Falcon 9 First Stage Landing Analysis & Prediction

An end-to-end Data Science project exploring SpaceX launch success

Presented by: MALOO BELL Loïck Michel

Project Context & Economic Impact

The Challenge

Predicting whether SpaceX's Falcon 9 first stage will successfully land is more than an engineering question—it's an economic imperative. The ability to reuse the first stage transforms the economics of space launch.

Cost Reduction Reality

Successful first stage recovery reduces launch costs from **$165M+** (competitors using expendable rockets) to just **$62M**—a revolutionary 62% cost savings that disrupts the entire industry.

$62M

Falcon 9 Launch Cost

With first stage reuse

$165M

Competitor Cost

Expendable rockets

62%

Cost Reduction

Through reusability

Project Objectives & Methodology

Primary Goal

Develop a classification model to predict first stage landing success, enabling accurate launch cost estimation and competitive intelligence for the commercial space industry.

01

Data Collection & Wrangling

Web scraping from Wikipedia and SpaceX REST API integration with comprehensive data cleaning

02

Exploratory Data Analysis

Statistical analysis and visualization to uncover patterns and relationships between variables

03

Interactive Visualization

Dynamic dashboards for intuitive data exploration and insight discovery

04

Machine Learning Modeling

Training and evaluation of multiple algorithms to select the optimal prediction model

Phase 1: Web Scraping Wikipedia Data

Notebook: jupyter-labs-webscraping.ipynb

We extracted historical launch data from Wikipedia's comprehensive list of Falcon 9 and Falcon Heavy launches, transforming raw HTML tables into structured, analyzable data.

Technical Approach

- **requests**: HTTP GET operations
- **BeautifulSoup**: HTML parsing engine
- **pandas**: Data manipulation framework

1

HTTP Request

Query static Wikipedia page with launch history

2

HTML Parsing

Create BeautifulSoup object from response content

3

Table Extraction

Identify and isolate launch data table structure

4

Data Cleaning

Extract fields using custom helper functions

5

DataFrame Creation

Export to spacex_web_scraped.csv

Phase 1: SpaceX REST API Integration

Notebook: jupyter-labs-spacex-data-collection-api-v2.ipynb

We enriched our dataset by querying SpaceX's public REST API, retrieving detailed information unavailable through web scraping alone.

Primary Endpoint

GET request to `api.spacexdata.com/v4/launches/past` for complete launch history

Helper Functions

`getBoosterVersion()`, `getLaunchSite()`, `getPayloadData()`, `getCoreData()`

Data Assembly

Structured dictionary converted to comprehensive Pandas DataFrame filtered for Falcon 9 only

Booster Details

Version identifiers, serial numbers, and block information

Launch Sites

Names, coordinates (longitude/latitude), and location metadata

Core Recovery

Landing outcomes, reuse history, and flight counts

Phase 1: Data Cleaning & Target Variable Creation

Notebook: labs-jupyter-spacex-Data wrangling-v2.ipynb

Missing Value Treatment

We identified gaps in **LandingPad** and **PayloadMass** columns. PayloadMass null values were imputed using the column mean, preserving data integrity while maintaining statistical validity.

Outcome Analysis

The original `Outcome` column contained descriptive strings like "True ASDS", "False Ocean", "None None"—requiring transformation into a binary classification target.

1

Define Failure Set

Created failure criteria: False ASDS, False Ocean, False RTLS, None ASDS, None None

2

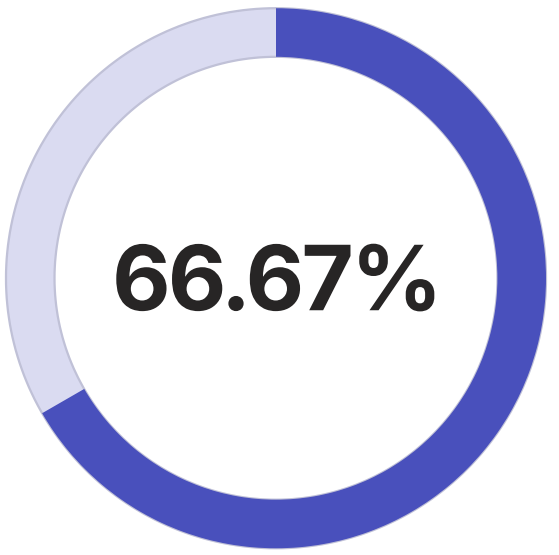
Binary Transform

Generated `landing_class`: 1 for success, 0 for failure

3

Export Dataset

Saved cleaned data to dataset_part_2.csv



Overall Success Rate

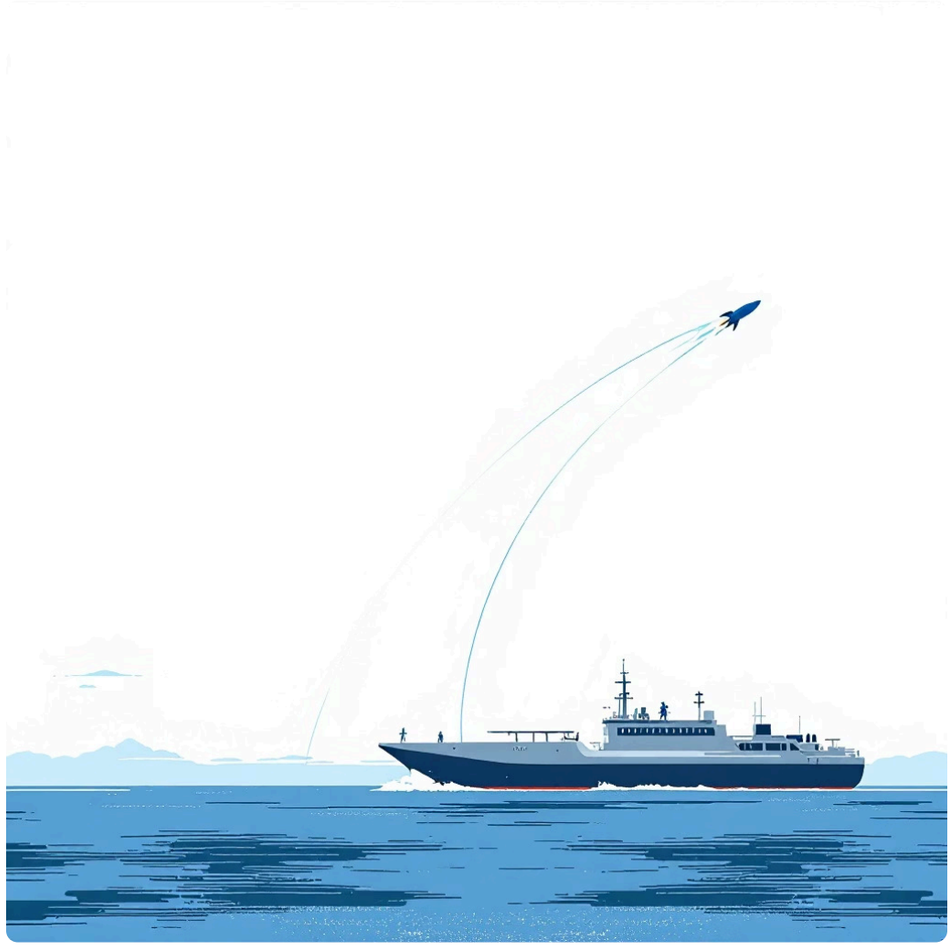
Landing success across all Falcon 9 launches in dataset

Complete Dataset Structure

Our final dataset integrates web-scraped and API data into a comprehensive analytical framework with 18 features plus our target variable.

Column	Description	Type
FlightNumber	Sequential flight identifier	Numeric
Date	Launch date and time	DateTime
BoosterVersion	Falcon 9 booster variant	Categorical
PayloadMass	Payload mass in kilograms	Numeric
Orbit	Target orbit (LEO, GTO, ISS, etc.)	Categorical
LaunchSite	Launch facility location	Categorical
Outcome	Textual landing result description	Categorical
GridFins	Grid fin deployment status	Boolean
Reused	Booster reuse indicator	Boolean
Legs	Landing leg presence	Boolean
Longitude/Latitude	Launch site coordinates	Numeric
landing_class	**Target: Success (1) or Failure (0)**	Binary

Phase 1 Completion: Data Foundation Established



What We've Accomplished

Dual-Source Integration

Combined Wikipedia scraping with SpaceX API for comprehensive coverage

Clean, Structured Data

Resolved missing values and standardized formats across all features

Prediction-Ready Target

Binary `landing_class` variable enables supervised learning

Success Metrics Foundation

With **66.67% baseline success rate**, our model must demonstrate improvement over naive prediction strategies to prove value in production environments.

Next Steps: From Data to Insights

1

Exploratory Data Analysis

Uncover correlations between launch site, orbit type, payload mass, and landing success through statistical visualization

2

Interactive Dashboards

Build Plotly Dash applications enabling dynamic filtering and real-time exploration of launch patterns

3

Machine Learning Pipeline

Train and evaluate multiple algorithms—Logistic Regression, SVM, Decision Trees, Random Forests—to identify optimal predictor

Visualization Goals

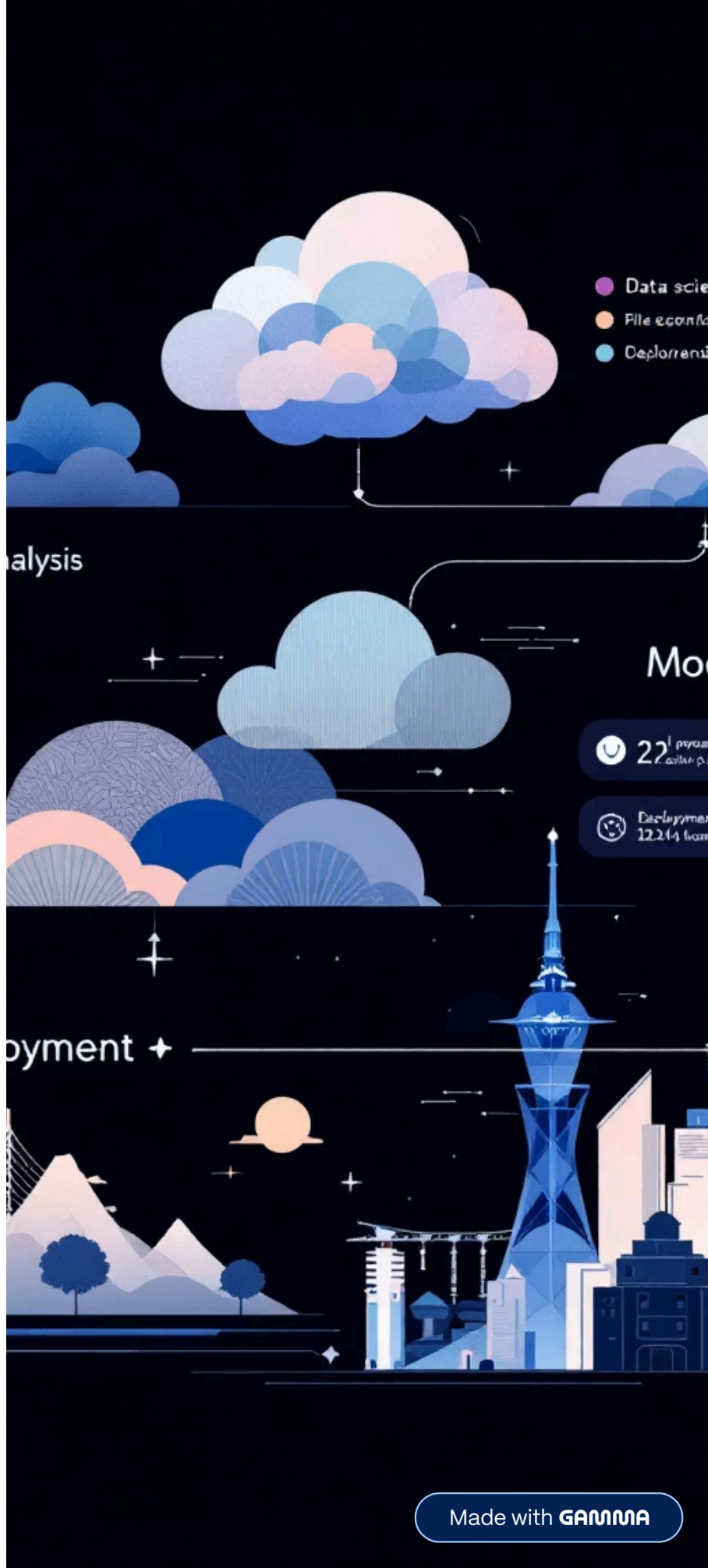
- Site success rates
- Orbit correlation analysis
- Temporal trends

Model Candidates

- Logistic Regression
- Support Vector Machines
- Tree-based ensembles

Success Metrics

- Accuracy & precision
- F1 score optimization
- ROC-AUC evaluation



Project Impact & Vision

This end-to-end data science project demonstrates the complete analytical lifecycle—from raw data acquisition through predictive modeling. By accurately forecasting Falcon 9 first stage landing success, we enable:

Cost Estimation Precision

Competitive intelligence for pricing launch services and understanding SpaceX's cost advantages in the commercial space market

Risk Assessment

Quantitative evaluation of mission success probability based on launch parameters, site selection, and payload characteristics

Strategic Insights

Data-driven understanding of factors influencing reusability success—informing future aerospace engineering decisions

> "The future belongs to those who can transform data into actionable intelligence. This project bridges the gap between raw information and strategic decision-making in the new space economy."