IBM Developer
SKILLS NETWORK

# Winning Space Race with Data Science

<Britta Smith>
<January 5, 2023>

# Outline: Space X to Y

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

# Executive Summary: Space X to Y

- The commercial space industry utilizes partially reusable medium lift launch vehicles (MLVs). MLVs are a type of rocket launch vehicle capable of transporting payloads of 2,000 to 20,000 kg to low Earth orbit (LEO), per NASA definition. Reuse relies upon landing of the first stage (booster or core) after the second stage, and payload separate and continue on the mission. One such vehicle is the SpaceX Falcon 9 which began testing in 2010 with version v1.0, first mission in 2012. In 2013, Falcon 9 v1.1 was introduced, then in 2015 Falcon 9 Full Thrust of which the Block 5 version has been flying since 2018. Falcon 9 Full Thrust is 230 ft tall and 12 ft wide. SpaceY has a partially reusable MLV of similar design and capabilities. Using SpaceX historical launch data can inform SpaceY business decisions.

- Methodologies occurred in a logical and sequential process. As SpaceY is new to the industry, SpaceX 2010 to 2020 was deemed most appropriate. To collect data the SpaceX REST API and web scrapping were used; and supplemented collection utilizing API sources. This data was then organized, processed, and cleaned using both Python and Microsoft Excel. Exploratory data analysis (EDA), visualization, and statistical analysis was performed using Python, IBM DB2 databases, and SQL. Interactive visuals were created using Folium for maps, and Plotly Dash. Model building, choice and evaluation involved algorithm parameter optimization, encoding of categorical variables, feature selection to choose independent variables, and scoring metric calculation to evaluate model accuracy. Lastly, analysis of available contract data used Python and Tableau to translate insights into a Dashboard tool.

- Results can be summarized by independent variable relationship to the target landing success or failure. Flight number seems to represent an increase in landing success over time. Launch site is more highly correlated with Orbit. Payload masses between 2,500 – 5,000 kg were the most successful. The orbits with the highest success rates and five or more flights include: SSO, VLEO, and LEO. Landing success occurs with nearly equal frequency with both non-reused cores and reused cores. The presence of grid fins and legs likely reflects the intention to land the core. From 2010 to 2020 landing success rate increased from 0% to 84.2%. The Support Vector Machine (SVM or SVC) model was able to achieve an F1 score weighted average of 100% with best parameters and feature selection. EDA of contract data allowed for cost/revenue/profit estimates and the creation of a Tableau dashboard tool to further aide in competitive contract bidding.

# Introduction: Space X to Y

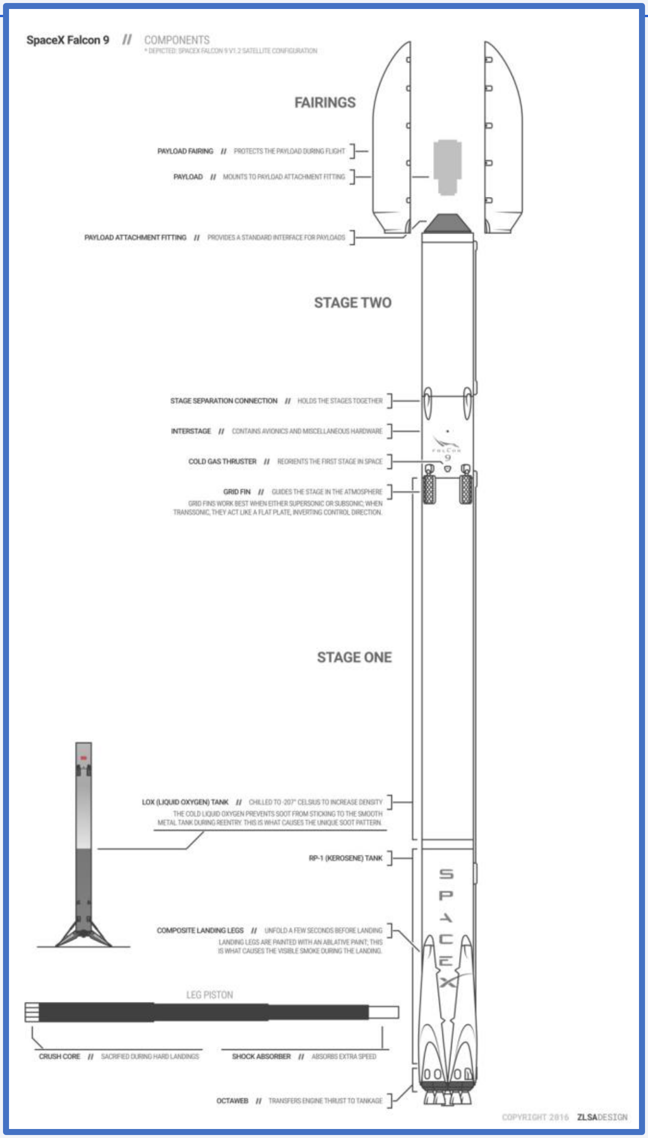
- **Project background and context**
  - SpaceX currently leads the commercial space transport market with the Falcon 9 rocket.
  - SpaceY has a reusable rocket with even more reusable components, and is entering into the marketplace. They aim to transport cargo from Earth to various orbital locations with a higher degree of success. Customers they are vying for include commercial telecommunications companies, governments, and research.
  - In order for SpaceY to offer SpaceX customers financially appealing contract alternatives, SpaceY costs must also be considered. This project will help SpaceY use SpaceX data as intelligence to ensure a successful entry into this commercial space sector.
- Business questions (problems)
  - How much should SpaceY bid on a mission contract?
  - How likely is the Falcon 9 booster (first stage or core) to land successfully?
  - How much will the mission cost?
- Hypotheses
  - SpaceX historical data provides a combination of variables that influence the target variable, landing success or failure. This prediction along with SpaceX financial analysis will provide SpaceX with probable mission costs, which can guide contract formulation.

# Background: Falcon 9 Components

**Booster Version**:
The most recent Falcon 9 booster version is Falcon 9 Full Thrust Block 5
There have been Blocks 1, 2, 3, 4, and 5 numbered sequentially

**Core**:
Also referred to as stage one, first stage, and booster
This is the section that disconnects from the upper portion and lands

**Serial**:
A number assigned to the core, B1XXX.Y, B = Booster, 1 = 1st stage, XXX incremental number starting at 001, .Y= mission information




**Fairings**:
Two piece protective shell enclosing the **payload**
They are shed when aerodynamic stresses no longer present a danger to the payload

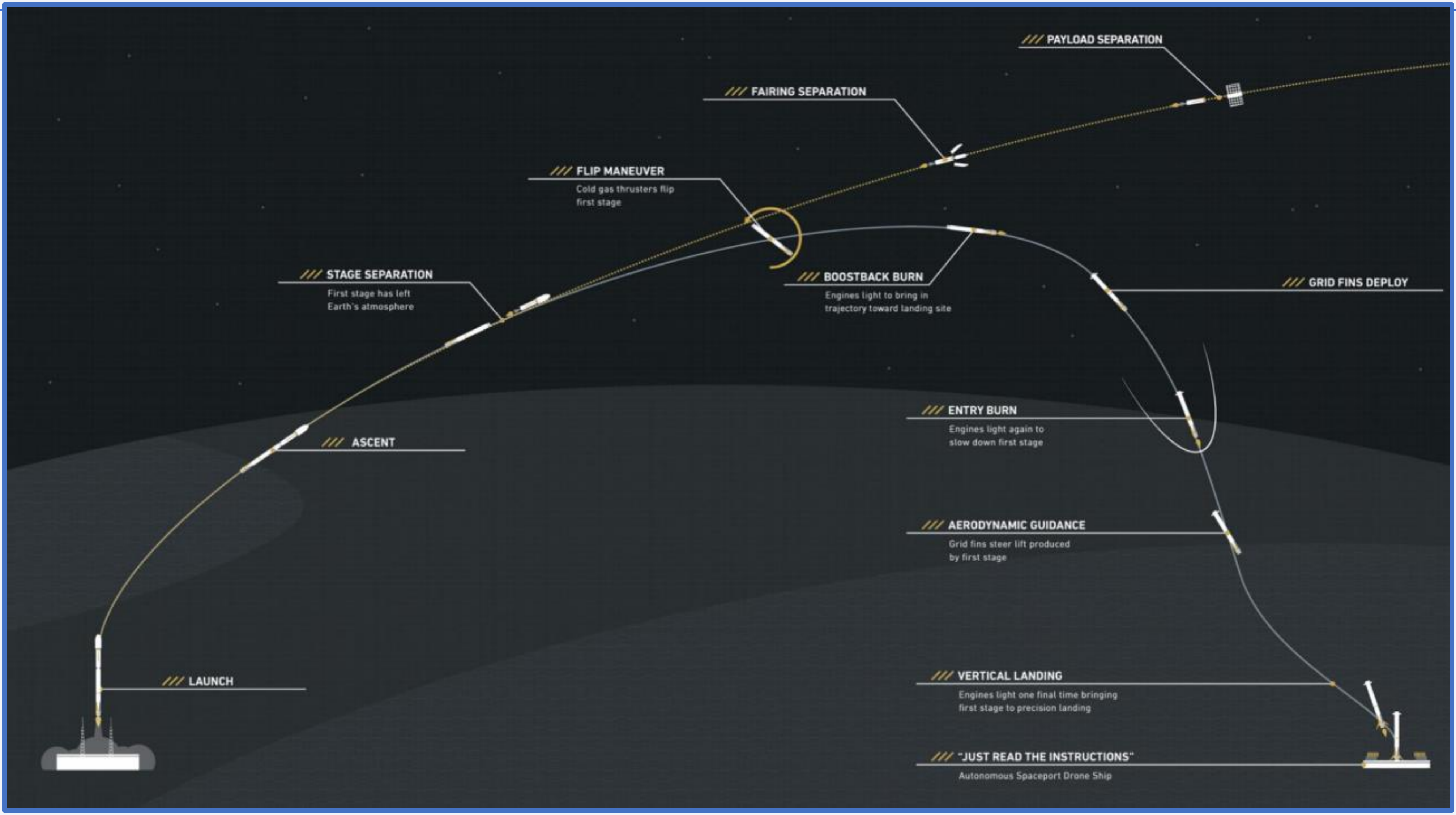
**Grid Fins**:
Deployed during re-entry to guide landing

**Legs**:
For landing in an upright position
Deployed shortly before touchdown



https://zlsadesign.com/infographic/vehicle/spacex-falcon9-components/

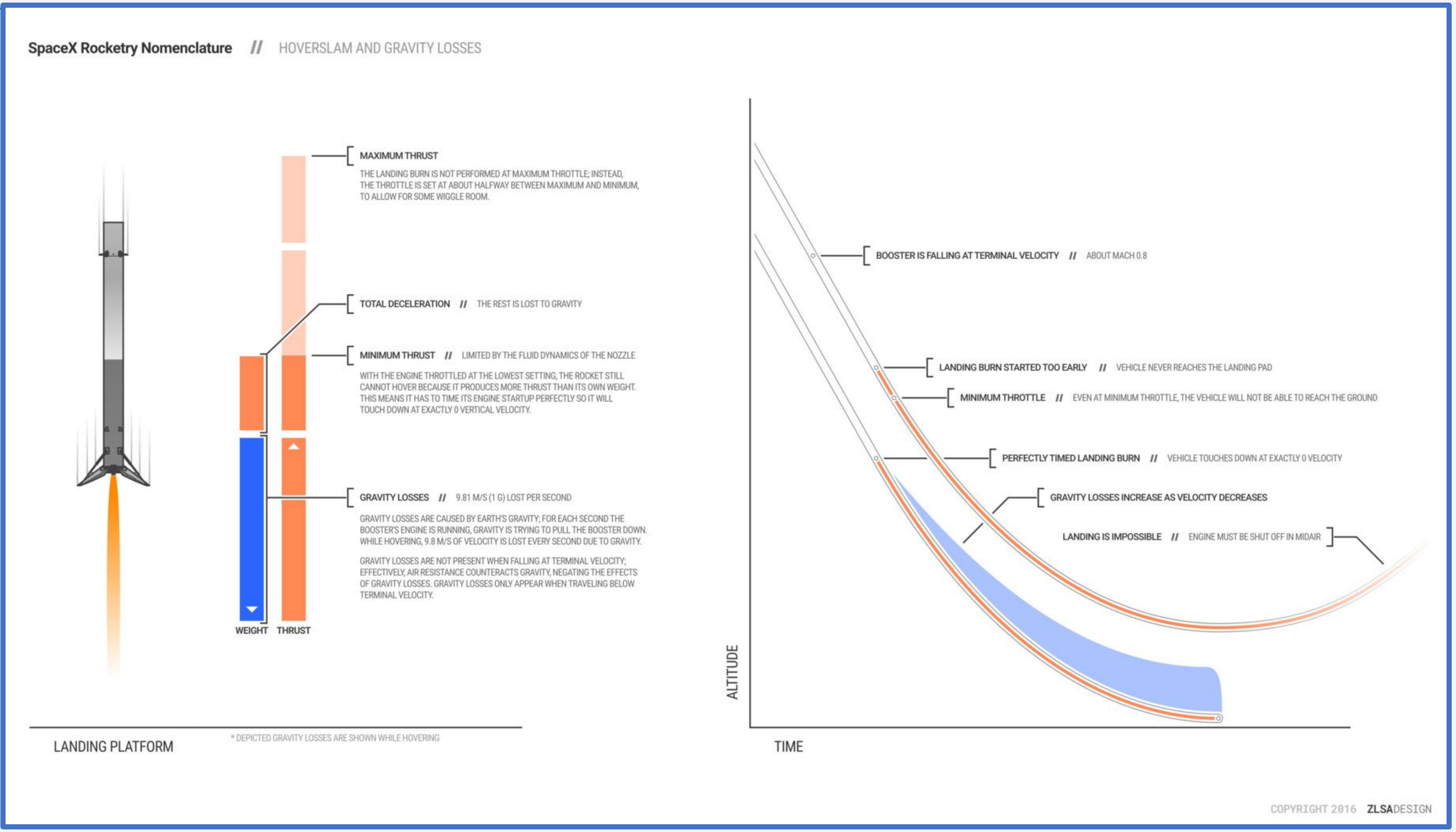
# Background: Launch and Landing




https://www.spacex.com/media/falcon-users-guide-2021-09.pdf

# Background: Landing

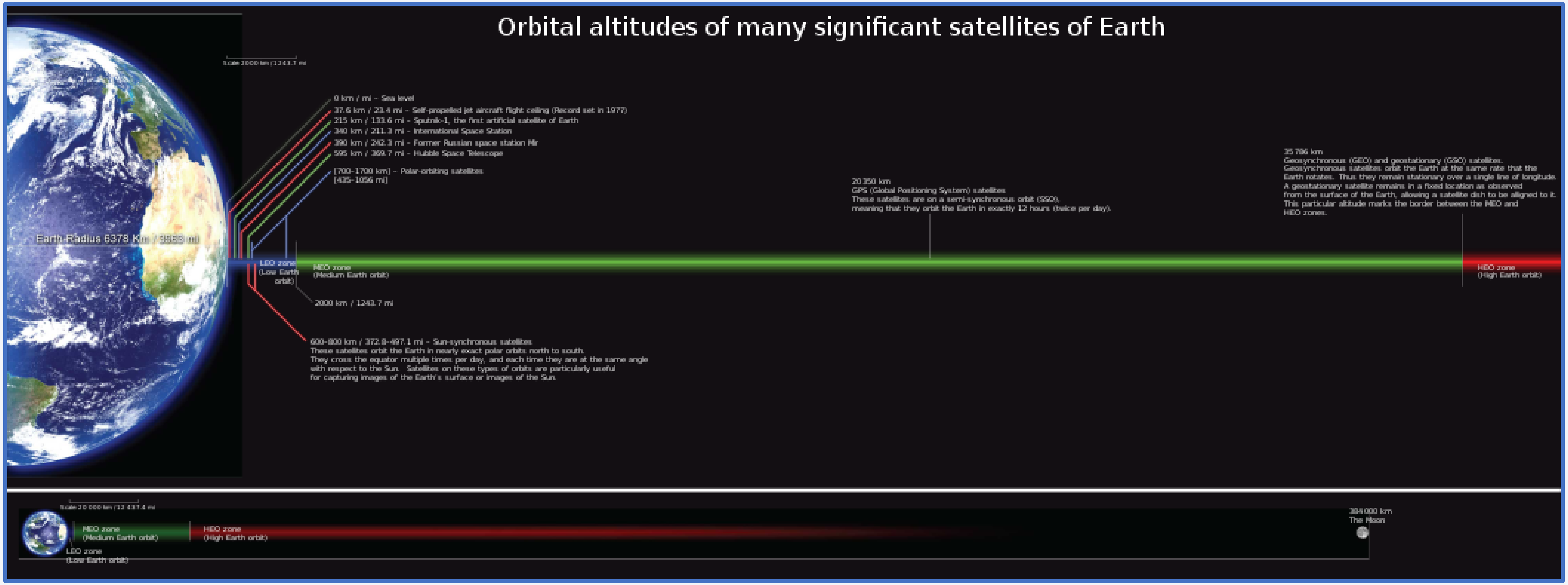
**Flights**:
Launch number of core (booster)
**Reused Count:**
Total number flights - 1 = number of times core reused
**Reused Core**:
Using a reused or new first stage
**Class**:
Successful or unsuccessful landing outcome
**Outcomes**:
None None: no attempt
True/False Ocean: controlled/uncontrolled into a specific region of the ocean
True/False RTLS = successfully/unsuccessfully landed to a ground pad
True/False ASDS = successfully/unsuccessfully landed on a drone ship






https://zlsadesign.com/infographic/trajectory/spacex-falcon9-booster-hoverslam/

# Background: Mission Orbits

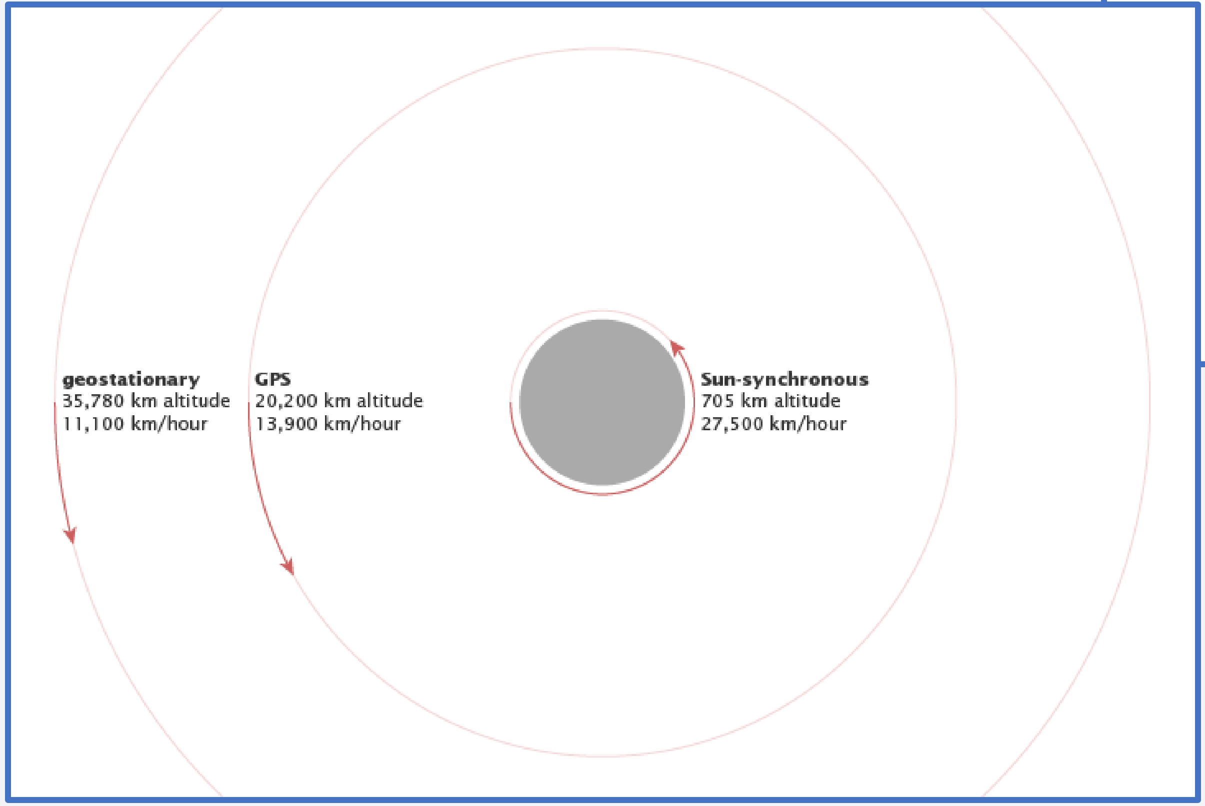


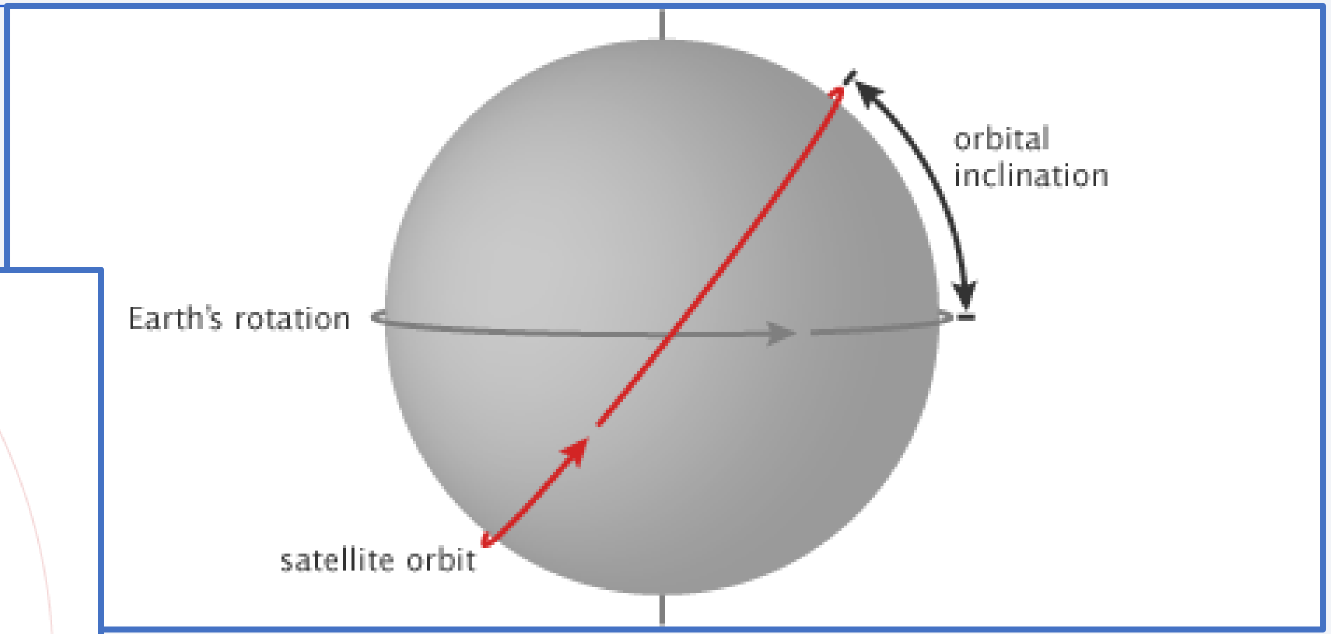

**Orbit: Orbit destination of payload**

GTO (geostationary transfer orbit)
ISS (International Space Station)
VLEO (Very low Earth orbit)
PO (polar orbit)
LEO (low Earth orbit)
SSO (Sun-synchronous orbit)
MEO (medium Earth orbit)
ES-L1 (type of Lagrange point orbit)
HEO (highly elliptical orbit)
SO (heliocentric orbit (HCO))
GEO (geosynchronous equatorial orbit)

https://en.wikipedia.org/wiki/Low_Earth_orbit#/media/File:Orbitalaltitudes.svg

# Background: Mission Orbits




- **For more information:**
- https://earthobservatory.nasa.gov/features/OrbitsCatalog
- https://solarsystem.nasa.gov/basics/chapter5-1/
- https://aerospace.csis.org/aerospace101/earth-orbit-101/

# Background: Recovery and Reuse

- **Reused Fairings:**
- Fairings deploy parachutes after separation and Mr. Steven(s) attempts to catch them before they hit the ocean
- https://www.youtube.com/watch?v=uuOwTNsCUjE&t=21s

- **Recovery and Reuse:**
- Once the first stage and fairings are recovered they are then refurbished and reused if deemed flight worthy
- This reduces cost substantially

https://www.space.com/41614-spacex-mr-steven-catcher-boat-up-close.html

Section 1

# Methodology

# Methodology

## Executive Summary

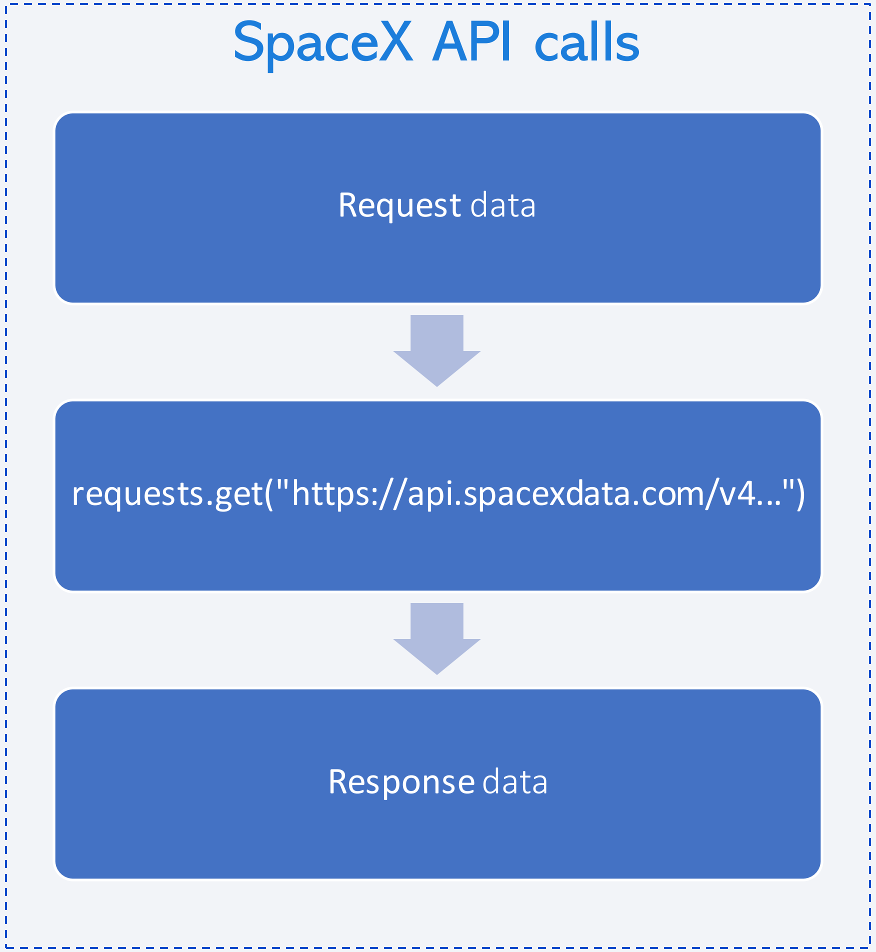
- Data collection methodology:
  - Utilize SpaceX REST API and web scrapping to obtain SpaceX historical launch data
- Perform data wrangling
  - Address missing values, obtain tabular organization, convert categorical variables
- Perform exploratory data analysis (EDA) using Python, visualization, and SQL
- Perform interactive visual analytics using Folium, Plotly Dash, and Tableau
- Perform predictive analysis using classification models
  - Build, tune, and evaluate classification models
  - Choose independent variables whose combination is statistically sound

# Data Collection

- Data Collection Methodology
  - API Requests
    - Using the Requests library to make HTTP requests from the API https://api.spacexdata.com/v4
  - Web Scrapping
    - Using the BeautifulSoup and Requests package to extract HTML tables from Wikipedia

# Data Collection – SpaceX API

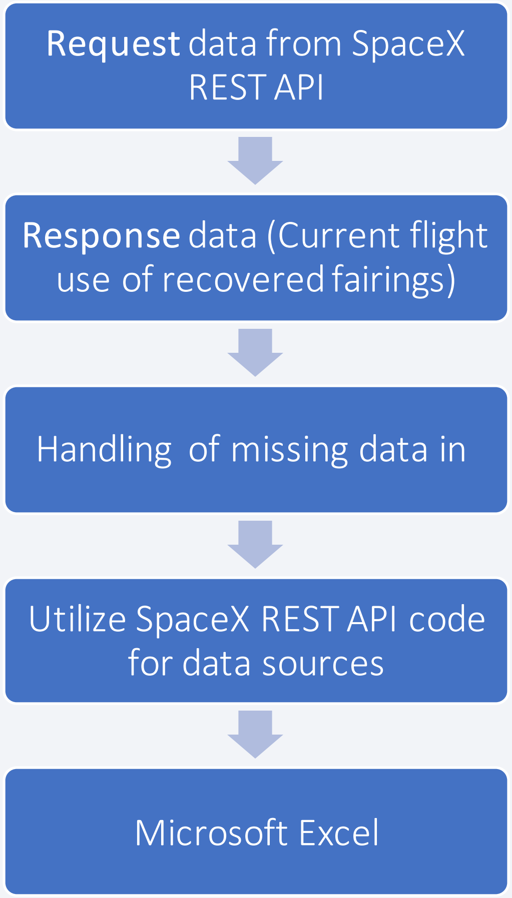
- Requests sent to the REST API URL https://api.spacexdata.com/v4/
- Endpoint launches/past after v4/ to request comprehensive launch information
- Response in the form of a list of json objects assigned to variable called response
- Json objects to Pandas data frame using json_normalize
- Obtain information not available in the launches/past endpoint and add to data frame by making requests to other endpoints (/rockets, /launchpads, /payloads, /cores), appending lists and mapping into data frame columns
- Responses include data on: launch specifications, booster specifications, payload mass, orbit, landing outcomes, and reuse
- SpaceX API Notebook



# Fairing Data Collection – SpaceX API and Excel

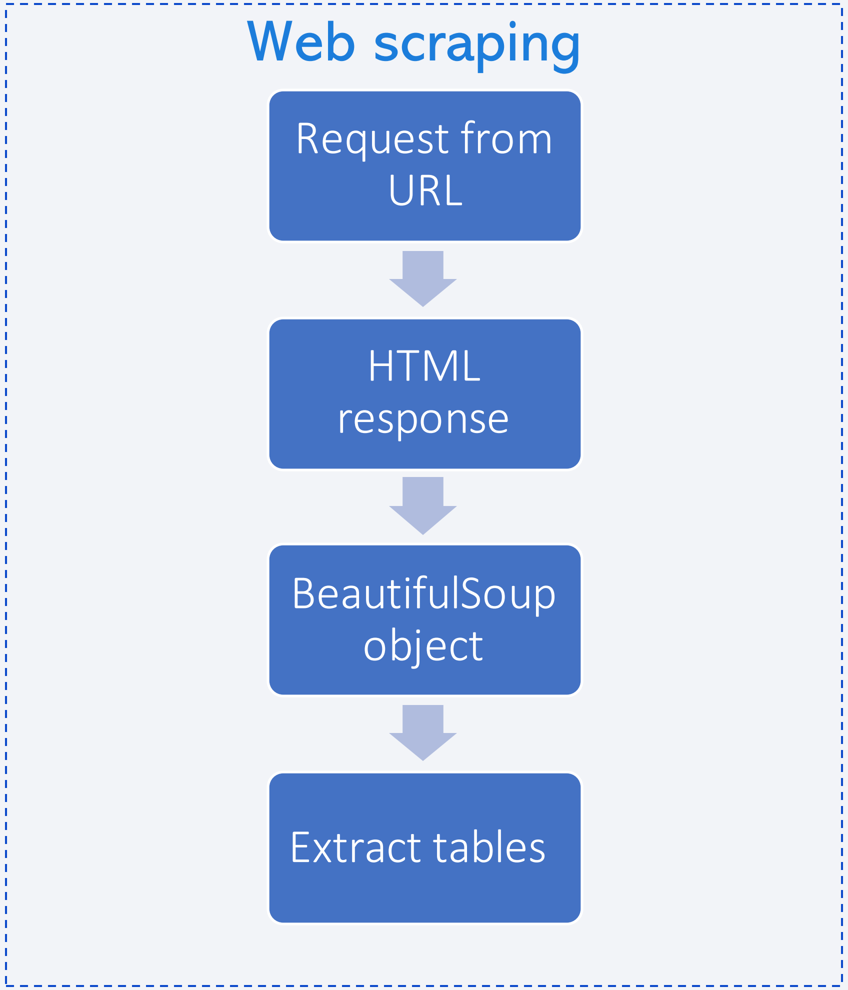
- Requests sent to the REST API URL [https://api.spacexdata.com/v4/](https://api.spacexdata.com/v4/)
- Endpoint launches/past after v4/ to request fairing reuse data
- Response in the form of a list of json objects assigned to variable called response
- Json objects to Pandas data frame using json_normalize
- Obtain missing/null data: identify and investigate SpaceX REST API source (r/spacex), identify and investigate r/spacex source (ElonX)
- Microsoft Excel: input missing values, merge based on flight number, date, and booster serial
- Fairing Data Collection Cleaning Notebook
- Fairing Null Handling Cleaning in Excel

## Fairing Data Collection and Wrangling



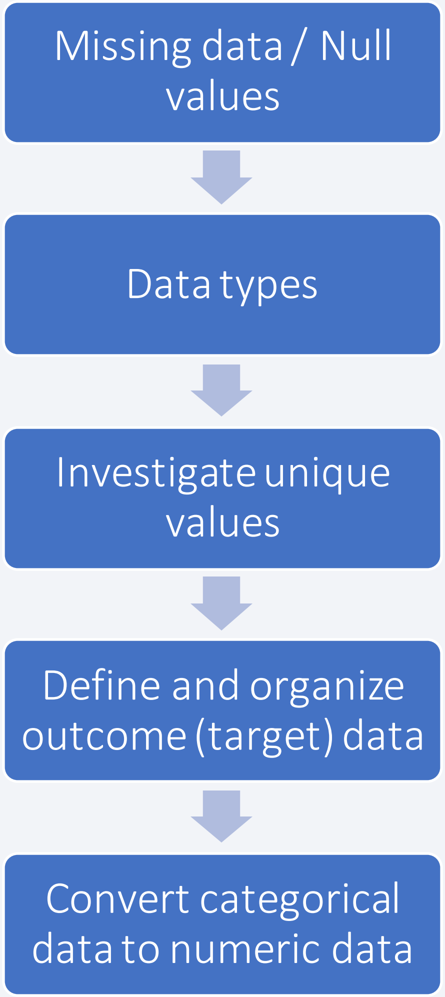
# Data Collection - Scraping

- Collect Falcon 9 historical launch records from https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches
- Make a get request to the URL for an HTML response
- Use BeautifulSoup to extract tables, rows and columns of desired data
- Create a dictionary of desired data and use it to create a Pandas data frame
- Response data: launch specifications, booster specifications, payload mass and contents, orbit, landing outcomes, some reuse information, and customers
- Web Scraping Notebook



# Data Wrangling

- Obtain an overview: looking at a small sections of the data frame for familiarization
- Missing data: .isnull().sum() and handle missing payloads by replacing with mean payload
- Data types: .dtypes to see data type by column, specifically which are categorical and numeric
- Unique values and number of occurrence by column: investigate using .value_counts()
- Target data: organize multiple unique categories into 2 that can be transformed into a binary numeric column by obtaining a list from the column and using a loop to append corresponding 0 or 1 in a new list which is added to the data frame as a new column
- [Data Wrangling Notebook]()



# EDA with Data Visualization

- Catplot of FlightNumber vs. PayloadMass with Outcome Overlay:
  - Looking at the landing outcomes of different payload masses overtime provided insight into both how success and payload mass changed overtime and into a relationship between payload mass and success
- Catplot of FlightNumber vs. LaunchSite with Outcome Overlay:
- To look deeper into prior discovery that launch sites had different success rates, looking at this information overtime provided insight into the reason for the differences in success rates by launch site
- Catplot of Payload vs. Launch Site with Outcome Overlay:
  - To look deeper even deeper into differing success rates by launch site, looking at the various payloads used by site provided further insight as to why success rates varied by launch site
- EDA and Visualization Notebook

# EDA with Data Visualization

- Barplot of Orbit vs. Success Rate:
  - Looking at landing outcomes by orbit along with flight counts by orbit showed that various orbits had higher success rates, which gained significance if that rate was out of a larger flight number
- Catplot of FlightNumber vs. Orbit with Outcome Overlay:
  - Looking at landing outcomes by orbit overtime provided another angle to check if time was also a factor in the differing success rates between mission orbit objectives
- Catplot of Payload vs. Orbit with Outcome overlay:
  - Looking at the success rate of various payloads going to the different orbits provided insight into the relationship between these two variables
- Lineplot of Success Rate by Year:
  - Visualizing success rate over time allowed observation of successful landing trends by showing how they have changed over time
- EDA and Visualization Notebook

# Supplemental EDA and Statistical Analysis

- Compare landing success distribution
  - KDE distribution graph of flight number by landing outcome
  - Histograms and analysis of flight subsets to investigate irregularities in distribution
- Compare landing success distribution by payload mass
  - KDE distribution graph and boxplot
  - Histograms and analysis of payload mass subsets to investigate irregularities in distribution
  - Compare payload mass subsets by landing outcome versus time, block, and orbit
- Compare landing success distribution by orbit
  - Histogram of orbits by landing outcome
  - Catplot of orbits by landing outcome versus time and block to investigate distribution irregularities
- Supplemental EDA and Statistical Analysis Notebook

# Supplemental EDA and Statistical Analysis

- Compare landing outcome distribution between new and reused cores
  - KDE graph to identify any irregularities
  - Countplot for further investigation
- Compare landing outcome distribution by blocks 1 through 5
  - KDE graph and countplot to identify any irregularities
  - Catplots to investigate irregularities over time and by the number of times the core was reused
- Compare landing outcome distribution by presence of grid fins and legs upon launch
  - Histogram to identify unexpected patterns
  - Subset analysis versus categorical outcome to investigate further
- Supplemental EDA and Statistical Analysis Notebook

# Supplemental EDA and Statistical Analysis

- Investigate relationships between features
  - Use Pandas corr() method to calculate pairwise correlation between features
  - Heatmap to visualize pearson correlation coefficients
- Delve deeper into specific orbits and blocks
  - Utilize functools.reduce to obtain a data frame with addition columns for specific orbits and blocks
  - Compare relationship between these additional features and landing outcomes using corr() and heatmap
- Preview redundant data
  - Heatmap showing features highly correlated with one another

- Supplemental EDA and Statistical Analysis Notebook

# Contract EDA and Financial Calculations

- Analyze known revenue of SpaceX missions to obtain estimates
  - Obtain known contract amounts, calculate per launch amounts, add per launch revenue via Excel
  - Distplot to visualize per launch revenue distribution
  - Histograms, countplots, and jointplot to visualize distribution of customers and missions, and compare distributions patterns to per launch revenue
- Build estimates of Revenue, Cost, and Profit
  - Use known costs to obtain unknown revenue, per kg revenue by launch
  - Use jointplot and boxplot by customer to obtain revenue per kg
  - Use known costs and known revenue to calculate SpaceX allocation of revenue to cost (expenses)
- Contract EDA and Financial Calculations Notebook

# EDA with SQL: Python Queries to DB2 Database

- Prepare
  - Python Jupyter Notebook: use ipython-sql extension to utilize SQL "Magic," connect to DB2 using credentials (ibm_db_sa://my-username:my-password@hostname:port/BLUDB?security=SSL) in first query
  - IBM DB2 Database Table: store the SPACEX dataset as a table in the database by uploading the .CSV file, formatting the schema, and creating a table in DB2
- SQL Queries
  - EDA first steps: Overview of the SPACEX dataset, Launch sites: Unique sites
  - Date: Monumental date of the first successful landing , Months in 2015 and boosters with failed landings to a drone ship , Number of successful landings between 04-06-2010 and 20-03-2017 (descending order, mut point as you will later see)
  - Payload mass and Boosters: Total payload mass by a specific customer and average payload mass by a specific booster , Boosters and payload mass for flights with payloads 4000-6000 kg and successful landings , Boosters and payload mass of the heaviest flight
  - Mission outcomes: Number of occurrences of successes and failures to reach or not reach the customer defined mission
  - Landing outcomes: Assigning a landing outcome frequency distribution query result to a variable followed by a data frame for visualization
- EDA with SQL Notebook

# Build an Interactive Map with Folium

- Circles with Popup: red circles to show the location of SpaceX launch sites and NASA and a popup to display the name of the site
- Markers: to display a readily readable shortened name to identify each site, to display the distance from a launch site to highway/railway/city/coastline
- MarkerCluster: to display a large number of points representing launch outcomes by site without cluttering the map
- MousePosition: to reveal the coordinates in the top right corner of the map corresponding to where your cursor is over the map, further allowing for proximity calculations
- PolyLine: to display a line from the launch site to nearest highway/railway/city/coastline
- Folium Interactive Maps

# Plotly Dash Dashboard Components

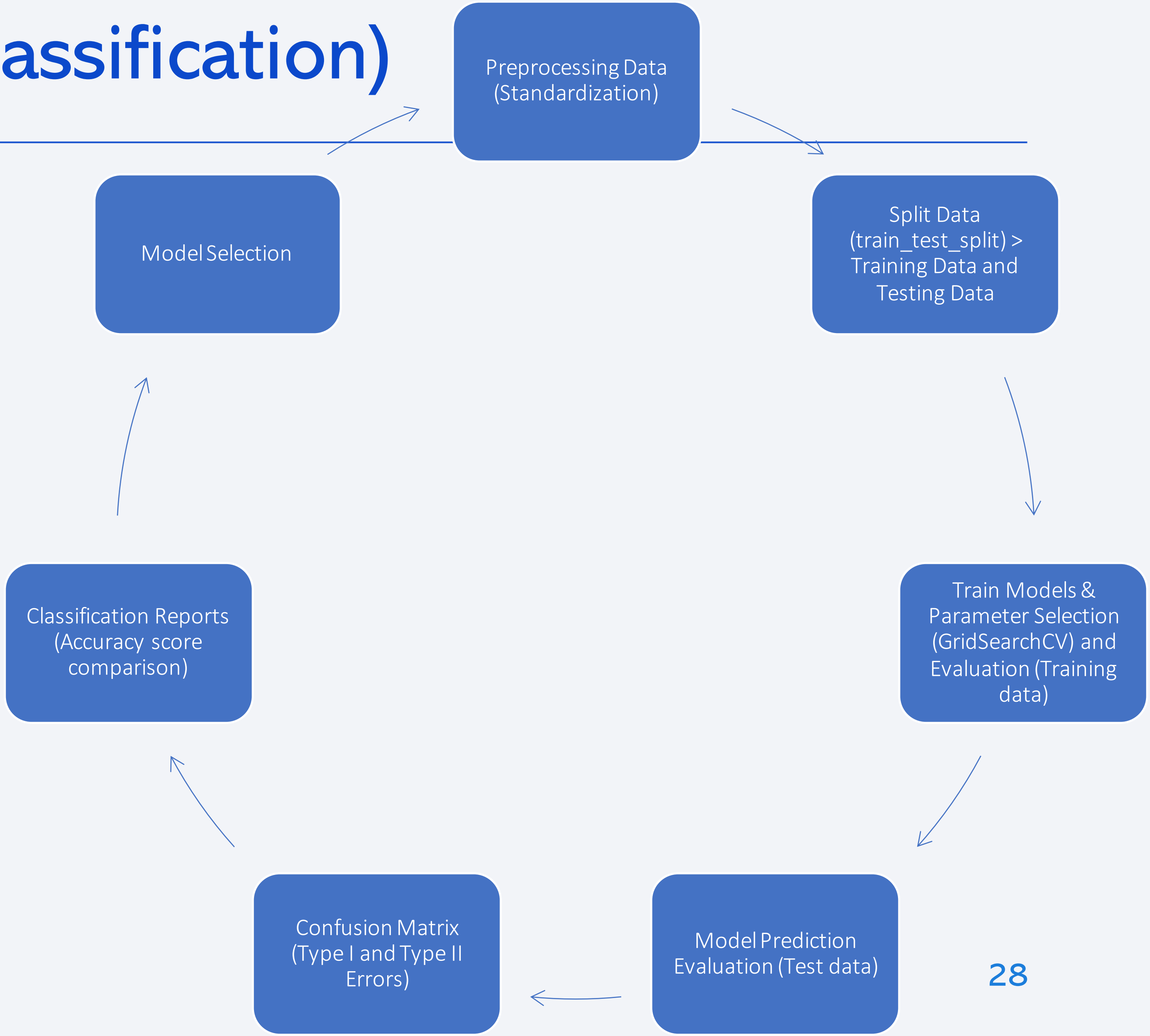
- Dropdown: To enable launch site selection, thus enabling data and visual selection. The dropdown interacts with both the pie and scatter charts to hone in on observed trends.
- Pie Chart: For all sites, the chart displays successful launch count by site and the sites' percentage of all successes. When the dropdown is used to select a specific site, the chart shows success/failure launch count and the corresponding percentage of all launches at that site.
- Scatter Plot (Scatter Chart): To visualize the relationship between payload mass and launch outcome with payload mass on the x-axis and class of 0 for launch failure or 1 for launch success on the y-axis. A booster version overlay allows filtering displayed data by booster version and visualization of the relationship between booster versions and launch outcomes.
- Slider: To select payload ranges displayed on the scatter plot to visualize launch outcomes within specified payload ranges.
- Plotly Dash PDF
- Dash Python Source Code

# Tableau Dashboard Components

- Relative Date Filter:
  - To enable choosing a date period for the dashboard display, date period being the filter choice of month and year versus filter choice of number of months before that date
- KPI over time: To compare Cost/Profit/Revenue and Reuse Quantity for a given period
  - Indicators: visualize the change of the KPI over the period compared to the start of the period
  - Area charts: display the trend of that KPI over the period displaying the KPI at period start and period end
- Bar graph by "Type":
  - Type filter to choose which bar graph, cost/profit/revenue or reuse quantity displayed over time with KPI amount per bar (month and year)
- Detailed Table:
  - A drilldown table to enable KPI breakdown into its components based on the chosen filters
- Other Filters
  - To enable further investigation
- Space X to Y Competitive Mission Bidding Dashboard

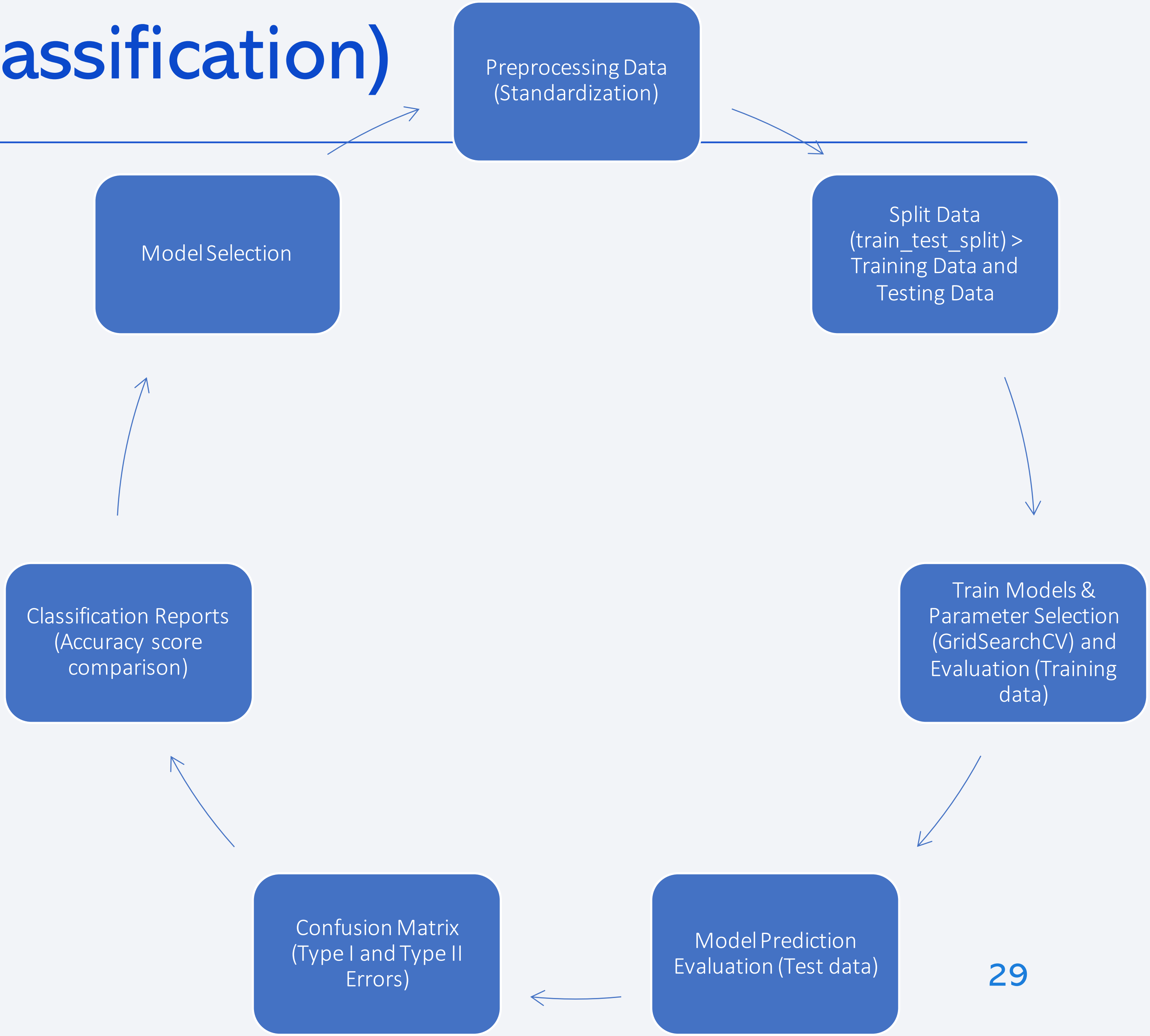
# Predictive Analysis (Classification)

- Classification Models: Logistic Regression, Support Vector machines, Decision Tree Classifier, and K-nearest neighbors
- Preprocessing: Separate feature inputs (X) and target outputs (Y), use StandardScaler to standardize/normalize input variables with different units of measure to $\mu$ = 0 and $\sigma$ = 1
- Split Data: Use train_test_split to obtain training and testing subset of data, the former to build the model, and the later to assess model classification performance (80%/20% split)
- Parameter Selection: Create a model object of each type, use GridSearchCV to create a grid search object of each type providing a dictionary of parameter options, fit and train the grid search objects on training data, obtain the best parameters corresponding to the best accuracy score of the algorithm can obtain based on training data
- …
- Predictive Analysis Notebook



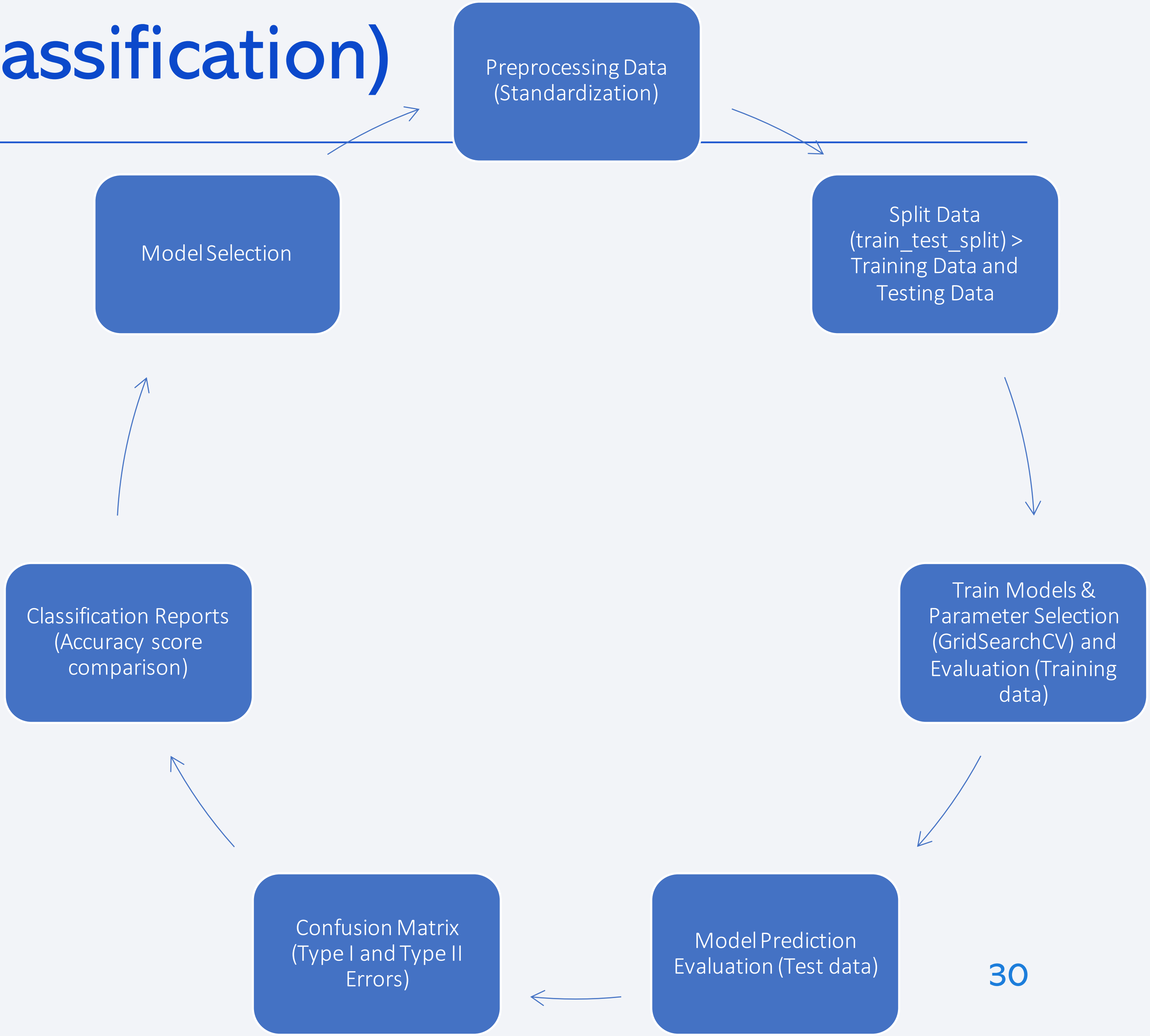
# Predictive Analysis (Classification)

- …
- Model Prediction Evaluation (Test Data): Expose the model to data it has not seen (test data). Use a confusion matrix to assess true positive, true negative, false positive, and false negatives. Use classification reports to compare accuracy scores (weighted average F1 score to take into account per class performance and number of occurrences)
- Model Selection: Select the combination of best parameters and classification algorithm that provides the best classification, the best false positive and false negative rate, along with the best weighted average F1 score.
- Predictive Analysis Notebook



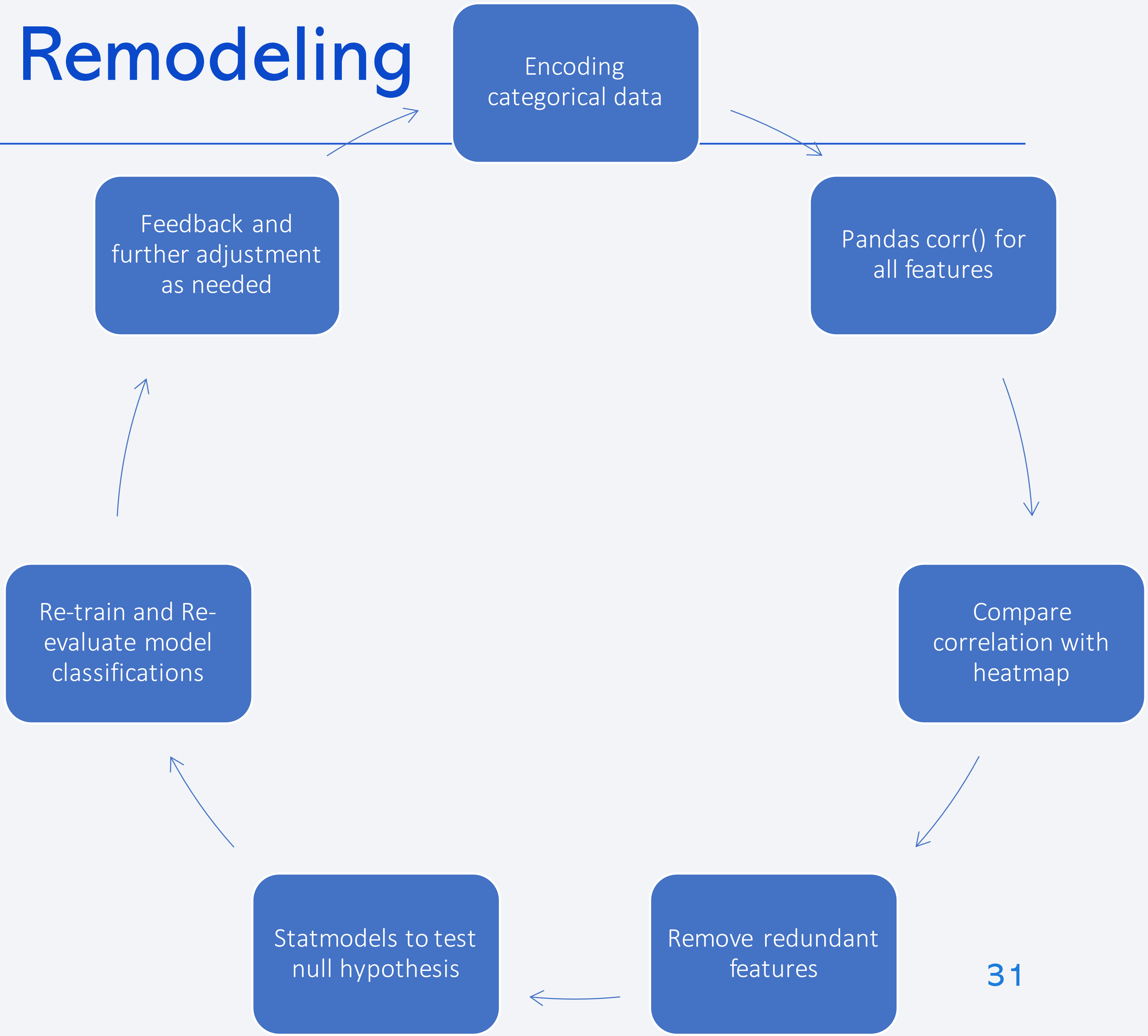
# Predictive Analysis (Classification)

- …
- Model Prediction Evaluation (Test Data): Expose the model to data it has not seen (test data). Use a confusion matrix to assess true positive, true negative, false positive, and false negatives. Use classification reports to compare accuracy scores (weighted average F1 score to take into account per class performance and number of occurrences)
- Model Selection: Select the combination of best parameters and classification algorithm that provides the best classification, the best false positive and false negative rate, along with the best weighted average F1 score.
- Predictive Analysis Notebook



# Feature Selection and Remodeling

- An iterative process…
- Feature selection to improve model:
    - Using the pairwise correlation between variables to remove redundant features highly correlated with one another to remove noise/overfitting from the algorithm and improve its classification abilities
- Using statmodels to test the null hypothesis
    - H0 = the selected combination of dependent variables do not have any effect on the independent variable
- Model re-evalutation: Confusion matrix and Classification report
- Feature Selection and Remodeling Notebook
- New Model Evaluation Notebook



# Results: Space X to Y

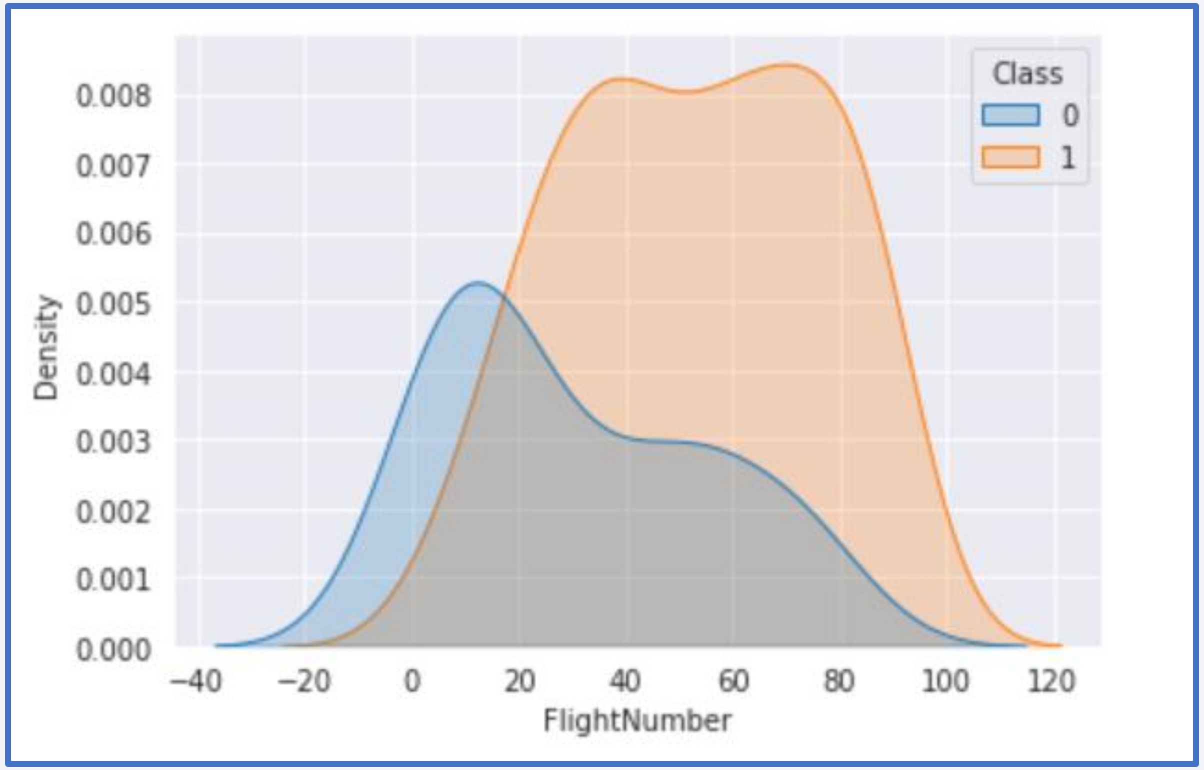
- The following results will be presented:
  - Exploratory data analysis results
  - Interactive analytics demo in screenshots
  - Launch cost calculations and Tableau Dashboard
  - Predictive analysis results
  - Model improvement results

Section 2

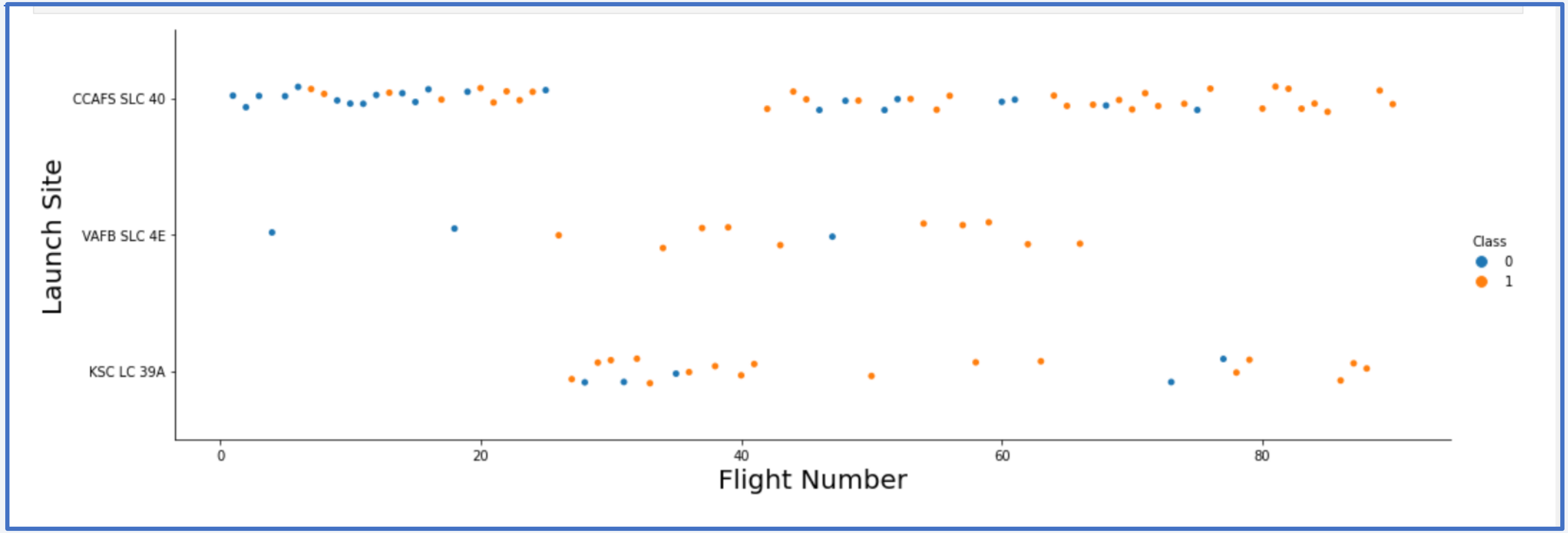
# Insights drawn from EDA

# Flight Number Distribution by Landing Outcome

- **Landing failure:**
- Distribution of failed landings peaks just before flight 20 (4/8/2016) and steadily decreases as flight number increases
- **Landing success:**
- Distribution of landing success begins to predominate just before flight 20 and begins to peak near flight 40. This continues to through flight 90, which is throughout the dataset
- *There is a trough is landing success between flights 40 and 60 (between 10/11/2017 and 12/5/2018)*
- *What is the nature of this trough? (see appendix)*

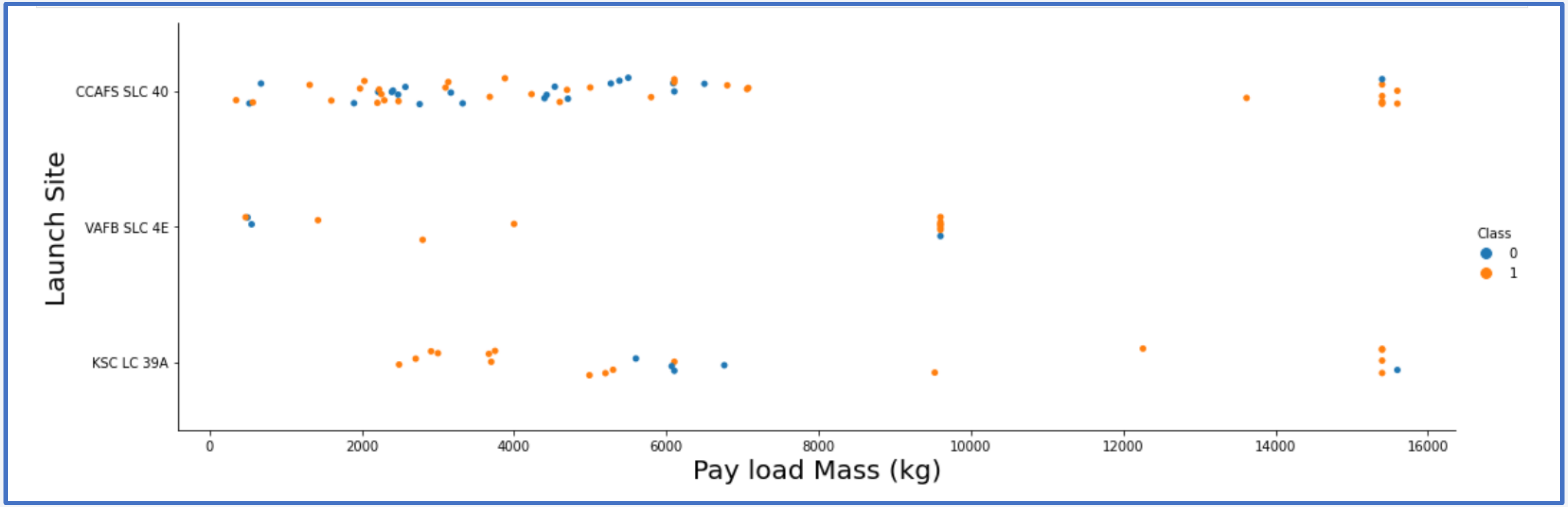


# Flight Number vs. Launch Site




- The launch site with the highest success rate is KSC LC-39A at 77.3% followed by VAFB SLC-4E at 76.9% and CCAFS SLC-40 at 60%. Success by site appears to improve over time, *as SpaceX moved beyond early testing (see appendix)*

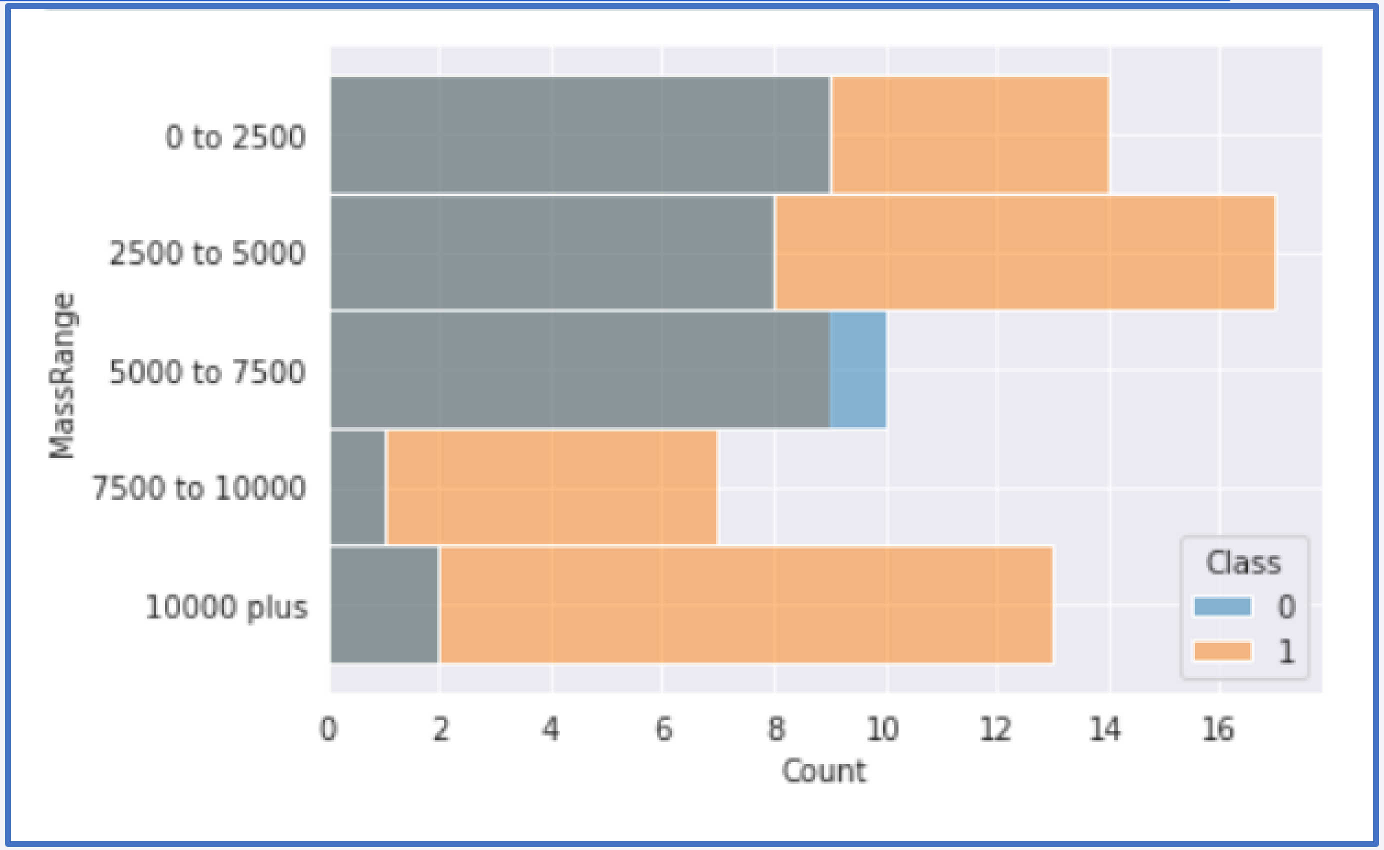
# Payload vs. Launch Site




- Payload masses > 10,000 kg did not launch from VAFB-SLC-4E
  - 9/15 launched from CCAFS SLC-40
  - 6/15 launched from KSC LC-39A

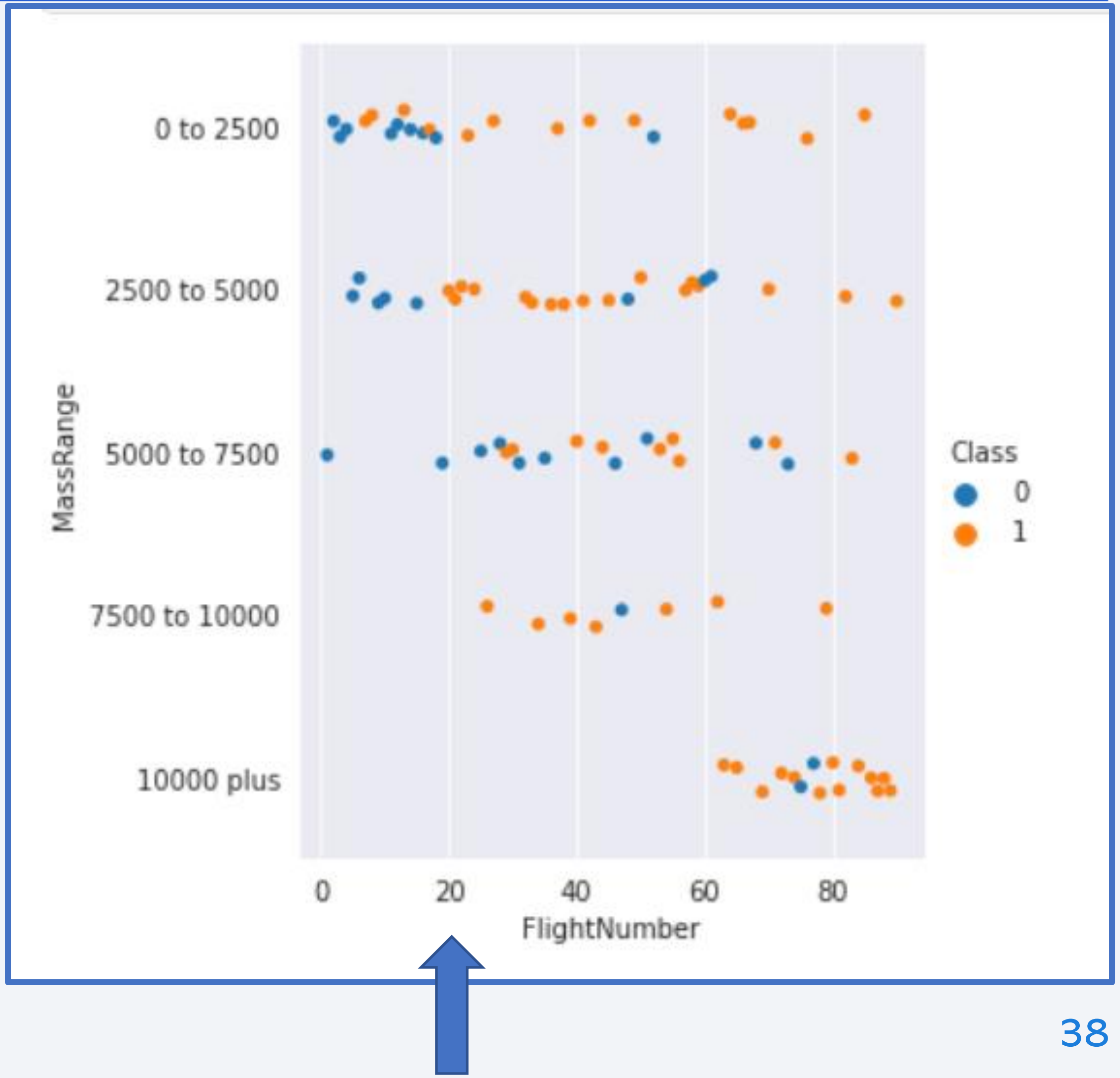
# Payload Mass Distribution by Landing Outcome

- 0 – 2,500 kg: Low success to failure ratio
- 2,500 – 5,000 kg: Most successful landings & Low success to failure ratio
- 5,000 – 7,500 kg: Most failures of all ranges & More failures than successes
- 7,500 – 10,000 kg: High success to failure ratio
- 10,000 kg plus: High success to failure ratio



# Payload Mass Distribution Over Time

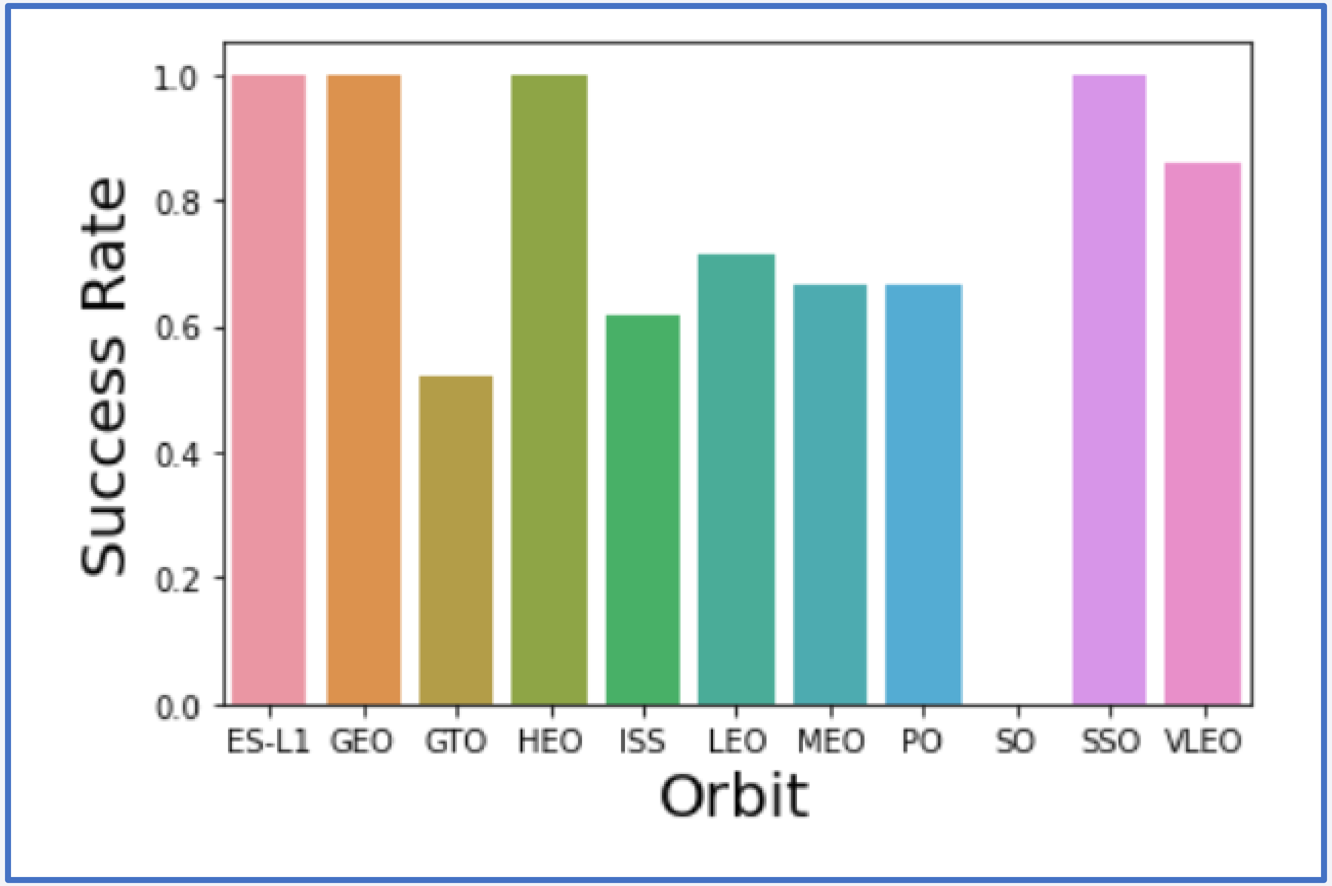
- 0 – 2,500 kg: many flights < flight 20
- 2,500 – 5,000 kg: many flights < flight 20
- 5,000 – 7,500 kg:
  - 2 failures < flight 20
  - failures sprinkled across all flight numbers
- 7,500 – 10,000 kg: flights all >20
- 10,000 kg plus: flights all > 60
- Before 4/8/2016 (flight 20): payloads ranged from 0 to 7,500 kg, and failures were more common than in later flights




*(for more see appendix)*

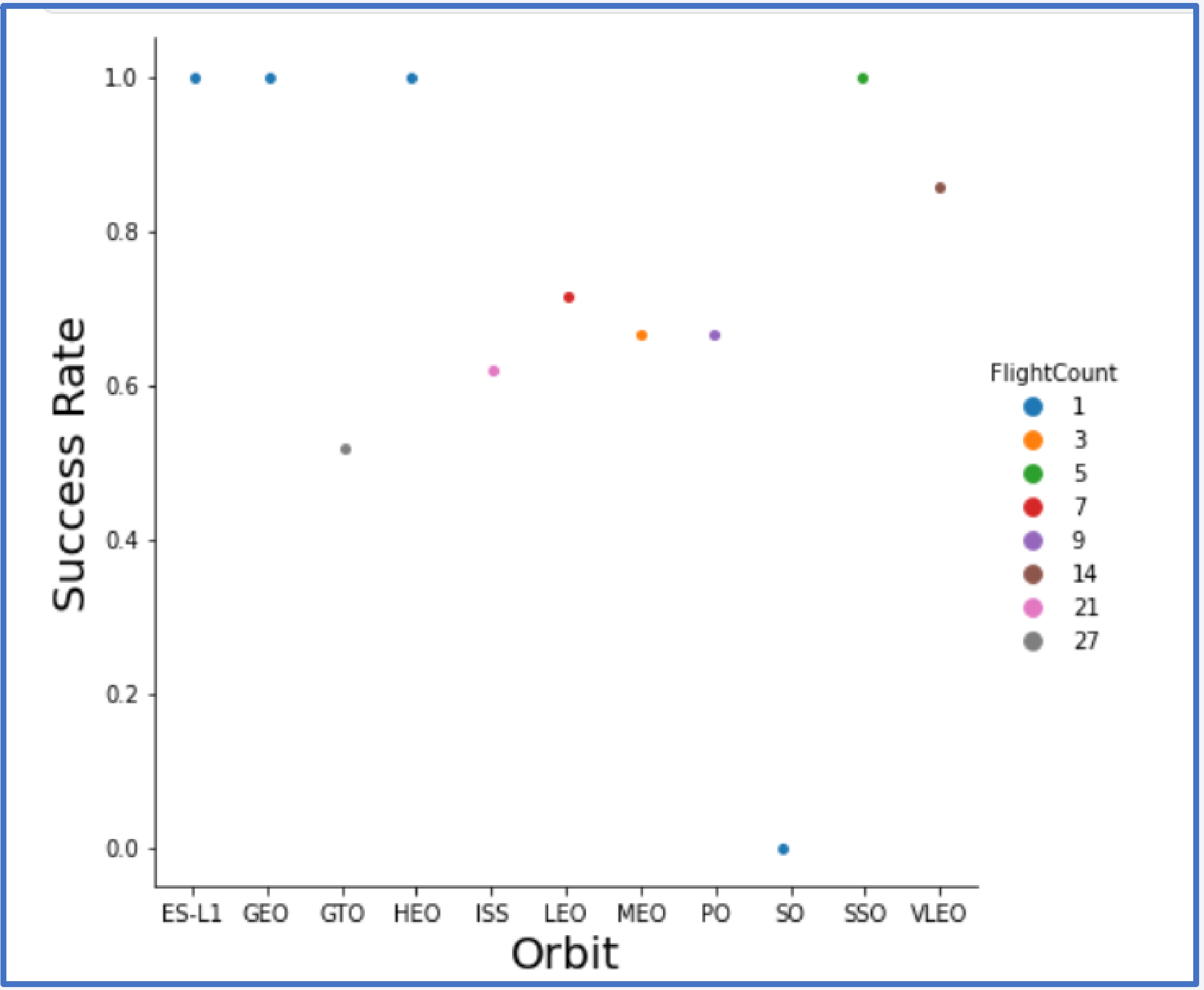
# Success Rate vs. Orbit Type

- 100% success rates for:
  - ES-L1
  - GEO
  - HEO
  - SSO
- Other orbit success rates
  - VLEO 86%
  - LEO 71%
  - MEO and PO 67%
  - ISS 62%
  - GTO 52%

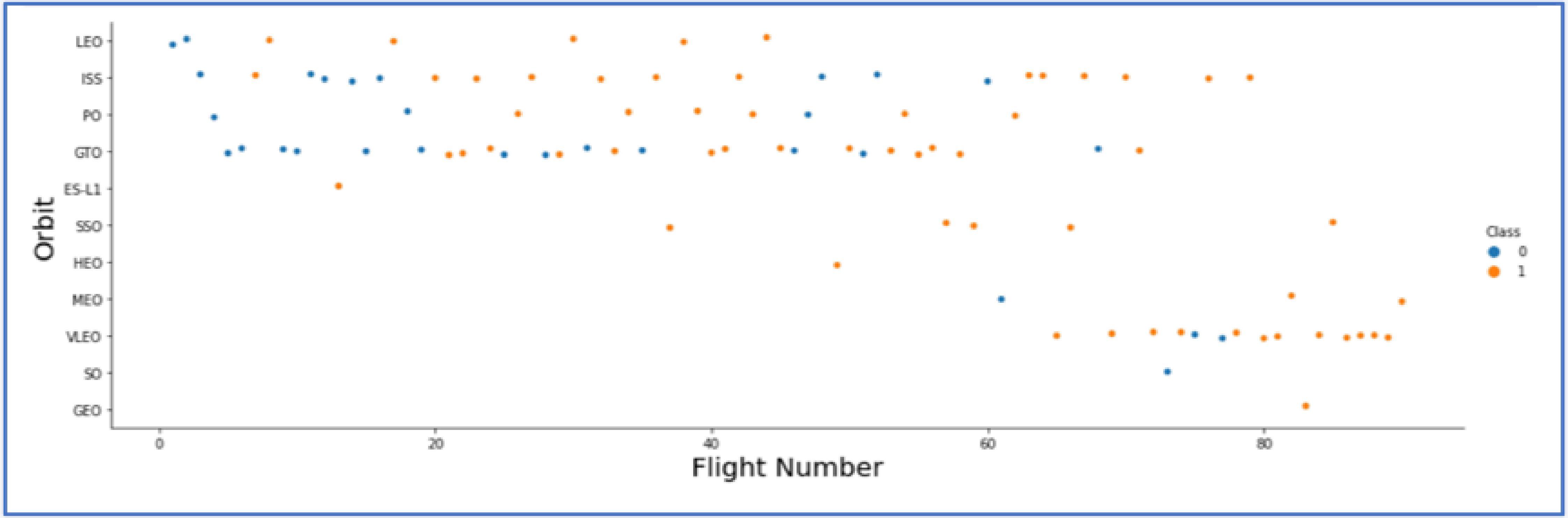


# Success Rate vs. Orbit Type

- 100% success rates for:
  - ES-L1, GEO, HEO: out of 1
  - SSO: out of 5
- Other orbit success rates
  - VLEO 86% out of 14
  - LEO 71% out of 7
  - MEO 67% out of 3
  - PO 67% out of 9
  - ISS 62% out of 21
  - GTO 52% out of 27

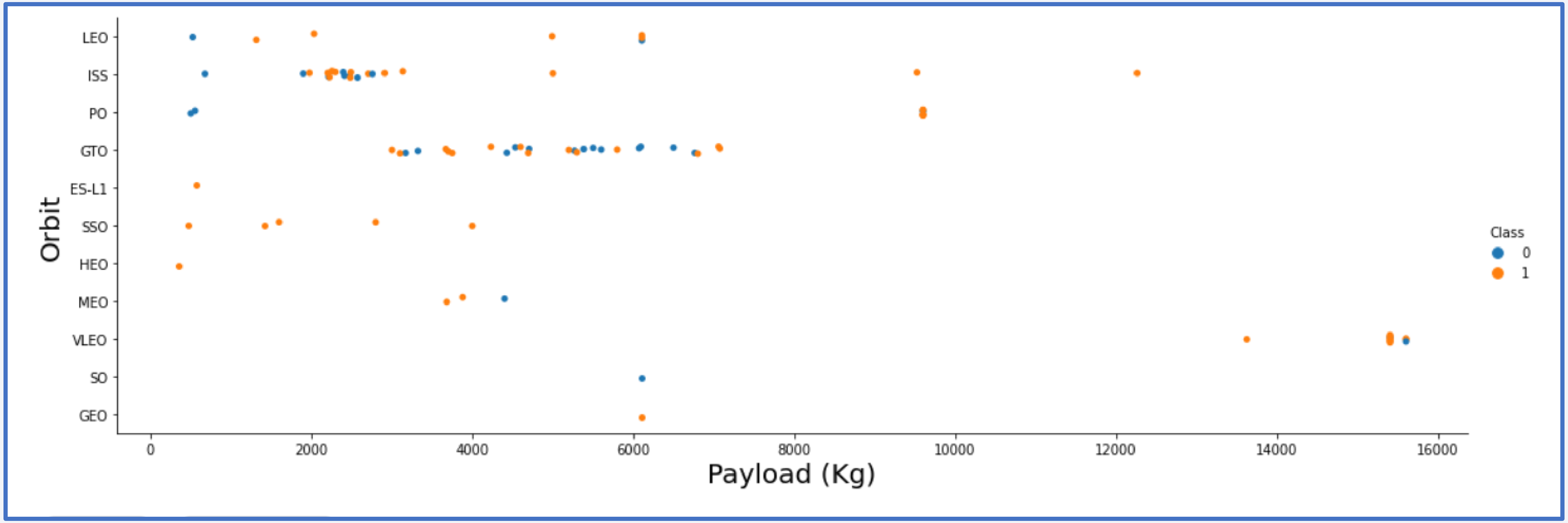


# Flight Number vs. Orbit Type



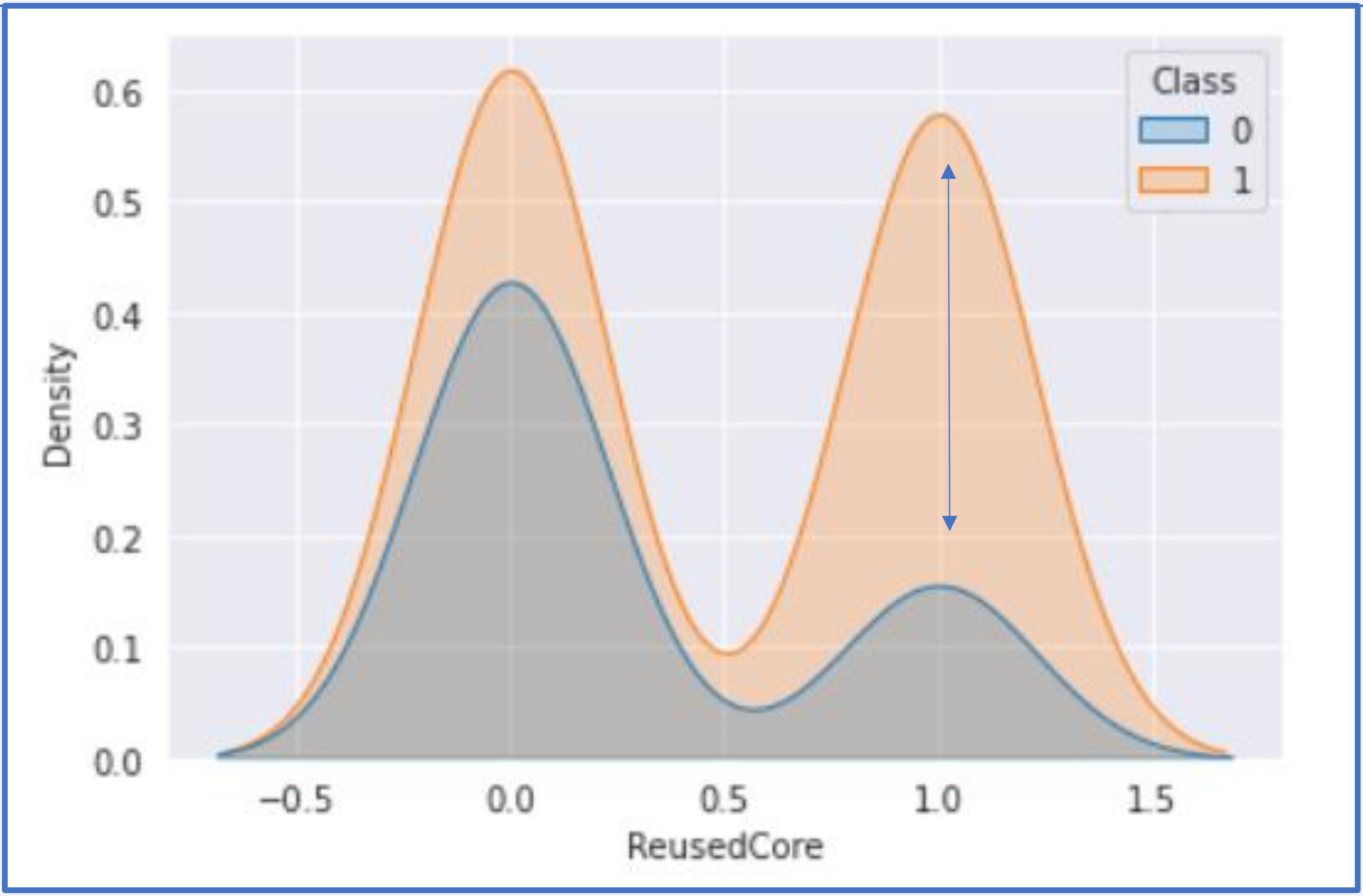

- Success appears to increase over time (success rate): LEO 71%, MEO 67%
- Success does not appear to be related to time: VLEO 86%, PO 67% , ISS 62%, GTO 52%

# Payload vs. Orbit Type



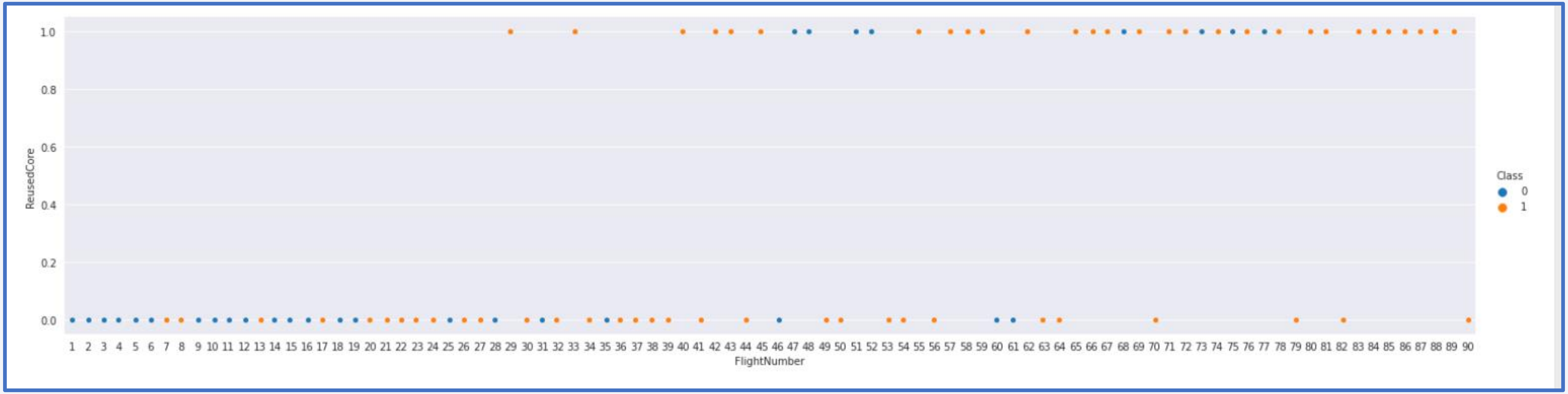

- Looking at payloads >8,000 kg, successful landings appear to be most common for ISS and PO. VLEO landing success appears to also be high.
- Looking at payloads between 5,000 kg and 8,000 kg, GTO the likelihood of landing success cannot be interpreted clearly (*See appendix for more)*

# Reused Core by Landing Outcome




- A reused core does not appear to detrimentally affect landing outcomes, successfully reused core distribution has a slightly higher peak for non-reused core; however, unsuccessfully reused core distribution has a much lower peak with core reuse

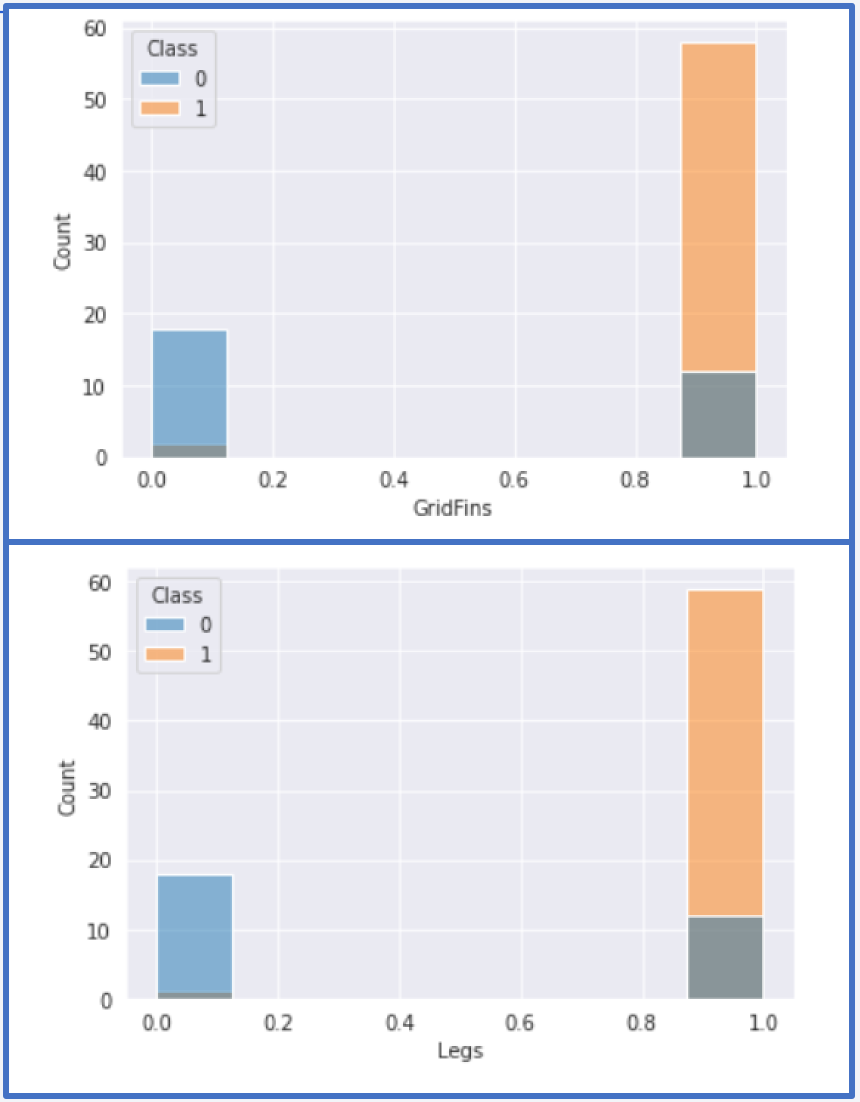
# Reused Core by Landing Outcome and Flight Number




- <u>Landing success seems to increase over time with new cores</u>
- Reused cores did not begin until Flight 29:
    - Success rates for reused core flights are 78.4%, while success rates for new cores are 58.5%
- When looking only at Flights >or= 29:
    - Success rates for reused cores remain 78.4%, and success rates for new cores increase to 80%

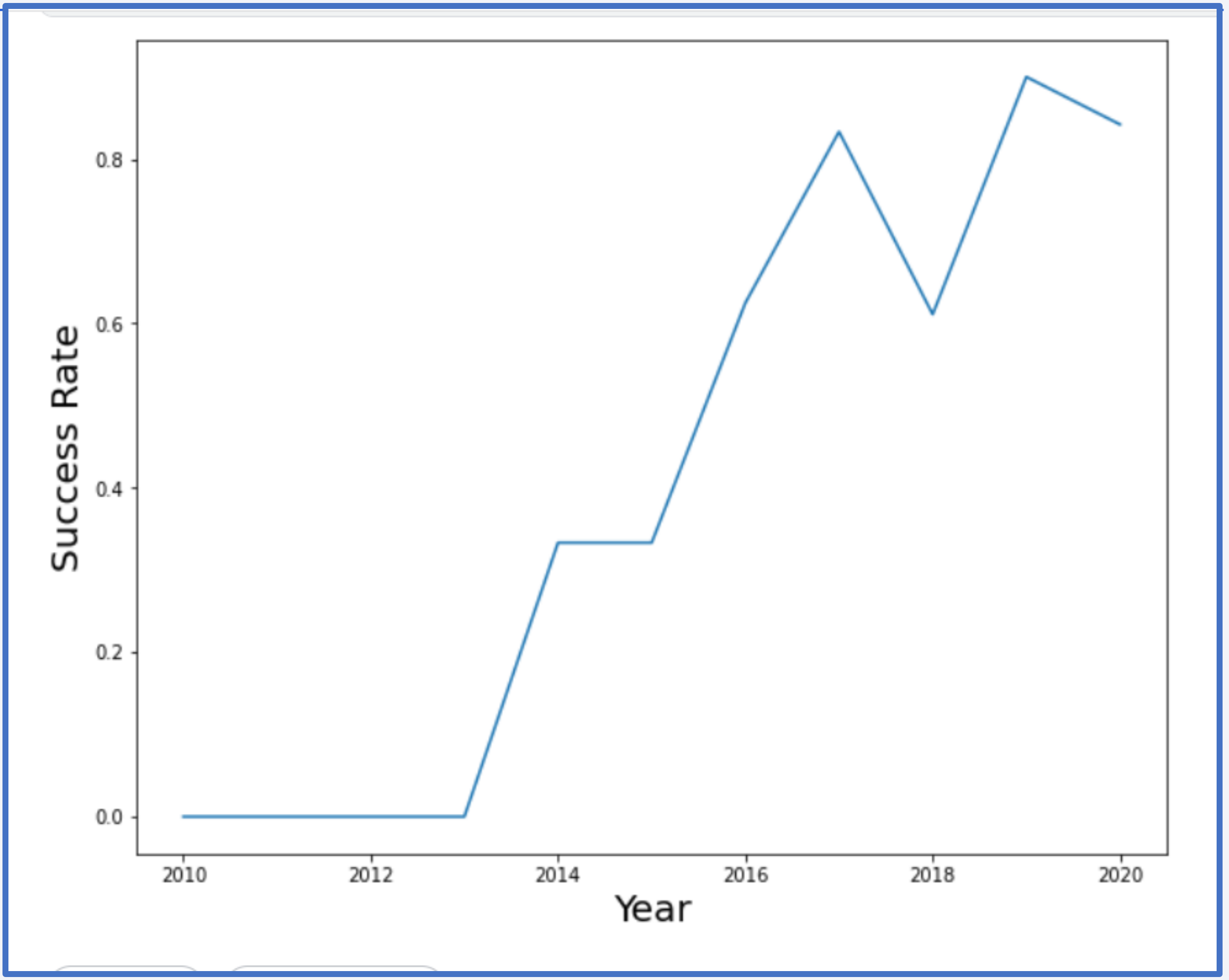
# Grid Fins & Legs by Landing Outcome

- Flights without Grid Fins / Legs
  - Landing Outcome:
    - False Ocean 2 / 2
    - None None 16 / 16
    - True Ocean 2 / 1
  - <u>Success rate 80%</u>
- The presence or absence of Grid Fins and Legs at launch have outcomes related to testing or intention to expend.



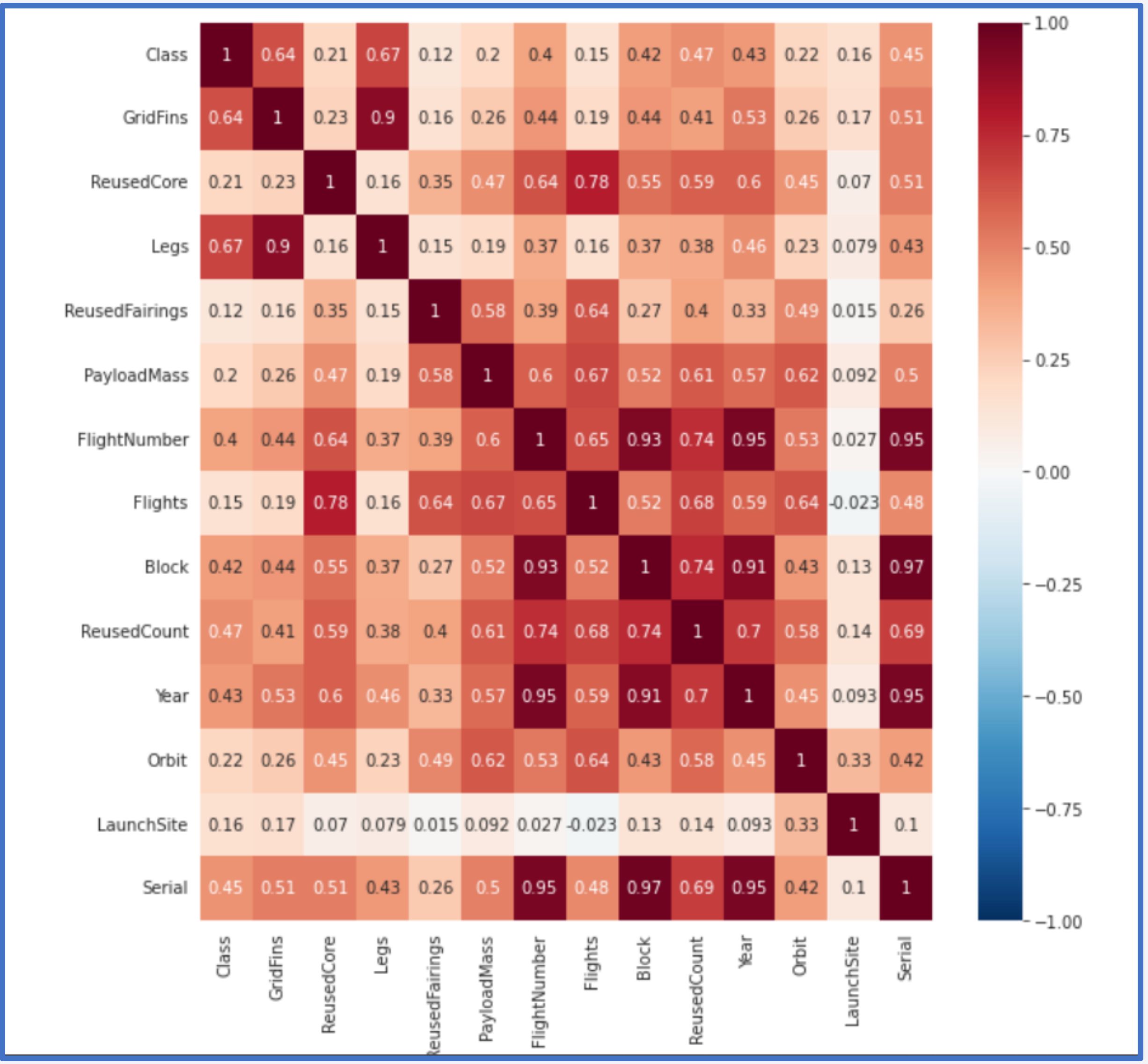
# Landing Success Yearly Trend

- <u>After 2013, the successful landing rate increases over time</u>
  - 0% in 2013 to 84.2% in 2020
- 2014 and 2015 had equal success rates
  - 33.3%
- Success rates declined between 2 years:
  - 83% in 2017 to 61% in 2018
  - 90% in 2019 to 84.2% in 2020



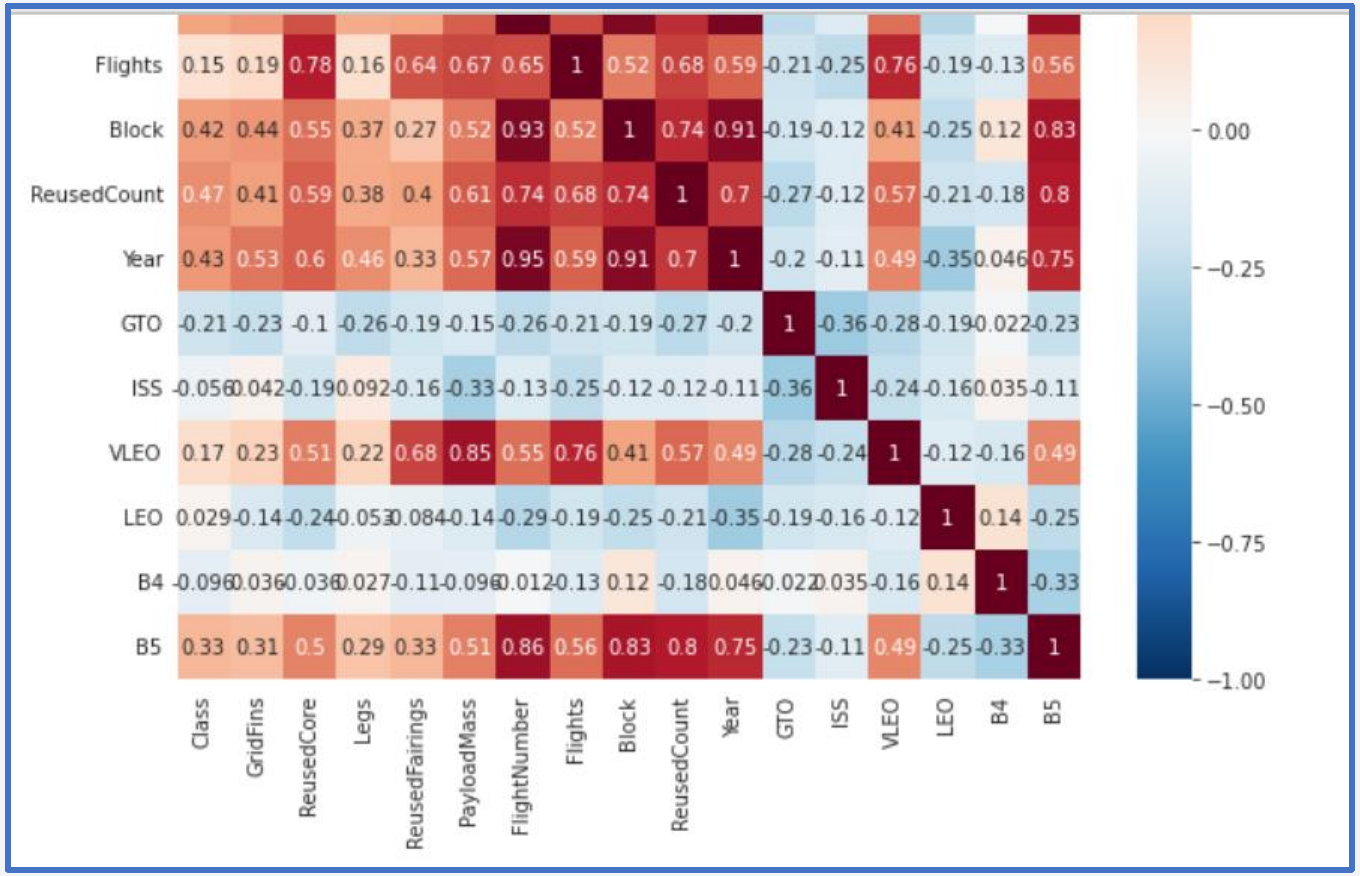
# Relationship between Features and Landing outcome

- Pearson correlation coefficients to evaluate the relationship between independent and dependent variables
- Strong Relationships vs. Class
    - None
- Moderate Relationships vs. Class
    - Legs and GridFins
- **Weak Relationships vs. Class**
    - **ReusedCore, PayloadMass, FlightNumber, Block, ReusedCount, Year, Orbit**
- Redundant Data
    - Year - FlightNumber - Block
    - Legs - GridFins



# Additional Exploration of Relationships between Features and Landing outcome

- Digging into the following values of features:
  - Orbits: GTO, ISS, VLEO, LEO
  - Blocks: 4 and 5
- Class of Landing Outcome
  - **GTO - weak negative**
  - ISS - no relationship
  - VLEO - no relationship
  - LEO - no relationship
  - B4 - no relationship
  - **B5 - weak positive**



# All Launch Site Names

- Find the names of the unique launch sites
  - %sql select distinct LAUNCH_SITE from SPACEX
- Query result
  - Results show 4 unique launch sites in the dataset
  - CCAFS LC-40 later became CCAFS SLC-40

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

# Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with \`CCA\`
    - %sql select LAUNCH_SITE from SPACEX where LAUNCH_SITE like 'CCA%' limit 5
- Query result
    - The first 5 records encountered in the dataset beginning with CCA were CCAFS SLC-40

| launch_site |
|---|
| CCAFS SLC-40 |
| CCAFS SLC-40 |
| CCAFS SLC-40 |
| CCAFS SLC-40 |
| CCAFS SLC-40 |

# Total Payload Mass

- Calculate the total payload carried by boosters from NASA
    - %sql select sum(PAYLOAD_MASS_KG) as Total_Payload_Mass_Kg_of_All_Falcon 9_Launches from SPACEX where CUSTOMER = 'NASA (CRS)'
- Query result
    - 45,596 kg is the mass of all payloads launched for NASA (CRS) within the dataset

| total_payload_mass_kg_of_all_falcon9_launches |
| --- |
| 45596 |

# Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1
  - %sql select avg(PAYLOAD_MASS_KG) as average_payload_mass_kg_for_F9_v1_1 from SPACEX where BOOSTER_VERSION = 'F9 v1.1'
- Query result
  - F9 v1.1 carried an average of 2,928 kg calculated from launches in the dataset

| average_payload_mass_kg_for_f9_v1_1 |
| ---: |
| 2928 |

# First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad
  - %sql select min(DATE) as "First Landing Success" from SPACEX where LANDING__OUTCOME = 'Success (ground pad)'
- Query result
  - On December 22, 2015 there was the 1st truly successful landing onto a ground pad
  - Flight number 20 landing outcome was True RTLS indicating a fully successful ground pad landing

| First Landing Success |
| --- |
| 2015-12-22 |

# Successful Drone Ship Landing with Payload between 4,000 and 6,000 kg

- List the names of boosters which have successfully landed on drone ship and had payload mass greater than 4,000 but less than 6,000 kg
  - %sql select distinct BOOSTER_VERSION, LANDING__OUTCOME, PAYLOAD_MASS_KG from SPACEX where (PAYLOAD_MASS_KG between 4000 and 6000) and LANDING__OUTCOME = 'Success (drone ship)'
- Query result
  - B1021, B1031, B1022, and B1026 all landed successfully on a drone ship after having launched payloads between 4,000 and 6,000 kg
  - All 4 of these where transported by F9 Full Thrust also known as F9 v1.2 with blocks 1-5

| booster_version | landing__outcome | payload_mass_kg |
|---|---|---|
| F9 FT B1021.2 | Success (drone ship) | 5300 |
| F9 FT B1031.2 | Success (drone ship) | 5200 |
| F9 FT B1022 | Success (drone ship) | 4696 |
| F9 FT B1026 | Success (drone ship) | 4600 |

# Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes
  - %sql select count(*), MISSION_OUTCOME as "Launch and/or Objective Success" from SPACEX group by MISSION_OUTCOME
- Query result
  - Frequency of customer defined mission outcome was analyzed
  - 99 clearly successful missions from launch to target occurred in the dataset, 1 failure, and 1 unclear outcome (Zuma Controversy)

| 1 | Launch and/or Objective Success |
|---|---|
| 1 | Failure (in flight) |
| 99 | Success |
| 1 | Success (payload status unclear) |

# Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass
    - %sql select distinct BOOSTER_VERSION, PAYLOAD_MASS_KG from SPACEX where PAYLOAD_MASS_KG = (select max(PAYLOAD_MASS_KG) from SPACEX)
- Query result
    - 15,600 kg was the heaviest payload mass in the dataset
    - B5 blocks transported the heaviest payloads

| booster_version | payload_mass_kg |
|---|---|
| F9 B5 B1048.4 | 15600 |
| F9 B5 B1048.5 | 15600 |
| F9 B5 B1049.4 | 15600 |
| F9 B5 B1049.5 | 15600 |
| F9 B5 B1049.7 | 15600 |
| F9 B5 B1051.3 | 15600 |
| F9 B5 B1051.4 | 15600 |
| F9 B5 B1051.6 | 15600 |
| F9 B5 B1056.4 | 15600 |
| F9 B5 B1058.3 | 15600 |
| F9 B5 B1060.2 | 15600 |
| F9 B5 B1060.3 | 15600 |

# 2015 Launch Records

- List the failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015
  - %sql select month(DATE) as month, year(DATE) as year, BOOSTER_VERSION, LAUNCH_SITE, LANDING__OUTCOME from SPACEX where LANDING__OUTCOME = 'Failure (drone ship)' and year(DATE) = '2015'
- Query result
  - In January 2015 and April 2015 there were failed attempts to land on a drone ship
  - Both were transported by block 1, also known as v1.1
  - Both launched from CCAFS LC-40

| MONTH | YEAR | booster_version | launch_site | landing__outcome |
| --- | --- | --- | --- | --- |
| 1 | 2015 | F9 v1.1 B1012 | CCAFS LC-40 | Failure (drone ship) |
| 4 | 2015 | F9 v1.1 B1015 | CCAFS LC-40 | Failure (drone ship) |

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

- 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

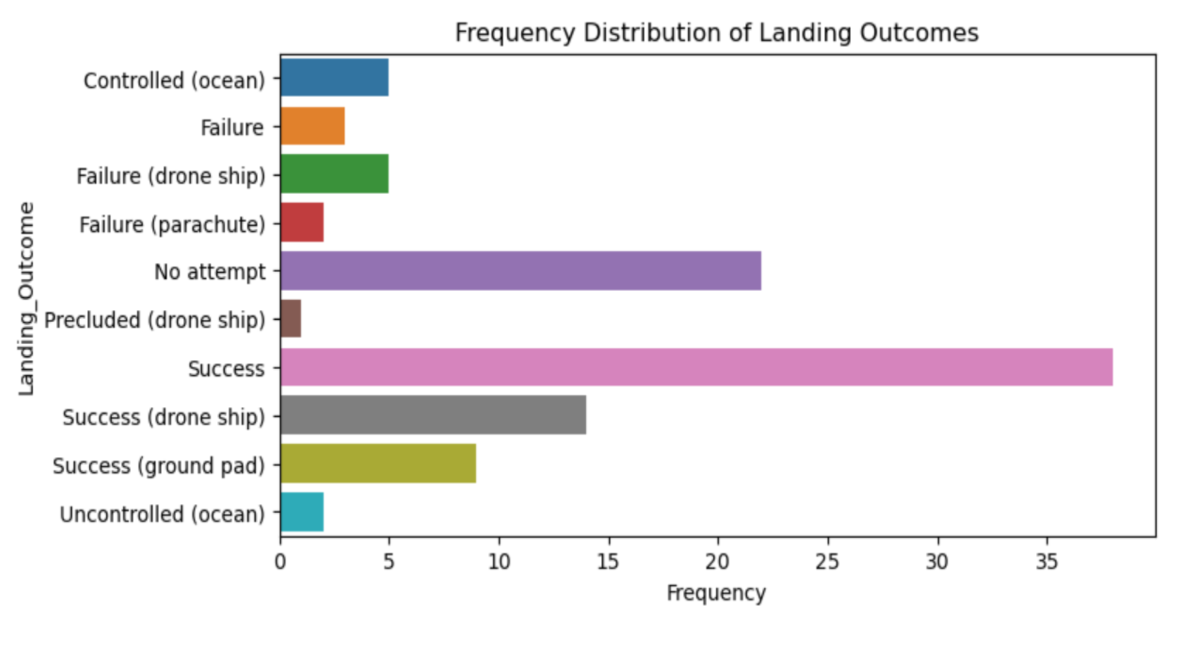
```
%%sql
    select count(LANDING__OUTCOME) as outcome_count, LANDING__OUTCOME
    from (
            select * from SPACEX where DATE between '2010-04-06' and '2017-03-20'
             )
    group by LANDING__OUTCOME
    having LANDING__OUTCOME like 'Success%'
    order by outcome_count DESC
```

- Query result
  - Between these dates there where 10 successful landings
  - 5 onto a drone ship and 5 onto a ground pad

| 1 | landing__outcome |
|---|---|
| 5 | Success (drone ship) |
| 5 | Success (ground pad) |

# Landing Outcome Distribution

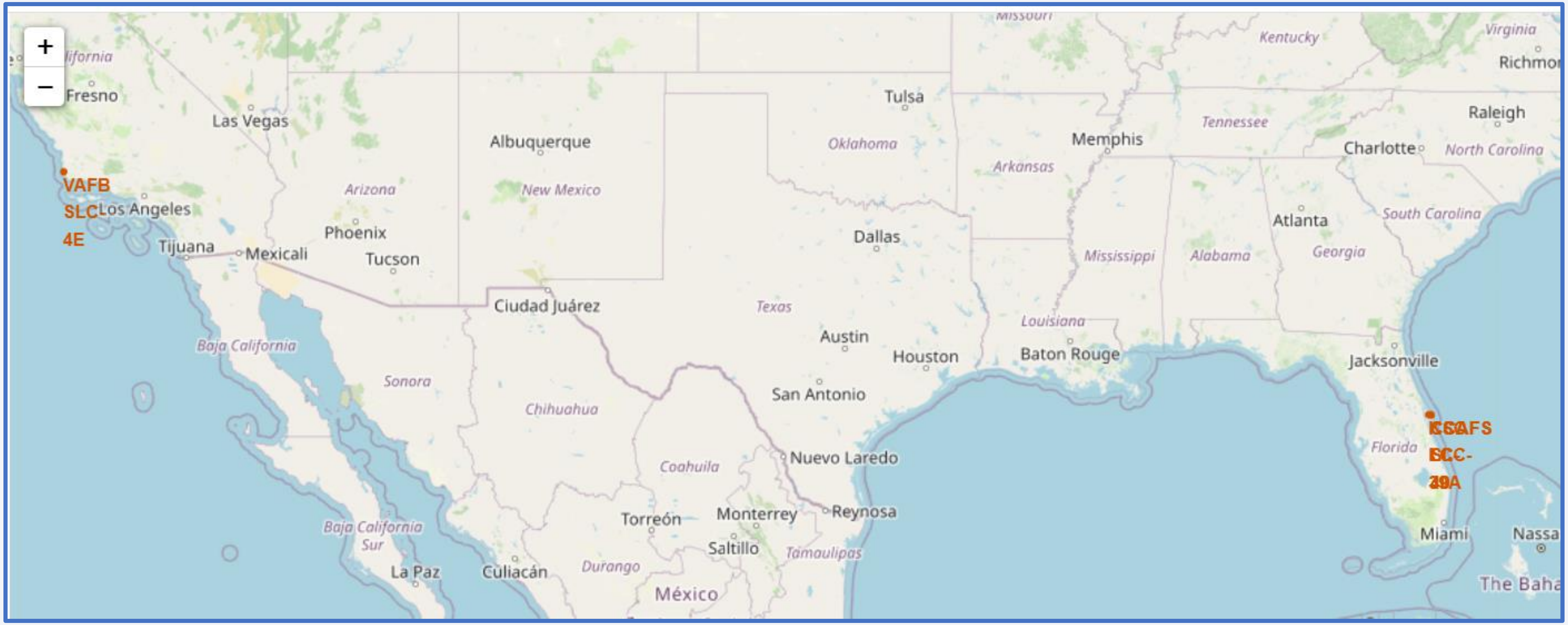
- Assigning query result to a variable followed by a data frame for visualization
  - landing_outcome_distribution = %sql select LANDING__OUTCOME as "Landing_Outcome", count(*) as "Frequency" from SPACEX group by LANDING__OUTCOME
- Query result
  - Most frequent landing outcome in the dataset is 'Success'
  - This is followed by 'No attempt', 'Success (drone ship)', then 'Success (ground pad)'



Section 3

# Launch Sites Proximities Analysis

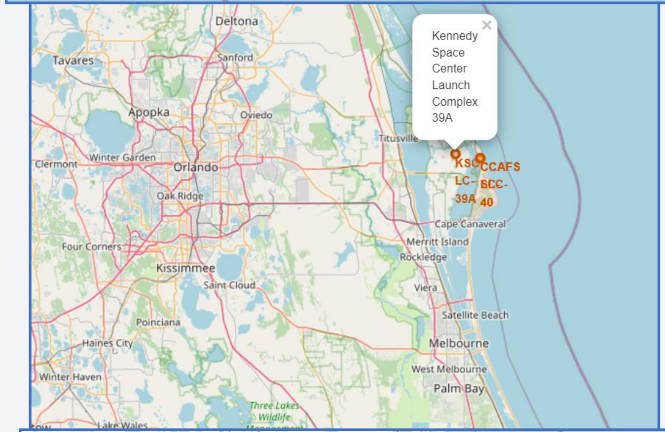
# Launch Site Folium Map

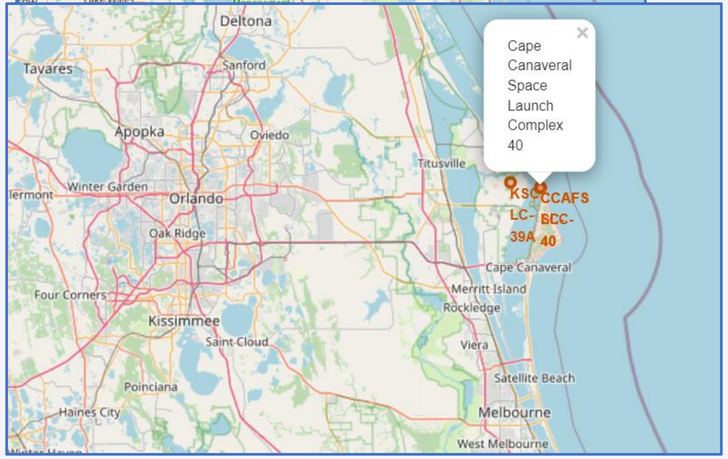



- This interactive Folium map displays the all SpaceX utilized launch sites with red circles, markers with abbreviated names and pop-ups that display the full name…

# Launch Site Folium Map: Zoom

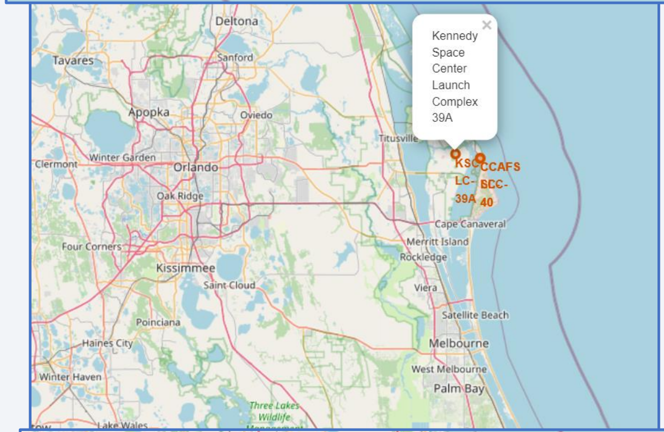
- Launch sites are in close proximity to the Equator line
  - Haversine Distances to the Equator:
  - 3175.99 km for CCAFS LC-40
  - 3176.09 km for CCAFS SLC-40
  - 3177.21 km for KSC LC-39A
  - 3851.00 km for VAFBS LC-4E
- Launch sites are also very close proximity to the coast
  - Haversine Distances to the coastlines:
  - 1.18 km for CCAFS LC-40 to FL Coastline
  - 1.07 km for CCAFS SLC-40 to FL Coastline
  - 7.50 km for KSC LC-39A to FL Coastline
  - 1.53 km for VAFBS LC-4E to CA Coastline

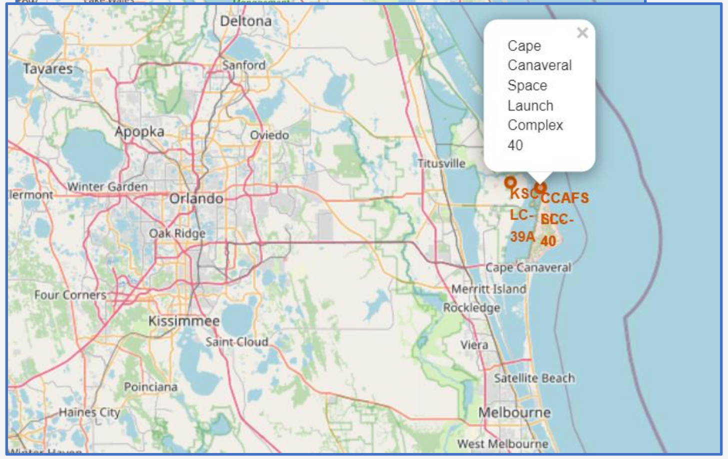



# Launch Site Folium Map: Zoom

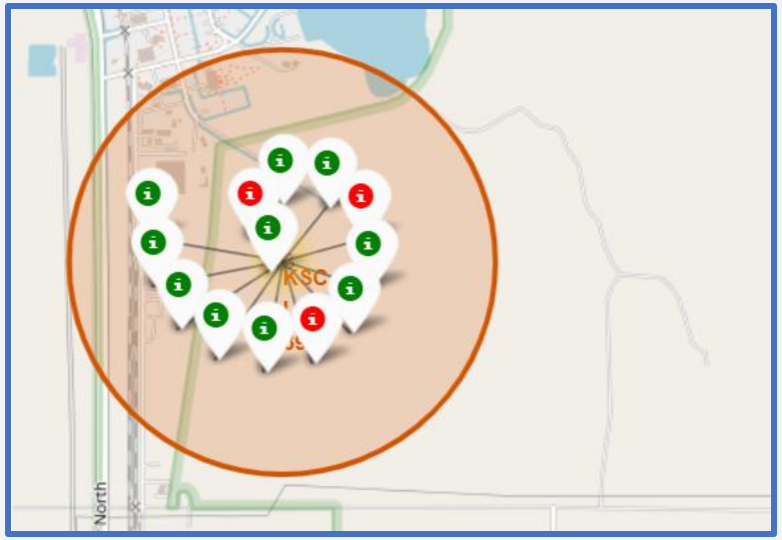
- Equator proximity:
  - Maximum rotational speed of lies on the Equator (average equatorial speed: 1,040 mph or 1,674 kmph); this saves reduces launch cost by decreasing need for extra fuel and boosters.
- Coast proximity:
  - In case of launch failure, coastal proximity mitigates risk from burning debris.
- Eastern coast launch sites:
  - The rotational direction is west to east, thus launching eastward uses rotational speed and direction to reduce launch cost. An eastward traveling launch failure would be falling into the ocean.

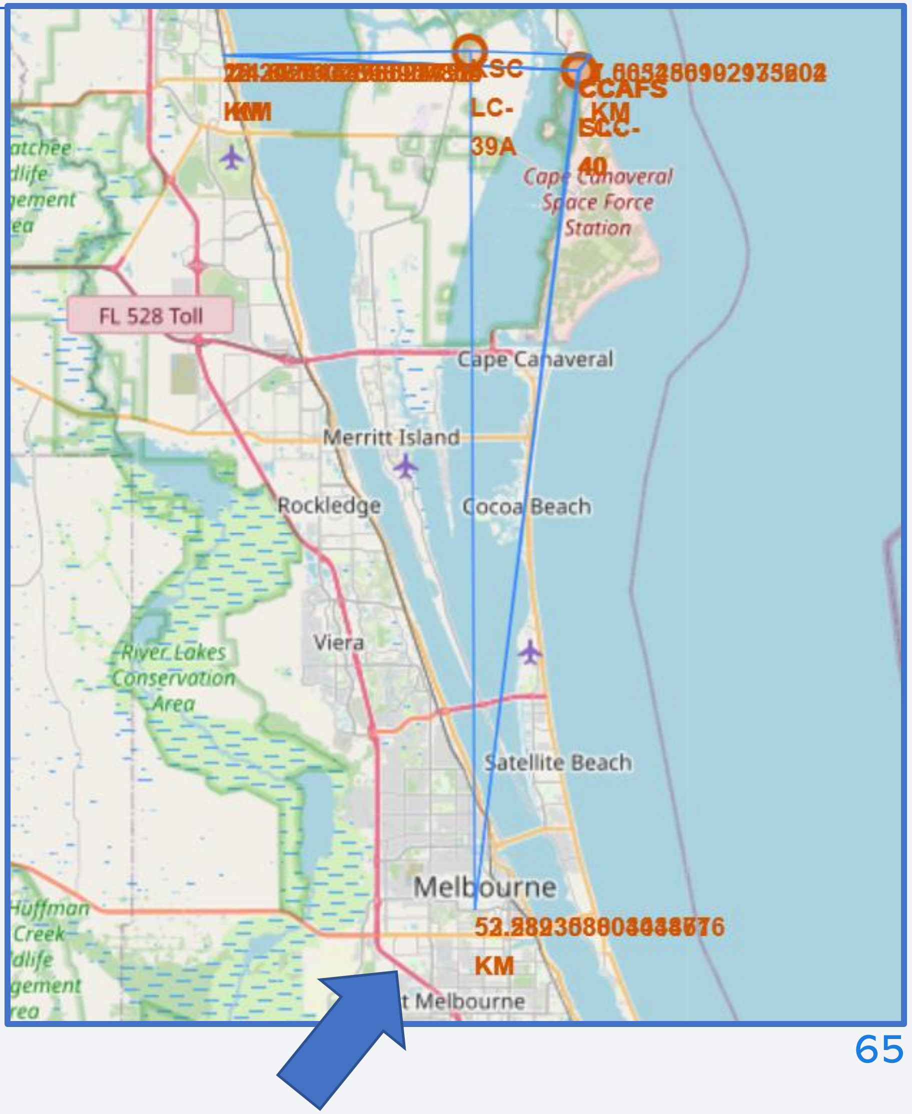
# Launch Site Folium Map: Launch Outcome Clusters

- Green markers indicate successful launches
- Red markers indicate launch failures
- Yellow circles display a launch count at each site
- Clicking on the yellow circle reveals unique launch instances and outcomes
- KSC LC-39A (Kennedy Space Center) clusters easily show the launch site's high launch success rate ~ 77%

# Launch Site Folium Map: Proximity to Cities

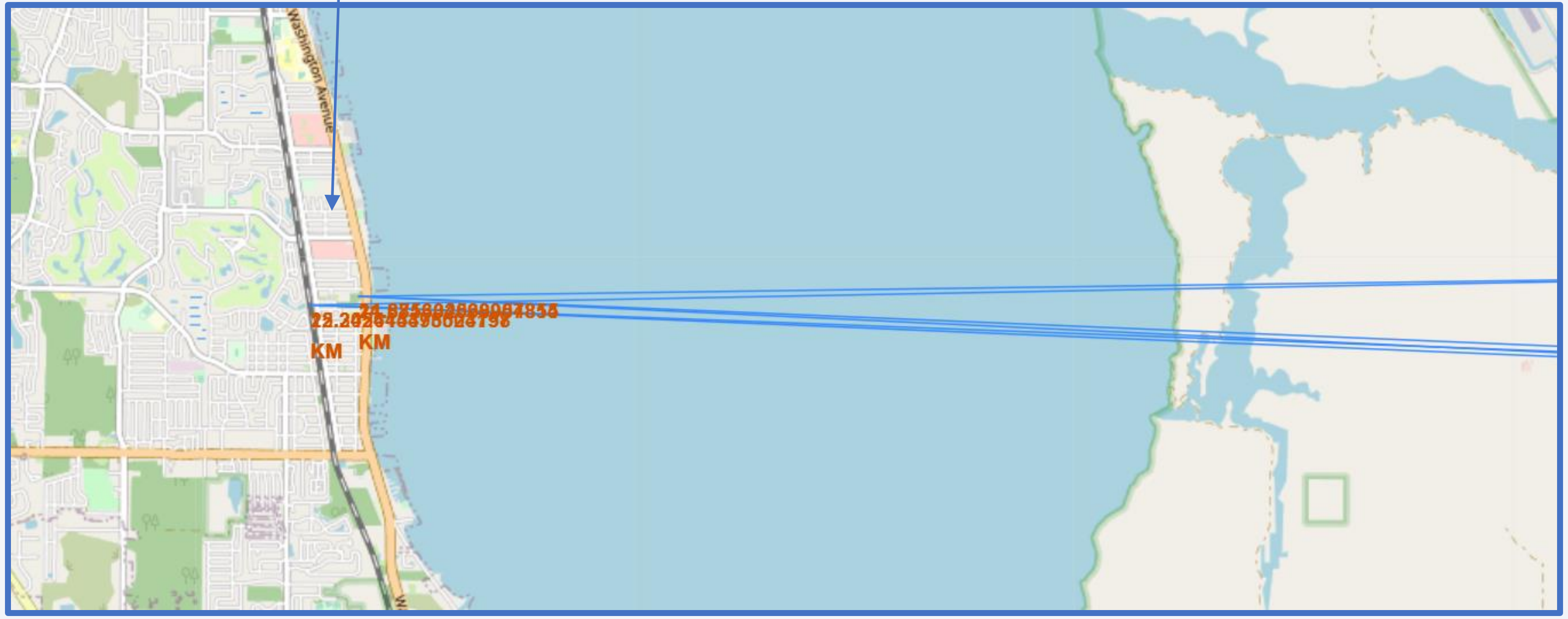
- Blue line marks launch site to nearest city
- The distance to the city is displayed in kilometers at the coordinates of the city
- Haversine distances to closest cities:
- CCAFS LC-40 to Melbourne, FL = 52.58 km
- KSC LC-39 to Melbourne, FL = 53.29 km
- VAFB SLC-4 to Lompoc, CA = 14.01 km
- Launch sites in Florida have Melbourne ~50 km south, the launch site in California has a small city ~ 15 km to the east, and Santa Barbara more than 5 times further away; major cities of Orlando and Los Angelas are given an even wider berth



# Launch Site Folium Map: Proximity to Highway and Rail

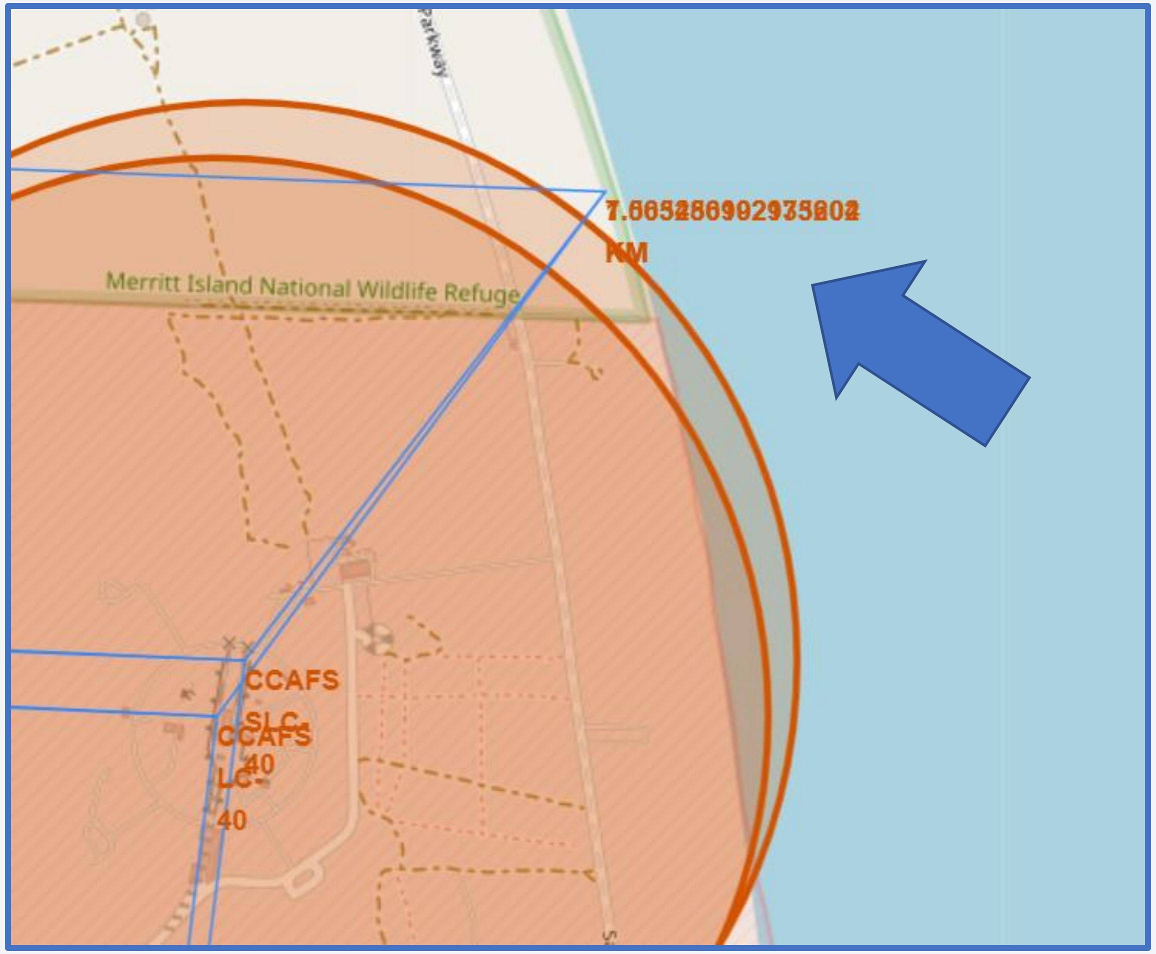
- Haversine distances to closest highway:
- CCAFS SLC-40 to US1 = 21.69 km
- KSC LC39-A to US1 = 14.87 km
- VAFB SLC-4E to CA1= 16.50 km
- Haversine distances to closest railway:
- CCAFS SLC- 40 to Rail = 22.20 km
- KSC LC-39A to Rail = 15.35 km
- VAFB SLC-4E to Rail = 14.00 km

- Highways: further distance is allocated between highways and launch sites, Florida launch sites are east of highways, in California closest proximity is north of the launch site with a wide berth to the east
- Railways: launch sites are located to the east of railways, the risk of launch failure damage is low



# Launch Site Folium Map: Proximity to Coastline

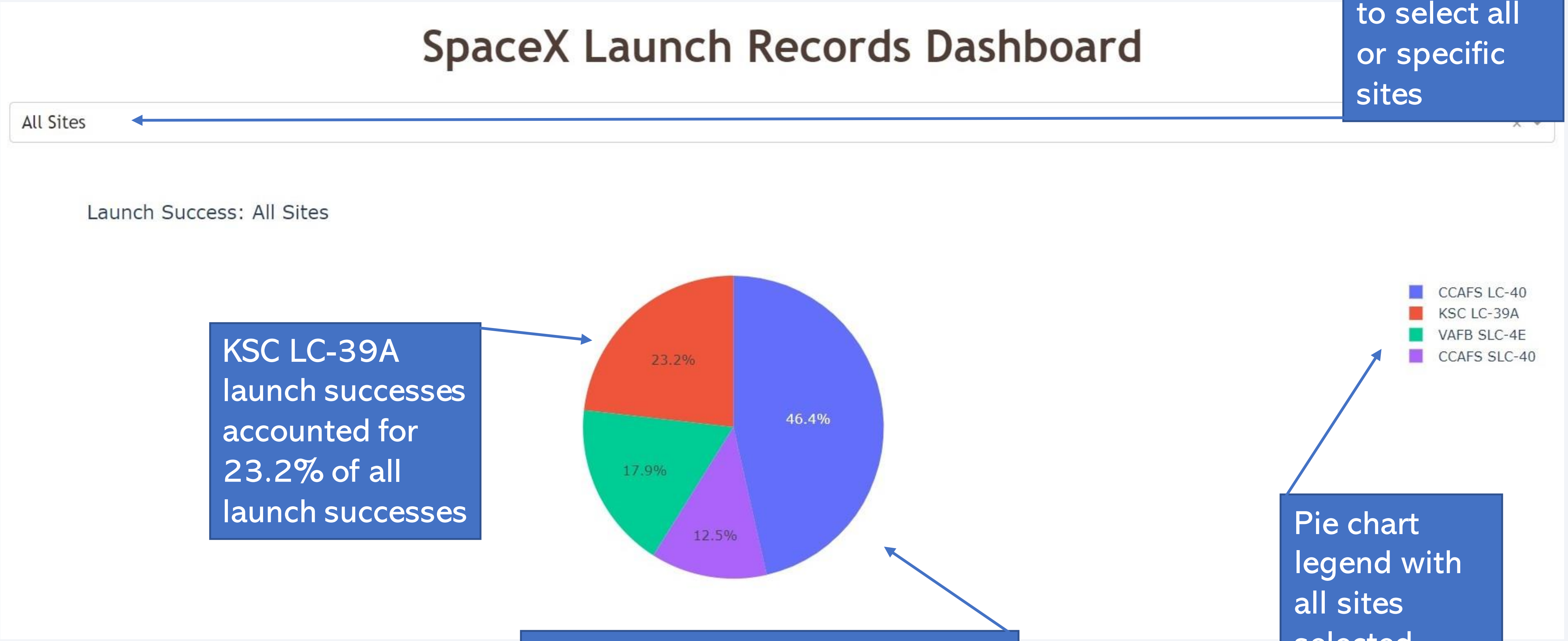
- Coastlines:
- Launch sites are close to coastlines, for reasons mentioned previously
- Additionally this guarantees a non-populous adjacent region should launch failure occur along with the resulting potential damage from falling debris



Section 4

# Build a Dashboard with Plotly Dash

# Dashboard: Pie Chart All Sites



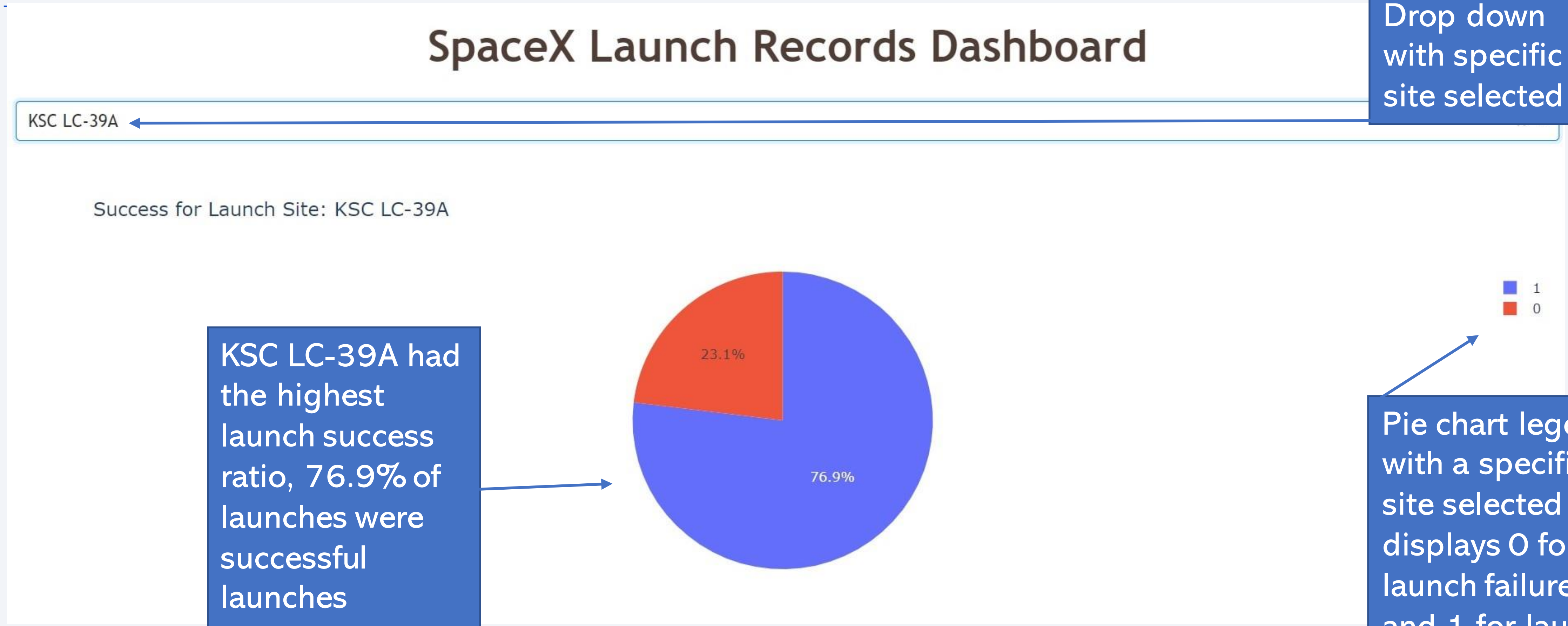

Drop down to select all or specific sites

KSC LC-39A launch successes accounted for 23.2% of all launch successes

CCAFS LC-40 launch successes accounted for 46.4% of all launch successes

Pie chart legend with all sites selected

# Dashboard: Site with Highest Launch Success Ratio

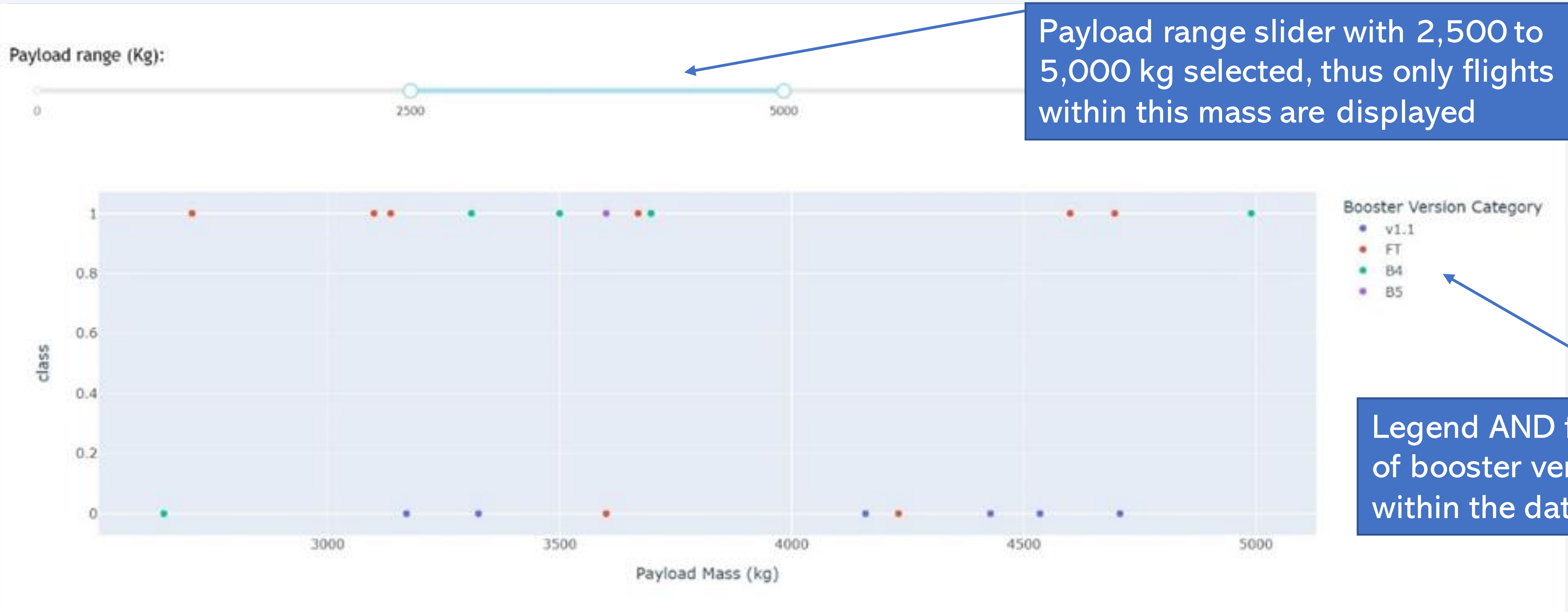



Drop down with specific site selected

KSC LC-39A had the highest launch success ratio, 76.9% of launches were successful launches

Pie chart legend with a specific site selected it displays 0 for launch failure and 1 for launch success

# Dashboard: Payload Range with Highest Success Rate




Payload range slider with 2,500 to 5,000 kg selected, thus only flights within this mass are displayed

Legend AND filter of booster versions within the data

Drop down to select all sites are selected, thus this scatterplot displays data for all sites

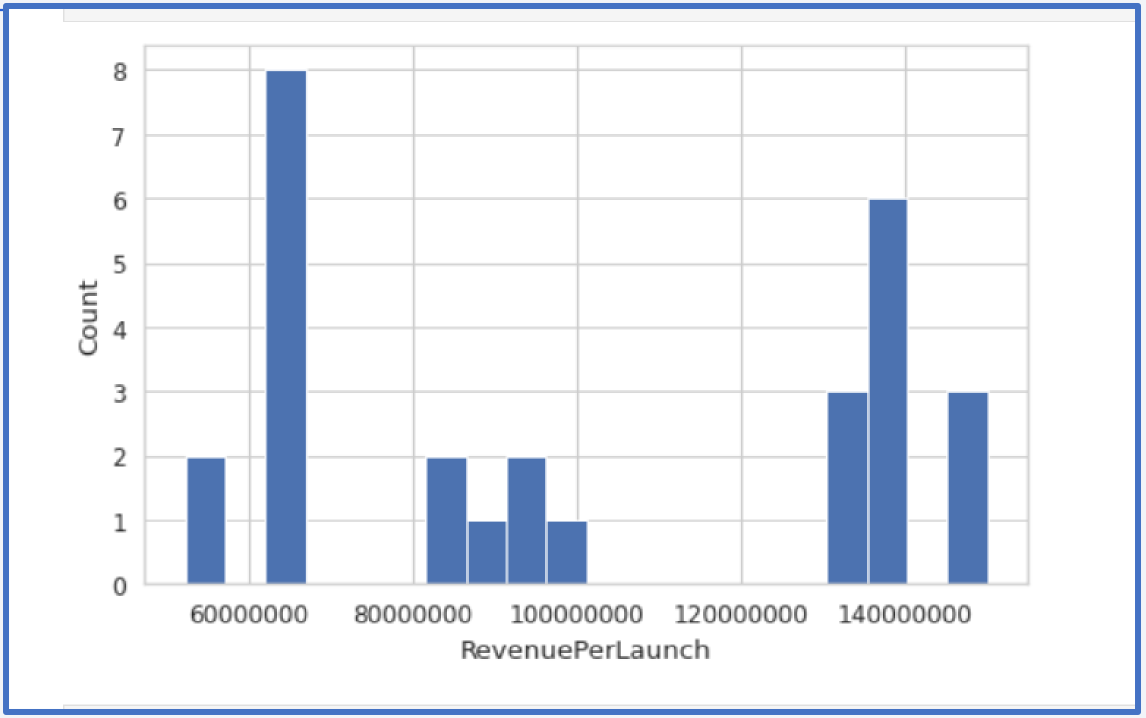
Payload range 2,500 to 5,000 kg had the highest success rate of 55%
This is 11 successful launches (class 1) out of 20 attempted launches (class 0 + class 1) in this range.

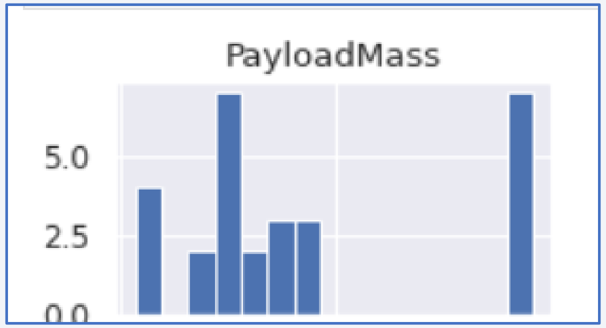
Section 4b

# Tableau: Competitive Mission Bidding

# Contract EDA and Distribution Comparison

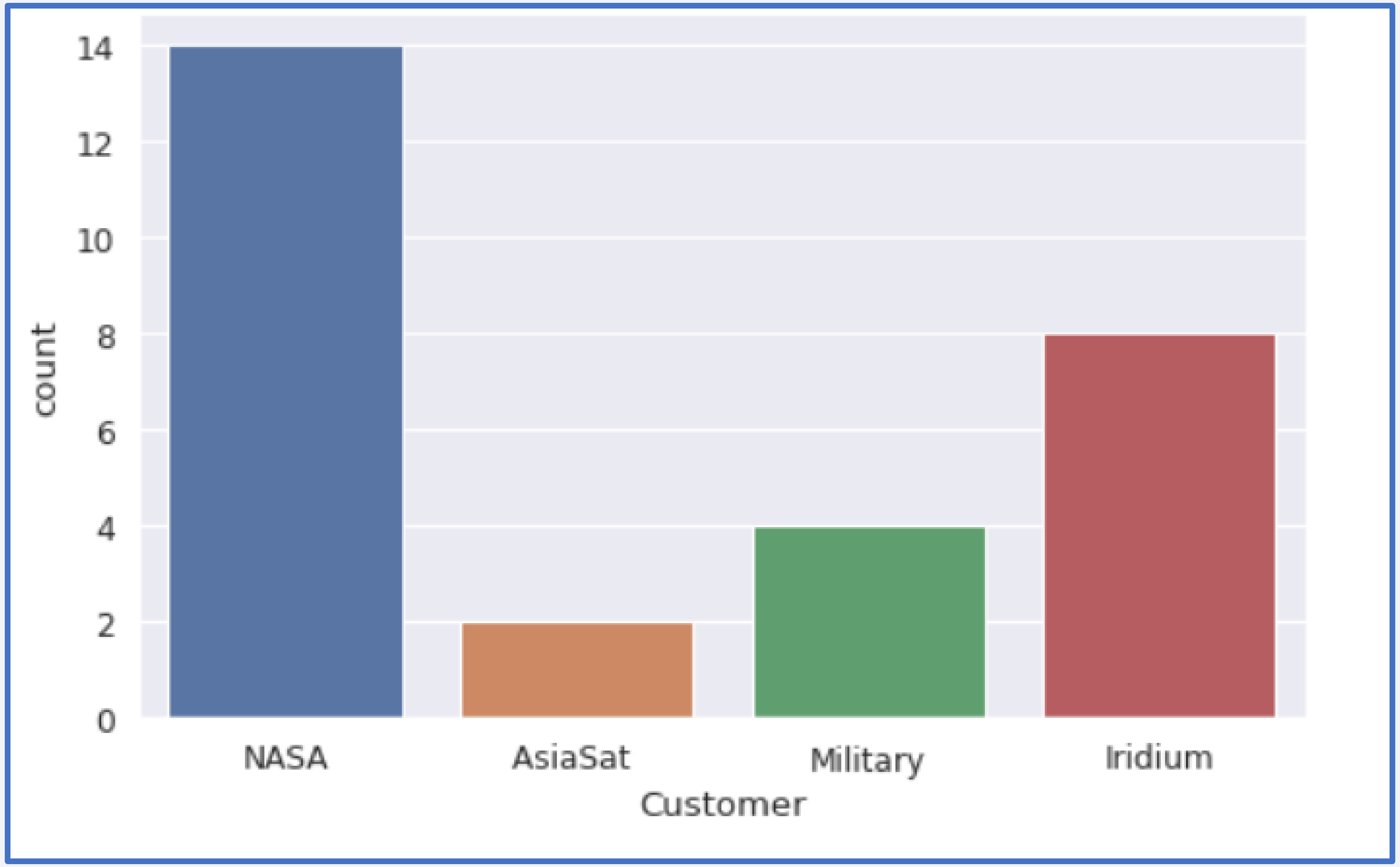
- **Contract data added for 28 flights**
  - **Mission, Customer, Contract Revenue per Launch**
- **Revenue per launch:**
  - **Mean $101,641,428.57**
  - Median $92,600,000.00
  - Range $52,200,000.00 to $150,000,000.00
- Payload mass and Revenue per launch seem to have similar distributions
- Contract Data Source



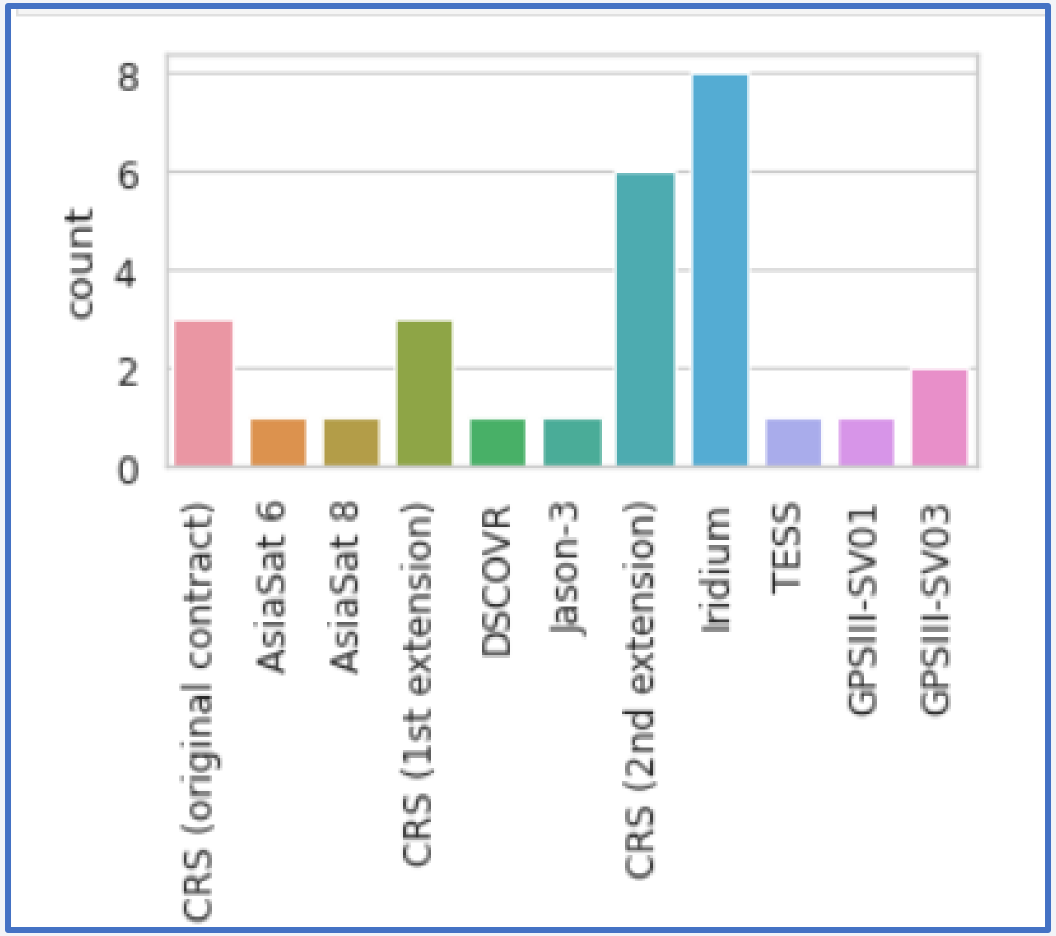
# Contract EDA: Customer Distribution

- Top private sector customer in the dataset: Iridium
- Top public sector customer in the dataset: NASA

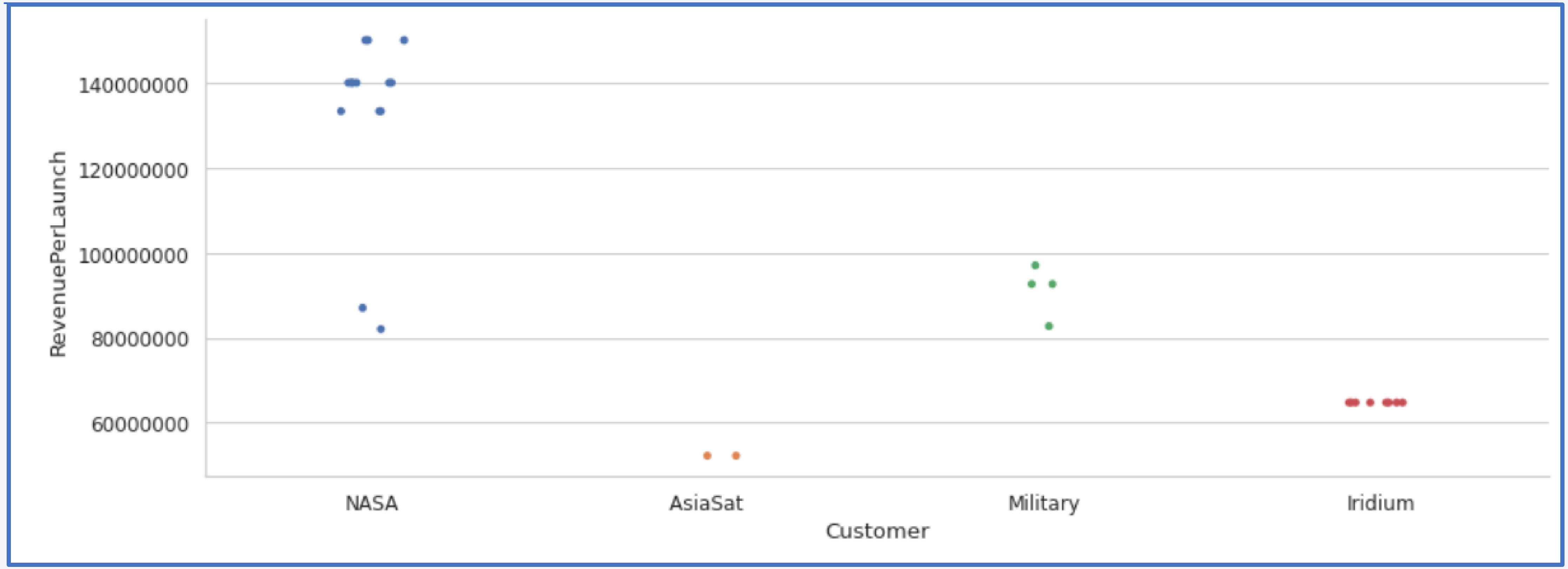


# Contract EDA: Mission Distribution

- Most missions in the dataset: CRS
  - Includes two extensions
  - NASA contract
- Followed by: Iridium

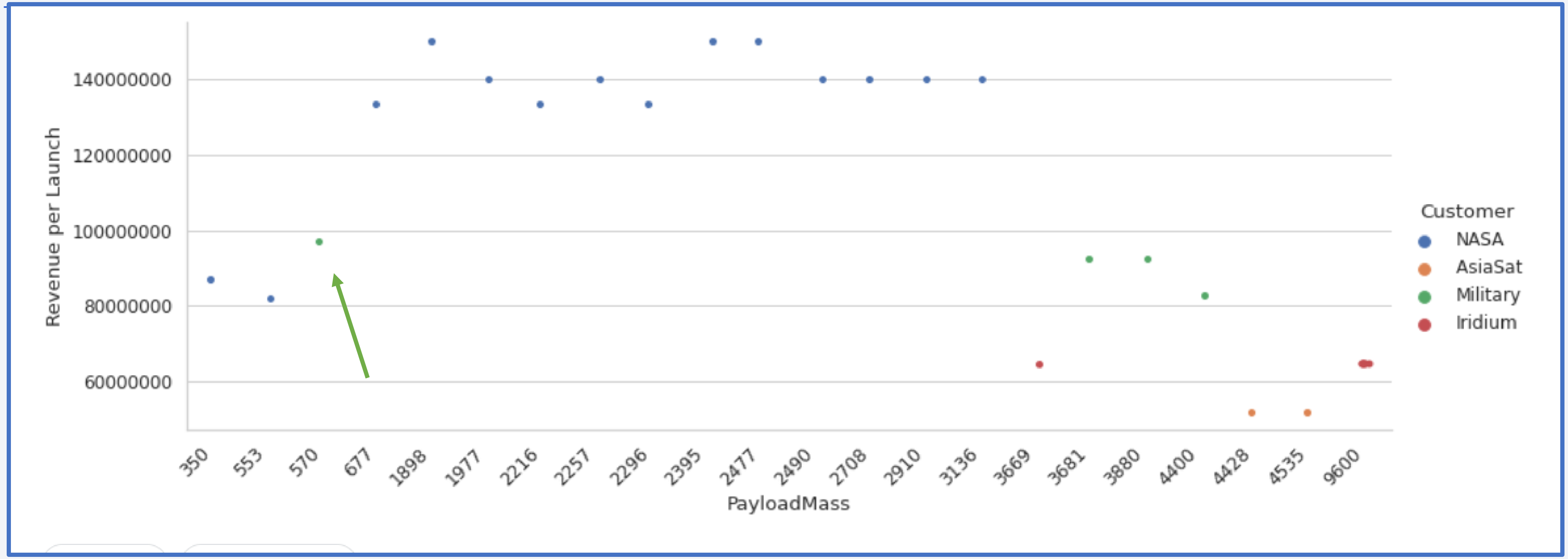


# Contract EDA: Revenue by Customer




- NASA and the military appear to be charged more for missions

# Contract EDA: Revenue by Customer and Payload Mass




- A relationship between public and private sector contract pricing differences cannot be determined from the contract data, as commercial missions all appear to be of greater mass
- Payload mass and price appear to have a quadratic relationship; however, the size of the dataset is too small to determine the exact nature of the relationship

# Contract EDA and Financial Calculations: Cost Sources

- SpaceX website
  - 2020 SpaceX Falcon 9 **base price of $62 million**
    - Increased to $67 million March 2022 due to inflation, 2020 price more relevant to dataset
- Elon Musk on Twitter: *"Payload reduction due to reusability of booster & fairing is <40% for F9 & recovery & refurb is <10%, so you're roughly even with 2 flights, definitely ahead with 3."*
- Elon Musk interviews such as:
  - May 2018 interview for CNBC: *"You've got the boost stage is probably close to 60 percent of the cost, the upper stage is about 20 percent of the cost, fairing is about 10 percent and then about 10 percent which is associated with the launch itself…We may be able to get down to a marginal cost for a Falcon 9 launch down, fully considered, down under $5 million or $6 million."*
  - May 2020 interview for Aviation Week: podcast with specific figures

# Contract EDA and Financial Calculations: Cost Estimates

- **Costs** extrapolated from sources:
  - **New core = $30,000,000**
    - Reused core = $250,000 for refurbishment
  - **New second stage = $10,000,000**
  - **New fairings = $6,000,000**
    - Reused fairings = $750,000 for refurbishment
  - **Fuel and other launch expenses = $4,000,000**
  - *Note: Costs and revenue are ever changing*
    - *Mission offerings: Rideshare program*
    - *Changes in component costs: Economic considerations*
    - *Changes in reuse success: Innovation and experience*

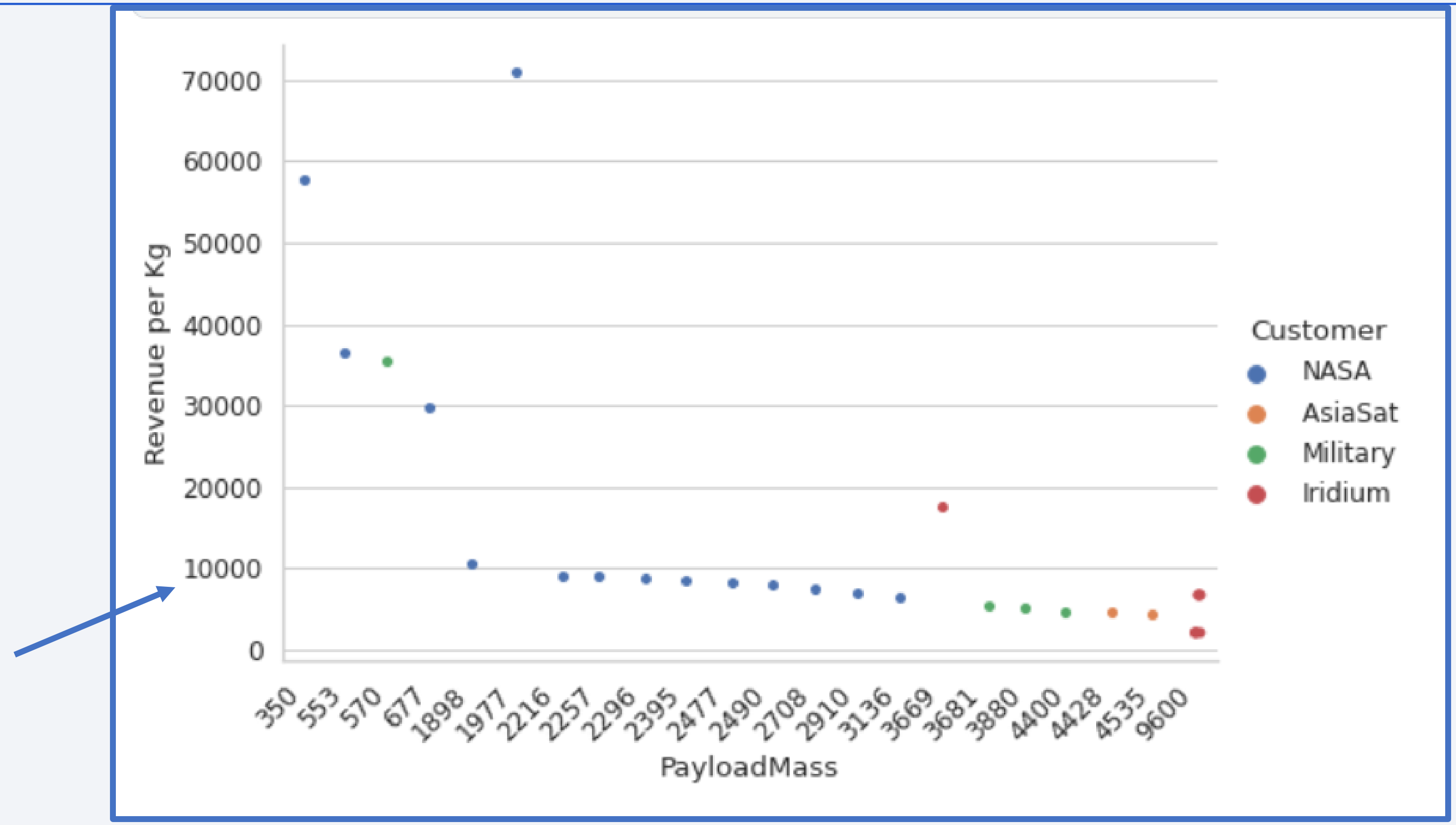
# Contract EDA and Financial Calculations: Cost Estimates

- **New Cost = $50,000,000**
  - $30,000,000 + $10,000,000 + $6,000,000 + $4,000,000
- **Reused Core Cost = $20,250,000**
  - $250,000 + $10,000,000 + $6,000,000 + $4,000,000
  - *No flights with reused fairings were available in the contract dataset*

# Contract EDA and Financial Calculations: Revenue per Kg

- **Compare flight dataset vs contract dataset**: estimated costs to flight information (new cores vs reused cores)
  - Revenue per launch – Cost (New/Reused) = "Revenue Minus"
    - Mean revenue minus = $70,761,785.89
  - Revenue per launch – Revenue Minus = "Revenue Other"
    - Mean revenue other = $30,879,642.68
  - Revenue Other / Payload Mass = "Revenue Other Per Kg"
    - Mean payload mass = 4310.82 kg
- **Revenue** Other **Per Kg**: **Mean $13,737.38 / Median $7,218.30**
  - Range: $2,109.38 to $70,814.36

# Contract EDA and Financial Calculations: Revenue per Kg




**$10,000 per kg** seems to be a reasonable estimate for SpaceX charge per kg

# Contract EDA and Financial Calculations: Allocation

- Average Revenue = $101,641,428.57
  - Mean Revenue per Launch
- **Allocation pattern = 0.3038 (~30%)**
  - Average Revenue Other / Average Revenue
- An allocation estimate of 30%, and **70% to cost** seems appropriate

- 0.70 x $10,000 x Payload Mass = **Cost from Payload Mass**

# Financial Calculations to Columns

**CostCoreMil**
- Reused Core = $250,00
- New Core = $30,000,000

**CostFairingsMil**
- Reused = $750,000
- New = $ 6,000,000

**CostOtherMil**
- New Second Stage = $10,000,000
- Fuel and Other Costs = $4,000,000

**CostKgMil**
- 70% of RevenueKgMil

**CostTotalMil**
- Sum of Cost columns

**RevenueReuseMil**
- New First Stage = $62,000,000 base price
- Reused Core = $50,000,000 base price

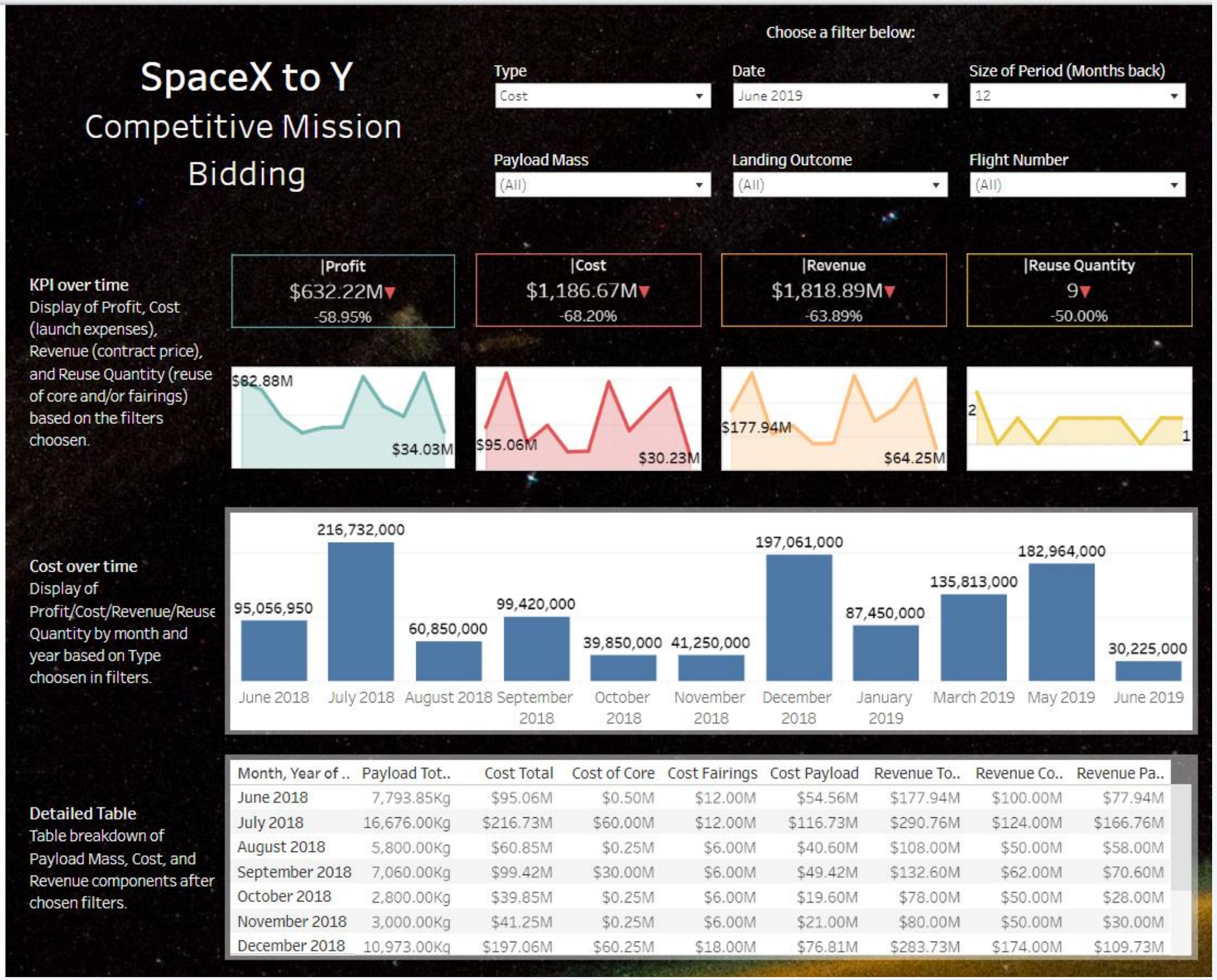
**RevenueKgMil**
- Cost per Kg of Payload Mass = $10,000
- Payload Mass x 10,000

**RevenueTotalMil**
- Sum of Revenue columns

**ProfitMil**
- RevenueTotalMil - CostTotalMil

# Tableau Dashboard: Competitive Mission Bidding

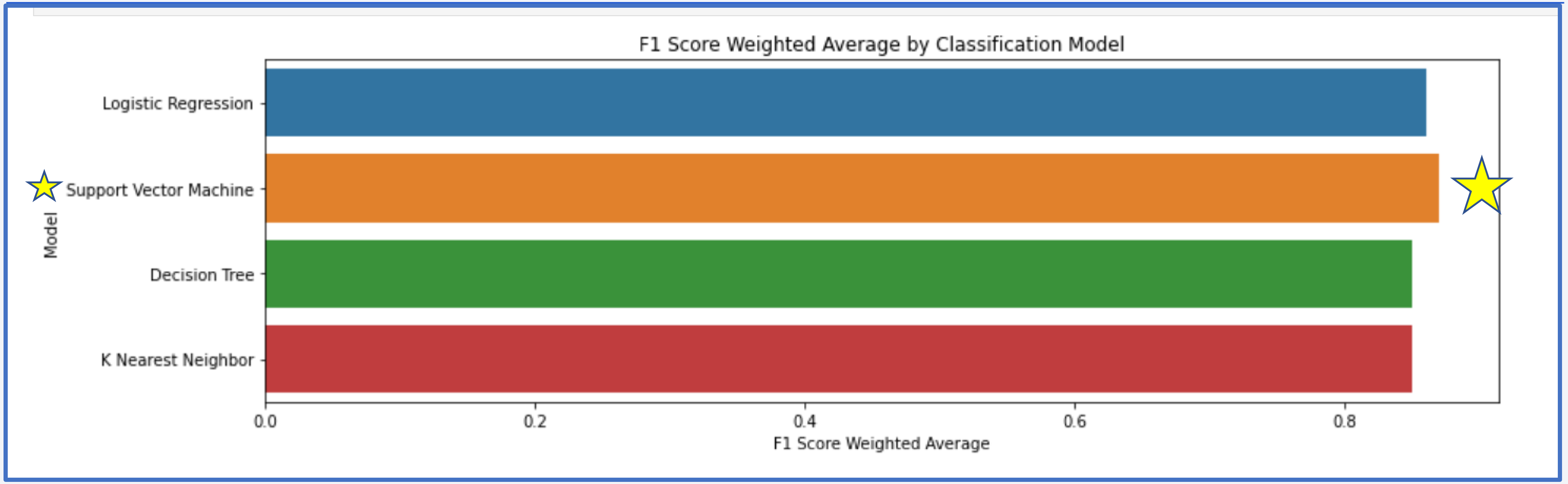
- Design aimed to aide decision makers in choosing an amount to bid on any given mission
- Filters enable choice of payload mass to match up launch Cost/Revenue/Profit/Reuse estimates for comparable missions
- Detailed table allows access to finetune a bid amount based on KPI component breakdown
- **See and use the dashboard:**
- Space X to Y Competitive Mission Bidding Dashboard



Section 5

# Predictive Analysis (Classification)

# Evaluation of Classification Reports (Test Data Classification Performance)




Precision: SVM 85%, a greater percentage of SVM predictions were correct vs. all other models, greater accuracy of positive predictions (TP/(TP + FP)). Recall: SVM 100% of successful landings were caught, greater ability to find all positive instances (TP/(TP+FN)) compared to Decision Tree and KNN

**F1 Score Weighted Average: SVM 87%, greatest of all models**

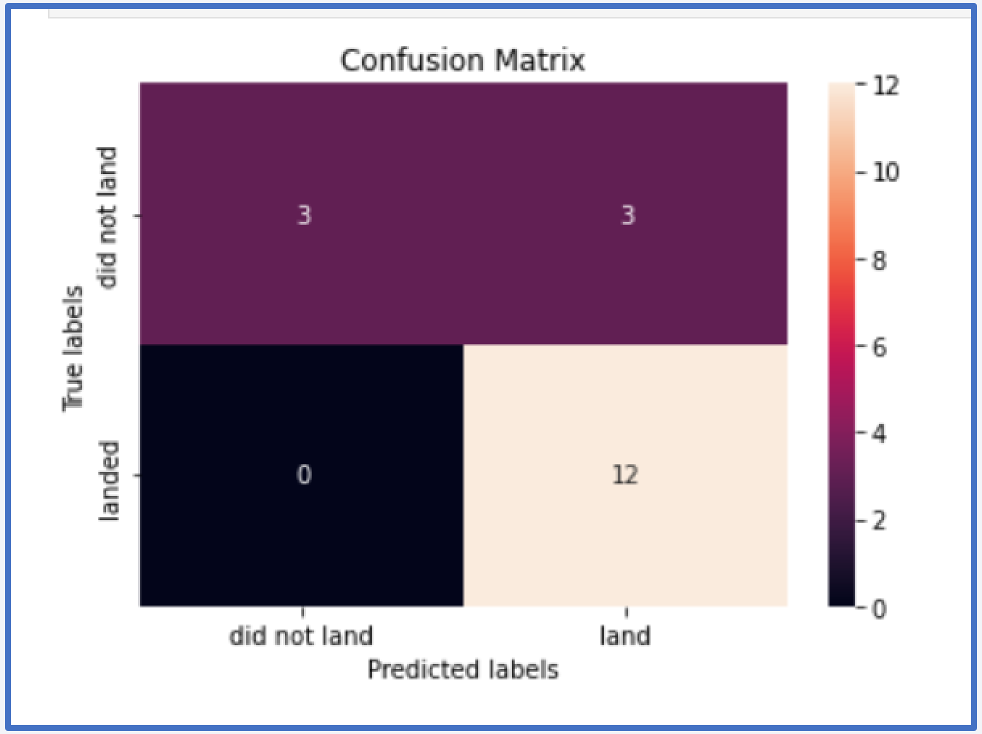
*Weighted Average F1 Score: it is suggested to compare models on the weighted average of F1 as opposed to global accuracy, as this calculation takes the mean of all per-class F1 scores and considers each class's support*

# Confusion Matrix: Support Vector Machine

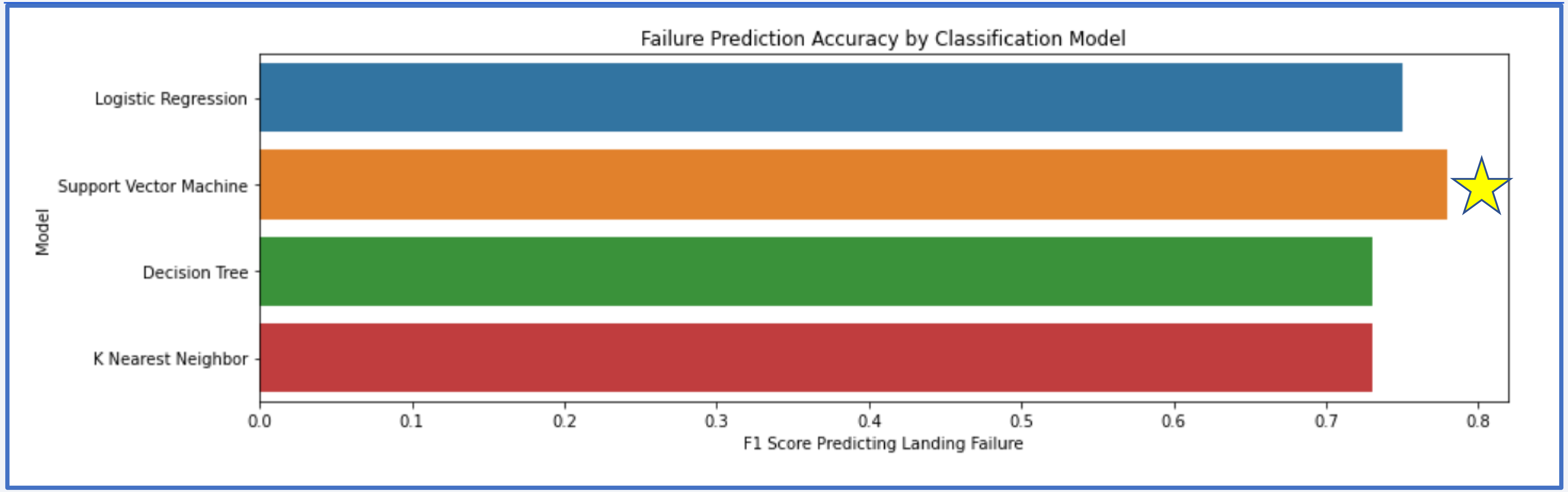
[[TN FP]

[FN TP]]

- False positives: 3 predicted to land, but failed
- False negatives: 0 predicted to fail, but landed
- Great at predicting: landing
- Not so great at predicting: failure (false positive issue) …



# Classification Accuracy: False Positive Comparison

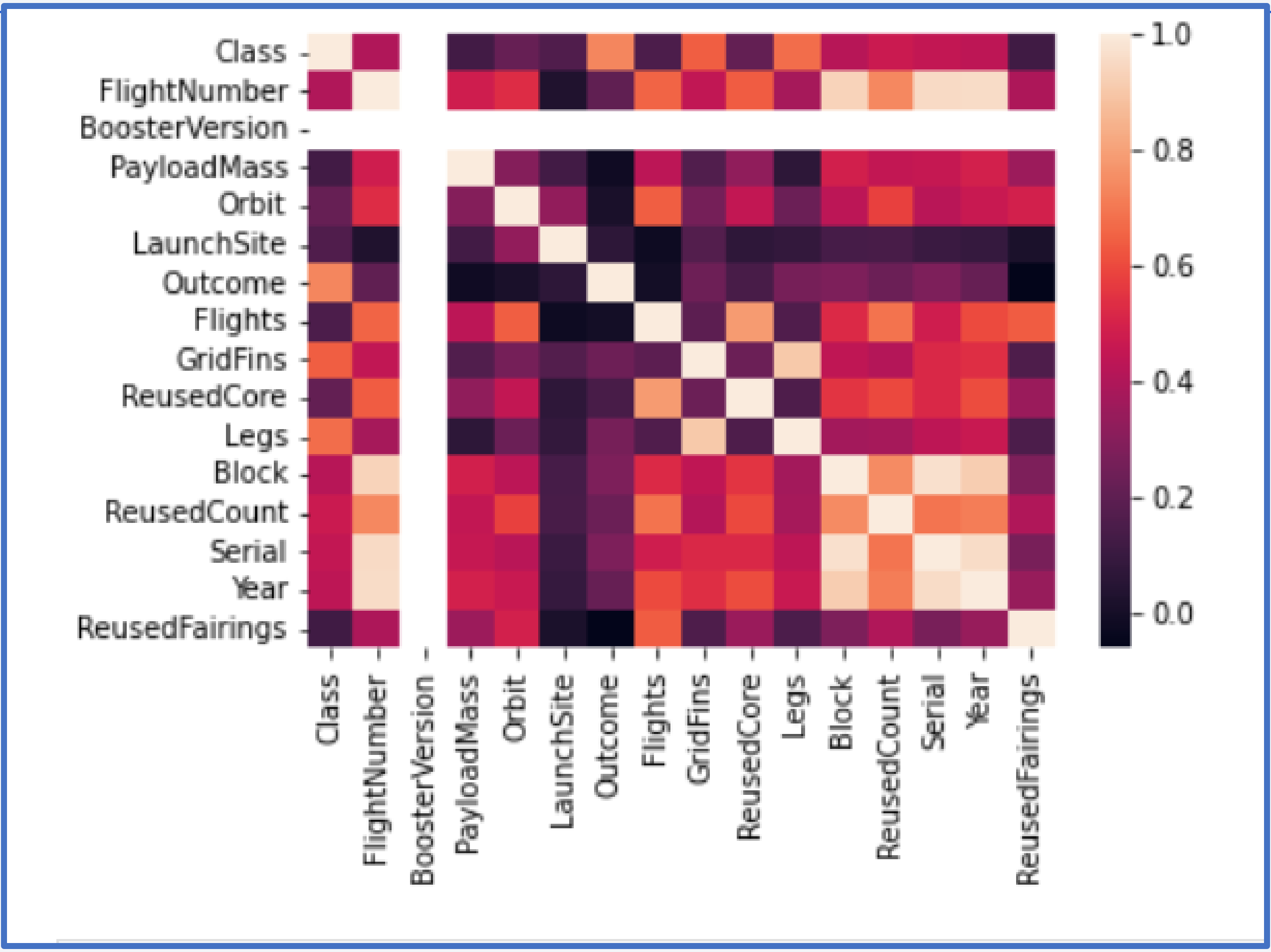



- F1 Score Predicting Failure:
- SVM 78%, the greatest percentage of correct failure to land predictions

Section 5b

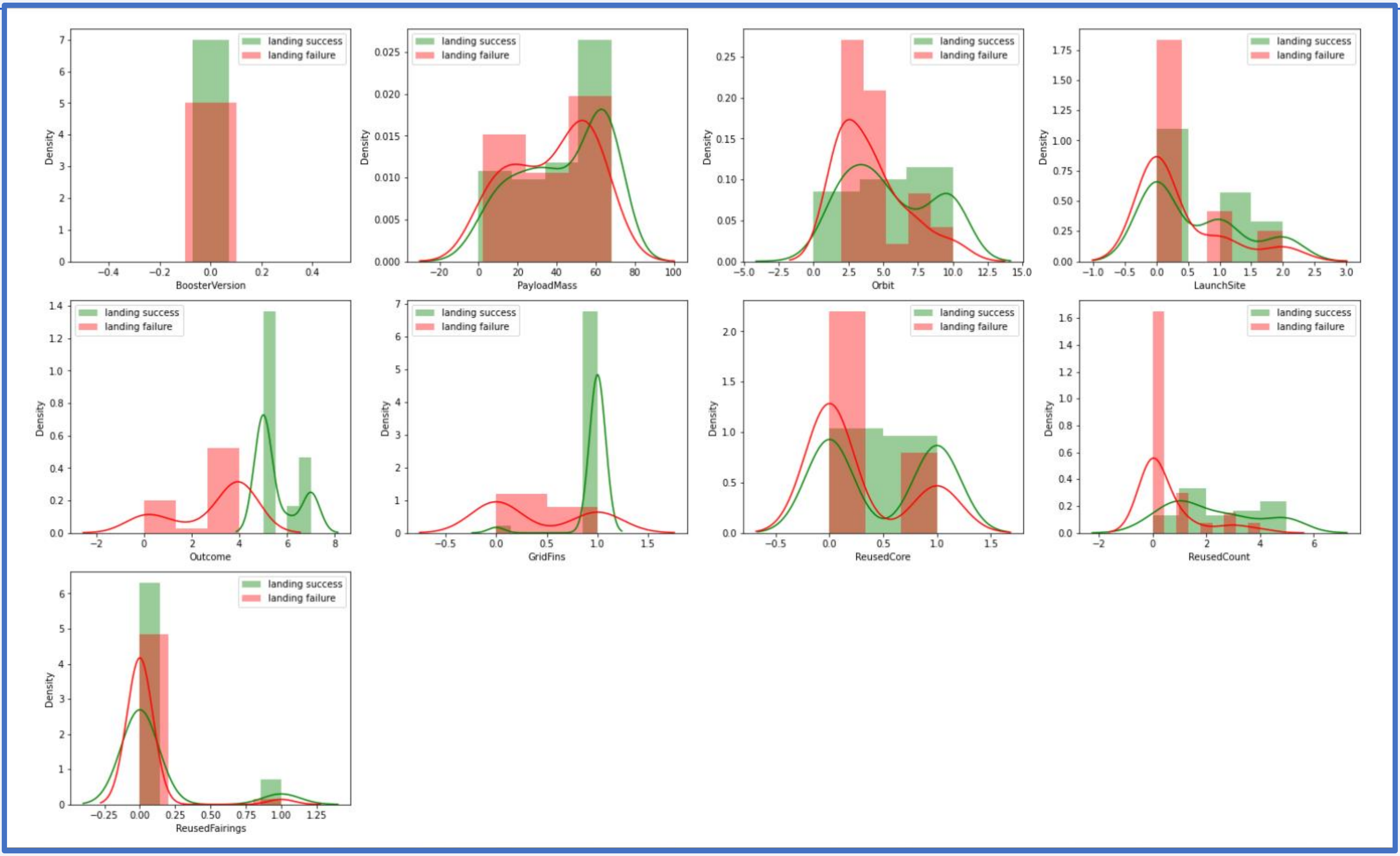
# Model Improvement
# Feature Selection

# Feature Selection and Remodeling




- **Assess redundant data and overfitting** by calculating correlation coefficients, removing one or two features with correlation coefficients > 0.9 between one another: Block, Serial, Legs, and Year

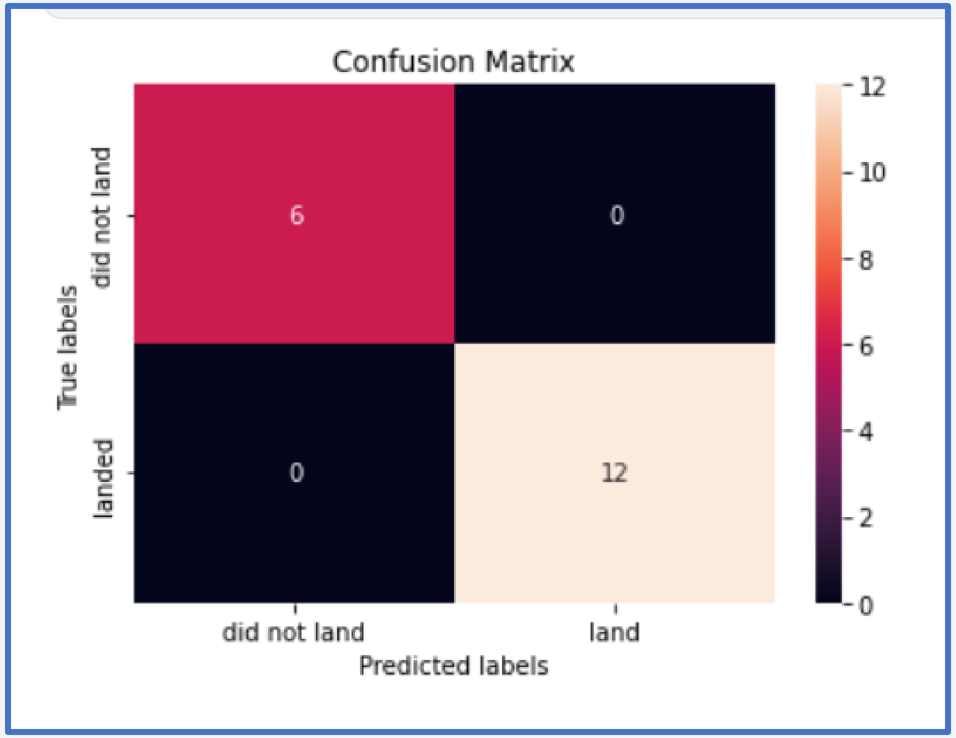
# Feature Selection and Remodeling: Distribution of the Remaining Features



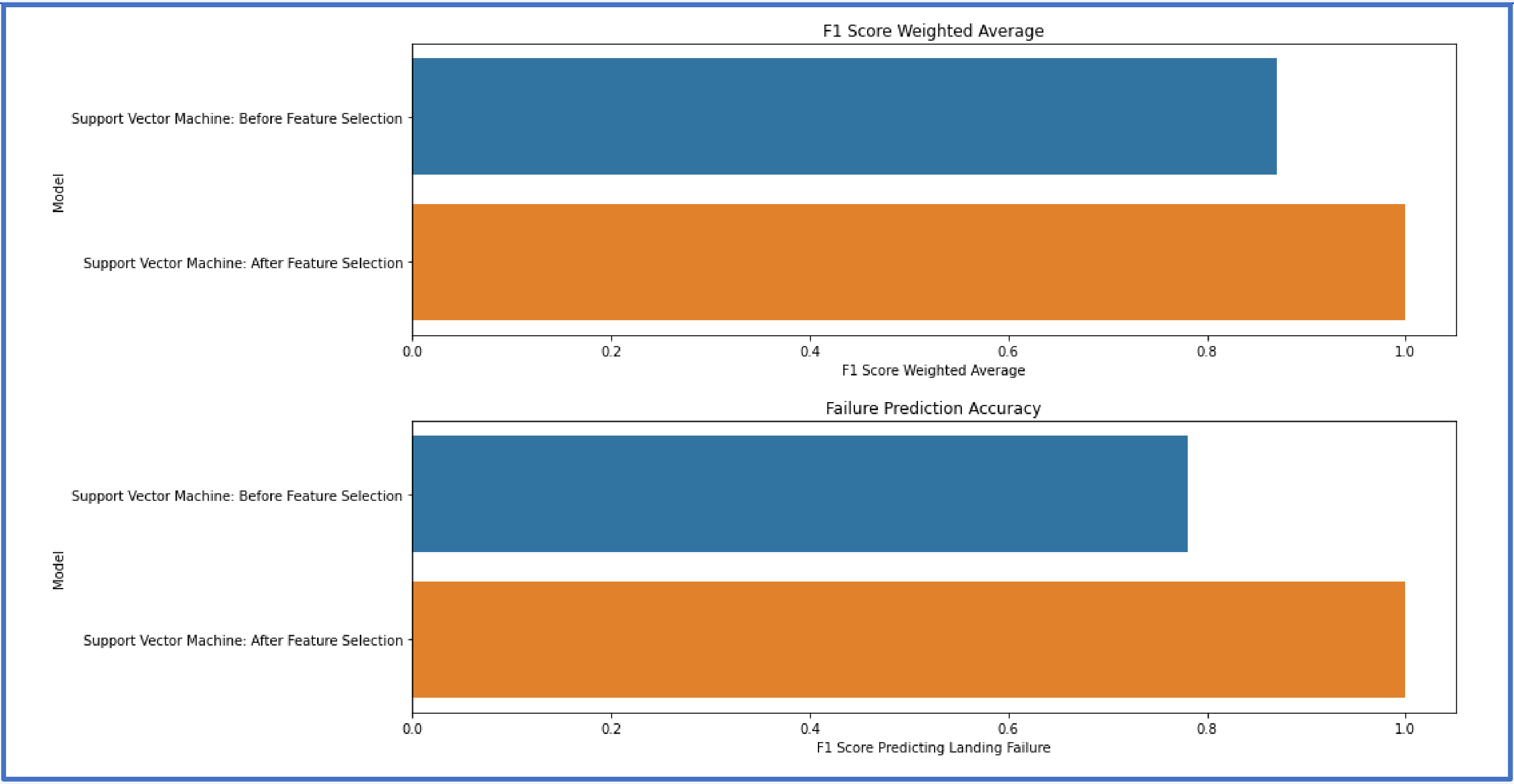

- **Test the null hypothesis:** “The selected feature combination does not have any effect on the target variable" and remove features with p-values >0.05, resulting in the additional removal of Class, FlightNumber, Flights

# Feature Selection and Remodeling: Re-Evaluation

- After Repeating:
  - GridSearchCV to identify best parameters
  - Rebuilding and Retraining the SVM
- Re-Evaluation:
  - Confusion Matrix
    - **False Negatives: 0**
    - **False Positives: 0**
  - Classification Report
    - See next slide



# Feature Selection and Remodeling: Re-Evaluation




- Feature selection further improved the SVM model, with **100% accuracy across all metrics** in the classification report

# Conclusions: Space X to Y

# Conclusions: Space X to Y

- How much should SpaceY bid on a mission contract?
  - How likely is the Falcon 9 booster (first stage or core) to land successfully?
    - How much will the mission cost?

| Exploratory Data Analysis | Landing Predictions | Launch Costs and Bidding |
| --- | --- | --- |
| Key Findings | Model Accuracy and Meaning | Strategic Decisions and Tools |

# Conclusions: Flight Number: Relationships to Landing Success

- Landing success increases overtime with a pivotal comparative distribution change ~flight 20, which is sustained throughout the dataset. This coincides with landmarks:
  - 12/22/2015 (Flight 17) the first Falcon 9 Full Thrust (v1.2) flight, and the first successful landing on a ground pad
  - 4/8/2016 (Flight 20) the first successful landing on a drone ship, and the first booster to be later re-flown
- Flight number is weakly correlated (correlation coefficient 0.401923) with landing success, and is strongly correlated with Year and Block (0.95 and 0.93 respectively)
  - Flight number seems to represent an increase in landing success over time and with advances in rocket design.

# Conclusions: Launch Site: Relationships to Landing Success

- Launch site with the highest success rate and lowest success rate, KSC LC-39A 77.3% and CCAFS SLC-40 60%, are located next to one another along the Florida coastline, while VAFB SLC-4E located on the west coast of California ranks in the middle 76.9%.
- Launch site does not have a linear relationship with landing success (correlation coefficient 0.160706). Launch site is more highly correlated with Orbit (0.33).
  - This points to the nature of launch site choice. High-inclination orbits such as PO and SSO necessitate western coast launch sites, as their unique trajectory avoids populous locations. They also require more fuel and non-equatorial locations such as VAFB SLC-4E.
    - Launch site and payload mass: As reaching these orbits requires more fuel, this limits the payload mass available for the flight.
  - Florida is more well suited, and utilized in the data set, for low-inclination orbits and higher payload masses.
  - https://solarsystem.nasa.gov/basics/chapter14-1/#sites

# Conclusions: Payload Mass: Relationships to Landing Success

- Payload masses between 2,500 – 5,000 kg were the most successful range; however, it also had a high failure to success ratio. Payload masses between 5,000 – 7,500 kg had both had the most failures and highest failure to success ratio.
- Payload masses >7,500 kg did not occur until later flights, and flights 10,000 kg or more did not occur until even later.
- Payload mass does not have a linear relationship with landing outcome (correlation coefficient 0.199992). Payload mass is more strongly correlated with VLEO, Flights, and Reused Count (0.85, 0.67, and 0.61 respectively).
    - Payload mass is more related to the allowable mass, such as with VLEO, larger payloads can be carried to orbits without inclination restraints.
    - Payload mass is also more related with time, which is also related to more successful landings and more reuse, thus we see high success to failure ratios even amongst the heaviest payloads and more reuse as payload increases.

# Conclusions: Orbit Type: Relationships to Landing Success

- The orbits with the highest success rates and five or more flights include:
  - SSO, VLEO, and LEO.
  - For some orbits this appears to increase overtime, and for others it does not. When also looking at payload mass, a possible reason for this observation is visualized. The latter have experience with larger payload masses, while the former do not.
- Orbit type is weakly correlated with landing success (correlation coefficient 0.217868). Orbit type is more strongly correlated with Flights, Payload mass and Reused count (0.64, 0.62, and 0.58 respectively).
  - Missions to some orbit types such as SSO were not attempted until 8/24/2017; it seems over time more mission type options emerged.
  - The relationship with payload mass seems to be both as discussed before and a factor of time. Time being reflected in the moderate correlation with both Flights and Reused count.

# Conclusions: Reused Core: Relationships to Landing Success

- Comparative distribution of core reuse by landing outcome suggests that landing success occurs with nearly equal frequency with both non-reused cores and reused cores. Additionally, landing failure seems to be less frequent with core reuse.
- Reused core is weakly correlated with landing success (correlation coefficient 0.207582). Reused cores may have a linear relationship with Flights as there is a strong correlation (0.78). This is further supported by moderate relationships with Flight Number and Year (0.64 and 0.6 respectively), which are variable indicative of time.
  - The first flight using a reused and refurbished core was on 3/30/2017 (Flight 29). When looking only at flights >or=29 the success rates of new and reused cores converge to 80% and 78.4% respectively.
  - One must also consider that with reused cores, a landing may not be attempted as it is standard operation to either retire the core or to send it on a mission where it will be expended based on the nature of the mission.
    - For example, 6/4/2018 (Flight 51) was an SES satellite deployment mission for delivery into geostationary transfer orbit (GTO), as it was a block 4 version after the transition to block 5 and had already been flown twice no attempt to recover the core was made; however, an attempt to recover the fairings did occur.
    - https://www.nasaspaceflight.com/2018/06/falcon-9-cape-canaveral-night-launch-ses-12/

## Conclusions: GridFins and Legs: Relationships to Landing Success

- If gridfins and/or legs are present at launch, landing success rate is 80%. Flights on which grid fins and/or legs were not present at launch had landing outcomes, “No Attempt”, “Controlled Ocean”, “Uncontrolled Ocean,” associated with testing or expending.
- Grid fins and legs both may have linear relationships with landing success (correlation coefficients 0.56). They are more strongly correlated with one another than with class (correlation coefficient 0.9). They are also moderately correlated with year (0.46).
  - These relationships suggest that the presence of grid fins and/or legs reflects the intention to land the core.

# Conclusions: Year: Relationships to Landing Success

- From 2010 to 2020 landing success rate increased from 0% to 84.2%. From 2017 to 2018 there was a dip in success rate of 22%.
- Year is weakly correlated with landing success (correlation coefficient 0.433912), and is more highly correlated with Flight number and Block (0.95 and 0.91 respectively). This points to redundancy of these features as well as the intuitive relationship with time like features of ordinal numbers.
  - In 2018 many changes were occurring that seem related to a decrease in success rate.
    - This was a transition period from Block 4 to Block 5, associated with core expenditure and retirement.
    - Larger payload masses began to be introduced, but also limited with additional mission types.
    - Expansion of missions to other orbits, such as GTO which has a comparatively low success rate of 52%.
    - GTO missions involve more fuel, which leaves less reserve for the braking burn upon reentry and presents the challenge of landing a booster returning at higher velocity.
    - https://spaceflight101.com/falcon-9-ses-9/falcon-9-exceeds-expectations-in-high-energy-delivery-of-ses-9/
    - https://web.archive.org/web/20190727151524/https://www.spacex.com/sites/spacex/files/spacex_ses9_press_kit_final.pdf

# Conclusions: Landing Predictions: Model Accuracy and Meaning

- **Predictive Analysis and Classification of Landing Outcome: Mission Bidding Success**
  - Landing success is a multivariate dependent variable.
    - This points to the need for consideration of multiple independent variables and a beneficial role machine learning models can play in accurately predicting whether landing success will occur.
  - The **support vector machine model (SVM)** initially built with parameter optimization resulted in a F1 score weighted average of 87%.
  - Further improvement of the model was achieved through utilizing feature selection and removal of redundant independent variables. This **SVM model was able to achieve a F1 score weighted average of 100%**.
    - As this model <u>can accurately predict both successful landing and landing failure</u>, it shows promise once deployed.
    - This will **aid us in making pivotal mission bidding decisions that can be both competitive and avoid excessive expenditure.**
  - Best parameters: C = 1.0, gamma = 0.001, kernel = 'linear'
  - Selected independent variables = ['BoosterVersion', 'PayloadMass', 'Orbit', 'LaunchSite', 'Outcome', '', 'GridFins', 'ReusedCore', 'ReusedCount', 'ReusedFairings']

# Conclusions: Launch Cost and Bidding: Stategic Decisions

- **Contract EDA, Estimates, and Competitive Bidding Dashboard**
  - To further aide decision making the analysis of contract information was able to provide a foundation for more detailed guidance.
  - This was summarized into a dashboard that can put these calculations into action.
    - Being able to breakdown costs/revenue/profit/reuse will hopefully aide decision makers in fine tuning mission bid amounts.
  - *As the company gains more experience, and thus more data, we can translate this tool into one based on both internal and external information.*

  - https://public.tableau.com/views/SpaceXDashboard_16743477717270/Dashboard1?:language=en-US&:display_count=n&:origin=viz_share_link

# Next Steps: Space X to Y

# Next Steps: Space X to Y

## Exploratory Data Analysis

Continue data collection and EDA

Provide new insights on new information

Explore new aspects that emerge over time

Provide feedback concerning angles to consider

## Landing Predictions

Model deployment

Ongoing assessment

Feedback

Improvement

Repeat

Expand to additional models

## Launch Costs and Bidding

Continue to gather contract data

Improve estimates as our company matures

Provide feedback

Adjust to economic ecosystem and marketplace landscape

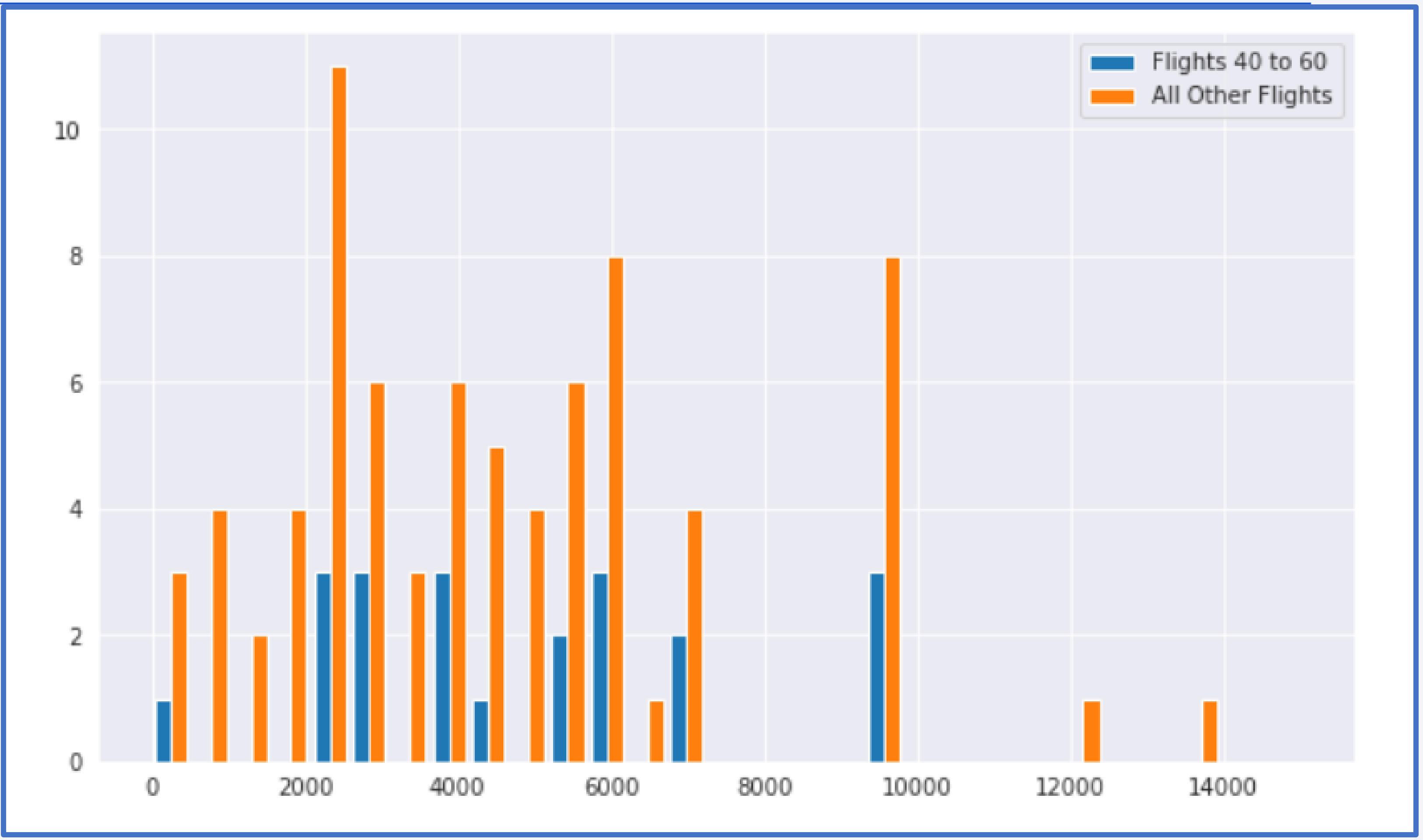
Improve decision aides

# Appendix

- Python code, SQL queries, charts, and Notebook outputs
  - https://github.com/brittabeta
- Data sets and Metadata
  - https://www.kaggle.com/datasets/brittasmith/spacextoy-dataanalysis-launchprediction
- Tableau Dashboard
  - https://public.tableau.com/views/SpaceXDashboard_16743477717270/Dashboard1?:language=en-US&:display_count=n&:origin=viz_share_link
- Additional Slides
  - Flights 40 to 60
  - Launch sites
  - Payload mass range subsets
  - Block 4

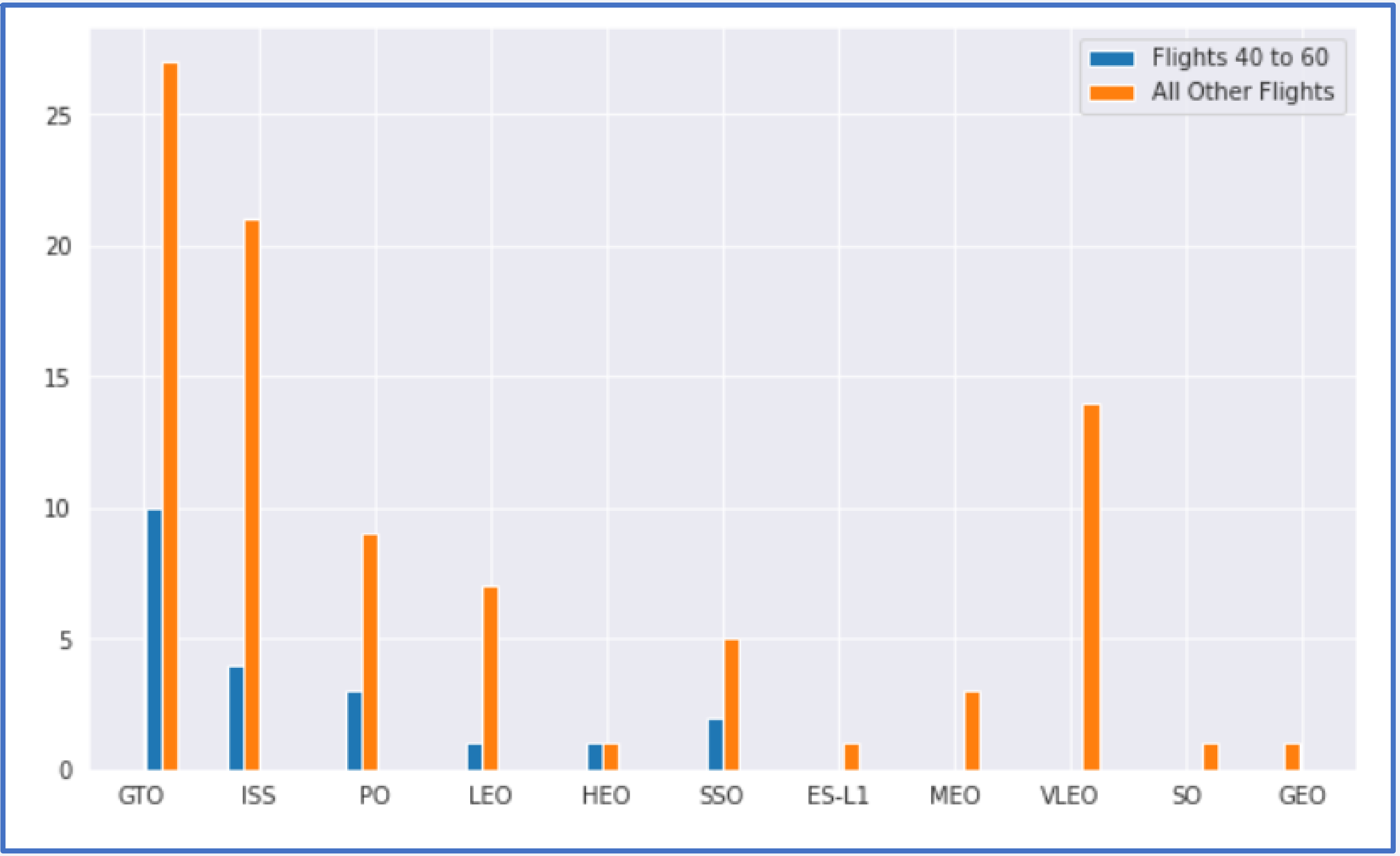
# Flights 40 to 60: Largest Differences – Payload Mass

- Flights 40-60:
  - Most payload masses are between 2500-7500 kg
- All other flights:
  - Most payload masses are between 0-7500 kg



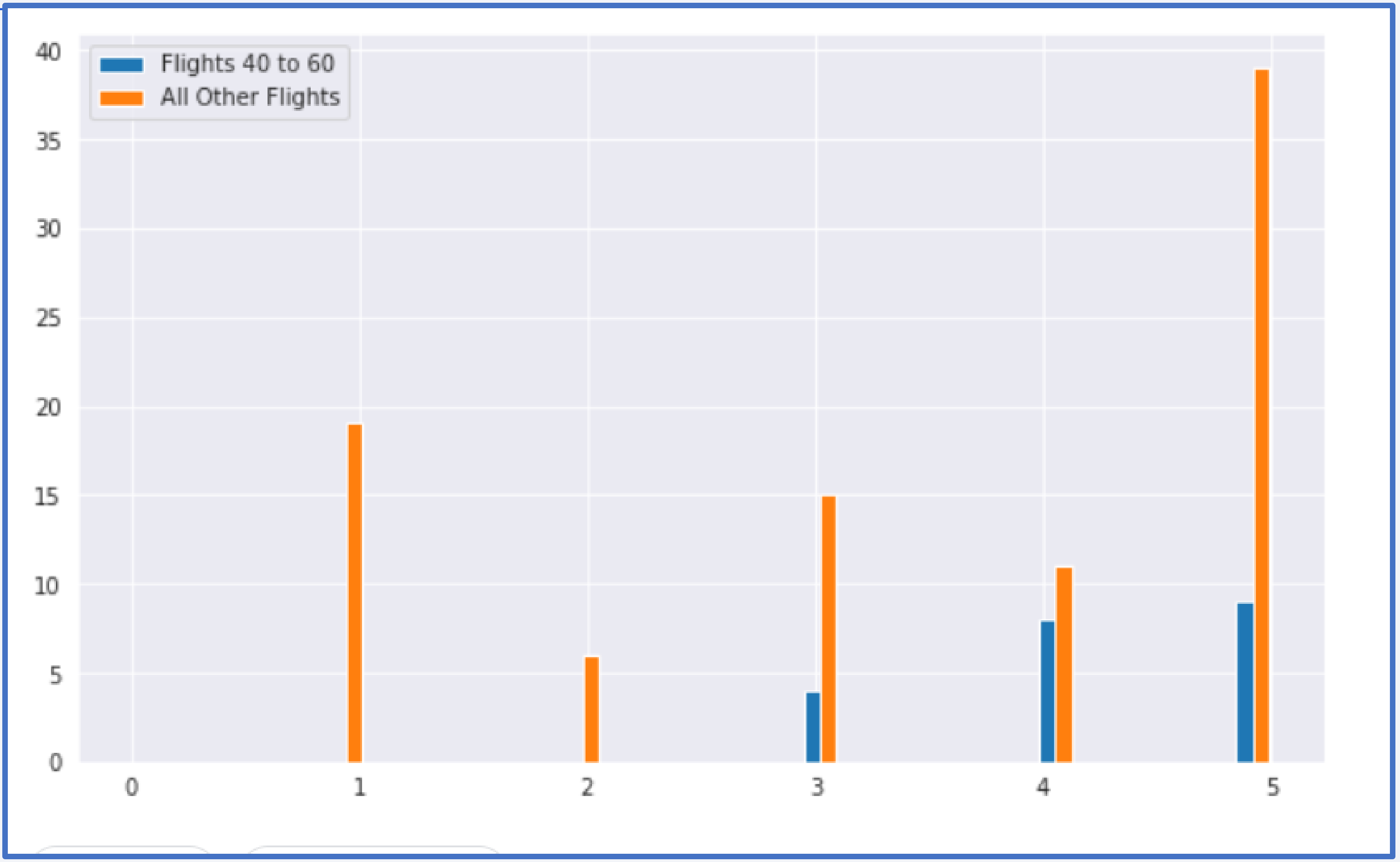
# Flights 40 to 60: Largest Differences – Orbits

- Flights 40-60:
  - GTO 47.62%
    - vs 39.13%
  - ISS 19.05%
    - vs 30.43%
  - VLEO 0%
    - vs 20.29%
  - LEO 4.76%
    - vs 10.14%



# Flights 40 to 60: Largest Differences – Blocks

- Flights 40-60:
  - Block 4 38.10%
    - vs 15.94%
  - Block 5 42.86%
    - vs 56.52%



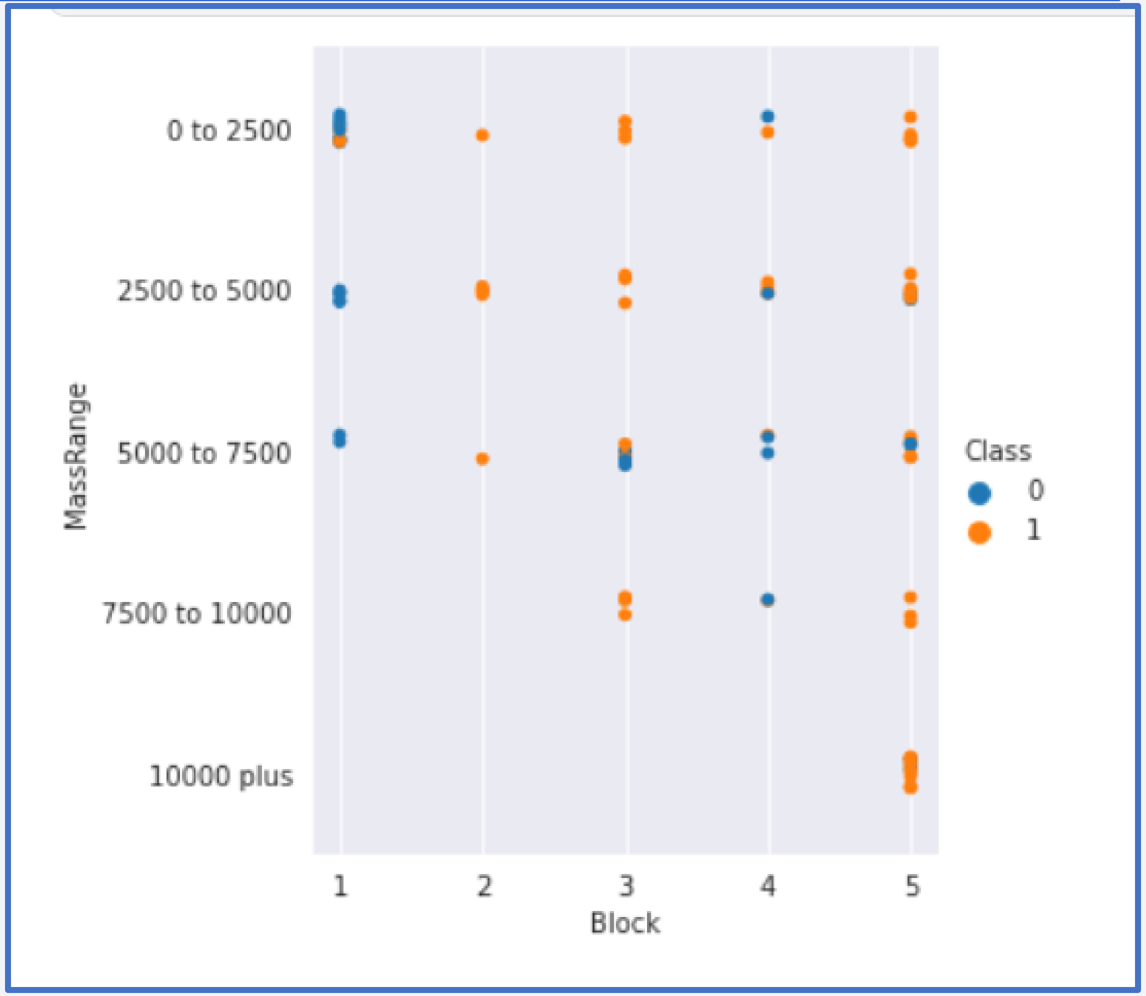
# Flight Number vs. Launch Site: Success Rates

| LaunchSite | SuccessRateAllFlights | LaunchSite | SuccessRatePostTesting |
|---|---|---|---|
| CCAFS SLC 40 | 0.600000 | CCAFS SLC 40 | 0.750000 |
| KSC LC 39A | 0.772727 | KSC LC 39A | 0.772727 |
| VAFB SLC 4E | 0.769231 | VAFB SLC 4E | 0.833333 |

- Success rates at CCAFS SLC-40 and VAFB SLC-4E increase if the early testing flights are removed from the calculation, as KSC LC-39A was not yet a launch site its success rate remains the same
    - Ranking site by success rate remains the same

# Payload Mass: Mass Range by Block

- Block 4 (vs Block 5):
- 0 to 2,500 kg
  - 2.22% (5.55%)
- 2,500 to 5,000 kg
  - 4.44% (10%)
- 5,000 to 7,500 kg
  - 3.33% (7.78%)
- 7,500 to 10,000 kg
  - 2.22% (3.33%)
- 10,000 kg plus
  - 0% (16.66%)
  -



# Payload Mass: Mass Range by Orbit

- Highest Percentage of Flights withing a given range were to orbit:

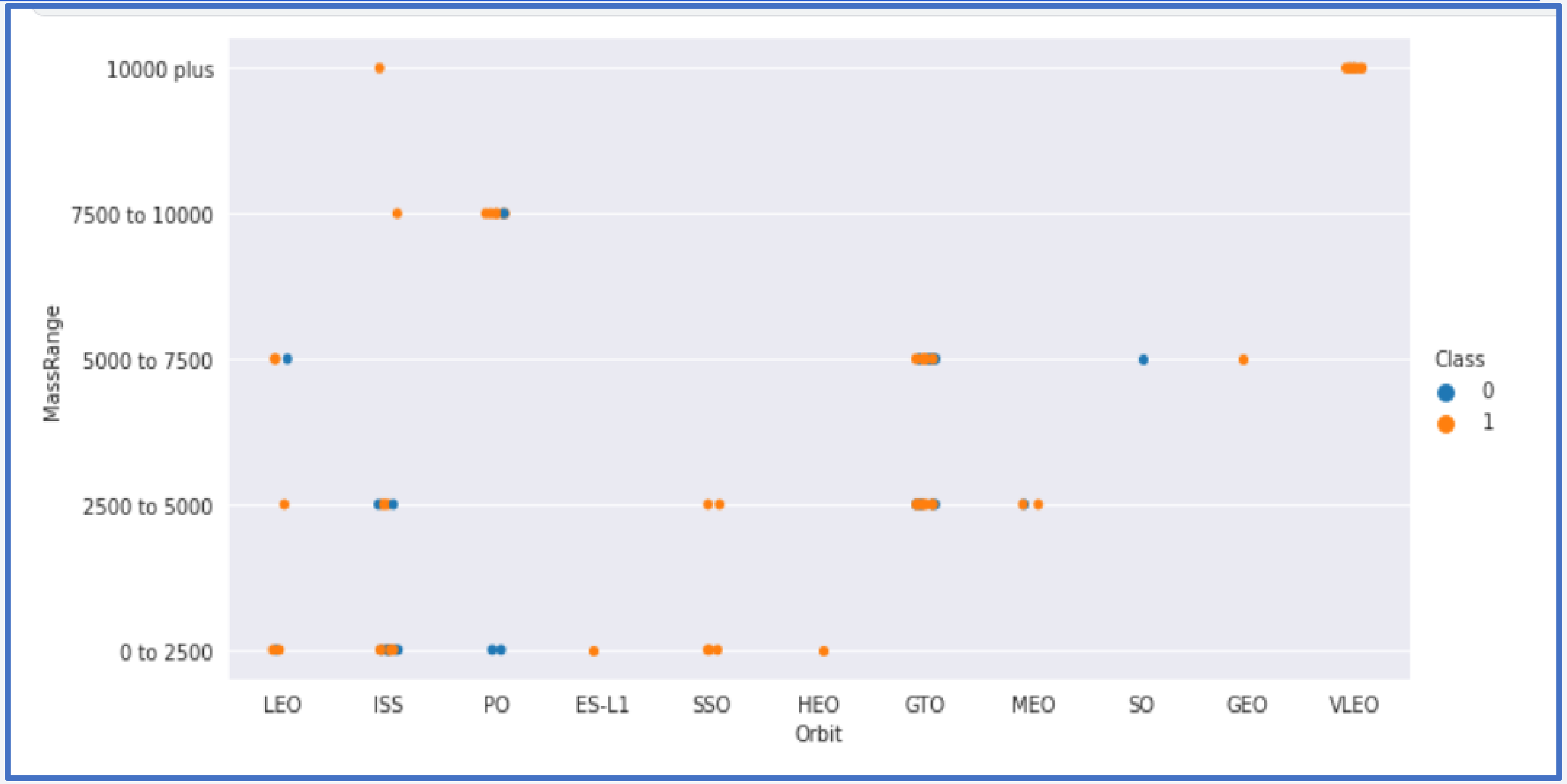
- 0 to 2,500 kg
  - ISS 14.44%

- 2,500 to 5,000 kg
  - GTO 14.44%

- 5,000 to 7,500 kg
  - GTO 15.56%

- 7,500 to 10,000 kg
  - PO 7.78%

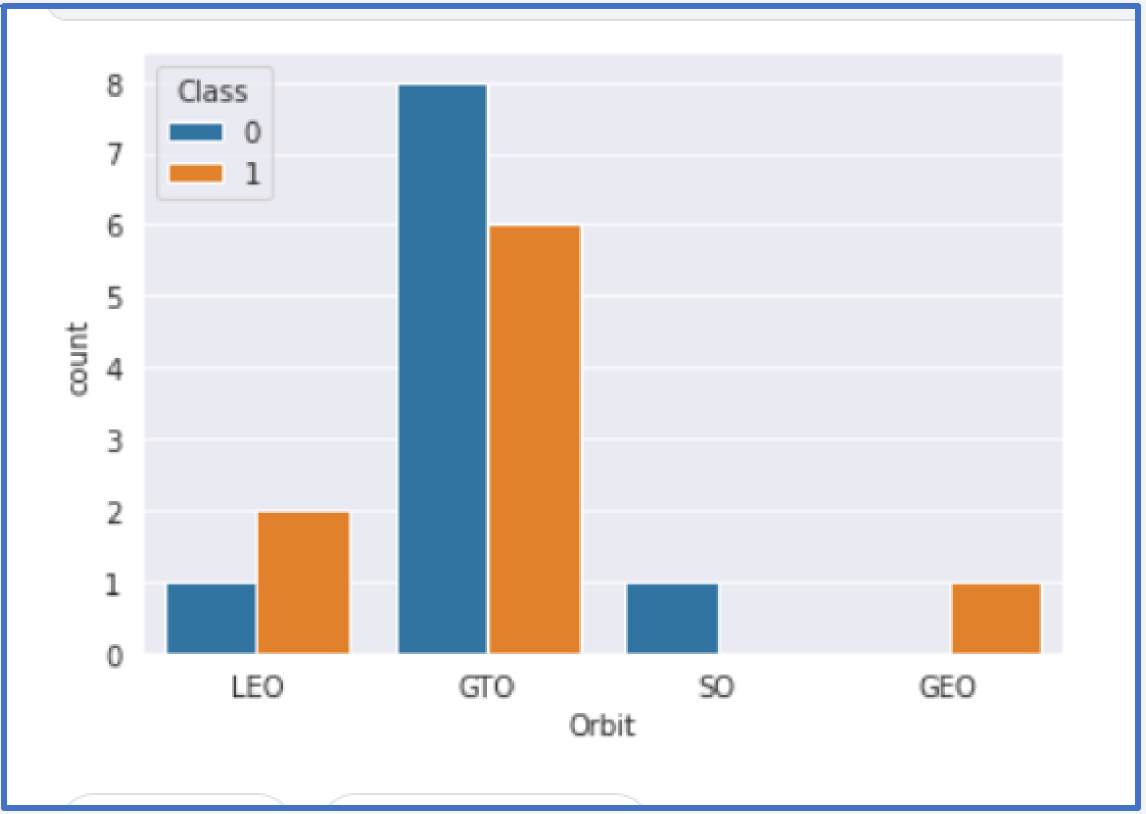
- 10,000 kg plus
  - VLEO 15.55%



# Payload vs. Orbit Type: Payload Range 5,000 to 7,500 kg

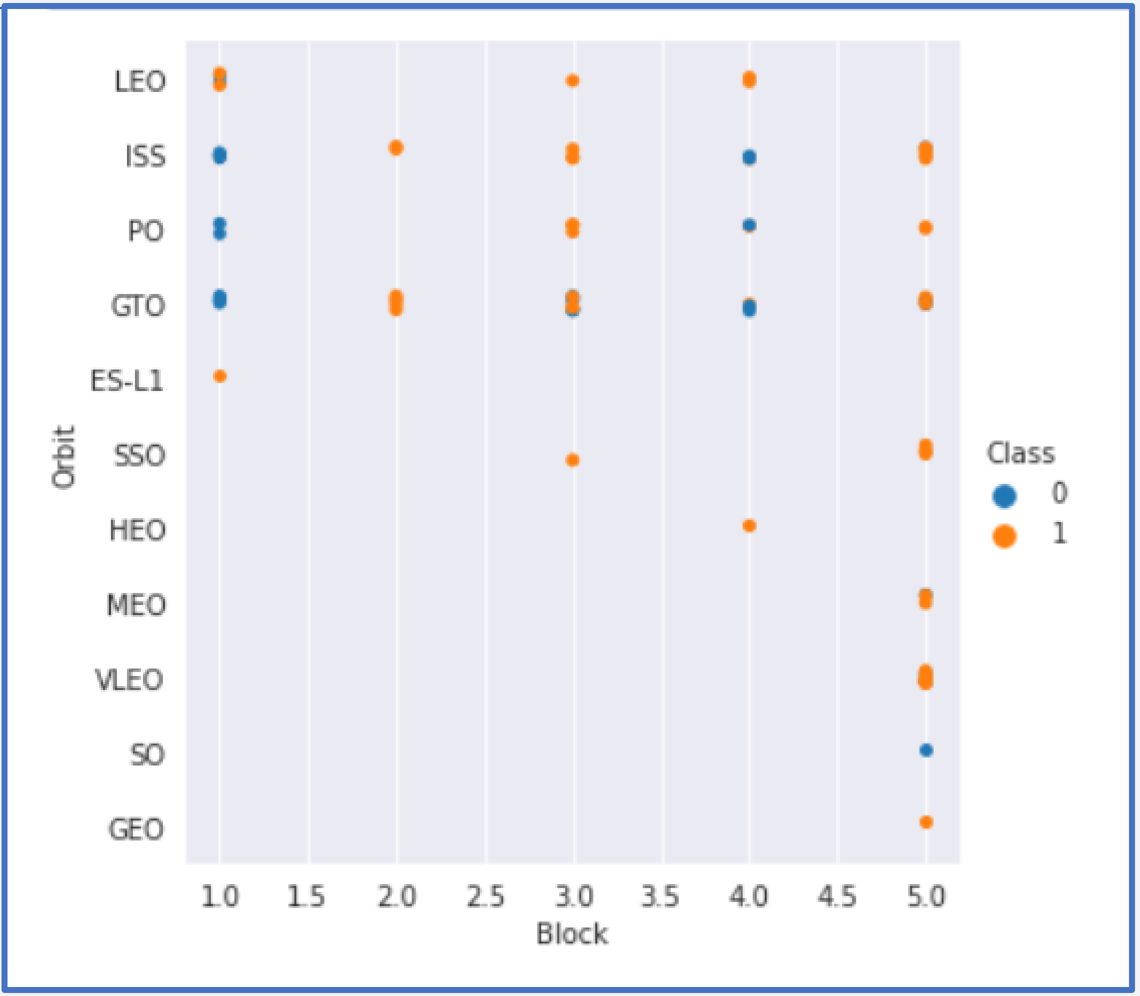
- Looking at payloads between 5,000 kg to 7,500 kg and GTO:
  - 6 successful landings
  - 8 failed landings
  - Success rate 42.86%

- Overall GTO
  - Success rate 51.85%



# Orbit vs Block by Landing Outcome

- Orbits with Block 4 missions:
  - GTO 3 flights
    - Success rate: 33.33% - Low
  - ISS 3 flights
    - Success rate: 33.33% - Low
  - LEO 2 flights
    - Success rate: 100%
  - PO 2 flights
    - Success rate: 50%
  - HEO 1 flight
    - Success rate: 100%



# Thank you!