SpaceX Report

Author: Tes

Date Published: 4/26/2024

Contact: XYZ@gmail.com

#FOR EDUCATIONAL PURPOSES

Table of Contents

EXECUTIVE SUMMARY

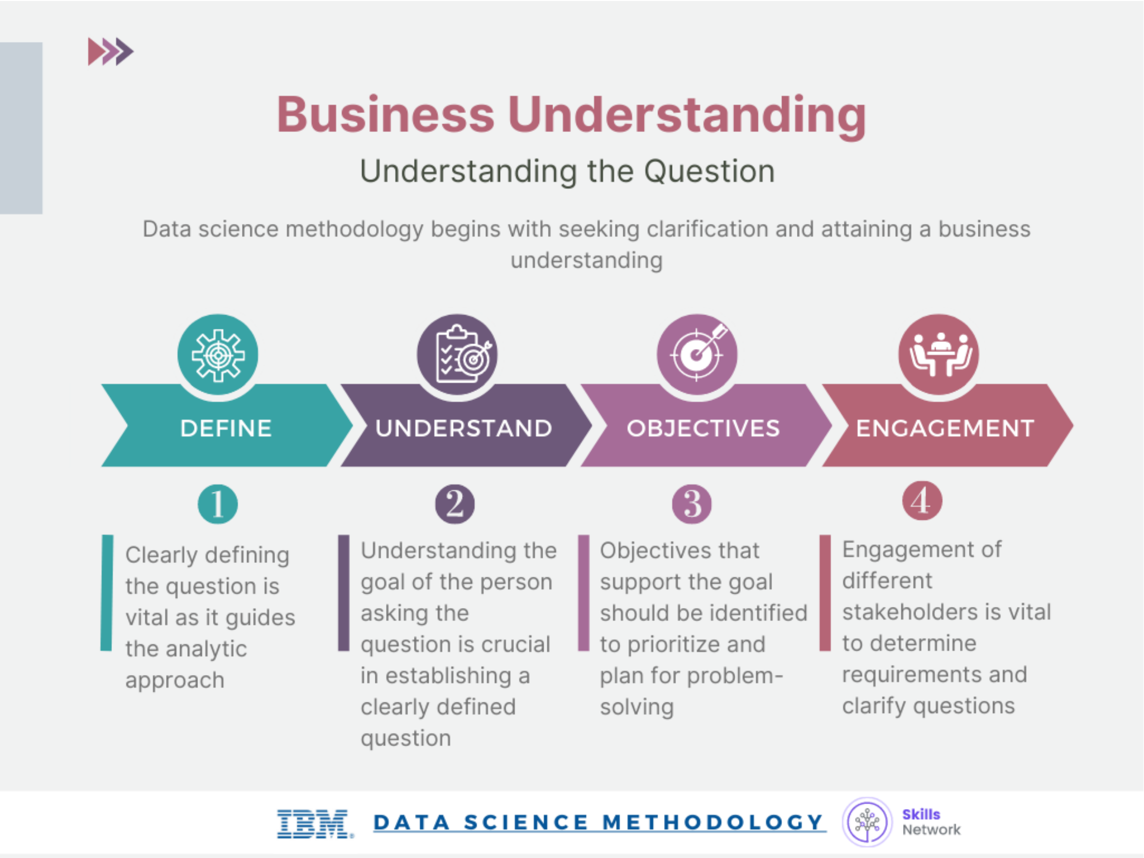
In this Project, I will predict if the Falcon 9 first stage will land successfully. SpaceX advertises Falcon 9 rocket launches on its website, with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, **much of the savings is because SpaceX can reuse the first stage**. 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 SpaceX for a rocket launch.

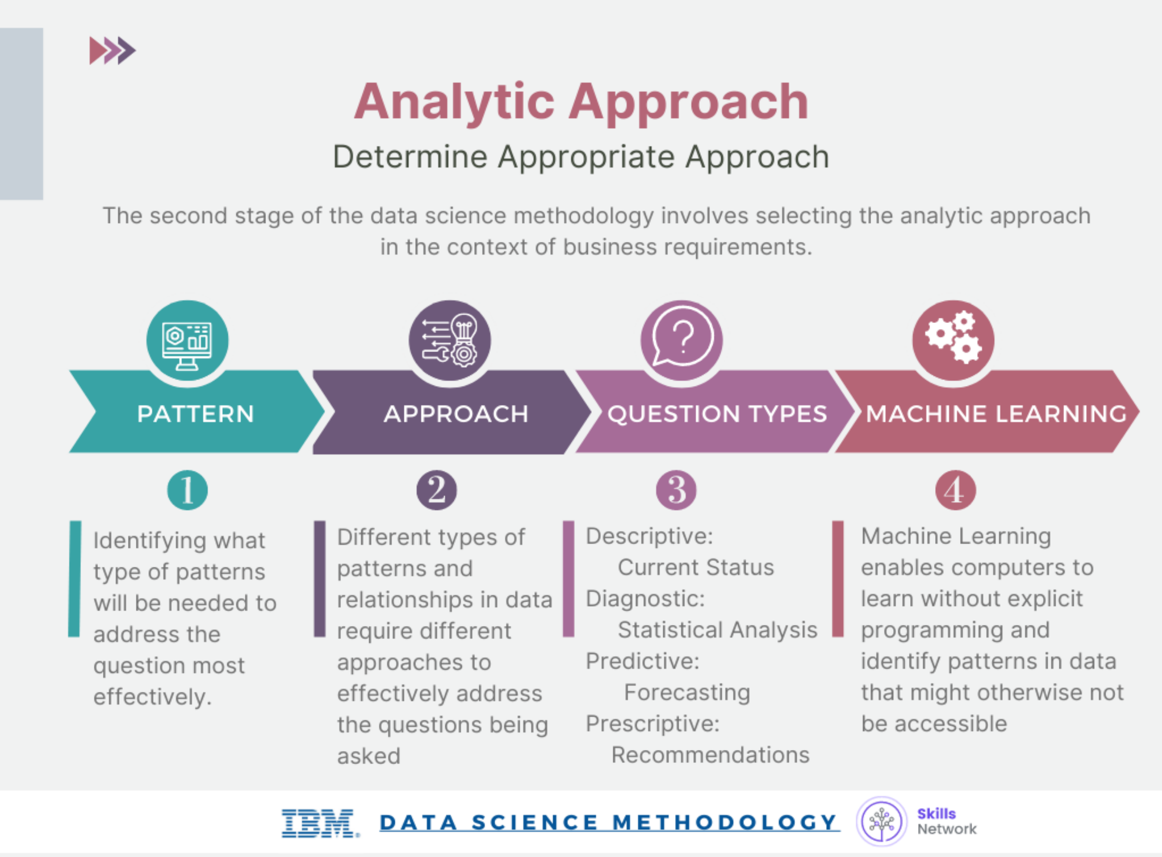
Instead of using rocket science to determine if the first stage will land successfully, I will train several different machine learning models and use public information to predict if SpaceX will reuse the first stage.

INTRODUCTION

- Methodology
  - Quick review of Business Understanding and Analytic Approach
- Methodology (Our Methodology)
  - Data Collection: I collected data using the SpaceX REST API and publicly available data. This data forms the backbone of our analysis and prediction model.
  - Data Wrangling: This step involved cleaning and organizing the data to prepare it for further analysis. This is a crucial step as the quality of data affects the performance of the model.
  - Results and Exploratory Data Analysis (EDA): After preparing the data, we conducted an EDA using SQL, Pandas, and Matplotlib. This helped us understand the patterns in the data, identify any outliers or anomalies, and form hypotheses about potential relationships between variables.

METHODOLOGY





METHODOLOGY (Our Methodology)

Objective

- Determine the price of each launch.
- Gather information about SpaceX and create a dashboard for the team
- Determine if SpaceX will reuse the first stage
- Train various ML models to predict if SpaceX will reuse the first stage

Task 1: Data Collection
- Request to the SpaceX API
- Clean the requested data

Task 2: Data Collection Using BeautifulSoup
- Request Historical data from public source
- Clean the requested data

Task 3: EDA Data Wrangling
- Perform Exploratory Data Analysis
- Determine Training Labels

Task 4: EDA with SQL
- Understand the SpaceX Data
- Load the dataset into the corresponding table in a Db2 database

Task 5: EDA with Pandas
- Exploratory Data Analysis
- Preparing Data Feature Engineering

Task 6: Launch Site Location Analysis with Folium
- Mark all launch sites on a map
- Mark the success/failed launches for each site on the map
- Calculate the distances between a launch site to its proximities

Task 7: Dashboard
- Create a Dashboard

Task 8: Model training And Evaluation
- Standardize the data
- Split into training data and test data
- Find best Hyperparameter for SVM, Classification Trees and Logistic Regression
- Find the method performs best using test data

RESULTS

With the instructions given for this project, we can say things like from Site X with Payload Y it should “Succeed” and which model brought us to this conclusion. But with the data we collected I believe the model is underfitted. We don’t have weather data(windspeeds, temperature),nor do we have data on fuel types, fuel mixture used during launches, which I believe would greatly increase accuracy.

“We’ll go into further details of our findings later on.”

Task 1: Data Collection

Data (Uncleaned)

	static_fire_date_utc	static_fire_date_unix	net	window	rocket	success	failures	details	crew	ships	capsules	payloads	launchpad	flight_number	nam
0	2006-03-17T00:00:00.000Z	1.142554e+09	False	0.0	5e9d0d95eda69955f709d1eb	False	[{'time': 33, 'altitude': None, 'reason': 'merlin engine failure'}]	Engine failure at 33 seconds and loss of vehicle	[]	[]	[]	[5eb0e4b5b6c3bb0006eeb1e1]	5e9e4502f5090995de566f86	1	FalconSa
1	None	NaN	False	0.0	5e9d0d95eda69955f709d1eb	False	[{'time': 301, 'altitude': 289, 'reason': 'harmonic oscillation leading to	Successful first stage burn and transition to second stage, maximum altitude 289 km, Premature engine shutdown	[]	[]	[]	[5eb0e4b6b6c3bb0006eeb1e2]	5e9e4502f5090995de566f86	2	DemoSa

Data (Cleaned)

FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCo	Serial	Longitude	Latitude
6	6/4/2010	Falcon 9	6123.547647	LEO	CCSFS SLC 40	None None	1	FALSE	FALSE	FALSE		1	0	B0003	-80.5774	28.5618
8	5/22/2012	Falcon 9	525	LEO	CCSFS SLC 40	None None	1	FALSE	FALSE	FALSE		1	0	B0005	-80.5774	28.5618
10	3/1/2013	Falcon 9	677	ISS	CCSFS SLC 40	None None	1	FALSE	FALSE	FALSE		1	0	B0007	-80.5774	28.5618
11	9/29/2013	Falcon 9	500	PO	VAFB SLC 4E	False Ocean	1	FALSE	FALSE	FALSE		1	0	B1003	-120.611	34.6320
12	12/3/2013	Falcon 9	3170	GTO	CCSFS SLC 40	None None	1	FALSE	FALSE	FALSE		1	0	B1004	-80.5774	28.5618
13	1/6/2014	Falcon 9	3325	GTO	CCSFS SLC 40	None None	1	FALSE	FALSE	FALSE		1	0	B1005	-80.5774	28.5618
14	4/18/2014	Falcon 9	2296	ISS	CCSFS SLC 40	True Ocean	1	FALSE	FALSE	TRUE		1	0	B1006	-80.5774	28.5618
15	7/14/2014	Falcon 9	1316	LEO	CCSFS SLC 40	True Ocean	1	FALSE	FALSE	TRUE		1	0	B1007	-80.5774	28.5618
16	8/5/2014	Falcon 9	4535	GTO	CCSFS SLC 40	None None	1	FALSE	FALSE	FALSE		1	0	B1008	-80.5774	28.5618
17	9/7/2014	Falcon 9	4428	GTO	CCSFS SLC 40	None None	1	FALSE	FALSE	FALSE		1	0	B1011	-80.5774	28.5618
18	9/21/2014	Falcon 9	2216	ISS	CCSFS SLC 40	False Ocean	1	FALSE	FALSE	FALSE		1	0	B1010	-80.5774	28.5618
19	1/10/2015	Falcon 9	2395	ISS	CCSFS SLC 40	False ASDS	1	TRUE	FALSE	TRUE	5e9e3032383ecb761	1	0	B1012	-80.5774	28.5618
20	2/11/2015	Falcon 9	570	ES-L1	CCSFS SLC 40	True Ocean	1	TRUE	FALSE	TRUE		1	0	B1013	-80.5774	28.5618
22	4/14/2015	Falcon 9	1898	ISS	CCSFS SLC 40	False ASDS	1	TRUE	FALSE	TRUE	5e9e3032383ecb761	1	0	B1015	-80.5774	28.5618
23	4/27/2015	Falcon 9	4707	GTO	CCSFS SLC 40	None None	1	FALSE	FALSE	FALSE		1	0	B1016	-80.5774	28.5618
24	6/28/2015	Falcon 9	2477	ISS	CCSFS SLC 40	None ASDS	1	TRUE	FALSE	TRUE	5e9e3032383ecb6bb	1	0	B1018	-80.5774	28.5618
25	12/22/2015	Falcon 9	2034	LEO	CCSFS SLC 40	True RTLS	1	TRUE	FALSE	TRUE	5e9e3032383ecb267	1	0	B1019	-80.5774	28.5618
26	1/17/2016	Falcon 9	553	PO	VAFB SLC 4E	False ASDS	1	TRUE	FALSE	TRUE	5e9e3033383ecbb9e	1	0	B1017	-120.611	34.6320
27	3/4/2016	Falcon 9	5271	GTO	CCSFS SLC 40	False ASDS	1	TRUE	FALSE	TRUE	5e9e3032383ecb6bb	1	0	B1020	-80.5774	28.5618
28	4/8/2016	Falcon 9	3136	ISS	CCSFS SLC 40	True ASDS	1	TRUE	FALSE	TRUE	5e9e3032383ecb6bb	2	1	B1021	-80.5774	28.5618
29	5/6/2016	Falcon 9	4696	GTO	CCSFS SLC 40	True ASDS	1	TRUE	FALSE	TRUE	5e9e3032383ecb6bb	2	0	B1022	-80.5774	28.5618

Task 2: Data Collection Using BeautifulSoup

Data (Uncleaned)

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)

Data (Cleaned)

Flight No.	Launch site	Payload	Payload mass	Orbit	Customer	Launch outcome	Version Booster	Booster landing	Date	Time
1	CCAFS	Dragon Spacecraft Qualif	0	LEO	SpaceX		F9 v1.1	Failure	4-Jun-10	18:45
1	CCAFS	Dragon	0	LEO	NASA	Success	F9 v1.1	Failure	4-Jun-10	18:45
2	CCAFS	Dragon	525 kg	LEO	NASA	Success	F9 v1.1		8-Dec-10	15:43
3	CCAFS	SpaceX CRS-1	4,700 kg	LEO	NASA		F9 v1.1	No attempt	22-May-12	7:44
4	CCAFS	SpaceX CRS-2	4,877 kg	LEO	NASA		F9 v1.1		8-Oct-12	0:35
5	VAFB	CASSIOPE	500 kg	Polar orbit	MDA	Success	F9 v1.1	Uncontrolled	1-Mar-13	15:10
6	CCAFS	SES-8	3,170 kg	GTO	SES	Success	F9 v1.1	No attempt	29-Sep-13	16:00
7	CCAFS	Thaicom 6	3,325 kg	GTO	Thaicom	Success	F9 v1.1	No attempt	3-Dec-13	22:41
8	Cape Canaveral	SpaceX CRS-3	2,296 kg	LEO	NASA		F9 v1.1	Controlled	6-Jan-14	22:06
9	Cape Canaveral	Orbcomm-OG2	1,316 kg	LEO	Orbcomm	Success	F9 v1.1	Controlled	18-Apr-14	19:25
10	Cape Canaveral	AsiaSat 8	4,535 kg	GTO	AsiaSat	Success	F9 v1.1	No attempt	14-Jul-14	15:15
11	Cape Canaveral	AsiaSat 6	4,428 kg	GTO	AsiaSat	Success	F9 v1.1		5-Aug-14	8:00
12	Cape Canaveral	SpaceX CRS-4	2,216 kg	LEO	NASA	Success	F9 v1.1	Uncontrolled	7-Sep-14	5:00
13	Cape Canaveral	SpaceX CRS-5	2,395 kg	LEO	NASA	Success	F9 v1.1	Failure	21-Sep-14	5:52
14	Cape Canaveral	DSCOVR	570 kg	HEO	USAF		F9 v1.1	Controlled	10-Jan-15	9:47
15	Cape Canaveral	ABS-3A	4,159 kg	GTO	ABS		F9 v1.1	No attempt	11-Feb-15	23:03
16	Cape Canaveral	SpaceX CRS-6	1,898 kg	LEO	NASA		F9 v1.1	Failure	2-Mar-15	3:50
17	Cape Canaveral	TÃ¼rkmenÃ„lem 52Â°E /	4,707 kg	GTO			F9 v1.1	No attempt	14-Apr-15	20:10
18	Cape Canaveral	SpaceX CRS-7	1,952 kg	LEO	NASA	Failure	F9 v1.1	Precluded	27-Apr-15	23:03
19	Cape Canaveral	Orbcomm-OG2	2,034 kg	LEO	Orbcomm		F9 v1.1	Success	28-Jun-15	14:21
20	VAFB	Jason-3	553 kg	LEO	NASA		F9 v1.1	Failure	22-Dec-15	1:29

Task 3: EDA Data Wrangling

Data (Uncleaned)

Flight No.	Launch site	Payload	Payload mass	Orbit	Customer	Launch outcome	Version Booster	Booster landing	Date	Time
1	CCAFS	Dragon Spacecraft Qualif	0	LEO	SpaceX		F9 v1.1	Failure	4-Jun-10	18:45
1	CCAFS	Dragon	0	LEO	NASA	Success	F9 v1.1	Failure	4-Jun-10	18:45
2	CCAFS	Dragon	525 kg	LEO	NASA	Success	F9 v1.1		8-Dec-10	15:43
3	CCAFS	SpaceX CRS-1	4,700 kg	LEO	NASA		F9 v1.1	No attempt	22-May-12	7:44
4	CCAFS	SpaceX CRS-2	4,877 kg	LEO	NASA		F9 v1.1		8-Oct-12	0:35
5	VAFB	CASSIOPE	500 kg	Polar orbit	MDA	Success	F9 v1.1	Uncontrolled	1-Mar-13	15:10
6	CCAFS	SES-8	3,170 kg	GTO	SES	Success	F9 v1.1	No attempt	29-Sep-13	16:00
7	CCAFS	Thaicom 6	3,325 kg	GTO	Thaicom	Success	F9 v1.1	No attempt	3-Dec-13	22:41
8	Cape Canaveral	SpaceX CRS-3	2,296 kg	LEO	NASA		F9 v1.1	Controlled	6-Jan-14	22:06
9	Cape Canaveral	Orbcomm-OG2	1,316 kg	LEO	Orbcomm	Success	F9 v1.1	Controlled	18-Apr-14	19:25
10	Cape Canaveral	AsiaSat 8	4,535 kg	GTO	AsiaSat	Success	F9 v1.1	No attempt	14-Jul-14	15:15
11	Cape Canaveral	AsiaSat 6	4,428 kg	GTO	AsiaSat	Success	F9 v1.1		5-Aug-14	8:00
12	Cape Canaveral	SpaceX CRS-4	2,216 kg	LEO	NASA	Success	F9 v1.1	Uncontrolled	7-Sep-14	5:00
13	Cape Canaveral	SpaceX CRS-5	2,395 kg	LEO	NASA	Success	F9 v1.1	Failure	21-Sep-14	5:52
14	Cape Canaveral	DSCOVR	570 kg	HEO	USAF		F9 v1.1	Controlled	10-Jan-15	9:47
15	Cape Canaveral	ABS-3A	4,159 kg	GTO	ABS		F9 v1.1	No attempt	11-Feb-15	23:03
16	Cape Canaveral	SpaceX CRS-6	1,898 kg	LEO	NASA		F9 v1.1	Failure	2-Mar-15	3:50
17	Cape Canaveral	TÃ¼rkmenÃ„lem 52Â°E /	4,707 kg	GTO			F9 v1.1	No attempt	14-Apr-15	20:10
18	Cape Canaveral	SpaceX CRS-7	1,952 kg	LEO	NASA	Failure	F9 v1.1	Precluded	27-Apr-15	23:03
19	Cape Canaveral	Orbcomm-OG2	2,034 kg	LEO	Orbcomm		F9 v1.1	Success	28-Jun-15	14:21
20	VAFB	Jason-3	553 kg	LEO	NASA		F9 v1.1	Failure	22-Dec-15	1:29

Data (Cleaned)

FlightNumber	Date	BoosterVe	PayloadM	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPa	Block	ReusedCo	Serial	Longitude	Latitude	Class
1	6/4/2010	Falcon 9	6104.959	LEO	CCAFS SLC 40	None None	1	FALSE	FALSE	FALSE		1	0	B0003	-80.577366	28.5618571	0
2	5/22/2012	Falcon 9	525	LEO	CCAFS SLC 40	None None	1	FALSE	FALSE	FALSE		1	0	B0005	-80.577366	28.5618571	0
3	3/1/2013	Falcon 9	677	ISS	CCAFS SLC 40	None None	1	FALSE	FALSE	FALSE		1	0	B0007	-80.577366	28.5618571	0
4	9/29/2013	Falcon 9	500	PO	VAFB SLC 4E	False Ocean	1	FALSE	FALSE	FALSE		1	0	B1003	-120.610829	34.632093	0
5	12/3/2013	Falcon 9	3170	GTO	CCAFS SLC 40	None None	1	FALSE	FALSE	FALSE		1	0	B1004	-80.577366	28.5618571	0
6	1/6/2014	Falcon 9	3325	GTO	CCAFS SLC 40	None None	1	FALSE	FALSE	FALSE		1	0	B1005	-80.577366	28.5618571	0
7	4/18/2014	Falcon 9	2296	ISS	CCAFS SLC 40	True Ocean	1	FALSE	FALSE	TRUE		1	0	B1006	-80.577366	28.5618571	1
8	7/14/2014	Falcon 9	1316	LEO	CCAFS SLC 40	True Ocean	1	FALSE	FALSE	TRUE		1	0	B1007	-80.577366	28.5618571	1
9	8/5/2014	Falcon 9	4535	GTO	CCAFS SLC 40	None None	1	FALSE	FALSE	FALSE		1	0	B1008	-80.577366	28.5618571	0
10	9/7/2014	Falcon 9	4428	GTO	CCAFS SLC 40	None None	1	FALSE	FALSE	FALSE		1	0	B1011	-80.577366	28.5618571	0
11	9/21/2014	Falcon 9	2216	ISS	CCAFS SLC 40	False Ocean	1	FALSE	FALSE	FALSE		1	0	B1010	-80.577366	28.5618571	0
12	1/10/2015	Falcon 9	2395	ISS	CCAFS SLC 40	False ASDS	1	TRUE	FALSE	TRUE	5e9e3032	1	0	B1012	-80.577366	28.5618571	0
13	2/11/2015	Falcon 9	570	ES-L1	CCAFS SLC 40	True Ocean	1	TRUE	FALSE	TRUE		1	0	B1013	-80.577366	28.5618571	1
14	4/14/2015	Falcon 9	1898	ISS	CCAFS SLC 40	False ASDS	1	TRUE	FALSE	TRUE	5e9e3032	1	0	B1015	-80.577366	28.5618571	0
15	4/27/2015	Falcon 9	4707	GTO	CCAFS SLC 40	None None	1	FALSE	FALSE	FALSE		1	0	B1016	-80.577366	28.5618571	0
16	6/28/2015	Falcon 9	2477	ISS	CCAFS SLC 40	None ASDS	1	TRUE	FALSE	TRUE	5e9e3032	1	0	B1018	-80.577366	28.5618571	0
17	12/22/2015	Falcon 9	2034	LEO	CCAFS SLC 40	True RTLS	1	TRUE	FALSE	TRUE	5e9e3032	1	0	B1019	-80.577366	28.5618571	1
18	1/17/2016	Falcon 9	553	PO	VAFB SLC 4E	False ASDS	1	TRUE	FALSE	TRUE	5e9e3033	1	0	B1017	-120.610829	34.632093	0
19	3/4/2016	Falcon 9	5271	GTO	CCAFS SLC 40	False ASDS	1	TRUE	FALSE	TRUE	5e9e3032	1	0	B1020	-80.577366	28.5618571	0
20	4/8/2016	Falcon 9	3136	ISS	CCAFS SLC 40	True ASDS	1	TRUE	FALSE	TRUE	5e9e3032	2	1	B1021	-80.577366	28.5618571	1
21	5/6/2016	Falcon 9	4696	GTO	CCAFS SLC 40	True ASDS	1	TRUE	FALSE	TRUE	5e9e3032	2	0	B1022	-80.577366	28.5618571	1

Task 4: EDA with SQL

Q1: Display the names of the unique launch sites in the space mission

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

```
%sql select distinct "Launch_Site" from SPACEXTABLE
```

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

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

Q2: Display 5 records where launch sites begin with the string 'CCA'

```
%sql select  "Launch_Site" from SPACEXTABLE where "Launch_Site" like 'CCA%' limit 5
```

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

Launch_Site
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40

Q3: Display the total payload mass carried by boosters launched by NASA (CRS)

```
%sql select SUM("PAYLOAD_MASS__KG_") from SPACEXTABLE where "Customer" like 'NASA (CRS)'
```

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

SUM("PAYLOAD_MASS__KG_")
45596

Q4: Display average payload mass carried by booster version F9 v1.1

```
%sql select AVG("PAYLOAD_MASS__KG_") from SPACEXTABLE where "Booster_Version" == 'F9 v1.1'
```

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

AVG("PAYLOAD_MASS__KG_")
2928.4

Q5: List the date when the first successful landing outcome in ground pad was achieved.

Hint: Use min function

```
%sql select MIN("Date") from SPACEXTABLE where "Landing_Outcome" == "Success (ground pad)"
```

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

MIN("Date")
2015-12-22

Q6: List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

```
%sql select "Booster_Version" from SPACEXTABLE where "PAYLOAD_MASS__KG_" > 4000 and "PAYLOAD_MASS__KG_" < 6000 and  "Landing_Outcome" = "Success (drone ship)"
```

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

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

Q7: List the total number of successful and failure mission outcomes

```
%sql SELECT SUM(CASE WHEN "Mission_Outcome" LIKE 'Success%' THEN 1 ELSE 0 END) AS "Successful Missions",SUM(CASE WHEN "Mission_Outcome" LIKE 'Failure%' THEN 1 ELSE 0 END) AS "Failed Missions" FROM SPACEXTABLE;
```

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

Successful Missions	Failed Missions
100	1

Q8: List the names of the booster_versions which have carried the maximum payload mass. Use a subquery

```
%sql SELECT "Booster_Version" FROM SPACEXTABLE WHERE "PAYLOAD_MASS__KG_" = (SELECT MAX("PAYLOAD_MASS__KG_") FROM SPACEXTABLE);
```

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

Q9: List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.

Note: SQLLite does not support monthnames. So you need to use substr(Date, 6,2) as month to get the months and substr(Date,0,5)='2015' for year.

```
%sql SELECT SUBSTR("Date", 6, 2) AS "Month", "Booster_Version", "Launch_Site", "Landing_Outcome" FROM SPACEXTABLE WHERE "Landing_Outcome" LIKE 'Failure (drone ship)' AND SUBSTR("Date", 0, 5) = '2015';
```

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

Month	Booster_Version	Launch_Site	Landing_Outcome
01	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
04	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

Q10: 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.

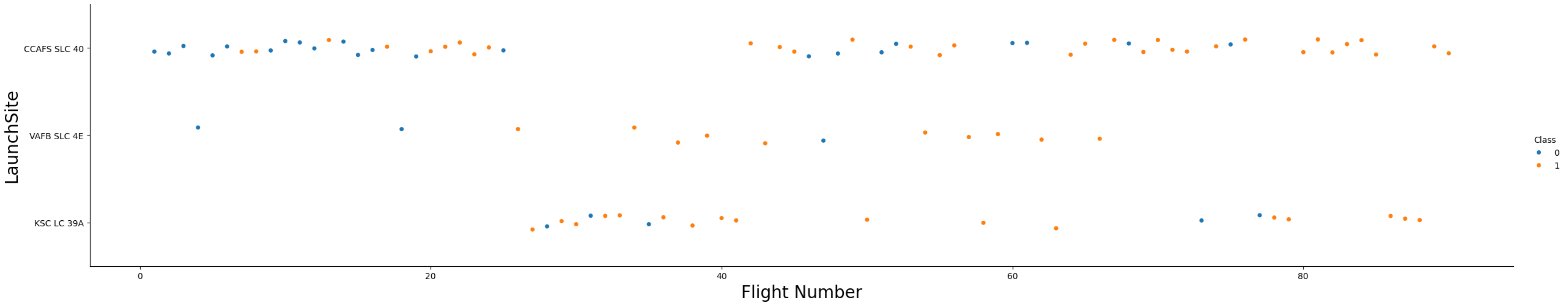
```
%sql SELECT "Landing_Outcome", COUNT(*) AS "Count"FROM SPACEXTABLE WHERE "Date" BETWEEN '2010-06-04' AND '2017-03-20'GROUP BY "Landing_Outcome"ORDER BY "Count" DESC;
```

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

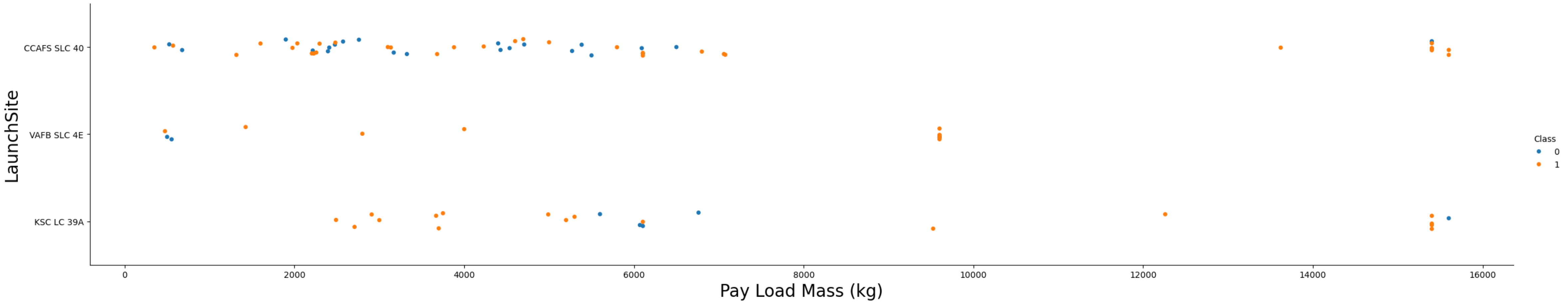
Landing_Outcome	Count
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Task 5: EDA with Pandas and Matplotlib

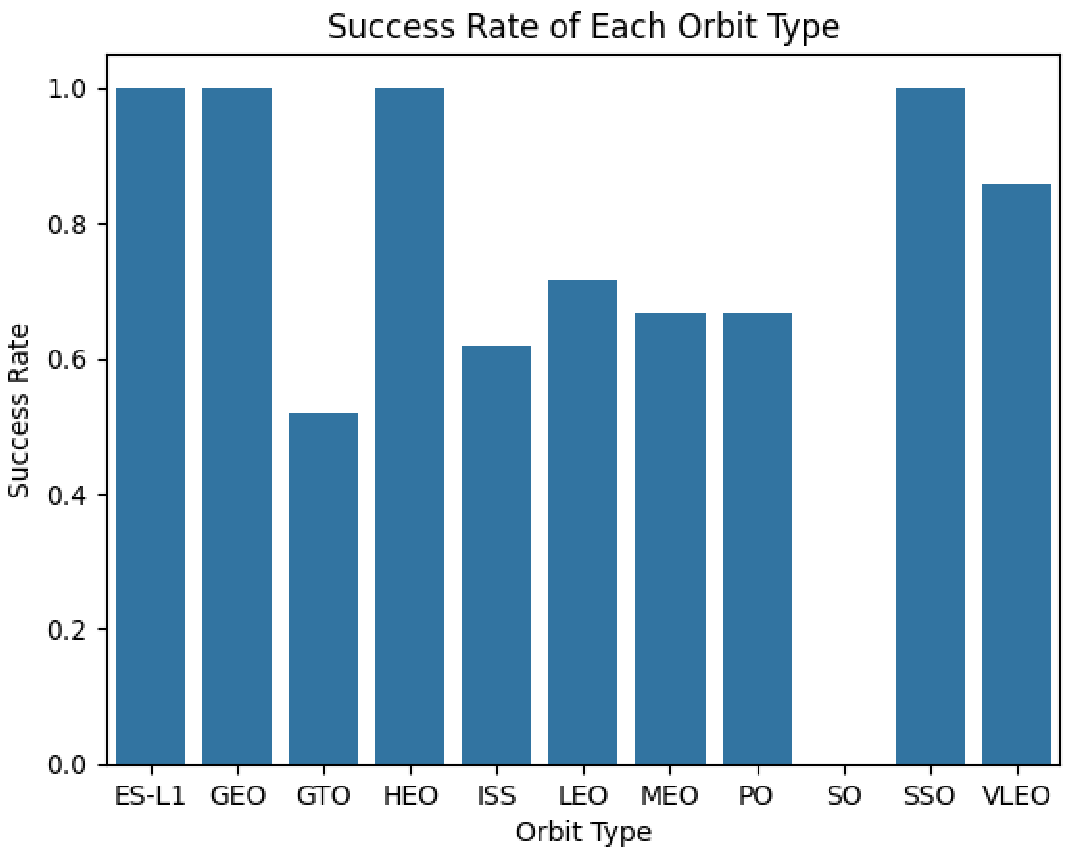
Visualize Data 1: Visualize the relationship between Flight Number and Launch Site



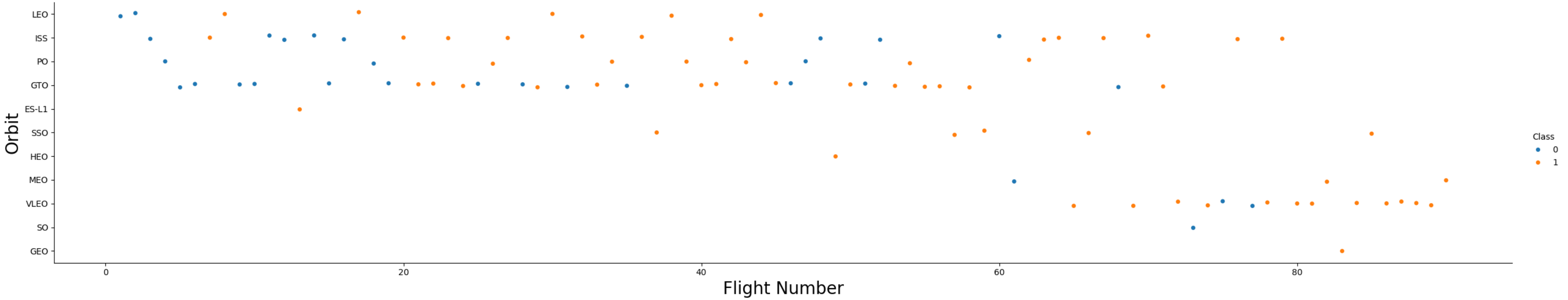
Visualize Data 2: Visualize the relationship between Payload and Launch Site



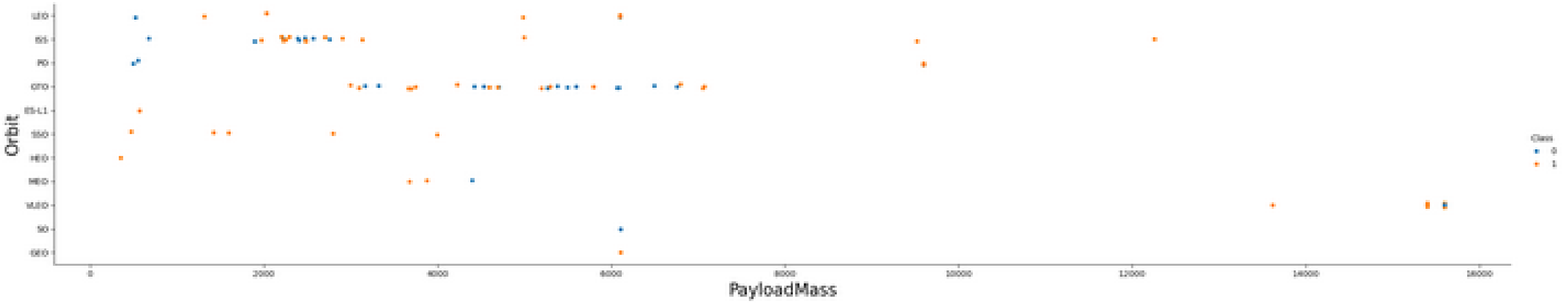
Visualize Data 3: Visualize the relationship between success rate of each orbit type



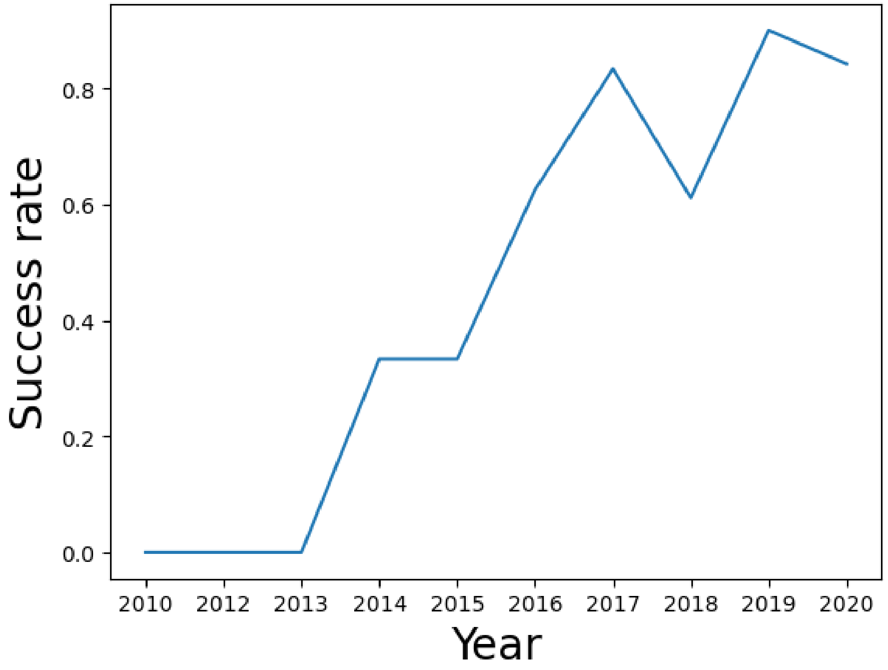
Visualize Data 4: Visualize the relationship between FlightNumber and Orbit type



Visualize Data 5: Visualize the relationship between Payload and Orbit type



Visualize Data 6: Visualize the launch success yearly trend

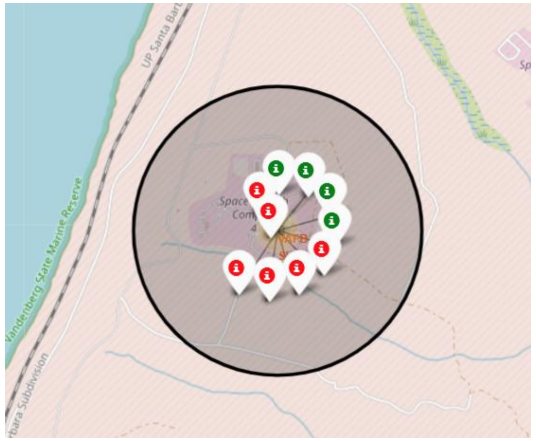


Task 6: Launch Site Location Analysis with Folium

Step 1: Mark all launch sites on a map

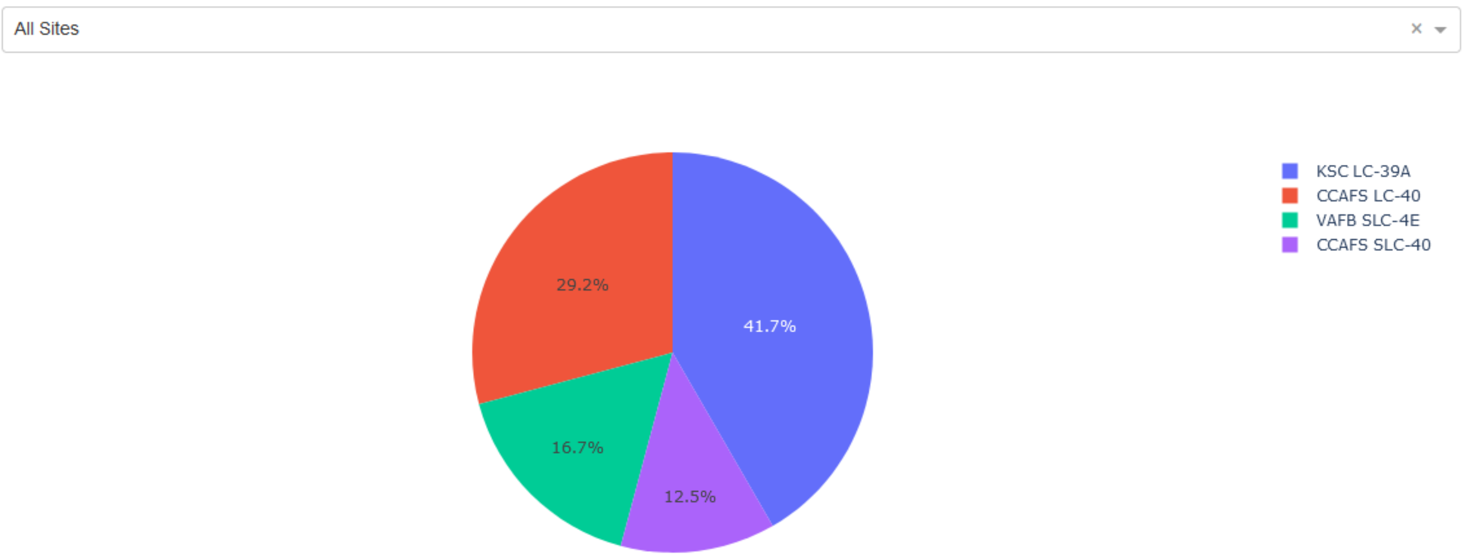
Step 2: Mark the success/failed launches for each site on the map

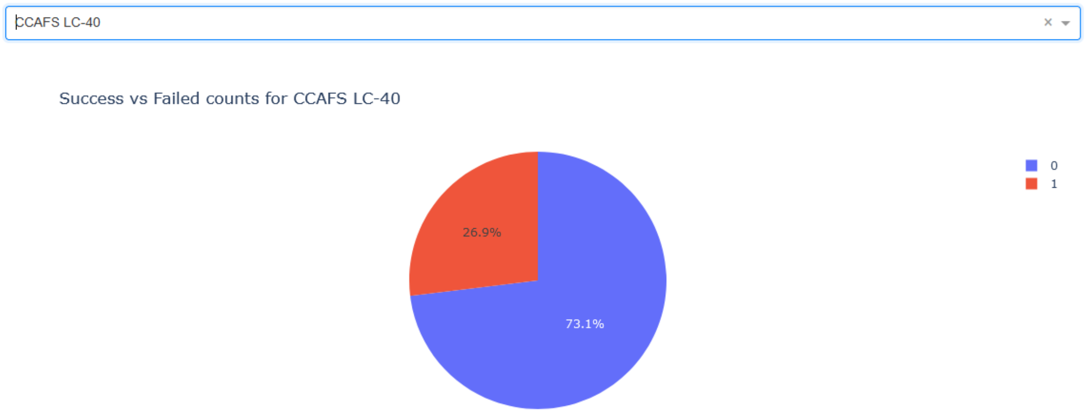
West Coast

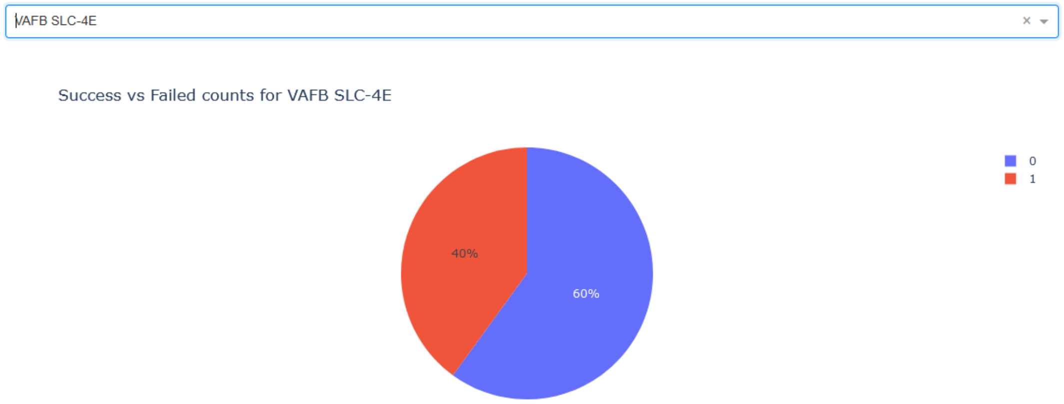


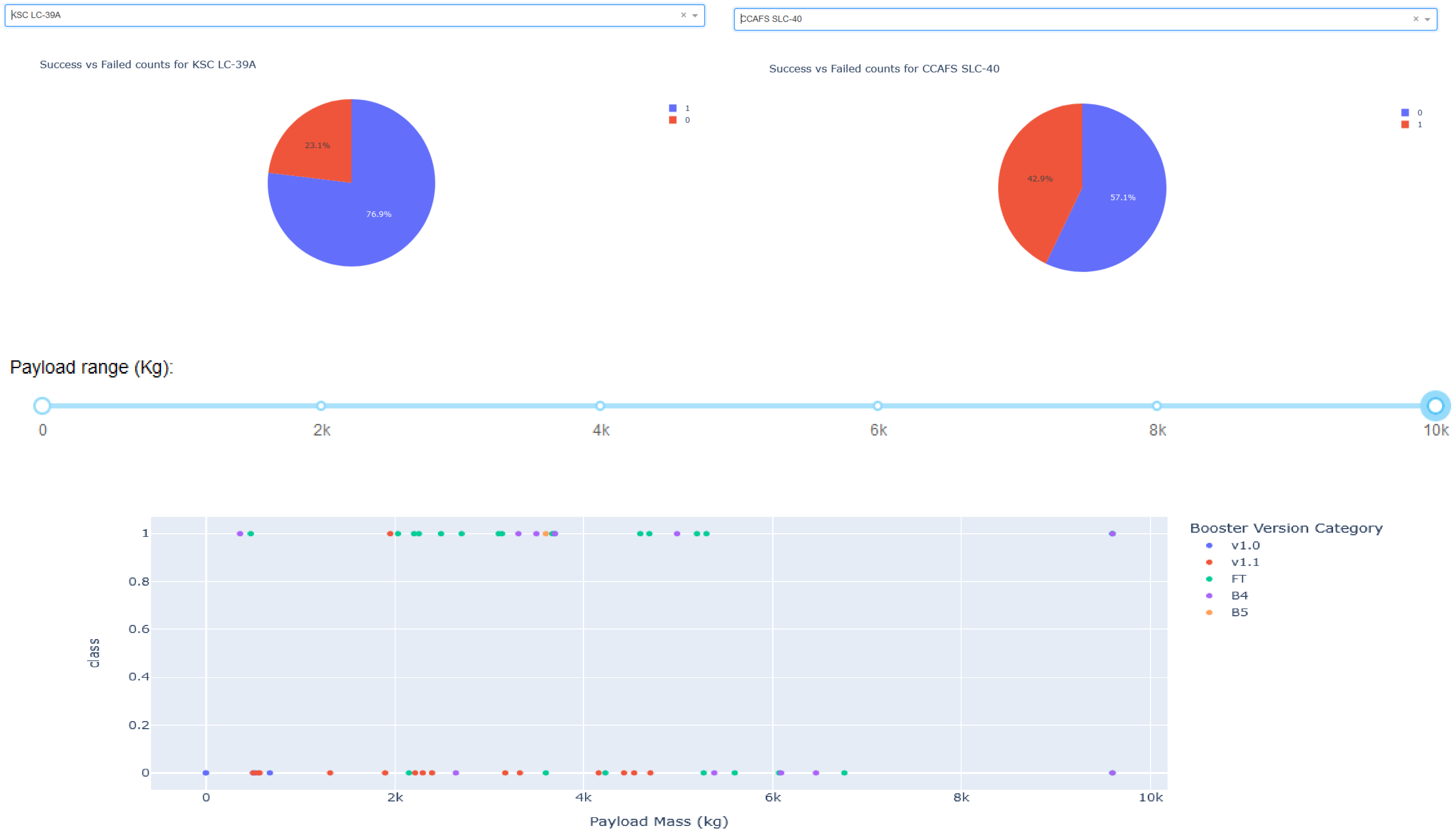
East Coast

Task 7: Dashboard



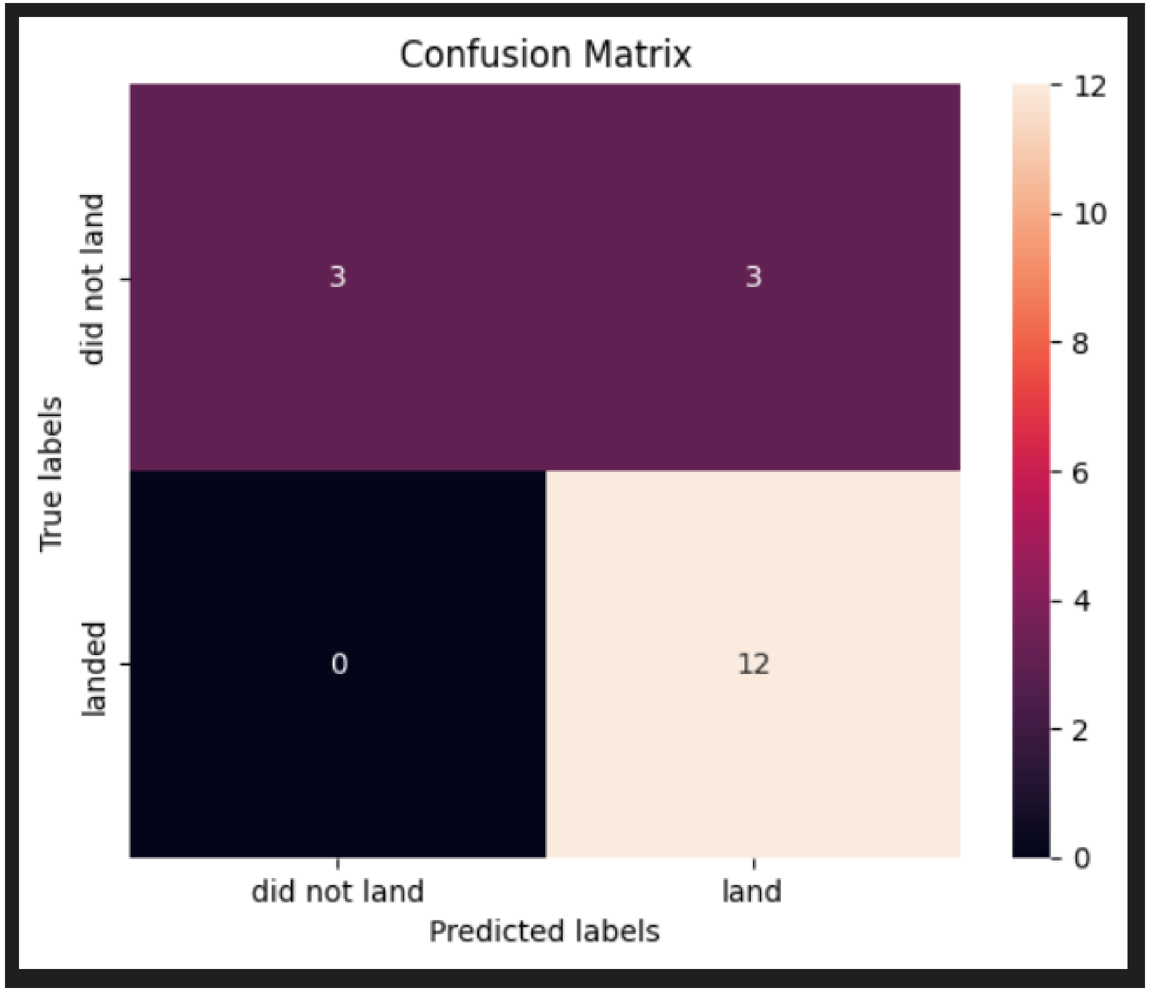





KSC LC-39A
CCAFS SLC-40
Success vs Failed counts for KSC LC-39A
1
0
23.1%
76.9%
Success vs Failed counts for CCAFS SLC-40
0
1
42.9%
57.1%
Payload range (Kg):
0
2k
4k
6k
8k
10k
class
1
0.8
0.6
0.4
0.2
0
Payload Mass (kg)
Booster Version Category
v1.0
v1.1
FT
B4
B5

Task 8: Model training And Evaluation

```
GridSearchCV
GridSearchCV(cv=10, estimator=LogisticRegression(),
             param_grid={'C': [0.01, 0.1, 1], 'penalty': ['l2'],
                         'solver': ['lbfgs']})
estimator: LogisticRegression
LogisticRegression()
LogisticRegression
LogisticRegression()
```



Examining the confusion matrix, we see that logistic regression can distinguish between the different classes. We see that the major problem is false positives.

GridSearchCV

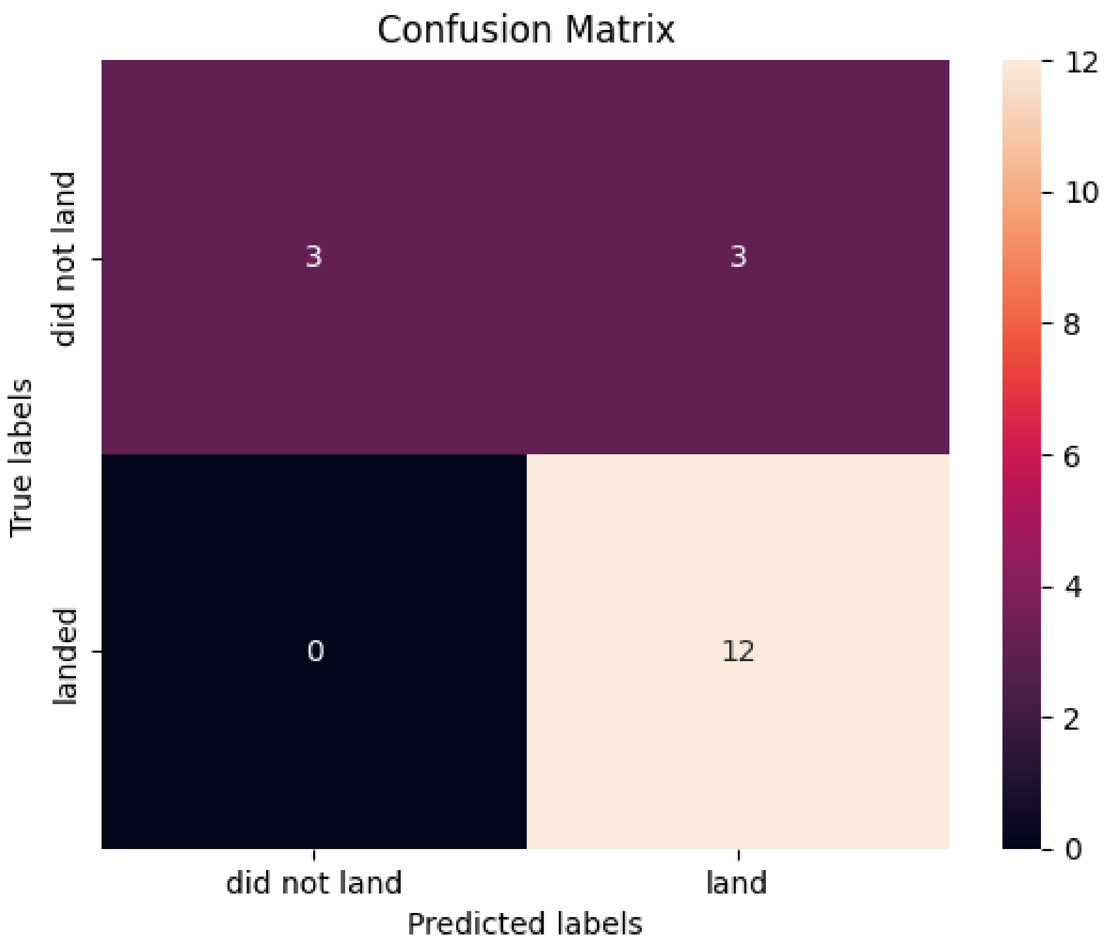
```
GridSearchCV(cv=10, estimator=SVC(),
             param_grid={'C': array([1.00000000e-03, 3.16227766e-02, 1.00000000e+00, 3.16227766e+01,
       1.00000000e+03]),
                         'gamma': array([1.00000000e-03, 3.16227766e-02, 1.00000000e+00, 3.16227766e+01,
       1.00000000e+03]),
                         'kernel': ('linear', 'rbf', 'poly', 'rbf', 'sigmoid')})
```

estimator: SVC

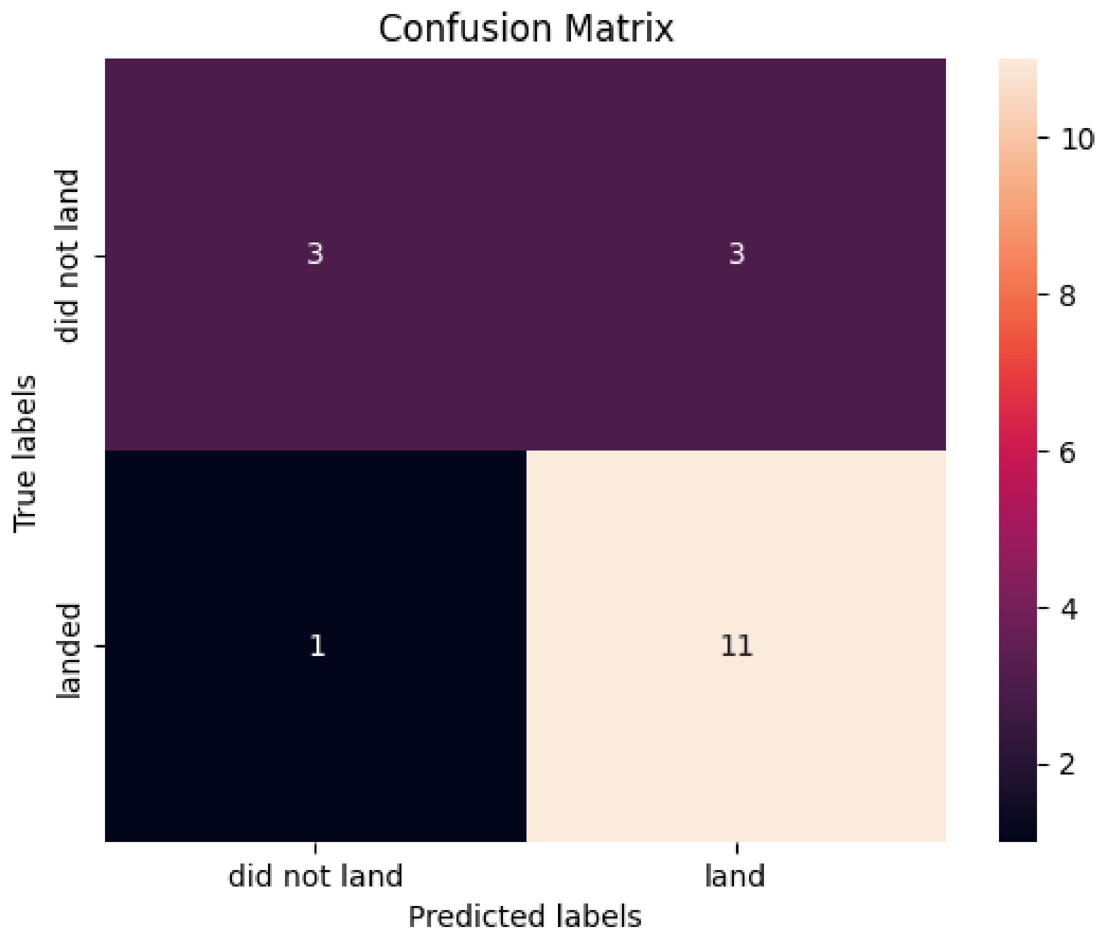
```
SVC()
```

SVC

```
SVC()
```



```
GridSearchCV
GridSearchCV(cv=10, estimator=DecisionTreeClassifier(),
             param_grid={'criterion': ['gini', 'entropy'],
                         'max_depth': [2, 4, 6, 8, 10, 12, 14, 16, 18],
                         'max_features': ['auto', 'sqrt'],
                         'min_samples_leaf': [1, 2, 4],
                         'min_samples_split': [2, 5, 10],
                         'splitter': ['best', 'random']})
estimator: DecisionTreeClassifier
DecisionTreeClassifier()
DecisionTreeClassifier
DecisionTreeClassifier()
```



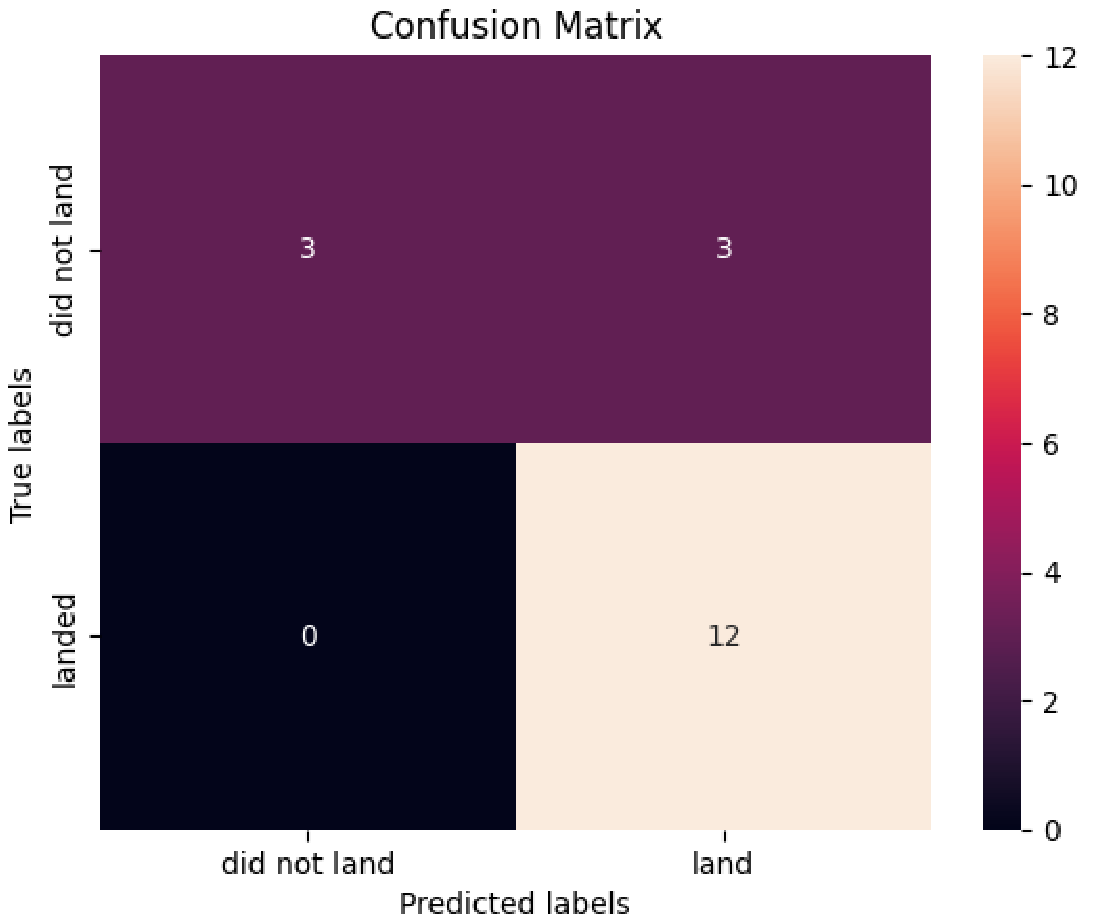
GridSearchCV

```
GridSearchCV(cv=10, estimator=KNeighborsClassifier(),
             param_grid={'algorithm': ['auto', 'ball_tree', 'kd_tree', 'brute'],
                         'n_neighbors': [1, 2, 3, 4, 5, 6, 7, 8, 9, 10],
                         'p': [1, 2]})
```

estimator: KNeighborsClassifier

```
KNeighborsClassifier()
```

▸ KNeighborsClassifier



OVERALL FINDINGS & IMPLICATIONS

Task 1 – Task 4

Findings

- That a single data source may not be sufficient to validate our hypothesis, highlighting the importance of being able to scrape data via APIs and use web scraping tools.
- Data Wrangling is a continuous process that occurs throughout the data collection phase.

Implications

- The need to cross-reference data from multiple sources in various formats implies that parsing skills are crucial for extracting necessary data.
- By answering a variety of questions, I've been able to identify data that could potentially help answer our overall objective.

Task 5 – Task 8

E
D
A

M
L
M

**Machine Learning Models = MLM

Examining the confusion matrix, we see that or Regression, SVM, Tree, and KNN models can distinguish between the different classes. But we see that the major problem is false positives.

DASHBOARD

https://github.com/teslonobo/Notebook.git

DISCUSSION

- If you have any questions feel free to reach out to me on github!

- If there are things I could of done better you are invited to leave constructive criticism.

CONCLUSION

You can spend time playing with the models and come up with different conclusions. You could be overfitting, you could be underfitting, you can be using K-fold instead of how split test train is used. But the Tasks breaks every part down into manageable and digestible parts. Making the project easy to follow along and do. Very fun project!

APPENDIX

- There are a total of 40 points possible for the final assessment, and you will be graded by your peers, who are also completing this assignment.
- The main grading criteria will be:
- Uploaded the URL of your GitHub repository including all the completed notebooks and Python files (1 pt)
- Uploaded your completed presentation in PDF format (1 pt)
- Completed the required Executive Summary slide (1 pt)
- Completed the required Introduction slide (1 pt)
- Completed the required data collection and data wrangling methodology related slides (1 pt)
- Completed the required EDA and interactive visual analytics methodology related slides (3 pts)
- Completed the required predictive analysis methodology related slides (1 pt)
- Completed the required EDA with visualization results slides (6 pts)
- Completed the required EDA with SQL results slides (10 pts)
- Completed the required interactive map with Folium results slides (3 pts)
- Completed the required Plotly Dash dashboard results slides (3 pts)
- Completed the required predictive analysis (classification) results slides (6 pts)
- Completed the required Conclusion slide (1 pts)
- Applied your creativity to improve the presentation beyond the template (1 pts)
- Displayed any innovative insights (1 pts)