Winning Space Race with Data Science

Taian Fonseca Feitosa
02/03/2023

Outline

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

Executive Summary

- Summary of methodologies
- Summary of all results

Introduction

- In this project we will predict if the Falcon 9 first stage will land successfully. 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. This information can be used if an alternate company wants to bid against SpaceX for a rocket launch.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
- Perform data wrangling
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

Data Collection

- To collect data were used 2 methods:
  - SpaceX API
  - Web Scrapping

Data Collection – SpaceX API

- Data collection with SpaceX REST
- [IBM-Data-Science/data-collection1-jupyter-labs-spacex-data-collection-api.ipynb at main · taianf/IBM-Data-Science (github.com)]

```
spacex_url="https://api.spacexdata.com/v4/launches/past"
```

```
response = requests.get(spacex_url)
```

Check the content of the response

```
print(response.content)
```


```
b'[{"fairings":{"reused":false,"recovery_attempt":false,"recovered":false,"ships":[]},"links":{"patch":{"small":"https://images2.imgbox.com/94/f
2/NN6Ph45r_o.png","large":"https://images2.imgbox.com/5b/02/QcxHUb5V_o.png"},"reddit":{"campaign":null,"launch":null,"media":null,"recovery":nul
l},"flickr":{"small":[],"original":[]},"presskit":null,"webcast":"https://www.youtube.com/watch?v=0a_00nJ_Y88","youtube_id":"0a_00nJ_Y88","artic
le":"https://www.space.com/2196-spacex-inaugural-falcon-1-rocket-lost-launch.html","wikipedia":"https://en.wikipedia.org/wiki/DemoSat"},"static_
fire_date_utc":"2006-03-17T00:00:00.000Z","static_fire_date_unix":1142553600,"net":false,"window":0,"rocket":"5e9d0d95eda69955f709d1eb","succes
s":false,"failures":[{"time":33,"altitude":null,"reason":"merlin engine failure"}],"details":"Engine failure at 33 seconds and loss of vehicl
e","crew":[],"ships":[],"capsules":[],"payloads":["5eb0e4b5b6c3bb0006eeb1e1"],"launchpad":"5e9e4502f5090995de566f86","flight_number":1,"name":"F
alconSat","date_utc":"2006-03-24T22:30:00.000Z","date_unix":1143239400,"date_local":"2006-03-25T10:30:00+12:00","date_precision":"hour","upcomin
g":false,"cores":[{"core":"5e9e289df35918033d3b2623","flight":1,"gridfins":false,"legs":false,"reused":false,"landing_attempt":false,"landing_su
ccess":null,"landing_type":null,"landpad":null}],"auto_update":true,"tbd":false,"launch_library_id":null,"id":"5eb87cd9ffd86e000604b32a"},{"fair
ings":{"reused":false,"recovery_attempt":false,"recovered":false,"ships":[]},"links":{"patch":{"small":"https://images2.imgbox.com/f9/4a/ZboXReN
b_o.png","large":"https://images2.imgbox.com/80/a2/bkWotCIS_o.png"},"reddit":{"campaign":null,"launch":null,"media":null,"recovery":null},"flick
r":{"small":[],"original":[]},"presskit":null,"webcast":"https://www.youtube.com/watch?v=Lk4zQ2wP-Nc","youtube_id":"Lk4zQ2wP-Nc","article":"http
s://www.space.com/3590-spacex-falcon-1-rocket-fails-reach-orbit.html","wikipedia":"https://en.wikipedia.org/wiki/DemoSat"},"static_fire_date_ut
c":null,"static_fire_date_unix":null,"net":false,"window":0,"rocket":"5e9d0d95eda69955f709d1eb","success":false,"failures":[{"time":301,"altitud
e":289,"reason":"harmonic oscillation leading to premature engine shutdown"}],"details":"Successful first stage burn and transition to second st
age, maximum altitude 289 km, Premature engine shutdown at T+7 min 30 s, Failed to reach orbit, Failed to recover first stage","crew":[],"ship
s":[],"capsules":[],"payloads":["5eb0e4b6b6c3bb0006eeb1e2"],"launchpad":"5e9e4502f5090995de566f86","flight_number":2,"name":"DemoSat","date_ut
c":"2007-03-21T01:10:00.000Z","date_unix":1174439400,"date_local":"2007-03-21T13:10:00+12:00","date_precision":"hour","upcoming":false,"cores":
[{"core":"5e9e289ef35918416a3b2624","flight":1,"gridfins":false,"legs":false,"reused":false,"landing_attempt":false,"landing_success":null,"land
ing_type":null,"landpad":null}],"auto_update":true,"tbd":false,"launch_library_id":null,"id":"5eb87cdaffd86e000604b32b"},{"fairings":{"reused":f
alse,"recovery_attempt":false,"recovered":false,"ships":[]},"links":{"patch":{"small":"https://images2.imgbox.com/6c/cb/na1tzhHs_o.png","larg
```

Data Collection - Scraping

- Web scraping process
- [IBM-Data-Science/data-collection2-jupyter-labs-webscraping.ipynb at main · taianf/IBM-Data-Science (github.com)]

First, let's perform an HTTP GET method to request the Falcon9 Launch HTML page, as an HTTP response.

```
# use requests.get() method with the provided static_url
# assign the response to a object
html_data = requests.get(static_url).text
```

Create a `BeautifulSoup` object from the HTML `response`

```
# Use BeautifulSoup() to create a BeautifulSoup object from a response text content
soup = BeautifulSoup(html_data, "html.parser")
```

Print the page title to verify if the `BeautifulSoup` object was created properly

```
# Use soup.title attribute
soup.title
```

```
<title>List of Falcon 9 and Falcon Heavy launches - Wikipedia</title>
```

Data Wrangling

- Using pandas library, we checked some data properties
- [IBM-Data-Science/data-wrangling-labs-jupyter-spacex-Data wrangling.ipynb at main · taianf/IBM-Data-Science (github.com)]

```
df=pd.read_csv("https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/dataset_part_1.csv")
df.head(10)
```

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitu
0	1	2010-06-04	Falcon 9	6104.959412	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0003	-80.577366	28.561
1	2	2012-05-22	Falcon 9	525.000000	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0005	-80.577366	28.561
2	3	2013-03-01	Falcon 9	677.000000	ISS	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0007	-80.577366	28.561
3	4	2013-09-29	Falcon 9	500.000000	PO	VAFB SLC 4E	False Ocean	1	False	False	False	NaN	1.0	0	B1003	-120.610829	34.632
4	5	2013-12-03	Falcon 9	3170.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B1004	-80.577366	28.561
5	6	2014-01-06	Falcon 9	3325.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B1005	-80.577366	28.561
6	7	2014-04-18	Falcon 9	2296.000000	ISS	CCAFS SLC 40	True Ocean	1	False	False	True	NaN	1.0	0	B1006	-80.577366	28.561
7	8	2014-07-14	Falcon 9	1316.000000	LEO	CCAFS SLC 40	True Ocean	1	False	False	True	NaN	1.0	0	B1007	-80.577366	28.561
8	9	2014-08-05	Falcon 9	4535.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B1008	-80.577366	28.561
9	10	2014-09-07	Falcon 9	4428.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B1011	-80.577366	28.561

EDA with Data Visualization

- To analyze the relation between variables, some scatterplots were used.
- A bar graph was used also to analyze the success rate.
- IBM-Data-Science/jupyter-labs-eda-dataviz.ipynb at main · taianf/IBM-Data-Science (github.com)

EDA with SQL

- Some queries run:
  - SELECT distinct launch_site FROM spacex
  - SELECT * FROM spacex where launch_site like 'CCA%' limit 5
  - SELECT sum(payload_mass_kg_) FROM spacex where customer like '%NASA%'
  - SELECT avg(payload_mass_kg_) FROM spacex where Booster_Version like '%F9 v1.1%'
  - SELECT min(Date) FROM spacex where Landing_Outcome = 'Success'
- IBM-Data-Science/eda-sql-jupyter-labs-eda-sql-coursera.ipynb at main · taianf/IBM-Data-Science (github.com)

Build an Interactive Map with Folium

- Markers were used to Launch Sites
- Circles were used to highlight areas
- Lines were used to show distances
- IBM-Data-Science/lab_jupyter_launch_site_location.ipynb at main · taianf/IBM-Data-Science (github.com)

Build a Dashboard with Plotly Dash

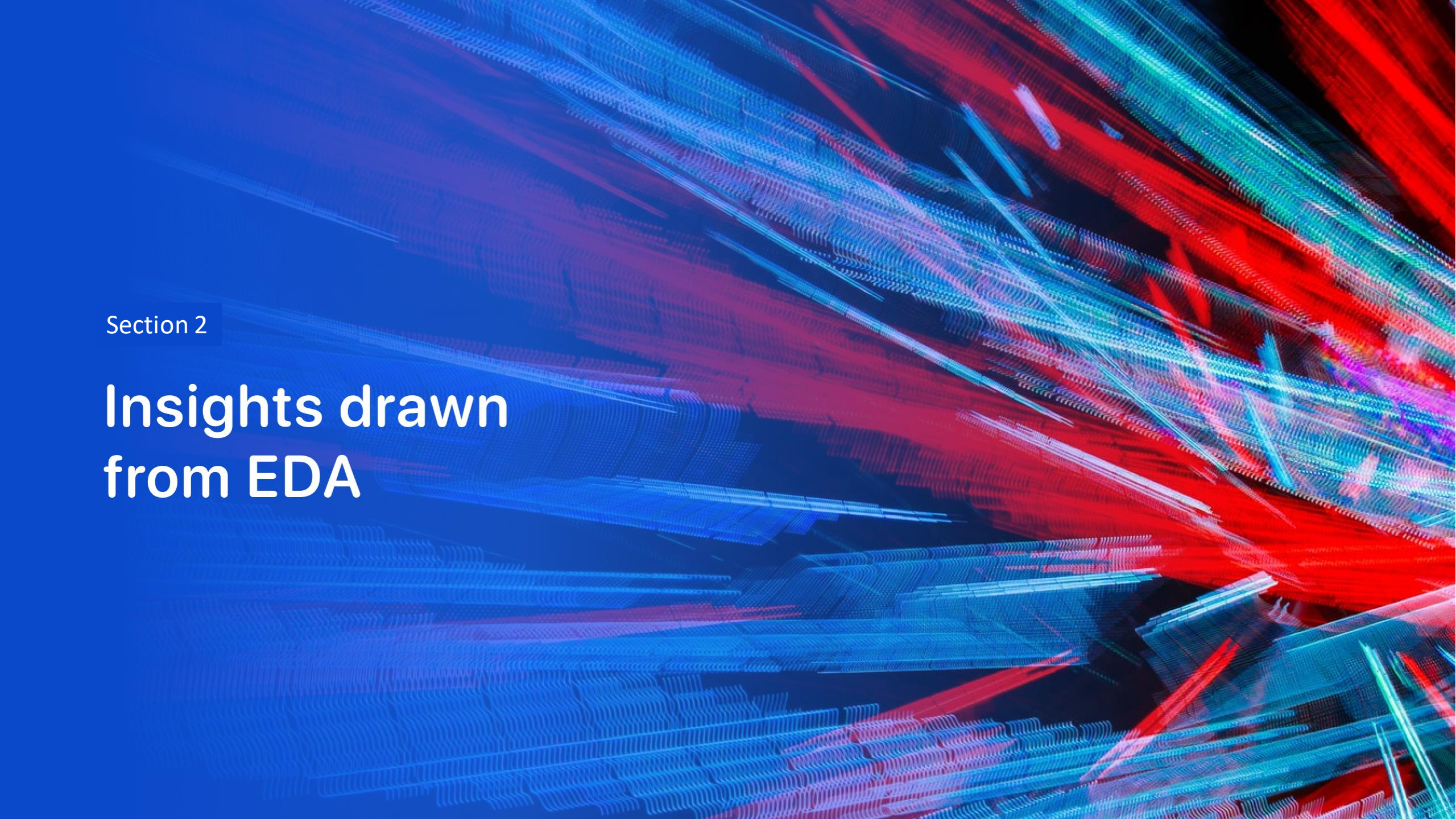
- Pie charts were used to show and compare success rate
- Scatter plots were used also
- IBM-Data-Science/spacex_dash_app.py at main · taianf/IBM-Data-Science (github.com)

Predictive Analysis (Classification)

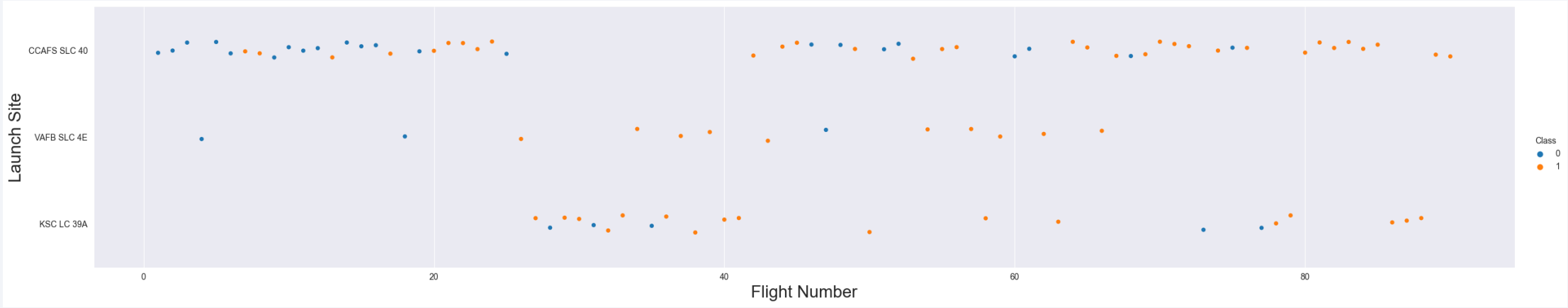
- The variables (x) were standardized
- The dataset were split in train and test
- Algorithms used to predict the success of launches:
  - Logistic Regression
  - Support Vector Machine
  - Tree
  - KNN
- [IBM-Data-Science/SpaceX_Machine Learning Prediction_Part_5.ipynb at main · taianf/IBM-Data-Science (github.com)]

Section 2

Insights drawn from EDA

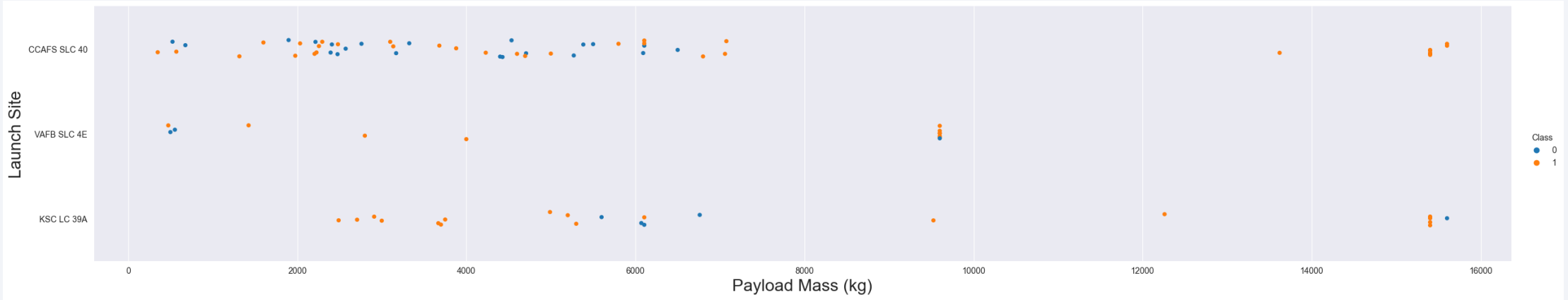


Flight Number vs. Launch Site



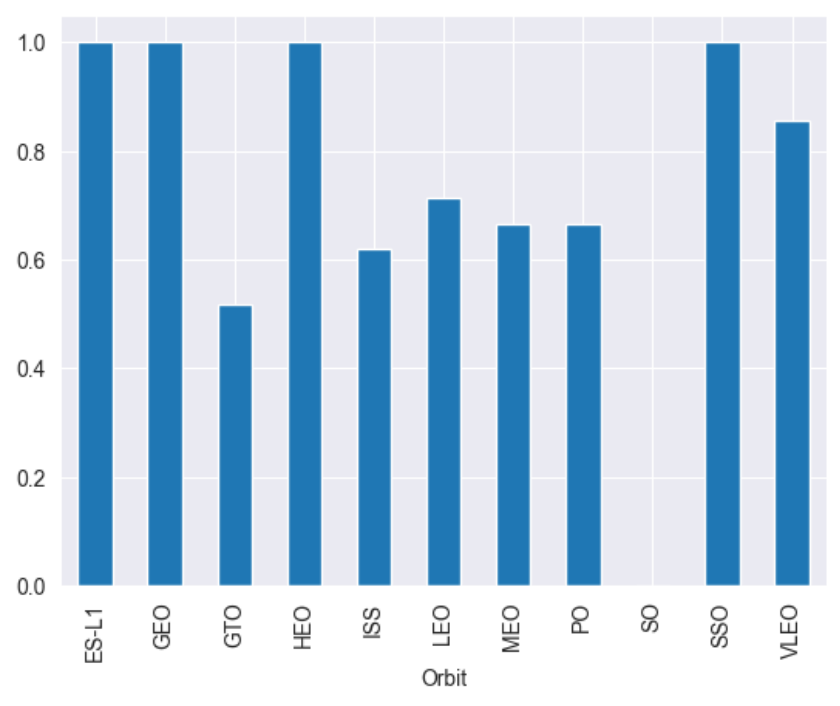
- It seems to be more success when the flight number were bigger
- It seems the team were improving the launches making the more successful

Payload vs. Launch Site



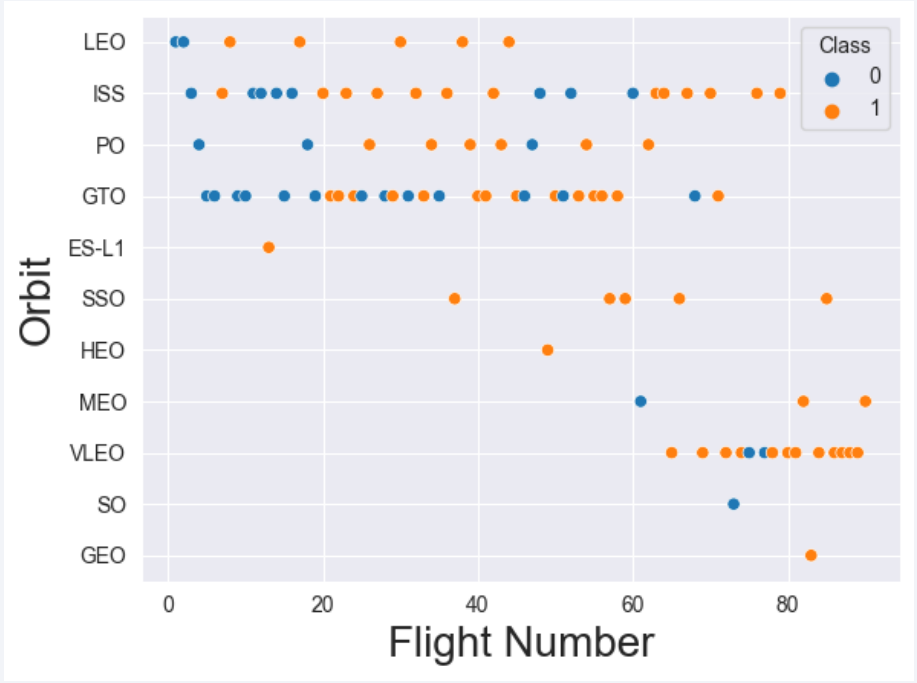
- There are no rockets launched for heavy payload mass(greater than 10000) for VAFB-SLC launchsite.
- Heavy payload tends to be successful, but lighter payload does not mean failure.

Success Rate vs. Orbit Type



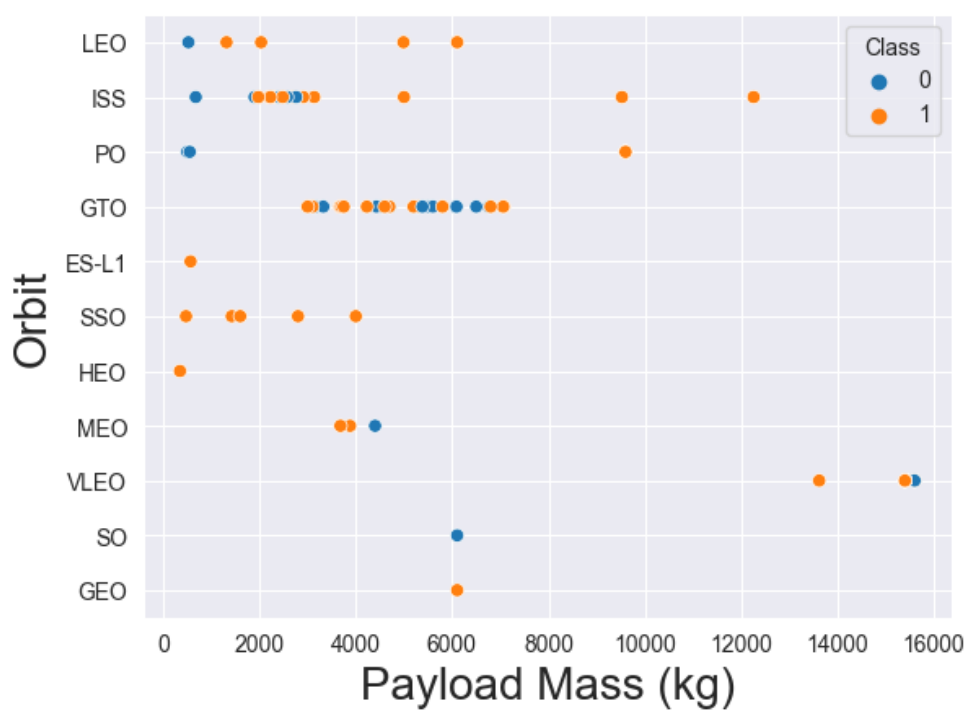
- SO orbit had no successful launch
- ES-L1, GEO, HEO and SSO had success on all launches

Flight Number vs. Orbit Type



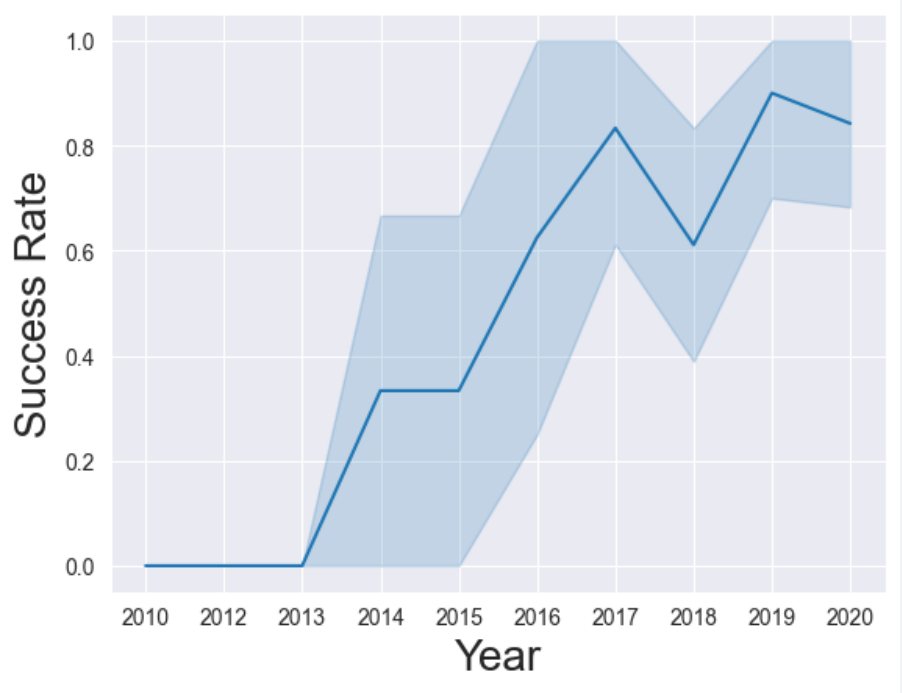
- In the LEO orbit the Success appears related to the number of flights
- On the other hand, there seems to be no relationship between flight number when in GTO orbit.

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.

Launch Success Yearly Trend



- We can observe that the success rate since 2013 kept increasing till 2020

All Launch Site Names

- Names of launch sites

```
%%sql
SELECT distinct launch_site FROM spacex
```

	Launch_Site
0	CCAFS LC-40
1	VAFB SLC-4E
2	KSC LC-39A
3	CCAFS SLC-40

Launch Site Names Begin with 'CCA'

- 5 records where launch sites begin with `CCA`

```
%%sql
SELECT * FROM spacex where launch_site like 'CCA%' limit 5
```

	Date	Time_UTC	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS_KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
0	2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
1	2010-12-08	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of...	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
2	2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
3	2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
4	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 carried by boosters from NASA

```
%%sql
SELECT sum(payload_mass_kg_) FROM spacex where customer like '%NASA%'
```

	sum(payload_mass_kg_)
0	107010

Average Payload Mass by F9 v1.1

- Average payload mass carried by booster version F9 v1.1

```
%%sql
SELECT avg(payload_mass_kg_) FROM spacex where Booster_Version like '%F9 v1.1%'
```

	avg(payload_mass_kg_)
0	2534.666667

First Successful Ground Landing Date

- Date of the first successful landing outcome on ground pad

```
%%sql
SELECT min(Date) FROM spacex where Landing_Outcome = 'Success'
```

	min("Date")
0	2018-07-22

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

```
%%sql
SELECT Booster_Version FROM spacex where Landing_Outcome = 'Success (drone ship)' and PAYLOAD_MASS_KG_ > 4000 and PAYLOAD_MASS_KG_ < 6000;
```

	Booster_Version
0	F9 FT B1022
1	F9 FT B1026
2	F9 FT B1021.2
3	F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

- Total number of successful and failure mission outcomes

```
%%sql
SELECT mission_outcome, count(1) FROM spacex group by mission_outcome
```

	Mission_Outcome	count(1)
0	Success	99
1	Failure (in flight)	1
2	Success (payload status unclear)	1

Boosters Carried Maximum Payload

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

```
%%sql
select Booster_Version from spacex where PAYLOAD_MASS_KG_ =( SELECT max(PAYLOAD_MASS_KG_) FROM spacex)
```

	Booster_Version
0	F9 B5 B1048.4
1	F9 B5 B1049.4
2	F9 B5 B1051.3
3	F9 B5 B1056.4
4	F9 B5 B1048.5
5	F9 B5 B1051.4
6	F9 B5 B1049.5
7	F9 B5 B1060.2
8	F9 B5 B1058.3
9	F9 B5 B1051.6
10	F9 B5 B1060.3
11	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 Landing_Outcome,booster_version,launch_site from spacex where year(date)=2015
```

	Landing_Outcome	Booster_Version	Launch_Site
0	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
1	Controlled (ocean)	F9 v1.1 B1013	CCAFS LC-40
2	No attempt	F9 v1.1 B1014	CCAFS LC-40
3	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40
4	No attempt	F9 v1.1 B1016	CCAFS LC-40
5	Precluded (drone ship)	F9 v1.1 B1018	CCAFS LC-40
6	Success (ground pad)	F9 FT B1019	CCAFS LC-40

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

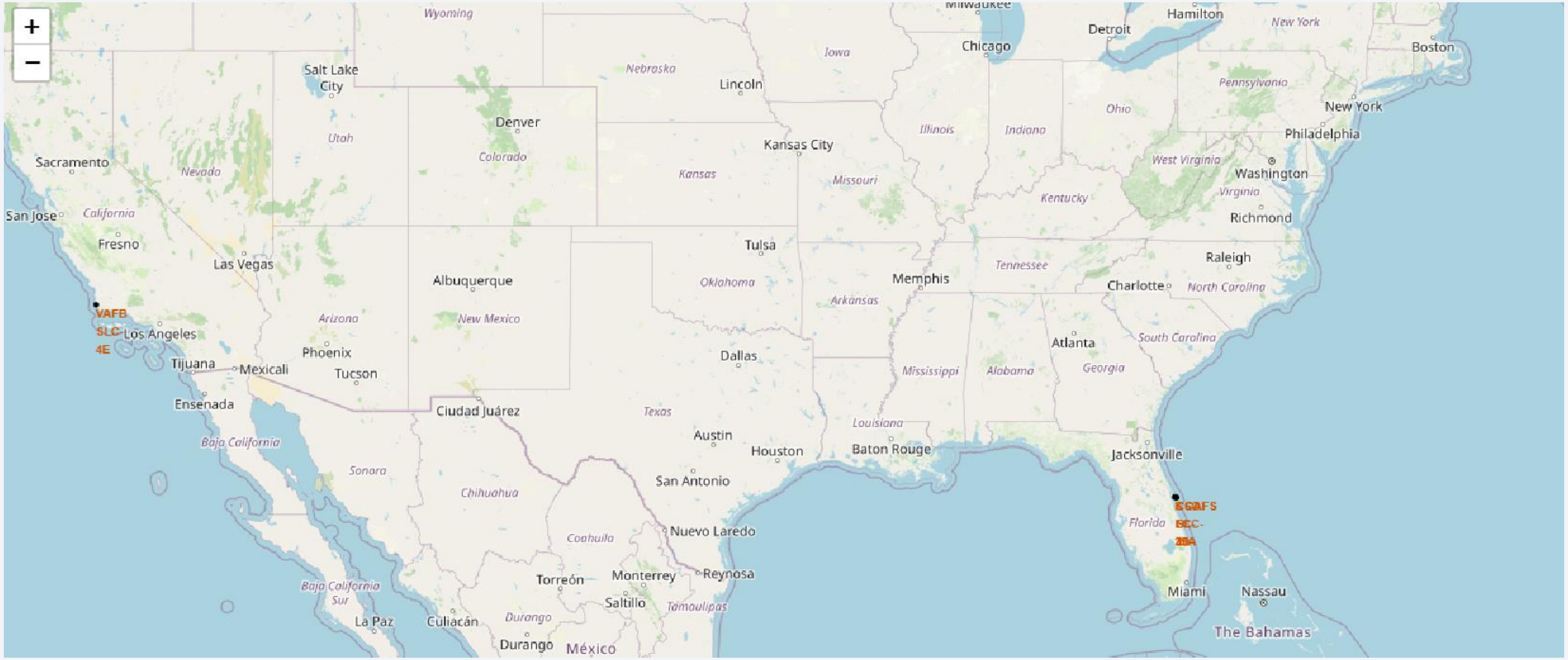
- Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

```
%%sql
select Landing_Outcome , count(1) from spacex where date between '2010-06-04' and '2017-03-20' group by Landing_Outcome order by count(1) desc
```

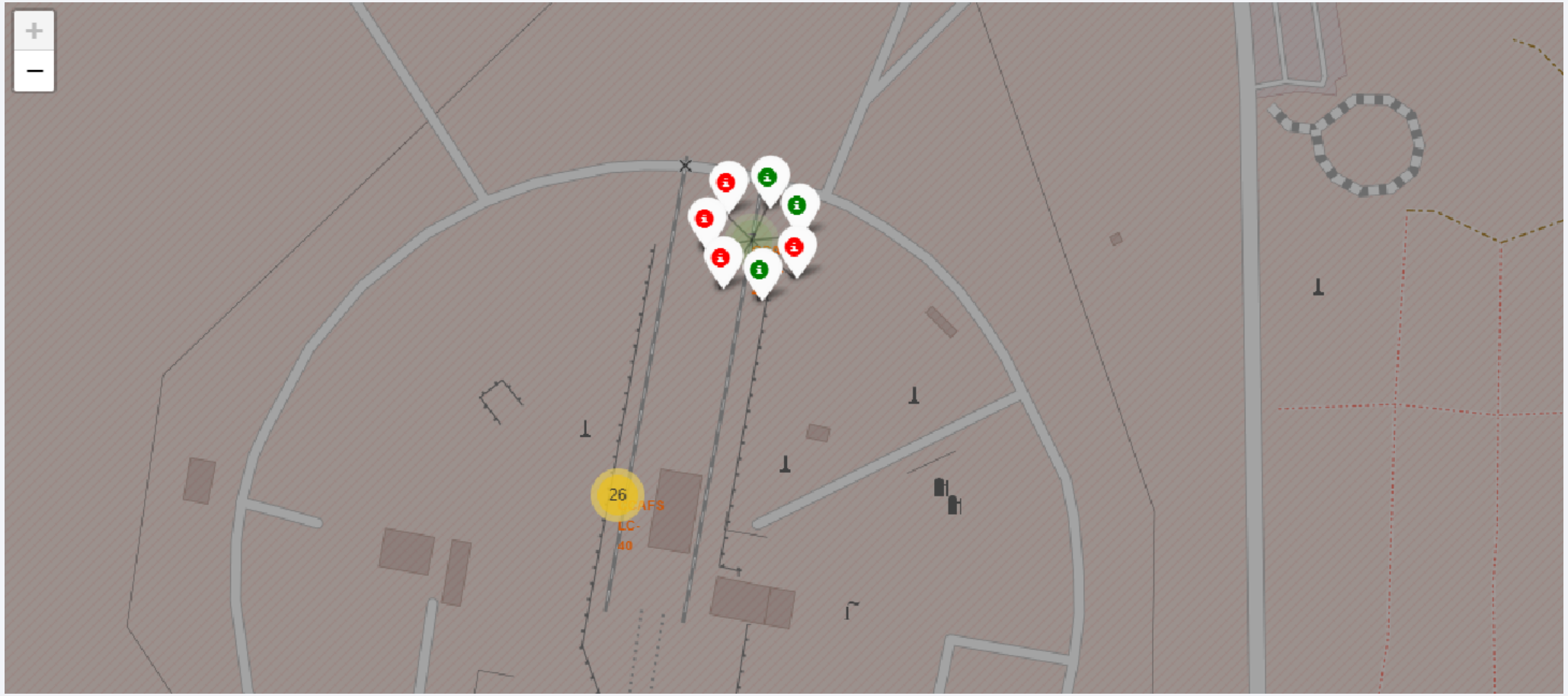
	Landing_Outcome	count(1)
0	No attempt	10
1	Failure (drone ship)	5
2	Success (drone ship)	5
3	Controlled (ocean)	3
4	Success (ground pad)	3
5	Failure (parachute)	2
6	Uncontrolled (ocean)	2
7	Precluded (drone ship)	1


Section 3
Launch Sites
Proximities Analysis

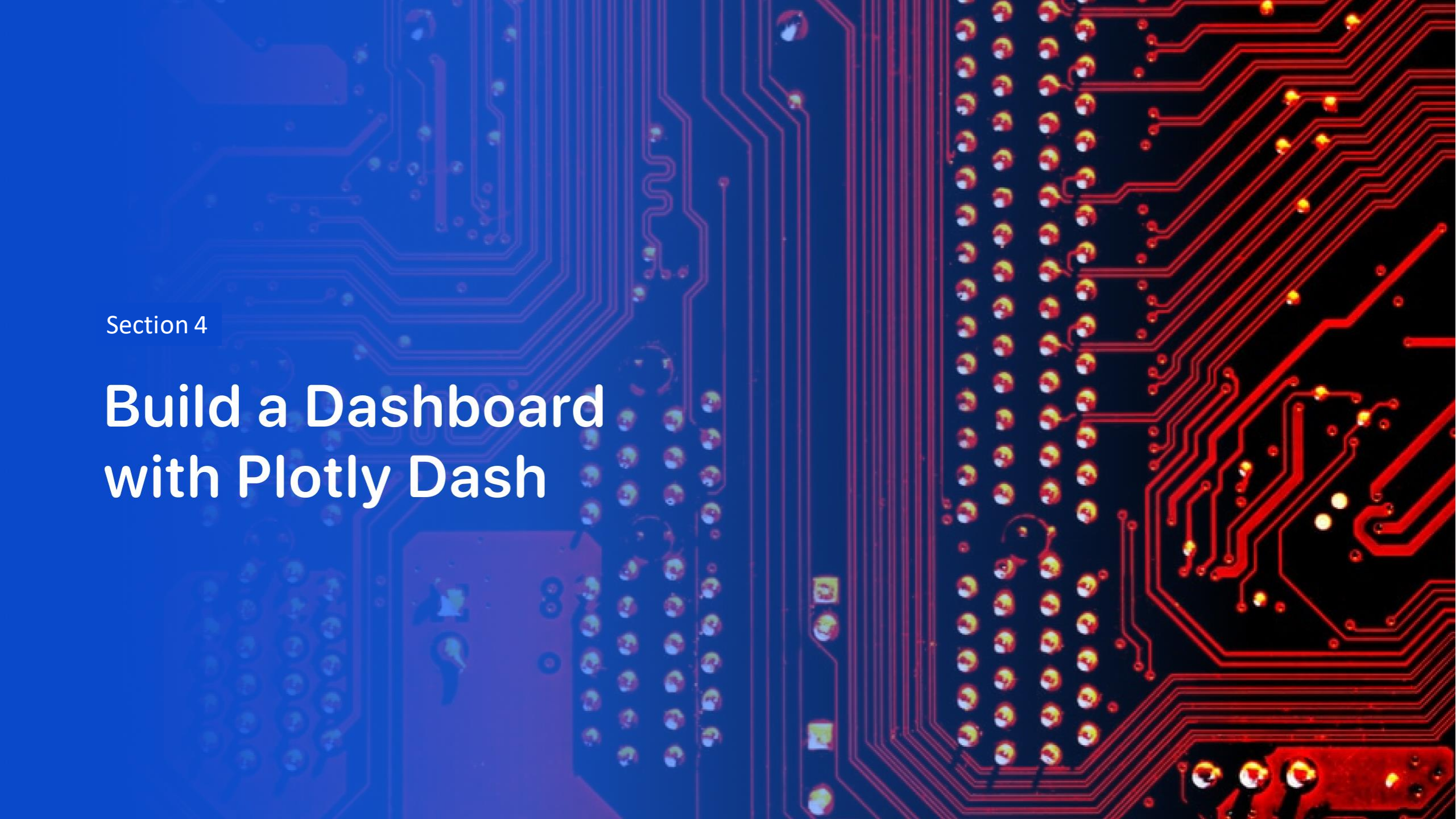
All launch sites



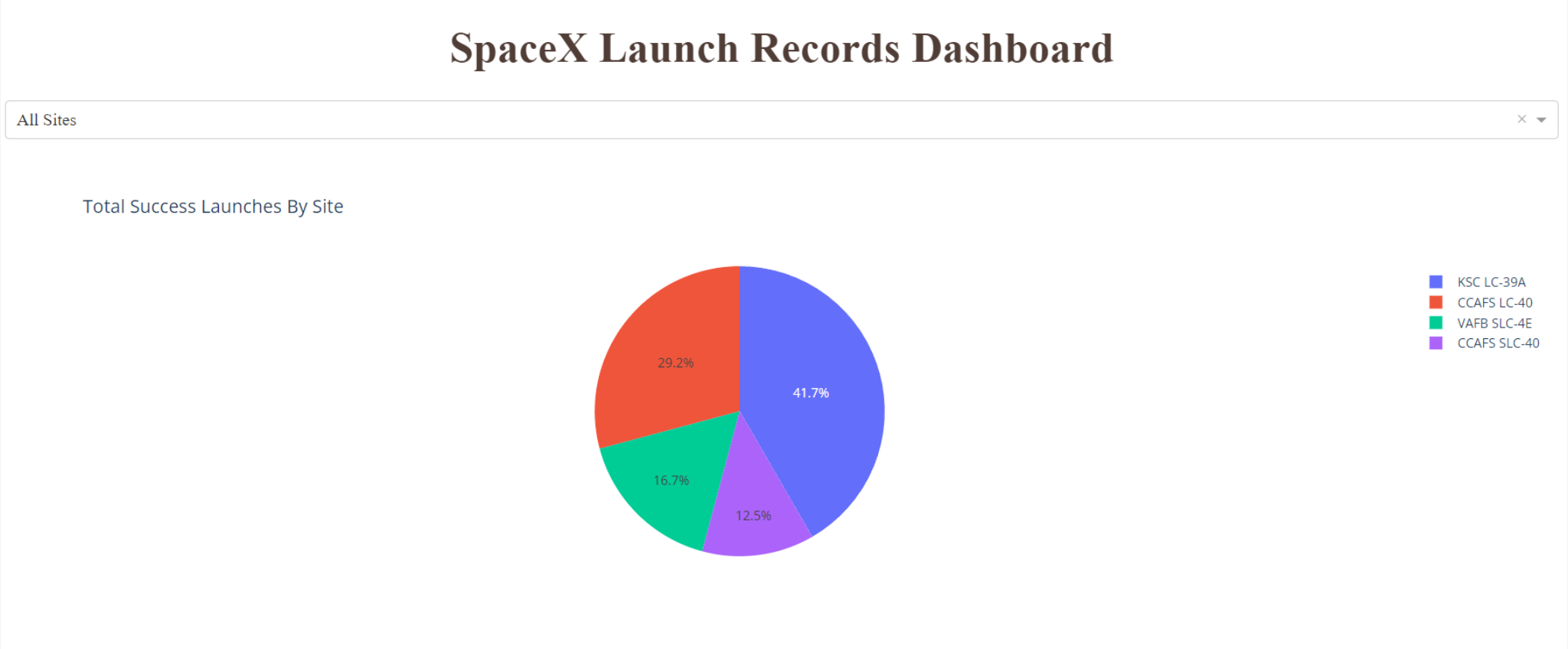
Success/failed launches for each site



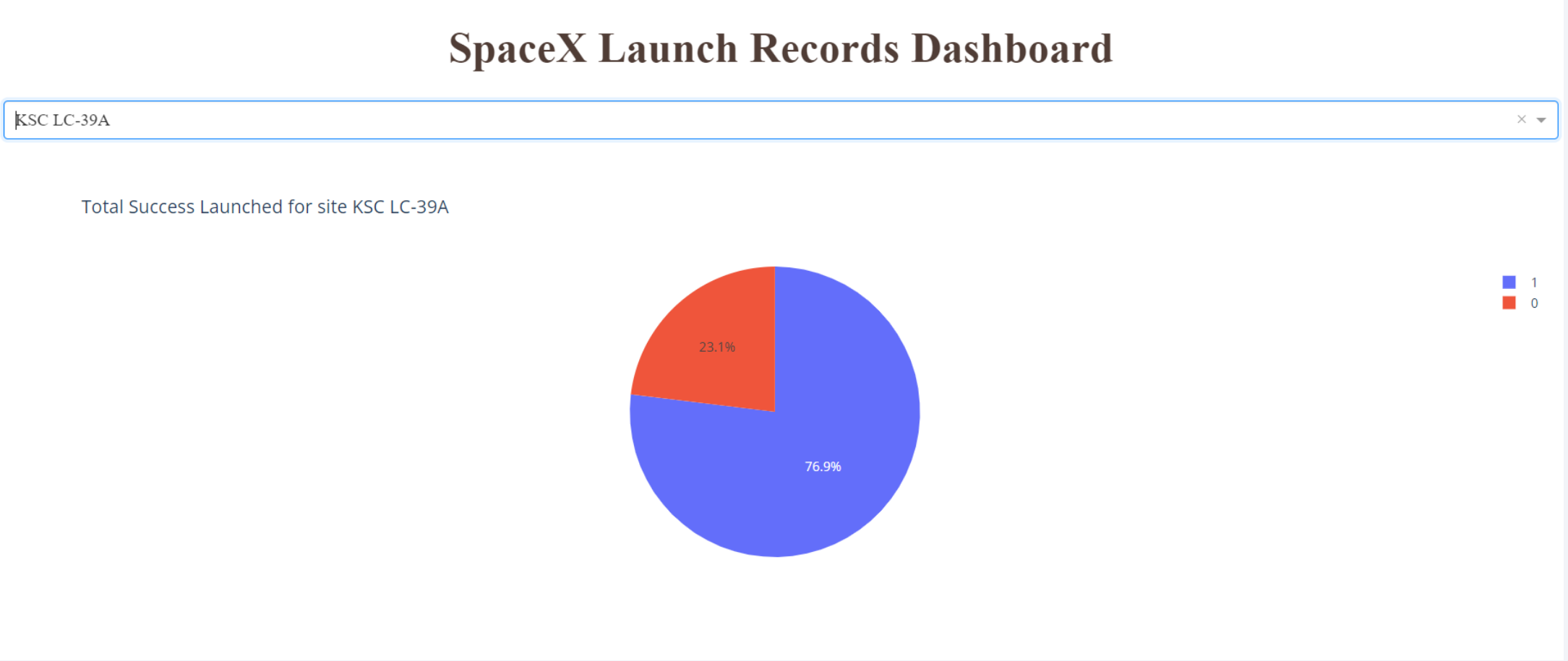
Distance between a launch site and coastline


Section 4
Build a Dashboard with Plotly Dash

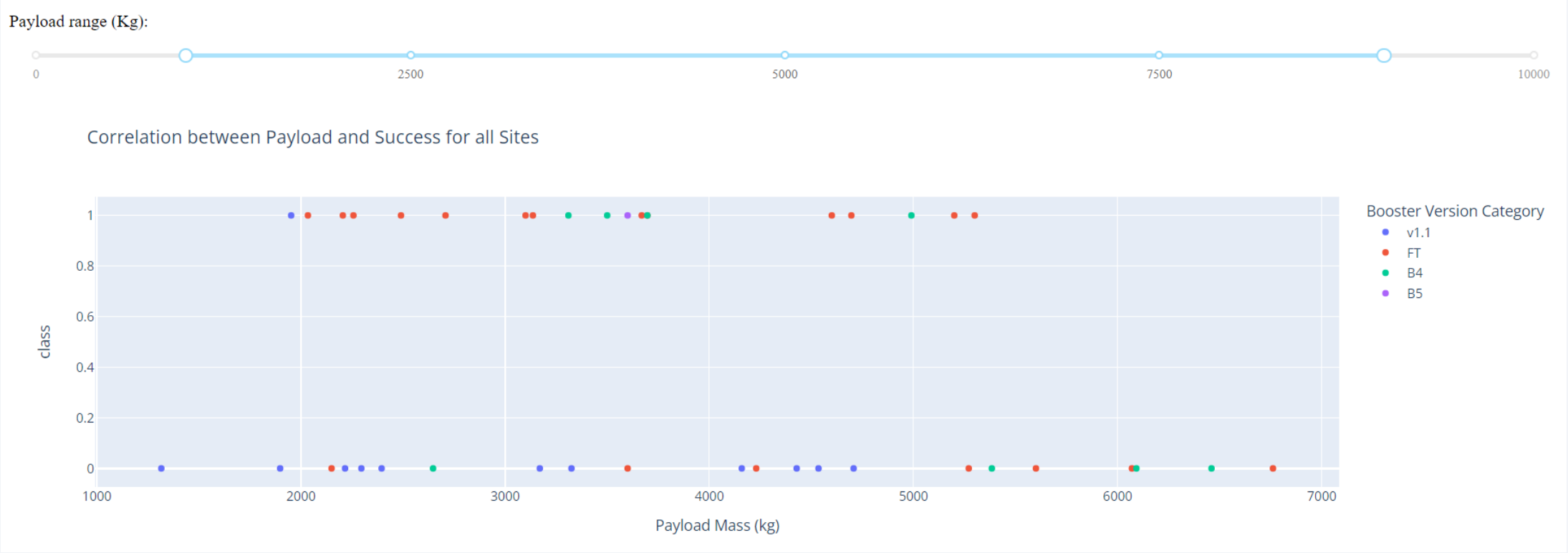
Total success Launches by site



Most successful site



Correlation between payload and success rate for all sites

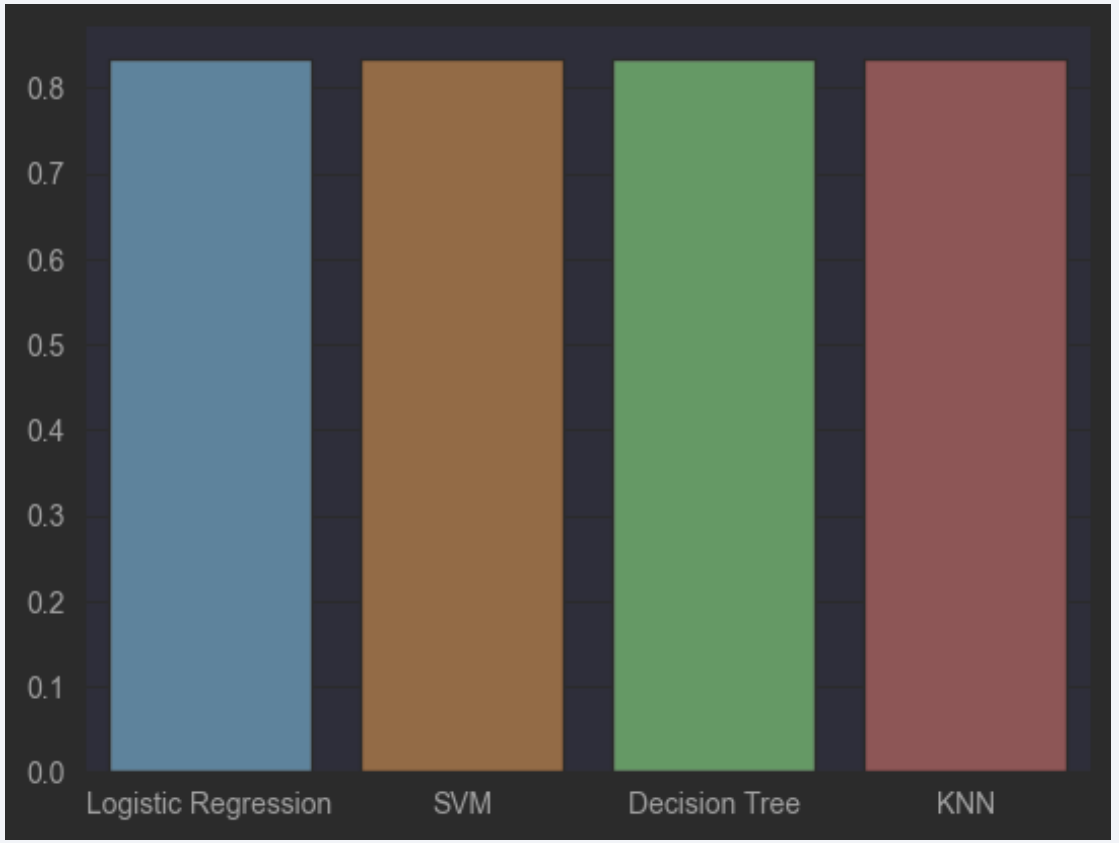


Section 5

Predictive Analysis (Classification)

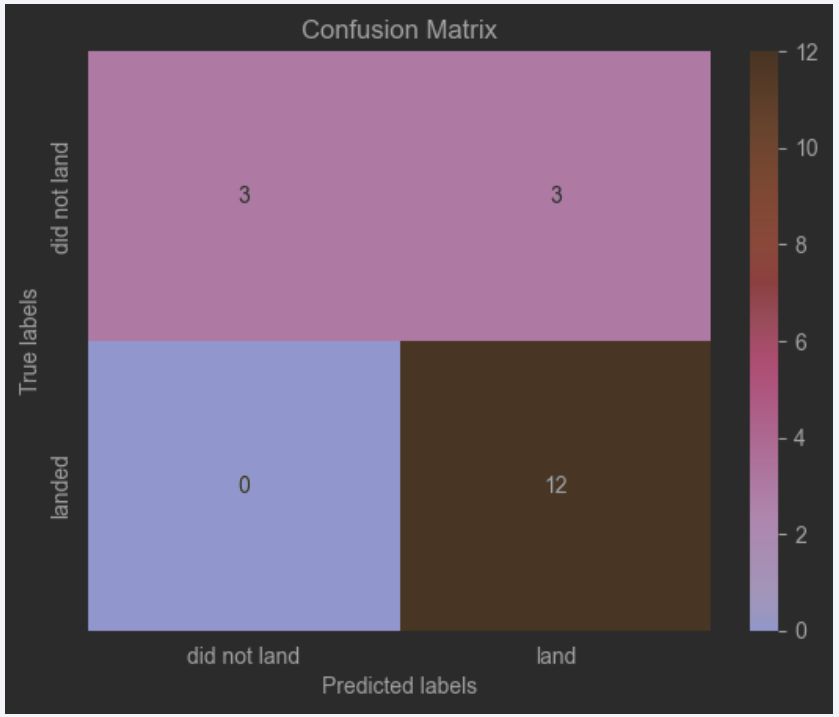
Classification Accuracy

- All modes had the same accuracy



Confusion Matrix

- 12 true-positives classifications
- 3 true-negatives classifications
- 3 false-positives classifications
- 0 false-negatives classifications



Conclusions

- All the models had the same accuracy for this dataset
- A bigger dataset may produce better results
- The models predicted some false-positives but none false-negative

Appendix

- All notebooks in this GitHub: IBM-Data-Science/Applied Data Science Capstone at main · taianf/IBM-Data-Science (github.com)

Thank you!