IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Relly Pangilinan
October 22, 2024

Outline

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

Executive Summary

- Summary of methodologies
  - **Collect raw data** using SpaceX REST API and web scraping.
  - **Wrangle the data** to create a variable to represent success or failed outcome
  - **Explore the data** with data visualization techniques, using the factors of flight number, launch site, payload mass and yearly trend.
  - **Analyze the data** with SQL to calculate the total number of successful and failed outcomes related to launch sites and payload.
  - **Compare** launch site success rates and their proximity to geographical markers.
  - **Visualize** the launch sites with most number of success.
  - **Build Models to predict** landing outcomes of the stage 1 rocket using logistic regression, support vector machine (SVM), decision tree and K-nearest neighbor (KNN).

- Summary of Results
  - **Exploratory Data Analysis:** Launch success has improved over time
  - KSC LC-39A has the highest success rate among landing sites
  - Payload mass and booster selection can be a factor in overall success
  - Orbits ES -L1, GEO, HEO, and SSO have a 100% success rate
  - **Visualization/Analytics:** Most launch sites are near the equator, and all are close to the coast
  - **Predictive Analytics:** All models performed similarly on the test set. The decision tree model slightly outperformed

Introduction

- Project background and context
  - SpaceX is an American spacecraft manufacturer and launch service provider company headquartered at the SpaceX Star base near Brownsville Texas. It currently produces and operates the Falcon 9 and Falcon Heavy rockets. SpaceX is now the leader in the space industry and strives to make space travel affordable for everyone. Its accomplishments include sending spacecraft to the international space station, launching a satellite constellation that provides internet access and sending manned missions to space.

- Findings we are interested:
  - Determine if the first stage of SpaceX Falcon 9 will land successfully
  - How payload mass, launch site, number of flights, and orbits affect first-stage landing success
  - Correlations between launch sites and success rates

Section 1

Methodology

Methodology

Data Collection

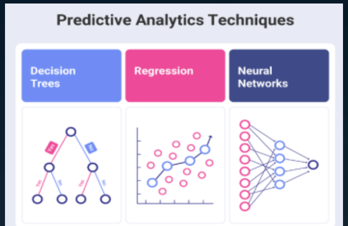
- Collect data using SpaceX REST API and web scraping techniques

Perform Data Wrangling

- Determined labels for training the supervised models by converting mission outcomes in to training labels

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
- Classification is a supervised machine learning method where the model tries to predict the correct label of a given input data.

Data Collection – SpaceX API

Request Data
Request and Parse the SpaceX launch data set
Convert to a dataframe using JSON normalized

↓

Create Dictionary
Construct data using dictionary. Combined columns into a dictionary

↓

Create DataFrame
Convert dictionary to Pandas Dataframe

↓

Filter DataFrame
Filter dataframe for Falcon9 launches

↓

Replace missing values and Export data to CSV
Replace missing values of payload mass
Export the data to CSV

```
[6]: spacex_url="https://api.spacexdata.com/v4/launches/past"

[7]: response = requests.get(spacex_url)
```

```
# Create a data from launch_dict
# Assuming launch_dict is the dictionary containing launch data
launch_dict = {
    'rocket': [1, 2, 3],
    'payloads': ['Payload 1', 'Payload 2', 'Payload 3'],
    'launchpad': ['Launchpad A', 'Launchpad B', 'Launchpad C'],
    'cores': ['Core 1', 'Core 2', 'Core 3'],
    'flight_number': [101, 102, 103],
    'date_utc': ['2022-01-01', '2022-01-02', '2022-01-03']
}

# Create a DataFrame from launch_dict
data = pd.DataFrame.from_dict(launch_dict)

# Print the DataFrame
print(data)
```

```
# Hint data['BoosterVersion']!='Falcon 1'
data_falcon9=data[data['BoosterVersion']=='Falcon 9']
data_falcon9.head()
```


Github link: Data Collection

Data Collection - Scraping

Request Data
Request Falcon 9 Launch data (Wiki)

↓

Create Beautifulsoup object from the HTML response

↓

Extract column names from HTML table header

↓

Create Dictionary
Create dictionary from the collected data

↓

Create Dataframe
Transpose dictionary to Dataframe

```
# use requests.get() method with the provided static_url
# assign the response to a object
response=requests.get(static_url).text
```

```
# Use BeautifulSoup() to create a BeautifulSoup object from a response text content
soup = BeautifulSoup(response, 'html.parser')
```

```
# Check if the key exists before deleting
if 'Date and time (UTC)' in launch_dict:
    del launch_dict['Date and time (UTC)']

# Initialize the launch_dict with each value as 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'] = []
launch_dict['Version Booster'] = []  # Added new column
launch_dict['Booster landing'] = []  # Added new column
launch_dict['Date'] = []  # Added new column
launch_dict['Time'] = []  # Added new column
```


Github link: [Scraping]

Data Wrangling

- Perform EDA and data labels
- Calculate number of launches for each site, occurrence of orbit and mission outcome per orbit
- Different mission outcomes were converted into Training Labels with 1 meaning the booster successfully landed and 0 meaning the booster failed to land. A number of landing scenarios were considered when creating labels :
    1. True Ocean means the mission outcome was successfully landed to a specific region of the ocean
    2. False Ocean means the mission outcome was unsuccessfully landed to a specific region of the ocean
    3. RTLS means the mission outcome was successfully landed to a ground pad
    4. False RTLS means the mission outcome was unsuccessfully landed to a ground pad
    5. True ASDS means the mission outcome was successfully landed on a drone ship
    6. False ASDS means the mission outcome was unsuccessfully landed on a drone ship

Data Wrangling

Calculate the number of launches on each site

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

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

Calculate the number and occurrence of each orbit

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

```
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
```

Calculate the number and occurence of mission outcome of the orbits

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

Create a landing outcome label from Outcome column

```
# landing_class = 0 if bad_outcome
# landing_class = 1 otherwise
landing_class=[]
for outcome in df['Outcome']:
    if outcome in bad_outcomes:
        landing_class.append(0)
    else:
        landing_class.append(1)
```

Github link: [Wrangling]

EDA with Data Visualization

- **To visualize the relationship between Flight Number and Launch Site, relationship between Payload Mass and Launch Site the following charts were plotted**

1. Scatter plot

- Flight Number vs. Payload
- Flight Number vs. Launch Site
- Payload Mass (kg) vs. Launch Site
- Payload Mass (kg) vs. Orbit type

```
# Plot a scatter point chart with x axis to be Flight Number and y axis to be the launch site, and hue to be the class value
sns.catplot(x='FlightNumber', y='LaunchSite', hue='Class', data=df)
plt.title('Flight Number vs Launch Site')
plt.xlabel('Flight Number')
plt.ylabel('Launch Sites')
plt.show()
```

2. Bar Chart:

- Plotted following Bar chart to visualize:

Relationship between success rate of each orbit type

```
# HINT use groupby method on Orbit column and get the mean of Class column
orbit_success = df.groupby('Orbit')['Class'].mean()
orbit_success.plot(kind='bar')
plt.title('Success Rate by Orbit Type')
plt.xlabel('Orbit Type')
plt.ylabel('Success Rate')
plt.show()
```

3. Line Chart:

- Plotted following Line chart to observe:

Average launch success yearly trend

```
# Plot a line chart with x axis to be the extracted year and y axis to be the success rate
sns.lineplot(x='Date', y='Class', data=df)
plt.title('Success Reate per Year')
plt.xlabel('Year')
plt.ylabel('Success')
plt.show()
```

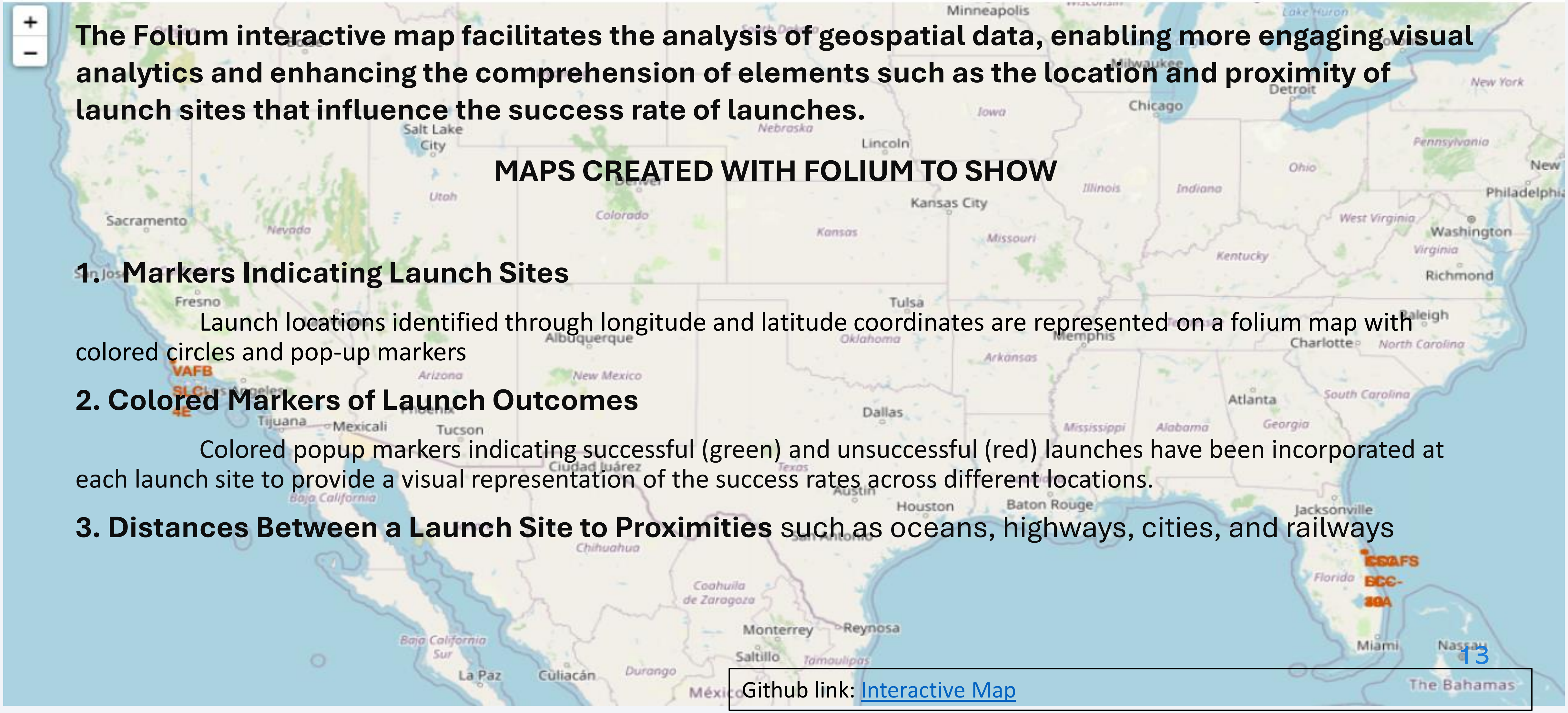
Github link: [EDA]

EDA with SQL

- Using an IBM DB2 cloud instance, the following SQL queries were performed to better understand the SpaceX data set
  - The names of unique launch sites
  - 5 records with 'CCA' as the launch site
  - The total mass of payload carried by NASA's boosters (CRS)
  - The average payload mass carried by booster version F9 v1.1.
  - List the date of the first successful landing on the ground pad.
  - Identify the names of the boosters that have successfully landed on the drone ship and possess a payload mass exceeding 4000 but less than 6000.
  - List the total number of successful and failure mission outcomes
  - List the names of the booster_versions which have carried the maximum payload mass.
  - List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015
  - Rank the count of landing outcomes

Github link: [EDA SQL]

Build an Interactive Map with Folium



The Folium interactive map facilitates the analysis of geospatial data, enabling more engaging visual analytics and enhancing the comprehension of elements such as the location and proximity of launch sites that influence the success rate of launches.

MAPS CREATED WITH FOLIUM TO SHOW

1. Markers Indicating Launch Sites

Launch locations identified through longitude and latitude coordinates are represented on a folium map with colored circles and pop-up markers

2. Colored Markers of Launch Outcomes

Colored popup markers indicating successful (green) and unsuccessful (red) launches have been incorporated at each launch site to provide a visual representation of the success rates across different locations.

3. Distances Between a Launch Site to Proximities such as oceans, highways, cities, and railways

Github link: [Interactive Map]

Build a Dashboard with Plotly Dash

Developed a Plotly Dash web application to facilitate real-time interactive visual analytics of SpaceX launch data. Incorporated a Launch Site drop-down menu, a pie chart, a payload range slider, and a scatter chart into the dashboard.

- A Launch Site Drop-down Input component has been incorporated into the dashboard, enabling users to filter the dashboard visuals by all launch sites or by a specific launch site. **This enable the user to choose either all launch sites or a specific launch site.**
- A Pie Chart has been incorporated into the Dashboard to display the total number of successful launches when the 'All Sites' option is chosen, and to present the counts of successful and failed launches when a specific site is selected. **This enable the user to view the percentage of successful and unsuccessful launches relative to the total.**
- A Payload range slider has been incorporated into the Dashboard, facilitating the selection of various payload ranges to effectively identify visual patterns. **Enable the user to choose a range for the payload mass.**
- A Scatter chart has been incorporated to analyze the potential correlation between payload and mission outcomes for the chosen site(s). Each scatter point is color-coded according to the Booster version, illustrating the mission outcomes associated with various boosters. **Enable the user to observe the relationship between Payload and Launch Success.**

Github link: [Plotly Dash]

Predictive Analysis (Classification)

DATA COLLECTED, TRAINED, TESTED and DISPLAYED

- Create NumPy array from the ‘Class’ column
- Standardize the data with *StandardScaler*
- Fit and transform the data
- Split the data using the *train_test_split* function.
- Create a *GridSearchCV* object with cv=10 for parameter optimization
- Apply *GridSearchCV* on different algorithms:
    - logistic regression(LogisticRegression())
    - support vector machine (SVC())
    - decision tree (DecisionTreeClassifier())
    - K-Nearest Neighbor (KNeighborsClassifier())
- Calculate accuracy on the test data using *.score()* for all models
- Study the confusion matrix for all four models
- Identify the best model if possible

```
[9]: y = data['Class'].to_numpy()
```

```
# students get this relly pangilinan
transform = preprocessing.StandardScaler()
x = transform.fit(X).transform(X)
```

```
parameters ={'C':[0.01,0.1,1],
             'penalty':['l2'],
             'solver':['lbfgs']}
```

```
parameters ={"C":[0.01,0.1,1],'penalty':['l2'], 'solver':['lbfgs']}# l1 lasso l2 ridge
lr=LogisticRegression()
logreg_cv = GridSearchCV(estimator=lr, cv=10, param_grid=parameters).fit(X_train, Y_train)
```

```
parameters = {'kernel':('linear', 'rbf','poly','rbf', 'sigmoid'),
              'C': np.logspace(-3, 3, 5),
              'gamma':np.logspace(-3, 3, 5)}
svm = SVC()
```

```
svm_cv = GridSearchCV(estimator=svm, cv=10, param_grid=parameters).fit(X_train, Y_train)
```

```
print("tuned hpyerparameters :(best parameters) ",svm_cv.best_params_)
print("accuracy :",svm_cv.best_score_)
```

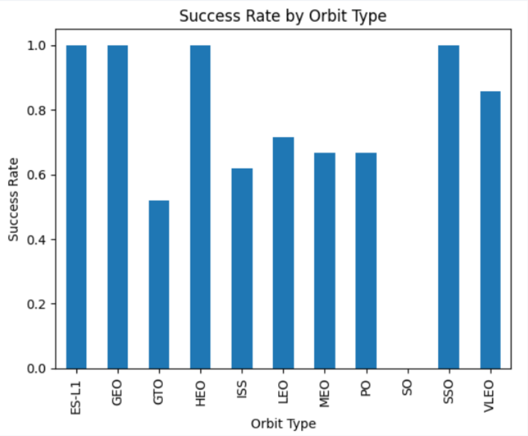
```
tuned hpyerparameters :(best parameters)  {'C': 1.0, 'gamma': 0.03162277660168379, 'kernel': 'sigmoid'}
accuracy : 0.8482142857142856
```

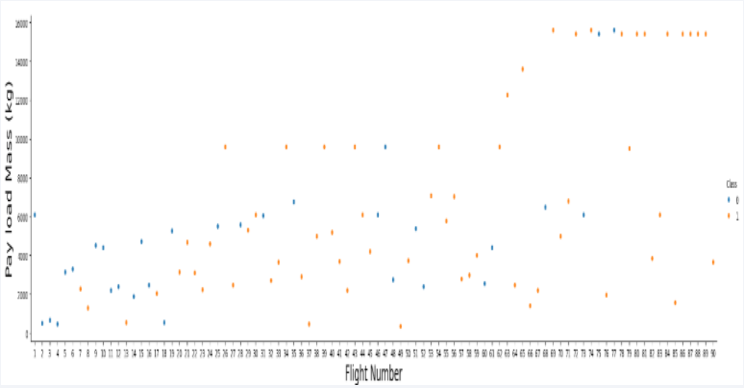

Github link: [Predictive Analysis]()

Results

ExploratoryData Analysis

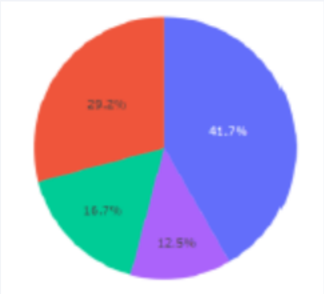
- The success of launches has progressively enhanced over time.
- KSC LC-39A boasts the highest success rate among all landing locations.
- The orbits ES-L1, GEO, HEO, and SSO have achieved a perfect success rate of 100%.





Visual Analytics

- The majority of launch sites are situated in proximity to the equator, with all being located near coastal areas.
- These sites are strategically positioned at a sufficient distance from potential hazards that a failed launch could impact, such as urban areas, highways, or railways, while remaining accessible for the transportation of personnel and materials.



Predictive Analytics

- The Decision Tree model is the most effective predictive model for the dataset.

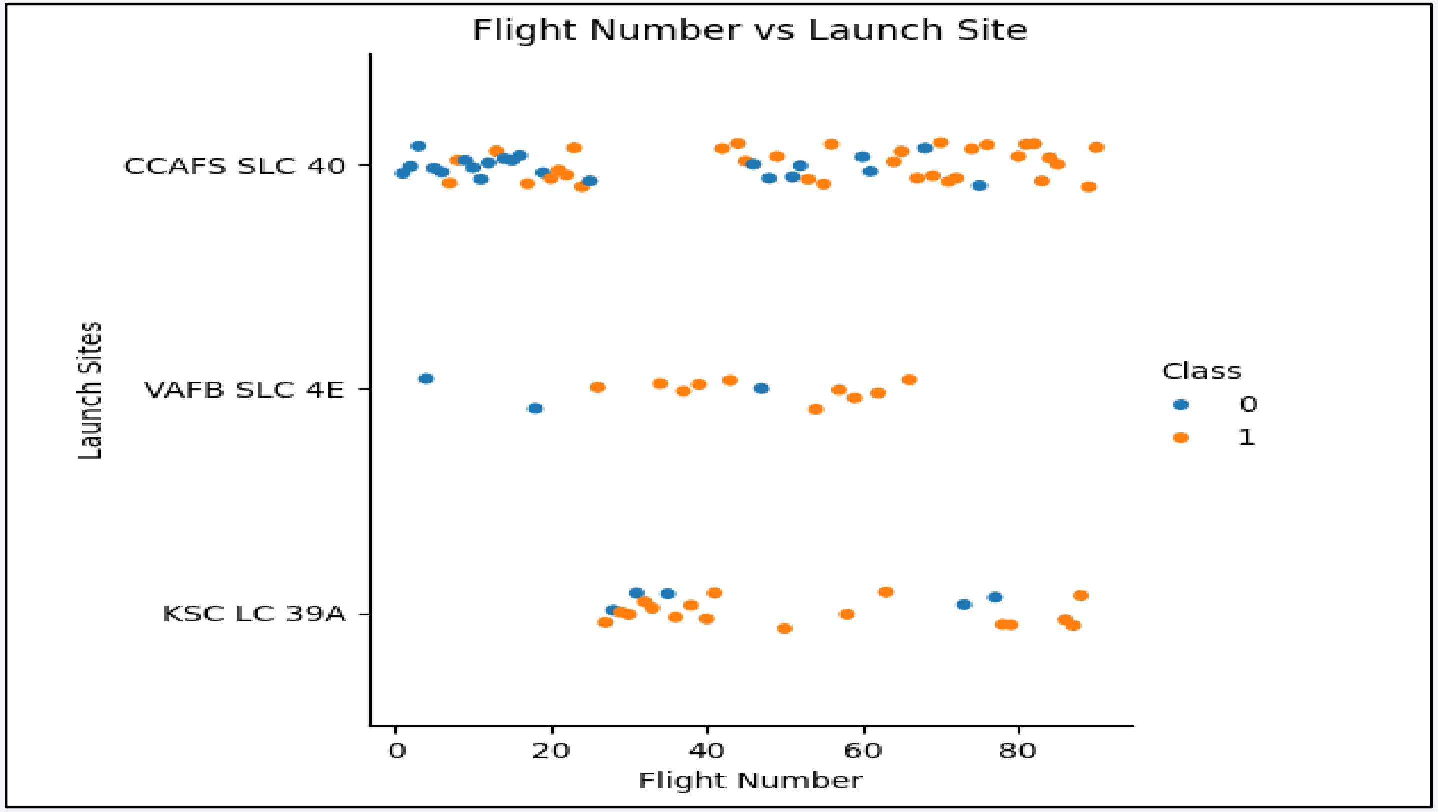
	ML Method	Accuracy Score (%)
0	Support Vector Machine	83.333333
1	Logistic Regression	83.333333
2	K Nearest Neighbour	83.333333
3	Decision Tree	66.666667

Section 2

Insights drawn from EDA

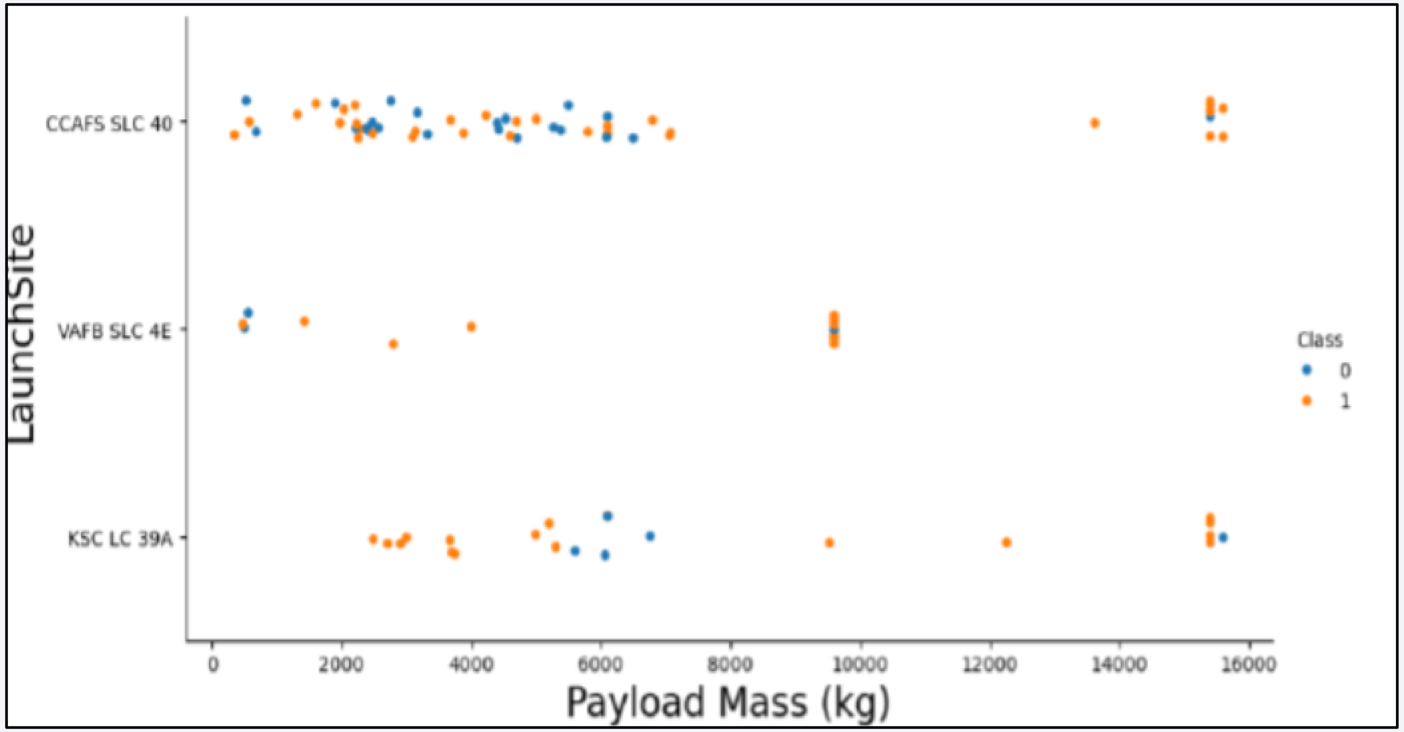
Flight Number vs. Launch Site

- **Exploratory Data Analysis**
- Earlier flights had a lower success rate (blue = fail). Later flights had a higher success rate (orange = success)
- Approximately fifty percent of the launches originated from the CCAFS SLC 40 launch site.
- VAFB SLC 4E and KSC LC 39A exhibit superior success rates.
- The data indicates that the success rate of flight launches improved as the number of flights increased.



Payload vs. Launch Site

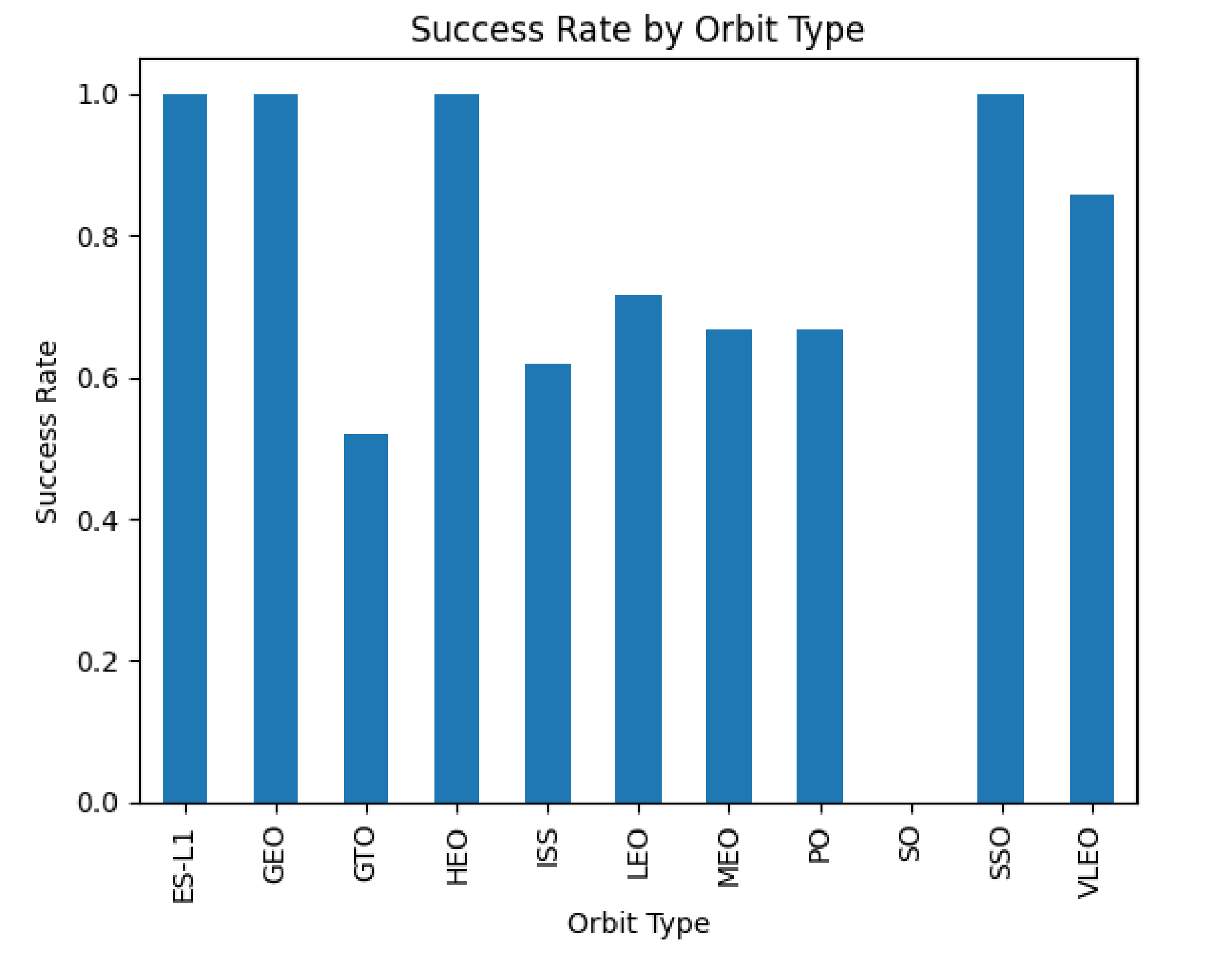
- At the launch site 'VAFB SLC 4E', there have been no rocket launches conducted for payloads exceeding 10,000 kg.
- The proportion of successful launches (Class=1) at the launch site 'VAFB SLC 4E' rises as the payload mass increases.
- There is no evident relationship or trend between the launch site and the payload.



Success Rate vs. Orbit Type

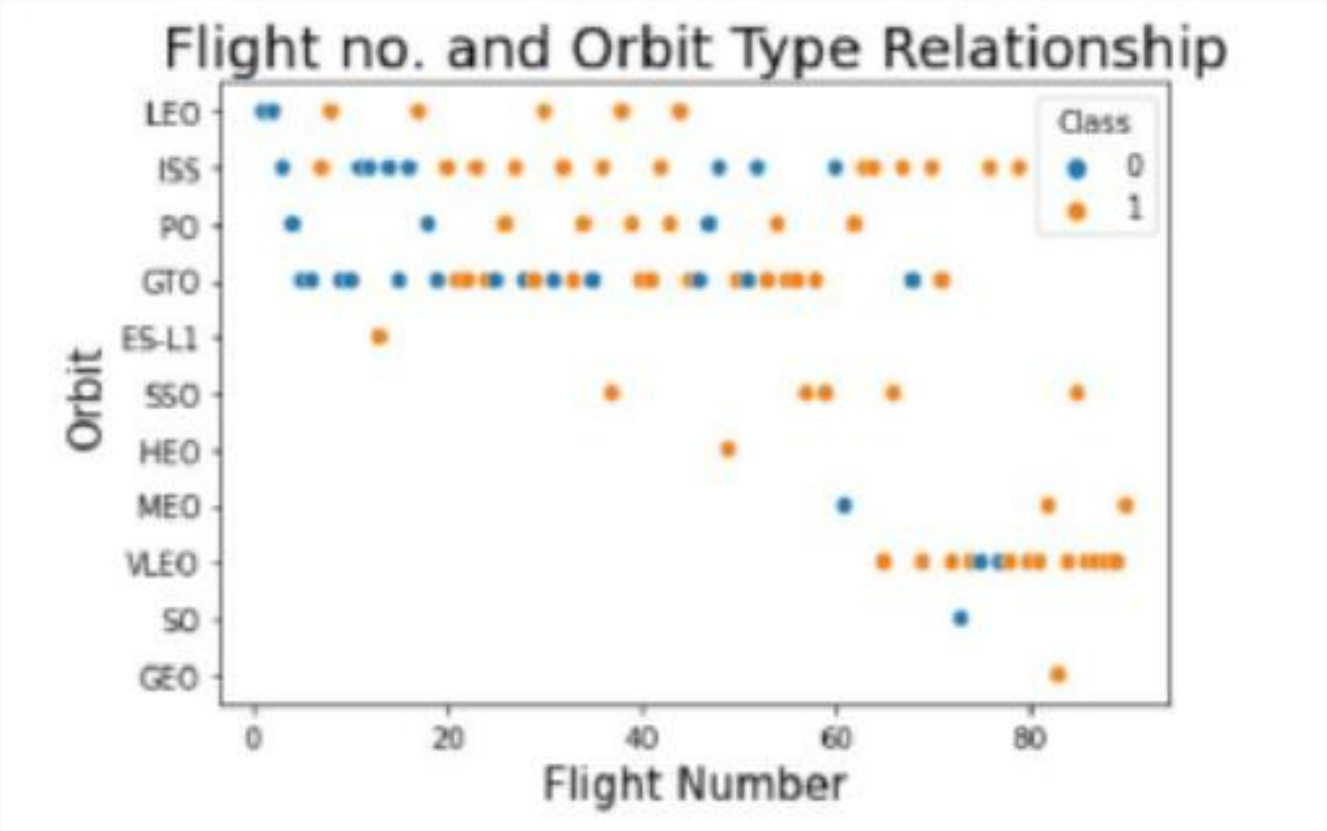
- Orbits ES-LI, GEO, HEO, and SSO have the highest success rates

- •GTO orbit has the lowest success rate



Flight Number vs. Orbit Type

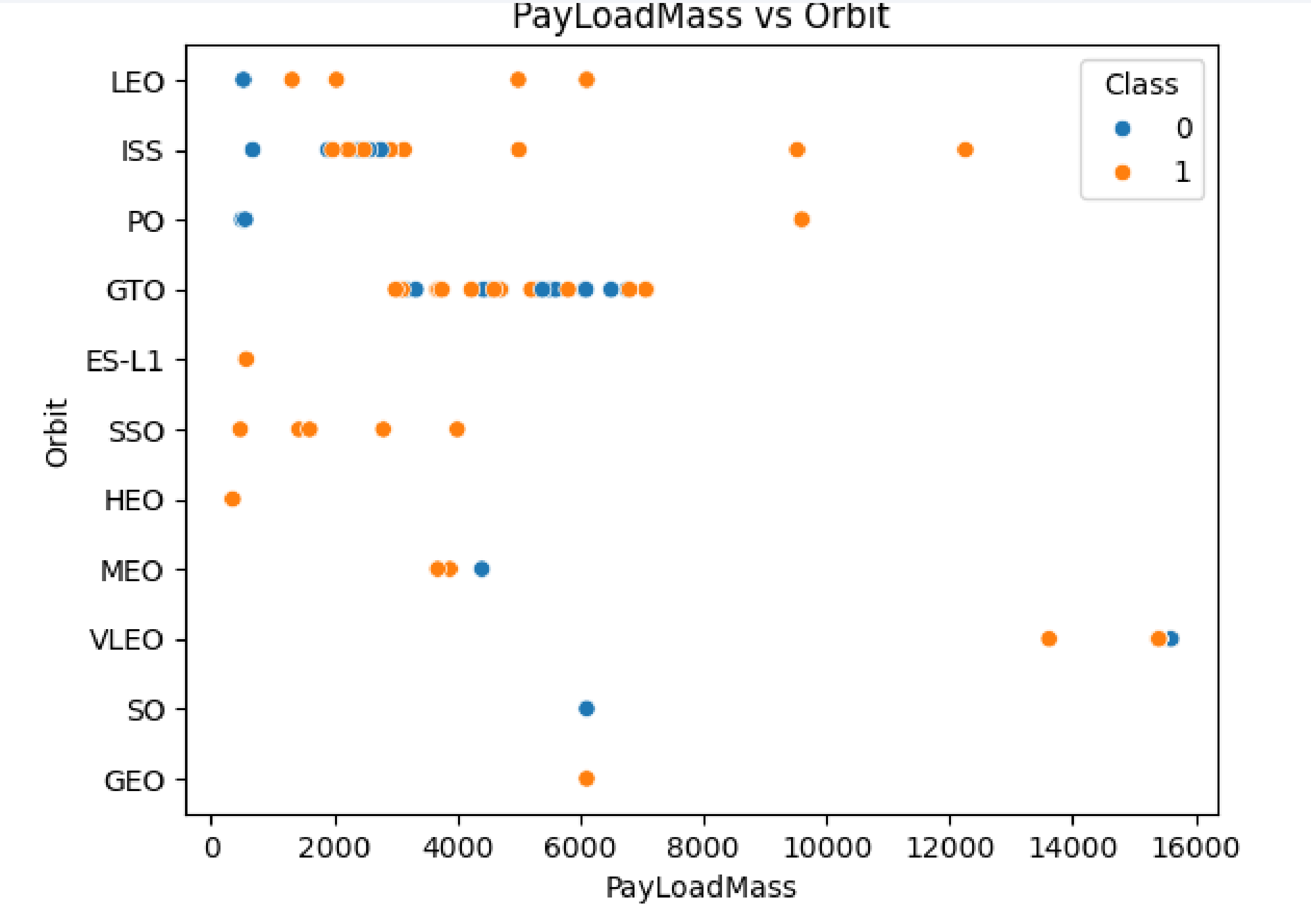
- The success rate typically increases with the number of flights for each orbit
- And its apparent on LEO orbit
- There is no relationship between flight number and orbit for GTO



Payload vs. Orbit Type

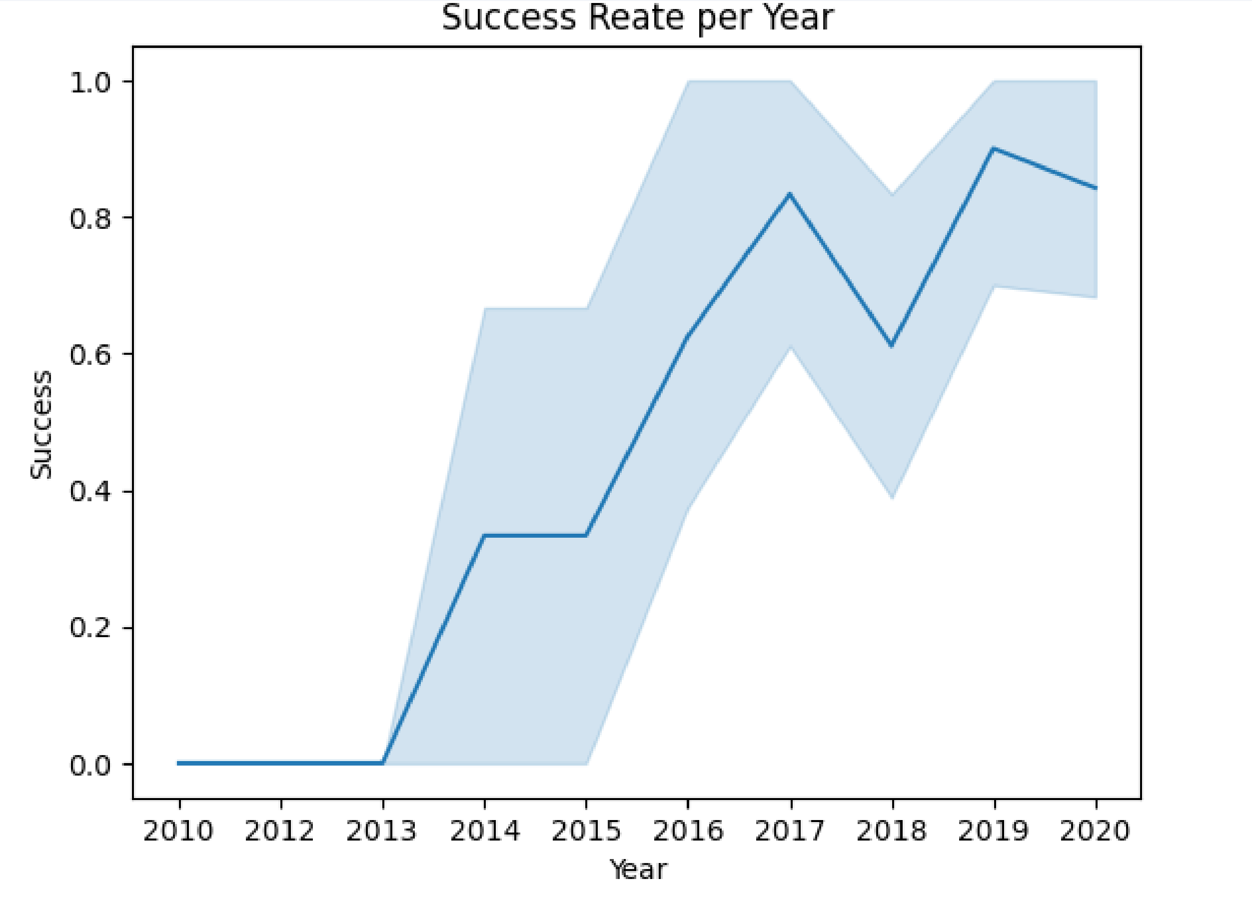
- Successful landing rates (Class=1) seem to improve as payload increases for the orbits of LEO, ISS, PO, and SSO.

- The GTO orbit experiences varying degrees of success when accommodating heavier payloads.

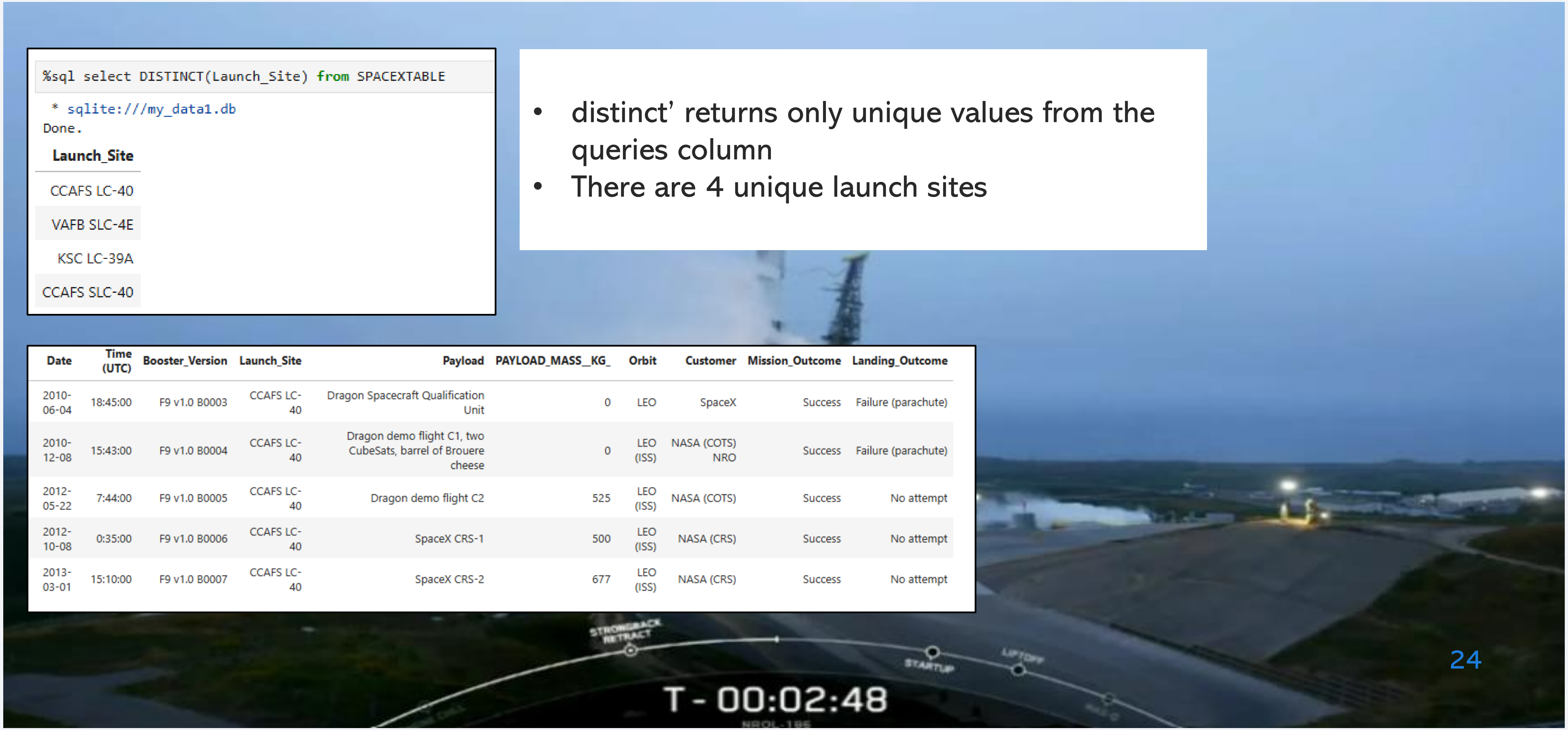


Launch Success Yearly Trend

- The success rate showed an enhancement between the periods of 2013-2017 and 2018-2019.
- The success rate experienced a decline between the years 2017 and 2018, as well as from 2019 to 2020.
- Overall, the success rate has improved since 2013



All Launch Site Names




```
%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

- distinct' returns only unique values from the queries column
- There are 4 unique launch sites

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

Launch Site Names Begin with 'CCA'

```
%sql select * from SPACEXTABLE WHERE Launch_Site LIKE 'CCA%' LIMIT 5;
```

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

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	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

- Using keyword ‘Like’ and format ‘CCA%’, returns records where ‘Launch_Site’ column starts with “CCA”.
- •Limit 5, limits the number of returned records to 5

Total Payload Mass

- **45,596 kg** (total) carried by boosters launched by NASA (CRS)

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

Average Payload Mass by 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

- **2,928 kg** (average) carried by booster version F9 v1.1

First Successful Ground Landing Date

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

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

MIN(Date)
2015-12-22

- The first successful ground landing of a Falcon 9 Stage One was on December 12, 2015.

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%sql SELECT DISTINCT(Booster_Version) FROM SPACEXTABLE WHERE Landing_Outcome='Success (drone ship)' AND PAYLOAD_MASS__KG_>4000 AND PAYLOAD_MASS__KG_<600
```

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

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

- The query identifies booster versions with a payload mass exceeding 4000 but less than 6000, while also ensuring that the landing outcome is successful on the drone ship.
- The use of the 'and' operator in the where clause retrieves booster versions that satisfy both conditions.

Total Number of Successful and Failure Mission Outcomes

```
%sql SELECT MISSION_OUTCOME, COUNT(*) as total_number \
FROM SPACEXTBL \
GROUP BY MISSION_OUTCOME;
```

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

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

- The 'group by' keyword organizes similar data within a column into distinct groups.
- In this instance, the various types of mission outcomes are categorized in the column based on their respective outcomes.‘counts’

Boosters Carried Maximum Payload

```
%sql SELECT DISTINCT(Booster_Version) FROM SPACEXTABLE WHERE PAYLOAD_MASS__KG_= (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTABLE);
```

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

Booster_Version
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

- The sub query returns the maximum payload mass by using keyword ‘max’ on the pay load mass column
- The main query returns booster versions and respective payload mass where payload mass is 15600

2015 Launch Records

```
SELECT substr(Date,6,2), Booster_Version, Launch_Site, Landing_Outcome FROM SPACEXTABLE WHERE Landing_Outcome='Failure (drone ship)' AND substr(Date,0,5)
```

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

substr(Date,6,2)	Booster_Version	Launch_Site	Landing_Outcome
01	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
04	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

- The inquiry details the landing results, booster version, and the launch site associated with a failed landing outcome on a drone ship, specifically for the year 2015.
- The use of the 'and' operator in the where clause retrieves booster versions that satisfy both conditions stated in the clause.
- The 'year' keyword is utilized to extract the year from the 'Date' column. The findings indicate that the launch site is 'CCAFS LC-40' and the booster version is F9 v1.1 B1012.

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

```
%sql SELECT Landing_Outcome, COUNT(*) AS Counting FROM SPACEXTABLE WHERE Date>='2010-06-04' AND Date <= '2017-03-20' GROUP BY Landing_Outcome ORDER BY
```

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

Landing_Outcome	Counting
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

We have conducted a query to obtain the count of landing outcomes from June 4, 2010, to March 20, 2017, arranged in descending order.

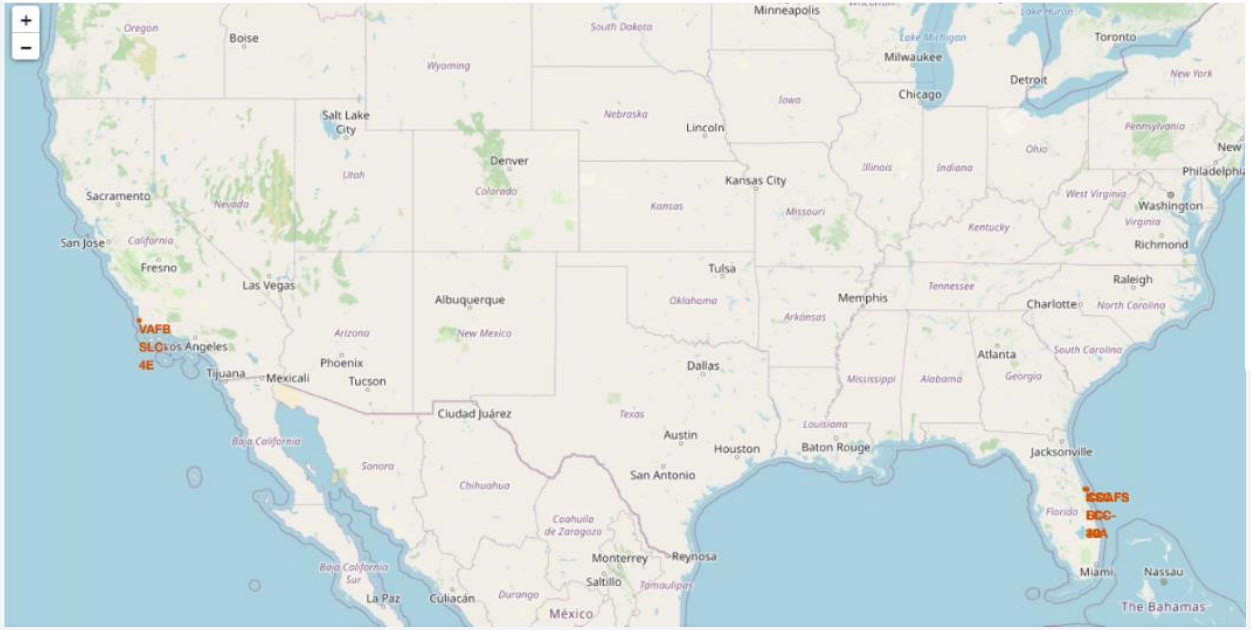
During this time period, there were several attempts at landings.

- 13 were successful
- 3 were planned failures
- 4 were unsuccessful
- And 1 was precluded (prevented)

Section 3

Launch Sites Proximities Analysis

LAUNCH SITES



	Launch Site	Lat	Long
0	CCAFS LC-40	28.562302	-80.577356
1	CCAFS SLC-40	28.563197	-80.576820
2	KSC LC-39A	28.573255	-80.646895
3	VAFB SLC-4E	34.632834	-120.610745

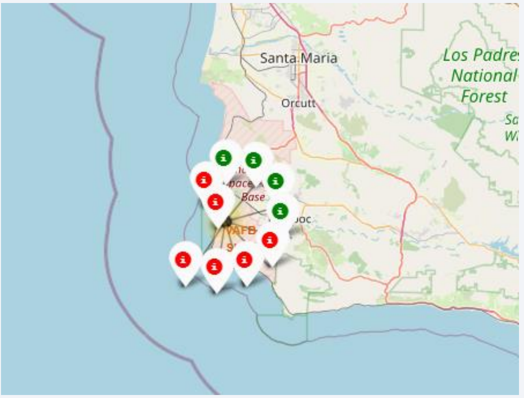
Visual/Interactive Analysis

The markers indicate the positions of the four SpaceX launch sites, with one located in California and three situated in Florida.

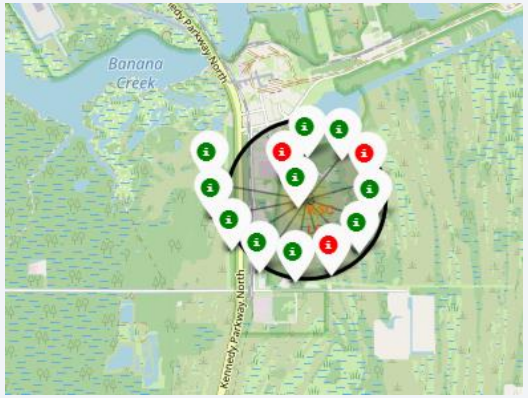
All are as near the equator

The proximity of the launch site to the equator significantly facilitates the process of launching a rocket into orbit. Launching rockets from locations near the equator benefits from an additional natural advantage, attributed to the Earth's rotational speed. This advantage contributes to cost savings by reducing the need for extra fuel and boosters.

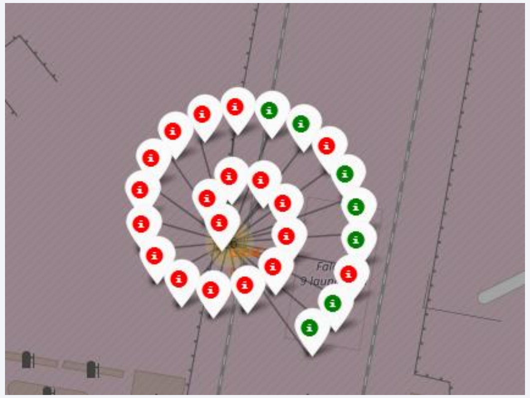
LAUNCH OUTCOME



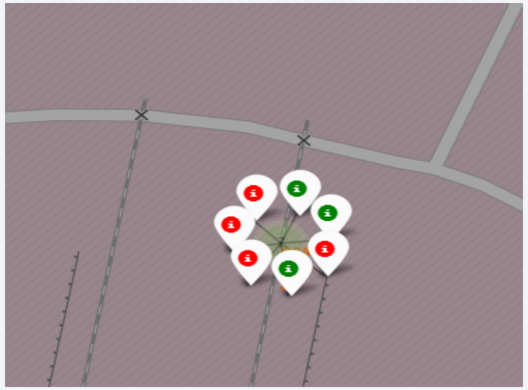
VAFB SLC-4E



KSC LC-39A



CCAFS SLC-40



CCAFS SLC-40

- This illustration depicts the tally of successful and unsuccessful launch outcomes (stage one landings) at each launch site.
- Successful launches are indicated by green markers, while unsuccessful launches are represented by red markers.
- Launch site KSC LC-39 boasts the highest number of successful launches, achieving 20 out of 27, which corresponds to a success rate of 76.9%.

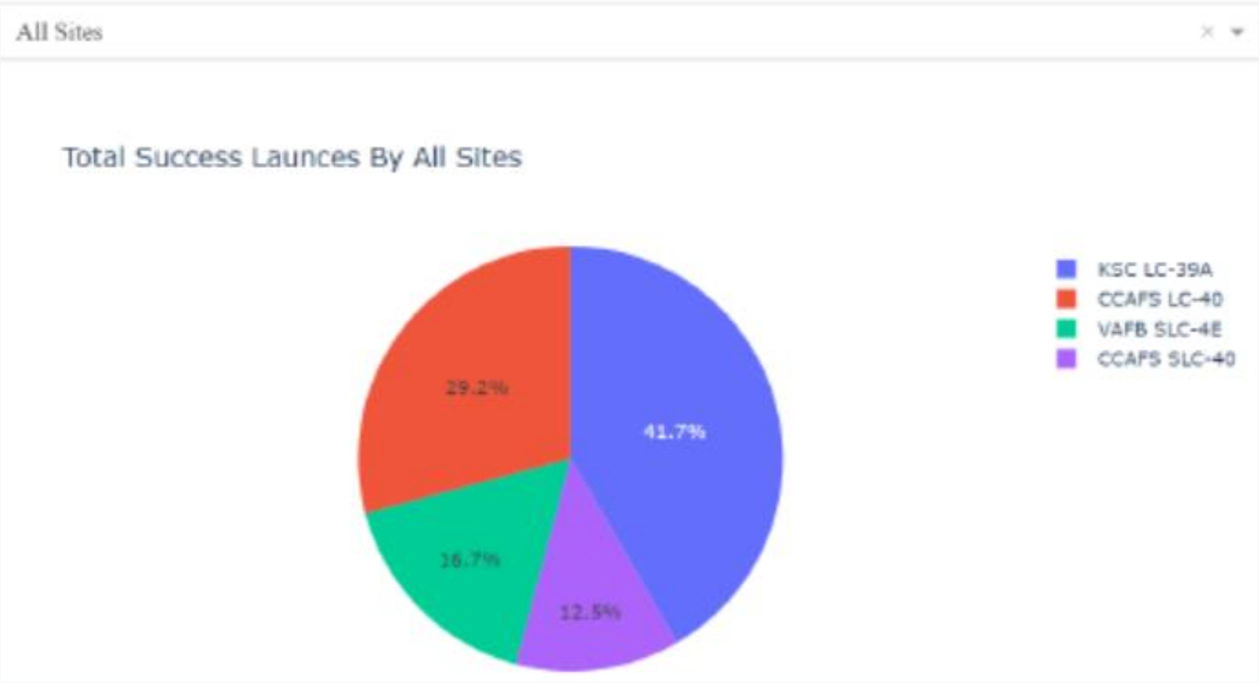
DISTANCE TO PROXIMITIES

The Vandenberg Space Launch Site, located in California, is positioned near the ocean and sufficiently distanced from the nearest city to ensure safety for the general public, while still being accessible for the operational requirements of the space station. This characteristic is consistent across all four launch sites.

Section 4

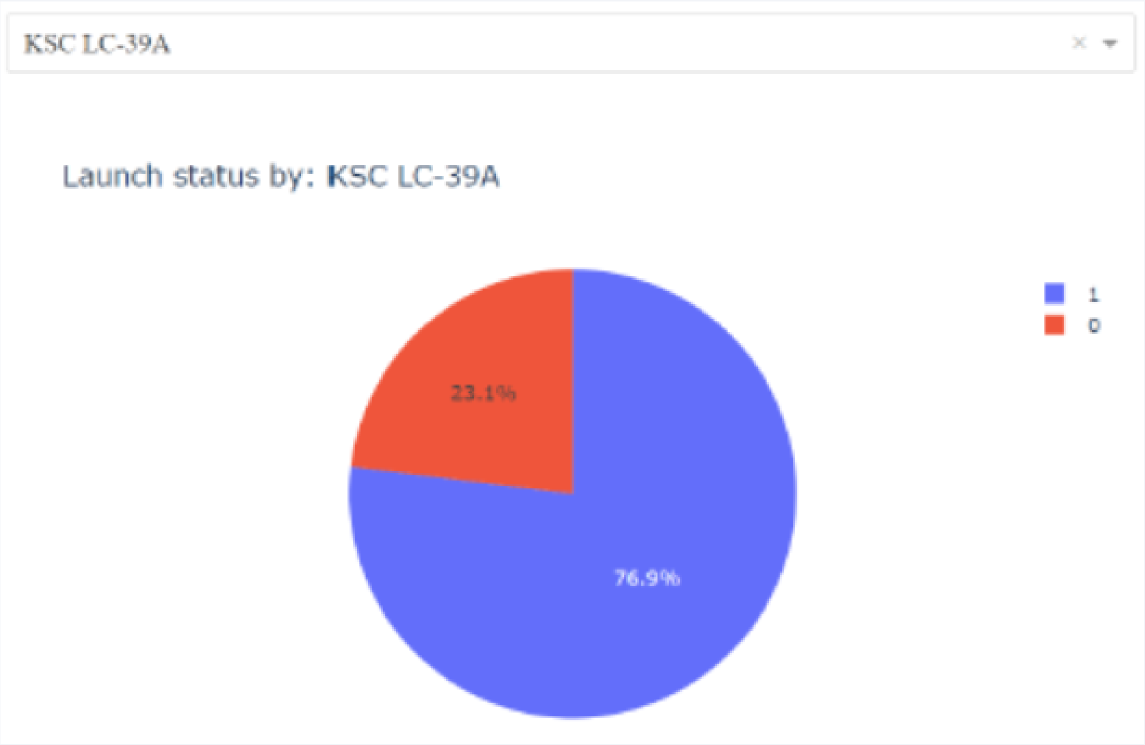
Build a Dashboard with Plotly Dash

LAUNCH SUCCESS BY SITE



- •Launch Site ‘KSC LC-39A’ has the highest launch success rate
- •Launch Site ‘CCAFS SLC-40’ has the lowest launch success rate

LAUNCH SITE WITH HIGHEST SUCCESS RATIO



- KSC LC-39A Launch Site has the highest launch success rate and count
- Launch success rate is 76.9%
- Launch successfailurerateis23.1%

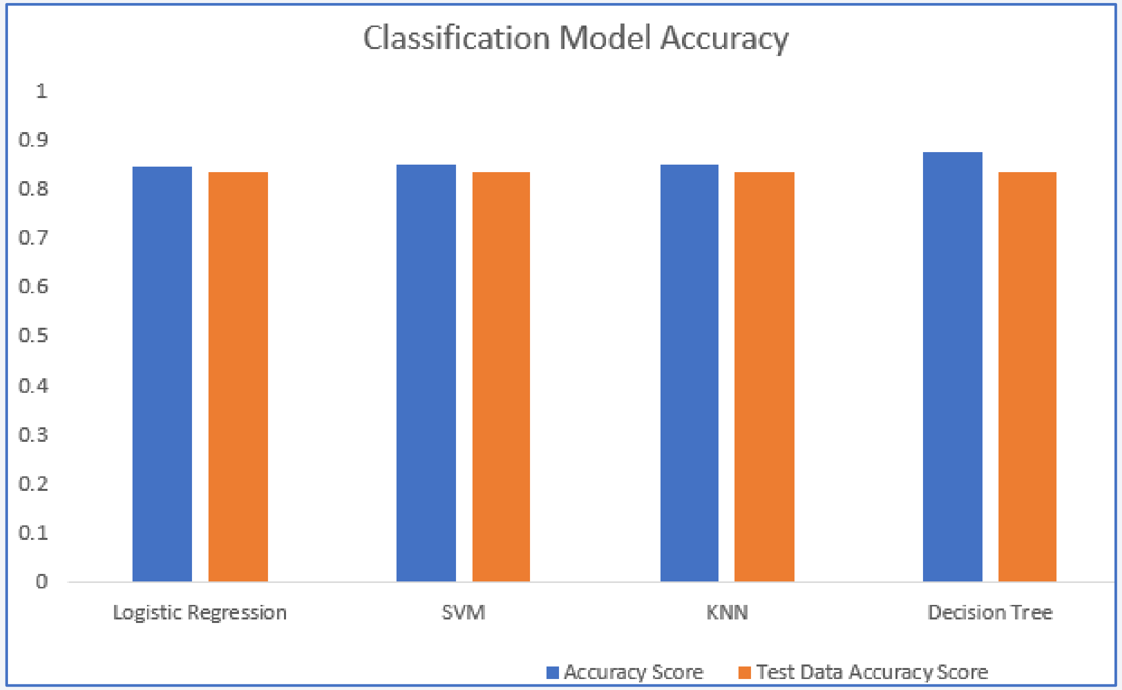
PAYLOAD MASS AND OUTCOME

- Payloads between 2,000 kg and 4,000 kg have the highest success rate
- 1 indicating successful outcome and 0 indicating an unsuccessful outcome
- Most of the successful outcomes were from launches using the FT Booster Version

Section 5

Predictive Analysis (Classification)

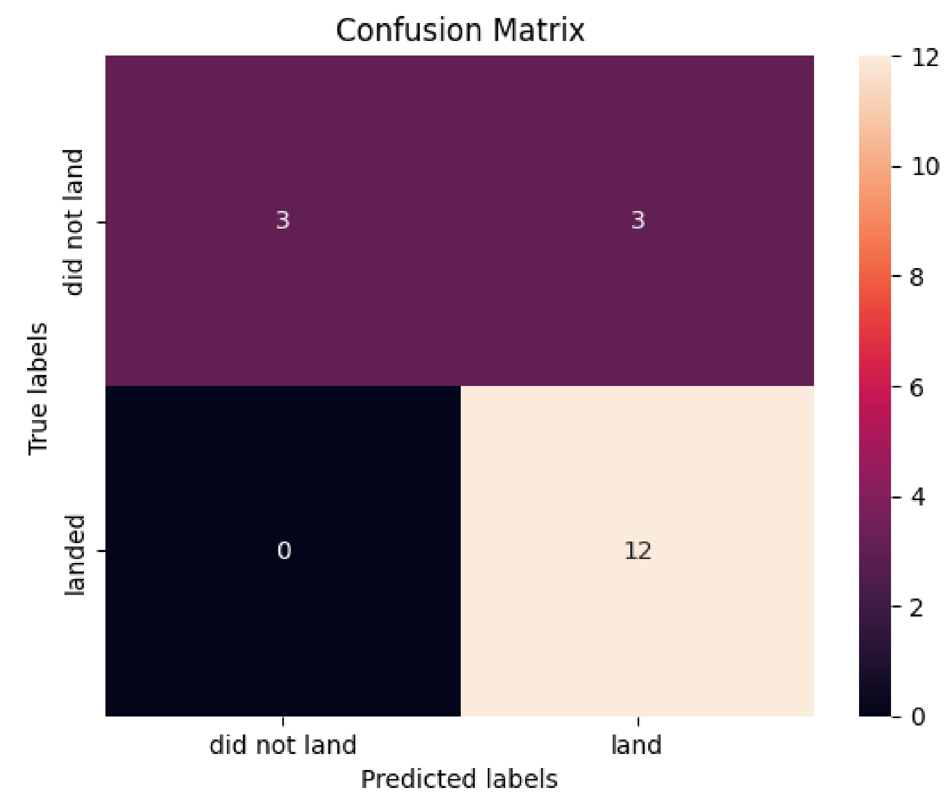
Classification Accuracy



	Algo Type	Accuracy Score	Test Data Accuracy Score
2	Decision Tree	0.875000	0.833333
3	KNN	0.848214	0.833333
1	SVM	0.848214	0.833333
0	Logistic Regression	0.846429	0.833333

- The Decision Tree algorithm demonstrates the highest classification score, as indicated by the Accuracy scores and illustrated in the bar chart, achieving a value of 0.8750.
- Furthermore, the Accuracy Score for the test data remains consistent across all classification algorithms within the dataset, recorded at a value of 0.8333.
- .best_score_ is the average of all cv folds for a single combination of the parameters

Confusion Matrix



- A confusion matrix provides a comprehensive overview of the effectiveness of a classification algorithm
- All confusion matrices derived from the test data were the same across all four models.
- **Confusion Matrix Outputs:**
    - 12 True positive
    - 3 True negative
    - 3 False positive
    - 0 False Negative

Conclusions

- Predictive analysis indicates that the performance of all four models on the test set is comparable, with the decision tree model exhibiting a marginal advantage when evaluated using the .best_score_ function. Each of the four models achieved a success rate of 83.3% with the test data.

- The models exhibited comparable results on the test set, with the decision tree model demonstrating a marginal advantage.

- A majority of the launch sites are situated close to the equator, providing an extra natural advantage due to the Earth's rotational velocity.

- As the frequency of flights rises, the likelihood of a successful landing during the initial phase also increases. Furthermore, while success rates seem to improve with an increase in payload, there is no evident correlation between payload mass and success rates.

- KSC LC-39A boasts the highest success rate of any launch site, achieving a flawless 100% success rate for missions involving payloads under 5,500 kg. Additionally, it has maintained a perfect success rate for orbits including ES-L1, GEO, HEO, and SSO.

Conclusions

Things to Consider

- A more extensive data set could enhance the effectiveness of predictive analytics; however, as the data set expands over time, the influencing factors also evolve and improve.

- The average number of reuses for a Falcon 9 stage one booster is a topic of interest in the aerospace community.

Appendix

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

Thank you!