Winning Space Race with Data Science

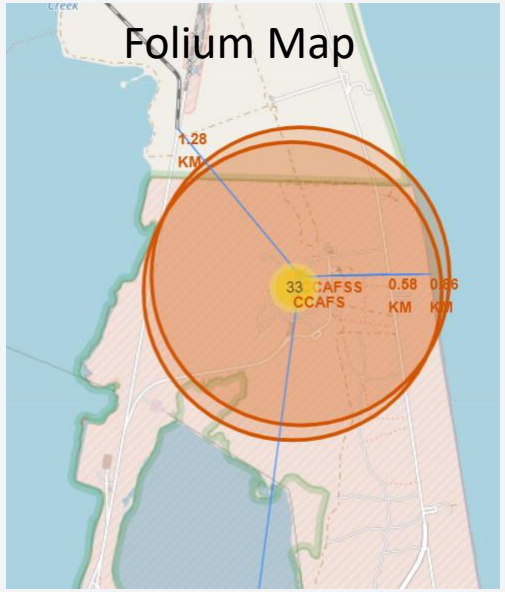
Prepared by: Sayli Shiradkar
Date: 11th May,2022

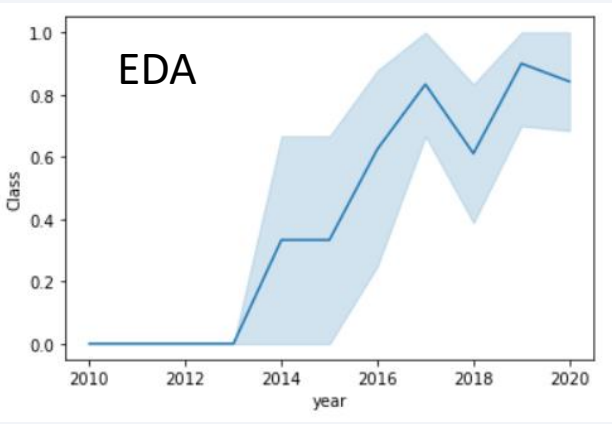
Outline

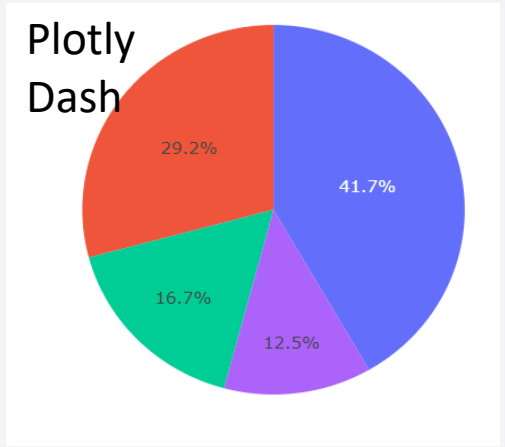
- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion
- Appendix

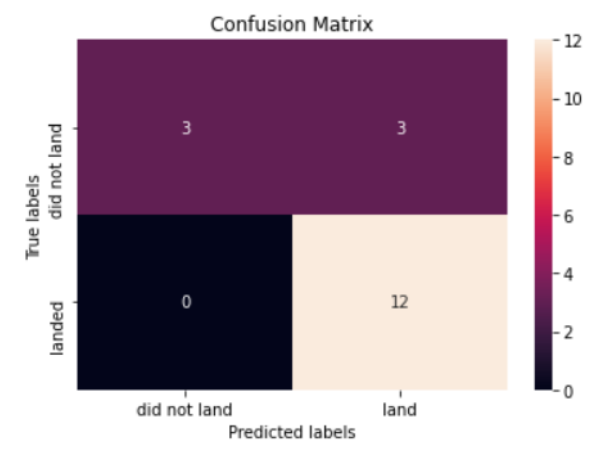
Executive Summary

- Summary of methodologies
  - ➢ Data Collection with API and Web scraping
  - ➢ Data Wrangling
  - ➢ Exploratory Data Analysis and Visualization
  - ➢ Interactive Maps with Folium and Dashboard with Plotly Dash
  - ➢ Predictive Analysis for various classification models
- Summary of all results
  - Interactive Data Visualization
  - Best Classification Model for Predictive Analysis









Model evaluation

Introduction

- SpaceX Falcon 9 rocket launches cost $62 M
- Other providers cost $165M
- Why are costs so different?
- because SpaceX can reuse the first stage!

Problems of interest:

- What factors determine successful rocket landing?
- Understanding Relationship among different predictor variables and outcome
- Predict if the first stage will land successfully or not

Landing Success

Motivation

- Graduated from a university in Florida and lived in Florida for about five years
- Have witnessed rocket launches in KSC!
- Data Science project with SpaceX launches gave great opportunity to explore data and factors behind launch success and failure

Section 1

Methodology

Methodology

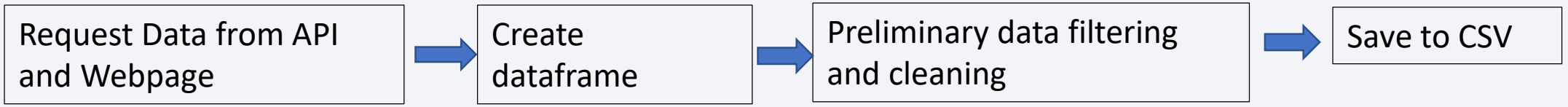
Executive Summary

- Data collection:
  - SpaceX API was requested to get information about SpaceX launches
  - Web scraping of Wikipedia page* to collect Falcon 9 historical launch records
- Data wrangling
  - Converting outcome labels and other data for training supervised models
- Exploratory data analysis (EDA) using visualization and SQL
- Interactive visual analytics using Folium and Plotly Dash
- Predictive analysis using classification models
  - Split Data into train and test, use train data for training different classification models, evaluate model accuracy using test data

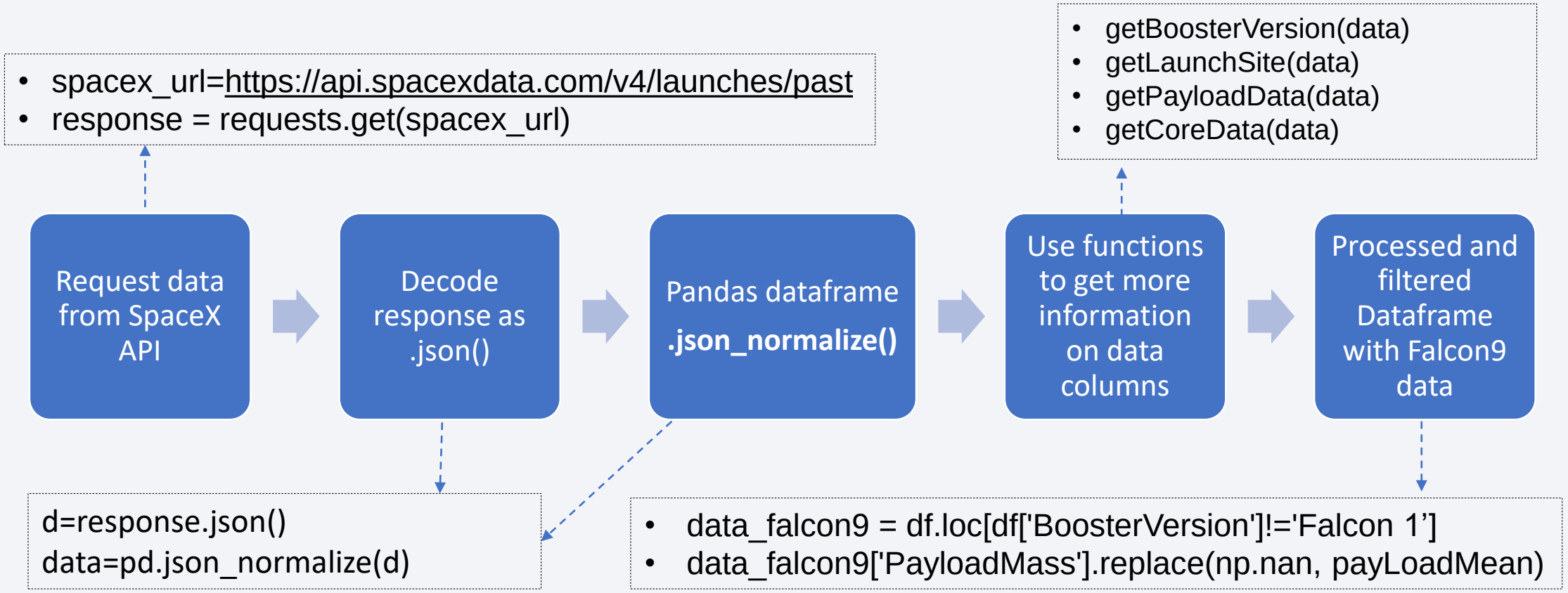
- References: *https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches

Data Collection

- Relevant data on SpaceX to predict rocket launch outcomes was gathered
- SpaceX API was requested to get launches related data
- Web scraping of SpaceX Wikipedia page to collect Falcon 9 historical launch records



Data Collection – SpaceX API



GitHub URL

Data Collection - Scraping

Request the Falcon9 Launch HTML page

Create a BeautifulSoup object

Extract Tables

Extract columns

Create a dictionary and append data to keys

Convert dictionary to dataframe

Save to csv

GitHub URL scraping

```
static_url=https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922
html_data  = requests.get(static_url).text
```

```
soup = BeautifulSoup(html_data, 'html5lib')
```

```
html_tables=soup.find_all('table')
```

```
column_names = []
for row in first_launch_table.find_all('th'):
    name = extract_column_from_header(row)
    if name != None and len(name) > 0:
        column_names.append(name)
```

```
launch_dict= dict.fromkeys(column_names)
```

Flight No.	Launch site	Payload	Payload mass	Orbit	Customer	Launch outcome	Version Booster	Booster landing	Date	Time
1	CCAFS	Dragon Spacecraft Qualification Unit	NaN	LEO	SpaceX	Success\n	F9 v1.0B0003.1	Failure	4 June 2010	18:45
2	CCAFS	Dragon	NaN	LEO	NASA	Success	F9 v1.0B0004.1	Failure	8 December 2010	15:43

Data Wrangling

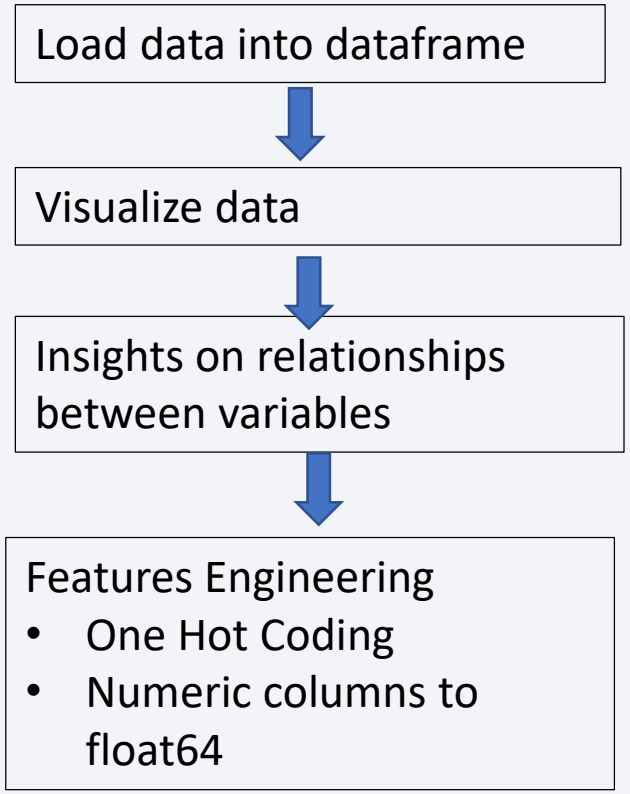
- Data was preprocessed, explored and Outcome labels for training models determined -
  - % missing values in each attribute: df.isnull().sum()/ df.count()*100
  - Identify which columns are numerical and categorical: df.dtypes
  - Number of launches per orbit type: df['Orbit'].value_counts()
  - Number of landing outcomes: df['Outcome'].value_counts()
  - Landing Class was created with 0 as Success and 1 as failure in landing
  - Dummy variables were created for categorical columns

GitHub URL wrangling

EDA with Data Visualization

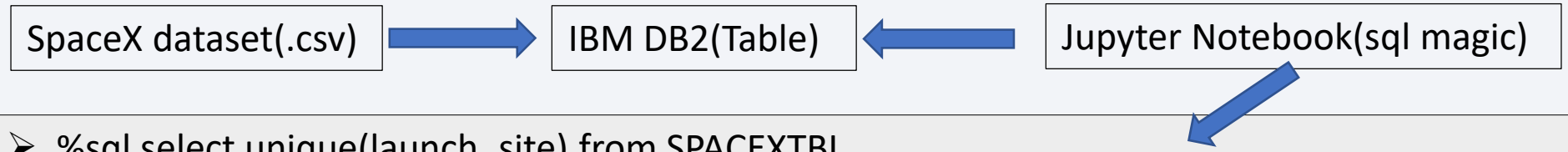
- Various charts were plotted to visualize relationships between various variables of interest
  - FlightNumber vs PayloadMass
  - Flight Number vs Launch Site
  - Payload and Launch Site
  - Success rate per Orbit Type
  - FlightNumber and Orbit type
  - Launch success yearly trend from 2013 to 2020

GitHub URL EDA Visualization



EDA with SQL

- The SQL queries performed to gain insights on payloads, mission outcomes, launch sites and landing outcomes



- %sql select unique(launch_site) from SPACEXTBL
- %sql select avg(payload_mass__kg_) as avg_payload_mass from SPACEXTBL where (booster_version) like 'F9 v1.1';
- %sql select min(DATE) as firstSuccessGroundDate from SPACEXTBL where LANDING__OUTCOME='Success (ground pad)';
- %sql select BOOSTER_VERSION from SPACEXTBL where PAYLOAD_MASS__KG_=(select max(PAYLOAD_MASS__KG_) from SPACEXTBL);
- %sql select MISSION_OUTCOME,count(MISSION_OUTCOME) as cnt from SPACEXTBL GROUP BYMISSION_OUTCOME;

GitHub URL EDA SQL

Build an Interactive Map with Folium

- Does launch success depend on the location and proximities of a launch site?
- What factors decide building a launch site?
- Folium Map objects created to visualize launch sites on global map
  - Circle – highlighting circle area with location name on coordinates
  - Marker Cluster and markers- mark for each site launch success (green) and failure(red)
  - Mouse Position -get the coordinate (Lat, Long) for a mouse over on the map
  - PolyLine –draw a line between a launch site to city, coast, railway, highway locations

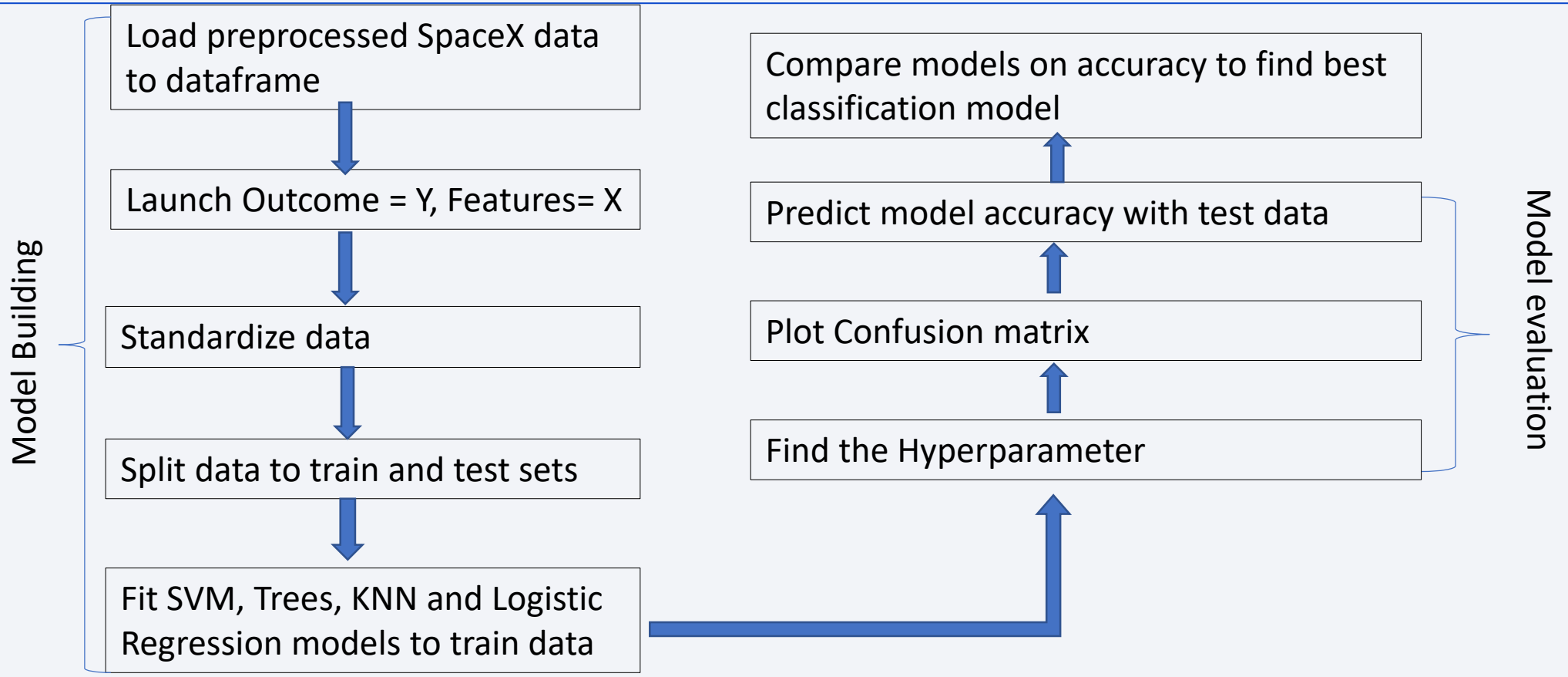
GitHub URL Folium Map

Build a Dashboard with Plotly Dash

- Interactive Dashboard with Plotly Dash developed to visualize and compare
  - Total successful launches for all launch sites
  - Launch site wise success and failure ratio
  - Effect of Payload mass on success/ failure rates and booster versions
- Dropdown menu was used to display success/failure rate for sites
- Slider was added to select payload range

GitHub URL PlotlyDash

Predictive Analysis (Classification)



GitHub URL predictive analysis

Results

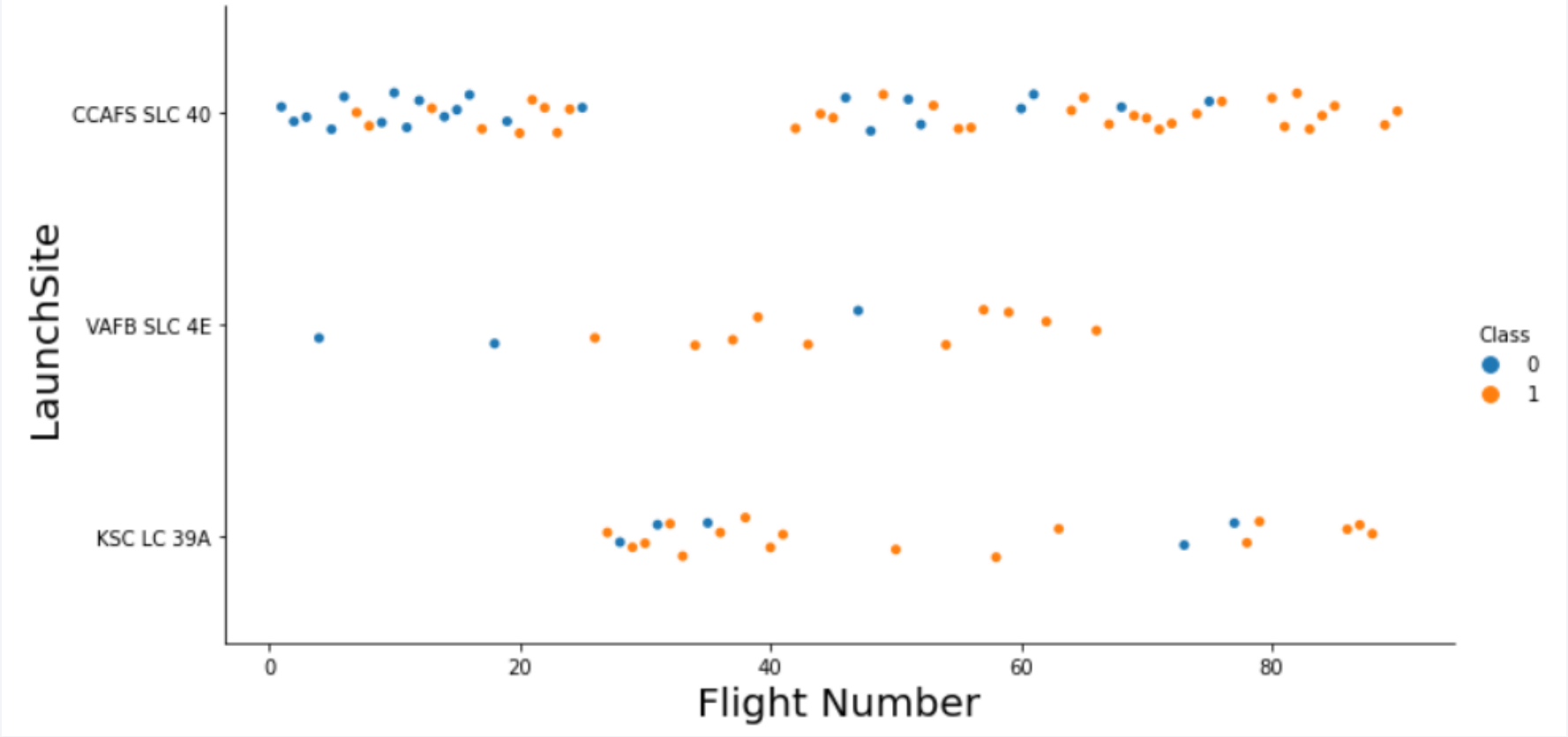
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

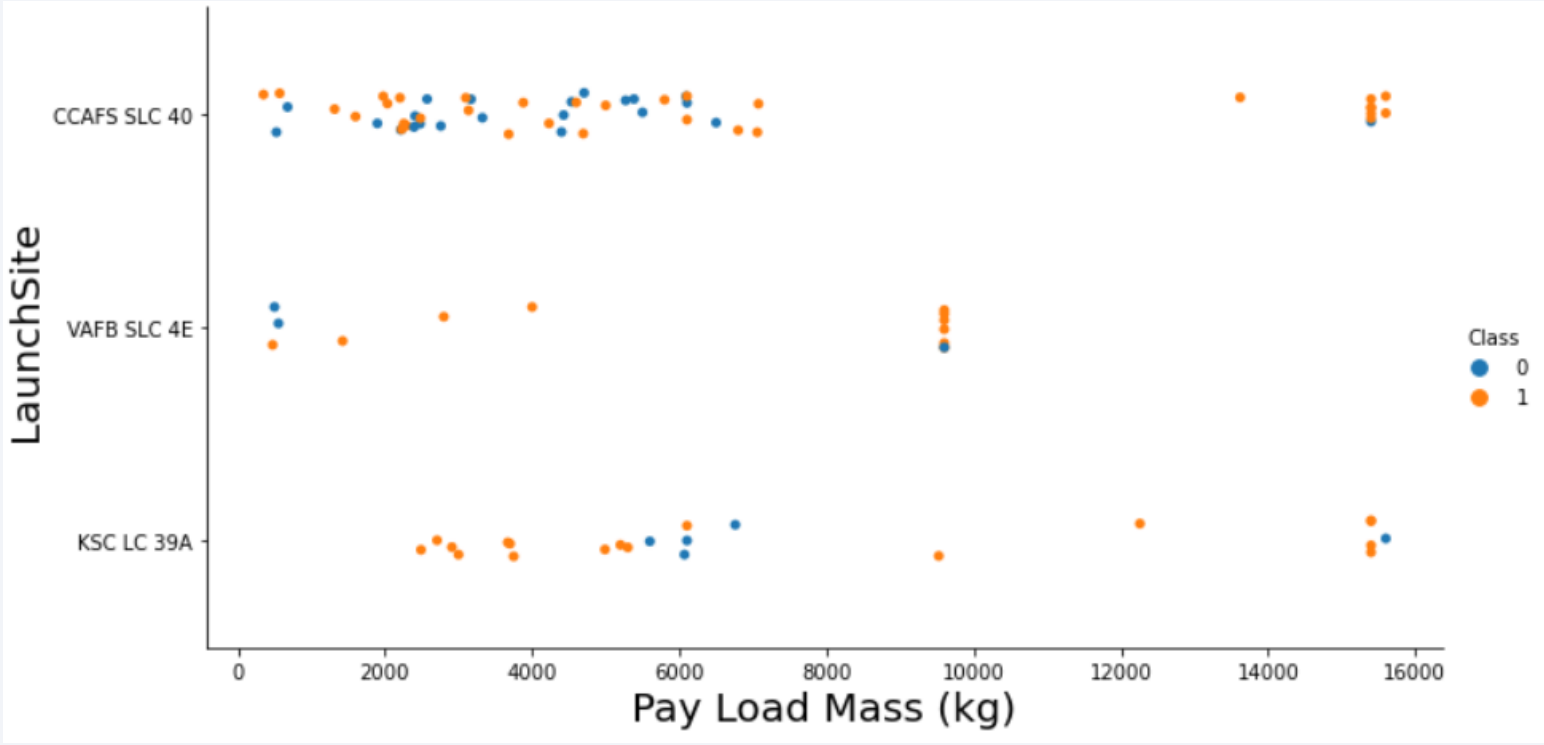
Flight Number vs. Launch Site

- CCAFS: Success rate increases as flight numbers increase beyond 65
- VAFB: Success rate increases after 20 Flight Number
- Many launches between range of 0 to 25 Flight Numbers for CCAFS site while KSC has none



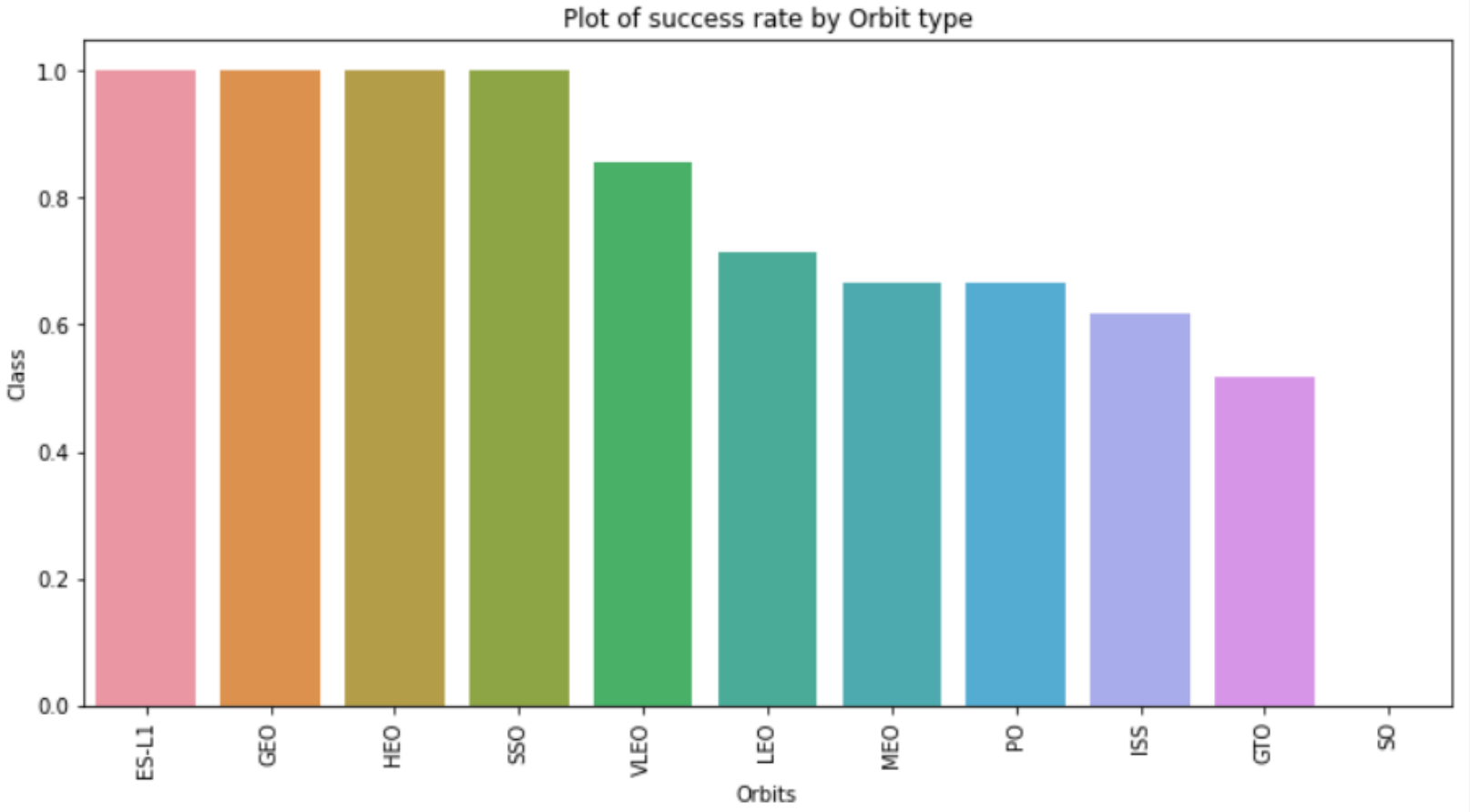
Payload vs. Launch Site

- VAFB-SLC launch site: no rockets launched for heavy pay load mass (greater than 10000)



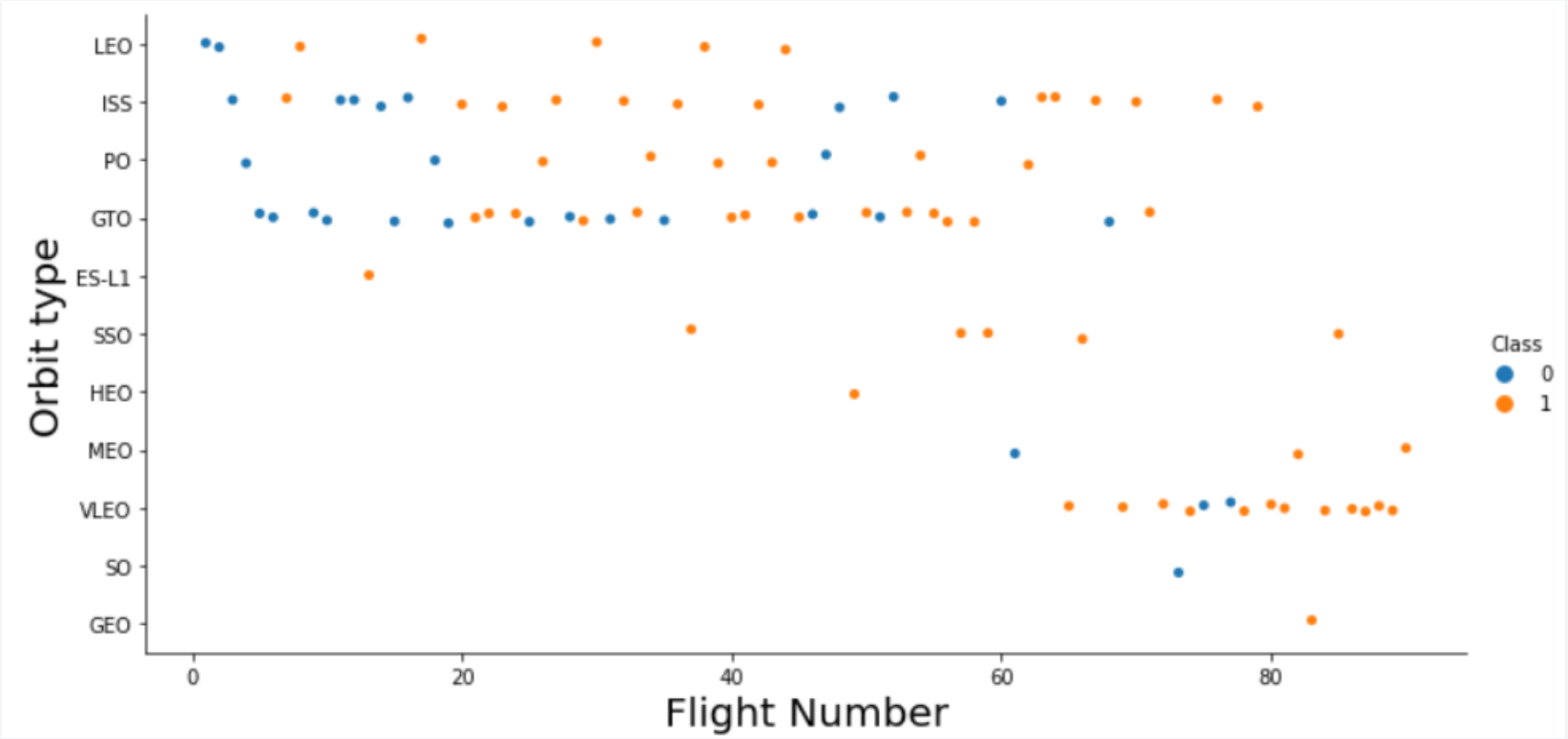
Success Rate vs. Orbit Type

- High Success rate Orbits:
  - ES-L1
  - GEO
  - HEO
  - SSO



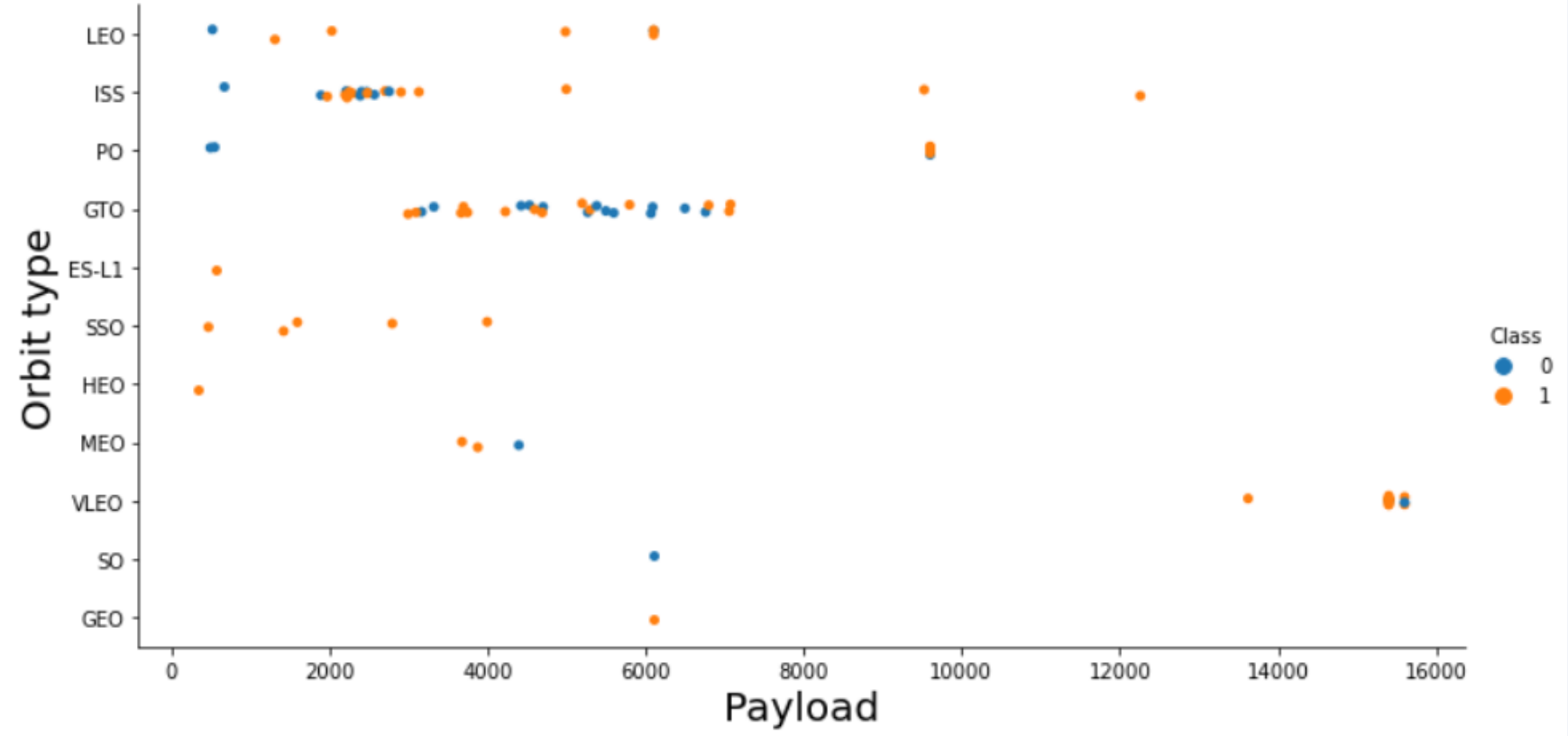
Flight Number vs. Orbit Type

- LEO orbit: Success rate increases with Flight Number
- GTO orbit: no relationship with Flight Number



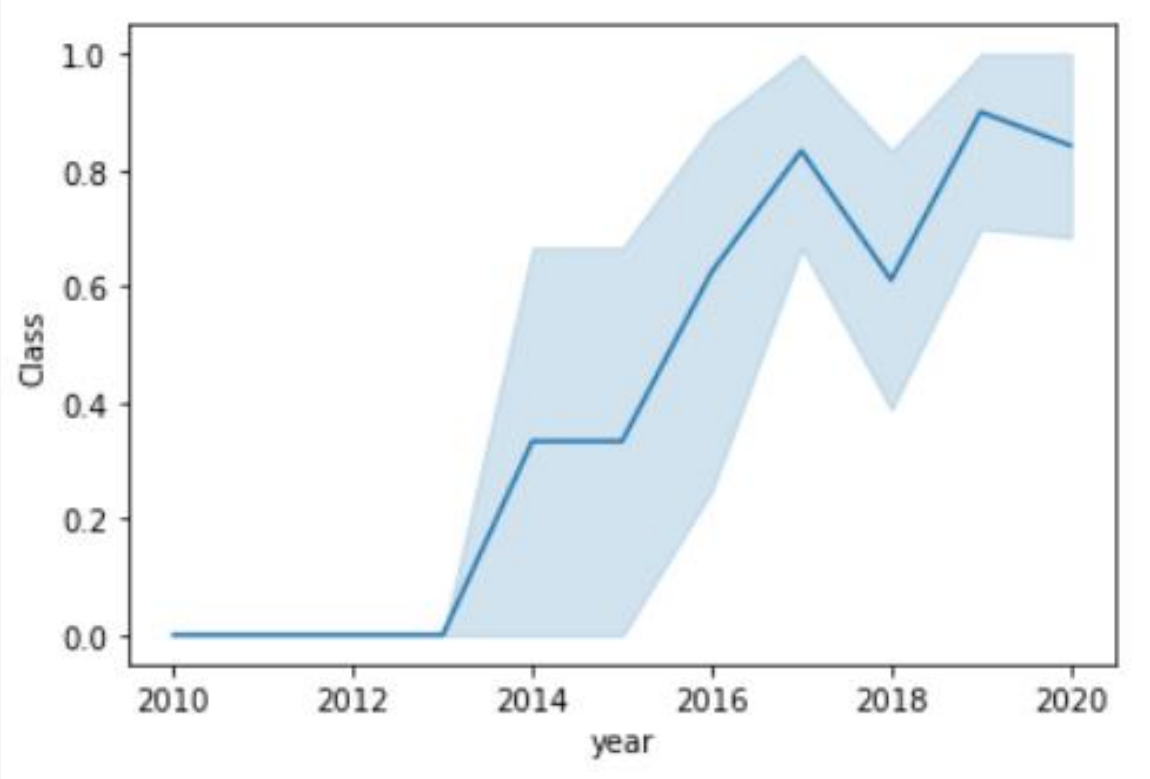
Payload vs. Orbit Type

- With heavy payloads the successful landing or positive landing rate are more for PO, LEO and ISS
- No specific pattern for GTO



Launch Success Yearly Trend

- The success rate increasing since 2013 till 2020



All Launch Site Names

- There are four different lauching sites of SpaceX in US

SQL magic query

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

Result

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

Launch Site Names Begin with 'CCA'

- There are 2 launch sites that begin with CCA
- First 5 records fetched from sql query are from CCAFS LC-40 site only!

Result from sql query

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

- Total payload mass carried by boosters from NASA was

SQL magic query

```
%sql select sum(payload_mass__kg_) as payloadMass from
SPACEXTBL where customer= 'NASA (CRS)';
```

Result

payloadmass
45596

Average Payload Mass by F9 v1.1

- Average Payload Mass carried by booster version is F9 v1.1 was 2928 kg

SQL magic query

```
%sql select avg(payload_mass__kg_) from
SPACEXTBL where (booster_version) like 'F9
v1.1';
```

Result

avg_payload_mass
2928

First Successful Ground Landing Date

- First successful landing outcome on ground pad was on 2015-12-22

SQL magic query

```
%sql select min(DATE) as firstSuccessGroundDate from
SPACEXTBL where LANDING__OUTCOME ='Success
(ground pad)';
```

Result

firstsuccessgrounddate
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- Four Boosters have successfully landed on drone ship and had payload mass between 4000 and 6000

SQL magic query

```
%sql select BOOSTER_VERSION from SPACEXTBL
where LANDING__OUTCOME='Success (drone ship)'
and PAYLOAD_MASS__KG_ BETWEEN 4000 and 6000;
```

Result

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 there were 101 mission outcomes in the database
- Only one of the outcome is Failure while 100 are a Success
- Among 100 success outcomes, payload status of one mission was unclear

SQL magic query

```
%sql select
MISSION_OUTCOME,count(MISSION_OUTCOME) as
counts from SPACEXTBL GROUP BY
MISSION_OUTCOME;
```

Result

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

Boosters Carried Maximum Payload

- There are 12 boosters which have carried the maximum payload mass

SQL magic query

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

Result

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

- There were two failed landing outcomes in drone ship from CCAFS launch site in 2015

SQL magic query

```
%sql SELECT BOOSTER_VERSION,LAUNCH_SITE FROM
SPACEXTBL where
landing__outcome='Failure (drone ship)'AND
EXTRACT(YEAR FROM DATE)='2015';
```

Result

booster_version	launch_site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

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

- Between the date 2010-06-04 and 2017-03-20, there are 31 landing outcome records
- Among 31 landing outcomes, ‘No attempt’ has highest rank (10) while landing outcome ‘Precluded (drone ship)’ has lowest rank (1)

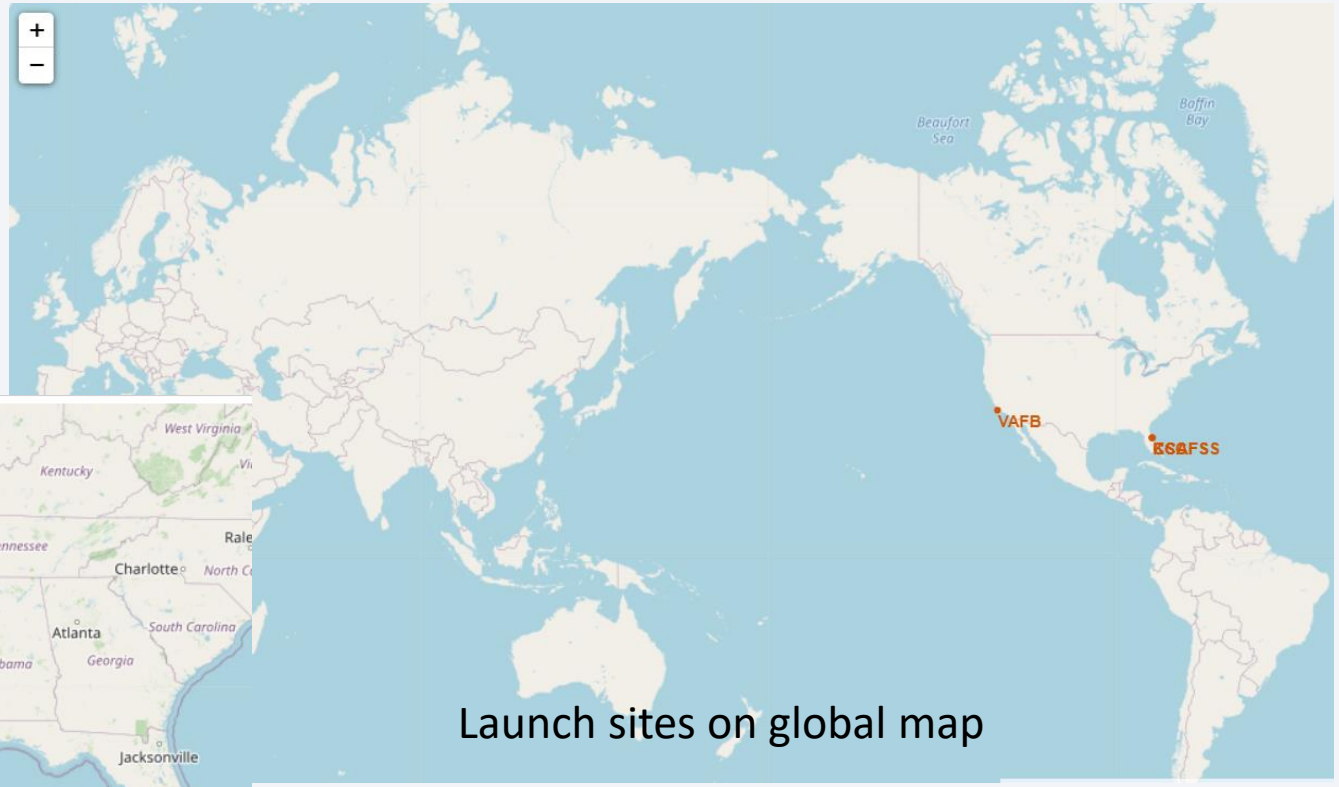
landing__outcome	con
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

Section 3

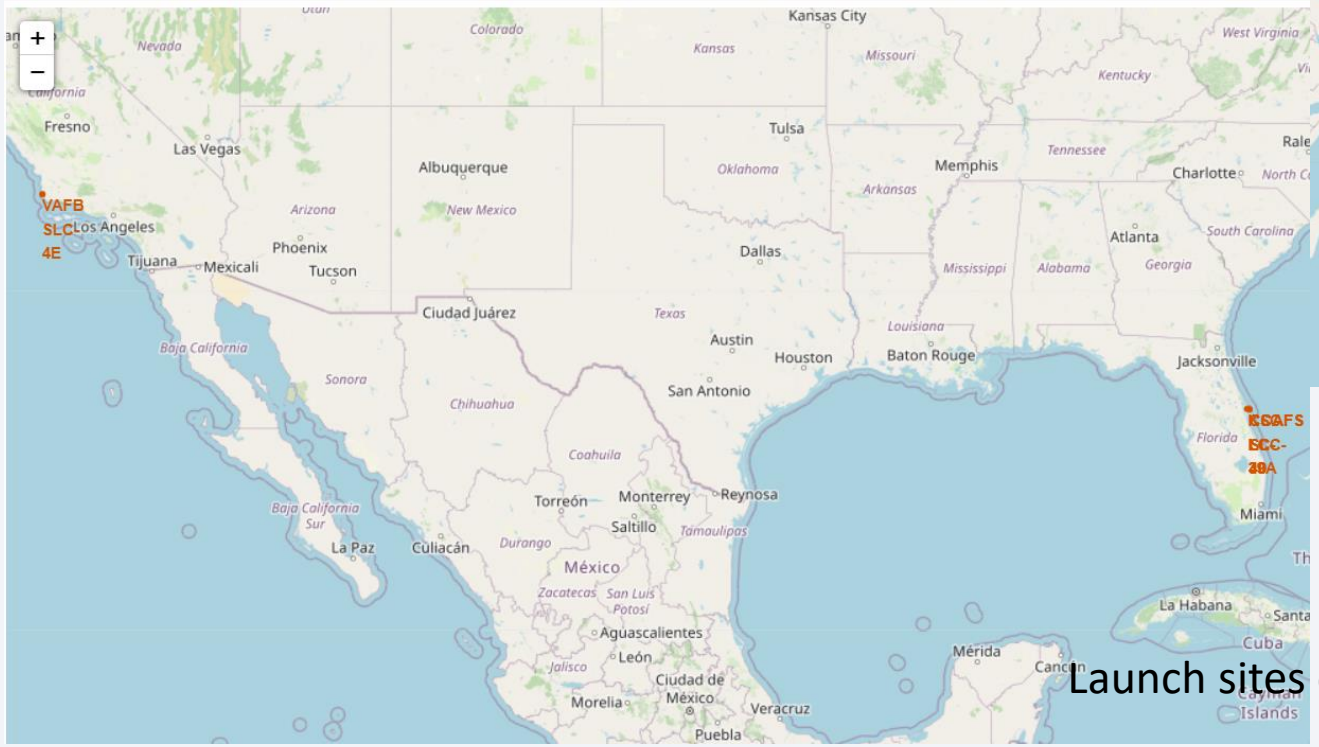
Launch Sites Proximities Analysis

SpaceX Launching Sites on Global Map

- 4 launch sites of SpaceX in USA
- 3 launch sites: East Coast
- 1 launch site: West Coast



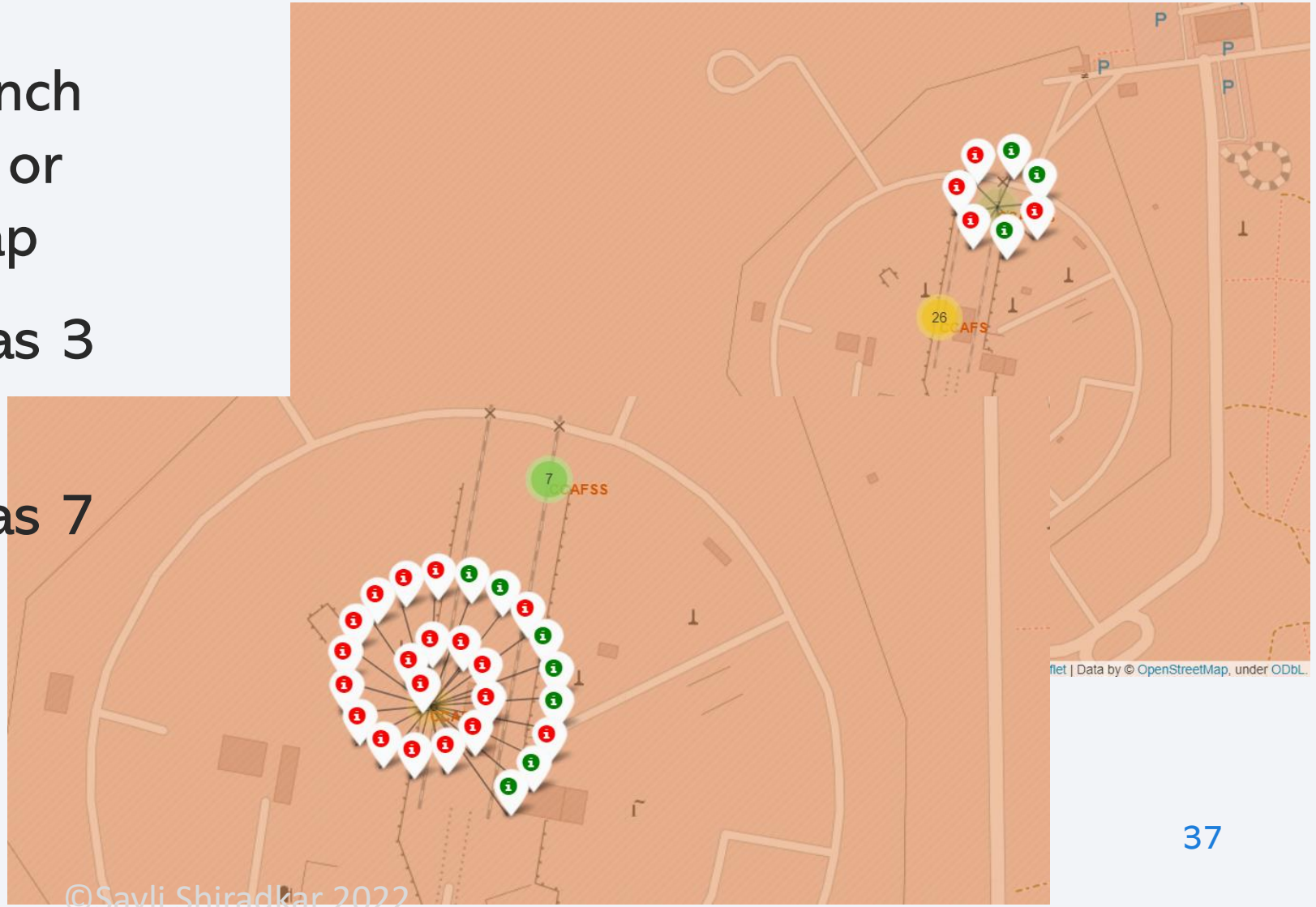
Launch sites on global map



Launch sites on USA map

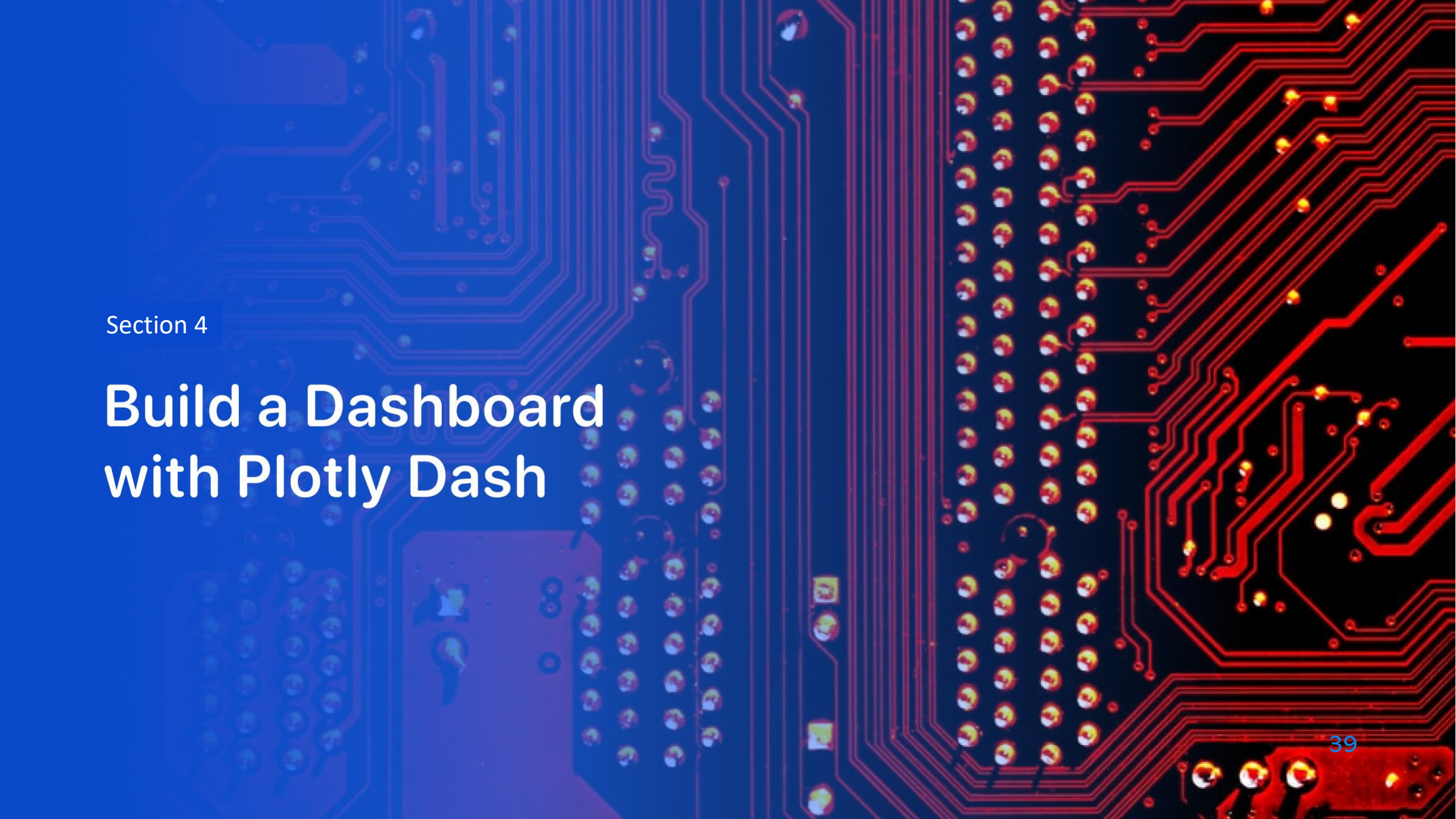
Launch Outcomes per Launch Sites

- We can visualize launch outcome as Success or Failure on folium map
- CCAFS LC-40 site has 3 successes
- CCAFS LC-40 site has 7 successes



Distance between Launch Site and its proximities

- Folium map shows distance between Launch Site CCAFS SLC-40 and its proximities
- Launch site to
  - Coastline = 0.86 KM
  - Railway track=1.28 KM
  - Melbourne city = 51.43 KM
  - Highway=0.58 KM
- Launch site is away from city for public safety
- Launch site is closer to coastline to provide additional boost to the rocket and in case of catastrophe, debris can fall in ocean

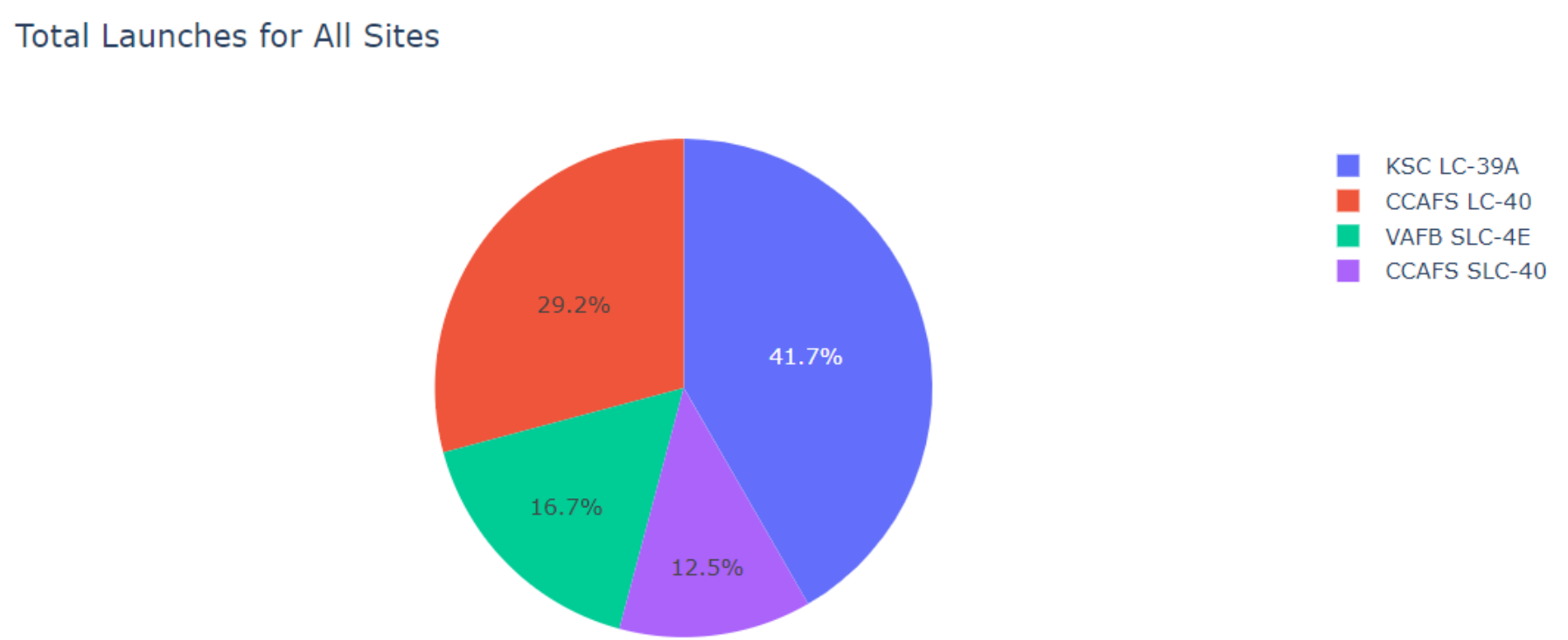


Section 4

Build a Dashboard with Plotly Dash

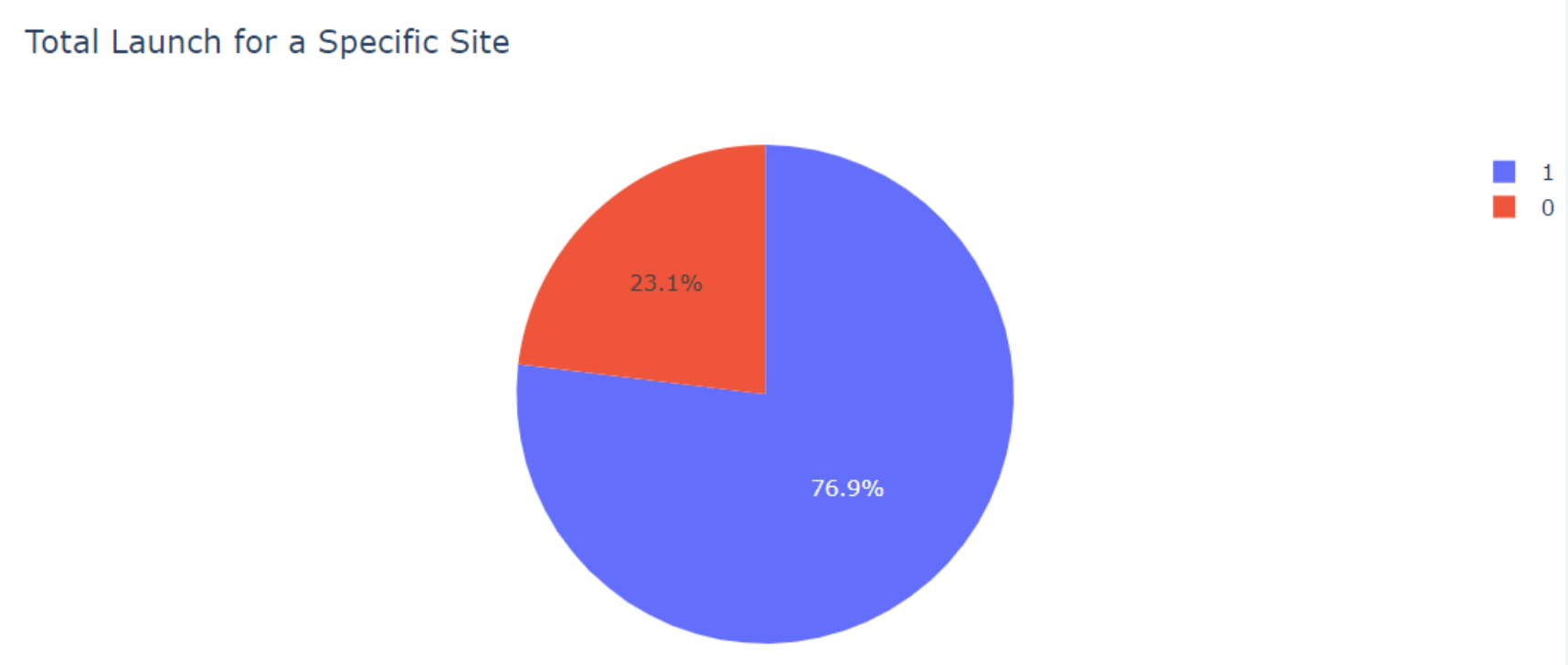
launch success count for all sites

- Number of launches with Success for all sites of SpaceX is shown in pie-chart
- Highest Success in KSC LC-39A site while lowest success rate is observed in CCAFS SLC-40



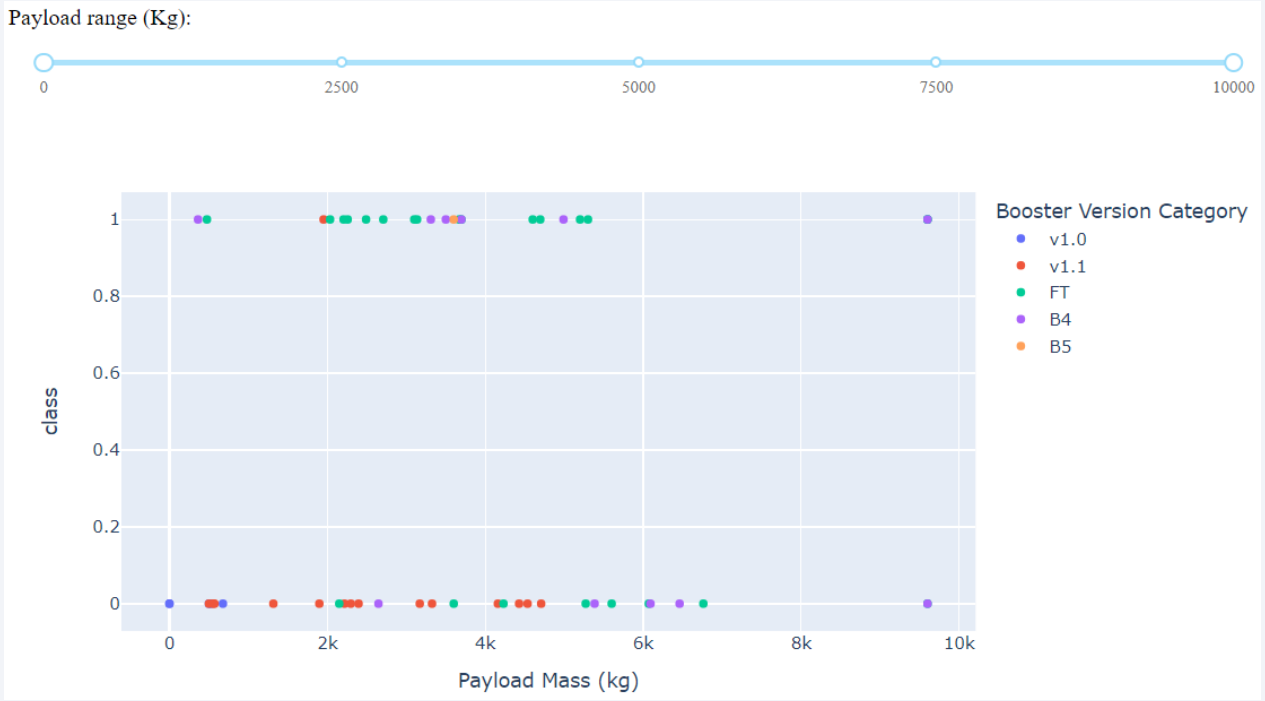
Launch site with highest success ratio

- KSC LC-39A site has highest launch success ratio with 76.9% launch Success rate and 23.1% failure rate



Payload vs. Launch Outcome

- Payload range varies between 362 to 9600 Kg for the highest launch success and 0 to 9600 Kg for the lowest launch success rate
- As the payload increases, success rate decreases
- Success rate of Launches is highest for Booster version category FT

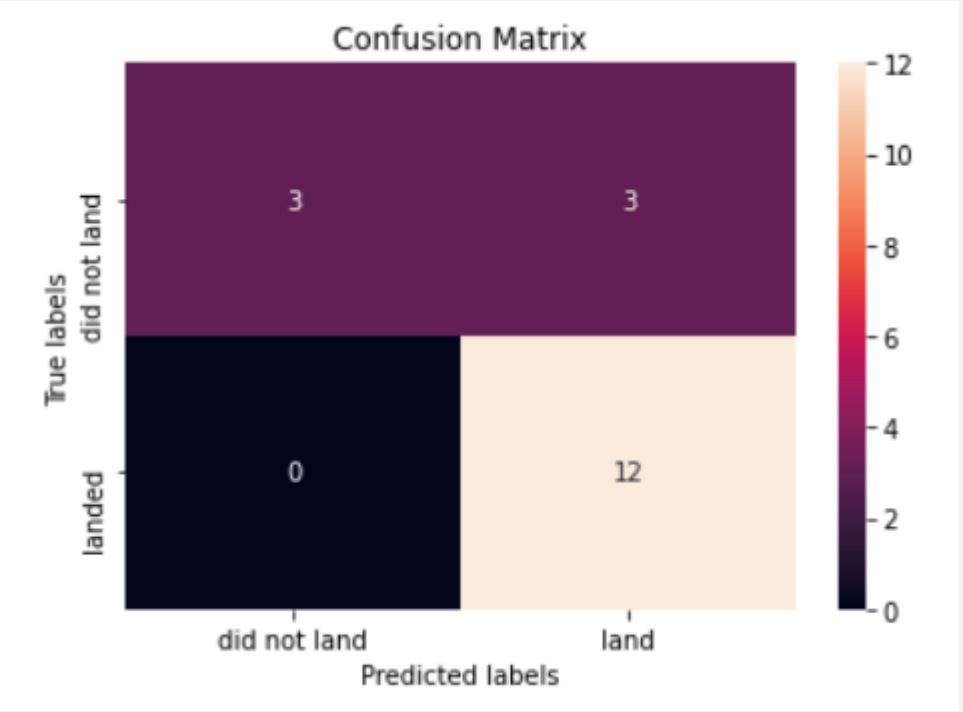


Section 5

Predictive Analysis (Classification)

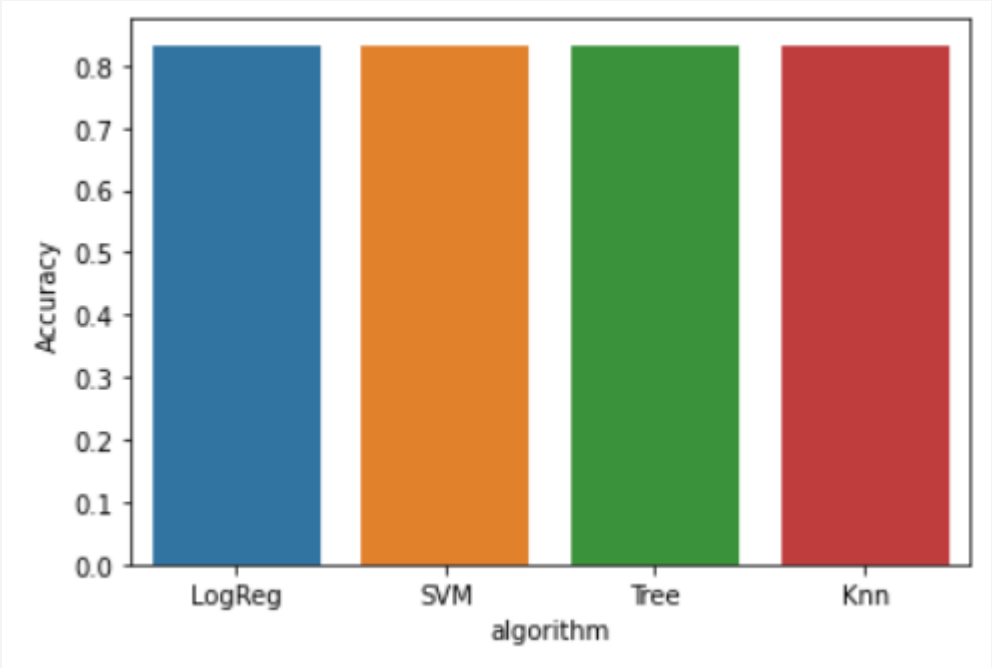
Confusion Matrix

- Confusion matrix for KNN, Tree, SVM and logistic Regression algorithms is same
- It correctly predicts landing Success outcomes
- However, for landing failures, models predict half of the times correctly
- False positive prediction rate 50%



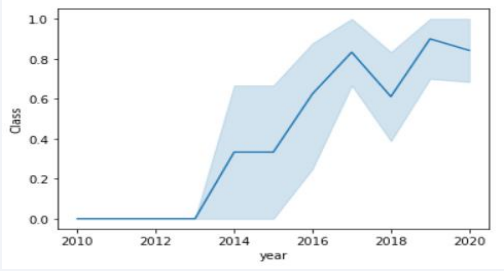
Classification Accuracy

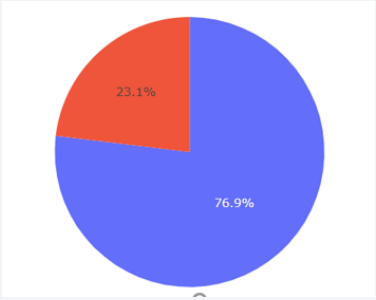
- Four classification models are built for first stage landing prediction :
  - Logistic Regression
  - SVM
  - Tree
  - Knn
- All built classification models have same classification accuracy=0.8334

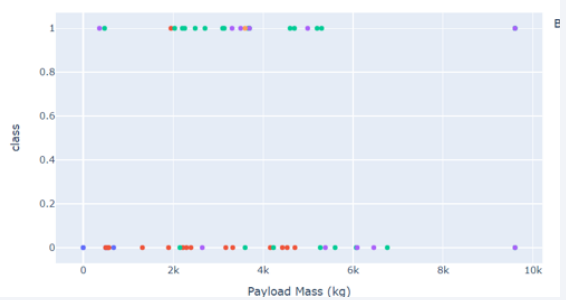


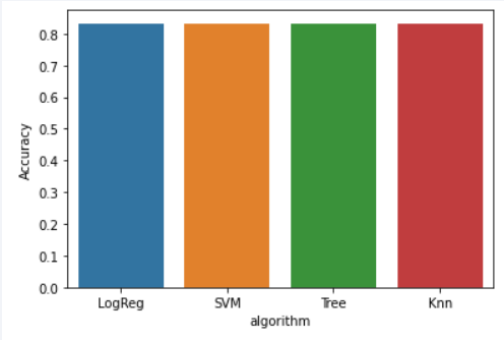
Conclusions

- SpaceX Launch Success Rate has been increasing since 2013
- High Success Rate Orbits: ES-L1, GEO, HEO, SSO
- Highest launch Success: KSC LC-39A site
- As payload increases, success rate seem to decrease
- KNN, SVM, LogReg and Decision Tree models have same accuracy (0.8334) for predicting launch outcome
- Future work: need to refine models further to improve accuracy









Thank you!

References

- GitHub URL of the completed SpaceX API calls notebook- https://github.com/sayliNS/DataScienceCapstone/blob/main/1.jupyter-labs-spacex-data-collection-api%20(1).ipynb
- GitHub URL of the web scraping notebook- https://github.com/sayliNS/DataScienceCapstone/blob/main/2.%20jupyter-labs-webscraping.ipynb
- GitHub URL of data wrangling notebook- https://github.com/sayliNS/DataScienceCapstone/blob/main/3.labs-jupyter-spacex-Data%20wrangling.ipynb
- GitHub URL EDA with data visualization notebook-https://github.com/sayliNS/DataScienceCapstone/blob/main/5.jupyter-labs-eda-dataviz%20(1)new.ipynb
- GitHub URL of EDA with SQL notebook https://github.com/sayliNS/DataScienceCapstone/blob/main/4.jupyter-labs-eda-sql-coursera%20(1)%20(1).ipynb
- GitHub URL of interactive map with Folium map - https://github.com/sayliNS/DataScienceCapstone/blob/main/6.lab_jupyter_launch_site_location%20(1).ipynb
- GitHub URL of Plotly Dash lab-https://github.com/sayliNS/DataScienceCapstone/blob/main/spacex_dash_app.py