IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Mitchell Kelly
08/16/23

Table of Contents

Executive Summary

Space X Falcon 9 launches save millions of dollars reusing their first rocket stage instead of it going to waste by destruction. With historical Falcon 9 flight data it is possible to determine through the data alone whether the first stage will land successfully or not.

The flight data was collected entirely through a SpaceX API and web scrapping flight history data from Wikipedia. Once the required information was pulled from these sources, the data was processed by standardizing a binary class system which was either successful or unsuccessful for landing outcome. Several placeholder variables were used to allow for a more streamline statistical and machine learning model analysis after exploring the data for hidden insights.

It was determined through SQL querying, graphing in Python, and Dashboarding that Payload Mass and Flight number were the most significant indicators of landing outcome. It was not possible to observe a favorable outcome trend with data visualization alone, so several machine learning models were trained with optimal hyperparameters. KNN was observed to be the best model in terms of training time and accuracy score (0.77). A more expansive database of flight history is needed to improve the accuracy of the model.

Introduction

- SpaceX Falcon 9 launches cost 62 million dollars compared to 165 million of its competitors
- Reusing the first stage of the rocket launch accounts for most of the price savings
- Can we determine if the first stage will land using data driven insights and machine learning?
- Is it possible to find a sufficiently accurate model that can determine if the first stage will land based upon historical flight data?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Describe how data was collected
- Perform data wrangling
  - Describe how data was processed
- 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

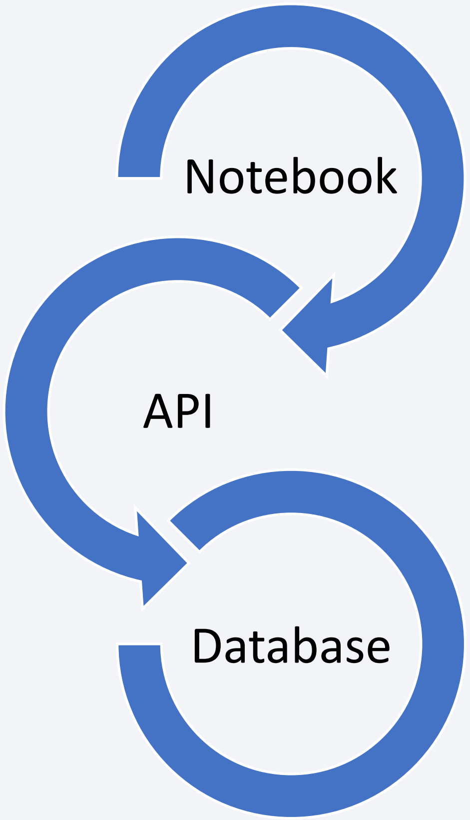
Historical flight data was acquired using:

1) SpaceX API: https://docs.spacexdata.com/

- Get requests with python functions to extract pertinent data

2) Wikipedia Web scrapping: https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches

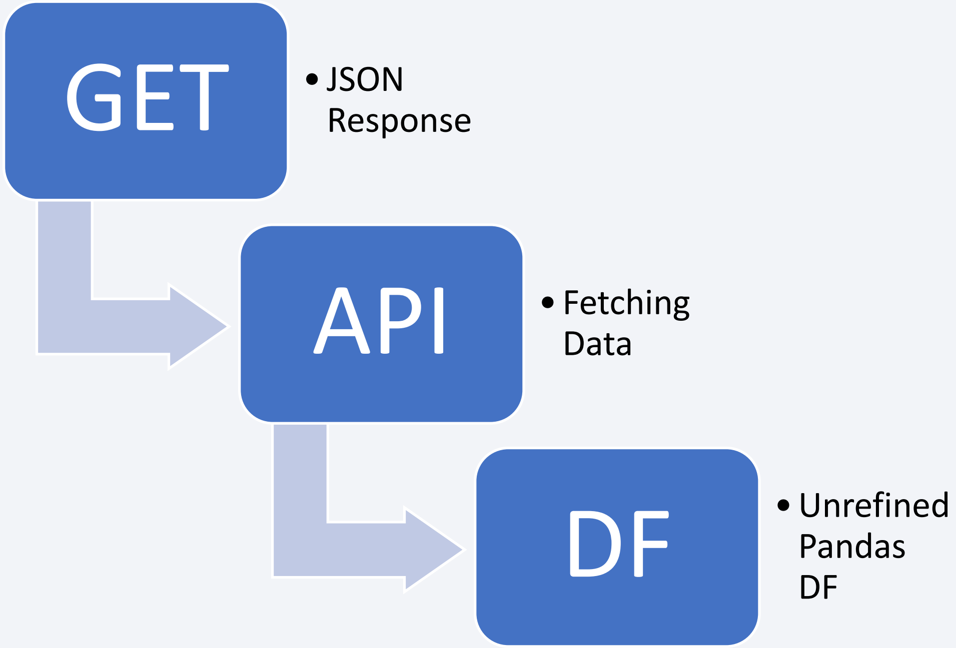
- BeautifulSoup library data extraction techniques



Data Collection – SpaceX API

1. Requesting and Parsing through SpaceX launch data with GET request
2. Filtering Pandas Dataframe (DF) for only Falcon 9 launces

- https://github.com/mkelly44/Applied-Data-Science-Capstone/blob/main/jupyter-labs-spacex-data-collection-api.ipynb
- https://github.com/mkelly44/Applied-Data-Science-Capstone/blob/main/dataset_part_1.csv



Data Collection - Scraping

1. Request call from the Falcon 9 Launch Wiki Page
2. Extracting all column names from HTML table header
3. Creating a new DF by iterating through launch HTML tables

- https://github.com/mkelly44/Applied-Data-Science-Capstone/blob/main/jupyter-labs-webscraping.ipynb
- https://github.com/mkelly44/Applied-Data-Science-Capstone/blob/main/spacex_web_scraped.csv

2020 [edit]

In late 2019, Gwynne Shotwell stated that SpaceX hoped for as many as 24 launches for Starlink satellites in 2020,[490] in addition to 14 or 15 non-Starlink launches. At 26 launches, 13 of which for Starlink satellites, Falcon 9 had its most prolific year, and Falcon rockets were second most prolific rocket family of 2020, only behind China's Long March rocket family.[491]

[hide] Flight No.	Date and time (UTC)	Version, Booster[1]	Launch site	Payload[c]	Payload mass	Orbit	Customer	Launch outcome	Booster landing
78	7 January 2020, 02:19:21[492]	F9 B5 △ B1049.4	CCAFS, SLC-40	Starlink 2 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Success (drone ship)
	Third large batch and second operational flight of Starlink constellation. One of the 60 satellites included a test coating to make the satellite less reflective, and thus less likely to interfere with ground-based astronomical observations.[493]								
79	19 January 2020, 15:30[494]	F9 B5 △ B1046.4	KSC, LC-39A	Crew Dragon in-flight abort test[495] (Dragon C205.1)	12,050 kg (26,570 lb)	Sub-orbital[496]	NASA (CTS)[497]	Success	No attempt
	An atmospheric test of the Dragon 2 abort system after Max Q. The capsule fired its SuperDraco engines, reached an apogee of 40 km (25 mi), deployed parachutes after reentry, and splashed down in the ocean 31 km (19 mi) downrange from the launch site. The test was previously slated to be accomplished with the Crew Dragon Demo-1 capsule,[498] but that test article exploded during a ground test of SuperDraco engines on 20 April 2019.[419] The abort test used the capsule originally intended for the first crewed flight.[499] As expected, the booster was destroyed by aerodynamic forces after the capsule aborted.[500] First flight of a Falcon 9 with only one functional stage — the second stage had a mass simulator in place of its engine.								
80	29 January 2020, 14:07[501]	F9 B5 △ B1051.3	CCAFS, SLC-40	Starlink 3 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Success (drone ship)
	Third operational and fourth large batch of Starlink satellites, deployed in a circular 290 km (180 mi) orbit. One of the fairing halves was caught, while the other was fished out of the ocean.[502]								
81	17 February 2020, 15:05[503]	F9 B5 △ B1056.4	CCAFS, SLC-40	Starlink 4 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Failure (drone ship)
	Fourth operational and fifth large batch of Starlink satellites. Used a new flight profile which deployed into a 212 km × 386 km (132 mi × 240 mi) elliptical orbit instead of launching into a circular orbit and firing the second stage engine twice. The first stage booster failed to land on the drone ship[504] due to incorrect wind data.[505] This was the first time a flight proven booster failed to land.								
82	7 March 2020, 04:50[506]	F9 B5 △ B1059.2	CCAFS, SLC-40	SpaceX CRS-20 (Dragon C112.3 △)	1,977 kg (4,359 lb)[507]	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
	Last launch of phase 1 of the CRS contract. Carries *Bartolomeo*, an ESA platform for hosting external payloads onto ISS.[508] Originally scheduled to launch on 2 March 2020, the launch date was pushed back due to a second stage engine failure. SpaceX decided to swap out the second stage instead of replacing the faulty part.[509] It was SpaceX's 50th successful landing of a first stage booster, the third flight of the Dragon C112 and the last launch of the cargo Dragon spacecraft.								
83	18 March 2020, 12:16[510]	F9 B5 △ B1048.5	KSC, LC-39A	Starlink 5 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Failure (drone ship)
	Fifth operational launch of Starlink satellites. It was the first time a first stage booster flew for a fifth time and the second time the fairings were reused (Starlink flight in May 2019).[511] Towards the end of the first stage burn, the booster suffered premature shut down of an engine, the first of a Merlin 1D variant and first since the CRS-1 mission in October 2012. However, the payload still reached the targeted orbit.[512] This was the second Starlink launch booster landing failure in a row, later revealed to be caused by residual cleaning fluid trapped inside a sensor.[513]								
84	22 April 2020, 19:30[514]	F9 B5 △ B1051.4	KSC, LC-39A	Starlink 6 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Success (drone ship)

Out[39]:

	Flight No.	Launch site	Payload	Payload mass	Orbit	Customer	Launch outcome	Version Booster	Booster landing	Date	Time
0	1	CCAFS	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success\n	F9 v1.0B0003.1	Failure	4 June 2010	18:45
1	1	CCAFS	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success\n	F9 v1.0B0003.1	Failure	4 June 2010	18:45
2	1	CCAFS	Dragon	0	LEO	NASA	Success	F9 v1.0B0003.1	No attempt\n	4 June 2010	18:45
3	1	CCAFS	Dragon	525 kg	LEO	NASA	Success	F9 v1.0B0004.1	No attempt	4 June 2010	18:45
4	2	CCAFS	SpaceX CRS-1	4,700 kg	LEO	NASA	Success\n	F9 v1.0B0005.1	No attempt\n	8 December 2010	15:43
...	...	...	...	...	...	...	...	...	...	...	...
225	117	CCSFS	Starlink	15,600 kg	LEO	NASA (CRS)	Success\n	F9 B5B1058.8	Success	9 May 2021	06:42
226	118	KSC	SpaceX CRS-22	3,328 kg	LEO	Sirius XM	Success\n	F9 B5B1063.2	NaN	15 May 2021	22:56
227	119	CCSFS	SXM-8	7,000 kg	GTO	NaN	NaN	F9 B5B1067.1	NaN	26 May 2021	18:59
228	120	NaN	NaN	NaN	NaN	NaN	NaN	F9 B5	NaN	3 June 2021	17:29
229	121	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	6 June 2021	04:26

Data Wrangling

1. Dealing with Missing Values
2. Calculating the number of launches on each site
3. Determining the frequency of each specific orbit track
   - LEO, VLEO, GTO, SSO, ES-L1, HEO, ISS, MEO, HEO, GEO, and PO
4. Calculating the occurrence and amount of mission outcomes for each orbit
5. Creating a labeled outcome list based upon mission outcomes
6. Features Engineering with placeholder variables

- https://github.com/mkelly44/Applied-Data-Science-Capstone/blob/main/IBM-DS0321EN-SkillsNetwork_labs_module_1_L3_labs-jupyter-spacex-data_wrangling_jupyterlite.jupyterlite.ipynb

```
In [21]:
# landing_class = 0 if bad_outcome
# landing_class = 1 otherwise

landing_class = []
for outcome in df['Outcome']:
    if outcome == ('False ASDS' or 'False Ocean' or 'False RTLS' or 'None ASDS' or 'None None'):
        landing_class.append(0)
    else:
        landing_class.append(1)
```

```
In [22]:
print(landing_class)

[1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 0, 1, 1, 1, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 0, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1]
```

Out[24]:

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block
0	1	2010-06-04	Falcon 9	6104.959412	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0
1	2	2012-05-22	Falcon 9	525.000000	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0
2	3	2013-03-01	Falcon 9	677.000000	ISS	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0
3	4	2013-09-29	Falcon 9	500.000000	PO	VAFB SLC 4E	False Ocean	1	False	False	False	NaN	1.0
4	5	2013-12-03	Falcon 9	3170.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0

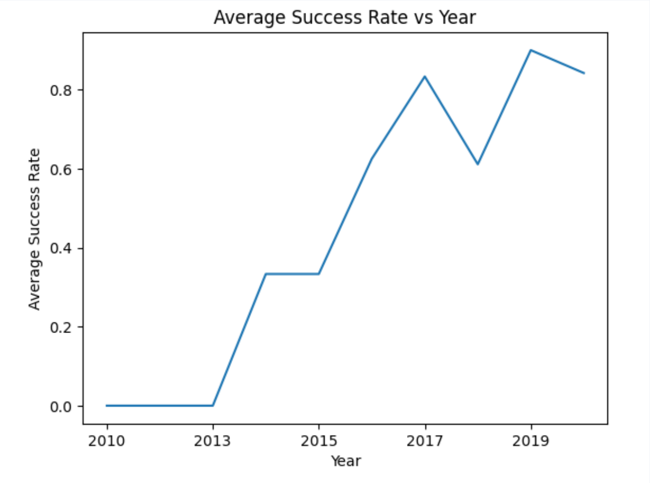
EDA with Data Visualization

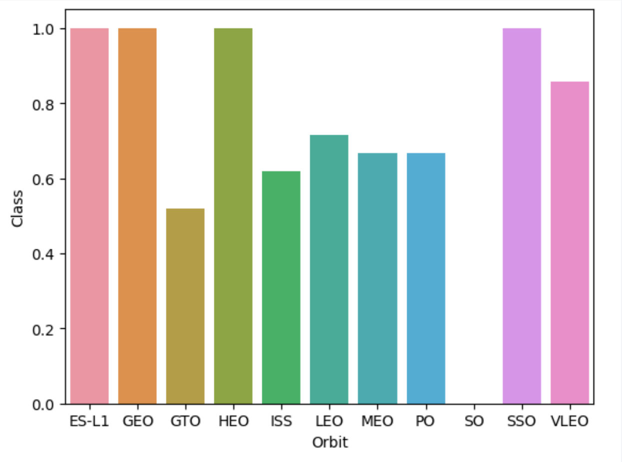
- Each Chart plots two features against each other to determine their correlation and if a specific trend can be observed

- An additional element of Successful Mission (1) or Unsuccessful Mission (0) is present so the correlations can be related to outcomes

1. Categorical Plot of Flight Number and Payload Mass
2. Categorical Plot of Flight Number and Launch Site
3. Scatter Plot of Payload Mass and Launch Site
4. Bar Plot of Orbit Type and Mean Success Rate
5. Scatter Plot of Flight Number and Orbit Type
6. Scatter Plot of Payload Mass and Orbit Type
7. Line Plot of Average Success Rate over Years

- https://github.com/mkelly44/Applied-Data-Science-Capstone/blob/main/IBM-DS0321EN-SkillsNetwork_labs_module_2_jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb





EDA with SQL

- SQL Queries seek to display pertinent exploratory data analysis findings in an easily digestible manner

1. Displaying Names of Unique Launch Sites
2. Selecting 5 Launch Sites Beginning with CCA
3. Displaying Total Payload Mass from Rockets Launched by NASA
4. Displaying Average Payload Mass from booster version F9 v1.1
5. Listing Date of First Successful Ground Pad Landing
6. Finding Booster Names with Payload Mass greater than 4000, but less than 6000
7. Calculating Total Successful and Failure Mission Outcomes
8. Finding Booster Names with Maximum Payload Mass
9. Displaying Failure Landing Outcomes in 2015
10. Calculating Frequency of Landing Outcome Type

- https://github.com/mkelly44/Applied-Data-Science-Capstone/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

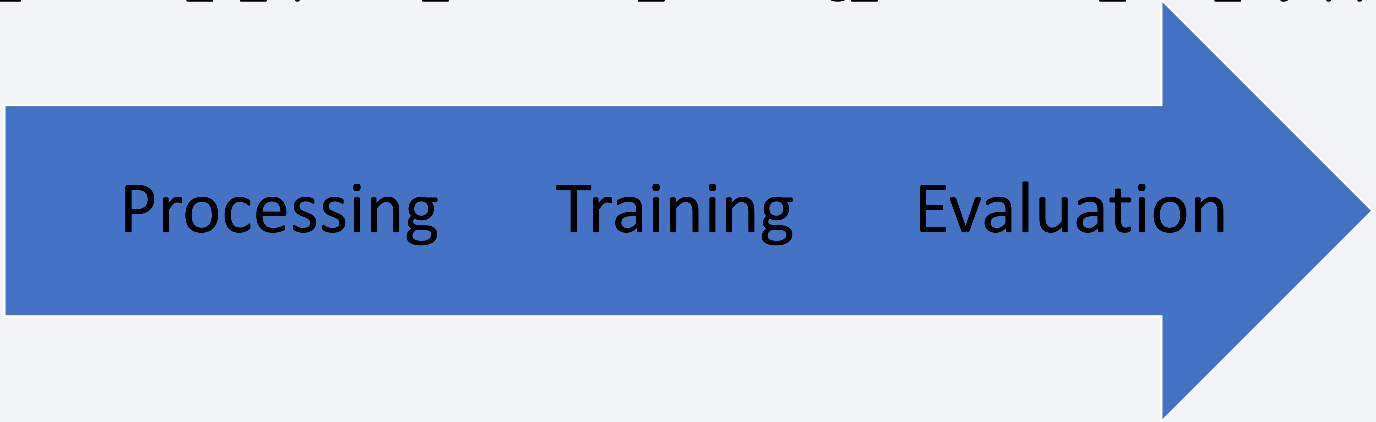
Build a Dashboard with Plotly Dash

1. Created Pie Chart Showing Successful Launches for All or Specific Launch Sites
    - Can effectively see what launch site has the most successful landing outcomes
2. Showcased the Correlation Between Payload Mass and Outcome for All or Specific Booster Versions with a Category Plot
    - Can see what booster type correlates the most with successful landing outcomes, and a payload mass slider was added to refine results further

- https://github.com/mkelly44/Applied-Data-Science-Capstone/blob/main/spacex_dash_app%20(1).py

Predictive Analysis (Classification)

- Segmented columns of Dataframe into target variable Y and normalized features set X
- Split X and Y into train and test data portions
- Used GridSearchCV from sklearn with several parameters to optimize results for each method employed: *Logistic Regression, SVC, Decision Tree, and KNN*
- Each machine learning method was fit with best hyperparameters to maximize the accuracy of prediction
- SVM was chosen as the best performing model based upon accuracy score and minimal training time
- https://github.com/mkelly44/Applied-Data-Science-Capstone/blob/main/IBM-DS0321EN-SkillsNetwork_labs_module_4_SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb



Results

Exploratory Data Analysis Results:

- ES-L1, GEO, HEO, and SSO orbits have had successful first stage landing for every mission
- CCAFS LC-40 has a 60% success rate while both KSC LC-39A and VAFB SLC 4E have a success rate of 77%
- LEO orbit had 100% success rate after less than 10 flights up to 40+
- From 2013 to 2019 average success rate has increased roughly 10% each year

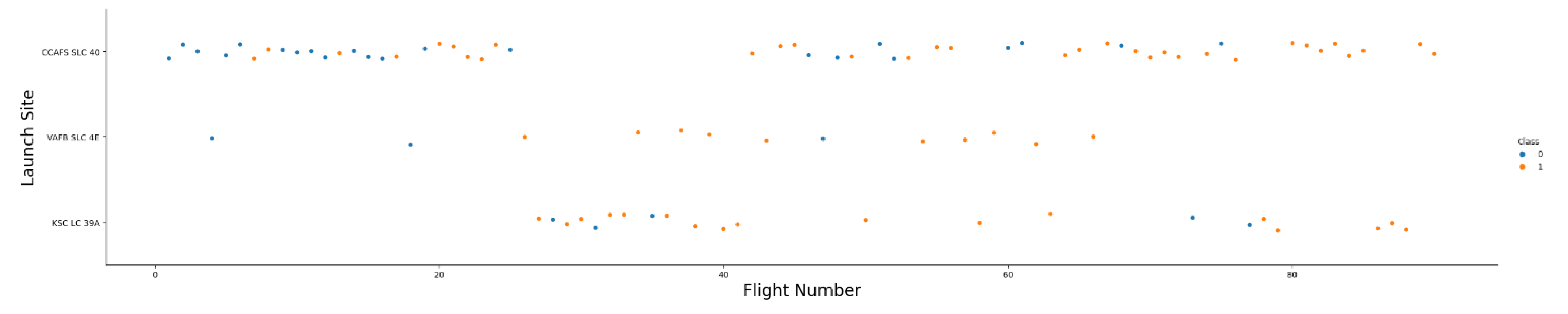
- Predictive Analysis Results:
  - KNN and SVM had an accuracy score of 0.77 each
  - Logistic Regression and Decision Tree has an accuracy score of 0.72 each

Section 2

Insights drawn from EDA

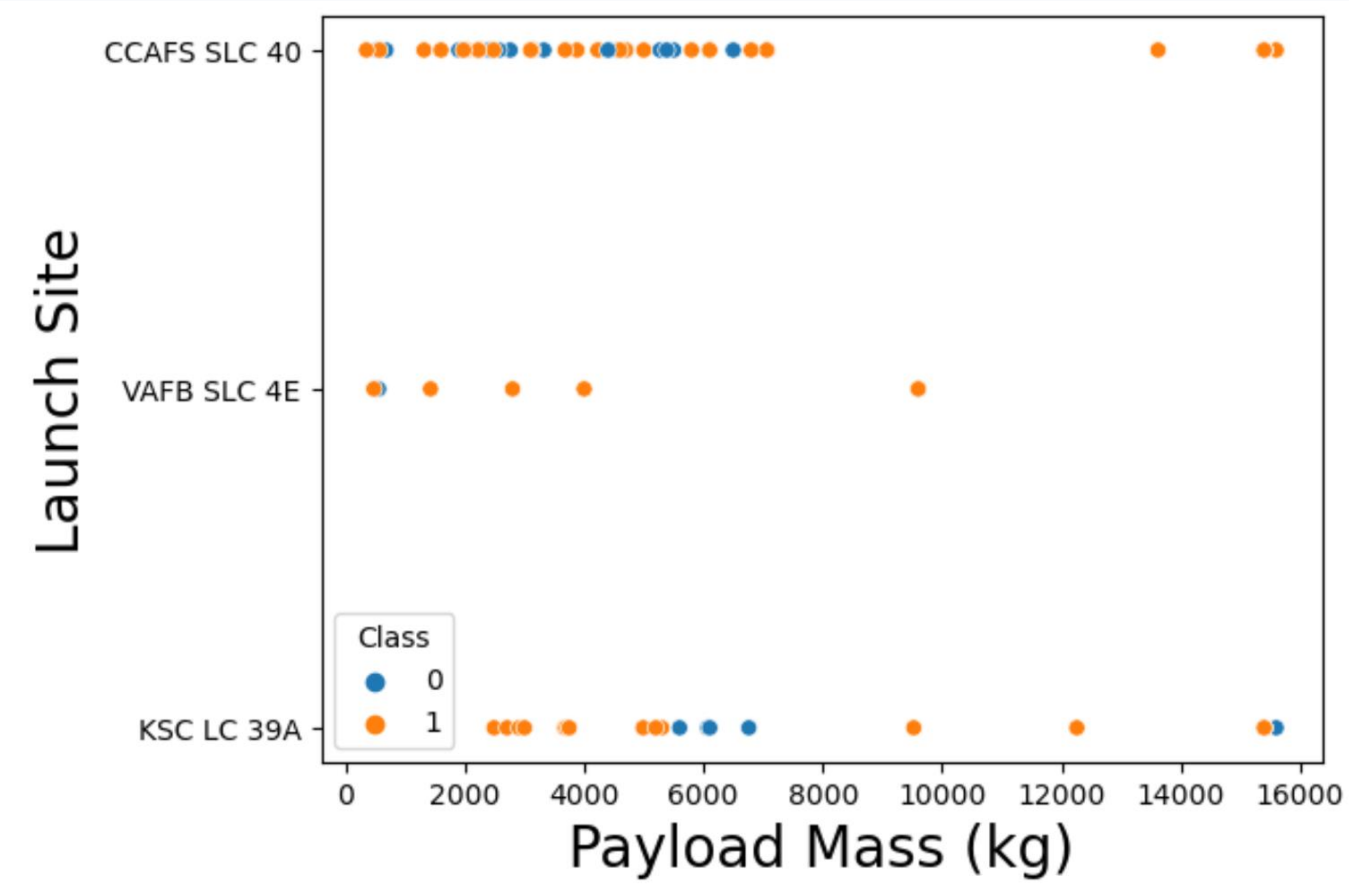
Flight Number vs. Launch Site

- After Flight Number 80, every launch site takeoff has a successful landing outcome.
- CCAFS SLC-40's unsuccessful launches are heavily distributed towards lower flight number



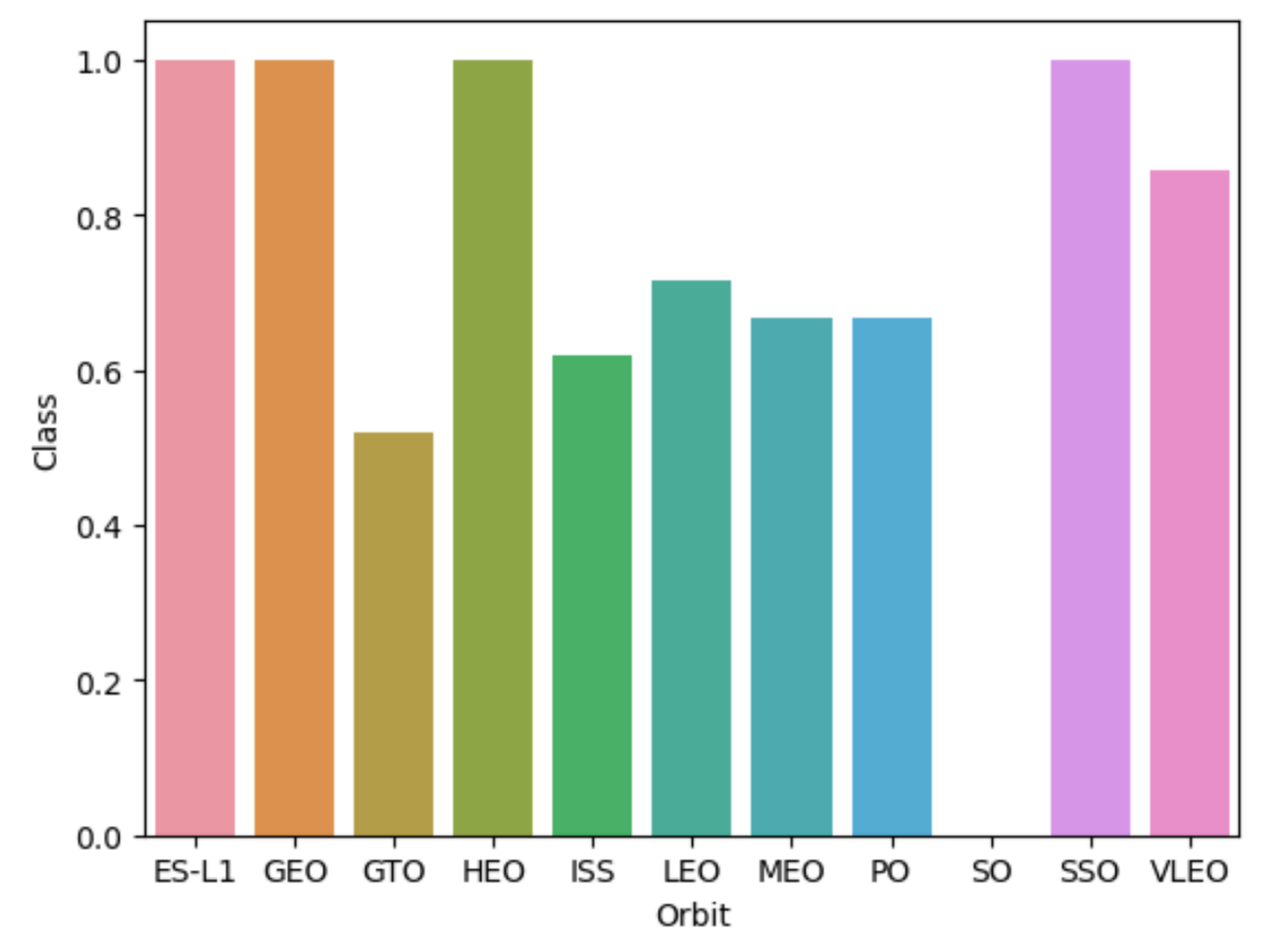
Payload vs. Launch Site

- For VAFB SLC 4E there are no payload sizes greater than 10,000 kg.
- Large payload size (12,000+ kg) for CCAFS SLC-40 has 100% success rate, yet most of its launches occurred in the less than 8,000 kg range.



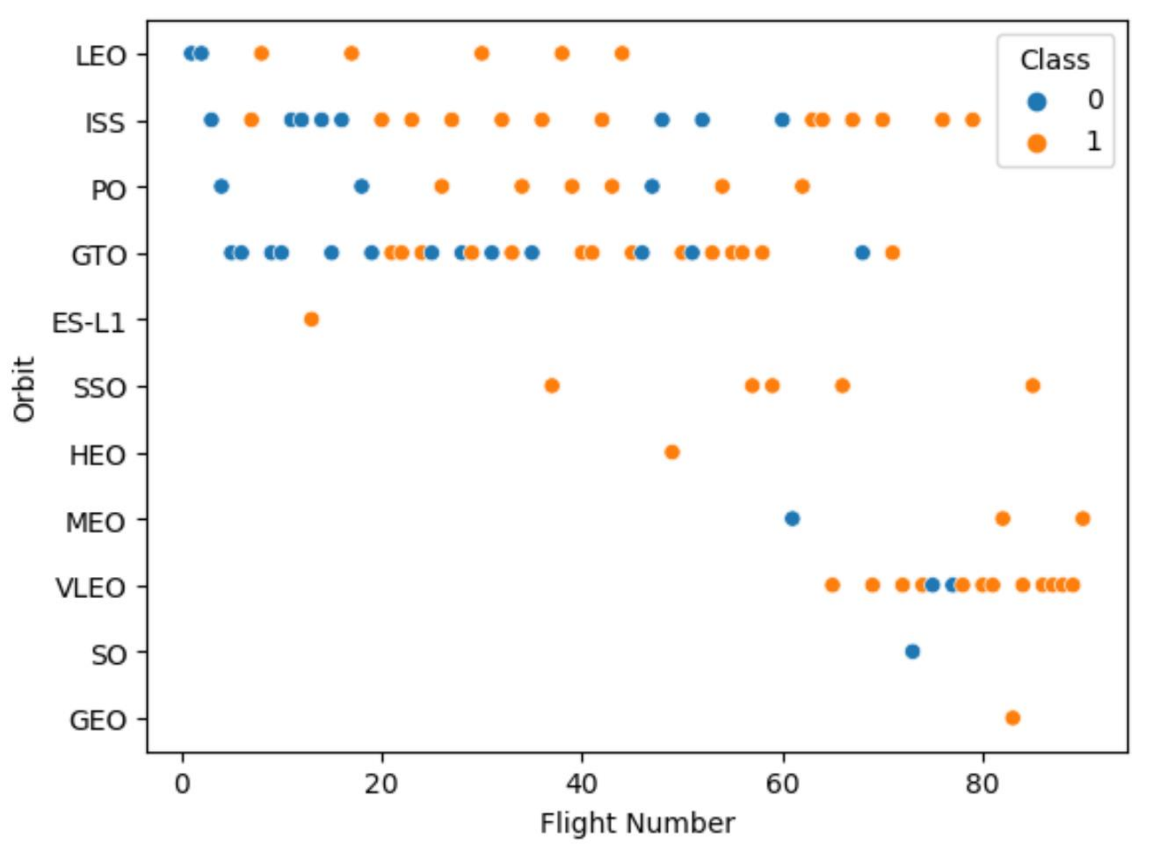
Success Rate vs. Orbit Type

- GTO orbits have an average success rate of 0.5
- ES-L1, GEO, HEO, and SSO orbits have a success rate of 1.0
- SO orbit launches have never had a success.



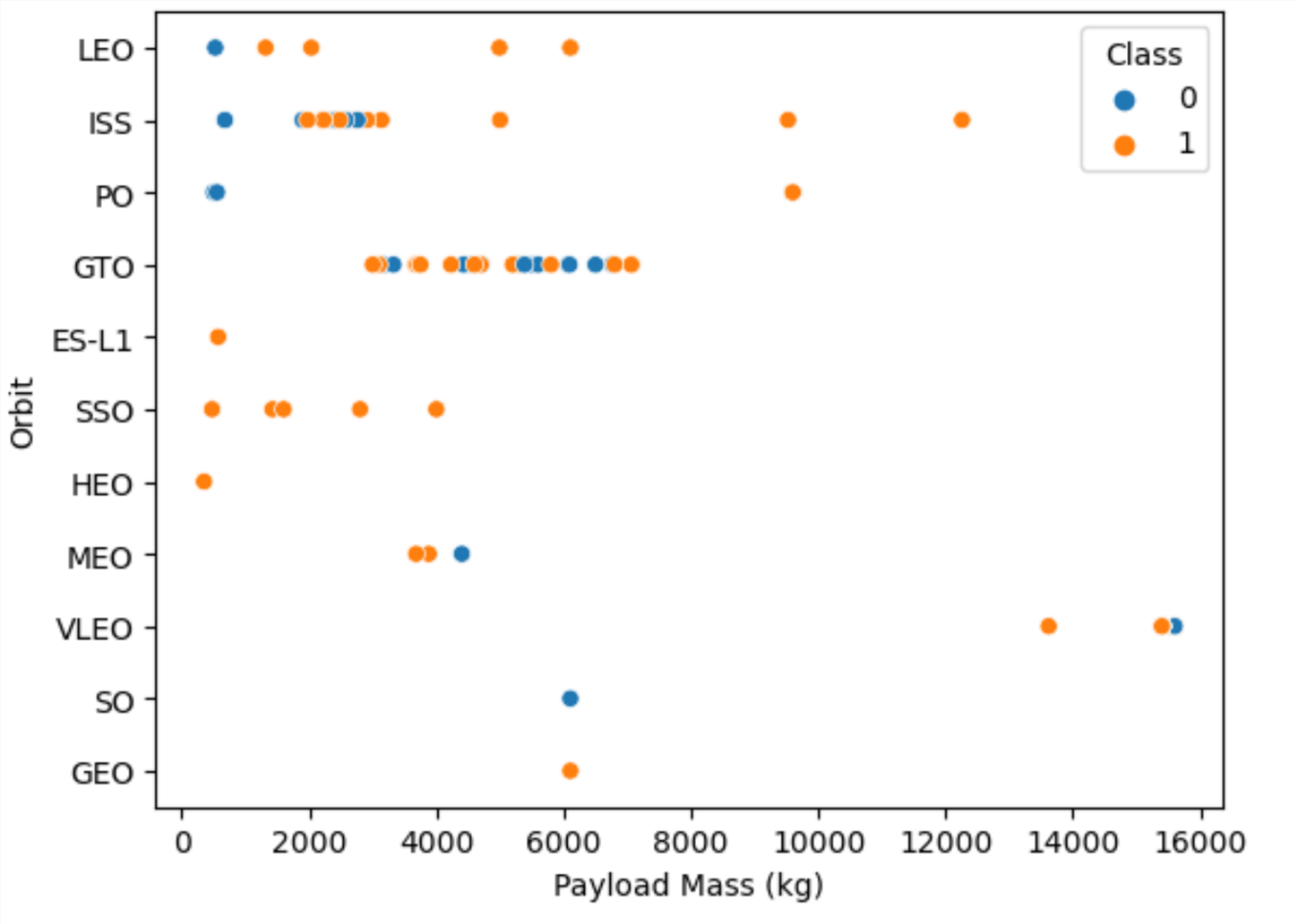
Flight Number vs. Orbit Type

- LEO orbit success rate depends entirely on increasing Flight Number.
- GTO orbits have the most flights, yet the relationship between success rate and flight number is inconclusive.



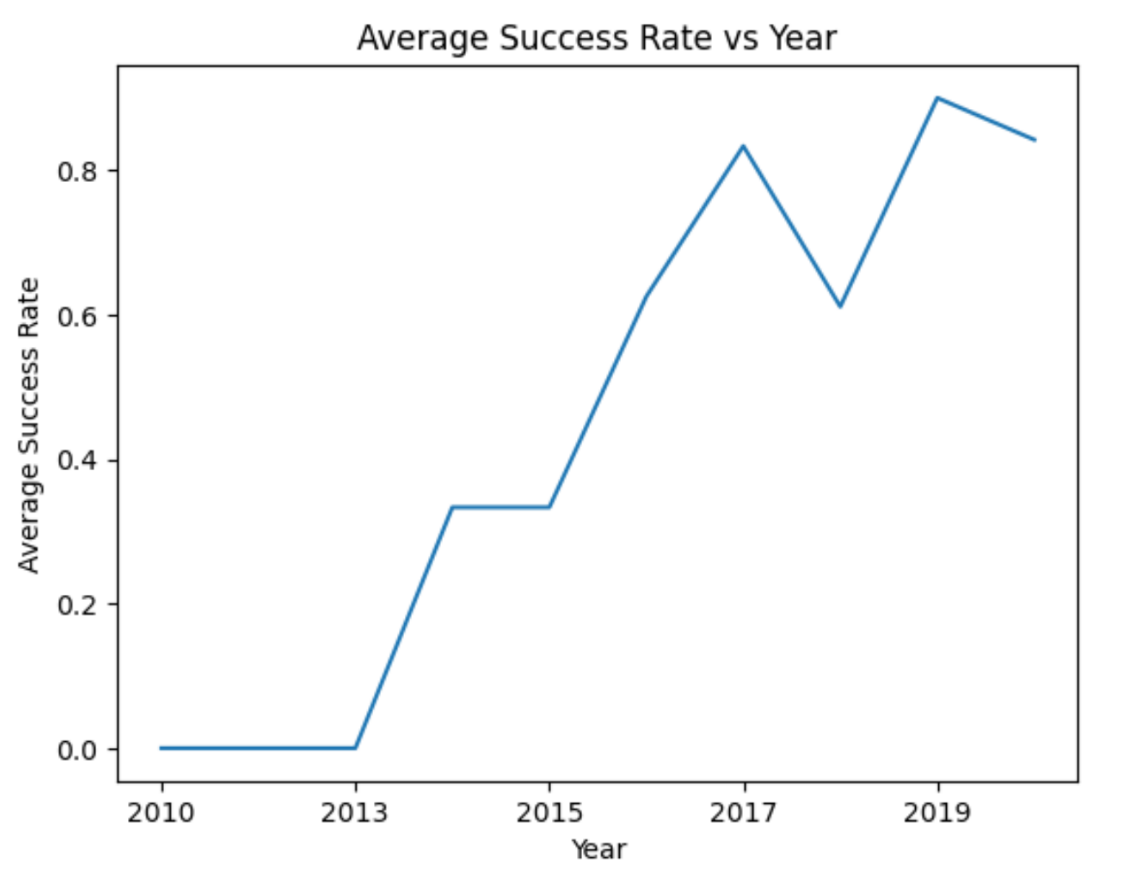
Payload vs. Orbit Type

- For heavier payloads in PO, LEO, and ISS orbit, the success rate is higher.
- GTO orbits have an inclusive relationship with payload mass.



Launch Success Yearly Trend

- From 2013 to 2020, the average success rate increases around 0.1 per year.
- A dip in average success rate was exhibited in 2018, but it did not affect the long term outlook.



All Launch Site Names

- All the unique launch site names were needed to conceptually simplify where rockets would launch from.

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

In [8]:
```
%%sql
SELECT DISTINCT Launch_Site FROM SPACEXTABLE
```

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

Out[8]:

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

Launch Site Names Begin with 'CCA'

- Query was completed to get a general picture of CCAFS LC-40 launch site data.

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

In [9]:
```
%%sql
SELECT * FROM SPACEXTABLE WHERE Launch_Site LIKE 'CCA%' LIMIT 5
```

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

Out[9]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mi
2010-04-06	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	
2010-08-12	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0	LEO (ISS)	NASA (COTS) NRO	
2012-05-22	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	
2012-08-10	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	
2013-01-03	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	

Total Payload Mass

- Total NASA customer payload was queried to compare with other customers

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

```
%%sql
SELECT SUM(PAYLOAD_MASS__KG_) AS 'NASA (CRS) Total Payload Mass' FROM SPACEXTABLE WHERE Customer = 'NASA (CRS)'
```

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

NASA (CRS) Total Payload Mass
45596

Average Payload Mass by F9 v1.1

- Average payload mass from booster version F9 v1.1 rockets was queried for data analysis.

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

```
%%sql
SELECT AVG(PAYLOAD_MASS__KG_) AS 'Average Payload Mass of F9 v1.1' FROM SPACEXTABLE WHERE Booster_Version = 'F9 v1.1'
```

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

Average Payload Mass of F9 v1.1
2928.4

First Successful Ground Landing Date

- The first successful ground landing date was queried to signifiy the start of when the first stage of Falcon9 rockets began to land.

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

Hint:Use min function

```
%%sql
SELECT MIN(Date) AS 'Date of First Successful Ground Pad Landing' FROM SPACEXTABLE WHERE Landing_Outcome = 'Success (ground pad)'
```

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

Date of First Successful Ground Pad Landing
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- Query of successful drone ship landing with payload between 4000 and 6000 shows what are the popular booster versions for successful landing outcomes.

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

```
%%sql
SELECT Booster_Version FROM SPACEXTABLE WHERE (Landing_Outcome = 'Success (drone ship)') AND (PAYLOAD_MASS__KG_ > 4000 AND PAYLOAD_MASS__KG_ < 6000)
```

```
 * sqlite:///my_data1.db
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 failed mission outcomes query was needed to see how many individual instances there were.

List the total number of successful and failure mission outcomes

```
%%sql
SELECT COUNT(*) AS 'Total Number of Successes and Failures' FROM SPACEXTABLE WHERE Mission_Outcome LIKE 'Succ%' OR Mission_Outcome LIKE 'Fail%'
```

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

Total Number of Successes and Failures
101

Boosters Carried Maximum Payload

- A query containing the booster versions of a maximum payload launch helps differentiate between the individual launches better.

List the names of the booster_versions which have carried the maximum payload mass. Use a subquery

```
%%sql
SELECT Booster_Version AS 'Booster Versions with Maximum Payload' FROM SPACEXTABLE WHERE PAYLOAD_MASS__KG_ IN (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTABLE)
```

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

Booster Versions with Maximum Payload
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

2015 Launch Records

- Querying information about failure landing outcomes in 2015 indicates what booster and what launch site was responsible for the undesired outcome.

List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.

Note: SQLLite does not support monthnames. So you need to use substr(Date, 4, 2) as month to get the months and substr(Date,7,4)='2015' for year.

```
%%sql
SELECT Date, Booster_Version, Launch_Site FROM SPACEXTABLE WHERE Landing_Outcome = 'Failure (drone ship)' AND Date LIKE '2015%'
```

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

Date	Booster_Version	Launch_Site
2015-10-01	F9 v1.1 B1012	CCAFS LC-40
2015-04-14	F9 v1.1 B1015	CCAFS LC-40

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

- Querying a descending list of landing outcomes from 2010 to 2017 shows the complete picture of the data for that time period.

```
%%sql
SELECT Landing_Outcome AS 'Landing Outcome',COUNT(Landing_Outcome) AS 'Frequency' FROM SPACEXTABLE WHERE (Date > '2010-06-04' AND Date < '2017-03-20') GROUP BY Landing_Outcome ORDER BY 2 DESC
```

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

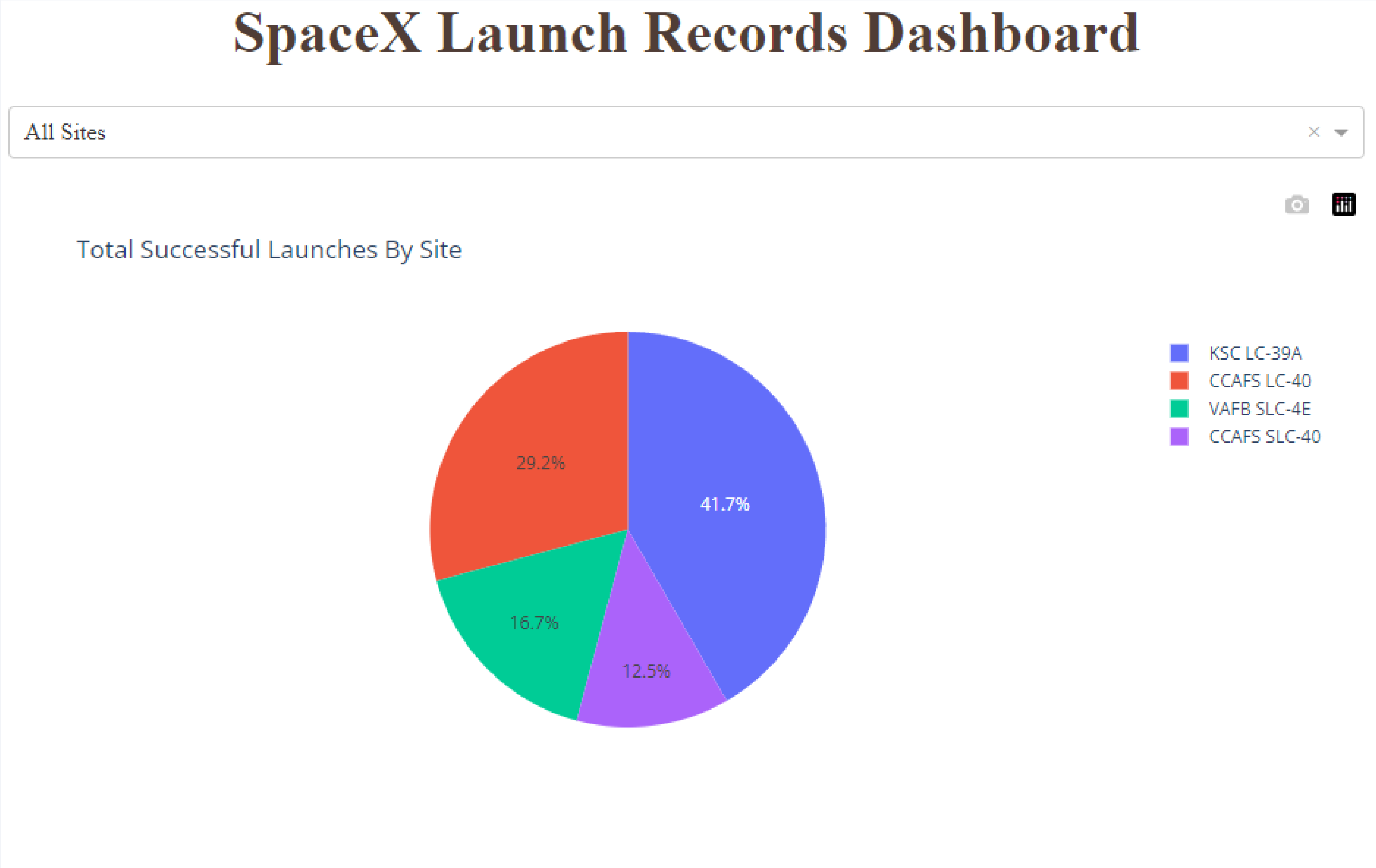
Landing Outcome	Frequency
No attempt	10
Success (ground pad)	5
Success (drone ship)	5
Failure (drone ship)	5
Controlled (ocean)	3
Uncontrolled (ocean)	2
Precluded (drone ship)	1
Failure (parachute)	1

Section 3

Build a Dashboard with Plotly Dash

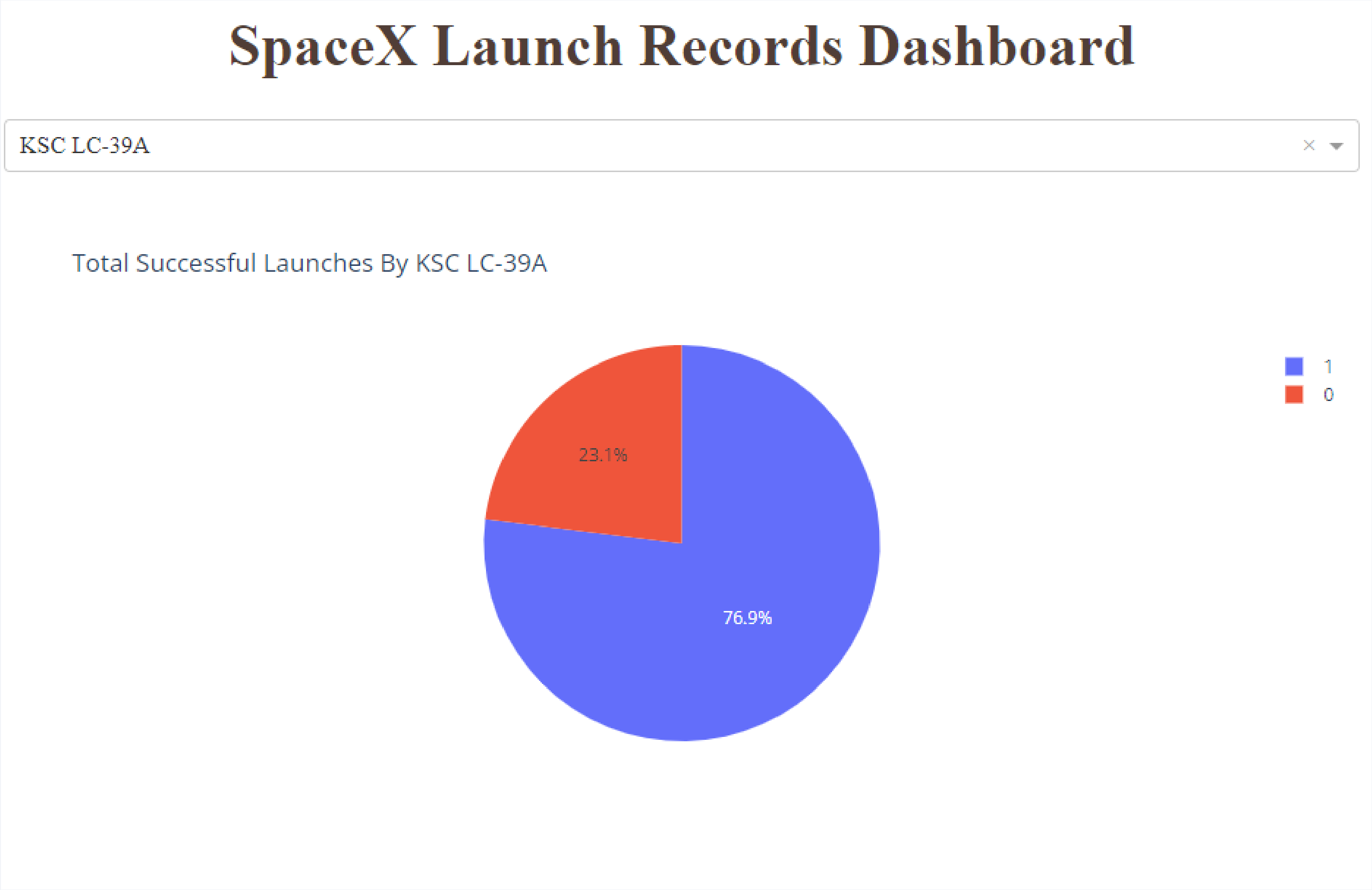
Total Successful Launches By Site Pie Chart

- KSC LC-39A has the most amount of successful launches
- CCAFS SLC-40 has the least amount of successful launches
- https://github.com/mkelly44/Applied-Data-Science-Capstone/blob/main/spacex_dash_app%20(1).py



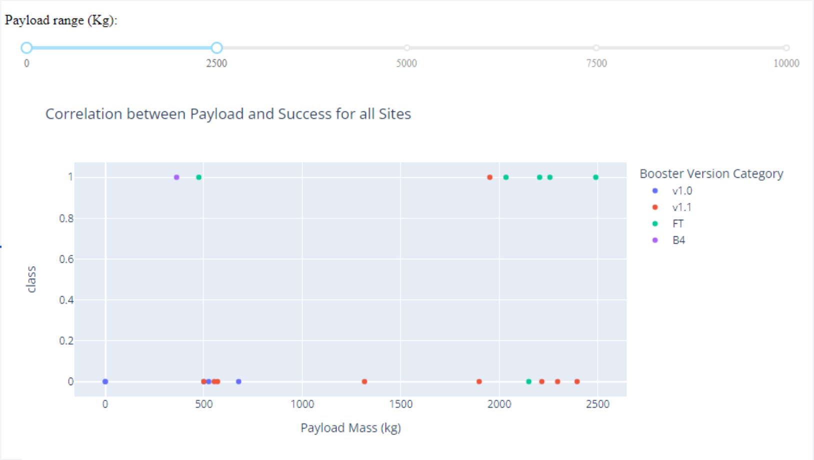
Total Successful Launches By KSC LC-39A

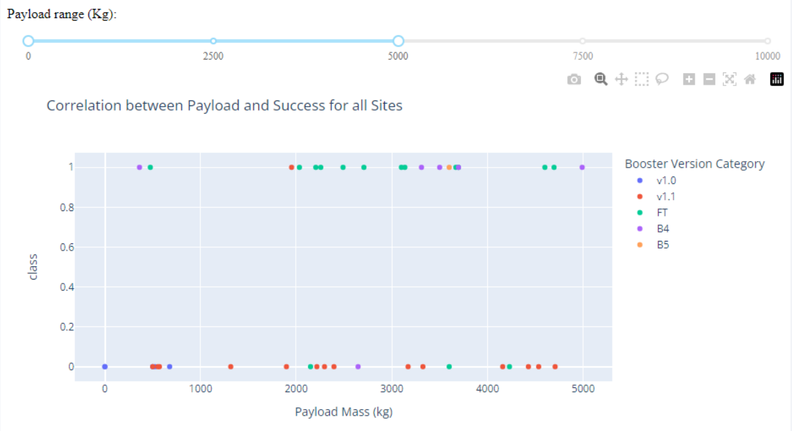
- 76.9% of all launches from KSC LC-39A launch site have the first stage land successfully.
- Shows that 3 out of every 4 rockets will have a landing first stage if they took off from KSC LC-39A.



Correlation Between Payload and Success for all Sites

- In the range of 0-2500 kg of payload mass, the FT booster has the most successful outcomes.
- For both the 0-2500 kg and 0-5000 kg range of payload mass, booster version v1.1 has only a single successful landing.
- Although FT and B4 consistently result in a successful outcome, their unsuccessful outcomes are equally as frequent.
- The B5 booster has a 100% success rate, but only one launch was recorded.



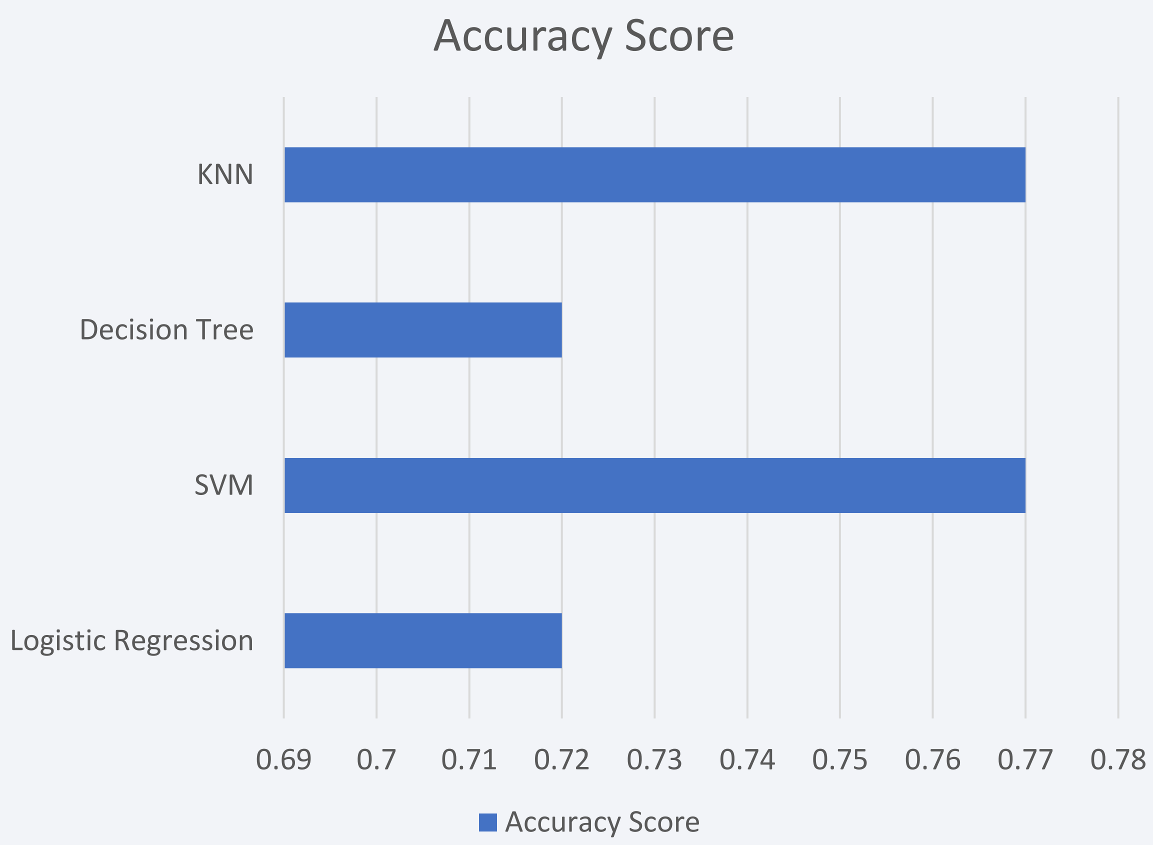


Section 4

Predictive Analysis (Classification)

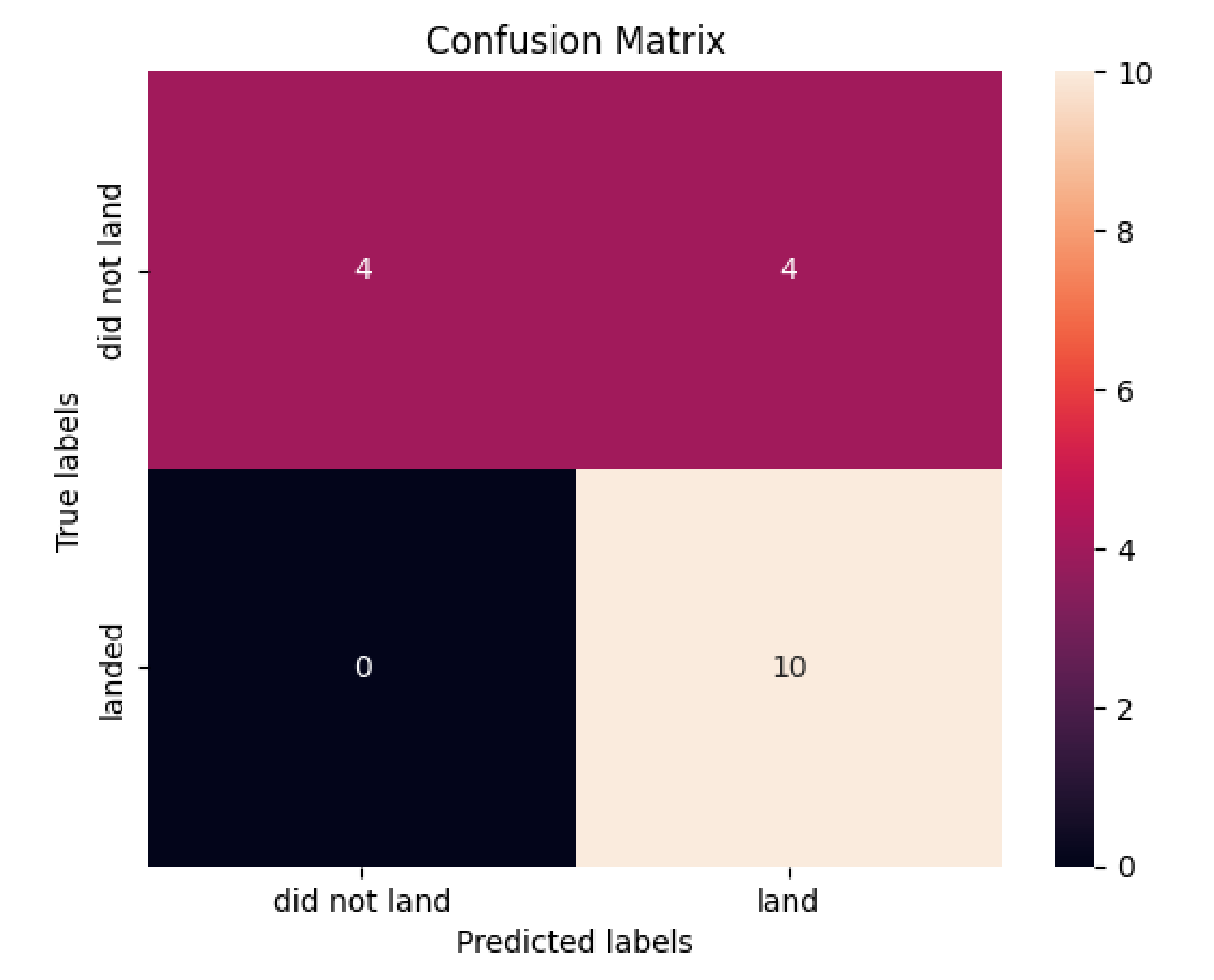
Classification Accuracy

- KNN and SVM tied for highest accuracy model
- KNN best model since hyperparameter based accuracy the highest, and training time is 2x faster than SVM



Confusion Matrix

- KNN predicted all10 landing labels correctly and none incorrectly
- KNN split the correct labeling for first stage not landing 4 to 4
- KNN model is very good at predicting true “landed” labels but not true “did not land” labels.

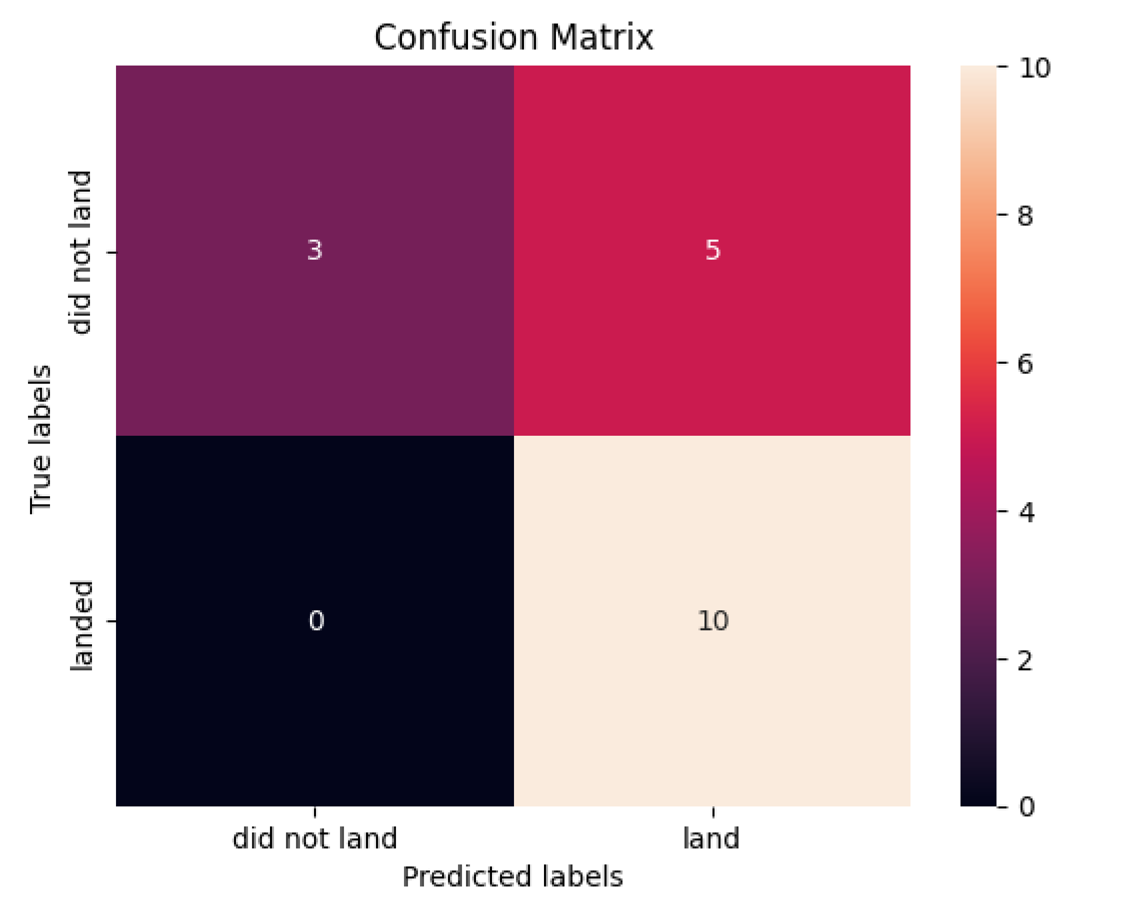


Conclusions

- The KNN machine learning model for first stage rocket landing outcome was the most accurate with a score of 0.77
- As more rockets launch from SpaceX, more data will be available to increase the accuracy of the model even further, for more accurate predictions
- Flight Number and Payload Mass were the most impactful features because more flights and heavier payloads tended to have their first stage land successfully, more often

Appendix

- Confusion Matrix for Logistic Regression and Decision Tree (both had exact same values)



Thank you!