Unlocking Insights from SpaceX Launch Data

Presented by :
BenSaad Amine

OUTLINE

- Executive Summary
- Introduction
- Results
    - Visualization – Charts
    - Dashboard
- Discussion
    - Findings & Implications
- Conclusion

EXECUTIVE SUMMARY

- **Objective:** Predict the success of SpaceX's Falcon 9 first stage landings
- **Data Collection:** Aggregated data from SpaceX API and web scraping
- **Data Processing:** Performed data wrangling to handle missing values and engineered features relevant to landing success.
- **Exploratory Data Analysis (EDA):** Identified key factors influencing landing success, such as launch sites and payload masses, using visual analytics
- **Predictive Modeling:** Developed and evaluated multiple classification models

INTRODUCTION

- **Background:** SpaceX aims to reduce space travel costs by reusing rocket components, notably the Falcon 9 first stage. Predicting landing success is crucial for operational efficiency and cost reduction.
- **Problem Statement:** Accurately forecasting the landing outcome of Falcon 9's first stage to inform decision-making and improve reusability rates.
- **Project Scope:** Collect and preprocess relevant data, conduct exploratory data analysis, build predictive models, and identify the most accurate model for predicting landing success.
- **Significance:** Enhancing prediction accuracy supports SpaceX's goal of sustainable and cost-effective space missions through improved reusability strategies.

Data Collection and Data Wrangling

1. **Data Source**:
   - SpaceX API: https://api.spacexdata.com/v4/launches/past
   - Retrieved past launch data in JSON format
2. **Data Extraction Process**:
   - Used .json() to decode the response
   - Converted JSON data into a Pandas DataFrame using .json_normalize().
   - Found that many columns contained only IDs (e.g., rocket, payloads, launchpad, cores).
3. **Additional API Requests**:
   - Extracted detailed launch information using IDs.
   - Mapped key features:
     - **Rocket** → Booster version
     - **Payloads** → Payload mass, Orbit.
     - **Launchpad** → Launch site name, Longitude, Latitude.
     - **Cores** → Landing outcome, Landing type, Flights, GridFins, Reused status, Legs, Landing pad, Block, Reused count, Serial.

4. Data Wrangling:

- Created a structured dataset using extracted data.
- Combined selected columns into a new DataFrame.
- Filtered data to include only Falcon 9 launches.
- Handled missing values.

5.Web scraping:

- perfor web scraping to collect Falcon 9 historical launch records from a Wikipedia

[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)

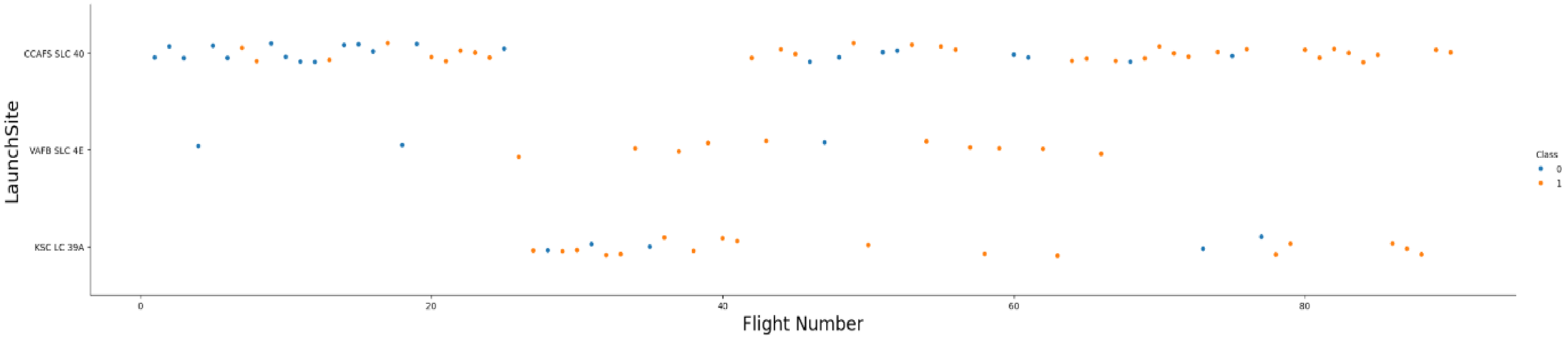
	Flight No.	Launch site	Payload	Payload mass	Orbit	Customer	Launch outcome	Version Booster	Booster landing	Date	Time
0	1	CCAFS	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success\n	F9 v1.07B0003.18	Failure	4 June 2010	18:45
1	2	CCAFS	Dragon	0	LEO	NASA	Success	F9 v1.07B0004.18	Failure	8 December 2010	15:43
2	3	CCAFS	Dragon	525 kg	LEO	NASA	Success	F9 v1.07B0005.18	No attempt\n	22 May 2012	07:44
3	4	CCAFS	SpaceX CRS-1	4,700 kg	LEO	NASA	Success\n	F9 v1.07B0006.18	No attempt	8 October 2012	00:35
4	5	CCAFS	SpaceX CRS-2	4,877 kg	LEO	NASA	Success\n	F9 v1.07B0007.18	No attempt\n	1 March 2013	15:10
...	...	...	...	...	...	...	...	...	...	...	...
222	117	CCSFS	Starlink	15,600 kg	LEO	SpaceX	Success\n	F9 B5B1051.10657	Success	9 May 2021	06:42
223	118	KSC	Starlink	~14,000 kg	LEO	SpaceX	Success\n	F9 B5B1058.8660	Success	15 May 2021	22:56
224	119	CCSFS	Starlink	15,600 kg	LEO	NASA	Success\n	F9 B5B1063.2665	Success	26 May 2021	18:59
225	120	KSC	SpaceX CRS-22	3,328 kg	LEO	Sirius XM	Success\n	F9 B5B1067.1668	Success	3 June 2021	17:29
226	121	CCSFS	SXM-8	7,000 kg	GTO	NaN	NaN	F9 B5	NaN	6 June 2021	04:26

227 rows × 11 columns

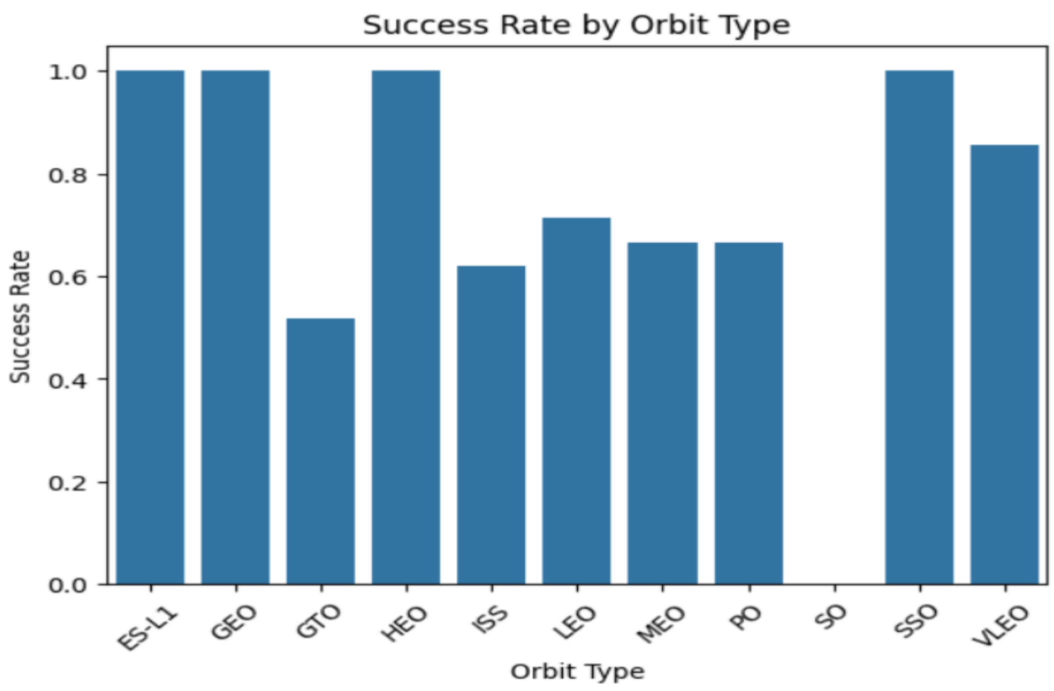
	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitu
0	1	2010-06-04	Falcon 9	6104.959412	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0003	-80.577
1	2	2012-05-22	Falcon 9	525.000000	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0005	-80.577
2	3	2013-03-01	Falcon 9	677.000000	ISS	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0007	-80.577
3	4	2013-09-29	Falcon 9	500.000000	PO	VAFB SLC 4E	False Ocean	1	False	False	False	NaN	1.0	0	B1003	-120.610
4	5	2013-12-03	Falcon 9	3170.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B1004	-80.577
5	6	2014-01-06	Falcon 9	3325.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B1005	-80.577
6	7	2014-04-18	Falcon 9	2296.000000	ISS	CCAFS SLC 40	True Ocean	1	False	False	True	NaN	1.0	0	B1006	-80.577
7	8	2014-07-14	Falcon 9	1316.000000	LEO	CCAFS SLC 40	True Ocean	1	False	False	True	NaN	1.0	0	B1007	-80.577
8	9	2014-08-05	Falcon 9	4535.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B1008	-80.577
9	10	2014-09-07	Falcon 9	4428.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B1011	-80.577

Exploring and Preparing Data

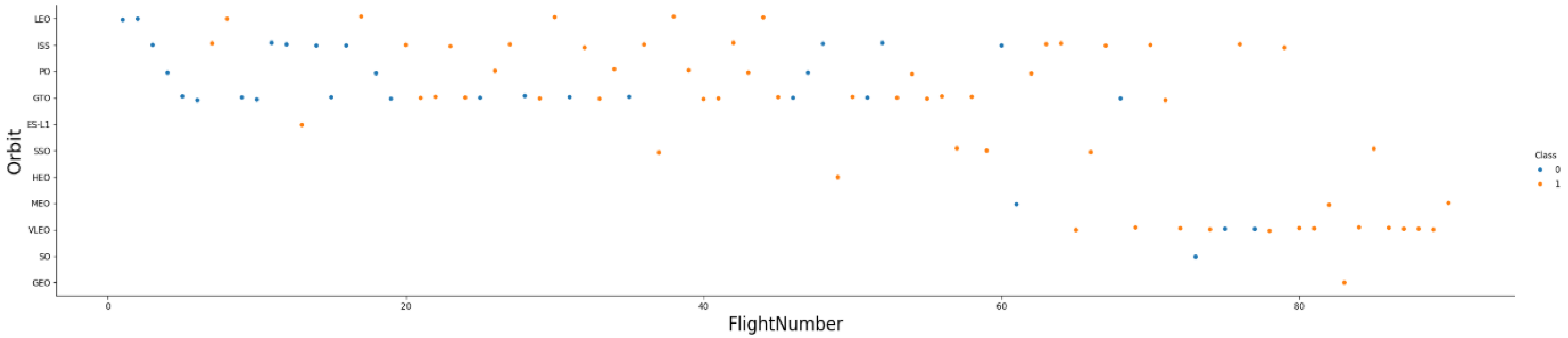
In this section we tried seeing how diffrent features would affect the launch outcome:

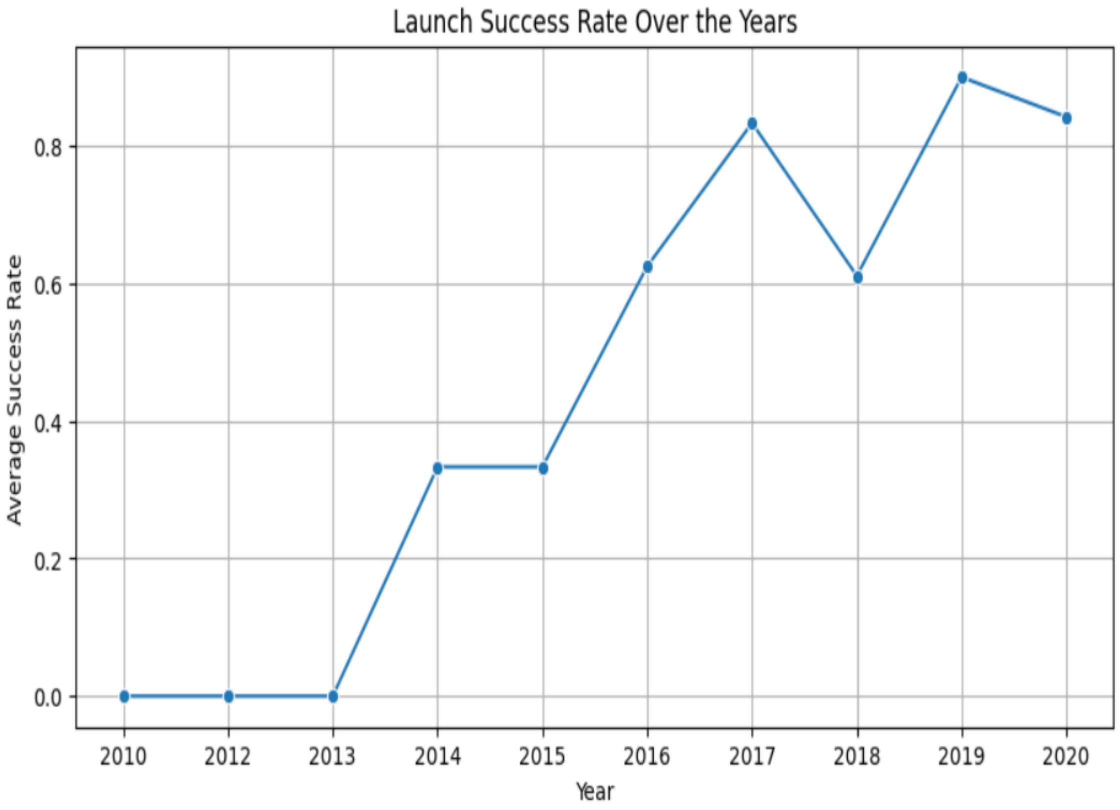


From this graph we could interpret that ccafs followed by ksc lc have the most number of launches as they have a heighr concentration and they have nearly continuous spread indicating that launches have been conducted **regularly** over time.



From those two graphes we can interpret that ES-L1,GEO,HEO,SSO are the orbits with heighest sucess rate but we can observe that in the LEO orbit, success seems to be related to the number of flights. Conversely, in the GTO orbit, there appears to be no relationship between flight number and success.





Also we can observe that the sucess rate since 2013 kept increasing till 2020 explained by :

- **Technological Improvements**: Continuous advancements in rocket design, landing technology, and reusable boosters.
- **Increased Experience**: More launches allowed SpaceX to refine and optimize landing techniques.
- **More Stringent Testing**: SpaceX likely improved pre-launch testing and quality control measures to minimize failures.
- **Enhanced Data Utilization**: Improved machine learning models and simulations helped predict and mitigate failures.

SQL results slides

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 count of landing outcomes (Failure or Success)between the date 2010 and 2017

Landing_Outcome	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

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

sum(PAYLOAD_MASS__KG_)
45596

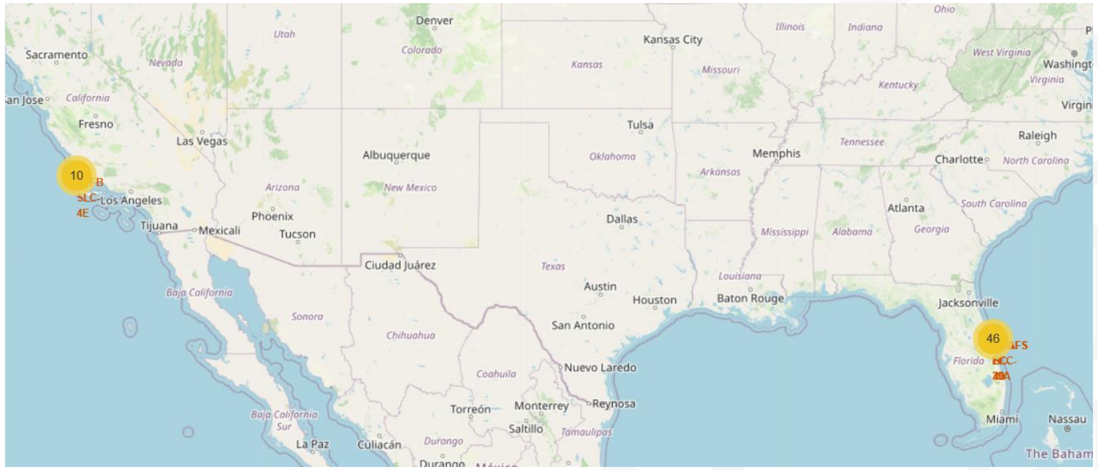
average payload mass carried by booster version F9 v1.1

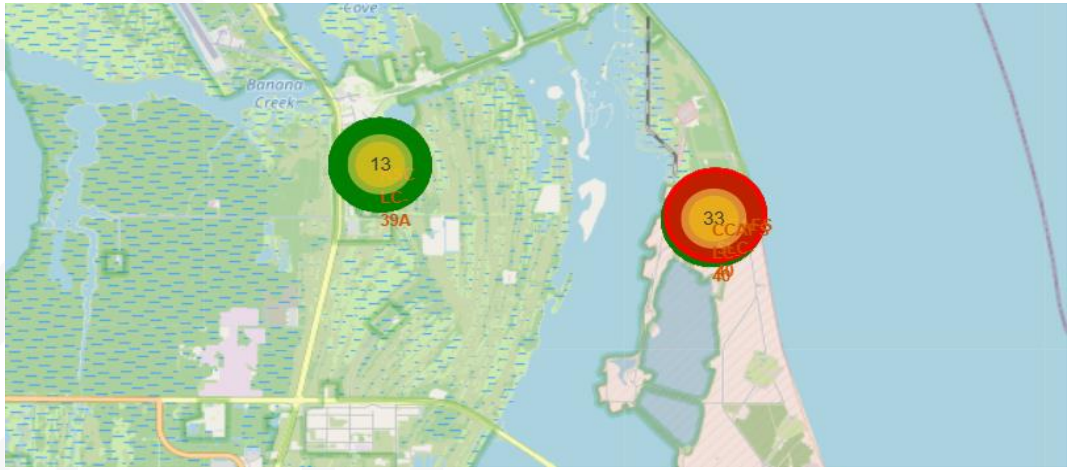
avg(PAYLOAD_MASS__KG_)
2928.4

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

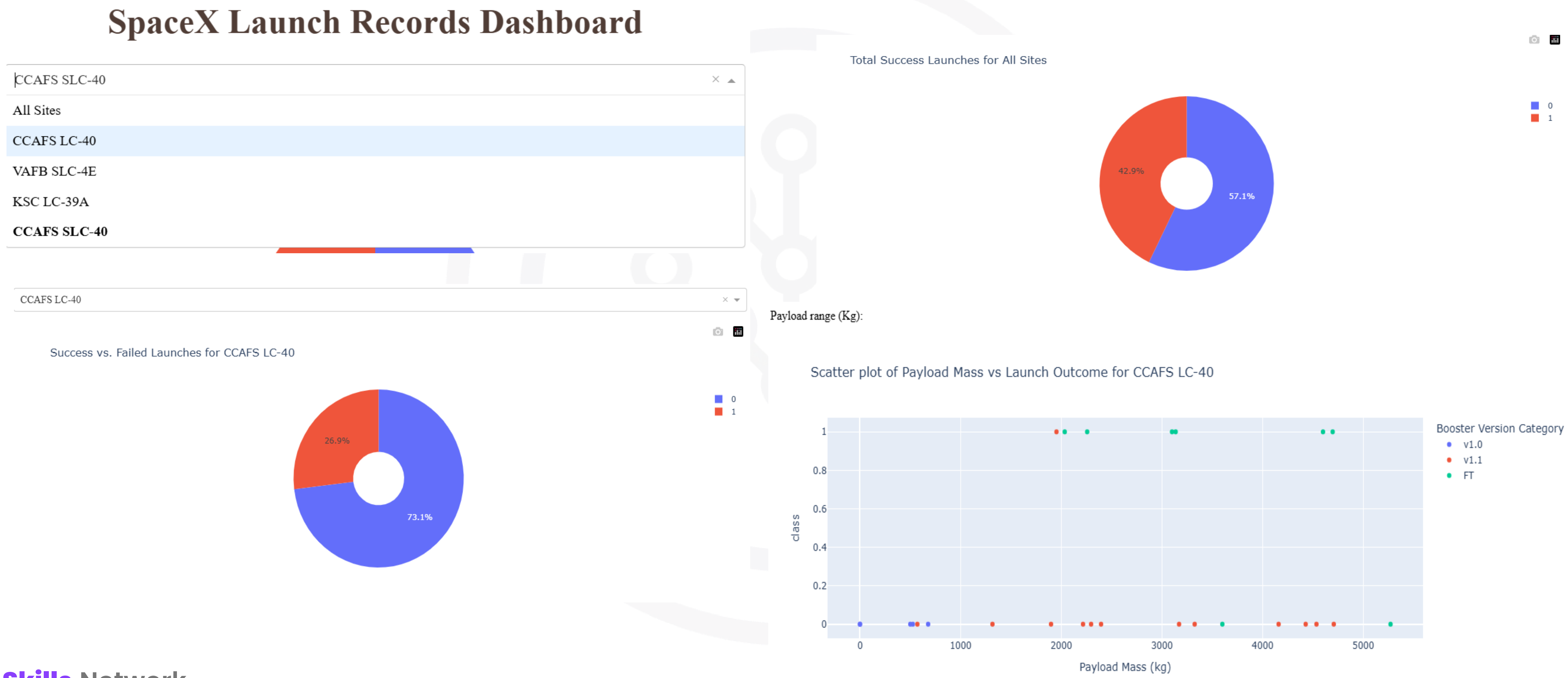
min(Date)
2010-06-04

Interactive Visual Analytics with Folium

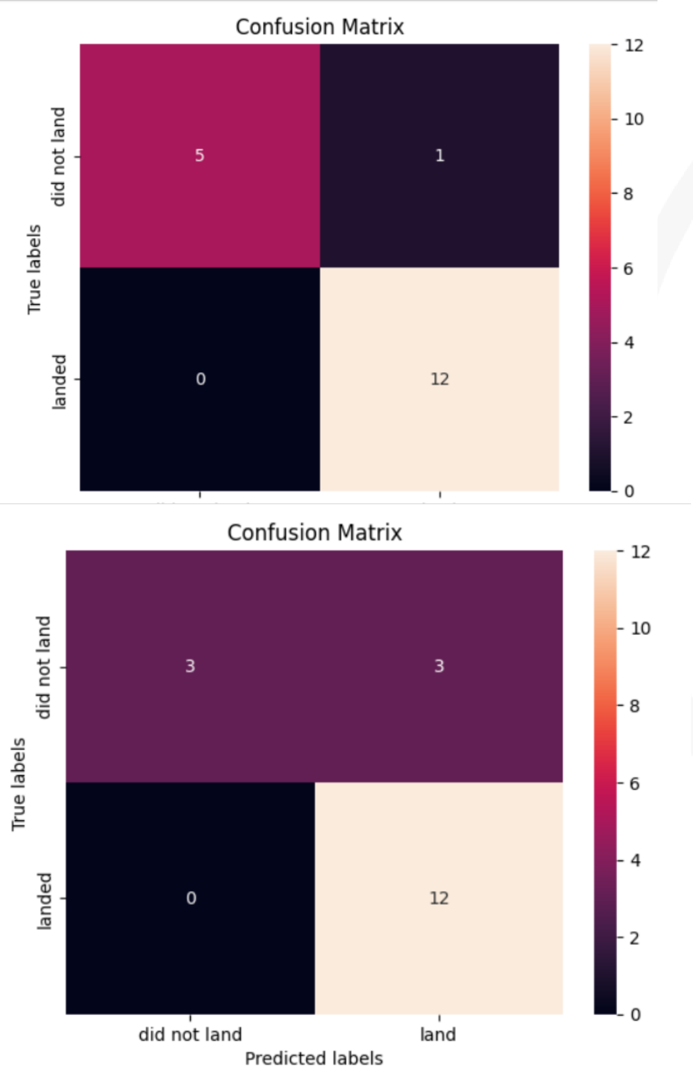




Plotly Dash dashboard results



predictive analysis



Based on our analysis, the Decision Tree Classifier demonstrated the highest accuracy among the models evaluated, achieving approximately 94.4% on the test data. This suggests that the Decision Tree model is well-suited for our classification task, likely due to its ability to handle both numerical and categorical data, its interpretability, and its robustness against co-linearity. Therefore, we recommend utilizing the Decision Tree Classifier for predicting the successful landing of the Falcon 9 first stage.

GitHub link

<https://github.com/bensaadamine/Capstone-Project.git>

CONCLUSION

In this project, we developed and evaluated several classification models to predict the successful landing of SpaceX's Falcon 9 first stage. Among the models tested—Logistic Regression, Support Vector Machine (SVM), Decision Tree, and K-Nearest Neighbors (KNN)—the Decision Tree classifier demonstrated the highest accuracy on the test set, achieving a score of 94.44%. This superior performance suggests that the Decision Tree model effectively captures the underlying patterns in the data related to successful landings. Therefore, we recommend utilizing the Decision Tree classifier for predicting Falcon 9 first stage landing outcomes.