IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

David Lockner
14 January 2022

Outline

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

Executive Summary

- Historical data obtained from the SpaceX API and Wikipedia page
- Data was cleaned to obtain a binary yes/no for launch success in order to simplify the dataset.
- Based on yearly trends, SpaceX's launch success is on track to be near 100% in the near future
- Launch Site KSC LC-39A has the highest historical launch success rate
- VAFB SLC 4E is located close to a residential area
- The best model to predict launch outcomes based on payload mass, destination orbit, and booster version is the Decision Tree.

Introduction

- For this project, we are whether a future launch of the SpaceX Falcon 9 rocket will be successful, given criteria such as launch site, payload mass, booster version, and destination orbit.
- We will generate a machine learning algorithm to make future predictions. In order to train the algorithm, we will use publicly available historical launch data. We will train multiple different types of algorithms, and select the algorithm with the highest accuracy when tested against historical launches.

Section 1

Methodology

Methodology

Executive Summary

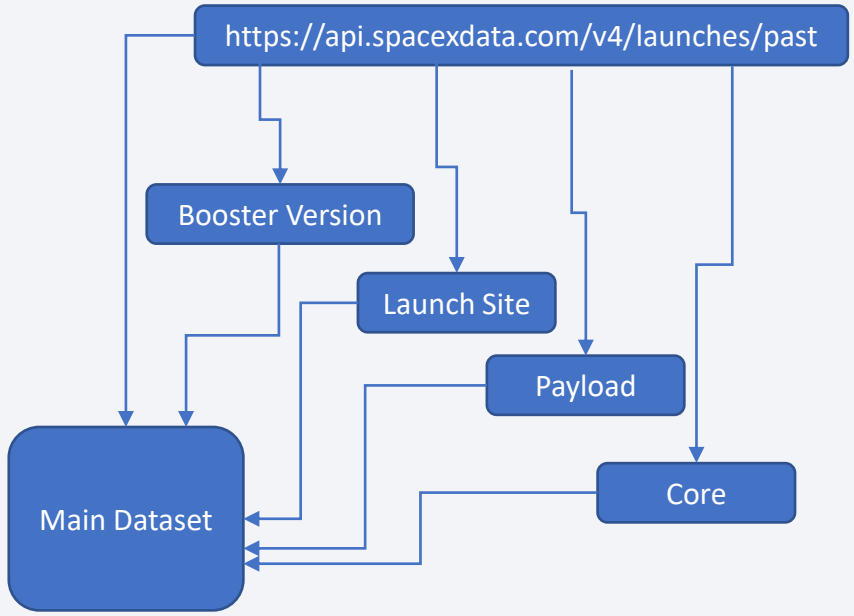
- Data collection methodology:
  - Data was collected from the SpaceX API and from scraping the Wikipedia page on SpaceX launch information.
- Perform data wrangling
  - A field was generated to obtain a binary yes/no for launch success in order to simplify the dataset.
- Perform exploratory data analysis (EDA) using visualization and SQL
  - Charts and queries were used to gauge initial trends in the data, to find fields that were most relevant to launch outcome.
- Perform interactive visual analytics using Folium and Plotly Dash
  - A dashboard was created to dissect launch success vs launch site and payload mass. Folium maps were created to visually represent launch sites in relation to other geographic features and launch success in a visual way.
- Perform predictive analysis using classification models
  - Several predictive models supplied by SciKit-Learn were trained using the historical dataset and then evaluated on that dataset using a train/test split. The test results were evaluated using a SciKit-Learn provided accuracy score. The model with the best accuracy was selected.

Data Collection

- Data was collected through the SpaceX API and through web scraping of launch records.
- Data was wrangled to simplify data set when analyzing, eliminating fields not related to exploring the hypothesis

Data Collection – SpaceX API

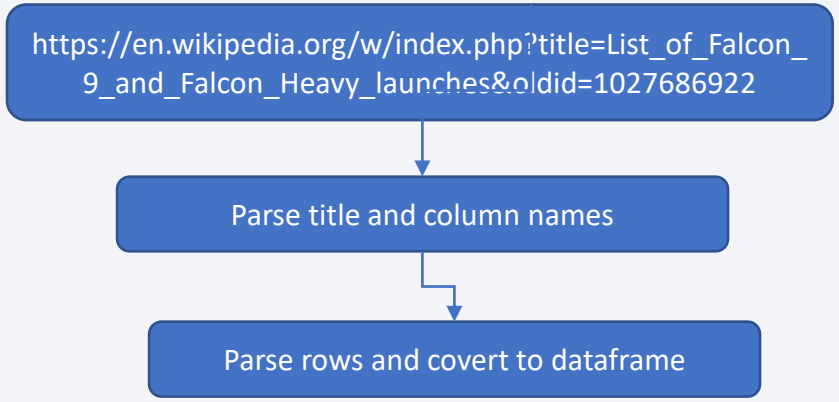
- Data retrieved from SpaceX API
- Response = Requests.get
- https://api.spacexdata.com/v4/launches/past
- Fields retrieved listed below:
- Github link: https://github.com/dlockner5/DataScienceCapstone/blob/894c27388d960623ecda3631c229e17abc380807/Data%20Collection%20API%20capstone.ipynb



FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude

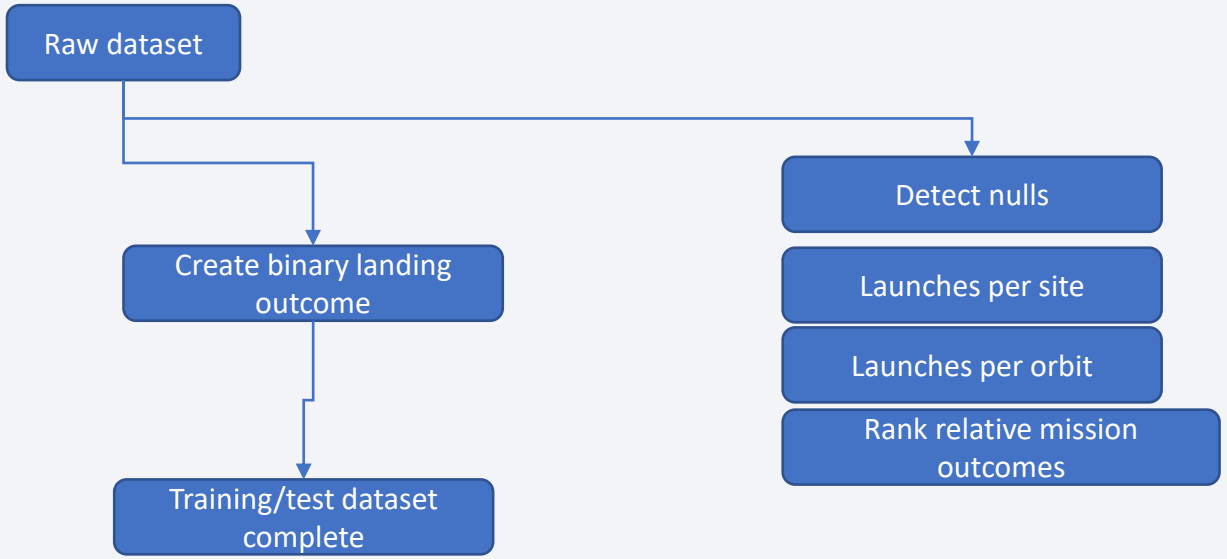
Data Collection - Scraping

- Data retrieved from https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922 using beautifulsoup
- Field names scaped from table headers, then the data was scraped from each row by iterating through each table item in beautifulsoup.
- Github URL: https://github.com/dlockner5/DataScienceCapstone/blob/833b0a803f2835276e99a4497a126f59fe337ce0/Data%20Collection%20with%20Web%20Scraping%20lab.ipynb



Data Wrangling

- Null values were explored, finding that the Landing Pad field was blank 40% of the time
- Launches per site were calculated
- Launches to each orbit were calculated
- Mission outcomes were summarized to rank relative amounts
- Binary landing outcome column created to simplify dataset
- GitHub URL: https://github.com/dlockner5/DataScienceCapstone/blob/02424d1cfc46f8c758503d974454c1ee45bca13d/EDA.ipynb



EDA with Data Visualization

- I first charted payload mass vs flight number with success (yes/no) as a color to show that the payload mass and success rate appeared to increase over time (as the flight number increased). Then, I plotted launch site vs flight number, with success (yes/no) as a color to show launch site utilization over time, and whether that impacted success rate. I then charted launch site vs payload mass to see if there was a correlation between the two. Then, I looked at the success rate of launches to each orbit type, to determine which orbits had the best success. I then plotted orbit vs flight number with success as a color to see if certain orbits were chosen as more flights were launched and success rate improved. Then, I plotted orbit vs payload mass with success as a color to see if certain orbits fared better for certain payload masses. Finally, I plotted the success rate for each year to see if SpaceX was trending towards or away from more successful launches.
- Github URL: https://github.com/dlockner5/DataScienceCapstone/blob/a980e5b87ed494db3f5b74e554de10ab11bfedf9/EDA%20with%20Data%20Visualization.ipynb

EDA with SQL

- Queries were run to gather baseline information on the dataset such as:
  - All launnch sites, to determine how many sites are in the data, and a query to look at a few launches from CCA
  - Total payload mass launched and average payload mass for a specific booster type to determine total cargo launched so far (for revenue calculation purposes)
  - Landing successes (first time and on drone ships with a certain payload mass)
  - Success ratios overall and for 2015, and outcomes ranked by number
- Github URL: https://github.com/dlockner5/DataScienceCapstone/blob/e4c0b3a0033dbd52cf6fe417f01ebdabc3d60e88/EDA%20with%20SQL.ipynb

Build an Interactive Map with Folium

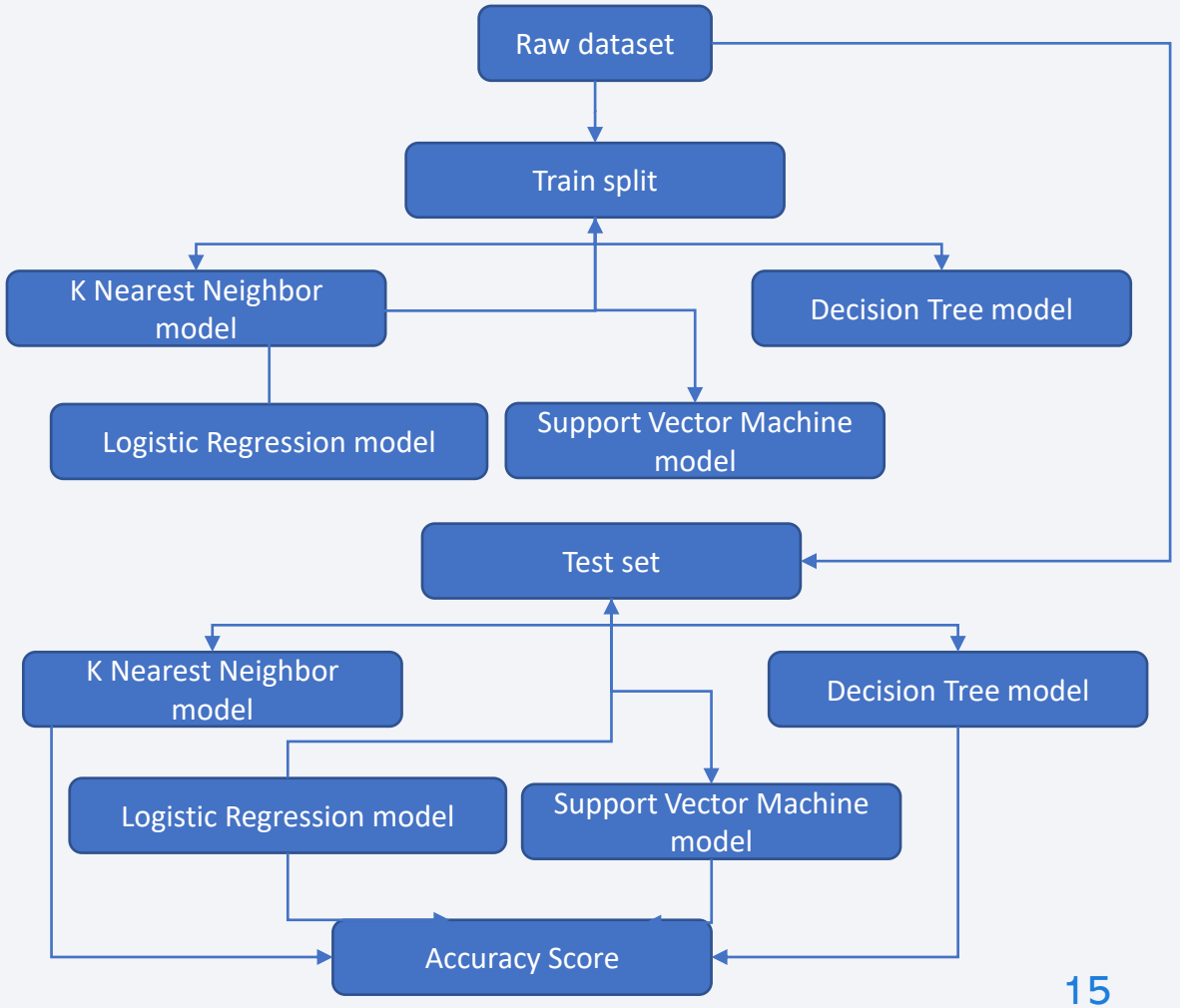
- A Folium map was created to show launch sites to visualize which areas may be used for future launches. Circles were added to the launch sites to help locate them.. Markers were added to a launch site to give a visual representation of launch outcomes from that landing site.
- Another map was created to show that the nearest town of Titusville is 15.5 km to the west of the launch site via the straight line depicted.
- GitHub URL: https://github.com/dlockner5/DataScienceCapstone/blob/02424d1cfc46f8c758503d974454c1ee45bca13d/Data%20Visualization%20with%20Folium.ipynb

Build a Dashboard with Plotly Dash

- A dashboard was created with a pie chart depicting successful launch outcomes. A dropdown list enables the user to select individual launch sites to get the specific launch success rate for that site so that the user can see which sites have the best success rate.
- A scatter plot shows payload mass vs outcome with a slider to limit displayed payload mass, drilled down to selected launch site, with booster version shown in different colors. This allows the user to see the ideal payload mass and booster version for each launch site.
- GitHub link to source code: https://github.com/dlockner5/PlotlyDash/blob/69aecc4fcc9431f0082c421380b2b688b40406bd/Plotldash_launches

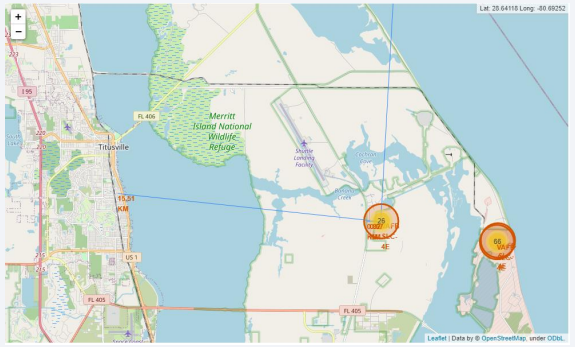
Predictive Analysis (Classification)

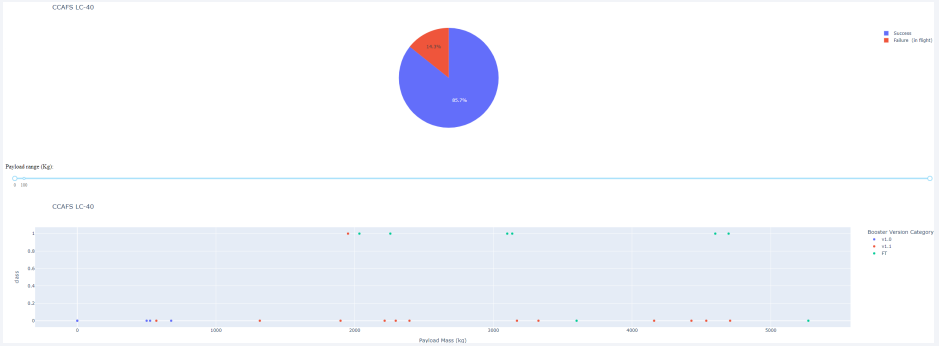
- Predictive models were built to assess launch outcomes based on variables such as payload mass, destination orbit, and launch site.
- Models were built using K-nearest-neighbor, decision tree, logistic regression, and support vector machines.
- Models were trained using 72 records and tested using 18 records from the original dataset.
- Models were evaluated using their accuracy score, of which the decision tree model scored the highest.
- You need present your model development process using key phrases and flowchart
- GitHub URL: https://github.com/dlockner5/DataScienceCapstone/blob/0d496a398b09ef9b18771b9fa3fd0b812ee7df05/Capstone%20Machine%20Learning.ipynb



Results

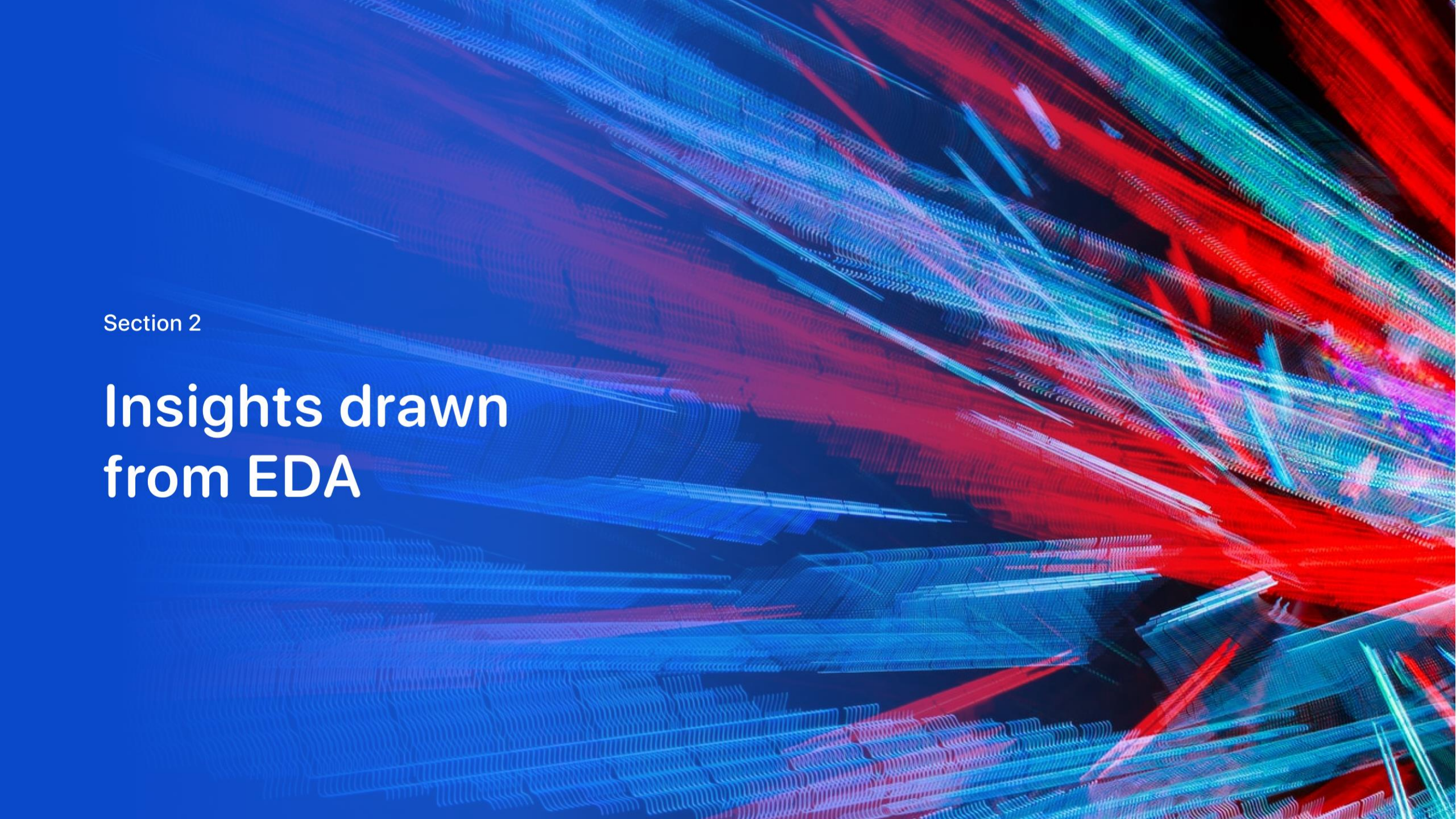
- Based on SQL queries and visualizations, the variables selected as inputs were payload mass, destination orbit, and whether a booster has been reused
- The proximity of launch sites to geographic areas such as towns was evaluated in selecting a launch site
- A Decision Tree model was selected as the most accurate model in predicting future launch success of the SpaceX Falcon 9 booster rocket when the accuracy score was measured.



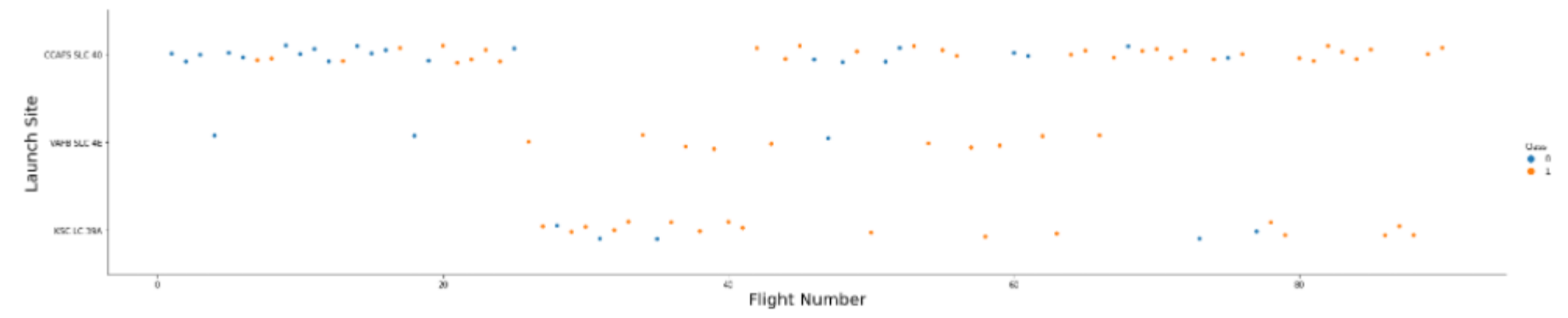


Section 2

Insights drawn from EDA

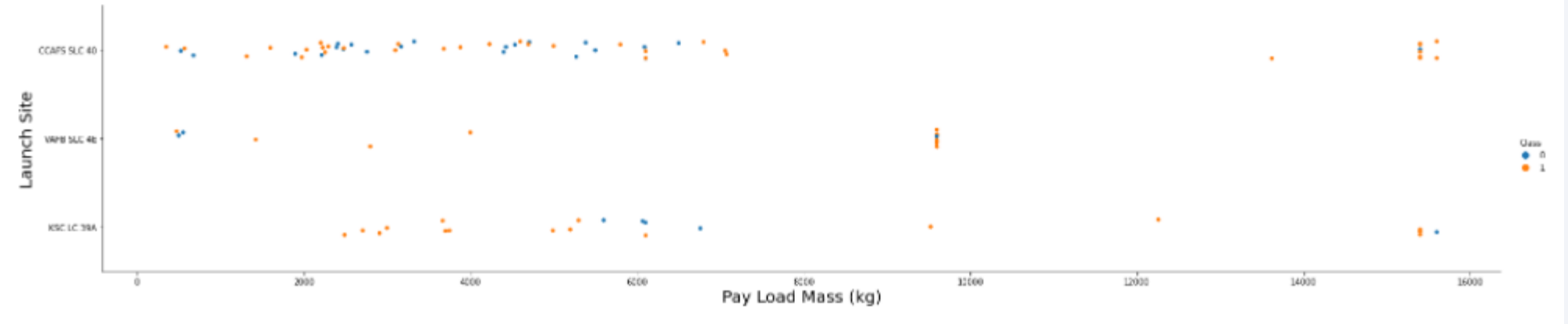


Flight Number vs. Launch Site



- This graph shows launches over time and which site the flights launched from, as well as the success of recovery. This plot shows that VAFB SLC 40 has not been used in the most recent launches. Also, the most recent launches have a high success rate.

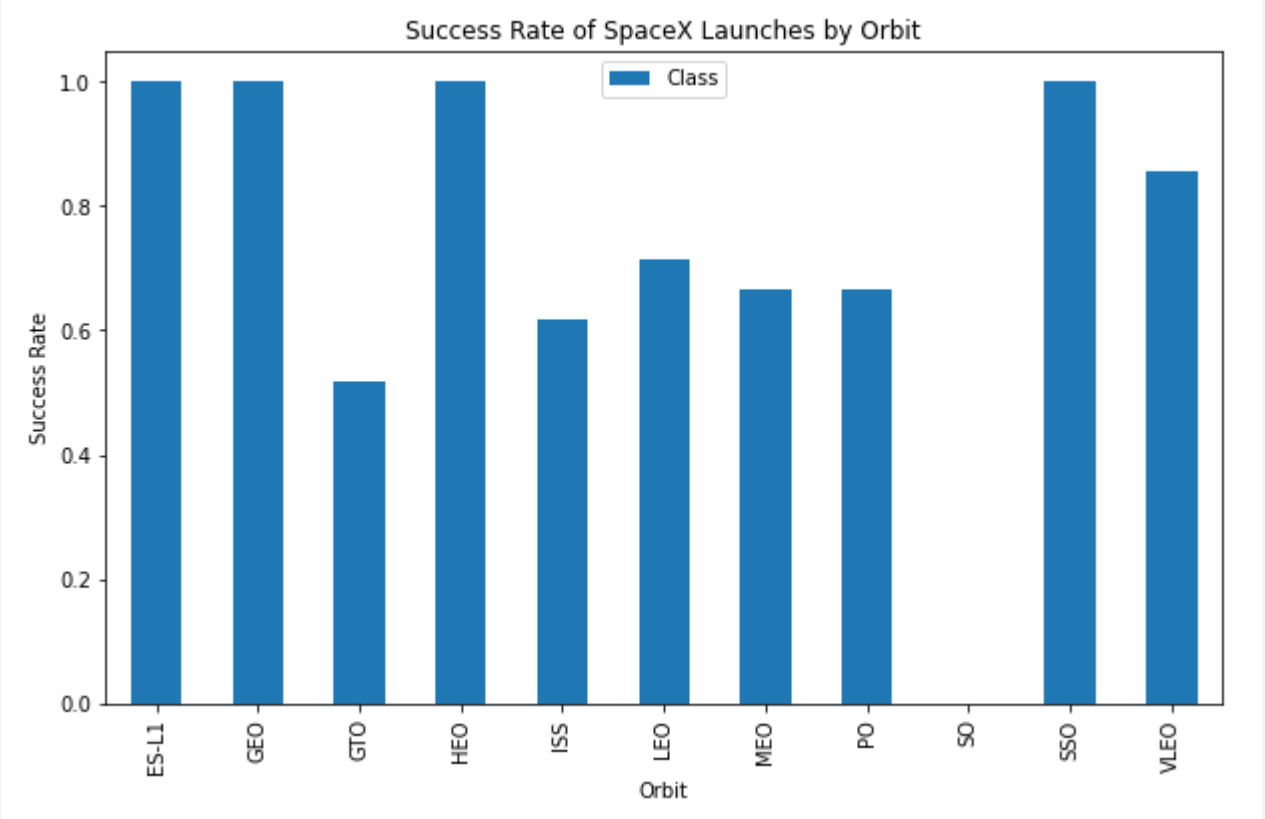
Payload vs. Launch Site



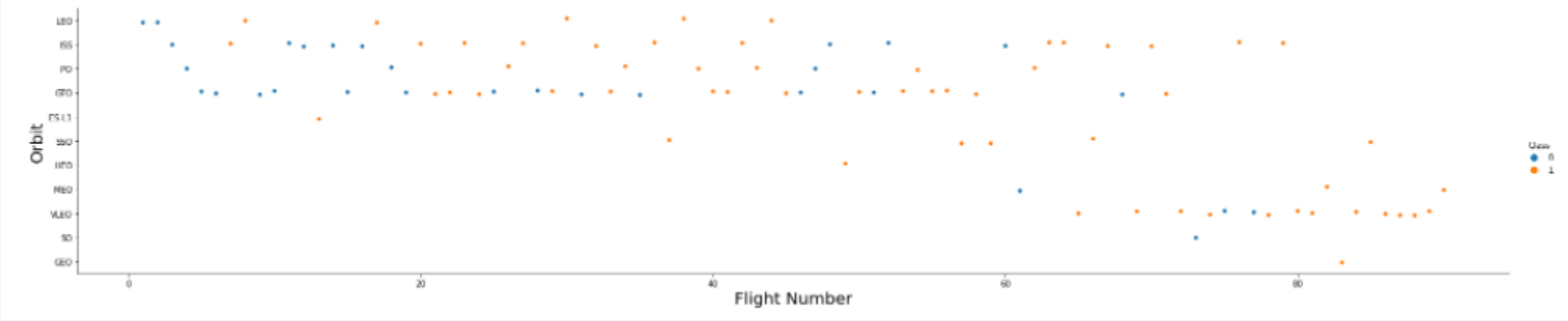
- In this plot, we show that there is little correlation between launch site and payload mass and launch success, although VAFB SLC 40 has not been used to launch the heaviest payloads.

Success Rate vs. Orbit Type

- The orbit destinations with the highest success rate are ES-L1, GEO, HEO, and SSO, with those orbits achieving 100% success rate.

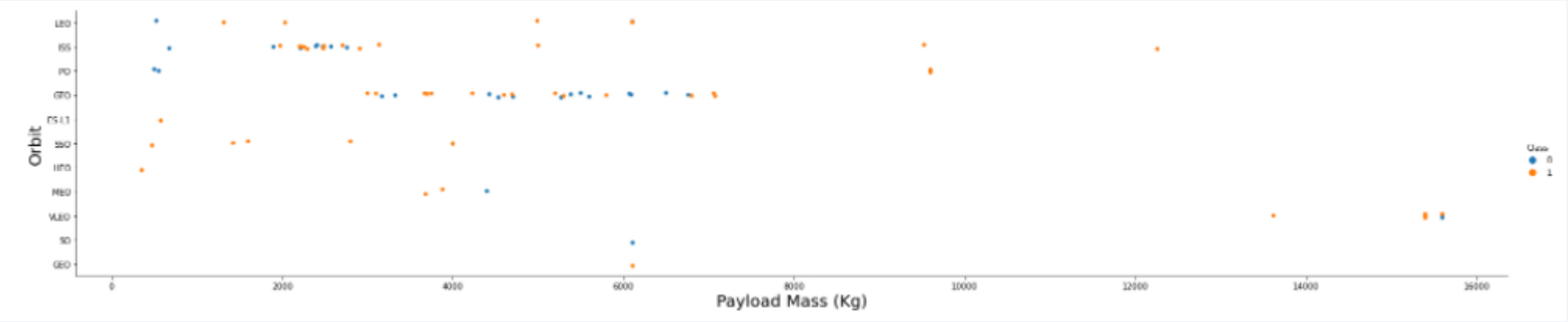


Flight Number vs. Orbit Type



- In this plot, we see that different orbits have been launched for over time. LEO were some of the first launches, but LEO has not been launched for since approximately flight 45. VLEO started around flight 65 and has been the majority of the most recent launches.

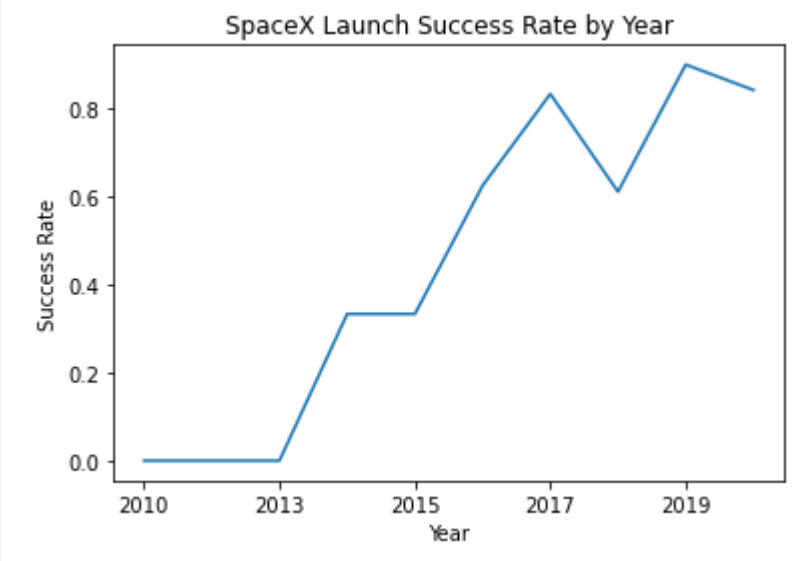
Payload vs. Orbit Type



- In this plot, we see that VLEO has had some of the heaviest payloads, and MEO has some of the lightest payloads.

Launch Success Yearly Trend

- The overall success rate of launches for SpaceX has risen from 0 in 2013 to over 80% in 2019, with a slight dip in 2018. it may be worth further investigation to determine the cause of the dip in 2018 in future studies.



All Launch Site Names

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

Here is a select distinct query to obtain each individual launch site used by SpaceX. There are four different launch sites: two CCAFS, one KSC, and one VAFB. Now we know all of the launch sites used by SpaceX.

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

- Here are a sampling of 5 launches from CCA sites to see generally what kind of data we are looking at.

Total Payload Mass

```
%sql select sum(PAYLOAD_MASS__KG_) from SPACEXTBL WHERE (CUSTOMER = 'NASA (CRS)')
```

```
 * db2://ytf13914:***@2f3279a5-73d1-4859-88f0-a6c3e6b4b907.c3n41cmd0nqnrk39u98g.da
Done.
    1
45596
```

- This query shows the total amount of weight launched to orbit for NASA. This can be useful in calculating historical use or revenue from NASA.

Average Payload Mass by F9 v1.1

```
%sql select avg(PAYLOAD_MASS__KG_) from SPACEXTBL WHERE (BOOSTER_VERSION LIKE 'F9 v1.1%')
```

```
 * db2://ytf13914:***@2f3279a5-73d1-4859-88f0-a6c3e6b4b907.c3n41cmd0nqnrk39u98g.databases.a
Done.
```

1
2534

- From this query we can calculate the average weight carried by the F9 booster. This will help us get an idea for future needs when using the F9 booster.

First Successful Ground Landing Date

```
%sql select min(DATE) from SPACEXTBL WHERE (LANDING__OUTCOME = 'Success (ground pad)')
```

```
 * db2://ytf13914:***@2f3279a5-73d1-4859-88f0-a6c3e6b4b907.c3n41cmd0nqnrk39u98g.databas
Done.
```

1
2015-12-22

- This query shows us the date the milestone of a successful landing on land was achieved. This can be used to gauge the relative difficulty of the accomplishment.

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%sql select DISTINCT BOOSTER_VERSION from SPACEXTBL where ((LANDING__OUTCOME = 'Success (drone ship)') AND (PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000))
```

```
 * db2://ytf13914:***@2f3279a5-73d1-4859-88f0-a6c3e6b4b907.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:30756/bludb
Done.
```

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

- The only booster to have a successful outcome of landing on the drone ship with a payload mass between 4,000 and 6,000 kg was the F9 booster, showing that future launches with a payload of that size require the F9 booster.

Total Number of Successful and Failure Mission Outcomes

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

- Here we see that over all launches, the mission was successful approximately 99% of the time, with only 1 total failure and 1 mission where the outcome was in question.

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

- Here are the booster versions carrying the heaviest payload weight attempted by SpaceX. The results show the F9 B5 has been used to carry the maximum payload.

2015 Launch Records

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

- In 2015, these were the missions that resulted in a failed landing on the drone ship. The F9 v1.1 was launched from CCAFS LC-40 in both instances. This indicates a potential problem with the F9 v1.1 when attempting a drone ship landing.

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

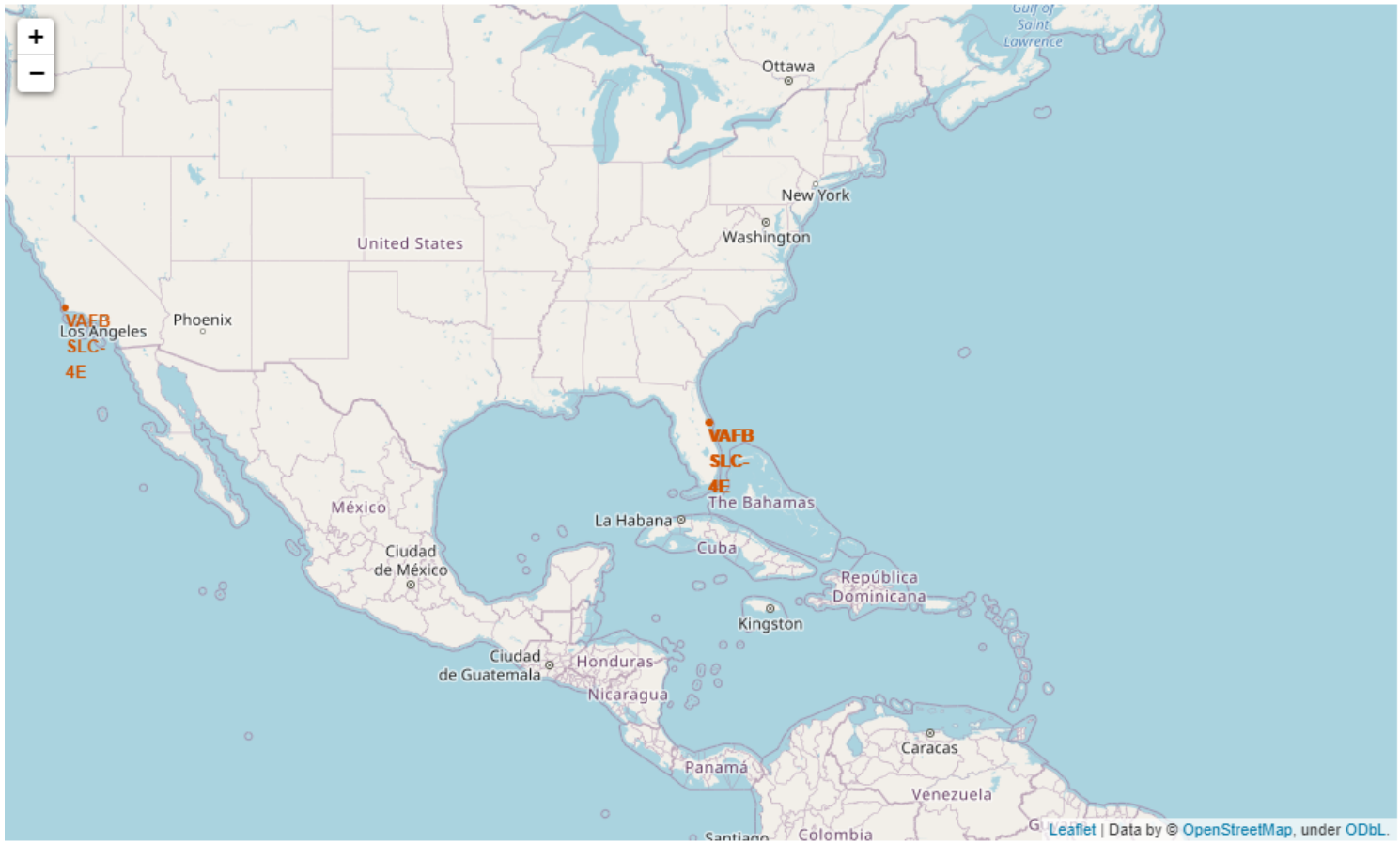
landing__outcome	count_of_outcome
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

- This query shows that for many of the launches between 2010-06-04 and 2017-03-20, a landing was not attempted. For those launches with a landing attempt, most were attempted on a drone ship, and the success rate was 50%.

Section 4

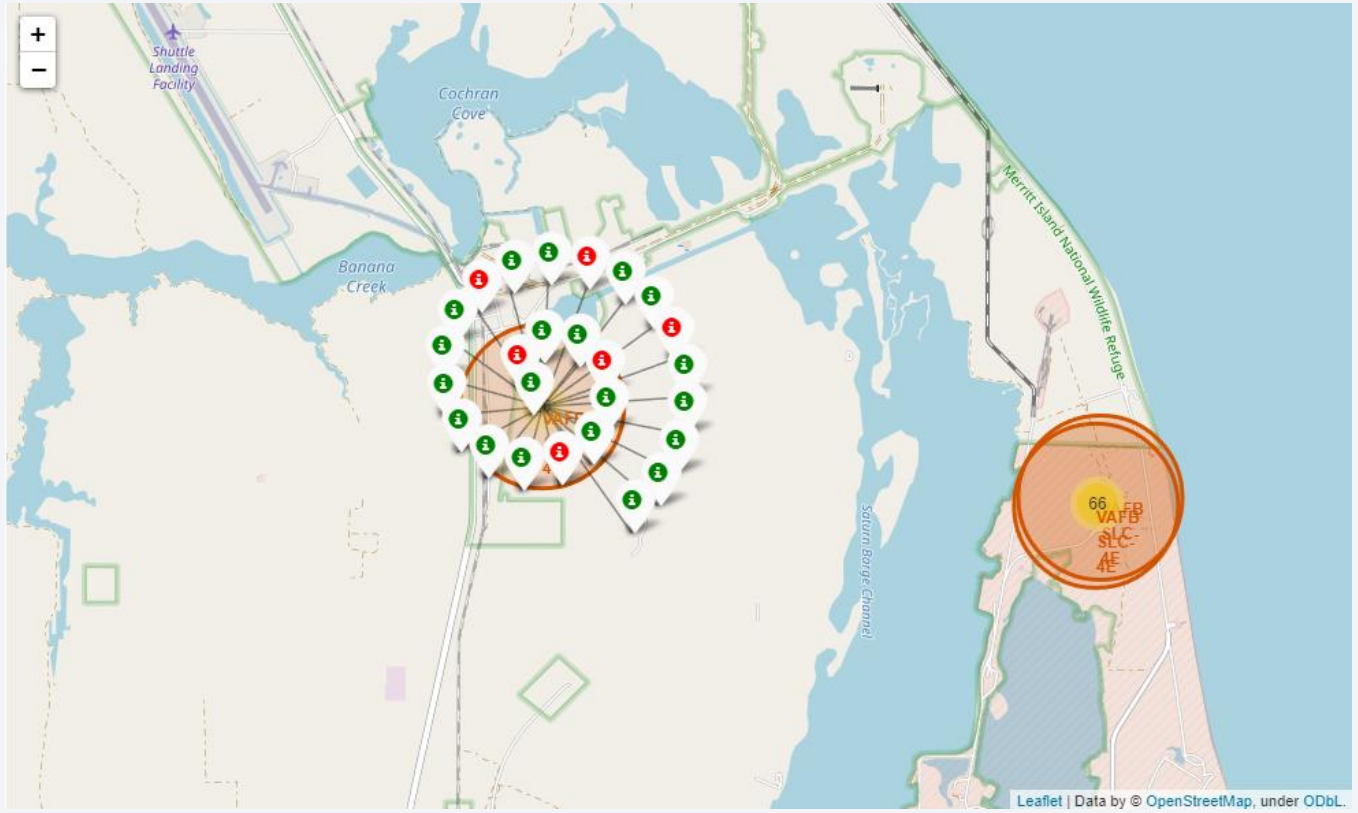
Launch Sites Proximities Analysis

Launch Site Locations



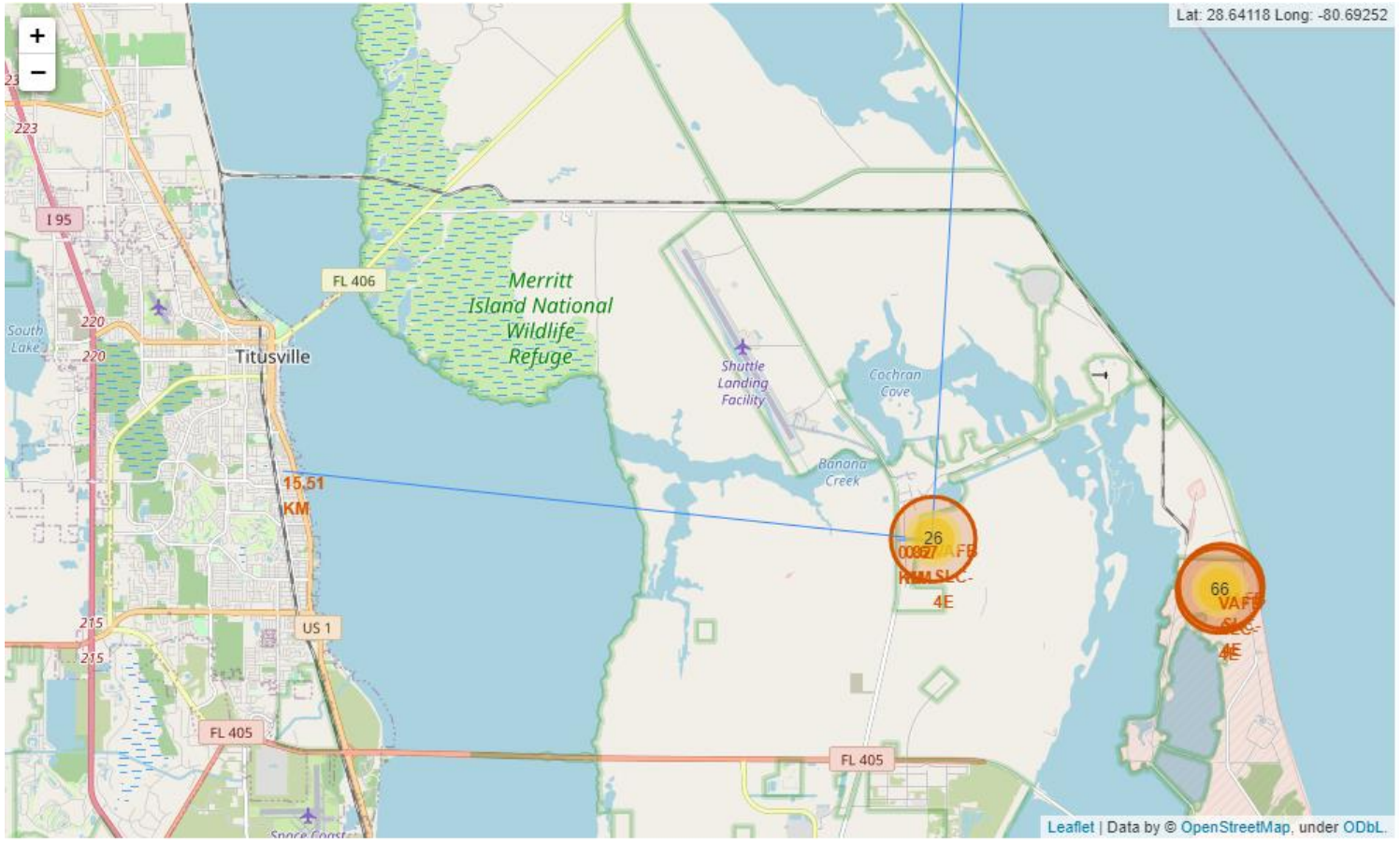
- This maps shows the two main geographic locations for SpaceX launches, the coast of California and the coast of Florida.

Launch outcomes at VAFB SLC 4E



- This maps graphically shows the launch outcomes at VAFB SLC 4E. The successful launches are in green, and the failures are in red.

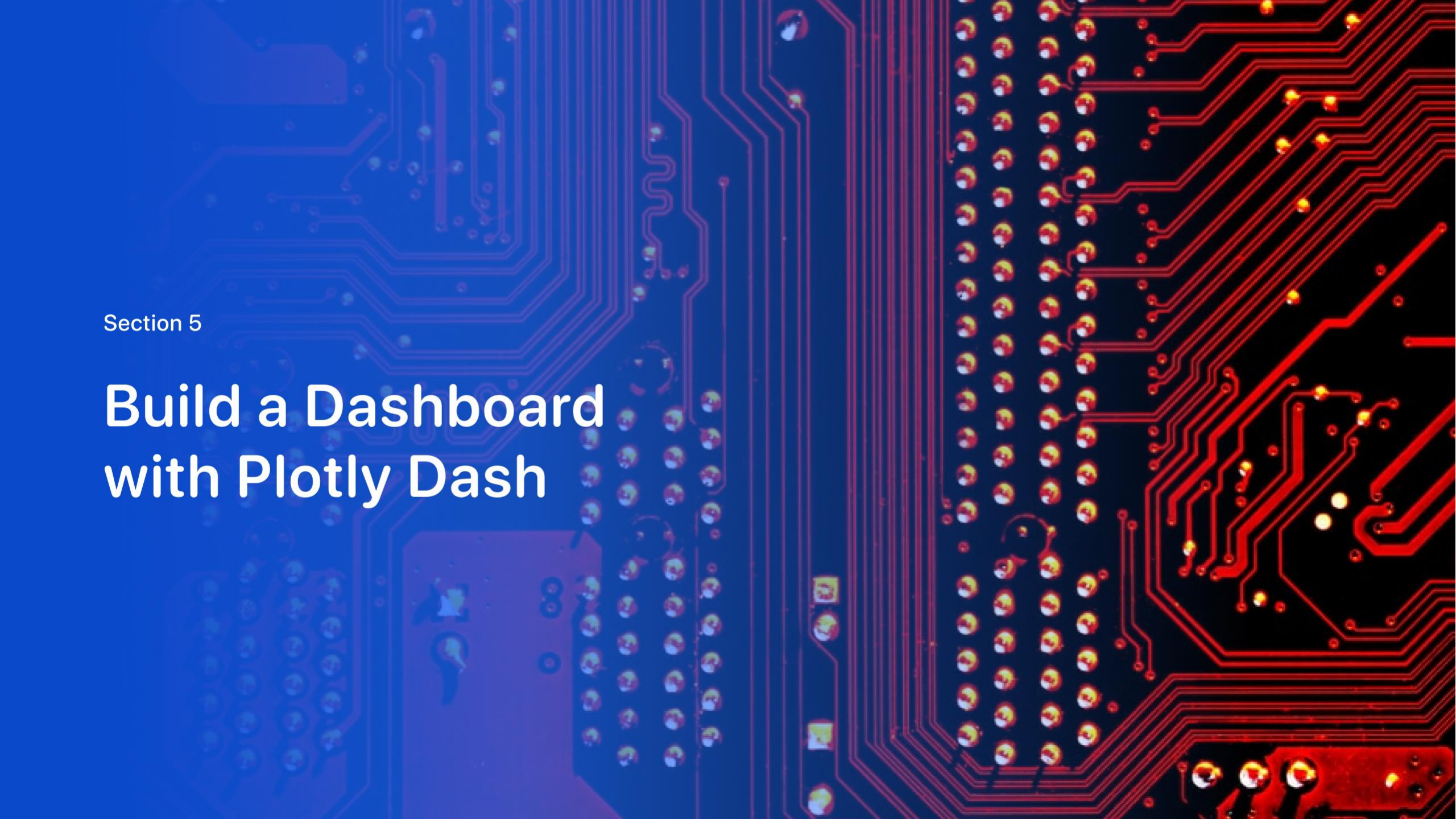
Launch site proximity to important features



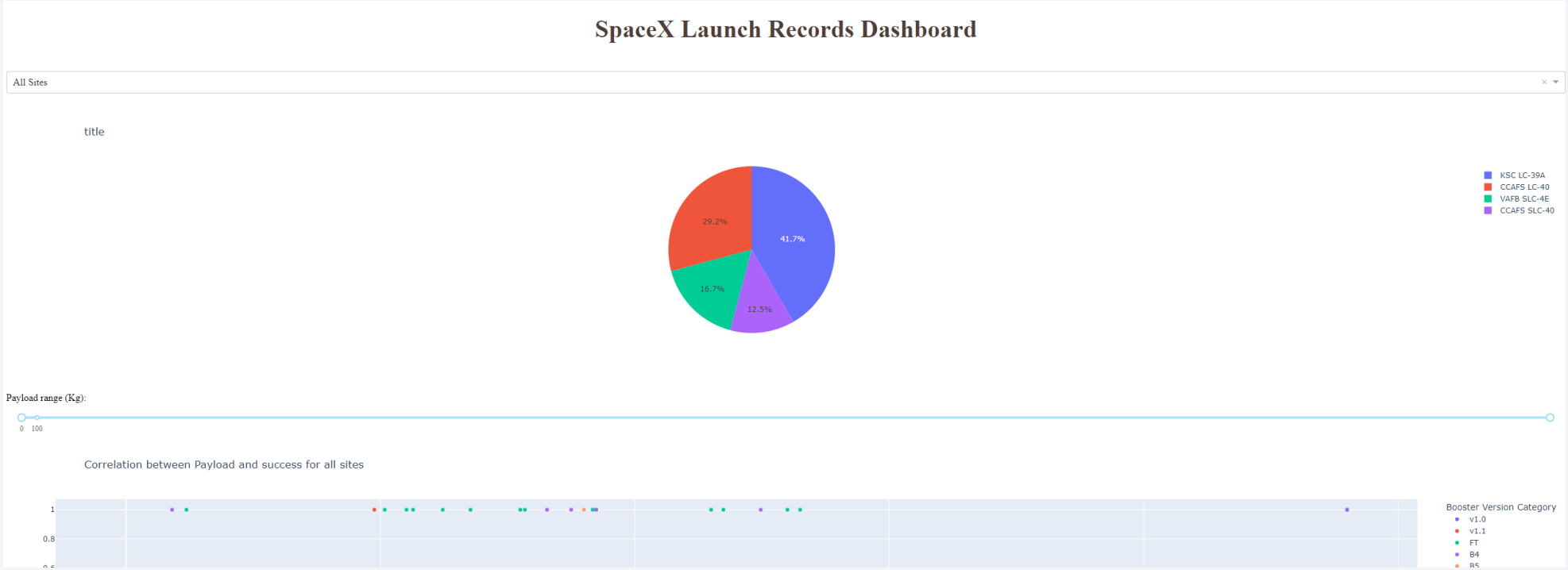
- The VAFB SLC 4E launch site is 15.51 km from the nearest town of Titusville.

Section 5

Build a Dashboard with Plotly Dash

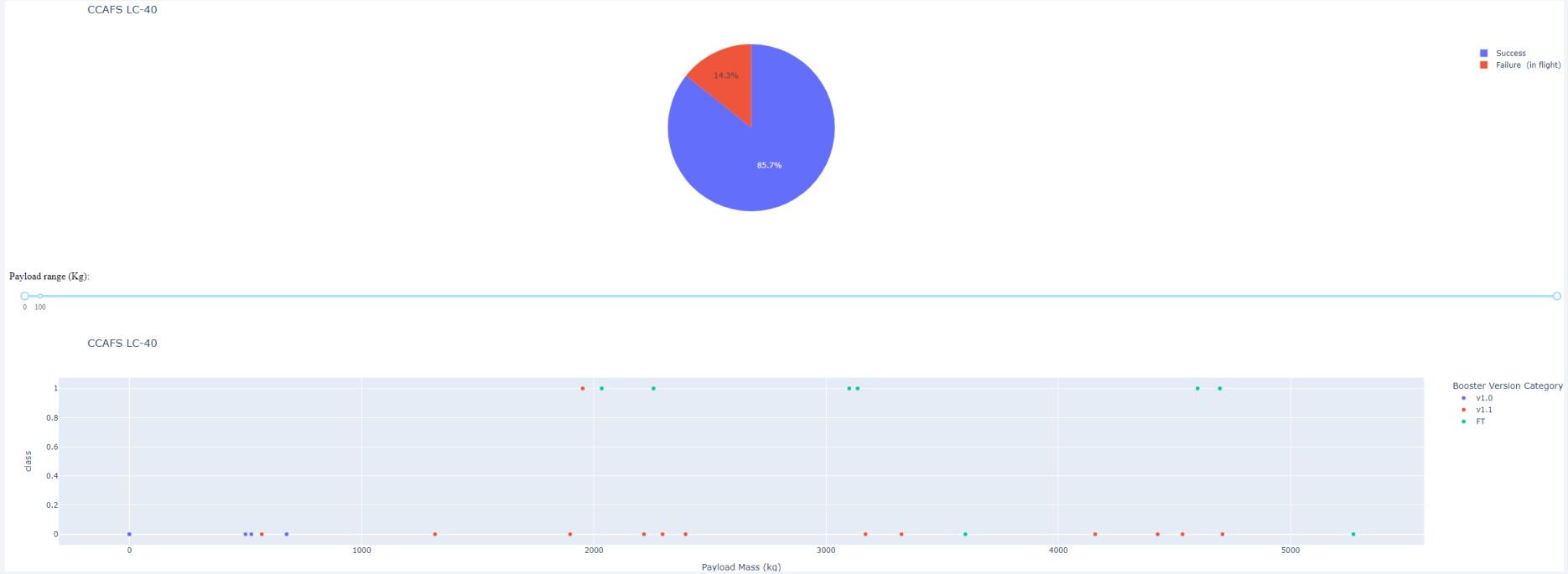


Launch Success Pie Chart



- This pie chart depicts the contribution of successful launches from each site.

Success Rate of Most Successful Launch Site



- This pie chart and scatter plot depict the success rate of CCAFS LC-40, the launch site with the highest success rate.

Payload Mass and Booster Version Success Rate

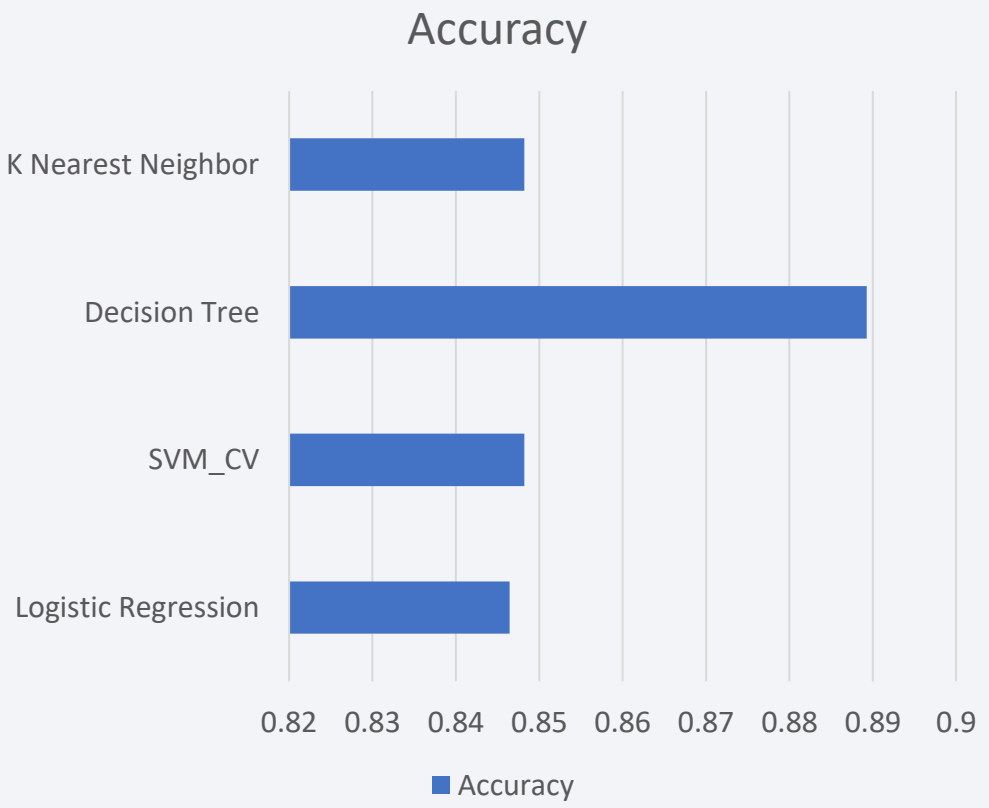
- This scatter plot shows the success of launches plotted against the payload mass, with the booster version color coded.

Section 6

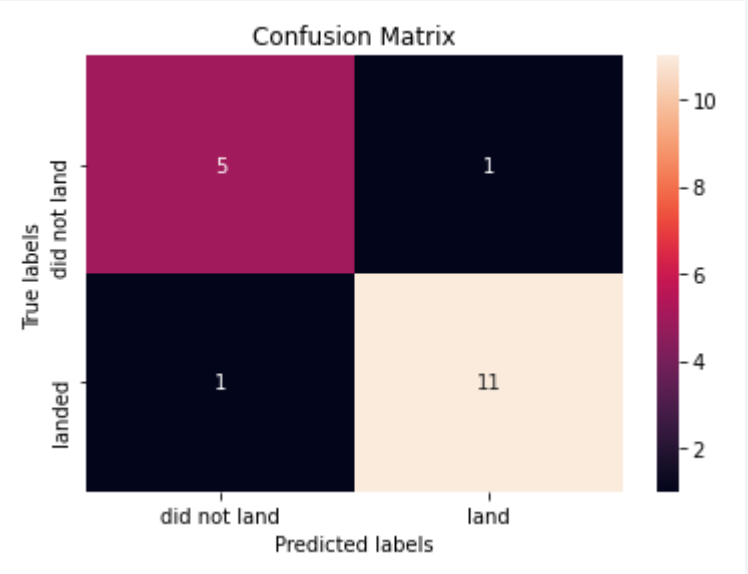
Predictive Analysis (Classification)

Classification Accuracy

- According to the accuracy scores for each model, the decision tree resulted in the highest accuracy against the test set when trained with a sample set.
- For predicting the success of future landings when considering variables such as launch site, booster version, orbit, and payload mass.



Confusion Matrix



- Here we see the decision tree's relative successes and failures on accurately predicting whether a landing would have been a success or failure. The model correctly guessed 11 landings and 5 non-landings, and incorrectly guessed the outcome two times.

Conclusions

- Based on yearly trends, SpaceX's launch success is on track to be near 100% in the near future
- Launch Site KSC LC-39A has the highest historical launch success rate
- VAFB SLC 4E is located close to a residential area
- The best model to predict launch outcomes based on payload mass, destination orbit, and booster version is the Decision Tree.

Thank you!