

Report

Topic: SpaceX Launch Performance and Trends (2006–2025)

Goal: studying the frequency of launches, the success and trends in the use of rockets.

Collecting real API data and one static web page

Data Collection :

API Source: GET <https://api.spacexdata.com/v5/launches/latest> (from <https://github.com/r-spacex/SpaceX-API>)

Data fields: mission name, rocket type, launch date, and success status.

Web Scraping: https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches(from Wikipedia – List of Falcon 9 and Falcon Heavy launches)

Data fields: launch date, rocket version, payload, and outcome

Cleaning, combining, and converting these datasets

1.Initial data inspection

After data collection from two sources —

SpaceX API and Wikipedia table —

the first step was to inspect the structure of both datasets using: `df.head()`, `df.info()`, `df.columns`

```
df_api.shape=(205, 43)
```

```
df_web.shape=(272, 4)
```

2.Data cleaning

Handling missing and duplicated values

- Checked for nulls and duplicates with:
`df.isna().sum()`, `df.duplicated().sum()`
- Removed empty or duplicated rows:
`df.drop_duplicates(inplace=True)`

3.Column renaming and selection

Source	Selected Columns	Renamed As
API	name, date_utc, rocket, success	name, date_utc, rocket, result
Wikipedia	Date and time (UTC), Version, booster[j], Payload[k], Launch outcome	date_utc, rocket, name, result

4. Date conversion and standardization

The API provided ISO 8601 timestamps (e.g., `2006-03-24T22:30:00.000Z`), while Wikipedia had natural-language formats (e.g., `January 3, 2024 03:44[23]`).

- All non-numeric characters and references like `[23]` were removed
- Converted both columns to the same datetime format
- Unified timezones (API had `UTC`, Wikipedia was naive)

5. Merging two data sources

Because both datasets describe SpaceX launches but have different coverage, they were merged *vertically* (not by key)

This combined older missions (API, 2006–2023) and recent launches (Wikipedia, 2024–2025). Duplicates by date and mission name were removed.

Result

After cleaning and merging:

The final dataset contained ~300+ launches from 2006 to 2024

The data was ready for EDA — grouping by year, analyzing success rates, and comparing rockets.

EDA and visualizations:

1. Number of launches by year

- It shows how SpaceX has grown over time.
- Visualization: linear graph (`plt.plot`)
- Interpretation: the growth of launches after 2015 and stabilization after 2020 are visible.

2,Percentage of successful missions

- It shows how much launch reliability has improved.
- Visualization: linear graph (plt.plot)
- Interpretation: in 2006-2010 there were many failures, then almost 100% successes.

3.Which missiles were used most often

- Shows the share of Falcon 9, Heavy, etc.
- Visualization: bar chart (plt.bar)
- Interpretation: Falcon 9 is the main workhorse of SpaceX.

Key Insights

- SpaceX significantly increased launch frequency from <5 per year in 2010 to >90 per year in 2024.
- Reliability has improved drastically, showing nearly 100% success in recent years.
- Falcon 9 remains the most frequently used and reliable rocket, supporting Starlink and other missions.
- The strong positive trend indicates technological maturity and operational scalability.

Tools and Methods

- **Python libraries:** requests, pandas, matplotlib, seaborn
- **Data sources:** REST API + HTML Table (Web scraping & API)
- **Visualization methods:** bar chart, line plot