Winning Space Race with Data Science

Rogério Yamada
26 September 2025

IBM Developer SKILLS NETWORK

Outline

1. Executive Summary
2. Introduction
3. Methodology
4. Results
5. Conclusion
6. Appendix

Executive Summary I

- Objective: Predict and explain SpaceX Falcon 9 landing outcomes to inform cost and mission planning
- Data: Combined SpaceX API and Wikipedia; cleaned and standardised key fields (payload, orbit, site, outcomes)
- Methods: SQL for quick insights; visual EDA; Folium maps for geospatial context; Plotly Dash for interactive analysis; baseline classification models

Key Insights

- Success rates improve over time and with flight experience
- Payload mass and orbit type relate to landing outcomes
- Launch sites exhibit distinct success/failure profiles

Executive Summary II

Results

- Actionable EDA on sites, payloads, and orbits
- Interactive dashboard for scenario exploration
- Baseline models demonstrate predictive potential for landings

Business Value

- Better risk/cost estimation and mission planning
- Data-driven selection of payload/orbit/site combinations

Next Steps

- Enrich with weather/sea-state features and latest flights
- Model tuning and deployment with live data refresh

Introduction I

Context

- Falcon 9 reusability drives major cost savings; predicting landing outcomes informs mission economics

Problem

- What factors influence landing success, and can we predict outcomes ahead of launch?

Data Sources

- SpaceX REST API (launch, payload, rocket, launchpad)
- Wikipedia launch records (historical details)

Approach

- Data wrangling and schema standardization
- SQL queries and visual EDA to uncover patterns
- Geospatial exploration with Folium

Introduction II

- Interactive analytics with Plotly Dash
- Baseline classification models for outcome prediction

Scope

- Falcon 9 missions through the available dataset window; focus on payload, orbit, site, and temporal trends

Deliverables

- Clean, analyzable dataset
- Key insights and visualizations
- Interactive dashboard
- Initial predictive models and recommendations

Section 1

Methodology

- Data collection methodology
  - Data about rocket launches was obtained from a SpaceX API and web scraping Wikipedia pages
- Perform data wrangling
  - Missing data was handled, a preliminary Exploratory Data Analysis was performed, and the variable `Outcome Class` was defined for training the supervised models
- Perform Exploratory Data Analysis (EDA) using visualisation and SQL
- Perform interactive visual analysis using Folium and Plotly Dash
- Perform predictive analysis using classification models

Data Collection

- SpaceX API data extraction and Wikipedia pages web scraping were combined to produce a dataset of SpaceX Falcon 9 landings information

data_falcon9

[34]

11 rows · 90 rows × 17 cols

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude
0	1	2010-06-04	Falcon 9	6123.547647	LEO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0003	-80.577366	28.561857
1	2	2012-05-22	Falcon 9	525.000000	LEO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0005	-80.577366	28.561857
2	3	2013-03-01	Falcon 9	677.000000	ISS	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0007	-80.577366	28.561857
3	4	2013-09-29	Falcon 9	500.000000	PO	VAFB SLC 4E	False Ocean	1	False	False	False	None	1.0	0	B1003	-120.610829	34.632093
4	5	2013-12-03	Falcon 9	3170.000000	GTO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B1004	-80.577366	28.561857
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
85	86	2020-09-03	Falcon 9	15600.000000	VLEO	KSC LC 39A	True ASDS	2	True	True	True	5e9e3032383ecb6bb234e7ca	5.0	12	B1060	-80.603956	28.608058
86	87	2020-10-06	Falcon 9	15600.000000	VLEO	KSC LC 39A	True ASDS	3	True	True	True	5e9e3032383ecb6bb234e7ca	5.0	13	B1058	-80.603956	28.608058
87	88	2020-10-18	Falcon 9	15600.000000	VLEO	KSC LC 39A	True ASDS	6	True	True	True	5e9e3032383ecb6bb234e7ca	5.0	12	B1051	-80.603956	28.608058
88	89	2020-10-24	Falcon 9	15600.000000	VLEO	CCSFS SLC 40	True ASDS	3	True	True	True	5e9e3033383ecbb9e534e7cc	5.0	12	B1060	-80.577366	28.561857
89	90	2020-11-05	Falcon 9	3681.000000	MEO	CCSFS SLC 40	True ASDS	1	True	False	True	5e9e3032383ecb6bb234e7ca	5.0	8	B1062	-80.577366	28.561857

Data Collection – SpaceX API

- From the SpaceX API endpoint `https://api.spacexdata.com/v4/` we probed the following data sources:
  - rocket
  - payload
  - launchpad
  - cores
- Jupyter Notebook's GitHub URL

Data Collection – Scraping

- From the Wikipedia List of Falcon 9 and Falcon Heavy launches web page we collected Falcon 9 historical launch records
- Jupyter Notebook's GitHub URL

Data Wrangling

- Through data wrangling, the variable `Outcome Class` was defined for training the supervised models
- Jupyter Notebook's GitHub URL

EDA with SQL

- Several SQL queries have been processed to gain insights about the landing outcomes:
  - Launching sites
  - Total payload mass carried by specific boosters in specific sites
  - Successful and failed landing outcomes
- Jupyter Notebook's GitHub URL

EDA with Data Visualisation

- Several charts have been drawn to gain insights about the landing outcomes:
  - Flight Number versus Launch Site by Class
  - Payload Mass versus Launch Site by Class
  - Success Rate by Orbit
  - Flight Number versus Orbit by Class
  - Payload Mass versus Orbit by Class
  - Yearly Launch Success Rate
- Jupyter Notebook's GitHub URL

Build an Interactive Map with Folium

- A geographical analysis has been performed to gain insights about the dependencies of landing outcomes and location and surrounding of launching sites:
  - Marking all launch sites on a map
  - Marking the success/failed launches for each site on the map
  - Calculating the distances between a launch site to its proximities
- Jupyter Notebook's GitHub URL

Build a Dashboard with Plotly Dash

- A dashboard has been implemented to perform real-time analysis about landing outcomes considering:
  - Launch site drop-down menu
  - Interactive successful landing outcome pie chart
  - Range slider for selecting payload mass
  - Interactive successful landing outcome scatter plot
- Jupyter Notebook's GitHub URL

Predictive Analysis (Classification)

- The following predictive models have been considered:
  - Logistics Regression
  - Support Vector Machine
  - Decision Tree
  - K Nearest Neighbours
- Jupyter Notebook's GitHub URL

Results

- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

Flight Number versus Launch Site

- There is a concentration of unsuccessful landing outcomes in the site CCAFS SLC 40
- As the number of flights increase, the successful landing outcomes are likely to increase

Payload versus Launch Site

- There is a concentration of unsuccessful landing outcomes in CCAFS SLC 40 for payloads lighter than 7000 kg

Success Rate versus Orbit Type

- The success rate is higher when the orbit type is ES-L1, GEO, HEO, and SSO
- The success rate is lower when the orbit type is GTO and SO

Flight Number versus Orbit Type

- The success rate is higher as the number of flights increases for orbit type LEO, ISS, and PO
- There seems to be no relationship between flight number when the orbit type is GTO

Payload versus Orbit Type

- With heavy payloads, the successful landing rate is higher for PO, LEO, and ISS orbit types
- However, regarding GTO, we cannot distinguish this well as both positive landing rate and negative landing (unsuccessful mission) occur almost evenly

Launch Success Yearly Trend

- The success rate since 2013 kept increasing until 2017 (stable in 2014) and after 2015 it started increasing

All Launch Site Names

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

- This is a sorted list of unique, non-null values from the Launch_Site column in the SPACEXTABLE table

Launch Site Names Begin with 'CCA'

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

- This is a set of up to 5 rows from the SPACEXTABLE table where the Launch_Site column starts with the text "CCA" (e.g., "CCAFS . . ."). The LIKE 'CCA%' filter matches any value beginning with "CCA"

Total Payload Mass

Total Payload Mass = 45596

- This is the total payload mass (sum of PAYLOAD_MASS__KG_) for all rows in SPACEXTABLE where the Customer is exactly 'NASA (CRS)'. The result is a single value labeled total_payload_mass

Average Payload Mass by F9 v1.1

Average Payload Mass = 2534.6666666666665

- This is the average payload mass (PAYLOAD_MASS__KG_) over rows in SPACEXTABLE whose Booster_Version starts with "F9 v1.1", returning a single value named avg_payload_mass

First Successful Ground Landing Date

First Ground Pad Landing Success Date = 2015-12-22

- This is the earliest Date (minimum) among rows where Landing_Outcome is exactly "Success (ground pad)", labeled first_ground_pad_success_date.

Successful Drone Ship Landing with Payload between 4000 and 60000

booster_name
F9 FT B1021.2
F9 FT B1031.2
F9 FT B1022
F9 FT B1026

- These are the unique booster versions (as booster_name) that had a "Success (drone ship)" landing and carried payloads strictly between 4000 kg and 6000 kg, sorted alphabetically

Total Number of Successful and Failure Mission Outcomes

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

- These are rows grouped by Mission_Outcome, counting how many are successes or failures (strings starting with "Success" or "Failure"), and returns each outcome with its count, sorted by count descending

Boosters Carried Maximum Payload

Booster_Version
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

- These are the unique booster versions that carried the maximum payload mass found in the table. The subquery gets the global MAX(PAYLOAD_MASS__KG_), and the outer query lists distinct Booster_Version rows matching that value

2015 Launch Records

month	Landing_Outcome	Booster_Version	Launch_Site
01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

- This is a list of 2015 records where Landing_Outcome is "Failure (drone ship)"

Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

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

- This is a calculation of how many launches fall into each Landing_Outcome between 2010-06-04 and 2017-03-20, returning each outcome with its count, sorted by count descending

Section 4

Launch Sites Proximities Analysis

All launch sites map

- All launch sites are plotted on the map

Success/Failed launches

- The successful and failed launches are clustered on the map

Coastline point map

- There a distance of 0.51 km between the CCAFS SLC-40 launch site and the coastline

Section 5

Build a Dashboard with Plotly Dash

Launch site drop-down menu

- Launch site drop-down menu

Interactive successful landing outcome pie chart

- Interactive successful landing outcome pie chart

Complete Dashboard

- This is the complete dashboard with the following elements:
  - Launch site drop-down menu
  - Interactive successful landing outcome pie chart
  - Range slider for selecting payload mass
  - Interactive successful landing outcome scatter plot

Section 6

Predictive Analysis (Classification)

Classification Accuracy

- There is a draw between the Logistic Regression, Support Vector Machine (SVM) and K Nearest Neighbours models

Confusion Matrix

- Logistic Regression confusion matrix

Conclusions I

- We built a unified SpaceX launches dataset via API + web scraping and cleaned it for analysis
- SQL and visual EDA revealed patterns: site- and orbit-specific success rates, payload/flight effects, and yearly improvement trends
- Interactive Folium maps and a Plotly Dash app enabled exploratory insights by site, payload, and outcomes
- Classification models provided baseline predictability for landing outcomes

Key Takeaways

- Landing success improves over time and with experience (flight count)
- Payload mass and orbit type correlate with outcome patterns
- Certain launch sites show distinct success/failure profiles.

Conclusions II

Business Impact

- Better launch planning and risk estimation
- Data-driven targeting of missions (payload/orbit/site combinations)
- Foundation for cost optimization via landing outcome prediction.

Next Steps

- Incorporate weather and wind data for richer features
- Expand to more recent missions and Starship data
- Model tuning and feature engineering for higher accuracy
- Deploy the dashboard with live data refresh.

Appendix

Dashboard

127.0.0.1:8050
WhatsApp
Toggl Track
Bitwarden
Portal do Empreen...
Restituição do IRP...
Calculadora do CL...
Check IP AWS
Date Calculator
Media Converter
SpaceX Launch Records Dashboard
All Sites
Total Successful Launches by Site
29.2%
41.7%
16.7%
12.5%
KSC LC-39A
CCAFS LC-40
VAFB SLC-4E
CCAFS SLC-40
Payload range (Kg):
0
1k
2k
3k
4k
5k
6k
7k
8k
9k
10k
Correlation between Payload and Success for all Sites
Class
1
0.8
0.6
0.4
0.2
0
0
2k
4k
6k
8k
10k
Payload Mass (kg)
Booster Version Category
v1.0
v1.1
FT
B4
B5

Thank you!