IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Ilmars Bels

17-Aug-2024

Outline

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

Executive Summary

- **Summary of methodologies**
  - Data Collection through API
  - Data Wrangling
  - EDA with SQL
  - EDA with Data Visualization
  - Machine learning prediction

- **Summary of all results**
  - EDA result
  - Interactive analysis in screenshots
  - Predictive analysis results

Introduction

- **Project background and context**

  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

- **Problems you want to find answer**
  - Does the success or failure depend on the launch site? Does it depend on the landing target? -
  - Is it possible to predict a successful first stage landing of Falcon9?
  - Which machine learning model would work best (highest accuracy) to predict the outcome of Falcon9 first stage landing in a future launch?

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

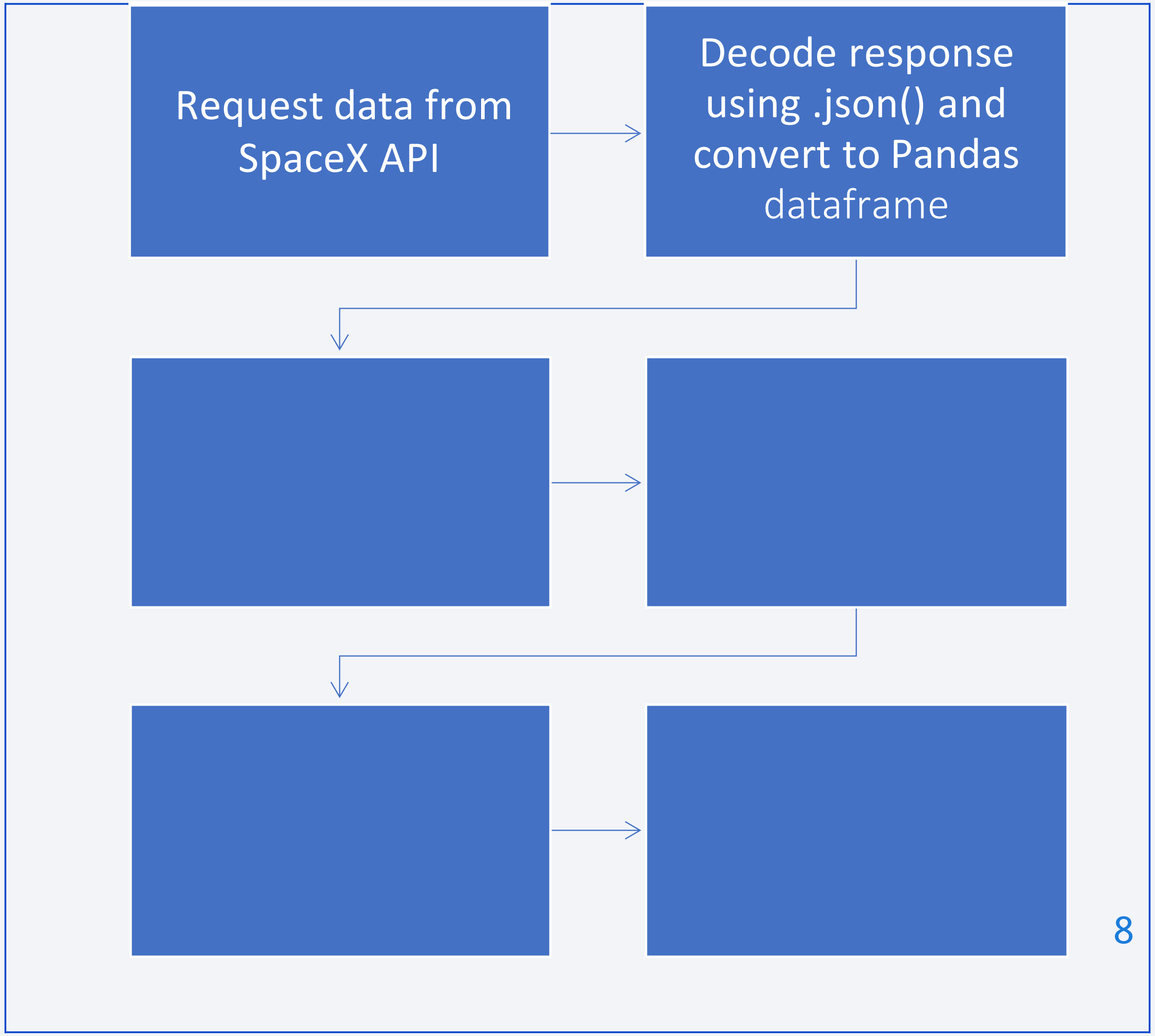
Data Collection

- The data was collected using various methods
  - Data collection was conducted using GET requests to the SpaceX API.
  - The response content was decoded as JSON using the '.json()' function and converted into a pandas DataFrame using '.json_normalize()'.
  - The data was then cleaned, checked for missing values, and any missing values were filled as necessary.
  - Additionally, web scraping was performed using BeautifulSoup to obtain Falcon 9 launch records from Wikipedia

Data Collection – SpaceX API

- We utilized GET requests to the SpaceX API to collect data, followed by data cleaning, basic wrangling, and formatting

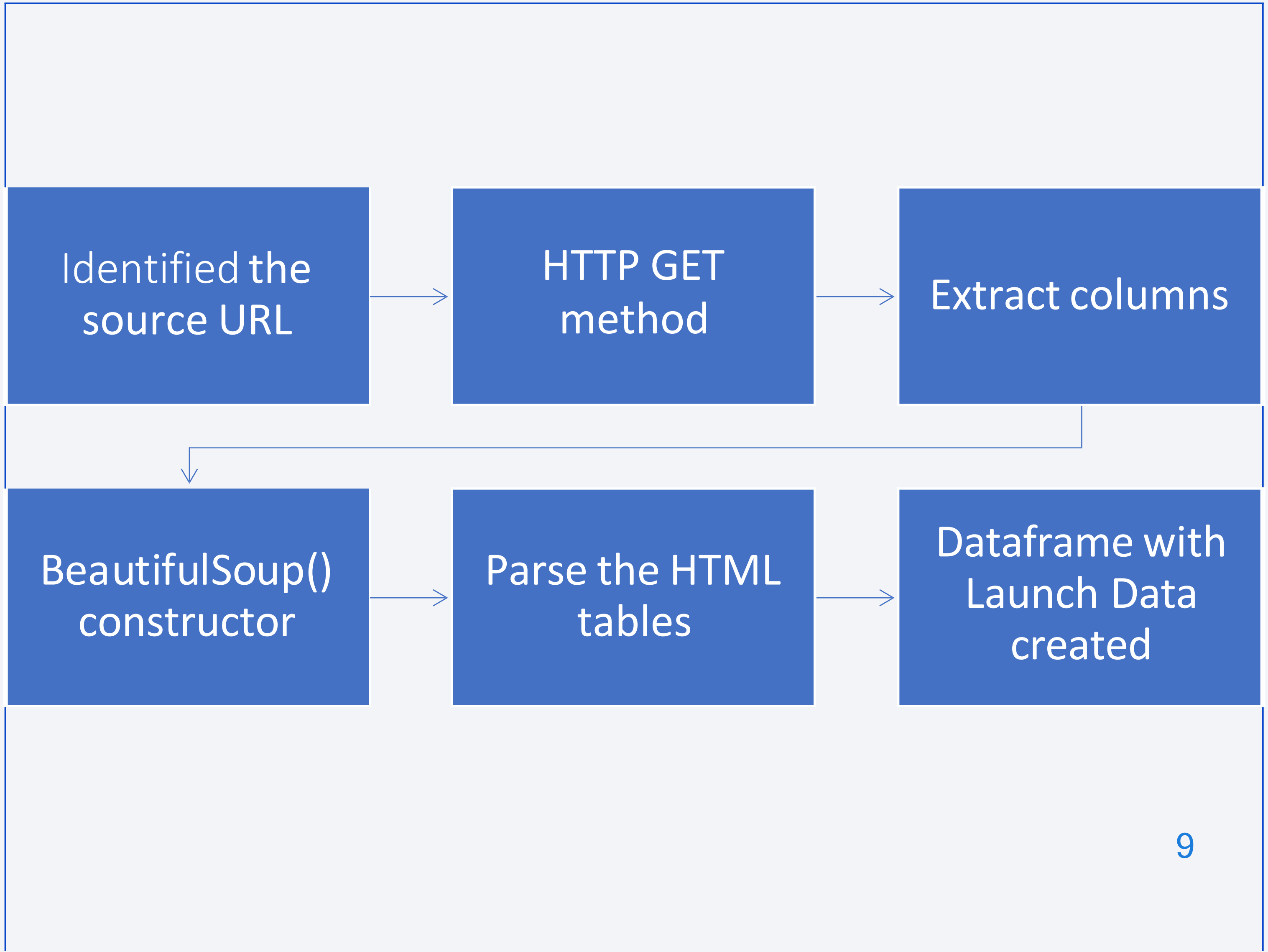
- [Data Collection notebook]



Data Collection - Scraping

- We used web scraping with BeautifulSoup to gather Falcon 9 launch records.

- [Here is web scrapping notebook]()



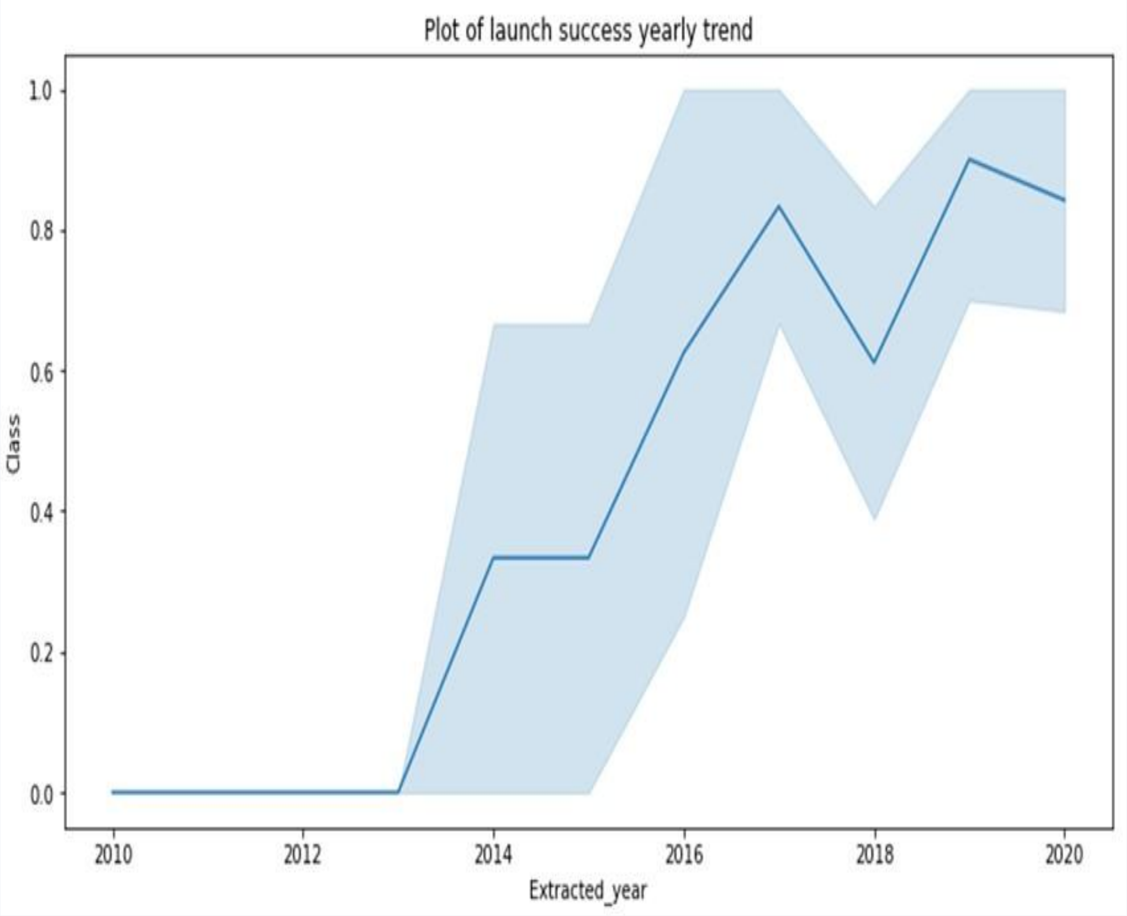
Data Wrangling

- We performed exploratory data analysis and determined the training labels.
- We calculated the number of launches at each site, and the number and occurrence of each orbits
- We created landing outcome label from outcome column and exported the results to csv

- Here is data wrangling notebook

EDA with Data Visualization

- We explored the data by visualizing various relationships, including flight number and launch site, payload and launch site, success rate of each orbit type, flight number and orbit type, and the yearly trend of launch success



- [Here is EDAviz notebook]

EDA with SQL

- EDA was executed with SQL to extract information about:
  - Launch Sites
  - Payload masses
  - Dates
  - Booster Types
  - Mission outcomes

- HERE IS SQL NOTEBOOK

Build an Interactive Map with Folium

- Mark all launch sites on a map
  - Marker with popup on NASA Johnson Space Center
  - Circle for each launch site
- Mark the success/failed launches for each site on the map
  - Marker Cluster marking success and failed launches for each site
- Calculate the distances between a launch site to its proximities
  - Calculate distance to coastline and Cape Canaveral
  - Draw line from launch site to points

- [Here is link]

Build a Dashboard with Plotly Dash

- Dropdown List with Launch Sites
  - Select all launch sites or certain launch site
- Pie Chart Showing Successful Launches
  - Displays successful and unsuccessful launches as percent of total launches per site
- Slider of Payload Mass Range
  - Select payload mass range
- Scatter Chart Showing Payload Mass vs. Success Rate by Booster Version
  - Shows correlation between Payload and Launch Success

Predictive Analysis (Classification)

- We loaded the data using numpy and pandas, then transformed and split it into training and testing sets.
- We built various machine learning models and fine-tuned hyperparameters using GridSearchCV.
- Accuracy was used as the performance metric.
- Through feature engineering and algorithm tuning, we improved the model and identified the best-performing classification model.

- predictive analysis link

Results

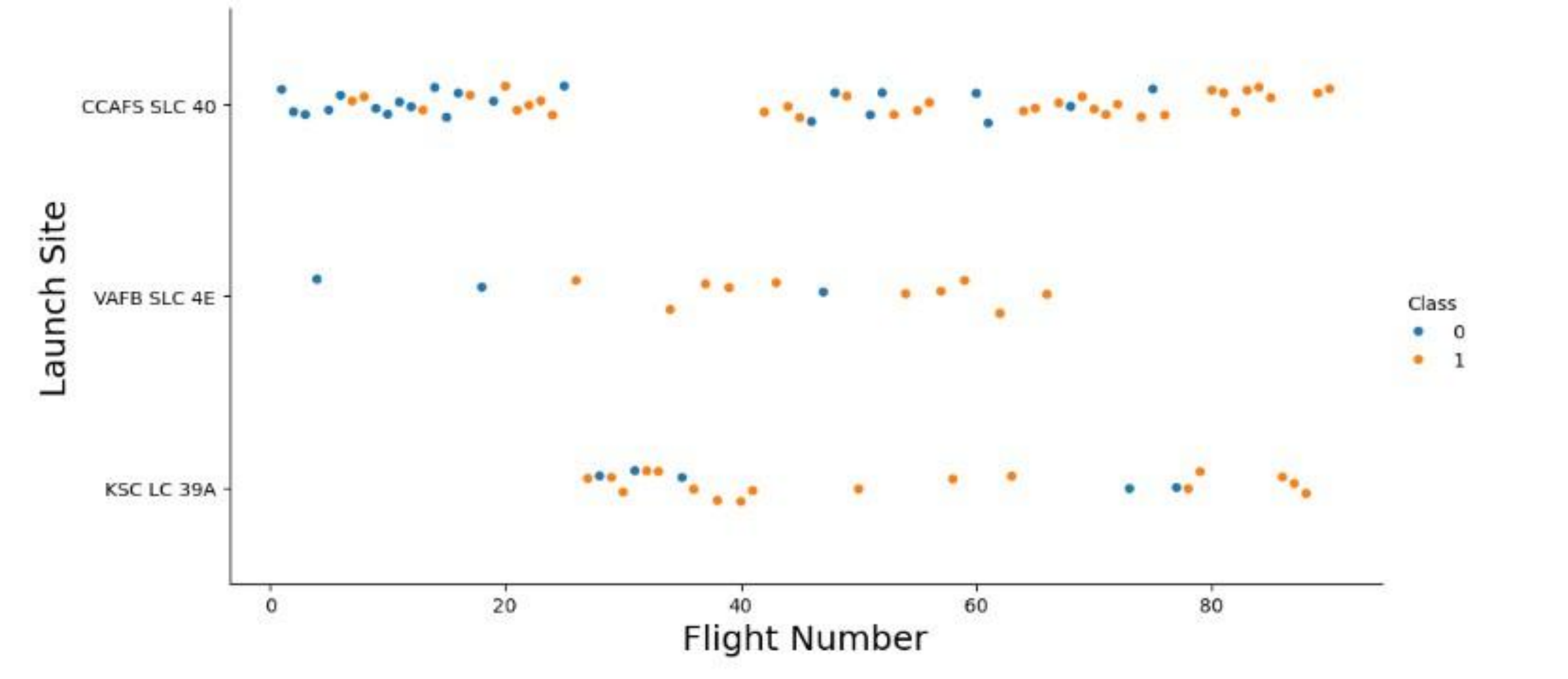
- Exploratory data analysis results in Section 2
- Interactive analytics demo in screenshots in Sections 3 and 4
- Predictive analysis results in Section 5

Section 2

Insights drawn from EDA

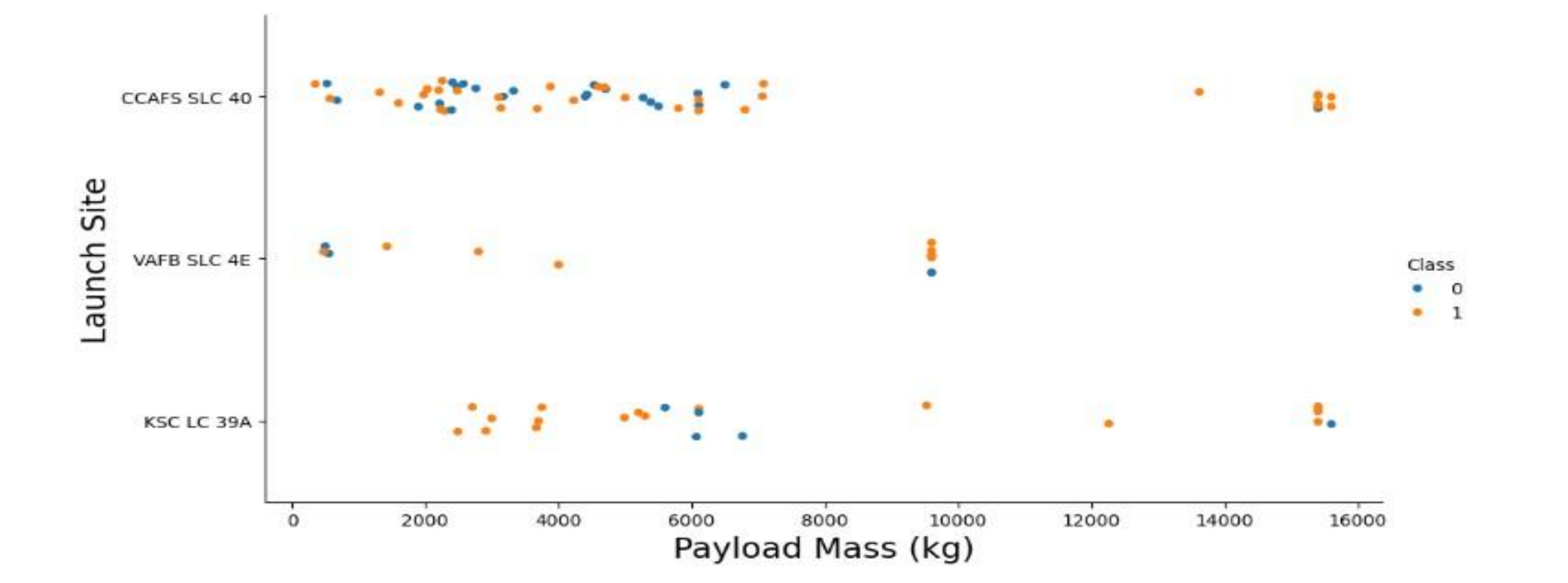
Flight Number vs. Launch Site

The plot revealed that higher flight volumes at a launch site correlate with increased success rates



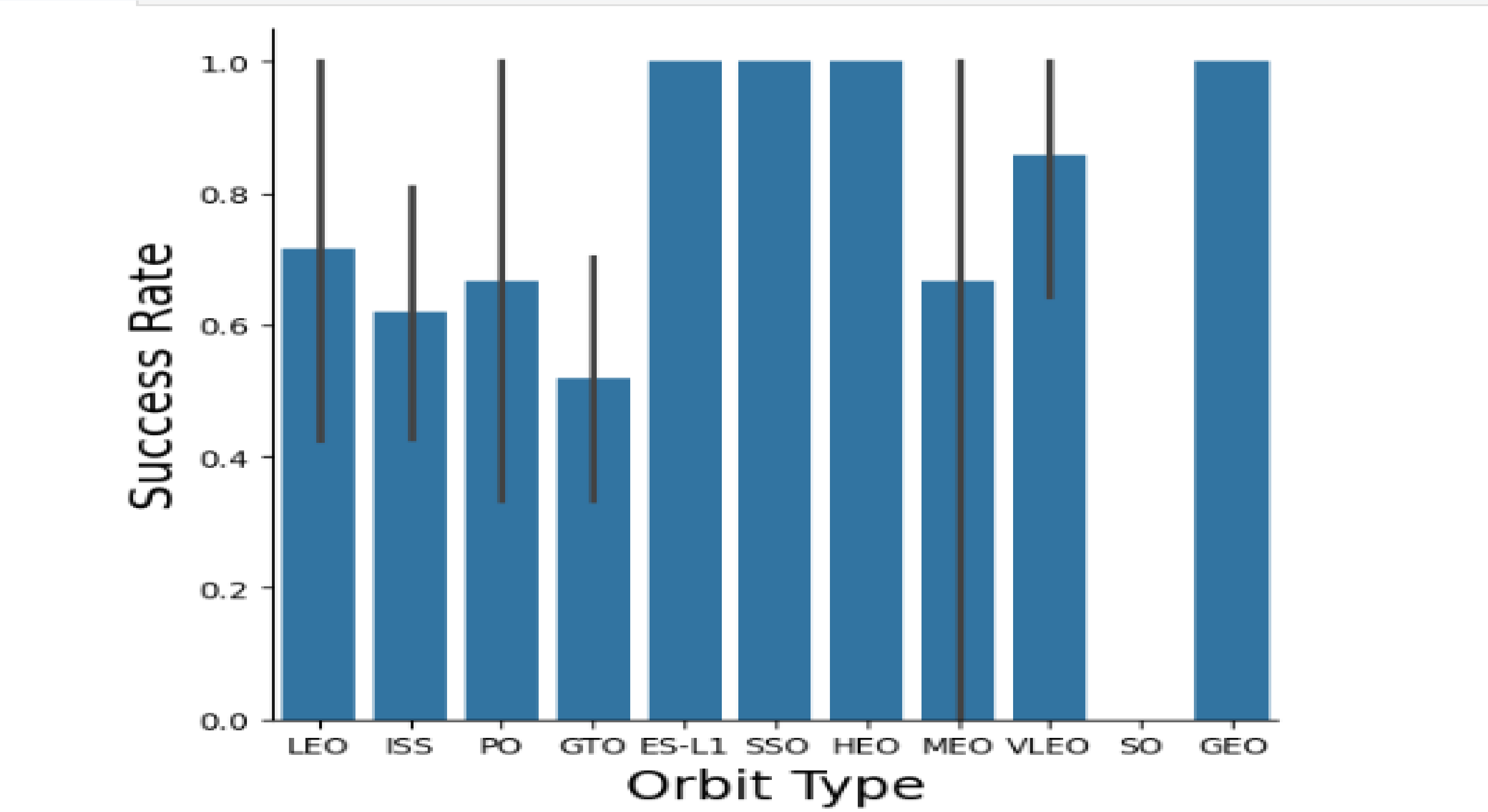
Payload vs. Launch Site

- KSC LC 39A has a 100% successrate for launches less than 5,500 kg
- VAFB has not launched anything greater than 10,000 kg



Success Rate vs. Orbit Type

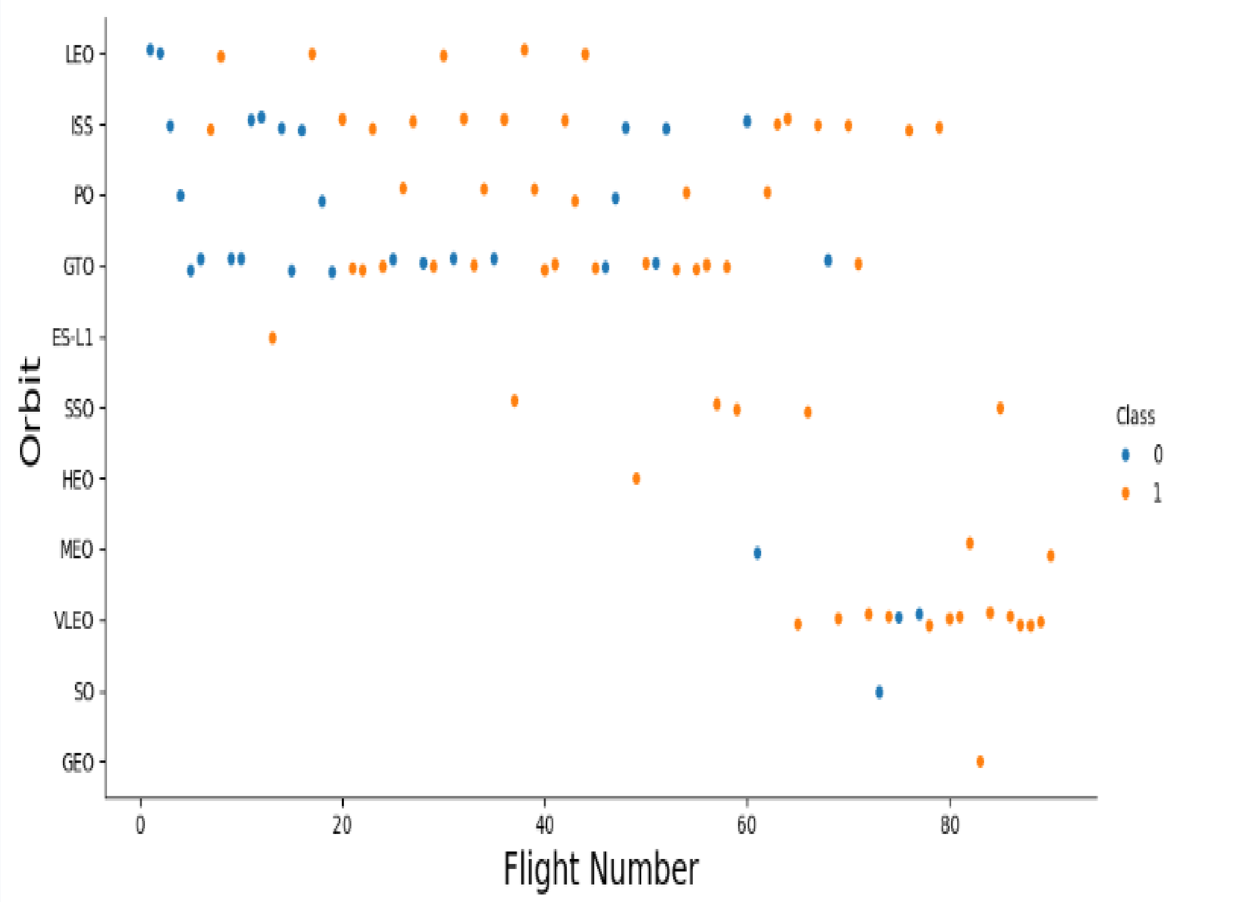
The plot indicates that ES-L1, GEO, HEO, SSO, and VLEO had the highest success rates5



Flight Number vs. Orbit Type

- Visually, can be observed that:
  - For LEO Orbit, the success seems to be related to the Number of Flights.
  - For GTO Orbit, it seems that there is no relationship between the number of flights and the orbit.

•It looks like there is a correlation between Fligth Number and Success Rate with larger flights number being associated with higher success rates.

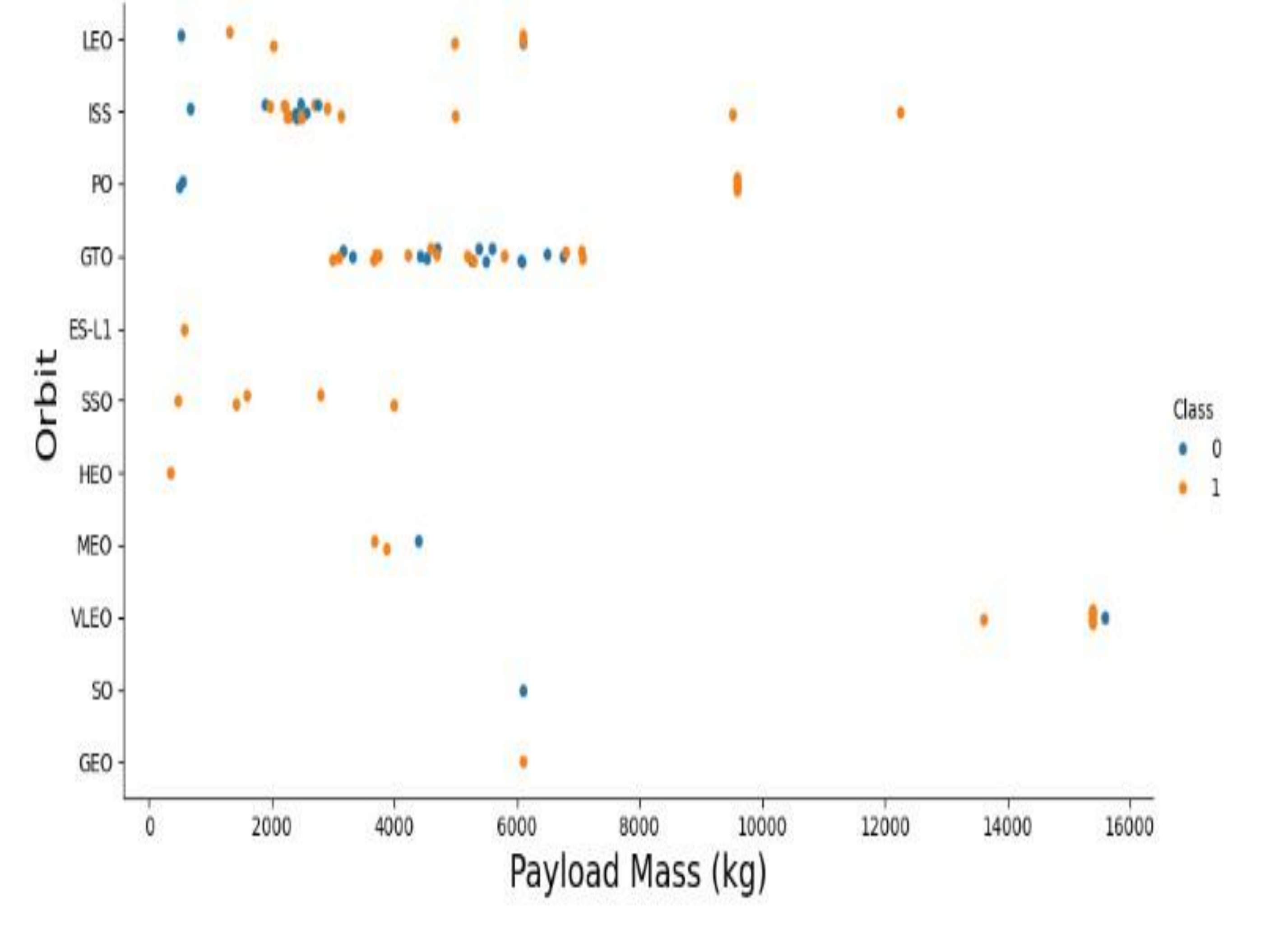


Payload vs. Orbit Type

For Orbits PO, LEO and ISS, successful landing (or positive landing) looks related to the heavier payloads.

•For GTO Orbit, it is not possible to identify impacts of payload (positive or negative).

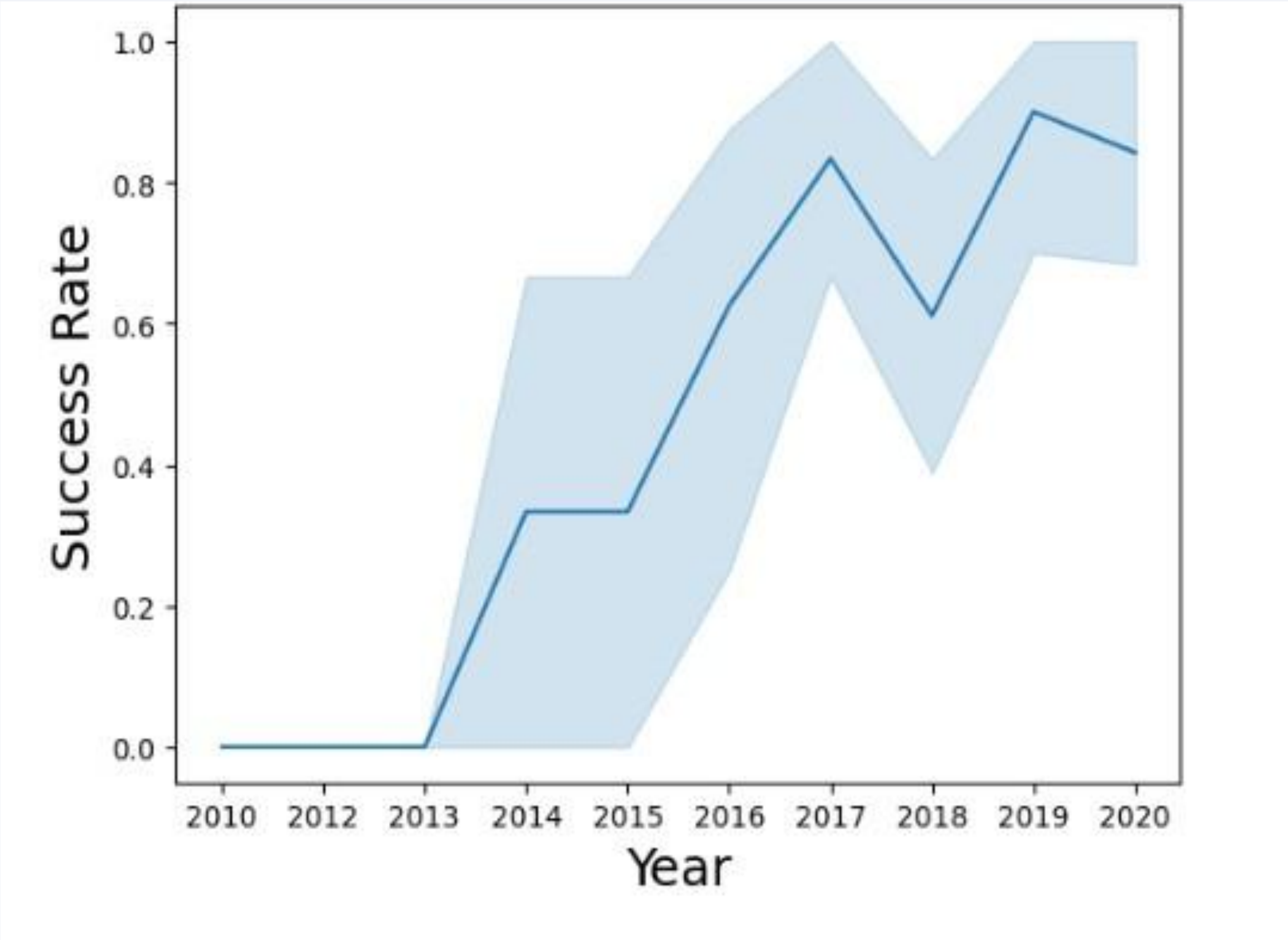
•Considering all orbit types, it is not possible to establishif there is a correlation between Success Rate and Payload Mass



Launch Success Yearly Trend

Sucess Rate has increased significantly over the Years, starting on 2013

- There was a small decrease in 2018. Unknown causes without deeper analysis.



All Launch Site Names

- Launch Sites:
  - CCAFS LC-40
  - VAFB SLC-4E
  - KSC LC-39A
  - CCAFS SLC-40

Used keyword DISTINCT

In [19]:

```
%%sql
SELECT DISTINCT LAUNCH_SITE FROM SPACEXTBL;
```

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

Out[19]:

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

Launch Site Names Begin with 'CCA'

Launch Sites, with name starting with ‘CCA’:

Used the ”like” query to display records that begin with ‘CCA

In [20]:
```
%%sql
select * from SPACEXTBL where LAUNCH_SITE like "CCA%" limit 5
```

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

Out[20]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-12-08	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

In [21]:

```
%%sql
select sum(PAYLOAD_MASS__kg_) from SPACEXTBL where customer = 'NASA (CRS)'
```

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

Out[21]:

sum(PAYLOAD_MASS__kg_)
45596

45596 KG total

Average Payload Mass by F9 v1.1

In [24]:
```
%%sql
select avg(PAYLOAD_MASS__kg_) from SPACEXTBL where BOOSTER_VERSION = 'F9 v1.1'
```

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

Out[24]:

avg(PAYLOAD_MASS__kg_)
2928.4

2,928 kg average

First Successful Ground Landing Date

In [30]:

```
%%sql
select min(DATE) from SPACEXTBL where Landing_Outcome = 'Success (ground pad)'
```

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

Out[30]:

min(DATE)
2015-12-22

12/22/15 is first successful ground landing date

Successful Drone Ship Landing with Payload between 4000 and 6000

- Booster version
  - F9 FT B1022
  - F9 FT B1026
  - F9 FT B1021.2
  - F9 FT B1031.2

In [40]:

```
%%sql
select Booster_Version
from SPACEXTBL
where Landing_Outcome = 'Success (drone ship)'
and Payload_Mass__kg_ > 4000
and Payload_Mass__kg_ < 6000
```

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

Out[40]:

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

Total Number of Successful and Failure Mission Outcomes

- 1 failure in flight
- 99 successes
- 1 success (payload status unclear)

In [47]:

```
%%sql

select MISSION_OUTCOME, count(*) as total_number
from SPACEXTBL
group by MISSION_OUTCOME;
```

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

Out[47]:

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

Boosters Carried Maximum Payload

- F9 B5 B1048.4
- F9 B5 1049.4
- F9 B5 1051.3
- F9 B5 1056.4
- F9 B5 1048.5
- F9 B5 1051.4
- F9 B5 1049.5
- F9 B5 1060.2
- F9 B5 1058.3
- F9 B5 1051.6
- F9 B5 1060.3
- F9 B5 1049.7

In [52]:

```
%%sql
select BOOSTER_VERSION
from SPACEXTBL
where PAYLOAD_MASS__KG_ = (
    select MAX(PAYLOAD_MASS__KG_) from SPACEXTBL
    );
```

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

Out[52]:

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

2015 Launch Records

Shows Month, date, booster version, launch site, and landing outcome

```
In [57]: %%sql
select substr(Date,6,2) as month, DATE,BOOSTER_VERSION, LAUNCH_SITE, Landing_Outcome
from SPACEXTBL
where Landing_Outcome = 'Failure (drone ship)' and substr(Date,0,5)='2015';

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

Out[57]:

month	Date	Booster_Version	Launch_Site	Landing_Outcome
01	2015-01-10	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
04	2015-04-14	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

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

Ranked in descending order

The most common landing outcome was "No Attempt

In [64]:

```
%%sql

select Landing_Outcome, count(*) as count_outcomes
from SPACEXTBL
where date between '2010-06-04' and '2017-03-20'
group by Landing_Outcome
order by count(Landing_Outcome) desc;
```

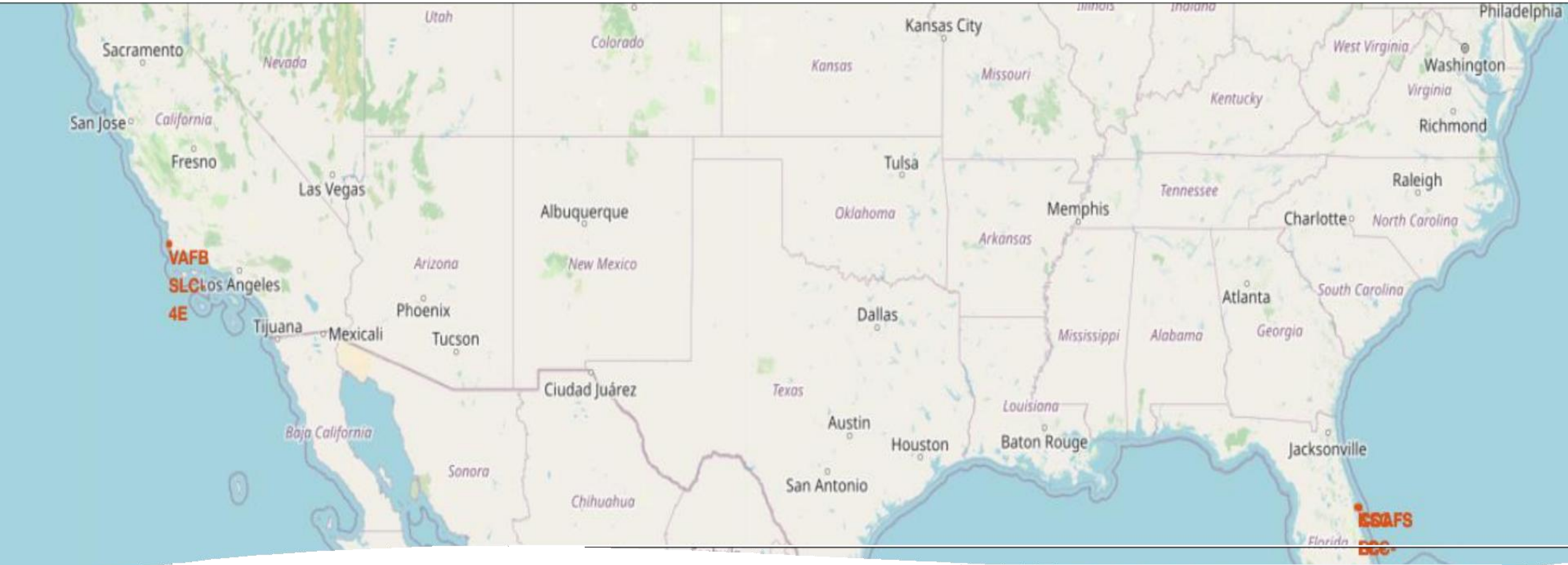
```
 * sqlite:///my_data1.db
Done.
```

Out[64]:

Landing_Outcome	count_outcomes
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

Section 3

Launch Sites Proximities Analysis



All launch sites are in low latitude, closest to Equator. This may be explained that rockets can most easily reash satellite orbits if launched near the Equator in an easterly direction (as this maximizes use of Earth’s rotational speed) - this also provides a desirable orientation for arriving at GEO

Out[8]:

	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

Folium – Success/Failed Launches

Green Marks: Successful launches
Red Marks: Failed launches

These views and analysis allow a quick understanding of success rate for each site When launching the rockets. While it does not explain the why some sites have more failures than others, it allows this understanding and provide the basis for deeper furture checks

Distance from Launch Site to other geographical marks

- The launch sites are not in close proximities to railways and highways
- They are in close proximity to the coastline and do keep distance from cities

Section 4

Build a Dashboard with Plotly Dash

Dashboard – Total Launch Success Count by Site

- A dropdown menu, by default, showing “All Site”, select the data for the Pie Chart below.

•With this selection, the pie chart displays the distribution of successful launches per site.

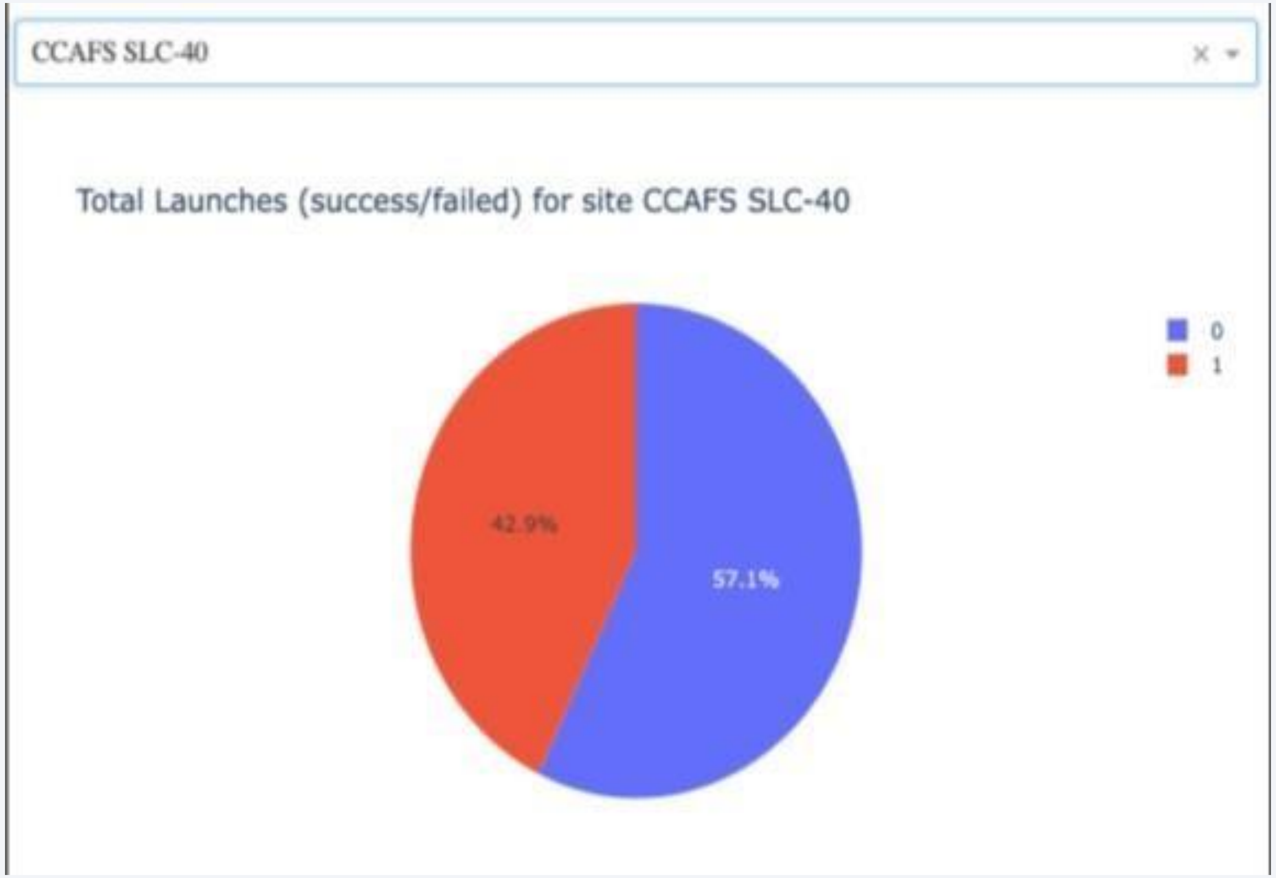
- It allows a quick view that:

    •Greatest success rate is for Site KSC LC-39A, with 41.7%;

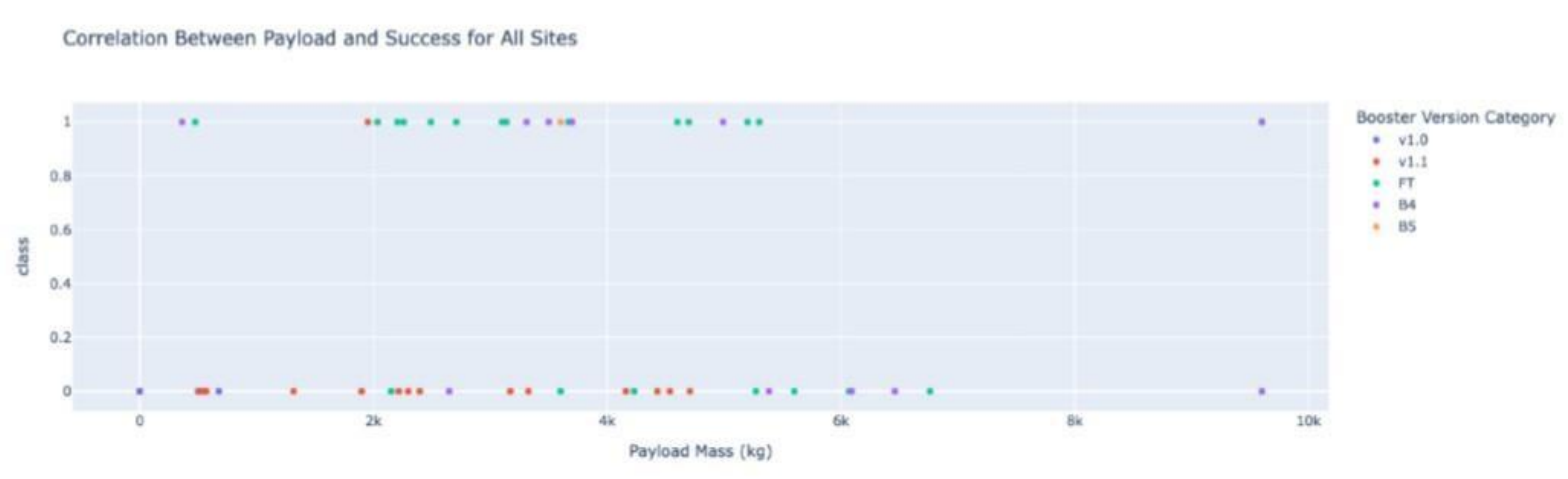
    •Lowest success rate is for Site CCAFS SLC-40, with 12.5%.

Dashboard – Total Successful launches for KSC LC-39A

- 76.9% of launches from site KSC LC-39A were successful
- Only 23.1% of launches were unsuccessful



<Dashboard Screenshot 3>



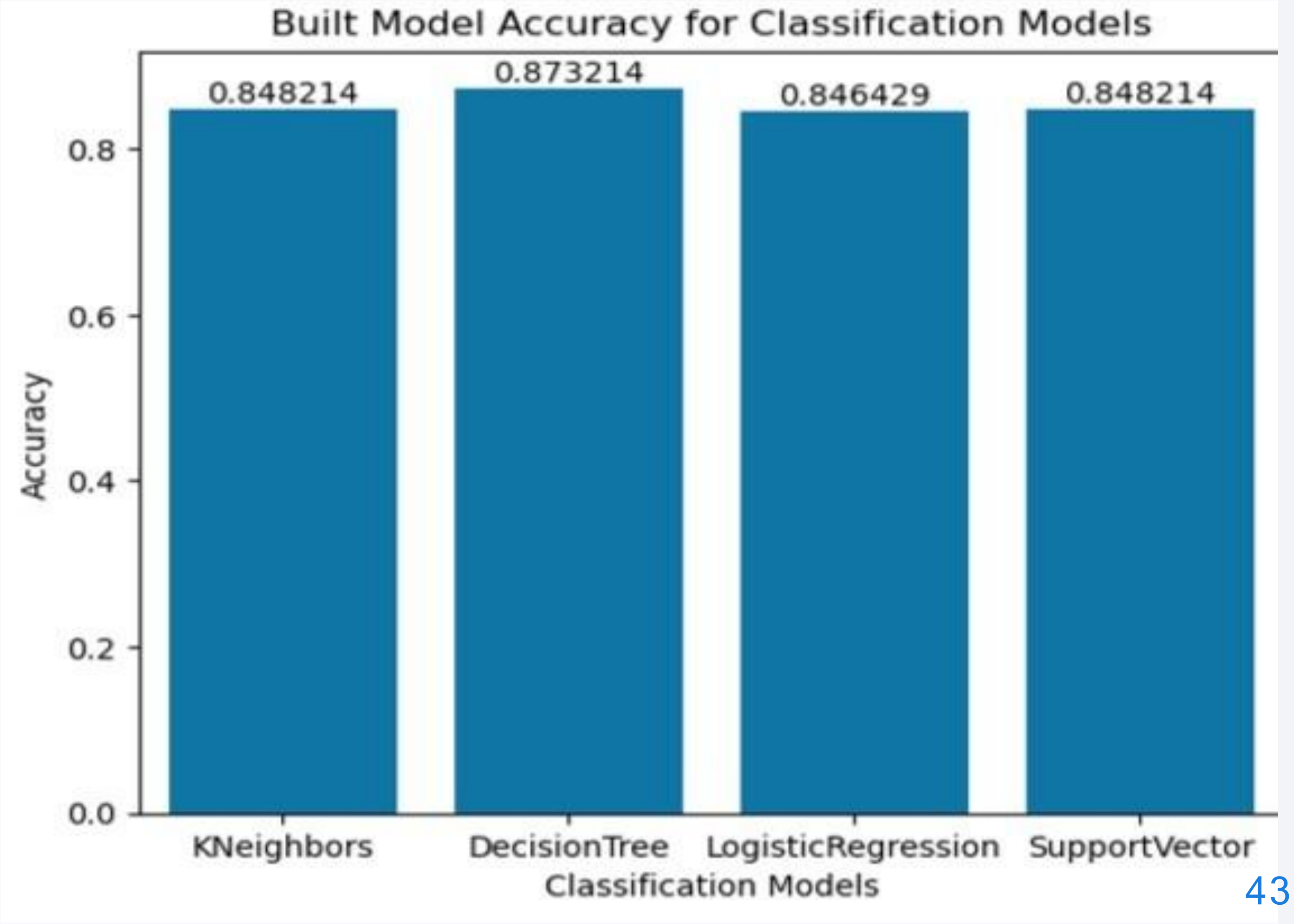
- Most launches with Booster Version 1.1 failed at all payloads
- The launch with the highest payload that was successful was launched from VAFB SLC-4E with a B4 booster, and weighed 9600 kg

Section 5

Predictive Analysis (Classification)

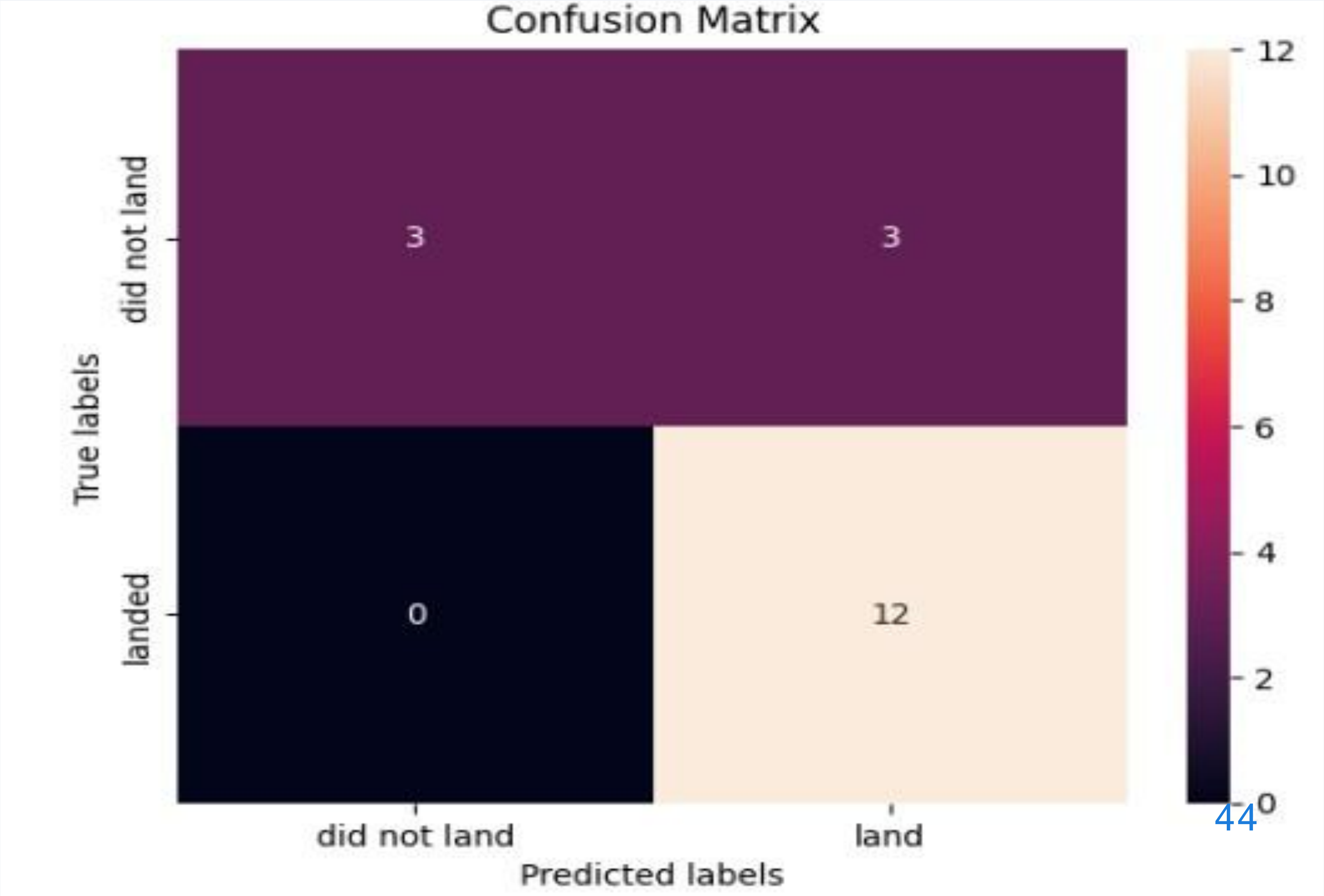
Classification Accuracy

- The Decision Tree model has the best model accuracy with 87.3214%



Confusion Matrix

- The classifier can distinguish between the different classes
- The biggest issue is with false positives i.e. unsuccessful landings being marked as successful by the classifier.



Conclusions

- How payload mass, launch site, number of flights, and orbits affect if the first-stage will land successfully
  - The higher the payload mass, the higher the success rate
  - ES-L1, GEO, HEO, and SSO orbits have a 100% success rate
- Rate of successful landings over time
  - Launch success increases over time
- Best predictive model for successful landing (binary classification)
  - The Decision tree model slightly outperforms other models

Thank you!