IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Sabine G.
06.10.2024

Outline

Executive Summary

- Background/Aim:
  - Analyzing the factors that contribute to successful landings of Falcon 9 space vehicles and utilizing those factors to train a classifier to predict landing success.

- Summary of methodologies:
  - **Data collection** (from SpaceX API and web-scraped from Wikipedia), **Data Wrangling** to create successful/unsuccessful outcome variable, **Data Visualizations** to explore the role of the factors Year/Flight Number, Payload Mass, Orbit Type and Launch Site; Transfer to **SQL database** to allow search queries; **Interactive Map** creation to explore the proximities of Launch Sites, **Interactive Dashboard** creation to visualize and explore landing success depending on Launch Site, Payload Mass, and Booster Version; **Predictive Analysis** (Machine Learning Classification) to predict landing success using four different types of classifiers

- Summary of results:
  - Several factors seem to influence landing success, such as Year/Flight Number, Payload Mass, Orbit Type, and possibly Launch Site. Yet, some of the predictors might be correlated.
  - Overall, landing success can be predicted using machine learning techniques with an accuracy of around 84 %.

Introduction

- The **Falcon 9** is a **two-stage-to-orbit** launch vehicle, which **can be partially reused** when the first stage of the rocket is successfully landed.
- Thus, **successful landings** immensely **reduce costs** of future missions.
- Analyzing the factors that contribute to successful landings and utilizing them to predict landing success is the core aim of the present report.

Falcon 9 first stage landing in 2015
Source: https://en.wikipedia.org/wiki/SpaceX_reusable_launch_system_development_program#/media/File:ORBCOMM-2_(23282658734).jpg

Section 1

Methodology

Methodology Overview

- Data collection methodology:
  - Using the SpaceX API
  - Web scraping the Wikipedia page *List of Falcon 9 and Falcon Heavy launches*
- Perform data wrangling
  - Binary coding of landing outcomes to serve as training labels
- Exploratory data analysis (EDA) using visualization and SQL
- Interactive visual analytics using Folium and Plotly Dash
- Predictive analysis using classification models
  - Comparing Logistic Regression, Support Vector Machine (SVM), Decision Tree, and K-Nearest Neighbours (KNN) classifiers in their prediction accuracy for the landing outcome

Data Collection – SpaceX API

- Data were requested from the SpaceX API under the following URL: https://api.spacexdata.com/v4/
- For 90 listed Falcon 9 launches between 2010 and 2022, the following variables were retrieved: Flight Number, Date, Booster Version, Payload Mass, Orbit, Launch Site, Longitude, Latitude, Number of Flights, Grid Fins, Reused, Legs, Landing Pad, Block, Reused Count, Serial, Outcome

Request Falcon launches from:
https://api.spacexdata.com/v4/launches/past

Use identification numbers to retrieve information on
rocket: https://api.spacexdata.com/v4/rockets/
launchpad: https://api.spacexdata.com/v4/launchpads/
payload: https://api.spacexdata.com/v4/payloads/
other core data: https://api.spacexdata.com/v4/cores/

Notebook on GitHub:
https://github.com/spongefrog/dsCapstone/blob/2692d1f7d3bdb8fc0abc45568ad5043fef656bf5/DS01_Capstone_DataCollection_SpaceX_API.ipynb

Data Collection – Web Scraping

- Data were scraped from the the *List of Falcon 9 and Falcon Heavy launches* at the URL: https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922
- For 121 listed Falcon 9 launches between 2010 and 2021 (June 6), the following variables were retrieved: Flight Number, Date, Time, Launch site, Payload, Payload mass, Orbit, Customer, Booster version, Booster landing, Launch outcome

HTTP GET: request the Falcon 9 Launch HTML page

↓

BeautifulSoup: parse the retrieved HTML data

↓

Iteration through the <th> elements of all tables

Notebook on GitHub:
https://github.com/spongefrog/dsCapstone/blob/2692d1f7d3bdb8fc0abc45568ad5043fef656bf5/DS02_Capstone_DataCollection_WebScraping.ipynb

Data Wrangling

- Data were retrieved from https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/dataset_part_1.csv:
  - i.e., the dataset imported via SpaceX API
  - filtered for the 90 entries of Falcon 9 launches
  - imputation of missing payload mass entries: substitution by mean

- Landing Outcomes which are originally coded as categories (True ASDS, None None, True RTLS, False ASDS, True Ocean, False Ocean, None ASDS, False RTLS) were binarized (0 = not successful; 1 = successful) to later serve as training labels.

Notebook on GitHub:
https://github.com/spongefrog/dsCapstone/blob/2692d1f7d3bdb8fc0abc45568ad5043fef656bf5/DS03_Capstone_DataWrangling.ipynb

EDA with Data Visualization

- Data were retrieved from https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/dataset_part_2.csv:
  - i.e., the dataset imported via SpaceX API
  - filtered for the 90 entries of Falcon 9 launches
  - imputation of missing payload mass entries: substitution by mean
  - coding of landing outcomes as 0 (not successful) or 1 (successful)
- The libraries Matplotlib and Seaborn were used to create bar plots (e.g., success rate per Orbit type), line plots (e.g., development of success rate over the years), and scatter plots (e.g., success rate for different orbit types depending on flight number) (for more details, see Jupyter Notebook)

Notebook on GitHub:
https://github.com/spongefrog/dsCapstone/blob/2692d1f7d3bdb8fc0abc45568ad5043fef656bf5/DS04_Capstone_DataVisualization.ipynb

EDA with SQL

- Data for the SQL exercise were retrieved from https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/labs/module_2/data/Spacex.csv:
  - other than in the previous sections, this dataset contains 101 entries of Falcon 9 launches
- The library SQLite (sqlite3) was used to allow for accessing and manipulating databases using a variant of the SQL query language.
- The retrieved dataframe was converted to a table (SPACEXTBL) of a SQL database (my_data1.db).
  - blank rows were removed from the table
- Specific SQL queries were performed on the table SPACEXTBL to answer the given questions (for more details, see Jupyter Notebook)

Notebook on GitHub:
https://github.com/spongefrog/dsCapstone/blob/2692d1f7d3bdb8fc0abc45568ad5043fef656bf5/DS05_Capstone_SQL.ipynb

Build an Interactive Map with Folium

- Data for the Interactive Map were retrieved from https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/spacex_launch_geo.csv:
  - other than in the previous sections, this dataset contains 56 entries of Falcon 9 launches
- The library Folium was used to create interactive leaflet maps:
  - using Circle and Marker objects, all Launch Sites were marked on the map
  - MarkerCluster objects were created, to indicate the number of successful/unsuccessful launches at each launch site
  - Line and Marker objects were added, to depict distances to certain landmarks in the proximity of the Launch Site
  - (for more details, see Jupyter Notebook)

Notebook on GitHub:
https://github.com/spongefrog/dsCapstone/blob/2692d1f7d3bdb8fc0abc45568ad5043fef656bf5/DS06_Capstone_Folium.ipynb

Build a Dashboard with Plotly Dash

- Data for the Dashboard were retrieved from https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/spacex_launch_dash.csv:
  - similar as for the Interactive Map exercise, the dataset contains 56 entries of F9 launches
- The library Plotly Dash was used to create an interactive dashboard:
  - a dropdown menu allows to select a specific Launch Site or All Launch Sites
  - a pie chart depicts the distribution of successful missions across the four Launch Sites (when All Launch Sites are selected) or the percentage of successful/unsuccessful missions (when a specific Launch Site is selected)
  - a slider allows to select the lower and upper limit of a payload mass range (from 0 to 10,000 kg in steps of 1,000 kg) used for a scatter plot
  - a scatter plot shows the outcome as a function of payload mass for different booster versions (for the payload mass range selected with the slider

Plotly Dash/Python Script on GitHub:
https://github.com/spongefrog/dsCapstone/blob/cfc9cc4997c726b4dac577d6f162254ff2d63fae/DS07_Capstone_Dashboard.py

Predictive Analysis (Classification)

- Data for the Predictive Analysis were retrieved from https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/dataset_part_3.csv:
  - This is the filtered and prepared dataset from the SpaceX API with 90 entries of F9 launches.
  - It contains the numerical features Flight Number, Payload Mass, Number of Flights, Block Version, Count of Reusage, and the one-hot encoded categorical features Orbit Type, Serial, Reusage (0/1), Grid Fins (0/1), Legs (0/1).
  - The landing outcome was coded binarily (0 = unsuccessful / 1 = successful) and served as label.
- The machine learning library Scikit-learn was used to:
  - Standardize the data.
  - Split the data into training data (80%) and test data (20%).
  - To select hyperparameters for for each model (using the training data). -> see next slide
  - To calculate accuracy, Jaccard index (relative number of true positive (TP) to the sum of true positive and false positive (FP) predictions) and F1 score (harmonic mean between the classification precision and classification sensitivity) for each trained model being applied to the test data.

Predictive Analysis (Classification)

- The following models were trained:
  - a **Logistic regression** classifier (using ridge regression and L-BFGS optimization and a 5-fold cross validation procedure to tune the complexity penalty C between the values 0.01, 0.1, or 1)
  - a **Support vector machine** classifier (using a 10-fold cross validation procedure to tune the hyperparameters: kernel type ['linear', 'rbf','poly','rbf', 'sigmoid'], the penalty C [out of five logspaced values between 0.001 and 1000] and the curvature gamma for the RBF kernel [out of five logspaced values between 0.001 and 1000])
  - a **Decision tree** classifier (using a 10-fold cross validation procedure to tune the hyperparameters: criterion [gini or entropy], splitter [best or random], maximum depth [2, 4, 6, 8, 10, 12, 14, 16, 18], maximum features [auto or sqrt], minimum samples leaf [1, 2, 4], min_samples_split [2, 5, 10])
  - a **K-Nearest-Neighbors** classifier (using a 10-fold cross validation procedure to tune the hyperparameters: number of neighbors [1 to10], algorithm [auto, ball_tree, kd_tree, brute], and p [Euclidean or Manhattan distance])

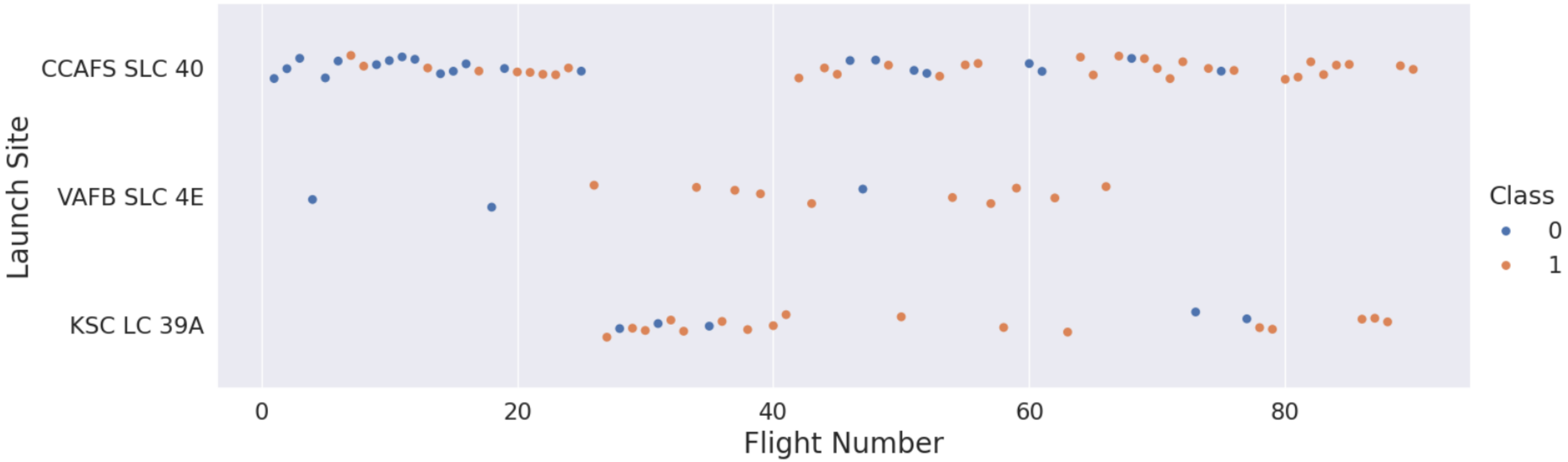
Notebook on GitHub:
https://github.com/spongefrog/dsCapstone/blob/f82b49150fda0e871b32d59506a03ea05ac6e9b2/DS08_Capstone_ML_Prediction.ipynb

Section 2

Insights drawn from EDA

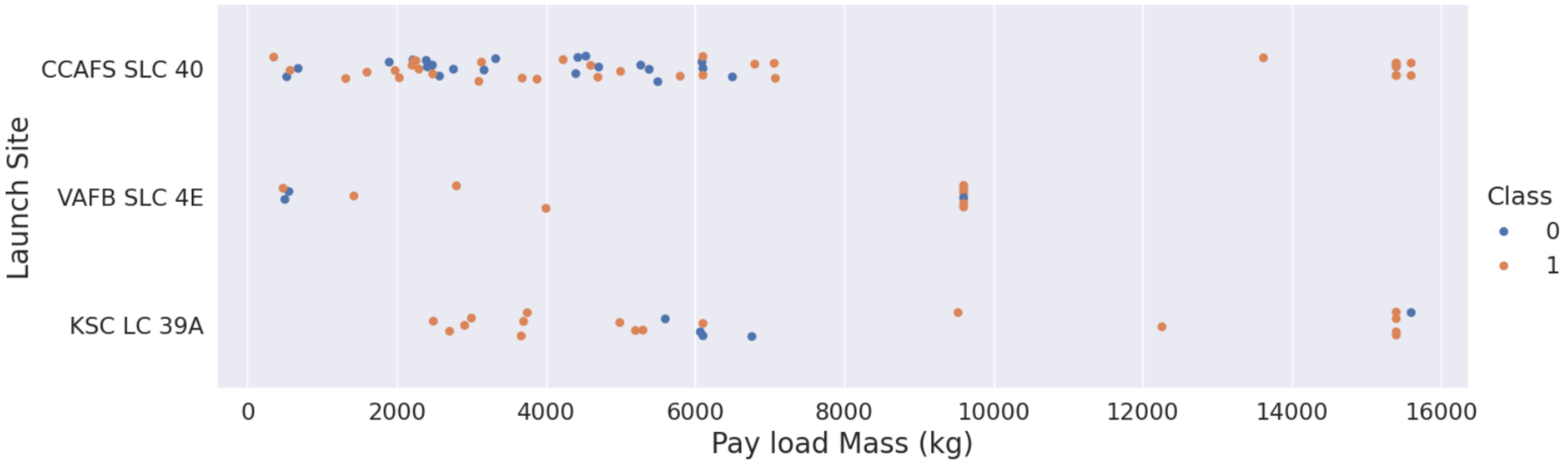
Flight Number vs. Launch Site

- Success rate was rather low for early flights (flight numbers 1-20) and increased with flight number
- Most launches (including the very first launch) of Falcon 9 were accomplished from Space Launch Complex 40 (CCAFS SLC-40) located on Cape Canaveral



Payload vs. Launch Site

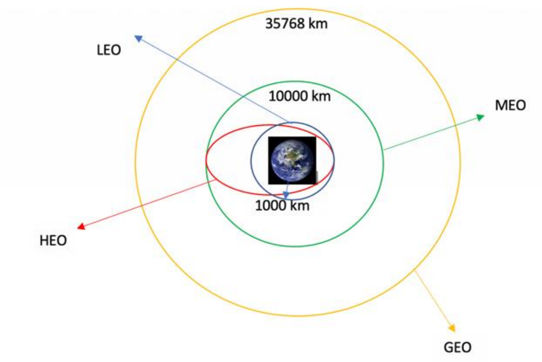
- Carriers with payload mass above 10,000 kg were only launched from Florida (CCAFS SLC-40 and KSC LC-39A)
- Launches with high payload mass (>10,000 kg) had an overall high success rate (note, that flight number might still be a mediating factor in the relationship)



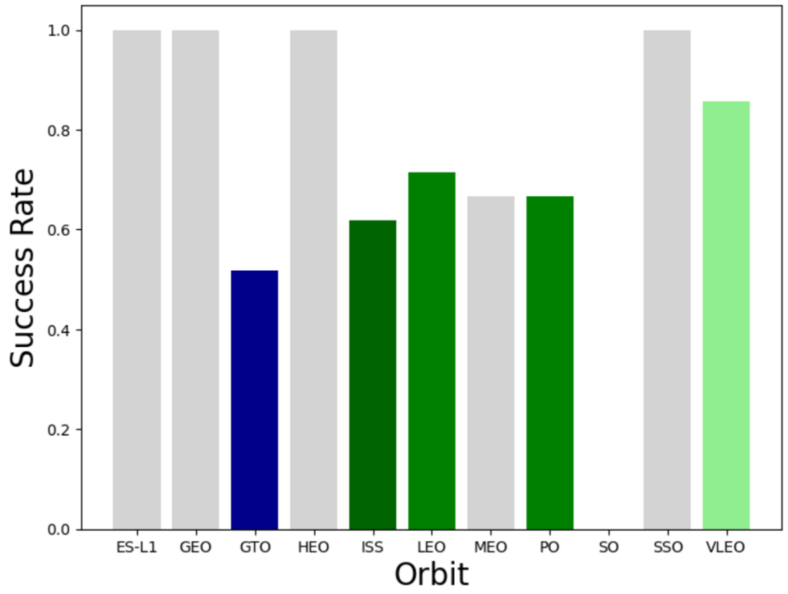
Success Rate vs. Orbit Type

- Most of the launches were aimed at geostationary transfer orbit (GTO), the ISS orbit (400 km), and Very low Earth orbit (VLEO, < 400 km).
- Success rate for landings seems to increase the lower the orbit (when considering orbits for which a reasonable amount of data is available, i.e. N>5, see colored bars).
- Yet, the higher success rate for VLEO launches could partly be explained by flight number (see next slide)

Orbit types:

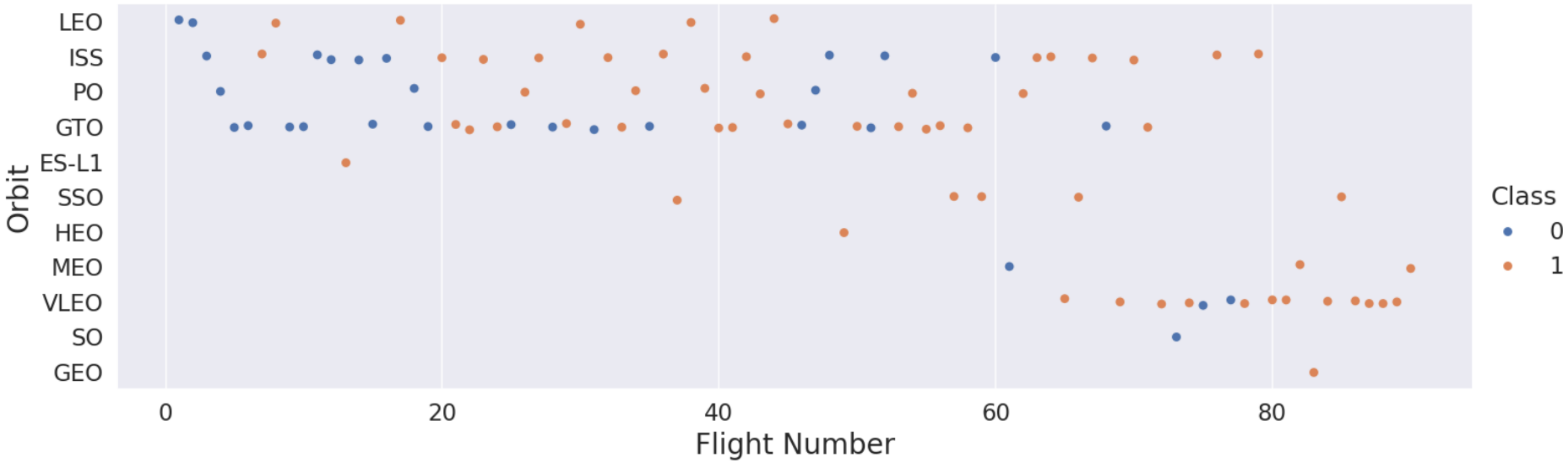


GTO	27
ISS	21
VLEO	14
PO	9
LEO	7
SSO	5
MEO	3
ES-L1	1
HEO	1
SO	1
GEO	1



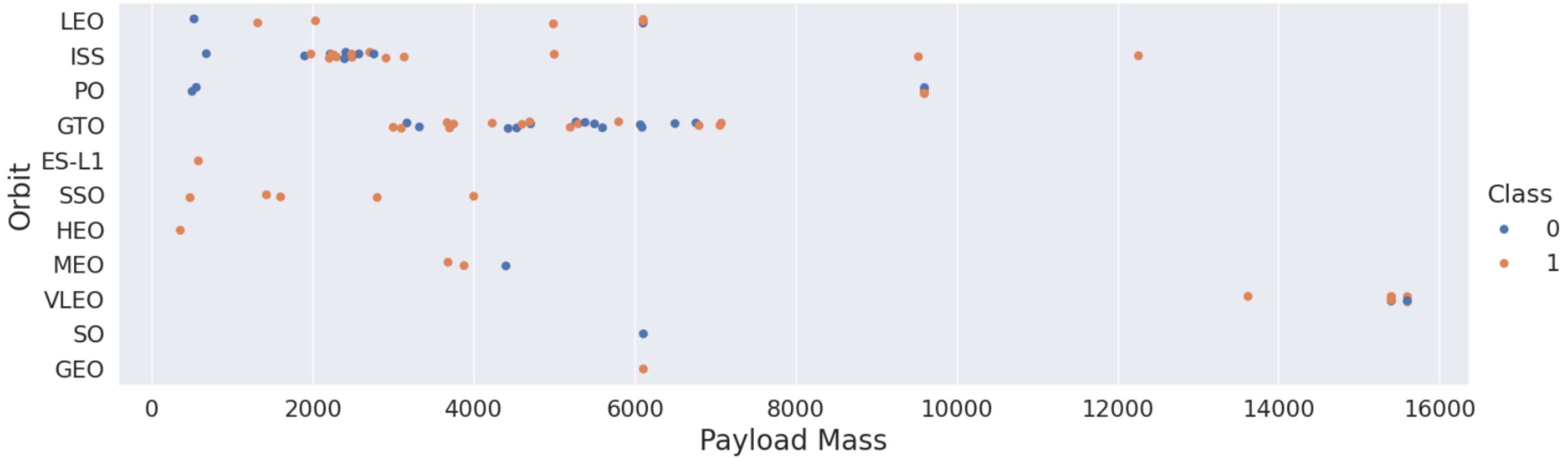
Flight Number vs. Orbit Type

- Launches to Very low Earth orbit (VLEO, < 400 km) have a high success rate (only 2 failures out of 14 flights), and they only occurred with flight numbers > 64
- Other orbit types have a success rate of even 100% (ES-L1, SSO, HEO, GEO), but the number of flights is too limited to reliably estimate the “true” success rate



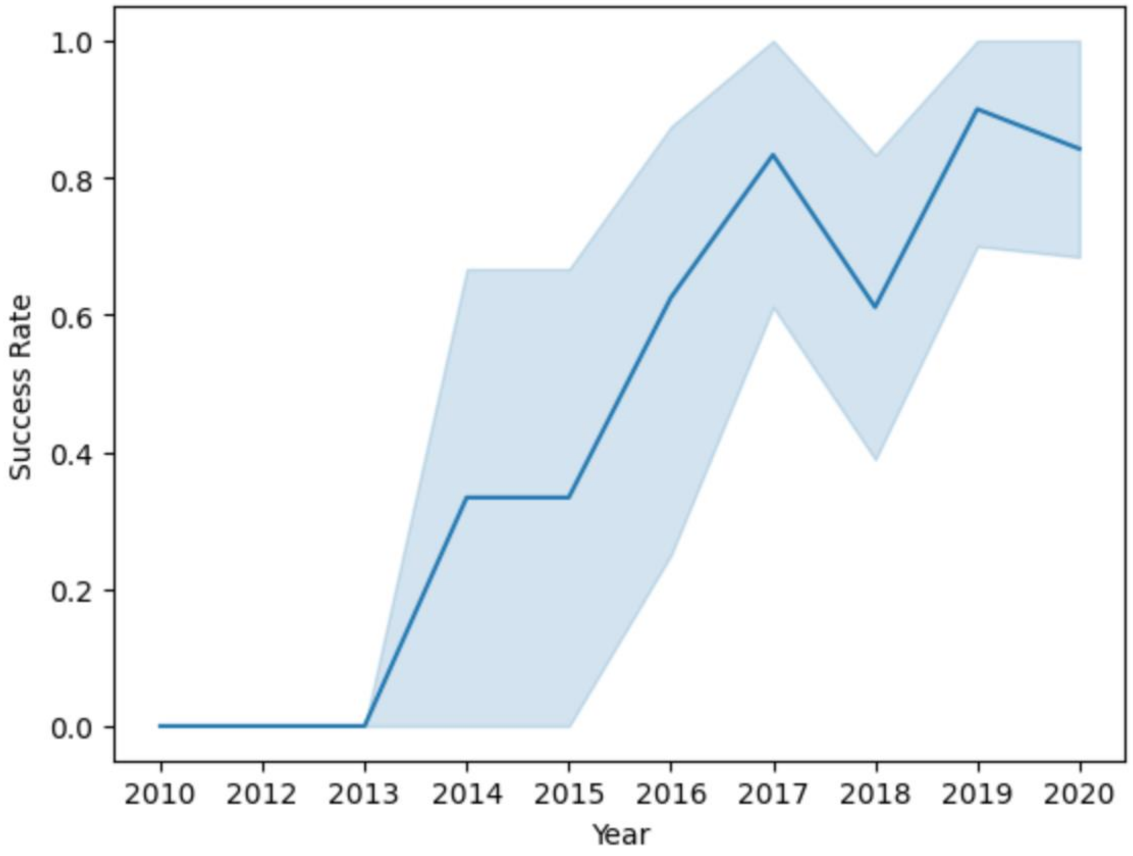
Payload vs. Orbit Type

- Launches to VLEO all carried payload mass above 13,000 kg.
- Median payload mass for launches to ISS orbit was around 2,500 kg, to GTO orbit around 5,000 kg.



Launch Success Yearly Trend

- Landing Success Rate increased over the years.
- The **first successful controlled landing** was achieved in 2014 (please note, that controlled ocean touchdowns are included in the category of successful landing).
- In 2020, a Success Rate of 84.2 % was reached.



All Launch Site Names

Lauches were carried out from three launch sites in Florida and California (one of them occurs with its current and its former name in the list):

Florida:

- Cape Canaveral Air Force Station: Space Launch Complex 40 (CCAFS SLC-40), formerly known as Launch Complex 40 (CCAFS LC-40)
- Kennedy Space Center Launch Complex 39A (KSC LC-39A)

California:

- Vandenberg AFB Space Launch Complex 4 (VAFB SLC-4E)

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

Launch Site Names Begin with 'KSC'

- Launch Complex 39 at the John F. Kennedy Space Center (KSC LC-39A) was first used on February 19th, 2017 to launch a Falcon 9.
- The first five launches of Falcon 9 from KSC LC-39A had the following specifications:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2017-02-19	14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
2017-03-16	6:00:00	F9 FT B1030	KSC LC-39A	EchoStar 23	5600	GTO	EchoStar	Success	No attempt
2017-03-30	22:27:00	F9 FT B1021.2	KSC LC-39A	SES-10	5300	GTO	SES	Success	Success (drone ship)
2017-05-01	11:15:00	F9 FT B1032.1	KSC LC-39A	NROL-76	5300	LEO	NRO	Success	Success (ground pad)
2017-05-15	23:21:00	F9 FT B1034	KSC LC-39A	Inmarsat-5 F4	6070	GTO	Inmarsat	Success	No attempt

Total Payload Mass

- NASA (CRS) launched in sum a total payload mass of 45,596 kg.

SUM("PAYLOAD_MASS__KG_")
45596

Image source: https://www.nasa.gov/wp-content/uploads/2018/07/spacex_crs-9_mission_overview2.pdf?emrc=6c82e4

Average Payload Mass by F9 v1.1

- Missions using the booster version Falcon 9, v1.1 carried on average 2,928.4 kg.

AVG("PAYLOAD_MASS__KG_")
2928.4

Falcon 9 v1.1

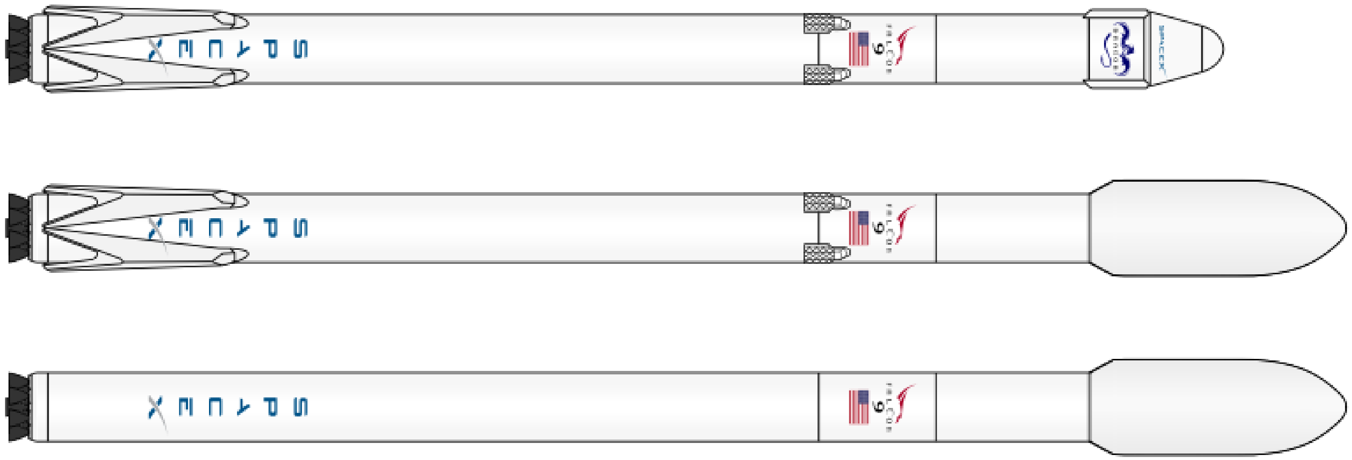


Image source: https://en.wikipedia.org/wiki/List_of_Falcon_9_first-stage_boosters#/media/File:F9_and_Heavy_visu.png

First Successful Ground Landing Date

- The first succesful landing on a drone ship was achieved on April 8th, 2016.

MIN("Date")
2016-04-08

First successful landing of a Falcon 9 first stage on a drone ship. The drone ship's name was "Of Course I Still Love You".

Image source: https://en.wikipedia.org/wiki/Autonomous_spaceport_drone_ship#/media/File:CRS-8_(26239020092).jpg

Successful Drone Ship Landing with Payload between 4000 and 6000

- The boosters with a success in ground pad and a payload mass greater than 4000 but less than 6000 kg are F9 FT B1032.1, F9 B4 B1040.1, F9 B4 B1043.1.

Booster_Version
F9 FT B1032.1
F9 B4 B1040.1
F9 B4 B1043.1

Image source: https://www.nasaspaceflight.com/2017/08/falcon-9-block-4-debut-success-dragon-station-berthing/

Total Number of Successful and Failure Mission Outcomes

- For all listed Falcon 9 missions, the mission outcomes were in 100 cases success and in only one case failure.
- (Note, that mission outcome does not necessarily require a successful landing outcome; e.g. when not attempted. That is the reason why the success rate related to overall mission outcome is larger than the success rate related to landing outcome.)

Boosters Carried Maximum Payload

- Booster versions which have carried the maximum payload mass are Falcon 9, Block 5.

Image source: https://en.wikipedia.org/wiki/List_of_Falcon_9_first-stage_boosters#/media/File:F9_and_Heavy_visu.png

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 February, May, June, August, September and December of 2017, six successfully landed missions were carried out (all landing on ground pad; using Falcon 9, Block 4 or Falcon 9 FT rockets, five of them launched on KSC LC-39A).

Month	Landing_Outcome	Booster_Version	Launch_Site
02	Success (ground pad)	F9 FT B1031.1	KSC LC-39A
05	Success (ground pad)	F9 FT B1032.1	KSC LC-39A
06	Success (ground pad)	F9 FT B1035.1	KSC LC-39A
08	Success (ground pad)	F9 B4 B1039.1	KSC LC-39A
09	Success (ground pad)	F9 B4 B1040.1	KSC LC-39A
12	Success (ground pad)	F9 FT B1035.2	CCAFS SLC-40

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

- Between the date 2010-06-04 and 2017-03-20, 5 successful landings on drone ships and 3 successful landings on ground pad were achieved.

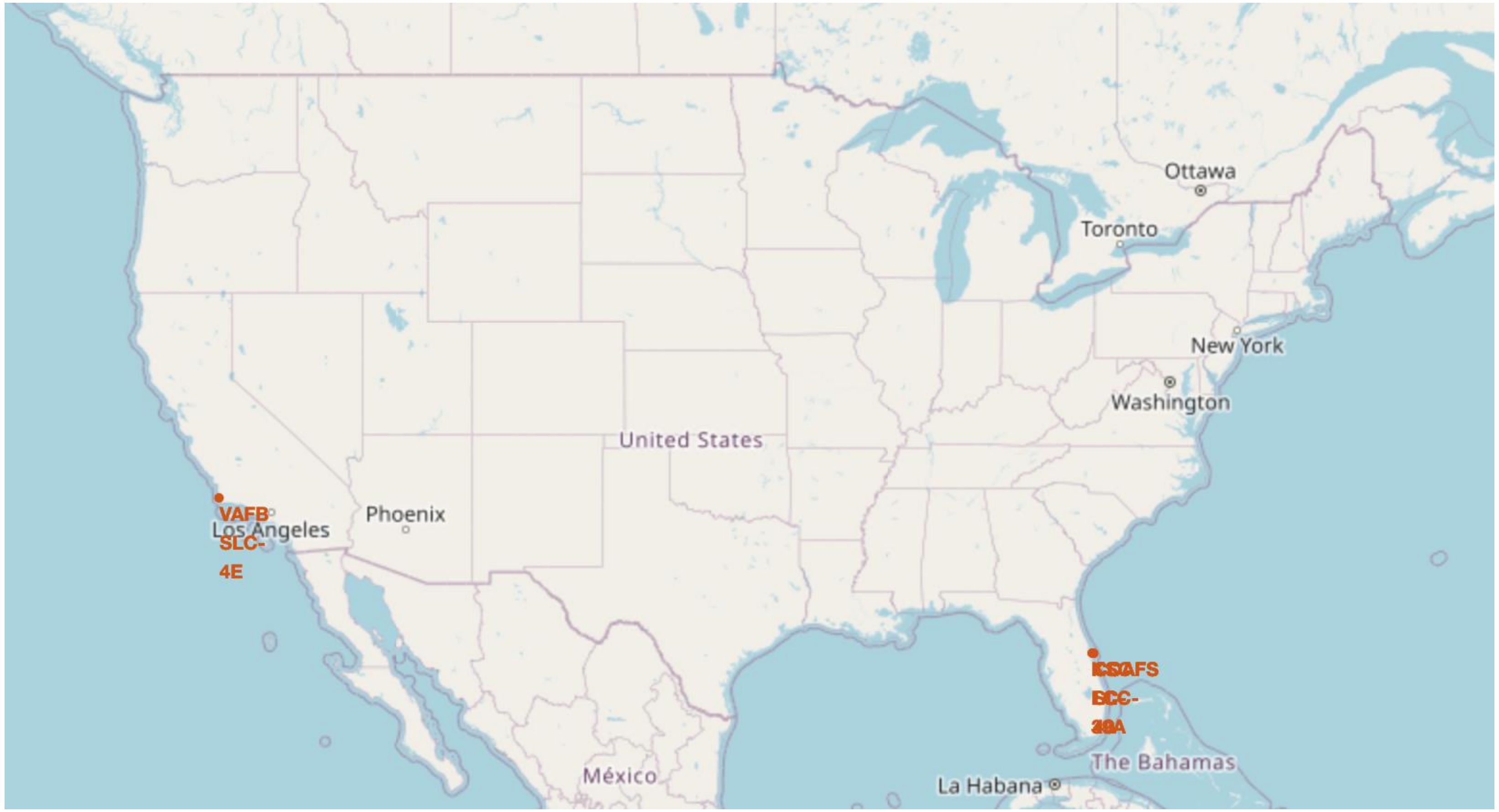
Landing_Outcome	COUNT(*)
Success (drone ship)	5
Success (ground pad)	3

Section 3

Launch Sites Proximities Analysis

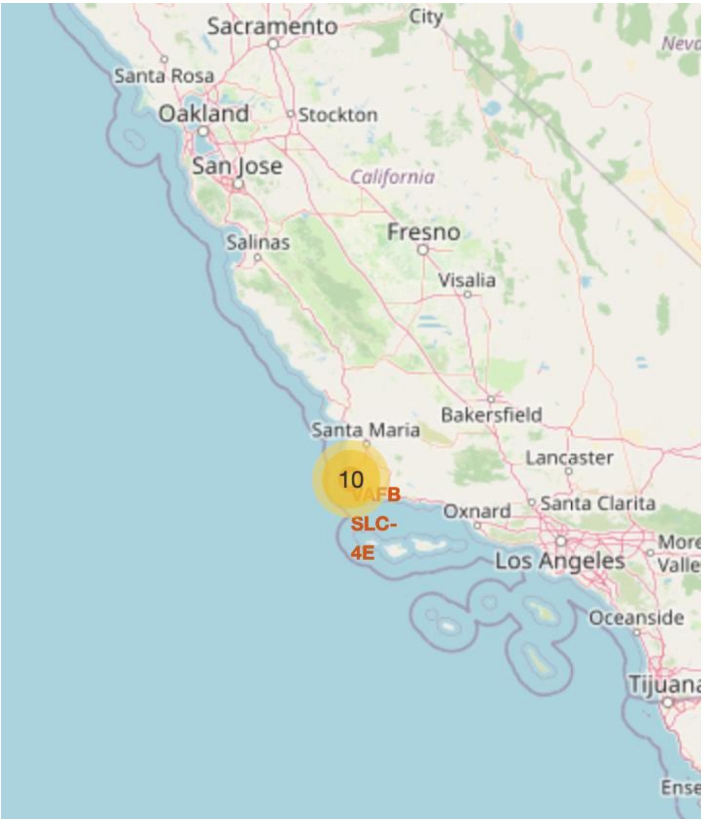
Locations of Launch Sites

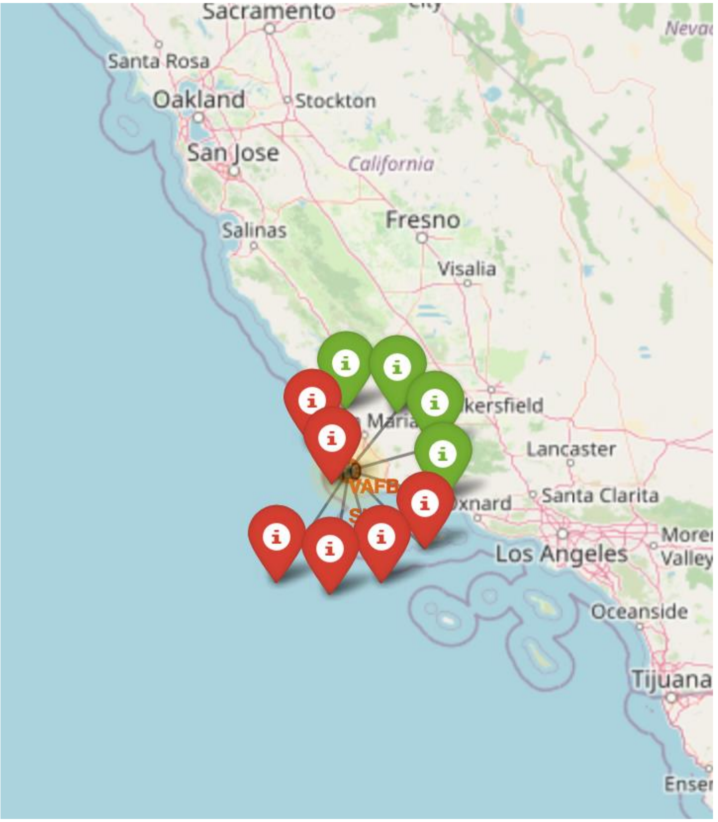
- The three launch sites CCAFS SLC-40/LC-40, KSC LC-39A, and VAFB SLC-4E are located at the US east and west coast in Florida and California.



Launches at specific Launch Sites

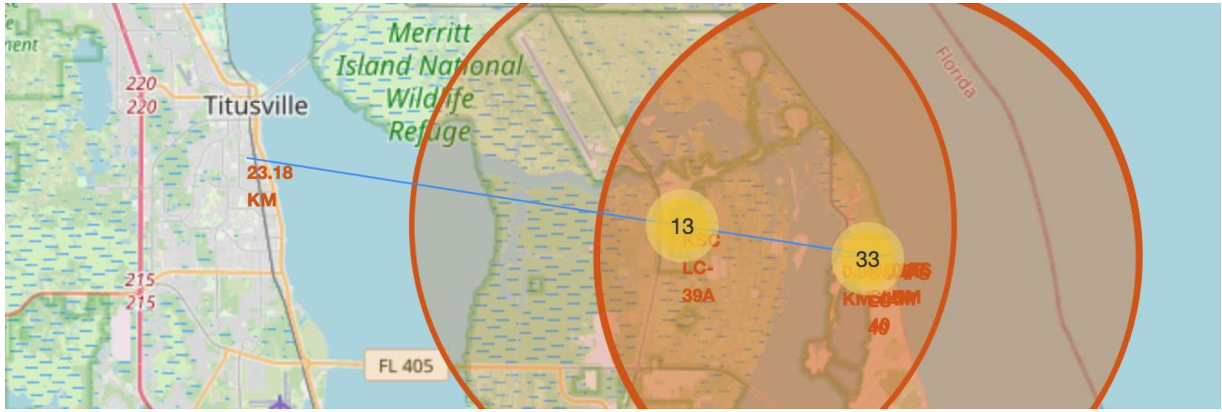
- The maps to the right depict the location of the VAFB SLC-4E launch site.
- Interactively, the maps indicate the number of launches (yellow circle) and upon mouse click the successful (green flag) and unsuccessful (red flag) missions launched from VAFB SLC-4E

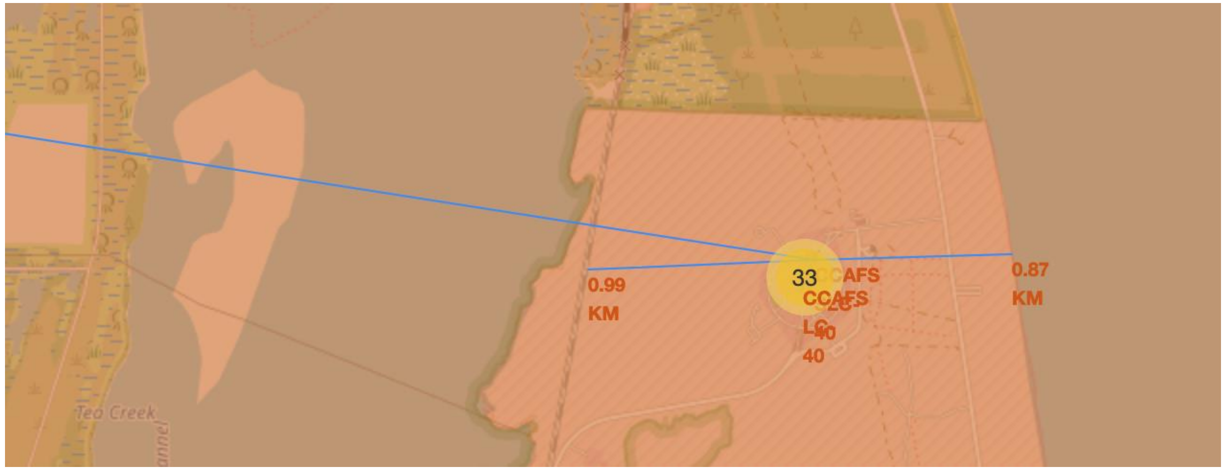




Launch site proximities

- The upper map to the right shows the locations of the Florida launch sites.
- Blue lines indicate the distance from CCAFS SLC-40/LC-40 to the nearest city Titusville (23.18 km), to the nearest coast line (0.87 km), and to the nearest railway line (0.99 km).
- **Launch sites are far from cities, near the coast and near relevant transport infrastructure.**



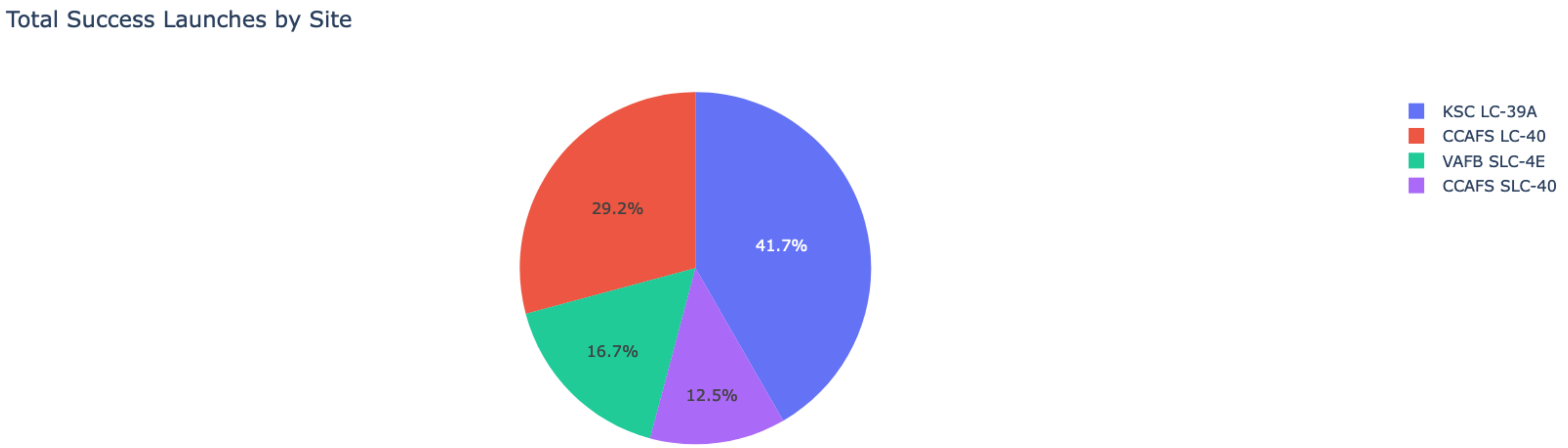


Section 4

Build a Dashboard with Plotly Dash

Distrubution of successful launches over launch sites

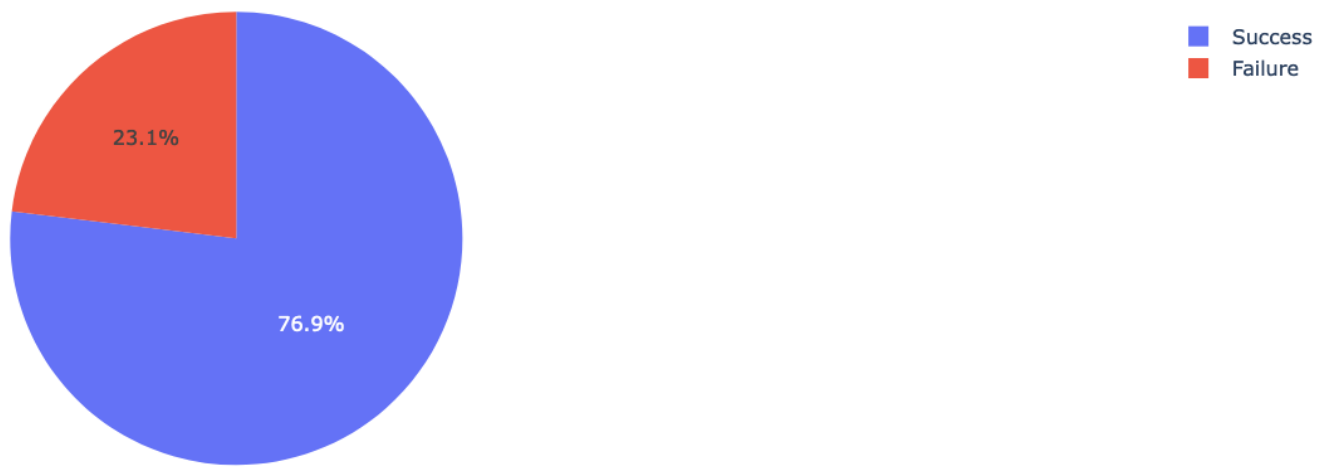
- Out of the total number of successful launches, 41.7 % occurred from KSC LC-39A.



Success rate of launches for specific launch site

- At KSC LC-39A, 76.9% of the Falcon 9 launches were successful.

Total Success Launches for Site KSC LC-39A



Launch outcome for specific payload mass ranges

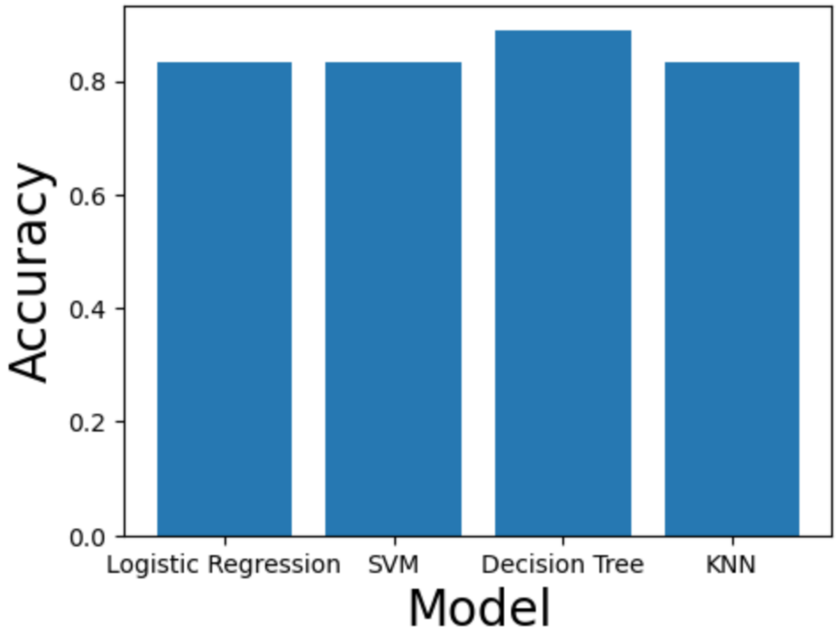
- For payload masses between 2,000 and 4,000 kg, there were more successful than unsuccessful mission outcomes (left graph).
- For payload masses between 5,000 and 9,000 kg, there were more unsuccessful than successful mission outcomes (right graph).

Section 5

Predictive Analysis (Classification)

Classification Accuracy

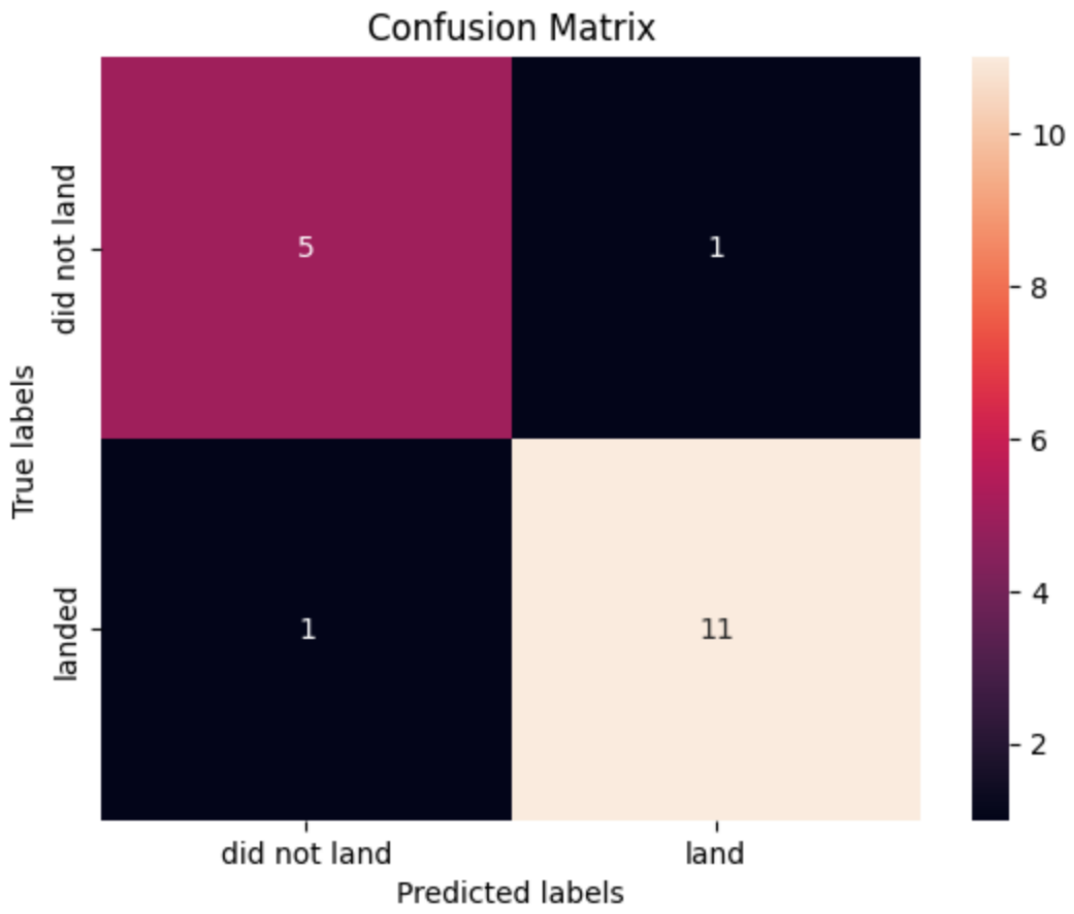
- Test accuracies for the built classification models (Logistic regression, SVM, Decision tree, KNN) were 83.3 % or larger.
- The Decision tree model had a slightly larger test accuracy (88.8%) than the other models (all 83.3%). Yet, the accuracy value could vary quite a bit depending on randomisation.



	Model	Accuracy	Jaccard	F1-Score
0	Logistic Regression	0.833333	0.714286	0.833333
1	SVM	0.833333	0.714286	0.833333
2	Decision Tree	0.888889	0.800000	0.888889
3	KNN	0.833333	0.714286	0.833333

Confusion Matrix

- Confusion matrix of the best performing model, i.e., the Decision tree classifier.
- The classifier achieved a true positive rate of 11 out of 12, while having a false negative rate of 1 out of 6.
- Overall, 2 out of 18 cases are misclassified.



Conclusions

- Since the first controlled ocean touchdown of a Falcon 9 v1.1's first stage in 2014, the landing success rate continuously increased over the years (reaching 84.2 % in 2020).
- The more recent development of newer Booster versions (e.g., the Falcon 9, Block 5) allowed the vehicles to carry up to 15,600 kg of payload mass to low earth orbit.
- Landing success rate is comparably high for flights carrying more payload mass and for flights to low and very low earth orbit (LEO, VLEO), likely with flight number as a moderating variable.
- Machine Learning Classifiers can predict the Landing Success based on predictors, such as Flight Number, Payload Mass, Orbit Type and others, with an accuracy of around 84%.

Appendix

- Python code used during this project can be found on GitHub:
  - Data Collection API:
    https://github.com/spongefrog/dsCapstone/blob/f82b49150fda0e871b32d59506a03ea05ac6e9b2/DS01_Capstone_DataCollection_SpaceX_API.ipynb
  - Data Collection WebScraping:
    https://github.com/spongefrog/dsCapstone/blob/f82b49150fda0e871b32d59506a03ea05ac6e9b2/DS02_Capstone_DataCollection_WebScraping.ipynb
  - Data Wrangling:
    https://github.com/spongefrog/dsCapstone/blob/f82b49150fda0e871b32d59506a03ea05ac6e9b2/DS03_Capstone_DataWrangling.ipynb
  - Data Visualization:
    https://github.com/spongefrog/dsCapstone/blob/f82b49150fda0e871b32d59506a03ea05ac6e9b2/DS04_Capstone_DataVisualization.ipynb
  - SQL queries:
    https://github.com/spongefrog/dsCapstone/blob/f82b49150fda0e871b32d59506a03ea05ac6e9b2/DS05_Capstone_SQL.ipynb
  - Interactive Location Maps (Folium):
    https://github.com/spongefrog/dsCapstone/blob/f82b49150fda0e871b32d59506a03ea05ac6e9b2/DS06_Capstone_Folium.ipynb
  - Interactive Dashboard (Plotly Dash):
    https://github.com/spongefrog/dsCapstone/blob/f82b49150fda0e871b32d59506a03ea05ac6e9b2/DS07_Capstone_Dashboard.py
  - ML Predictions (Scikit-learn):
    https://github.com/spongefrog/dsCapstone/blob/f82b49150fda0e871b32d59506a03ea05ac6e9b2/DS08_Capstone_ML_Prediction.ipynb

Thank you!