IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Mark Venuto
10/10/21

Outline

Executive Summary

- Data collection methodology:
  - Collect rocket launch data from SpaceX API
- Perform data wrangling
  - Modify and add columns, useful for training models
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Break dataset into training and test data
  - Develop various models (KNN, Decision Tree, etc…) using train data
  - Assess accuracy scores of models and their best parameters using test data

Executive Summary

- SpaceX’s successful recoveries generally have the following properties:
  - A launch date in the year 2017 or later
  - Light payload (in the range 2000-4000kg)
  - Launched from site KSC LC-39A
  - Successfully recovered via drone ship

- Our machine learning model can predict the outcome of a given recovery with a reasonable degree of accuracy, 83.33%

Introduction

- Most pace exploration companies may spend up to $165 million per rocket launch

- With their flagship Falcon 9 rocket, SpaceX has cut down this cost to only $62 million per launch

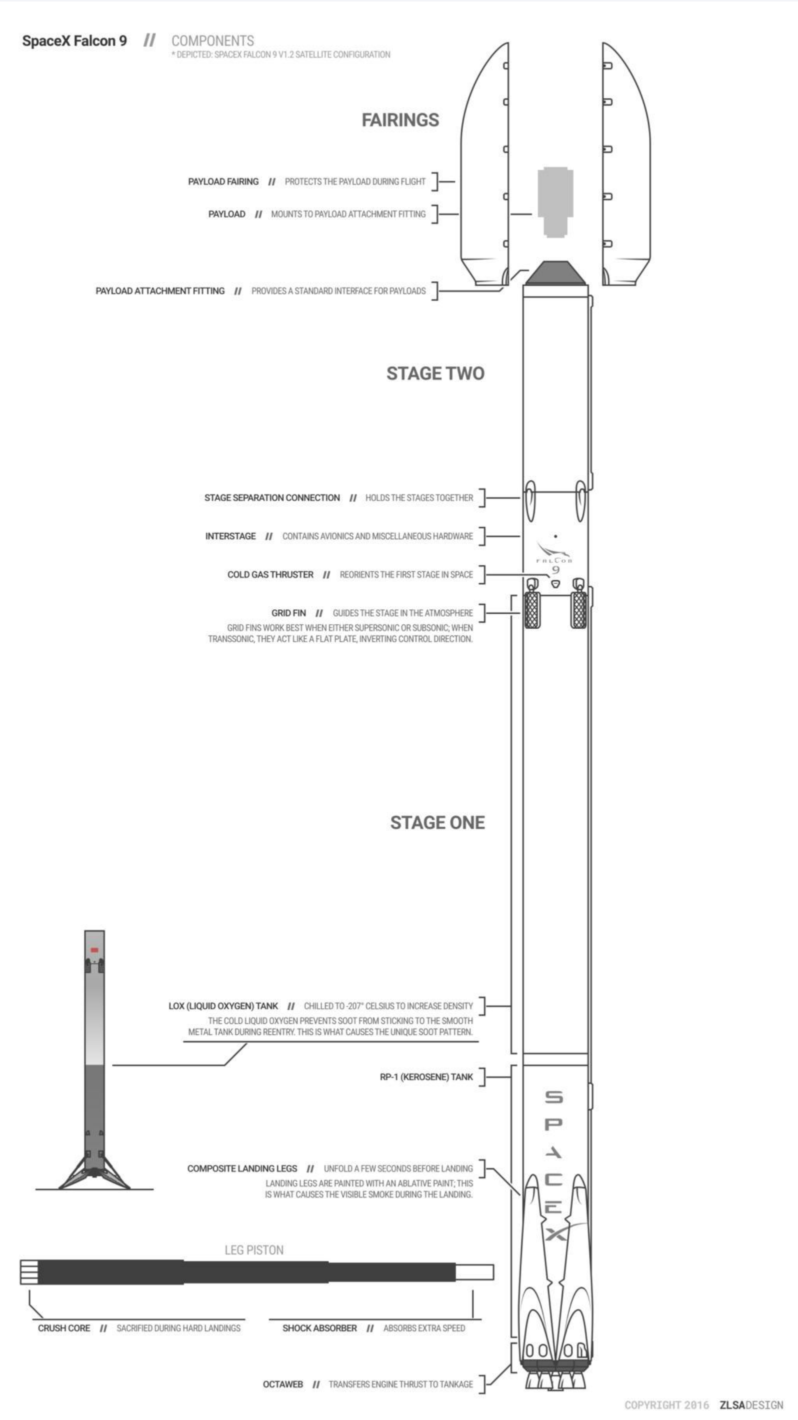
- How?

Introduction (cont.)

- It’s not rocket science!

- With the Falcon 9, SpaceX is able to recover and reuse Stage One, a large and expensive component of the rocket

- Recovery of Stage One means **drastically** reduced manufacturing costs



(Created by - Forest Katsch, zlsadesign.com)

Introduction (cont.)

- Of course, a successful Stage One recovery is not guaranteed...

Introduction (cont.)

- The outcome of a landing will have a significant effect on the overall cost of a launch
- So, we will be training machine learning models to predict whether or not a SpaceX Stage One recovery will be successful
- With this knowledge, we will be able to copy the aspects of successful recoveries with our own SpaceY rockets and ensure that we will be able to save millions

Section 1

Methodology

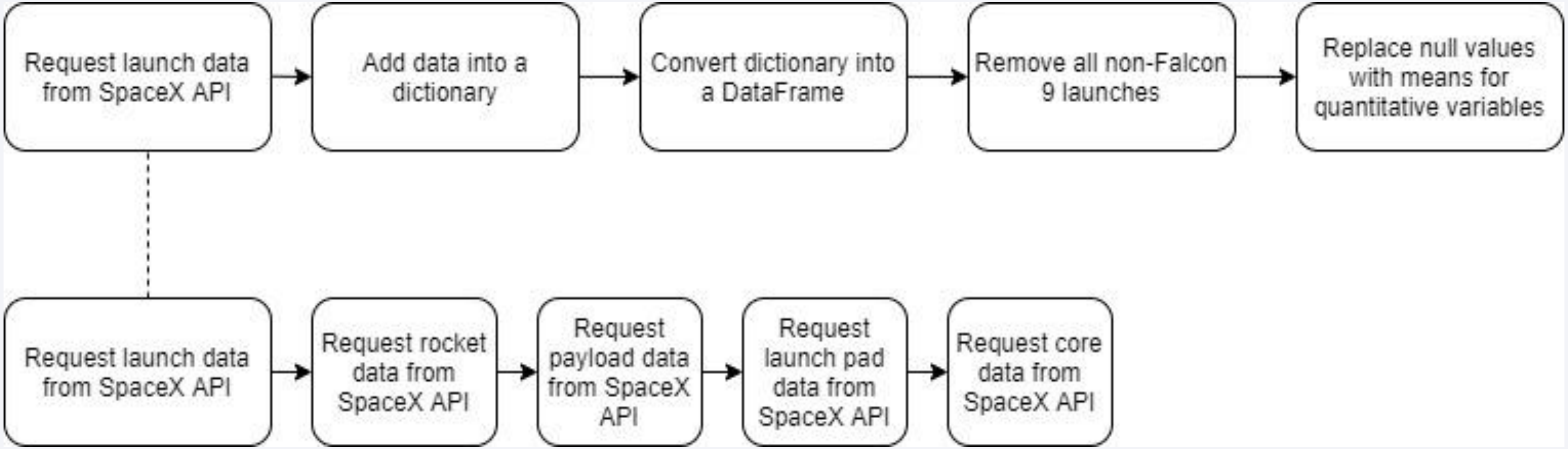
Methodology

- Data collection methodology:
  - Collect rocket launch data from SpaceX API
- Perform data wrangling
  - Modify and add columns, useful for training models
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Break dataset into training and test data
  - Develop various models (KNN, Decision Tree, etc…) using train data
  - Assess accuracy scores of models and their best parameters using test data

Data Collection

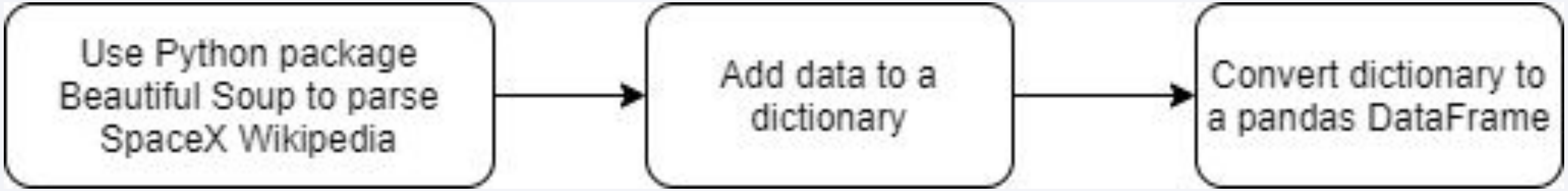
- To collect our data, we need to:
  - Download a .json file containing rocket launch data from the SpaceX API
  - Parse the information into a dataframe
  - Convert null values of quantitative variables into the mean of the rest of the column

Data Collection – SpaceX API



- [Link to Github repo]()

Data Collection - Scraping



- [Link to Github repo]

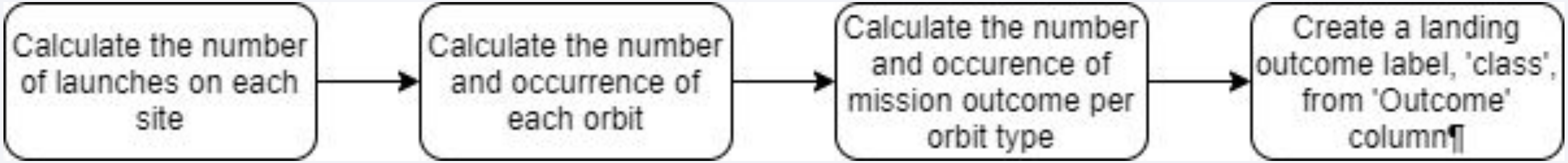
Data Wrangling

- A key point of this project is to determine whether a recovery was successful or not
- So how do we distinguish between ‘successful’ or ‘unsuccessful’?
- According to our dataframe, there are 8 different outcomes, each of which can be considered a Success or Failure:
  - True ASDS → Successful landing to drone ship
  - True RTLS → Successful landing on a ground pad
  - True Ocean → Successful landing in ocean
  - None None → Failed to land
  - None ASDS → Failed to land
  - False ASDS → Failed to land on drone ship
  - False RTLS → Failed to land on ground pad
  - False Ocean → Failed to land in ocean

Data Wrangling

- From here, we could create a new column, ‘class’, to delineate between successful and unsuccessful recoveries
- 1 → successful recovery
  0 → unsuccessful recovery
- From here, we are ready to do some more advanced EDA

Data Wrangling



- [Link to Github repo]

EDA with Data Visualization

- Here, a number of plots showing relationships between different variables:
    - Flight Number vs. Payload (Cat plot)
    - Flight Number vs. Launch Site (Cat plot)
    - Launch Site vs. Payload (Scatter plot)
    - Success Rate vs. Orbit type (Bar plot)
    - Orbit type vs. Flight Number (Scatter plot)
    - Orbit type vs. Payload (Scatter plot)
    - Success rate vs. Time in years (Line plot)
- An explanation of these various plots are shown in Section 2
- We also perform some one-hot encoding in this section
- Link to Github repo

EDA with SQL

- Using SQL allows us to make complicated queries without much difficulty.

- Here, we make some basic queries on our data to get a better sense for the relationships between variables, particularly the following:
    - Launch Site
    - Payload Mass (kg)
    - Mission Outcome
    - Booster Version
    - Date

- Link to Github repo

Build an Interactive Map with Folium

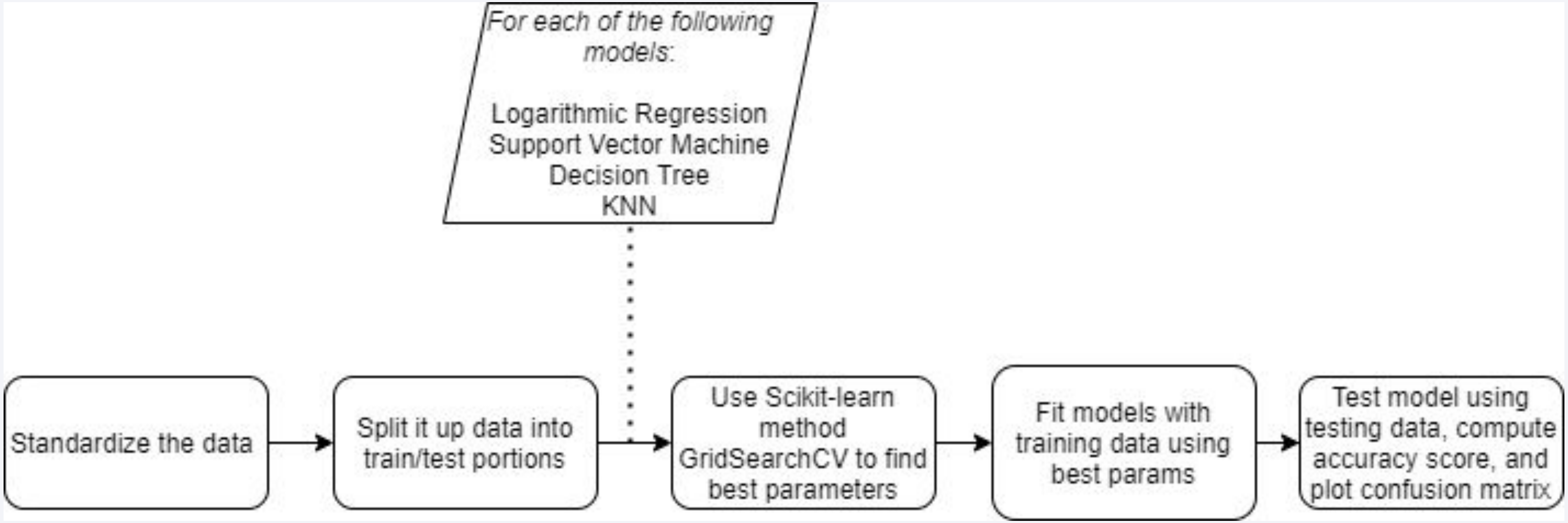
- Using the Python package Folium, we created an interactive map where one can:
    - View where each Falcon 9 launch site is located, represented by a circle
    - Learn how many launches occurred at each location, represented by markers. Green markers represent a successful recovery while red markers represent unsuccessful one
    - Determine distances to the closest coastline, city, railway, and highway, each represented by a blue line.

- Link to Github repo
- Note: Github does not support Folium maps natively

Build a Dashboard with Plotly Dash

- Through Plotly Dash, we made a dashboard that shows:
    - A **pie chart** showing the proportion of successful recoveries to unsuccessful ones for each site (which can be changed via a dropdown menu)
    - A Recovery Outcome vs. Payload Mass **scatter plot** with a range (0-10000kg) with bounds that can be changed by the user
- This dashboard provides insight into the launch sites' and payload masses' relationships with the recovery outcomes.
- Link to Github repo

Predictive Analysis (Classification)



- [Link to Github repo]

Results

EDA Results:

- SpaceX has gotten better at launching rockets overtime. Launches are most successful when launched in 2017 or later
- Light payloads are evidently easier to recover, as most successful recoveries occur when the payload has a mass between 2000kg and 4000kg)
- Each launch site is reasonably successful in their own right, but site KSC LC-39A appears to be ideal as it has a success rate of over 75%
- The best recovery method appears to be via drone ships. This would make sense as SpaceX has much more control over the recovery maneuvering a ship in the ocean as opposed to a stationary landing pad on land.

Results (cont.)

Interactive Demo

- The demo .py file can be downloaded from page 20, or alternatively refer to Section 5 for screenshots

Predictive analysis

- Each model performed about equally, correctly predicting a recovery outcome at a rate of 83.33%

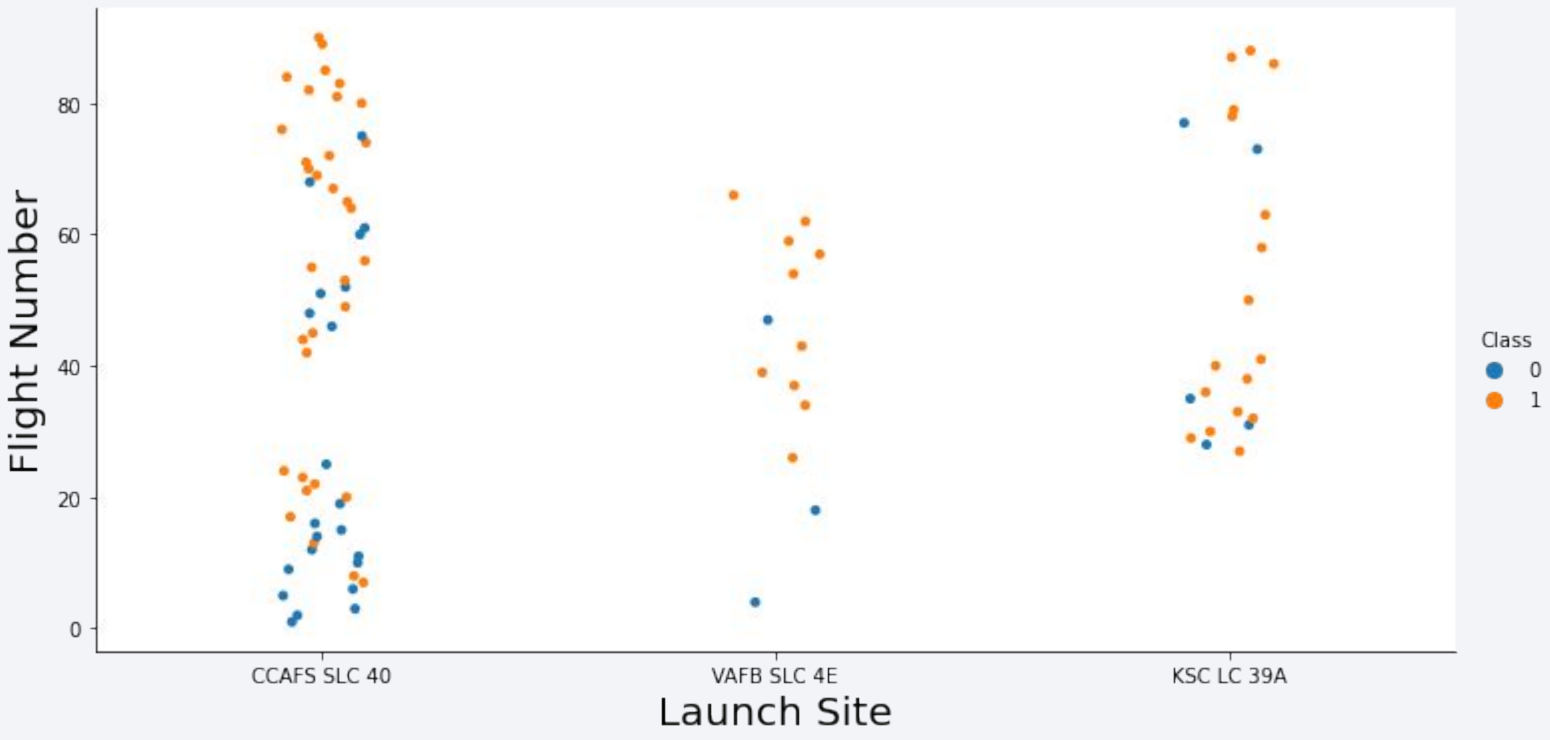
Section 2

Insights drawn from EDA

EDA with Data Visualization

- Rate of success has grown over time at each site, though KSC LC-39A seems to be the most consistent

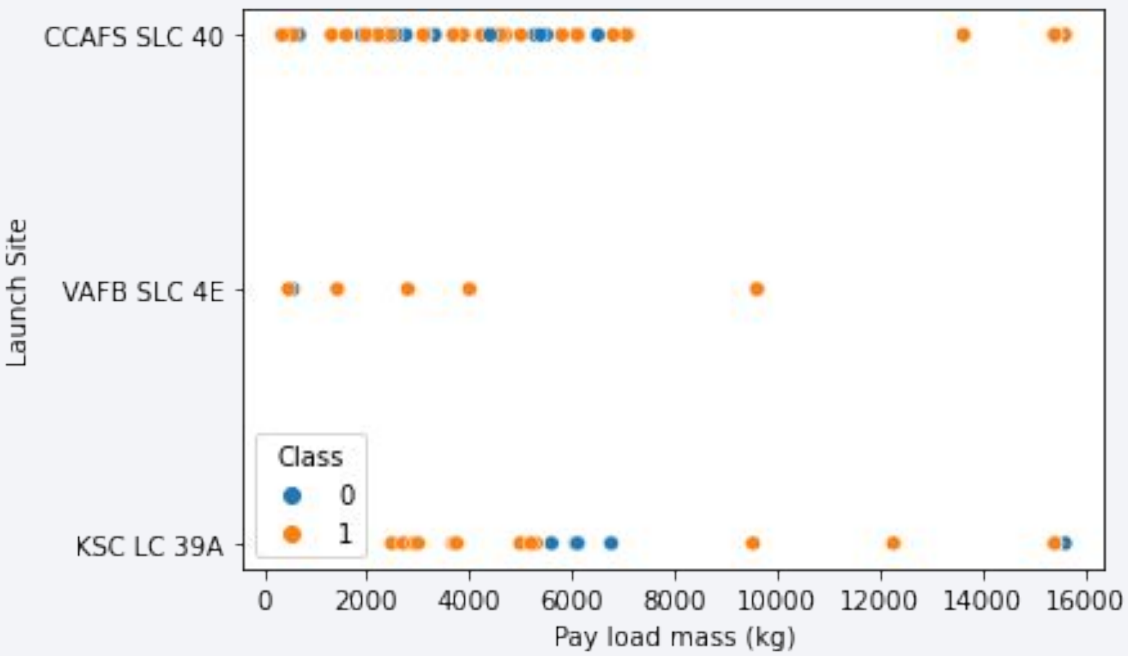
Flight Number vs Launch Site (Cat Plot)



EDA with Data Visualization

- Smaller payloads (<6000 kgs) seem to correlate with higher success rate

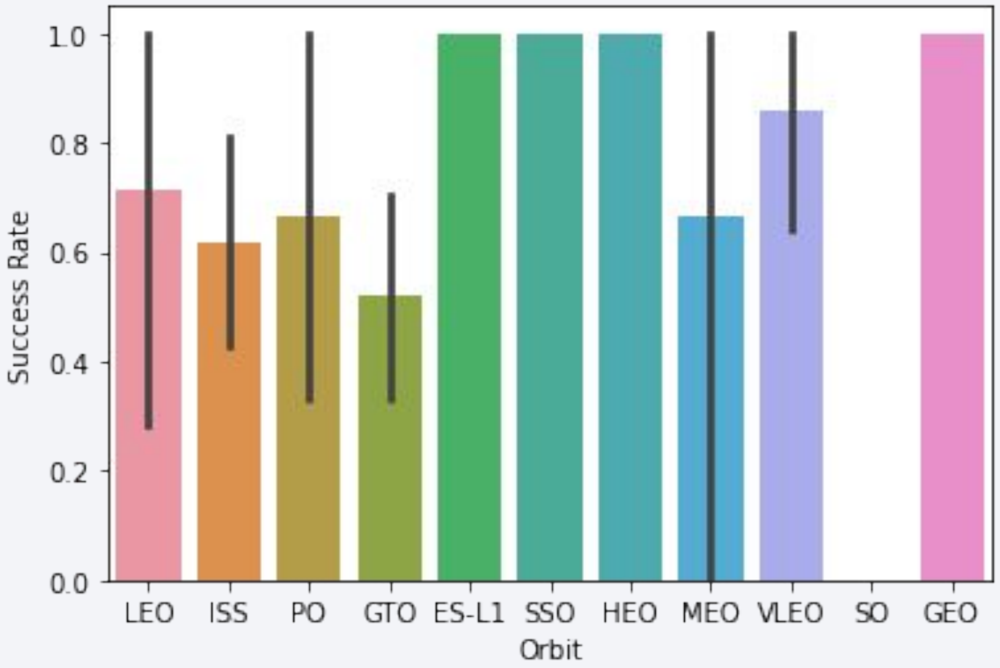
Launch Site vs Payload Mass (Scatter Plot)



EDA with Data Visualization

- ES-L1, SSO, HEO, and GEO orbits are very reliable

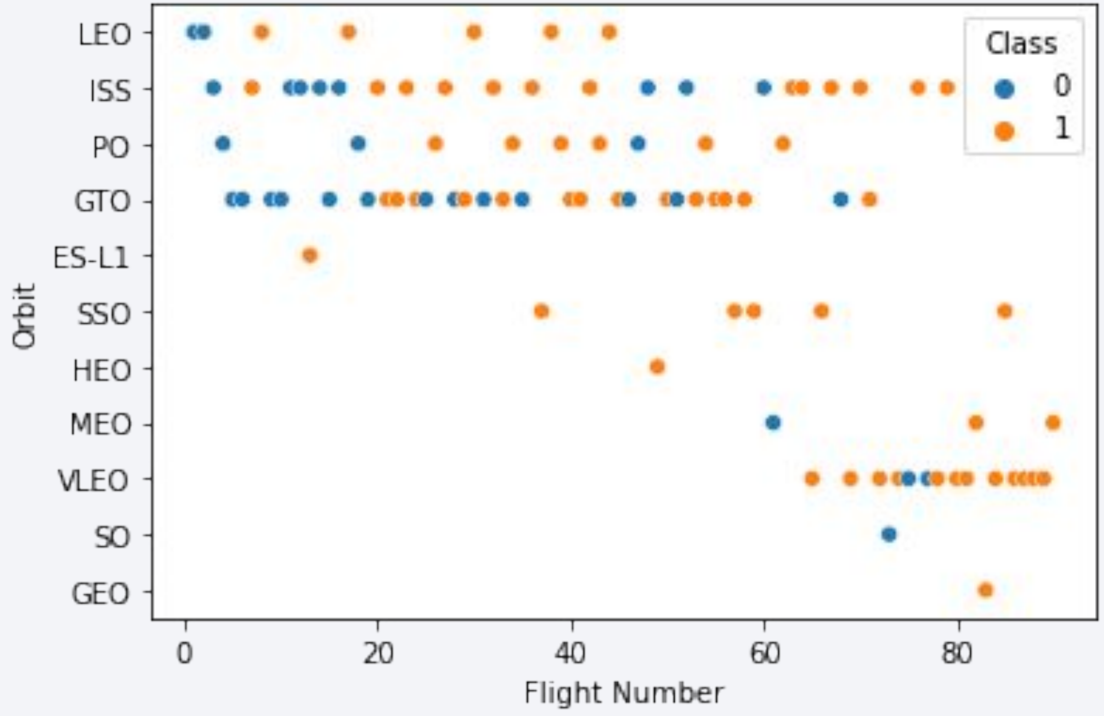
Success Rate vs. Orbit Type (Bar plot)



Flight Number vs. Orbit Type

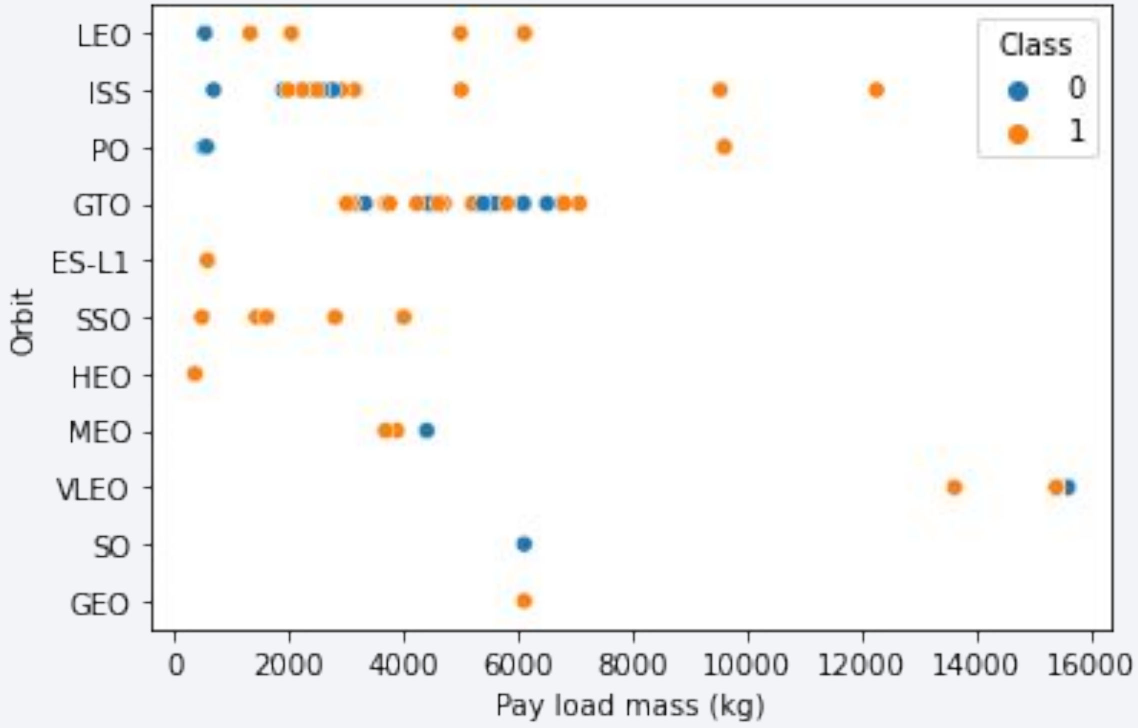
- However, ES-L1, HEO, and GEO success rate may be skewed due to each having only 1 launch

- LEO, SSO, and VLEO seem all have high success rates while having high sample sizes



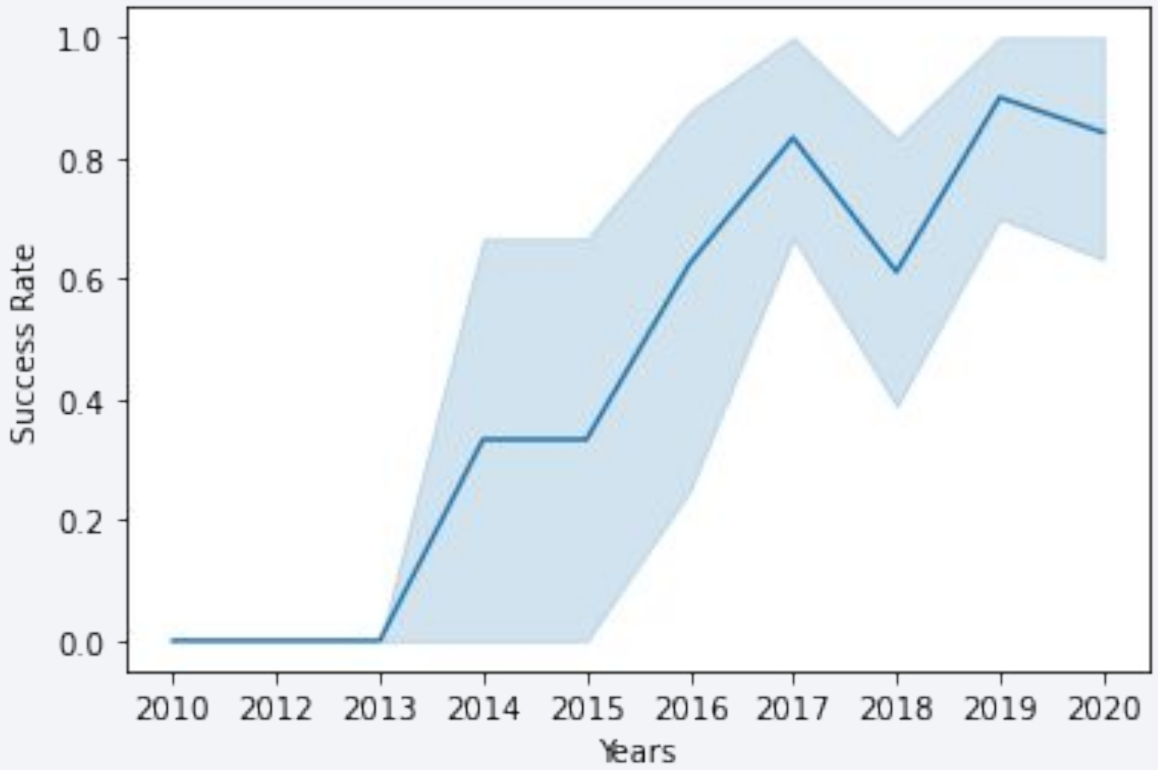
Payload vs. Orbit Type

- LEO and SSO orbits' success may be due to light payloads



Launch Success Yearly Trend

- Overall success rate increased from 2013 to 2017, where it has more or less stagnated



All Launch Site Names

- According to the SQL Query on the right, we have 4 unique launch site locations:

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

Launch Site Names Begin with 'CCA'

- These are the first 5 records for launch sites that begin with ‘CCA’:

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	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	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

Total Payload Mass

- Here is the total payload mass (in kg) carried by boosters from NASA:

1
111268

Average Payload Mass by F9 v1.1

- On average, rockets with by booster version F9 v1.1 carry a mass of:

1
2534

First Successful Ground Landing Date

- The first successful Stage One recovery landing occurred on:

1
2010-06-04

Successful Drone Ship Landing with Payload between 4000 and 6000

- This is a list of the names of boosters which have successfully landed on a drone ship and had payload mass greater than 4000kg but less than 6000kg:

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

Total Number of Successful and Failure Mission Outcomes

- In total, there were 101 missions recorded in this database
- These 101 missions were either a:
  - Success

1
100

  - Failure

1
1

Boosters Carried Maximum Payload

- Here is a list of the names of boosters which have carried the maximum payload mass:

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, there were two launches which resulted in a failed Stage One recovery:

DATE	landing__outcome	booster_version	launch_site
2015-01-10	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
2015-04-14	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

- This is a rank of the types and number of landing outcomes (in descending order) between dates 2010-06-04 and 2017-03-20:

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

Section 4

Launch Sites Proximities Analysis

Map of Launch Site Locations

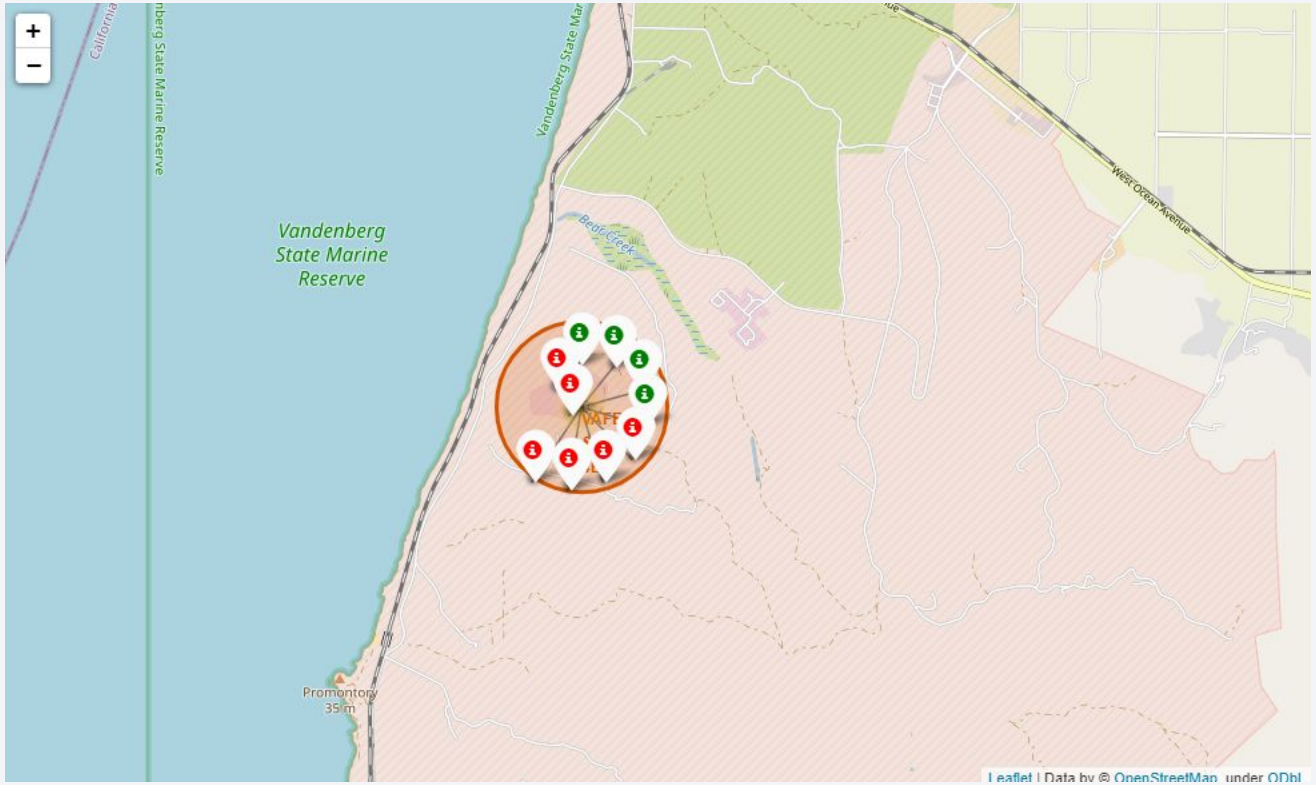
- As can be seen, SpaceX has one launch site on the Pacific coast of Southern California

- The rest of the launch sites are located on the Atlantic coast of Florida

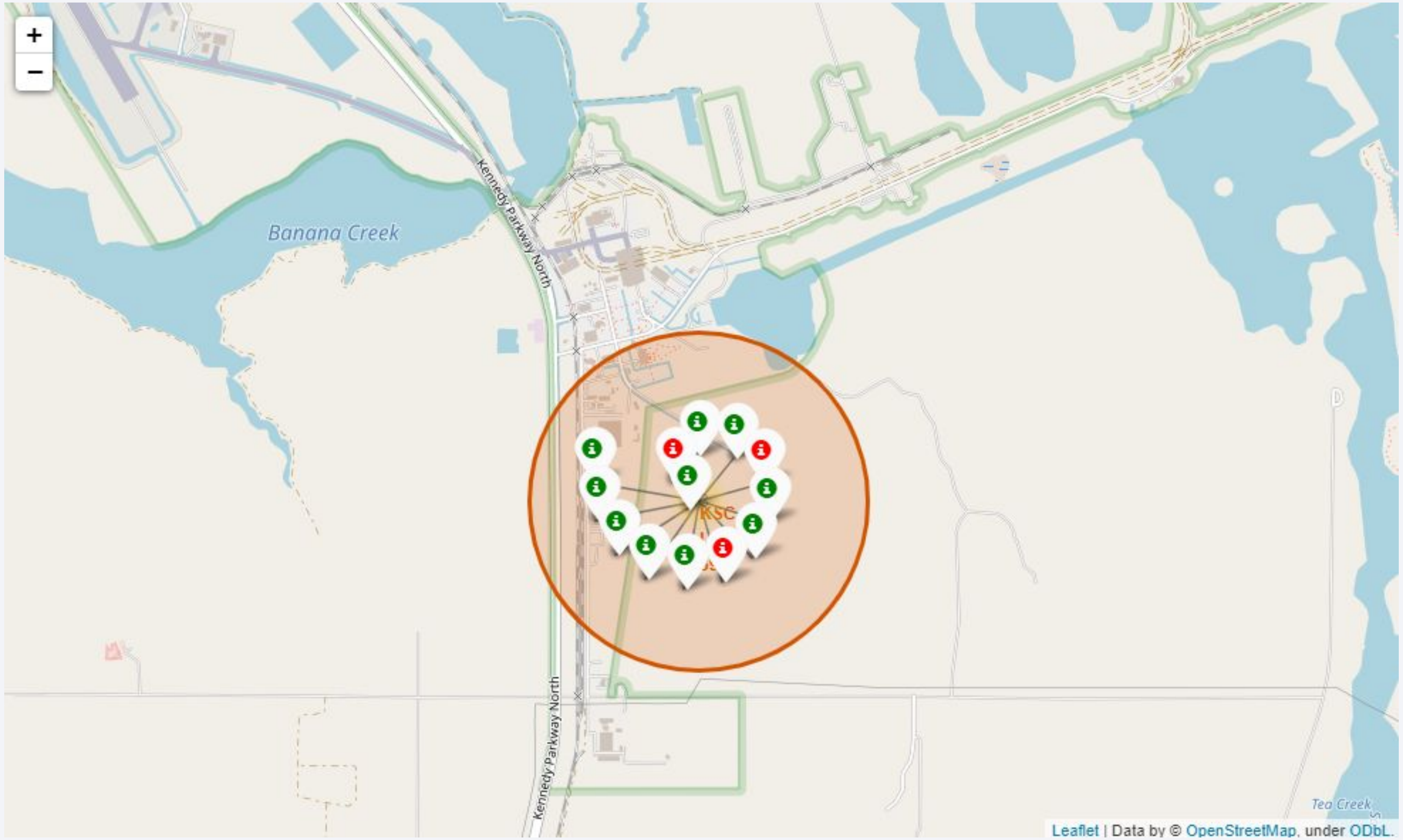
VAFB SLC-4E Recovery Outcomes

- Each marker represents a Falcon 9 Stage One recovery

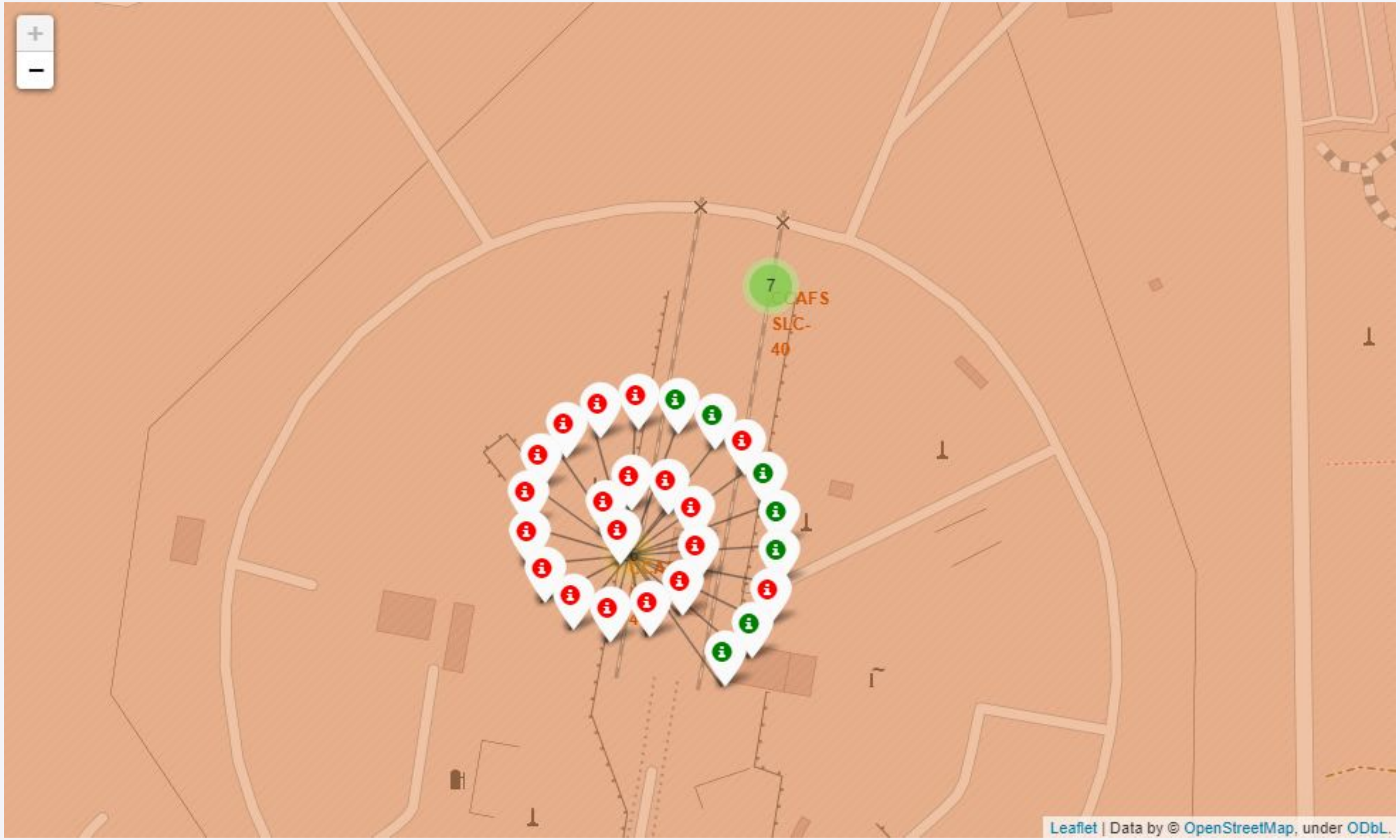
- Green - successful recovery
- Red - unsuccessful recovery



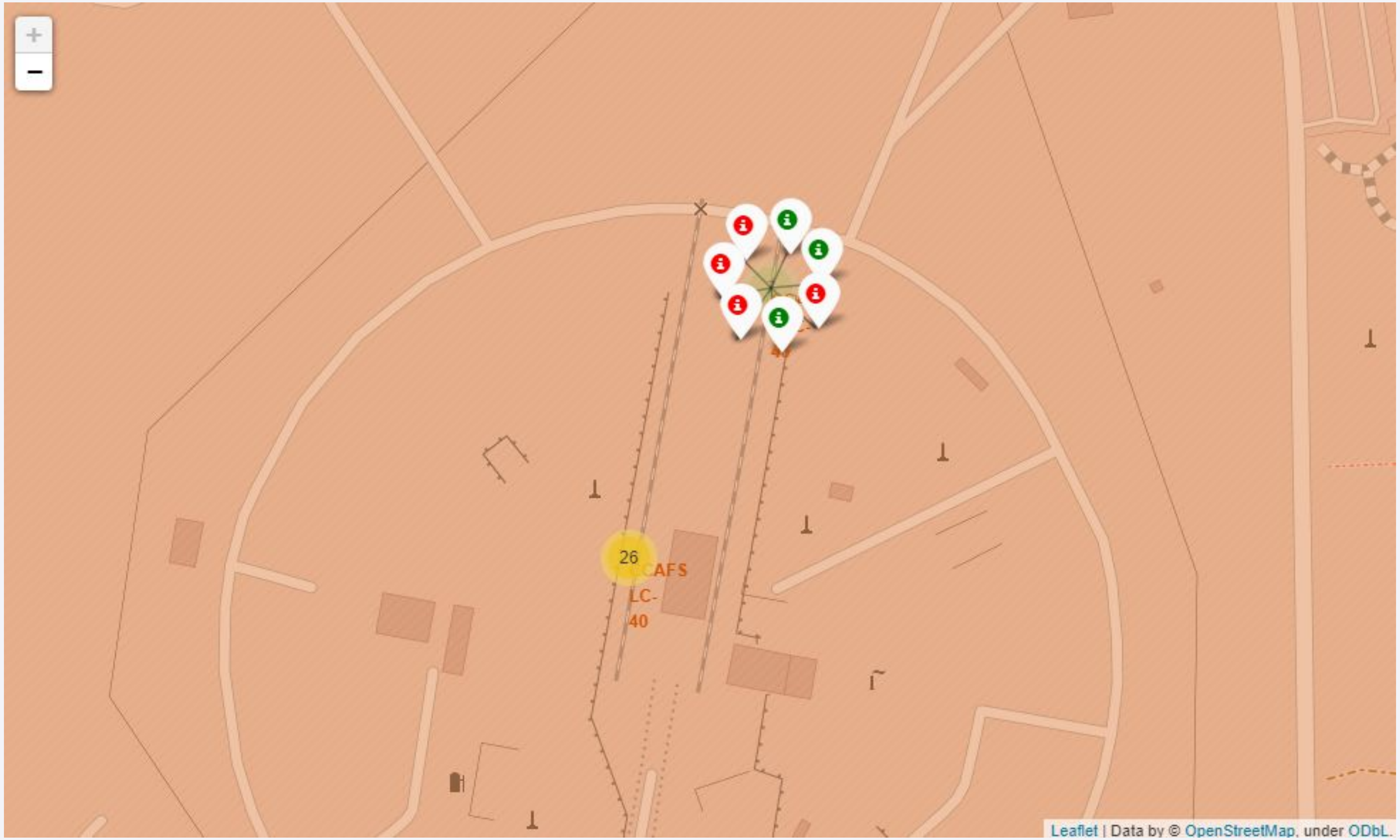
KSC LC-39A Recovery Outcomes



CCAFS LC-40 Recovery Outcomes

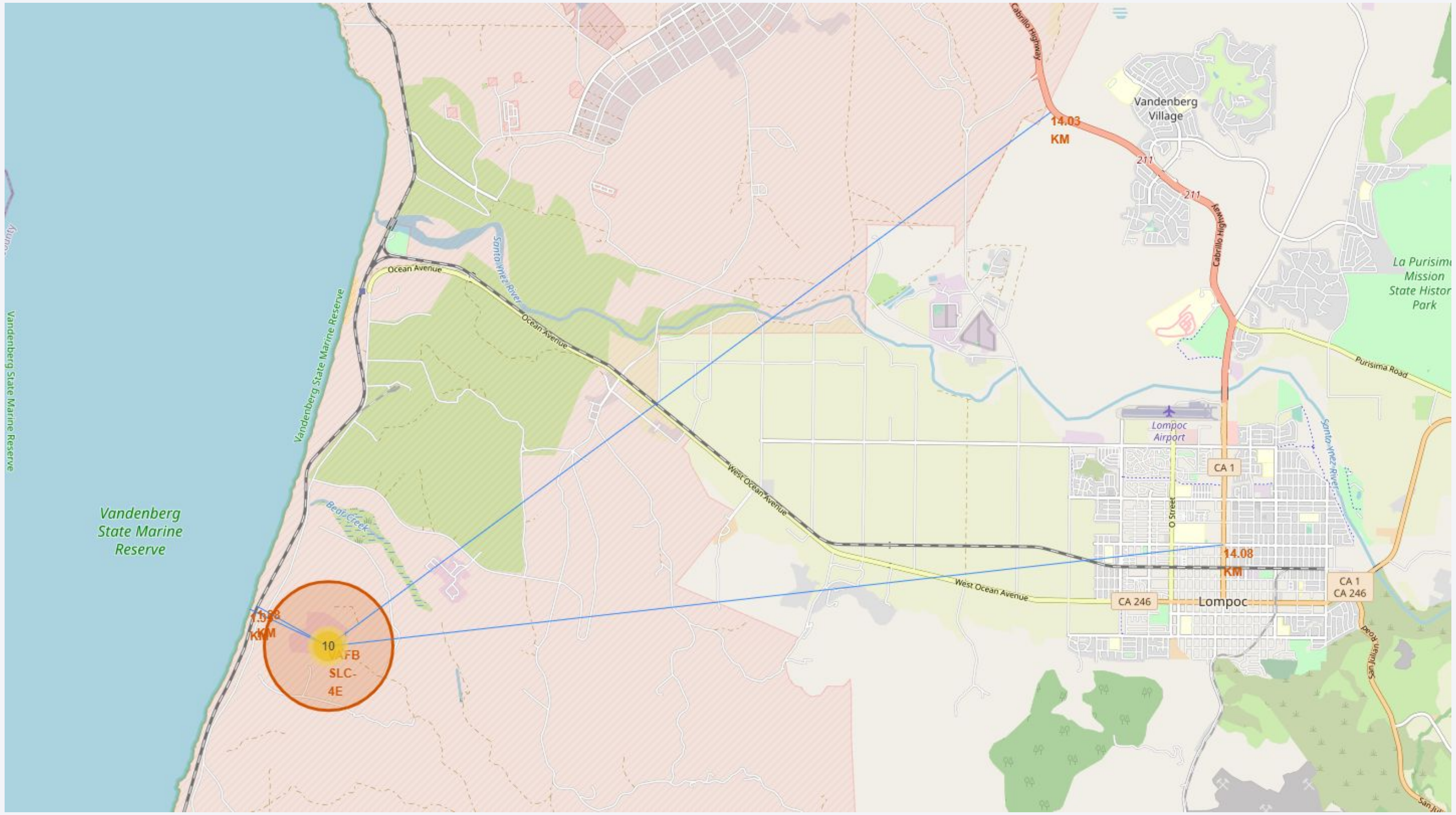


CCAFS SLC-40 Recovery Outcomes



VAFB SLC-4E Nearby Locations

- Each blue line represents the distance to the nearest
  - coastline
  - city/town
  - railway
  - highway

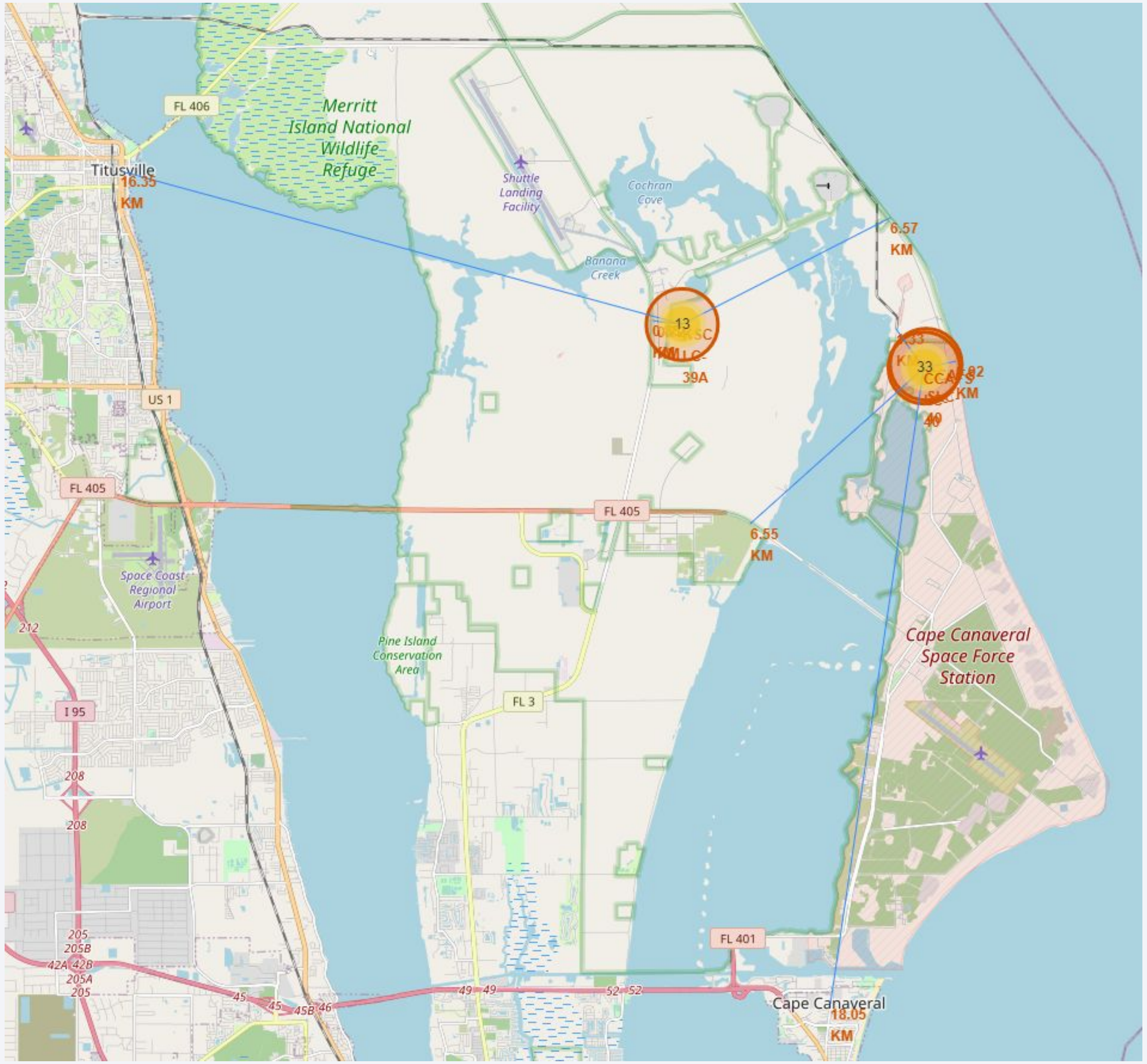


Notes:

- Reasonable distance from closest city and highway (both ~14km)
- Very close to coast and nearest railway (both <2km)

KSC LC-39A, CCAFS LC-40, and CCAFS SLC-40 Nearby Locations

- Reasonable distance to nearest towns (>15km)
- Launch sites located very close to coast and surprisingly close to railways and highways (<7km)



Section 5

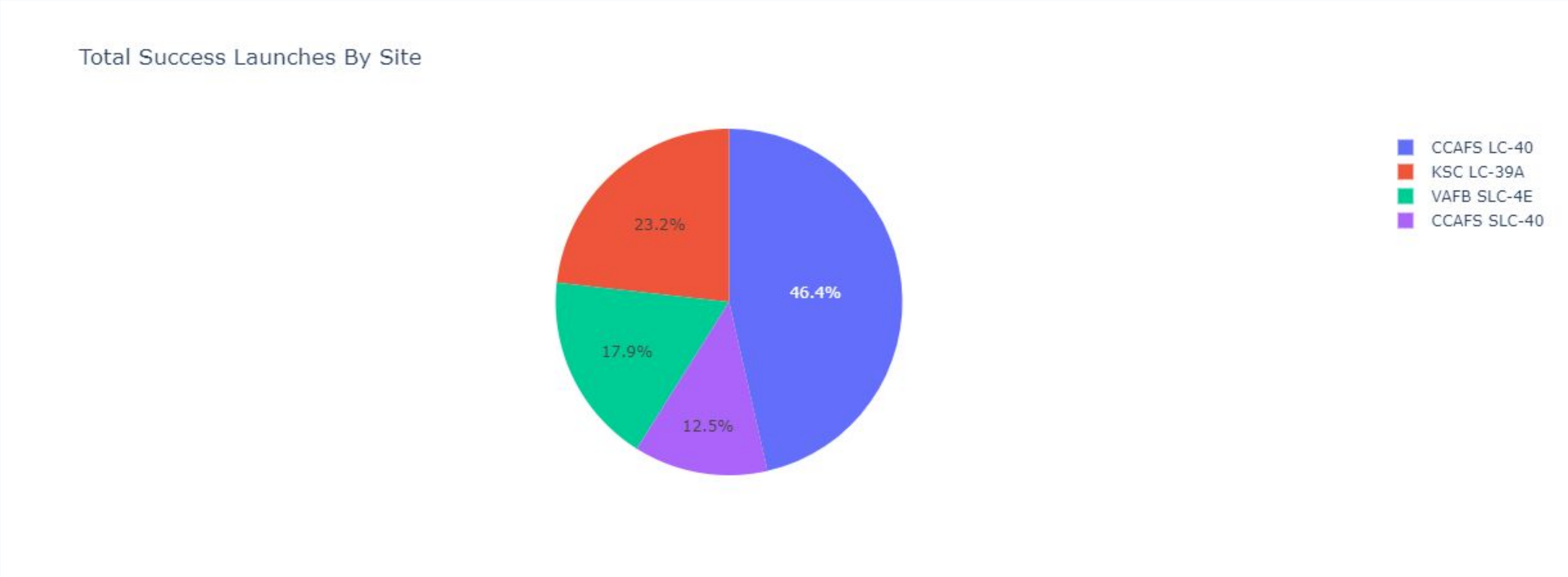
Build a Dashboard with Plotly Dash

Pie Chart of Successful Launches by Site

Here is a breakdown of the total amount of successful launches by site:

- CCAFS LC-40 is the location of over half of SpaceX's successful launches

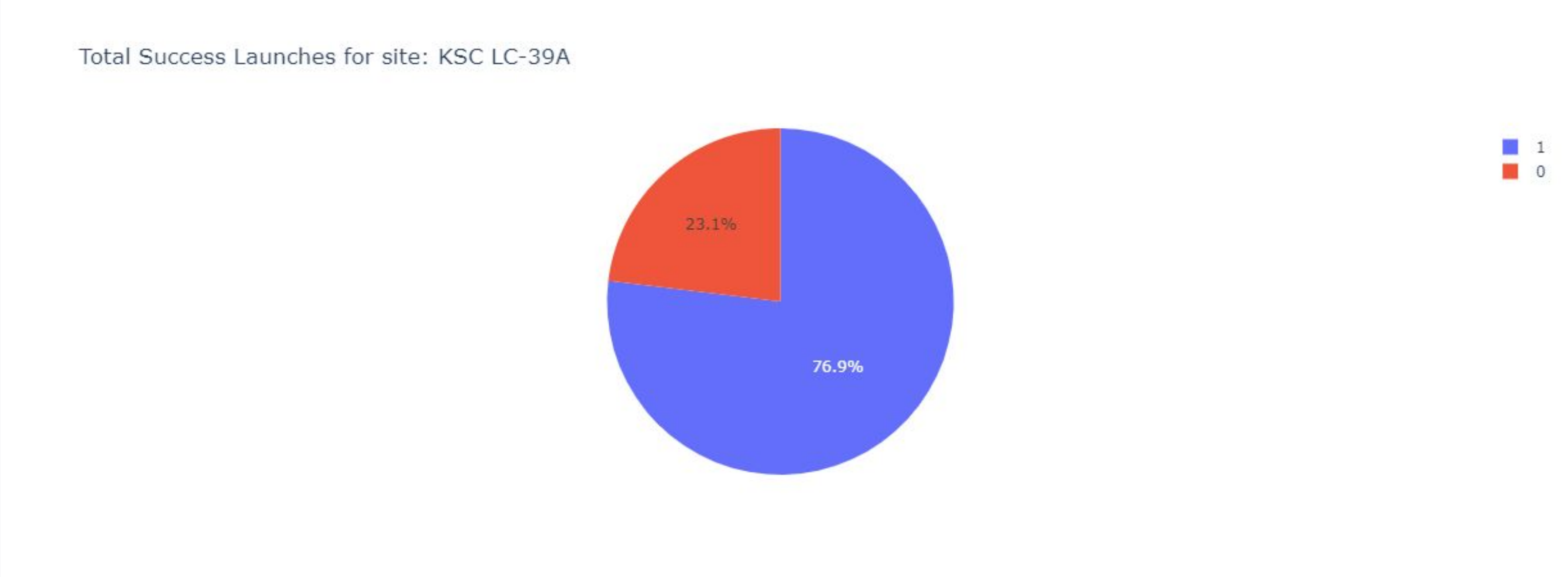
- However, this does not tell the full story...



Most Proportionally Successful Launch Site

- CCAFS LC-40 has the most launches, so naturally it has the most successful launches

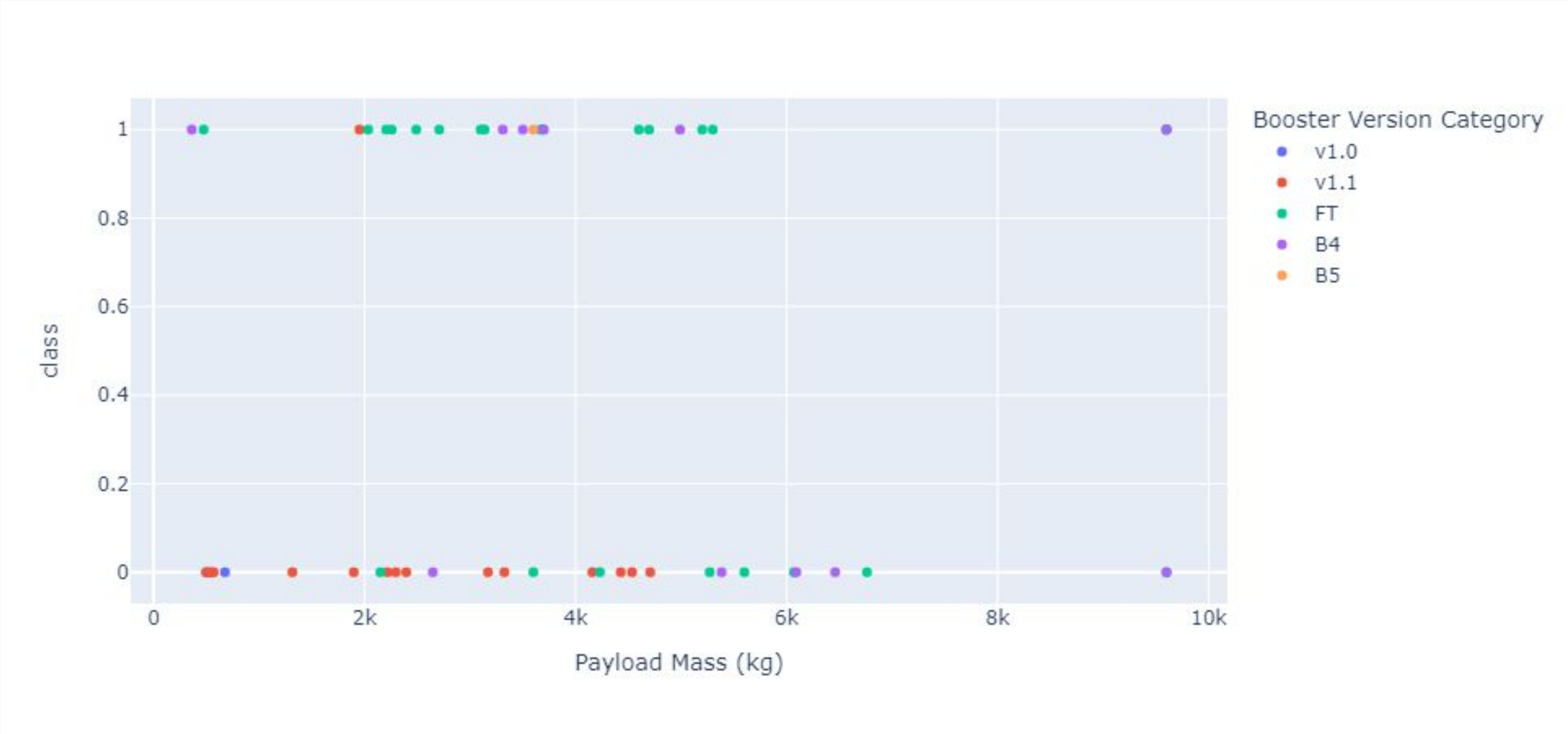
- KSC LC-39A, comparatively, has fewer launches overall but the proportion of success is higher



Recovery Outcome vs. Payload Mass Scatter Plot

- The most successful payload range appears to be from 2000-4000kg

- The v1.1 booster seems to have failed the most while the FT booster appears to be very successful

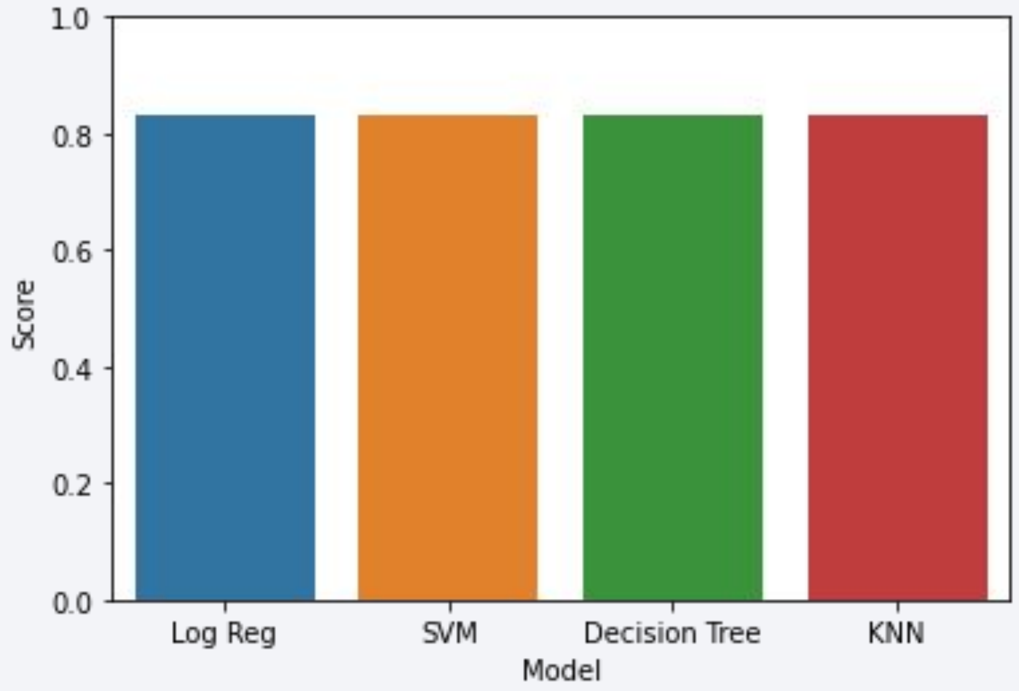


Section 6

Predictive Analysis (Classification)

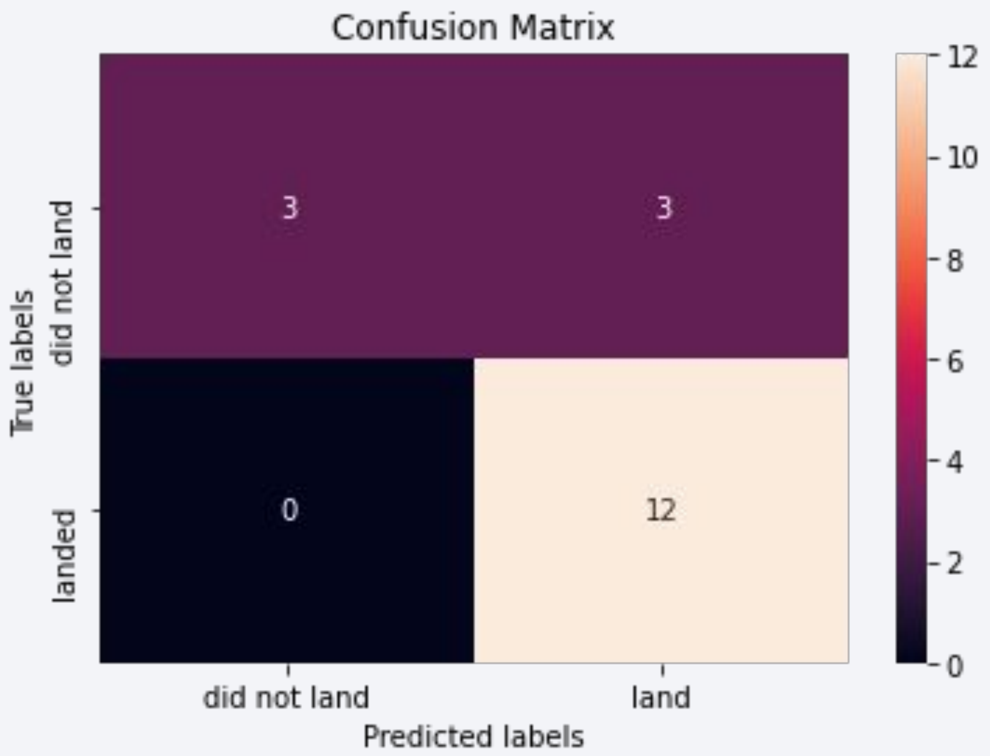
Classification Accuracy

- Here is a barplot for the accuracy score of each model. As can be seen, when testing them on our test data, they all exhibit identical scores (83.33%)



Confusion Matrix

- Our model was mostly accurate, however it wrongly predicted three landings as being successful where they were not



Conclusions

- SpaceX's successful recoveries generally have the following properties:
  - A launch date in the year 2017 or later
  - Light payload (in the range 2000-4000kg)
  - Launched from site KSC LC-39A
  - Successfully recovered via drone ship

- Our model can predict the outcome of a given recovery with a reasonable degree of accuracy, 83.33%

Thank you!