IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Alexander DeLoach
February 19, 2023

Outline

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

Executive Summary

- Summary of methodologies
- Summary of all results

Introduction

- SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars, much cheaper than many competitors.
- Most of the savings is because SpaceX can reuse the first stage. If we can determine if the first stage will land, we can determine the cost of a launch.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Webscraping & Requests to the SpaceX API
- Perform data wrangling
  - Data was put into a Pandas DataFrame and then filtered to show Falcon 9 Launches only
- 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

- Data collected using requests to the SpaceX API & filtered to include only Falcon 9 launches.
- Data was then filtered to only include Falcon 9 launches
- Missing Values in the column "Payload Mass" were replaced by the average Payload Mass for Falcon 9 Launches

Data Collection – SpaceX API

- The JSON file was assigned to the static response object
- The file was then decoded and turned into a Pandas Dataframe
- [IBM-Applied-Data-Science-Capstone-Project/Lab 1 - Data Collection API.ipynb at main · kingloach/IBM-Applied-Data-Science-Capstone-Project (github.com)]

```
static_json_url='https://cf-courses-data.s3.us.cloud-object-storage.
```

We should see that the request was successfull with the 200 status response code

```
response.status_code
```

```
200
```

Now we decode the response content as a Json using `.json()` and turn it into a Pandas dataframe using `.json_normalize()`

```
# Use json_normalize meethod to convert the json result into a dataf
response1 = requests.get(static_json_url)
data = response1.json()

data = pd.json_normalize(data, max_level=0)
```

Data Collection - Scraping

- An HTTP GET method was used to request the Falcon9 Launch HTML page from Wikipedia
- A BeautifulSoup object was then created from the HTML response
- [IBM-Applied-Data-Science-Capstone-Project/Lab 2 - Data Collection with Webscraping.ipynb at main · kingloach/IBM-Applied-Data-Science-Capstone-Project (github.com)]

```
# use requests.get() method with the provided static_url
# assign the response to a object

x = requests.get(static_url).text
```

Create a `BeautifulSoup` object from the HTML `response`

```
# Use BeautifulSoup() to create a BeautifulSoup object from a respon
soup = BeautifulSoup(x, 'html.parser')
```

Print the page title to verify if the `BeautifulSoup` object was created properly

```
# Use soup.title attribute
title = soup.title
print(title)
```

```
<title>List of Falcon 9 and Falcon Heavy launches - Wikipedia</title>
```

Data Wrangling

- First the number of missing values was calculated for each attribute
- Then, the number of launches per site was calculated
- The number and occurrence of each orbit was calculated, and then the number and occurrence of each mission outcome per orbit type using the .value_counts() method
- IBM-Applied-Data-Science-Capstone-Project/Lab 3 - Data Wrangling.ipynb at main · kingloach/IBM-Applied-Data-Science-Capstone-Project (github.com)

EDA with Data Visualization

- Multiple Charts were created to visualize the data
- First chart was the Flight Number vs. PayloadMass chart
- Another chart that was created was a chart to visualize the relationship between Flight Number and Launch Sites
- IBM-Applied-Data-Science-Capstone-Project/Lab 5 - EDA Visualization.ipynb at main · kingloach/IBM-Applied-Data-Science-Capstone-Project (github.com)

EDA with SQL

- I ran SQL queries to find out to following (among other insights):
  - The names of each launch site
  - Average payload mass carried by booster version F9 v1.1
  - The first successful Ground Pad Landing
  - The failed landings in 2015
  - Successful landings between 06/04/2010 and 03/20/2017
- IBM-Applied-Data-Science-Capstone-Project/Lab 4 - EDA with SQL.ipynb at main · kingloach/IBM-Applied-Data-Science-Capstone-Project (github.com)

Build an Interactive Map with Folium

- In the interactive map, I added "Marker Clusters", indicating the locations of each Launch Site, as well as the number of launches and indicators signifying the outcome of each mission
- Lines and distance indicators were added to show launch site locations' proximity to other landmarks, such as cities, coastline, and railroads
- IBM-Applied-Data-Science-Capstone-Project/Lab 6 - Interactive Analytics with Folium (2).ipynb at main · kingloach/IBM-Applied-Data-Science-Capstone-Project (github.com)

Build a Dashboard with Plotly Dash

- Summarize what plots/graphs and interactions you have added to a dashboard
- Explain why you added those plots and interactions
- Add the GitHub URL of your completed Plotly Dash lab, as an external reference and peer-review purpose

Predictive Analysis (Classification)

- A NumPy array was created from the Mission Outcome data, and then the data was standardized and split into Train & Test sets
- Created models using Logistic Regression, Support Vector Machines, Decision Tree Classifier, and K-Nearest Neighbor objects to see which model had to highest accuracy
- Confusion Matrix was then created for each Predictive Analysis Model
- IBM-Applied-Data-Science-Capstone-Project/Lab 8 - Machine Learning Prediction.ipynb at main · kingloach/IBM-Applied-Data-Science-Capstone-Project (github.com)

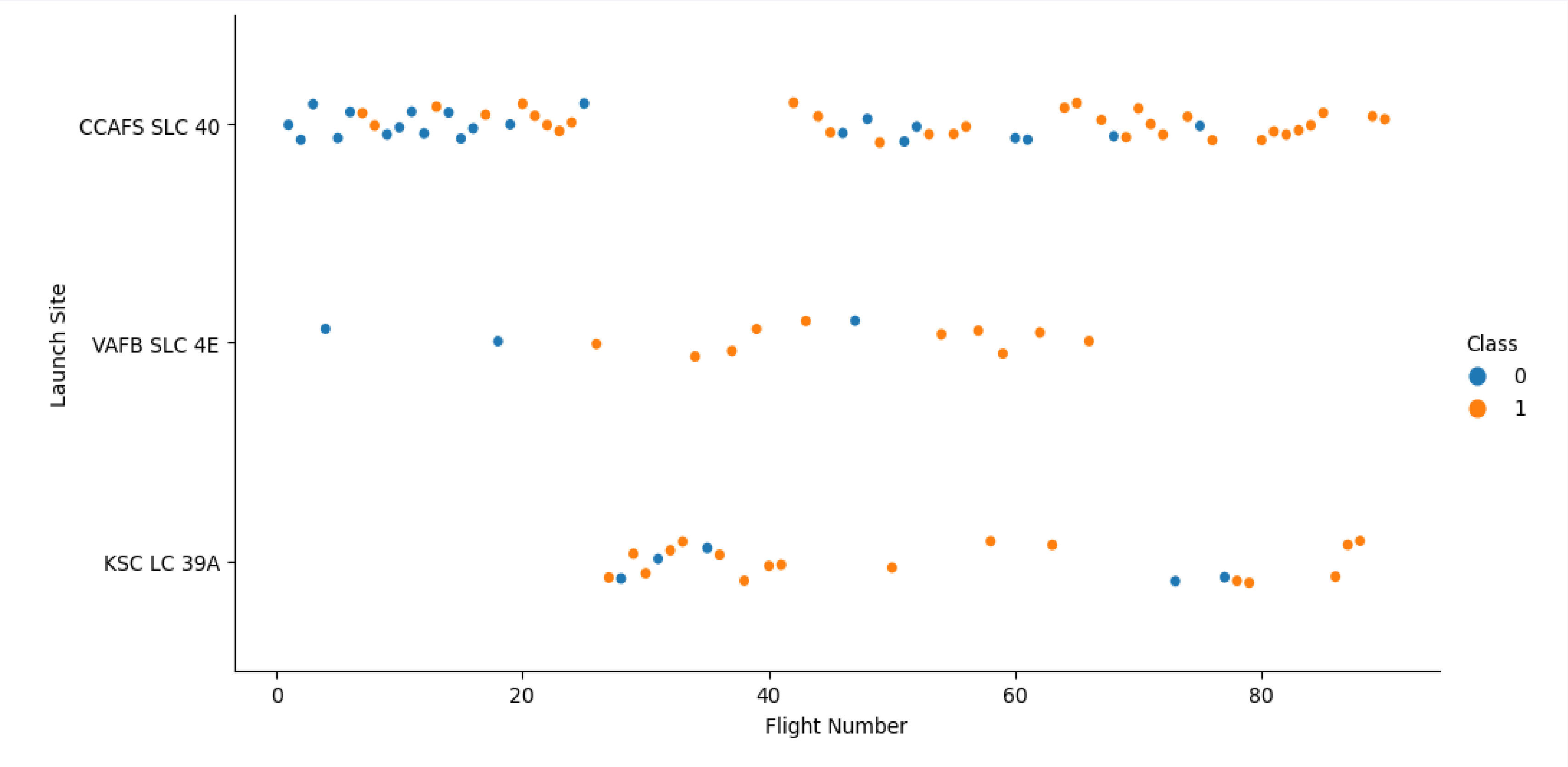
Results

- Certain Orbits have higher successful landing rates with heavy payloads
- Success Rate has generally been increasing since 2013
- The Decision Tree Classifier Model had the highest accuracy among the 4 models

Section 2

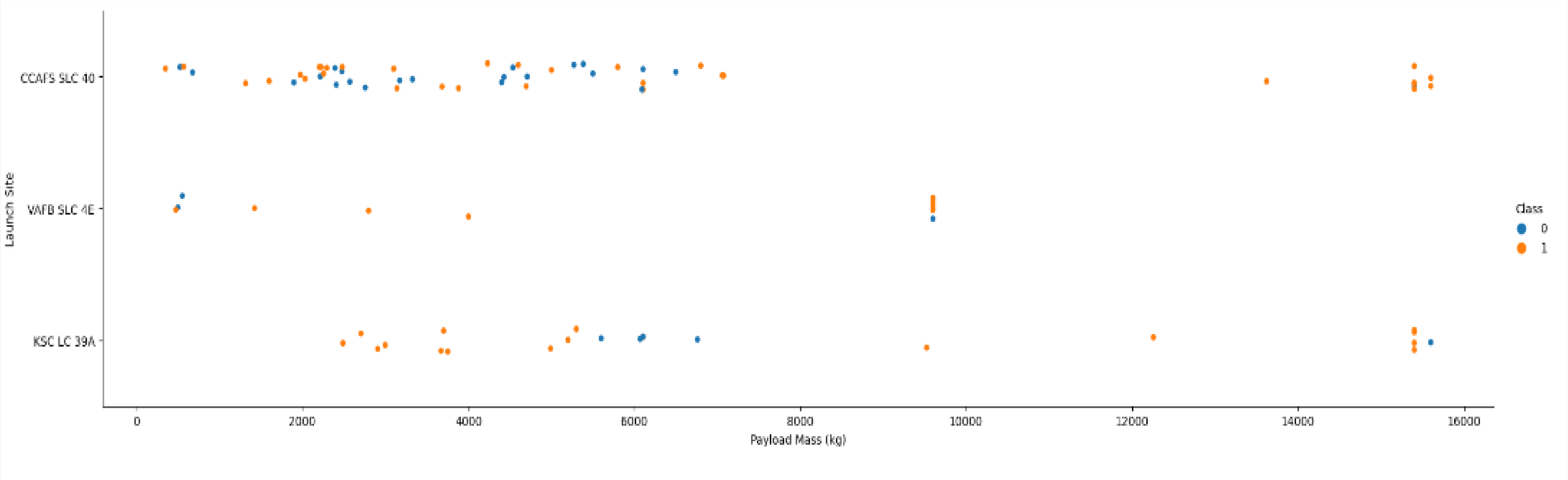
Insights drawn from EDA

Flight Number vs. Launch Site



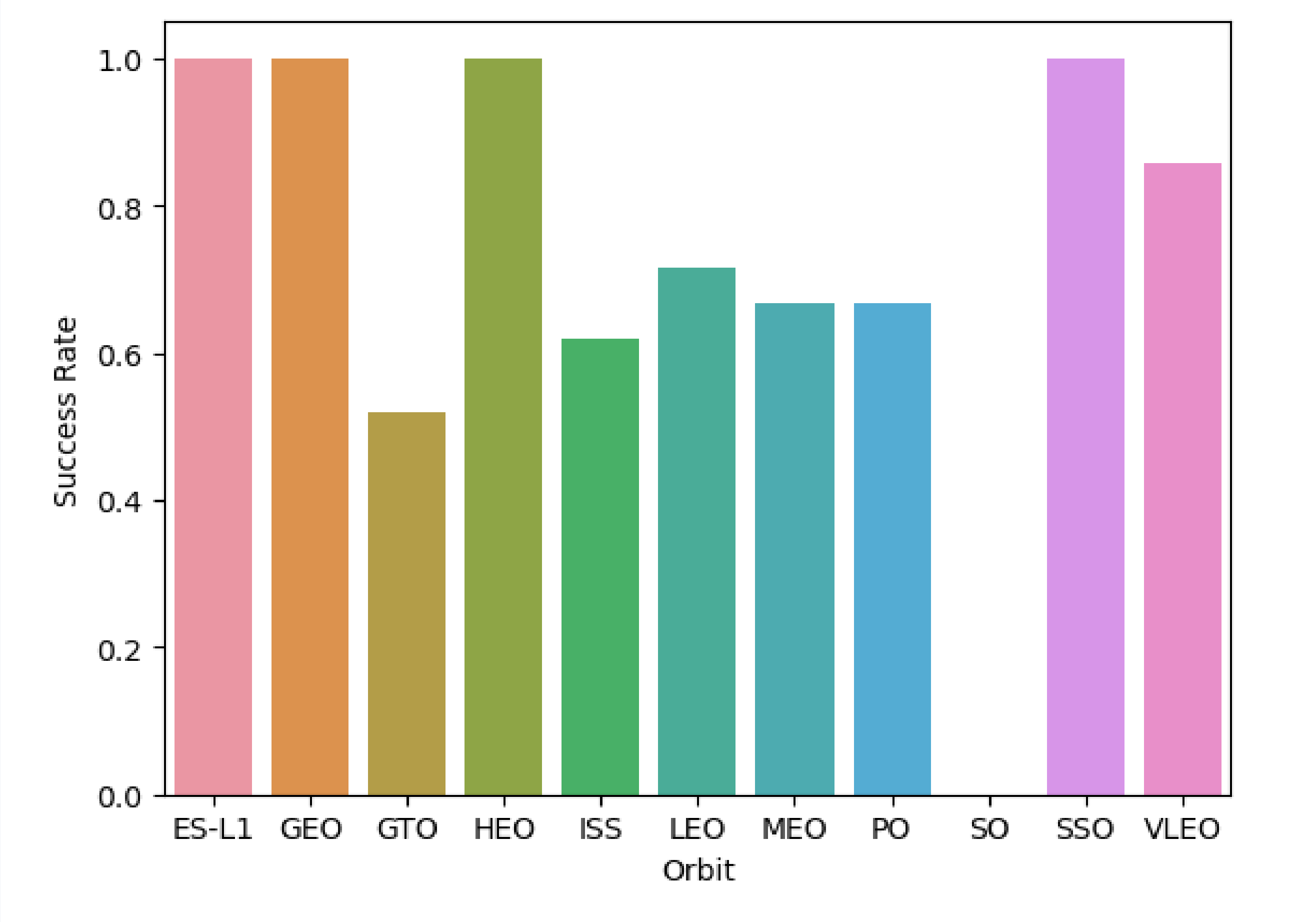
- Datapoints of Class 0 (blue points) correspond with bad landing outcomes
- Datapoints of Class 1 (orange points) correspond with successful landings
- Higher flight numbers appear more likely to have successful landings

Payload vs. Launch Site



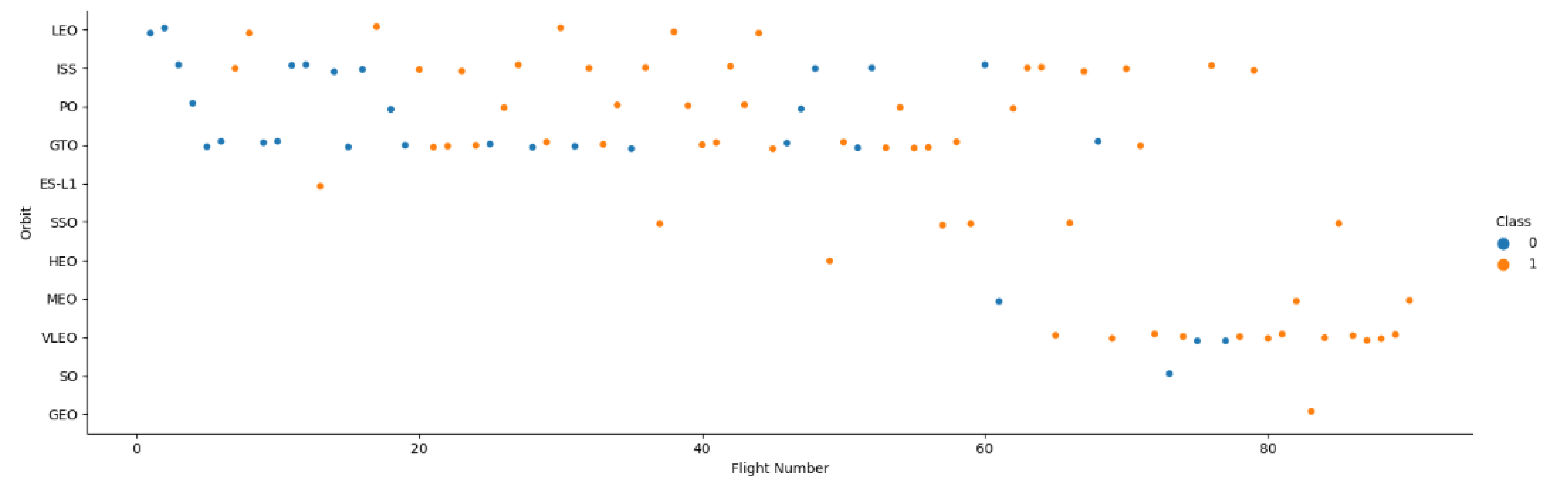
- Relationship between the Launch Site & Payload Mass
- Class 0 (blue points) correspond with unsuccessful landings
- Class 1 (orange points) correspond with successful landings

Success Rate vs. Orbit Type



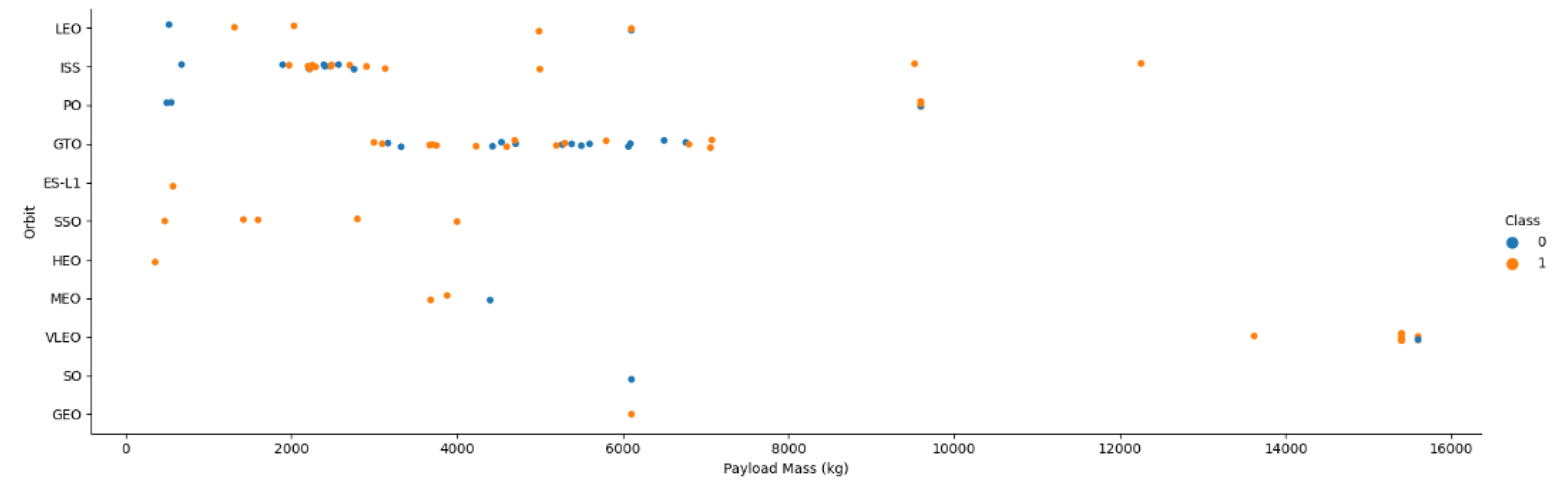
- Chart shows relationship between the type of Orbit the rocket was sent to and the average Success Rate of the mission
- ES-L1, GEO, HEO, and SSO orbits tend to have higher success rate.

Flight Number vs. Orbit Type



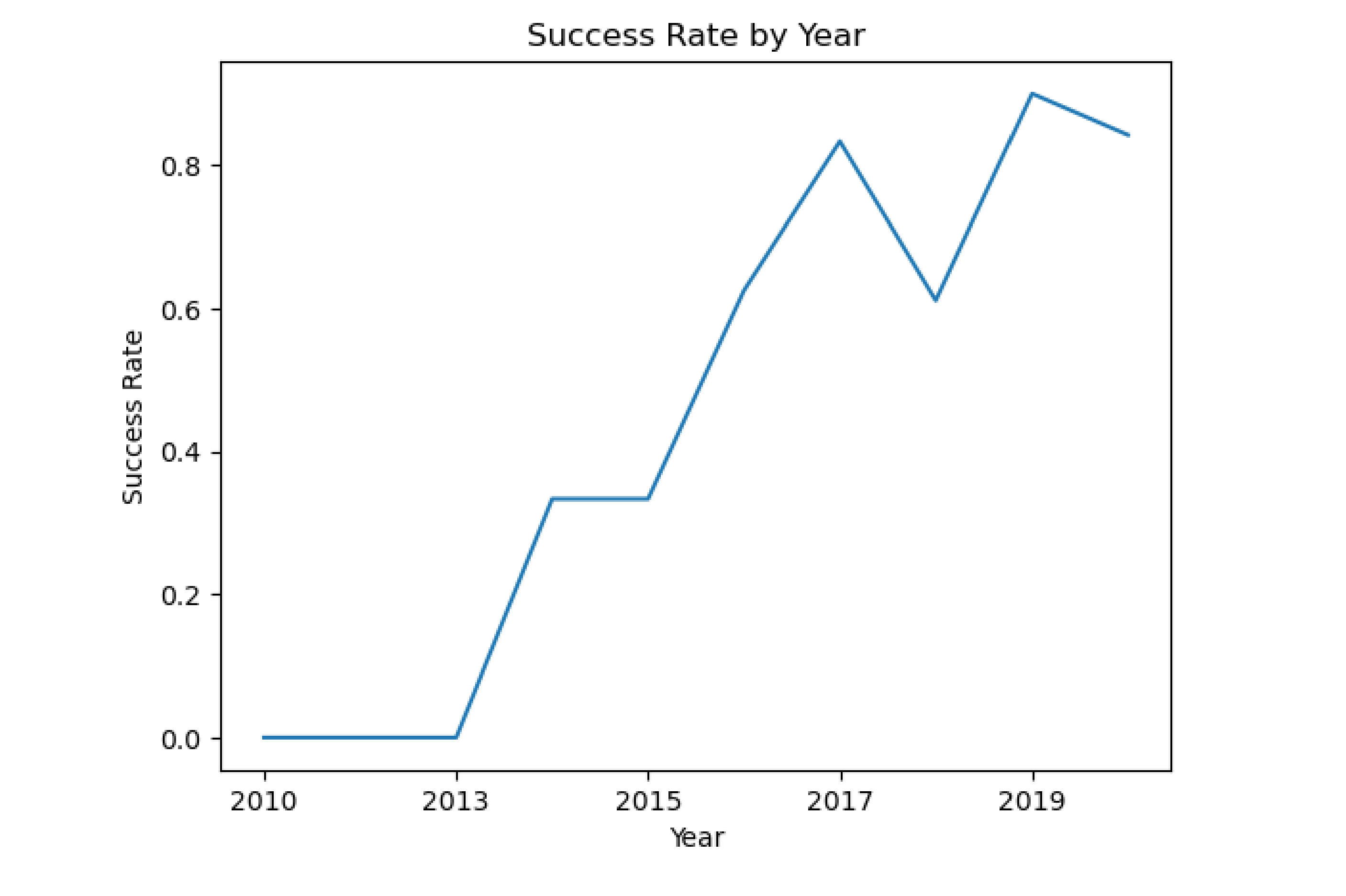
- Plot shows relationship between Orbit Type and Number of Flights
- In LEO orbit, success appears to correlate to Number of Flights

Payload vs. Orbit Type



- Heavier Payloads are more successful in ISS, LEO and Polar orbit

Launch Success Yearly Trend



- Success generally trending upward since 2013
- 2018 and 2020 only years with a lower success rate than the year prior

All Launch Site Names

- Using a SQL query, the names of the 4 launch sites used were displayed. They are as follows:

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

Launch Site Names Begin with 'CCA'

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome
04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
08-12-2010	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)
22-05-2012	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
01-03-2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

- 5 occurrences where the Launch Site names begin with ‘CCA’

Total Payload Mass

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

```
%sql SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE Customer = 'NASA(CRS)';
```

```
 * sqlite:///my_data1.db
Done.
```

SUM(PAYLOAD_MASS__KG_)
None

- Result of query to display total Payload Mass

Average Payload Mass by F9 v1.1

- The average Payload Mass carried by Booster Version F9 v1.1 is displayed below

AVG(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

- The first successful Ground Pad Landing date was:

MAX(Date)
22-12-2015

Successful Drone Ship Landing with Payload between 4000 and 6000

- The following Boosters Have had successful Drone Ship landings while carrying Payload Mass between 4000 kg and 6000 kg:

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

Total Number of Successful and Failure Mission Outcomes

- The total number of Successful and Failure Mission Outcomes is:

COUNT (Mission_Outcome)
100

Boosters Carried Maximum Payload

- The following boosters have carried the Maximum Payload:

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

2015 Launch Records

- The following are the failed Drone Ship landings, their booster version, their launch sites, and the month number they were launched in.

substr(Date, 4, 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

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

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome
19-02-2017	14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
18-10-2020	12:25:57	F9 B5 B1051.6	KSC LC-39A	Starlink 13 v1.0, Starlink 14 v1.0	15600	LEO	SpaceX	Success	Success
18-08-2020	14:31:00	F9 B5 B1049.6	CCAFS SLC-40	Starlink 10 v1.0, SkySat-19, -20, -21, SAOCOM 1B	15440	LEO	SpaceX, Planet Labs, PlanetIQ	Success	Success
18-07-2016	04:45:00	F9 FT B1025.1	CCAFS LC-40	SpaceX CRS-9	2257	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
18-04-2018	22:51:00	F9 B4 B1045.1	CCAFS SLC-40	Transiting Exoplanet Survey Satellite (TESS)	362	HEO	NASA (LSP)	Success	Success (drone ship)
17-12-2019	00:10:00	F9 B5 B1056.3	CCAFS SLC-40	JCSat-18 / Kacific 1, Starlink 2 v1.0	6956	GTO	Sky Perfect JSAT, Kacific 1	Success	Success
16-11-2020	00:27:00	F9 B5B1061.1	KSC LC-39A	Crew-1, Sentinel-6 Michael Freilich	12500	LEO (ISS)	NASA (CCP)	Success	Success
15-12-2017	15:36:00	F9 FT B1035.2	CCAFS SLC-40	SpaceX CRS-13	2205	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)

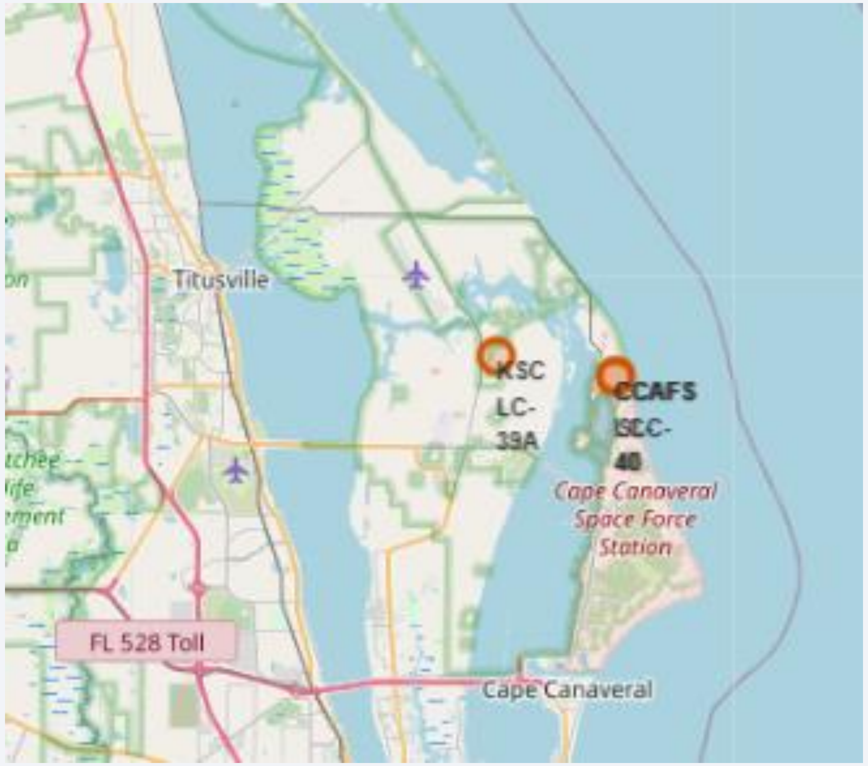
- Successful landing outcomes between 2010-06-04 and 2017-03-20, sorted in Descending order by date. (first 8 displayed)

Section 3

Launch Sites Proximities Analysis

Launch Site Locations on Map

- Locations of the Launch Sites shown below

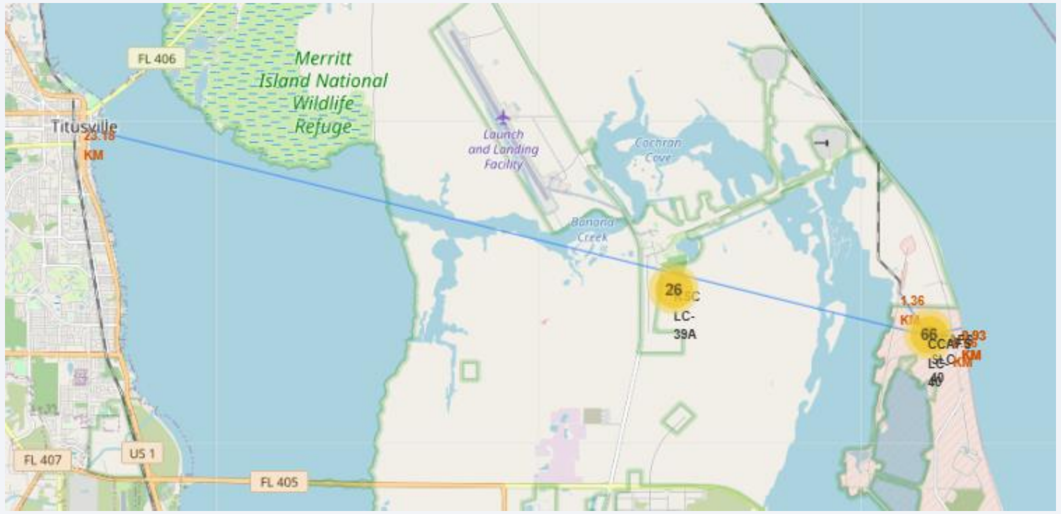


Locations with Occurrence Number

- Locations shown on a world map, including the number of occurrences for each Launch Site

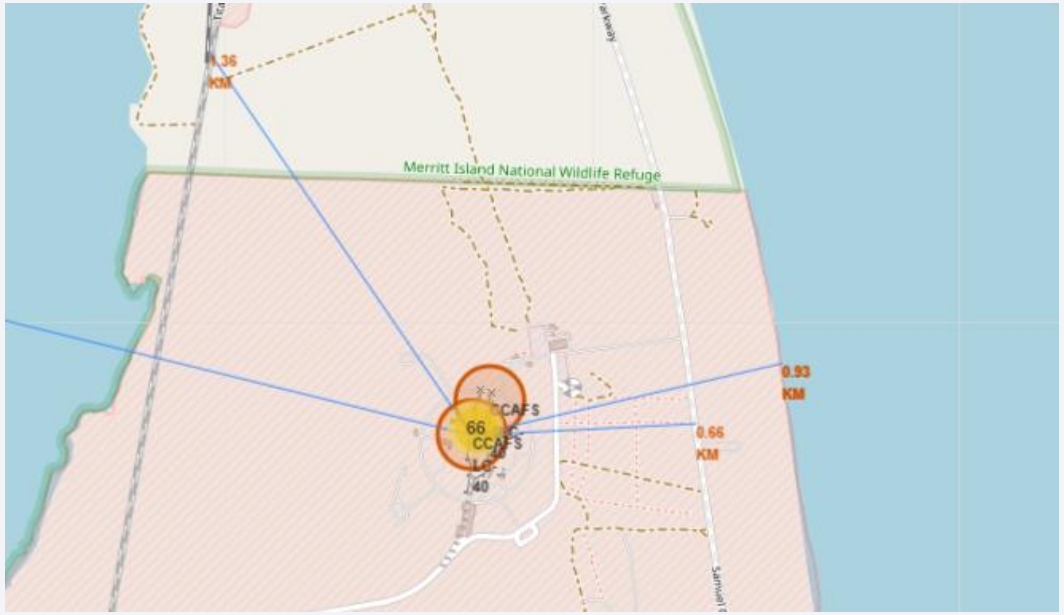
- In each cluster, the occurrence were color-coded to indicate the landing outcome (green = successful landing, red = unsuccessful landing)

Launch Site Proximities to Other Landmarks



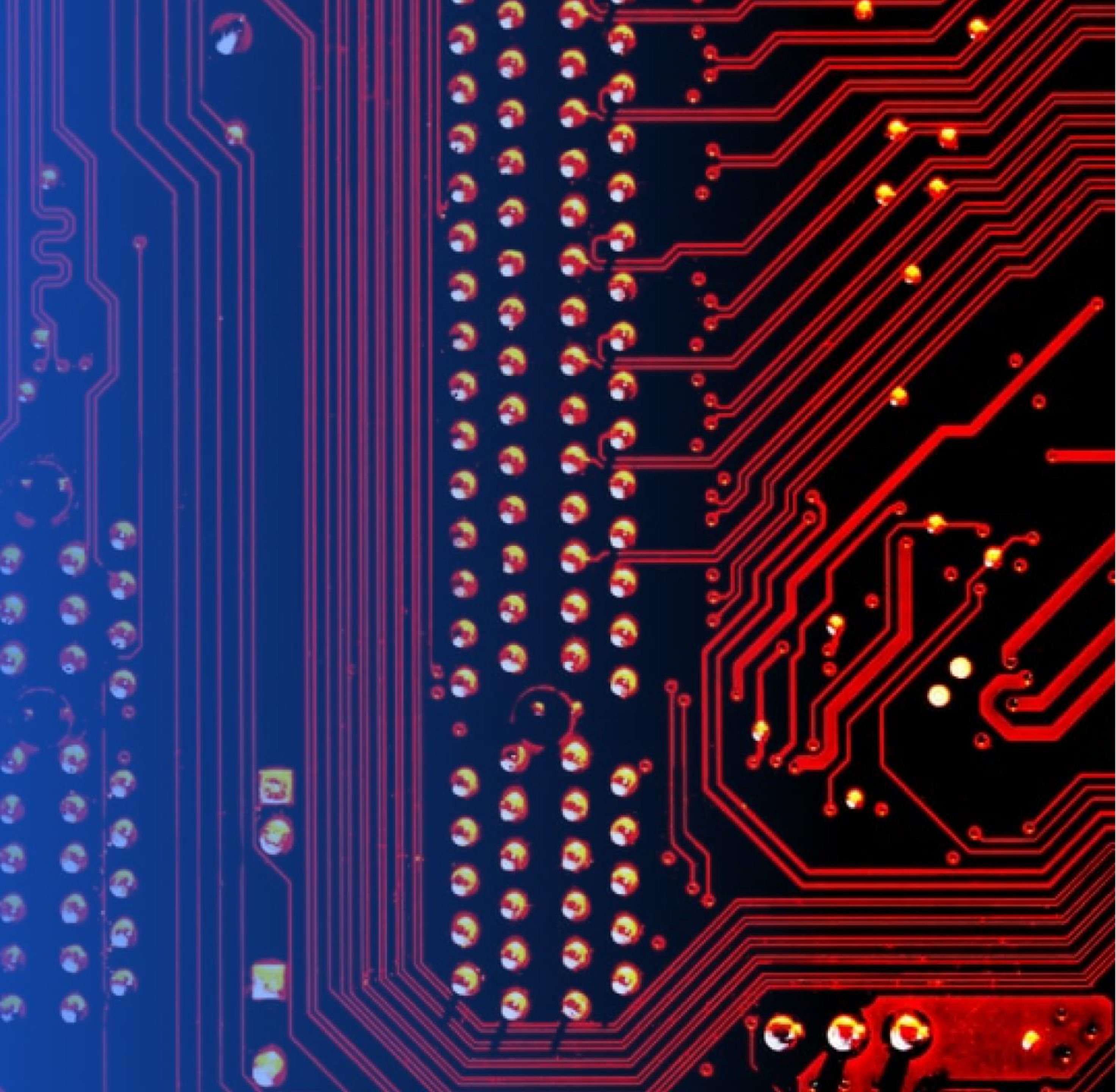
- See the lines, along with the distances to the nearest city, railroad, coastline, and highway

- The launch sites are generally close to coastlines, railroads, or highways
- Tend to keep distance from cities



Section 4

Build a Dashboard with Plotly Dash



<Dashboard Screenshot 1>

- Replace <Dashboard screenshot 1> title with an appropriate title
- Show the screenshot of launch success count for all sites, in a piechart
- Explain the important elements and findings on the screenshot

<Dashboard Screenshot 2>

- Replace <Dashboard screenshot 2> title with an appropriate title

- Show the screenshot of the piechart for the launch site with highest launch success ratio

- Explain the important elements and findings on the screenshot

<Dashboard Screenshot 3>

- Replace <Dashboard screenshot 3> title with an appropriate title

- Show screenshots of Payload vs. Launch Outcome scatter plot for all sites, with different payload selected in the range slider

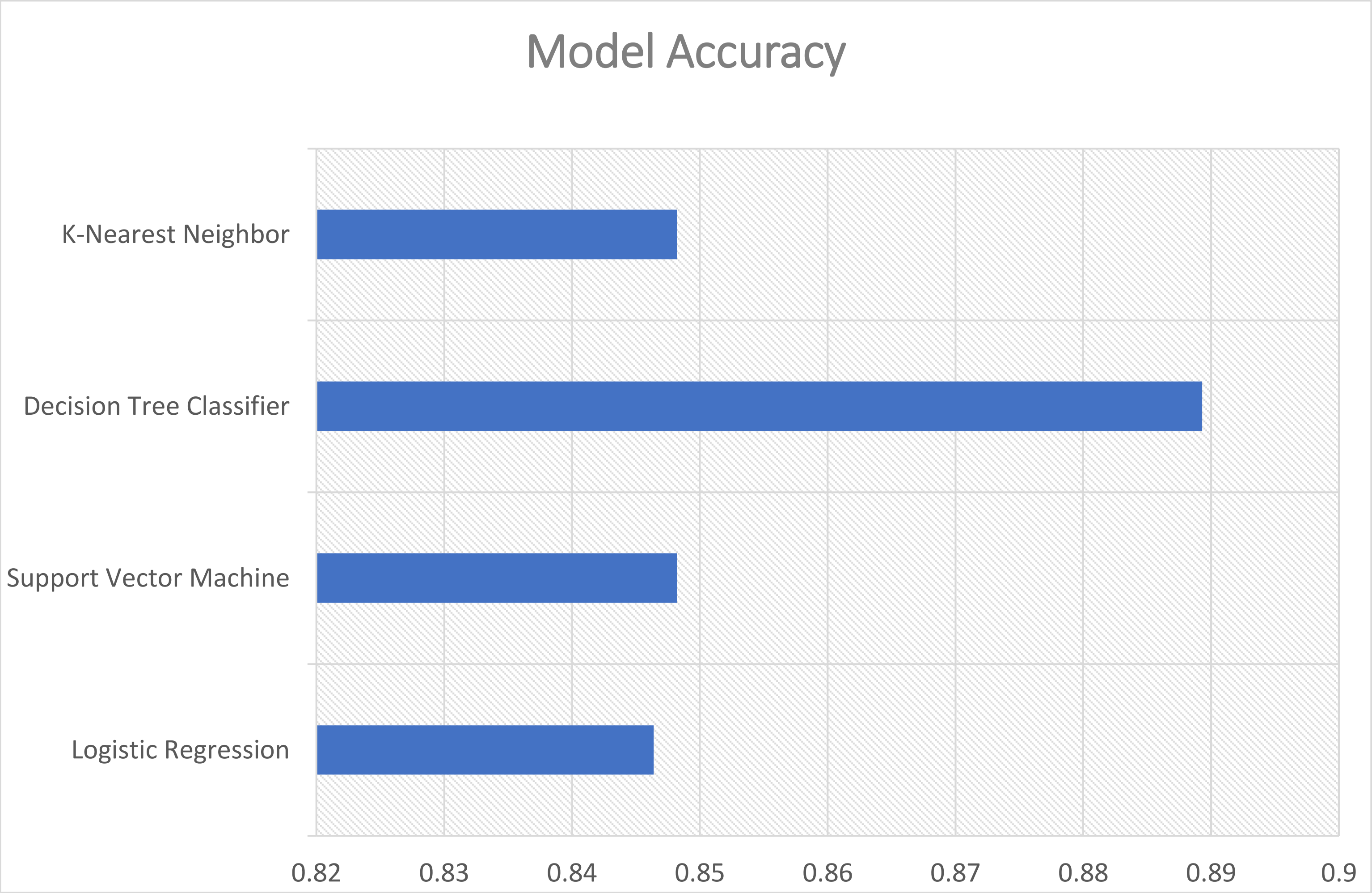
- Explain the important elements and findings on the screenshot, such as which payload range or booster version have the largest success rate, etc.

Section 5

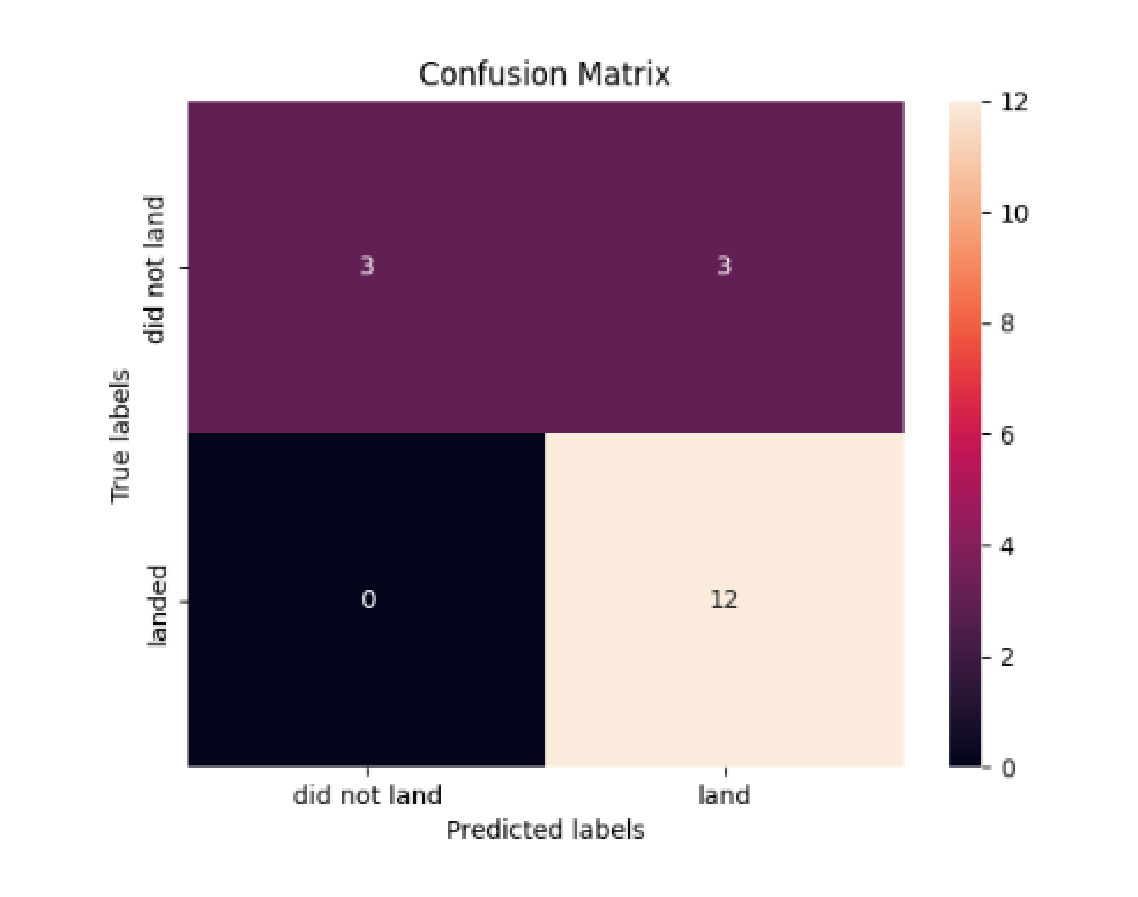
Predictive Analysis (Classification)

Classification Accuracy

- The bar chart shows the Accuracy of each of the four Machine Learning Models that were used for Prediction
- Of the 4 models, Decision Tree Classifier had the highest accuracy



Confusion Matrix



- Confusion Matrix for the Decision Tree Classifier Model
- Model correctly predicted all12 successful landings in the test data
- The model correctly predicted 3 of the 6 unsuccessful landings in the test data

Conclusions

- Classification Trees perform best when predicting whether or not Stage 1 will land
- Heavy Payloads tend to have more successful landings
- ES-L1, GEO, HEO, and SSO orbit launches have the highest success rate in the data
- Landing Success rate has been trending upwards

Thank you!