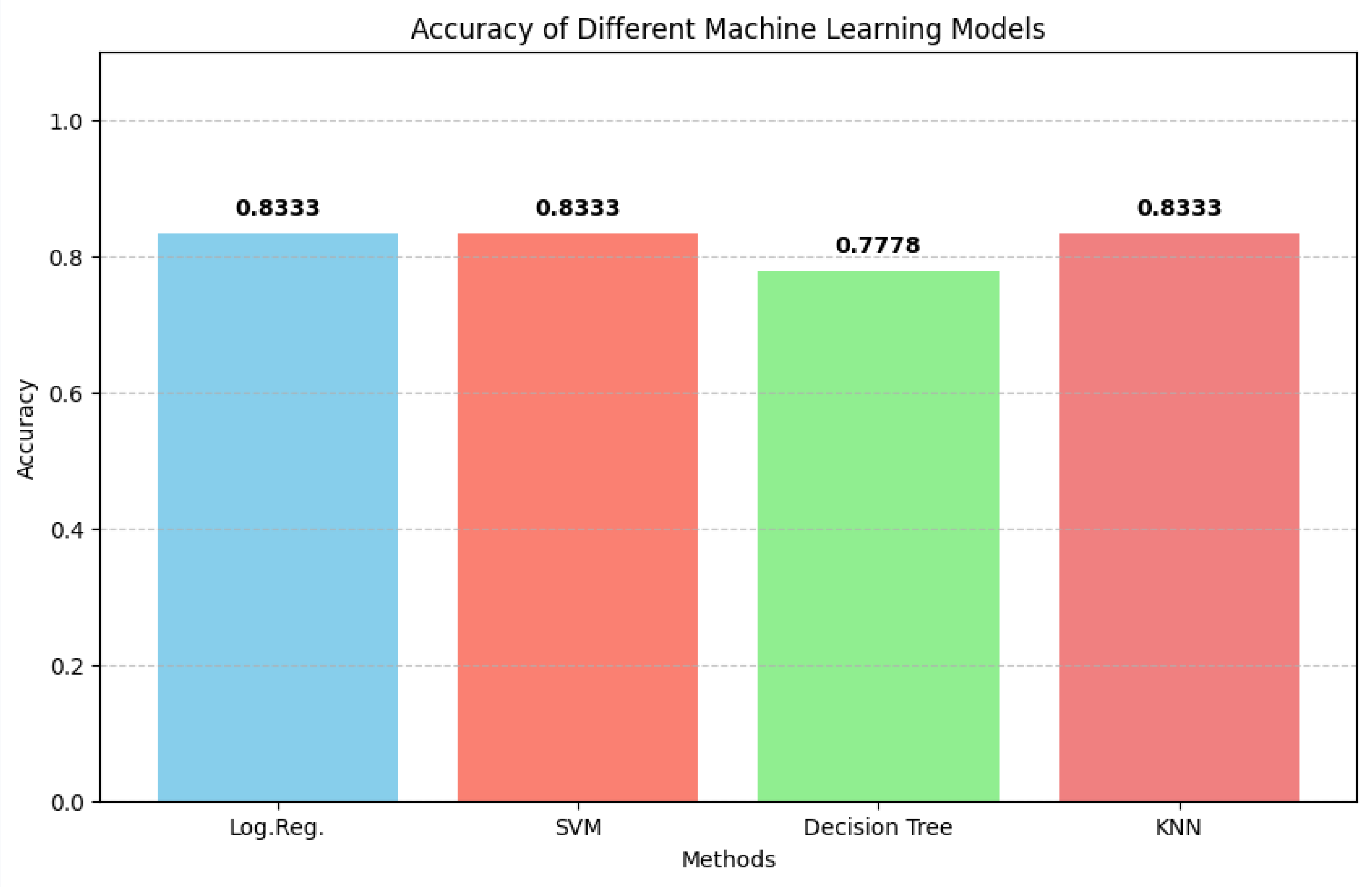
- From 2000 to 7000 kg of payload, the success rate is not high. The best version is the FT (Full Thrust), the first version of the Falcon 9 to successfully land.

Section 5

Predictive Analysis (Classification)

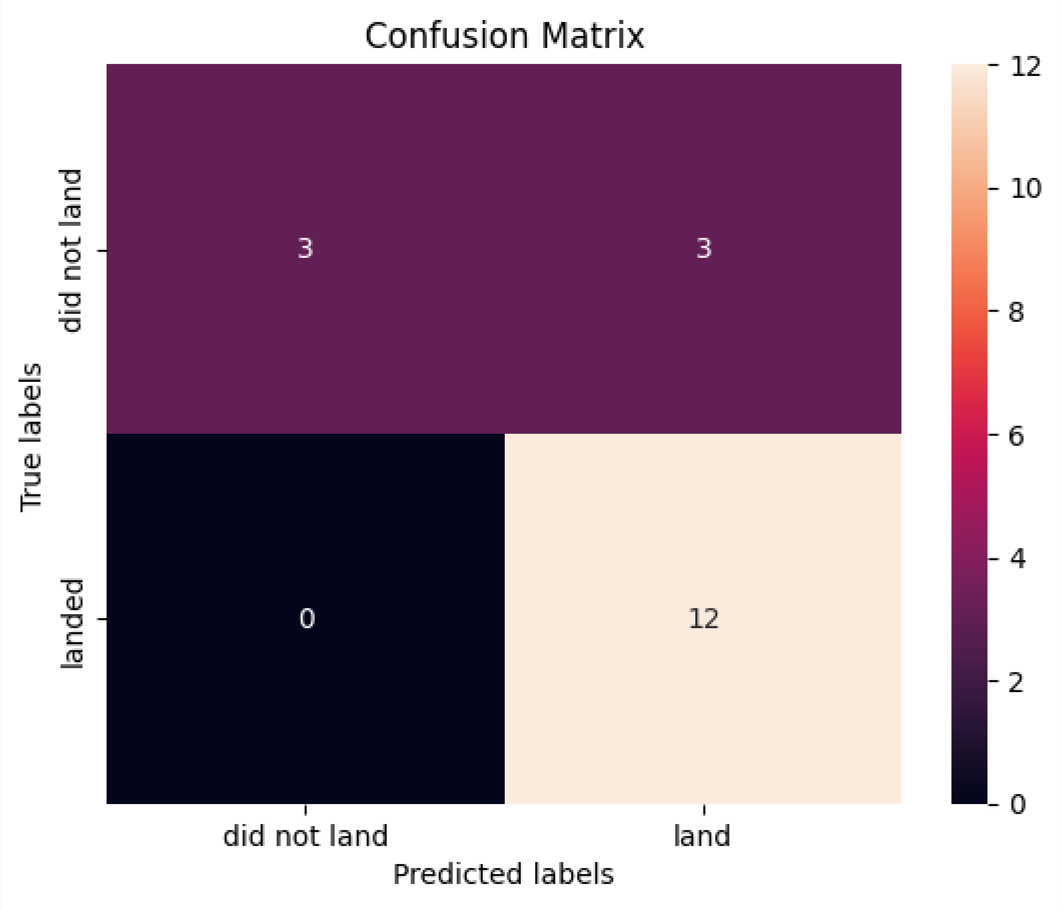
Classification Accuracy

- All models perform well and are aligned. Only the Decision Tree scored slightly lower.



Confusion Matrix

- The figure shows the Confusion Matrix for the three models with the best score (Logistic Regression, SVM, KNN).
- The models are able to predict all successful missions, while showing difficulty in determining unsuccessful ones (high precision, low recall).



Conclusions

- The annual trend of successful launches shows a significant improvement from 2014 to 2015, due to SpaceX's increased expertise in the recovery of first-stage rockets.
- The ability to launch from multiple sites and the diversity of missions, along with the ability to place payloads in various orbits, underscore SpaceX's versatility in responding to the needs of its customers.
- The predictive models used produced similar scores. A high score can be explained by the fact that the test set employed is skewed in favor of successful throws.

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!