IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Sanela Mehanovic
December 18, 2023

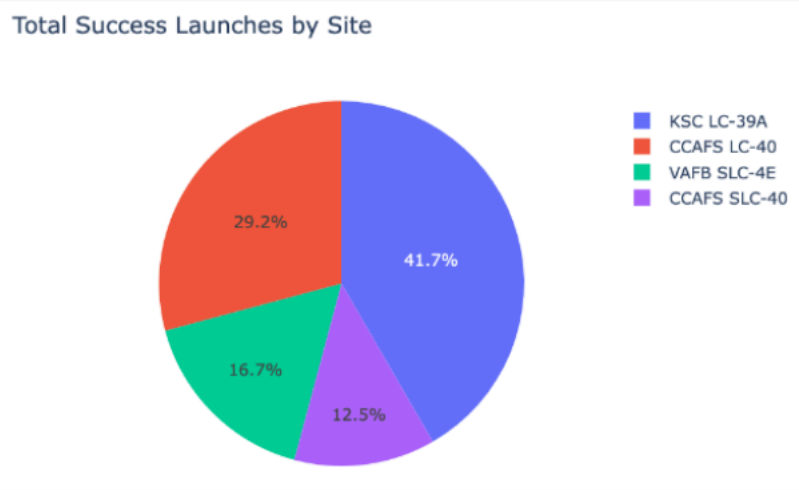
Outline

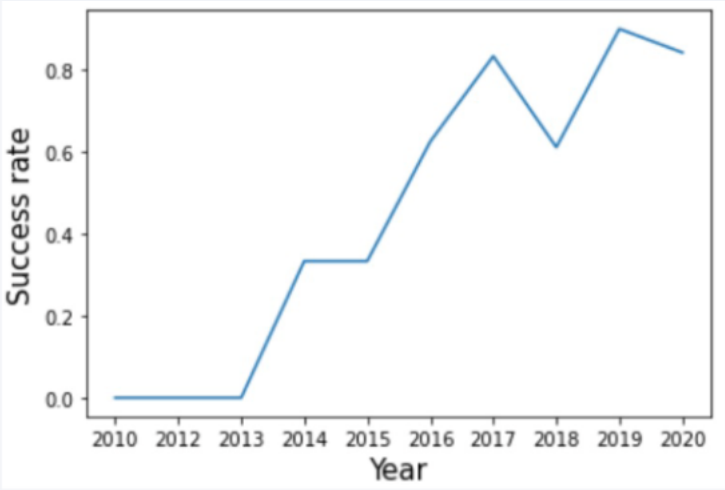
- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion
- Appendix

Executive Summary

Summary of methodologies:
- data collection API to get SpaceX data sets
- exploratory data analysis and visualisation of data
- grid search method, machine learning model to predict next landing

Summary of all results:





Introduction

- Background

For different providers, a rocket launch can cost upward of 165 million dollars each.
On the other hand, based on the reuse of the first stage, SpaceX Falcon 9 rocket is able to launch with a cost of 62 million dollars.

- Problems you want to find answers

If we can determine whether the first stage will land, we can determine the cost of a launch. Than we can use this information in order to predict if an alternate company will bid against SpaceX for a rocket launch.

- We are using the historical launch data in order to determine a new launch with success.
- We also aim to determine the best choice for a successful launch

Section 1

Methodology

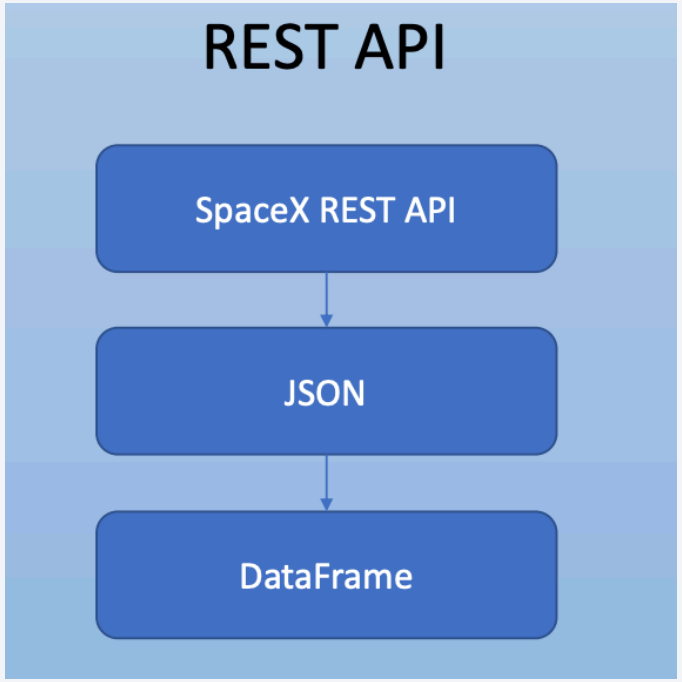
Methodology

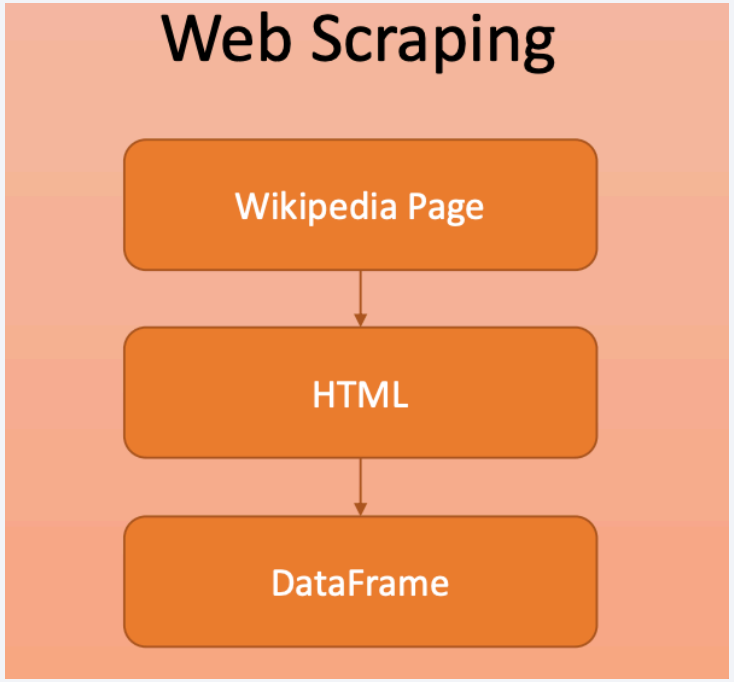
Executive Summary

- Data collection methodology:
  - SpaceX REST API
  - Web Scraping (Wikipedia)
- Perform data wrangling
  - Generate landing class from outcome column
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Use grid search to find the best model

Data Collection

- A brief description of how data sets were collected.





Data Collection – SpaceX API

- SpaceX API repository
  https://github.com/r-spacex/SpaceX-API
- Main Endpoint
  https://api.spacexdata.com/v4/launches/past
- My Notebook
  https://github.com/SanelaMehanovic/c10_applied_data-science_capstone/blob/main/1-jupyter-labs-spacex-data-collection-api.ipynb

```
import requests
import pandas as pd
spacex_url="https://api.spacexdata.com/v4/launches/
past"
response = requests.get(spacex_url)
data = pd.json_normalize(response.json())
```

	static_fire_date_utc	static_fire_date_unix	net	window	rocket	success	failures	details	crew
0	2006-03-17T00:00:00.000Z	1.142554e+09	False	0.0	5e9d0d95eda69955f709d1eb	False	[{'time': 33, 'altitude': None, 'reason': 'merlin engine failure'}]	Engine failure at 33 seconds and loss of vehicle	[]
1	None	NaN	False	0.0	5e9d0d95eda69955f709d1eb	False	[{'time': 301, 'altitude': 289, 'reason': 'harmonic oscillation leading to premature engine shutdown'}]	Successful first stage burn and transition to second stage, maximum altitude 289 km, Premature engine shutdown at T+7 min 30 s, Failed to reach orbit, Failed to recover first stage	[]
2	None	NaN	False	0.0	5e9d0d95eda69955f709d1eb	False	[{'time': 140, 'altitude': 35, 'reason': 'residual stage-1 thrust led to collision between stage 1 and stage 2'}]	Residual stage 1 thrust led to collision between stage 1 and stage 2	[]
								Ratsat was carried to orbit on the first successful	

Data Collection - Scraping

- Wikipedia Falcon Page. https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches

- My Notebook https://github.com/SanelaMehanovic/c10_applied_data-science_capstone/blob/main/2-jupyter-labs-webscraping.ipynb

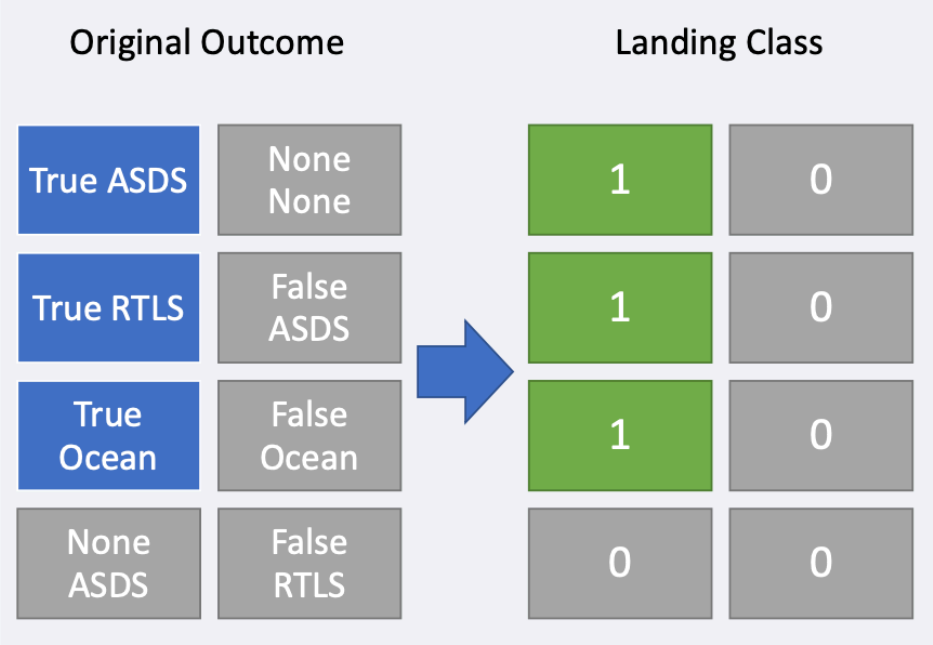
```
import requests
from bs4 import BeautifulSoup
url="https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches"
response = requests.get(url)
html_data = response.text
soup = BeautifulSoup(html_data)
```

```
<tr>
<th scope="col">Flight No.
</th>
<th scope="col">Date and<br/>time (<a href="/wiki/Coordinated_Universal_Time" title="Coordinated Universal Time">UTC</a>)
</th>
<th scope="col"><a href="/wiki/List_of_Falcon_9_first-stage_boosters" title="List of Falcon 9 first-stage boosters">Version,<br/>Booster</a> <sup class="reference" id="cite_ref-booster_11-0"><a href="#cite_note-booster-11">[b]</a></sup>
</th>
<th scope="col">Launch site
</th>
<th scope="col">Payload<sup class="reference" id="cite_ref-Dragon_12-0"><a href="#cite_note-Dragon-12">[c]</a></sup>
</th>
<th scope="col">Payload mass
</th>
<th scope="col">Orbit
</th>
<th scope="col">Customer
</th>
<th scope="col">Launch<br/>outcome
</th>
<th scope="col"><a href="/wiki/Falcon_9_first-stage_landing_tests" title="Falcon 9 first-stage landing tests">Booster<br/>landing</a>
</th></tr>
```

Next, we just need to iterate through the `<th>` elements and apply the provided `extract_column_from_header()` to extract column name one by one

Data Wrangling

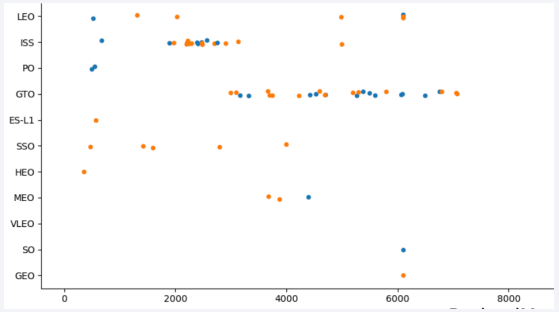
- Transform raw data to useful data, which means, convert original outcome labels into landing classification that represent our new landing prediction target in a following way:
- 1 for success
- 0 for failure.

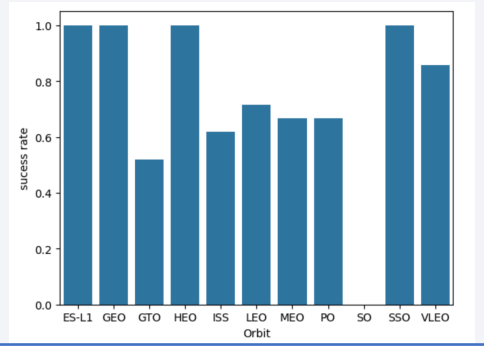


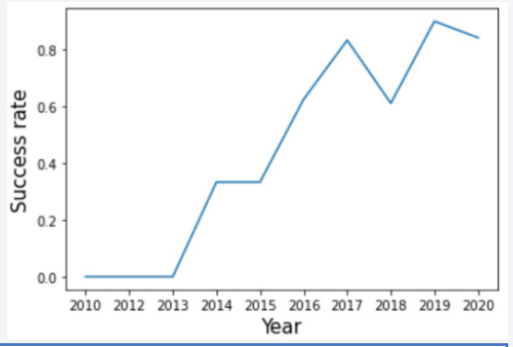
- My Notebook https://github.com/SanelaMehanovic/c10_applied_data-science_capstone/blob/main/3-jupyter-labs-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- My Visualisation Notebook https://github.com/SanelaMehanovic/c10_applied_data-science_capstone/blob/main/5-jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb







Scatter plot to show relationship between:
- FlightNumber vs Orbit type
- Payload vs. Orbit type
- FlightNumber vs. PayloadMass
- FlightNumber vs. Launch Site

Bar plot to plot success rate of each orbit

Line chart to show the yearly average launch success trend-

EDA with SQL

- Query the names of the **unique launch sites** in the space mission
- Query the names of the **booster_versions** which have carried the maximum payload mass.
- List the total number of **successful** and **failure** mission outcomes
- List the names of the boosters which have **success in drone ship** and have **payload mass** in some range
- Rank the count of successful **landing_outcomes** in date range in descending order.

- My Notebook https://github.com/SanelaMehanovic/c10_applied_data-science_capstone/blob/main/4-jupyter-labs-eda-sql-coursera_sqllite.ipynb

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

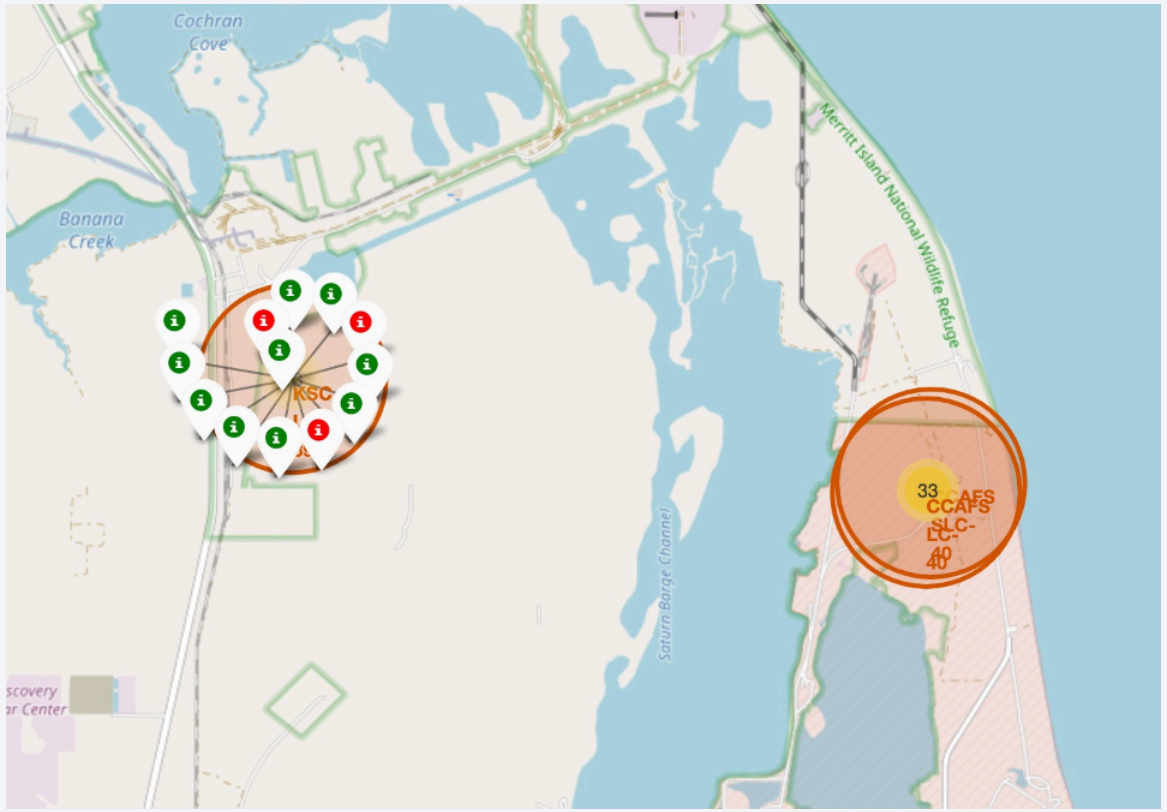
Booster_Version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

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

Build an Interactive Map with Folium

- Add **Circles** for Launch sites and **Markers** for labels
- Add **MarkerCluster** for successful and failed launches
- Add **Lines** for calculate distance between launch sites and their proximities

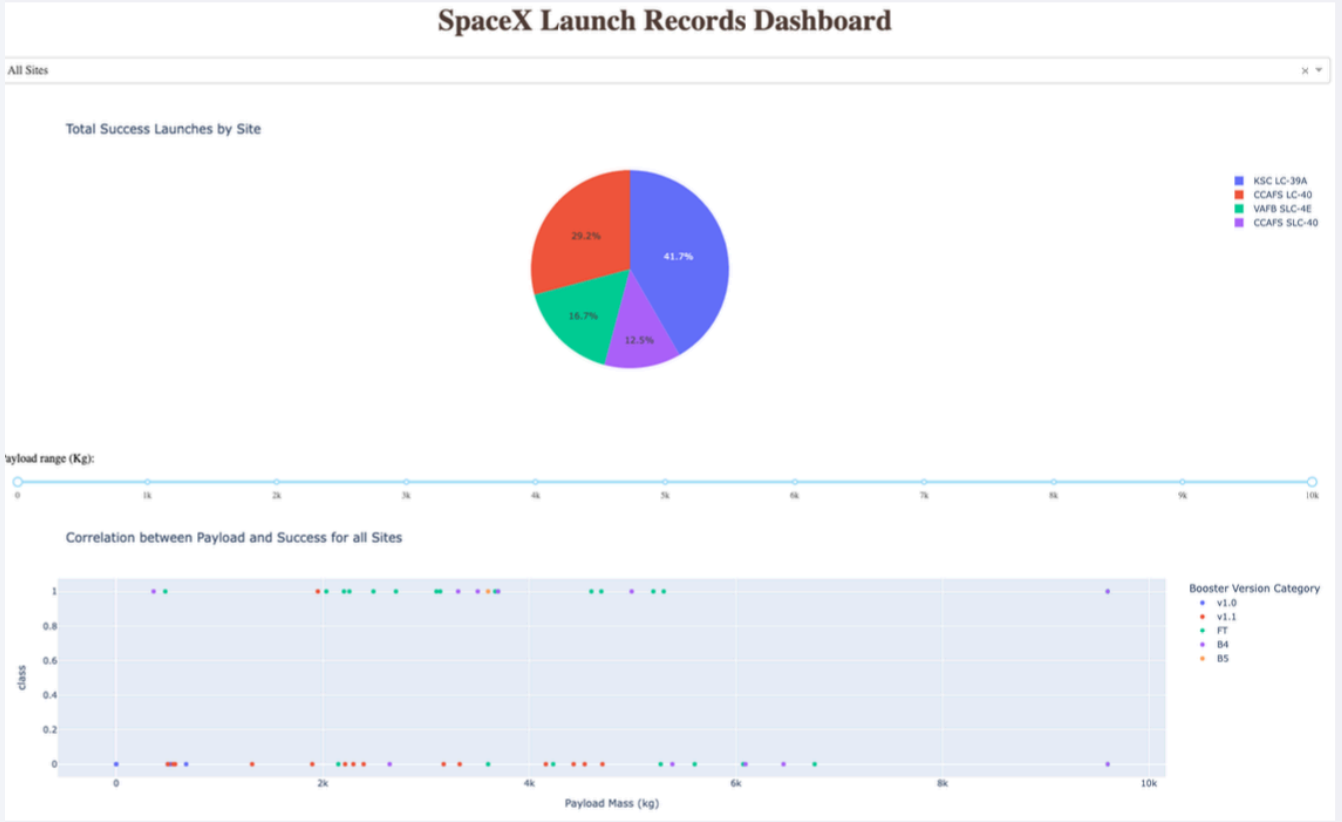
- My Notebook https://github.com/SanelaMehanovic/c10_applied_data-science_capstone/blob/main/6-jupyter-labs-data-analysis-with-folium.jupyterlite.ipynb



Build a Dashboard with Plotly Dash

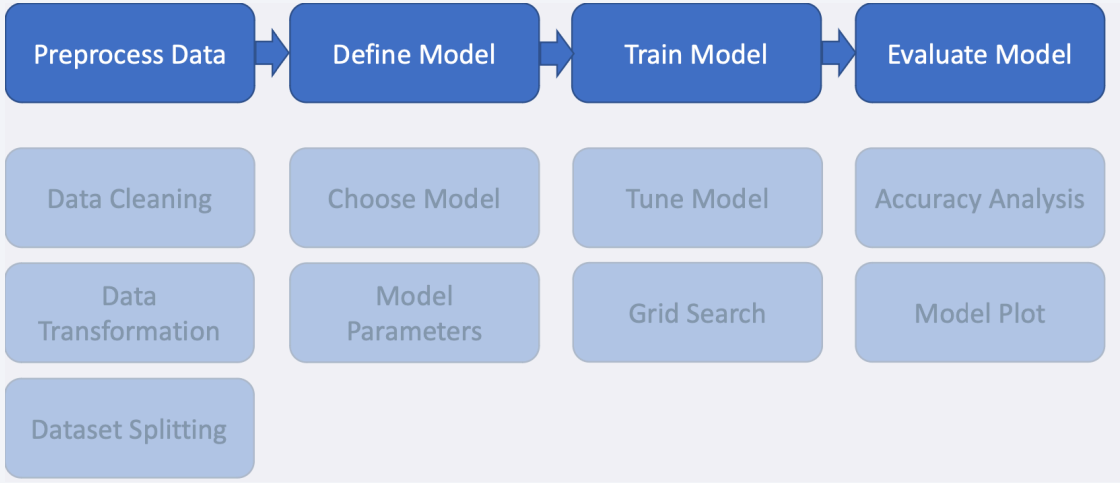
- **Dropdown menu** and a **Pie Chart** can show success launches distribution by launch site
- **Range Slider** and a **Scatter Plot** can show the correlation between Payload and Success for different launch sites

- My Plotly Dash lab https://github.com/SanelaMehanovic/c10_applied_data-science_capstone/blob/main/spacex_dash_app.py



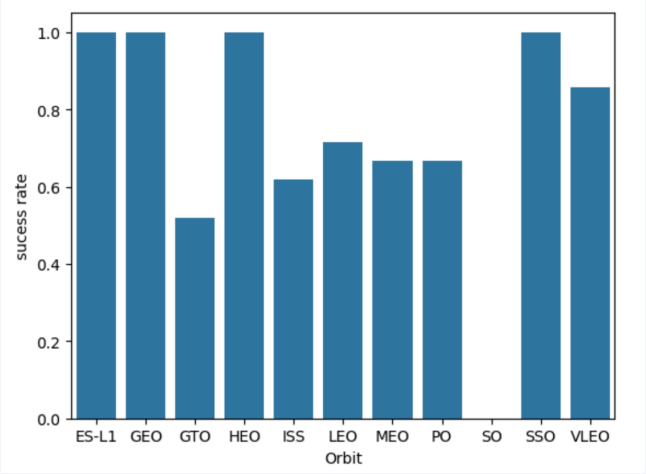
Predictive Analysis (Classification)

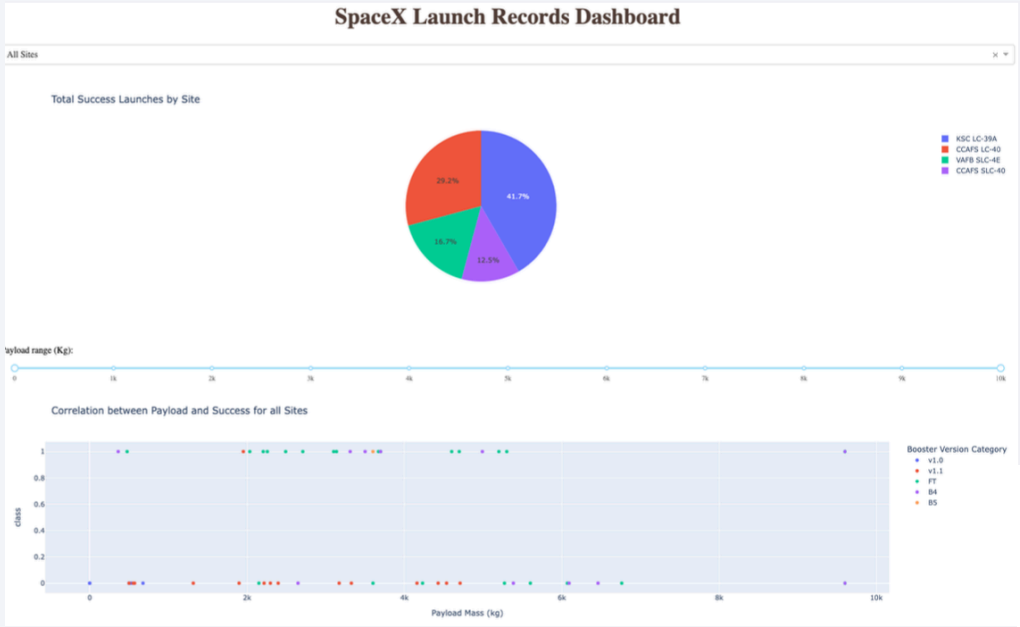
- **Prepare** data
- **Create** a column for the class
- **Standardize** the data
- **Split** into training data and test data
- **Define** model and parameters
- **Train** and Grid Search for best parameters
- **Evaluation**

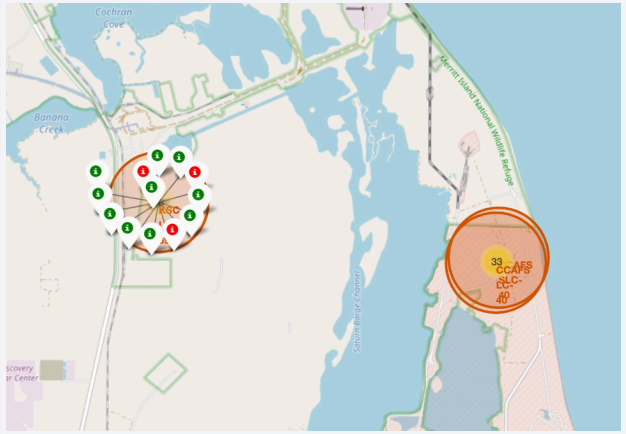


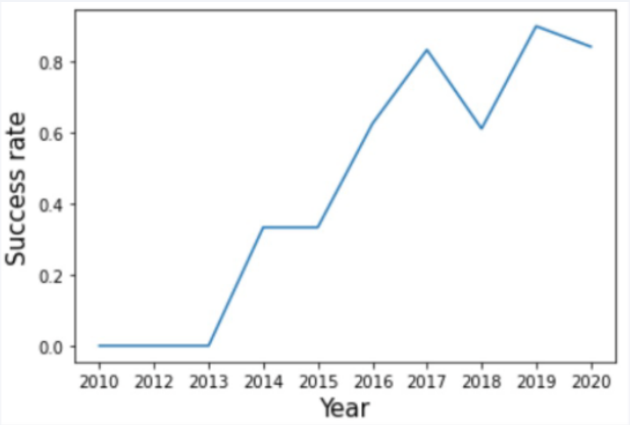
- My Predictive Analysis Notebook https://github.com/SanelaMehanovic/c10_applied_data-science_capstone/blob/main/7-jupiter-labs-machine-learning-prediction.jupyterlite.ipynb

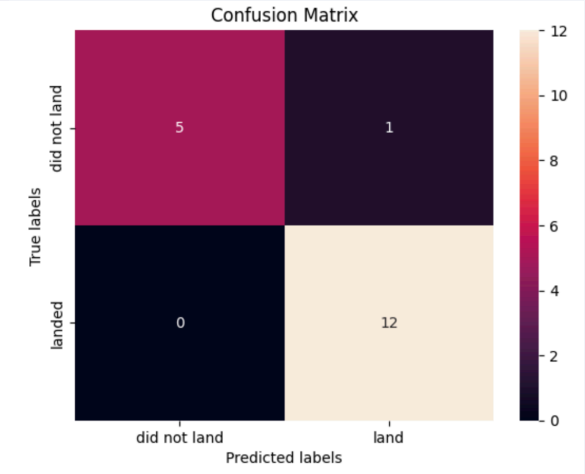
Results


sucess rate
1.0
0.8
0.6
0.4
0.2
0.0
ES-L1 GEO GTO HEO ISS LEO MEO PO SO SSO VLEO
Orbit


SpaceX Launch Records Dashboard
All Sites
Total Success Launches by Site
KSC LC-39A
CCAFS LC-40
VAFB SLC-4E
CCAFS SLC-40
41.7%
29.2%
16.7%
12.5%
Correlation between Payload and Success for all Sites
Booster Version Category
v1.0
v1.1
FT
B4
B5
class
Payload Mass (kg)


Cochran Cove
Banana Creek
KSC
CCAFS SLC-40


Success rate
0.8
0.6
0.4
0.2
0.0
2010 2012 2013 2014 2015 2016 2017 2018 2019 2020
Year

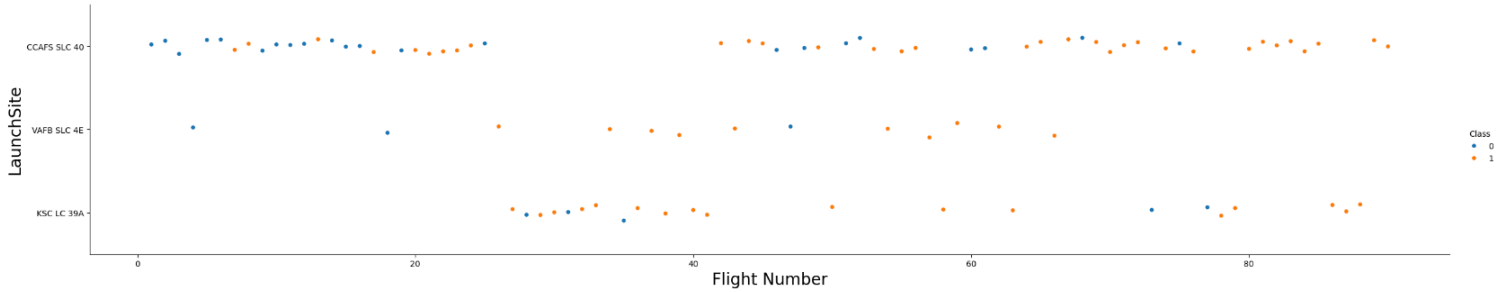

Confusion Matrix
True labels
did not land
landed
5
1
0
12
did not land
land
Predicted labels
12
10
8
6
4
2
0

Section 2

Insights Drawn from EDA

Flight Number vs. Launch Site

```
# Plot a scatter point chart with x axis to be Flight Number and y axis to be the launch site, and hue to
sns.catplot(y="LaunchSite", x="FlightNumber", hue="Class", data=df, aspect = 5)
plt.xlabel("Flight Number",fontsize=20)
plt.ylabel("LaunchSite",fontsize=20)
plt.show()
```

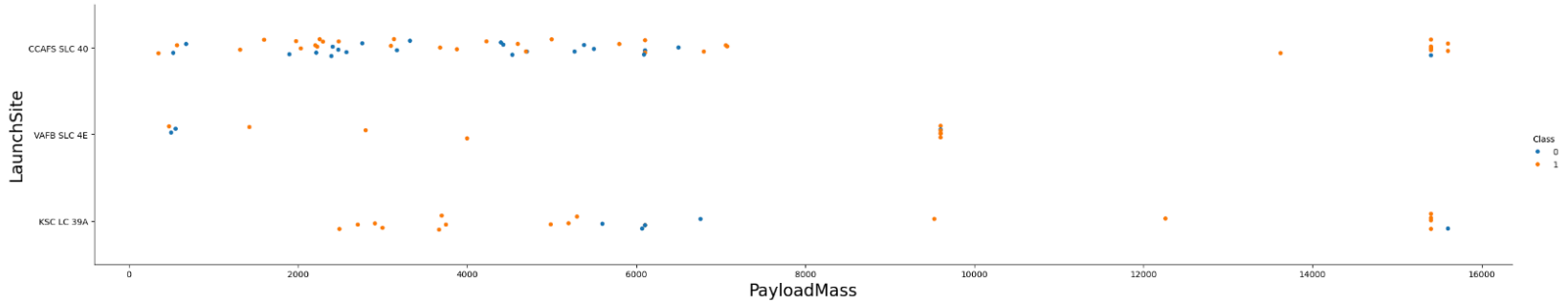


Now try to explain the patterns you found in the Flight Number vs. Launch Site scatter point plots.

```
### TASK 2: Visualize the relationship between Payload and Launch Site
The scatter plot shows a more successful first stage
landing as flight number increases. With small flight numbers, launches happens more in the site
CCAFS SLC 40 and with much lower success rate. Even if there are less launches in VAFB SLC 4E and
KSC LC39A, higher success rate can be seen in these two sites.
```

Payload vs. Launch Site

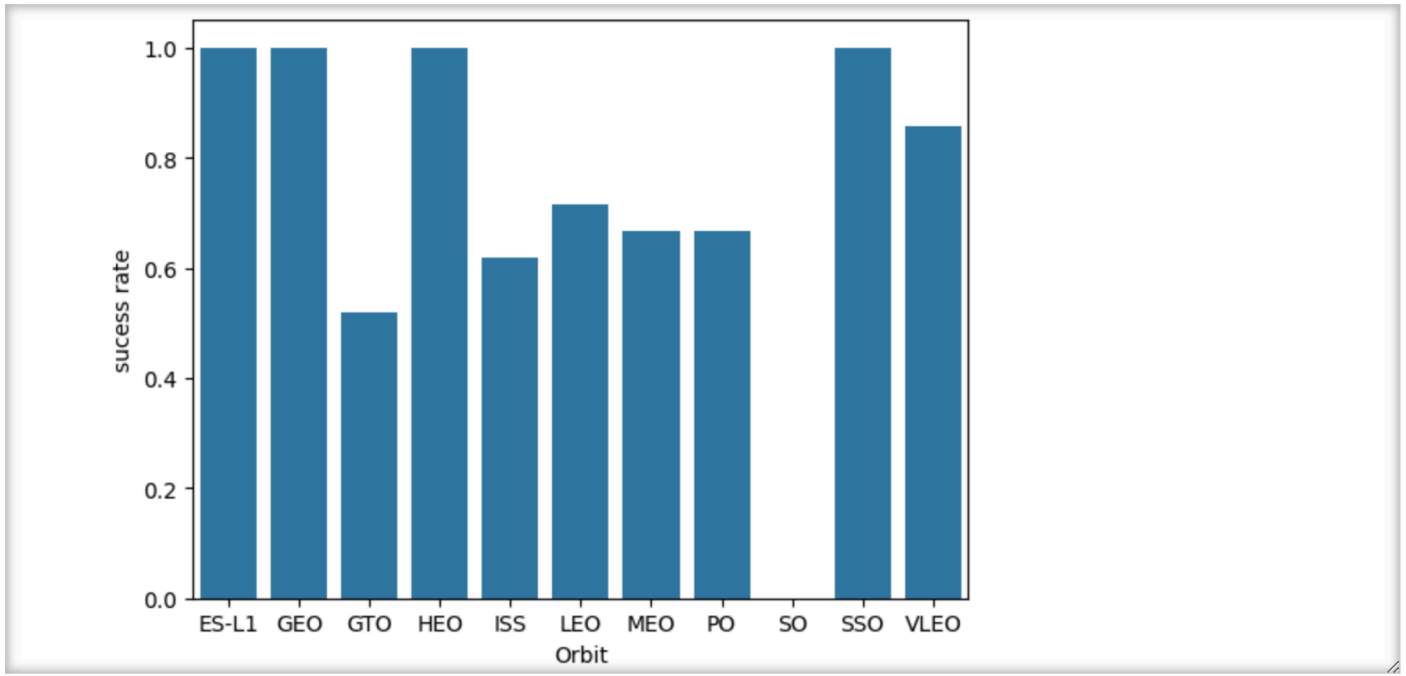
```
# Plot a scatter point chart with x axis to be Pay Load Mass (kg) and y axis to be the launch site, and hu
sns.catplot(y="LaunchSite", x="PayloadMass", hue="Class", data=df, aspect = 5)
plt.xlabel("PayloadMass",fontsize=20)
plt.ylabel("LaunchSite",fontsize=20)
plt.show()
```



Now if you observe Payload Vs. Launch Site scatter point chart you will find for the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000).

```
### TASK  3: Visualize the relationship between success rate of each orbit type
The success rate is much higher with higher Payload. In KSC LC39A launchsite we can
see much higher success rate with low Payload. This rate is much
lower in CCAFS SLC 40 launchsite. Also, there are no rockets launched in VAFB-SLC
for Payload greater than 10000. Finally, there are high success rate overall when Payload
higher than 9500.
```

Success Rate vs. Orbit Type

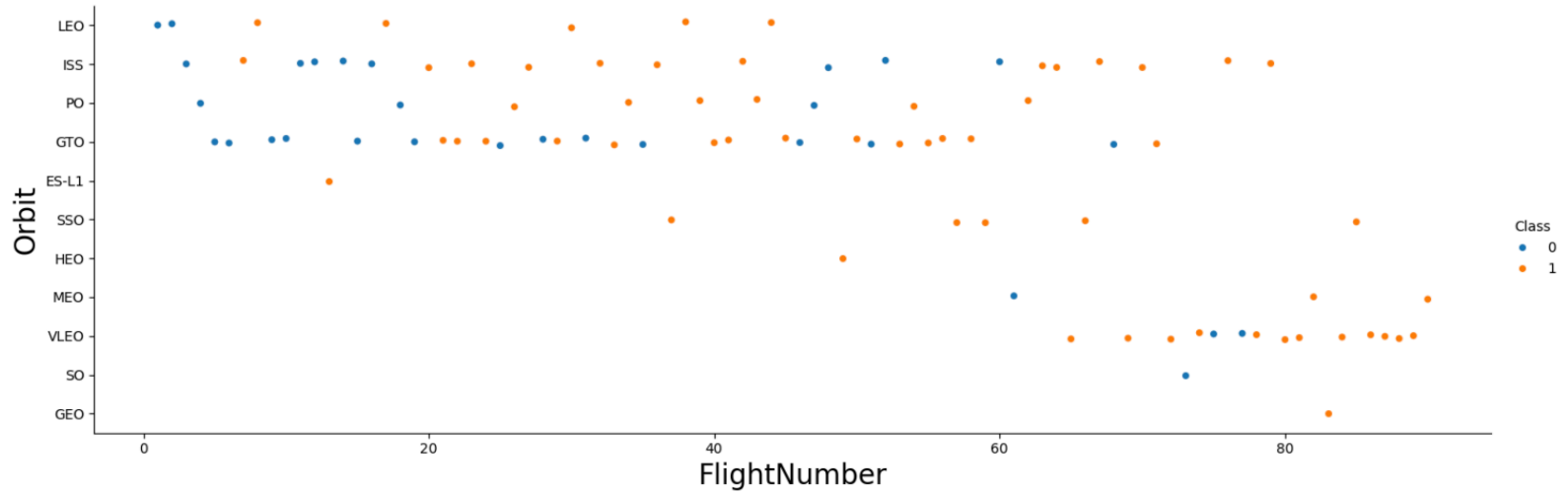



```
[ ]: Analyze the ploted bar chart try to find which orbits have high sucess rate.
```

TASK 4: Visualize the relationship between FlightNumber and Orbit type

Bar Plot shows that Orbit type ES-L1, GEO, HEO, and SSO have the highest success rate (100%). The lowest rate (zero)is for SO orbit.

Flight Number vs. Orbit Type

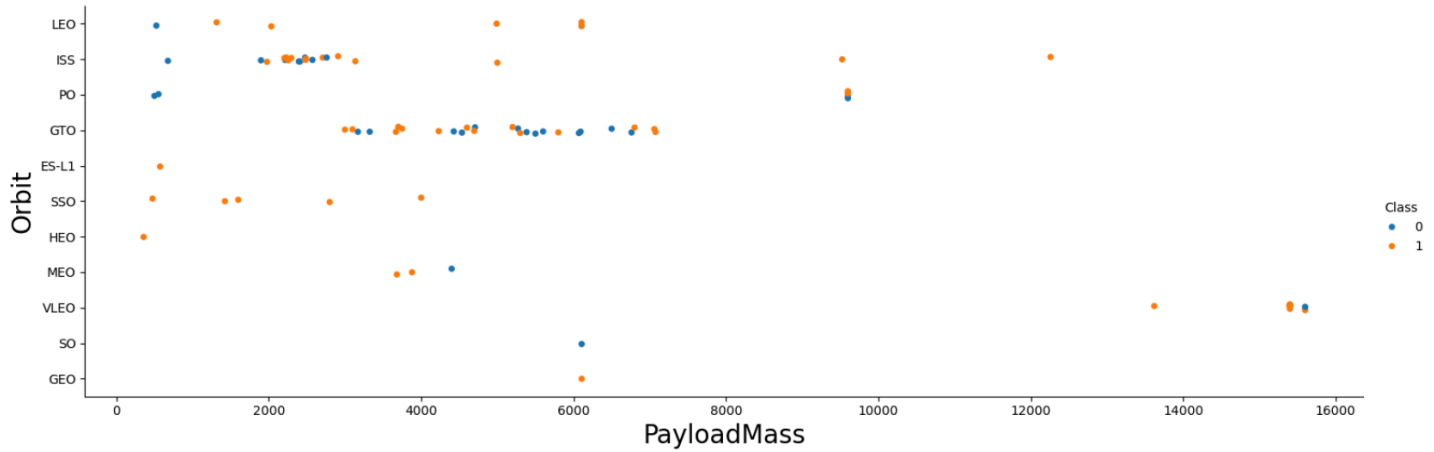


You should see that in the LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.

TASK 5: Visualize the relationship between Payload and Orbit type ¶

All launches are successful In ES-L1, GEO, HEO, and SSO orbits. There is clear relationship between flight number and success rate in LEO orbit since as flightnumber increases, the success rate increases. There is no such obvious relationship in GTO orbit.

Payload vs. Orbit Type



With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.

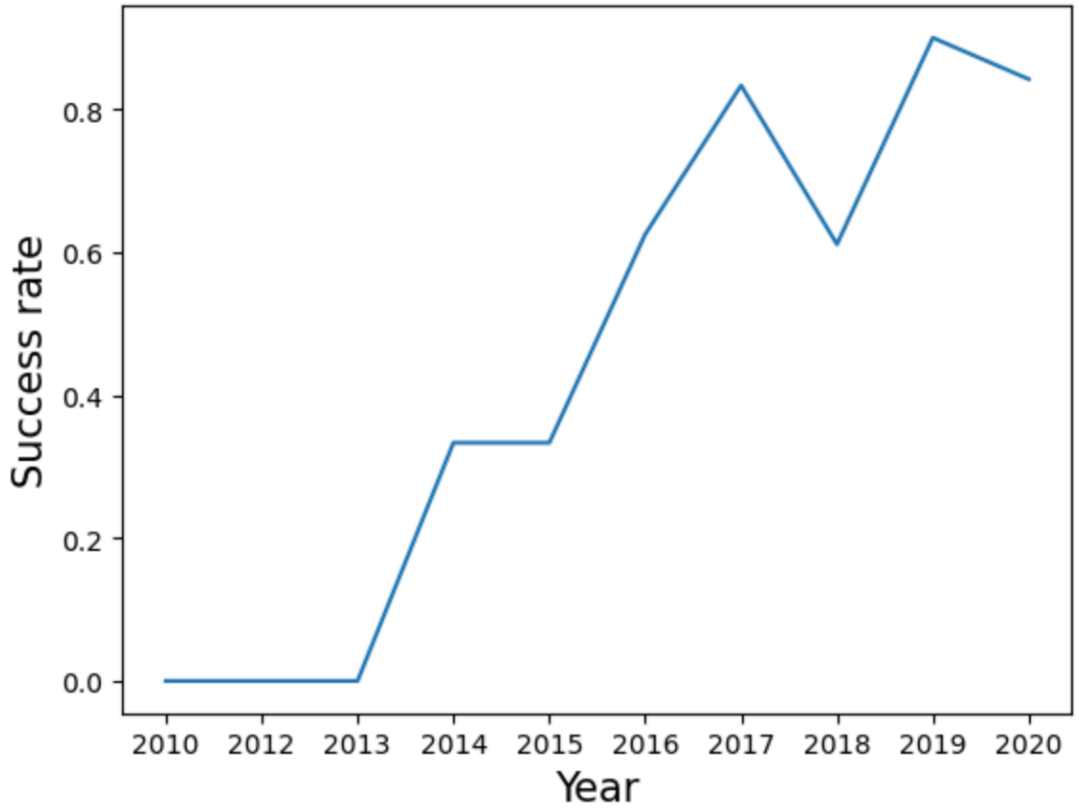
However for GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both there here.

TASK 6: Visualize the launch success yearly trend

With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.

regarding GTO, both positive landing rate and negative landing (unsuccessful mission) are there.

Launch Success Yearly Trend



you can observe that the sucess rate since 2013 kept increasing till 2020

All Launch Site Names

Four launch Sites:

- VAFB SLC-4E (western coast)
- CCAFS LC-40 (eastern coast)
- KSC LC-39A (eastern coast)
- CCAFS SLC-40 (eastern coast)

Launch Site Names Begin with 'CCA'

5 launches in LEO orbit. 4 of them NASA costumer.

```
%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

The total payload carried by boosters from NASA is 99980.

```
%sql select sum(PAYLOAD_MASS__KG_) from SPACEXTBL where Customer like 'NASA%'
```

```
 * sqlite:///my_data1.db
Done.
sum(PAYLOAD_MASS__KG_)
99980
```

Average Payload Mass by F9 v1.1

The average payload mass carried by booster version F9 v1.1 is 2534,67.

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

```
 * sqlite:///my_data1.db
Done.
avg(PAYLOAD_MASS__KG_)
2534.6666666666665
```

First Successful Ground Landing Date

The first successful landing outcome on ground pad happened on 22-12-2015.

```
%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

Names of 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" = "Success (drone ship)"
    and PAYLOAD_MASS__KG_ > 4000
    and PAYLOAD_MASS__KG_ < 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

The total number of successful mission outcomes is 100 and failure mission outcomes is 1.

```
%%sql

select count(*) from SPACEXTBL
where "Mission_Outcome" like "Success%"
```

```
 * sqlite:///my_data1.db
Done.
count(*)
100
```

```
%%sql

select count(*) from SPACEXTBL
where "Mission_Outcome" like "Failure%"
```

```
 * sqlite:///my_data1.db
Done.
count(*)
1
```

Boosters Carried Maximum Payload

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)
```

```
 * 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

Failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015:

```
%%sql select substr(Date, 6, 2) as Month, Booster_Version, Launch_Site from SPACEXTBL
where substr(Date,0,5)='2015' and "Landing_Outcome" = "Failure (drone ship)"
```

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

Month	Booster_Version	Launch_Site
01	F9 v1.1 B1012	CCAFS LC-40
04	F9 v1.1 B1015	CCAFS LC-40

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

Rank the count of landing outcomes between the date 2010-06-04 and 2017-03-20:

```
%%sql select "Landing_Outcome",
    count("Landing_Outcome") as landings
from SPACEXTBL
where Date >= "2010-06-04" and Date <= "2017-03-20"
group by "Landing_Outcome"
order by landings desc
```

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

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

Section 3

Launch Sites Proximities Analysis

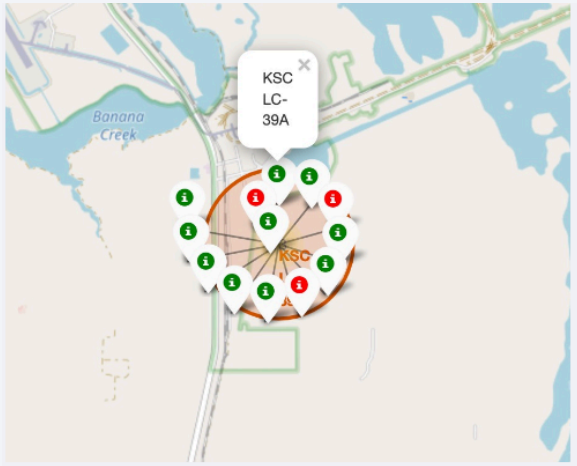
Locations of Launch Sites on Maps

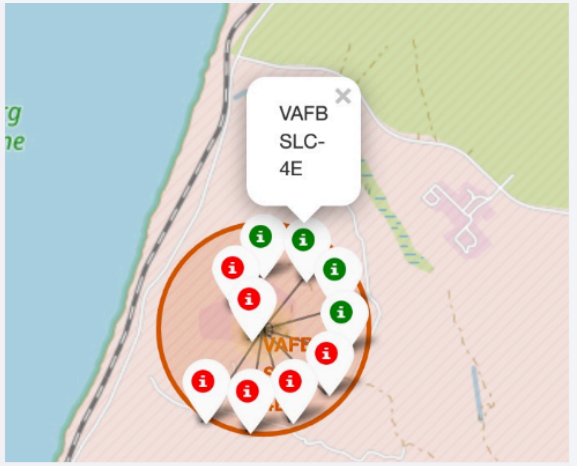
Launch Site	Lat	Long
CCAFS LC-40	28.56230197	-80.57735648
CCAFS SLC-40	28.56319718	-80.57682003
KSC LC-39A	28.57325457	-80.64689529
VAFB SLC-4E	34.63283416	-120.6107455

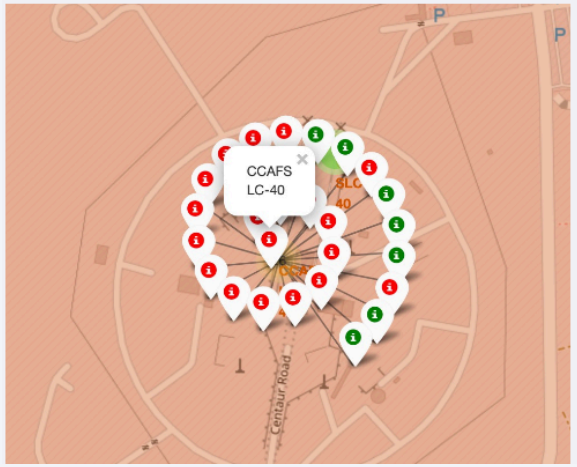
Launch Outcome Displayed by Color

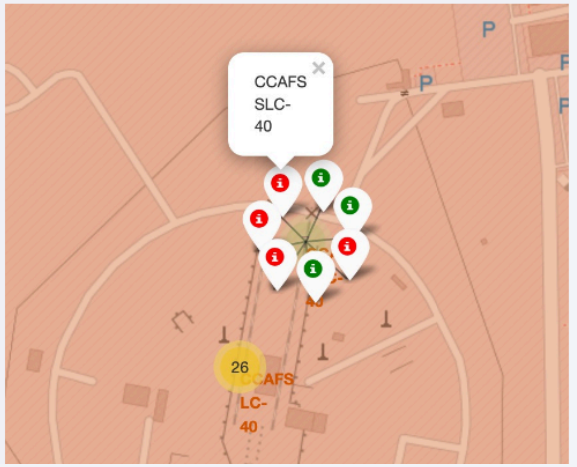
Color labels indicate that:

- KSC LC-39A has a higher success rate
- CCAFS LC-40 and CCAFS SLC-40 have much lower rate



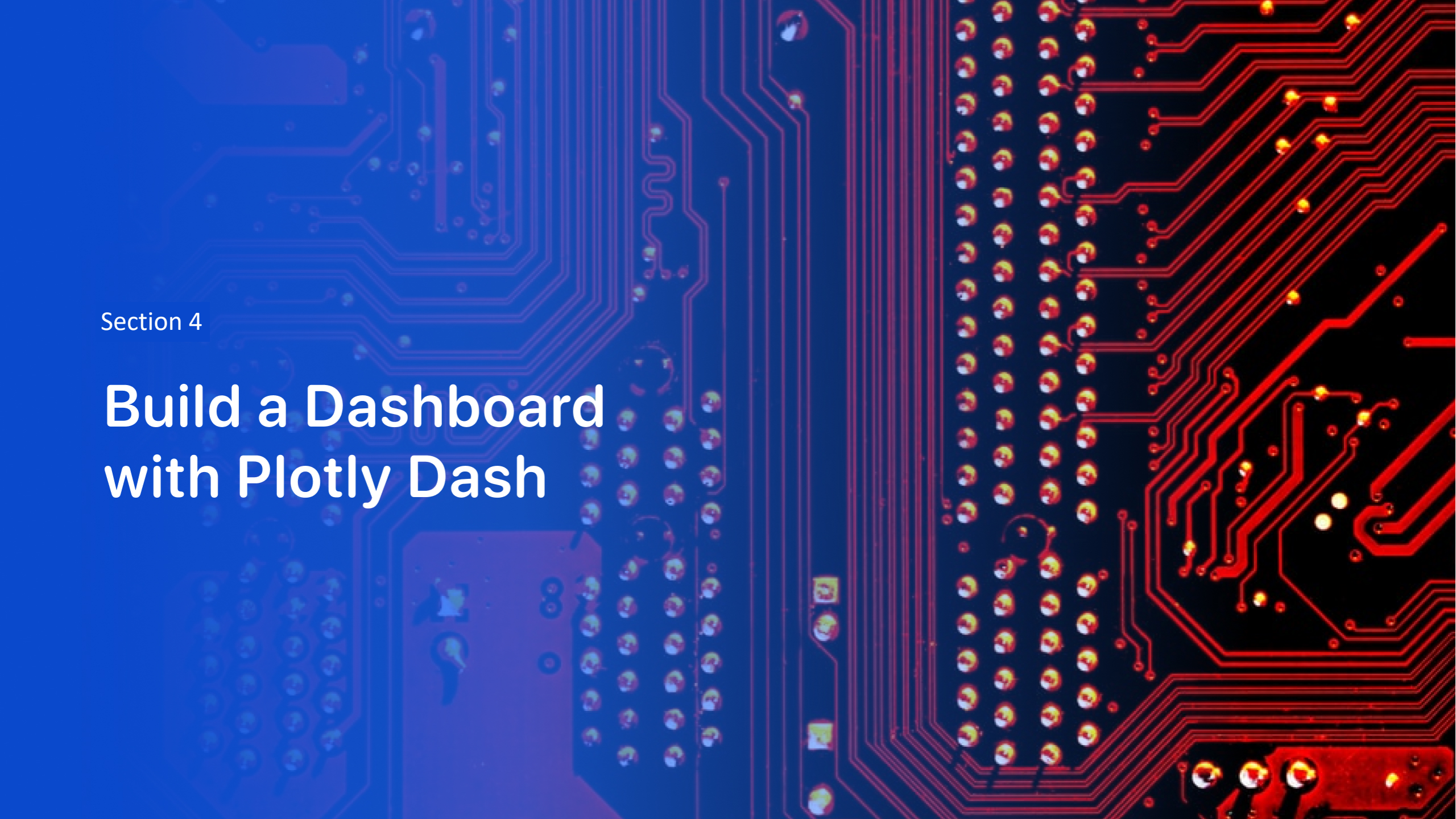






Distance to Proximities

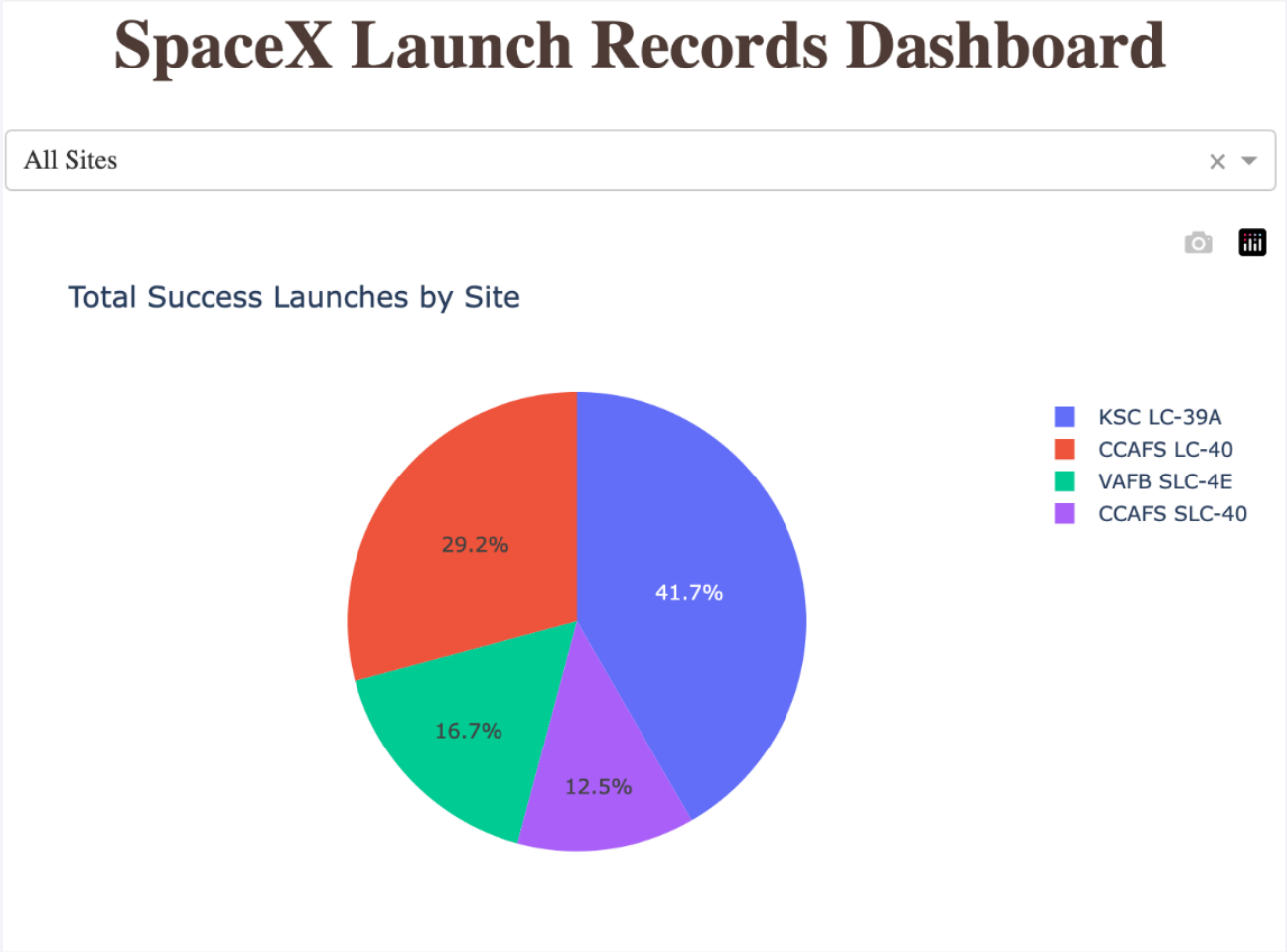
- The distance from KSC LC-39A to the coastline is approximatively 6.5km.
- The distance from KSC LC-39A to the nearest shuttle landing facility is approximatively 4.16km.
- The distance from KSC LC-39A to the nearest highway is approximatively 1km.


Section 4
Build a Dashboard with Plotly Dash

Total Success Launches for All Sites

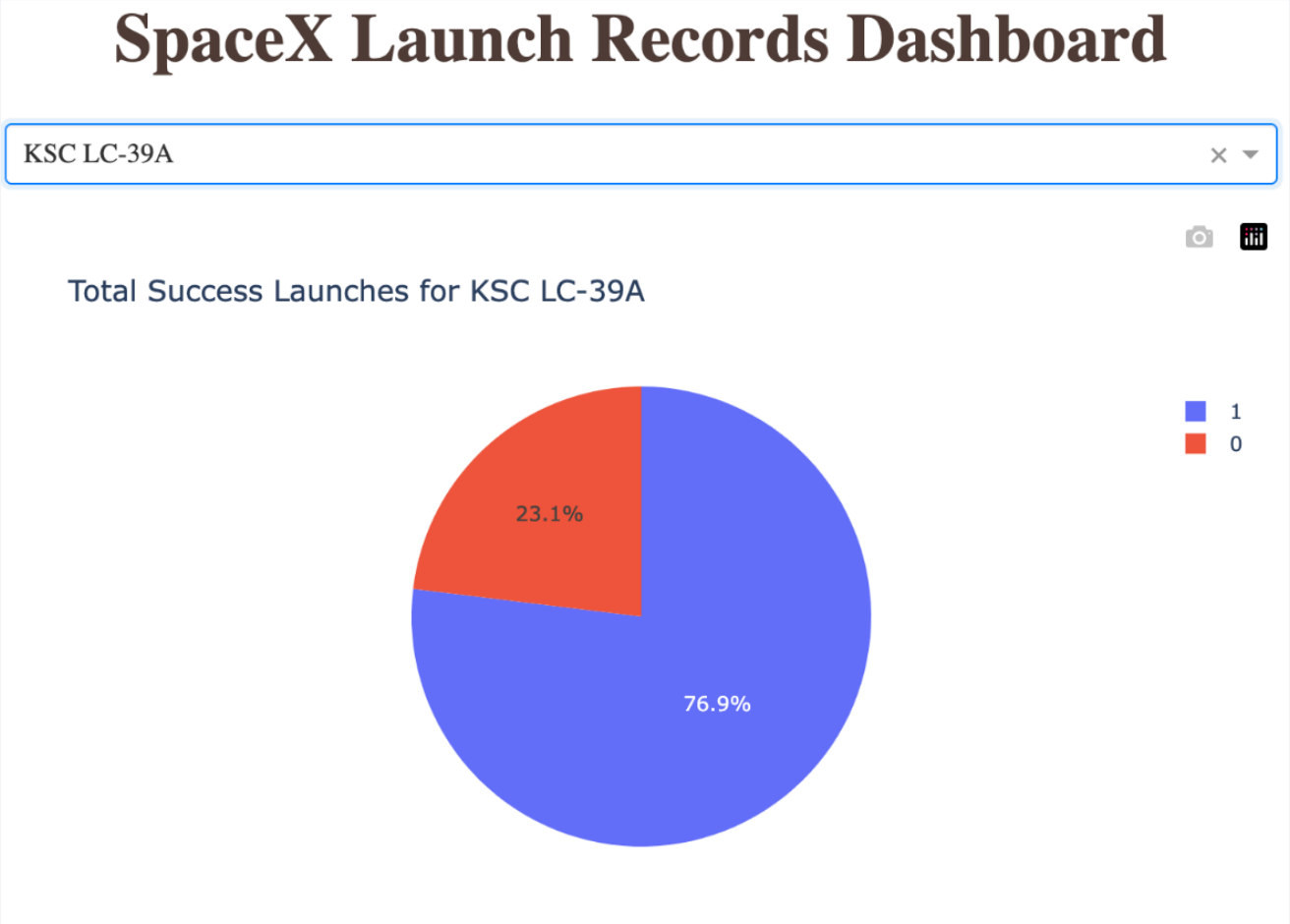
Total Success Launches by Site:

- KSC LC-39A: 41.7%
- CCAFS LC-40: 29.2%
- VAFB SLC-4E: 16.7%
- CCAFS SLC-40: 12.5%



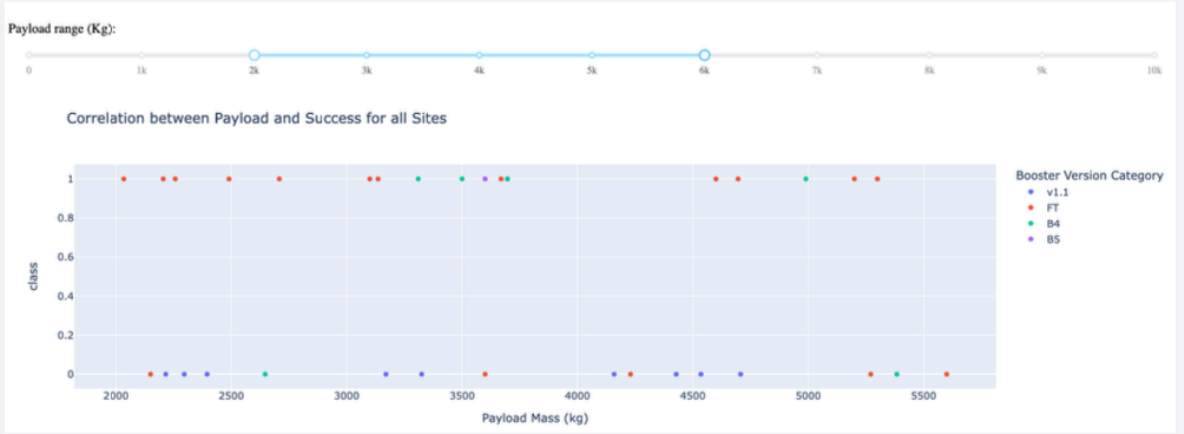
The launch site with highest launch success ratio

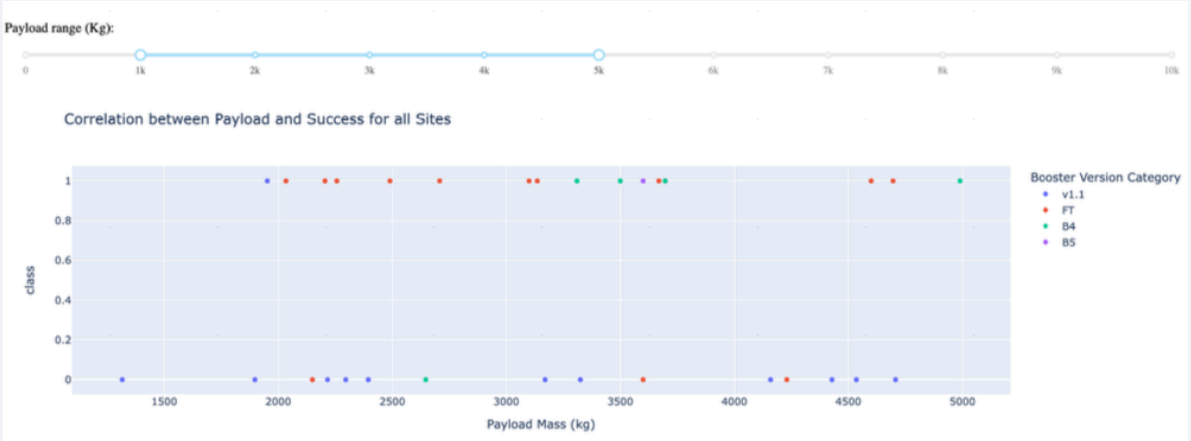
- The launch site with highest launch success ratio is **KSC LC-39A**.
- It has a success rate of **76.9%**.



Correlation Between Payload and Success

- Payload range in **[3000, 4000]** has the largest success rate.
- Booster version of **FT** has the largest success rate.



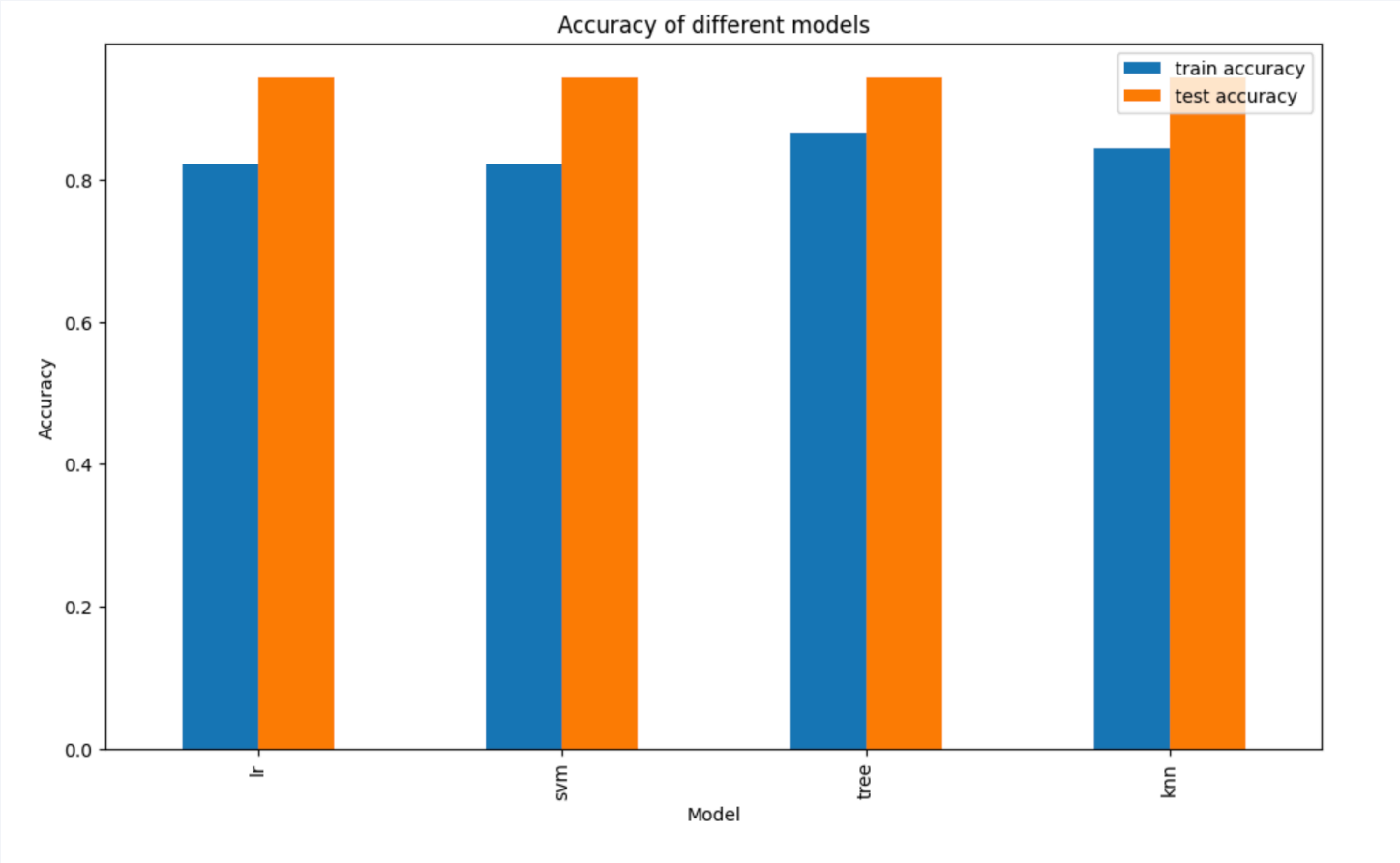


Section 5

Predictive Analysis (Classification)

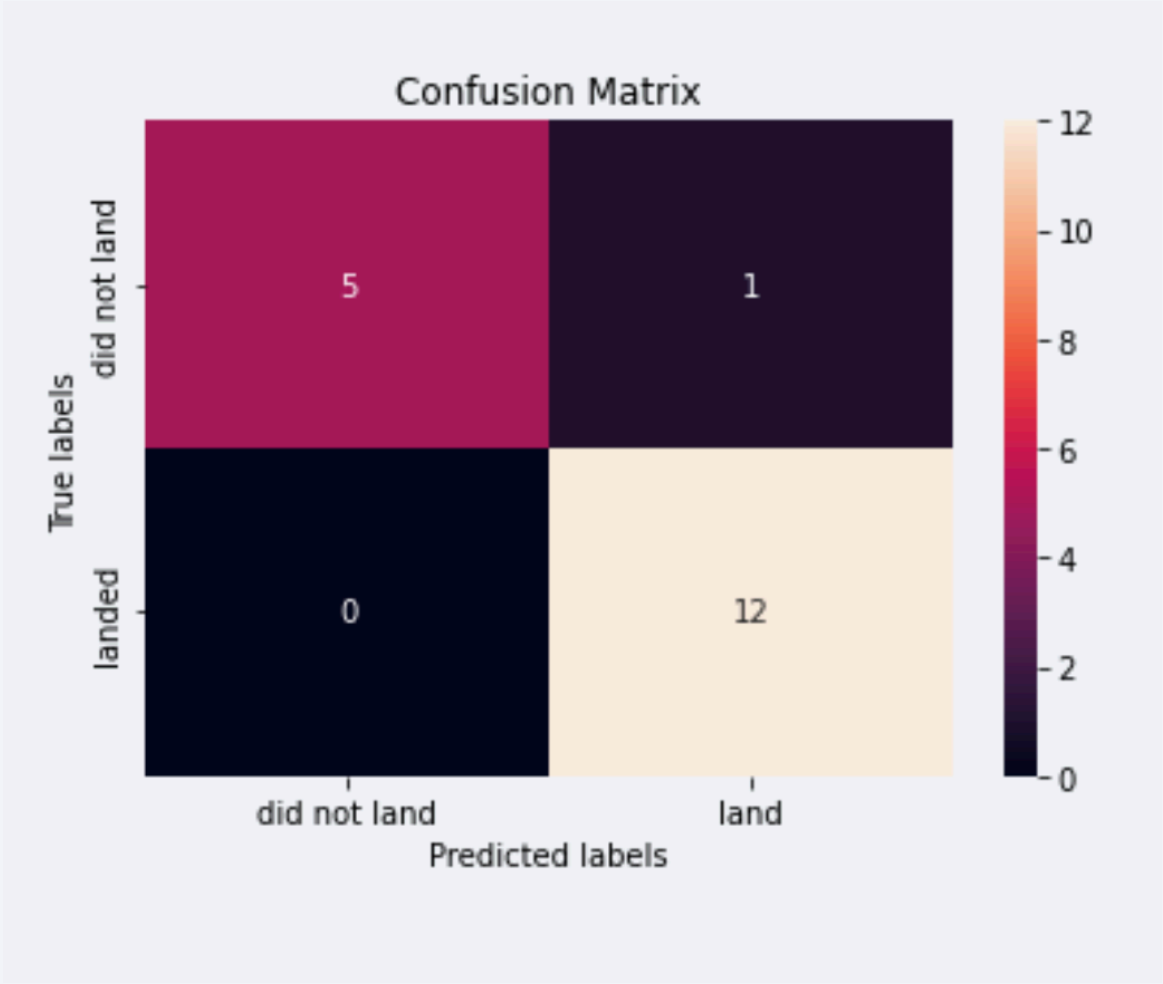
Classification Accuracy

- **Decision Tree model** has the highest classification accuracy
- Training accuracy 0.9, **testing accuracy 0.94**



Confusion Matrix

- Decision Tree model can distinguish between the different classes.
- The major problem is **false positives**.



Conclusions

- The dataset contains 90 rows 83 columns. Spliting data in 80/20 gives us 72 rows of training data and 18 rows of testing data.
- Enhancing by GridSearchCV, four models are trained which have the best performance on test data set.
- Of these models, Decision Tree is chosen as the best model for predicting landing outcome of rocket.
- By the decision tree, there might be some problems with false positives might impact the estimation of the next bid for rocket launch.

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+

Thank You!