IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

John Kopp
April, 25 2025

Outline

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

Executive Summary

Overview

SpaceX advertises its successful Falcon 9 launches are 62 million dollars while its competitors cost upward of 165 million dollars for the same service. SpaceX is able to offer these savings because it successfully can land and reuse the first stage of the Falcon 9 booster rocket from many of its launches.

Data collection was performed on SpaceX historical and launch data to determine whether retrieval of the first stage booster rocket can be predicted with a model built with supervised learning.

Technical launch data was retrieved from SpaceX APIs and Historical launch data was scraped from a Wikipedia website.

Executive Summary

Methods and Results, Page 1

Exploratory Data Analysis was performed using both descriptive statistics and visualizations. Relationship between variables and successful outcomes were noted.

- Payload vs Launch Site visualization showed that higher success rates were associated with heavier payloads. Additionally, the Vanderburg Space Force Base only had launches with payload under 10000 kg.
- Certain orbit types were found to be correlated with successful missions. ES-L1, GEO, HEO and SSO has 100% success rates.
- Flight number with success/failure status was plotted against orbit type. It was seen seen that for VLEO the rate of success seems dependent on flight number. The relationship between success and flight number for other orbit types such as GTO and VLEO is less clear.
- Success as a function of Payload vs Orbit type was visualized. Successful landing for heavy payloads are more likely for Polar, LEO and ISS. For GTO, a correlation between payload and success cannot be seem as the data is mixed.
- The success rate was seen to generally increase with year. Rates kept rising from 2013 to 2017. Slight dips are seen in 2018 and 2020, but the overall increasing rate is clear.

Executive Summary

Methods and Results, Page 2

SQL Analysis was performed on the data set with some interesting results.

- Total payloads brought to space by SpaceX during the period were 45596 kg.
- The average payload for the F9 V1.1 booster was 2535 kg.
- Successful drone ship landings were seen with 4 different F9 booster versions.
- A total of 98 missions were con2010-06-04 and 2017-03-20ducted over the period.
- It is seen that between 2010 and 2017 the successful landings are in the minority. The most common outcome was no attempt at landing.

Executive Summary

Methods and Results, Page 3

Launch site locations and proximities to nearby highways, railways and coasts were studied and visualized using Folium.

- The location of the launch sites and launch outcomes were plotted on a map of the United States to foster better understanding.
- Proximity analysis was conducted. It was found that the launch sites were close to highways, railways and the coast.
- Close railways and highways support transport of needed resources.
- Close oceans are needed to support ocean landings as are sometimes conducted.

Executive Summary

Methods and Results, Page 4

Interactive dashboards of launch outcomes and payload vs outcome were built using Plotly and Dash.

- It was seen that the most successful launches occurred at KSC LC-39A site.
- The CCAFS SLC-40 site was found to have the highest success ratio with 42.9% successful launches.
- Analysis of payload vs success for various booster models showed the FT booster version has the best success across payloads from 2000 kg to 5500 kg. The v1.1 and B5 boosters have poor success and the B4 booster has mixed success across this payload range.

Executive Summary

Methods and Results, Page 5

Predictive analysis was performed and 4 classification models were built and evaluated

- Four models, Logistic Regression, SMV, Binary Tree and KNN were developed.
- A grid search method was used to find optimal hyperparameters for each.
- The accuracies of each of the models were found to be equivalent and equal to 0.833.
- The confusion matrix of each model was graphed. As with accuracy, the confusion matrix for each model was found to be identical. The models predicted landed missions perfectly with 12 true positives and no false positives. Performances on unsuccessful missions was poor with 3 true negatives and 3 false negatives.
- The equal performance of all the models was probably due to the small size of the data set.

Introduction

Project background and context

SpaceX advertises its successful Falcon 9 launches are 62 million dollars while its competitors cost upward of 165 million dollars for the same service. SpaceX is able to offer these savings because it successfully can land and reuse the first stage of the Falcon 9 booster rocket from many of its launches.

Question to be answered

Successful launches depend on many factors and sources of data. Data was collected from both SpaceX via API and from summary data on Wikipedia to perform data analysis and modeling. This study determines which factors influence successful retrieval of the first stage of the rocket and forms a prediction model to predict successful retrievals. With this information, competitors to SpaceX could offer similar services.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Describe how data was collected
- Perform data wrangling
  - Describe how data was processed
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection

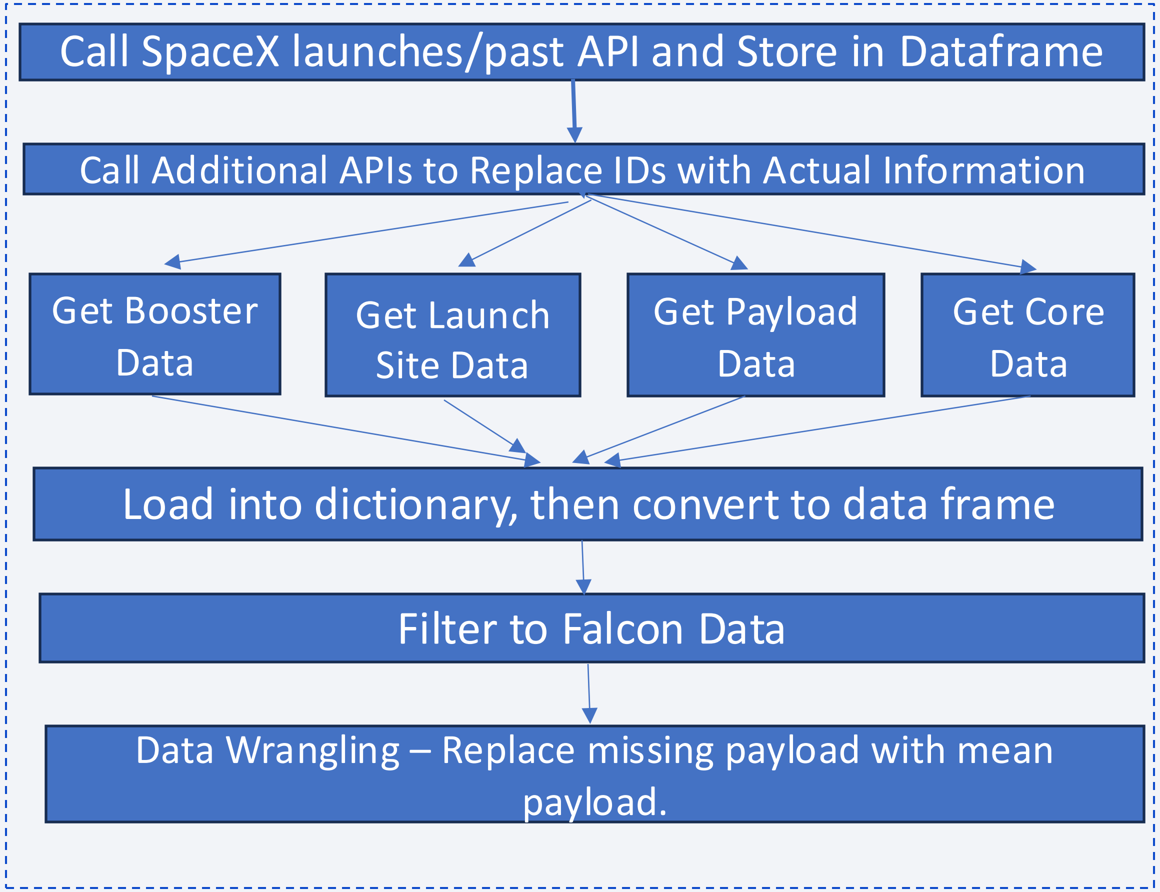
Overview

The data that drove analysis and modeling was collected from two main sources.

- SpaceX APIs were utilized to retrieve key technical details about each launch. These APIs provided technical details about each launch such as booster version, payload mass, launch pad, type of landing. Data from several APIs were combined to form a dataset.
- Historical records were scraped from a Wikipedia page, "List of Falcon 9 and Falcon Heavy Launches". This historical data has information about the launch date, launch site, booster details, landings and success or failure of the mission and landing.

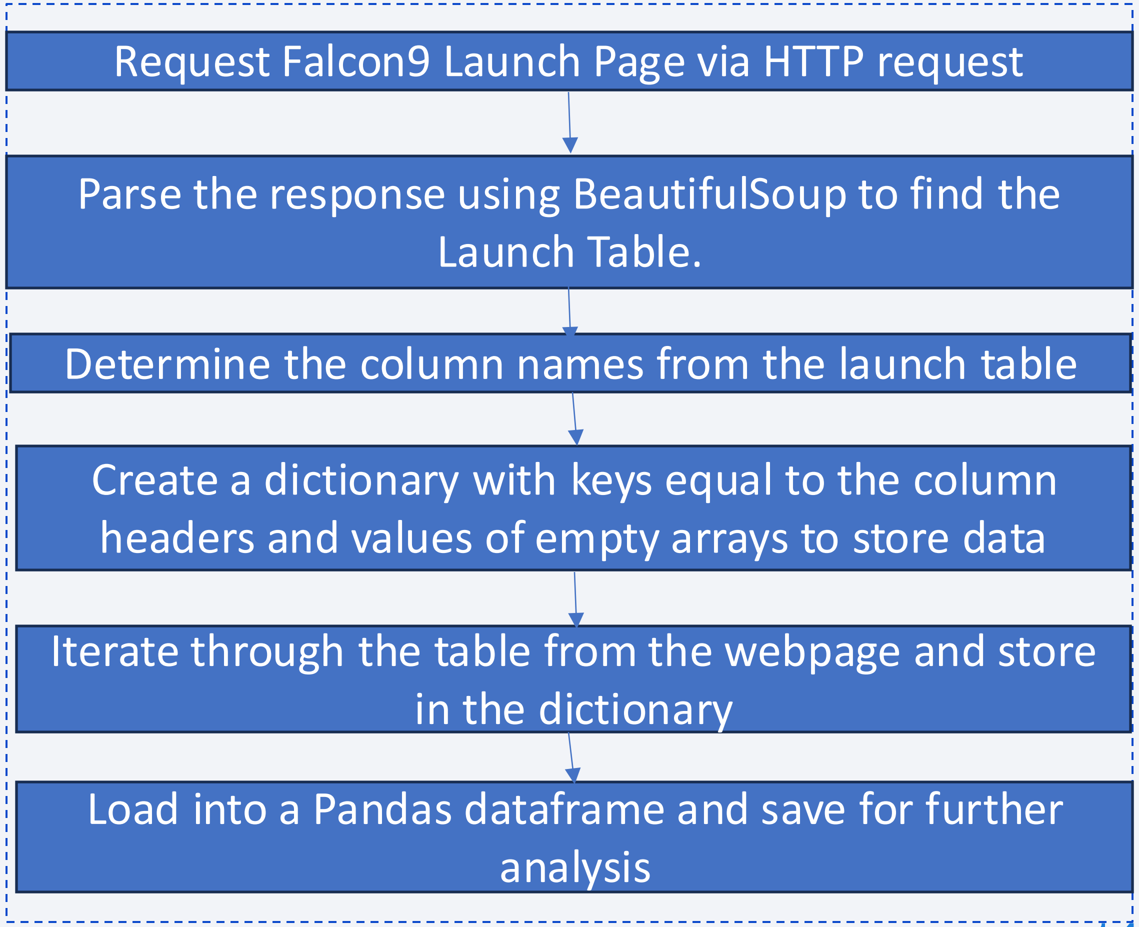
Data Collection – SpaceX API

- Data was collected from SpaceX APIs. Data from several API calls was combined to form a data set that included BoosterVersion, PayloadMass, Orbit, LaunchSite, Outcome, Flights, GridFins, Reused, Legs, LandingPad, Block, ReusedCount, Serial, Longitude, Latitude. Data was obtained using a HTTP get request, converted to JSON and loaded in Pandas dataframes.
- The Jupyter Notebook containing details of the processing of SpaceX API calls is located at data-science-capstone/jupyter-labs-spacex-data-collection-api-v2.ipynb at master · jrkopp/data-science-capstone



Data Collection - Scraping

- Historical launch data was scraped from the Wikipedia page, "List of Falcon 9 and Falcon Heavy Launches". Data was pulled via HTTP GET request, parsed with BeautifulSoup and loaded into a dataframe. Data included date, flight no, launch site, payload, orbit, customer, launch outcome, booster version and outcome.
- The Jupyter notebook containing details on the web scraping is located at [data-science-capstone/jupyter-labs-webscraping.ipynb at master · jrkopp/data-science-capstone]()



Data Wrangling

- The starting point for the data wrangling effort was a pandas dataframe containing data from SpaceX APIs that was stored within a CSV file. The goal of this exercise was to determine labels (targets) for each launch that could be used for training supervised models. Each record in the dataset contained an outcome that depended on both where the booster landing site was (ocean, drone ship, land) and the outcome. These were mapped into binary labels which indicate success or failure. Preliminary exploratory data analytics was also performed on the data to understand its fit to answering the business question of predicting successful retrievals of the first stage booster.
- Full details of the data wrangling are given in the Jupyter Notebook stored at this link. data-science-capstone/labs-jupyter-spacex-Data wrangling-v2.ipynb at master · jrkopp/data-science-capstone
- See next page for visualizations.

Data Wrangling, Continued

- Load dataframe containing SpaceX Launch Data from CSV file
- Perform Exploratory Data Analysis
- Look for null values and determine the data type of each field
- Determine the number of launches at each site
- Determine the number and occurrences of each orbit
- Determine the number and occurrences of mission outcomes
- Create a landing outcome label from the mission outcomes

Mean Outcome: 0.667

Success: 60
Failure: 30
Total Missions: 90

Nulls found only in Launch Site column

Launch Site	Count
CCAFS SLC 40	55
KSC LC 39A	22
VAFB SLC 4E	13

Outcome	Count
True ASDS	41
None None	19
True RTLS	14
False ASDS	6
True Ocean	5
False Ocean	2
None ASDS	2
False RTLS	1

Orbit Type	Count
GTO	27
ISS	21
VLEO	14
PO	9
LEO	7
SSO	5
MEO	3
HEO	1
ES-L1	1
SO	1
GEO	1

EDA with Data Visualization

- Exploratory data analysis was conducted on the SpaceX Launch Data. Visualizations were used to understand the relationships and correlations between variables. Feature Engineering was done to one-hot encode categorical variables into numeric for use in modeling.

Visualization	Plot Type
Launch Site vs Flight Number	Scatter (with hue)
Launch Site vs Payload	Scatter (with hue)
Success Rate vs Orbit	Bar
Orbit vs Flight Number	Scatter (with hue)
Payload vs Orbit	Scatter (with hue)
Launch Success vs Year	Line

- The following variables were one-hot encoded to create numeric values for later modeling: Orbits, LaunchSite, LandingPad, Serial.
- Full details on the data visualization and feature engineering can be found at this link: data-science-capstone/jupyter-labs-eda-dataviz-v2.ipynb at master · jrkopp/data-science-capstone

EDA with SQL

SQL analysis and queries were performed on the SpaceX dataset to further understand the data. The following queries were performed.

- Display the names of the unique launch sites in the space mission
- Display five records where launch sites begin with the string "CCA"
- 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 data when the first successful landing outcome in ground was achieved
- 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 all the booster versions that have carried the maximum payload mass. Use a subquery.
- 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.

Full details on the SQL analysis and queries can be found at this link: [data-science-capstone/jupyter-labs-eda-sql-coursera_sqllite.ipynb at master · jrkopp/data-science-capstone]

Build an Interactive Map with Folium

The launch success rate of SpaceX missions is dependent on technical factors, but also on the location of the launch. The location and proximities of the launch determines the trajectory the rocket must take and also determines the how/where the booster may land. To analyze these geographical properties the Folium python library for maps was used. Key useful features of Folium are:

- A map was plotted focused on the continental United States.
- Folium circles and markers were added to the map to allow easy identification of the launch sites. This allowed the sites to be easily identified on the map and allowed visual inspection and understanding of where the locations of the sites and their closeness to the coast and cities. Distance to the equator could be estimated.
- Marker clusters, with markers of green for success and red for failure, were added to each launch site on the map. This allows to simply identify the relative success rates of each site.
- Mouse Position and Markers was used to mark and label geographically significant locations on the map such as the nearest ocean and the nearest city, highway, rails to each launch site. A distance function was built and used to measure the distance between these landmarks and the launch site. Distance lines were drawn between these locations using Polyline.
- Launch site proximities for the coast, highways, railways and cities were considered.

Full details showing the maps and analysis can be found in the Jupyter Notebook at this link: data-science-capstone/lab-jupyter-launch-site-location-v2.ipynb at master · jrkopp/data-science-capstone

Build a Dashboard with Plotly Dash

A dashboard to interactively analyze and understand the launch data was built using Plotly and Dash. The dashboard contained both controls and graphs.

- A dropdown was added to allow selection of the launch site either All, CCAFS LC-40, VAFB SLC-4E, KSC LC-39A, CCAFS SLC-40. This selection was used to drive the display on each graph.
- A pie chart was added. Based on the location selected in the dropdown it displayed either total success/failure or success/failure at a given launch site.
- A slider was added to allow examination of the impact of different payload masses. It, along with the launch location, were used to allow interactive analysis of the relationship between payload and success.
- A scatter plot was added to allow analysis of the correlation between payload and launch success.

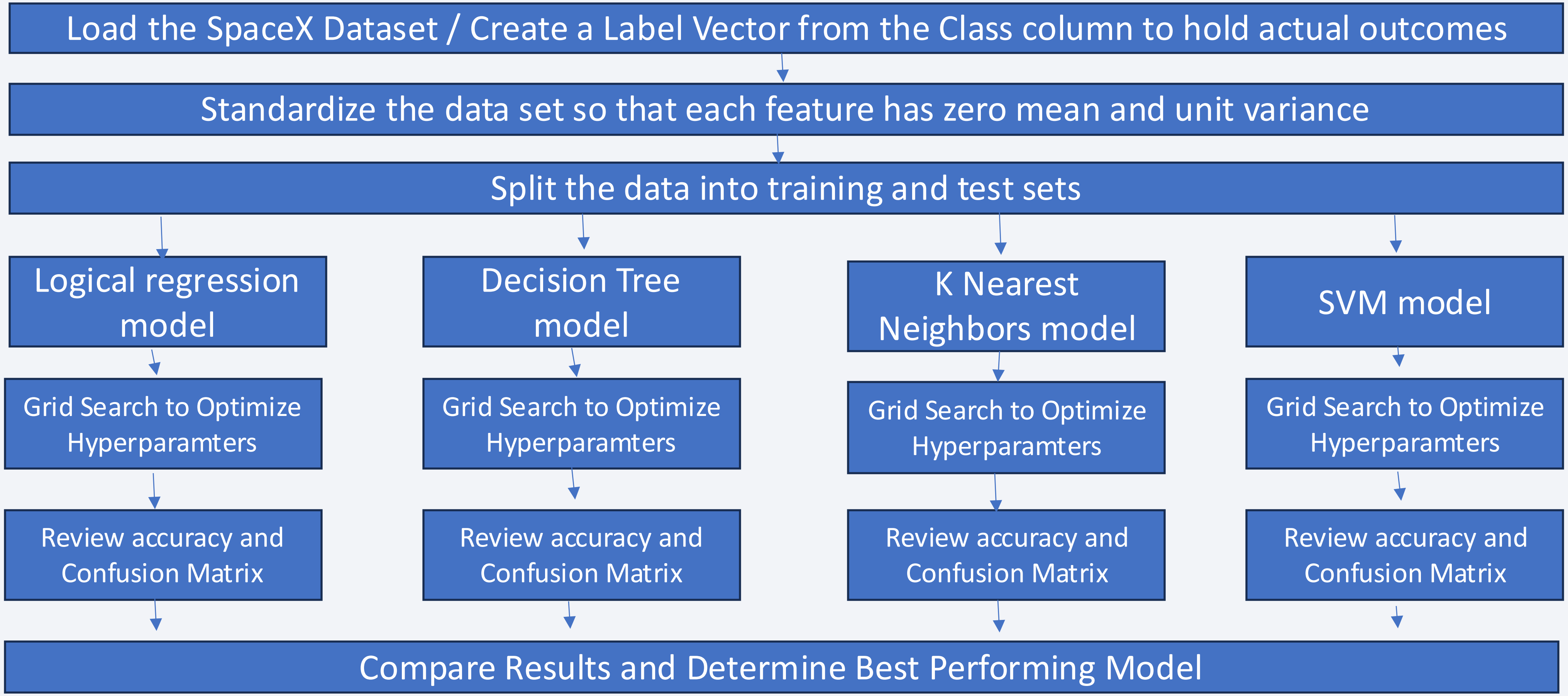
Full details on the dashboard can be found in the python script located at: data-science-capstone/spacex_dash_app.py at master · jrkopp/data-science-capstone

Predictive Analysis (Classification)

- Exploratory data analysis was performed and training labels were assigned to the SpaceX data set. Data was standardized to remove the effects of features having different scales. The dataset and training labels were split into training and test data using train_test_split.
- Supervised machine learning was used to build SVM, Decision Tree, KNN Classification Trees and Logistic Regression models.
- Grid Searches were performed to find the best hyperparameters for each model using the training data.
- The test data was used to evaluate model performance. Both accuracy and confusion models were used to determine the model that best predicted launch outcome. The confusion matrices were plotted and reviewed.
- *A flow chart of the process is presented on the next page.*

Full details of the exploratory data analysis, standardization, model development and evaluation can be found in the Jupyter notebook at this link: data-science-capstone/SpaceX-Machine-Learning-Prediction-Part-5-v1.ipynb at master · jrkopp/data-science-capstone

Predictive Analysis (Classification) (Continued)



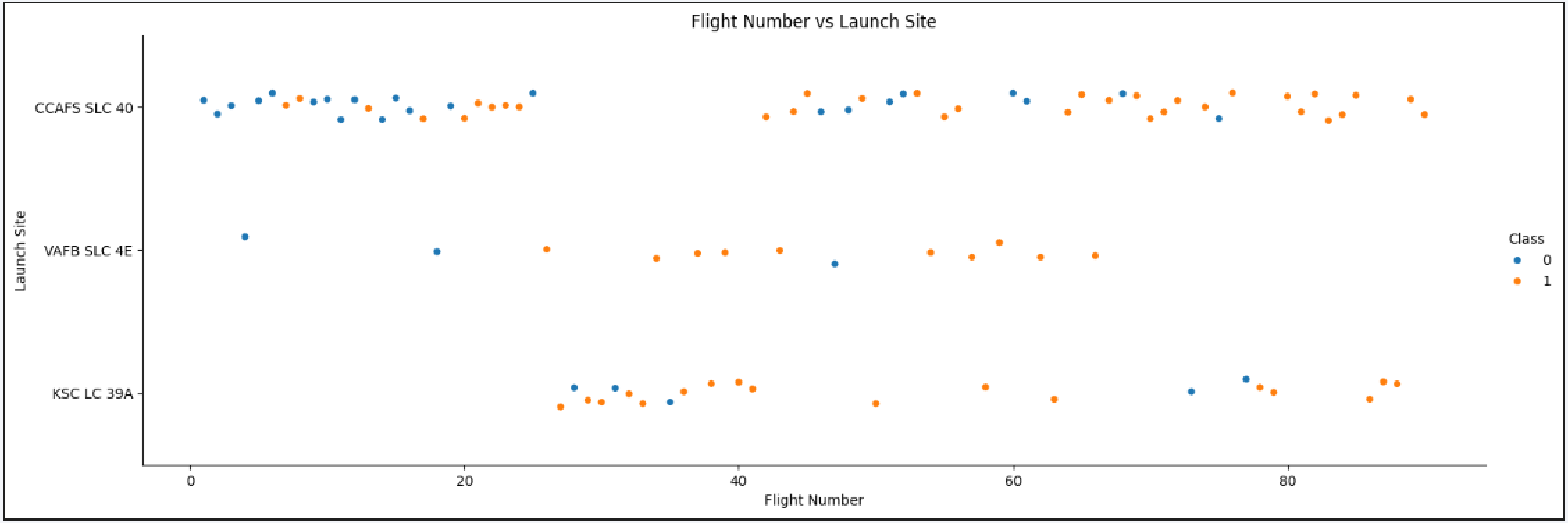
Results

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

Section 2

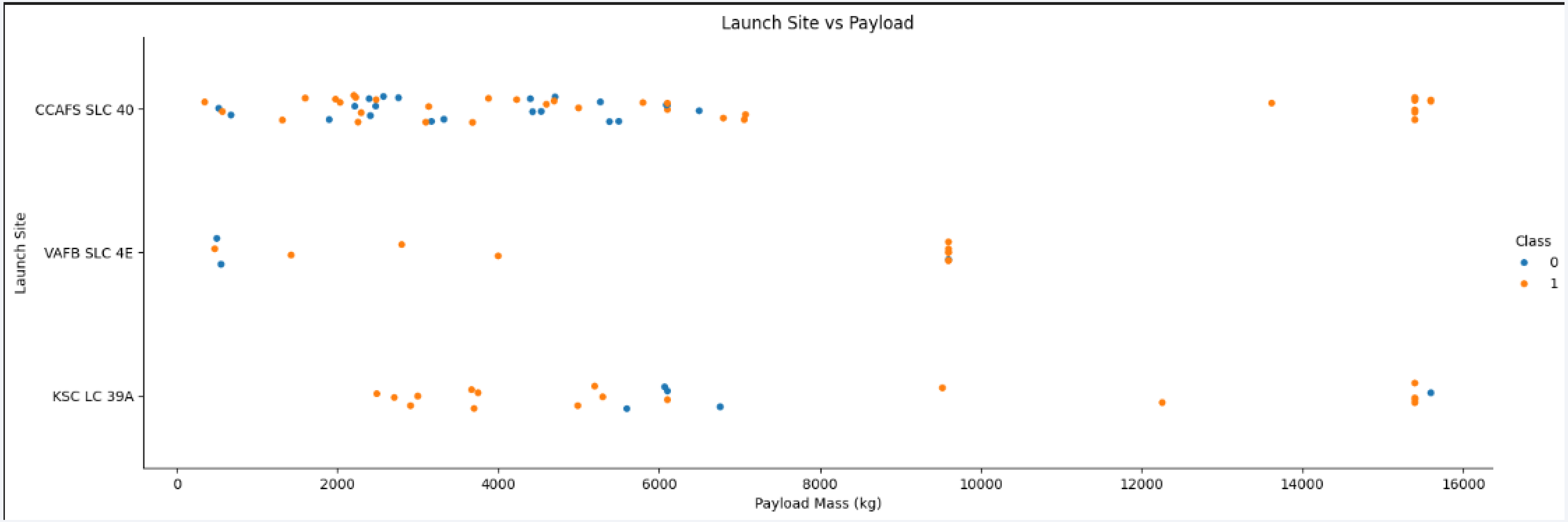
Insights drawn from EDA

Flight Number vs. Launch Site



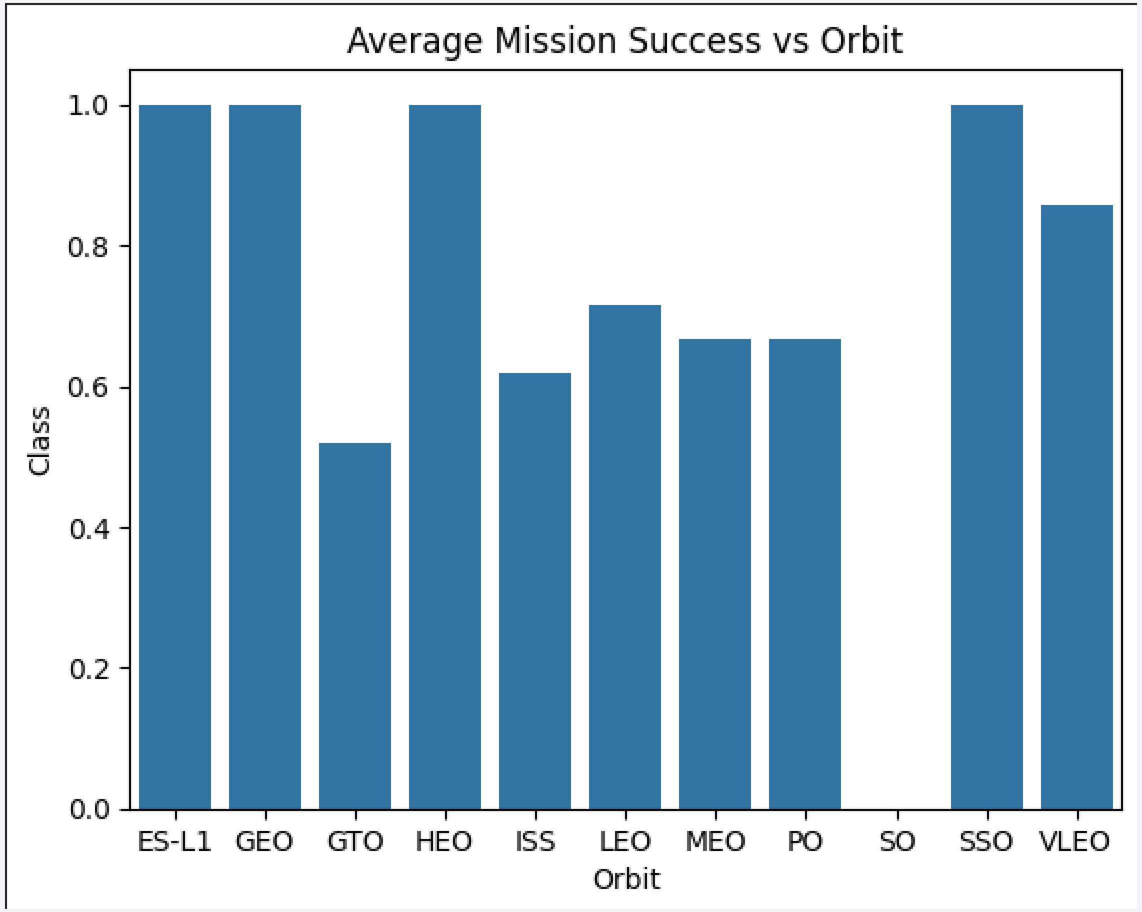
This scatter plot shows Launch Site vs Flight Number along with the success or failure of each plot. It is seen that more recent flights (higher flight number) have more successes. The CCAFS SLC 40 launch site was used for many early launches and the VAFB SLC 4E site hasn't been used recently. Most recent launches are at KSC LC 39A and CCAFS SLC 40.

Payload vs. Launch Site



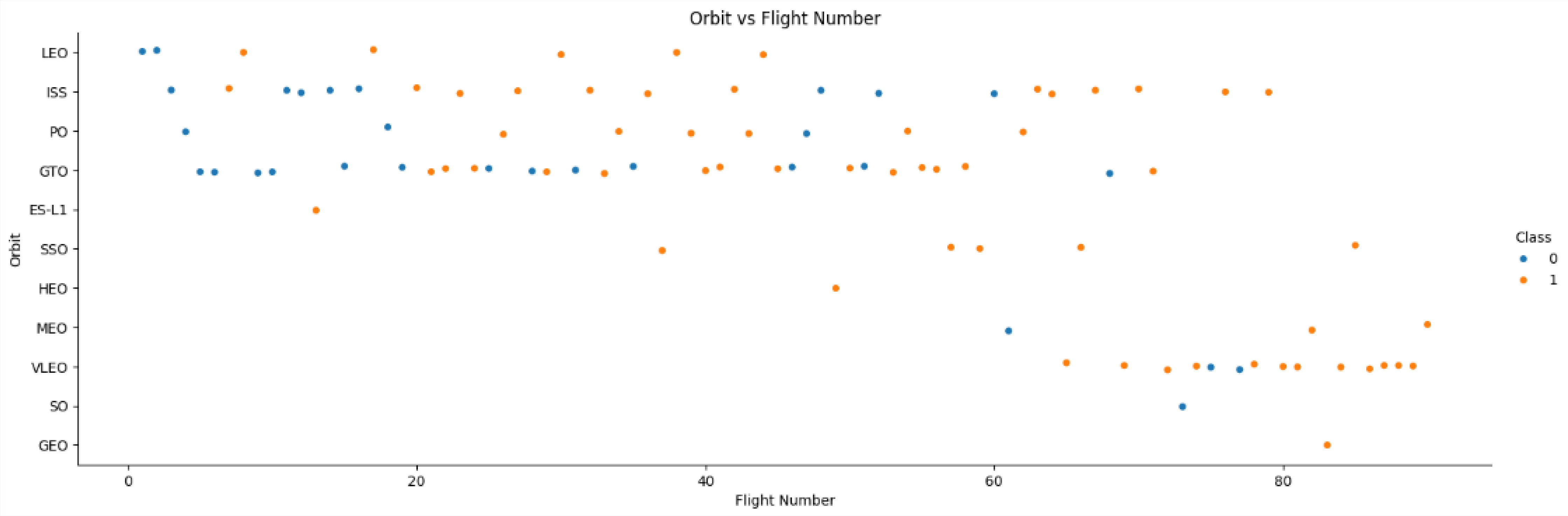
This scatter plot shows launch site with success or failure against payload mass (kg). The Vanderburg Space Force Base has no launches with payload greater than 10000 kg.

Success Rate vs. Orbit Type



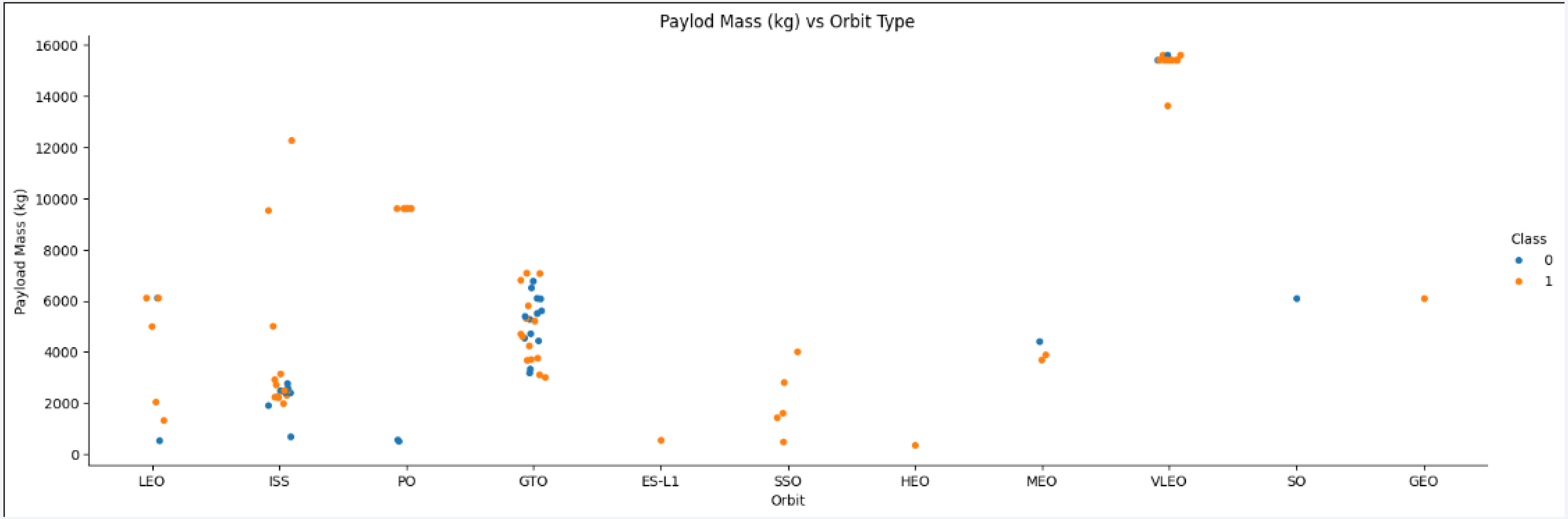
This bar chart shows average mission success vs Orbit. It is seen that success varies based on orbit type. ES-L1, GEO, HEO and SSO are seen to have perfect success rates (class = 1).

Flight Number vs. Orbit Type



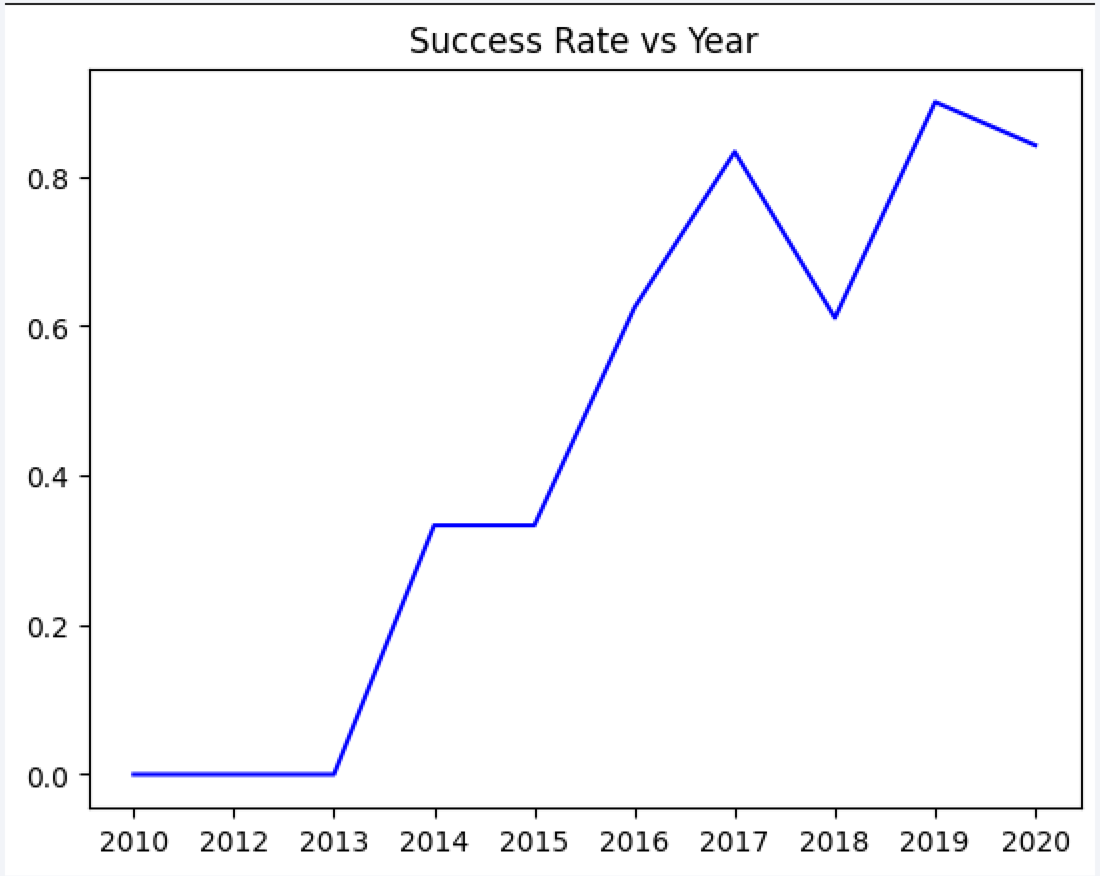
This scatter plot shows the relationship between orbit (with success/failure) vs flight number. It is seen that for VLEO the rate of success seems dependent on flight number. The relationship between success and flight number for other orbit types such as GTO and VLEO is less clear.

Payload vs. Orbit Type



Successful landing for heavy payloads are more likely for Polar, LEO and ISS. For GTO, a correlation between payload and success cannot be seem as the data is mixed.

Launch Success Yearly Trend



This line plot shows success rate vs year. It is seen that there is a positive correlation with success generally increasing with year. Rates kept rising from 2013 to 2017. Slight dips are seen in 2018 and 2020, but the overall increasing rate is clear.

All Launch Site Names

SpaceX has used 4 unique launch sites.

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

The following query was run against the SpaceX dataset loaded into a sqlite database. The distinct keyword was used to eliminate duplicate rows from the query result set.

```
SELECT DISTINCT Launch_Site
FROM SPACEXTABLE
```

Launch Site Names Begin with 'CCA'

The SpaceX dataset was loaded into a SQLite database and the following query was run. Limit was used to get only the first 5 rows shown below. All columns are not shown in order to fit on the slide.

```
SELECT *
FROM SPACEXTABLE
WHERE Launch_Site
LIKE "CCA%"
LIMIT 5;
```

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success
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
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success

Total Payload Mass

The SpaceX dataset was loaded into a SQLite database and the following query was run to determine the total payload mass carried by boosters launched by NASA (CRS). The query run and results are given below.

```
SELECT SUM(PAYLOAD_MASS__KG_)
FROM SPACEXTABLE
WHERE Customer = "NASA (CRS)";
```

Total Payload Mass = 45596 kg

The query is seem to use the aggregate sum function with a where clause limiting the customer to NASA (CRS)

Average Payload Mass by F9 v1.1

The SpaceX dataset was loaded into a SQLite database and the following query was run to determine the average payload mass carried by the booster version F9 v1.1. The query run and results are given below.

```
SELECT AVG(PAYLOAD_MASS__KG_)
FROM SPACEXTABLE
WHERE Booster_Version
LIKE "F9 V1.1%";
```

The average payload mass in kilograms was found to be 2534.67 kg. This is relatively low given the range of payloads up to 16000kg showing that many payloads were small. The aggregate function average was used in the query with the results limited to booster version like "F9 v1.1".

First Successful Ground Landing Date

The SpaceX dataset was loaded into a SQLite database and the following query was run to determine the first successful landing outcome on ground. The query run and results are given below.

```
SELECT MIN(Date)
FROM SPACEXTABLE
WHERE Landing_Outcome = "Success (ground pad)";
```

The minimum date was found to be 2015-12-22.

The SQL min function was used to find the minimum from the rows where outcome as "Success (ground pad)".

Successful Drone Ship Landing with Payload between 4000 and 6000 Kg

The SpaceX dataset was loaded into a SQLite database and the following query was run to determine the names of the boosters with successful drone ship landings and payloads between 4000 and 6000 kg. The query run and results are given below.

```
SELECT DISTINCT Booster_Version
FROM SPACEXTABLE
WHERE Landing_Outcome = "Success (drone ship)"
AND PAYLOAD_MASS__KG_ > 4000
AND PAYLOAD_MASS__KG_ < 6000;
```

Results:

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

A distinct was used in the query to remove duplicate results. The where clause limited results to those with successful drone ship landing and payloads between 4000 and 6000 kg.

Total Number of Successful and Failure Mission Outcomes

The SpaceX dataset was loaded into a SQLite database and the following query was run to calculate the total number of successful and failure mission outcomes. The query run and results are given below.

```
SELECT COUNT(*)
FROM SPACEXTABLE;
WHERE Mission_Outcome
IN ('Success','Failure');
```

A total of 98 missions were found.

The count function was used to determine the number of rows with success or failure. This number is actually the total number of rows since the only outcomes were success or failure. The outcome criteria was included in the query in case some rows had other outcomes.

Boosters Carried Maximum Payload

The SpaceX dataset was loaded into a SQLite database and the following query was run to list the names of the boosters which have carried the maximum payload mass. The query run and results are given below.

QUERY

```
SELECT Booster_Version
FROM SPACEXTABLE
WHERE PAYLOAD_MASS__KG_ =
  (SELECT MAX(PAYLOAD_MASS__KG_)
  FROM SPACEXTABLE);
```

RESULTS

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

DISCUSSION

It is seen that there are multiple versions of the F9 booster that carried the maximum payload. The query uses a subquery to first find the maximum payload and then finds all boosters that have carried that payload.

2015 Launch Records

The SpaceX dataset was loaded into a SQLite database and the following query was run to list the failed landing outcomes on drone ships, their booster versions, and launch site names for the year 2015. The query run and results are given below.

```
SELECT SUBSTR(Date,6,2), Landing_Outcome, Booster_Version, Launch_Site
FROM SPACEXTABLE
WHERE SUBSTR(Date,0,5) = '2015'
AND Landing_Outcome = 'Failure (drone ship)';
```

SUBSTR(Date,6,2)	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

Because SQLite does not support month names, substrings were used to get the month and the year as well. It is seem that there were two failed landings on drone ships in 2015.

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

The SpaceX dataset was loaded into a SQLite database and the following query was run to 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. The query run and results are given below.

```
SELECT COUNT(Landing_Outcome), Landing_Outcome
FROM SPACEXTABLE
WHERE Date BETWEEN '2010-06-04' AND '2017-03-20'
GROUP BY Landing_Outcome
ORDER BY COUNT(Landing_Outcome) DESC;
```

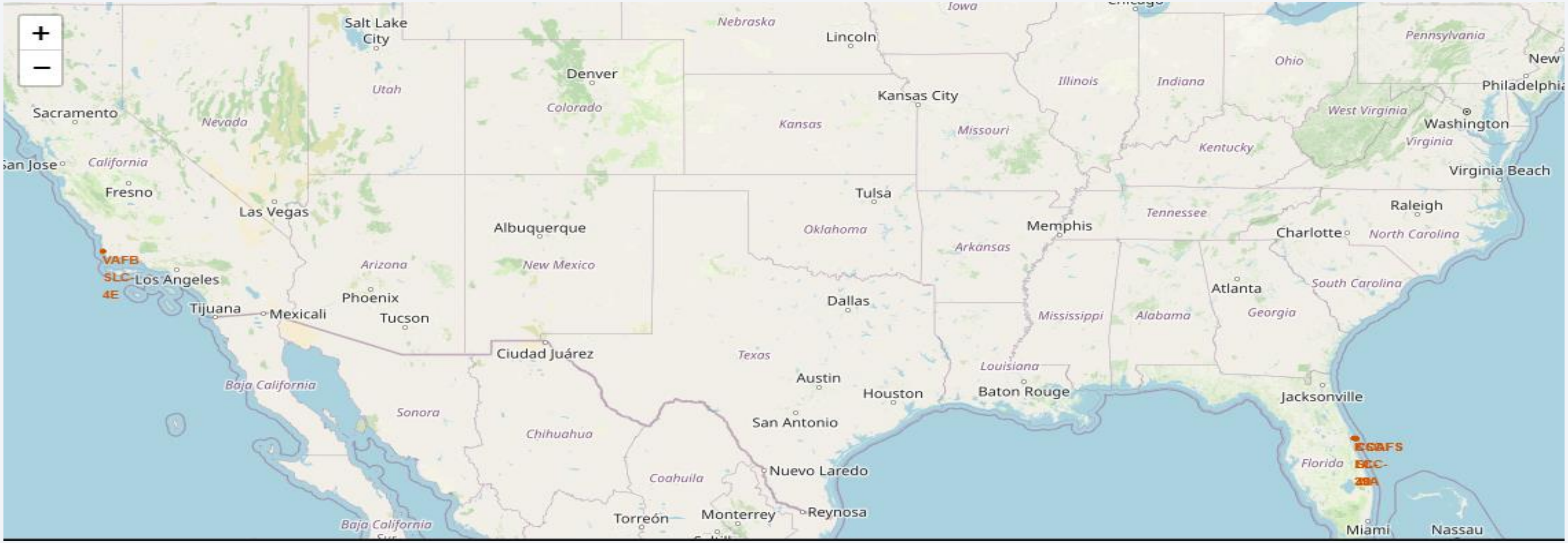
It is seen that between 2010 and 2017 the successful landings are in the minority. The most common outcome was no attempt at landing. Ranking in the query was achieved by an order by. Outcomes were grouped and counted

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

Section 3

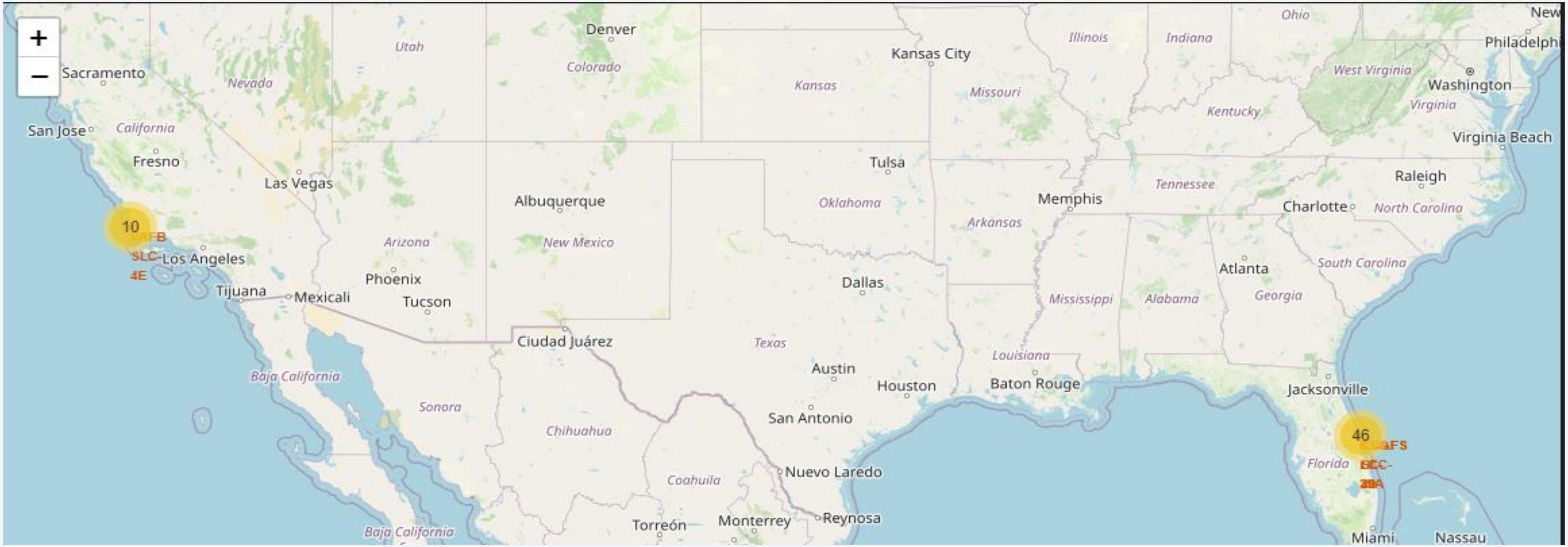
Launch Sites Proximities Analysis

Map of All SpaceX Launch Sites in the United States



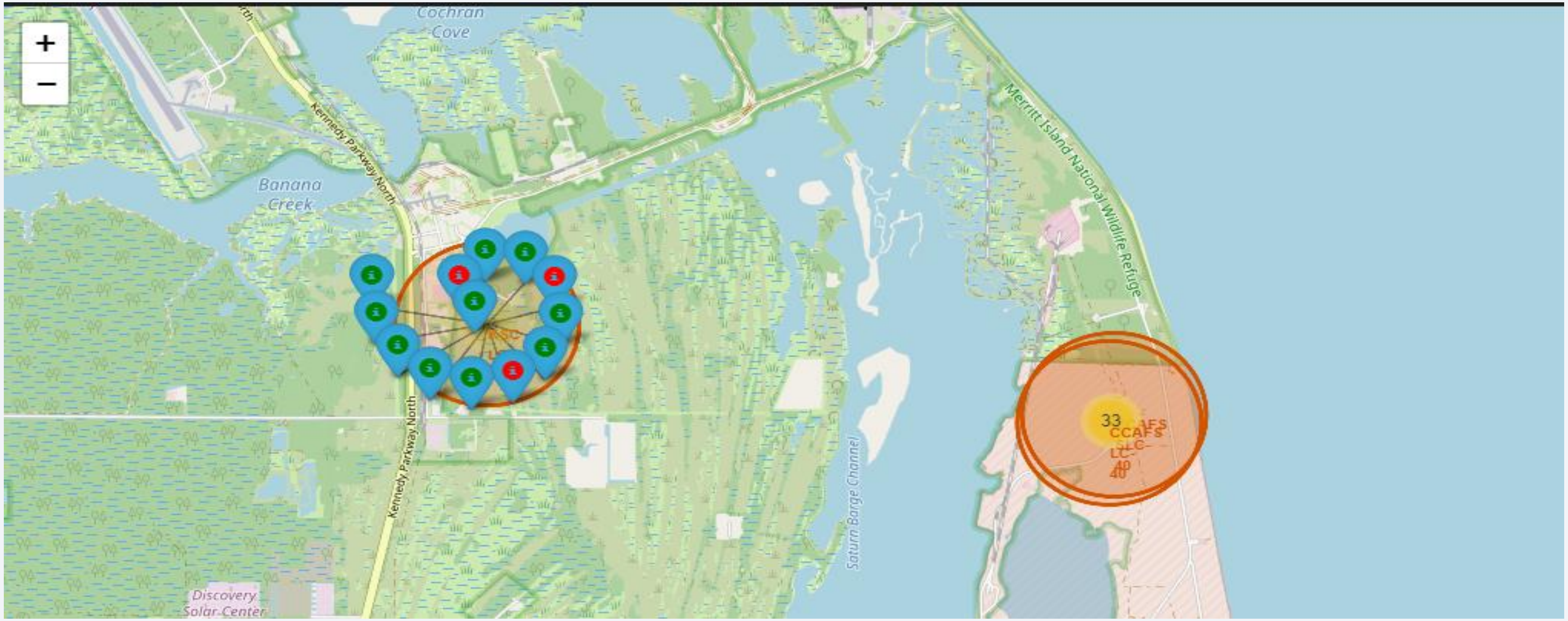
This folium map displays the launch sites on a map of the United States. The location of each launch site is surrounded by a red cirle and each site is labelled by a marker with the site name.

Success and Failed Launces for Each Launch Site (Page 1)



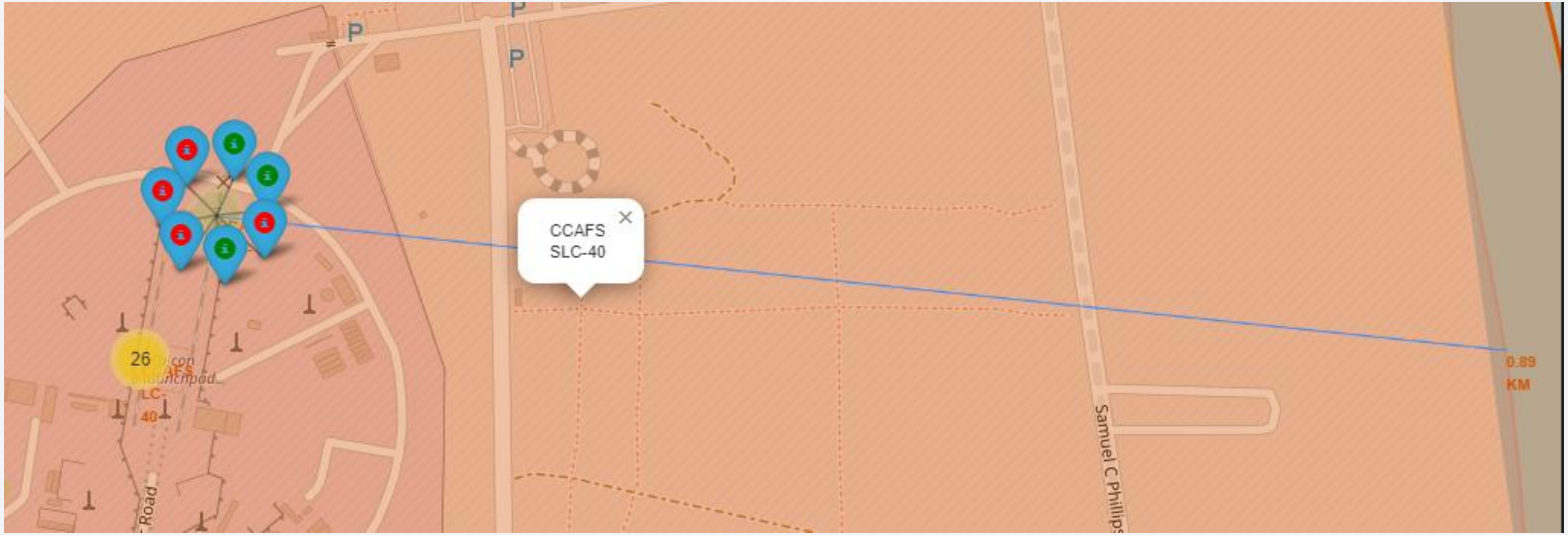
See analysis on next page.

Success and Failed Launces for Each Launch Site (Page 2)



A marker cluster was used for each launch site. The location of each launch can be seen by drilling in with the success or failure indicated by the color of the marker, red or green.

Distances between CCAFS SLC-40 and Proximities (Coast)

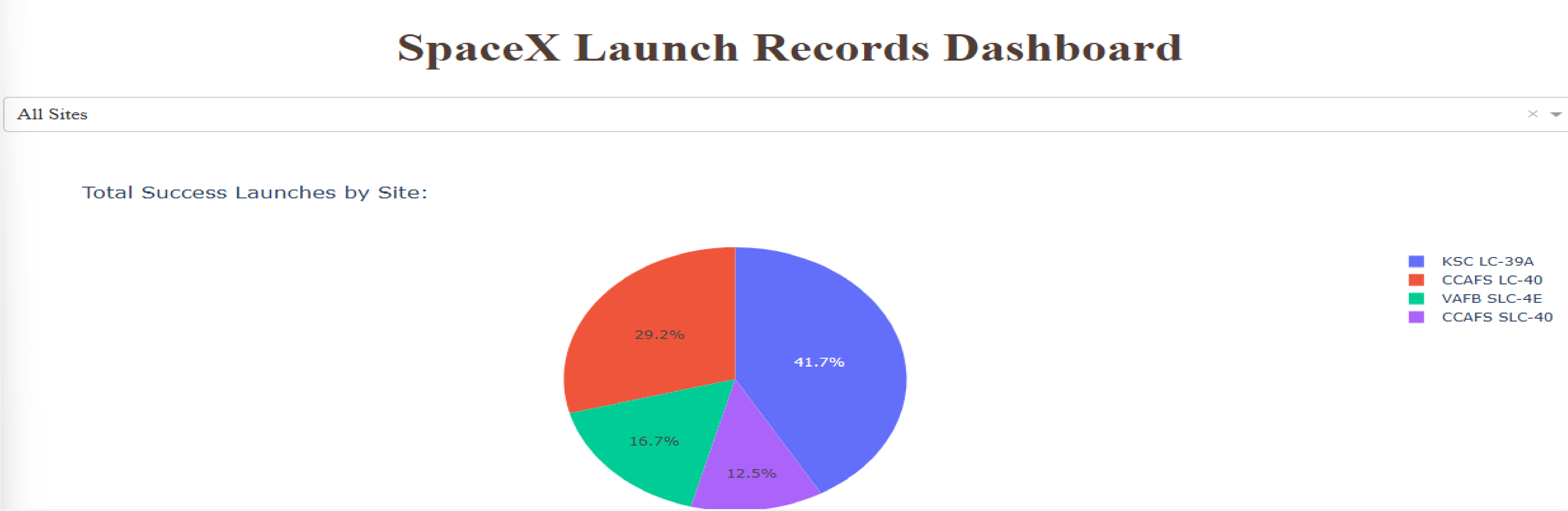


The map was updated to show the distance between a launch site and nearby proximities such as highways, railways and coast. The distance between CCAFS SLC-40 launch site and the coast was calculated using measurements taken from the map with MousePosition and a distance function. The proximities were labelled with the calculated distance and a line was drawn between the proximity and launch site.

Section 4

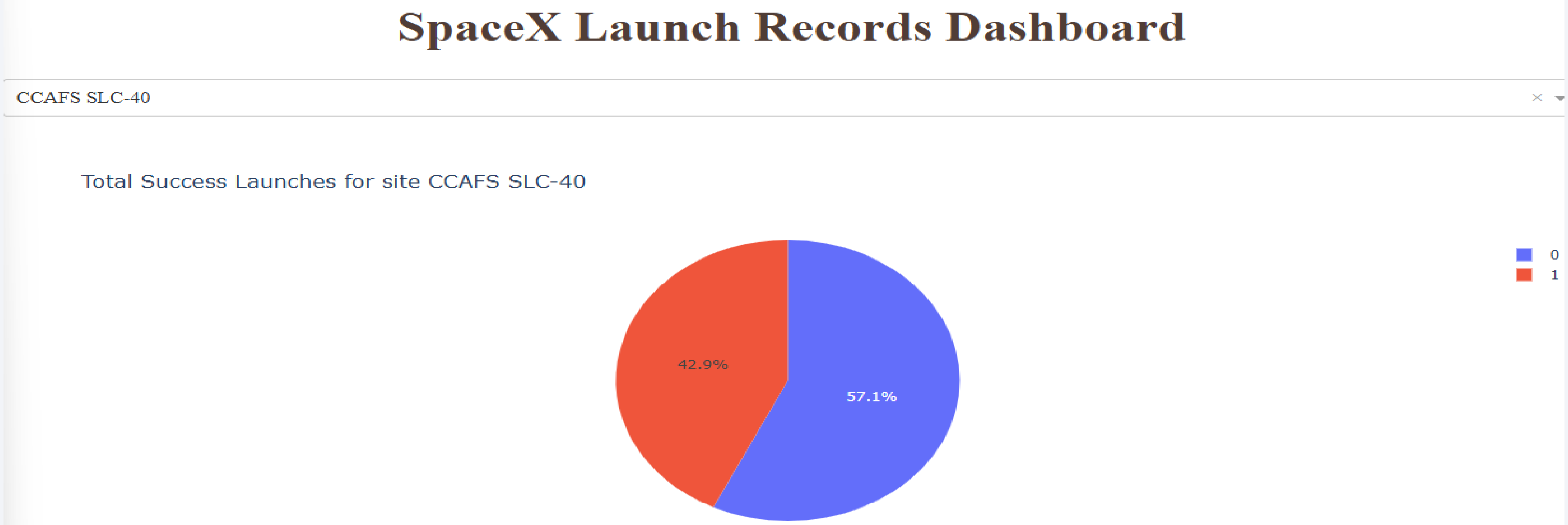
Build a Dashboard with Plotly Dash

Launch Successes By Site Across All Sites.



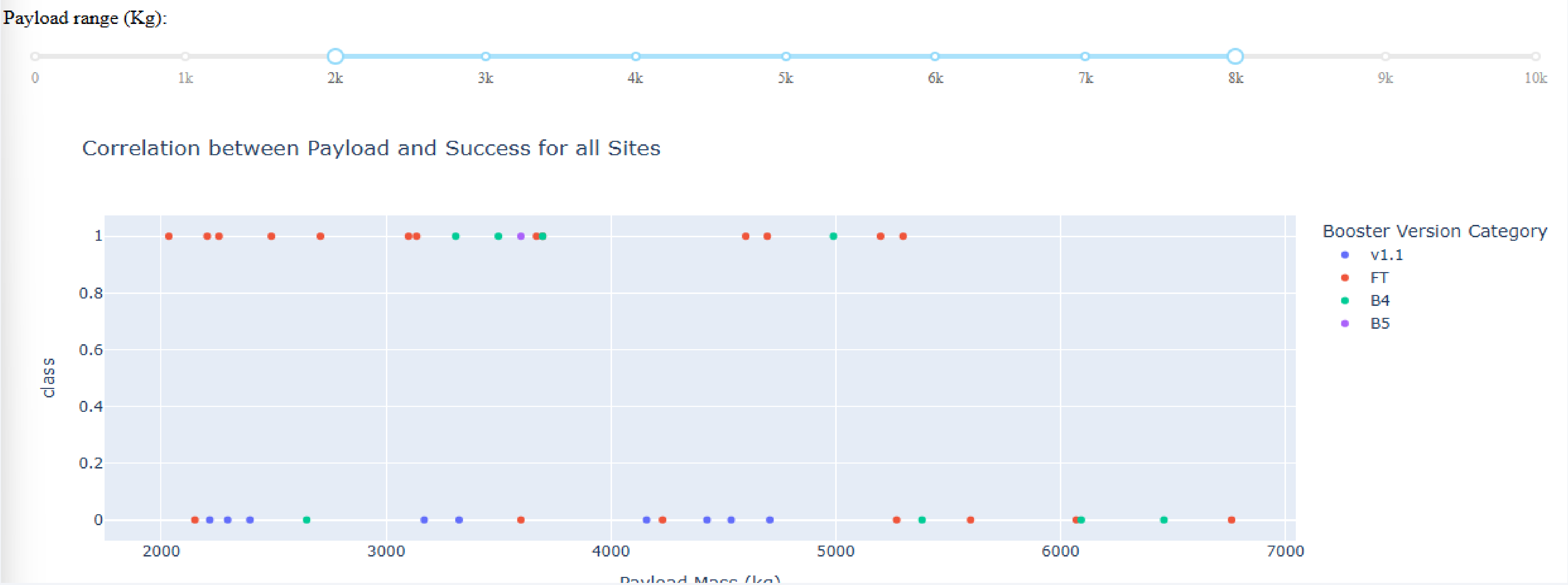
The pie chart on the dashboard shows successful launches by site for all four sites by percentage. It is seen that the most successful launches occurred at KSC LC-39A site. Of note is the dropdown. It allows selection of either all sites (shown) or a single site. When a single site is selected, the counts of successes and failures are shown.

CCAFS SLC-40 Has the Highest Launch Success Ratio



The CCAFS SLC-40 site was found to have the highest success ratio with 42.9% successful launches. The pie chart shows success or failure by the colors shown in the legend. Percentages of percent and failures are shown in the pie chart. The dash board can show this breakdown for any of the four launch sites by selecting it in the dropdown.

Payload vs Launch Scatter Plot for All Sites

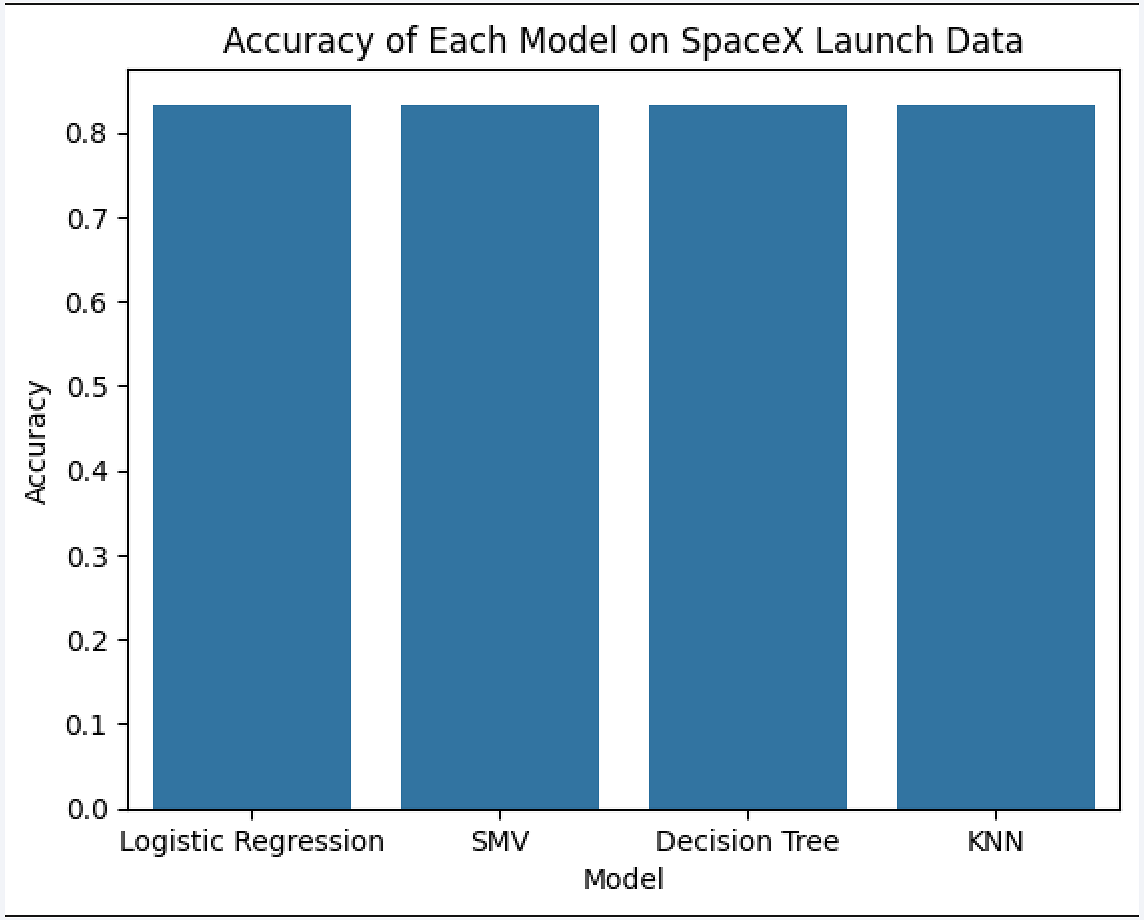


This scatter plot shows the FT booster version has the best success across payloads from 2000 kg to 5500 kg. The v1.1 and B5 boosters have poor success and the B4 booster has mixed success across this payload range.

Section 5

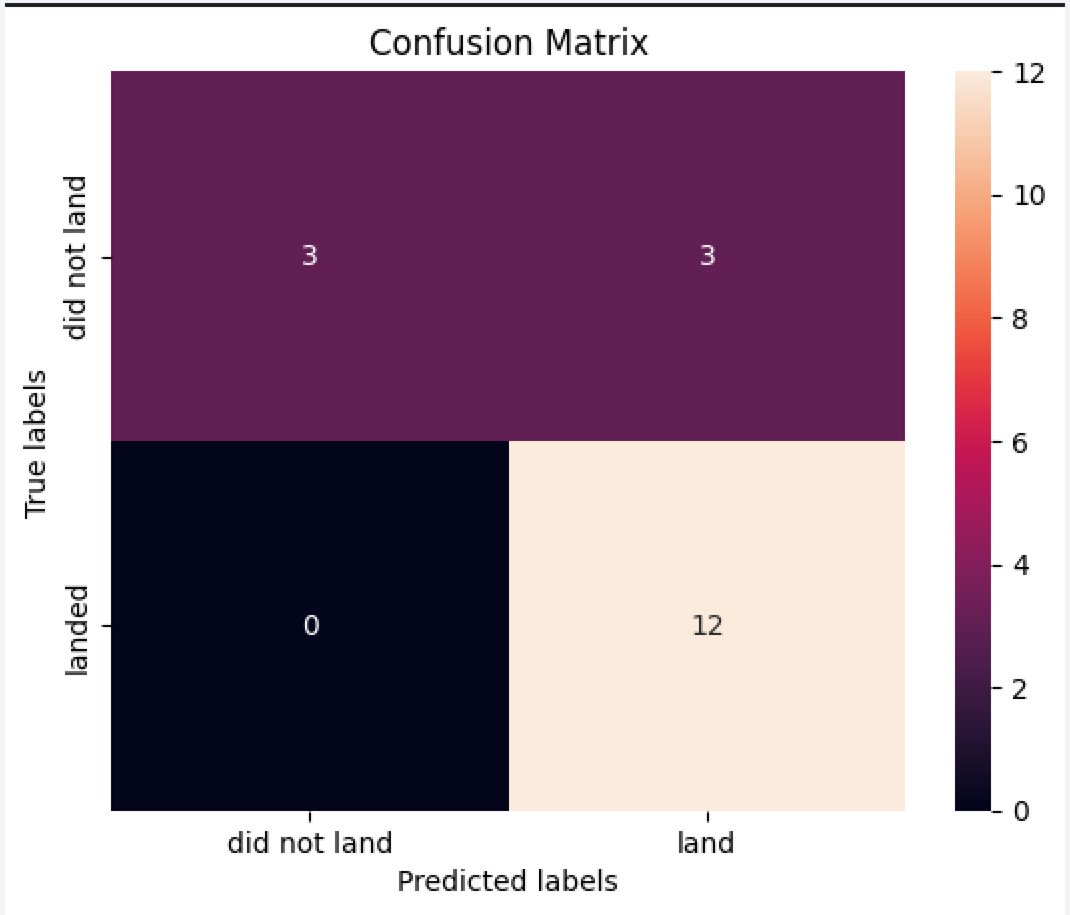
Predictive Analysis (Classification)

Classification Accuracy



Four models were developed and tested against the SpaceX launch data. For each model a Grid Search was performed to tune its hyperparamaters. After training and finding the optimal hyperparameters for each, the accuracy of each model was measured using the test data set. Each model produced an accuracy of 0.833. This was most likely due to the small data set used for modeling and evaluation.

Confusion Matrix



As with accuracy, the confusion matrix for each model was found to be identical. The models predicted landed missions perfectly with 12 true positives and no false positives. Performances on unsuccessful missions was poor with 3 true negatives and 3 false negatives. As with accuracy this was probably due to too small a data set for training and testing.

Conclusions

SpaceX advertises its successful Falcon 9 launches are 62 million dollars while its competitors cost upward of 165 million dollars for the same service. SpaceX is able to offer these savings because it successfully can land and reuse the first stage of the Falcon 9 booster rocket from many of its launches. Data collection, wrangling, exploratory data analysis, geographic visualization, interactive dashboards and predictive analysis was performed and determined the following.

- Payload vs Launch Site visualization showed that higher success rates were associated with heavier payloads. Additionally, the Vanderburg Space Force Base only had launches with payload under 10000 kg.
- Certain orbit types were found to be correlated with successful missions. ES-L1, GEO, HEO and SSO has 100% success rates.
- Flight number with success/failure status was plotted against orbit type. It was seen seen that for VLEO the rate of success seems dependent on flight number. The relationship between success and flight number for other orbit types such as GTO and VLEO is less clear.

Conclusions

- Success as a function of Payload vs Orbit type was visualized. Successful landing for heavy payloads are more likely for Polar, LEO and ISS. For GTO, a correlation between payload and success cannot be seem as the data is mixed.
- The success rate was seen to generally increase with year. Rates kept rising from 2013 to 2017. Slight dips are seen in 2018 and 2020, but the overall increasing rate is clear.
- Total payloads brought to space by SpaceX during the period were 45596 kg.
- The average payload for the F9 V1.1 booster was 2535 kg.
- Successful drone ship landings were seen with 4 different F9 booster versions.
- A total of 98 missions were con2010-06-04 and 2017-03-20ducted over the period.
- It is seen that between 2010 and 2017 the successful landings are in the minority. The most common outcome was no attempt at landing.

Conclusions

- The location of the launch sites and launch outcomes were plotted on a map of the United States to foster better understanding.
- Proximity analysis was conducted. It was found that the launch sites were close to highways, railways and the coast.
- Close railways and highways support transport of needed resources.
- Close oceans are needed to support ocean landings as are sometimes conducted.
- It was seen that the most successful launches occurred at KSC LC-39A site.
- The CCAFS SLC-40 site was found to have the highest success ratio with 42.9% successful launches.
- Analysis of payload vs success for various booster models showed the FT booster version has the best success across payloads from 2000 kg to 5500 kg. The v1.1 and B5 boosters have poor success and the B4 booster has mixed success across this payload range.

Conclusions

- Four models, Logistic Regression, SMV, Binary Tree and KNN were developed.
- A grid search method was used to find optimal hyperparameters for each.
- The accuracies of each of the models were found to be equivalent and equal to 0.833.
- The confusion matrix of each model was graphed. As with accuracy, the confusion matrix for each model was found to be identical. The models predicted landed missions perfectly with 12 true positives and no false positives. Performances on unsuccessful missions was poor with 3 true negatives and 3 false negatives.
- The equal performance of all the models was probably due to the small size of the data set.

Appendix

All data, python and artifacts are located in GitHub at jrkopp/data-science-capstone and are available for reference.

Thank you!