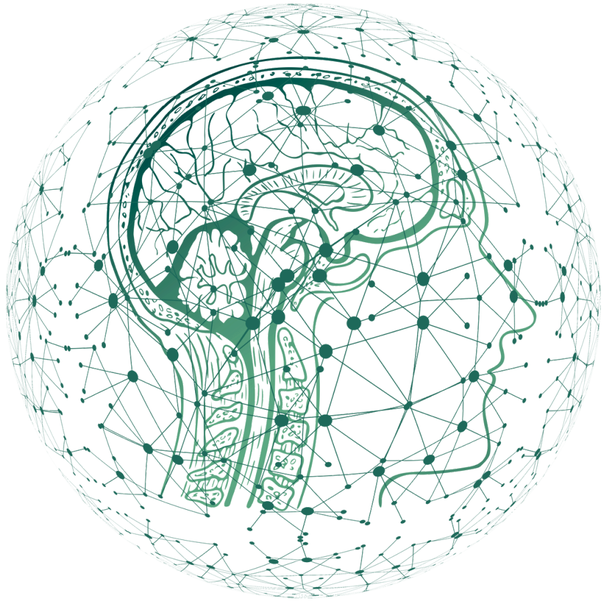
رواد مصر الرقمية



IBM Data Science
Applied Data Science Capstone

Presented by: Reham Hassan

Project Overview

Course Objective:
Assume the role of a Data Scientist working
for Space Y, a startup competing with SpaceX.

Key Tasks:

- Data collection, wrangling, and exploration
- Predictive modeling for Falcon 9 first-stage landing success
- Interactive dashboard creation
- Stakeholder reporting

View The Full Code Notebooks From Here:

Project Scenario

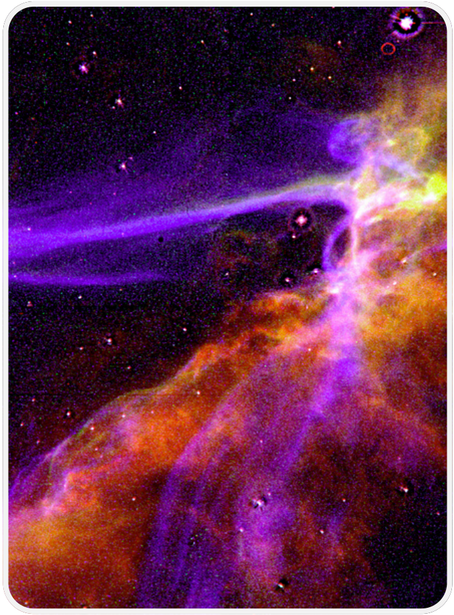
Space Industry Landscape:

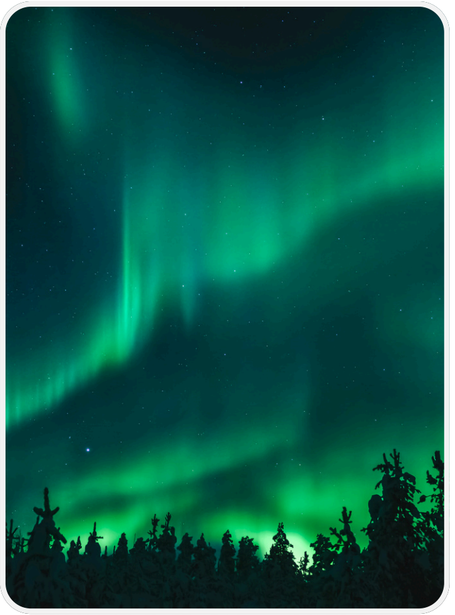
- Major players: SpaceX (62Mperlaunch),VirginGalactic,BlueOrigin(62Mperlaunch),VirginGalactic,BlueOrigin(165M+ per launch).
- SpaceX’s cost advantage: Reusable first stage.

Space Y’s Goal:

- Determine competitive launch pricing.
- Analyze first-stage reuse feasibility.

Data Collection & Wrangling





01

Data Sources:

- SpaceX REST API (launches, rockets, payloads).
- Web scraping with BeautifulSoup.

02

Challenges Addressed:

- Handling null values (e.g., 26 missing LandingPad entries).
- Feature engineering (e.g., converting Outcome to binary Class: 1=landed, 0=not landed).

Exploratory Data Analysis (EDA)

01

Key Insights:

- Launch sites: CCAFS SLC 40, VAFB SLC 4E, KSC LC 39A.
- Payload mass vs. orbit type: LEO (low-cost) vs. GTO (high-cost).
- Success rate trends (2010–2020).

02

Visualization:

- Folium maps for launch site locations.
- Plotly interactive charts for success rates by year.

Predictive Modeling

Goal: Predict first-stage landing success.

Pipeline:

1. **Preprocessing**: sklearn.preprocessing for feature scaling.
2. **Train-Test Split**: 80/20 ratio.
3. **Models Tested:**
   - Logistic Regression (Accuracy: 0.85)
   - Decision Tree (Accuracy: 0.83)
   - SVM (Accuracy: 0.84)
4. **Best Model**: Logistic Regression.

Interactive Dashboard

Tools Used: Streamlit Dashboard.

Features:

- Dropdown filters for launch sites and years.
- Dynamic graphs showing success rates and payload ranges.
- Geo-map highlighting optimal launch locations.

Purpose: Enable stakeholders to explore data-driven insights.

Interactive Dashboard

🚀 SpaceX Launch Analytics Dashboard

Data preview:

	FlightNumber	Date	Version Booster	Payload mass	Orbit	Launch site	Launch outcome
0	1	2010-06-04	Falcon 9	6,104.9594	LEO	CCAFS SLC 40	None None
1	2	2012-05-22	Falcon 9	525	LEO	CCAFS SLC 40	None None
2	3	2013-03-01	Falcon 9	677	ISS	CCAFS SLC 40	None None
3	4	2013-09-29	Falcon 9	500	PO	VAFB SLC 4E	False Ocean
4	5	2013-12-03	Falcon 9	3,170	GTO	CCAFS SLC 40	None None

Select Launch Site:

ALL

Select Payload Range (kg):

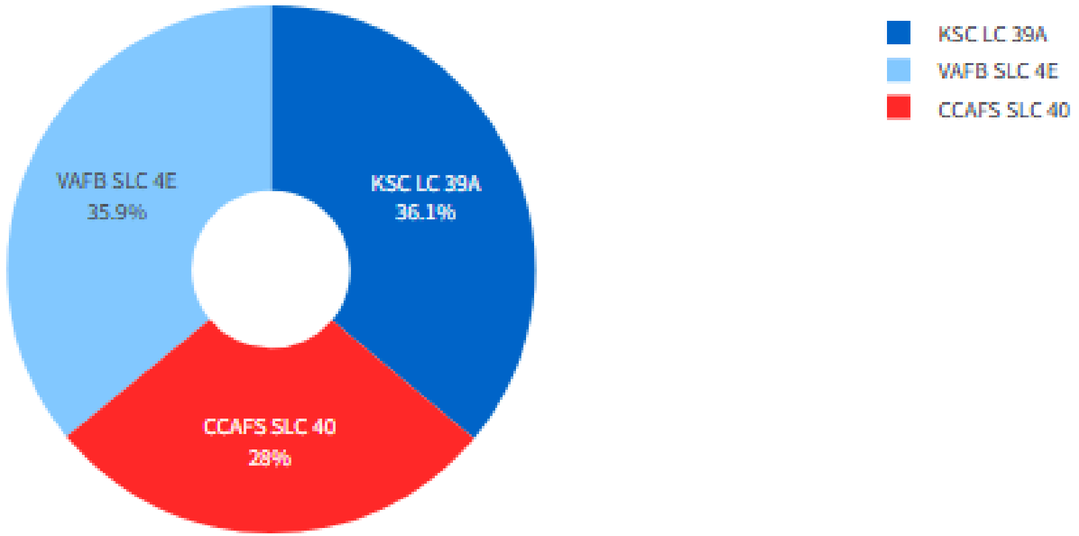
1582 — 15400

0 — 16100

Interactive Dashboard

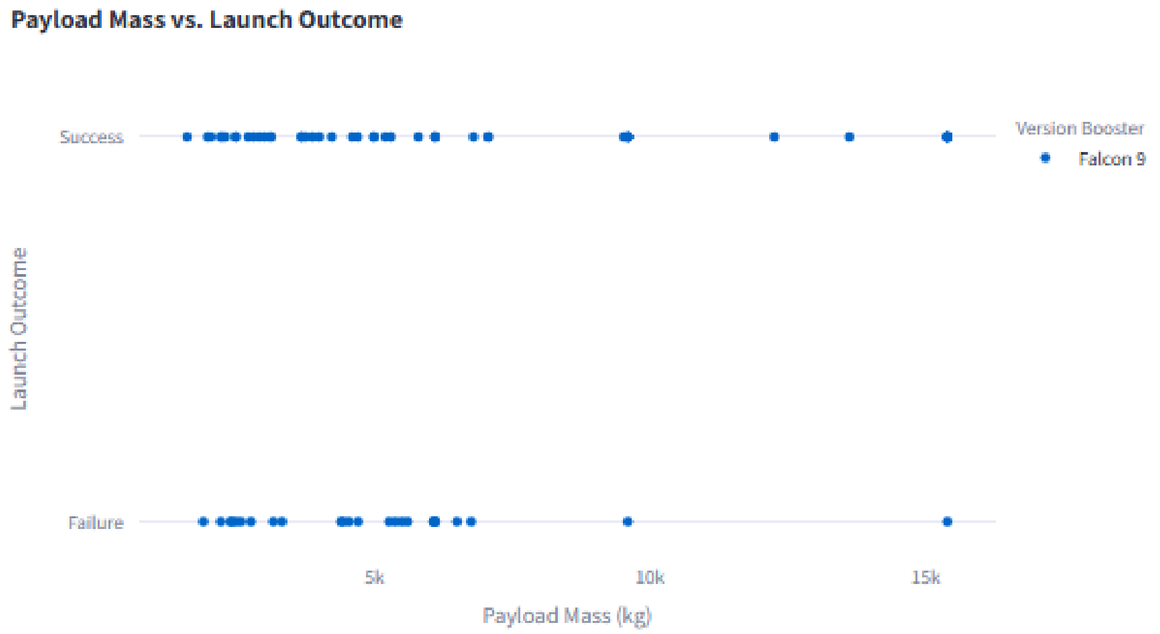
Launch Success Rate

Success Rate by Launch Site



Interactive Dashboard

📈 Payload vs. Launch Outcome



ℹ Dataset Summary

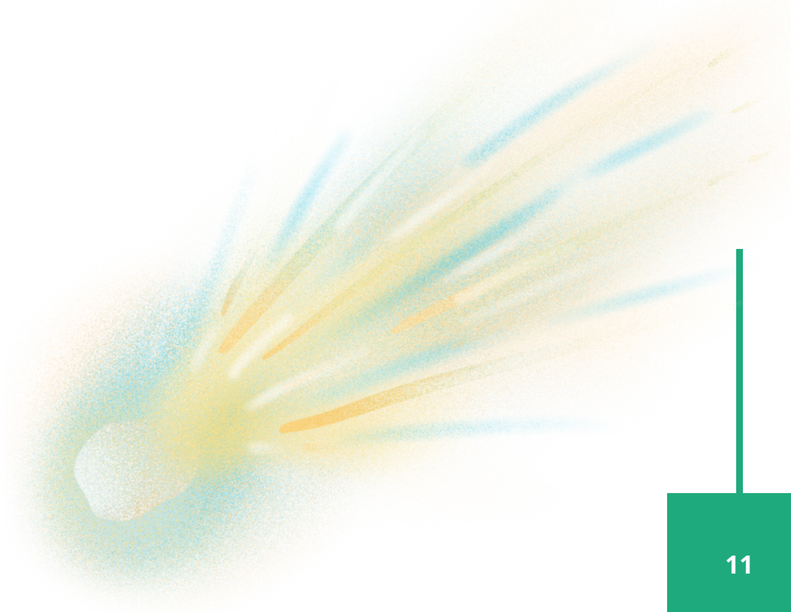
Conclusion & Recommendations

Findings:

- Reusability reduces costs by ~60% (vs. competitors).
- High success rates for LEO missions with payloads <5,000 kg.

Next Steps:

- Optimize launch parameters for Space Y’s first-stage recovery.
- Target LEO missions to undercut SpaceX’s pricing.



CONTACT US

linkedIn

reham552002@gmail.com