IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Aleksandra Schilis
March, 2024

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection
  - Data Wrangling
  - Explanatory Data Analysis
    - Using SQL
    - Using Pandas and Matplotlib
  - Visual Data Analysis
    - Interactive map with Folium
    - Interactive dashboard with Plotly Dash
  - Predictive Analysis
- Summary of all results
  - EDA results
  - Interactive analysis
  - Predictive analysis

Introduction

- Project background and context
  - SpaceX advertises Falcon 9 rocket launches on its website. A launch costs 62 million dollars, which is cheaper than other providers since SpaceX can reuse the first stage. The success of the first-stage landing can predict the cost of a rocket launch. Space Y wants to bid against SpaceX for a rocket launch.
- Problems you want to find answers
  - To determine the success of the first-stage landing, we need to find factors that help us predict its success.

Section 1

Methodology

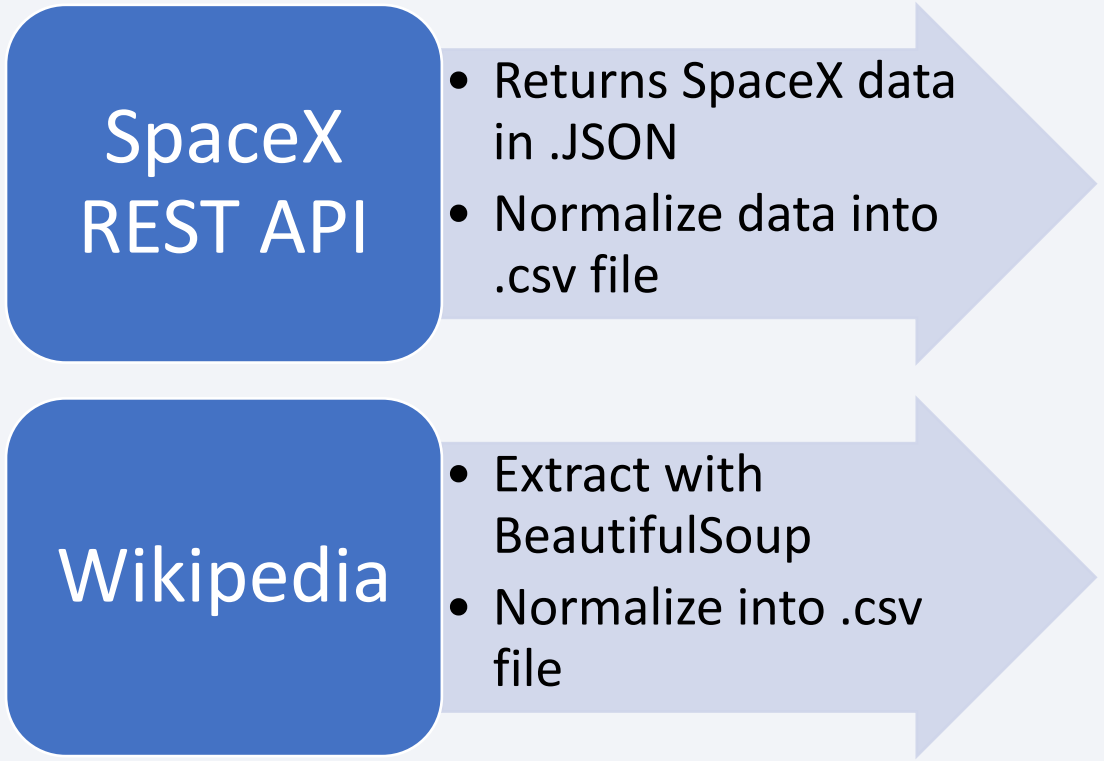
Methodology

Executive Summary

- Data collection methodology:
  - SpaceX Rest API
  - Web Scrapping - Wikipedia
- Perform data wrangling
  - Landing outcomes were converted to classes, mull values and irrelevant columns were eliminated
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Build LR, KNN, SVM, and DT models to evaluate for the best classifier

Data Collection

- The following data sets were collected:
  - SpaceX launch data was gathered from the SpaceX Rest API.
    - api.spacexdata.com/v4/
    - api.spacexdata.com/v4/launches/past
  - Web Scraping from Wikipedia using Beautiful Soup.



Data Collection – SpaceX API

- Data collection with SpaceX REST calls

Steps	Calls
Request data from SpaceX API with the given url	`spacex_url="https://api.spacexdata.com/v4/launches/past"`
Request and parse SpaceX launch data using GET	`static_json_url='https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/`

Decode the response content using .json() and turn it into a Pandas dataframe using .json_normalize()	`jlist = requests.get(static_json_url).json()`
	`df2 = pd.json_normalize(jlist)`
	
	`# Create a data from launch_dict`
	`data_falcon9 = pd.DataFrame(launch_dict)`

export it to a **CSV**	`data_falcon9.to_csv('dataset_part_1.csv', index=False)`

- Link: https://github.com/aschilis/SpaceY/blob/main/jupyter-labs-spacex-data-collection-api%20(3).ipynb

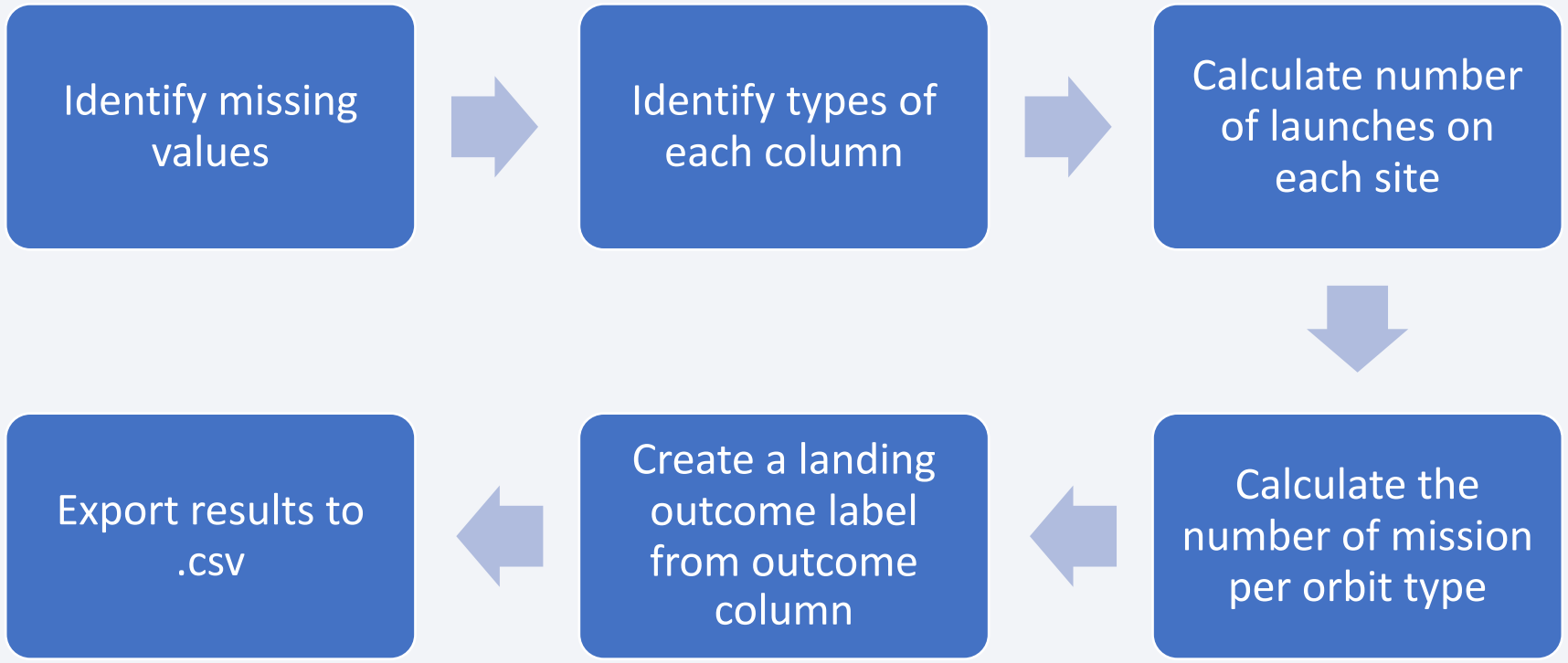
Data Collection - Scraping

- WebScraping

STEPS	CALLS
Getting response from HTML	data = requests.get(static_url).text
Creating BeautifulSoup Object	soup = BeautifulSoup(data, 'html5lib')
Finding Tables	html_tables = soup.find_all('table')
Getting Column Names	column_names = [] **for** row **in** first_launch_table.find_all('th'): name = extract_column_from_header(row)
Creating Dictionary	launch_dict= dict.fromkeys(column_names) #
Converting to Dataframe	df=pd.DataFrame(launch_dict)
Saving to.CSV	df.to_csv('spacex_web_scraped.csv', index=False)

https://github.com/aschilis/SpaceY/blob/main/jupyter-labs-webscraping.ipynb

Data Wrangling



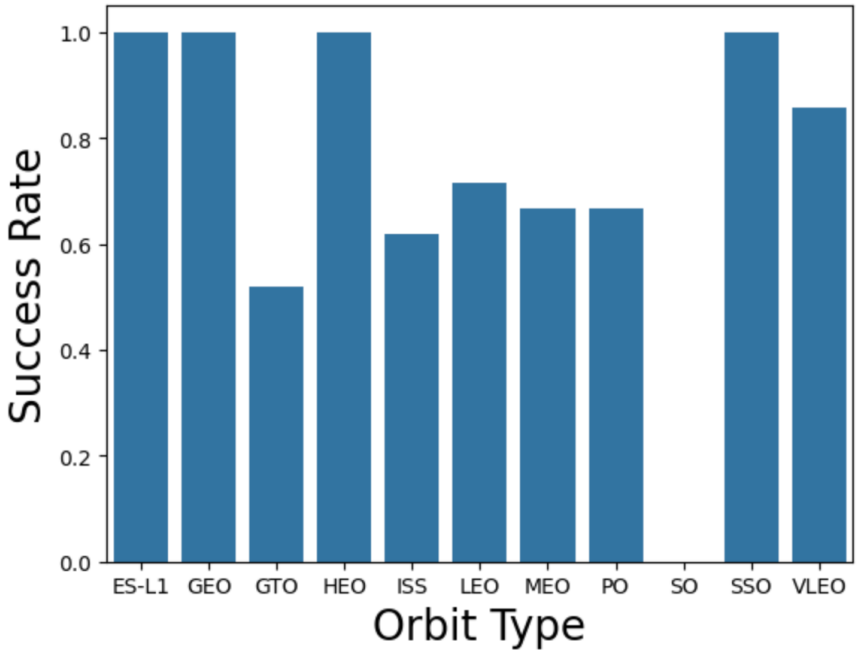
- https://github.com/aschilis/SpaceY/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

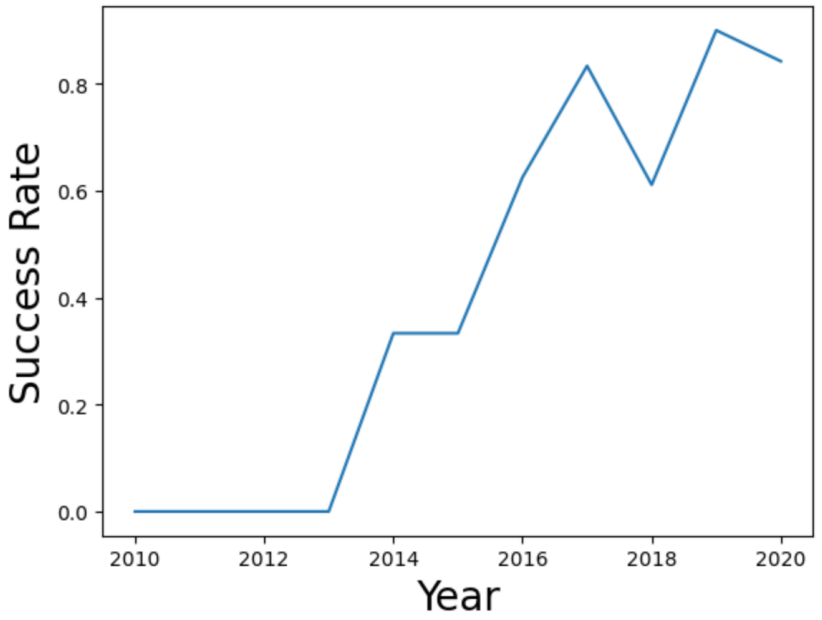
EDA with Data Visualization

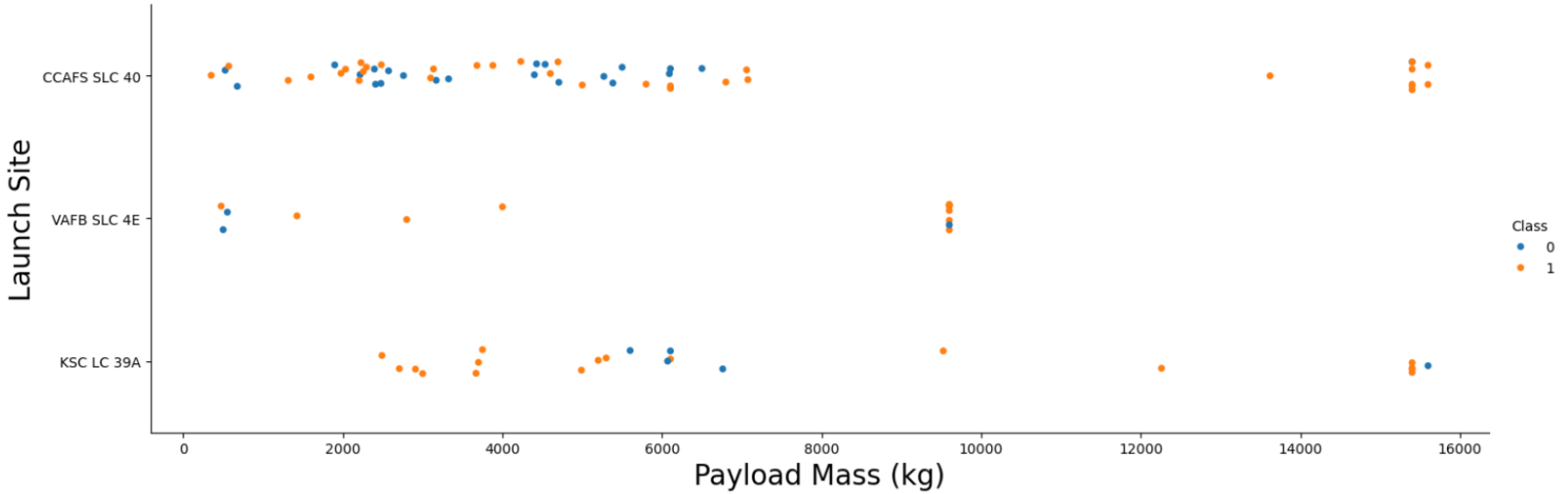
- Graphs and charts were created to explore the relationships between:
  - Flight Number and Payload (Catplot)
  - Flight Number and Launch Site (Catplot)
  - Payload and Launch Site (Catplot)
  - Orbit type and success rate (Bar Chart)
  - Flight Number and Orbit Type (Catplot)
  - Payload and Orbit Type (Catplot)
  - Year and Launch Success (Line Chart)

  - https://github.com/aschilis/SpaceY/blob/main/jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb

EDA with Data Visualization – Sample Graphs







EDA with SQL

SQL queries performed:

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

```
select distinct(LAUNCH_SITE) from SPACEXTBL
```

- Display 5 records where launch sites begin with the string ‘CCA’:

```
select * from SPACEXTBL where LAUNCH_SITE like 'CCA%' limit 5
```

- Display the total payload mass carried by boosters launched by NASA:

```
select sum(PAYLOAD_MASS__KG_) from SPACEXTBL where CUSTOMER = 'NASA (CRS)'
```

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

```
select avg(PAYLOAD_MASS__KG_) from SPACEXTBL where BOOSTER_VERSION = 'F9
```


Link: https://github.com/aschilis/SpaceY/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

EDA with SQL cont.

- List the date when the first successful landing outcome in ground pad was achieved:

```
select min(DATE) from SPACEXTBL where Landing_Outcome = 'Success (ground pad)'
```

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

```
select BOOSTER_VERSION from SPACEXTBL where Landing_Outcome = 'Success (drone ship)'
and PAYLOAD_MASS__KG_ > 4000 and PAYLOAD_MASS__KG_ < 6000
```

- List the total number of successful and failure mission outcomes:

```
select count(MISSION_OUTCOME) from SPACEXTBL where MISSION_OUTCOME = 'Success'
or MISSION_OUTCOME = 'Failure (in flight)'
```

EDA with SQL cont.

- List the names of the booster versions which have carried the maxiumum payload mass. Use a subquery:

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

- 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:

```
select substr(Date,4,2) as month, DATE,BOOSTER_VERSION, LAUNCH_SITE, [Landing_Outcome] from SPACEXTBL where [Landing_Outcome] = 'Failure (drone ship)' and substr(Date,7,4)='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:

```
select [Landing_Outcome], count(*) as count_outcomes from SPACEXTBL where DATE between '04-06-2010' and '20-03-2017' group by [Landing _Outcome] order by count_outcomes DESC;
```

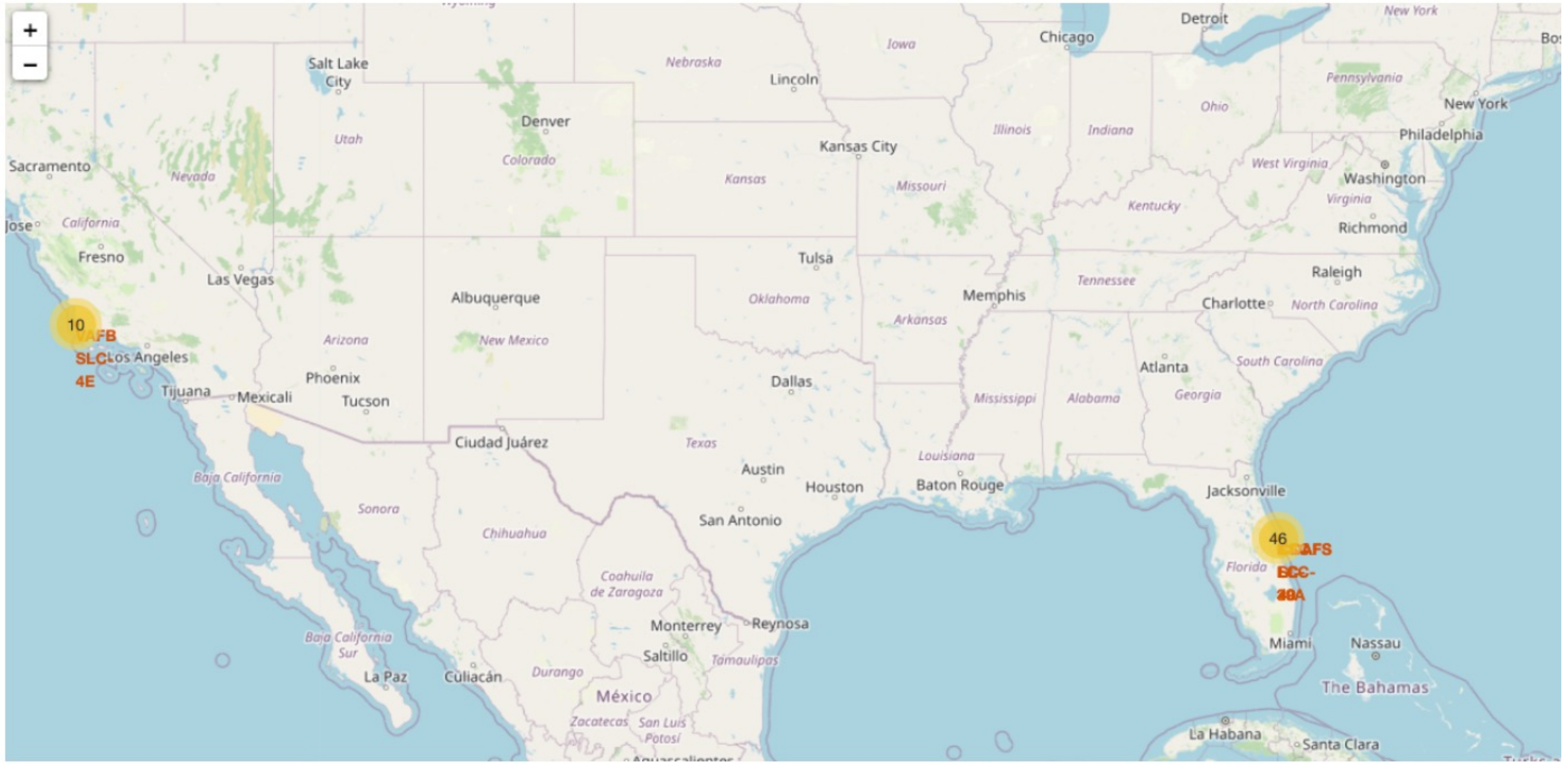
Build an Interactive Map with Folium

Map objects created and added to the Folium map:

- A highlighted circle with a text label on a specific coordinate to label each launch site on the map – with *folium.Circle* and *folium.Marker.*
- An object for combining markers with the same coordinate – with *MarkerCluster.*
- An object to draw a line between a launch site to its closest city, railway and highway – with *folium.PolyLine*.
- A mouse over a point on the map - with *MousePosition.*

- Link: https://github.com/aschilis/SpaceY/blob/main/lab_jupyter_launch_site_location.jupyterlite.ipynb

Build an Interactive Map with Folium



Predictive Analysis (Classification)

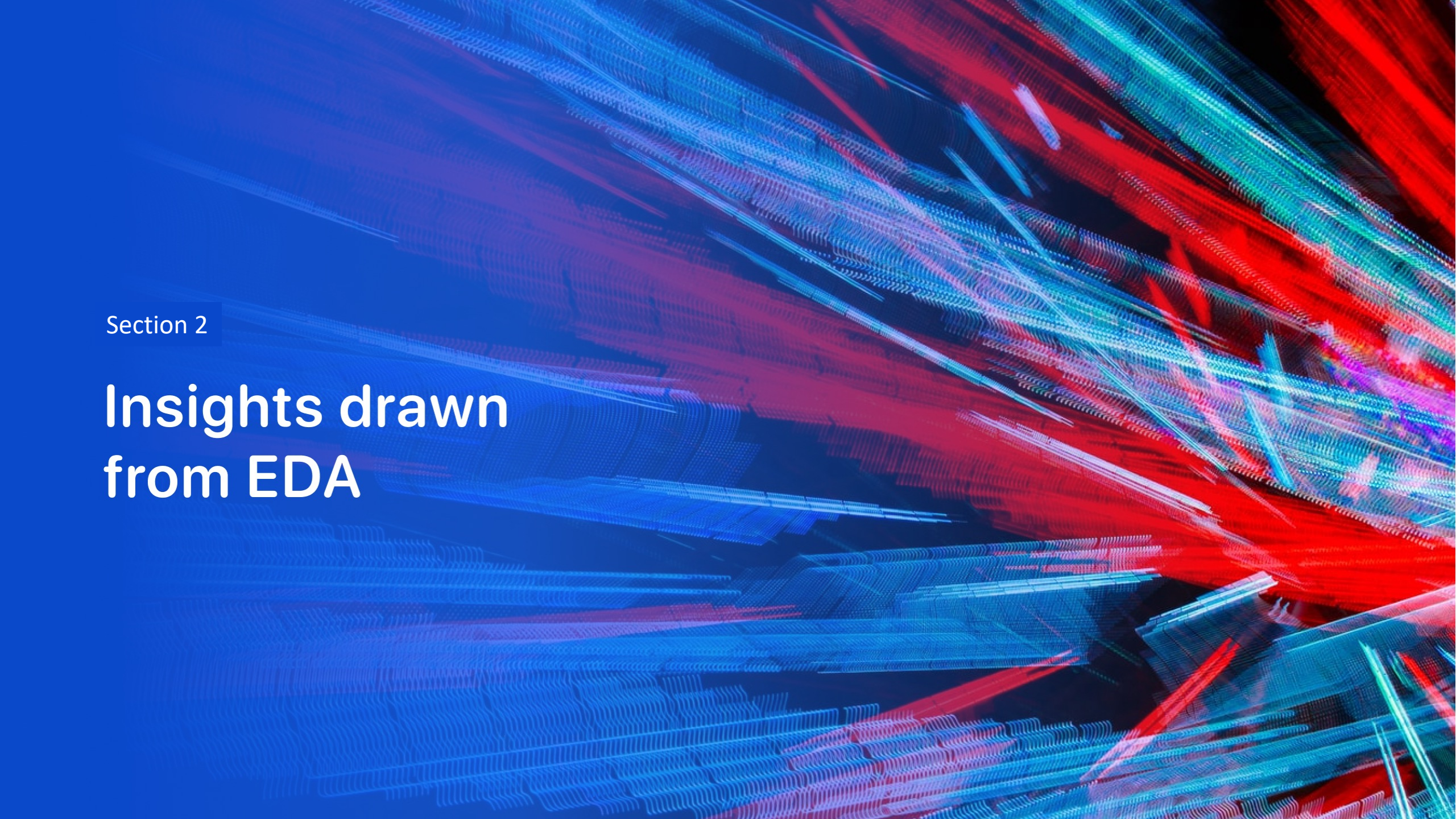
Perform Predictive Analysis:

- Upload data with numpy and pandas
- Transform data and split it into training and testing.
- Use different machine learning models and tune different hyperparameters using GridSearchCV
- Search for the best hyperparameters for Logistic Regression, SVM, Decision Tree and KNN classifiers
- Find the method that performs best using test data.

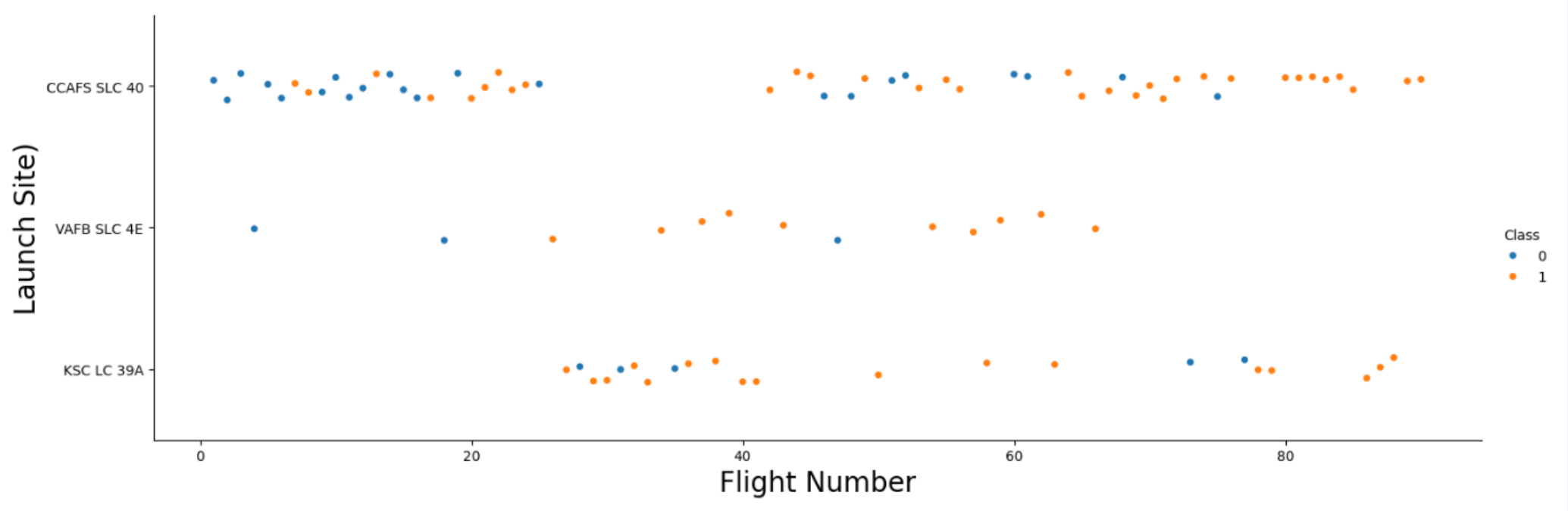
- Link: https://github.com/aschilis/SpaceY/blob/main/SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

Results

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

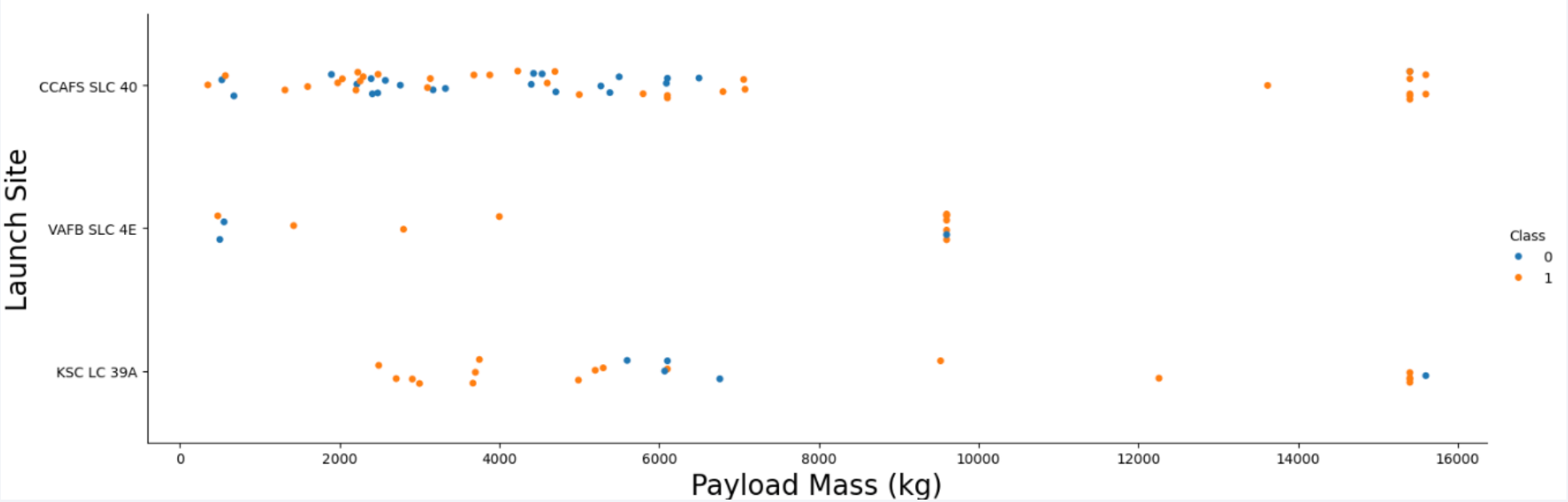

Section 2
Insights drawn from EDA

Flight Number vs. Launch Site



- From the plot, we see that the higher flight number at a launch site, the higher the success rate.
- Launches from the site of CCAFS SLC 40 had the most launches, the early flights mostly ended with a failure.

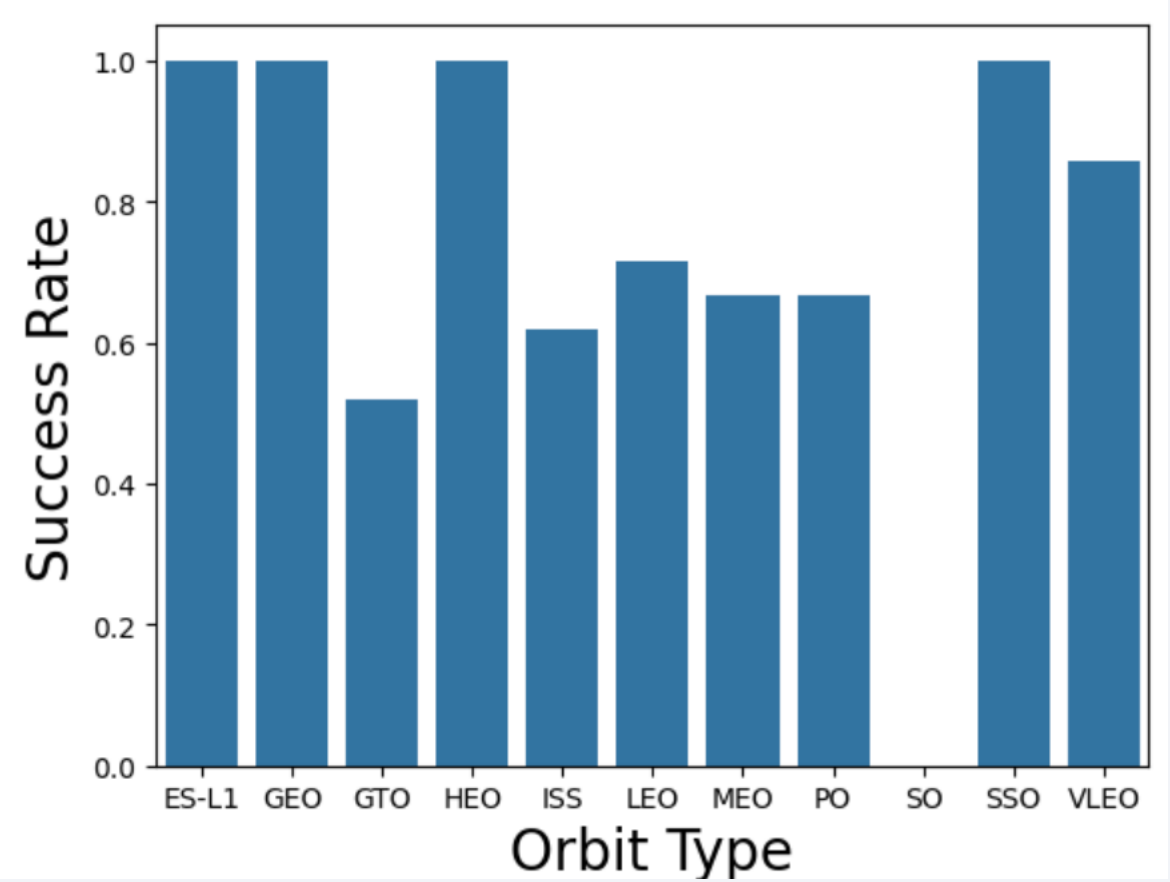
Payload vs. Launch Site



- The greater the payload mass, the higher the success rate for the rocket
- Launch site KSC LC has no rockets with payload mass less than 2500 kg
- VAFB-SLC launch site has no rockets launched with payload mass greater than 10000kg
- The majority of Payload Mass has been launched from CCAFS SLC 40

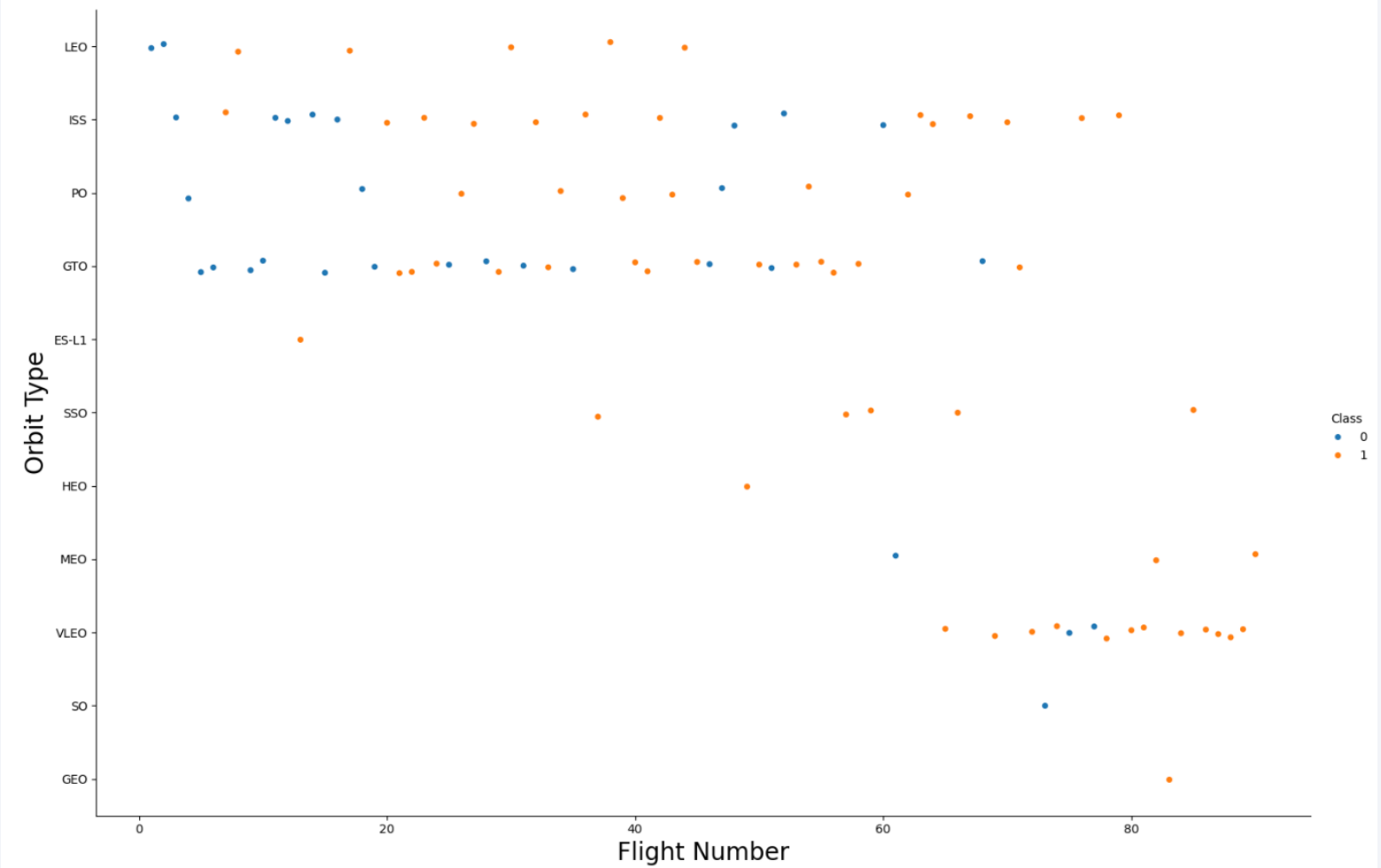
Success Rate vs. Orbit Type

- The following orbit types: ES-L1, GEO, HEO, SSO have the most successful rate.



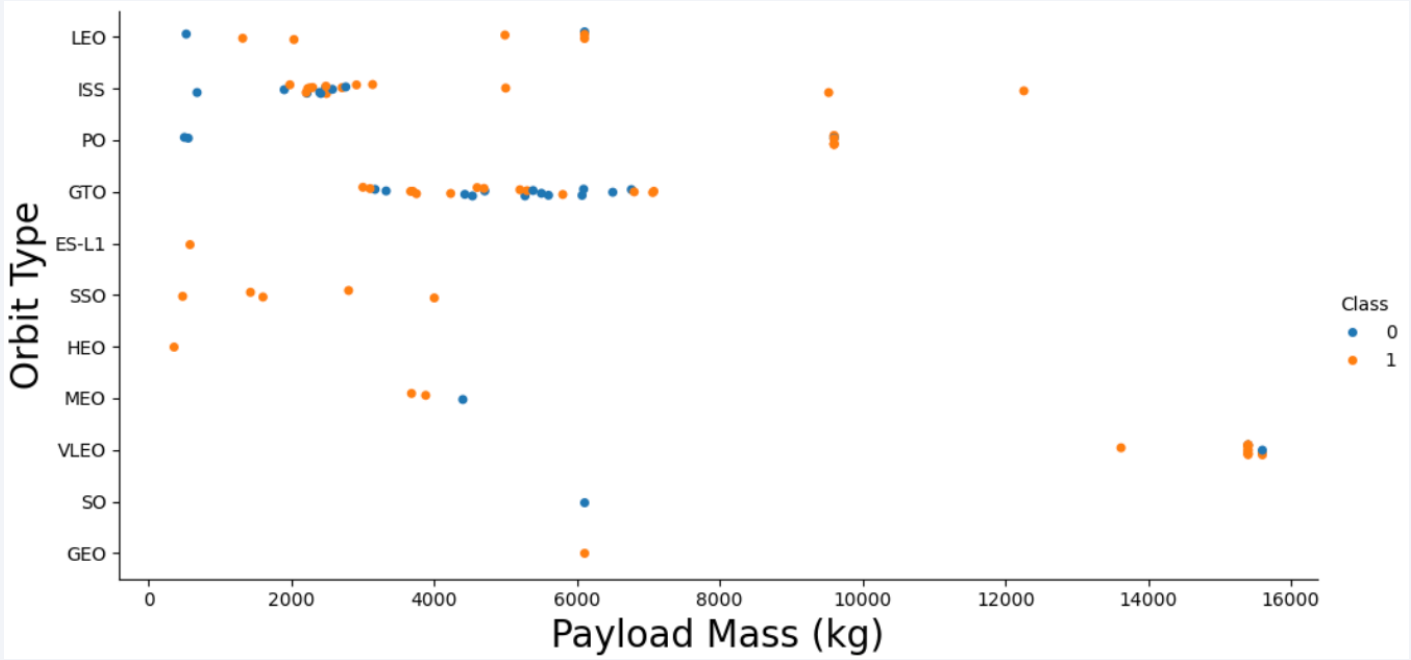
Flight Number vs. Orbit Type

- The initial launches were less successful, but as the flight number increases, the success of the launch becomes more visible.
- Most recent rocket launches were to VLEO orbit
- LEO orbit has the highest percentage rate of success.



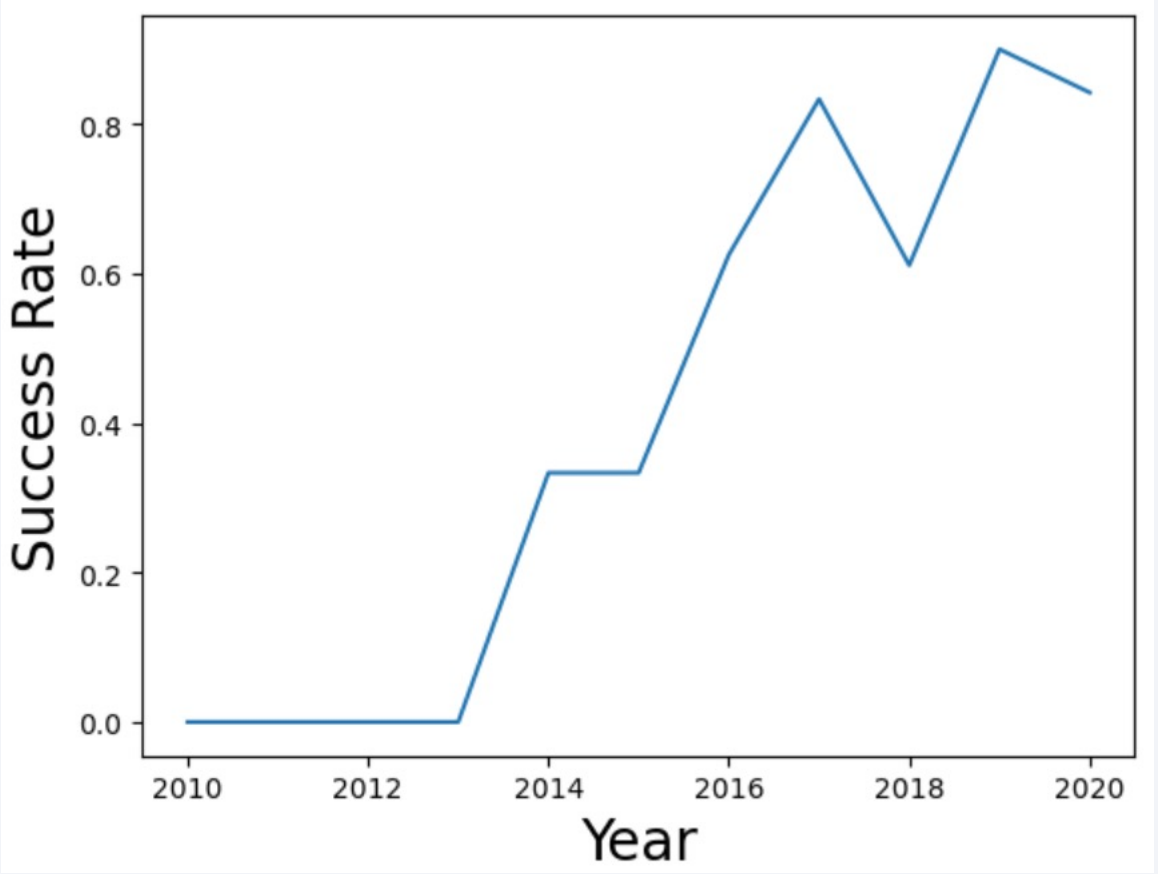
Payload vs. Orbit Type

- There is some correlation between the payload mass, orbit, and successful landing.
- The heavier the payload mass, the higher the successful landing rate.
- The orbits with the most successful landing:
- For Sso, all launches were a success; however, the payload mass was below 4000kg



Launch Success Yearly Trend

- The success rate has been increasing since 2013.
- There has been a slight drop in the success rate in 2018.



All Launch Site Names

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

In [12]:
```
%sql select distinct(LAUNCH_SITE) from SPACEXTBL
```

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

Out[12]:

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.

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

```
]: %sql select * from SPACEXTBL where LAUNCH_SITE like 'CCA%' limit 5
```

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

]:

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

Total Payload Mass

- Calculate the total payload carried by boosters from NASA

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

```
%sql select sum(PAYLOAD_MASS__KG_) from SPACEXTBL where CUSTOMER =
```

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

sum(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1

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

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

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

avg(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on a ground pad

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

Hint:Use min function

```
%sql select min(DATE) from SPACEXTBL where Landing_Outcome = 'Success (ground pad)'
```

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

min(DATE)
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.

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

```
%sql select BOOSTER_VERSION from SPACEXTBL where Landing_Outcome = 'Success (drone ship)' and PAYLOAD_MASS__KG_ >
```

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

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.

List the total number of successful and failure mission outcomes

```
%sql select count(MISSION_OUTCOME) from SPACEXTBL where MISSION_OUTCOME = 'Success' or MISSION_OUTCOME = 'Failure
```

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

count(MISSION_OUTCOME)
99

Boosters Carried Maximum Payload

```
select BOOSTER_VERSION from SPACEXTBL where PAYLOAD_MASS__KG_ = (select max(PAYLOAD_MASS__KG_) from SPACEXTB
```

- List the names of the booster which have carried the maximum payload mass

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

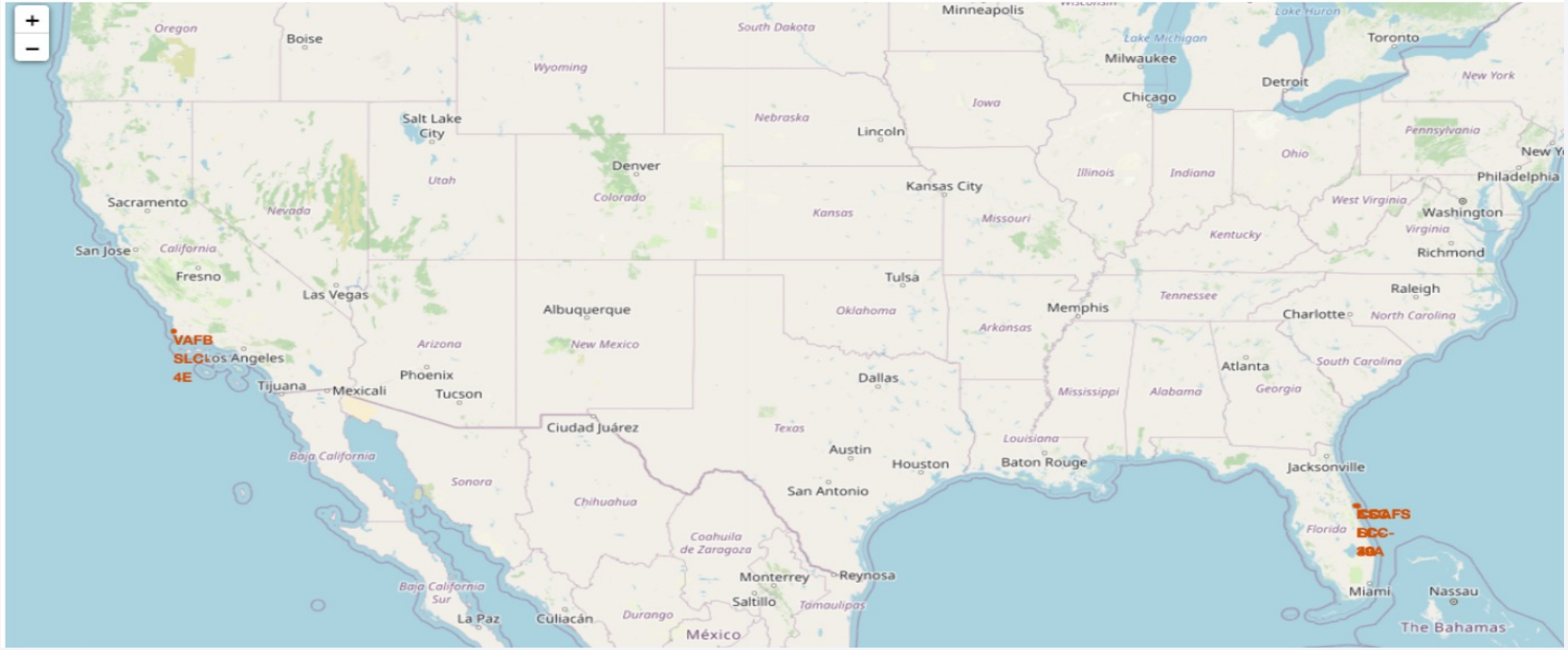
```
%sql SELECT substr(Date,4,2) as month, DATE,BOOSTER_VERSION, LAUNCH_SITE, [Landing_Outcome] from SPACEXTBL where
```

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

time__utc_	booster_version	launch_site	payload	payload_mass__kg_	orbit	customer	mission_outcome	landing__outcome
14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
17:54:00	F9 FT B1029.1	VAFB SLC-4E	Iridium NEXT 1	9600	Polar LEO	Iridium Communications	Success	Success (drone ship)
05:26:00	F9 FT B1026	CCAFS LC-40	JCSAT-16	4600	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
04:45:00	F9 FT B1025.1	CCAFS LC-40	SpaceX CRS-9	2257	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
21:39:00	F9 FT B1023.1	CCAFS LC-40	Thaicom 8	3100	GTO	Thaicom	Success	Success (drone ship)


Section 3
Launch Sites
Proximities Analysis

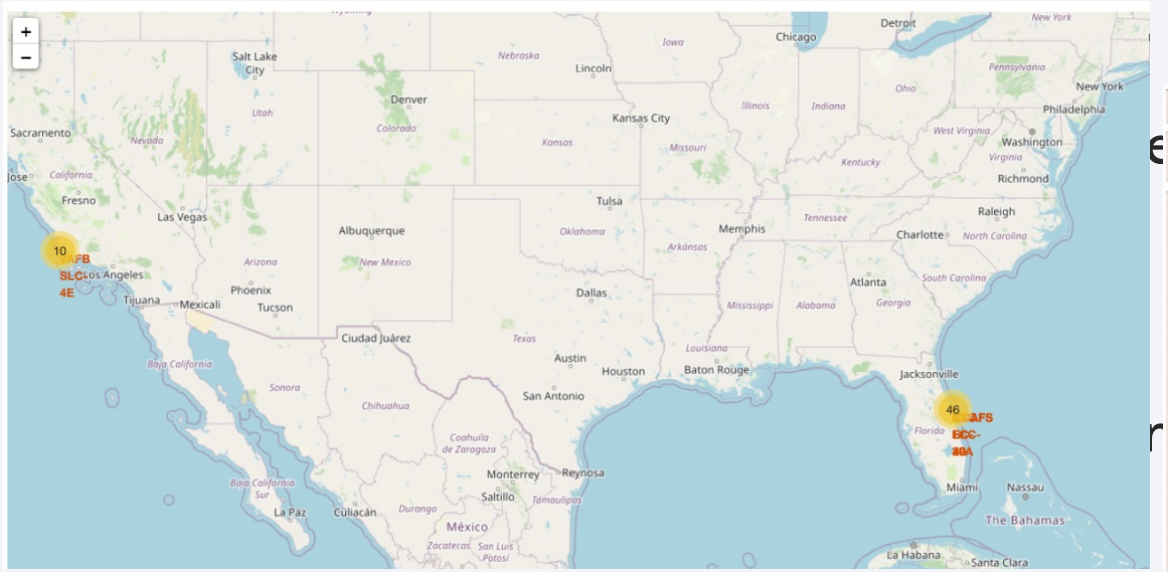
SpaceX Launch Sites

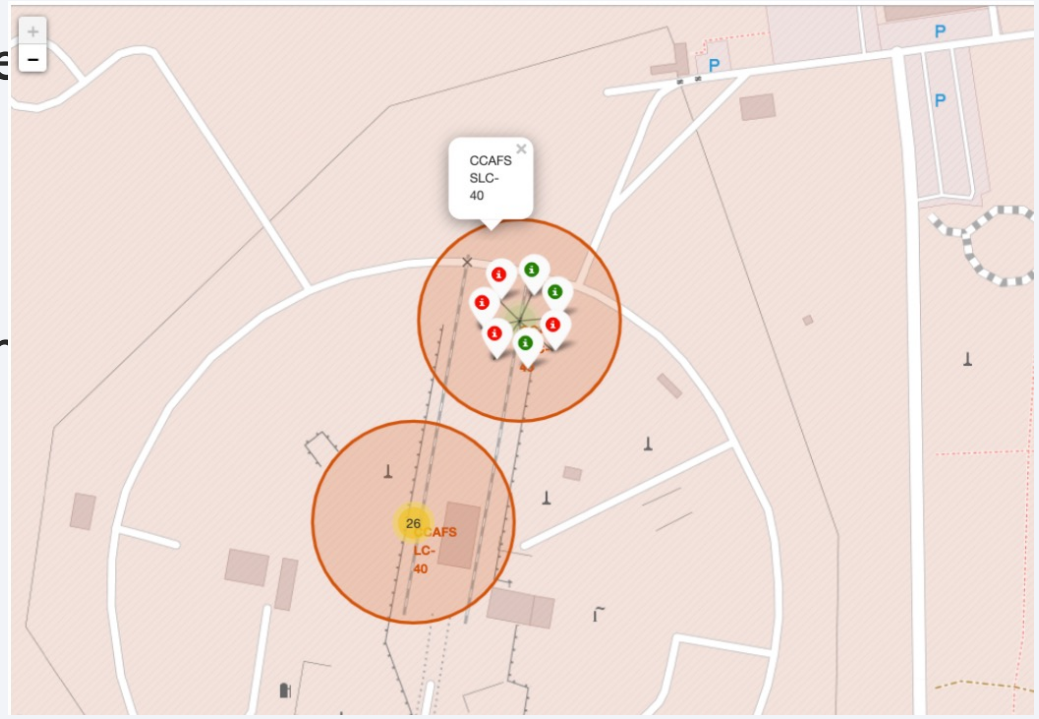


- All the SpaceX launch sites are in the United States of America. All sites are in close proximity to the coasts.

Launch Sites

- Markers showing successful launches (green color) and failures (red color).





Distance to Landmarks

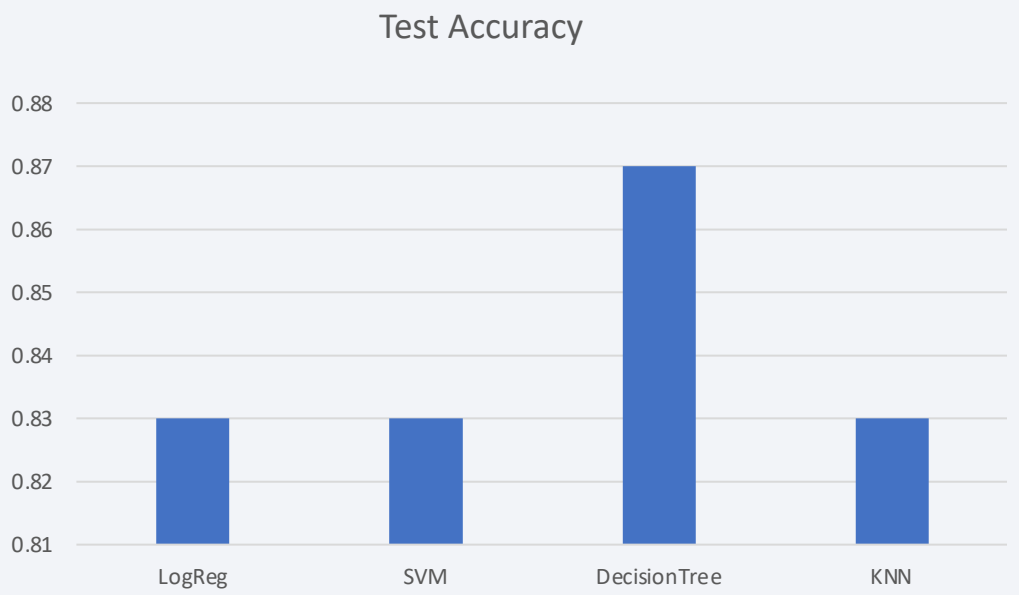
- Markers showing distance to specific landmarks:
  - Railway station
  - Closest highway
  - Coastline
  - City.

Section 5

Predictive Analysis (Classification)

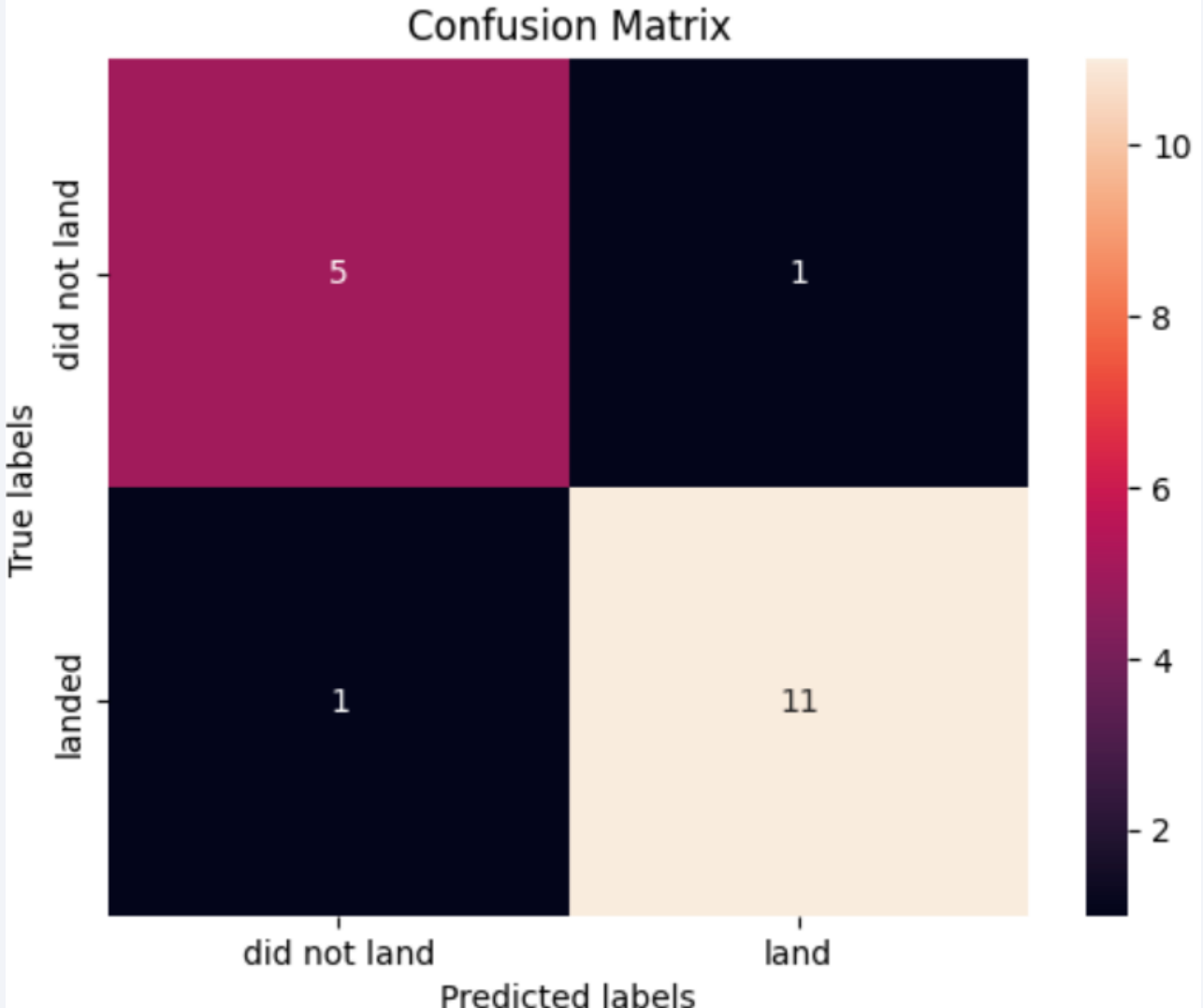
Classification Accuracy

Model	Test Accuracy
LogRegression	0.8333333333333334
SVM	0.8333333333333334
Decision Tree	0.8888888888888888
KNN	0.8333333333333334



Confusion Matrix

```
yhat = tree_cv.predict(X_test)
plot_confusion_matrix(Y_test,yhat)
```



Conclusions

Based on our analysis, we can conclude:

- The higher the number of flights at a given launch site, the greater its rate of success.
- The success rate has been increasing since 2013.
- Orbits ES-L1, GEO, HEO, SSO, VLEO have the highest success rate.
- KSC LC – 39A has been the most successful launch site
- The Decision tree classifier had the best machine-learning algorithm for this data set.
- Based on our analysis, we can predict the successful rate of rocket landing and determine the cost of a launch.

Appendix

- For notebooks, datasets, and explanations follow this GitHub repository link:
- https://github.com/aschilis/SpaceY

Thank you!