Predicting SpaceX Launch Successes

—

Mira Callahan

Introduction

SpaceX, the leading space travel company today, advertises their rocket trips as affordable. Each launch costs 62 million USD compared to 165 million for their competitors - largely due to their dual launch system where the first part of the launch can be reused. But how cost effective is this really? The following study aims to determine how often the first stage lands, and can then be reused, as a means of determining how affordable SpaceX launches are.

Data Collection

Data was collected by web scraping Wikipedia for information on the SpaceX launches and from IBM's SpaceX REST API. It was then cleaned by:

- Transforming JSON into the proper format
- Adding missing values
- Filtering data to only include Falcon 9 launches
- Normalizing the data

print(response.content)

b'[{"fairings":{"reused":false,"recovery_attempt":false,"recovered":false,
e":"https://images2.imgbox.com/5b/02/QcxHUb5V_o.png"},"reddit":{"campaign"
t":null,"webcast":"https://www.youtube.com/watch?v=0a_00nJ_Y88","youtube_i
unch.html","wikipedia":"https://en.wikipedia.org/wiki/DemoSat"},"static_fi
w":0,"rocket":"5e9d0d95eda69955f709d1eb","success":false,"failures":[{"tim
s and loss of vehicle","crew":[],"ships":[],"capsules":[],"payloads":["5eb
Sat","date_utc":"2006-03-24T22:30:00.000Z","date_unix":1143239400,"date_lo
e289df35918033d3b2623","flight":1,"gridfins":false,"legs":false,"reused":f
update":true,"tbd":false,"launch_library_id":null,"id":"5eb87cd9ffd86e0006
nks":{"patch":{"small":"https://images2.imgbox.com/f9/4a/ZboXReNb_o.png","
ll,"media":null,"recovery":null},"flickr":{"small":[],"original":[]},"pres
c","article":"https://www.space.com/3590-spacex-falcon-1-rocket-fails-reac
l","static_fire_date_unix":null,"net":false,"window":0,"rocket":"5e9d0d95ed
lation leading to premature engine shutdown"}],"details":"Successful first
at T+7 min 30 s, Failed to reach orbit, Failed to recover first stage","cr
090995de566f86","flight_number":2,"name":"DemoSat","date_utc":"2007-03-21T
n":"hour","upcoming":false,"cores":[{"core":"5e9e289ef35918416a3b2624","fl
null,"landing_type":null,"landpad":null}],"auto_update":true,"tbd":false,"
attempt":false,"recovered":false,"ships":[]},"links":{"patch":{"small":"ht
0k_o.png"},"reddit":{"campaign":null,"launch":null,"media":null,"recovery"
atch?v=v0w9p3U8860","youtube_id":"v0w9p3U8860","article":"http://www.space
wiki/Trailblazer_(satellite)"},"static_fire_date_utc":null,"static_fire_da
res":[{"time":140,"altitude":35,"reason":"residual stage-1 thrust led to c
ween stage 1 and stage 2","crew":[],"ships":[],"capsules":[],"payloads":["
light_number":3,"name":"Trailblazer","date_utc":"2008-08-03T03:34:00.000Z"
ing":false,"cores":[{"core":"5e9e289ef3591814873b2625","flight":1,"gridfin
pe":null,"landpad":null}],"auto_update":true,"tbd":false,"launch_library_i
e,"recovered":false,"ships":[]},"links":{"patch":{"small":"https://images2
eddit":{"campaign":null,"launch":null,"media":null,"recovery":null},"flick
EH6G0","youtube_id":"dLQ2tZEH6G0","article":"https://en.wikipedia.org/wiki
T00:00:00.000Z","static_fire_date_unix":1221868800,"net":false,"window":0,
to orbit on the first successful orbital launch of any privately funded an
[],"capsules":[],"payloads":["5eb0e4b7b6c3bb0006eeb1e5"],"launchpad":"5e9e
Z","date_unix":1222643700,"date_local":"2008-09-28T11:15:00+12:00","date_p

Sample of the information retrieved from Wikipedia

Data Collection - Launch Information Sample

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude
0	1	2006-03-24	Falcon 1	20.0	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	Merlin1A	167.743129	9.047721
1	2	2007-03-21	Falcon 1	NaN	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	Merlin2A	167.743129	9.047721
2	4	2008-09-28	Falcon 1	165.0	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	Merlin2C	167.743129	9.047721
3	5	2009-07-13	Falcon 1	200.0	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	Merlin3C	167.743129	9.047721
4	6	2010-06-04	Falcon 9	NaN	LEO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0003	-80.577366	28.561857

Data Collection - Falcon 9 Data Sample
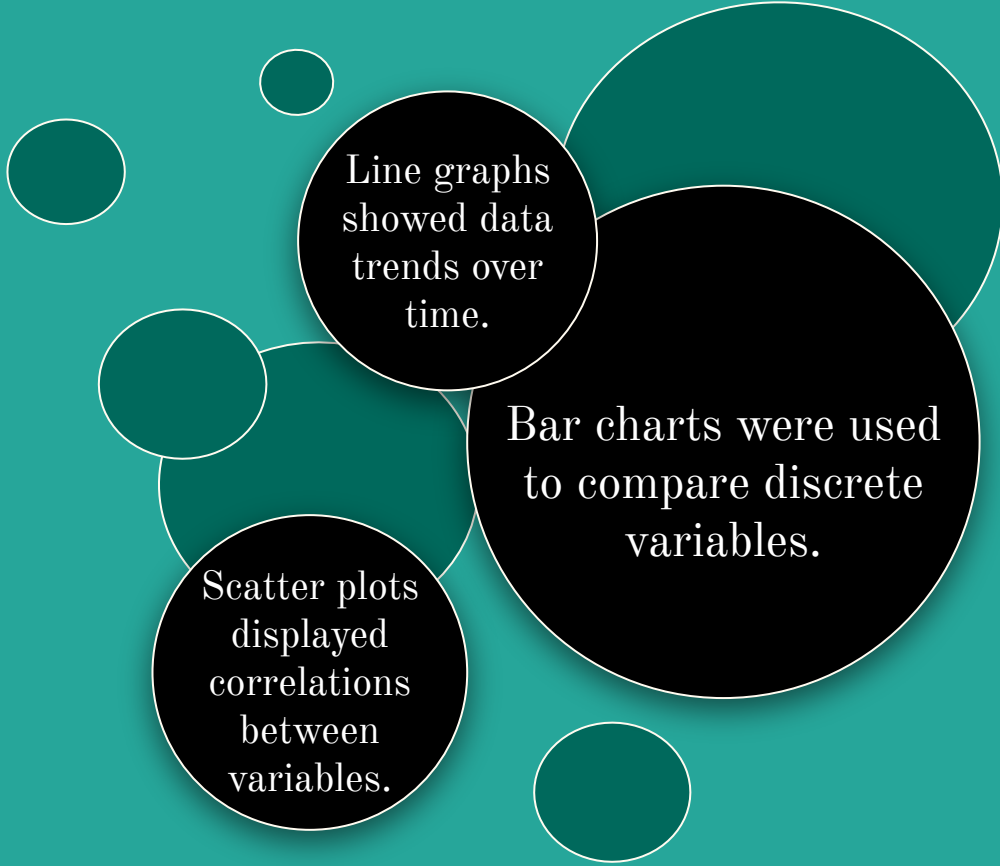
	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude
4	1	2010-06-04	Falcon 9	NaN	LEO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0003	-80.577366	28.561857
5	2	2012-05-22	Falcon 9	525.0	LEO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0005	-80.577366	28.561857
6	3	2013-03-01	Falcon 9	677.0	ISS	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0007	-80.577366	28.561857
7	4	2013-09-29	Falcon 9	500.0	PO	VAFB SLC 4E	False Ocean	1	False	False	False	None	1.0	0	B1003	-120.610829	34.632093
8	5	2013-12-03	Falcon 9	3170.0	GTO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B1004	-80.577366	28.561857
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
89	86	2020-09-03	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	2	True	True	True	5e9e3032383ecb6bb234e7ca	5.0	12	B1060	-80.603956	28.608058
90	87	2020-10-06	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	3	True	True	True	5e9e3032383ecb6bb234e7ca	5.0	13	B1058	-80.603956	28.608058
91	88	2020-10-18	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	6	True	True	True	5e9e3032383ecb6bb234e7ca	5.0	12	B1051	-80.603956	28.608058
92	89	2020-10-24	Falcon 9	15600.0	VLEO	CCSFS SLC 40	True ASDS	3	True	True	True	5e9e3033383ecbb9e534e7cc	5.0	12	B1060	-80.577366	28.561857
93	90	2020-11-05	Falcon 9	3681.0	MEO	CCSFS SLC 40	True ASDS	1	True	False	True	5e9e3032383ecb6bb234e7ca	5.0	8	B1062	-80.577366	28.561857

Data Wrangling - Feature Engineering a Target Variable

Within the SpaceX dataset, launch outcomes are not classified exclusively as a success or failure. Rather, each booster landing was assigned to a landing environment (drone ship (ASDS), ocean, and ground pad (RTLS)) and assessed as to whether it successfully landed there. Since this project is solely interested in whether the landing was successful, a categorical variable was derived from these outcomes to assess this.

Original Value	New Landing Outcome
True ASDS	Success (1)
False ASDS	Failure (0)
None ASDS	Failure (0)
None None	Failure (0)
True Ocean	Success (1)
False Ocean	Failure (0)
True RTLS	Success (1)
False RTLS	Failure (1)

Line graphs showed data trends over time.

Bar charts were used to compare discrete variables.

Scatter plots displayed correlations between variables.

Exploratory Data Analysis with Visualization

EDA with Visualization - Key Insights

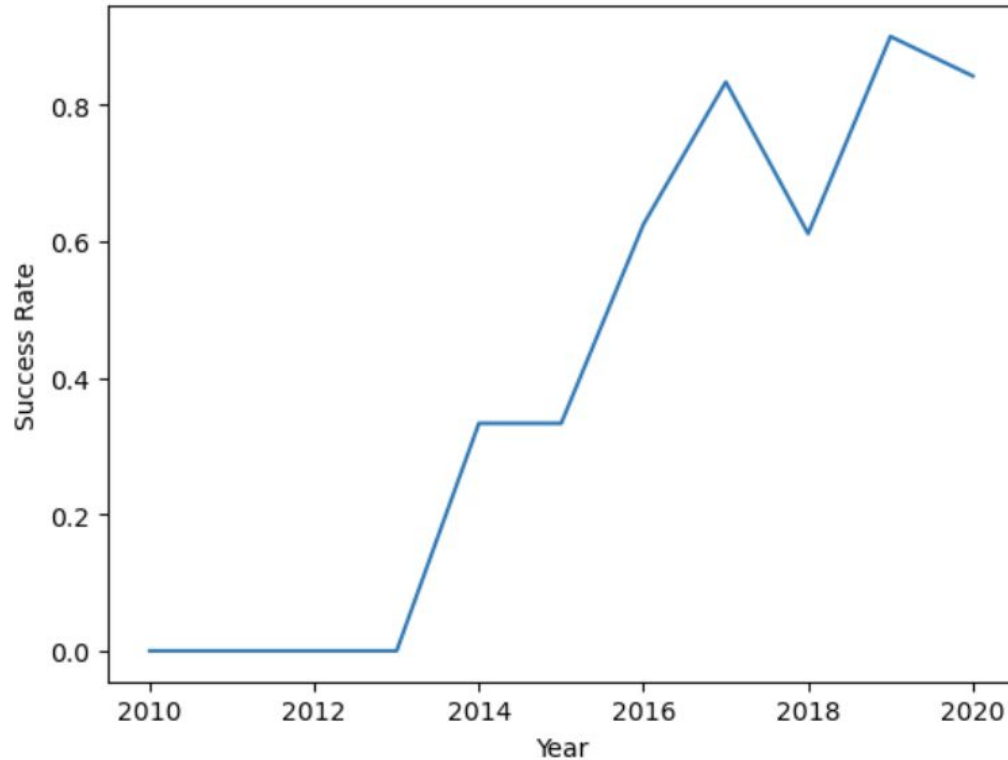Launches with over 7000 kg of payload mass tend to be the most successful.
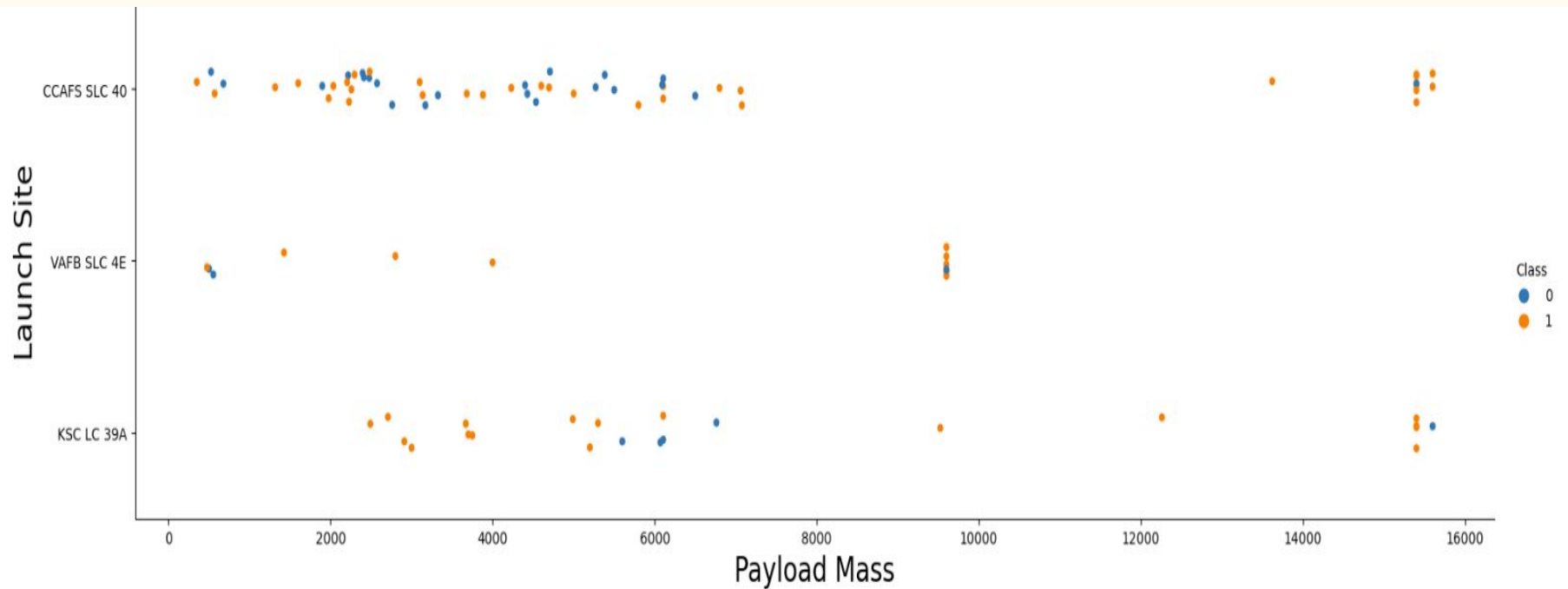
EDA with Visualization - Key Insights

The booster landing for launches to the ES-L1, GEO, HEO and SSO orbits tended to be the most successful, while the first launch components to SO never landed successfully.
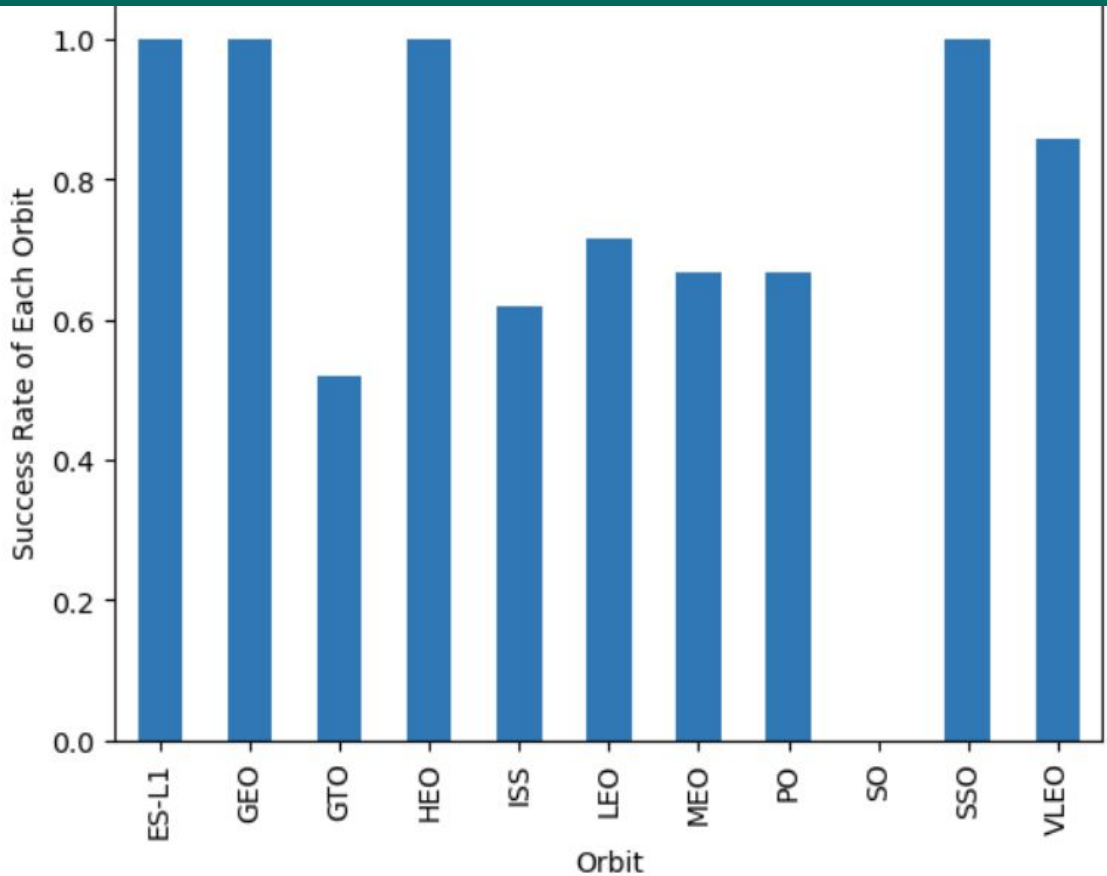
EDA with Visualization - Key Insights

Since 2013, the number of successful SpaceX launches has steadily increased overall - barring a major dip from 2017 - 2018.

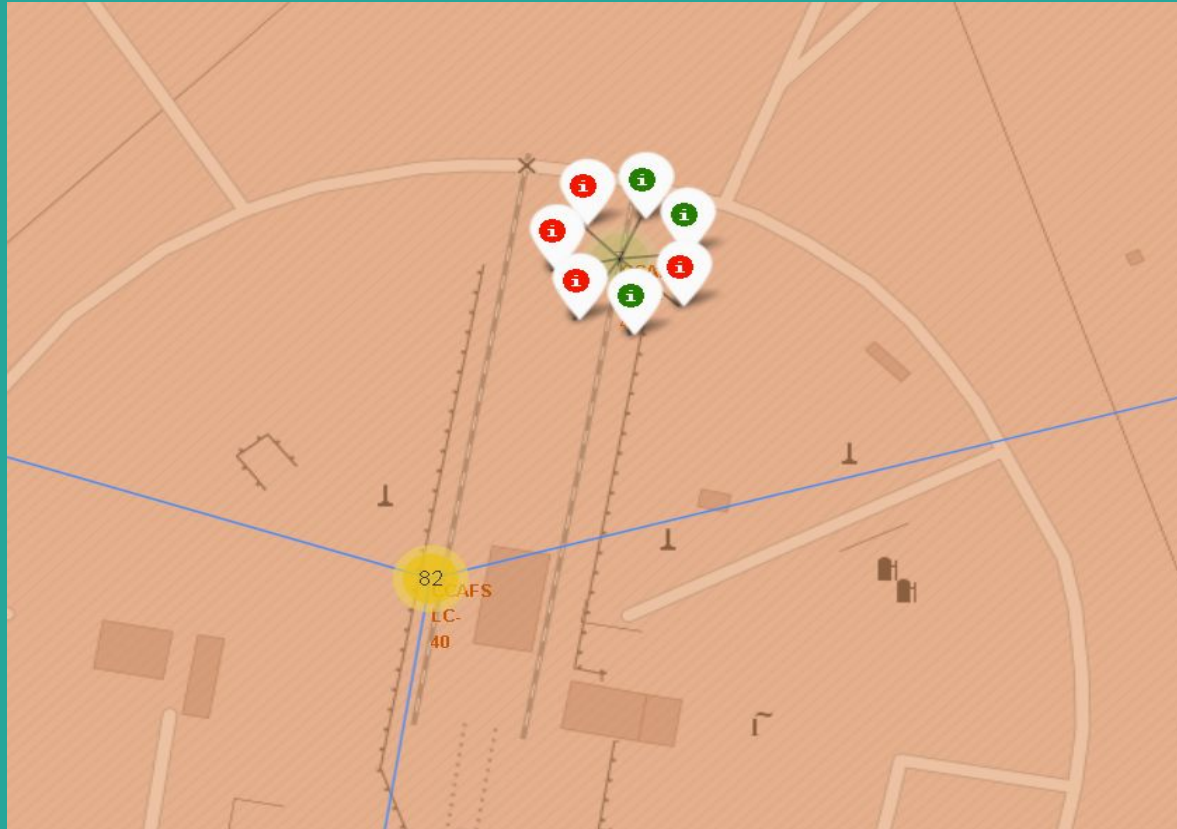# Exploratory Data Analysis with SQL

SQL queries were used to examine specific portions of the dataset and gather further insights. For example, the query and outcome to the right rank the outcome counts for launch failures/successes between 6/4/2010 and 3/20/2017.

```
%sql SELECT LANDING_OUTCOME, COUNT(LANDING_OUTCOME) FROM SPACEX
WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20' \
GROUP BY  LANDING_OUTCOME ORDER BY COUNT(LANDING_OUTCOME) DESC;
```

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

The red and green markers display the successful and failed launch sites and the blue lines demonstrate the distance of the SpaceX center from nearby oceans and cities.

Data Visualization with Folium

To get a better idea of the environment surrounding the launch sites, Folium was used to create an interactive map displaying the SpaceX launches and their distances from nearby oceans, cities, etc.

Data Visualization with Plotly

To gain further insights through data visualization, Plotly was used create scatter plots displaying the correlation between payload and launch site success and pie charts with the overall first stage success across launch sites. As seen below, KSC LC-39A is the most successful launch site.
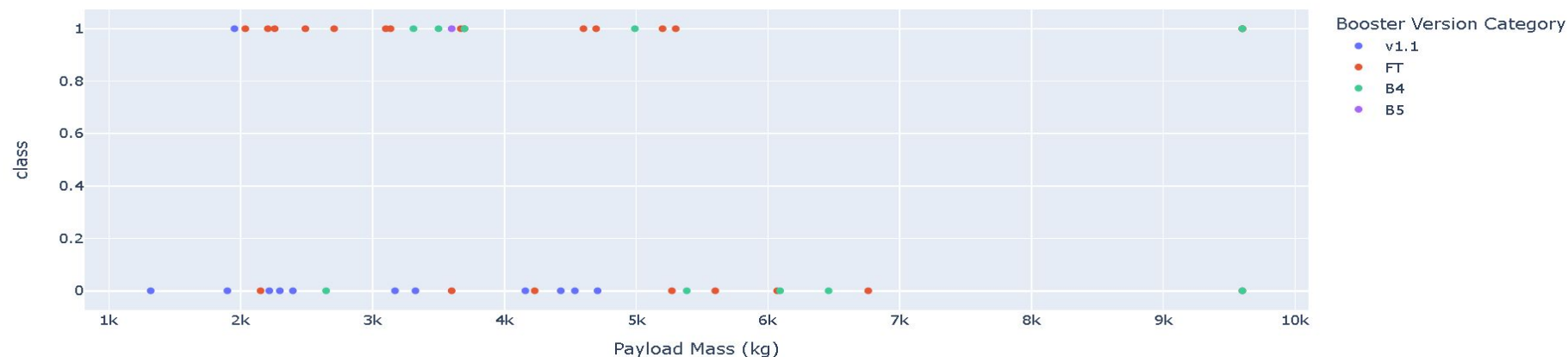

Success Count for All Launch Sites

Data Visualization with Plotly

Launches with a payload mass between 2000 and 5500 kilograms tended to be the mosts successful.

Machine Learning Predictions

The final SpaceX launch predictions were performed by using 4 different classification models: a decision tree, support vector machine, logistic regression and the k-nearest neighbor algorithm. From the confusion matrices of the models, it is unclear which model performs the best - largely due to a high number of false positives.
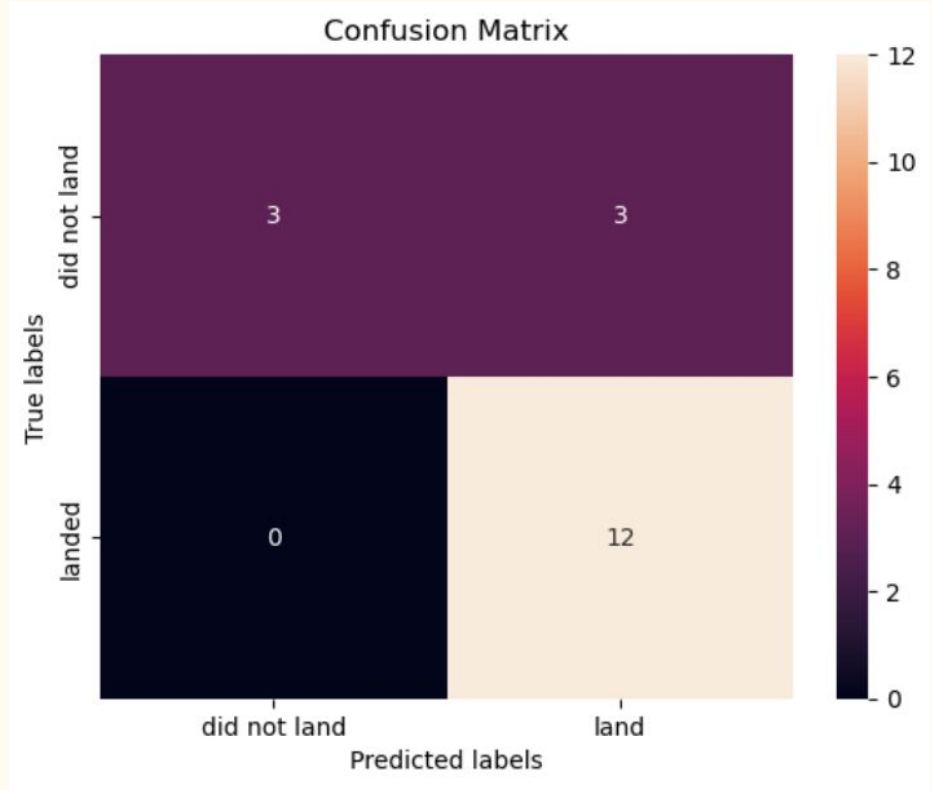
Machine Learning Predictions

As demonstrated by the overall accuracy tests, the decision tree model performed best. However, given our small sample size (18 test cases) it would be best to perform more tests going forward.

Models	Score
Logistic Regression	84.64
Support Vector Machine	84.82
Decision Tree	87.14
K Nearest Neighbor	84.82