

# Winning Space Race with Data Science

Amritansh Kushwaha  
15-01-2026



# Outline

---

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

# Executive Summary

---

- **Methodology**
- Extracted data using a combination of web scraping and the SpaceX API.
- Conducted EDA and data wrangling to prepare the dataset for interactive visualization.
- Applied machine learning techniques to build a predictive framework.
- **Key Findings**
- Confirmed the viability of using public data sources for complex analysis.
- Used visualization and EDA to isolate the most important predictors of launch success.
- Determined the most effective model for analyzing how specific mission parameters drive successful outcomes.

# Introduction

---

- **Objective** To determine the viability of **Space Y** as a direct competitor to **SpaceX** by analyzing cost-drivers and infrastructure requirements.
- **Primary Objectives**
- Develop a methodology to estimate total launch costs by predicting the probability of successful first-stage landings.
- Determine the most advantageous sites for launch operations to minimize fuel consumption and maximize payload capacity.

Section 1

# Methodology

# Methodology

---

## Executive Summary

- Data collection methodology:
  - Data from Space X was obtained from 2 sources:
  - Space X API (<https://api.spacexdata.com/v4/rockets/>)
  - WebScraping ([https://en.wikipedia.org/wiki/List\\_of\\_Falcon/\\_9/\\_and\\_Falcon\\_Heavy\\_launches](https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches))
- Perform data wrangling
  - Collected data was enriched by creating a landing outcome label based on outcome data after summarizing and analyzing features
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Data that was collected until this step were normalized, divided in training and test data sets and evaluated by four different classification models, being the accuracy of each model evaluated using different combinations of parameters

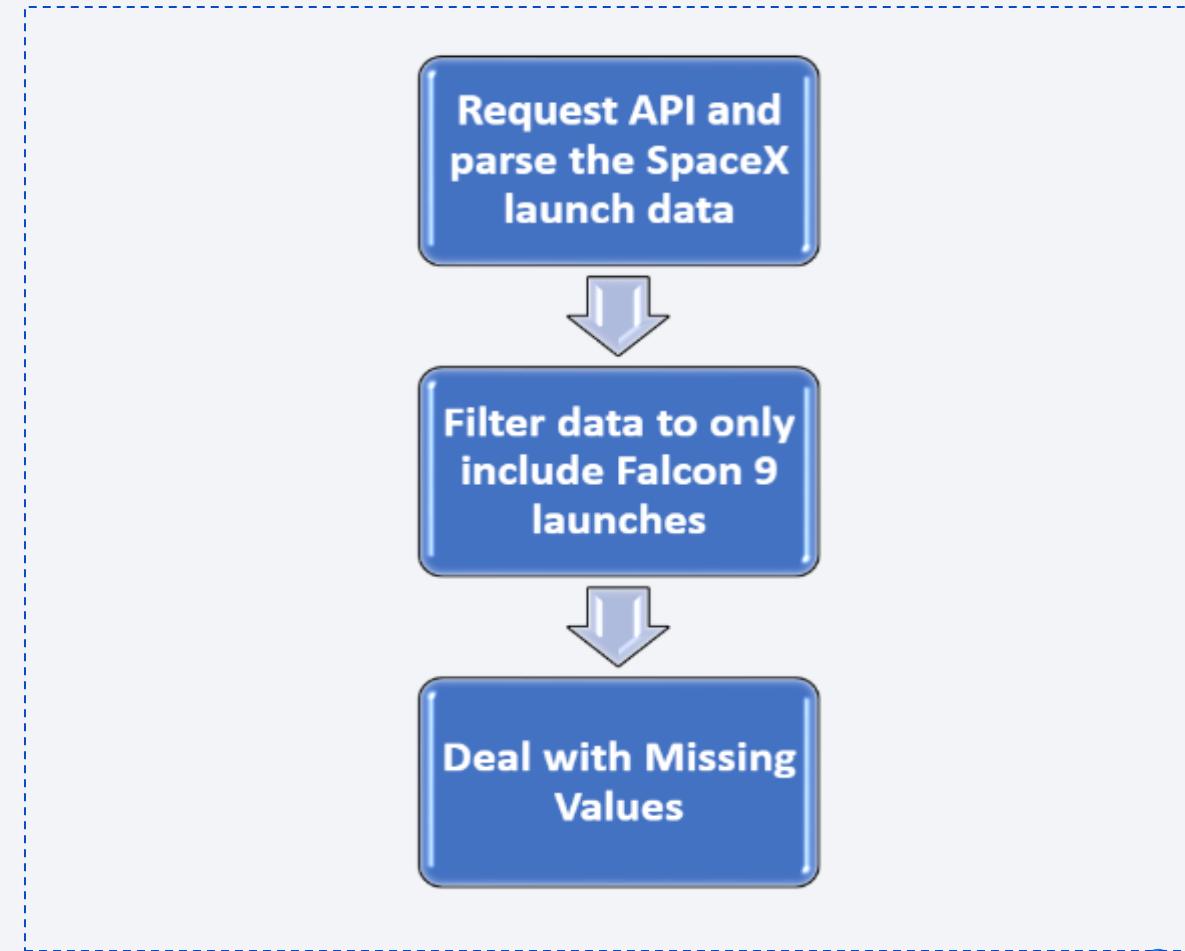
# Data Collection

---

- Data sets were collected from Space X API (<https://api.spacexdata.com/v4/rockets/>) and from Wikipedia ([https://en.wikipedia.org/wiki/List\\_of\\_Falcon\\_9\\_and\\_Falcon\\_Heavy\\_launches](https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches)), using web scraping technics.

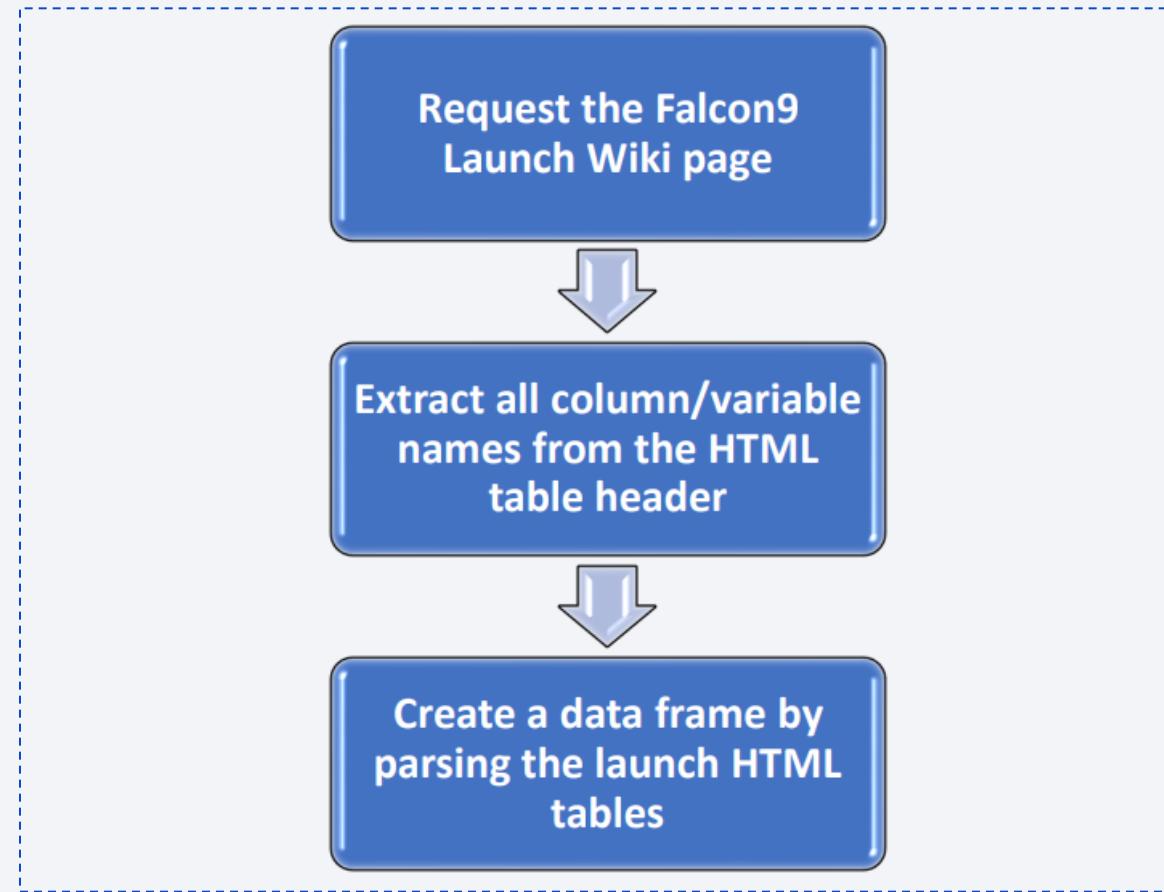
# Data Collection – SpaceX API

- SpaceX offers a public API from where data can be obtained and then used; • This API was used according to the flowchart beside and then data is persisted.
- Source:  
[https://github.com/Amritansh236/  
Data-Science-  
Capstone/blob/main/Data%20Coll  
ection.ipynb](https://github.com/Amritansh236/Data-Science-Capstone/blob/main/Data%20Collection.ipynb)



# Data Collection - Scraping

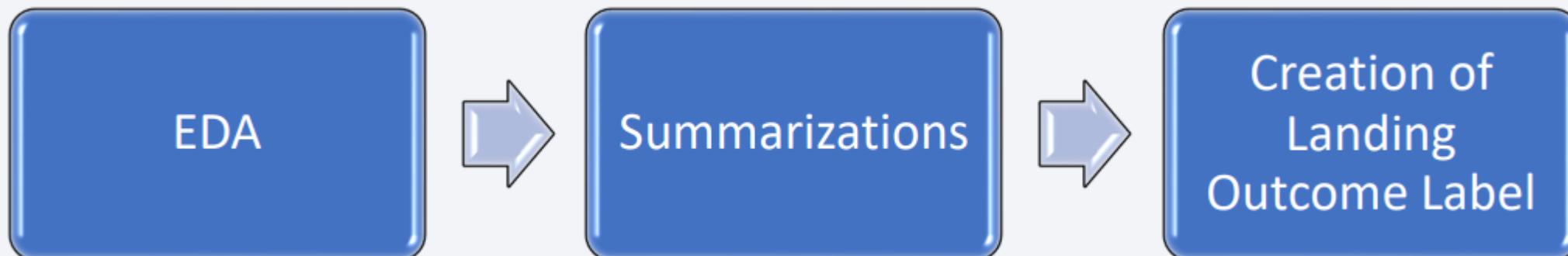
- Data from SpaceX launches can also be obtained from Wikipedia;
- Data are downloaded from Wikipedia according to the flowchart and then persisted.
- Source:  
[https://github.com/Amritansh236/Data-Science-Capstone/blob/main/Data%20Collection\\_Webscraping.ipynb](https://github.com/Amritansh236/Data-Science-Capstone/blob/main/Data%20Collection_Webscraping.ipynb)



# Data Wrangling

---

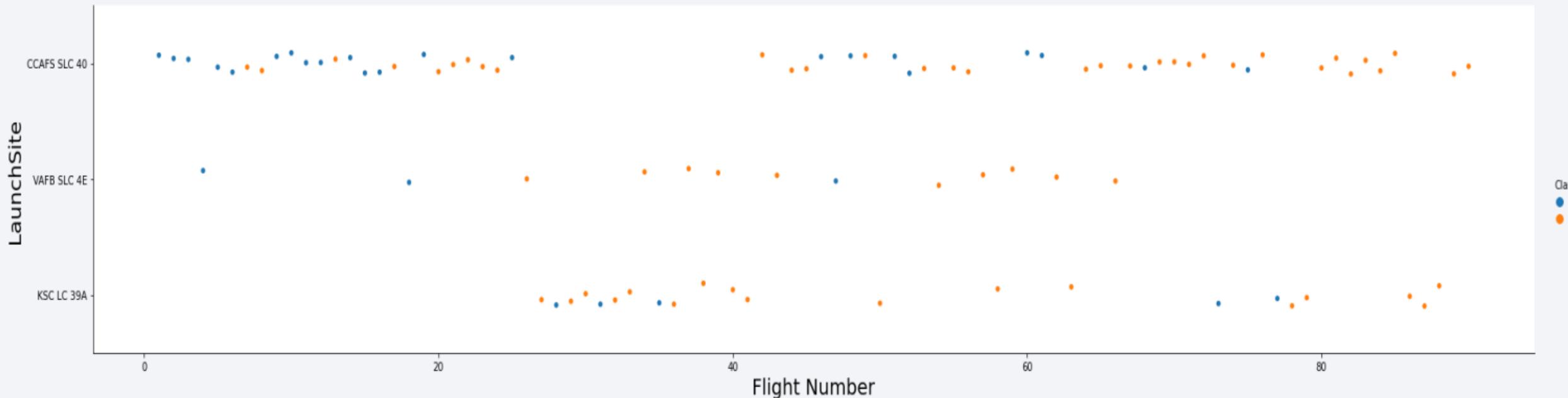
- Performed EDA to identify patterns and anomalies within the SpaceX flight data.
- Quantified mission metrics by site and orbit, specifically analyzing how different orbital trajectories impacted the final mission results.
- Derived a new target variable for the predictive models by labeling landing outcomes based on the existing status descriptions.
- Source: <https://github.com/Amritansh236/Data-Science-Capstone/blob/main/Data%20Wrangling.ipynb>



# EDA with Data Visualization

---

- To explore data, scatterplots and barplots were used to visualize the relationship between pair of features:
- Source: <https://github.com/Amritansh236/Data-Science-Capstone/blob/main/EDA%20with%20Data%20Visualization.ipynb>



# EDA with SQL

---

The following SQL queries were performed:

- Names of the unique launch sites in the space mission;
- Top 5 launch sites whose name begin with the string 'CCA';
- Total payload mass carried by boosters launched by NASA (CRS);
- Average payload mass carried by booster version F9 v1.1;
- Date when the first successful landing outcome in ground pad was achieved;
- Names of the boosters which have success in drone ship and have payload mass between 4000 and 6000 kg;
- Total number of successful and failure mission outcomes;
- Names of the booster versions which have carried the maximum payload mass;
- Failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015
- Rank of the count of landing outcomes

# Build an Interactive Map with Folium

---

- Markers, circles, lines and marker clusters were used with Folium Maps
- Markers indicate points like launch sites
- Circles indicate highlighted areas around specific coordinates, like NASA Johnson Space Center
- Marker clusters indicates groups of events in each coordinate, like launches in a launch site
- Lines are used to indicate distances between two coordinates.
- Source: [https://github.com/Amritansh236/Data-Science-Capstone/blob/main/LaunchSite\\_Location.ipynb](https://github.com/Amritansh236/Data-Science-Capstone/blob/main/LaunchSite_Location.ipynb)

# Build a Dashboard with Plotly Dash

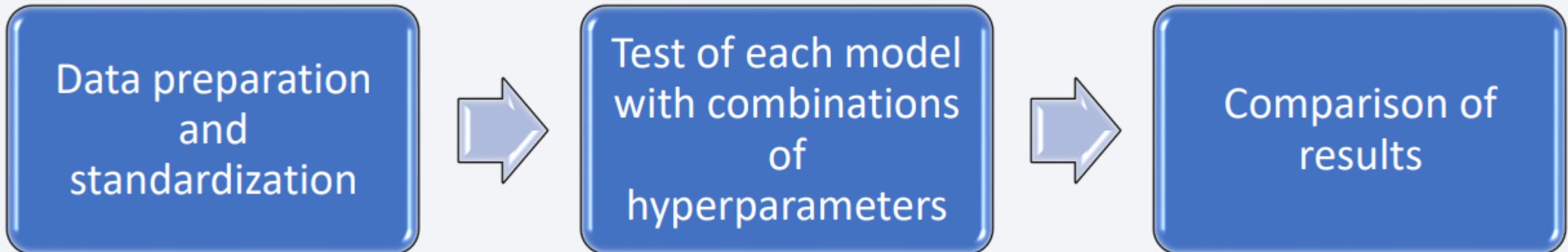
---

- Launch Site Metrics: We visualized the percentage of total missions handled by each site to understand operational density.
- Payload Profiling: We mapped out payload ranges to identify the typical mission profiles.
- Synthesis: Combining these two perspectives allowed for an immediate understanding of how payload size influences the choice of launch site, highlighting the best locations for different mission scales.
- Source: <https://github.com/Amritansh236/Data-Science-Capstone/blob/main/Machine%20Learning%20Prediction.ipynb>

# Predictive Analysis (Classification)

---

- Four classification models were compared: logistic regression, support vector machine, decision tree and k nearest neighbors
- Source:



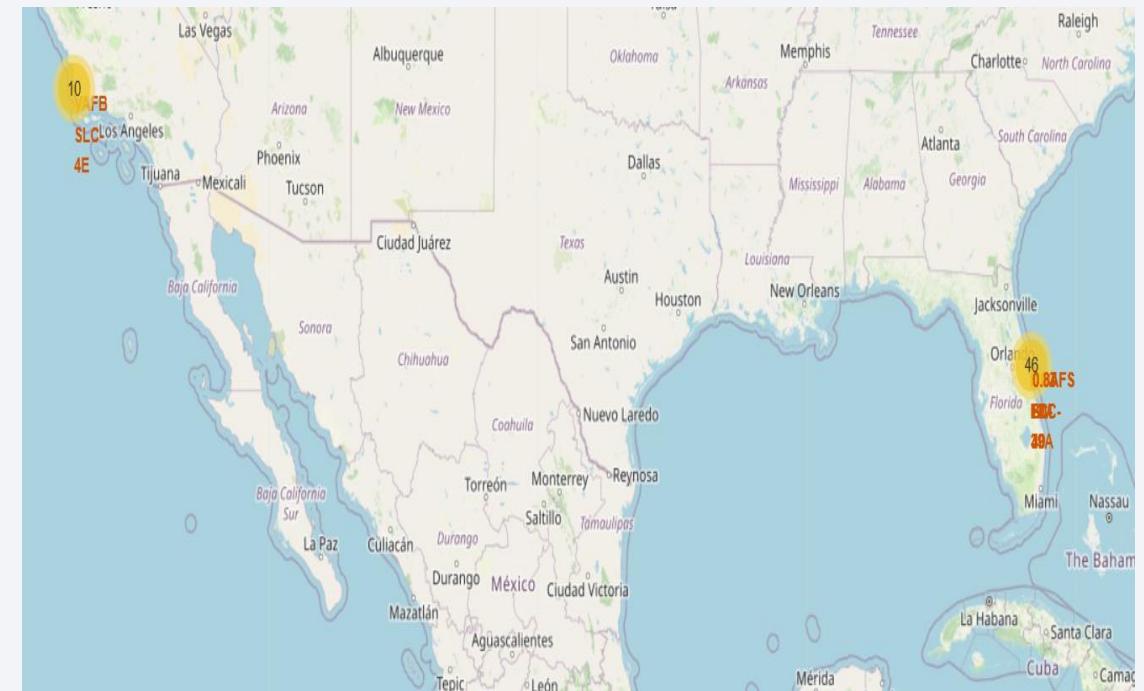
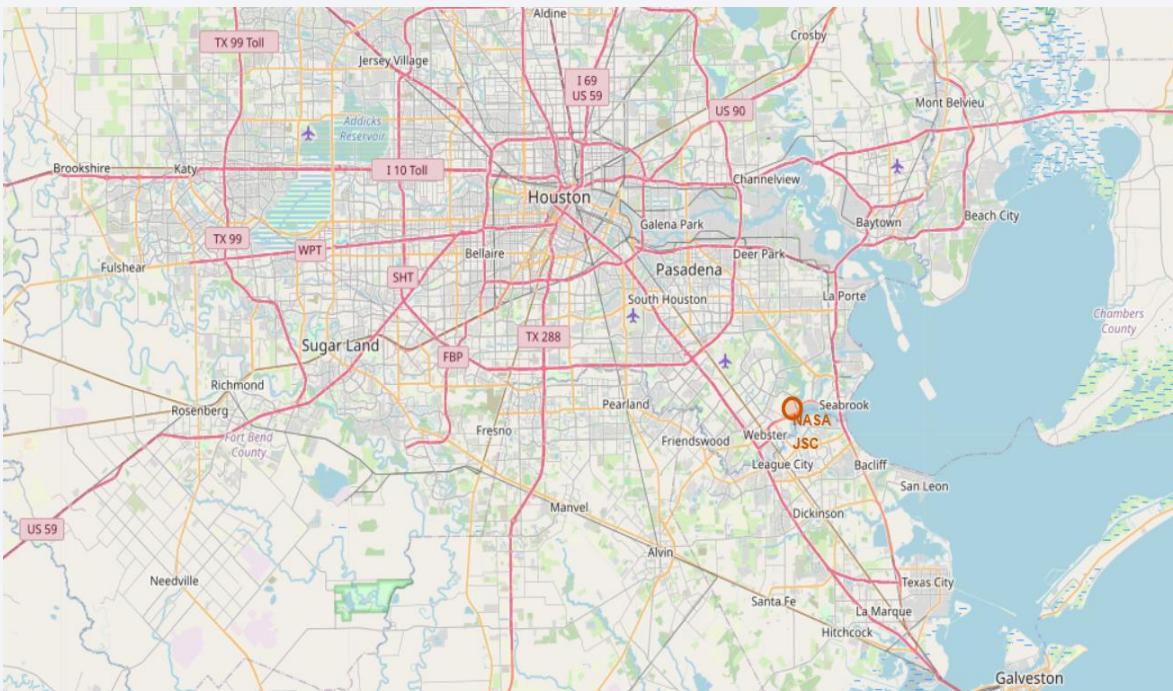
# Results

---

- Site Diversity: Data confirms the utilization of four primary launch complexes.
- Benchmark Payload: Calculated a baseline average payload of 2,928 kg for the F9 v1.1 booster.
- Success Timeline: Identified a five-year gap between the first launch and the first successful landing (2010 vs. 2015).
- Recovery Analysis: Observed that landing on autonomous spaceport drone ships (ASDS) was frequently successful for boosters carrying above-average payload weights.
- Outliers: 2015 saw two notable recovery failures involving the B1012 and B1015 boosters.
- Site Analytics: Visualization confirmed that coastal proximity is a standard requirement for site safety and infrastructure efficiency.
- Model Validation: The Decision Tree Classifier outperformed other algorithms, yielding a 94% accuracy score on unseen test data.

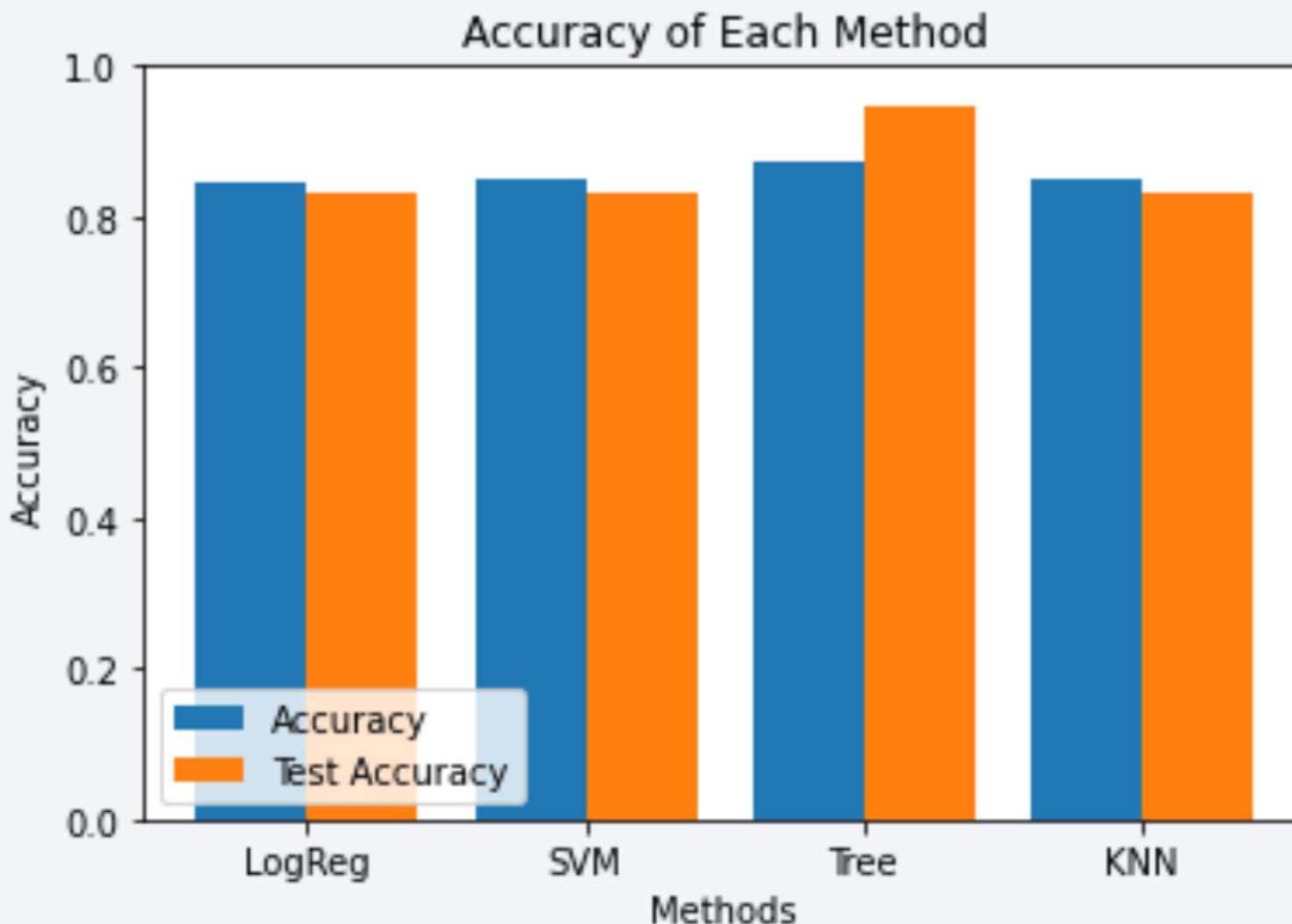
# Results

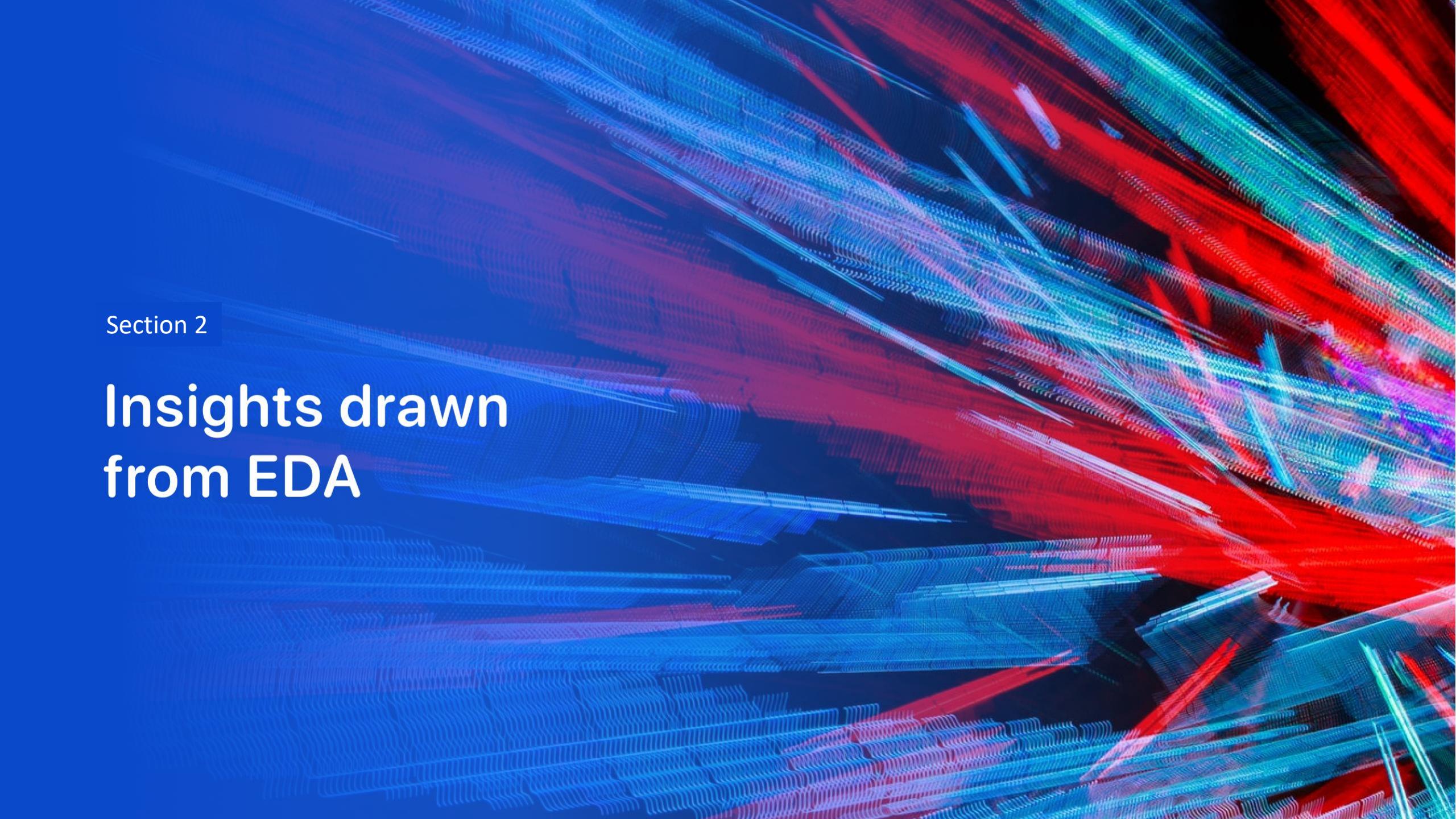
---



# Results

---



The background of the slide features a complex, abstract digital visualization. It consists of numerous thin, glowing lines that create a sense of depth and motion. The lines are primarily blue and red, with some green and purple highlights. They form a grid-like structure that curves and twists across the frame, resembling a 3D wireframe or a network of data points. The overall effect is futuristic and dynamic, suggesting concepts like data flow, digital communication, or complex systems.

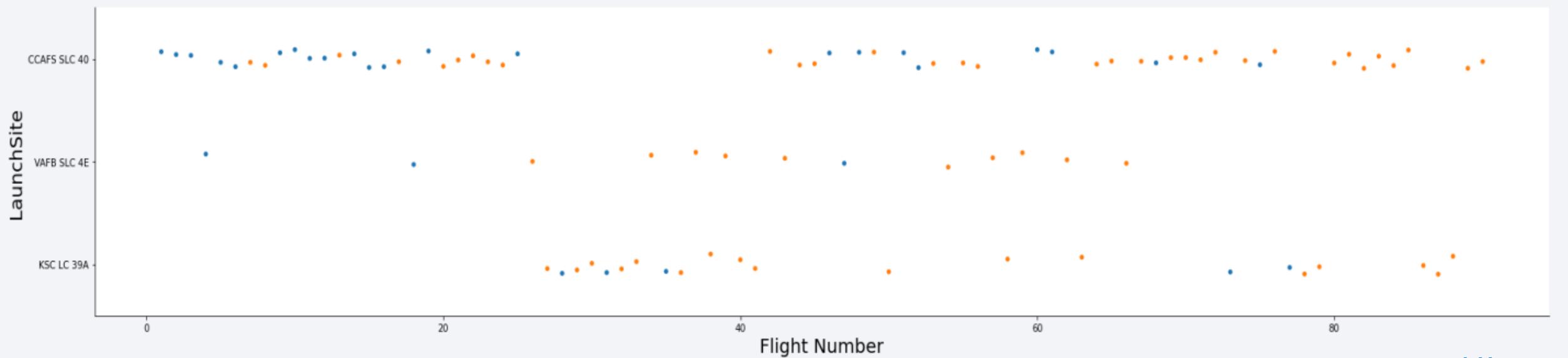
Section 2

## Insights drawn from EDA

# Flight Number vs. Launch Site

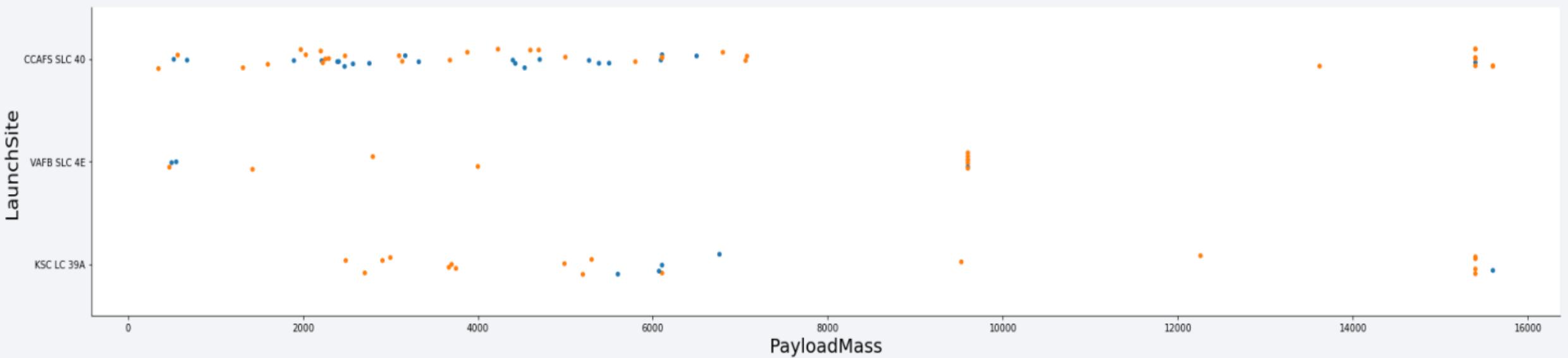
---

- According to the plot above, it's possible to verify that the best launch site nowadays is CCAF5 SLC 40, where most of recent launches were successful;
- In second place VAFB SLC 4E and third place KSC LC 39A;
- It's also possible to see that the general success rate improved over time.



# Payload vs. Launch Site

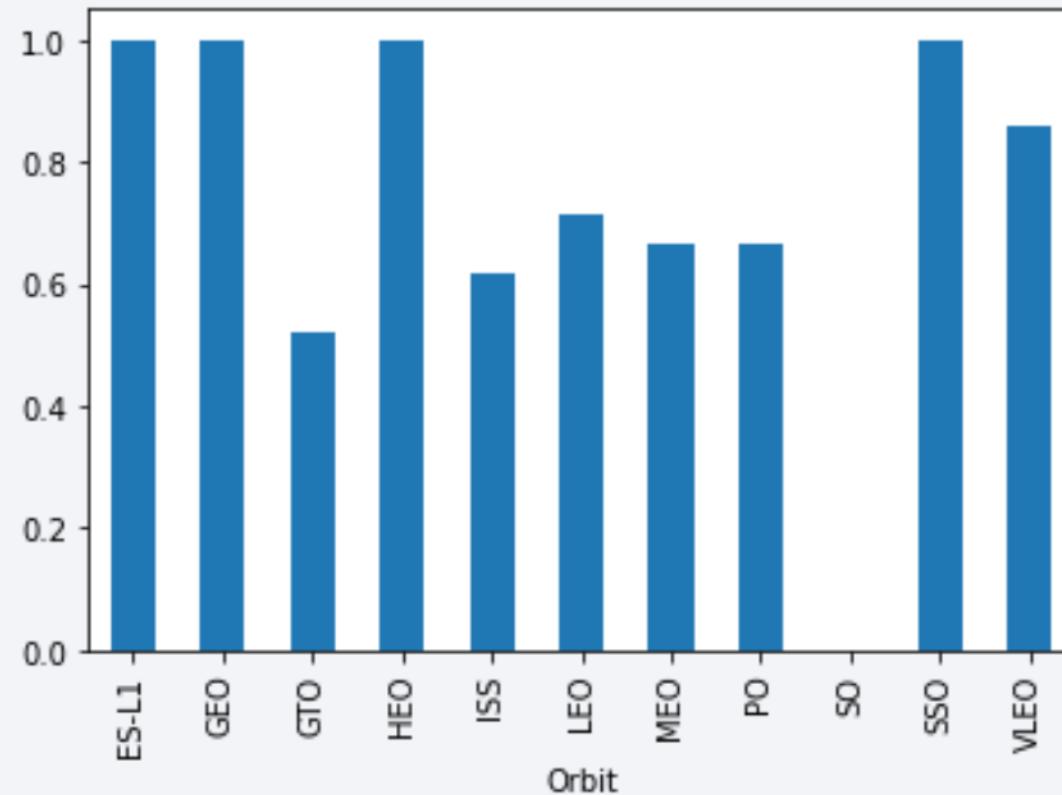
- Payloads over 9,000kg (about the weight of a school bus) have excellent success rate;
- Payloads over 12,000kg seems to be possible only on CCAFS SLC 40 and KSC LC 39A launch sites.



# Success Rate vs. Orbit Type

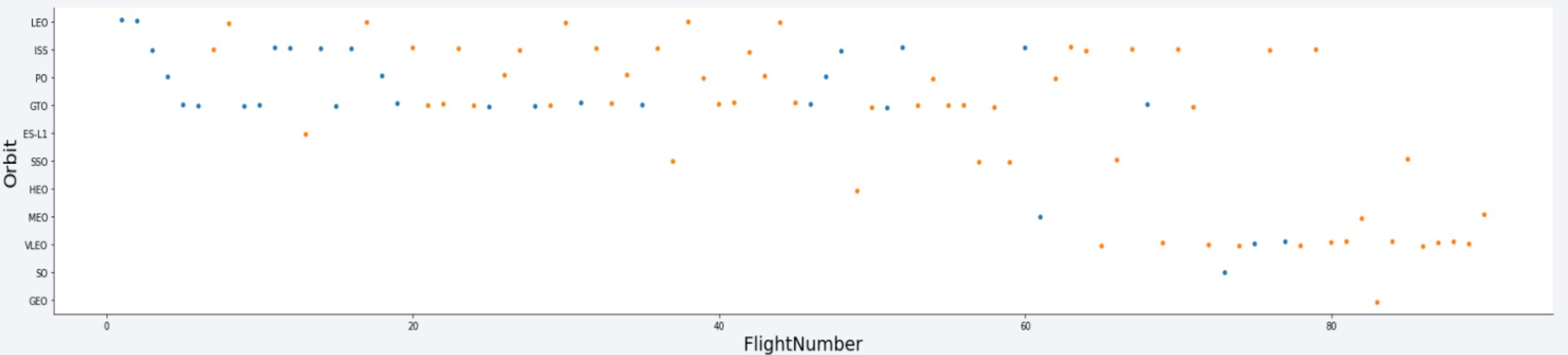
---

- The biggest success rates happens to orbits:
- ES-L1;
- GEO;
- HEO
- SSO.
- **Followed by:**
- VLEO (above 80%)
- LFO (above 70%).



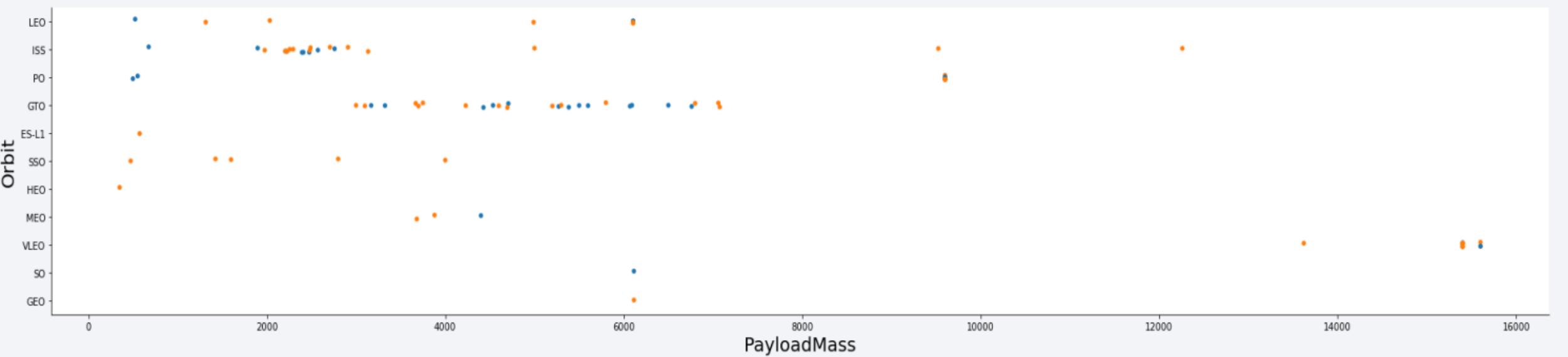
# Flight Number vs. Orbit Type

- Apparently, success rate improved over time to all orbits;
  - VLEO orbit seems a new business opportunity, due to recent increase of its frequency.



# Payload vs. Orbit Type

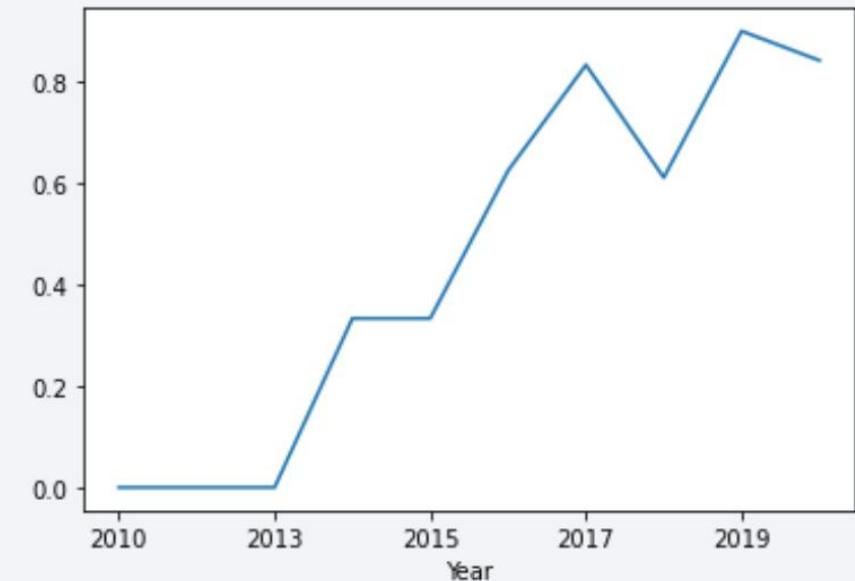
- Apparently, there is no relation between payload and success rate to orbit GTO;
- ISS orbit has the widest range of payload and a good rate of success;
- There are few launches to the orbits SO and GEO.



# Launch Success Yearly Trend

---

- Success rate started increasing in 2013 and kept until 2020;
- It seems that the first three years were a period of adjusts and improvement of technology.



# All Launch Site Names

---

- Obtained by selecting unique occurrences of “launch\_site” values from the dataset.

Launch Site
CCAFS LC-40
CCAFS SLC-40
KSC LC-39A
VAFB SLC-4E

# Launch Site Names Begin with 'CCA'

---

- Five samples of Cape Canaveral launches

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	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	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 attemp

# Total Payload Mass

---

- Total payload carried by boosters from NASA:

Total Payload (kg)
111.268

# Average Payload Mass by F9 v1.1

---

- Average payload mass carried by booster version F9 v1.1:

Avg Payload (kg)
2.928

# First Successful Ground Landing Date

---

- First successful landing outcome on ground pad:

**Min Date**

2015-12-22

## Successful Drone Ship Landing with Payload between 4000 and 6000

---

- Boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

Booster Version
F9 FT B1021.2
F9 FT B1031.2
F9 FT B1022
F9 FT B1026

## Total Number of Successful and Failure Mission Outcomes

---

- Number of successful and failure mission outcomes:

Mission Outcome	Occurrences
Success	99
Success (payload status unclear)	1
Failure (in flight)	1

# Boosters Carried Maximum Payload

---

- Boosters which have carried the maximum payload mass

Booster Version (...)
F9 B5 B1048.4
F9 B5 B1048.5
F9 B5 B1049.4
F9 B5 B1049.5
F9 B5 B1049.7
F9 B5 B1051.3

Booster Version
F9 B5 B1051.4
F9 B5 B1051.6
F9 B5 B1056.4
F9 B5 B1058.3
F9 B5 B1060.2
F9 B5 B1060.3

# 2015 Launch Records

---

- Failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015

Booster Version	Launch Site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

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

---

- Ranking of all landing outcomes between the date 2010-06-04 and 2017-03-20:

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

The background of the slide is a photograph taken from space at night. It shows the curvature of the Earth against the dark void of space. City lights are visible as numerous small white and yellow dots, primarily concentrated in the lower right quadrant where the United States appears. In the upper left quadrant, the green and blue glow of the aurora borealis is visible in the upper atmosphere.

Section 3

# Launch Sites Proximities Analysis

# <Folium Map Screenshot 1>

---

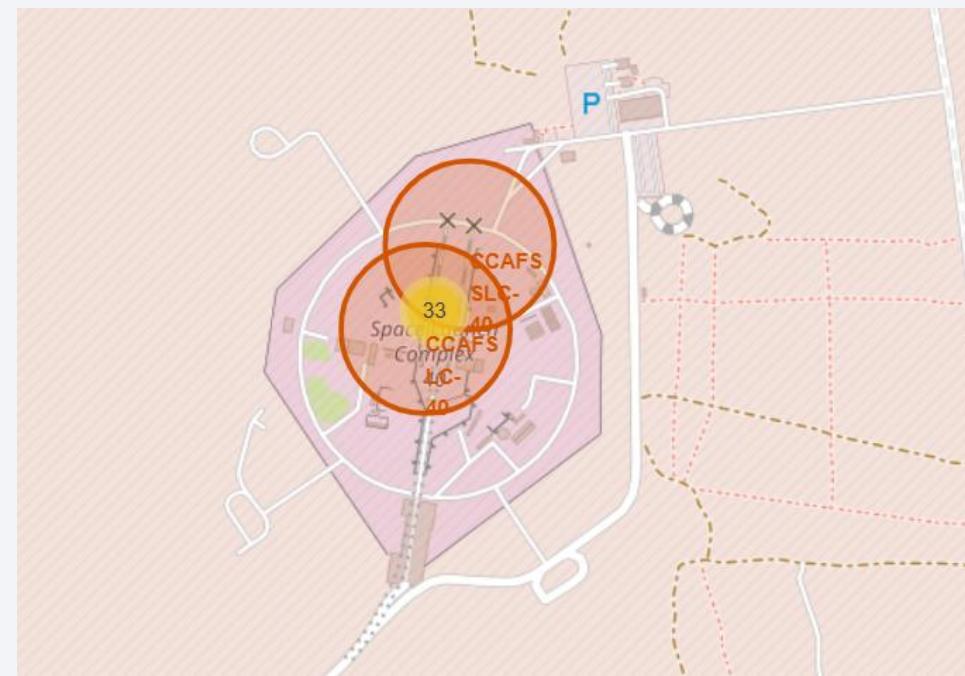
- Launch sites are near coastlines



## <Folium Map Screenshot 2>

---

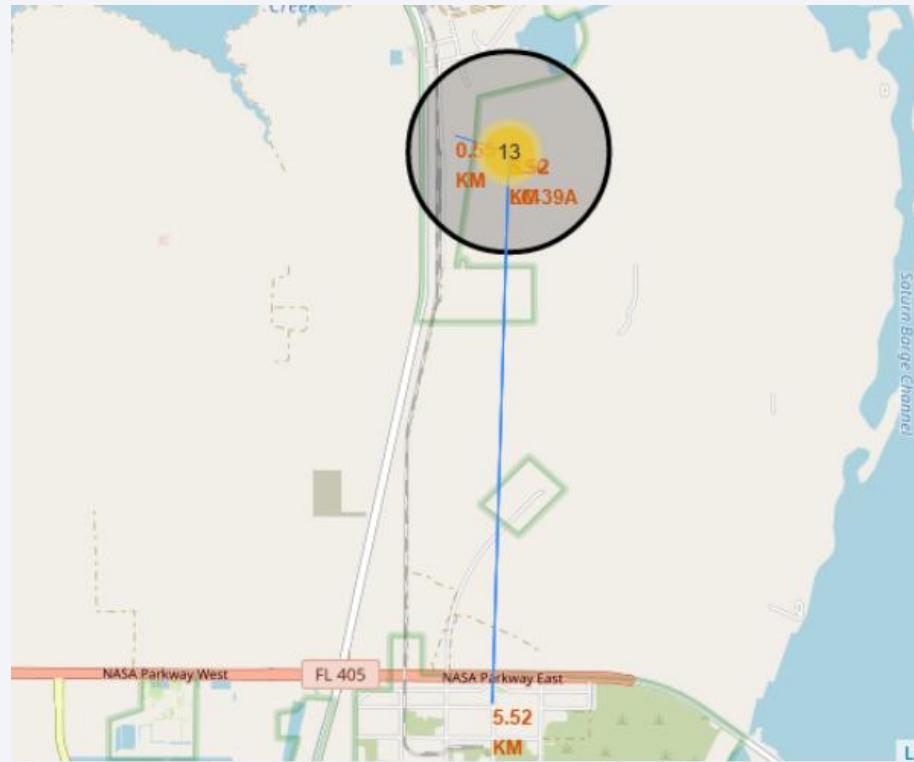
- Green markers indicate successful and red ones indicate failure.



## <Folium Map Screenshot 3>

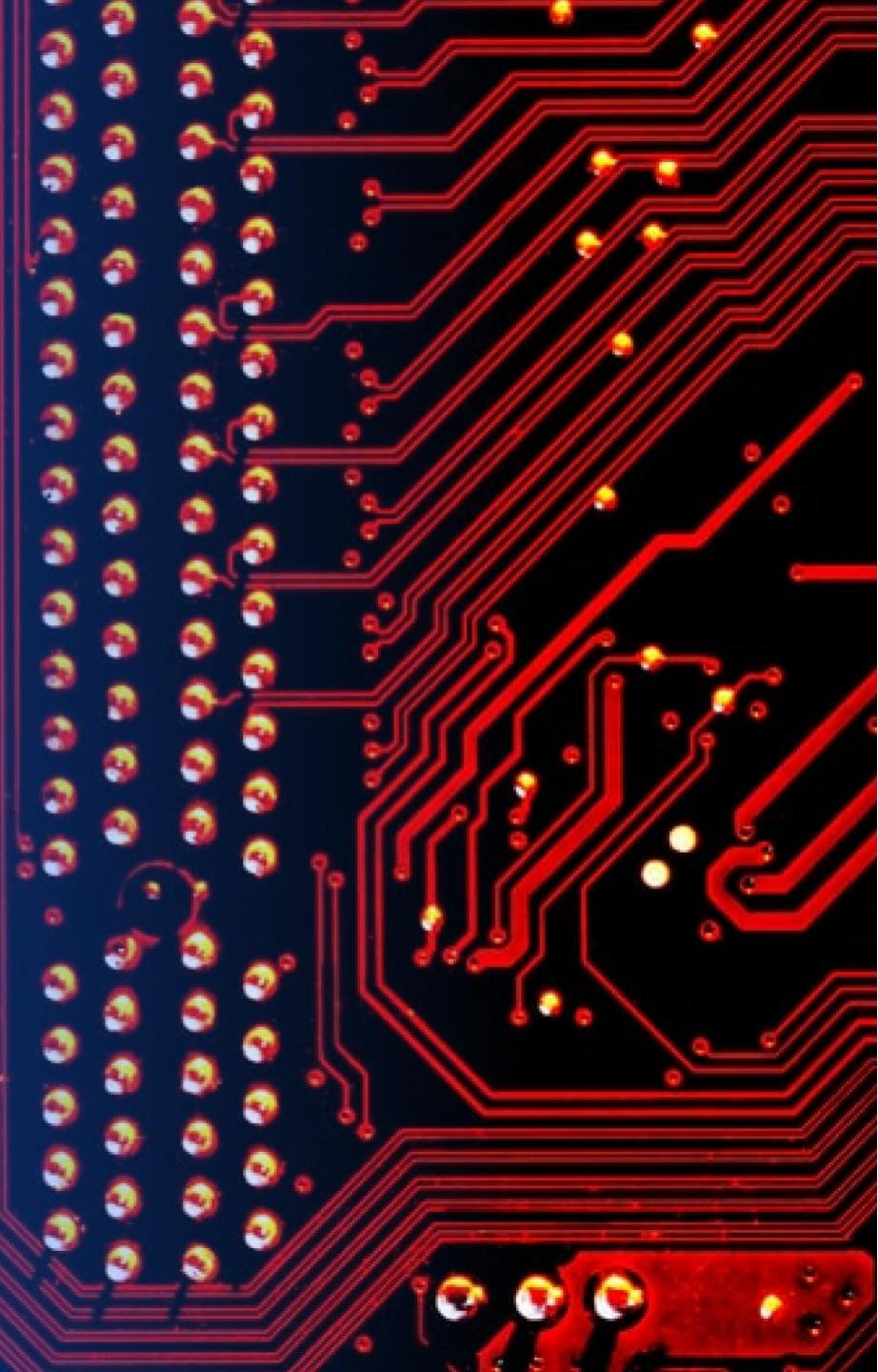
---

- Launch site KSC LC-39A has good logistics aspects, being near railroad and road and relatively far from inhabited areas



Section 4

# Build a Dashboard with Plotly Dash



# <Dashboard Screenshot 1>

---

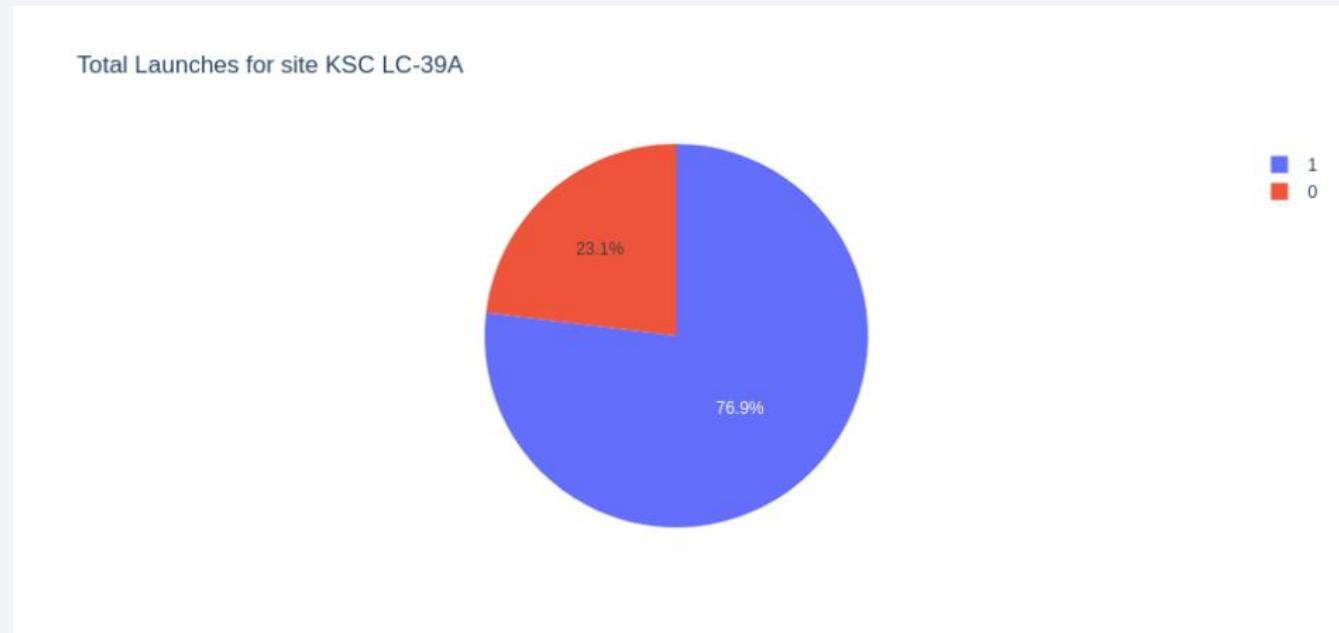
- The place from where launches are done seems to be a very important factor of success of missions.



## <Dashboard Screenshot 2>

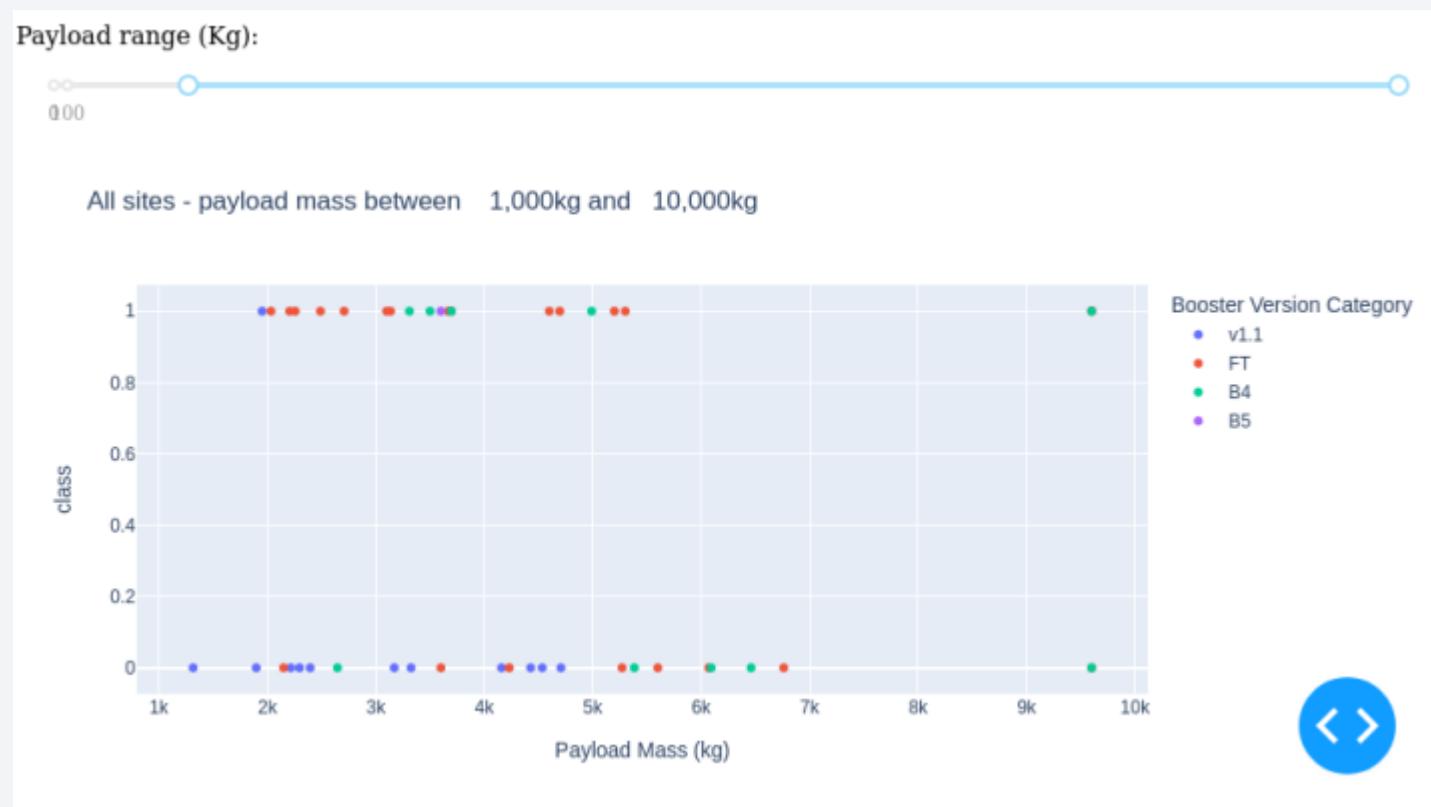
---

- 76.9% of launches are successful in this site and 23.1% failed.



## <Dashboard Screenshot 3>

- Payloads under 6,000kg and FT boosters are the most successful combination.



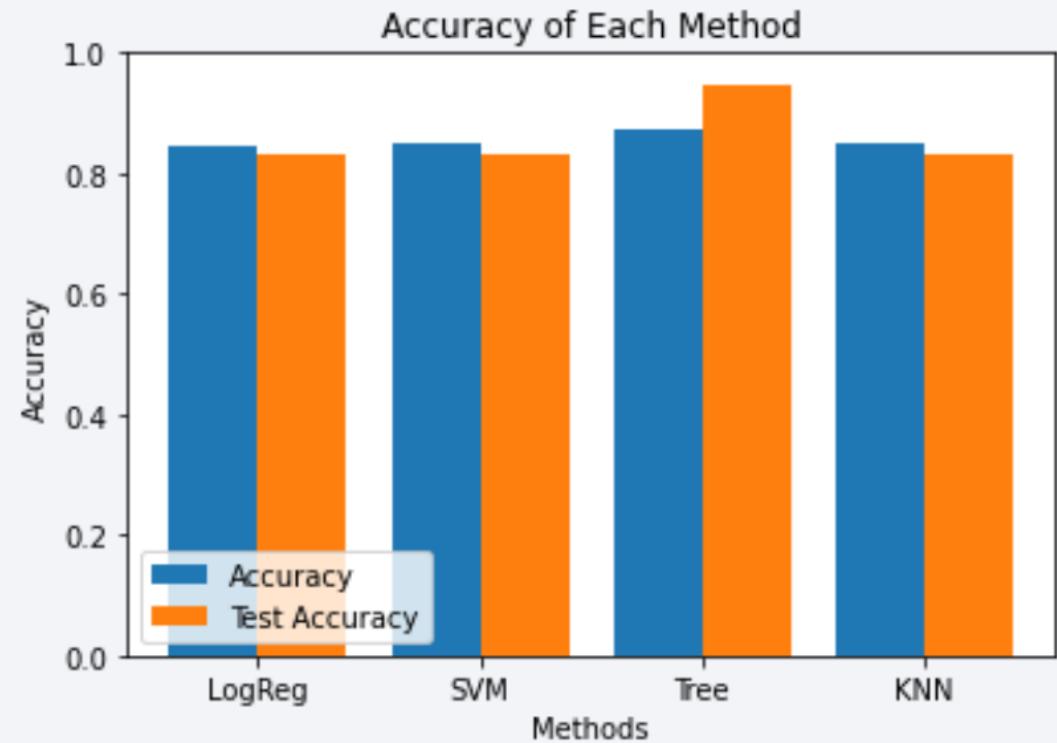
Section 5

# Predictive Analysis (Classification)

# Classification Accuracy

---

- The model with the highest classification accuracy is Decision Tree Classifier, which has accuracies over than 87%.



# Confusion Matrix

---

- Confusion matrix of Decision Tree Classifier proves its accuracy by showing the big numbers of true positive and true negative compared to the false ones.



# Conclusions

---

- Different data sources were analyzed, refining conclusions along the process;
- The best launch site is KSC LC-39A;
- Launches above 7,000kg are less risky;
- Although most of mission outcomes are successful, successful landing outcomes seem to improve over time, according the evolution of processes and rockets;
- Decision Tree Classifier can be used to predict successful landings and increase profits.

# Appendix

---

- Folium doesn't generate outputs in Github. Please download the notebook to see the outputs.

Thank you!

