IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Freddy Montiel
06/18/2025

OUTLINE

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

Executive Summary

Summary of methodologies

- Data Collection
- Data Wrangling
- Exploratory Data Analysis w Data Visualization
- EDA w SQL
- Interactive Map with Folium
- Dashboard with Ploty Dash
- Predictive Analysis (Classification)

Summary of all results

- EDA results
- Snapshots/summary of results
- Predictive analysis results

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
    - SpaceX Rest API
    - Web Scrapping Wikipedia
- Perform data wrangling
    - Filtering data
    - Filling in missing values
    - Utilized Hot One Encoding to prep data for binary classification.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
    - Built different predictive analysis classification models and compared performance and accuracy to choose the best one.

Data Collection

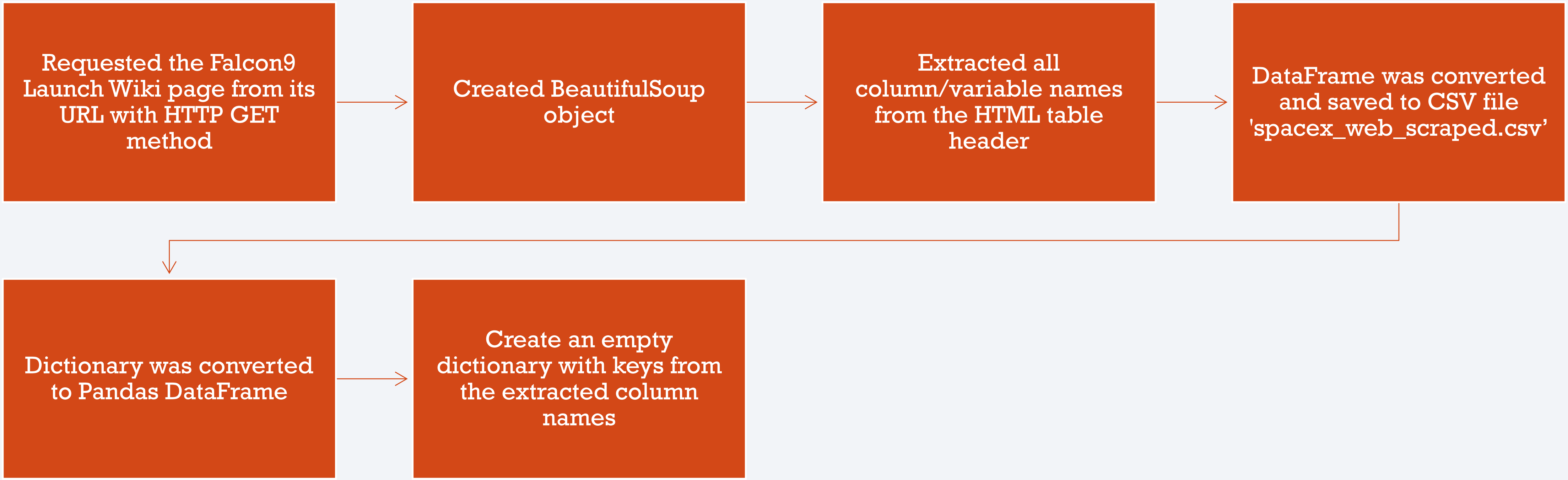
- The data for this project was collected from the SpaceX API with the use of requests.get(). The data was requested in three separate instances which included booster versions, launch sites, payloads, and cores, etc. The requested data was then stored in lists which were subsequently combined into a dictionary and then into a DataFrame.
- Records pertaining to Falcon 9 were filtered for in our DataFrame. Missing values were filled in with the column average.

Data Collection – SpaceX API

- Requested launch data from SpaceX API
- Now we decode the response content as a Json using .json() and turn it into a Pandas dataframe using .json_normalize()
- DataFrame was subsequently parsed for data of interest and stored in lists.
- Next, these lists were combined into a dictionary which was then converted into a Pandas DataFrame.
- DataFrame was filtered to include Falcon 9 data exclusively
- Missing values pertaining to Payload Mass were filled in with the use of .mean() and .replace()
- DataFrame was saved as 'dataset_part_1.csv'

GitHub URL for SpaceX API notebook: https://github.com/montielf11/IBM-Capstone/blob/76be61db1188255973e38dfaaa0859f59e974ae4/jupyter-labs-spacex-data-collection-api.ipynb

Data Collection - Scraping



GitHub URL for Webscraping notebook: https://github.com/montielf11/IBM-Capstone/blob/76be61db1188255973e38dfaaa0859f59e974ae4/jupyter-labs-webscraping.ipynb

Data Wrangling

Data from Scraping lab was loaded.

→ Number of missing values for each attribute identified.

→ Number of launches from each site was calculated.

→ Number and occurrence of mission outcomes were determined.

→ Landing outcomes were enumerated to make it possible to differentiate and label positive and negative outcomes

→ Data was saved in CSV format "dataset_part_2.csv"

GitHub URL for data wrangling notebook: https://github.com/montielf11/IBM-Capstone/blob/76be61db1188255973e38dfaaa0859f59e974ae4/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

The objective of Exploratory Data Analysis was to determine which factors influenced the success rate of Falcon 9 launches by SpaceX

- Payload Mass
- Flight Number
- Launch Site

Graphs and scatterplots created using Matplotlib and Pandas.

- Scatter: Payload Mass v. Flight Number
- Scatter: Launch Site vs Flight Number
- Scatter: Launch Site vs Payload Mass
- Scatter: Orbit vs Flight Number
- Scatter: Orbit vs Payload Mass
- Histogram: Average success rate per orbit type
- Line Chart: Success rate as time progressed

GitHub URL for data wrangling notebook:
https://github.com/montielf11/IBM-Capstone/blob/76be61db1188255973e38dfaaa0859f59e974ae4/edadataviz.ipynb

EDA with SQL

- SQL queries ran included:
  - SELECT DISTINCT Launch_Site FROM SPACEXTABLE;
  - SELECT * FROM SPACEXTABLE WHERE Launch_Site LIKE 'CCA%' LIMIT 5;
  - SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXTABLE WHERE CUSTOMER LIKE '%NASA%';
  - SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTABLE WHERE BOOSTER_VERSION LIKE "%F9 v1.1%"
  - SELECT MIN(DATE) FROM SPACEXTABLE WHERE Landing_Outcome LIKE "%ground pad%"
  - Select Booster_Version from SPACEXTABLE WHERE PAYLOAD_MASS__KG_ BETWEEN '4000' AND '6000' AND Mission_Outcome ='Success';
  - SELECT Mission_Outcome, Count(Mission_Outcome) from SPACEXTABLE group by Mission_Outcome;
  - SELECT DISTINCT Booster_Version FROM SPACEXTABLE WHERE PAYLOAD_MASS__KG_ =(SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTABLE);
  - SELECT substr("Date", 6, 2) AS Month, "Landing_Outcome", "Booster_Version", "Launch_Site" FROM SPACEXTABLE WHERE substr("Date", 0, 5) = '2015'AND "Landing_Outcome" LIKE '%drone ship%'AND "Landing_Outcome" LIKE '%Failure%';
  - SELECT Landing_Outcome, count(Landing_Outcome), Date FROM SPACEXTABLE WHERE Date BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY Landing_Outcome ORDER BY COUNT(Landing_Outcome) DESC;

- Conclusion of this lab led to summary of successful and failed launches

Landing_Outcome	count(Landing_Outcome)	Date
No attempt	10	2012-05-22
Success (drone ship)	5	2016-04-08
Failure (drone ship)	5	2015-01-10
Success (ground pad)	3	2015-12-22
Controlled (ocean)	3	2014-04-18
Uncontrolled (ocean)	2	2013-09-29
Failure (parachute)	2	2010-06-04
Precluded (drone ship)	1	2015-06-28

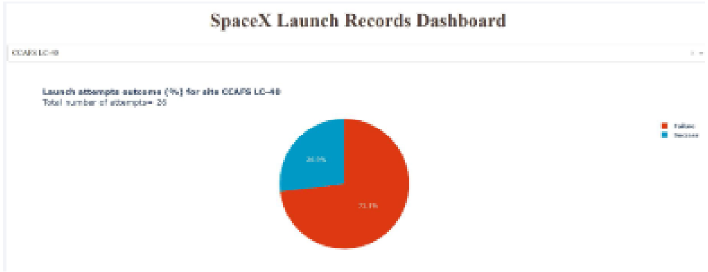
Lab URL link: https://github.com/montielf11/IBM-Capstone/blob/76be61db1188255973e38dfaaa0859f59e974ae4/jupyter-labs-eda-sql-coursera_sqllite.ipynb

- Add the GitHub URL of your completed EDA with SQL notebook, as an external reference and peer-review purpose

Build an Interactive Map with Folium

- Summarize what map objects such as markers, circles, lines, etc. you created and added to a folium map
- Explain why you added those objects
- Add the GitHub URL of your completed interactive map with Folium map, as an external reference and peer-review purpose

DASHBOARD WITH PLOTLY DASH



- Built an interactive dashboard with Ploty:
- Dropdown menu for selecting launch sites
- Pie charts displaying success rate
- Scatter chart displaying launch site, payload mass, success and failure
- Range slider for selecting range of payload mass

- Utilized dashboard to analyze SpaceX launches to determine:
- Site with largest successful launches
- Payload ranges with highest launch success rate
- Payload ranges with lowest launch success rate
- F9 Booster versions with highest success rates

Predictive Analysis (Classification)

Model Training

Logistic Regression

Support Vector Machine

Decision Tree Classifier

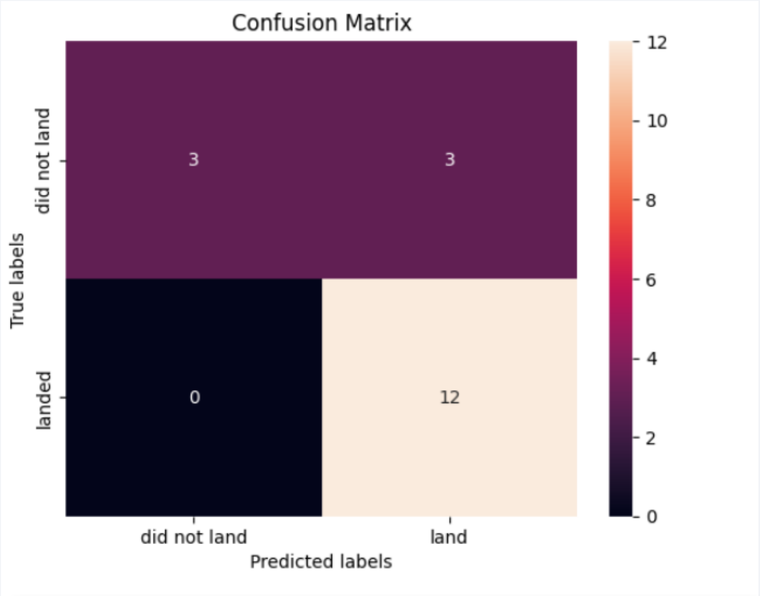
KNN Classifier

Supervised Classification Models
Launcher Landing Labels
Class 0: Fail
Class 1: Success

Imported Falcon 9 data to CSV

Data Normalization

Split Data: Training and Test Sets



Predictive Analysis
Accuracy
Confusion Matrix

Selection of Best Model

EXPLORATORY DATA ANALYSIS RESULTS

INTERACTIVE ANALYTICS DEMO IN SCREENSHOTS

PREDICTIVE ANALYSIS RESULTS

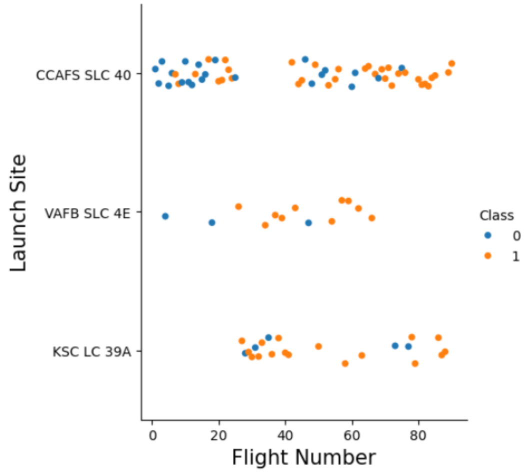
Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

- All sites had higher levels of success as Flight Number was higher
- CCAFS SLC 40 had the most mixed results.

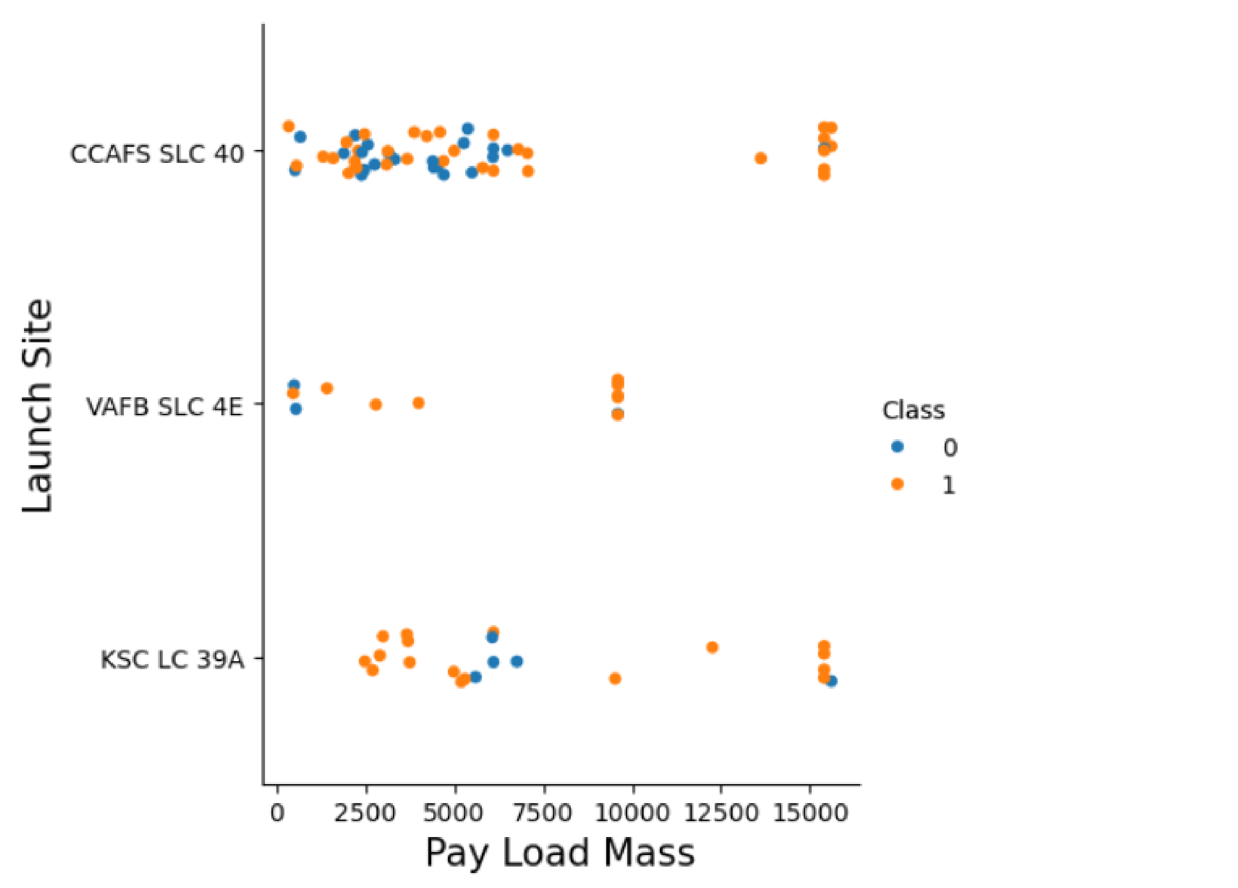
```
# Plot a scatter point chart with x axis to be Flight Number and y axis to be the launch site, and hue to be the class val
sns.catplot(y='LaunchSite', x='FlightNumber', data=df, hue='Class')
plt.xlabel('Flight Number', fontsize=15)
plt.ylabel('Launch Site', fontsize=15)
plt.show()
```



Payload vs. Launch Site

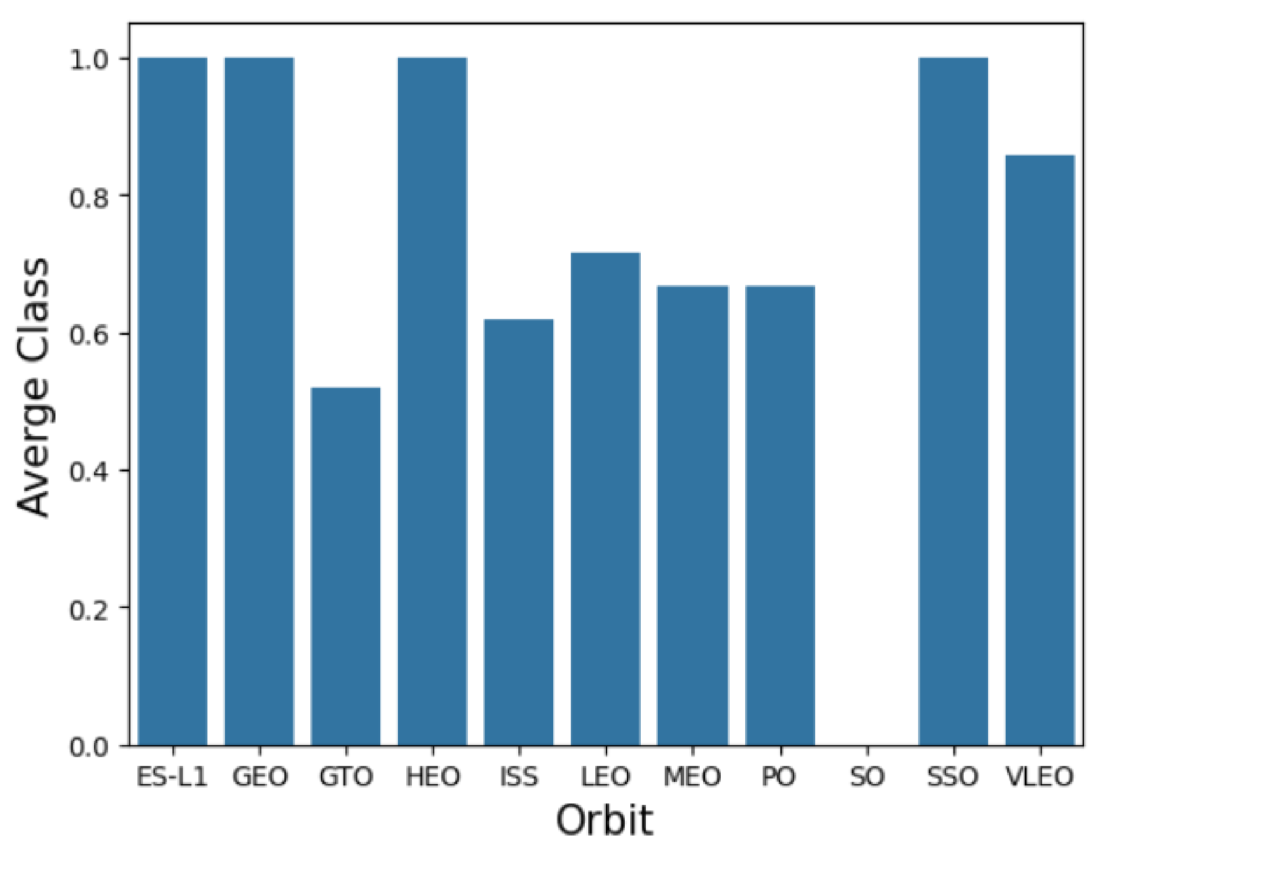
- All three launch sites had greater levels of success at relatively greater pay load masses.

- No Payload Mass was greater than 10000 for VAFB



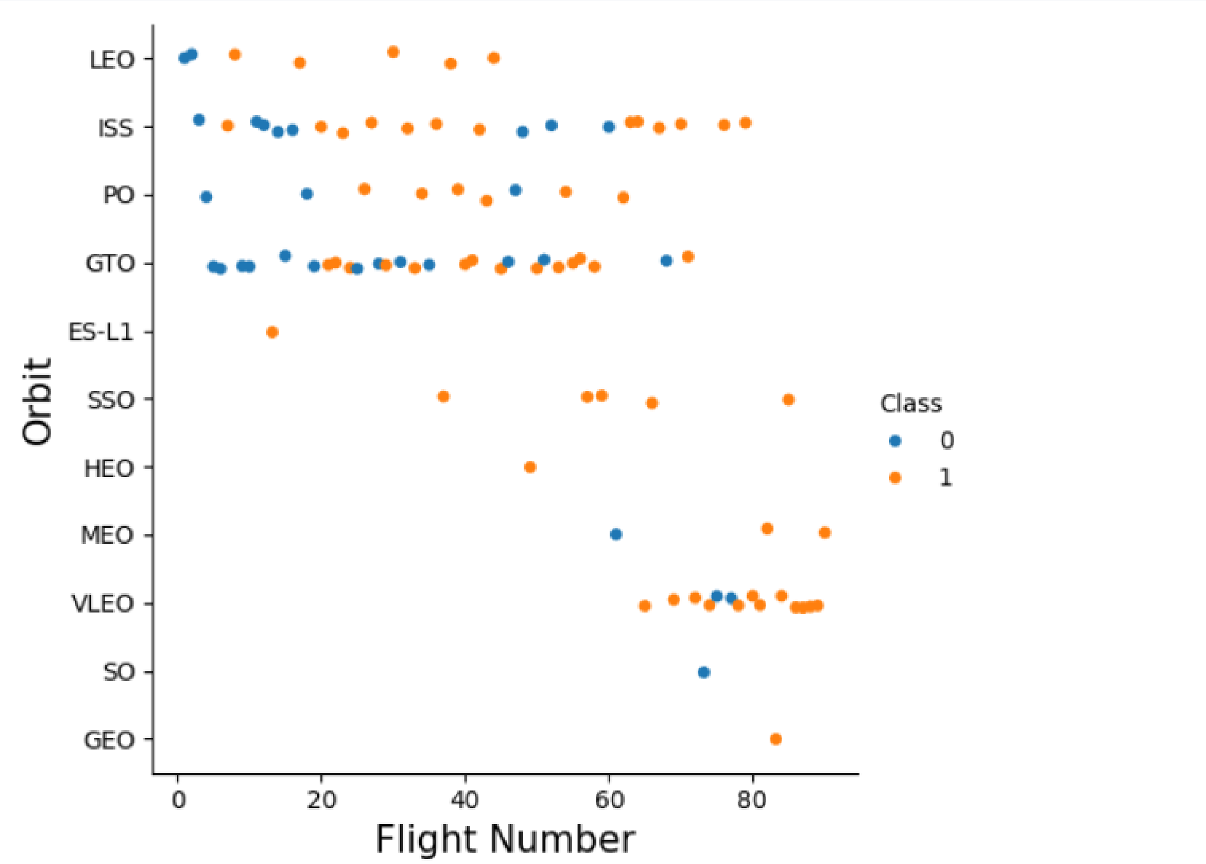
Success Rate vs. Orbit Type

- The orbit types with the highest success rates are ES-L1, GEO, HEO, SSO, VLEO



Flight Number vs. Orbit Type

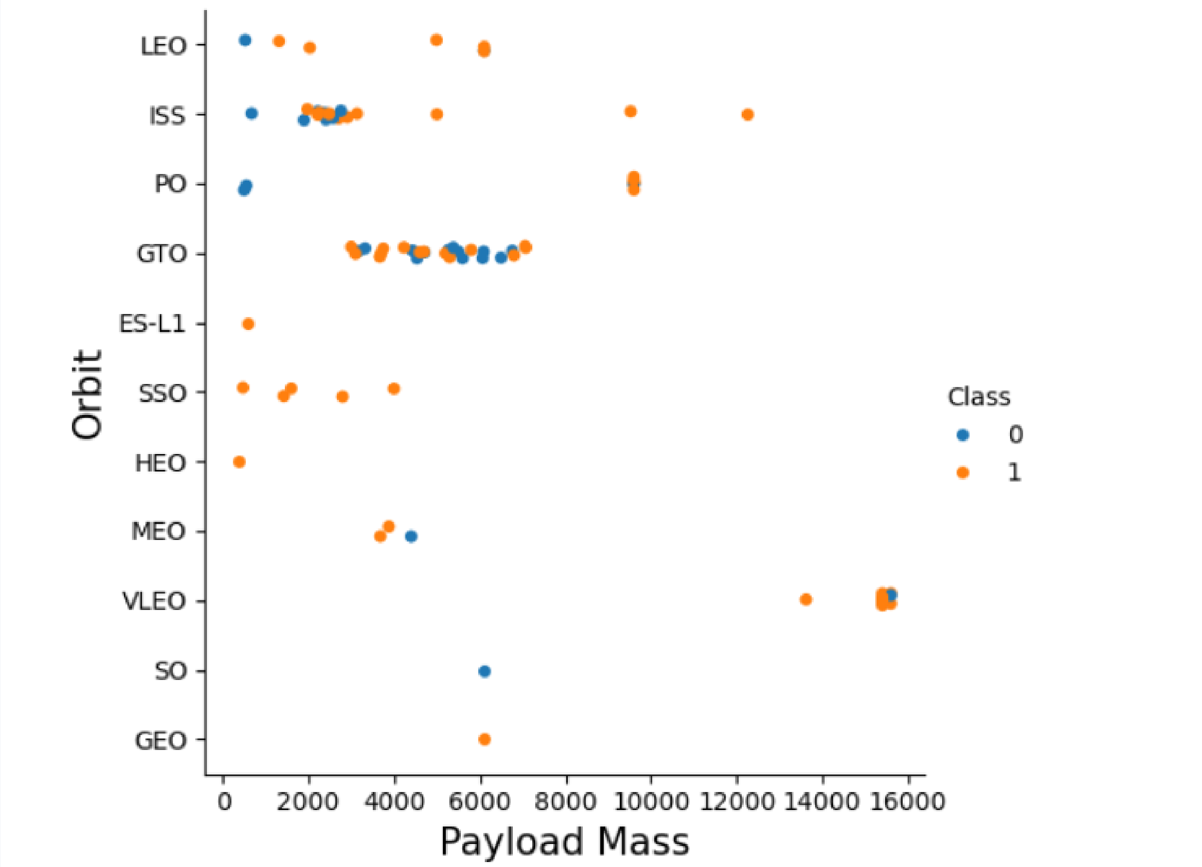
- There is a higher success rate for higher flight number for LEO flights.
- Some orbit types, such as GTO, and ISS show not correlation between Flight Number and Orbit type.



Payload vs. Orbit Type

- LEO, ISS, PO show higher success rates associated with greater Payload Mass.

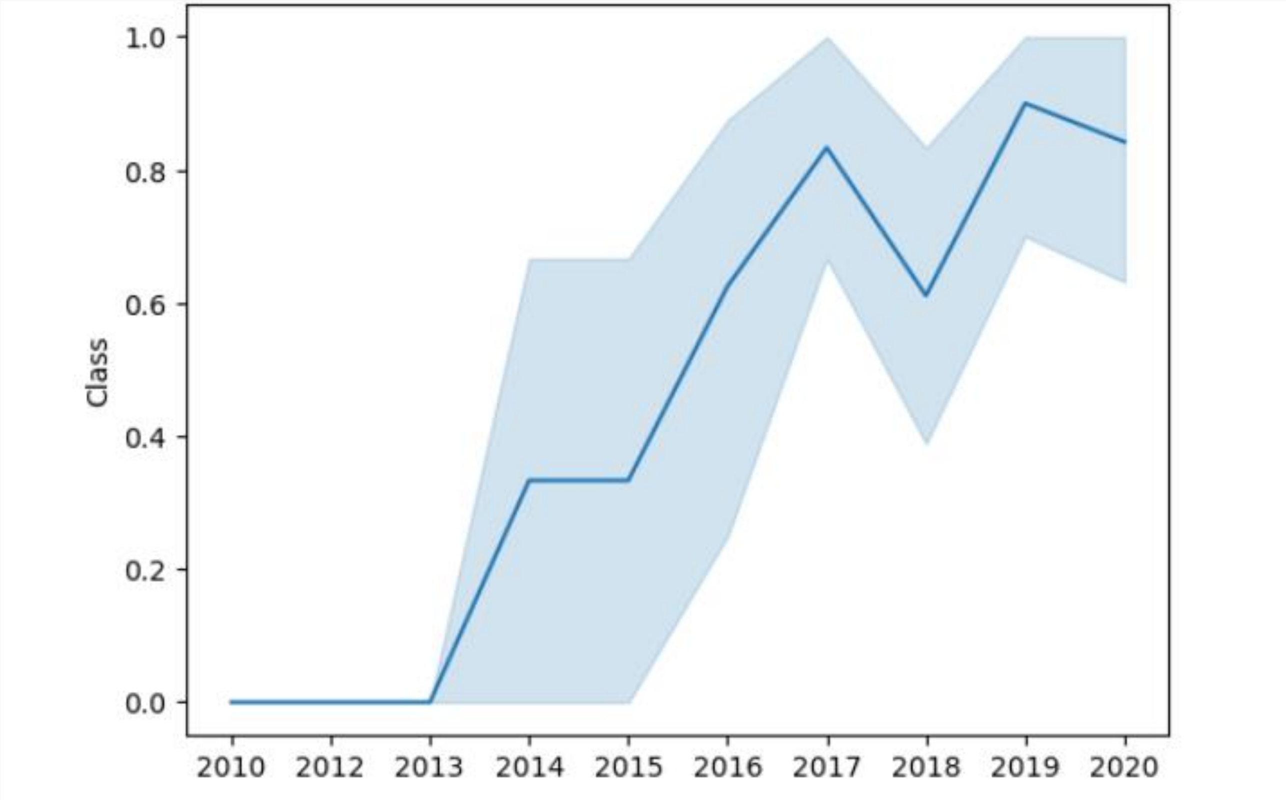
- Some Orbit types such as MEO, GTO, VLEO show no improvement as Payload Mass becomes greater.



Launch Success Yearly Trend

- After 2013 the launch success rate increased steadily each year.

- The only year that saw a decrease was 2018.



All Launch Site Names

- The unique launch sites were:
  - CCAFS LC-40
  - VAFB SLC-4E
  - KSC LC-39A
  - CCAFS SLC-40
- SQL code to obtain this result: SELECT DISTINCT Launch_Site FROM SPACEXTABLE;

Task 1

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

```
%sql SELECT DISTINCT Launch_Site FROM SPACEXTABLE;
```

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

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

Launch Site Names Begin with 'CCA'

- SQL query to obtain 5 entries:
  SELECT * FROM SPACEXTABLE WHERE Launch_Site LIKE ‘CCA%’ LIMIT 5;

```
]: %sql SELECT * FROM SPACEXTABLE 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

- Total Payload Mass was calculated at 107010 KG.
- SQL script needed: SELECT SUM(PAYLOAD_MASS__KG) FROM SPACEXTABLE WHERE CUSTOMER LIKE '%NASA%'

```
%sql SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXTABLE WHERE CUSTOMER LIKE '%NASA%';
```

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

SUM(PAYLOAD_MASS__KG_)
107010

Average Payload Mass by F9 v1.1

- Average Payload Mass carried by F9 V1.1 is 2534.66
- SQL script needed: SELECT SUM(PAYLOAD_MASS__KG) FROM SPACEXTABLE WHERE CUSTOMER LIKE ‘%NASA%’

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

```
]: %sql SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTABLE WHERE BOOSTER_VERSION LIKE "%F9 v1.1%"
```

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

AVG(PAYLOAD_MASS__KG_)
2534.6666666666665

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad
- Present your query result with a short explanation here

```
]: %sql SELECT MIN(DATE) FROM SPACEXTABLE WHERE Landing_Outcome LIKE "%ground pad%"

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

]:

MIN(DATE)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

SQL code required: Select .Booster_Version from SPACEXTABLE WHERE PAYLOAD_MASS__KG_ BETWEEN '4000' AND '6000' AND Mission_Outcome ='Success';

```
%sql Select Booster_Version from SPACEXTABLE WHERE PAYLOAD_MASS__KG_ BETWEEN '4000' AND '60
```

Booster_Version
F9 v1.1
F9 v1.1 B1011
F9 v1.1 B1014
F9 v1.1 B1016
F9 FT B1020
F9 FT B1022
F9 FT B1026
F9 FT B1030
F9 FT B1021.2
F9 FT B1032.1
F9 B4 B1040.1
F9 FT B1031.2
F9 FT B1032.2
F9 B4 B1040.2
F9 B5 B1046.2
F9 B5 B1047.2
F9 B5 B1046.3
F9 B5 B1048.3
F9 B5 B1051.2
F9 B5B1060.1
F9 B5 B1058.2
F9 B5B1062.1

Total Number of Successful and Failure Mission Outcomes

- The following table shows landing outcomes for SpaceX and respective counts.
- The SQL code is listed above.

```
7]: %sql SELECT Mission_Outcome, Count(Mission_Outcome) from SPACEXTABLE group by Mission_Outc
```

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

7]:

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

Boosters Carried Maximum Payload

- The following are the Booster Versions that carried the maximum payloads.
- Query code is as follows: SELECT DISTINCT Booster_Version FROM SPACEXTABLE WHERE PAYLOAD_MASS__KG_ =(SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTABLE);

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

- The following table lists the month, landing outcome, booster version, and launch site for 2015:

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

- Both failed launches on record were from launch site CCAFS LC-40. Booster versions were different, however.

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

- Following code was used to produce desired query: %sql select Landing_Outcome, count(Landing_Outcome), Date FROM SPACEXTABLE WHERE Date BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY Landing_Outcome ORDER BY COUNT(Landing_Outcome) DESC;
- The highest landing outcome was No attempt followed by Success by drone ship.

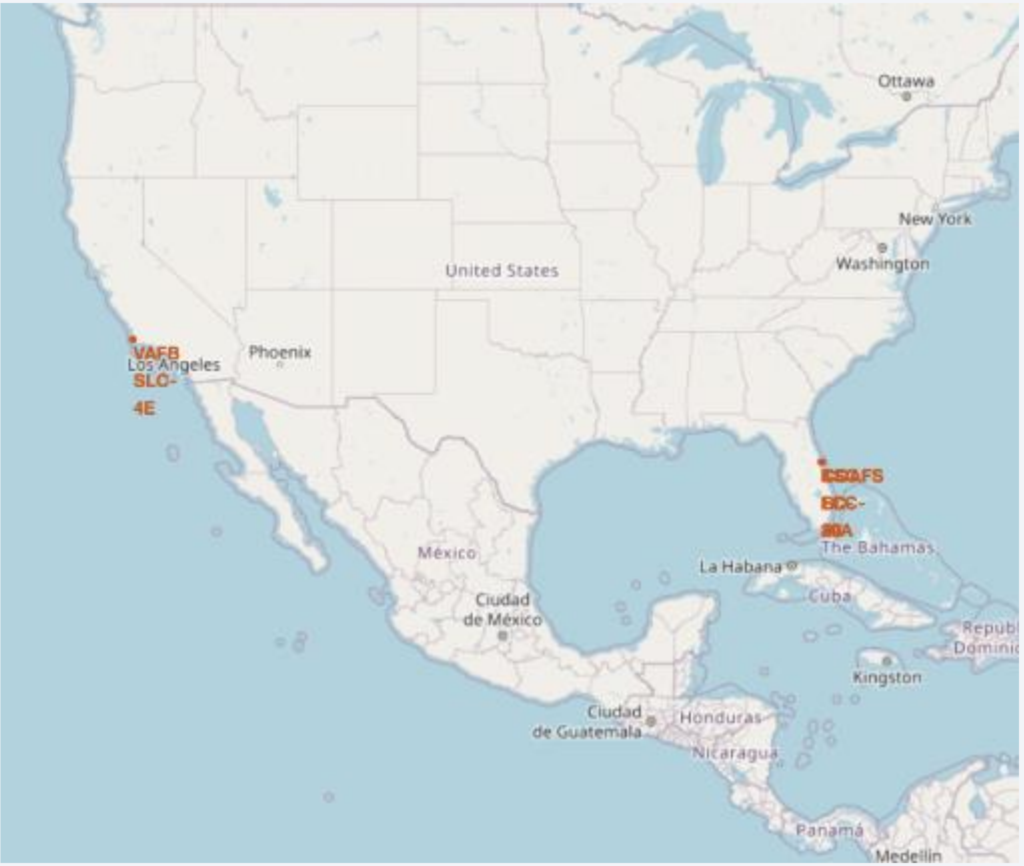
Landing_Outcome	count(Landing_Outcome)	Date
No attempt	10	2012-05-22
Success (drone ship)	5	2016-04-08
Failure (drone ship)	5	2015-01-10
Success (ground pad)	3	2015-12-22
Controlled (ocean)	3	2014-04-18
Uncontrolled (ocean)	2	2013-09-29
Failure (parachute)	2	2010-06-04
Precluded (drone ship)	1	2015-06-28

Section 3

Launch Sites Proximities Analysis

Site Locations on Global Map

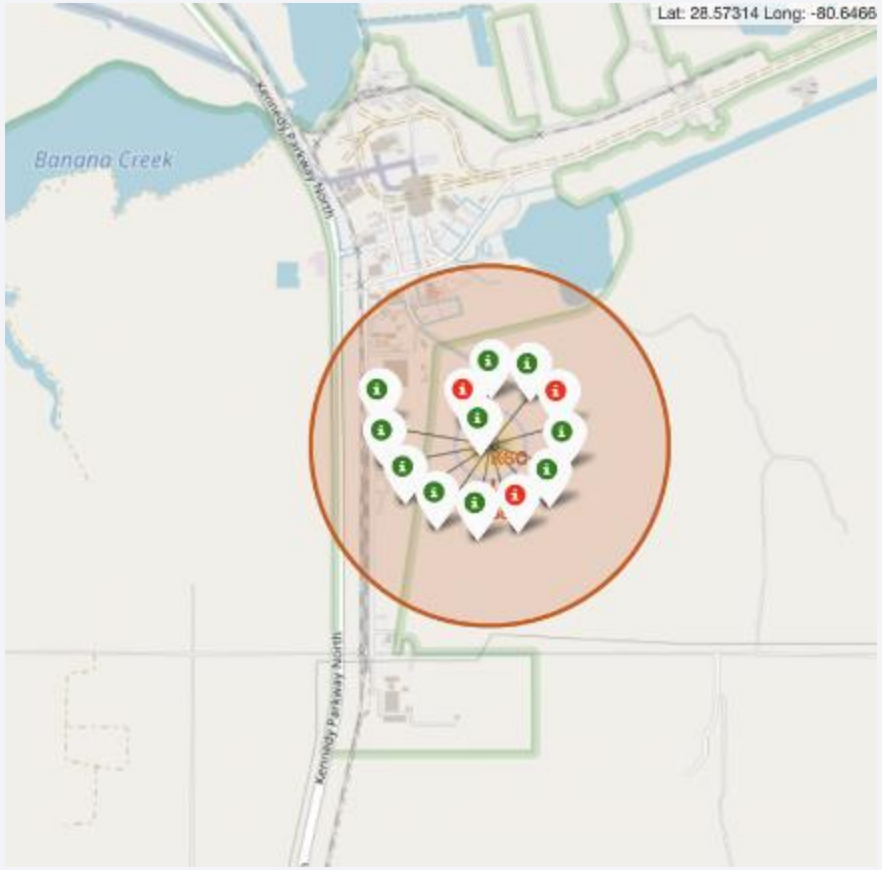
- The launch sites have been purposely chosen to be by the coast.
- Minimizes the chance that debris can fall on populated areas.



Color Coded Success Rates

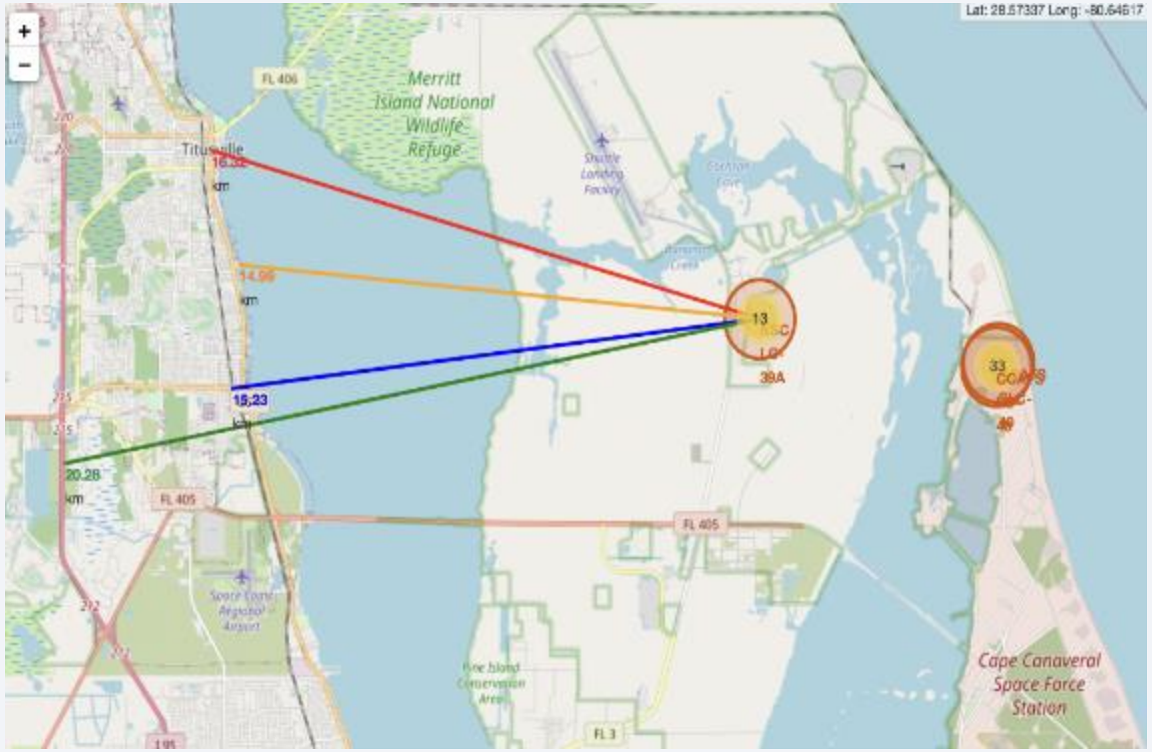
- Success and failure are easily identified by red and green markers.

- KSC LC 39A had a high success rate.



Distance from Launch Sites

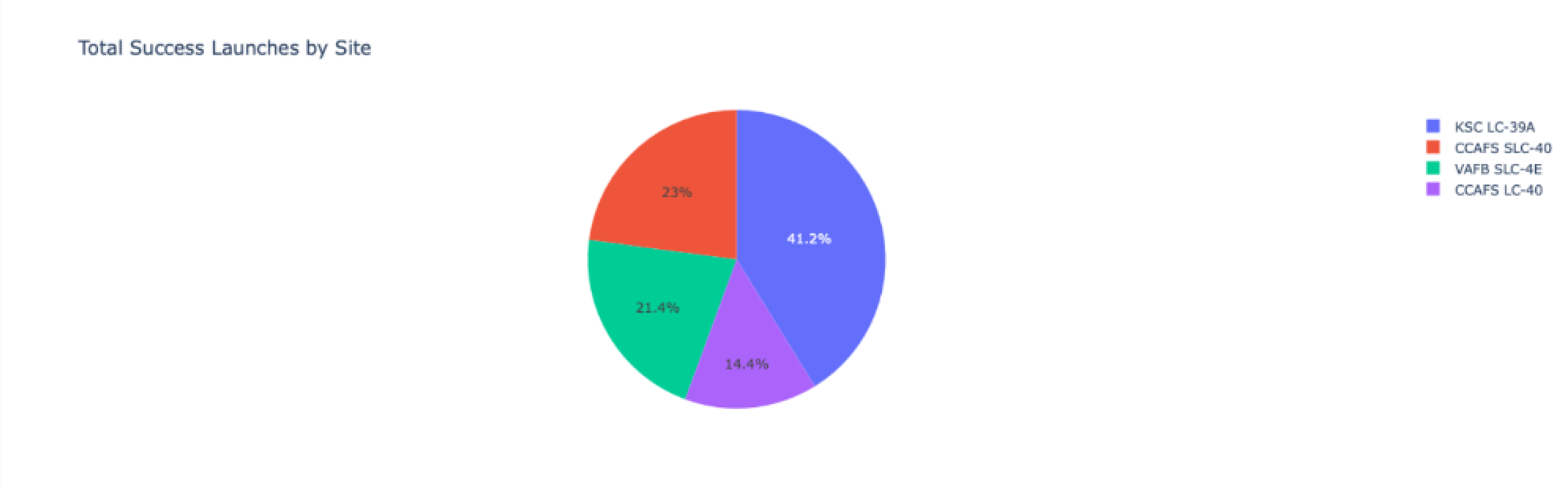
- Railway is15.23 km away.
- Highway is 20.28 km away.
- Coastline is 14.99 km away.



Section 4

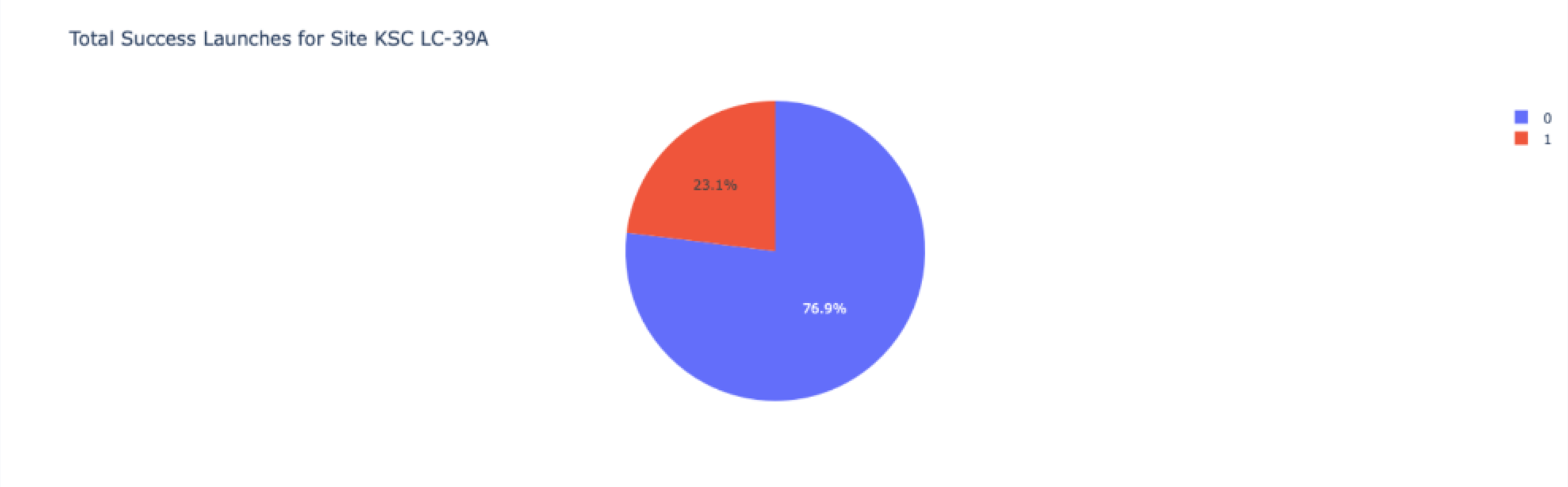
Build a Dashboard with Plotly Dash

Total Successful Launches by Site



- With a success rate of 41.2%, KSC LC-39A was the most successful site.

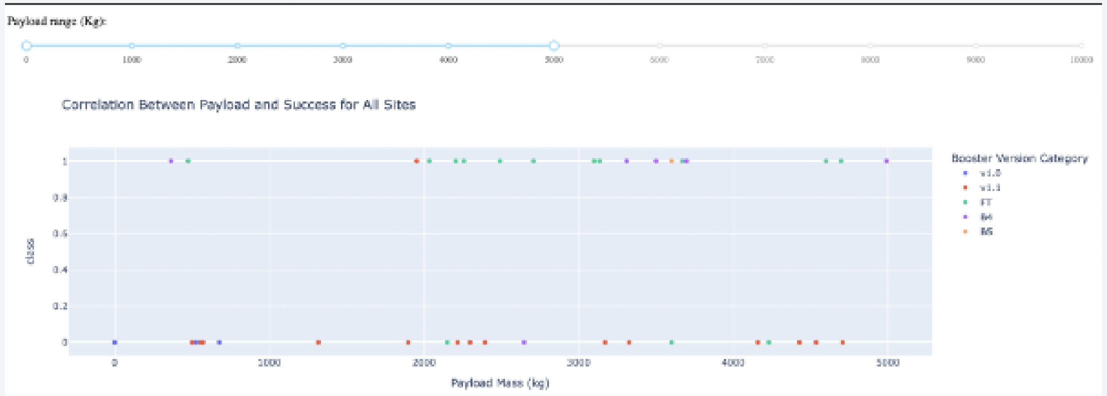
Stats for Most Successful Launch Site

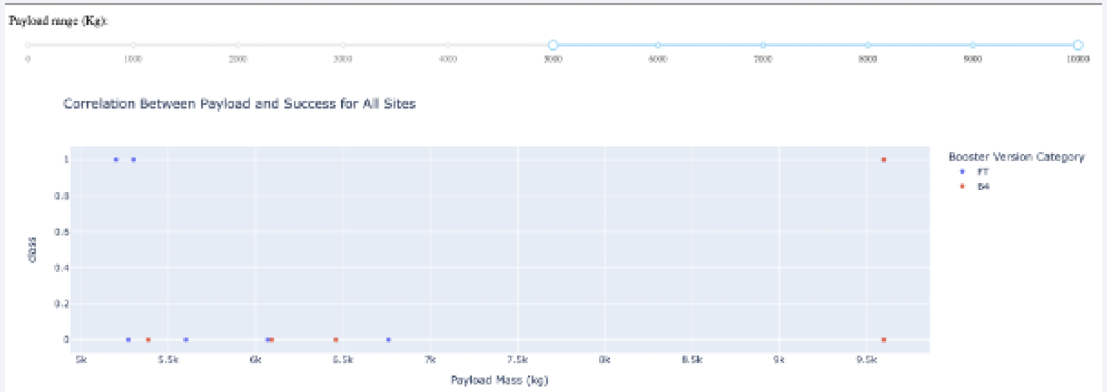


- KSC LC-39A had a success rate of 76.9% and failure rate of 23.1%.

Successful Payload Mass for All Sites

- Charts show that payloads that were between 2000 and 5500 kg had the highest rates of success.





Section 5

Predictive Analysis (Classification)

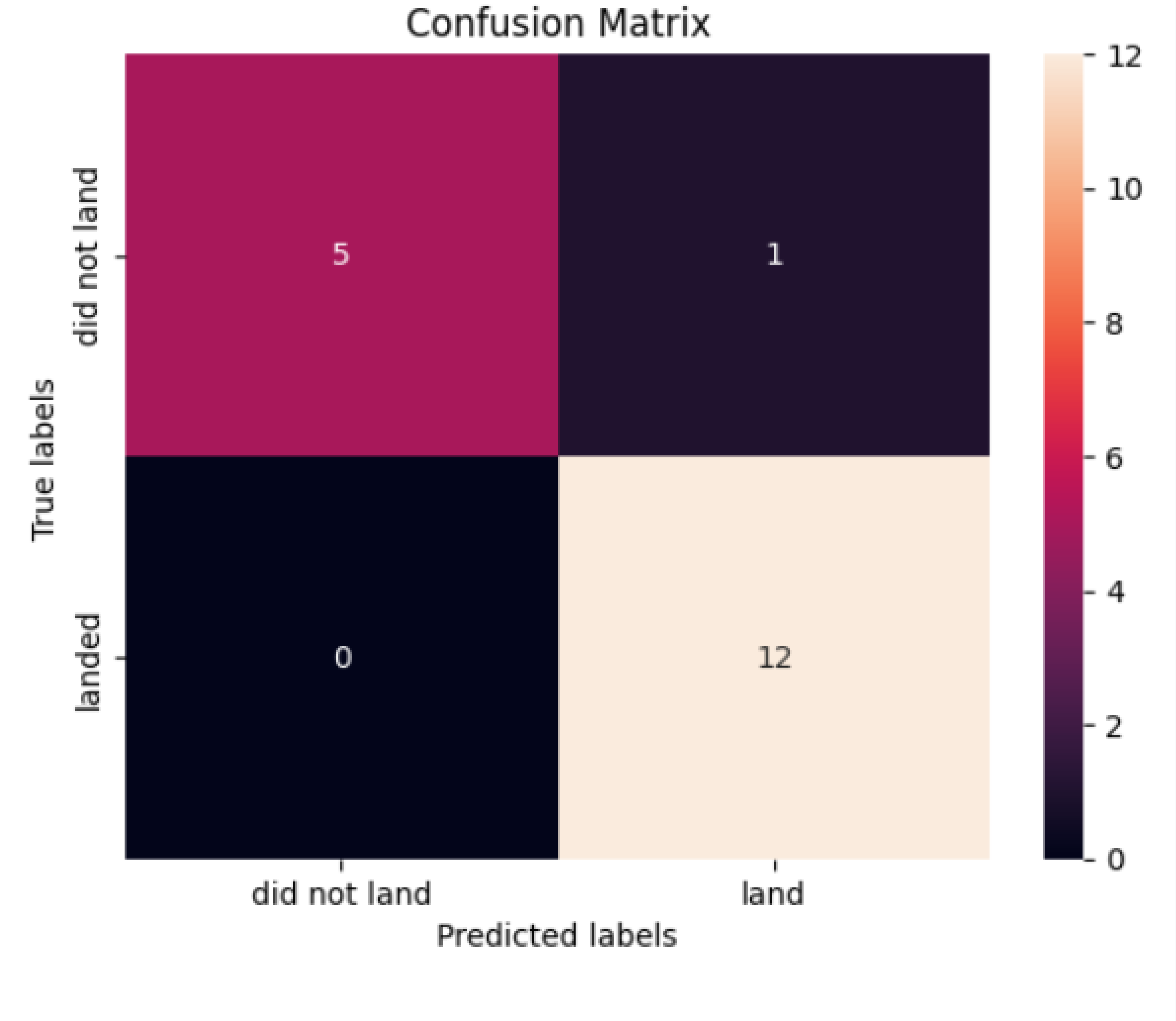
Classification Accuracy

- The model with the highest level of Accuracy was the Decision Tree.

```
Logistic Regression Accuracy: 0.8333333333333334
Decision Tree Accuracy: 0.9444444444444444
SVM Accuracy: 0.8333333333333334
```

Decision Tree Confusion Matrix

- The confusion matrix for Decision Tree model shows only 1 error out of 18 (False Positive).



CONCLUSIONS

- Decision Tree model had the best accuracy for the relevant data set
- Launches with Pay load mass between 2000 and 5500 kg had the highest rates of success.
- Launch sites are close to the coastline and as close to the equator as possible within the US.
- Successful rate for launches had increased dramatically over time.
- Site KSC LC-39A had the highest launch success rate.


Thank you!