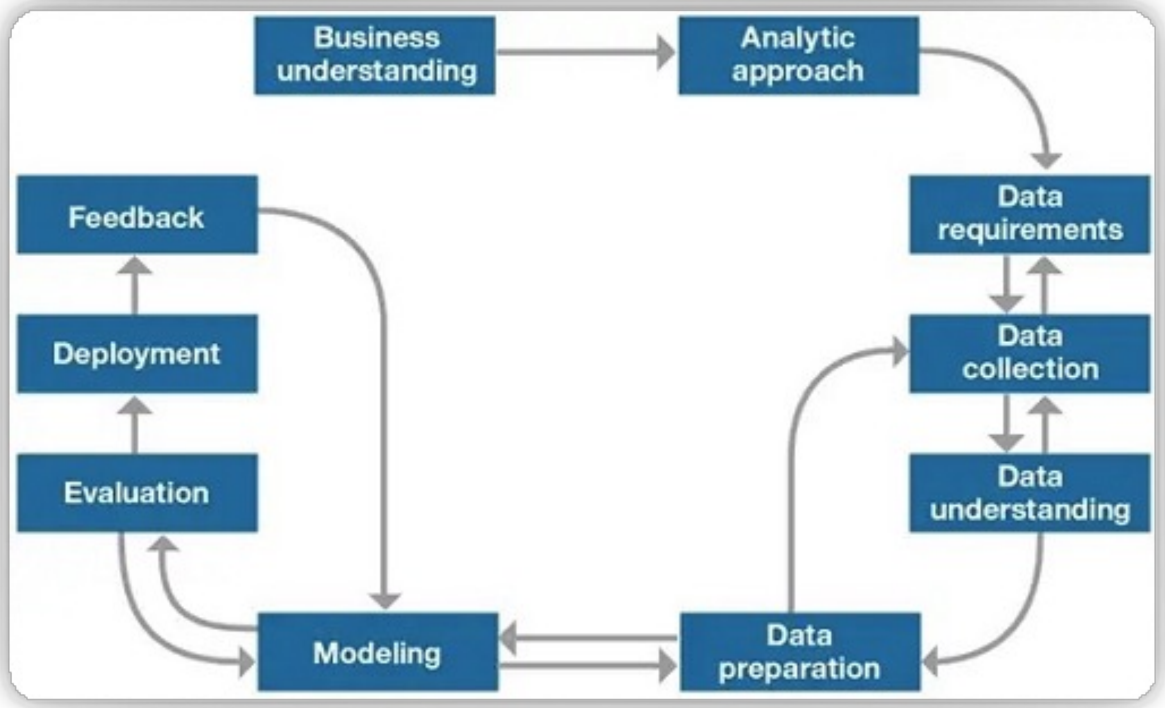
IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Matt Caron
Jan 2023

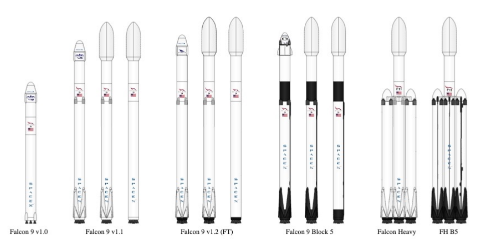
Project Outline

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



Foundational methodology for data science (Source: John Rollins, SDS IBM)

Executive Summary



According the Space Foundation, the global space industry was worth approximately $447 billion in 2020. Industry analysts estimate, **SpaceX** generated over $2 billion in revenue in 2020 at an estimated 40% profit.

SpaceX is the only company that employs re-usable rocket hardware and is their core value of their business model. Over time, SpaceX has innovated many Falcon 9 rocket and operational features (see figure 1).

SpaceX advertises Falcon 9 mission **cost $62 million**. Much of the savings is because Space X can reuse the first stage. To compete, SpaceY (Tiercel) rocket program will need to meet key features and predict landing outcomes.

We use **"Foundational Methodology for Data Science"**(John Rollins, SDS IBM). It consists of five phases: Frame the problem, Collect the data, Explore the data, Model the data, and Communicate the results - the methodology is iterative. A supervised classification model using machine learning pipelines, when tuned and refined, we found **Logistic Regression predicted historical landings Falcon 9 boosters with 0.94 accuracy rate** (see details figure 2.).

We use this to be confident of the key features and outcomes for SpaceY Tiercel rocket program.

Findings from our study of key rocket features, program resources and target financials necessary for competitive business as an alternate supplier, SpaceY must meet the following:

- **Payload Capacity** - **engines** capacity of **22,800 kg to LEO and 8,300 kg to GTO orbits**.
- **Rocket Reuse** - **landing legs and grid fins**, thermal protection system, **avionics** systems for reusability, reliability, and maintainability.
- **Operational Facilities** - **lower latitude, oceanside launch sites** with close **landing proximities**. And close **road /rail networks** for effective movement of mission hardware, near to key human resources and suppliers.
- **Pricing** – with rapid re-use model of operations SpaceY Tiercel missions must be a **maximum of $62 million for each flight**.

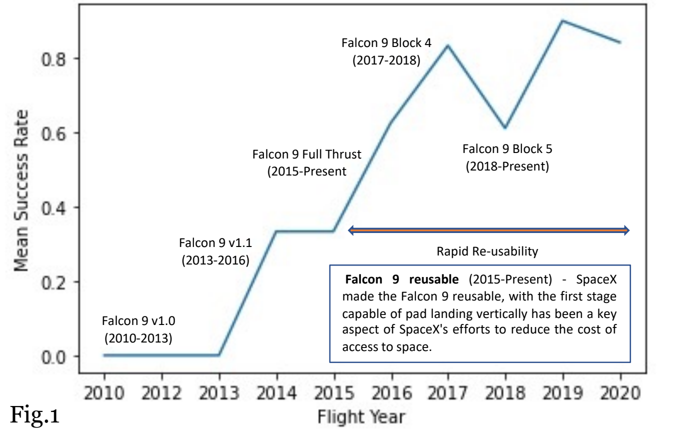
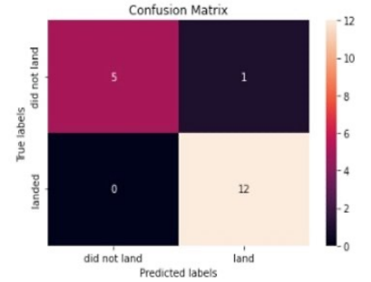


Fig.1



Precision: Out of all the flights that the model predicted would land, **92%** actually did.

Recall: Out of all the flights that actually did land, the model predicted the actual landing correctly for **100%** of those flights.

F1 Score calculated as: 2 * (Precision * Recall) / (Precision + Recall) = **0.96**.

Logistic Regression Metrics Report

	precision	recall	f1-score	support
Did Not Land	1.00	0.83	0.91	6
Landed	0.92	1.00	0.96	12
accuracy			0.94	18
macro avg	0.96	0.92	0.93	18
weighted avg	0.95	0.94	0.94	18

SVM Metrics Report

	precision	recall	f1-score	support
Did Not Land	0.50	0.50	0.50	6
Landed	0.75	0.75	0.75	12
accuracy			0.67	18
macro avg	0.62	0.62	0.62	18
weighted avg	0.67	0.67	0.67	18

Decision Tree Metrics Report

	precision	recall	f1-score	support
Did Not Land	0.83	0.83	0.83	6
Landed	0.92	0.92	0.92	12
accuracy			0.89	18
macro avg	0.88	0.88	0.88	18
weighted avg	0.89	0.89	0.89	18

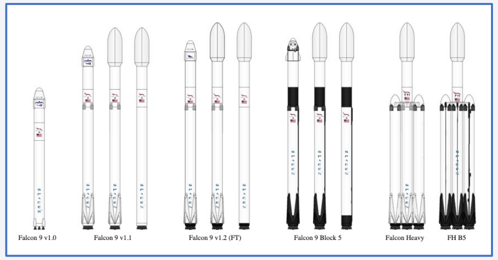
KNN Metrics Report

	precision	recall	f1-score	support
Did Not Land	0.50	0.50	0.50	6
Landed	0.75	0.75	0.75	12
accuracy			0.67	18
macro avg	0.62	0.62	0.62	18
weighted avg	0.67	0.67	0.67	18

scikit-learn.org

Fig.2

Executive Summary



Global space industry worth $447 billion in 2020, in 2020 SpaceX generated over $2 billion revenue at 40% profit.

SpaceX's produces the only reusable rocket so that Falcon 9 mission cost $62 million vs. industry cost of up to $165 million.

For competitive advantage, SpaceY's Tiercel missions must have key features that produce successful landing outcomes. Using the Foundational Methodology for Data Science, a supervised classification model, machine learning pipeline predicted historical landings Falcon 9 boosters with 0.94 accuracy rate.

SpaceY, Tiercel must meet key features of the model: Payload Capacity, Rocket Reuse, Operational Facilities, and Pricing to be a competitive alternate supplier:

- **Payload Capacity**: engine thrust/fuel, 22,800 kg to LEO, 8,300 kg to GTO.
- **Rocket Reuse**: Landing legs, grid fins, thermal protection, and avionics systems for reusability, reliability, and maintainability.
- **Operational Facilities**: Lower latitude, oceanside launch sites with close landing proximity, near key infrastructure and resources.
- **Pricing**: SpaceY Tiercel missions must be a maximum of $62 million per flight with rapid reuse operations.

The Logistics Regression Model predicted with accuracy of 94% (more detail fig2.)

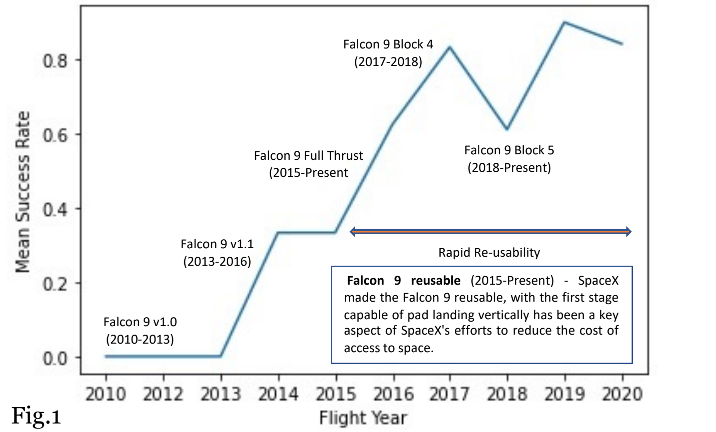
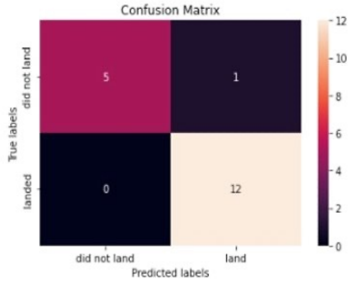


Fig.1



Precision: Out of all the flights that the model predicted would land, **92%** actually did.

Recall: Out of all the flights that actually did land, the model predicted the actual landing correctly for **100%** of those flights.

F1 Score calculated as: 2 * (Precision * Recall) / (Precision + Recall) = **0.96**.

Logistic Regression Metrics Report

	precision	recall	f1-score	support
Did Not Land	1.00	0.83	0.91	6
Landed	0.92	1.00	0.96	12
accuracy			0.94	18
macro avg	0.96	0.92	0.93	18
weighted avg	0.95	0.94	0.94	18

SVM Metrics Report

	precision	recall	f1-score	support
Did Not Land	0.50	0.50	0.50	6
Landed	0.75	0.75	0.75	12
accuracy			0.67	18
macro avg	0.62	0.62	0.62	18
weighted avg	0.67	0.67	0.67	18

Decision Tree Metrics Report

	precision	recall	f1-score	support
Did Not Land	0.83	0.83	0.83	6
Landed	0.92	0.92	0.92	12
accuracy			0.89	18
macro avg	0.88	0.88	0.88	18
weighted avg	0.89	0.89	0.89	18

KNN Metrics Report

	precision	recall	f1-score	support
Did Not Land	0.50	0.50	0.50	6
Landed	0.75	0.75	0.75	12
accuracy			0.67	18
macro avg	0.62	0.62	0.62	18
weighted avg	0.67	0.67	0.67	18

scikit-learn.org

Fig.2

Introduction

In this capstone project, we use Data Science methods to predict if a Falcon 9 first stage will land successfully. Using SpaceX mission records, and public information we must target a competitive flight pricing for the rocket Space Y - headed by founder and CEO, Allon Mask.

SpaceX advertises Falcon 9 rocket launches cost 62 million dollars while other providers cost up to 165 million dollars each. Much of the cost savings is because SpaceX can quickly reuse the rocket's booster. Therefore, to determine the cost of a launch, we will employ Machine Learning to calculate the probability of a first stage landing.

To understand the key conditions when a first stage most likely to land successfully, the project will train a machine learning model and public information. Using IBM Watson Studio, we collect SpaceX market and technical data, extract key features using python tools and employ machine learning tools to predict first stage outcomes.

Finally, we conclude with the technical features, resources and target cost for SpaceY's business model as an alternate supplier for the space industry.

https://github.com/mattjcaron/Applied-Data-Science-Capstone/blob/master/README.md

Launch Financial Breakdown	**Cost**
Flight Revenue (New Booster)	$62 million
Flight Revenue (Restored Booster)	$50 million
Total Marginal Cost (per extra flight)	$12 million
SpaceX Revenue (for two flights)	$112 million
Net Profit (for two flights)	$47 million (42% profit)

Section 1

Methodology

Data Science Methodology - Executive Summary

- **Data collection:** Python "**request**" SpaceX R for EST API HPTT. Use **BeautifulSoup** for Web Scraping HTML historical Wikipedia tables .
- **Data wrangling: Pandas** and **NumPy** for Exploratory Data Analysis (EDA) and determine training labels and classify outcomes as "1" or "0".
- **Exploratory Data Analysis (EDA): SQL** queries from tables loaded in **DB2** tables to understand key statistics and **Matplotlib** visualization of patterns and trends for EDA and feature engineering .
- **Interactive visual analytics: Folium** for important **geo-locations/patterns** and **Plotly Dash** to interactively explore relationships key mission features.
- **Predictive analysis:** Supervised classification methods to create a machine learning pipeline that best predict if the first stage will land; finding best Hyperparameter for SVM, Classification Trees and Logistic Regression.

We iteratively build, tune, evaluate classification models split the data into training testing data, train different classification models, hyperparameter grid search and score model.

Data Collection

1. **SpaceX REST API** - SpaceX Falcon 9 launch data including:

- Rocket used,
- Payload delivered,
- Launch specifications,
- Landing specifications,
- Landing outcome.

SpaceX REST API
- Request/Get SpaceX data
- Parse key feature into dataframe

SpaceX REST API
- Filter data keeping only Falcon 9 rocket missions

SpaceX REST API
- Identify missing values
- Calculate/replace missing values with mean

2. **Web Scraping** - HTTPS Wiki pages using **BeautifulSoup** to extract **HTML tables** that contain Falcon 9 launch records, and **parse** data from tables and convert to Pandas data frame:

- Flight No., Date and time (UTC),
- Version, Booster [b],
- Launch site,
- Payload[c], Payload mass,
- Orbit and Customer,
- Launch outcome, Booster landing.

Web Scraping
- Request/Get Wiki HTML data
- Create BeautifulSoup Object

Web Scraping
- Extract Featured Column/Variable Names

Web Scraping
- Create DF by parsing HTML table.

1. **Data collection: Python "request"** SpaceX R for EST API HPTT. Use **BeautifulSoup** for Web Scraping HTML historical Wikipedia tables .

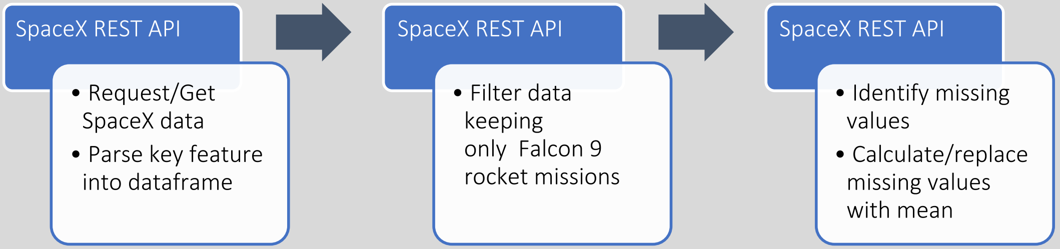
2. **Data wrangling: Pandas** and **NumPy** for Exploratory Data Analysis (EDA) and determine training labels and classify outcomes as "1" or "0".

Data Collection – SpaceX API

SpaceX API:

1. Define **REST API URL** endpoint/parameters of request.
   - Send server GET request using **requests.get()** method and get response.
   - Check request using **response.status_code**.
   - Parse JSON data using **response.json()**.
   - Store **pd.DataFrame()**.
2. Filter ‘**BoosterVersion**’ to **data_falcon9** where: != ‘Falcon_1’
3. Replace ‘**PayloadMass**’ missing values using .**mean()** and .**replace()** for **np.nan** values with the mean.

SpaceX REST API, key phrases and flowcharts:



- Click for GitHub URL of the completed SpaceX API calls notebook.

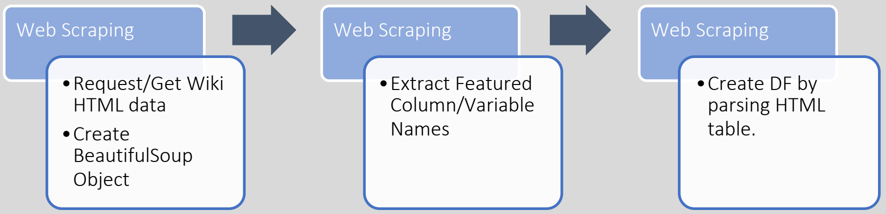
https://github.com/mattjcaron/Applied-Data-Science-Capstone/blob/master/Data Collection SpaceX API.ipynb

Data Collection - Scraping

Web Scraping:

1. Define **Web Page URL**.
2. Send server HTTP GET request HTML using **requests.get()**.
3. Create **soup** object from **response** using **Beautiful Soup**.
4. Find element data to scrape using **soup.find_all()**.
5. Extract data from elements using **extract_column_from_header()**.
6. Create launch_dict with **dict.fromkeys()**.
7. **Parse data elements to launch_dict**
8. Create df using **pd.DataFrame()**.

The Web Scraping process key phrases and flowcharts:



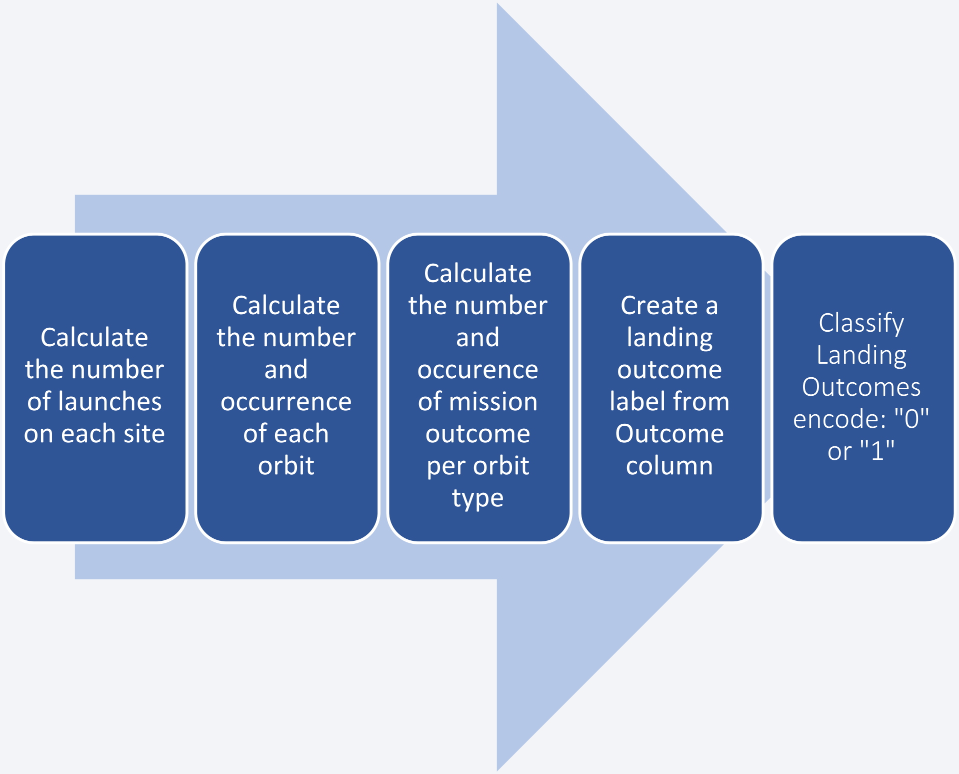
Flow Chart

https://github.com/mattjcaron/Applied-Data-Science-Capstone/blob/master/Data Collection Web Scraping.ipynb

Data Wrangling

Exploratory Data Analysis (EDA) to find data patterns and define training labels for supervised models (classification).

- Exploratory Data Analysis Calculate:
  - Number of Launch Sites
  - Number and occurrence of orbit
  - Number and occurrence of mission outcome per orbit
- Determine and Training Labels:
  - Classify landing data categories as "good' landing as 1 or "bad" landing 0 list
  - Create landing outcome column in our data frame



https://github.com/mattjcaron/Applied-Data-Science-Capstone/blob/master/Data Wrangling.ipynb

Key method phrases: pd.read_csv(), df.isnull(), df.dtype, df.value_counts(), enumerate()

EDA with Data Visualization

Key features of the data are used to classify our model and predict mission outcomes. Using our SpaceX database, we visualize data and perform feature engineering;

Data Visualizations - relationships between:

- Flight Number/Payload Mass – key time series feature shows for booster design improvements.
- Flight Number/Launch Site – key time series feature shows capability facility and location.
- Payload/Launch Site – key booster performance feature for facility and location.
- Success rate of each Orbit type – booster capability for orbit requirements.
- Flight Number and Orbit type – shows time series improvement for booster improvements.
- Payload and Orbit type – capability of all boosters for all orbit requirements.
- Launch success yearly trend..
- Pair plot key features sns.pairplot()

Features Engineering – process key features:

- Create dummy variables to categorical columns
- Cast all numeric columns to float6

https://github.com/mattjcaron/Applied-Data-Science-Capstone/blob/master/EDA Visual Features.ipynb

EDA with SQL

We use IBM Db2 and SQL libraries to survey our data set with summary of SQL queries and displays:

1. Unique launch sites names.
2. Five launch sites 'CCA' records.
3. Total payload mass for NASA (CRS).
4. Average payload mass by booster version F9 v1.1.
5. Date of first successful landing on a ground pad.
6. Booster drone ship success and payload mass > 4000 and < 6000 kg.
7. Total number of successful and failure mission outcomes.
8. Names of 'booster_versions' which carried the maximum payload mass.
9. Failed 'landing_outcomes" in drone ship, booster versions, and launch site names for 2015.
10. Rank the count of landing outcomes (Failure (drone ship), Success (ground pad) between 2010-06-04 and 2017-03-20 in descending order.

https://github.com/mattjcaron/Applied-Data-Science-Capstone/blob/master/EDA SQL.ipynb

Build an Interactive Map with Folium

Optimal launch site locations done by analyzing existing site results and proximities using Folium map data objects:

1: **Mark all launch sites on a map.**
2: Mark the success/failed launches for each site on the map.
3: Calculate the distances between a launch site to important proximities.

Map objects for booster landing success/re-use featured:

Launch site proximities with success/fail markers:

- **Orbital -** advantage long./lat. to equator
- **Coastline** – close/safe recovery and re-use.

Distance lines proximities people, materials and technology.

- **Railways -** heavy transport of hardware.
- **Highways -** light transportation for materials and people.
- **Coastline -** safe efficient launch and landing.
- **Cities -** safe efficient distance, for personnel and services

https://github.com/mattjcaron/Applied-Data-Science-Capstone/blob/master/Interactive Visual Analytics Folium.ipynb

Build a Dashboard with Plotly Dash

We designed and launched interactive plots and graphs with live interactions to visualize mission relationships. Using the dashboard, we obtain insights to the following success rates.

We want to see which:

1. Site has the largest number successful launches.
2. Site has the highest launch success rate.
3. Payload range(s) has the highest launch success rate.
4. Payload range(s) has the lowest launch success rate.
5. F9 Booster version (v1.0, v1.1, FT, B4, B5, etc.) has the highest success rate?

We added plots/interactions to visually facilitate finding key trends and relationships within for our prediction model.

https://github.com/mattjcaron/Applied-Data-Science-Capstone/blob/master/Dash Interactive Model.ipynb

Predictive Analysis (Classification)

Create a machine learning pipeline to predict if the first stage will land given the data collected.

Preprocess the Falcon 9 Data Set
- Classify Launch, "Y"
- Transform Data, "X"
- Train Test Split

Grid Search CV LR,SVM,DT,KNN
- GridSearch best parameters
- Test CV Accuracy
- Check Confusion Matrix

Evaluate LR, SVM,DT KNN
- Compare model perormance
- Generate report and comparison graphs.

Each method's best hyperparameters and accuracy scores are used for model selection – LR, SVN, DT KNN.

Preprocess the Falcon 9 Data Set
- Load Data Frame "X":
  - X =pd.read_csv('https://...)
  - Classify Launch, "Y" Y= pd.Series(data['Class'])
  - Y= Y.to_numpy()
- Transform "X";
  - transform = preprocessing.StandardScaler()
  - X = transform.fit_transform(X)
- Train Test Split X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.2, random_state=2)

Grid Search CV: LR,SVM,DT,KNN – LR example
- logreg = LogisticRegression()
- logreg_cv = GridSearchCV(logreg, parameters, cv=10).fit(X_train, Y_train)
- logreg_acc = logreg_cv.score(X_test, Y_test)
- logreg_yhat = logreg_cv.predict(X_test)
- logreg_yhat = logreg_cv.predict(X_test)
- plot_confusion_matrix(Y_test, logreg_yhat)

Evaluate Model: LR,SVM,DT,KNN, Classification Report and Graph
- target_names = ['Did Not Land', 'Landed']
- logreg_report = classification_report(Y_test, logreg_yhat, target_names=target_names)
- print('\t\t Logistic Regression Metrics Report')
- print(logreg_report,'\n')
- svm_report = classification_report(Y_test, svm_yhat, target_names=target_names)
- print('\t\t SVM Metrics Report')
- print(svm_report,'\n')
- tree_report = classification_report(Y_test, tree_yhat, target_names=target_names)
- print('\t\t Decision Tree Metrics Report')
- print(tree_report,'\n')
- knn_report = classification_report(Y_test, knn_yhat, target_names=target_names)
- print('\t\t KNN Metrics Report')
- print(svm_report,'\n')
- Graph comparing accuracy

https://github.com/mattjcaron/Applied-Data-Science-Capstone/blob/master/Machine Learning Prediction.ipynb

EDA with Data Visualization

Using our collected SpaceX database, we visualize the data and perform feature engineering used to classify our model and predict mission outcomes.

Data Visualizations - relationships between:

- Flight Number and Payload Mass – shows increased booster capability and success over time.
- Flight Number and Launch Site – shows geography overtime for site selection.
- Payload and Launch Site – shows geographic results for booster design capability.
- Success rate of each orbit type – orbits accessibility rates.
- Flight Number and Orbit type - shows improved outcomes for higher orbits overtime
- Payload and Orbit type – booster capabilities for missions.
- Launch success yearly trend – booster capability overtime.
- Booster version and Payload Capacity – increasing capability with versions.
- Pair plot key features sns.pairplot().

Features Engineering – process key features:

- Create dummy variables for modelling categorical columns.
- Cast all numeric columns to float6 for better data processing.

https://github.com/mattjcaron/Applied-Data-Science-Capstone/blob/master/EDA Visual Features.ipynb

Results

<u>Exploratory Data Analysis Results</u> key patterns/trends for feature engineering and modelling successful landings:

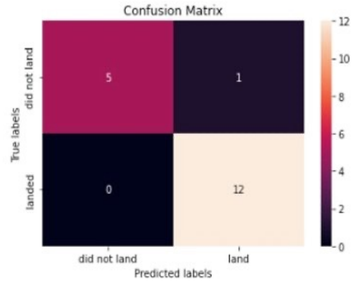
- Time based analysis using flight number with launch site, orbits and payload mass.
- Average success rate for flight number booster version and orbit were explored.

<u>Interactive Analytics</u> for key machine learning features our model:

- Confirming key patterns for classification modelling include operational proximities to resources include important features such as:
  - launch sites close to access orbital latitudes, existing facilities purpose-built for space programs, transportation networks for the movement of materials and people and safe distance from cities.
- Dashboard analysis of key features from exploratory analysis confirm important relationships between launch sites, payload mass and booster versions.
- Pair-plot matrix for numerical features were analyzed to confirm key patterns and trends for landing outcomes seeing outcome distribution and scatter plots in one graphic.

<u>Predictive Analysis Results</u>:

Key features were preprocessed and tune validate prediction model for successful landings.



Precision: Out of all the flights that the model predicted would land, **92%** actually did.

Recall: Out of all the flights that actually did land, the model predicted the actual landing correctly for **100%** of those flights.

F1 Score calculated as: 2 * (Precision * Recall) / (Precision + Recall) = **0.96**.

Logistic Regression Metrics Report

	precision	recall	f1-score	support
Did Not Land	1.00	0.83	0.91	6
Landed	0.92	1.00	0.96	12
accuracy			0.94	18
macro avg	0.96	0.92	0.93	18
weighted avg	0.95	0.94	0.94	18

SVM Metrics Report

	precision	recall	f1-score	support
Did Not Land	0.50	0.50	0.50	6
Landed	0.75	0.75	0.75	12
accuracy			0.67	18
macro avg	0.62	0.62	0.62	18
weighted avg	0.67	0.67	0.67	18

Decision Tree Metrics Report

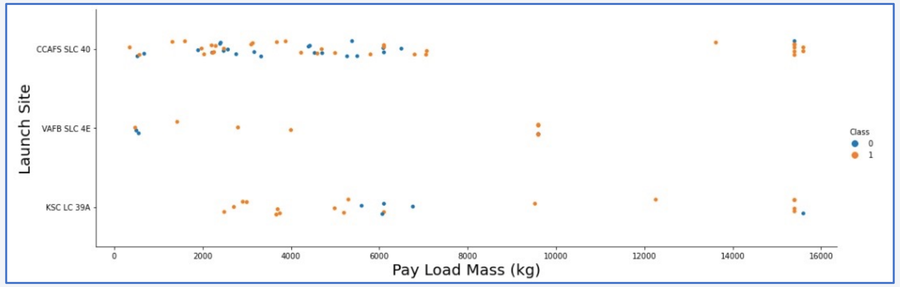
	precision	recall	f1-score	support
Did Not Land	0.83	0.83	0.83	6
Landed	0.92	0.92	0.92	12
accuracy			0.89	18
macro avg	0.88	0.88	0.88	18
weighted avg	0.89	0.89	0.89	18

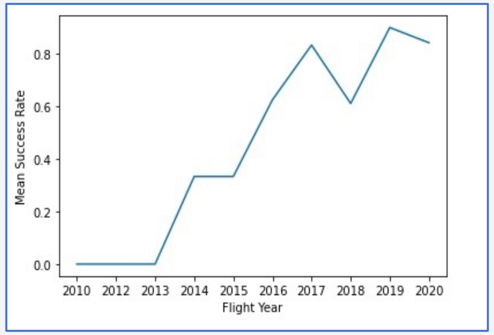
KNN Metrics Report

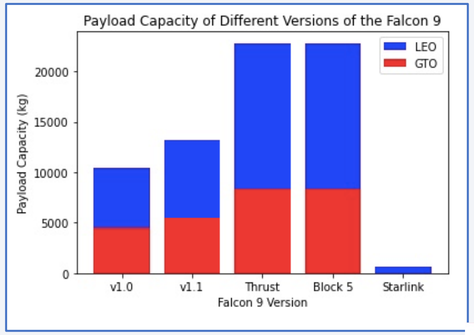
	precision	recall	f1-score	support
Did Not Land	0.50	0.50	0.50	6
Landed	0.75	0.75	0.75	12
accuracy			0.67	18
macro avg	0.62	0.62	0.62	18
weighted avg	0.67	0.67	0.67	18

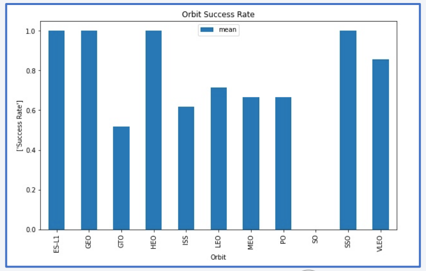
scikit-learn.org

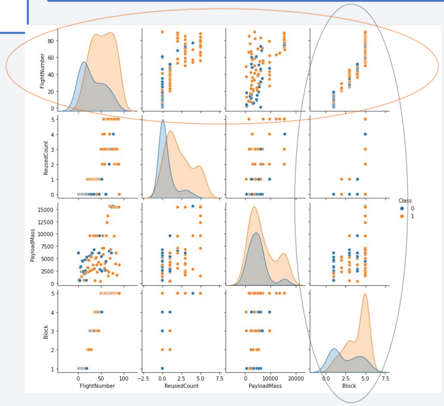
EDA with Data Visualization

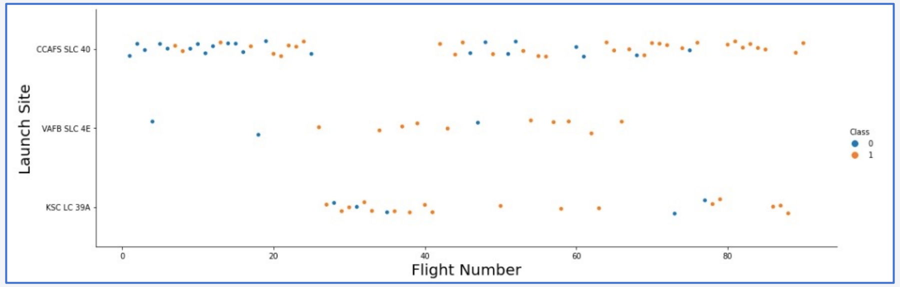


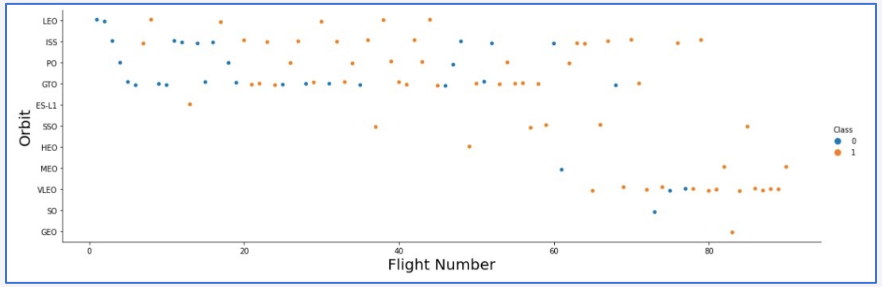


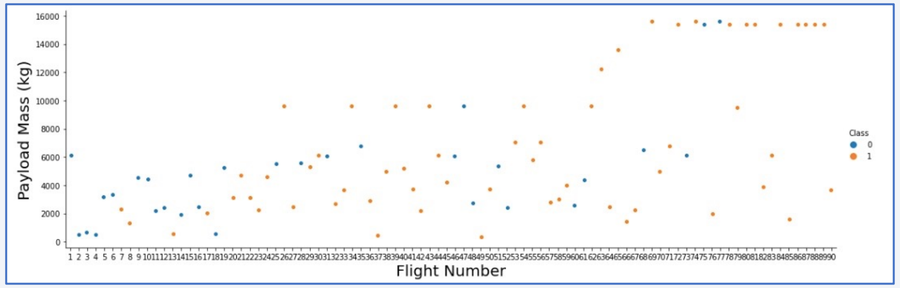


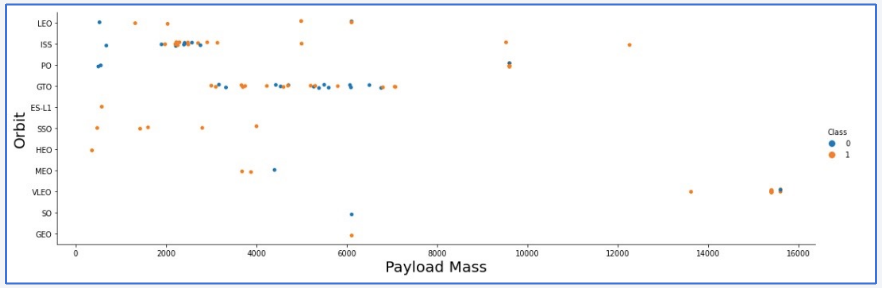


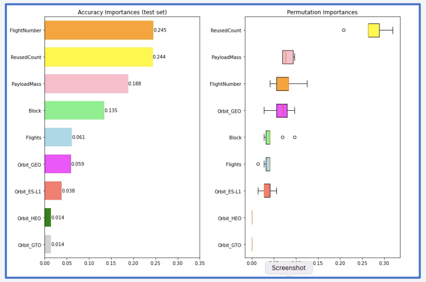








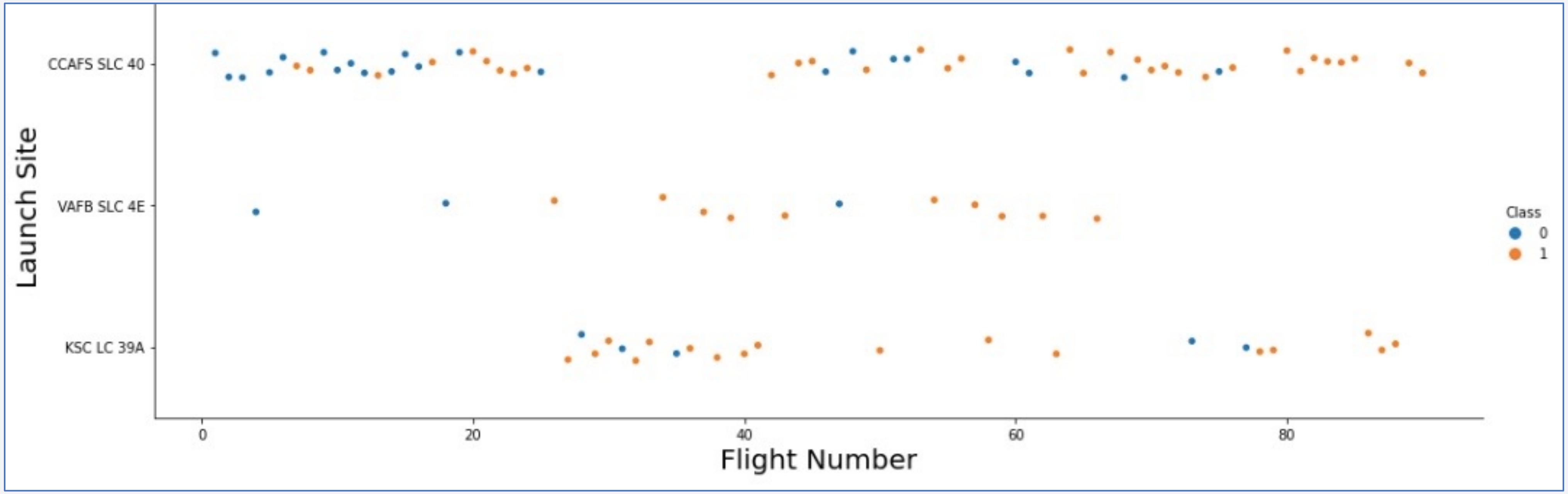




Section 2

Insights drawn from EDA

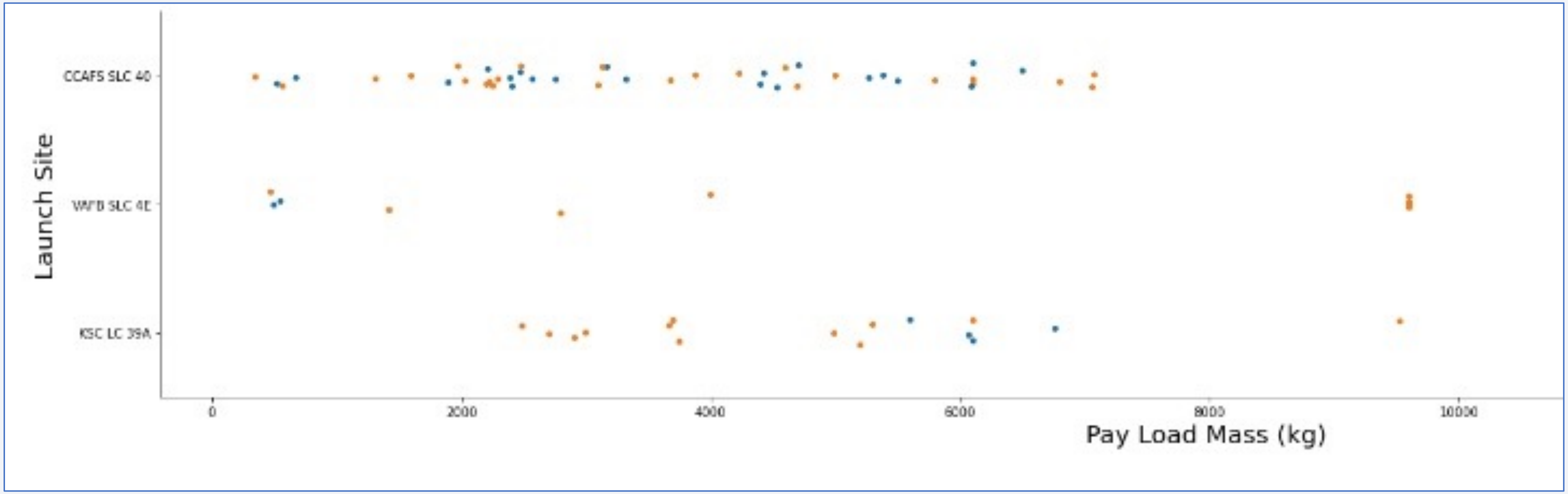
Flight Number vs. Launch Site



We see that outcomes for these launch sites have improved success rates over time:

- CCAFS SLC 40 has the highest flight count.
- KSC LC 39A began use at around flight number 30.
- VAFB has the fewest launches.

Payload vs. Launch Site



Scatter plot shows:

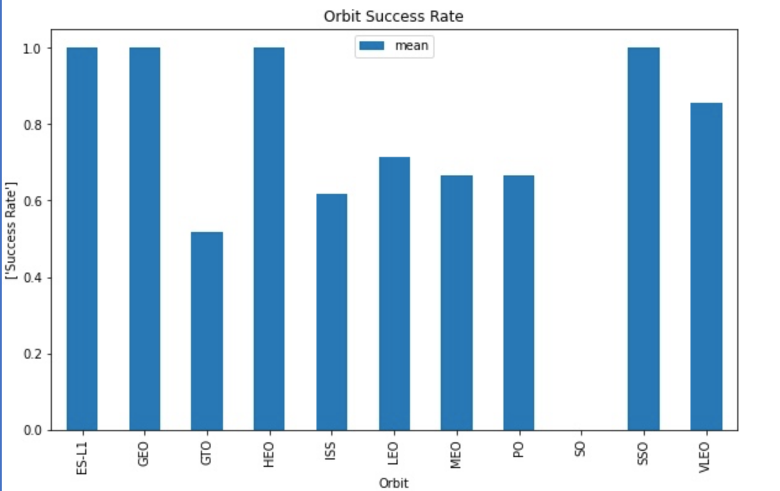
- Better outcomes at greater than 6000kg payload mass all site, 100% at 15,00kg (logarithmic).
- high success for lower payloads at KSC LC 39A.

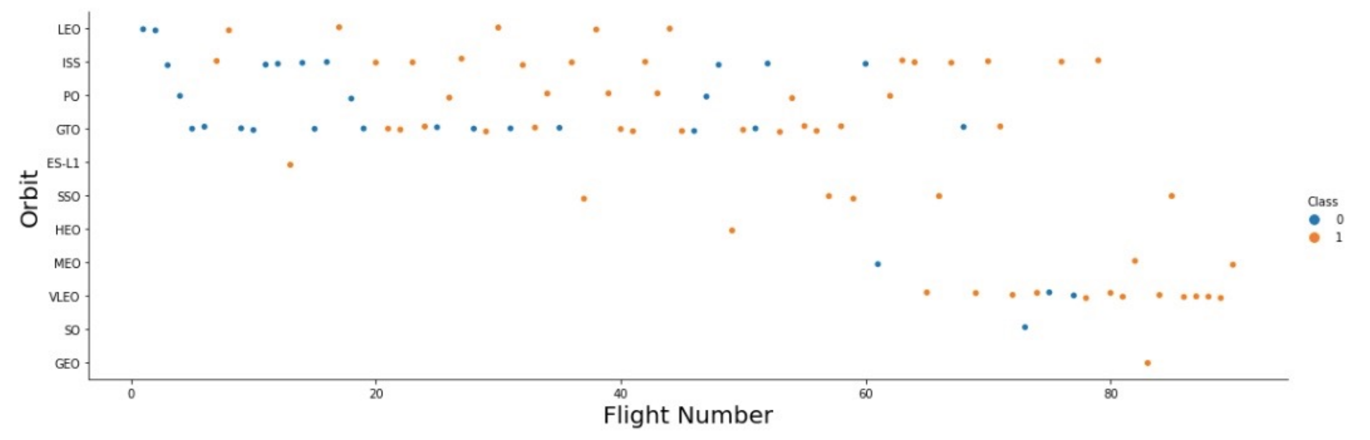
Success Rate vs. Orbit Type

Success rates for lower orbits like VLEO and LEO have much better success rates than higher orbits like GTO.

Note that ES-L1, GEO, HEO orbits have less than three or missions and do not have enough data to conclude rates confidently.

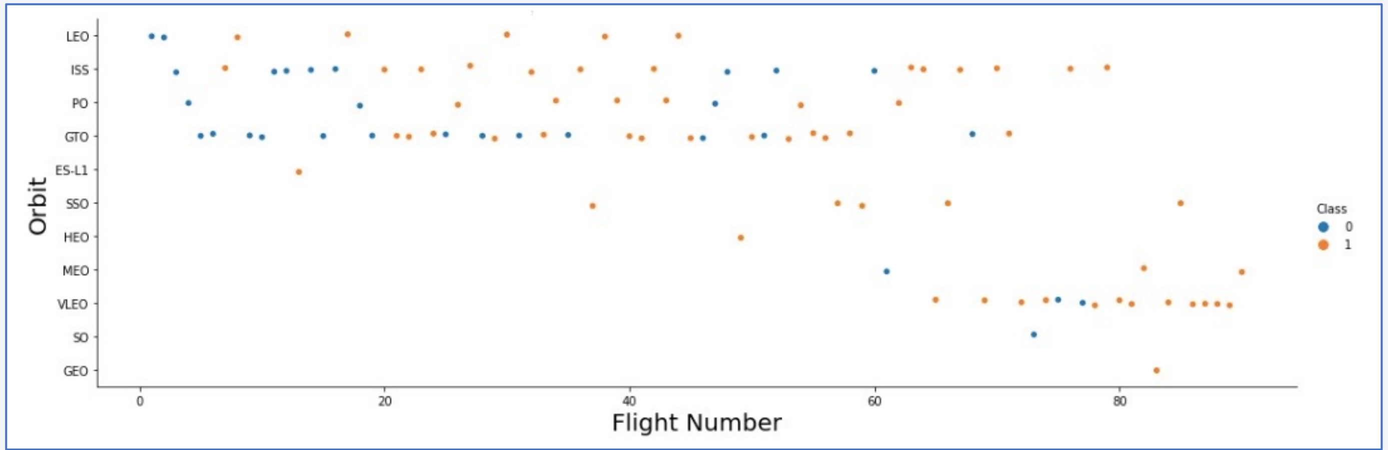
Orbit	count	mean	std	min	25%	50%	75%	max
ES-L1	1.0	1.000000	NaN	1.0	1.0	1.0	1.0	1.0
GEO	1.0	1.000000	NaN	1.0	1.0	1.0	1.0	1.0
GTO	27.0	0.518519	0.509175	0.0	0.0	1.0	1.0	1.0
HEO	1.0	1.000000	NaN	1.0	1.0	1.0	1.0	1.0
ISS	21.0	0.619048	0.497613	0.0	0.0	1.0	1.0	1.0
LEO	7.0	0.714286	0.487950	0.0	0.5	1.0	1.0	1.0
MEO	3.0	0.666667	0.577350	0.0	0.5	1.0	1.0	1.0
PO	9.0	0.666667	0.500000	0.0	0.0	1.0	1.0	1.0
SO	1.0	0.000000	NaN	0.0	0.0	0.0	0.0	0.0
SSO	5.0	1.000000	0.000000	1.0	1.0	1.0	1.0	1.0
VLEO	14.0	0.857143	0.363137	0.0	1.0	1.0	1.0	1.0





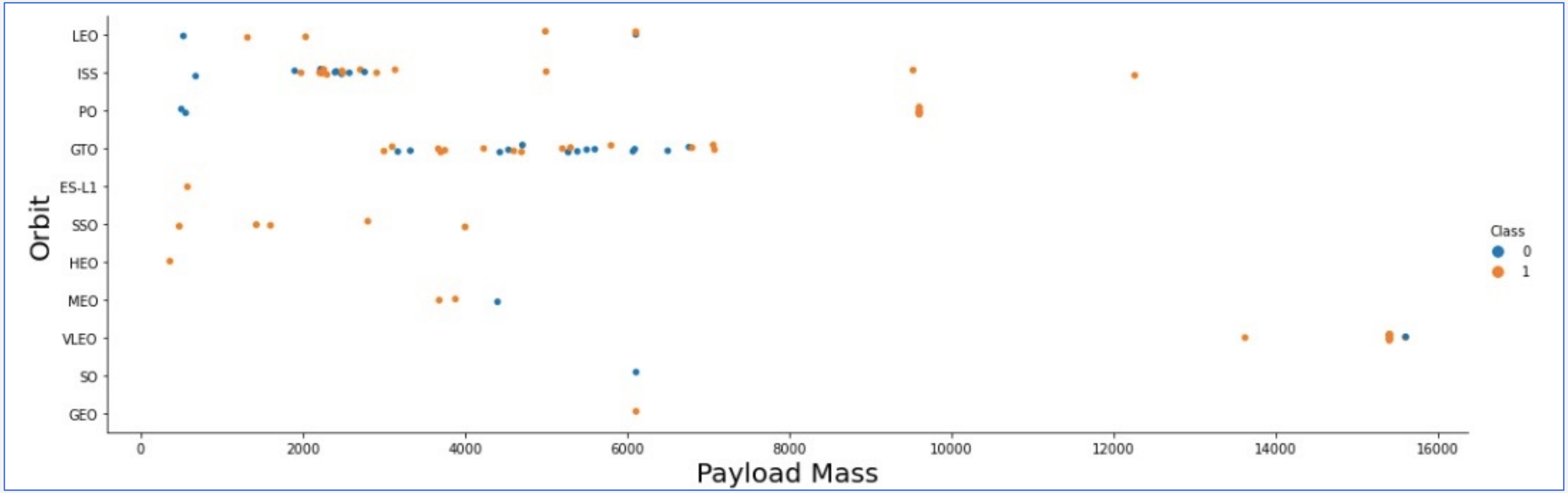
Flight Number vs. Orbit Type

Orbit	count	mean	std	min	25%	50%	75%	max
ES-L1	1.0	1.000000	NaN	1.0	1.0	1.0	1.0	1.0
GEO	1.0	1.000000	NaN	1.0	1.0	1.0	1.0	1.0
GTO	27.0	0.518519	0.509175	0.0	0.0	1.0	1.0	1.0
HEO	1.0	1.000000	NaN	1.0	1.0	1.0	1.0	1.0
ISS	21.0	0.619048	0.497613	0.0	0.0	1.0	1.0	1.0
LEO	7.0	0.714286	0.487950	0.0	0.5	1.0	1.0	1.0
MEO	3.0	0.666667	0.577350	0.0	0.5	1.0	1.0	1.0
PO	9.0	0.666667	0.500000	0.0	0.0	1.0	1.0	1.0
SO	1.0	0.000000	NaN	0.0	0.0	0.0	0.0	0.0
SSO	5.0	1.000000	0.000000	1.0	1.0	1.0	1.0	1.0
VLEO	14.0	0.857143	0.363137	0.0	1.0	1.0	1.0	1.0



- Success rate have improved for All Orbits over time.
- More missions are launched at VLEO after flight number 60.

Payload vs. Orbit Type

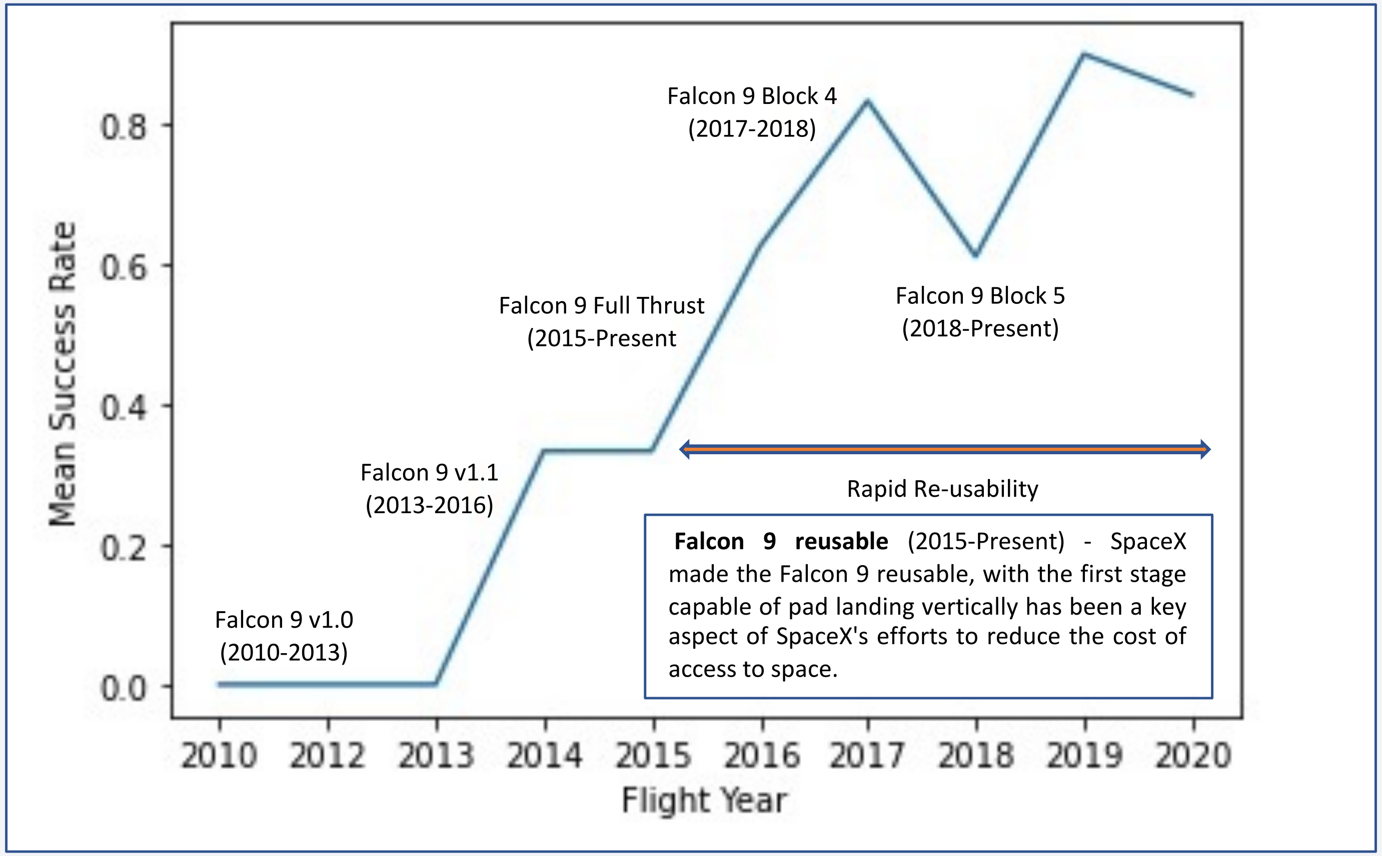


Scatter point of payload vs. orbit type:

- High success rates for greater than 8,000 kg
- High success for 0kg-400kg SSO

Launch Success Yearly Trend

Yearly average success rate with a strong improvement trends in flight success from 2013 through the present. Improvement rate coincides with improved rocket capabilities for all orbits like payload mass and booster re-use (see version blocks).



All Launch Site Names

There are 4 unique launch sites used by SpaceX for the Falcon 9 rocket.

Launch Site Names:
CCAFS LC-40
CCAFS SLC-40
KSC LC-39A
VAFB SLC-4E

Launch Site Names Begin with 'CCA'

Five records beginning with CCA:	booster_version
CCAFS LC-40	F9 v1.0 B0003
CCAFS LC-40	F9 v1.0 B0004
CCAFS LC-40	F9 v1.0 B0005
CCAFS LC-40	F9 v1.0 B0006
CCAFS LC-40	F9 v1.0 B0007

- We find 5 records where launch sites begin with `CCA`
- Present your query result with a short explanation here
- %sql select LAUNCH_SITE as "Five records beginning with CCA:", Booster_Version from SPACEXTBL where (LAUNCH_SITE) like 'CCA%' limit 5;

Total Payload Mass

Total Payload (kg) for NASA (CRS):
45596

Total payload carried by boosters from NASA:

- Query shows the extent of missions with NASA.
- %sql select sum (PAYLOAD_MASS__KG_) as "Total Payload (kg) for NASA (CRS):" from SPACEXTBL where Customer = 'NASA (CRS)';

Average Payload Mass by F9 v1.1

Average Payload (kg) for F9 v1.1:
2928

Calculated the average payload mass carried by booster version F9 v1.1

- Results show the payload average is low relative to success rate for recent launches.
- %sql select avg(PAYLOAD_MASS__KG_) as "Average Payload (kg) for F9 v1.1:" from SPACEXTBL where (Booster_Version) like 'F9 v1.1';

First Successful Ground Landing Date

First Successful Landing Outcome in Ground Pad:
2015-12-22

Dates of the first successful landing outcome on ground pad

- We see the ground pad data starts with Falcon 9 at the end of 2015.
- %sql select Min(DATE) as "First Successful Landing Outcome in Ground Pad:" from SPACEXTBL where (LANDING__OUTCOME) = 'Success (ground pad)'

Successful Drone Ship Landing with Payload between 4000 and 6000

Boosters Success with 4000kg - 6000kg:
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

Names of boosters versions which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

- F9 BT version ship landing has only four successes.
- %sql select BOOSTER_VERSION as "Boosters Success with 4000kg - 6000kg:" from SPACEXTBL where LANDING__OUTCOME = 'Success (drone ship)' and PAYLOAD_MASS__KG_ between 4000 and 6000

Total Number of Successful and Failure Mission Outcomes

Mission Outcomes	Count
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

- Total number of successful and failure mission outcomes
- Falcon 9 has very high mission success rate in comparison to landings success.

Boosters Carried Maximum Payload

F9 B5 boosters which have successfully landed on drone ship and had payload mass maximum.

The B5 boosters are with maximum payload can be most cost effect.

%sql Select Booster_Version as "Booster Version", PAYLOAD_MASS__KG_ as "Maximum Payload Mass" from Spacextbl where PAYLOAD_MASS__KG_ = (select MAX(PAYLOAD_MASS__KG_) from SPACEXTBL)

Booster Version	Maximum Payload Mass
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

2015 Launch Records

Failed landing outcomes on drone ship, booster versions, and launch site names for first in year 2015.

First year failures on drone ship with F9 v1.1

Landing Outcomes	Booster Version	Launch Sites	Dates 2015
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40	2015-01-10
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40	2015-04-14

- %sql select Landing__Outcome as "Landing Outcomes", Booster_Version as "Booster Version", Launch_site as "Launch Sites", Date as "Dates 2015" from Spacextbl where date between '2015-01-01' and '2015-12-31' and Landing__Outcome = 'Failure (drone ship)';

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

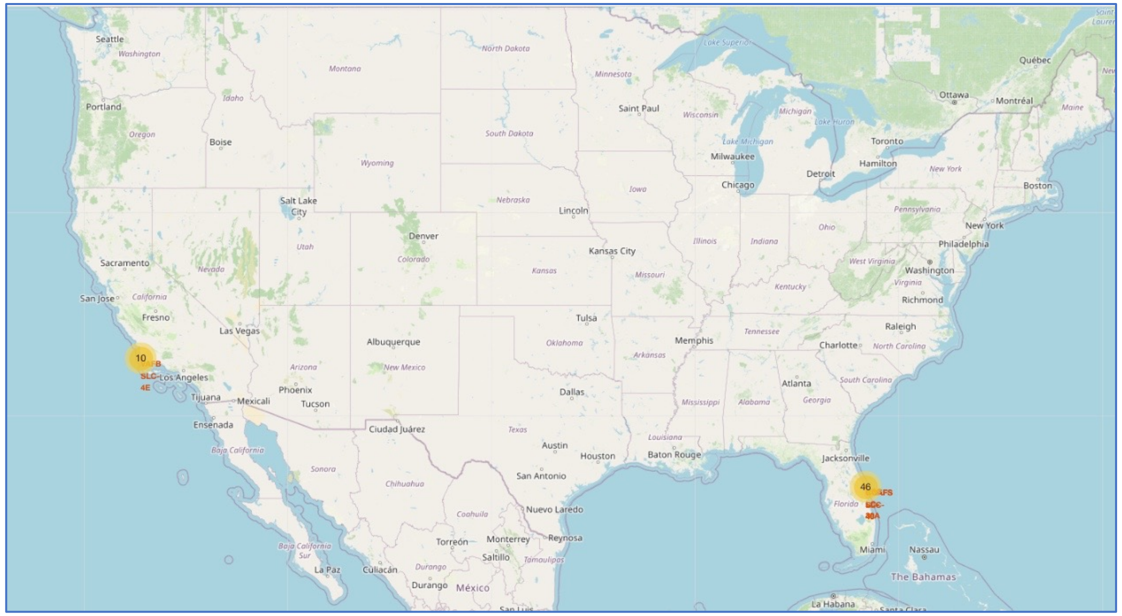
- Rank count of landing outcomes Failure/Success. between the date 2010-06-04 and 2017-03-20, in descending order
- Present your query result with a short explanation here
- sql select LANDING__OUTCOME as "Landing Outcomes", COUNT (LANDING__OUTCOME) AS "Count" from Spacextbl where date between '2010-06-04' and '2017-03-20' group by LANDING__OUTCOME

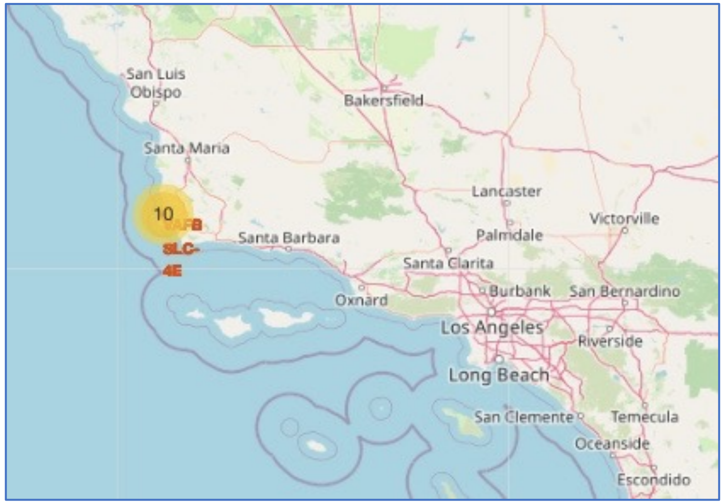
Landing Outcomes	Count
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

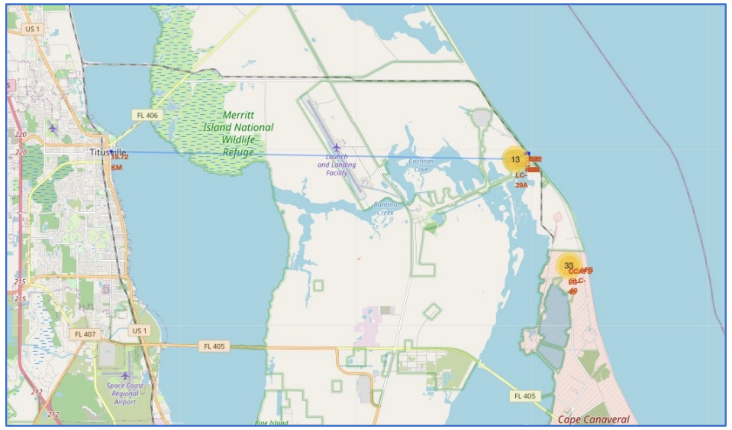
Section 3

Launch Sites Proximities Analysis

Regional and National Mapping of Launch Site Locations







Launch Sites are located:

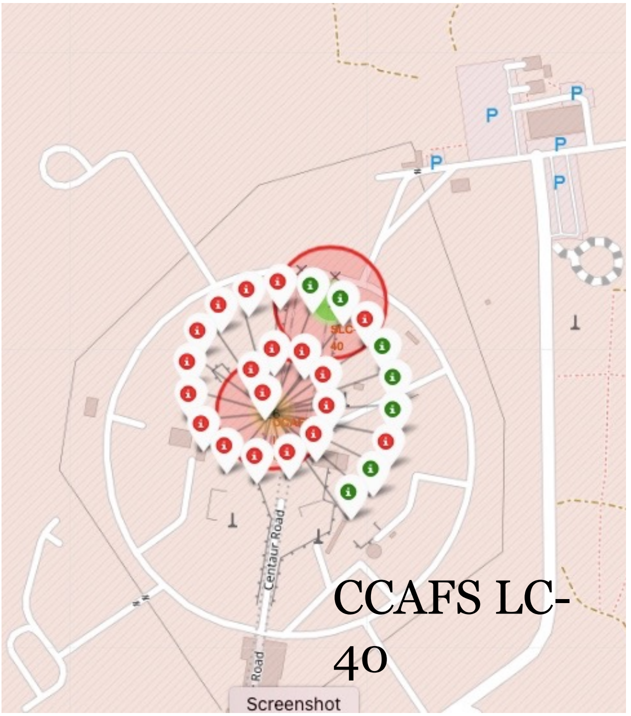
- Southern regions of US,
- Coastal proximity,
- Two sites in Florida 46 Missions,
- One site California 10 missions.

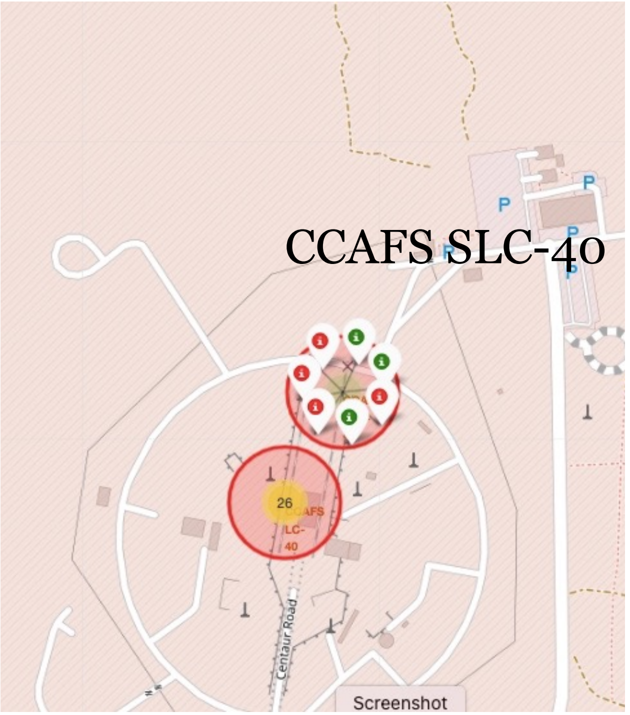
Launch Pad Outcomes

Patterns/trends found in Launch Site Geo's and proximities:

- Mission sites are at long established aerospace facilities.
- Optimal position for target orbits.
- Ocean and land-based landings are available.
- Infrastructure and resources for movement of heavy hardware are established.

* Green = Landed, Red = Did Not Land

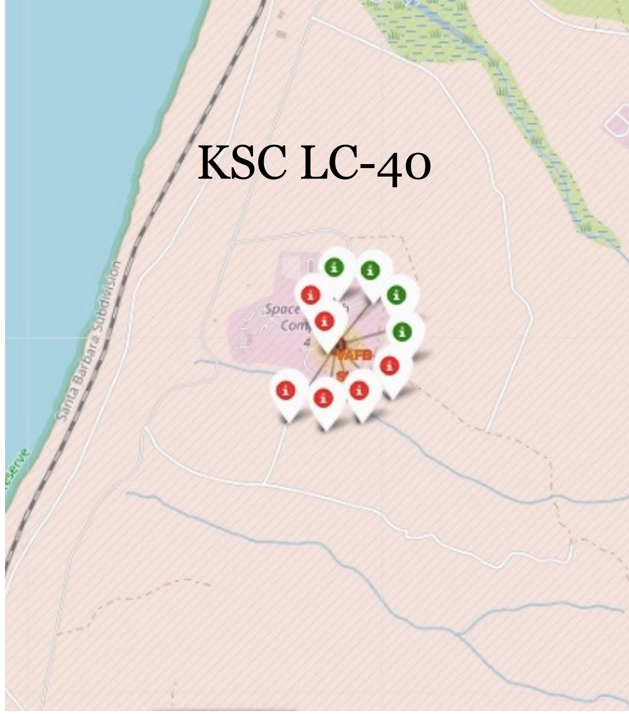


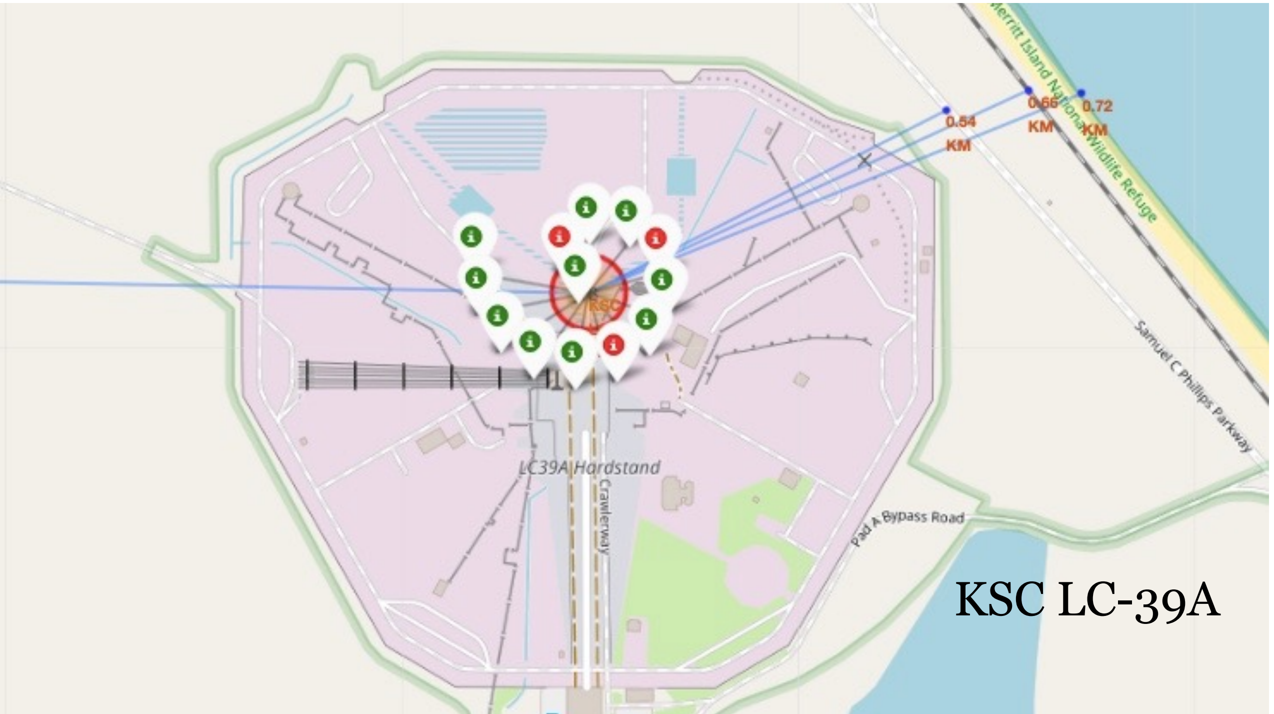


KSC LC-39A:
- Most successful outcomes,
- closest proximities to the ocean,
- near roads, rail and nearest to city.

CCAFS SLC/LC-40:
- Most flights, lower success rate
- Close to roads/rail.

KSC LC-40:
- Least flights and lower success rate,
- Close to roads, rail and ocean.
- Not close to cities.



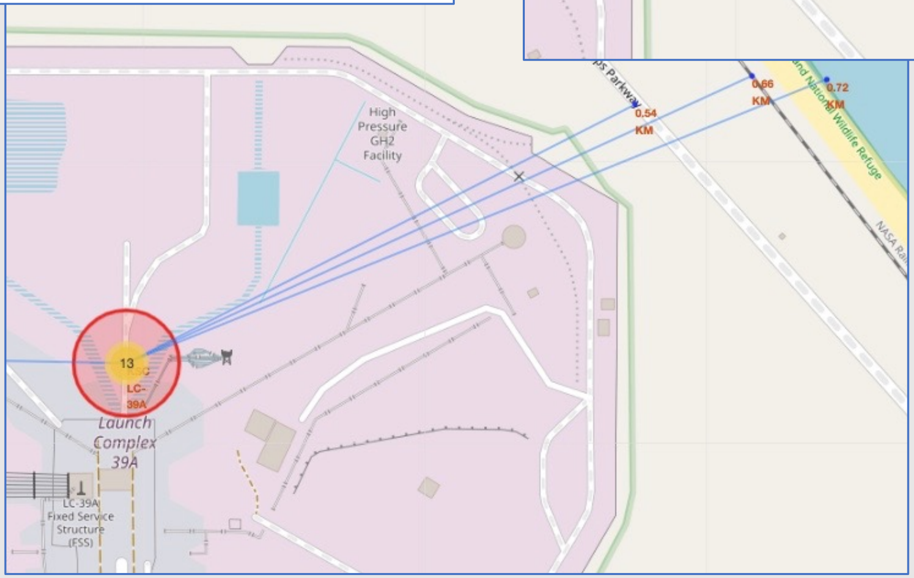


Key Launch Site Proximities

KSC LC-39A

- 19.72 km from Titusville FL
- 0.54 km from nearest road
- 0.66 km from rail line
- 0.72 km from coastline

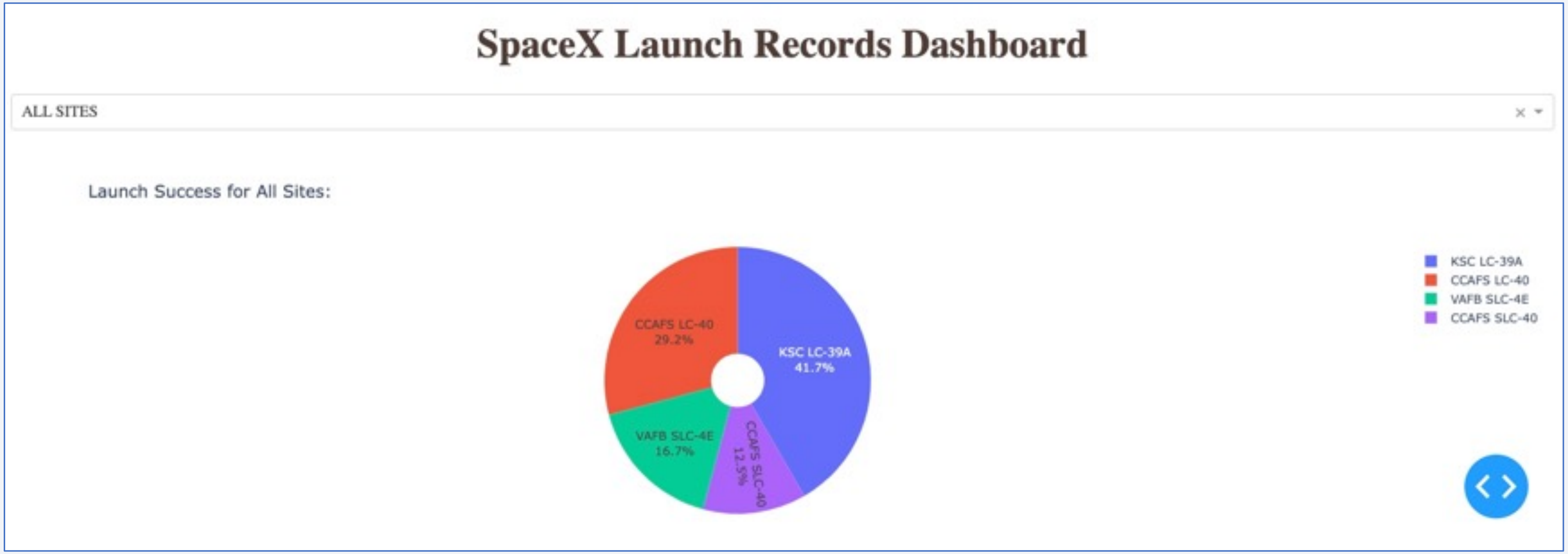
Rapid reuse of boosters requires close proximities to resources, hardware transportation and landing facilities.



Section 4

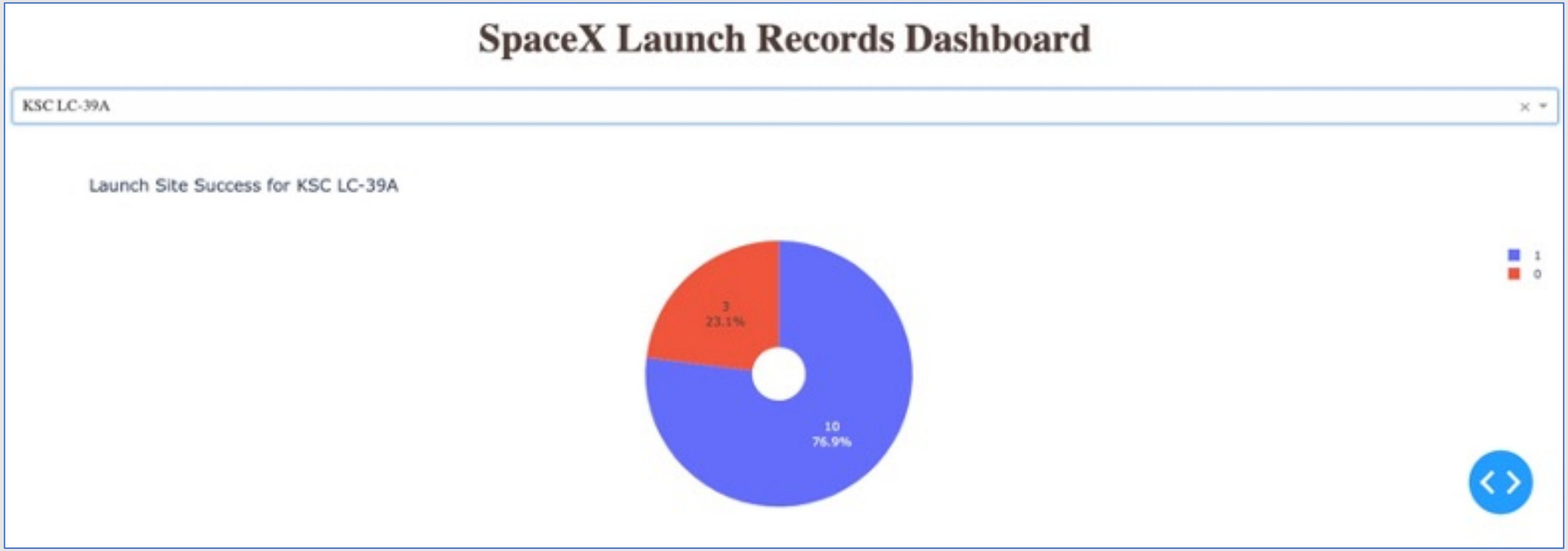
Build a Dashboard with Plotly Dash

Dashboard Screenshot – Launch Site Success Rate



KSC LC - 39A has best launch site success 41.7% of all successful launches.

Dashboard Screenshot – Best Launch Site Success Rate



Most Successful Launch Site, KSC LC-39A, 76.9% of flights.

Payload vs. Launch Outcomes: 2,000 - 5,000 kg for all booster versions

The scatter plot shows high success rates for FT, B4 and B5.

Section 5

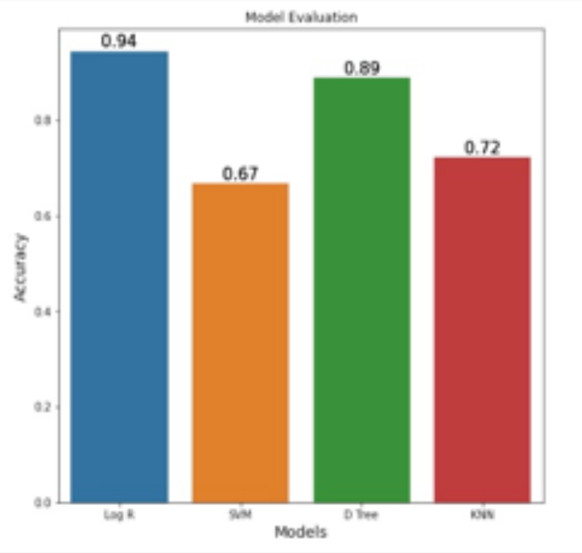
Predictive Analysis (Classification)

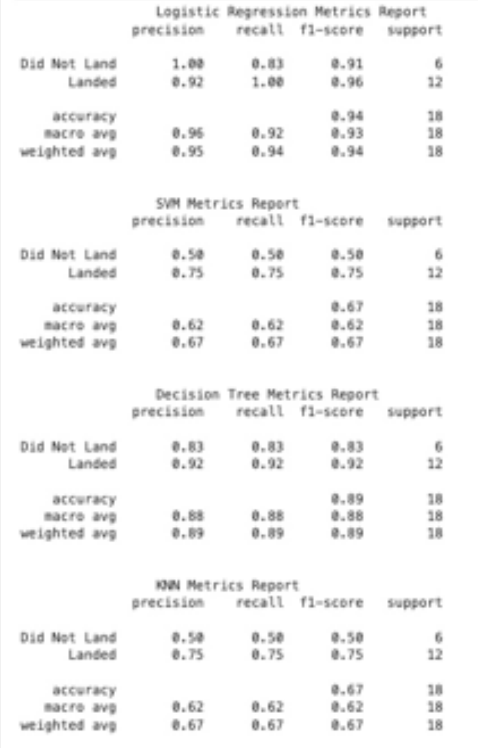
Classification Accuracy

Bar Chart, Model Accuracy for all built **models.**

Logistic Regression, **best method for** predicting successful landings.

Model Accuracy = 0. 94




Logistic Regression Metrics Report

	precision	recall	f1-score	support
Did Not Land	1.00	0.83	0.91	6
Landed	0.92	1.00	0.96	12
accuracy			0.94	18
macro avg	0.96	0.92	0.93	18
weighted avg	0.95	0.94	0.94	18

SVM Metrics Report

	precision	recall	f1-score	support
Did Not Land	0.50	0.50	0.50	6
Landed	0.75	0.75	0.75	12
accuracy			0.67	18
macro avg	0.62	0.62	0.62	18
weighted avg	0.67	0.67	0.67	18

Decision Tree Metrics Report

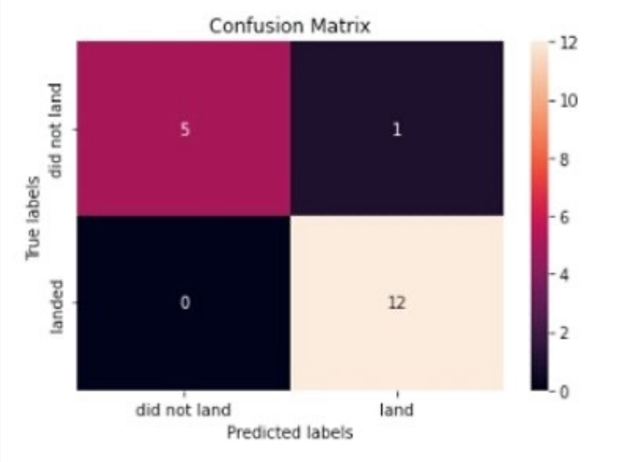
	precision	recall	f1-score	support
Did Not Land	0.83	0.83	0.83	6
Landed	0.92	0.92	0.92	12
accuracy			0.89	18
macro avg	0.88	0.88	0.88	18
weighted avg	0.89	0.89	0.89	18

KNN Metrics Report

	precision	recall	f1-score	support
Did Not Land	0.50	0.50	0.50	6
Landed	0.75	0.75	0.75	12
accuracy			0.67	18
macro avg	0.62	0.62	0.62	18
weighted avg	0.67	0.67	0.67	18

Confusion Matrix

To predict successful landings, Logistic Regression is our best performing model, with accuracy at 94%.



Precision: Out of all the flights that the model predicted would land, **92%** actually did.

Recall: Out of all the flights that actually did land, the model predicted the actual landing correctly for **100%** of those flights.

F1 Score calculated as: 2 * (Precision * Recall) / (Precision + Recall) = **0.96**.

Logistic Regression Metrics Report

	precision	recall	f1-score	support
Did Not Land	1.00	0.83	0.91	6
Landed	0.92	1.00	0.96	12
accuracy			0.94	18
macro avg	0.96	0.92	0.93	18
weighted avg	0.95	0.94	0.94	18

SVM Metrics Report

	precision	recall	f1-score	support
Did Not Land	0.50	0.50	0.50	6
Landed	0.75	0.75	0.75	12
accuracy			0.67	18
macro avg	0.62	0.62	0.62	18
weighted avg	0.67	0.67	0.67	18

Decision Tree Metrics Report

	precision	recall	f1-score	support
Did Not Land	0.83	0.83	0.83	6
Landed	0.92	0.92	0.92	12
accuracy			0.89	18
macro avg	0.88	0.88	0.88	18
weighted avg	0.89	0.89	0.89	18

KNN Metrics Report

	precision	recall	f1-score	support
Did Not Land	0.50	0.50	0.50	6
Landed	0.75	0.75	0.75	12
accuracy			0.67	18
macro avg	0.62	0.62	0.62	18
weighted avg	0.67	0.67	0.67	18

scikit-learn.org

Conclusions

Over time, SpaceX has innovated Falcon 9 rocket and operational features (see figure 1) to achieve mission objectives. To compete, the SpaceY (Tiercel) rocket program will need to meet key features and landing outcomes of the Falcon 9 program.

A supervised classification model using machine learning pipelines, when tuned and refined, we found **Logistic Regression predicted historical landings Falcon 9 boosters with 0.94 accuracy rate** (see details figure 2.).
We use this to be confident of the key features and predict outcomes for SpaceY Tiercel rocket program.
Key rocket features, program resources and target financials necessary for SpaceY's business as an alternate supplier:

- <u>**Payload Capacity**</u> - **engines** capacity of <u>**22,800 kg to LEO and 8,300 kg to GTO orbits**</u>.
- <u>**Rocket Reuse**</u> - **landing legs and grid fins**, thermal protection system, **avionics** systems for reusability, reliability, and maintainability.
- <u>**Operational Facilities**</u> - **lower latitude, oceanside launch sites** with **close landing proximities**. And close **road /rail networks** for effective movement of mission hardware, near to key human resources and suppliers.
- <u>**Pricing**</u> – with rapid re-use model of operations SpaceY Tiercel missions must be a <u>**maximum of $62 million for each flight**</u>.

Appendix – SpaceX Innovation Timeline

The Evolution of Falcon 9 Rocket:

1. **Introduction of the original Falcon 9 v1.0 (2010-2013):**

   First launched in 2010 and retired in 2013, marked the beginning of SpaceX's goal to make space travel cheaper, more frequent and reliable by developing reusable rockets.

   Single Merlin 1C engine powering first stage, max payload capacity: 10,450 kg to LEO and 4,540 kg to GTO.

2. **Falcon 9 v1.1 (2013-2016):**

   Featured a longer first stage, and more powerful **Merlin 1D engines**, which improved its performance

   Maximum payload capacity of **13,150 kg to LEO and 5,500 kg to GTO**

3. **Full Thrust (2015-Present):**

   Upgraded engines, aerodynamic interstage, and stronger stage separation**.**

   Featured several upgrades such as the use of densified propellants, the addition of **landing legs and grid fins**, which allowed the first stage to land back on Earth after launch.

   Maximum payload capacity of 22,800 kg to LEO and 8,300 kg to GTO.

4. **Falcon 9 Block 4 (2017-2018):**

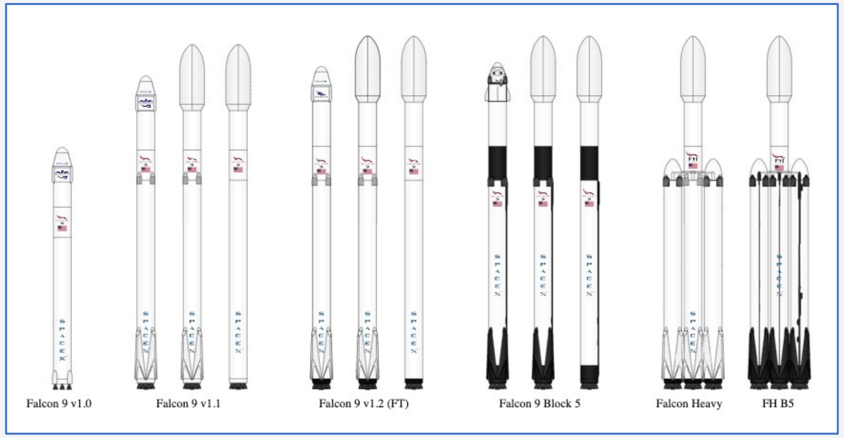
   Improvements to thermal protection system, guidance, navigation, and control systems.

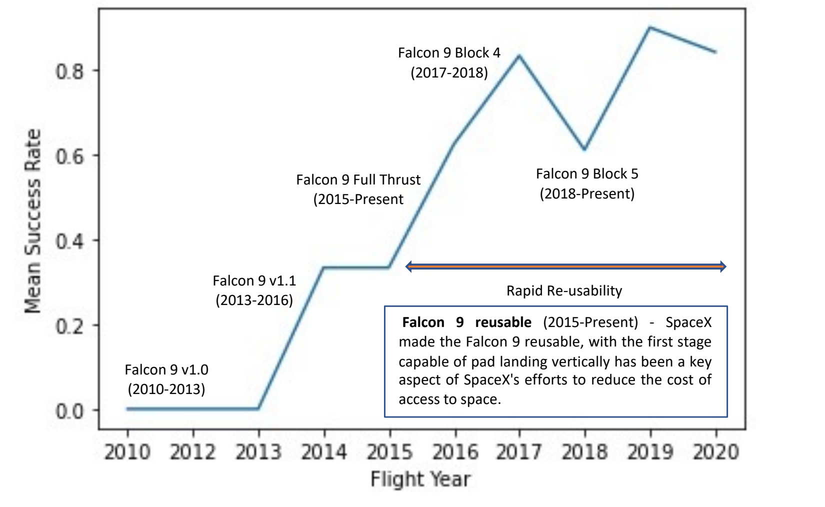
5. **Falcon 9 Block 5 (2018-Present):**

   The last major upgrade to the Falcon 9, featured several improvements such as more powerful engines and more durable structures, which allowed it to perform more launches and landings before needing maintenance.

   Maximum payload capacity of **22,800 kg to LEO and 8,300 kg to GTO**

   Upgraded heat shielding on first stage, improved landing legs, enhanced avionics system, improved reusability, reliability, and maintainability.





Appendix - Business Analysis

Launch Financial Breakdown	Cost
Flight Revenue (New Booster)	$62 million
Flight Revenue (Restored Booster)	$50 million
Total Marginal Cost (per extra flight)	$15.25 million
SpaceX Revenue (for two flights)	$112 million
Net Profit (for two flights)	$47 million (42% profit)
Potential Loss (for one failed mission)	$47 million (76% profit loss)

Appendix

Descriptive Statistics for Landing Success = “1” .

Note rates for orbits with less than three flights are not useable for performance patterns.

Orbit	count	mean	std	min	25%	50%	75%	max
ES-L1	1.0	1.000000	NaN	1.0	1.0	1.0	1.0	1.0
GEO	1.0	1.000000	NaN	1.0	1.0	1.0	1.0	1.0
GTO	27.0	0.518519	0.509175	0.0	0.0	1.0	1.0	1.0
HEO	1.0	1.000000	NaN	1.0	1.0	1.0	1.0	1.0
ISS	21.0	0.619048	0.497613	0.0	0.0	1.0	1.0	1.0
LEO	7.0	0.714286	0.487950	0.0	0.5	1.0	1.0	1.0
MEO	3.0	0.666667	0.577350	0.0	0.5	1.0	1.0	1.0
PO	9.0	0.666667	0.500000	0.0	0.0	1.0	1.0	1.0
SO	1.0	0.000000	NaN	0.0	0.0	0.0	0.0	0.0
SSO	5.0	1.000000	0.000000	1.0	1.0	1.0	1.0	1.0
VLEO	14.0	0.857143	0.363137	0.0	1.0	1.0	1.0	1.0

Appendix

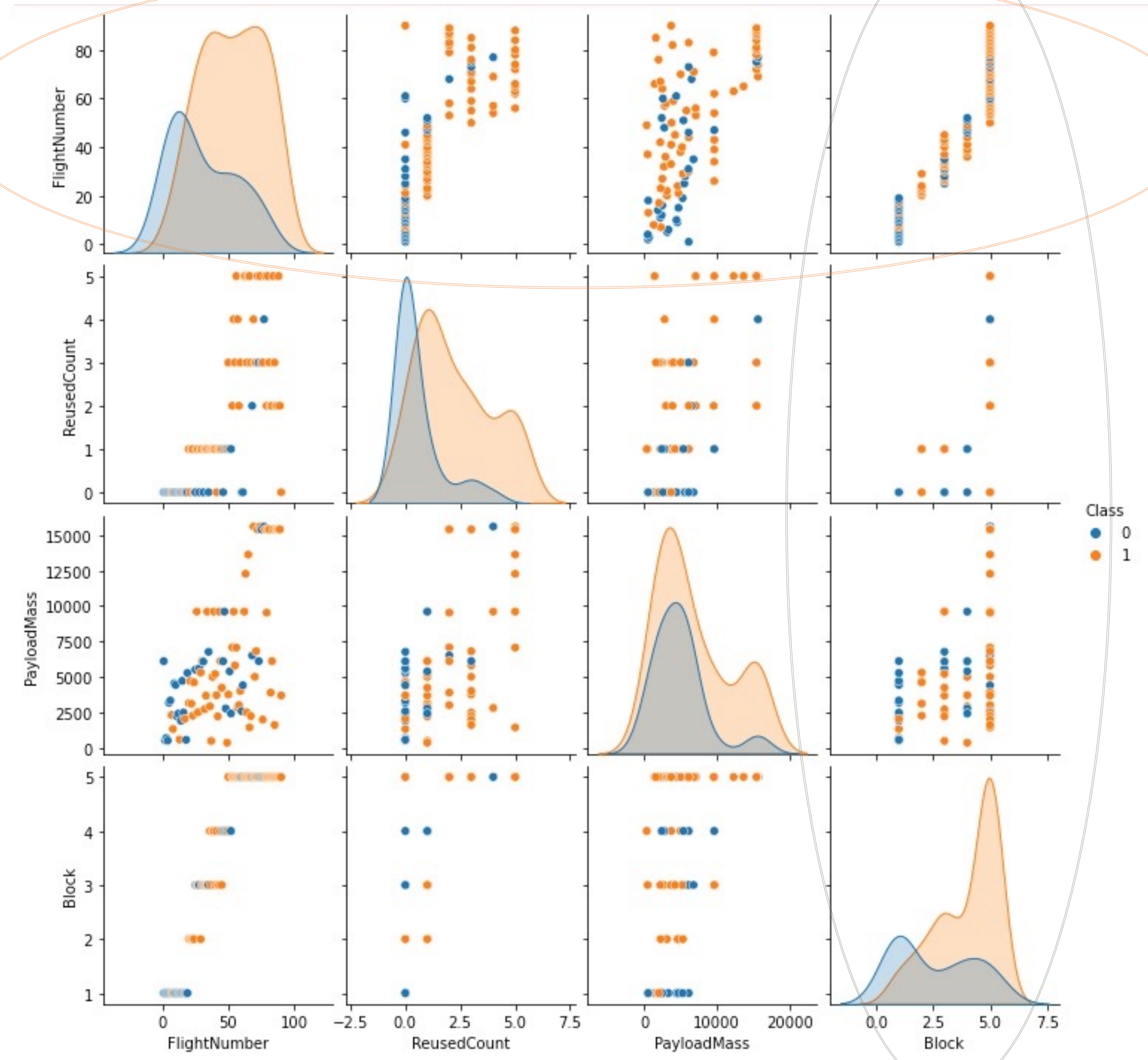
Pair Plots, key numerical features landing class outcomes showing distributions and scatter plots.

Significant improvement in outcomes for all payloads can be seen over time with flight number and block number.

Block innovation has been key for success:

Payload Capacity - engines capacity of 22,800 kg to LEO and 8,300 kg to GTO orbits.

Rocket Reuse - landing legs and grid fins, thermal protection system, avionics systems for reusability, reliability, and maintainability.



Appendix – Feature Analysis

Data sets were refined at the predictive classification modeling to refine key features.

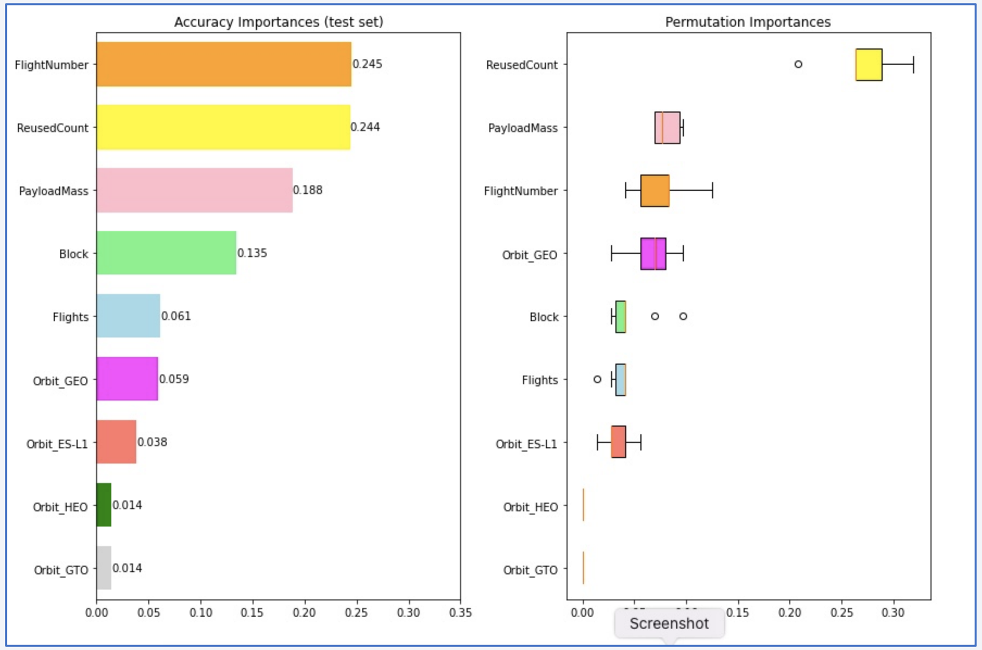
Aditional feature engineering to optimize features and improve accuracy.

```
# "VarianceThreshold is a simple baseline approach to improved feature selection whose variance doesn't meet threshold for significance",
# We found that if we remove features that are either one or zero in more than 70% of our samples, we optimize feature selection (see visu
from sklearn.feature_selection import VarianceThreshold
X_new = VarianceThreshold(threshold=(.7 * (1 - .7))).fit_transform(X)
X = X_new
print(X.shape)
```

```
(90, 9)
```

Graphical analysis of features selected - we see ranked "feature importance scores" and "feature box plots" to support our selection.

```
# Showing the key features importance for accuracy and permutation (see reference sklearn website)
from sklearn.ensemble import RandomForestClassifier
from sklearn.inspection import permutation_importance
```



Thank you!