IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Mukesh Chapagain
08/29/2025

Outline

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

Executive Summary

- **Summary of methodologies**
  - Data Collection - with API & Web Scraping
  - Data Wrangling
  - Exploratory Data Analysis (EDA) - with SQL & Visualization
  - Interactive Visual Analytics (with Folium) and Dashboard (with Plotly Dash)
  - Predictive Analysis (Classification)
- **Summary of all results**
  - EDA Results
  - Interactive Visual Analytics Results
  - Predictive Analytics Results

Introduction

- **Project background and context**
  - Space X advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because Space X can reuse the first stage.
  - Therefore if we can determine if the first stage will land, we can determine the cost of a launch.
  - This information can be used if an alternate company wants to bid against space X for a rocket launch.
  - The goal of the project is to create a machine learning pipeline to predict if the first stage will land given the history of the launch data.
- **Problems you want to find answers**
  - Find the best parameter influencing the launch.
  - Find the best machine learning model which can predict if the first stage will land successfully.

Section 1

Methodology

Methodology

Executive Summary

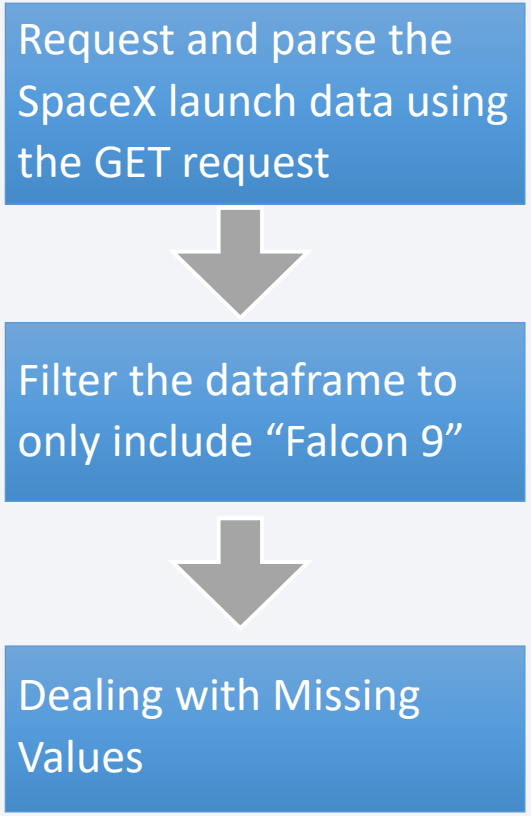
- Data collection methodology
- Perform data wrangling
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

Data Collection

- Data were collected from SpaceX API endpoints and from Wikipedia Page.
  - SpaceX API Endpoints
    - https://api.spacexdata.com/v4/rockets/
    - https://api.spacexdata.com/v4/launchpads/
    - https://api.spacexdata.com/v4/payloads/
    - https://api.spacexdata.com/v4/cores/
    - https://api.spacexdata.com/v4/launches/past
  - Wikipedia Page
    - https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches

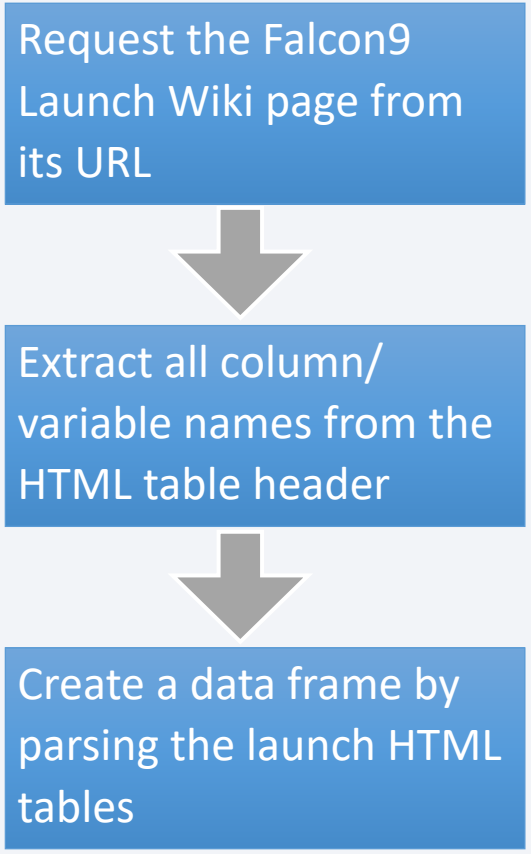
Data Collection – SpaceX API

- Data collection with SpaceX REST calls using key phrases and flowcharts —>
- GitHub URL of the completed SpaceX API calls notebook:
  - https://github.com/chapagain/data-science-practice/blob/main/applied-data-science-capstone/jupyter-labs-spacex-data-collection-api.ipynb



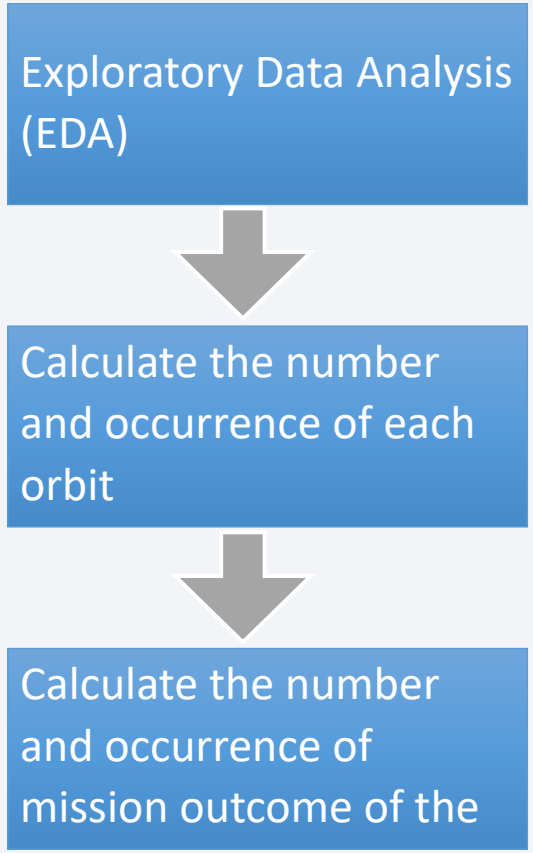
Data Collection - Scraping

- The data was scraped from Wikipedia Page.
- GitHub URL of the completed web scraping notebook:
  - https://github.com/chapagain/data-science-practice/blob/main/applied-data-science-capstone/jupyter-labs-webscraping.ipynb



Data Wrangling

- Some Exploratory Data Analysis (EDA) were performed to find some patterns in the data and determine what would be the label for training supervised models.
- GitHub URL of the completed data wrangling related notebook:
  - https://github.com/chapagain/data-science-practice/blob/main/applied-data-science-capstone/labs-jupyter-spacex-Data%20wrangling.ipynb



EDA with Data Visualization

- EDA & Data Visualization was done with Scatter plot, Bar chart, and Line chart.
    1. Visualize the relationship between *Flight Number* and *Launch Site*
    2. Visualize the relationship between *Payload Mass* and *Launch Site*
    3. Visualize the relationship between *Success rate* of each *Orbit type*
    4. Visualize the relationship between *FlightNumber* and *Orbit type*
    5. Visualize the relationship between *Payload Mass* and *Orbit type*
    6. Visualize the launch success yearly trend
- GitHub URL of the completed EDA with data visualization notebook:
    - https://github.com/chapagain/data-science-practice/blob/main/applied-data-science-capstone/edadataviz.ipynb

EDA with SQL

- List of the SQL queries performed:
  - Display the names of the unique launch sites in the space mission
  - Display 5 records where launch sites begin with the string ‘CCA'
  - Display the total payload mass carried by boosters launched by NASA (CRS)
  - Display average payload mass carried by booster version F9 v1.1
  - List the date when the first succesful landing outcome in ground pad was acheived.
  - List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
  - List the total number of successful and failure mission outcomes
  - List all the booster_versions that have carried the maximum payload mass, using a subquery with a suitable aggregate function. %
  - 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.
  - 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.
- GitHub URL of the completed EDA with SQL notebook:
  - https://github.com/chapagain/data-science-practice/blob/main/applied-data-science-capstone/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

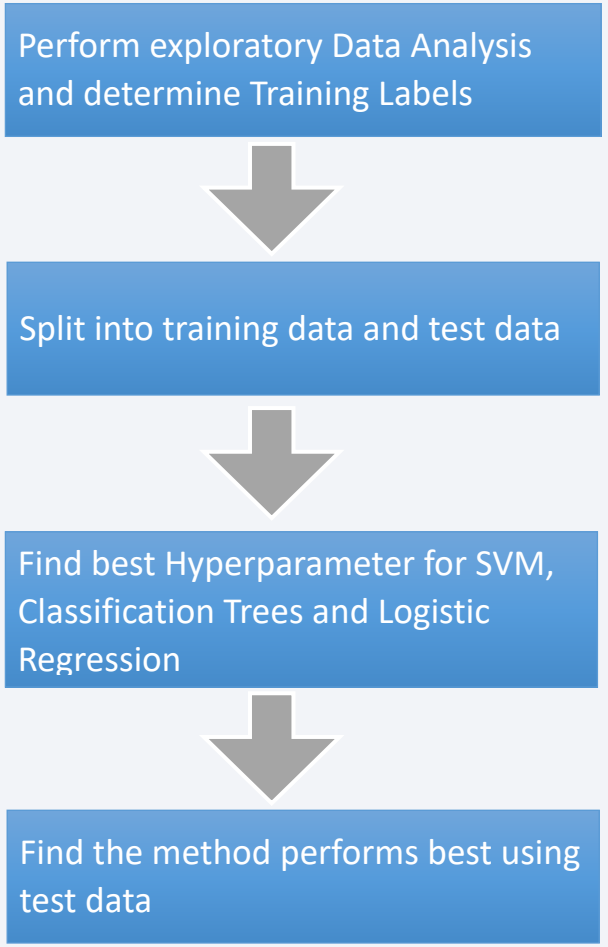
- We added Markers, Circles, PolyLines & MarkerClusters to the Folium Map.
- The launch success rate may also depend on the location and proximities of a launch site, i.e., the initial position of rocket trajectories. Hence, this analysis with interactive map with Folium was done.
- GitHub URL of the completed interactive map with Folium map:
  - https://github.com/chapagain/data-science-practice/blob/main/applied-data-science-capstone/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- An interactive dashboard was developed using Plotly Dash.
  - Added a dropdown list to enable Launch Site selection
  - Added a pie chart to show the total successful launches count for all sites
  - Added a slider to select payload range
  - Added a scatter chart to show the correlation between payload and launch success
- Explain why you added those plots and interactions
- GitHub URL of the completed Plotly Dash lab:
  - https://github.com/chapagain/data-science-practice/blob/main/applied-data-science-capstone/spacex-dash-app.py

Predictive Analysis (Classification)

- Summary of Predictive Analysis:
  - Imported data using Pandas and Numpy libraries.
  - Created ‘Class’ column and normalized the data.
  - Created Training and Testing data set.
  - Built different machine learning models (logistic regression, SVM, Decision Tree, KNN) and adjusted their hyperparameters using GridSearchCV.
  - Calculated accuracy of all the models to figure out the best model with the highest accuracy score.
- GitHub URL of the completed predictive analysis lab:
  - https://github.com/chapagain/data-science-practice/blob/main/applied-data-science-capstone/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb



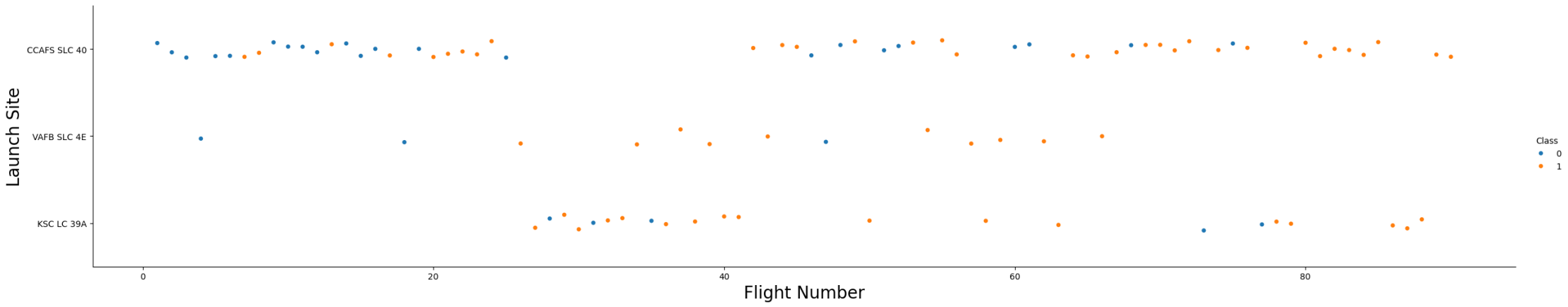
Results

- Exploratory data analysis results
  - The success rate since 2013 kept increasing till 2020
  - With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS Orbits.
  - The LEO orbit, success seems to be related to the number of flights. Conversely, in the GTO orbit, there appears to be no relationship between flight number and success.
  - The first successful landing outcome in ground pad was achieved in 2015-12-22.
- Interactive analytics demo in screenshots
  - The findings were that the launch sites were located in safe areas and far from cities.
  - The launch sites are located near highways and railroads for the transport of needed facilities and people.
- Predictive analysis results
  - The classification accuracy of all the models analyzed had same accuracy score.

Section 2

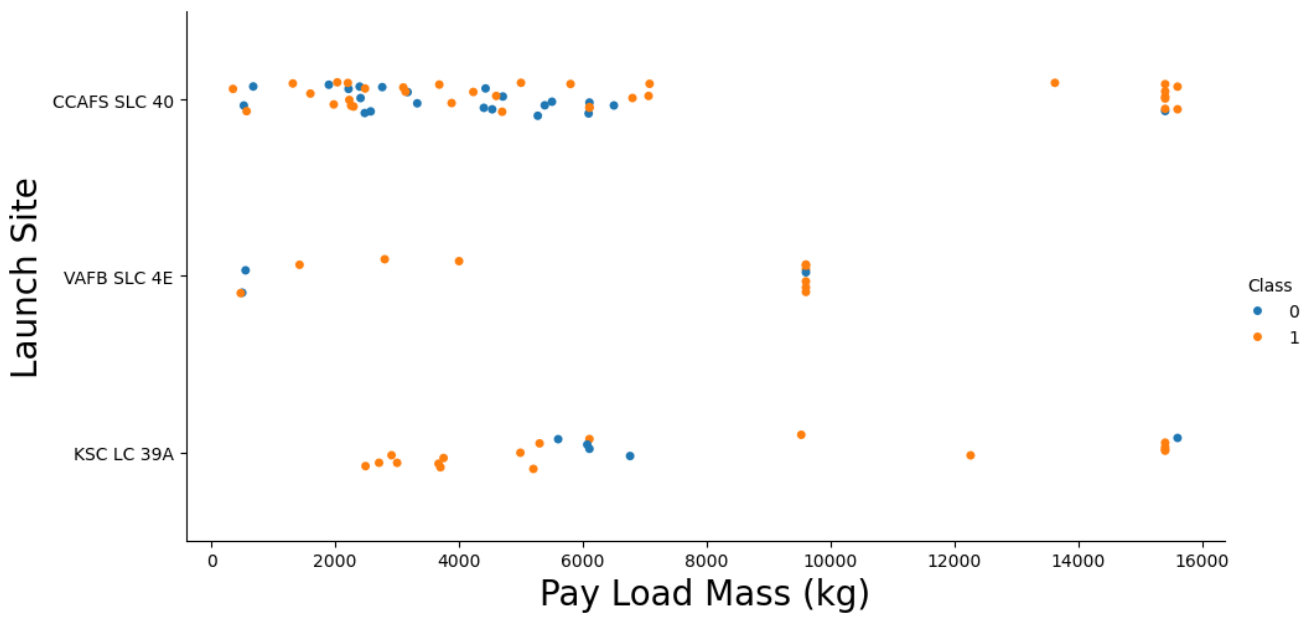
Insights drawn from EDA

Flight Number vs. Launch Site



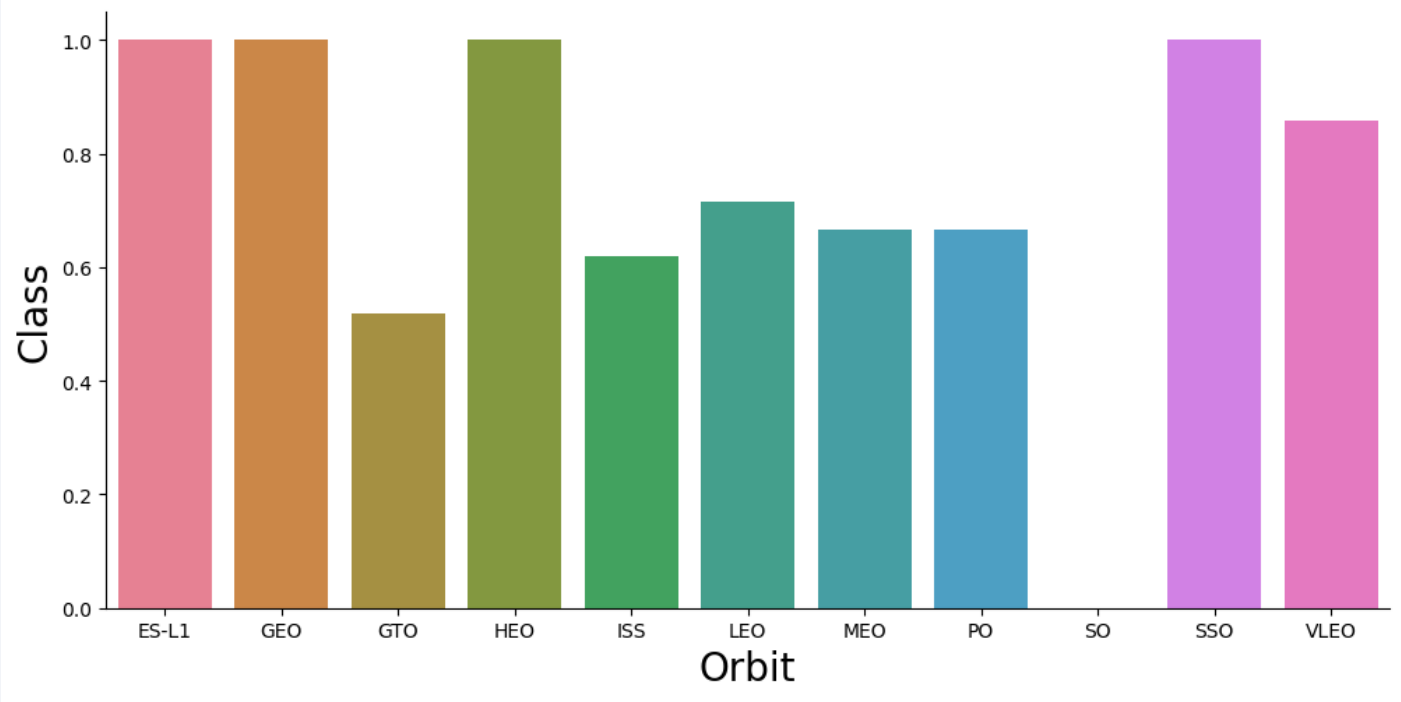
- Scatter plot of Flight Number vs. Launch Site
- The probability of success outcome is highest for VAFB SLC 4E followed by KSC LC 39A. The success outcome of CCAFS SLC 40 is the lowest among the three.

Payload vs. Launch Site



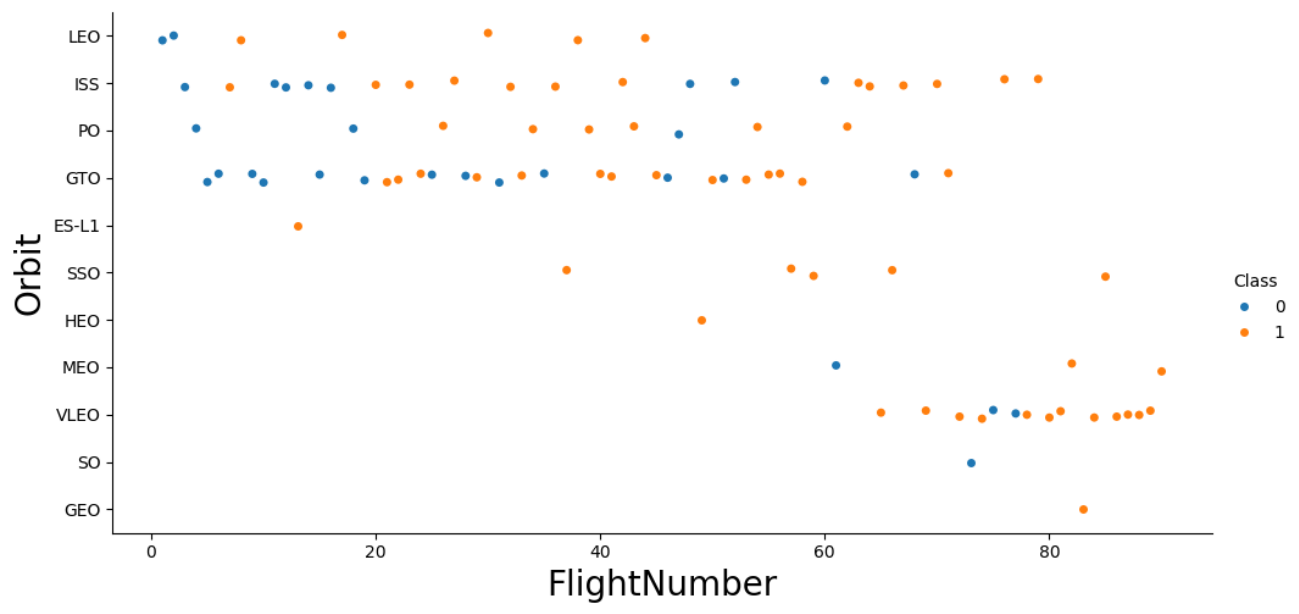
- Scatter plot of Payload vs. Launch Site
- For the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000).

Success Rate vs. Orbit Type



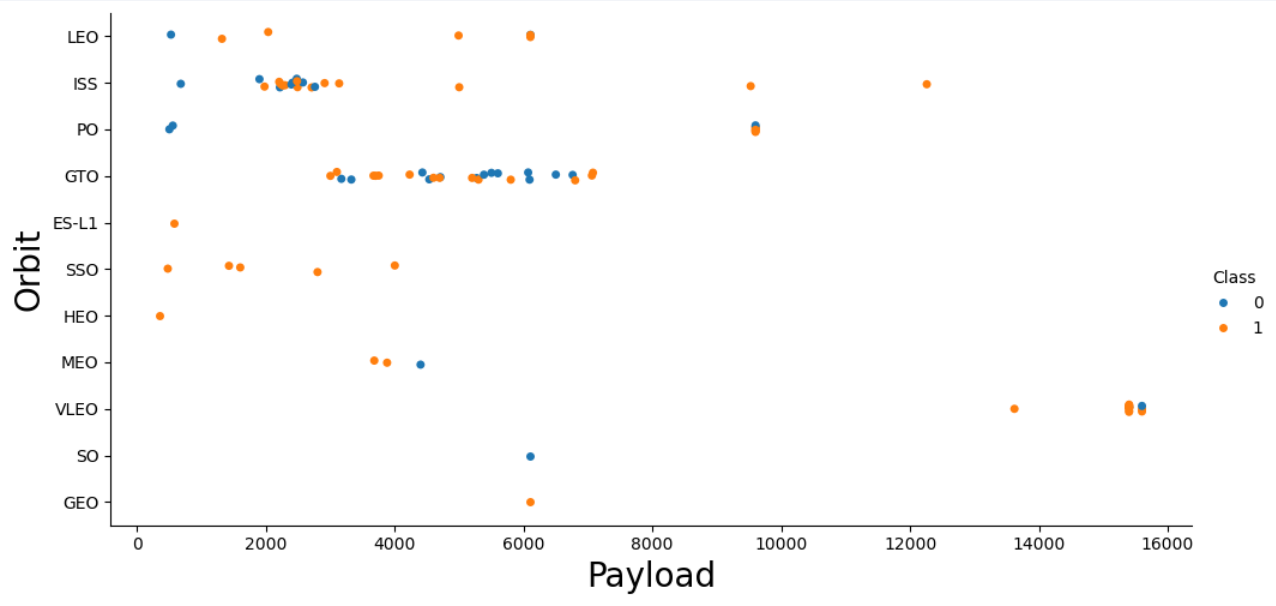
- Bar chart for the success rate of each orbit type
- The orbits with the highest success rates are ES-L1, GEO, HEO, and SSO.

Flight Number vs. Orbit Type



- Scatter point of Flight number vs. Orbit type
- In the LEO orbit, success seems to be related to the number of flights. Conversely, in the GTO orbit, there appears to be no relationship between flight number and success.

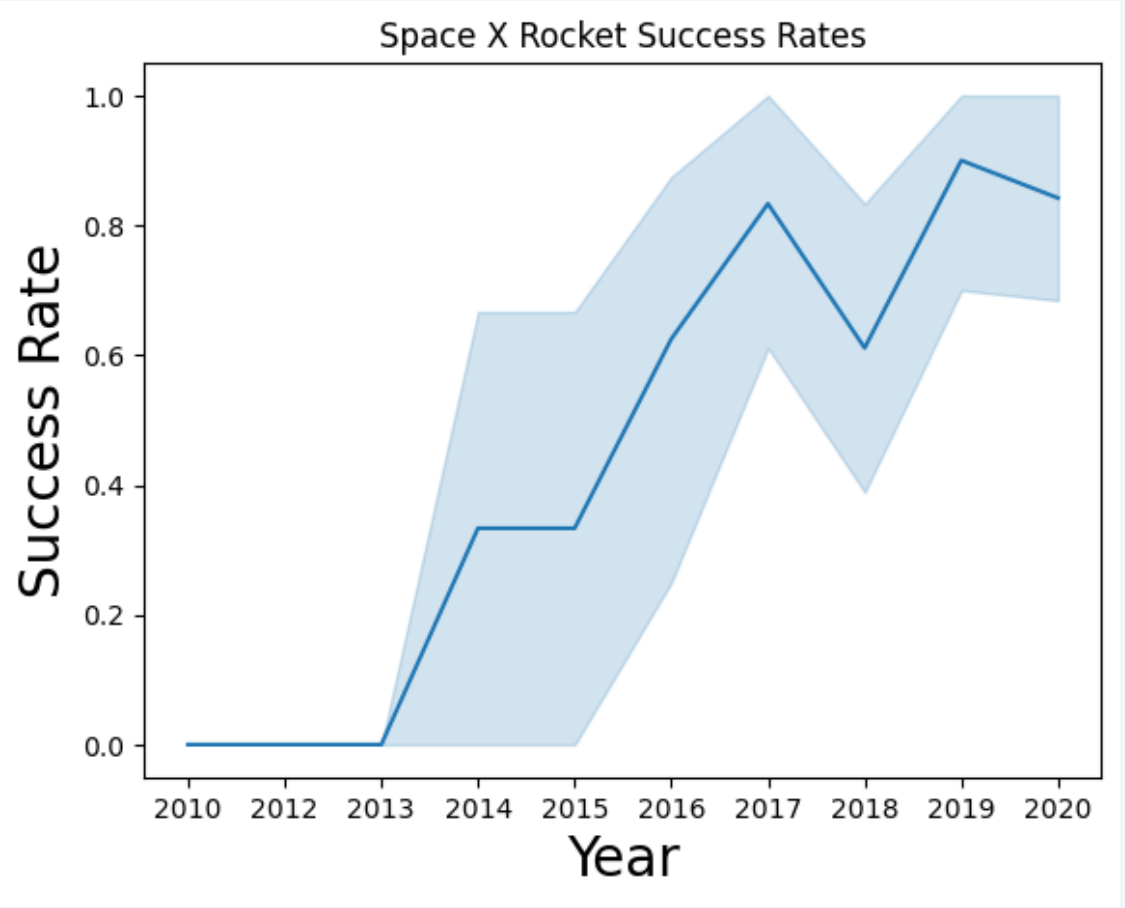
Payload vs. Orbit Type



- Scatter point of payload vs. orbit type
- With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS. However, for GTO, it's difficult to distinguish between successful and unsuccessful landings as both outcomes are present.

Launch Success Yearly Trend

- Line chart of yearly average success rate
- The success rate since 2013 kept increasing till 2020.



All Launch Site Names

- “DISTINCT” SQL Query was used to get the names of all unique launch site names.
- Display the names of the unique launch sites in the space mission

```
In [10]: %sql SELECT DISTINCT Launch_Site FROM SPACEXTBL
 * sqlite:///my_data1.db
Done.
Out[10]:
```

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

Launch Site Names Begin with 'CCA'

- “LIKE” query was used to get the launch sites names beginning with ‘CCA’
- Display 5 records where launch sites begin with the string 'CCA'

```
In [11]: %sql SELECT Launch_Site FROM SPACEXTBL WHERE Launch_Site LIKE 'CCA%' LIMIT 5
 * sqlite:///my_data1.db
Done.
```

Out[11]:

Launch_Site
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40

Total Payload Mass

- SUM() function is used to calculate the total payload.
-

Task 3

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

```
In [12]: %sql SELECT Customer, SUM(PAYLOAD_MASS__KG_) Total_Mass FROM SPACEXTBL GROUP BY Customer HAVING Customer = 'NASA
```

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

Out[12]:

Customer	Total_Mass
NASA (CRS)	45596

Average Payload Mass by F9 v1.1

- AVG() function was used to calculate the average payload mass.
-

Task 4

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

```
In [13]: %sql SELECT Booster_Version, AVG(PAYLOAD_MASS__KG_) FROM SPACEXTBL GROUP BY Booster_Version HAVING Booster_Versic
 * sqlite:///my_data1.db
Done.
```

Out[13]:

Booster_Version	AVG(PAYLOAD_MASS__KG_)
F9 v1.1	2928.4

First Successful Ground Landing Date

- MIN() function is used to find the first successful landing date.
-

Task 5

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

Hint:Use min function

```
In [14]: %sql SELECT min(Date) AS Start_Date, Landing_Outcome FROM SPACEXTBL GROUP BY Landing_Outcome HAVING Landing_Outcc
 * sqlite:///my_data1.db
Done.
```

Out[14]:

Start_Date	Landing_Outcome
2015-12-22	Success (ground pad)

Successful Drone Ship Landing with Payload between 4000 and 6000

- DISTINCT query with WHERE clause was used to get the result.

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

In [15]:
```
%sql SELECT DISTINCT Booster_Version, Landing_Outcome, PAYLOAD_MASS__KG_ from SPACEXTBL \
WHERE (Landing_Outcome = 'Success (drone ship)') AND (PAYLOAD_MASS__KG_ between 4000 and 6000)
```

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

Out[15]:

Booster_Version	Landing_Outcome	PAYLOAD_MASS__KG_
F9 FT B1022	Success (drone ship)	4696
F9 FT B1026	Success (drone ship)	4600
F9 FT B1021.2	Success (drone ship)	5300
F9 FT B1031.2	Success (drone ship)	5200

Total Number of Successful and Failure Mission Outcomes

- COUNT function and GROUP BY clause were used to get the result.
- List the total number of successful and failure mission outcomes

```
In [16]: %sql SELECT Mission_Outcome, COUNT(Mission_Outcome) AS 'Total' FROM SPACEXTBL GROUP BY Mission_Outcome
 * sqlite:///my_data1.db
Done.
```

Out[16]:

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

Boosters Carried Maximum Payload

- Subquery with MAX() aggregate function was used to get the result.
- List all the booster_versions that have carried the maximum payload mass, using a subquery with a suitable aggregate function.

```
In [19]: %sql SELECT Booster_Version FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTE
 * sqlite:///my_data1.db
Done.
```

Out[19]:

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

- substr() function was used to get the result.
- 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.

  Note: SQLLite does not support monthnames. So you need to use substr(Date, 6,2) as month to get the months and substr(Date,0,5)='2015' for year.

```
In [22]: %sql SELECT substr(Date, 6, 2) AS Month, Booster_Version, Launch_Site, Landing_Outcome FROM SPACEXTBL \
         WHERE Landing_Outcome = 'Failure (drone ship)' AND substr(Date, 0, 5) = '2015'
```

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

Out[22]:

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

- COUNT, BETWEEN query, GROUP BY & ORDER BY clauses were used to obtain the result.
- 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.

```
In [24]: %sql SELECT Landing_Outcome, Count(*) Total_Count FROM SPACEXTBL \
         WHERE Date BETWEEN '2010-06-04' AND '2017-03-20' \
         GROUP BY Landing_Outcome \
         ORDER BY Total_Count DESC;

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

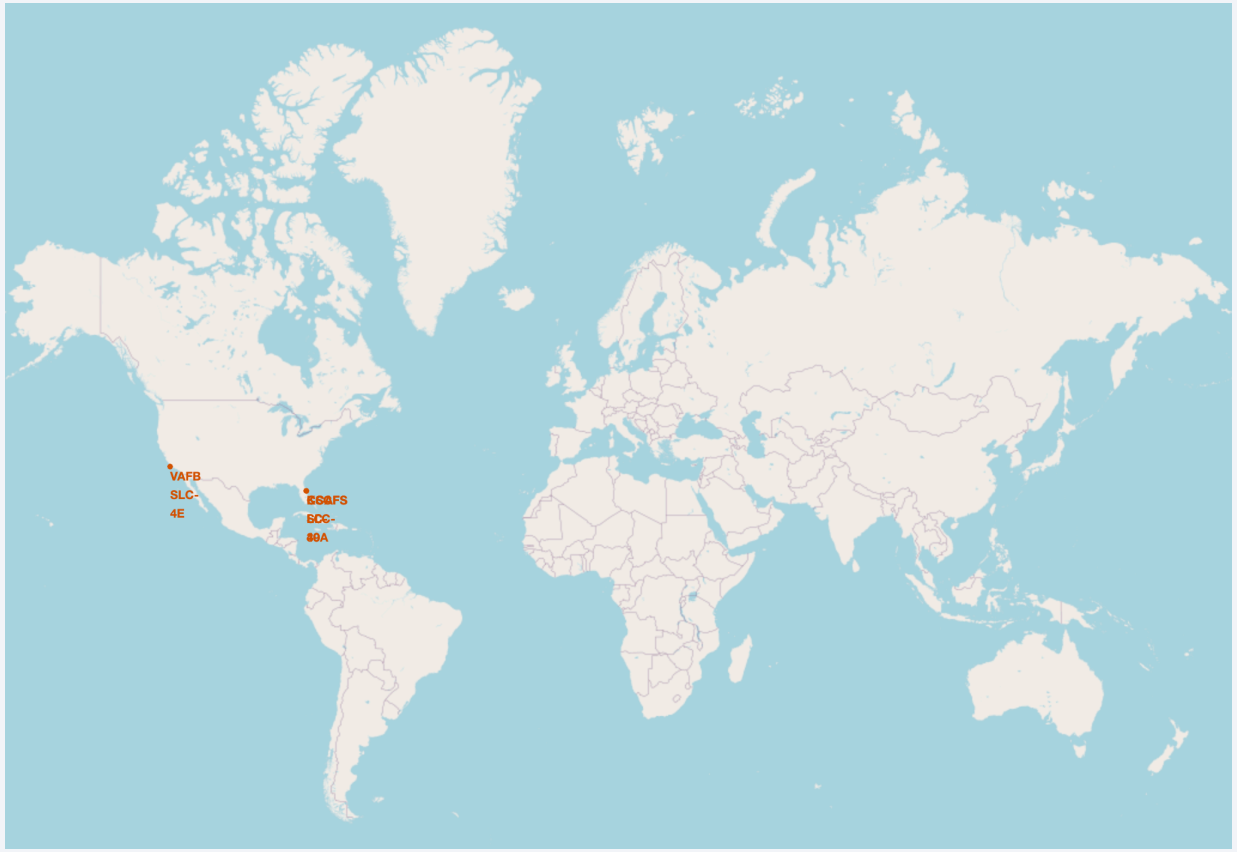
Out[24]:

Landing_Outcome	Total_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

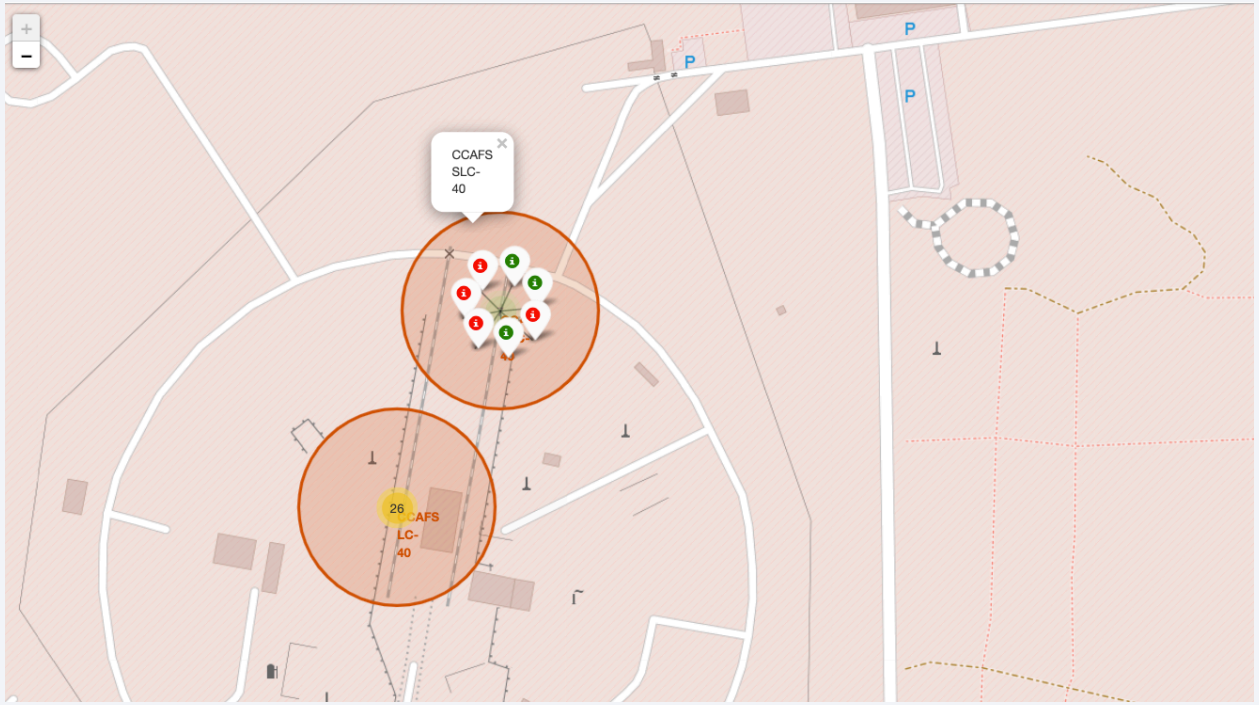
Section 3

Launch Sites Proximities Analysis

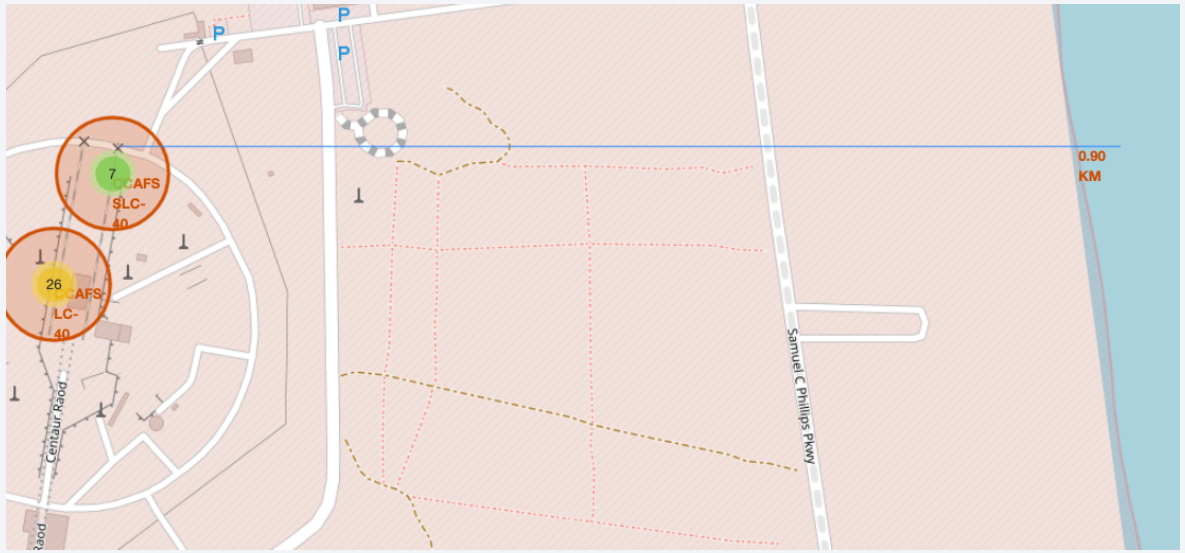
All Launch Sites on a Global Map

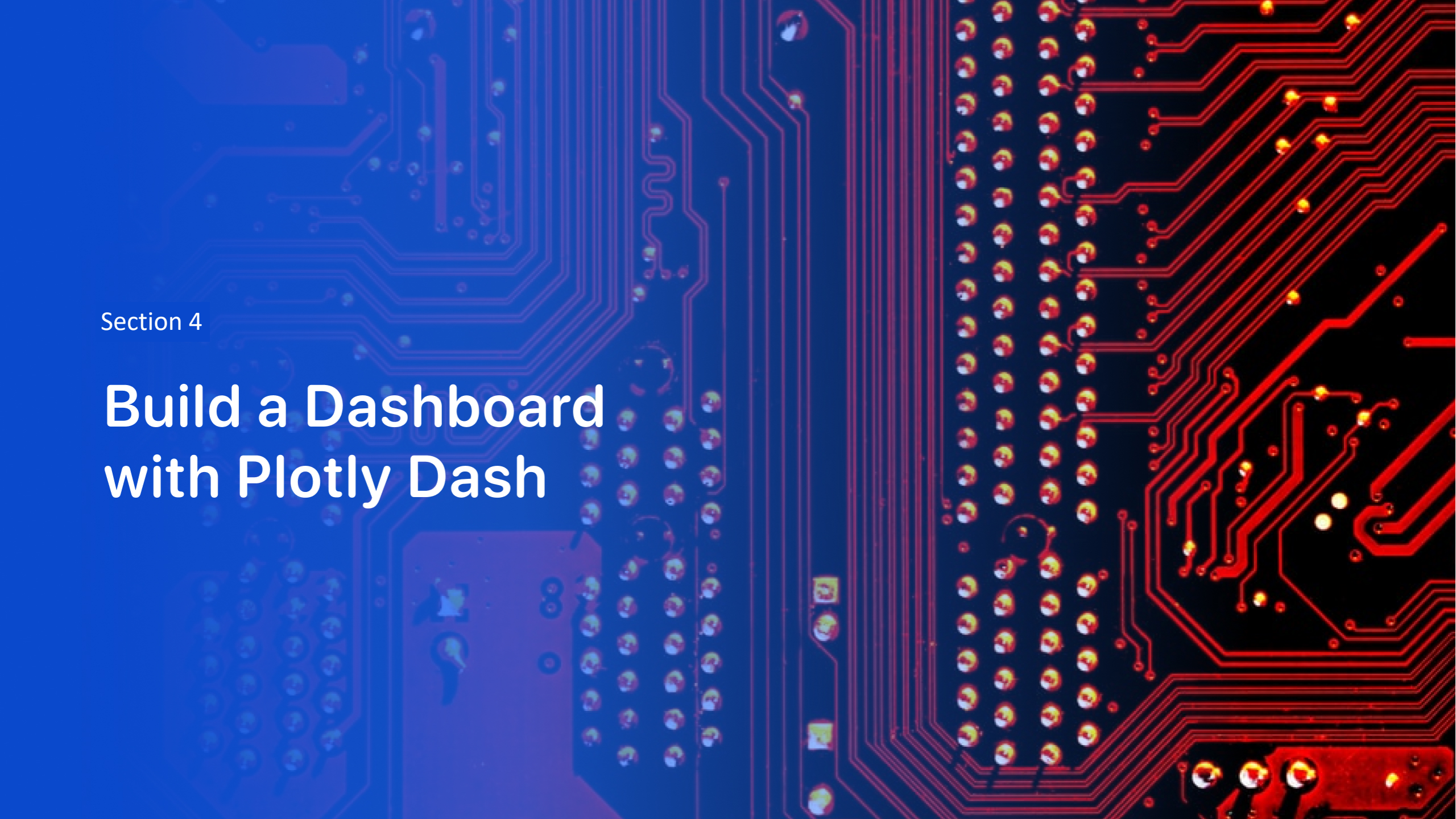
- All launch sites are located in the coasts of the USA.
- 

Color-labeled Launch Markers

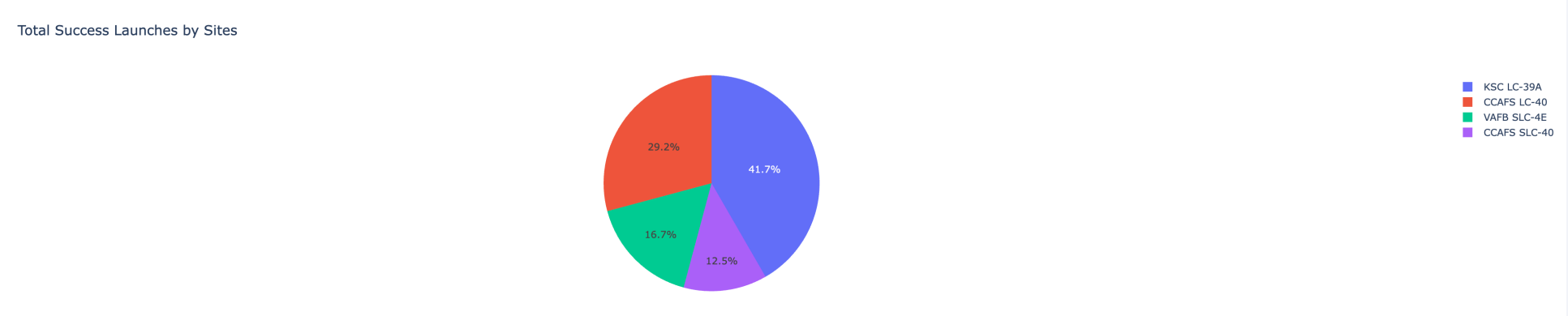
- Green Marker = Successful Launch, Red Marker = Failed Launch
- 

Distance of Launch Site to its Proximities

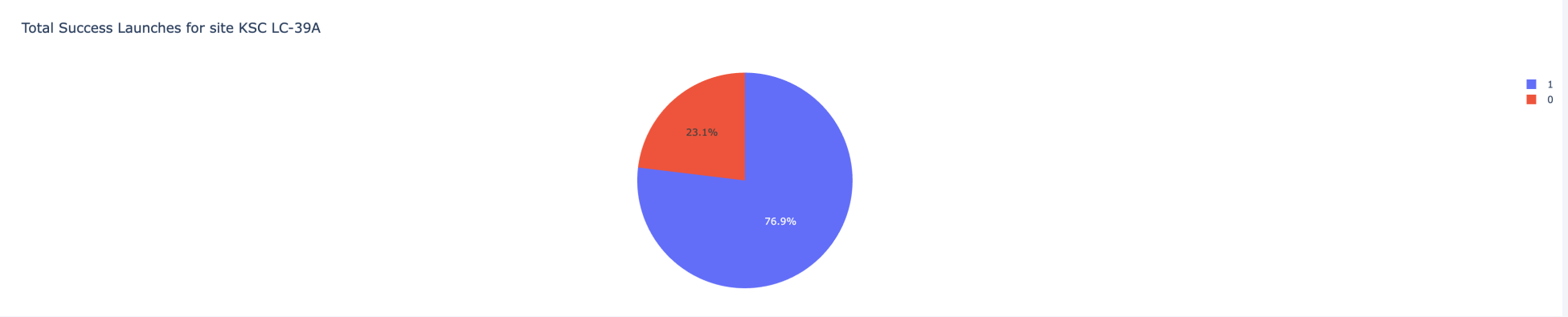
- Launch sites are close to coastline, railway and highway but far from cities.
- 


Section 4
Build a Dashboard with Plotly Dash

Launch success of all sites

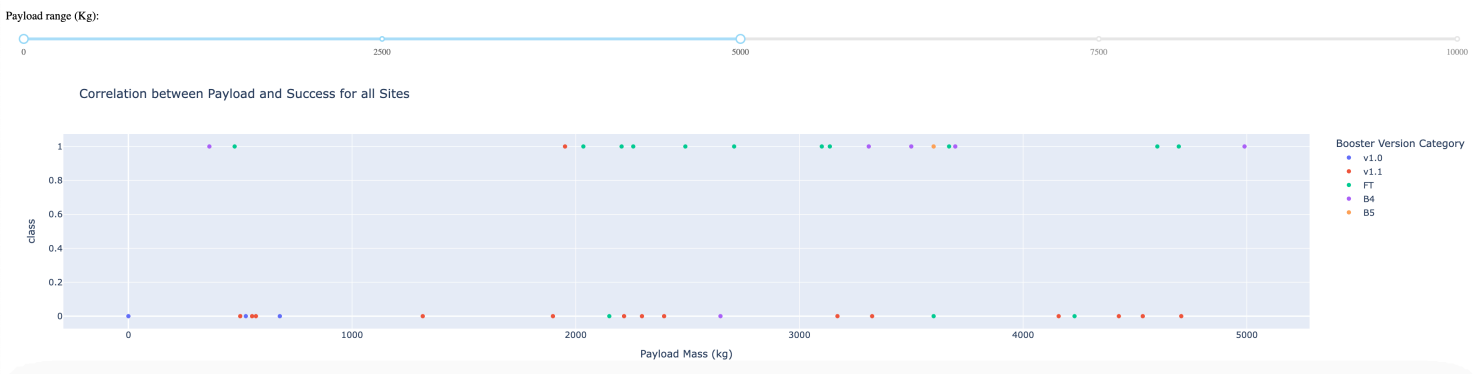
- KSC LC-39A has the most successful launches
- VAFB SLC-4E has the least successful launches
- 

Launch site with highest success ratio

- KSC LC-39A has the launch success ratio of 76.9%
- 

Payload vs. Launch Outcome for all sites

- Light Payload have higher success rate than Heavy Payload.
- FT booster version has the largest success rate.
- Light Payload



- Heavy Payload

Section 5

Predictive Analysis (Classification)

Classification Accuracy

```
In [43]: df_report = pd.DataFrame({'Method': ['Log Reg', 'SVM', 'Tree', 'KNN'],
                                   'Best Accuracy': [logreg_cv.best_score_, svm_cv.best_score_, tree_cv.best_score_, knn_cv.b
                                   'Accuracy': [lr_score, svm_score, tree_score, knn_score]})
         df_report
```

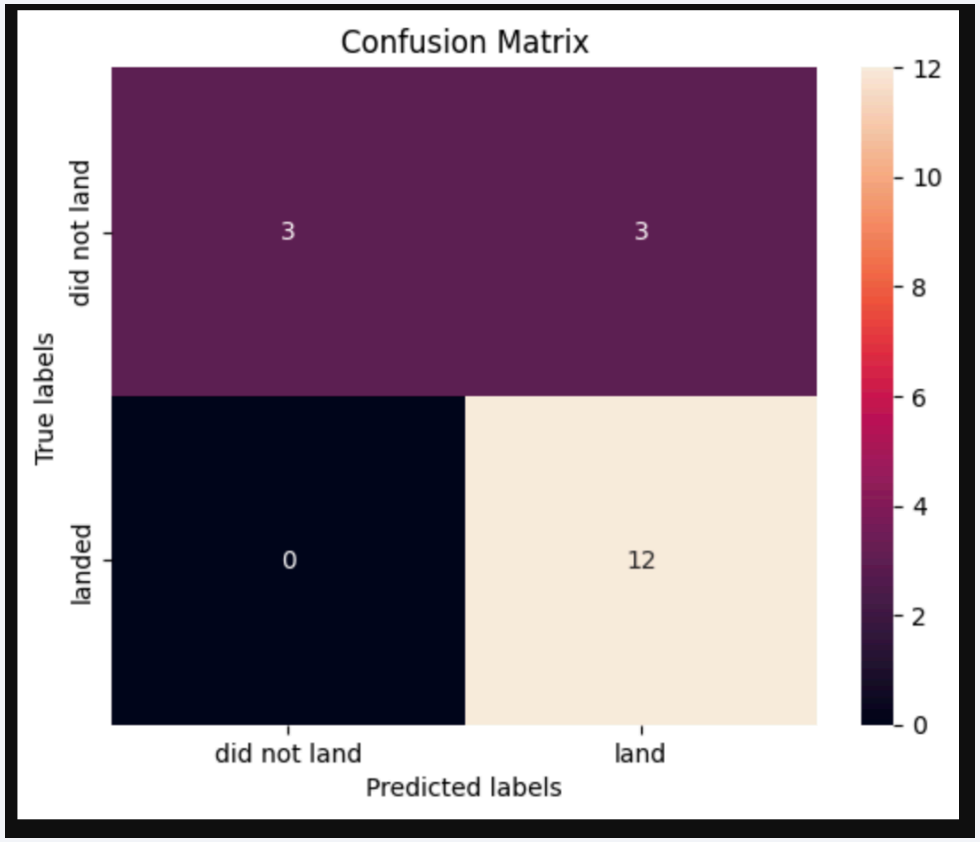
Out[43]:

	Method	Best Accuracy	Accuracy
0	Log Reg	0.846429	0.833333
1	SVM	0.848214	0.833333
2	Tree	0.864286	0.833333
3	KNN	0.848214	0.833333

- All models have the same prediction accuracy score.
- However, Decision Tree Classifier had the best_score_ when tested with the best_params_.

Confusion Matrix

- The prediction score was same for all the 4 models.
- Hence, their confusion matrix is also the same.



Conclusions

- The launch success rate since 2013 kept increasing till 2020.
- The first successful landing outcome in ground pad was achieved in 2015-12-22.
- Launch sites are close to coastline but far from cities.
- KSC LC-39A has the most successful launches whereas VAFB SLC-4E has the least successful launches.
- Light Payload have higher success rate than Heavy Payload.
- FT booster version has the largest success rate.
- All classifiers had the same accuracy score.

Thank you!