IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Aladdin Alswaifi
22/06/2023

Outline

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

Executive Summary

- SpaceY is a new company that is planning to compete with SpaceX by making space trips more affordable.

Summary of methodologies:	Summary of all results:
- Data collection: using SpaceX API and webscraping from Wikipedia. - Data wrangling. - Exploratory Data Analysis (EDA): using SQL, Pandas, and MatplotLib. - Interactive Data Analytics and Dashboarding with Folium and Ploty Dash. - Predictive Analysis using Machine Learning (Classification).	- Logistic regression, SVM, Decision Tree, and KNN Classification. - Defining the best method using Test Data.

Introduction

- Project background and context:
  - In SpaceY, we try to predict if the Falcon 9 first stage will land successfully. SpaceX 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 SpaceX 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 SpaceX for a rocket launch. In this module, you will be provided with an overview of the problem and the tools you need to complete the course.

- Problems you want to find answers:
  - Will Space X will attempt to land a Rocket or not?
  - Can Space Y launch cheaper rockets by finding the factors that can save launch costs?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:

  - Space-X REST API - Web scraping from web-related Wiki pages.

- Perform data wrangling:
  - We replaced the missing values with the average values.
  - We removed the irrelevant data such as data of Faclon1.
  - We completed the data from other sources when required (e.g. replacing the rocked codes with rocket names.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Using Machine Learning to determine if the first stage of Falcon 9 will land successfully.

Data Collection

Data was collected with 2 methods:

From Space-X website using API and web scraping from Wikipedia

SpaceX REST API endpoint → Get request to receive the information from tSpaceX Endpoint → Receive the data in a form of JSON file → Convert JSON file into a DataFrame

Webscraping Falcon 9 data from wikipedia using BeautifulSoup → Parse Data from the scraped tables → Convert tables into DataFrames

Data Collection – SpaceX API

SpaceX REST API endpoint

Get request to receive the information from tSpaceX Endpoint

Receive the data in a form of JSON file

Convert JSON file into a DataFrame

[Collecting the data using API Notebook]

Data Collection - Scraping

[Collecting Data using webscrapping Notebook]

Data Wrangling

- Identify the missing values and replace it with the mean
- **Define numerical and categorical data.**
- **Calculate number of launches in each site.**
- Calculate the number and occurrence of each orbit.
- **Create a landing outcome label from outcome column**
- Calculate the number and occurrence of mission outcome per orbit type.

[Data Wrangling Notebook]()

EDA with Data Visualization

Charts that were plotted:

- Catplot of the relationship between flight number and Payload.
- Catplot of the relationship between flight number and launch site.
- Catplot of the relationship between the Payload and Launch site.
- Catplot of the relationship between flight number and Orbit type.
- Catplot of the relationship between Payload and Orbit type.
- Bar chart to visualize relationship between success rate of each Orbit type.
- Line chart to visualize the launch success yearly trend.

- Data Visualization Notebook

EDA with SQL

- SQL queries performed:
  - Display the names of launch sites.
  - Display 5 records where launch sites being with “CCA”.
  - Display the total Payload mass carried by boosters launched by NASA.
  - Display average Payload mass carried by booster version F9 v1.1
  - List the data when the first successful landing outcome achieved.
  - List the names of the boosters which had success in drone ship with Payload mass between 4000 kg and 6000 kg.
  - List of total numbers of successful and failure outcomes.
  - List of the names of the boosters that carried the maximum Paybload mass.
  - List of failed landing outcome.
  - Rank of the count of landing outcome.
  - [EDA with SQL Notebook]

Build an Interactive Map with Folium

- Summary of map objects created and added to the folium map:
  - Launch sites.
  - Marker clusters used to simplify a map containing many markers.
  - Mouse position on the map to get coordinate of the mouse over a point on the map.
  - Folium.Polyline object was used to draw a line between launch site to the closest city, railway, and highway.

  - Interactive visual analytics with Folum Notebook

Build a Dashboard with Plotly Dash

- Summary of interactive tools that were added to the Dashboard:
  - A launch site drop-down input component.
  - A callback function to render success-pie-chart based on selected site dropdown
  - A range slider to select payload.
  - A callback function to render the success-payload-scatter-chart plot

[Interactive visual analytics]

Predictive Analysis (Classification)

- Summary of Model Development process:
  - Numpy array created from the ‘Class’ Column.
  - Data standardized.
  - Function used to split data into training and test sets.
  - Searching for the best hyperparameters for logistic regression, SVM, decision tree and KNN Classifiers.
  - Searching for the best performing method using test data.

  - Machine Learning prediction Notebook

Predictive Analysis (Classification)

- Load of data CSV files
- Variables Y & X created from the column class
- Data splatted into a training and test sets
- Models created
- Models created
- **Selection of the best models**

[Machine Learning prediction Notebook]

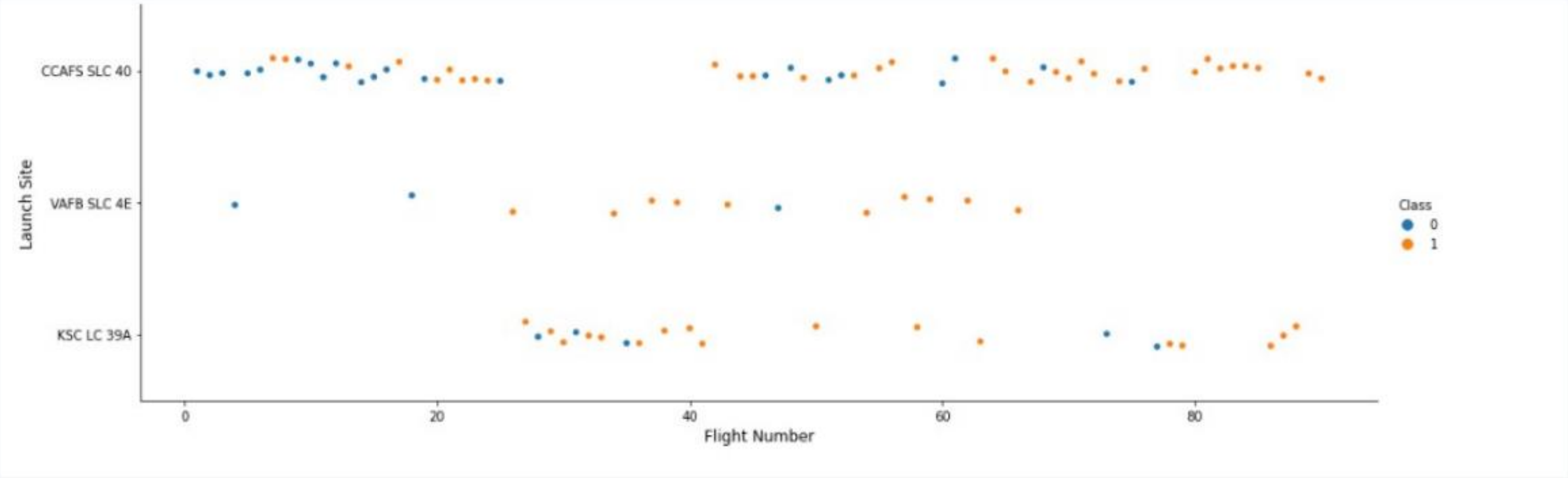
Results

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

Section 2

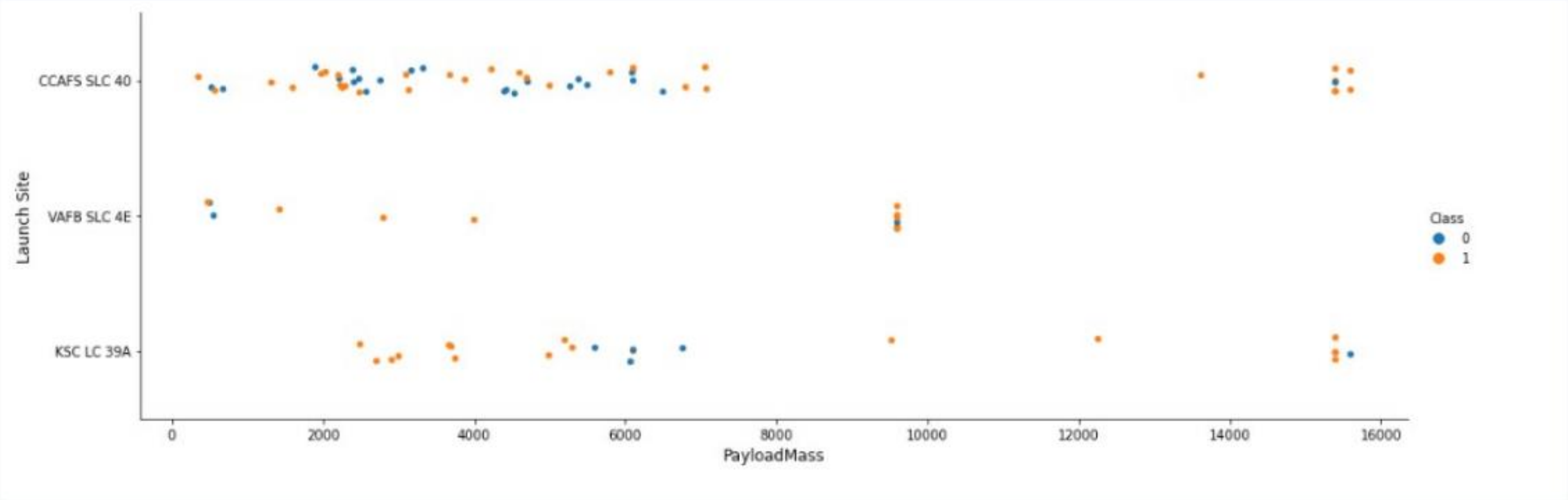
Insights drawn from EDA

Flight Number vs. Launch Site



- Success rate increased with time for all sites.
- VAFB SLC 4E and KSC LC 39 A had a higher success rate.

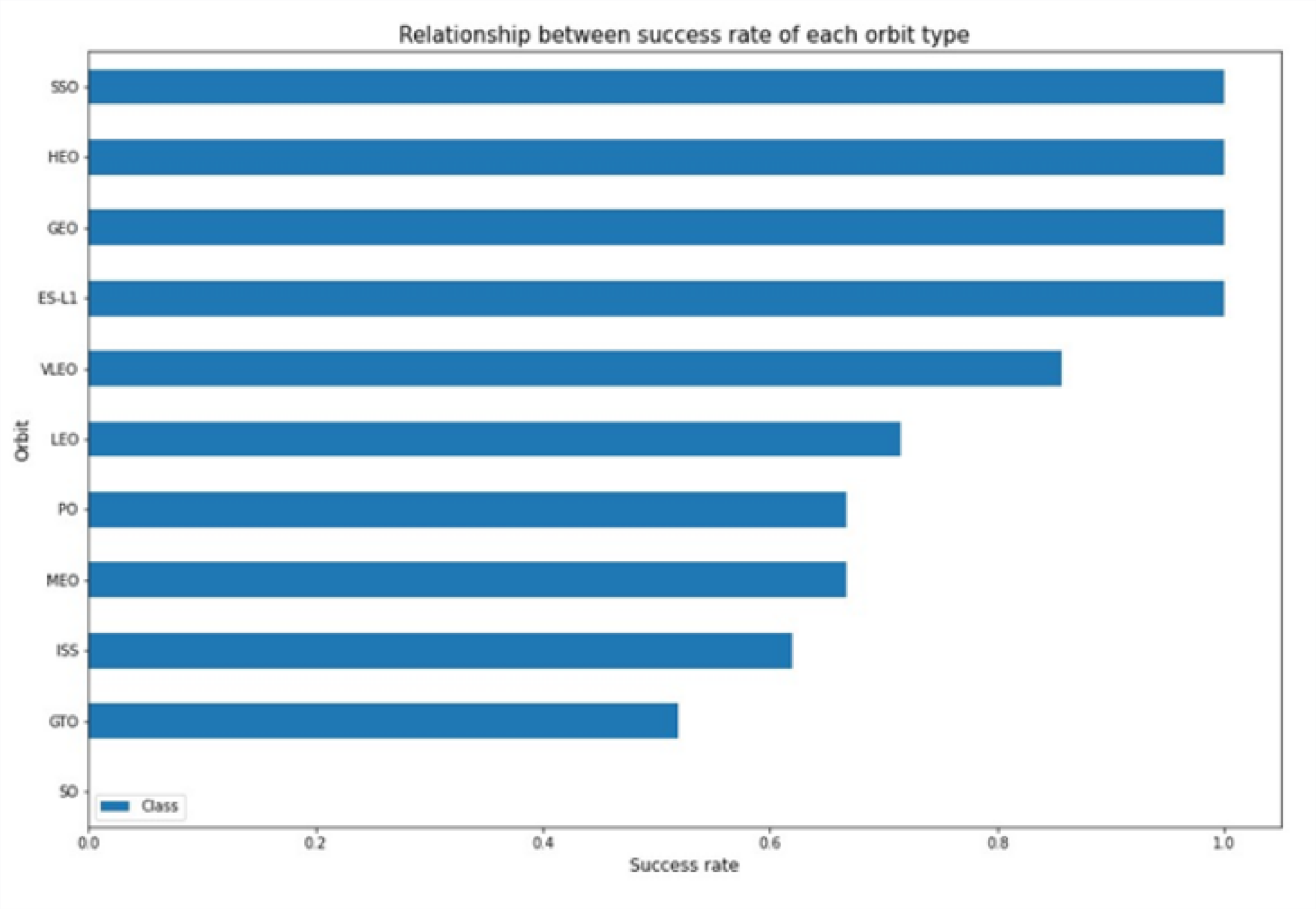
Payload vs. Launch Site



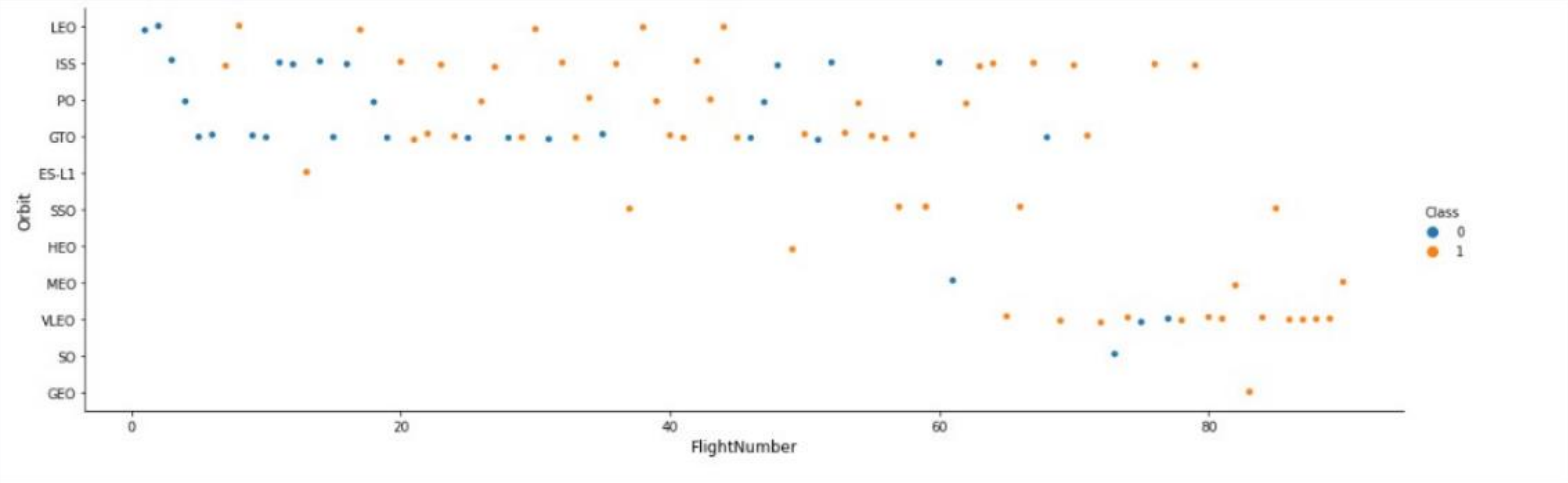
- In VAFB-SLC site: no rockets launched with heavy payload mass (more than 10000 Kg)
- In KSC LC launch site: no rockets launched with small preload mass (less than 2500 kg)
- CCAFS SLC has launched rockets either less than 7500 kg or more than 1300 kg.

Success Rate vs. Orbit Type

- The first 4 orbit types have the best successful rate.

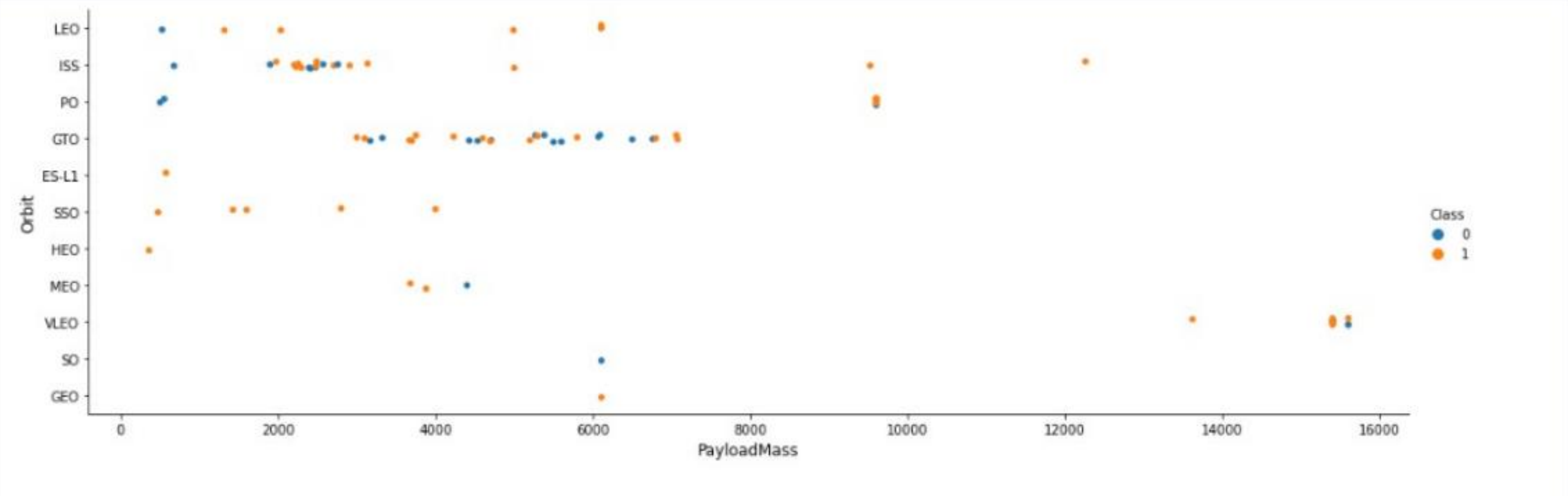


Flight Number vs. Orbit Type



- Success rate increased after the first 40 launches.
- GTO and ISS has the higher number of launches with the lowest successful landings.

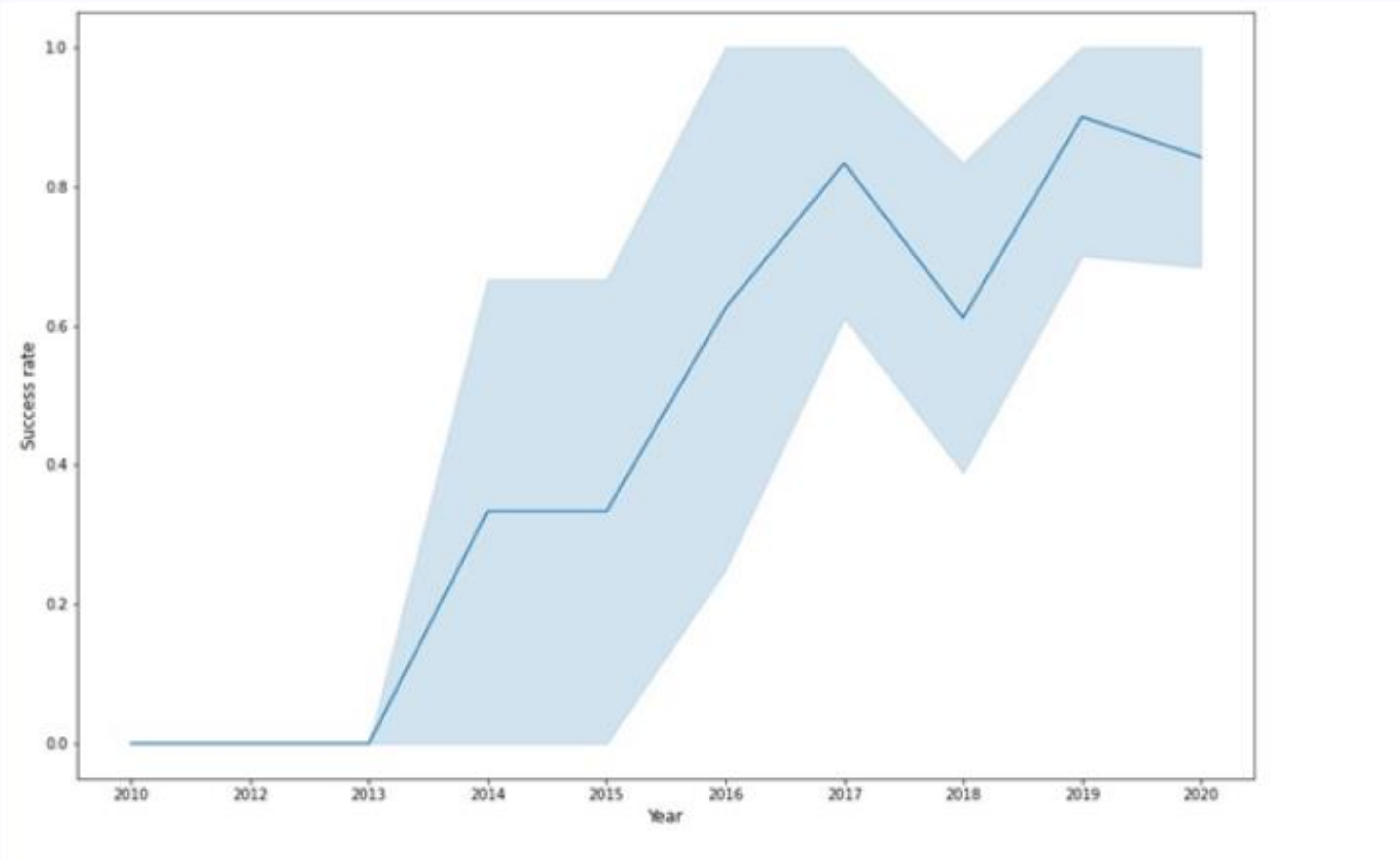
Payload vs. Orbit Type



- Successful landing rate with heavy payload are more for Polar, LEO, and ISS.
- Less than 10 launches exceeded 1600 kg.

Launch Success Yearly Trend

- Success rate was increasing from 2013 till 2020.



All Launch Site Names

- Select Statement used

Task 1

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

```
[7]: %sql  SELECT DISTINCT LAUNCH_SITE  FROM SPACEXTBL;
```

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

[7]:

Launch_Site
CCAFS LC-40
VAFB SLC-4E
KSC LC-39A
CCAFS SLC-40
None

Launch Site Names Begin with 'CCA'

Task 2

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

```
[8]: %sql SELECT * FROM SPACEXTBL WHERE LAUNCH_SITE LIKE 'CCA%' LIMIT 5;
```

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

[8]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_M.
06/04/2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	
12/08/2010	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	
22/05/2012	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	
10/08/2012	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	
03/01/2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	

Total Payload Mass

Task 3

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

[9]:
```
%sql SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE CUSTOMER = 'NASA (CRS)';
```

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

[9]:

SUM(PAYLOAD_MASS__KG_)
45596.0

Average Payload Mass by F9 v1.1

Task 4

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

```
[10]: %sql SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE BOOSTER_VERSION = 'F9 v1.1'

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

[10]:

AVG(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

Task 5

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

Hint:Use min function

[11]:
```
%sql select min(DATE) from SPACEXTBL where Landing_Outcome = 'Success (ground pad)';
```

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

[11]:

min(DATE)
01/08/2018

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

[12]:
```
%sql SELECT BOOSTER_VERSION from SPACEXTBL WHERE LANDING_OUTCOME = 'Success (drone ship)' and  PAYLOAD_MASS__KG_  >4000 and PAYLOAD_MASS__KG_ <6000;
```

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

[12]:

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

Total Number of Successful and Failure Mission Outcomes

Task 7

List the total number of successful and failure mission outcomes

```
[25]: %sql select MISSION_OUTCOME, count(MISSION_OUTCOME) As Total from SPACEXTBL where MISSION_OUTCOME =  'Success' or MISSION_OUTCOME = 'Failure (in flight)'  Group By MI
```

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

[25]:

Mission_Outcome	Total
Failure (in flight)	1
Success	98

Boosters Carried Maximum Payload

Task 8

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

[14]:
```
%sql SELECT BOOSTER_VERSION FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ = (SELECT max(PAYLOAD_MASS__KG_) FROM SPACEXTBL);
```

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

[14]:

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

Task 9

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, 4, 2) as month to get the months and substr(Date,7,4)='2015' for year.

```
[24]: %sql SELECT substr(Date, 4, 2) as Month , BOOSTER_VERSION,LAUNCH_SITE,LANDING_OUTCOME FROM SPACEXTBL WHERE LANDING_OUTCOME = 'Failure (drone ship)' and substr(Date,
```

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

[24]:

Month	Booster_Version	Launch_Site	Landing_Outcome
10	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

Task 10

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.

[27]:
```
%%sql
select `Landing_Outcome`, count(`Landing_Outcome`) as 'Number of landings' from SPACEXTBL
where `Date` between '04-06-2010' and '20-03-2017'
group by `Landing_Outcome`
order by  'Number of landings' desc
```

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

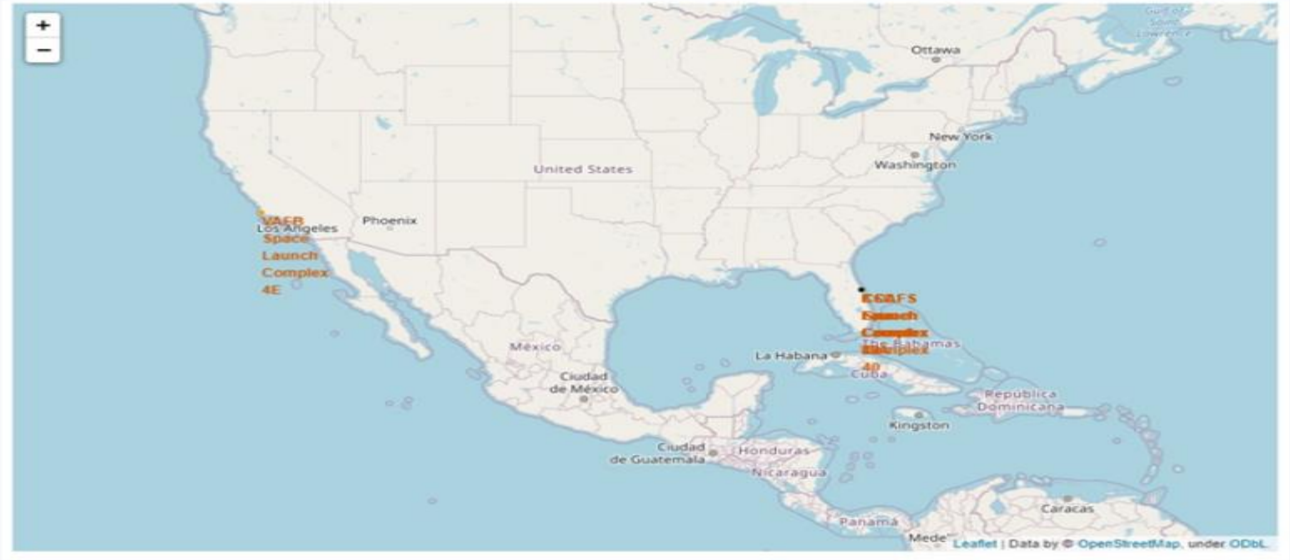
[27]:

Landing_Outcome	Number of landings
Success (ground pad)	7
Success (drone ship)	8
Success	20
No attempt	1
No attempt	10
Failure (parachute)	2
Failure (drone ship)	3
Failure	3
Controlled (ocean)	2

Section 3

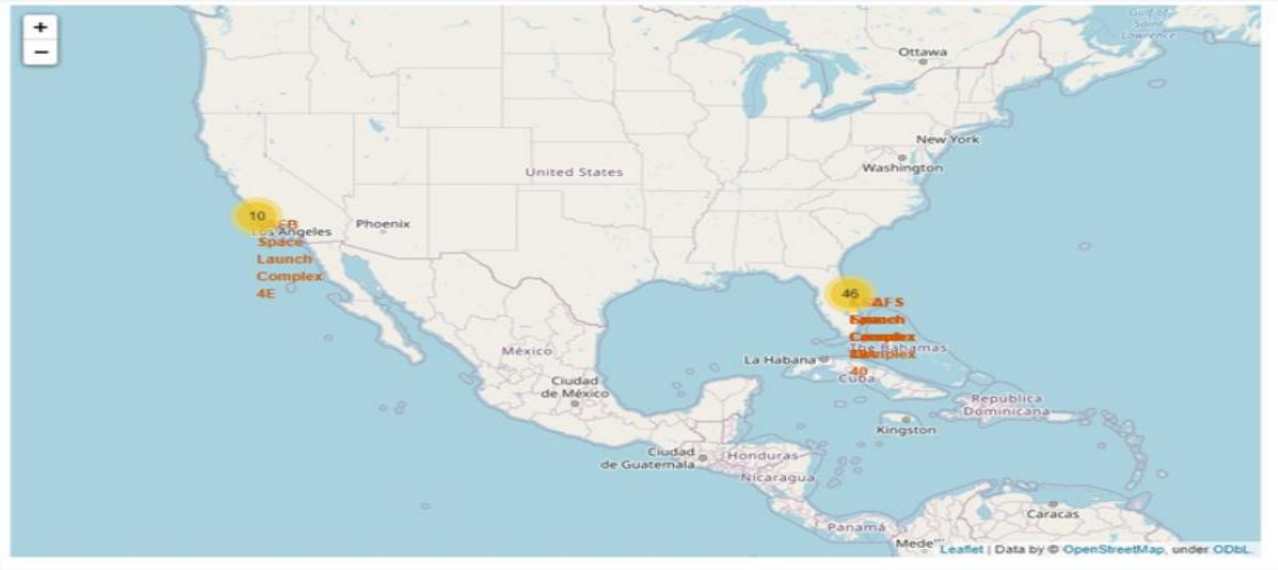
Launch Sites Proximities Analysis

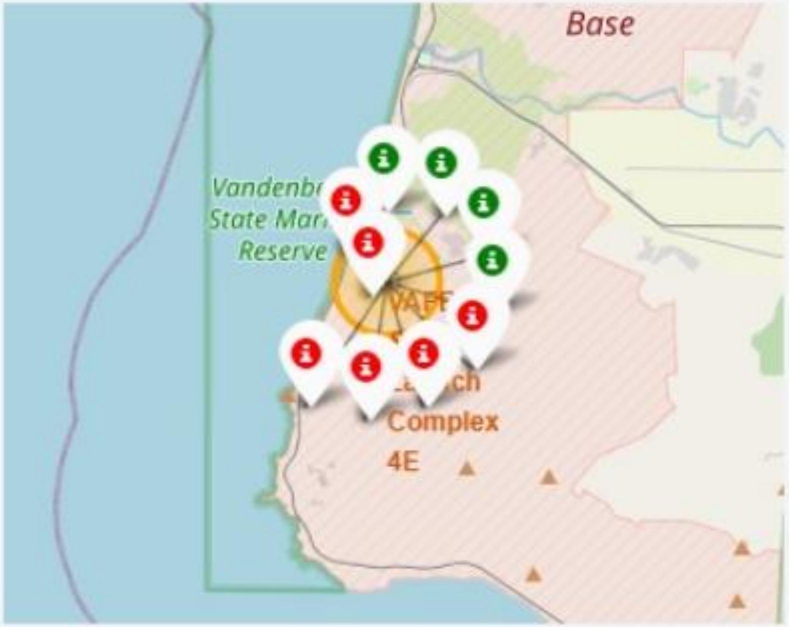
All Launch sites



All launch sites are close to coast and away from public areas.

Successful and failed launches





- Red are the failed launches.
- Green are the successful launches.

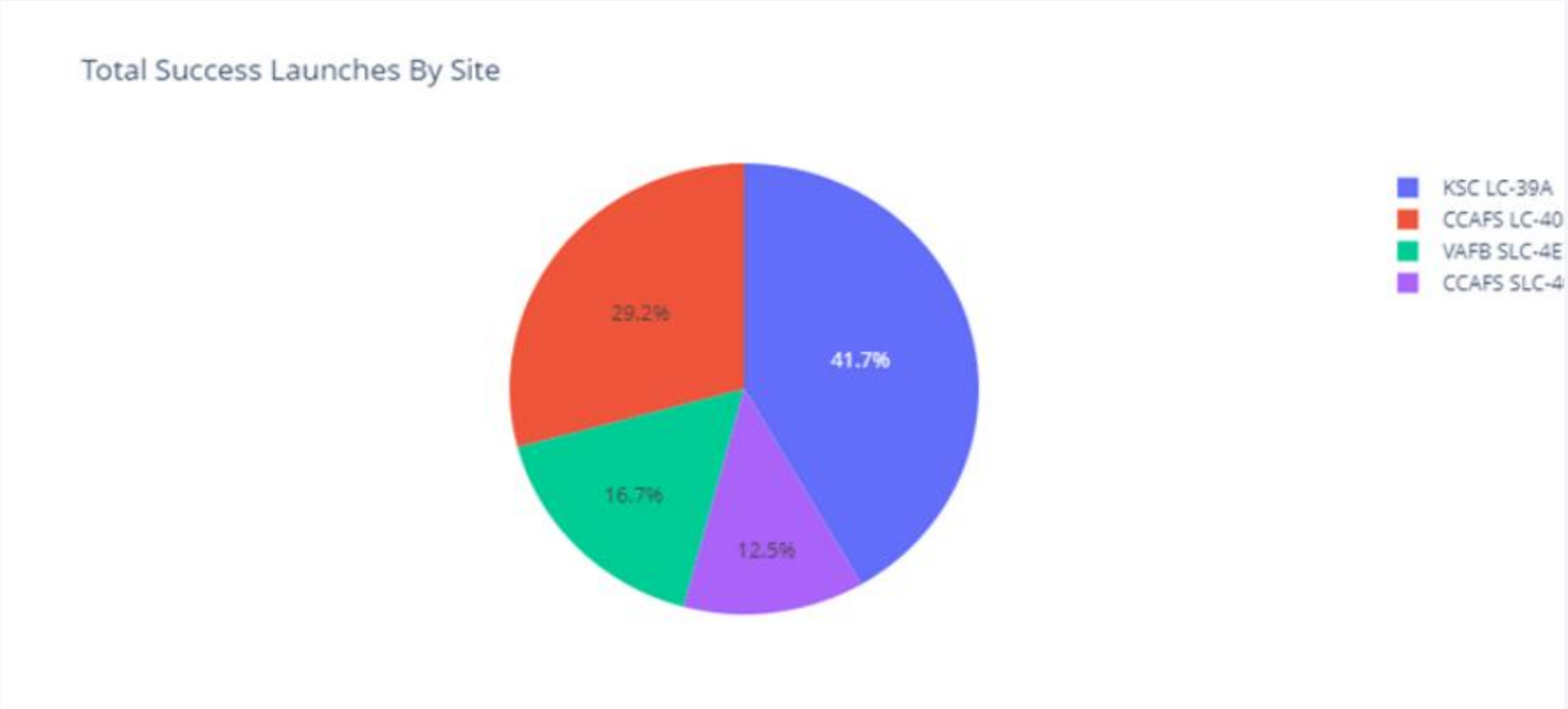
Launch sites and Nearby locations

All launch sites are near railways, roads, highways, and coastlines, but it is far from inhabitant areas.

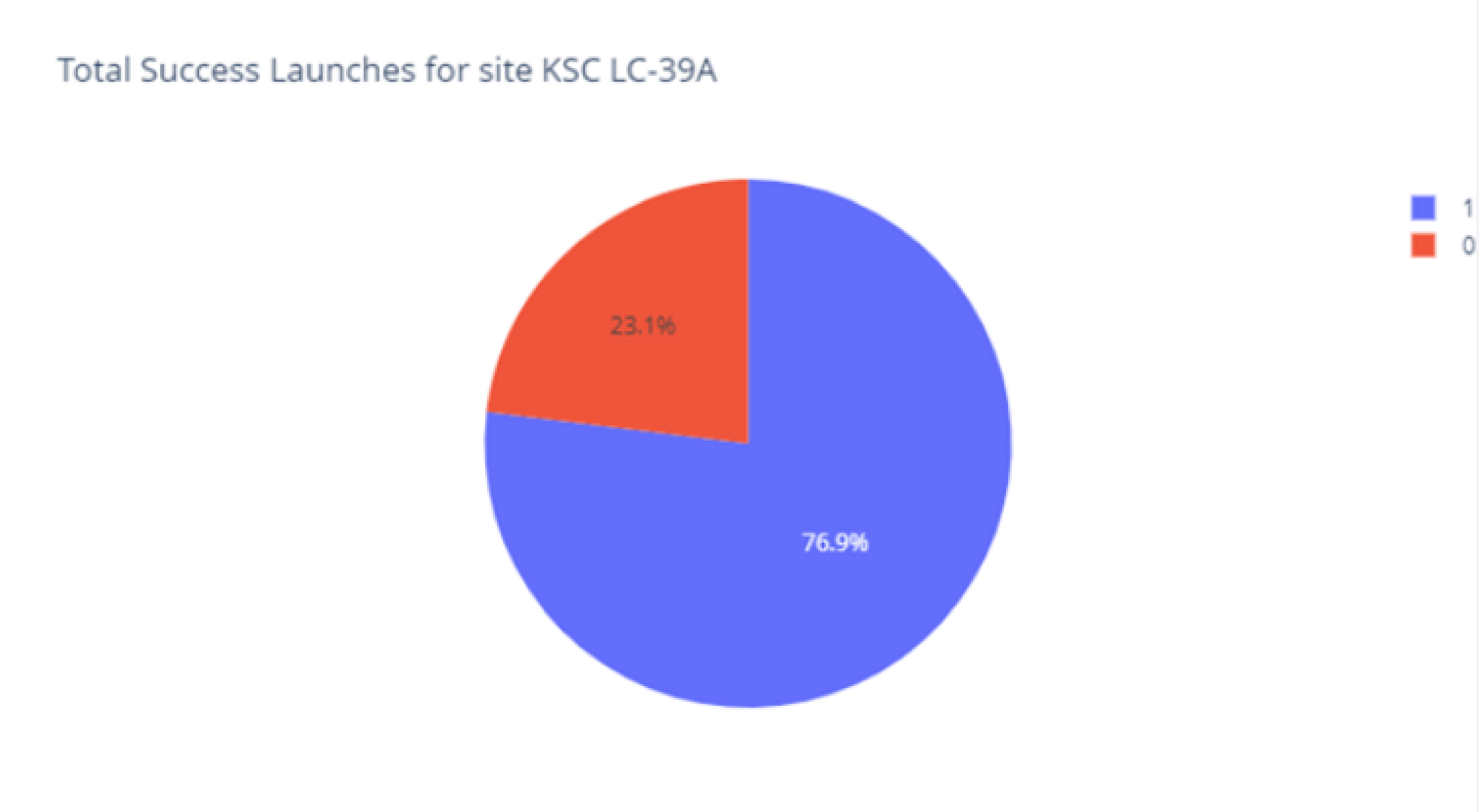
Section 4

Build a Dashboard with Plotly Dash

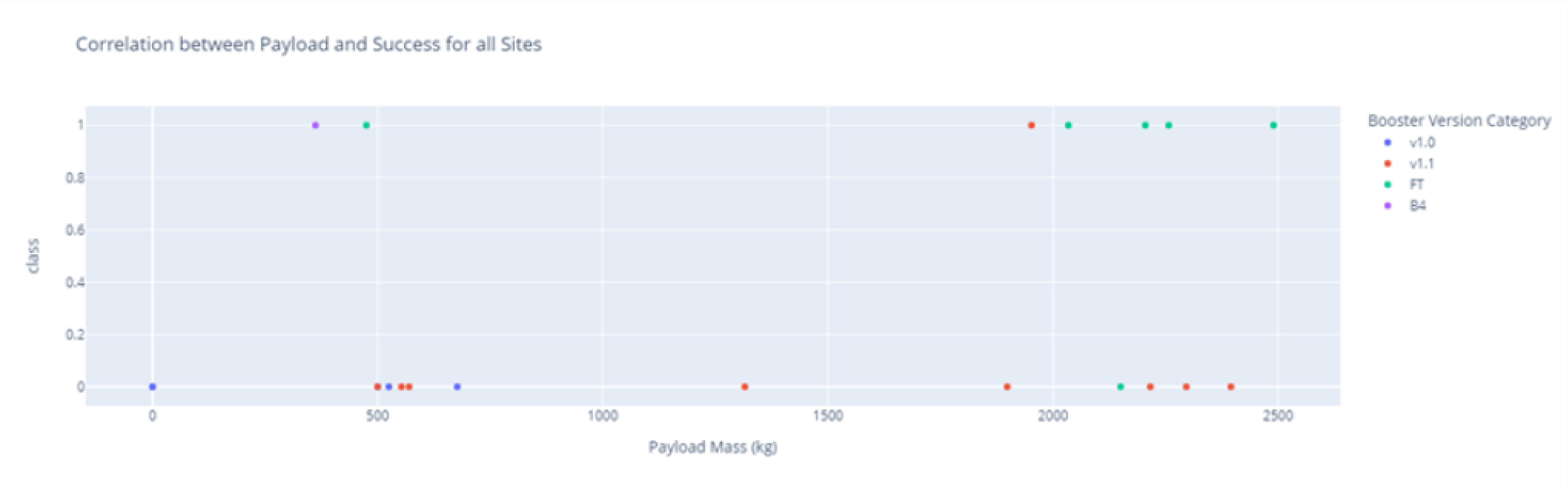
Success launches number per site



KSC LC-39A launching site



Payload Vs launch Outcome

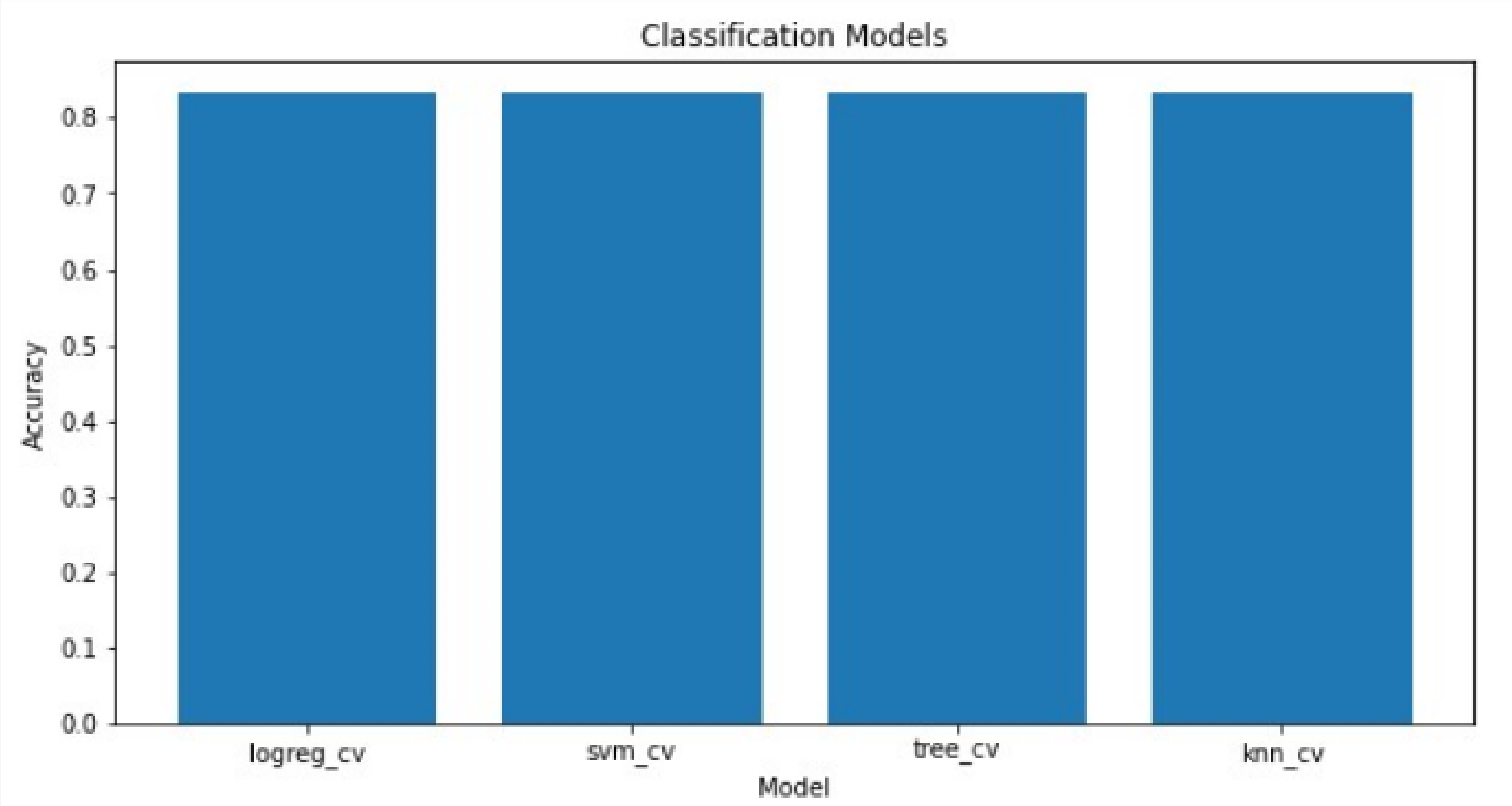


- Most successful launches were with a payload between 2500 kg and 5000 kg.
- Moat failed launches were with payload less than 2500 kg.

Section 5

Predictive Analysis (Classification)

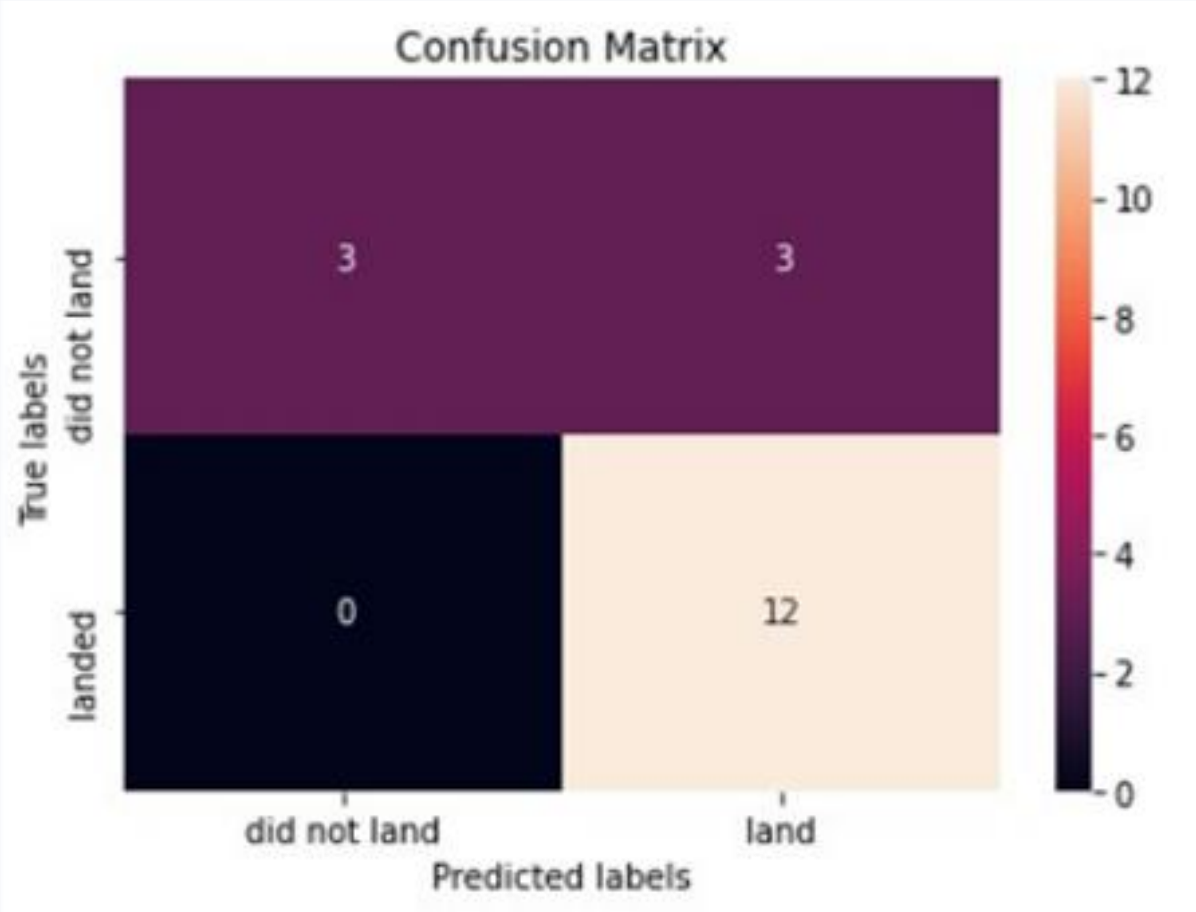
Classification Accuracy



- All models had similar accuracy.

Confusion Matrix

- Confusion matrix is the same for all model



Conclusions

- Any of our Machine Learning models can be used for predicting if the first stage of our rocket will land successfully and it can help with predicting the cost.

Appendix

Project GitHub link:

Data Science Cap Stone

Thank you!