IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

James Daniel Knapp
February 17, 2023

Outline

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

Executive Summary

- Ramjett Rocket Industries has hired me, James Daniel Knapp, Data Scientist, to evaluate the feasibility of expanding into the market space currently dominated by SpaceX.
- SpaceX's enormous cost advantages are realized by re-using the first stage rocket.
- Accurately predicting SpaceX success rates in re-using the first stage rocket is paramount to providing Ramjett Rocket with competent advice.
- Methodologies Utilized to Determine SpaceX First Stage Rocket Re-Use Prediction
    - Data Collection
    - Data Wrangling
    - Exploratory Data Analysis
    - Interactive Visual Analytics
    - Dashboard
    - Predictive Analysis

Executive Summary

- SpaceX has cut costs due to re-use of the stage 1 rocket, making their space missions much cheaper than the competition..
- SpaceX has experienced success and has improved the program over the years. Predictive analysis seems to show they will attain greater success in the future.
- Competing with SpaceX at this point in time, without a similar program, similar cost saving technology, and similar results will result in loss of revenue.

Introduction

- SpaceX rocket launches cost far less than competitors (\$62 million versus competitors >= \$165 million)
- Risk for competitors to enter the market are high if SpaceX can re-use the stage one rocket
- Analysis Objectives
  - Predict SpaceX probability of landing first stage rocket successfully
  - Access launch site success rates to determine the best launch site
  - Factor in other variables such as payload weight, booster models, orbit, etc., to determine impact upon landing outcomes

Section 1

Methodology

Executive Summary

- Data collection methodology:
  - Utilize "Get request" to SpaceX API "https://api.spacexdata.com/v4"
  - Collect rocket, launchpad, cores, and payload data
  - Web scrape Falcon 9 launch records with BeautifulSoup
  - Goal
    - From Rockets get booster name
    - From Payload get payload mass and orbit
    - From Launchpad get launch site lattitude and longitude
    - From Cores get landing outcome plus additional info

Executive Summary

- Perform data wrangling
    - Filter the data to only include Falcon 9 launches
    - Replacing missing values with the mean of the column
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
    - Determine the best classification algorithm

Data Collection SpaceX API

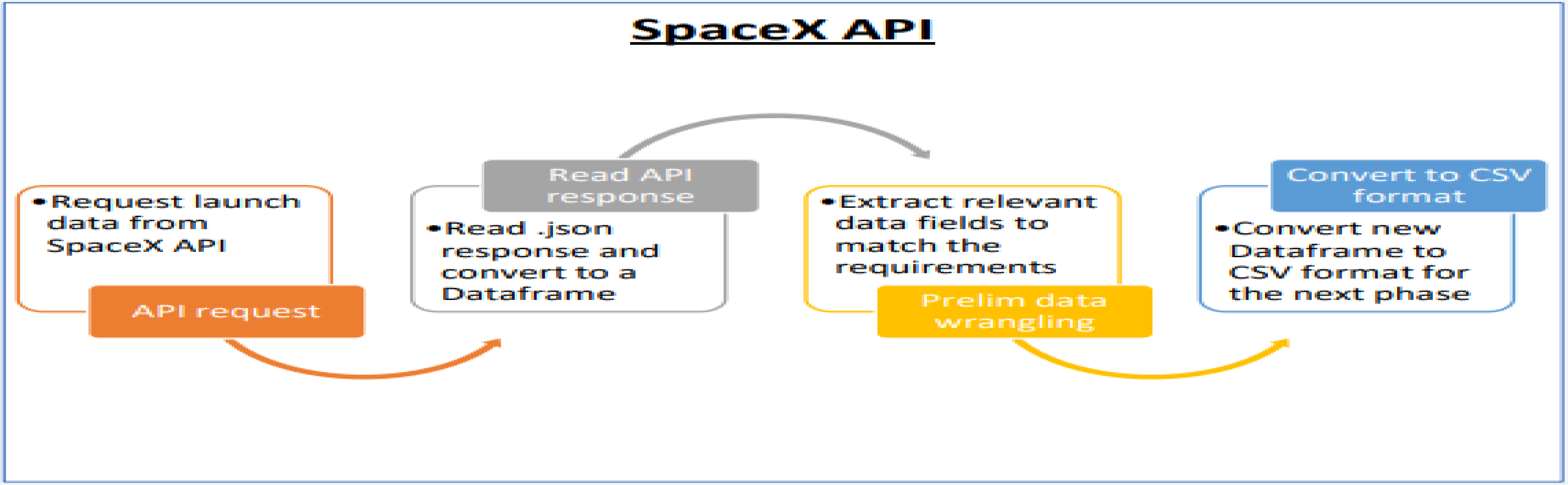
- Data collection was via the get request to the SpaceX API
- Data was collected, cleaned and formatted.
- A dataframe was created.
- Data was wrangled to only include Falcon 9 launches.
- Missing values were replaced with the mean of the column.
- https://github.com/jdjeep57/JamesDanielKnapp-Capstone-Project.git

Data Collection Web Scraping

- Accessed a Falcon 9 Launch HTML page
- Utilized BeautifulSoup to extract columns from the HTML page.
- Created the structure of file and then populated with launch records.
- Converted the launch dictionary to a dataframe.

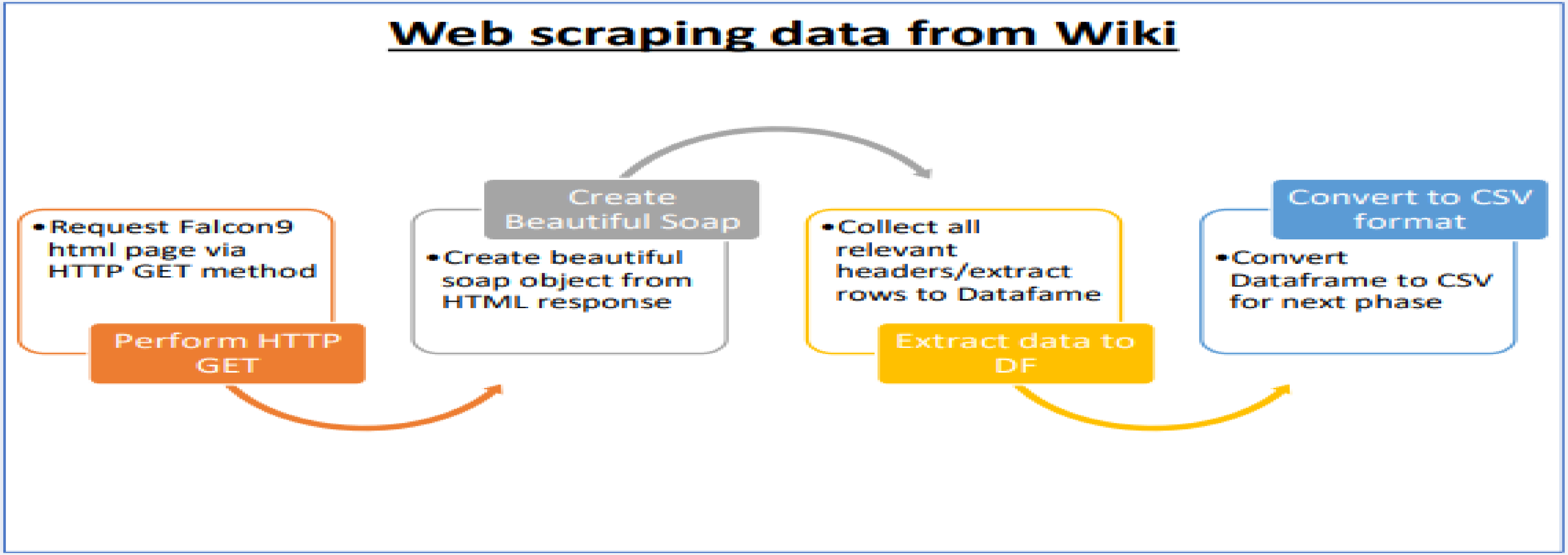
https://github.com/jdjeep57/JamesDanielKnapp-Capstone-Project.git

Data Collection – SpaceX API



- https://github.com/jdjeep57/JamesDanielKnapp-Capstone-Project.git

Data Collection - Scraping



- https://github.com/jdjeep57/JamesDanielKnapp-Capstone-Project.git

Data Wrangling

- Filter the data to only include Falcon 9 launches
- Replace missing PayloadMass values with the mean of the column
- Perform detailed analysis on the data
- https://github.com/jdjeep57/JamesDanielKnapp-Capstone-Project.git

Exploratory Data Analysis with Data Visualization

- Flight Number and Payload variables effect on launch outcome
- Flight Number and Launch Site Relationship
- Payload and Launch Site Relationship
- Orbit Type Success Rate
- Flight Number and Orbit Type Relationship
- Payload and Orbit Type Relationship
- https://github.com/jdjeep57/JamesDanielKnapp-Capstone-Project.git

EDA with SQL

- Display the unique launch site names
- Display 5 records where launch sites are like 'CCA'
- Display the total payload mass carried by boosters
- Display average payload mass carried by booster version F9 v1.1 5.
- Identify first successful landing outcome date
- Identify all boosters which have landed successfully on the drone ship with 4000 <= payload mass < 6000
- Identify total successful and failed missions
- List the names of the booster_versions which have carried the maximum payload mass.
- List the failed landing outcomes
- Rank landing outcomes between 2010-06-04 and 2017-03-20, in descending order
- https://github.com/jdjeep57/JamesDanielKnapp-Capstone-Project.git

Build an Interactive Map with Folium

- Mark all launch sites on the map to visually view launch sites
- Added ‘folium.circle’ and ‘folium.marker’ to highlight each circle area with a text label
- Added a ‘MarkerCluster()’ to visually determine success (green) versus failure (red)
- Calculated distances between a launch site to its proximities (e.g., railroad, highway, population centers)
- Added ‘MousePosition() to get coordinate for a mouse position over a point on the map
- Added ‘folium.Marker()’ to display distance
- Added ‘folium.Polyline()’ to draw a line between the point on the map and the launch site
- Repeated same steps above to add markers and draw lines between launch sites and proximities such as railroad, highway, population centers
- Questions:
- Are launch sites in close proximity to railways? YES •
- Are launch sites in close proximity to highways? YES •
- Are launch sites in close proximity to coastline? YES •
- Do launch sites keep certain distance away from cities? YE
- https://github.com/jdjeep57/JamesDanielKnapp-Capstone-Project.git

Build a Dashboard with Plotly Dash

- Built a Plotly Dash Web Application with Launch drop down, Pie Chart, Range slide for Payload, and Scatter Chart capabilities to answer the following questions:
- Which site has the largest successful launches? KSC LC-39A with 10
- Which site has the highest launch success rate? KSC LC-39A with 76.9% success
- Which payload range(s) has the highest launch success rate? 2000 – 5000 kg Which payload range(s) has the lowest launch success rate? 0-2000 and 5500 - 7000
- Which F9 Booster version (v1.0, v1.1, FT, B4, B5, etc.) has the highest launch success rate? FT
- https://github.com/jdjeep57/JamesDanielKnapp-Capstone-Project.git

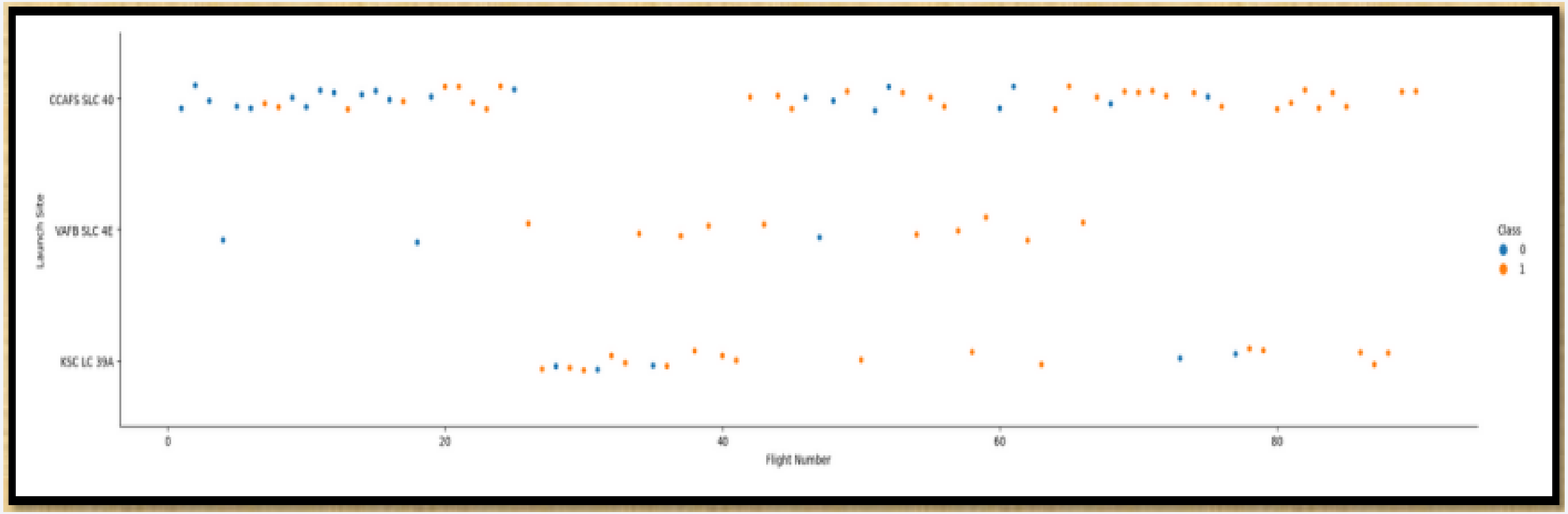
Predictive Analysis (Classification)

- Loaded SpaceX dataset into a DataFrame and created NumPy array
- Standardized the data
- Train/Test/Split data into training and test data sets
- Created and refined Models based on classification algorithms
- Identify the best performing Model
- https://github.com/jdjeep57/JamesDanielKnapp-Capstone-Project.git

Section 2

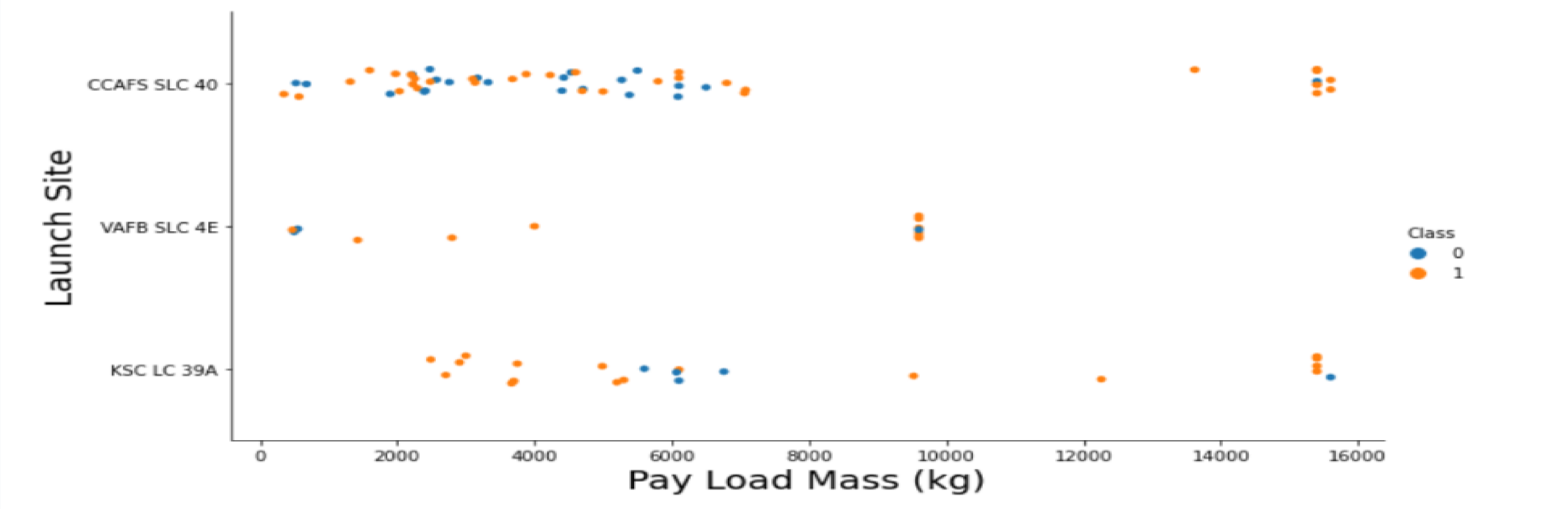
Insights drawn from EDA

Launch Site vs Flight Number



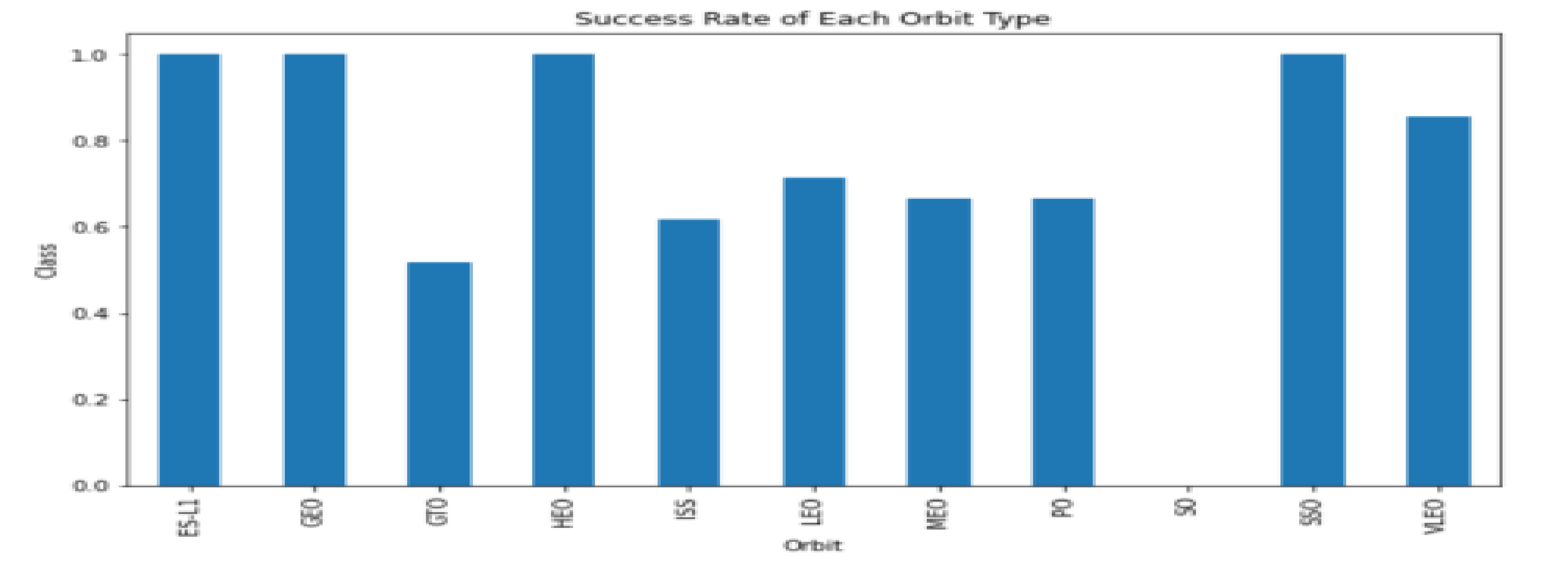
- Launch site ‘KSC LC 39A’h has the highest success rate.

Payload vs. Launch Site



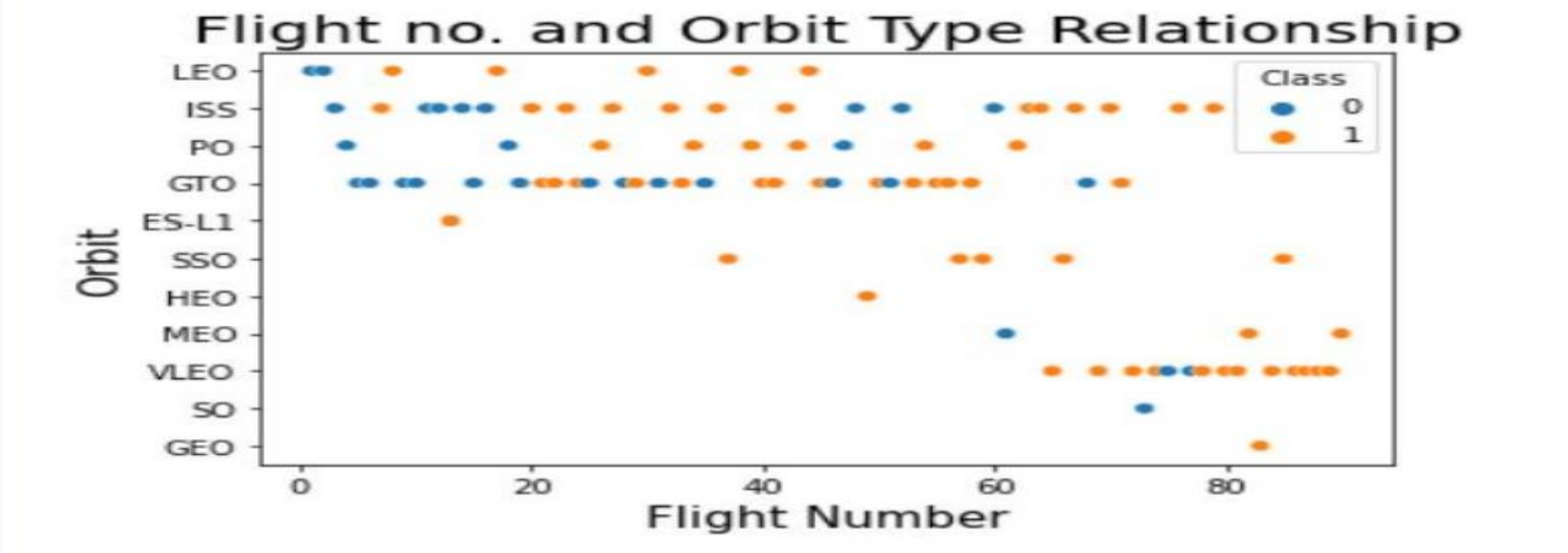
- Launch Site and size of payload do not appear to be correlated.
- Launch Site VAFB SLC 4E has not handled pay load greater than 10,000 kg but has been successful at loads just shy of 10,000 kg

Success Rate vs. Orbit Type



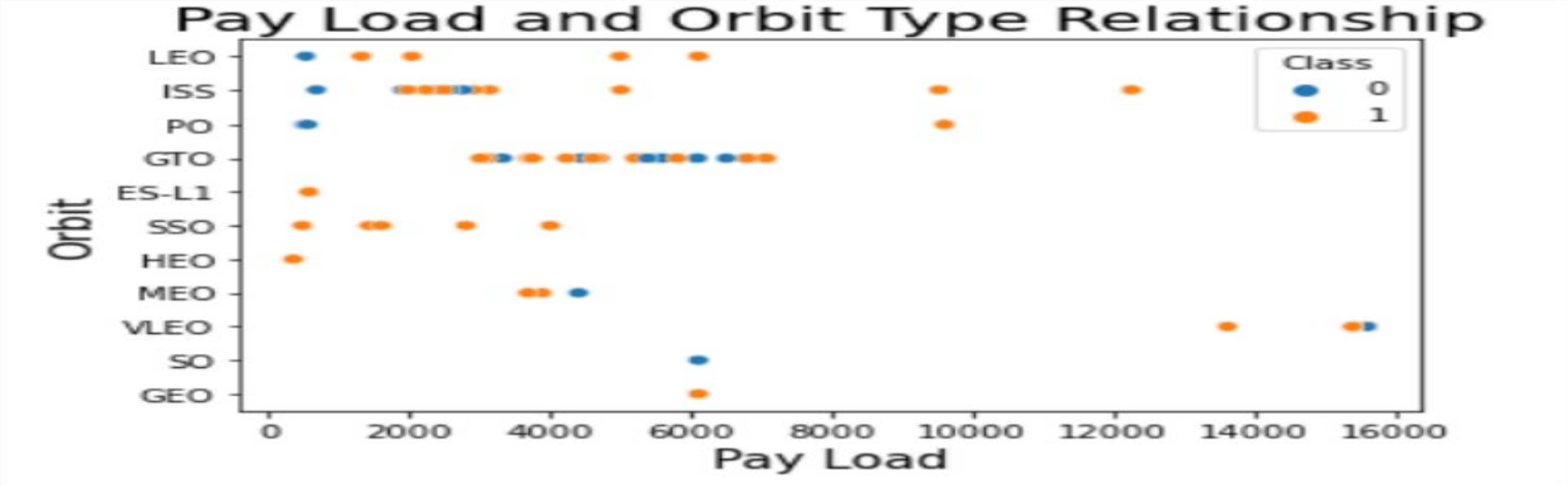
- Orbit type 'SO' has the lowest success rate
- 'ES-LI', 'GEO', 'HEO', and 'SSO' have the highest success rates.

Flight Number vs. Orbit Type



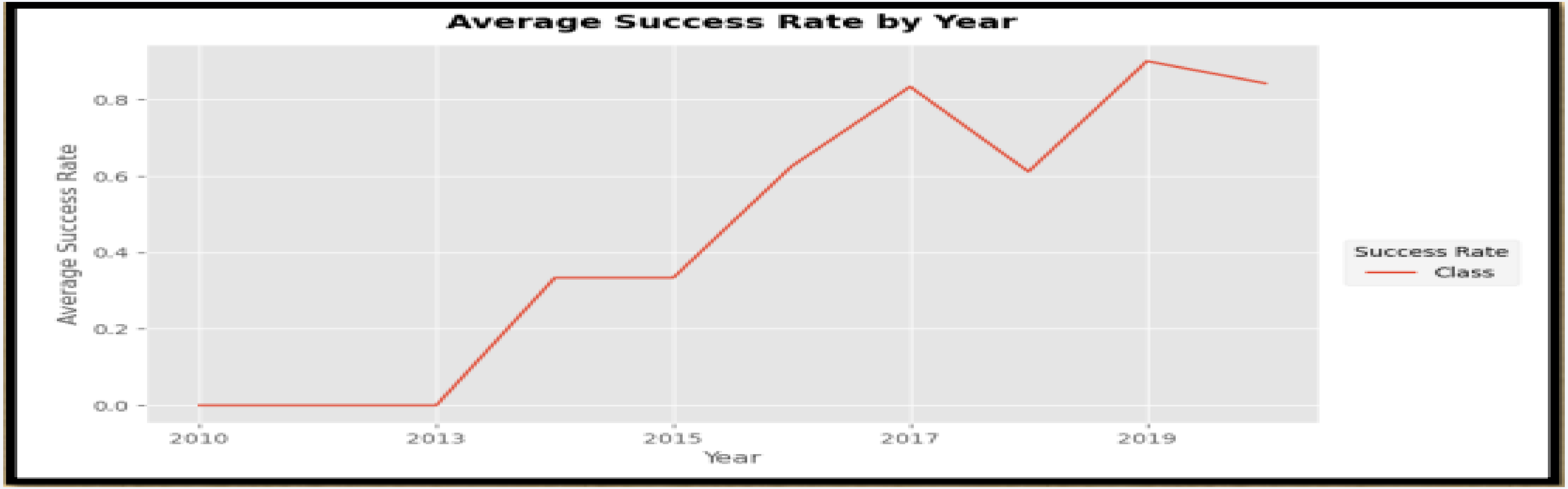
- Success rates are higher with the later flight numbers for most of the orbits.

Payload vs. Orbit Type



- In the majority of orbits as payload size increased, mission success was high.

Launch Success Yearly Trend



- Mission success rates started increasing materially after 2013 and continued to improve as time goes on.

All Launch Site Names

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

```
%%sql
SELECT DISTINCT(launch_site) FROM SPACEXTBL
```

Launch Site Names Begin with 'CCA'

DATE	time__utc_	booster_version	launch_site	payload	payload_mass__kg_	orbit	customer	mission_outcome	landing__outcome
2010-04-06	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
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	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
2012-08-10	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-01-03	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

```
select * from spacextbl where Launch_Site LIKE 'CCA%' limit 5;
```

Total Payload Mass

Total mass
45596

```
select sum(PAYLOAD_MASS__KG_) from spacextbl where Customer = 'NASA (CRS)'
```

Average Payload Mass by F9 v1.1

Booster Version	Average mass
F9 v1.1	2928.4

```
select avg(PAYLOAD_MASS__KG_) from spacextbl where Booster_Version LIKE 'F9 v1.1'
```

First Successful Ground Landing Date

min_date
2015-12-22

```
select min(Date) as min_date from spacextbl where Landing_Outcome = 'Success (ground pad)'
```

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

```
select Booster_Version from spacextbl where (PAYLOAD_MASS__KG_ > 4000 and PAYLOAD_MASS__KG_ < 6000)
and (Landing__Outcome = 'Success (drone ship)');
```

Total Number of Successful and Failure Mission Outcomes

success	failure
61	10

```
%%sql
SELECT (SELECT COUNT(*) FROM SPACEXTBL WHERE LANDING_OUTCOME LIKE 'Success%') as Success, (SELECT COUNT(*) FROM SPACEXTBL WHERE LAND
ING_OUTCOME LIKE 'Failure%') as Failure FROM SPACEXTBL LIMIT 1;
```

Boosters Carried Maximum Payload

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

```
select Booster_Version, PAYLOAD_MASS__KG_ from spacextbl where PAYLOAD_MASS__KG_ = (select max(PAYLOAD_MASS__KG_) from spacextbl)
```

2015 Launch Records

landing__outcome	booster_version	launch_site
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

```
%%sql
select landing_outcome, booster_version, launch_site from (select * from spacextbl where year(date) = '2015') where landing_outcome
= 'Failure (drone ship)';
```

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

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

```
%%sql
select count (*) as total, landing_outcome from spacextbl where date >= '2010-06-04' AND date <= '2017-03-20' group by landing_outc
ome order by total desc
```

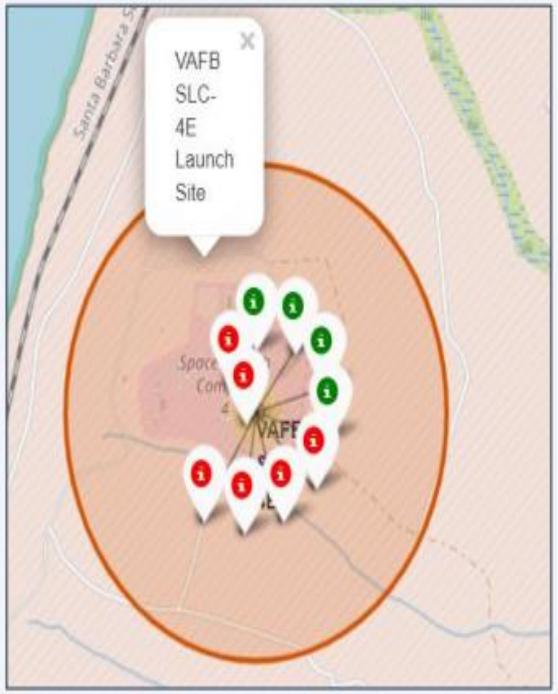
Section 3

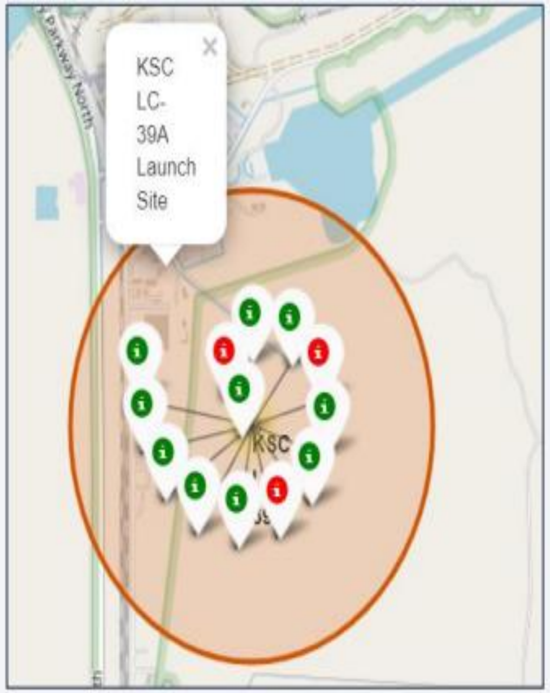
Launch Sites Proximities Analysis

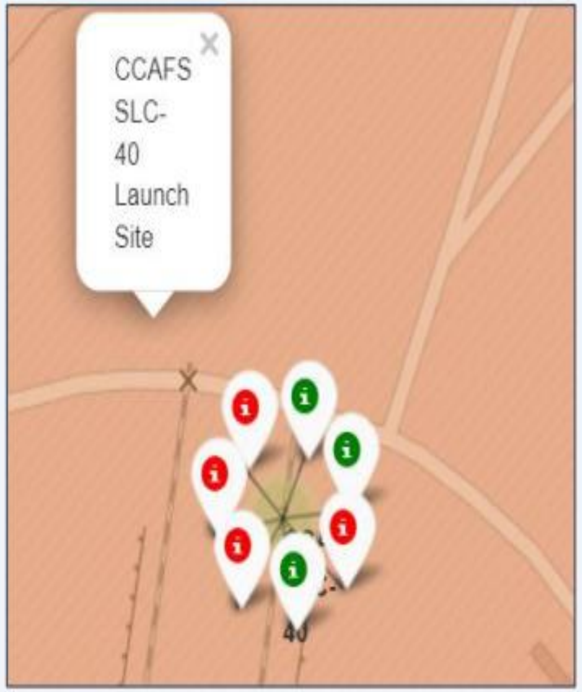
Map of Launch Sites

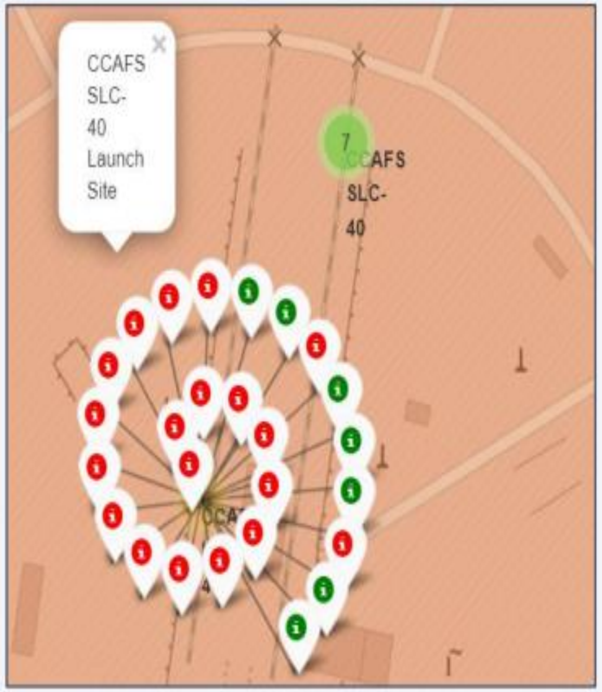
All launch sites have been identified in the map

Successful and Failed Landings per Launch Site









- Green = Success
- Red = Failure

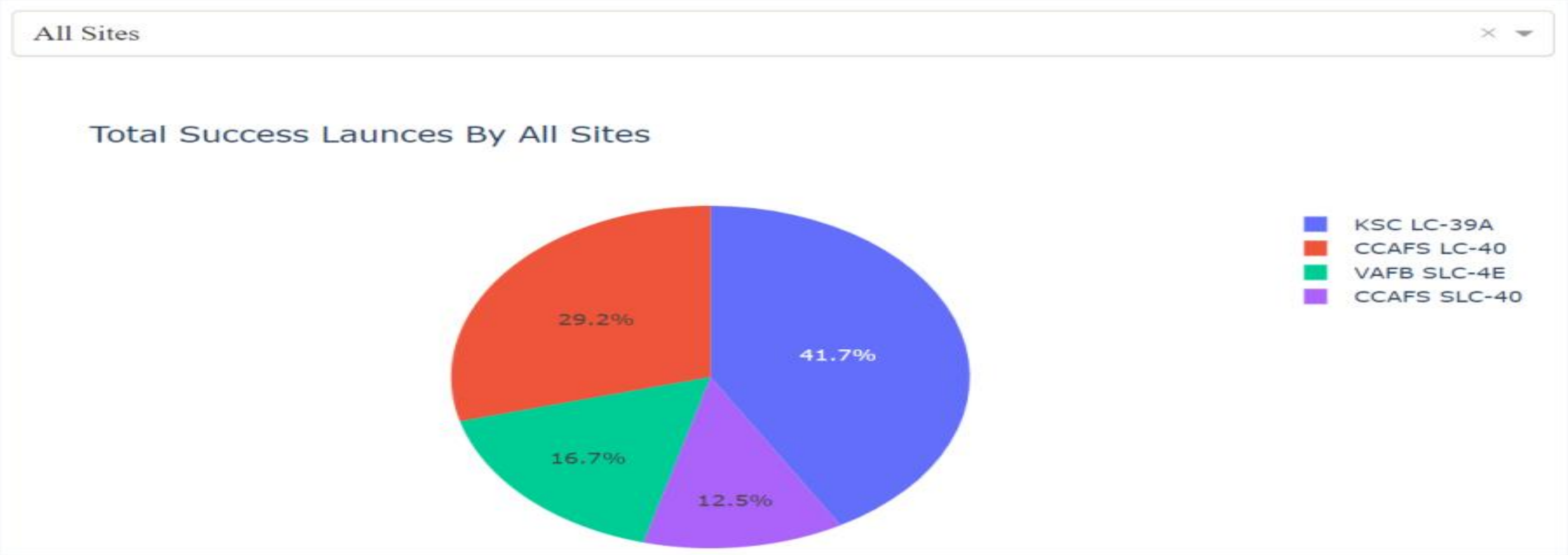
Proximity of Launch Sites

- This shows how close railroads, airports, etc., are from launch sites, for safety purposes.

Section 4

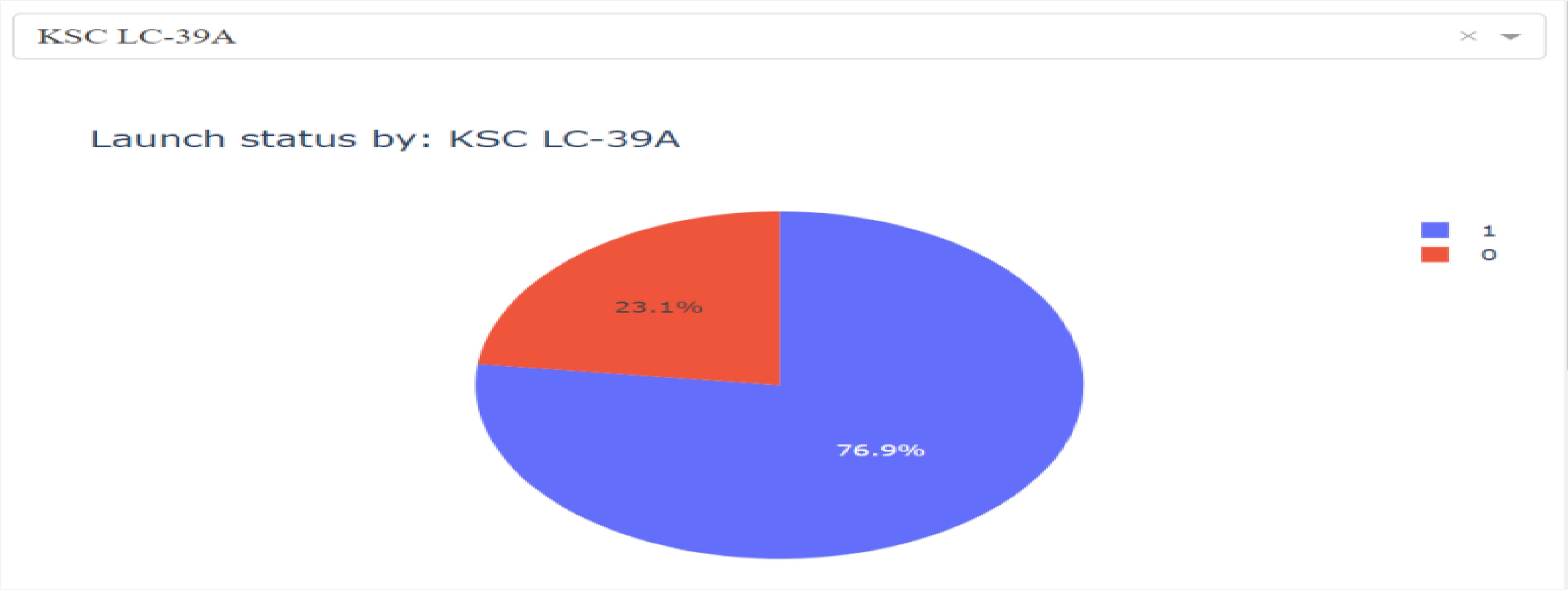
Build a Dashboard with Plotly Dash

Launch Site Sucess



- Interactive dashboard can be used to easily determine KSC LC-39A has the highest success rate and CCAFS SLC-40 has the lowest success rate.

Most Successful Launch Site



- The success rate is 76.9%

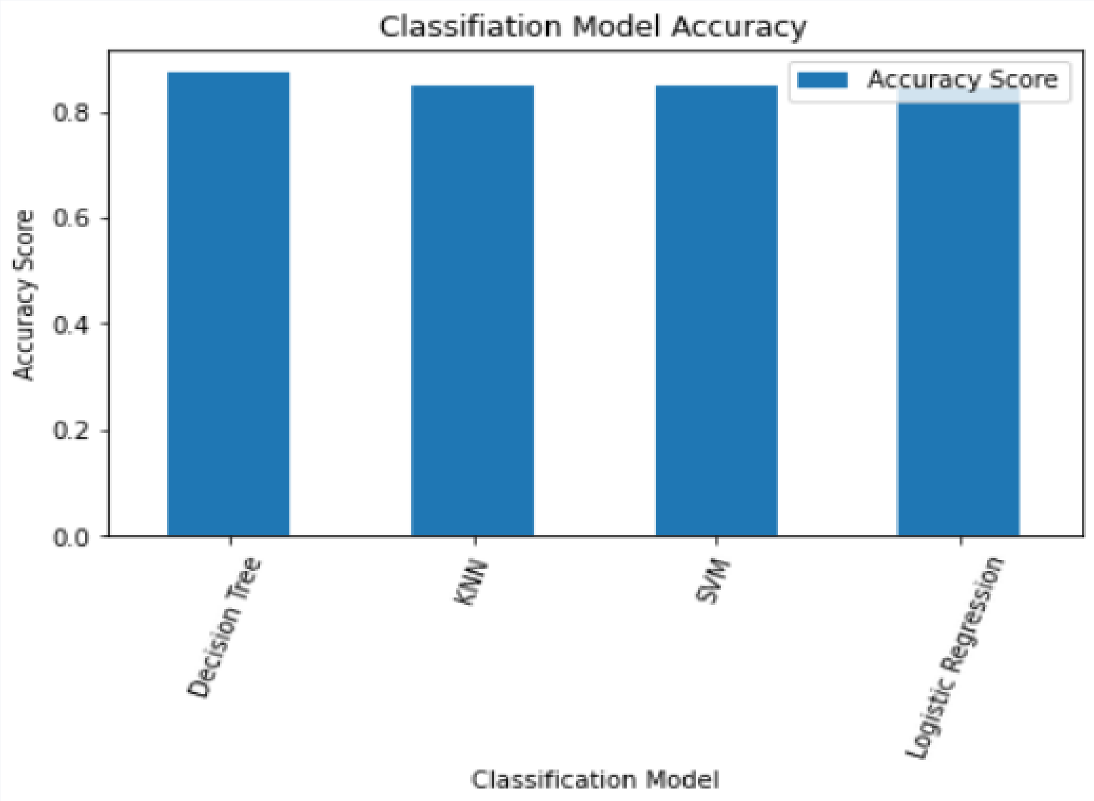
Payload vs Outcome

- The dashboard allows instant answers using the slide, for example, to show Booster FT is the most successful.

Section 5

Predictive Analysis (Classification)

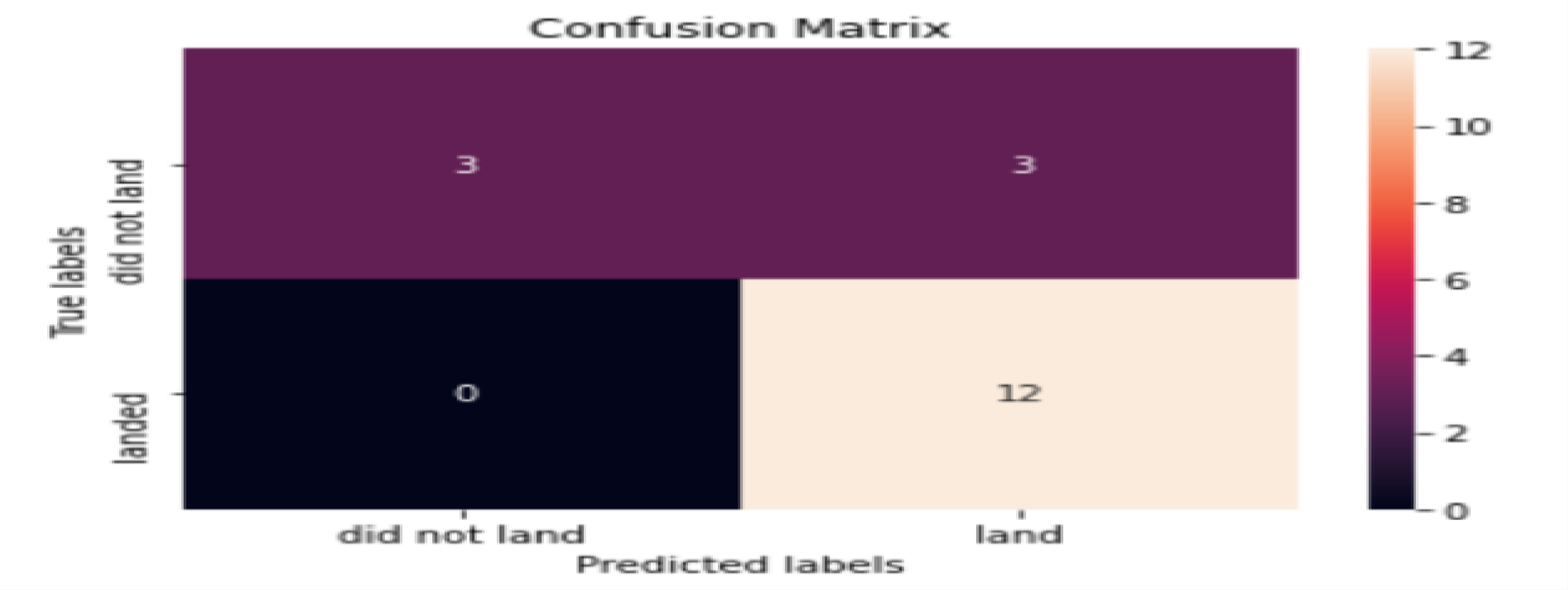
Classification Accuracy



	Algo Type	Accuracy Score	Test Data Accuracy Score
2	Decision Tree	0.875000	0.833333
3	KNN	0.848214	0.833333
1	SVM	0.848214	0.833333
0	Logistic Regression	0.846429	0.833333

- The Decision Tree is most accurate at 0.875

Confusion Matrix



- Out of 18 predictions, 12 were accurate (83% accuracy).

Conclusions

- Success rates have been increasing as time goes by and will most likely continue to increase.
- KSC LC-39A is the most successful launch site.
- Although Decision Tree has the most success, all prediction algorithms are accurate.
- Locating launch sites near the coast and away from railroads, airports, and people are done for safety purposes.
- I would recommend my (fictional) employer, Ramjett Rocket Industries not compete directly with SpaceX in less they can create a similar rocket strategy, where they can re-use the stage 1 rocket, as it appears in future years SpaceX will be successful with their code reduction efforts.

Appendix

- https://github.com/jdjeep57/JamesDanielKnapp-Capstone-Project.git

Thank you!