IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Tyler Trzecki
December 28, 2021

Outline

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

Executive Summary

SpaceX is a revolutionary company, disrupting the rocket and satellite industry through successful landings of first stage boosters. SpaceX was able to achieve this by optimizing parameters such as launch site, payload mass, and orbit for their missions to maximize primary mission success rate (payload reaches desired orbit) and secondary mission success rate (land first stage booster). SpaceX was able to achieve a 66% success rate for their secondary mission.

Introduction

SpaceX is the first rocket company to land a first stage booster that reached orbit, revolutionizing the economics and engineering of the rocket industry. Along their journey they encountered many successes and many failures. The goal of this study is to identify the reasons for SpaceX's successes and failures based upon parameters such as payload mass, orbit, launch site.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - SpaceX mission data was collected from Wikipedia and was organized into a dataframe
- Perform data wrangling
  - Data was filtered to specifically look at the Falcon 9 booster
  - Where payload mass data was unavailable the mean value was used in its place
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - The Scikit Learn model selection GridSearchCV was used to optimize and find the best parameters for each classification model

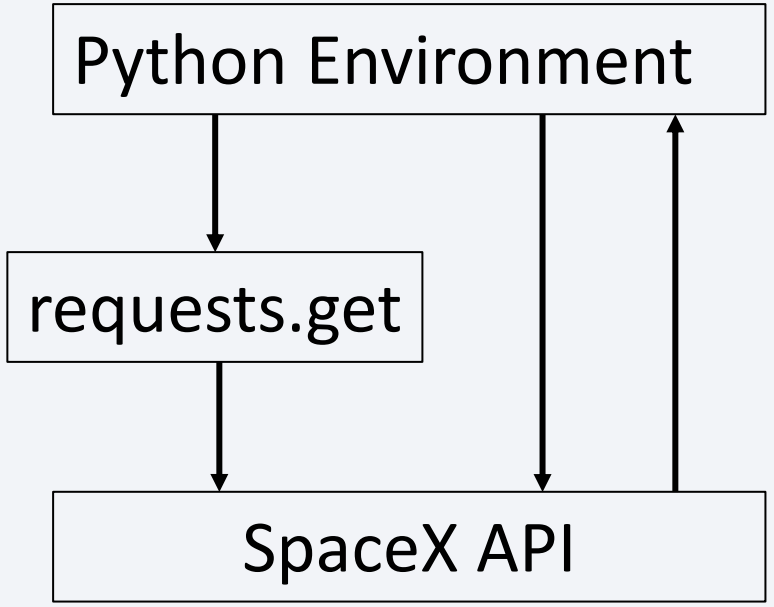
Data Collection

- The data sets were collected by performing get requests on SpaceX's rocket launch data API (https://api.spacexdata.com/v4/launches/past)
- Additional data was also scraped from the Wikipedia's URL to collect a list of Falcon 9 and Falcon Heavy Launches (https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922)
- A response was requests was sent to each URL
  - To organize data from Wikipedia, the BeautifulSoup library was used to parse the HTML
- SpaceX's API organized the data in a .json file which could be read directly into a pandas dataframe

Data Collection – SpaceX API

- Requests.get was used to connect and store data from the SpaceX API
- Pandas was able to store the data organized as a JSON file into a dataframe with pandas.read_json()

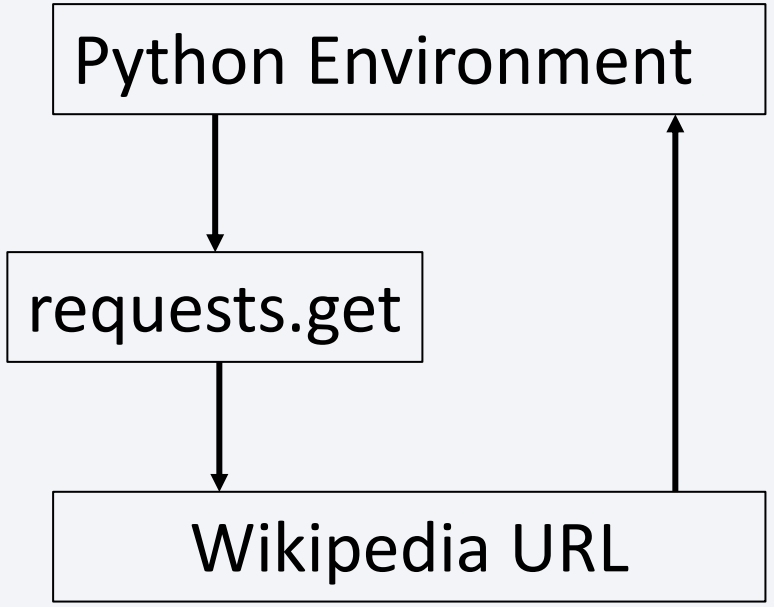
- https://github.com/tytrzecki/IBMDataScience/blob/main/IBM%20Capstone.ipynb



Data Collection - Scraping

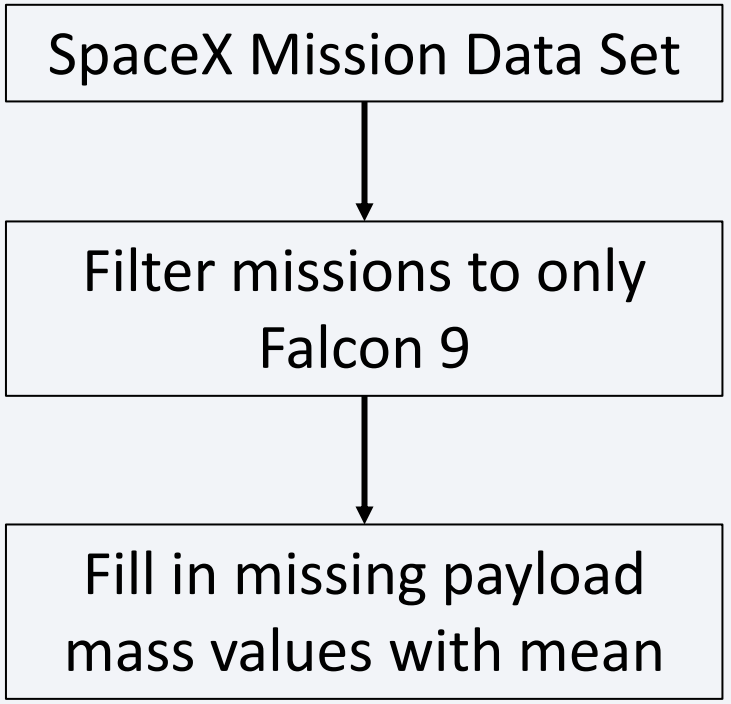
- Requests.get was used to collect data from the Wikipedia URL
- The data was stored in HTML and BeautifulSoup was used to parse the HTML

- https://github.com/tytrzecki/IBMDataScience/blob/main/IBM%20Capstone%20-%20Web%20Scraping.ipynb



Data Wrangling

- The data wrangling process collected and organized the relevant parameters and filled in missing data points where necessary
- The data was filtered to only review Falcon 9 rockets
- Where payload mass data was unavailable the mean value of the provided payload mass within the dataset was used
- https://github.com/tytrzecki/IBMDataScience/blob/main/IBM%20Capstone.ipynb



EDA with Data Visualization

- Flight Number vs Payload Mass was plotted to see how SpaceX's missions evolved as the company gained more experience and how that contributed to their overall success rate
  - As their experience grew, they took on larger missions and ultimately had more success
- Flight number vs Launch site illustrated their success rate at each launch site over time
  - Approximately 25 of their first launches took place at CCAF SLC 40 where their secondary mission success rate was 36%

- https://github.com/tytrzecki/IBMDataScience/blob/main/EDA%20with%20Data%20Visualization.ipynb

EDA with SQL

- The launch sites were extracted from SQL database
  - 5 launch sites were extracted that contained 'CCA'
- Calculated the payload mass carried by the boosters launching from NASA (CRS)
- Calculated the average payload mass carried by booster version F9 v1.1
- Found the date of the first successful landing
- Determined the boosters that successfully landed on a drone ship and had a payload mass between 4000 and 6000 lbs
- Found the total number of successful missions
- Found the max payload for each booster version
- Found the failed landings from the year 2015
- Rank landing outcomes between 2010 and 2017
- https://github.com/tytrzecki/IBMDataScience/blob/main/EDA%20with%20SQL.ipynb

Build an Interactive Map with Folium

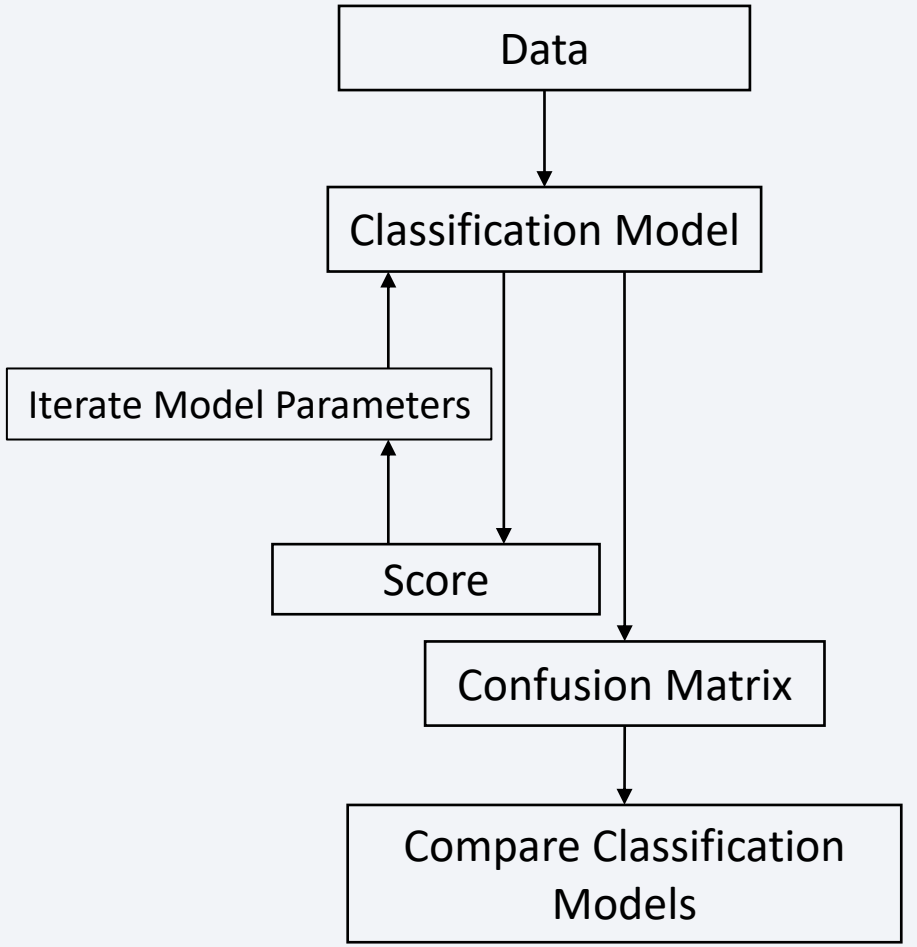
- An interactive map was created with Folium
- Markers, circles and lines were used to identify key features on the map
  - Markers were used to identify the launch sites on the map
    - Marker clusters were used to identify locations that had multiple launches
  - Circles were used to identify a radius around the launch site that could be used to determine nearby transportation methods such as railroads, highways, etc.
  - Lines were used to calculate the distance between objects
    - Specifically, a line was used to determine the distance between the launch site and the nearest coast line
- https://github.com/tytrzecki/IBMDataScience/blob/main/Interactive%20Visual%20Analytics%20with%20Folium%20lab.ipynb

Predictive Analysis (Classification)

- Data was collected via a request command on the SpaceX data API
- Data was organized under the following features:
    - Flight Number, date, booster, payload mass, orbit, launch site, outcome, flights, grid fins, reused, legs, landing pad, and block
- Data wrangling was performed on Payload Mass parameter where there were 5 missing values. The average payload mass was input where data was not available.
- A column labeled Class was created to identify if the mission was successful and is used to train and test our models on
- The parameters listed above were used as features to build the model and were transformed and scaled using the preprocessing library and standard scaler method
- https://github.com/tytrzecki/IBMDataScience/blob/main/IBM%20Capstone.ipynb

Predictive Analysis (Classification)

- A Logistic Regression, Support Vector Machine, Decision Tree, and K-Nearest Neighbors models were created
  - Each model had its parameters iterated over in order to maximize the accuracy of the model
- These models were evaluated using the score method and a multiclassification confusion matrix
- The models were evaluated individually as each parameter in the model was iterated through to determine the parameters that maximized accuracy
- Then each model was compared based on its accuracy measured using the score method and a confusion matrix
- https://github.com/tytrzecki/IBMDataScience/blob/main/Machine%20Learning%20Prediction.ipynb

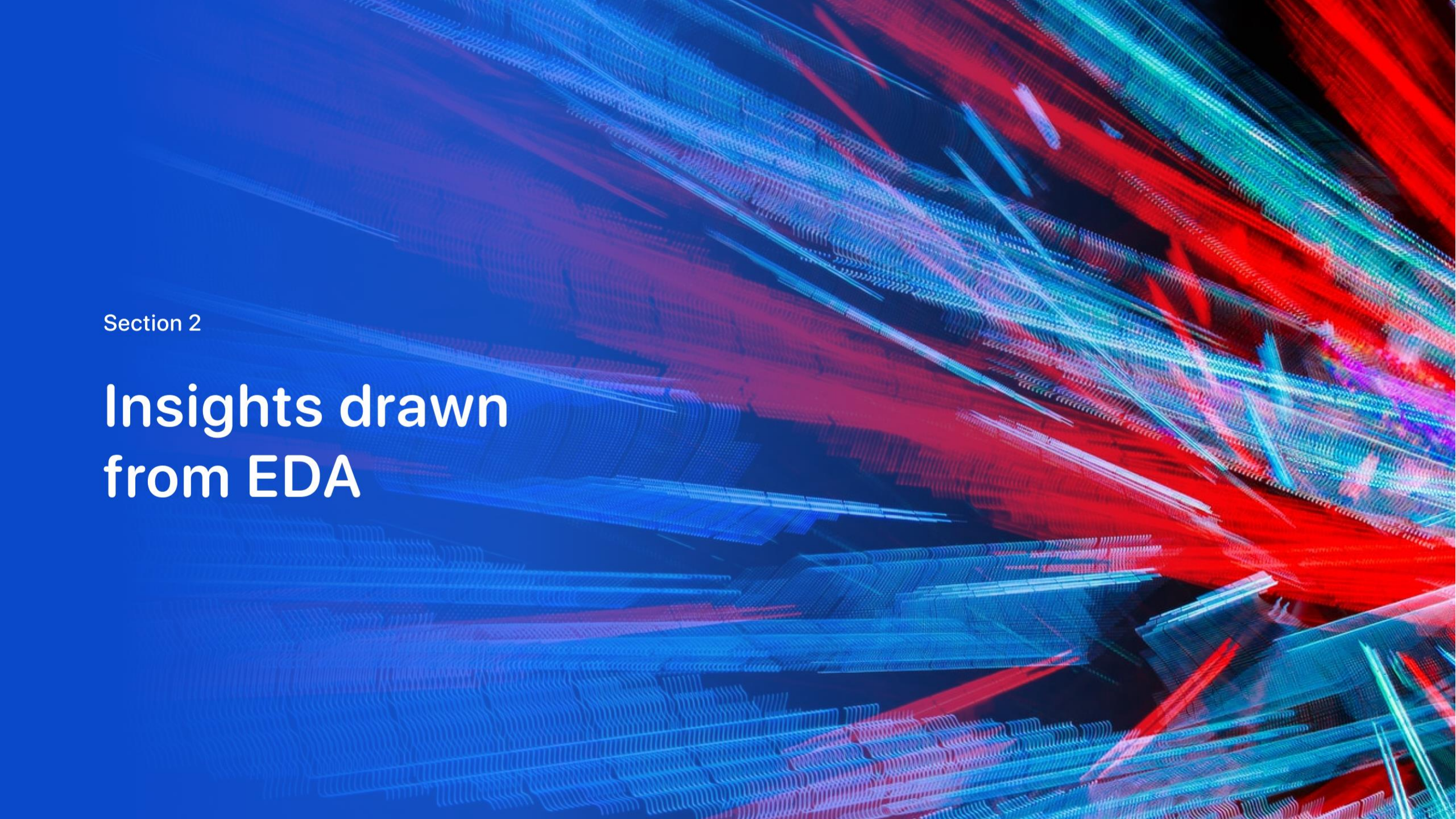


Results

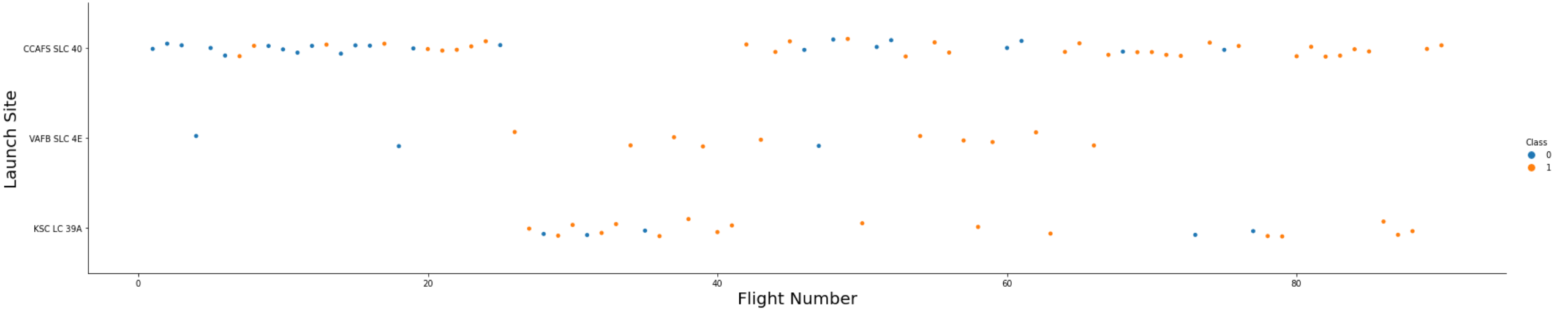
- The following slides will highlight:
  - Exploratory data analysis results
  - Interactive analytics demo in screenshots
  - Predictive analysis results

Section 2

Insights drawn from EDA

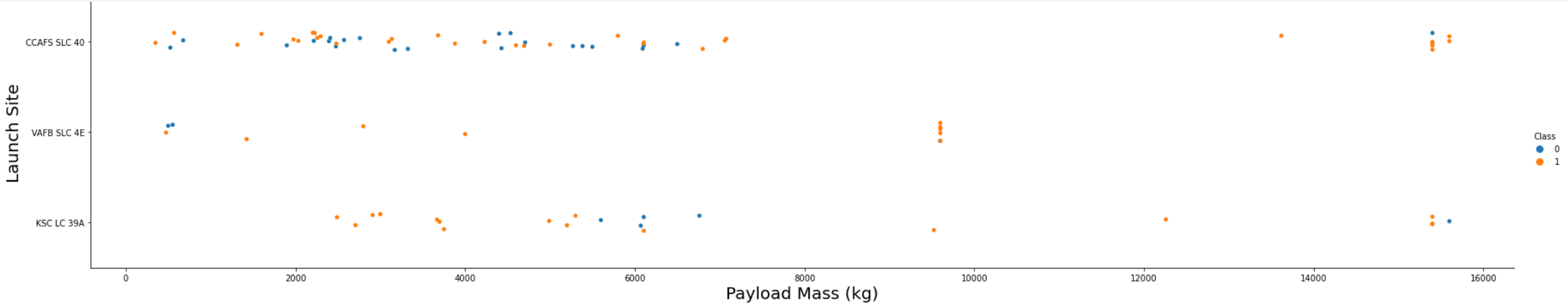


Flight Number vs. Launch Site



Many of the earlier launches were performed at Cape Canaveral Airforce Base. When SpaceX was first beginning they did not have much experience or knowledge of how to land a rocket evidenced by the first 6 attempts failing, and among the first 25 attempts a landing success rate of 36%. As they gained more experience and knowledge their successful landings increased at each launch site.

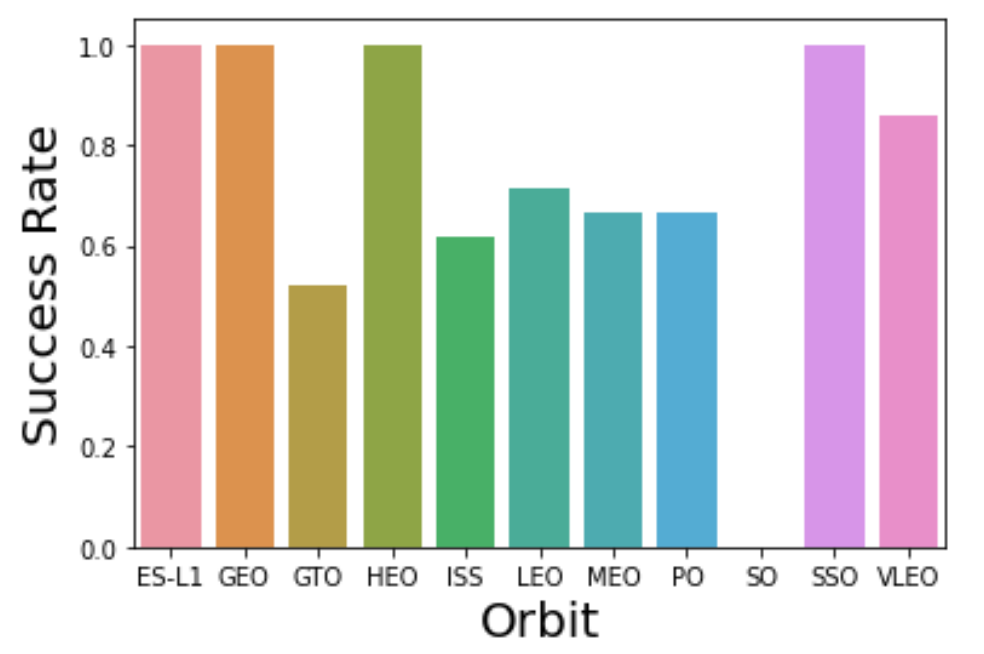
Payload vs. Launch Site



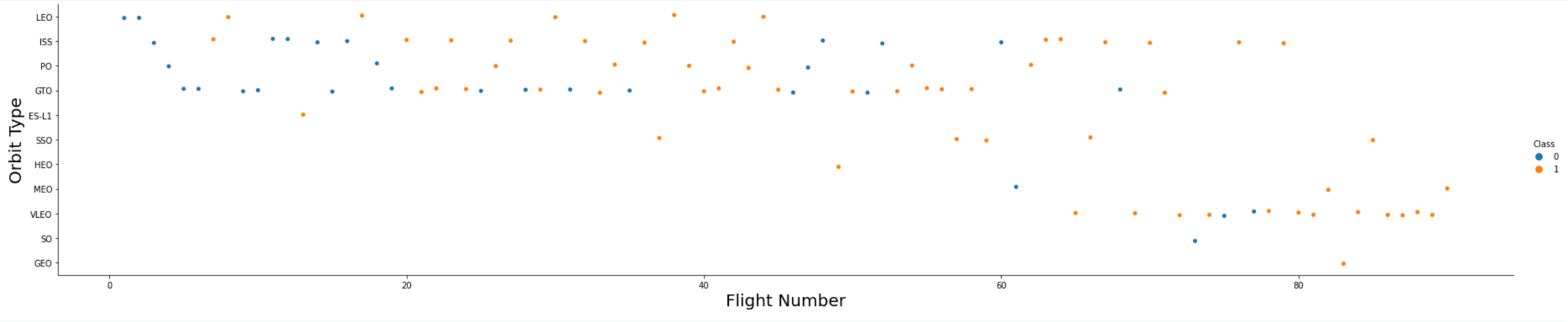
Most missions had a payload less than 10000 kg. Vandenberg Airforce Base carried out very specific missions with set payloads and a high success rate. Kennedy Space Center carried out a number of missions with an average payload mass and a good success rate. CCAFS carried out many missions with a scatter of payload masses, and success and failures.

Success Rate vs. Orbit Type

There are four orbits in which the mission success rate is a perfect 100%: ES-L1,GEO, HEO, and SSO. The SO orbit has the lowest success rate at 0%. All other orbits have a success rate of 50% or above.

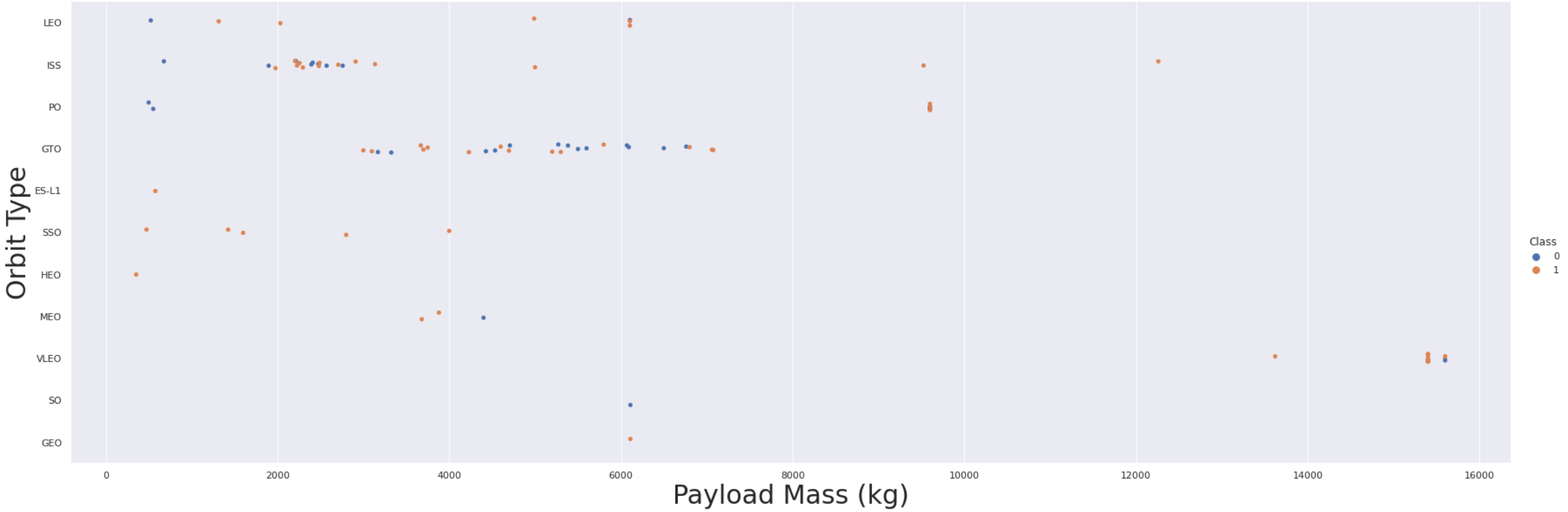


Flight Number vs. Orbit Type



The earlier missions were flown to LEO, ISS, PO, and GTO with a low landing success rate. As their experience grew their mission capability and probability of success grew.

Payload vs. Orbit Type

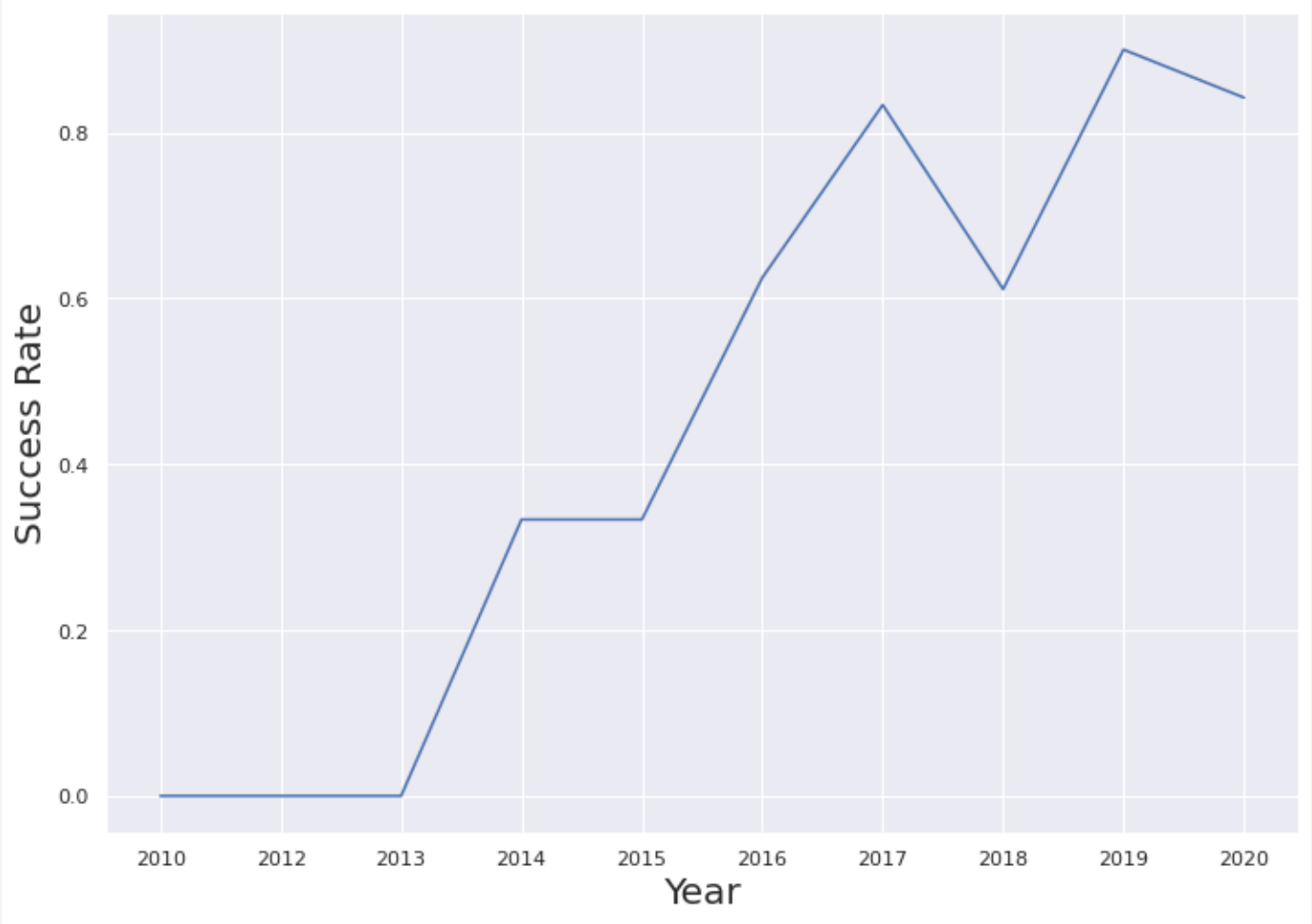


With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.

However for GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both there.

Launch Success Yearly Trend

It is observed that the success rate since 2013 kept increasing until 2020, with a small dip in 2018.



All Launch Site Names

Task 1

Display the names of the unique launch sites in the space mission

In [11]:
```
%%sql

Select Distinct LAUNCH_SITE From SPACEXTBL
```

```
 * ibm_db_sa://vqp18483:***@54a2f15b-5c0f-46df-8954-7e38e612c2bd.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32733/bludb
Done.
```

Out[11]:

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

- There are three major launch sites in the US and they are Cape Canaveral Airforce Base (CCAFS LC-40 & CCAFS SLC-40), Kennedy Space Center (KSC LC-39A), and Vandenberg Airforce Base (VAFB SLC-4E).

Launch Site Names Begin with 'CCA'

Task 2

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

In [26]:
```
%%sql

Select * From SPACEXTBL
WHERE LAUNCH_SITE LIKE 'CCA%'
LIMIT 5;
```

```
 * ibm_db_sa://vqp18483:***@54a2f15b-5c0f-46df-8954-7e38e612c2bd.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32733/bludb
Done.
```

Out[26]:

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

- Above, 5 launches with launch sites beginning with string ‘CCA’

Total Payload Mass

- The total mass of the payloads launched by NASA (CRS) is 45,596 lbs

Task 3

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

In [28]:
```
%%sql

SELECT SUM(PAYLOAD_MASS__KG_)
From SPACEXTBL
WHERE CUSTOMER = 'NASA (CRS)';
```

```
 * ibm_db_sa://vqp18483:***@54a2f15b-5c0f-46df-8954-7e38e612c2bd.c1ogj3s
Done.
```

Out[28]:

1
45596

Average Payload Mass by F9 v1.1

- The average payload mass carried by the F9 v1.1 booster is 2,534 lbs

Task 4

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

In [31]:
```
%%sql

SELECT AVG(PAYLOAD_MASS__KG_)
From SPACEXTBL
WHERE BOOSTER_VERSION like 'F9 v1.1%';
```

```
 * ibm_db_sa://vqp18483:***@54a2f15b-5c0f-46df-8954-7e38e612
Done.
```

Out[31]:

1
2534

First Successful Ground Landing Date

- The date of the first successful landing was July 22nd, 2018

Task 5

List the date when the first successful landing outcome in ground pad was acheived.

Hint:Use min function

In [32]:
```
%%sql

SELECT MIN(DATE)
From SPACEXTBL
WHERE LANDING__OUTCOME = 'Success';
```

```
 * ibm_db_sa://vqp18483:***@54a2f15b-5c0f-46df-8954-7e38e612c2bd.c1ogj3sd0t
Done.
```

Out[32]:

1
2018-07-22

Successful Drone Ship Landing with Payload between 4000 and 6000

Task 6

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

In [33]:
```
%%sql

SELECT *
From SPACEXTBL
WHERE LANDING__OUTCOME = 'Success (drone ship)'
    AND PAYLOAD_MASS__KG_ > 4000
    AND PAYLOAD_MASS__KG_ < 6000;
```

```
 * ibm_db_sa://vqp18483:***@54a2f15b-5c0f-46df-8954-7e38e612c2bd.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32733/bludb
Done.
```

Out[33]:

DATE	time__utc_	booster_version	launch_site	payload	payload_mass__kg_	orbit	customer	mission_outcome	landing__outcome
2016-05-06	05:21:00	F9 FT B1022	CCAFS LC-40	JCSAT-14	4696	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
2016-08-14	05:26:00	F9 FT B1026	CCAFS LC-40	JCSAT-16	4600	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
2017-03-30	22:27:00	F9 FT B1021.2	KSC LC-39A	SES-10	5300	GTO	SES	Success	Success (drone ship)
2017-10-11	22:53:00	F9 FT B1031.2	KSC LC-39A	SES-11 / EchoStar 105	5200	GTO	SES EchoStar	Success	Success (drone ship)

- Listed above are the names of the boosters with successful landings on drone ships with payloads between 4000 and 6000 pounds.

Total Number of Successful and Failure Mission Outcomes

- The total number of successful and failure mission outcomes is 61.

Task 7

List the total number of successful and failure mission outcomes

In [35]:
```
%%sql

SELECT COUNT(*)
From SPACEXTBL
WHERE LANDING__OUTCOME LIKE 'Success%';
```

```
 * ibm_db_sa://vqp18483:***@54a2f15b-5c0f-46df-8954-7e38e612c2
Done.
```

Out[35]:

1
61

Boosters Carried Maximum Payload

Shown is a list of the booster versions that have carried the maximum payload.

Task 8

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

In [30]:
```
%%sql

select BOOSTER_VERSION, PAYLOAD_MASS__KG_ from SPACEXTBL
where PAYLOAD_MASS__KG_=(select max(PAYLOAD_MASS__KG_) from SPACEXTBL);
```

```
 * ibm_db_sa://vqp18483:***@54a2f15b-5c0f-46df-8954-7e38e612c2bd.c1ogj3sd0tgtu0lqde00.databa
Done.
```

Out[30]:

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

- There were two booster versions that failed to land on drone ships in 2015. Shown is the booster version name and their launch site.

Task 9

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

In [38]:
```
%%sql

SELECT BOOSTER_VERSION, LAUNCH_SITE
From SPACEXTBL
WHERE LANDING__OUTCOME LIKE 'Failure%'
    AND DATE LIKE '2015%';
```

```
 * ibm_db_sa://vqp18483:***@54a2f15b-5c0f-46df-8954-7e38e612c2bd.c1ogj3sd0tgtu0lqde00.databases.a
Done.
```

Out[38]:

booster_version	launch_site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

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

- Shown is a list of landing outcomes between June 4th, 2010 and March 20th, 2017 in descending order.

In [31]:
```
%%sql

SELECT LANDING__OUTCOME FROM SPACEXTBL
WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20' ORDER BY DATE DESC;
```

```
 * ibm_db_sa://vqp18483:***@54a2f15b-5c0f-46df-8954-7e38e612c2bd.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32733/bludb
Done.
```

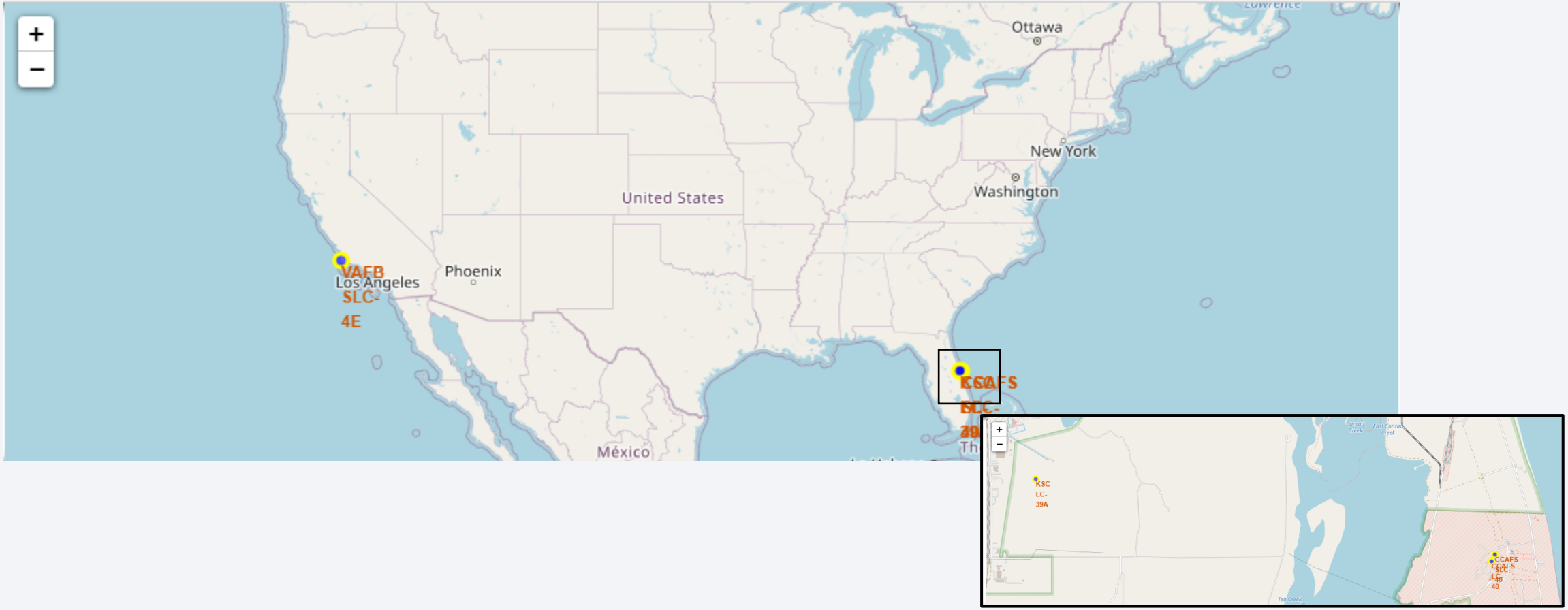
Out[31]:

landing__outcome
No attempt
Success (ground pad)
Success (drone ship)
Success (drone ship)
Success (ground pad)
Failure (drone ship)
Success (drone ship)
Success (drone ship)
Success (drone ship)
Failure (drone ship)
Failure (drone ship)
Success (ground pad)
Precluded (drone ship)
No attempt
Failure (drone ship)
No attempt
Controlled (ocean)
Failure (drone ship)
Uncontrolled (ocean)
No attempt
No attempt
Controlled (ocean)
Controlled (ocean)
No attempt
No attempt
Uncontrolled (ocean)
No attempt
No attempt
No attempt
Failure (parachute)
Failure (parachute)

Section 4

Launch Sites Proximities Analysis

Launch Sites



- In the map above the launch sites are displayed with markers and labels for every mission of the F9 v1.1 booster.

Launch Site Clusters and Launch Outcome

- This map displays all of the launches at their respective launch sites as clusters that can be expanded and minimized on click. It also shows the launch outcome displaying red if the launch failed or green if it was successful.

Distance From Coast Line to Launch Site

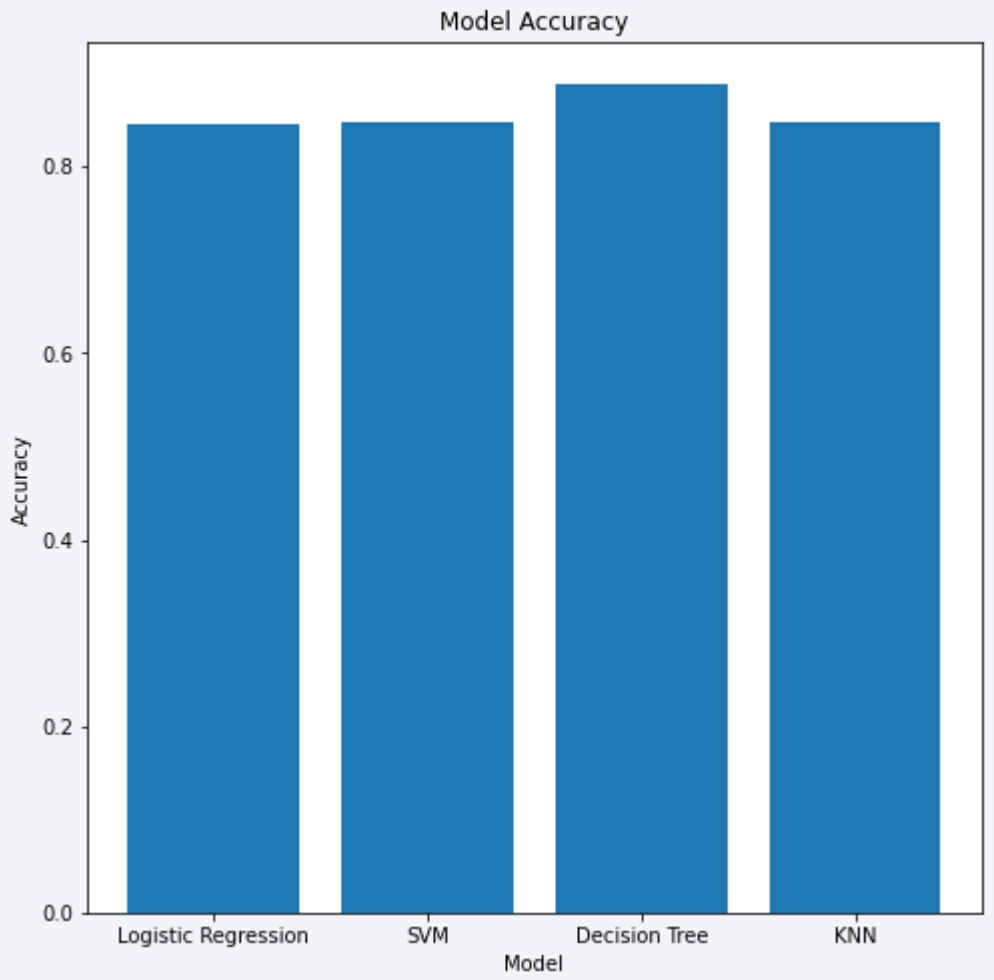
- The Map shows the distance from Vandenberg Airforce Base to the coast line illustrated by a line and shown to be approximately 1.34 kilometers. Additionally, the upper right hand corner of the map displays the latitude and longitude based on where the mouse is on the map.

Section 6

Predictive Analysis (Classification)

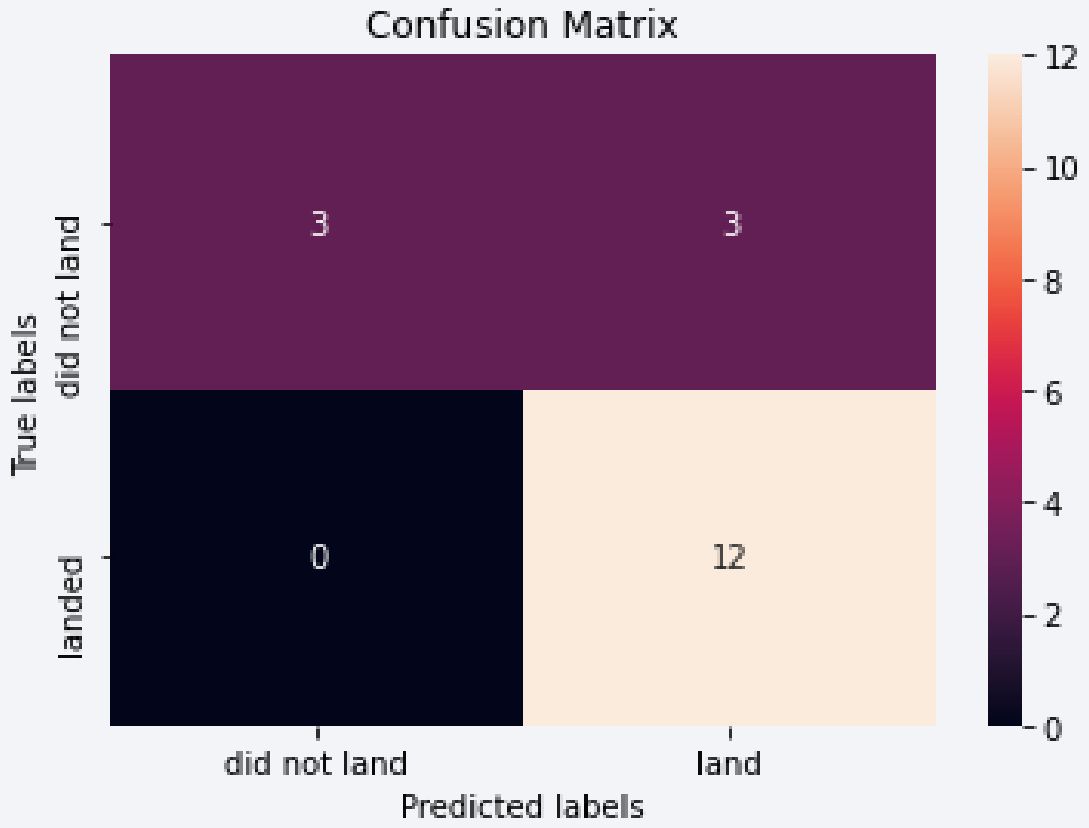
Classification Accuracy

- All classification models performed above 80% accuracy
- The decision tree model had the best accuracy at 88.9%



Confusion Matrix

- The confusion matrix for the Decision Tree model is shown
- The matrix shows that the model had no False Negatives and only 3 false positives



Conclusions

- Our study revealed several interesting points about SpaceX's demonstrated success in landing rockets

1. With experience the likelihood that the booster would successfully land increased
2. The success rate of landing correlated with the desired orbit of the mission
3. The number of successful missions improved for missions with a smaller payload mass
4. The interactive map identified how large scale transportation influences the location of launch sites
5. The model that best predicted landing outcomes was a decision tree classifier with the following parameters:
    - Criterion: Entropy; Max Depth: 12; Max Features: SQRT; Min Samples Leaf: 4; Min Samples Split: 2; Splitter: best

Thank you!