Winning Space Race with Data Science

Robert Turner
07-24-2023

Outline

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

Executive Summary

- Data was gathered from the SpaceX API
- Launches from site KSC LC-39A succeeded roughly 77% of the time
- Fins and legs on the stage 1 rocket are effective means of aiding the recovery of the 1st stage
- Missions targeted at GTO, ISS, LEO, MEO, or PO orbits were less likely to be recovered
- Stage 1 recoveries have become more successful over time

Introduction

- Many rocket providers are available on the market to launch payloads, but SpaceX typically provides cheaper launches
- Unlike the majority of providers, SpaceX is capable of recovery the first stage of the rocket
  - The first stage accounts for most of the launch work, and each successful recovery cuts down on costs
- Not all recoveries are successful
- This analysis attempted to determine whether a launch will result in a stage 1 recovery, and what factors influence the recovery

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Version 4 of the SpaceX API provided data for past rocket launches
- Performed data wrangling
  - Missing data was replaced with the mean of its feature
- Performed exploratory data analysis (EDA) using visualization and SQL
- Performed interactive visual analytics using Folium and Plotly Dash
- Performed predictive analysis using classification models
  - Used GridCVSearch to estimate optimal parameters for all models
  - Fit Logistic Regression, Support Vector Machine, Decision Tree Classifier, and K-Nearest Neighbor models to compare efficacy

Data Collection

- The SpaceX API was used for all data gathering for past rocket launches
  - Webscraping of SpaceX's Wikipedia article was completed, but this data was not used in later analysis
- Code for all collection and wrangling available on the project GitHub

Data Collection – SpaceX API

- Launches/past API used to download primary data table for launches
- Rockets, launchpads, payloads, and cores APIs subsequently used to map raw IDs to useful features
- Notebook with completed scraping available on GitHub.

1. Raw rocket data collected through launches/past API, including raw IDs to direct following calls
2. Rocket IDs used to gather versions info for each booster from rockets API
3. Launchpad IDs used to gather Launch Site name, latitude, and longitude from launchpads API
4. Payload IDs used to gather payload weight and targeted orbit from payloads API
5. Core IDs used to gather specific information for stage 1 rockets from cores API
    1. Leg info, fin info, whether or not the stage had been reused, target landing pad

Data Collection - Scraping

- Past launches were scraped from the Falcon 9 and Heavy Launches Wikipedia page
- All data in this table is usable without referencing other data sources
- Notebook with completed scraping available on GitHub

1. Scraped past launches from the linked Wikipedia page
2. Payload mass converted from string, all masses were in KG
3. Landing status converted to boolean from string
4. Booster version was reformatted from table string

Data Wrangling

- Most data wrangling was performed during collection to map IDs to usable data, and clean up unusual strings
- Afterward, outcomes were converted into boolean variables for later model fitting
  - Any confirmed landing was considered a successful landing, regardless of intended landing destination
  - Landings that were not reported or confirmed to fail were considered failures
- Some payload data was missing; that data was interpolated with the average payload mass
- Data wrangling notebook available on GitHub.

EDA with Data Visualization

- Flight Number vs. Payload Mass shows the change in payload mass over time
- Flight Number vs. Launch Site shows how SpaceX has used each launch site over time
- Payload Mass vs. Launch Site shows which launch sites SpaceX uses for heavier / lighter loads
- Orbit vs. Success Percentage shows which orbits are most likely to succeed
- Flight number vs. Orbit shows the trend in orbit targets over time
- Payload Mass vs. Orbit shows which payload weights tend to target each orbit
- Year vs. Success Rate shows how recovery success rate has changed over time
- All graphs colored with success/failure to show how compared factors impacted recovery success
- EDA code available on GitHub

EDA with SQL

- Queried for all unique launch site names to list our launch sites
- Queried for 5 sample records from launch sites starting with "CCA"
- Calculated total payload mass carried by boosters launched from NASA
- Displayed the average payload mass carried by a specific booster version for comparison
- Listed the date of the first successful recovery
- Listed boosters used to carry 4000 – 6000 kg weights
- Showed the success rate of missions (ignoring recovery result)
- Showed the boosters capable of carrying the largest possible payload
- Showed all recovery failures in 2015 and relevant features to those failures
- Broke down landing results for landings between 2010 and 2017
- SQL Code available on GitHub

Build an Interactive Map with Folium

- Created a map with all launch sites and all recovery results from each launch sites
  - Attempting to see whether specific launch sites have a higher failure rate
- Drew distance lines from the CCAFS LC-40 launch site to nearby coastlines, trains, highways, and cities
  - Illustrated the fact that while launch sites after close to trains and highways for easy component delivery, launch sites also need to be near coastlines and away from cities for safety reasons
- Map notebook available on GitHub.

Build a Dashboard with Plotly Dash

- Created a dashboard which can view data for all sites or for individual sites
  - All sites view showed mission frequency for each launch site in a pie chart, and a scatter plot of payload mass against the success or failure of the recovery colored by launch site
  - Single site view showed frequency of recovery success in a pie chart, and a scatter plot of payload mass against the success or failure of the recovery colored by the booster used
- The goal for each graph was the same; we were looking for launch sites which were more prone to recovery failure or a relationship between payload size and failure rate.
- Python code available on GitHub.

Predictive Analysis (Classification)

- For each model, GridCVSearch was used to select optimal parameters
  - GridCVSearch used 10-fold cross-validation to test a range of parameters for each model and selected the best
- The following classifiers were attempted on this data:
  - Logistic Regression
  - Support Vector Machine
  - Decision Tree Classifier
  - K-Nearest Neighbors:
- Notebook available on GitHub.

Results

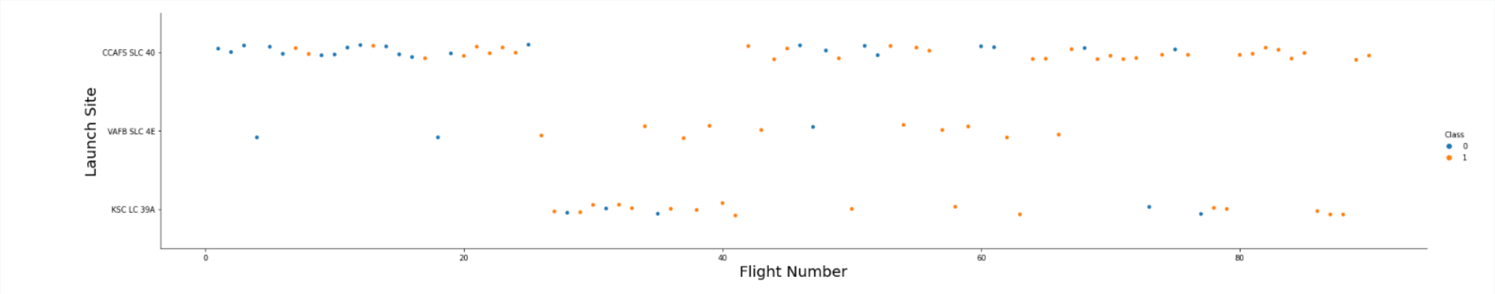
- Higher weight payloads appear more prone to failure
- Orbit targets GTO, ISS, LEO, MEO, PO, and SO are mor prone to failure
- The presence of grid fins or legs on the stage 1 rocket increases the odds of recovery
- Dashboard shows site CCAFS LC-40 is a heavily used site and prone to failure
  - Also shows some boosters are not used for higher weight payloads
- All models show roughly 85% rate of successful predictions and performed equally well on the test set
  - False positives were the most likely error made by each model

Section 2

Insights drawn from EDA

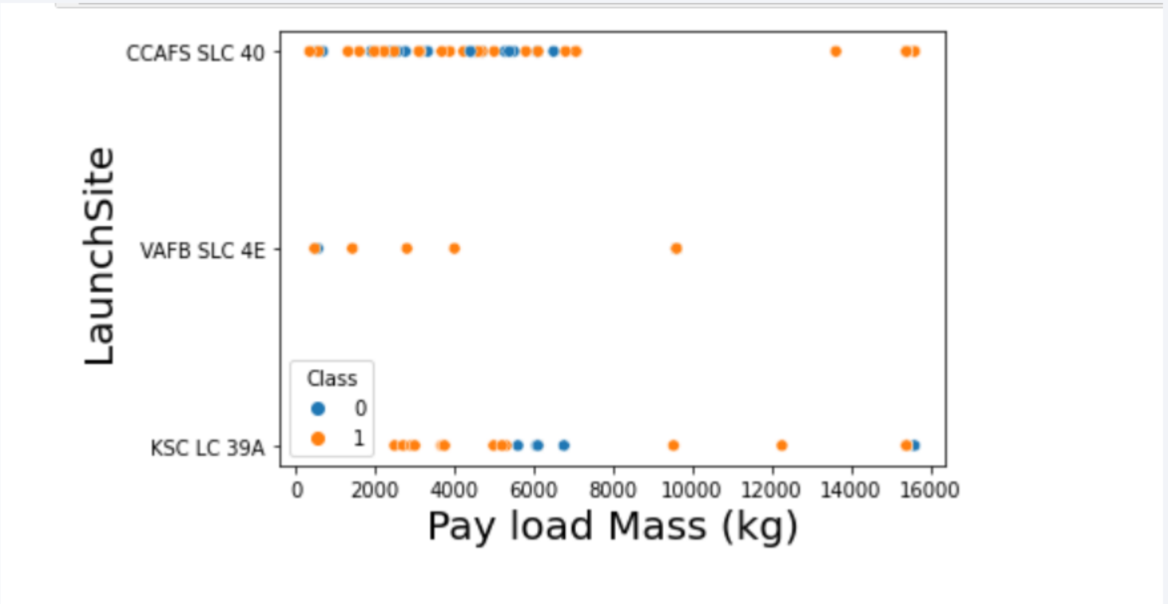
Flight Number vs. Launch Site

- The launch site most commonly used by SpaceX varied over time
- VAFB SLC 4E produced the highest rate of success, but also was not frequently used
- As flight number increased, the recoveries became more successful



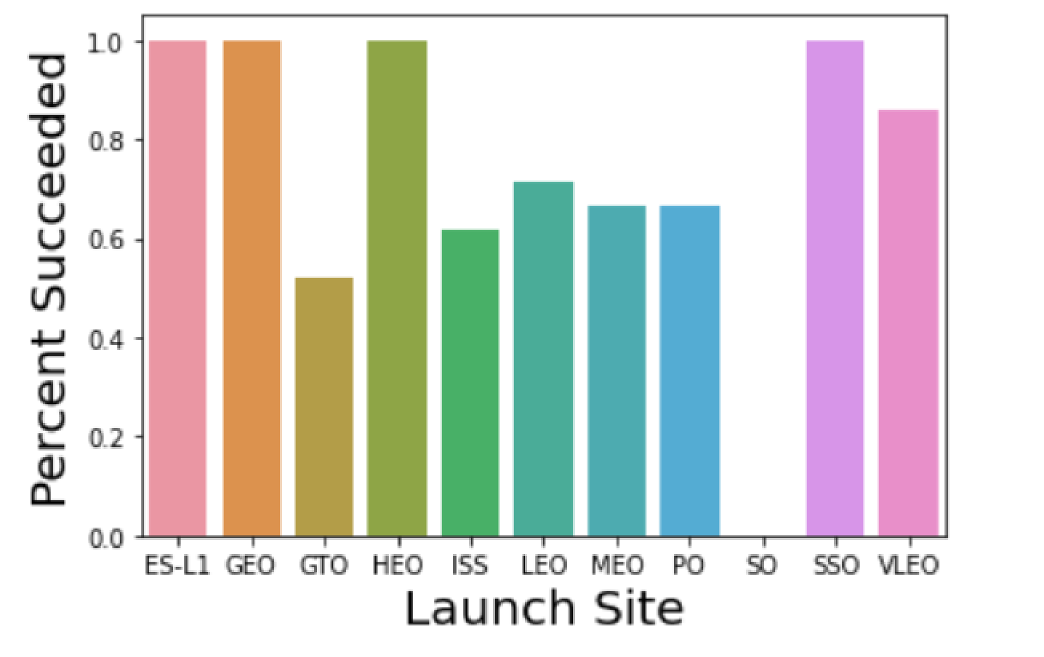
Payload vs. Launch Site

- Most launches performed had a payload of less than 10,000 kg
- VAFB SLC 4E was primarily used for smaller launches
- Larger payloads tend to produce more failures than smaller payloads



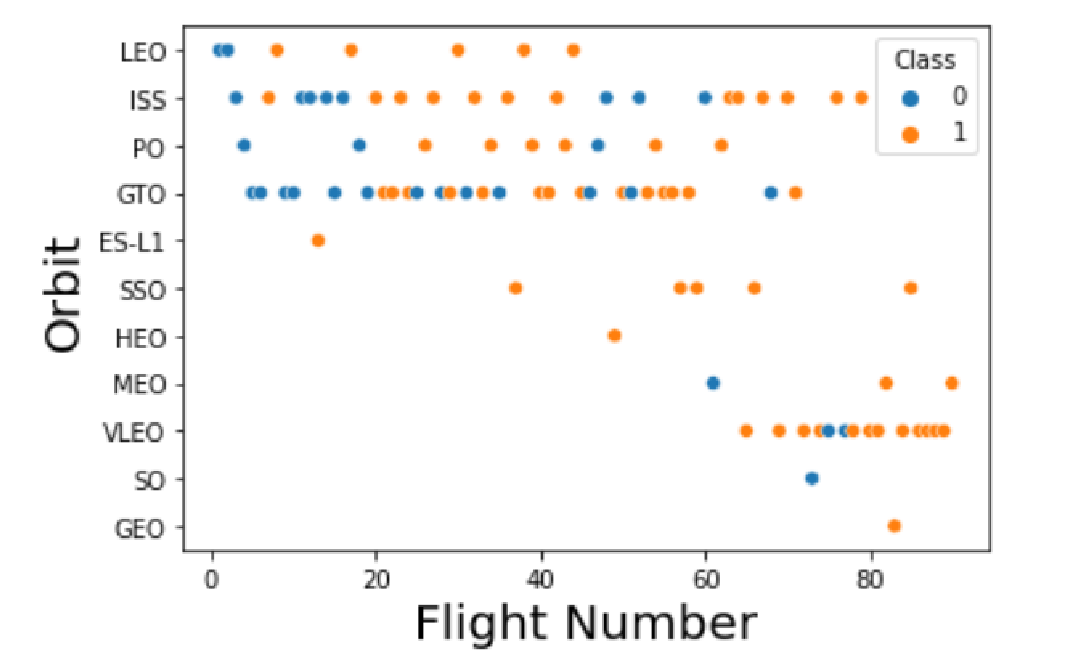
Success Rate vs. Orbit Type

- Missions targeted at ES-L1, GEO, HEO, and SSO orbits were always successfully recovered
- Missions targeted at GTO, ISS, LEO, MEO, or PO orbits were less likely to be recovered.
- No missions targeted at a SO orbit were recovered



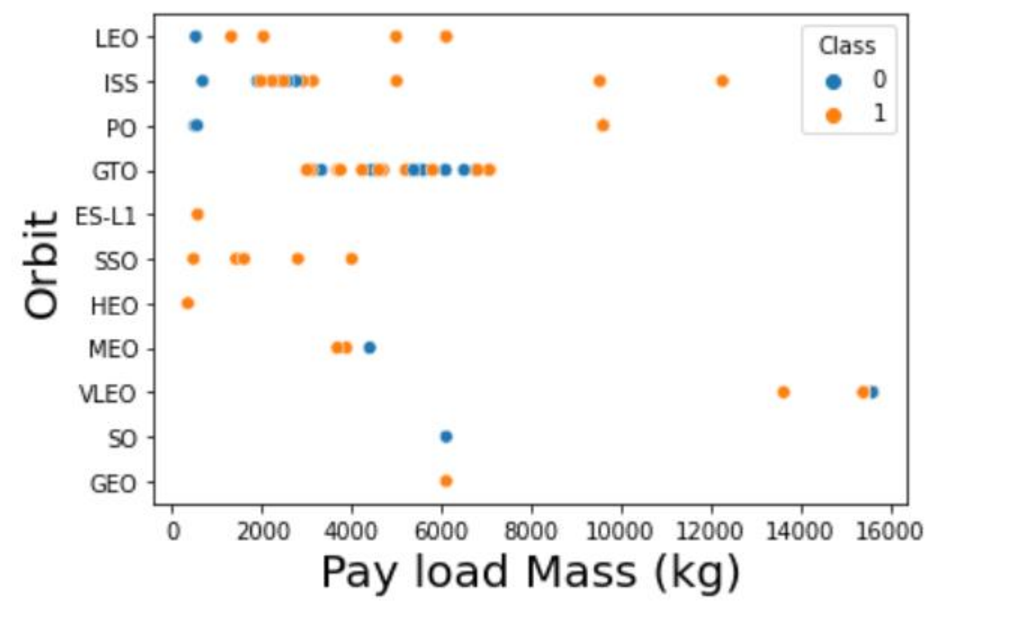
Flight Number vs. Orbit Type

- GTO was the most common orbit target that also included a high failure rate
- Only 1 orbit was attempted for the SO, ES-L1, or GEO target
- The LEO, SSO, and PO targets were commonly used, and had a high success rate



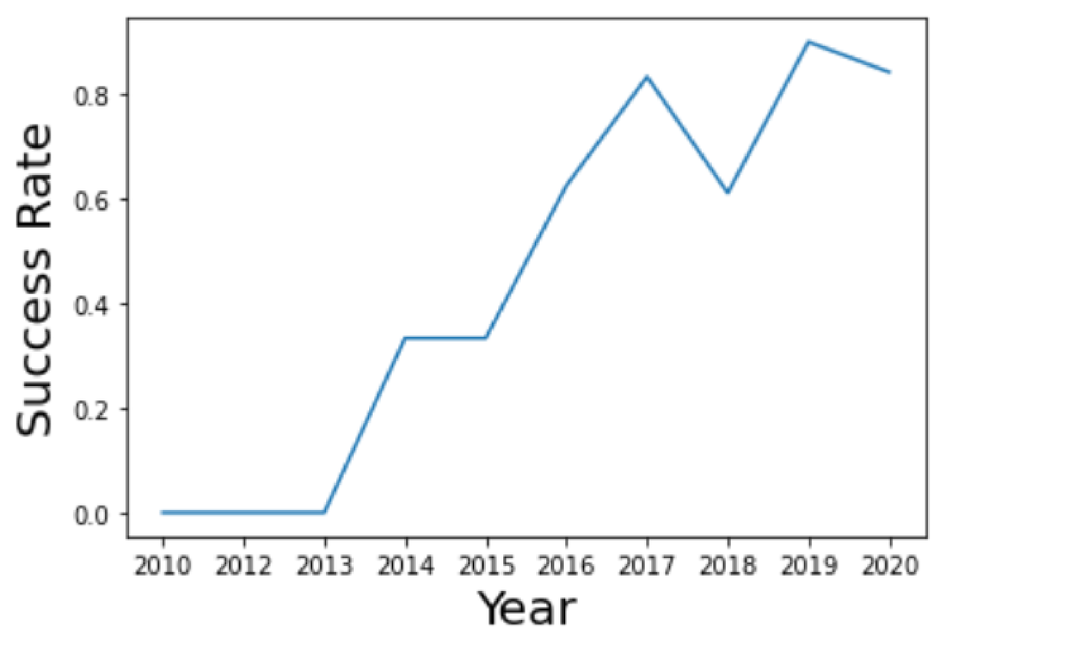
Payload vs. Orbit Type

- No payloads heavier than 8,000 kg were sent to LEO, GTO, ES-L1, SSO, HEO, MEO, SO, or GEO orbits
- For GTO orbits, payload mass appears to not make a major difference
- For ISS orbits, larger payloads were more likely to have a successful stage 1 recovery



Launch Success Yearly Trend

- Since 2013, successful stage 1 recoveries have increased along with the year
- In 2018 there was a drop in success rate back to 2016 levels



All Launch Site Names

- There are only 4 unique launch sites present in this dataset
- Launch Sites:
    - CCAFS LC-40
    - VAFB SLC-4E
    - KSC LC-39A
    - CCAFS SLC-40

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

Launch Site Names Begin with 'CCA'

- Launches from CCAFS LC-40 were highly varied, as this is a frequently used launch site

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2010-04-06 00:00:00	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0.0	LEO	SpaceX	Success	Failure (parachute)
2010-08-12 00:00:00	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0.0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
2012-05-22 00:00:00	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525.0	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-08-10 00:00:00	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500.0	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-01-03 00:00:00	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677.0	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

- For NASA requested missions, 45,596 kg of payload has been moved by SpaceX
- Most of these deliveries likely went to the ISS based on the query from the previous slide

total_payload_mass
45596.0

Average Payload Mass by F9 v1.1

- The average payload sent with an F9 v1.1 booster was 2928 kg
- This visually appears to be roughly average for the entire dataset per our previous graphs.

average_payload_mass
2928.4

First Successful Ground Landing Date

- Prior to September 2013, no attempted stage 1 recoveries were successful

min(Date)
2013-09-29 00:00:00

Successful Drone Ship Landing with Payload between 4000 and 6000

- Many boosters have experienced a successful landing on a drone ship with average sized payloads

Booster_Version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1032.1
F9 B4 B1040.1
F9 FT B1031.2
F9 B4 B1043.1
F9 FT B1032.2
F9 B5 B1046.2
F9 B5 B1047.2
F9 B5 B1046.3
F9 B5 B1048.3
F9 B5 B1051.2
F9 B5B1060.1
F9 B5 B1058.2
F9 B5B1062.1

Total Number of Successful and Failure Mission Outcomes

- Most SpaceX missions are successful, even if the stage 1 recovery is not

- There has only been one SpaceX mission that has ended in failure during the flight

Mission_Outcome	count(Mission_Outcome)
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- Many versions of boosters have successfully delivered the largest attempted payload from SpaceX
- Each of these boosters appears to be part of the B5 series of F9 boosters

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

2015 Launch Records

- In 2015, only two missions resulted in a failed stage 1 recovery
- Both were targeted at drone ship landings
- Both used F9 v1.1 rockets, and were launch from CCAFS LC-40

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

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

- The most likely landing outcome was not attempting to land the stage 1 rocket in the first place
- Landings aimed at the ground always succeeded, but were less often attempted
- Drone ship landings had a 50% failure rate, and were the most commonly used landing site
- Ocean landings were marginally more likely to succeed than fail

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

Section 3

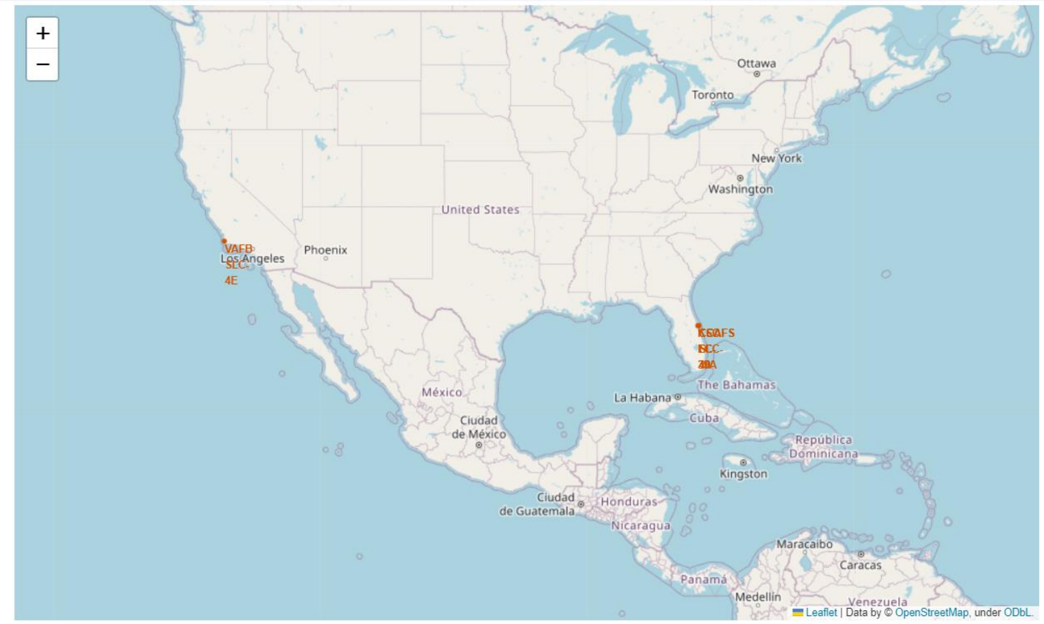
Launch Sites Proximities Analysis

Launch Site Locations

- All launch sites are near the equator, and near coastlines

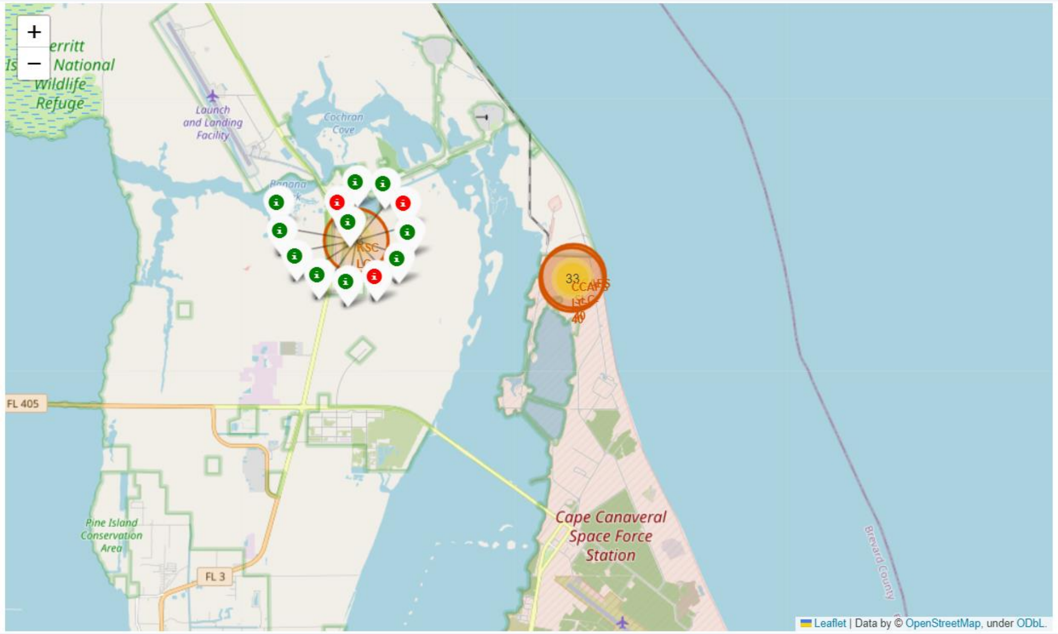
- There are no launch sites in middle America

- Locations are likely selected to make it easy for failed launches or landings to crash in the ocean safely



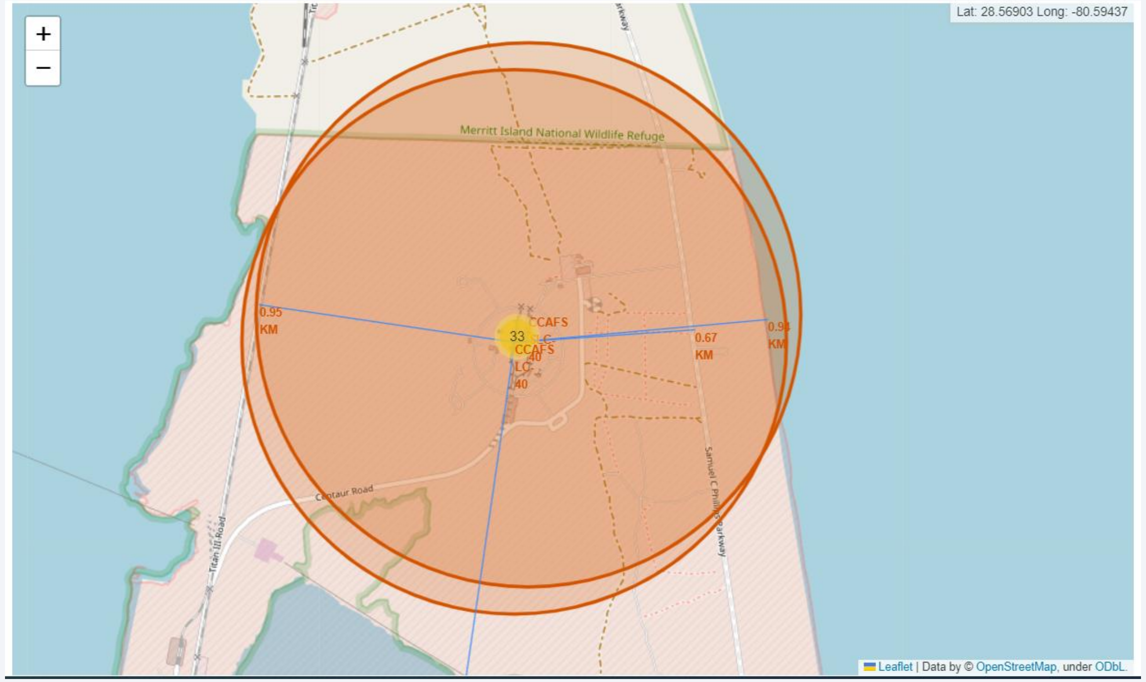
Launch Site Success Breakdown

- Most launches are from the launch sites on the east coast
- The CCAF launch sites are responsible for more failed recoveries than other sites
- The KSC LC-39A launch site has only had 3 failed recoveries across its lifetime



Distances from Important Landmarks

- Railways and highways useful for delivering large rocket parts are less than 1 km from the launch site
- The coast line is also within 1 km of the launch site to allow for safe crashes
- The nearest city, Cape Canaveral is 18 km from the launch site
- This supports the conclusion that launch sites need to be easy to reach and close to the ocean, but farther from cities for safety reasons



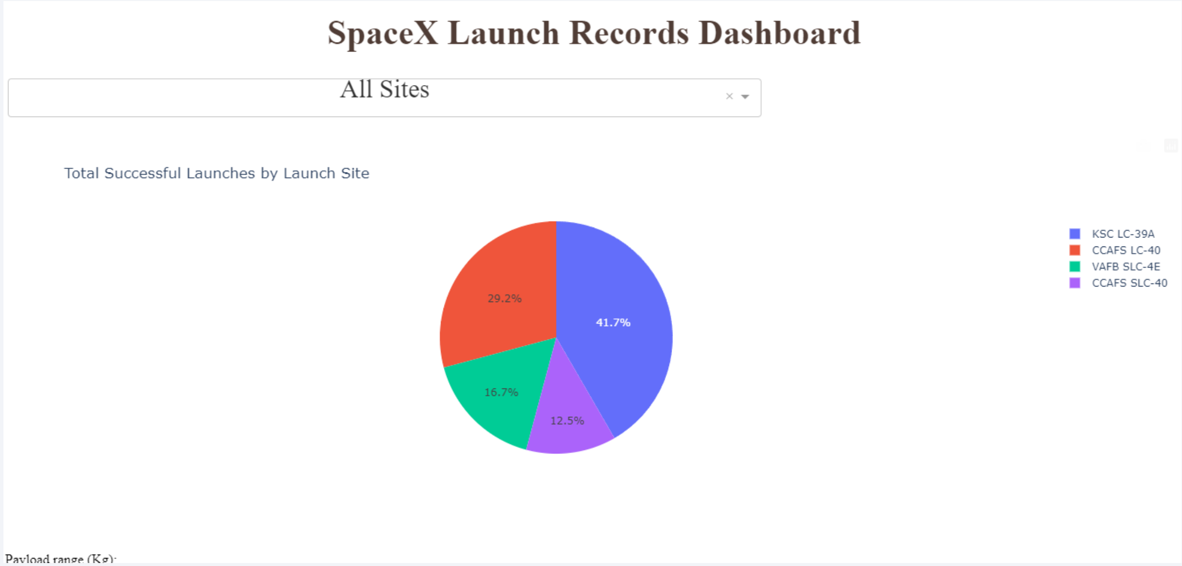
Section 4

Build a Dashboard with Plotly Dash

Launch Successes by Launch Site

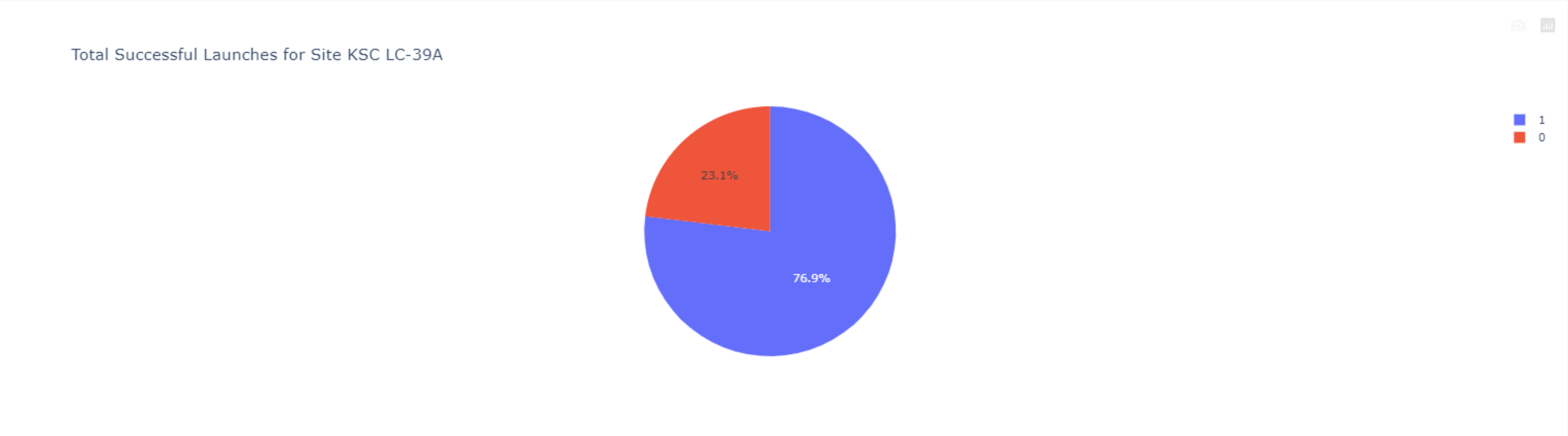
- Most successful recoveries were from the KSC LC-39A launch site

- The CCAFS LC-40 launch site is responsible for many of the launches, but contributes few successful recoveries to the total



Most Successful Launch Site – KSC LC-39A

- KSC LC-39A is the most successful launch site in the dataset
- All other launch sites show <50% success rate for recovery



Recovery Success by Payload Mass

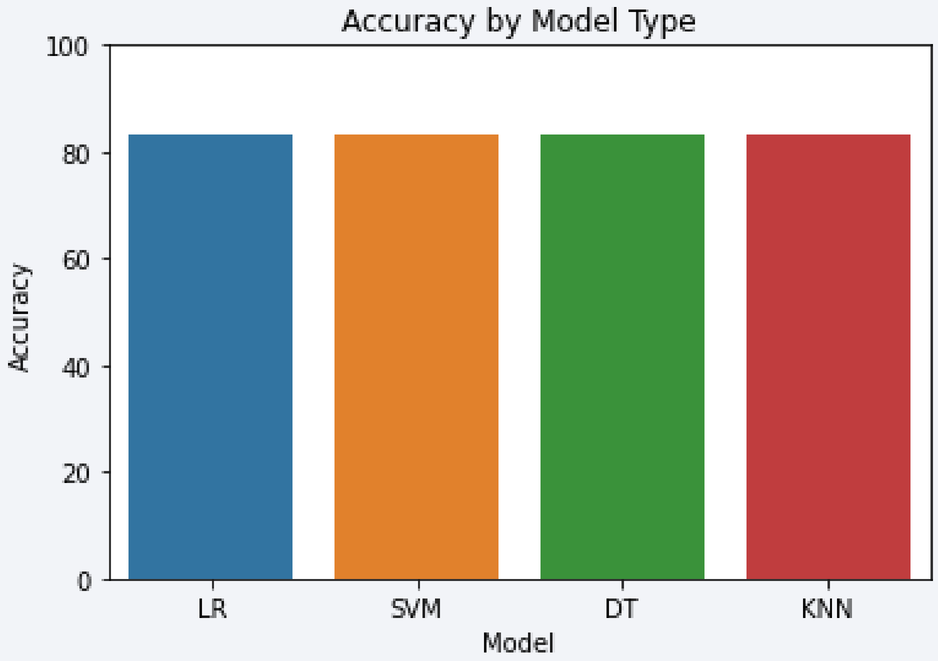
- The mass of the majority of payloads was under 7000 kg
- The smallest and largest payloads failed more often than mid-sized paylods
- The FT version booster has a high recovery success rate

Section 5

Predictive Analysis (Classification)

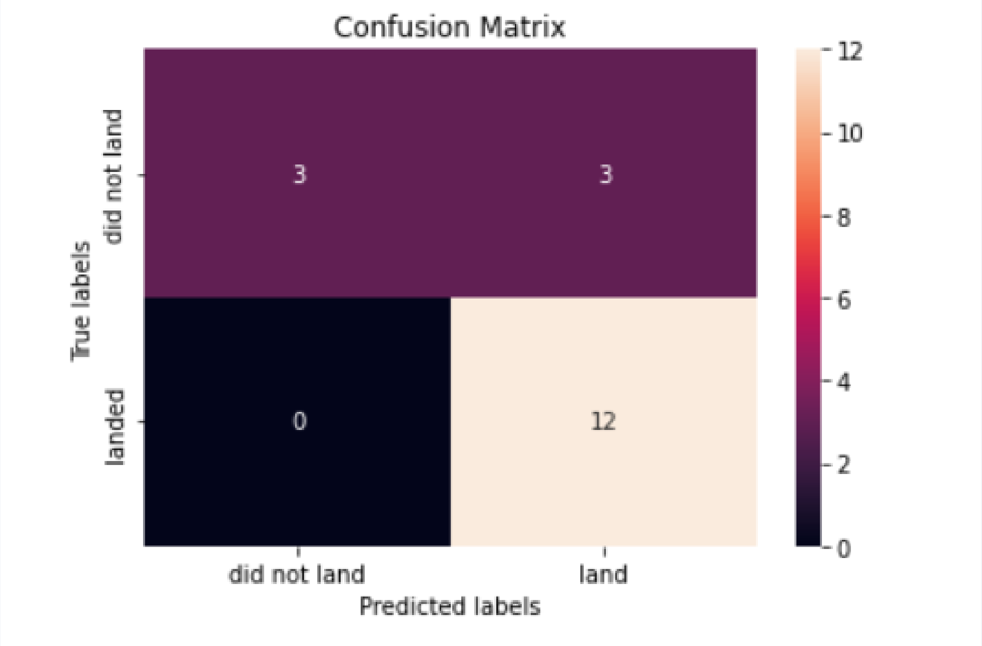
Classification Accuracy

- All models have the same measured accuracy in our testing set
- Within our training set, each model has roughly an 85% accuracy
- The logistic regression model is most effective as it fits quickly and provides probability estimates



Confusion Matrix

- There are no false negatives reported in the testing set for this model
- Roughly 50% of the negative recovery outcomes are mislabeled
- This model is most accurate when predicting that a launch will fail



Conclusions

- Per the logistic regression model coefficients, stage 1 legs and fins are both effective at increasing recovery rate (appendix A)
- Most launches from KSC LC-39A include a successful stage 1 recovery
- Missions targeted at GTO, ISS, LEO, MEO, or PO orbits were less likely to be recovered.
- While we can fairly accurately predict if a recovery will succeed or fail, our model is most accurate at determining that a recovery will fail
- Launch sites are placed near coastlines for safe failure as even stage 1 landings we predict to succeed are prone to failure

Appendix A – Logistic Regression Coefficients

- As all features are scaled to the same range, coefficient magnitudes can be directly compared
- The presence of fins or legs provide the largest improvement to recovery success
- The improvement of rockets over time is shown by the flight number feature

	0	1
0	FlightNumber	4.758620e-02
1	PayloadMass	2.361485e-02
2	Flights	-5.835493e-04
3	Block	5.327613e-02
4	ReusedCount	7.829705e-02
5	Orbit_ES-L1	2.381759e-02
6	Orbit_GEO	1.710794e-02
7	Orbit_GTO	-4.561903e-02
8	Orbit_HEO	-6.927145e-07
9	Orbit_ISS	-4.281832e-04
10	Orbit_LEO	3.397952e-02
11	Orbit_MEO	-7.026510e-03
12	Orbit_PO	-8.432809e-04
13	Orbit_SO	-4.191237e-02
14	Orbit_SSO	2.392699e-02
15	Orbit_VLEO	2.241229e-02
16	LaunchSite_CCAFS SLC 40	-2.430906e-02
17	LaunchSite_KSC LC 39A	1.541051e-02
18	LaunchSite_VAFB SLC 4E	1.487126e-02
19	LandingPad_5e9e3032383ecb267a34e7c7	5.314832e-02
20	LandingPad_5e9e3032383ecb554034e7c9	2.408398e-02
21	LandingPad_5e9e3032383ecb6bb234e7ca	5.131611e-02
22	LandingPad_5e9e3032383ecb761634e7cb	-9.851967e-07
23	LandingPad_5e9e3033383ecbb9e534e7cc	2.689504e-02
24	Serial_B0003	-6.927145e-07
25	Serial_B0005	-3.219401e-02
26	Serial_B0007	-3.000595e-02
27	Serial_B1003	-6.927145e-07
28	Serial_B1004	-2.860708e-02
29	Serial_B1005	-2.865109e-02
30	Serial_B1006	3.674099e-02
31	Serial_B1007	3.444603e-02
32	Serial_B1008	-2.884689e-02
33	Serial_B1010	-3.039290e-02
34	Serial_B1011	-2.886868e-02
35	Serial_B1012	-6.927145e-07
36	Serial_B1013	2.381759e-02
37	Serial_B1015	-6.927145e-07
38	Serial_B1016	-6.927145e-07
39	Serial_B1017	-5.481515e-02
40	Serial_B1018	-5.370292e-02
41	Serial_B1019	2.282856e-02
42	Serial_B1020	-5.218510e-02
43	Serial_B1021	1.849808e-02
44	Serial_B1022	2.620833e-02
45	Serial_B1023	-6.927145e-07
46	Serial_B1025	2.337087e-02
47	Serial_B1026	-6.927145e-07
48	Serial_B1028	-5.371085e-02
49	Serial_B1029	1.696712e-02
50	Serial_B1030	-3.218980e-02
51	Serial_B1031	3.078819e-02
52	Serial_B1032	3.133325e-02
53	Serial_B1034	-3.232945e-02
54	Serial_B1035	3.020117e-02
55	Serial_B1036	3.758108e-02
56	Serial_B1037	-3.252182e-02
57	Serial_B1038	-6.927145e-07
58	Serial_B1039	-2.671768e-02
59	Serial_B1040	-1.192253e-02
60	Serial_B1041	-2.787127e-02
61	Serial_B1042	2.333102e-02
62	Serial_B1043	1.973540e-02
63	Serial_B1044	-5.313058e-02
64	Serial_B1045	-2.793050e-02
65	Serial_B1046	1.129356e-02
66	Serial_B1047	2.273132e-03
67	Serial_B1048	-8.914255e-03
68	Serial_B1049	3.629896e-02
69	Serial_B1050	-5.864533e-02
70	Serial_B1051	1.511750e-02
71	Serial_B1054	-3.391297e-02
72	Serial_B1056	3.063910e-02
73	Serial_B1058	3.049767e-02
74	Serial_B1059	2.881416e-02
75	Serial_B1060	2.074052e-02
76	Serial_B1062	2.188091e-02
77	GridFins_False	-1.091954e-01
78	GridFins_True	1.091954e-01
79	Reused_False	-1.981756e-02
80	Reused_True	1.981756e-02
81	Legs_False	-1.213370e-01
82	Legs_True	1.213370e-01

Thank you!