IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Dasi Yashasvi
29-03-2025

Outline

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

Executive Summary

- This capstone predicts the Falcon 9 first-stage landing success using machine learning. I collected and cleaned launch data, performed exploratory analysis, and trained models. Decision trees achieved the highest accuracy. Insights from this study can help competitors assess launch costs and improve bidding strategies.
- Initial findings suggest that probability of success has improved with time since 2013, as newer flight numbers proved to be more successful, and orbits ES-L1, GEO, HEO, SSO being highly successful while SO and GTO saw the least success. Launch stations like KSC saw the most success.
- I dealt as to find the factors which were affecting the above results

Introduction

- SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upward 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. This information can be used if an alternate company wants to bid against SpaceX for a rocket launch.
- Answer to the questions like what all factors affect the success of a mission
- Success rate which will tell us if the upcoming launch will be successful of not referencing to the previous launches
- Factors which doesn’t affect the success of a mission such that spaceX can reduce their cost of launches

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was scraped from Wikipedia and spaceX official API using requests and parsed using beautifulSoup
- Performed data wrangling
  - After the API response was cleaned, parsed and converted into a dataframe, the null values in payload mass were replaced by the mean of the overall payload mass
- Performed exploratory data analysis (EDA) using visualization and SQL
- Performed interactive visual analytics using Folium and Plotly Dash
- Performed predictive analysis using classification models
  - Comparing Logistic Regression, SVM, Decision Tree, K-nearest neighbour algorithms

Data Collection

The data collection involved two primary methods:

1. Web Scraping from Wikipedia
    1. Source: "List of Falcon 9 and Falcon Heavy launches" Wikipedia page
    2. Tool: BeautifulSoup library in Python
    3. Data extracted: Launch records including dates, booster versions, payloads, outcomes
    4. https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches

2. SpaceX API Requests
    1. Multiple endpoints used to gather detailed information
    2. Collected data about rockets, payloads, launchpads, and cores
    3. Enriched dataset with technical specifications and outcomes
    4. https://api.spacexdata.com/v4/cores/

Data Collection – SpaceX API

SpaceX API

- Initial request to SpaceX API's past launches endpoint
- JSON data normalized into pandas Data Frame
- Core information filtered to focus on relevant data points
- Additional API requests for detailed information about:
    - Booster specifications
    - Launch site coordinates
    - Payload mass and orbit
    - Core landing details

```
spacex_url="https://api.spacexdata.com/v4/launches/past"
response = requests.get(spacex_url)
data = pd.json_normalize(response.json())
data = data[['rocket', 'payloads', 'launchpad', 'cores', 'flight_number', 'date_utc']]
data = data[data['cores'].map(len)==1]
data = data[data['payloads'].map(len)==1]
data['cores'] = data['cores'].map(lambda x : x[0])
data['payloads'] = data['payloads'].map(lambda x : x[0])
data['date'] = pd.to_datetime(data['date_utc']).dt.date
data = data[data['date'] <= datetime.date(2020, 11, 13)]


launch_dict = {'FlightNumber': list(data['flight_number']),
'Date': list(data['date']),
'BoosterVersion':BoosterVersion,
'PayloadMass':PayloadMass,
'Orbit':Orbit,
'LaunchSite':LaunchSite,
'Outcome':Outcome,
'Flights':Flights,
'GridFins':GridFins,
'Reused':Reused,
'Legs':Legs,
'LandingPad':LandingPad,
'Block':Block,
'ReusedCount':ReusedCount,
'Serial':Serial,
'Longitude': Longitude,
'Latitude': Latitude}

df= pd.DataFrame(launch_dict)
```


Github link: [DataAnalysisCapstone/jupyter-labs-spacex-data-collection-api.ipynb at main · Yashasvidasi/DataAnalysisCapstone]()

Data Collection - Scraping

Wikipedia Web Scraping

- HTTP GET request to Wikipedia page snapshot
- BeautifulSoup used to parse HTML content
- Helper functions created to extract specific data points from table cells
- Extraction of flight numbers, dates, booster versions, and landing outcomes

```python
static_url = "https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922"
response = requests.get(static_url)
soup = BeautifulSoup(response.content)
html_tables = soup.find_all('table')
first_launch_table = html_tables[2]
column_names = []
column_data = first_launch_table.find_all('th')

for a in column_data:
    x = extract_column_from_header(a)
    if x is not None and len(x)>0:
        column_names.append(x)
launch_dict= dict.fromkeys(column_names)

# Remove an irrelvant column
del launch_dict['Date and time ( )']

# Let's initial the launch_dict with each value to be an empty list
launch_dict['Flight No.'] = []
launch_dict['Launch site'] = []
launch_dict['Payload'] = []
launch_dict['Payload mass'] = []
launch_dict['Orbit'] = []
launch_dict['Customer'] = []
launch_dict['Launch outcome'] = []
# Added some new columns
launch_dict['Version Booster']=[]
launch_dict['Booster landing']=[]
launch_dict['Date']=[]
launch_dict['Time']=[]

df= pd.DataFrame({ key:pd.Series(value) for key, value in launch_dict.items() })
```


Github link: DataAnalysisCapstone/jupyter-labs-webscraping.ipynb at main · Yashasvidasi/DataAnalysisCapstone

Data Wrangling

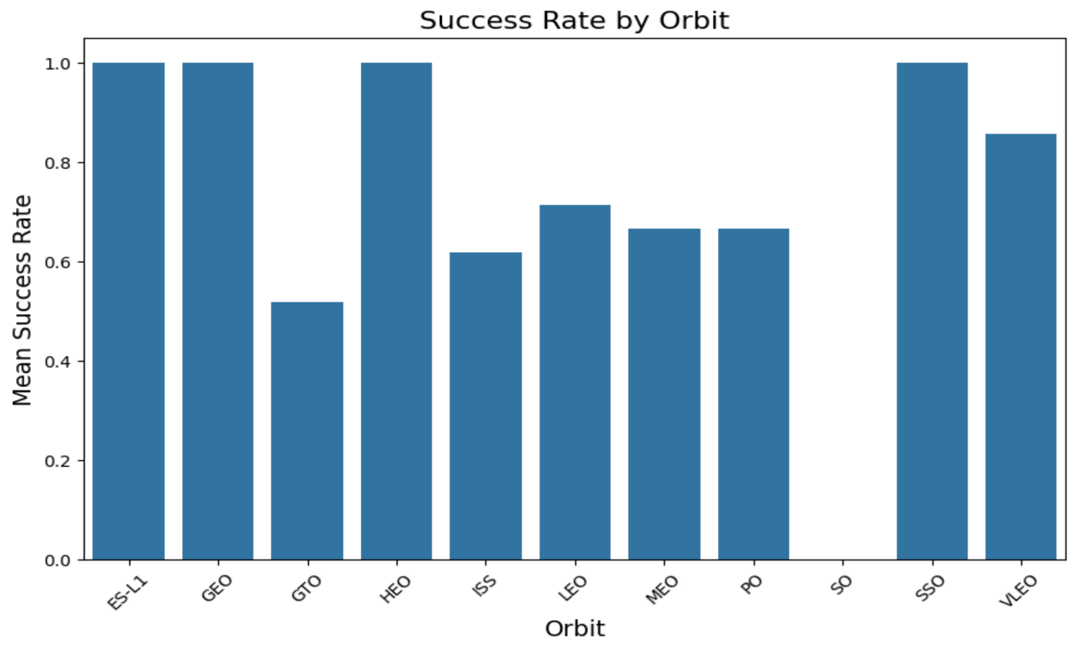
```
#inlcude only Falcon 9 boosters
df_9 = df[df['BoosterVersion'] == 'Falcon 9']

#renumber the flight numbers
df_9.loc[:,'FlightNumber'] = list(range(1, df_9.shape[0]+1))

#replace the Null payload masses with average
df_9['PayloadMass'].fillna(value=df['PayloadMass'].mean(), inplace=True)

```

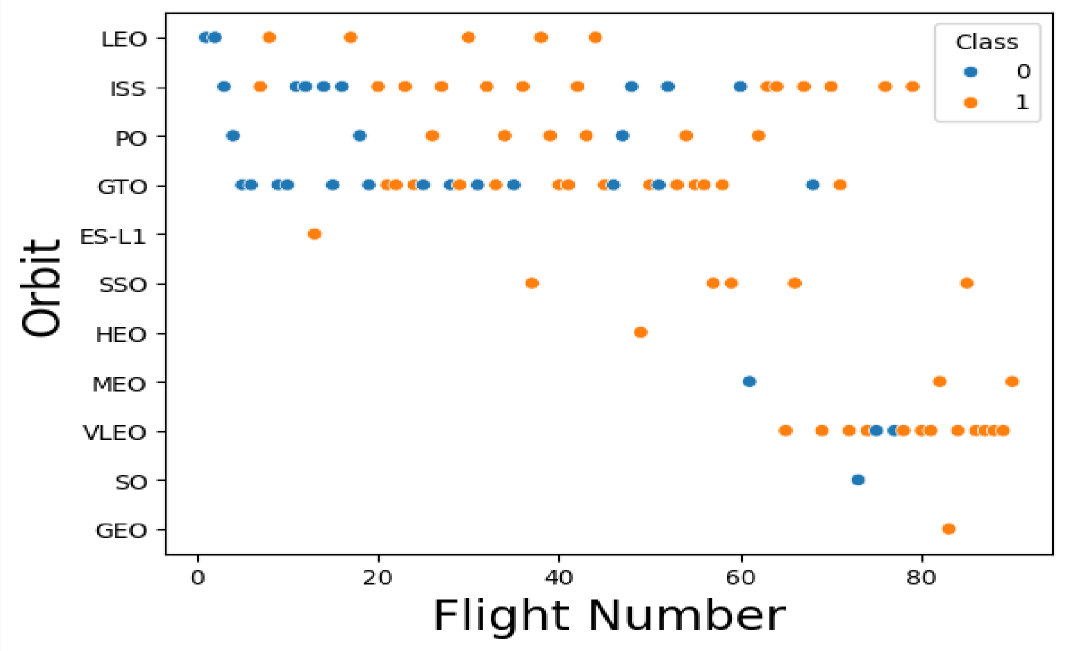
Selecting only Falcon 9 boosters and storing them in a new Data Frame called df_9.

Reassigning flight numbers for the Falcon 9 boosters, to ensure a sequential numbering system starting from 1.

Replacing any 26 null values in the 'PayloadMass' column with the mean payload mass, ensuring there are no missing values for this attribute.

EDA with Data Visualization





SSO, HEO, GEO, ES-L1 were mostly successful but HEO, GEO, ES-L1 saw only 1 mission which was a success, so we can't say for sure they are reason for success of launch, in that rate SSO saw success 5 out of 5 times being a better indicator. In the LEO, VLEO orbit, success seems to be related to the number of flights. Conversely, in the GTO orbit, there appears to be no relationship between flight number and success. Lastly SO saw only 1 mission which was a failure so we can't say for sure SO is reason for failure of launch.

Github link: [DataAnalysisCapstone/edadataviz.ipynb at main · Yashasvidasi/DataAnalysisCapstone]

EDA with SQL

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

```
%sql SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTABLE WHERE Booster_Version = 'F9 v1.1'
```

* sqlite:///my_data1.db
Done.

AVG(PAYLOAD_MASS__KG_)
2928.4

Display the names of the unique launch sites in the space mission

```
%sql SELECT DISTINCT Launch_Site FROM SPACEXTABLE
```

* sqlite:///my_data1.db
Done.

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

Rank the count of landing outcomes between the date 2010-06-04 and 2017-03-20 in descending order.

```
%%sql SELECT Landing_Outcome, COUNT(*)
    FROM SPACEXTABLE
    WHERE Date BETWEEN '2010-06-04' AND '2017-03-20'
    GROUP BY Landing_Outcome
    ORDER BY COUNT(*) DESC;
```

* sqlite:///my_data1.db
Done.

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

Finding the average mass for spaceX missions, Failure and success of landings count, unique launch sites to get a better scope of the data we are dealing with.

Github link: [DataAnalysisCapstone/jupyter-labs-eda-sql-coursera_sqllite.ipynb at main · Yashasvidasi/DataAnalysisCapstone]

Build an Interactive Map with Folium


```
spacex_df=pd.read_csv(spacex_csv_file)
def get_marker_color(class_value):
    return 'green' if class_value == 1 else 'red'

spacex_df['marker_color'] = spacex_df['class'].apply(get_marker_color)
site_map = folium.Map(location=[spacex_df['Lat'].mean(), spacex_df['Long'].mean()], zoom_start=5)
marker_cluster = MarkerCluster().add_to(site_map)
for _, record in spacex_df.iterrows():
    marker = folium.Marker(
        location=[record['Lat'], record['Long']],
        popup=f"Site: {record['Launch Site']}<br>Status: {'Success' if record['class'] == 1 else 'Failure'}",
        icon=folium.Icon(color=record['marker_color'])  # Use the marker_color column
    )
    marker_cluster.add_child(marker)  # Add marker to the cluster
site_map.add_child(marker_cluster)

site_map
```

If a launch was successful `(class=1)`, then we use a green marker and if a launch was failed, we use a red marker `(class=0)`. Geolocation helps us to visualize the real-world location and the though process of selecting the location for building a launch pad

Github: [DataAnalysisCapstone/lab_jupyter_launch_site_location.ipynb at main · Yashasvidasi/DataAnalysisCapstone]

Build a Dashboard with Plotly Dash

I have used :

- Total successful launch ratios for each site
- Success vs failure ratios for each site
- Payload vs success for each site for variable ranges

Analyzing **total successful launch ratios** helps identify the most reliable sites, ensuring efficient resource allocation and improved launch strategies. Comparing **success vs. failure ratios** highlights potential risks, allowing for better failure mitigation and site optimization. Examining **payload vs. success rates** reveals site-specific capabilities, ensuring the right locations are chosen for different payload ranges. Together, these analyses enhance decision-making, improve reliability, and optimize future space missions.

Github link: [DataAnalysisCapstone/spacex-dash-app.py at main · Yashasvidasi/DataAnalysisCapstone]

Predictive Analysis (Classification)

Flow of the Machine Learning Project

•The notebook aims to predict whether SpaceX's Falcon 9 first stage will land successfully, which directly impacts launch costs.
•The dataset contains information about past Falcon 9 launches with various features like flight number, payload mass, orbit type, and landing outcomes.
•The target variable 'Class' indicates landing success (1) or failure (0).
•The data was preprocessed by standardizing features using Standard Scaler.
•The dataset was split into training (80%) and testing (20%) sets with a random state of 2.

Results and Model Performance

- Linear Regression: tuned hyperparameters :(best parameters) {'C': 0.01, 'penalty': 'l2', 'solver': 'lbfgs'} accuracy : 83.33%
- SVM: tuned hyperparameters :(best parameters) {'C': 1.0, 'gamma': 0.03162277660168379, 'kernel': 'sigmoid'}, accuracy: 83.33%
- tuned hyperparameters :(best parameters) {'criterion': 'gini', 'max_depth': 6, 'max_features': 'sqrt', 'min_samples_leaf': 2, 'min_samples_split': 2, 'splitter': 'random'}, accuracy: 88.88%
- KNN: tuned hyperparameters :(best parameters) {'algorithm': 'auto', 'n_neighbors': 10, 'p': 1}, accuracy: 83.33%

Github :DataAnalysisCapstone/SpaceX_Machine Learning Prediction_Part_5.ipynb at main · Yashasvidasi/DataAnalysisCapstone

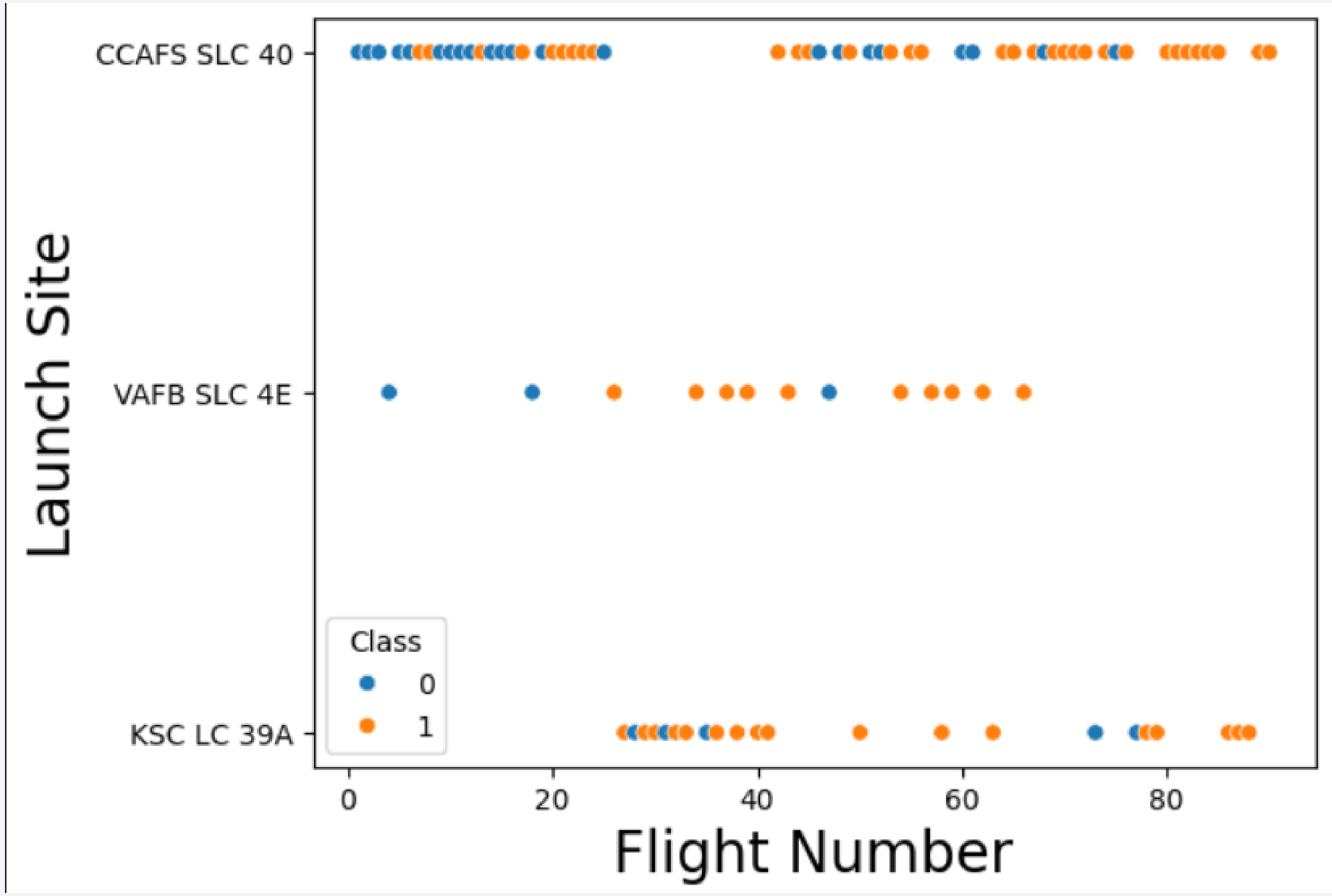
Results

- SSO, HEO, GEO, ES-L1 were mostly successful but HEO, GEO, ES-L1 saw only 1 mission which was a success, so we can't say for sure they are reason for success of launch, in that rate SSO saw success 5 out of 5 times being a better indicator. In the LEO, VLEO orbit, success seems to be related to the number of flights. Conversely, in the GTO orbit, there appears to be no relationship between flight number and success. Lastly SO saw only 1 mission which was a failure so we can't say for sure SO is reason for failure of launch.
- Launch sites are located in distant locations, like Cape Canaveral Space Force Station (CCSFS) and Kennedy Space Center (KSC) in Florida are situated along the Atlantic coast, providing direct access to supply chains and transportation routes while maintaining a safe distance from densely populated areas. Their coastal location allows for eastward launches over the ocean, reducing the risk of debris falling over land. Additionally, launch sites are typically positioned away from major railroads and civilian infrastructure to prevent accidents and minimize disruptions.
- Decision tree was the highest accuracy with tuned hyperparameters :(best parameters) {'criterion': 'gini', 'max_depth': 6, 'max_features': 'sqrt', 'min_samples_leaf': 2, 'min_samples_split': 2, 'splitter': 'random'}, accuracy: 88.88%

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

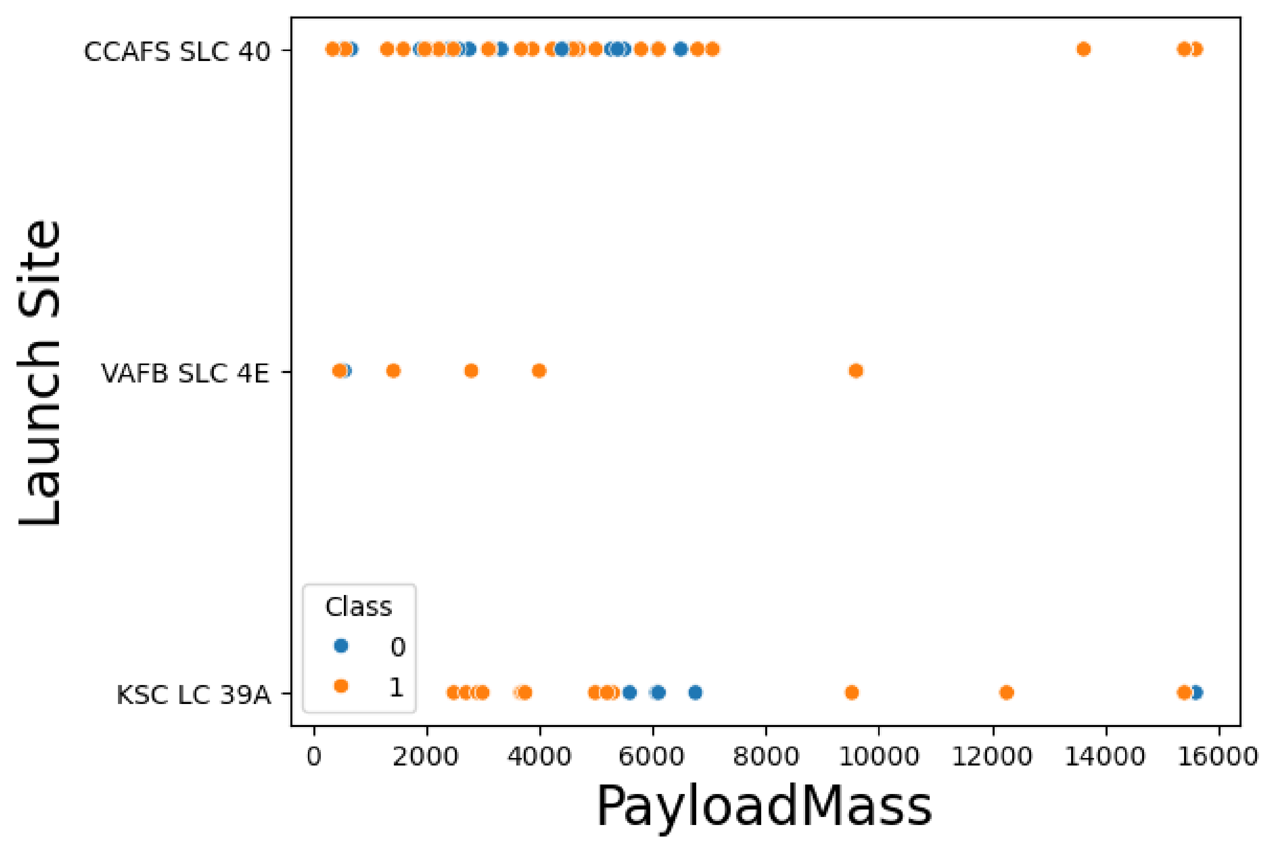


CCAFS SLC 40 has the highest number of launches, showing a mix of successes and failures, but success rates improve over time.

KSC LC 39A has a high success rate, with most flights being successful, likely due to its use for critical missions.

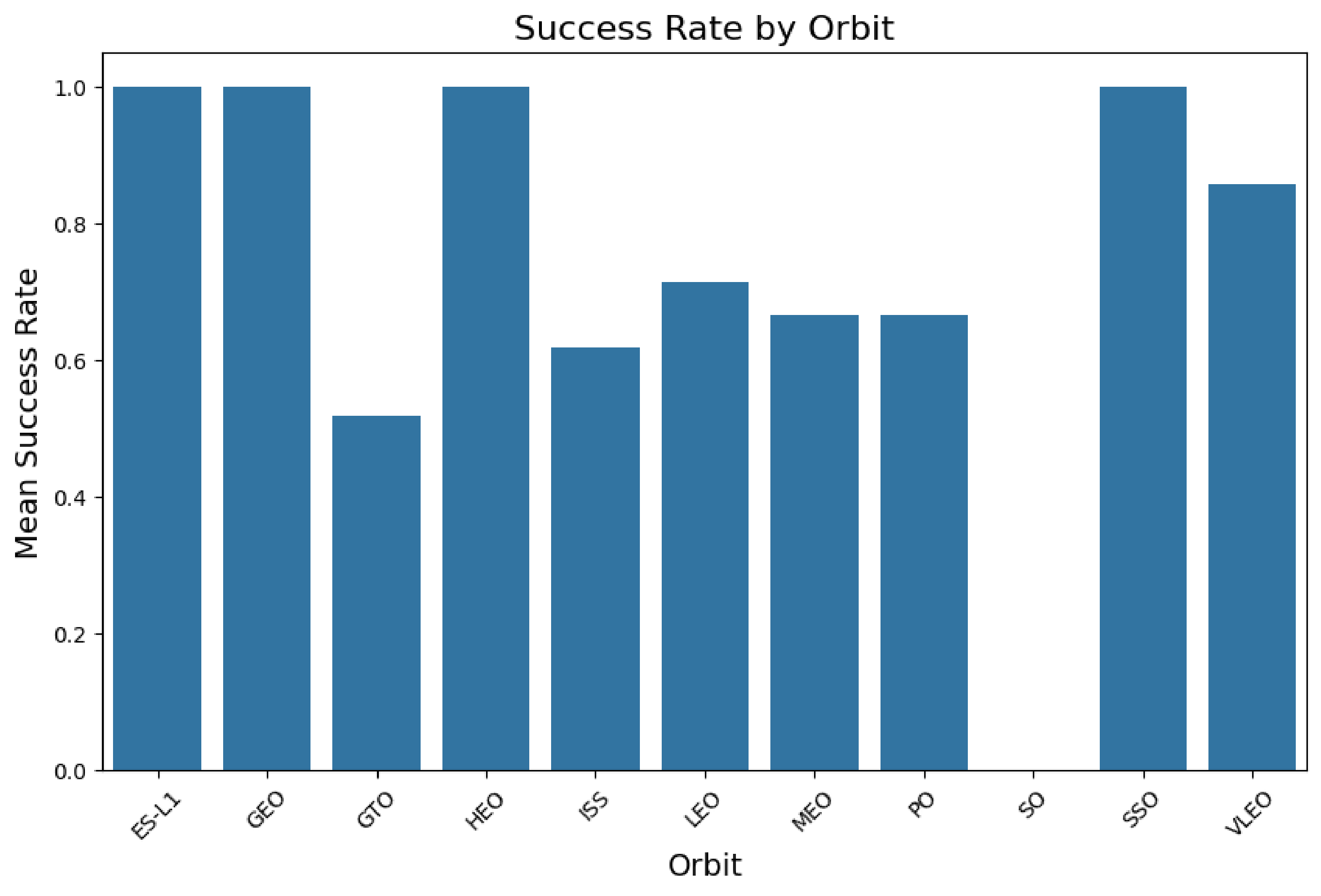
VAFB SLC 4E has fewer launches and a lower success rate, possibly due to different mission profiles like polar orbits.

Payload vs. Launch Site



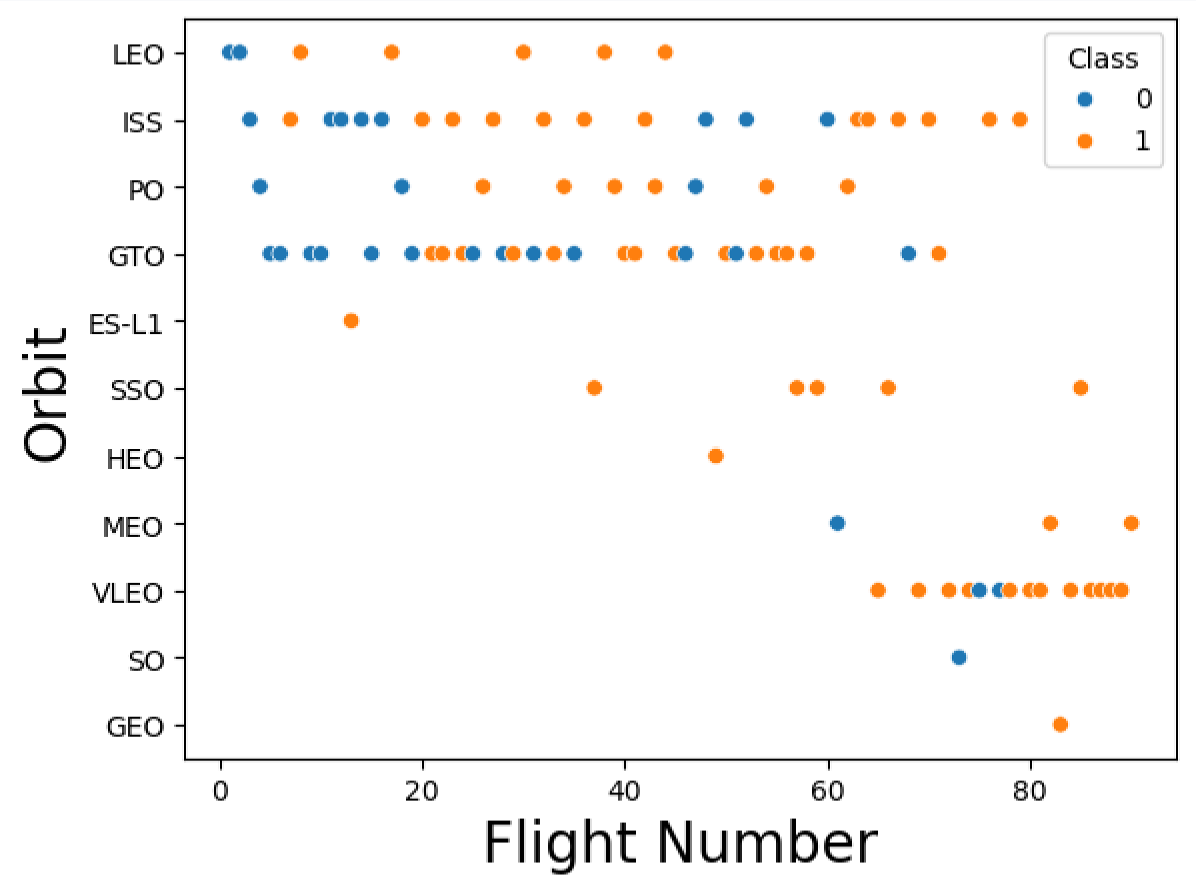
1. The chart shows payload mass distribution across three launch sites - CCAFS SLC 40 (Cape Canaveral Air Force Station Space Launch Complex 40), VAFB SLC 4E (Vandenberg Air Force Base Space Launch Complex 4E), and KSC LC 39A (Kennedy Space Center Launch Complex 39A).
2. The payload masses range from approximately 0 to 16,000 units (likely kilograms). Most launches across all sites carry payloads under 8,000 units, with CCAFS SLC 40 showing the most consistent launch activity. There are a few heavier payloads (12,000-16,000 range) primarily launched from CCAFS SLC 40 and KSC LC 39A.
3. VAFB SLC 4E saw most of the success also the average mass of payload is much lower than rest of the 2

Success Rate vs. Orbit Type



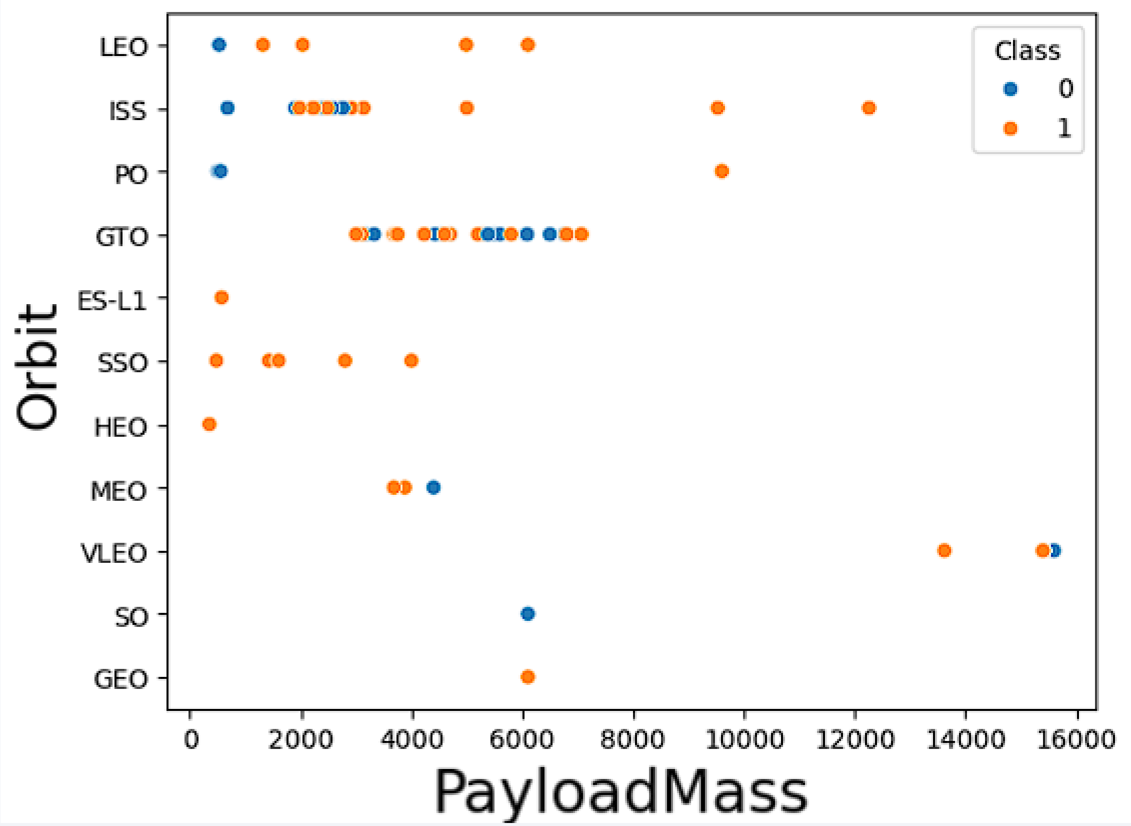
1. Perfect Success Rates: Several orbits (ES-L1, GEO, HEO, and SO) show a 100% success rate, indicating flawless mission execution for these specific orbital destinations.
2. Lowest Success Rate: The GTO (Geostationary Transfer Orbit) has the lowest success rate among all orbits, at approximately 52%, suggesting it may be the most challenging orbit to achieve.
3. Varied Performance: Other orbits like ISS, LEO, MEO, and PO show success rates ranging from about 65% to 75%, indicating room for improvement in mission reliability for these destinations.

Flight Number vs. Orbit Type



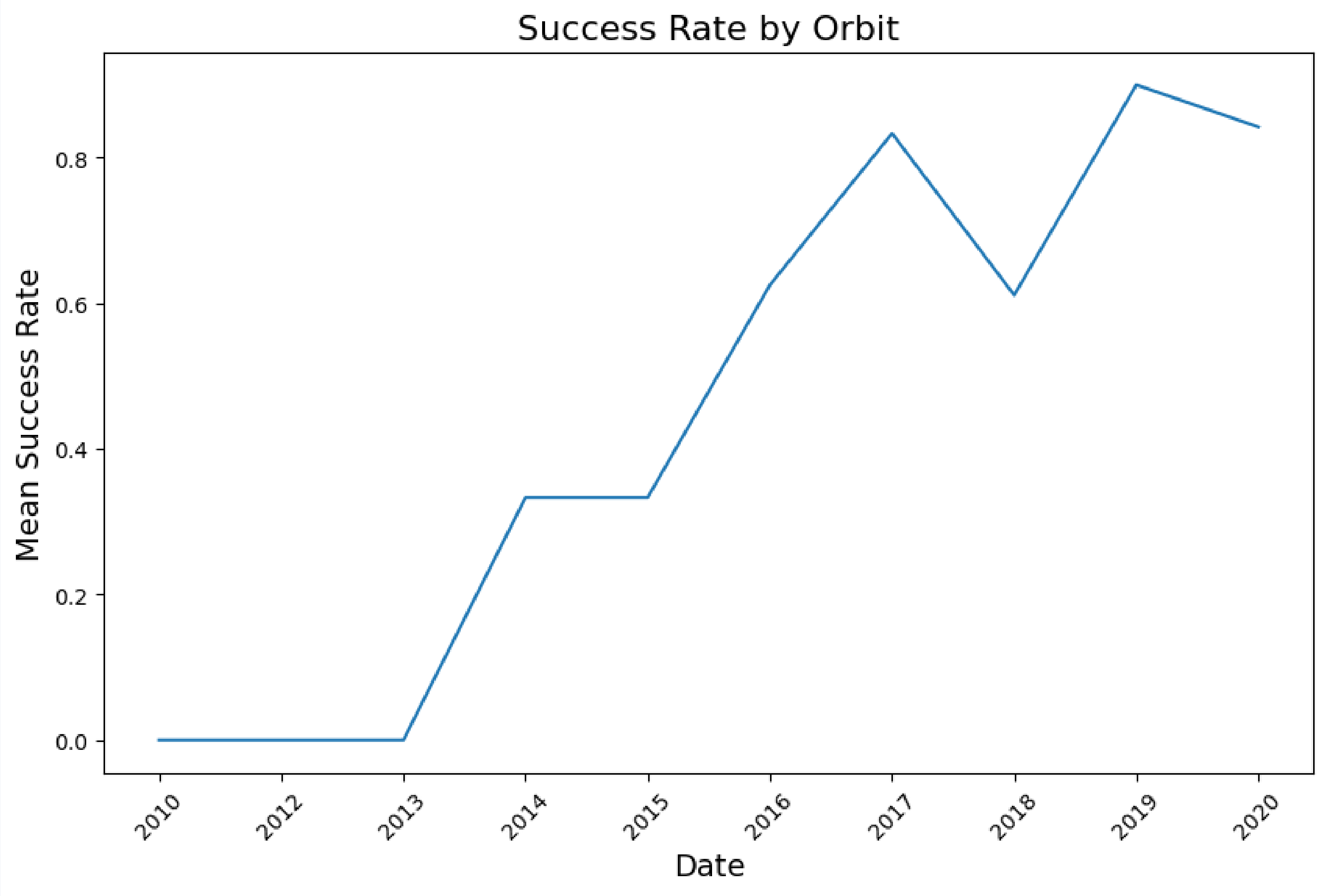
- We can see as the flight number increased the spaceX targeted orbits such as MEO, VLEO, SO, GEO
- They stopped targeting LEO after 45 flight number
- Only 1 mission to GEO
- SSO was most successful
- GEO was the newest destination
- SpaceX concentrated on VLEO mostly in recent Flight Launches and was fairly successful

Payload vs. Orbit Type



1. Higher payload masses (above ~10,000 kg) are predominantly associated with successful launches (Class 1), especially in GTO and GEO orbits.
2. Certain orbits like LEO and ISS show a mix of success (Class 1) and failure (Class 0), while others like GEO and SO have fewer data points but are mostly successful.
3. Lower payload masses (<5,000 kg) exhibit varied outcomes across different orbits, indicating that success is influenced by factors beyond payload mass alone.

Launch Success Yearly Trend



1. The mean success rate has shown a steady increase from 2013 to 2019, indicating significant advancements in orbital launch reliability over time.
2. Peak Performance: The highest success rate was achieved around 2019, nearing 90%, followed by a slight decline in 2020.
3. There was dip in success rate during 2018 probably due to the shift to the newer version of falcon 9 booster
4. Dip in 2020 is probably due to covid

All Launch Site Names

Display the names of the unique launch sites in the space mission

```
%sql SELECT DISTINCT Launch_Site FROM SPACEXTABLE
```

✓ 0.0s

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

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

- CAAFS LC-40 (Cape Canaveral Air Force Launch Complex 40)
- VAFB SLC-4E (Vandenberg Air Force Base Space Launch Complex 4 East)
- KSC LC-39A (Kennedy Space Center Launch Complex 39A)
- CAAFS SLC-40 (Cape Canaveral Air Force Space Launch Complex 40)

First and the last are the same hence merged in later queries

Launch Site Names Begin with 'CCA'

Display 5 records where launch sites begin with the string 'CCA'

```
%sql SELECT * FROM SPACEXTABLE WHERE Launch_Site LIKE 'CCA%' LIMIT 5;
```

Python

* sqlite:///my_data1.db
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

We got 2 rocket launches of 2010 with payload mass 0 kg both landings were failures

2 more on 2012, they didn’t attempt to land them with approx. 500kg payload both carried out by nasa

Another 1 by nasa on 2013 with a payload of 677 kg

All of the launches destinations were ISS (international space station)

Total Payload Mass

Task 3

Display the total payload mass carried by boosters launched by NASA (CRS)

```
%sql SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXTABLE where Customer = 'NASA (CRS)'
```

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

SUM(PAYLOAD_MASS__KG_)
45596

Total Payload of 45596kg were carried out

CRS (commercial Resupply Services) they are carried out by NASA to Resupply the astronauts and gear in the international space station ISS mostly targeted on the LEO orbit.

Average Payload Mass by F9 v1.1

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

```
%sql SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTABLE WHERE Booster_Version = 'F9 v1.1'
```

* sqlite:///my_data1.db
Done.

AVG(PAYLOAD_MASS__KG_)
2928.4

V1.1 carried out average of 2928.4KG

While the overall average was around 6000 kg so booster F1 V1.1 was is less effective in carrying out heavy payload to the space

First Successful Ground Landing Date

List the date when the first succesful landing outcome in ground pad was acheived.

Hint:Use min function

```
1  %sql SELECT MIN(Date) FROM SPACEXTABLE WHERE Landing_Outcome = 'Success (ground pad)'
```

✓ 0.0s

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

MIN(Date)
2015-12-22

Date: 2015-12-22 recorded the successful landing on ground pad

The Data we have is from 2010, but the first successful landing on the ground pad happened in 2015, nearly 6 year after the earliest launch record

Successful Drone Ship Landing with Payload between 4000 and 6000

Booster_Version	PAYLOAD_MASS__KG_
F9 v1.1	4535
F9 v1.1 B1011	4428
F9 v1.1 B1014	4159
F9 v1.1 B1016	4707
F9 FT B1020	5271
F9 FT B1022	4696
F9 FT B1026	4600
F9 FT B1030	5600
F9 FT B1021.2	5300
F9 FT B1032.1	5300
F9 B4 B1040.1	4990
F9 FT B1031.2	5200
F9 B4 B1043.1	5000
F9 FT B1032.2	4230
F9 B4 B1040.2	5384
F9 B5 B1046.2	5800
F9 B5 B1047.2	5300
F9 B5 B1046.3	4000
F9 B5B1054	4400
F9 B5 B1048.3	4850
F9 B5 B1051.2	4200
F9 B5B1060.1	4311
F9 B5 B1058.2	5500
F9 B5B1062.1	4311

List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

```
1  %sql SELECT DISTINCT Booster_Version, PAYLOAD_MASS__KG_ FROM SPACEXTABLE WHERE PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000
```

✓ 0.0s

Pytho

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

There are a total of 24 distinct Boosters which were able to carry a payload greater than 4000 and less than 6000

Majority of them being the B5 and FT boosters, there is no order to them, both of the versions carried out payload similarly, while V1.1 carried out towards the lower end of the payload mass.

Total Number of Successful and Failure Mission Outcomes

List the total number of successful and failure mission outcomes

```
%sql SELECT Mission_Outcome, COUNT(*) FROM SPACEXTABLE GROUP BY Mission_Outcome
```

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

Mission_Outcome	COUNT(*)
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Missions were dominantly successful only failure were during landing, this shows that Rocket launches are safer and most of the risk is in landing the rocket safely

Boosters Carried Maximum Payload

List all the booster_versions that have carried the maximum payload mass. Use a subquery.

```
1  %sql SELECT * FROM SPACEXTABLE where PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTABLE)
✓  0.0s

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

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2019-11-11	14:56:00	F9 B5 B1048.4	CCAFS SLC-40	Starlink 1 v1.0, SpaceX CRS-19	15600	LEO	SpaceX	Success	Success
2020-01-07	2:33:00	F9 B5 B1049.4	CCAFS SLC-40	Starlink 2 v1.0, Crew Dragon in-flight abort test	15600	LEO	SpaceX	Success	Success
2020-01-29	14:07:00	F9 B5 B1051.3	CCAFS SLC-40	Starlink 3 v1.0, Starlink 4 v1.0	15600	LEO	SpaceX	Success	Success
2020-02-17	15:05:00	F9 B5 B1056.4	CCAFS SLC-40	Starlink 4 v1.0, SpaceX CRS-20	15600	LEO	SpaceX	Success	Failure
2020-03-18	12:16:00	F9 B5 B1048.5	KSC LC-39A	Starlink 5 v1.0, Starlink 6 v1.0	15600	LEO	SpaceX	Success	Failure
2020-04-22	19:30:00	F9 B5 B1051.4	KSC LC-39A	Starlink 6 v1.0, Crew Dragon Demo-2	15600	LEO	SpaceX	Success	Success
2020-06-04	1:25:00	F9 B5 B1049.5	CCAFS SLC-40	Starlink 7 v1.0, Starlink 8 v1.0	15600	LEO	SpaceX, Planet Labs	Success	Success
2020-09-03	12:46:14	F9 B5 B1060.2	KSC LC-39A	Starlink 11 v1.0, Starlink 12 v1.0	15600	LEO	SpaceX	Success	Success
2020-10-06	11:29:34	F9 B5 B1058.3	KSC LC-39A	Starlink 12 v1.0, Starlink 13 v1.0	15600	LEO	SpaceX	Success	Success
2020-10-18	12:25:57	F9 B5 B1051.6	KSC LC-39A	Starlink 13 v1.0, Starlink 14 v1.0	15600	LEO	SpaceX	Success	Success
2020-10-24	15:31:34	F9 B5 B1060.3	CCAFS SLC-40	Starlink 14 v1.0, GPS III-04	15600	LEO	SpaceX	Success	Success
2020-11-25	2:13:00	F9 B5 B1049.7	CCAFS SLC-40	Starlink 15 v1.0, SpaceX CRS-21	15600	LEO	SpaceX	Success	Success

There were a total of 12 such launches, Where the payload was maximum of 15600 Kg, all of them were carried out by the F9 B5 boosters, and were predominantly successfully carried out by spaceX, with only 2 failures of landing out of 12. All of them were target to LEO orbit which means their Payload was a satellite.

2015 Launch Records

```
%%sql SELECT CASE substr(Date, 6, 2) WHEN '01' THEN 'January' WHEN '02' THEN 'February' WHEN '03' THEN 'March'
WHEN '04' THEN 'April' WHEN '05' THEN 'May' WHEN '06' THEN 'June' WHEN '07' THEN 'July' WHEN '08' THEN 'August'
WHEN '09' THEN 'September' WHEN '10' THEN 'October' WHEN '11' THEN 'November' WHEN '12' THEN 'December'
END AS Month_Name, Booster_Version, Launch_Site, Landing_Outcome FROM SPACEXTABLE WHERE substr(Date, 0, 5) = '2015'
AND Landing_Outcome LIKE '%Failure%' AND Landing_Outcome LIKE '%drone ship%';

```

✓ 0.0s

* sqlite:///my_data1.db
Done.

Month_Name	Booster_Version	Launch_Site	Landing_Outcome
January	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
April	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

2015 Recorded 7 missions, out of them only 1 successfully landed on December, all the drone ship landing attempts were a failure (January, April), boosters used in 2015 were predominantly V1.1 of falcon 9 launched from cape caraneval.

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

```
%%sql SELECT Landing_Outcome, COUNT(*)
    FROM SPACEXTABLE
    WHERE Date BETWEEN '2010-06-04' AND '2017-03-20'
    GROUP BY Landing_Outcome
    ORDER BY COUNT(*) DESC;
```

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

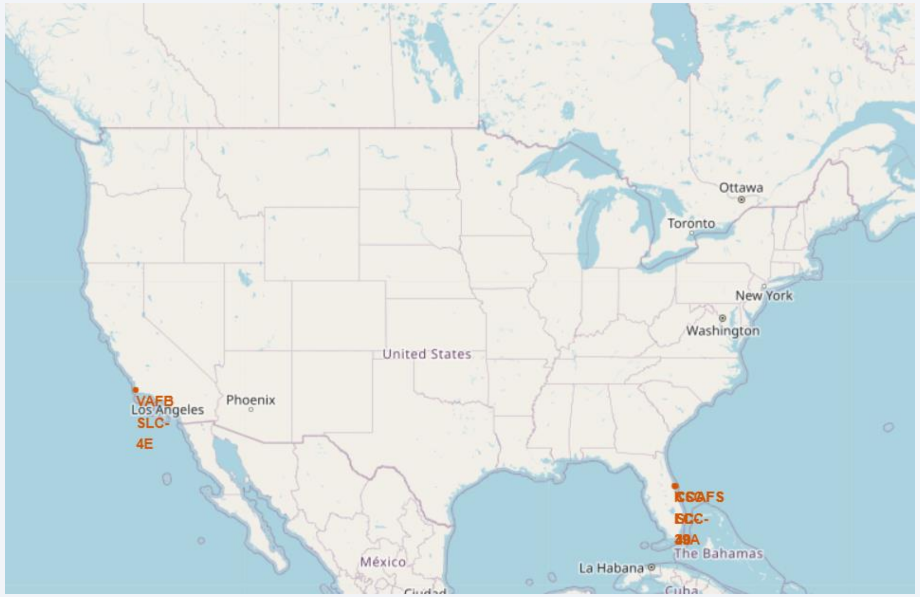
Landing_Outcome	Count(*)
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

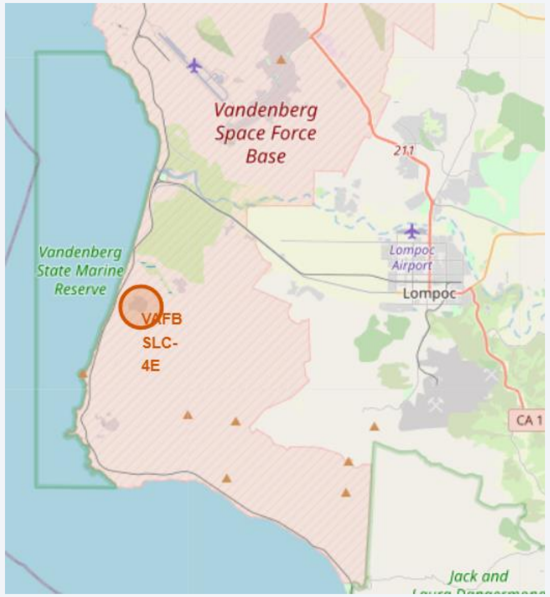
- Count of landing outcomes between the date 2010-06-04 and 2017-03-20, in descending order
- They predominantly didn't attempt landings, their first landing success with a falcon 9 booster was recorded in 2015, so total 8 successful landings between December 2015-march 2017.

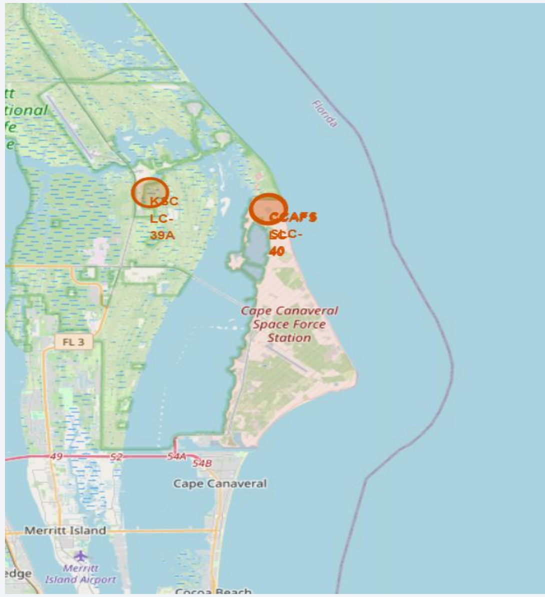
Section 3

Launch Sites Proximities Analysis

Launch Locations







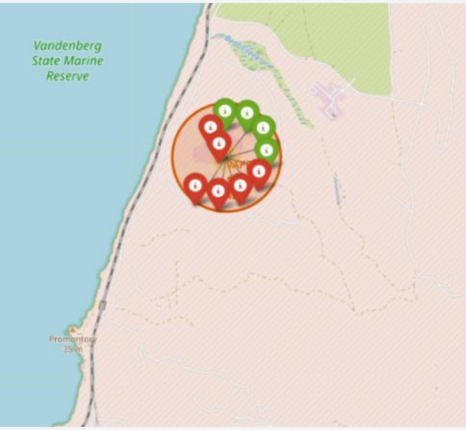
CAAFS LC-40 (Cape Canaveral Air Force Launch Complex 40) Florida US

VAFB SLC-4E (Vandenberg Air Force Base Space Launch Complex 4 East) Nebraska US

KSC LC-39A (Kennedy Space Center Launch Complex 39A) Florida US

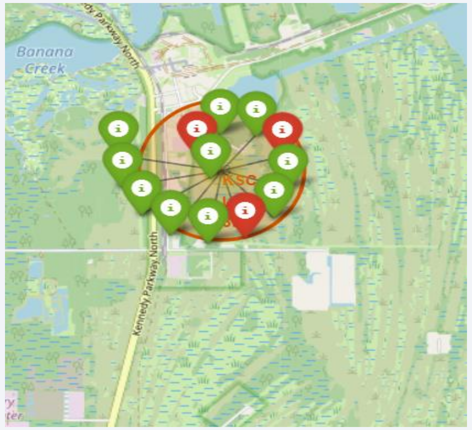
They are situated along the Atlantic and Pacific coast, providing direct access to supply chains and transportation routes. Their coastal location allows for eastward launches over the ocean, westwards for Vandenberg reducing the risk of debris falling over land.

Launch Outcomes in Launch bases



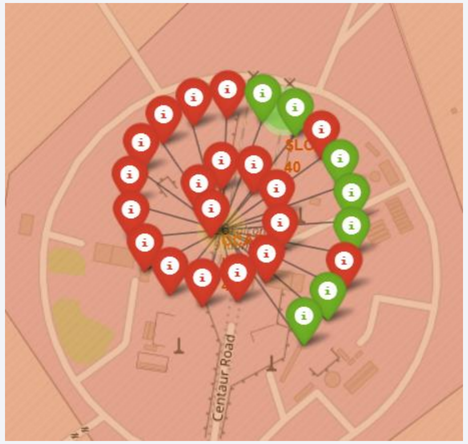
VAFB SLC-4E (Vandenberg Air Force Base Space Launch Complex 4 East) Nebraska US

Total of 10 attempts
Success ratio 2:3



KSC LC-39A (Kennedy Space Center Launch Complex 39A) Florida US

Total of 13 attempts
Success ratio 10:3



CAAFS LC-40 (Cape Canaveral Air Force Launch Complex 40) Florida US

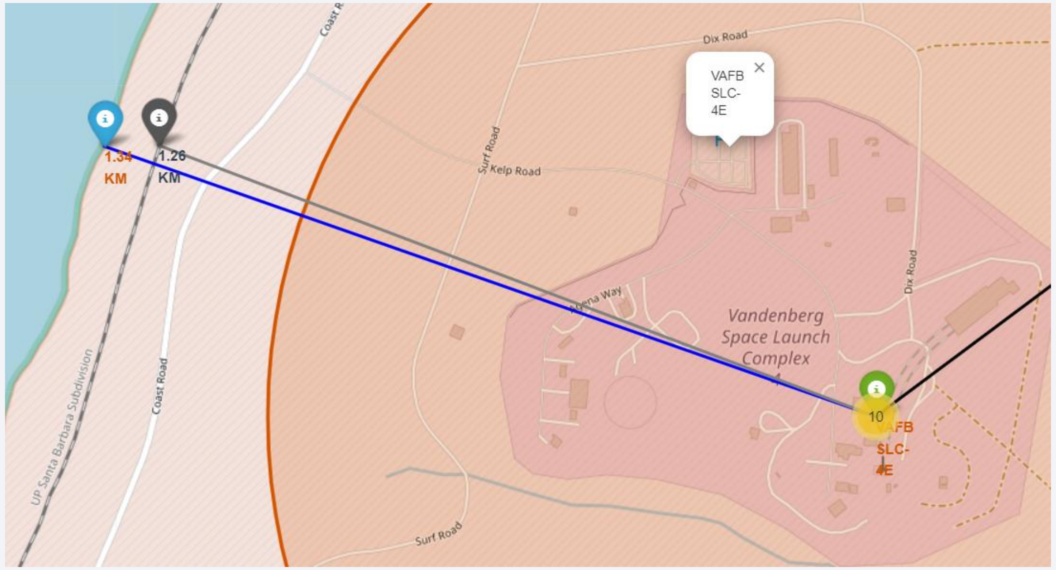
Total of 26 attempts
Success ratio 7:19

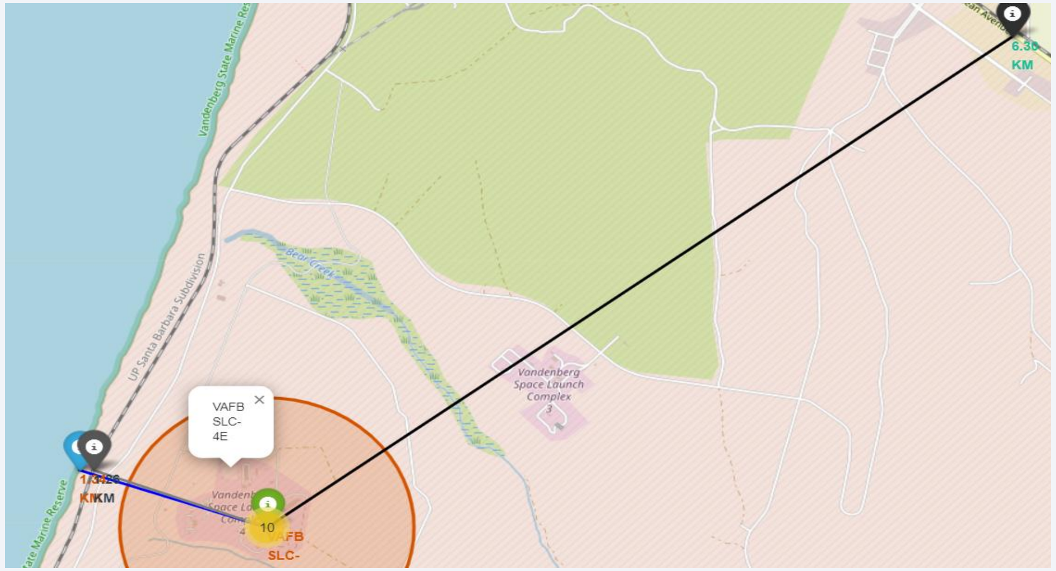
CAAFS SLC-40 (Cape Canaveral Air Force Space Launch Complex 40) Florida US

Total of 7 attempts
Success ratio 3:4

Kennedy Space center Florida was mostly successful

Launch bases distances from coastlines, railways and highways





Closest railway is at a distance of 1.26 Km
Closest Coastline is at a distance of 1.34 Km

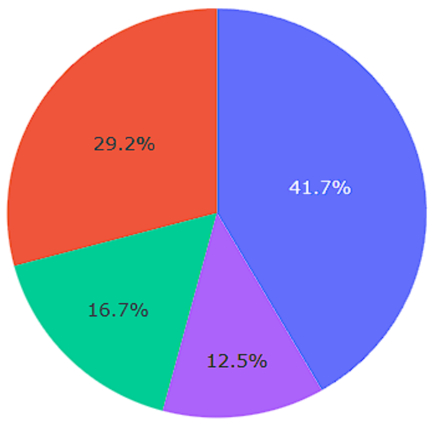
Closest Highway is 6.3 Km away

Vandenberg space launch complex taken
for calculation

Section 4

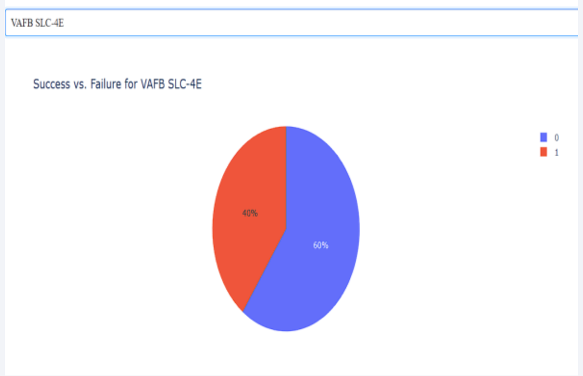
Build a Dashboard with Plotly Dash

DashBoard for success ratio of each site

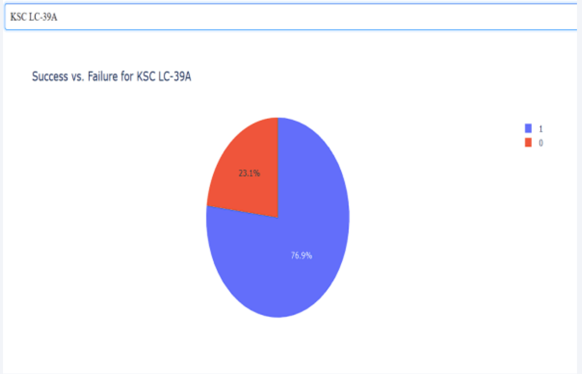


KSC LC-39A (Kennedy Space Center Launch Complex 39A) Florida US saw the highest success ratio, followed by Cape Canaveral, they both are located near to each other, while Vandenberg and the Cape Canaveral Space launch saw the least success ratio

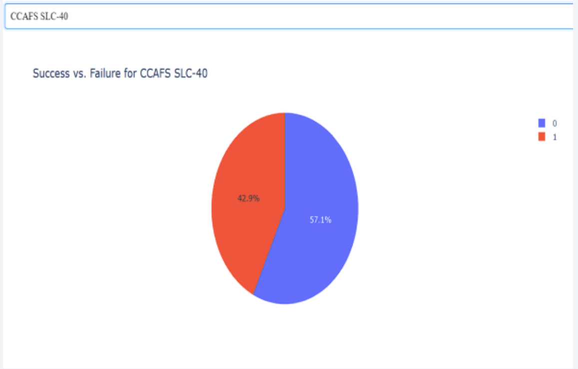
Success Ratio of each site



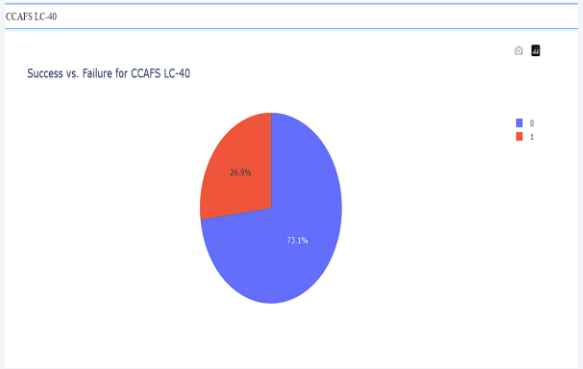
VAFB SLC-4E (Vandenberg Air Force Base Space Launch Complex 4 East) Nebraska US



KSC LC-39A (Kennedy Space Center Launch Complex 39A) Florida US

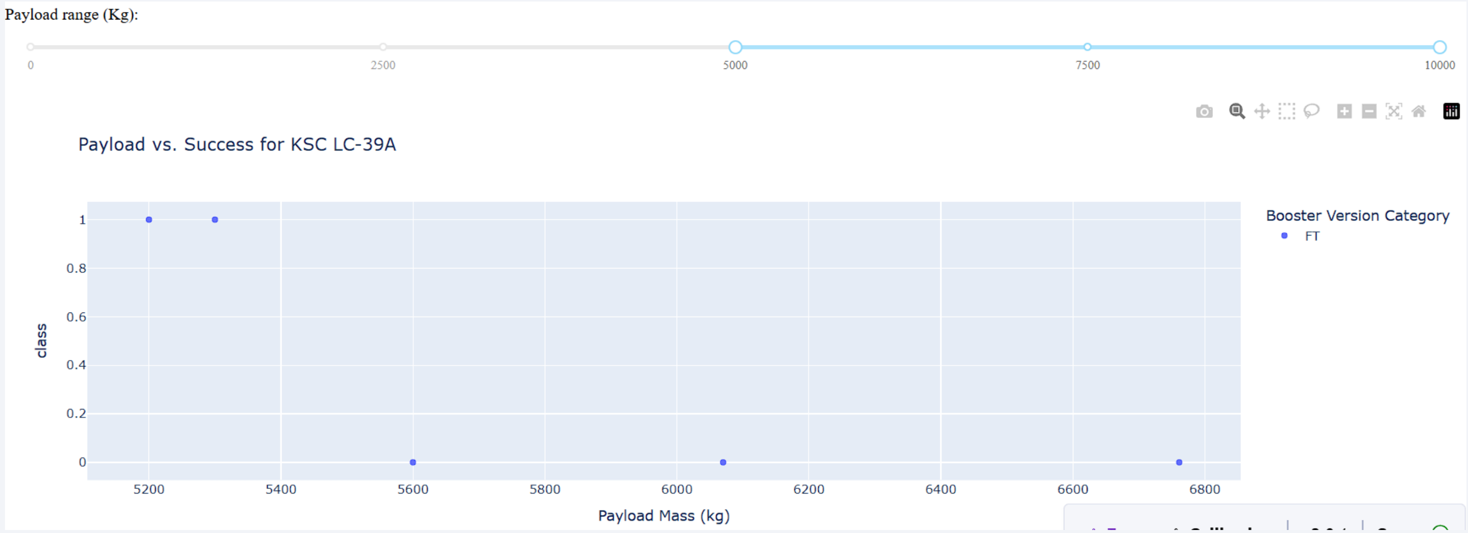


CAAFS LC-40 (Cape Canaveral Air Force Launch Complex 40) Florida US

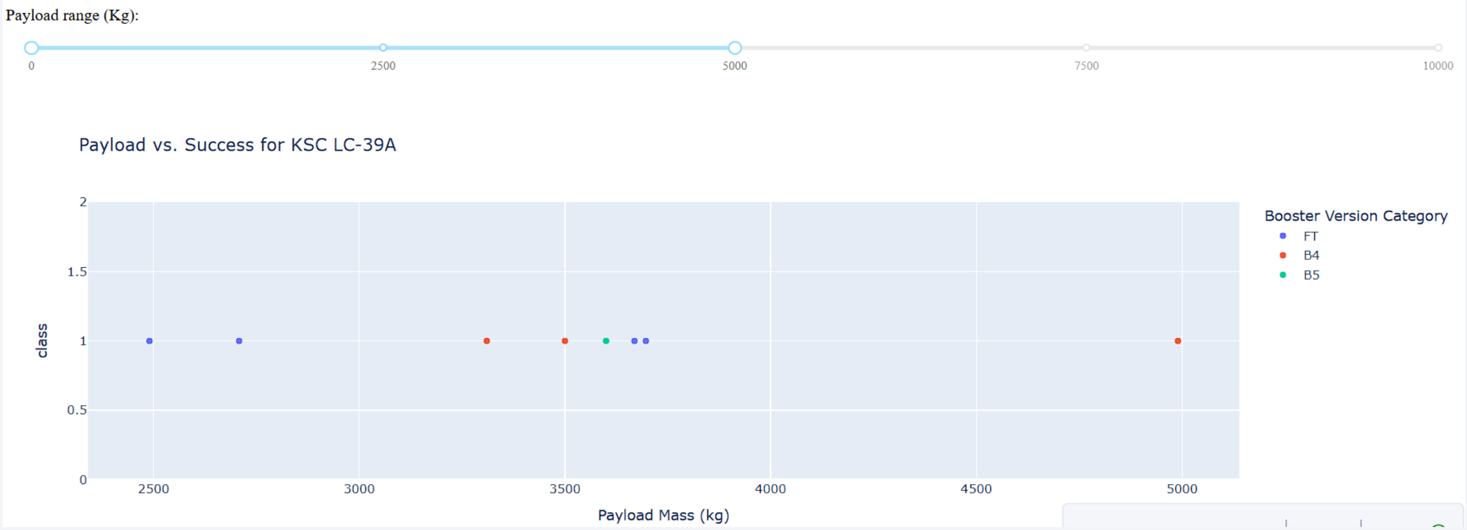


CAAFS SLC-40 (Cape Canaveral Air Force Space Launch Complex 40) Florida US

Payload vs Success of various boosters for KSC (Kennedy Space Center)



Higher payloads were carried out by only FT booster of the falcon 9 family, where majority of success was achieved when payload was below 5400 Kg, higher payload resulted in failures

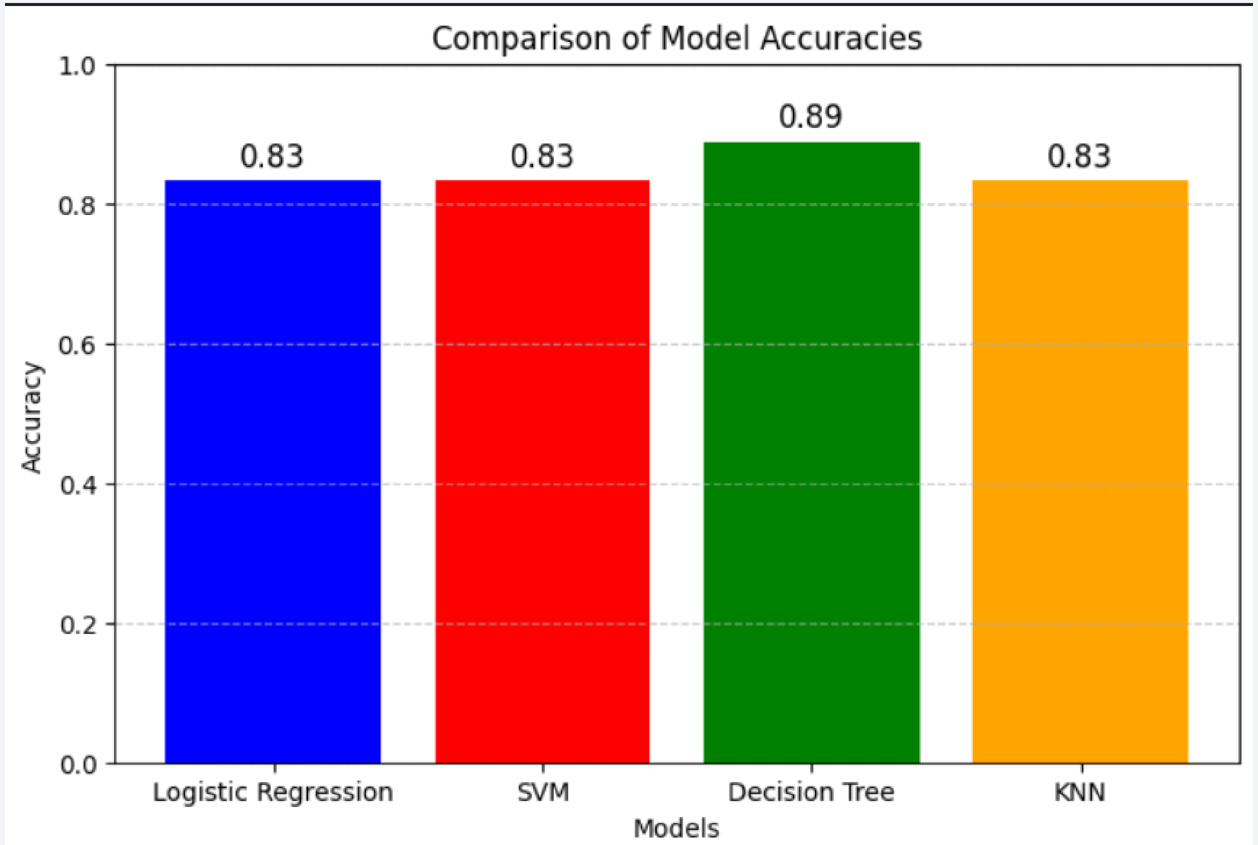


All the payloads below 5000 kg were a success, given that all the 3 of the 5 Booster versions of falcon 9 (FT, B4, B5) were used. Majorly the payload weighed around 3500 Kg, lowest being 2500Kg.

Section 5

Predictive Analysis (Classification)

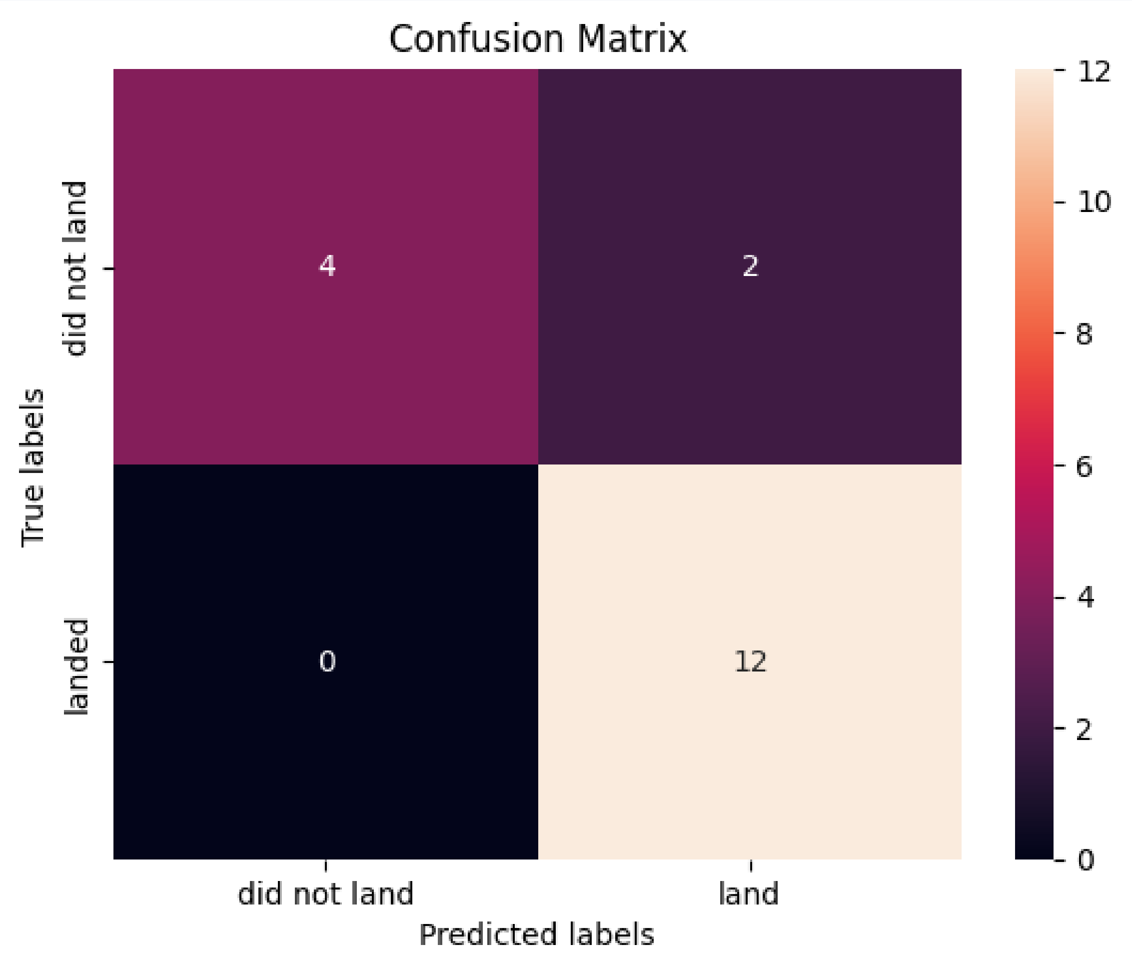
Classification Accuracy




```python
models = ['Logistic Regression', 'SVM', 'Decision Tree', 'KNN']
accuracies = [accuracy_lr, accuracy_svm, accuracy_tree, accuracy_knn]

# Plot bar graph
plt.figure(figsize=(8, 5))
plt.bar(models, accuracies, color=['blue', 'red', 'green', 'orange'])
plt.xlabel("Models")
plt.ylabel("Accuracy")
plt.title("Comparison of Model Accuracies")
plt.ylim(0, 1)
plt.grid(axis='y', linestyle='--', alpha=0.7)

for i, acc in enumerate(accuracies):
    plt.text(i, acc + 0.02, f"{acc:.2f}", ha='center', fontsize=12)

plt.show()
```

Linear Regression, SVM, KNN all had an accuracy of 83.33% on the test set, whereas Decision Tree had an accuracy of 88.88%.

Confusion Matrix of Decision Tree



True Positive (12)
Predicted Label = landed Successfully
Actual Label = landed Successfully

True Negative (4)
Predicted Label = landing Failed
Actual Label = landing Failed

False Positive (2)
Predicted Label = landed Successfully
Actual Label = landing Failed

True Negative (0)
Predicted Label = landing Failed
Actual Label = landing Successfully

Conclusions

```python
from sklearn.metrics import silhouette_score, f1_score
X_test = X_test.reshape(-1, 1) if X_test.ndim == 1 else X_test

sil_score = silhouette_score(X_test, yhat)
f_score = f1_score(Y_test, yhat, average='weighted')

print("Silhouette Score:", sil_score)
print("F1 Score:", f_score)
```

✓ 0.0s

```
Silhouette Score: 0.1301208364328537
F1 Score: 0.882051282051282
```

High F1 score suggests that the model is performing well on classification tasks and F1 score varies from 0-1

Low Silhouette Score means that the data is well cluster able and that is the data points lie close to each other in the n-dimensional space. Silhouette score lies between -1 to 1, -1 being the poorest.

Appendix

All relevant data to be found in Github: [Yashasvidasi/DataAnalysisCapstone]

Thank you!