

## Winning Space Race with Data Science

Reeva Gadhavi Jan 26, 2025



### **Outline**

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

### **Executive Summary**

In today's fast-paced and competitive space industry, SpaceX has completely changed the game with its reusable launchers and boosters, like the Falcon 9 and Falcon Heavy. The big win here? A massive cut in the cost per kilogram of payload, making space more accessible than ever before. Of course, it's not all smooth sailing. Compared to traditional launch vehicles like the Soyuz or Ariane-5, reliability is still a bit of a challenge. For SpaceX, though, success isn't just about getting the payload into orbit—it's also about bringing the booster back safely, which keeps their costs low and their edge sharp.

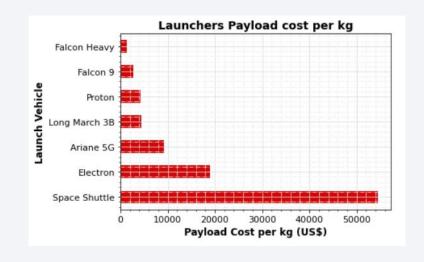
Whether a Falcon 9 booster lands successfully depends on factors like:

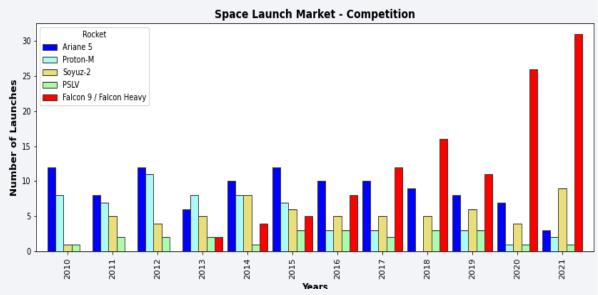
- The orbit it's headed to
- How heavy the payload is
- Which version of the booster is being used
- •Where the launch happens

Using these details, this report showcases a Machine Learning model that predicts booster recovery with nearly 94% accuracy. It's a step closer to making reusable rockets the norm in space exploration!

### Introduction - Business understanding

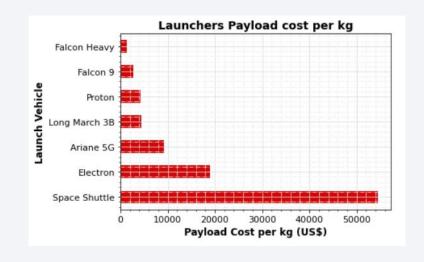
- The commercial rocket launch market is highly competitive. From 2000 to 2017, Ariane 5 (ESA - Arianespace) led with low costs, high payload capacity, and reliability.
- China, India, and Russia offer cheaper, but less reliable options like LongMarch, Proton, Soyuz-2, and PSLV. New entrants like Blue Origin are still developing.
- Since 2017, SpaceX's Falcon 9 and Falcon
  Heavy have dominated with lower cost per
  kg, thanks to reusable boosters.
- Competition will heat up with the upcoming Ariane 6 and SpaceX Starship.

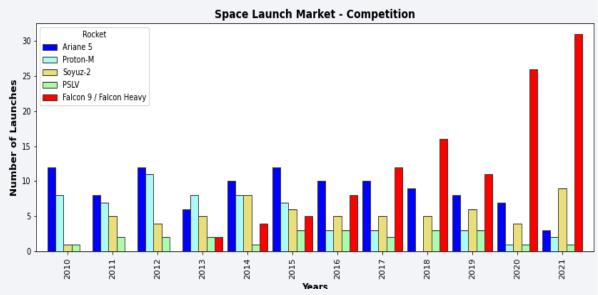




### Introduction - Business understanding

- The commercial rocket launch market is highly competitive. From 2000 to 2017, Ariane 5 (ESA - Arianespace) led with low costs, high payload capacity, and reliability.
- China, India, and Russia offer cheaper, but less reliable options like LongMarch, Proton, Soyuz-2, and PSLV. New entrants like Blue Origin are still developing.
- Since 2017, SpaceX's Falcon 9 and Falcon
  Heavy have dominated with lower cost per
  kg, thanks to reusable boosters.
- Competition will heat up with the upcoming Ariane 6 and SpaceX Starship.







### Methodology

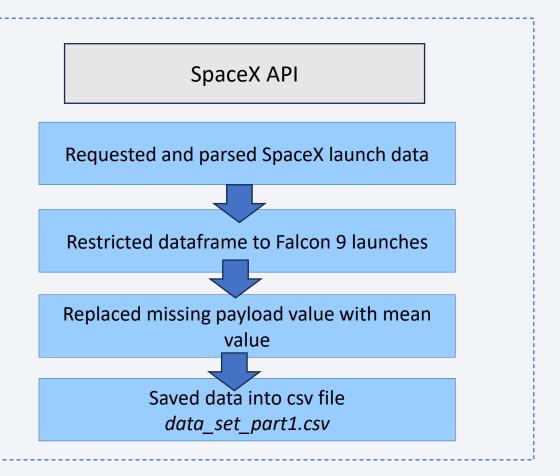
#### **Executive Summary**

- Data collection methodology:
  - I collected data from open database and Wikipedia (Falcon9, Ariane-5).
- Perform data wrangling
- 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 Collection - SpaceX API

Collected data from SpaceX REST API.
 First, I requested and parsed SpaceX launch data. Then, restricted the dataframe to Falcon 9 only. Replaced some missing payload value with mean value. And, at last, saved data into a csv file.

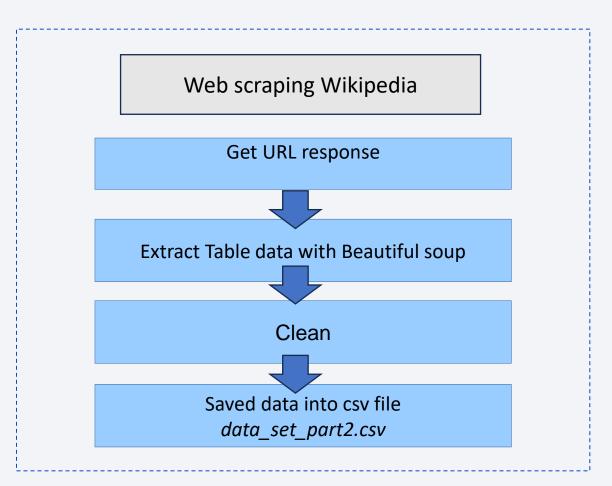
# GitHub URL: https://github.com/reeva111/Applied Data Science Capstone SpaceX/blob/main/jupyter-labs-spacex-data-collection-api.ipynb



### **Data Collection - Scraping**

- I collected data through web scraping in Wikipedia. First, got an URL response. Then, extracted table data with Beautiful soup, cleaned it. And, at last, saved data into a csv file.
- GitHub URL:

   https://github.com/reeva111
   /Applied Data Science Caps
   tone SpaceX/blob/main/jupy
   ter-labs-webscraping.ipynb



### **Data Wrangling**

#### Dataframe From SpaceX API

Identified missing values

Replaced missing "PayloadMass" value with mean value

SpaceX dataset
Further Data wrangling

#### Identified:

- null values for each feature
- numerical and categorical features

#### Calculated:

- Number of launches on each site
- number and occurrence of each orbit
- number and occurrence of missions outcome per orbit type

Created a set of 1 stage booster landing outcomes

O True ASDS: successful landing on a drone ship

1 None None: failure to land

2 True RTLS: successful landing to a ground pad

3 False ASDS: failed landing on a drone ship

4 True Ocean: successful landing, specific region of the ocean

5 False Ocean: failed landing, specific region of the ocean

6 None ASDS: failure to land

7 False RTLS: failed landing to a ground pad

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Class
49	50	2018-05-11	Falcon 9	3750.00	GTO	KSC LC 39A	True ASDS	1
47	48	2018-04-02	Falcon 9	2760.00	ISS	CCAFS SLC 40	None None	0
50	51	2018-06-04	Falcon 9	5383.85	GTO	CCAFS SLC 40	None None	0
44	45	2018-01-31	Falcon 9	4230.00	GTO	CCAFS SLC 40	True Ocean	1
11	12	2015-01-10	Falcon 9	2395.00	ISS	CCAFS SLC 40	False ASDS	0

Created a Training label: 'Class'

**Class = 0: booster landing failure** 

**Class = 1: booster landing success** 

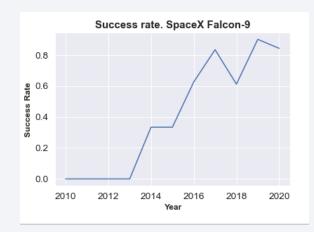
### **EDA** with Data Visualization

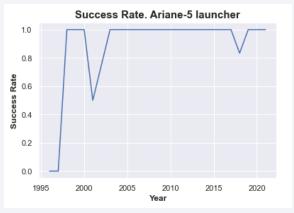
#### Methodology:

- Utilized Matplotlib and Seaborn for visual exploration through diverse charts.
- Categorical outcomes labeled as "0" (Unsuccessful) and "1" (Successful) to interpret results.

#### Visual Insights:

- Payload Mass vs. Flight Sequence Observing performance variations over time.
- Launch Site vs. Flight Sequence Identifying site-specific trends.
- Launch Site vs. Payload Mass Assessing mass handling capabilities.
- Orbit Type vs. Flight Sequence Tracking orbital performance across missions.
- Orbit Type vs. Payload Mass Exploring mass distribution patterns by orbit.





### EDA with SQL

- Using bullet point format, summarize the SQL queries you performed
- Add the GitHub URL of your completed EDA with SQL notebook, as an external reference and peer-review purpose

### Build an Interactive Map with Folium

#### **Geospatial Analysis with Folium:**

- •Utilized an interactive map to study SpaceX launch sites across the United States.
- •Mapped sites in Florida and California, integrating precise latitude and longitude data.
- Key Insights from Location Analysis:
- Visualized successful and unsuccessful launches at each site.
- •Examined the influence of site location and surrounding features on launch performance.

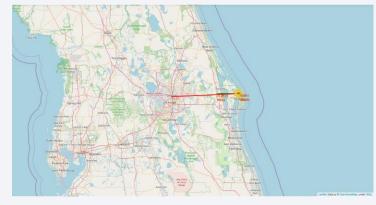
#### Proximity Calculations for Cape Canaveral Site (CCAFS\_SLC40):

Nearest coastline – measured the distance to evaluate potential environmental factors.

Railway proximity – calculated the distance from the Florida East Coast Railway, a high-traffic corridor.

Highway access – determined the distance to Interstate I-95, a major transportation route.

Urban area impact – assessed the distance to Orlando, FL, to analyze possible urban influences.



CCAFS\_SLC40 in Cape Canaveral FL Coordinates: -80.577°, 28.563°



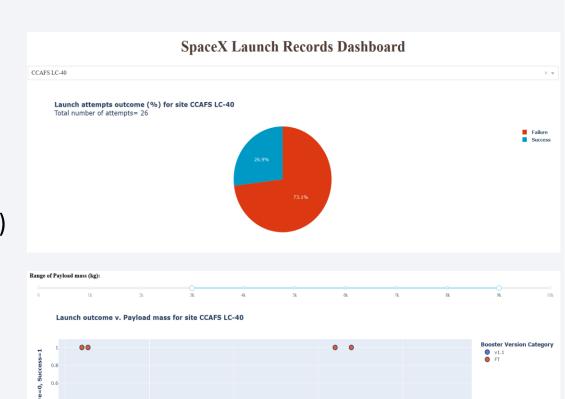
Ariane launch pad - Kourou in French Guiana Coordinates: -52.792°, 5.265° (~ Equator)

13

### Build a Dashboard with Plotly Dash

We developed an interactive dashboard using Plotly that provides a dynamic view of SpaceX launch data. The dashboard includes:-

- A dropdown menu to choose different launch sites
- Pie charts illustrating the distribution of successful launches
- Scatter plots showing launch sites, payload masses, and launch outcomes
- A range slider to filter launches based on payload mass (in kg) The analysis focuses on key insights, such as:-
- The launch site with the highest number of successful launches
- The location with the best success rate overall
- Payload ranges that are most and least successful
- The Falcon 9 booster version with the best performance across missions



### Predictive Analysis (Classification)

Imported Falcon9 launching data from .csv to Pandas dataframe **Supervised Classification models** Launcher landing Labels • Class 0: fail • Class 1: success Data normalization Split Data: Training, Testing Sets Shuffling option!

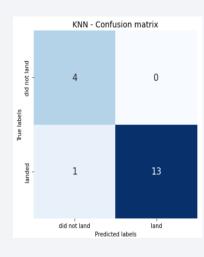
#### **Model Training**

**Logistic Regression** 

**Support Vector Machine (SVM)** 

**Decision Tree Classifier** 

K Nearest Neighbors Classifier



#### **Testing set**

Predictive analysis
Accuracy
Confusion matrix

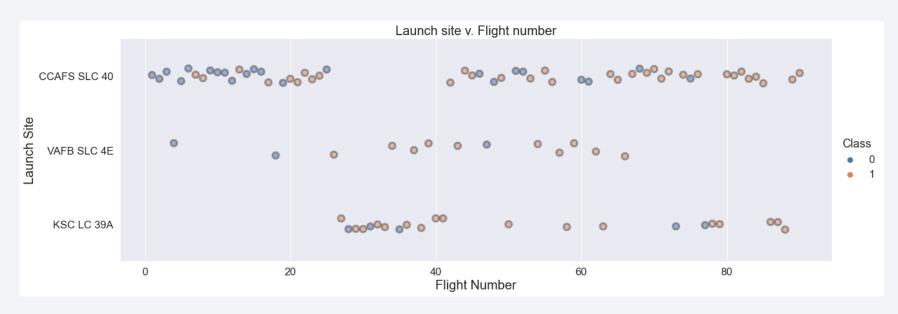
Selection of best model

### Results

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



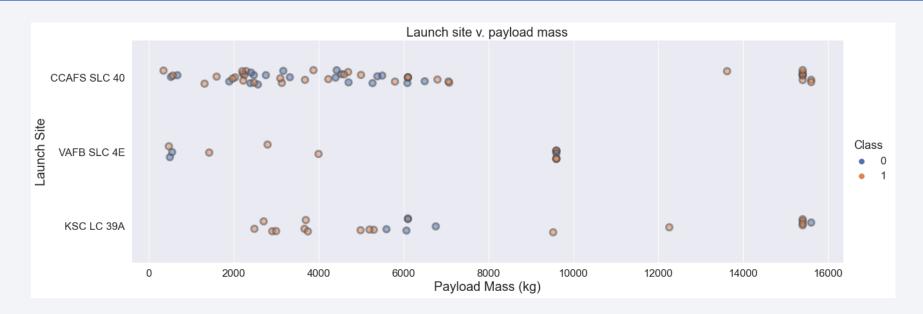
### Flight Number vs. Launch Site



#### The chart displays valuable info about:

- Chronology: flight numbers over time
- Launch Site Utilization: Number of flights per launch site
- Mission Outcome: Success/Failure per launch site
- Cape Canaveral CCAFS-SLC 40 is the most frequently used launch site.
- Failure rates are highest at CCAFS-SLC 40, especially during the early stages of the Falcon 9 project.
- Due to its southern location, CCAFS-SLC 40 may be preferred for high-risk GTO (Geostationary Transfer Orbit) and GEO 18
   (Geostationary Earth Orbit) missions.

### Payload vs. Launch Site



#### The chart brings additional info:

- Payload Mass Distribution: Payload mass per launch site
- Mission Outcomes: Success/Failure per payload mass
- Given Falcon 9's capabilities, heavy payloads (>10,000 kg) are primarily sent to low or medium orbits (LEO/MEO).
- The failure rate appears lower for heavy payloads, suggesting that low-orbit missions might be less risky for mission success and booster recovery.
- Light payloads are not exclusively launched to GTO/GEO, indicating varied deployment strategies.

### Success Rate vs. Orbit Type

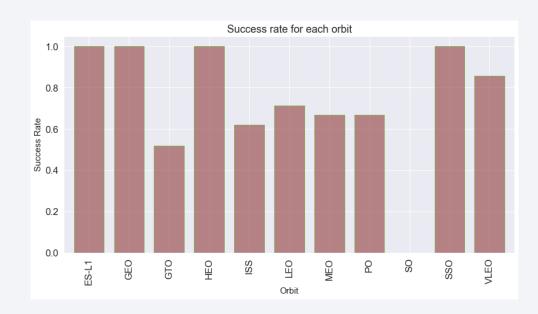
- •GTO is a transfer orbit to GEO. The payload's low-thrust engines (satellite) complete the orbiting phase.
- •Results for GEO, SO, HEO, ESL-1, and MEO are ignored due to an insignificant number of flights.
- •GTO has the lowest success rate, as suggested in the previous slide.
- •SSO (polar low orbit) has the highest success rate.

#### **Key Factors Influencing Success Rate:**

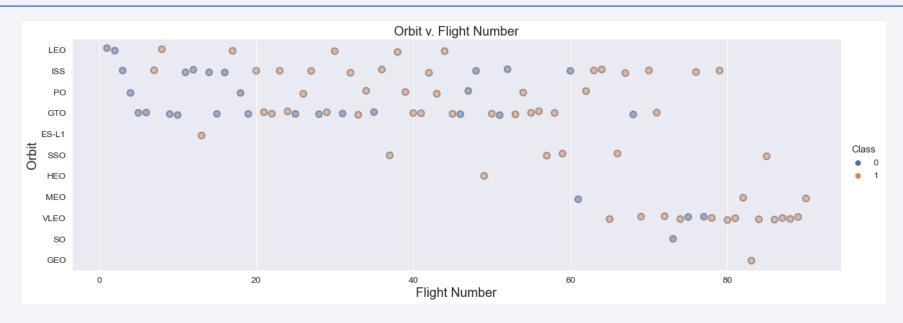
- Payload mass
- •Orbit

#### **Impact of Energy at Lift-Off:**

- •High energy at lift-off can cause strong noise and vibrations, potentially damaging satellites.
- •Vibrations may also affect booster electronics and inertial guidance systems, leading to booster recovery or landing failures.



### Flight Number vs. Orbit Type



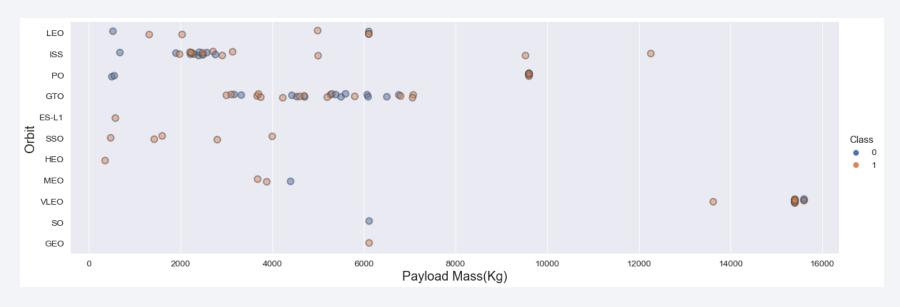
#### The chart provides additional insights, including:

- •The distribution of flights across various orbit types
- The success rate associated with each orbit category

For certain orbits like GEO, SO, HEO, ESL-1, and MEO, the number of launches is relatively small, making it difficult to draw solid conclusions about their success rates.

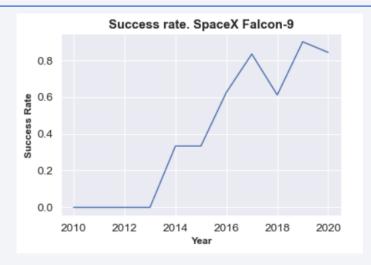
Lower orbits such as PO, SSO, ISS, and VLEO are more frequently used and generally involve less risk. GTO, being a transfer orbit to GEO, appears to have a higher risk compared to these lower orbits.

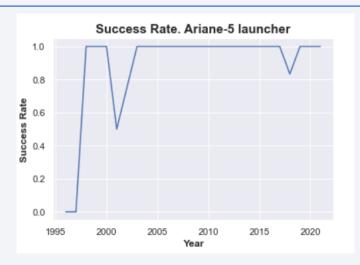
### Payload vs. Orbit Type



- Higher success rates are generally observed for lighter payloads in lower orbits, with the ISS being an exception.
- The ISS experienced most of its failures (5 out of 8) during the initial stages of the Falcon 9
  program when reliability was still developing.
- For GTO, the success rate appears relatively stable when payloads range between 2000 and 7500 kg.
- Regardless of payload mass, GTO remains a more challenging orbit, often impacting mission outcomes. While Falcon 9 has become more dependable over time, booster recovery failures still occasionally occur after GTO launches.

### Launch Success Yearly Trend





The reliability of the Falcon 9 rocket has shown remarkable improvement over time. Key factors influencing booster recovery success:

- •Payload mass: Heavier payloads can impact performance.
- •Orbit: The target orbit plays a significant role in recovery success.
- •Additional variables: Further analysis will explore other contributing factors.

#### Comparative insights:

- •Ariane 5: Demonstrates near-perfect reliability, with a success rate close to 100% across 82 launches since 2003.
- •Falcon 9: Achieves an average booster recovery success rate of 66%, a figure that ensures SpaceX's financial sustainability despite the challenges.

### All Launch Site Names

```
df_unique_launchsites=pd.read_sql_query("Select distinct Launch_Site from spacex_v11 ",conn)
print(df_unique_launchsites)

Launch_Site
0 CCAFS LC-40
1 VAFB SLC-4E
2 KSC LC-39A
3 CCAFS SLC-40
```

So, there are 4 distinct launch sites

### SpaceX: All launch sites



### Launch Site Names Begin with 'CCA'

df\_launchsites\_CCA5=pd.read\_sql\_query("Select \* from spacex\_v11 where Launch\_Site Like 'CCA%' Limit 5",conn)
df\_launchsites\_CCA5

id	Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS_KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
0 1	2010-04-06	0 days 18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
1 2	2010-08-12	0 days 15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
<b>2</b> 3	2012-05-22	0 days 07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
<b>3</b> 4	2012-10-08	0 days 00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
<b>4</b> 5	2013-03-01	0 days 15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

• Here are the 5 records where launch sites begin with `CCA'.

### Launch outcome v. Payload mass (all sites)

The initial Falcon 9 versions, v1.0 and v1.1, had relatively low reliability.

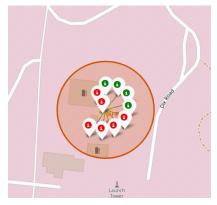
The v1.1 model introduced landing legs for the first time, though it never achieved a successful landing and was retired in 2015.

The Full Thrust (FT) variant followed as the next generation, demonstrating the highest success rates for payloads below 6 tons, even with drone ship landings (details on the next slide).

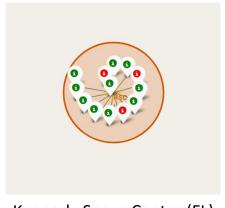
Many FT missions have utilized reused boosters, which have consistently shown strong performance. Missions carrying heavier payloads remain more challenging and inherently riskier.



### Falcon 9 Success/Failed launches for each site



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



Kennedy Space Center (FL) KSC LC 39A



Cape Canaveral (FL) CCAFS-LC40



Cape Canaveral (FL) CCAFS-SLC40

La	un	ch	S	ite	C	ass

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

**Table: Synthesis of launches outcomes** 

Class 0= failure
Class 1= success

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

- List the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000 kg.
- Recent Full Thrust (FT) boosters exhibit the highest success rate on drone ship landing.

### **Total Payload Mass**

Calculate the total payload carried by boosters from NASA (2 methods)

```
# For validation purposes... sum in df NASA CRS 'PAYLOAD MASS KG' column
df NASA CRS=pd.read sql query("Select * from spacex v11 where Customer='NASA (CRS)'",conn)
print(df NASA CRS.head(2))
print('----')
print('Total payload mass, customer= NASA (CRS):', df NASA CRS['PAYLOAD MASS KG '].sum(),' kg')
   id
                      Time (UTC) Booster Version Launch Site
                                                                  Payload \
            Date
0 4 2012-10-08 0 days 00:35:00 F9 v1.0 B0006 CCAFS LC-40 SpaceX CRS-1
1 5 2013-03-01 0 days 15:10:00 F9 v1.0 B0007 CCAFS LC-40 SpaceX CRS-2
                         Orbit Customer Mission Outcome Landing Outcome
   PAYLOAD MASS KG
                500 LEO (ISS) NASA (CRS)
                                                  Success
                                                              No attempt
                677 LEO (ISS) NASA (CRS)
                                                              No attempt
                                                  Success
Total payload mass, customer= NASA (CRS): 45596 kg
# Based on SOL only...
sql nasa crs mass= """ Select sum(PAYLOAD MASS KG ) as 'Total payload mass (kg) NASA CRS'
                   from spacex v11
                   where Customer='NASA (CRS)' """
payload NASA CRS=pd.read sql query(sql nasa crs mass,conn)
print(payload NASA CRS)
   Total payload mass (kg) NASA CRS
                           45596.0
```

### Average Payload Mass by F9 v1.1

Calculate the average payload mass carried by booster version F9 v1.1

```
payload_F9v11=pd.read_sql_query("Select avg(PAYLOAD_MASS__KG_) as 'avg mass (kg)' from spacex_v11 where Booster_Version='F9 v1.1'",conn)
print(payload_F9v11)

avg mass (kg)
0 2928.4
```

### First Successful Ground Landing Date

• Find the dates of the first successful landing outcome on ground pad

```
min_Date_success_landing=pd.read_sql_query("select min(Date) from spacex_v11 where Landing_Outcome = 'Success (ground pad)'",conn)
print(min_Date_success_landing)
    min(Date)
0 2015-12-22
```

### **Boosters Carried Maximum Payload**

• List the names of the booster which have carried the maximum payload mass.



### 2015 Launch Records

 List the failed landing\_outcomes in drone ship, their booster versions, and launch site names for in year 2015

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

• Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in

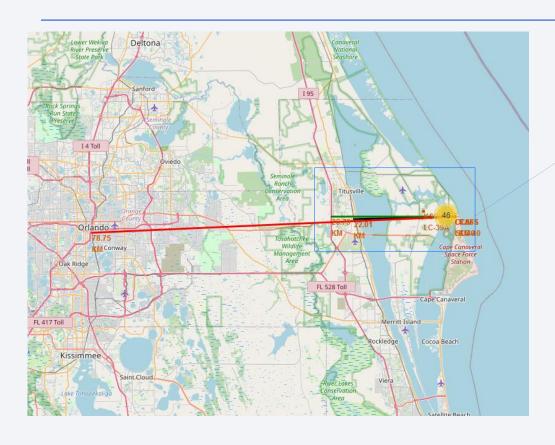
descending order

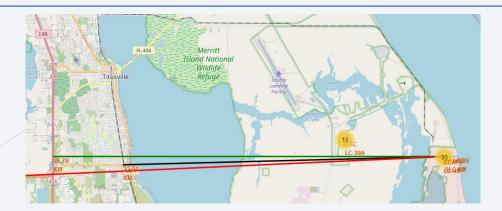
# sql query

```
q count landing= """ Select Landing Outcome, count(*) as count landings
                     from spacex v11
                     where Date between '2010-06-04' and '2017-03-20'
                     group by Landing Outcome
                     order by count landings desc
count landing= pd.read sql query(q count landing,conn)
count landing.head(10)
     Landing Outcome count landings
            No attempt
                                 10
      Failure (drone ship)
                                  5
    Success (drone ship)
                                  5
      Controlled (ocean)
                                  3
4 Success (ground pad)
     Uncontrolled (ocean)
                                  2
      Failure (parachute)
7 Precluded (drone ship)
```



### Distances between a launch site to its proximities





Launch sites are strategically located near coasts to ensure safety in case of early-flight failures.

Key distances from CCAFS\_SLC40:

•Coast: ~900 m •Railway: 22 km

•Highway I-95: 26.8 km

•Orlando: 78.75 km

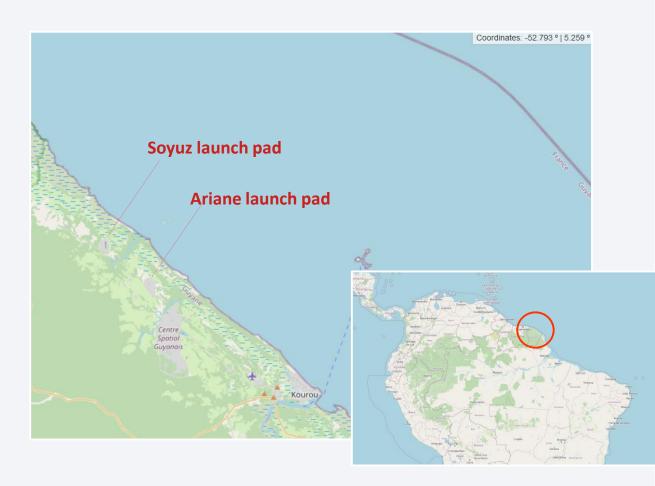
Florida launches head east over the ocean, while California launches follow north-south paths for polar orbits. Sites are kept distant from populated areas to 37 minimize risk during liftoff incidents.

### ESA Unique Launch site – Kourou, French Guiana

The European Space Agency (ESA) and ArianeEspace operate launch sites in Kourou, French Guiana, offering key advantages:

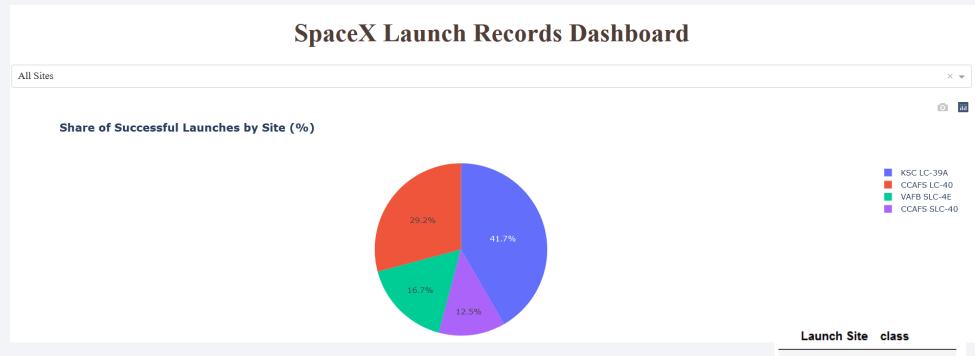
- •Equatorial Location: Ideal for GTO/GEO flights, reducing energy needs at liftoff.
- •Remote & Safe: Positioned away from densely populated areas and major infrastructure.
- •Coastal Proximity: Ensures safer trajectories over the ocean.

This setup provides an edge over SpaceX for GEO missions. SpaceX could counter this advantage by adopting a "Sea Launch" approach near the equator.





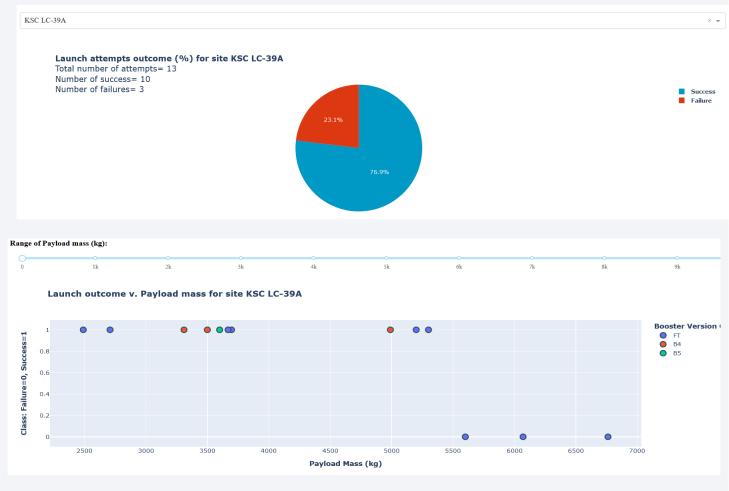
### SpaceX Falcon 9: Launch success count for all sites



The dashboard offers an interactive way to explore Falcon launch successes. It combines data from various charts into a single view and helps identify the launch site with the highest success rate, which is the Kennedy Space Center in Florida.

### SpaceX Falcon9 Launch site with highest launch success ratio

Launch Complex 39A at the Kennedy Space Center in Florida has conducted 13 launches, with 10 of them being successful. Missions carrying heavy payloads are considered high risk. For payloads weighing less than 5500 kg, success does not appear to be strongly influenced by the version of the booster. The Block 5 and Full Thrust boosters are the most frequently reused, and while the available data is limited, it suggests that these reused boosters may be just as reliable as those used for the first time.





### **Classification Accuracy**

Model performance evaluation using the test dataset.

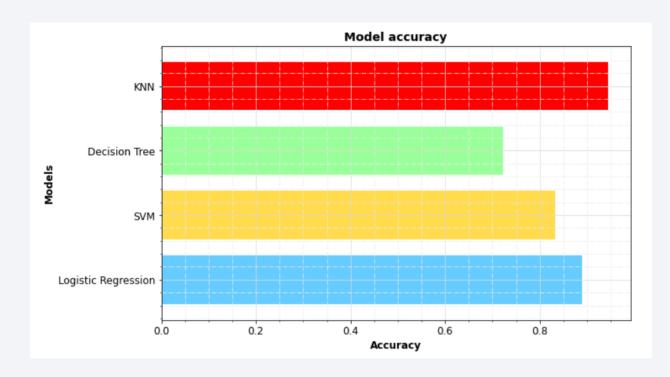
The results were obtained using a train-test split with a random state set to 3.

Hyperparameter tuning for SVM and Logistic Regression was carried out to improve training accuracy.

However, these adjustments did not significantly enhance performance on the test data.

The test set size is relatively small, limiting reliable conclusions.

Among the models evaluated, KNN delivered the highest accuracy, reaching approximately 94%.

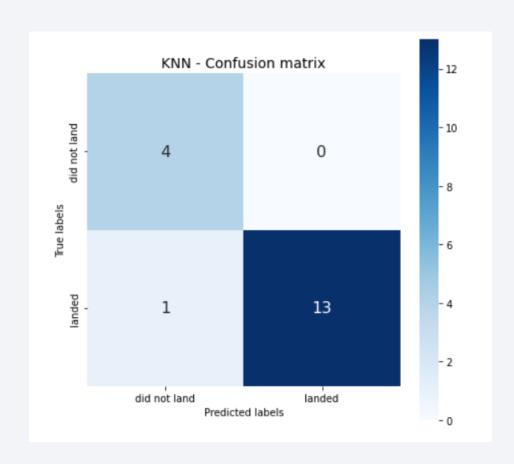


### **Confusion Matrix**

The k-nearest neighbors (k-NN) algorithm demonstrated the most reliable predictive performance.

The model accurately identified all mission failures.

There was only one instance of a false negative when predicting a successful booster landing.



### **Conclusions**

We defined a successful outcome as the Falcon 9 booster being recovered intact.

To explore the factors influencing recovery success, we collected and analyzed data from Falcon 9 mission records and Wikipedia.

Key factors examined include payload mass, orbital destination, booster version, and launch site. Using these features, we implemented a supervised classification model, achieving a predictive accuracy of 94%.

Falcon 9 booster recovery presents greater challenges compared to traditional rockets. However, its performance:

- Improves consistently over time
- •With a 66% success rate, it offers competitive cost efficiency per kilogram relative to Ariane 5 and Ariane 6

Geostationary Transfer Orbit (GTO) missions remain riskier, potentially due to:

- •High energy demands at liftoff
- •Vibrational stress affecting electronics and control systems

Competitors like ESA/Arianespace and Soyuz benefit from equatorial launch sites, reducing energy needs for GTO missions.

SpaceX could enhance its GTO success rate with ocean-based launch platforms near the equator.

Despite recent GTO failures, SpaceX continues to lead in cost efficiency. If Starship achieves similar 45 reliability, this advantage could grow significantly.

### Jupyter Notebooks

- https://github.com/reeva111/Applied Data Science Capstone SpaceX/blob/main/Build a Dashboard Application with Plotly Dash v10.ipynb
- <a href="https://github.com/reeva111/Applied Data Science Capstone SpaceX/blob/main/SpaceX Machine%20Lear-ning%20Prediction Part 5.ipynb">https://github.com/reeva111/Applied Data Science Capstone SpaceX/blob/main/SpaceX Machine%20Lear-ning%20Prediction Part 5.ipynb</a>
- <a href="https://github.com/reeva111/Applied Data Science Capstone SpaceX/blob/main/edadataviz%20(1).ipynb">https://github.com/reeva111/Applied Data Science Capstone SpaceX/blob/main/edadataviz%20(1).ipynb</a>
- <a href="https://github.com/reeva111/Applied Data Science Capstone SpaceX/blob/main/edasql.ipynb">https://github.com/reeva111/Applied Data Science Capstone SpaceX/blob/main/edasql.ipynb</a>
- <a href="https://github.com/reeva111/Applied Data Science Capstone SpaceX/blob/main/jupyter-labs-spacex-data-collection-api.ipynb">https://github.com/reeva111/Applied Data Science Capstone SpaceX/blob/main/jupyter-labs-spacex-data-collection-api.ipynb</a>
- https://github.com/reeva111/Applied Data Science Capstone SpaceX/blob/main/jupyter-labswebscraping.ipynb
- https://github.com/reeva111/Applied Data Science Capstone SpaceX/blob/main/lab jupyter launch site lo cation.ipynb
- https://github.com/reeva111/Applied Data Science Capstone SpaceX/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

