Optimizing First-Stage Rocket Recovery: A Feasibility Study

Deborah L. Boxall
November 2024

SPACE Y

Photo credit: Craig Bailey/FLORIDA TODAY

Table of Contents

Executive Summary

Summary of Methodologies

- Data analysis was employed to identify the factors that contribute to a successful landing of reusable first stage boosters. The following methodologies were utilized:
- Collect data using SpaceX API and web scraping techniques
- Wrangle the data in order to correct for missing values and to generate a binary column to track the success (value = 1) or failure (value = 0) of an attempted landing
- Examine how payload, launch site, flight number and yearly trend are correlated to successful landings using data visualization and statistical (SQL) techniques.
- Create an interactive map to visualize geographic trends in landing success data
- Compare the training/test results for logistic regression, support vector machine, decision tree and K-nearest neighbor to identify the best model for predictive analysis of successful landings

Key results

- Launch success has improved over time
- Kennedy Space Center has the highest success rates among landing sites
- Boosters used for Low Earth Orbits have a higher success rate than geosynchronous transfer orbits
- The Decision Tree predictive analysis model highest accuracy score for validation data and test data

Introduction

When it comes to lowering launch costs for commercial applications, SpaceX is a leader in the space industry. One of the key ways that they have brought costs down is by having the first stage booster for the rocket land on either land or ocean drones. The retrieved booster can then be refurbished and relaunched. To date, SpaceX's turnaround time to refurbish a Falcon 9 booster is on the order of 20 days. An additional cost saving strategy that SpaceX employs is to keep all construction, launching and refurbishing in-house. The only component that is not always constructed by SpaceX is the payload, which is furnished by the customer.

Objectives for this study

- Determine the relationship between successful landings and payload mass, launch site, number of flights and orbit
- Quantitatively determine the landing success rate for each of the factors listed above
- Develop a machine learning model to predict the probability of a successful landing from given values for the above factors.

NSF : March 18, 2022

SpaceX achieves a new milestone in reusable boosters with 12th successful launch of the same Falcon 9 booster.
[Alcantarilla, 2022]

TESLARATI : April 29, 2022

SpaceX sets new turnaround record by refurbishing Falcon 9 booster in 9 days; significantly lower than the previous refurbishment time of 21 days.
[Ralph, 2022]

ARK INVEST : September 26, 2022

It is estimated that it costs ~$800/kg to launch a refurbished Falcon 9 to low Earth orbit as compared to ~$2700/kg for a new Falcon 9. *[Korus, 2022]*

Methodology

Steps:

1. Data collection using SpaceX REST API and web scraping techniques
2. Data wrangling to prepare collected data for analysis
3. Exploratory data analysis with SQL and data visualization techniques
4. Data visualization using Folium to prepare an interactive map and Plotly Dash to prepare an interactive dashboard
5. Build and train models to determine best model for predicting landing outcomes

Attribution: gtstudioimagen on Freepik

Data Collection: SpaceX API

Steps:

1. Request data from SpaceX API, decode using .json_normalize, and convert to dataframe
2. Filter dataframe to contain only Falcon 9 boosters
3. Replace missing values in payload mass column with the calculated mean
4. Export data to csv file

Falcon 9 launch

(By Joel Kowsky - https://www.flickr.com)

Data Collection: Webscraping

Steps:

1. Request Falcon 9 launch data from Wikipedia and create Beautiful Soup object using HTML parser
2. Extract column names from HTML header information
3. Create dictionary from parsing the HTML data tables
4. Create dataframe from the dictionary
5. Export data to csv file

Data Wrangling

Steps:

- Determine the number of launches from each site and if had a successful landing of booster
- Determine the different types of orbitals
- Create binary landing outcome column
- Export data to csv file

Landing outcome values: consist of two terms

First term: True/False for landing success

Second term: location of landing

Examples:

- True Ocean = landed successfully in designated ocean site
- False RTLS = did not land on ground pad

Predictive Analytics

Steps:

1. Create Numpy array from Class column
2. Standardize data with StandardScaler, then fit and transform data
3. Split the data into 80% train and 20% test sets
4. Apply a GridSearchCV on algorithims listed below and determine accuracy of each model
   - Logistic Regression
   - Support Vector Machine
   - Decision Tree Classifier
   - K-Nearest Neighbor
5. Assess confusion matrix for each model
6. Identify the best model using the Jaccard Score, F1 Score and Accuracy

Results: Initial EDA using SQL

- The first successful landing of a first stage booster rocket occurred on December 22, 2015
- Three Falcon 9 Full Thrust booster versions were successfully landed (on land or on drone ship)
  - **Full Thrust** boosters (F9 FT) were used for 24 launches from December 2015 to February 2018. If the rocket was intended to deliver the payload to LEO, a successful landing was achieved 38% of the time. If a GTO was attempted, the first stage of the rocket was only able to land successfully 25% of the time. There was also an 8% failure rate for GTO orbits as compared to 0% for LEO, suggesting that the extra power needed to reach the higher GTO may have been compromising the performance of the rocket.
  - **Block 4** boosters (F9 B4) were only used 12 times from August 2017 to June 2018. While they did have a 33% success rate of landing, this was only for LEO orbit. No landing attempt for the other two orbital types, GTO and HEO, were made with these boosters.
  - **Block 5** boosters (F9 B5)were used for over half (55%) of the launches between May 2018 to December 2020, and were used for the greatest variety of orbital types, with LEO having the highest landing success rate of 56%. In addition, 18% of the GTO also were also landed successfully with no failures.

Results: Exploratory Data Analysis

Exploratory Data Analysis

- Launch success has increased over time ***(see Figure 1)***

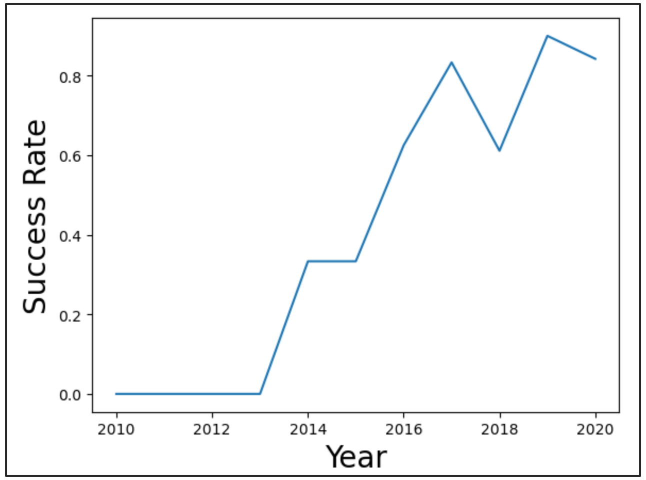
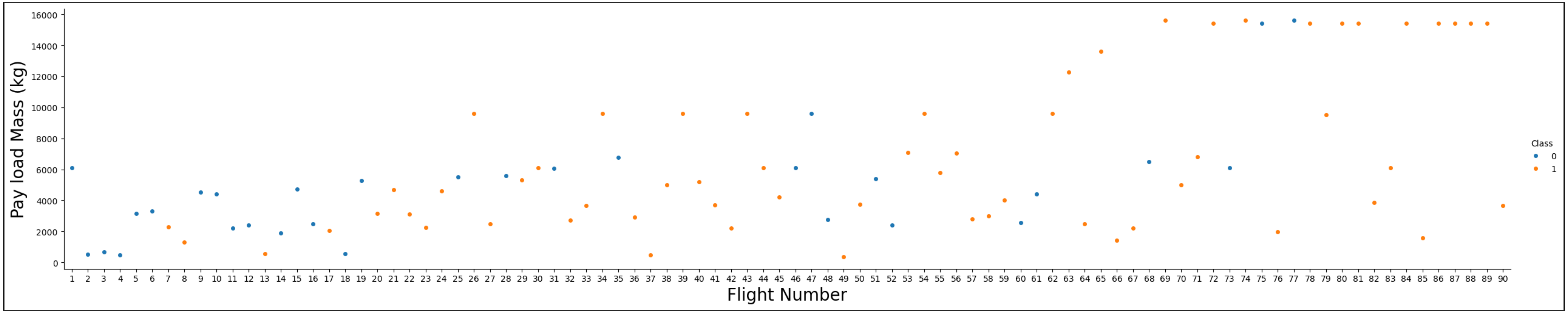


Figure 1: Success rate vs. Launch Date

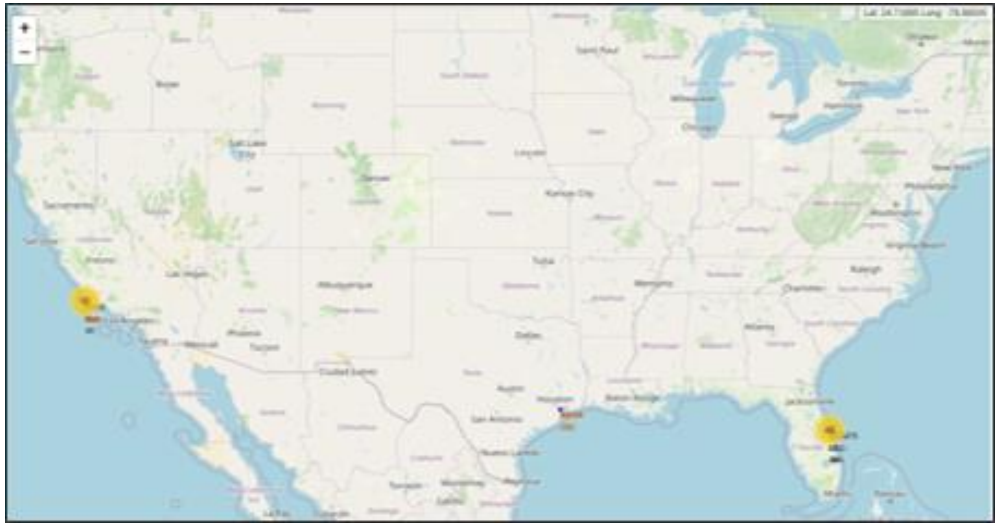
- The number of successful landings with high payloads increased with time (see figure below)



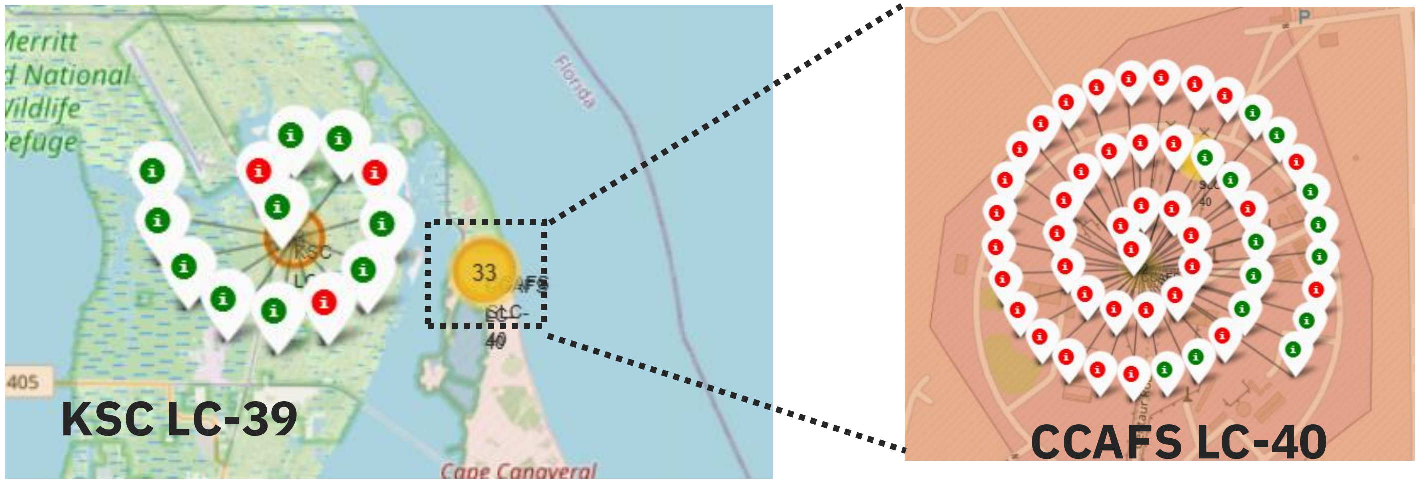
Results: Interactive Map of Launch Sites

Steps:

- Indicated site of NASA Johnson Space Center with **blue** marker, and used **red** markers to indicate SpaceX launch sites

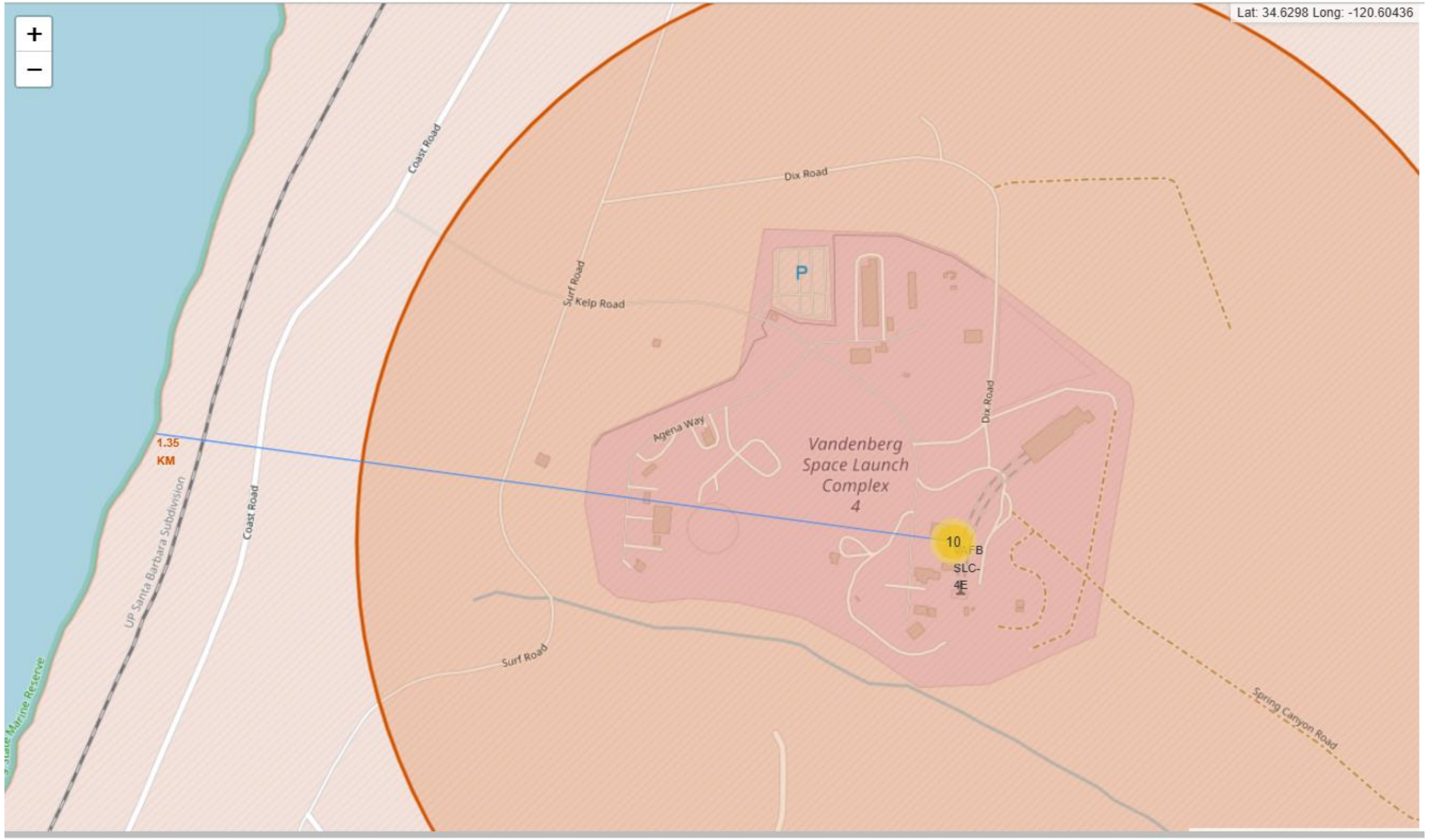


- Each SpaceX launch site was marked with popup **green** markers for successful landings and **red** markers for unsuccessful landings



Interactive Map of Launch Sites *(cont.)*

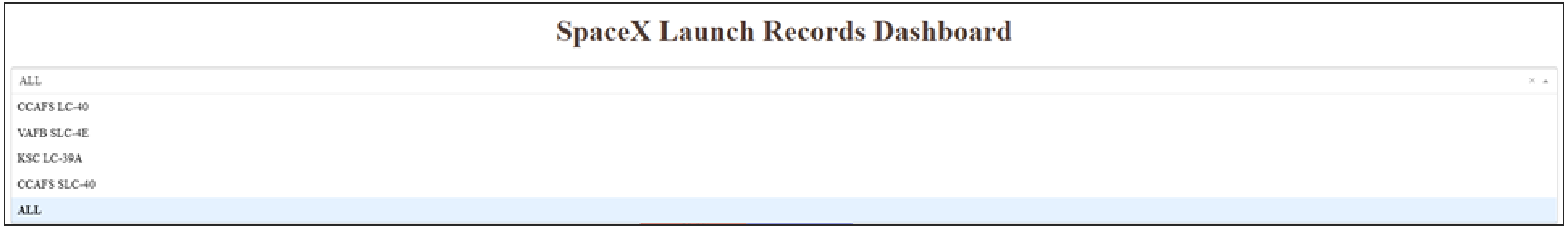
- A function was defined to calculate the distance between Vandenberg Space Force Complex (VAFB) and another location identified by coordinates.
- A second function was defined to draw a line connecting VAFB to the second location.



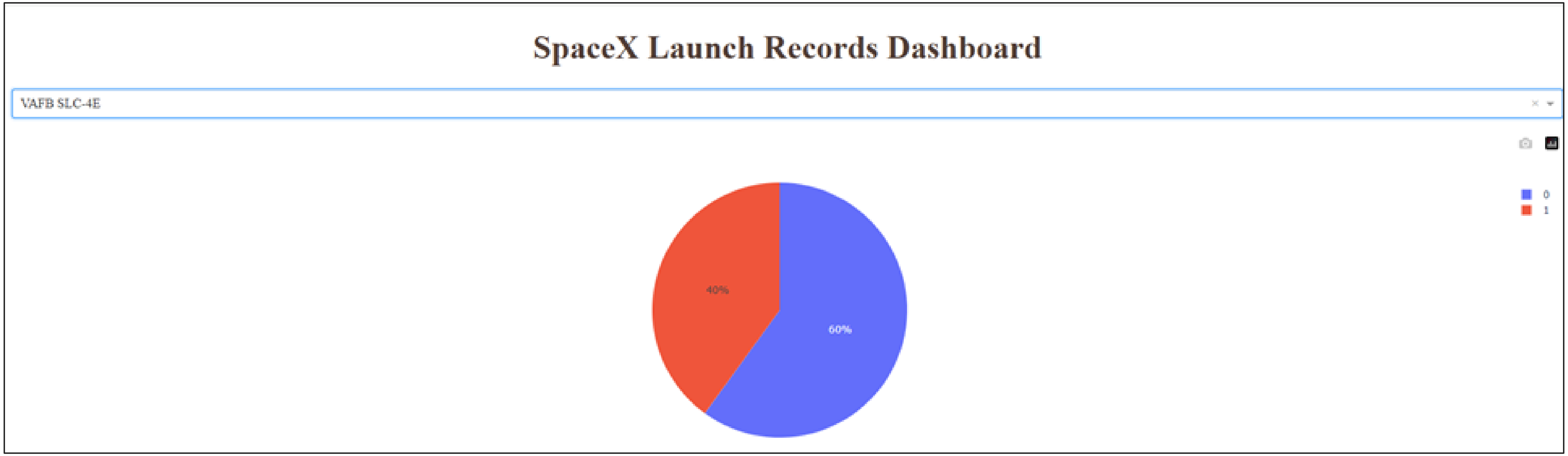
Results: Dashboard with Plotly Dash

Dashboard consists of:

- Dropdown list of launch sites to allow user selection of specific site

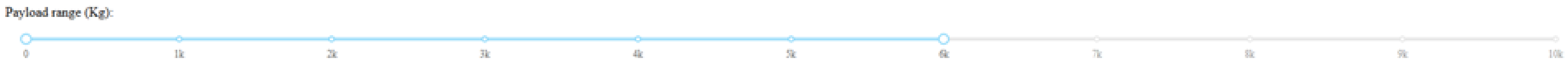


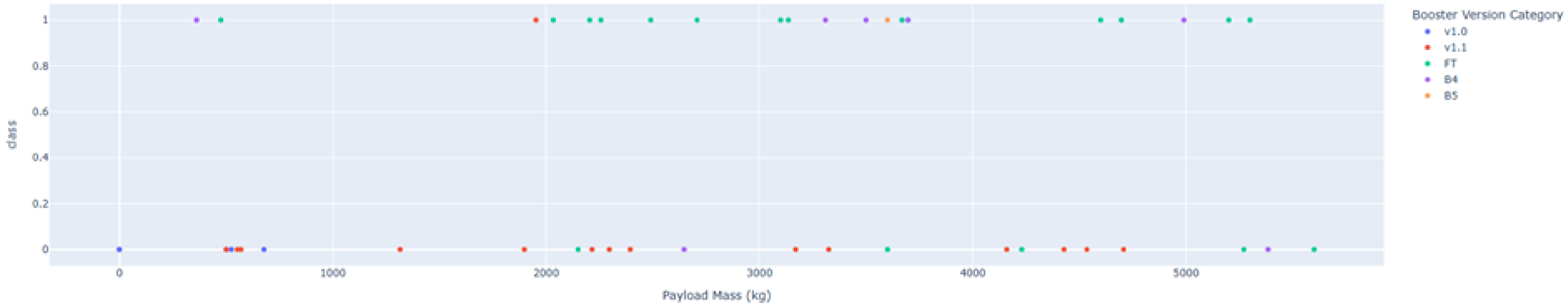
- Pie chart of successful launches that updates depending on launch site selected



Dashboard with Plotly Dash *(cont.)*

- Slider for payload mass range

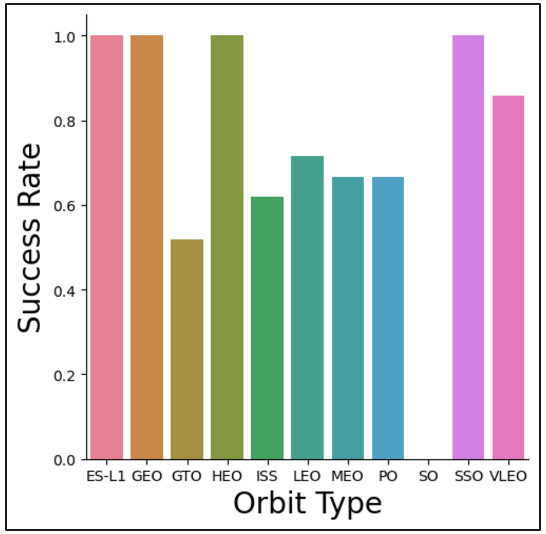




- Scatter chart shows successful landings (y = 1) and unsuccessful landings (y = 0) for each Falcon 9 booster that was used to lift the user selected payload mass.

Results: Orbit Types and Predictive Analytics

- Orbits ES-L1, GEO, HEO and SSO had the highest landing success rate for the booster used
- Most launch sites are close to coast and located in areas with low population density



<u>Predictive Analytics</u>

The Decision Tree model is the best predictive model used as can be seen *(below)* from the accuracy values for both the validation and test data sets

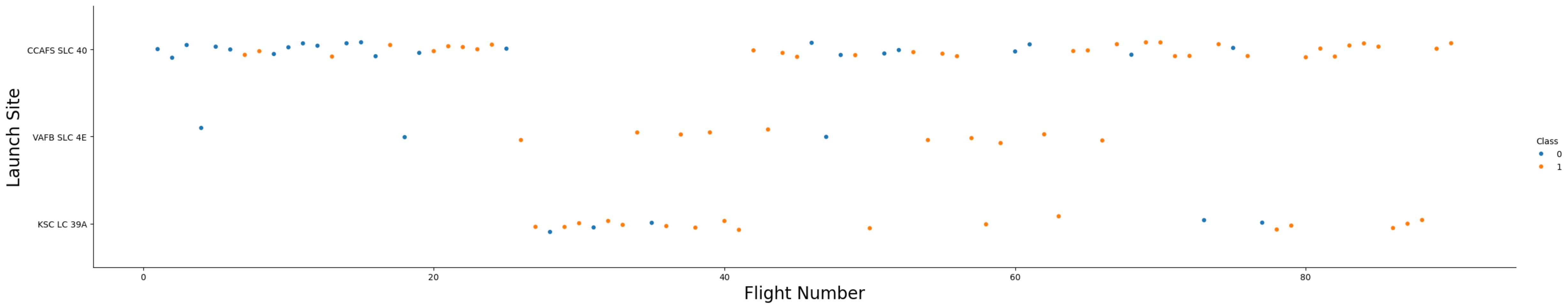
Machine Learning Models

Accuracy of model using:	Logistic Regression (LogReg)	Support Vector Machine (SVM)	Decision Tree (DT)	K-Nearest Neighbor (KNN)
Validation data	0.846	0.848	0.875	0.848
Test data	0.833	0.833	0.944	0.833

Results: Flight Number vs. Launch Site

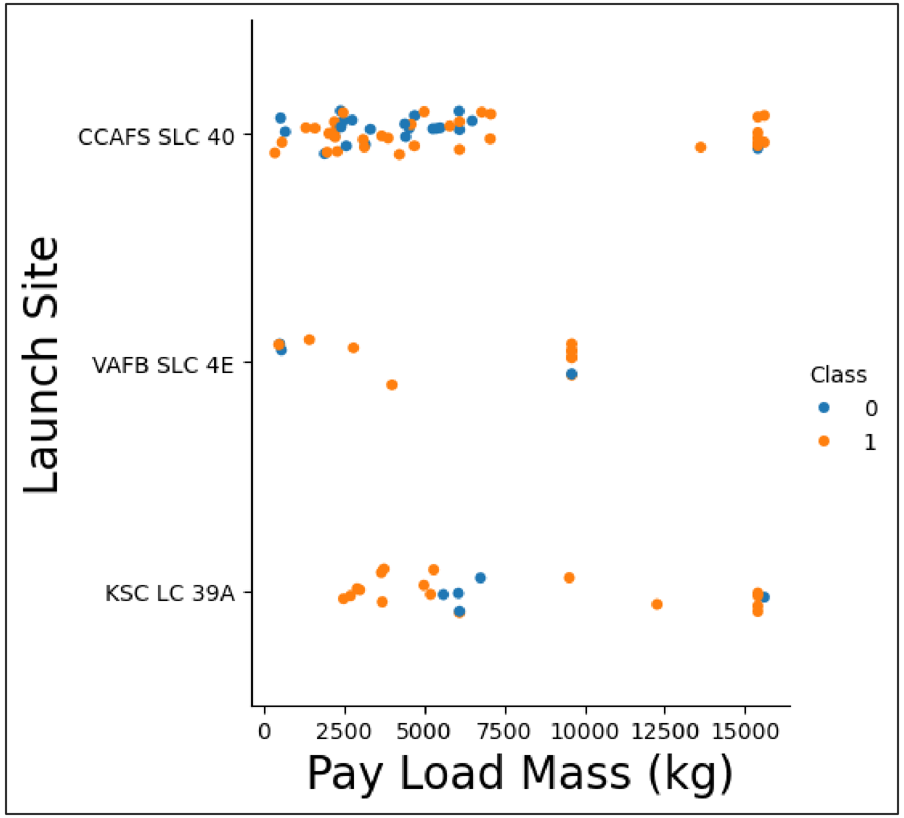
In the figure below, successful landings are denoted with orange markers while unsuccessful landings are denoted with blue markers.

- The probability of a successful landing increased with time (higher flight numbers)
- The majority of the launches occurred at Cape Canaveral LC-40
- However, Vandenberg and Kennedy Space Center had an overall higher success rate.



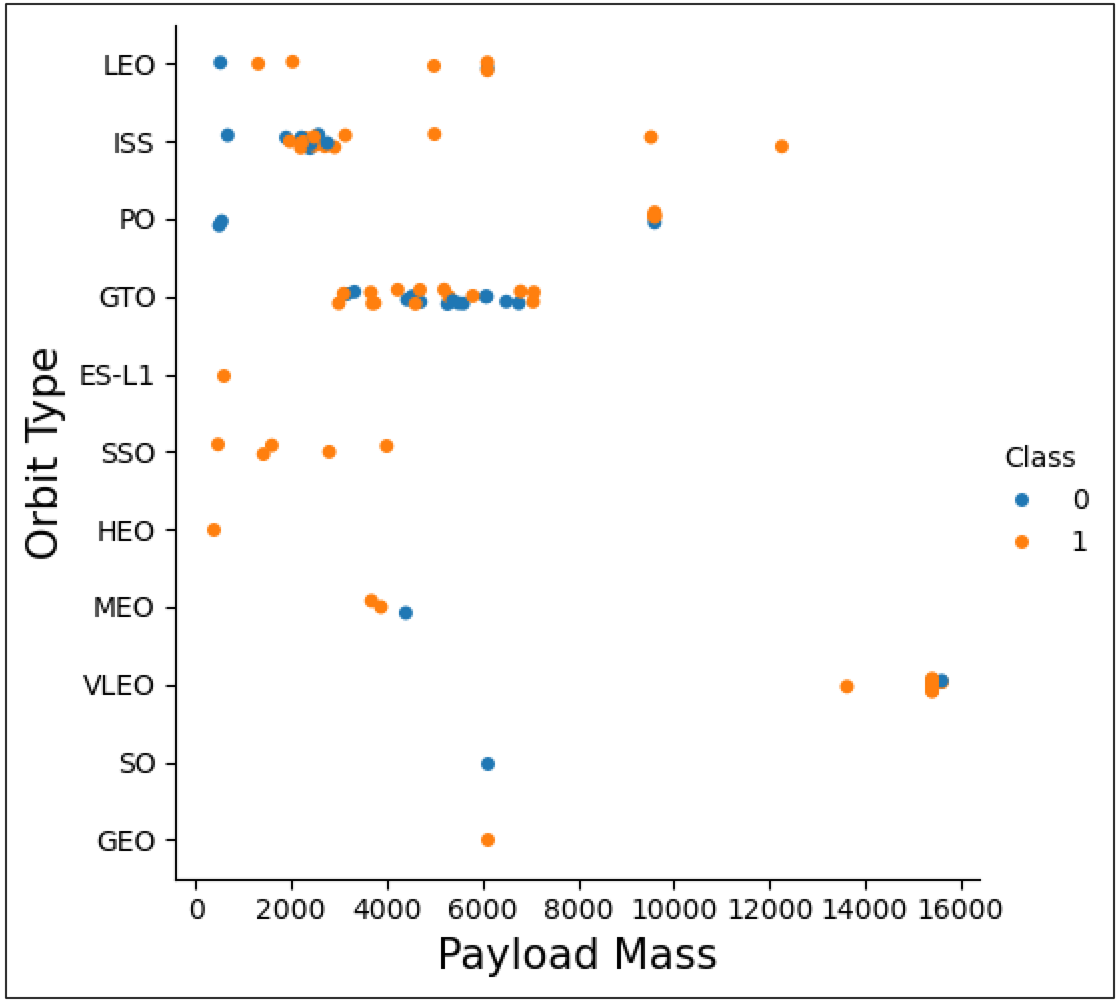
Results: Payload vs. Launch Site

- In general, success rate increases with payload mass.
- Kennedy Space Center has a 100% success rate for launches with payloads below 5500 kg
- Vandenberg Air Force Base is primarily used for low payload mass (<10,000 kg) launches



Results: Payload vs. Orbit

- Heavy payloads (>6000 kg) are more likely to be successful when launched for lower elevation orbitals (LEO, PO, VLEO)
- GTO had mixed results for heavy payloads
- SSO was 100% successful but was only used for light payloads (<6000 kg)



Appendix

Sources:

Alcantarilla, A. *SpaceX Sets New Booster Reuse Record on Starlink Mission*. **Nasa Space Flight** (March 18, 2022) https://www.nasaspaceflight.com/2022/03/spacex-booster-reuse-record-starlink. (accessed 11/19/24)

Korus, S. *The Turnaround Time in Rocket Reuse Suggests that the Cost of Refurbishing the First Stage of the Falcon 9 Has Dropped from Roughly 13 Million t $1 Million in the Last Five Years*, **ARK Invest Newsletter**, (September 26, 2022). https://www.ark-invest.com/newsletters/issue-335 (accessed 11/19/24)

Ralph, E. *SpaceX Smashes Falcon 9 Booster Turnaround Record*. **Teslarati**. (April 29, 2022) https://www.teslarati.com/spacex-falcon-9-new-booster-turnaround-record-21-days. (accessed 11/19/24)

Extras: SQL EDA

Landing Outcome from 2010-06-04 to 2020-12-06 (SQL EDA)

Landing_Outcome	COUNT_LAUNCHES
Success	38
No attempt	21
Success (drone ship)	14
Success (ground pad)	9
Failure (drone ship)	5
Controlled (ocean)	5
Failure	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1
No attempt	1

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

[66]:
```
%sql SELECT DATE, BOOSTER_VERSION, PAYLOAD_MASS__KG_ FROM SPACEXTBL WHERE LANDING_OUTCOME = 'Success (drone
```

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

[66]:

Date	Booster_Version	PAYLOAD_MASS__KG_
2016-05-06	F9 FT B1022	4696
2016-08-14	F9 FT B1026	4600
2017-03-30	F9 FT B1021.2	5300
2017-10-11	F9 FT B1031.2	5200

EDA SQL: Booster Version, Payload Mass, Orbit and Landing Outcome

Date	Launch_Site	Booster_Version	PAYLOAD_MASS__KG_	Orbit	Landing_Outcome
2015-12-22	CCAFS LC-40	F9 FT B1019	2034	LEO	Success (ground pad)
2016-03-04	CCAFS LC-40	F9 FT B1020	5271	GTO	Failure (drone ship)
2016-04-08	CCAFS LC-40	F9 FT B1021.1	3136	LEO (ISS)	Success (drone ship)
2016-05-06	CCAFS LC-40	F9 FT B1022	4696	GTO	Success (drone ship)
2016-05-27	CCAFS LC-40	F9 FT B1023.1	3100	GTO	Success (drone ship)
2016-06-15	CCAFS LC-40	F9 FT B1024	3600	GTO	Failure (drone ship)
2016-07-18	CCAFS LC-40	F9 FT B1025.1	2257	LEO (ISS)	Success (ground pad)
2016-08-14	CCAFS LC-40	F9 FT B1026	4600	GTO	Success (drone ship)
2017-01-14	VAFB SLC-4E	F9 FT B1029.1	9600	Polar LEO	Success (drone ship)
2017-02-19	KSC LC-39A	F9 FT B1031.1	2490	LEO (ISS)	Success (ground pad)
2017-03-16	KSC LC-39A	F9 FT B1030	5600	GTO	No attempt
2017-03-30	KSC LC-39A	F9 FT B1021.2	5300	GTO	Success (drone ship)
2017-05-01	KSC LC-39A	F9 FT B1032.1	5300	LEO	Success (ground pad)
2017-05-15	KSC LC-39A	F9 FT B1034	6070	GTO	No attempt
2017-06-03	KSC LC-39A	F9 FT B1035.1	2708	LEO (ISS)	Success (ground pad)
2017-06-23	KSC LC-39A	F9 FT B1029.2	3669	GTO	Success (drone ship)
2017-06-25	VAFB SLC-4E	F9 FT B1036.1	9600	LEO	Success (drone ship)
2017-07-05	KSC LC-39A	F9 FT B1037	6761	GTO	No attempt
2017-08-24	VAFB SLC-4E	F9 FT B1038.1	475	SSO	Success (drone ship)
2017-10-11	KSC LC-39A	F9 FT B1031.2	5200	GTO	Success (drone ship)
2017-12-15	CCAFS SLC-40	F9 FT B1035.2	2205	LEO (ISS)	Success (ground pad)
2017-12-23	VAFB SLC-4E	F9 FT B1036.2	9600	Polar LEO	Controlled (ocean)
2018-01-31	CCAFS SLC-40	F9 FT B1032.2	4230	GTO	Controlled (ocean)
2018-02-22	VAFB SLC-4E	F9 FT B1038.2	2150	SSO	No attempt

Date	Launch_Site	Booster_Version	PAYLOAD_MASS__KG_	Orbit	Landing_Outcome
2017-08-14	KSC LC-39A	F9 B4 B1039.1	3310	LEO (ISS)	Success (ground pad)
2017-09-07	KSC LC-39A	F9 B4 B1040.1	4990	LEO	Success (ground pad)
2017-10-09	VAFB SLC-4E	F9 B4 B1041.1	9600	Polar LEO	Success (drone ship)
2017-10-30	KSC LC-39A	F9 B4 B1042.1	3500	GTO	Success (drone ship)
2018-01-08	CCAFS SLC-40	F9 B4 B1043.1	5000	LEO	Success (ground pad)
2018-03-06	CCAFS SLC-40	F9 B4 B1044	6092	GTO	No attempt
2018-03-30	VAFB SLC-4E	F9 B4 B1041.2	9600	Polar LEO	No attempt
2018-04-02	CCAFS SLC-40	F9 B4 B1039.2	2647	LEO (ISS)	No attempt
2018-04-18	CCAFS SLC-40	F9 B4 B1045.1	362	HEO	Success (drone ship)
2018-05-22	VAFB SLC-4E	F9 B4 B1043.2	6460	Polar LEO	No attempt
2018-06-04	CCAFS SLC-40	F9 B4 B1040.2	5384	GTO	No attempt
2018-06-29	CCAFS SLC-40	F9 B4 B1045.2	2697	LEO (ISS)	No attempt

Date	Launch_Site	Booster_Version	PAYLOAD_MASS__KG_	Orbit	Landing_Outcome
2018-05-11	KSC LC-39A	F9 B5 B1046.1	3600	GTO	Success (drone ship)
2018-07-22	CCAFS SLC-40	F9 B5B1047.1	7075	GTO	Success
2018-07-25	VAFB SLC-4E	F9 B5B1048.1	9600	Polar LEO	Success
2018-08-07	CCAFS SLC-40	F9 B5 B1046.2	5800	GTO	Success
2018-09-10	CCAFS SLC-40	F9 B5B1049.1	7060	GTO	Success
2018-10-08	VAFB SLC-4E	F9 B5 B1048.2	3000	SSO	Success
2018-11-15	KSC LC-39A	F9 B5 B1047.2	5300	GTO	Success
2018-12-03	VAFB SLC-4E	F9 B5 B1046.3	4000	SSO	Success
2018-12-05	CCAFS SLC-40	F9 B5B1050	2500	LEO (ISS)	Failure
2018-12-23	CCAFS SLC-40	F9 B5B1054	4400	MEO	No attempt
2019-01-11	VAFB SLC-4E	F9 B5 B1049.2	9600	Polar LEO	Success
2019-02-22	CCAFS SLC-40	F9 B5 B1048.3	4850	GTO	Success
2019-03-02	KSC LC-39A	F9 B5B1051.1	12055	LEO (ISS)	Success
2019-05-04	CCAFS SLC-40	F9 B5B1056.1	2495	LEO (ISS)	Success
2019-05-24	CCAFS SLC-40	F9 B5 B1049.3	13620	LEO	Success
2019-06-12	VAFB SLC-4E	F9 B5 B1051.2	4200	SSO	Success

Date	Launch_Site	Booster_Version	PAYLOAD_MASS__KG_	Orbit	Landing_Outcome
2019-07-25	CCAFS SLC-40	F9 B5 B1056.2	2268	LEO (ISS)	Success
2019-08-06	CCAFS SLC-40	F9 B5 B1047.3	6500	GTO	No attempt
2019-11-11	CCAFS SLC-40	F9 B5 B1048.4	15600	LEO	Success
2019-12-05	CCAFS SLC-40	F9 B5B1059.1	2617	LEO (ISS)	Success
2019-12-17	CCAFS SLC-40	F9 B5 B1056.3	6956	GTO	Success
2020-01-07	CCAFS SLC-40	F9 B5 B1049.4	15600	LEO	Success
2020-01-19	KSC LC-39A	F9 B5 B1046.4	12050	Sub-orbital	No attempt
2020-01-29	CCAFS SLC-40	F9 B5 B1051.3	15600	LEO	Success
2020-02-17	CCAFS SLC-40	F9 B5 B1056.4	15600	LEO	Failure
2020-03-07	CCAFS SLC-40	F9 B5 B1059.2	1977	LEO (ISS)	Success
2020-03-18	KSC LC-39A	F9 B5 B1048.5	15600	LEO	Failure
2020-04-22	KSC LC-39A	F9 B5 B1051.4	15600	LEO	Success
2020-05-30	KSC LC-39A	F9 B5B1058.1	12530	LEO (ISS)	Success
2020-06-04	CCAFS SLC-40	F9 B5 B1049.5	15600	LEO	Success
2020-06-13	CCAFS SLC-40	F9 B5 B1059.3	15410	LEO	Success
2020-06-30	CCAFS SLC-40	F9 B5B1060.1	4311	MEO	Success
2020-07-20	CCAFS SLC-40	F9 B5 B1058.2	5500	GTO	Success
2020-08-07	KSC LC-39A	F9 B5 B1051.5	14932	LEO	Success
2020-08-18	CCAFS SLC-40	F9 B5 B1049.6	15440	LEO	Success
2020-08-30	CCAFS SLC-40	F9 B5 B1059.4	3130	SSO	Success
2020-09-03	KSC LC-39A	F9 B5 B1060.2	15600	LEO	Success
2020-10-06	KSC LC-39A	F9 B5 B1058.3	15600	LEO	Success
2020-10-18	KSC LC-39A	F9 B5 B1051.6	15600	LEO	Success
2020-10-24	CCAFS SLC-40	F9 B5 B1060.3	15600	LEO	Success
2020-11-05	CCAFS SLC-40	F9 B5B1062.1	4311	MEO	Success
2020-11-16	KSC LC-39A	F9 B5B1061.1	12500	LEO (ISS)	Success
2020-11-21	VAFB SLC-4E	F9 B5B1063.1	1192	LEO	Success
2020-11-25	CCAFS SLC-40	F9 B5 B1049.7	15600	LEO	Success
2020-12-06	KSC LC-39A	F9 B5 B1058.4	2972	LEO (ISS)	Success