SpaceX: A Data Science Odyssey

Mickey A. Couvertier
May 2023

Outline

- Executive Summary
- Introduction
- Methodology
- Findings
- Conclusion
- Appendix

Executive Summary

- The data was gathered using the SpaceX API and by scraping Wikipedia. Pandas and Numpy were used to normalize data and data types, and to transform the data for analysis. Various methods were used for analysis including in an SQL database, in Python using Matplotlib and Seaborn, and by building an interactive dashboard using Plotly. We analyzed location data using Folium maps and annotated landmarks close to each launch site.

- The data was ultimately evaluated and standardized using Sci-Kit Learn, which we also used to evaluate 4 classification models. Each model was optimized to the best hyperparameters, and we were ultimately successful in achieving a standard 83.3% accuracy across the board, with similar results in the confusion matrix. Each model was far more successful in accurately predicting a successful landing than failed landings per the confusion matrices for each model.

Introduction

- In recent years, SpaceX has become a leader in the space industry, achieving historic milestones and pushing the boundaries of what is possible in space exploration. One of SpaceX's most notable accomplishments is their ability to reuse the first stage of their F9 rockets, which has led to significant cost savings compared to other providers in the industry. However, determining whether the first stage will land successfully remains a challenge.

- The goal of this project is to use data science techniques to analyze the SpaceX dataset, which includes a record for each payload carried during a SpaceX mission into outer space and determine if there are any factors that can help predict the successful landing of the first stage. By doing so, we aim to provide insights that can be used to determine the cost of a launch and used as a reference for alternate companies that want to bid against SpaceX for a rocket launch.

- Through this project, we hope to gain a better understanding of the factors that contribute to the success of a SpaceX mission and provide valuable insights for the space industry. We will use a variety of data science techniques, including exploratory data analysis, machine learning algorithms, and data visualization, to analyze the SpaceX dataset and uncover hidden patterns and relationships in the data.

Section 1

Methodology

- Executive Summary
- Data collection methodology:
  - Parts of the data were collected using the SpaceX API and the rest was scraped from Wikipedia
- Perform data wrangling
  - Predefined methods were used to ensure specific data was included and formatted as required. We further analyzed the data looking for patterns and creating new features
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection

The data was collected using the SpaceX API. First, we defined functions that help us use the API to extract information using identification numbers in the launch data.

Next, we request the data from the API which is returned as a JSON object. We normalize this data and create a data frame. We use the functions we defined to get more information from the API such as launch site, payload, booster version, landing outcome, and more. Finally, we filter the data to only list F9 launches.

We then got more data on heavy payload launches by scraping the Wikipedia page for Falcon 9 and Falcon Heavy launches. We then used BeautifulSoup to parse the data, extract tables, and create new columns using the table header names (th).

Request launch data → SpaceX API returns JSON Object → Normalize and filter data → Save dataset_part_1.csv

Enter URL for scraping → Parse HTML using BeautifulSoup → Extract table names as columns → Save spacex.csv

Data Collection – SpaceX API

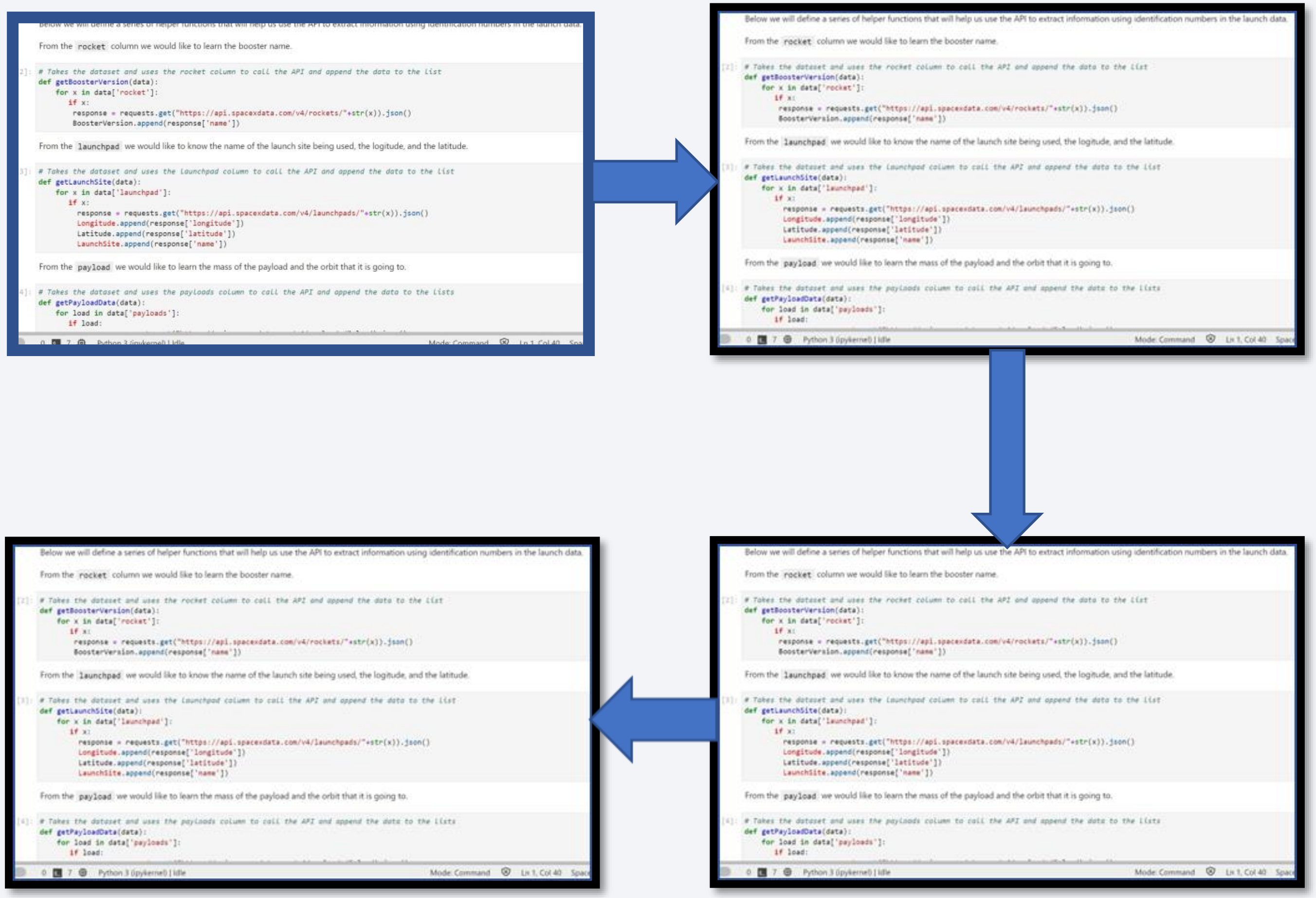
First we define some methods used to extract the data we want

We then request the data from the SpaceX API

The returned json data is normalized and our defined methods are applied

After some cleaning we export the data as dataset_part_1.csv

[Github Link to SpaceX API Notebook]



Below we will define a series of helper functions that will help us use the API to extract information using identification numbers in the launch data.

From the `rocket` column we would like to learn the booster name.

```
[2]: # Takes the dataset and uses the rocket column to call the API and append the data to the list
def getBoosterVersion(data):
    for x in data['rocket']:
       if x:
        response = requests.get("https://api.spacexdata.com/v4/rockets/"+str(x)).json()
        BoosterVersion.append(response['name'])
```

From the `launchpad` we would like to know the name of the launch site being used, the logitude, and the latitude.

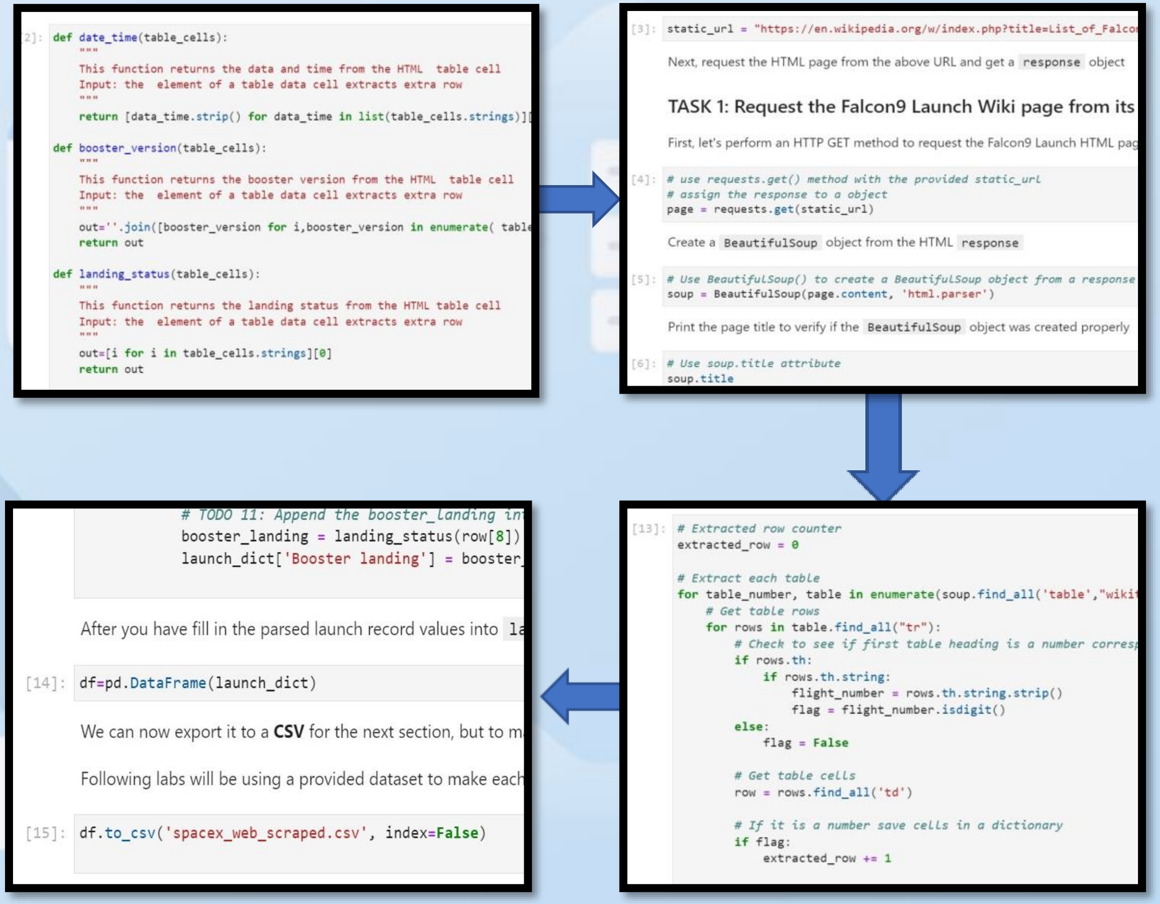
```
[3]: # Takes the dataset and uses the launchpad column to call the API and append the data to the list
def getLaunchSite(data):
    for x in data['launchpad']:
       if x:
         response = requests.get("https://api.spacexdata.com/v4/launchpads/"+str(x)).json()
         Longitude.append(response['longitude'])
         Latitude.append(response['latitude'])
         LaunchSite.append(response['name'])
```

From the `payload` we would like to learn the mass of the payload and the orbit that it is going to.

```
[4]: # Takes the dataset and uses the payloads column to call the API and append the data to the lists
def getPayloadData(data):
    for load in data['payloads']:
       if load:
```

Data Collection - Scraping

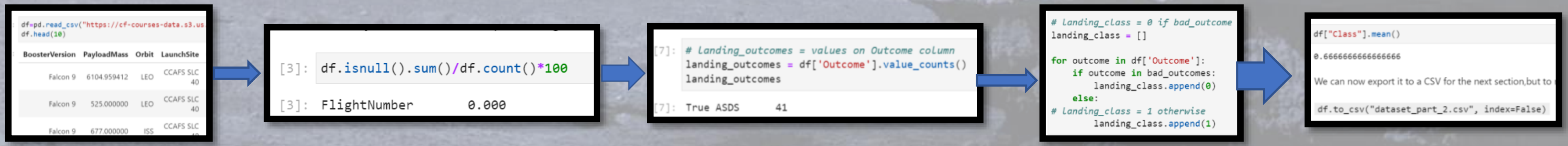
- First we define methods for some basic data wrangling
- Next we request the Falcon 9 launches Wikipedia page using the HTTP GET method
- Then we create a Soup object using BeautifulSoup
- We then use our predefined methods to further parse and categorize our data
- Finally we export a file as spacex.csv



[SpaceX Webscraping Notebook]()

Data Wrangling

- Use Pandas to import dataset_part_1.csv, which is the data we extracted in the SpaceX API notebook, as a dataframe and use Numpy to look at our null values, value counts, and dtypes to better understand our data
- Next we clean the data by addressing null values and fixing dtypes as needed
- We then transform the data by doing some feature engineering. We create new columns by aggregating existing ones and merging the data
- We look for patterns in the data, including statistics like the launch success rate
- Once we understand our data and use this knowledge to create features within our data, it is prepared for further analysis as dataset_part_2.csv



[SpaceX Data Wrangling Notebook]

EDA with Data Visualization

OUR FIRST SCATTERPLOT SHOWS US THE LAUNCH NUMBER ON THE X AXIS AND PAYLOAD MASS ON THE Y AXIS. WE CAN SEE A CORRELATION BETWEEN PAYLOAD MASS, FLIGHT NUMBER, AND SUCCESS RATE.

WE USE A BAR CHART TO OBSERVE THE RELATIONSHIP BETWEEN LAUNCH ORBIT AND SUCCESS RATE. WE SEE SOME ORBITS WITH A MUCH HIGHER SUCCESS RATE THAN OTHERS.

ANOTHER SCATTERPLOT SHOWS THE RELATIONSHIP BETWEEN PAYLOAD MASS AND ORBIT TYPE. ISS ORBITS, FOR EXAMPLE, AVERAGE 2000KG PAYLOAD MASS.

A LINE PLOT SHOWS US THE YEARLY TREND OF SUCCESSFUL LAUNCHES, WITH THE LAST 5 YEARS OR SO HOVERING AROUND A 100% SUCCESS RATE.

SPACEX EDA WITH DATA VISUALIZATION

EDA with SQL

We log into IBM Db2 and load our spacex.csv into a table. We use the ibm_db_sa API and SQL Magic to run queries in our Jupyter Notebook in Python. We ran the following:

- Names of the unique launch sites in the space mission
- 5 records where launch sites begin with the string 'CCA’
- Total payload mass carried by boosters launched by NASA (CRS)
- Average payload mass carried by booster version F9 v1.1
- Date when the first successful landing outcome in ground pad was achieved
- Names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- Total number of successful and failure mission outcomes
- Names of the booster_versions which have carried the maximum payload mass
- Failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015
- Rank the count of landing outcomes between the date 2010-06-04 and 2017-03-20, in descending order

[SpaceX EDA with SQL]

Build an Interactive Map with Folium

First we mark every SpaceX F9 launch location with a 1km radius circle and popup marker showing the name of each launch station.

Next we create a MarkerCluster object we'll use to map each launch attempt to its launch site as well as a green or red popup denoting a successful or failed attempt, respectively.

Finally, we calculate the distances to different landmarks such as the nearest coastline, the nearest railroad, the nearest highway, and the nearest city. We create markers for the location of each of those objects in reference to each launch location, and we create a line marking the distance in kilometers from each launch site to each nearest landmark.

[SpaceX Location Analysis with Folium]

Build a Dashboard with Plotly Dash

- We first create our dashboard title ‘SpaceX Launch Records Dashboard’ by creating an HTML division and giving it an H1 header styled to the center.
- We create a dropdown with each launch site as an option, and all launch sites as the default option.
- We create a division for a pie chart, a paragraph header with ‘Payload range (Kg):’ for our range slider object, we create the range slider division and define our range slider behavior, and finally a scatter plot division.
- We create a callback object for our pie chart with the dropdown as our input component, and success-pie-chart as our output component. We then define a pie chart that shows all the launch sites by success rate as the default and shows the success vs failure rate in percentage and color – green and red, respectively – when a launch site is chosen from the dropdown.
- Lastly, we create a callback object for the scatter plot with both the payload slider and the launch site dropdown as input components, and success-payload-scatter-chart as the output component. We map each booster version to a color using the colors.qualitative.Alphabet method from Plotly Express, and we create a legend giving the booster version for each color assigned.

SpaceX Plotly Dashboard Script

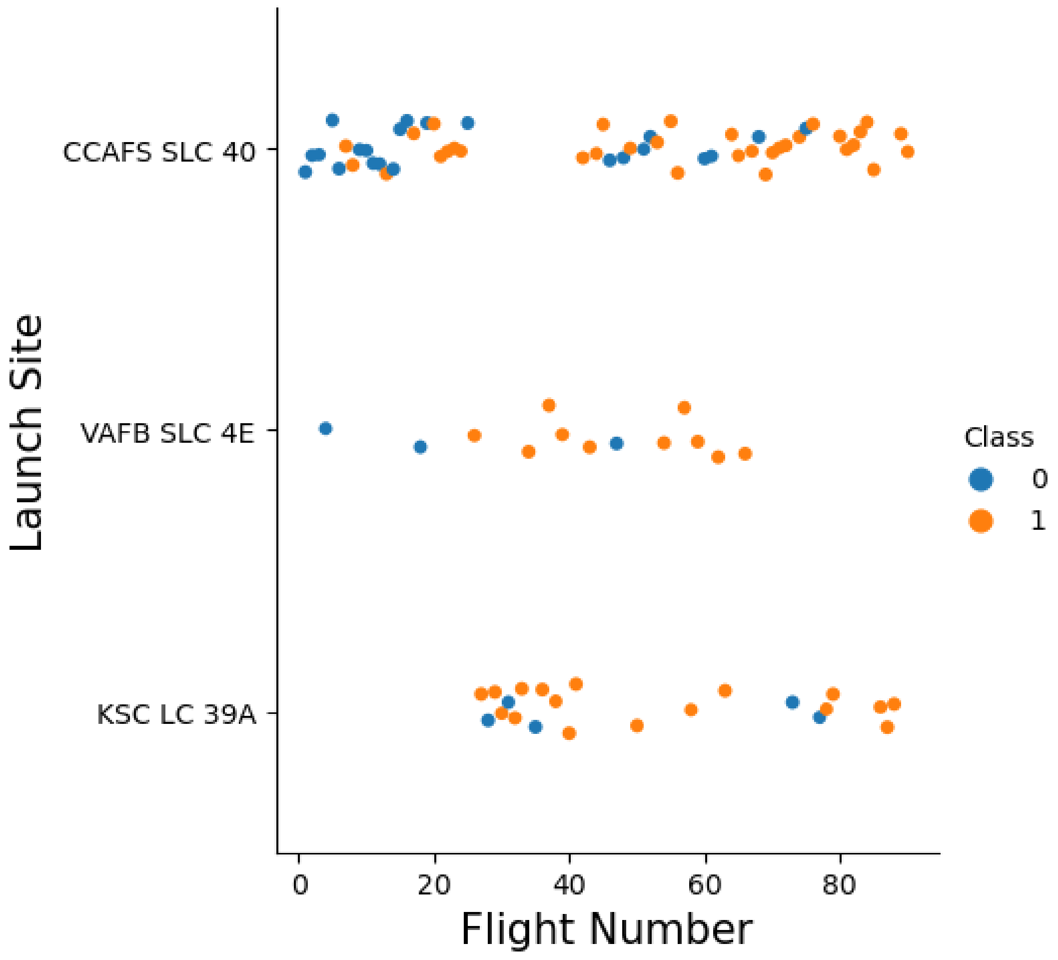
Predictive Analysis (Classification)

- Our end goal is to predict the probability of a SpaceX F9 rocket having a successful stage 1 landing and recovery.
- We will evaluate logistic regression, support vector machines, decision tree classifier, and K nearest neighbor algorithms in Sci-Kit Learn.
- We used the pre-processing library in SK Learn to standardize our data, train_test_split to create our train and test data, and GridSearchCV to test the best parameters for and ultimately the best classification model.
- We ran each model individually using different hyperparameters and compared against a confusion matrix. All 4 models tested performed equally with 83.3% accuracy.

Preprocessing and Standardization → Split train and test data → Create GridSearchCV object to test each model → Run each algorithm and tune hyperparameters → Evaluate and adjust until optimized

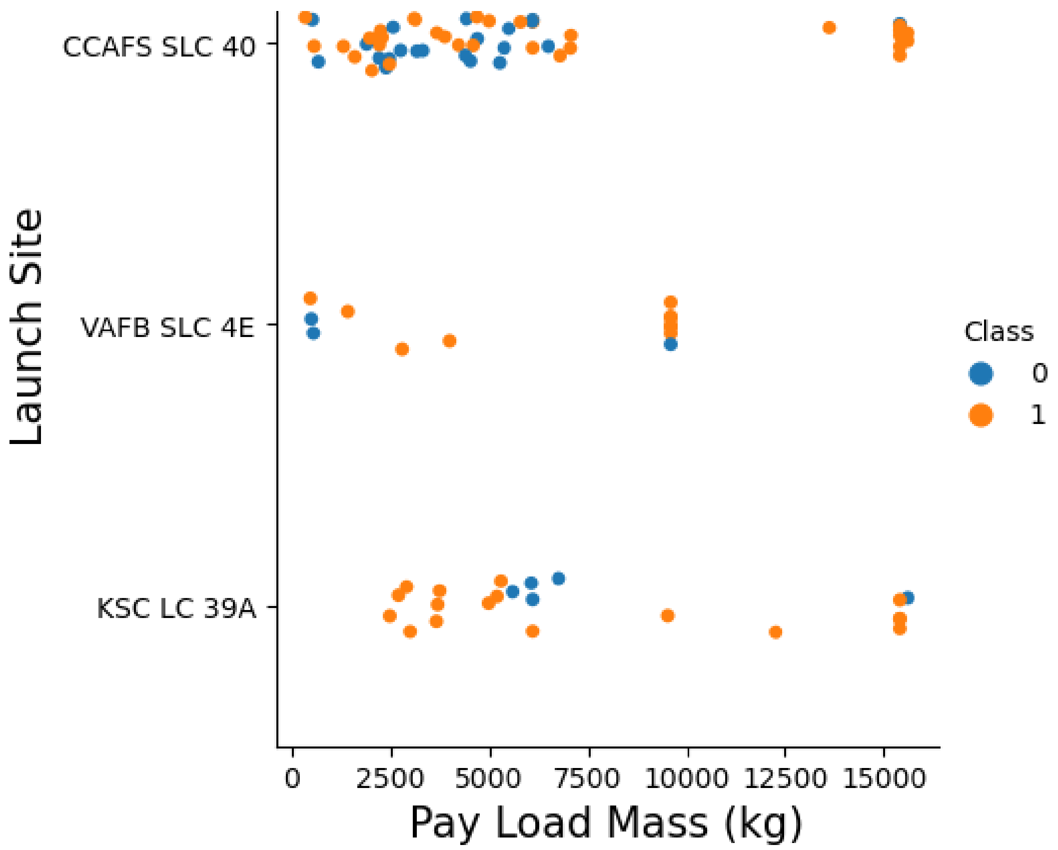
SpaceX Predictive Analysis Github Link

Section 2
Findings



Flight Number vs. Launch Site

- Earlier launches were all at CCAFS SLC 40 and had a high landing fail rate.
- Around the 20^{th} launch the success rate seems to have improved. The earliest successful landing was queried in another slide.
- CCAFS SLC 40 has more failed landings than the other launch sites – this may be due to it having the most launches total.

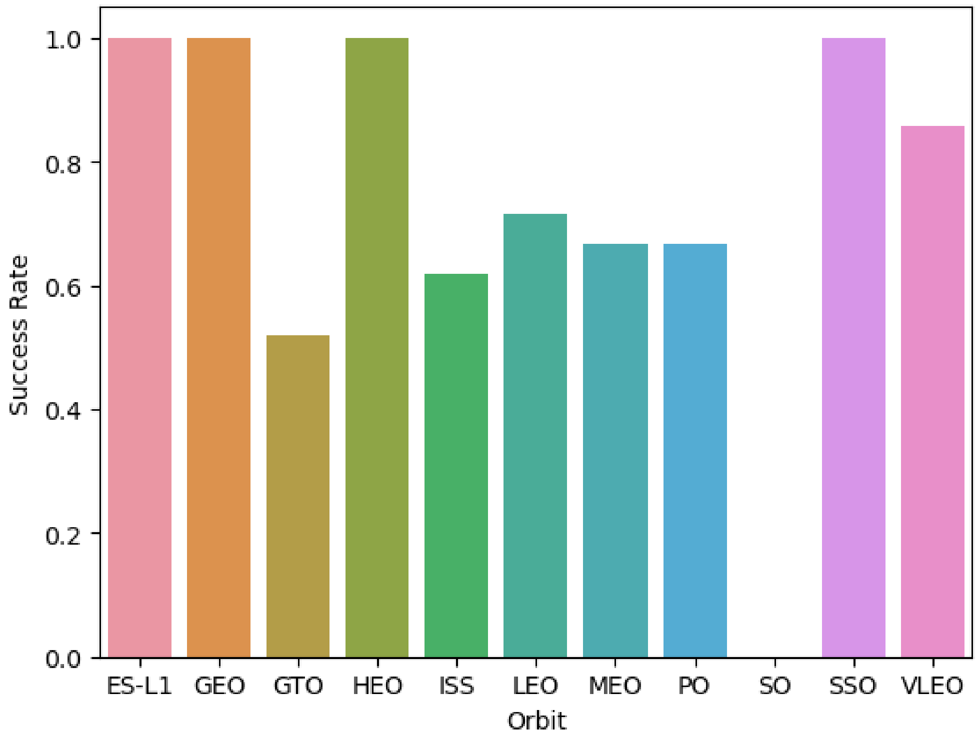


Payload vs. Launch Site

- CCAFS SLC 40 shows a payload range of about 100-7500kg, with an oulier at 12500, and a cluster of launches at 15000kg
- VAFB SLC 4E has had a few launches in the 100-5000kg range, but most of its launches have 10000kg of payload
- KSC LC 39A has no launches below 2500kg, has had failed landings at around 6000kg and one at 15000kg making it highly successful in most payload ranges.

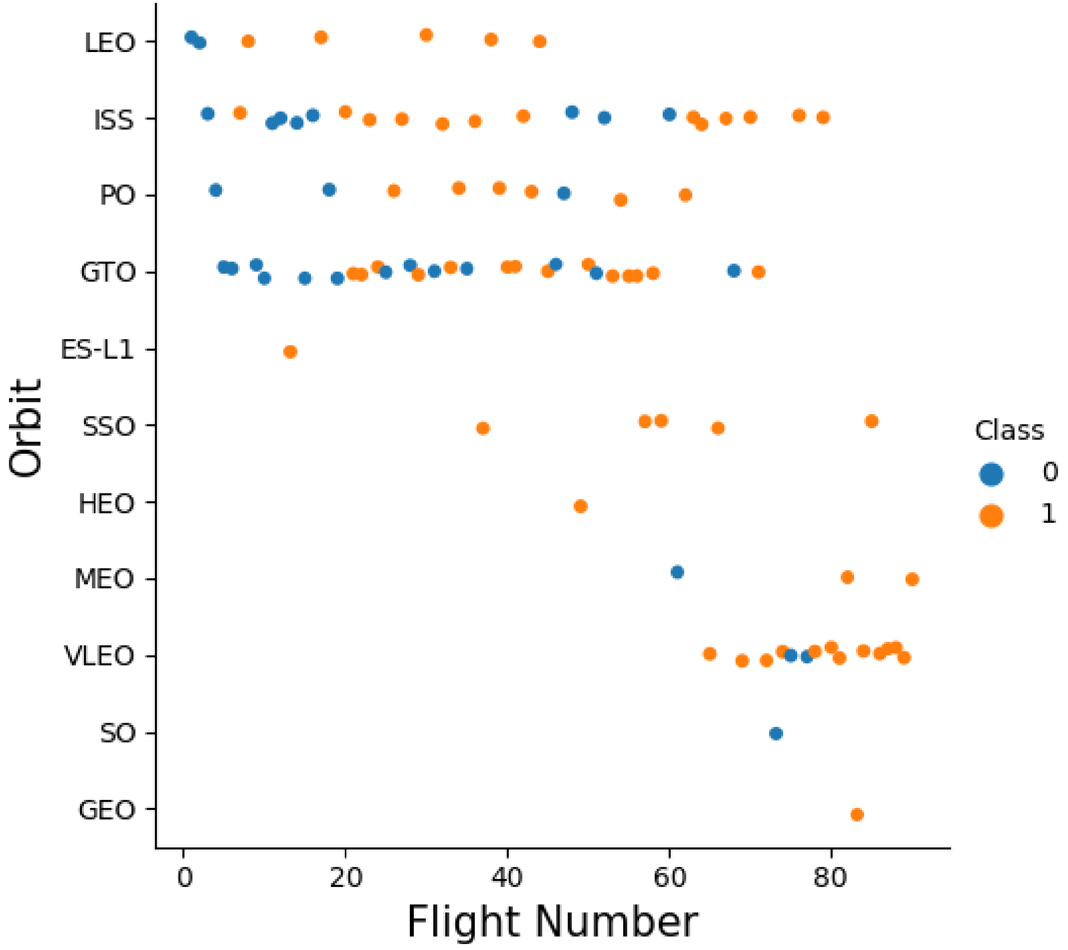
Success Rate vs. Orbit Type

- We see that four of the orbit types have 100% success rates
- VLEO has a lower than 100% success rate but still higher than the rest – we may later see some data explaining this better
- We see GTO has the lowest success rate
- Many of the early launches were ISS – this is one explanation for having the 2nd lowest success rate



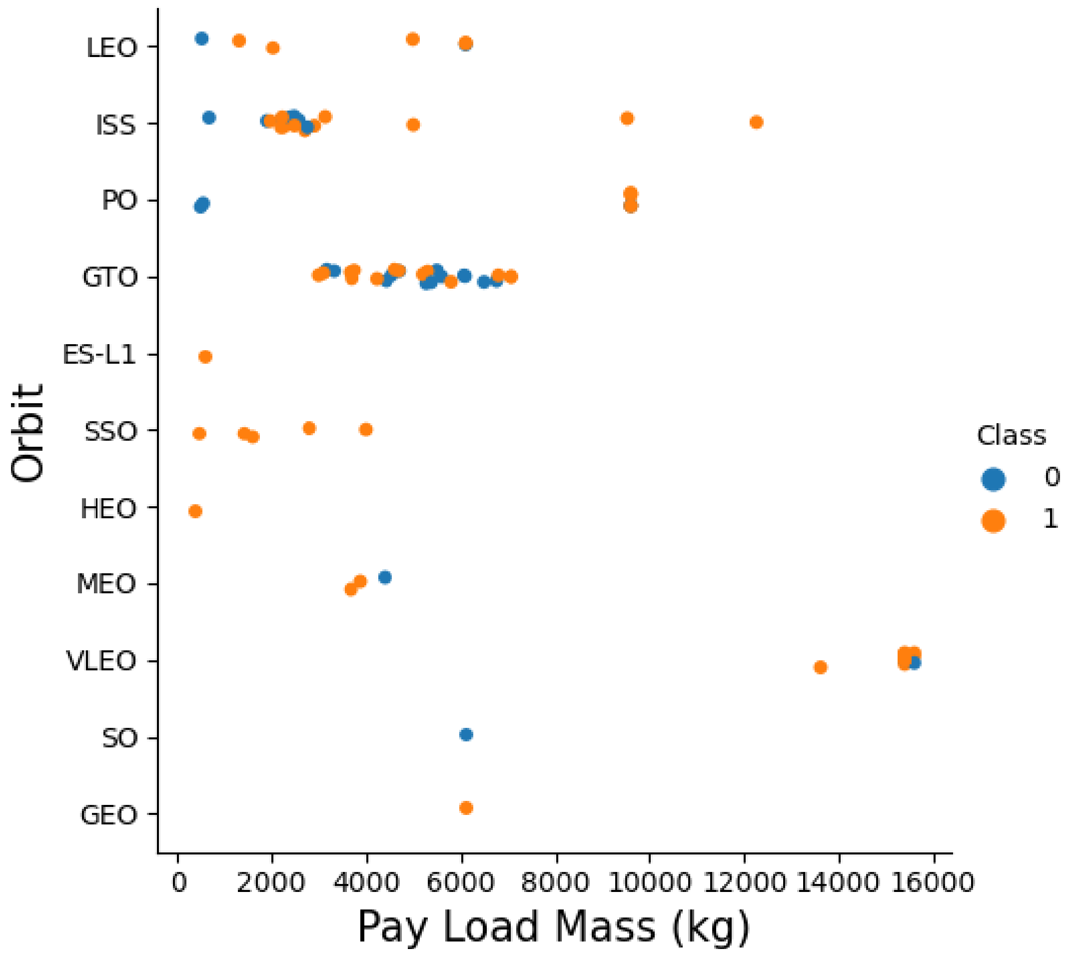
Flight Number vs. Orbit Type

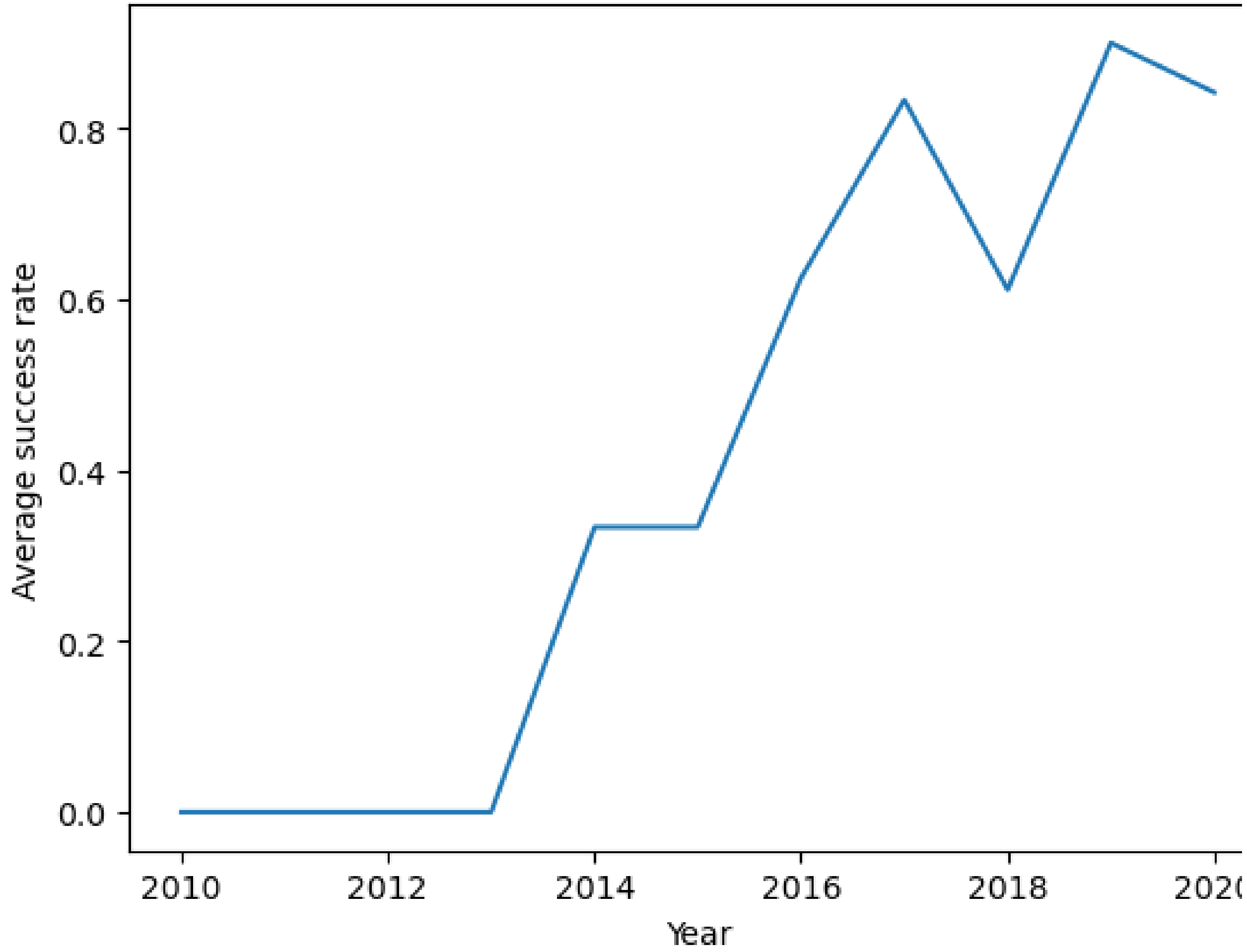
- We can see an improvement across the board with regard to flight number and success except with the GTO orbit, which has a more mixed success rate
- All SSO missions have been successful, as have all HEO and ES-L1 orbit type missions, and the one GEO mission recorded
- VLEO missions have a high success rate – they also account for most of the recent missions



Payload vs. Orbit Type

- SSO has a payload range of about 500-4000kg with a 100% success rate
- GTO has a range of about 2500kg-8000kg and a very mixed success rate, but also most launches in that payload range are of this orbit type
- PO, LEO, and ISS have better success rates with higher payloads
- VLEO, again, has a high success rate, but also seems to carry the heaviest loads





Launch Success Yearly Trend

- The landing success rate drop for 2018 is due to a Falcon Heavy launch that resulted in 2 of 3 boosters being recovered. The two outer boosters landed, but the core, a third booster, failed to land. Another launch later that year failed to recover the boosters
- SpaceX started landing attempts in 2013, and had its first successful attempt in December 2015
- Since 2016 SpaceX has dramatically improved its success rate in recovering boosters

All Launch Site Names

This SQL query selects the distinct launch sites from the SpaceX table. The four launch sites listed in the results are:

- 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.
- KSC LC-39A: Kennedy Space Center Launch Complex 39A, located in Florida, USA.
- VAFB SLC-4E: Vandenberg Air Force Base Space Launch Complex 4E, located in California, USA.

```
%sql SELECT DISTINCT(LAUNCH_SITE) FROM SPACEX;
```

```
 * ibm_db_sa://your-user-name:your-password@your-ho
Done.
```

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

Launch Site Names Begin with 'CCA'

- This SQL query selects the first 5 records from the SpaceX table where the launch site begins with the string 'CCA’
- We can see that all 5 launch sites begin with the string 'CCA', and they are all located in Cape Canaveral Air Force Station Launch Complex 40, Florida, USA

```
%sql SELECT * FROM SPACEX WHERE LAUNCH_SITE like 'CCA%' LIMIT 5;
```

```
* ibm_db_sa://ghv31291:***@6667d8e9-9d4d-4ccb-ba32-21da3bb5aafc.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:30376/bludb?security=SSL
Done.
```

DATE	time_utc	booster_version	launch_site	payload	payload_mass_kg_	orbit	customer	mission_outcome	landing_outcome
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-12-08	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

- This SQL query calculates the total payload mass in kilograms launched by SpaceX for NASA under the CRS program
- The results show that the total payload mass for NASA under CRS program is 45596 kg

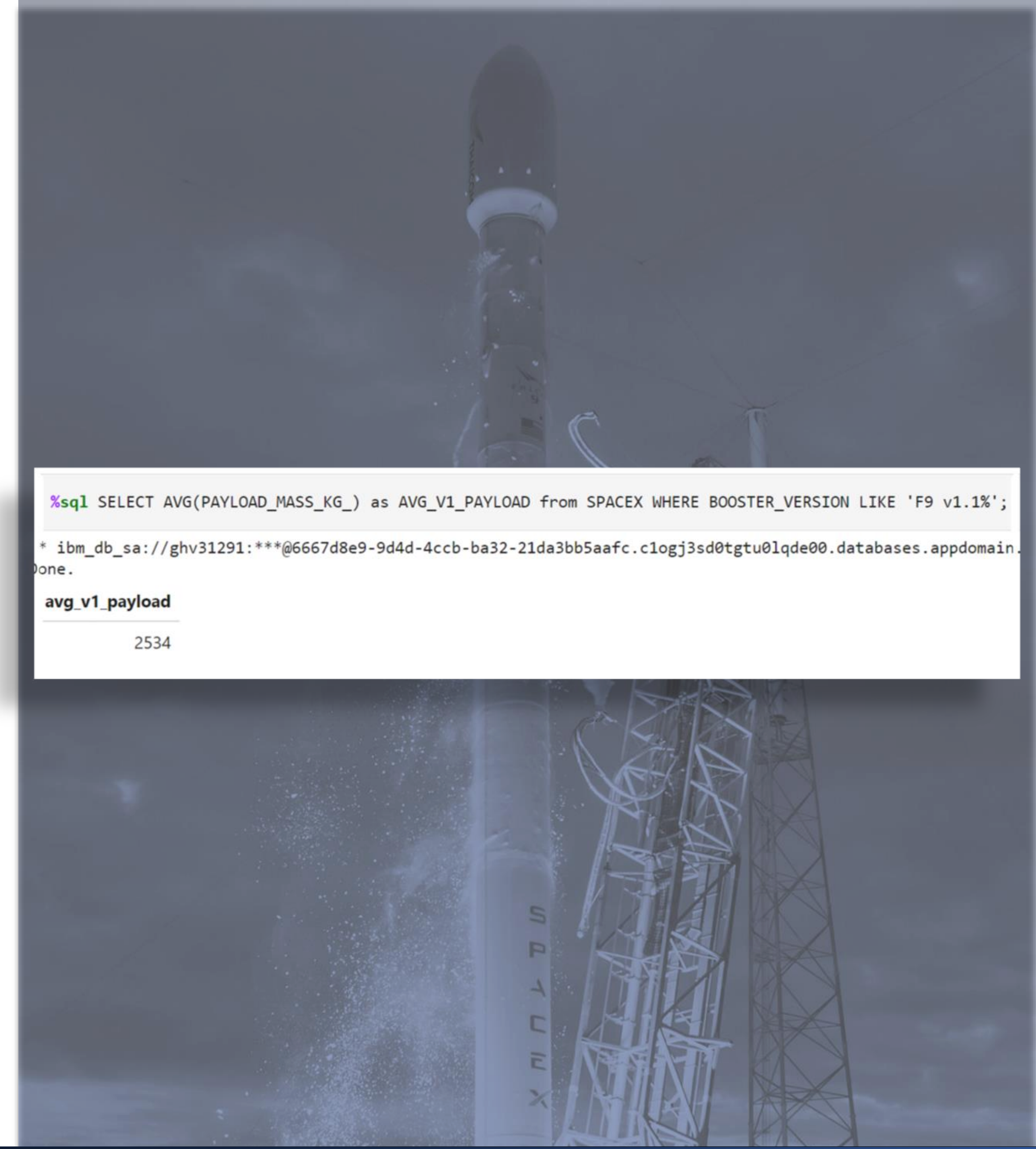
```
%sql SELECT SUM(PAYLOAD_MASS_KG_) as Total_NASA_Payload from SPACEX WHERE CUSTOMER = 'NASA (CRS)';
```

```
 * ibm_db_sa://ghv31291:***@6667d8e9-9d4d-4ccb-ba32-21da3bb5aafc.c1ogj3sd0tgtu0lqde00.databases.appdoma
Done.
```

total_nasa_payload
45596

Average Payload Mass by F9 v1.1

- This SQL query calculates the average payload mass in kilograms for all SpaceX launches where the booster version starts with 'F9 v1.1’
- The query is structured to select the average payload mass by using the AVG() function to calculate the average of the PAYLOAD_MASS_KG_ column
- The WHERE clause is used to filter the results to only include launches where the booster version starts with 'F9 v1.1’
- The result of the query is a single value, which represents the average payload mass in kilograms for all SpaceX launches where the booster version starts with 'F9 v1.1'. The result is 2534 kg, which means that on average, the payload mass for these launches was 2534 kg



```
%sql SELECT AVG(PAYLOAD_MASS_KG_) as AVG_V1_PAYLOAD from SPACEX WHERE BOOSTER_VERSION LIKE 'F9 v1.1%';
* ibm_db_sa://ghv31291:***@6667d8e9-9d4d-4ccb-ba32-21da3bb5aafc.c1ogj3sd0tgtu0lqde00.databases.appdomain.
Done.
```

avg_v1_payload
2534

First Successful Ground Landing Date

- This SQL query selects the earliest date of a successful landing outcome for SpaceX's booster landings
- The query is structured to select the minimum date from the SpaceX table where the landing outcome starts with the string 'Success' using the MIN() function
- The result of the query is a single value, which represents the earliest successful landing outcome for SpaceX's booster landings
- The result is 2015-12-22, which means that the first successful landing of a SpaceX booster occurred on December 22, 2015

This landing was accomplished during the ORBCOMM-2 mission, in which the first stage of the Falcon 9 rocket successfully landed on a ground pad at Landing Zone 1 (formerly known as Landing Complex 1) at Cape Canaveral, Florida, USA. Prior to this landing, SpaceX had attempted several unsuccessful landing attempts of their first stage booster, making this the first successful landing and a significant milestone for the company in their efforts to develop reusable rockets.

```
%sql SELECT MIN(DATE) from SPACEX WHERE LANDING_OUTCOME LIKE 'Success%';
```

```
* ibm_db_sa://ghv31291:***@6667d8e9-9d4d-4ccb-ba32-21da3bb5aafc.c1ogj3sd0tg
one.
```

1
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

This SQL query selects the distinct booster versions used in SpaceX launches where the landing outcome was 'Success (drone ship)' and the payload mass was between 4000 and 6000 kg. The query is structured to select the distinct booster versions using the DISTINCT keyword and filter the results using the WHERE clause with three conditions:

- LANDING_OUTCOME = 'Success (drone ship)': This condition filters the results to only include successful landings on a drone ship
- PAYLOAD_MASS_KG_ > 4000: This condition filters the results to only include launches with a payload mass greater than 4000 kg
- PAYLOAD_MASS_KG_ < 6000: This condition filters the results to only include launches with a payload mass less than 6000 kg

```
%%sql SELECT DISTINCT BOOSTER_VERSION
FROM SPACEX
WHERE LANDING_OUTCOME = 'Success (drone ship)'
AND PAYLOAD_MASS_KG_ > 4000
AND PAYLOAD_MASS_KG_ < 6000;

* ibm_db_sa://ghv31291:***@6667d8e9-9d4d-4ccb-ba
one.
```

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

These booster versions were used in launches that met the filter criteria, which included successful landings on a drone ship with a payload mass between 4000 and 6000 kg. These landings were a significant milestone for SpaceX as they demonstrated the company's ability to recover and reuse their rocket boosters, which can significantly reduce the cost of spaceflight.

Total Number of Successful and Failure Mission Outcomes

This SQL query selects the total number of successful and failed missions in the SpaceX table. The query is structured to select the total number of successful and failed missions using the SUM() function and the CASE statement with two conditions:

- WHEN MISSION_OUTCOME LIKE '%Success%' THEN 1 ELSE 0 END: This condition increments the counter by 1 if the mission outcome contains the word 'Success’
- WHEN MISSION_OUTCOME LIKE '%Failure%' THEN 1 ELSE 0 END: This condition increments the counter by 1 if the mission outcome contains the word 'Failure’

The results of the query show that there were 100 successful missions and 1 failed mission in the SpaceX table. This means that the success rate for SpaceX missions in this table is approximately 99%.

It is worth noting that this query only considers the missions that are present in the SpaceX table, which may not be a comprehensive representation of all SpaceX missions. Additionally, the mission outcomes listed in the table are not always mutually exclusive, meaning that some missions may have multiple outcomes listed, and the categorization of successful or failed may be subjective. Our overall focus is on landing outcomes, which are different from mission outcomes, such as with the failed landings in 2018 mentioned earlier.

```
%%sql SELECT
SUM(CASE WHEN MISSION_OUTCOME LIKE '%Success%' THEN 1 ELSE 0 END) AS Total_Successful_Missions,
SUM(CASE WHEN MISSION_OUTCOME LIKE '%Failure%' THEN 1 ELSE 0 END) AS Total_Failed_Missions
FROM SPACEX;
```

```
* ibm_db_sa://ghv31291:***@6667d8e9-9d4d-4ccb-ba32-21da3bb5aafc.c1ogj3sd0tgtu0lqde00.databases.a
one.
```

total_successful_missions	total_failed_missions
100	1

Boosters Carried Maximum Payload

Booster versions which have carried the max payload:

booster_version
F9 B5 B1048.4
F9 B5 B1048.5
F9 B5 B1049.4
F9 B5 B1049.5
F9 B5 B1049.7
F9 B5 B1051.3
F9 B5 B1051.4
F9 B5 B1051.6
F9 B5 B1056.4
F9 B5 B1058.3
F9 B5 B1060.2
F9 B5 B1060.3

```
%sql SELECT DISTINCT(BOOSTER_VERSION) FROM SPACEX WHERE PAYLOAD_MASS_KG_ = (SELECT MAX(PAYLOAD_MASS_KG_) from SPACEX);
```

This SQL query selects the distinct booster versions used in SpaceX launches where the payload mass was the maximum payload mass in the table. The query is structured to select the distinct booster versions using the DISTINCT keyword and filter the results using the WHERE clause with a subquery:

- (SELECT MAX(PAYLOAD_MASS_KG_) from SPACEX): This subquery selects the maximum payload mass in the SpaceX table

The results show twelve booster versions that meet the filter criteria. These booster versions were used in launches that carried the maximum payload mass

2015 Launch Records

This SQL query selects the landing outcome, booster version, and launch site for SpaceX launches in 2015 where the landing outcome was a failure on a drone ship. The query is structured to select the landing outcome, booster version, and launch site using the SELECT statement and filter the results using the WHERE clause with two conditions:

- LANDING_OUTCOME LIKE '%Failure (drone ship)%': This condition filters the results to only include landing outcomes that contain the string 'Failure (drone ship)'.
- YEAR(DATE) = 2015: This condition filters the results to only include launches that occurred in the year 2015.
- The result of the query is a table with two rows, each representing a failed landing attempt on a drone ship in 2015. The table shows the landing outcome, booster version, and launch site for each failed landing attempt.
- Both landing attempts occurred at the Cape Canaveral Air Force Station Launch Complex 40 in Florida, USA, and involved the Falcon 9 version 1.1 boosters with the booster versions of F9 v1.1 B1012 and F9 v1.1 B1015, respectively.

These landing failures were significant events for SpaceX as they demonstrated the challenges of landing on a drone ship in rough seas and helped inform future landing attempts as they were partially attributed to the rough seas encountered during the landing attempts.

```
%%sql SELECT LANDING_OUTCOME, BOOSTER_VERSION, LAUNCH_SITE
FROM SPACEX
WHERE LANDING_OUTCOME LIKE '%Failure (drone ship)%' AND YEAR(DATE) = 2015;
```

```
* ibm_db_sa://ghv31291:***@6667d8e9-9d4d-4ccb-ba32-21da3bb5aafc.c1ogj3sd0tg
one.
```

landing_outcome	booster_version	launch_site
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

The query is structured to select the landing outcome and count of occurrences using the SELECT statement, group the results by landing outcome using the GROUP BY clause, and sort the results by count of occurrences in descending order using the ORDER BY clause.

The table shows that the most common landing outcome during this period was 'No attempt', which occurred 10 times. This is followed by 'Failure (drone ship)' and 'Success (drone ship)', each occurring 5 times. The other landing outcomes listed in the table occurred less frequently during this period

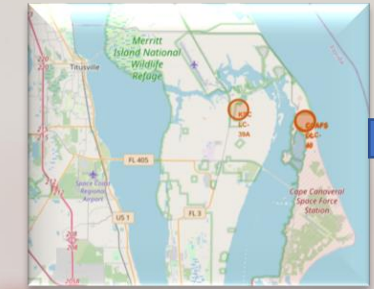
landing_outcome	COUNT
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

```
%%sql
SELECT LANDING_OUTCOME, COUNT(*) AS count
FROM SPACEX
WHERE DATE >= '2010-06-04' AND DATE <= '2017-03-20'
GROUP BY LANDING_OUTCOME
ORDER BY count DESC;
```

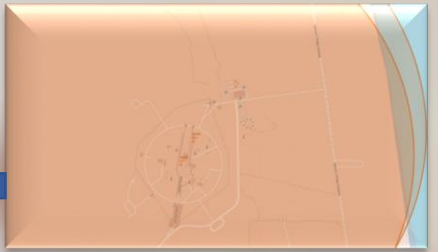
Section 3

Launch Sites Proximities Analysis

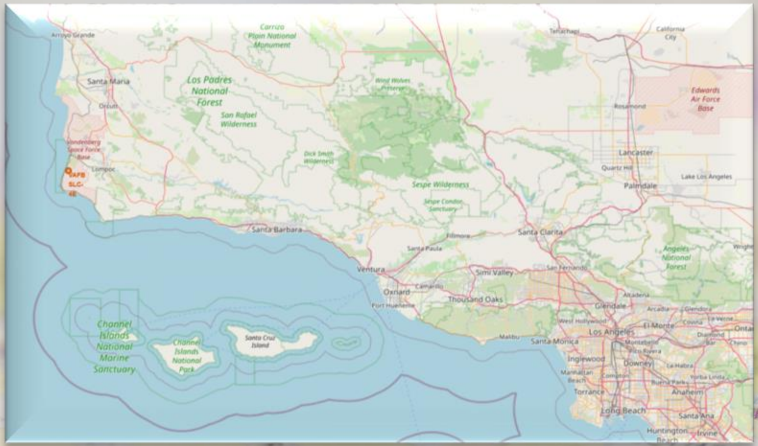
Launch Site Locations on Map



- KSC LC-39A: Kennedy Space Center Launch Complex 39A
- CCAFS LC-40: Cape Canaveral Air Force Station Launch Complex 40
- CCAFS SLC-40: Cape Canaveral Air Force Station Space Launch Complex 40



Here we can see the proximity of CCAFS LC-40 to CCAFS SLC-40



VAFB SLC-4E in California not too far from Los Angeles

CCAFS LC-40 and CCAFS SLC-40 both refer to launch complexes located at the Cape Canaveral Air Force Station in Florida.

CCAFS LC-40 stands for "Cape Canaveral Air Force Station Launch Complex 40". This launch complex was originally built for the Titan rocket program in the 1960s and was later leased to SpaceX in 2007. SpaceX has since used the launch complex for many of its Falcon 9 and Falcon Heavy launches, including the launch of the Dragon spacecraft to the International Space Station (ISS) as part of NASA's Commercial Resupply Services program.

CCAFS SLC-40 stands for "Cape Canaveral Air Force Station Space Launch Complex 40“. It was also built for the Titan program and later leased to SpaceX. Both launch complexes have been key facilities for SpaceX's launch operations, particularly for missions to the ISS and other orbits.

KSC LC-39A is the Kennedy Space Center Launch Complex 39A located in Florida. This launch complex was originally built in the 1960s for the Apollo program and later modified for the Space Shuttle program. After the retirement of the Space Shuttle program, the launch complex was leased to SpaceX in 2014. SpaceX made significant modifications to the complex to support their Falcon 9 and Falcon Heavy rockets, and the launch complex has since been used for many SpaceX launches, including the Falcon Heavy test flight and the launch of NASA astronauts to the International Space Station as part of the Commercial Crew Program.

VAFB SLC-4E is the Vandenberg Air Force Base Space Launch Complex 4E located in California, USA. This launch complex was originally built for the Titan III and Atlas-Centaur rocket programs in the 1960s and has since been modified for various launch vehicles, including the Delta II and Delta IV rockets. SpaceX has also used this launch complex for some of its Falcon 9 launches, including the Iridium NEXT communications satellites and the Sentinel-6 Michael Freilich oceanography satellite.

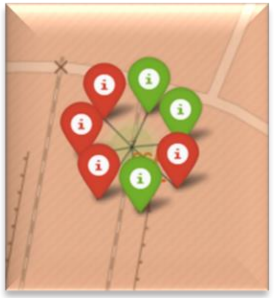
All four launch sites are near the coastline and are close enough to cities with large hospitals systems and good response in case of emergencies. Coastal cities also have greater tolerance for noise and noise level ordinances that accommodate the launches.

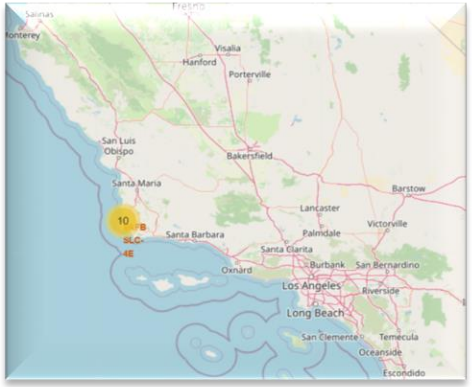
Mapping Launch Outcomes

There were a total of 33 launches for CCAFS LC-40 and CCAFS SLC-40

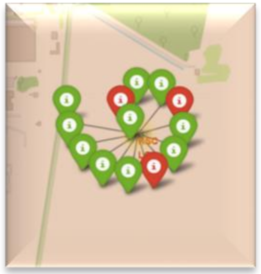
By now we should see a few patterns. KSC wasn't leased to SpaceX until 2014 and SpaceX had no successful landing attempts until December 2015 at CCAFS. CCAFS has a 30.3% success rate compared to a 23% fail rate for KSC. VAFB also has a low success rate at 40% due largely to failed ocean landings.

Here we see 7 of those were launched from SLC-40, and the other 26 from LC-40

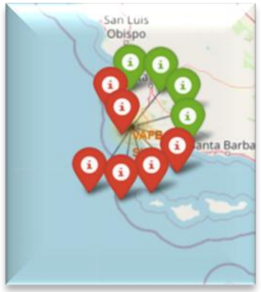




KSC LC-39A has a total of 13 launches



There were a total of 10 launches recorded for VAFB SLC-4E and from our markers we can observe a 60% failed landings.

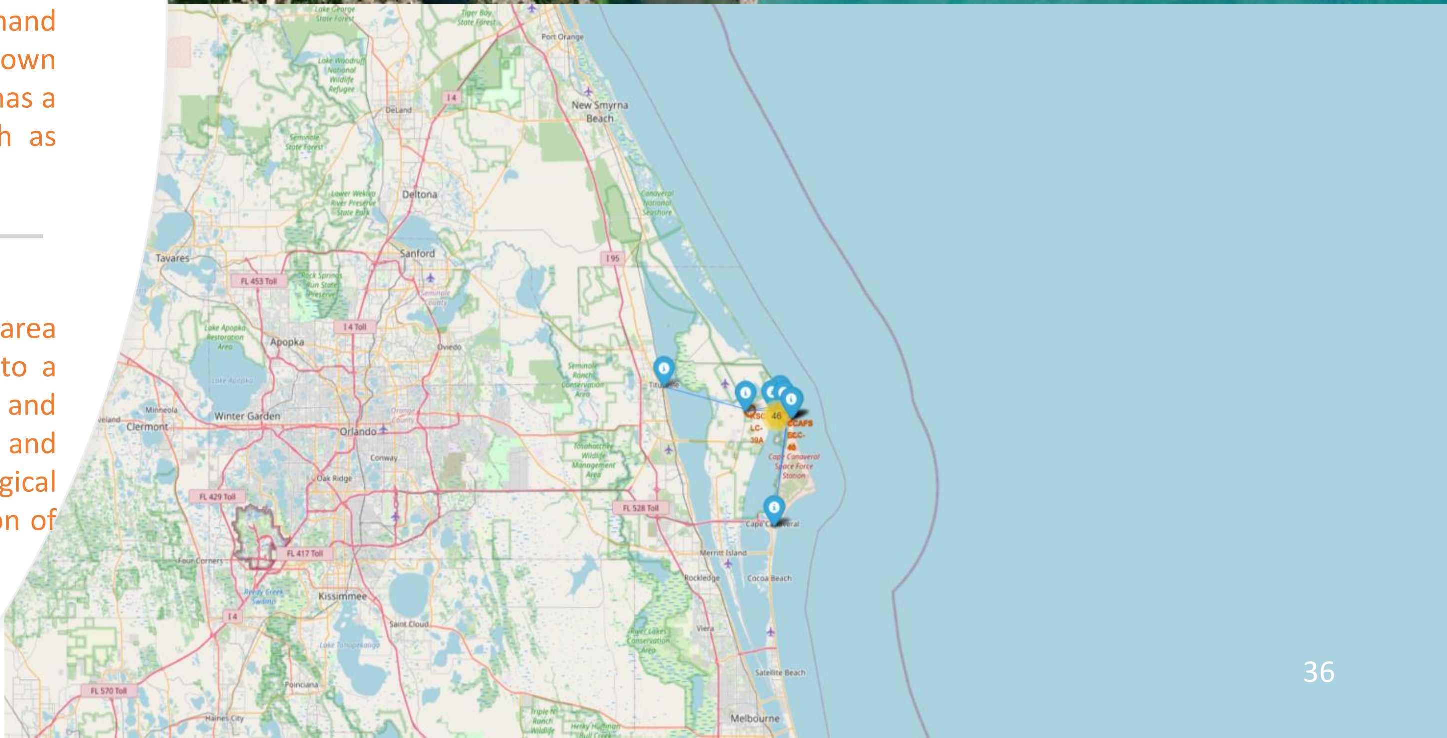


Florida’s Space Coast

Cape Canaveral is a small city known for its connection to the space industry, as it is near the site of the Cape Canaveral Air Force Station and the Kennedy Space Center, which have been used for many historic space launches. The city is also home to several beaches, parks, and wildlife preserves, making it a popular tourist destination. Additionally, Cape Canaveral is situated close to the larger city of Cocoa Beach, which has a thriving tourism industry.

Titusville is a larger city located just north of Cape Canaveral. Like its neighbor, Titusville has strong ties to the space industry and is located near the Kennedy Space Center. The city is also home to a few historic sites and museums, including the American Space Museum & Space Walk of Fame and the Valiant Air Command Warbird Museum. Titusville is situated along the Indian River Lagoon and is known for its scenic waterfront views and outdoor recreation opportunities. The city has a small-town feel but is located within driving distance of larger cities such as Orlando and Daytona Beach.

Together, Cape Canaveral and Titusville make up a significant portion of the area known as the Space Coast which, in addition to CCAFS and KSC, is home to a number of aerospace and defense companies, as well as research institutions and universities. The region has a strong economy and is a hub for innovation and technology. With its unique blend of history, natural beauty, and technological innovation, the Space Coast continues to be a fascinating and important region of the United States.

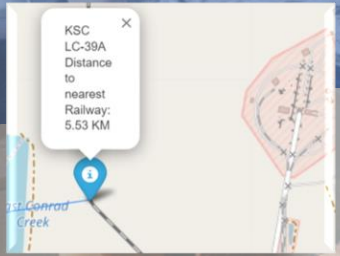


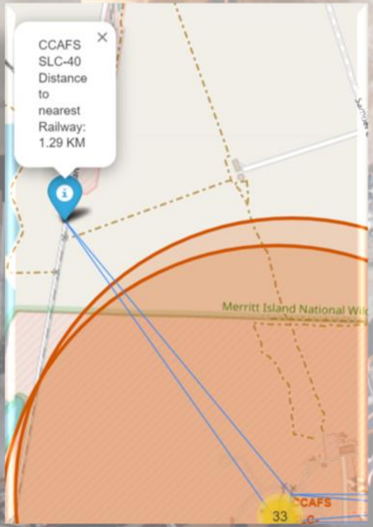
Kennedy Space Center and Cape Canaveral Air Force Station

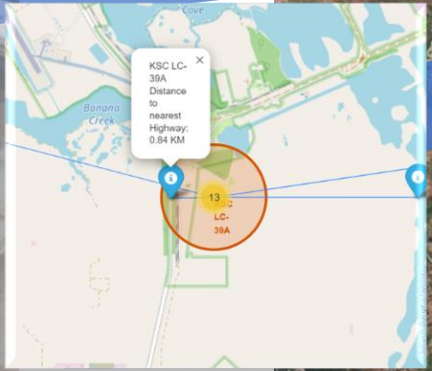
The NASA Railroad is a railway system that operates within the Kennedy Space Center (KSC) at Cape Canaveral, Florida. It is primarily used to transport heavy equipment and payloads, such as spacecraft, rocket boosters, and fuel tanks, to and from the launch pads and various other facilities within the KSC. The railroad was used to transport the massive Saturn V rocket from the Vehicle Assembly Building to the launch pad for the Apollo missions.

The railroad spans a total of about 38 miles, with tracks connecting various locations within the KSC, such as the Vehicle Assembly Building, Launch Control Center, and Launch Complex 39. It also connects with the mainline rail network outside the KSC, allowing for the transport of equipment to and from other parts of the country. Its closest point is less than 6km from LC-39A, and just a little over a kilometer from the CCAFS launch sites.

The NASA Railroad is considered a critical infrastructure asset for the KSC, as it provides a reliable and efficient means of transporting equipment and payloads for space missions. It is also an important part of the history and legacy of the U.S. space program.

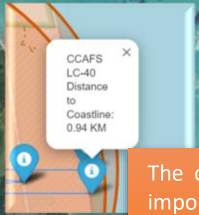




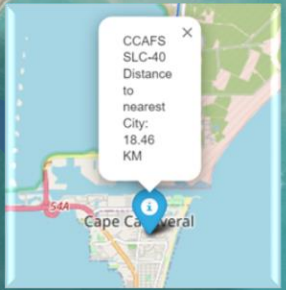


Kennedy Parkway North is a four-lane divided highway, with a speed limit of 45 miles per hour. It is heavily guarded and monitored by security personnel, as access to the KSC is restricted for safety reasons. It is a vital artery for space operations at the KSC. It is used to transport rockets, spacecraft, and other equipment to and from the launch pads, and to move personnel to various areas of the facility. The road is also used for security patrols and emergency response vehicles.

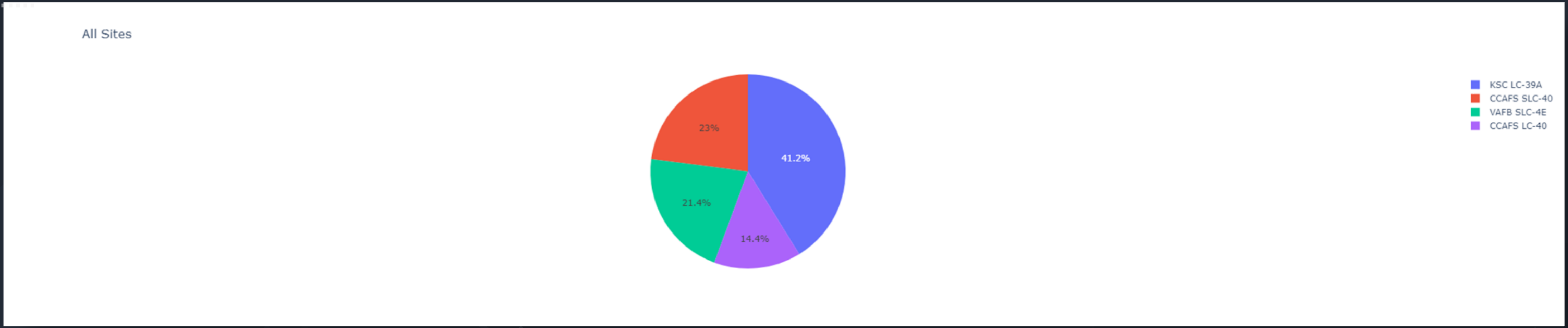
Titusville is the nearest city to KSC. Cape Canaveral is the closest city to CCAFS. While both launch sites are technically considered part of Cape Canaveral, they are a good distance from residential areas.



The coastline near CCAFS is also an important habitat for a variety of plant and animal species, including sea turtles, shorebirds, and dune vegetation. The area is subject to conservation efforts and regulations aimed at protecting these species and their habitats.



Section 4

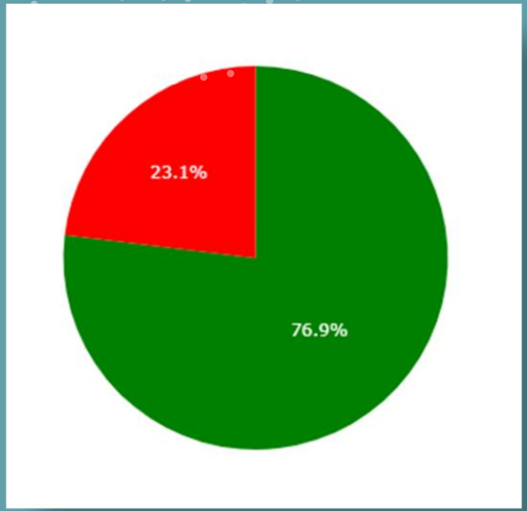


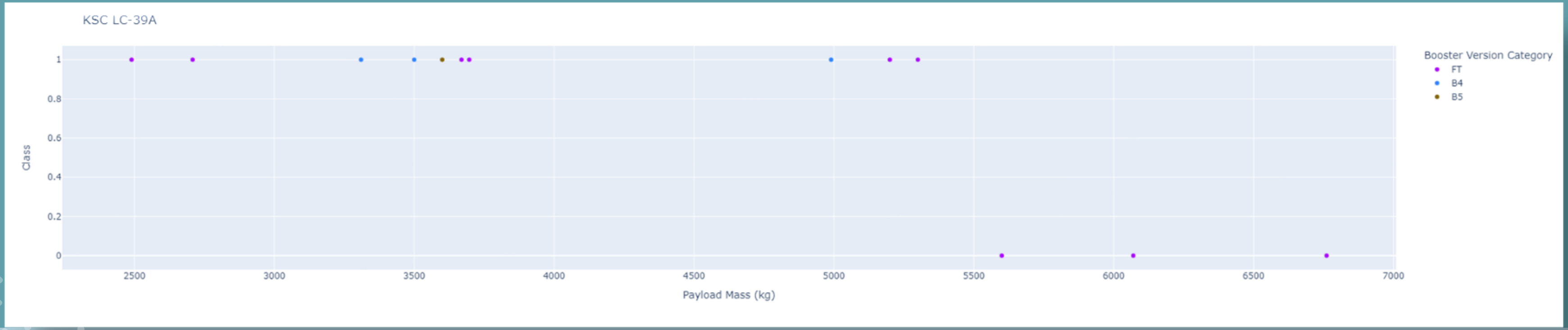
All Successful F9 Booster Recoveries by Launch Site

- As previously noted, KSC has the highest recovery success rate for various reasons, including the fact that it wasn't used for a launch by SpaceX until after the company had already completed a successful recovery at CCAFS.
- CCAFS accounts for the second highest total successful recoveries. It was the site of the first successful booster landing on a ground-based landing pad in December of 2015.
- Flight number, payload, and booster type all affect the launch site success rates given changes over time, including improvements made to the F9 rocket.

Kennedy Space Center

- The high success rate at Kennedy Space Center is largely due to the time SpaceX began using it for launches. 2015 is when SpaceX first introduced the Falcon 9 Full Thrust (FT), which included the ability to land and reuse the first stage.
- KSC has had 3 failed FT landings but has had an impeccable record since the introduction of the B4, and later the B5, which is designed to be recovered and reused 10 times with minimal refurbishing.
- We also see that those failed landing attempts involved payloads much higher than the average for payloads launched from Kennedy Space Center.



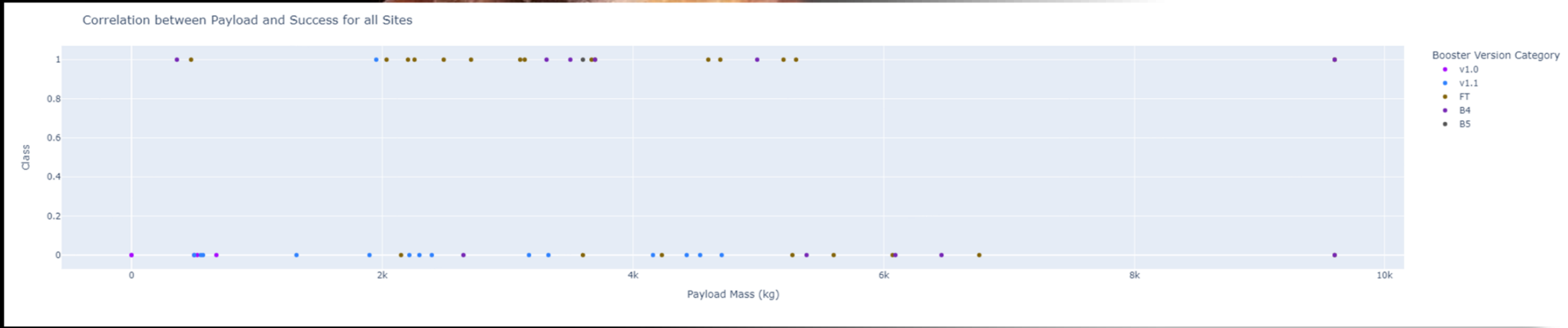


Payload vs Landing Outcome

When we set the dropdown to all sites and select 8000kg to 10000kg payloads, we see that only two payloads, both equaling 9600kg, have been launched over the 8000kg range. Of these there has been a 50% landing failure/success rate.

The fact that we've been asked to look at the relation between landing outcome and payload is a little misleading. When we look at 9600kg, for example, we see they were both B4 boosters where one failed the landing, and the other was successfully recovered.

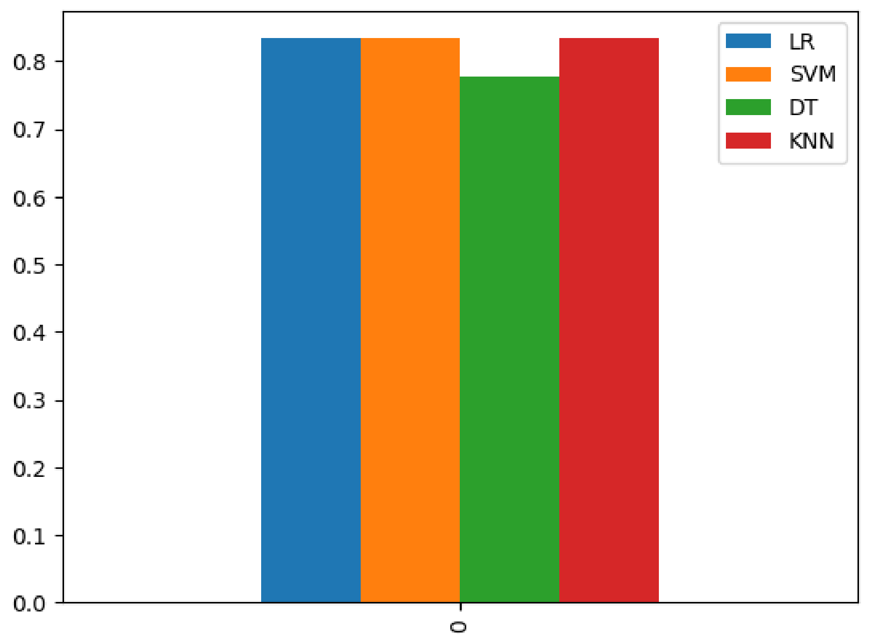
Another factor which skews the payload success rate is the band in which SpaceX rockets carry payloads for Starlink. While not viewable here each of those launches has been in the 15,600kg range, and have been largely B5 rockets launched after 2015.



Section 5

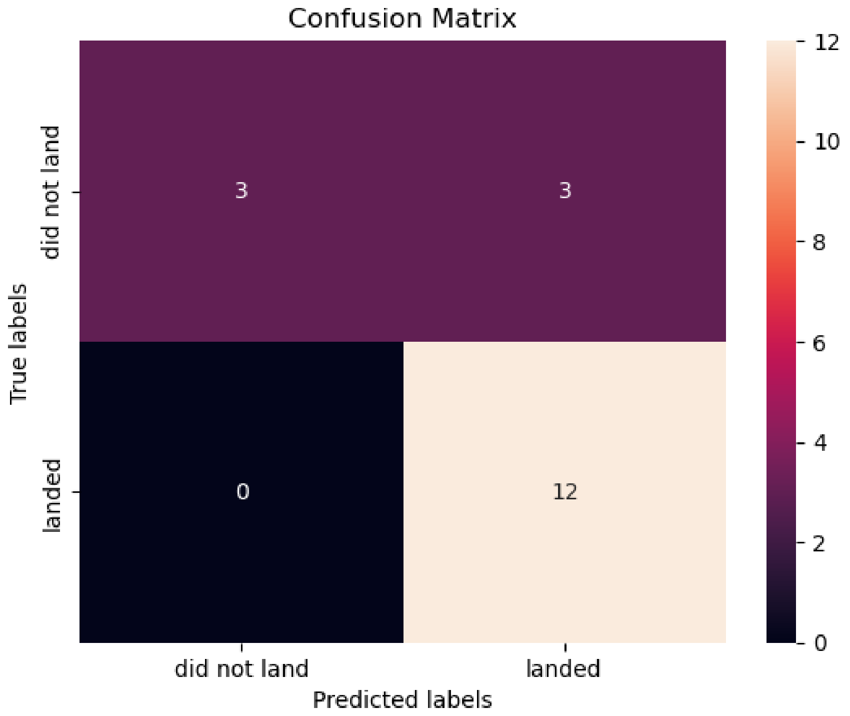
Classification Accuracy

- The average accuracy for each model after several tests was 83.3%
- The Decision Tree Classifier had the least reliable results varying in accuracy from 60% to 90% but averaging around 83% for all the tests.
- Attempting to tune different parameters doesn’t permanently increase the accuracy due to the number of variables, but also due to the mixture of data.
- Other than GridSeachCV we played around with RandomizedSearchCV which increased the accuracy of the Decision Tree Classifier, but it was too inconsistent.



Confusion Matrix

- This is how the confusion matrix was generated for each model.
- The accuracy of landing predictions and other factors imply issues with the data.
- One big issue with the data is the fact that it includes launches before 2015, and that includes non-attempts for recovery as failures.
- Furthermore, the B4 and B5 have performed nearly flawlessly, but some of the missions listed as having ended in a landing failure were not launched with the intent to recover the booster.



Conclusions

- SpaceX successfully landed the first F9 rocket booster in December 2015
- The Falcon 9 with Full Thrust booster (FT) was introduced in 2015 with
- SpaceX has completed over 30 missions for Starlink with a payload of 15,600kg each, or 60 Starlink Satellites per trip
- All Starlink launches have been carried out with B5 boosters.
- The F9 B5 booster is designed to be reused 10 times with minimal refurbishing, and it is the primary booster currently used for all SpaceX launches.
- As of May 4, 2023, SpaceX has had 228 launches, 189 successful landings, and 163 reflights.

Appendix

- SpaceX Launch Manifest Website
- SpaceX API Documentation
- NASA's Falcon 9 Data Sheet
- F9 v1.0 and v1.1 Data Sheet

Thank you!