IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Brian McCarthy

September 19, 2024

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection through API and with Web Scraping
  - Data Wrangling
  - Exploratory Data Analysis with SQL and Data Visualization
  - Interactive Visual Dashboards with Folium and Dash
  - Predictive Analysis
- Summary of all results
  - Exploratory Data Analysis results
  - Visual Analytics results
  - Predictive Analytics results

Introduction

- Project background and context
  - SpaceX advertises its Falcon 9 rocket with a cost of $62 million, which is far less than other rockets costing upwards of $165 million. The cost savings depend on the ability for the first stage of the rocket to land so that it can be reused.
- Problems you want to find answers
  - Which factor among launch site, booster type, landing site, payload and orbit contributes most to predicting whether the first stage will land successfully
  - Can we create an accurate model to predict the success of the first stage landing successfully using machine learning?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:

- Perform data wrangling
  - Data from SpaceX was filtered to isolate Falcon 9 Rockets, cleaned, and combined with historical data from Wikipedia
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

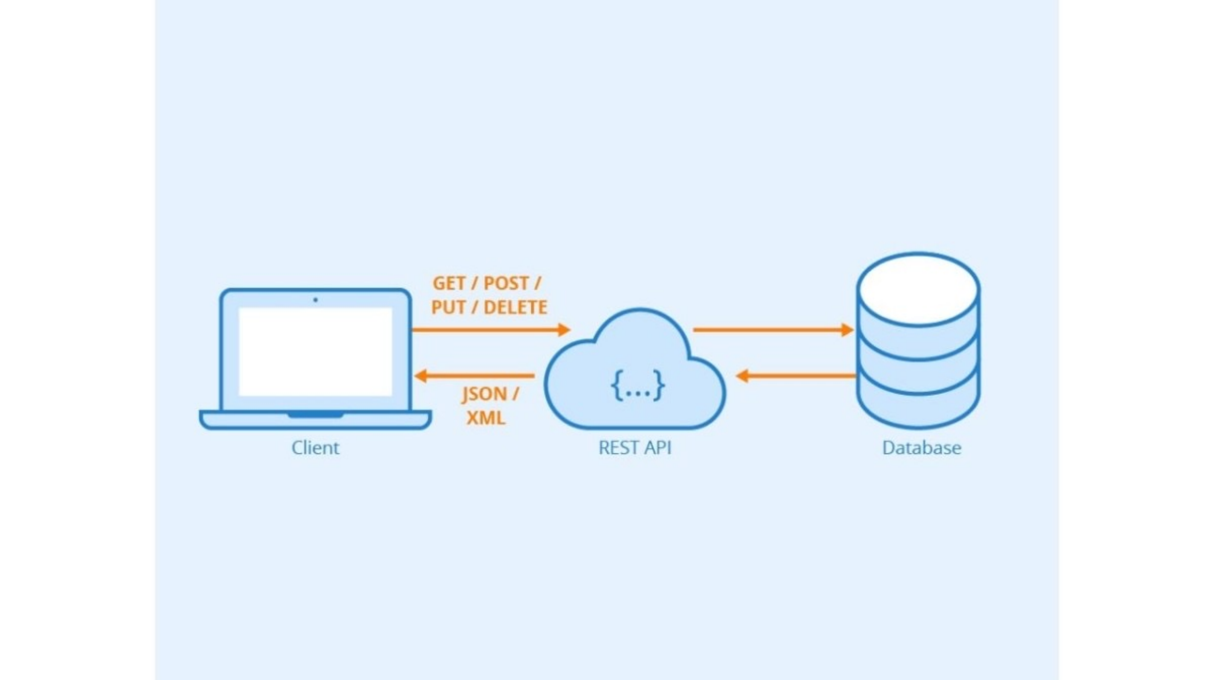
Data Collection

- Data was collected from SpaceX
    - GET requests were made to SpaceX's API to extract information about past launches
        - A Pandas dataframe was created with features including `'rocket'`, `'payloads'`, `'launchpad'`, `'cores'`, `'flight_number'`, `'date_utc'`
    - Data was cleaned and filtered
        - Rockets with multiple cores and payloads were filtered out
        - `'date_utc'` was cleaned to extract only the date
        - Dates were filtered to exclude those after November 13, 2020

Data Collection – SpaceX API

- SpaceX REST were made to the following API's:
- `Launchpads, Payloads, Cores,`
- `Rockets, and Past Launches`

- GitHub URL of the completed SpaceX API calls notebook: https://github.com/bmccarthy12/Capstone/blob/main/jupyter-labs-spacex-data-collection-api.ipynb



Source: iron.io

Data Collection - Scraping

- Data on past launches was scraped from Wikipedia
  - Data was scraped using Beautiful Soup
  - Data was stored in a dataframe

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)

- Github URL: https://github.com/bmccarthy12/Capstone/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

Source: Kaşif, Ahmet & Bilgin, Turgay. (2018). Classification of Turkish Universities by quantity and titles of Academic Staff. 1-4. 10.1109/ISMSIT.2018.8567279.

Data Wrangling

- Describe how data were processed
- Collected data was first filtered to isolate Falcon 9 Data
- Data with multiple entries for ‘payload’ and ‘cores’ was excluded
- Missing data in ‘PayloadMass’ was replaced by the mean mass

- GitHub URL: https://github.com/bmccarthy12/Capstone/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- Summarize what charts were plotted and why you used those charts
  - Used categorical plots to visualize the relationship between
    - Flight Number and Launch Site
    - Flight Number and Payload Mass
  - Used scatter plots to visualize the relationship between
    - Payload Mass and Launch Site
    - Flight Number and Orbit type
    - Payload and Orbit type
  - Used a bar chart to compare success rates of each orbit
  - Used a line chart to visualize launch success trends by year
- GitHub URL: https://github.com/bmccarthy12/Capstone/blob/main/jupyter-labs-eda-dataviz.ipynb

EDA with SQL

- Used the following SQL Queries:
  - `SELECT DISTINCT Launch_Site FROM SPACEXTABLE`
  - `SELECT SUM(Payload_Mass__kg_) FROM SPACEXTABLE WHERE Customer='NASA (CRS)';`
  - `SELECT AVG(Payload_Mass__kg_) FROM SPACEXTABLE WHERE Booster_Version LIKE "F9 V1.1";`
  - `SELECT MIN(Date) FROM SPACEXTABLE WHERE Landing_Outcome LIKE "Success (ground pad)";`
  - `SELECT COUNT(Mission_Outcome) FROM SPACEXTABLE WHERE Mission_Outcome LIKE 'Failure%';`
  - `SELECT DISTINCT Booster_Version FROM SPACEXTABLE WHERE Payload_Mass__kg_ = (SELECT MAX(Payload_Mass__kg_) FROM SPACEXTABLE`
- GitHub URL: https://github.com/bmccarthy12/Capstone/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Summarize what map objects such as markers, circles, lines, etc. you created and added to a folium map
  - Circles and markers were added for each location of launch sites
  - Lines were drawn to show distance between the launch site and the coast and the site and Orlando
- Objects were added to visualize locations of launch sites
- Add the GitHub URL https://github.com/bmccarthy12/Capstone/blob/main/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- Summarize what plots/graphs and interactions you have added to a dashboard
  - A dashboard was created to show the following:
    - Pie graph comparing success to failure
    - Scatter chart comparing Payload mass and success rate
- Explain why you added those plots and interactions
- GitHub URL: https://github.com/bmccarthy12/Capstone/blob/main/spacex_dash_app.py

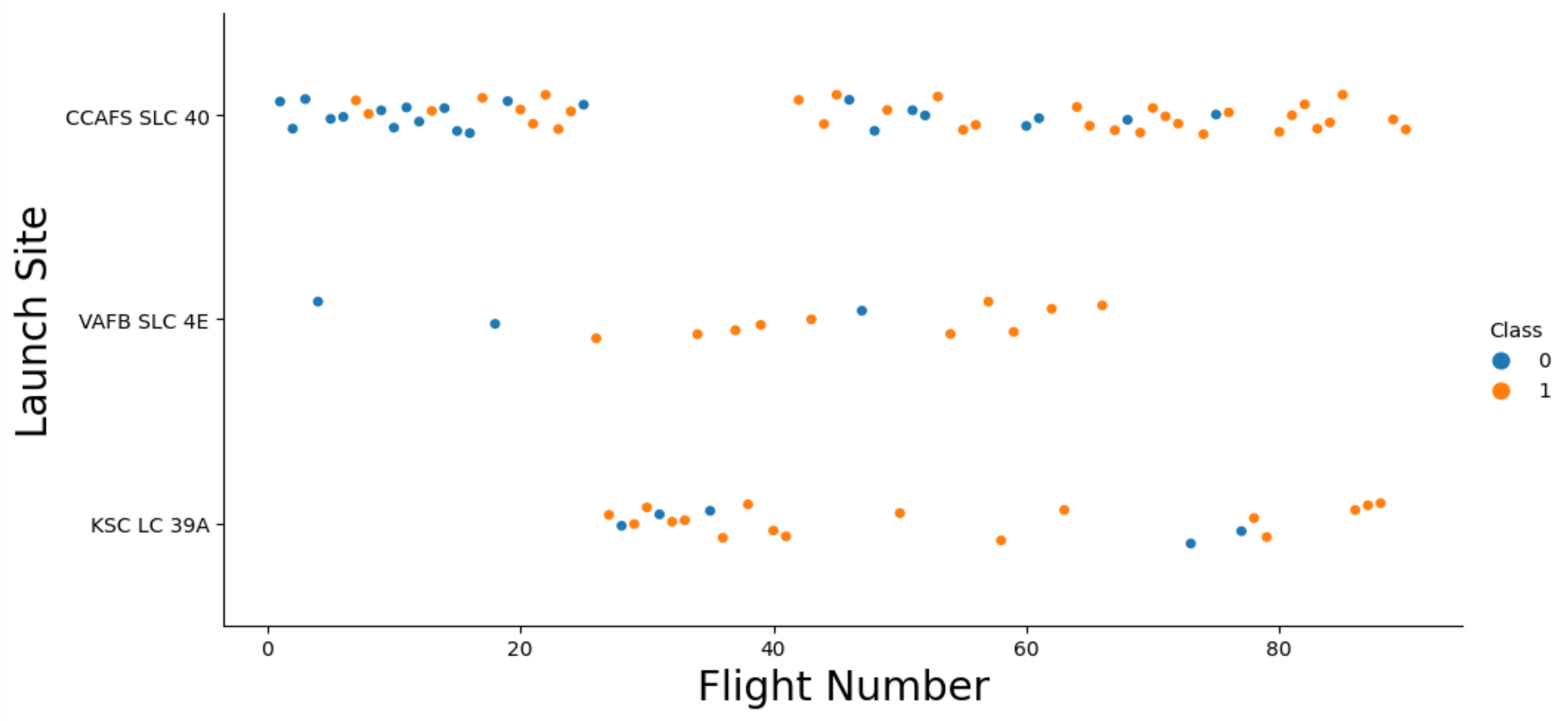
Predictive Analysis (Classification)

- Data was loaded using numpy and pandas and split into training and testing.
- Used GridSearchCV to optimize hyperparameters in making modes
- Linear Regression, Logistic Regression, K Nearest Neighbors, Support Vector Machines and Decision Trees were used to create models.
- X_Train and Y_Train were used to fit data to the models and were evaluated using Test data
- GitHub URL: https://github.com/bmccarthy12/Capstone/blob/main/SpaceX-Machine-Learning-Prediction-Part-5-v1.ipynb

Section 2

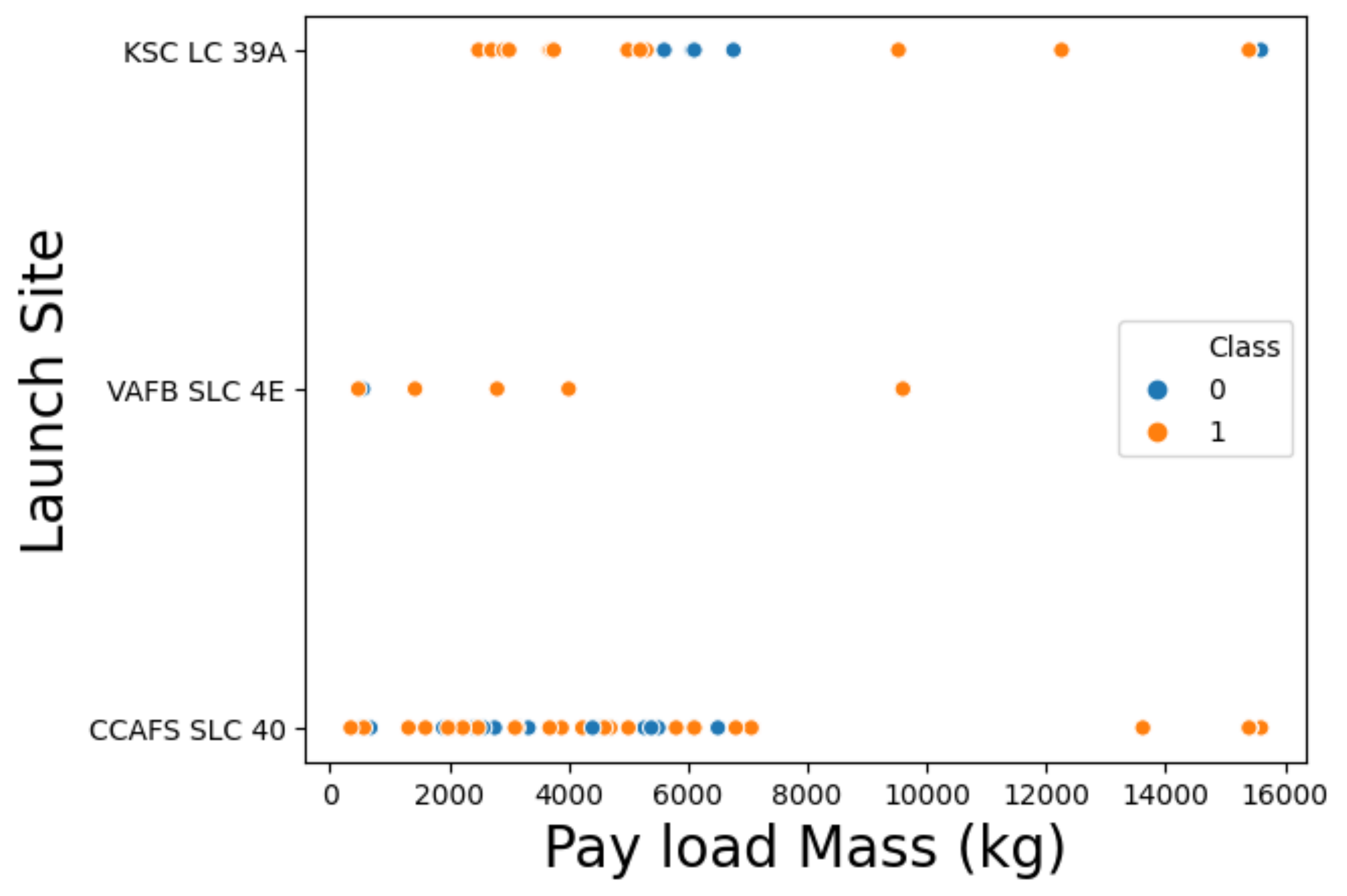
Insights drawn from EDA

Flight Number vs. Launch Site



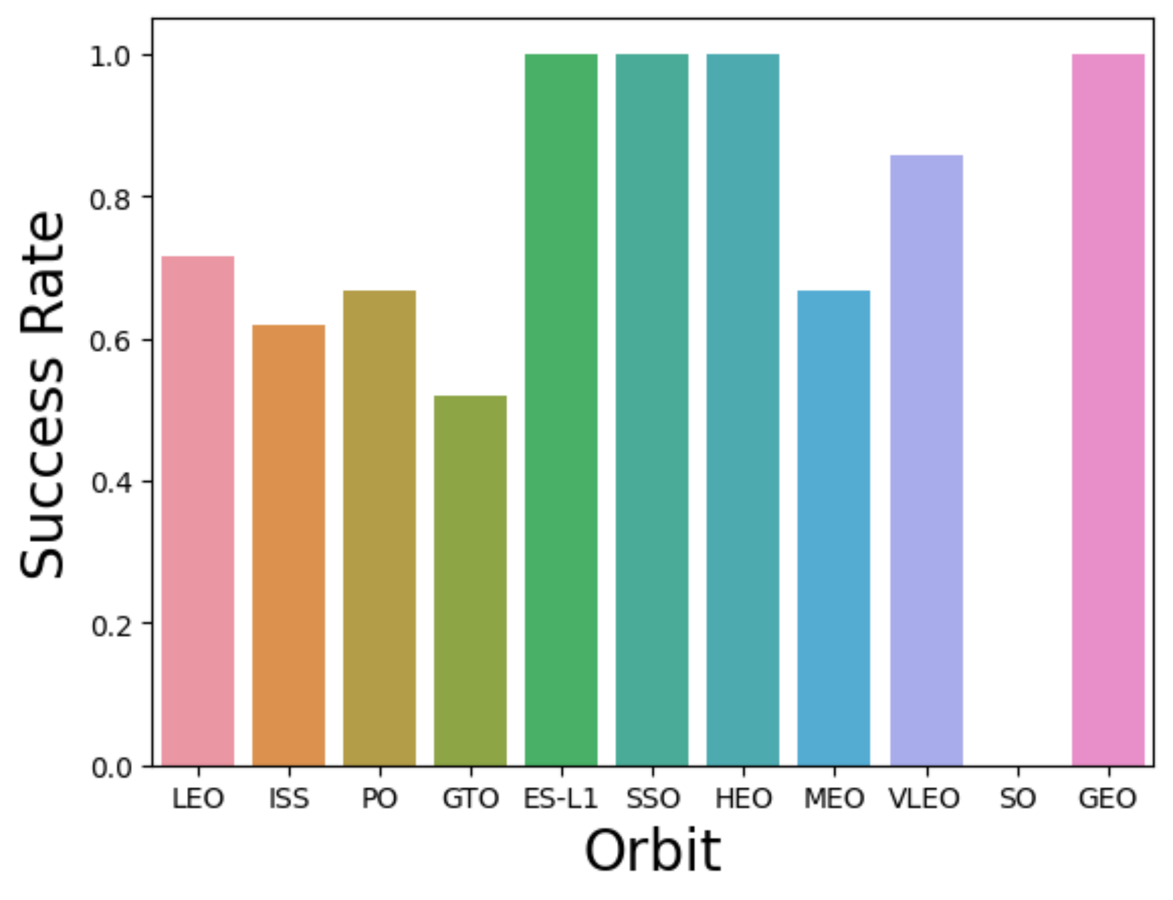
- Successful landings (Class = 1) became more common with later flight numbers

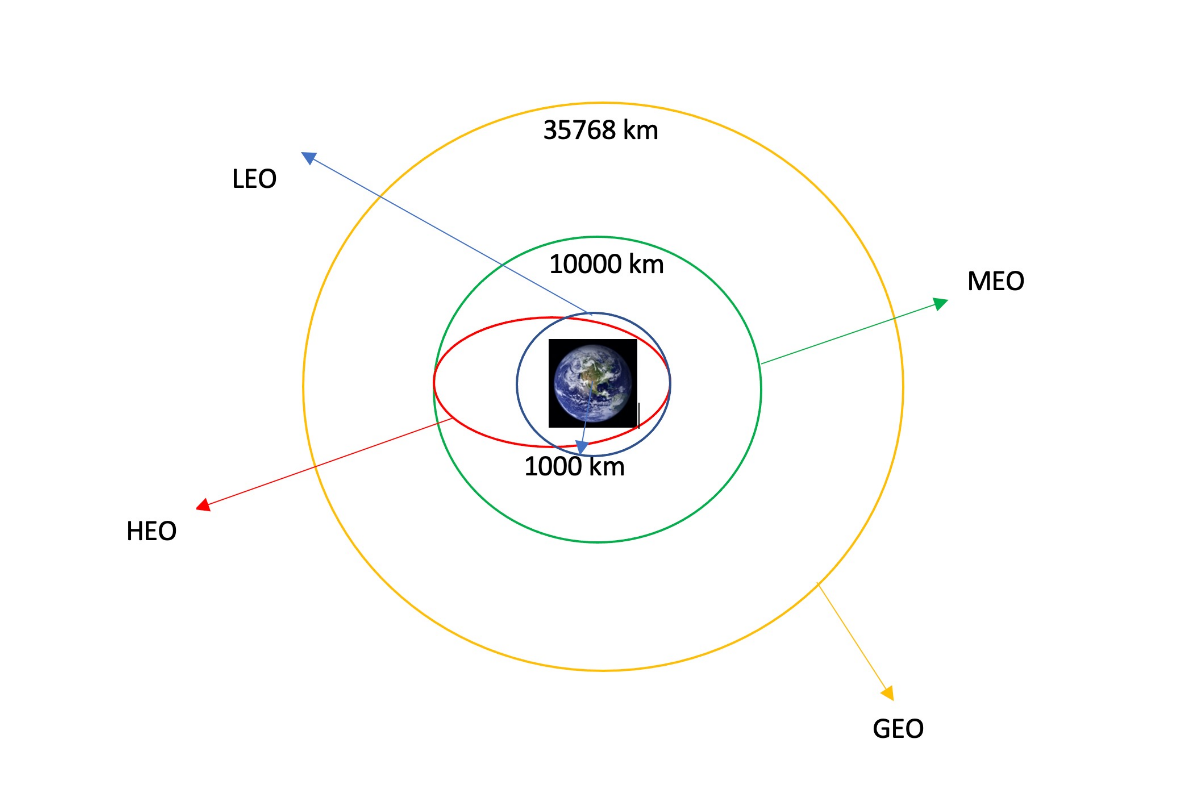
Payload vs. Launch Site



- Launch site VAFB had the fewest launches and only launched lighter payloads than the other sites

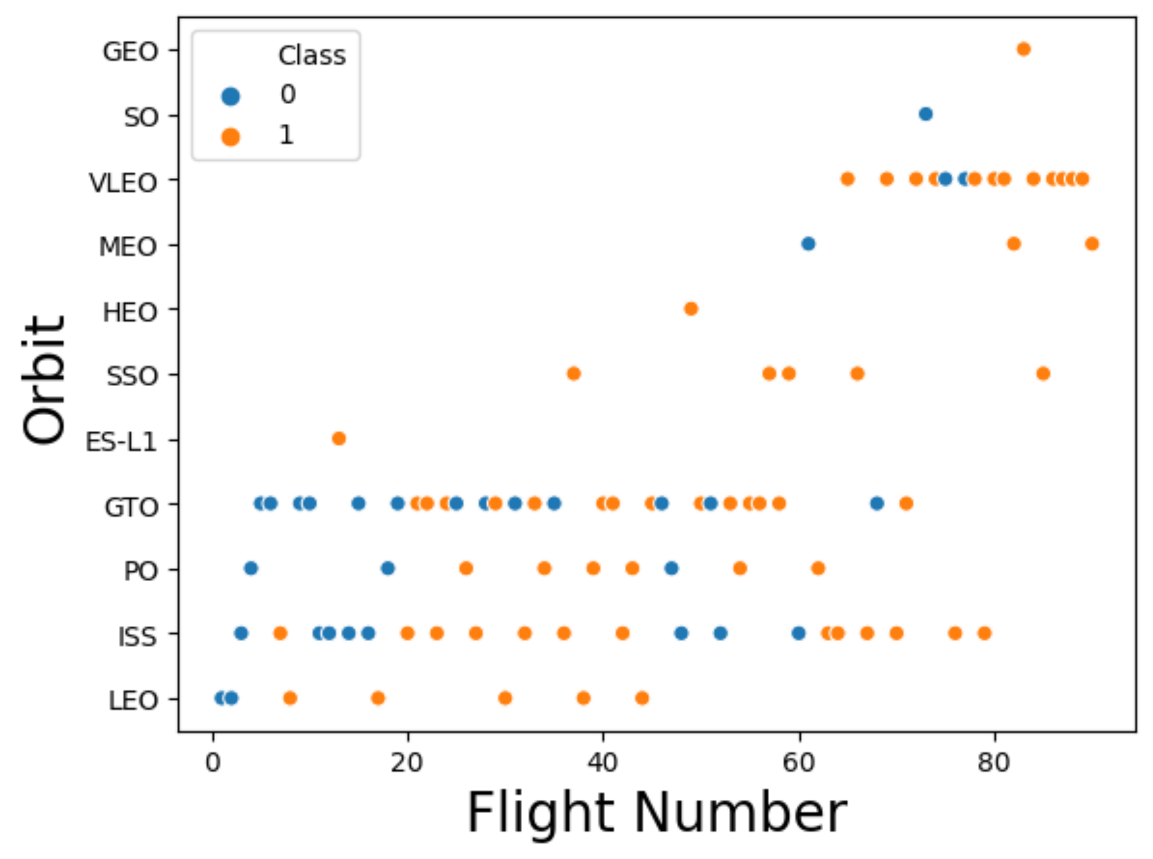
Success Rate vs. Orbit Type





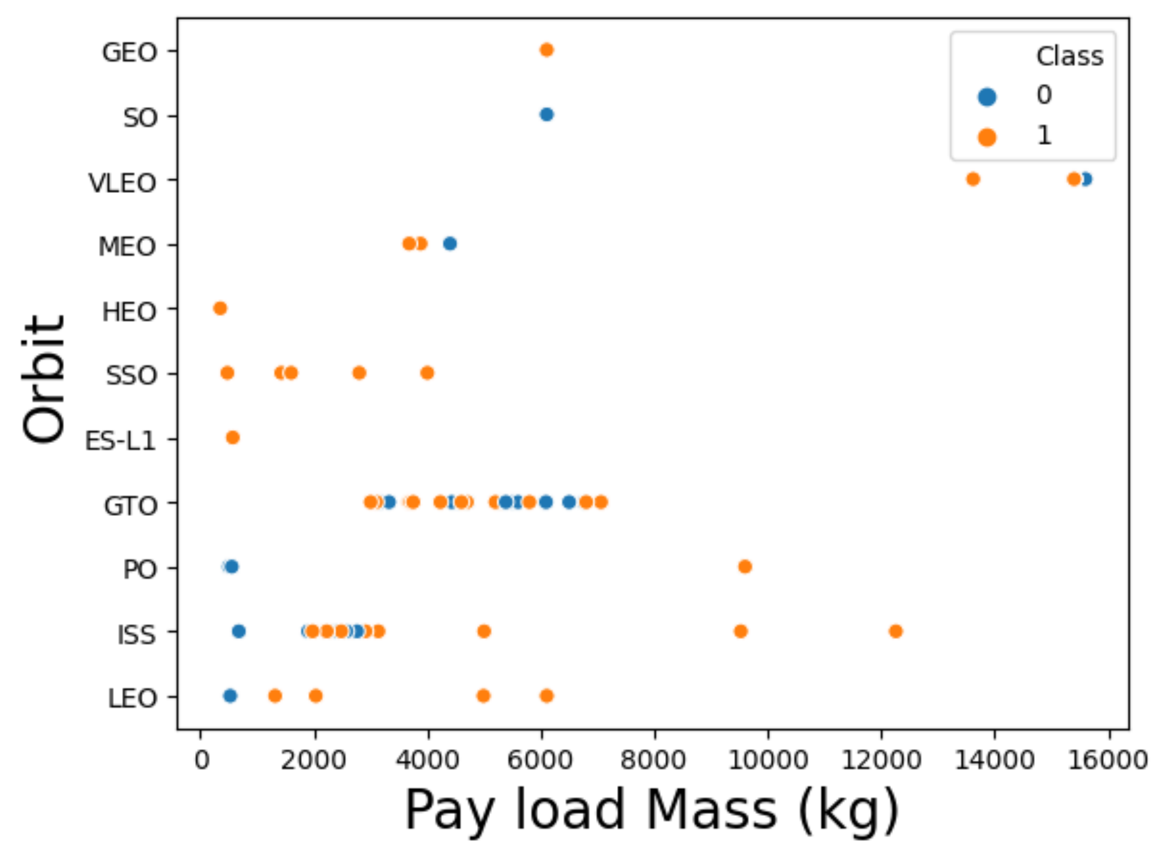
ES-L1, SSO, HEO, and GEO orbits have a 100% success rate

Flight Number vs. Orbit Type



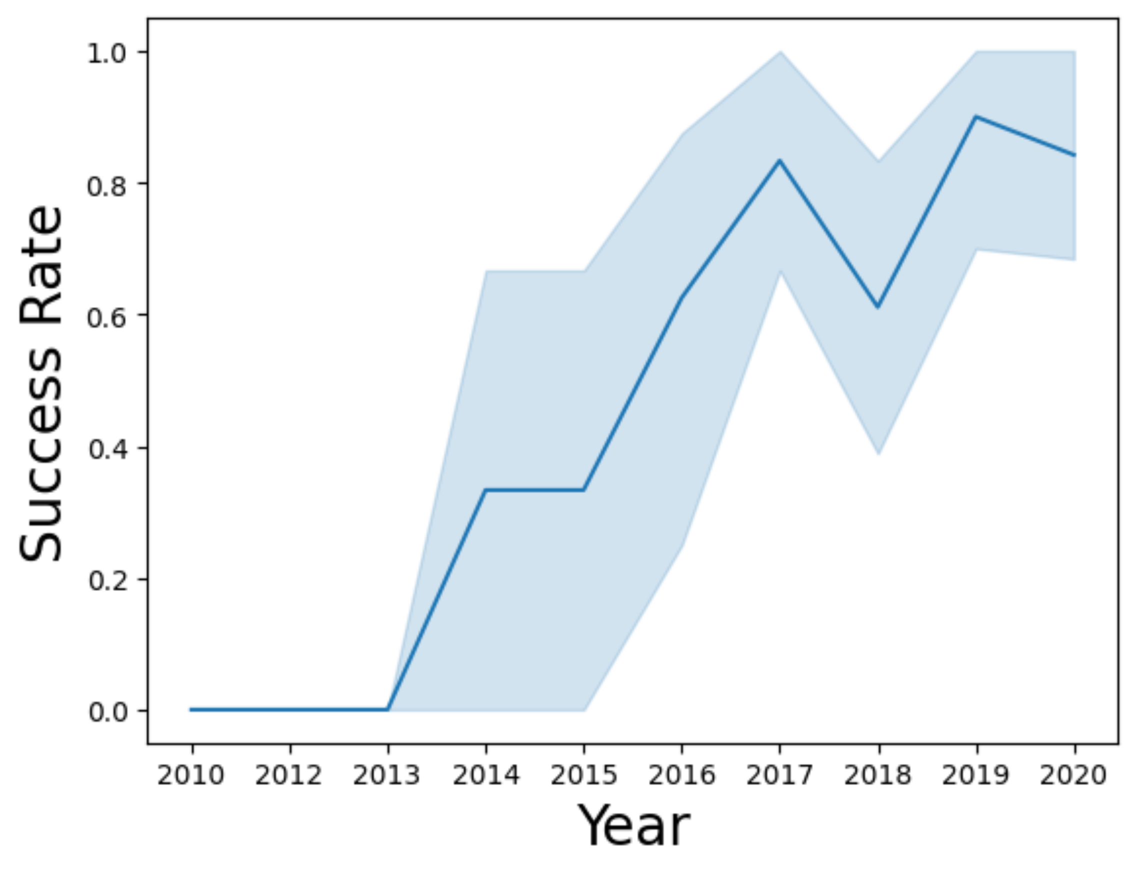
- LEO orbits became more successful as flight number increased suggesting improvement over time.

Payload vs. Orbit Type



- Heavier rockets are successful in PO, ISS, and LEO, but there is little correlation between payload mass and success in GTO orbits

Launch Success Yearly Trend



- Other than a brief decline in 2018, the success rate has been increasing or stayed level since 2013

All Launch Site Names

```
%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

- A query of SpaceX data shows four distinct launch sites

Launch Site Names Begin with 'CCA'

```
%sql SELECT * FROM SPACEXTABLE WHERE Launch_Site LIKE 'CCA%'LIMIT 5;
```

✓ 0.0s Python

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

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	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	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

- The first 5 records from launch site CCAFS LC-40 are shown. Launches from CCAFS SLC-40 took place later

Total Payload Mass

```
%sql SELECT SUM(Payload_Mass__kg_) FROM SPACEXTABLE WHERE Customer='NASA (CRS)';
```

✓ 0.0s

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

SUM(Payload_Mass__kg_)
45596

- The total mass carried by rockets launched by NASA (CRS) is 45596

Average Payload Mass by F9 v1.1

```
%sql SELECT Booster_Version,AVG(Payload_Mass__kg_) FROM SPACEXTABLE WHERE Booster_Version LIKE "F9 V1.1";
```

✓ 0.0s

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

Booster_Version	AVG(Payload_Mass__kg_)
F9 v1.1	2928.4

- The average payload of F9 v1.1 has a mass of 2928.4 Kg

First Successful Ground Landing Date

```
%sql SELECT MIN(Date) FROM SPACEXTABLE WHERE Landing_Outcome LIKE "Success (ground pad)";
```

✓ 0.0s

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

MIN(Date)
2015-12-22

- The first successful landing on a ground pad was in 2015

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%sql SELECT Booster_Version FROM SPACEXTABLE WHERE Landing_Outcome='Success (drone ship)' AND Payload_Mass__kg_ BETWEEN 4000 AND 6000;
```

✓ 0.0s

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

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

- List the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000
  - F9 FT B1022
  - F9 FT B1026
  - F9 FT B1021.2
  - F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

```
%sql SELECT COUNT(Mission_Outcome) FROM SPACEXTABLE WHERE Mission_Outcome LIKE 'Failure%';
```

✓ 0.0s

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

COUNT(Mission_Outcome)
1

```
%sql SELECT COUNT(Mission_Outcome) FROM SPACEXTABLE WHERE Mission_Outcome LIKE 'Success%';
```

✓ 0.0s

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

COUNT(Mission_Outcome)
100

- All but one of the 101 missions were successful

Boosters Carried Maximum Payload

```
%sql SELECT DISTINCT Booster_Version FROM SPACEXTABLE WHERE Payload_Mass__kg_ = (SELECT MAX(Payload_Mass__kg_) FROM SPACEXTABLE);
```

✓ 0.0s

```
* 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

- The Booster that carried the most massive payloads are:

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

2015 Launch Records

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

✓ 0.0s

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

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

- 2015 saw two unsuccessful landings on drone ships

Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

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

✓ 0.0s

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

Landing_Outcome	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

- 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

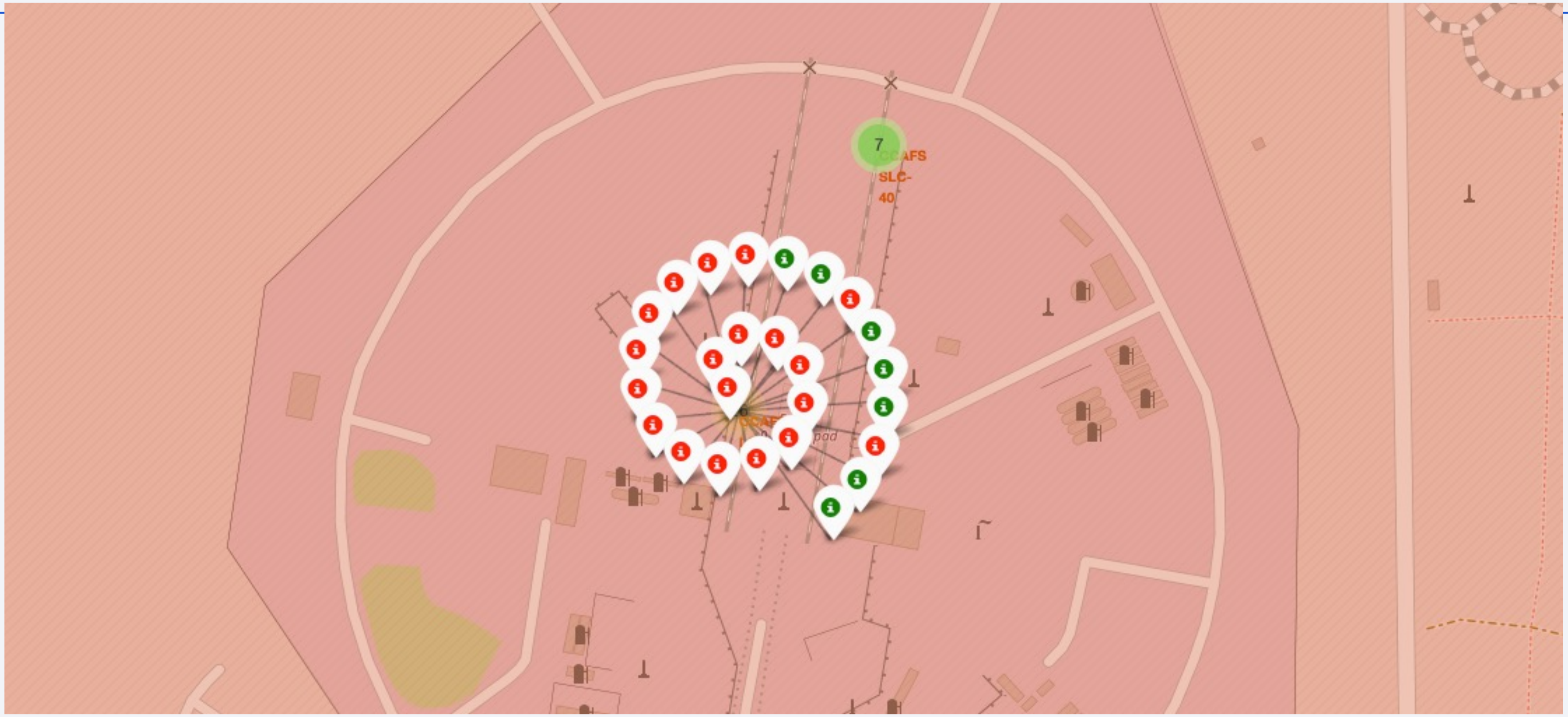
Section 3

Launch Sites Proximities Analysis

Locations of launch sites

- Launch sites were clustered in Florida and California. More ships launched from Florida

Launches from Site CCAFS LC-40



- This map shows the launch sites from Florida. Launches with successful landing are labeled in green

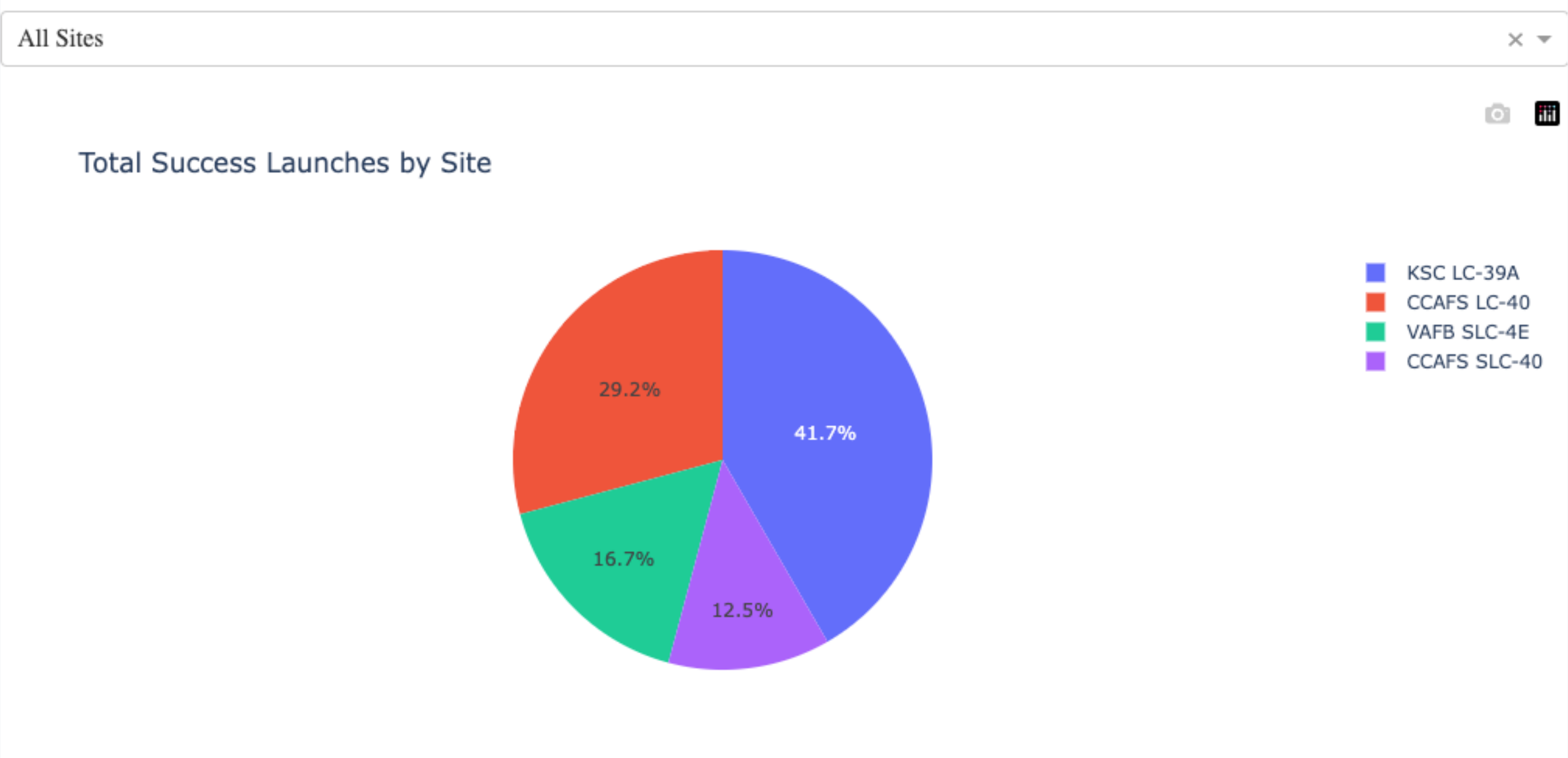
Distance between Orlando and Launch Site

- Line showing distance from launch site to Orlando

Section 4

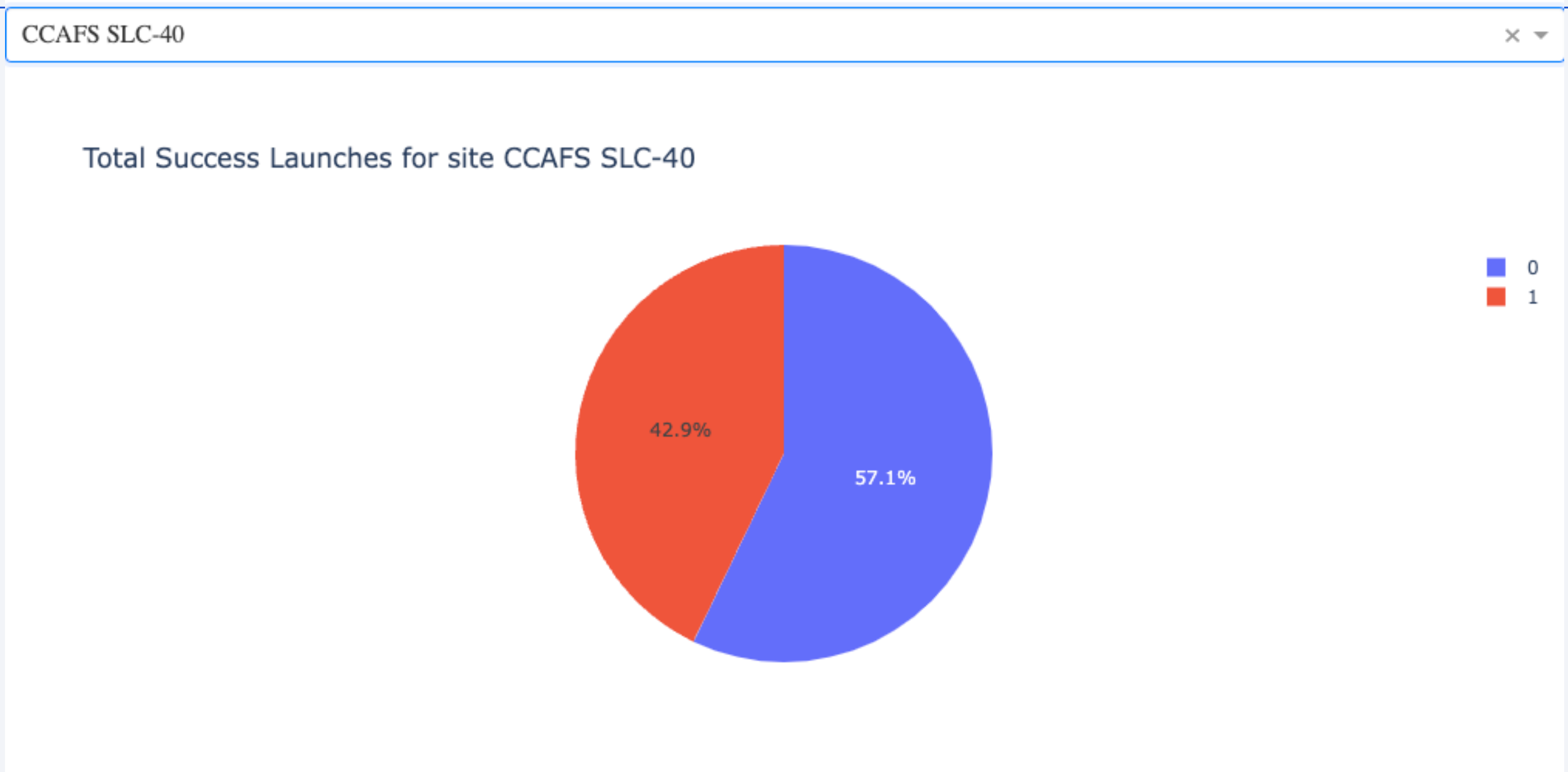
Build a Dashboard with Plotly Dash

Launch Success



- Launch Site KSC LC-39A had the most successful launches

<Dashboard Screenshot 2>



- Site CCAFS SLC-40 had the highest success ration

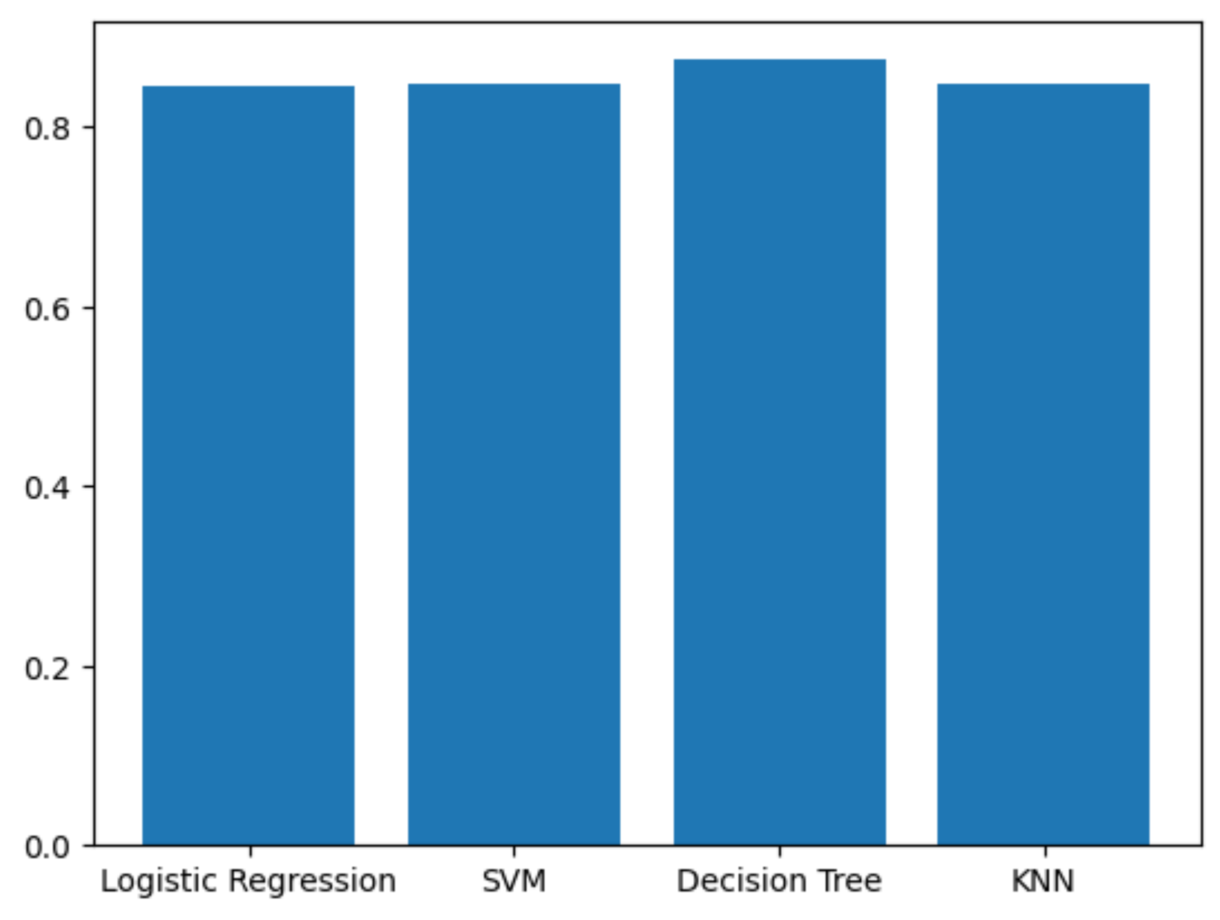
Payload vs. Launch Outcome

- FT Boosters carrying lighter payloads have a high success rate

Section 5

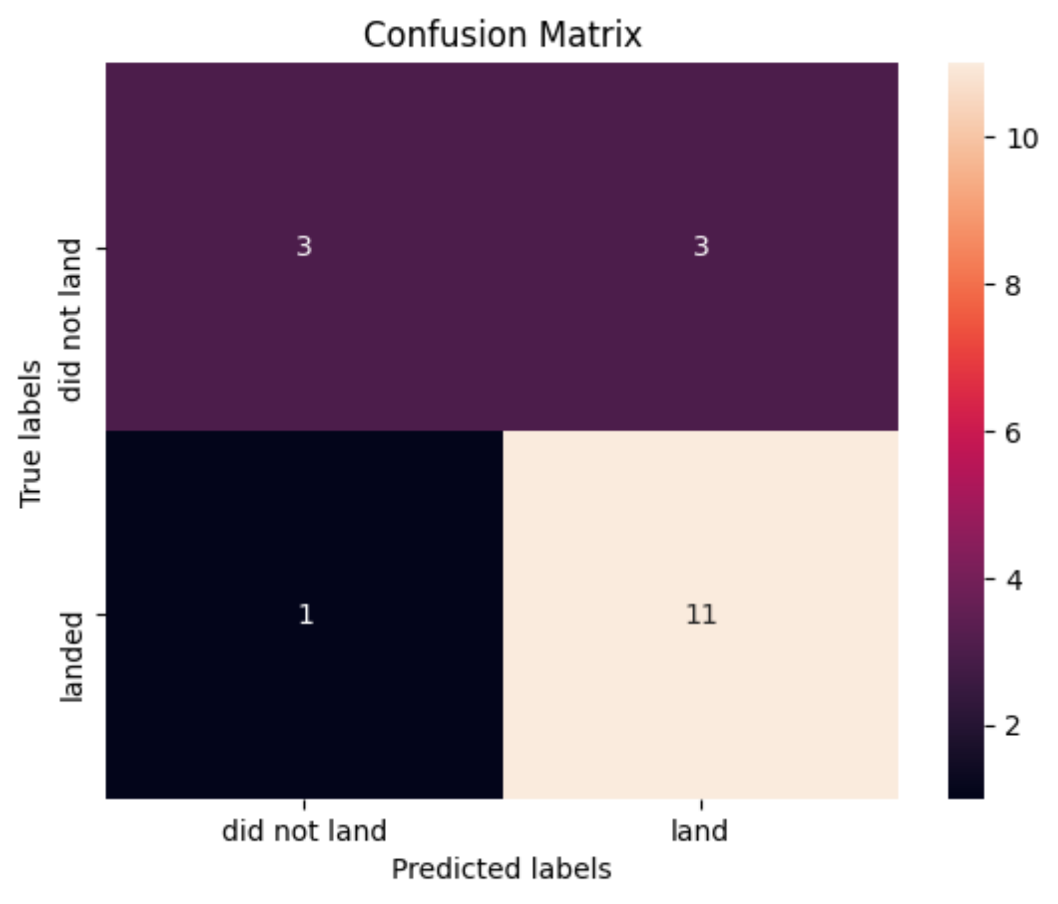
Predictive Analysis (Classification)

Classification Accuracy



- The model accuracy was high for all models, but the decision tree model was most accurate

Confusion Matrix



- The confusion matrix for the decision tree model led to 3 false positive and 11 true positives on the test data

Conclusions

- The Decision Tree Model is best at predicting a successful landing of the first stage
- Launch success rate started increasing in 2013 and other than one year has increased or remained level until 2020.
- Orbits ES-L1, GEO, HEO, SSO, and VLEO had the most success rate
- Launch Site KSC LC-39A had the highest success rate

Appendix

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project
- All notebooks used in creation of this report can be found at:
  - https://github.com/bmccarthy12/Capstone

Thank you!