Space Race with Data Science

Randy Rankin

12/23/2023

EXECUTIVE SUMMARY

- Summary of Methodologies
  - Data Collection: SpaceX-API + Webscraping of SpaceX Wikipedia page
  - Data Wrangling: Missing Values replaced by mean values
  - Exploratory Data Analysis:
    - Analyze outcome by orbit type
    - Analyze outcome by payload mass and booster versions with SQL
    - Visualize Analysis with charts by payload mass, time, orbit type and launch site
    - Visualize Analysis with map by site
  - Interactive Dashboard: Analysis by Site, Payload and booster version
  - Predictive Analysis Using Classification: Logistic Regressions, SVM, Decision Tree, KNN
- Summary of all results
  - Launch success rate increases over time
  - Higher success rate for higher orbits
  - Higher success rate for higher payload mass
  - Low success rate for booster versions v1.0, v1.1, high success rate for FT, B4, B5
  - Higher success rate for Kennedy Space Center and recent starts at Cape Canaveral

INTRODUCTION

- Project Background and Context
  - SpaceX advertises low-cost Falcon 9 Rocket Launches(average of $62m vs. $165m of competitors
  - This success if because of the reusablilty of the first stage
- Problems we need answers for
  - Can we determine if the first stage will land, we can determine the cost of a launch.
  - This information can be used if an alternate company wants to bid against SpaceX for a rocket launch.

Data Collection

- SpaceX Restful API
  - Restful Interface
  - Get Core Data
  - Get Booster Version
  - Get Launch Site Data
  - Get Payload
- SpaceX Wikipedia Page Webscraping
  - HTML Requests (HTTP-Get)
  - Python/BeautifulSoup
  - Extract Column Names from HTML table Header

SpaceX API Data Collection

- Send Get Request to SpaceX API interface website
- Parse data into Pandas Dataframe
- Extract Data with Specific Functions:
  - Core Data
  - Launch Site Data
  - Payload Mass
  - Booster Version
- Filtered Data for Falcon 9 only

Data Collection – Scraping & Data Wrangling

- Scraping
  - Send HTTP Request to SpaceX Wikipedia website
  - Parse Data with BeautifulSoup into Pandas Dataframe
  - Extract Data with find_all method
  - Store data into dataframe for further use
- Wrangling
  - Dealing with missing values: The payload mass column was missing values. Replaced them with mean value.

METHODOLOGY

Executive Summary

- Data Collection Methodology:
  - SpaceX-API
  - Webscraping of SpaceX Wikipedia page
- Perform Data Wrangling
  - Missing values replace by mean values (Payload Mass)
- Perform Exploratory Data Analysis (EDA) using Visualization and SQL
  - Analyse outcome by orbit type
  - Analyze outcome by payload mass and booster versions with SQL
  - Visual Analysis with charts by payload mass, time, orbit type and launch site
- Perform Interactive visual analytics using Folium and Plotly Dash
  - Visual Analysis with map by site
  - Interactive Dashboard, Analysis by Site, Payload and Booster Version.
- Perform Predictive Analysis using Classification Models
  - Logistic Regression, SVM, Decision Tree, KNN
  - Parameter Tuning with Grid Search

EDA with Data Visualization

- Charts:
  - Payload Mass vs Flight Number vs Success Rate: Shows the development of the payload mass and the success rate over time
  - Launch Site vs Flight Number vs Success Rate: Shows us the success rate of each launch site over time
  - Launch Site vs Payload Mass vs Success Rate: Shows which payload is best to have success at a specific launch site
  - Orbit Type vs Success Rate: Can give an idea on which orbit types have the highest success rates.
  - Orbit Type vs Flight Number vs Success Rate: Shows the development of orbit types over time
  - Orbit Type vs Payload Mass vs Success Rate: Show the success rate for specific orbit types and payload mass clusters
  - Success Rate vs Year: Shows the success development over time

EDA with SQL

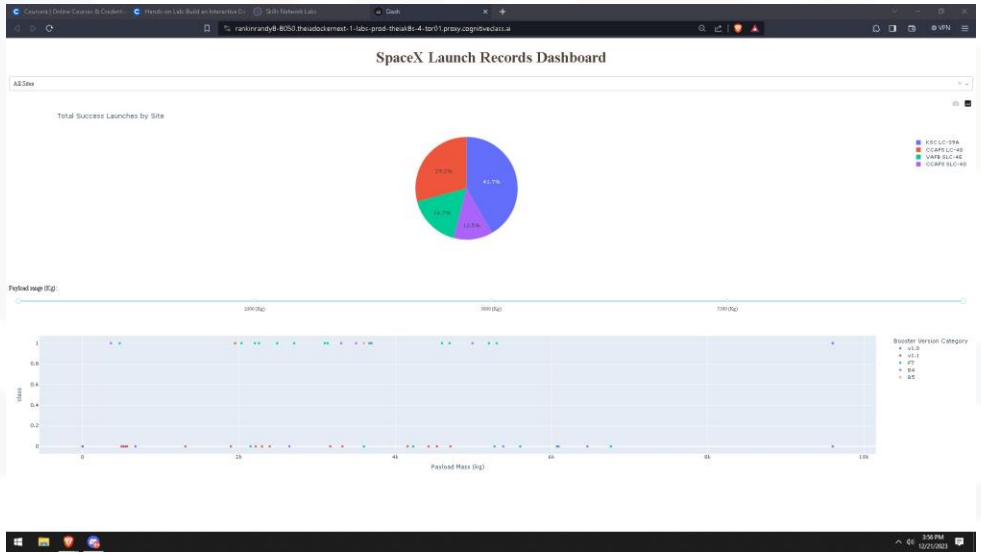
- SQL Queries
  - Extract a list of launch sites
  - Display the total payload mass carried by boosters launched by NASA (CRS)
  - Display average payload mass carried by booster version F9 v1.1
  - List date when the first successful landing outcome in ground pad was accomplished
  - List names of boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
  - List the number of successful and failed mission outcomes
  - List the names of the booster versions which carried the maximum payload mass
  - List the failed landing outcomes in drone ship, their booster versions, and launch site names for the year 2015
  - Rank the count of landing outcomes in descending order

Interactive Map with Folium

- Map Objects
  - Edged Circles (radius 1000m): Space launch sites
  - Markers: for labeling all objects
  - MarkerCluster: for creating a bunch of markers around space launch sites to indicate success or failure of the landing of the rocket's first stage
  - Lines: Measure the distance between the launch site and the next coast or city

DashBoard with Plotly Dash

- Input Elements:
  - Dropdown list for the launch site
  - RangeSlider for selection the payload mass
- Output Elements:
  - PieChart: for showing the success rate of each launch site, or number of successful landing outcomes
  - Scatterplot: Show success or failure by payload and booster version



Predictive Analysis (Classification)

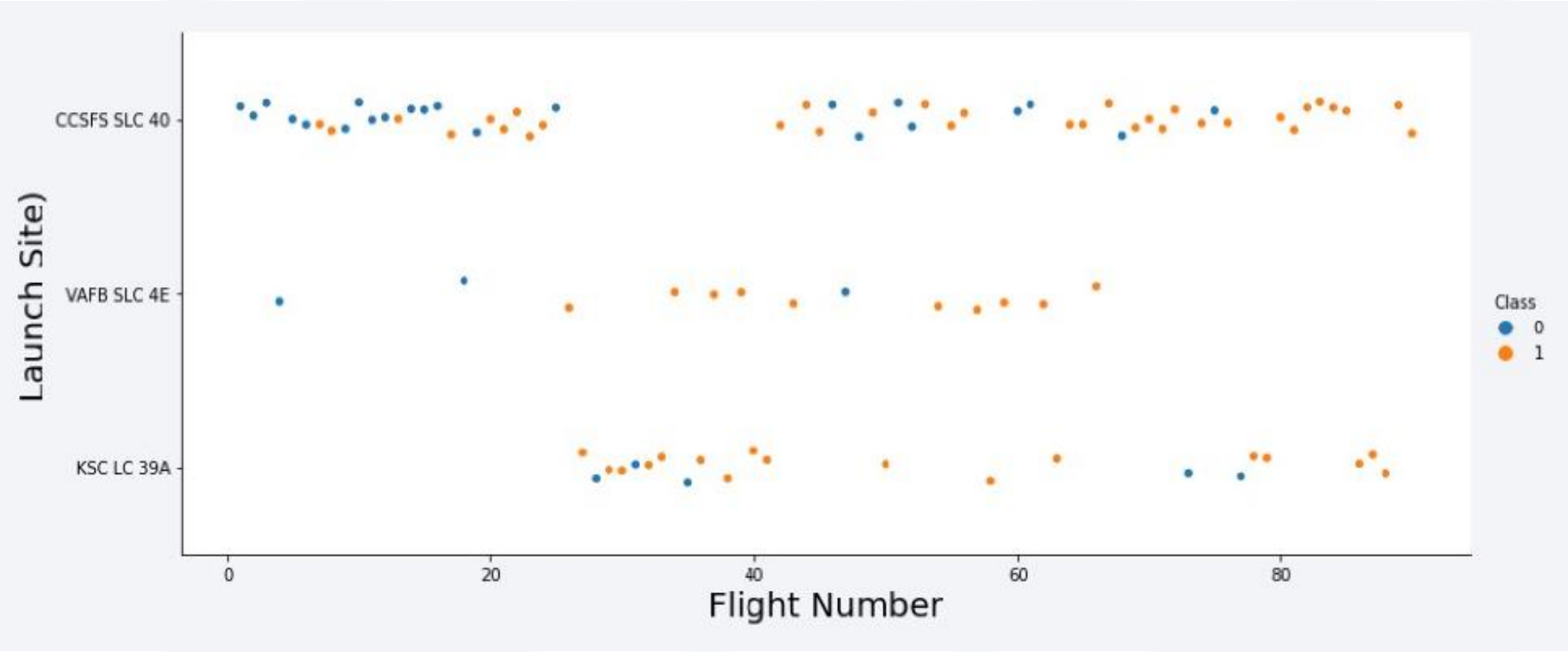
- Preprocessing
  - One-Hot-Encoding for Categorial Features
  - Split data into Dependent/Independent variables and train/test data
  - Scale Data with StandardScaler
- Model Building for each Method
  - Logistic Regression
  - Support Vector Machine
  - Decision Tree
  - K-Nearest Neighbor
- Optimization
  - Use Gridsearch for optimizing the models based on their hyperparameters
- Evaluation
  - Use Accuracy of Gridsearch for selecting the best parameter
  - Use Score to compare each classification Method

Results

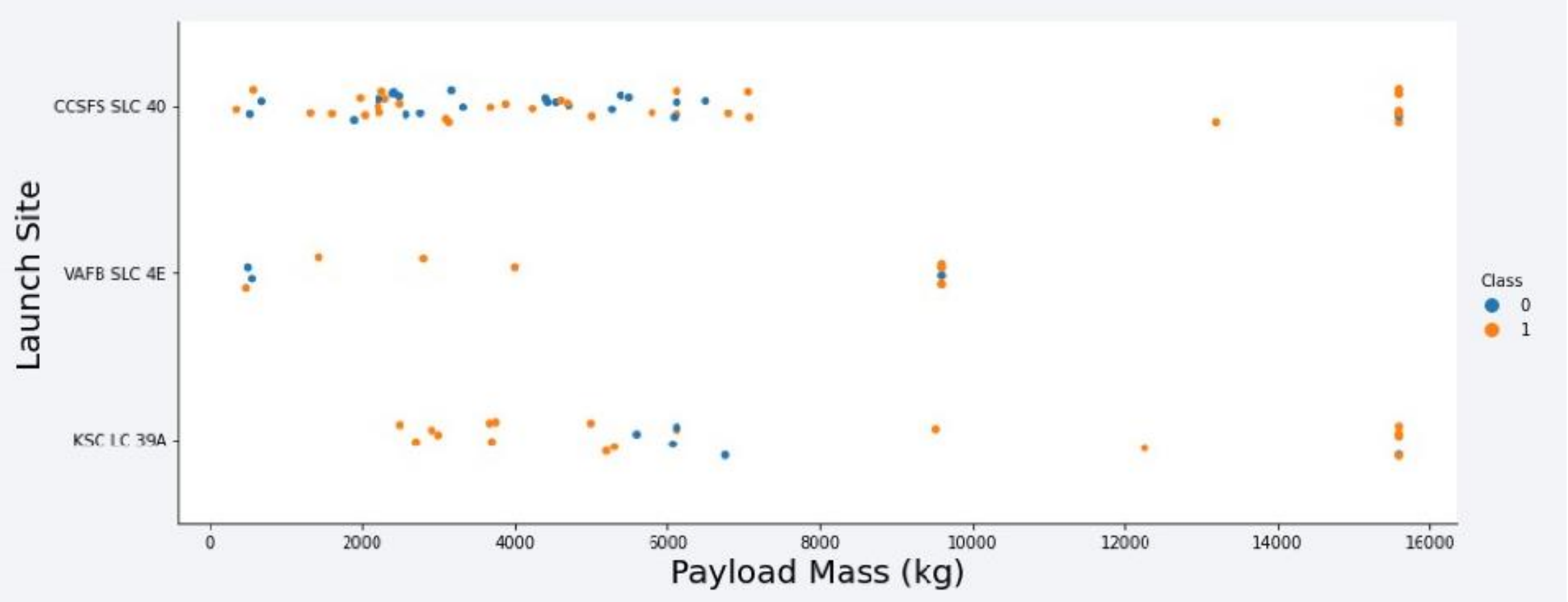
- Exploratory Data Analysis Results
  - Launch success rate increases over time
  - Higher success rate for higher orbits
- Interactive analytics demo in screenshots
  - Higher success rate for higher payload mass
  - Low success rate for booster versions v1.0, v1.1, high success rate for FT, B4, B5
  - Higher success rate for Kennedy Space Center and recent starts at Cape Caneveral
- Predictive analysis results
  - Best prediction results with Logistic Regression and Support Vector Machine

INSIGHTS GAINED FROM EDA

Flight Number vs Launch Site

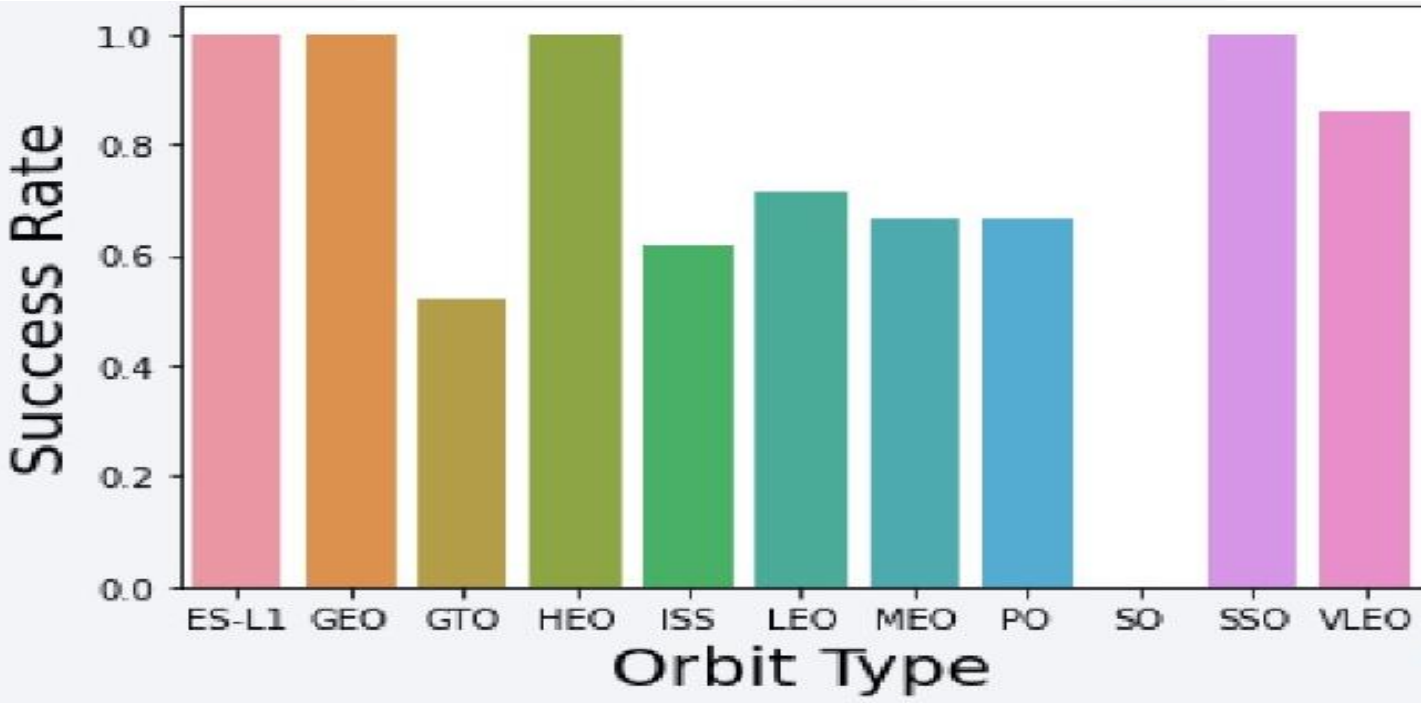


Payload vs Launch Site



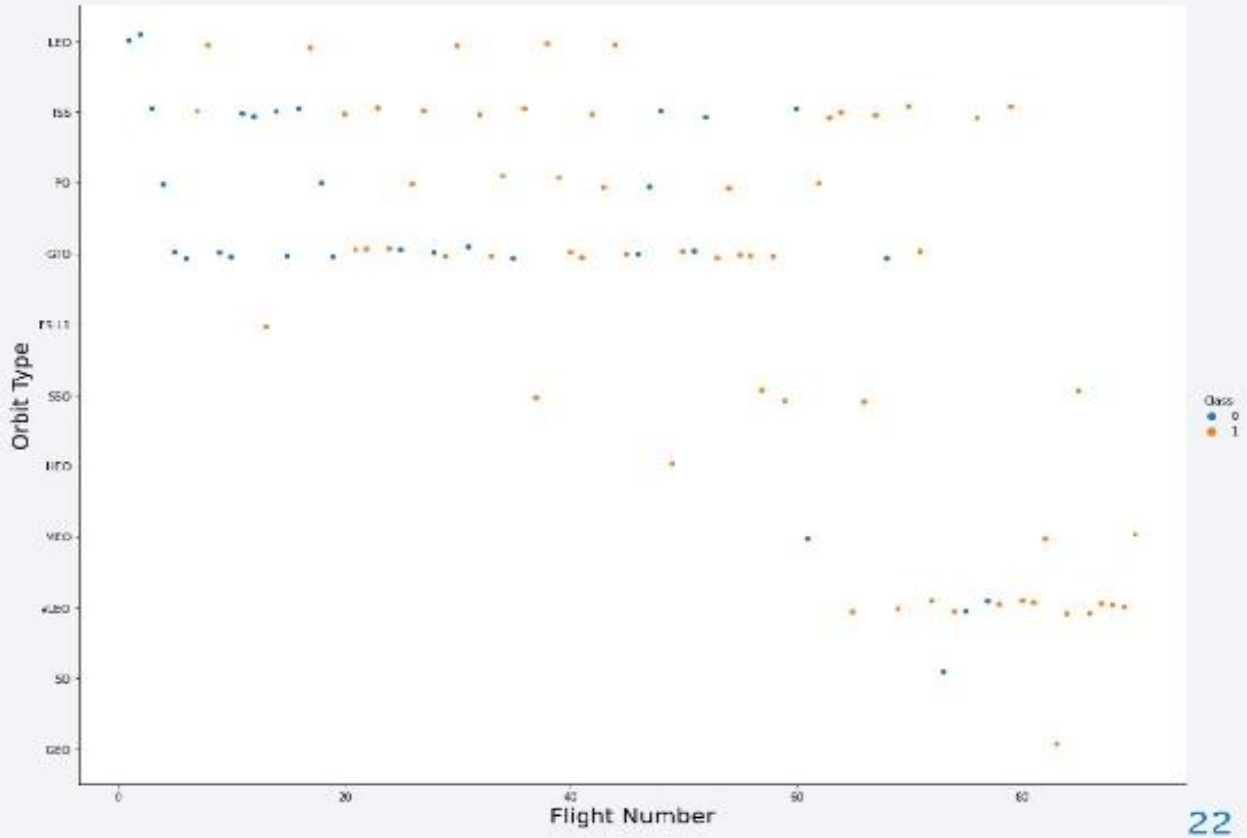
Success Rate vs Orbit Type

- Low Earth Orbits
  - GTO
  - ISS
  - LEO
  - MEO
  - PO
  - VLEO
- High Earth Orbits
  - ES-L1
  - GEO
  - HEO
  - SSO

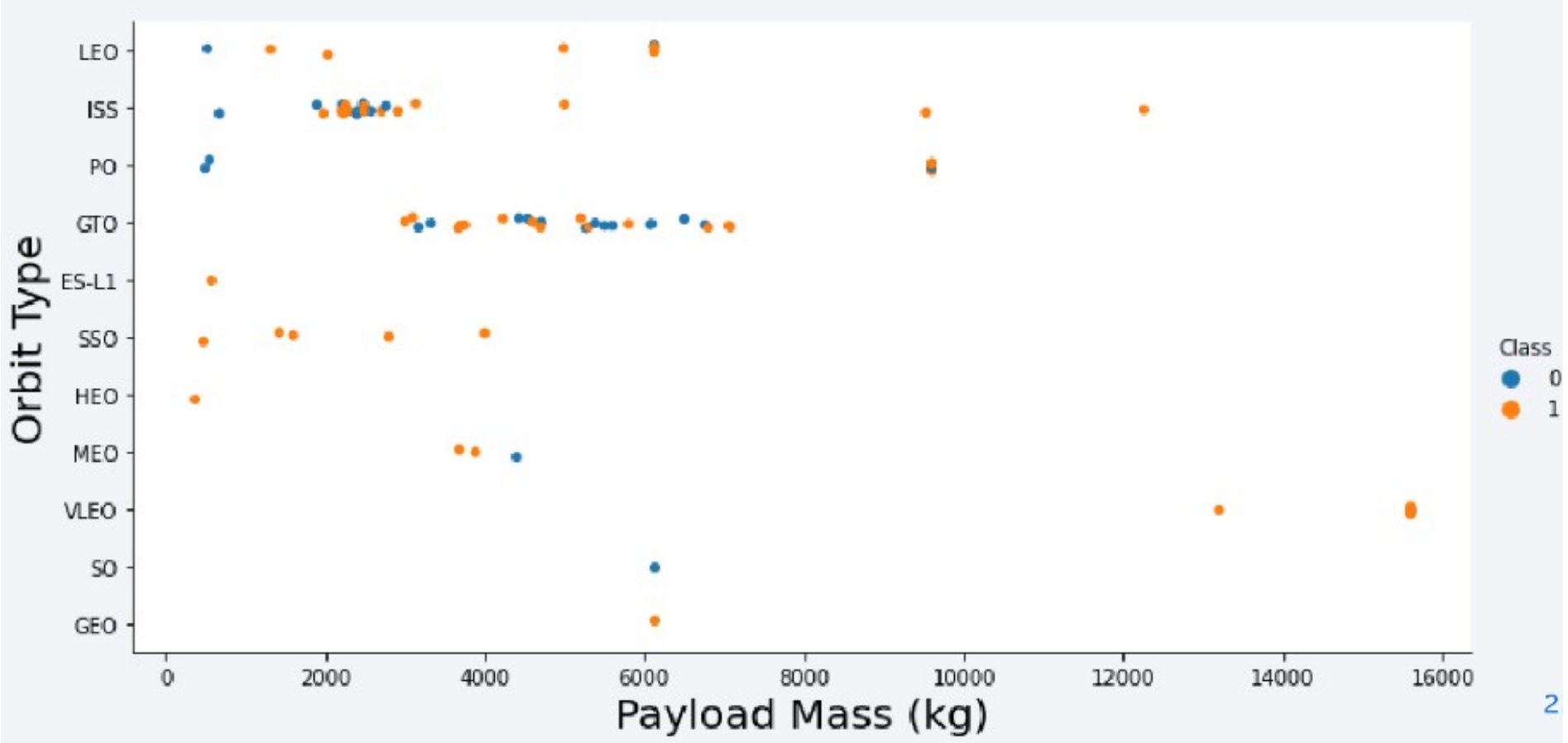


Flight Number vs Orbit Type

- The predominant orbit types have changed over time. Success rate has increased over time for all orbit types.

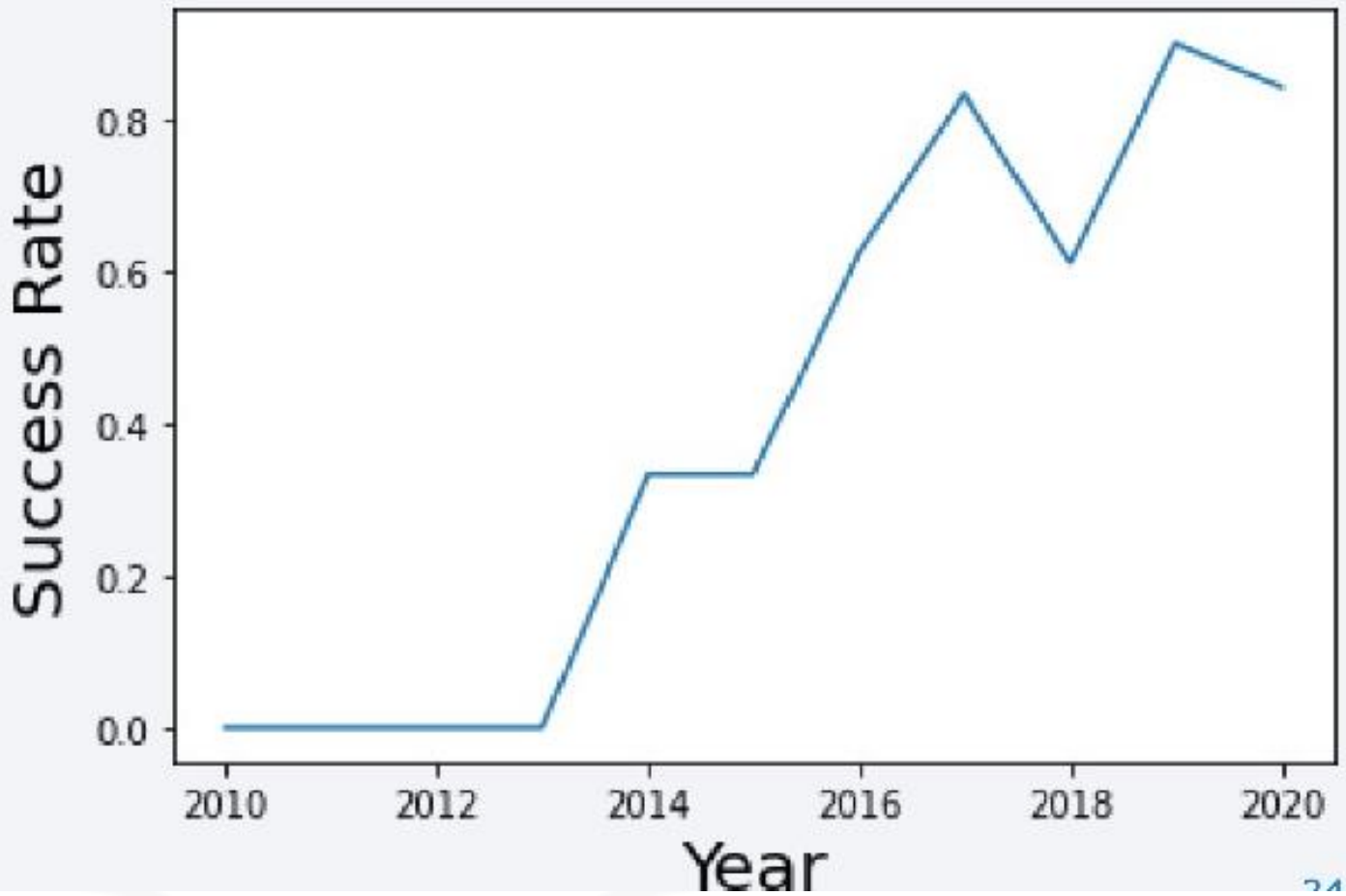


Payload vs Orbit Type



Launch Success Yearly Trend

- Launch success has increased over the years



All Launch Site Names

- KSC: Kennedy Space Center
- CCA...: Cape Canaveral Launch Center
- VAFB: Vandenburg Air Force Base

	Launch_Site
0	CCAFS LC-40
1	VAFB SLC-4E
2	KSC LC-39A
3	CCAFS SLC-40

Launch Sites Names Begin with 'CAA'

- Some sample records for starts at Cape Canaveral Space Center

Date	Time_(UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_C
2010-06-04 00:00:00	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success
2010-12-08 00:00:00	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0	LEO (ISS)	NASA (COTS) NRO	Success
2012-05-22 00:00:00	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success
2012-10-08 00:00:00	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success
2013-03-01 00:00:00	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success

Total Payload Mass

- Total payload carried by boosters from NASA

sum(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

- Average payload mass carried by booster version F9 v1.1

	avg(PAYLOAD_MASS__KG_)
0	2928.4

First Successful Ground Landing

- Date of the first successful landing outcome on ground pad

min(Date)
2015-12-22

Successful Drone Ship Landing Payload between 4000 and 6000

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

	Booster_Version
0	F9 FT B1022
1	F9 FT B1026
2	F9 FT B1021.2
3	F9 FT B1031.2

Outcome of Missions

- Total number of successful and failure mission outcomes (not to be confused with the outcome of the stage-1 landing)

Mission_Outcome	count(*)
Failure	1
Success	100

Boosters Carried Maximum Payload

- Names of the booster which have carried the maximum payload mass

	Booster_Version
0	F9 B5 B1048.4
1	F9 B5 B1049.4
2	F9 B5 B1051.3
3	F9 B5 B1056.4
4	F9 B5 B1048.5
5	F9 B5 B1051.4
6	F9 B5 B1049.5
7	F9 B5 B1060.2
8	F9 B5 B1058.3
9	F9 B5 B1051.6
10	F9 B5 B1060.3
11	F9 B5 B1049.7

2015 Launch Records

- List of failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015

Landing__Outcome	Booster_Version	Launch_Site
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40
Failure (drone ship)	F9 v1.1 B1017	VAFB SLC-4E
Failure (drone ship)	F9 FT B1020	CCAFS LC-40
Failure (drone ship)	F9 FT B1024	CCAFS LC-40

Rank Landing Outcomes Between 6/4/2010 & 3/20/2017

- Rank of 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

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

Launch Sites Proximities Analysis

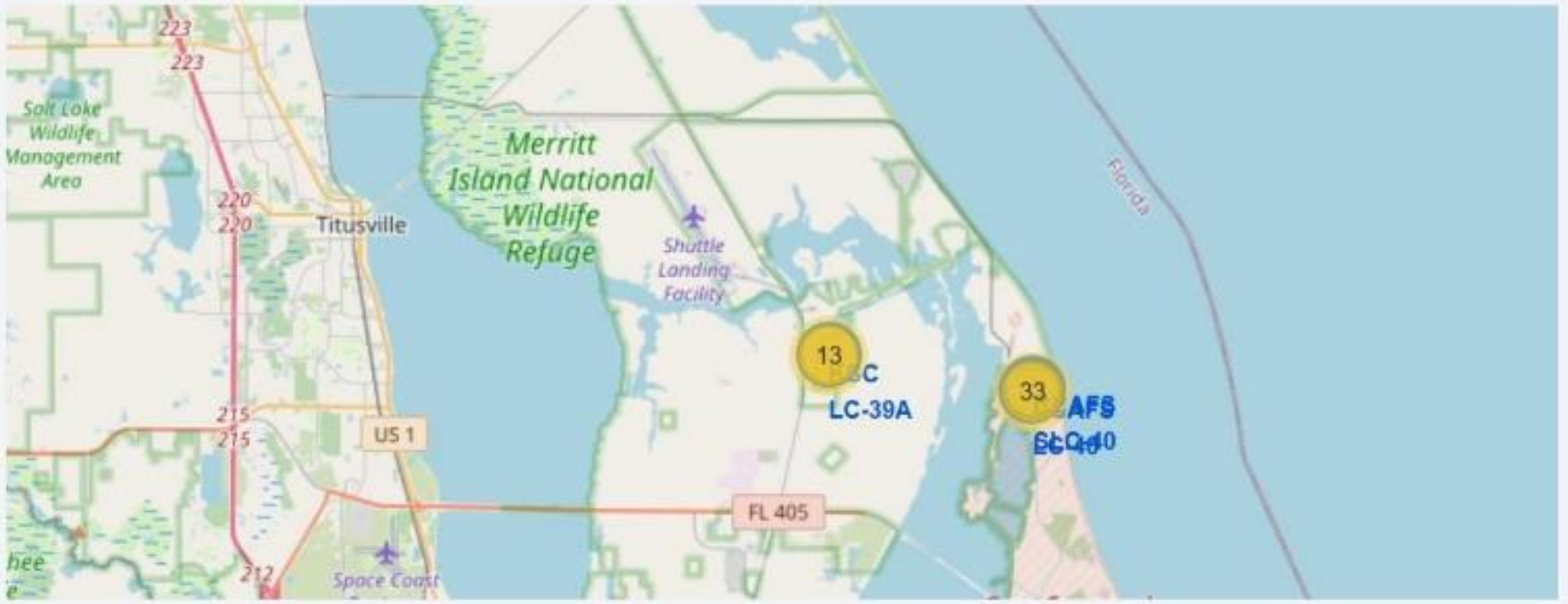
Folium Map: Launch Sites

- Launch sites are at the East and West coast, near the southernmost U.S. mainland area, which is Florida and California

Folium Map: Proximity Vandenburg AFB

- Close to the Vandenburg AFB is the town of Lompoc. This might be an issue, if the stage-1 landing cannot be controlled, since rockets would usually start towards Eastern direction

Folium Map: Proximity Kennedy SC/Cape Canaveral



- No city towards the Eastern Direction, ideal place for testing rocket launches

Folium Map: Stage-1 Landing Success by Launch Site

- Vandenburg AFB

- Kennedy Space Center

- Cape Canaveral

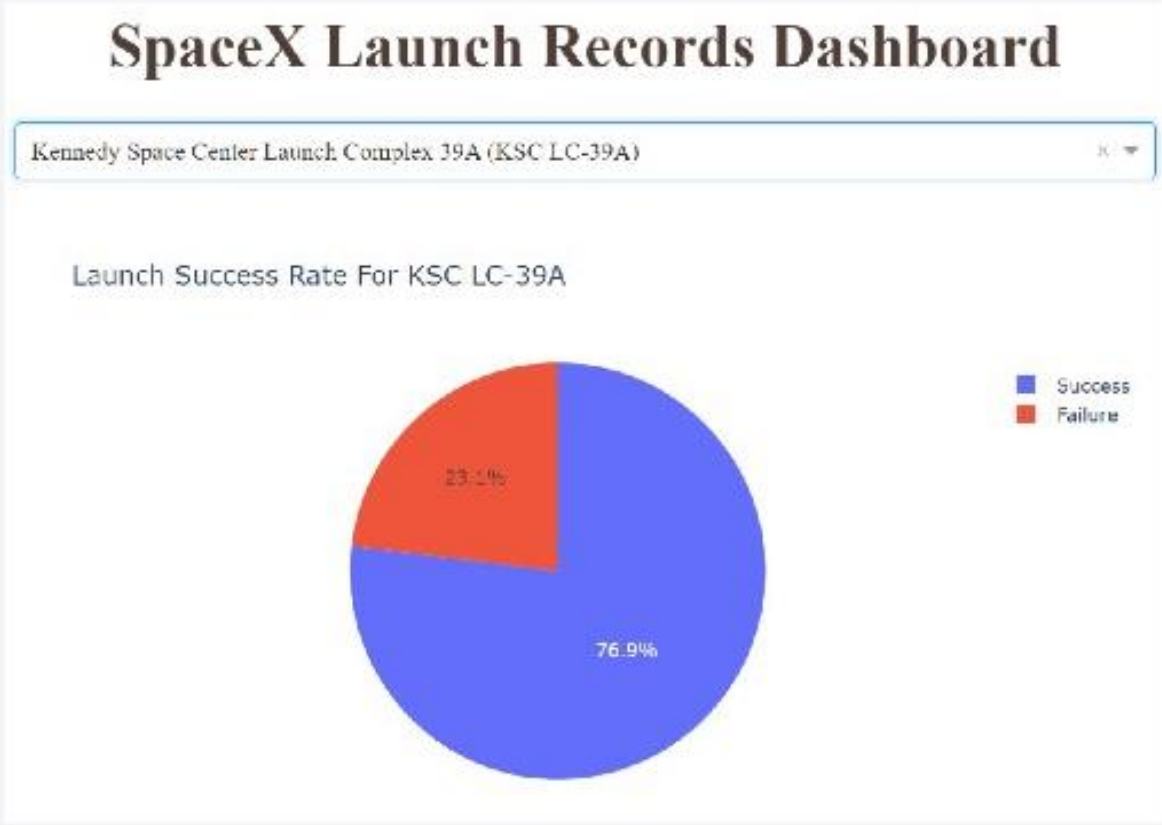
Dashboard with Plotly Dash

Dashboard: Launch Success Count (All Sites)

- Kennedy Space Center (KSC LC-39A) hast the most successful stage-1 landings
- Vandenberg Air Force Base (VAFB SLC-4E) has the least number of successful stage-1 landings

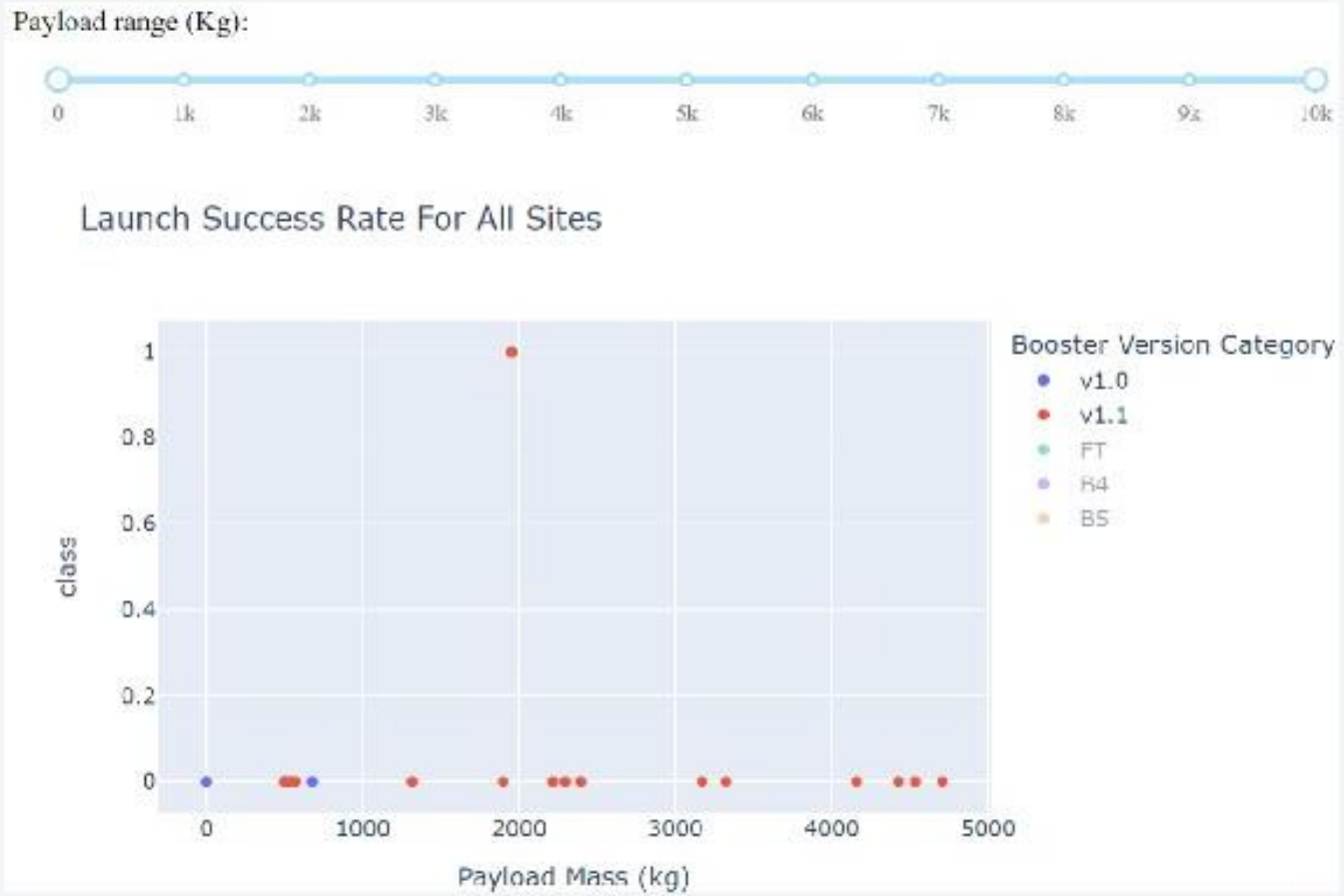
Dashboard: Success Rate Kennedy Space Center

- More than 3 of 4 landings have been successful at Kennedy Space Center



Dashboard: Booster Versions V1.0, V1.1

- Success rate forBooster versions v1.0 and v1.1 is quite small in the payload range to 10000kg

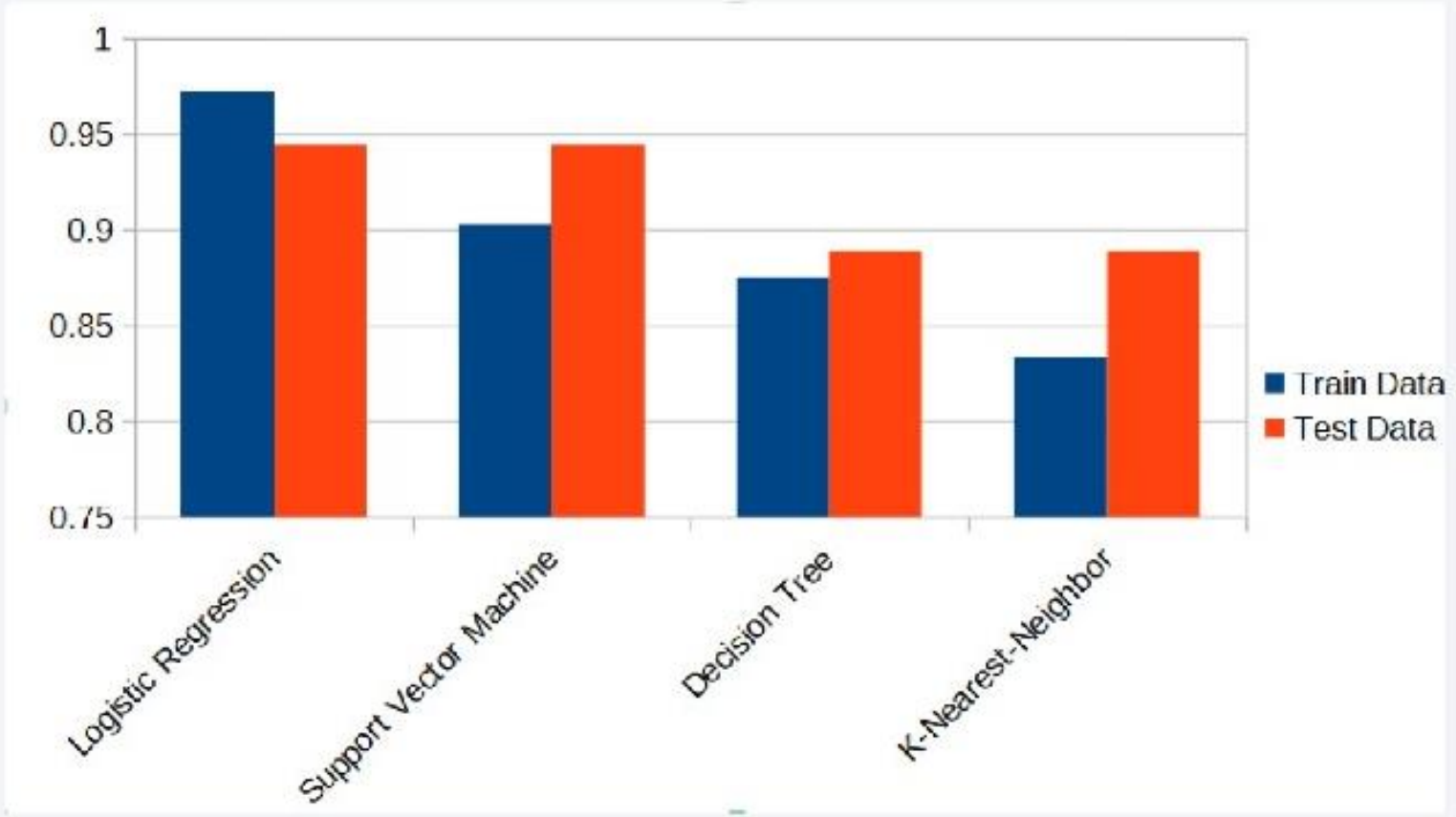


Dashboard: Booster Versions V1.0, V1.1

- Success rate forBooster versions FT, B4 and B5 is better in the payload range to 10000kg

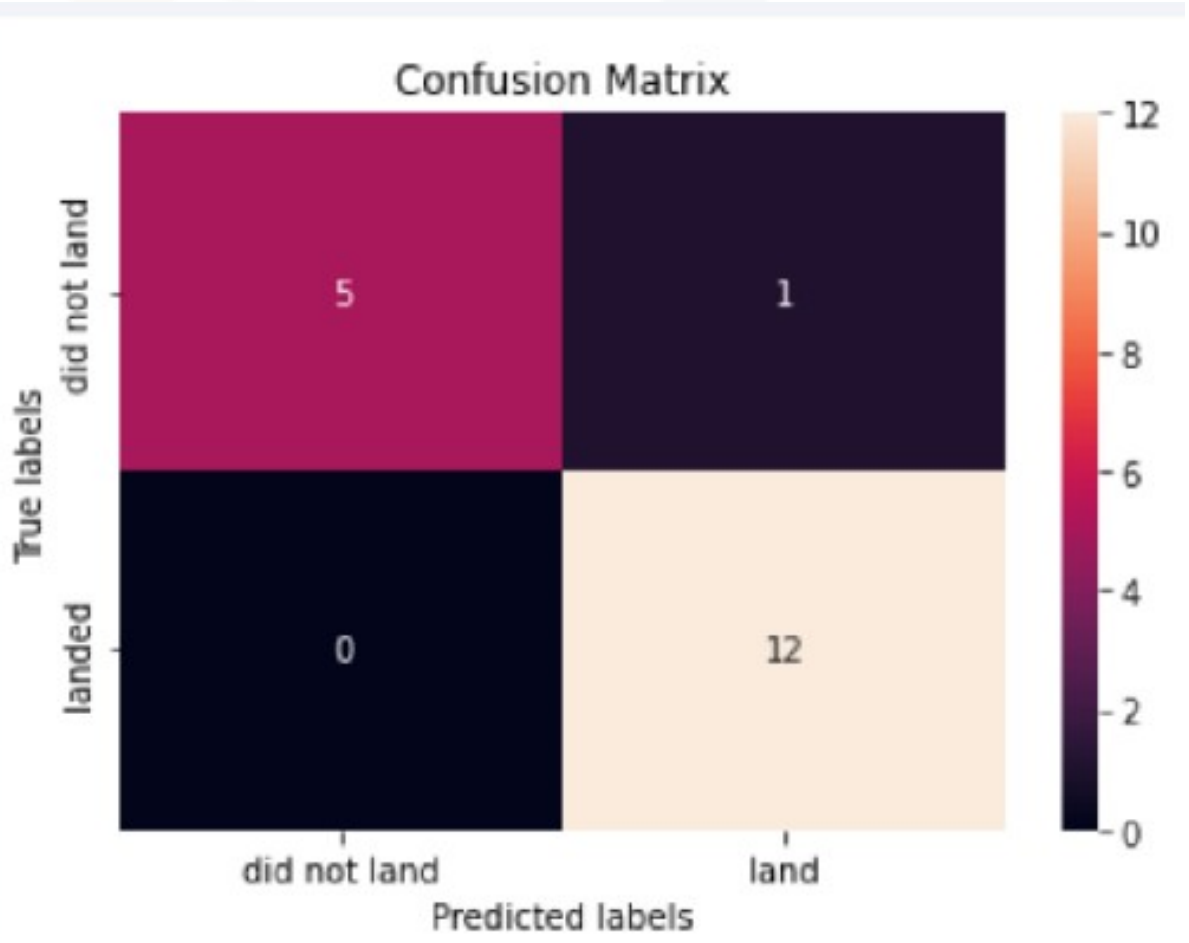
Predictive Analysis (Classification)

- Logistic Regression has the best result for train data
- Logistic Regression and Support Vector Machines have the best results on test data



Confusion Matrix

- True Positives: 12
- True Negatives: 5
- False Positives: 1
- False Negatives: 0



Conclusions

- Prediction with Logistic Regression is quite accurate
- Support Vector Machine also provides a good result for predicting the landing outcome
- None of the Models had false negative
- All the Models had as least one false positive