Applied Data Science Capstone Project Report

SpaceX Competitor Analysis: Predicting Falcon 9 First-Stage Landing Success

Author: Reham Hassan

Email: reham552002@gmail.com

Institution: IBM Data Science Professional Certificate

Table of Contents

---

1. Executive Summary

This report details a data science project analyzing SpaceX's Falcon 9 rocket launches to predict first-stage landing success—a critical factor in cost efficiency. The goal was to provide **Space Y**, a hypothetical competitor, with actionable insights to optimize launch strategies.

Key Findings:

- **Cost Advantage:** SpaceX's reusable Falcon 9 costs **62Mperlaunch∗∗,comparedtocompetitors(62*Mperlaunch*∗∗,*comparedtocompetitors*(165M+).
- **Success Factors:** Payload mass, launch site, and orbit type significantly impact landing success.
- **Best Model:** Logistic Regression achieved **85% accuracy** in predicting successful landings.

Recommendations for Space Y:

- Focus on **Low Earth Orbit (LEO)** missions with payloads <5,000 kg for higher success rates.
- Invest in **first-stage recovery technology** to reduce costs.

---

2. Introduction

Project Background

SpaceX dominates the commercial space industry due to reusable rockets. **Space Y**, a startup founded by industrialist "Allon Musk," aims to compete by optimizing launch pricing and reusability.

Problem Statement

- Can we predict whether the **Falcon 9's first stage will land successfully** based on historical data?
- What factors (payload, launch site, orbit) most influence landing success?

Objectives

1. Collect and clean SpaceX launch data.
2. Perform EDA to identify trends.
3. Build a predictive model for landing success.
4. Create an interactive dashboard for stakeholders.

---

3. Methodology

Data Collection

- **Primary Source:** SpaceX REST API (launches, rockets, payloads).
- **Secondary Source:** Web scraping with **BeautifulSoup** for supplementary data.

Data Wrangling

- **Handling Missing Data:**
  - 26 missing values in `LandingPad` (filled with "None").
- **Feature Engineering:**

- Converted `Outcome` to binary `Class` (1=successful landing, 0=failure).
- Extracted `Year` from `Date` for trend analysis.

Exploratory Data Analysis (EDA)

- **Launch Sites:**
  - **CCAFS SLC 40** (Most active), **VAFB SLC 4E**, **KSC LC 39A**.
- **Success Rate by Orbit:**
  - **LEO (Low Earth Orbit):** 92% success rate.
  - **GTO (Geostationary Orbit):** 58% success rate.
- **Payload vs. Success:**
  - Missions with payloads <5,000 kg had **higher landing success**.

Machine Learning Modeling

Pipeline:

1. **Preprocessing:** Scaled features using `StandardScaler`.
2. **Train-Test Split:** 80% training, 20% testing.
3. **Models Evaluated:**
   - Logistic Regression
   - Decision Tree
   - Support Vector Machine (SVM)
   - K-Nearest Neighbors (KNN)

Evaluation Metric: Accuracy

Model	Accuracy
Logistic Regression	85%
Decision Tree	83%
SVM	83%

Model	Accuracy
KNN	80%

Best Model: Logistic Regression (85% accuracy).

---

4. Results & Analysis

Key Insights

1. **Reusability Saves Costs:**
   - SpaceX's Falcon 9 is **60% cheaper** due to reusable first stages.
2. **Optimal Launch Conditions:**
   - **LEO missions** with **lighter payloads** have the highest success rates.
3. **Launch Site Matters:**
   - **CCAFS SLC 40** had the highest number of successful landings.

Interactive Dashboard (Plotly Dash)

- **Features:**
  - Dropdown filters for launch sites, years, and payload ranges.
  - Real-time success rate visualizations.

---

5. Discussion

Business Implications

- **Space Y should:**
  - Prioritize **LEO missions** to maximize landing success.
  - Develop **recovery infrastructure** for first-stage reuse.

Limitations

- Data limited to **publicly available SpaceX launches**.
- Assumes **similar conditions** for Space Y's rockets.

---

6. Conclusion & Recommendations

Conclusion

- **Logistic Regression** reliably predicts first-stage landing success (85% accuracy).
- **LEO missions with payloads <5,000 kg** are most viable for cost efficiency.

Recommendations

1. **Focus on LEO launches** to improve reusability.
2. **Invest in drone ships** for ocean landings (higher success than ground landings).
3. **Monitor competitor data** for ongoing strategy adjustments.

---

7. Appendix

Code Snippets

python

Copy

Download

```python
# Data Wrangling: Handling Missing Values
df['LandingPad'].fillna("None", inplace=True)
```

```python
# Logistic Regression Model
from sklearn.linear_model import LogisticRegression
model = LogisticRegression()
model.fit(X_train, y_train)
```

Data Samples

FlightNumber	Date	BoosterVersion	PayloadMass (kg)	Orbit	LaunchSite	Class
1	2010-06-04	Falcon 9	6104.93	LEO	CCAFS SLC 40	0
2	2012-05-22	Falcon 9	525.00	LEO	CCAFS SLC 40	1

References

1. SpaceX API Documentation: https://api.spacexdata.com/v4/
2. Scikit-learn: Machine Learning in Python.

- **PowerPoint Presentation**
- **Interactive Dashboard** : https://spacex.streamlit.app/