IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

John Mangiaracina

Mar 31, 2022

Outline

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

Executive Summary

- Accepted and state of the art Data Science and Machine Modeling techniques developed mathematical models for the success of launch vehicle recovery for the SpaceX booster program. Launch vehicle recovery is CRITICAL to reducing costs and ensuring SpaceX's leadership role in the industry.
- With a 77% success rate of recovery from the top performing launch sites , SpaceX has a clear advantage over all competitors.
- The Decision Tree Classifier Model was selected as the model with the highest level of accuracy over all models considered, at 87.5%, a very high level.
- The company is showing a generally increasing trend in success rate recovery over time. A very good sign that the organization is learning and implementing continuous improvement!

Introduction

- Unprecedented need to place Communication Satellites, Intelligence Gathering Satellites, Military Hardware, Satellites for the Needs of Pure Science, and Manned Spacecraft in Earth Orbit , Solar Orbit, and Beyond
- The Future holds incredible promise for Advanced Defense Systems and both Un-Manned and Manned Missions to Mars … and that is Just the Beginning!
- The Key Factor in advancing space exploration, travel, and economic development is having a launch cost structure that is FAR cheaper than all competition. This requires reusable First Stages and knowing the probability of reuse.

Section 1

Methodology

Methodology, part 1

Methodology Executive Summary

- Data collection methodology:
    - Web scraping techniques were utilized on the SpaceX Launches data site and the SpaceX Wikipedia page
- Data wrangling
    - Data wrangling and data cleaning were conducted in jupyter notebooks utilizing the python programming language with various open source tools
- Exploratory Data Analysis (EDA) was conducted using various data science open source libraries and queries were conducted with SQL

Methodology, part 2

Methodology Executive Summary

- Interactive visual analytics were conducted using Folium, Plotly, and Dash
- Predictive analysis using classification machine models
    - Cleaned and processed dataset of all launches classified by payload, launch site, and success/failure of reuse of first stage was split into testing and training datasets. Logistic Regression, Support Vector Machine, Decision Classifier Tree, and K Nearest Neighbors Machine Learning Models were Evaluated to find the best Model. We found the Decision Tree Classifier Model to be the most accurate.

Data Collection

- Data sets were collected from the SpaceX Launch data web site:

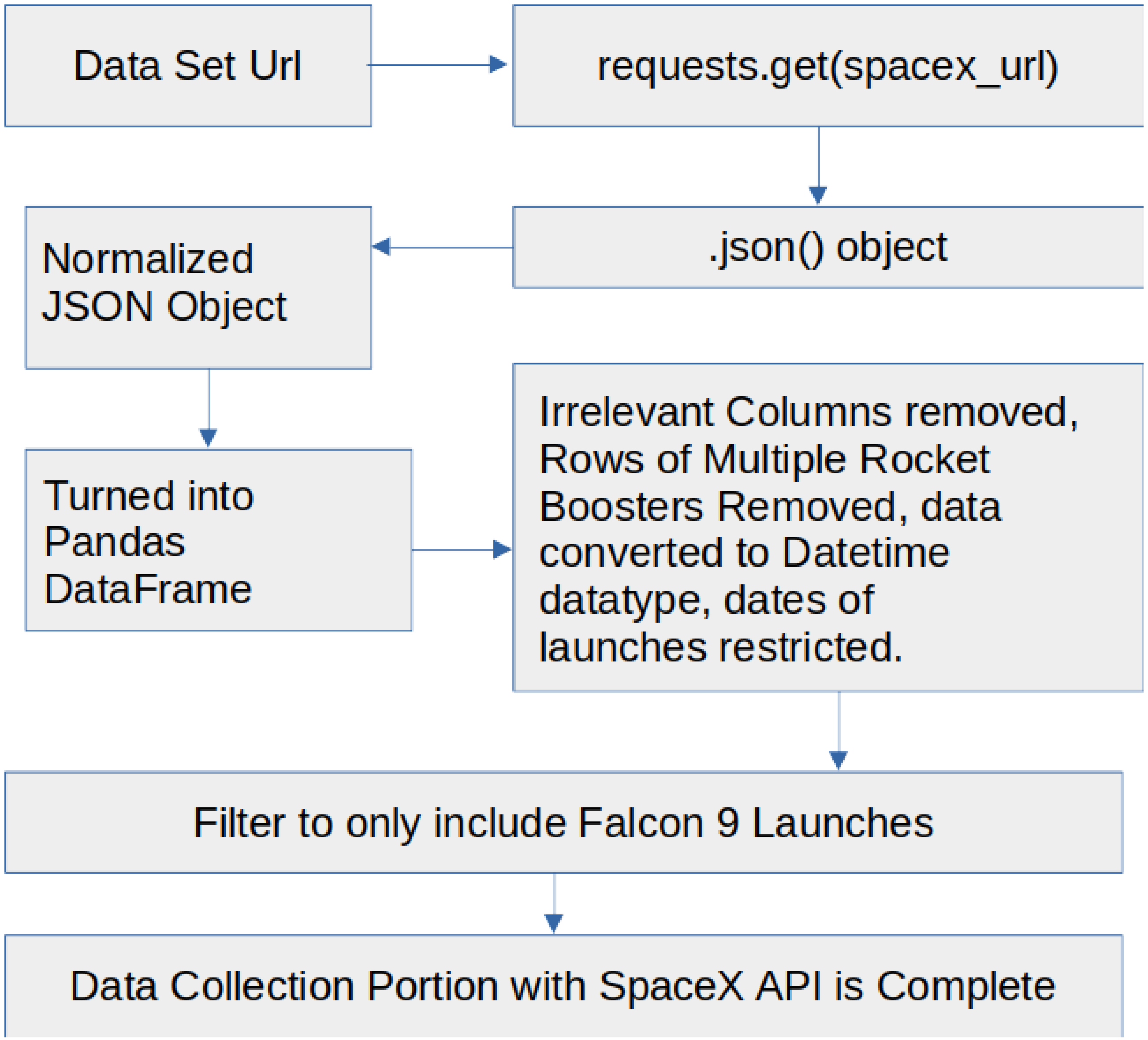
  "https://api.spacexdata.com/v4/launches/past"

- Falcon9 Data was scraped from the web site:

  "https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922"

Data Collection – SpaceX API

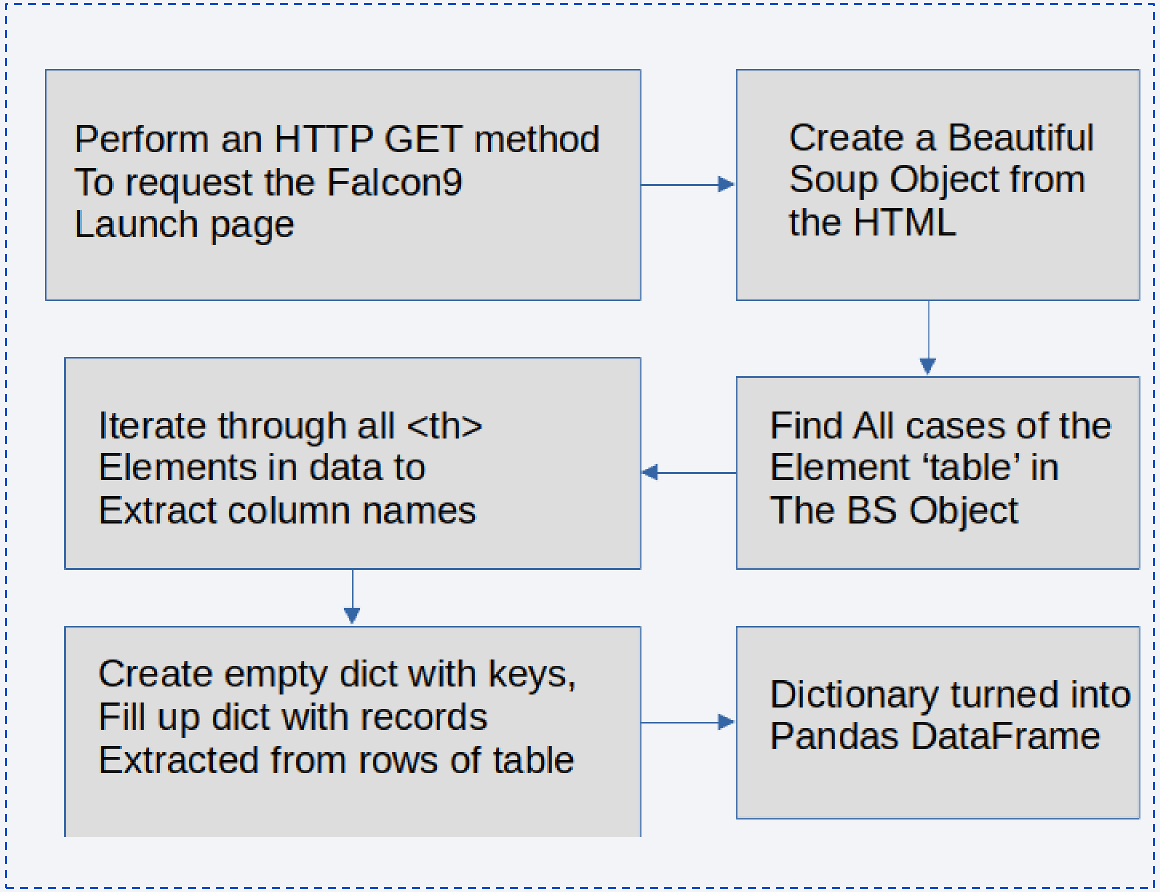
- Flowchart of Data Collection Process is to the right
- The Jupyter Notebook of the entire SpaceX Data Collection Process is available at:
  - https://github.com/EasyAs12345/coursera-capstone-project/blob/master/Completed_Lab_Data%20Collection%20API%20Lab.ipynb



Data Collection - Scraping

- Webscraping Process Flowchart is to the right
- The web scraping jupyter notebook is available at:

  https://github.com/EasyAs12345/coursera-capstone-project/blob/master/Completed_Lab_Data%20Collection%20with%20Web%20Scraping.ipynb



Data Wrangling

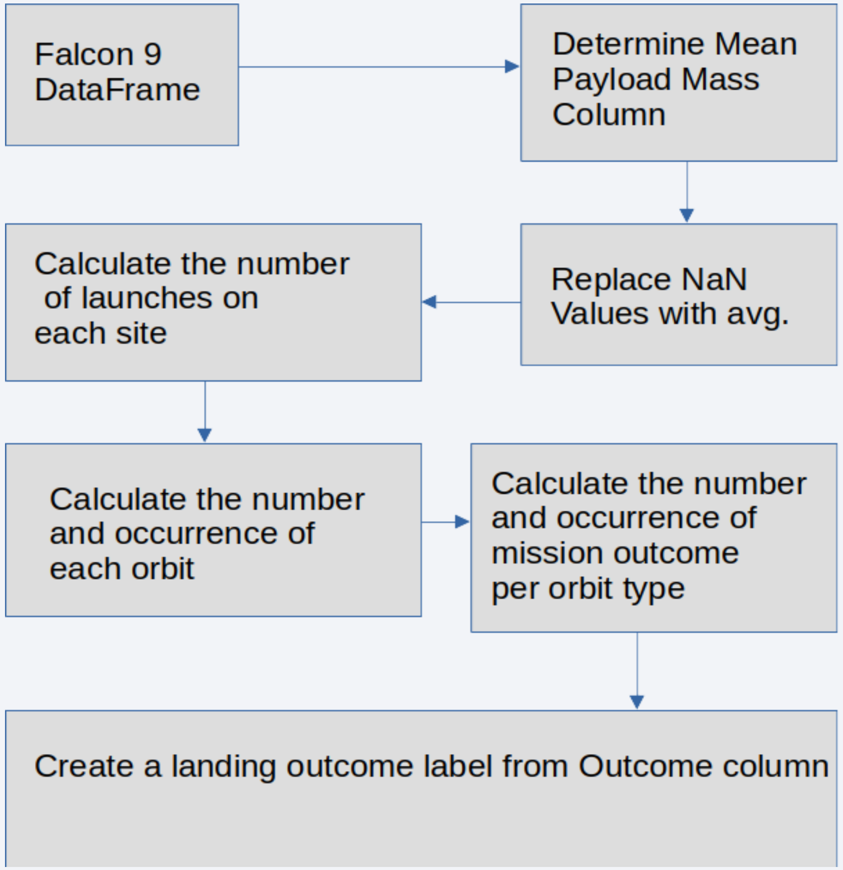
- Data Wrangling Focused on determining which key values of interest were missing in the dataset.
- We applied the .isnull() method to the Falcon9 data to determine this
- It was decided to apply the .mean() method to take the average of the data that does exist for a particular attribute, and then replace the missing data of a particular attribute with the mean of said attribute.
- Data Wrangling can be found in the Jupyter Notebook located at:

https://github.com/EasyAs12345/coursera-capstone-project/blob/master/Completed_Lab_Data%20Collection%20API%20Lab.ipynb

And

https://github.com/EasyAs12345/coursera-capstone-project/blob/master/Completed_Lab_EDA.ipynb

Data Wrangling Flow Chart



EDA with Data Visualization, Part 1

- Summary of Charts Plotted as part of EDA
  - Payload Mass (kg) vs. Flight Number: To see if the Rate of success of recovery trended over successive launches (if the system learned and got better), but also to see whether there was any change in the quantity of the payload over successive launches.
    - CCAFS LC-40, has a success rate of 60 %, while KSC LC-39A and VAFB SLC 4E have a success rate of 77%.
  - Launch Site vs. Flight Number: To see if there was a difference over time in success of launch recovery based on where the Rocket was launched from, and to see how launch site selection varied over time.
    - Launch Recovery Rate increase over time at VAFB and KSC, not as clear cut at CCAFS
  - Launch Site vs. Payload Mass: To see any preferences for Launch Site based on the amount of mass to be launched, and to see how success rate of launch vehicle recovery varied by launch site for groupings of payload mass.
    - No Launches above 10KG from VAFB, slight advantage of successful recovery with very high Payload mass at CCAFS

EDA with Data Visualization, Part 2

- Summary of Charts Plotted as part of EDA, cont.
  - Orbit Class vs. Orbit: To see the sucess rate of each type of Orbit.
    - ES-L1, SSO, HEO, and GEO had the highest success rate of 1.0 = 100%!
  - Orbit vs. Flight Number: To see variation over time of actual orbit of payload over time, and also success of launch vehicle recovery over time (Flight Number).
    - With Geo and MEO, greater success over time. Others see no difference over time.
  - Orbit vs. PayloadMass: To see any trends in Orbit selection based on differences in the Mass of the Payload (a larger mass requires more Energy to drive it higher in the atmosphere before reaching outer space.)
    - Polar, LEO, and ISS Orbits, when launching heavy payloads, have a greater success rate of successful launch vehicle recovery

EDA with Data Visualization, Part 3

- Summary of Charts Plotted as part of EDA, cont.
  - Success Rate vs Time: To see if Launch Vehicle Recovery improved or not over time. To see if the organization is getting better, worse, or neutral over time.
    - Overall, the success rate of launch vehicle recovery has been on a trending upward curve from 2013 – 2020!
    - Very good news for SpaceX as a whole … although always room for improvement!
- The complete EDA with data visualization notebook is available at:

  https://github.com/EasyAs12345/coursera-capstone-project/blob/master/Completed_Lab%20Complete%20the%20EDA%20with%20Visualization%20lab.ipynb

EDA with SQL, Part 1

- Utilizing SQL queries, I explored:
  - What are the unique launch sites that were used:
    - CCAFS LC-40, CCAFS SLC-40, KSC LC-39A, and VAFB SLC-4E
  - To see launch sites that begin with the string 'CCA'
  - Display the total payload mass carried by boosters launched by NASA (CRS)
    - 48213 kg.
  - Display average payload mass carried by booster version F9 v1.1
    - 3890 kg.
  - List the date when the first successful landing outcome in ground pad was achieved.
    - 22 Dec 2015

EDA with SQL, Part 2

- Utilizing SQL queries, I explored (cont):
  - List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

booster_version	landing__outcome	payload_mass__kg_
F9 FT B1022	Success (drone ship)	4696
F9 FT B1026	Success (drone ship)	4600
F9 FT B1021.2	Success (drone ship)	5300
F9 FT B1031.2	Success (drone ship)	5200

  - List the total number of successful and failure mission outcomes (No aspect of success at all)
    - 100 Successful
    - 1 Unsuccessful

EDA with SQL, Part 3

- Utilizing SQL queries, I explored (cont):
    - List the names of the booster versions which have carried the maximum payload mass. (List on the right)

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

EDA with SQL, Part 4

- Utilizing SQL queries, I explored (cont):
  - List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

landing__outcome	booster_version	launch_site	DATE
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40	2015-01-10
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40	2015-04-14

EDA with SQL, Part 5

- Utilizing SQL queries, I explored (cont):
  - 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

number	landing__outcome	ranking
10	No attempt	1
5	Failure (drone ship)	2
5	Success (drone ship)	2
3	Controlled (ocean)	4
3	Success (ground pad)	4
2	Failure (parachute)	6
2	Uncontrolled (ocean)	6
1	Precluded (drone ship)	8

EDA with SQL, Part 6

- Utilizing SQL queries, I explored (cont):
  - List Launches launched on a Friday

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)
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt
2014-04-18	19:25:00	F9 v1.1	CCAFS LC-40	SpaceX CRS-3	2296	LEO (ISS)	NASA (CRS)	Success	Controlled (ocean)
2016-03-04	23:35:00	F9 FT B1020	CCAFS LC-40	SES-9	5271	GTO	SES	Success	Failure (drone ship)
2016-04-08	20:43:00	F9 FT B1021.1	CCAFS LC-40	SpaceX CRS-8	3136	LEO (ISS)	NASA (CRS)	Success	Success (drone ship)

  - Jupyter Notebook is available at:

    https://github.com/EasyAs12345/coursera-capstone-project/blob/master/Completed_Lab%20EDA%20with%20SQL.ipynb

Build an Interactive Map with Folium

- A series of map objects were added to a folium map
  - site_map.add_child(circle) was used to add a circle to the map for NASA in Texas, as well as the three launch sites in Florida and the one launch site in California
  - site_map.add_child(marker) was used to add markers such as text designations for the various sites
  - site_map.add_child(lines) was used to indictate a line marking the distance of the Vandenburg launch site to the Ocean shore, as well as from the launch site to the US Forest Service Jet Center in nearby Santa Maria, CA
- These indicators were added to the map to greater clarify and explain the exact locations of the launch sites discussed in the study. Also, to demonstrate the capability, such as determining distances based of (lat, log) data and then drawing a line to indicate said distance (as the crow flies.) Very cool!
- The GitHub URL is available at:

  https://github.com/EasyAs12345/coursera-capstone-project/blob/master/Completed_Lab_Interactive%20Visual%20Analytics%20with%20Folium%20lab.ipynb

Build a Dashboard with Plotly Dash

- A Dashboard was created entitled "SpaceX Launch Records Dashboard".
    - The Dashboard includes a dropdown menu to select either all launch sites for consideration, or the option to choose an individual launch site.
    - The first graphic is a pie chart showing the percentage breakdown for total successes for launches based on site option from the dropdown.
    - There is a slide to choose the total Payload range mass (0 – 10,000 kg)
    - The bottom graphic is a scatter plot showing Orbit Class vs Payload Mass. Launch site data is selected based on the top dropdown. This allows us to see the success (1) or failure (0) of the mission based on payload mass. Color coding is used to indicate the booster version category.
- This visual is very helpful in allowing a user to see how success and failure of launch vehicle recovery varies based on the total mass of the payload and based on launch category. Being able to explore this helps understand where the greatest success is and where efforts need to be focused to improve.

Build a Dashboard with Plotly Dash, part 2

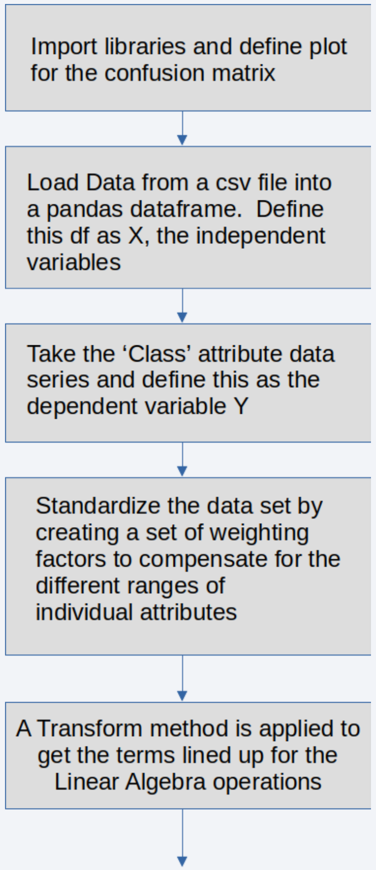
- A copy of the Python source code for this is available at:

  https://github.com/EasyAs12345/coursera-capstone-project/blob/master/spacex_launch_dashboard_code.py

  Note: It has been tested on both the IBM Skills Network Docker based setup and on a Workstation running Linux Fedora with the latest version of the Anaconda Data Science developers suite.

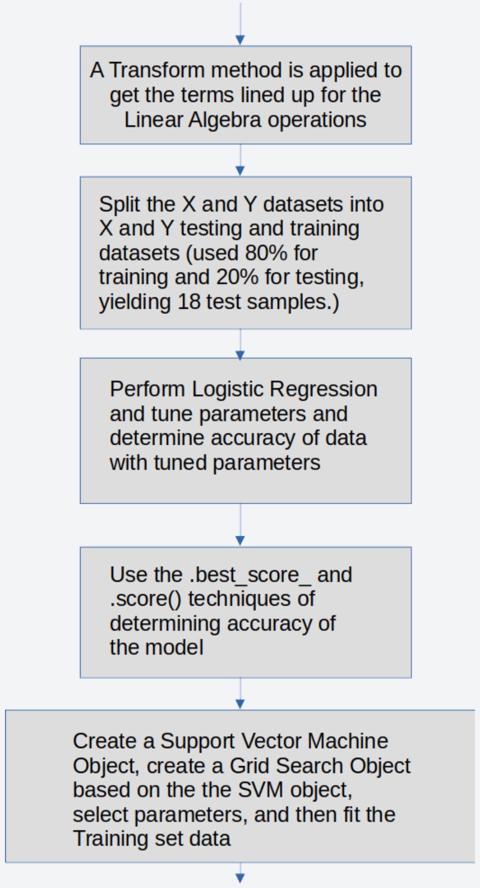
Predictive Analysis (Classification), part 1

- The top portion of the Flow Chart is shown on the right and on successive pages
- We imported necessary libraries, define out confusion matrix plot function, loaded the dataset from a csv file, turned it into a dataframe, broke it out to define dependent and independent variables.
- We standardized the dataset by depending appropriate weighting factors for the attributes, and conducted a transform.



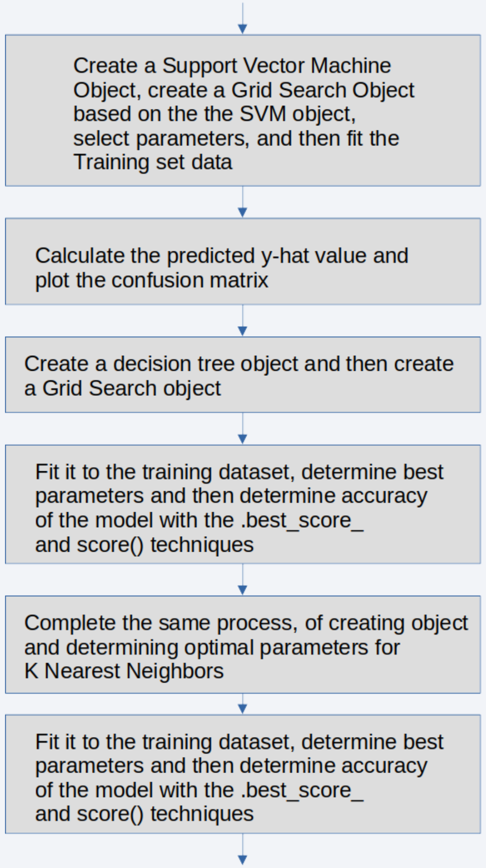
Predictive Analysis (Classification), part 2

- We split the dependent and independent variables into training and testing datasets
- We then performed a Logistic Regression as the first of our series of Algorithms.
- We tuned parameters, found best case, and then determined the accuracy with two different techniques.
- We same approach was then repeated for Support Vector Machine



Predictive Analysis (Classification), part 3

- The same approach was then taken for Decision Tree and for K Nearest Neighbors



Predictive Analysis (Classification), part 4

- Based on the dataset and our breakdown of training/testing data, we determined the Decision Tree Classifier had the highest accuracy, at 87.5%, a very acceptable number.
- The jupyter notebook for this analysis is available at:

https://github.com/EasyAs12345/coursera-capstone-project/blob/master/Completed_Machine%20Learning%20Prediction%20Project%20Capstone%20Course%20FINAL.ipynb

Fit it to the training dataset, determine best parameters and then determine accuracy of the model with the .best_score_ and score() techniques

Determine the method that is best. My analysis showed Decision Tree classifier was best with an accuracy of 0.875. This says 87.5% of the time the model is accurate, which is very acceptable.

Results

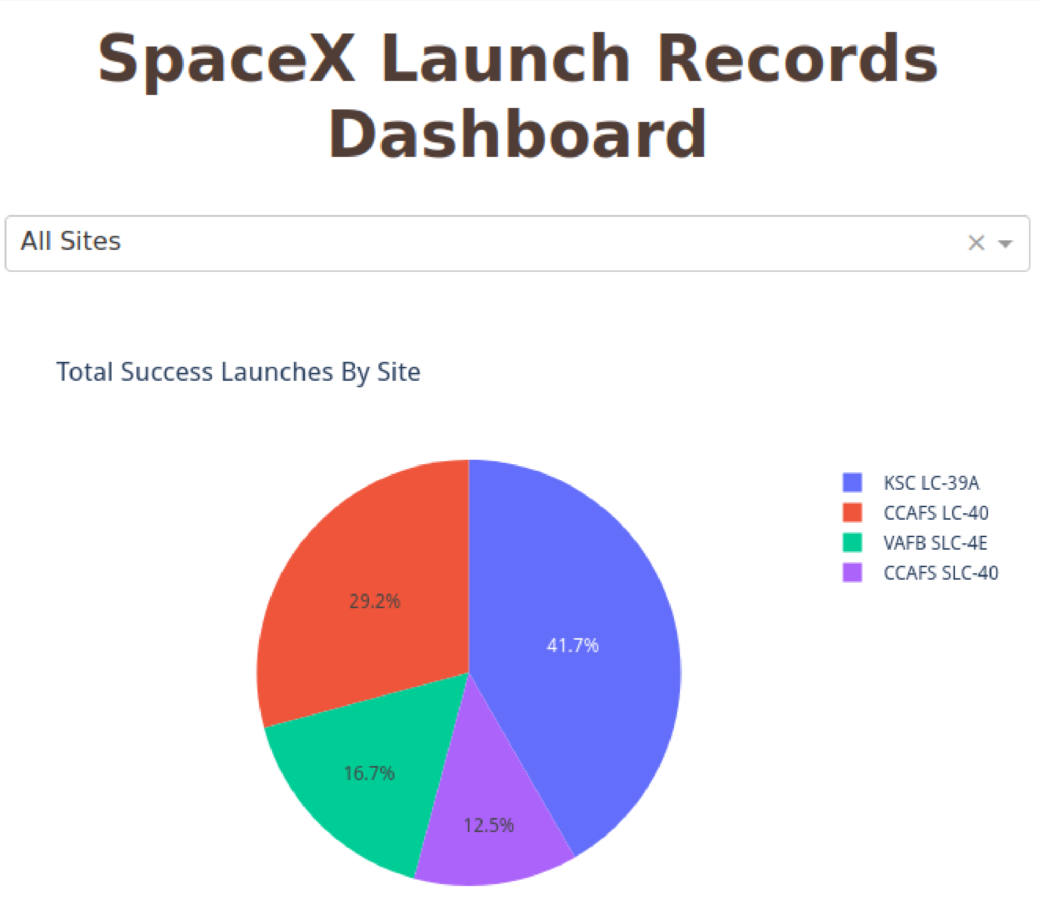
- Summary of results from the EDA:
  - CCAFS LC-40, has a success rate of 60 %, while KSC LC-39A and VAFB SLC 4E have a success rate of 77%.
  - Launch Recovery Rate increase over time at VAFB and KSC, not as clear cut at CCAFS
  - No Launches above 10KG from VAFB, slight advantage of successful recovery with very high Payload mass at CCAFS

Results, cont.

- Summary of results from the EDA (cont):
  - ES-L1, SSO, HEO, and GEO had the highest success rate of 1.0 = 100%!
  - With Geo and MEO, greater success over time. Others see no difference over time.
  - Polar, LEO, and ISS Orbits, when launching heavy payloads, have a greater success rate of successful launch vehicle recovery

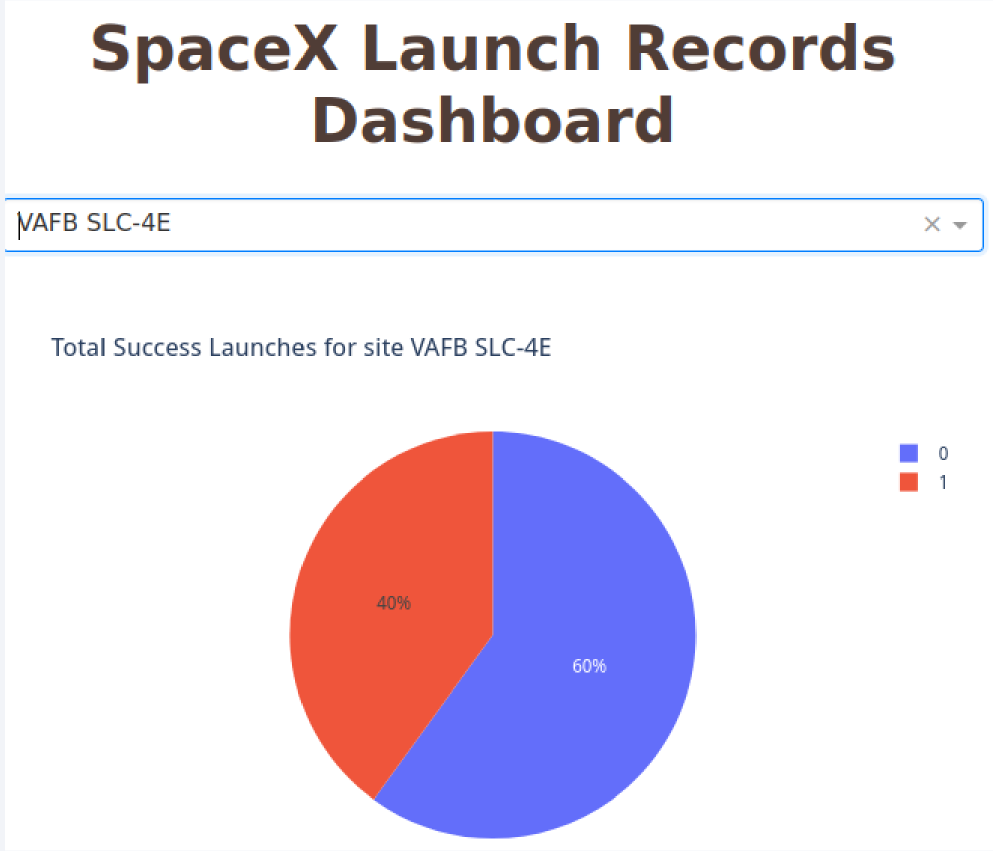
Results, cont.

- Interactive analytics demo screenshot of the SpaceX Launch Records Dashboard, ALL Sites selected from the dropdown.



Results, cont.

- Interactive analytics demo screenshot of the SpaceX Launch Records Dashboard, Vandenburg VAFB selected from the dropdown.

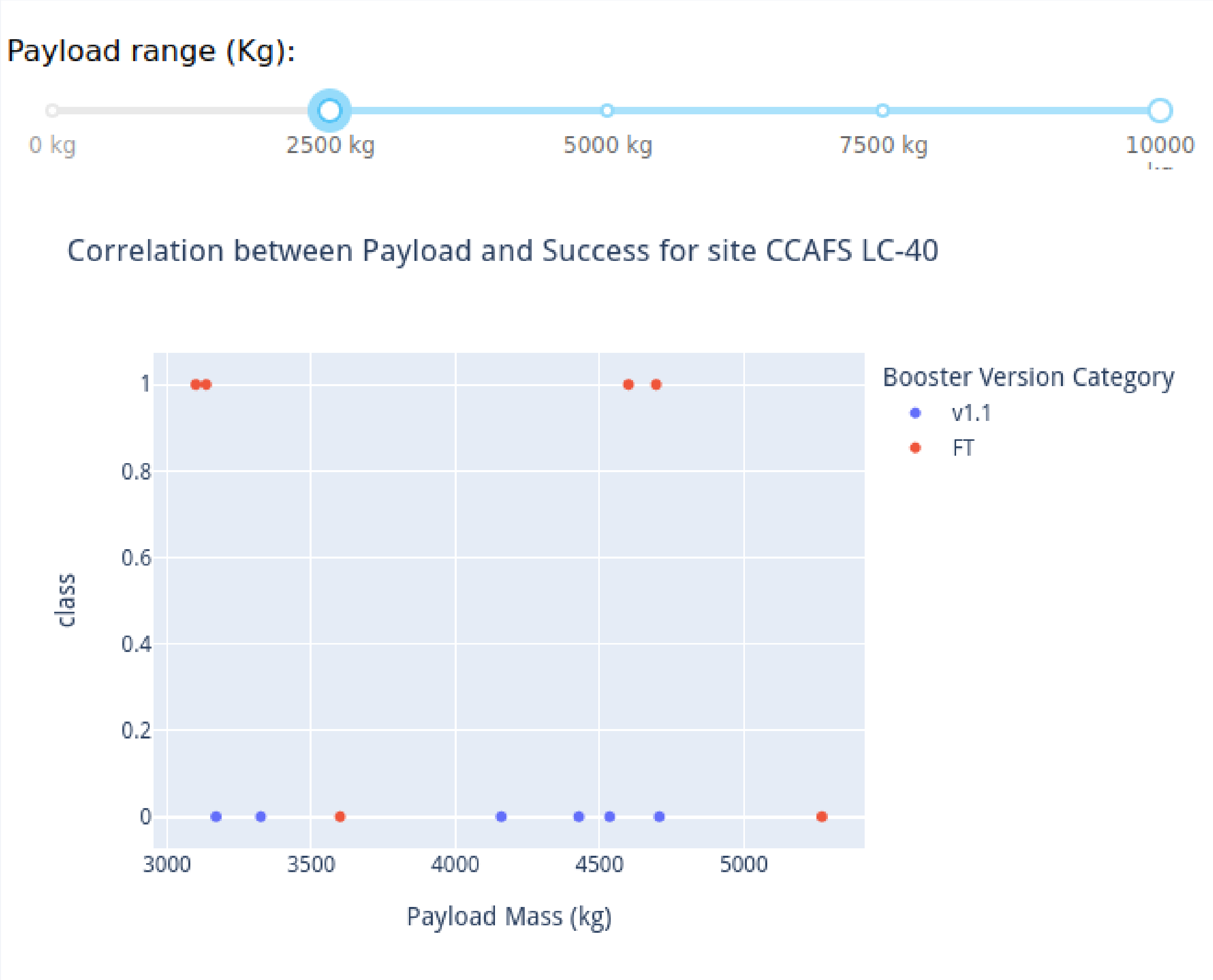


Results, cont.

- Interactive analytics demo screenshot of the SpaceX Launch Records Dashboard, All Sites selected from the dropdown, showing the Full Slider Range and the scatter plot for correlation between Payload and Success for ALL Sites, color coded by booster version

Results, cont.

- Interactive analytics demo screenshot of the SpaceX Launch Records Dashboard, CCAFS LC-40 selected from the dropdown, showing the 2500kg to 10000 kg and the scatter plot for correlation between Payload and Success for only CCAFS LC-40, color coded by booster version



Results, cont.

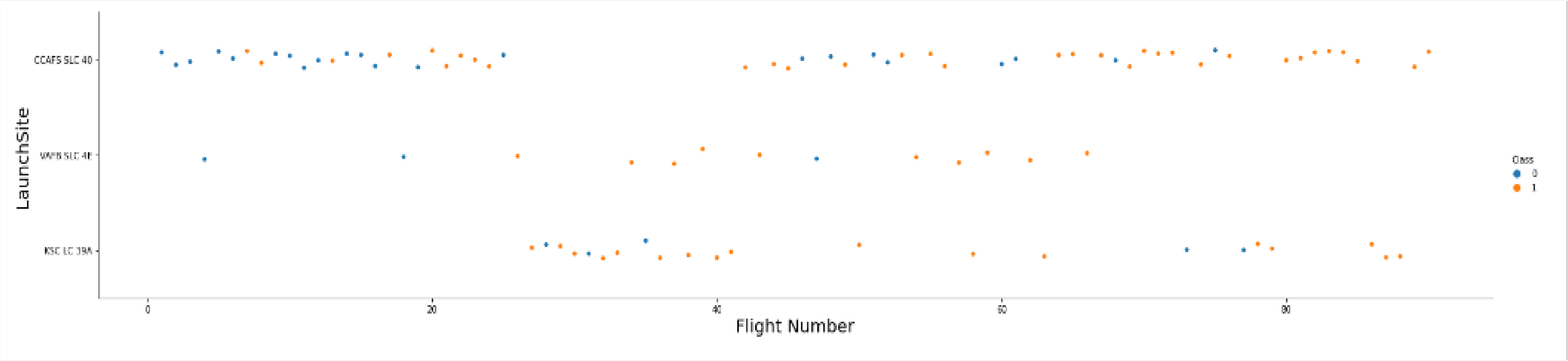
- Based on the dataset and our breakdown of training/testing data, the Predictive Analytics determined the Decision Tree Classifier had the highest accuracy, at 87.5%, a very acceptable number.

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

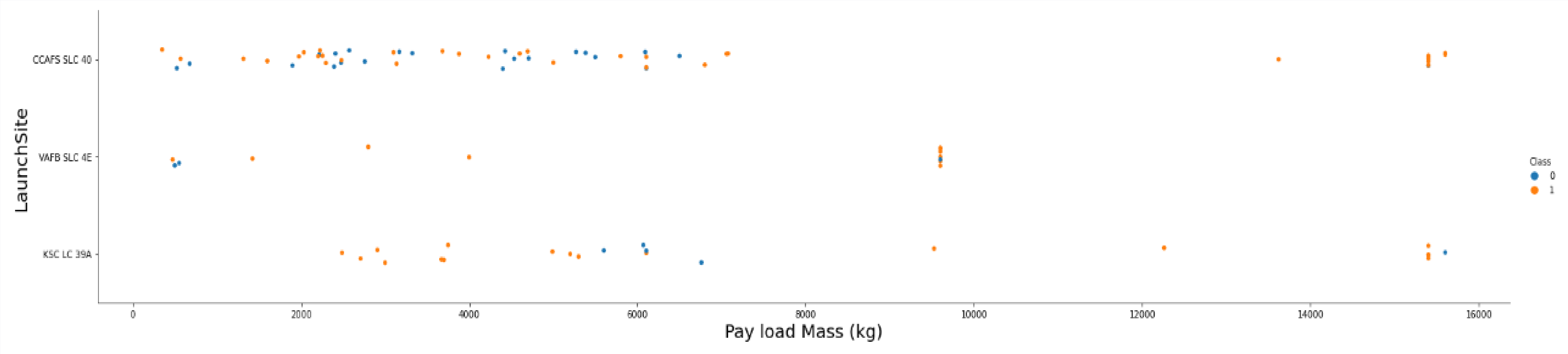
- Scatter plot of Flight Number vs. Launch Site



This Scatter Plot groups launches by launch site and displays it over time, while color coding the success (1) or failure (0) of recovery

Payload vs. Launch Site

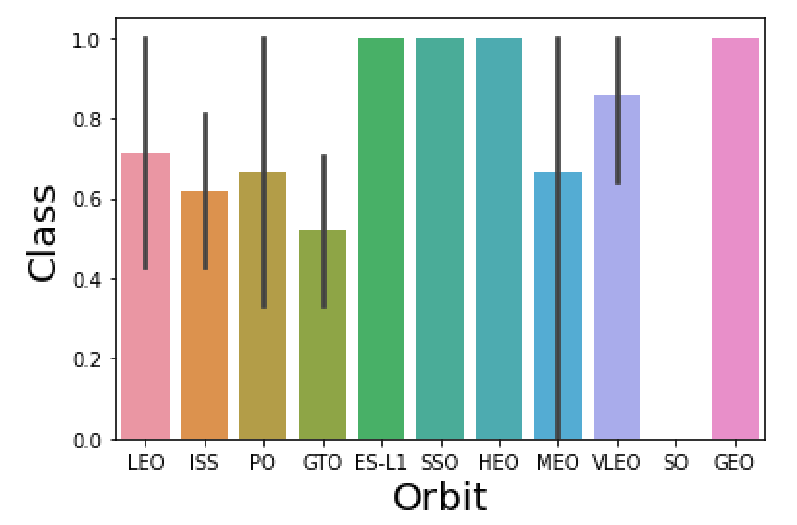
- Scatter plot of Payload vs. Launch Site



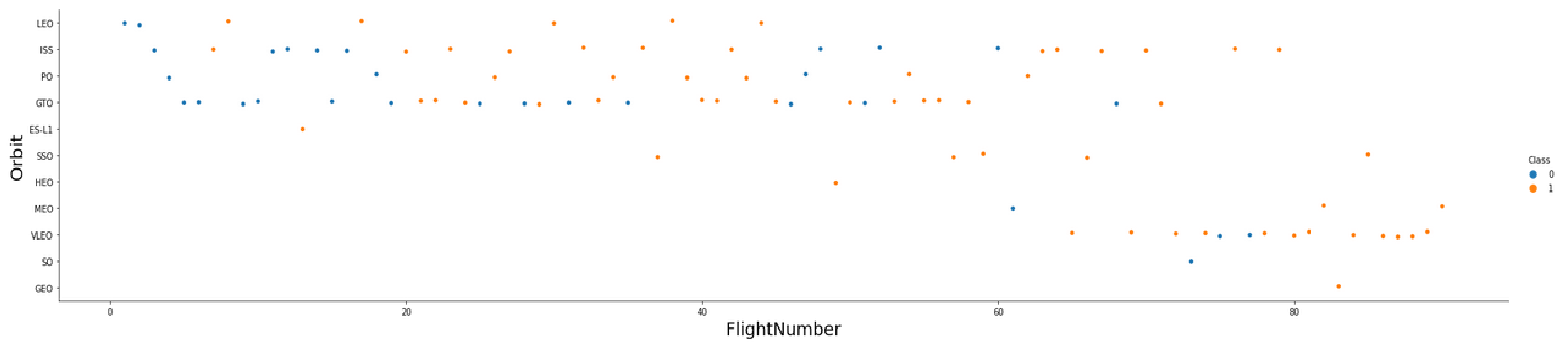
- This Scatter plot of Launch Site vs. Payload Mass groups missions by Payload across the different launch sites, with color coding by success (1) or failure (0) of recovery.

Success Rate vs. Orbit Type

Bar chart for the success rate of each orbit type:

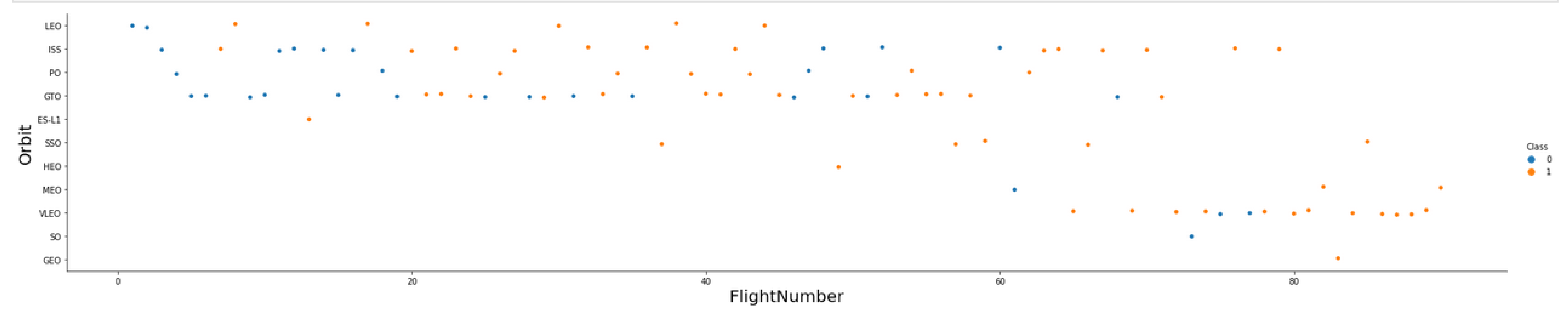


Scatter plot of Flight Number vs. Launch Site:



Flight Number vs. Orbit Type

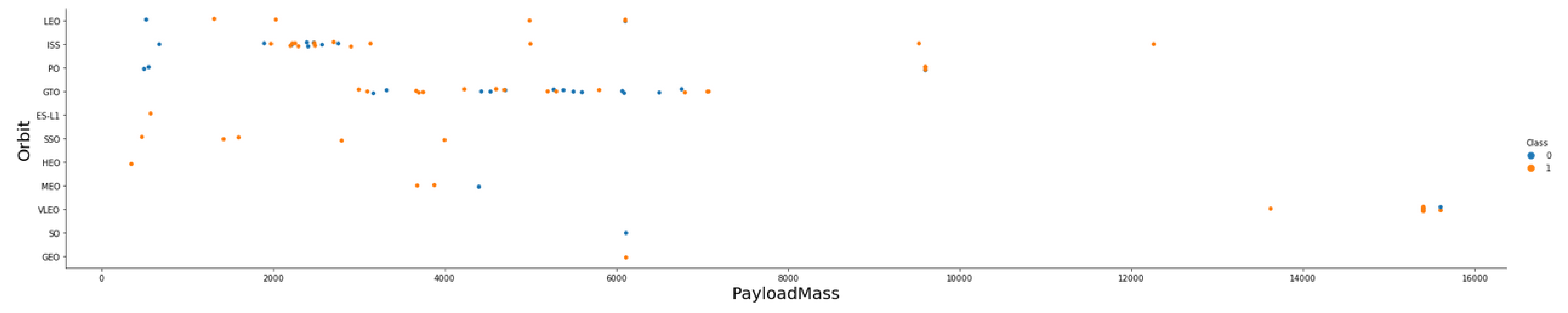
Scatter plot of Flight number vs. Orbit type



This scatter plot shows us how the success of launch vehicle recovery varied over time (Flight Number), as well as selection for different Orbit Class varied over time (Flight Number). This allows us to see quickly how trends may have appeared of greater degrees of success over time for each of the Orbits!

Payload vs. Orbit Type

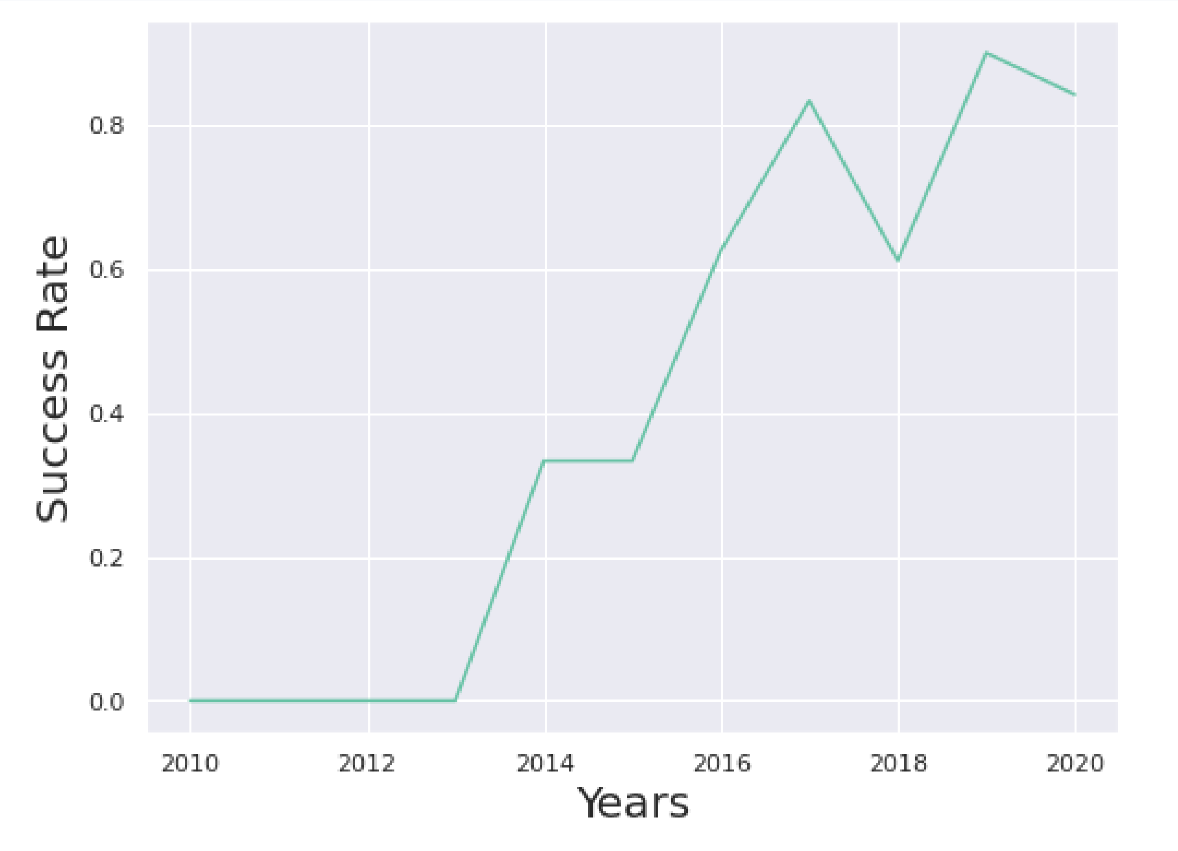
- Scatter Plot of Payload vs. Orbit type



This plot shows us a very interesting grouping of ranges of payload masses for the different orbits that said masses are launched into. There appears to be no payloads launched into the GTO orbit above 7000 kg, whereas the VLEO (Very Low Earth Orbit) seems to have all payloads above 13000 kg. This makes sense since for a given amount of propulsion, a greater mass would not be able to be launched as high in the atmosphere!

Launch Success Yearly Trend

Line chart of yearly average success rate is shown at the right:



The plot shows how there is a generally overall uptrend in success rate from 2013 – 2020, with some hiccups.

This shows the organization is learning and improving their technology and processes and is a good overall sign for SpaceX.

All Launch Site Names

- Find the names of the unique launch sites
  - The unique launch sites are in the table to the right
- The SQL code to determine this is:

```
[5]: %%sql

SELECT UNIQUE(LAUNCH_SITE) FROM SPACEXDATASET;
```

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

- This SQL code will return the launch site names from the table of data. The UNIQUE command will assure that there are no duplicates returned.

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`:

Display 5 records where launch sites begin with the string 'CCA'

[6]:
```
%%sql

SELECT *
FROM SPACEXDATASET
WHERE LAUNCH_SITE LIKE 'CCA%' LIMIT 5;
```

```
 * ibm_db_sa://nfk04348:***@b70af05b-76e4-4bca-a1f5-23dbb4c6a74e.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32716/bludb
Done.
```

[6]:

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

- This query returns a limit of 5 records where the launch site entry starts with the string 'CCA'

Total Payload Mass

- Calculate the total payload carried by boosters from NASA:

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

[8]:
```
%%sql

SELECT SUM(payload_mass__kg_) AS Total_payload_mass
FROM SPACEXDATASET
WHERE CUSTOMER LIKE '%CRS%';
```

```
 * ibm_db_sa://nfk04348:***@b70af05b-76e4-4bca-a1f5-23dbb4c6a74e.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32716/bludb
Done.
```

[8]:

total_payload_mass
48213

- This query adds up all of the payload masses which contains the customer designation 'CRS' for NASA and determined it was 48,213 kg.

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1:

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

[9]:
```
%%sql

SELECT AVG(payload_mass__kg_) AS average_payload_mass
FROM SPACEXDATASET
WHERE BOOSTER_VERSION LIKE '%F9%1.1%';
```

```
 * ibm_db_sa://nfk04348:***@b70af05b-76e4-4bca-a1f5-23dbb4c6a74e.clogj3sd0tgtu0lqde00.databases.appdomain.cloud:32716/bludb
Done.
```

[9]:

average_payload_mass
3890

- The average payload mass for the designated booster was 3890 Kg.

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad:

```
%%sql

SELECT MIN(date) AS first_successful_landing_outcome
FROM SPACEXDATASET
WHERE LANDING__OUTCOME LIKE '%Success (ground pad)%';
```

```
 * ibm_db_sa://nfk04348:***@b70af05b-76e4-4bca-a1f5-23
Done.
```

first_successful_landing_outcome
2015-12-22

- The minimum (lowest number) date for a successful ground pad landing was 22 Dec 2015. We used the like command to with wild card characters to capture any strings that contained the key phrase.

Successful Drone Ship Landing with Payload between 4000 and 6000

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

```
%%sql

SELECT BOOSTER_VERSION, LANDING__OUTCOME, payload_mass__kg_
FROM SPACEXDATASET
WHERE LANDING__OUTCOME LIKE '%Success (drone ship)%' AND payload_mass__kg_ > 4000 AND payload_mass__kg_ < 6000;
```

```
 * ibm_db_sa://nfk04348:***@b70af05b-76e4-4bca-a1f5-23dbb4c6a74e.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:
Done.
```

booster_version	landing__outcome	payload_mass__kg_
F9 FT B1022	Success (drone ship)	4696
F9 FT B1026	Success (drone ship)	4600
F9 FT B1021.2	Success (drone ship)	5300
F9 FT B1031.2	Success (drone ship)	5200

- This query is similar to others before, we just this time use logical AND statements and < or > signs to limit the payload mass to specified range

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes to the right:
- Here we just count the number of times the mission outcome includes success to determine the #, which is 100 success. We then just negate this to obtain the 1 unsuccessful outcome (this criteria includes ANY and ALL forms that can be defined as success.

```
%%sql

SELECT COUNT(MISSION_OUTCOME) AS TOTAL_NUMBER_of_SUCCESSFUL_OUTCOMES
FROM SPACEXDATASET
WHERE MISSION_OUTCOME LIKE '%Success%';
```

```
 * ibm_db_sa://nfk04348:***@b70af05b-76e4-4bca-a1f5-23dbb4c6a74e.c1ogj3
Done.
```

total_number_of_successful_outcomes
100

```
%%sql

SELECT COUNT(MISSION_OUTCOME) AS TOTAL_NUMBER_of_UNSUCCESSFUL_OUTCOMES
FROM SPACEXDATASET
WHERE MISSION_OUTCOME NOT LIKE '%Success%';
```

```
 * ibm_db_sa://nfk04348:***@b70af05b-76e4-4bca-a1f5-23dbb4c6a74e.c1ogj3
Done.
```

total_number_of_unsuccessful_outcomes
1

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass:
- Here we determined the maximum payload that was carried, and then used this to limit and only match the attribute payload_mass_kg values that matched this. The results are shown to the right.

```
%%sql

SELECT BOOSTER_VERSION, payload_mass__kg_ from SPACEXDATASET
WHERE (SELECT max(payload_mass__kg_) FROM SPACEXDATASET) = payload_mass__kg_
```

```
 * ibm_db_sa://nfk04348:***@b70af05b-76e4-4bca-a1f5-23dbb4c6a74e.clogj3sd0tgt
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

- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015
- Here we do a basic query where we limit the returned values to landing outcome that meets the drone ship requirement, as well as in the year 2015

```
%%sql

SELECT LANDING__OUTCOME, BOOSTER_VERSION, LAUNCH_SITE, date
from SPACEXDATASET
WHERE (LANDING__OUTCOME = 'Failure (drone ship)') AND (DATE LIKE '%2015%');
```

```
 * ibm_db_sa://nfk04348:***@b70af05b-76e4-4bca-a1f5-23dbb4c6a74e.c1ogj3sd0tg
Done.
```

landing__outcome	booster_version	launch_site	DATE
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40	2015-01-10
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40	2015-04-14

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 2017-03-20, in descending order
- Here we limit the date range and group data by the landing outcome, in descending order. I used the RANK() command over the Order by statement to get the cool ranking

```
%%sql

SELECT COUNT(*) AS number, LANDING__OUTCOME,
       RANK() OVER (Order BY COUNT(*) DESC) as ranking
FROM SPACEXDATASET S
WHERE S.DATE >= '2010-06-04' AND
      S.DATE <= '2017-03-20'
GROUP BY S.LANDING__OUTCOME
ORDER BY COUNT(*) DESC;
```

```
 * ibm_db_sa://nfk04348:***@b70af05b-76e4-4bca-a1f5-23dbb
Done.
```

number	landing__outcome	ranking
10	No attempt	1
5	Failure (drone ship)	2
5	Success (drone ship)	2
3	Controlled (ocean)	4
3	Success (ground pad)	4
2	Failure (parachute)	6
2	Uncontrolled (ocean)	6
1	Precluded (drone ship)	8

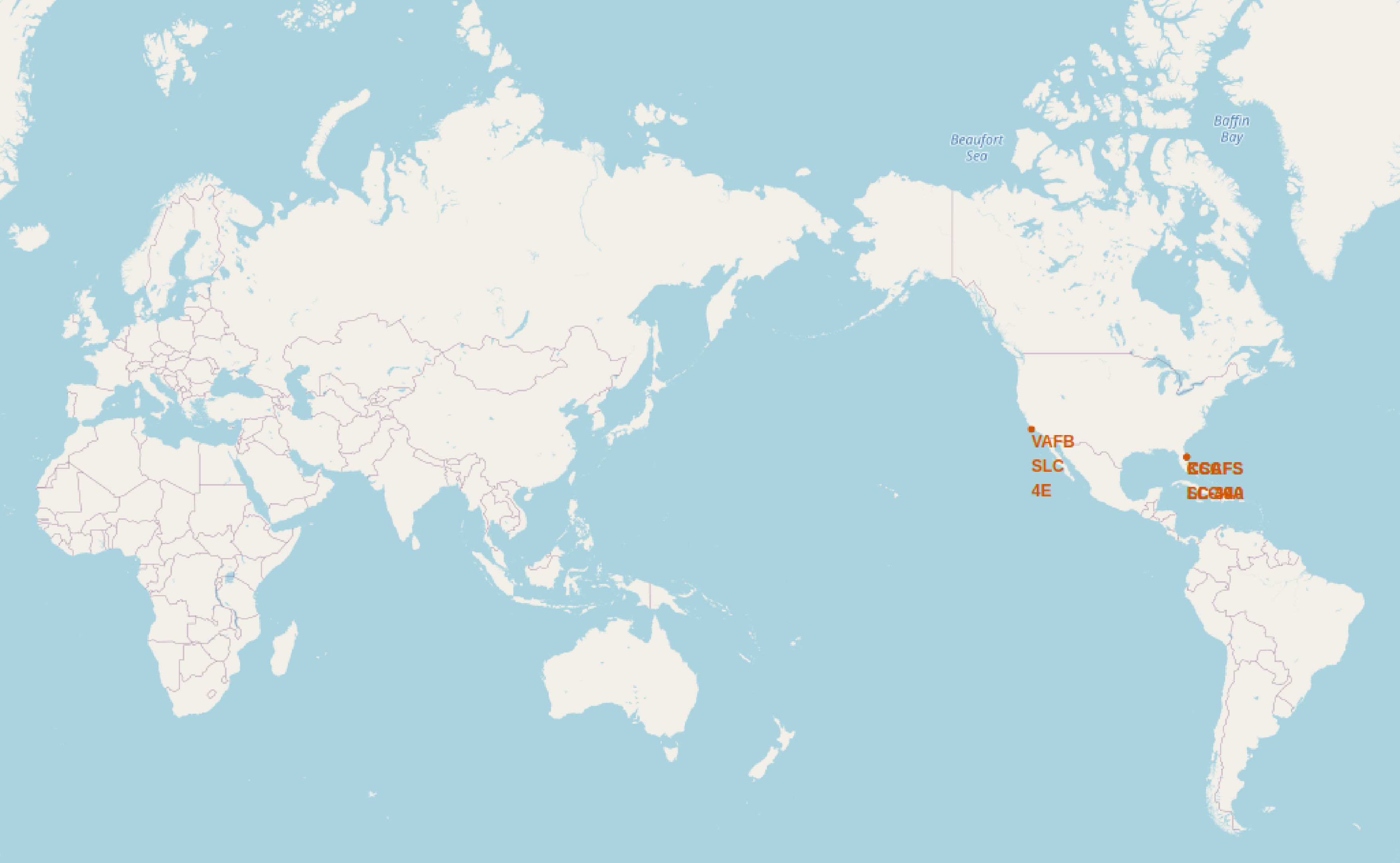
Section 3

Launch Sites Proximities Analysis

Folium Map Screenshot launch sites' with all location markers on a global map

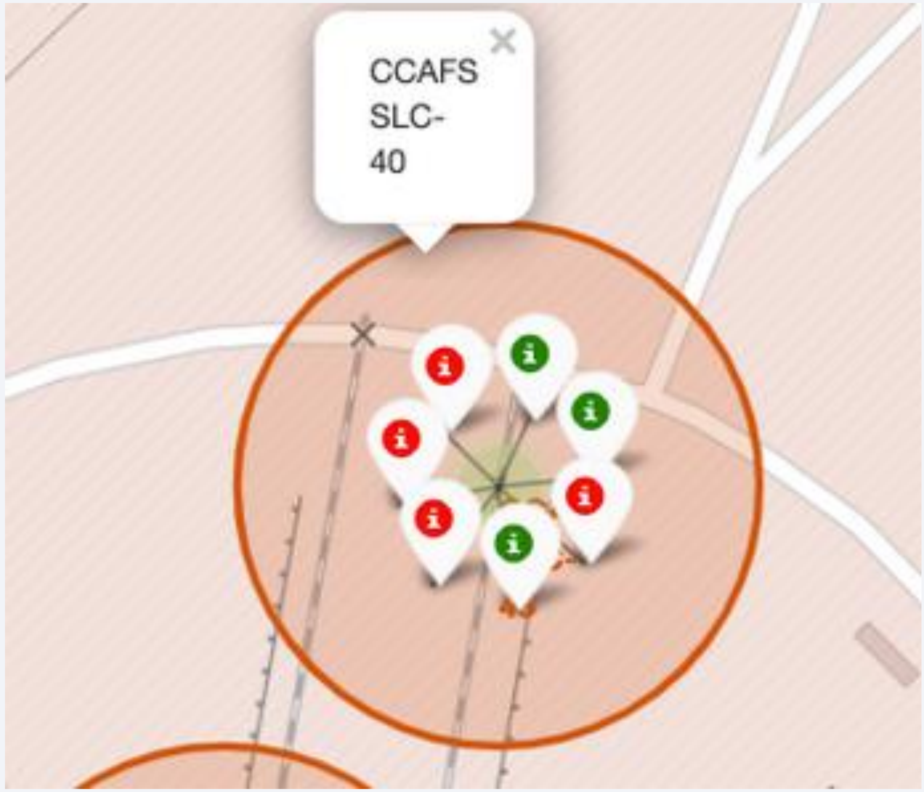
Folium map with all launch sites' location markers on a global map (right):

- Our map shows the four launch location, one in CA and three in Fl. Note that all launch sites are in the 28.5 deg to 34.7 deg lat range



Folium Map with an example of the color-labeled launch outcomes

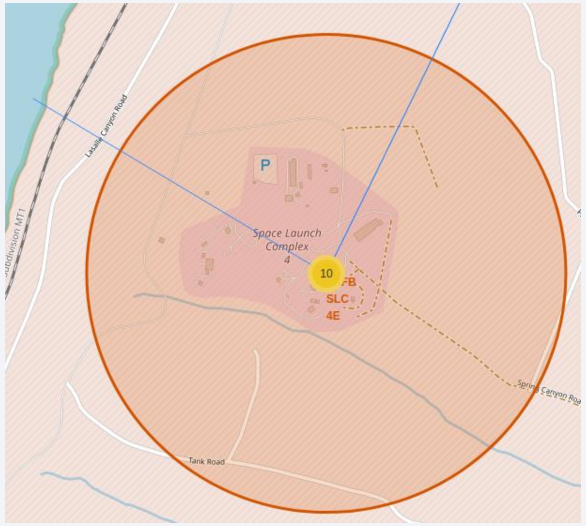
- A folium map with an example of the color-labeled launch outcomes
- This example uses markers to demonstrate red for failure, green for success for individual launches from the CCAFS SLC-40 launch site



Folium map showing lines from Vandenburg launch to the coastline and to the US Forest Service Jet Center in Santa Maria, CA

Folium map showing lines from Vandenburg launch to the coastline and to the US Forest Service Jet Center in Santa Maria, CA

- Distance from Vandenburg Launch to Land's End 1.427 km.
- Distance from Vandenburg Launch to USFS Santa Maria Air Base 32.116 KM
-

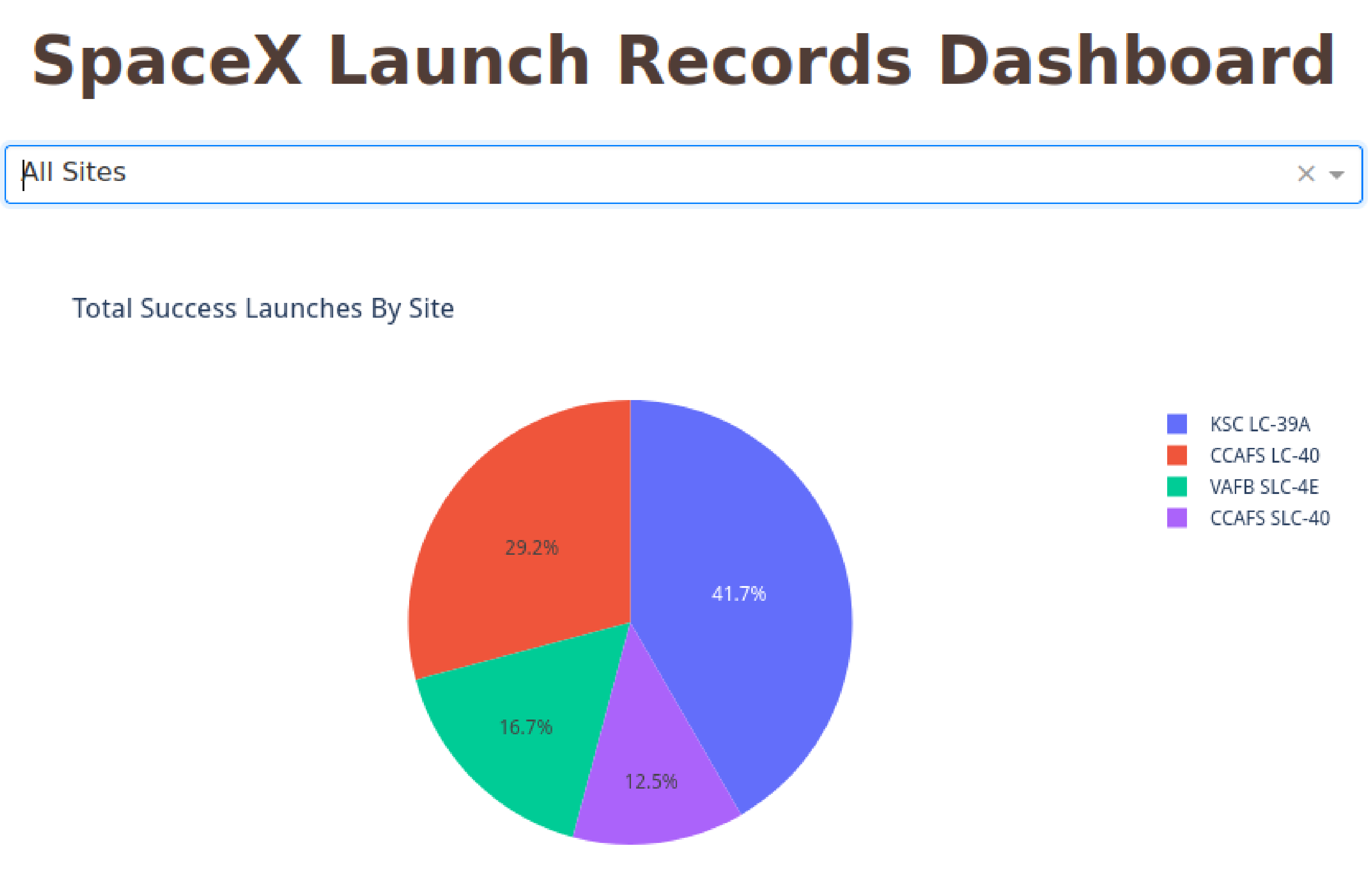


Section 4

Build a Dashboard with Plotly Dash

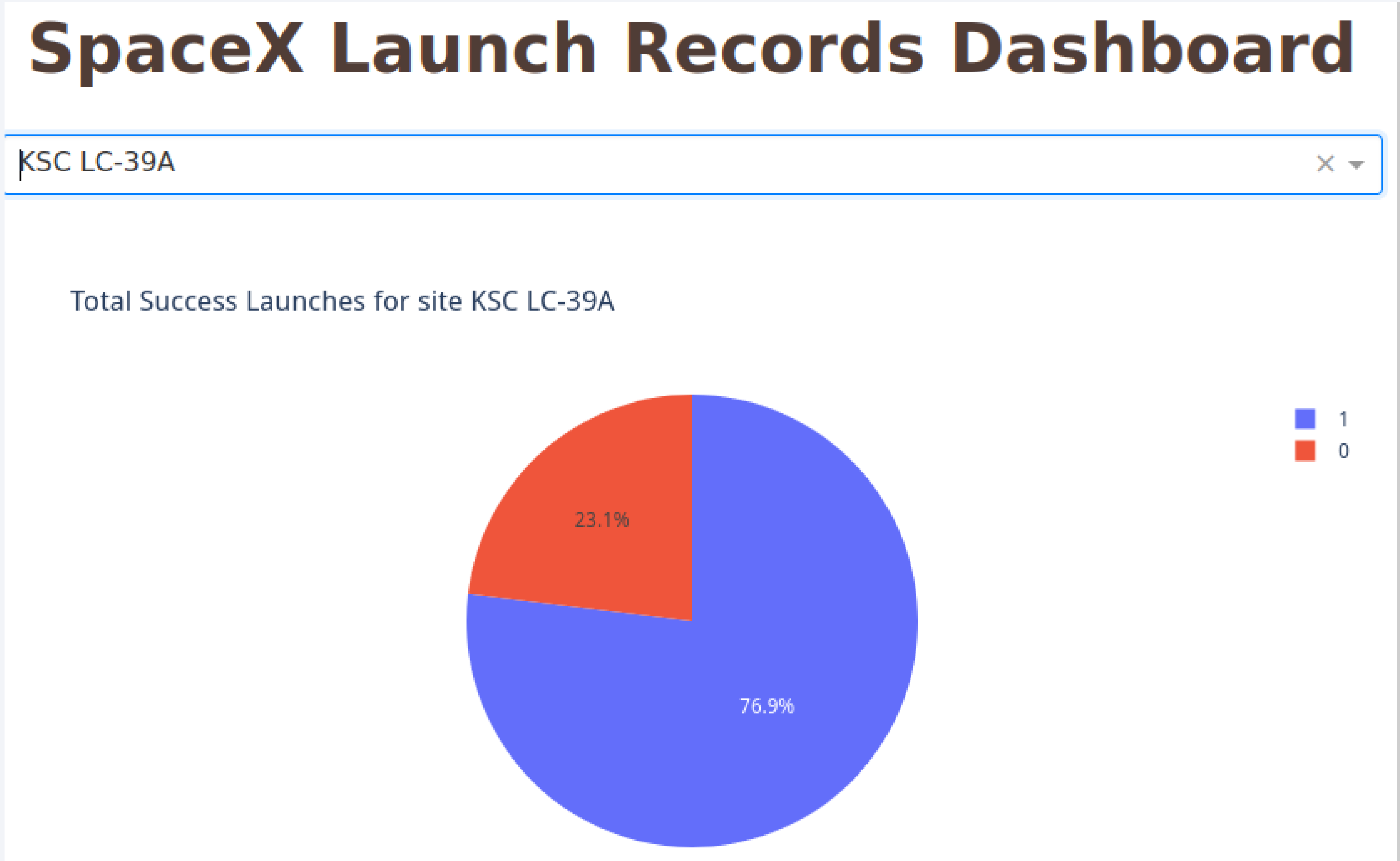
Pie chart showing Total Launch Success by launch site for all sites

- Pie chart showing Total Success Launches by launch site for all sites
- 41.7% (the highest) were launched from KSC



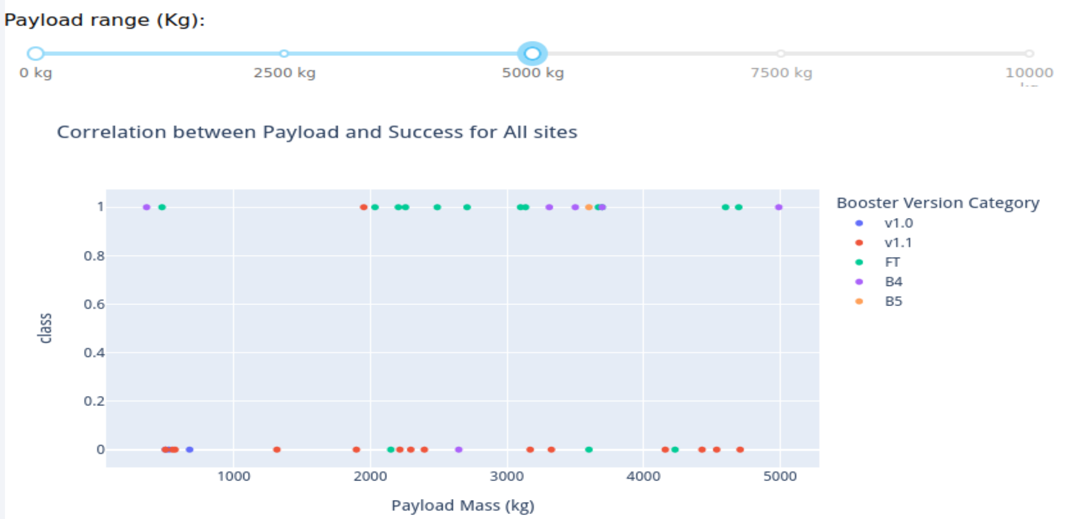
Pie chart for the launch site with highest launch success ratio

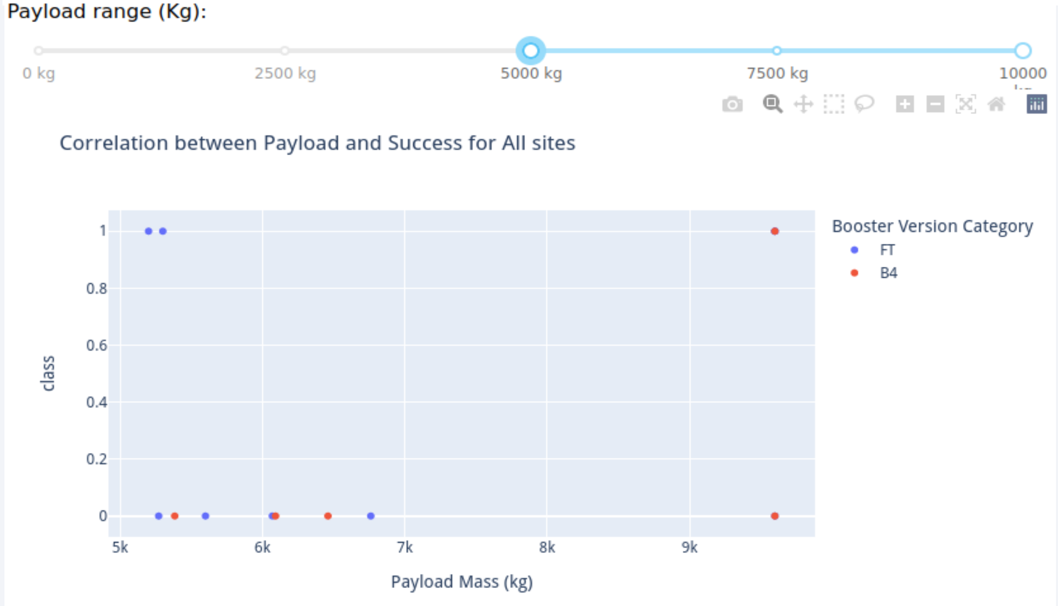
- Pie chart for the launch site with highest launch success ratio is KSC
- 76.9% of all launches were successful!



Payload vs. Launch Outcome scatter plot for all sites, with different payloads selected in the range slider

- Payload vs. Launch Outcome scatter plot for all sites, with different payloads selected in two examples on the range slider
- Analysis shows that the lower/mid range payloads had a higher success rate than the higher payload ranges, as shown at the right.
- The FT booster looks good above 5000 kg, whereas the v 1.1 shows far more failures than successes at lower payloads.





Section 5

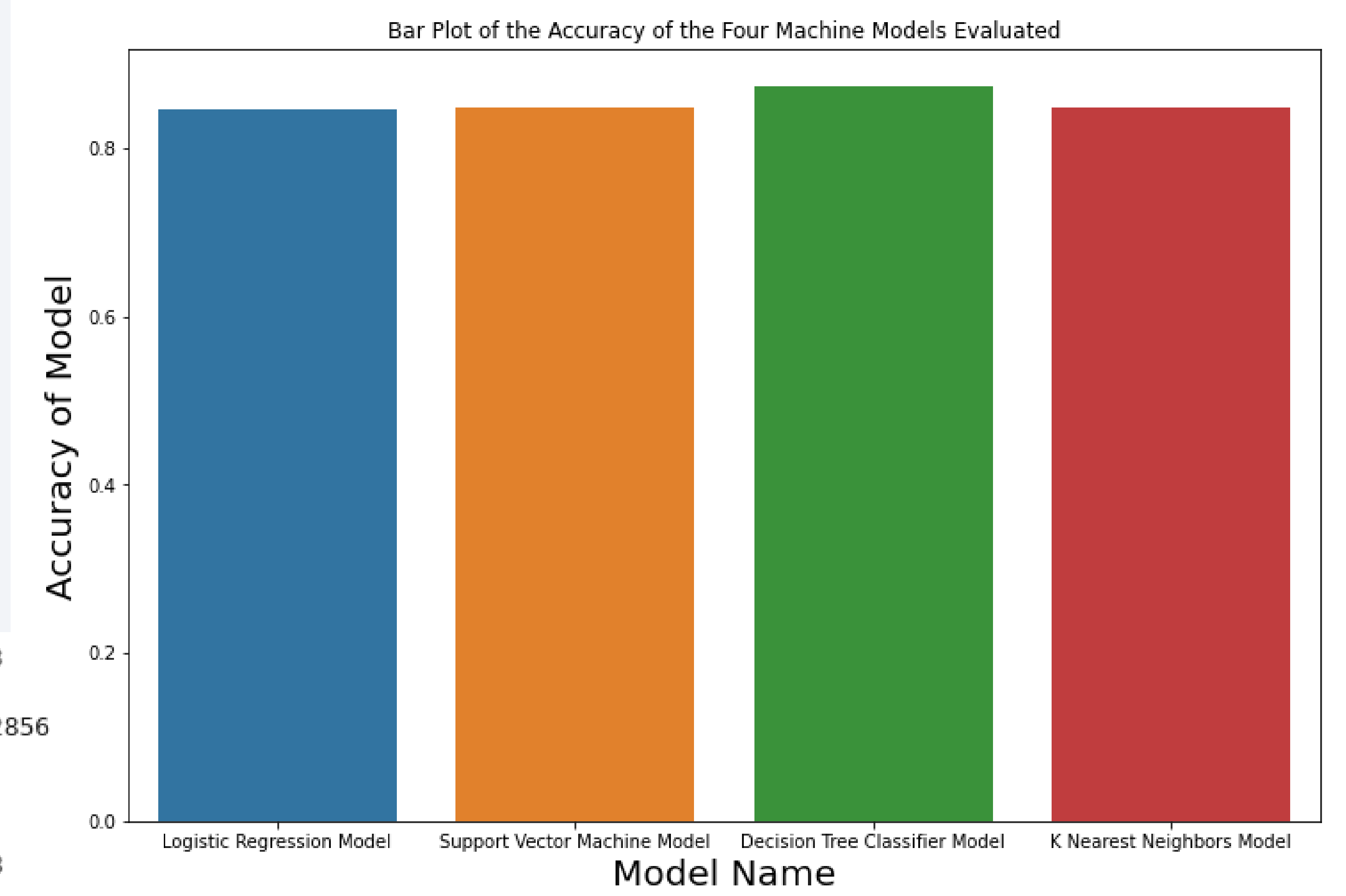
Predictive Analysis (Classification)

Classification Accuracy

A bar chart is shown to the right for the accuracy of the four machine models evaluated

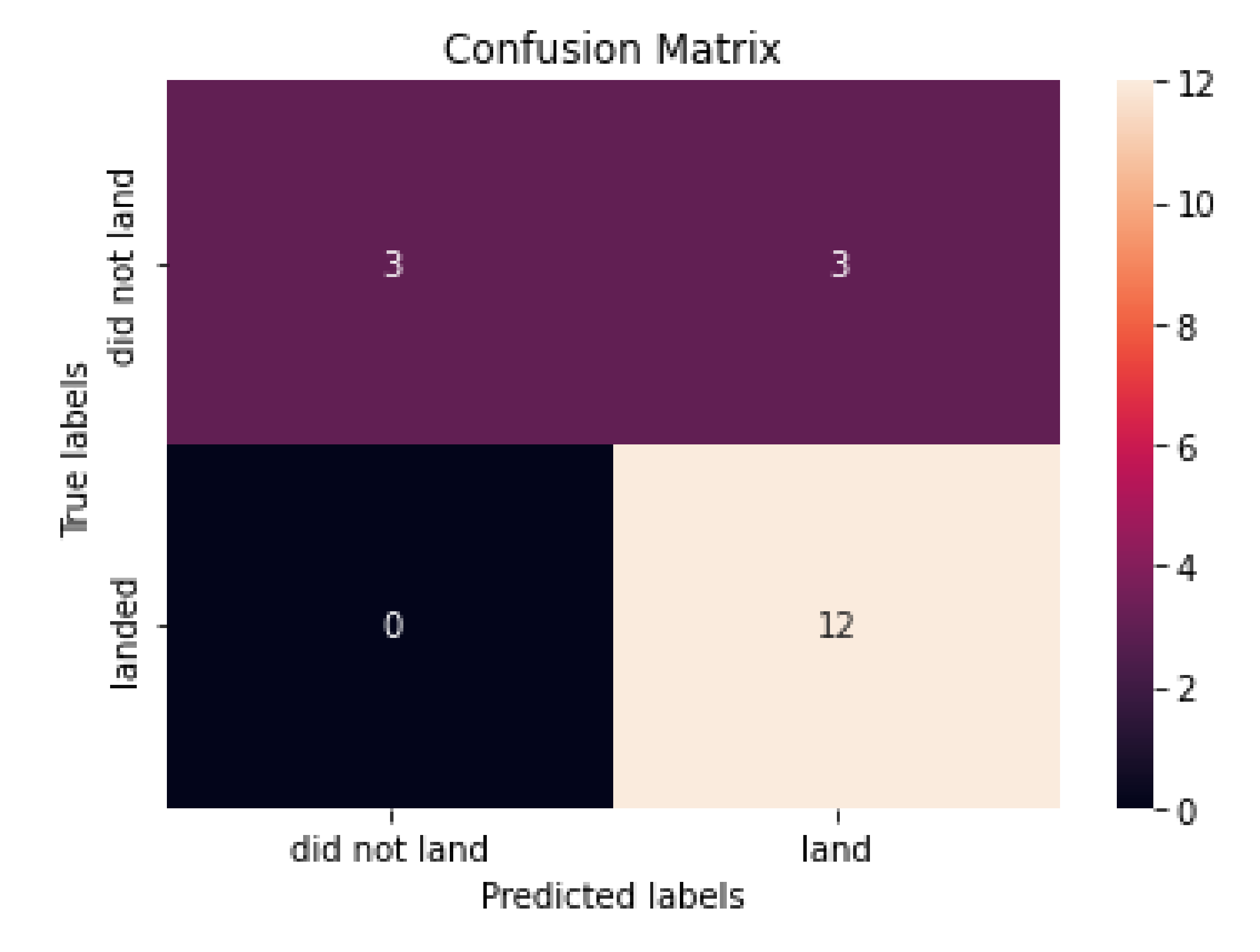
- In my analysis, the machine model that has the highest classification accuracy is the Decision Tree Classifier at 87.5 %

```
Accuracy of the Logistic Regression Model:  0.8464285714285713

Accuracy of the Support Vector Machine Model:  0.8482142857142856

Accuracy of the Decision Tree Classifier Model:  0.875

Accuracy of the K Nearest Neighbors Model:  0.8482142857142858
```



Confusion Matrix

- The confusion matrix for the Decision Tree Classifier is shown to the right.
- We see the same issues here as we did in all of the confusion matrix's evaluated – the models are very good at correctly predicting 'did land' or 'did not land'. The issue is relatively small percentage wise, but it is in false positives (3/18) = 16.7%



Conclusions

- The best performing machine model in terms of accuracy is the Decision Tree Classifier at 87.5%
- The success rate is generally increasing over time, a very good sign that the organization is successfully implementing redesigns and continuous improvement programs!
- CCAFS LC-40, has a success rate of 60 %, while KSC LC-39A and VAFB SLC 4E have a success rate of 77%.
- Unfortunately, the Project doesn't provide more information in terms of fix cost and variable that is involved in a launch. But with the above assumptions, we can still clearly conclude A CLEAR COST ADVANTAGE for SpaceX in a very costly business.
- Cost Advantages -> Higher Margins, Pricing Power -> Industry Leadership -> Industry Domination

Acknowledgements

I would like to thank the Palo Alto City Library for featuring their Upskill 2022 program. Thanks to the California State Library's American Rescue Plan Act (ARPA) program, Palo Alto City Library patrons have access to Coursera memberships for free.

For more information, check out:

https://library.cityofpaloalto.org/upskill-2022/

Appendix

- Python code and other materials are available at my GitHub repo:

  https://github.com/EasyAs12345

- Additional resources include:
  - https://plotlygraphs.medium.com/
  - https://dash.gallery/Portal/
  - https://dash.plotly.com/
  - https://dash.plotly.com/dash-core-components

- And of course …
  - https://www.coursera.org/learn/applied-data-science-capstone/home/welcome

Thank you!