IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Derek Doane
1-18-2024

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection
  - Data Wrangling
  - EDA with SQL
  - EDA with Data Viz
  - Visualizations with Folium
  - Machine Learning
- Summary of all results
  - EDA Results
  - Interactive Analytics
  - Predictive Analytics Results

Introduction

- Project background and context
  - We are trying to understand how and why Falcon 9 landings are successful, so we can support the creation of Space Y company and achieve a low cost
- Problems you want to find answers
  - What characteristics define a successful landing and correlate to a successful landing?
  - Can we predict landing success? If so, with what model?

Section 1

Methodology

Methodology

Executive Summary

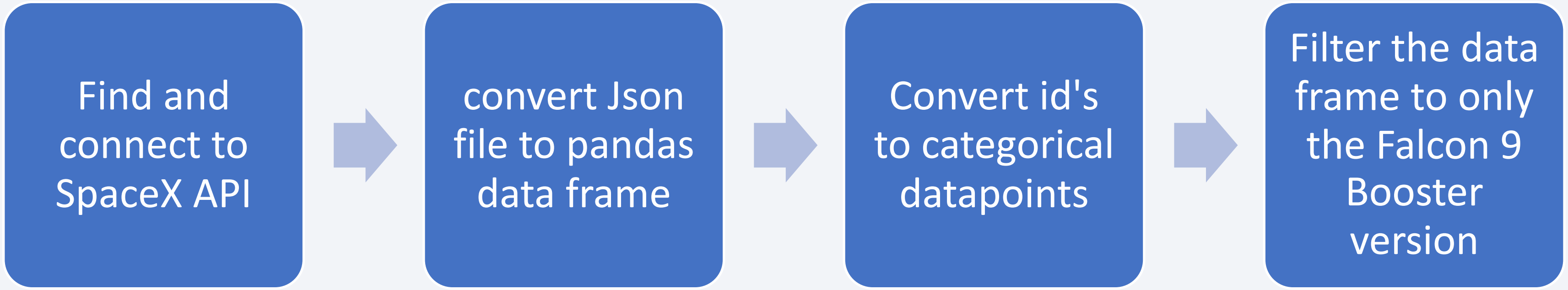
- Data collection methodology:
  - Data was collected via both web scrapping and the Space X API
- Perform data wrangling
  - Creating the Class column to define success and failure
  - Convert categorical columns to numerical columns
- Perform exploratory data analysis (EDA) using visualization and 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

- Describe how data sets were collected.
- You need to present your data collection process use key phrases and flowcharts

Data Collection – SpaceX API

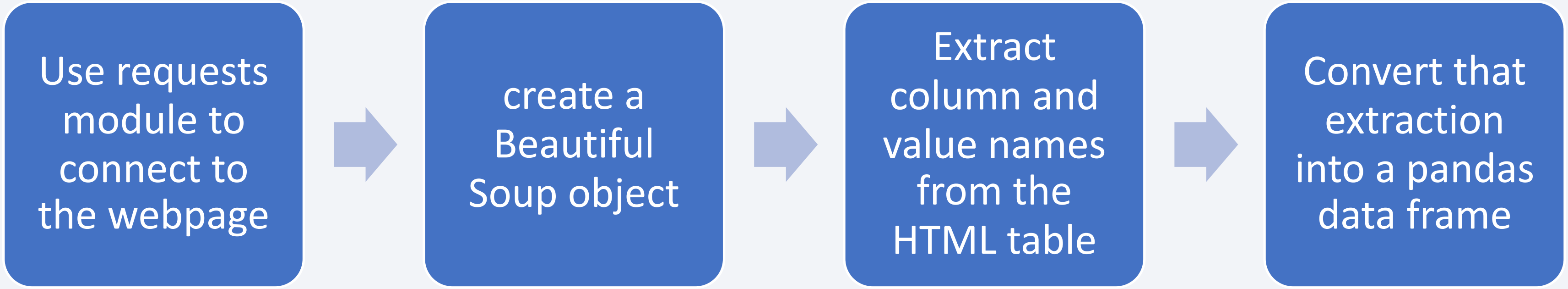
- Below is the methodology for collection the data source for SpaceX Falcon 9 launches via the SpaceX API



- [Github API Workbook]

Data Collection - Scraping

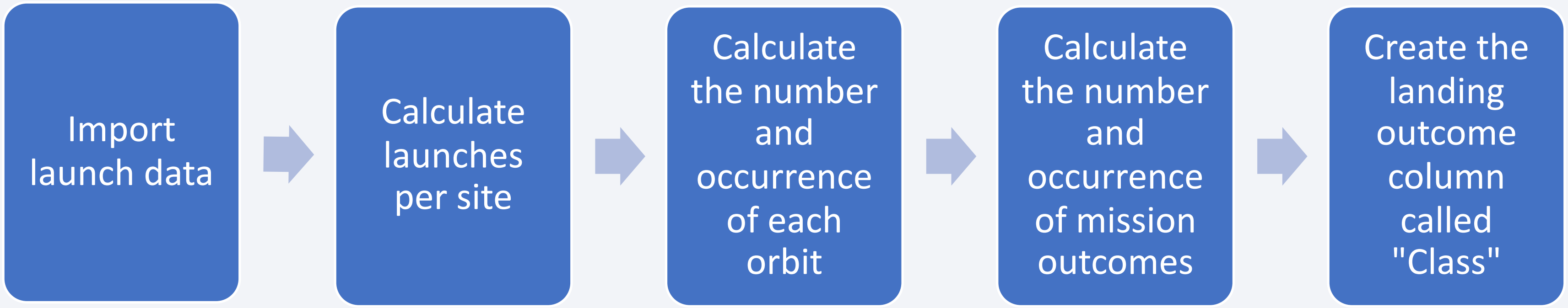
- Below is the methodology for collection the data source for SpaceX Falcon 9 launches via Wikipedia



- Github BS Workbook

Data Wrangling

- The objective of these steps was to create a final “success or failure” column to append to the data frame we created in the previous steps and to briefly explore the data available to us currently



- [Github Data Wrangling Workbook]

EDA with Data Visualization

- Summarize what charts were plotted and why you used those charts
- Scatter Plots – used to show relationships between (generally) 2 numerical variables
  - Payload Mass vs Launch Site
  - Flight Number vs Launch Site
- Bar plots – used to show numerical data compared against different categories
  - Success rate by orbit type
- Line plots – best for showing numerical data trends over time (on x axis)
  - Year vs success rate

- Github EDA Workbook

EDA with SQL

- SQL queries used for exploratory data analysis
  - Display unique launch sites
  - Display 5 launch records where launch site begins with CCA
  - Display the total payload mass carried by NASA (CRS) boosters
  - Display average payload mass carried by F9 v1.1 boosters
  - List the first successful landing via ground pad as a date
  - List the names of boosters which have been successful in drone ships with payload mass between 4k and 6k
  - List the total number of successful and failed outcomes
  - List the names of the boosters which have carried the max payload mass
  - List failed missions in 2015 with the drone ship booster type
  - Rank the count of landing outcomes between 2010-06-04 and 2017-03-20
- [Github EDA SQL Workbook]

Build an Interactive Map with Folium

- Summarize what map objects such as markers, circles, lines, etc. you created and added to a folium map
- Mark launch sites on the map
  - To give a visual representation of where these 3 sites were in the US
- Mark the successful / failed launches for each site
  - To see if there were trends based off site
- Calculate the distance between launch site to proximities
  - To understand standards for where launch sites can be located

- [Github Folium Workbook]()

Build a Dashboard with Plotly Dash

- Summarize what plots/graphs and interactions you have added to a dashboard
- The Plotly dashboard utilized many elements to make it easier to consume
  - Launch site drop-down to filter by the launch site
  - A pie-chart to show success based off the launch site selector
  - A range slider for payload
  - A scatter chart of payload vs success % based off the range selector

Predictive Analysis (Classification)

- Create columns for success and failure
- Standardize the data
- Split into training and testing data
  - Dependent variable was success rate
  - Payload mass, orbit, launch site, landing page, etc were all used an independent variables for training
- Find the best ML method and use cross validation to tune
  - Logistical regression model had a score of 83%
  - Support vector machine had a score of 83%
  - Decision tree had a score of 83%
  - K-nearest neighbor had a score of 100%

- Github ML Workbook

Results

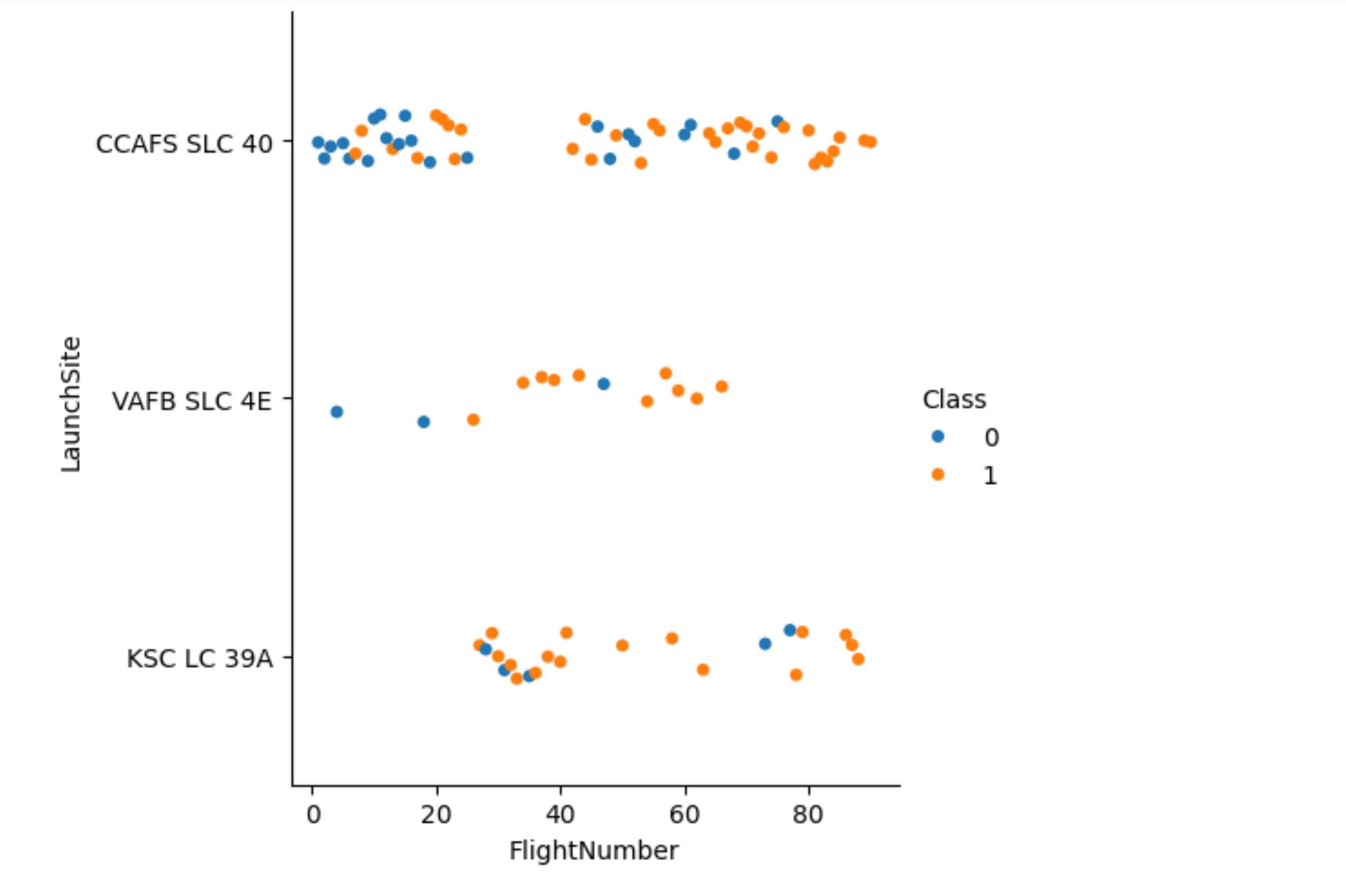
- Exploratory data analysis results
  - Except for CCAFS SLC 40 site, increase flights increases success
  - ES-L1, GEO, HEO, and SSO are the most success types of orbits (100%)
  - For the yearly trend, 2018 was a dark time but overall time has made launches more success
- Predictive analysis results
  - K-nearest neighbor using 1 neighbor, algo = auto, and p = 1 created the best results

Section 2

Insights drawn from EDA

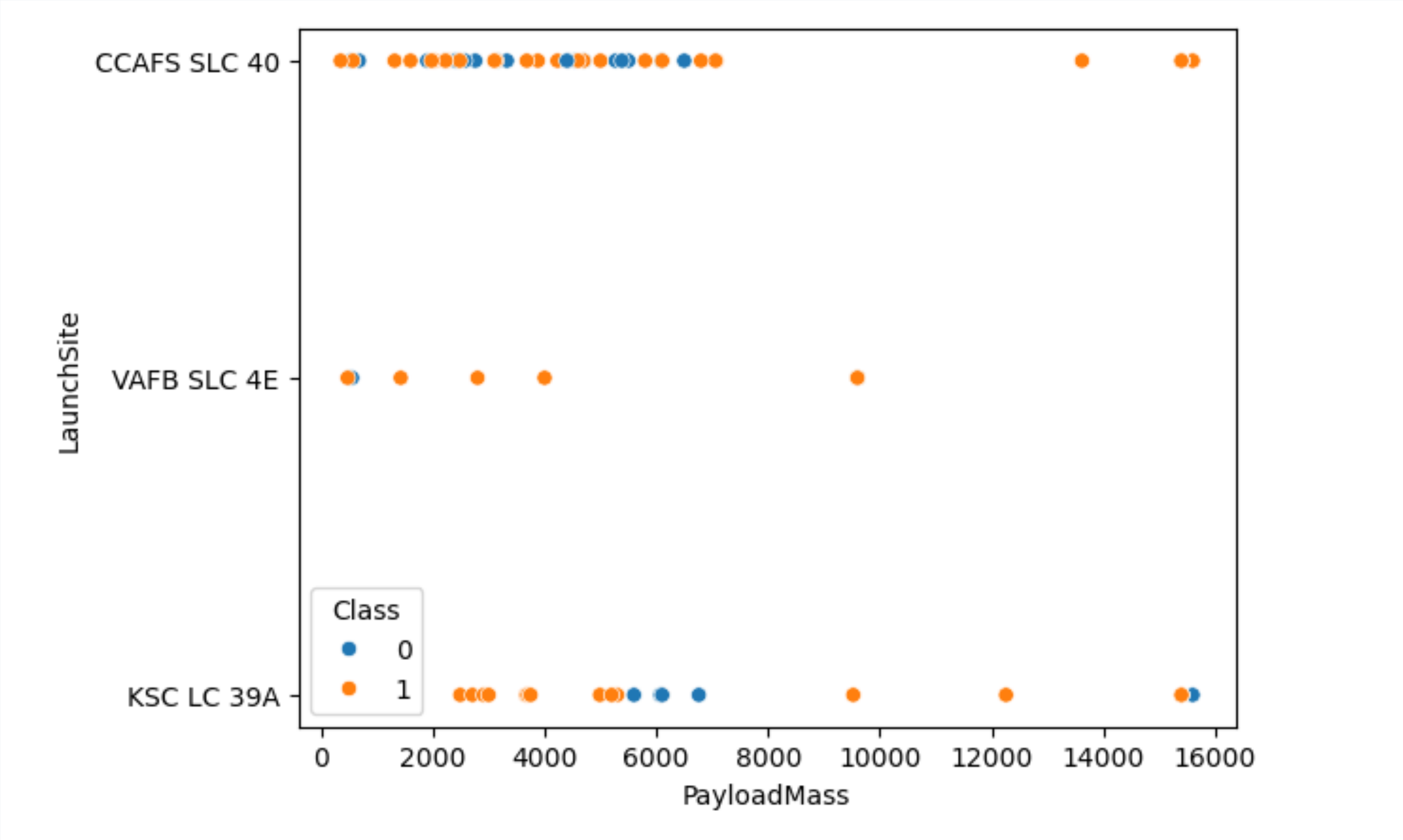
Flight Number vs. Launch Site

- Success rate appears to increase for every site as flight number increases



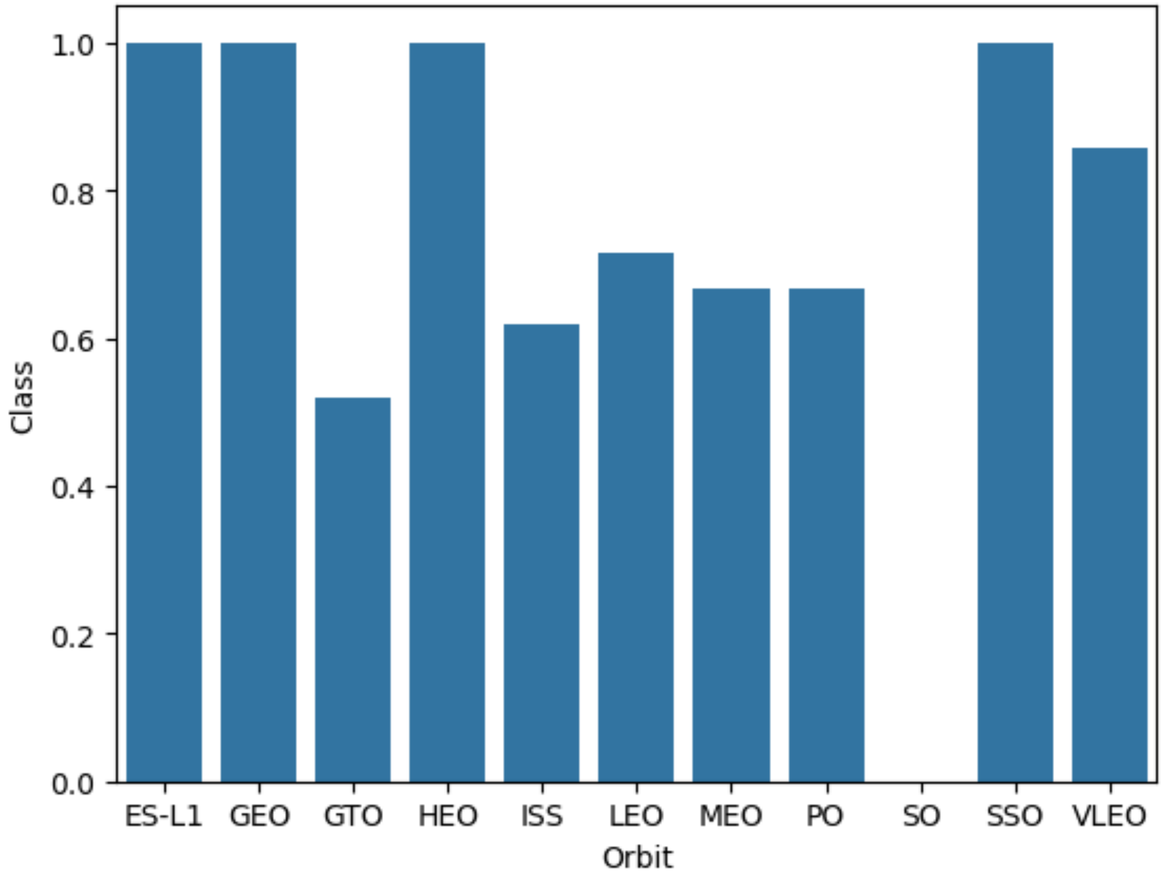
Payload vs. Launch Site

- Payloads above 8k do seem to have a higher success rate but the heaviest payload recorded should be investigated more as it appears to be a 50% success rate



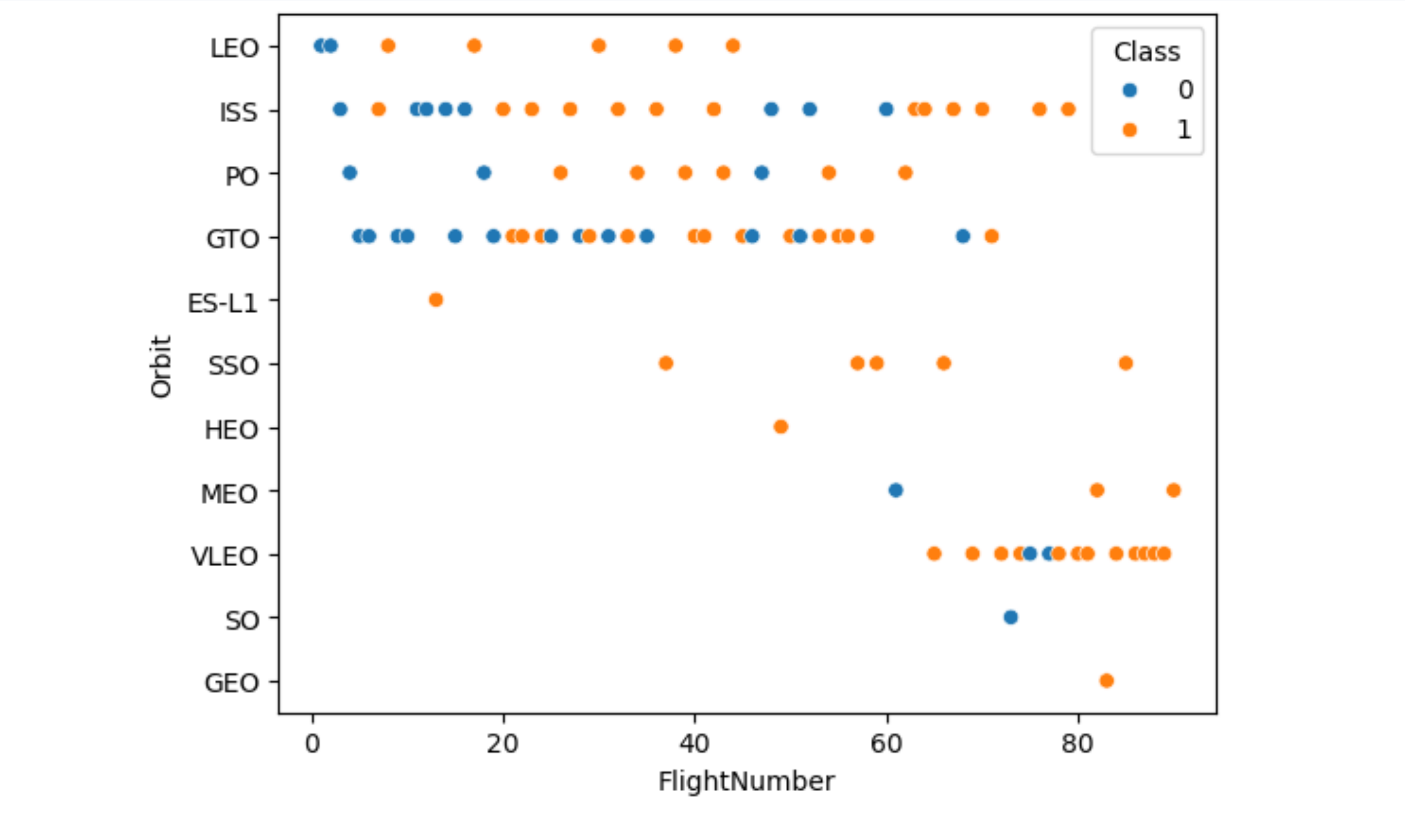
Success Rate vs. Orbit Type

- Certain orbit types have better success rates than others. GTO, ISS, MEO, and PO have the lowest success rate



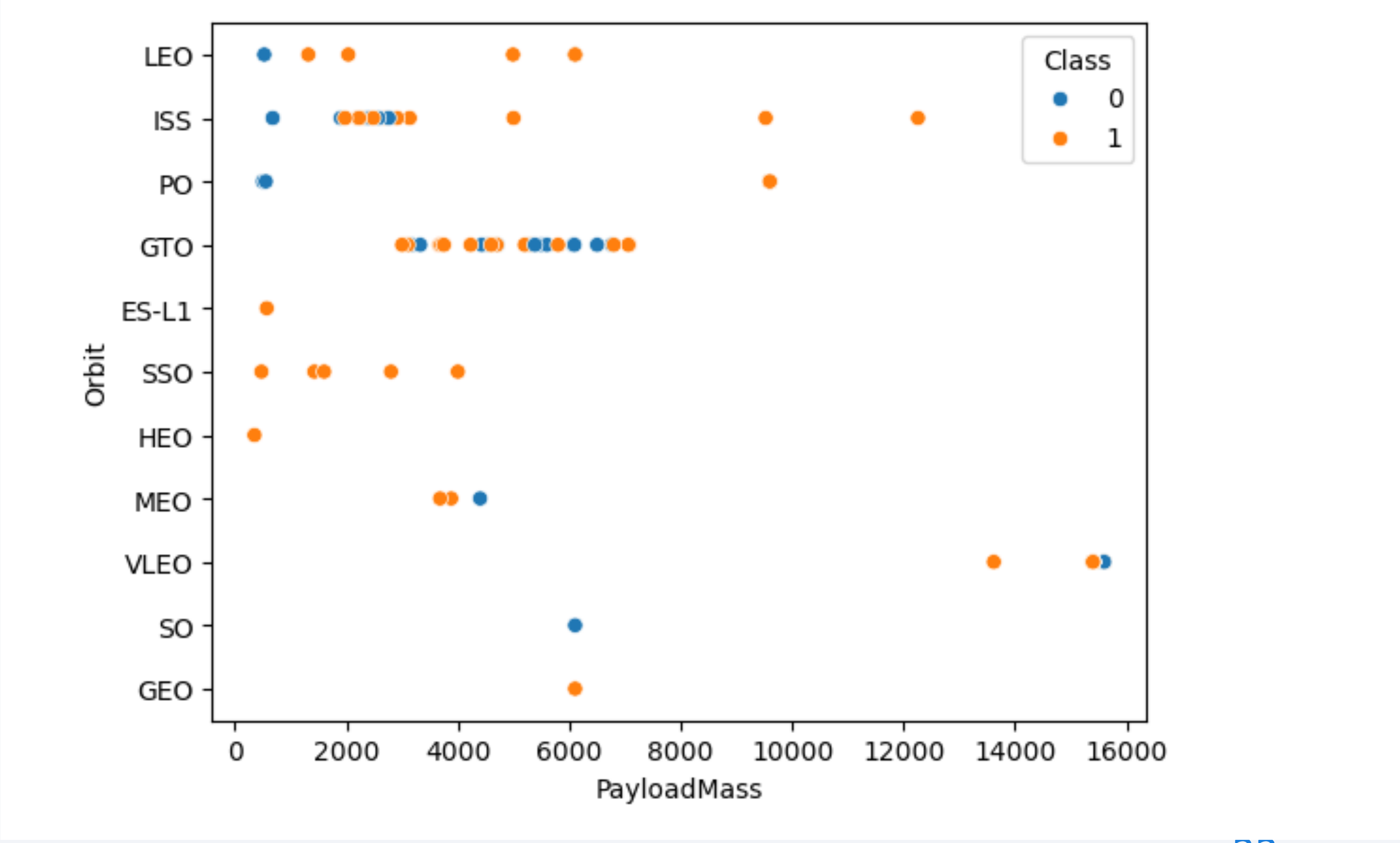
Flight Number vs. Orbit Type

- Flight number does seen to have an impact on some orbit types, but random impact on other



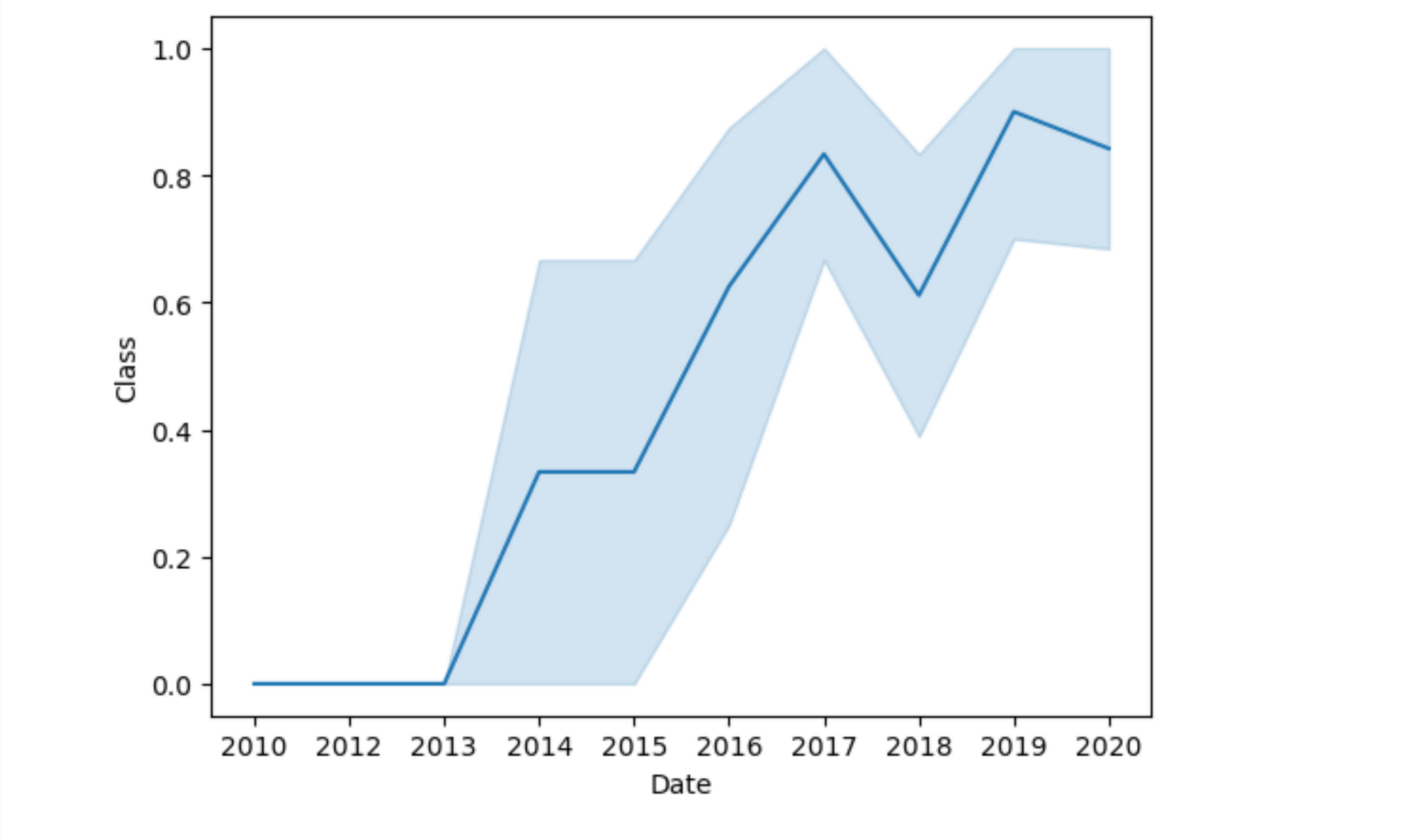
Payload vs. Orbit Type

- Heavy payloads for Polar, LEO, and ISS do lead to a higher success rate



Launch Success Yearly Trend

- Something serious happened in 2018 to impact success rate but overall, success rate has been climbing with time



All Launch Site Names

- Find the names of the unique launch sites
- Select distinct only returns unique rows for a specific column

```
In [9]: %sql select distinct(Launch_Site) from SPACEXTABLE

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

Out[9]:

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

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`
- Like is used on Launch_Site with special character % to return these results

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

In [10]:
```
%sql select * from SPACEXTABLE where Launch_Site like 'CCA%' LIMIT 5
```

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

Out[10]:

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

Total Payload Mass

- Calculate the total payload carried by boosters from NASA
- Sum against payload mass with the Customer filter of NASA

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

In [13]:
```
%sql select sum(PAYLOAD_MASS__KG_) as total_payload_mass from SPACEXTABLE where Customer = 'NASA (CRS)'
```

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

Out[13]:

total_payload_mass
45596

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1
- Avg can be used after filtering to F9 on payload mass

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

In [14]:
```
%sql select avg(PAYLOAD_MASS__KG_) as avg_payload_mass from SPACEXTABLE where Booster_Version = 'F9 v1.1'
```

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

Out[14]:

avg_payload_mass
2928.4

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad
- Min on date where landing outcome is the ground pad and it's successful

No attempt

[17]:

```
%sql select min(Date) as first_ground_pad_landing from SPACEXTABLE where Landing_Outcome = 'Success (ground pad)'
```

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

[17]:

first_ground_pad_landing
2015-12-22

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

- I used the between statement for the payload mass to be between 4k and 6k

List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

[18]:
```
%sql select Booster_Version from SPACEXTABLE where Landing_Outcome = 'Success (drone ship)' and PAYLOAD_MASS__KG_
```

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

[18]:

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

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes
- Counting rows grouped by the mission outcome

List the total number of successful and failure mission outcomes

[22]:
```
%sql select Mission_Outcome, count(*) from SPACEXTABLE group by 1 order by 1
```

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

t[22]:

Mission_Outcome	count(*)
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass
- I used a subquery to capture the max payload and then used that in my where statement to return boosters having used the max payload

List the names of the booster_versions which have carried the maximum payload mass. Use a subquery

In [23]:
```
%sql select Booster_Version from SPACEXTABLE where PAYLOAD_MASS__KG_ = (select max(PAYLOAD_MASS__KG_) from SPACEX
```

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

Out[23]:

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

- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

- Using the documentation provided to return years, then filtering appropriately

```
[26]: %sql select substr(Date, 6,2) as month, Landing_Outcome, Booster_Version, Launch_Site from SPACEXTABLE where subs

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

[26]:

month	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

- 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

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.

```
31]: %sql select Landing_Outcome, Count(*) from SPACEXTABLE where Date between '2010-06-04' and '2017-03-20' group by
```

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

31]:

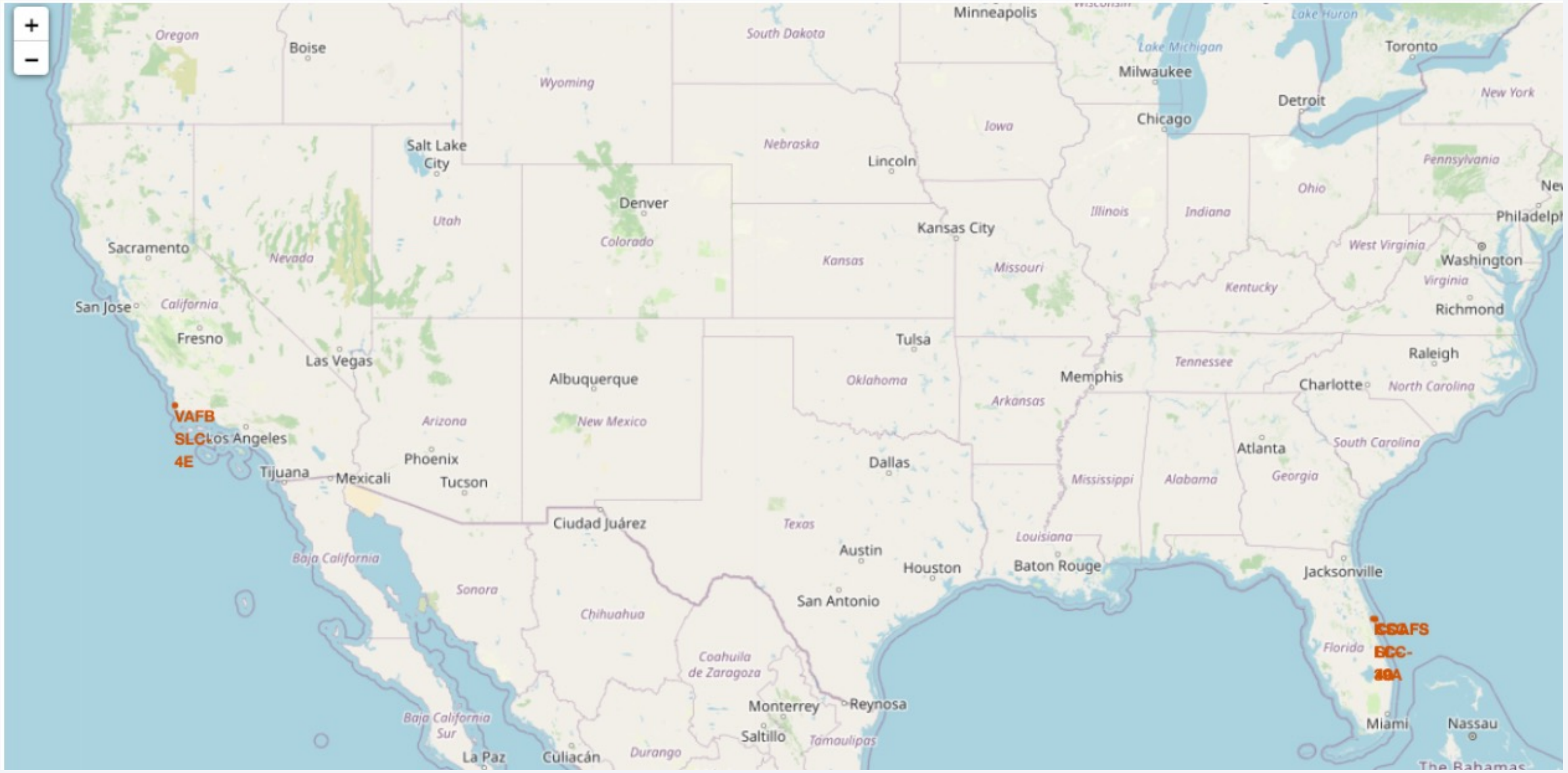
Landing_Outcome	Count(*)
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

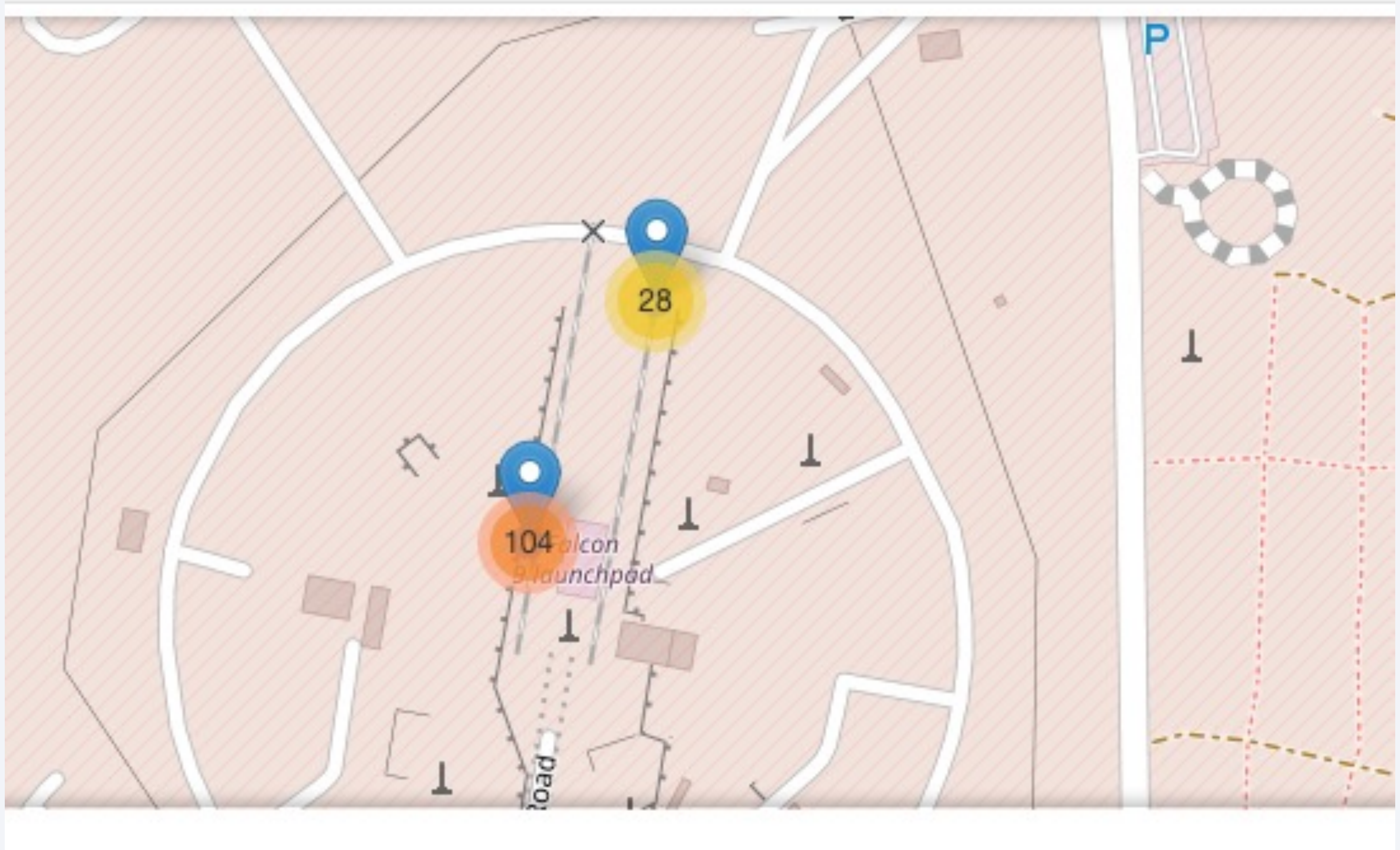
Space X Launch Sites

- Marking the launch sites on the map

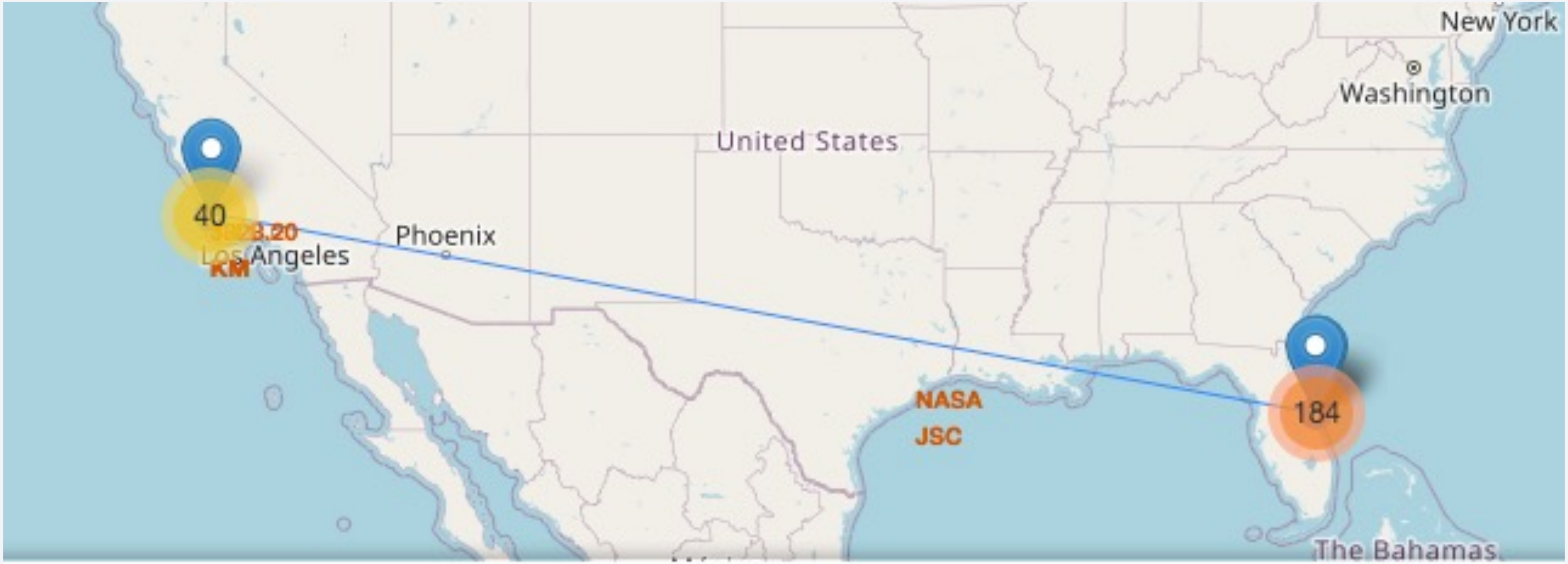


Marking Launch Success and Failure

- Marking success and failure by launch site using the Class field



Distance Between Launch Sites and Proximities



Line between a launch site to the selected coastline point

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>

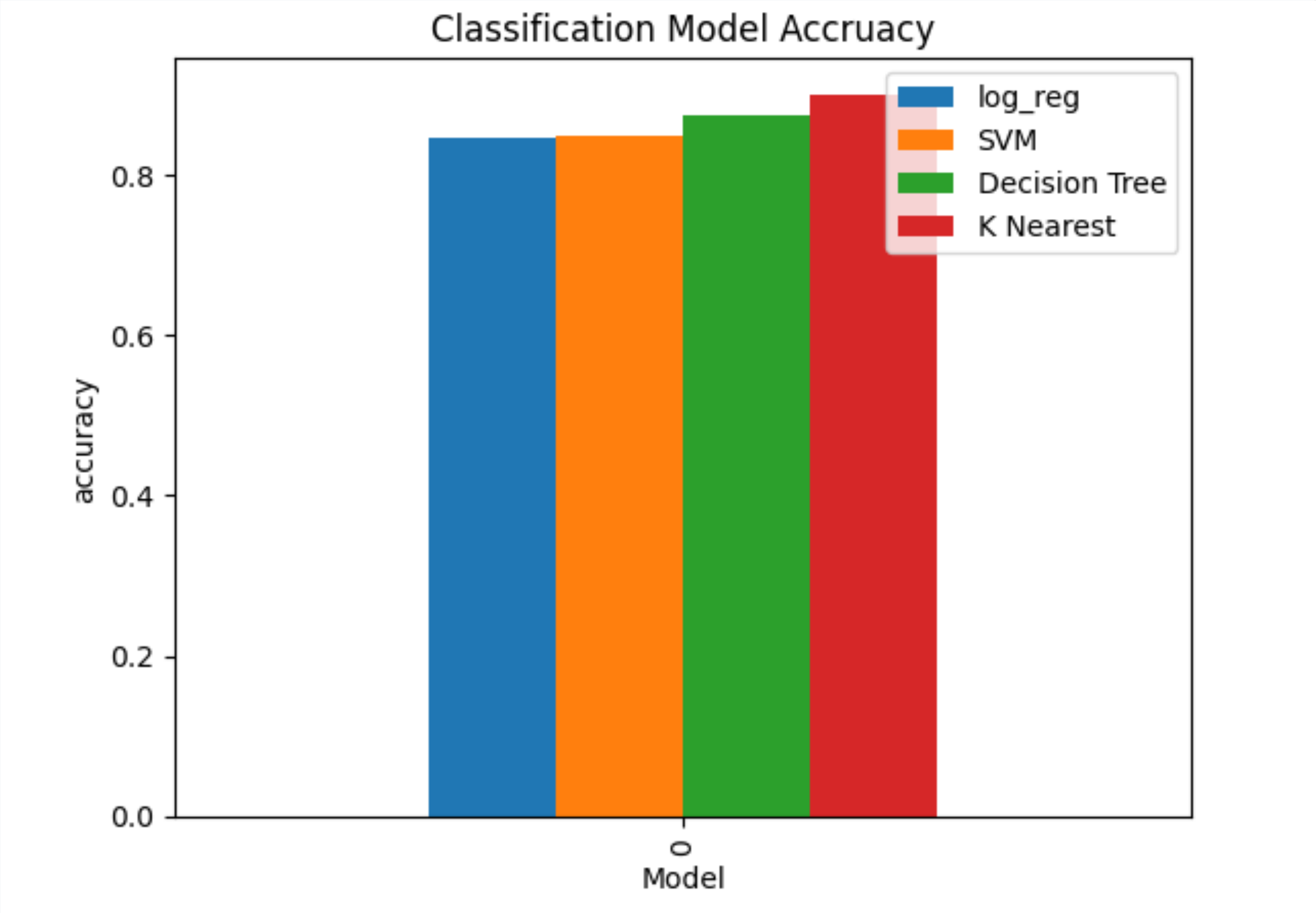
- Replace <Dashboard screenshot 3> title with an appropriate title
- Show screenshots of Payload vs. Launch Outcome scatter plot for all sites, with different payload selected in the range slider
- Explain the important elements and findings on the screenshot, such as which payload range or booster version have the largest success rate, etc.

Section 5

Predictive Analysis (Classification)

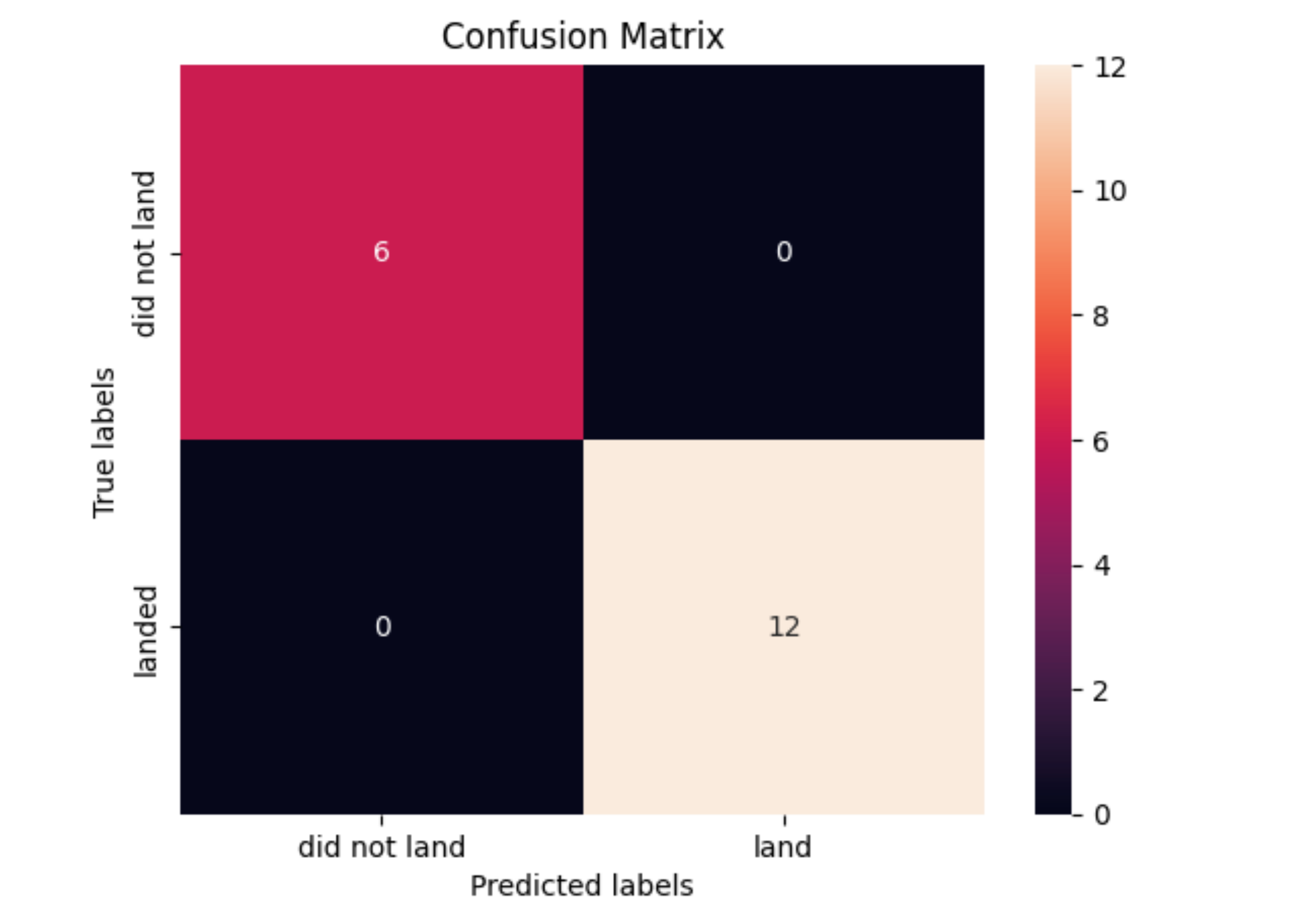
Classification Accuracy

- K nearest had the highest accuracy



Confusion Matrix

- There were no wrong predictions



Conclusions

- Many factors contribute to launch success including: payload mass, orbit, site, etc
- GEO, HEO, SSO, and ES-L1 had some of the highest success rates and should be heavily leveraged
- Depending on the type of orbit, payload mass can play a big factor in success
- If we want to classify upcoming launches, we would use the nearest neighbor classification model to help understand success or failure

Appendix

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

Thank you!