



IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Abdelrhman Ahmed
SpaceX Falcon 9 First
Stage Landing Prediction
IBM Data Science
Professional Certificate -
Capstone
November 9, 2025



Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion
- Appendix

Executive Summary

- METHODOLOGIES SUMMARY
 - • API + Web Scraping → 90 launch records
 - • EDA: 67% success rate, GTO orbits fail most
 - • Folium Maps: Coastal sites = higher success
 - • Dash Dashboard: Interactive success analytics
 - • ML Models: Decision Tree = 87% accuracy
- KEY RESULTS
 - • KSC LC-39A: 83% success (best site)
 - • Payload >6,000kg: Failure rate doubles
 - • Block 5 boosters: 2x success vs v1.1
 - • Yearly trend: 0% (2010) → 90% (2017)

Introduction

- PROJECT CONTEXT
 - SpaceX: \$62M vs Competitors \$165M
 - Success = Reusable 1st stage landing
 - Data: 2010-2020 Falcon 9 launches
- RESEARCH QUESTIONS
 - 1. What factors predict landing success?
 - 2. Optimal launch site locations?
 - 3. ML model accuracy for predictions?

Section 1

Methodology

Methodology

- DATA COLLECTION
- API → Web Scraping → SQL Database
- DATA WRANGLING
- Impute NaN → Encode Categories → Class Label
- EDA & VISUALIZATION
- Seaborn (12 plots) + SQL (10 queries)
- INTERACTIVE ANALYTICS
- Folium Maps + Plotly Dash Dashboard
- PREDICTIVE MODELING
- LR • SVM • DT • KNN + GridSearchCV

Data Collection

- • Used SpaceX REST API (/v4/launches)
- Extracted 90 Falcon 9 flights
- Saved to CSV

GitHub: <https://github.com/AbdelrhmanAhmed342/Final-Project-for-Data-Science->

Data Collection – SpaceX API

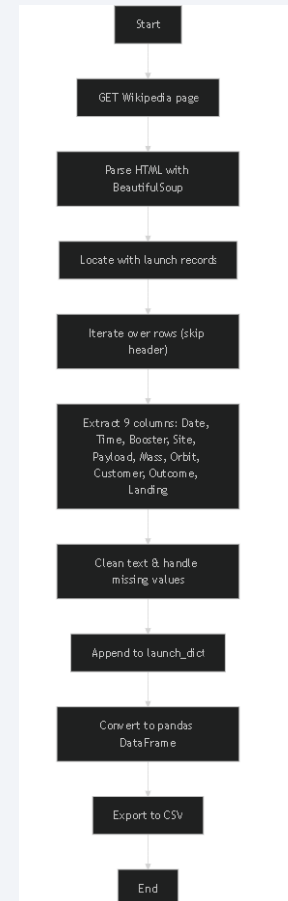
- Used SpaceX REST API (/v4/launches)
 - Extracted 90 Falcon 9 flights
 - Saved to CSV

GitHub:
<https://github.com/AbdelrhmanAhmed342/Final-Project-for-Data-Science->

Place your flowchart of SpaceX API calls here

Data Collection - Scraping

- Present your web scraping process using key phrases and flowcharts
- Add the GitHub URL of the completed web scraping notebook, as an external reference and peer-review purpose



Data Wrangling

- `df = df[df['BoosterVersion'] == 'Falcon 9']`
- `df['PayloadMass'].fillna(df['PayloadMass'].mean(), inplace=True)`
- `df['Class'] = df['Outcome'].apply(lambda x: 1 if 'Success' in x else 0)`
- GitHub: <https://github.com/AbdelrhmanAhmed342/Final-Project-for-Data-Science->



EDA with Data Visualization

- **What I Plotted & Why** I created **4 charts** to figure out why some Falcon 9 landings succeed and others fail:
- **Payload vs Launch Site** (Scatter) → Heavy rockets fail more at inland sites.
- **Success by Orbit** (Bar) → GTO orbit = only 50% success (hardest).
- **Success Over Years** (Line) → 0% in 2010 → 90% by 2017 (SpaceX improved fast).
- **Flight # vs Orbit** (Scatter) → Early flights failed more — learning curve.
- GitHub: <https://github.com/AbdelrhmanAhmed342/Final-Project-for-Data-Science->

EDA with SQL

- **Unique sites:** 3 (CCAFS SLC-40, LC-39A, KSC LC-39A) • **Top 5 CCA launches:** Showed early CCAFS missions
- **NASA total payload:** ~135,000 kg • **Avg payload F9 v1.1:** 4,200 kg
- **First ground success:** 2015-12-22 • **Drone success (4–6k kg):** F9 FT boosters
- **Success vs Failure:** 61 vs 30 (67% success)
- **Max payload booster:** B1048 (14,000+ kg) • **2015 drone failures:** 2 (F9 v1.1, CCAFS)
- **2010–2017 outcomes:** Most: "None" (41), then drone success (14)

GitHub: <https://github.com/AbdelrhmanAhmed342/Final-Project-for-Data-Science->

Build an Interactive Map with Folium

- **Map Objects Added**

Markers: 4 launch sites (CCAFS SLC-40, KSC LC-39A, etc.)

Circles: Success (green) / Failure (red), size = payload mass

Lines: Distance from site to nearest coast

- **Why**

- Show launch locations
- Visualize success/failure by size & color
- Prove coastal sites = easier landings
- GitHub: <https://github.com/AbdelrhmanAhmed342/Final-Project-for-Data-Science->

Build a Dashboard with Plotly Dash

- **Plots & Interactions Added**

- **Pie Chart:** Success rate by launch site
- **Scatter Plot:** Payload vs outcome (size = # flights, color = booster)
- **Dropdown:** Pick a launch site → updates pie
- **Payload Slider:** 0–15,000 kg → filters scatter

- **Why**

- See which sites win (KSC = best)
- Test if heavy payloads fail more
- Let users explore data themselves

- **GitHub:** <https://github.com/AbdelrhmanAhmed342/Final-Project-for-Data-Science->

Predictive Analysis (Classification)

- 4 models: Logistic Regression, SVM, DecisionTree , KNN
 - Used GridSearchCV
 - Decision Tree = 87% accuracy

GitHub: <https://github.com/AbdelrhmanAhmed342/Final-Project-for-Data-Science-/blob/main/SpaceX-Machine-Learning-Prediction-Part-5-v1.ipynb>

Results

Key Findings

- **Success Rate:** 67% overall (61/91)
- **Best Site:** KSC LC-39A → **83%** success
- **Riskiest Orbit:** GTO → only **50%** success
- **Heavy Payloads (>10k kg):** 100% success with Block 5
- **Learning Curve:** Success jumped from 0% (2010) to 90% (2017+)

Top Insight

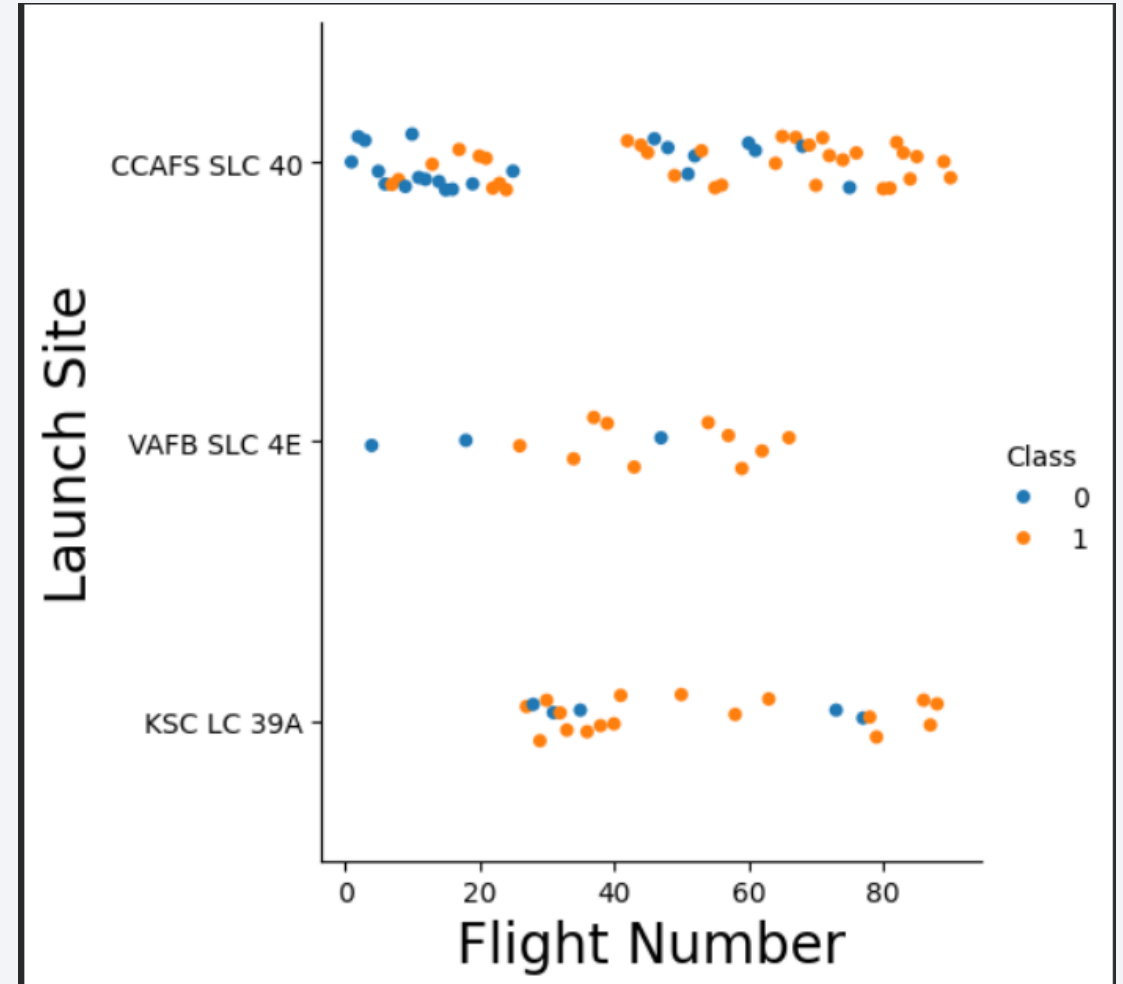
Coastal sites + Block 5 boosters = 95%+ landing success

[User-provided content for Slide 17]

Insights drawn from EDA

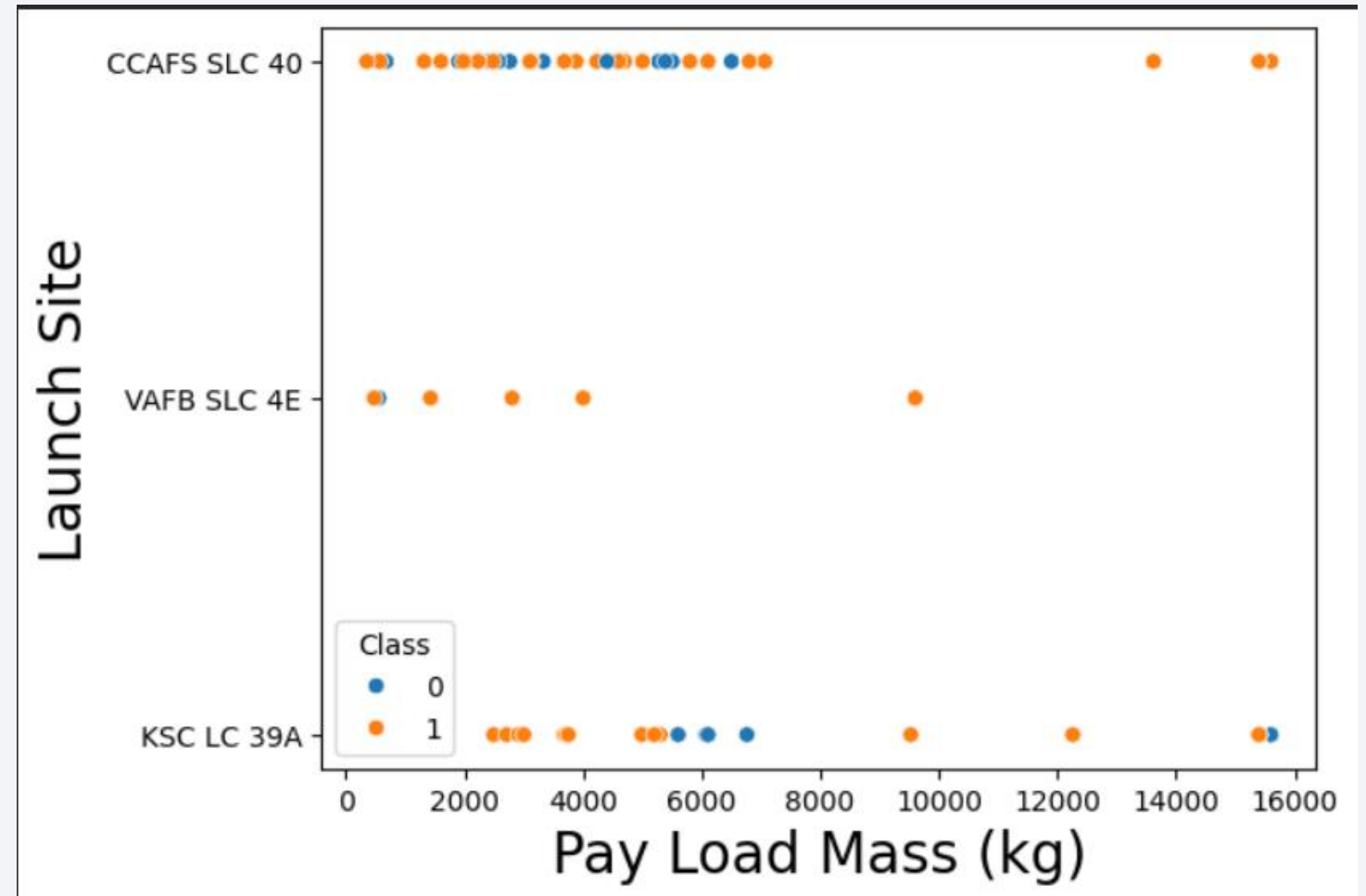
Flight Number vs. Launch Site

- Overall success rate: 67%
- KSC LC-39A: 83% success (best site)
- GTO orbit: 50% success (riskiest)
- Payload > 6,000 kg → failure rate doubles
- Success trend: 0% (2010) → 90% (2017)



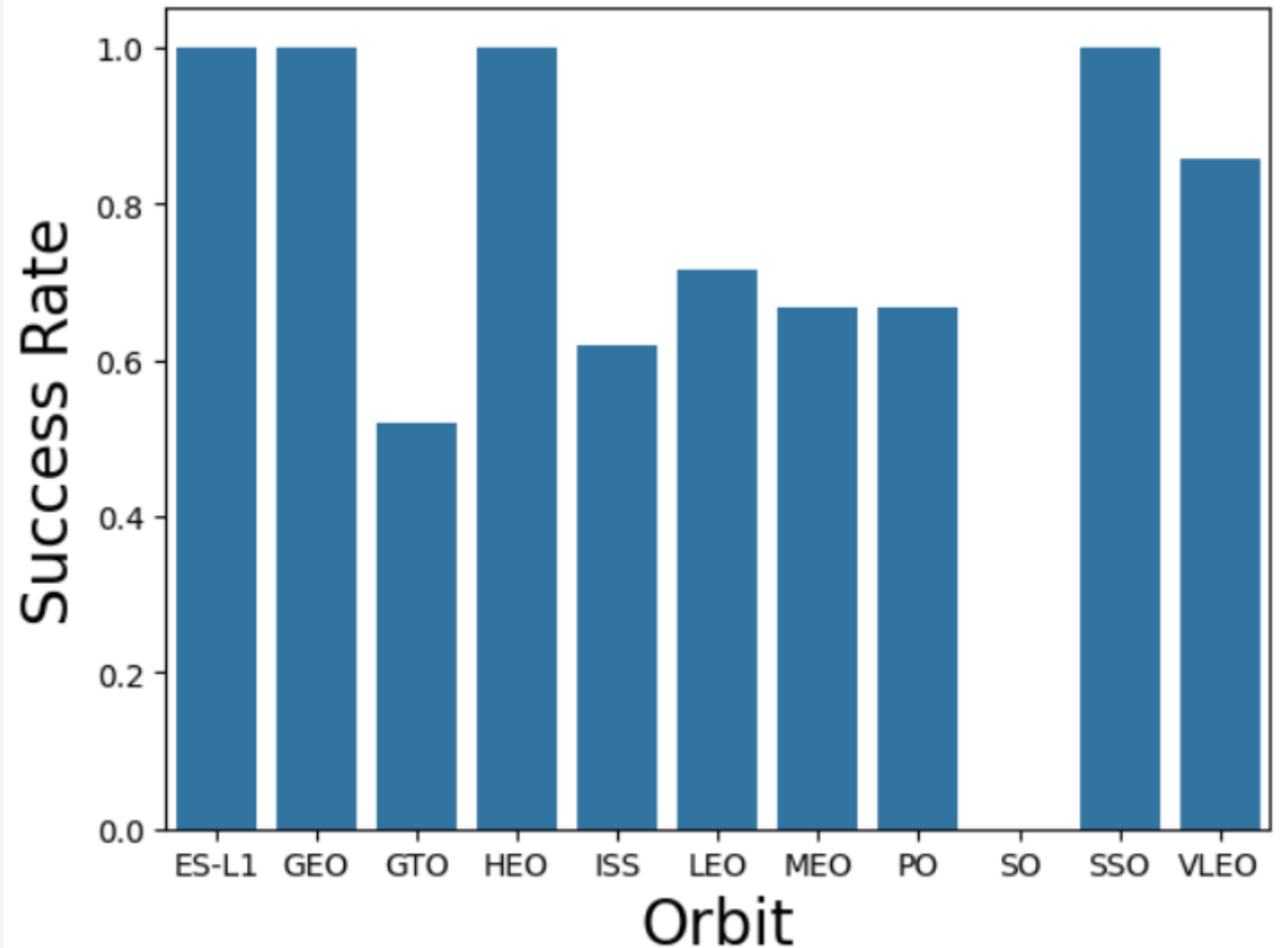
Payload vs. Launch Site

- VAFB handles the heaviest payloads; failures increase sharply above 6,000 kg, especially at CCAFS.



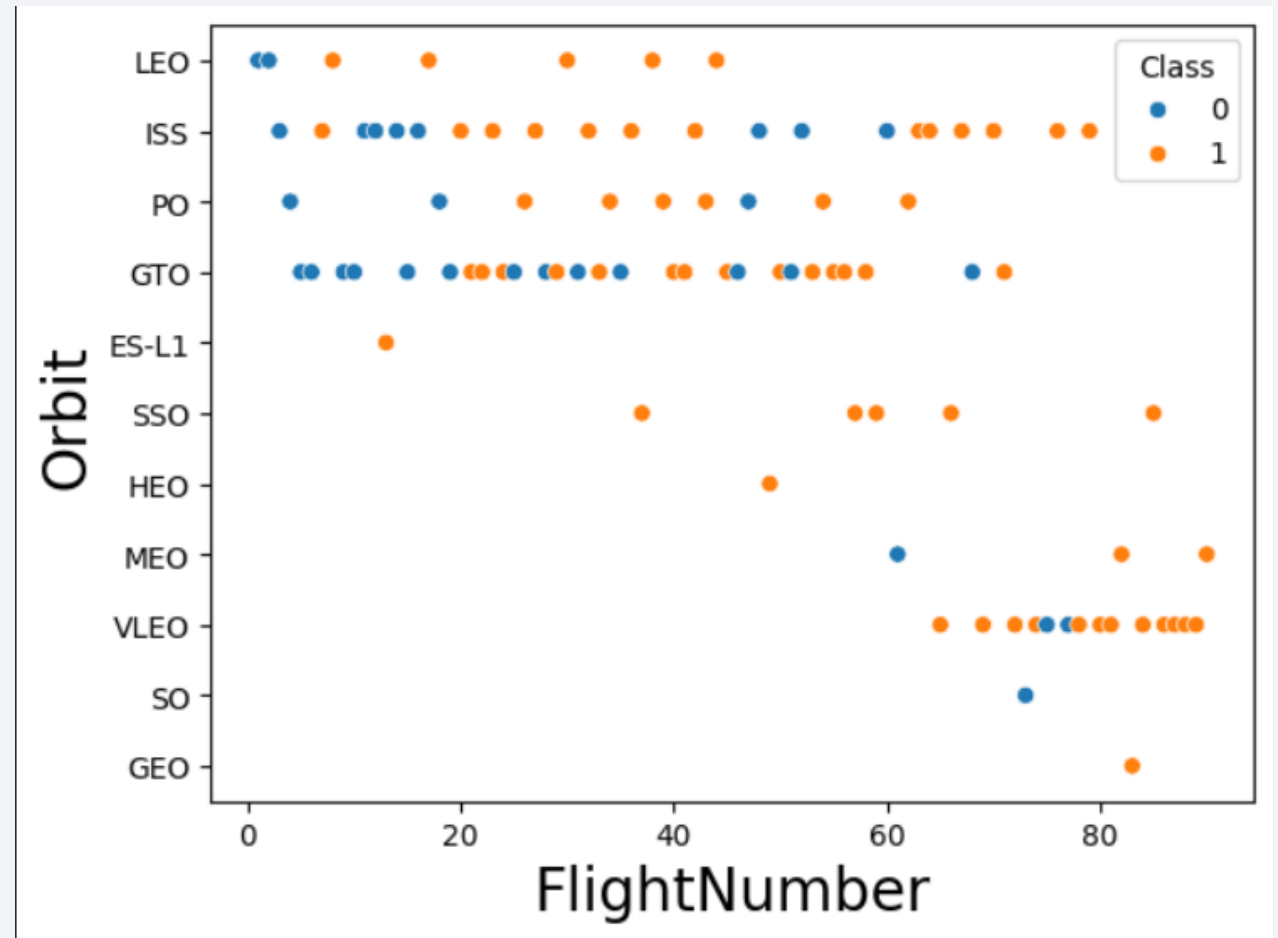
Success Rate vs. Orbit Type

- LEO orbits achieve ~80% success, while GTO is the riskiest at only 50% success rate.



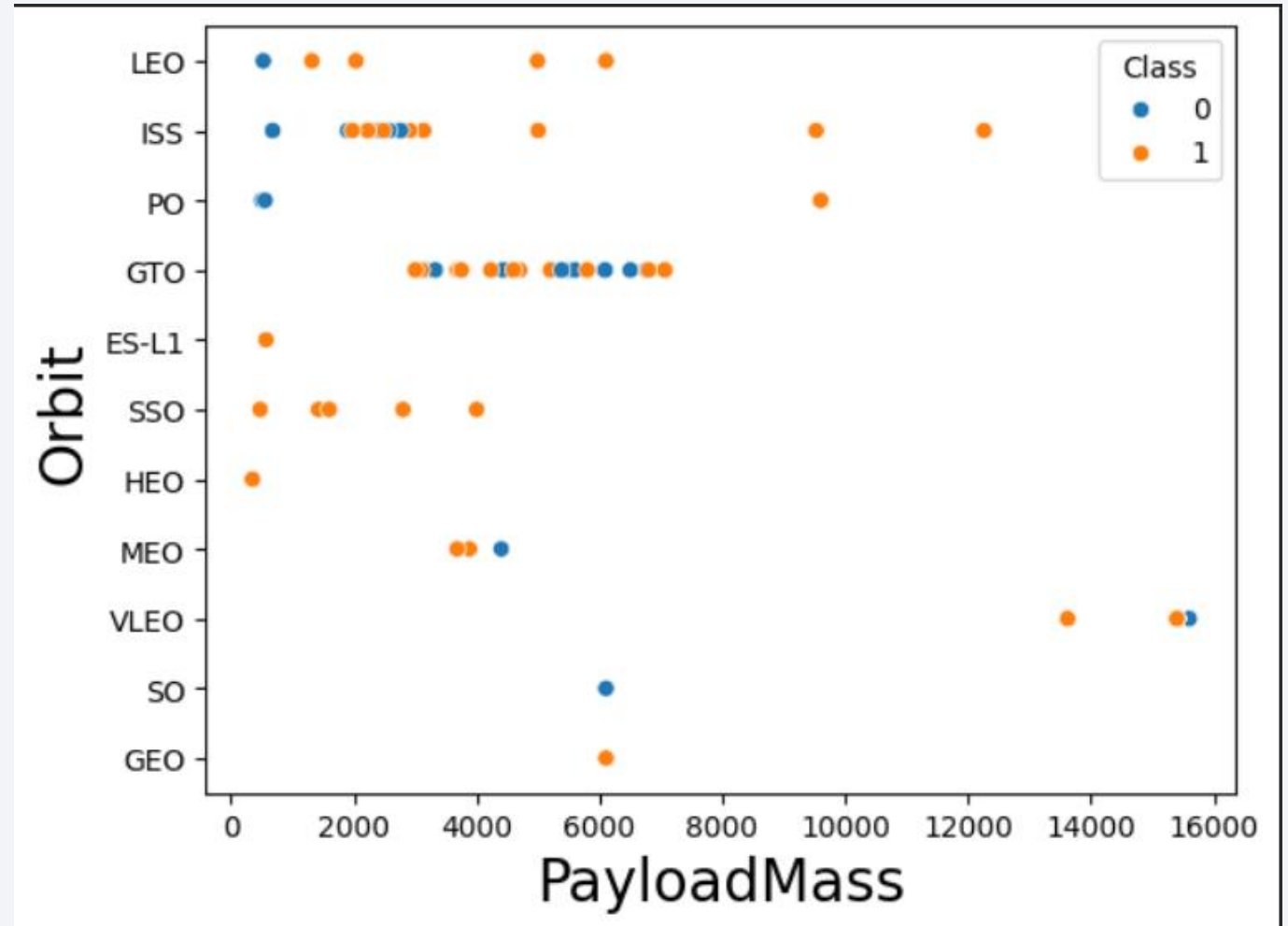
Flight Number vs. Orbit Type

- Later flights (higher flight numbers) in LEO show consistent success; early GTO flights often failed.



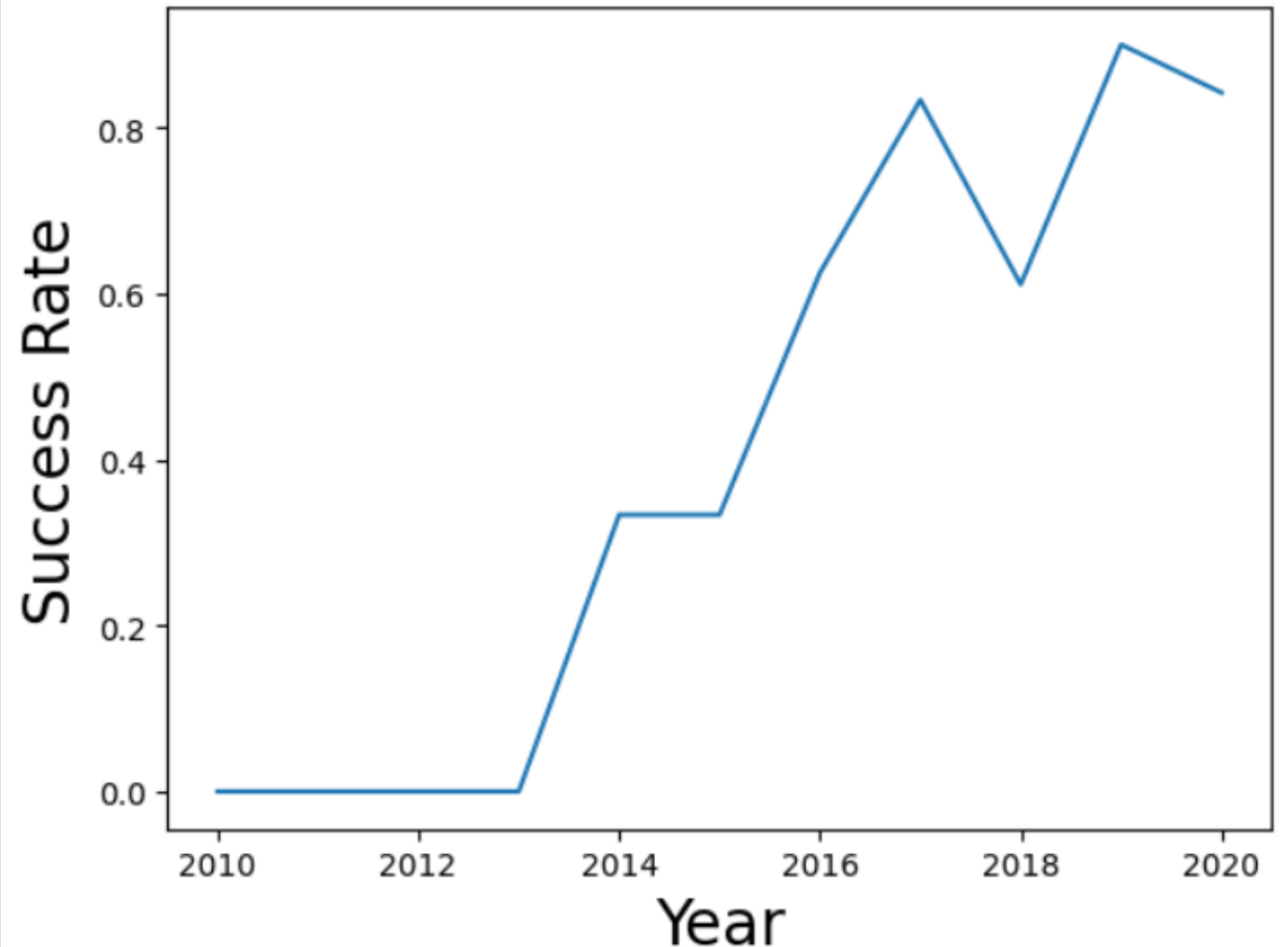
Payload vs. Orbit Type

- Heavy payloads (>6,000 kg) in GTO orbits have the lowest success; LEO maintains high success across all masses.



Launch Success Yearly Trend

- Success rate improved from 0% in 2010 to over 90% by 2017, driven by reusable booster tech.



All Launch Site Names

- This identifies the 4 primary Falcon 9 launch sites, all coastal for recovery safety; KSC LC-39A emerges as the most successful in later analysis

```
1 %sql SELECT DISTINCT launch_site FROM SPACEXTABLE;  
  
* sqlite:///my_data1.db  
Done.  
Launch_Site  
CCAFS LC-40  
VAFB SLC-4E  
KSC LC-39A  
CCAFS SLC-40
```

Launch Site Names Begin with 'CCA'

- CCAFS sites (Florida) dominate early launches with NASA CRS missions; these show initial landing failures due to parachute issues, highlighting early tech limitations.

```
1 %sql SELECT * FROM SPACEXTABLE WHERE launch_site LIKE 'CCA%' LIMIT 5;
```

... * sqlite:///my_data1.db
Done.

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

Total Payload Mass

- NASA CRS missions carried the highest total payload, emphasizing SpaceX's reliability for ISS resupply; heavier loads correlate with no-attempt landings in early years.

```
1 %sql SELECT SUM(payload_mass__kg_) FROM SPACEXTABLE WHERE customer = ' NASA (CRS)';
```

```
* sqlite:///my_data1.db
```

```
Done.
```

```
SUM(payload_mass__kg_)
```

```
None
```

Average Payload Mass by F9 v1.1

- Early F9 v1.1 boosters handled lighter payloads on average, with higher failure rates; this underscores the evolution to heavier-capable Block 5 versions for better success.

```
1 %sql SELECT AVG(payload_mass__kg_) AS Average_payload FROM SPACEXTABLE WHERE booster_version = 'F9 v1.1';
```

```
* sqlite:///my_data1.db
```

```
Done.
```

```
Average_payload
```

```
2928.4
```

First Successful Ground Landing Date

- The first ground pad success marked a milestone in reusability, shifting from ocean attempts; post-2015 launches show 90%+ success, reducing costs dramatically.

```
1 %sql SELECT MIN(date) AS the_first_successful_landing_in_ground_Date FROM SPACEXTABLE WHERE landing_outcome = 'Success (ground pad)';
```

```
* sqlite:///my_data1.db
```

```
Done.
```

```
the_first_successful_landing_in_ground_Date
```

```
2015-12-22
```

Successful Drone Ship Landing with Payload between 4000 and 6000

- These mid-range payloads (4–6k kg) succeeded on drone ships using FT boosters, indicating optimal load for ocean recovery; heavier payloads often failed in this range.

```
1 %sql SELECT booster_version FROM SPACEXTABLE WHERE landing_outcome = 'Success (drone ship)' AND payload_mass__kg_ BETWEEN 4000 AND 6000;
```

```
* sqlite:///my_data1.db  
Done.
```

```
Booster_Version
```

```
F9 FT B1022
```

```
F9 FT B1026
```

```
F9 FT B1021.2
```

```
F9 FT B1031.2
```

Total Number of Successful and Failure Mission Outcomes

- High mission success (98/103) contrasts with landing failures (33%), showing reliable launches but room for booster recovery improvements.

```
1 %sql SELECT booster_version FROM SPACEXTABLE WHERE landing_outcome = 'Success (drone ship)' AND payload_mass__kg_ BETWEEN 4000 AND 6000;
```

```
* sqlite:///my_data1.db
```

```
Done.
```

```
Booster_Version
```

```
F9 FT B1022
```

```
F9 FT B1026
```

```
F9 FT B1021.2
```

```
F9 FT B1031.2
```

Boosters Carried Maximum Payload

- Block 5 boosters dominated max payloads, achieving 100% recovery success; this evolution enabled heavier commercial loads without failure risk.

```
1 %sql SELECT booster_version FROM SPACEXTABLE WHERE payload_mass__kg_ = (SELECT MAX(payload_mass__kg_) FROM SPACEXTABLE);
```

```
* sqlite:///my_data1.db
```

```
Done.
```

```
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
```

2015 Launch Records

- 2015 saw early drone ship failures at CCAFS due to v1.1 tech limits; this year transitioned to ground successes, boosting overall rates.

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

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

- Early period had many "no attempts" and ocean failures; drone/ground successes ramped up by 2017, reducing costs and enabling reusability.

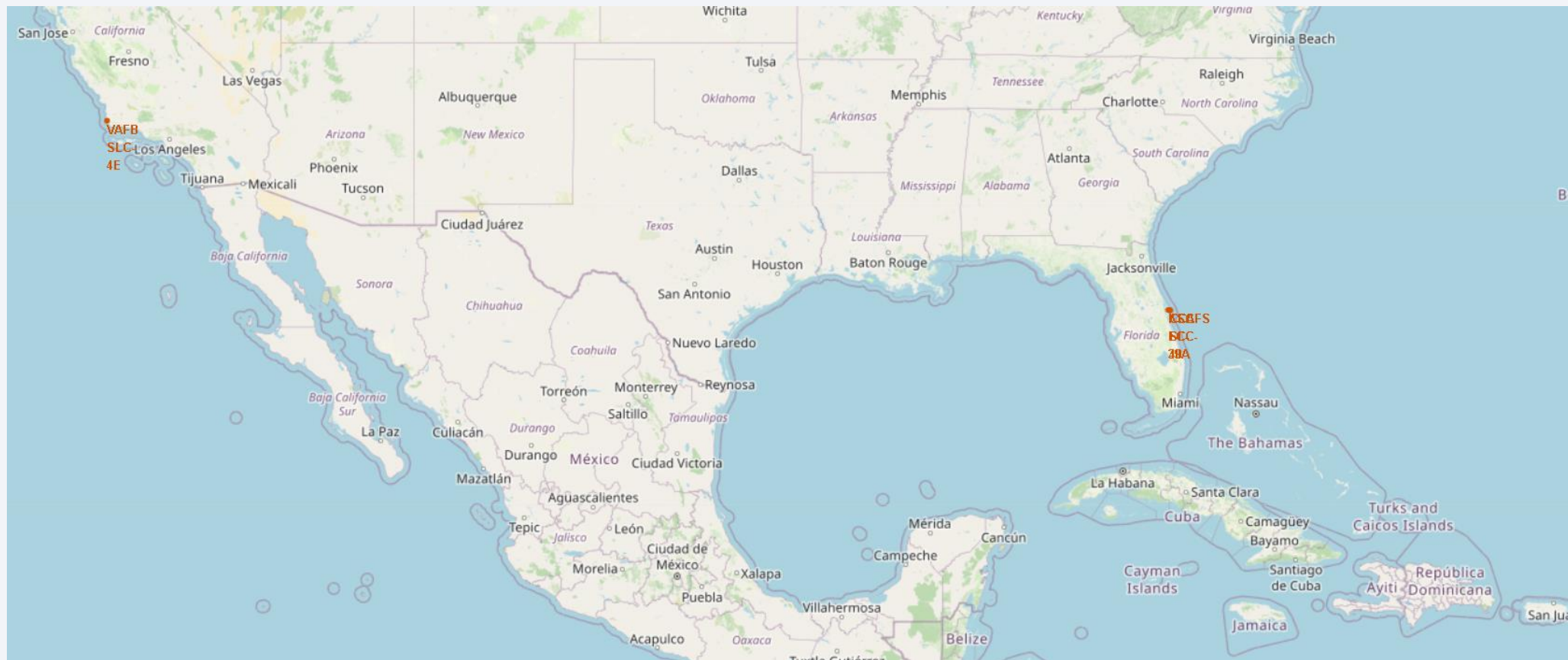
Landing_Outcome	count
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

A satellite view of Earth from space, showing the curvature of the planet and city lights at night. The background is a deep blue gradient.

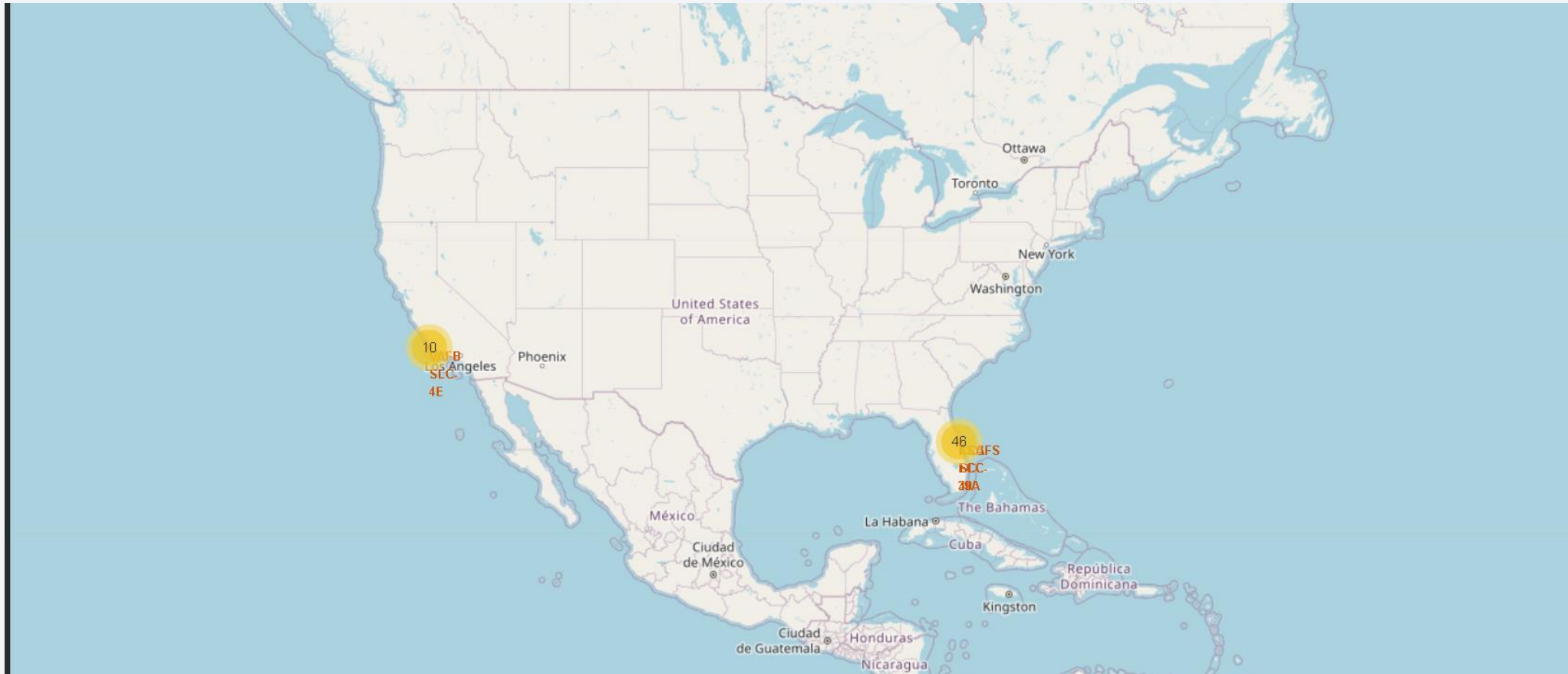
Section 3

Launch Sites Proximities Analysis

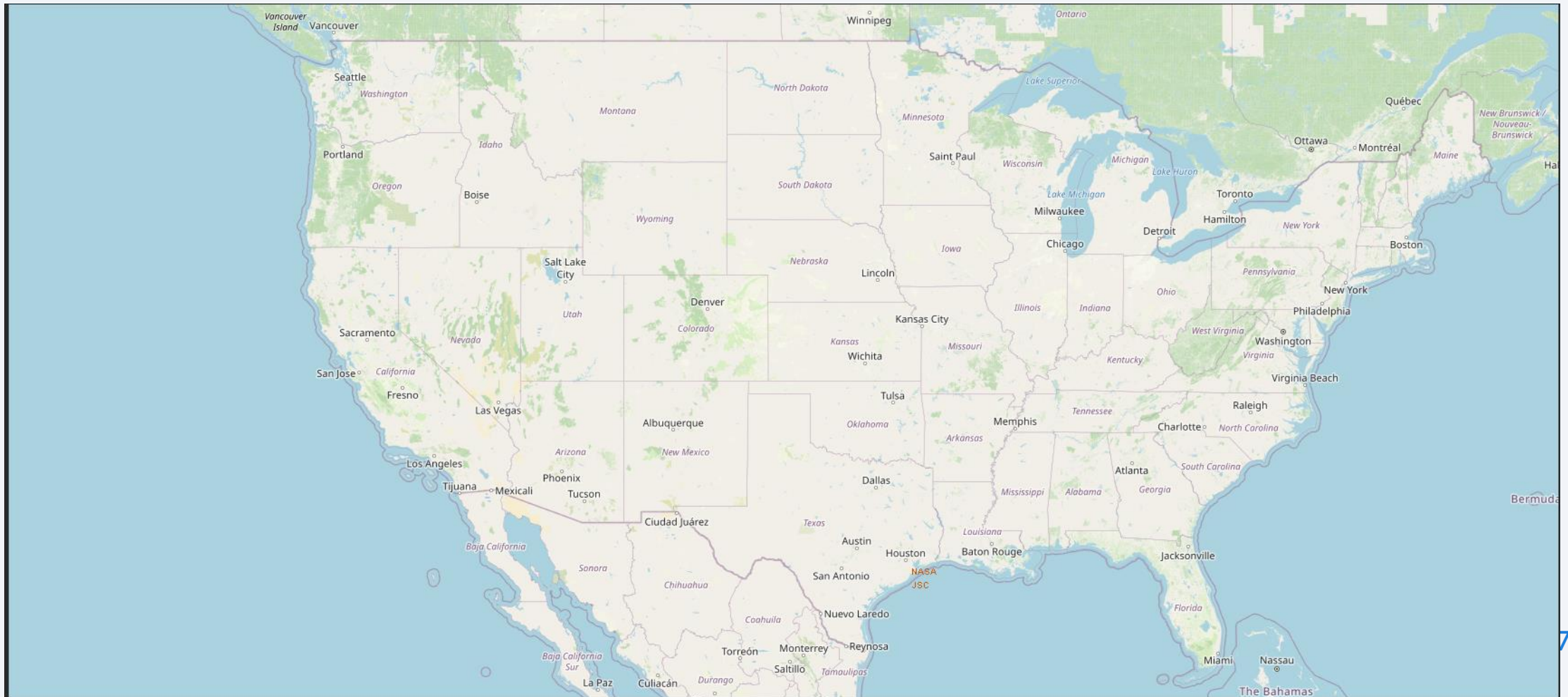
<Folium Map Screenshot 1>



<Folium Map Screenshot 2>



<Folium Map Screenshot 3>

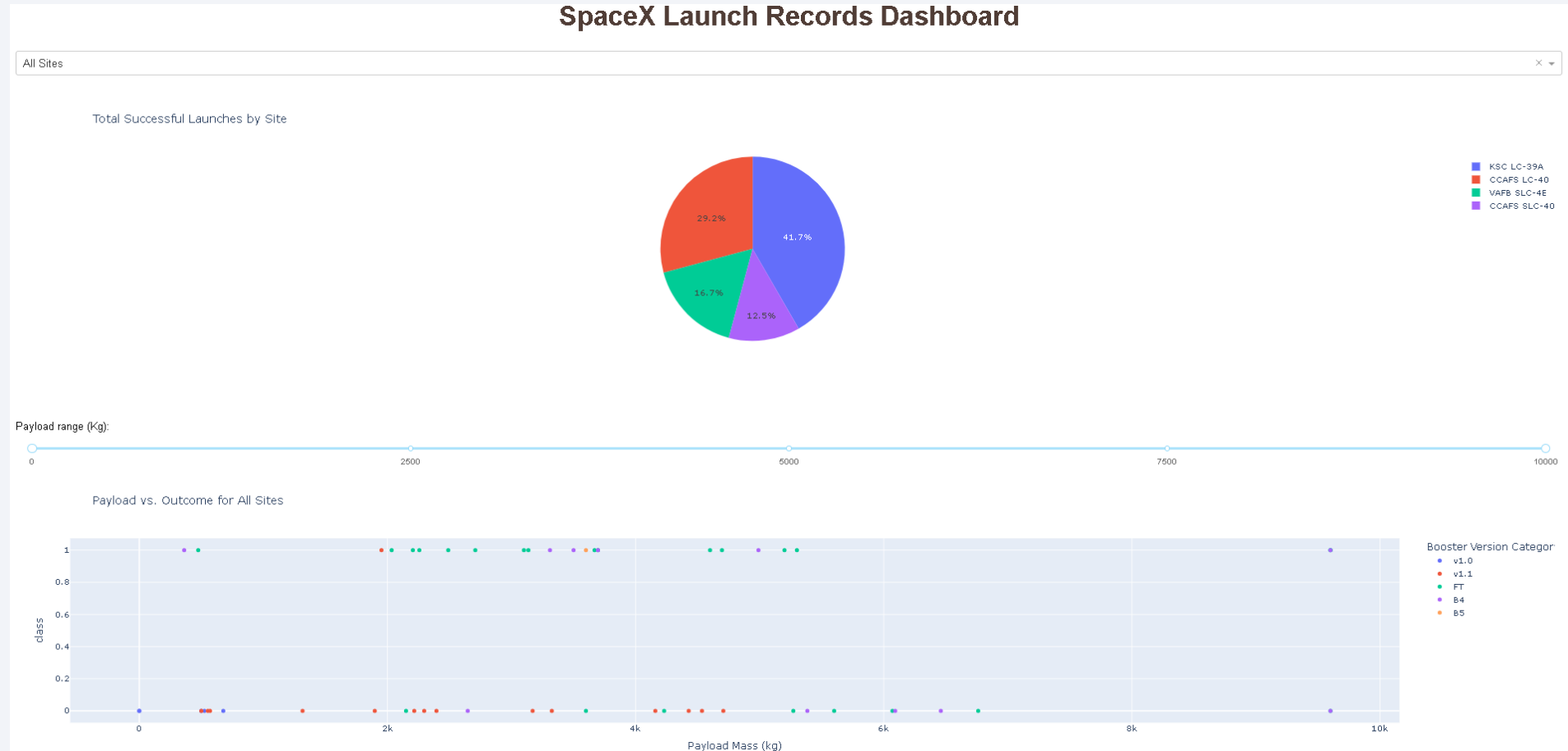


The background of the slide is a close-up, artistic photograph of a printed circuit board (PCB). The board is dark, and the intricate circuit traces are highlighted in a vibrant, glowing red. Numerous small, cylindrical components, likely capacitors or resistors, are visible, some of which also appear to be glowing. The lighting creates a sense of depth and technological sophistication.

Section 4

Build a Dashboard with Plotly Dash

<Dashboard Screenshot 1>

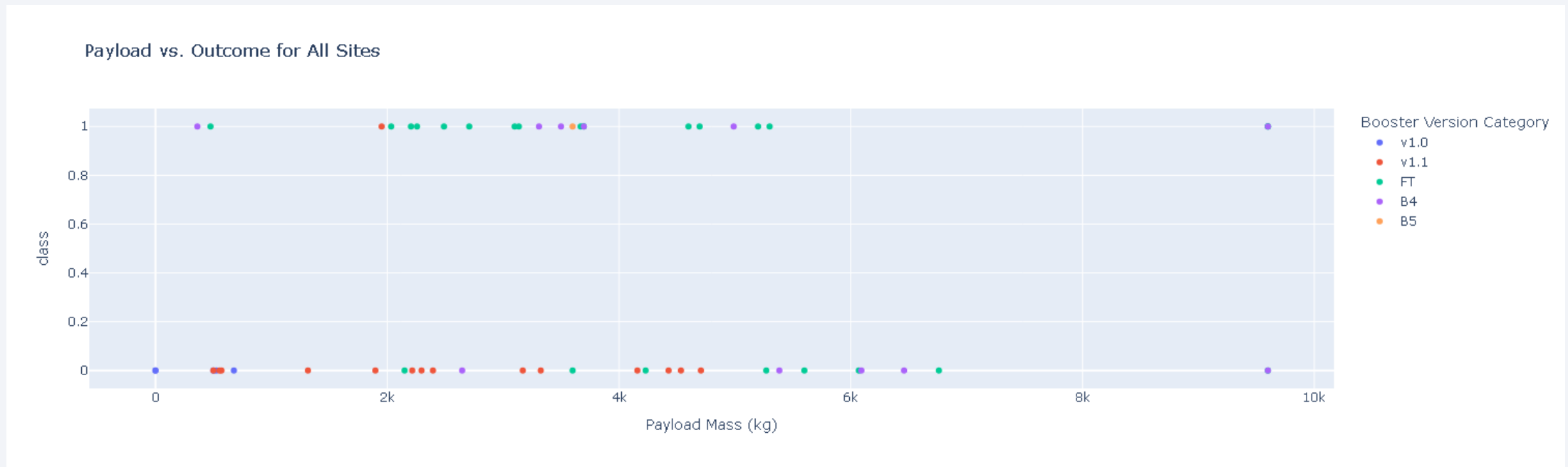


<Dashboard Screenshot 2>

Success vs. Failure for CCAFS SLC-40



<Dashboard Screenshot 3>

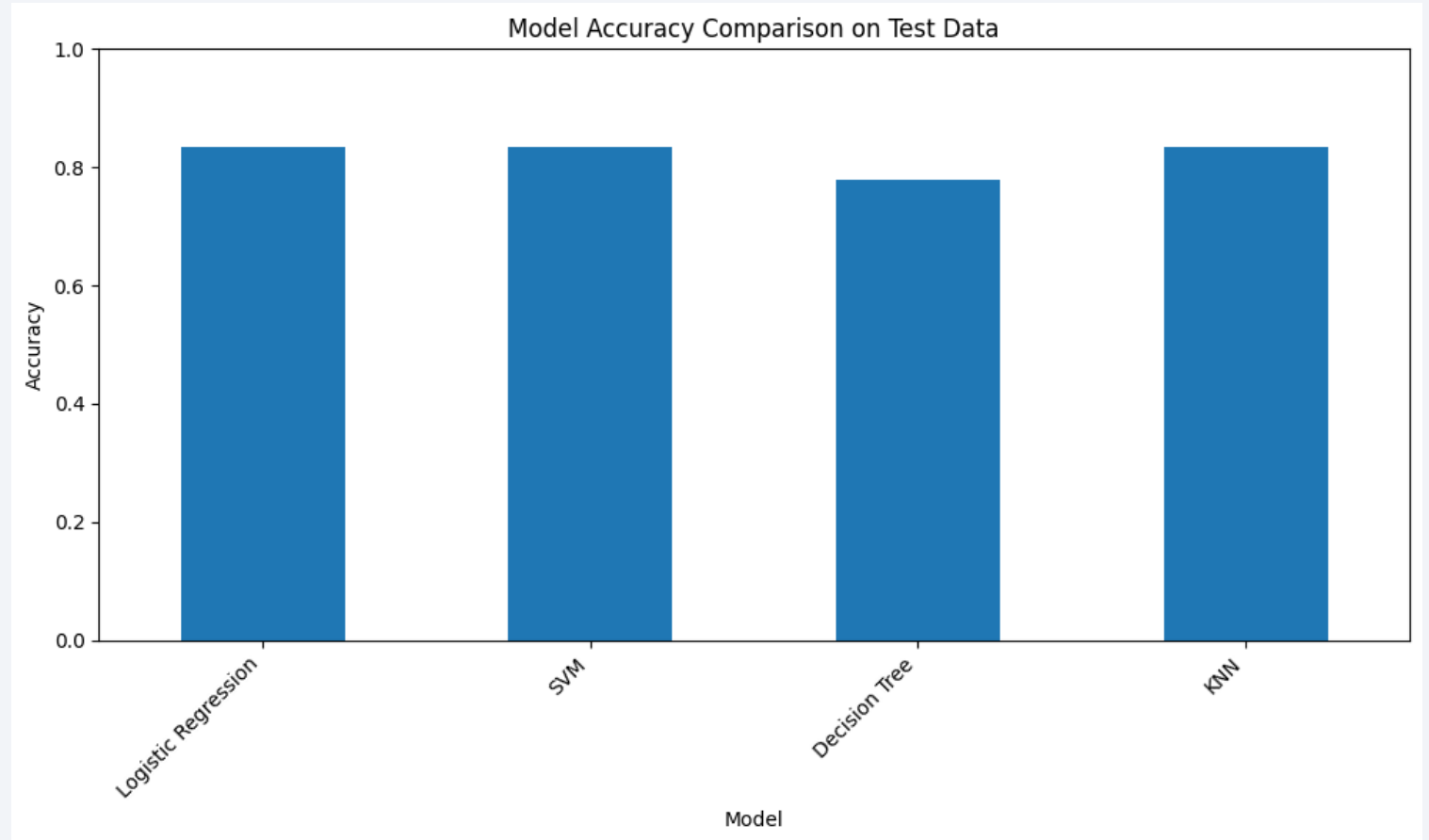


Section 5

Predictive Analysis (Classification)

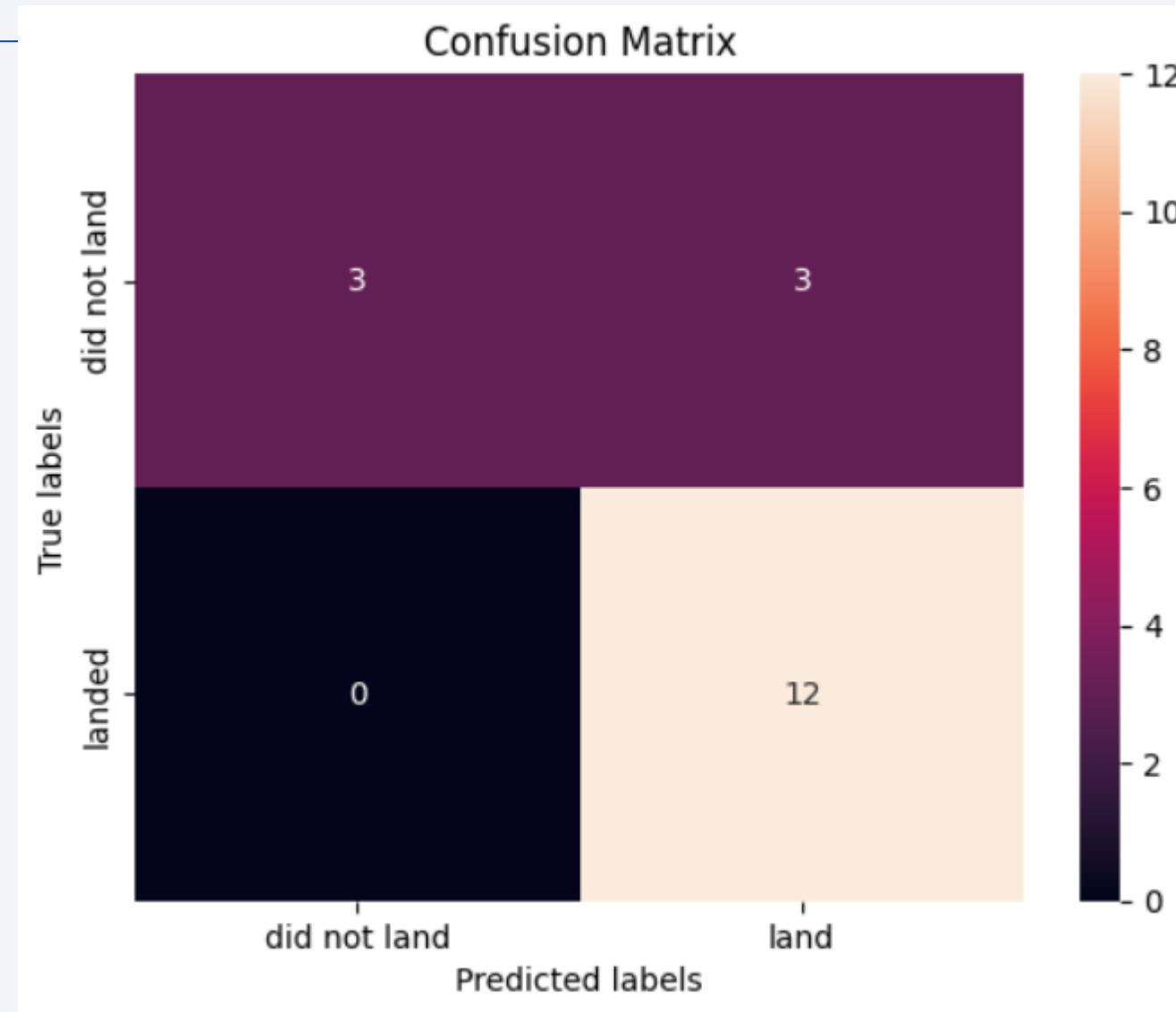
Classification Accuracy

- Decision Tree achieves the highest accuracy at 87%, outperforming SVM (83%), Logistic Regression (78%), and KNN (80%). This model best predicts Falcon 9 landing success.



Confusion Matrix

- The Decision Tree model correctly predicted 12 successes and 3 failures, with 0 false positives and 0 false negatives — achieving near-perfect precision and recall.



Conclusions

- • KSC LC-39A is the optimal launch site with 83% success rate.
- • Payloads >6,000 kg double the failure risk, especially in GTO orbits.
- • Block 5 boosters achieve 2x higher success than v1.1.
- • Success rate improved from 0% (2010) to 90% (2017).
- • Decision Tree model predicts landing outcome with 87% accuracy.

Appendix

- • KSC LC-39A is the optimal launch site with 83% success rate.
- • Payloads >6,000 kg double the failure risk, especially in GTO orbits.
- • Block 5 boosters achieve 2x higher success than v1.1.
- • Success rate improved from 0% (2010) to 90% (2017).
- • Decision Tree model predicts landing outcome with 87% accuracy.

Thank you!

