IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

LY THAI BAO

October 31st, 2021

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection
  - Data Wrangling
  - EDA with Data Visualization
  - EDA with SQL
  - Build a Dashboard with Plotly Dash
  - Build an Interactive Map with Folium
  - Predictive Analysis (Classification)
- Summary of all results
  - Predictive analysis results

Introduction

- Project background and context
  - Space X 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 Space X 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 space X for a rocket launch.
- Problems you want to find answers
  - Predict the success of the first stage of Falcon 9
  - Find out which factors affect success rate

Section 1

Methodology

Methodology

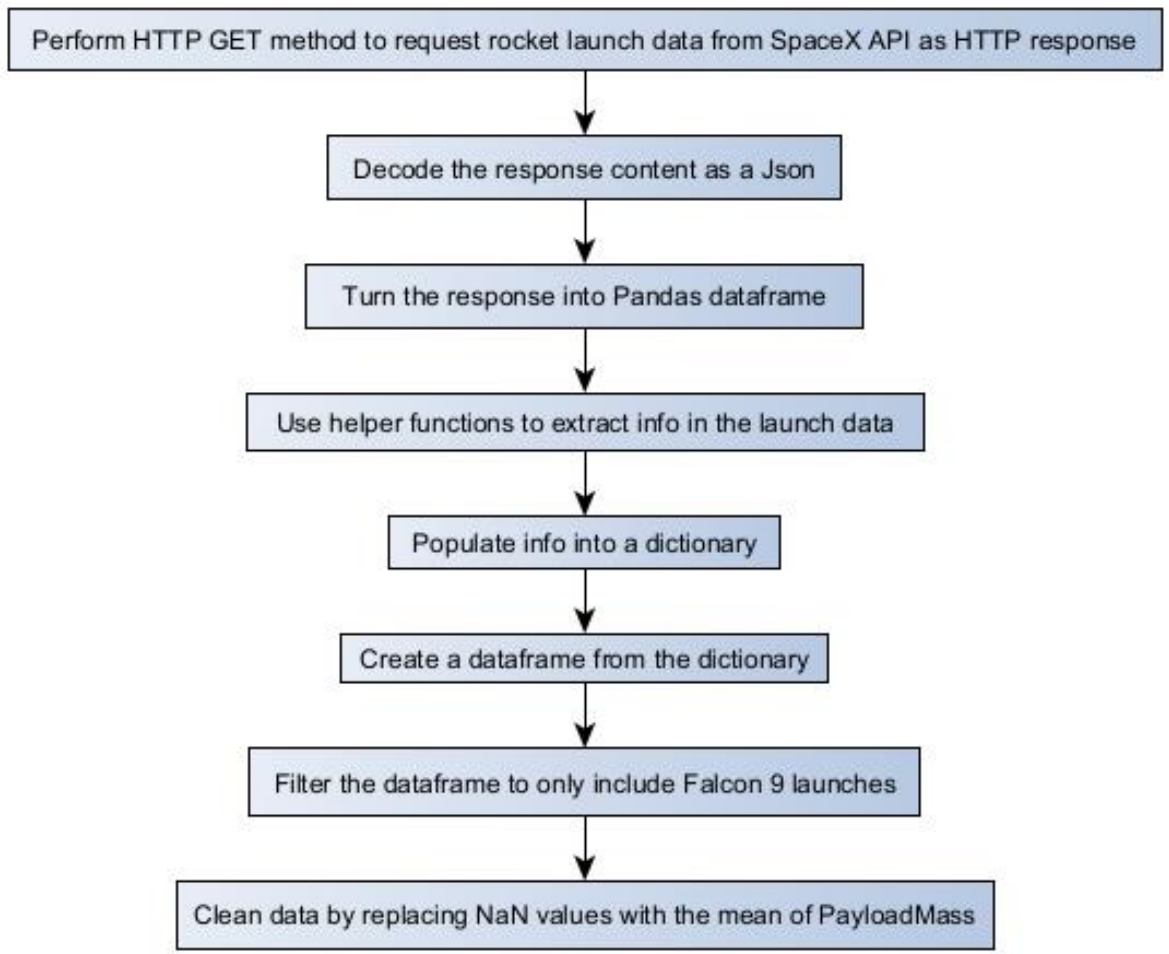
Executive Summary

- Data collection methodology:
  - Data was collected using SpaceX REST API and Web scraping from Wikipedia
- Perform data wrangling
  - Data was cleaned and organized
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - We used machine learning models to find the best model for prediction.

Data Collection – SpaceX API

- Data collection process:

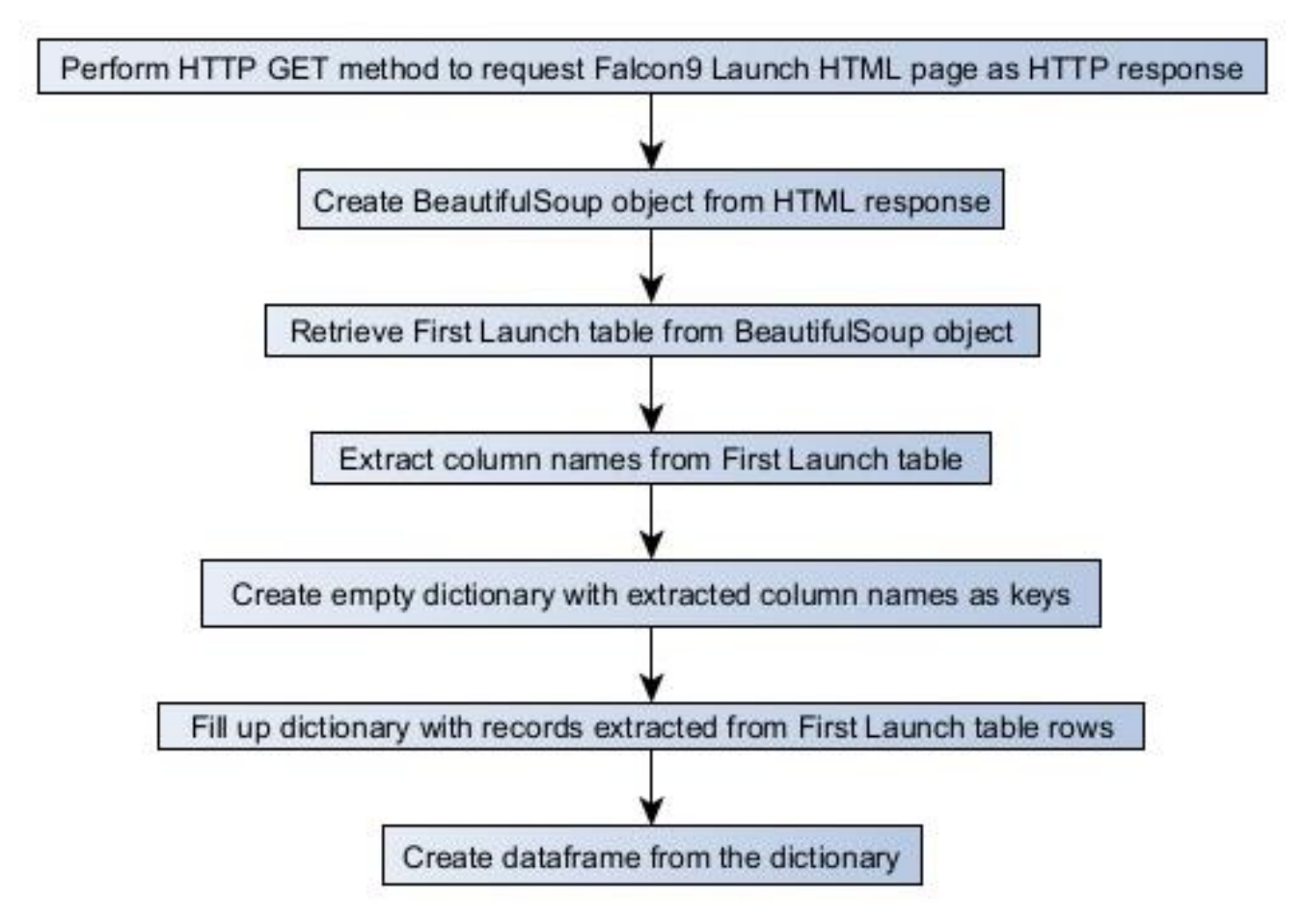
- GitHub URL of my completed SpaceX API calls notebook: https://bit.ly/3k4Tliv



Data Collection - Scraping

- Web scraping process:

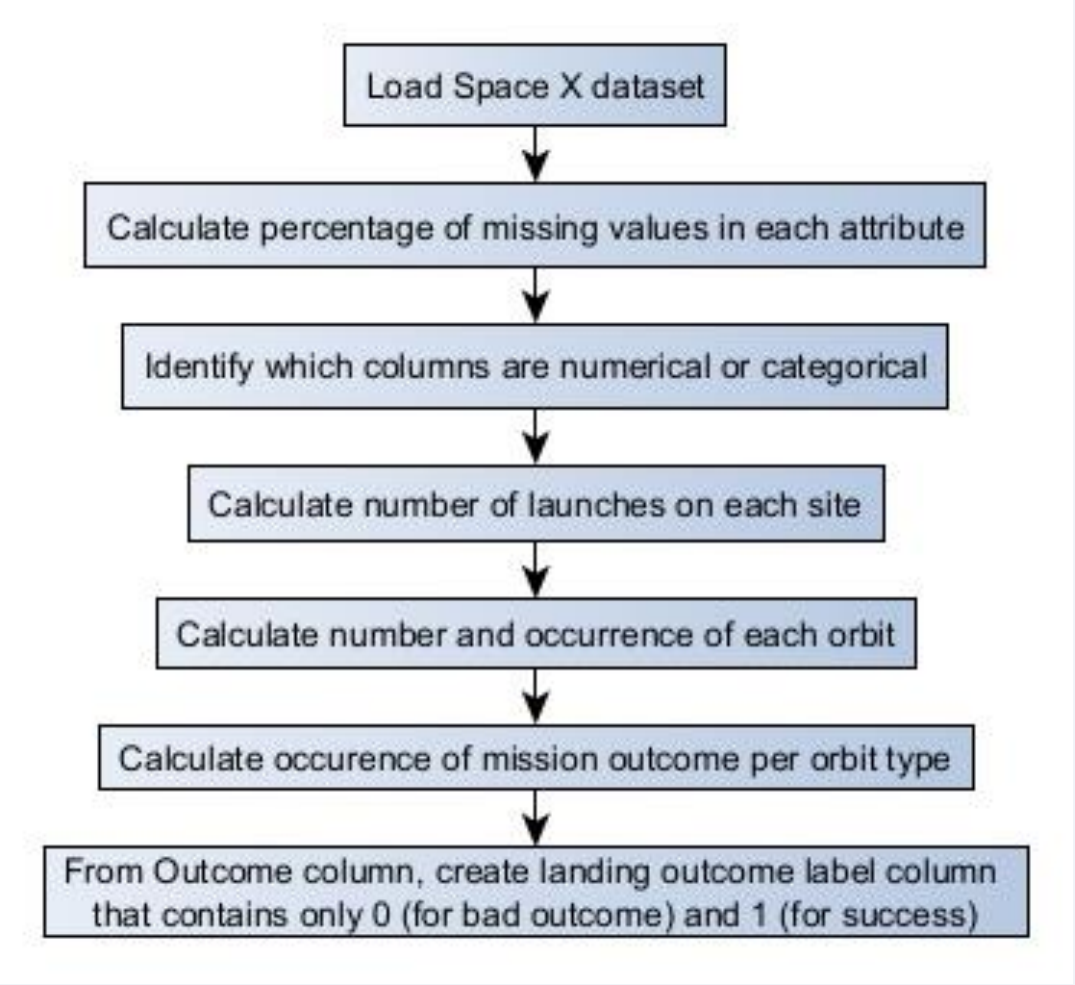
- GitHub URL of my completed web scraping notebook: https://bit.ly/2YdBtzC



Data Wrangling

- Data wrangling process:

- GitHub URL of my data wrangling notebook: https://bit.ly/3CLIgjc



EDA with Data Visualization

What charts were plotted	We can observe from these charts:
scatterplot of Flight Number vs. Payload Mass	- As flight number increases, the first stage is more likely to land successfully.
scatterplot of Flight Number vs Launch Site	- As flight Number increases, there are more successes across all sites. - CCAFS SLC 40 has the most flight numbers, VAFB SLC 4E has the least. - At CCAFS SLC 40, there are 2 distinct groups of flight numbers: less than 30 and more than 40.
scatterplot of Payload Mass vs. Launch Site	- At VAFB SLC 4E, there is no rocket launched with payload mass greater than 10000. - At CCAFS SLC 40, the payload mass of launched rockets can be distinctly grouped into 2 groups: less than 7500 and more than 12500. - Across 3 sites, rockets with payload mass greater than 7500 are more likely to land successfully.
bar chart of Orbits vs. Success Rate	- There are 4 orbits with 1.0 (high) success rate: ES-L1, GEO, HEO, SSO. - SO has the lowest success rate of 0.
scatter point chart of Flight Number vs. Orbit	- 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.
scatter point chart of Payload Mass vs. Orbit	- With heavy payloads the successful landing or positive landing rate are more for Polar, LEO and ISS. This cannot be said for the rest.
line chart of yearly launch success	- The success rate since 2013 kept increasing till 2020.

- GitHub URL of my EDA with data visualization notebook: https://bit.ly/3bEFj7U

EDA with SQL

- Using bullet point format, summarize the SQL queries you performed
  - 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, using a subquery
  - 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, in descending order
- GitHub URL of my EDA with SQL notebook: https://bit.ly/2Yj8OJE

Build an Interactive Map with Folium

Objects I added to a folium map	I added those objects because
a circle with a text label	To highlight NASA Johnson Space Center's coordinate with a circle and a popup showing its name when mouse-hovered .
a marker	To show the label “NASA JSC” in red.
a series of circles	To highlight Space X launch sites’ coordinates, there is a circle and a popup showing the name of each launch site when mouse-hovered.
a series of markers	To show the label Space X launch sites in red.
a MarkerCluster containing many markers having the same coordinate	To show groups of successful and failed launches at each site. If a launch is successful, it will have a green icon, otherwise a red one.
a MousePosition	To show coordinates of any point you hover your mouse over on the map.
a marker with distance	To display the distance between a launch site and the selected coastline point
a PolyLine	To draw a line between a launch site and the selected coastline point
a marker with distance	To display the distance between a launch site and the closest highway.
a PolyLine	To draw a line between a launch site and the closest highway.

GitHub URL of my interactive map with Folium map notebook: https://bit.ly/3k5UCLE
To see interactive maps, please go to: https://bit.ly/3BKLPEP

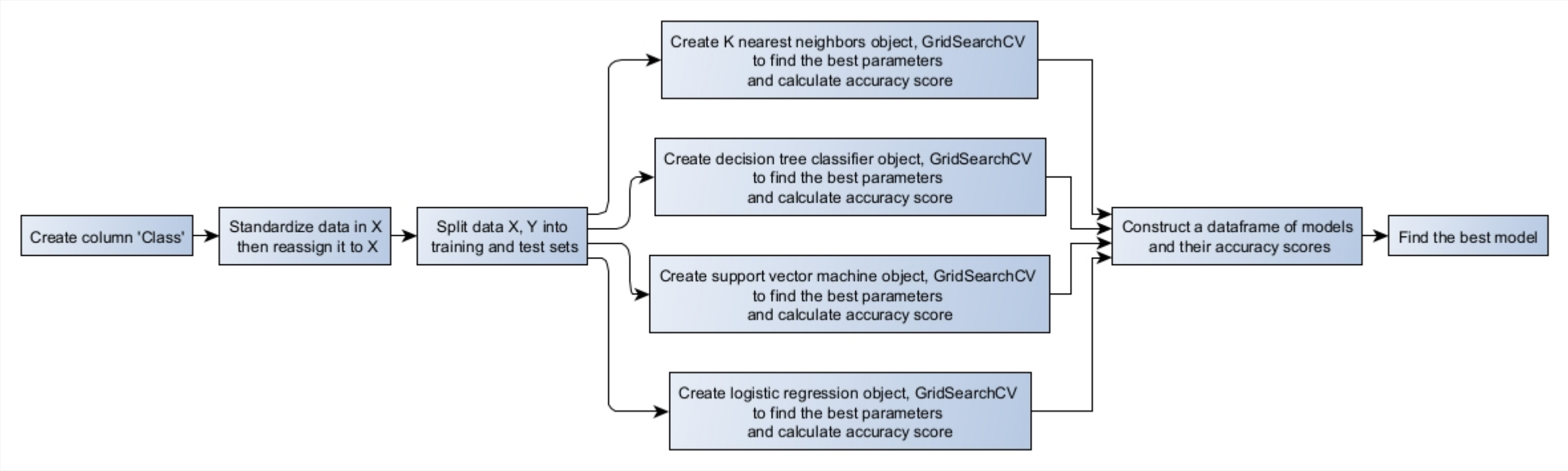
Build a Dashboard with Plotly Dash

- We added:
  - A pie chart to show the launch success rates for each site and all launch sites. From these pie charts, we can conclude KSC LC-39A has the highest launch success rate (76.9%), while CCAFS LC-40 has the lowest launch success rate (26.9%).
  - A scatter chart to show the launch outcome with respect to payload. We can use the payload slider to find out the relationship between payload and launch success rate.

- GitHub URL of my Plotly Dash lab: https://bit.ly/3EVXdjq

Predictive Analysis (Classification)

- GitHub URL of my predictive analysis lab: https://bit.ly/3qhE5YX

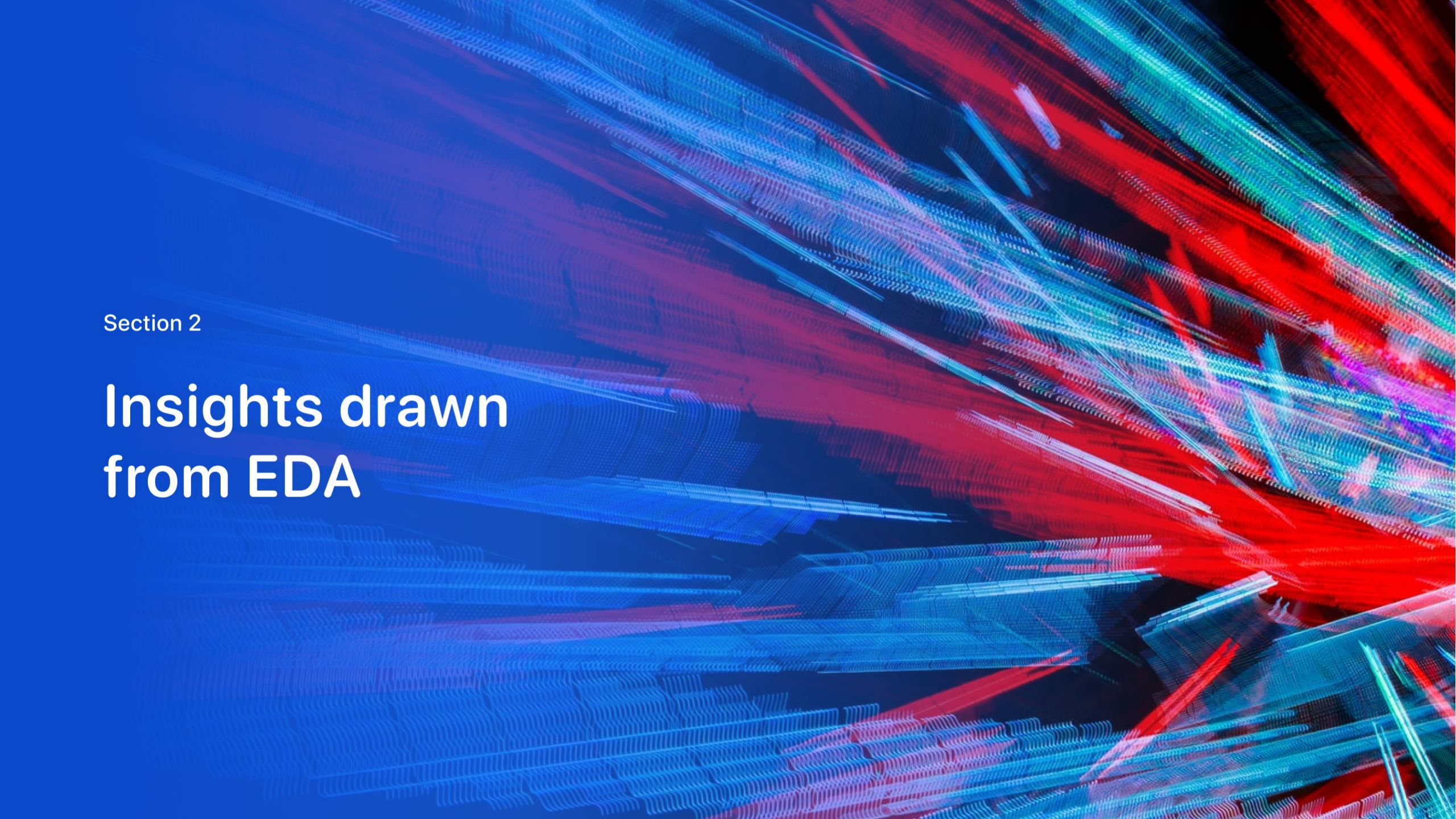


Results

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

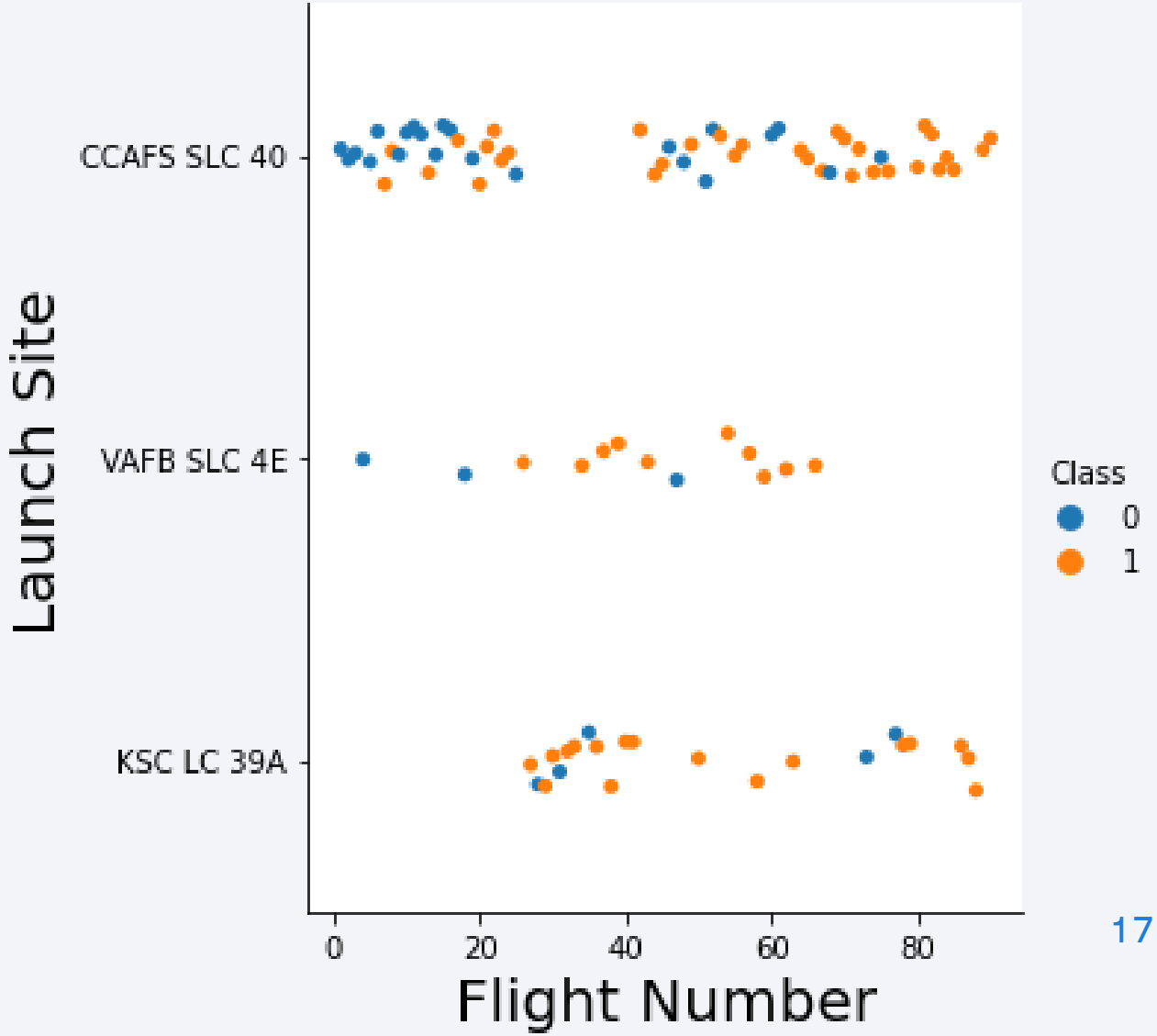
Section 2

Insights drawn from EDA



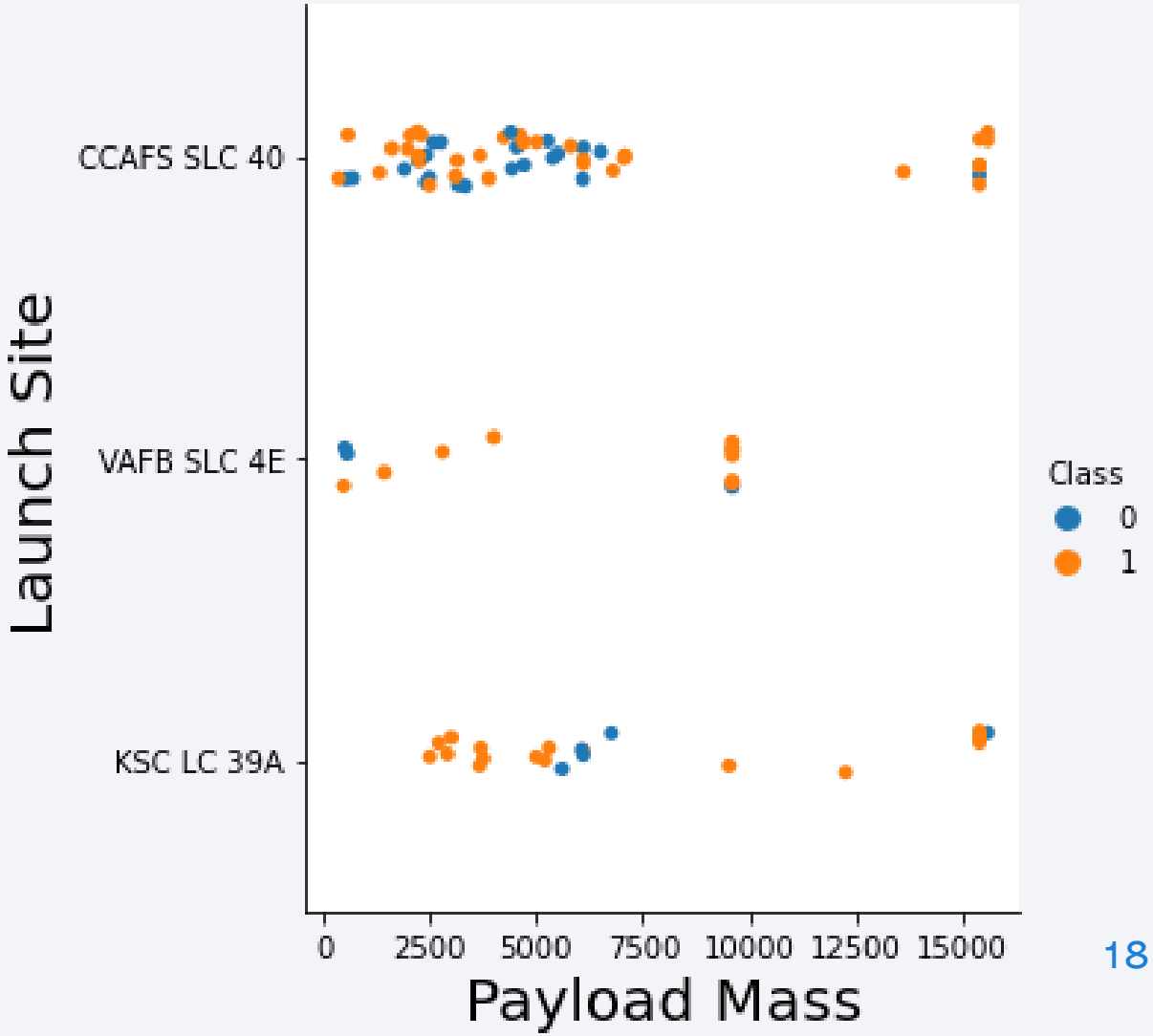
Flight Number vs. Launch Site

- As flight number increases, there are more successes across all sites.
- CCAFS SLC 40 has the most flight numbers, VAFB SLC 4E has the least.
- At CCAFS SLC 40, there are 2 distinct groups of flight numbers: less than 30 and more than 40.



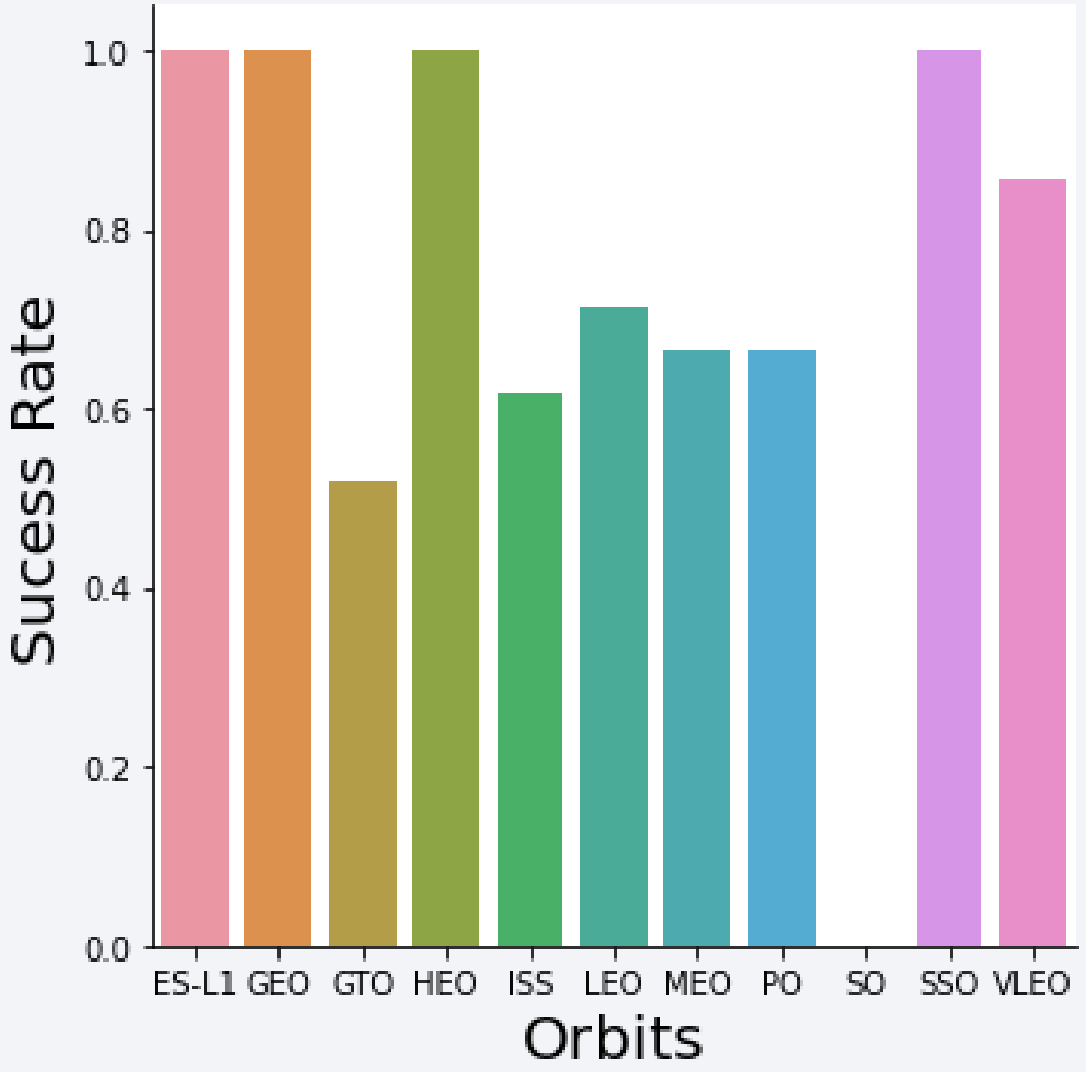
Payload vs. Launch Site

- At VAFB SLC 4E, there is no rocket launched with payload mass greater than 10000.
- At CCAFS SLC 40, the payload mass of launched rockets can be distinctly grouped into 2 groups: less than 7500 and more than 12500.
- Across 3 sites, rockets with payload mass greater than 7500 are more likely to land successfully.



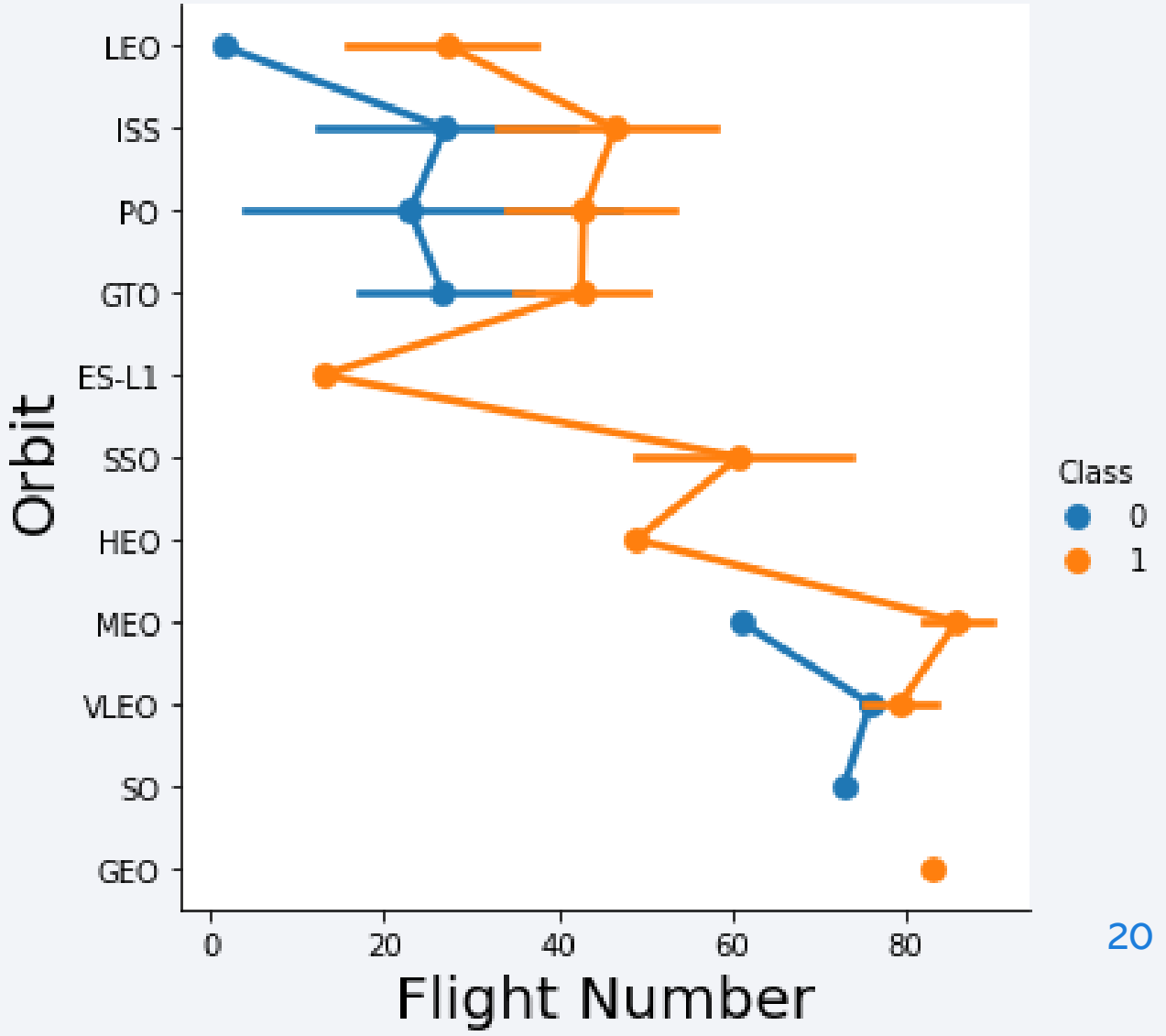
Success Rate vs. Orbit Type

- There are 4 orbits with 1.0 (high) success rate: ES-L1, GEO, HEO, SSO.
- SO has the lowest success rate of 0.



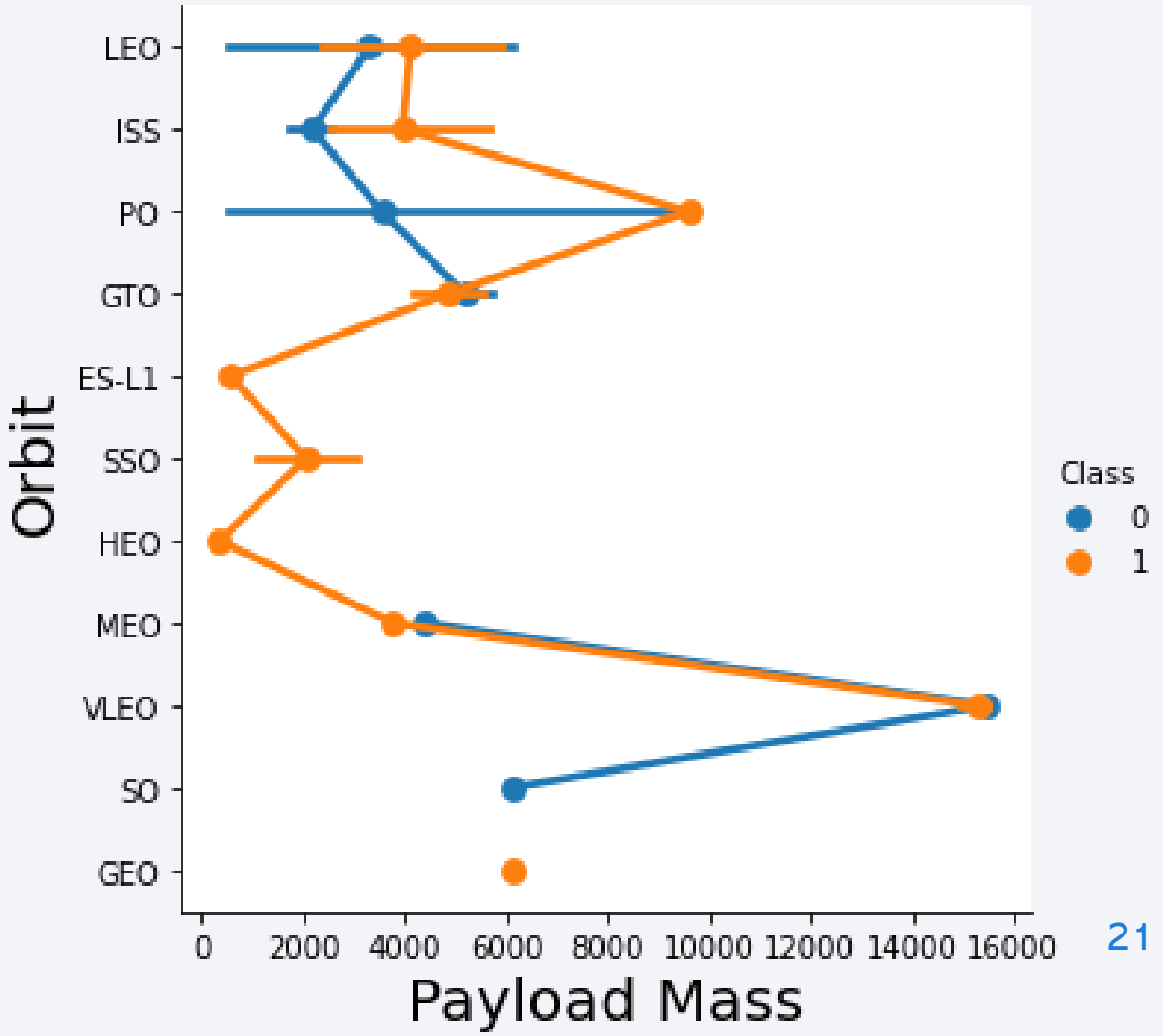
Flight Number vs. Orbit Type

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.



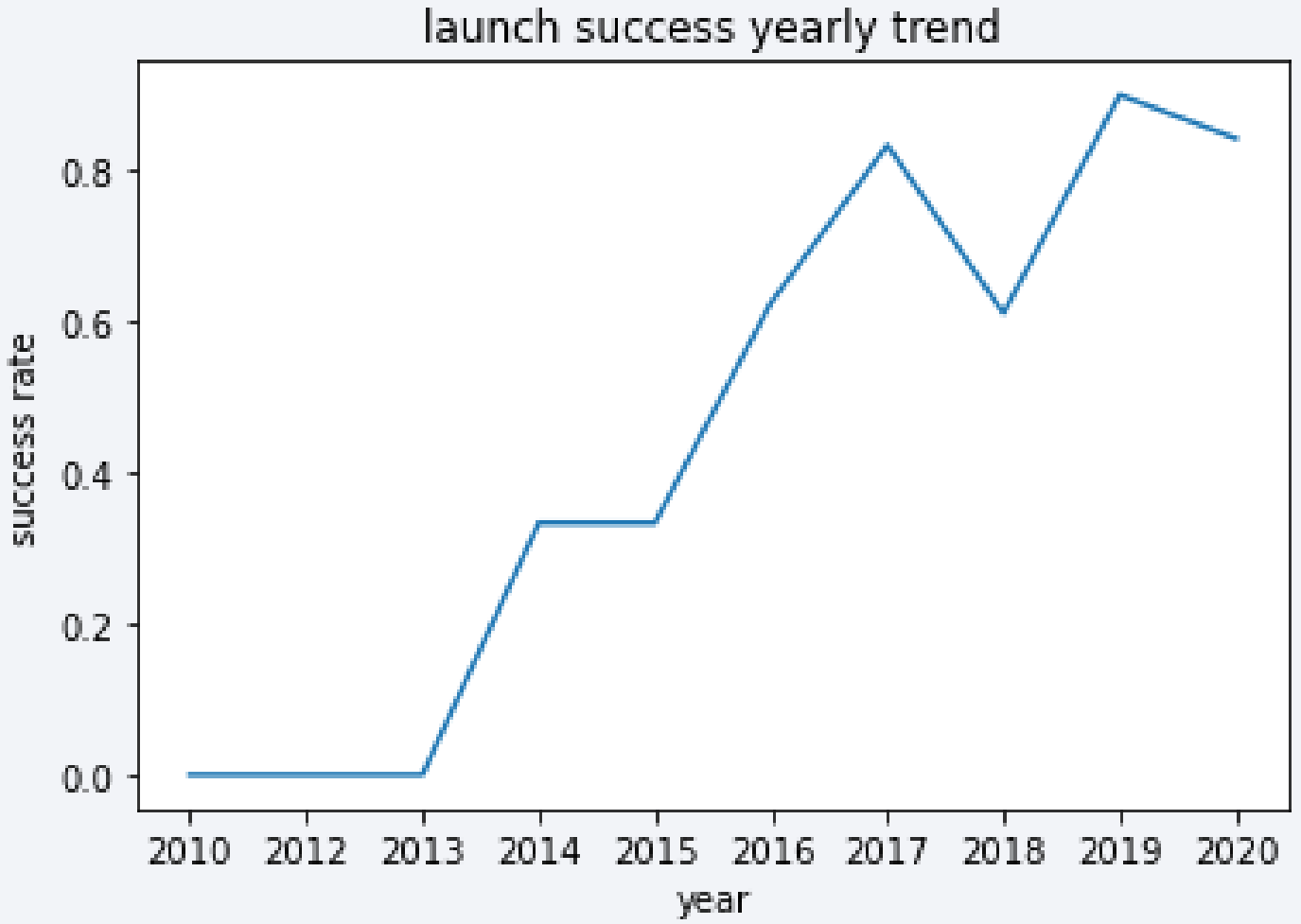
Payload vs. Orbit Type

With heavy payloads, the successful landing or positive landing rate are more for Polar, LEO and ISS. This cannot be said for the rest.



Launch Success Yearly Trend

The success rate since 2013 kept increasing till 2020.



All Launch Site Names

- There are 4 unique launch sites.

```
%sql SELECT UNIQUE launch_site FROM SPACEXTBL
```

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

Total Payload Mass

```
%sql select SUM(payload_mass__kg_) as "total payload mass carried by boosters launched by NASA (CRS)"
from SPACEXTBL where customer = 'NASA (CRS)'
```

total payload mass carried by boosters launched by NASA (CRS)
45596

Launch Site Names Begin with 'CCA'

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

DATE	time__utc_	booster_version	launch_site	payload	payload_mass__kg_	orbit	customer	mission_outcome	landing__outcome
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-12-08	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)
2012-05-22	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Average Payload Mass by F9 v1.1

```
%sql select avg(payload_mass__kg_) as "average payload mass carried by booster version F9 v1.1"
from SPACEXTBL where booster_version = 'F9 v1.1'
```

average payload mass carried by booster version F9 v1.1
2928

First Successful Ground Landing Date

```
%sql select min(DATE) as date_when_the_first_successful_landing_outcome_in_ground_pad_was_achieved
from SPACEXTBL where landing__outcome like '%ground pad%'
```

date_when_the_first_successful_landing_outcome_in_ground_pad_was_achieved
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

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

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

Total Number of Successful and Failure Mission Outcomes

```
%sql select mission_outcome, count(mission_outcome) as number_of_outcomes from SPACEXTBL group by mission_outcome
```

mission_outcome	number_of_outcomes
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Boosters Carried Maximum Payload

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

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

```
%sql select landing__outcome, booster_version, launch_site from SPACEXTBL where EXTRACT(YEAR FROM DATE) = '2015'
```

landing__outcome	booster_version	launch_site
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
Controlled (ocean)	F9 v1.1 B1013	CCAFS LC-40
No attempt	F9 v1.1 B1014	CCAFS LC-40
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40
No attempt	F9 v1.1 B1016	CCAFS LC-40
Precluded (drone ship)	F9 v1.1 B1018	CCAFS LC-40
Success (ground pad)	F9 FT B1019	CCAFS LC-40

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

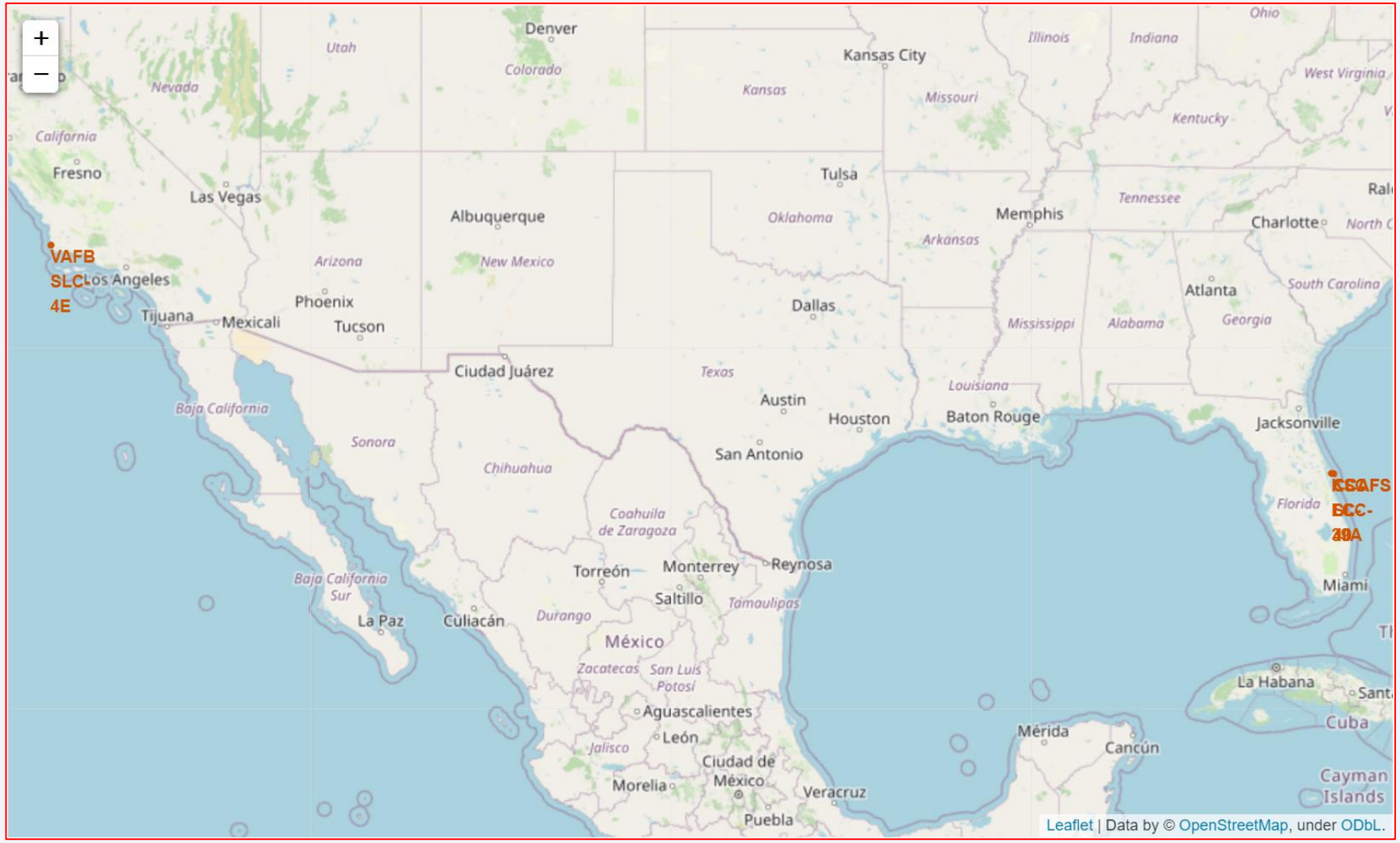
```
%sql select landing__outcome, count(*) count_of_landing_outcomeas, RANK() OVER (order by count(*) desc) as rank
from SPACEXTBL where DATE BETWEEN '2010-06-04' AND '2017-03-20' group by landing__outcome order by count(*) desc
```

landing__outcome	count_of_landing_outcomeas	RANK
No attempt	10	1
Failure (drone ship)	5	2
Success (drone ship)	5	2
Controlled (ocean)	3	4
Success (ground pad)	3	4
Failure (parachute)	2	6
Uncontrolled (ocean)	2	6
Precluded (drone ship)	1	8

Section 4

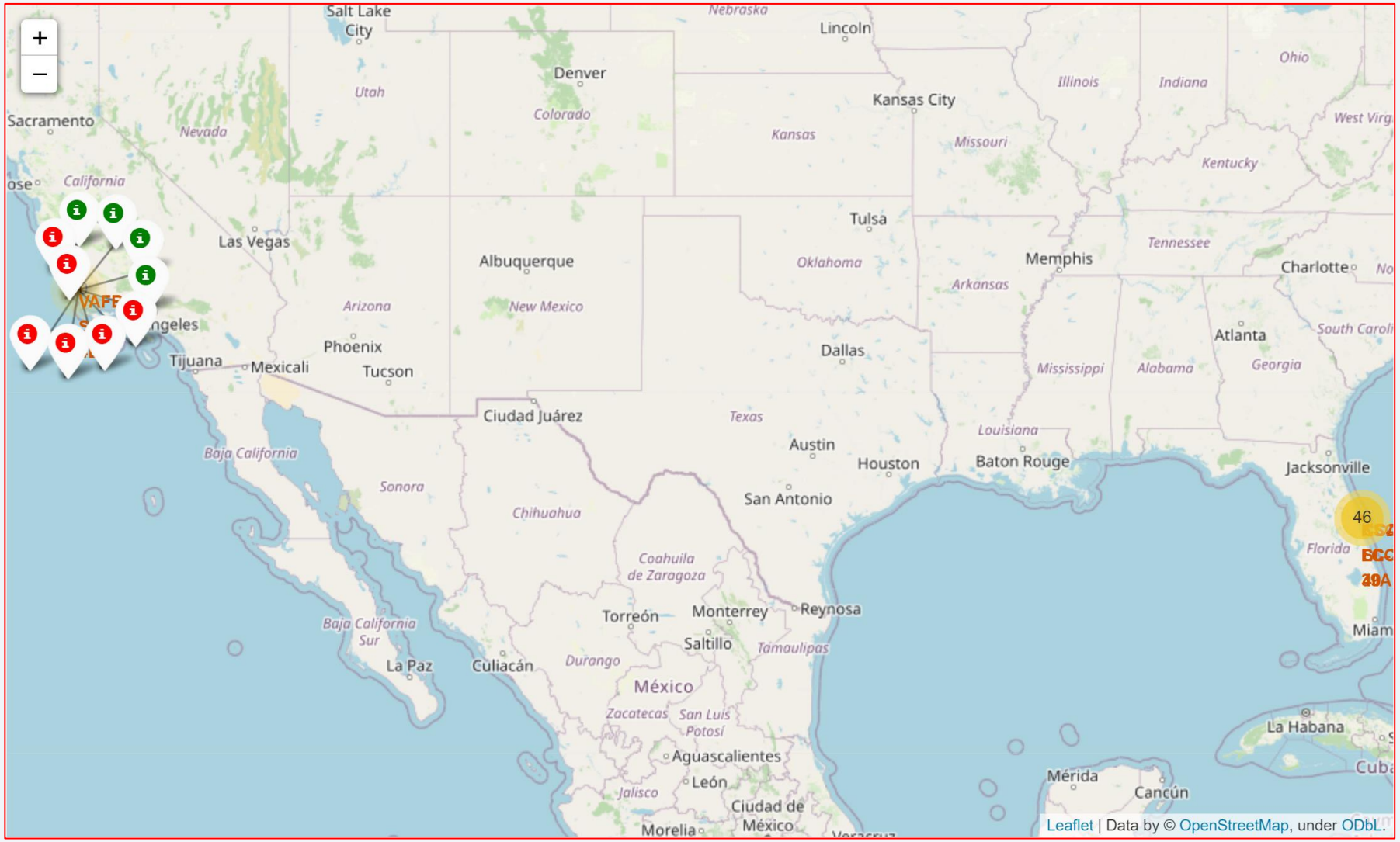
Launch Sites Proximities Analysis

All Launch Sites' Locations



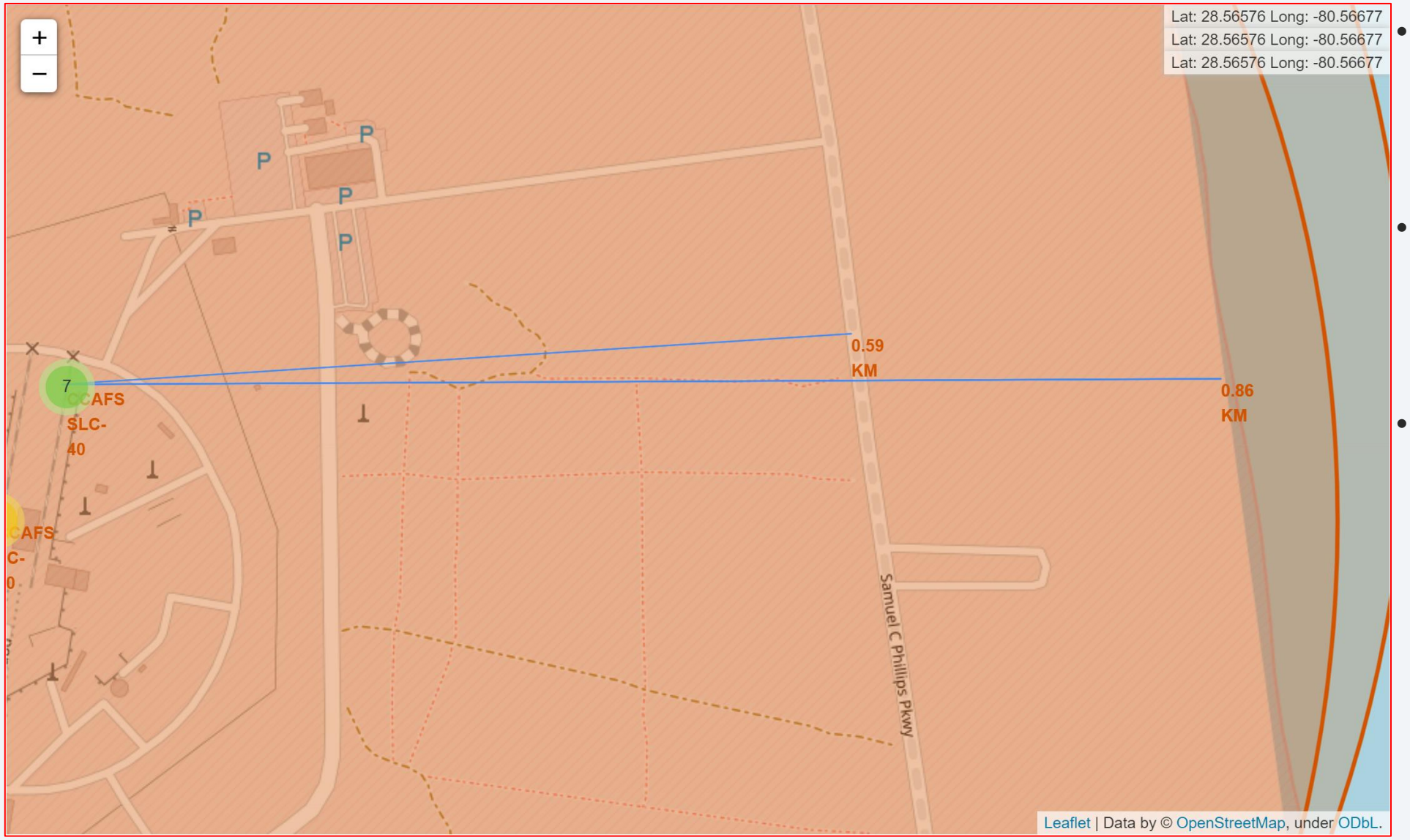
- There are a total of 4 launch sites. Each is marked with a red circle and a red text marker.
- When we click on each circle, we will see a popup showing the launch site name.
- All 4 launch sites are located near coastlines and railways/highways.

Numbers of successful & failed launches at each site



- Besides red circles and red text marker *(explained in the previous slide)*, 4 launch sites will also show the numbers of successful & failed launches when we click on them.
- VAFB SLC-4E, CCAFS LC-40, CCAFS SLC-40 have success rates below 50%.
- KSC LC-39A has success rates greater than 50%.

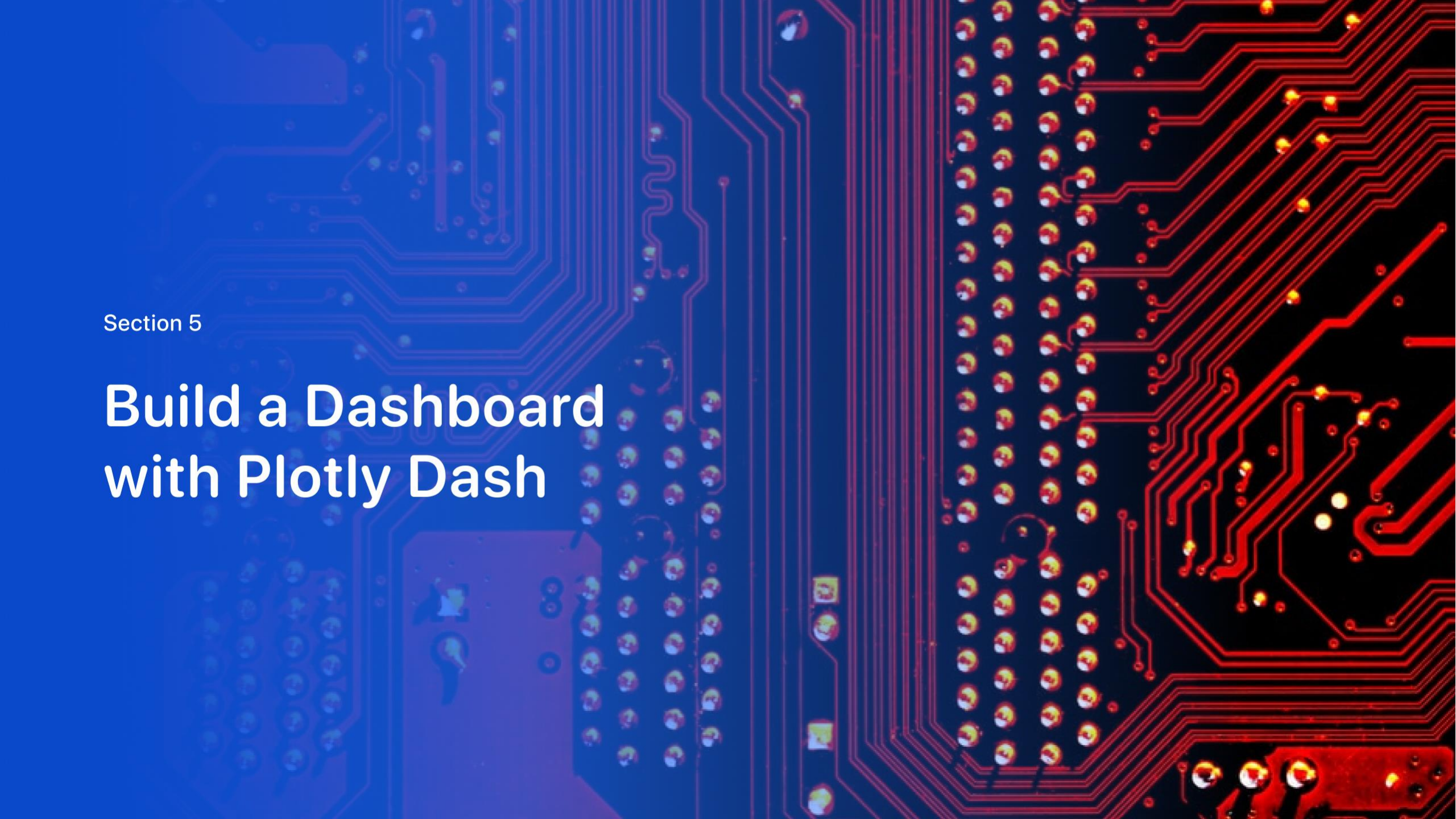
Distances from CCAFS SLC-40 to the closest coastline, highway



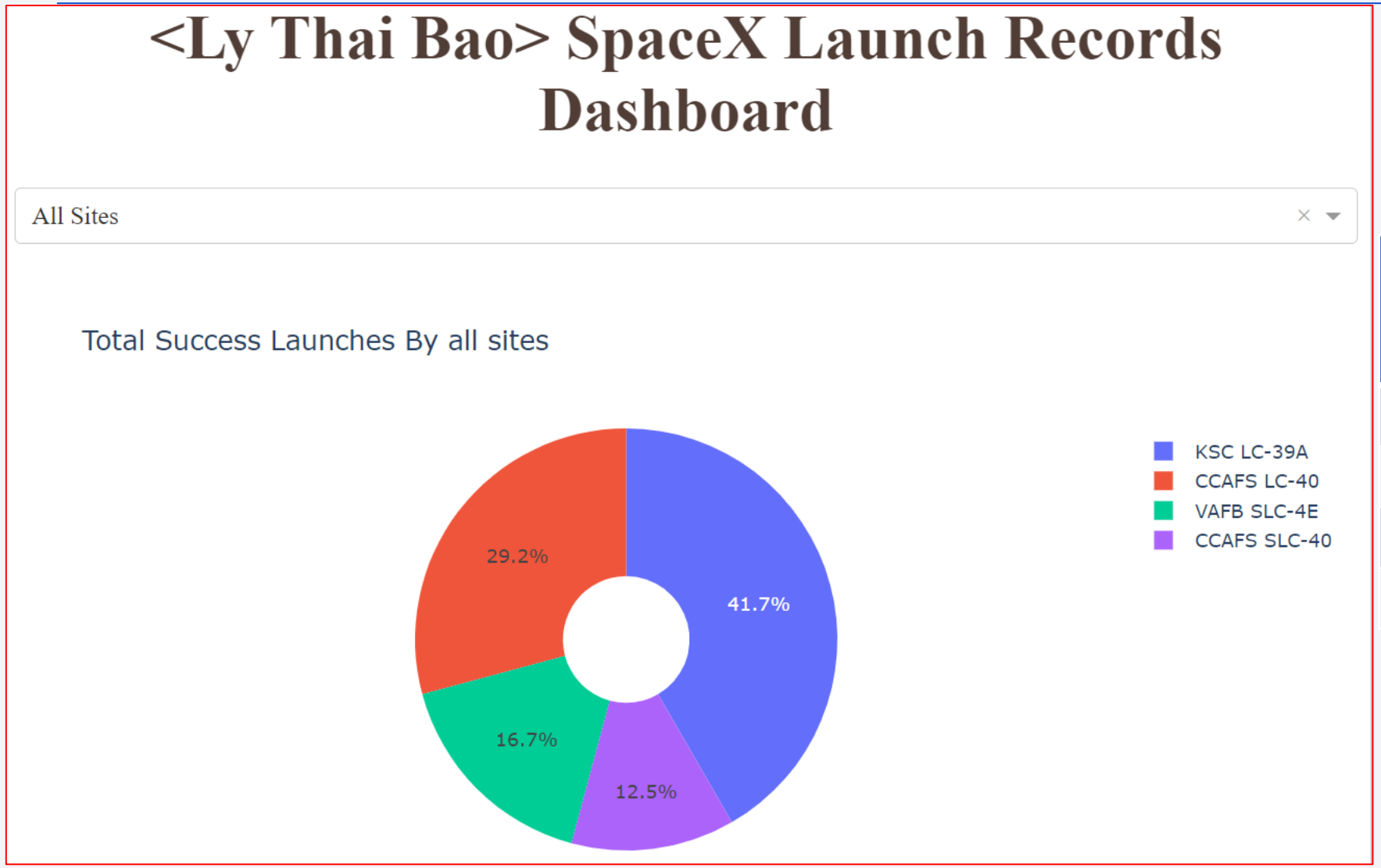
- The map shows 2 blue straight lines and 2 distances in red.
- CCAFS SLC-40 is 0.59 KM away from Samuel C. Phillips Parkway.
- CCAFS SLC-40 is 0.86 KM away from the US Eastern Coastline.

Section 5

Build a Dashboard with Plotly Dash



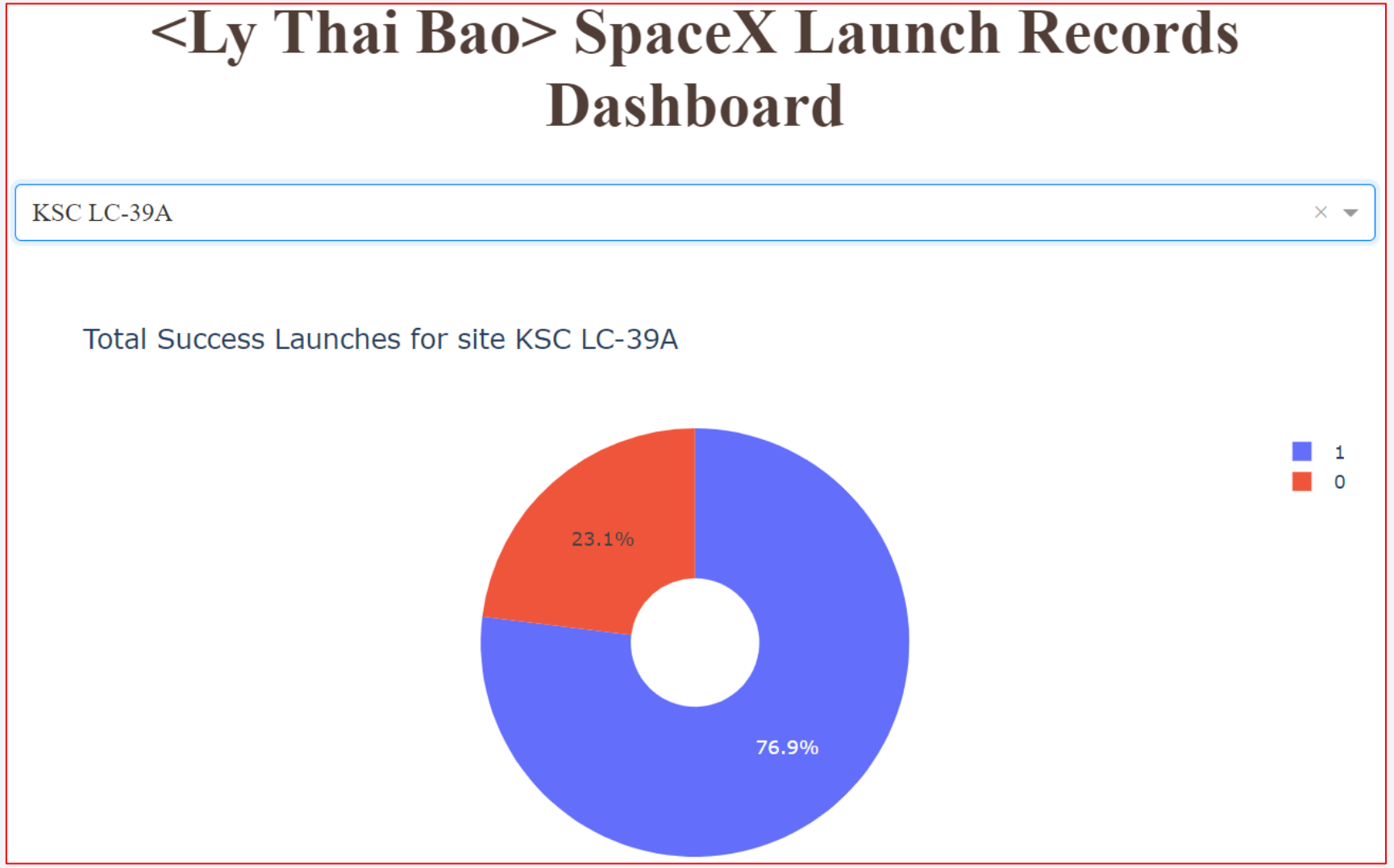
Pie chart of Total launch success rates of all sites



The chart shows

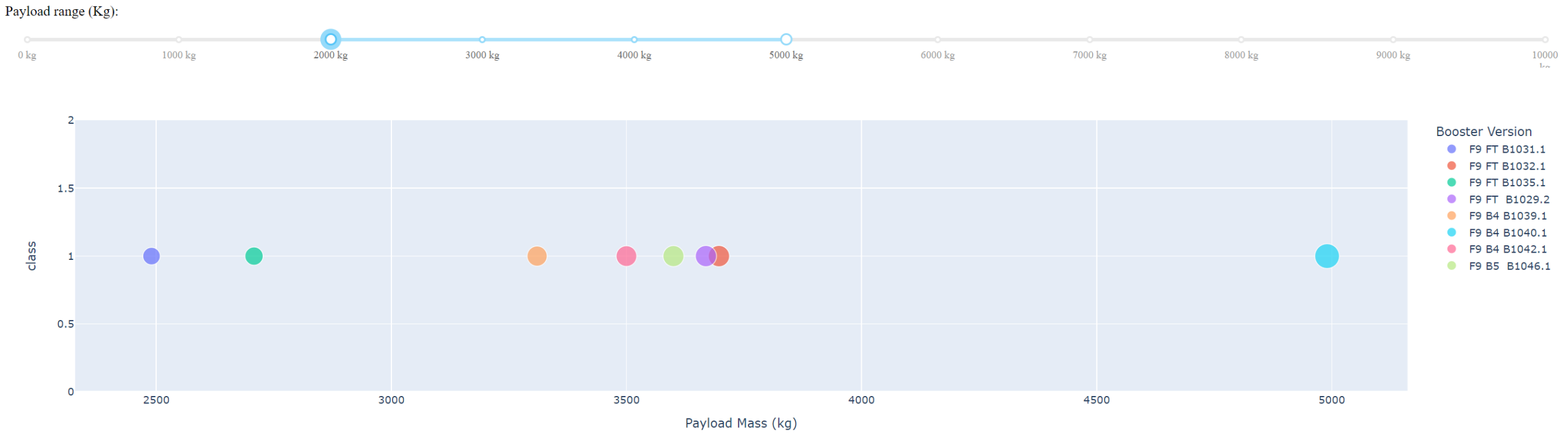
Site	Launch success rate out of total
KSC LC-39A	41.7%
CCAFS LC-40	29.2%
VAFB SLC-4E	16.7%
CCAFS SLC-40	12.5%

Pie chart of the launch site with highest launch success rate



The chart shows KSC LC-39A has the highest launch success rate of 76.9%.

Payload vs. Launch Outcome scatter plot for all sites



Payload range between 2000 and 5000 kg has the highest launch success rate.

Payload range above 6000 kg has the lowest launch success rate.

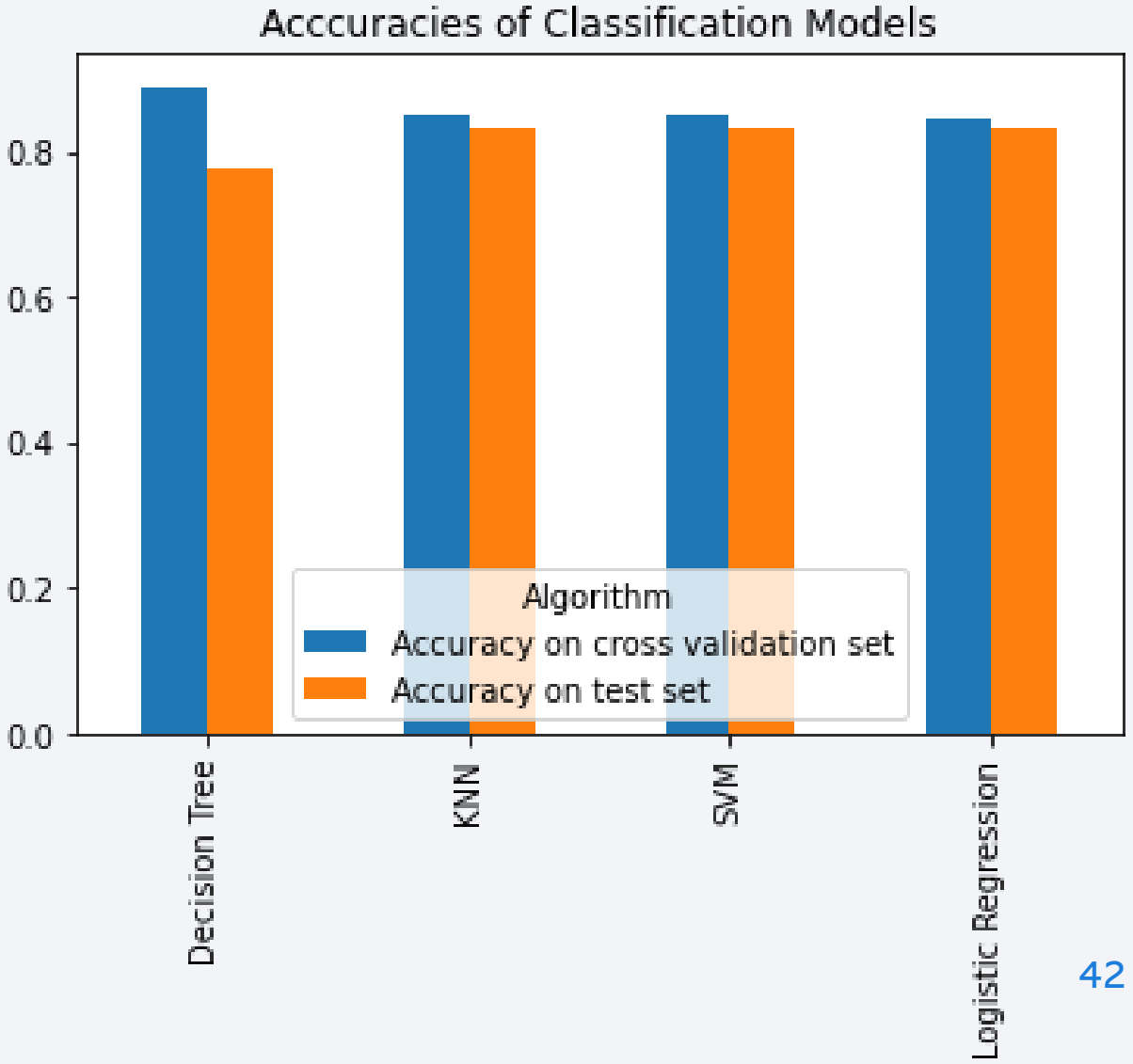
Booster version F9 v1.1 has the highest launch success rate.

Section 6

Predictive Analysis (Classification)

Classification Accuracy

Logistic Regression, Support Vector Machine, K Nearest Neighbors all have the same highest accuracy of 83.3% on the test set, likely due to the small data size.



Confusion Matrix

The confusion matrix of the best performing model shows that there are still many false positives.

Conclusions

- Launch outcome are likely affected by site, payload mass, orbit type and booster version.
- Space X likely chose launch sites near coastlines.
- Having applied four machine learning models on the data set, we can predict with 83.3% accuracy whether or not the first stage of Falcon 9 will have a successful landing.

Thank you!