Winning Space Race with Data Science

Milot M.
05 Dec 2023

Outline

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

Executive Summary

- The highest success ratio has KSC LC-39A site
- The most successful version that has the highest success ratio is FT Booster
- Payload Mass has the highest success ratio
- The most cost reducing is the Reusable Falcon 9 rocket

Introduction

- The aim of the project is to build an appropriate model for forecasting the likelihood of successful landing. It is as result of leveraging predictive data based on machine learning field
- Cost effectivity is very important for aerospace industry players and for the environment as well. In this project, SpaceX's Falcon 9 resulted as most effective one
- The methodology used in this project includes data collection, data wrangling, model selection and model evaluation
- Cost competitiveness in aerospace industry is very significant due to direct environment impact

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was collected using SpaceX API and Web Scraping from Wikipedia.
- Perform data wrangling
  - Missing values were replaced with Mean Value.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Different classification models were chosen and the best parameters were found for tunning. Also, the last evaluate accuracy and precision is presented.

Data Collection

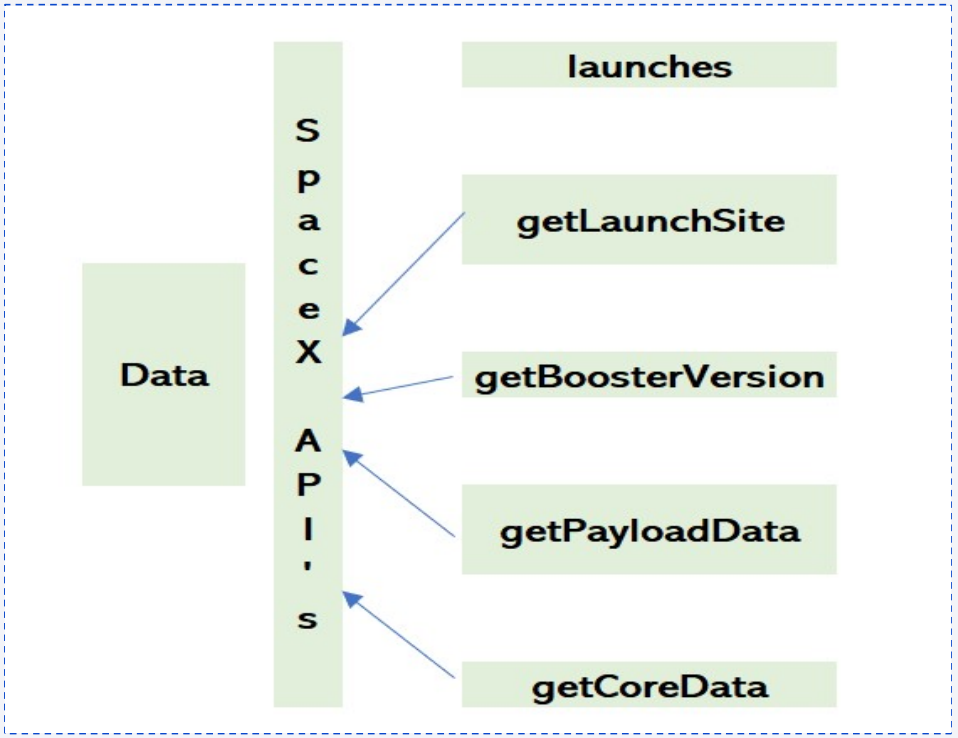
Data was collected using:

- SpaceX API and
- Web Scraping from Wikipedia.

Data Collection – SpaceX API

- Couple of API's were used for data collection, as presented on the right side.

- [testrepo/jupyter-labs-spacex-data-collection-api.ipynb at main · mmilotm/testrepo (github.com)]

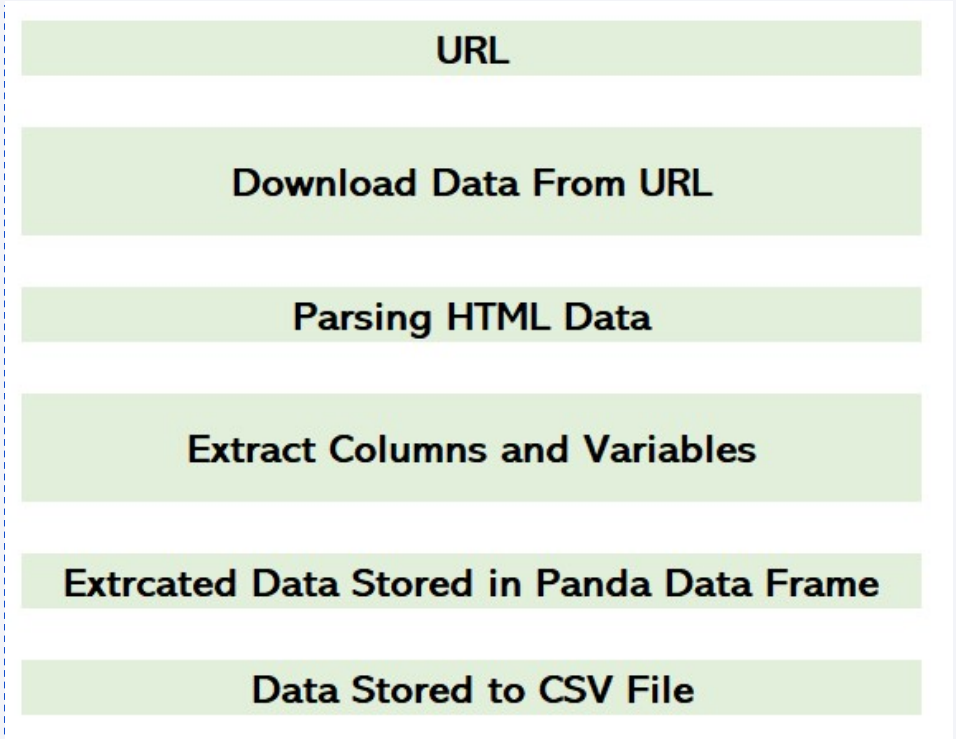


Data Collection - Scraping

- BeautifulSoup used for web scraping from Wikipedia.
- URL:

https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches

testrepo/jupyter-labs-webscraping.ipynb at main · mmilotm/testrepo (github.com)



Data Wrangling

- Cleaned the data
- Structured the data in a suitable form for analysis
- testrepo/labs-jupyter-spacex-Data wrangling.ipynb at main · mmilotm/testrepo (github.com)

EDA with Data Visualization

- Following Charts Plotted:
  - Flight Number vs Payload Mass
  - Flight Number vs Launch Site
  - Payload vs Launch Site
  - Flight Number vs Orbit type
  - Payload vs Orbit type
  - Success rate of each orbit type
- testrepo/jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb at main · mmilotm/testrepo (github.com)

EDA with SQL

- Summary of the SQL queries performed as below:
  - Unique launch site names in the space mission
  - 5 records where launch sites begin with the string 'CCA'
  - The total payload mass carried by boosters launched by NASA (CRS)
  - Average payload mass carried by booster version F9 v1.1
  - List the date when the first successful landing outcome in ground pad was achieved
  - List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
  - List the total number of successful and failure mission outcomes
  - The names of the booster versions which have carried the maximum payload mass using sub query
- [testrepo/jupyter-labs-eda-sql-coursera_sqllite.ipynb at main · mmilotm/testrepo (github.com)]

Build an Interactive Map with Folium

- Summary of the map objects used as below:
  - Marker – to specify locations on the map
  - Circle – presenting the circle around the point on the map
  - Line – to connect multiple points

- testrepo/jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb at main · mmilotm/testrepo (github.com)

Build a Dashboard with Plotly Dash

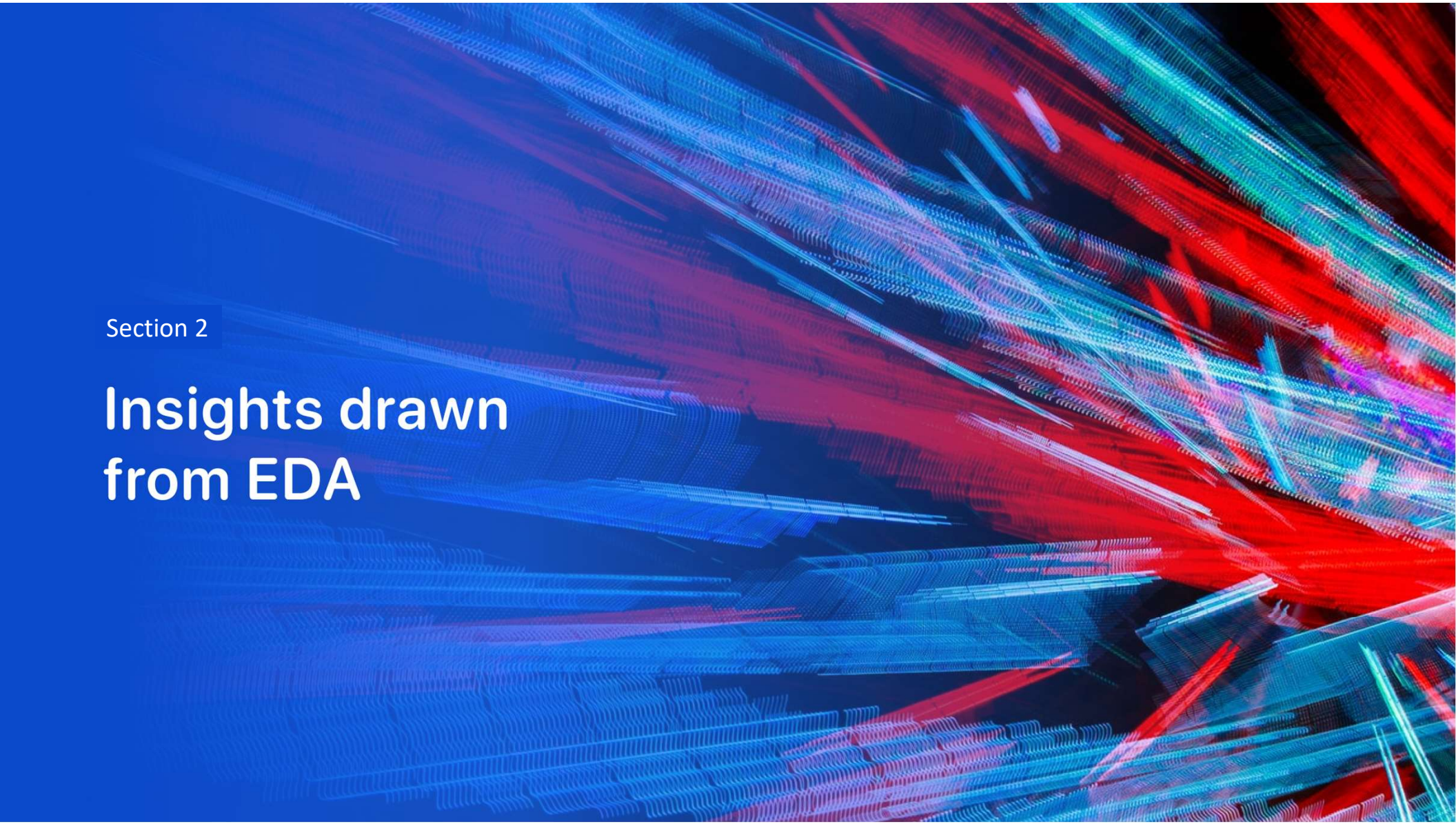
- Summarize what plots/graphs and interactions you have added to a dashboard
- Explain why you added those plots and interactions
- testrepo/jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb at main · mmilotm/testrepo (github.com)

Predictive Analysis (Classification)

- Created an output variable
- Standardized the data, cleaned, trained and tested
- Classification Models used: SVM, Classification Trees and Logistic Regression
- Grid search used to find the best tuned parameters for prediction
- Accuracy and Confusion Matrix is checked for model validation
- testrepo/SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb at main · mmilotm/testrepo (github.com)

Results

- Exploratory data analysis results
  - The success rate kept increasing till 2020 since 2013
- Interactive analytics demo in screenshots
  - KSC LC-39A site as the highest success ratio
- Predictive analysis results
  - Best Model Confusion Matrix

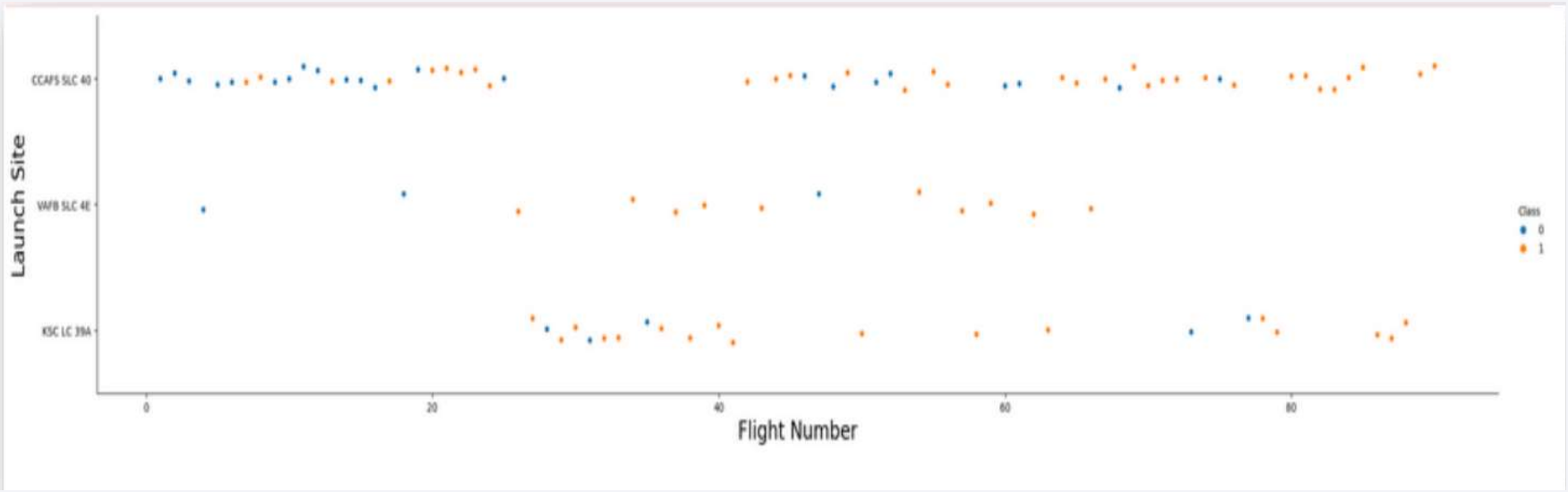


Section 2

Insights drawn from EDA

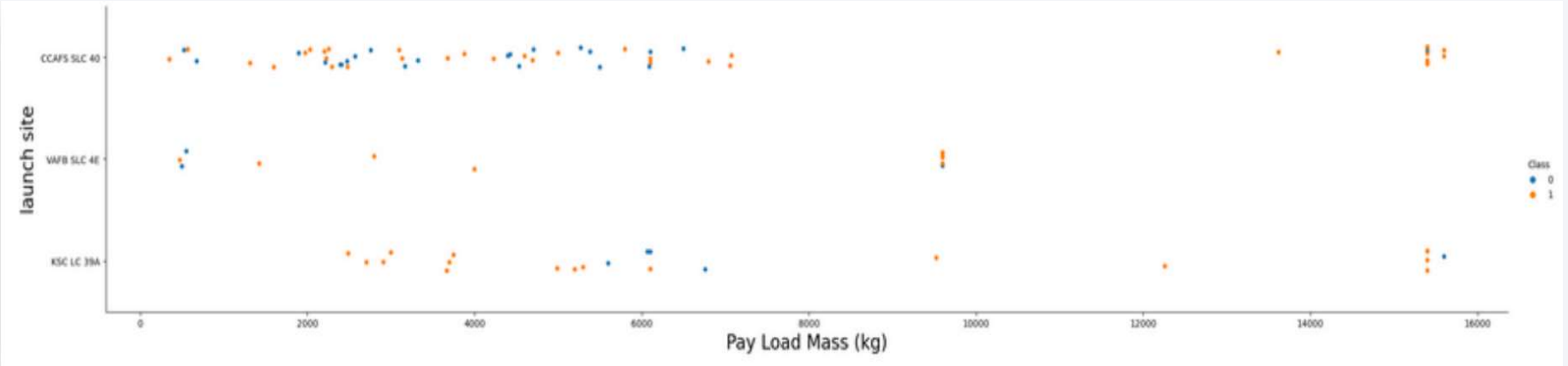
Flight Number vs. Launch Site

- Show a scatter plot of Flight Number vs. Launch Site
- Success ration is increasing along with flight number



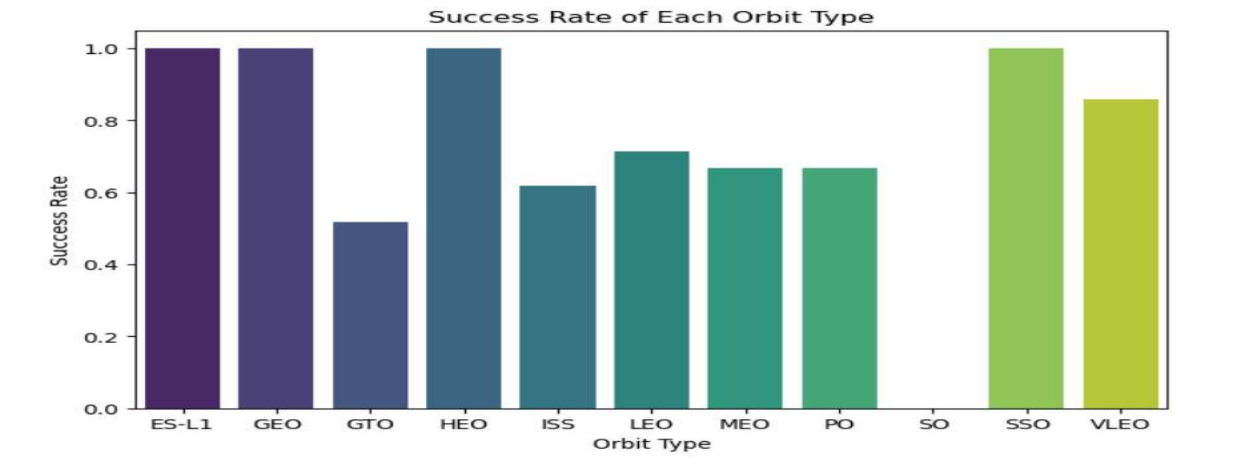
Payload vs. Launch Site

- Show a scatter plot of Payload vs. Launch Site
- VAFB-SLC launch site there are no rockets launched for heavy payload mass(greater than 10000).



Success Rate vs. Orbit Type

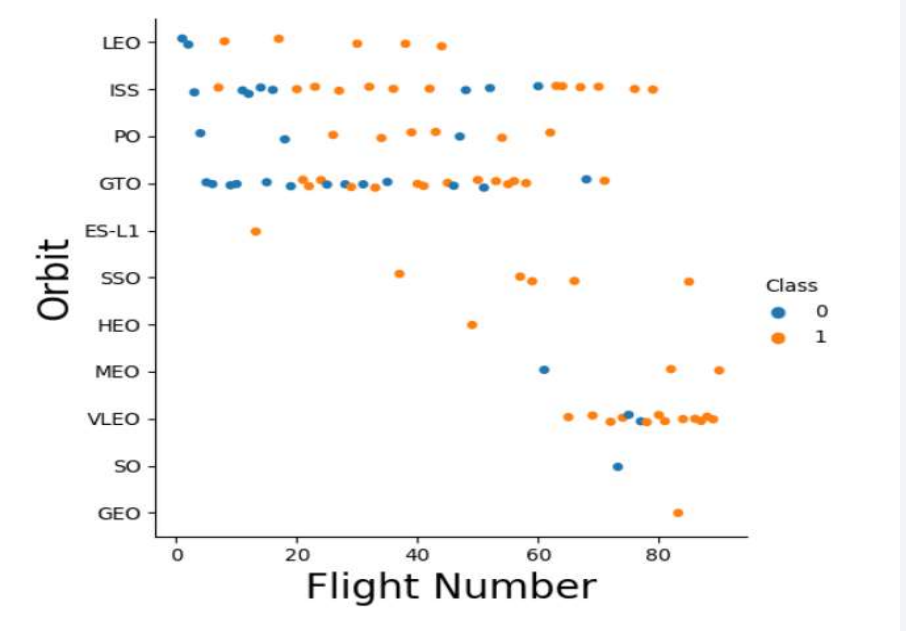
- Show a bar chart for the success rate of each orbit type
- ES-L1, GEO, HEO, SSO having success ratio as compared to others



Flight Number vs. Orbit Type

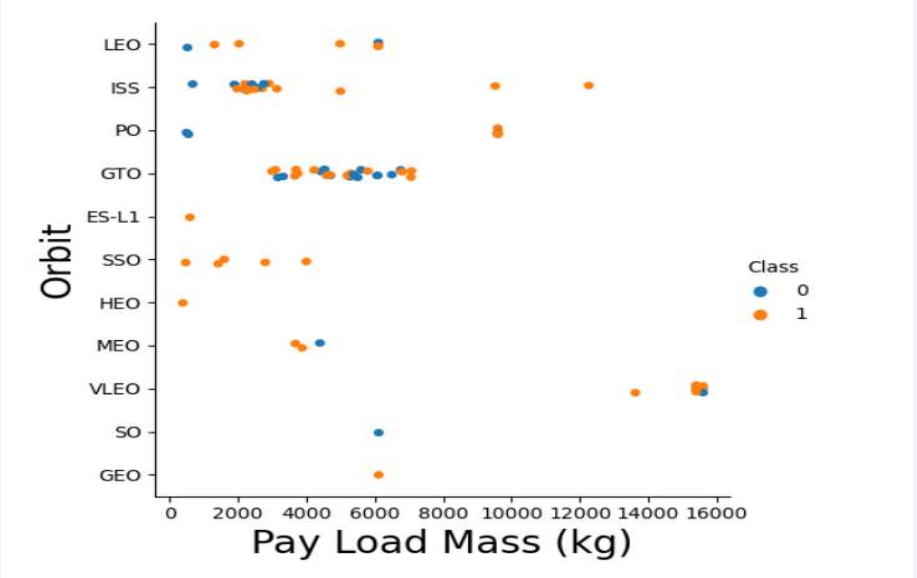
- Show a scatter point of Flight number vs. Orbit type

- LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit



Payload vs. Orbit Type

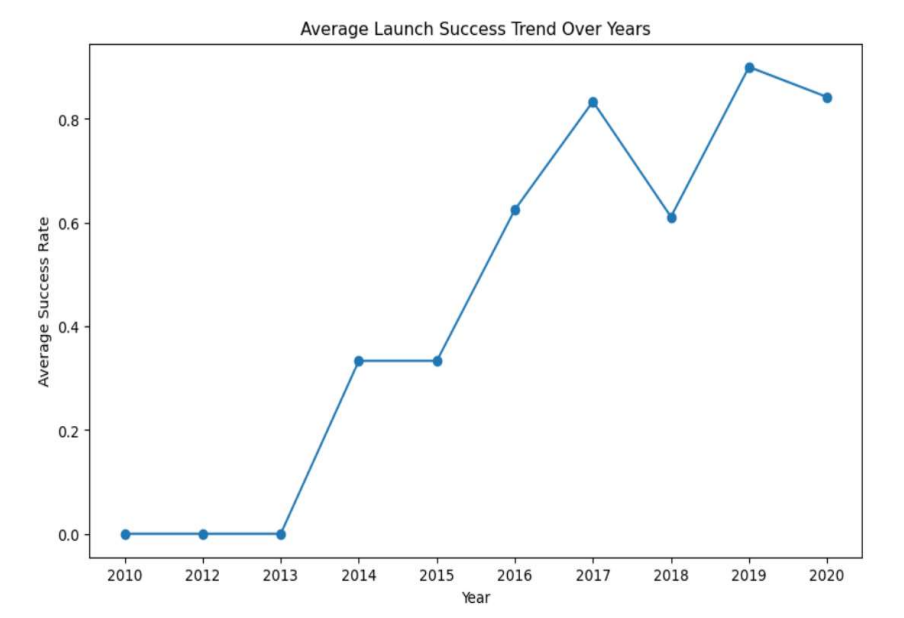
- Show a scatter point of payload vs. orbit type
- GTO we cannot distinguish as we have both values
- With heavy payloads the successful landing or positive landing rate are more for Polar, LEO and ISS plot with explanations



Launch Success Yearly Trend

- Show a line chart of yearly average success rate

- The success ratio kept increasing till 2020, since 2013.



All Launch Site Names

- Names of the unique launch sites are listed below:
  - CCAFS LC-40
  - VAFB SLC-4E
  - KSC LC-39A
  - CCAFS SLC-40

- The query result is presented as below:

```
%sql select distinct Launch_Site from  SPACEXTABLE
```

Launch Site Names Begin with 'CCA'

- 5 records where launch sites begin with `CCA`

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
6/4/2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
12/8/2010	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)
2/05/2012	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
10/8/2012	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
3/1/2013	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 carried by boosters from NASA
- The query result is presented below

```
%sql select sum(PAYLOAD_MASS__KG_) as total_payload_mass from SPACEXTABLE where Customer='NASA (CRS)'
```

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

total_payload_mass
45596

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1
- The query result is presented below

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

```
%%sql SELECT AVG(payload_mass__kg_) AS average_payload_mass
FROM SPACEXTBL
WHERE booster_version = 'F9 v1.1'
```

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

average_payload_mass
2928.4

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad
- The query result is presented below

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

Hint:Use min function

```
%%sql

SELECT Date as first_successful_landing_date
--min(Date) as first_successful_landing_date
FROM SPACEXTBL
WHERE Landing_Outcome = 'Success (ground pad)' limit 1;
```

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

first_successful_landing_date
22/12/2015

Successful Drone Ship Landing with Payload between 4000 and 6000

- List the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000
  - F9 FT B1022
  - F9 FT B1026
  - F9 FT B1021.2
  - F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes
- The query result is presented below

List the total number of successful and failure mission outcomes

```
%sql select Mission_Outcome,count(1) from SPACEXTBL group by Mission_Outcome
```

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

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

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass
- The query result is presented below

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

```
%%sql
SELECT Booster_Version
FROM SPACEXTBL
WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL);
```

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

- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015
- The query result is presented below

```
%%sql
SELECT substr(Date, LENGTH(Date) - 3) year, Date,Landing_Outcome,Booster_Version, Launch_Site
FROM SPACEXTBL
WHERE substr(Date, LENGTH(Date) - 3)='2015' and Landing_Outcome = 'Failure (drone ship)'
```

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

year	Date	Landing_Outcome	Booster_Version	Launch_Site
2015	1/10/2015	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
2015	14/04/2015	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

- 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
- The query result is presented below

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.

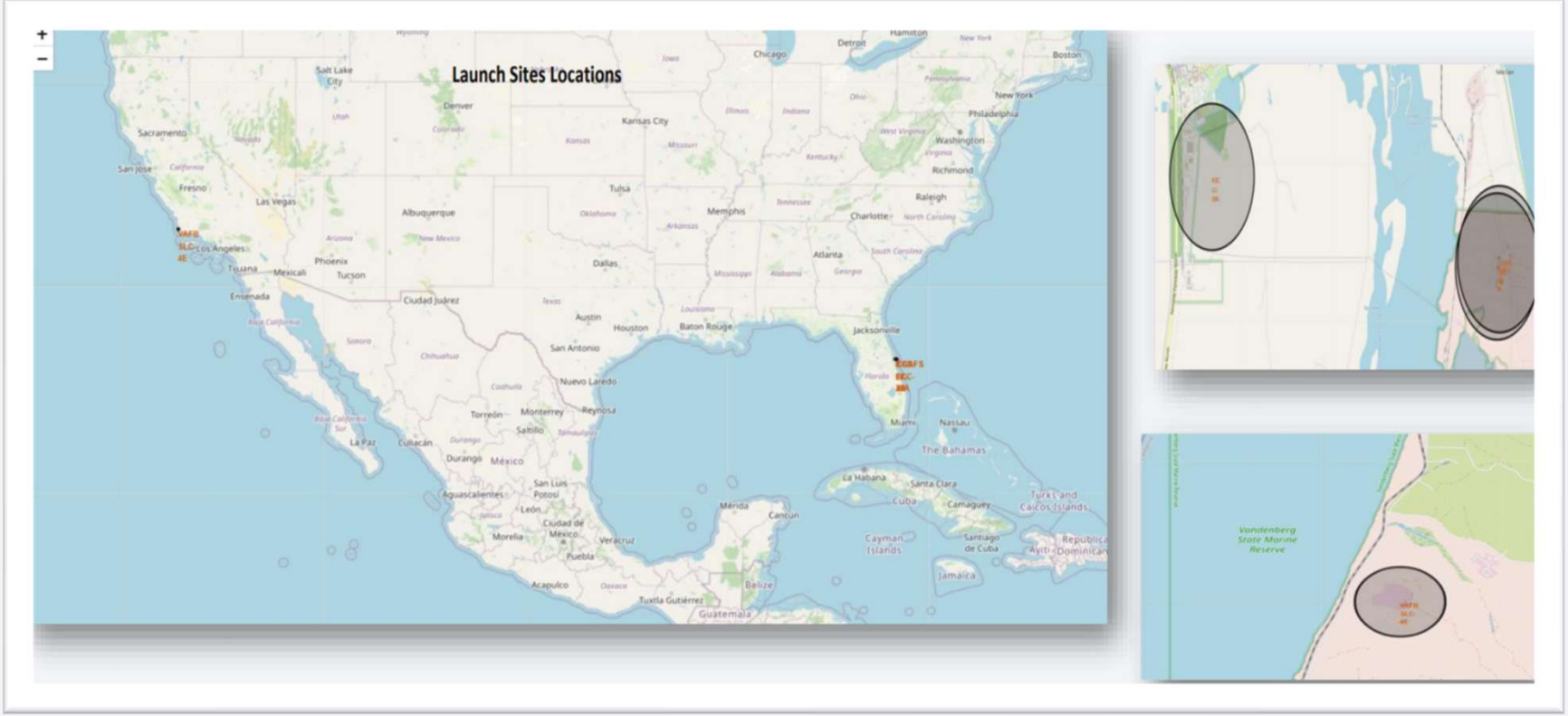
```
%%sql
WITH OutcomeCounts AS (
    SELECT
        Landing_Outcome,
        COUNT(*) AS outcome_count
    FROM
        SPACEXTBL
    WHERE
        Date BETWEEN '2010-06-04' AND '2017-03-20'
    GROUP BY
        Landing_Outcome
)

SELECT
    Landing_Outcome,
    DENSE_RANK() OVER (ORDER BY outcome_count DESC) AS ranking
FROM
    OutcomeCounts
ORDER BY
    ranking;
```

Section 3

Launch Sites Proximities Analysis

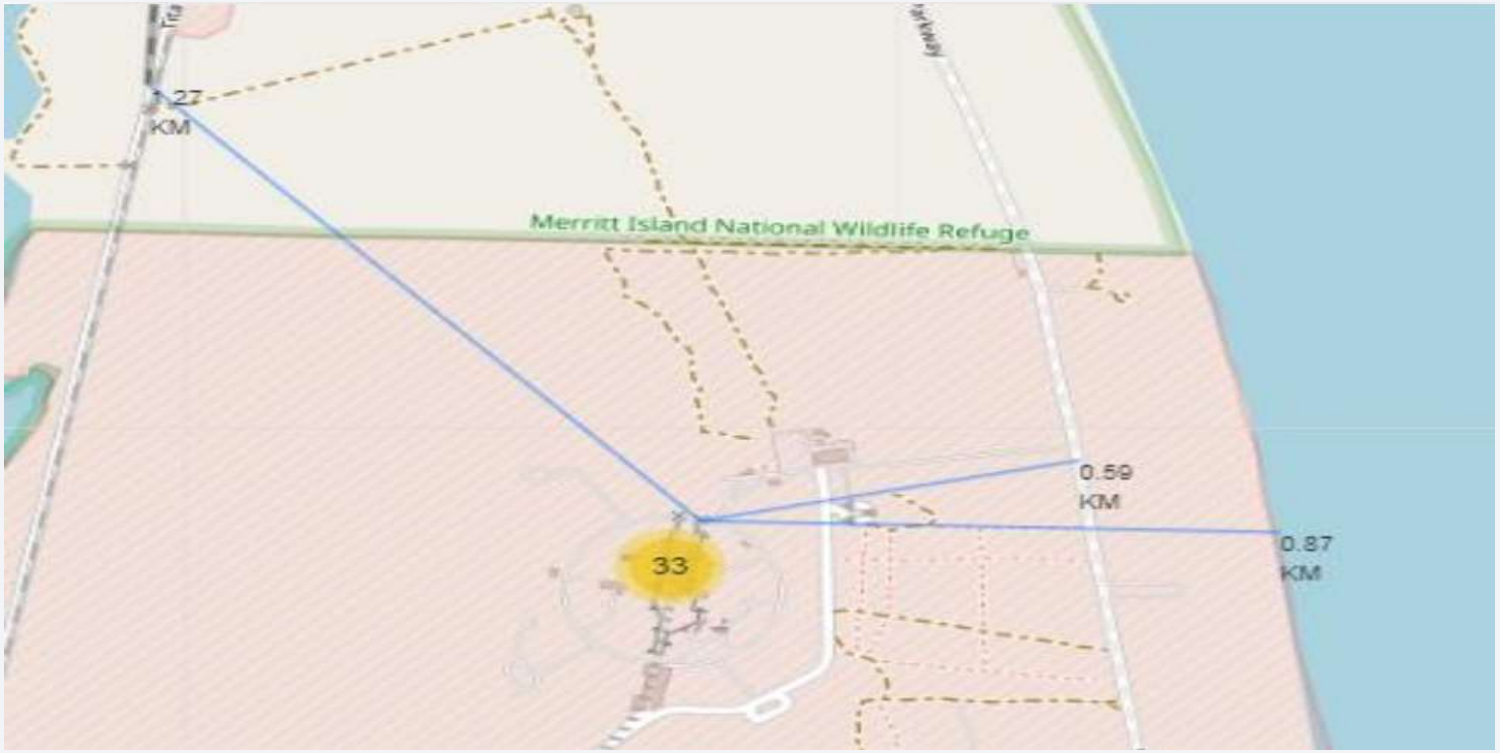
ALL SITES LOCATIONS

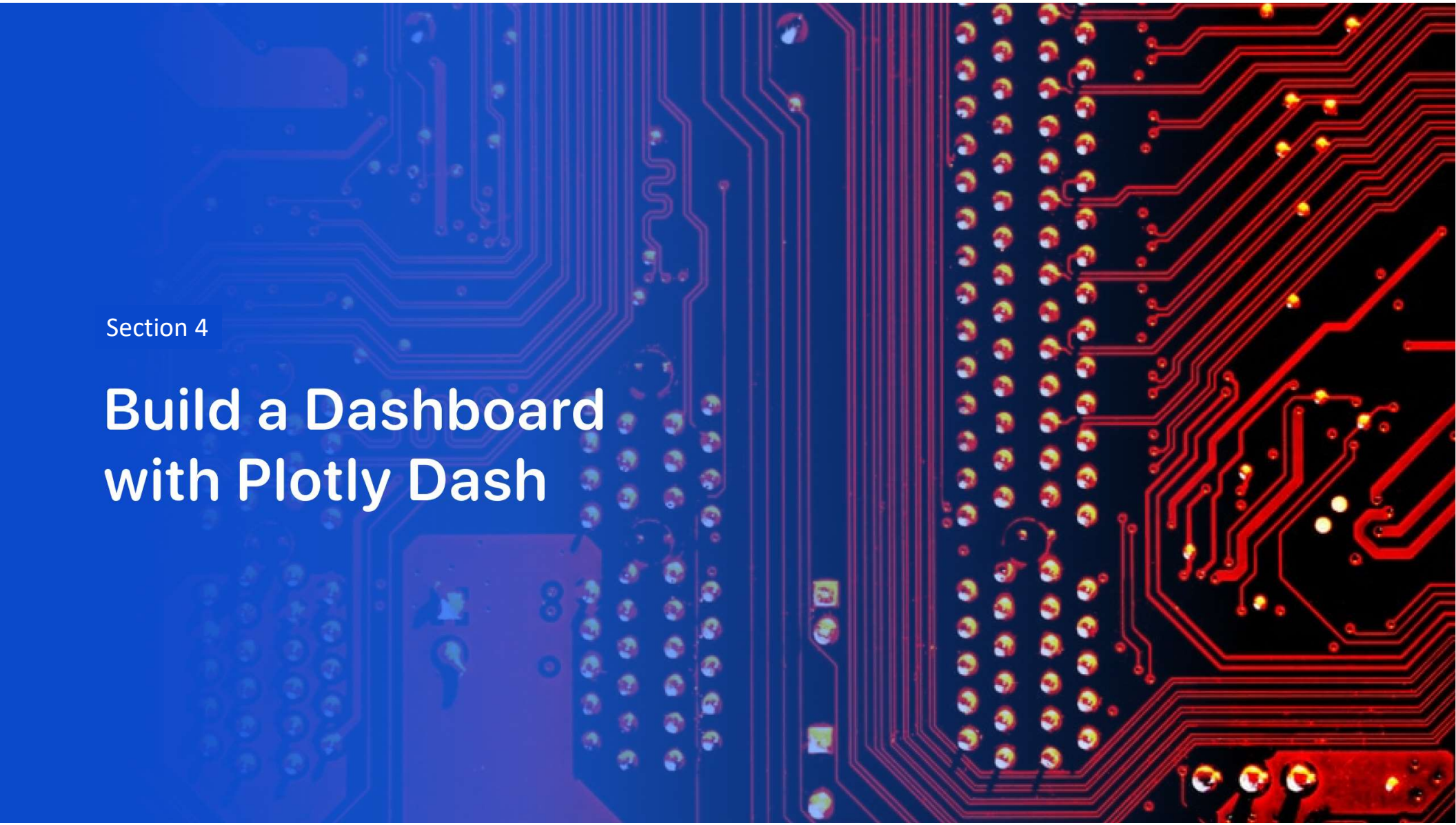


SITE SLC 40


VAFB SLC-4E
VAFB SLC-4E
VAFB SLC-4E
VAFB SLC-4E
VAFB SLC-4E
VAFB SLC-4E
VAFB SLC-4E
26
Space Launch Complex 40

SITE SLC 40

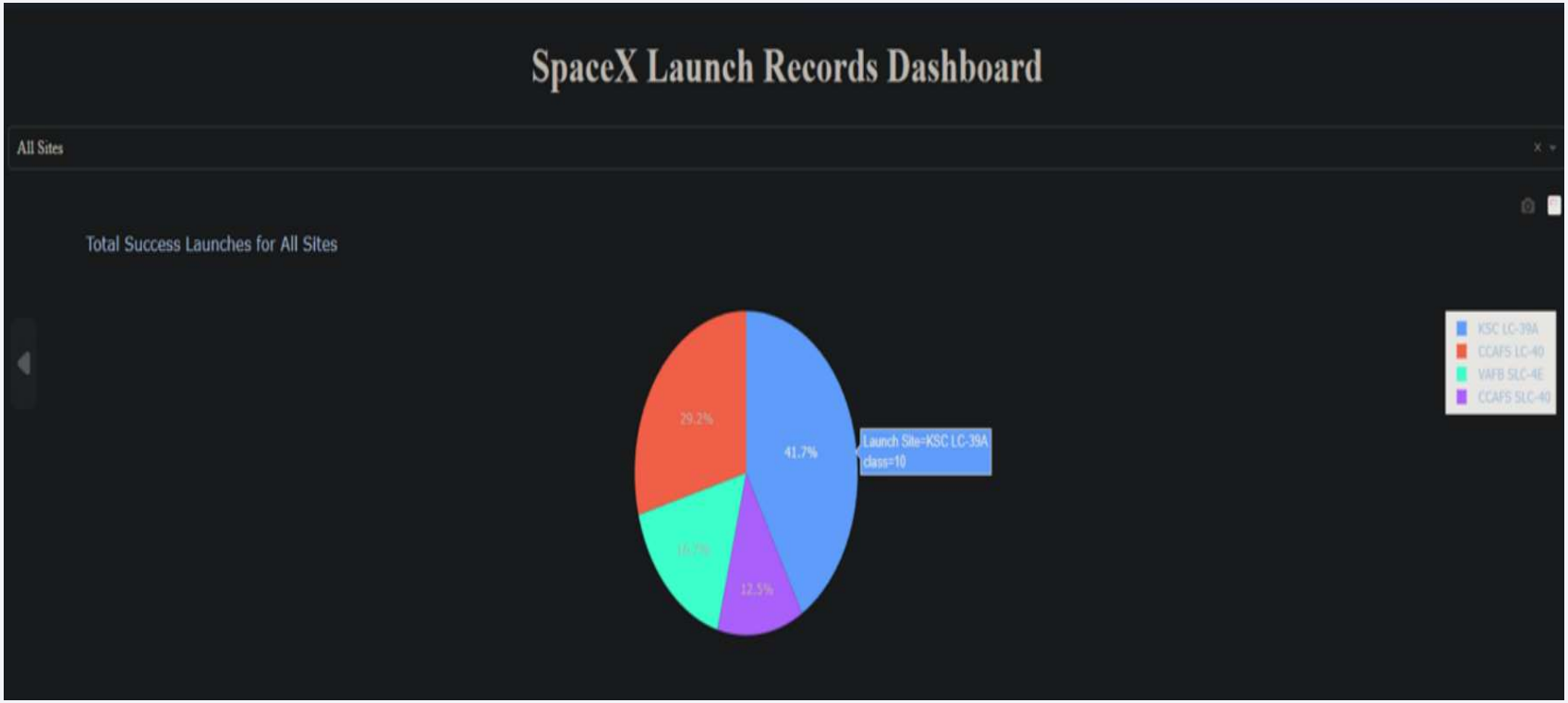

1.27
KM
Merritt Island National Wildlife Refuge
0.59
KM
0.87
KM
33



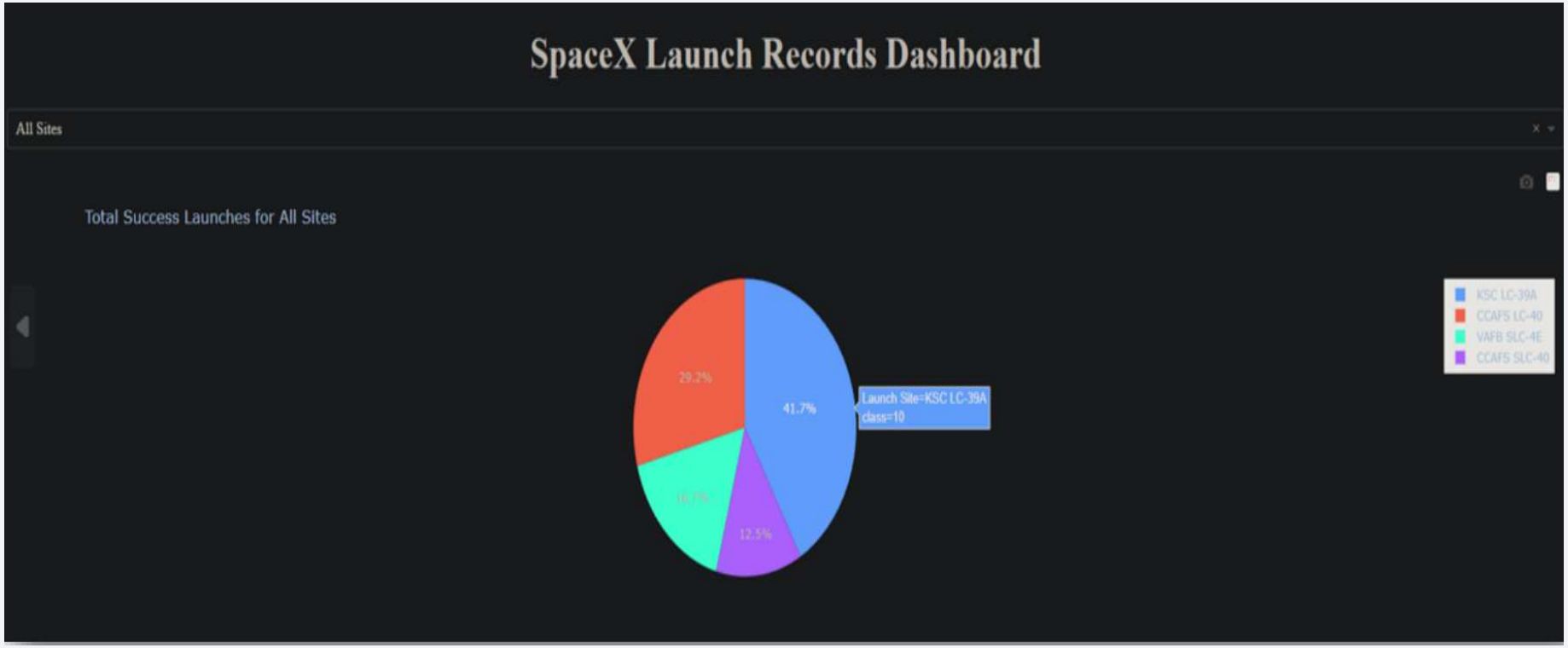
Section 4

Build a Dashboard with Plotly Dash

ALL SITES


SpaceX Launch Records Dashboard
All Sites
Total Success Launches for All Sites
29.2%
41.7%
12.5%
Launch Site=KSC LC-39A
class=10
KSC LC-39A
CCAFS LC-40
VAFB SLC-4E
CCAFS SLC-40

KSC LC-39A – 41.7% Success Ratio



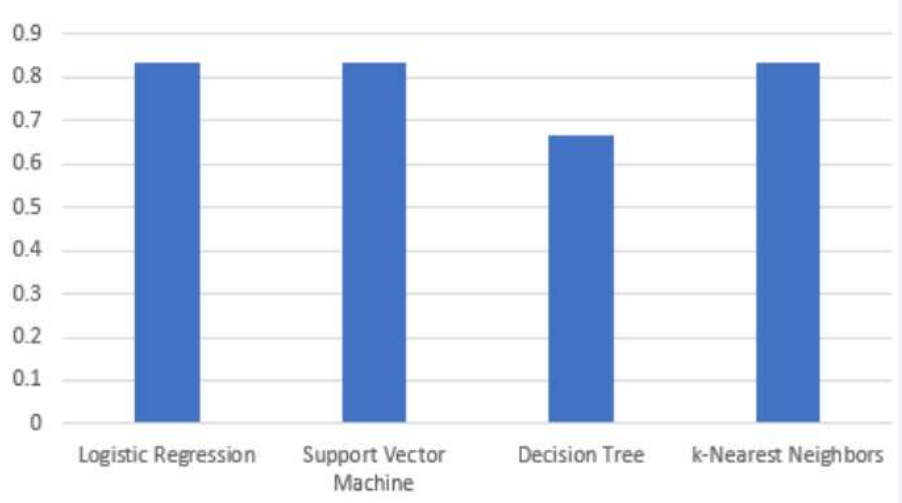
PAYLOAD vs. LAUNCH OUTCOME

Section 5

Predictive Analysis (Classification)

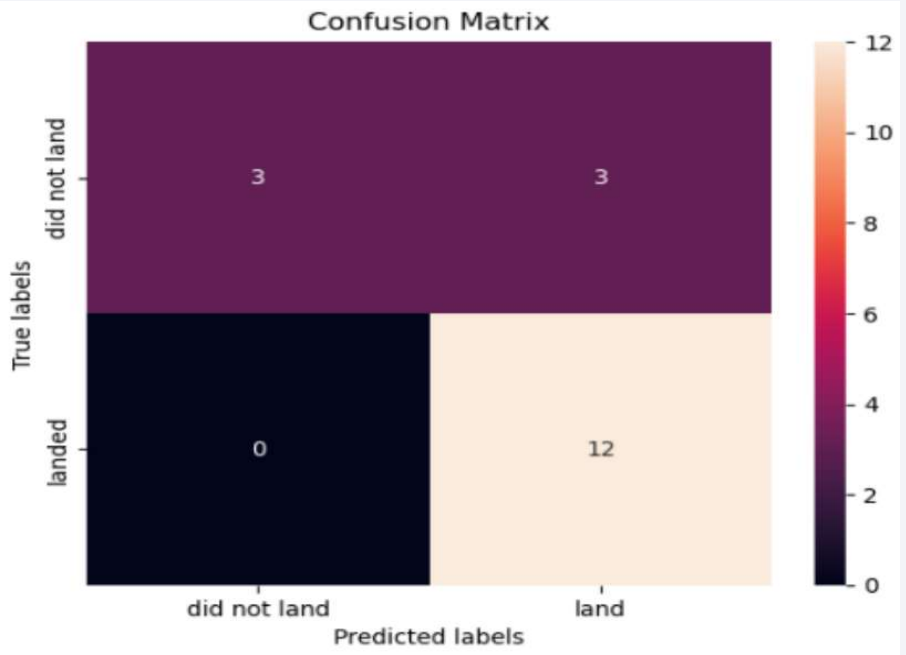
Classification Accuracy

- Highest accuracy of classification models 0.833333333333334
  - Logistic Regression
  - Support Vector Machine
  - KNN



Confusion Matrix

Confusion matrix with the best performing model has the highest sum of true positives and true negatives



Conclusions

- The highest success ratio has KSC LC-39A site
- The most successful version that has the highest success ratio is FT Booster
- Payload Mass has the highest success ratio
- The most cost reducing is the Reusable Falcon 9 rocket

Thank you!