Applied Data Science Capstone

Falcon 9 first stage successful landing prediction

Martina Oravcova

1.11.2022

Executive Summary

- In this capstone, I predict if SpaceX Falcon 9 first stage will land successfully
- I train a machine learning model and use public information to predict if SpaceX will reuse the first stage

Table of contents

Introduction

- SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars
- other providers cost upwards of 165 million dollars each
- much of the savings is because SpaceX can reuse the first stage
- therefore, if I can determine if the first stage will land, we can determine the cost of a launch

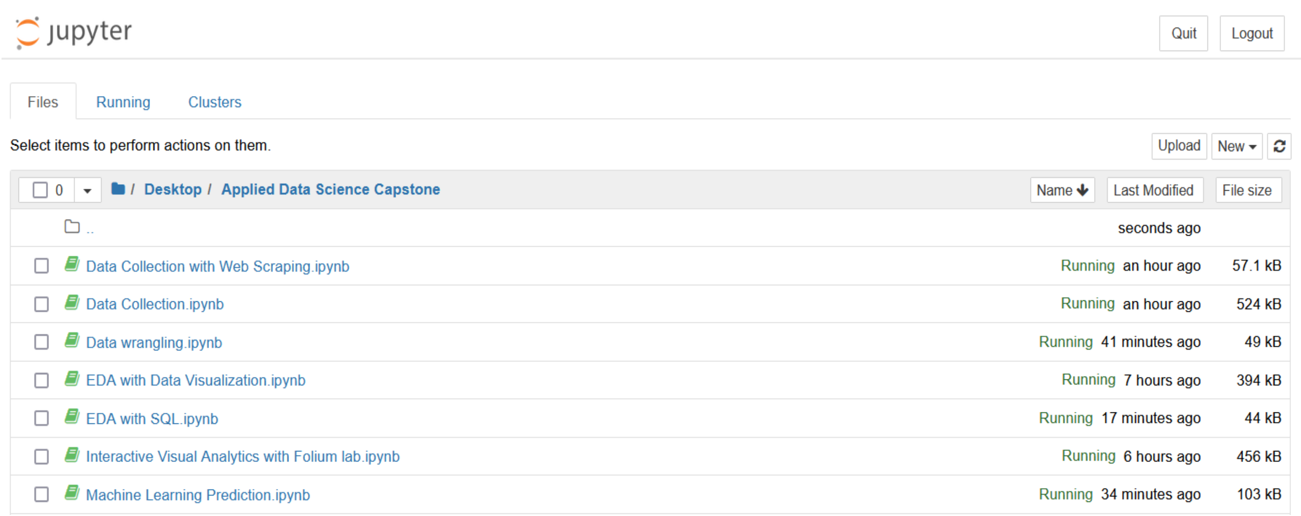
- First stage is quite large and expensive
- Unlike other rocket providers, SpaceX's Falcon 9 can recover the first stage
- Sometimes the first stage does not land
- Sometimes it will crash
- Other times, Space X will sacrifice the first stage due to the mission parameters like payload, orbit, and customer

Main question for analysis:

Will SpaceX Falcon 9 first stage land successfully ?

Methodology 1

- General methods
- Write Python code to manipulate data in a Pandas data frame
- Convert a JSON file into a Pandas data frame
- Create a Jupyter notebook and make it sharable using GitHub
- Use data science methodologies like web scraping to define and formulate a real-world business problem
- Use data analysis tools to load a dataset, clean it, and find out interesting insights from it



Methodology 2

Data sources and methods in details:

- Data collection by Web scraping Falcon 9 and Falcon Heavy Launches Records from Wikipedia page titled List of Falcon 9 and Falcon Heavy launches

- Data collection by Request and parse the SpaceX launch data using the GET request to the SpaceX API

- Wrangle data by creating a Python Pandas data frame from JSON file gained from API

- Wrangle data by replacing missing values of PayloadMass with mean

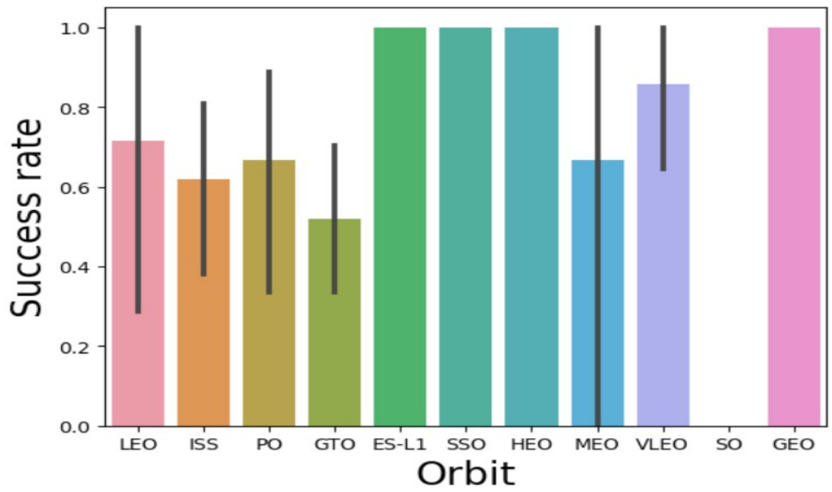
2020 [edit]

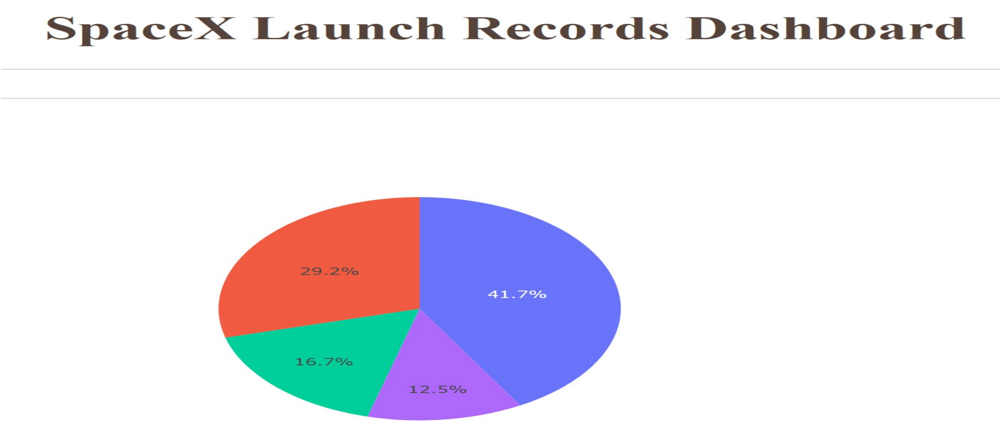
In late 2019, Gwynne Shotwell stated that SpaceX hoped for as many as 24 launches for Starlink satellites in 2020,[490] in addition to 14 or 15 non-Starlink launches. At 26 launches, 13 of which for Starlink satellites, Falcon 9 had its most prolific year, and Falcon rockets were second most prolific rocket family of 2020, only behind China's Long March rocket family.[491]

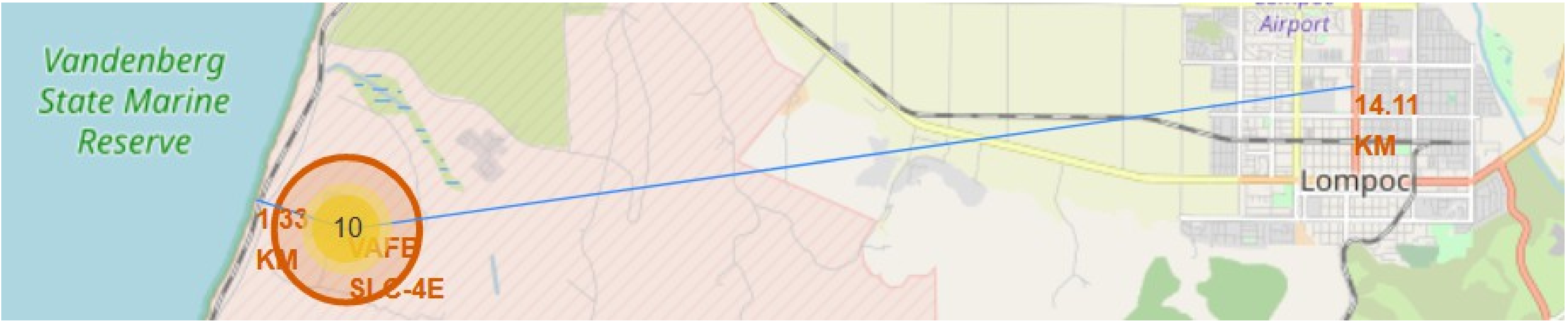
[hide] Flight No.	Date and time (UTC)	Version, Booster[b]	Launch site	Payload[c]	Payload mass	Orbit	Customer	Launch outcome	Booster landing
78	7 January 2020, 02:19:21[492]	F9 B5 ♺ B1049.4	CCAFS, SLC-40	Starlink 2 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Success (drone ship)
	Third large batch and second operational flight of Starlink constellation. One of the 60 satellites included a test coating to make the satellite less reflective, and thus less likely to interfere with ground-based astronomical observations.[493]								
79	19 January 2020, 15:30[494]	F9 B5 ♺ B1046.4	KSC, LC-39A	Crew Dragon in-flight abort test[495] (Dragon C205.1)	12,050 kg (26,570 lb)	Sub-orbital[496]	NASA (CTS)[497]	Success	No attempt
	An atmospheric test of the Dragon 2 abort system after Max Q. The capsule fired its SuperDraco engines, reached an apogee of 40 km (25 mi), deployed parachutes after reentry, and splashed down in the ocean 31 km (19 mi) downrange from the launch site. The test was previously slated to be accomplished with the Crew Dragon Demo-1 capsule;[498] but that test article exploded during a ground test of SuperDraco engines on 20 April 2019.[419] The abort test used the capsule originally intended for the first crewed flight.[499] As expected, the booster was destroyed by aerodynamic forces after the capsule aborted.[500] First flight of a Falcon 9 with only one functional stage — the second stage had a mass simulator in place of its engine.								
80	29 January 2020, 14:07[501]	F9 B5 ♺ B1051.3	CCAFS, SLC-40	Starlink 3 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Success (drone ship)
	Third operational and fourth large batch of Starlink satellites, deployed in a circular 290 km (180 mi) orbit. One of the fairing halves was caught, while the other was fished out of the ocean.[502]								
81	17 February 2020, 15:05[503]	F9 B5 ♺ B1056.4	CCAFS, SLC-40	Starlink 4 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Failure (drone ship)
	Fourth operational and fifth large batch of Starlink satellites. Used a new flight profile which deployed into a 212 km × 386 km (132 mi × 240 mi) elliptical orbit instead of launching into a circular orbit and firing the second stage engine twice. The first stage booster failed to land on the drone ship[504] due to incorrect wind data.[505] This was the first time a flight proven booster failed to land.								
82	7 March 2020, 04:50[506]	F9 B5 ♺ B1059.2	CCAFS, SLC-40	SpaceX CRS-20 (Dragon C112.3 ♺)	1,977 kg (4,359 lb)[507]	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
	Last launch of phase 1 of the CRS contract. Carries *Bartolomeo*, an ESA platform for hosting external payloads onto ISS.[508] Originally scheduled to launch on 2 March 2020, the launch date was pushed back due to a second stage engine failure. SpaceX decided to swap out the second stage instead of replacing the faulty part.[509] It was SpaceX's 50th successful landing of a first stage booster, the third flight of the Dragon C112 and the last launch of the cargo Dragon spacecraft.								
83	18 March 2020, 12:16[510]	F9 B5 ♺ B1048.5	KSC, LC-39A	Starlink 5 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Failure (drone ship)
	Fifth operational launch of Starlink satellites. It was the first time a first stage booster flew for a fifth time and the second time the fairings were reused (Starlink flight in May 2019).[511] Towards the end of the first stage burn, the booster suffered premature shut down of an engine, the first of a Merlin 1D variant and first since the CRS-1 mission in October 2012. However, the payload still reached the targeted orbit.[512] This was the second Starlink launch booster landing failure in a row, later revealed to be caused by residual cleaning fluid trapped inside a sensor.[513]								
84	22 April 2020, 19:30[514]	F9 B5 ♺ B1051.4	KSC, LC-39A	Starlink 6 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Success (drone ship)

Methodology 3
EDA and visualization methods in details:

- Exploratory data analysis done with Python and SQL
- Visualizations were created with Seaborn, Folium, Dash







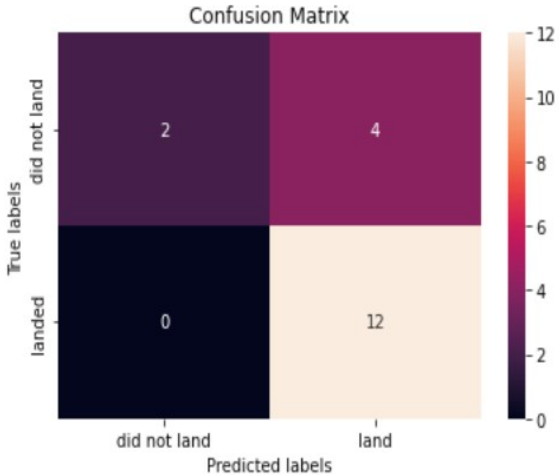
Methodology 4

Prediction methods in details:

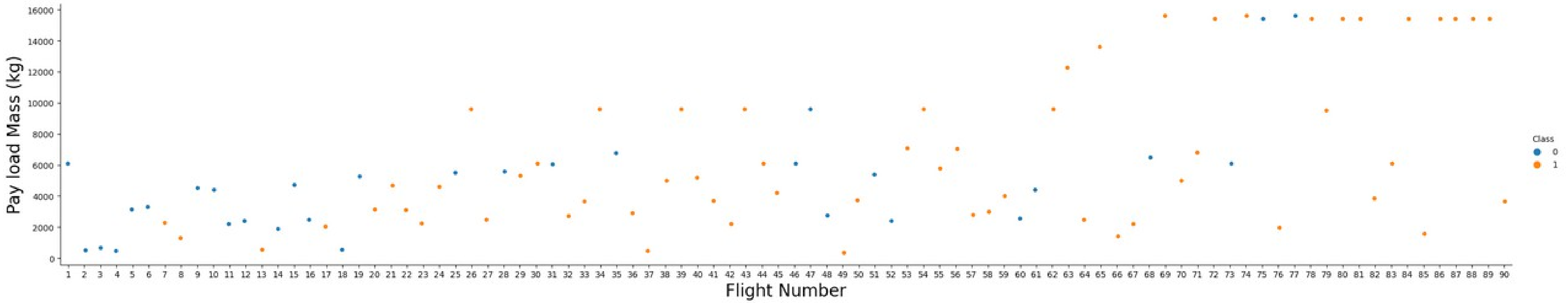
Prediction with sklearn library using following models:

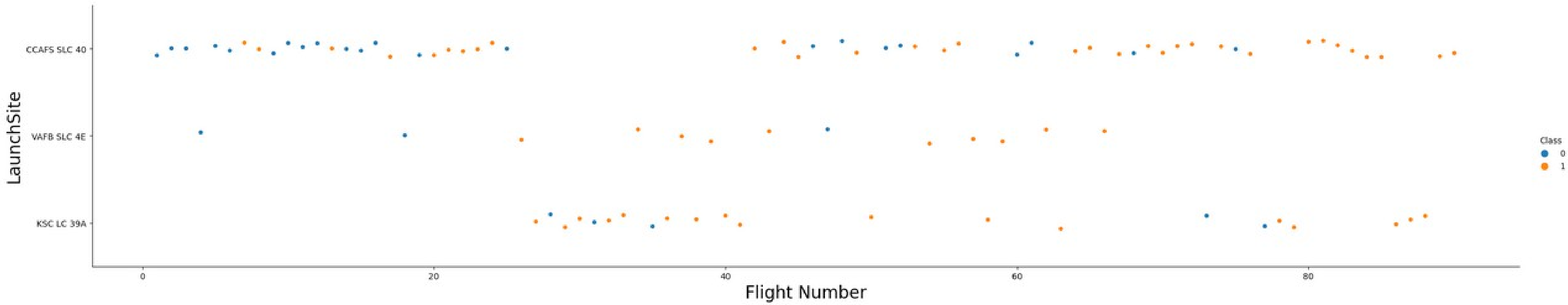
- logistic regression
- support vector machine
- decision tree classifier
- k nearest neighbors

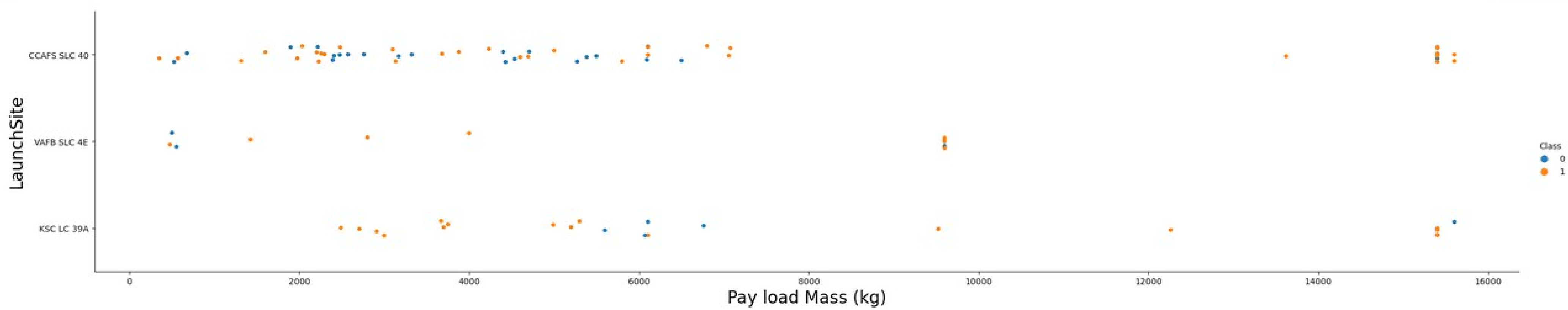
Model performance was compared with f1_score, jaccard_score Confusion Matrix



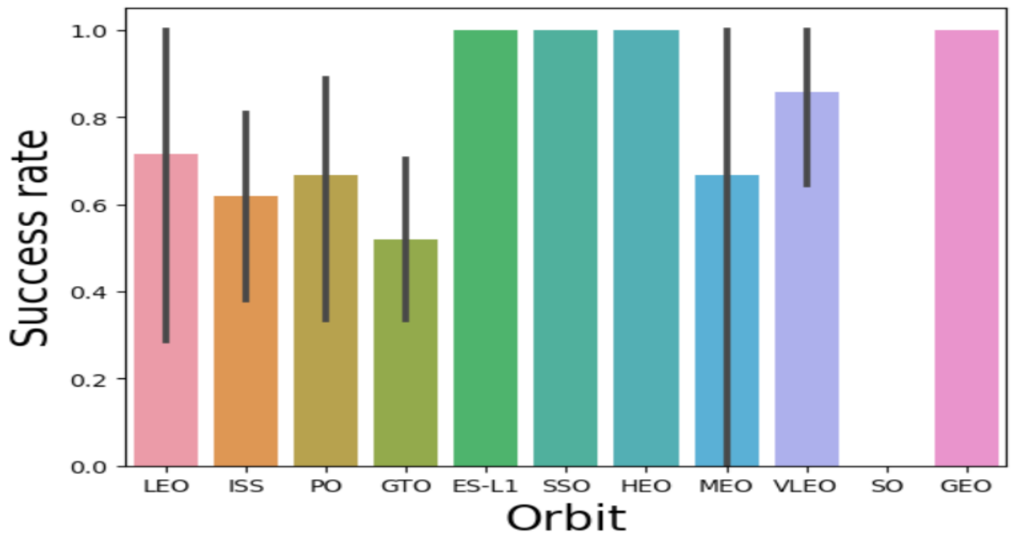
EDA with visualization results 1

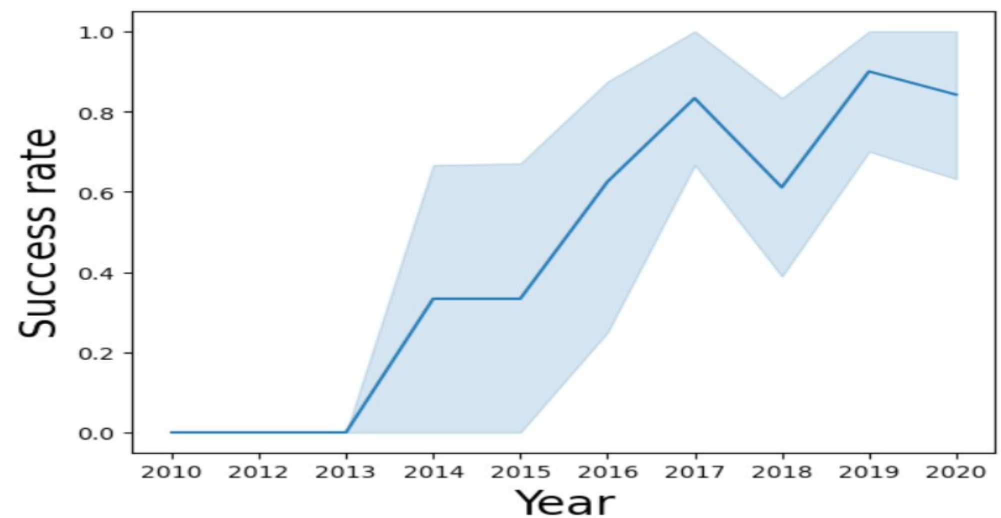


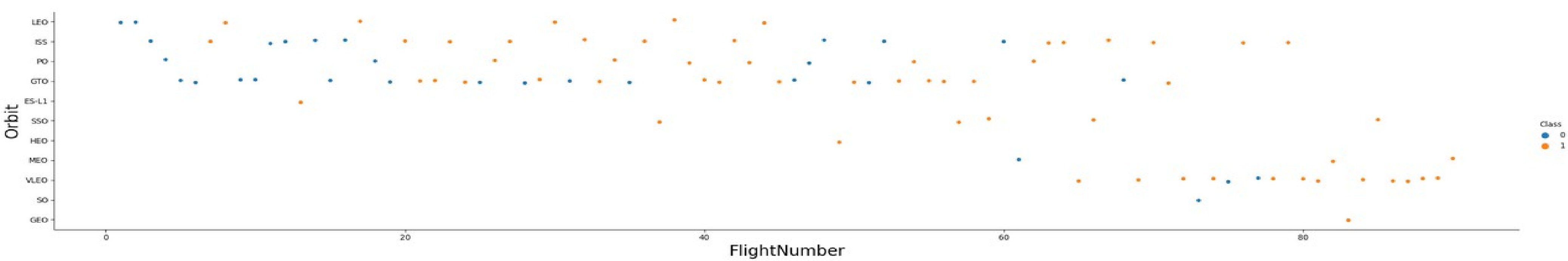


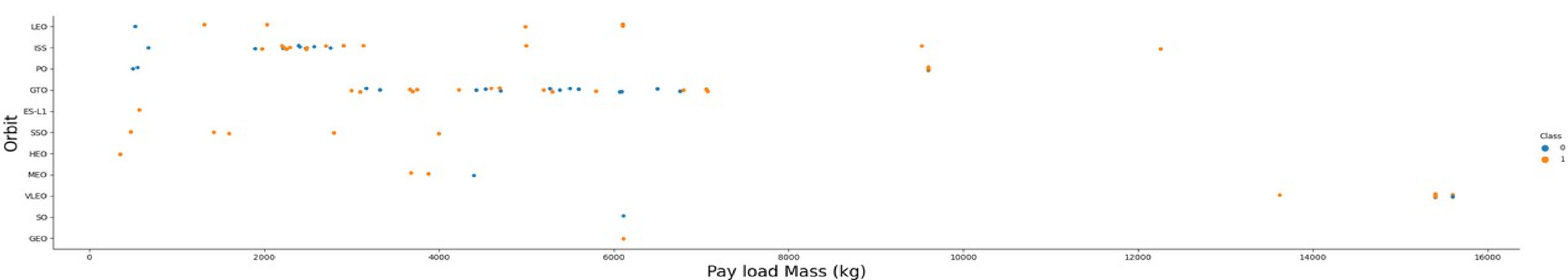


EDA with visualization results 2









EDA with SQL results 1

Task 1

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

In [5]:
```
%sql SELECT DISTINCT LAUNCH_SITE FROM TABLE
```

```
 * ibm_db_sa://mdx72742:***@3883e7e4-18f5-4afe-be8c-fa31c41761d2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31498/bludb
Done.
```

Out[5]:

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

Task 2

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

In [6]:
```
%sql SELECT * FROM TABLE WHERE LAUNCH_SITE LIKE 'CCA%' LIMIT 5
```

```
 * ibm_db_sa://mdx72742:***@3883e7e4-18f5-4afe-be8c-fa31c41761d2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31498/bludb
Done.
```

Out[6]:

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

EDA with SQL results 2

Task 3

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

In [7]:
```
%sql SELECT SUM(payload_mass__kg_) FROM TABLE WHERE CUSTOMER = 'NASA (CRS)'
```

```
 * ibm_db_sa://mdx72742:***@3883e7e4-18f5-4afe-be8c-fa31c41761d2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31498/bludb
Done.
```

Out[7]:

1
45596

Task 4

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

In [8]:
```
%sql SELECT AVG(payload_mass__kg_) FROM TABLE WHERE booster_version LIKE 'F9 v1.1%'
```

```
 * ibm_db_sa://mdx72742:***@3883e7e4-18f5-4afe-be8c-fa31c41761d2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31498/bludb
Done.
```

Out[8]:

1
2534

EDA with SQL results 3

Task 5

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

Hint:Use min function

In [9]:
```
%sql SELECT DISTINCT landing__outcome FROM TABLE
```

```
 * ibm_db_sa://mdx72742:***@3883e7e4-18f5-4afe-be8c-fa31c41761d2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31498/bludb
Done.
```

Out[9]:

landing__outcome
Controlled (ocean)
Failure
Failure (drone ship)
Failure (parachute)
No attempt
Precluded (drone ship)
Success
Success (drone ship)
Success (ground pad)
Uncontrolled (ocean)

In [10]:
```
%sql SELECT MIN(DATE) FROM TABLE WHERE landing__outcome = 'Success (ground pad)'
```

```
 * ibm_db_sa://mdx72742:***@3883e7e4-18f5-4afe-be8c-fa31c41761d2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31498/bludb
Done.
```

Out[10]:

1
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

In [11]:
```
%sql SELECT * FROM TABLE WHERE landing__outcome = 'Success (drone ship)' AND payload_mass__kg_ > 4000 AND payload_mass__kg_ <
```

```
 * ibm_db_sa://mdx72742:***@3883e7e4-18f5-4afe-be8c-fa31c41761d2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31498/bludb
Done.
```

Out[11]:

DATE	time__utc_	booster_version	launch_site	payload	payload_mass__kg_	orbit	customer	mission_outcome	landing__outcome
2016-05-06	05:21:00	F9 FT B1022	CCAFS LC-40	JCSAT-14	4696	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
2016-08-14	05:26:00	F9 FT B1026	CCAFS LC-40	JCSAT-16	4600	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
2017-03-30	22:27:00	F9 FT B1021.2	KSC LC-39A	SES-10	5300	GTO	SES	Success	Success (drone ship)
2017-10-11	22:53:00	F9 FT B1031.2	KSC LC-39A	SES-11 / EchoStar 105	5200	GTO	SES EchoStar	Success	Success (drone ship)

EDA with SQL results 4

Task 7

List the total number of successful and failure mission outcomes

In [12]:
```
%sql SELECT DISTINCT mission_outcome FROM TABLE
```

```
 * ibm_db_sa://mdx72742:***@3883e7e4-18f5-4afe-be8c-fa31c41761d2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31498/bludb
Done.
```

Out[12]:

mission_outcome
Failure (in flight)
Success
Success (payload status unclear)

In [13]:
```
%sql SELECT COUNT(*) FROM TABLE WHERE mission_outcome = 'Success'
```

```
 * ibm_db_sa://mdx72742:***@3883e7e4-18f5-4afe-be8c-fa31c41761d2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31498/bludb
Done.
```

Out[13]:

1
99

In [14]:
```
%sql SELECT COUNT(*) FROM TABLE WHERE mission_outcome LIKE 'Failure%'
```

```
 * ibm_db_sa://mdx72742:***@3883e7e4-18f5-4afe-be8c-fa31c41761d2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31498/bludb
Done.
```

Out[14]:

1

[15]:
```
%sql SELECT COUNT(*) FROM TABLE
```

```
 * ibm_db_sa://mdx72742:***@3883e7e4-18f5-4afe-be8c-fa31c41761d2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31498/bludb
Done.
```

Out[15]:

1

EDA with SQL results 5

Task 8

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

In [16]:
```
%sql SELECT booster_version FROM TABLE WHERE payload_mass__kg_ = (SELECT MAX(TABLE.payload_mass__kg_)FROM TABLE);
```

```
 * ibm_db_sa://mdx72742:***@3883e7e4-18f5-4afe-be8c-fa31c41761d2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31498/bludb
Done.
```

Out[16]:

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

EDA with SQL results 6

Task 9

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

7]:

```
%sql SELECT landing__outcome, booster_version, launch_site FROM TABLE WHERE Date LIKE '2015%'
```

```
 * ibm_db_sa://mdx72742:***@3883e7e4-18f5-4afe-be8c-fa31c41761d2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31498/bludb
Done.
```

ut[17]:

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

EDA with SQL results 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

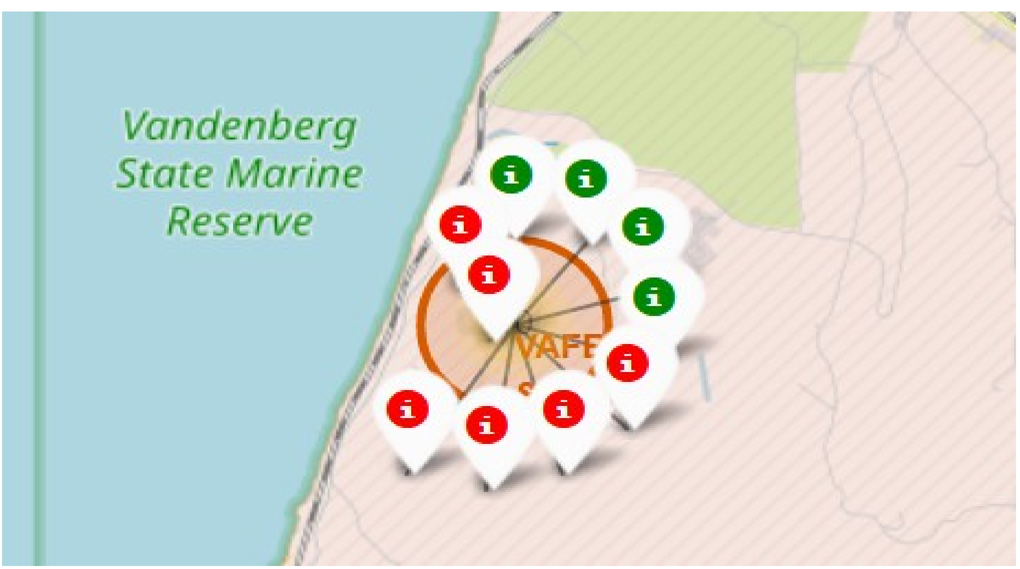
8]:
```
%sql SELECT * FROM TABLE WHERE landing__outcome = 'Failure (drone ship)' OR landing__outcome = 'Success (ground pad)' AND (DA
```

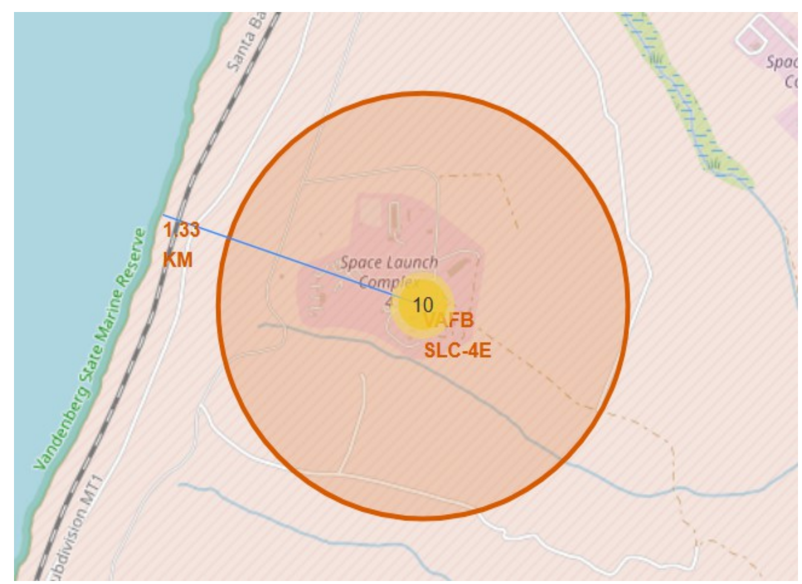
```
 * ibm_db_sa://mdx72742:***@3883e7e4-18f5-4afe-be8c-fa31c41761d2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31498/bludb
Done.
```

ut[18]:

DATE	time__utc_	booster_version	launch_site	payload	payload_mass__kg_	orbit	customer	mission_outcome	landing__outcome
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)
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)
2016-06-15	14:29:00	F9 FT B1024	CCAFS LC-40	ABS-2A Eutelsat 117 West B	3600	GTO	ABS Eutelsat	Success	Failure (drone ship)
2016-03-04	23:35:00	F9 FT B1020	CCAFS LC-40	SES-9	5271	GTO	SES	Success	Failure (drone ship)
2016-01-17	18:42:00	F9 v1.1 B1017	VAFB SLC-4E	Jason-3	553	LEO	NASA (LSP) NOAA CNES	Success	Failure (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)
2015-04-14	20:10:00	F9 v1.1 B1015	CCAFS LC-40	SpaceX CRS-6	1898	LEO (ISS)	NASA (CRS)	Success	Failure (drone ship)
2015-01-10	09:47:00	F9 v1.1 B1012	CCAFS LC-40	SpaceX CRS-5	2395	LEO (ISS)	NASA (CRS)	Success	Failure (drone ship)

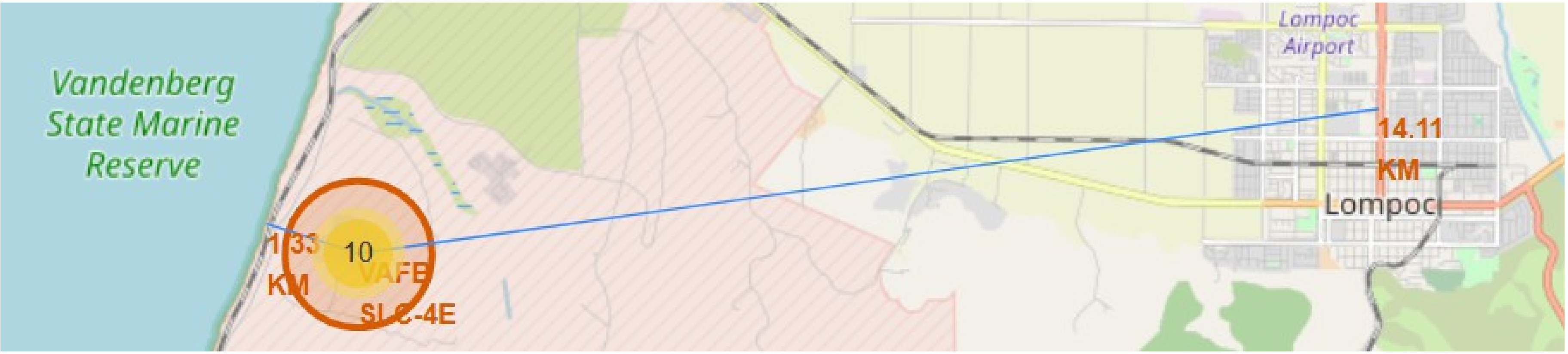
Interactive map with Folium results 1



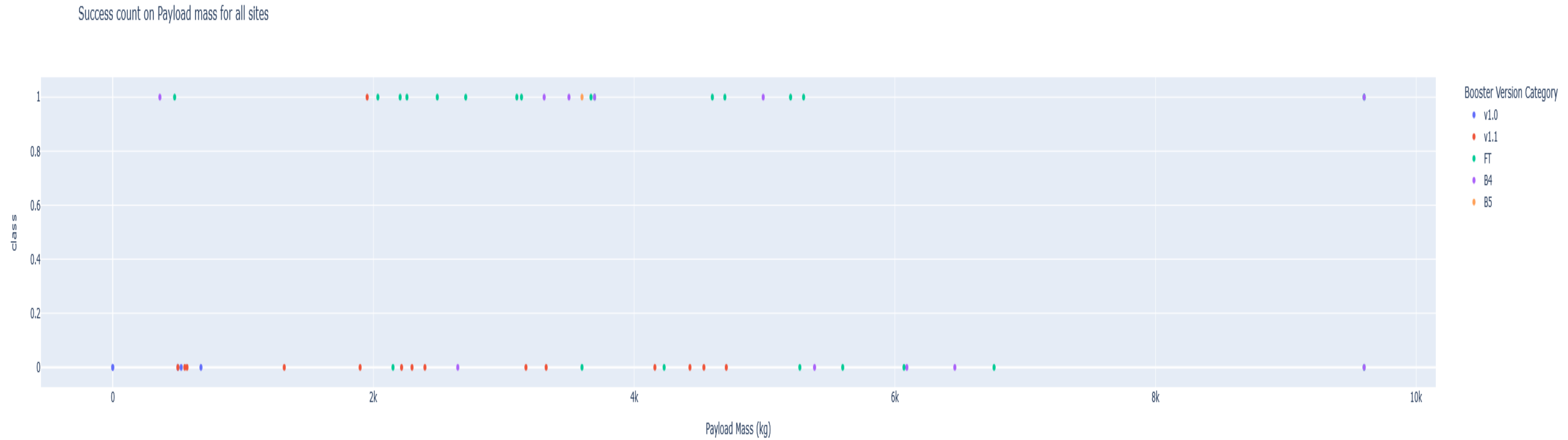


Interactive map with Folium results 2

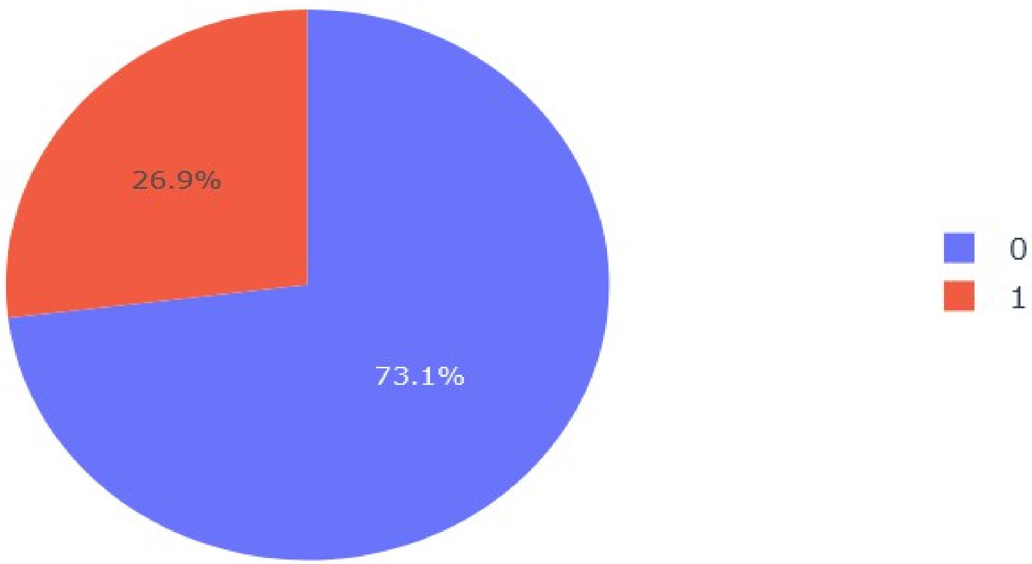
- Launch sites are close to railways, highways providing easy access for people and suppliers.
- Launch sites are close to coastline for safety reasons - easy access to water in case of a fire.
- Launch sites keep the distance away from cities to protect the citizens, for example the cosest city to VAFB launchsite is city Lompoc that is 14.11 KM away from the lauch site.

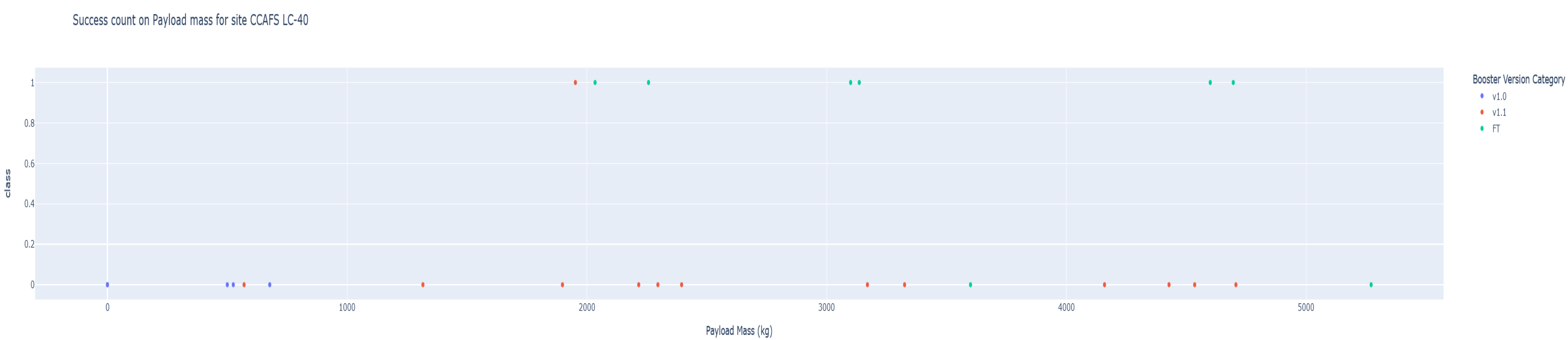


Plotly Dash dashboard results 1
Total success launches all sites

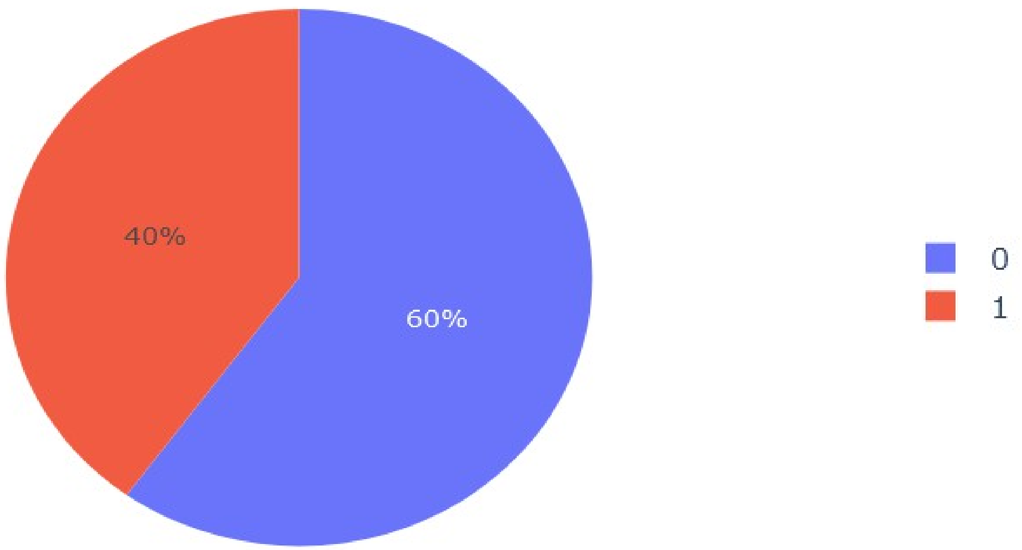


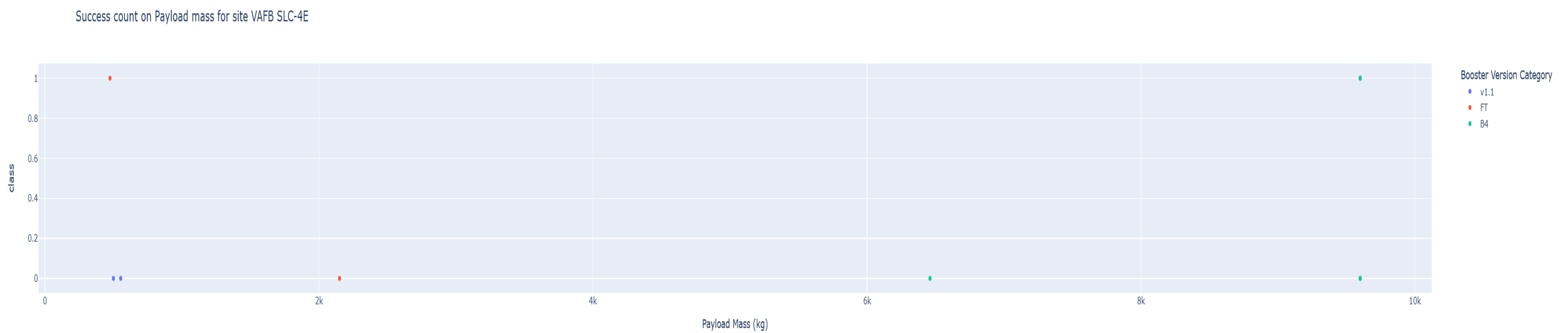
Plotly Dash dashboard results 2
CCAFS LC-40



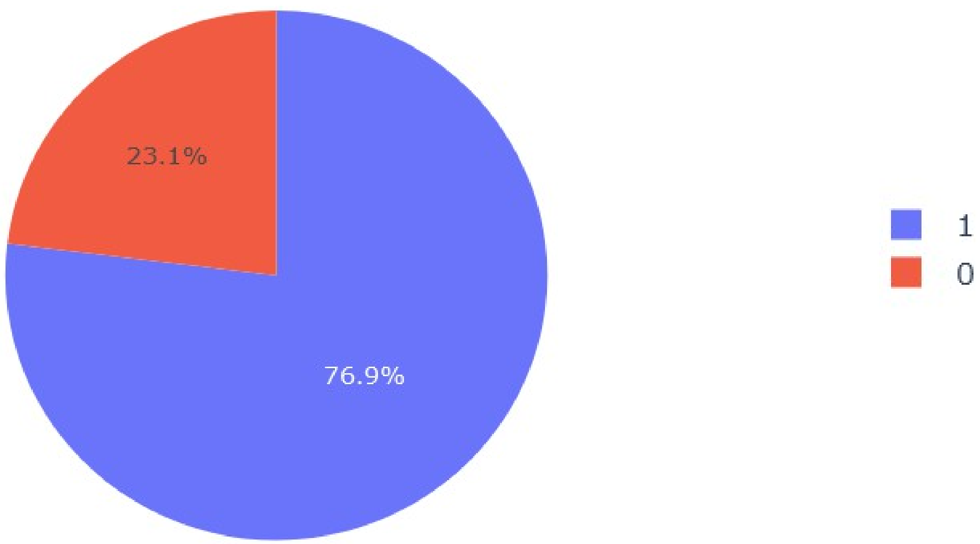


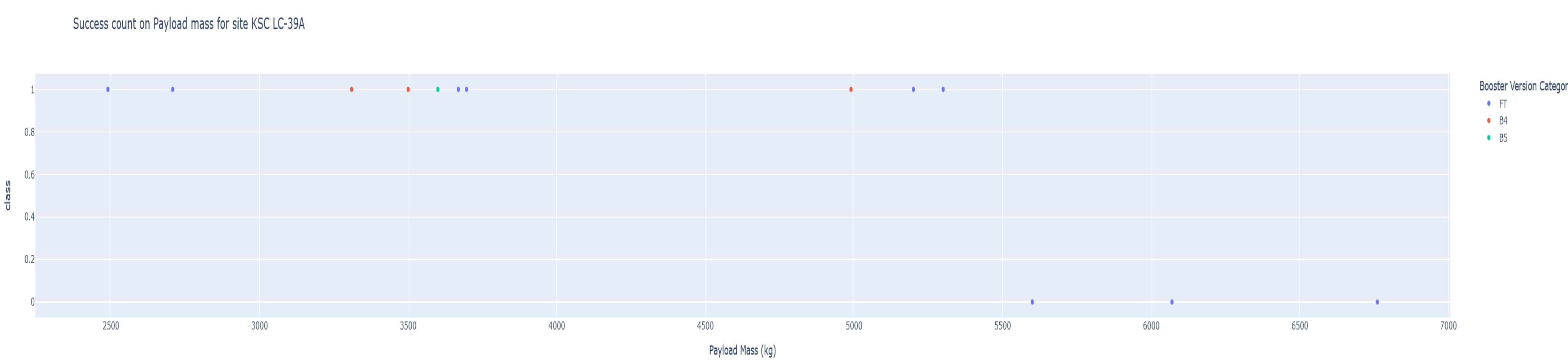
Plotly Dash dashboard results 3 VAFB SLC-4E



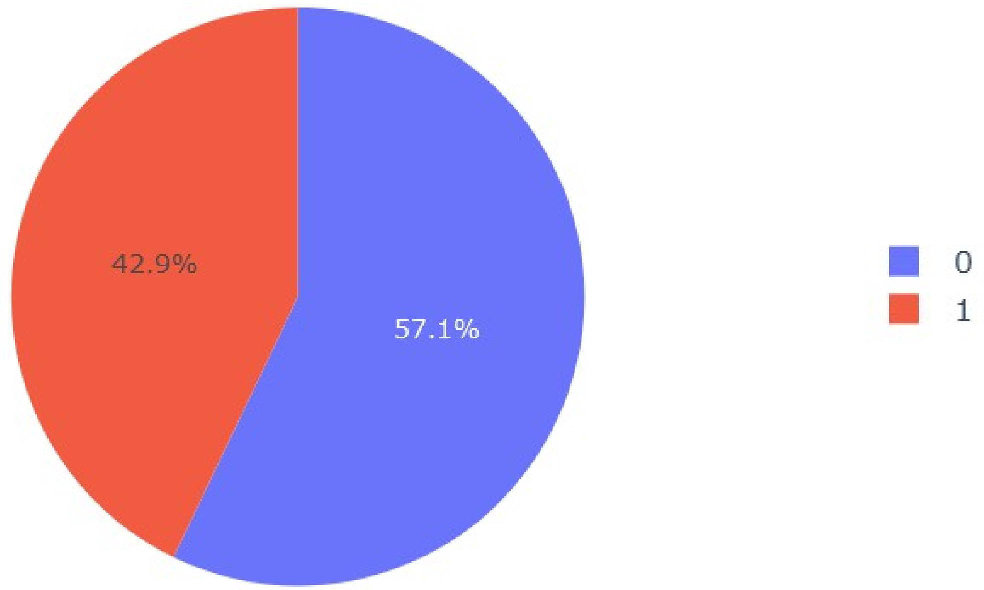


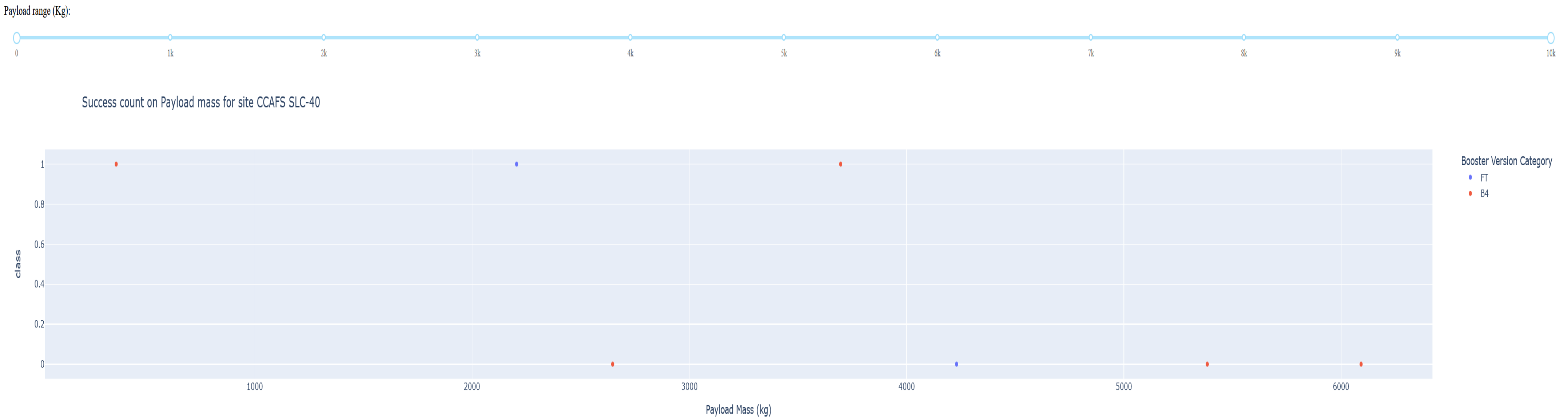
Plotly Dash dashboard results 4 KSC LC-39A





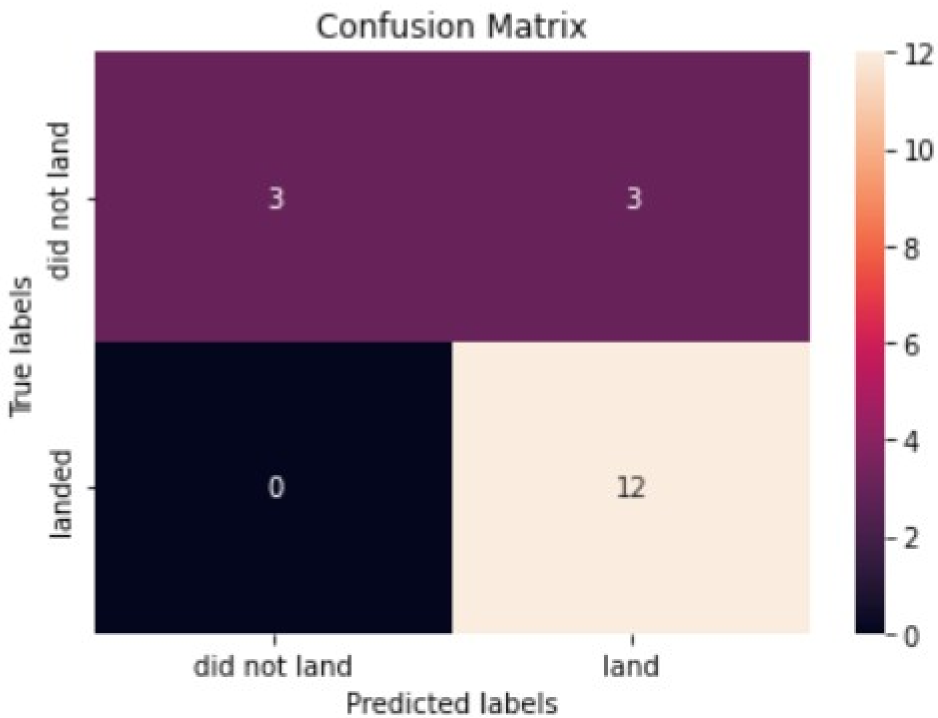
Plotly Dash dashboard results 5
CCAFS SLC-40





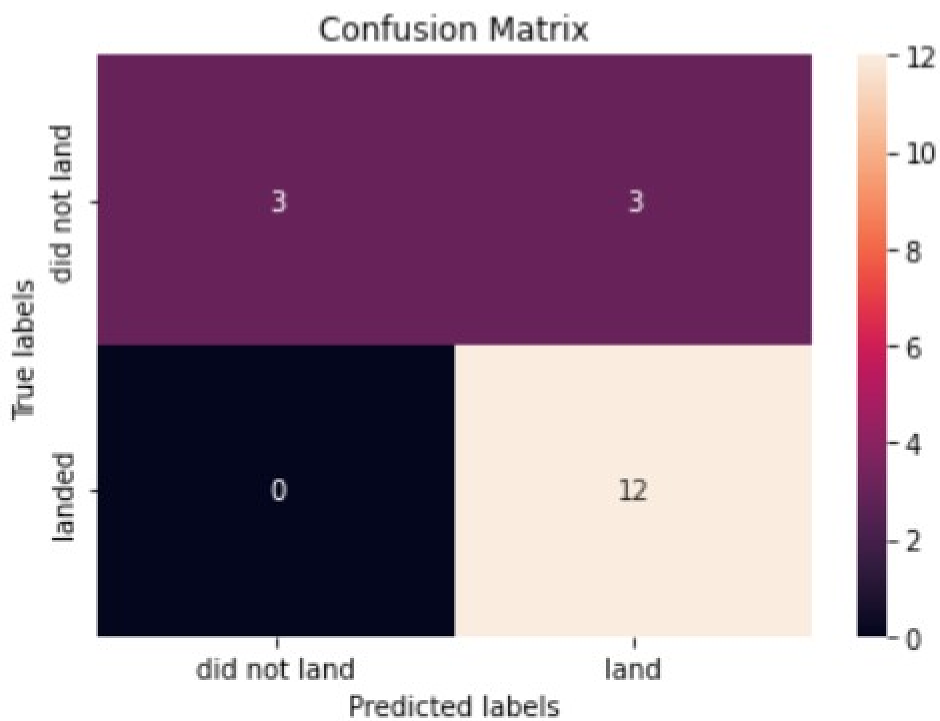
Predictive analysis (classification) results 1

- Logistic regression
- jaccard score: 0.8
- f1_score: 0.8148148148148149



Predictive analysis (classification) results 2

- support vector machine
- jaccard_score: 0.8
- f1_score: 0.8148148148148149



Predictive analysis (classification) results 3

- decision tree classifier
- jaccard_score: 0.75
- f1_score: 0.738095238095238

Predictive analysis (classification) results 4

- k nearest neighbors
- jaccard_score: 0.75
- f1_score: 0.738095238095238

Conclusion

- Launch site with the highest count of successful launches is KSC LC-39A with booster version FT
- Best method for prediction are either logistic regression or SVM, because their Jaccard index and F1 score are the highest
- The closer is the site to sea, roads and railways and the further from cities it is, the better.

Resources

- https://www.coursera.org/learn/applied-data-science-capstone
- https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches
- https://api.spacexdata.com/v4/launches/past