Winning Space Race with Data Science

Stephanie Watson
November 16, 2021

Outline

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

Executive Summary

- Summary of methodologies
  - Data collection included parsing SpaceX APIs and Falcon 9 Wikipedia page for information regarding rockets utilized, payloads, landing outcomes, launch sites, etc.
  - Data wrangling included separating out the Falcon 9 rocket data. Null values were replaced with the column mean and landing outcome data was reclassified to aid in EDA and ML activities.
  - Exploratory data analysis (EDA) using visualization (charts) and SQL was performed
  - Interactive visual analytics using Folium and Plotly Dash was performed
  - Predictive analysis using classification models was performed (Sci-kit learn library)
- Summary of all results
  - We can correctly identify if a Falcon 9 launch will fail or succeed with 83% accuracy.
- GitHub Repository: https://github.com/SwatsonDS/IBM_DS_Capstone_Project

Introduction

- The commercial space age is here! Companies are making space travel affordable for everyone.
  - Some rocket providers cost upwards of 165 million dollars per launch
  - SpaceX advertises on its website that Falcon 9 rocket launches cost 62 million dollars[1], resulting in substantial savings
  - One reason why SpaceX can offer rocket launches at a much lower price is because it can reuse the first stage of the Falcon 9 rocket
- Problems Statement:
  - This project aims to predict if the Falcon 9 first stage will land successfully
  - By determining if the first stage will land successfully, (and thus can be reused) the cost of the rocket launch can be determined.

1:https://www.spacex.com/media/Capabilities&Services.pdf

Section 1

Methodology

Methodology

Executive Summary

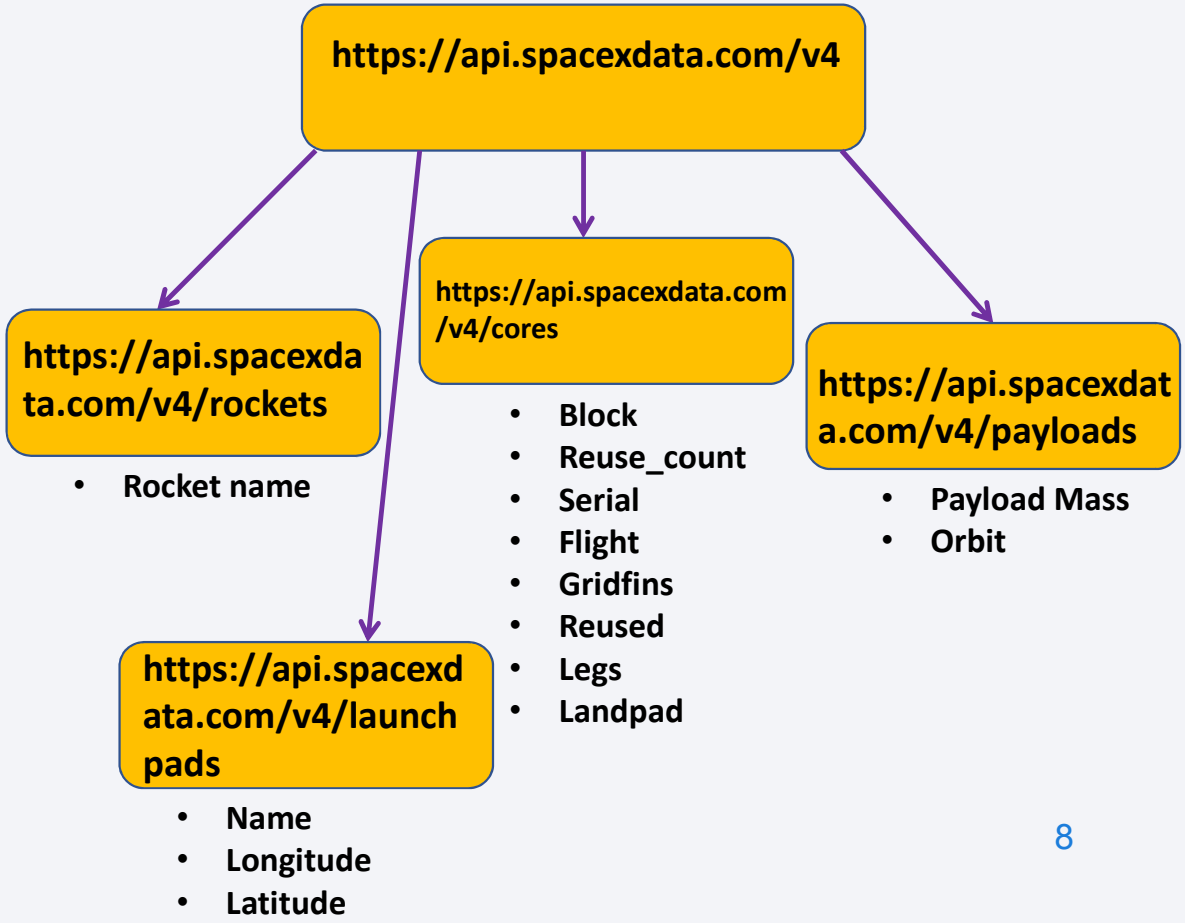
- Data collection methodology:
  - Past launch data was collected from the SpaceX REST API and the Falcon 9 Wikipedia page which provided information about rockets utilized, payloads, landing outcomes, etc
- Perform data wrangling
  - The data was sorted to only contain Falcon 9 data
  - Null values of the Payload Mass column were replaced with the mean this column
  - Outcome values were reclassified as success (1) or fail (0) to aid in EDA and ML classification
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Data was one hot encoded and standardized before GridSearch CV was utilized to determine best hyperparameters for each model

Data Collection

- Past launch data was collected from the SpaceX REST API and the Falcon 9 Wikipedia page which provided information about rockets utilized, payloads, landing outcomes, etc
- Datasets were converted to dataframes
- Dataset part 1 – Obtained from SpaceX API call
- Dataset part 2 – Part 1 dataset updated to classify success/fail outcomes
- Dataset part 3 – Part 2 dataset updated to with one hot encoding for key variables

Data Collection – SpaceX API

- ‘Get’ request was made to url to obtain launch data in the form of a .json file
- .json file was converted to dataframe
- Calls were made to various APIs to obtain launch information (see flowchart)
- This information was placed into list
- A dictionary was created to store keys and list of data from API calls
- This dictionary was converted to a dataframe
- Missing values in the Payload Mass column were replaced with the mean of the column
- Dataset labelled as part1.csv
- GitHub URL: https://github.com/SwatsonDS/IBM_DS_Capstone_Project/blob/main/Data%20Collection.ipynb



Data Collection - Scraping

- ‘Get’ request was made to wikipage url to obtain launch data in the form of a .HTML text file
- A Beautiful Soup object was created to parse the HTML data
- The HTML data was sorted to find the table containing past launch data
- This table was iterated though to find all the column headers, then a dictionary was created with the column header as the dictionary keys
- Each row of this table was iterated through to obtain launch information, this information was stored in respective list tied to the dictionary keys
- This dictionary was converted a dataframe
- Dataset labelled as space_web_scraped.csv
- GitHub URL: https://github.com/SwatsonDS/IBM_DS_Capstone_Project/blob/main/Data%20Collection%20with%20Web%20Scraping.ipynb

Get request to Wikipage

Pull data from Past Launch table

Data Wrangling

- The .json files obtained from the SpaceX API calls were converted to a dataframe
- Missing values in the Payload Mass column were replaced with the mean of the column
- Dataframe was updated to only include data on Falcon 9 rockets
- Dataset labelled part1.csv
- Outcome values were reclassified as success (1) or fail (0) to aid in EDA and ML classification
- Updated dataset labelled part2.csv
- GitHub URL: https://github.com/SwatsonDS/IBM_DS_Capstone_Project/blob/main/Exploratory%20Data%20Analysis.ipynb

EDA with Data Visualization

- Dataset was first loaded into a dataframe
- Utilizing the Seaborn library in Python:
  - Scatterplots were created to determine if a correlation existed between the variables, Flight Number, Payload Mass, Orbit
  - A Bar chart was utilized to visually display the success rate for each orbit type
  - A line chart was utilized to visually depict the success rate on a yearly basis.
- Based on the charts display, One Hot Encoding was performed on columns Orbit, Launch Site, Landing Pad, and Serial (saved as dataset_part3) to convert categorical data to binary to allow for better prediction in machine learning algorithms
- GitHub URL: https://github.com/SwatsonDS/IBM_DS_Capstone_Project/blob/main/EDA%20Visualization.ipynb

EDA with SQL

- A CSV dataset comprised of SpaceX launch information was uploaded to IBM Db2 SQL server
- The SQL extension was loaded and the database was then linked to the juptyer notebook file to allow for querying
- SQL queries were performed to:
  - Determine unique launch sites utilized by SpaceX
  - Calculate the total and average payloads carried by Falcon 9 for various versions and customers
  - Determine the date of the first successful landing
  - Determine number of successful and failed landings
- GitHub URL: https://github.com/SwatsonDS/IBM_DS_Capstone_Project/blob/main/Exploratory%20Data%20Analysis%20with%20SQL.ipynb

Build an Interactive Map with Folium

- Maps were created using the Folium library in Python
  - The coordinates from each site were collected from the dataset
  - A folium map object was created of the United States
  - Markers/Circles were created to depict the location of the various launch sites utilized by SpaceX and added to the map object along with popup site label
  - Marker Clusters were created to depict the number of success/fail launch attempts at each launch site and added to the map object
  - Distances were calculated from one launch site to various landmarks (city, highway, railway) to determine the proximity of each to the launch site
    - Lines and markers were added to the created Folium map to highlight the distances
- GitHub URL: https://github.com/SwatsonDS/IBM_DS_Capstone_Project/blob/main/Visualization%20with%20Folium.ipynb

Build a Dashboard with Plotly Dash

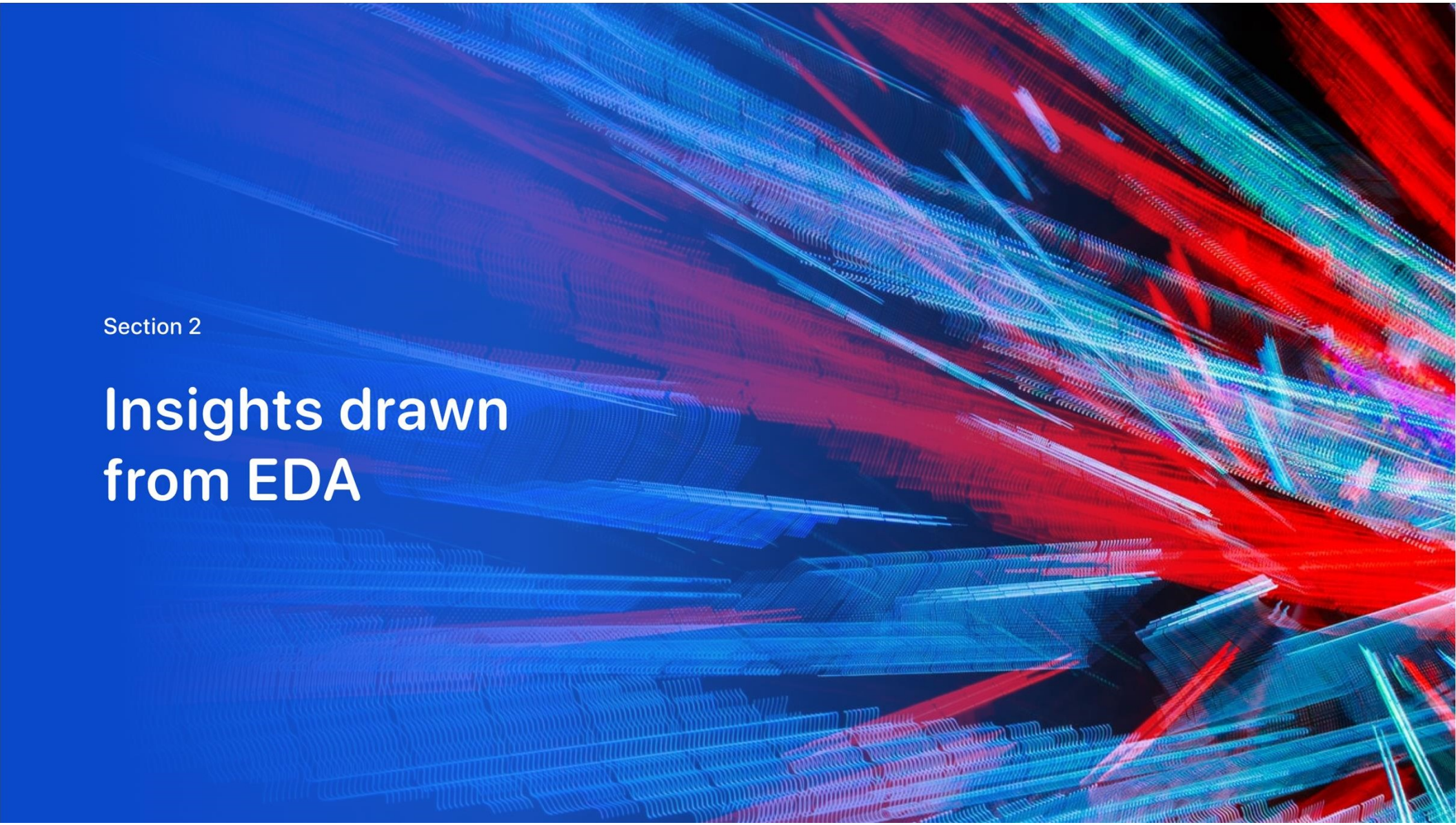
- Created Interactive Dashboard utilizing the Plotly library in Python
  - Created Dash application
  - Created Dash layout:
    - Created dropdown list to allow for launch site selection
    - Created interactive pie chart to depict the number of successful/failed launches per launch site. This will allow for a visual representation of successful launches for each site. Utilize callback function to link dropdown list to pie chart.
    - Created slider to select payload range.
    - Created interactive scatterplot that visualizes the payload mass vs. success rate by each Falcon 9 booster version type. This will allow for a visual representation of success rate for each booster type based on the payload. Utilize callback function to link slider to scatterplot.
- GitHub URL: https://github.com/SwatsonDS/IBM_DS_Capstone_Project/blob/main/spacex_dash_app_final.py

Predictive Analysis (Classification)

- Classification models were created utilizing the Sci-kit learn library in Python
  - The dataset utilized was part3.csv (after one hot encoding was performed)
  - The dataset was standardized utilizing the Standard Scalar transform method
  - The dataset was split into a training portion (80%) and a testing portion (20%)
  - To determine the best hyperparameters, a GridSearchCV object was created for each classification model to perform an exhaustive search of specified parameters
  - The best hyperparameters for each classification model were then used to test each model accuracy on the test data
- GitHub URL: https://github.com/SwatsonDS/IBM_DS_Capstone_Project/blob/main/Machine%20Learning%20Prediction.ipynb

Results

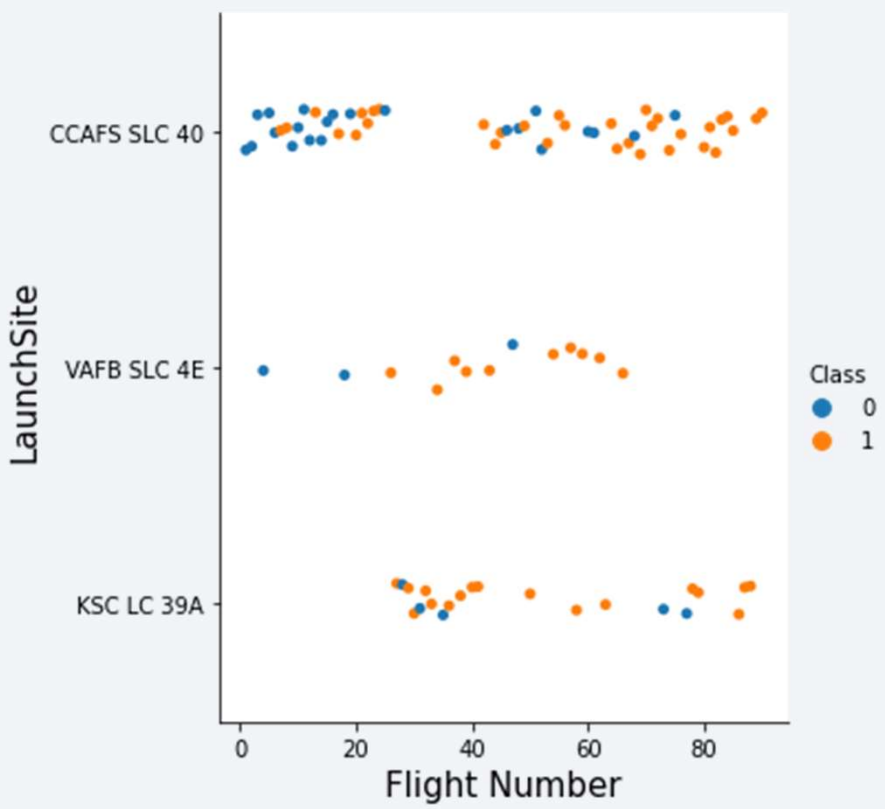
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results



Section 2

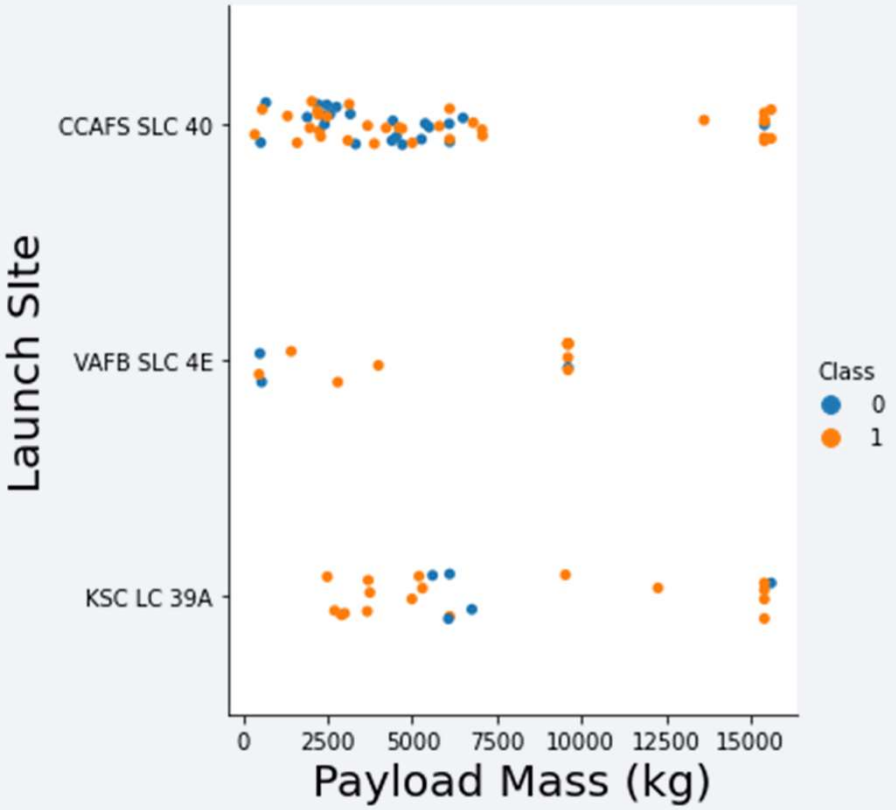
Insights drawn from EDA

Flight Number vs. Launch Site



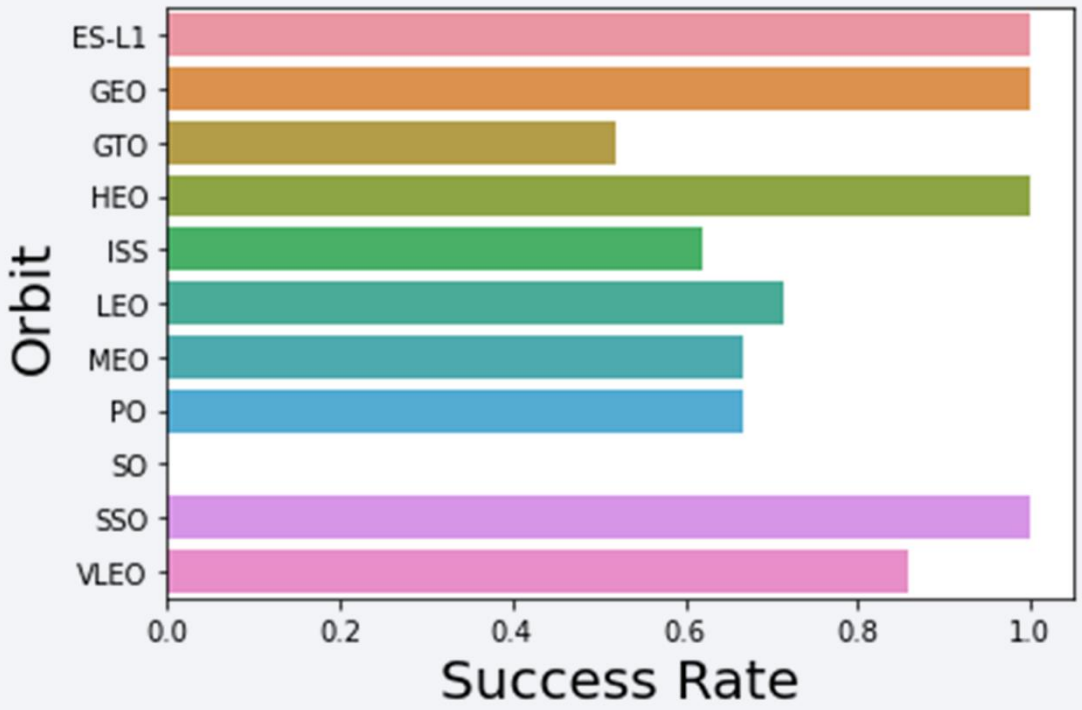
- Shown is a scatter plot of Flight Number vs. Launch Site
  - Class 0 – Failure
  - Class 1 – Success
- No correlation between flight number and launch site

Payload vs. Launch Site



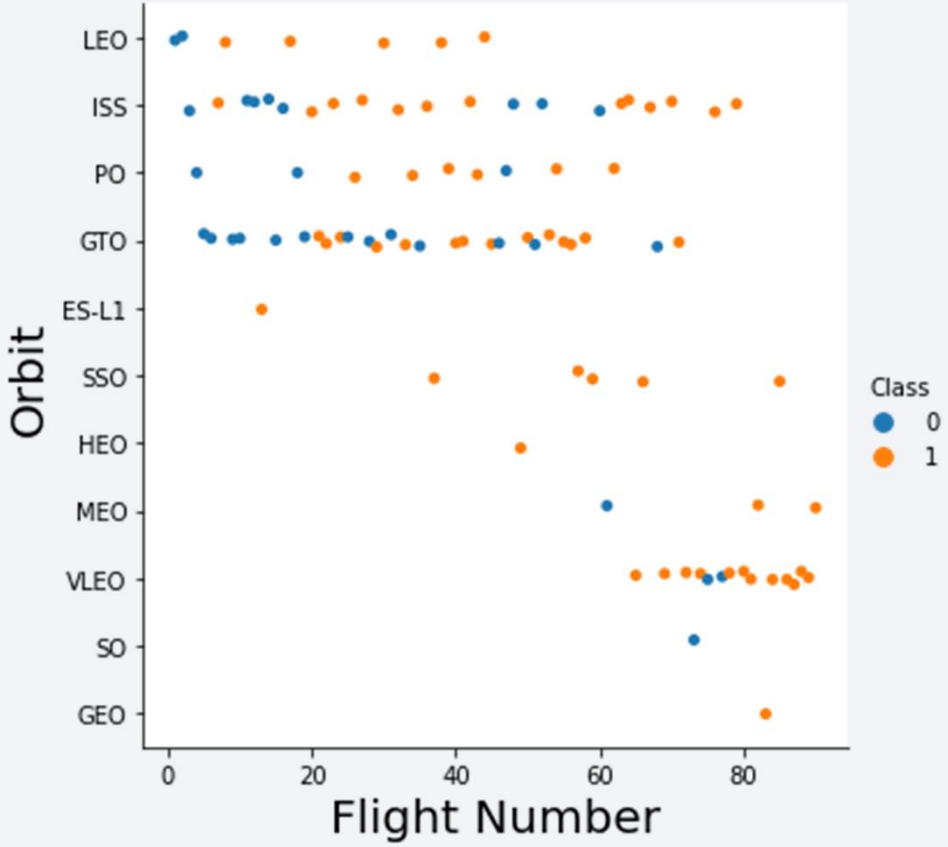
- Shown is a scatter plot of Payload vs. Launch Site
  - Class 0 – Failure
  - Class 1 – Success
- Higher payload mass result in a higher success rate
  - CCAFS/KSC Overall Success rate: 60% / 77%
  - CCAFS/KSC Success rate with Payload Mass above 10M kg: 89%/83%
- Site VAFB has no launches with payload above 10,000 kg
  - Success rate: 77%

Success Rate vs. Orbit Type



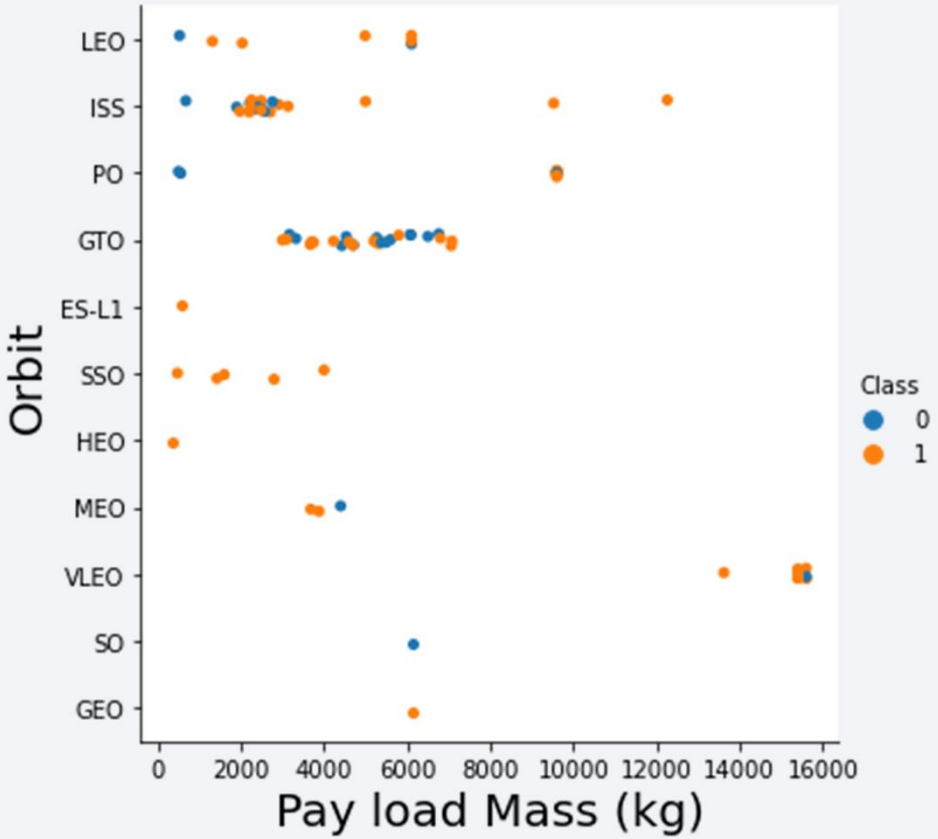
- Shown is a bar chart for the success rate of each orbit type
- Orbits ‘ES-L1’, ‘GEO’, ‘HEO’, and ‘SSO’ have high success rates (100%)
- There have been no successful launches to ‘SO’ orbit
  - Only 1 launch attempted

Flight Number vs. Orbit Type



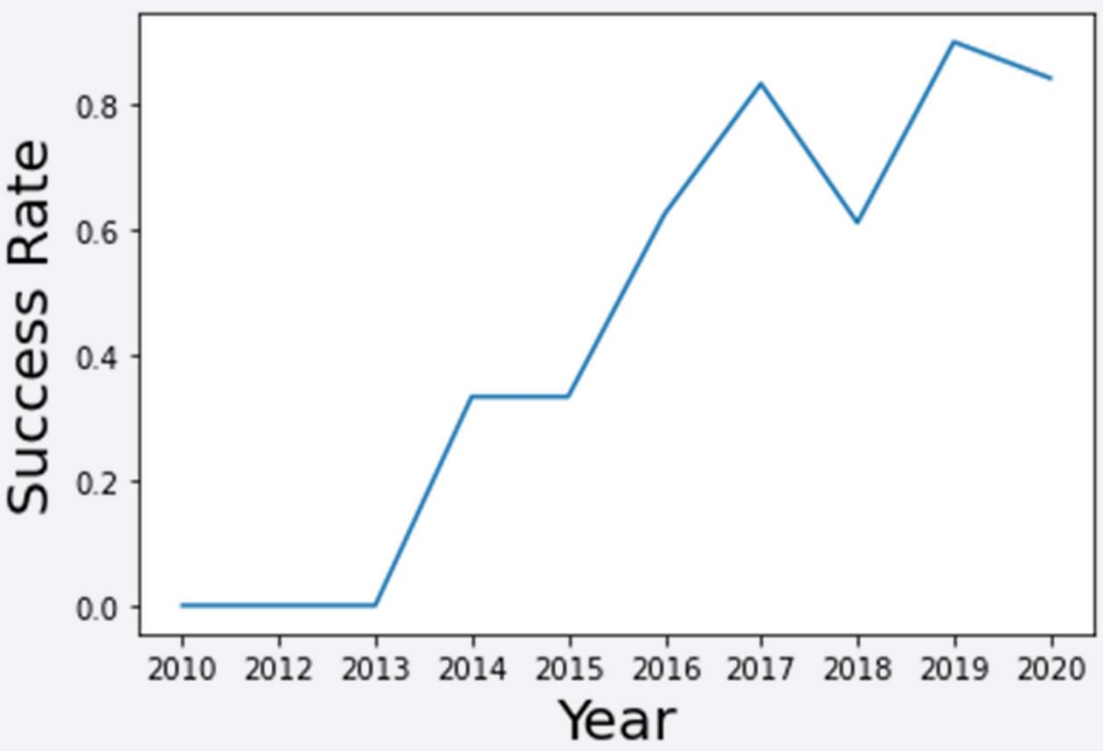
- Shown is a scatter plot of Flight number vs. Orbit type
  - Class 0 – Failure
  - Class 1 – Success
- Orbit ‘SSO’ has a high success rate throughout all flights
- After two failed attempts, Orbit ‘LEO’ success launches appear related to number of flights
- There is no correlation between flight number and the “GTO’ orbit

Payload vs. Orbit Type



- Shown is a scatterplot of payload vs. orbit type
  - Class 0 – Failure
  - Class 1 – Success
- The majority of the flights occur with payloads at or below 10,000 kg (83%)
- There is no correlation between payload mass and the “GTO’ orbit

Launch Success Yearly Trend



- Shown is a line chart of yearly average success rate
- The success rate trend of Falcon 9 launches has continued to increase since 2013

All Launch Site Names

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

- Shown is a table of all of the launch site utilized for SpaceX Falcon 9 launches
- In total, there have been four(4) site utilized for SpaceX launches

Launch Site Names Begin with 'CCA'

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	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	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

- Shown are 5 records where launch sites begin with `CCA’

Total Payload Mass for NASA

total_payload_mass_nasa
107010

- The total payload carried by boosters for NASA is 107,010 kg

Average Payload Mass by F9 v1.1

avg_payload_f9
2534

- The average payload mass carried by booster version F9 v1.1 is 2,534 kg

First Successful Ground Landing Date

first_success_ground
2015-12-22

first_success_drone
2016-04-08

- The first successful landing outcome on ground pad was on December 22, 2015
- The first successful landing outcome on drone ship was on April 8, 2016

Successful Drone Ship Landing with Payload between 4000 and 6000

booster_drone_4_6
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

- Shown are the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

Total Number of Successful and Failure Mission Outcomes

mission_outcome	total
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

- Shown are the total number of successful and failure mission outcomes

Boosters Carried Maximum Payload

booster_version
F9 B5 B1048.4
F9 B5 B1048.5
F9 B5 B1049.4
F9 B5 B1049.5
F9 B5 B1049.7
F9 B5 B1051.3
F9 B5 B1051.4
F9 B5 B1051.6
F9 B5 B1056.4
F9 B5 B1058.3
F9 B5 B1060.2
F9 B5 B1060.3

- Shown are the names of the booster which have carried the maximum payload mass of 15,600 kg

2017 Launch Records

MONTH	landing_outcome	booster_version	launch_site
February	Success (ground pad)	F9 FT B1031.1	KSC LC-39A
May	Success (ground pad)	F9 FT B1032.1	KSC LC-39A
June	Success (ground pad)	F9 FT B1035.1	KSC LC-39A
August	Success (ground pad)	F9 B4 B1039.1	KSC LC-39A
September	Success (ground pad)	F9 B4 B1040.1	KSC LC-39A
December	Success (ground pad)	F9 FT B1035.2	CCAFS SLC-40

- List displays the month successful landing outcomes in ground pad, booster versions, launch site for the months in year 2017

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

DATE	landing_outcome
2017-02-19	Success (ground pad)
2017-01-14	Success (drone ship)
2016-08-14	Success (drone ship)
2016-07-18	Success (ground pad)
2016-05-27	Success (drone ship)
2016-05-06	Success (drone ship)
2016-04-08	Success (drone ship)
2015-12-22	Success (ground pad)

- Shown is a table of successful landing outcomes between the date 2010-06-04 and 2017-03-20 in descending order

- The first successful outcome did not occur until December 2015


Section 4
Launch Sites
Proximities Analysis

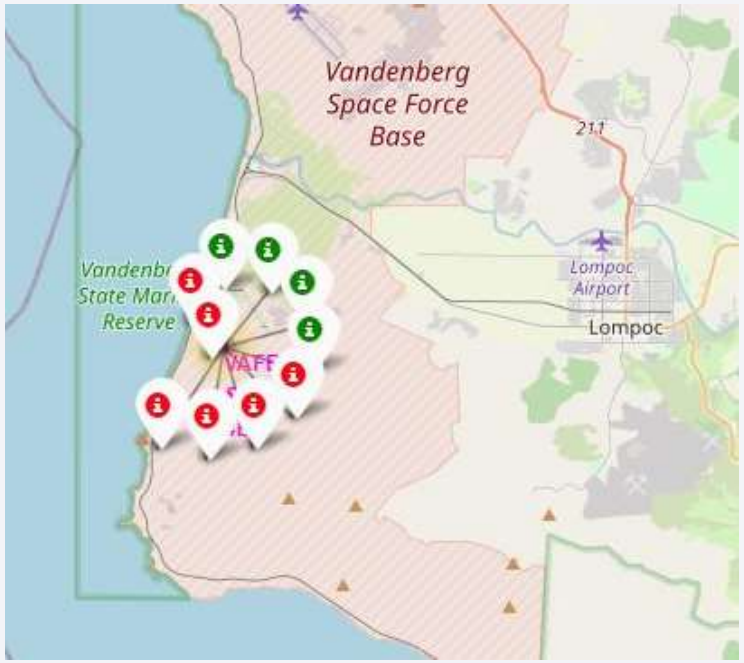
All SpaceX Launch Sites

- Shown is a map of all of the SpaceX launch sites
- Three launch sites are located on the east coast in Florida and one launch site is located on the west coast in California.

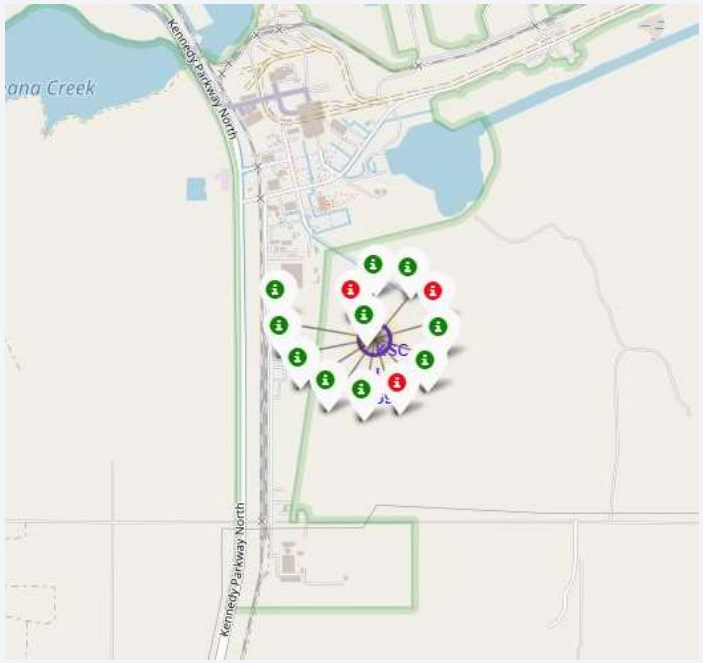
SpaceX Launch Site Success/Failures

- Shown is a map depicting the number of launches per launch site
- Four times as many launches have taken place at the Florida launch sites, this could potentially be due to the close proximity to 3 of the 4 launch sites

SpaceX Launch Site Success/Failures cont…



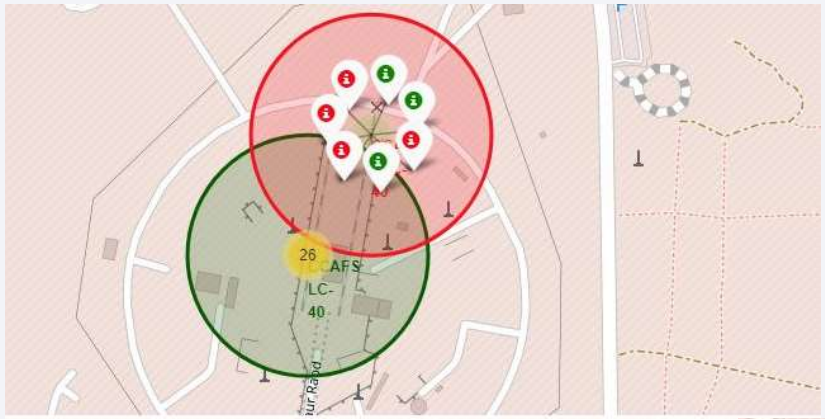
Launch Site: VAFB



Launch Site: KSC

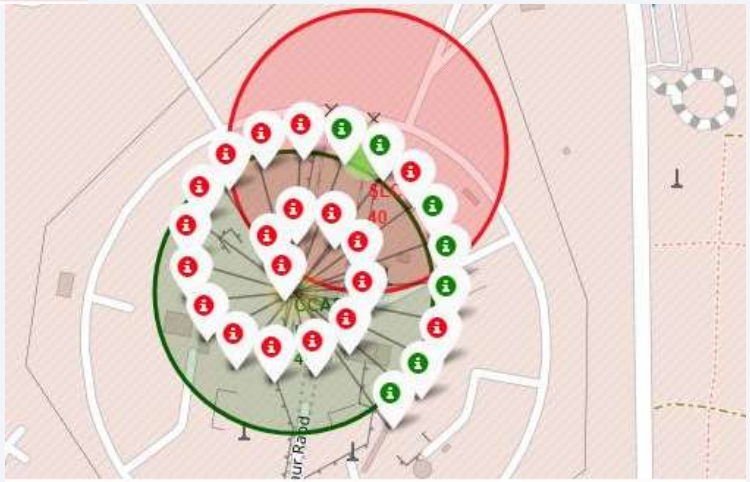
- Shown are maps of each launch site with color-labeled launch outcomes on the map
  - Green – Success
  - Red – Failure
- Site ‘KSC’ has the highest success rate

SpaceX Launch Site Success/Failures cont…



Launch Site: SLC-40

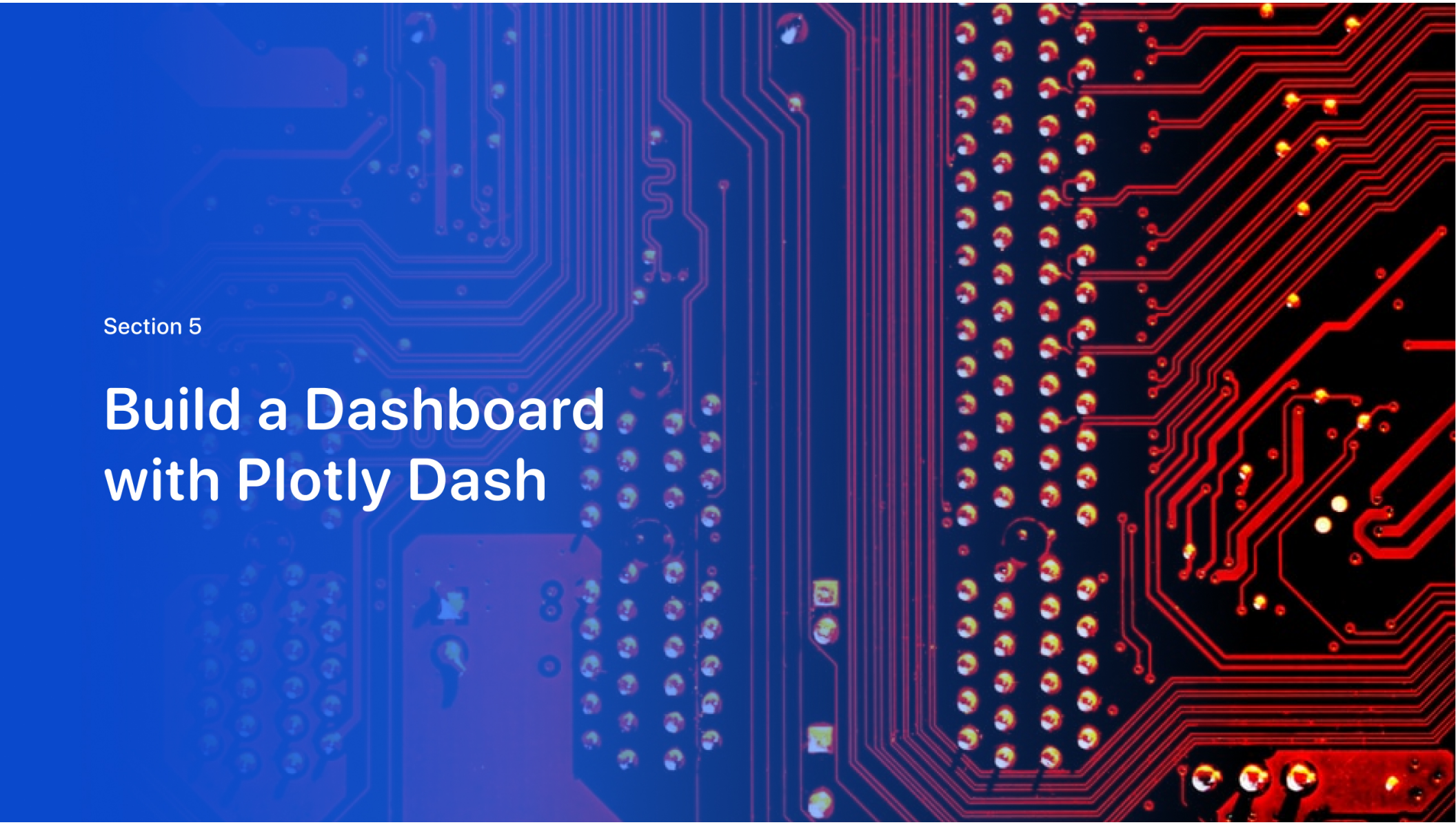
Launch Site: LC-40



- Shown are maps of each launch site with color-labeled launch outcomes on the map
  - Green – Success
  - Red – Failure
- Site ‘LC-40’ has the most launch attempts (26)

Launch Site Proximities

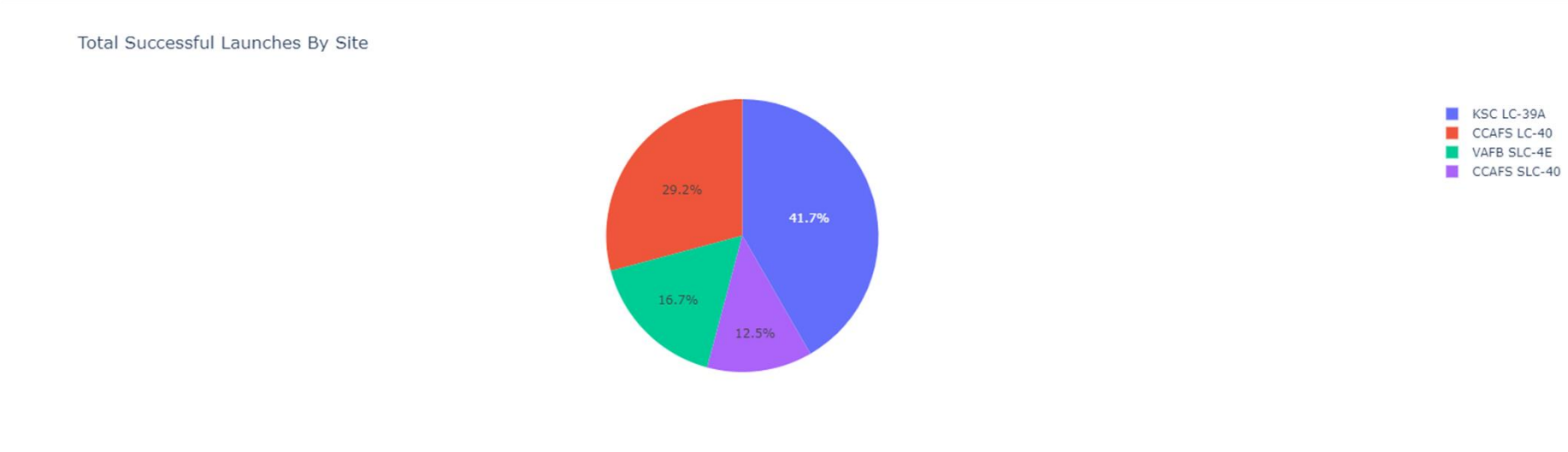
- Shown is a map detailing the proximity of launch site “VAFB” to critical areas (city, highway, railway, coastline)
- Launch site appears to be close to the coastline and railway but far away from highway and cities
  - Close proximity to railways allows for easy transport of rockets and payload to/from launch site
  - A far distance away from highways and cities reduces the likelihood of civilian causalities in the event of a launch incident (crash, explosion, loss payload, etc)



Section 5

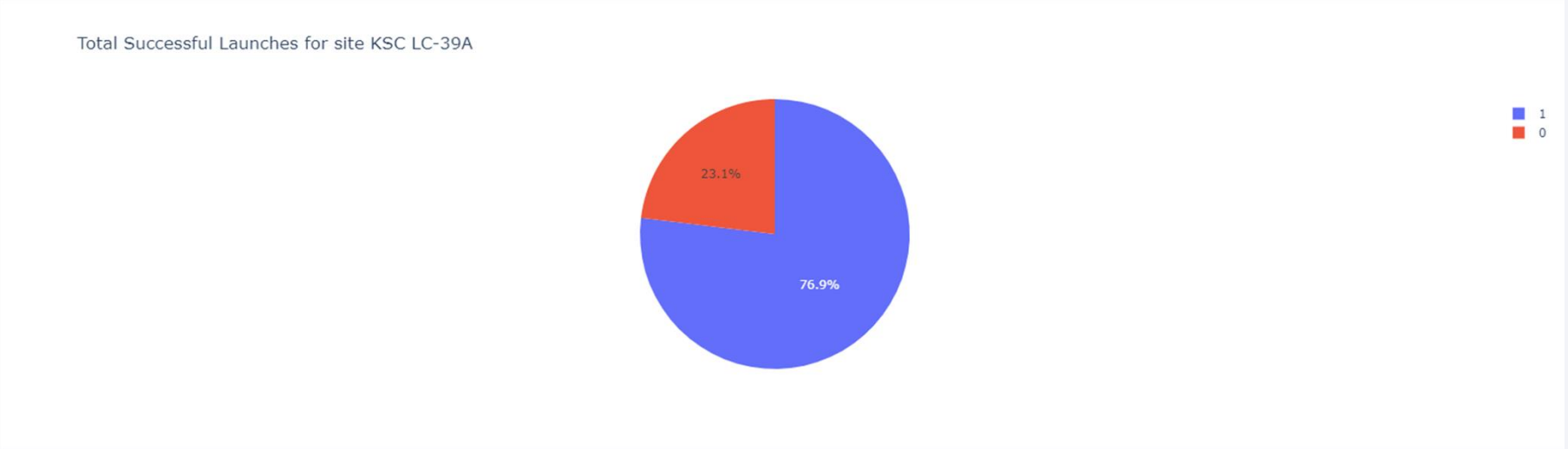
Build a Dashboard with Plotly Dash

Total Successful Launches for all Sites



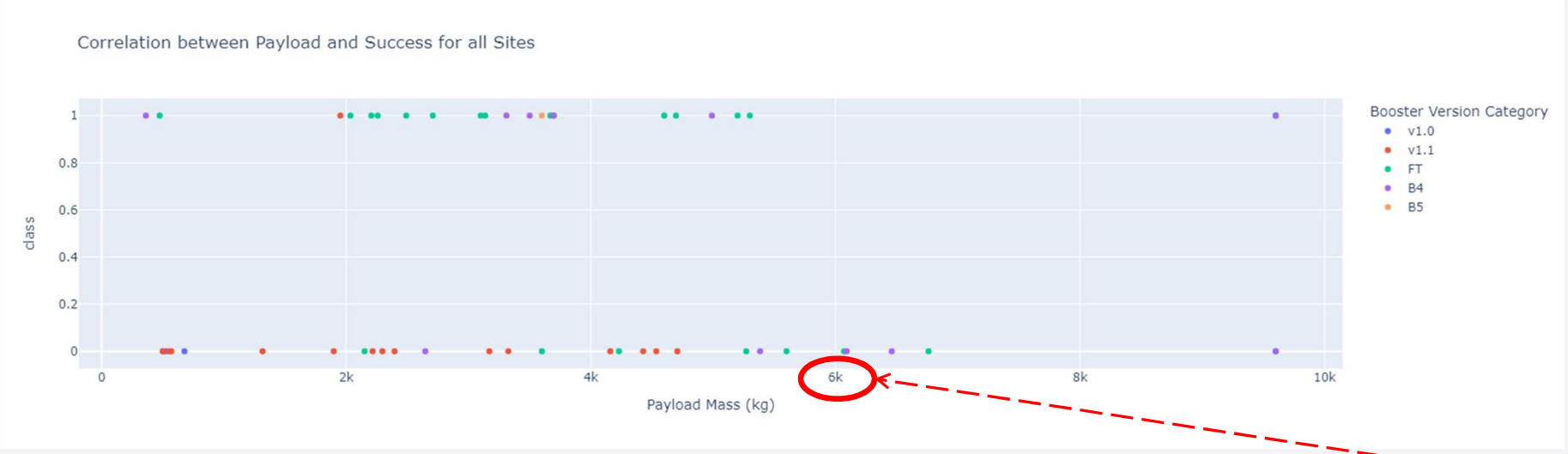
- Based on the dataset provided, there are a total of 24 successful launches out of 46 total launches from all four (4) sites resulting in an overall success rate of 52%.
- Note that the site with the largest amount of attempted launches is CCAFS LC-40 which also has the highest percentage of unsuccessful launches (73%).
- The site with the highest number of successful launches is KSC LC-39A.

Total Successful Launches for KSC LC-39A



- Site KSC LC-39A has the highest launch success rate of 77%, with 10 successful launches and 3 unsuccessful launches.
- Its three (3) unsuccessful attempts occurred with payload weights above 5500 kg.

Correlation between Payload and Success Rate

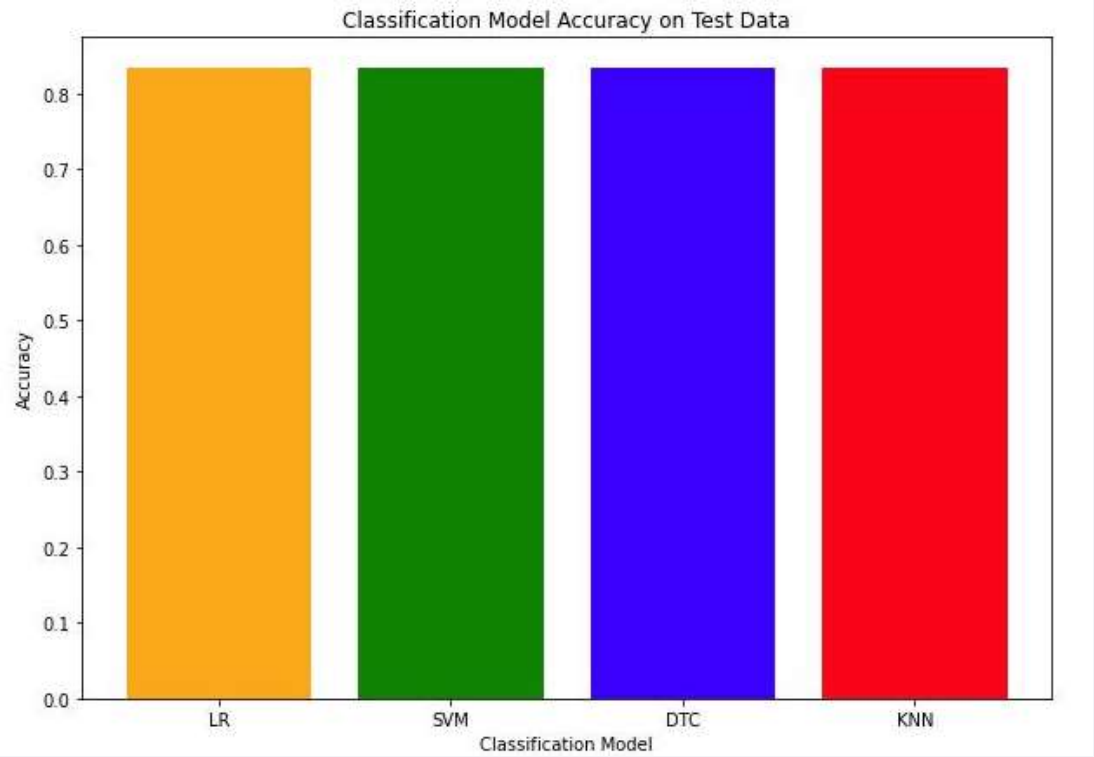


- The majority of the successful launches have occurred with payload weights below 6,000 kg.
  - The payload range with the highest launch success rate is 2,000 – 4000 kg
  - The payload range with the lowest launch success rate is 6,000 – 8000 kg (no successful launches)
- The F9 Booster version with the highest launch success rate is “FT”
  - Note that F9 Booster version “B5” has a success rate of 100% but it only has one attempted launch.

Section 6

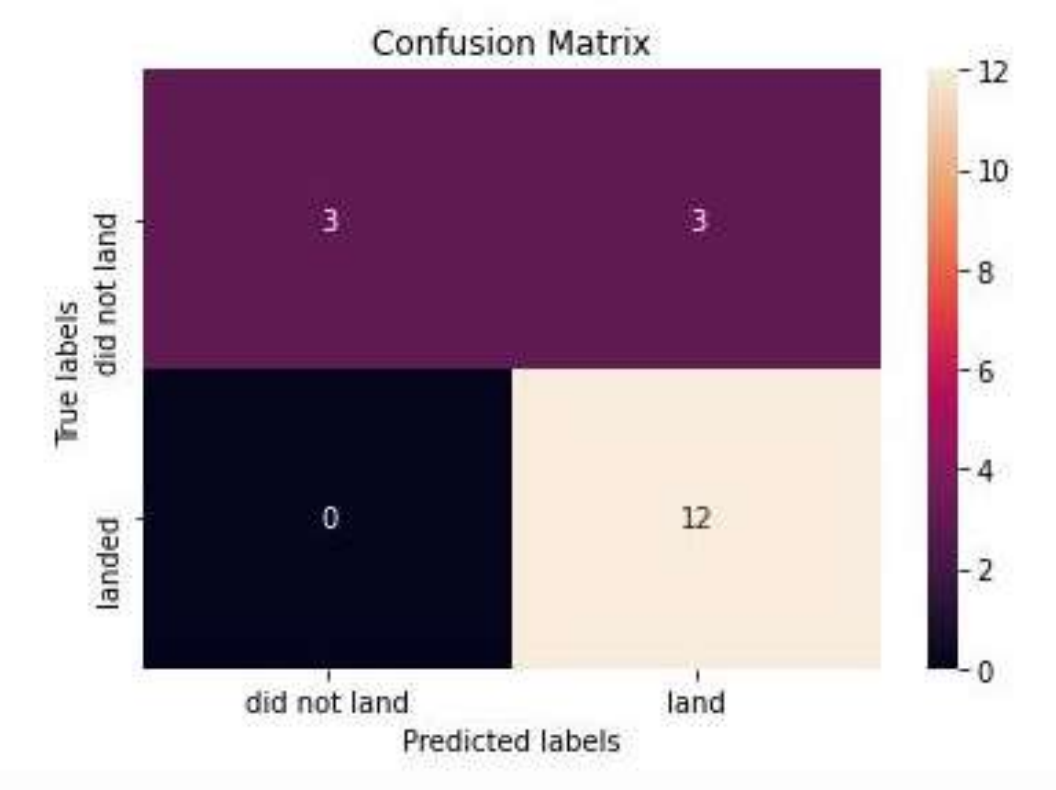
Predictive Analysis (Classification)

Classification Accuracy



- Each classification model utilized produces the same level of accuracy (83.33%) on the test data (see bar chart)
- Note that the dataset utilized from this analysis is small which could contribute to the lack of classification variation

Confusion Matrix

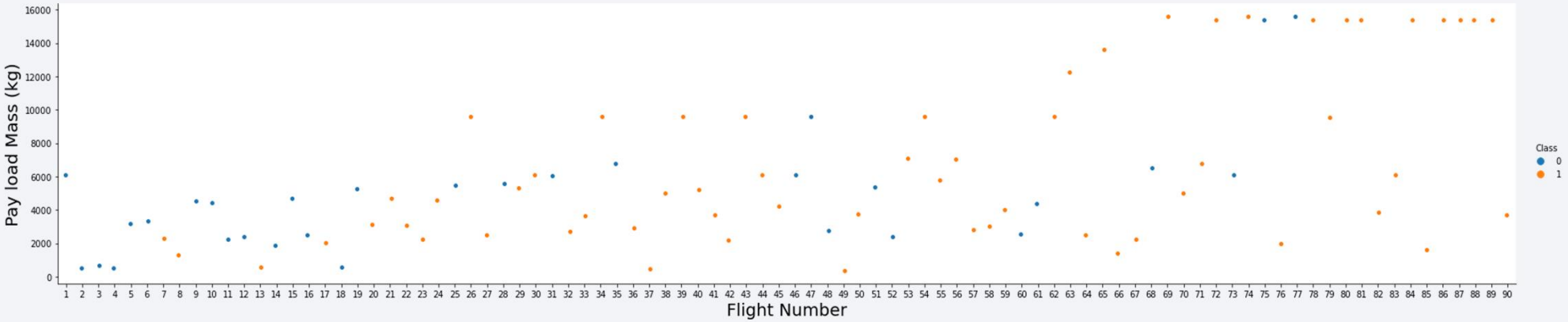


- Shown is the confusion matrix for one of the classification models utilized.
  - Note that each classification model produced the same matrix
- Out of the 18 samples utilized for the classification model testing, each model was able to correctly classify 15 samples.
  - The major issue with the models is false negatives
  - The models predicted 3 successful landings that were failed landings

Conclusions

- The majority of the flights occur with payloads at or below 10,000 kg, higher payload mass appear to result in a decrease likelihood of launch success
- All Launch Site are located relatively close to coastlines and railway and are located relatively far away from cities and highways. This is too allow for easy transportation of material to/from the launch site and to minimize civilian causalities in the event of a launch failure
- We can correctly identify if a Falcon 9 launch will fail or succeed with 83% accuracy.

Appendix： Flight Number vs. Payload Mass (kg)



Appendix： Launch Site Success Rates

Launch Site Success Rate (all instances)

Launch Site	Class
CCAFS SLC 40	0.600000
KSC LC 39A	0.772727
VAFB SLC 4E	0.769231

```
series_site = df.groupby('LaunchSite')['Class'].mean() #sorting dataframe by launchsite to find average success rate of each launchsite
df_site=pd.DataFrame(series_site)
df_site.reset_index(inplace=True)
df_site.head()
```

	LaunchSite	Class
0	CCAFS SLC 40	0.600000
1	KSC LC 39A	0.772727
2	VAFB SLC 4E	0.769231

Launch Site Success Rate (Payload > 10,000)

Launch Site	Class
CCAFS SLC 40	0.888889
KSC LC 39A	0.833333

```
series_10 = df[df['PayloadMass'] > 10000].groupby('LaunchSite')['Class'].mean() # sorting dataframe to find launches with mass above 10M kg, then find
df_10=pd.DataFrame(series_10)
df_10.reset_index(inplace=True)
df_10.head()
```

	LaunchSite	Class
0	CCAFS SLC 40	0.888889
1	KSC LC 39A	0.833333

Appendix： Max payload mass

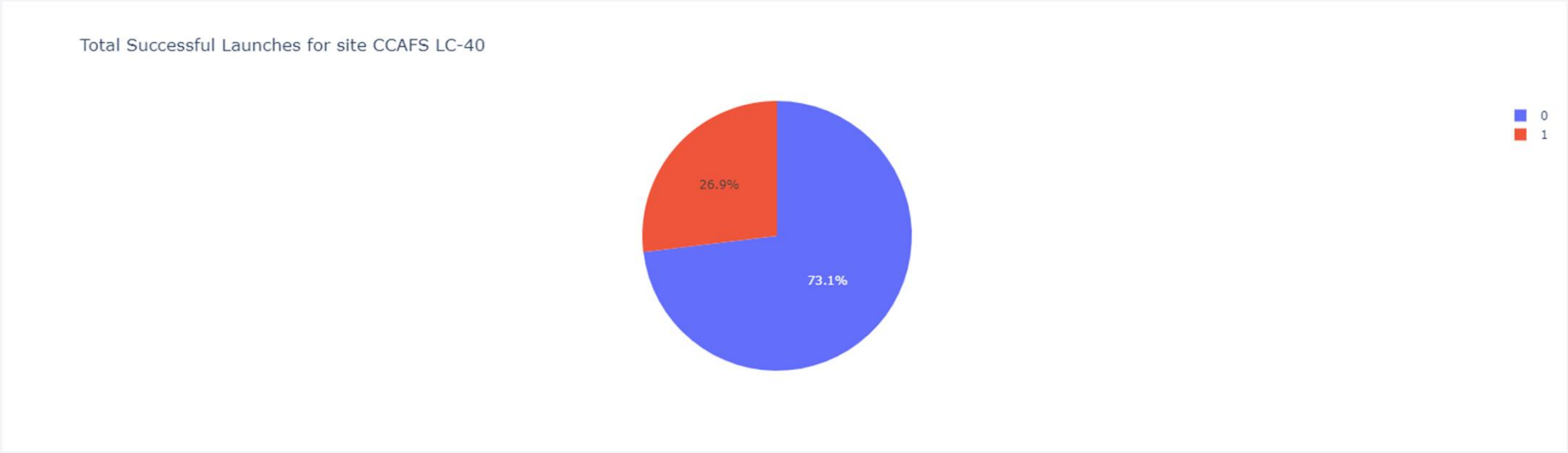
In [43]:
```
%sql SELECT MAX(PAYLOAD_MASS__KG_) AS MAX_PAYLOAD_MASS FROM SPACEXDATASET;
```

```
 * ibm_db_sa://vdx86203:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:31321/bludb
Done.
```

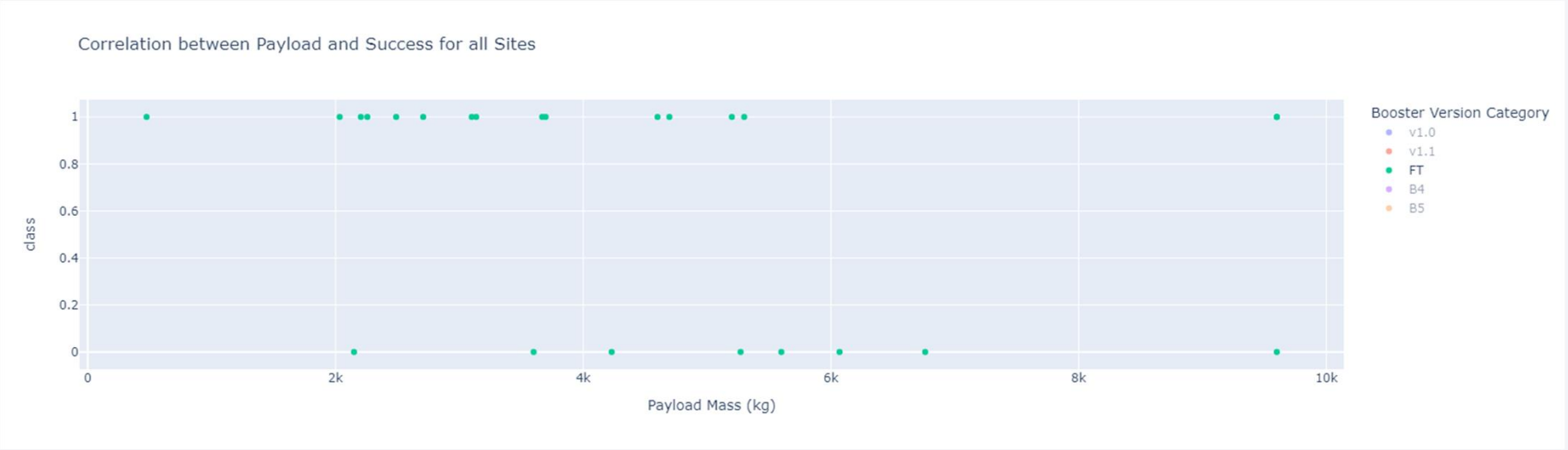
Out[43]:

max_payload_mass
15600

Appendix： Launch Site CCAFS LC-40 Success



Appendix： Correlation between Payload and Success ‘FT’



Thank you!