IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Tanmaya Joshi
18.07.2024

Outline

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

Executive Summary

This capstone is aimed at utilizing data science methodologies to determine the cost of a rocket launch from historical Falcon 9 launch and first stage landing records.

- Summary of methodologies

Data Collection

- Data Collection API
- Web Scraping
- Data Wrangling

Exploratory Data Analysis

- Exploratory analysis using SQL
- Exploratory analysis using Pandas and Matplotlib
- Interactive Visual Analytics and Dashboards

Predictive Analysis (Classification)

- Machine Learning Prediction

- Key results

Analysis revealed a correlation between outcome of the launches and features such as payload mass and launch sites.

Different launch sites have different success rates: CCAFS LC-40 has a success rate of 60 %, while KSC LC-39A and VAFB SLC 4E has a success rate of 77%.

Orbits ES-L1, SSO, HEO, GEO have high success rate.

Introduction

- SpaceX has gained worldwide attention for a series of historic milestones.
- SpaceX is the only private company ever to return a spacecraft from low-earth orbit, which it first accomplished in December 2010.
- It advertises Falcon 9 rocket launches with a cost of 62 million dollars whereas it costs upward of 165 million dollars each for the other providers.
- Much of the savings is because SpaceX can reuse the first stage.

Determining if the first stage will land can help to determine the cost of a launch. This information can be used if an alternate company wants to bid against SpaceX for a rocket launch.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - By making a get request to the SpaceX API
  - By performing web scraping to collect Falcon 9 historical launch records from a Wikipedia page titled List of Falcon 9 and Falcon Heavy launches: https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches
- Perform data wrangling
  - Using the ”Outcome”, a classification variable “landing_class“ was created that represents the outcome of each launch (0 = first stage did not land successfully; 1 = first stage landed successfully)
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using Support Vector Machine (SVM), Decision Tree, k nearest neighbors (KNN) and Logistic Regression classification models. The accuracy of each model was calculated on the test data and compared to each other to identify the best performing model

Data Collection - SpaceX API

1. Rocket launch data was requested from SpaceX API with the following URL: "https://api.spacexdata.com/v4/launches/past"
2. The response content was decoded as a Json using .json() and turned it into a Pandas dataframe using .json_normalize() as: `data = pd.json_normalize(response.json())`
3. Next, the data dataframe was filtered to only keep the Falcon 9 launches. The filtered data was saved to a new dataframe called data_falcon9
4. The missing values for PayloadMass in the data set were replaced with mean for the PayloadMass, calculated as: `PayLoadMean = data_falcon9[‘PayloadMass’].mean()`
5. In the final step, data was exported to a CSV for the next section as: `data_falcon9.to_csv('dataset_part_1.csv', index=False)`

Data Collection - SpaceX API

Flowchart of SpaceX API calls

First few rows of filtered dataframe “data_falcon9” containing only Falcon9 launches:

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude
4	1	2010-06-04	Falcon 9	NaN	LEO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0003	-80.577366	28.561857
5	2	2012-05-22	Falcon 9	525.0	LEO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0005	-80.577366	28.561857
6	3	2013-03-01	Falcon 9	677.0	ISS	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0007	-80.577366	28.561857
7	4	2013-09-29	Falcon 9	500.0	PO	VAFB SLC 4E	False Ocean	1	False	False	False	None	1.0	0	B1003	-120.610829	34.632093
8	5	2013-12-03	Falcon 9	3170.0	GTO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B1004	-80.577366	28.561857

- GitHub URL of the completed SpaceX API calls notebook: https://github.com/tanmayajoshi/IBM-DataScience-Capstone/blob/ec4fc2f47c6f90901b993951c6610834c50a2b55/1_jupyter-labs-spacex-data-collection-api.ipynb

Data Collection - Scraping

- Falcon 9 historical launch records were web scraped from a Wikipedia page titled List of Falcon 9 and Falcon Heavy launches : https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches

Flowchart for web scraping

1. Wikipedia source → 2. Extracting Falcon 9 launch records HTML table → 3. Parsing and converting to Pandas dataframe → 4. Export data set to CSV

First few rows of data frame created by parsing the launch HTML tables:

	Flight No.	Launch site	Payload	Payload mass	Orbit	Customer	Launch outcome	Version Booster	Booster landing	Date	Time
0	1	CCAFS	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success\n	F9 v1.0B0003.1	Failure	4 June 2010	18:45
1	2	CCAFS	Dragon	0	LEO	NASA	Success	F9 v1.0B0004.1	Failure	8 December 2010	15:43
2	3	CCAFS	Dragon	525 kg	LEO	NASA	Success	F9 v1.0B0005.1	No attempt\n	22 May 2012	07:44
3	4	CCAFS	SpaceX CRS-1	4,700 kg	LEO	NASA	Success\n	F9 v1.0B0006.1	No attempt	8 October 2012	00:35
4	5	CCAFS	SpaceX CRS-2	4,877 kg	LEO	NASA	Success\n	F9 v1.0B0007.1	No attempt\n	1 March 2013	15:10

- GitHub URL of the completed web scraping notebook: https://github.com/tanmayajoshi/IBM-DataScience-Capstone/blob/ec4fc2f47c6f90901b993951c6610834c50a2b55/2_jupyter-labs-webscraping.ipynb

Data Wrangling

- Exploratory Data Analysis (EDA) was performed to find some patterns in the data and determine training labels

- A classification variable was created that represents the outcome of each launch. If the value is zero, the first stage did not land successfully; one means the first stage landed Successfully. The values were added to the column 'Class' in the data frame.

Outcome	Value of landing_class variable
True ASDS	1
None None	0
True RTLS	1
False ASDS	0
True Ocean	1
False Ocean	0
None ASDS	0
False RTLS	0

```
df['Class']=landing_class
df[['Class']].head(8)
```

	Class
0	0
1	0
2	0
3	0
4	0
5	0
6	1
7	1

- GitHub URL of the completed data wrangling related notebook: https://github.com/tanmayajoshi/IBM-DataScience-Capstone/blob/ec4fc2f47c6f90901b993951c6610834c50a2b55/3_labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

For Data Visualization, following chart types were used:

1. **Categorical plot:**
   To see how the FlightNumber (indicating the continuous launch attempts) and Payload variables would affect the launch outcome; to visualize the relationship between Flight Number and Launch Site.

2. **Scatter plot**:
   To visualize the relationship between Payload and Launch Site, Flight Number and Orbit type, Payload and Orbit type.

3. **Bar plot:**
   To visualize the relationship between success rate of each orbit type.

4. **Line chart**:
   To visualize the launch success yearly trend.

- GitHub URL of the completed EDA with data visualization notebook: https://github.com/tanmayajoshi/IBM-DataScience-Capstone/blob/ec4fc2f47c6f90901b993951c6610834c50a2b55/5_edadataviz.ipynb

EDA with SQL

- SQL queries performed:
  - ❑ Names of the unique launch sites in the space mission
  - ❑ Launch sites beginning with the string ‘CCA’
  - ❑ Total payload mass carried by boosters launched by NASA (CRS)
  - ❑ Average payload mass carried by booster version F9 v1.1
  - ❑ Date for first successful landing outcome in ground pad
  - ❑ Names of boosters which have success in drone ship and have payload mass in given range
  - ❑ Total number of successful and failure mission outcomes
  - ❑ Names of the booster_versions which have carried the maximum payload mass
  - ❑ Records from year 2015 with failed landing
  - ❑ Count of landing outcomes

- GitHub URL of the completed EDA with SQL notebook: https://github.com/tanmayajoshi/IBM-DataScience-Capstone/blob/ec4fc2f47c6f90901b993951c6610834c50a2b55/4_jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

For **interactive visual analytics,** following map objects were created and added to a folium map:

1. All launch sites (**circles**)
2. Success/failed launches for each site (**marker clusters**)
3. The distances between a launch site to its proximities such as, railway, highway, coastline, city (**lines**)

The map objects were added to find some geographical patterns about launch sites. For example, using color-labeled markers in marker clusters, launch sites having relatively high success rates can be easily identified.

- GitHub URL of the completed interactive map with Folium map: https://github.com/tanmayajoshi/IBM-DataScience-Capstone/blob/ec4fc2f47c6f90901b993951c6610834c50a2b55/6_lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- Plotly Dash application was built to perform interactive visual analytics on SpaceX launch data in real-time.
- The dashboard application contains input components such as a launch site dropdown list and a payload range slider to interact with a pie chart and a scatter point chart.
- Payload range slider was added to visually observe how payload is correlated with mission outcomes for selected site(s).

From a dashboard point of view, it was intended to be able to easily select different payload range and identify visual patterns, if any.

- GitHub URL of the completed Plotly Dash lab: https://github.com/tanmayajoshi/IBM-DataScience-Capstone/blob/ec4fc2f47c6f90901b993951c6610834c50a2b55/7_spacex_dash_app.py

Predictive Analysis (Classification)

Steps taken to create a machine learning pipeline:

1. create a column for the class
2. standardize the data
3. Split data - Data was split into training and testing data using the function train_test_split.

4. The training data was divided into validation data, a second set used for training data; then the models were trained and hyperparameters selected using the function GridSearchCV.

5. Hyperparameters were tuned to obtain the best hyperparamters for SVM, decision tree, k nearest neighbors, and logistic regression classification models.
6. Test data was used to identify the best performing modelling algorithm.

- GitHub URL of the completed predictive analysis lab: https://github.com/tanmayajoshi/IBM-DataScience-Capstone/blob/4f30d1e0ebcc67de17076ab89efd1f5a86d22b6a/8_SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Results

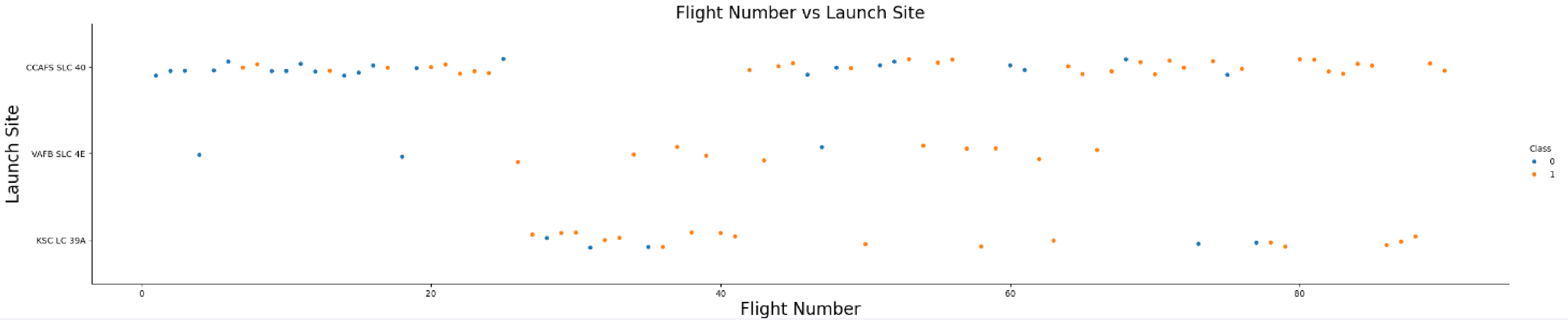
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

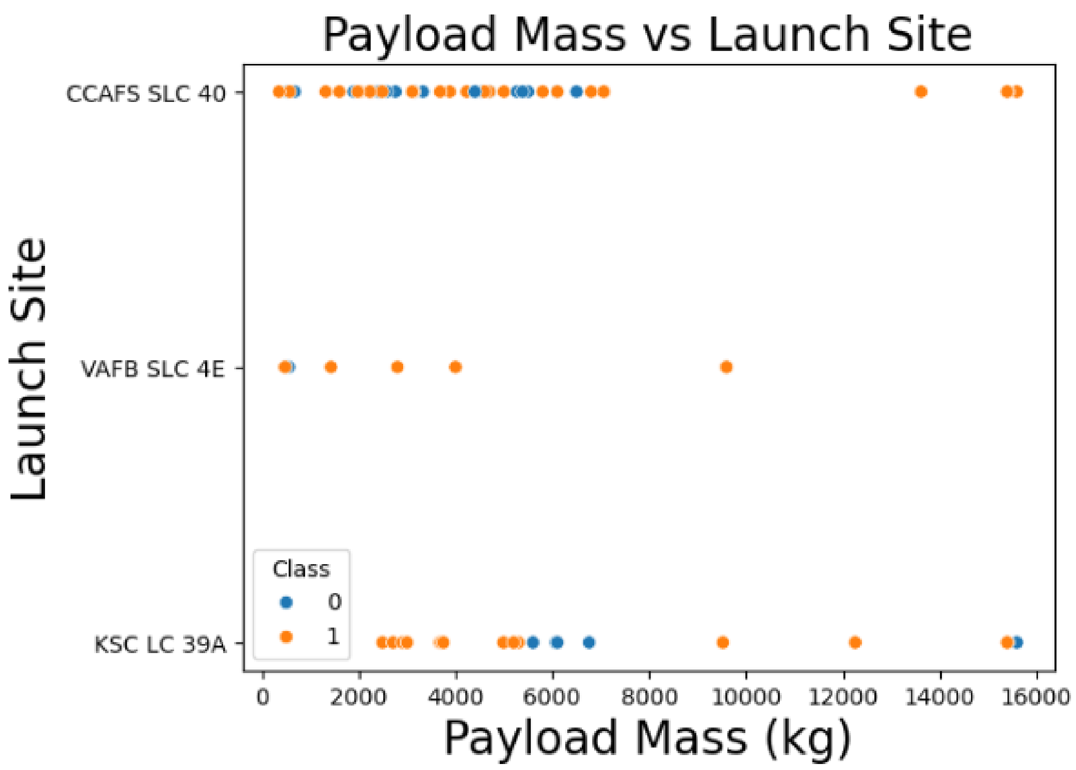
```
# Plot a scatter point chart with x axis to be Flight Number and y axis to be the launch site, and hue to be the class value
sns.catplot(y="LaunchSite", x="FlightNumber", hue="Class", data=df, aspect = 5)
plt.xlabel("Flight Number",fontsize=20)
plt.ylabel("Launch Site",fontsize=20)
plt.title("Flight Number vs Launch Site", fontsize=20)
plt.show()
```



- Explanation: We see that different launch sites have different continuous launch attempts (flight number), with more launches at small flight numbers from site CCAFS SLC 40. Launches from VAFB SLC 4E and KSC LC 39A have a higher success rate. Also, as the flight number increases, the likelihood of success with landing of first stage increases.

Payload vs. Launch Site

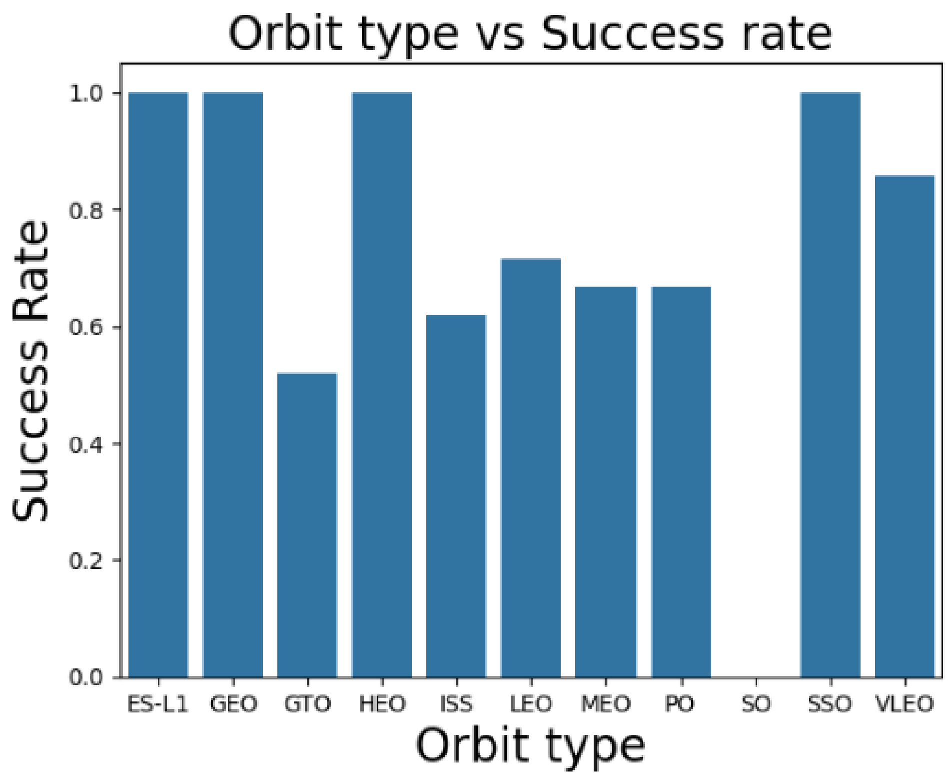
```
# Plot a scatter point chart with x axis to be Pay Load Mass (kg) and y axis to be the launch site, and hue to be the class value
sns.scatterplot(x="PayloadMass", y="LaunchSite", hue="Class", data=df)
plt.xlabel("Payload Mass (kg)", fontsize=20)
plt.ylabel("Launch Site", fontsize=20)
plt.title("Payload Mass vs Launch Site", fontsize=20)
plt.show()
```



- Explanations: For the VAFB-SLC launchsite there are no rockets launched for heavy payload mass (greater than 10000 Kg). Also, at CCAFS SLC 40 and KSC LC 39A launchsites, payload mass is mostly under 8000 Kg.

Success Rate vs. Orbit Type

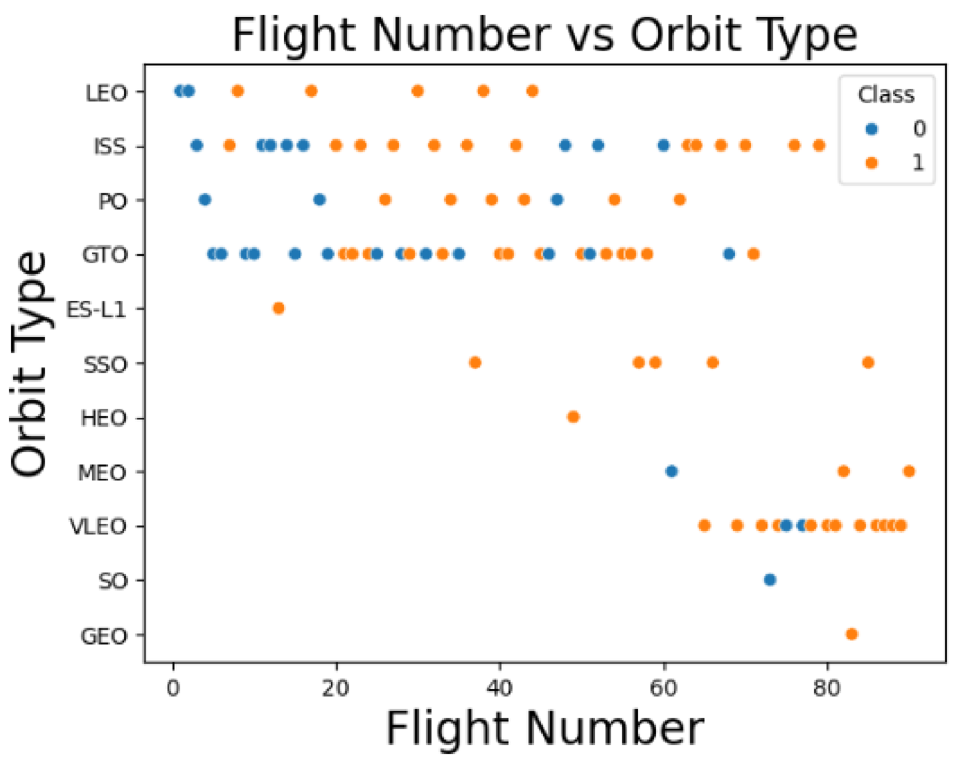
```
[27]: # HINT use groupby method on Orbit column and get the mean of Class column
Data_orbit = df.groupby('Orbit')['Class'].mean().reset_index()
sns.barplot(x="Orbit", y="Class", data=Data_orbit)
plt.xlabel("Orbit type", fontsize=20)
plt.ylabel("Success Rate", fontsize=20)
plt.title("Orbit type vs Success rate", fontsize=20)
plt.show()
```



- Explanation: Orbits ES-L1, GEO, HEO, and SSO have the highest success rate, which is 100%. Also, in SO orbit, the rate is zero.

Flight Number vs. Orbit Type

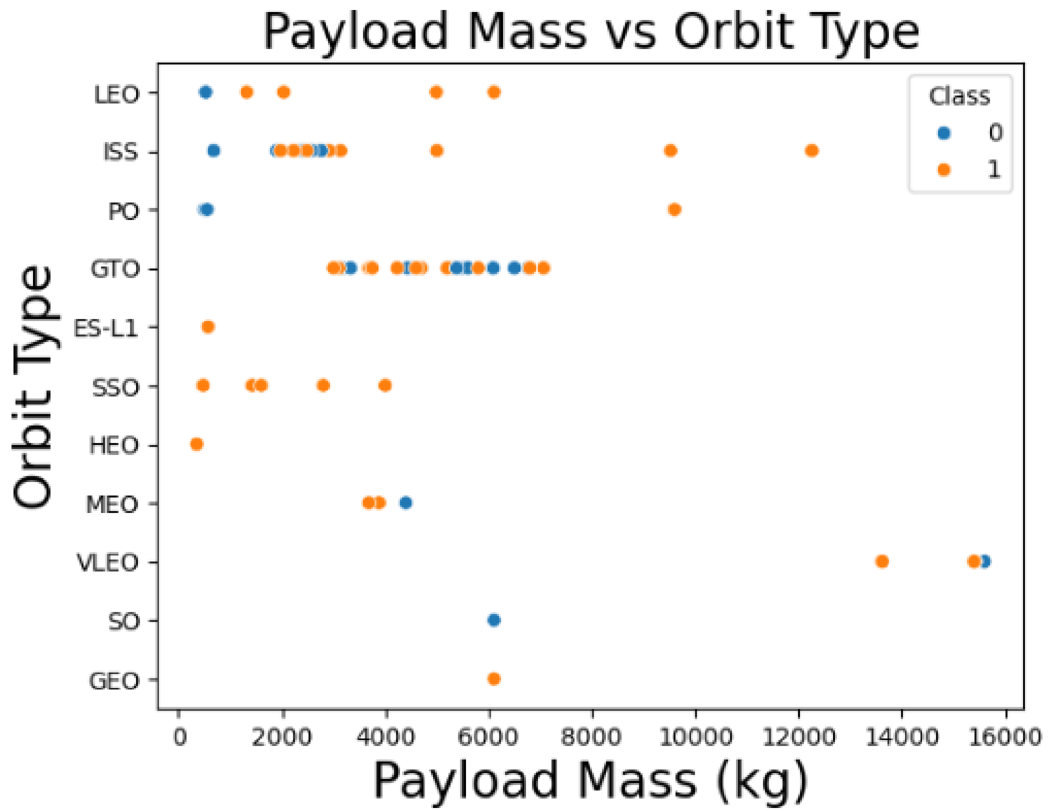
```
[28]: # Plot a scatter point chart with x axis to be FlightNumber and y axis to be the Orbit, and hue to be the class value

sns.scatterplot(x="FlightNumber", y="Orbit", hue="Class", data=df)
plt.xlabel("Flight Number", fontsize=20)
plt.ylabel("Orbit Type", fontsize=20)
plt.title("Flight Number vs Orbit Type", fontsize=20)
plt.show()
```



- Explanations: In the LEO orbit, success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.

Payload vs. Orbit Type

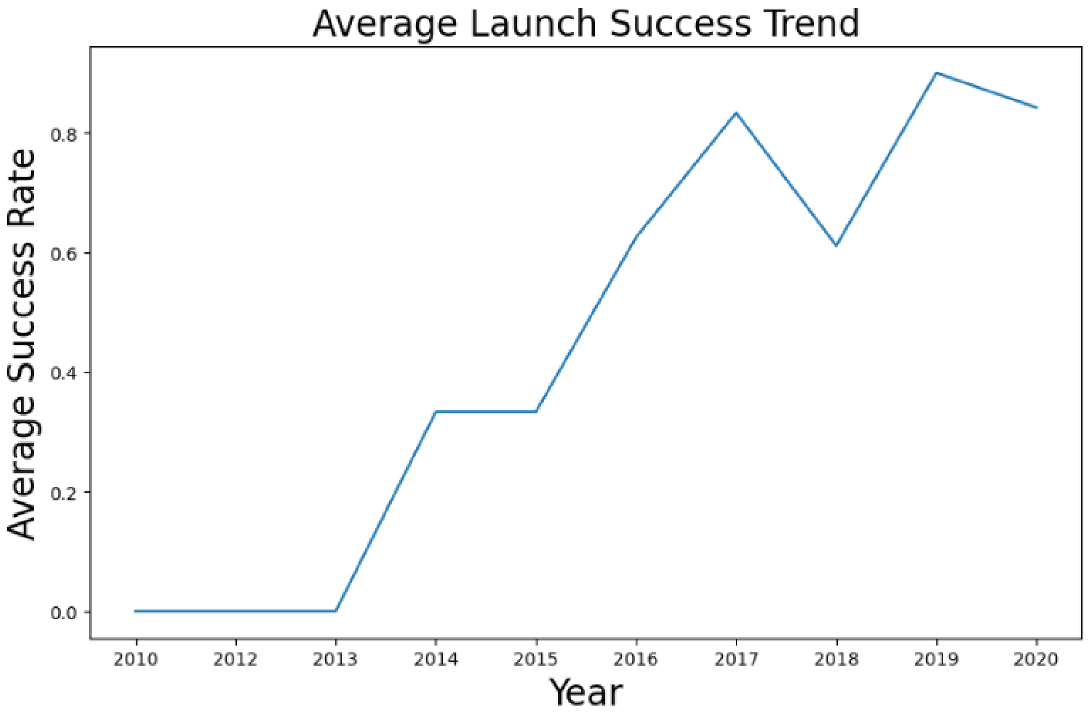
```
[29]: # Plot a scatter point chart with x axis to be Payload and y axis to be the Orbit, and hue to be the class value
sns.scatterplot(x="PayloadMass", y="Orbit", hue="Class", data=df)
plt.xlabel("Payload Mass (kg)", fontsize=20)
plt.ylabel("Orbit Type", fontsize=20)
plt.title("Payload Mass vs Orbit Type", fontsize=20)
plt.show()
```



- Explanations: 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

[35]:
```
# Plot a line chart with x axis to be the extracted year and y axis to be the success rate

Success_rate = df.groupby('Date')['Class'].mean().reset_index()

# Plot a line chart
plt.figure(figsize=(10, 6))
sns.lineplot(x='Date', y='Class', data = Success_rate)
plt.xlabel('Year', fontsize=20)
plt.ylabel('Average Success Rate', fontsize=20)
plt.title('Average Launch Success Trend', fontsize=20)
plt.show()
```



- Explanation: The success rate since 2013 kept increasing till 2020, with a marginal drop in year 2018.

All Launch Site Names

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

```
[11]: %sql select distinct launch_site from SPACEXTABLE;

 * sqlite:///my_data1.db
Done.
[11]: ,,,,,,,,,,,,,,,,,,,
```

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

SQL result show that there are **four** unique launch sites:

1. CCAFS LC-40
2. VAFB SLC-4E
3. KSC LC-39A
4. CCAFS SLC-40

Launch Site Names Begin with 'CCA'

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

```
[17]: %sql select * from SPACEXTABLE where launch_site like 'CCA%' LIMIT 5;
```

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

[17]:

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	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	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

SQL result show that there are **five** launch records where launch sites begin with ‘CCA’, of which four launches are with customer NASA and one is with SpaceX as customer. All five launches were in the LEO orbit.

Total Payload Mass

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

```
[15]: %sql select SUM(payload_mass__kg_) as "Total payload mass from NASA(CRS) in kg" from SPACEXTABLE where customer = 'NASA (CRS)';

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

Total payload mass from NASA(CRS) in kg
45596

SQL result show that the total payload carried by boosters from NASA (CRS) was 45596 Kg.

Average Payload Mass by F9 v1.1

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

```
[16]: %sql select AVG(payload_mass__kg_) as "Average payload mass carried by booster version F9 v1.1" from SPACEXTABLE where booster_version = 'F9 v1.1';
```

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

[16]:

Average payload mass carried by booster version F9 v1.1
2928.4

SQL result show that the average payload carried by booster version F9 v1.1 was 2928.4 Kg.

First Successful Ground Landing Date

```
[18]: %sql select MIN(date) as "First successful landing" from SPACEXTABLE where landing_outcome = 'Success (ground pad)';

 * sqlite:///my_data1.db

Done.

[18]:
```

First successful landing
2015-12-22

SQL result show that the first successful landing outcome on ground pad was on 22nd December, 2015.

Successful Drone Ship Landing with Payload between 4000 and 6000

```
[21]: %sql select booster_version from SPACEXTABLE where landing_outcome = 'Success (drone ship)' and payload_mass__kg_>4000 and payload_mass__kg_<6000;

 * sqlite:///my_data1.db

Done.

[21]:
```

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

SQL result show that there are **four** boosters which have successfully landed on drone ship and have payload mass greater than 4000 but less than 6000:

1. F9 FT B1022
2. F9 FT B1026
3. F9 FT B1021.2
4. F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

List the total number of successful and failure mission outcomes

[22]:
```
%%sql
select mission_outcome, COUNT(mission_outcome) as "Total Outcomes"
from SPACEXTABLE
group by mission_outcome;
```

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

[22]:

Mission_Outcome	Total Outcomes
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

SQL result show that the total number of successful and failure mission outcomes is 100 and 1, respectively.

Boosters Carried Maximum Payload

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

[25]:
```
%%sql
select booster_version
from SPACEXTABLE
where payload_mass__kg_ = (select MAX(payload_mass__kg_) from SPACEXTABLE);
```

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

[25]:

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

SQL results show that the following boosters carried the maximum payload mass:

- 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

```
[26]: %%sql
select substr(Date, 6, 2) as "Month", booster_version, launch_site, landing_outcome
from SPACEXTABLE
where substr(Date, 0, 5) = '2015' and Landing_Outcome = 'Failure (drone ship)'

 * sqlite:///my_data1.db
Done.
[26]: ,,,,,,,,,,,,,,,,,,,,,,
```

Month	Booster_Version	Launch_Site	Landing_Outcome
01	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
04	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

The failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015, as obtained from SQL query are:

Month	Booster_Version	Launch_Site	Landing_Outcome
01	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
04	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

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

[19]:
```
%%sql

select Landing_Outcome, count(Landing_Outcome) as "Count of Landing Outcomes"
from SPACEXTABLE
where date between '2010-06-04' and '2017-03-20'
group by Landing_Outcome
order by "Count of Landing Outcomes" desc
```

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

[19]:

Landing_Outcome	Count of Landing Outcomes
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

SQL result ranking the count of landing outcomes between the date 2010-06-04 and 2017-03-20, in descending order:

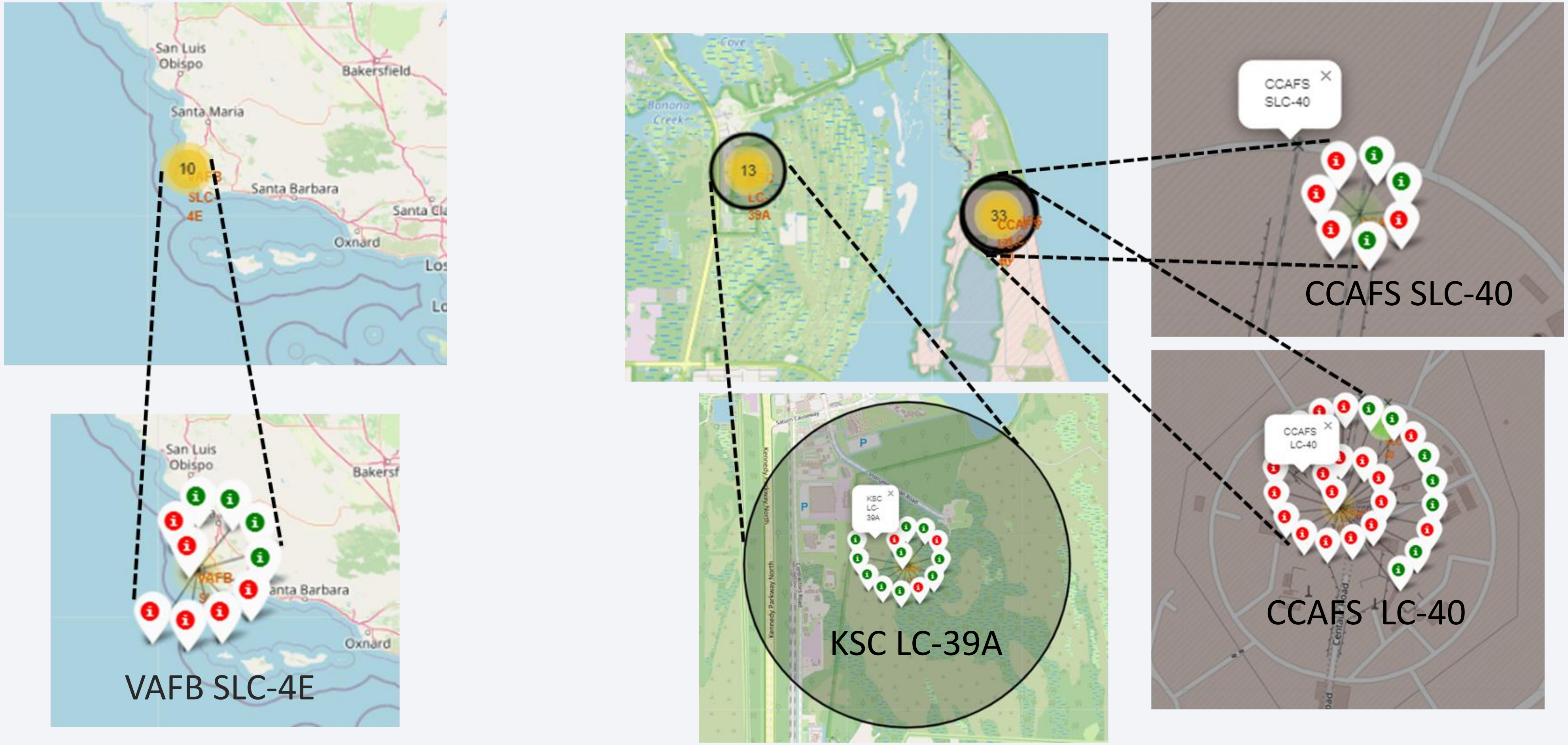
Landing _Outcome	**Count of landing outcomes**
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1


Section 3
Launch Sites
Proximities Analysis

Launch Site Locations

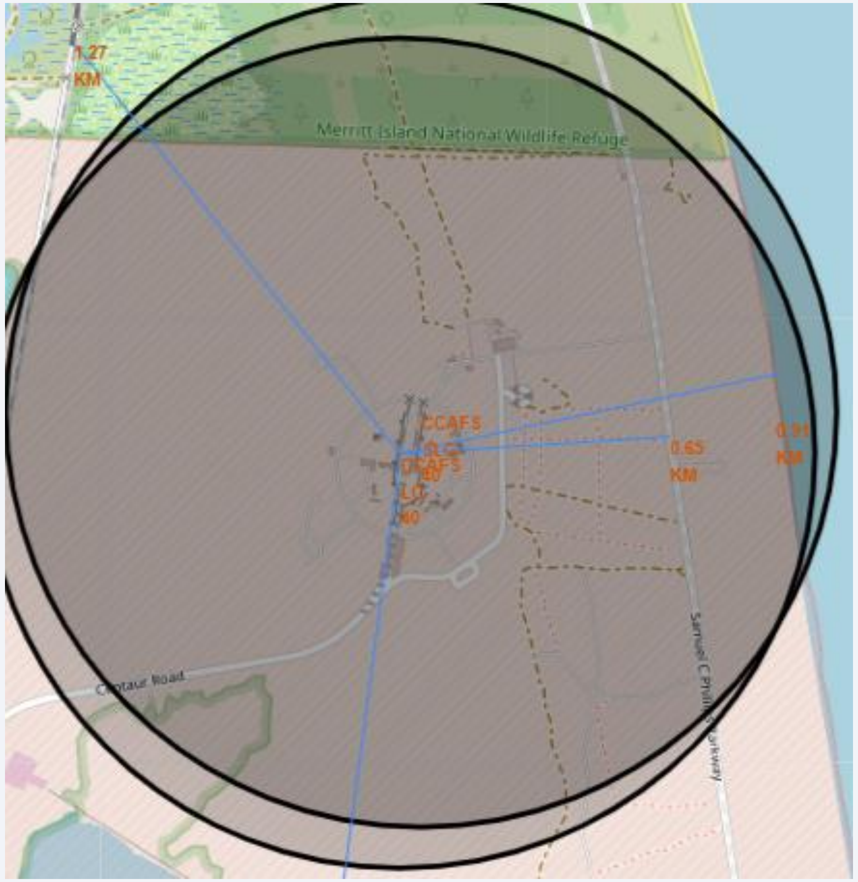
All launch sites' location as marked on a global map: Site VAFB SLC-4E is located in the west (Los Angeles) while CCAFS LC-40, KSC LC-39A and CCAFS SLC-40 are in the east (Florida).

Color-labeled Launch Outcomes



Explanation: Each green marker on the map represents a successful launch while each red marker represents a failed launch.

Distance from CCAFS LC-40



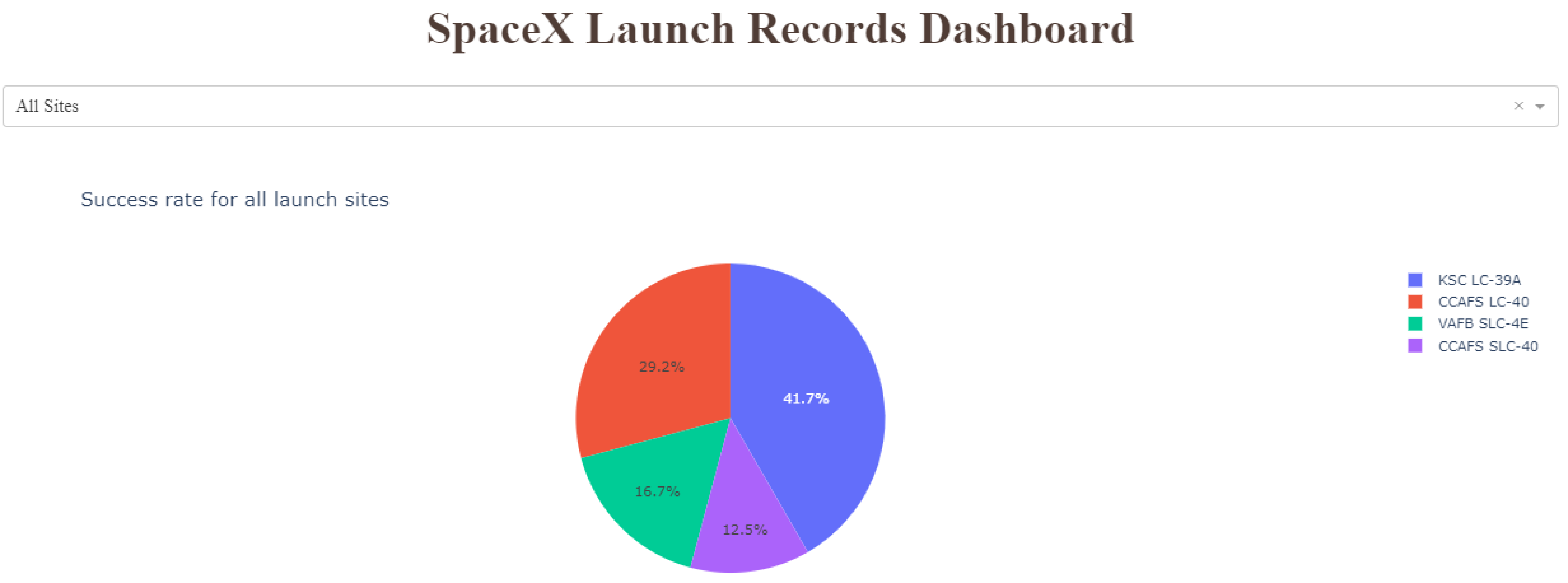
From the Launch site CCAFS LC-40
{coordinates = 28.562302, -80.577356}

- The nearest railway is about 1.27 Km
- The nearest highway is ca. 0.65 Km
- The nearest coastline is around 0.91 Km
- The nearest city is around 51.5 Km

Section 4

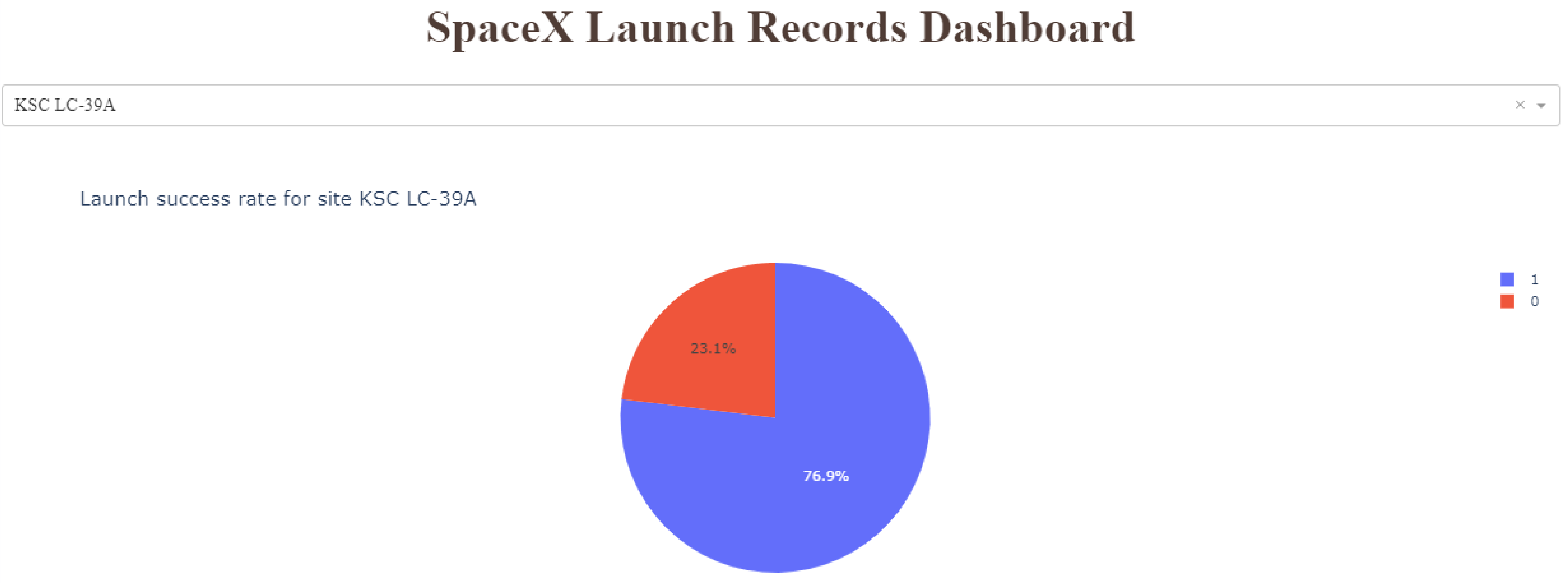
Build a Dashboard with Plotly Dash

Total Successful Launches For All Sites



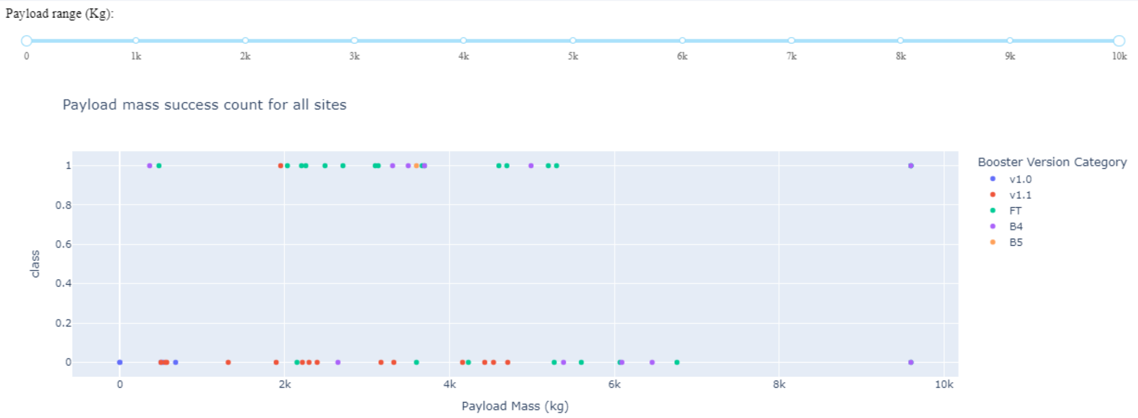
- The rendered pie chart shows that of all sites, KSC LC-39A has the highest success count of 41.7%

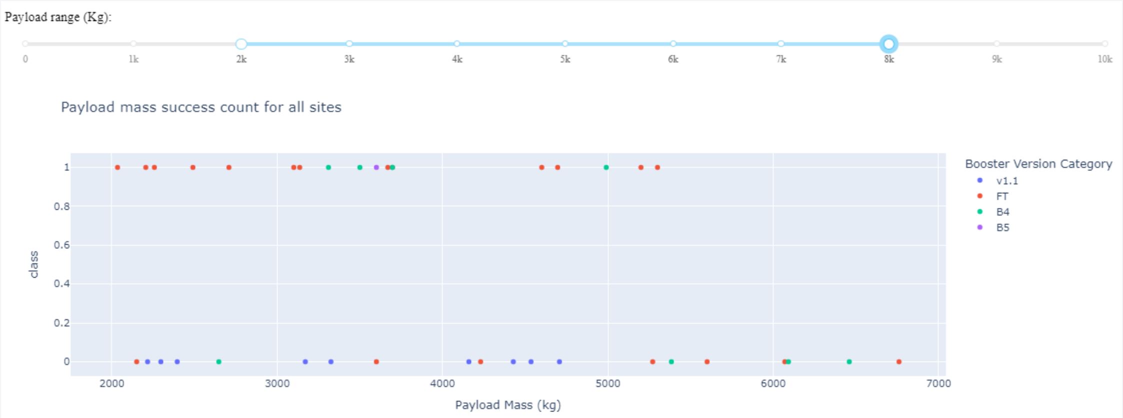
Launch Site With Highest Launch Success Ratio

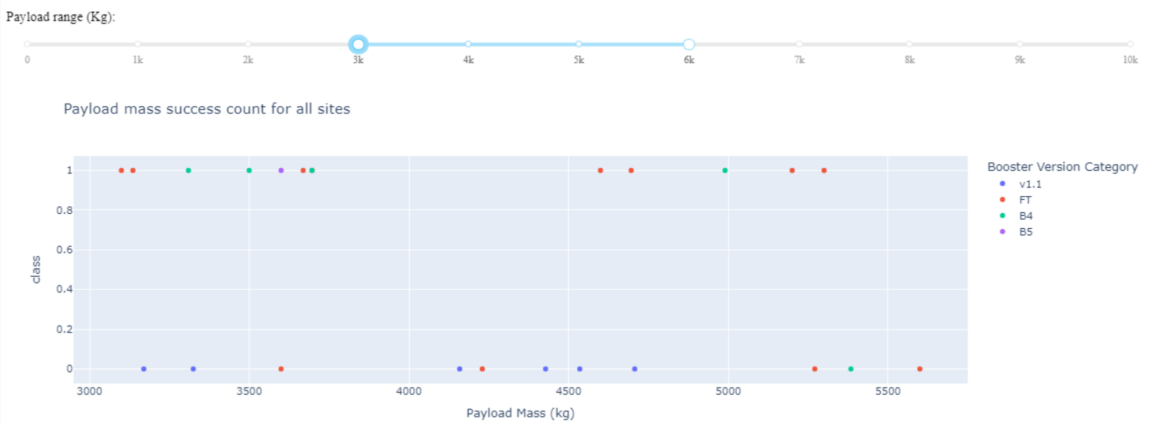


KSC LC-39A is the launch site with highest launch success ratio. The rendered pie chart shows that it has a success rate of **76.9%**.

Payload vs. Launch Outcome For All Sites







The rendered scattered plot show that the booster version FT has the largest success rate.

Section 5

Predictive Analysis (Classification)

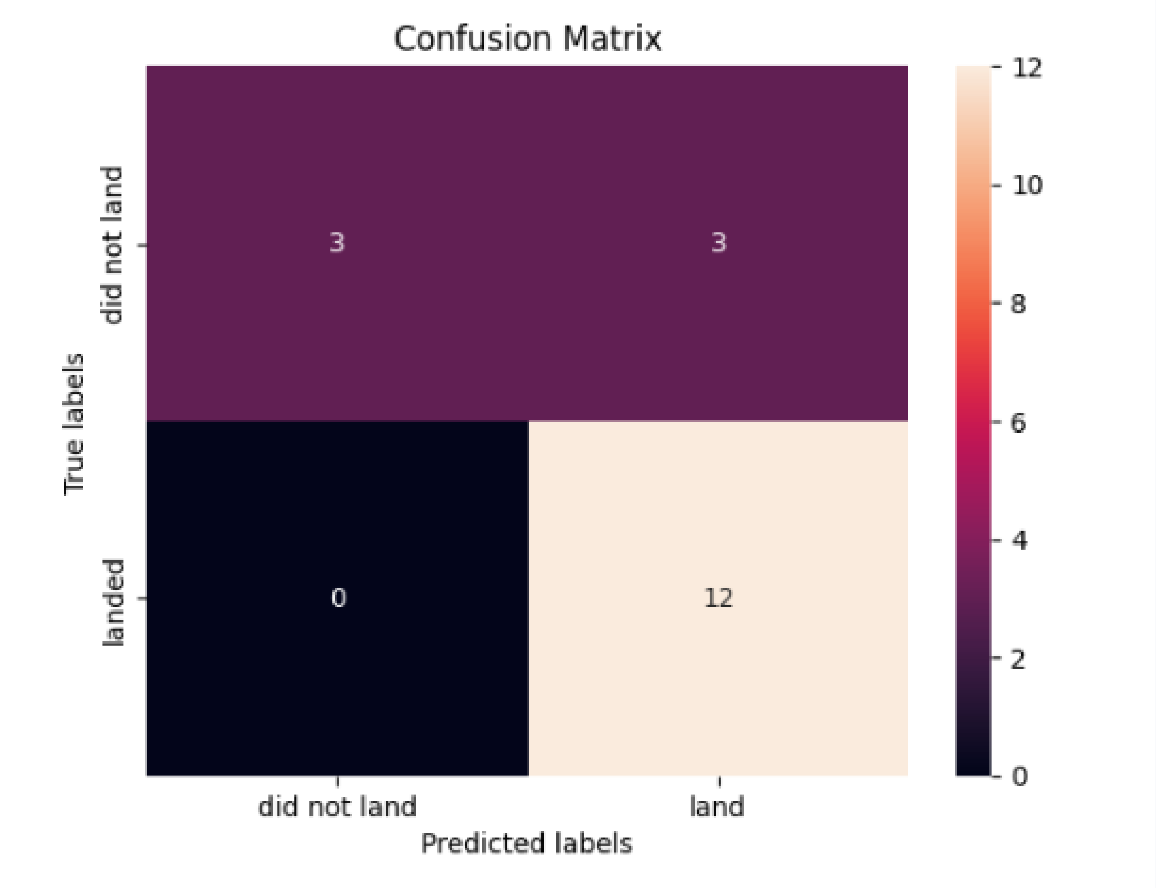
Classification Accuracy

Modelling Algorithm	Train Accuracy Score	Test Accuracy Score
Logistic Regression	0.846429	0.833333
SVM	0.848214	0.833333
Decision Trees	0.873214	0.833333
KNN	0.848214	0.833333

- GridSearchCV was used to test parameters of classification algorithms and find the best one
- For all classification model algorithms, the accuracy on the validation data was obtained by using the data attribute best_score_
- Accuracy on the test data was obtained using the method score

All methods showed same classification accuracy on the test data. However, as the decision tree method fits with slightly higher classification accuracy on the train data set, it was identified as the best performing model

Confusion Matrix



Confusion matrix for the decision tree model

- Confusion matrix for the decision tree model can distinguish between the different classes.
- Major problem is false positives.

Conclusions

Using exploratory data analysis (EDA), interactive visual analytics and predictive analysis methodologies, following can be concluded regarding Falcon 9 launches:

- Launch sites are in close proximity to coastline to allow for controlled landing in the ocean to be performed. They are also in close proximity to the highways and railways to allow for easy transport of cargo and far away from the cities to avoid danger to people and property.
- The success rate since 2013 kept increasing till 2020, with first successful landing outcome in ground pad achieved on 22nd Dec, 2015. A total number of 100 successful and 1 failure mission outcomes have taken place.
- Different launch sites have different success rates. Although there are less launches in VAFB SLC 4E and KSC LC 39A, higher success rate can be seen in these two sites. CCAFS LC-40, has a success rate of 60 %, while KSC LC-39A and VAFB SLC 4E has a success rate of 77%.
- As flight number increases, there are more successful first stage landing. With small flight numbers, launches happens more in the site CCAFS SLC 40 and with much lower success rate.
- With higher Payload the success rate is much higher. Furthermore, with Payload more than 9500, we can see very high success rate overall.
- In KSC LC39A launchsite we can see much higher success rate with low Payload whereas this rate is much lower in CCAFS SLC 40 launchsite. For the VAFB-SLC launchsite, there are no rockets launched for heavy payload mass(greater than 10000).
- In ES-L1, GEO, HEO, and SSO orbits, all launches are successful. In SO orbit, the rate is zero.
- With heavy payloads the successful landing or positive landing rate are more for Polar, LEO and ISS.
- Of all classification model algorithms used by us, Decision Tree is the best model for predicting landing outcome of rocket. It can distinguish between the different classes but its major problem is false positives.

Thank you!