Winning Space Race with Data Science

Priyanshu Sharma
10-04-2022

Outline

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

Executive Summary

- Data was collected from Spacex API and it was cleaned and preprocessed and converted into a csv file
- Observations were made related to input parameters
- Lunch Sites are closer to coastlines
- Payload Mass highly affects landing outcome
- First Successful Ground landing was made in year 2015, nearly after 5 years of SpaceX launching rockets.
- Classification model with an accuracy of 84% which can predict whether a rocket will land or not.

Introduction

- 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
- We want to predict if the Falcon 9 first stage will land successfully or not.

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

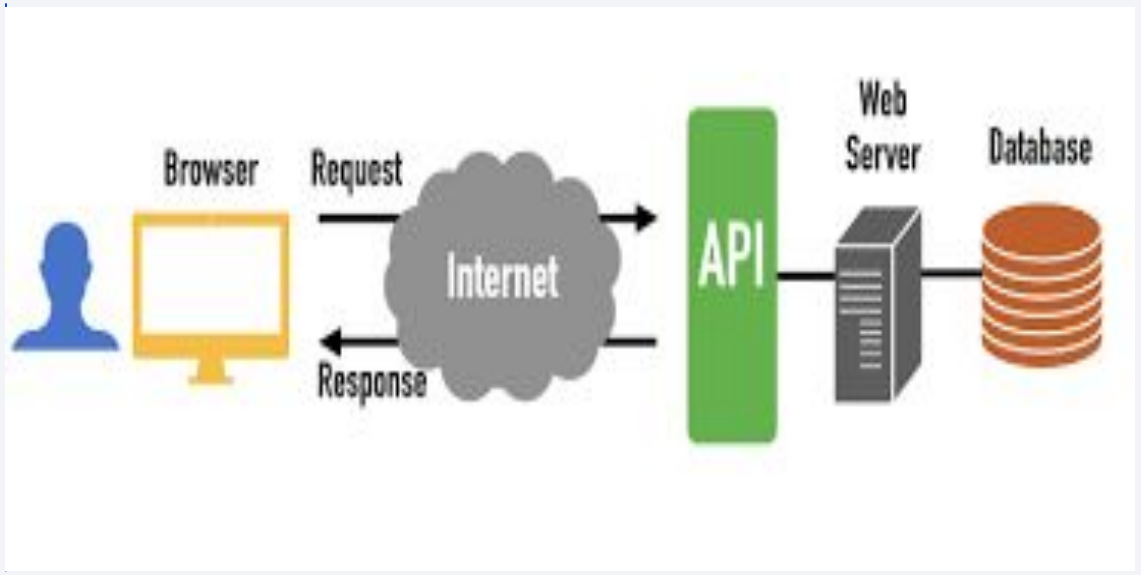
Data Collection

- DataSet was collected from Spacex api and Wikipedia
- Primary webscarping was performed using beautifulsoup on wikipedia to get falcon 9 launch related info
- To get more information about launches, spacex api was used.
- Data was parsed and stored in a dataframe after cleaning.

Data Collection – SpaceX API

- Request To SpaceX API
- Cleaning the processed Data

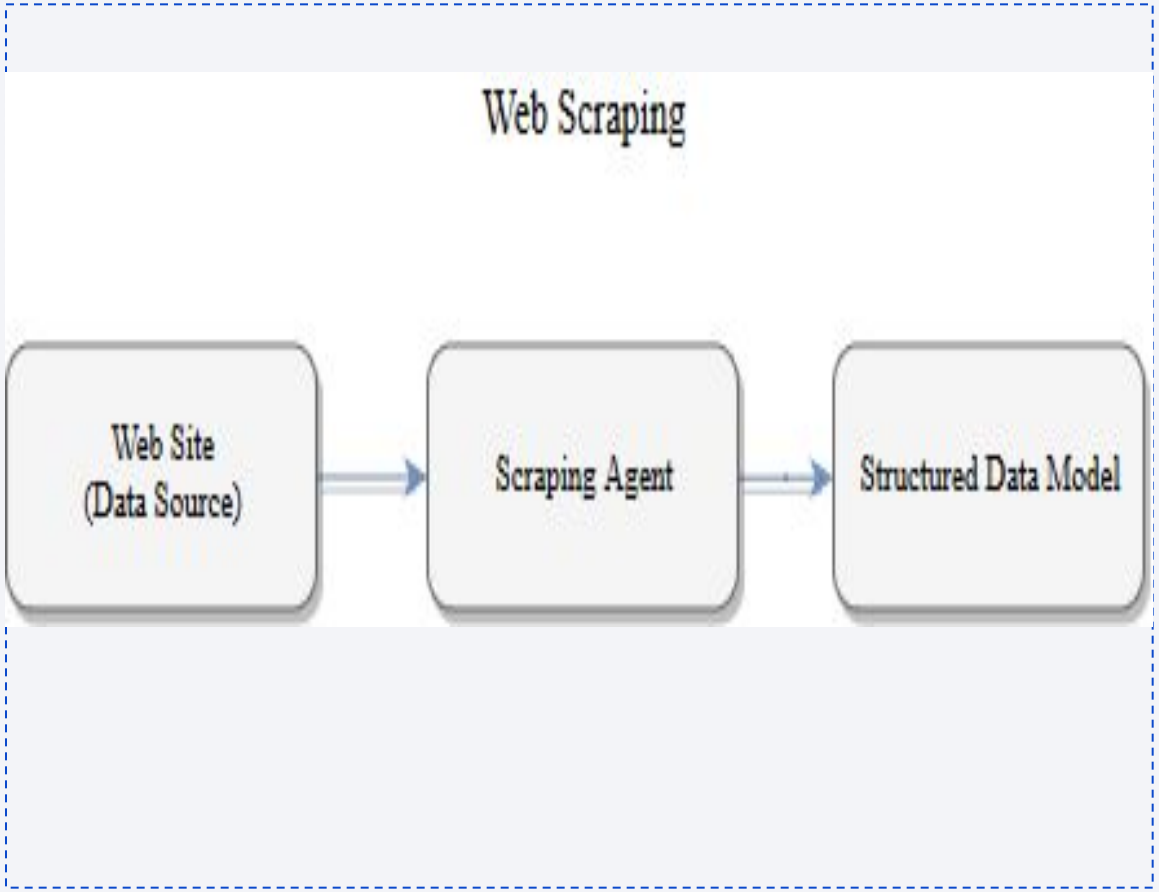
- Link :- https://github.com/xambert/Data-Science/blob/main/Week%201/jupyter_labs_spacex_data_collection_api.ipynb



SpaceX Api Call

Data Collection - Scraping

- Web Scraping done using BeautifulSoup
- Parsed the Falcon 9 data from wikipedia.
- Link :- https://github.com/xambert/Data-Science/blob/main/Week%201/jupyter_labs_webscraping.ipynb



Data Wrangling

- Web Scraping Done using BeautifulSoup
- Parsed the data and made a dataframe out of Lunch_Sites,Orbit,Payload Mass etc.
- Link :- https://github.com/xambert/Data-Science/blob/main/Week%201/labs_jupyter_spacex_Data_wrangling.ipynb

EDA with Data Visualization

- Scatter plots were used to find relationship between various parameters and Bar Chart for Success ratio of landing Outcomes
- Link :-https://github.com/xambert/Data-Science/blob/main/Week%202/jupyter_labs_eda_dataviz.ipynb

EDA with SQL

- Following Queries were performed
  - All Launch Site Names
  - 5 launch SItes Starting with CCA
  - Successful and Failure Mission Outcome
  - 2015 Launch Records
  - Landing Outcomes in 2010 - 2017
- Link :- https://github.com/xambert/Data-Science/blob/main/Week%202/jupyter_labs_eda_sql_coursera.ipynb

Build an Interactive Map with Folium

- Created Circle and marker objects to Folium map
- Circles gave us the radius i.e. range of sites and marker objects help pinpoint exact locations of sites on the map.

- Link :- https://github.com/xambert/Data-Science/blob/main/Week%203/lab_jupyter_launch_site_location_(1).ipynb

Build a Dashboard with Plotly Dash

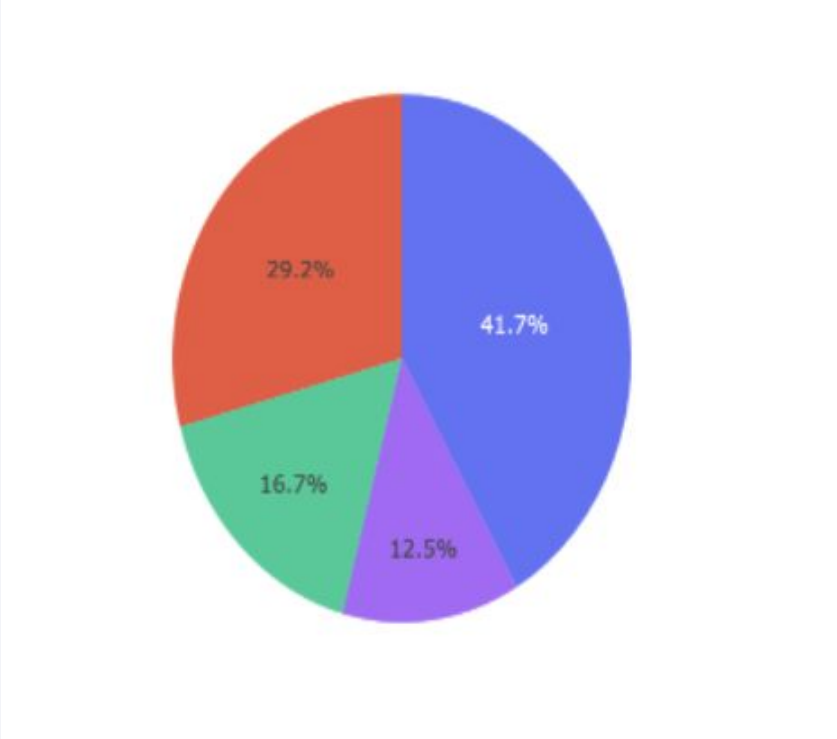
- Pie chart for Success Rate of
  - All sites
  - Individual Site
    To have a visualization Success Rate
- Scatter Plot of Payload Mass vs Launch Outcome based on Booster Version to find if there exist a relation between them
- Link :- https://github.com/xambert/Data-Science/tree/main/Week%203

Predictive Analysis (Classification)

- Predictive Models were built using Classification algorithms
  - Logistic Regression : Train Accuracy : 84.2%
  - SVM : Train Accuracy : 84.4%
  - Decision Tree : Train Accuracy : 84.7%
  - KNN : Train Accuracy : 84.5%
- Link :- https://github.com/xambert/Data-Science/tree/main/Week%204

Results

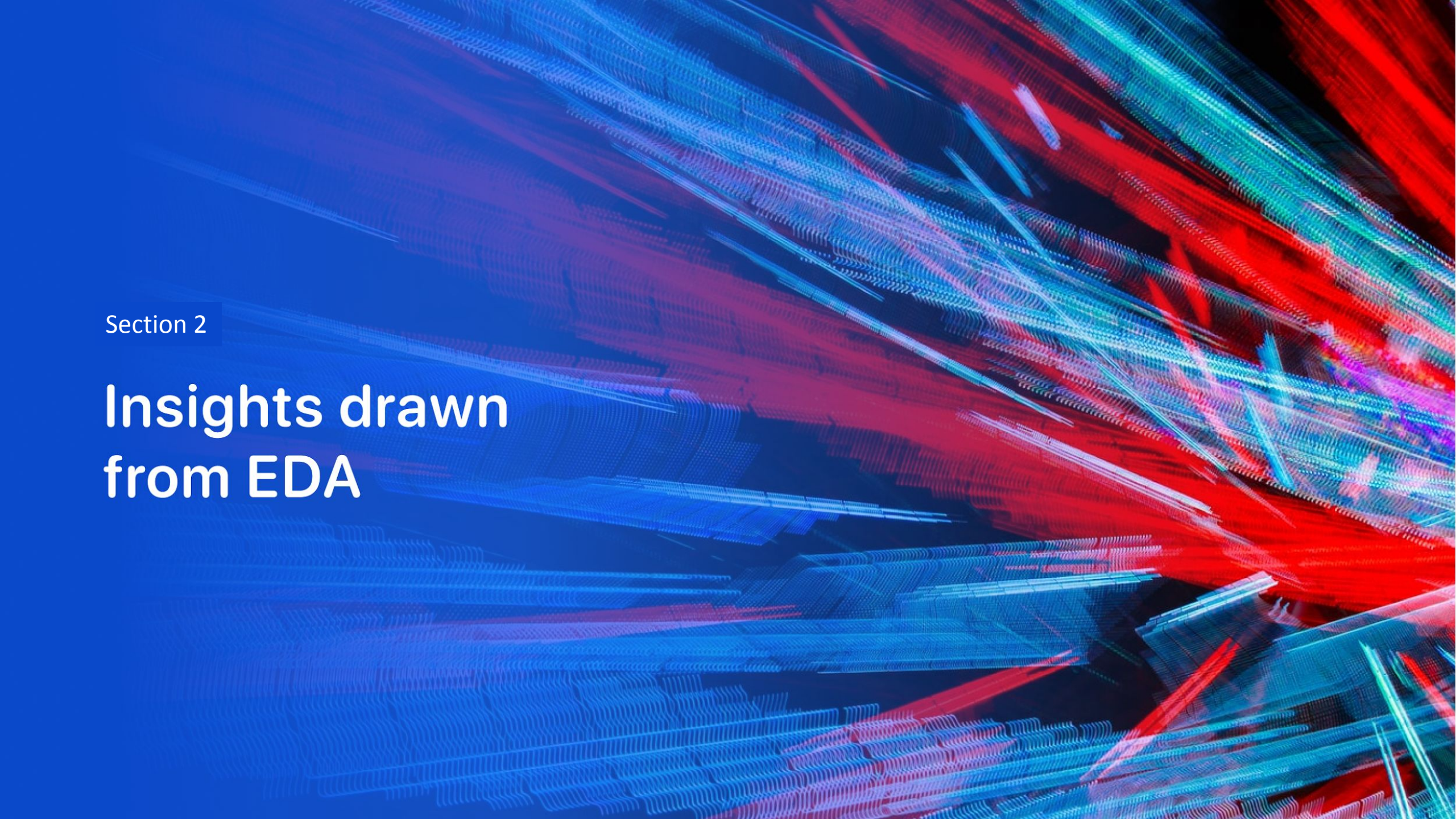
- In Exploratory Data Analysis, we found -
  - Four unique Launch Site
  - Higher Success Rate for lesser Payload Mass
  - Failure Rate was 100 % before 2013
  - First successful ground landing was in year 2015
  - Launch sites were in close proximity to coastlines
- Model has an accuracy of 83.3%.



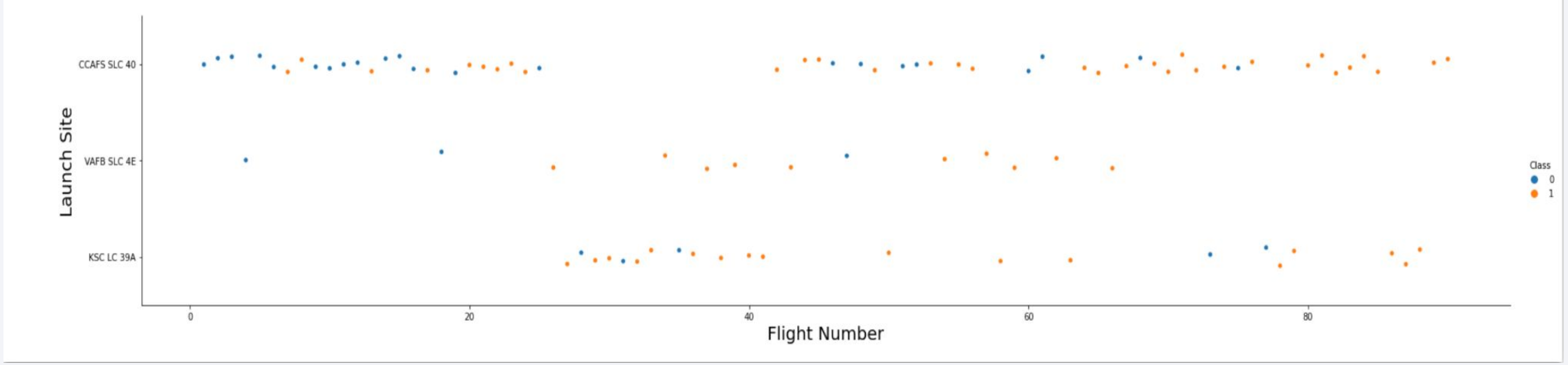
Success Rates for different sites.

Section 2

Insights drawn from EDA

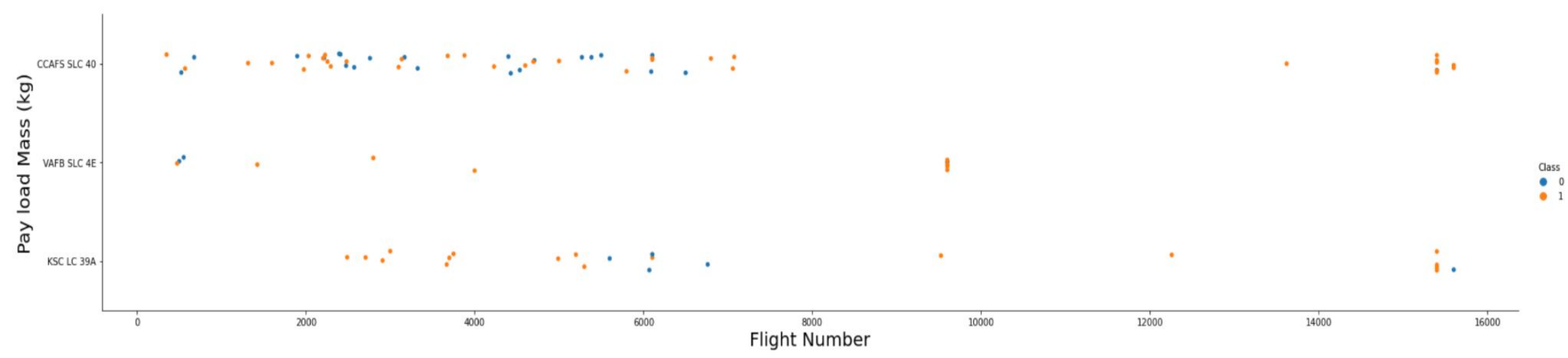


Flight Number vs. Launch Site



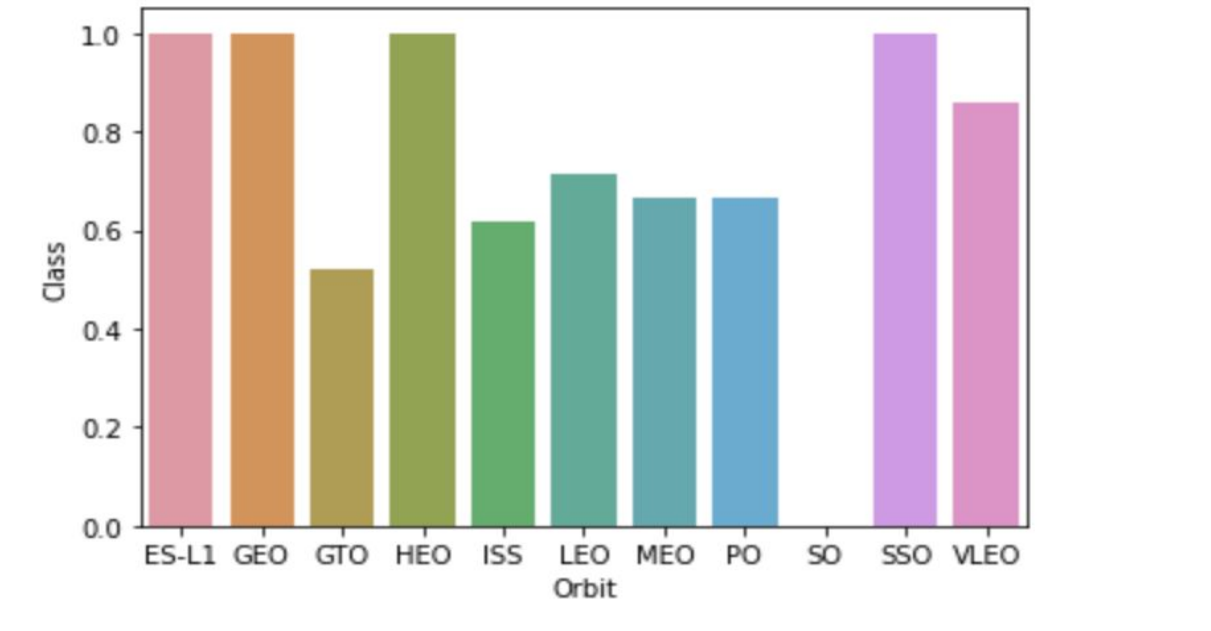
KSC LC-39A has more successful launches than CCAFS LC-40
VAFB SLC 4E has the highest success rate
As Flight Number **increases,** Chances of successful landing **increases.**

Payload vs. Launch Site



CCAFS LC-40 has 100% success rate for launches above 12000 kg.
VAFB SLC 4E has no launches with payload mass greater than 10000 kg.

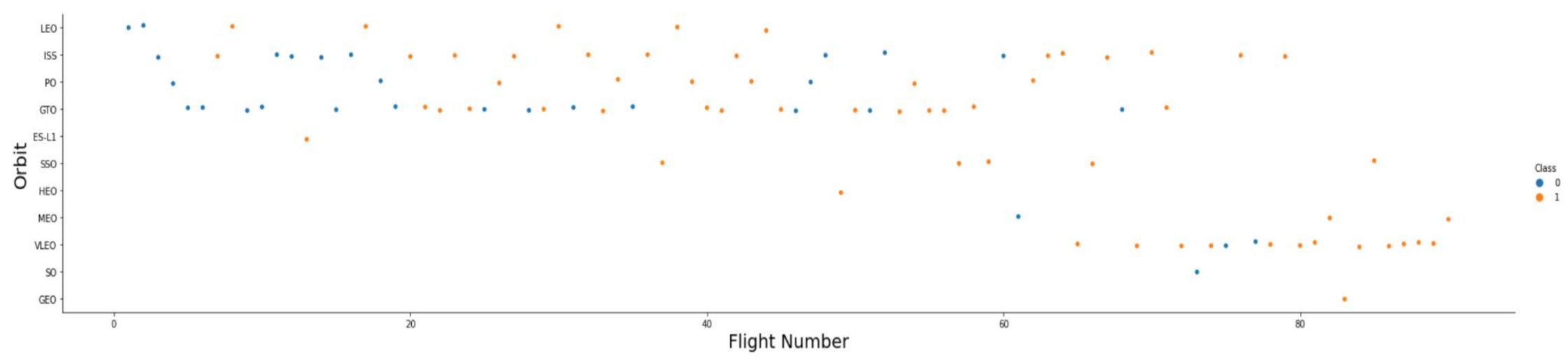
Success Rate vs. Orbit Type



We can see that ES-L1, GEO,HEO and SSO have 100% success rate

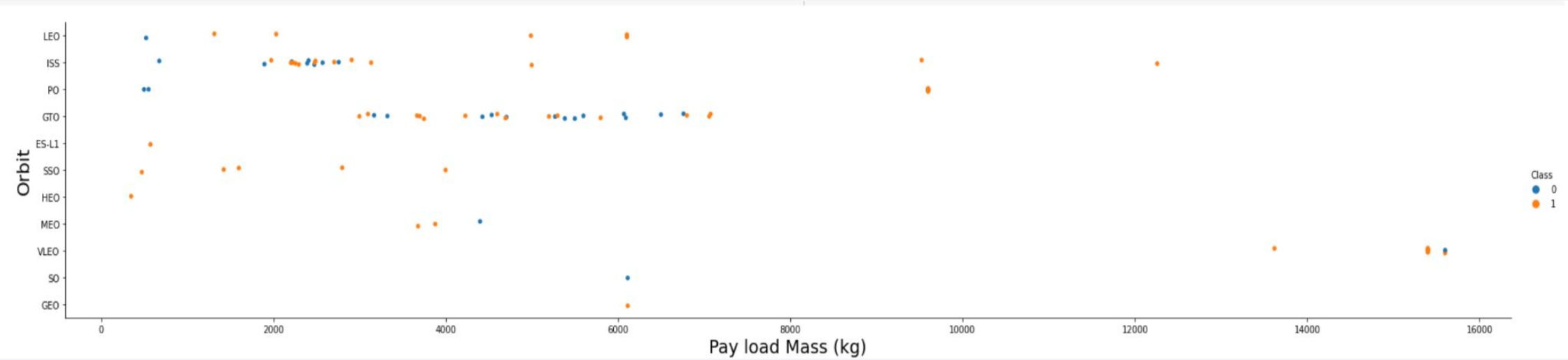
SO has lowest success rate

Flight Number vs. Orbit Type



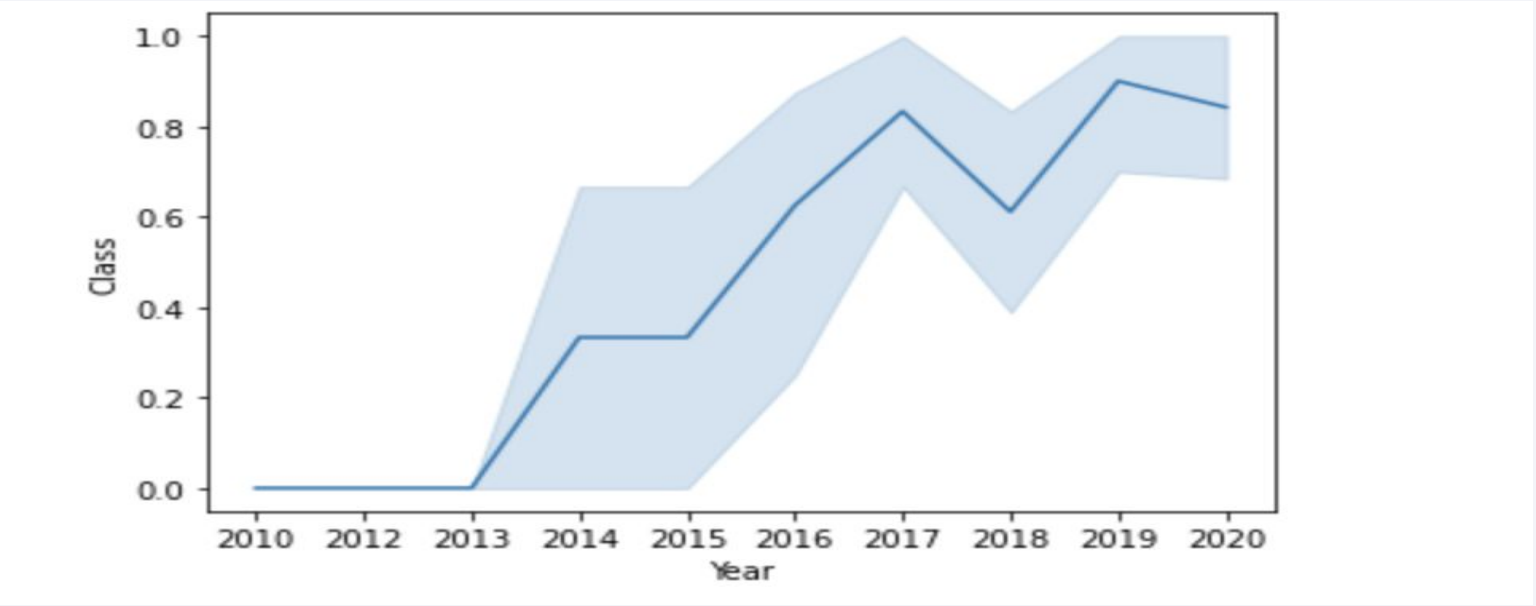
For **LEO** orbit, **success rate seems to be dependent on Flight Number.**

Payload vs. Orbit Type



LEO,ISS,PO orbits have increasing successful launches with respect to payload mass whereas **SSO,HEO,ES-LI** have successful launches with payload mass less than 6000

Launch Success Yearly Trend



Success rate began to improve since 2013, before 2013 failure rate 100%.

All Launch Site Names

```
%sql SELECT DISTINCT(LAUNCH_SITE) FROM SPACEXTBL;
```

```
 * ibm_db_sa://frc90031:***@824dfd4d-99de-440d-999
Done.
```

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

SPACEX has 4 different Launch Sites

Launch Site Names Begin with 'CCA'

```
%sql SELECT * FROM SPACEXTBL WHERE LAUNCH_SITE LIKE 'CCA%' LIMIT 5
```

```
 * ibm_db_sa://frc90031:***@824dfd4d-99de-440d-9991-629c01b3832d.bs2io90108kqb1od8lcg.databases.appdomain.cloud:30119/bludb
Done.
```

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	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	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

First 5 Records with Launch Sites beginning with CCA.

We can see that first 5 launches were unsuccessful.

Total Payload Mass

```
%%sql
SELECT CUSTOMER, SUM(PAYLOAD_MASS__KG_) AS TOTAL_PAYLOAD FROM SPACEXTBL
GROUP BY CUSTOMER
HAVING CUSTOMER = 'NASA (CRS)'
```

```
 * ibm_db_sa://frc90031:***@824dfd4d-99de-440d-9991-629c01b3832d.bs2io90
Done.
```

customer	total_payload
NASA (CRS)	45596

Total Payload Mass from Customer NASA(CRS) is 45596 kg.

Average Payload Mass by F9 v1.1

```
%%sql
SELECT AVG(PAYLOAD_MASS__KG_) AS AVERAGE_PAYLOAD FROM SPACEXTBL
where BOOSTER_VERSION LIKE '%F9 v1.1%'
```

```
 * ibm_db_sa://frc90031:***@824dfd4d-99de-440d-9991-629c01b3832d.b
Done.
```

average_payload
2534

We can see that average payload mass by **F9 v1.1** is **2534 kg**.

First Successful Ground Landing Date

```
%%sql
SELECT MIN(Date),LANDING__OUTCOME FROM SPACEXTBL
GROUP BY LANDING__OUTCOME
HAVING LANDING__OUTCOME like 'Success (ground pad)'
```

```
 * ibm_db_sa://frc90031:***@824dfd4d-99de-440d-9991-
Done.
     1        landing__outcome
2015-12-22 Success (ground pad)
```

First successful Ground Landing was done on **2015-12-22**

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%%sql
SELECT BOOSTER_VERSION FROM SPACEXTBL
WHERE LANDING__OUTCOME LIKE 'Success (drone ship)'
and PAYLOAD_MASS__KG_ BETWEEN 4001 AND 5999
```

```
 * ibm_db_sa://frc90031:***@824dfd4d-99de-440d-9991
Done.
```

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

Here is the list of booster version with Payload mass between 4000 kg and 6000kg and successful drone ship landing

Total Number of Successful and Failure Mission Outcomes

```
%%sql
SELECT MISSION_OUTCOME,COUNT(MISSION_OUTCOME)
FROM SPACEXTBL
GROUP BY MISSION_OUTCOME
```

```
 * ibm_db_sa://frc90031:***@824dfd4d-99de-440d-
Done.
```

mission_outcome	2
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Mission Outcome has a success rate of almost **99%**

Boosters Carried Maximum Payload

Here is the list of the names of the booster which have carried the maximum payload mass

```
%%sql
SELECT DISTINCT(BOOSTER_VERSION)
FROM SPACEXTBL
WHERE PAYLOAD_MASS__KG_ =
(SELECT MAX(PAYLOAD_MASS__KG_)
FROM SPACEXTBL)
```

```
 * ibm_db_sa://frc90031:***@824df
Done.
```

booster_version
F9 B5 B1048.4
F9 B5 B1048.5
F9 B5 B1049.4
F9 B5 B1049.5
F9 B5 B1049.7
F9 B5 B1051.3
F9 B5 B1051.4
F9 B5 B1051.6
F9 B5 B1056.4
F9 B5 B1058.3
F9 B5 B1060.2
F9 B5 B1060.3

2015 Launch Records

```
%%sql
SELECT LAUNCH_SITE, BOOSTER_VERSION, LANDING__OUTCOME
FROM SPACEXTBL
WHERE LANDING__OUTCOME LIKE 'Failure (drone ship)'
AND DATE LIKE '2015%'
```

```
 * ibm_db_sa://frc90031:***@824dfd4d-99de-440d-9991-629
Done.
```

launch_site	booster_version	landing__outcome
CCAFS LC-40	F9 v1.1 B1012	Failure (drone ship)
CCAFS LC-40	F9 v1.1 B1015	Failure (drone ship)

We can conclude that Site **CCAFS LC-40** has failure rate of **100%** for **drone ship** landing in the year **2015**

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

```
%%sql
SELECT LANDING__OUTCOME AS OUTCOME, COUNT(LANDING__OUTCOME) AS COUNT
FROM SPACEXTBL
WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20'
GROUP BY LANDING__OUTCOME
HAVING LANDING__OUTCOME IN ('Failure (drone ship)', 'Success (ground pad)')
ORDER BY COUNT DESC
```

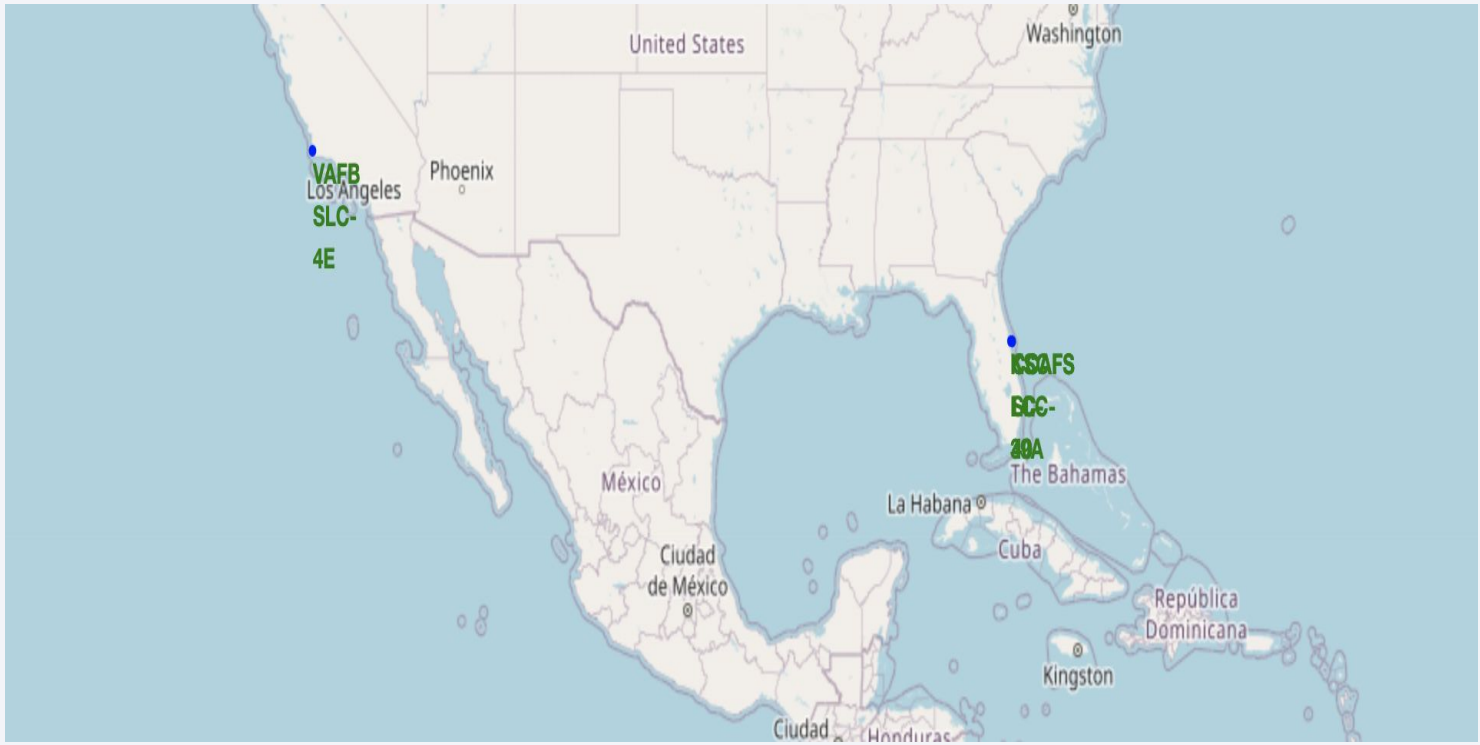
```
 * ibm_db_sa://frc90031:***@824dfd4d-99de-440d-9991-629c01b3832d.bs2io90108k
Done.
```

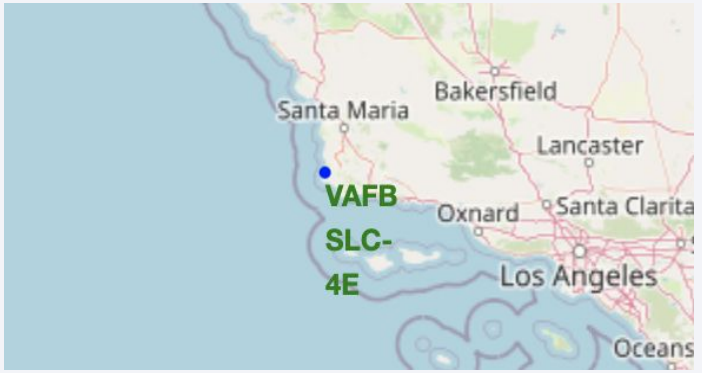
outcome	COUNT
Failure (drone ship)	5
Success (ground pad)	3

Failure Rate of drone ship is more than Success Rate of ground pad


Section 3
Launch Sites
Proximities Analysis

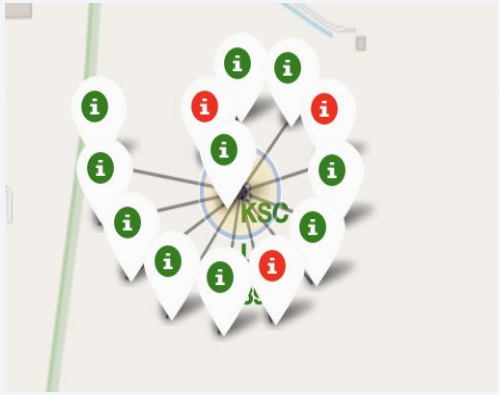
Launch Sites

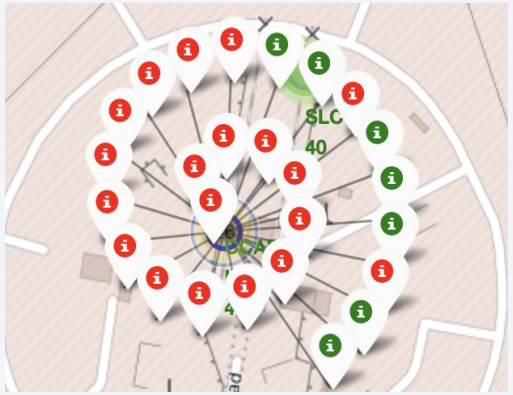


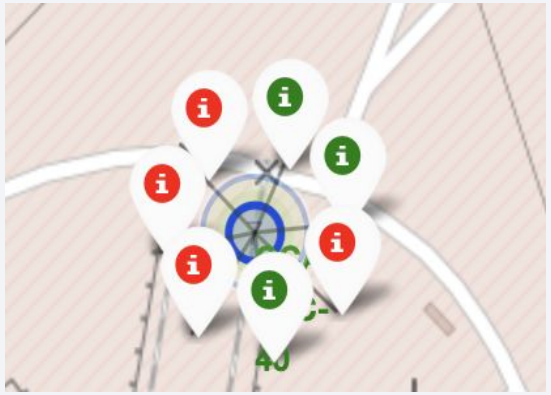


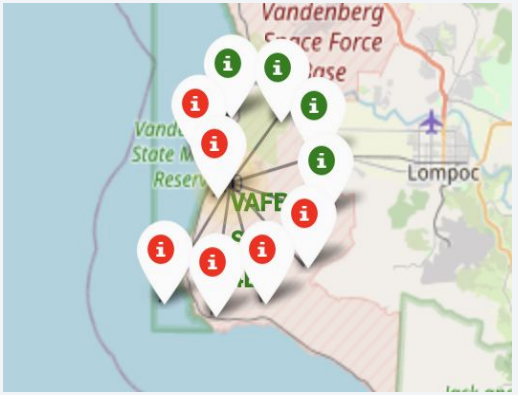
Launch Sites are in close proximities to the equator and coastlines

Lanch Outcomes









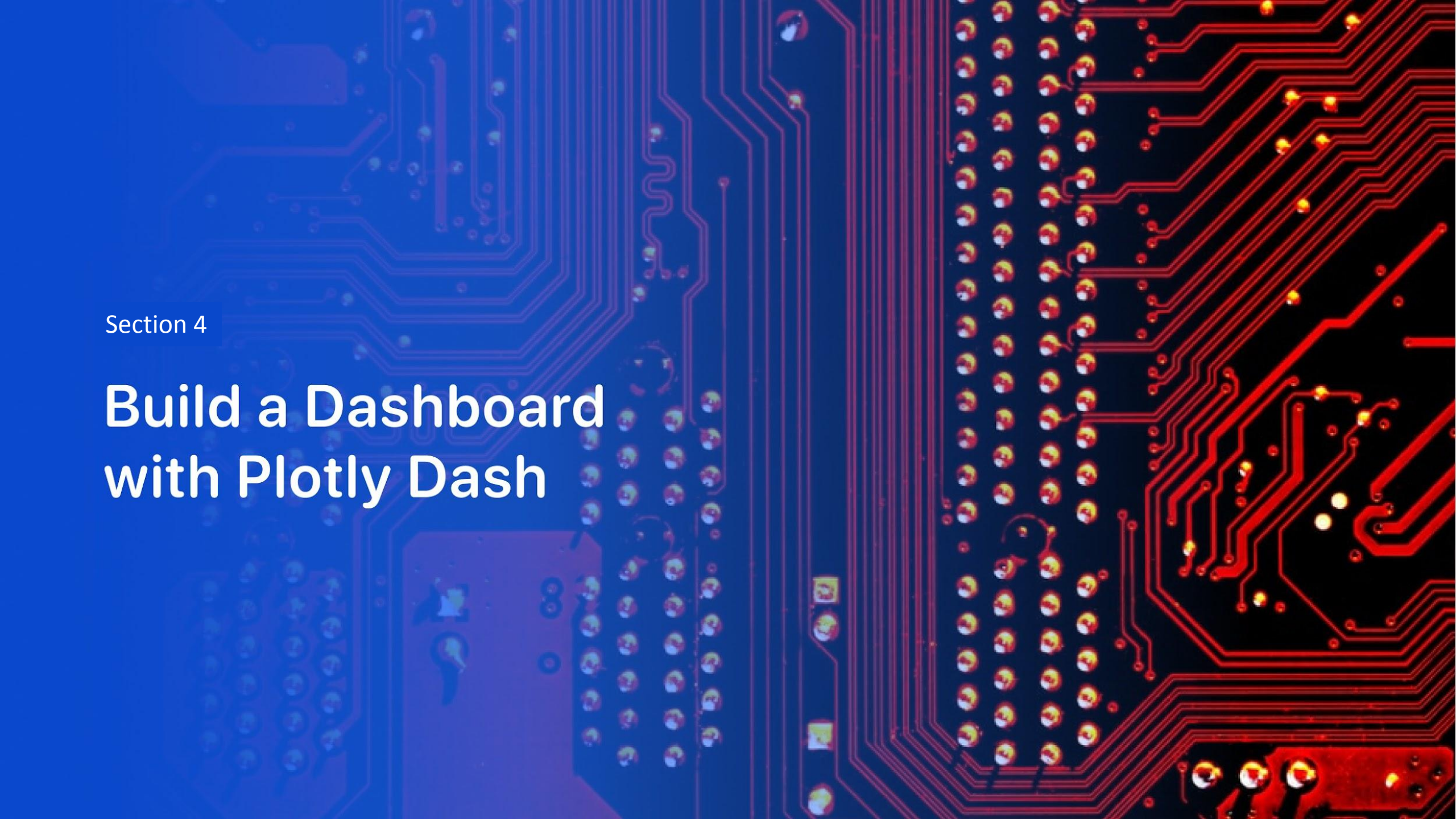
1. KSC LC-39A
2. CCAFS LC 40
3. CCAFS SLC 40
4. VAFB SLC-4E

Site **KSC LC_39A** has most successful launches whereas **CCAFS LC-40** having most launches with failures in major.

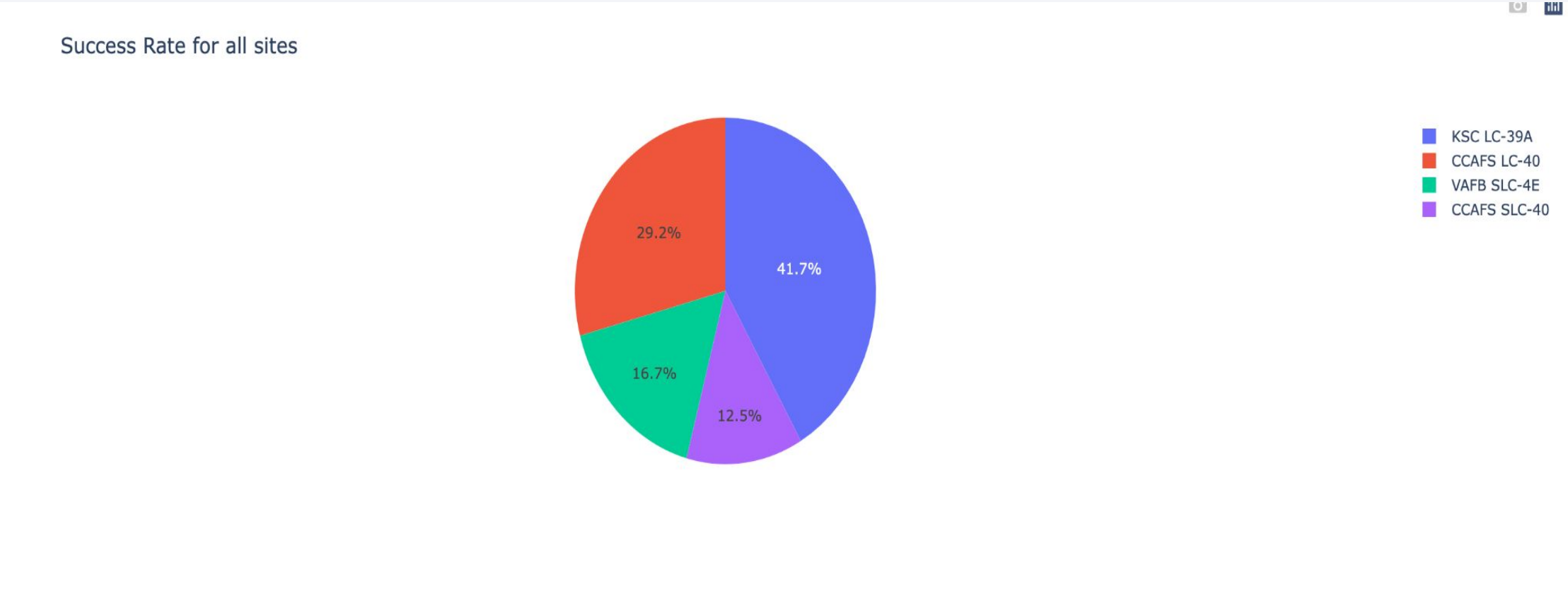
Thus, **KSC LC-39A** seems the optimal fit for Launch Site

Proximity Analysis

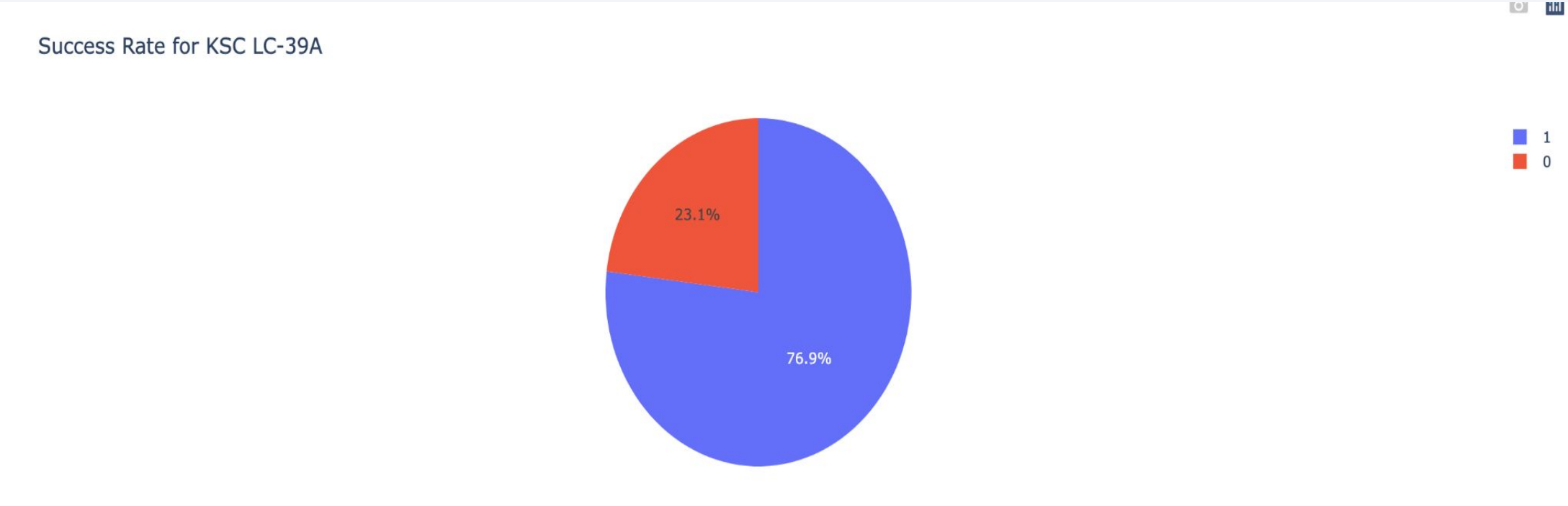
As we can see, Launch sites are in close proximities to coastlines


Section 4
Build a Dashboard with Plotly Dash

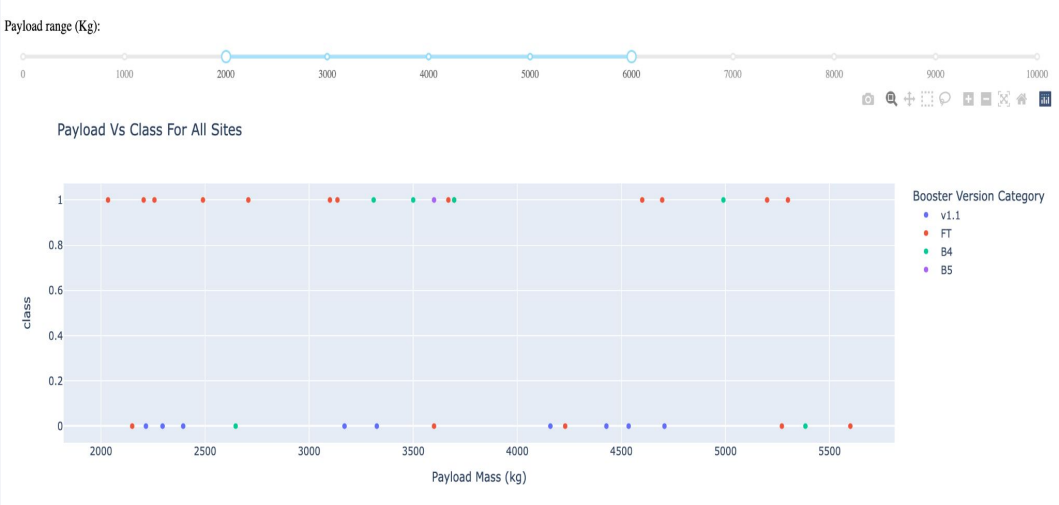
Success Rate for All Sites

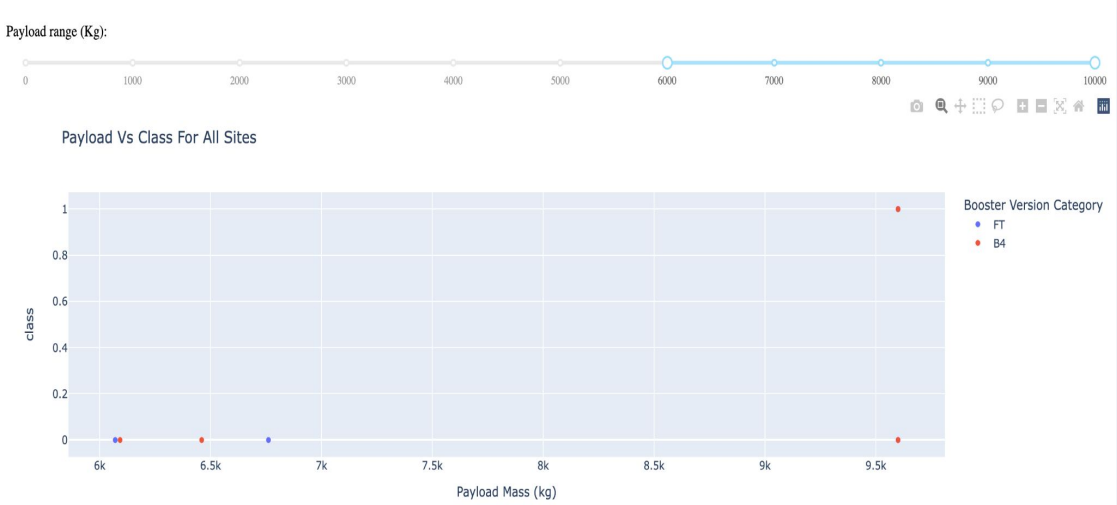


Site with Highest launch Success Ratio



Payload vs Launch Outcome





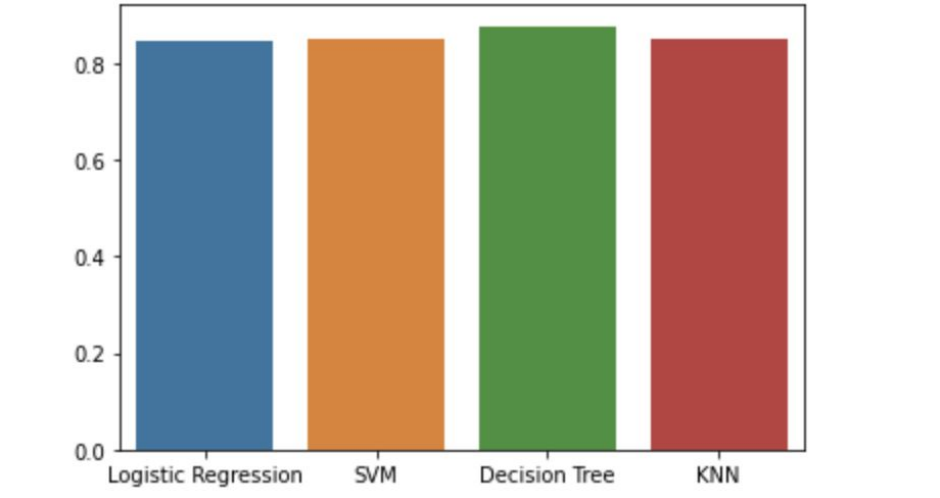
Booster Version FT seems to be the most successful in 2000 - 6000 kg range

Failure Rate is very high for Launches with Payload Mass above 6000 kg

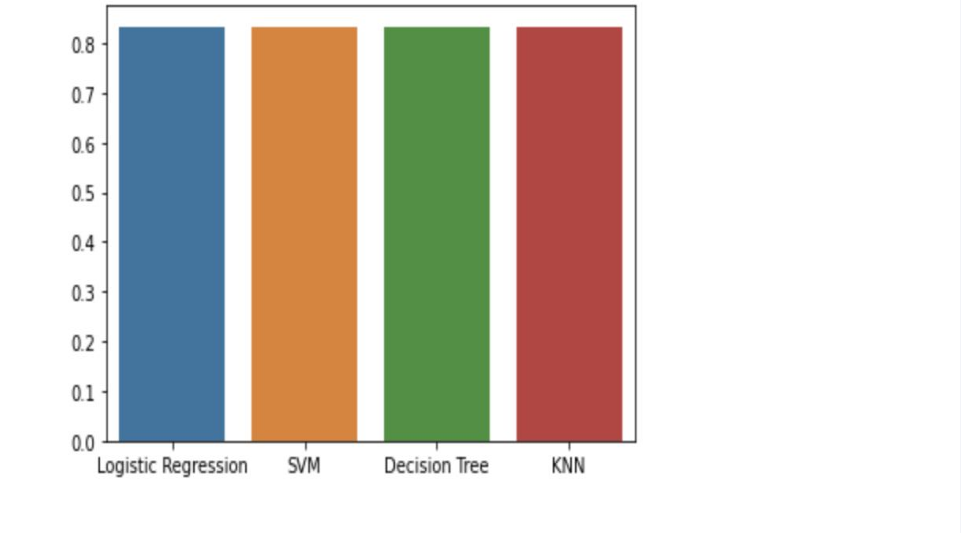
Section 5

Predictive Analysis (Classification)

Classification Accuracy

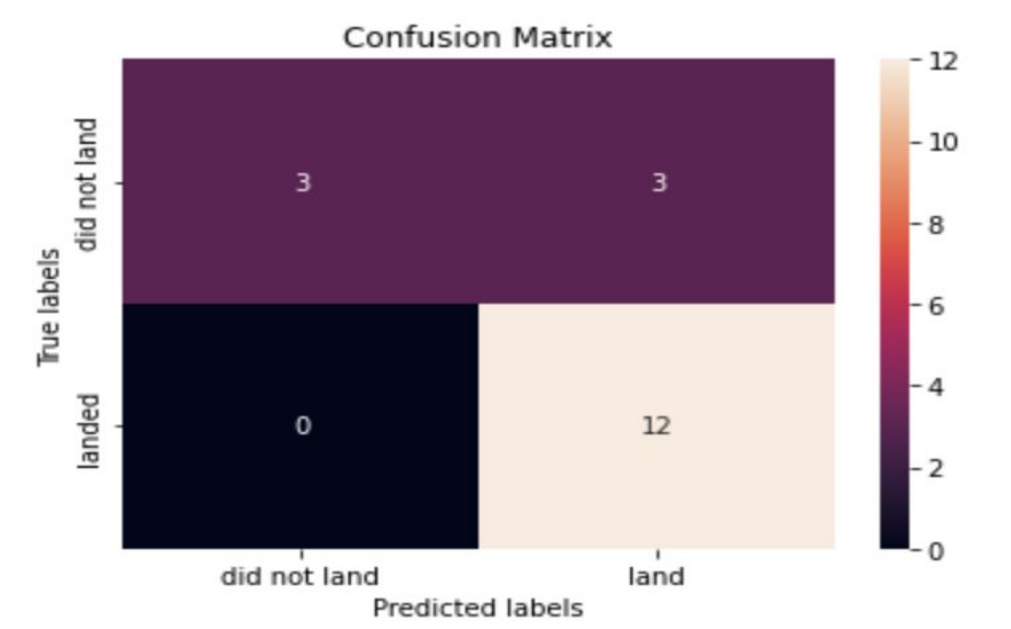


Training Accuracy



Test Accuracy

Confusion Matrix



Model is capturing True Positives and Negatives and False Positives as well.

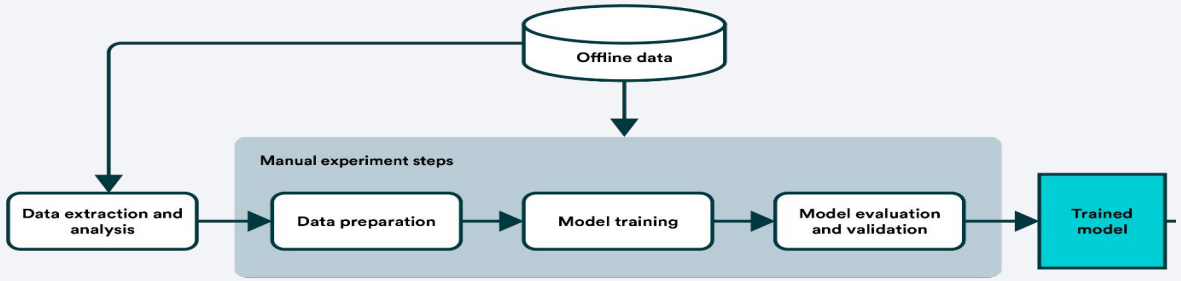
Conclusions

- KSC LC-39A is the best launch site
- Success Rate is higher for lower Payload Mass
- Launch Sites are in close proximity to coastlines
- LEO,ISS,PO orbits have increasing successful launches with respect to payload mass whereas SSO,HEO,ES-LI have successful launches with payload mass less than 6000
- FT is the most successful Booster Version
- Greater Flight number conveys higher chances of successful landing.

Appendix

- Link to whole project :- https://github.com/xambert/Data-Science

Flowchart of Machine Learning Pipeline used



ML model Pipeline

Thank you!