IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Name>
<Date>

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection from **API** and **Web Scraping**
  - Exploratory Data Analysis using **Visualizations** and **SQL**
  - Create interaction dashboards using **Folium** and **Plotly**
  - Use **machine learning** models for predictive analysis
- Summary of all results
  - Visualizations
  - Dashboard
  - Model Results

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.

Section
1

Methodology

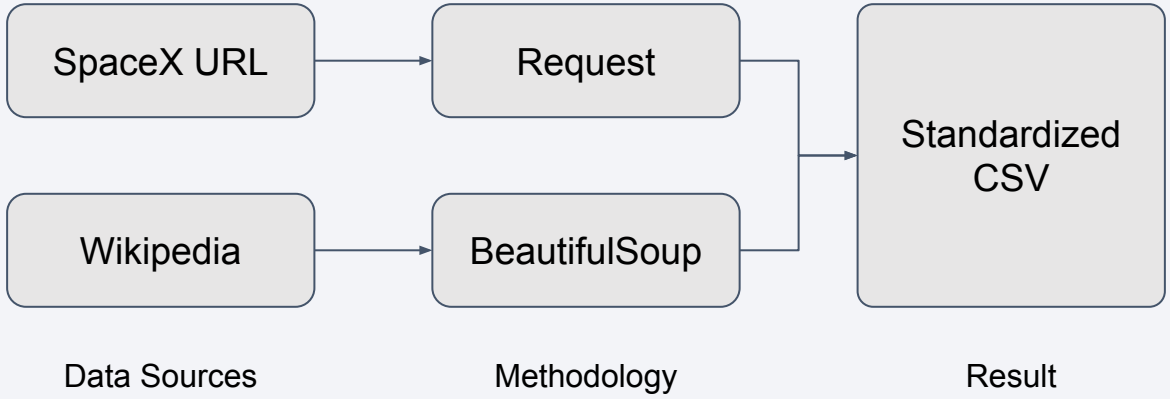
Methodology

Executive Summary

- Data collection methodology:
  - API and Web Scraping
- Perform data wrangling
  - Data Cleaning and Preprocessing
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Build and evaluation 4 machine learning models (Logistic regression, SVM, KNN, and Decision Tree)

Data Collection

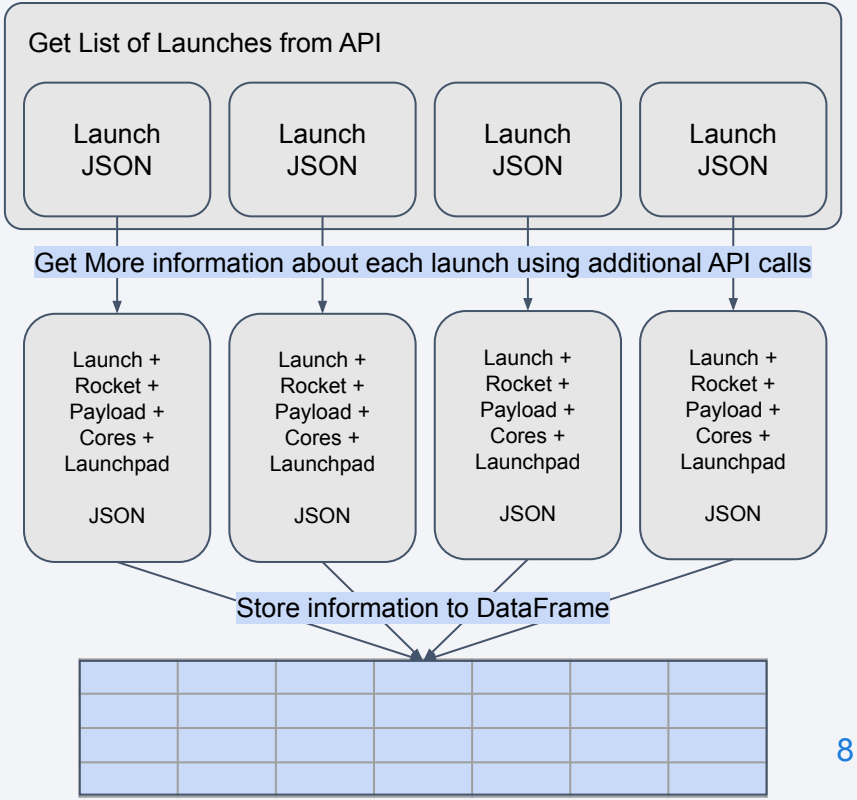
- Data was collected from two different sources:
  - SpaceX API
  - Wikipedia



https://github.com/liliwing/IBM-SpaceX-Falcon9/blob/main/Lab%2001%20-%20Data%20Collection%20through%20API.ipynb

Data Collection – SpaceX API

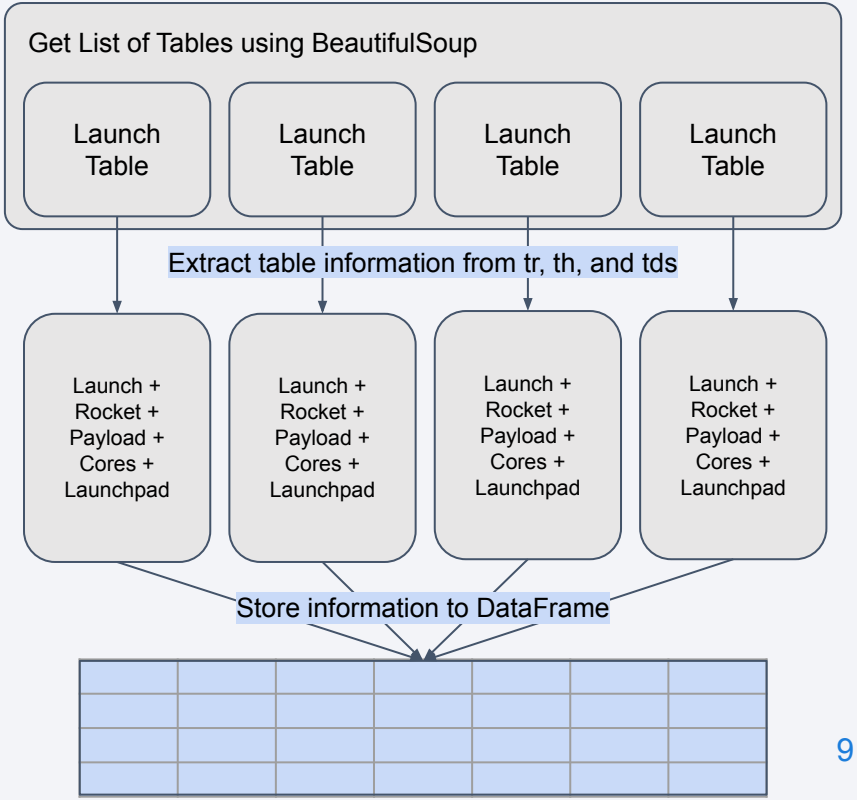
- Get all launches using the API
  https://api.spacexdata.com/v4/launches/past
- Get additional information about rockets
  https://api.spacexdata.com/v4/rockets/{rocket-id}
- Get additional information about launchpads
  https://api.spacexdata.com/v4/launchpads/{launchpad-id}
- Get additonal information about cores
  https://api.spacexdata.com/v4/cores/{core-id}
- Get additional information about payloads
  https://api.spacexdata.com/v4/payloads/{payload-id}
- Store all information in a DataFrame



https://github.com/liliwing/IBM-SpaceX-Falcon9/blob/main/Lab%2002%20-%20Data%20Collection%20through%20Web%20Scraping.ipynb

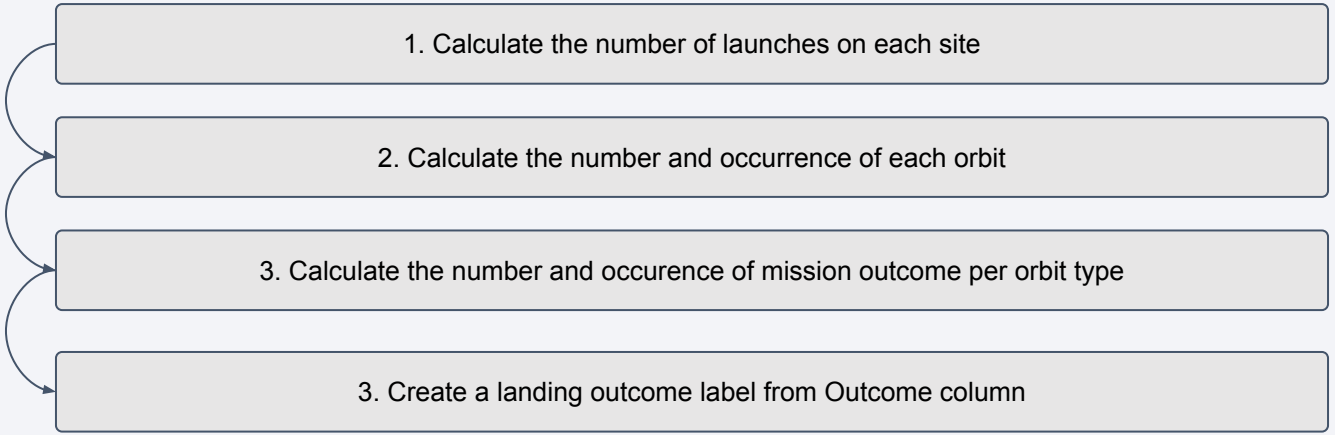
Data Collection - Scraping

- Scape Wikipedia web page
- Find all tables in the page with information about launches
- Process each table to extract detailed information for each launch



https://github.com/liliwing/IBM-SpaceX-Falcon9/blob/main/Lab%2003%20-%20Data%20Wrangling.ipynb

Data Wrangling



https://github.com/liliwing/IBM-SpaceX-Falcon9/blob/main/Lab%2004%20-%20EDA%20Visualization.ipynb

EDA with Data Visualization

Charts

- Flight Number vs Launch Sites
- Payload Mass vs Launch Sites
- Success Rates vs Orbit Type
- Flight Number vs Orbit Type
- Payload Mass vs Orbit Type
- Success Rates vs Year

https://github.com/liliwing/IBM-SpaceX-Falcon9/blob/main/Lab%2005%20-%20EDA%20SQL%20Query.ipynb

EDA with SQL

SQL Queries

- Display unique launch sites in the space mission
- Display 5 records where launch sites begin with the string 'CCA'
- Display the total payload mass carried by boosters launched by NASA (CRS)
- Display average payload mass carried by booster version F9 v1.1
- List the date when the first succesful landing outcome in ground pad was acheived.
- 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
- List the names of the booster_versions which have carried the maximum payload mass
- 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.
- Rank the count of successful landing_outcomes between the date 04-06-2010 and 20-03-2017 in descending order.

https://github.com/liliwing/IBM-SpaceX-Falcon9/blob/main/Lab%2006%20-%20Launch%20Site%20Map.ipynb

Build an Interactive Map with Folium

Interactive Map

- Mark all launch sites on a map
- Mark the success/failed launches for each site on the map
- Calculate the distances between a launch site to its proximities

https://github.com/liliwing/IBM-SpaceX-Falcon9/blob/main/Lab%2007%20-%20Plotly%20Dashboard.py

Build a Dashboard with Plotly Dash

Dashboard

- Pie Chart showing the success and failure launches in each launch site
- Scatter Plot showing Outcome vs Payload Mass

https://github.com/liliwing/IBM-SpaceX-Falcon9/blob/main/Lab%2008%20-%20Predictive%20Analysis.ipynb

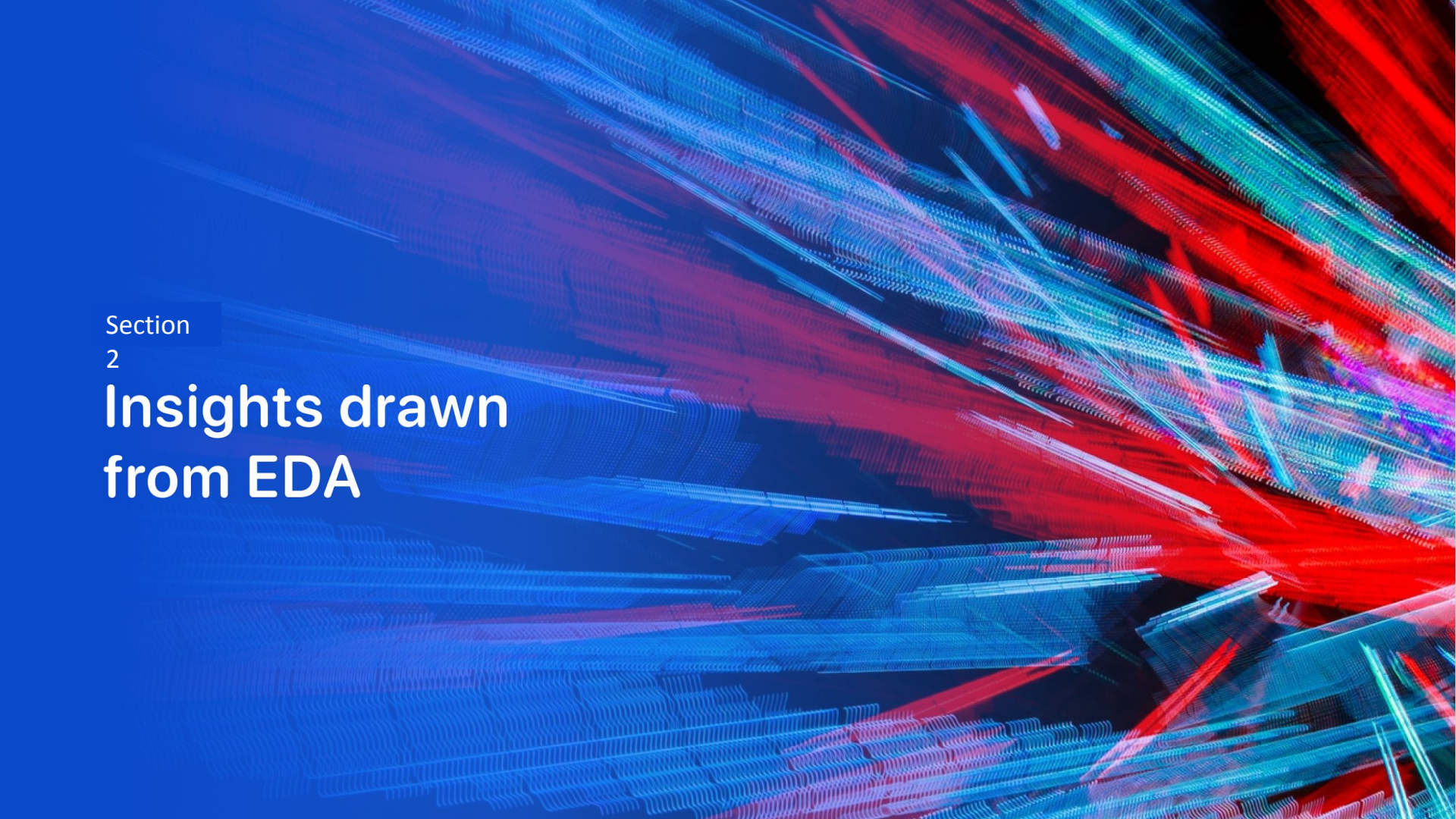
Predictive Analysis (Classification)

Modeling

- Trained and Evaluated four different machine learning models using Cross Validation
    - Logistic Regression
    - SVM
    - KNN
    - Decision Tree

Results

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

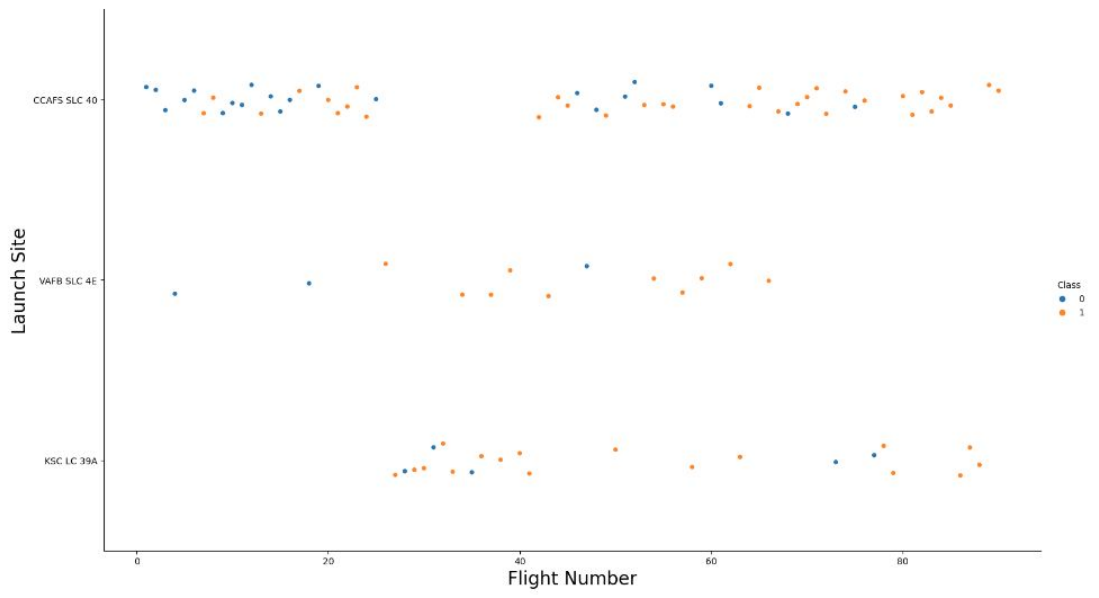


Section 2

Insights drawn from EDA

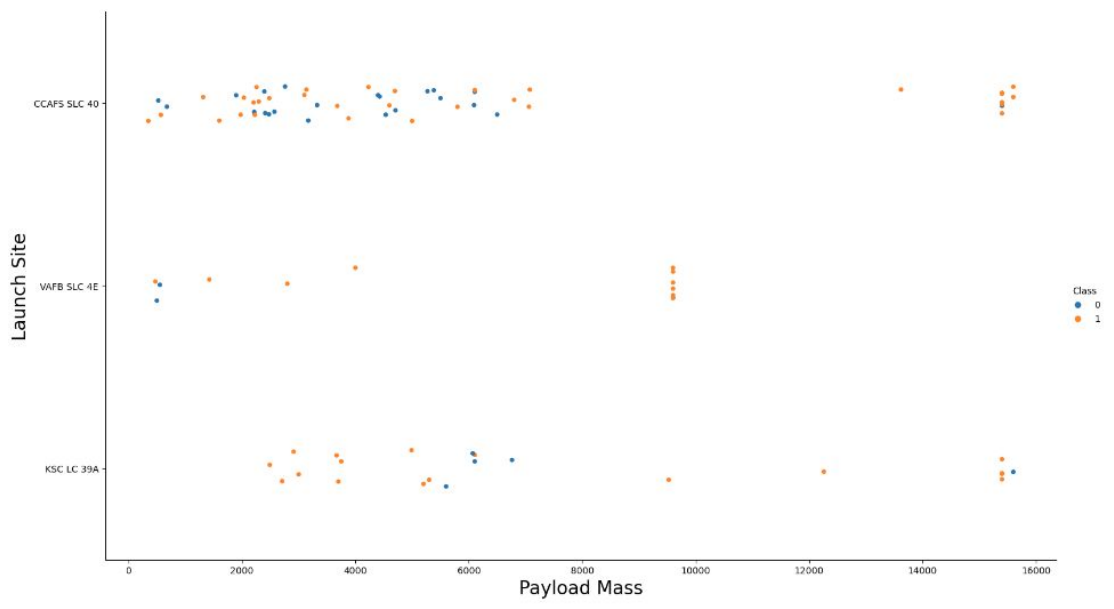
Flight Number vs. Launch Site

- Most flights are launched from launch site CCAFS SLC 40



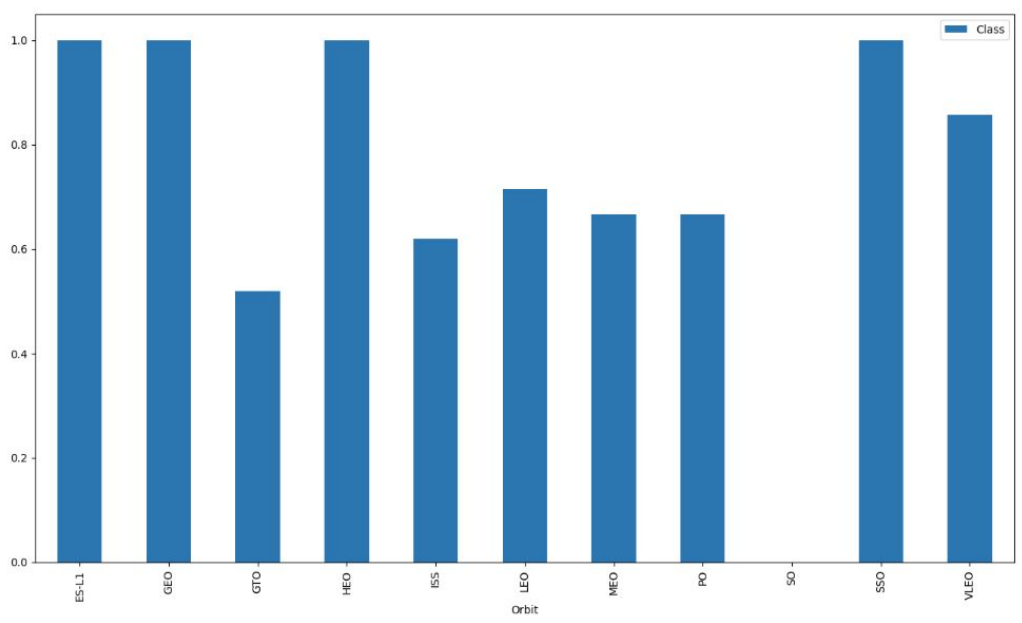
Payload vs. Launch Site

- Launches with payload mass above 8000 Kg has higher success rate
- Most launches with low payload mass are launched in CCAFS SLC 40



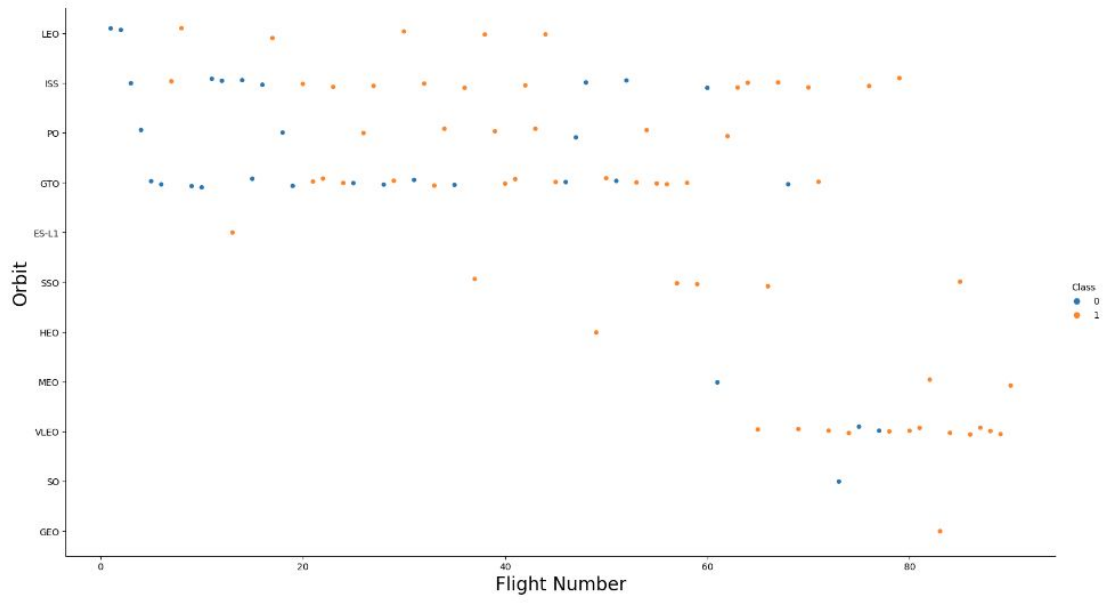
Success Rate vs. Orbit Type

- Orbit Type ESL1 GEO HEO has the highest success rate but only because there were only launched once
- SSO has the highest success rate



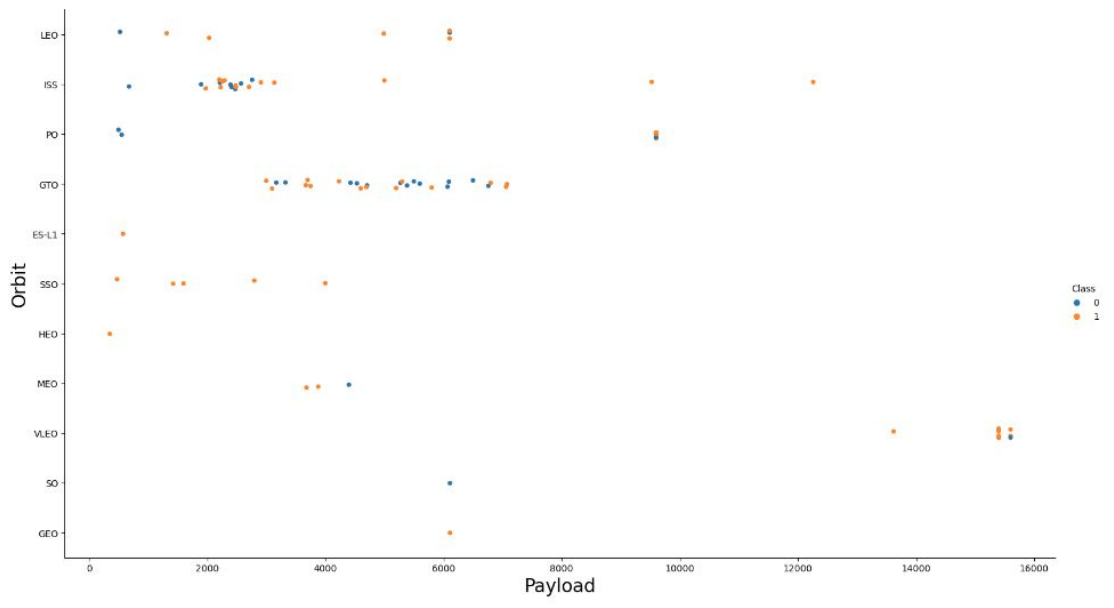
Flight Number vs. Orbit Type

- More launches are shifting to the VLEO orbit in recent years



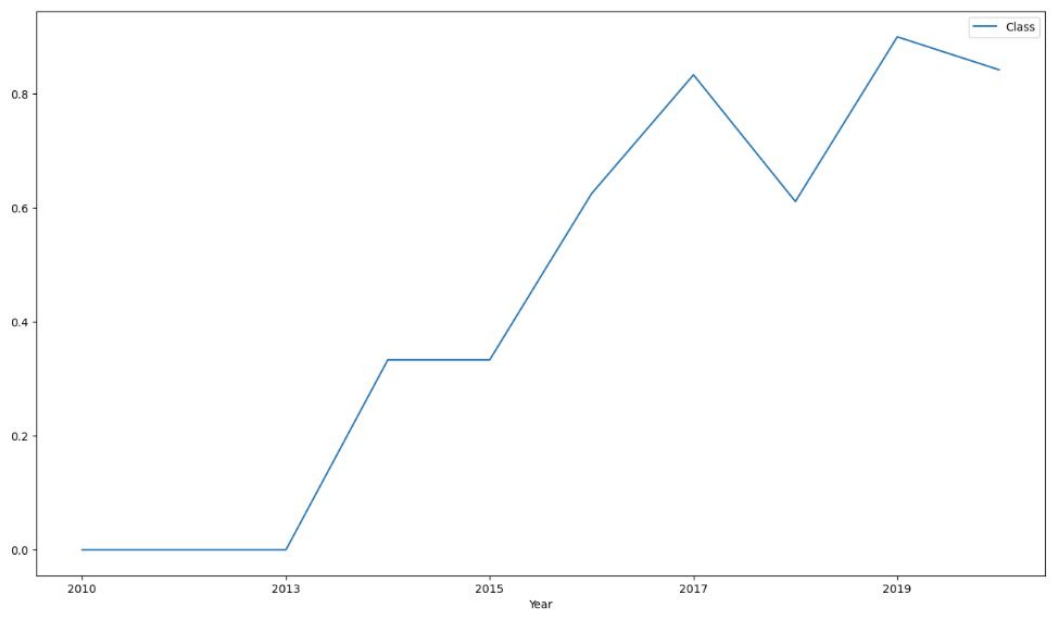
Payload vs. Orbit Type

- Most launches with payload mass in the range 3000 to 7000 uses GTO orbit
- VLEO is responsible for launches with higher payload mass



Launch Success Yearly Trend

- Success Rate increase significantly over the years



All Launch Site Names

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

```
%sql select distinct Launch_Site from SPACEXTBL;
```

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

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

Launch Site Names Begin with 'CCA'

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

```
%sql select * from SPACEXTBL where Launch_Site like 'CCA%' limit 5;
```

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

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome
04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
08-12-2010	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)
22-05-2012	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
01-03-2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

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

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

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

sum(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

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';
```

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

avg(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

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

Hint:Use min function

```
%sql select min(Date) from SPACEXTBL where "Landing _Outcome" = 'Success (ground pad)';
```

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

min(Date)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 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" = 'Success (drone ship)'
and "PAYLOAD_MASS__KG_" between 4000 and 6000;
```

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

List the total number of successful and failure mission outcomes

```
%%sql
select count(*) from SPACEXTBL
where "Mission_Outcome" like '%Success%'
or "Mission_Outcome" like '%Failure%';
```

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

count(*)
101

Boosters Carried Maximum Payload

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 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.

```
%%sql
select
substr(Date, 6, 2) as Month,
substr(Date, 1, 4) as Year,
"Landing _Outcome", "Booster_Version", "Launch_Site"
from SPACEXTBL
where Year='2015' and "Landing _Outcome" = 'Failure (drone ship)';
```

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

Month	Year	Landing _Outcome	Booster_Version	Launch_Site
01	2015	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
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 successful landing_outcomes between the date 04-06-2010 and 20-03-2017 in descending order.

```
%%sql
SELECT "Landing _Outcome", count(*)
FROM SPACEXTBL
WHERE Date Between '2010-06-04' and '2017-03-20'
and "Landing _Outcome" like '%Success%'
group by "Landing _Outcome"
order by count(*) desc;
```

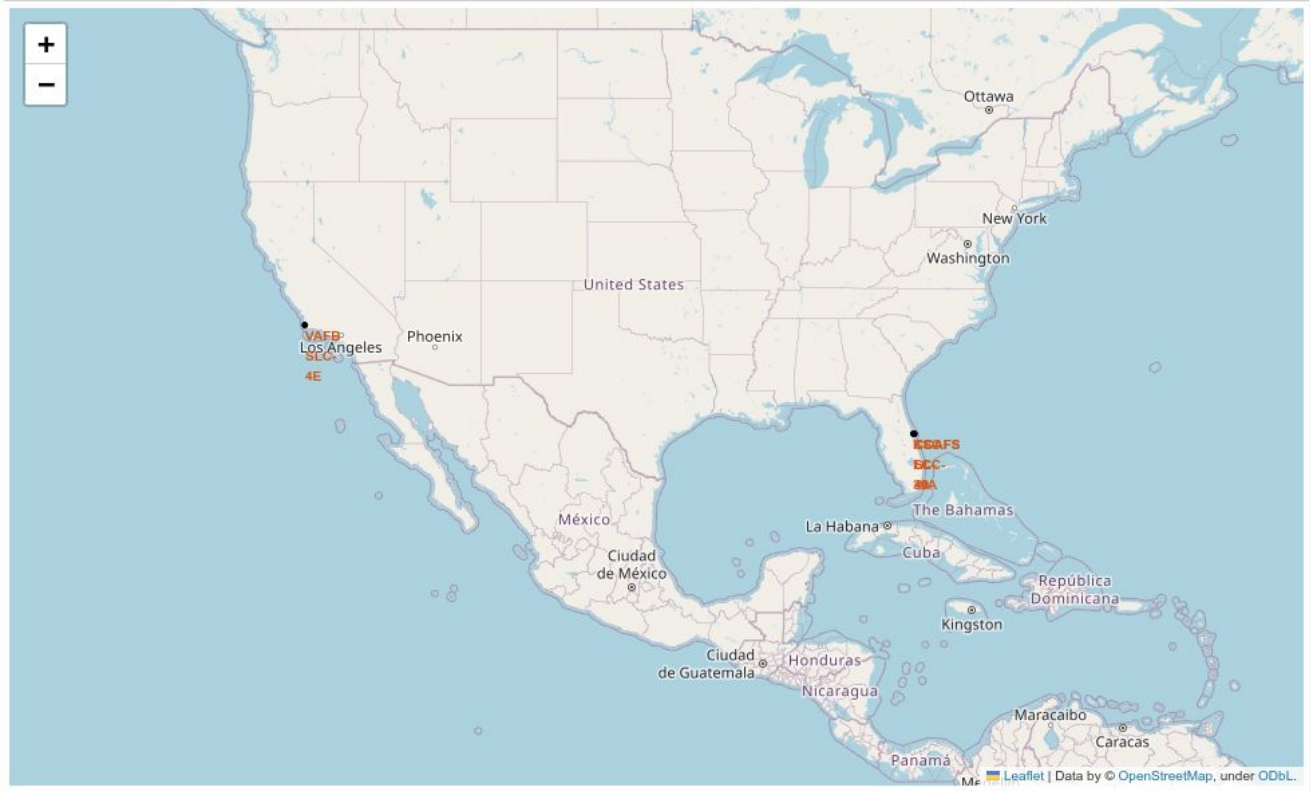
```
 * sqlite:///my_data1.db
Done.
```

Landing _Outcome	count(*)
Success (drone ship)	5
Success (ground pad)	3

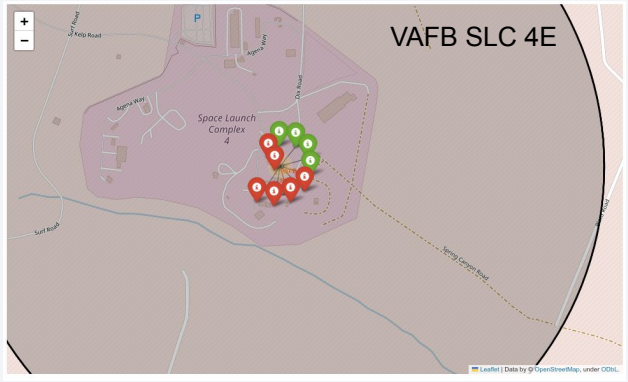
Section 3

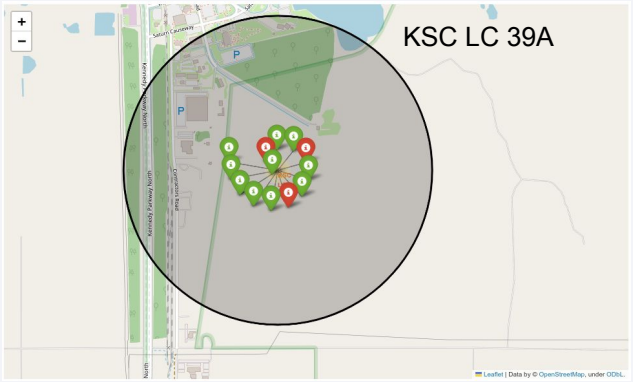
Launch Sites Proximities Analysis

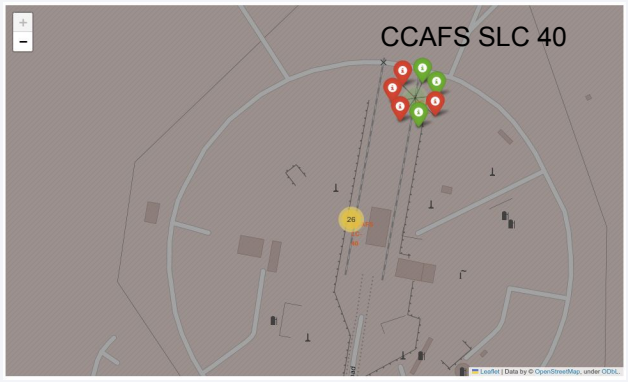
All Launch Sites

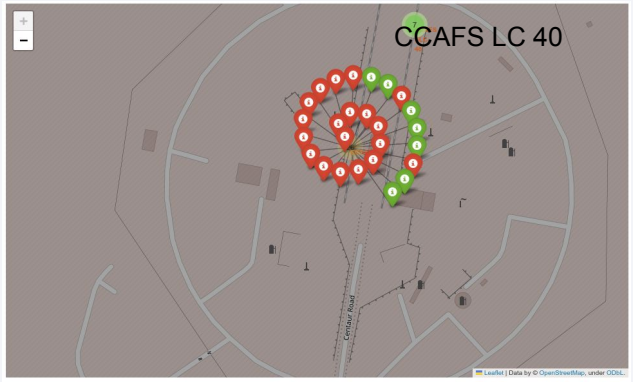


Successful / Failed Launches

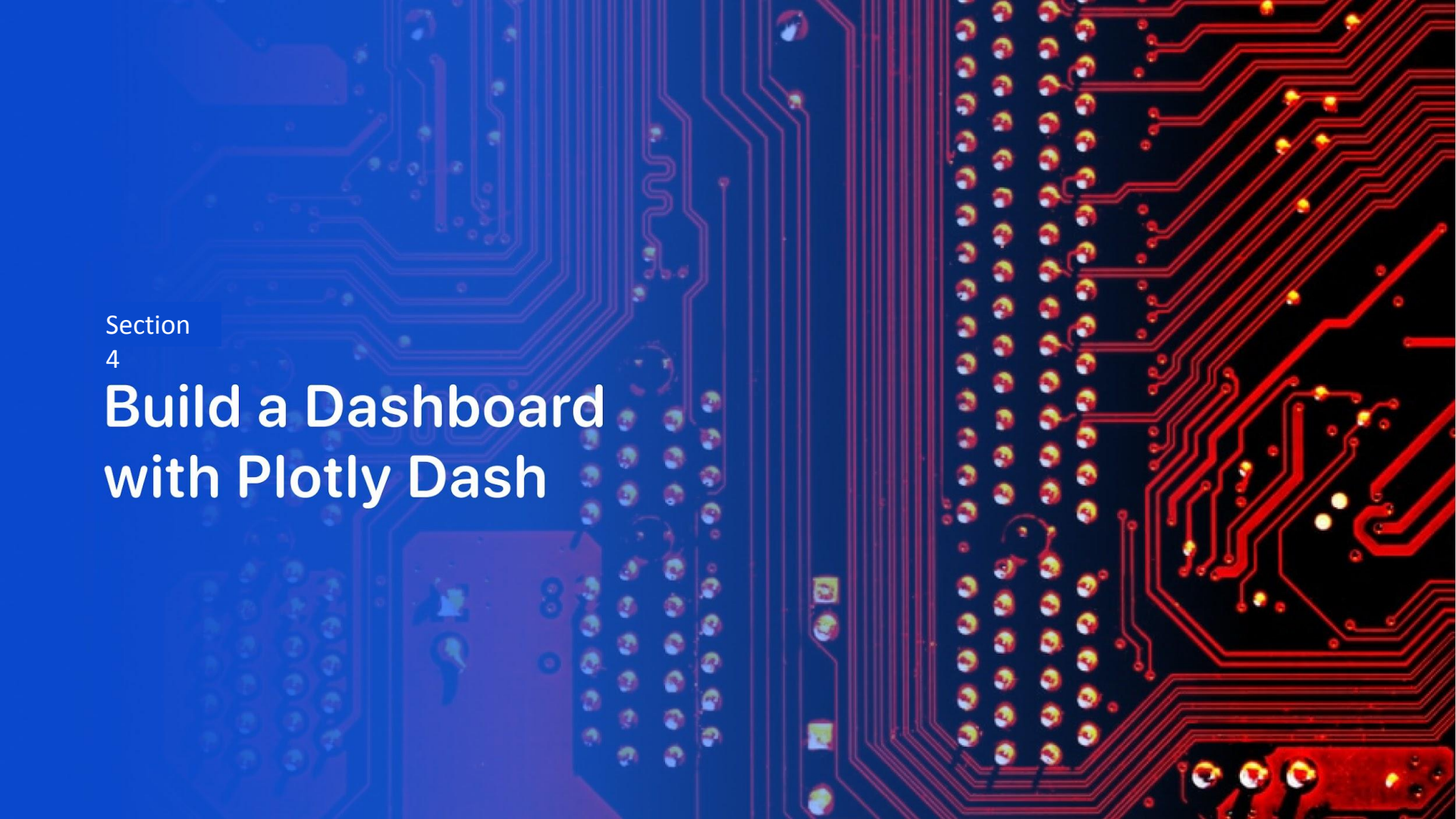




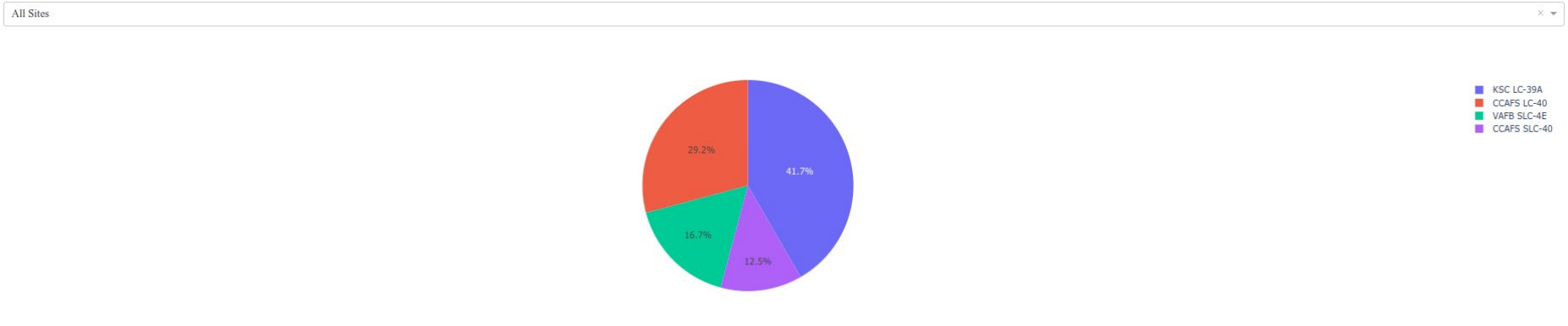




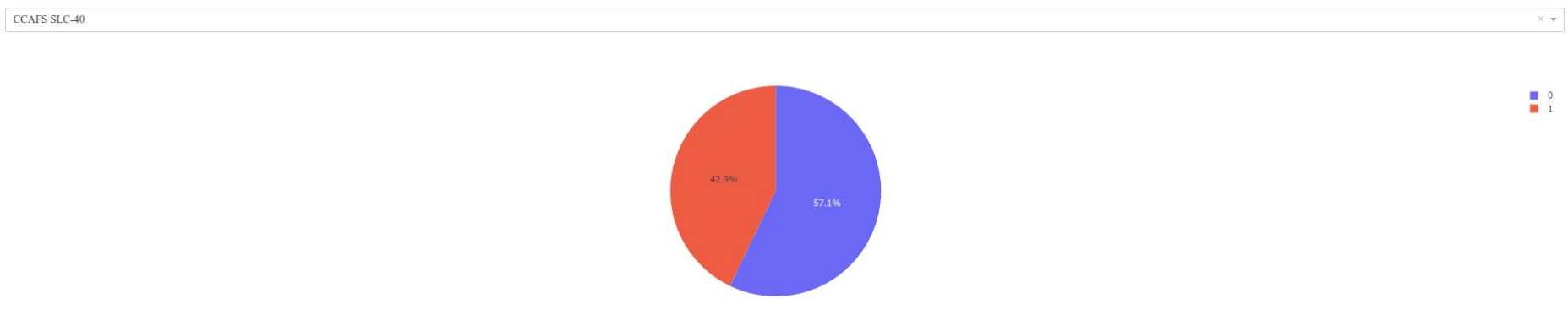
Distance to Coastline


Section
4
Build a Dashboard
with Plotly Dash

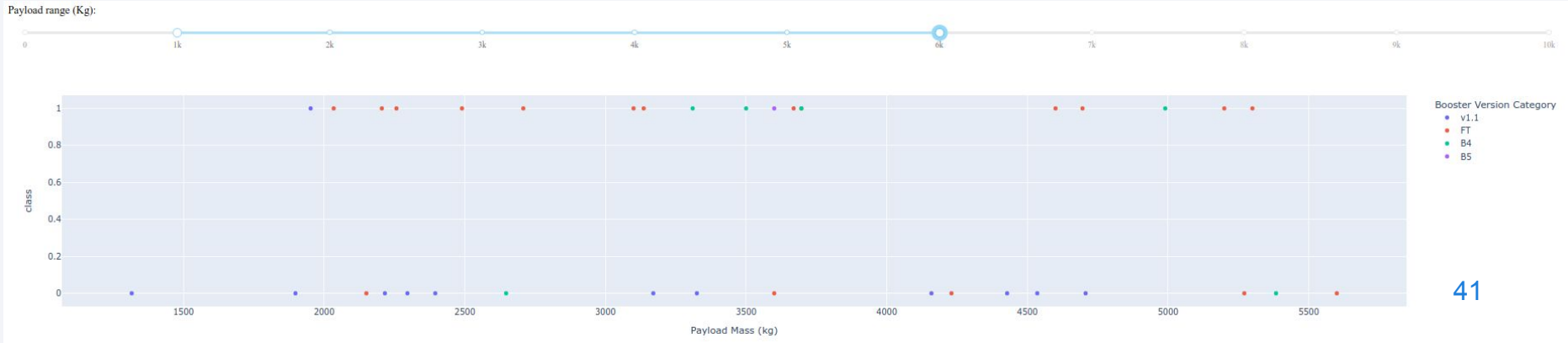
Successful Launches from All Sites



Launch Site with Highest Success Rate



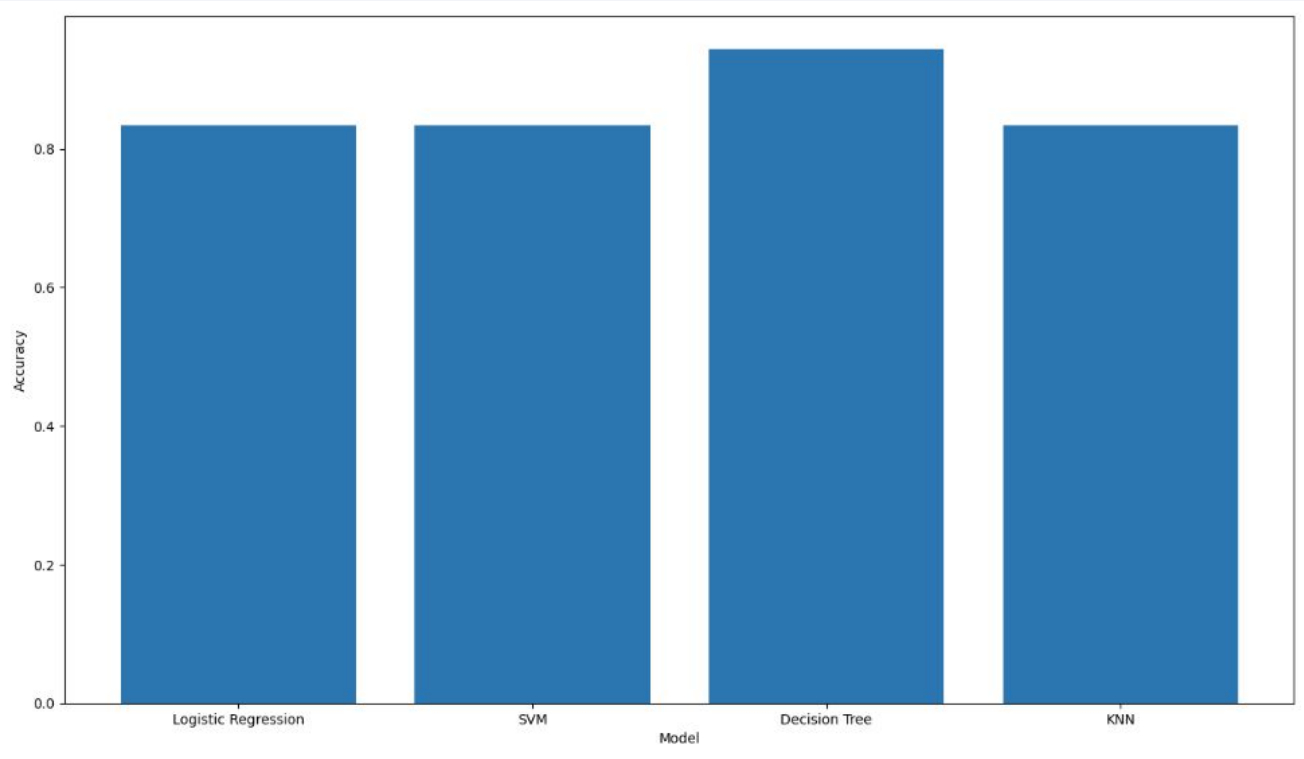
Success Rate of Different Payloads



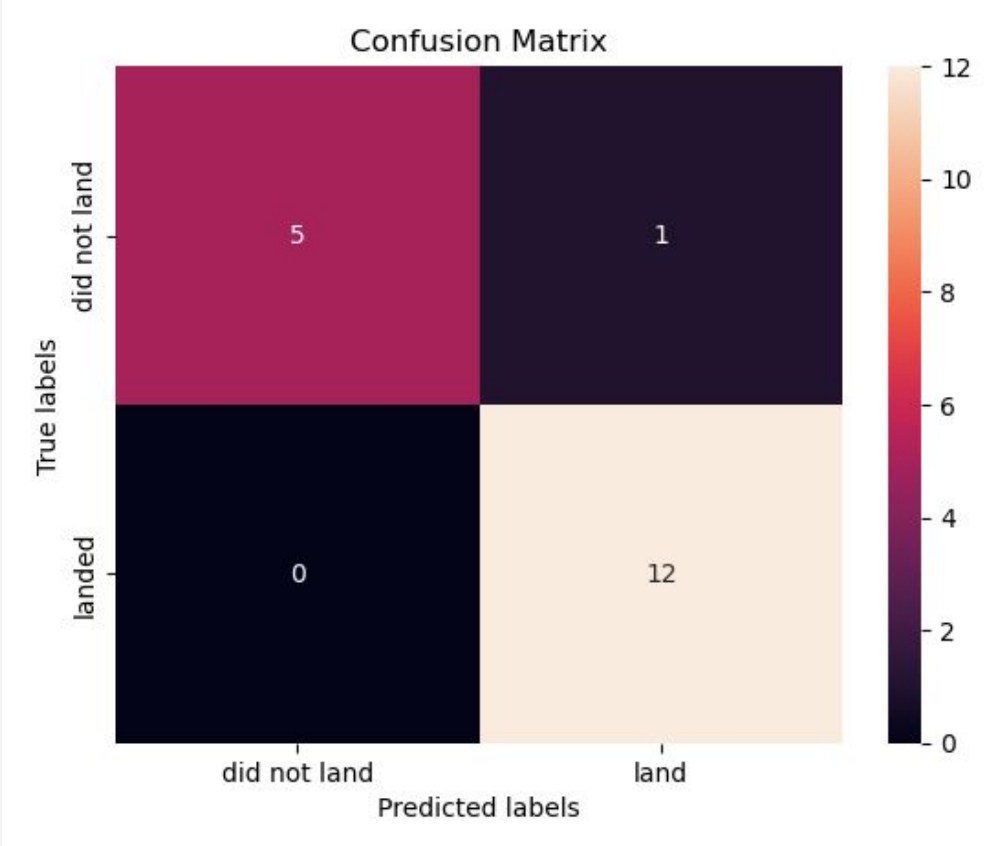
Section
5

Predictive Analysis (Classification)

Classification Accuracy



Confusion Matrix of Decision Tree Model



Conclusions

- Decision Tree models perform the best prediction
- SpaceX successful rate is improving year by year
- KSC LC 39A Launch site has the highest success rate

Appendix

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

Thank you!