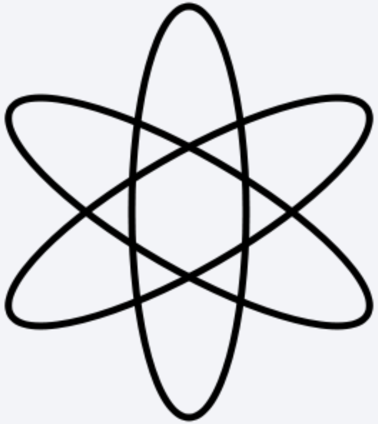
IBM Developer
SKILLS NETWORK

Falcon 9 Landing Predictor:
Cutting Costs with Precision

Daria Voronkov
June 11, 2024

Outline



- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion

Executive Summary

Project Overview: Predicting Falcon 9 Landing Success

- **Objective:** Predict successful landings of SpaceX's Falcon 9 first stage.
- **Cost Efficiency:** SpaceX rockets cost $62M vs. $165M for others, due to reusable first stages.

Methodology:

- **Data Collection:**
  - API integration and web scraping.
  - Data wrangling for restructuring.
- **EDA & Visualization:**
  - Uncovered insights, patterns, and trends.
- **Data Preprocessing:**
  - Feature engineering.
- **Predictive Modeling:**
  - Applied four ML algorithms.
  - Achieved ~83.33% accuracy.

Introduction

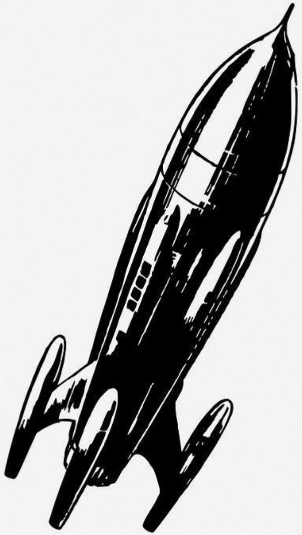
Project Background and Context

- **Commercial Space Age:**
  - SpaceX revolutionizes space travel, making it cost-effective and accessible.
  - Key innovation: reusable first stage of Falcon 9 rockets, reducing launch costs.
- **Project Aim:**
  - Develop a predictive model for Falcon 9 first stage landing success.
  - Improve launch cost estimation and offer competitive insights for new industry entrants.
- **Problems to Solve:**
  - Predict the successful landing of SpaceX Falcon 9's first stage.
  - Utilize historical launch data to develop a robust predictive model.
  - Help Space Y estimate launch costs accurately.
  - Enhance Space Y's competitive strategies against SpaceX.

Section 1

METHODOLOGY

Methodology



Executive Summary

1. **Data Collection Methodology:**
   - **Sources:**
     - SpaceX API
     - Web scraping from Wikipedia
   - **Processing:**
     - Data wrangling and cleaning
     - Conversion to tabular format using pandas
     - Addressing missing values
2. **Data Analysis:**
   - **Exploratory Data Analysis (EDA):**
     - Visualization and SQL queries
   - **Interactive Visual Analytics:**
     - Folium and Plotly Dash
3. **Predictive Analysis:**
   - **Steps:**
     - Data Preparation
     - Model Selection
     - Model Building
     - Hyperparameter Tuning
     - Model Evaluation
     - Model Interpretation
     - Model Deployment

Data Collection

Data collection was executed using various methods:

1. **Spacex API**
    - A GET request was sent to the SpaceX API to gather data.
    - The response content was decoded as JSON using the .json() function call and then converted into a pandas DataFrame using .json_normalize().
    - The data was cleaned by checking for and filling in missing values as necessary.

GET Request → Decode JSON, Pandas Dataframe → Data Cleaning → Save/Export Cleaned Data

2. Web Scraping from Wikipedia

- Additionally, web scraping was performed on Wikipedia for Falcon 9 launch records using BeautifulSoup.
- The goal was to extract the launch records as an HTML table, parse the table, and convert it to a pandas DataFrame for future analysis.

Web Scrape via BeautifulSoup → Extract HTML table → Parse HTML table → Convert to Pandas DataFrame

Data Collection – SpaceX API

- Used get request to the SpaceX API to collect data, clean the requested data and did some basic data wrangling and formatting.

- GitHub URL: [Data Collection via Spacex API_Notebook]

Data Collection – Scraping

- Employed web scraping via BeautifulSoup to gather Falcon 9 launch records from a website.
- Parsed the retrieved table.
- Converted the parsed data into a pandas dataframe.

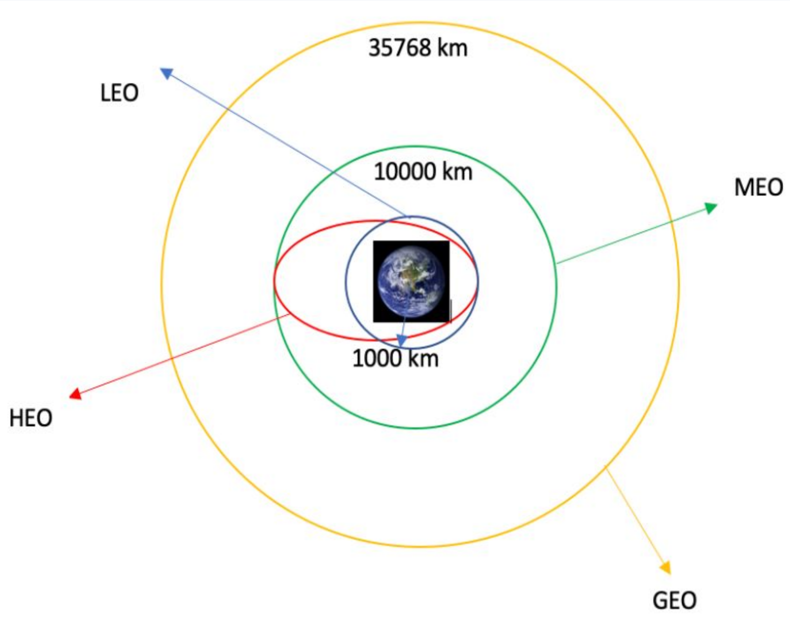
Wikipedia URL → Request URL page via HTTP Get method → BeautifulSoup Object → Extract columns names from HTML table → Create DataFrame → Export to CSV

- GitHub URL: Data Scraping Notebook

Data Wrangling

Process followed:

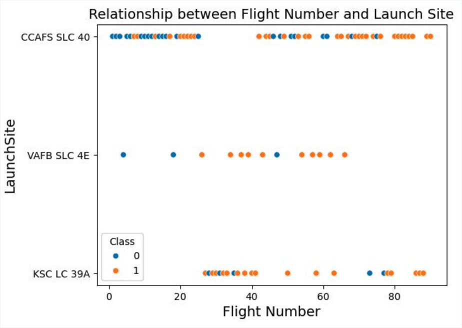
- Conducted exploratory data analysis to identify training labels.
- Computed the frequency of launches at individual sites and the occurrence of each orbit.
- Generated landing outcome labels based on the outcome column.
- Exported the results to a CSV file.

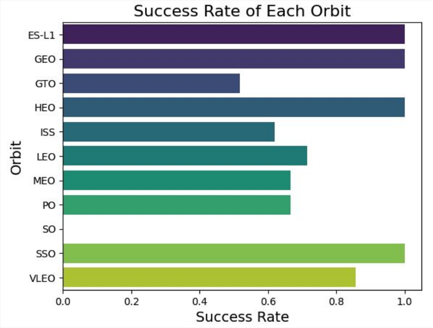


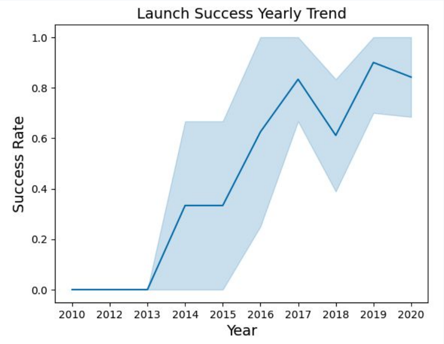
GitHub URL: [Data Wrangling]

EDA with Data Visualization

- Explored the data by visualizing the relationship between flight number and launch Site, payload and launch site, success rate of each orbit type, flight number and orbit type, the launch success







GitHub URL: [EDA w/Data Visualization]

EDA with SQL

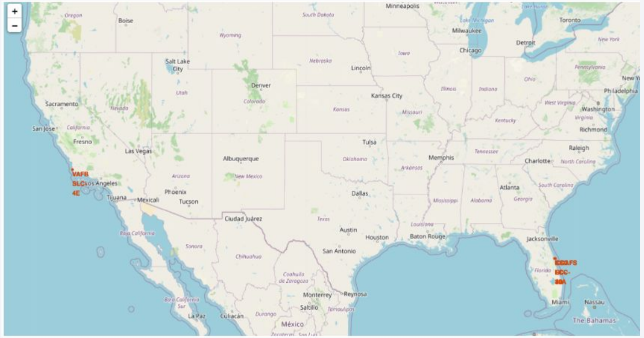
- Imported the SpaceX dataset into a PostgreSQL database directly within the Jupyter notebook.
- Utilized SQL for exploratory data analysis to gain insights from the dataset.
- Constructed **queries** to extract information such as:
  - *Unique launch site names in the space mission.*
  - *Total payload mass carried by boosters launched by NASA (CRS).*
  - *Average payload mass carried by booster version F9 v1.1.*
  - *Total number of successful and failed mission outcomes.*
  - *Failed landing outcomes on drone ships, including their booster version and launch site names.*

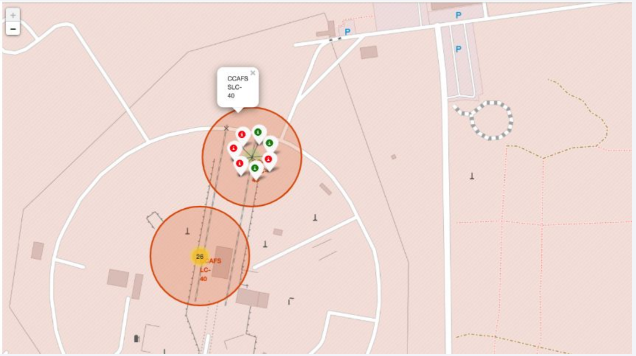
GitHub URL: [EDA with SQL]

Build an Interactive Map with Folium

- Tagged all launch sites and incorporated map elements such as markers, circles, and lines to indicate launch success or failure on the Folium map.
- Assigned launch outcomes (failure or success) to classes 0 and 1, where 0 represents failure and 1 represents success.
- Leveraged color-coded marker clusters to discern launch sites with notably high success rates.
- Computed distances between launch sites and nearby features, addressing **queries** such as:

    - *Proximity of launch sites to railways, highways, and coastlines.*
    - *Maintained distance between launch sites and urban areas.*



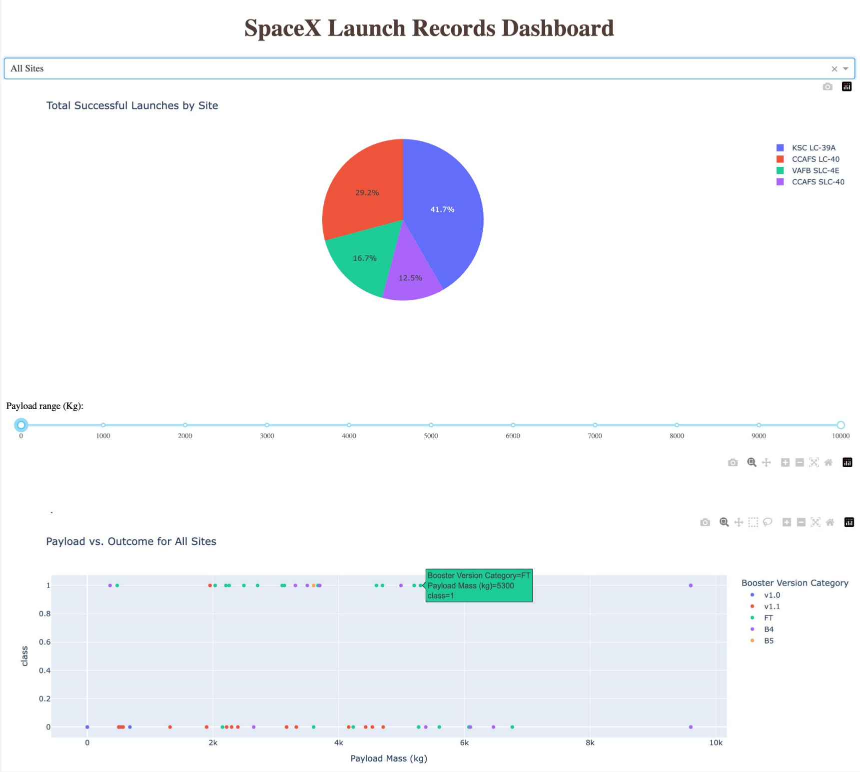


GitHub URL: [Interactive Map with Folium]

Build a Dashboard with Plotly Dash

- Developed an interactive dashboard using Plotly Dash.
- Illustrated total launches by specific sites through pie charts.
- Visualized the relationship between Outcome and Payload Mass (Kg) for various booster versions using scatter graphs.

GitHub URL: [Plotly Dash Notebook]



Predictive Analysis (Classification)

- Loaded data with numpy and pandas, performed data transformation, and split it into training and testing sets.
- Constructed various machine learning models and tuned hyperparameters using GridSearchCV.
- Evaluated model performance using accuracy, enhancing it through feature engineering and algorithm tuning.
- Identified the **best performing classification model.**

```
Logistic Regression Test Accuracy:  0.8333333333333334
SVM Test Accuracy:  0.8333333333333334
Decision Tree Test Accuracy:  0.7222222222222222
KNN Test Accuracy:  0.8333333333333334
Best performing model test accuracy:  0.8333333333333334
```

Load Data → Transform Data → Split Data → Build Models → Tune Hyperparameters → Evaluate Accuracy → Improve Model → Select Best Model

GitHub URL: Predictive Analysis (Classification)

Results

- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

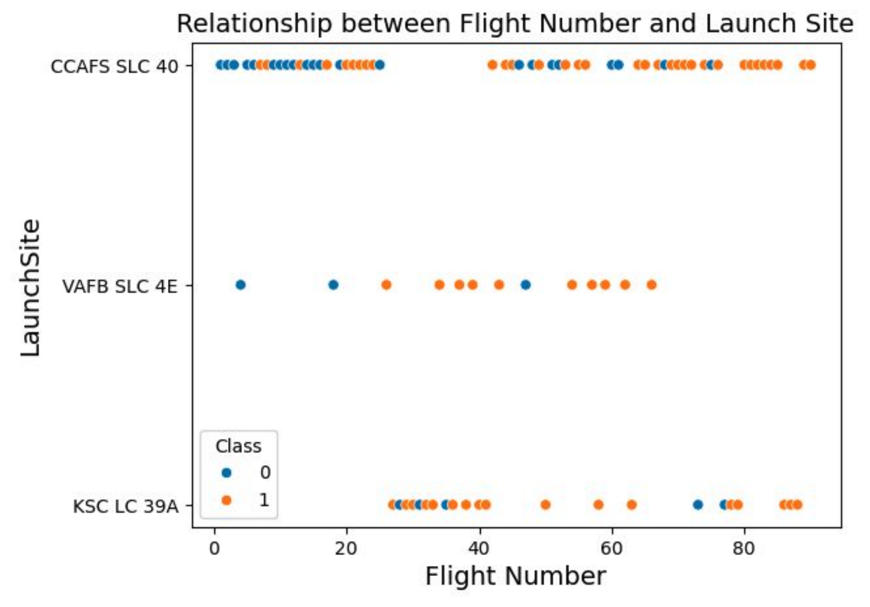
Section 2

INSIGHTS DRAWN FROM EDA

Flight Number vs. Launch Site

Observations on Falcon 9 Launches:

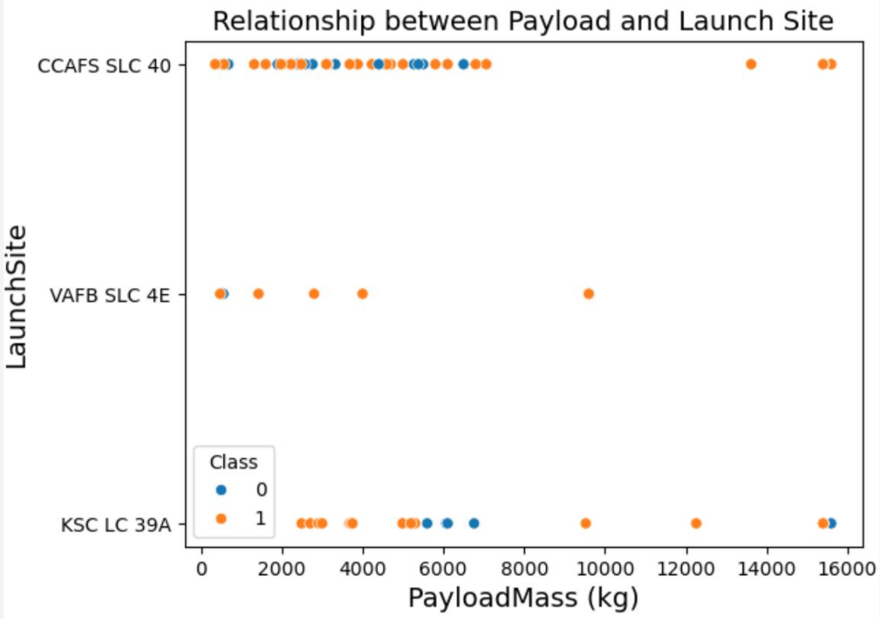
- **CCAFS LC-40:**
  - is the most used launch site
  - concentrates most of failures, particularly in the early stage of Falcon 9 project
  - success rate is around 60%
  - would be considered most “risky”
- **VAFB SLC-4E:**
  - least used launch site
  - success rate is around 77%
- **KSC LC 39A**
  - success rate around 77%
- **Pattern Observed:**
  - Higher flight numbers at a launch site correlate with higher success rates.



Payload vs. Launch Site

Observations on Falcon 9 Launches

- **Payload Distribution:**
  - Heavy payloads (>10,000 kg) are only sent to low or medium Earth orbits (LEO/MEO).
  - No heavy payload launches (>10,000 kg) from VAFB-SLC launch site.
- **Patterns Observed:**
  - Lower percentage of failures for heavier payloads.
  - Indicating that low orbits are less risky for mission success and booster recovery.



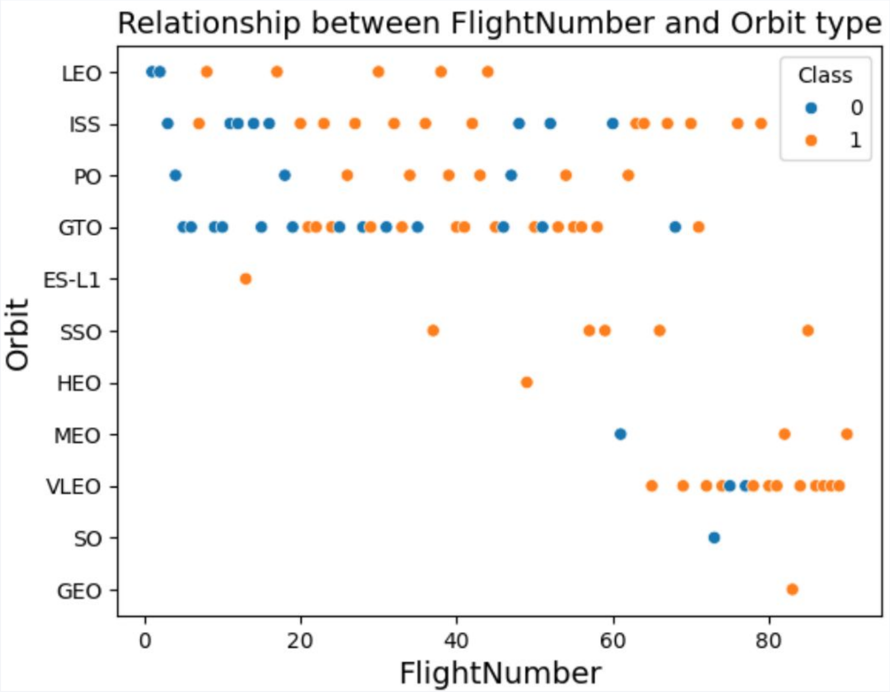
Flight Number vs. Orbit Type

To be noted:

- Results for GEO, SO, HEO, ESL-1, and MEO are disregarded due to their insignificant flight count.
- PO, SSO, ISS, VLEO are categorized as low orbits.
- GTO serves as a transfer orbit to GEO.

Observations:

- GTO missions appear to carry higher risk.
- Low orbits are associated with lower risk levels.



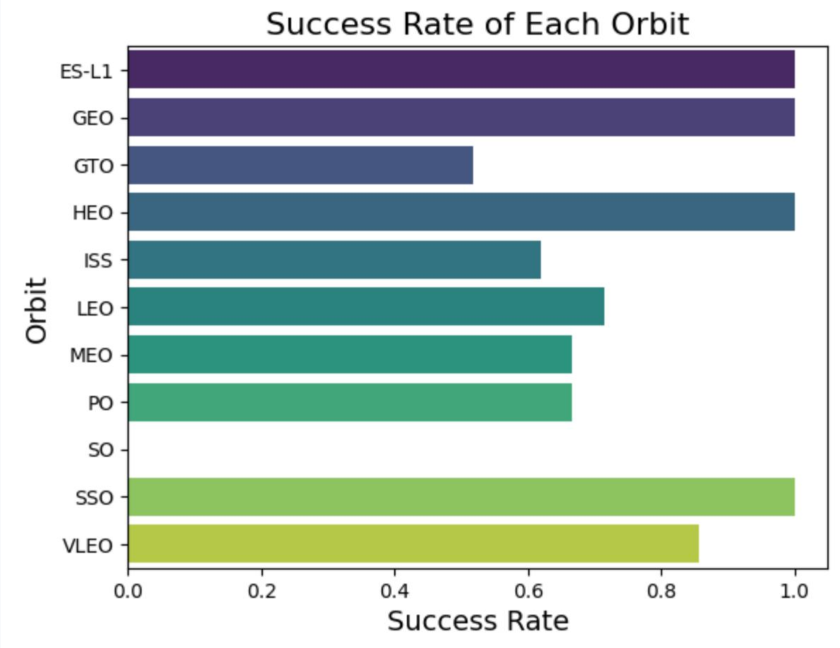
Success Rate vs. Orbit Type

To be noted:

- GTO (Geostationary Transfer Orbit) serves as a transition orbit to GEO (Geostationary Orbit). Payload thrusters, such as satellites, complete the orbital entry phase.
- Results for GEO, SO, HEO, ESL-1, and MEO are disregarded, as the number of flights holds insignificant value.

Observations:

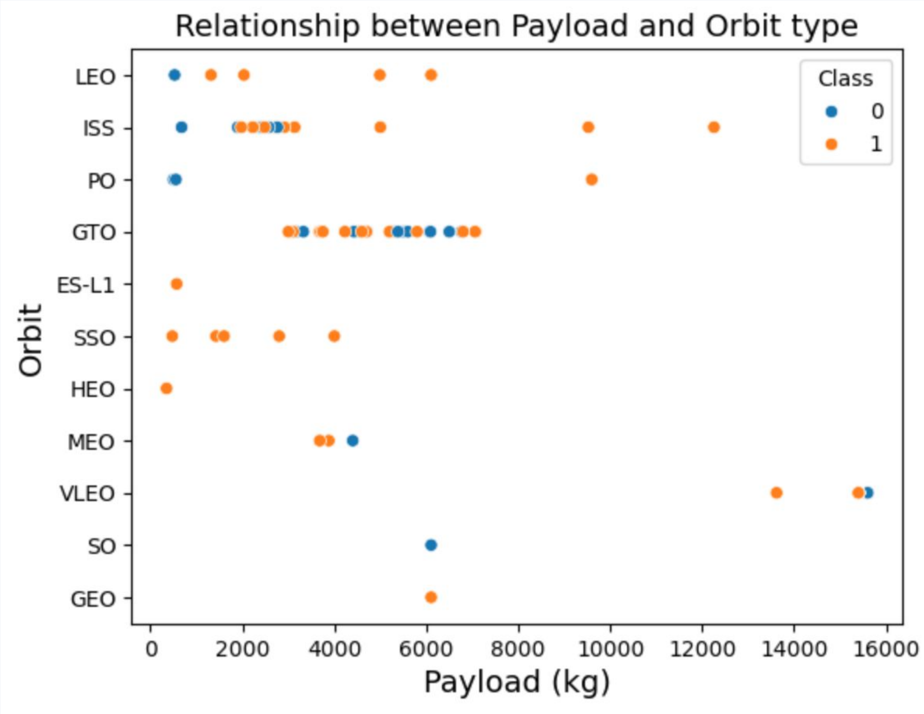
  - **GTO** exhibits the lowest success rate
  - **SSO** (Sun-Synchronous Orbit), or polar low orbit, boasts the highest success rate.
- Success probability hinges on both payload mass and orbit characteristics, with additional readily available information.



Payload vs. Orbit Type

Observations:

- **Optimal Success Rates**: Low orbit shows high success rates, with exceptions like ISS, especially with low payload masses.
- **Payload Impact**: Heavy payloads see higher success rates in Polar, LEO, and ISS orbits. GTO success is challenging to discern due to mixed outcomes.
- **ISS Analysis**: Early Falcon 9 failures (5/8) correlate with lower reliability.
- **GTO**: Risky regardless of payload mass; consistent success rate distribution observed.
- **Falcon 9 Reliability**: While the reliability of Falcon 9 shows improvement over time, challenges persist, particularly evident in post-GTO launch vehicle recovery attempts.



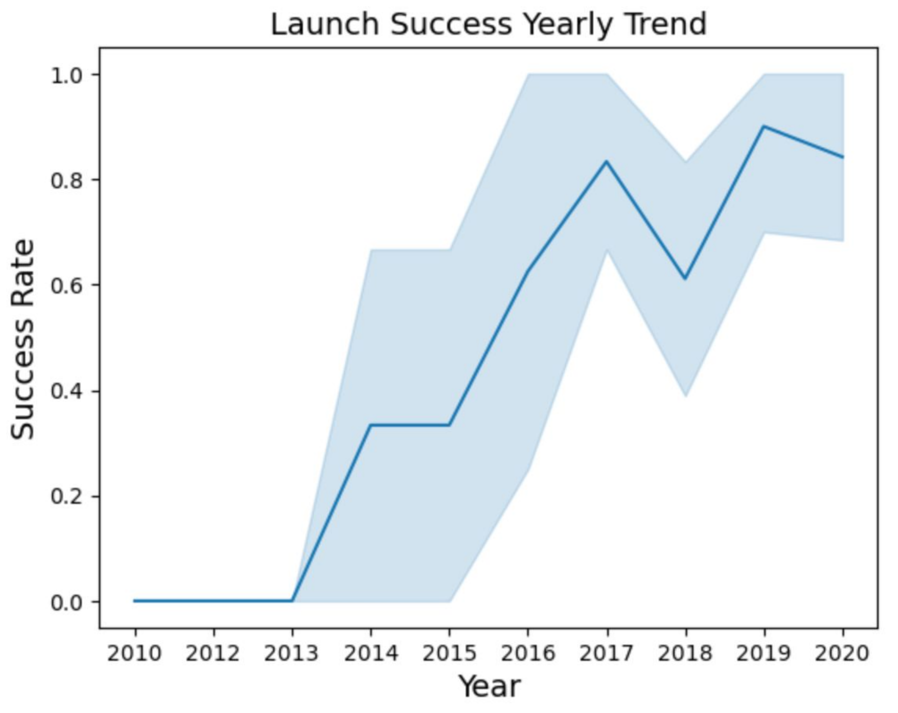
Launch Success Yearly Trend

Factors Affecting Success Rate:

- Success rate, defined as the successful recovery of the Falcon 9 booster, depends on:
  - Payload mass
  - Orbit
  - Other factors to be explored further.

Falcon 9 Booster Recovery Success Rate:

- The reliability of the Falcon 9 rocket has significantly increased over time.
- Since 2013, the success rate has been steadily rising.
- In 2018, there was a temporary drop in the success rate.
- The success rate rebounded and continued to increase in 2019.



All Launch Site Names and Records

4 distinct launch sites

- To analyze launch sites, we first compiled a list of all launch site names and retrieved some launch records using SQL queries.

```
%sql SELECT DISTINCT Launch_Site FROM SPACEXTBL;
```

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

Launch_Site
CCAFS LC-40
VAFB SLC-4E
KSC LC-39A
CCAFS SLC-40

5 records where launch sites begin with `CCA`

```
%sql SELECT * FROM SPACEXTBL 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
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success
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
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success

Total Payload Mass

- **Total payload** carried by boosters from NASA equals to **45,596 kg.**

```
%sql SELECT Customer, SUM(PAYLOAD_MASS__KG_) as "TOTAL_PAYLOAD_KG_" FROM SPACEXTBL WHERE Customer = 'NASA (CRS)';
```

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

Customer	TOTAL_PAYLOAD_KG_
NASA (CRS)	45596

Average Payload Mass by F9 v1.1

- Average payload mass carried by booster version F9 v1.1 equals to **2,534.66 kg**

Display average payload mass carried by booster version F9 v1.1

```
%sql SELECT Booster_Version, AVG(PAYLOAD_MASS__KG_) as "AVG_PAYLOAD_KG_" FROM SPACEXTBL WHERE Booster_Version LIKE "F9 v1.1%";
```

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

Booster_Version	AVG_PAYLOAD_KG_
F9 v1.1 B1003	2534.6666666666665

First Successful Ground Landing Date

- **December 22, 2015 -** is the date of the first successful landing outcome on ground pad

```
%sql SELECT MIN(Date), Landing_Outcome FROM SPACEXTBL WHERE Landing_outcome LIKE "Success%";
```

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

MIN(Date)	Landing_Outcome
2015-12-22	Success (ground pad)

Successful Drone Ship Landing with Payload between 4000 and 6000

- Below are the names of **boosters** which:
  - have **successfully landed** on drone ship
  - had **payload mass** greater than 4000 but less than 6000

```
%sql SELECT Booster_Version, PAYLOAD_MASS__KG_, Landing_Outcome FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ > 4000 AND PAYLOAD_MASS__KG_<6000 AND Landing_Outcome LIKE "Success (drone ship)";
```

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

Booster_Version	PAYLOAD_MASS__KG_	Landing_Outcome
F9 FT B1022	4696	Success (drone ship)
F9 FT B1026	4600	Success (drone ship)
F9 FT B1021.2	5300	Success (drone ship)
F9 FT B1031.2	5200	Success (drone ship)

Total Number of Successful and Failure Mission Outcomes

- Below is the total number of successful and failure mission outcomes

```
%sql SELECT Mission_Outcome, COUNT(*) AS "Total Mission Outcomes" FROM SPACEXTBL GROUP BY Mission_Outcome;
```

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

Mission_Outcome	Total Mission Outcomes
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- Below are the **names of the booster which have carried the maximum payload mass, which is 15,600 kg**

```
%sql SELECT Booster_Version, PAYLOAD_MASS__KG_ FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_)FROM SPACEXTBL);
```

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

Booster_Version	PAYLOAD_MASS__KG_
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

2015 Launch Records

- Below are listed:
    - failed landing outcomes in drone ship
    - their booster versions
    - launch site names for in year 2015

```
%sql SELECT substr(Date, 6, 2) AS Month,Landing_Outcome,Booster_Version,Launch_Site FROM SPACEXTBL WHERE substr(D
```

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

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

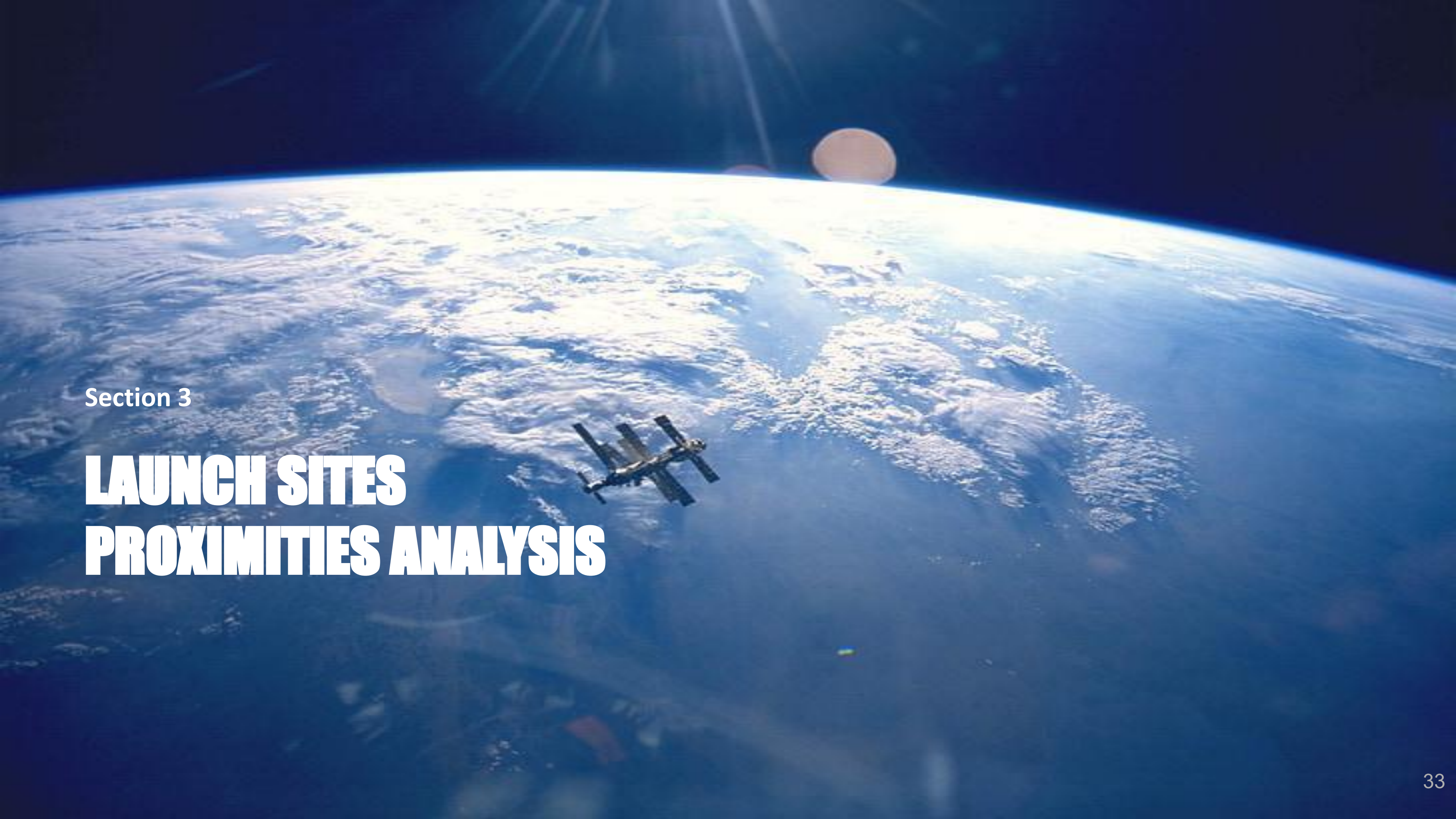
Key Insights:

- **Highest Frequency**: "No attempt" with 10 instances.
- **Drone Ship Landings**: Equal success and failure rates (5 each).
- **Ground Pad Success**: Achieved 3 successful landings.
- **Ocean Landings**: Controlled (3) and uncontrolled (2) landings were noted.
- **Parachute Failures**: Documented twice.
- **Least Occurrence**: "Precluded (drone ship)" with 1 instance.

```
%sql SELECT Landing_Outcome, COUNT(*) AS Outcome_Count FROM SPACEXTBL WHERE Date BETWEEN '2010-06-04' AND '2017-0
```

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

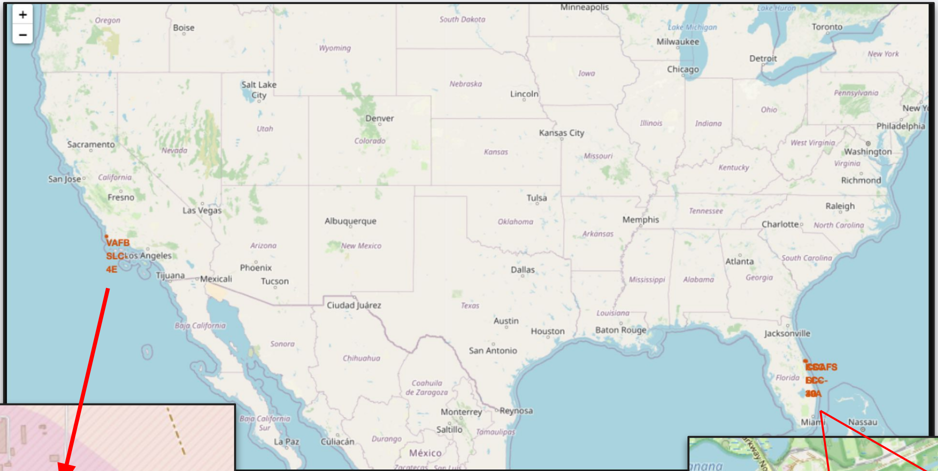
Landing_Outcome	Outcome_Count
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

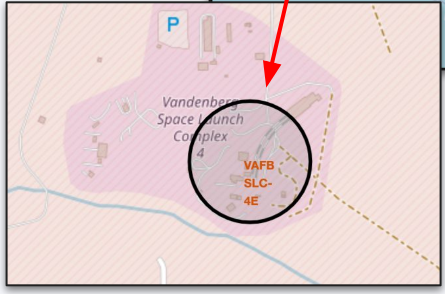


Section 3

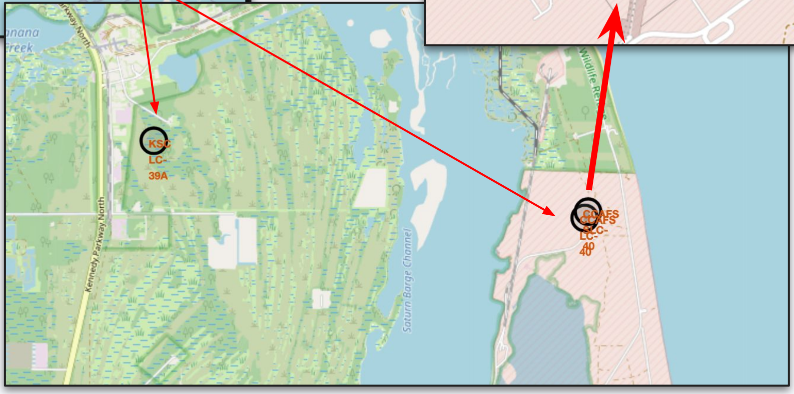
LAUNCH SITES PROXIMITIES ANALYSIS

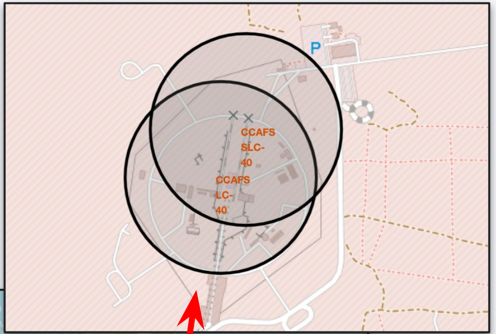
SpaceX Launch Sites





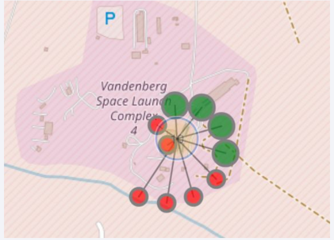
1. **VAFB SLC-4E: Vandenberg Space Launch Complex 4 (CA)**
2. **KSC-LC29A: Kennedy Space Center - Merritt Island (FL)**
3. **CCAFS-LC40: Cape Canaveral Launch Complex 40 (FL)**
4. **CCAF-SLC40: Cape Canaveral Space Launch Complex 40(FL)**



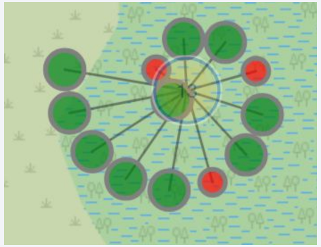


SpaceX Launch Outcomes

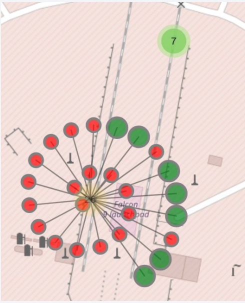
VAFB SLC-4E: Vandenberg Space Launch Complex 4 (CA)



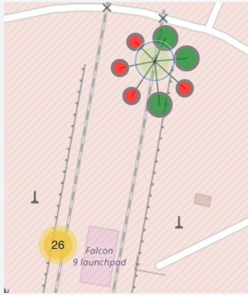
KSC-LC29A: Kennedy Space Center - Merritt Island (FL)



CCAFS-LC40: Cape Canaveral Launch Complex 40 (FL)



CCAF-SLC40: Cape Canaveral Space Launch Complex 40(FL)



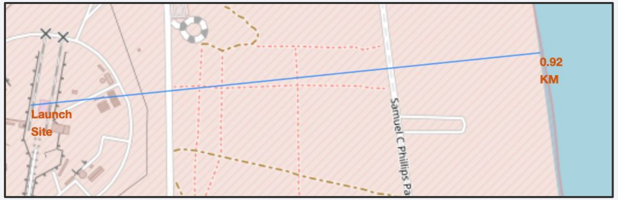
	Launch Site	Failures	Successes
0	CCAFS LC-40	19	7
1	CCAFS SLC-40	4	3
2	KSC LC-39A	3	10
3	VAFB SLC-4E	6	4

- The table summarizes the number of successes and failures for each launch site
- The map visualizes these performance metrics with markers:
  - Green markers denote successful launches.
  - Red markers denote failed launches.

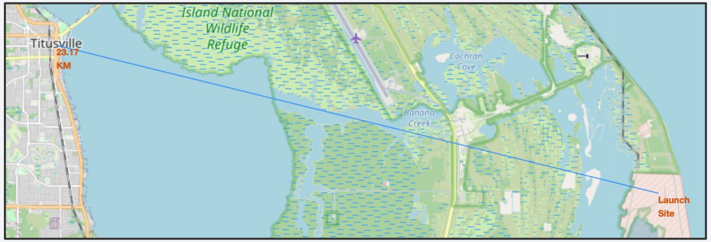
Geospatial Analysis of Launch Site Proximities

Distance from CCAFS-LC40:

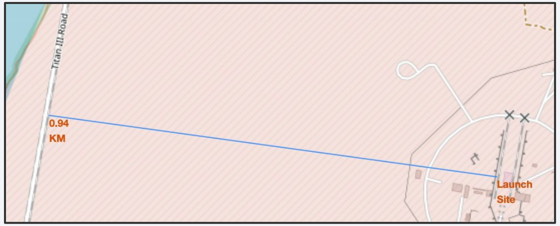
- Closest Coast: **920 m**



- Closest city - Titusville : **23.17 km**



- Closest Railway - Tran III Road: **940 m**



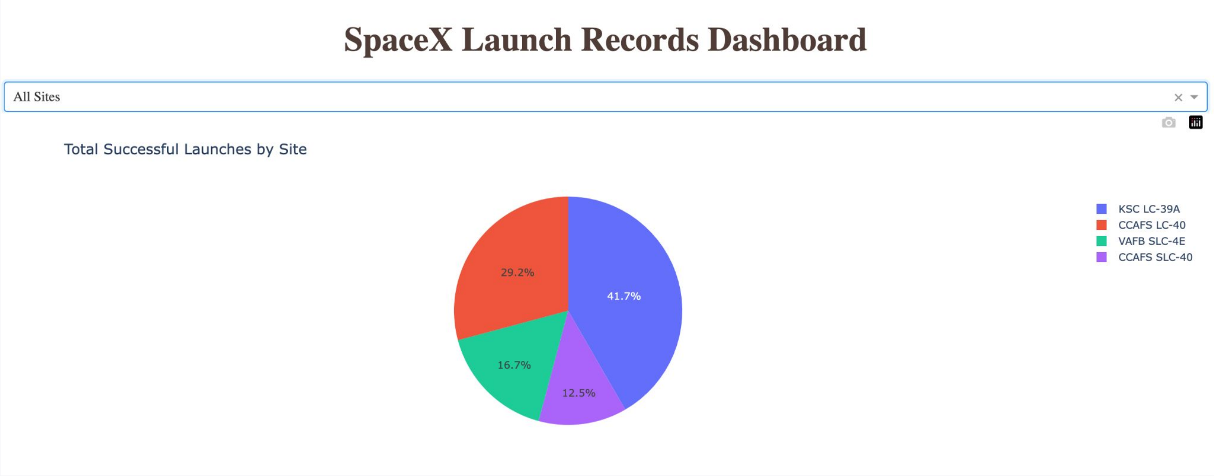
Section 4

BUILD A DASHBOARD WITH PLOTLY DASH

SpaceX Falcon 9: Launch success count for all sites

- The dashboard provides an interactive platform for visualizing and analyzing successful Falcon launches, consolidating scattered charts into one cohesive view.
- It allows users to easily identify the launch site with the highest success rate, which is the **Kennedy Space Center in Florida**.

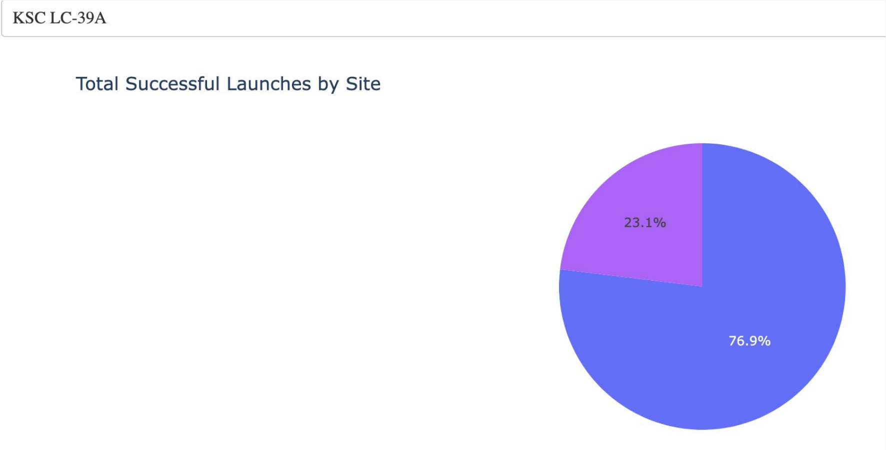
Launch Site	class	
CCAFS LC-40	0	19
	1	7
CCAFS SLC-40	0	4
	1	3
KSC LC-39A	0	3
	1	10
VAFB SLC-4E	0	6
	1	4



SpaceX Falcon 9: Launch Site with the Highest Success Rate

KSC LC-39A

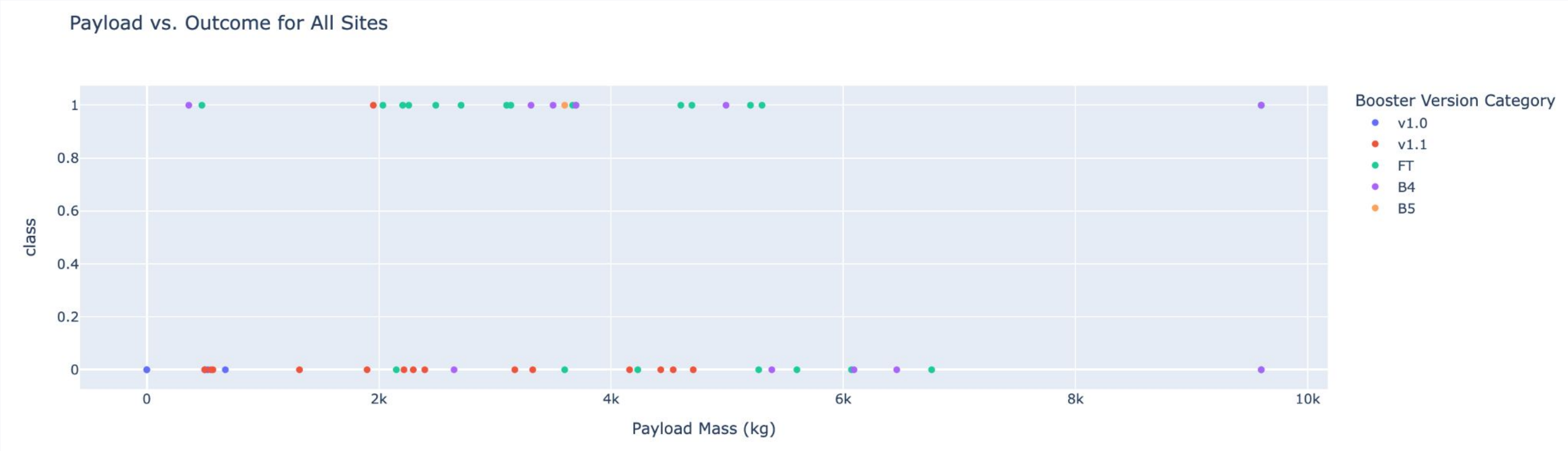
- Located at Kennedy Space Center in Florida.
- Conducted 13 flights, with 10 successful missions.
- Heavy payloads are considered high risk.
- Success rates are consistent regardless of booster versions for payloads <5500kg.
- Boosters B5 and FT show significant reusability, suggesting reliability comparable to first-time launchers based on available data.



Launch Outcome vs. Payload mass (All Sites)

Falcon 9 Launcher Evolution

- **V1.0 and V1.1:** Early versions with lower reliability.
- **Landing Legs:** Introduced on Falcon 9 V1.1 but never successfully landed intact; phased out in 2015.
- **FT (Full Thrust):** Next-generation version with the highest success rate for payloads under 6 tons.
  - Many FT flights utilize reused launchers, demonstrating good reliability.
  - Heavy payloads are considered high risk with this configuration.



Section 5

PREDICTIVE ANALYSIS (CLASSIFICATION)

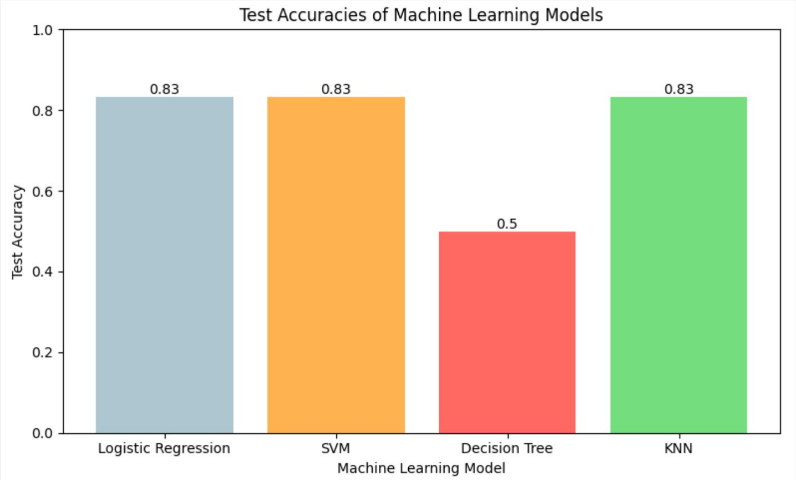
Classification Accuracy

Consistent Test Accuracy:

- Logistic Regression, SVM, and KNN models demonstrate identical test accuracies of 83.33%.
- Indicates robust predictive performance across different methodologies.

Challenges with Decision Tree:

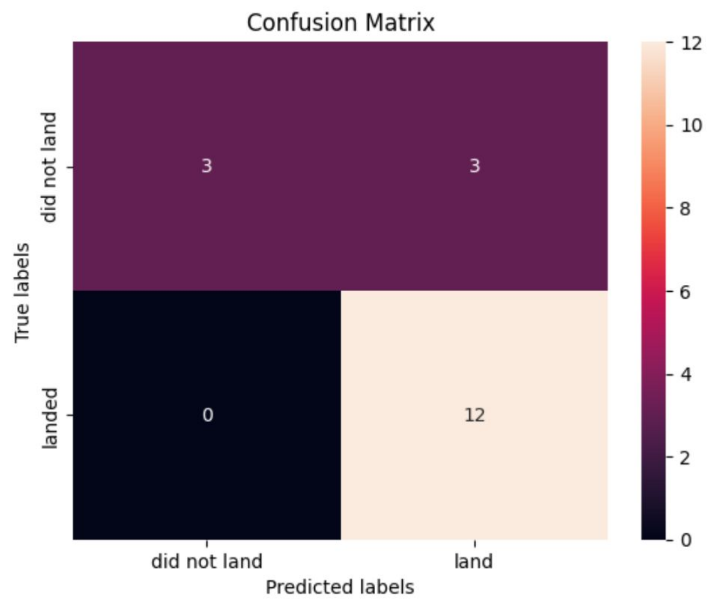
- Decision Tree model exhibits a lower test accuracy of 50%.
- Potential issues include suboptimal parameter tuning or variability in model performance.



Confusion Matrix - Logistic Regression

Confusion Matrix Insights:

- Logistic regression demonstrates capability in distinguishing between different classes.
- The primary issue observed is an elevated rate of false positives.



Conclusions

Booster Recovery Analysis:

- Success is defined as the successful recovery of Falcon 9 boosters.
- Utilized Falcon 9 and Wiki data to investigate factors influencing booster recovery: Payload, Orbit, Booster version, Launch sites.
- Employed supervised classification to predict launch outcomes with an accuracy of 84%.
- Falcon 9 booster recovery process is inherently riskier than traditional launchers, yet its success rate:
  - Shows improvement over time.
  - Stands at 66%, remaining competitive in cost per kg.
- SpaceX's GEO/GTO flights are identified as higher risk.
- Failures in these missions may relate to launch energy, vibrations affecting electronics, and control systems.
- SpaceX could enhance GEO/GTO success by exploring equatorial launch concepts like "Sea Launch."
- Recent successes with FT boosters landing on drone ships indicate potential for such innovations.
- Despite recent setbacks in GTO/GEO recovery, SpaceX maintains a competitive edge.
- Future competitiveness may be significantly bolstered by Starship's potential success rates.


THANK YOU!
SPACEX