IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

C. Hatala
July 1, 2023

Outline

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

Executive Summary

Summary of methodologies

Data Collection: SpaceX API & SpaceX Wikipedia Page
Data Wrangling: Missing Values replaced by mean values
Exploratory Analysis: Orbit Type; Outcomes; Payload Mass; Launch Sites; Time; Bit Type
Visual Analysis Site Mapping

Interactive Dashboard: Analysis by Site, Payload and Booster Version(s)
Predictive Analysis with Classification; Logic Regression, SVM, KNN, Decision Tree

Summary of all results

Launch success rate increased over time
Orbit height related to success rate
Higher success rate for higher payload mass
Low success rate for booster versions v1.0, 1.1; high rate for FT, B4, B5
Higher success rate for Kennedy v Canaveral

Introduction

Project background and context

- *SpaceX advertises low-cost launches of Falcon9*
- *Success due to reusability of stage 1*

Problems you want to find answers

- *Determine whether stage 1 will land, determine launch cost(s)*
- *Information can be used for bidding against SpaceX for a launch.*

Section 1

Methodology

Methodology

Executive Summary

- **Data collection methodology:**
  - SpaceX API
  - Webscrape of SpaceX Wiki Page
- **Perform data wrangling**
  - Missing values replaced with mean values (payload mass)
- **Perform exploratory data analysis (EDA) using visualization and SQL**
  - Analyze outcome by orbit type
  - Analyzed outcome by payload mass and booster version(s) w/ SQL
  - Visual Analysis with charts by payload mass, time, orbit, site.
- **Perform interactive visual analytics using Folium and Plotly Dash**
  - Visual analysis with map by launch site
  - Interactive dashboard with analysis by site, payload, and boosters.
- **Perform predictive analysis using classification models**
  - Logic Regression, SVM, KNN, Decision Tree
  - Parameter tuning w/ Grid Search

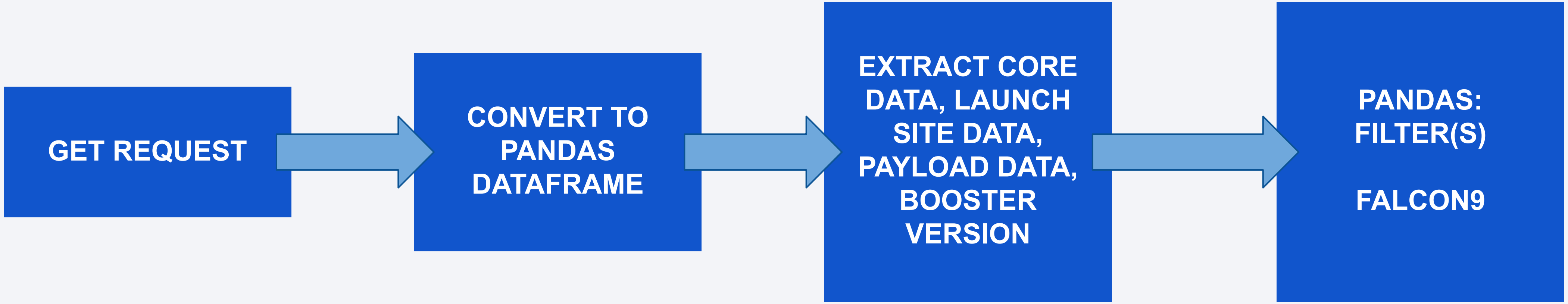
Data Collection

- SpaceX REST API
  - RESTful interface
  - Get core data
  - Get booster version
  - Get launch site data
  - Get payload data
- Webscrape of SpaceX wiki page
  - HTML Requests(HTTP-get)
  - Python/BS (BeautifulSoup)
  - Extract column names from HTML headers

Data Collection Jupyter Notebook

Data Collection – SpaceX API

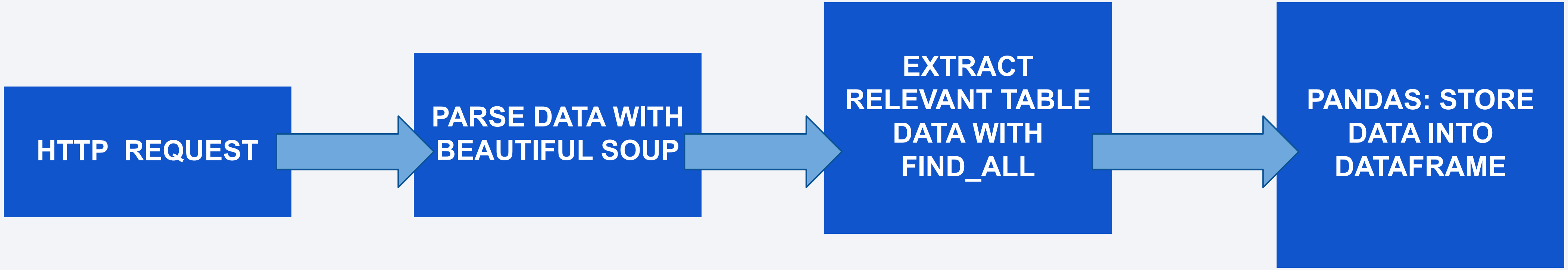
- Send get request to SpaceX API
- Parse data into pandas dataframe
- Extract data with specific functions for core launch data, site data, payload mass and booster versions.
- Since data contains other than Falcon9 data, we filter for Falcon9 data only.



Data Collection - Scraping

- Send HTTP request to SpaceX wiki page
- Parse data into Pandas dataframe with BeautifulSoup webscraper
- Extract data with find_all method
- Store data in Pandas dataframe for later use

- Add the GitHub URL of your completed EDA with data visualization notebook, as an external reference and peer-review purpose



Data Wrangling

- Dealing with missing values: The column for payload mass was missing several data points. They were filled with mean values.

[Data Wrangling]

EDA with Data Visualization

Charts

- Payload mass v Flight number v Success rate: This shows us the development of the payload mass and the success rate over time.
- Launch site v Flight number v Success rate: This shows us the success rate of each launch site over time.
- Launch site v Payload mass v Success rate: This shows us which payload is most likely to succeed at a specific site
- Orbit type v Success rate: This can provide insight into which orbits have the highest success rate
- Orbit type v Flight number v Success rate: This shows the development of orbit type over time
- Orbit type v Payload mass v Success rate: This shows the success rate for specific clusters of payload mass/orbit type
- Success rate v Year: This shows the success of development over time

Data Exploration

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
- Display the average payload mass carried by booster version F9 v1.1
- List the date when the first successful landing was achieved on ground pad
- List the names of boosters which have success in drone ship and have a payload mass great 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, booster versions, and the launch sites for 2015.
- Rank the count of landing_outcomes (ie. Failure(drone)) or success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

[Data Analysis]()

Build an Interactive Map with Folium

- Map Objects
  - Edged Circles; Launch Sites (1000m radius)
  - Markers: labeling all objects
  - MarkerCluster: markers around launch sites (red - failure; green - success) to indicate stage1 landing.
  - Lines: Measure the distance between the launch site and the next coast or next city

[Folium Interactive Dashboard]

Build a Dashboard with Plotly Dash

- **Input Elements**
  - Dropdown list for the launch site (w/ 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 selected) to show the number of successful landings
  - Scatterplot to show the success/failure rate by payload/booster

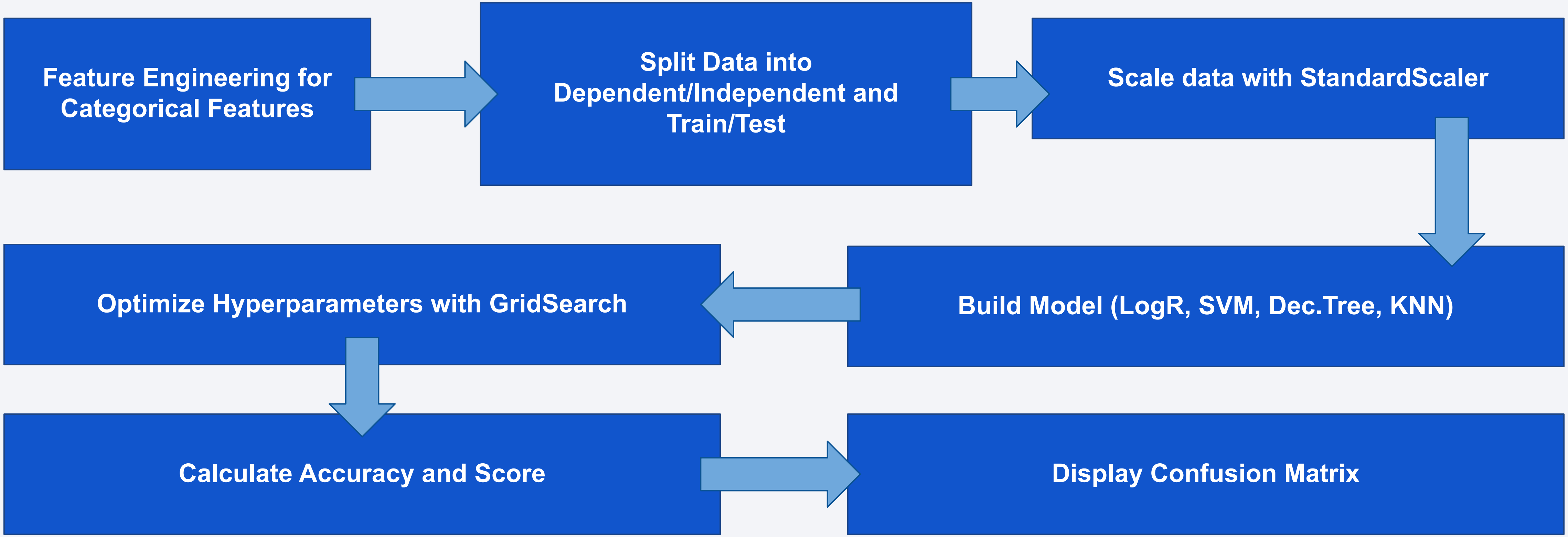
[PLOTLY Interactive Dashboard (Download)]

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 parameters
- **Evaluation**
  - Use accuracy of GridSearch for selecting the best parameter
  - Use score to compare each classification method

[Machine Learning]

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, 1.1**, high success rate for **FT, B4, B5**

  Higher success rate for **Kennedy Center v Canaveral**

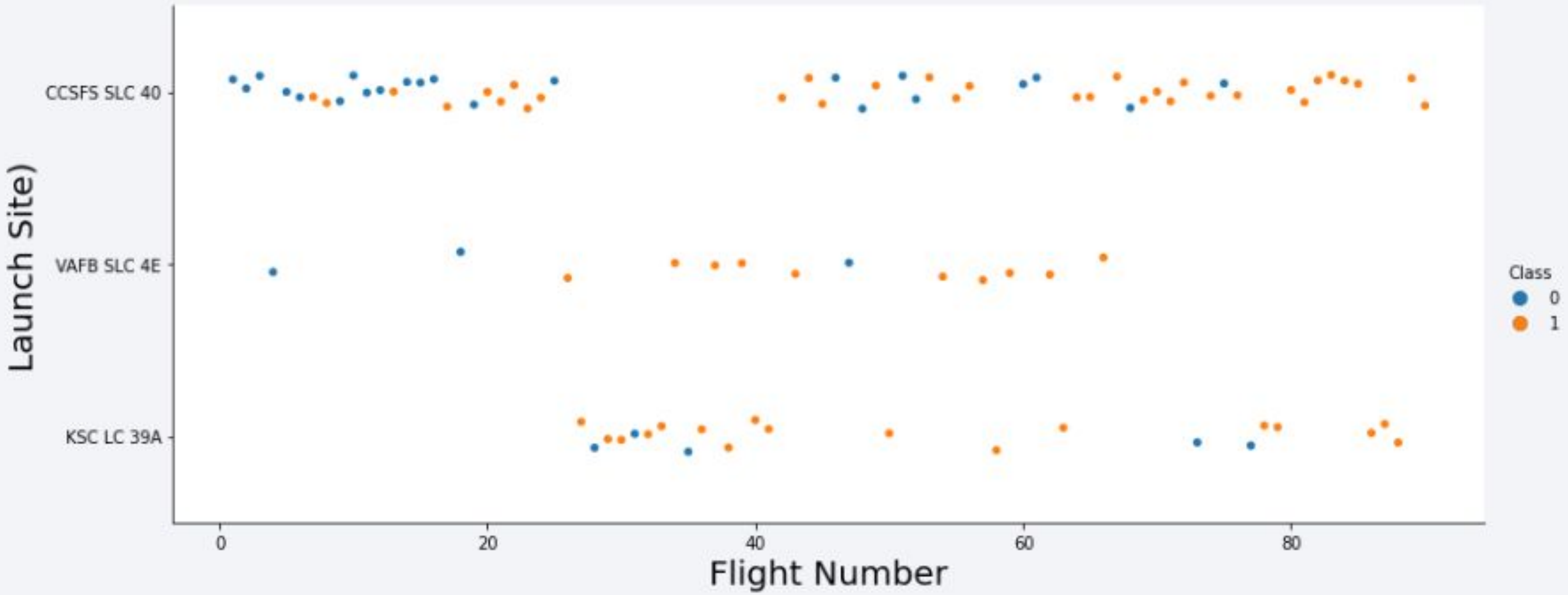
- **Predictive analysis results**

  Best prediction results with **logistic regression** and **support vector machine**

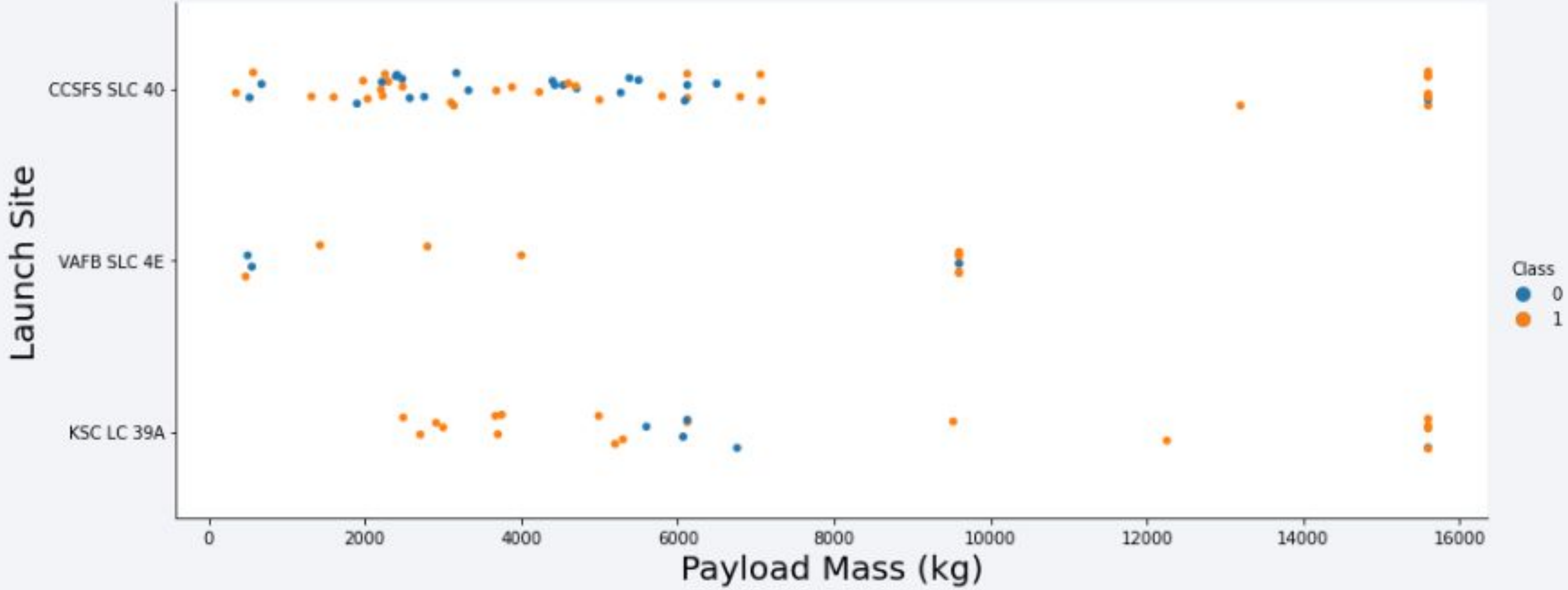
Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

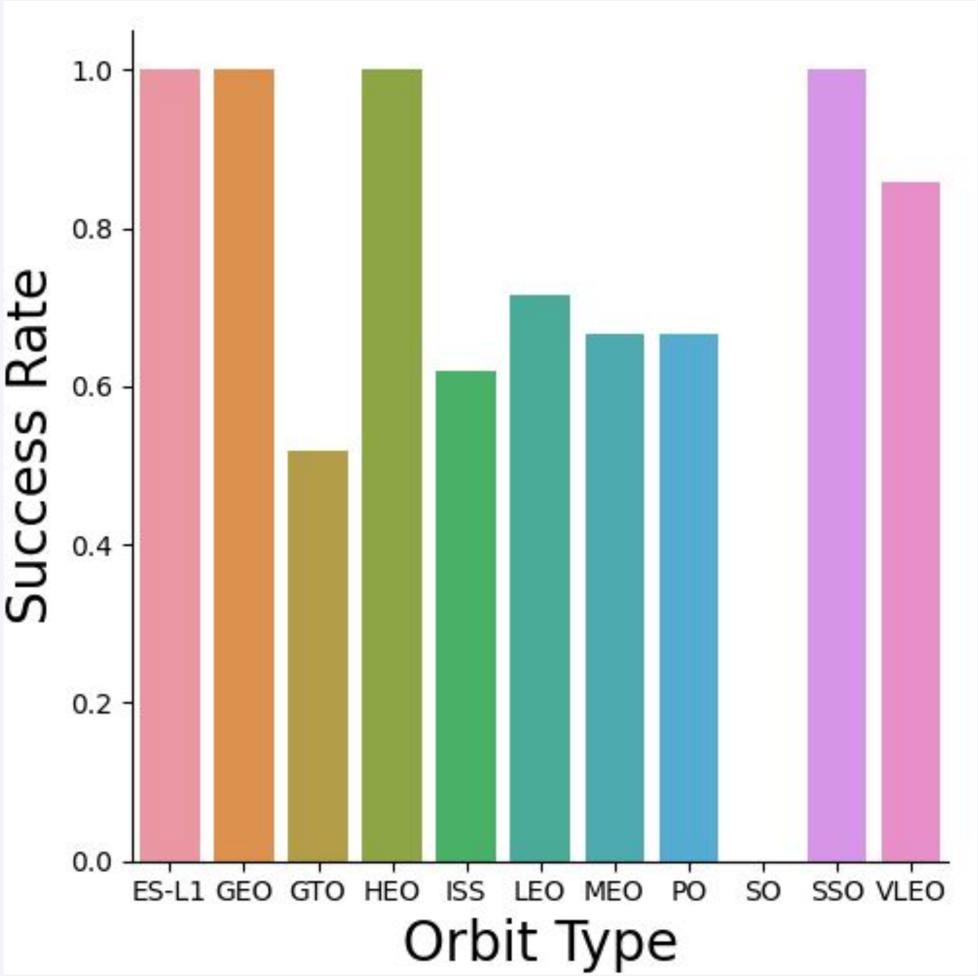


Payload vs. Launch Site



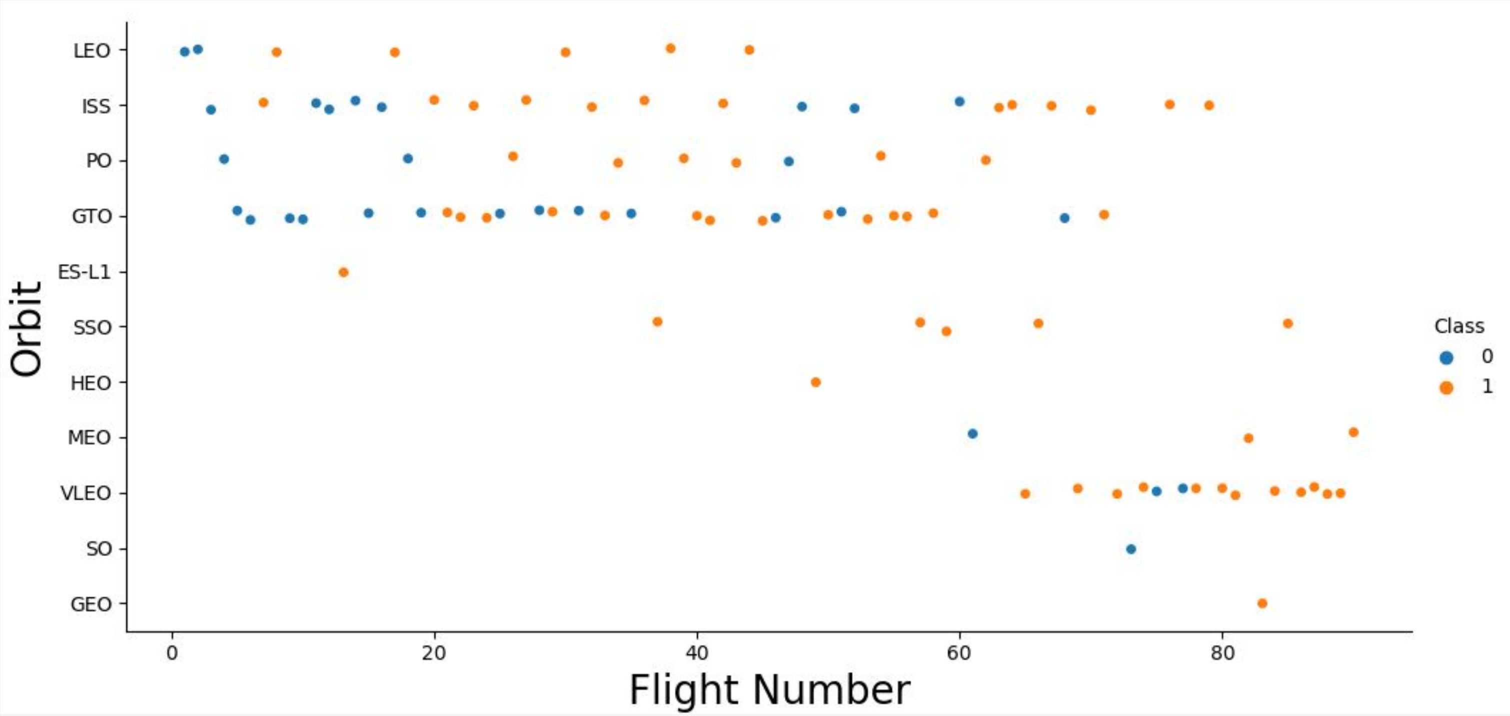
Success Rate vs. Orbit Type

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

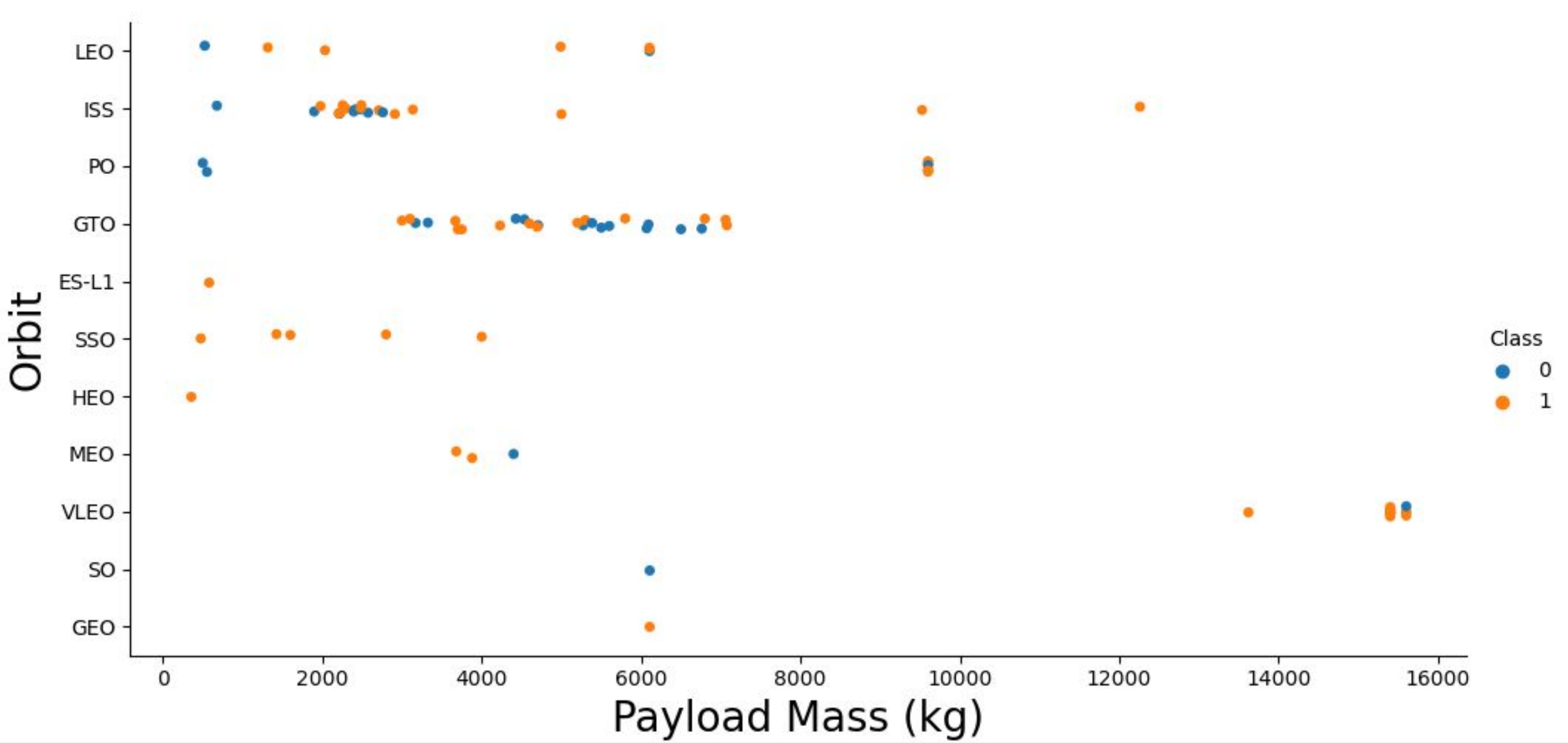


Flight Number vs. Orbit Type

- The predominant orbit types changed over time. The success rate increased over time for ALL orbit types.

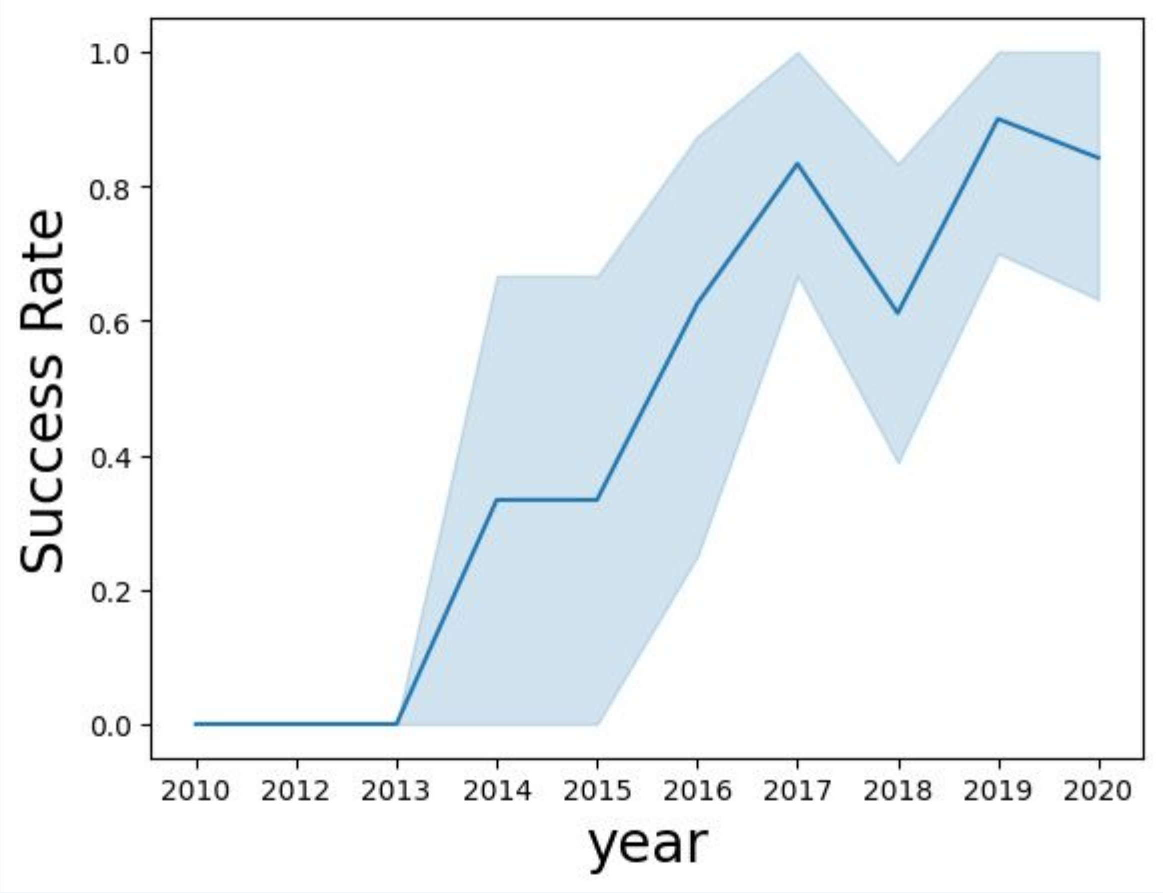


Payload vs. Orbit Type



Launch Success Yearly Trend

- Launch success has increased over years



All Launch Site Names

- KSC: Kennedy Space Center
- CCA: Cape Canaveral Launch Center
- VAFB: Vandenburg,AFB

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

Launch Site Names Begin with 'CCA'

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

$$sum(payload\_mass\_kg) = 45596$$

sum(PAYLOAD_MASS_KG)
45596

Average Payload Mass by F9 v1.1

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

$$Avg, Payload\_Mass\_KG = 2928.4$$

avg(PAYLOAD_MASS_KG)
2928.4

First Successful Ground Landing Date

`Date of the first successful landing outcome on ground pad`

$$min(Date)=2015\text{-}12\text{-}22$$

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 BT B1022
1	F9 BT B1026
2	F9 BT B1021.2
3	F9 BT B1031.2

Total Number of Successful and Failure Mission Outcomes

```
Total number of successful and failure mission outcomes
```

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 1049.5
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. Drone ship, Booster Version, Launch Site(s) accounted for. 2015.

landing_outcome	booster_version	launch_site
Failure (drone)	F9 v1.1B1012	CCAFS LC-40
Failure (drone)	F9 v1.1B1015	CCAFS LC-40
Failure (drone)	F9 v1.1B1017	VAFB SLC-4E
Failure (drone)	F9 FT B1020	CCAFS LC-40
Failure (drone)	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 on a drone ship v Success on a ground pad (Between the date 2010-06-04 and 2017-03-20) in descending order.

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

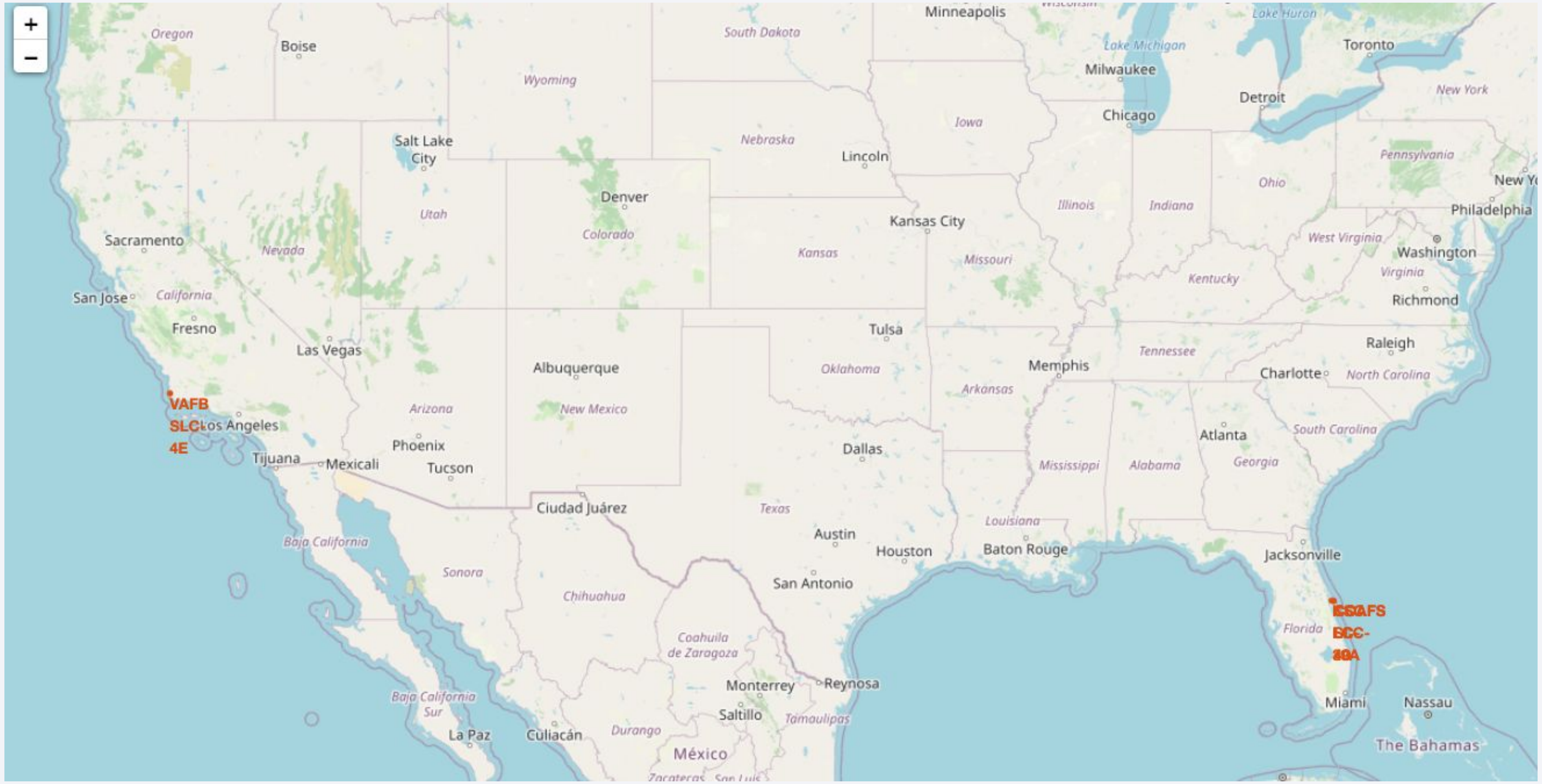
Section 3

Launch Sites Proximities Analysis

Folium Map - Launch Sites

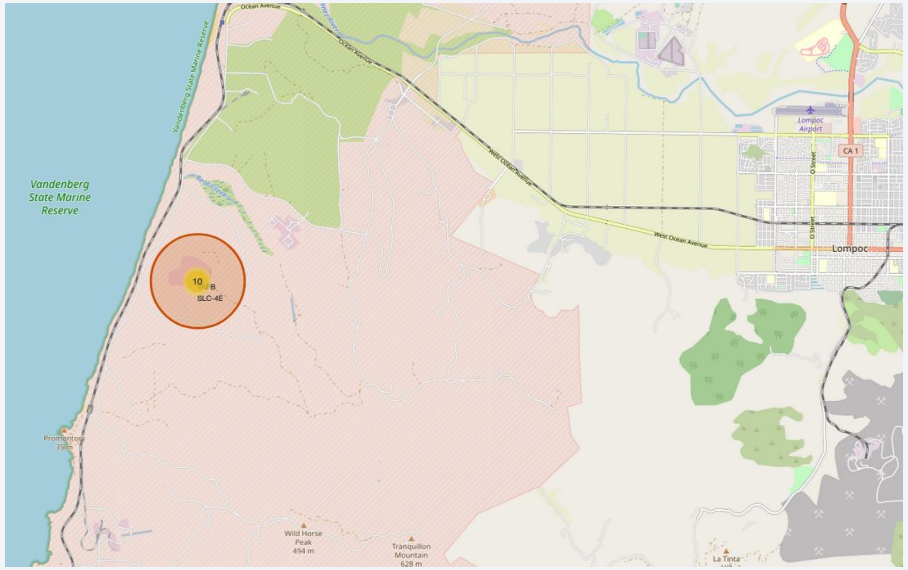
Launch sites are at the East and West coast, in Florida and California

[Interactive Map with Folium]



Folium Map - Vanderburg AFB

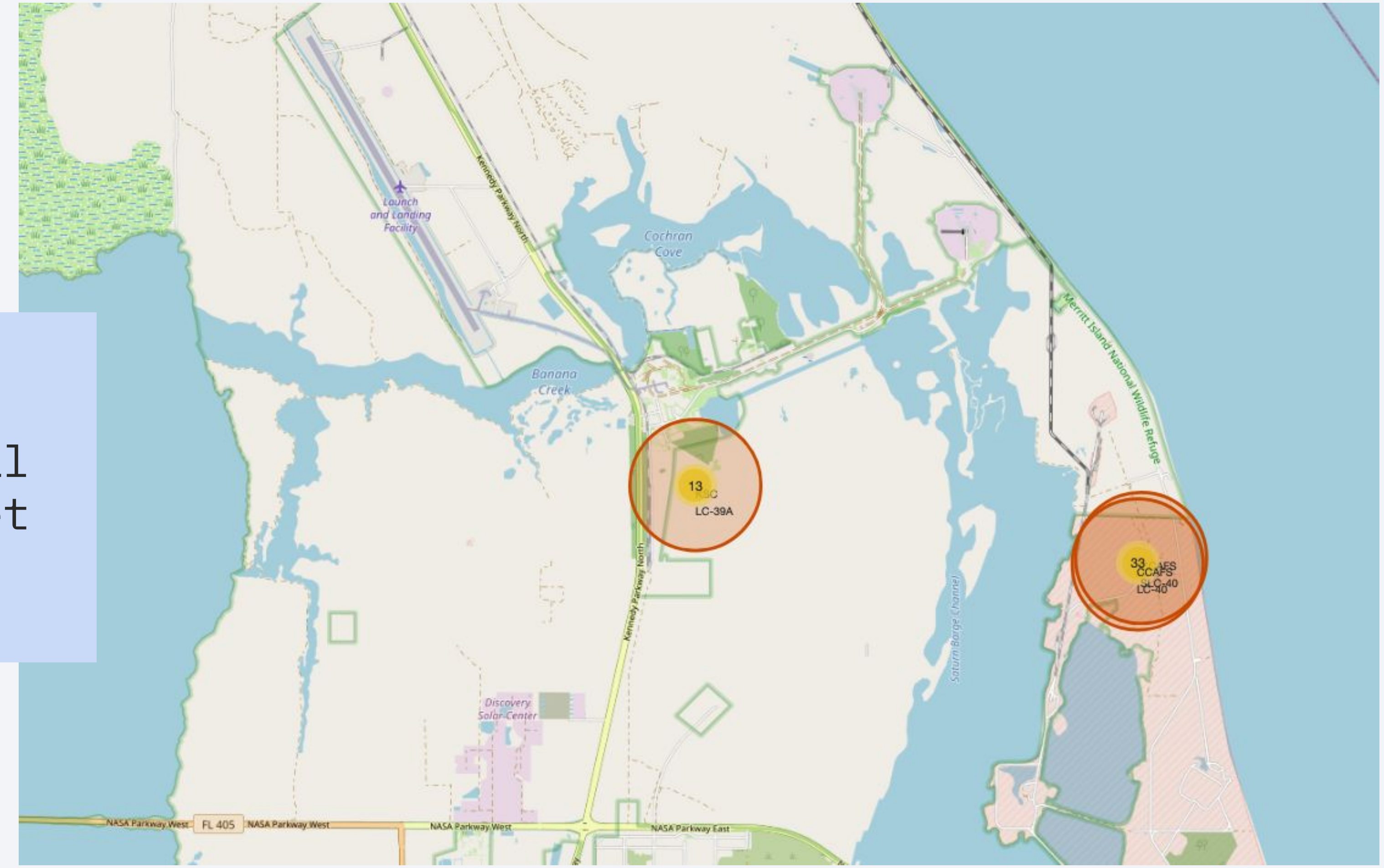
Close to the Vandenburg AFB is the town of Lompoc. This may be a safety issue if a stage 1 landing were to lose communication, since rockets generally come downward in an Eastern direction.



[Interactive Map with Folium]

Folium Map - Proximity KSC/CCA

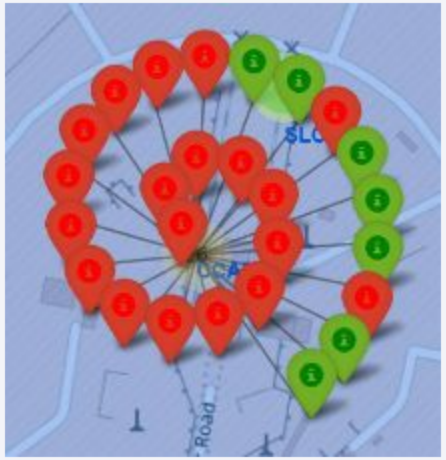
No city towards the Eastern Direction, ideal place for testing rocket launches



Folium: Stage 1 Landing Success by Site

Vandenburg 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

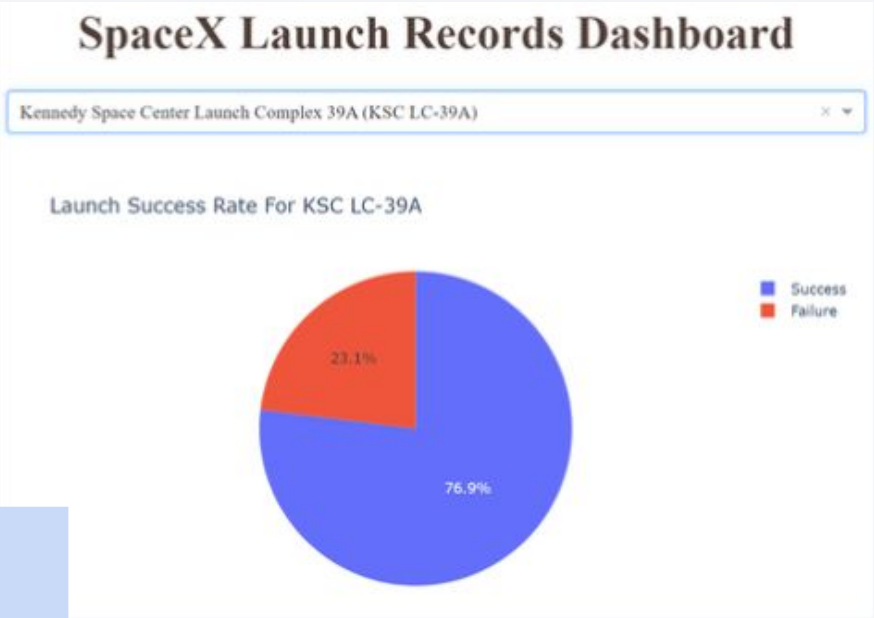
Has the most successful landings for Stage(1)

Vandenberg AFB

VAFB SLC-4E

Has the least successful landings for Stage(1)

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, 1.1 is **limited in the payload range to 10,000kg**

Dashboard - Booster Versions v1.0, v1.1

Success rate for Booster versions FT, B4 and B5 is **better in the payload range to 10,000kg**

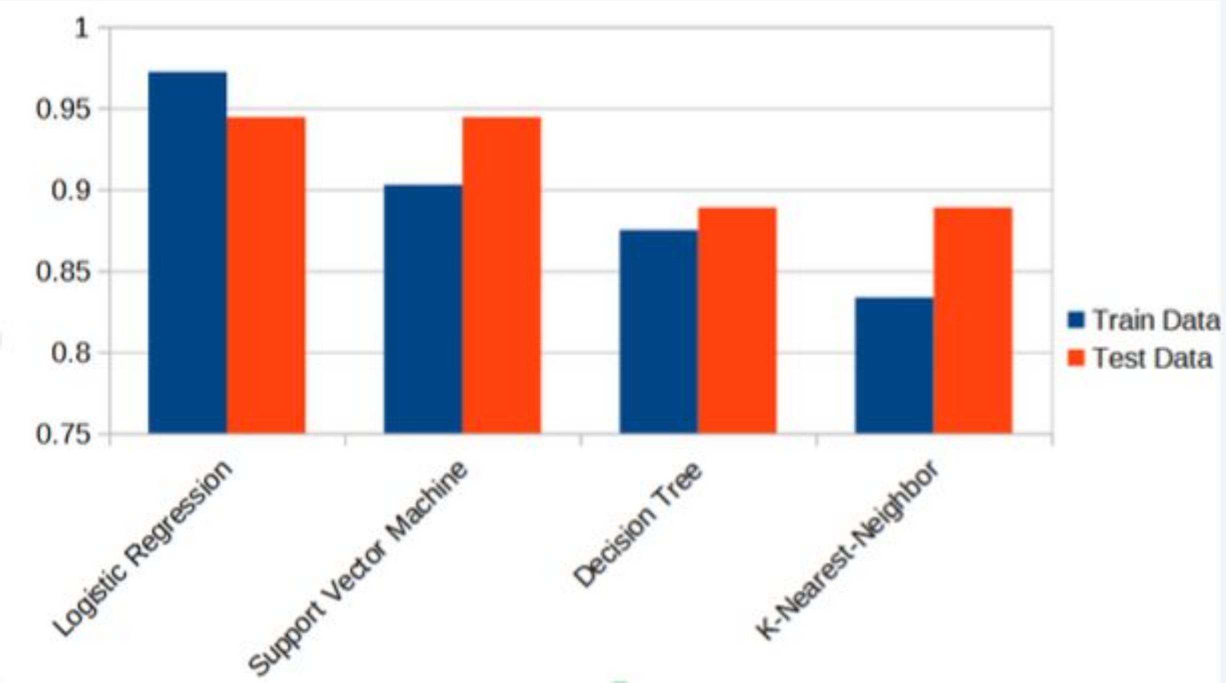
Section 5

Predictive Analysis (Classification)

Classification Accuracy

- Logistic Regression has the best result for train data

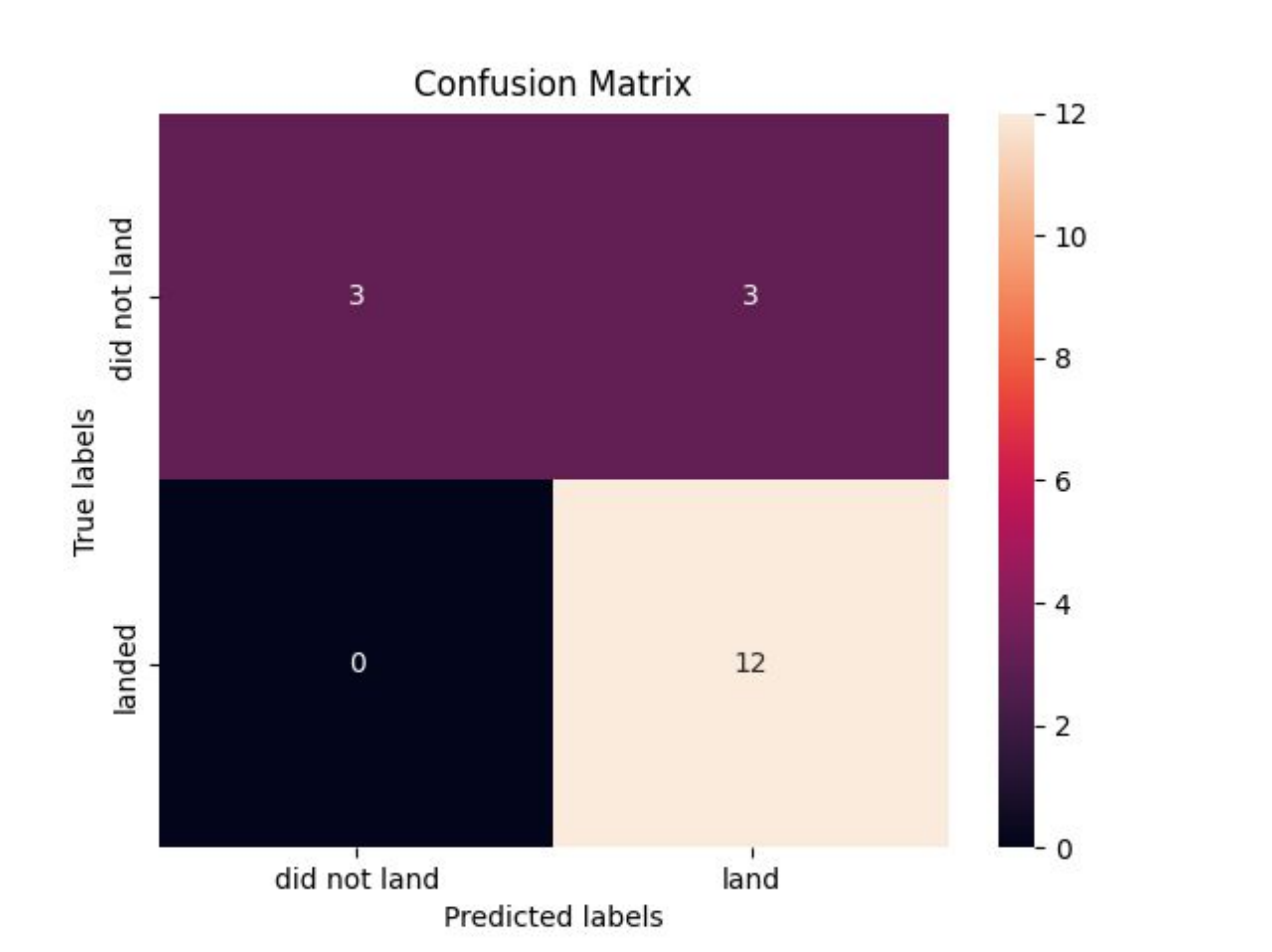
- Logistic Regression and Support Vector Machines have the best results on test data



Confusion Matrix

```
True Positives   12
True Negatives   3
```

```
False Positives  3
False Negatives  0
```



Conclusions

All models had at least one false positive

None of the models had false negatives

Prediction with Logistic Regression is relatively accurate

Support Vector Machine also provides a good analysis for predicting landing outcomes

Appendix

All Python code and SQL can be inspected here:

1. Jupyter Notebook
2. Plotly Dashboard

Current version of this document can be downloaded from the following link:

3. Analysis Presentation

Thank you!