IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Ryan Robertson

Outline

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

Executive Summary

- SpaceX advertises Falcon 9 rocket launches for $62 million for missions that allow for the 1st stage booster to be reused.
- The models in this report were able to predict an 83.3% success rate for the booster landing. Using this information competing companies can better adjust their cost predictions.

Introduction

- This report is part of the Applied Data Science Capstone course.
- In this project we are a data scientist working for SpaceY, a competing company to SpaceX.
- We are tasked with using data science to determine how to adjust our offers to better match SpaceX.
- This will be accomplished by calculating the likelihood that the first stage rocket will successfully land and be reused.

Section 1

Methodology

Methodology

- Data collection
  - API
    - Acquired rocket launch data from SpaceX API
      - Requested data using a GET
      - Filtered data frame to include only Falcon 9

Methodology

- Data wrangling
  - Web Scraping
    - Scraped data from the launch Wikipedia page
      - Requested data from page using URL
      - Extracted and converted to data frame

Methodology

- EDA
  - SQL
    - Used IBM DB2
    - Ran queries about
      - Launch site
      - Payload masses
      - Booster versions
      - Mission outcomes
      - Booster landings

```
[8]: %sql select distinct(LAUNCH_SITE) from SPACEXTBL
 * sqlite:///my_data1.db
Done.
[8]:
```

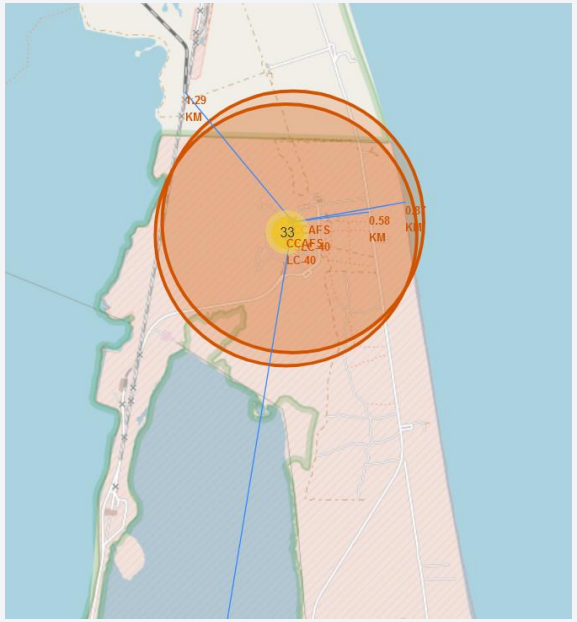
Launch_Site
CCAFS LC-40
VAFB SLC-4E
KSC LC-39A
CCAFS SLC-40

Methodology

- Visualization
  - Used Matplotlib and Seaborn to plot
    - FlightNumber x PayloadMass
    - FlightNumber x LaunchSite
    - PayloadMass x LaunchSite
    - OrbitType x SuccessRate
    - FlightNumber x OrbitType
    - PayloadMass x OrbitType
    - Year x Success Rate

Methodology

- Interactive visual analysis
  - Launch Site analysis
    - Folium
    - Marked all sites and success/fail
    - Calculated distance between a site to
      - Railways, Highways, Coastlines, Cities
  - Launch Records Dash
    - Plotly Dash
      - Pie chart showing success rate
      - Scatter chart showing payload v. landing outcome
      - Drop-down menu to pick between all sites or a specific site

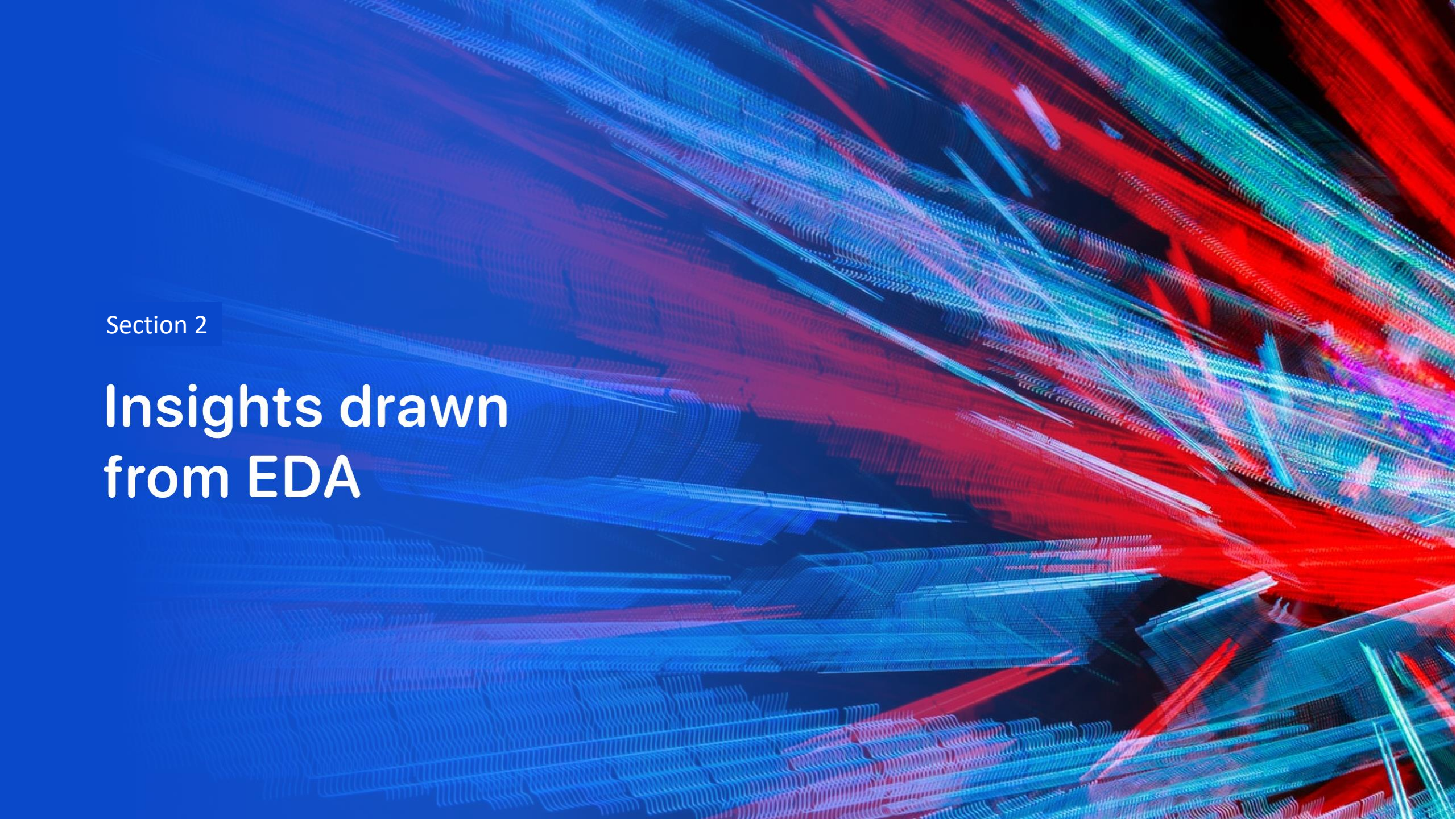


Methodology

- Predictive analysis
  - Created a confusion matrix
  - Load data, create column for training, standardized the data
  - Split data into test and training sets
  - Fit data using
    - Log Regression, Support Vector Machine, Decision Tree, K Nearest Neighbor
  - Evaluated accuracy of each to choose best

Section 2

Insights drawn from EDA



Results

- EDA with SQL

Task 1

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

```
[8]: %sql select distinct(LAUNCH_SITE) from SPACEXTBL
```

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

[8]:

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

Task 2

Display 5 records where launch sites begin with the string 'CCA'

```
[9]: %sql select * from SPACEXTBL where LAUNCH_SITE like 'CCA%' limit 5
```

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

[9]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2010-04-06	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-08-12	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
2012-08-10	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-01-03	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Launch Sites Used:
CCAFS LC-40
CCAFS SLC -40
VAFB SLC-4E
KSC LC-39A

Results

- EDA with SQL

Total Mass: 45596 kg
Avg Mass: 2928.4 kg

First successful landing: 12-22-2015

Task 3

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

```
[10]: %sql select sum(PAYLOAD_MASS__KG_) from SPACEXTBL where CUSTOMER = 'NASA (CRS)'
```

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

[10]:

sum(PAYLOAD_MASS__KG_)
45596

Task 4

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

```
[11]: %sql select avg(PAYLOAD_MASS__KG_) from SPACEXTBL where BOOSTER_VERSION = 'F9 v1.1'
```

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

[11]:

avg(PAYLOAD_MASS__KG_)
2928.4

Task 5

List the date when the first succesful landing outcome in ground pad was acheived.

Hint:Use min function

```
[13]: %sql select min(DATE) from SPACEXTBL where Landing_Outcome = 'Success (ground pad)'
```

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

[13]:

min(DATE)
2015-12-22

Task 6

List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

```
[15]: %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.
```

[15]:

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

Results

- EDA with SQL

Task 7

List the total number of successful and failure mission outcomes

```
[16]: %sql select count(MISSION_OUTCOME) from SPACEXTBL where MISSION_OUTCOME = 'Success' or MISSION_OUTCOME = 'Failure (in flight)'
```

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

[16]:

count(MISSION_OUTCOME)
99

Task 8

List the names of the booster_versions which have carried the maximum payload mass. Use a subquery

```
[17]: %sql select BOOSTER_VERSION from SPACEXTBL where PAYLOAD_MASS__KG_ = (select max(PAYLOAD_MASS__KG_) from SPACEXTBL)
```

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

[17]:

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

Task 10

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.

```
[20]: %sql select * from SPACEXTBL where Landing_Outcome like 'Success%' and (DATE between '2010-06-04' and '2017-03-20') order by date desc
```

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

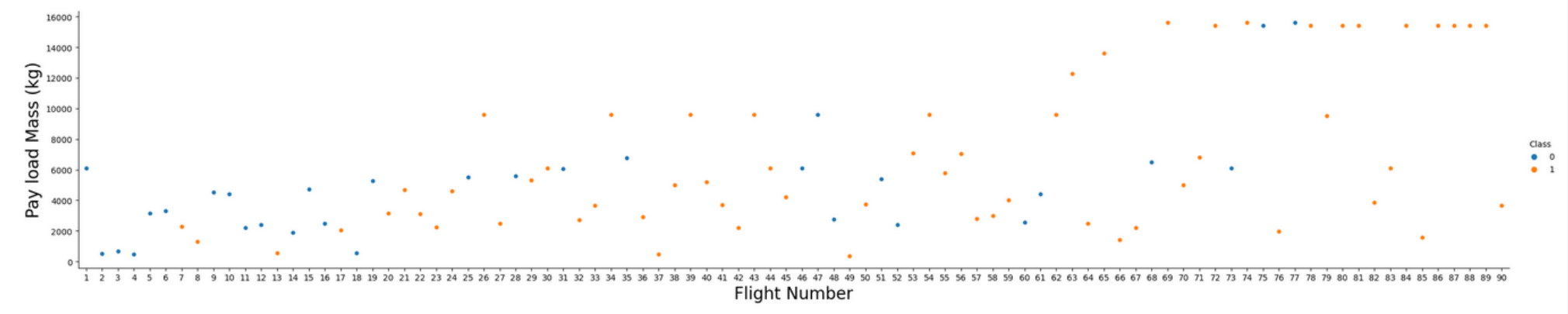
[20]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2017-03-06	21:07:00	F9 FT B1035.1	KSC LC-39A	SpaceX CRS-11	2708	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
2017-02-19	14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
2017-01-14	17:54:00	F9 FT B1029.1	VAFB SLC-4E	Iridium NEXT 1	9600	Polar LEO	Iridium Communications	Success	Success (drone ship)
2017-01-05	11:15:00	F9 FT B1032.1	KSC LC-39A	NROL-76	5300	LEO	NRO	Success	Success (ground pad)
2016-08-14	05:26:00	F9 FT B1026	CCAFS LC-40	JCSAT-16	4600	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
2016-08-04	20:43:00	F9 FT B1021.1	CCAFS LC-40	SpaceX CRS-8	3136	LEO (ISS)	NASA (CRS)	Success	Success (drone ship)
2016-07-18	04:45:00	F9 FT B1025.1	CCAFS LC-40	SpaceX CRS-9	2257	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
2016-06-05	05:21:00	F9 FT B1022	CCAFS LC-40	JCSAT-14	4696	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
2016-05-27	21:39:00	F9 FT B1023.1	CCAFS LC-40	Thaicom 8	3100	GTO	Thaicom	Success	Success (drone ship)
2015-12-22	01:29:00	F9 FT B1019	CCAFS LC-40	OG2 Mission 2 11 Orbcomm-OG2 satellites	2034	LEO	Orbcomm	Success	Success (ground pad)

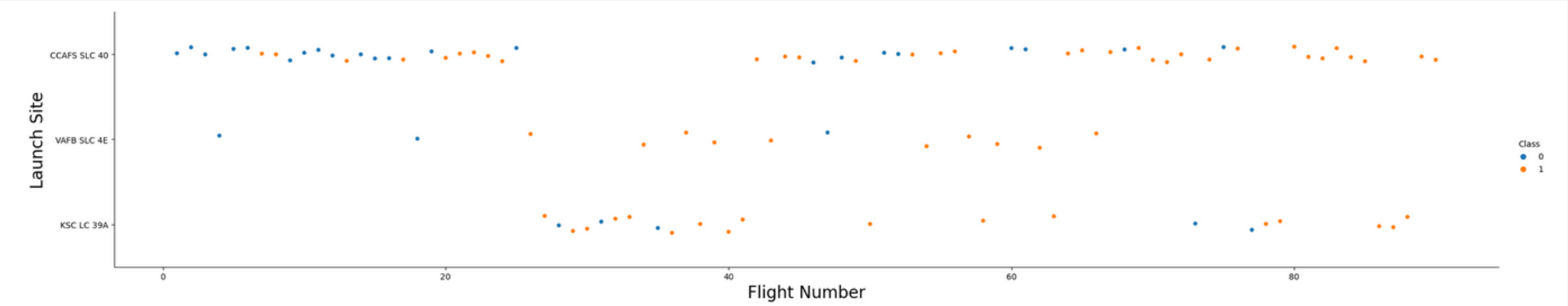
No Attempt: 10
Failure (drone): 5
Success (drone): 5
Controlled (ocean): 3
Success(ground pad): 3
Failure (parachute): 2
Uncontrolled: 2
Precluded: 1

Results

- EDA with Visualization



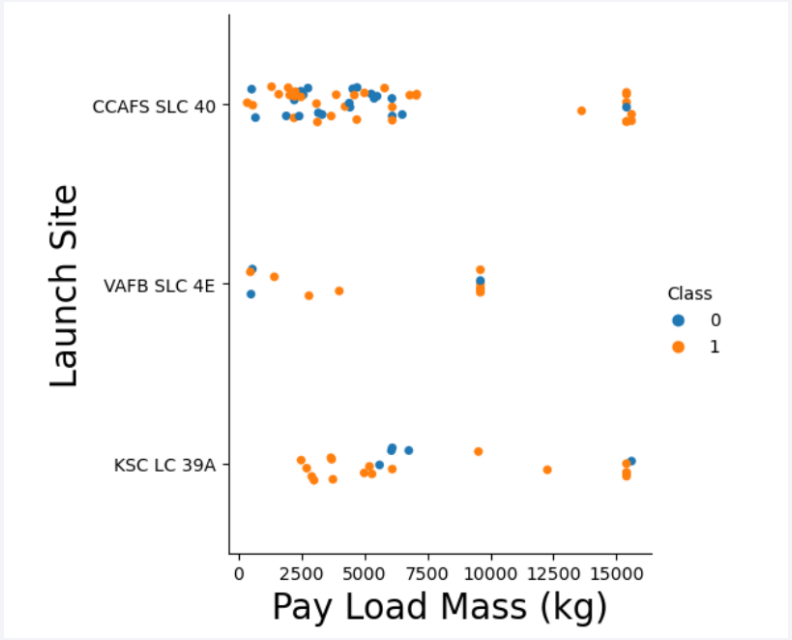
Positively correlated with continuous launch attempts

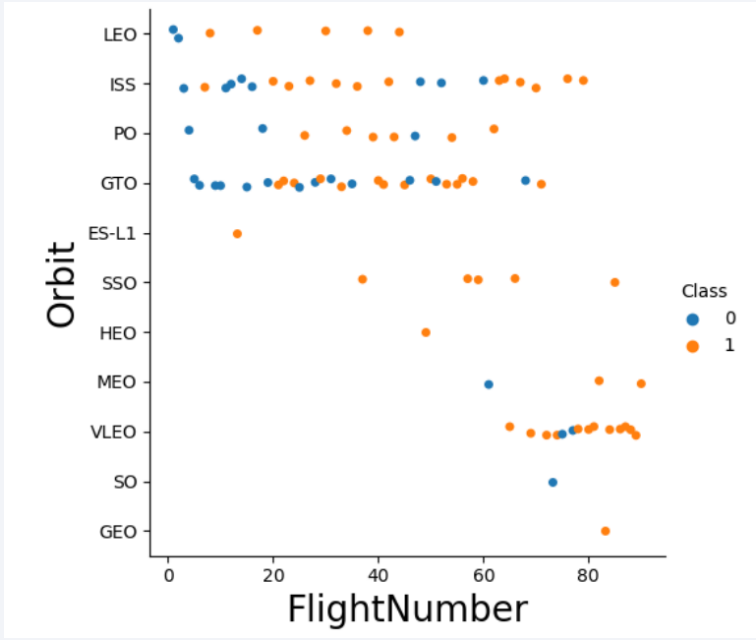


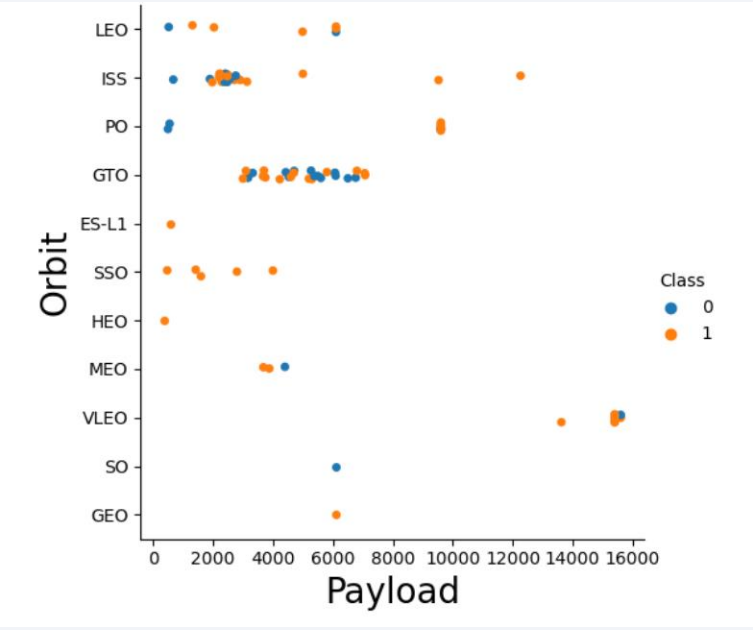
CCAFS SLC 40 is where most failures took place

Results

- EDA with visualization



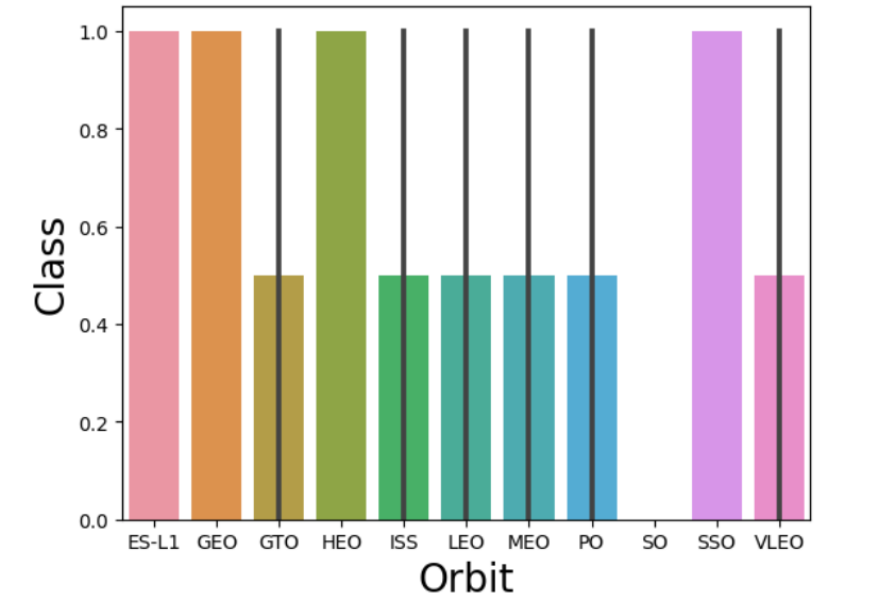


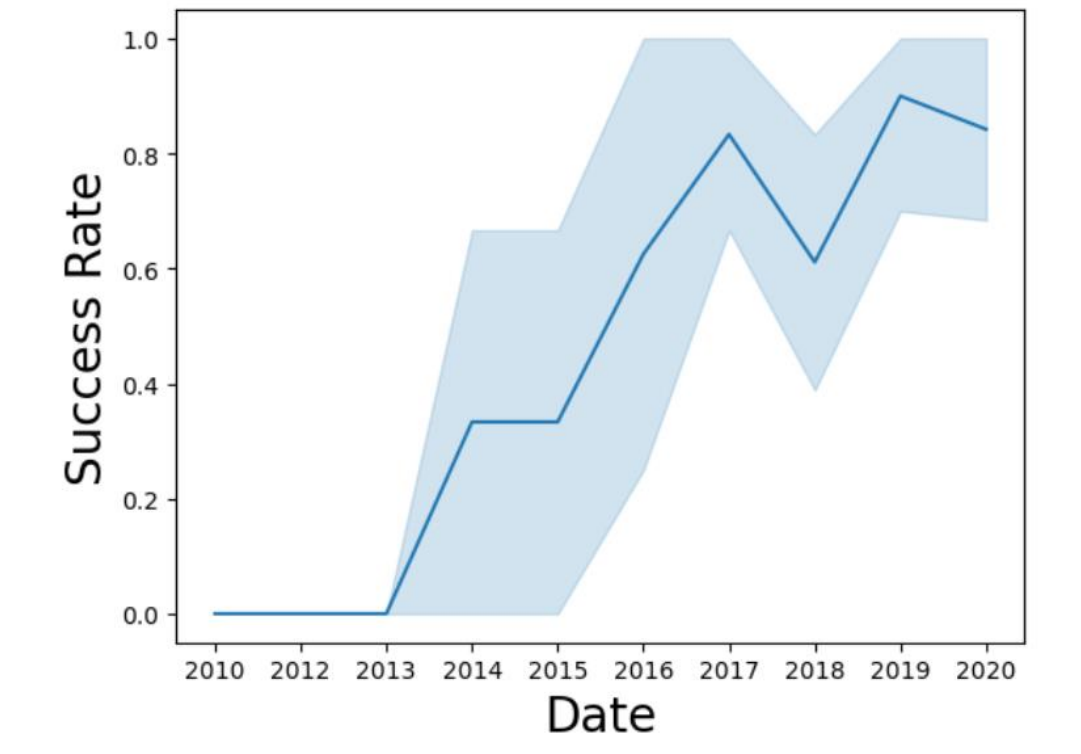


Orbit is negatively correlated with increased flight attempts and a heavier payload

Launch Success Yearly Trend

- EDA with visualization

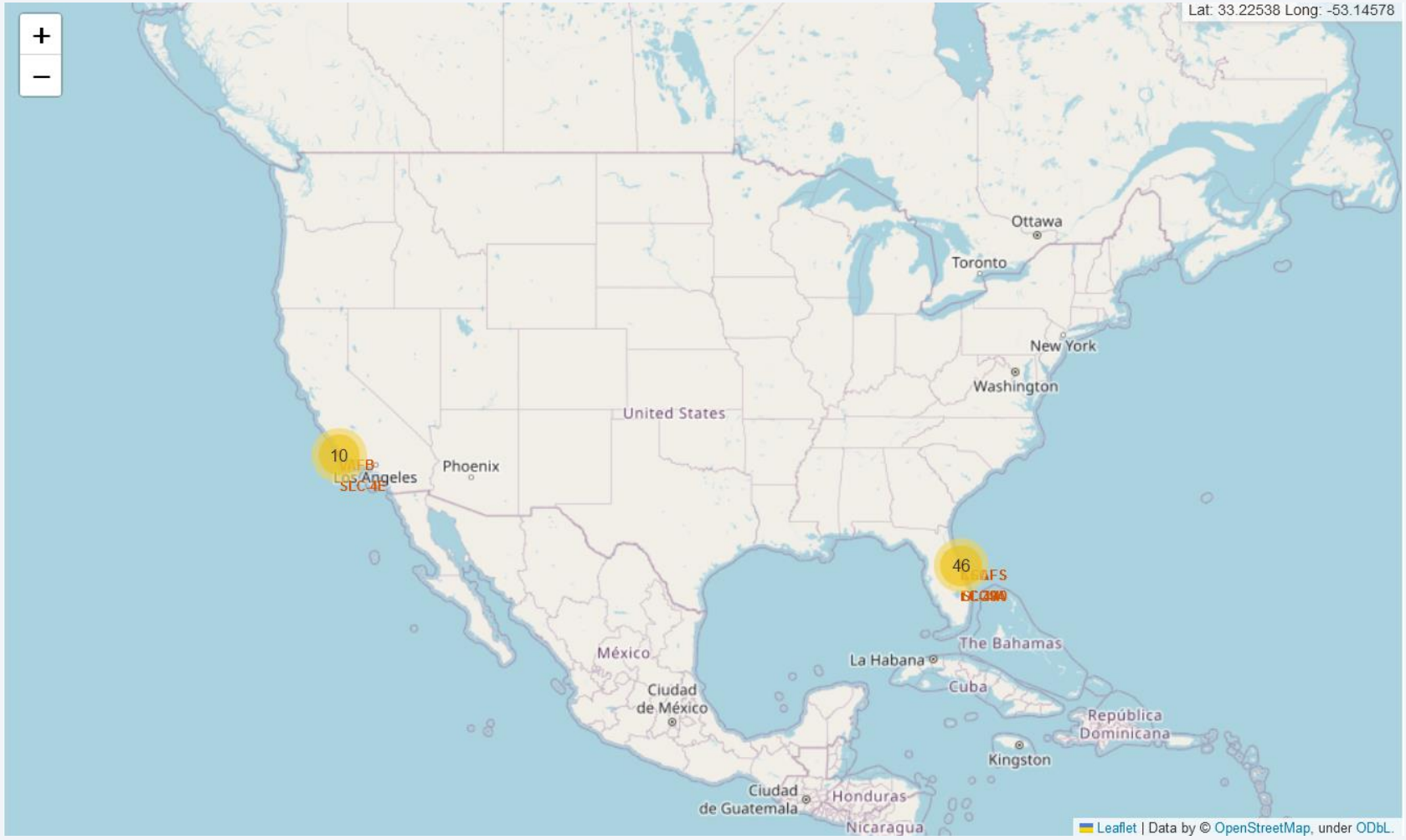




As the company has grown and learned from their mistakes the success rate has increased, with a few dips

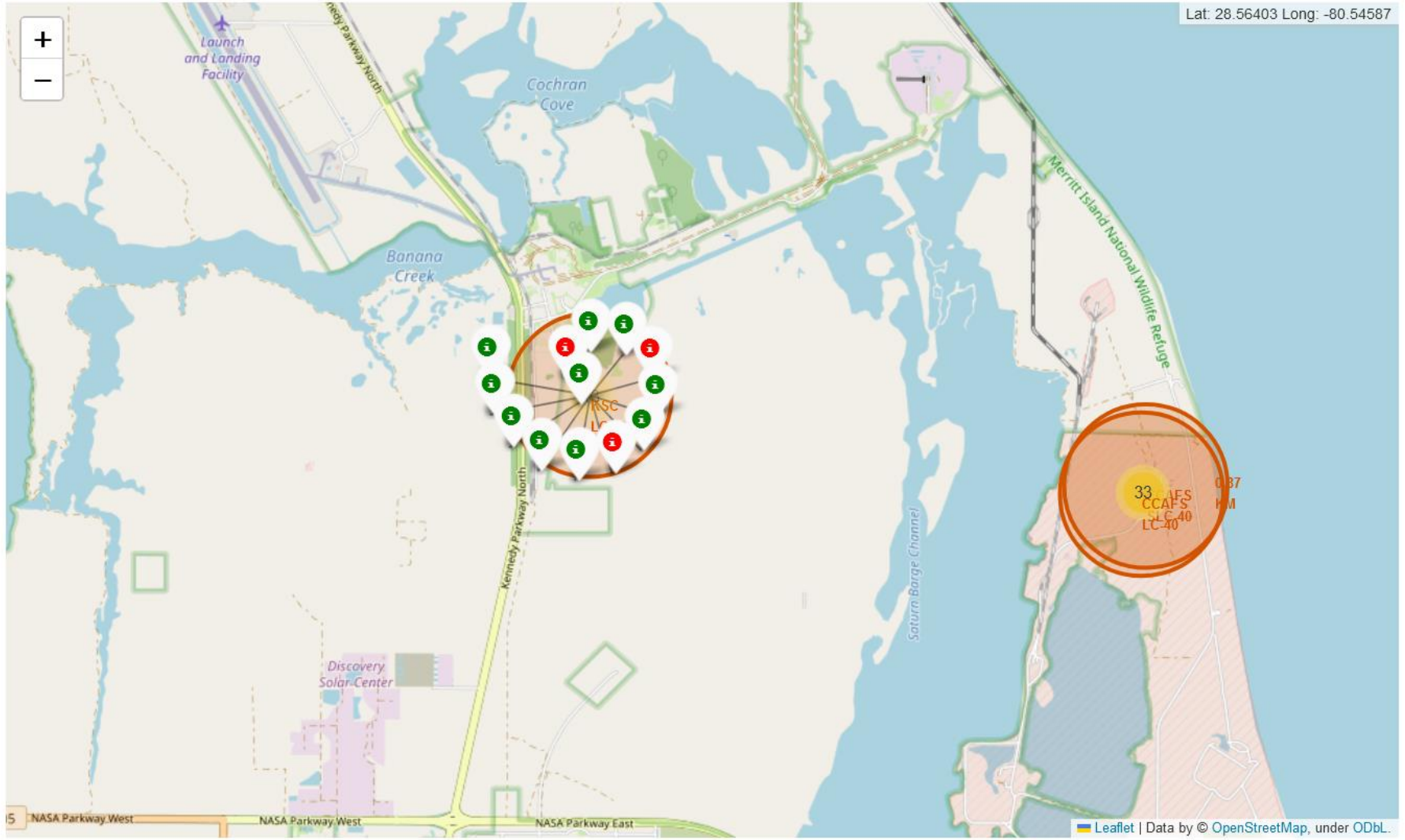

Section 3
Launch Sites
Proximities Analysis

Launch Site Location Analysis



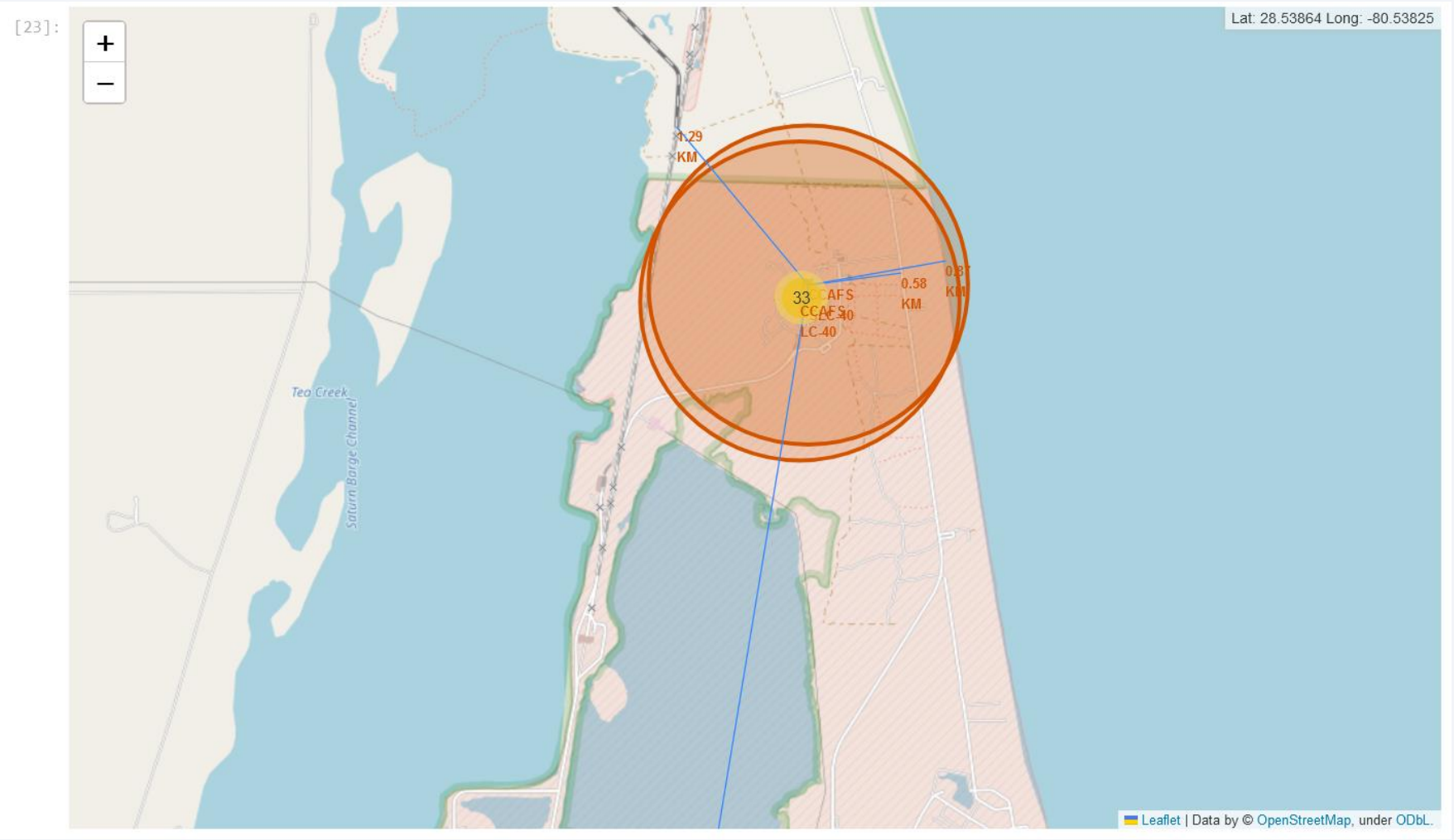
Launch Site Location Analysis

[19]:



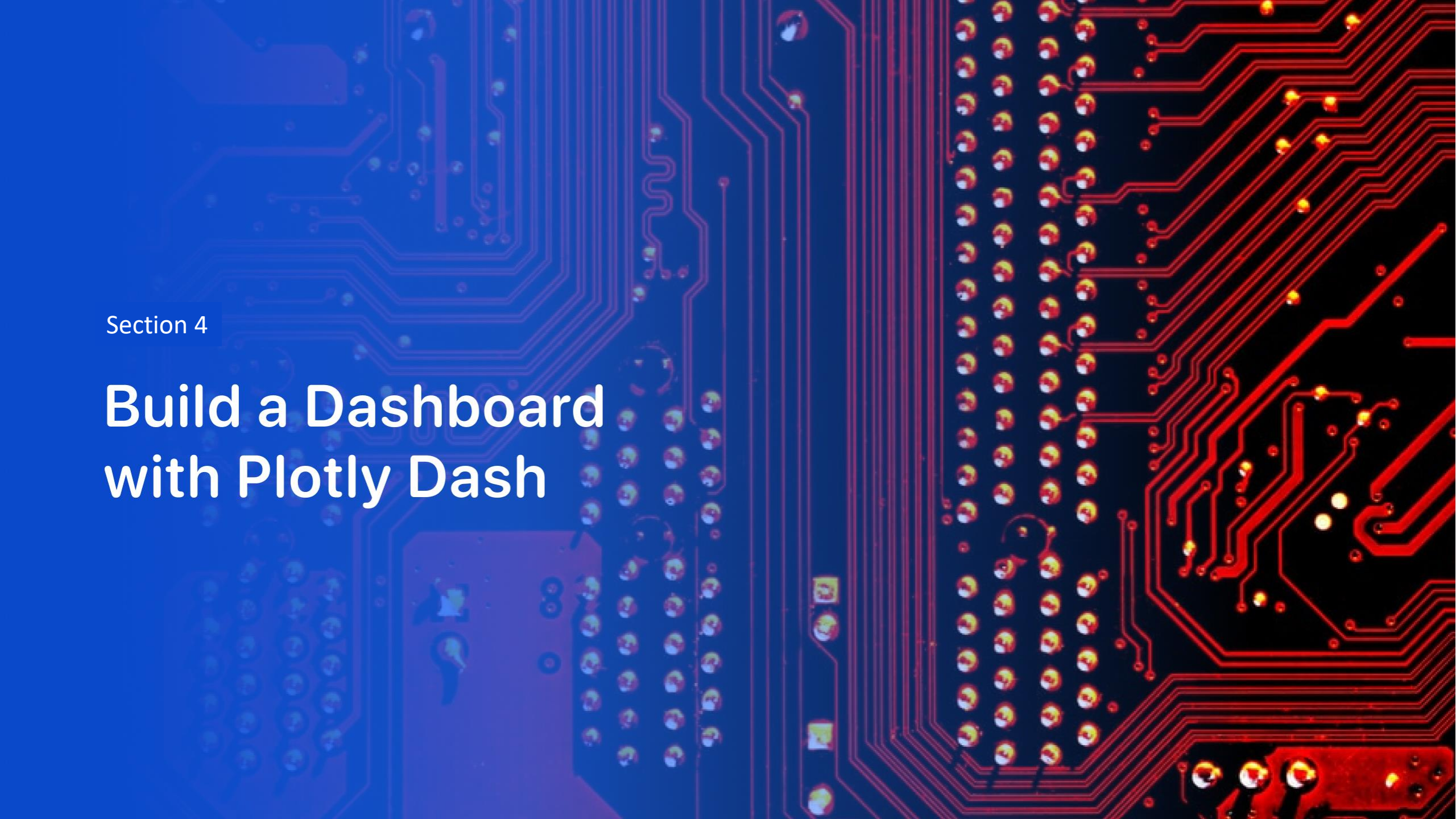
Shows booster landing success/fail for each site.

Launch Site Location Analysis

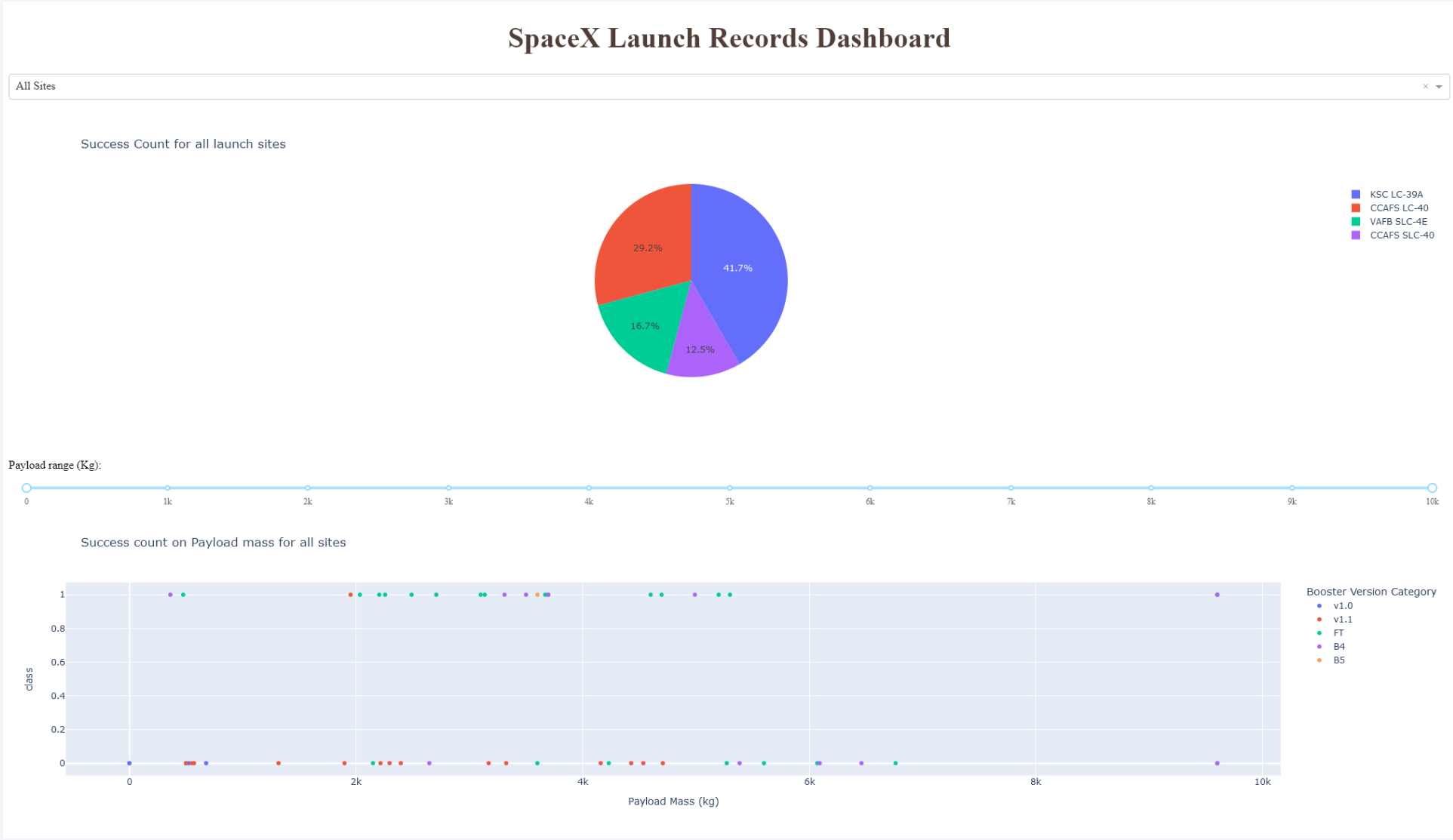


Visualizes distance to railway, highway, coast, and another city

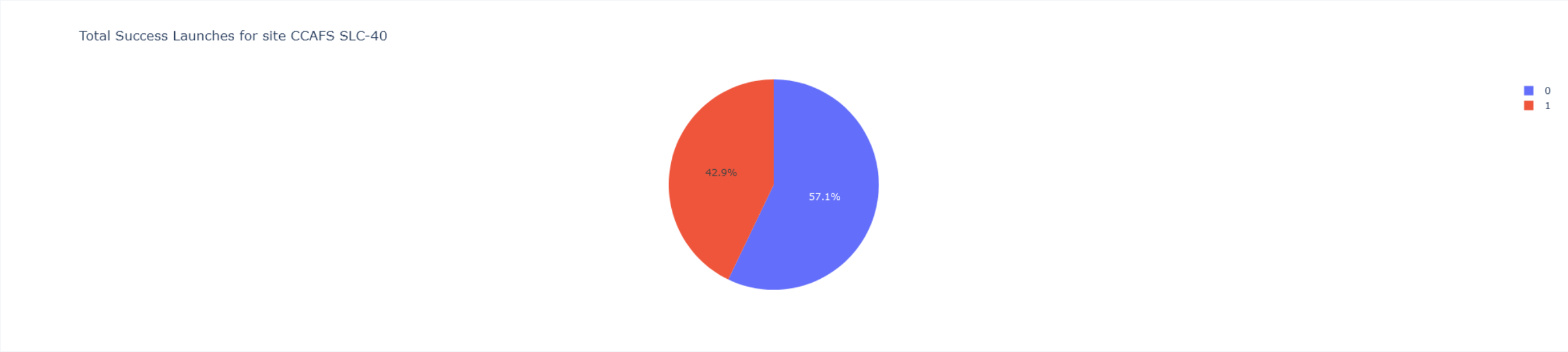
Railway: 1.29 km
Highway: .58 km
Coast: .87 km
City: 18.1 km


Section 4
Build a Dashboard with Plotly Dash

Launch Records Dash

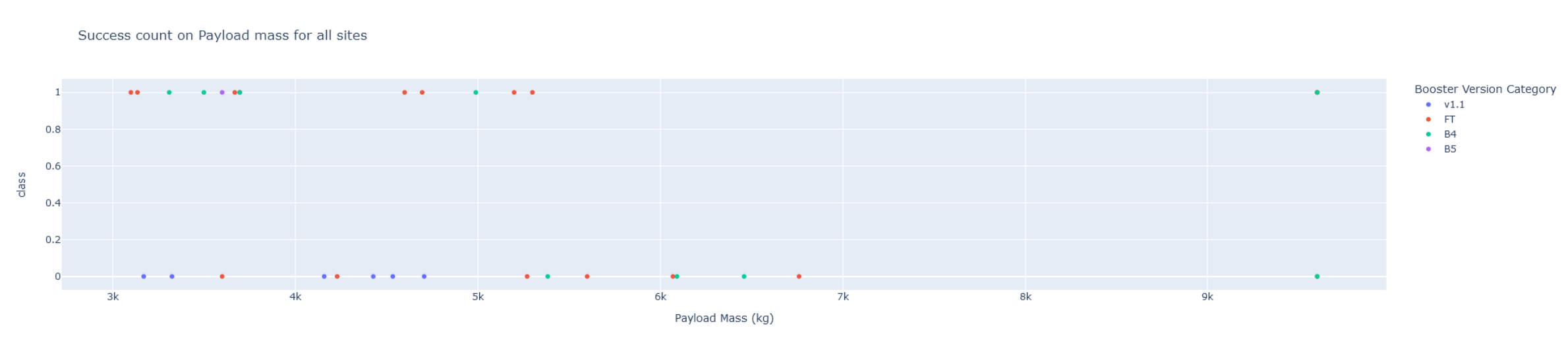


Launch Records Dash



CCAFS SLC-40 is the launch site with the highest success rate. At 42.9% success rate.

Launch Records Dash

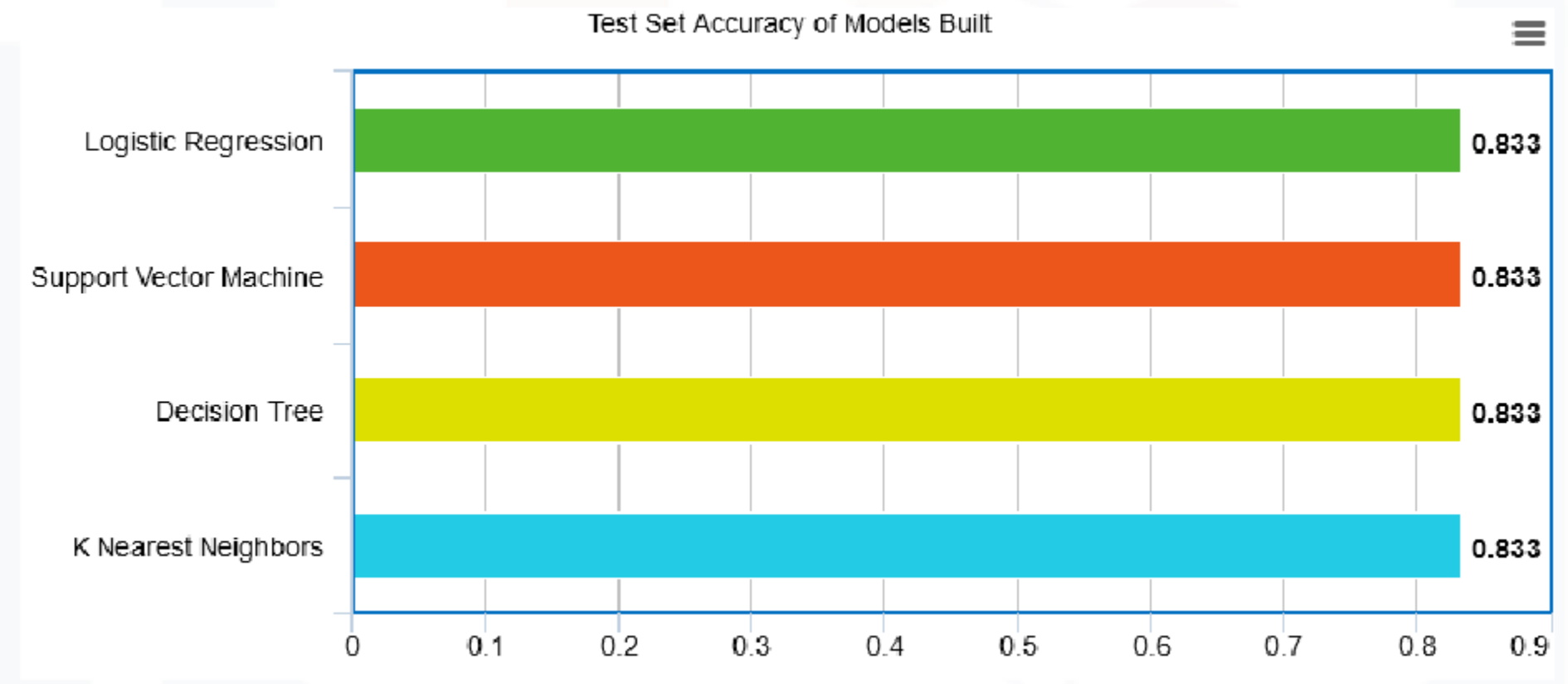


- Payloads greater than 5.3k kg had the highest booster landing success rate

Section 5

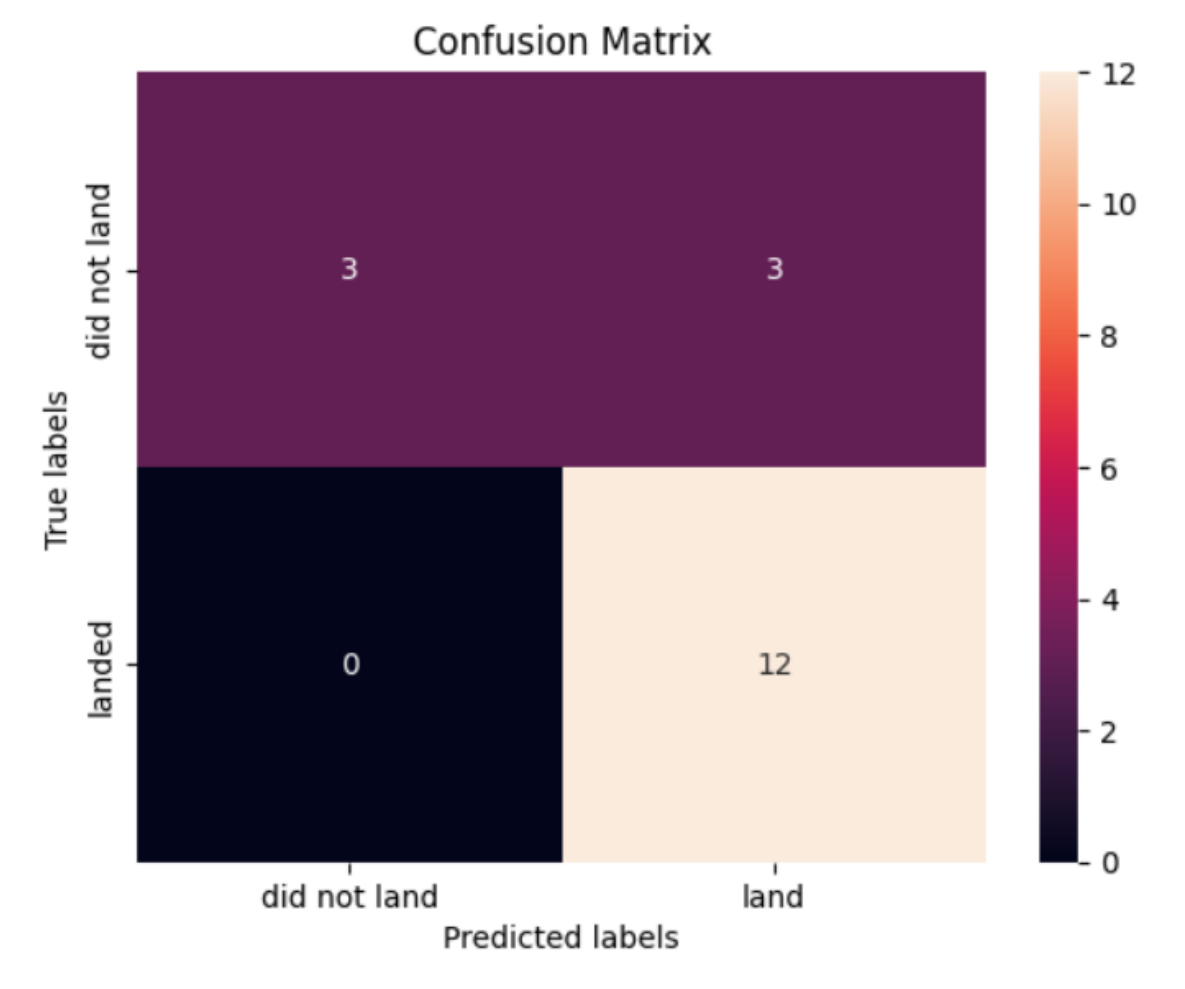
Predictive Analysis (Classification)

Classification Accuracy



All models produced the same accuracy

Confusion Matrix



As they all have the same accuracy, the confusion matrix is also the same.

Conclusions

- Using the models presented in this report SpaceY can predict if SpaceX will successfully land the first booster rocket.
- As SpaceX has stated that a booster costs $15 million, SpaceY can use this information along with the list price of $62 million to make a more competitive bid.
- Still opportunities to improve accuracy and be even more competitive.
  - Continually update as more launches occur
  - Re-fit using the whole data set instead of training sets.
  - Add a parameter that takes into account if the booster has already flown before

Thank you!