Space Rocket Launch Price Analysis for SpaceY

Deborah Leem

15 August 2022

Outline

Executive Summary

Background & context

Due to the ability to reuse the fist stage, SpaceX's costs for rocket launches are significantly lower than other providers.

- SpaceX's Falcon 9 rocket launch costs $62MM
- Other competitors cost upwards of $165MM

This project predicts if the first stage will land using a machine learning model to determine the cost of a launch.

Processes

- Data collection: ApaceX REST API & web scrap Falcon 9 launch records with 'Beautiful Soup'
- Data wrangling
- Exploratory Data Analysis: SQL & Visualisation
- Dashboarding: Folium & PlotlyDash
- Predictive Analysis: find best Hyperparameter for SVM, Classification Trees, KNN and Logistic Regression

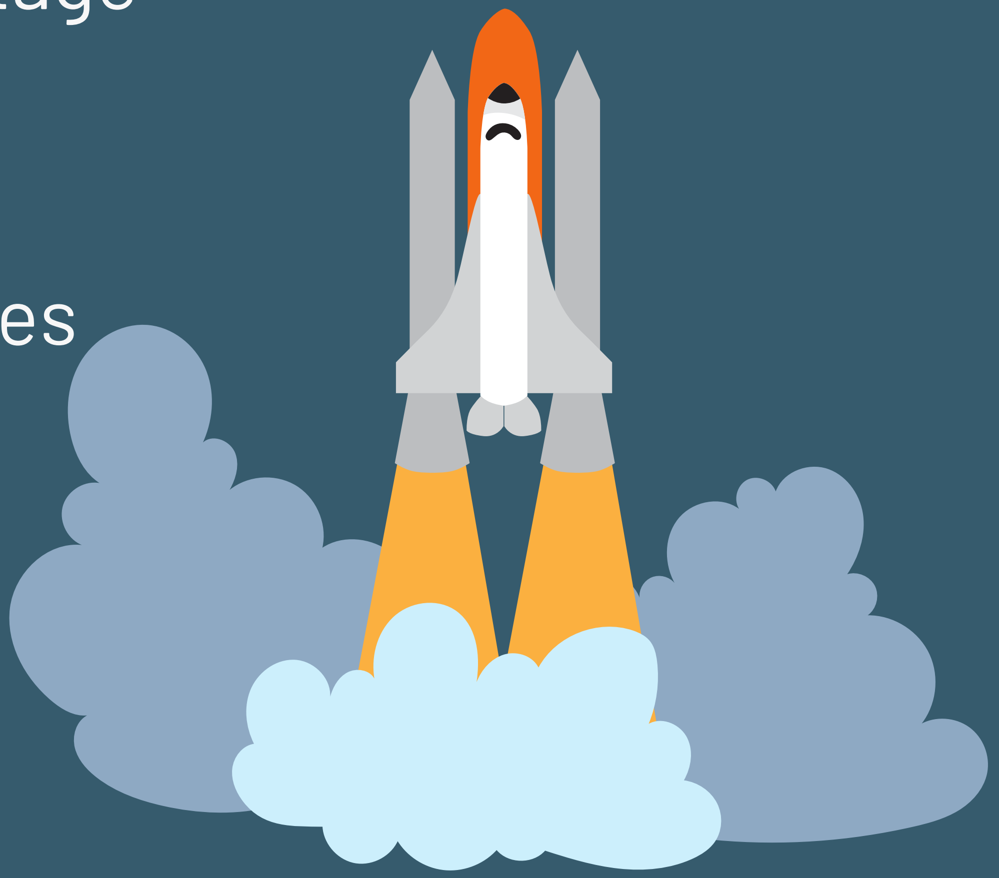
Findings

- There is a clear trend of success rate increase over time
- Launch site KSV LC-39A has the highest success rate but payload above 5500kg has negative impact on its success outcome
- CCAFS SLC 40 has the highest number of launches
- As the flight number increases, the first stage is more likely to land successfully
- The payload mass is also an important factor and it seems the more massive the payload, the less likely the first stage will return
- Majority of lower payloads launched from CCAFS SLC 40
- Orbits ES-L1, GEO, HEO, SSO and VLEO have the highest success rate
- Launch sites are also in close proximity to railways and highways for easy access
- Launch sites are not in close proximity to cities for safety reasons
- The best ML method is Decision Tree Classifier

Introduction

Overview of the project

- SpaceX's Falcon 9 rocket launch costs $62MM
- Other competitors cost upwards of $165MM
- Cost is significantly lower as the first stage can be reused by SpaceX
- Using a machine learning model, this project predicts if the first stage will land successfully
- The price of each launch is then calculated
- Information is used to keep SpaceY competitive for rocket launches



Methodology

Methods overview

01 Data Collection

- ApaceX REST API
- Web scrap Falcon 9 launch records with 'Beautiful Soup'
- Deal with missing values

02 Data Wrangling

- Perform Exploratory Data Analysis (EDA) to find patterns and label for training supervised models

03 Exploratory Data Analysis

- SQL
- Visualisation

04 Dashboarding

- Folium
- PlotlyDash

05 Predictive Analysis

- Best Hyperparameter for SVM, Classification Trees and Logistic Regression

Data Collection

- SpaceX launch data using SpaceX REST API
- Work with the endpoint api.spacexdata.com/v4/launches/past
- Web scrape Falcon 9 records from HTML tables from Wikipedia using the Python package BeautifulSoup

2020 [edit]

In late 2019, Gwynne Shotwell stated that SpaceX hoped for as many as 24 launches for Starlink satellites in 2020,[490] in addition to 14 or 15 non-Starlink launches. At 26 launches, 13 of which for Starlink satellites, Falcon 9 had its most prolific year, and Falcon rockets were second most prolific rocket family of 2020, only behind China's Long March rocket family.[491]

[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)

A table containing Falcon 9 launch data from https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches

Data Wrangling

- Wrangling data:
  - dealing with Nulls: calculate the mean of the PayloadMass data & replace the null values in PayloadMass with the mean
  - API: to obtain the missing values from some columns such as 'rocket'
  - sample data: to filter/sample the data to remove Falcon 1 launches

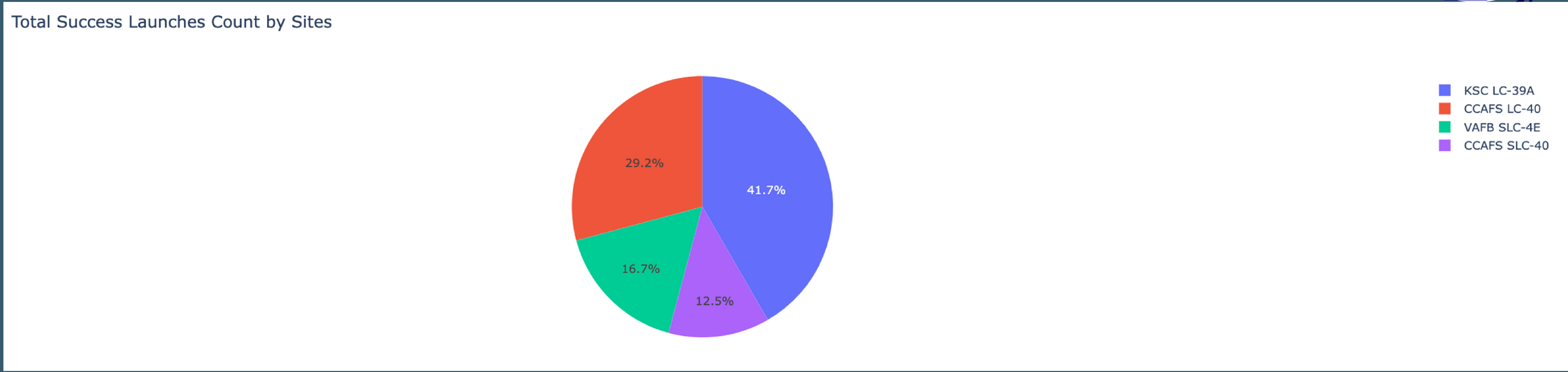
Exploratory Data Analysis with SQL

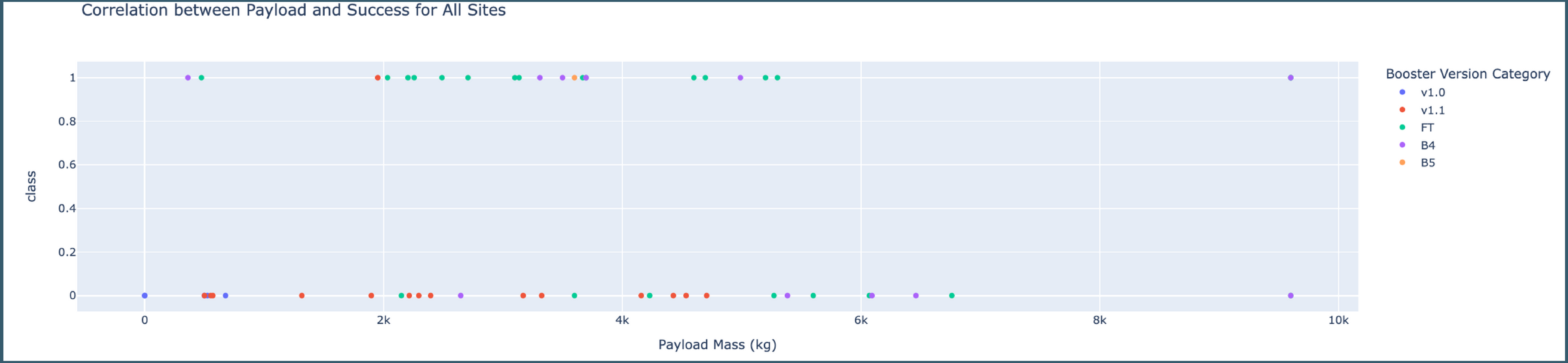
- SQL queries executed include:
  - total payload mass carried by boosters launched by NASA (CRS)
  - average payload mass carried by booster version F9 v1.1
  - the date when the first succesful landing outcome in ground pad was acheived
  - the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
  - total number of successful and failure mission outcomes
  - successful landing_outcomes between the date 04-06-2010 and 20-03-2017

Exploratory Data Analysis with Visualisation

- Try to see how the Flight Number (indicating the continuous launch attempts) and Payload variables would affect the launch outcome
- Visualise the relationship between Flight Number and Launch Site
- Visualise to observe if there is any relationship between launch sites and their payload mass
- Visually check if there are any relationship between success rate and orbit type
- Check if there is any relationship between Flight Number and Orbit type
- Visualise if there is any relationship between Payload and Orbit type
- Plot a line chart with x axis to be Year and y axis to be average success rate to get the average launch success trend

Dashboarding with Plotly Dash

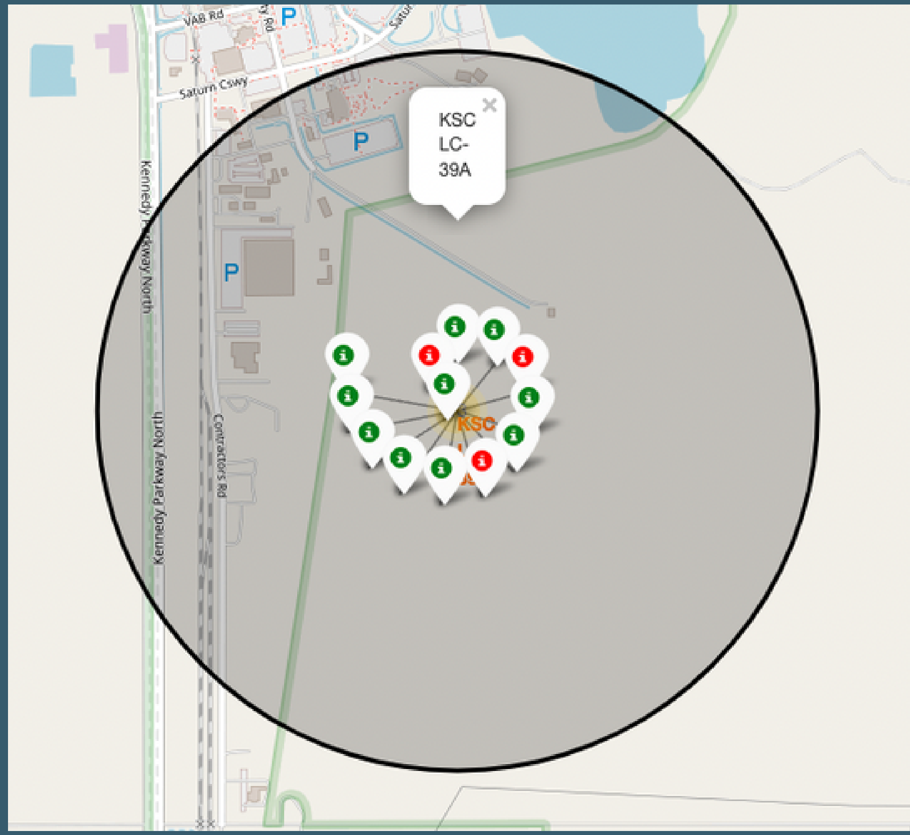




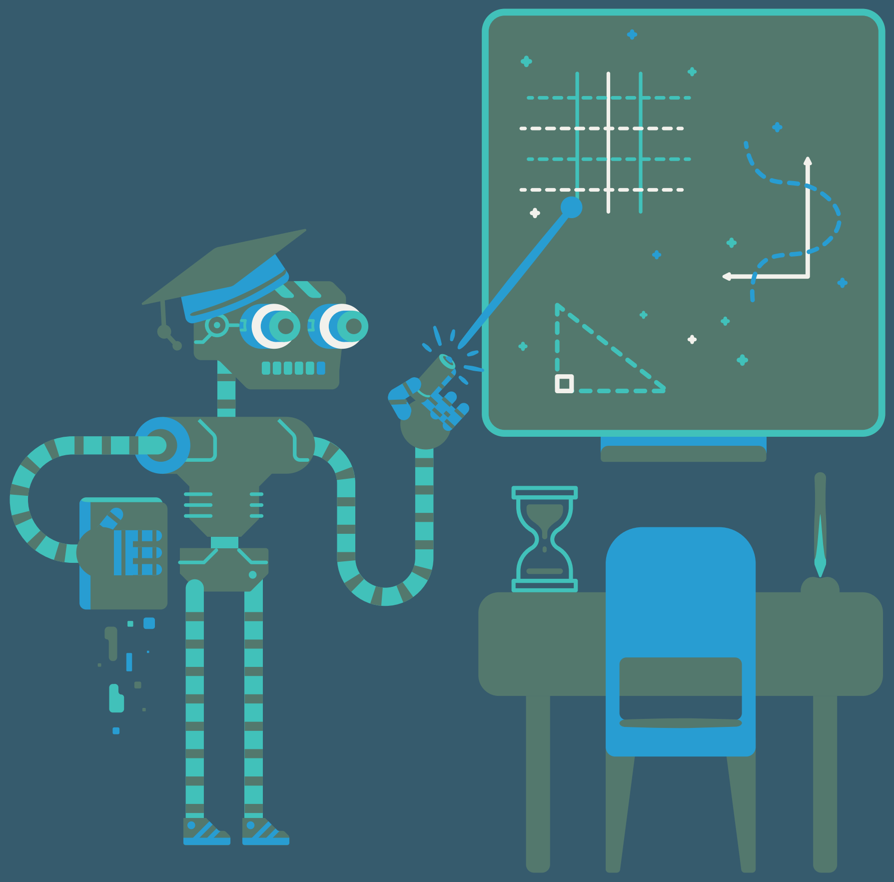
Launch Site Location Analysis with Folium

	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	-120.610746

- Interactive map with Folium
  - All launch sites are marked on the map using coordinates
  - Map enhanced with the launch outcomes (success or fail) for each launch sites to show which sites have high success rates
  - From the color-labeled markers in marker clusters, it is easy to identify launch sites have relatively high success rates .
  - Calculate the distances between a launch site and its proximities



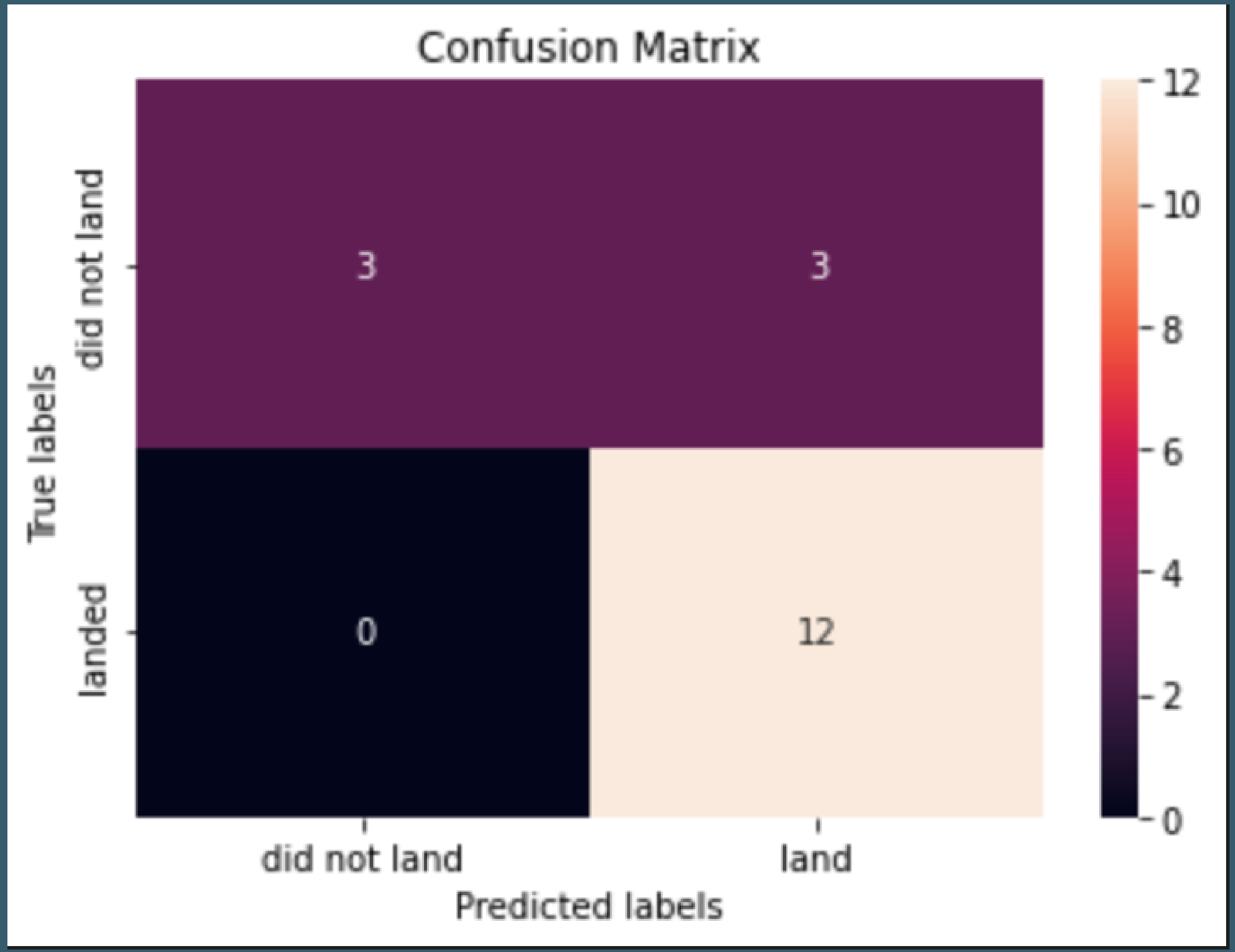
Predictive Analysis



Create a machine learning pipeline to predict if the first stage will land given the data from the preceding labs

- Perform exploratory Data Analysis and determine Training Labels
- Find best Hyperparameter for SVM, Classification Trees, Logistic Regression and K-nearest neighbour
- Find the method performs best using test data

Shape of X training set (72, 83) & Size of Y training set (72,)
Shape of X test set (18, 83) & Size of Y test set (18,)



Confusion Matrix for SVM, Classification Trees, Logistic Regression and K-nearest neighbour

Results

EDA with SQL

- The names of the unique launch sites in the space mission

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

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

22/12/2015

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

SUM(PAYLOAD_MASS__KG_)
45596

- the total number of successful and failure mission outcomes

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

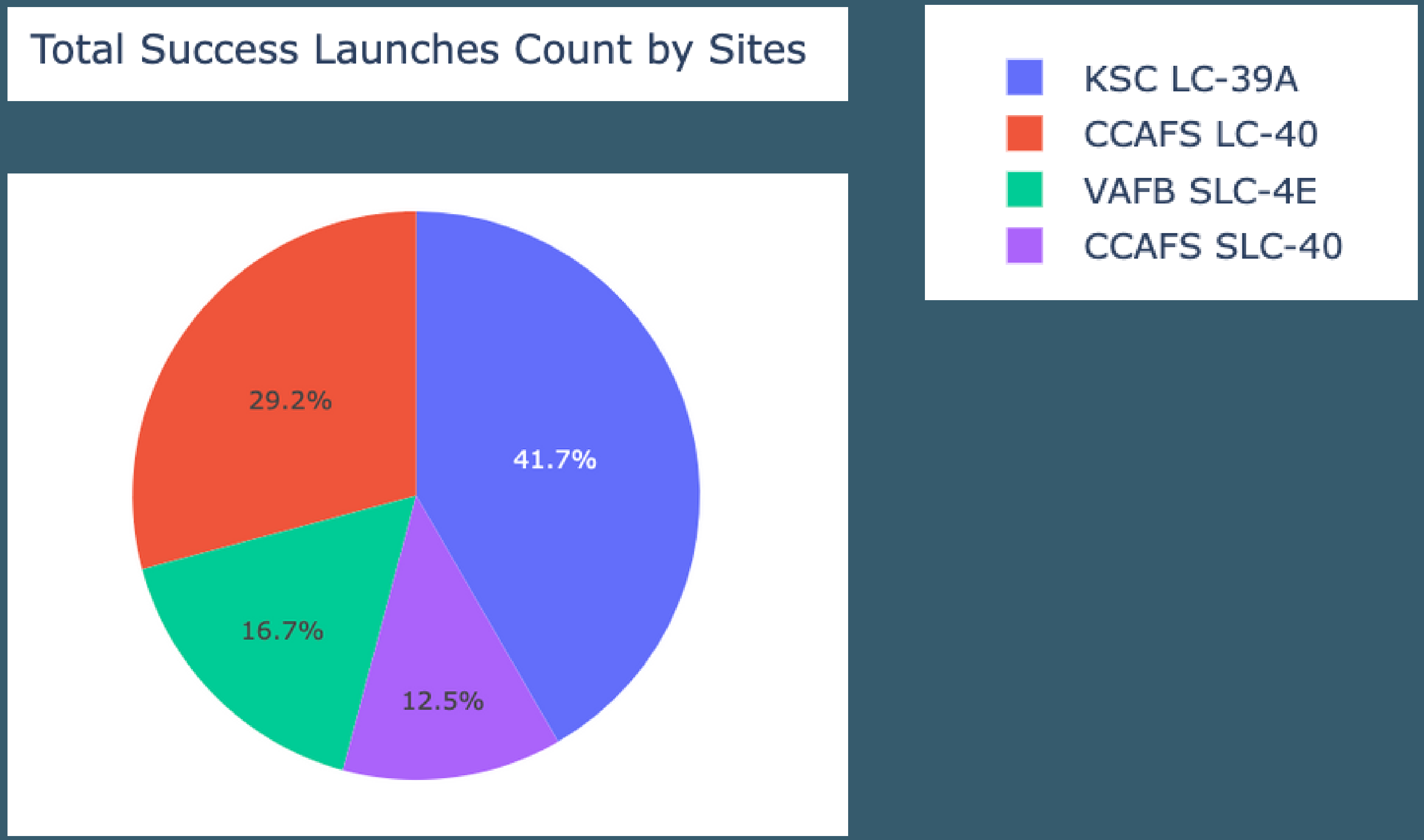
- average payload mass carried by booster version F9 v1.1

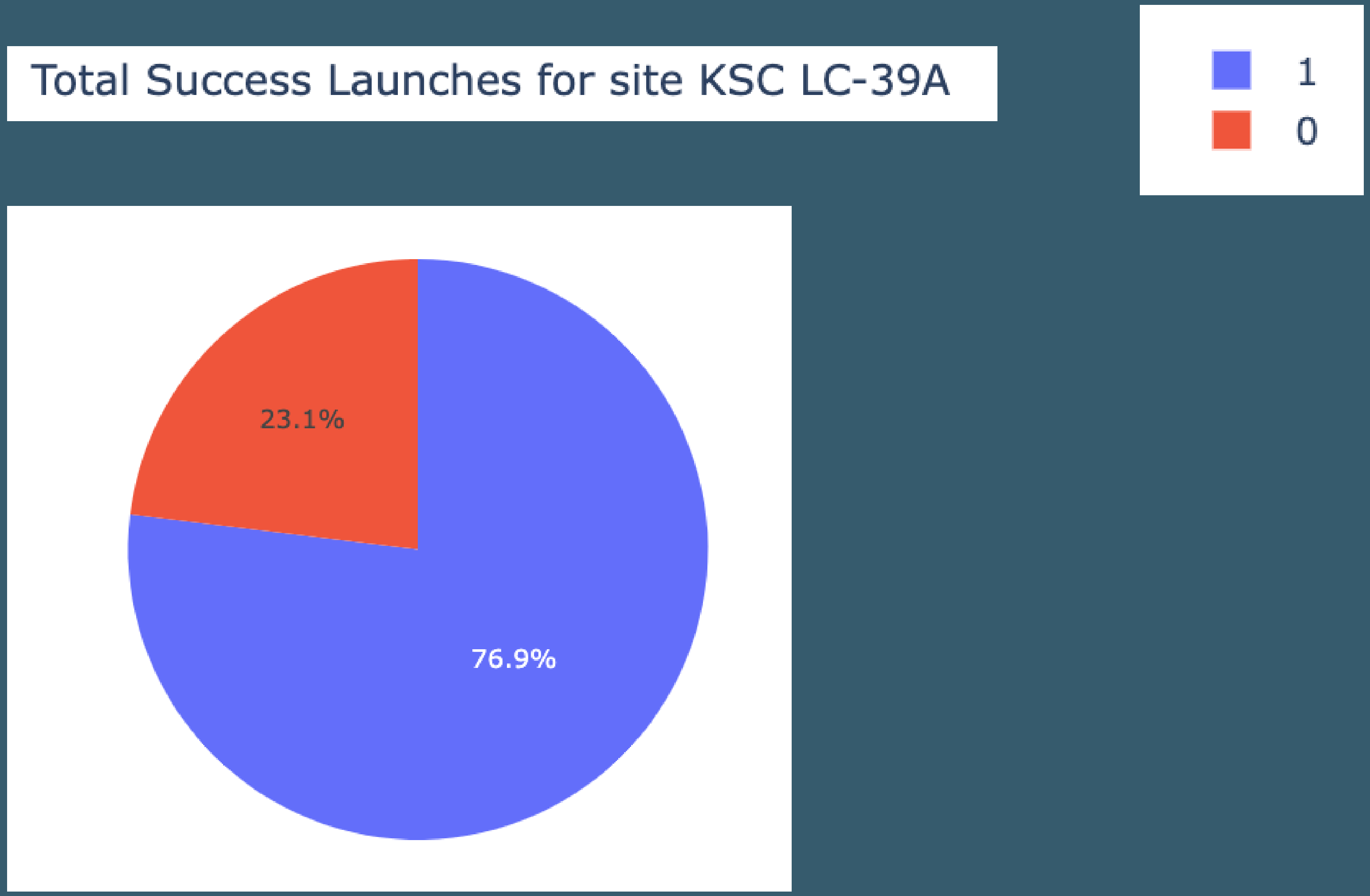
AVG(PAYLOAD_MASS__KG_)
2928.4

- Rank the count of successful landing_outcomes between the date 04-06-2010 and 20-03-2017 in descending order

Landing_Outcome	Total
Success	21
Success (drone ship)	8
Success (ground pad)	6

Total Success Launches by Site from Dashboard

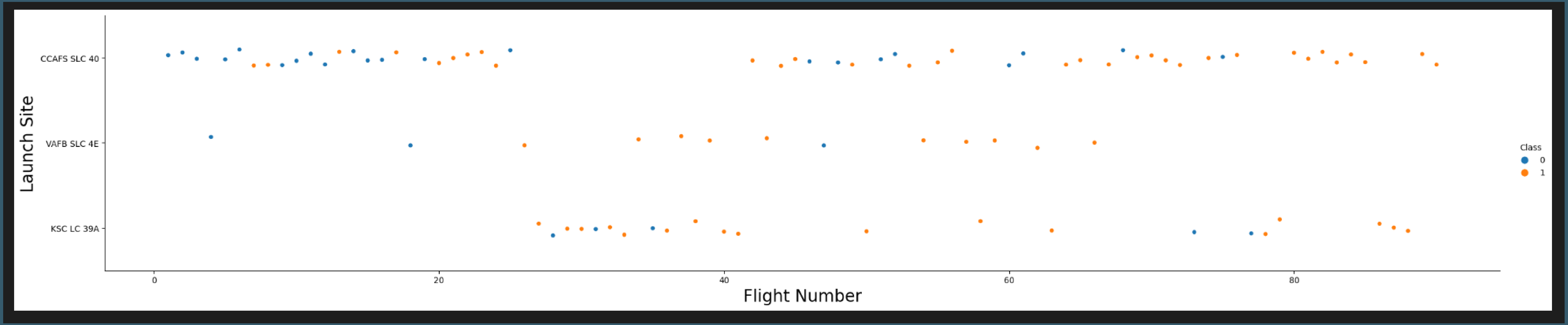




- Launches from KSC LC-39A have the highest success rate

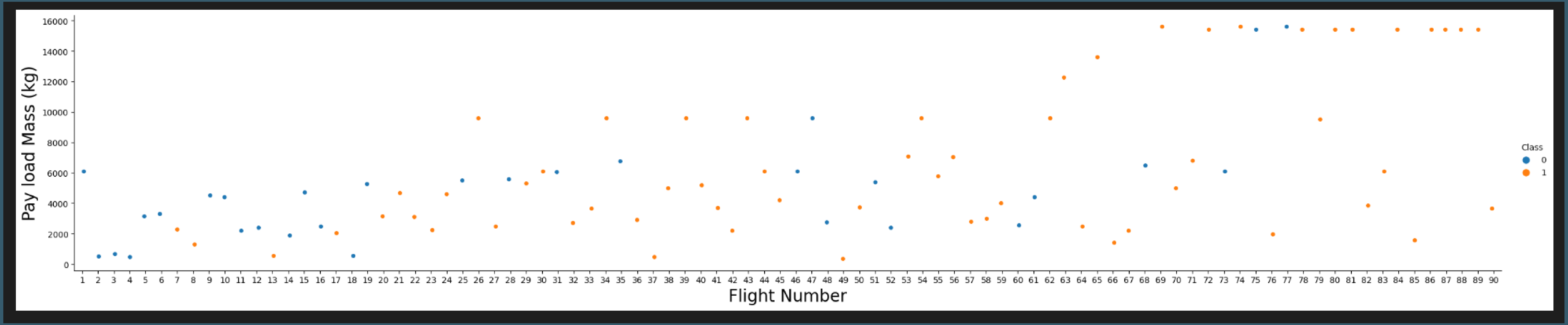
- Launches from KSC LC-39A have 76.9% success outcome

Flight Number vs Launch Site



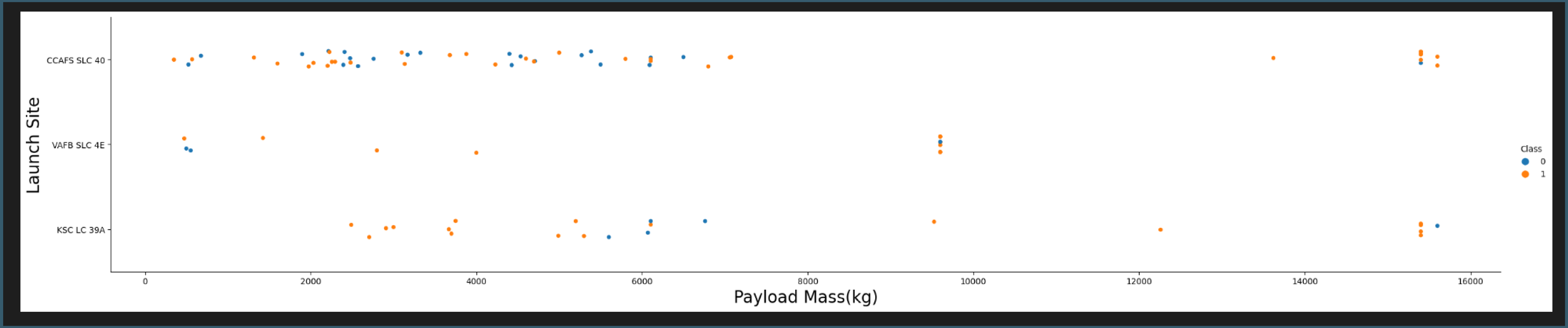
- Launches from CCAFS SLC 40 significantly higher than the other two sites
- Landing class 0 = bad outcome

Flight Number vs Payload Mass



- As the flight number increases, the first stage is more likely to land successfully
- The payload mass is also important and it seems the more massive the payload, the less likely the first stage will return

Payload vs Launch Site

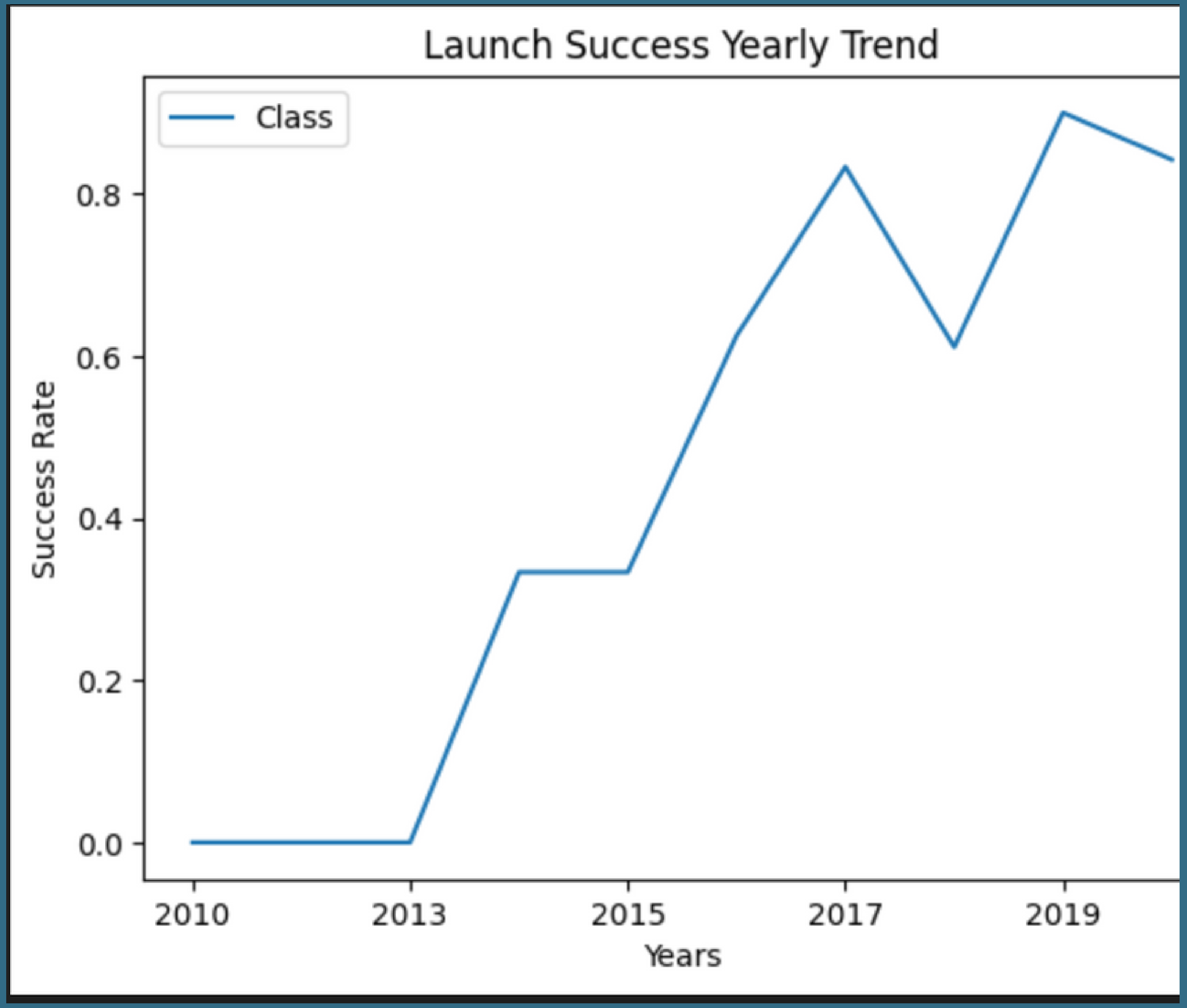


Payload vs Launch Site scatter point chart show that

- there are no rockets launched for heavy payload mass (greater than 10000) from VAFB SLC 4E
- heavy payloads launched from CCAFS SLC 40 & KSC LC 39A
- majority of lower payloads launched from CCAFS SLC 40
- the highest success between 2000kg and 6000kg

Launch Success Yearly Trend

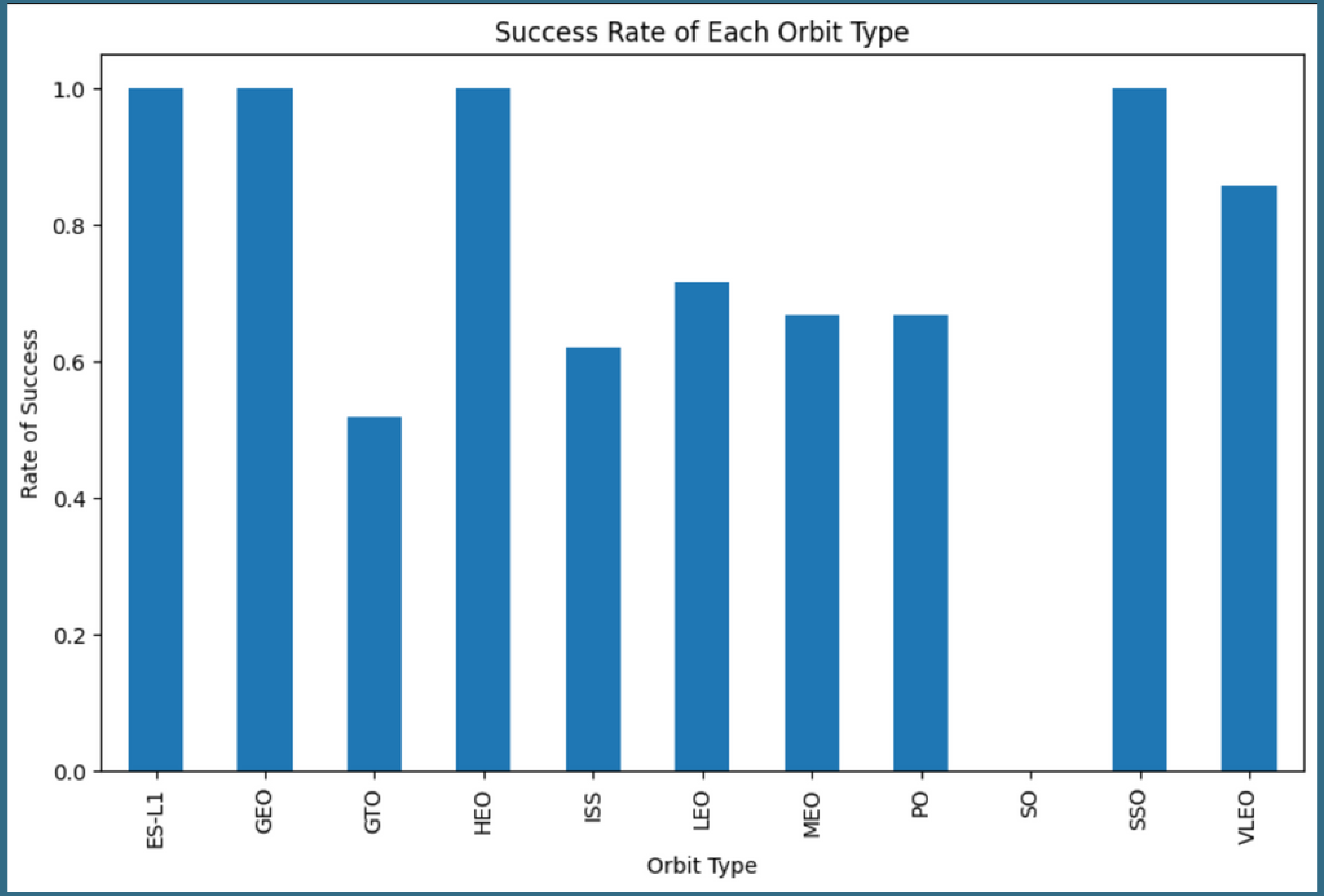
- The sucess rate since 2013 kept increasing till 2020
- There is a clear trend of success rate increase over time
- The success rate does not appear to be linear



Suceess Rate and Orbit Type

Create a `bar chart` for the success rate of each orbit to visually check if there are any relationship between success rate and orbit type.

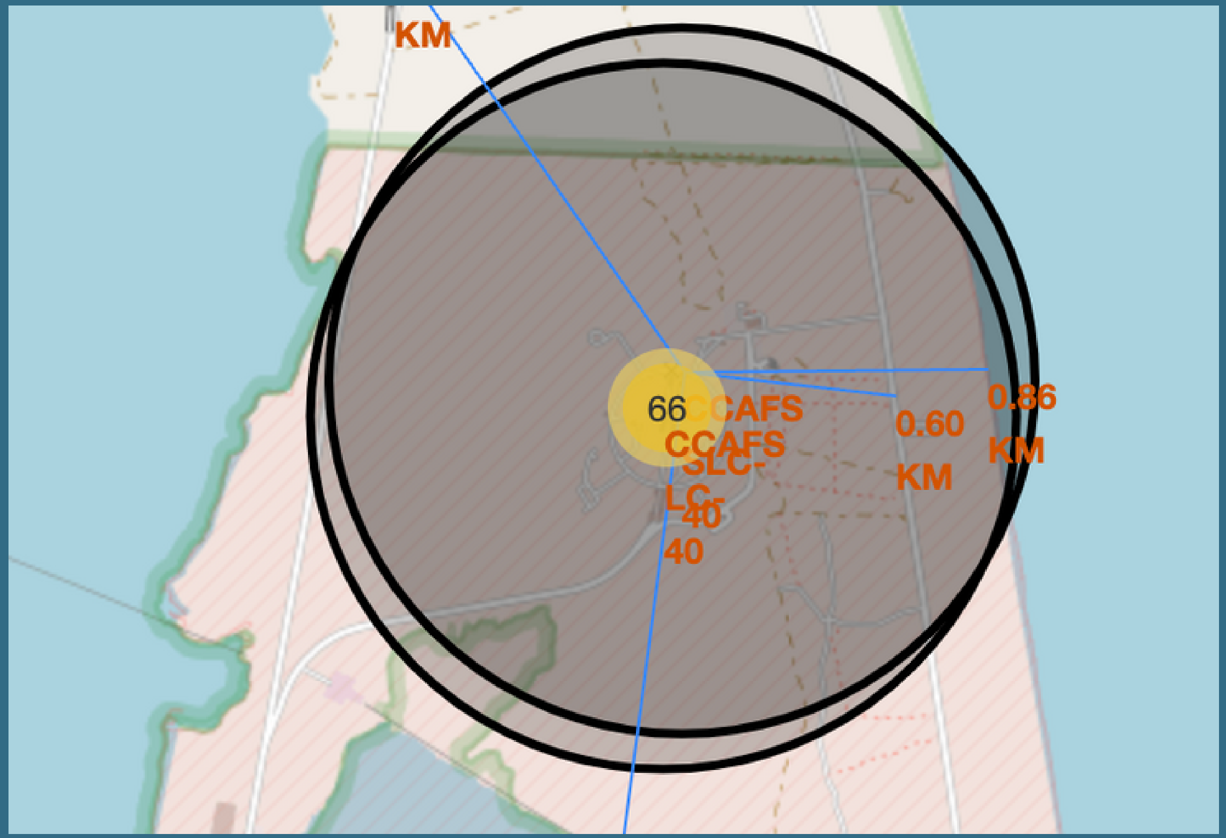
ES-L1, GEO, HEO, and SSO have the highest success rate

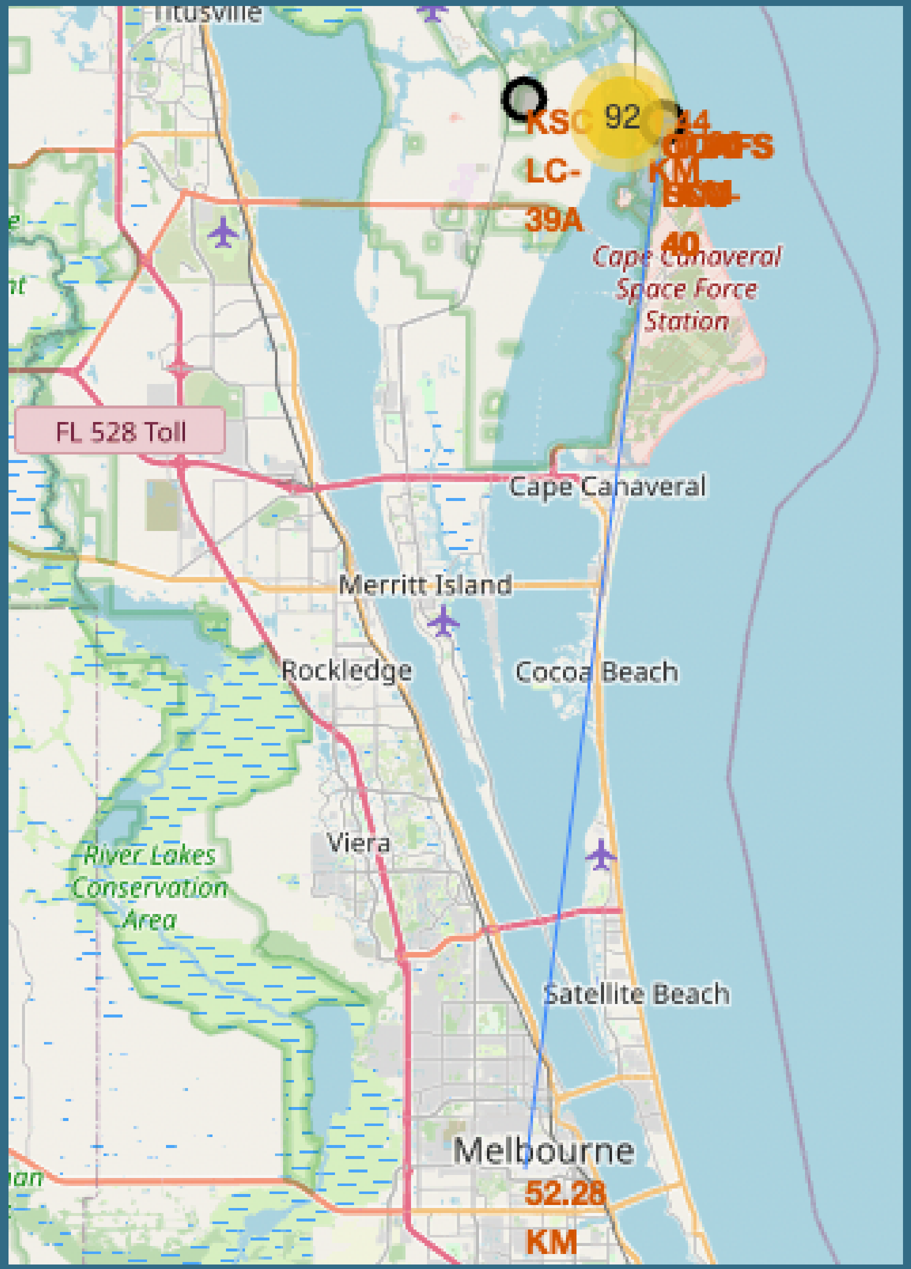


Launch site findings

- Launch sites are in close proximity to the equator line and the coast
- Safety - if something goes wrong during the ascent, the debris will fall into the ocean.
- Speed - launch sites near the equator get an additional natural boost that saves on fuel and boosters. It is because the surface of the Earth is traveling faster there so launching from the equator makes the rocket move faster once it is launched.
- Launch sites are also in close proximity to railways and highways for easy access
- Launch sites are not in close proximity to cities for safety reasons

Launch sites are in close proximity to the coast, railways and highways as shown





Launch sites are not in close proximity to cities as shown with Melbourne

Best Machine Learning Method

Winner is Decision Tree Classifier

Logistic Regression:

- tuned hpyerparameters :(best parameters){'C': 0.01, 'penalty': 'l2', 'solver': 'lbfgs'
- accuracy : 0.8464285714285713
- Test data accuracy: 0.8333333333333334

Decision Tree Classifier:

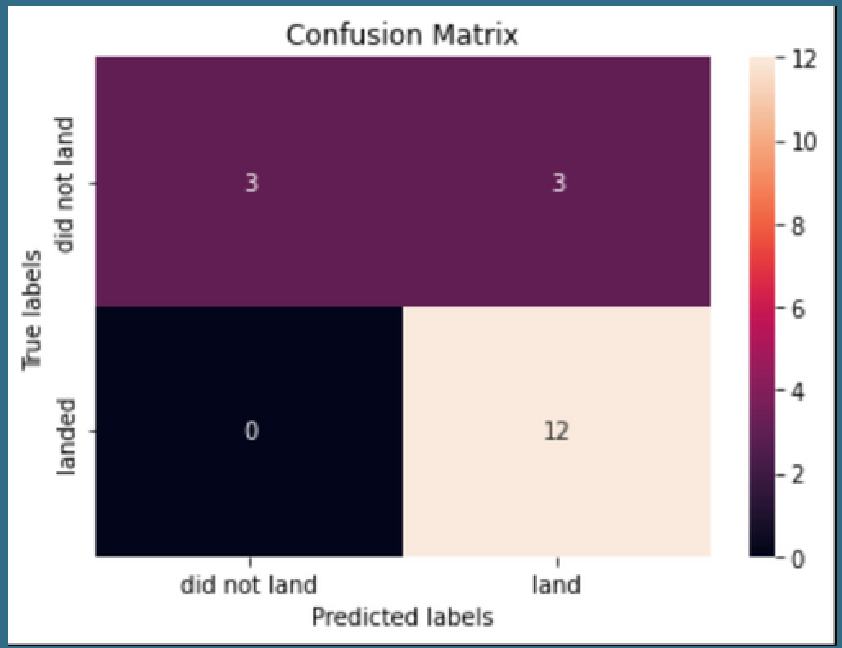
- tuned hpyerparameters :(best parameters){'criterion': 'entropy', 'max_depth': 12, 'max_features': 'auto', 'min_samples_leaf': 1, 'min_samples_split': 2, 'splitter': 'best'}
- accuracy : 0.9035714285714287
- Test data accuracy: 0.8333333333333334

SVM:

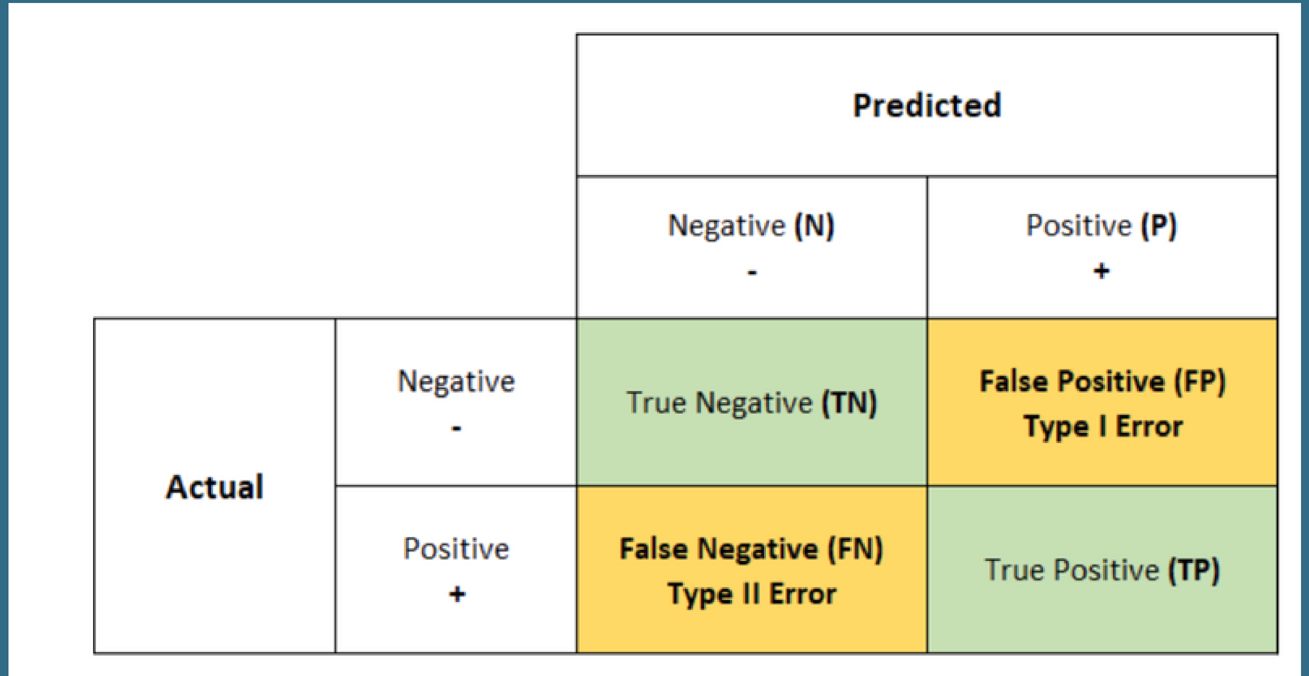
- tuned hpyerparameters :(best parameters){'C': 1.0, 'gamma': 0.03162277660168379, 'kernel': 'sigmoid'}
- accuracy : 0.8482142857142856
- Test data accuracy: 0.8333333333333334

K-nearest neighbour:

- tuned hpyerparameters :(best parameters){'algorithm': 'auto', 'n_neighbors': 10, 'p': 1}
- accuracy : 0.8482142857142858
- Test data accuracy: 0.8333333333333334



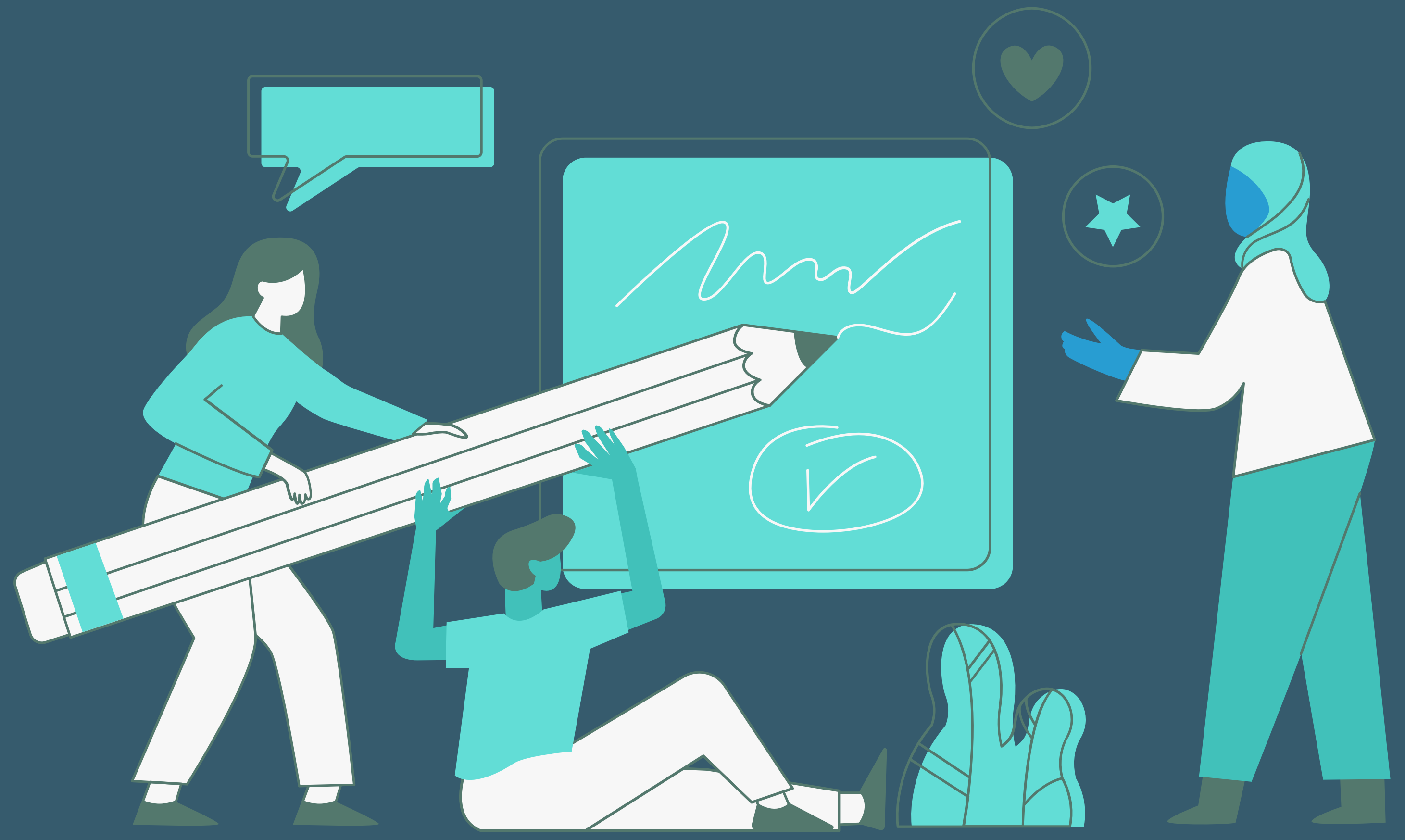
Same confusion matrix for all models



Binary Classification Problem (2×2 matrix)

Major problem is False Positives (Type 1 error)

Conclusion



- Decision Tree Classifier is the best model for the dataset provided
- Launch site KSV LC-39A has the highest success rate from all sites but it seems that the payload above 5500kg has negative impact
- As the flight number increases, the first stage is more likely to land successfully
- Orbits ES-L1, GEO, HEO, and SSO have the highest success rate

0.91
Decision Tree accuracy

76.9%
KSV LC-39A success rate

Thank You

Appendix

- GitHub
    - https://github.com/Violin1208/data_science_final_project
- Presentation
    - https://tinyurl.com/3ffavweb