Winning Space Race with Data Science

P N Narmadha
23-08-2022

Outline

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

Executive Summary

- Summary of methodologies

  End-end Process involved in this prediction task follows below :

  - Data Collection
  - Data Processing
  - EDA (Exploratory Data Analysis)
  - Data Preprocessing
  - Building Model
  - Validation

- Summary of all results

  - Prediction accuracy score is 83% which is reasonably good which adds some business value

Executive Summary

- Summary of all results
  - Accuracy is reasonable with score of 83%
  - Features related to successful launches are identified .
  - Success and failure ration calculated on various aspects

Introduction

- **Project background and context:**

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.

- **Business Benefit :** This cost information can be used if an alternate company wants to bid against SpaceX for a rocket launch.

- **Problem Statement :**

we will predict if the Falcon 9 first stage will land successfully or not.

Section 1

Methodology

Methodology

- Data collection methods:

  Collected data from various sources such as SpaceX Rest API end points and falcon 9 wiki page of falcon 9 launch

- Perform data wrangling

  Defined new attribute which indicates landing is successful or not

- Perform exploratory data analysis (EDA) by constructing plots and SQL queries .

  Using EDA , Identified what attributes are related to successful landing

- Perform interactive visual analytics using Folium and Plotly Dash App .

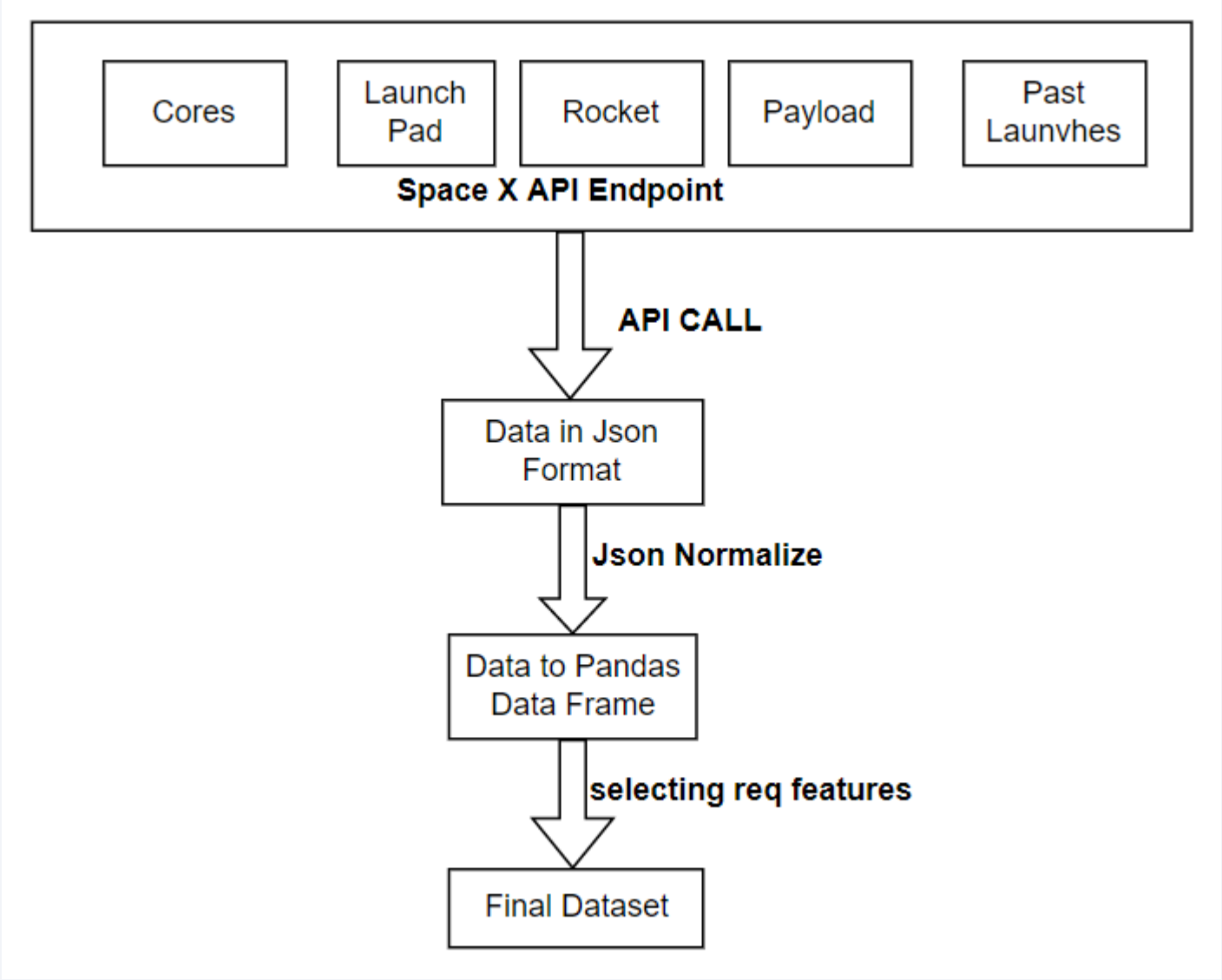
- Perform predictive analysis using classification models

Data Collection

- **Describe how data sets were collected.**
  - Source of data is SpaceX rest API end points and Wikipedia page of falcon 9
  - The SpaceX Rest API has various end points like cores, past launch, capsule,rocket, payload
  - Falcon 9 Historical Launch Data is collected through webscrapping technique
  - API data is download in json format which is handled using json_normalize module
  - Webscrapped data is in HTML format which is handled using Beautiful soup python library
  - Then data is converted to pandas data frame using above library packages .
  - Final Dataset is formed by gathering required features/columns from all the data sets collected

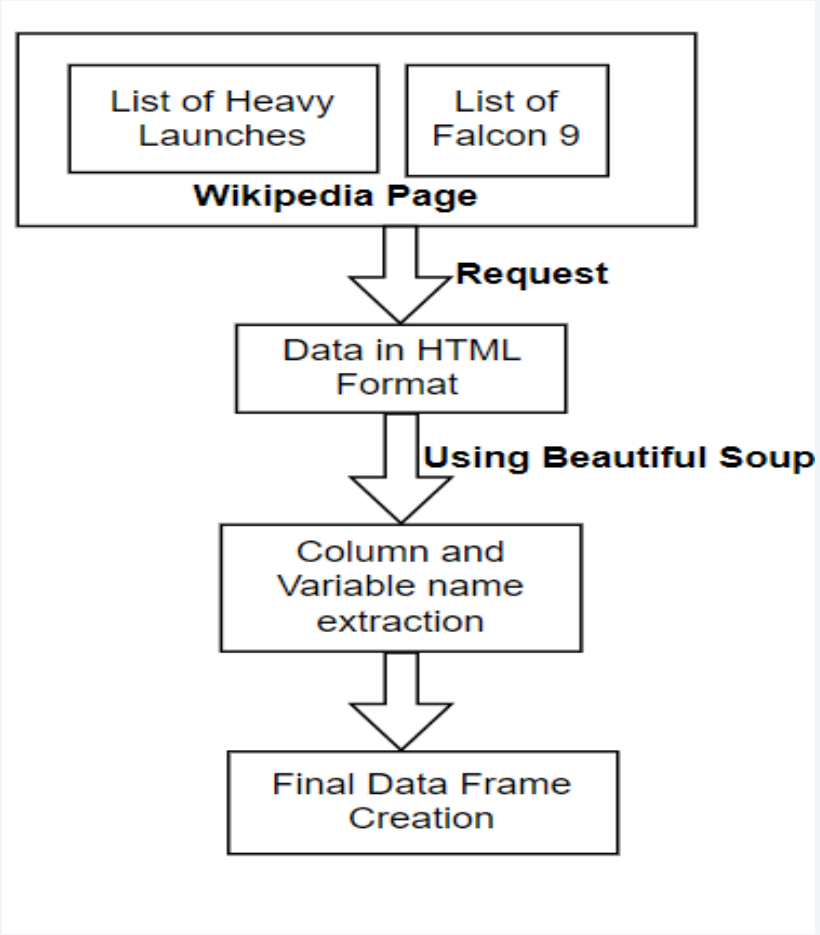
Data Collection – SpaceX API

- Data collected from spaceX API (cores, launch, payload and etc..)
- Collected is in json format which is parsed using json_normalize python library
- Data is convereted to data frame format for further processing
- [GitHub URL]



Data Collection - Scraping

- Information is downloaded in html format
- It is parsed using beautiful soup package in python
- Required information are extracted
- Final Data frame is constructed
- [Github URL]



Data Wrangling

- In data Wrangling, Some Exploratory Data Analysis (EDA) is done to find some patterns in the data and determine what would be the label for training supervised models.
- In the data set, there are several different cases in 'Landing Outcome' Column where the booster successful as well as unsuccessful .

True Ocean - Successful landing on Ocean

False Ocean - Unsuccessful landing on Ocean

True RTLS - Successful landing on Ground Pad

False RTLS - UnSuccessful landing on Ground Pad

True ASDS - Successful landing on DroneShip

False ASDS - UnSuccessful landing on Droneship

None ASDS - Failed to land

None RTLS - Failed to land

Data Wrangling (Contd.)

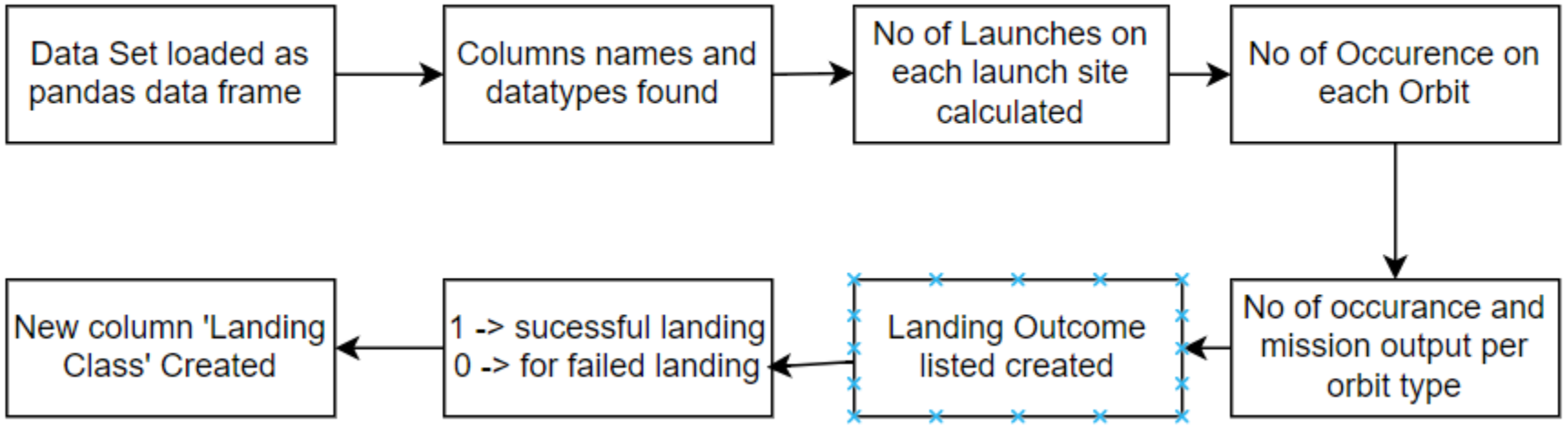
- Hence , it found to be 3 types of successful landing and 5 types of unsuccessful landing
- New column is created named “Class” where all succesfull landing are assigned ‘1’ and unsuccessful landing landings are assigned ‘0’

Successful Landing → 1

Unsuccesful Landing → 0

Github Link

Data Wrangling (Contd.)



EDA with Data Visualization

- Cat plot for categorical columns to find the frequency of each category .
- Scatter plot is drawn to find the relationship between different variables
- Bar plot is used to figure out relationship between numerical and categorical column t
- Plots Drawn :

    Flight Number Vs Payload (Cat Plot)

    Flight Number vs. Launch Site (Cat plot)

    Launch Site vs. Payload (Scatter plot)

    Success Rate vs. Orbit type (Bar plot)

    Orbit type vs. Flight Number (Scatter plot)

    Orbit type vs. Payload (Scatter plot)

    Success rate vs. Time in years (Line plot)

    GitHub URL

EDA with SQL

- we have used SQL query to get better insight about some variables
- Some of the queries executed:
  - ❑Average Payload mass carried by booster version f9
  - ❑Total number of success and failed mission outcome
  - ❑Booster which have success in droneship with payload mass > 6000
  - ❑Failed landing outcome in droneship
  - ❑Date of first successful landing on groundpad
  - ❑Names of booster_version which carried maximum payload Mass
  - ❑count of landing outcomes between a time period

GitHub URL

Build an Interactive Map with Folium

- Using the Python package Folium, designed an interactive map for below purpose.
- Map View where each Falcon 9 launch site is located, represented by a circle ○
- Learn how many launches occurred at each location, represented by markers.
- Green markers represent a successful recovery while red markers represent unsuccessful one
- Determine distances to the closest coastline, city, railway, and highway, each represented by a blue line.
- GitHub Link

Build a Dashboard with Plotly Dash

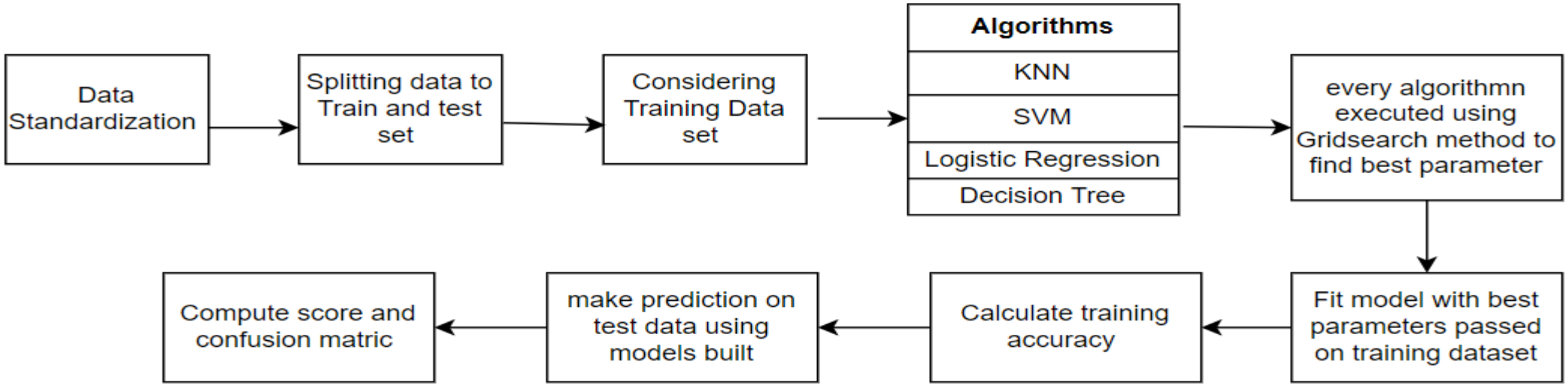
- Summarize what plots/graphs and interactions you have added to a dashboard
- Explain why you added those plots and interactions
- Github Link

Predictive Analysis (Classification)

- Data standardization is done on the Final dataset
- The dataset is then divided into training and test set
- Then Classification algorithm (KNN, SVM, Logistic regression and Decision tree) are executed using grid search on the training set to find the best params
- Then final model is final using best parameter on the training set
- Then prediction is done on the test dataset using the models build
- Accuracy score is calculated and confusion matrix is constructed to find the best performing the algorithm .

Predictive Analysis (Classification)

- Github Link



Results

Exploratory data analysis results:

- The SpaceX launch is more successful over a period of time after 2013
- Light payloads with mas between (2k to 6k) seems to have more successful launches
- Almost all the launch sites has good success rate , but KSC LC – 39A seems to be more reliable

Interactive analytics demo in screenshots:

- One launch site is on the east coast and remaining 3 are in west coast
- Nearby railroad and airport are calculated for all the launch sites

Predictive analysis results:

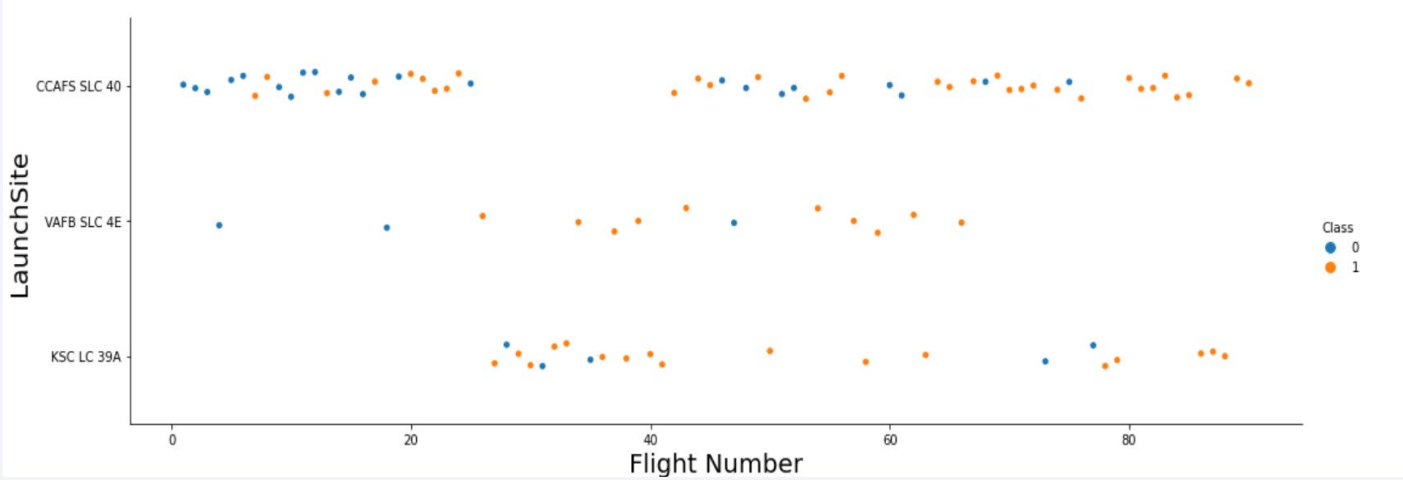
- Almost all the algorithm performed well with accuracy of 83% except decision tree which is bit lesser

Section 2

Insights drawn from EDA

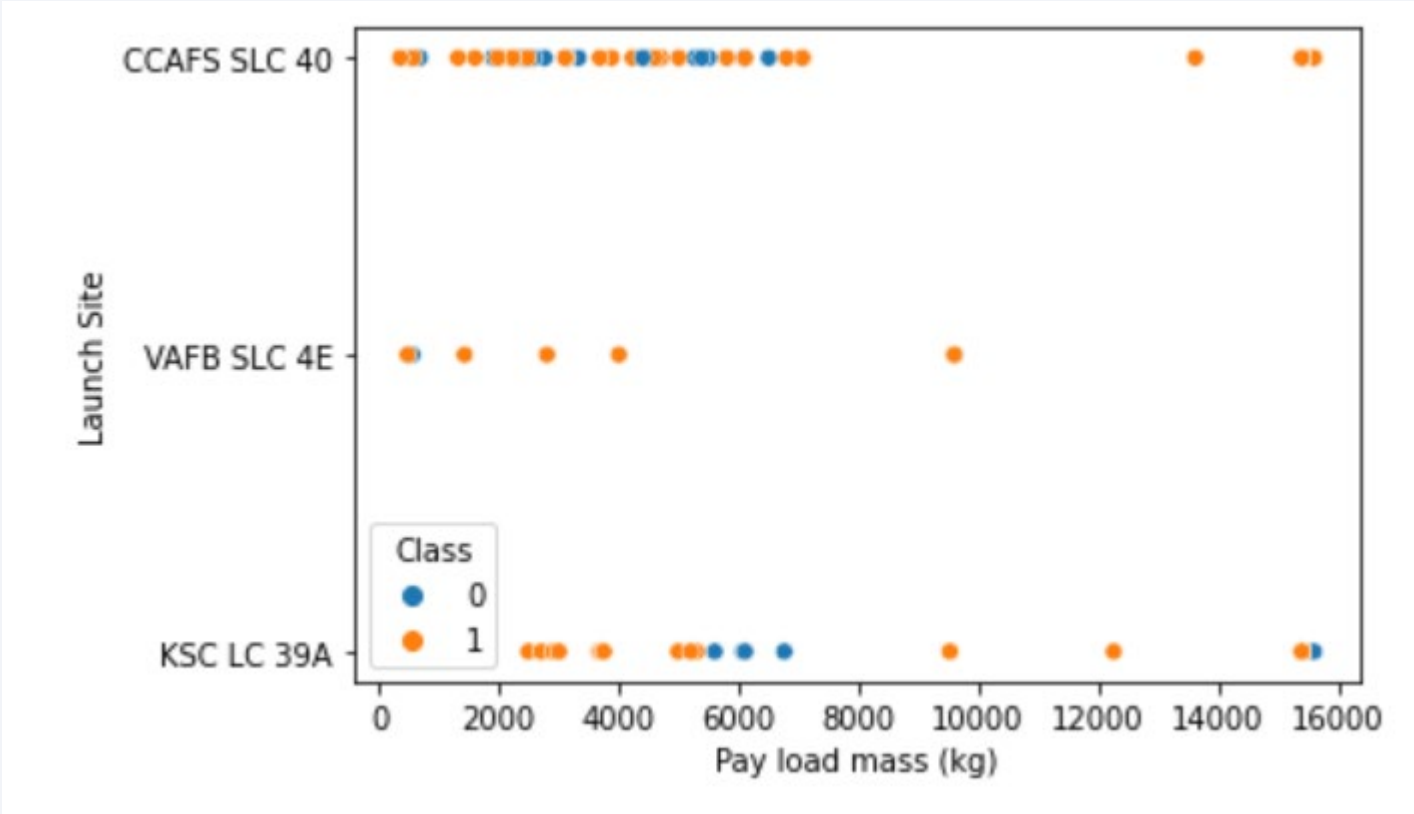
Flight Number vs. Launch Site

- The CCAFS_SLC 40 has more number of than other sites
- The launch Site CCAFS_SLC 40 has more successful launches than other launch sites
- The total launch count as well failed launches are less for VAFB SLC is less
- The successful launch is higher with flight number > 40



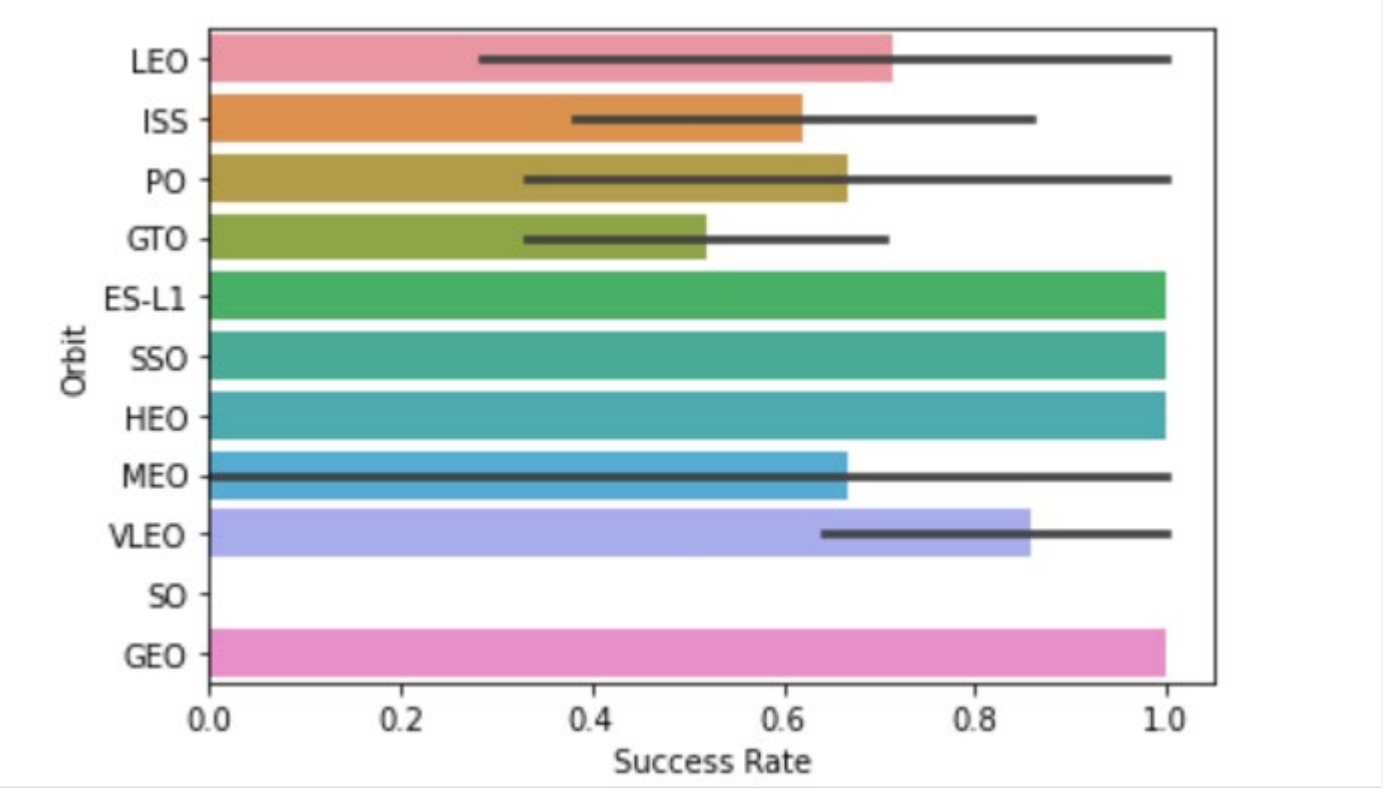
Payload vs. Launch Site

- CCAFS SLC 40 has more successful launches with payload mass between 2k to 4k
- Most of the payload has mass less than 8K
- Maximum payload mass of 16k was never launched from the site VAFB SLC 4E



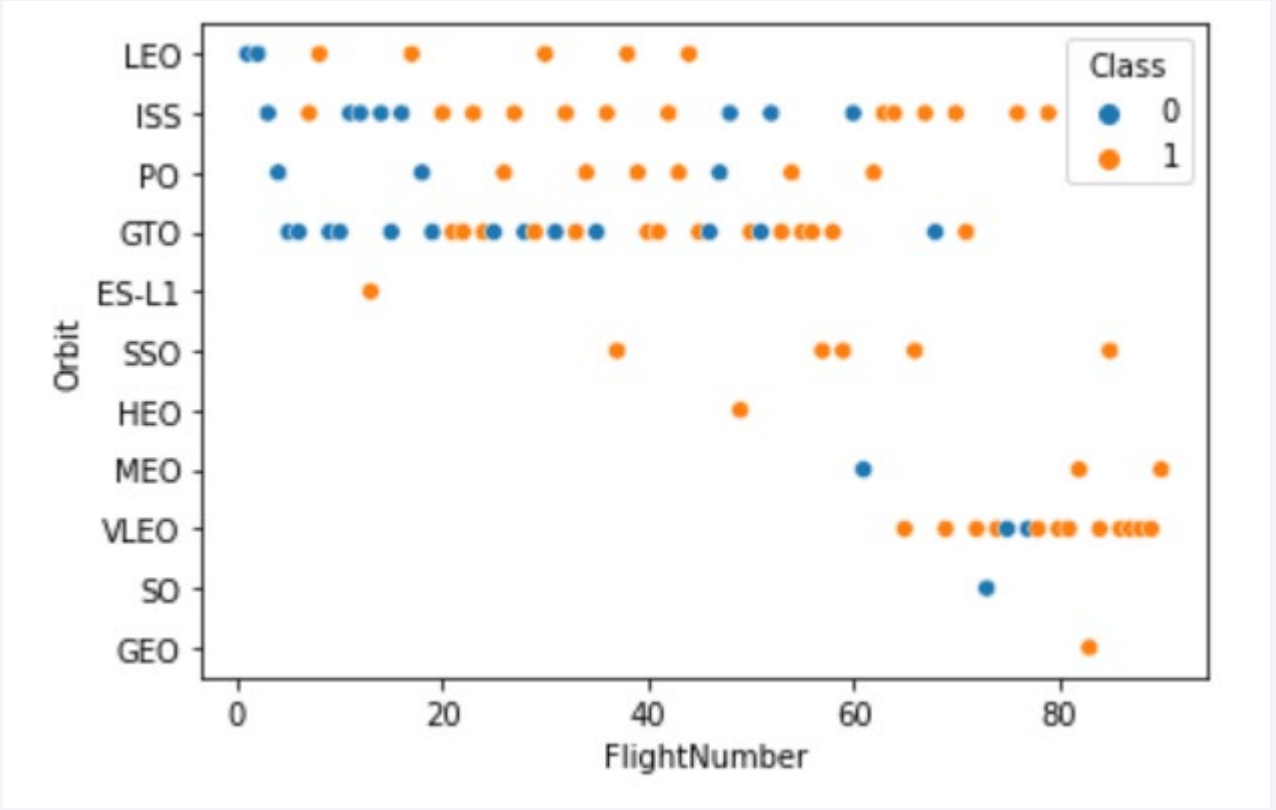
Success Rate vs. Orbit Type

- The Orbit ES-L1,SSO,HEO,GEO has higher success rate
- The Orbit GTO has lowest success rate



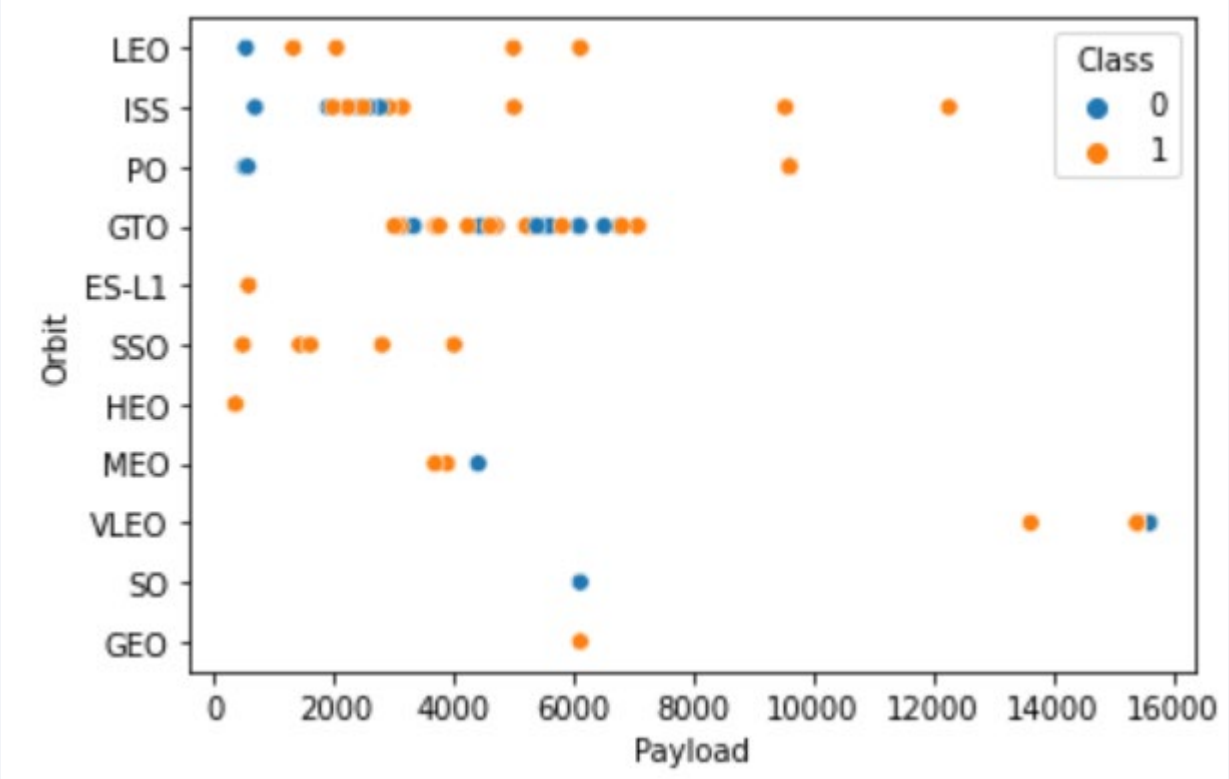
Flight Number vs. Orbit Type

- The Orbit type ISS,GTA has more of number of launches as well as successful launches .
- The orbit type ES-L1, SSO, HEO has no failures as well less number of launches



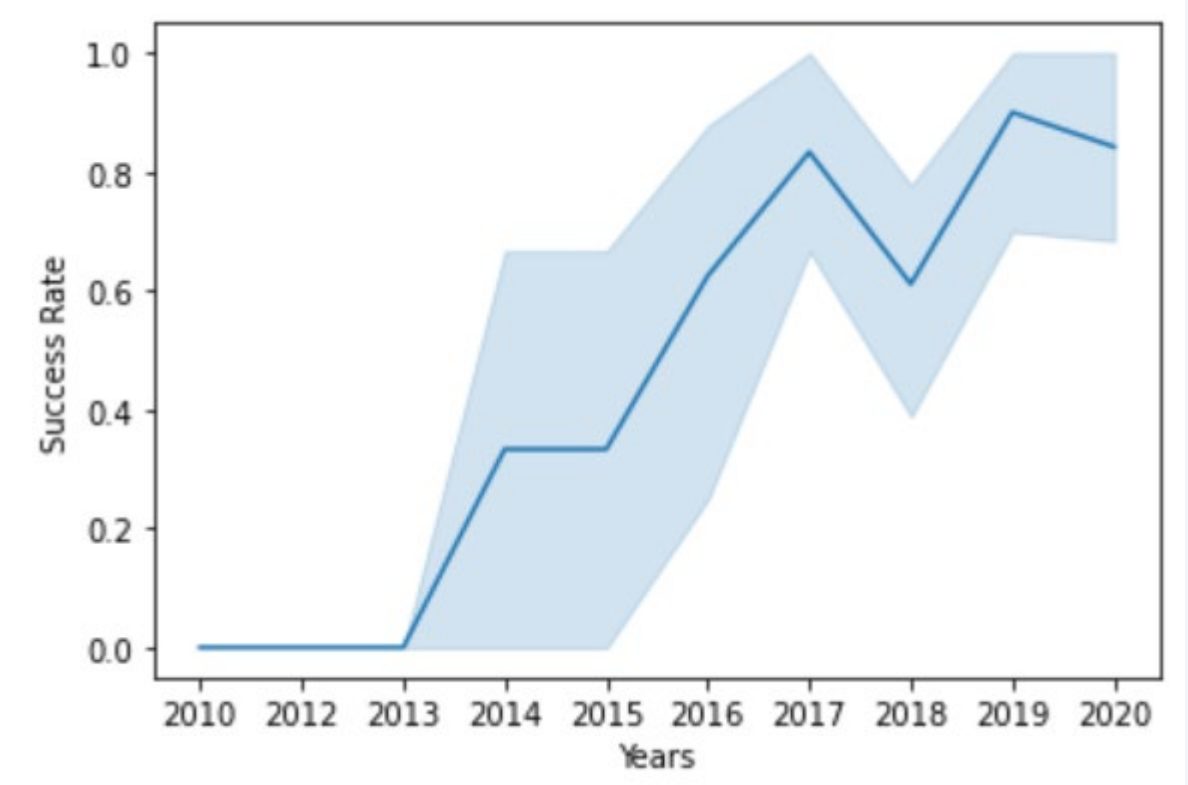
Payload vs. Orbit Type

- The orbit ES-L1, SSO HEO, has all the launches successful with payload mass less than 6k
- Only VLEO and ISS are used for high payload launches > 12k
- Most of the payloads launches are less than mass of 8k



Launch Success Yearly Trend

- The success rate is very much low between 2010 to 2013
- The overall success rate increased from 2013 to 2017
- Success rate again decreases between 2017 to 2018



All Launch Site Names

- There are four unique Launch sites ,
  - CCAFS LC-40
  - CCAFS SLC-40
  - KSC LC-39A
  - VAFB SLC-4E

```
%sql select unique(LAUNCH_SITE) from SPACEXTBL
```

```
 * ibm_db_sa://pfp77871:***@19af6446-6171-4641-8aba-9d
Done.
```

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

Launch Site Names Begin with 'CCA'

- First five records launch sites begin with `CCA` has launch site CCAFS LC-10
- Except one , all are launched by NASA and all has successful mission outcome

```
[ ]  %sql select *  from SPACEXTBL WHERE LAUNCH_SITE LIKE '%CCA%' LIMIT 5
```

```
 * ibm_db_sa://pfp77871:***@19af6446-6171-4641-8aba-9dcff8e1b6ff.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:30699/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

Total Payload Mass

- The total payload mass carried by boosters from NASA is 45596

```
%sql select sum(PAYLOAD_MASS__KG_) from SPACEXTBL where CUSTOMER='NASA (CRS)'
```

```
 * ibm_db_sa://pfp77871:***@19af6446-6171-4641-8aba-9dcff8e1b6ff.c1ogj3sd0tgtu0lq
Done.
   1
45596
```

Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1 is 2928

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

```
%sql select avg(PAYLOAD_MASS__KG_) from SPACEXTBL where BOOSTER_VERSION = 'F9 v1.1'
```

```
 * ibm_db_sa://pfp77871:***@19af6446-6171-4641-8aba-9dcff8e1b6ff.c1ogj3sd0tgtu0lqde00.database
99/bludb
Done.
```

1
2928

First Successful Ground Landing Date

- Dates of the first successful landing outcome on ground pad is 2010-06-04

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

Hint:Use min function

```
[ ]  %sql select min(DATE) from SPACEXTBL where MISSION_OUTCOME like 'Success'
```

```
 * ibm_db_sa://pfp77871:***@19af6446-6171-4641-8aba-9dcff8e1b6ff.c1ogj3sd0tgtu0l
Done.
     1
2010-06-04
```

Successful Drone Ship Landing with Payload between 4000 and 6000

- boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

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

```
[ ]  %sql SELECT BOOSTER_VERSION FROM SPACEXTBL WHERE LANDING__OUTCOME LIKE '%Success (drone ship)%' AND PAYLOAD_MASS__KG_ BETWEEN 4000 and 6000
```

```
 * ibm_db_sa://pfp77871:***@19af6446-6171-4641-8aba-9dcff8e1b6ff.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:30699/bludb
Done.
```

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

Total Number of Successful and Failure Mission Outcomes

- Total number of successful mission outcome is 100 and failure is 1

```
[12]: %sql SELECT COUNT(*) as success FROM SPACEXTBL WHERE MISSION_OUTCOME LIKE '%Success%'

 * ibm_db_sa://pfp77871:***@19af6446-6171-4641-8aba-9dcff8e1b6ff.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:30699/bludb
Done.
```

[12]:

success
100

```
[13]: %sql SELECT COUNT(*) as failure FROM SPACEXTBL WHERE MISSION_OUTCOME LIKE '%Failure%'

 * ibm_db_sa://pfp77871:***@19af6446-6171-4641-8aba-9dcff8e1b6ff.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:30699/bludb
Done.
```

[13]:

failure
1

Boosters Carried Maximum Payload

- The Names of the booster which have carried the maximum payload mass

```
%sql select BOOSTER_VERSION from SPACEXTBL where PAYLOAD_MASS__KG_=(select max(PAYLOAD_MASS__KG_) from SPACEXTBL)
```

```
 * ibm_db_sa://pfp77871:***@19af6446-6171-4641-8aba-9dcff8e1b6ff.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:30699/bludb
Done.
```

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 2015 failed landing_outcomes in drone ship, is launched from CCAFS LC-40 and its booster version is v1.1

```
%sql SELECT LANDING__OUTCOME,BOOSTER_VERSION,LAUNCH_SITE,DATE from SPACEXTBL where LANDING__OUTCOME like '%Failure (drone ship)%' and DATE like '%201
```

```
 * ibm_db_sa://pfp77871:***@19af6446-6171-4641-8aba-9dcff8e1b6ff.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:30699/bludb
Done.
```

landing__outcome	booster_version	launch_site	DATE
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40	2015-01-10
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40	2015-04-14

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

- 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
- Total number of success is 11, unattempted is 10, and failure is 9

```
%sql select LANDING__OUTCOME, COUNT(*) AS counts FROM SPACEXTBL WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY LANDING__OUTCOME ORDER BY
```

```
 * ibm_db_sa://pfp77871:***@19af6446-6171-4641-8aba-9dcff8e1b6ff.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:30699/bludb
Done.
```

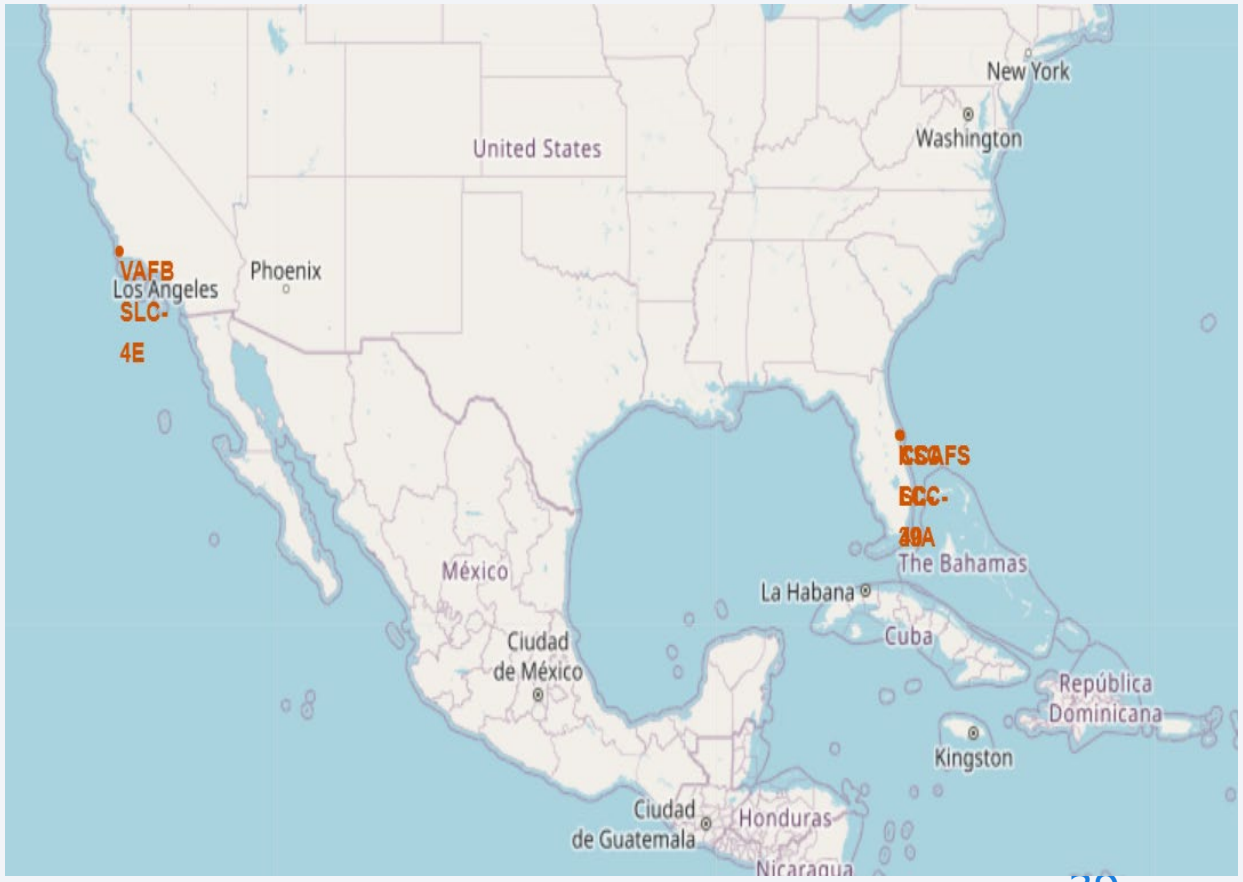
landing__outcome	counts
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

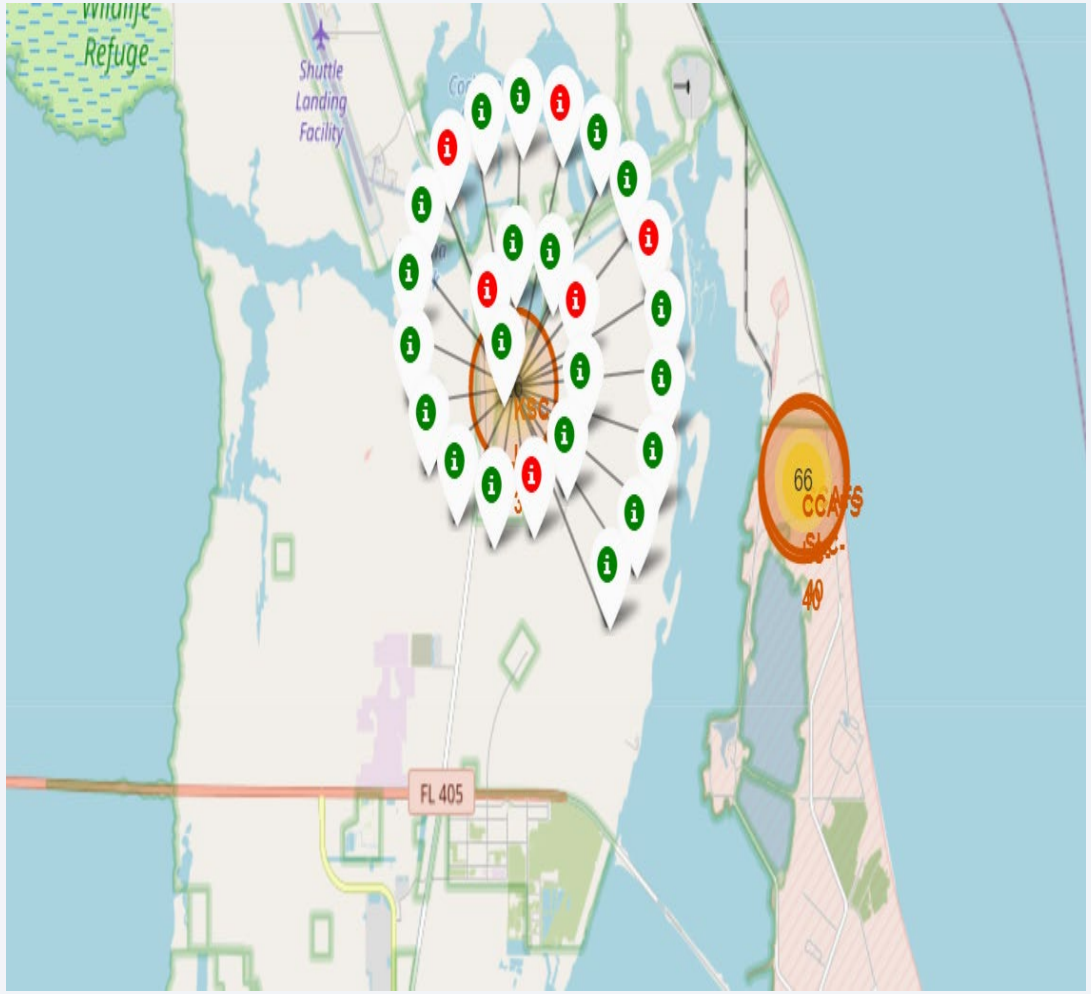
<Folium Map Screenshot 1>

- In this map , launch site of spaceX is plotted with red marker
- One launch site is near California(west)
- Rest of the launch sites are near Florida (east)



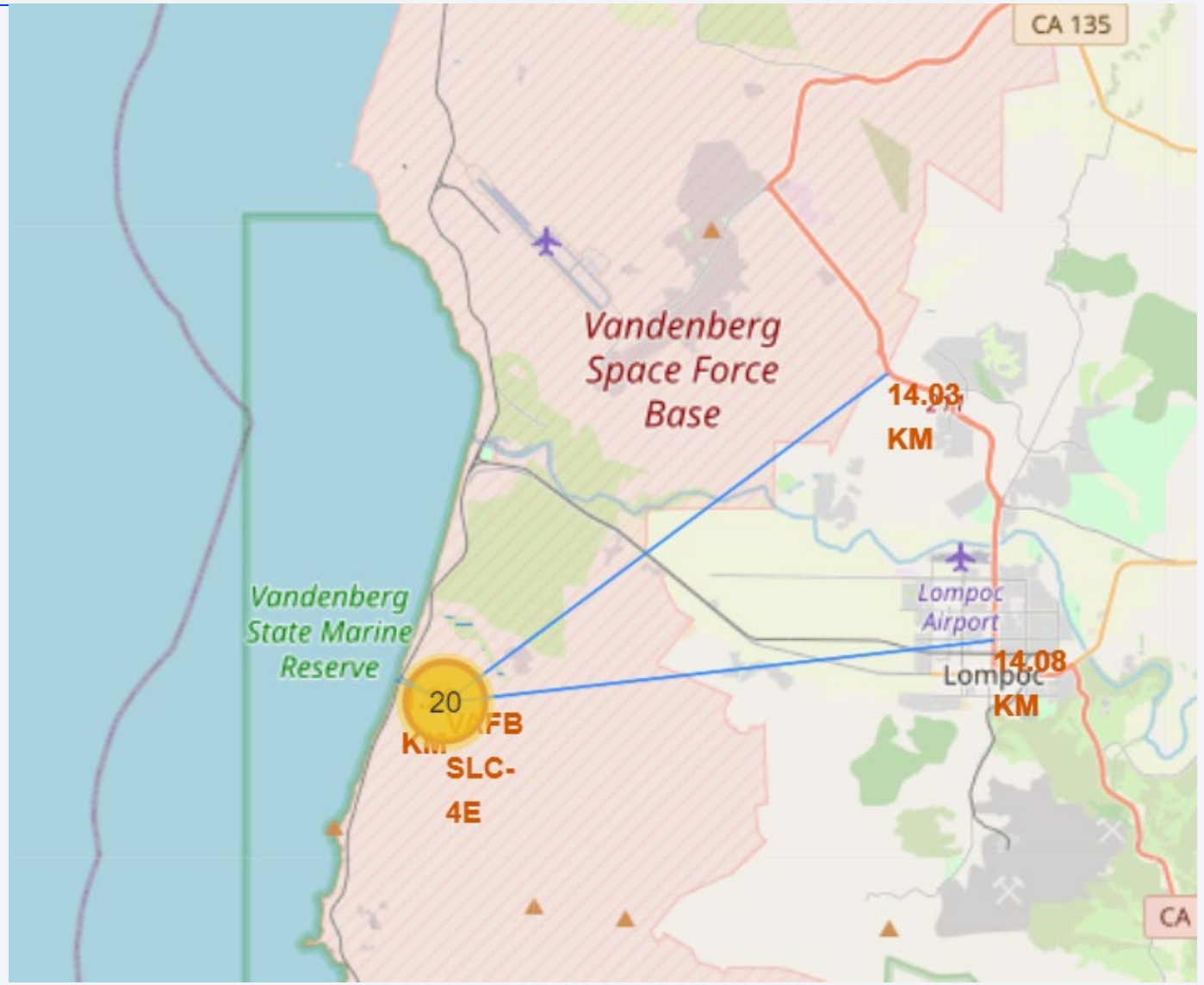
Folium Map showing successful and failed recovery

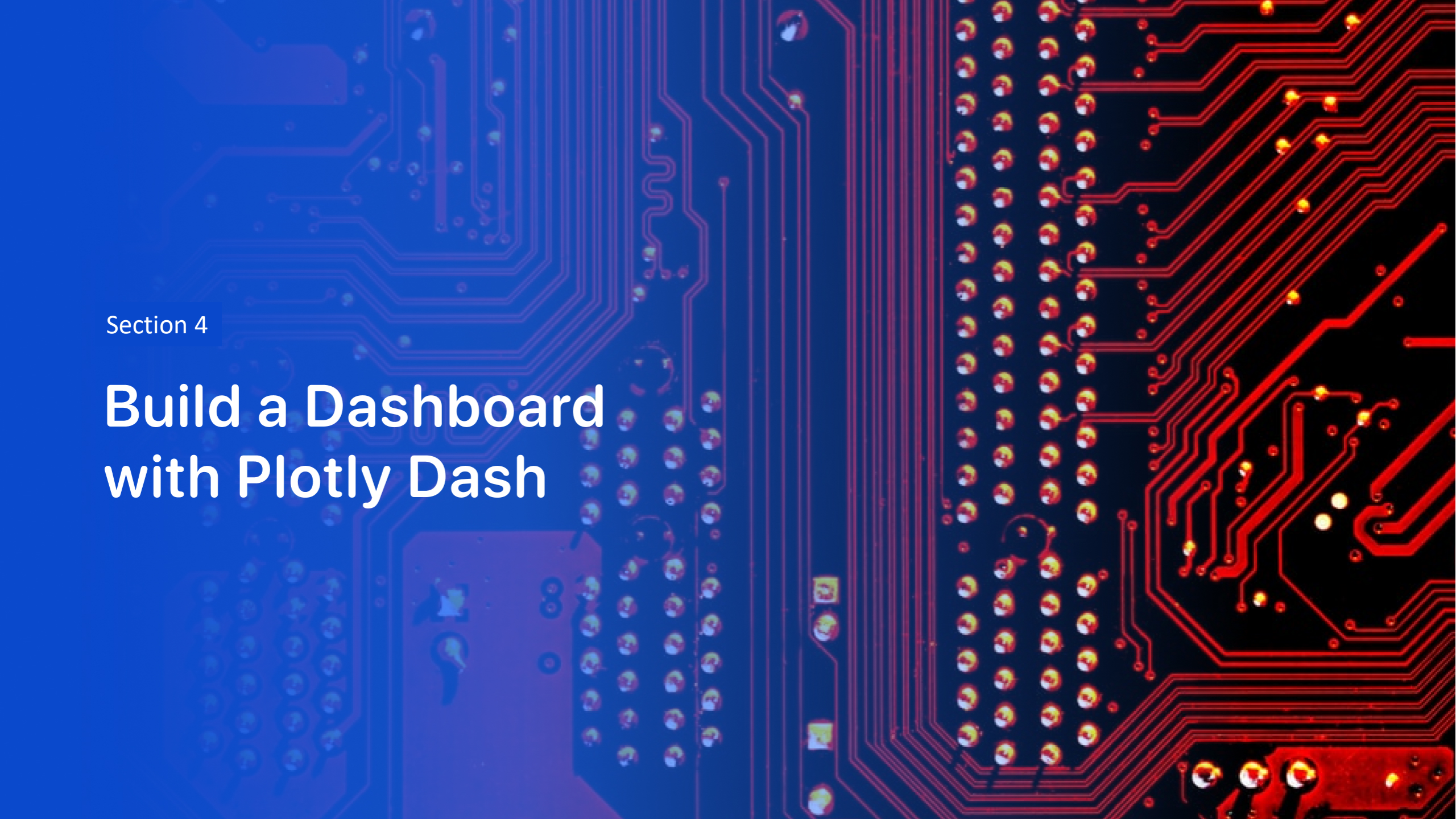
- This folium map shows successful and failed recovery of the launch site KSC LC-39A
- Green represents successful recovery
- Red Represents failure



Follium map showing nearby Location

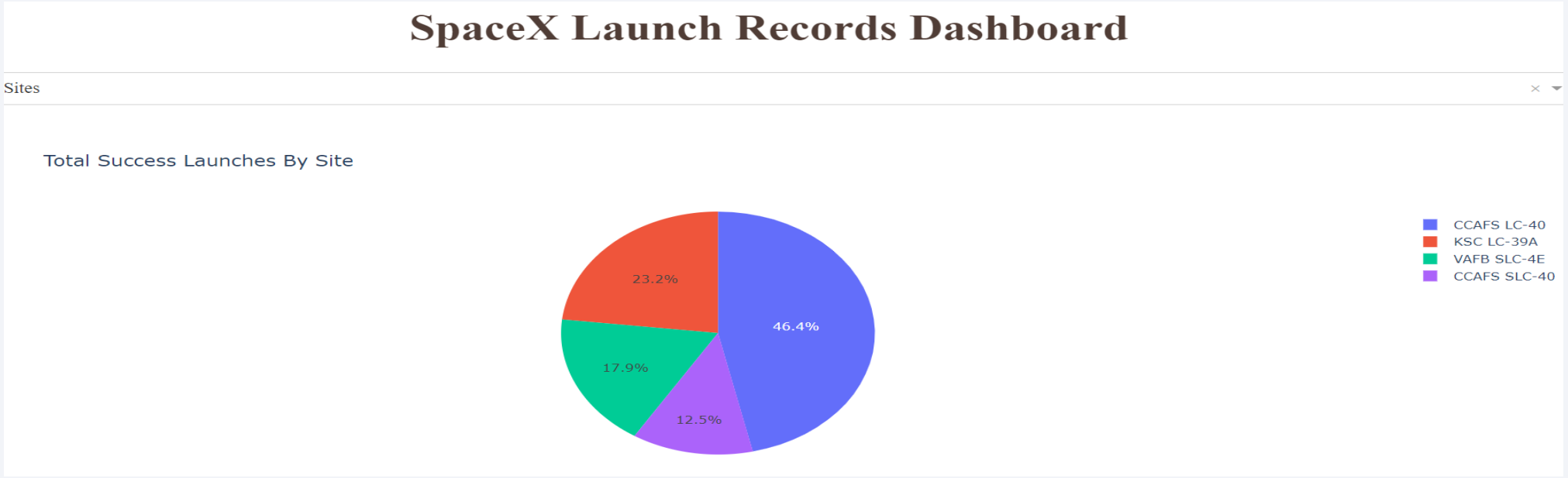
- The closest distance between airport and VAFB SLC-4E site is 14km
- The closest distance between roadways and VAFB SLC-4E site is 14km
- It is represented by blue connecting line on the map




Section 4
Build a Dashboard with Plotly Dash

<Dashboard Screenshot 1>

- From the screenshot it is found CCAFS LC – 40 has most successful launches and CCAFS SLC-40 has least successful launches
- But Failure launch percentage is not visible in this pie chart, hence conclusion cannot be drawn

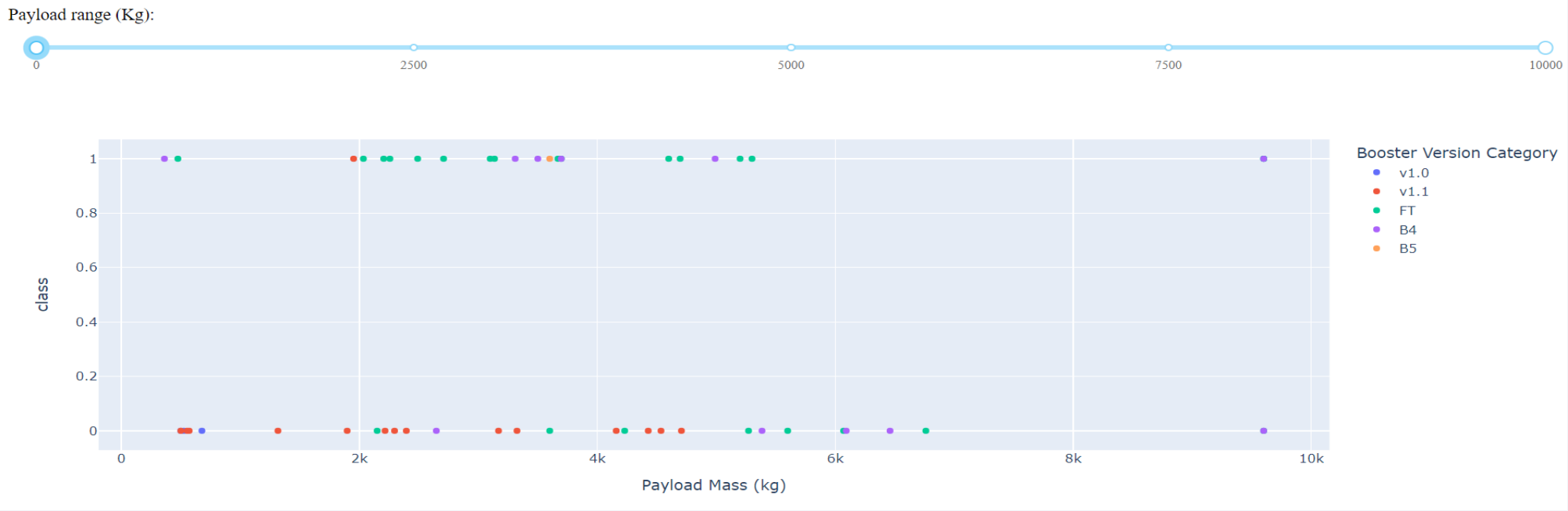


<Dashboard Screenshot 2>

- When every site is analyzed individually, KSC LC-39A has more success ratio than CCAFS LC-40, but its overall launch count is lesser than CCAFS LC-40.

<Dashboard Screenshot 3>

- The Booster version FT has more successful launches with payload range ranges between 2k to 4K, and others also
- The Booster Version V1.1 has more failures with payload mass less than 6K

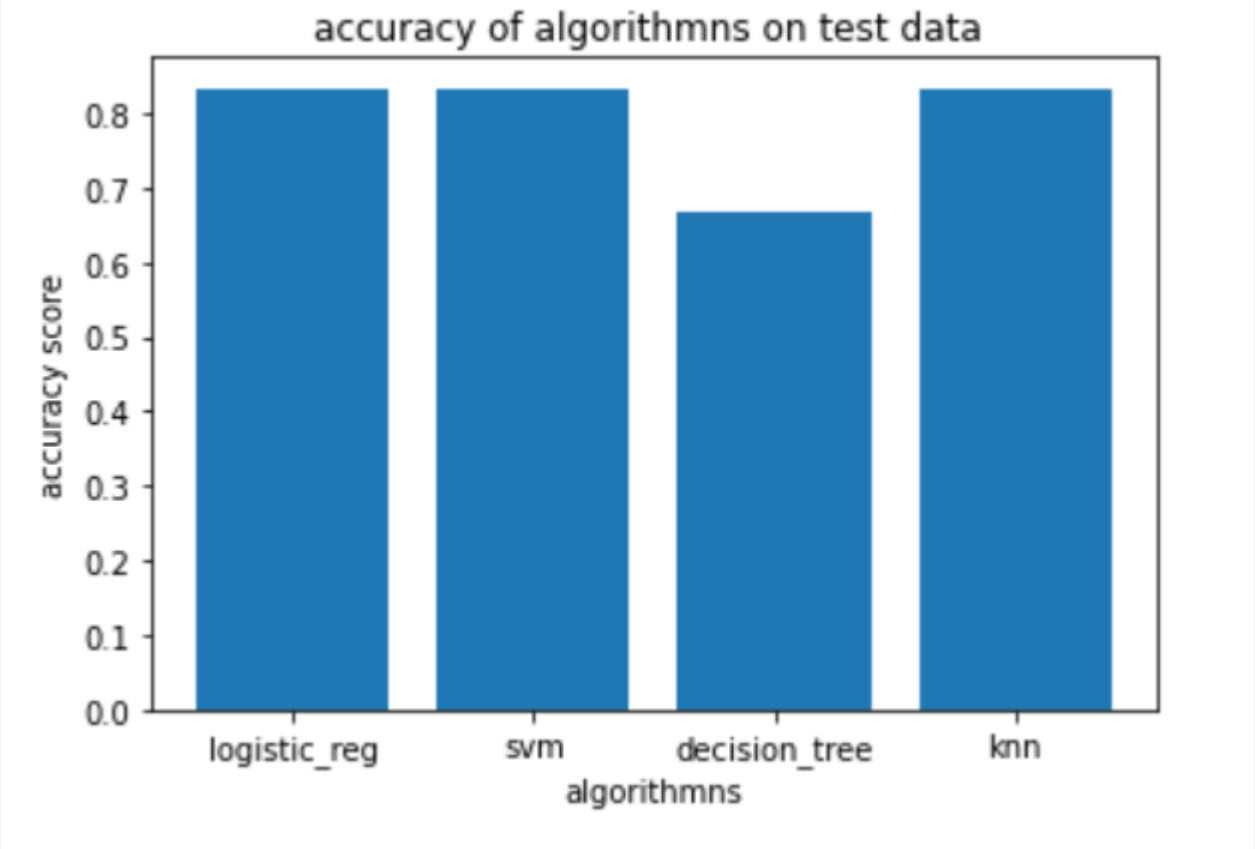


Section 5

Predictive Analysis (Classification)

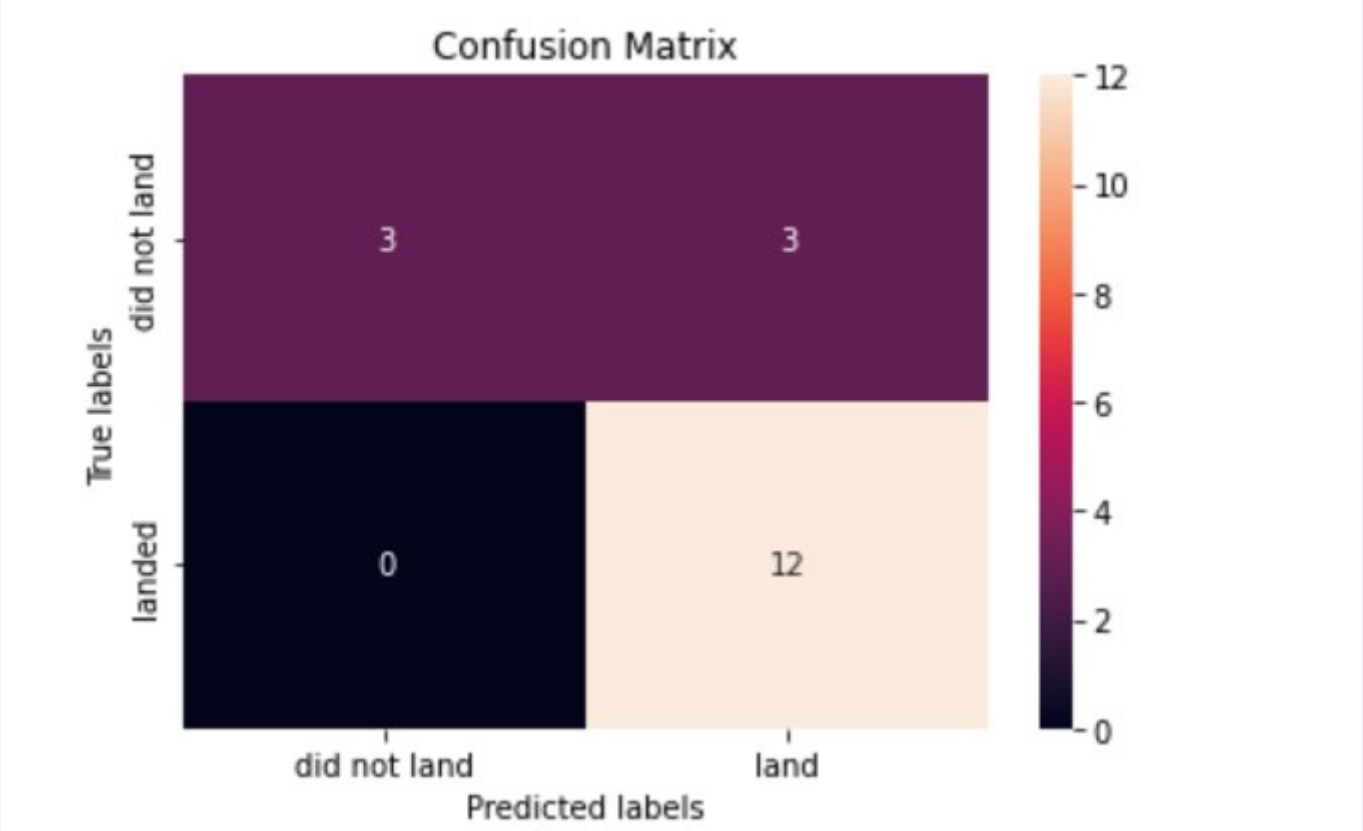
Classification Accuracy

- Almost all off the algorithm performed equally well with accuracy score of 83% on test data, except decision tree algorithm which has comparatively lesser accuracy



Confusion Matrix

- From the confusion matrix, it is found that successful landings are predicted correctly
- But , 3 of the unsuccessful landing were predicted wrongly as successful landed



Conclusions

From detailed analysis ,the SpaceX successful launches has following properties :

- Out of 4 launch sites , KSLC 39A has more number of successful recoveries.
- Light payload (mass between 2k to 6k) has high probability of success .
- The orbit type ISS typically has successful launch for both light and heavy payload
- The accuracy of the model is reasonably good with score of 83%

Thank you!