Winning the Space Race

Jason Pagan

Outline

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

Executive Summary

Summary of methodologies

- Collect first stage landing data form and API.
- Web scrape Falcon 9 historical launch records from Wikipedia.
- Exploratory Data Analysis to determine data for training supervised models.
- Explored data using SQL.
- Data Visualization.
- Launch Site Location Analysis with Folium
- Payload Dashboard
- Landing Prediction with Machine Learning

Summary of all results

- Landing success rate of 67%
- Decision tree classifier is the best predictor of First Stage landings (89% Accuracy).

Introduction

- **Background:**
- The SpaceY company would like to compete with Spacex.
- Falcon 9 rocket launches cost of 62 million dollars.
- Spacex has best pricing due to rocket stage recovery.

- **Problem:**
- Determine the price of each launch.
- Determine if Spacex will reuse the first stage

Section 1

Methodology

Methodology

- Executive Summary
- Data collection methodology:
  - Combined data from SpaceX public API and web scraped SpaceX Wikipedia page
- Perform data wrangling
  - Classified landings as successful and unsuccessful otherwise
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Models for SVM, Classification Trees and Logistic Regression are trained and hyperparameters are selected using the function *GridSearchCV*.

Executive Summary

- Data collection process involved a combination of API requests from SpaceX public API and web scraping data from a table in SpaceX's Wikipedia entry.

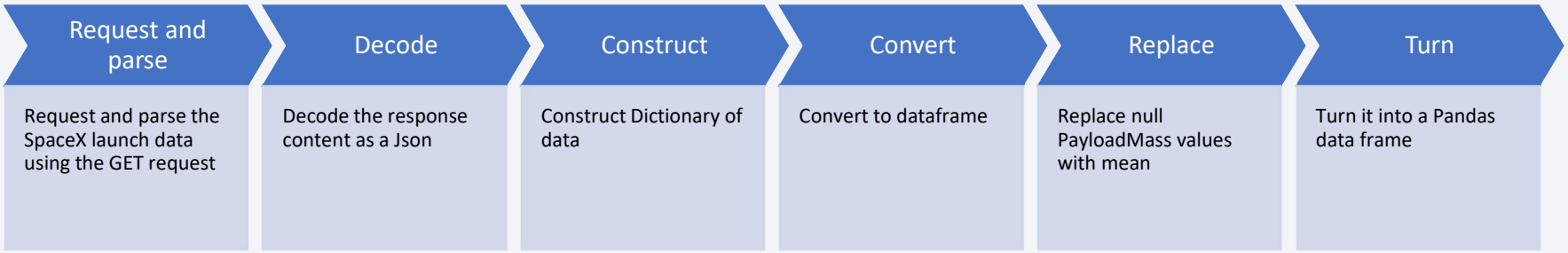
Space X API Data Columns:

- FlightNumber, Date, BoosterVersion, PayloadMass, Orbit, LaunchSite, Outcome, Flights, GridFins, Reused, Legs, LandingPad, Block, ReusedCount, Serial, Longitude, Latitude

Wikipedia Webscrape Data Columns:

- Flight No., Launch site, Payload, PayloadMass, Orbit, Customer, Launch outcome, Version Booster, Booster landing, Date, Time

Data Collection – SpaceX API



GitHub URL

https://github.com/DiscoveryUnlimited/IBM-Applied-Data-Science-Capstone/blob/main/Week%201%20jupyter-labs-spacex-data-collection-api.ipynb

Data Collection - Scraping

GitHub URL

https://github.com/DiscoveryUnlimited/IBM-Applied-Data-Science-Capstone/blob/main/Week%201/2%20jupyter-labs-webscraping.ipynb

Data Wrangling

GitHub URL

https://github.com/DiscoveryUnlimited/IBM-Applied-Data-Science-Capstone/blob/main/Week%201/3%20labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

Bar Chart

- Success rate of each orbit

Line Chart

- Success rate vs. Year

Scatter Plots

- Flight Number vs. Payload Mass
- Flight Number vs. Launch Site
- Payload Vs. Launch Site
- Orbit vs. Flight Number
- Orbit vs. Payload

GitHub URL

https://github.com/DiscoveryUnlimited/IBM-Applied-Data-Science-Capstone/blob/main/Week%202/5%20jupyter-labs-eda-dataviz.ipynb

EDA with SQL

Queried Information

- Launch site names
- Total payload mass carried by boosters launched by NASA
- Average payload mass carried by booster version
- Successful landing outcome in ground pad dates
- List the names of the boosters which have success in drone ship and have payload mass greater than 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 between the date 2010-06-04 and 2017-03-20, in descending order

GitHub URL

https://github.com/DiscoveryUnlimited/IBM-Applied-Data-Science-Capstone/blob/main/Week%202/4%20jupyter-labs-eda-sql-coursera.ipynb

Build an Interactive Map with Folium

- Red Circles mark Launch Sites
- Clusters of marker tags, color-labeled, for successful and unsuccessful landings
- Lines drawn and markers with distance to a closest city, railway, highway, and coastline

GitHub URL

https://github.com/DiscoveryUnlimited/IBM-Applied-Data-Science-Capstone/blob/main/Week%203/6%20lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

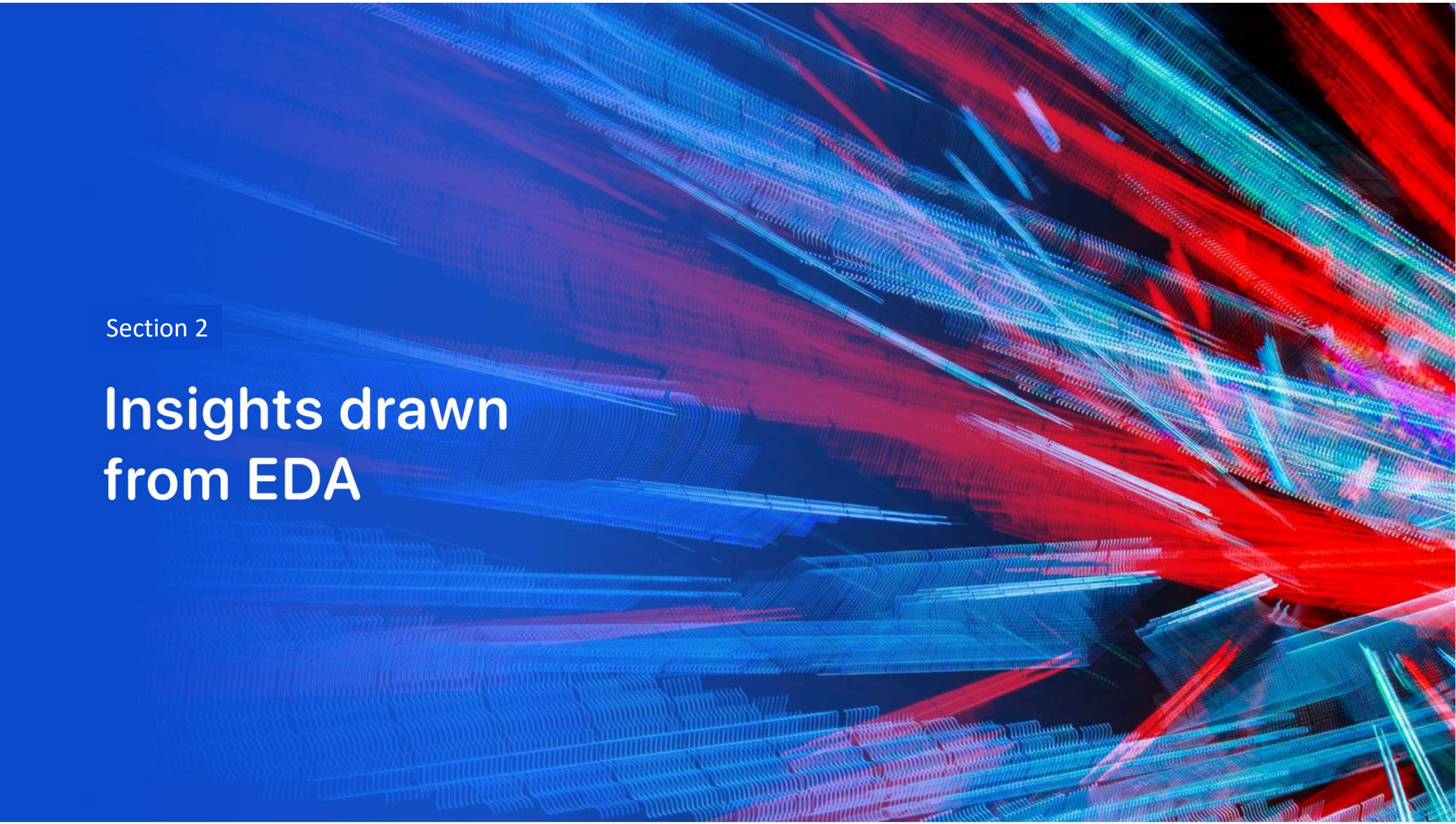
- Total Success Launches by Site (Pie Graph)
  - Dropdown Menu for site selection
- Class vs Payload Mass (Scatter Plot) with Booster Version Categories
  - Slider for payload range

GitHub URL

https://github.com/DiscoveryUnlimited/IBM-Applied-Data-Science-Capstone/blob/main/Week%203/7%20spacex_dash_app.py

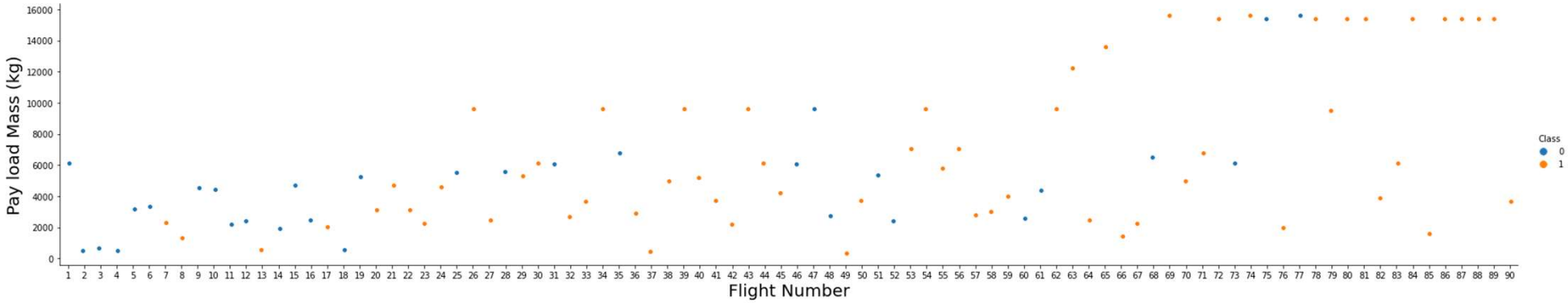
Predictive Analysis (Classification)

- Load dataframe
- Create a NumPy array from the column Class (Y)
- Standardize data (X)
- train_test_split to split the data X and Y
- Use GridSearchCV to evaluate ML models with parameters
  - logreg_cv with cv = 10
  - svm_cv
  - tree_cv
  - knn_cv
- Display best_params_ and best_score_
- Display accuracy of test data, .score()
- Confusion Matrix
- Determine best performed method



Section 2

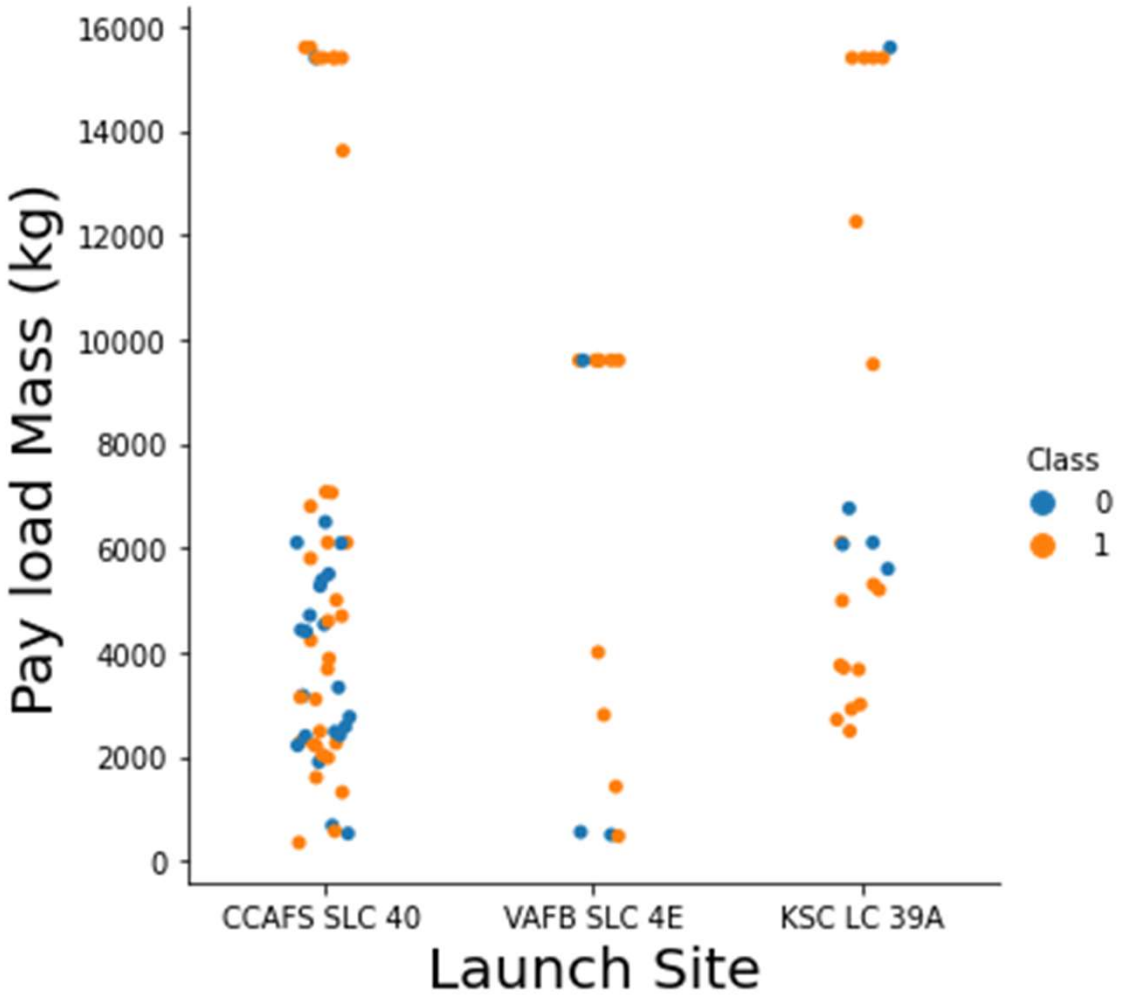
Insights drawn from EDA



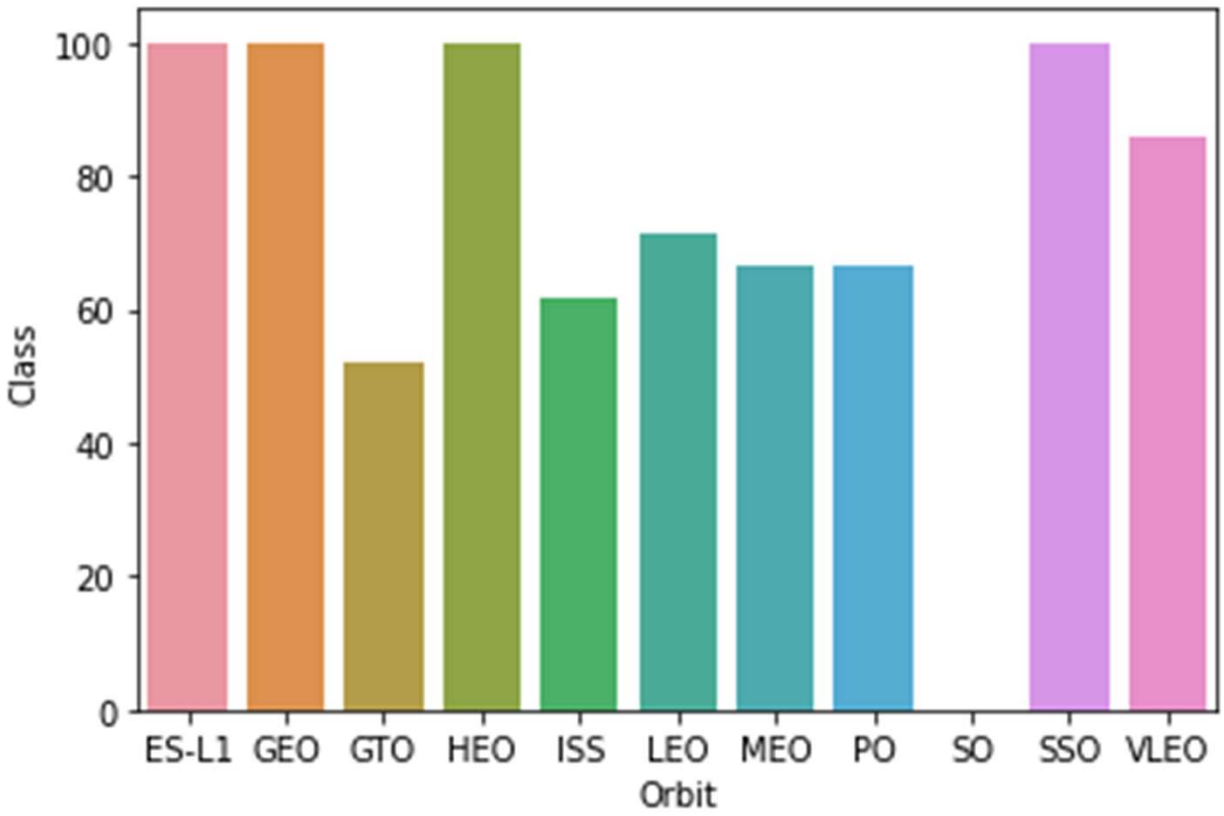
Flight Number vs. Launch Site

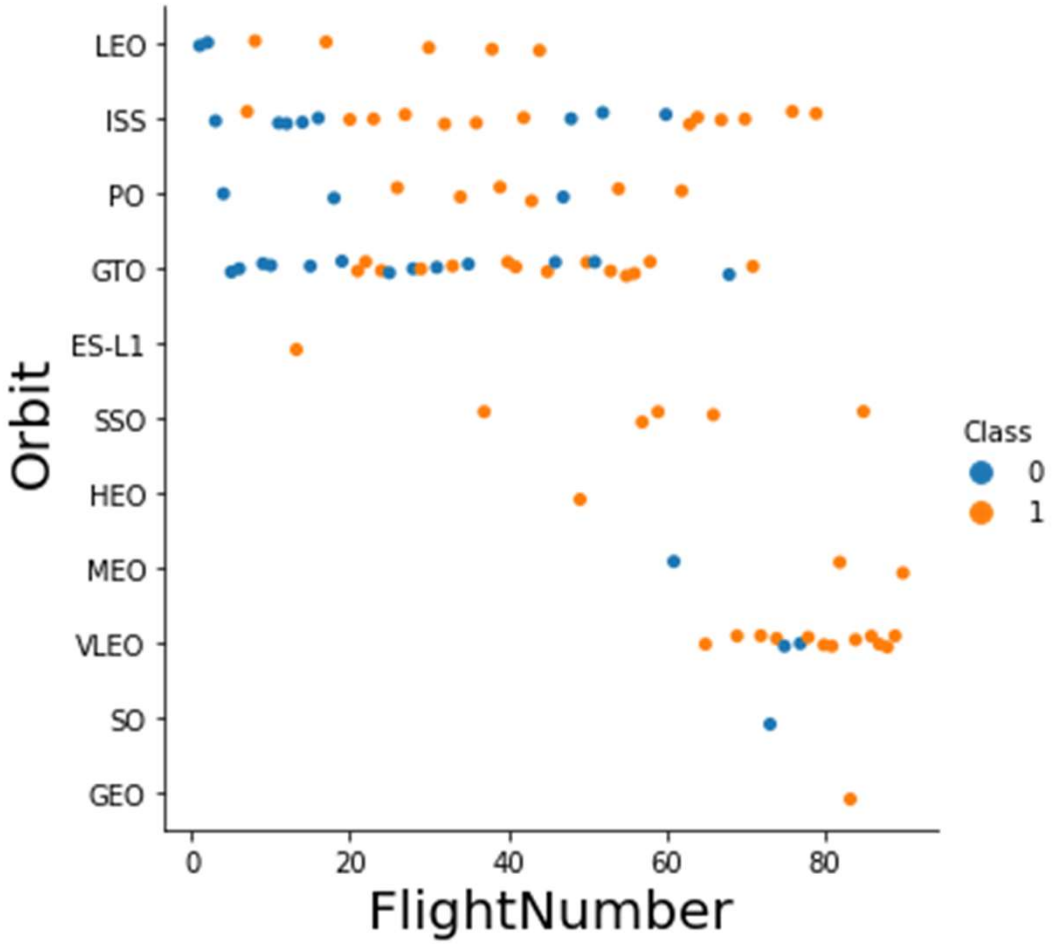
- Different launch sites have different success rates.
- CCAFS LC-40, has a success rate of 60 %
- KSC LC-39A and VAFB SLC 4E has a success rate of 77%.

Payload vs. Launch Site

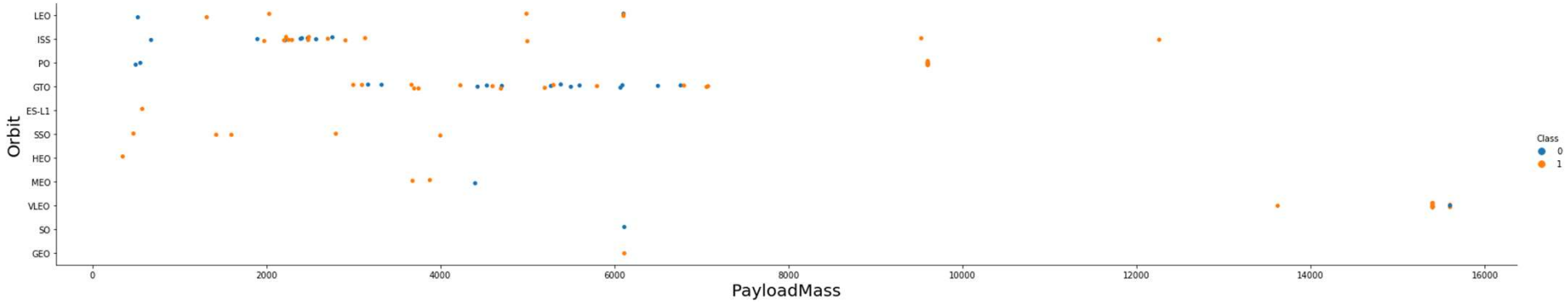


Success Rate vs. Orbit Type

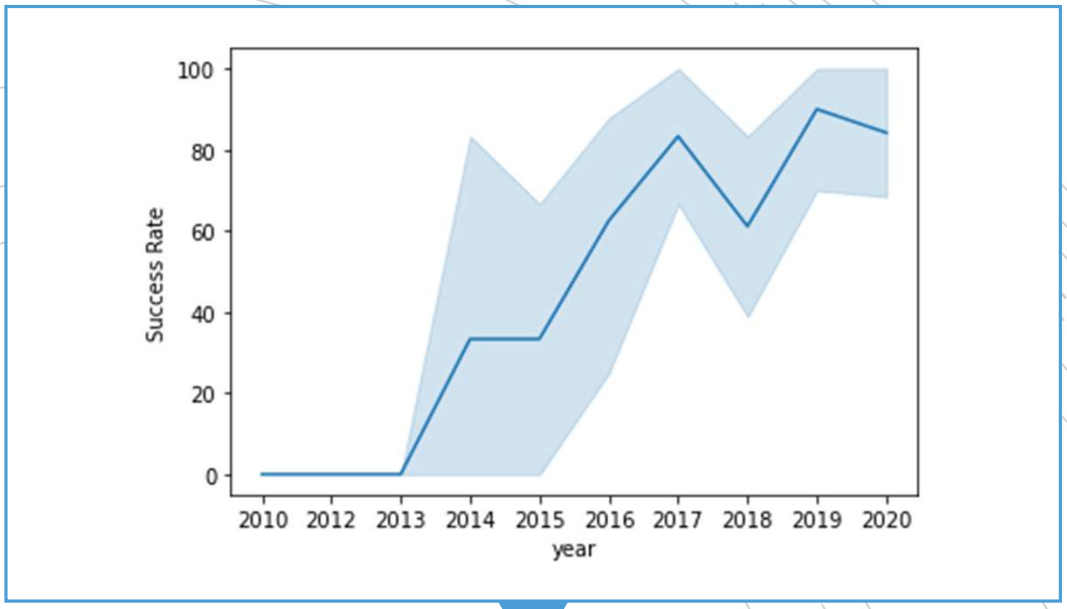




Flight Number vs. Orbit Type



Payload vs. Orbit Type



Launch Success Yearly Trend

All Launch Site Names

```
%sql select DISTINCT LAUNCH_SITE from SPACEXDATASET
```

```
 * postgresql+psycopg2://postgres:***@localhost/SpaceX
4 rows affected.
```

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

Launch Site Names Begin with 'CCA'

```
%sql select * from SPACEXDATASET where launch_site like 'CCA%' limit 5
```

```
 * postgresql+psycopg2://postgres:***@localhost/SpaceX
5 rows affected.
```

date	time_utc	booster_version	launch_site	payload	payload_mass_kg_	orbit	customer	mission_outcome	landing_outcome
04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
08-12-2010	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)
22-05-2012	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
01-03-2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

```
%%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

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

```
 * ibm_db_sa://ftb12020:***@0c77d6f2-5da9-48a9-81f8-8
Done.
```

avg_payload_mass_kg
2928

First Successful Ground Landing Date

```
%%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

```
%%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 Missio n 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

```
%%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

```
%%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

```
%%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;
```

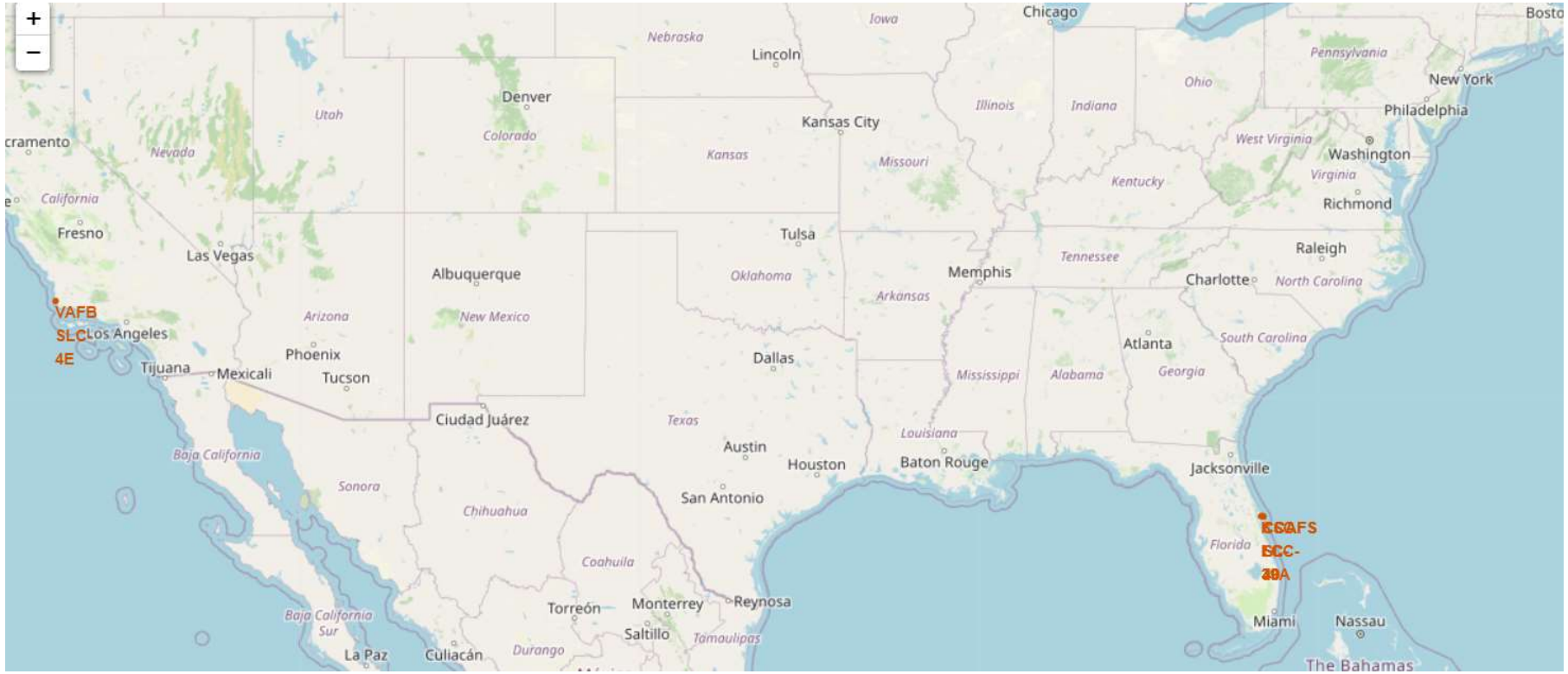
```
 * ibm_db_sa://ftb12020:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg
Done.
```

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

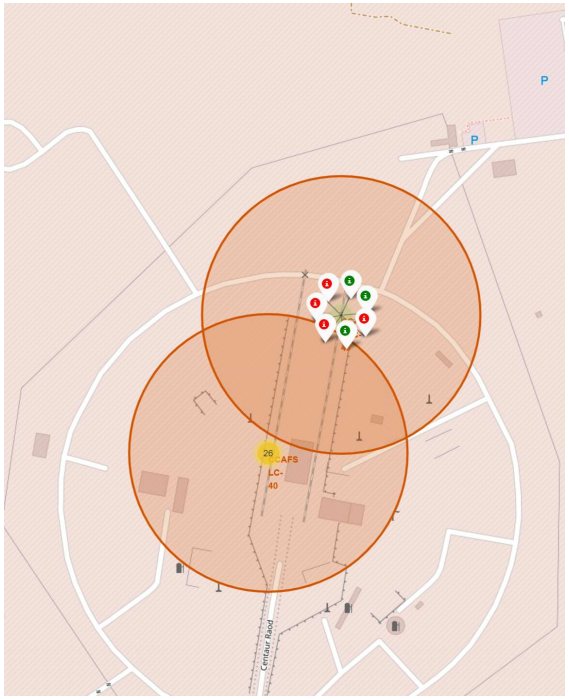
Section 3

Launch Sites Proximities Analysis

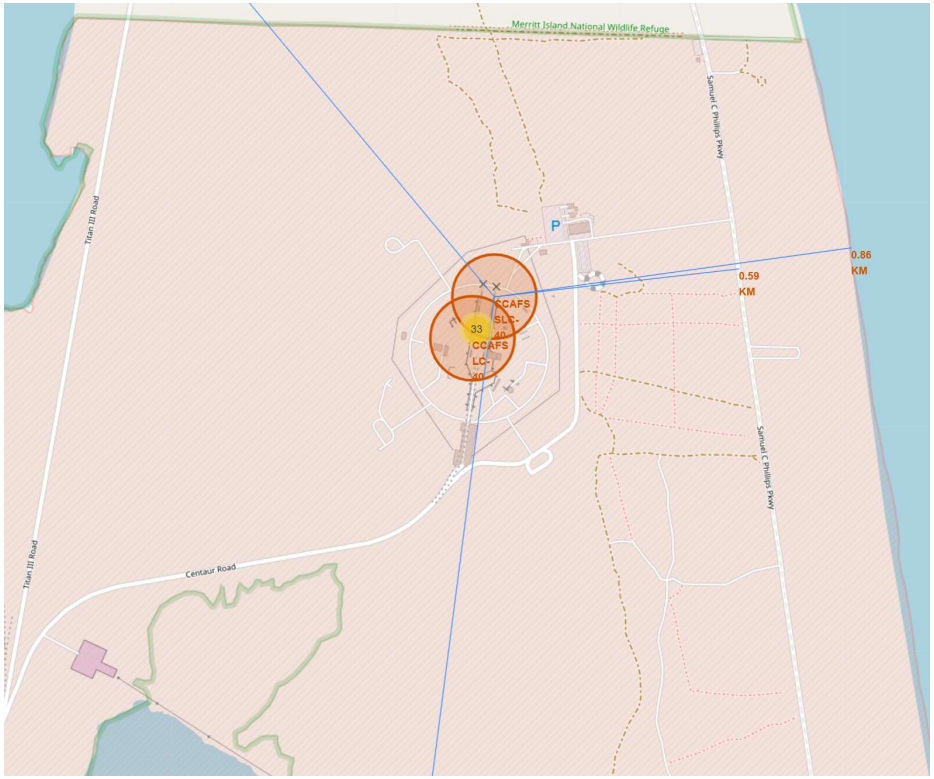
All Launch Sites on Map

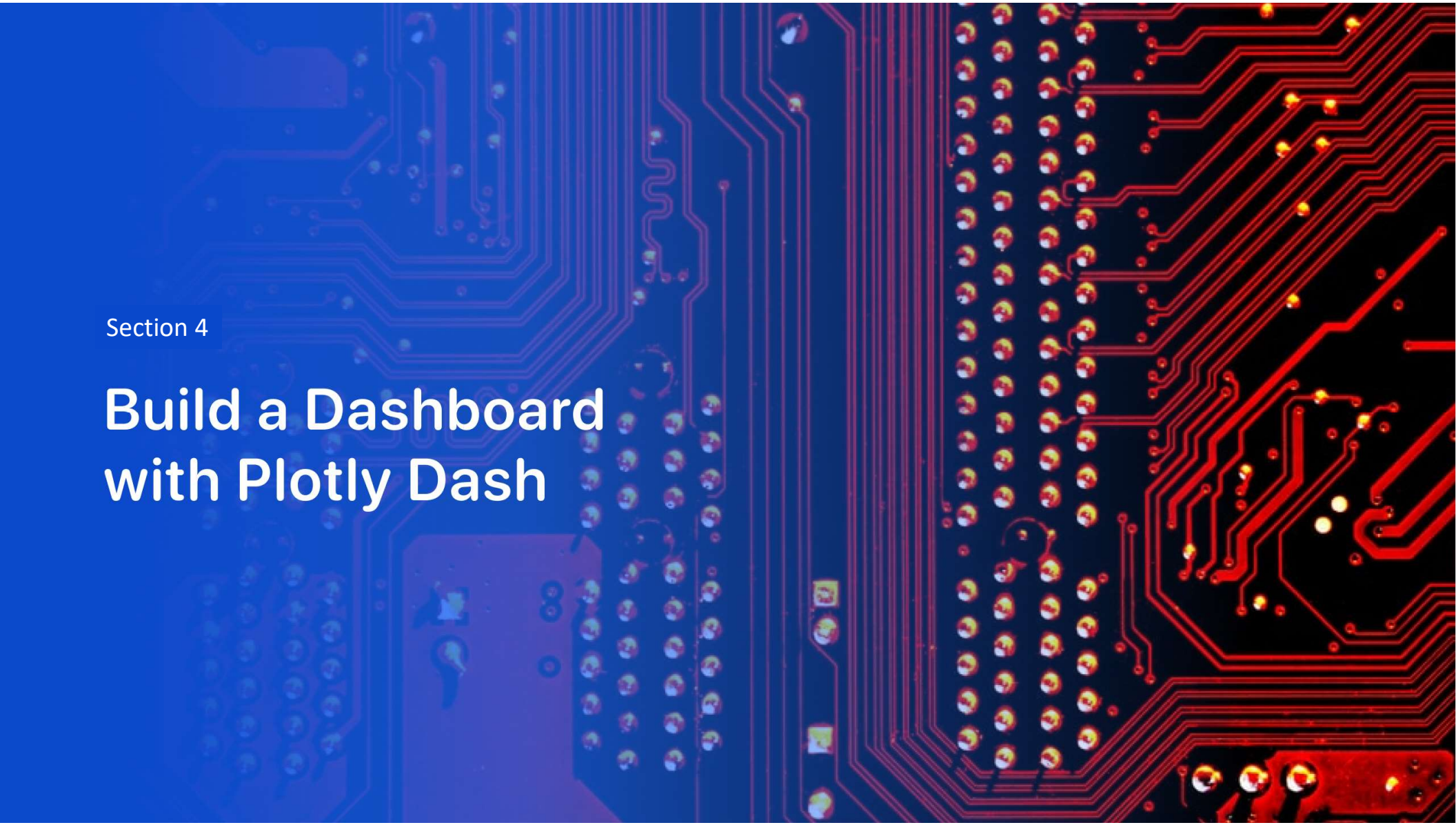


Launch Outcomes on Map



Launch Site Feature Proximities

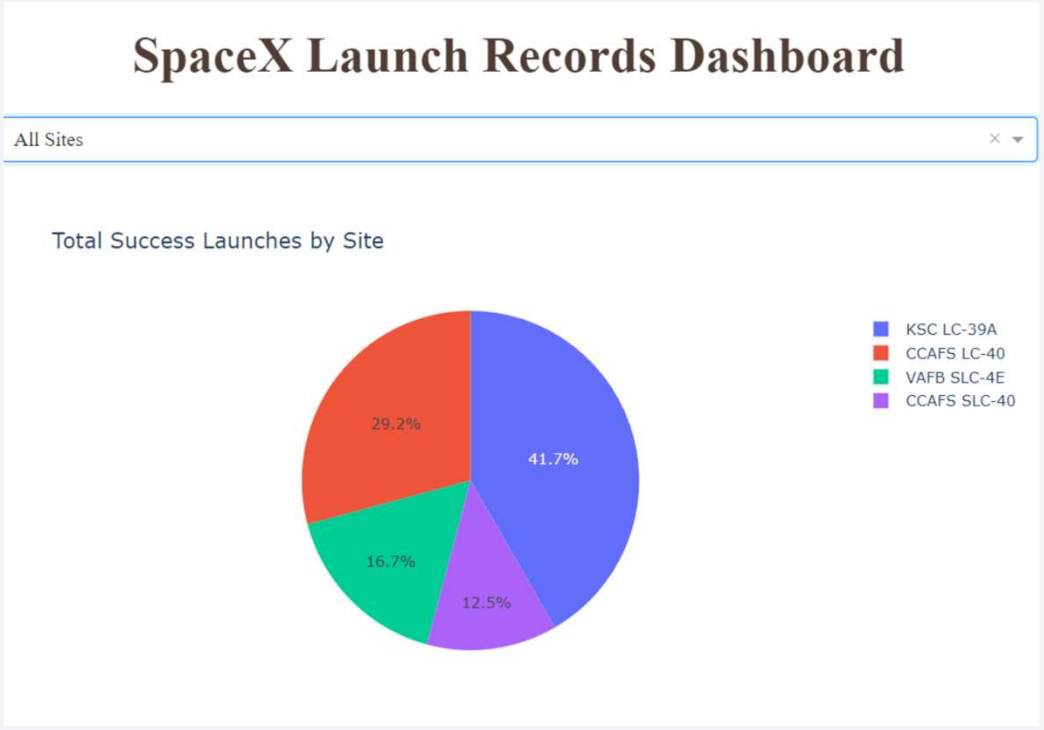




Section 4

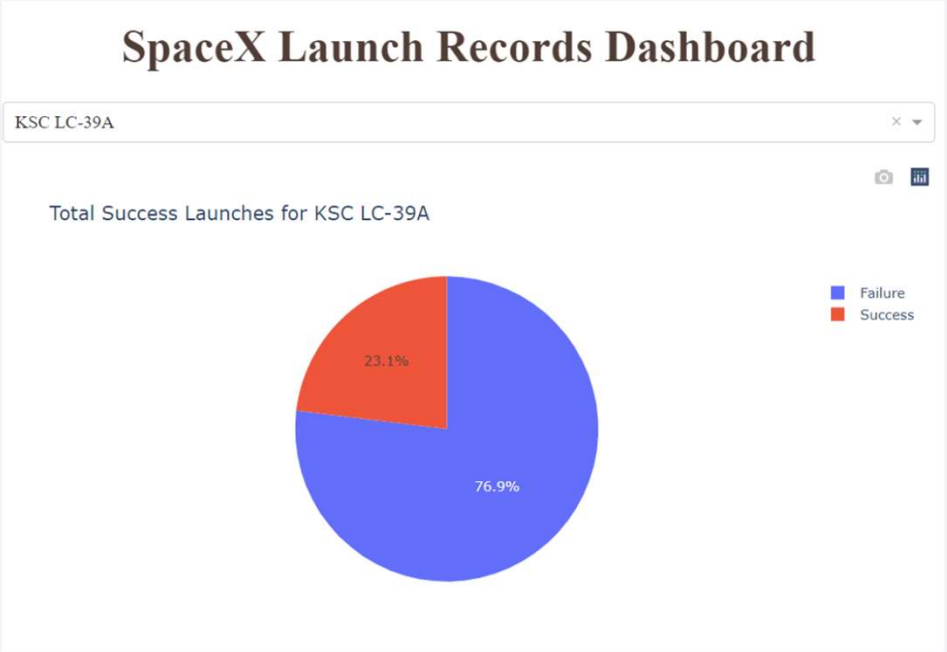
Build a Dashboard with Plotly Dash

Total Success Launches by Site



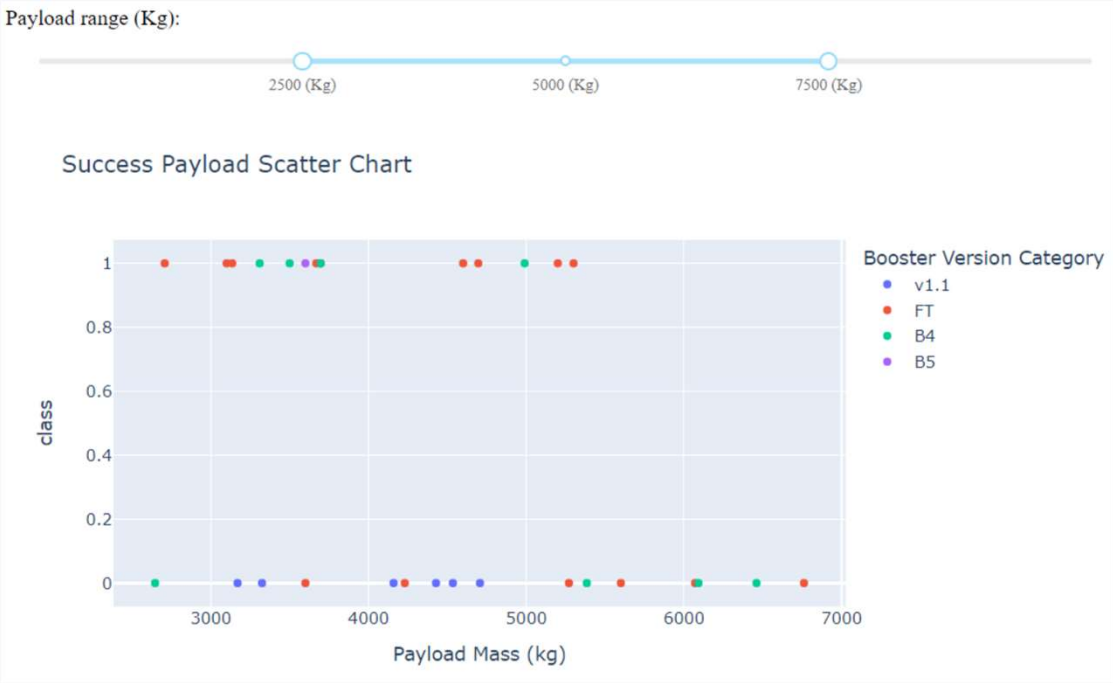
Distribution of successful landings across all launch sites

Highest Success Rate Launch Site



KSC LC-39A has the highest success rate with 10 successful landings and 3 failed landings.

Landing Success vs. Payload



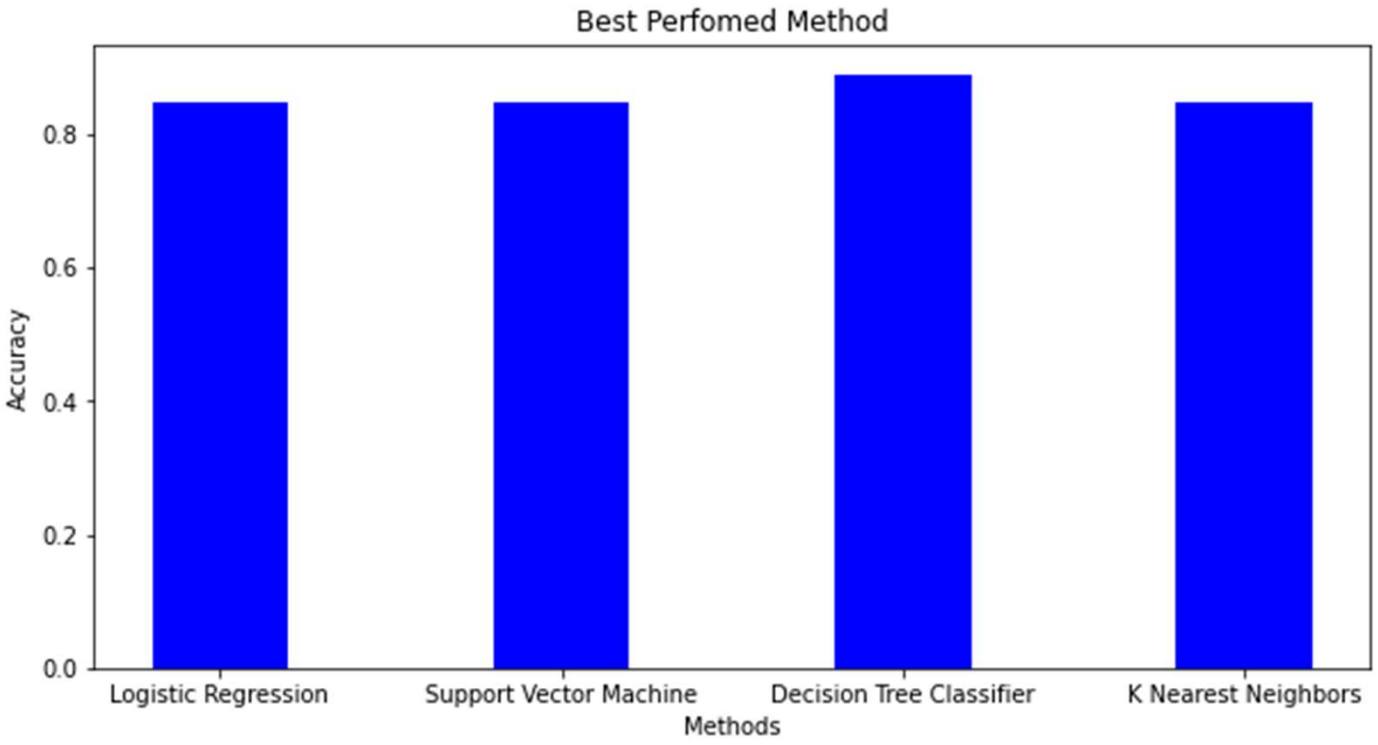
Landing Success vs. Payload scatter plot is for all launches. Range slider was selected for 2500-7500 kg.

Class indicates successful landing with 1 and failure with 0.

Section 5

Predictive Analysis (Classification)

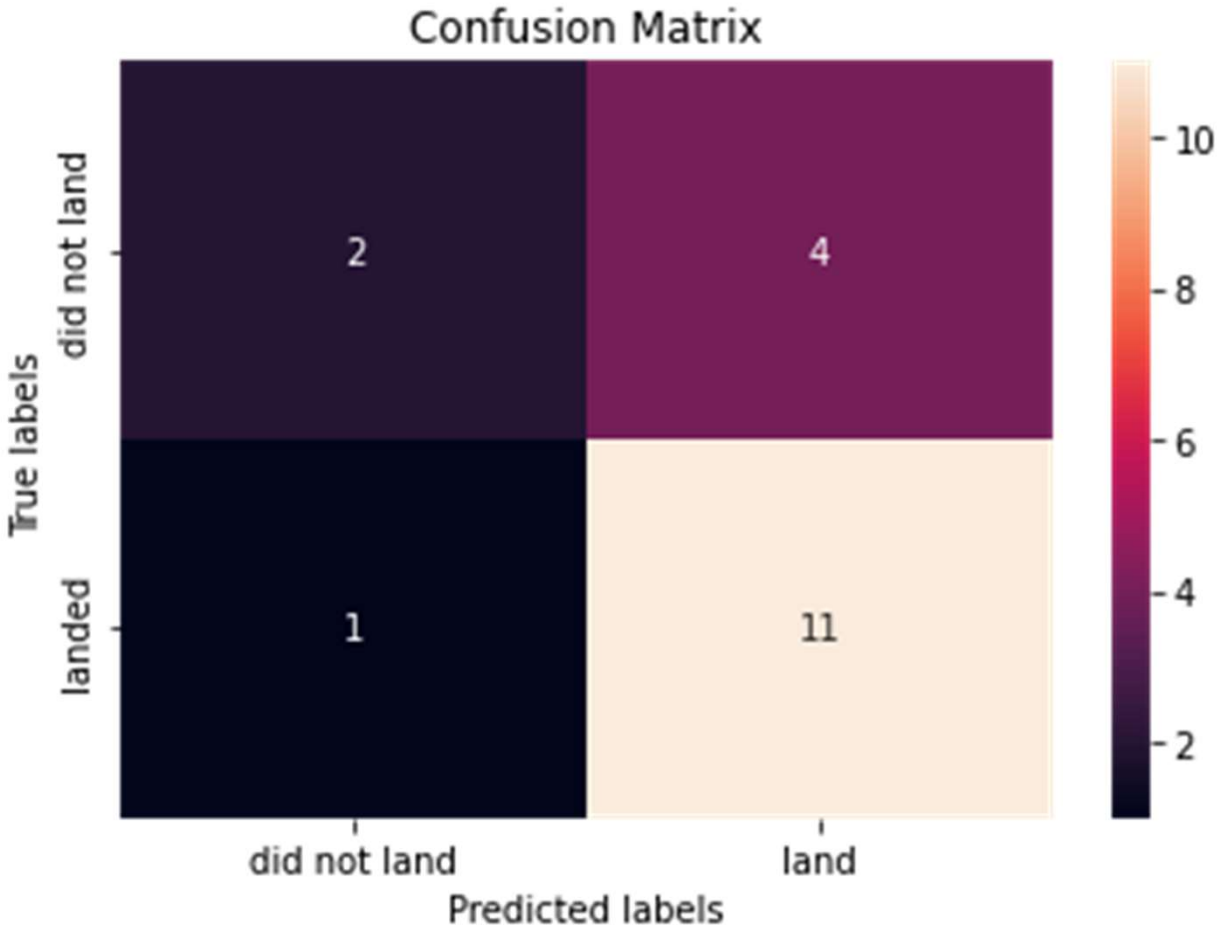
Classification Accuracy



Decision Tree model has the highest classification accuracy

88.9%

Confusion Matrix



Conclusions

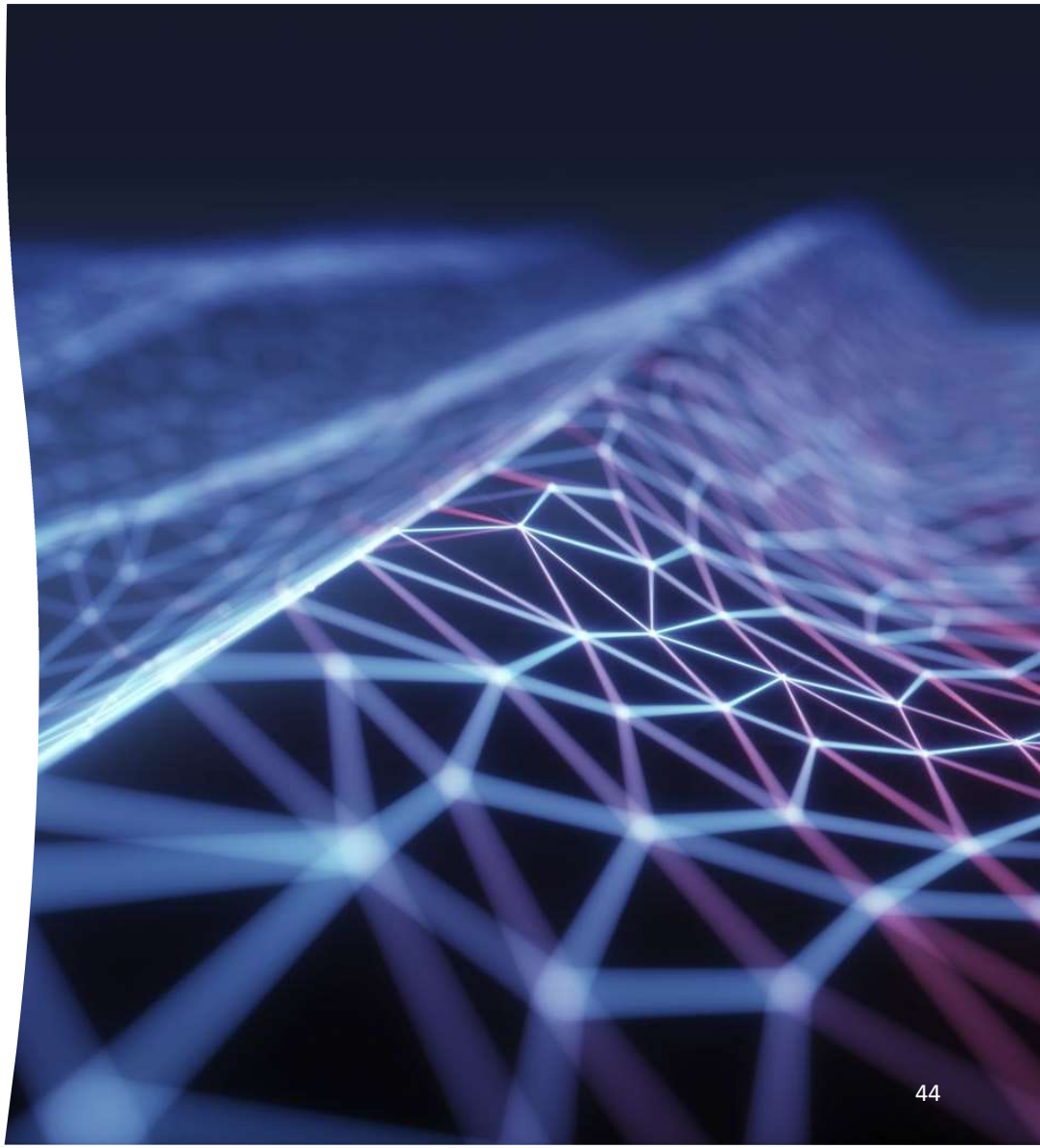
CCAFS LC-40, has a success rate of 60 %

KSC LC-39A and VAFB SLC 4E has a success rate of 77%.

Average Payload: 2928 kg

KSC LC-39A has the highest success rate with 10 successful landings and 3 failed landings

Decision Tree model has the highest classification accuracy 88.9%



Appendix

GitHub Repository URL:

https://github.com/DiscoveryUnlimited/IBM-Applied-Data-Science-Capstone

Thank you