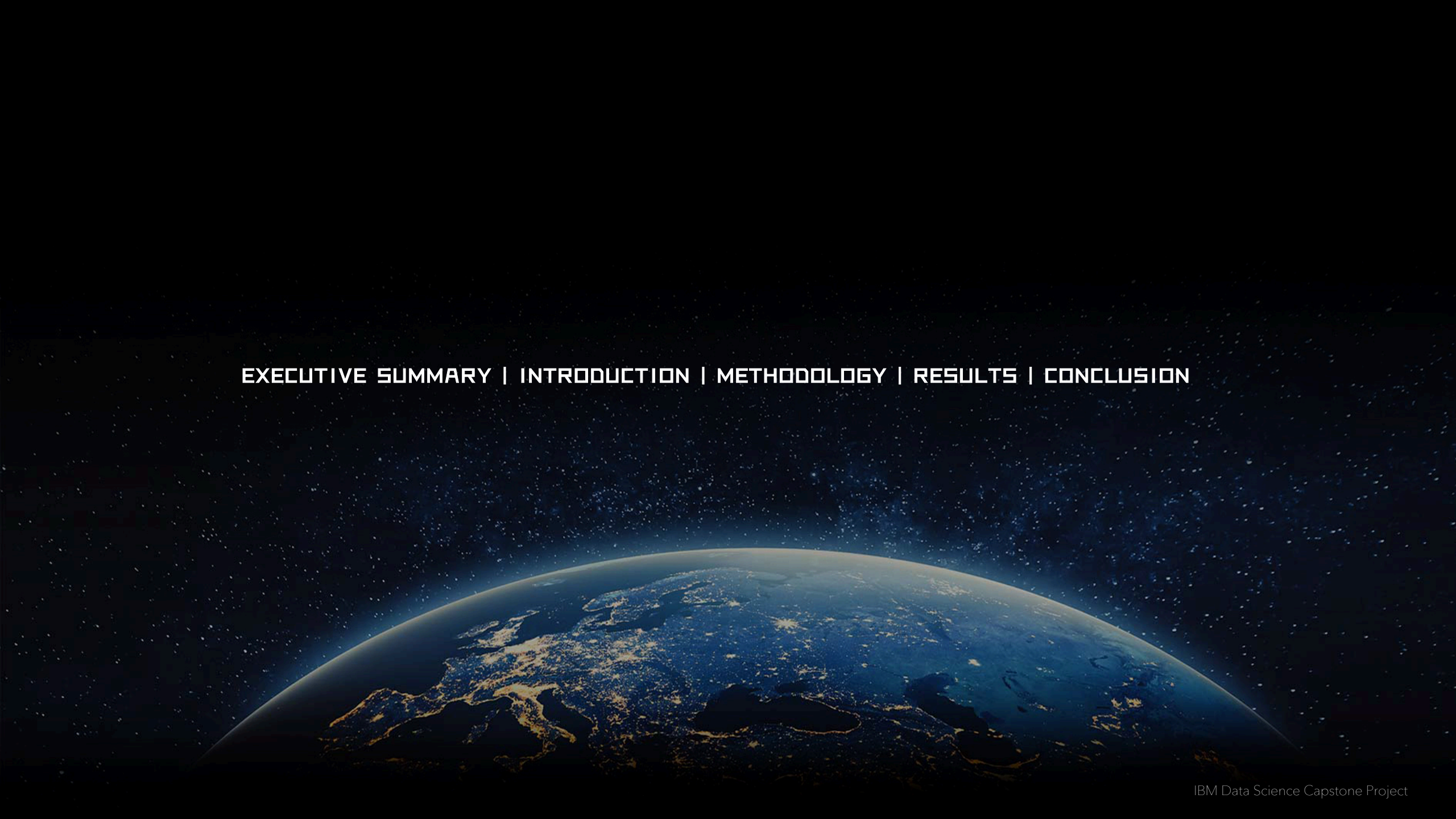
DATA SCIECEN
CAPSTONE PROJECT

PREDICTING THE LANDING
OF FALCON 9 OF SPACEX

By Zhuoquan Chen, October 2021

EXECUTIVE SUMMARY | INTRODUCTION | METHODOLOGY | RESULTS | CONCLUSION



EXECUTIVE SUMMARY

SUMMARY OF METHODOLOGIES

- Data collection via APIs & web scraping
- Data wrangling & Exploratory Data Analysis (EDA) & data visualization
- Dynamic data visualization & analysis with Folium
- Data Modeling with classification models

SUMMARY OF ALL RESULTS

- Analysis with SQL
- EDA
- Interactive visual analytics
- Predictive analysis

INTRODUCTION

PROJECT BACKGROUND

SpaceX advertised its Falcon 9 rocket launch at a cost of $62 million; Other companies cost more than $165 million. Unlike other rocket providers, SpaceX's Falcon 9 Can recover the first stage. Sometimes the first stage does not land. Sometimes it will crash. With most of the savings from first stage so if we can determine whether the first stage will land, we can determine the cost of the launch.

PROBLEMS I WANT TO FIND THE ANSWERS

- What is the successful rate of landing?
- Can weather affect the launch?
- Which condition is the most important for success?

METHODOLOGY |

DATA COLLECTION | DATA WRANGLING | EDA | INTERACTIVE VISUAL ANALYTICS | MODELS

APIs

Using SpaceX APIs to access the data we need, then via Request library to return a json format text.

Code snippet

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

response = requests.get(spacex_url)

print(response.content)
```

[output]

b'[{"fairings":{"reused":false,"recovery_attempt":false,"recovered":false,"ships":[]},"links":{"patch":{"small":"https://images2.imgbox.com/3c/0e/T8iJcSN3_o.png","large":"https://images2.imgbox.com/40/e3/GypSkayF_o.png"},"reddit":{"campaign":null,"launch":null,"media":null,"recovery":null},"flickr":{"small":[],"original":[]},"presskit":null,"webcast":"https://www.youtube.com/watch?v=0a_00nJ_Y88",”…

Then using these row data to transform to json format and product a data frame.

```
response = requests.get(static_json_url).json()
```

[output]

[{'fairings': {'reused': False, 'recovery_attempt': False, 'recovered': False, 'ships': []}, 'links': {'patch': {'small': 'https://images2.imgbox.com/3c/0e/T8iJcSN3_o.png', 'large': 'https://images2.imgbox.com/40/e3/GypSkayF_o.png'}, 'reddit': {'campaign': None, 'launch': None, 'media': None,…]

GitHub

Web Scraping

Scraping data from Wikipedia.

Code snippet

```
static_url = https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922

response = requests.get(static_url).text

print(response)
```

[output]

<!DOCTYPE html> <html class="client-nojs" lang="en" dir="ltr"> <head> <meta charset="UTF-8"/> <title>List of Falcon 9 and Falcon Heavy launches - Wikipedia</title> <script>document.documentElement.className="client-js";RLCONF={"wgBreakFrames":!1,"wgSeparatorTransformTable":["",""],"wgDigitTransformTable":["",""],"wgDefaultDateFormat":"dmy","wgMonthNames":["","January","February","Marc…

```
# Use BeautifulSoup() to create a BeautifulSoup object from a response text content

soup = BeautifulSoup(response, 'html.parser')

# Use the find_all function in the BeautifulSoup object, with element type `table`

html_tables = soup.find_all('table')

print(html_tables)
```

[output]

<table class="wikitable plainrowheaders collapsible" style="width: 100%;"> <tbody><tr> <th scope="col">Flight No.
</th> <th scope="col">Date and
time (UTC) </th>…

GitHub

DATA COLLECTION

	Flight No.	Launch site	Payload	Payload mass	Orbit	Customer	Launch outcome	Version Booster	Booster landing	Date	Time
0	1	CCAFS	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success\n	F9 v1.0B0003.1	Failure	4 June 2010	18:45
1	2	CCAFS	Dragon	0	LEO	NASA	Success	F9 v1.0B0004.1	Failure	8 December 2010	15:43
2	3	CCAFS	Dragon	525 kg	LEO	NASA	Success	F9 v1.0B0005.1	No attempt\n	22 May 2012	07:44
3	4	CCAFS	SpaceX CRS-1	4,700 kg	LEO	NASA	Success\n	F9 v1.0B0006.1	No attempt	8 October 2012	00:35
4	5	CCAFS	SpaceX CRS-2	4,877 kg	LEO	NASA	Success\n	F9 v1.0B0007.1	No attempt\n	1 March 2013	15:10

DATA WRANGLING

The process of cleaning, organizing, and unifying raw, complex, unorganized data sets so that they are more accessible for future data analysis.

1

The Number Of Launches on each site

```
df['LaunchSite'].value_counts()

[output]
CCAFS SLC 40     55
KSC LC 39A       22
VAFB SLC 4E      13
```

2

The number and occurrence of each orbit

```
df['LaunchSite'].value_counts()

[output]
CCAFS SLC 40     55
KSC LC 39A       22
VAFB SLC 4E      13
```

3

The Number and occurrence of mission outcome per orbit type

```
landing_outcomes = df['Outcome'].value_counts()
landing_outcomes

[output]
True ASDS        41
None None        19
True RTLS        14
False ASDS        6
True Ocean        5
None ASDS         2
False Ocean       2
False RTLS        1
```

4

Create a landing outcome label from Outcome column

```
landing_class =
df['Outcome'].apply(lambda x : 0
if x in bad_outcomes else 1)
landing_class

[output]
0      0
1      0
2      0
3      0
4      0
```

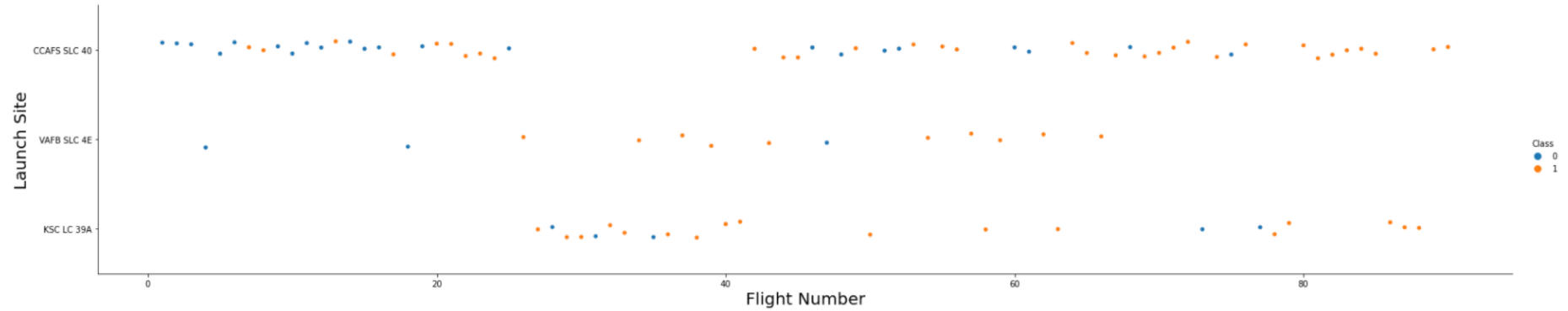
DATA WRANGLING

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial
0	1	2010-06-04	Falcon 9	6104.959412	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0003
1	2	2012-05-22	Falcon 9	525.000000	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0005
2	3	2013-03-01	Falcon 9	677.000000	ISS	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0007
3	4	2013-09-29	Falcon 9	500.000000	PO	VAFB SLC 4E	False Ocean	1	False	False	False	NaN	1.0	0	B1003
4	5	2013-12-03	Falcon 9	3170.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B1004

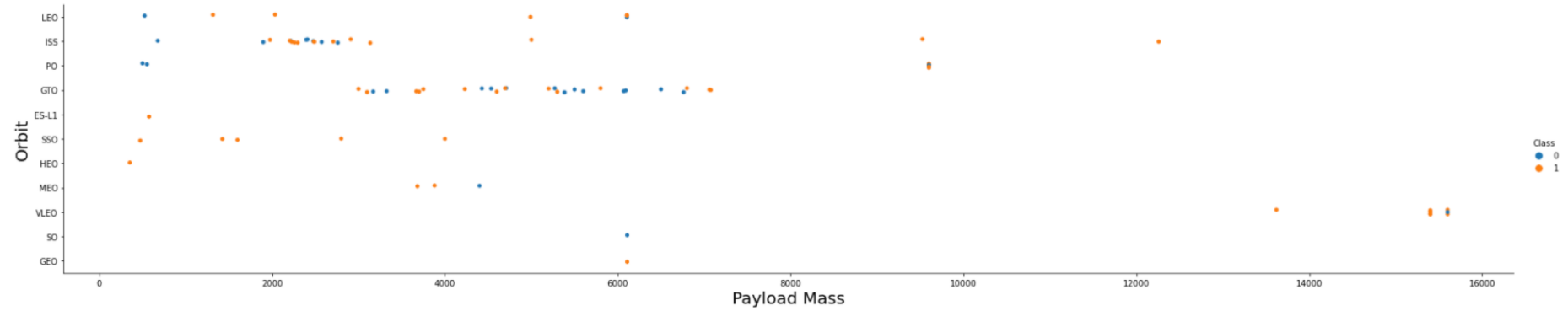
GitHub

EDA WITH DATA VISUALIZATION

Visualize the relationship between Flight Number and Launch Site



Visualize the relationship between Payload and Orbit type



GitHub

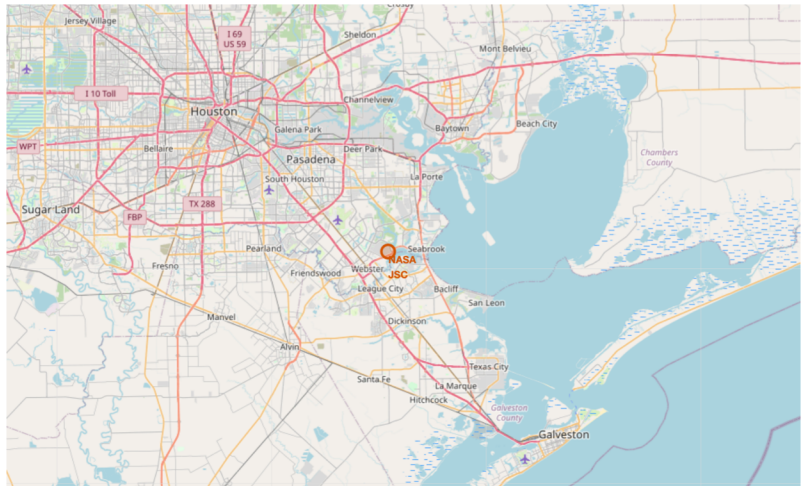
Executed SQL queries to solve the tasks

- Display the names of the 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 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
- List the names of the booster versions which have carried the maximum payload mass.
- List the failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015
- 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

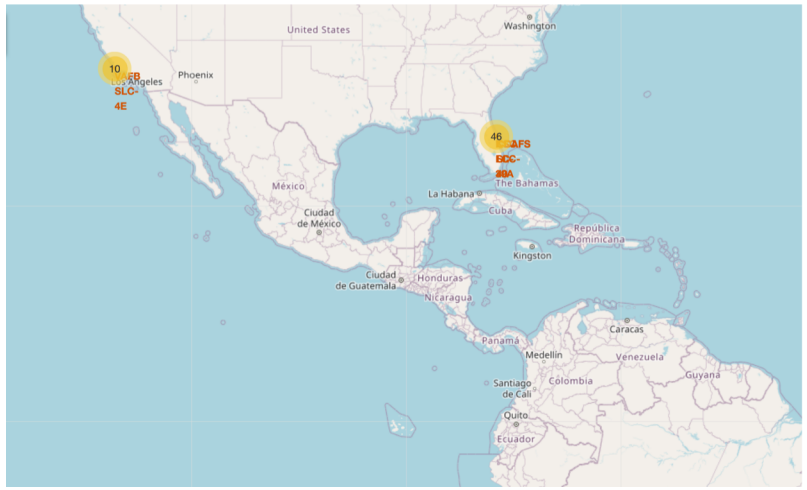
GitHub

BUILD AN INTERACTIVE MAP WITH FOLIUM

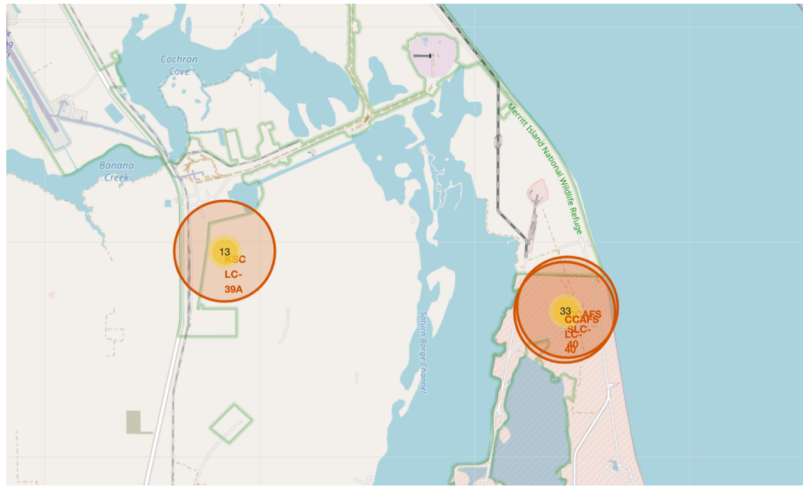
an initial center location to be NASA Johnson Space Center at Houston, Texas.



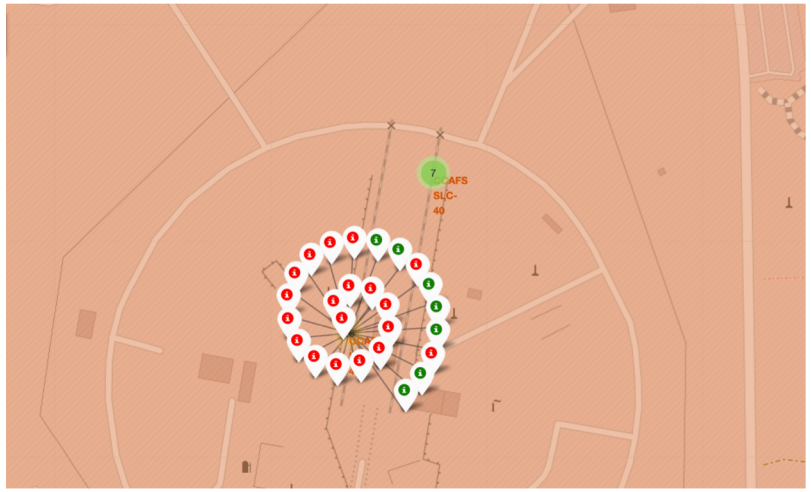
Mark all launch sites on a map



The number of the launches of each launch site



Mark the success/failed launches for each site on the map



[GitHub]

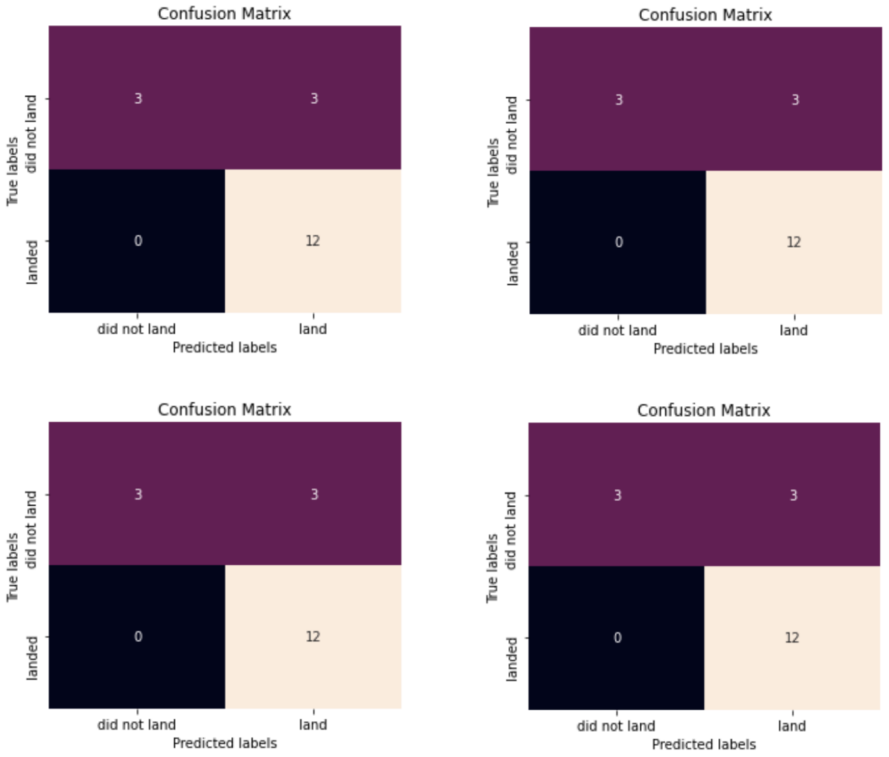
PREDICTIVE ANALYSIS (CLASSIFICATION)

1. Standardize the scale of all numeric type data
2. Split the dataset into X_train, X_test, Y_train, Y_test
3. Create a GridSearchCV object for search an optimal value of parameters
4. Create a confusing matrix for checking performance of the model

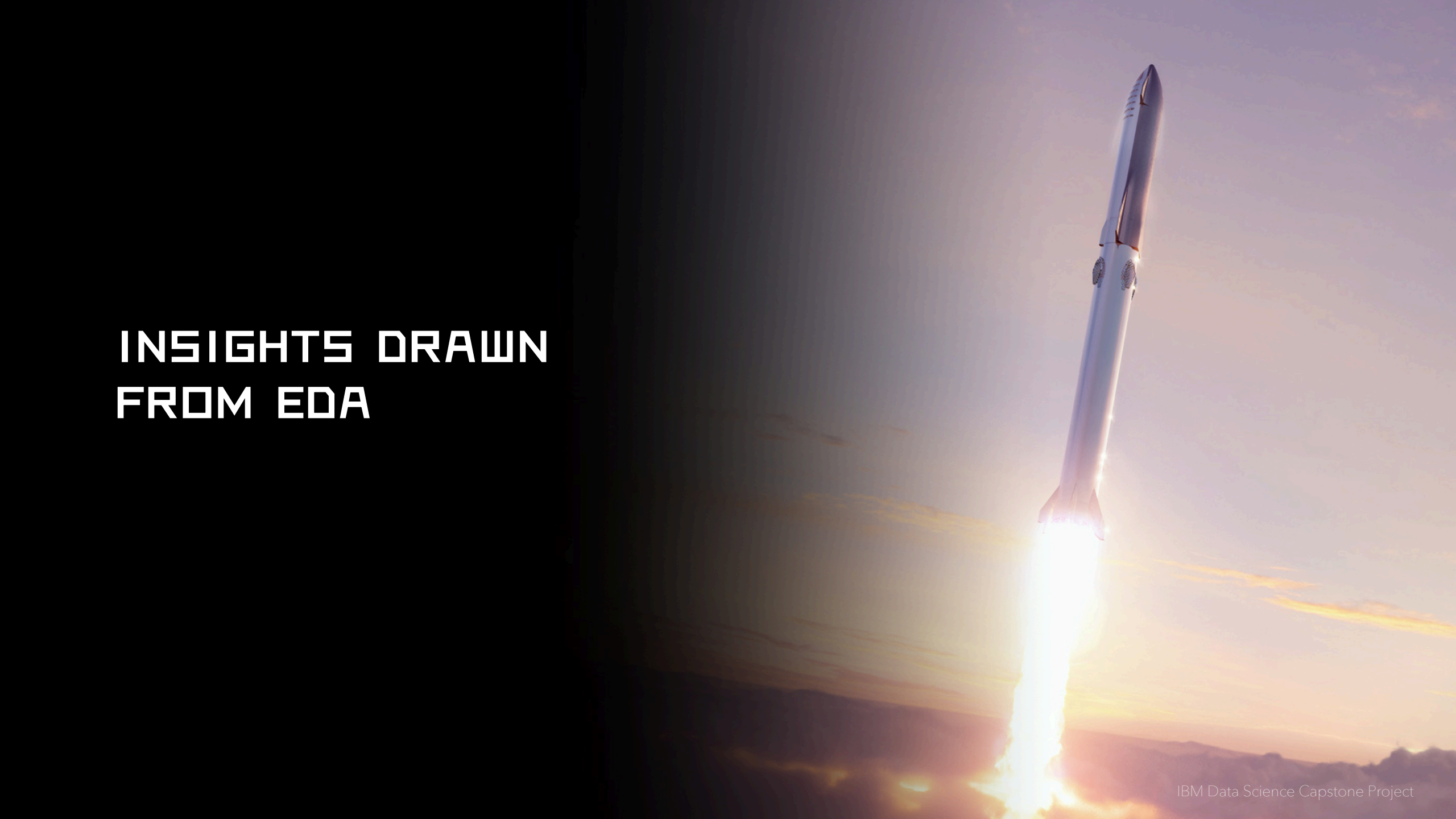
GitHub

RESULTS

	Accuracy
Decision Tree	0.887500
KNN	0.848214
SVM	0.848214
Logistic_Rgression	0.846429

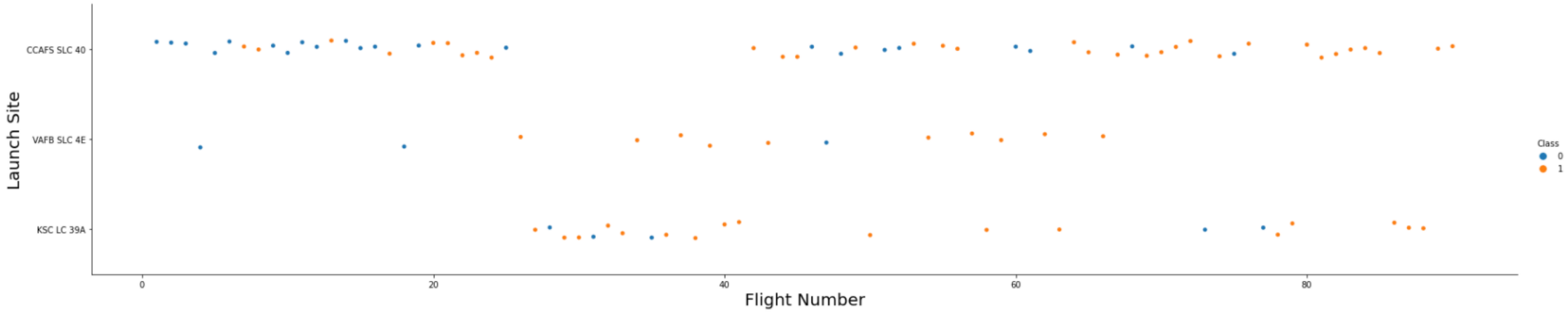


INSIGHTS DRAWN FROM EDA



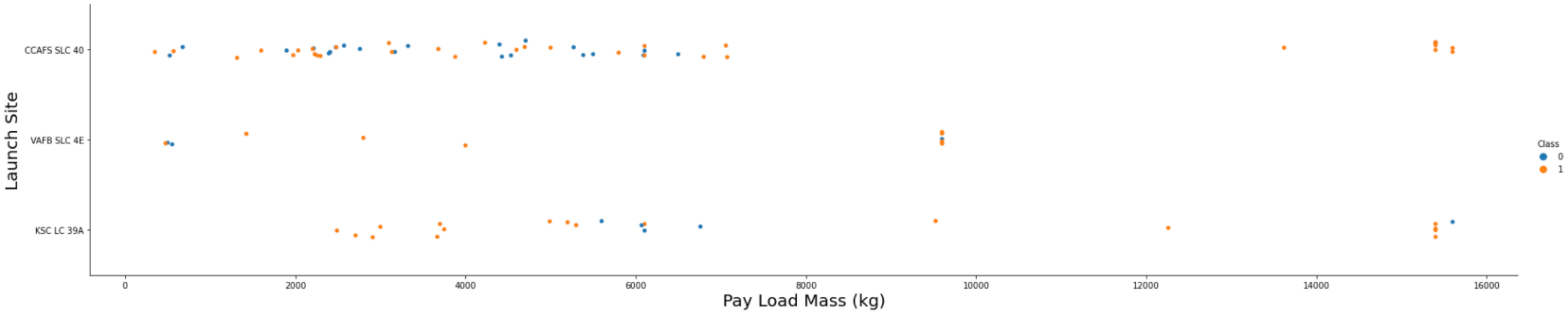
FLIGHT NUMBER vs. LAUNCH SITE

Different launch sites have different success rates. CCAFS LC-40, has a success rate of 60 %, while KSC LC-39A and VAFB SLC 4E has a success rate of 77%.



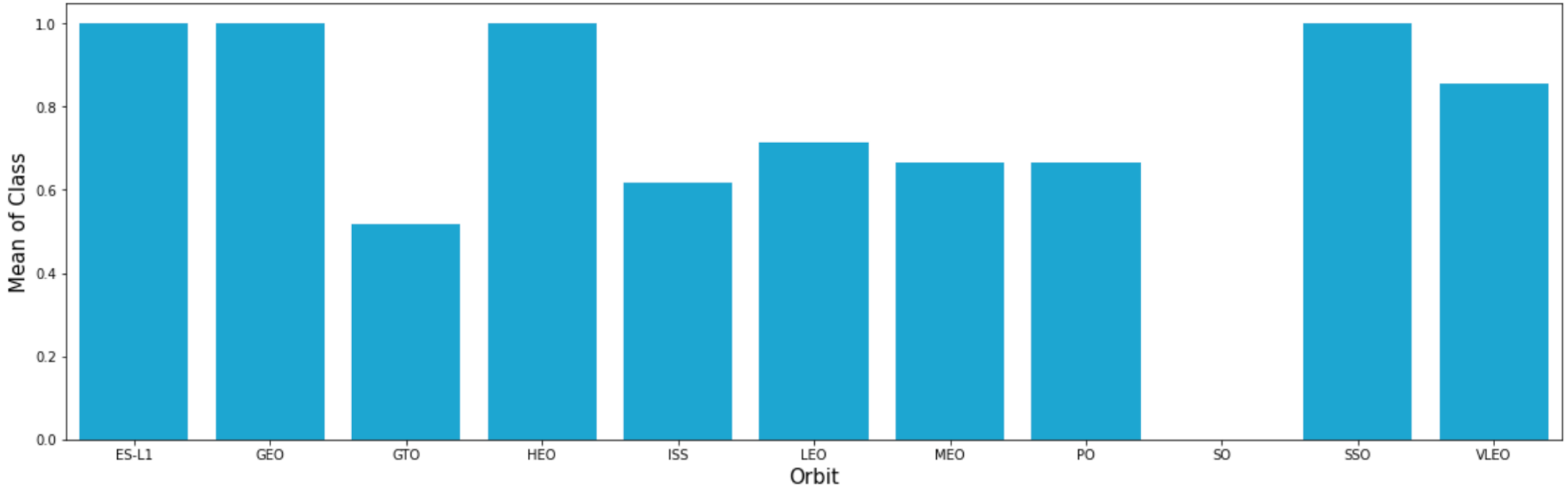
PAYLOAD vs. LAUNCH SITE

The VAFB-SLC launch site there are no rockets launched for heavy payload mass (greater than 10000).



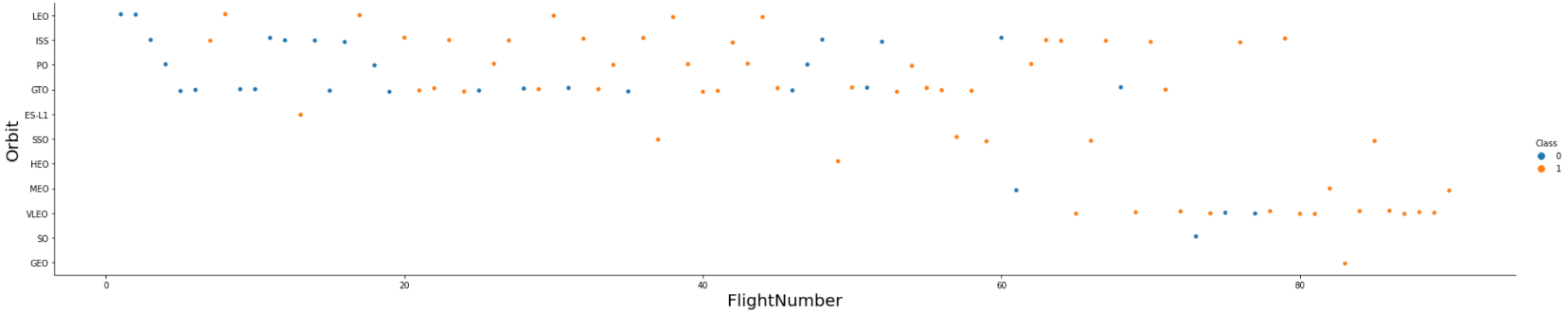
PAYLOAD vs. LAUNCH SITE

Success rate of each orbit type of SO was 0.



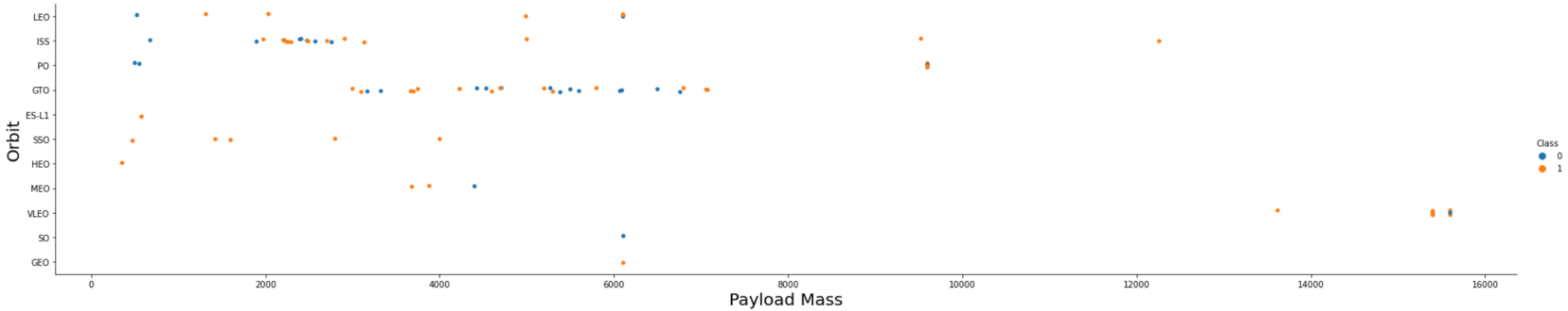
FLIGHT NUMBER vs. ORBIT TYPE

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.



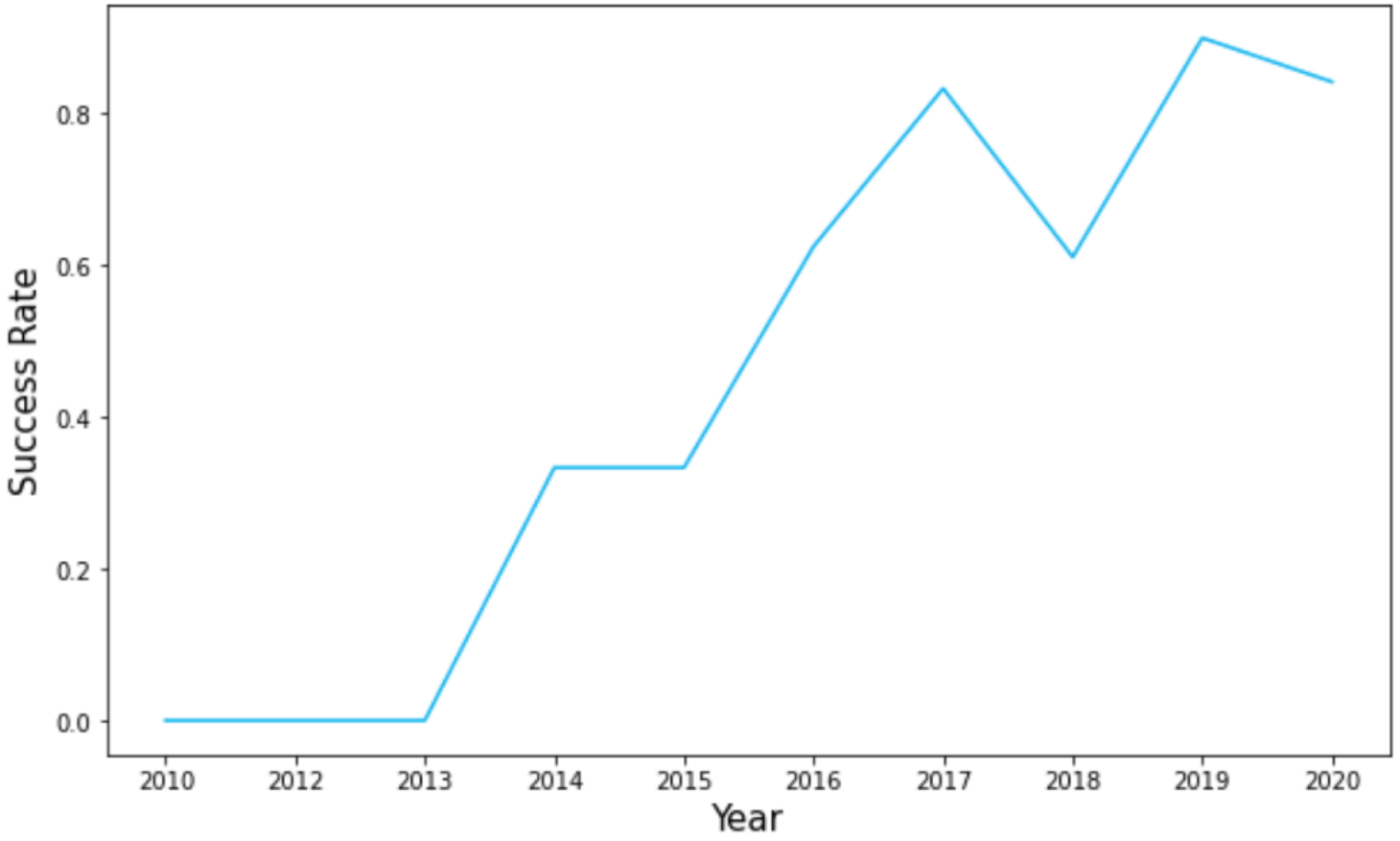
PAYLOAD vs. ORBIT TYPE

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.



LAUNCH SUCCESS YEARLY TREND

The success rate since 2013 kept increasing till 2020



SQL ANALYSIS

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

```
1 %sql select DISTINCT launch_site from SPACEXTBL;
```

```
 * ibm_db_sa://xss13292:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:3132
1/bludb
Done.
```

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

SQL ANALYSIS

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

```
1 %sql select * from SPACEXTBL where launch_site LIKE 'CCA%' limit 5;
```

```
 * ibm_db_sa://xss13292:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:3132
1/bludb
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

SQL ANALYSIS

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

```
1 %sql select sum(payload_mass__kg_) as total_payload_mass_by_NASA_CRS from SPACEXTBL where customer = 'NASA (CRS)';
```

```
 * ibm_db_sa://xss13292:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:3132
1/bludb
Done.
```

total_payload_mass_by_nasa_crs
45596

SQL ANALYSIS

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

```
1 %sql select booster_version, avg(payload_mass__kg_) as average from SPACEXTBL Group by booster_version having boost
```

```
 * ibm_db_sa://xss13292:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:3132
1/bludb
Done.
```

booster_version	average
F9 v1.1	2928

SQL ANALYSIS

List the date when the first successful landing outcome in ground pad was achieved

```
1 %sql select min(DATE) as "the date of first acheived" from SPACEXTBL where landing__outcome = 'Success (ground pad)
```

```
 * ibm_db_sa://xss13292:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:3132
1/bludb
Done.
```

the date of first acheived
01-05-2017

SQL ANALYSIS

List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 600

```
%%sql
select booster_version
from SPACEXTBL
where landing__outcome = 'Success (drone ship)' and payload_mass__kg_ between 4000 and 6000;
```

```
 * ibm_db_sa://xss13292:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:3132
1/bludb
Done.
```

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

SQL ANALYSIS

List the total number of successful and failure mission outcomes

```
1 %sql select count(mission_outcome) as "total number of successful and failure" from SPACEXTBL;
```

```
 * ibm_db_sa://xss13292:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:3132
1/bludb
Done.
```

total number of successful and failure
101

SQL ANALYSIS

List the names of the booster versions which have carried the maximum payload mass.

Use a subquery

```
%%sql
select booster_version, payload_mass__kg_
from SPACEXTBL
where payload_mass__kg_ = (select max(payload_mass__kg_) from SPACEXTBL);
```

```
 * ibm_db_sa://xss13292:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:3132
1/bludb
Done.
```

booster_version	payload_mass__kg_
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

SQL ANALYSIS

List the failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015

```
%%sql
select DATE, landing__outcome, booster_version, launch_site
from SPACEXTBL
where landing__outcome = 'Failure (drone ship)' and DATE like '%2015';
```

```
 * ibm_db_sa://xss13292:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:3132
1/bludb
Done.
```

DATE	landing__outcome	booster_version	launch_site
10-01-2015	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
14-04-2015	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

```
%%sql
select landing__outcome, booster_version, launch_site, DATE
from SPACEXTBL
where landing__outcome = 'Failure (drone ship)';
```

```
 * ibm_db_sa://xss13292:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:3132
1/bludb
Done.
```

landing__outcome	booster_version	launch_site	DATE
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40	10-01-2015
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40	14-04-2015
Failure (drone ship)	F9 v1.1 B1017	VAFB SLC-4E	17-01-2016
Failure (drone ship)	F9 FT B1020	CCAFS LC-40	04-03-2016
Failure (drone ship)	F9 FT B1024	CCAFS LC-40	15-06-2016

SQL ANALYSIS

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

```
%%sql
select landing__outcome, count(landing__outcome)
from SPACEXTBL
where DATE >= '04-06-2010' and DATE <= '20-03-2017'
group by landing__outcome
order by count(landing__outcome) desc;
```

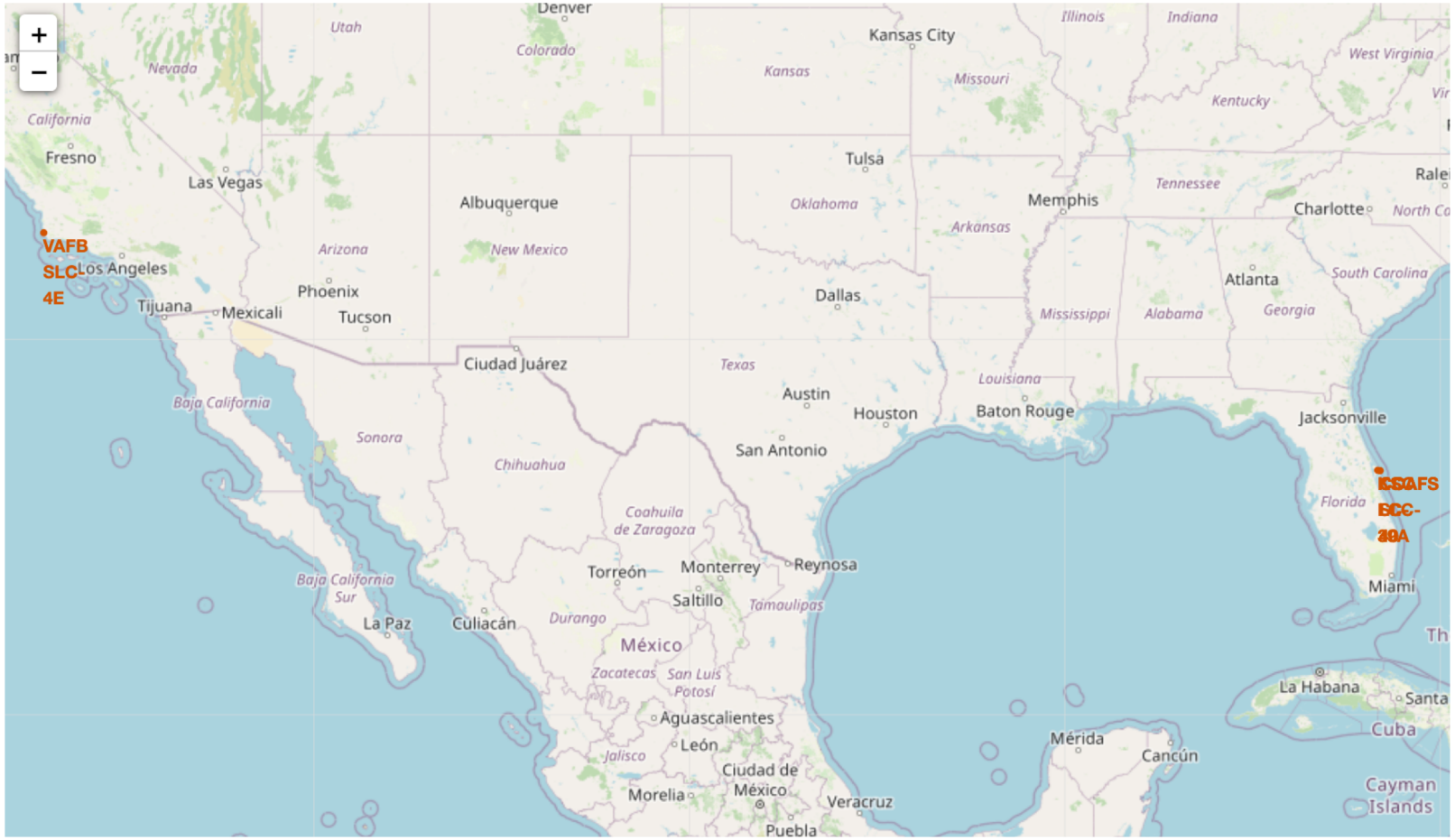
```
 * ibm_db_sa://xss13292:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:3132
1/bludb
Done.
```

landing__outcome	2
Success	20
No attempt	11
Success (drone ship)	8
Success (ground pad)	6
Failure (drone ship)	4
Controlled (ocean)	3
Failure	3
Failure (parachute)	2

LAUNCH SITES PROXIMITIES ANALYSIS

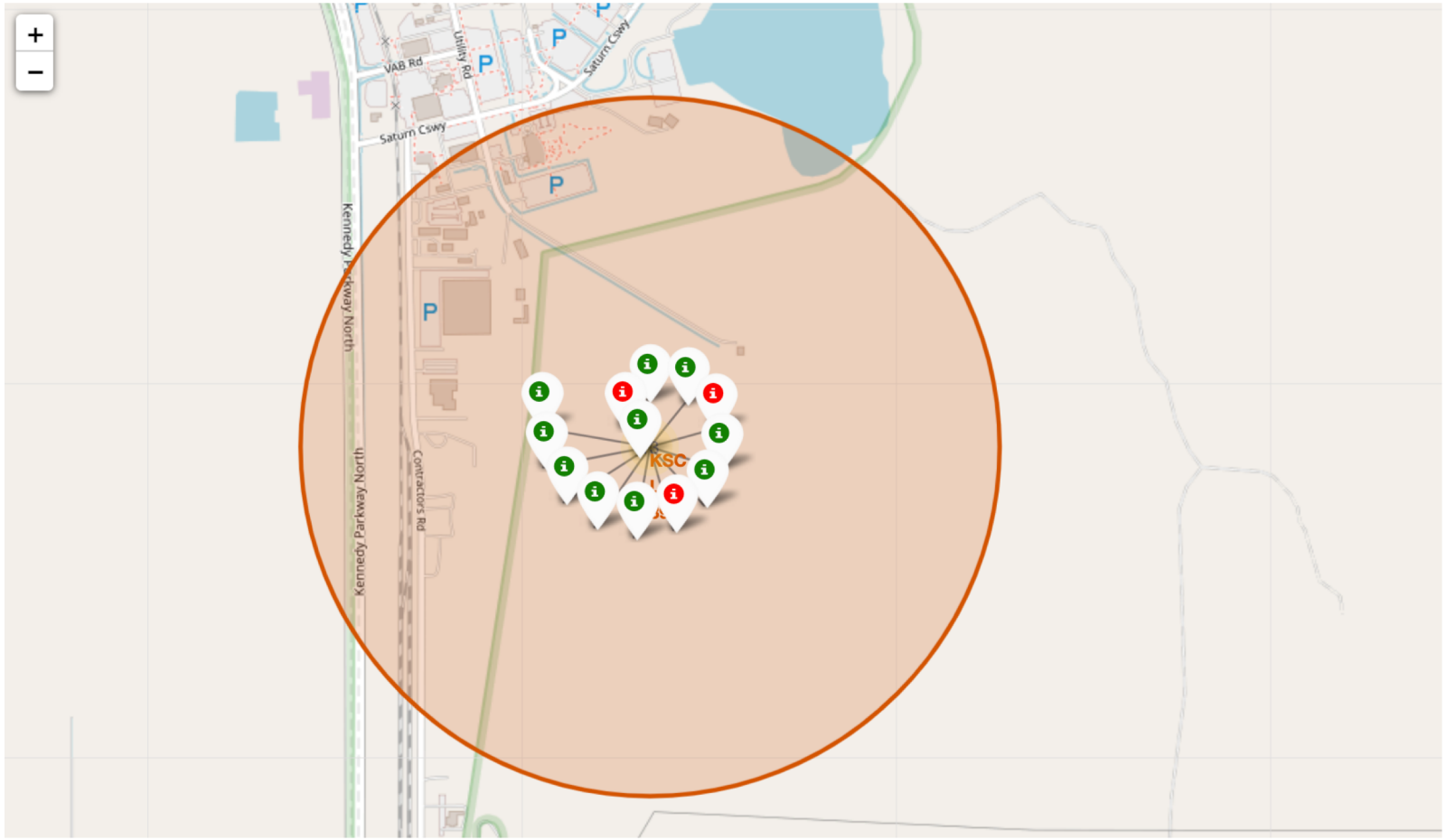
LAUNCH SITE LOCATIONS

All of launch sites are closed coasts of the United States.



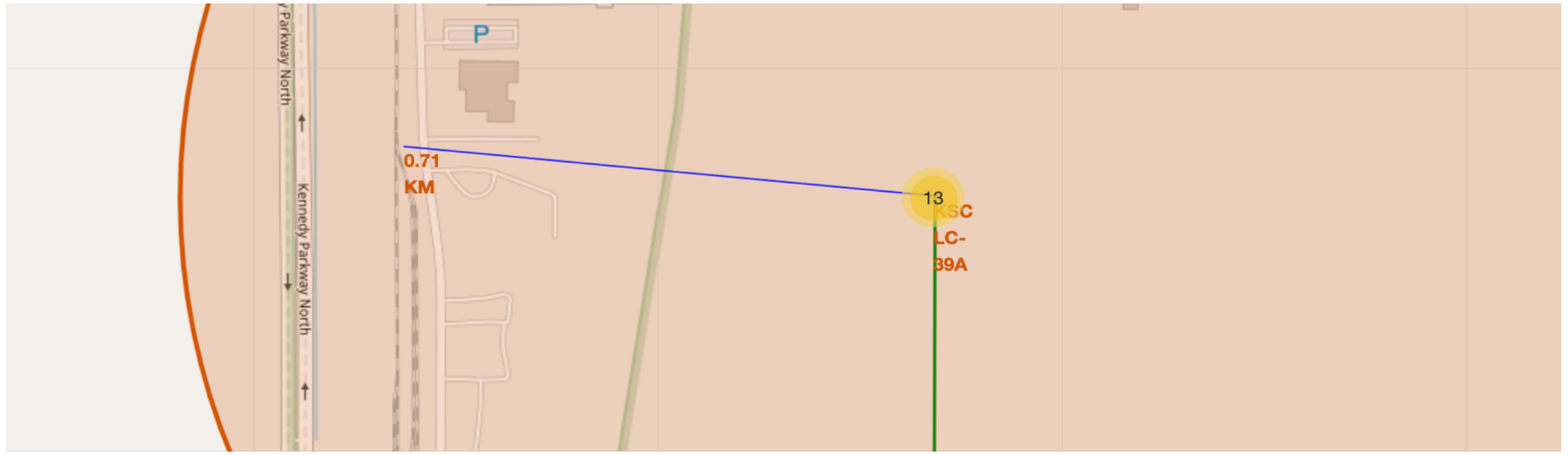
LAUNCH SITE LABELS

The green markers represent the successful launches, and the red markers represent the fail launches.

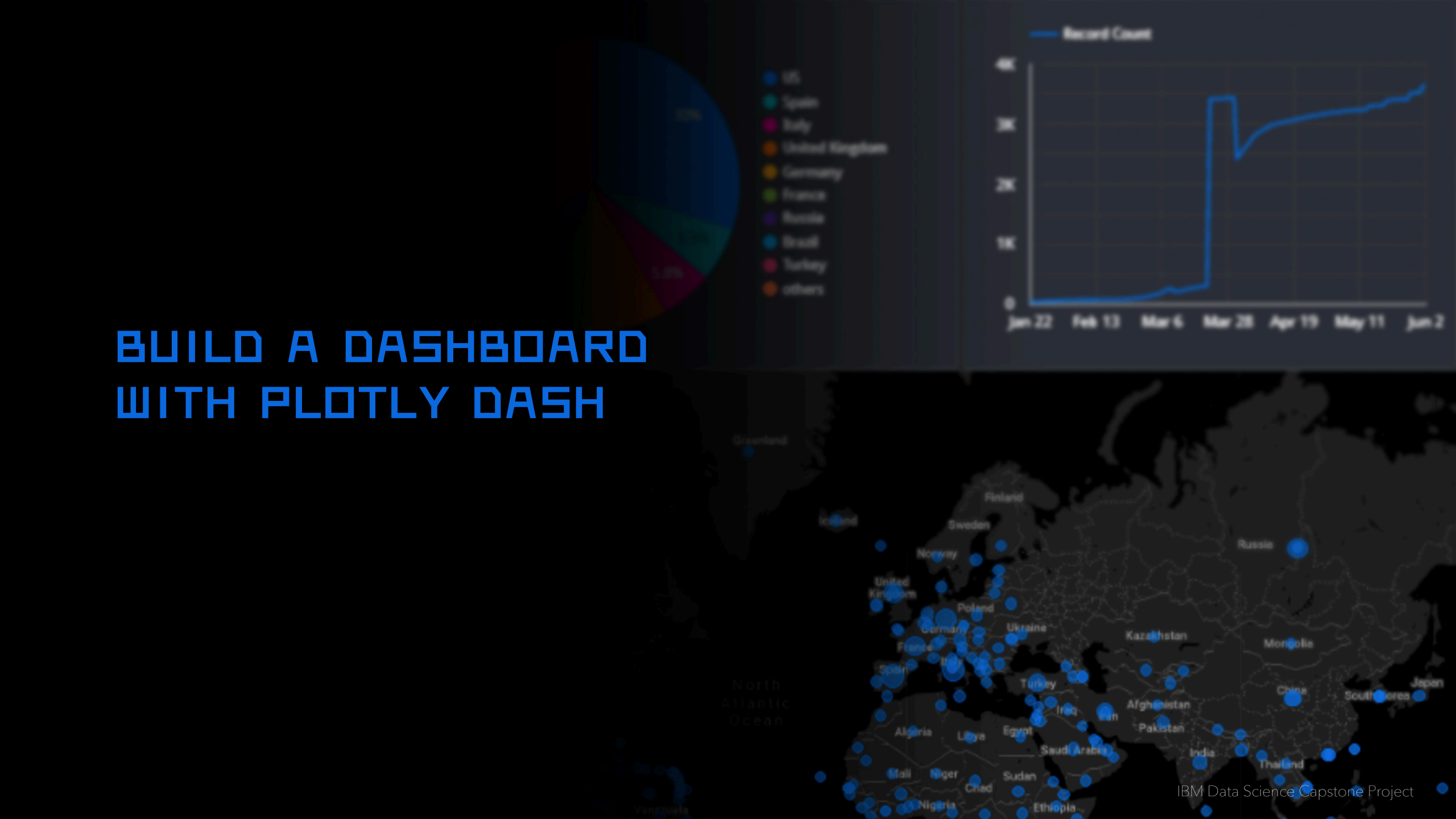


THE DISTANCES BETWEEN LAUNCH SITES AND OTHER PLACES

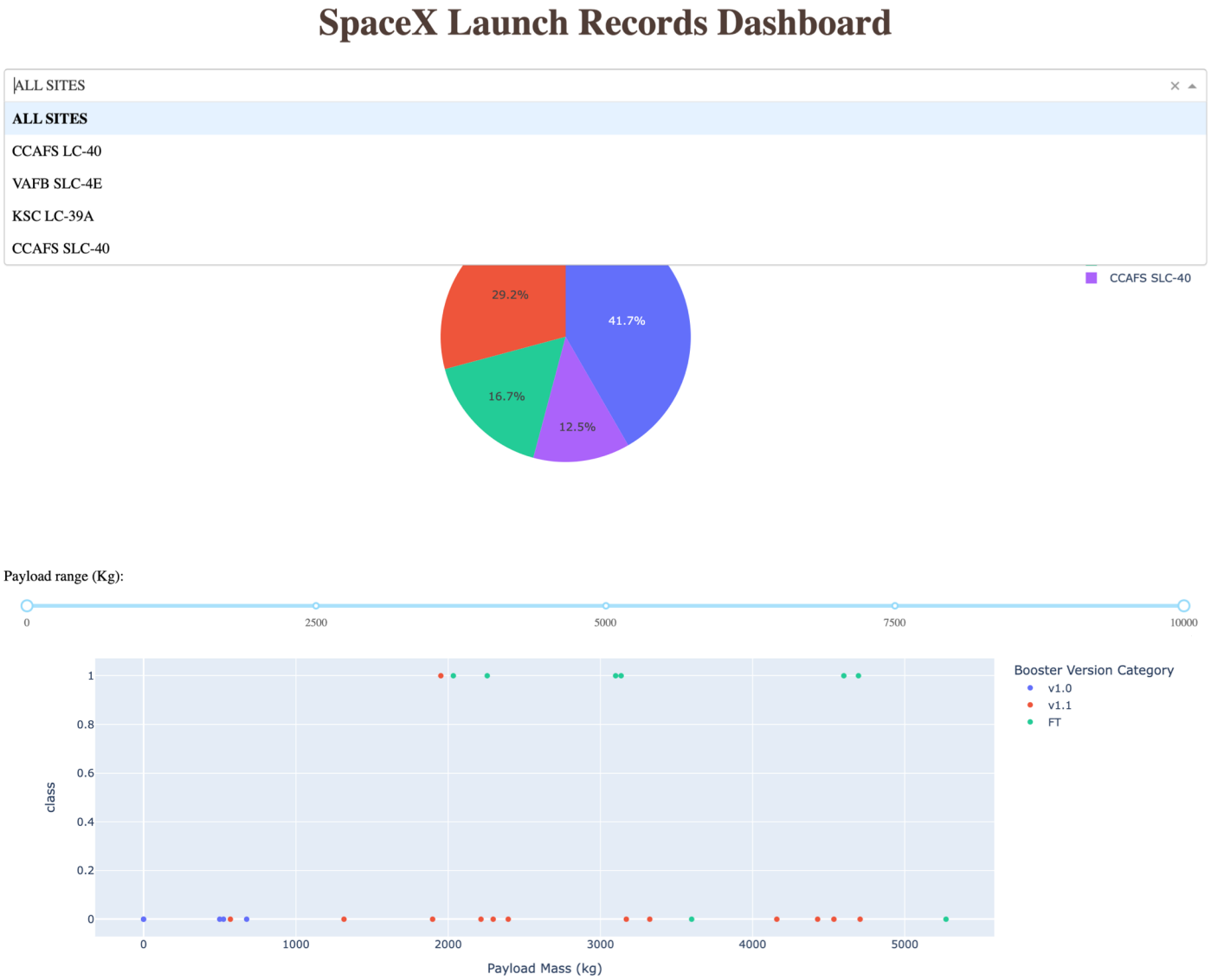
Show the distances between CCAFS SLC 40 and coast, KSC LC 39A and highways. etc.



BUILD A DASHBOARD WITH PLOTLY DASH

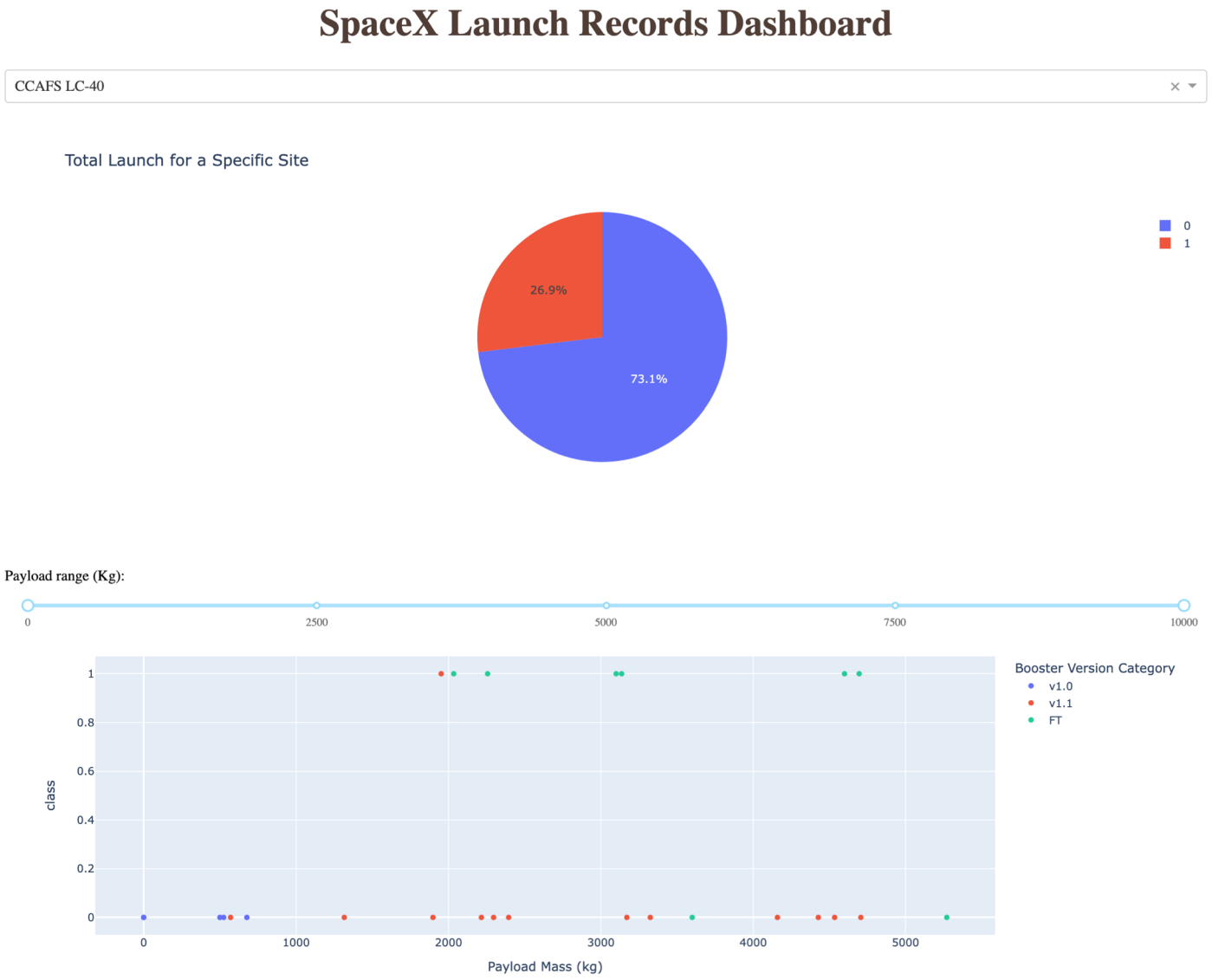


DASHBOARD



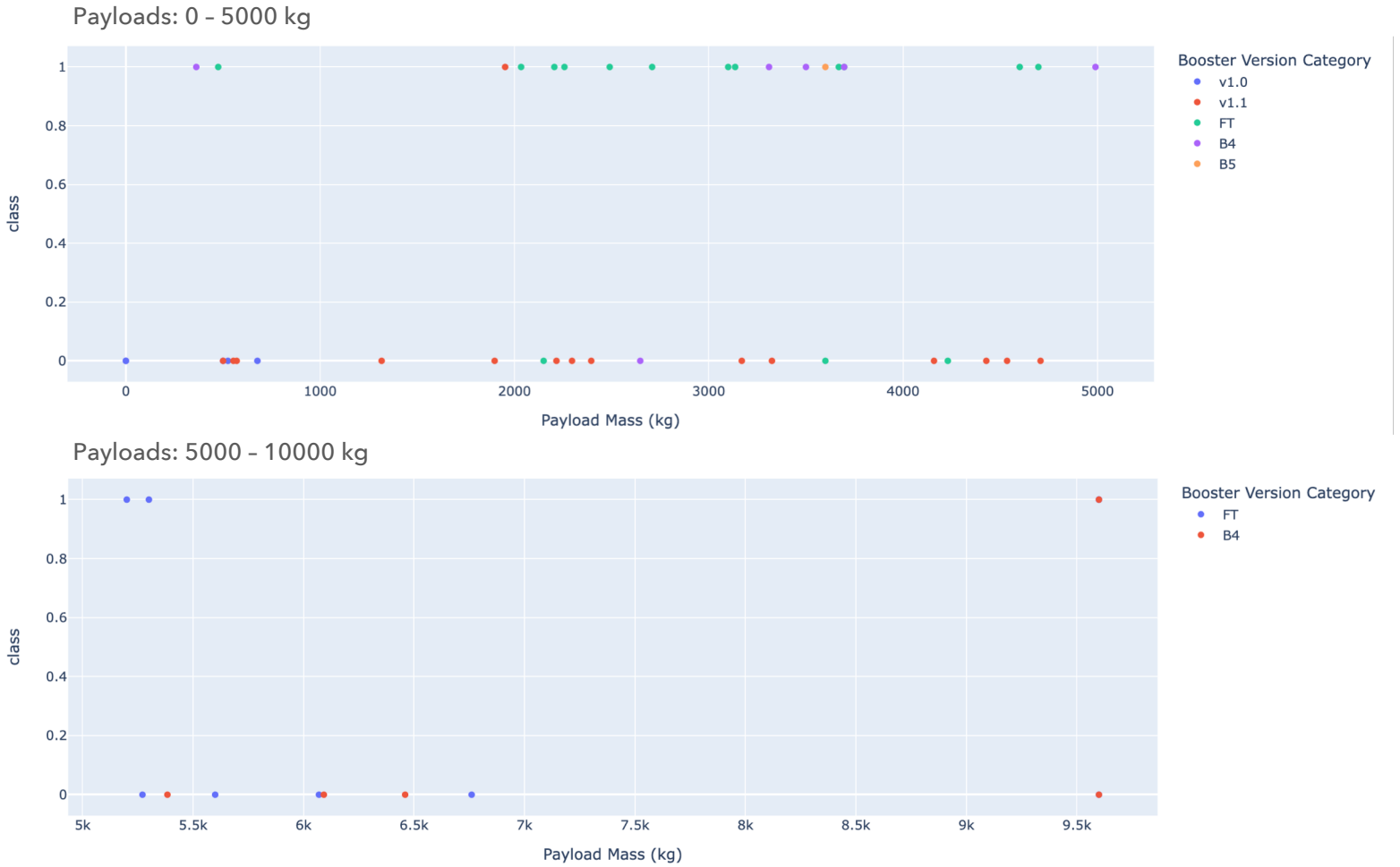
GitHub

DASHBOARD


SpaceX Launch Records Dashboard
CCAFS LC-40
Total Launch for a Specific Site
26.9%
73.1%
0
1
Payload range (Kg):
0
2500
5000
7500
10000
class
Payload Mass (kg)
Booster Version Category
v1.0
v1.1
FT

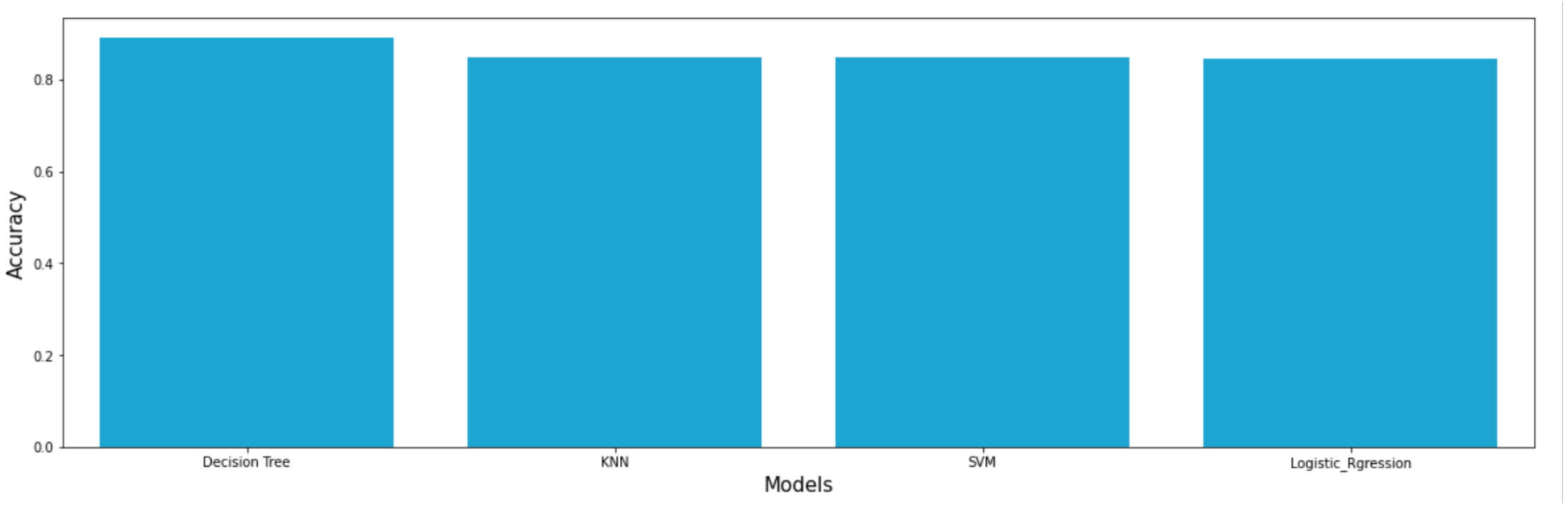
DASHBOARD

The successful rate of low weighted payloads is better than high weighted payloads.



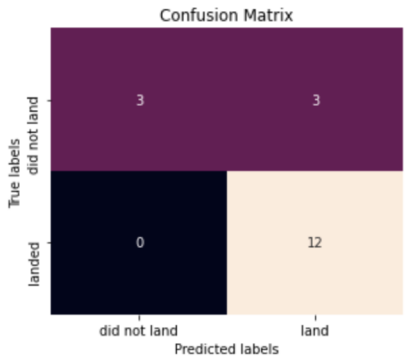
PREDICTIVE ANALYTICS (CLASSIFICATION)

CLASSIFICATION ACCURACY

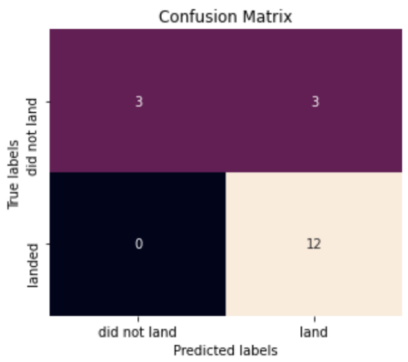


CONFUSION MATRIXS

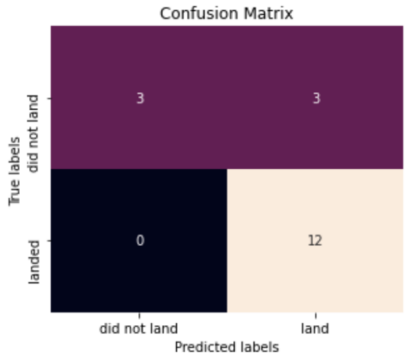
Logistic Regression



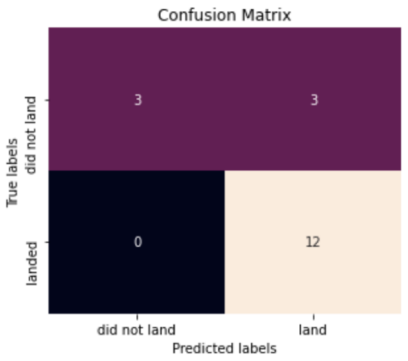
SAM



Decision Tree



KNN



All models' performance are same in this project:

Sensitivity: TP / (TP + FN) = 12 / (12 + 0) = 100%

Specificity: TN / (TN + FP) = 3 / (3 + 3) = 50%

Precision: TP / (TP + FP) = 12 / (12 + 3) = 80%

Accuracy: (TP + TN) / (TP + TN + FP + FN)

= (12 + 3) / (12 + 3 + 3 + 0) = 83.33%

CONCLUSION

Overview the project:

- different launch sites have different success rates. CCAFS LC-40, has a success rate of 60 %, while KSC LC-39A and VAFB SLC 4E has a success rate of 77%.
- the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000).
- 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.
- Orbits ES-L1, DEO, HEO, and SSO had highest successful rate.
- KSC LC-39A had the best successful launches, but with increasing payload mass it seems to have negative impact on success.
- the success rate since 2013 kept increasing till 2020.
- Decision tree model has a best performance.

THANK YOU!