IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Sergio R.

August 29, 2022

Outline

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

Executive Summary

CRISP Methodology:

1. Determine cost of each launch & whether SpaceX will reuse their first stage.

2. Collect Space X data via Wikipedia, Spacex.com

3. Format Data

4. Analyze for insights

5. Predict whether first stage will be reused.

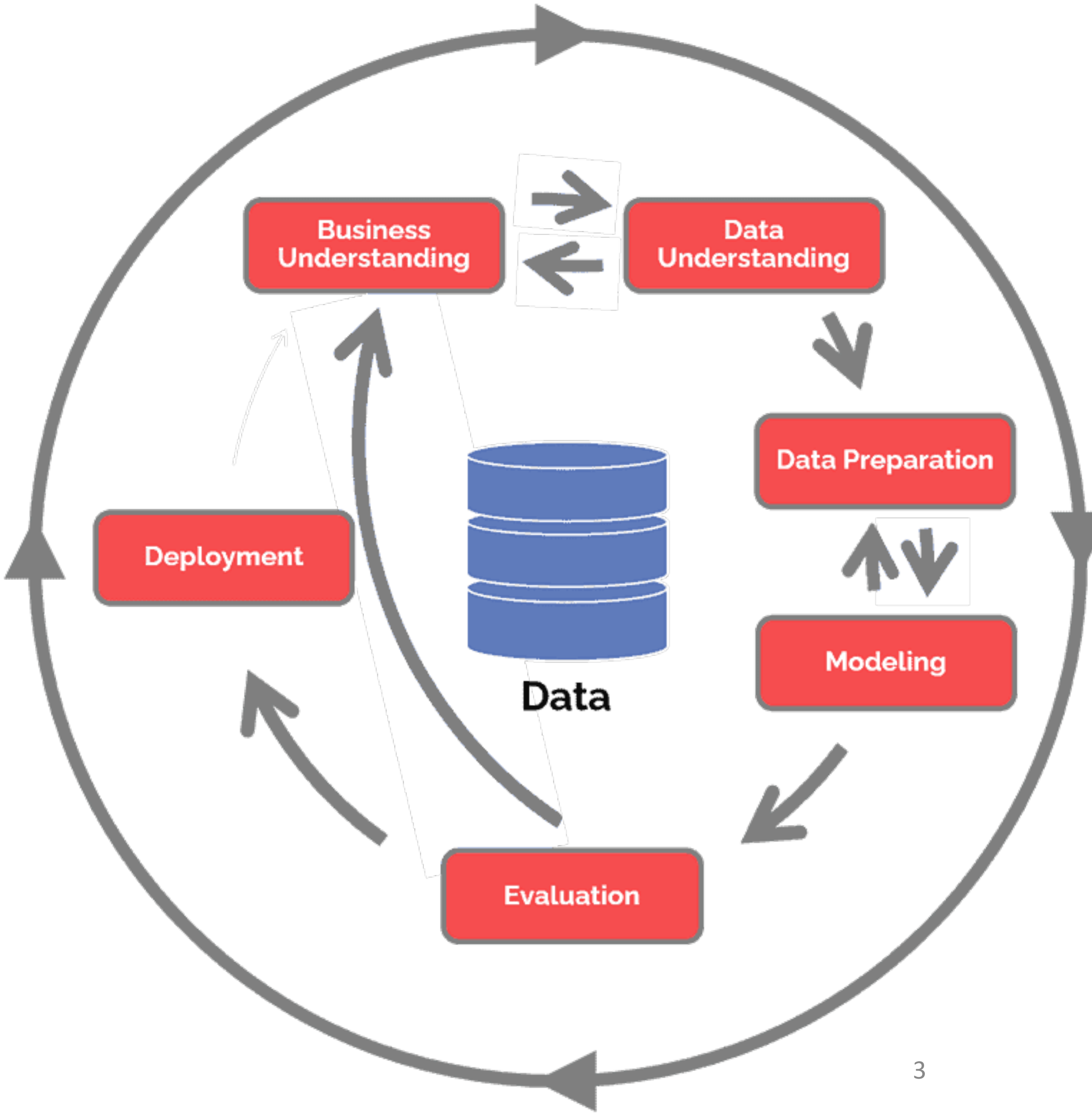
Summary of all results:

- Cost of launch decreases every reuse.

- SpaceX will reuse their first stage; perhaps, striving to reuse each stage.

- Over 96% of **Successful Landings** utilized Grid-Fins & Legs.

[CRISM-DM Diagram: Nick Hotz]



Introduction

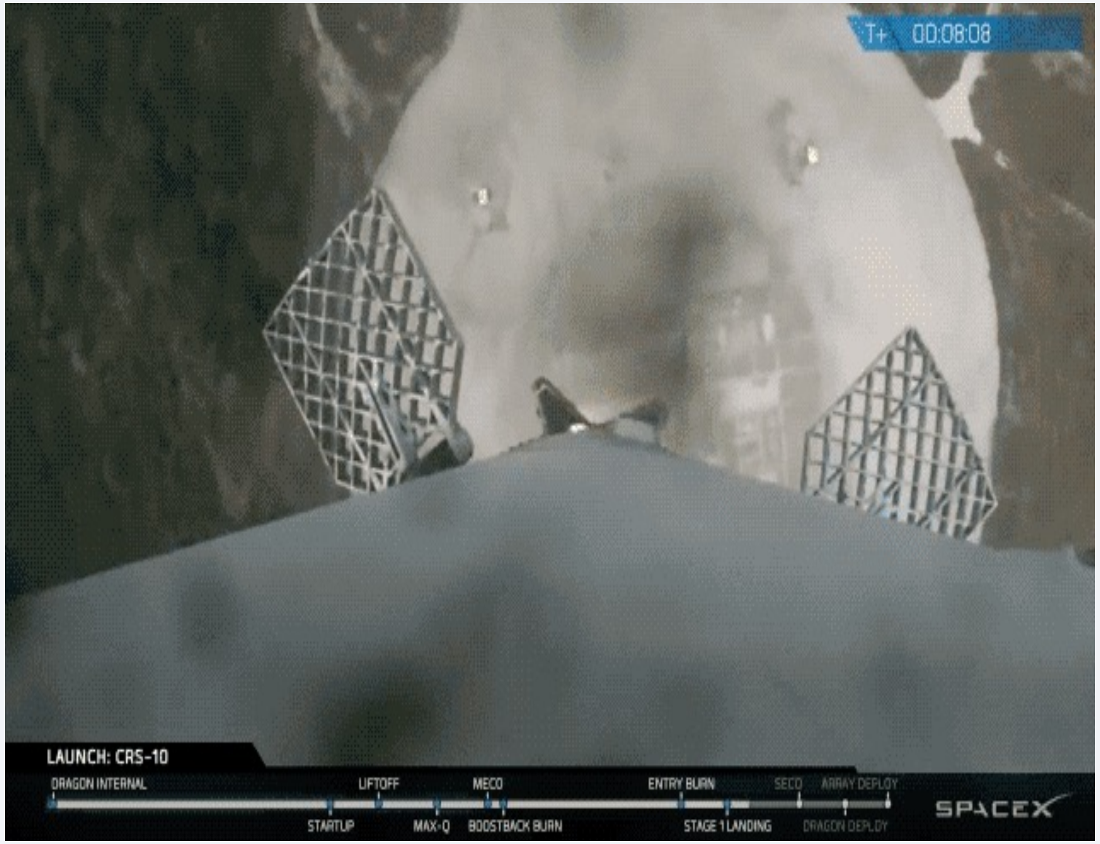
- Goal:

  - Company Space Y plans to compete in the Race to Space, our goal is to *minimize* the **Cost of Each Launch**

- Problems to Answer:

  - Why do SpaceX competitors spend $165 Million per Launch, relative to SpaceX's $62 million cost per Launch?

  - What variables contribute to a higher Success Rate of Landing, and why?



Section 1

Methodology

Methodology

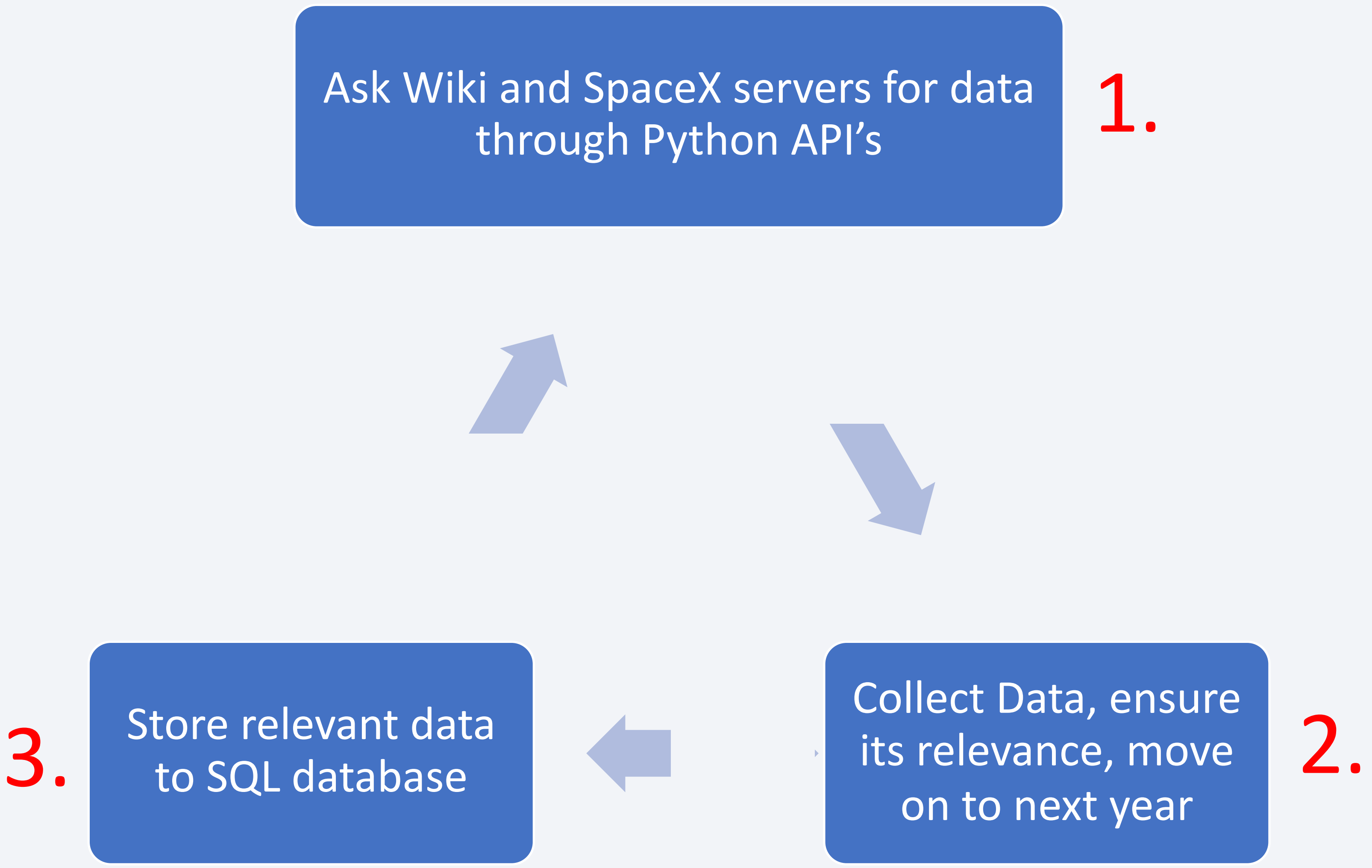
Executive Summary

- Data collection methodology:
  - SpaceX data was collected via Wikipedia & SpaceX website; data was collected with tools such as, Python and web-scraping HTML.
- Perform data wrangling
  - Collected data started off unstructured; we organized, cleaned, and visualized the collected data to find data-driven insights that lead to business-decisions.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - We build classification models using different variables and use evaluation methods to validate the out-of-sample accuracy of each Data Model.

Data Collection

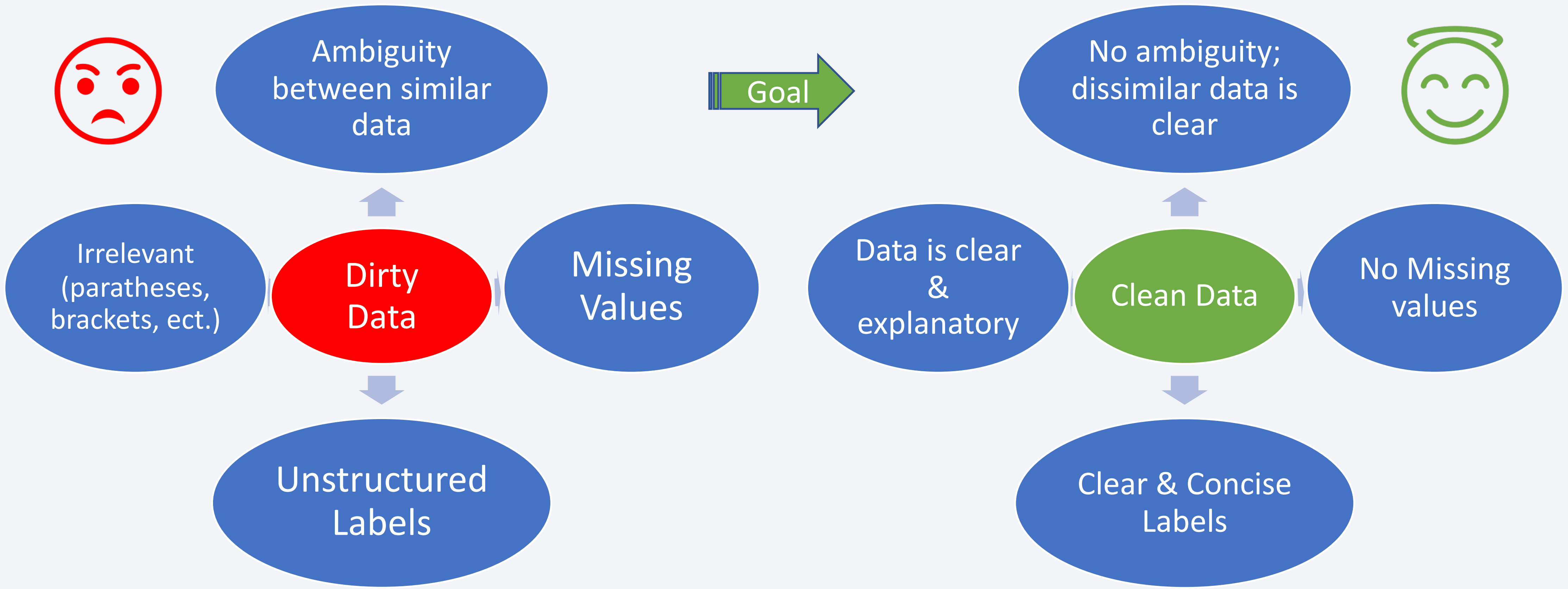
Proof of work:

- Data collection from SpaceX website
  - [SpaceX API]
- Data Collection from Wikipedia
  - [Wiki Collection]



Data Wrangling

- Proof of Work:
  - [Data Wrangling Notebook]



EDA with Data Visualization (1)

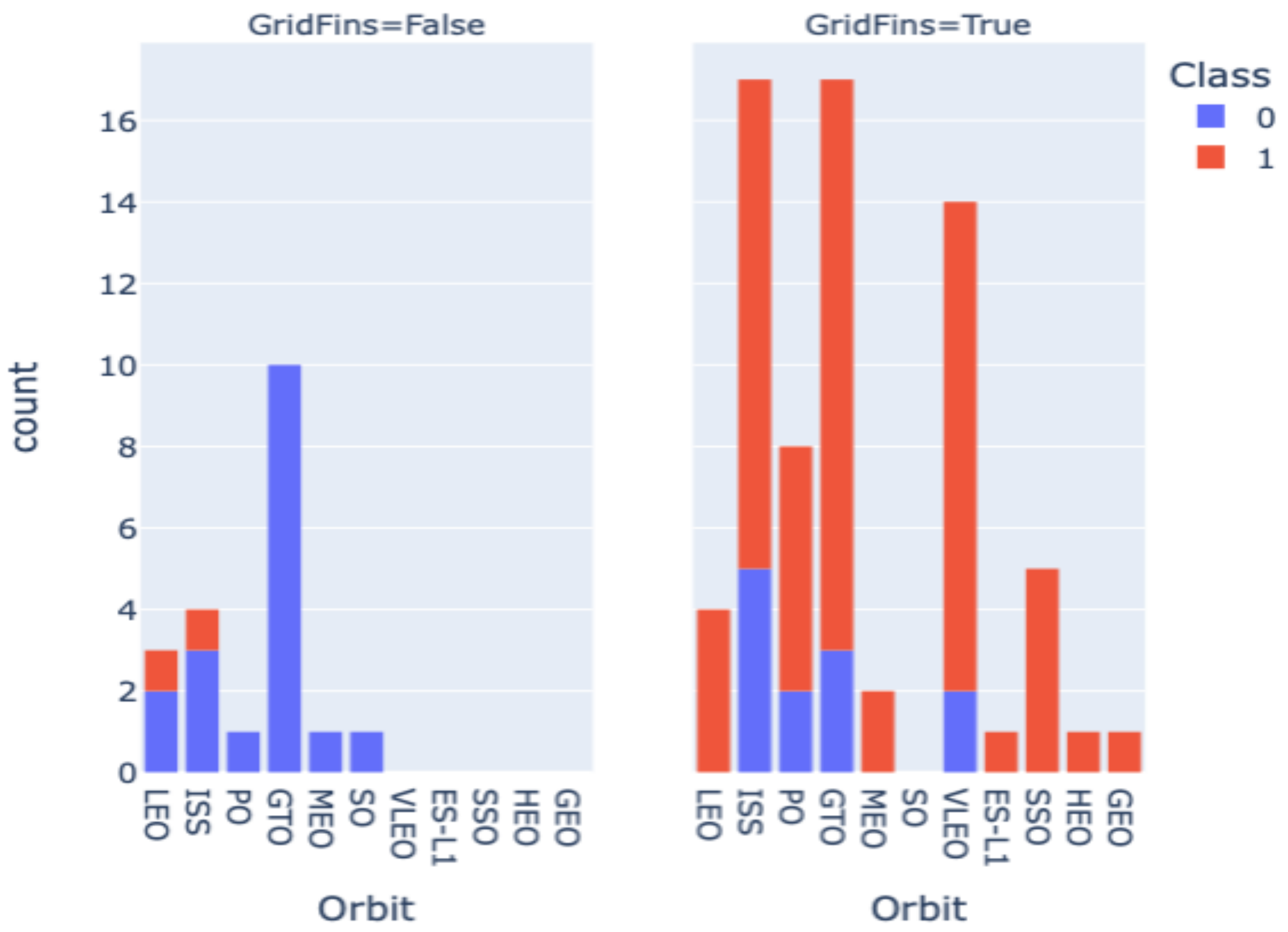
• **Goal**: DECREASE COST OF LAUNCH

- Probability of a **successful landing increases** when utilizing **Grid Fins**, allowing for diverse Orbital Missions.

- A successful landing decreases cost per launch, due to decreased **CAP EX, i.e. First Stage Booster cost**

• Proof of EDA

Relationship Analysis:
Orbit & Success Rate vs Grid Fins



EDA with Data Visualization (2)

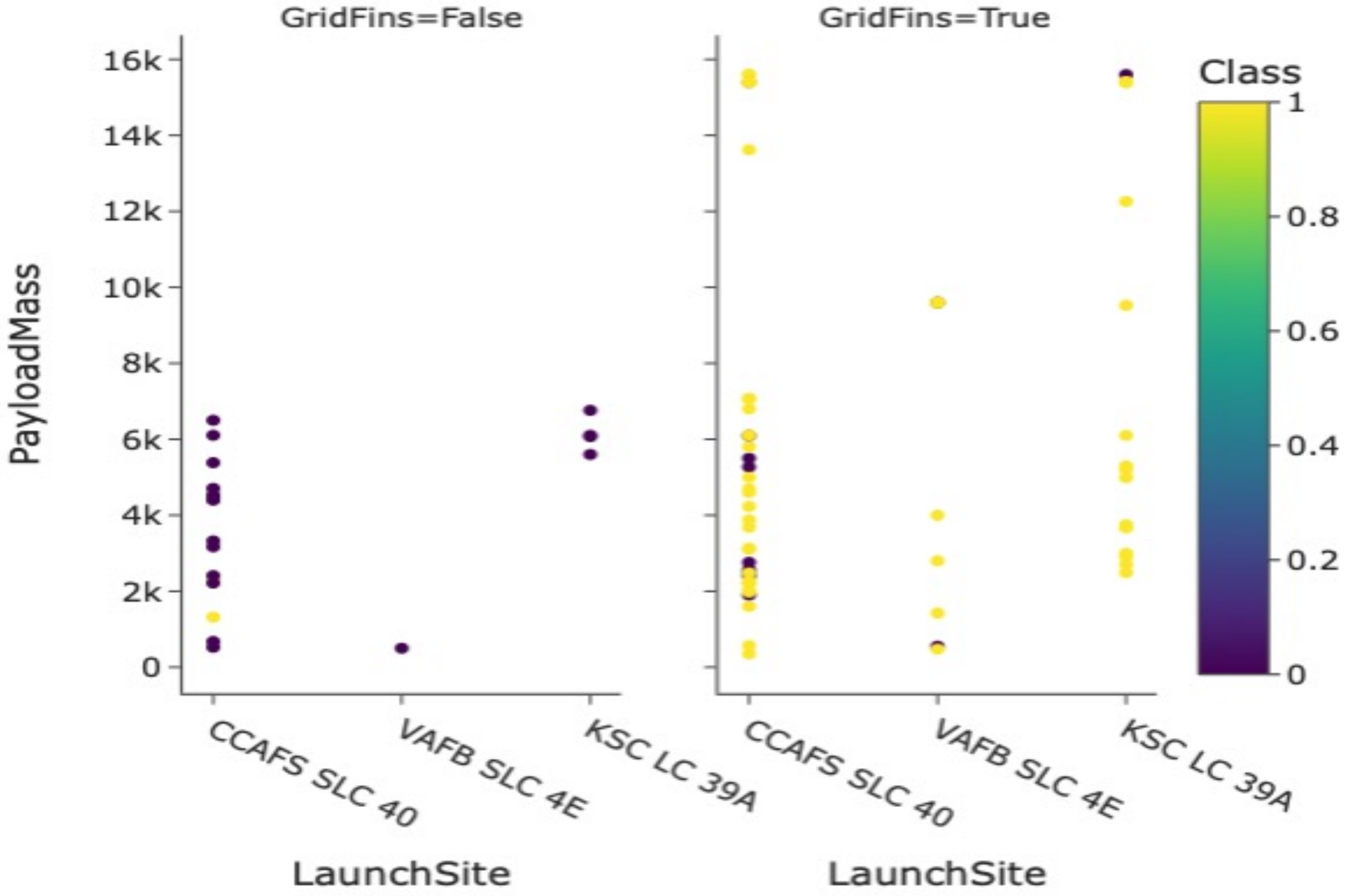
- **Goal**: Find Optimal Launch Site per Payload Mass

     - Vandenburg has fewer high payload flight attempts over 4,000 kg

     - Needs more research before coming to any conclusions

- [Proof of EDA]()

Relationship Analysis:
Payload Mass & Success Rate vs Grid Fins
(per Flight)



EDA with SQL

- **Historical Data Variance by Date:**

  - From 2010 to December of 2020

- **Minimum Payload Mass:**

  - 362 (kg)

- **Total payload mass by Booster Version:**

  - 619,967 (kg)

Build a Dashboard with Plotly Dash

- **Pie charts statistics to showcase the success rate of each launch landing**

  - Kennedy Space Launch Complex (KSC) had a 76.9% success rate, the most successful among other launch sites.

- **Scatterplot to find insights into whether variables Payload Mass and Success per Site had signs of correlation.**

  - Higher success rate per Payload Mass between **2,000 & 5,300 kg.**

  - Pearson Correlation (~ 0.2 Points), data size may be insufficient.

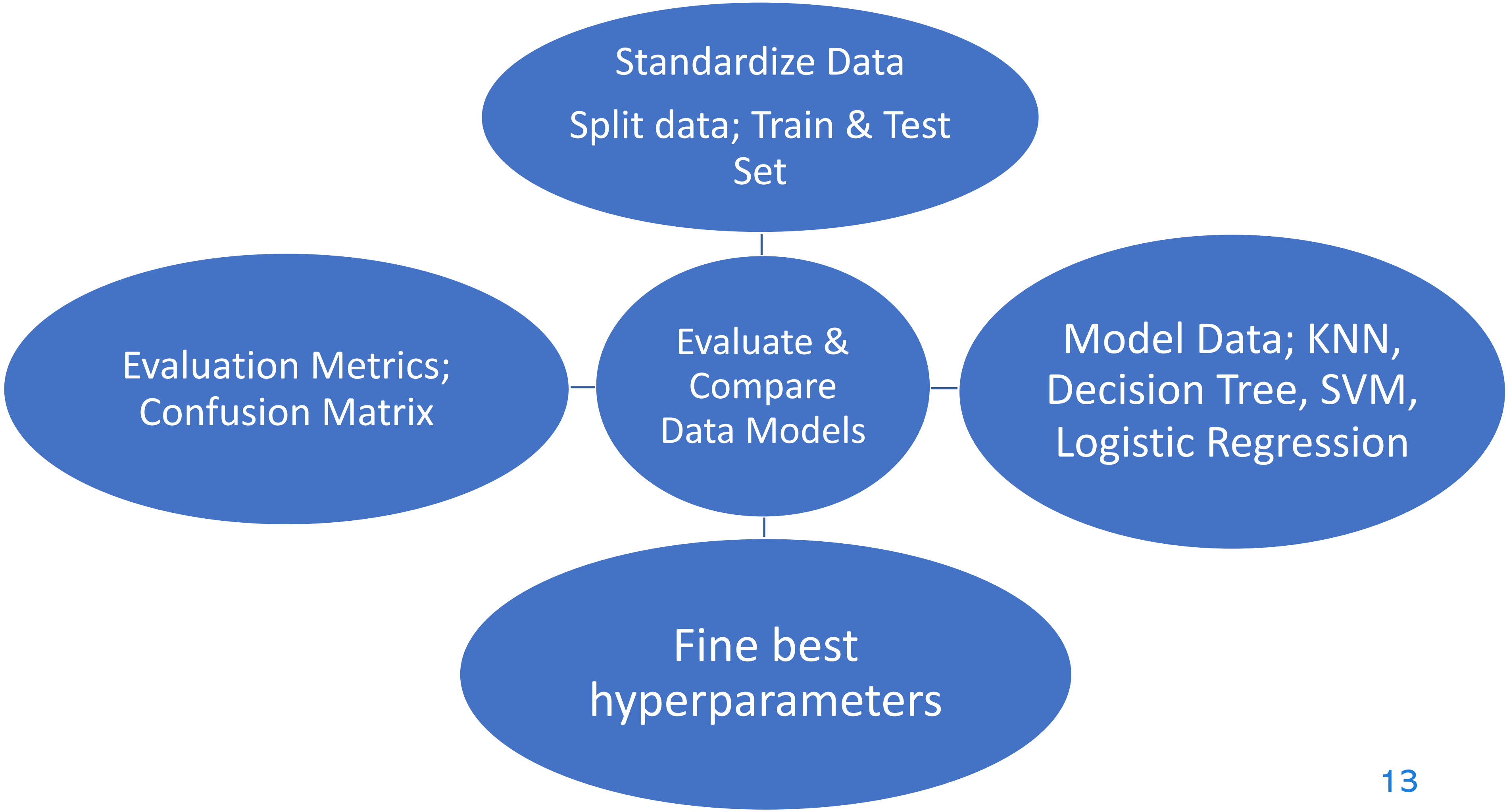
- Dashboard

Predictive Analysis (Classification)

- After tuning hyperparameters, every classification model had an **out-of-sample accuracy of 83%**

- Evaluation Metric: **Confusion Matrix**

- [Data Model Notebook]



Results (1)

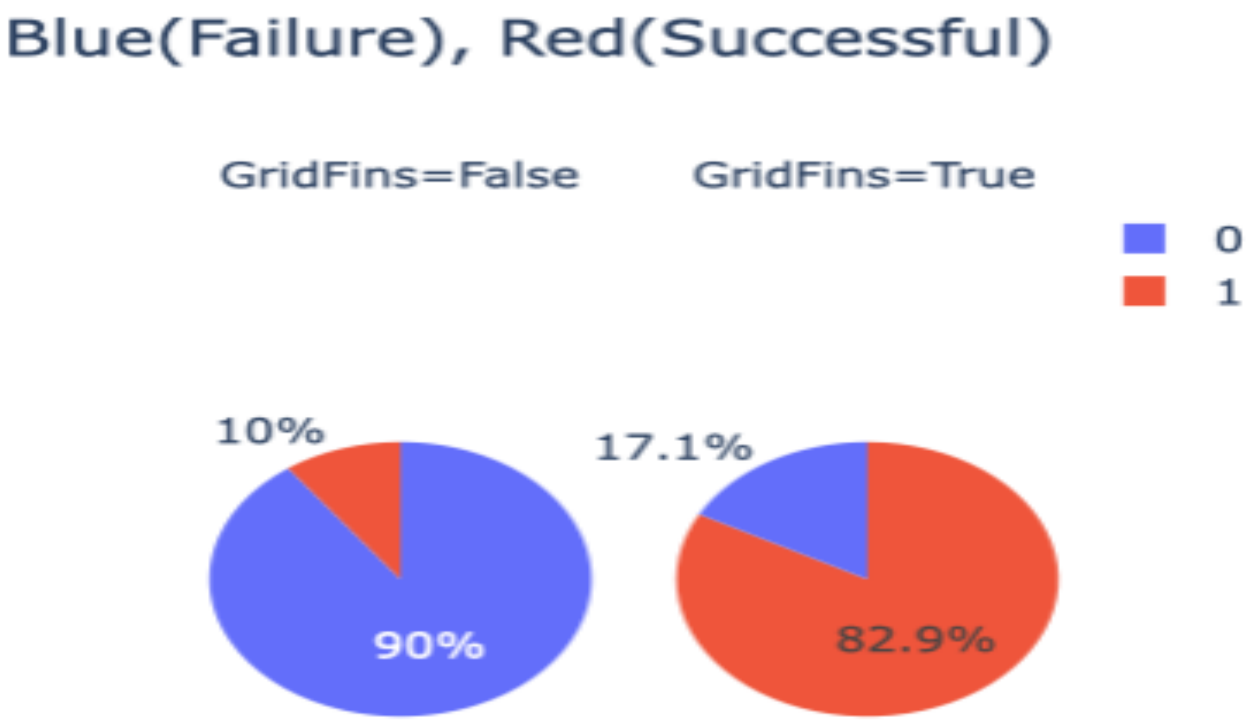
- **EDA: Are there any variables that INCREASE the probability of a SUCCESSFUL LANDING?**

  - Grid Fins and Legs **INCREASE** probability of a **SUCCESSFUL** landings by over 70%

- **Interactive analytics demo: Screenshots (next slide)**

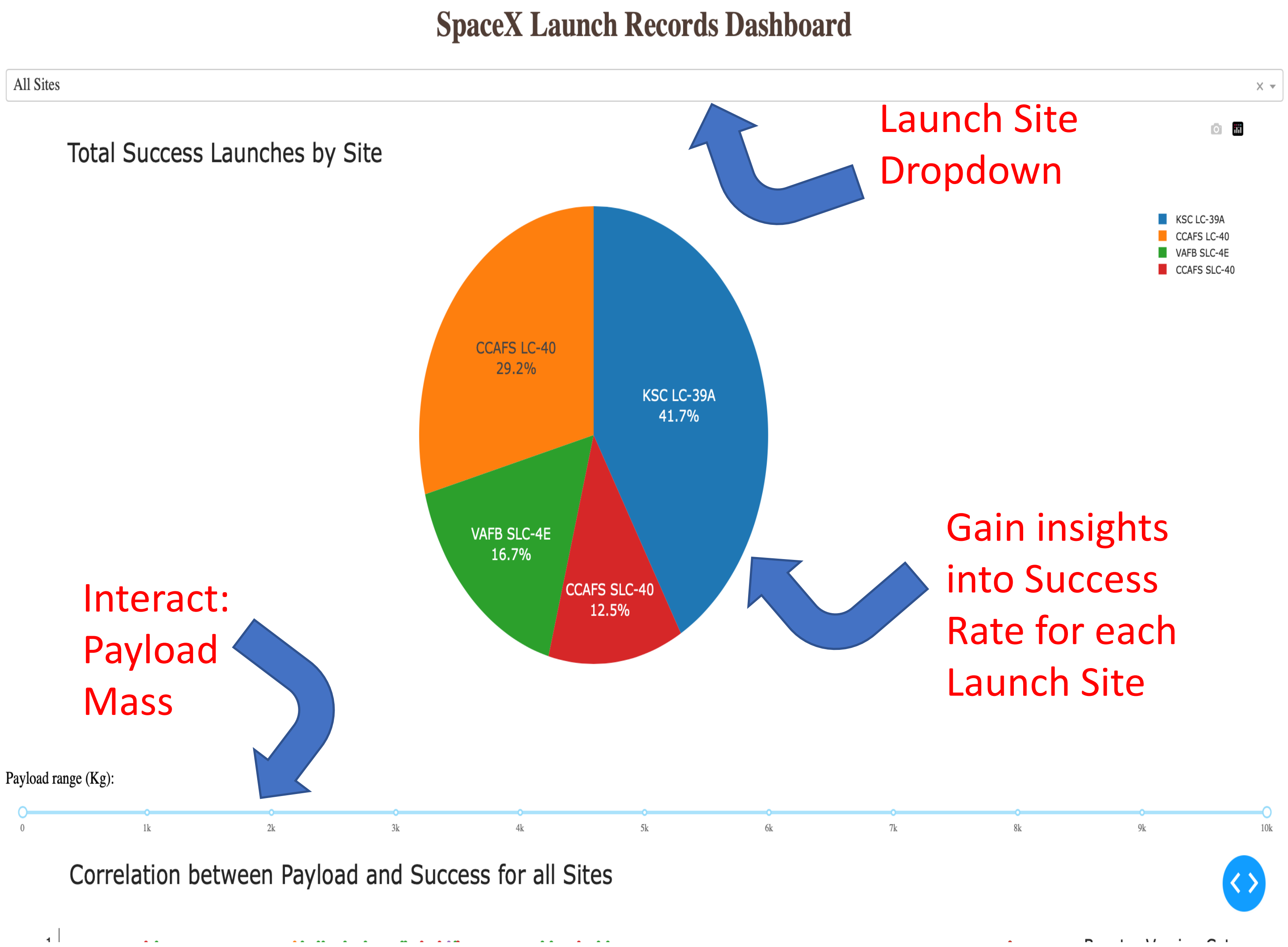
- **Predictive analysis results:**

  - **83% out-of-sample accuracy** (all data models)



Results (2)

- **Interactive analytics demo:**



Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

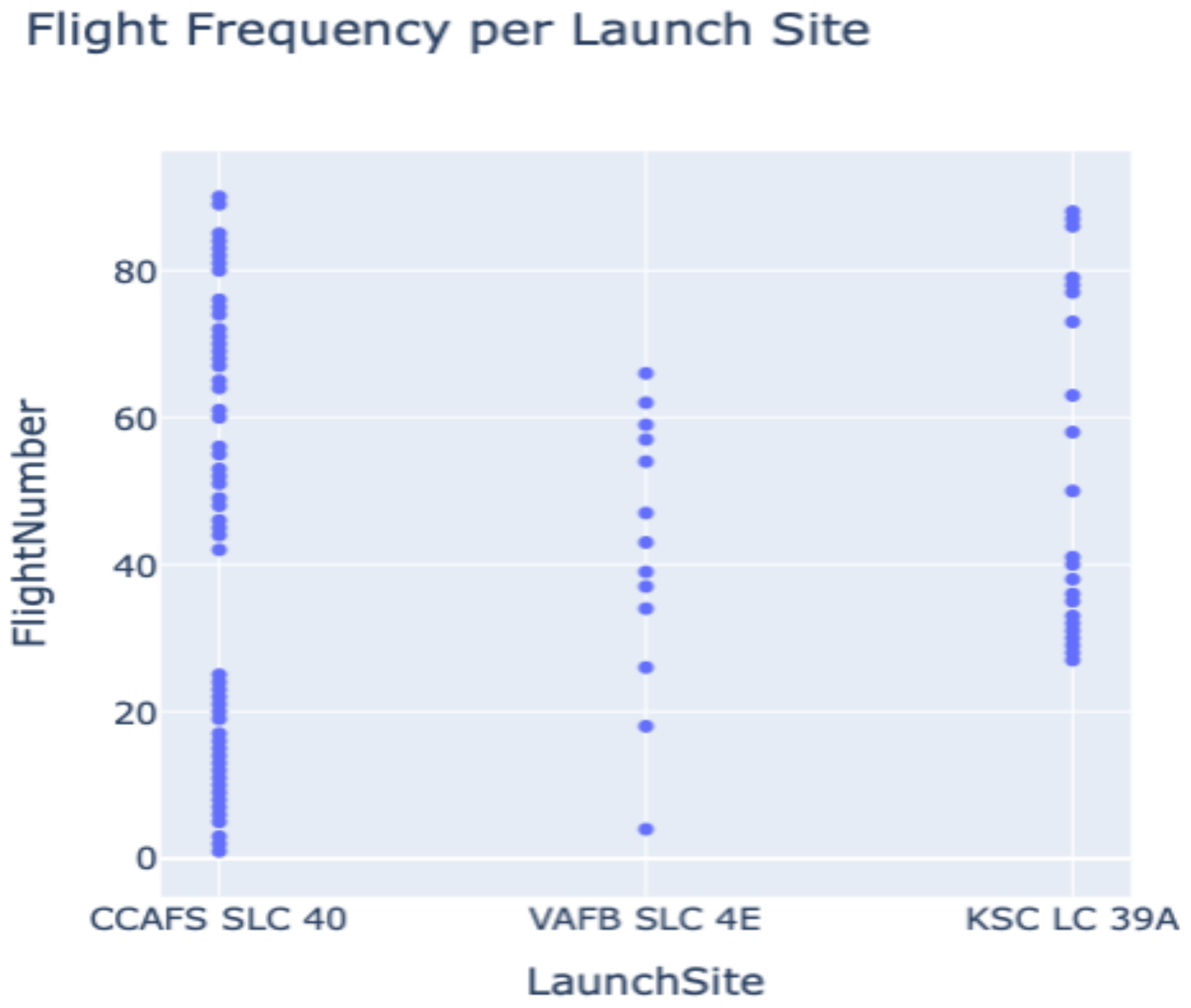
- 90 Flight Attempts (2010-2020)

- KSC accounted for ~25% of **Total Flights**

  - 77% Success Rate

- CCAFS accounted for ~61% of **Total Flights**

  - 60% Success Rate



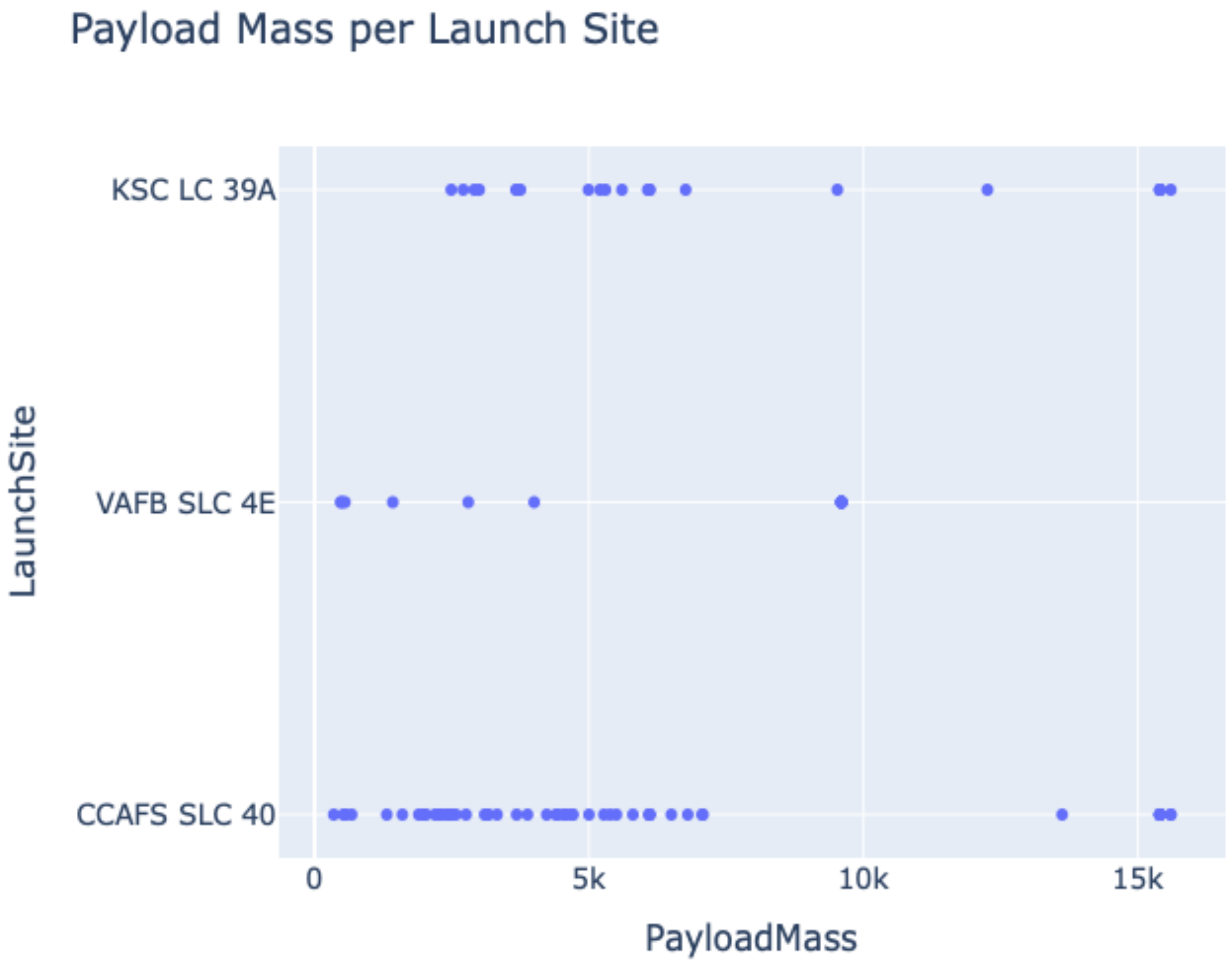
Payload vs. Launch Site

- CCAFS & KSC:

    - Majority of flights between 2500-6500 kg

- VAFB:

    - Majority of flights under 5000 kg



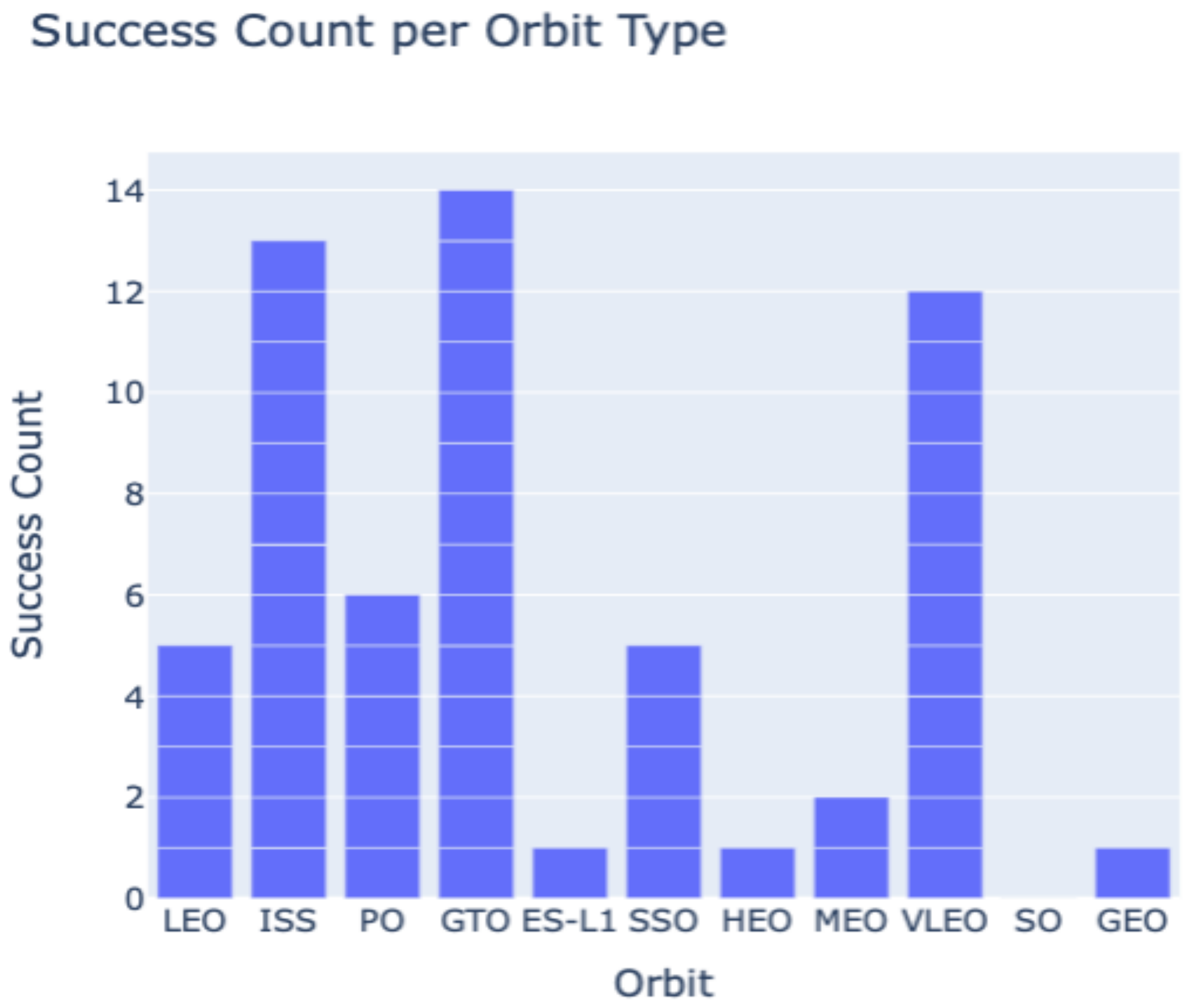
Success Rate vs. Orbit Type

• ISS, GTO, VLEO:

- Accounted for - **2/3, or ~66% of all successful launches**

• ES-L1, HEO, MEO, SO, GEO:

- Success rate **drops** to nearly **ZERO**

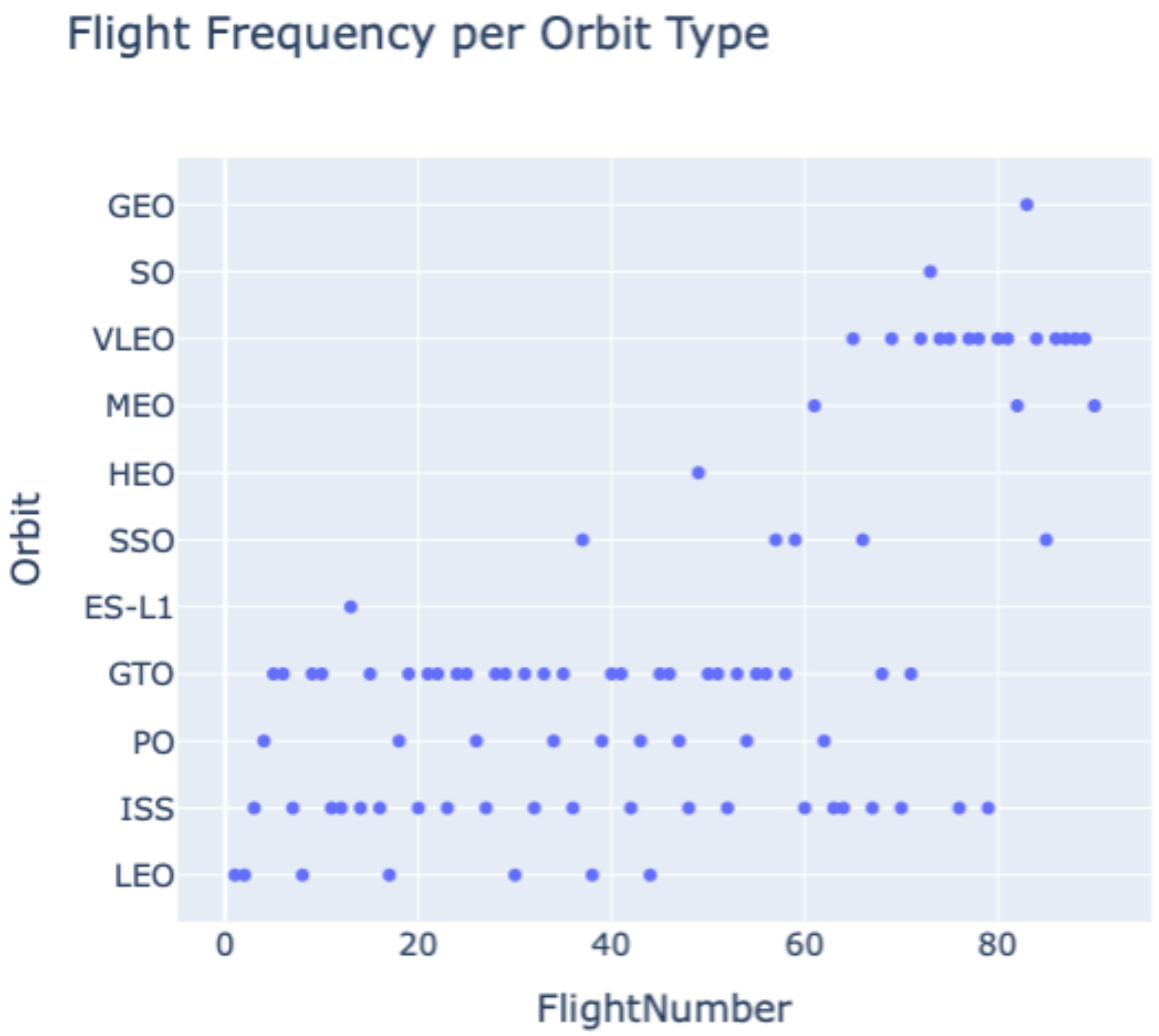


Flight Number vs. Orbit Type

• **Majority** of Flights per Orbit:

- ISS, LEO, GTO, VLEO

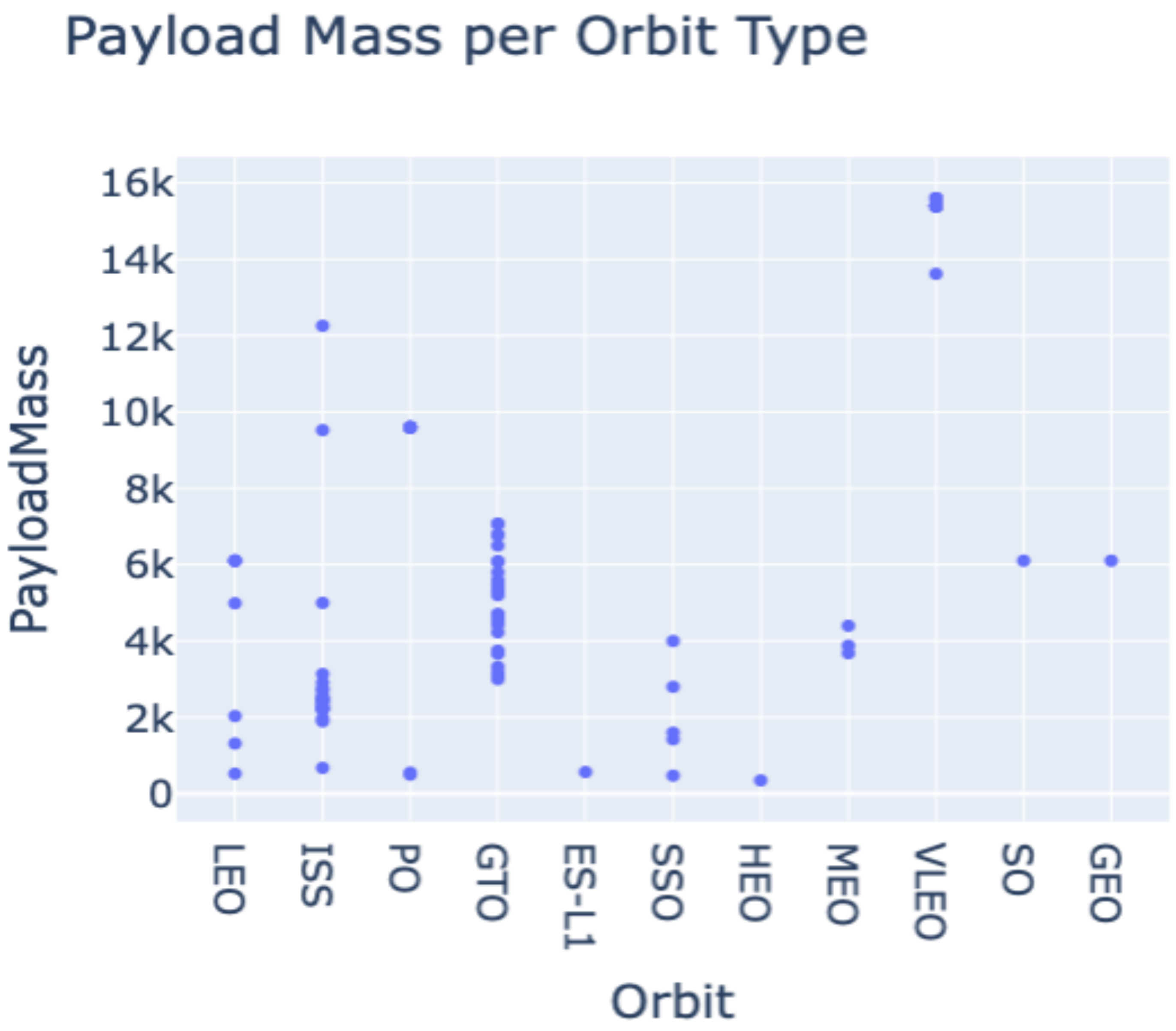
• **Large** sequence of VLEO orbit attempts started after **Flight #60**



Payload vs. Orbit Type

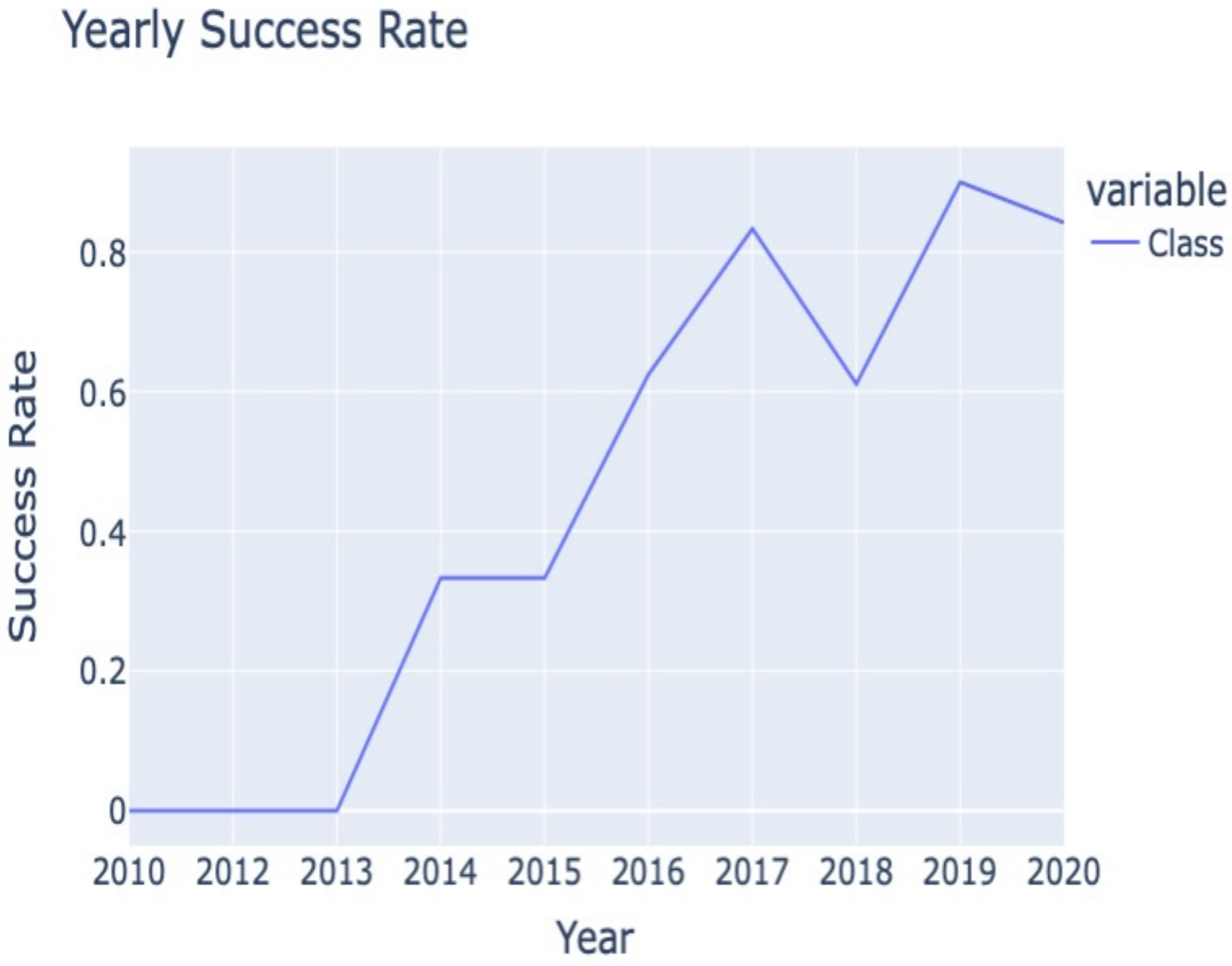
- Payload Mass distribution **concentrates** below 8,000 kg

- **GTO Payload range** between 3,000 & 7,000 kg



Launch Success Yearly Trend

- Over 80% increase in Success Rate from 2010 to 2020
- **60-90%** Success Rate since **2016**



All Launch Site Names

- Launch Sites:

  - KSC LC-39A

  - CCAFS LC-40

  - VAFB SLC-4E

  - CCAFS SLC-40

- QUERY:

```
SELECT DISTINCT
        LAUNCH_SITE
FROM 'launchData.spaceX'
```

Launch Site Names Begin with 'CCA'

- QUERY:

```
SELECT *
FROM `lightning-data-331802.launchData.spaceX`
WHERE LAUNCH_SITE
LIKE 'CCA%'
LIMIT 5
```

- Results:

Row	PAYLOAD_...	Orbit	Customer	Mission_Outcome	Landing__Outcome
1	3170	GTO	SES	Success	No attempt
2	3325	GTO	Thaicom	Success	No attempt
3	4535	GTO	AsiaSat	Success	No attempt
4	4428	GTO	AsiaSat	Success	No attempt
5	4159	GTO	ABS Eutelsat	Success	No attempt

Total Payload Mass

- Calculate the total payload carried by boosters from NASA

- QUERY:

```
SELECT SUM(PAYLOAD_MASS__KG_) AS NASA_PAYLOAD
FROM `lightning-data-331802.launchData.spaceX`
WHERE CUSTOMER
LIKE 'NASA%'
```

- RESULTS:

  - 99980 kg

Average Payload Mass by F9 v1.1

- QUERY:

  SELECT avg(PAYLOAD_MASS__KG_) AS avg_PAYLOAD

  FROM `lightning-data-331802.launchData.spaceX`

  WHERE BOOSTER_VERSION

  LIKE 'F9 v1.1’

- Results:

  - Booster Version F9 v1.1 carried a **total payload of 2928.4 kg**

First Successful Ground Landing Date

- QUERY:

```
SELECT
        DATE, Landing__Outcome
FROM `lightning-data-331802.launchData.spaceX`
WHERE LANDING__OUTCOME
LIKE 'Success (ground pad)’
ORDER BY DATE
LIMIT 5
```

- Results:

  First successful ground landing took place **December 22, 2015**

Successful Drone Ship Landing with Payload between 4000 and 6000

- QUERY:

```
SELECT DISTINCT
        Booster_Version, Landing__Outcome
FROM `lightning-data-331802.launchData.spaceX`
WHERE Landing__Outcome = 'Success (drone ship)’
        AND PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000
```

- Results: The only Version Boosters to successfully land on drone ship

Row	Booster_Version	Landing__Outcome
1	F9 FT B1021.2	Success (drone ship)
2	F9 FT B1031.2	Success (drone ship)
3	F9 FT B1022	Success (drone ship)
4	F9 FT B1026	Success (drone ship)

Total Number of Successful and Failure Mission Outcomes

- QUERY:

```
SELECT DISTINCT
        COUNT(Mission_Outcome) AS TOTAL,  Mission_Outcome
FROM `lightning-data-331802.launchData.spaceX`
GROUP BY Mission_Outcome
```

- RESULTS: All 101 Mission Outcomes were successful besides **one**

Row	TOTAL	Mission_Outcome
1	98	Success
2	1	Failure (in flight)
3	1	Success (payload status uncle...
4	1	Success

Boosters Carried Maximum Payload

- QUERY:

```
SELECT
        MAX(PAYLOAD_MASS__KG_)
AS MAX_PAYLOAD, Booster_Version
FROM `lightning-data-
        331802.launchData.spaceX`
GROUP BY Booster_Version
ORDER BY MAX_PAYLOAD DESC
```

- Results: **12 Boosters carried the maximum payload of 15,600 kg**

Row	MAX_PAYL...	Booster_Version
1	15600	F9 B5 B1048.5
2	15600	F9 B5 B1051.4
3	15600	F9 B5 B1060.2
4	15600	F9 B5 B1058.3
5	15600	F9 B5 B1051.6
6	15600	F9 B5 B1048.4
7	15600	F9 B5 B1049.4
8	15600	F9 B5 B1051.3
9	15600	F9 B5 B1056.4
10	15600	F9 B5 B1049.5
11	15600	F9 B5 B1060.3
12	15600	F9 B5 B1049.7

2015 Launch Records

- QUERY:

```
SELECT Booster_Version, Launch_Site, DATE, Landing__Outcome
FROM `lightning-data-331802.launchData.spaceX`
WHERE DATE BETWEEN '2015-01-01' AND '2015-12-31'
        AND Landing__Outcome = 'Failure (drone ship)'
```

- Results: **Two Booster Versions at CCAFS LC-40 failed to land on drone ship (2015)**

Row	Booster_Version	Launch_Site	DATE	Landing_Outcome
1	F9 v1.1 B1012	CCAFS LC-40	2015-01-10	Failure (drone ship)
2	F9 v1.1 B1015	CCAFS LC-40	2015-04-14	Failure (drone ship)

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

- QUERY:

SELECT DISTINCT

COUNT(Landing__Outcome) AS
TOTAL_COUNT, LANDING__OUTCOME

FROM `lightning-data-331802.launchData.spaceX`

WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20'

GROUP BY Landing__Outcome

ORDER BY TOTAL_COUNT DESC

Row	TOTAL_COU...	LANDING__OUTCOME
1	10	No attempt
2	5	Failure (drone ship)
3	5	Success (drone ship)
4	3	Success (ground pad)
5	3	Controlled (ocean)
6	2	Failure (parachute)
7	2	Uncontrolled (ocean)
8	1	Precluded (drone ship)

- Results:
  - Majority of landings were **NOT attempted**
  - 50% Success Rate landing on **drone ship**

Section 3

Launch Sites Proximities Analysis

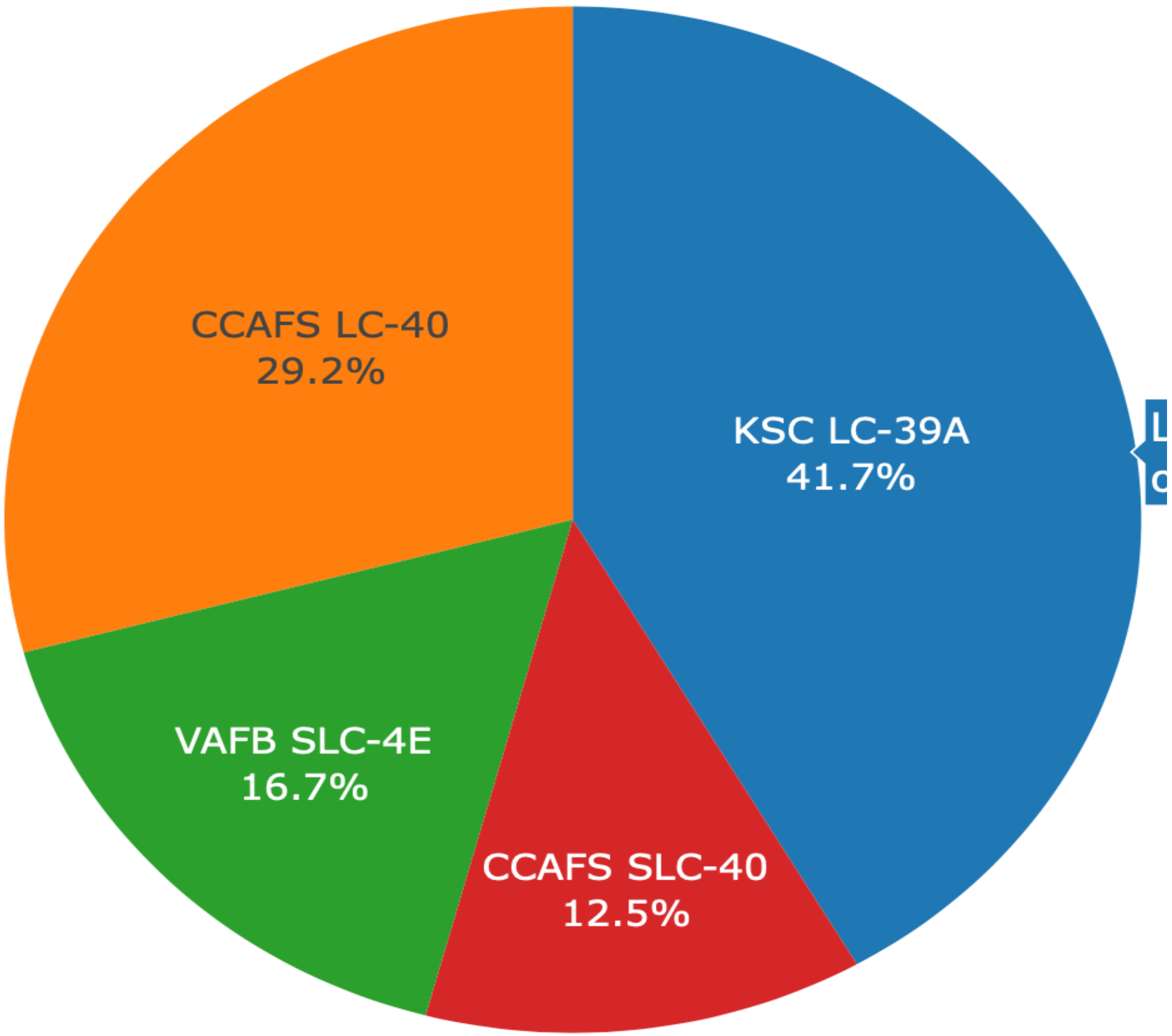
Section 4

Build a Dashboard with Plotly Dash

Success Rate: All Launch Site

- Nearly **42% of Successful Landings** performed at **Kennedy Space Complex**

- **CCAFS SLC-40** performed the **worst**, with a **12.5% Success Rate among other Launch Sites**

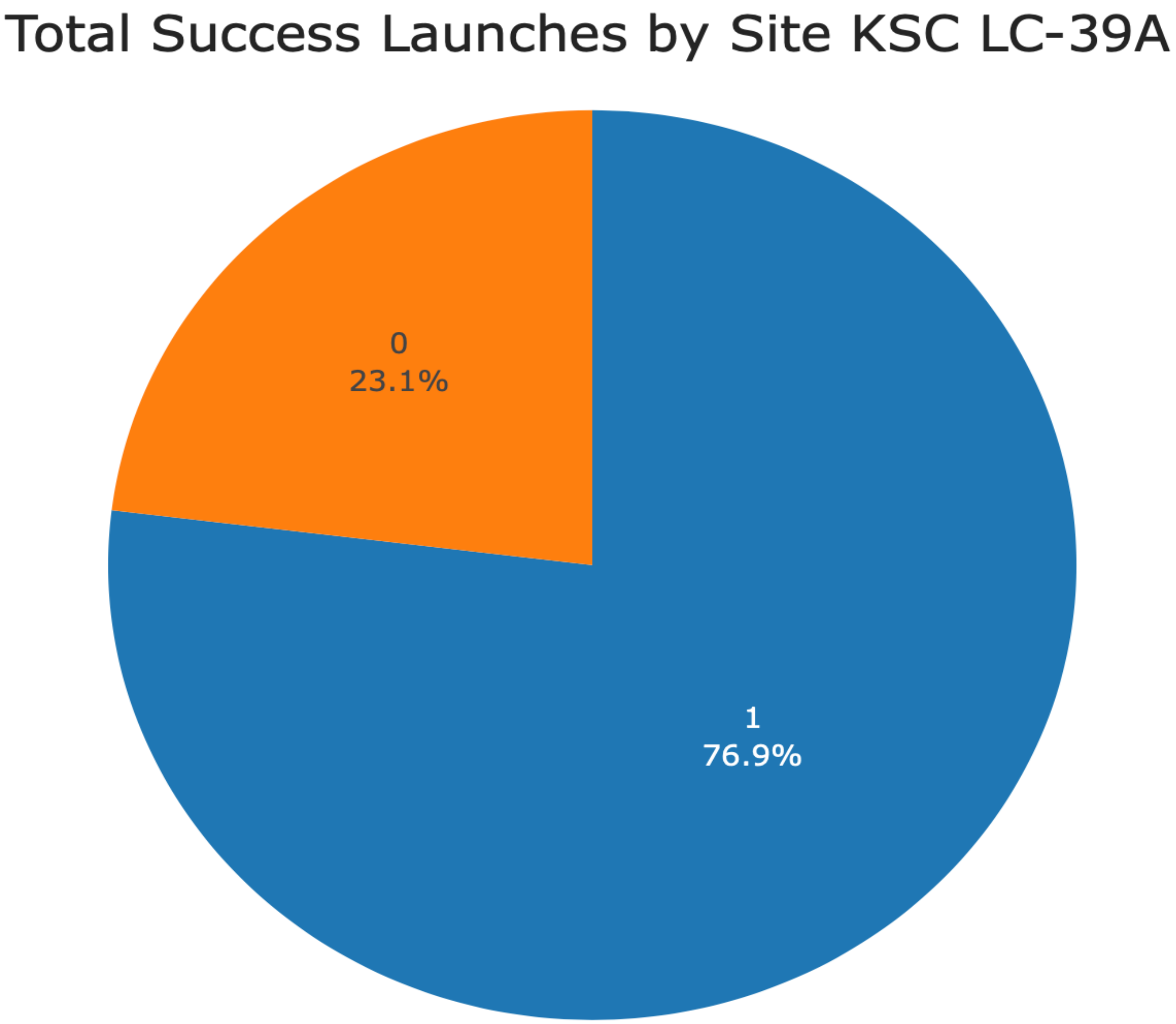


Kennedy Space Complex:
Highest Success Ratio

- KSC LC-39A has the **highest Success Ratio**

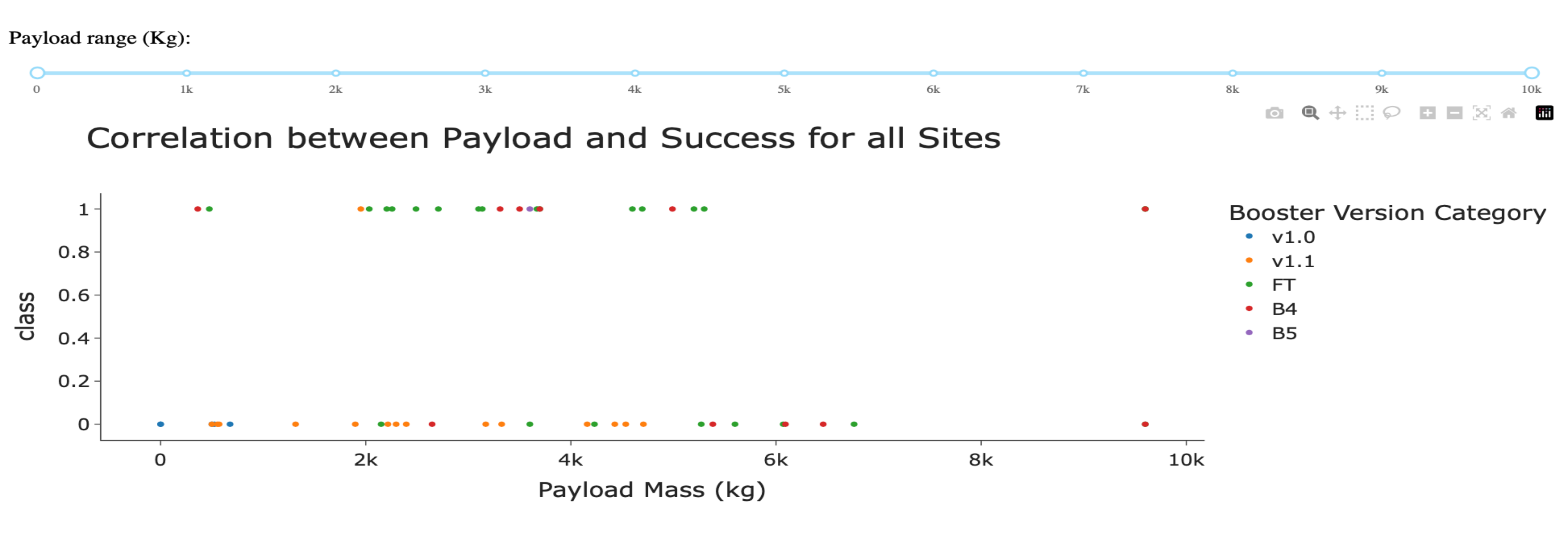
  - 76.9% Success

- **Less** than a **quarter** of launches were **failures**, compared to **VAFB's 60% Failure Ratio**



<Dashboard Screenshot 3>

- **Payload Mass ranges between 2,000 and 6,000 kg have high success**, especially among Booster Version (FT)
- Version Booster (v1.1) has **high failure among Payload Mass ranges between 500 and 5000 kg**



Section 5

Predictive Analysis (Classification)

Classification Accuracy:

- Highest out-of-sample accuracy **(~ 89%)**

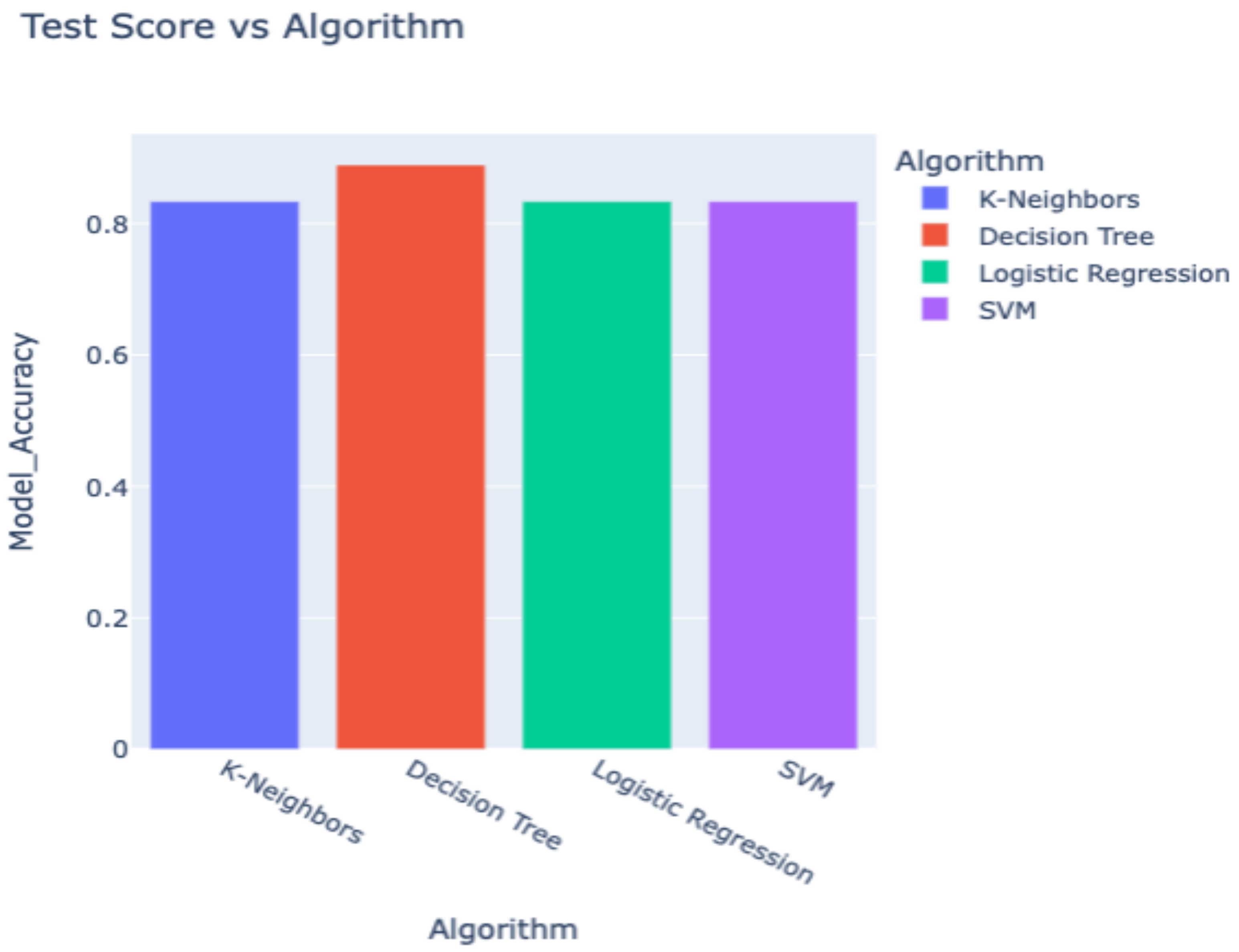
  - **Decision Tree**

- Evaluation Metric:

  - **Accuracy**

- Function:

(TP-TN) / (TP+FP+FN+TN)



Confusion Matrix

- Test set contained 18 observations / predictions:

  - Did Land / Did Not Land

- Land predictions:

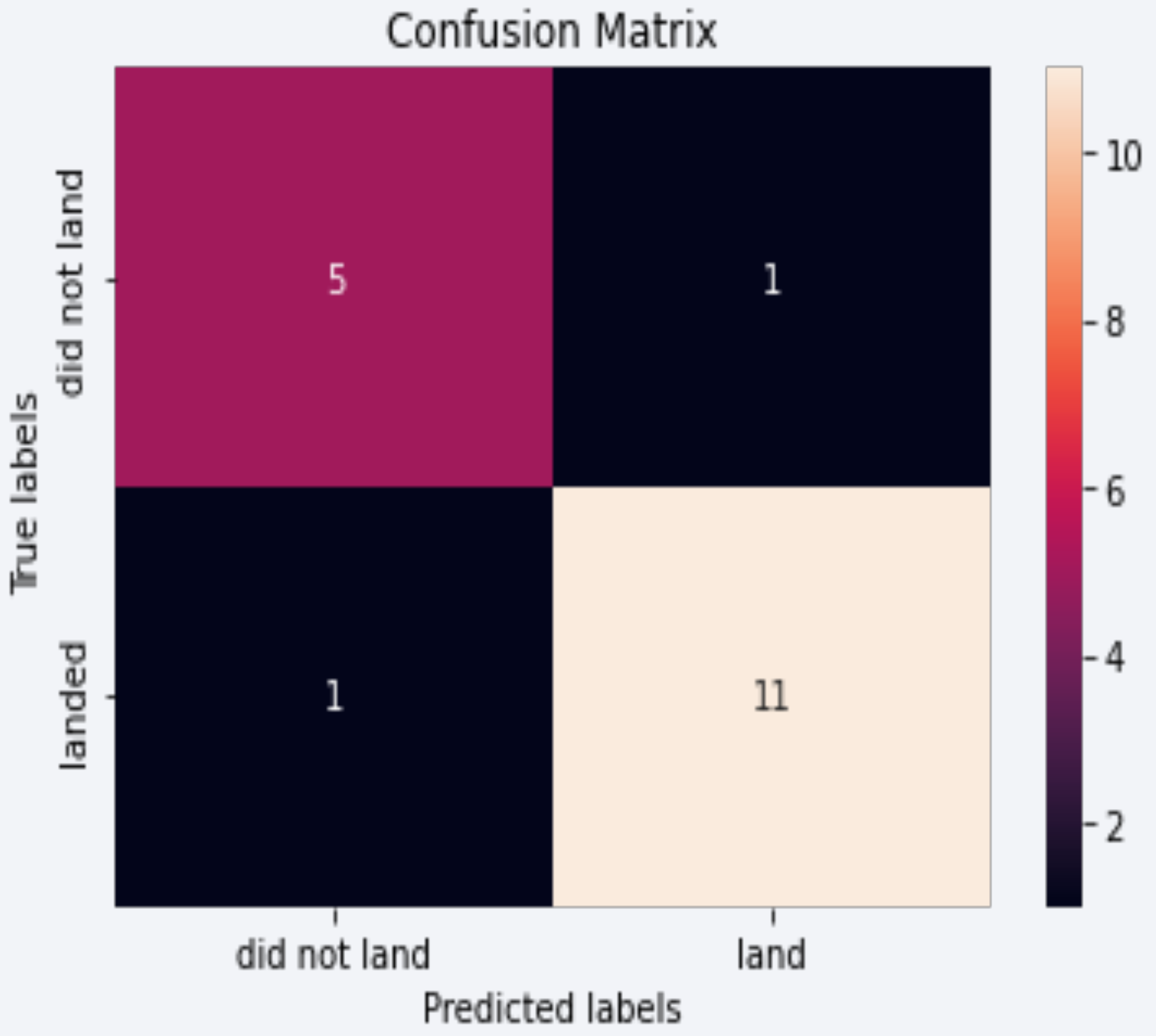
  - True Positive (11/12)

  - False Positive (1/12)

- Did Not Land predictions:

  - True Negative (5/6)

  - False Negative (1/6)



Conclusions

- **WHY** should SpaceY strive to build a **fully reusable rocket?**

  - Reusing rockets **reduces ALLOY COSTS** by reducing the amount of **alloy needed per launch**

  - Reduces **CAPITAL EXPENDITURES**

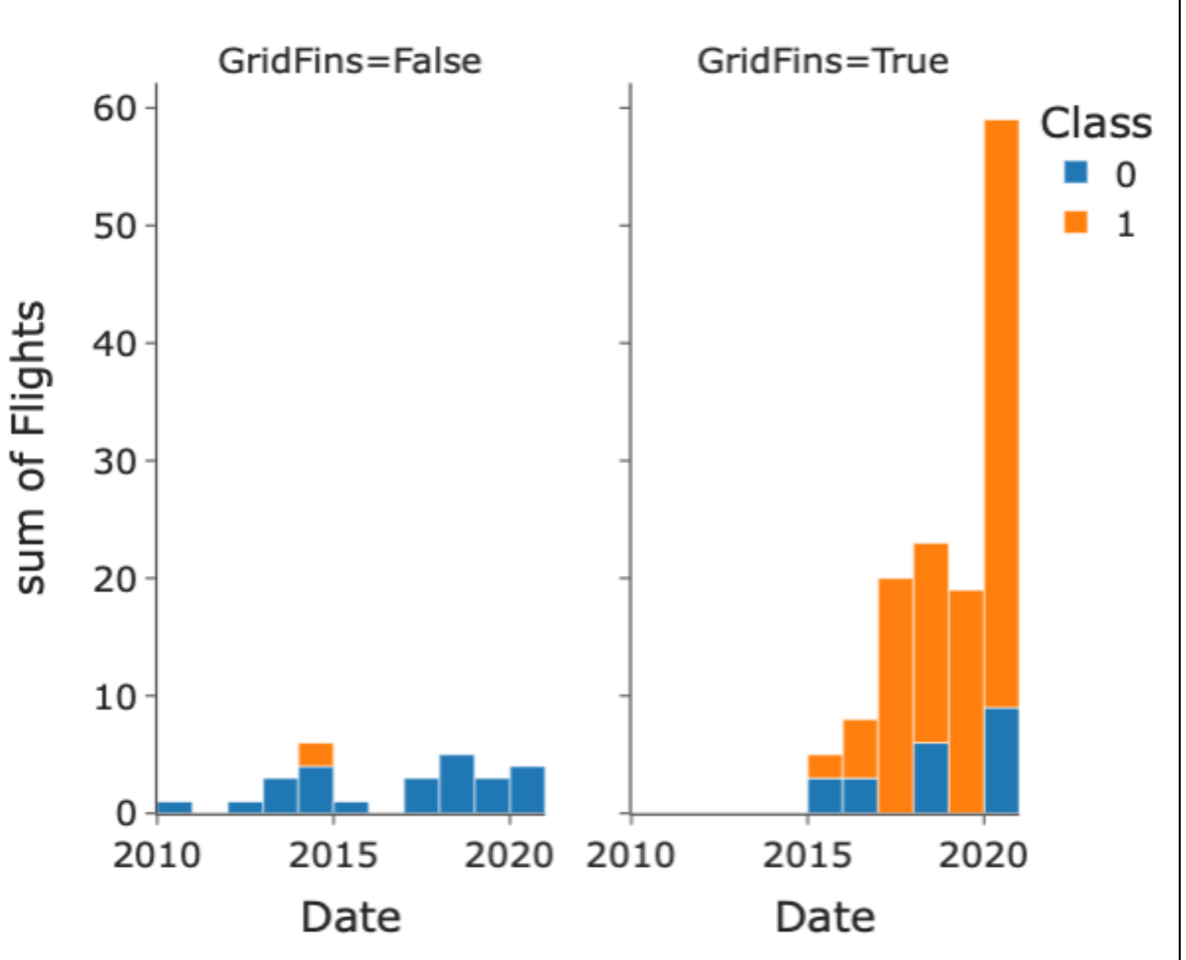
  - Reduces **TIME-TO-BUILD**, allowing for increased **flights per year**

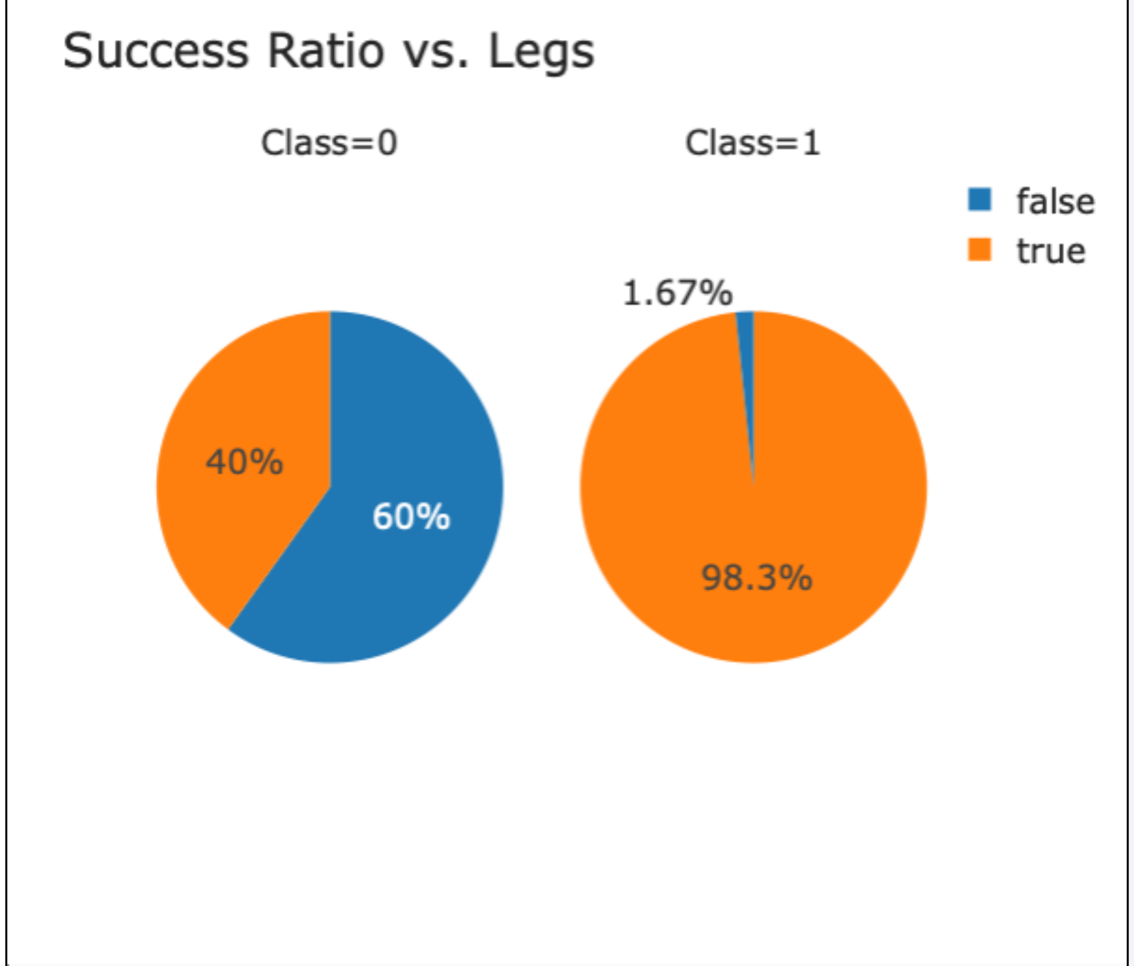
- **HOW** can SpaceY increase the **probability** of a **successful landing**?

  - Introducing **Grid Fins & Legs increased the Success Ratio from 10 - 82.9%**

Appendix

[EDA](#): BARCHART & PIE CHART– **Grid Fins & Legs dramatically INCREASED Success Ratio**





Thank you!