IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Sanjana Kumar
02/23/2024

Agenda

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

Executive Summary

- Objective : To predict if Space X Falcon 9 first stage will land successfully
- Methodology Steps: Data Collection, Data Wrangling, Exploratory Data Analysis, Interactive Visual Analytics, Predictive Analysis
- Summary of the Results: As the number of flights increase, so does the rate of successful landings. The machine learning models, a total of 4, all yielded an accuracy rate similar to 83.33%. Since the sample size of the test data was small, a best model was unable to be determined.

Introduction

- Space X is an American spacecraft manufacturer. The company currently operates the Falcon 9 rocket along with other spacecrafts. Space X launches Falcon 9 rockets at a cost of around $62 million. This cheaper compared to other providers that can cost $165 million dollars. This is due to Space X being able to recover part of the rocket when it lands allowing them to be able to reuse that part.

- Being able to make predictions if the first stage of a Falcon 9 rocket will land, we can determine the cost of a launch. Not only will we be able to determine the cost for Space X's financial needs but to use this information to assess whether or not an alternate company should work(fund) with Space X for a rocket launch.

Section 1

Methodology

Methodology

Data Collection

- Combined data from SpaceX REST API and web scraped data from the SpaceX Wikipedia page.

Data Wrangling

- Convert binary variables to binary values, 0’s and 1’s.
- Removing null values or replacing any missing data with their respective column’s mean

Exploratory Data Analysis

- Evaluated the dataset, visualize relationships between variables and determine values

Interactive Visual Analytics

- Geospatial Analytics of the launch sites

Data Modeling and Evaluation

- Implemented different machine algorithms such as logistic regression and decision trees and plotted Confusion Matrices for each model to assess the accuracy

Data Collection

- The data collected was a combination of API requests from Space X REST API and web scrapped data from a table in Space X's Wikipedia entry.
- Space X API Variables: FlightNumber, Date, BoosterVersion, PayloadMass, Orbit, LaunchSite, Outcome, Flights, GridFins, Reused, Legs, LandingPad, Block, ReusedCount, Serial, Longitude, Latitude
- Wikipedia Webscrape Data Variables: Flight No., Launch site, Payload, PayloadMass, Orbit, Customer, Launch outcome, Version Booster, Booster landing, Date, Time

Data Collection – SpaceX API

- Make Get Responses to the SpaceX REST API
- Convert the response to a JSON file
- Json-Normalize to a dataframe
- Created a dataframe from a dictionary dataset
- Filter the dataset to only contain Falcon 9 launches
- Impute the missing values with the mean of the respective column

GitHub Link

https://github.com/Sku002/IBM-Data-Science-Capstone-SpaceX/blob/main/SpaceX%20Data%20Collection%20API.ipynb

Data Collection – Scrapping

- Requst the HTML page from the static url
- Create a BeautifulSoup object from the html response object
- Find the launch information html table
- Find the column names from the tables found
- Use the column names as keys for a dictionary
- Convert the dictionary to a dataframe

GitHub Link

https://github.com/Sku002/IBM-Data-Science-Capstone-SpaceX/blob/main/SpaceX%20Webscrapping.ipynb

Data Wrangling

- Convert binary variables, landing_outcomes and mission_outcome, to binary values, 0's and 1's.
- Removing null values or replacing any missing data with their respective column's mean

GitHub Link
https://github.com/Sku002/IBM-Data-Science-Capstone-SpaceX/blob/main/SpaceX%20Data%20%20Wrangling.ipynb

EDA with Data Visualization

Scatter plots, line charts, and bar plots were used to compare relationships between variables to decide if a relationship exists so that they could be used in training the machine learning model

GitHub Link

https://github.com/Sku002/IBM-Data-Science-Capstone-SpaceX/blob/main/SpaceX%20EDA%20with%20Data%20Visualization%20Tools.ipynb

EDA with SQL

Queried information about launch sites, the mission outcomes, payload masses of customers, booster versions, and the landing outcome

GitHub

https://github.com/Sku002/IBM-Data-Science-Capstone-SpaceX/blob/main/SpaceX%20EDA%20with%20SQL.ipynb

Build an Interactive Map with Folium

- The folium map marked launch sites, success and fail landings and a proximity line to major locations such as railways, highways and the coast.
- Marking the map helps to visualize the area of launch and helps us to understand why the launch areas are located where they are.

GitHub Link

https://github.com/Sku002/IBM-Data-Science-Capstone-SpaceX/blob/main/SpaceX%20Interactive%20Map%20Visual.ipynb

Build a Dashboard with Plotly Dash

- The dashboard contains a pie chart and a scatter plot
- Pie chart can be selected to show the distribution of successful landing across the launch sites, making the visualization of success rates easier.
- The scatter plot shows how the successful landings vary across the the launch sites, payload masses, and the booster version variables.

GitHub Link

https://github.com/Sku002/IBM-Data-Science-Capstone-SpaceX/blob/main/SpaceX%20Interactive%20%20Plotly%20Dash

Predictive Analysis (Classification)

- Use sting setStandard Scaler to fit and transform the features(variables)
- Split the dataset into a training set and test
- Determine what machine learning models are appropriate for the dataset and objective
- Create a GridSearchCV object for each model
- Use the GridSearchCV on your models
- Score models on the split test set
- Create a Confusion Matrix for all the models
- Created a table to compare the accuracy results

GitHub Link

https://github.com/Sku002/IBM-Data-Science-Capstone-SpaceX/blob/main/SpaceX%20Machine%20Learning.ipynb

Results

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

Section 2

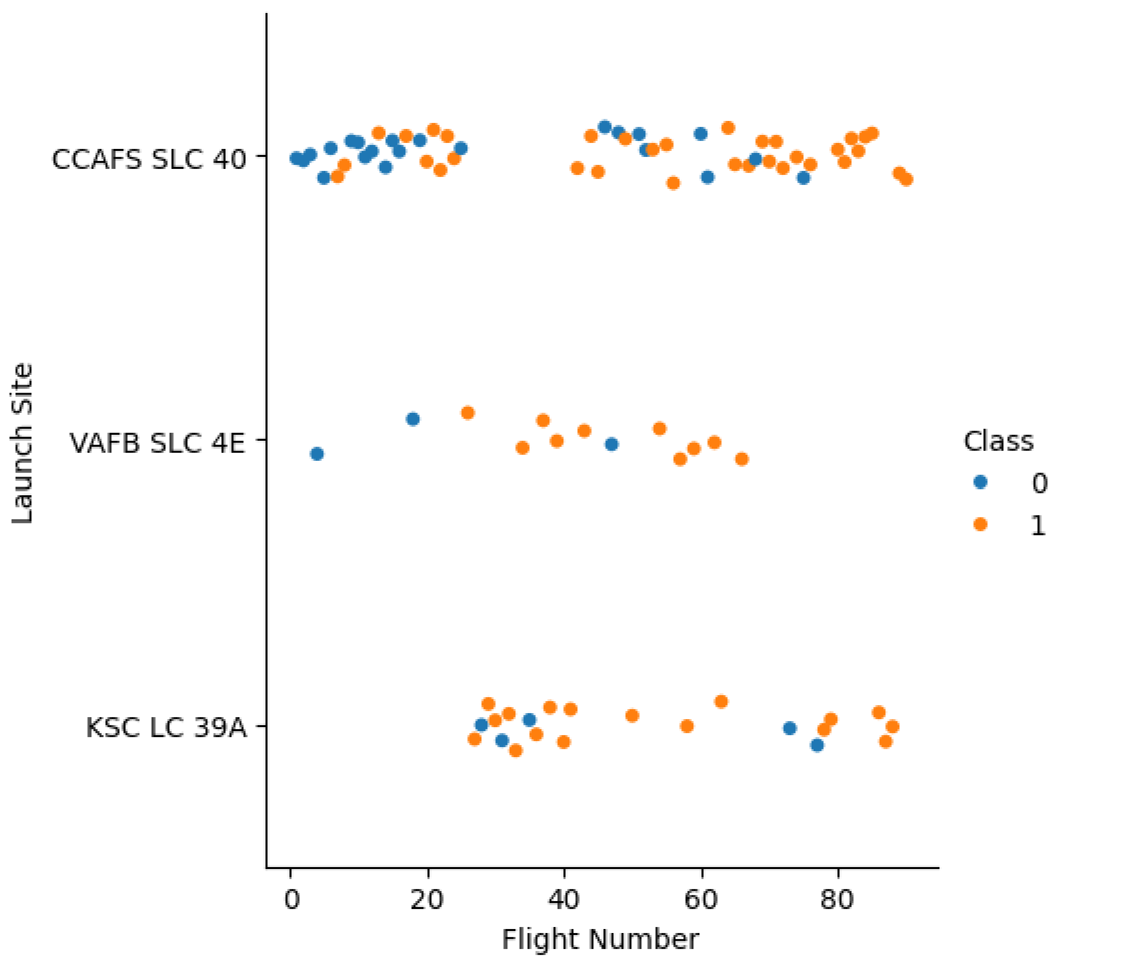
Insights drawn from EDA

Flight Number vs. Launch Site

Blue Dots = Failure

Orange Dots = Success

- As the number of flights increase, so does the launch site success
- Flight numbers less than 30 (early flights), as seen in CCAFS SLC 40 and VAFB SLC 4E, were generally unsuccessful in contrast to flight numbers greater than 30 (later flights), that had more successful landings.
- Above a flight number of 30, we can see that there are more successful landings.

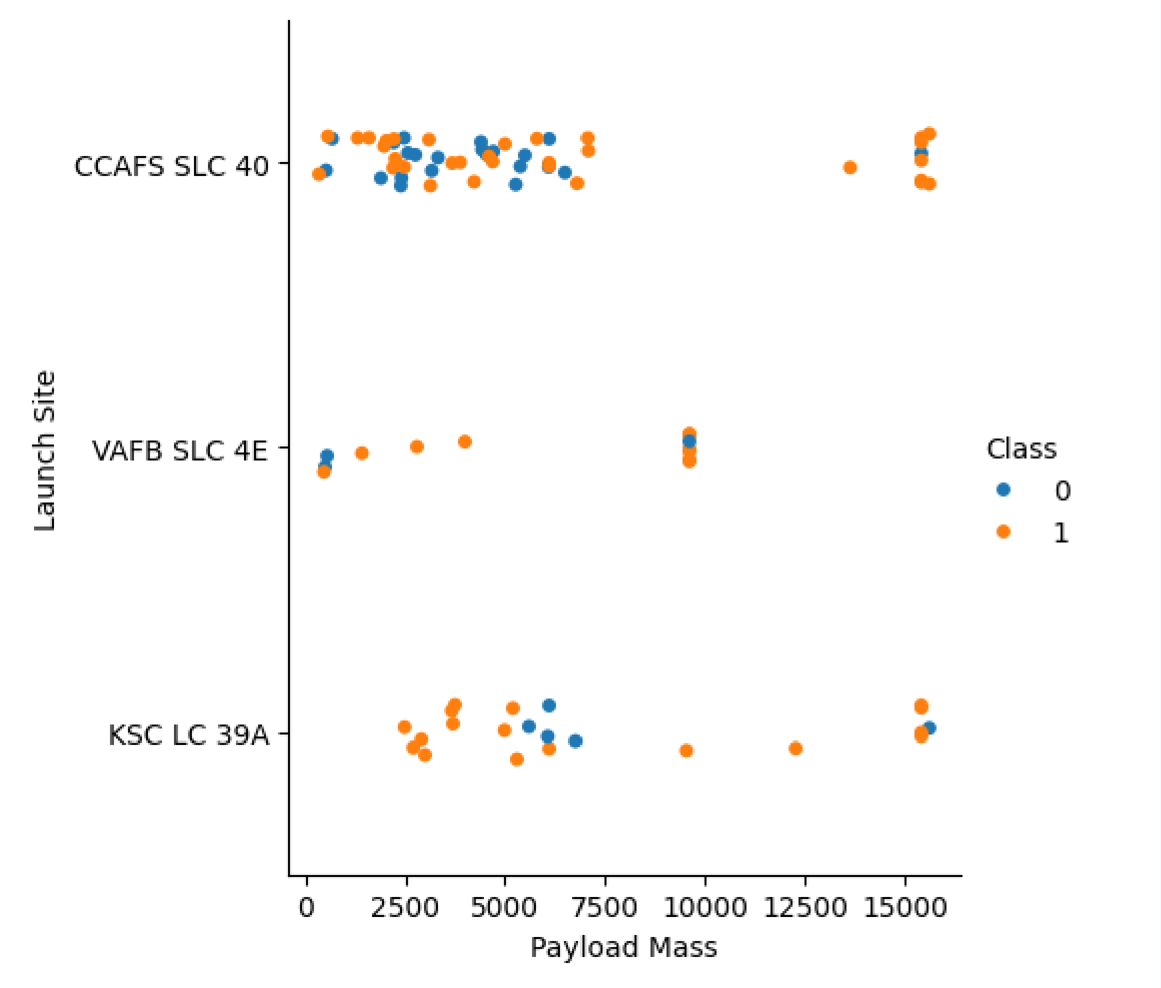


Payload vs. Launch Site

Blue Dots = Failure

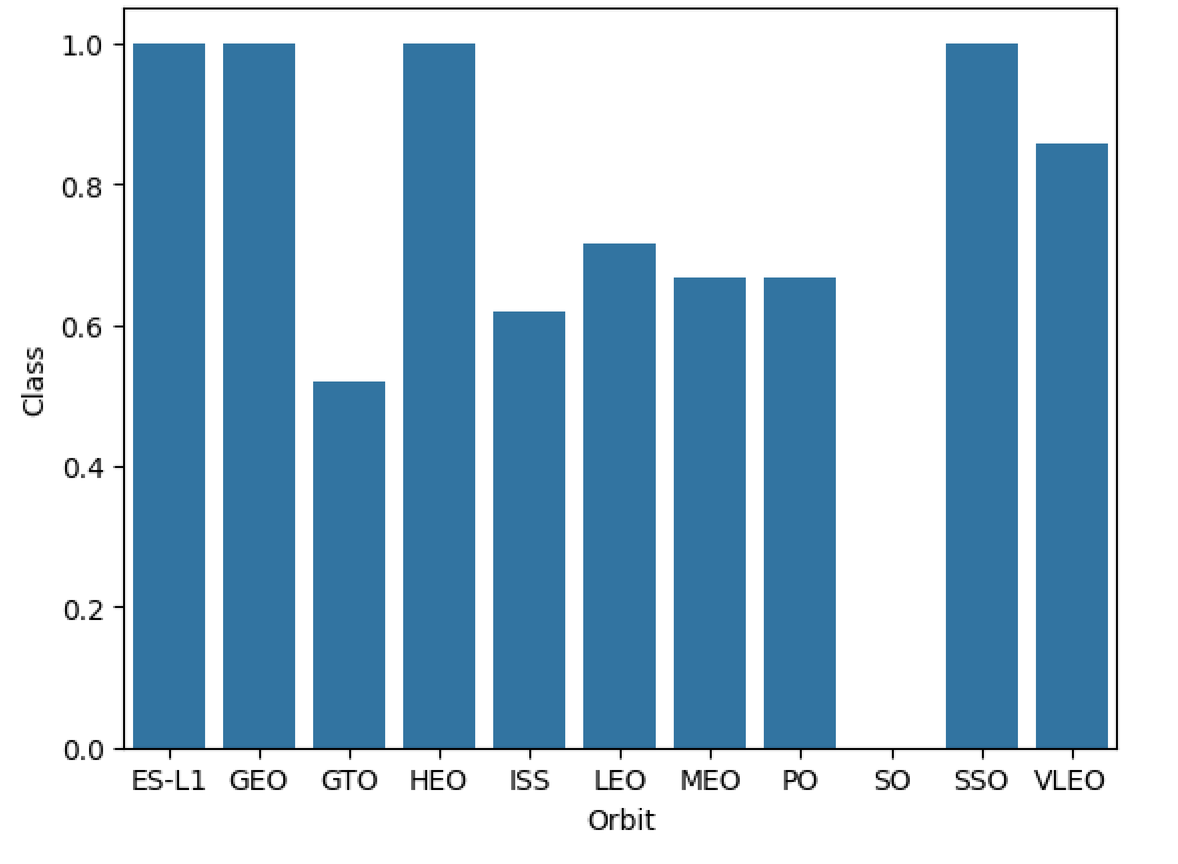
Orange Dots = Success

- There is no clear correlation between the payload mass and success rate for a launch site.
- Payload Masses greater than 8000 kg, there are very few unsuccessful landings however there is also far less data for heavier launches.
- Most of the launches came from CCAFS SLC 40 that were lighter launches however all sites have launched a variety of payload masses.



Success Rate vs. Orbit Type

- The 100% success rate of GEO, HEO, SO and ES-L1 is due to having only 1 flight into their respective orbits.
- The 100% success rate of SSO orbit is the most impressive with it's 5 successful flights and landings.
- Generally speaking, as flight numbers increase so does the success rate. However, LEO had unsuccessful landings occurring in very early flights and very quickly had successful flights.

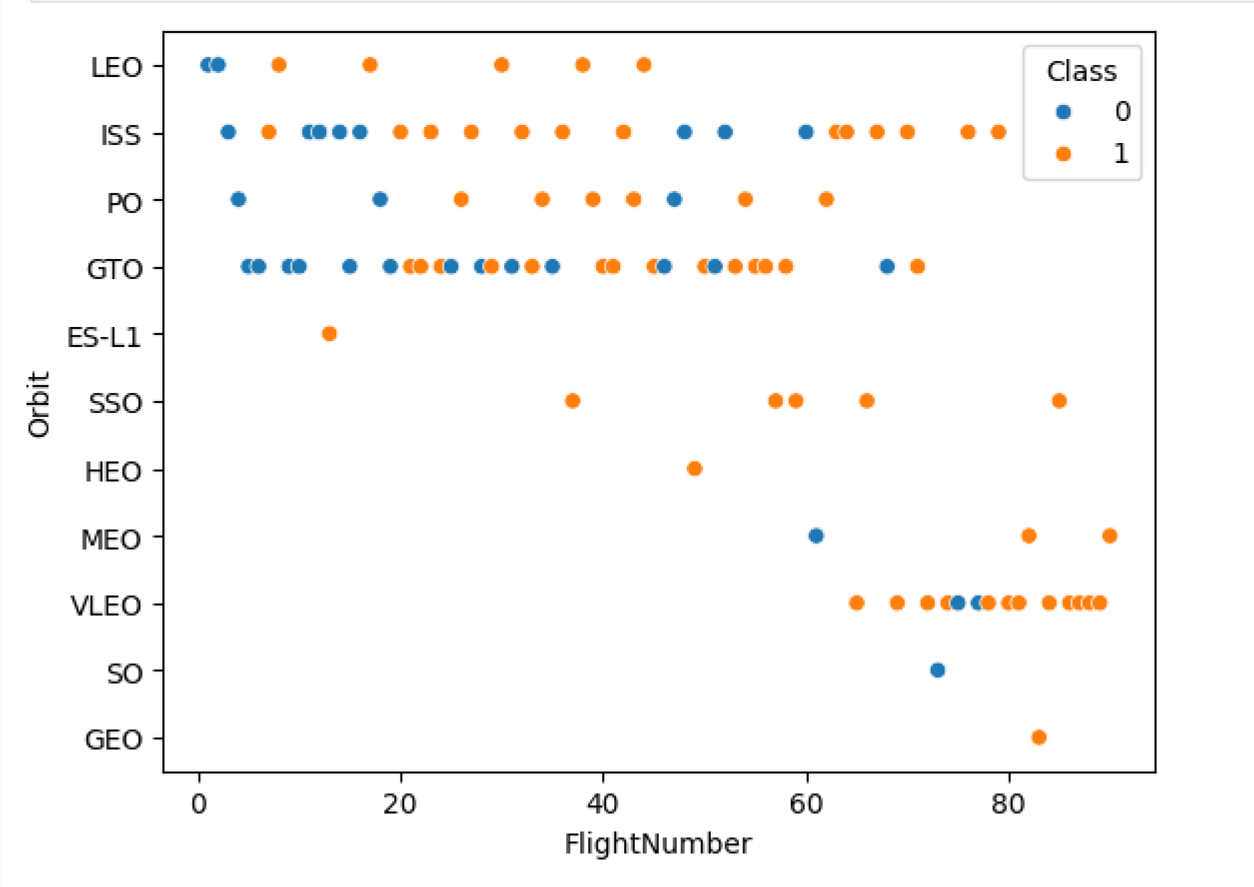


Flight Number vs. Orbit Type

Blue Dots = Failure

Orange Dots = Success

- The 100% success rate of GEO, HEO, SO and ES-L1 is due to having only 1 flight into their respective orbits.
- The 100% success rate of SSO orbit is the most impressive with it's 5 successful flights and landings.
- Generally speaking, as flight numbers increase so does the success rate. However, LEO had unsuccessful landings occurring in very early flights and very quickly had successful flights.

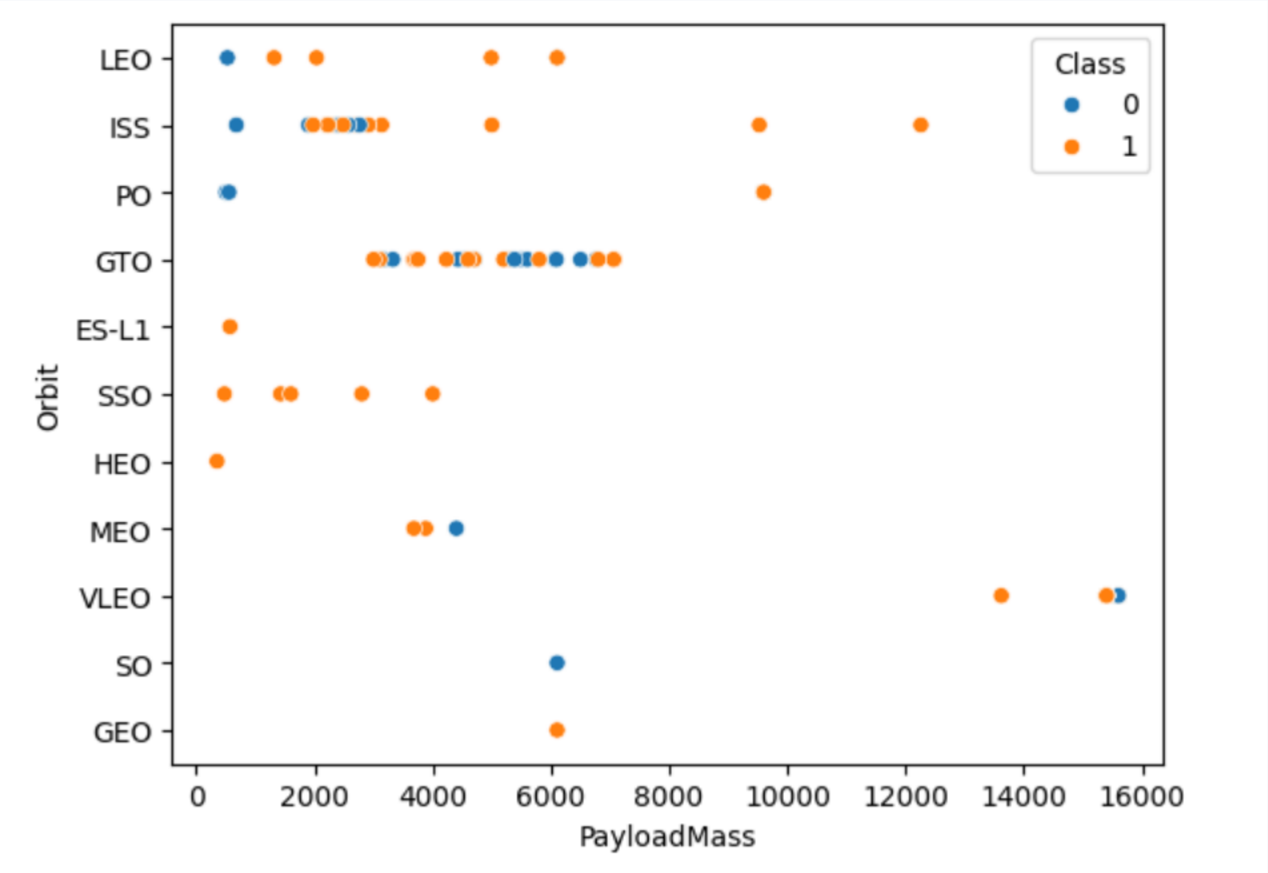


Payload vs. Orbit Type

Blue Dots = Failure

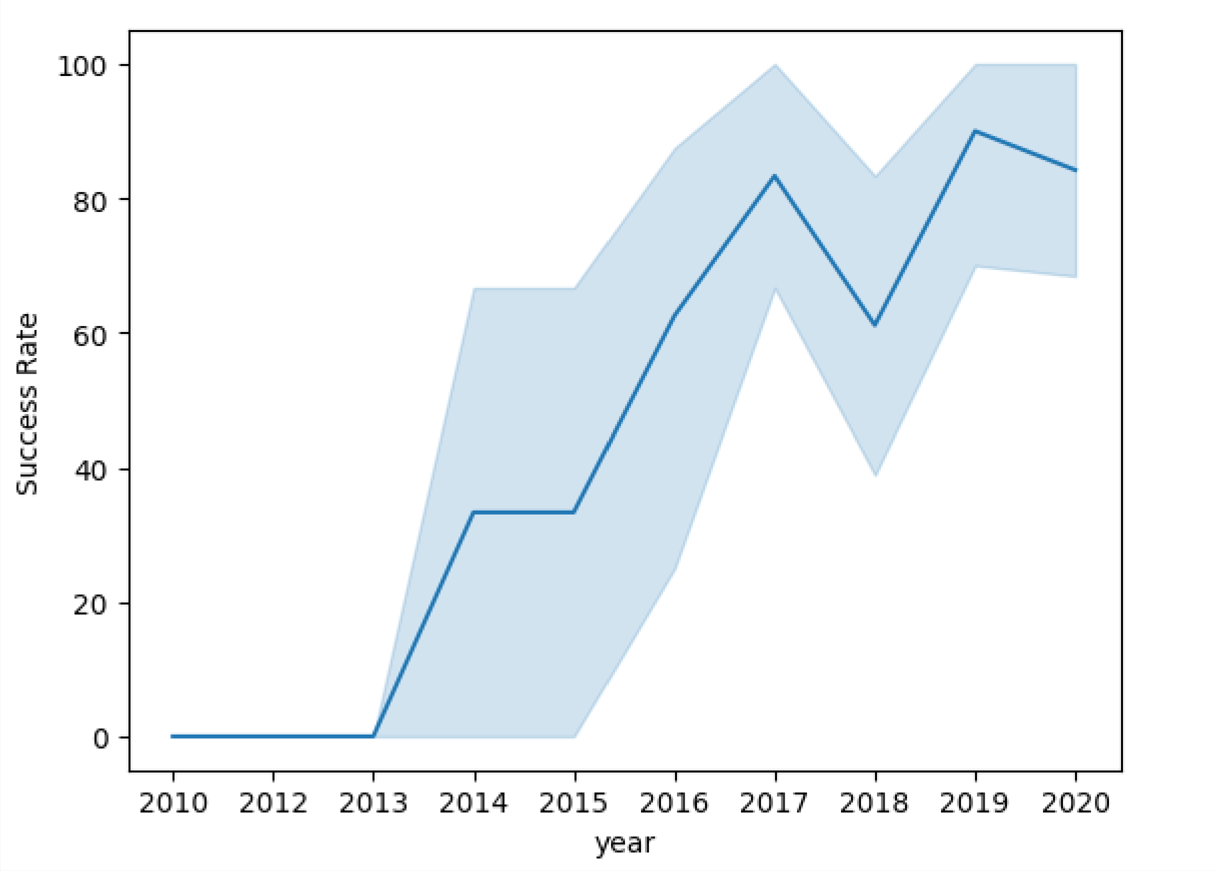
Orange Dots = Success

- The orbit types PO, LEO, ISS have more success with heavier payload masses.
- VLEO seems to only be associated with heavy payload masses, above 12,000kg, however it does have successful landings with them.



Launch Success Yearly Trend

- Between 2013 to 2017, the success rate continue to grow.
- Between 2010 to 2013 all the landings were unsuccessful.
- After 2016 and there on, there was always a greater chance of 50% success.



All Launch Site Names

- The UNIQUE key returns only the unique values from the Launch Site column of the SPACEXDATASET

- CCAFS LC -40, CCAFS SLC-40, CCAFSSLC-40, seem to be the same launch site based off previous analysis. It is likely that there are only 3 unique launch sites.

```
4]: %%sql
    SELECT UNIQUE LAUNCH_SITE
    FROM SPACEXDATASET;

 * ibm_db_sa://ftb12020:***@0c77d6f
Done.
4]:
```

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

Launch Site Names Begin with 'CCA'

First 5 records where launch sites begin with `CCA`

```
%%sql
SELECT *
FROM SPACEXDATASET
WHERE LAUNCH_SITE LIKE 'CCA%'
LIMIT 5;
```

```
 * ibm_db_sa://ftb12020:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31198/bludb
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	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	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

The total payload carried by boosters from NASA

```
%%sql
SELECT SUM(PAYLOAD_MASS__KG_) AS SUM_PAYLOAD_MASS_KG
FROM SPACEXDATASET
WHERE CUSTOMER = 'NASA (CRS)';
```

```
 * ibm_db_sa://ftb12020:***@0c77d6f2-5da9-48a9-81f8-86
Done.
```

sum_payload_mass_kg
45596

Average Payload Mass by F9 v1.1

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

```
%%sql
SELECT AVG(PAYLOAD_MASS__KG_) AS AVG_PAYLOAD_MASS_KG
FROM SPACEXDATASET
WHERE booster_version = 'F9 v1.1'
```

```
 * ibm_db_sa://ftb12020:***@0c77d6f2-5da9-48a9-81f8-86
Done.
```

avg_payload_mass_kg
2928

First Successful Ground Landing Date

The date of the first successful landing

```
%%sql
SELECT MIN(DATE) AS FIRST_SUCCESS
FROM SPACEXDATASET
WHERE landing__outcome = 'Success (ground pad)';
```

```
 * ibm_db_sa://ftb12020:***@0c77d6f2-5da9-48a9-81
Done.
```

first_success
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

The names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

```
%%sql
SELECT booster_version
FROM SPACEXDATASET
WHERE landing__outcome = 'Success (drone ship)' AND payload_mass__kg_ BETWEEN 4001 AND 5999;
```

```
 * ibm_db_sa://ftb12020:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.database
Done.
```

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

Total Number of Successful and Failure Mission Outcomes

The total number of successful and failure mission outcomes

```
%%sql
SELECT mission_outcome, COUNT(*) AS no_outcome
FROM SPACEXDATASET
GROUP BY mission_outcome;
```

```
 * ibm_db_sa://ftb12020:***@0c77d6f2-5da9-48a9-
Done.
```

mission_outcome	no_outcome
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Boosters Carried Maximum Payload

The names of the booster which have carried the maximum payload mass

```
%%sql
SELECT booster_version, PAYLOAD_MASS__KG_
FROM SPACEXDATASET
WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXDATASET);
```

```
 * ibm_db_sa://ftb12020:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1
Done.
```

booster_version	payload_mass__kg_
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

2015 Launch Records

The failed Landing Outcomes in drone ship, their booster versions, and launch site names for in year 2015

```
%%sql
SELECT MONTHNAME(DATE) AS MONTH, landing__outcome, booster_version, PAYLOAD_MASS__KG_, launch_site
FROM SPACEXDATASET
WHERE landing__outcome = 'Failure (drone ship)' AND YEAR(DATE) = 2015;
```

```
 * ibm_db_sa://ftb12020:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.app
Done.
```

MONTH	landing__outcome	booster_version	payload_mass__kg_	launch_site
January	Failure (drone ship)	F9 v1.1 B1012	2395	CCAFS LC-40
April	Failure (drone ship)	F9 v1.1 B1015	1898	CCAFS LC-40

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

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

```
%%sql
SELECT landing__outcome, COUNT(*) AS no_outcome
FROM SPACEXDATASET
WHERE landing__outcome LIKE 'Succes%' AND DATE BETWEEN '2010-06-04' AND '2017-03-20'
GROUP BY landing__outcome
ORDER BY no_outcome DESC;
```

```
 * ibm_db_sa://ftb12020:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg
Done.
```

landing__outcome	no_outcome
Success (drone ship)	5
Success (ground pad)	3

Section 3

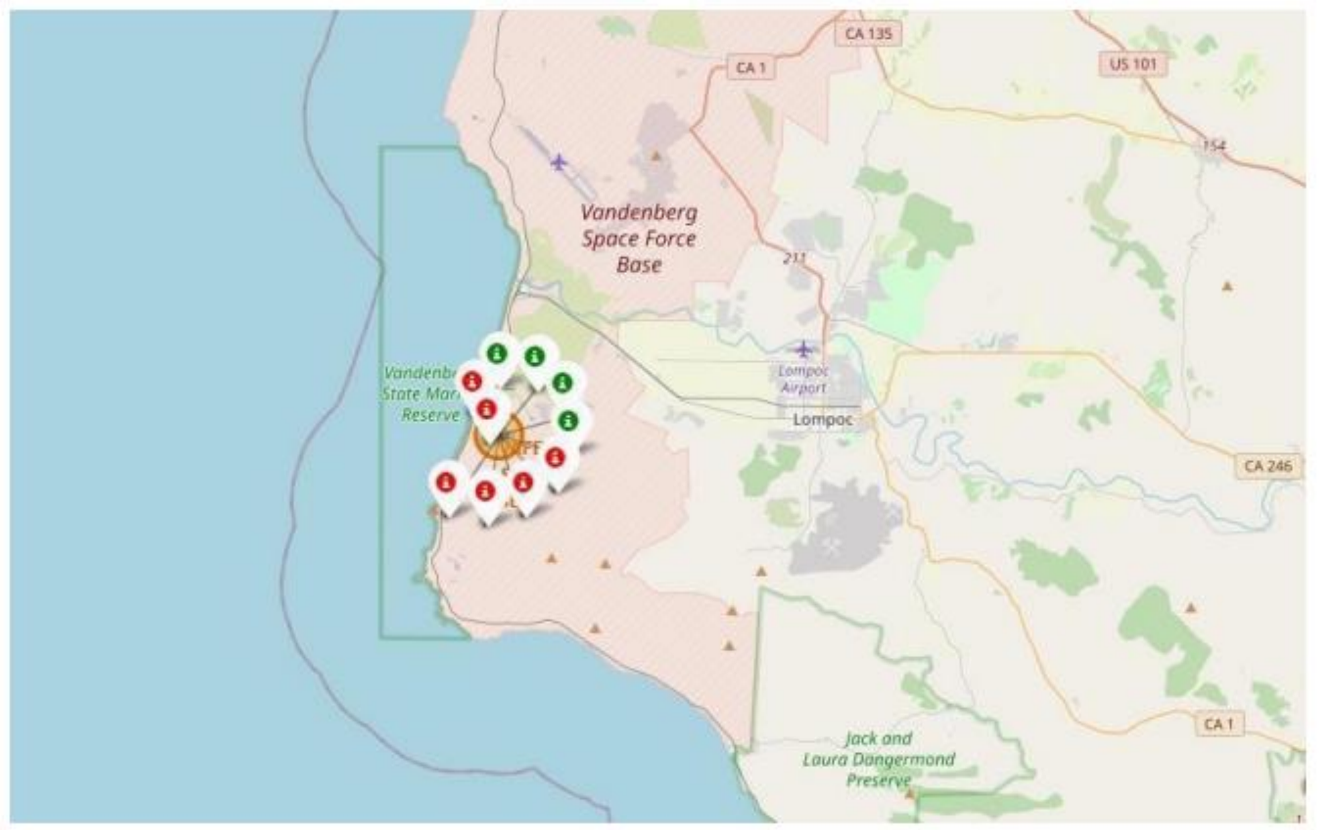
Launch Sites Proximities Analysis

Launch Site Locations

The map on the left shows where the launch sites locations from this it is clear that all the sites are located near the coastline and or near an ocean

Landing Markers

The markers show were the rocket landed and if it was a successful land (green) or if it was a failure (red).



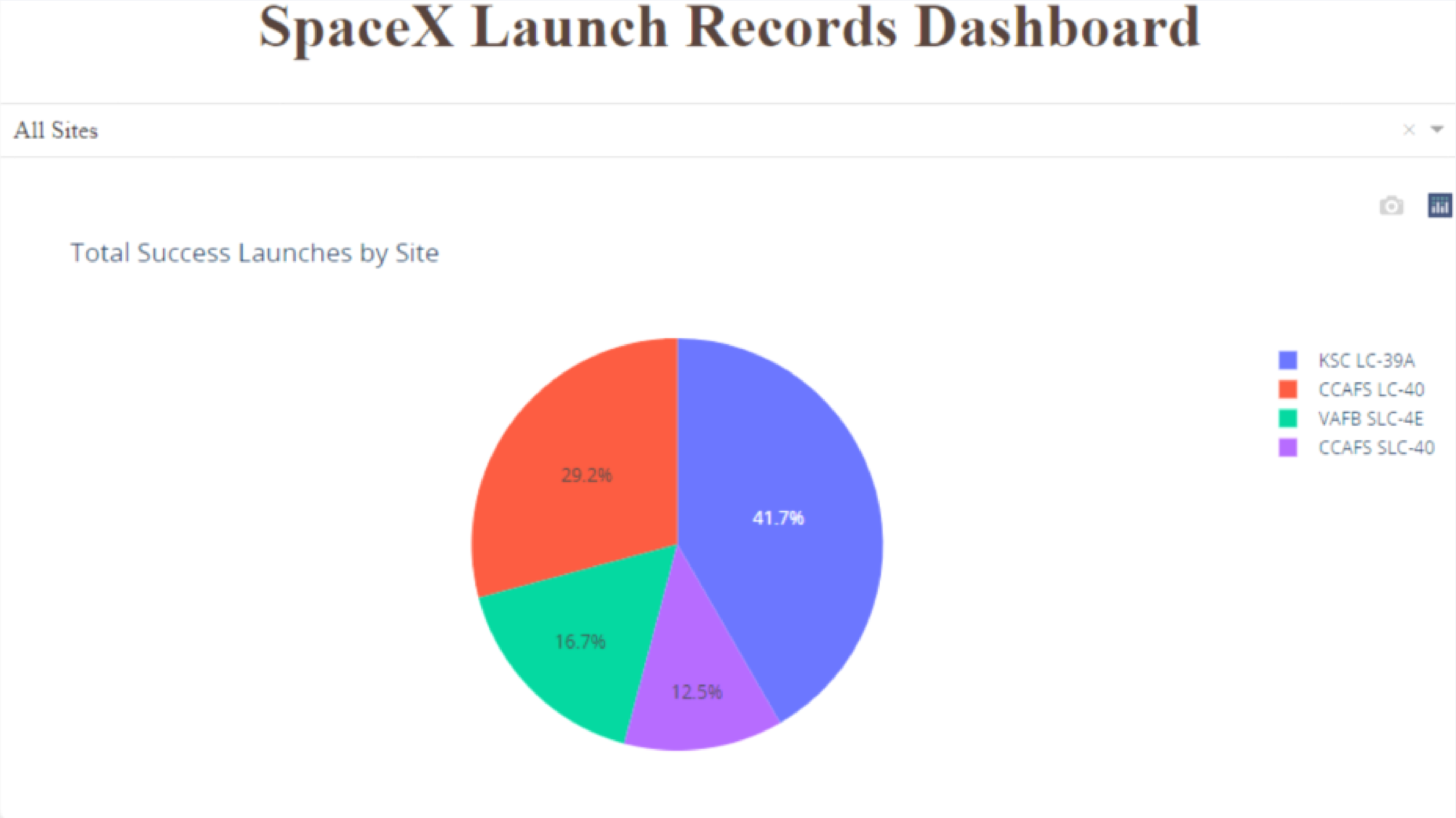
Key Location Proximity

Showing the distance, 14.805km, between KSC LC-39A from the city, coast and highway.

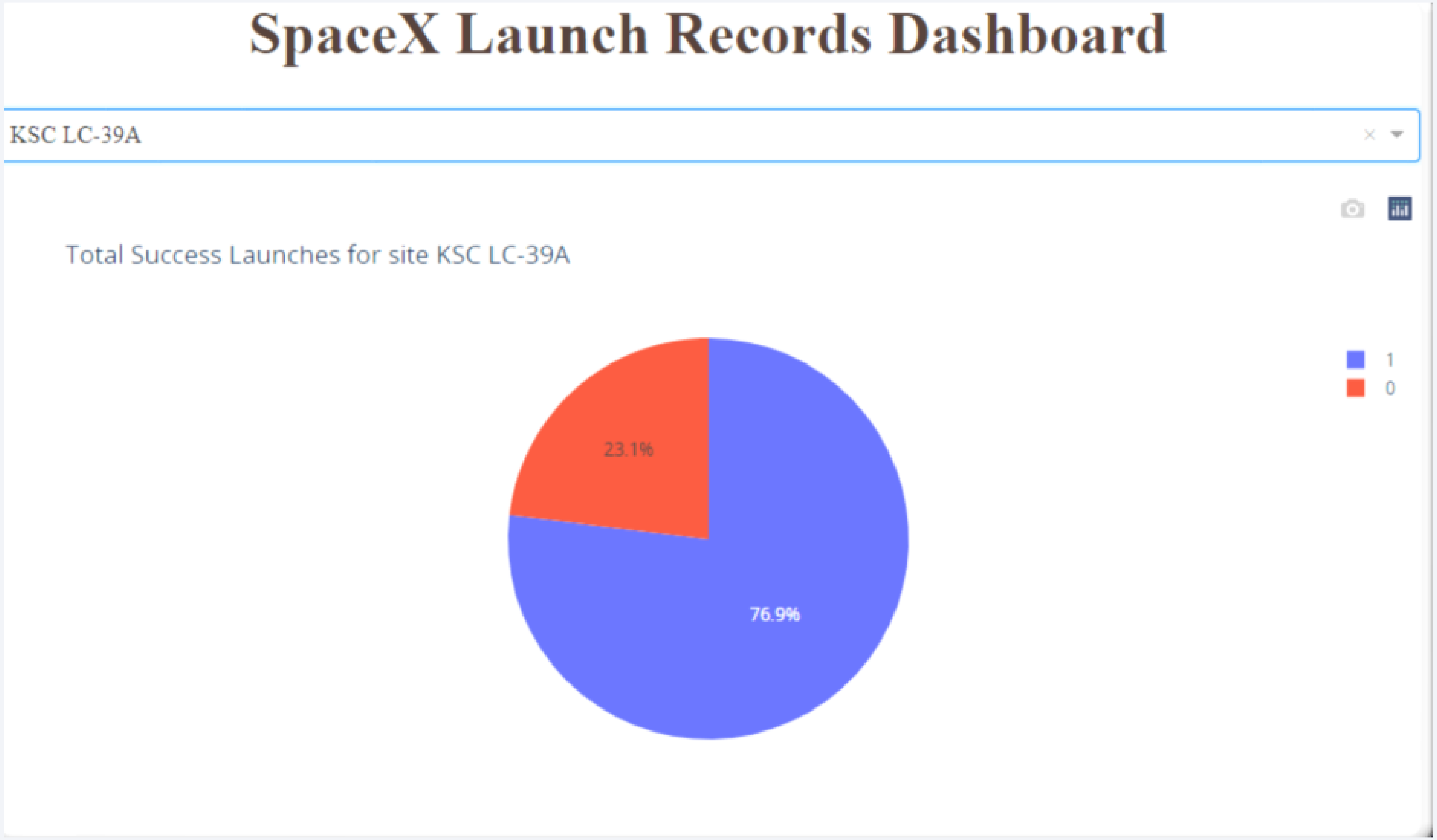
Section 4

Build a Dashboard with Plotly Dash

Success from Each Site



Success vs Failure Rate for Site KSC LC-39A



Payload Mass vs Success vs Booster Version

Section 5

Predictive Analysis (Classification)

Classification Accuracy

Logistic Regression, Support Vector Machine, K nearest Neighbor had the same accuracy, 0.83333, while the Decision Tree had the worst accuracy.

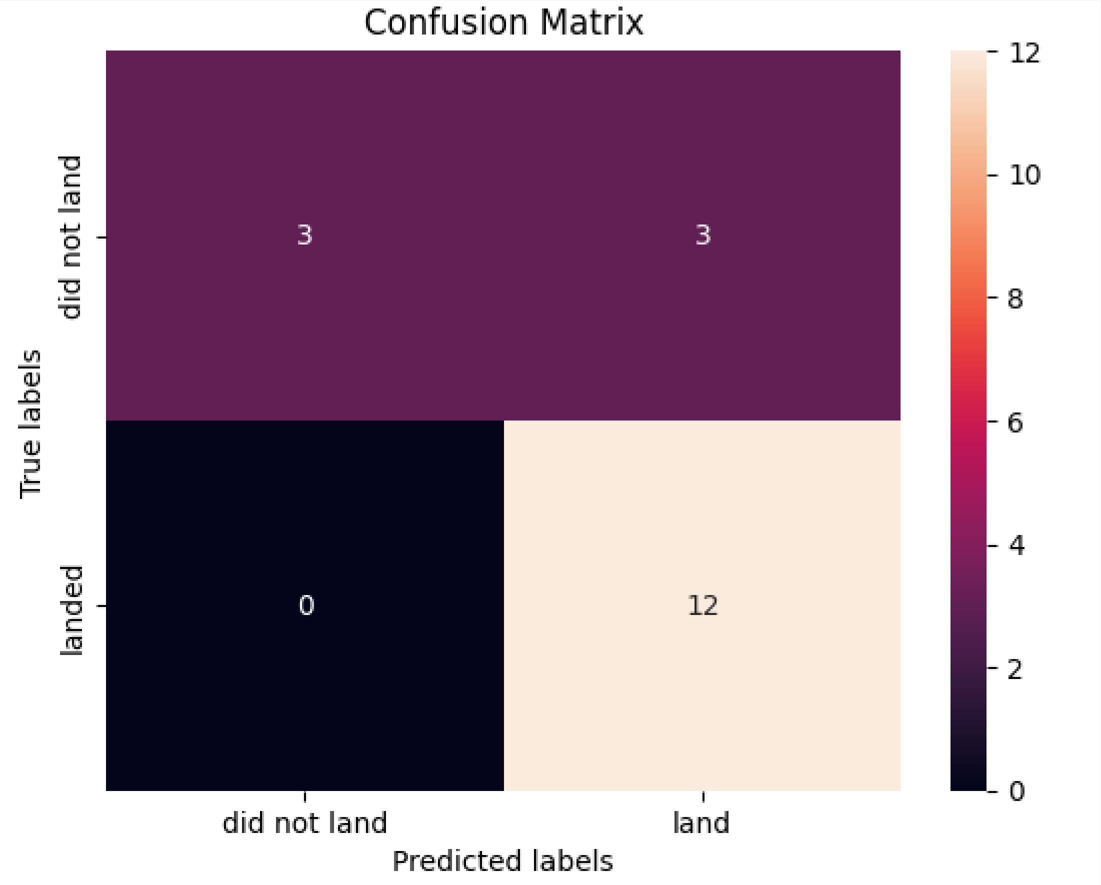
	Algorithm	Accuracy Score
0	Logistic Regression	0.833333
1	Support Vector Machine	0.833333
2	Decision Tree	0.666667
3	K Nearest Neighbours	0.833333

Notes:

- The test data was small, only a sample size of 18
- Having a small sample size can lead to a big variation in accuracy results. Which could explain why the Decision Tree had such a low accuracy compared to the others.
- A recommendation would be to have more data to determine which model is the best at predicting the success or fail of a Falcon 9 rocket landing.

Confusion Matrix

- Since most of the models accuracy were the same, the confusion matrix is also the same across them.
- Looking at the off diagonal, the models classified 15 data point correctly and misclassified 3
- Since all 3 data points misclassified as land when they were their true label was did not land, it implies that our model over predicts successful landings.



Conclusions

- As the number of flights increase so does the success rate. As in the beginning between 2010 and 2013 of a success rate of 0 to after 2016 when there was a greater chance of a 50% success rate.
- The Orbit Type SSO has an impressive success rate of 100% with their 5 successful flights. Orbit types PO, ISS, and LEO have more success with heavier flights, greater payload masses.
- The launch site KSC LC-39A had the most successful launches, with a 76.88% success rate.
- A model couldn’t be chosen as best performing; however Logistic Regression, Support Vector , and Kth nearest neighbor had the same accuracy rates of 0.8333%

Further Steps:

- More data to test the models further to determine which is the best.
- Further refine the models with the data and feedback received.

Appendix

GitHub Link

https://github.com/Sku002/IBM-Data-Science-Capstone-SpaceX/tree/main

Thank you!