Winning Space Race with Data Science

Miao Shi

05/13/2022

Outline

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

Executive Summary

- Summary of methodologies
- Summary of all results

Introduction

- Utilize available data from Space X, learn and build basis for creating a competitor company ”Space Y”, primarily to answer the following questions:
  - Where to build a launch site?
  - What payload, orbit, booster versions will have a higher success rate?
  - What options should we should for landing?
  - What predictive model should be used?

Section 1

Methodology

Methodology

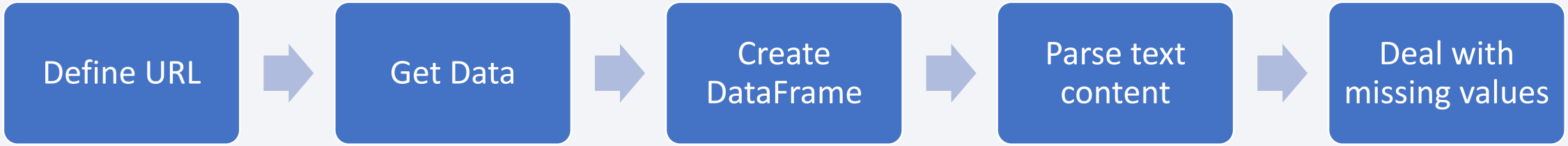
Executive Summary

- Data collection methodology:
  - Describe how data was collected
- Perform data wrangling
  - Describe how data was processed
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection

- Data is collected from API and Web Scraping

Data Collection – SpaceX API



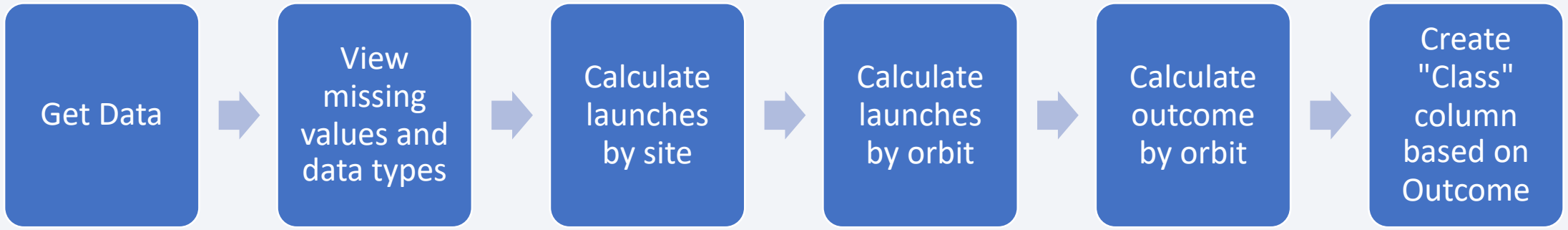
View On GitHub

Data Collection - Scraping

View On GitHub

Data Wrangling

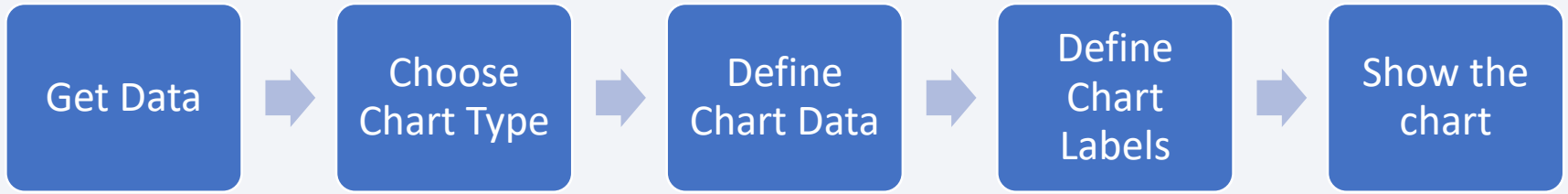
- Performed exploratory data analysis, by summarizing launch outcomes by Site, Orbit



View On GitHub

EDA with Data Visualization

- Used scatter plot to explore the relationship between Flight Number, Payload Mass, Launch Site, Orbit
- Used bar chart to explore the relationship between Orbit and Class
- Used line chart to explore the relationship between Date and Class



View On GitHub

EDA with SQL

- Get unique launch sites
- Display 5 records where launch site starts with ‘CCA’
- Display the total payload mass carried by boosters launched by NASA (CRS)
- Display average payload mass carried by booster version F9 v1.1
- ist the date when the first successful 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 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

View On GitHub

Build an Interactive Map with Folium

- Added Circle and ClusterMarker for launch sites
- Added colored successful and unsuccessful markers to compare success rate for each site
- Added Polylines to identify the distance between launch site and nearest coastline, highway, rail and city

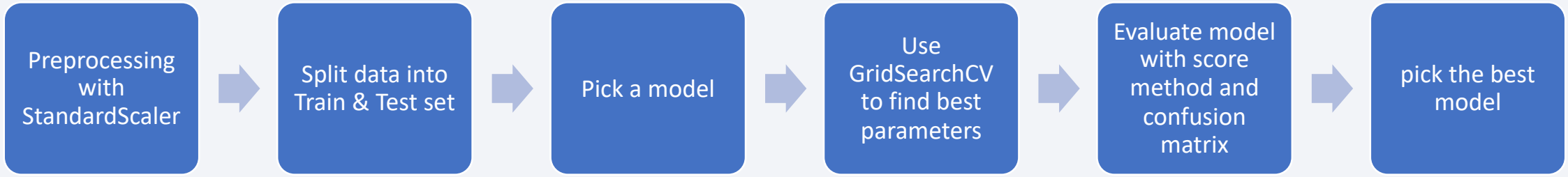
View On GitHub

Build a Dashboard with Plotly Dash

- Added Site dropdown to allow use easily filter result by site
- Added pie chart to show % of successful launches by site
- Added RangeSlider to filter result by payload
- Added scatter plot to show the success rate by Site, Payload, Booster Version

View On GitHub

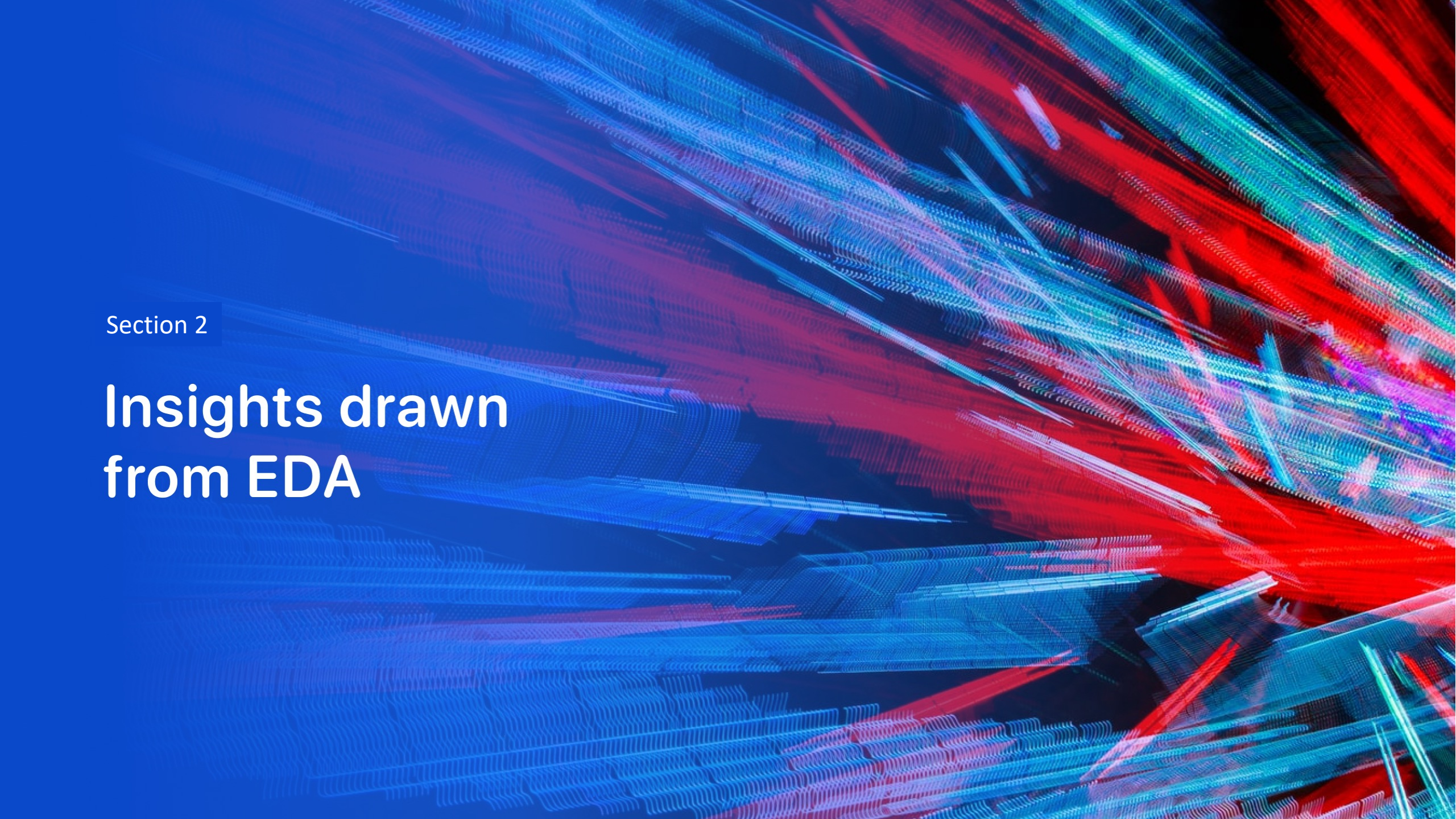
Predictive Analysis (Classification)



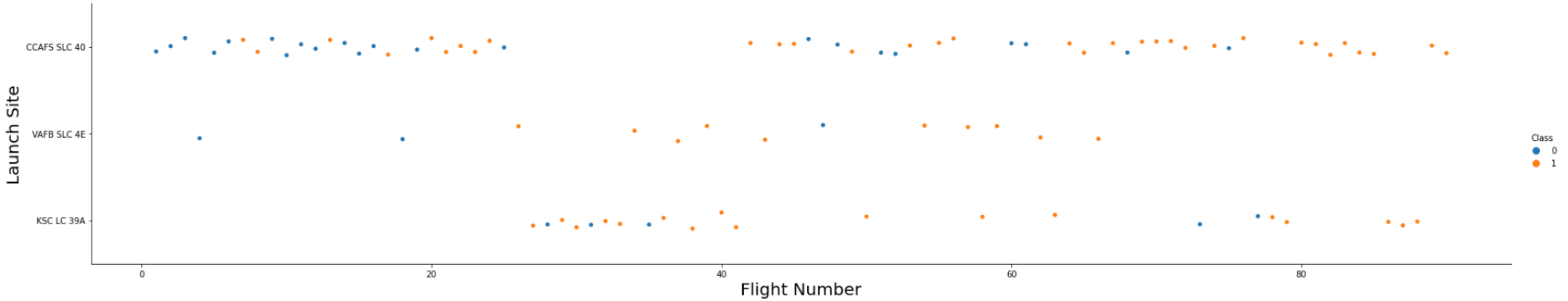
View On GitHub

Results

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

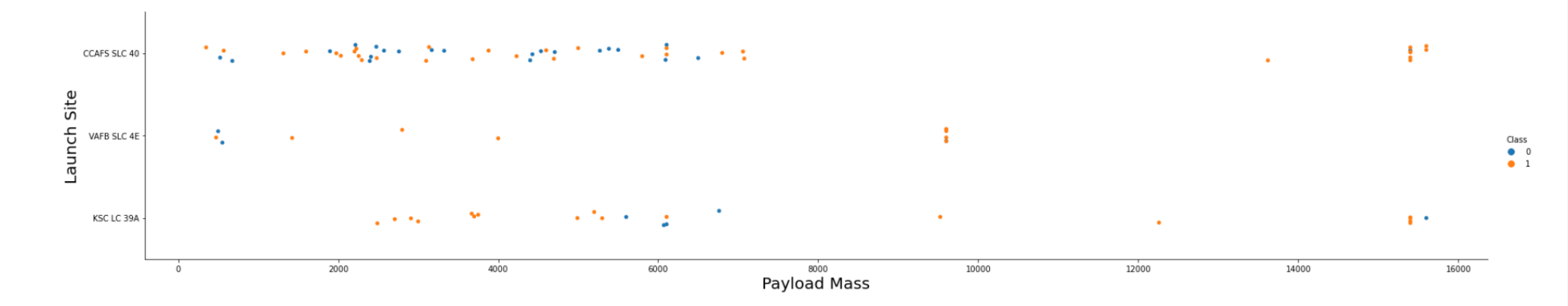

Section 2
Insights drawn from EDA

Flight Number vs. Launch Site



- Initial FlightNumber started at CCAFS SLC 40, then the flights shifted to KSC LC39A, then shifted back to CCAFS SLC 40
- Scattered flights from VAFB SLC 4E, and no activities for more recent flights

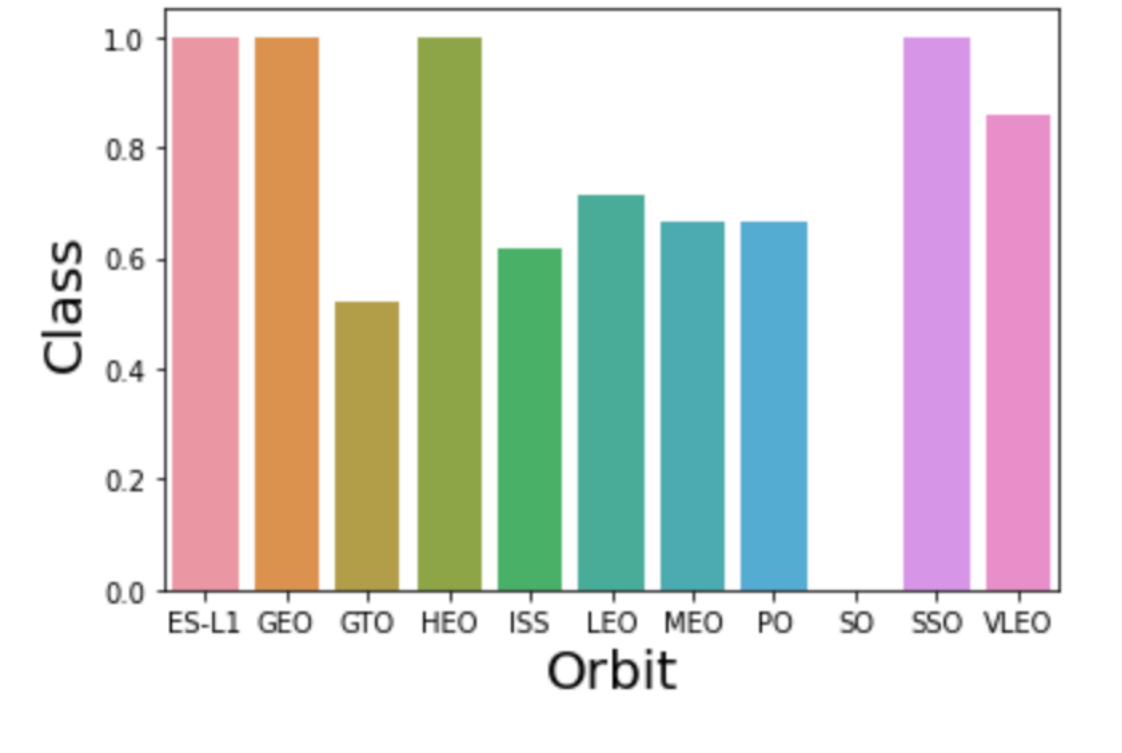
Payload vs. Launch Site



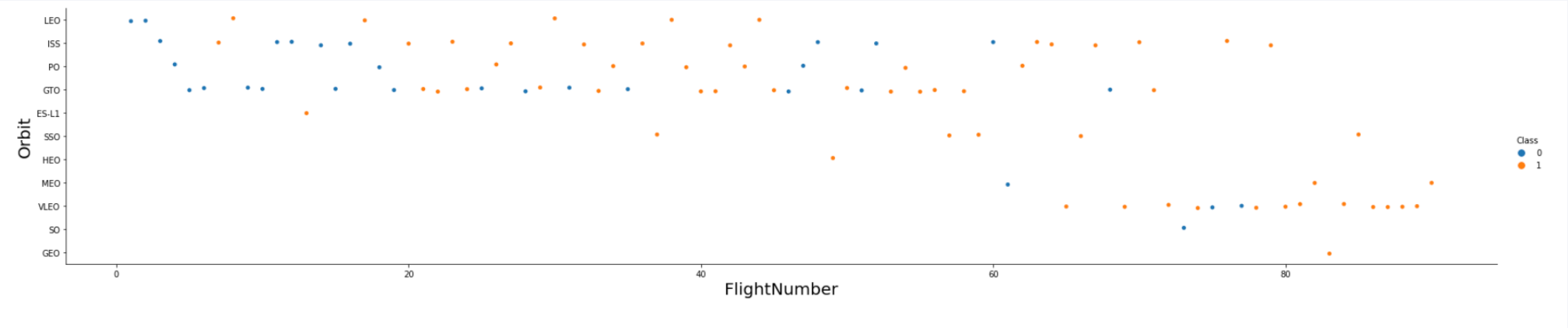
- CCAFS SLC 40 is used for lower and extra high payload mass
- VAFB SLC 4E is used for lower and mid-range payload mass
- KSC LC 39A has not been used for very low payload mass launches

Success Rate vs. Orbit Type

- EL-L1, GEO, HEO, SSO, VLEO are the orbits that have higher success rate
- GTO, ISS, LEO, MEO, PO and SO have lower success rate

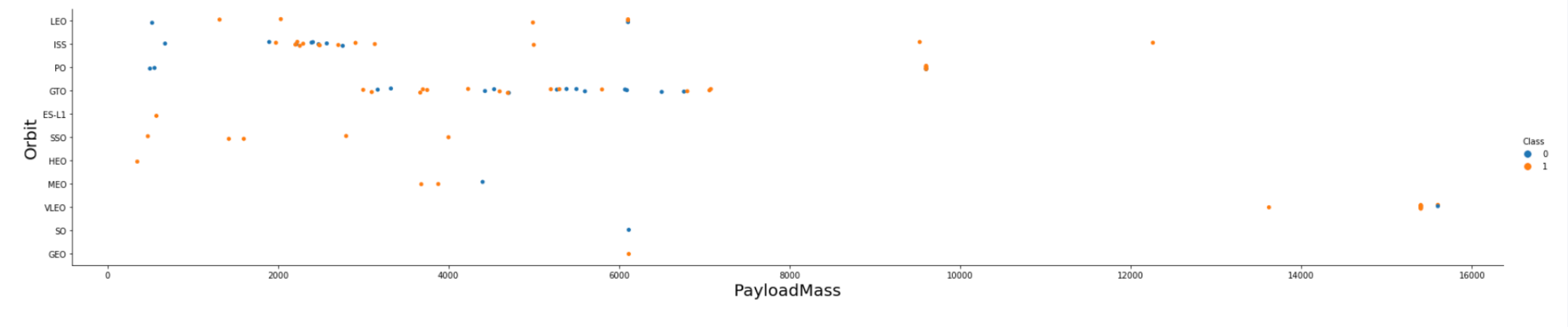


Flight Number vs. Orbit Type



- Recent flights re being shifted from LEO, ISS, PO, GRO, ES-L1, SSO to HEO, MEO, VLEO, SO and GEO

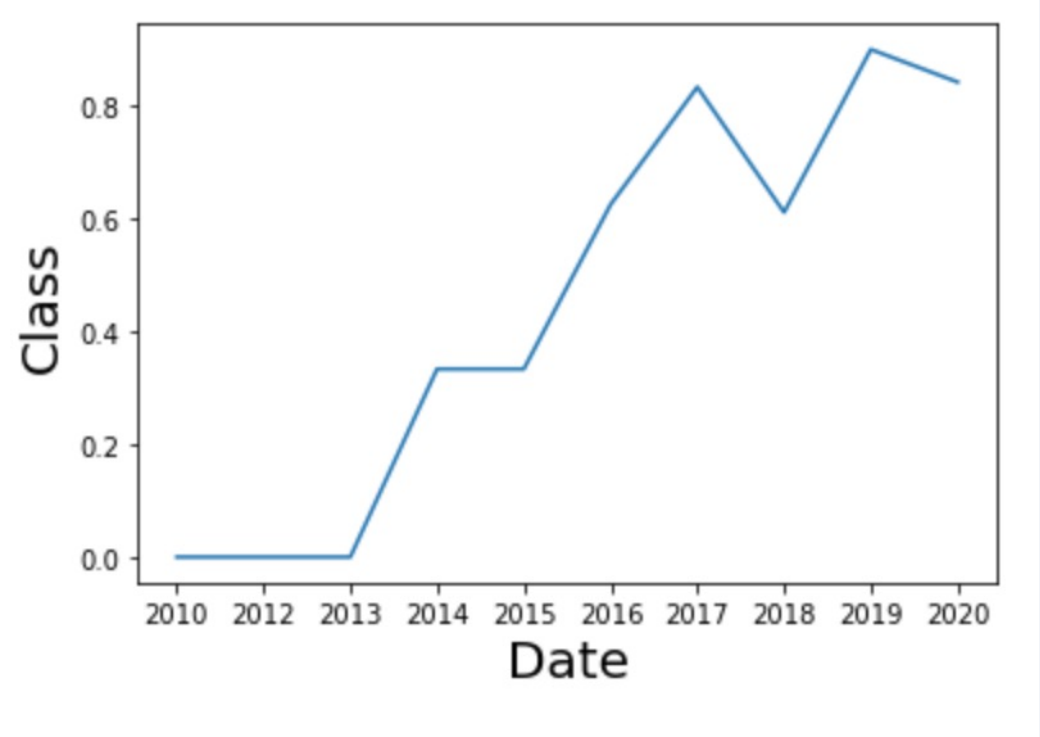
Payload vs. Orbit Type



- Lower payload mass launches use ISS orbit more frequently
- Mid-lower payload mass launches use GTO orbit more frequently
- High payload launches use VLEO orbit

Launch Success Yearly Trend

- Success rate has been increasing since 2013, with the exception of 2018 and 2020



All Launch Site Names

- Total of 4 unique launch sites

```
%sql select distinct launch_site from SPACEXDATASET
```

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

Launch Site Names Begin with 'CCA'

- Match launch_site using “like” statement and limit the result to top 5

```
%sql select * from SPACEXDATASET 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

Total Payload Mass

- Total Payload Mass launched from NASA (CRS) is 45596 KG

```
%sql select sum(PAYLOAD_MASS__KG_) FROM SPACEXDATASET WHERE CUSTOMER='NASA (CRS)'
```

```
 * ibm_db_sa://zgg83101:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:3119
8/bludb
Done.
```

1
45596

Average Payload Mass by F9 v1.1

- The average Payload Mass launched with Booster Version F9 v1.1 is 2928 KG

```
%sql select AVG(PAYLOAD_MASS__KG_) FROM SPACEXDATASET WHERE BOOSTER_VERSION='F9 v1.1'
```

```
 * ibm_db_sa://zgg83101:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:3119
8/bludb
Done.
```

1
2928

First Successful Ground Landing Date

- 12/22/2015 is the date for the first successful launch on ground pad

```
%sql SELECT MIN(DATE) FROM SPACEXDATASET WHERE LANDING__OUTCOME='Success (ground pad)'
```

```
 * ibm_db_sa://zgg83101:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:3119
8/bludb
Done.
```

1
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- The Booster Versions that have successful drone ship landing with Payload between 4000 and 6000 are: F9 FT B1021.2, F9 FT B1031.2, F9 FT B1022 and F9 FT B1026

```
%sql select distinct booster_version from spacexdataset where landing__outcome='Success (drone ship)' and payload_mass_
```

```
 * ibm_db_sa://zgg83101:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:3119
8/bludb
Done.
```

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

Total Number of Successful and Failure Mission Outcomes

- There are total of 1 failed missiong outcome and 100 successful missiong outcomes, including 1 payload status unclear.

```
%sql select Mission_outcome, count(*) from spacexdataset group by mission_outcome
```

```
 * ibm_db_sa://zgg83101:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:3119
8/bludb
Done.
```

mission_outcome	2
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- 12 Booster versions carried the maximum payload

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

```
 * ibm_db_sa://zgg83101:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31198/bludb
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

- Total of 2 failed landing outcomes, both at CCAFS LC-40, with 2 different booster versions

```
%sql select landing__outcome, booster_version, launch_site from spacexdataset where date between '2015-01-01' and '2016
```

```
 * ibm_db_sa://zgg83101:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:3119
8/bludb
Done.
```

landing__outcome	booster_version	launch_site
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

- The overall #1 landing outcome is “Successs”
- Drone ship is the #1 failed outcome

```
%sql select * from (select landing__outcome, count(*) as num from spacexdataset group by landing__outcome) ref order by
```

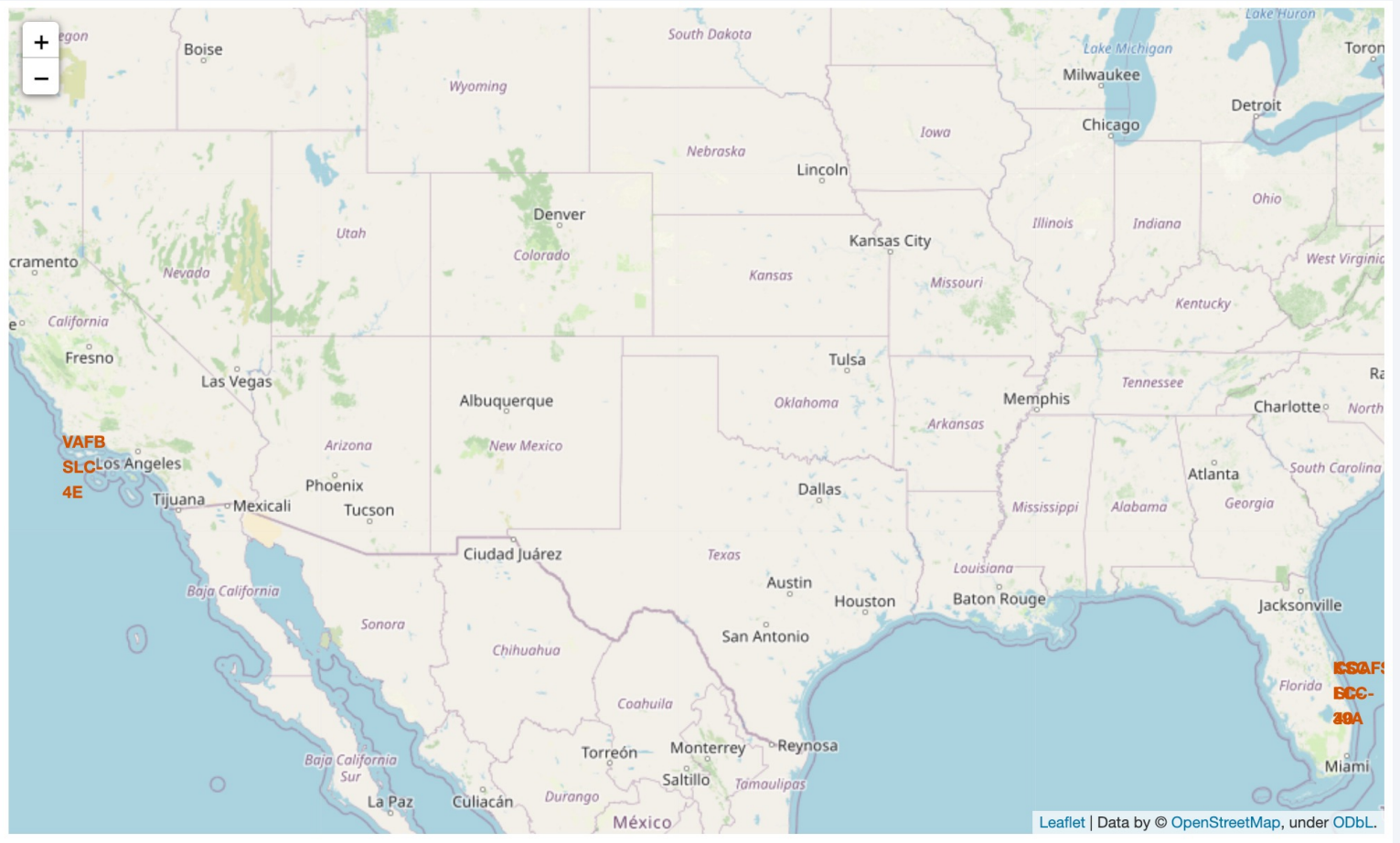
```
 * ibm_db_sa://zgg83101:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:3119
8/bludb
Done.
```

landing__outcome	num
Success	38
No attempt	22
Success (drone ship)	14
Success (ground pad)	9
Controlled (ocean)	5
Failure (drone ship)	5
Failure	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1


Section 3
Launch Sites
Proximities Analysis

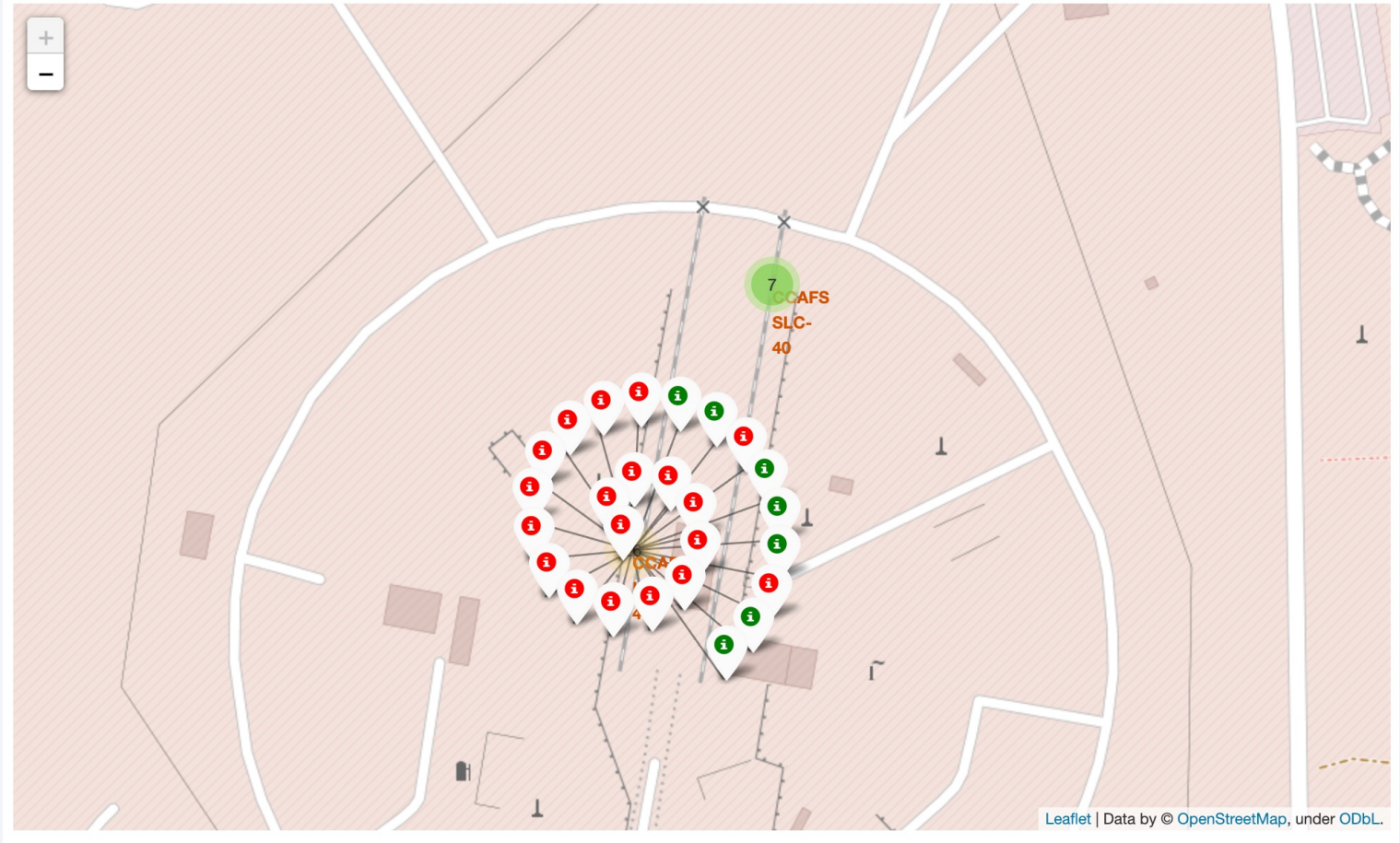
Launch Site Locations

- Launch sites are primarily along the coastline



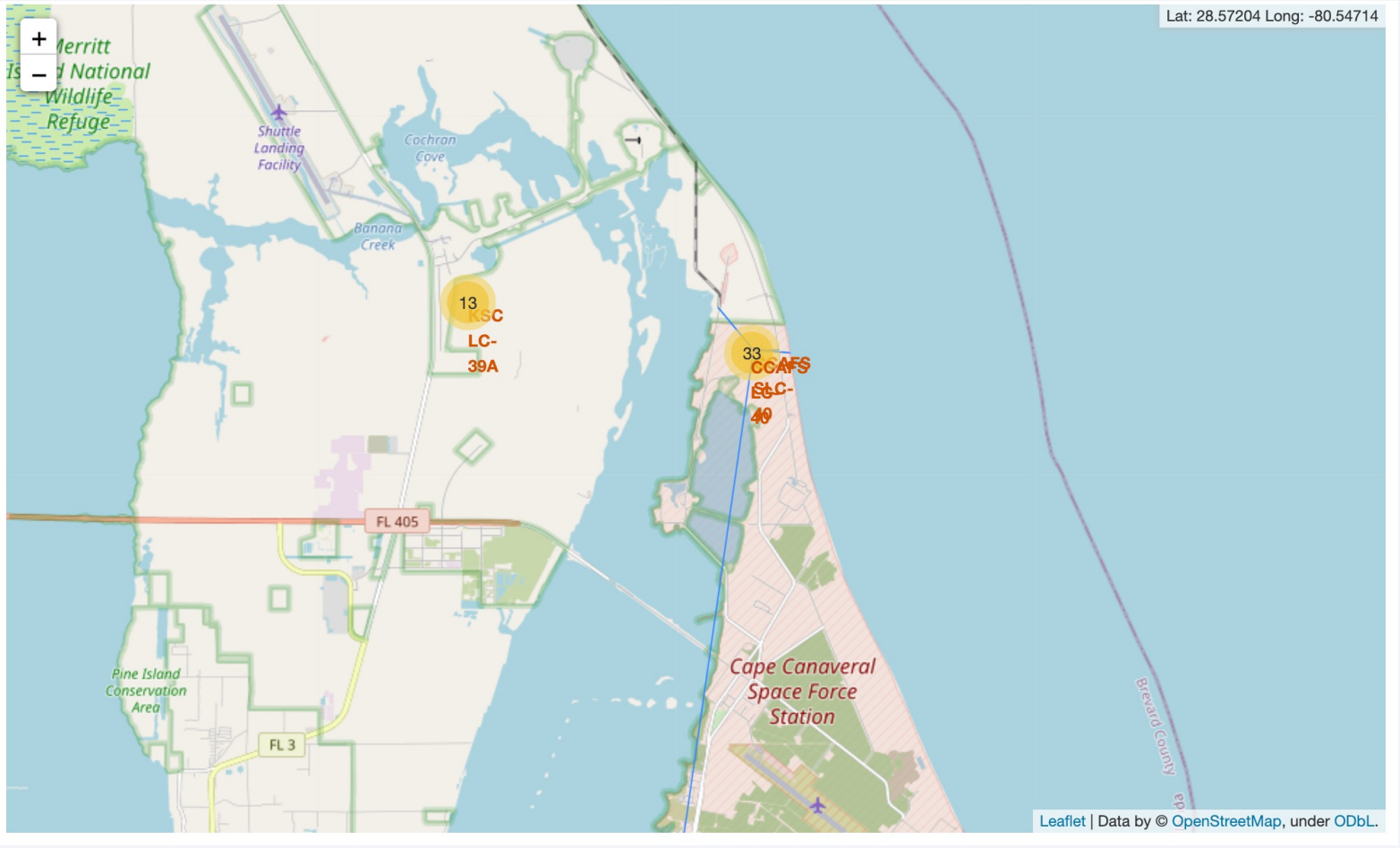
Success Rate at CCAFS LC-40

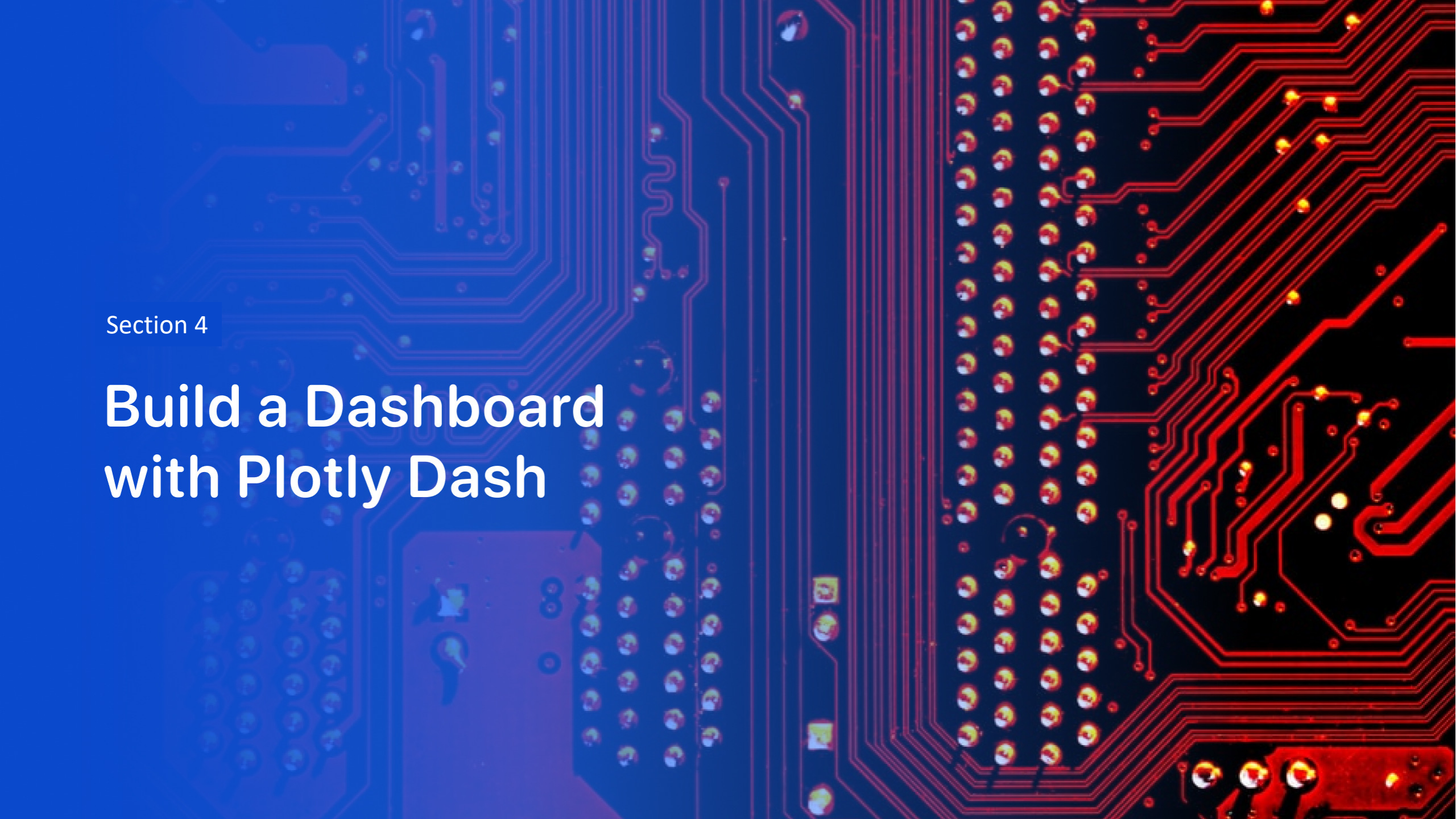
- More failed launches at CCAFS than successful launches



Launch Site Proximities to Landmarks

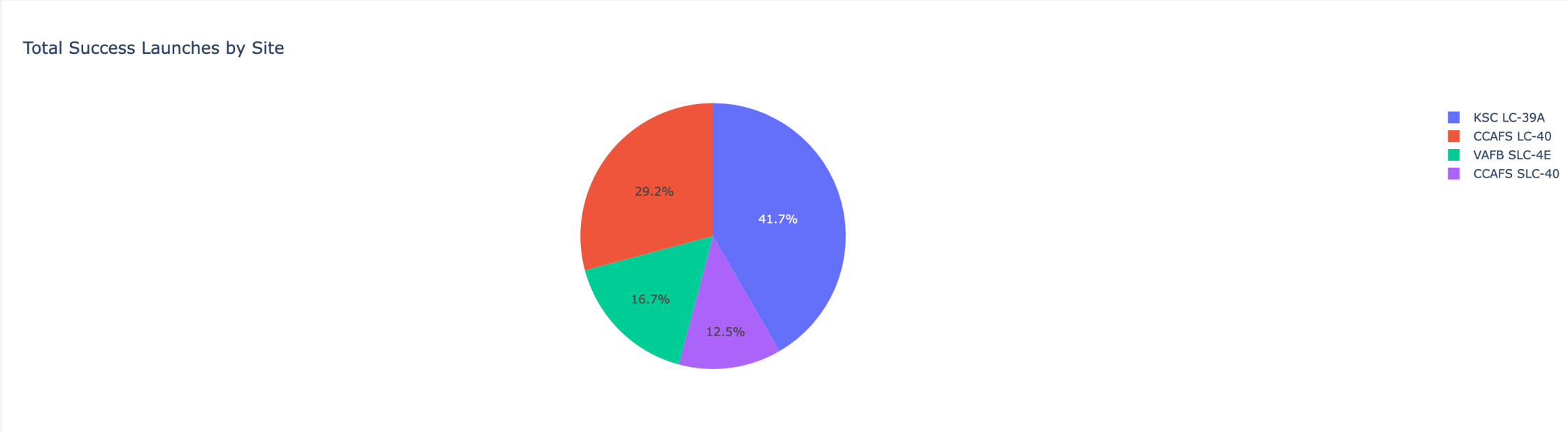
- Launches sites are chosen with close proximities to coastline, highway and railroad, but far from cities




Section 4
Build a Dashboard with Plotly Dash

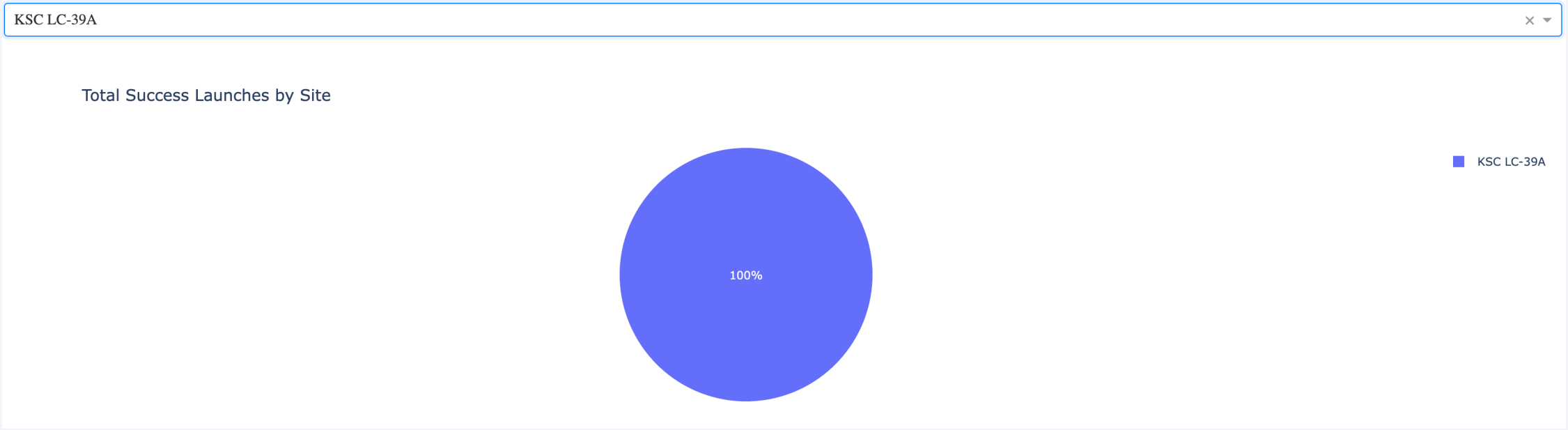
Success Rate by Launch Site

- KSC LC-39A has the highest success rate



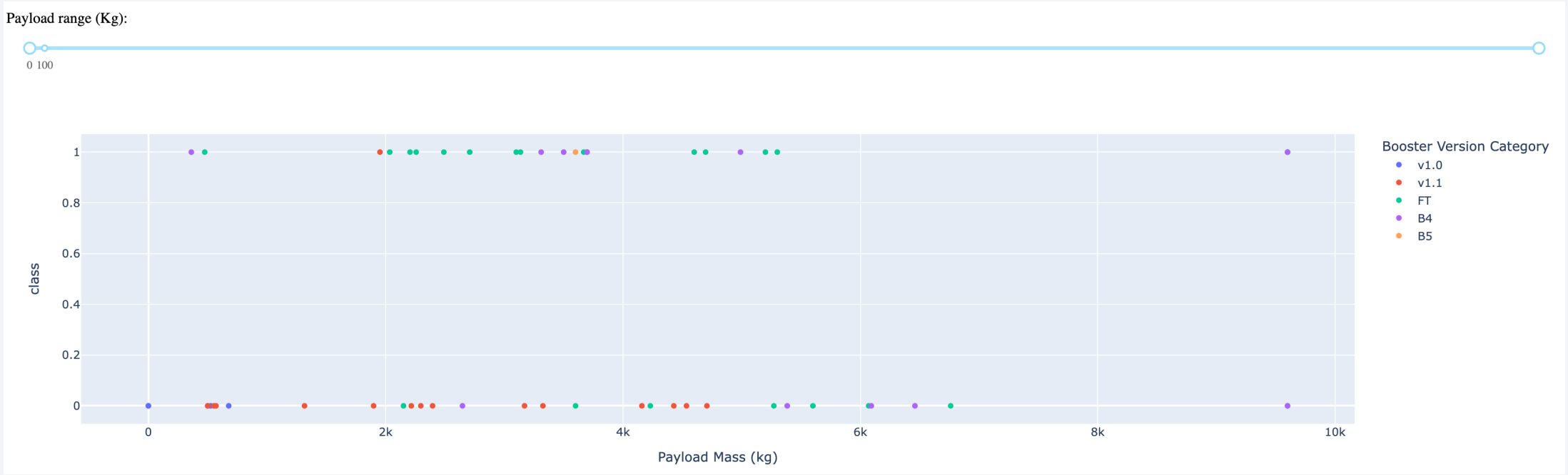
Launch Site with Highest Success Rate

- KSC LC-39A has the highest success rate



Success Rate by Payload & Booster Version

- Payload range between 4000 – 6000 with Booster Version FT has the highest success rate



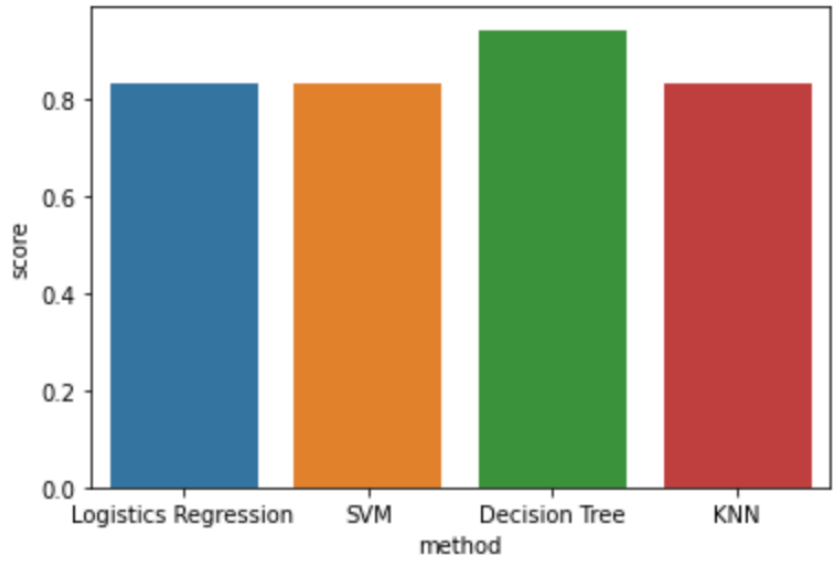
Section 5

Predictive Analysis (Classification)

Classification Accuracy

- Decision Tree has the highest accuracy out of all the models

```
df = pd.DataFrame(data={'method':['Logistics Regression', 'SVM', 'Decision Tree', 'KNN'],'score':[0.8333,0.8333,0.9444,
plot = sns.barplot(x="method", y="score", data=df)
```



Confusion Matrix

- Only 1 false positive result from the prediction

```
yhat = tree_cv.predict(X_test)
plot_confusion_matrix(Y_test,yhat)
```

Conclusions

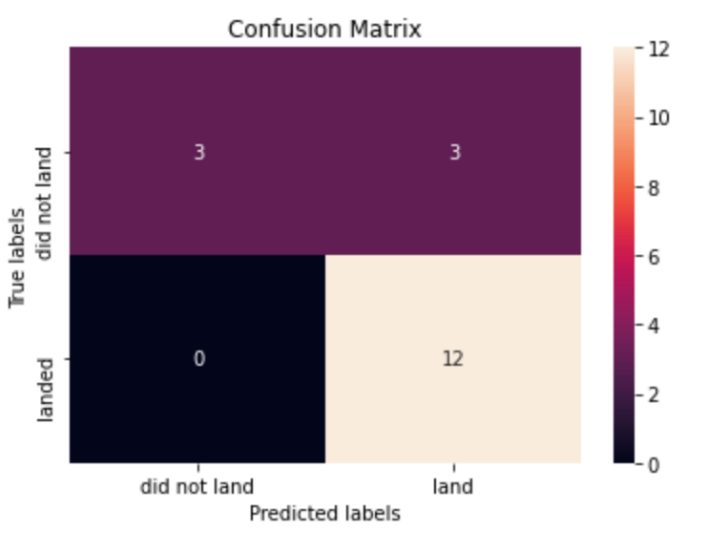
- There will be a learning curve in the beginning, launch success rate will improve overtime
- EL-L1, GEO, HEO, SSO, VLEO are the orbits for higher success rate
- Payload should be between 4000 – 6000 for higher success rate
- When choosing a launch site location, we should choose some where has proximity with coastline, highway and railroad, but away from city
- Drone Ship landing option is riskier
- Decision Tree model will work best for predicting whether if we can have a successful landing result

Appendix

- Average success rate from Space X data set is about 66.67%

```
In [19]: df["Class"].mean()
Out[19]: 0.6666666666666666
```

- Logistics Regression, SVM & KNN models generate same confusion matrix in this example



Thank you!