IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Aaron Pifanio

December 8, 2023

Outline

- Executive Summary
- Introduction
- Section 1: Methodology
- Section 2: Insights Drawn from EDA
- Section 3: Launch Sites Proximities Analysis
- Section 4: Build a Dashboard w/ Plotly Dash
- Section 5: Predictive Analysis
- Conclusion
- Appendix – GitHub Links

Introduction

- The commercial space age is here, companies are making space travel affordable for everyone. Virgin Galactic is providing suborbital spaceflights. Rocket Lab is a small satellite provider. Blue Origin manufactures sub-orbital and orbital reusable rockets. Perhaps the most successful is SpaceX. SpaceX's accomplishments include:
  - Sending spacecraft to the International Space Station.
  - Starlink, a satellite internet constellation providing satellite Internet access.
  - Sending manned missions to Space.
- One reason SpaceX can do this is the rocket launches are relatively inexpensive. SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upwards of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage.
- Therefore, if we can determine if the first stage will land, we can determine the cost of a launch.
- I will take the role of a data scientist working for a new rocket company, Space Y founded by Billionaire industrialist Allon Musk. Space Y would like to compete with SpaceX.
- My job is to determine the price of future launches. I will do this by gathering information about Space X and creating dashboards for my team.
- Instead of using rocket science to determine if the first stage will land successfully, I will train a machine learning model and use public information to predict if SpaceX will reuse the first stage.

Executive Summary

- To determine future launch prices and developing a machine learning model for predicting SpaceX's first stage reuse, I conducted a comprehensive analysis. This involved gathering SpaceX launch data from the SpaceX REST API and scraping Falcon 9 Launch records from Wikipedia. The collected data was meticulously processed for exploratory data analysis, and launch sites were visualized using Folium.
- To provide a holistic view of successful and failed launches, I designed a dashboard that breaks down outcomes by launch site. Subsequently, I constructed a machine learning pipeline to forecast the landing success of the first stage based on our processed data.
- Key findings from the analysis reveal that the KSC LC 39A Launch Site boasts the highest number of successful launches and the most favorable success rate for mission outcomes. Furthermore, the optimal booster version for achieving successful launch outcomes is identified as F9 FT. For heavier payloads exceeding 9,500 kg, the preferred booster version is B4.
- In terms of predictive modeling, the Decision Tree Classifier emerged as the most effective, yielding an overall accuracy of 83.33%. Notably, this model excelled in minimizing false positives and accurately capturing true positives. These insights serve as valuable inputs for strategic decision-making in future SpaceX launches.

Section 1

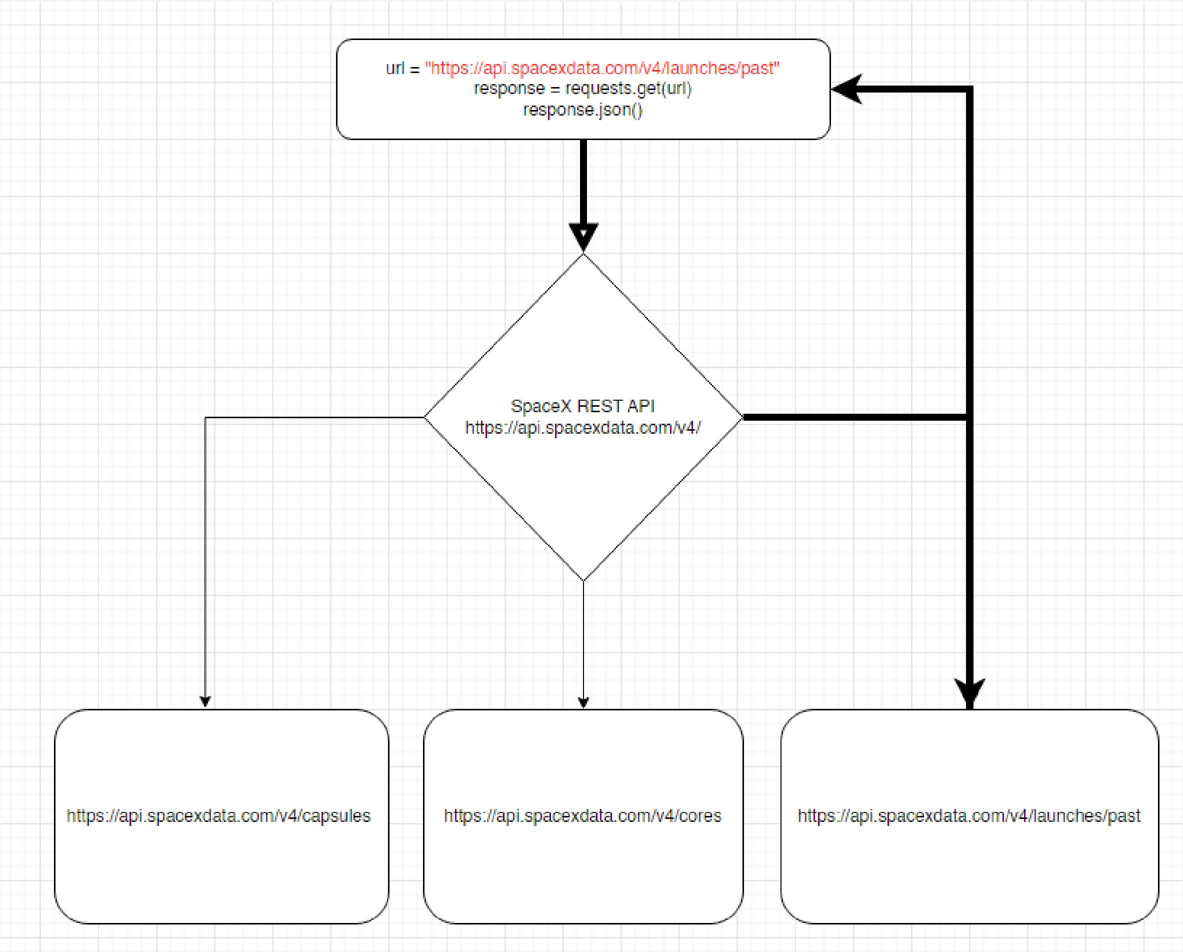
Methodology

Data Collection Methodology

1. Collected SpaceX launch data from the SpaceX REST API.
2. We also used the Python BeautifulSoup package to web-scrape some HTML tables that contain valuable Falcon 9 launch records from Wikipedia.

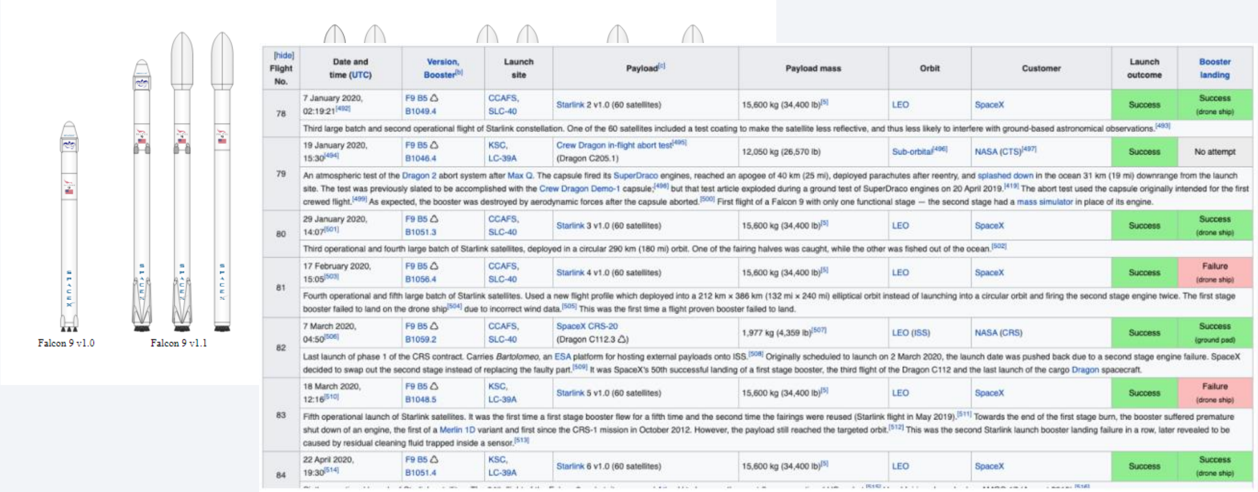
Data Collection – SpaceX REST API

- We gathered SpaceX Launch data from the SpaceX REST API.
- This API will give us data about launches, including information about the rocket used, payload delivered, launch and landing specifications, and landing outcome.
- We used the endpoint "api.spacexdata.com/v4/launches/past" to get data from past launches.
- We will perform a get request using the requests library to obtain the launch data, which we will use to get the data from the API. The result can be viewed by calling the .json() method.



Data Collection – Webscraping

- We also used the Python BeautifulSoup package to web-scrape some HTML tables that contain valuable Falcon 9 launch records from Wikipedia.
- Where after, we parsed the data and converted them into a Pandas data frame for further visualization and analysis.



Webscraping with BeautifulSoup

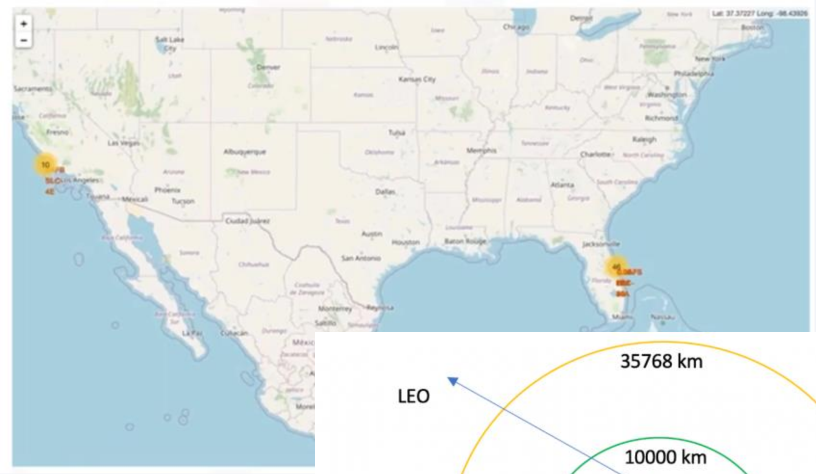
	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude
0	1	2006-03-24	Falcon 1	20.0	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	Merlin1A	167.743129	9.047721
1	2	2007-03-21	Falcon 1	NaN	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	Merlin2A	167.743129	9.047721
2	4	2008-09-28	Falcon 1	165.0	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	Merlin2C	167.743129	9.047721
3	5	2009-07-13	Falcon 1	200.0	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	Merlin3C	167.743129	9.047721
4	6	2010-06-04	Falcon 9	NaN	LEO	CCAFS SLC 40	None None	1	False	False	False	None	1.0	0	B0003	-80.577366	28.561857

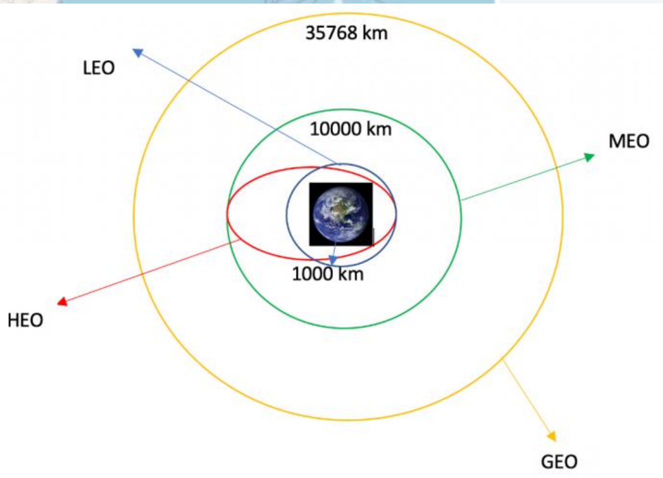
The Data We Obtained

Attributes:

- The column '**LaunchSite**' contains the different launch sites, including:
  - **Vandenberg AFB Space Launch**
  - **Kennedy Space Center**
  - **CCAF SLC 40**
  - **CCAFS SLC 40**

- The column '**Orbits**' are the different orbits of the payload.
  - **LEO**: Low Earth Orbit.
  - **VLEO**: Very Low Earth Orbits
  - **GTO**: Geosynchronous Orbit
  - **SSO**: Sun-Synchronous Orbit
  - **ES-L1**: Orbit at Lagrange
  - **HEO**: A highly elliptical orbit
  - **ISS**: International Space Station in Low Orbit Earth
  - **MEO**: Geocentric orbits below GTO
  - **HEO**: Geocentric orbits above GTO
  - **PO**: Polar Orbit





The Data We Obtained

Attributes:

- The column ‘Outcome’ indicates if the first stage successfully landed. There are 8 of them, for example.
  - **True Ocean** – The mission outcome was successfully landed to a specific region of the ocean.
  - **False Ocean** – The mission outcome was unsuccessfully landed to a specific region of the ocean.
  - **True RTLS** – The mission outcome was successfully landed to a ground pad
  - **False RTLS** – The mission outcome was unsuccessfully landed to a ground pad.
  - **True ASDS** – The mission outcome was successfully landed to a drone ship
  - **False ASDS** – The mission outcome was unsuccessfully landed to a drone ship.
  - **None ASDS & None None** – These represent a failure to land.

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude
0	1	2010-06-04	Falcon 9	6104.959412	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0003	-80.577366	28.561857
1	2	2012-05-22	Falcon 9	525.000000	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0005	-80.577366	28.561857
2	3	2013-03-01	Falcon 9	677.000000	ISS	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0007	-80.577366	28.561857
3	4	2013-09-29	Falcon 9	500.000000	PO	VAFB SLC 4E	False Ocean	1	False	False	False	NaN	1.0	0	B1003	-120.610829	34.632093
4	5	2013-12-03	Falcon 9	3170.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B1004	-80.577366	28.561857
5	6	2014-01-06	Falcon 9	3325.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B1005	-80.577366	28.561857
6	7	2014-04-18	Falcon 9	2296.000000	ISS	CCAFS SLC 40	True Ocean	1	False	False	True	NaN	1.0	0	B1006	-80.577366	28.561857
7	8	2014-07-14	Falcon 9	1316.000000	LEO	CCAFS SLC 40	True Ocean	1	False	False	True	NaN	1.0	0	B1007	-80.577366	28.561857
8	9	2014-08-05	Falcon 9	4535.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B1008	-80.577366	28.561857
9	10	2014-09-07	Falcon 9	4428.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B1011	-80.577366	28.561857

Data Wrangling Methodology

1. Store the data from ‘LaunchSite’, ‘Orbit’, ‘Outcome’ columns to determine **number of launches on each site**, **the occurrence of each orbit**, and **number of landing outcomes**.
2. Change the dependent variables from the ‘Outcome’ column to where **successful booster landings** are converted to the value **1** and **failed booster landings** are converted to the value **0**.

Data Wrangling

Step 1: Store the data from '[LaunchSite]', '[Orbit]', '[Outcome]' into variables using the method '.value_counts()' to determine the **number of launches on each site**, the **occurrence of each orbit**, and **number of landing outcomes**.

```python
# Apply value_counts() on column LaunchSite
launch_site_counts = df['LaunchSite'].value_counts()

print(launch_site_counts)
```

```
CCAFS SLC 40    55
KSC LC 39A      22
VAFB SLC 4E     13
Name: LaunchSite, dtype: int64
```

```python
# Apply value_counts on Orbit column
NumberofOrbits = df['Orbit'].value_counts()

print(NumberofOrbits)
```

```
GTO      27
ISS      21
VLEO     14
PO        9
LEO       7
SSO       5
MEO       3
ES-L1     1
HEO       1
SO        1
GEO       1
Name: Orbit, dtype: int64
```

```python
# landing_outcomes = values on Outcome column
landing_outcomes = df['Outcome'].value_counts()

print(landing_outcomes)
```

```
True ASDS      41
None None      19
True RTLS      14
False ASDS      6
True Ocean      5
False Ocean     2
None ASDS       2
False RTLS      1
Name: Outcome, dtype: int64
```

Data Wrangling

Step 2: Enumerate all variables in the ‘**outcome**’ variable.

.

```
for i,outcome in enumerate(landing_outcomes.keys()):
    print(i,outcome)
```

```
0 True ASDS
1 None None
2 True RTLS
3 False ASDS
4 True Ocean
5 False Ocean
6 None ASDS
7 False RTLS
```

Step 3: Create a set of outcomes where the second stage did not land successfully.

```
bad_outcomes=set(landing_outcomes.keys()[[1,3,5,6,7]])
bad_outcomes
```

```
{'False ASDS', 'False Ocean', 'False RTLS', 'None ASDS', 'None None'}
```

.

Step 4: Using the variable ‘**outcome**’, create a list where the element is 0 if the corresponding row in ‘**outcome**’ is in the set ‘**bad_outcome**’ otherwise, it's one. Then assign it to the variable ‘**landing_class**’

.

```
landing_class = []

# Iterate over each outcome in the 'Outcome' column of the DataFrame
for outcome in df['Outcome']:
    # Check if the outcome is in the bad_outcomes set
    if outcome in bad_outcomes:
        # If so, append 0 to the list
        landing_class.append(0)
    else:
        # If not, append 1 to the list
        landing_class.append(1)
```

Data Wrangling

Note: The ‘**landing_class**’ variable will represent the classification variable that represents the outcome of each launch. If the value is zero, the first stage did not land successfully; one means the first stage landed Successfully

Step 5: Add a new column named ‘**Class**’ to the data frame and assign the values from the variable ‘**landing_class**’ to this new column

```
df['Class']=landing_class
df[['Class']].head(8)
```

	Class
0	0
1	0
2	0
3	0
4	0
5	0
6	1
7	1

```
df.head(5)
```

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude	Class
0	1	2010-06-04	Falcon 9	6104.959412	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0003	-80.577366	28.561857	0
1	2	2012-05-22	Falcon 9	525.000000	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0005	-80.577366	28.561857	0
2	3	2013-03-01	Falcon 9	677.000000	ISS	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0007	-80.577366	28.561857	0
3	4	2013-09-29	Falcon 9	500.000000	PO	VAFB SLC 4E	False Ocean	1	False	False	False	NaN	1.0	0	B1003	-120.610829	34.632093	0
4	5	2013-12-03	Falcon 9	3170.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B1004	-80.577366	28.561857	0

Exploratory Data Analysis Methodology

1. Determine what attributes are correlated with successful landings using SQL and Data Visualization
2. The categorical variables will be converted using one hot encoding, preparing the data for a machine learning model that will predict if the first stage will land.

EDA with SQL

- SQL Queries Performed:

1. Display the names of the unique launch sites in the machines.
2. Display 5 records where launch sites begin with he string ‘CCA’.
3. Display the total payload mass carried by boosters launched by NASA (CRS).
4. Display average payload mass carried by booster version F9 v1.1.
5. List the date when the first successful landing outcome in ground pad was achieved.
6. List the names of the boosters which have success in drone ship and have a payload mass greater than 4,000 but less than 6,000.
7. List the total number of successful and failure mission outcomes.
8. List the names of the booster versions which have carried the maximum payload mass.
9. List the records which will display the months names, failed landing outcomes in drone ship, booster versions launch site for the months in year 2015.
10. 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.

EDA with Data Visualization

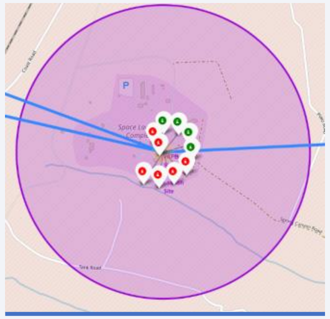
- Charts plotted:
  1. The relationship between flight number and launch sites.
     - Done to see how launch sites were used as more flights were launched and how that affected launch outcomes.
  2. The relationship between payload mass (kg) and launch sites.
     - Done to see if payload mass had any correlation with launch outcomes depending on the launch site.
  3. The relationship between the success rate of each orbit.
     - Done to see if orbits can affect success rates of launch outcomes.
  4. The relationship between flight number and orbit type.
     - Done to drill down on how many flights were made in each orbit and how many launch outcomes succeeded.
  5. The relationship between payload mass (kg) and orbit type.
     - Done to see if there was a correlation between payload mass(kg) and orbit type that affected success rates of launch outcomes.
  6. The average launch success rate over the years.
     - Done to observe the aggregate success rate of launch outcomes from 2010 to 2020.

Launch Sites Location Analytic Methodology w/ Folium

1. The launch success rate may depend on many factors such as payload mass, orbit type, and so on. It may also depend on the location and proximities of a launch site, i.e., the initial position of rocket trajectories. Finding an optimal location for building a launch site certainly involves many factors and hopefully we could discover some of the factors by analyzing the existing launch site locations and performing interactive visual analytics with Folium.

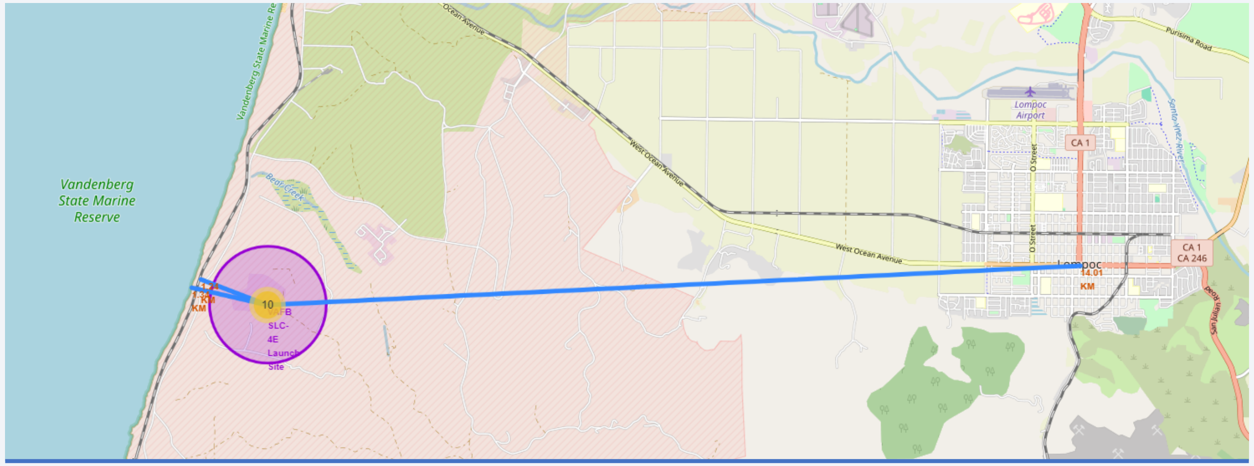
Launch Sites Location Analytic Methodology w/ Folium

- For our interactive map, we created markers for each launch site on the map.
- We also marked the number of successful/failed launches for each launch site on the map.
- We then calculated between a launch site and its distance to certain proximities such as cities, railways, and coastlines to see if any of these could be a correlated to how many successful/failed launches there were.



Launch Outcomes for VAFB SLC-4E Launch Site

Distance from KSC LC-39A Site to the coastline and railway



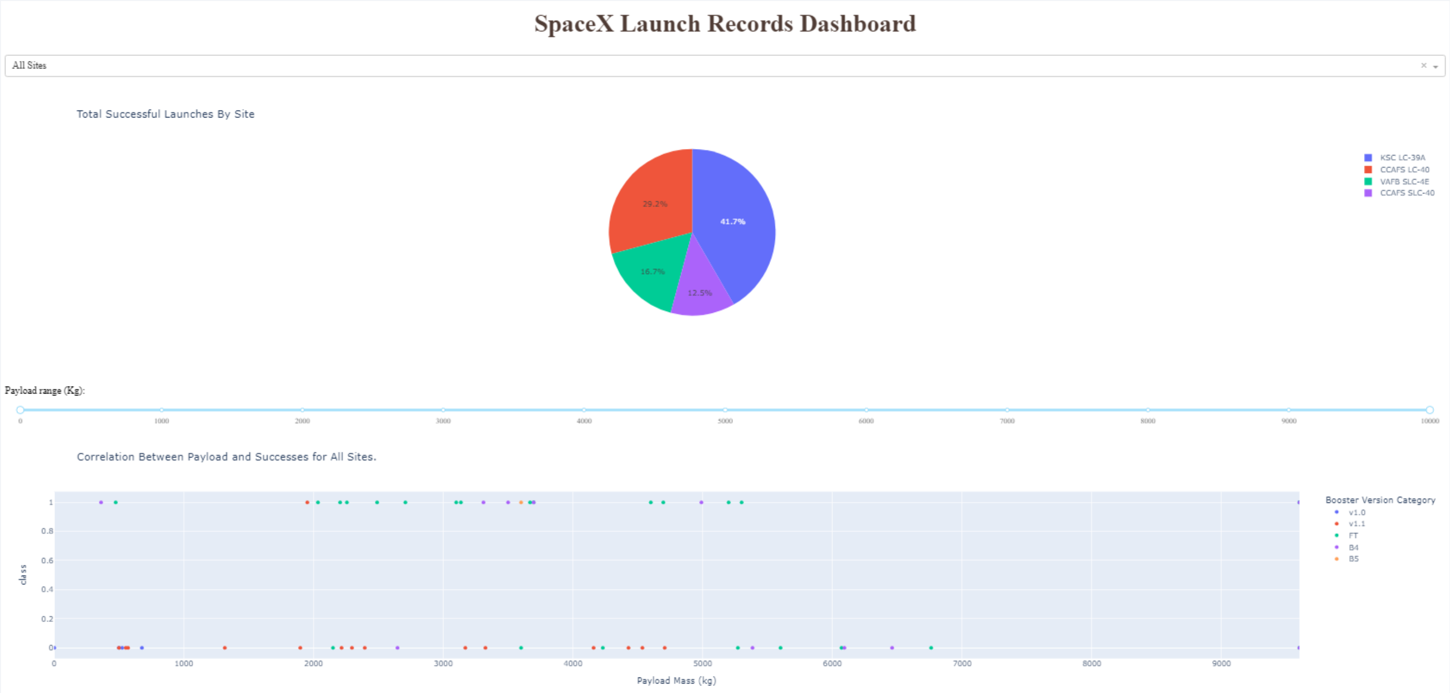
Distance from VAFB SLC-4E Launch Site to the City of Lompoc

Dashboard Methodology w/ Plotly Dash

1. I created an interactive and dynamic dashboard using Plotly Dash to analyze the total outcome of successful and failed launches by site through a pie chart and a scatterplot where we can observe how payload may be correlated with mission outcomes for selected sites.

Dashboard Methodology w/ Plotly Dash

- I also added a drop-down menu to easily select between each site or to see the total successful launches by site.
- A payload range slider was created to easily observe which the mission outcomes for each booster version depending on the payload that we want to see.

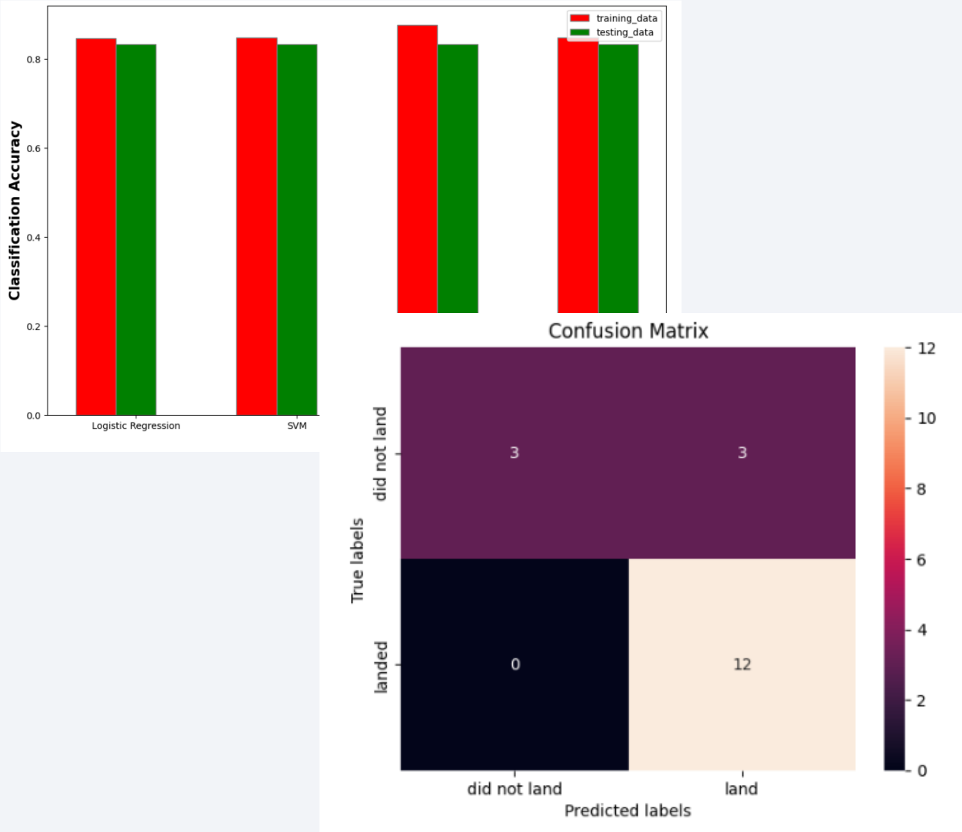


Predictive Analysis Methodology

1. I created a machine learning pipeline to predict if the first stage will land given the data that we have explored and manipulated previously

Predictive Analysis Methodology

- **Step 1:** I standardized the data into variable X and created a NumPy array from the column 'Class' in our data frame which I then assigned to variable Y.
- **Step 2:** I split the data X and Y into a training and test dataset. Setting the parameter test_size to 0.2 and random_state to 2.
- **Step 3:** I created a logistic regression, support vector machine, decision tree classifier, and k-nearest neighbors object.
- **Step 4:** I tested the classification accuracy for each object for the training and testing data to find the best model using a bar chart.
- **Step 5:** Once I found the best model, I then plotted its confusion matrix and assessed their true positives, true negatives, false positives, and false negatives.
- **Step 6:** I then analyzed the performance metrics for the best performing model.



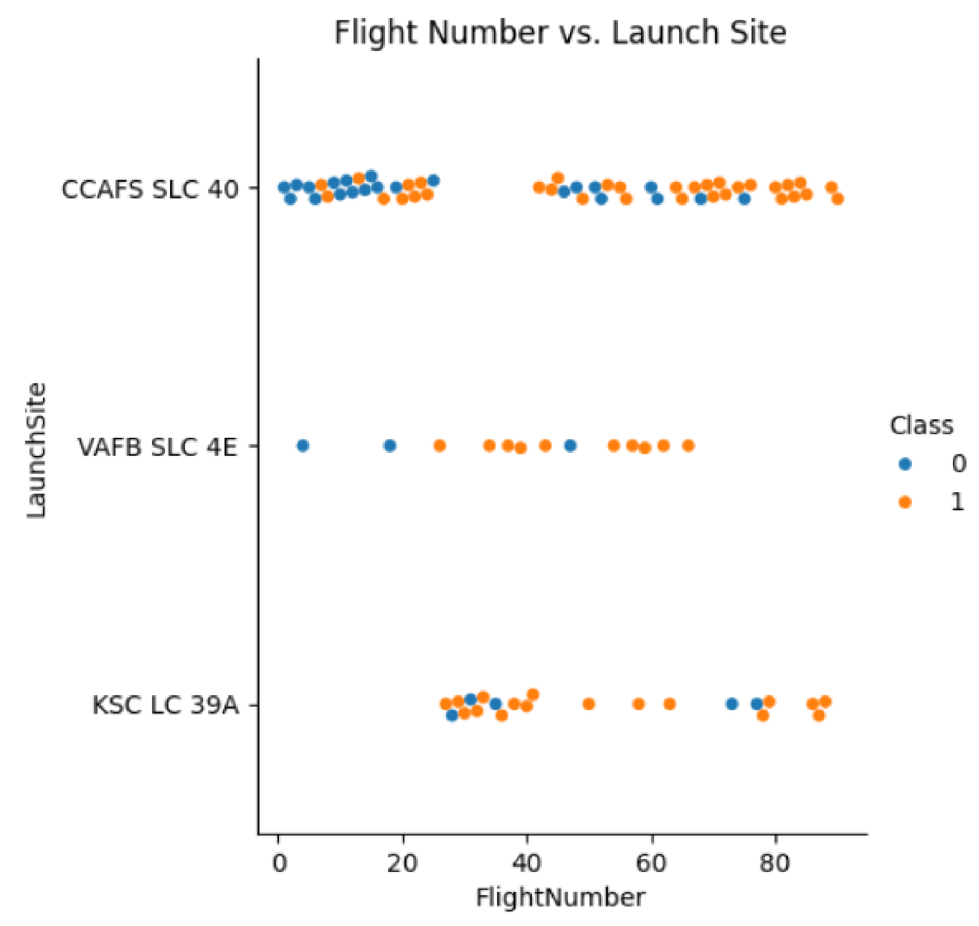
- Add the GitHub URL of your completed predictive analysis lab, as an external reference and peer-review purpose

Section 2

Insights drawn from EDA

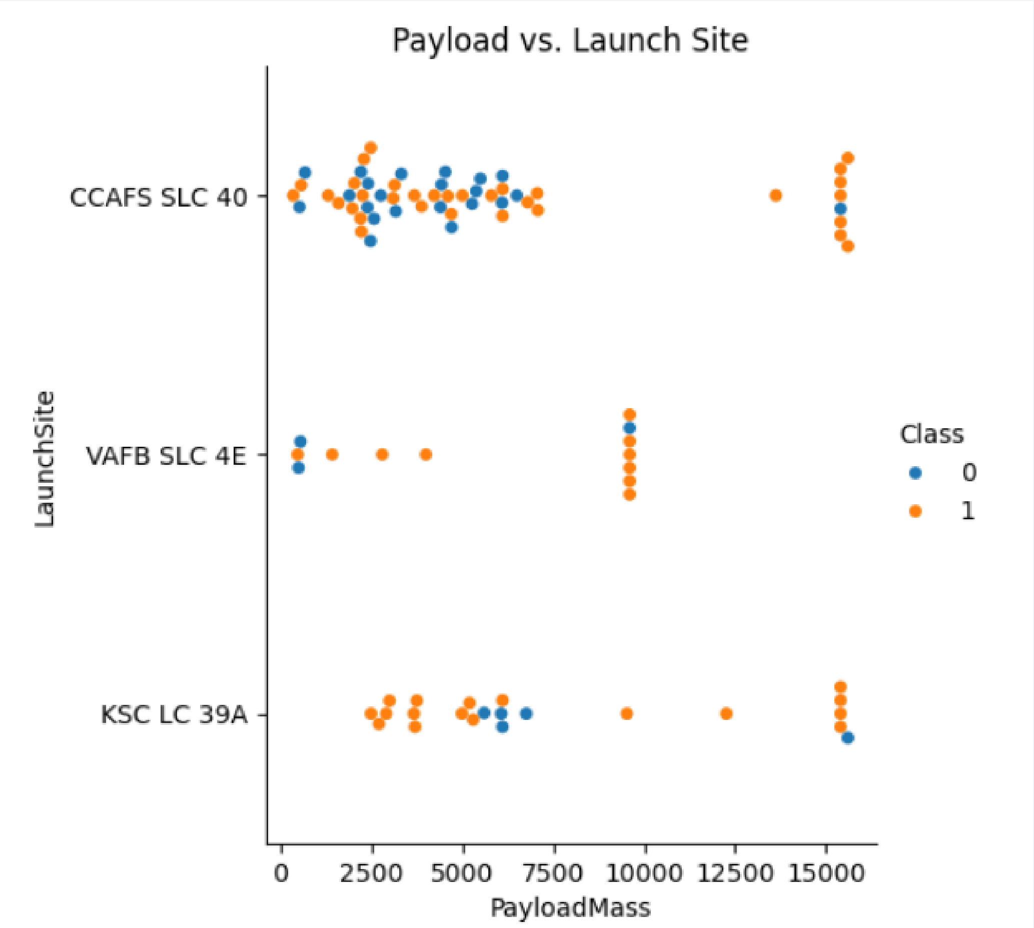
Flight Number vs. Launch Site

- From this scatter plot, we can see that the likelihood of a successful launch outcome increases as the flight number increases, indicating a <u>positive correlation between the flight number and successful launch outcome for all launch sites.</u>



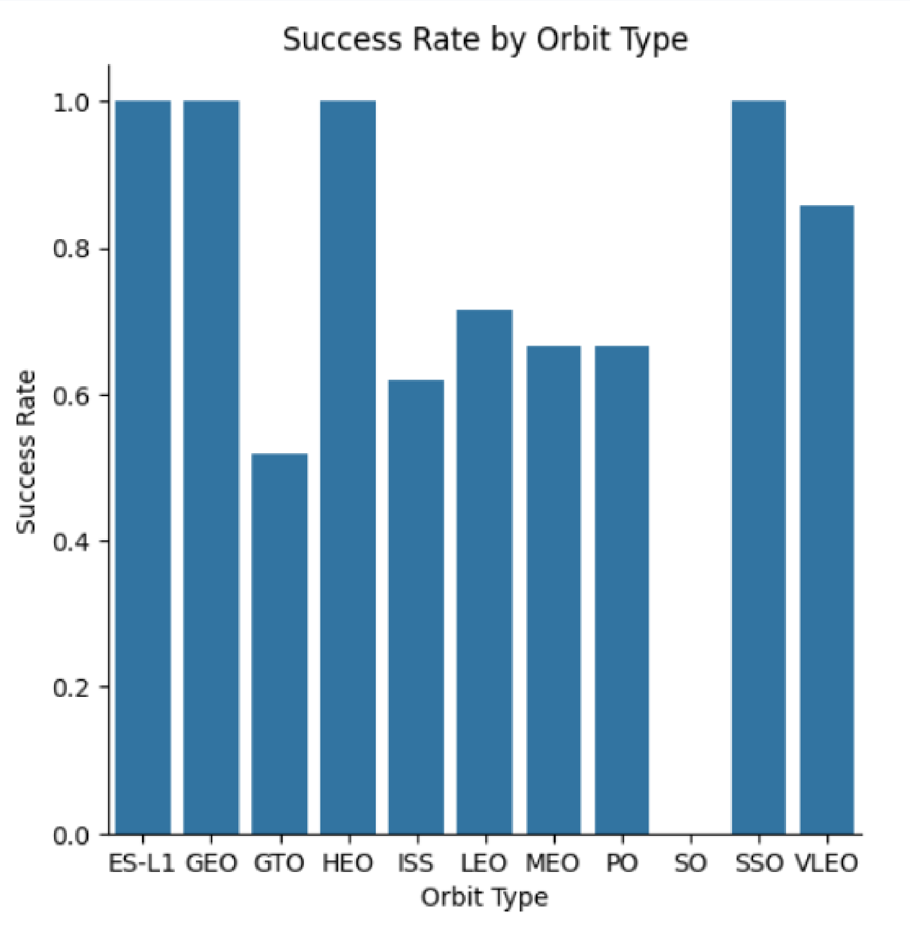
Payload vs. Launch Site

- Immediately, we can see that for launch site VAFB SLC 4E, there are no rockets launched for a payload mass greater than 10,000 kg
- Rockets in launch site CCAFS SLC 40 and KSC LC 39A can take heavier payloads just above 15,000 kg.
- The data shows that between launch site CCAFS SLC 40 and KSC LC 39A, CCAFS SLC 40 has a higher likelihood of a successful launch outcome with heavier payloads than KSC LC 39A.
- However, KSC LC 39A has a greater successful launch outcome ratio than CCAFS SLC 40.



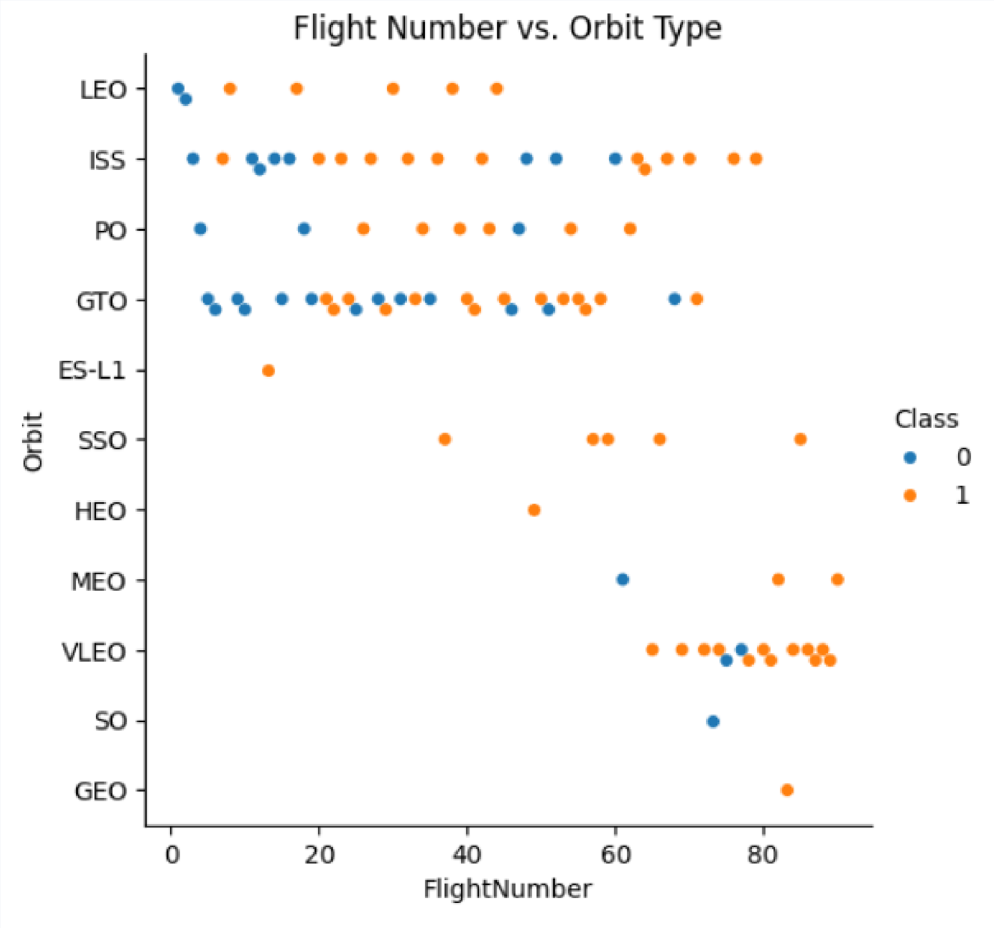
Success Rate vs. Orbit Type

- From this bar chart we can see that:
  - **ES-L1**, **GEO**, **HEO**, and **SSO** have a 100% success rate:
  - **VLEO** has around 90% success rate.
  - **LEO** has around a 70% success rate.
  - **MEO** and **PO** have around a 65% success rate.
  - **ISS** has around a 60% success rate.
  - **GTO** has a 50% success rate.
  - **SO** has a 0% success rate.
- But success rates don't tell us how many launches were dedicated for each orbit. For all we know, there could have been one failed launch that represents the **SO** orbit's 0% success rate… we need more information.



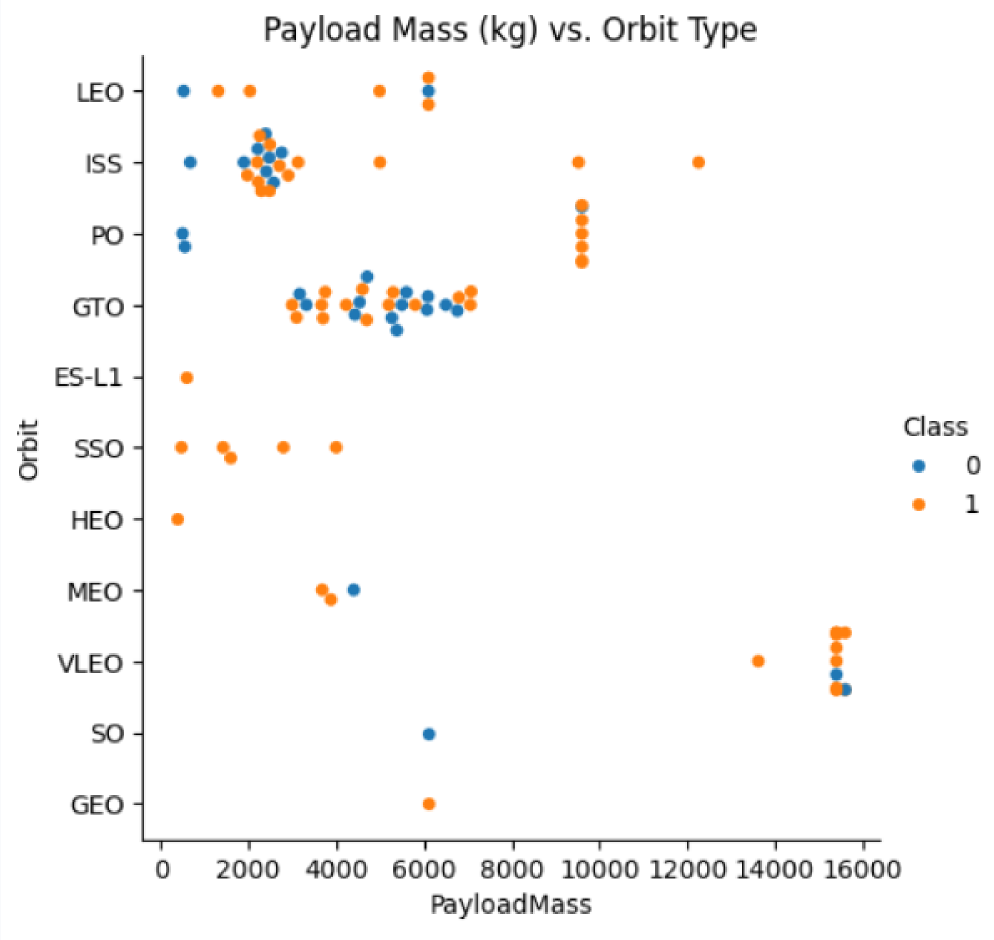
Flight Number vs. Orbit Type

- This scatter plot can give us a more detailed outlook on the relationship between flight numbers and orbit types.
- In the previous slide, we know that orbits ES-L1, HEO, and GEO have a 100% success rate while the SO orbit had a 0% success rate. In this plot, we can see that each orbit only has 1 flight to represent that statistic, emphasizing the limited sample size for each orbit type and underscoring the importance of caution in drawing definitive conclusions based on these results.
- For other orbits such as LEO, we can see that success does appear to be correlated to the number of flights while on the other hand, there seems to be no relationship between flight number when in GTO orbit.



Payload vs. Orbit Type

- For the relationship between payload and orbit type, we can see that Polar, LEO, and ISS have a relatively better positive landing rate with heavier payloads.
- However, for GTO, we can't distinguish between a positive or negative correlation for payload mass since the successful and failed launch outcomes are evenly distributed.



Yearly Trend of Successful Launches

- You can observe that the success rate since 2013 kept increasing until 2020.

All Launch Site Names

- There are four different launch sites:
    - **CCAFS LC-40**: Cape Canaveral Air Force Station Launch Complex 40, located in Florida, USA.
    - **CCAFS SLC-40**: Cape Canaveral Air Force Station Space Launch Complex 40, located in Florida, USA.
    - **VAFB SLC-4E**: Vandenberg Air Force Base Space Launch Complex 4E, located in California, USA.
    - **KSC LC-39A**: Kennedy Space Center Launch Complex 39A, located in Florida, USA.

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

Launch Site Names Begin with 'CCA'

- The first 5 records where the launch sites begin with ‘CCA’ give us information about the CCAFS LC-40 launch site.
- We can see that most of the launches were specific to the LEO (ISS) orbit when NASA was the designated customer.
- All mission outcomes were successful. However, there seems to be no success in having the first stage land for the first 5 records.

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2010-04-06	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-08-12	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
2012-08-10	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-01-03	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass for NASA (CRS)

- The total payload carried by boosters from NASA (CRS) is 45,596kg.

Customer	total_payload_mass
NASA (CRS)	45596

Average Payload Mass by F9 v1.1

- The average payload mass for Booster Version F9 v1.1 is 2,534.66kg

Booster_Version	AVG(PAYLOAD_MASS__KG_)
F9 v1.1 B1003	2534.6666666666665

First Successful Ground Landing Date

- The date when the first successful landing outcome in ground pad was achieved was in December 22, 2015.

MIN(Date)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- The boosters that have successfully landed on a drone ship with a payload mass greater than 4,000 but less than 6,000 are:
  - F9 FT B1022
  - F9 FT B1026
  - F9 FT B1021.2
  - F9 FT B1031.2

Booster_Version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

- Having calculated the total successes and failures of mission outcomes, we can conclude that there is a 99% success rate in mission outcomes.

TotalSuccess	TotalFailure
100	1

Boosters Carried Maximum Payload

- The maximum payload mass that a current boosters can carry are 15,600 kg. The following boosters that can carry the maximum payload are:
    - 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

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

- The failed Stage 1 landings by drone ship in 2015 occurred in April and October. The rockets were launched from CCAFS LC-40 and used boosters:
  - F9 v1.1 B1015
  - F9 v1.1 B1012

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

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

- By finding the landing outcomes between 2010-06-04 and 2017-03-20, we can see that there were 10 instances where no attempt was made to land Stage 1.
- However, there are 10 instances were Stage 1 did successfully land in ground pads and drone ships
- There were 5 failed landing attempts in a drone ship.
- Out of the 5 attempts to land Stage 1 in the ocean, 3 were controlled and 2 were uncontrolled.
- One outcome had a precluded landing in a drone ship. Suggesting that certain conditions or factors are in place, making landing unattainable or impractical.
- There was one attempt where Stage 1's parachute failed to launch.
- In total, 13 missions had a successful or controlled landing outcome while 9 missions had failed landing outcomes. (Not counting 'No attempt' landing_outcome.

Date	Landing_Outcome	OutcomeCount
2012-05-22	No attempt	10
2015-12-22	Success (ground pad)	5
2016-08-04	Success (drone ship)	5
2015-10-01	Failure (drone ship)	5
2014-04-18	Controlled (ocean)	3
2013-09-29	Uncontrolled (ocean)	2
2015-06-28	Precluded (drone ship)	1
2010-08-12	Failure (parachute)	1

EDA Conclusion

- There is a positive correlation between the flight number and successful launch outcome for all launch sites.
- There are 4 different launch sites: **CCAFS LC-40, CCAFS SLC-40, VAFB SLC-4E, KSC LC-39A.**
- CCAFS SLC 40 has a higher likelihood of a successful launch outcome with heavier payloads than KSC LC 39A.
- Orbits such as LEO, we can see that success does appear to be correlated to the number of flights while on the other hand, there seems to be no relationship between flight number when in GTO orbit.
- The boosters that have successfully landed on a drone ship with a payload mass greater than 4,000 but less than 6,000 are:
  - F9 FT B1022
  - F9 FT B1026
  - F9 FT B1021.2
  - F9 FT B1031.2
- Having calculated the total successes and failures of mission outcomes, we can conclude that there is a 99% success rate in mission outcomes.
- The success rate since 2013 kept increasing until 2020.

Section 3

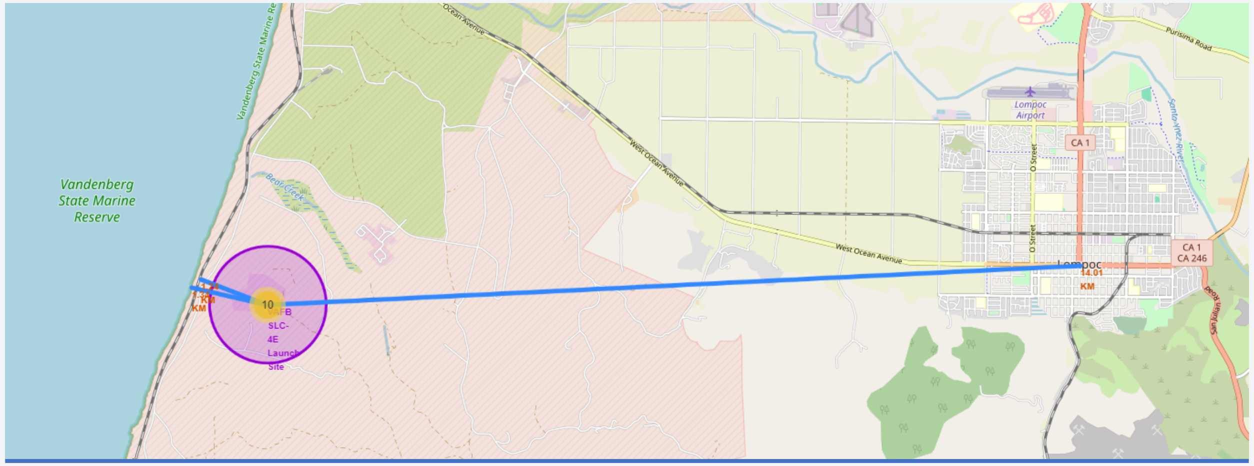
Launch Sites Proximities Analysis

VAFB SLC-4E Launch Site

- The VAFB SLCE-4E Launch Site is located within the Vandenberg Space Force Base in California USA.
- From the map, you can see that there are a total of 10 launch outcomes with 4 launches being successful and 6 launches having failed.
- The closest coastline is 1.34 km away, the closest railway is 1.24 km away, and the closest city, Lompoc, is 14.01 km away.

Launch Outcomes

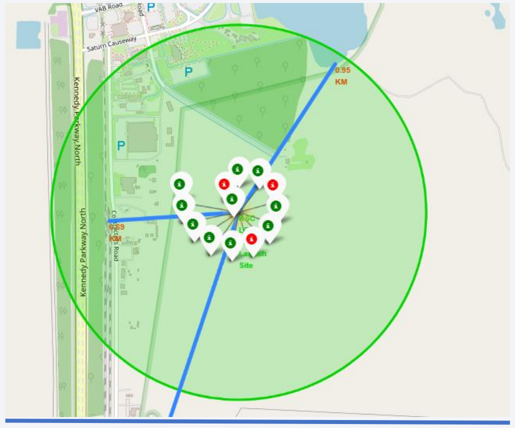
Distance from VAFB SLC-4E Launch Site to the coastline and railway



Distance from VAFB SLC-4E Launch Site to the City of Lompoc

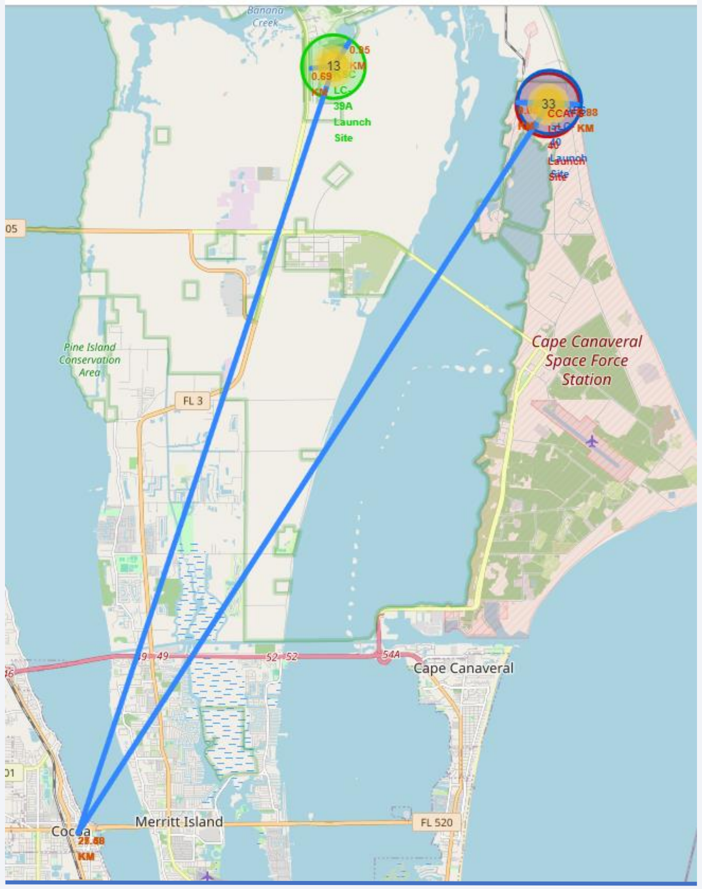
KSC LC-39A Launch Site

- The KSC LC-39A launch site is located at the Kennedy Space Center in Florida, USA. Specifically situated on Merritt Island.
- From the map, you can see that there are a total of 13 launch outcomes with 10 launches being successful and 3 launches having failed.
- The closest coastline is 0.95 km away, the closest railway is 0.69 km away, and the closest city, Cocoa, which is around 25-30 km away.



Launch Outcomes

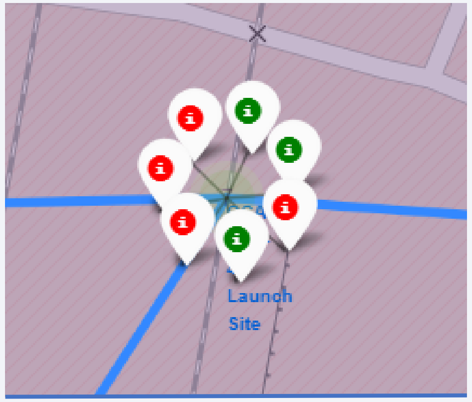
Distance from KSC LC-39A Site to the coastline and railway



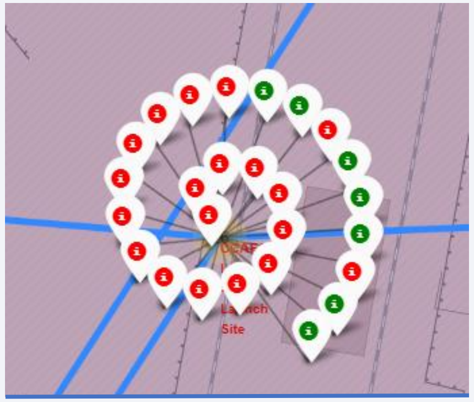
Distance from KSC LC-39A Site to the City of Cocoa

CCAFS LC-40 & CCAFS SLC-40 Launch Site

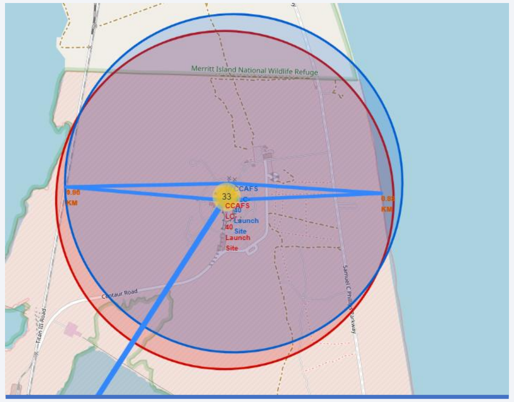
- The CCAFS LC-40 and CCAFS SLC-40 launch sites are located at Cape Canaveral Space Launch Complex in Florida, USA.
- For CCAFS LC-40, there were 7 launch outcomes with 3 launches being successful and 4 launches having failed.
- For CCAFS SLC-40, there were 26 launch outcomes with 7 launches being successful and 19 launches having failed.
- The closest coastline is 0.88 km away, the closest railway is 0.96 km away, and the closest city, Cocoa, which is around 25-30 km away.



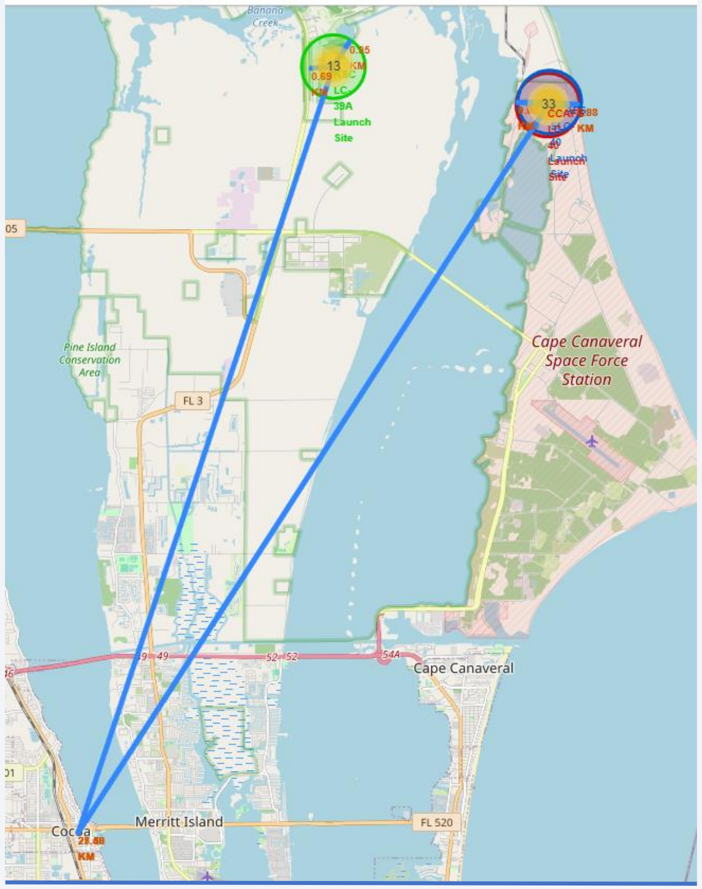
CCAFS SLC-40 Launch Outcomes



CCAFS LC-40 Launch Outcomes



Distance from CCAFS LC-40 and CCAFS SLC-40 to the coastline and railway



Distance from CCAFS LC-40 and CCAFS SLC-40 Site to the City of Cocoa

Launch Site Proximity Analysis Conclusion

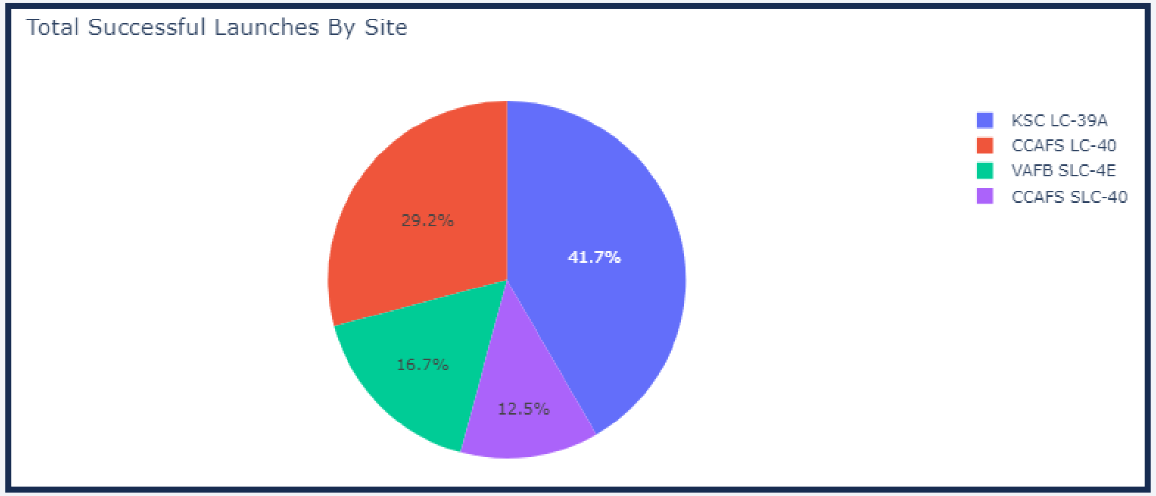
- KSC LC-39A is the launch site with the most successful launch outcomes and also happens to have the closest proximity to the nearest railway at 0.69 km.
- CCAFS SLC-40 happens to have the most failed launch outcomes with a 27% success rate and has the greatest number of launches.
- CCAF SLC-40 has a 47% success rate and is the launch site with the least amount of launches.
- VAFB SLC-4E is the launch site with the farthest proximity to a railway, coastline, and city compared to other launch sites. It has a 40% success rate and is based on the west coast compared to the other launch sites which are located on the east coast.

Section 4

Build a Dashboard with Plotly Dash

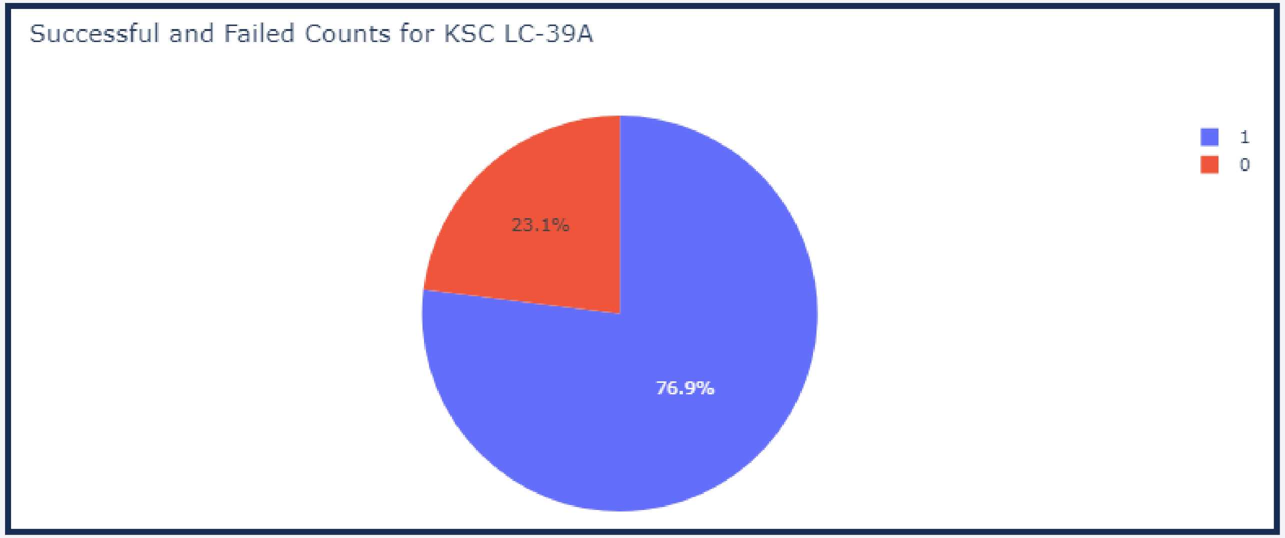
Total Successful Launches By Site

- Per the pie chart, it's evident to see that KSC LC-39A Launch Site has the most successful launches with a count of 10.
- Followed by CCAFS LC-40 with 7 successful launches
- VAFB SLC-4E with 4 successful launches
- And CCAFS SLC-40 with 3 successful launches.



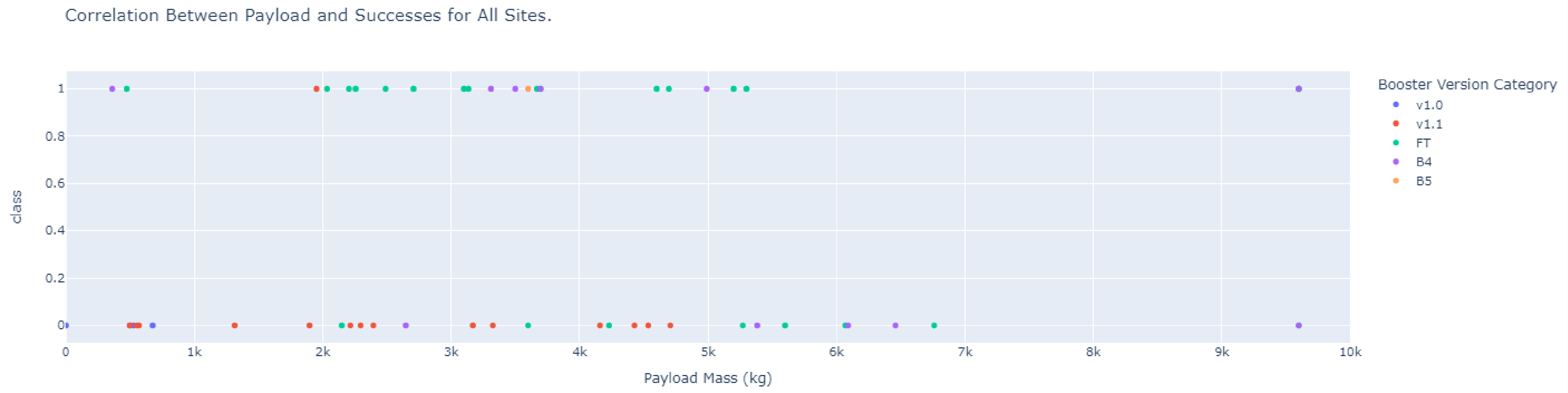
Launch Site With The Highest Launch Success Ratio

- The launch site with the highest launch success ratio is KSC LC-39A with a 76.9% success rate. This confirms that this launch site is the most successful out of the others and certain characteristics of this launch site need to be analyzed to conclude why it's so successful.



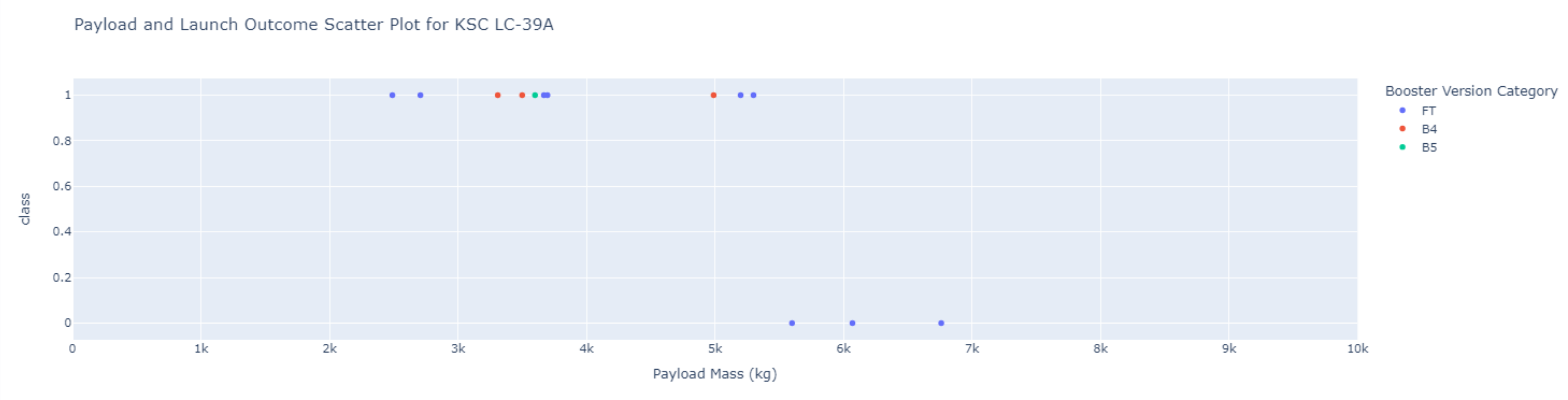
Payload vs. Launch Outcome For All Sites

- We can see that the booster that can carry over 9,500 kg of payload mass is B4.
- However, the booster version with the most successful launch outcome is FT for up to just above 5,000 kg of payload mass.



Payload vs. Launch Outcome For KSC LC-39A

- For launch site KSC LC-39A, the booster version that's most used is FT with a 66.66% success rate (6 successful launch outcomes and 3 failed launch outcomes). Looking at the graph, the maximum payload the FT booster can take is just below 5,500 kg until the booster fails. The other booster versions do not have a large enough sample size to make a conclusion. However, we can conclude that the B4 booster is able to take a 5,000 kg payload.



Dashboard Conclusion

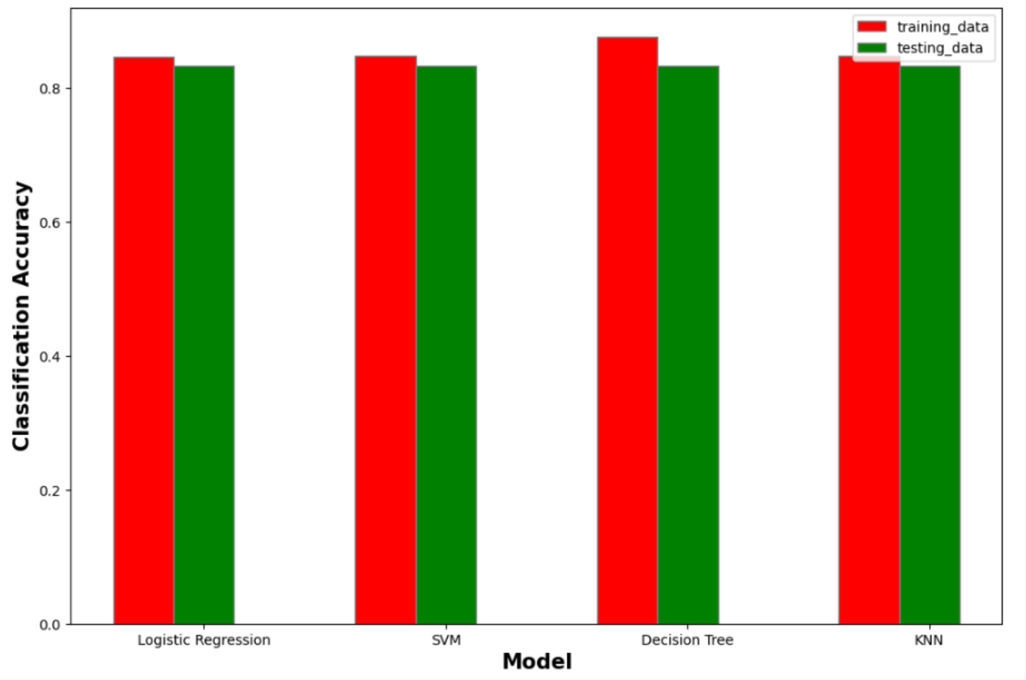
- KSC LC-39A Launch Site has the most successful launches with a count of 10.
- KSC LC-39A also has the highest success rate out of all the launch sites.
- The booster version that can carry the most payload mass at over 9,500 kg is B4.
- The booster version with the most successful launch outcome is FT for up to just above 5,000 kg of payload mass.
- FT also happens to be the booster version that KSC LC-39A uses the most.

Section 5

Predictive Analysis (Classification)

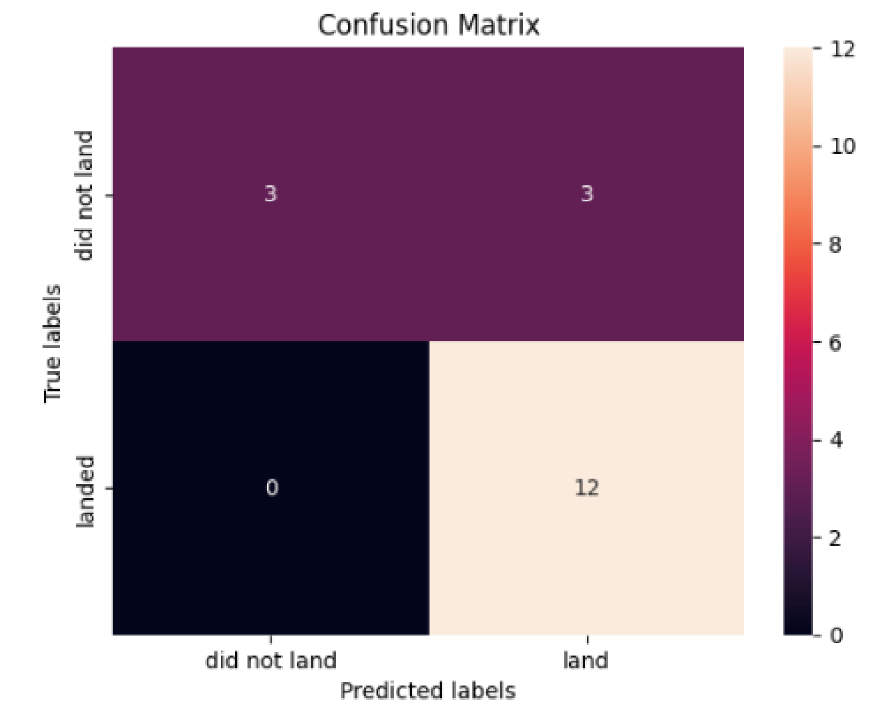
Classification Accuracy

- All models have a similar classification accuracy with marginally different accuracies for the training data set.
- For the testing data set, all models have the same accuracy of 83.33%. This means that all models are the same when it comes to making predictions in data it has never seen before.
- For the training data set, the model with the highest accuracy is the Decision Tree Classifier with 87.68%. This indicates that the Decision Tree Classifier model is the best when it comes to fitting the training set well and is the one that has learned the most patterns and relationships within the data it was trained on.



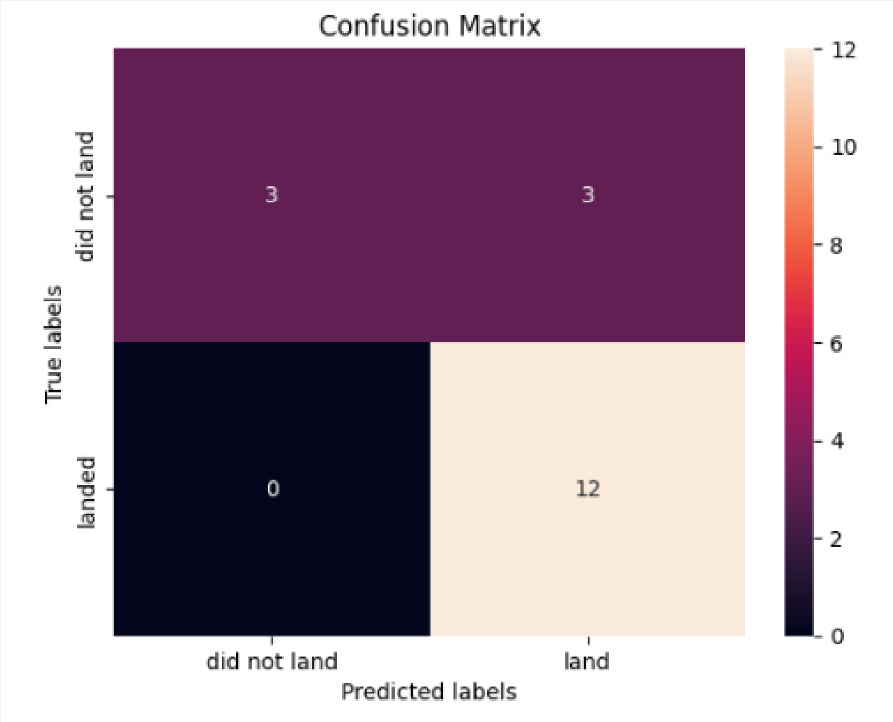
Confusion Matrix – Breakdown for Decision Tree Classifier

- **True Positive (TP):** The model was able to correctly predict 12 landings. TP = 12.
- **True Negative (TN):** The model was able to correctly predict 3 landings that failed. TN = 3.
- **False Positive (FP):** The model incorrectly predicted 3 landings as successful, but they did not actually land. FP = 3.
- **False Negative (FN):** The model did not incorrectly predict any landings that did not land which actually landed. FN = 0



Confusion Matrix – Performance Metrics

- **Accuracy:** (TP + TN) / (TP + TN + FP + FN) = 83.33%
  - The overall correctness of the model is 83.33%
- **Precision (Positive Predicted Value):** TP / (TP + FP) = 80%
  - The accuracy of successful landing predictions is 80%
- **Recall (True Positive Rate):** TP (TP + FN) = 100%
  - The ability of the model to capture al landings is 100%
- **Specificity (True Negative Rate):** TN / (TN + FP) = 50%
  - The ability of the model to avoid false positives is 50%
- **F1 Score:** 2*(Precision * Recall) / (Precision + Recall) = 88.89%
  - 88.89% is a high F1 score suggesting that the model is performing well in terms of minimizing false positives and capturing true positives.



Overall Conclusions

- KSC LC 39A Launch Site has the most successful launches and best success rate of mission outcomes.
- The best booster version to use with the most successful launch outcomes is F9 FT.
- For heavier payloads, it seems that the best booster version to use is B4 which can carry a payload mass at just over 9,500 kg.
- The best model to use to predict a successful landing outcome is the Decision Tree Classifier with an overall accuracy of 83.33%. The model performs well in terms of minimizing false positives and capturing true positives.

Thank you!

Appendix – GitHub Links

- Lab 1 - Complete the Data Collection API Lab
  - https://github.com/aarondukepifanio/ibmdatascience/blob/main/Lab%201%20-%20Complete%20the%20Data%20Collection%20API%20Lab.ipynb
- Lab 2 - Web Scraping Falcon 9 and Falcon Heavy Launches Records from Wikipedia
  - https://github.com/aarondukepifanio/ibmdatascience/blob/main/Lab%202%20-%20Web%20Scraping%20Falcon%209%20and%20Falcon%20Heavy%20Launches%20Records%20from%20Wikipedia%C2%B6.ipynb
- Lab 3 - Data Wrangling
  - https://github.com/aarondukepifanio/ibmdatascience/blob/main/Lab%203%20-%20Data%20Wrangling.ipynb
- Lab 4 - Exploratory Data w SQL
  - https://github.com/aarondukepifanio/ibmdatascience/blob/main/Lab%204%20-%20Exploratory%20Data%20w%20SQL.ipynb
- Lab 5 - EDA Visualization
  - https://github.com/aarondukepifanio/ibmdatascience/blob/main/Lab%205%20-%20EDA%20Visualization.ipynb
- Lab 6 - Folium Map
  - https://github.com/aarondukepifanio/ibmdatascience/blob/main/Lab%206%20-%20Folium%20Map.ipynb
- Lab 6.5 – SpaceX Dash App
  - https://github.com/aarondukepifanio/ibmdatascience/blob/main/Lab_6.5_spacex_dash_app.py
- Lab 7 - Machine Learning
  - https://github.com/aarondukepifanio/ibmdatascience/blob/main/Lab%207%20-%20Machine%20Learning.ipynb