IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Hua Lung Jao
December 31, 2023

Outline

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

Executive Summary

- Summary of methodologies

  Data Collection (API & Webscrapping)

  Data Wrangling

  Data Analysis (SQL, Visualization, Interactive)

  Prediction with Machine Learning

- Summary of all results

  - Exploratory Data
  - Screenshots of Interactive Analytics
  - Predictive Analytics Finding

Introduction

- Project background and context
  - Space X's Falcon 9 rocket costs 62 million dollars with each launch, while other providers cost upward of 165 million dollars. Much of this savings is a result of Space X reusing the first stage.
  - If an alternate company can predict whether the first stage will land successfully, we can determine the cost of a launch, and therefore bid against space X for a rocket launch.
- Problems you want to find answers
  - Factors which determine successful landing of the Falcon 9 rockets

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - API Provided by SpaceX
  - Webscrapping (wiki)
- Perform data wrangling
  - Data cleaned up and re-labeled landing outcome with success/failure category
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Evaluated different models (logistic regression, support vector machine, decision tree, k-nearest neighbor) using training and compared accuarcy

Data Collection

- Request: SpaceX API
- https://api.spacexdata.com/v4/launches/past

Data Collection – SpaceX API

- Data Flow:
  - Get – Response (REST API)
  - Format Raw Data in JSON
  - Normalize Data
  - Output to CSV
  - Ready for Wrangling

- Link To Source Code

Data Collection - Scraping

- Data Flow:
    - Beautiful Soup API
    - Format HTML data to Data Frame
- Link to Source Code

Data Wrangling

- Analyzed for Missing Values and Data Types
- Calculated # of Launches per Site, and per destination (orbit)
- Calculated # of occurence of mission outcome of the orbits
- Created landing outcome label from Outcome column

- Liink To Source Code

EDA with Data Visualization

- Charts Used:
    - Cat Plot: Visualizing Correlation between 2 variables
    - Scatter Plot: Visualizing Relationship between 2 variables
    - Bar Chart: Compare rate of success within one feature
    - Line Chart: Looking at historic trend
- edwardjao-Coursera_DSC10/module 2 - jupyter-labs-eda-dataviz.ipynb at main · edwardjao/edwardjao-Coursera_DSC10 (github.com)

EDA with SQL

- Created Table
- Display the total payload mass carried by boosters launched by NASA (CRS)
- Display average payload mass carried by booster version F9 v1.1
- 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 names of the booster_versions which have carried the maximum payload mass.
- 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
- Rank 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.
- edwardjao-Coursera_DSC10/module 2 - jupyter-labs-eda-sql-coursera_sqllite.ipynb at main · edwardjao/edwardjao-Coursera_DSC10 (github.com)

Build an Interactive Map with Folium

- Circle - Marked all launch sites on a map
- Marker - Marked the success/failed launches for each site on the map
- Distance Line – Marked the distances between a launch site to its proximity
- edwardjao-Coursera_DSC10/module_3_lab_jupyter_launch_site_location.jupyterlite.ipynb at main · edwardjao/edwardjao-Coursera_DSC10 (github.com)

Build a Dashboard with Plotly Dash

- Added Launch Site Drop-down Input Component
- Added callback function to render success-pie-chart based on selected site dropdown
- Added Range Slider to Select Payload
- Added callback function to render the success-payload-scatter-chart scatter plot
- [edwardjao-Coursera_DSC10/module_3_lab_theia_plotly_dash.md.py at main · edwardjao/edwardjao-Coursera_DSC10 (github.com)]

Predictive Analysis (Classification)

- Performed exploratory Data Analysis and determine Training Labels
  - Created a column for the class
  - Standardized the data
  - Split into training data and test data
- Found best Hyperparameter for SVM, Classification Trees and Logistic Regression
- Found the method performs best using test data
- edwardjao-Coursera_DSC10/module_4_SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb at main · edwardjao/edwardjao-Coursera_DSC10 (github.com)

Results

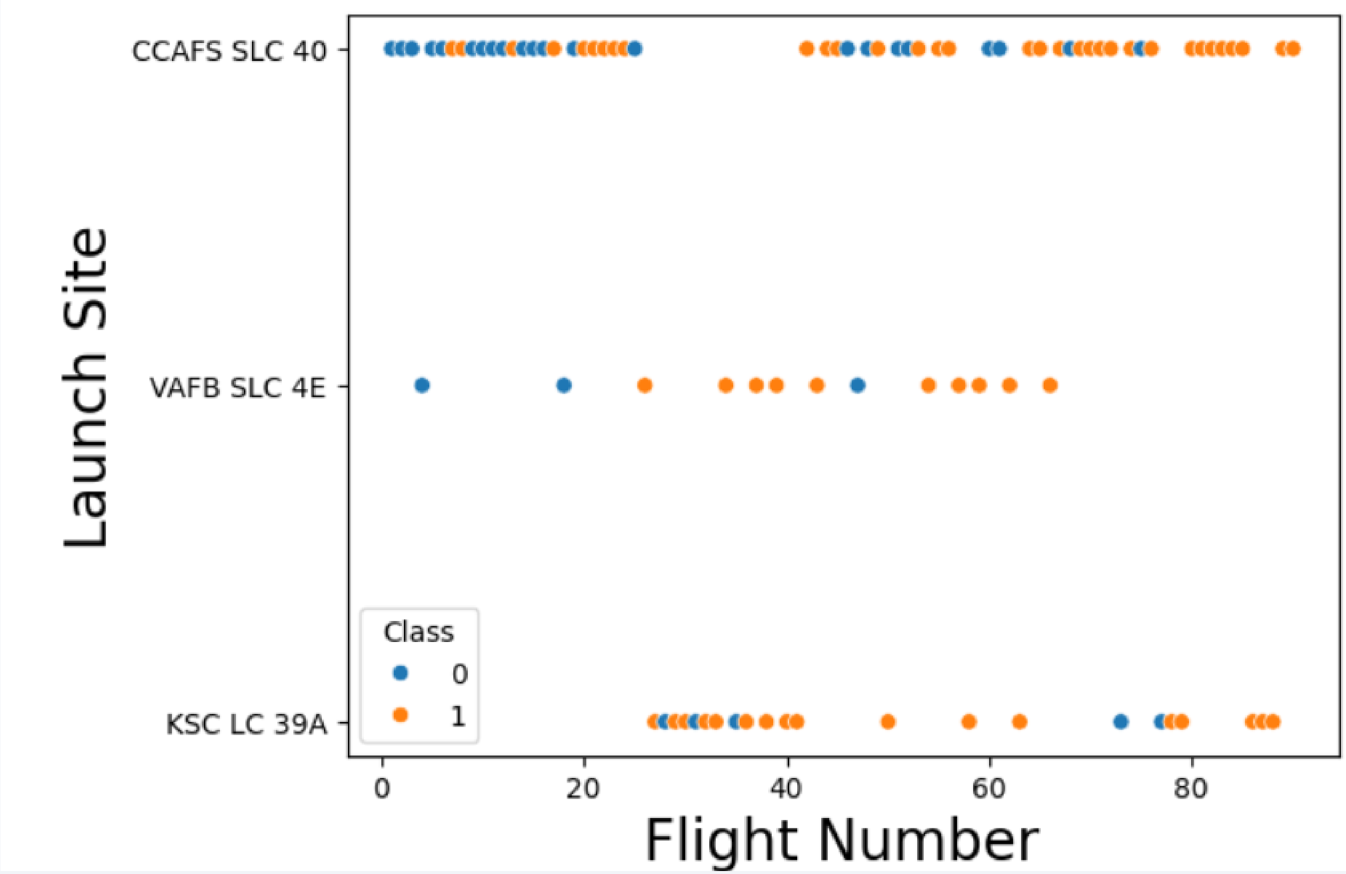
- Exploratory data analysis
  - SpaceX had increased success rate of landing from 2015 to 2020
  - Falcon 9 rockets is the most reliable booster version
  - Average payload is around 3,000 kg
- Interactive Screenshots on the right
- Predictive analysis
  - Decision Tree is the best predictor, with almost 90% accuracy
  - Best parameters are determined for each predictive model

Section 2

Insights drawn from EDA

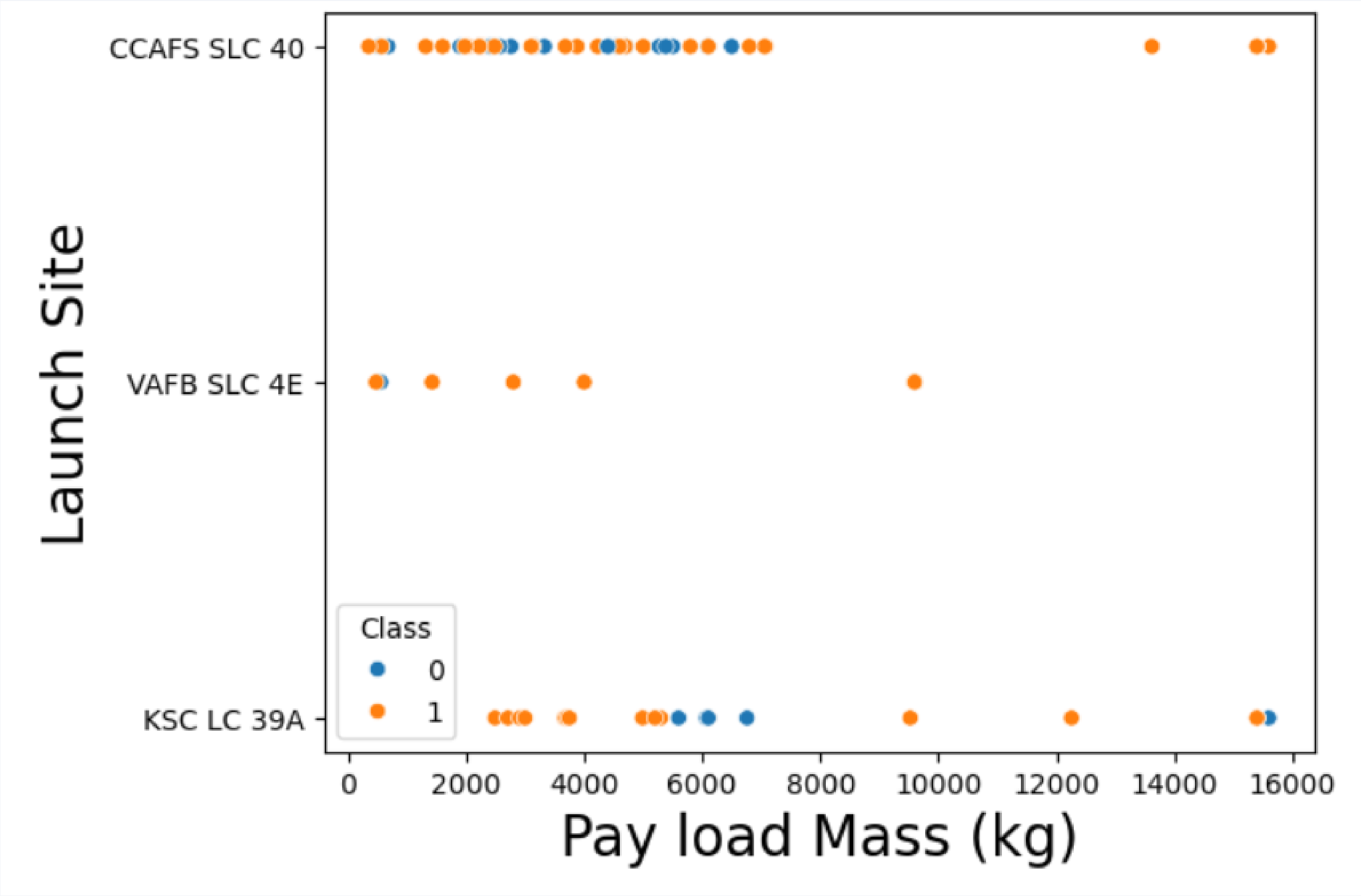
Flight Number vs. Launch Site

- SLC 40 is the most used launch site, with increased success over experience.



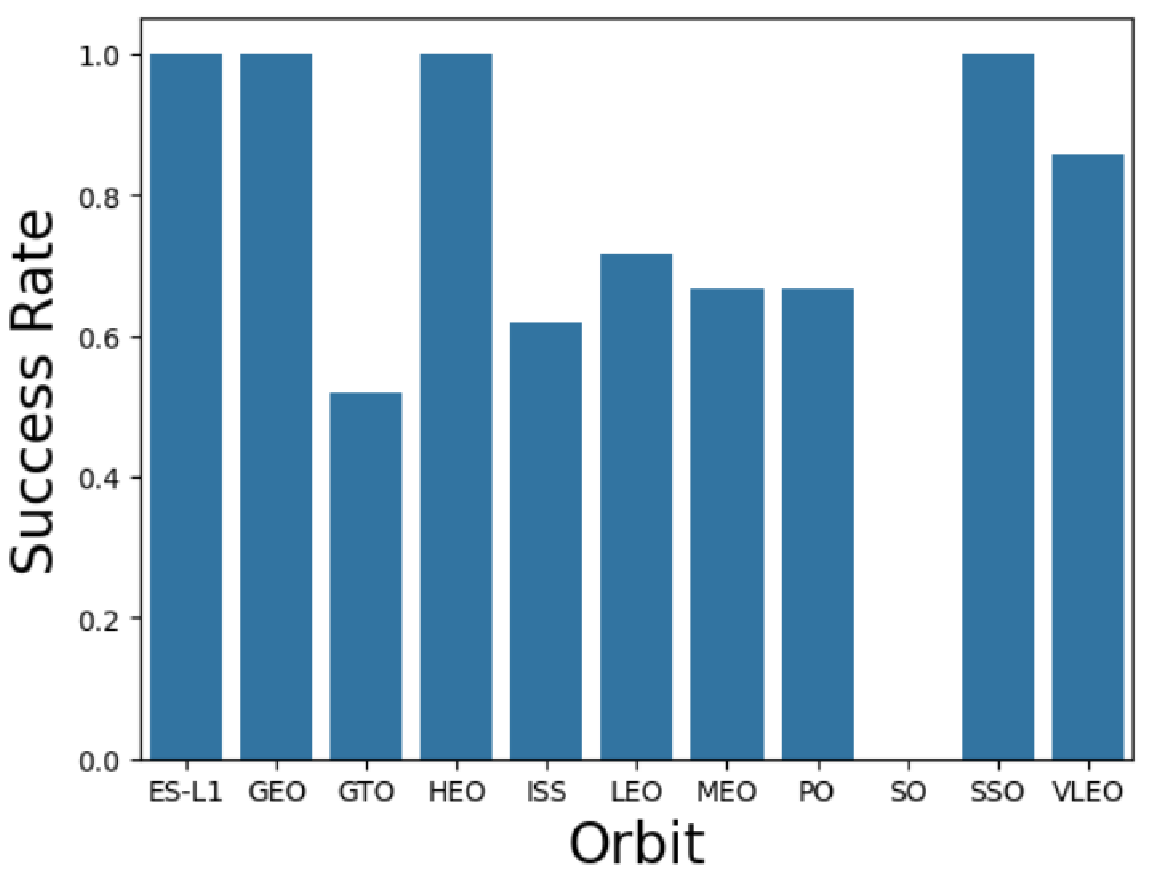
Payload vs. Launch Site

- Most payloads are under 6000 kg



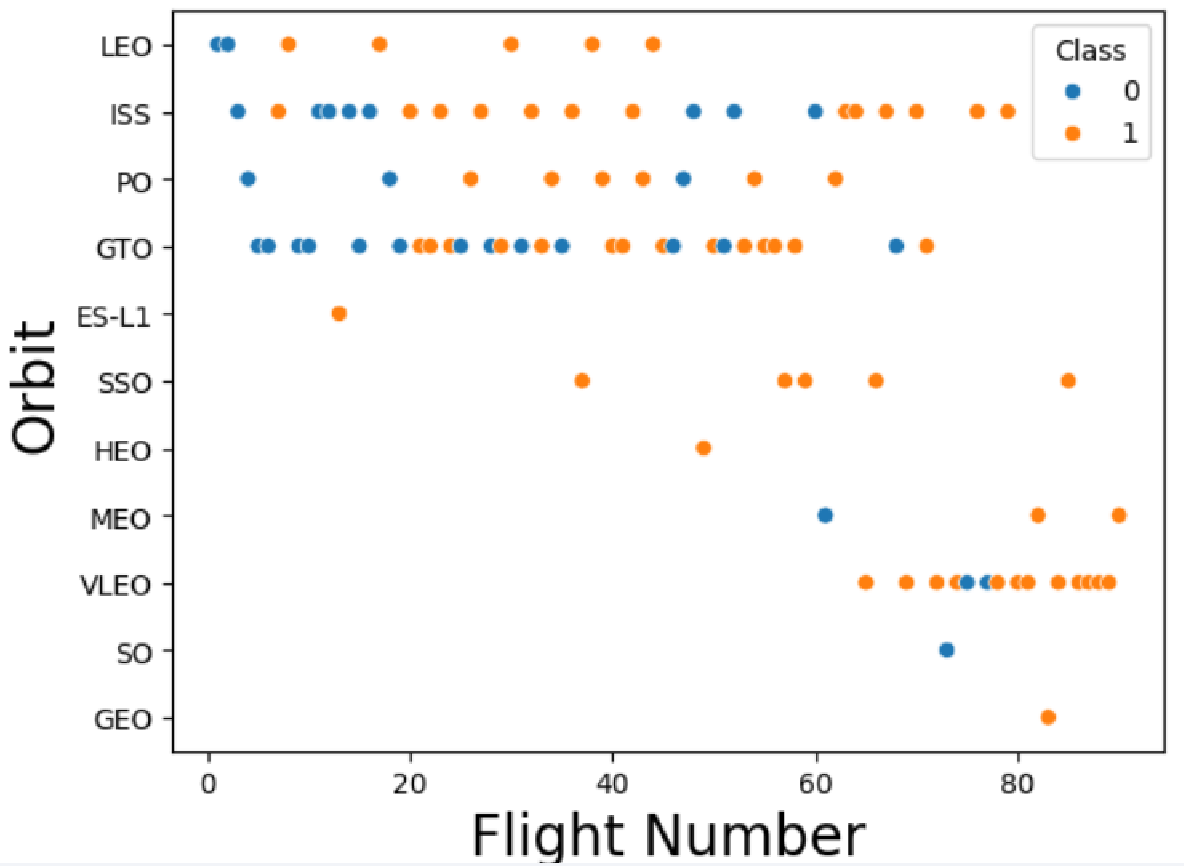
Success Rate vs. Orbit Type

- Certain Orbits has perfect success rate



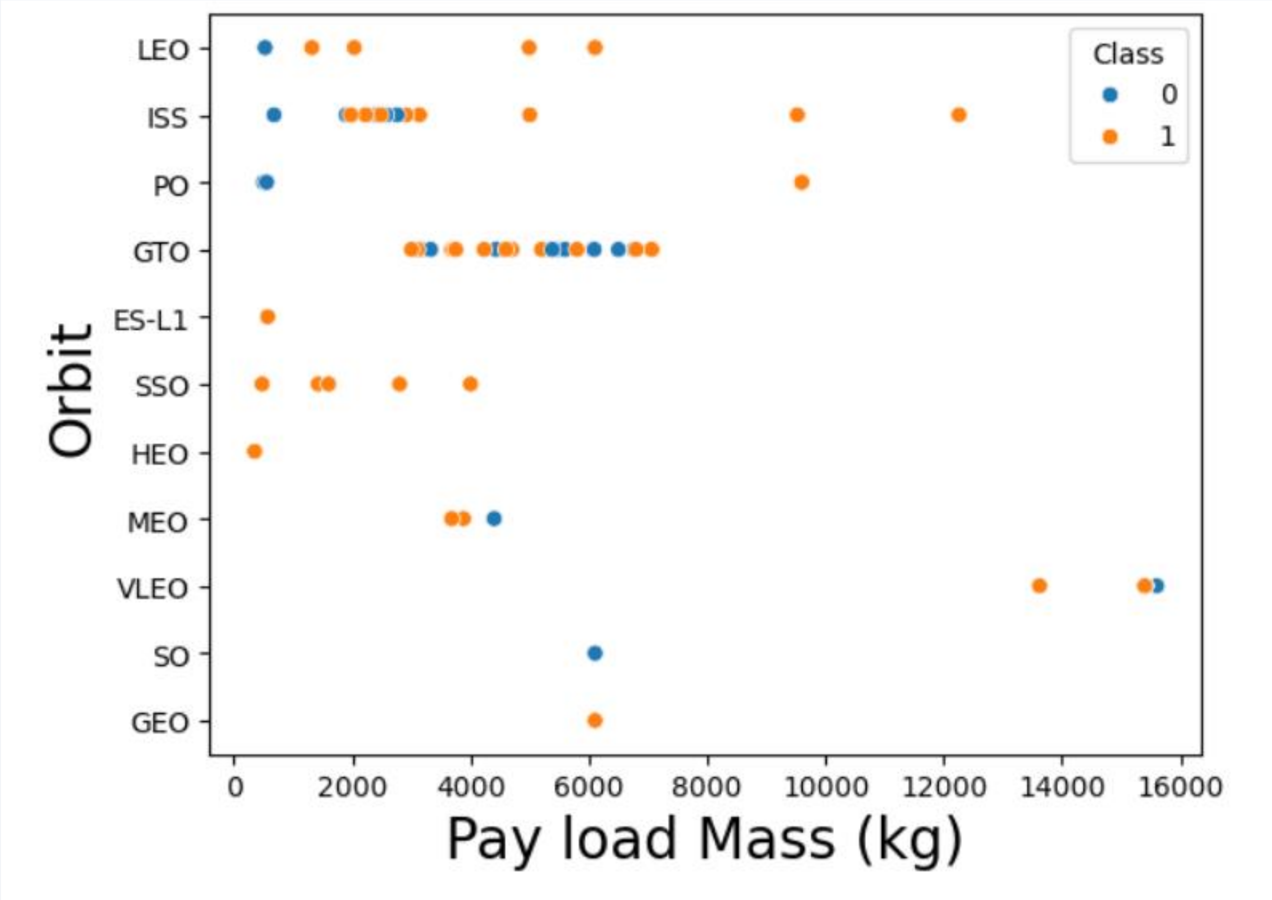
Flight Number vs. Orbit Type

- There's recent interest in VLEO orbit



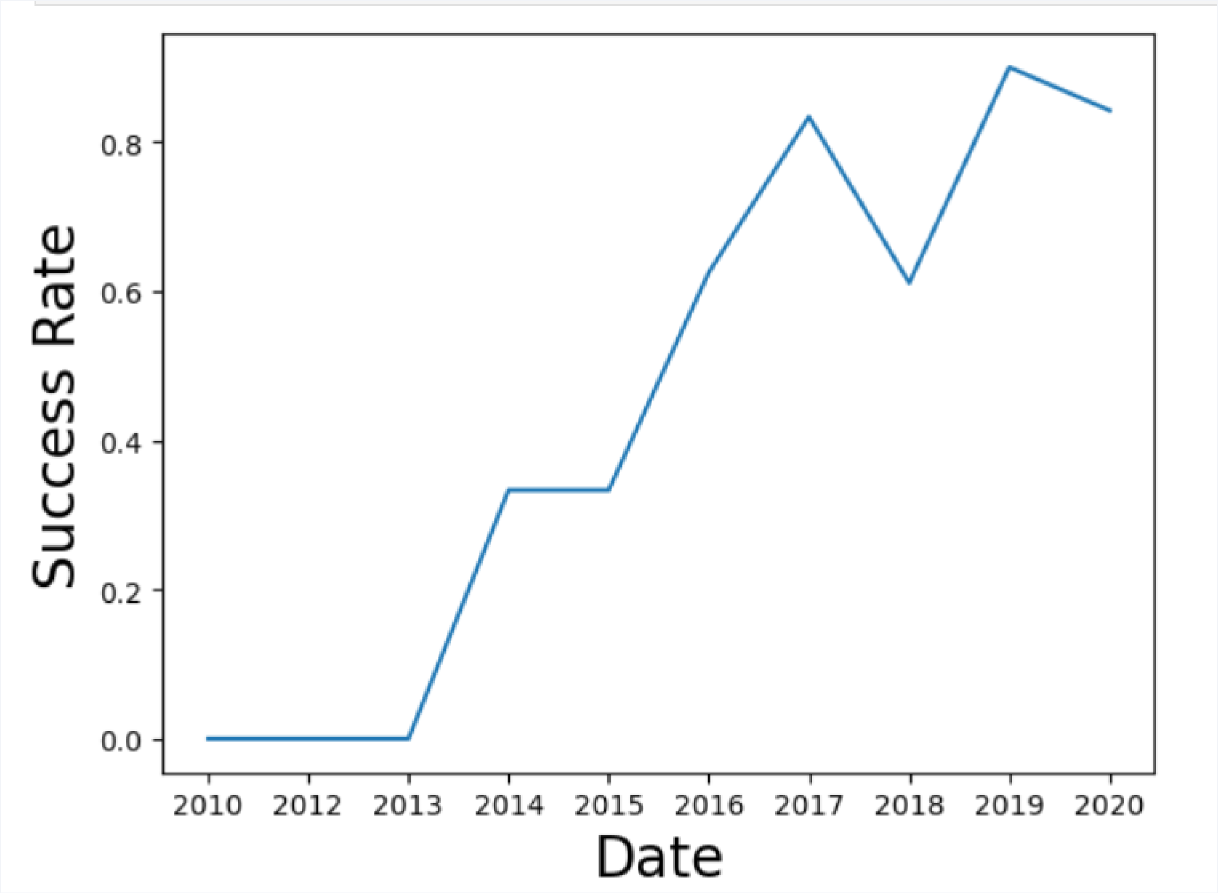
Payload vs. Orbit Type

- GTO contains the bulk of normal payloads



Launch Success Yearly Trend

- Success rate has increased over the years, and may be plateau'd



All Launch Site Names

- There are 4 unique launch sites

Task 1

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

```
%sql select distinct Launch_Site from SPACEXTBL
```

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

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

Launch Site Names Begin with 'CCA'

```
%sql select * from SPACEXTBL 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

Total Payload Mass

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

```
%sql select sum(PAYLOAD_MASS__KG_) from SPACEXTBL
```

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

sum(PAYLOAD_MASS__KG_)
619967

Average Payload Mass by F9 v1.1

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

```
%sql select avg(PAYLOAD_MASS__KG_) from SPACEXTBL where Booster_Version like 'F9 v1.1%'
```

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

avg(PAYLOAD_MASS__KG_)
2534.6666666666665

First Successful Ground Landing Date

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

Hint:Use min function

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

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

min(Date)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

Task 6

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 distinct Booster_Version from SPACEXTBL where Landing_Outcome = 'Success (drone ship)' and PAYLOAD_MASS__
```

```
 * 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

- SpaceX considers most missions "successful"

List the total number of successful and failure mission outcomes

```
%sql select Mission_Outcome, count(Mission_Outcome) from SPACEXTBL group by Mission_Outcome
```

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

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

Boosters Carried Maximum Payload

- Multiple booster versions had carried max payload

Task 8

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

```
%sql select distinct Booster_Version from SPACEXTBL where PAYLOAD_MASS__KG_ = ( select max(
```

```
 * 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

2015 Launch Records

- Two notable landing failures in 2015

```
%sql select substr(Date, 6,2) as monthname, Landing_Outcome, Booster_Version,
```

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

monthname	Landing_Outcome	Booster_Version	Launch_Site
01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

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

```
26]: %sql select Landing_Outcome, count(Landing_Outcome) from SPACEXTBL where Date between '2010-06-04' and '2017-03-20' gr
```

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

26]:

Landing_Outcome	count(Landing_Outcome)
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

Reference Links

Would you like to rec

Section 3

Launch Sites Proximities Analysis

Launch Site Locations

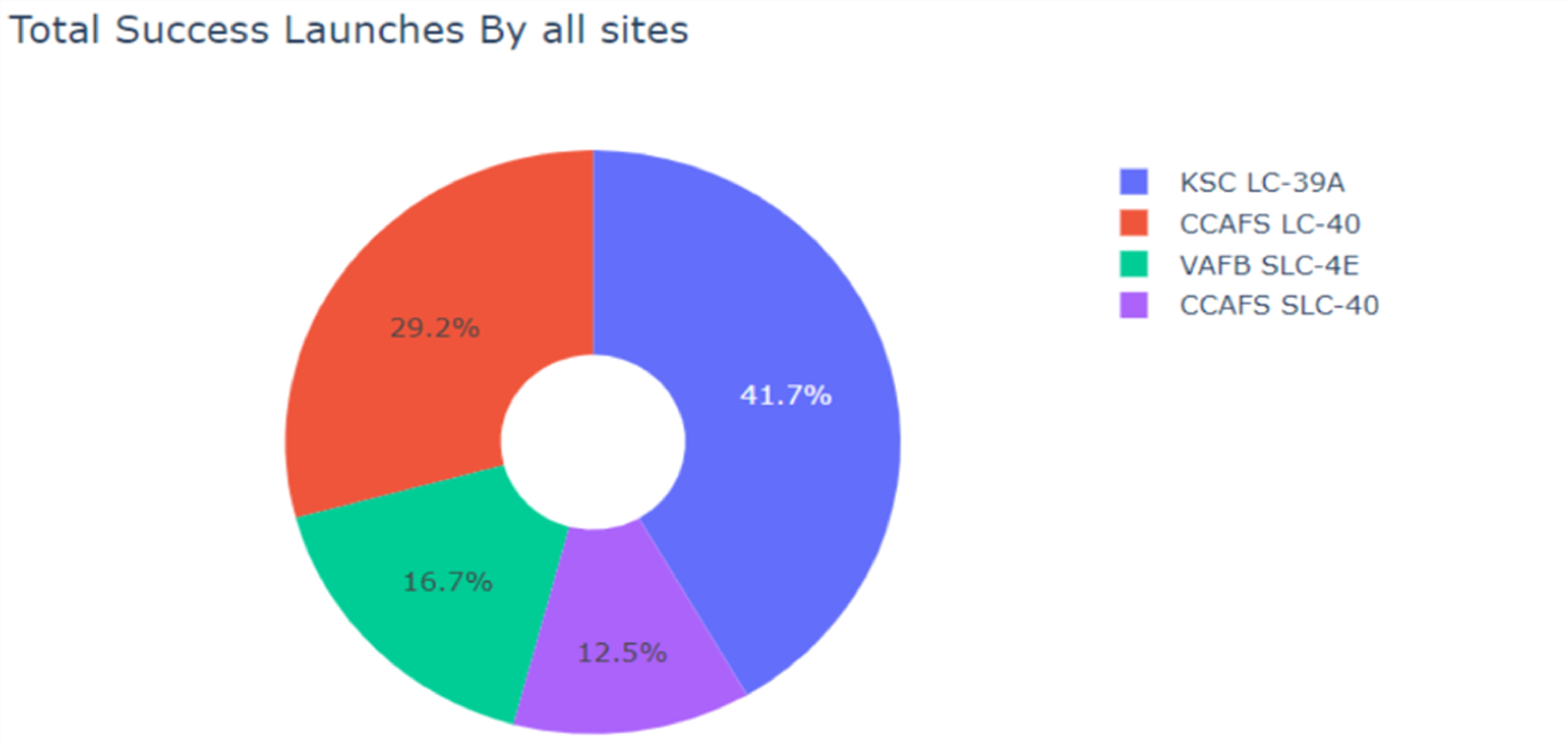
of Launches Per Location

Launch Site Distance From Shore

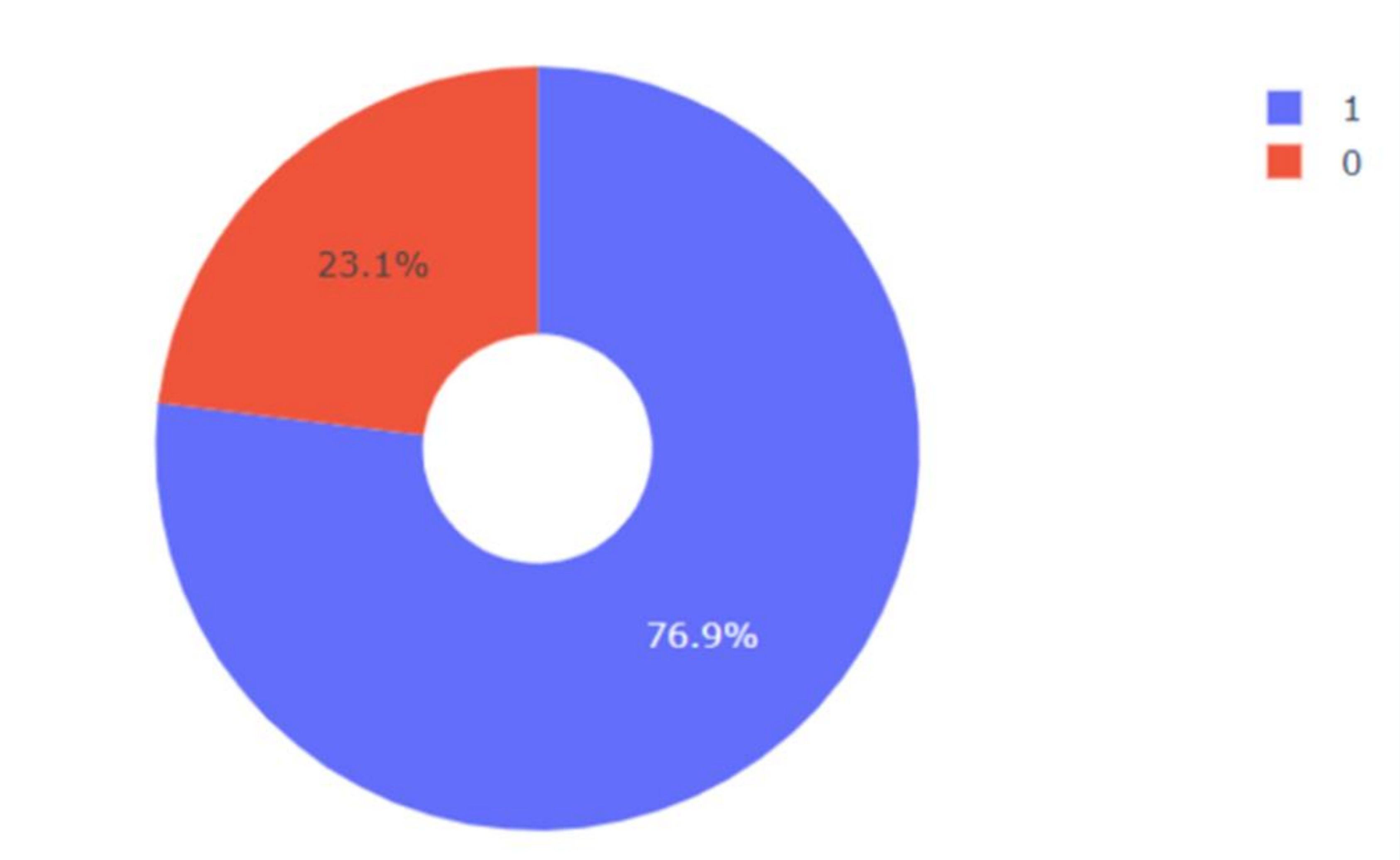
Section 4

Build a Dashboard with Plotly Dash

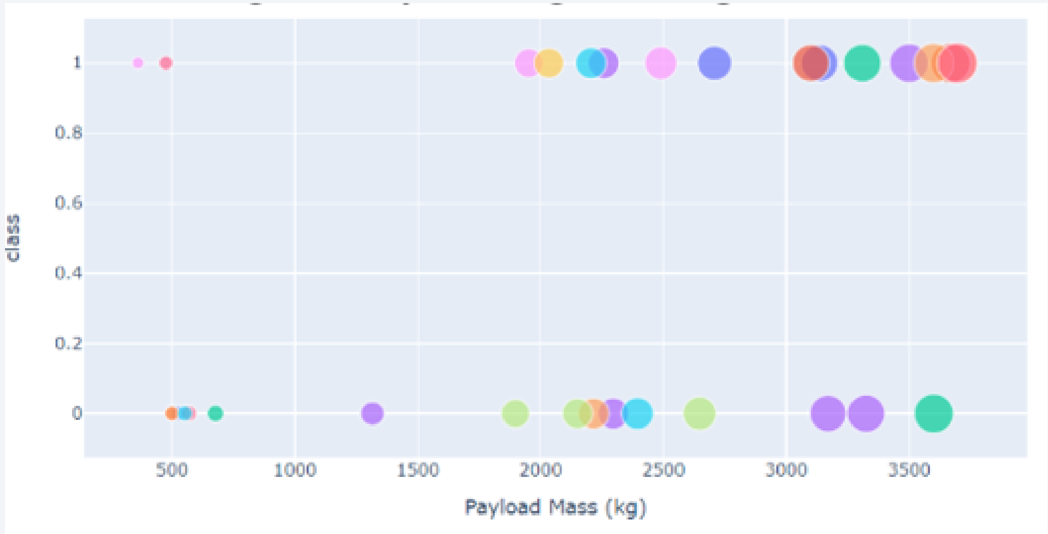
KSC LC-39A Has The Most Success

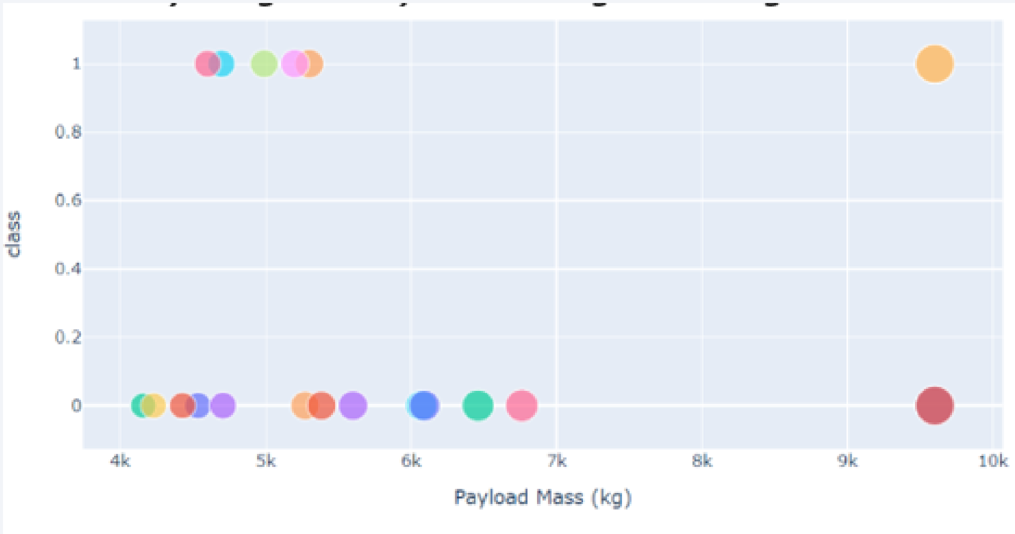


Deep Dive Into KSC LC-39A Success Rate



Lower Weight = More Success





Section 5

Predictive Analysis (Classification)

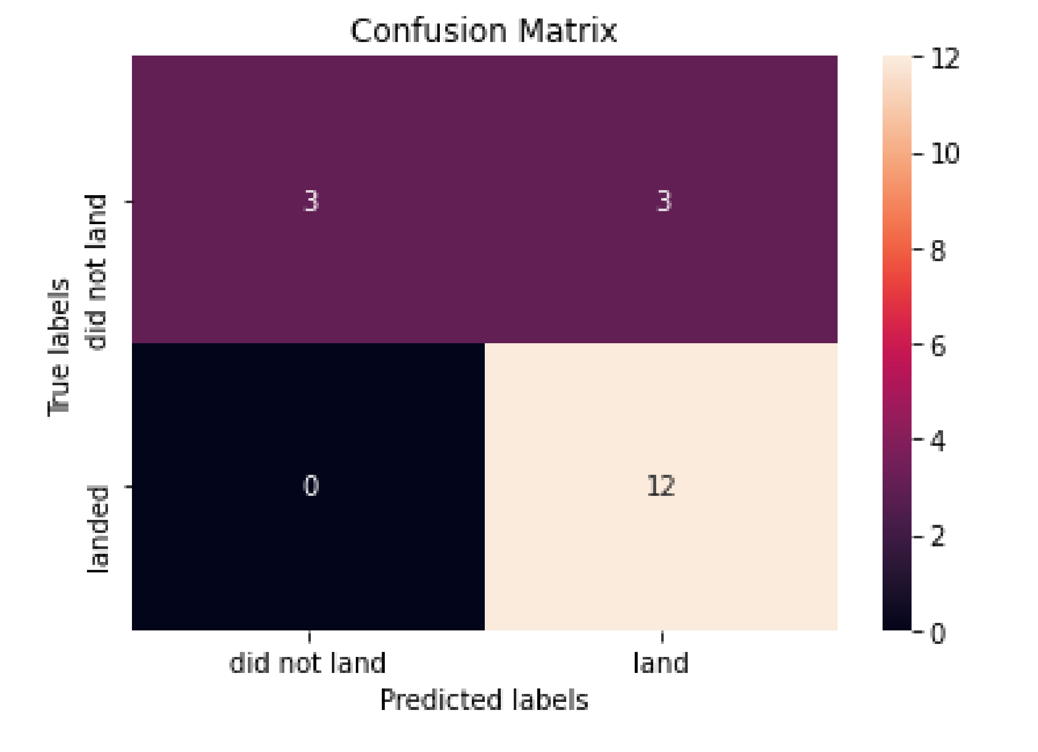
Classification Accuracy

- Visualize the built model accuracy for all built classification models, in a bar chart

- Find which model has the highest classification accuracy

Confusion Matrix – Decision Tree

- Decision tree had the least false-positive prediction



Conclusions

- Certain Launch Sites has better landing rate
- Certain Orbits leads to better success
- Payload contributes to success rate
- Decision Tree can better predict success

Appendix

[Python Data Science Handbook | Python Data Science Handbook (jakevdp.github.io)]

Thank you!