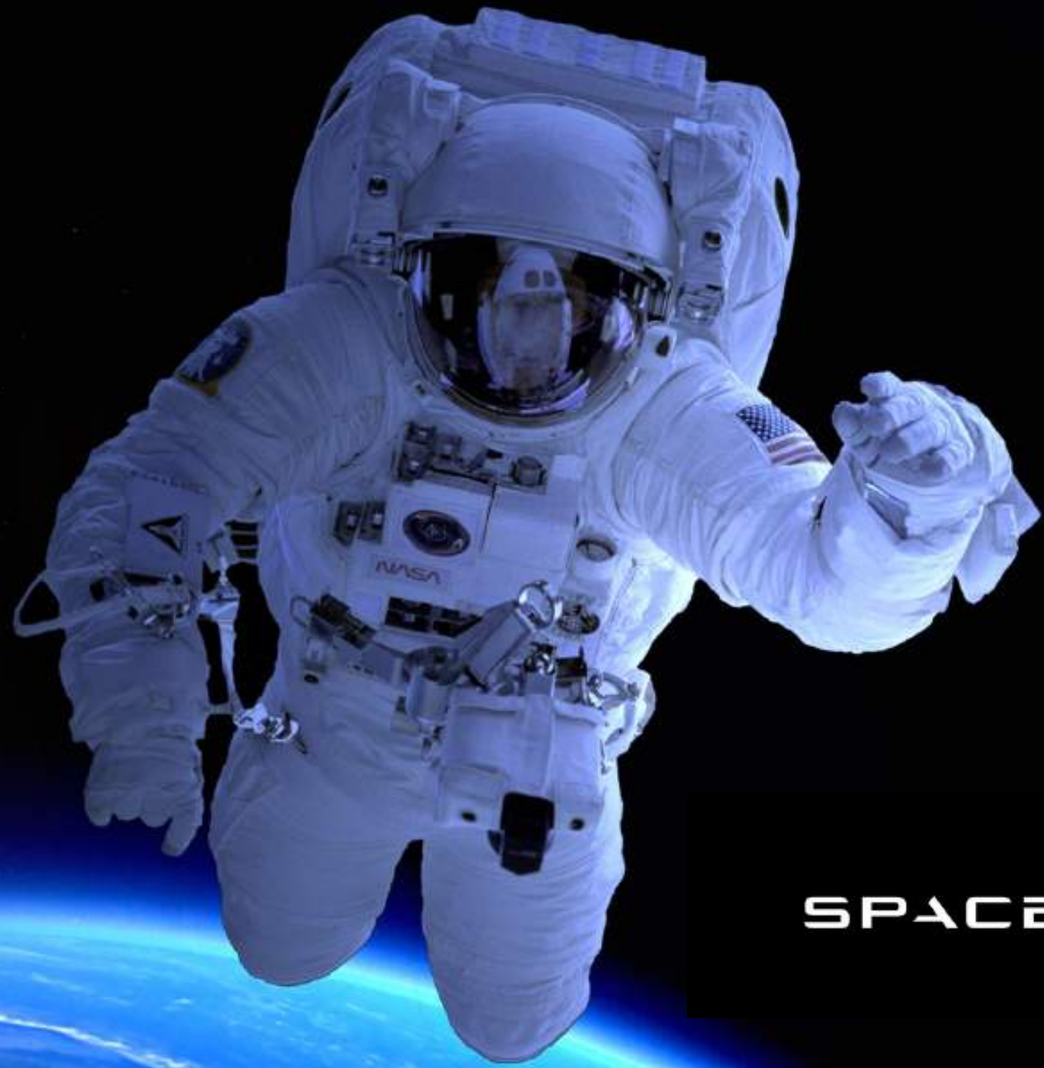
IBM Data Science Capstone

FALCON 9 FIRST STAGE LANDING PREDICTION



SPACEX

Chiao-Chieh, Deng

SPACEX

OUTLINE

01 EXECUTIVE SUMMARY

This project predicts whether the Falcon 9 first stage will land successfully, using SpaceX launch data.

By building a full ML pipeline — from API scraping to model evaluation — it demonstrates how reusable rockets affect cost efficiency in the space launch industry.

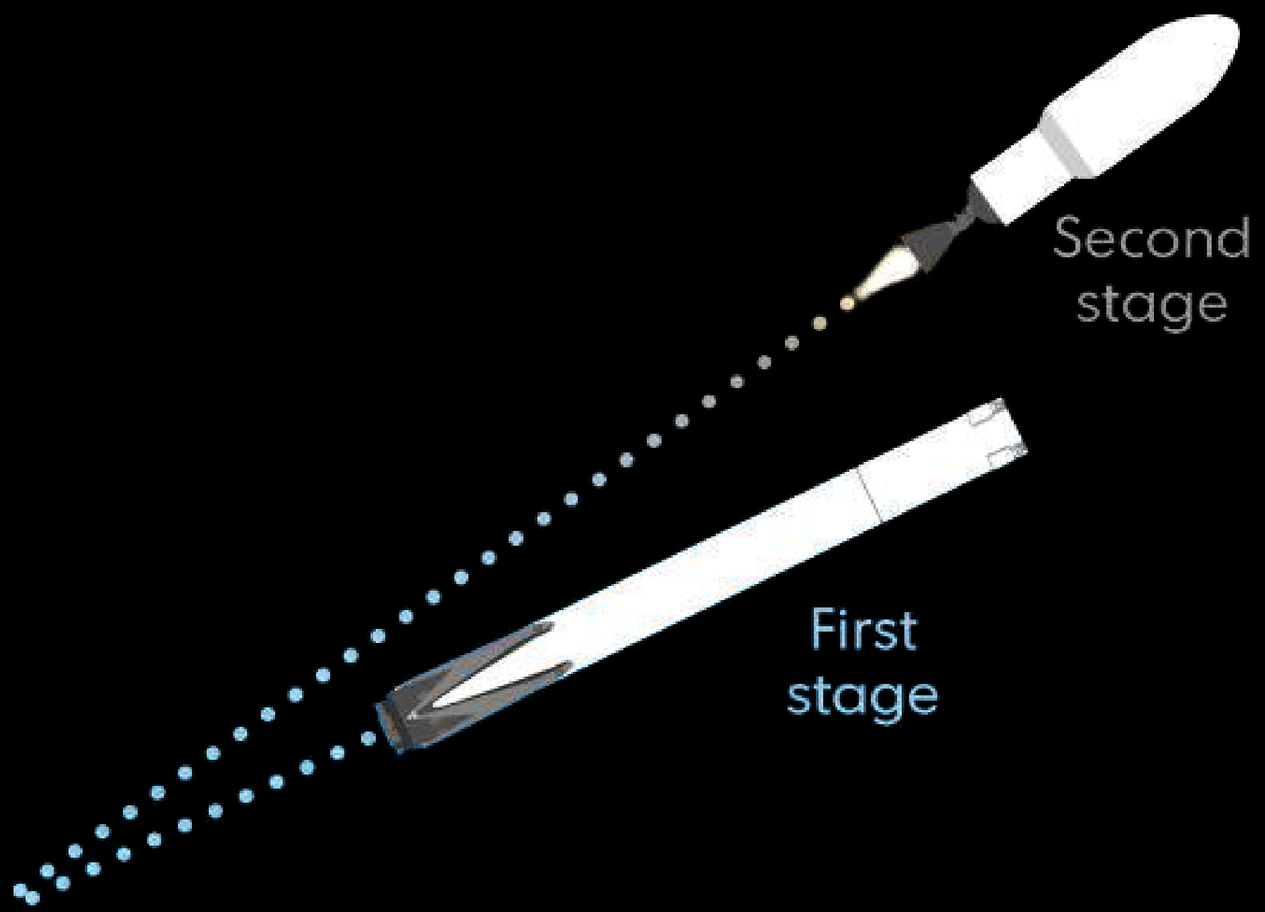
The final model achieved an accuracy of 83%.

02

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.

Thus, the goal of this project is to predict if the Falcon 9 first stage will land successfully.



The structure of the two-stage reusable rocket

03
METHODOLOGY

DATA COLLECTION

Data was collected using:

- **SpaceX API** (https://api.spacexdata.com/v4/rockets/)
- **Webscraping** (https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches)

DATA COLLECTION- SPACEX API

We used the “get request” to the SpaceX API to collect data, clean the requested data and did some basic data wrangling and formatting.

Request and parse the SpaceX launch data → Filter data to only include Falcon 9 launches → Deal with missing value

DATA COLLECTION- SCRAPING

Data could also be collected with web scraping from Wikipedia.

Request the Falcon 9 wiki page

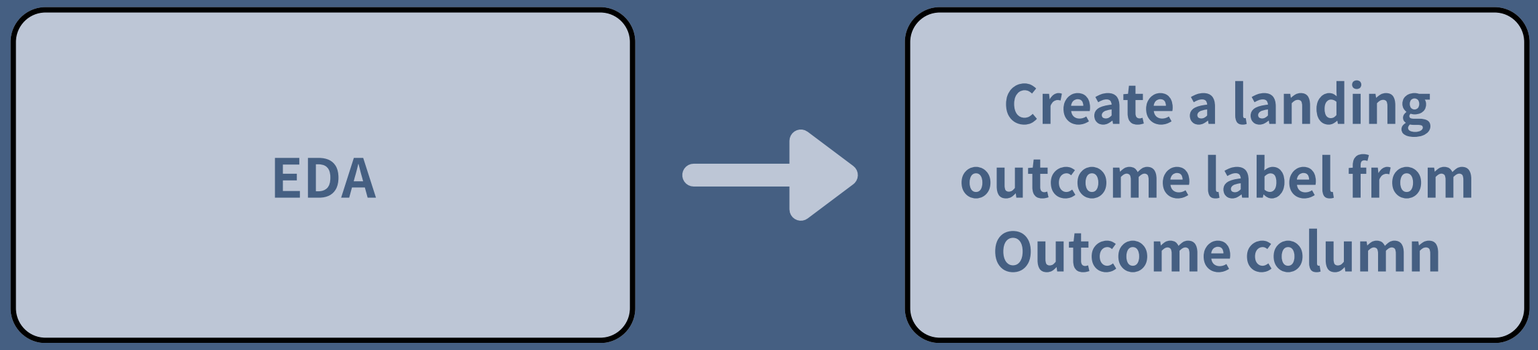
Extract all column/variable names from the HTML table header

Create a data frame by parsing the launch HTML tables

DATA WRANGLING

- Calculate the number of launches on each site, the number and occurrence of each orbit, and the number and occurence of mission outcome of the orbits

- Create a landing outcome. 0 for bad outcome, 1 otherwise.



EDA WITH DATA VISUALIZATION

Scatter plots and bar plots were used to visualize the relationship between two pairs of features.

EDA WITH SQL

The following SQL queries were performed:

🛰️ Launch Site Analysis

- Extracted all unique launch sites
- Filtered sites beginning with “CCA” for regional insights

📦 Payload & Customer Statistics

- Calculated total payloads by NASA (CRS missions)
- Found average payloads for booster version F9 v1.1
- Identified boosters with maximum payload capacity

EDA WITH SQL

🚀 Booster Landing Outcomes

- Found first successful landing on ground pad
- Summarized landing success/failure counts
- Filtered drone ship landings with payloads between 4000–6000kg
- Analyzed monthly landing failures in 2015
- Counted landing outcomes between 2010–2017 (sorted by frequency)

BUILD AN INTERACTIVE MAP WITH FOLIUM

- Circle highlighted circle area with a text label on a specific coordinate
- Marker clusters simplify a map containing many markers having the same coordinate.
- Lines calculate the distances between two coordinates

PREDICTIVE ANALYSIS (CLASSIFICATION)

Compare the result of four model with test accuracy: logistic regression, svm, decision tree, KNN

Data preparation and standardization → **Test 4 models with hyperparameter** → **Compare the result**

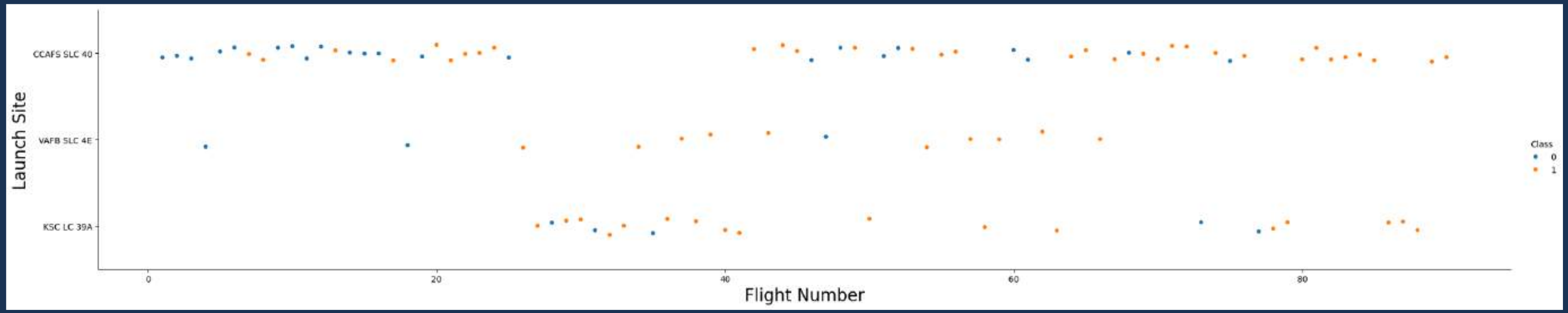
04
RESULTS

- Insights drawn from EDA
- Launch sites proximities analysis
- Predective analysis

INSIGHTS DRAWN FROM EDA

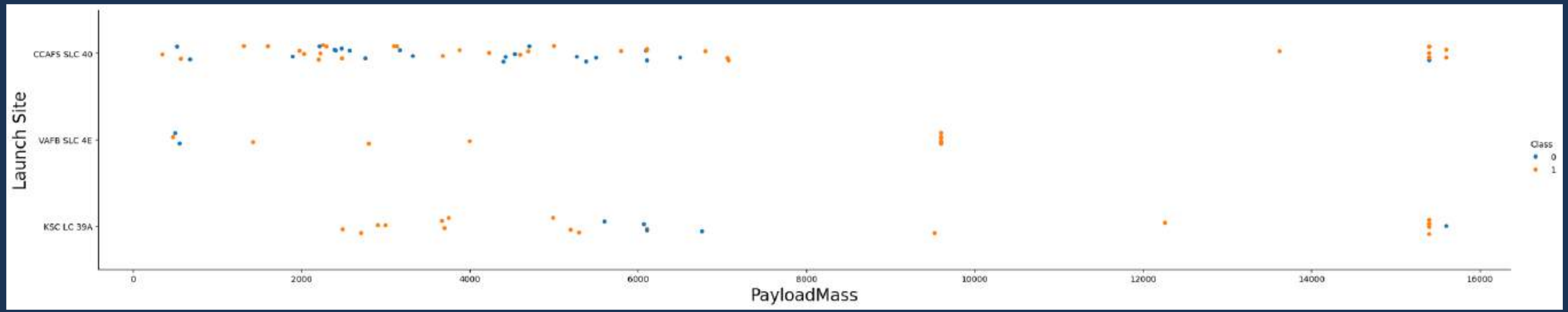
FLIGHT NUMBER VS. LAUNCH SITE

- CCAFS SLC 40 was the launch site that had the most flight number
- Successful outcomes increases over time

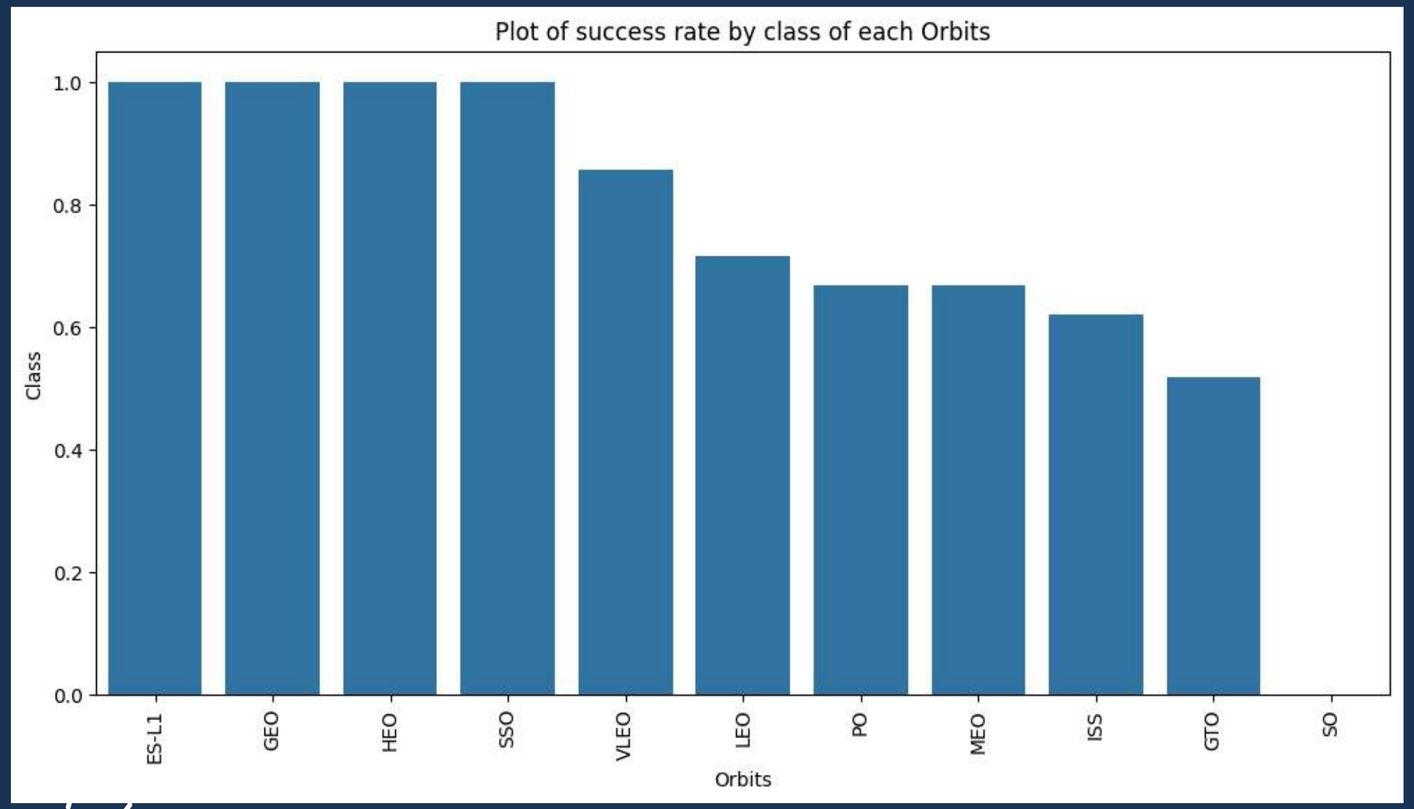


PAYLOAD VS. LAUNCH SITE

- Higher payload mass increases the chances of successful landings.
- No rockets launched at VAFB-SLC exceed a payload mass of 10,000.



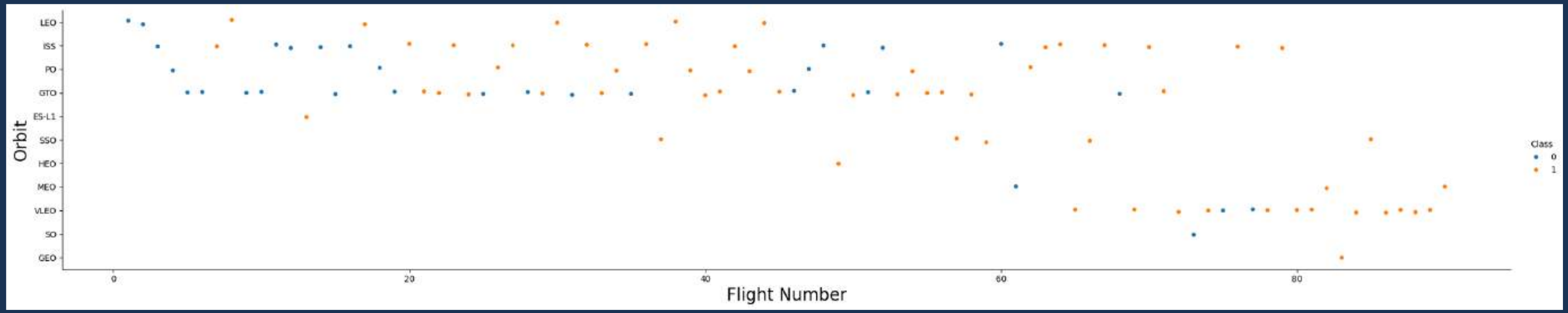
SUCCESS RATE VS. ORBIT TYPE



ES-L1, GEO, HEO and SSO had a 100% success rate, whereas SO failed entirely.

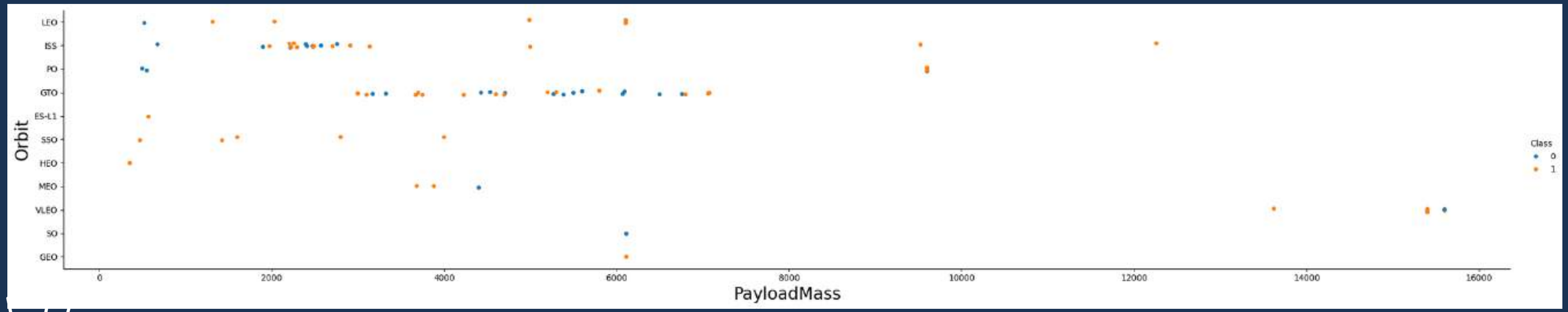
FLIGHT NUMBER VS. ORBIT TYPE

- VLEO is potential due to the highly successful outcome in recent years
- No relationship between flight number when in GTO orbit.

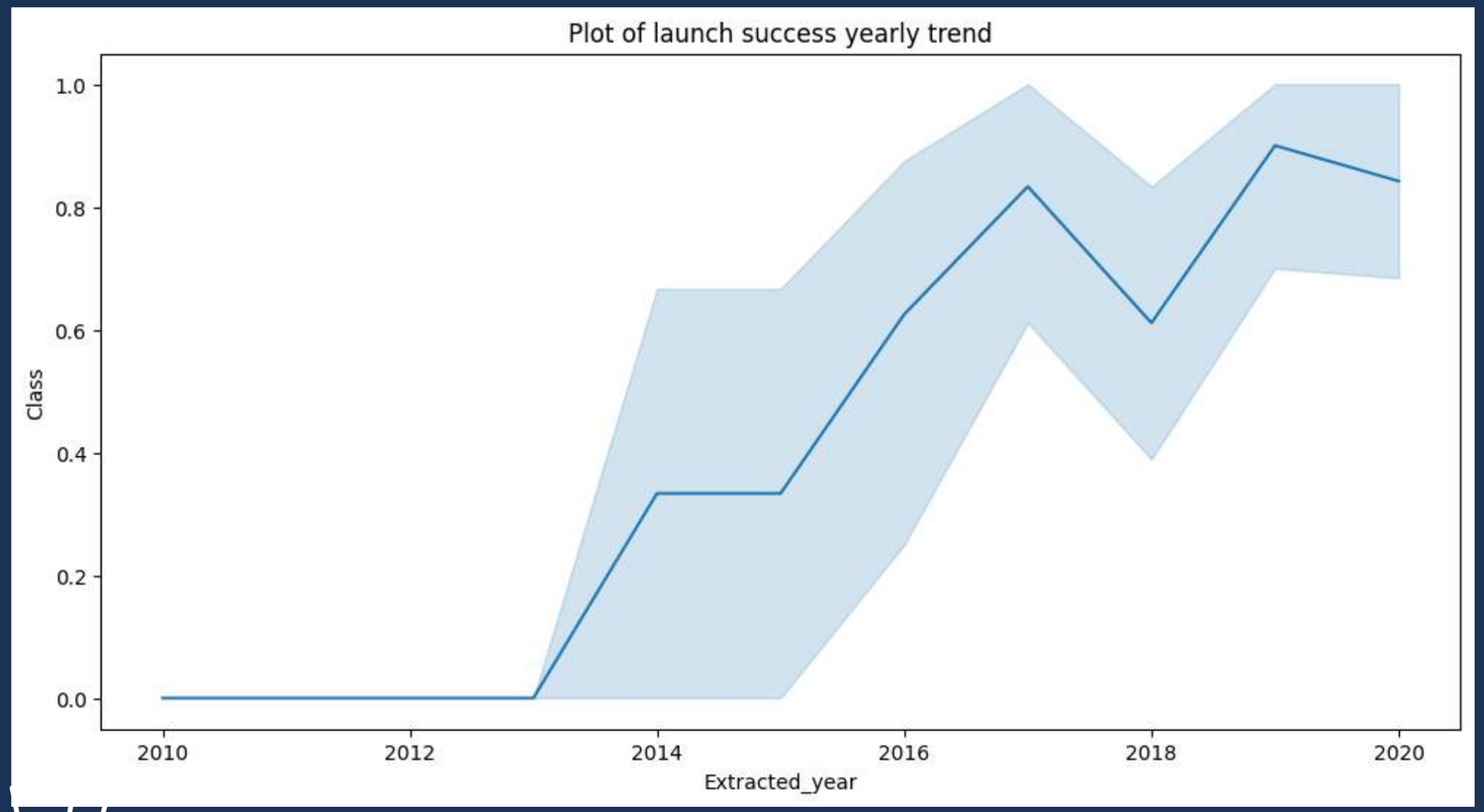


PAYLOAD VS. ORBIT TYPE

- With the heavy payload, ISS has achieved a higher success rate in landings compared to VLEO, which has experienced both successful and unsuccessful landings.



LAUNCH SUCCESS YEARLY TREND



Between 2010 and 2013, the launches failed entirely. Over time, the success rate gradually increased, reaching 0.8 in 2020.

ALL LAUNCH SITE NAMES

SpaceX used 4 different launch sites

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

LAUNCH SITE NAMES BEGIN WITH 'CCA'

The rocket was launched from CCAFSLC-40 mostly by SpaceX and NASA.

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	**CCA**FS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-12-08	15:43:00	F9 v1.0 B0004	**CCA**FS 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	**CCA**FS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	F9 v1.0 B0006	**CCA**FS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	**CCA**FS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attemp

TOTAL PAYLOAD MASS

The total payload of F9 v1.1 booster is 111,268 kg

Total Payload (kg)
111.268

AVERAGE PAYLOAD MASS BY F9 V1.1

The average payload of F9 v1.1 booster is 2,928 kg

Avg Payload (kg)
2.928

FIRST SUCCESSFUL GROUND LANDING DATE

The first success landing outcome happened in 2015 fiver year after the first launch

Min Date
2015-12-22

6

SUCCESSFUL DRONE SHIP LANDING WITH PAYLOAD BETWEEN 4000 AND 6000

Many Falcon 9 FT booster versions were successful at landing in drone ships having payload above the average.

Booster Version
F9 FT B1021.2
F9 FT B1031.2
F9 FT B1022
F9 FT B1026

TOTAL NUMBER OF SUCCESSFUL AND FAILURE MISSION OUTCOMES

Almost 100% of mission outcomes were successful

Mission Outcome	Occurrences
Success	99
Success (payload status unclear)	1
Failure (in flight)	1

BOOSTERS CARRIED MAXIMUM PAYLOAD

Many F9 B5 booster version carried the maximum payload

Booster Version (...)	Booster Version
F9 B5 B1048.4	F9 B5 B1051.4
F9 B5 B1048.5	F9 B5 B1051.6
F9 B5 B1049.4	F9 B5 B1056.4
F9 B5 B1049.5	F9 B5 B1058.3
F9 B5 B1049.7	F9 B5 B1060.2
F9 B5 B1051.3	F9 B5 B1060.3

2015 LAUNCH RECORDS

- Two booster versions failed at landing in drone ships in 2015
  - F9 v1.1 B1012
  - F9 v1.1 B1015

Booster Version	Launch Site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

RANK LANDING OUTCOMES BETWEEN 2010-06-04 AND 2017-03-20

No attempt had occurred the most within 7 years

Landing Outcome	Occurrences
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

LAUNCH SITES PROXIMITIES ANALYSIS

ALL LAUNCH SITES

Launch sites are near sea

LAUNCH OUTCOMES BY SITE

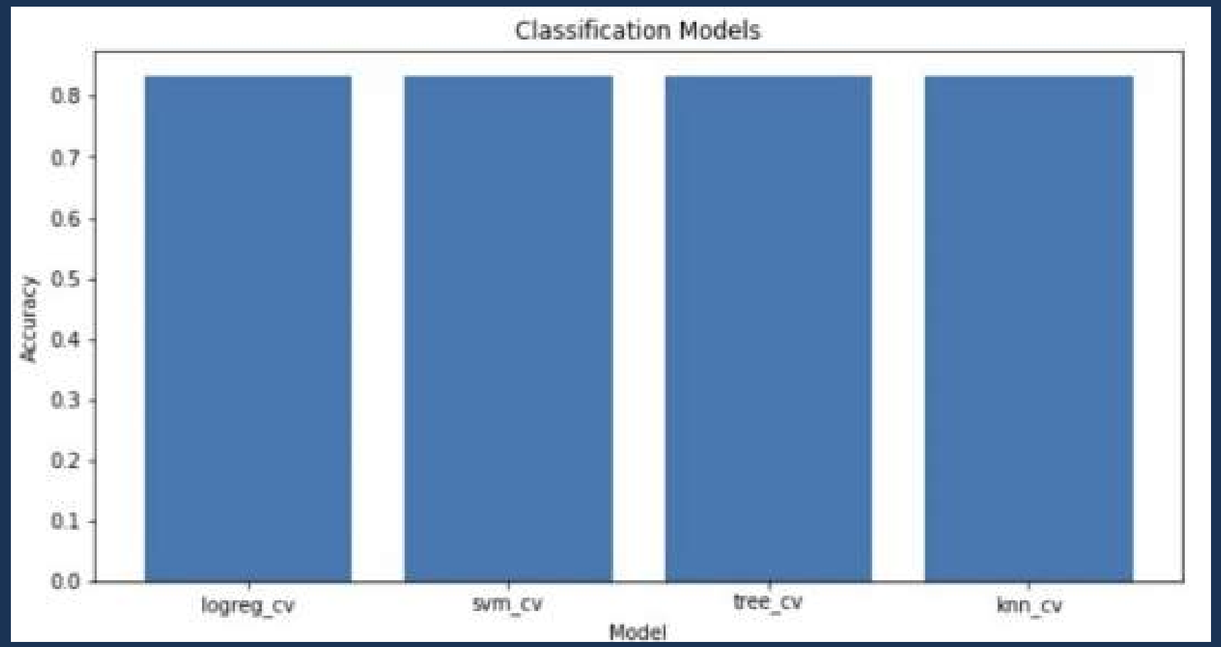
Green marks the successful outcomes, and red otherwise.

THE PROXIMITIES OF LAUNCH SITES

The CCAFSSLC-40 is 0.9km far away from the coastline, also being close to railroad and highway

PREDECTIVE ANALYSIS

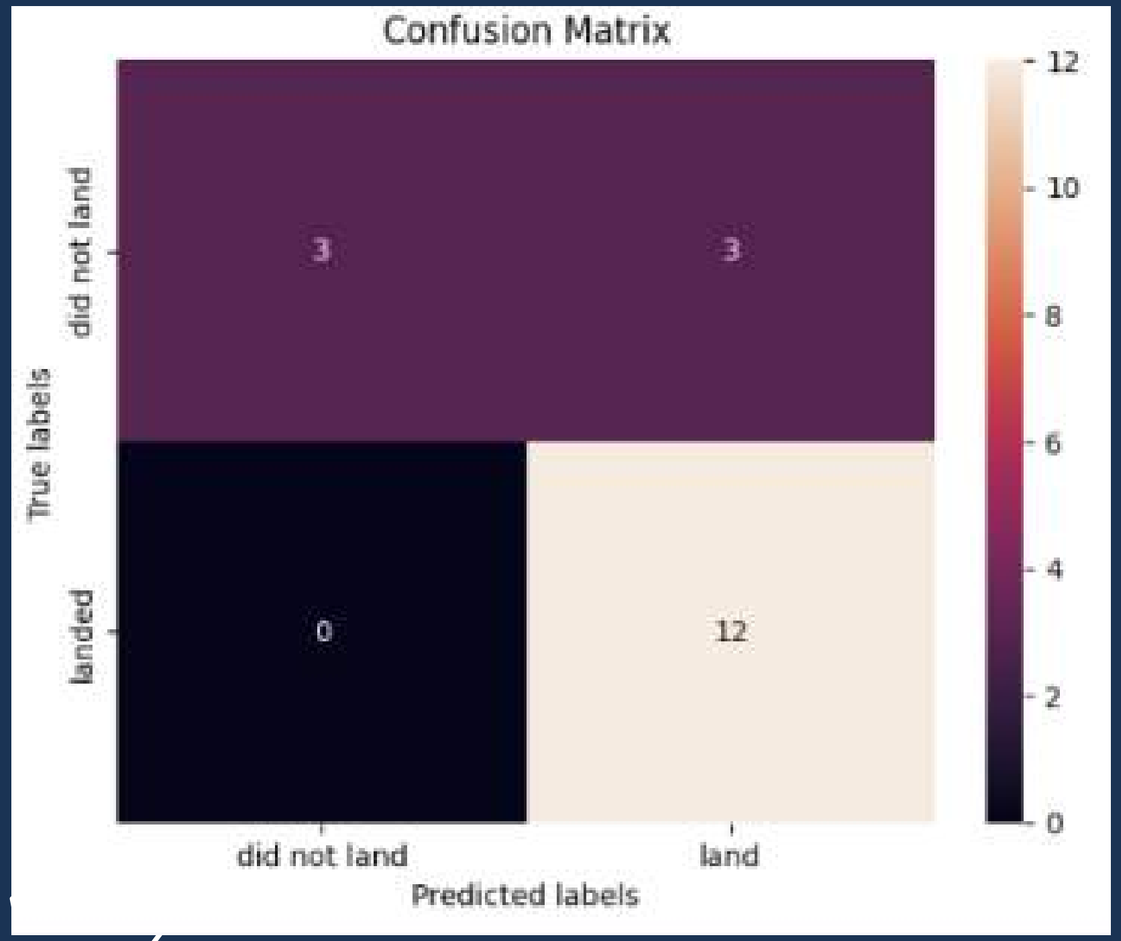
CLASSIFICATION ACCURACY



Model	Accuracy	TestAccuracy
LogReg	0.84643	0.83333
SVM	0.84821	0.83333
Tree	0.93036	0.83333
KNN	0.84821	0.83333

Four models(SVM, Classification Trees, KNN and Logistic Regression have same accuracy, which is **0.833**

CONFUSION MATRIX



Four models has the same accuracy, which means their confusion matrix was exactly the same.

05 CONCLUSIONS

- CCAFS SLC 40 was the most frequently used launch site, active from early to recent missions.
- VLEO has become the dominant orbit in recent years, with an exceptionally high success rate.
- Payloads over 15,000 kg (often in VLEO missions) surprisingly had high success rates, likely due to improved technology and mission planning.
- ISS missions were also highly reliable, though typically carried lighter payloads.

05 CONCLUSIONS

- The largest payloads were mostly carried by F9 Block 5 boosters in recent years.
- No booster recovery attempts were made between 2010–2017, reflecting the expendable phase of SpaceX.
- The first successful ground landing occurred in 2015, marking a shift toward reusability.
- Overall mission success improved over time, with most failures occurring in early missions.

06 APPENDIX

GitHub repository:

https://github.com/christine1117/Christine-Projects/tree/0a03a7c808d63025d1a2f27c8136348367731710a/Data-Science-Projects/SpaceX:

THANKS!