



IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Shriram
Govindarajan
03/05/2024



Outline

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

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
 - Visual Analysis with charts by payload mass, time, orbit type and launch site
 - Visual Analysis with map by site
 - Interactive Dashboard: Analysis by Site, Payload and booster version
 - Predictive Analysis Using Classification: Logistic Regression, 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 is because of the reusability of the first stage
- Problems we want to find answers
 - If we can 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.

Section 1

Methodology

Methodology

Executive Summary

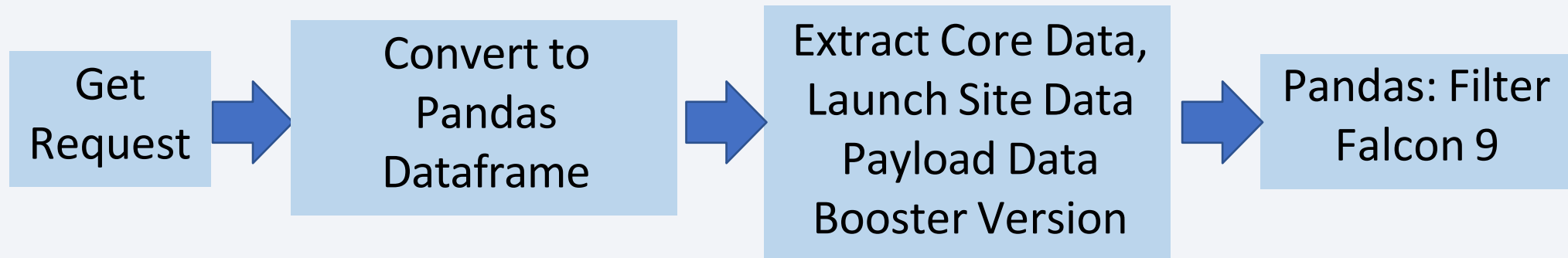
- Data collection methodology:
 - [SpaceX-API](#)
 - [Web scraping of SpaceX Wikipedia page](#)
- Perform data wrangling
 - [Missing Values replaced by mean values \(Payload Mass\)](#)
- Perform exploratory data analysis (EDA) using visualization and SQL
 - [Analyze outcome by orbit type](#)
 - [Analyze 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](#)

Data Collection

- SpaceX REST API
 - RESTful Interface
 - Get Core Data
 - Get Booster Version
 - Get Launch Site Data
 - Get Payload Data
- Webscraping of SpaceX Wikipedia Page
 - HTML Requests (HTTP-Get)
 - Python / BeautifulSoup (Package for Webscraping)
 - Extract Column Names from HTML table header
- [Data Collection Jupyter Notebook](#)

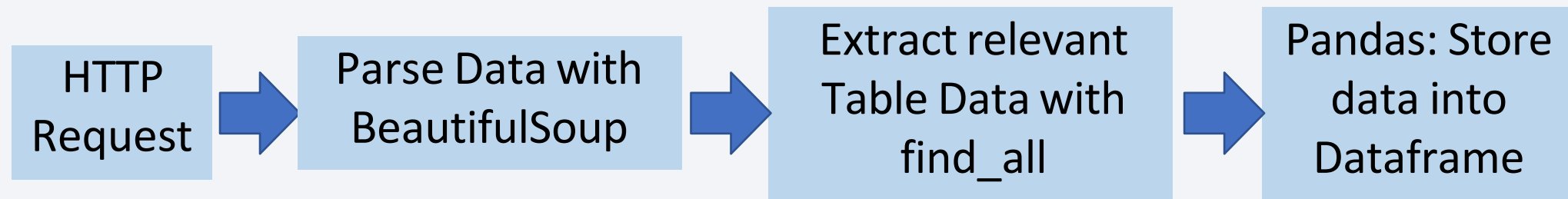
Data Collection – SpaceX API

- Send Get Request to SpaceX API interface website
- Parse data into Pandas dataframe
- Extract data with specific functions for:
 - Core data
 - Launch Site Data
 - Payload Mass
 - Booster Version
- Since Data contains other than Falcon 9 data, we filter for Falcon 9 data only
- [Data Collection \(RESTful API\)](#)



Data Collection - Scraping

- Send HTTP Request to SpaceX Wikipedia website
- Parse data into Pandas dataframe with BeautifulSoup Webscraper
- Extract data with find_all method
- Store data into Pandas dataframe for further use
- [Data Collection \(Web scraping\)](#)



Data Wrangling

- Dealing with missing values: The column for payload mass had a handful of missing values. We replaced them with the mean value (standard in Analysis / ML)
- Data Wrangling

EDA with SQL

- SQL queries

- Extract a list of all launch sites
- Display 5 records where launch sites begin with the string 'CCA'
- Display the total payload mass carried by boosters launched by NASA (CRS)
- Display average payload mass carried by booster version F9 v1.1
- List the date when the first successful landing outcome in ground pad was achieved
- List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- List the total number of successful and failure 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 in year 2015
- 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

- Data Analysis with SQL

EDA with Data Visualization

- Charts:
 - Payload mass vs. Flight number vs. Success rate: This shows us the development of the payload mass and the success rate over time
 - Launch site vs. Flight number vs. Success rate: This shows us the success rate of each launch site over time
 - Launch site vs. Payload mass vs. Success rate: This shows us which payload is best to have success at a specific launch site
 - Orbit type vs. Success rate: This can give us a hint which orbit types have the highest success rates
 - Orbit type vs. Flight number vs. Success rate: This shows us the development of orbit types over time
 - Orbit type vs. Payload mass vs. Success rate: Shows us the success rate for specific orbit type / payload mass clusters
 - Success rate vs. Year: Shows the success development over time
- Data Exploration

Build an 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 (green) or failure (red) of the landing of the rocket's first stage
 - Lines: Measure the distance between the launch site and the next coast or next city
- [Interactive Map with Folium](#)

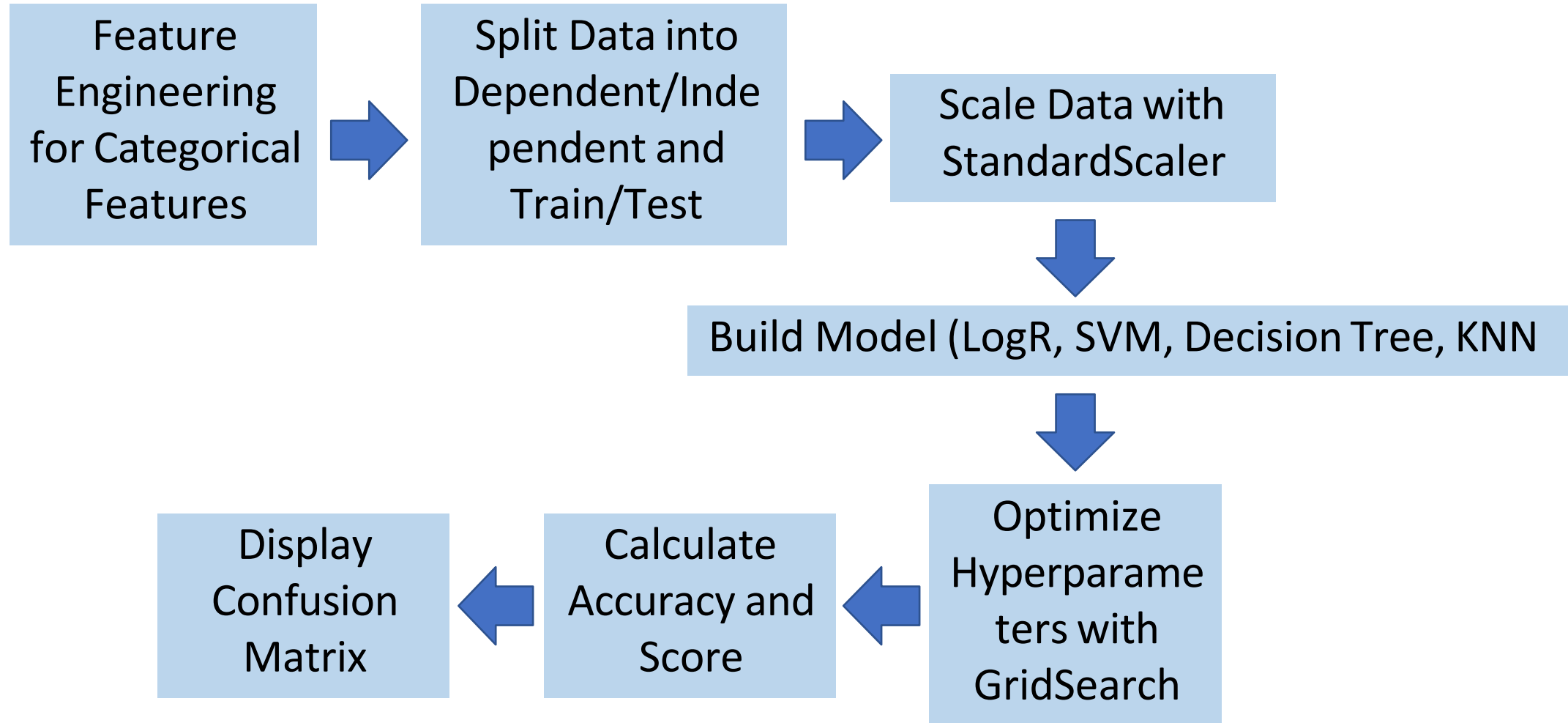
Build a Dashboard with Plotly Dash

- Input Elements:
 - Dropdown list for the launch site (with option to select all)
 - RangeSlider for selecting the payload mass
- Output Elements:
 - PieChart: for showing the success rate of each launch site, or (if all sites are selected) showing the number of successful landing outcomes
 - Scatterplot: Show success/failure by payload and booster version
- [Interactive Dashboard with Plotly](#)

Predictive Analysis (Classification)

- Preprocessing
 - One-Hot-Encoding for Categorical 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
- [Machine Learning Prediction](#)

Predictive Analysis (Classification)



Results

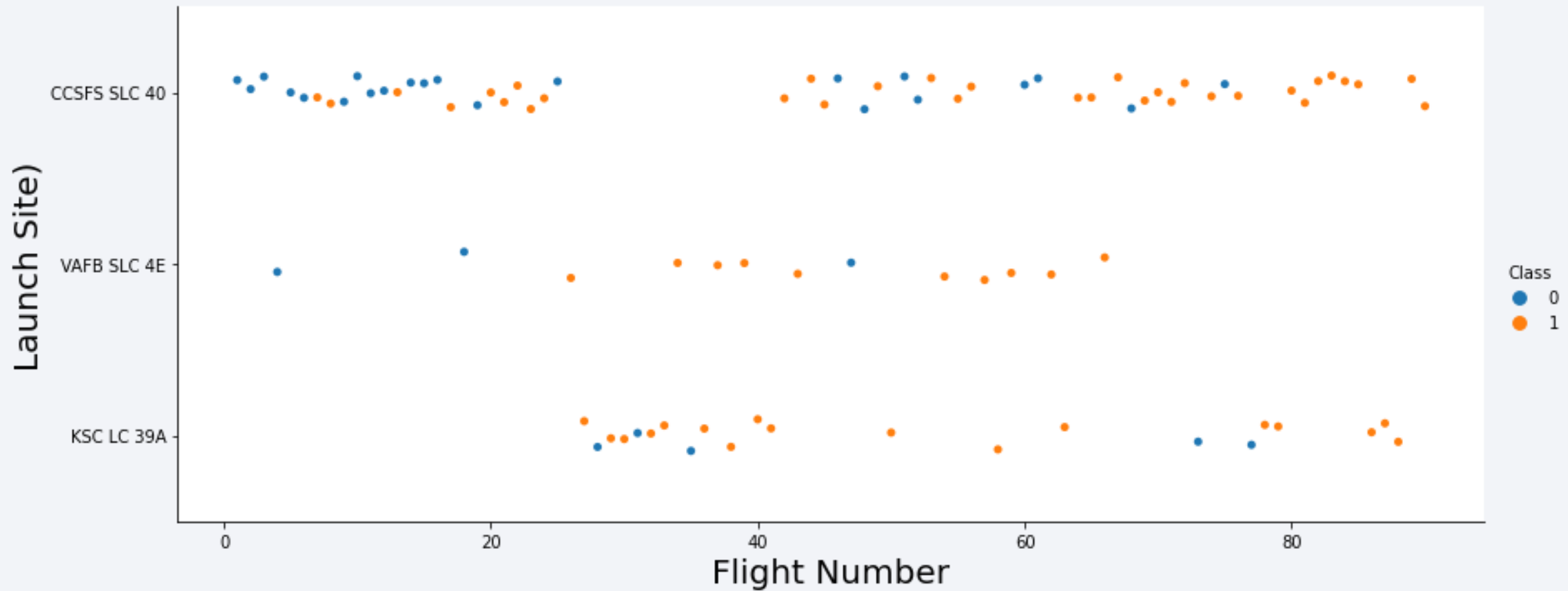
- Exploratory data analysis results
 - Launch success rate increases over time
 - Higher success rate for higher orbits
- Interactive analytics demo in screenshots
 - Higher success rate for higher payload mass
 - Low success rate for booster versions v1.0, v1.1, high success rate for FT, B4, B5
 - Higher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

The background of the slide is an abstract composition. It features a dark blue base color. Overlaid on this are numerous diagonal streaks in shades of blue and red, creating a sense of motion or data flow. A faint, light blue grid pattern is also visible, particularly in the lower-left quadrant. The overall effect is high-tech and modern.

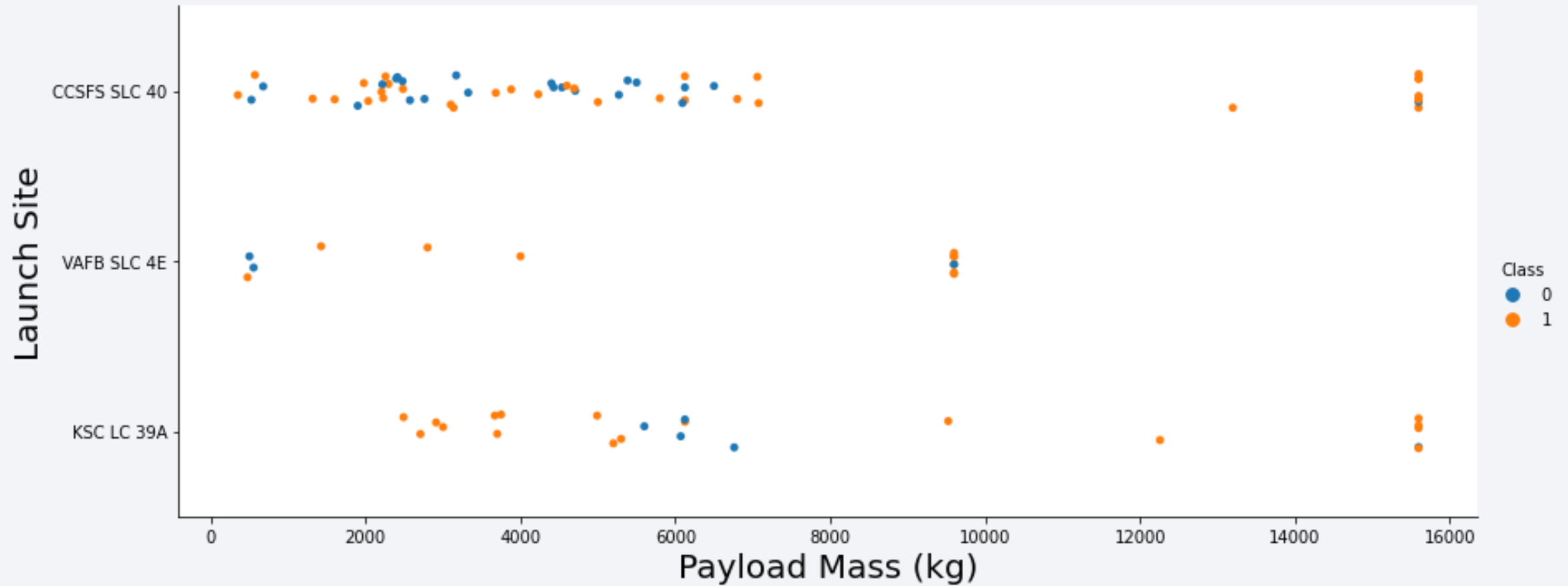
Section 2

Insights drawn from EDA

Flight Number vs. Launch Site



Payload vs. Launch Site



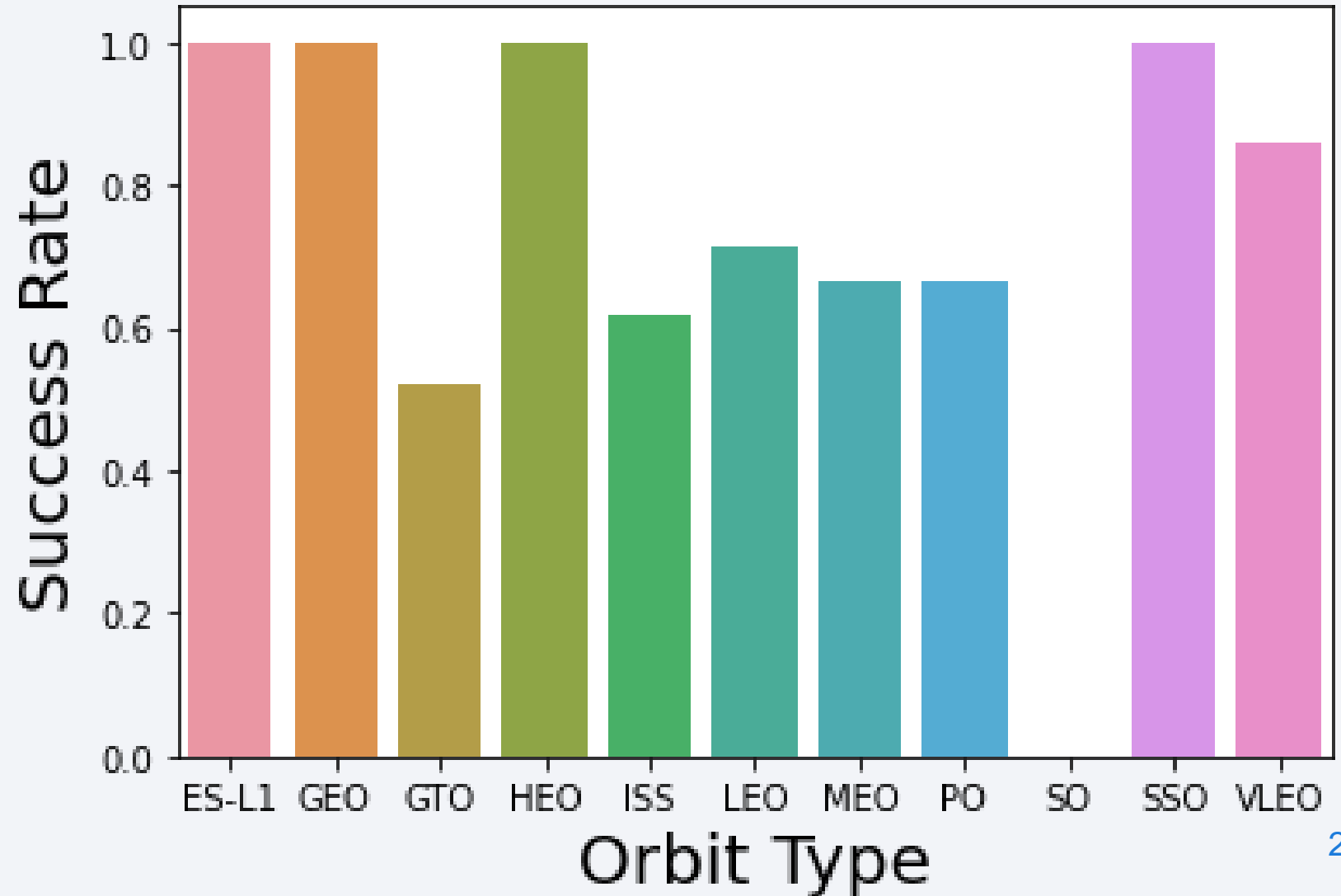
Success Rate vs. Orbit Type

- Low Earth Orbits

- GTO
- ISS
- LEO
- MEO
- PO
- VLEO

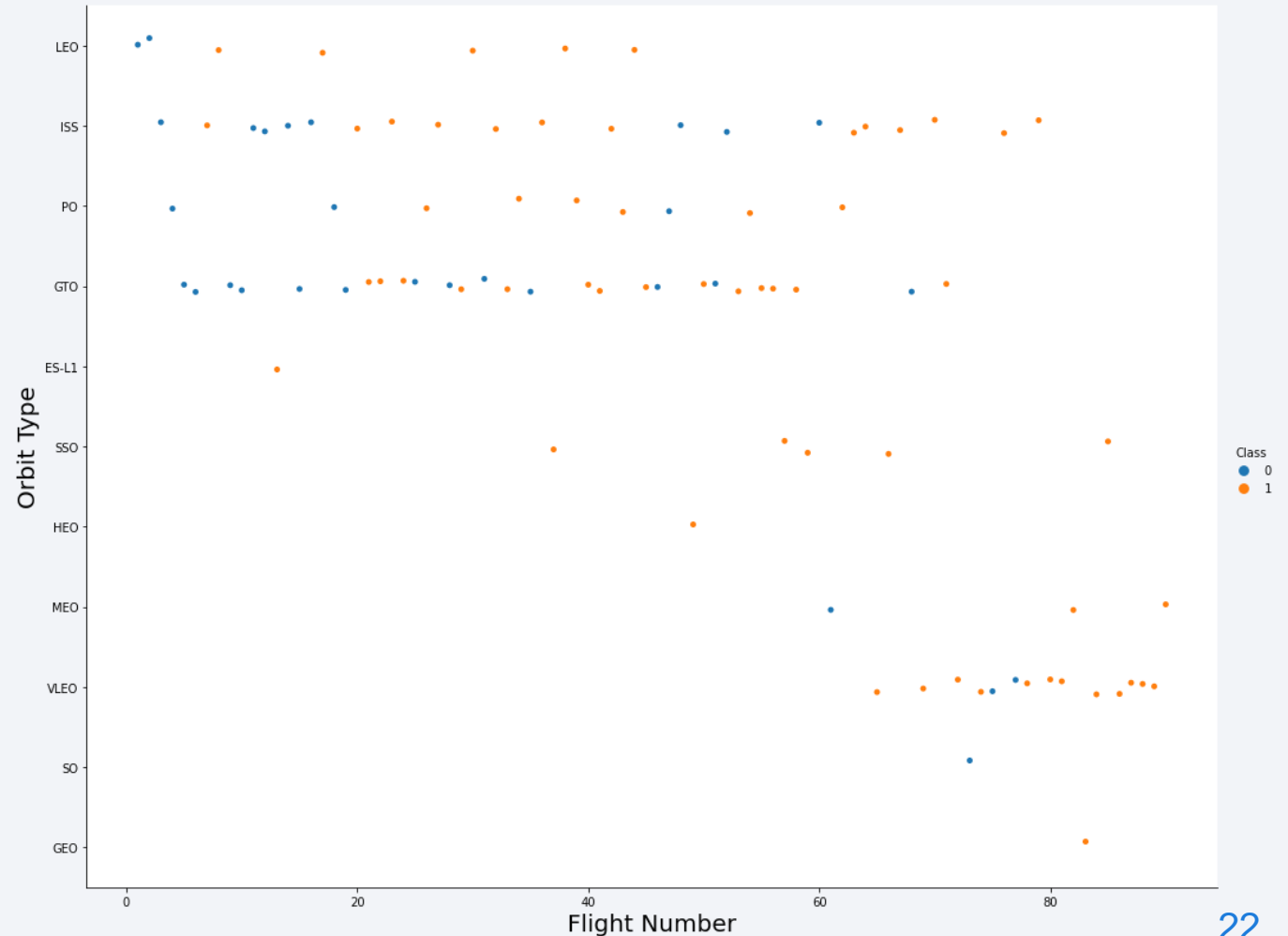
- High Earth Orbits

- ES-L1
- GEO
- HEO
- SSO

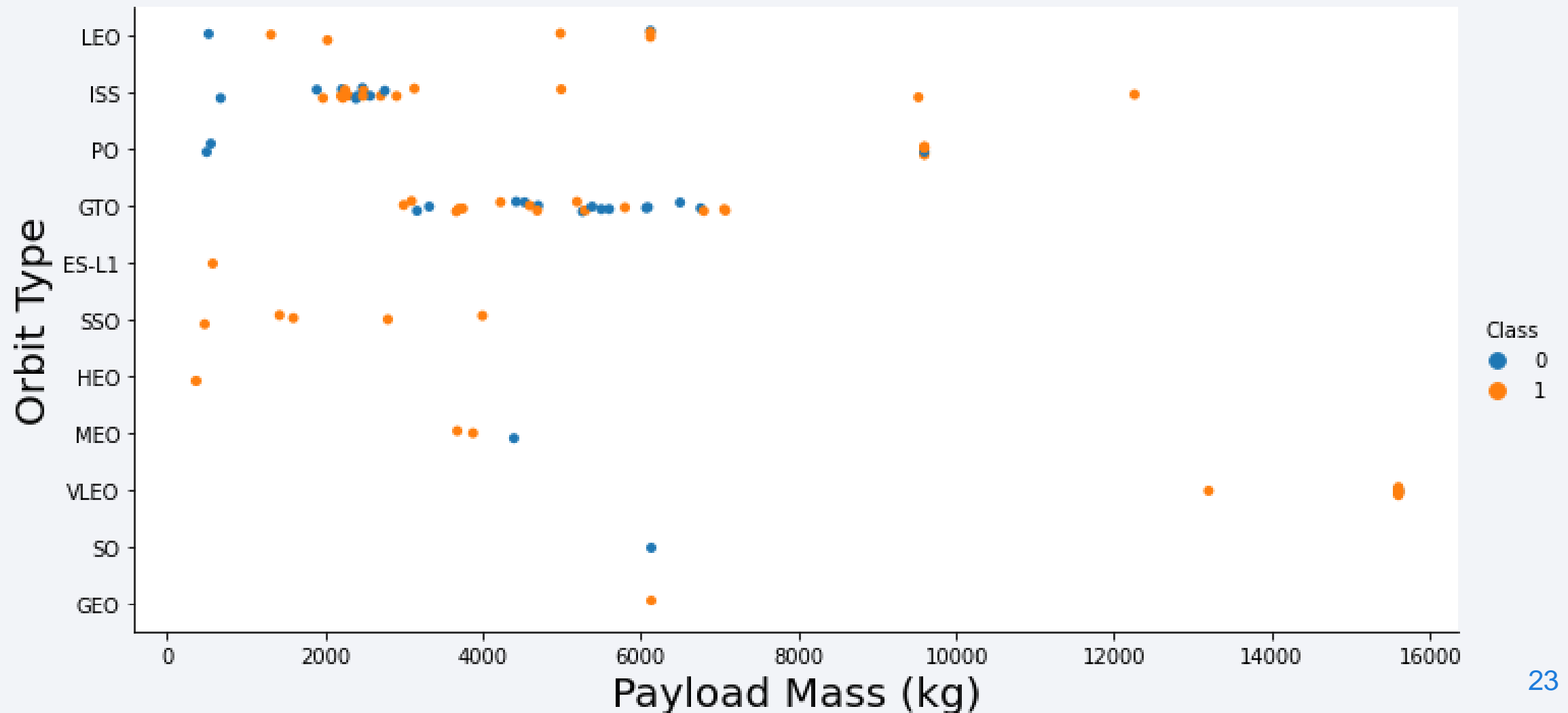


Flight Number vs. Orbit Type

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

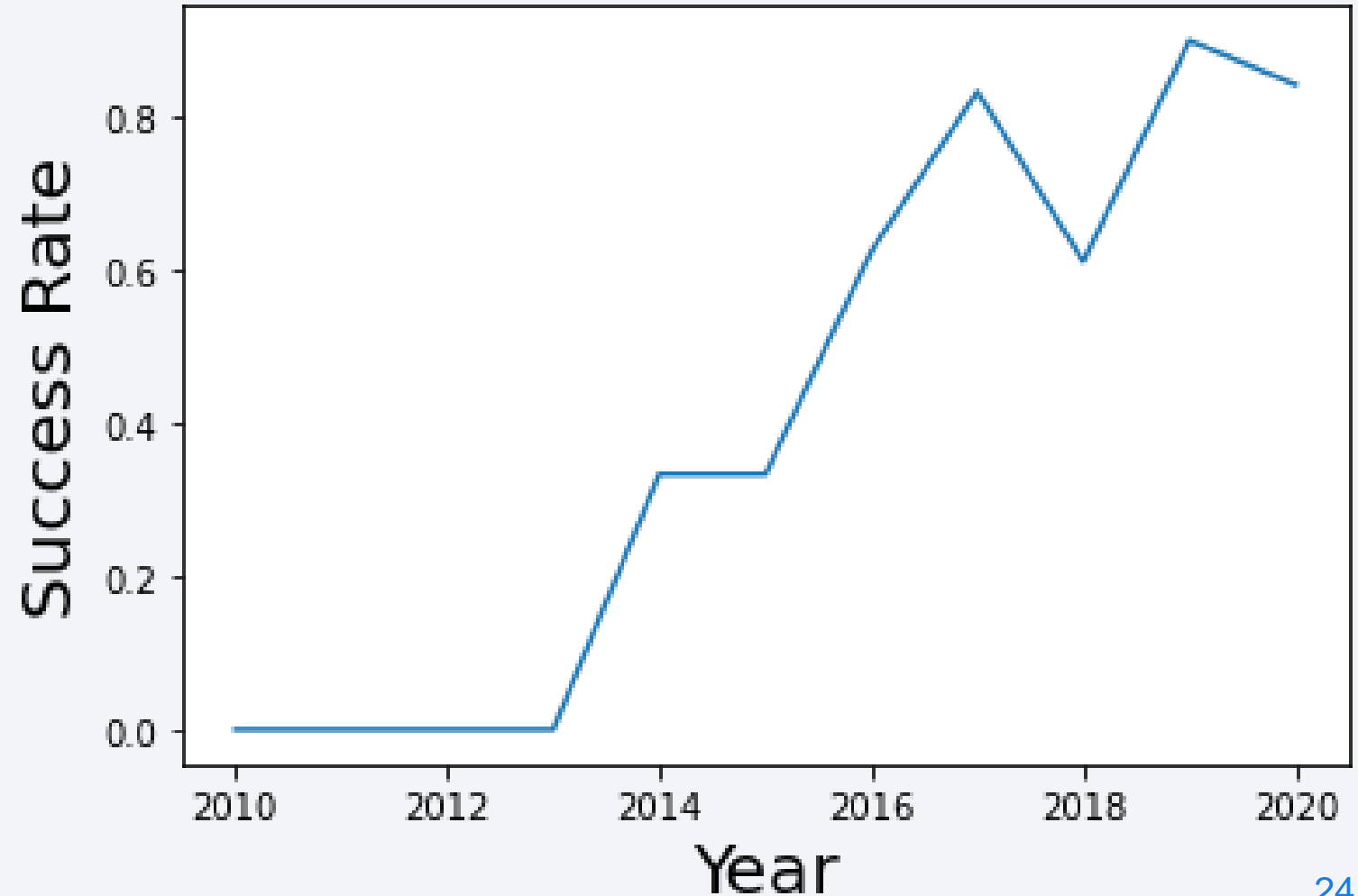


Payload vs. Orbit Type



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 Site Names Begin with 'CCA'

- 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

- Date of the first successful landing outcome on ground pad

	min(Date)
	2015-12-22

Successful Drone Ship Landing with 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

Total Number of Successful and Failure Mission Outcomes

- 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 2010-06-04 and 2017-03-20

- 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

A satellite view of Earth from space, showing the curvature of the planet and city lights at night. The image is a composite of a solid blue sky on the left and a satellite photograph of Earth on the right. The Earth's surface is dark, with numerous bright yellow and orange lights representing cities and urban areas. The horizon of the Earth is visible as a curved line separating the dark surface from the blue sky.

Section 3

Launch Sites Proximities Analysis

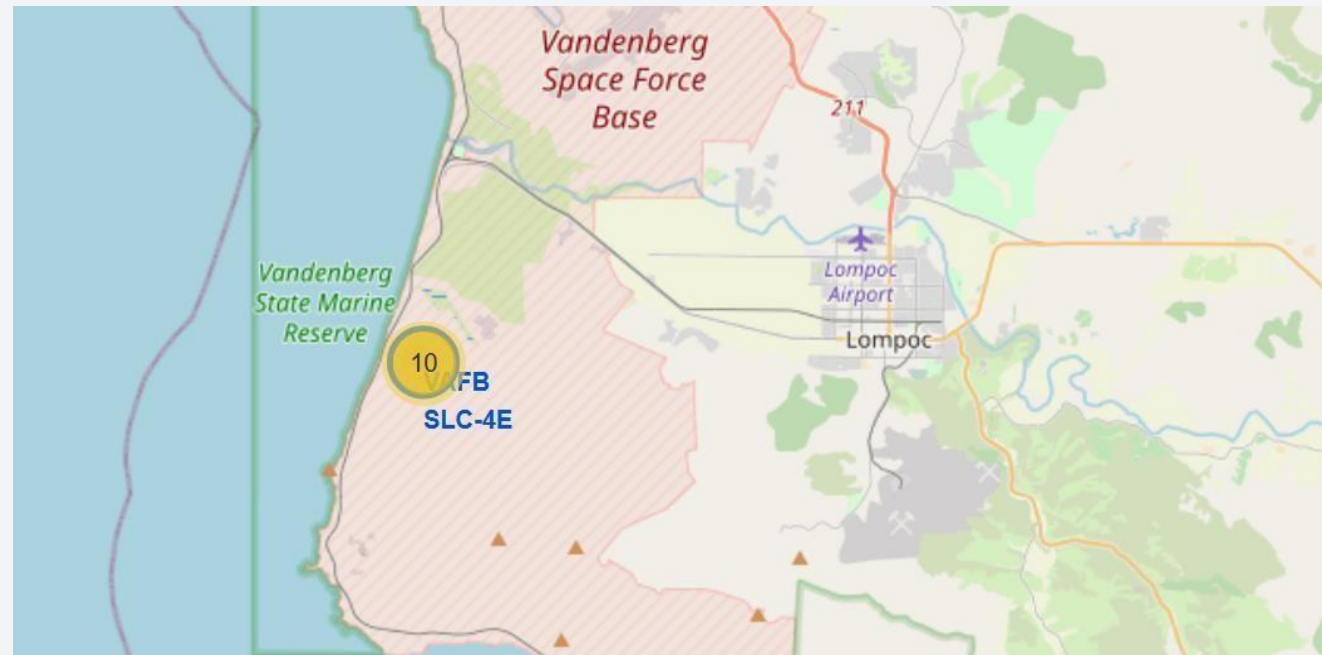
Folium Map: Launch Sites

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

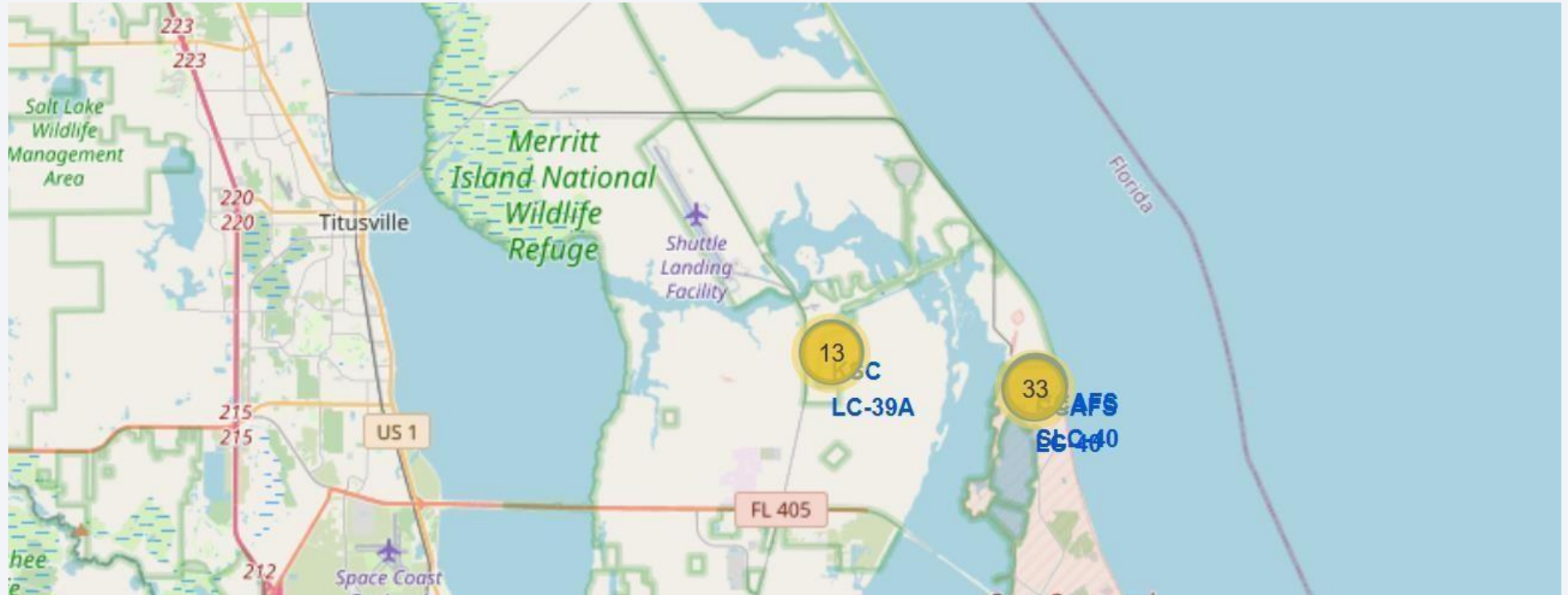


Folium Map: Proximity Vandenberg AFB

- Close to the Vandenberg 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

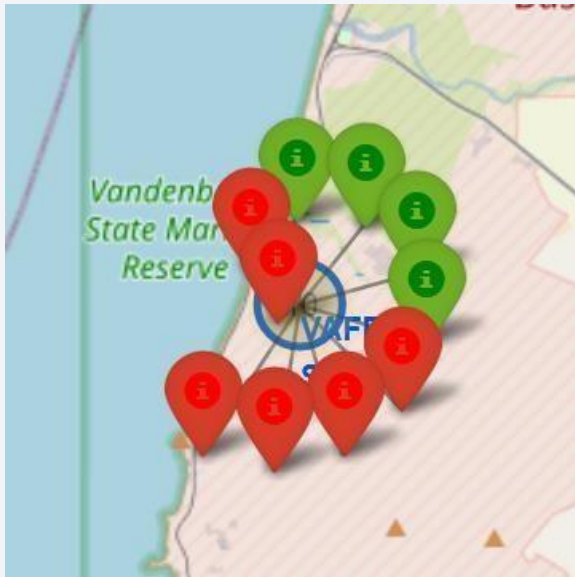


Folium Map: Proximity Kennedy SC/ Cape Canaveral



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

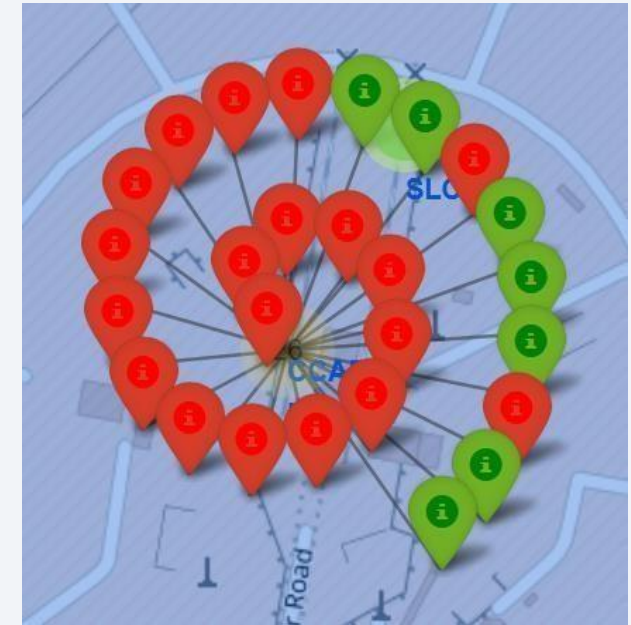
Folium Map: Stage-1 Landing Success by Launch Site



- Vandenberg AFB



- Kennedy Space Center



- Cape Canaveral

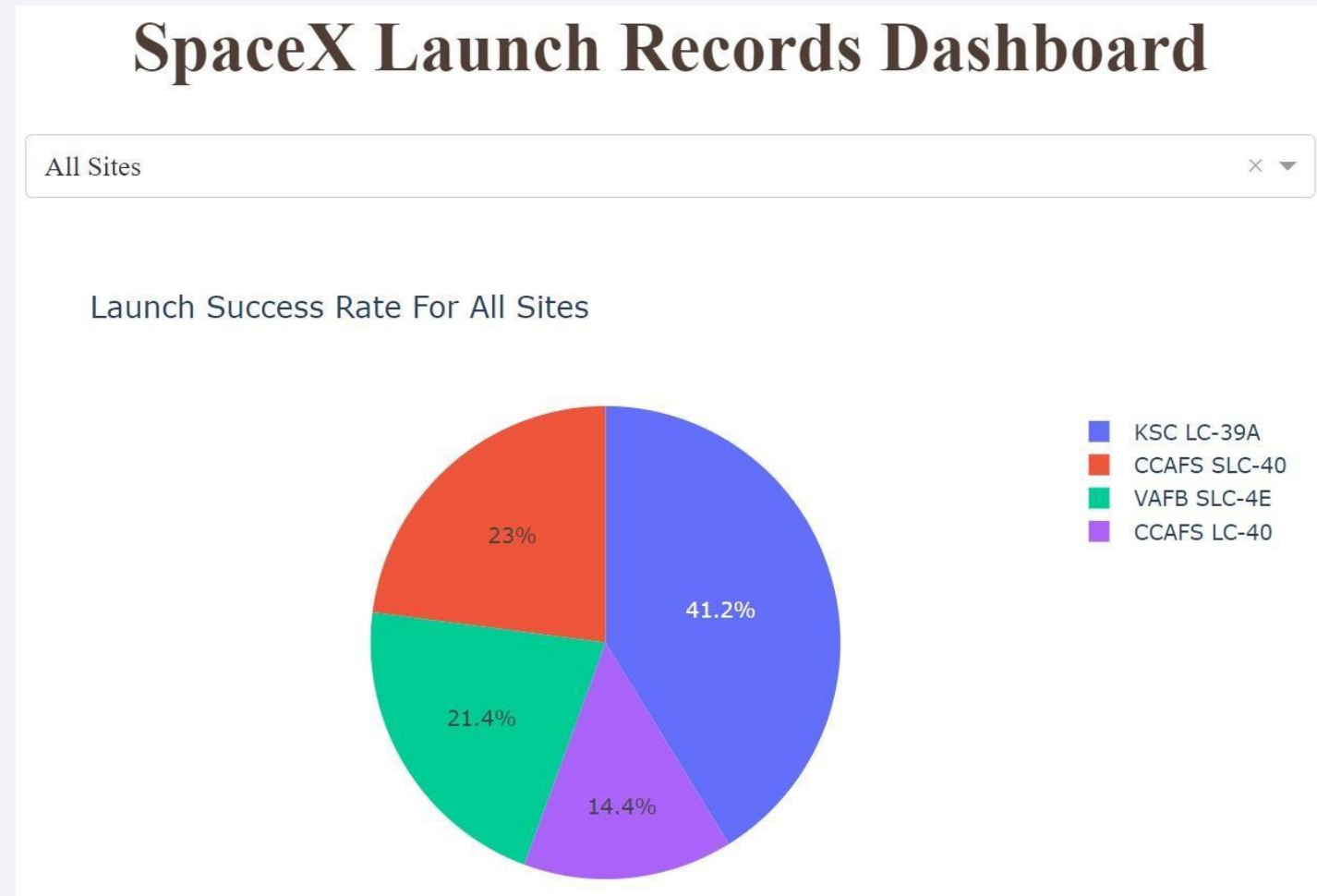


Section 4

Build a Dashboard with Plotly Dash

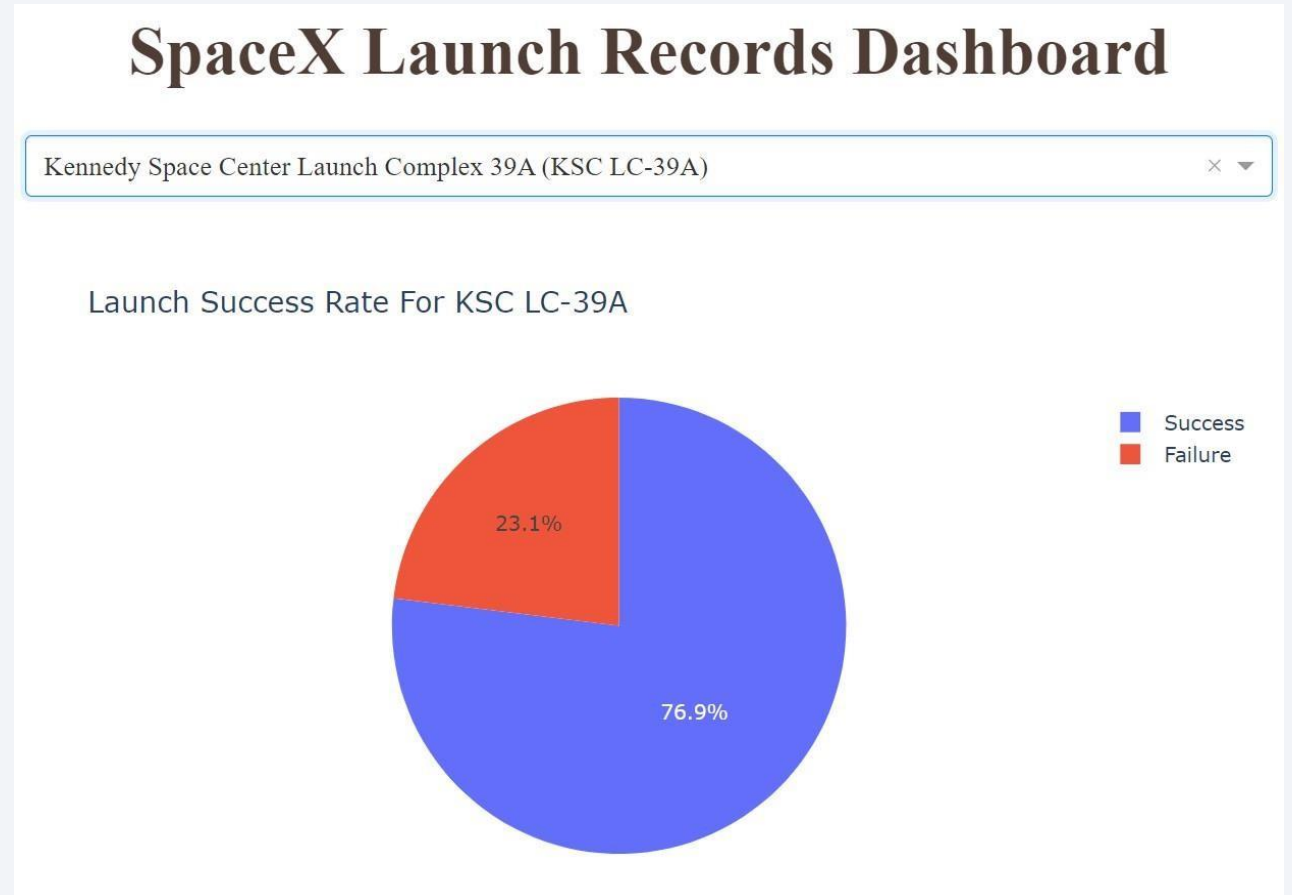
Dashboard: Launch Success Count For All Sites

- Kennedy Space Center (KSC LC-39A) has 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 for Booster versions v1.0 and v1.1 is quite small in the payload range to 10000kg



Dashboard: Booster Versions V1.0, V1.1

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



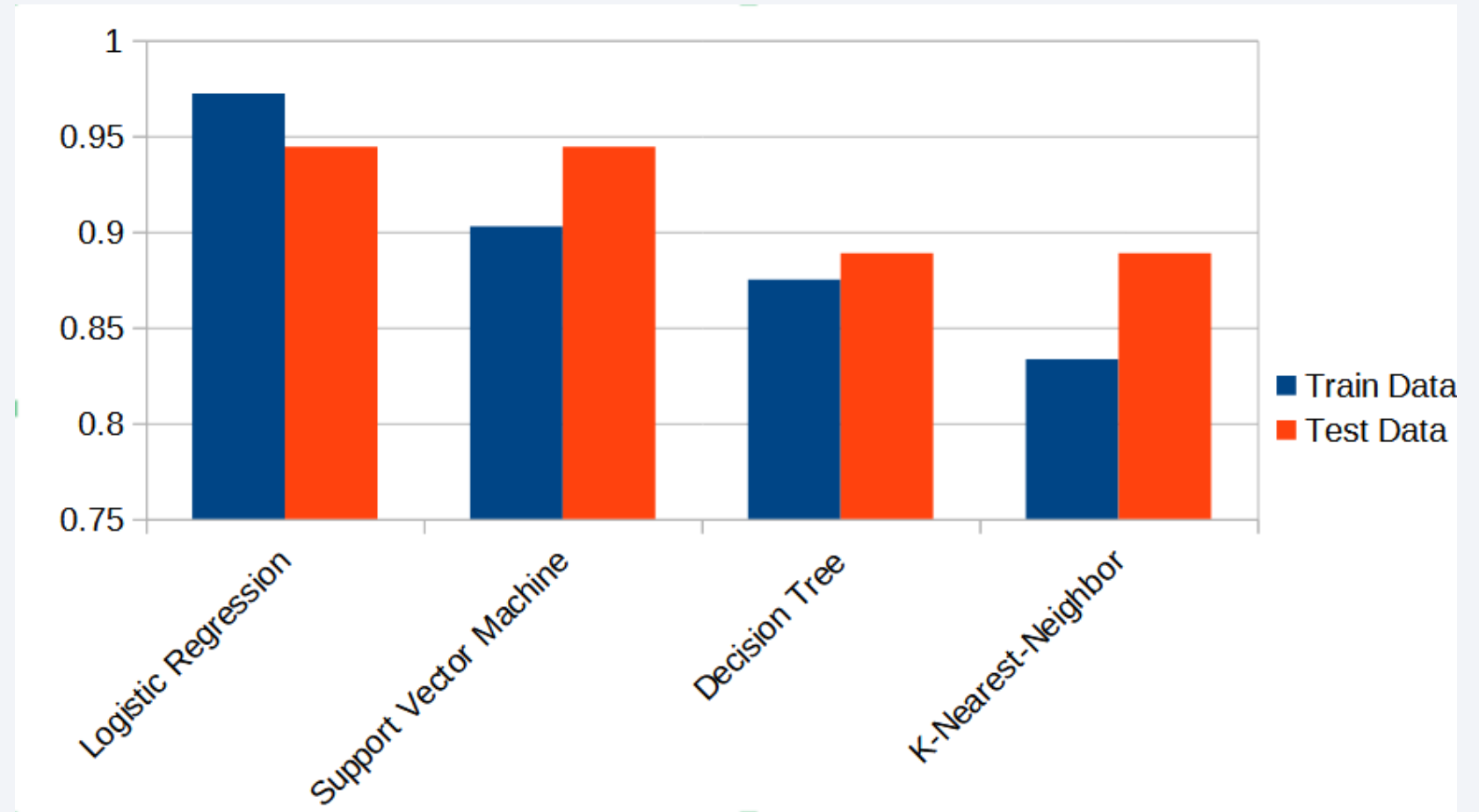


Section 5

Predictive Analysis (Classification)

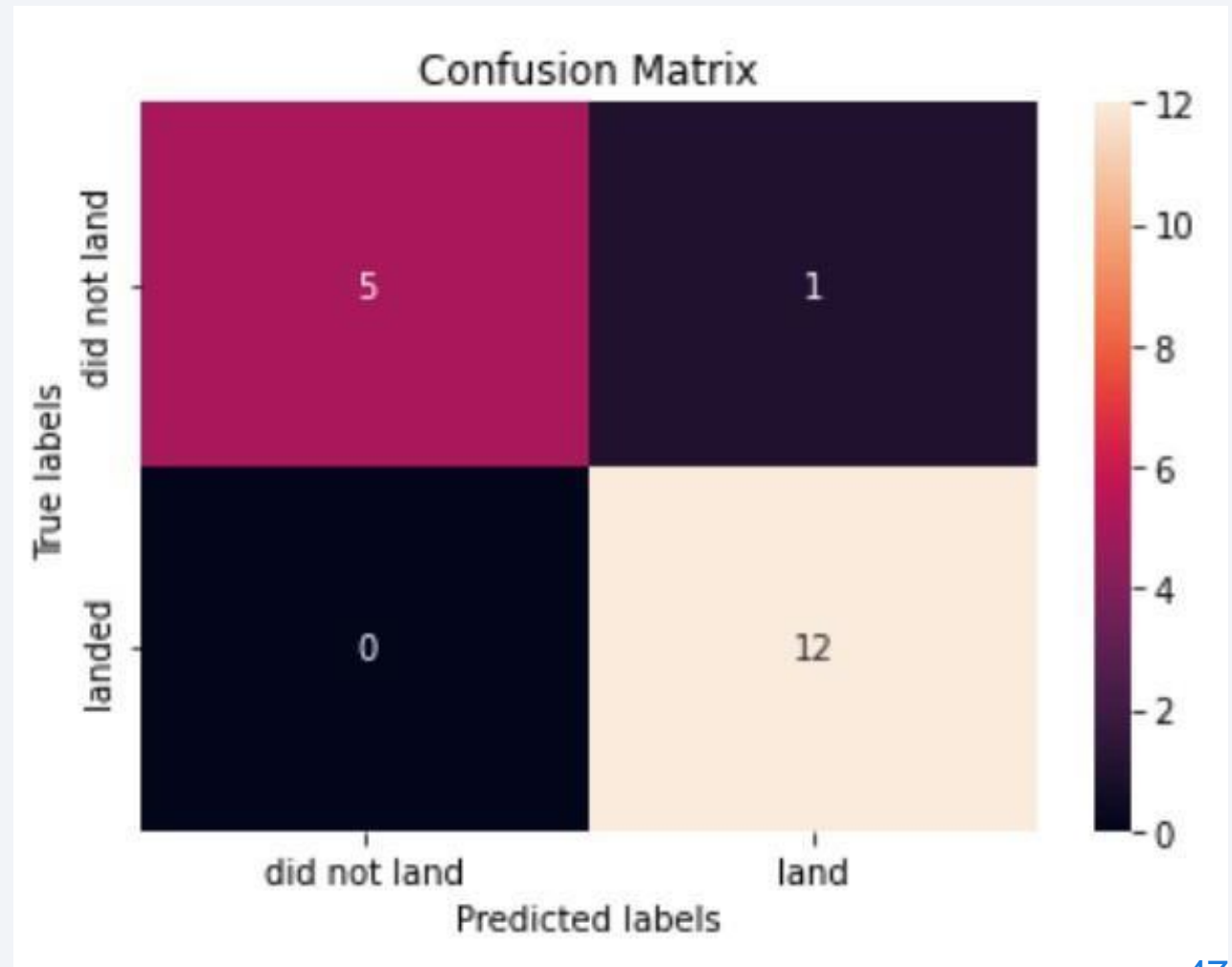
Classification Accuracy

- 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 provide a good result for predicting the landing outcome
- None of the models had false negatives
- All models had at least one false positive

Thank you!

