IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Jacob Waters
12.18.2023

Outline

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

Executive Summary

- Summary of methodologies
  - Data collection using an API
  - Data collection using web scraping
  - Data Wrangling
  - EDA using SQL
  - EDA using Python visualizations
  - Interactive maps
  - Machine learning to test Stage 1 reusability
- Summary of all results
  - Visualizations
  - Predictive analysis

Introduction

- Background
  - Space Y would like to compete with SpaceX, however understanding the cost and price of each launch is critical to the success of market entry
- Problem Statement
  - Will SpaceX reuse the first stage of their Falcon 9 rocket

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was collected using publicly available SpaceX launch data. The data was sourced using an API through Python.
- Perform data wrangling
  - Using Python, separate data tables were parsed together, and successful landings were categorized alongside unsuccessful landings.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection

Data source identified → Permissions granted → Connection established → Data quality viewing

Data source identified

SpaceX API

Permissions granted

Publicly available data

Connection established

Python connection using ‘get’ functions

Data quality viewing

Check for missing values, view data types, normalize json files

Data Collection – SpaceX API

- SpaceX REST calls
  - def getBoosterVersion(data): for x in data['rocket’]:
  - def getLaunchSite(data): for x in data['launchpad’]
  - def getPayloadData(data): for load in data['payloads’]:
  - def getCoreData(data): for core in data['cores']:
- Add the GitHub URL of the completed SpaceX API calls notebook (must include completed code cell and outcome cell), as an external reference and peer-review purpose

```
Def getXdata(data):

For x in data[‘column’]:


spacex_url=https://api.spacexdata.com/v4/launches/past

response = requests.get(spacex_url)

print(response.content)


data = pd.json_normalize(response.json())
```

Data Collection - Scraping

- Request the Falcon9 Launch Wiki page from its URL
- Add the GitHub URL of the completed web scraping notebook, as an external reference and peer-review purpose

Import required packages

‘def’ and ‘get’ functions to pull specific data from HTML

Define URL

Create a ‘Beautiful Soup’

Convert to a dataframe

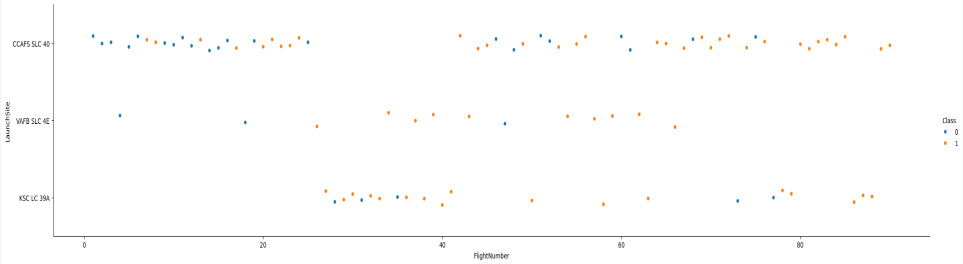
Data Wrangling

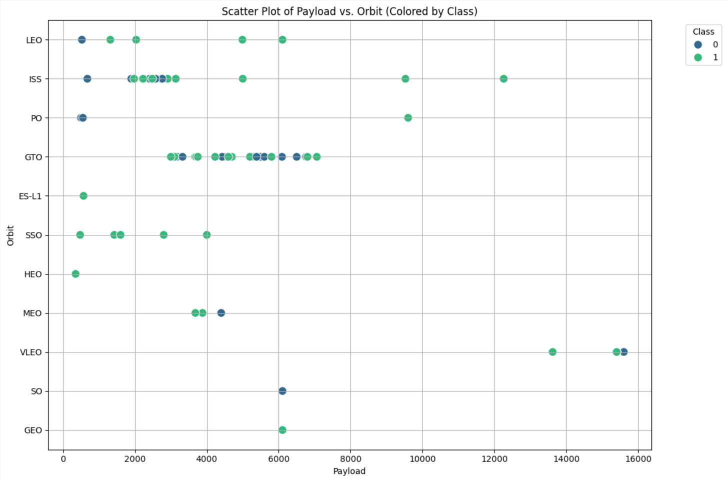
- Data was processed and cleaned of inconsistencies
  - Exploratory visualizations were created for better understanding of the data

- Connect to data > Missing data viewed > success class created > data normalized > data frame created
- Add the GitHub URL of your completed data wrangling related notebooks, as an external reference and peer-review purpose

EDA with Data Visualization

- Conducted an in-depth data analysis process using Python libraries for visualization such as Pandas and Seaborn
    - Scatter plots and bar graphs were utilized for EDA





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

EDA with SQL

- %load_ext sql
- import csv, sqlite3 con = sqlite3.connect("my_data1.db") cur = con.cursor()
- %sql sqlite:///my_data1.db
- %sql create table SPACEXTABLE as select * from SPACEXTBL where Date is not null
- SELECT DISTINCT launch_site FROM SPACEXTABLE; WHERE launch_site LIKE 'CCA%’ LIMIT 5;
- Add the GitHub URL of your completed EDA with SQL notebook, as an external reference and peer-review purpose

Build an Interactive Map with Folium

- Added circles around the launch sites and distinctive features to relate whether a successful launch occurred there
- This aided in visualizing which sites were optimal for a successful flight
- Add the GitHub URL of your completed interactive map with Folium map, as an external reference and peer-review purpose

Build a Dashboard with Plotly Dash

- Created an interactive dashboard which visualized launch site data by developing a callback function which was also used to create a pie chart
- This was done to better understand the data and the underlying problem statement
- Add the GitHub URL of your completed Plotly Dash lab, as an external reference and peer-review purpose

Predictive Analysis (Classification)

- Summarize how you built, evaluated, improved, and found the best performing classification model
- You need present your model development process using key phrases and flowchart
- Add the GitHub URL of your completed predictive analysis lab, as an external reference and peer-review purpose

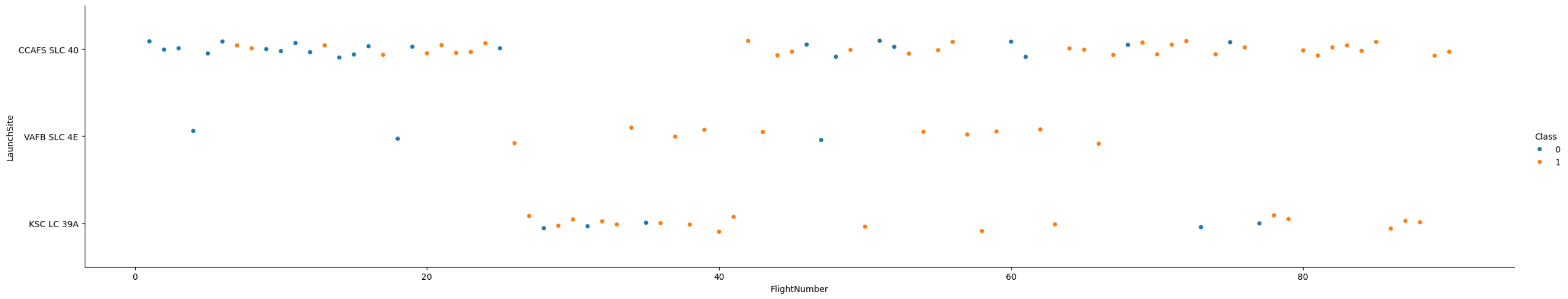
Results

- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

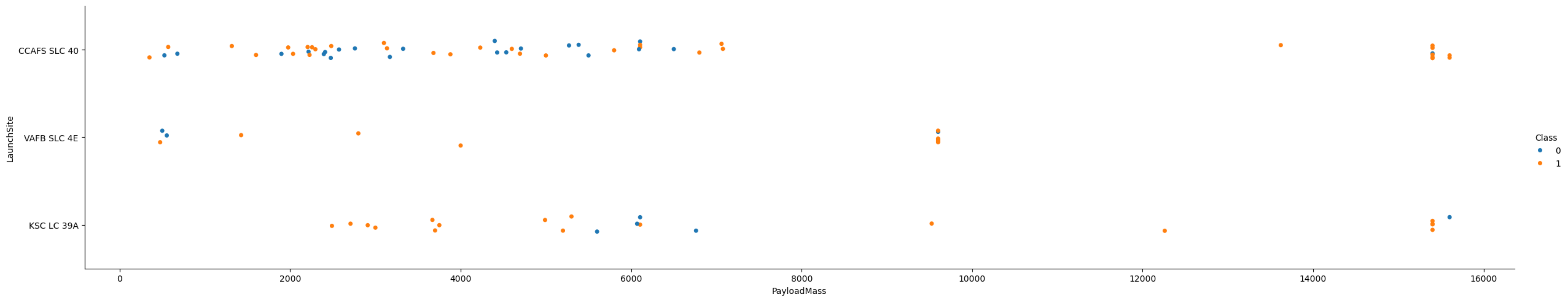
Insights drawn from EDA

Flight Number vs. Launch Site



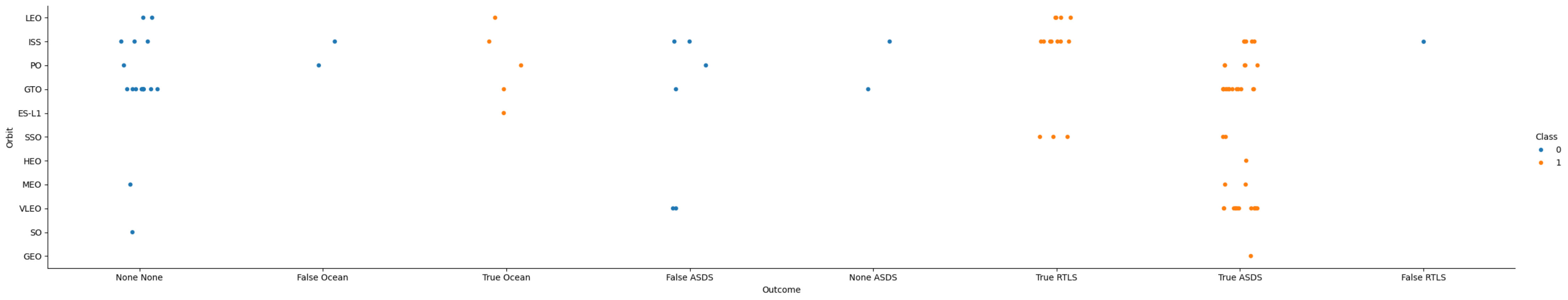
- This scatterplot shows the relationship between the place of launch and the flight number
- SLC-40 is heavily utilized as a launch site

Payload vs. Launch Site



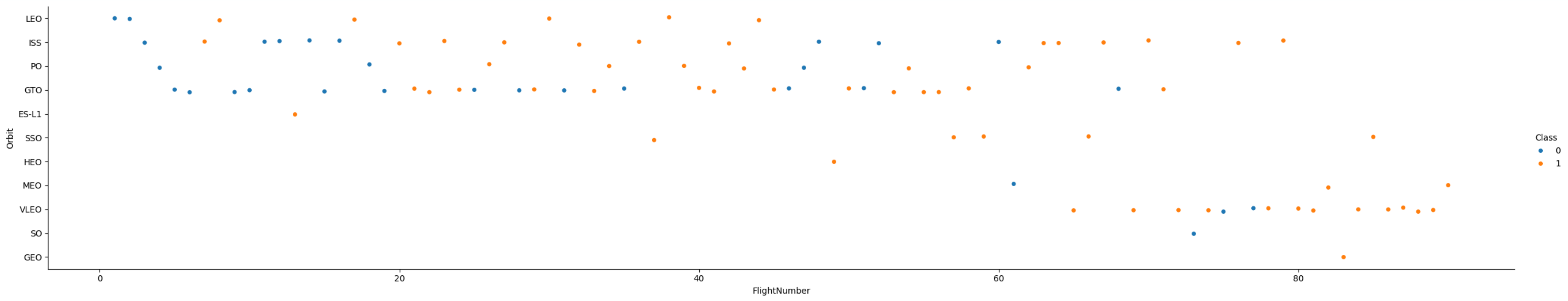
- This scatterplot shows the relationship between the payload size and the place of launch
- VAFB is seldom used and not at all at higher payloads
- Most payloads below 7000 happen at SLC-40

Success Rate vs. Orbit Type



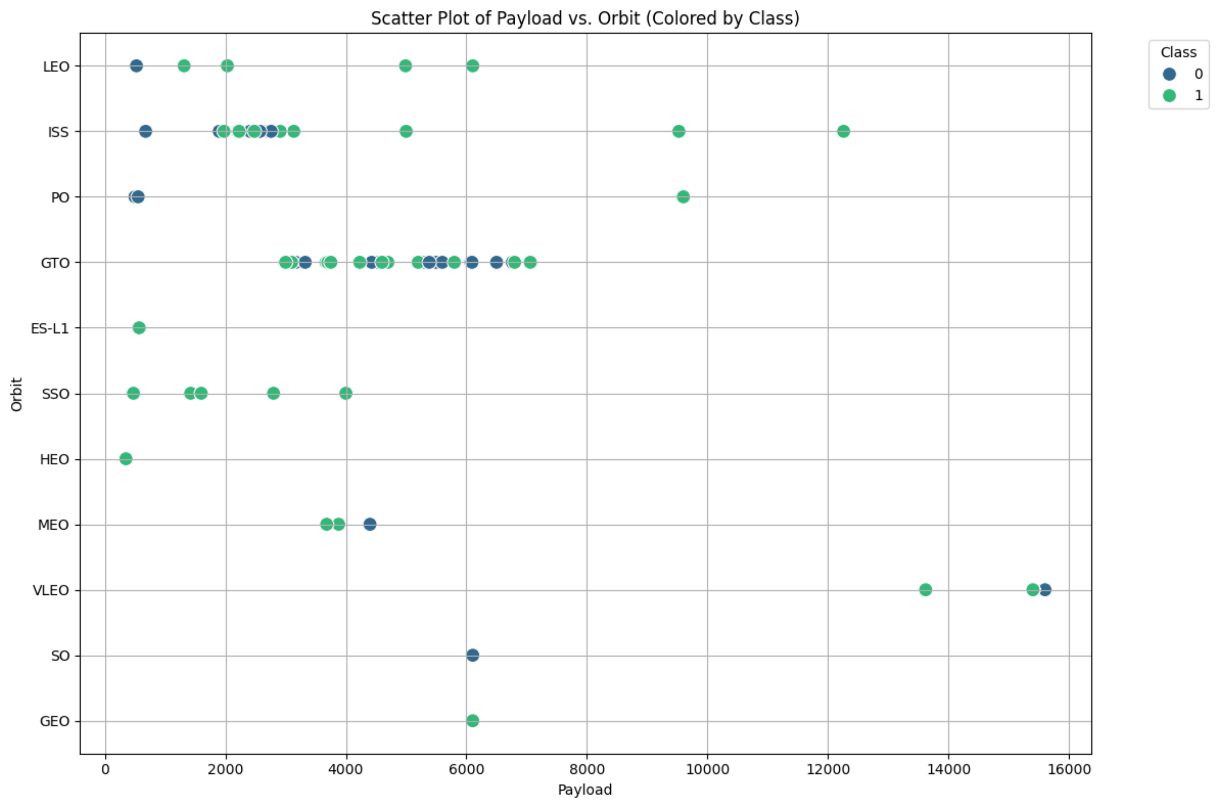
- This scatterplot shows the relationship between the success (0=unsuccessful) rate and the type of orbit the flight took.

Flight Number vs. Orbit Type



- This scatterplot shows the relationship between the flight number and type of orbit.
- In earlier flights only certain types of orbits were attempted with a mixed bag for whether the flight was successful

Payload vs. Orbit Type



- This scatterplot shows the relationship between payload and type of orbit.
- The VLEO orbit is used with heavy payloads and SSO has all successful missions but on the lighter side of the payload weight

All Launch Site Names

- %%sql
- SELECT DISTINCT launch_site
- FROM SPACEXTABLE;

Launch_Site
CCAFS LC-40
VAFB SLC-4E
KSC LC-39A
CCAFS SLC-40

Launch Site Names Begin with 'CCA'

- %%sql
- SELECT *
- FROM SPACEXTABLE
- WHERE launch_site LIKE 'CCA%'
- LIMIT 5;

Date	Time (UTC)	Booster_ Version	Launch_ Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_ Outcome	Landing_ Outcome
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-12-08	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	Failure (parachute)
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

- %%sql
- SELECT SUM(payload_mass__kg_)
- FROM SPACEXTABLE
- WHERE customer = 'NASA (CRS)';

SUM(payload_mass__kg_)
45596

Average Payload Mass by F9 v1.1

- %%sql
- SELECT AVG(payload_mass__kg_)
- FROM SPACEXTABLE
- WHERE Booster_Version = 'F9 v1.1';

AVG(payload_mass__kg_)
2928.4

First Successful Ground Landing Date

- %%sql
- SELECT MIN(Date)
- FROM SPACEXTABLE
- WHERE Mission_Outcome = 'Success';

MIN(Date)

2010-06-04

Successful Drone Ship Landing with Payload between 4000 and 6000

- %%sql

- SELECT DISTINCT Booster_Version
- FROM SPACEXTABLE
- WHERE Landing__Outcome = 'Success (drone ship)'
- AND Payload_Mass__kg_ > 4000
- AND Payload_Mass__kg_ < 6000;

Total Number of Successful and Failure Mission Outcomes

- %%sql
- SELECT Mission_Outcome, COUNT(*) as Total_Count
- FROM SPACEXTABLE
- GROUP BY Mission_Outcome;

Mission_Outcome	Total_Count
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- %%sql
- SELECT DISTINCT Booster_Version
- FROM SPACEXTABLE
- WHERE Payload_Mass__kg_ = (
-     SELECT MAX(Payload_Mass__kg_)
-     FROM SPACEXTABLE

2015 Launch Records

- %%sql
- SELECT
-     substr(Date, 6, 2) AS Month,
-     Landing__Outcome,
-     Booster_Version,
-     Launch_Site
- FROM SPACEXTABLE
- WHERE substr(Date, 0, 5) = '2015'
- AND Landing__Outcome = 'Failure (drone ship)'
- ORDER BY Month;

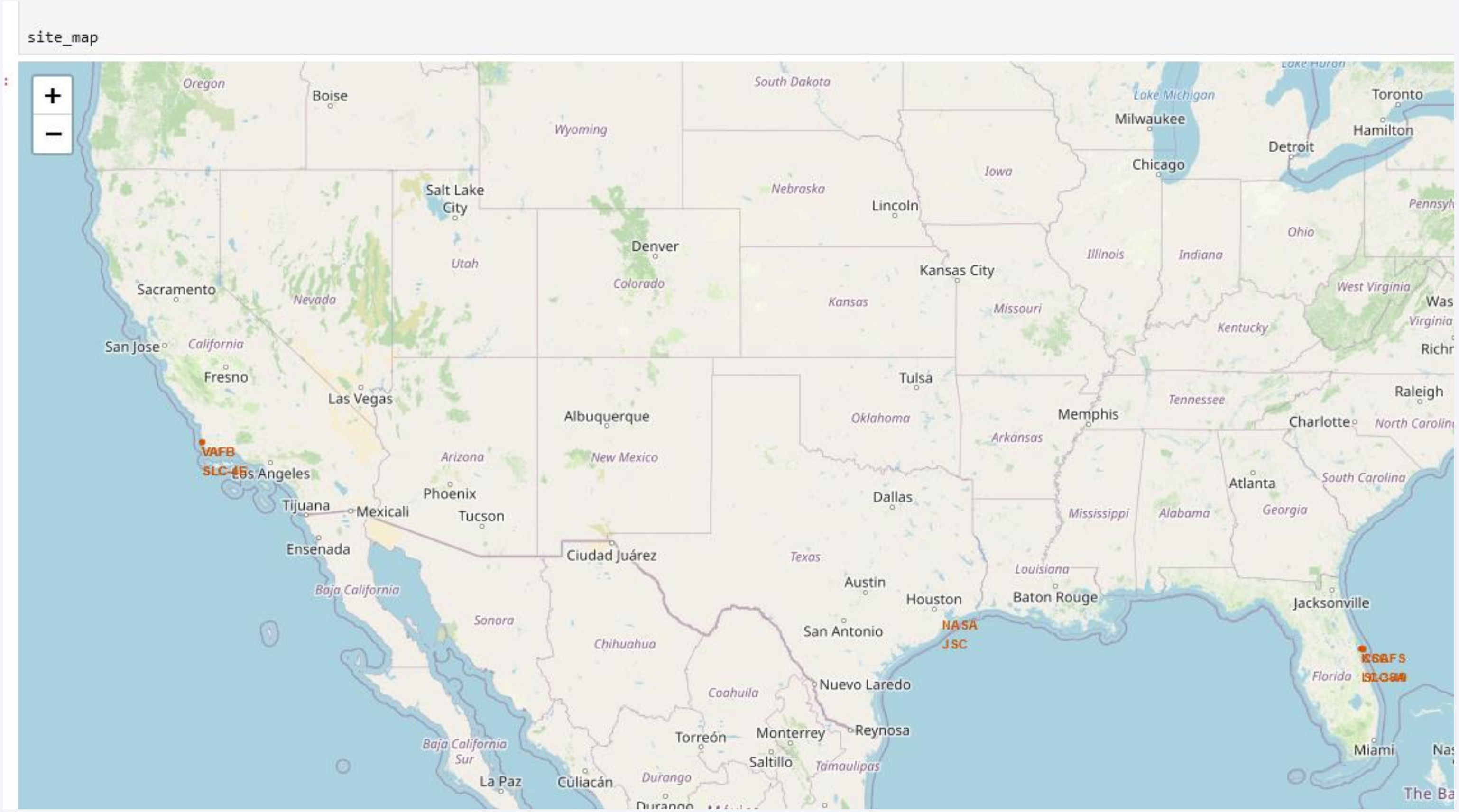
Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

- %%sql
- SELECT
-     Landing__Outcome,
-     COUNT(*) as Count_of_Outcomes
- FROM SPACEXTABLE
- WHERE Date BETWEEN '2010-06-04' AND '2017-03-20'
- GROUP BY Landing__Outcome
- ORDER BY Count_of_Outcomes DESC;

Section 3

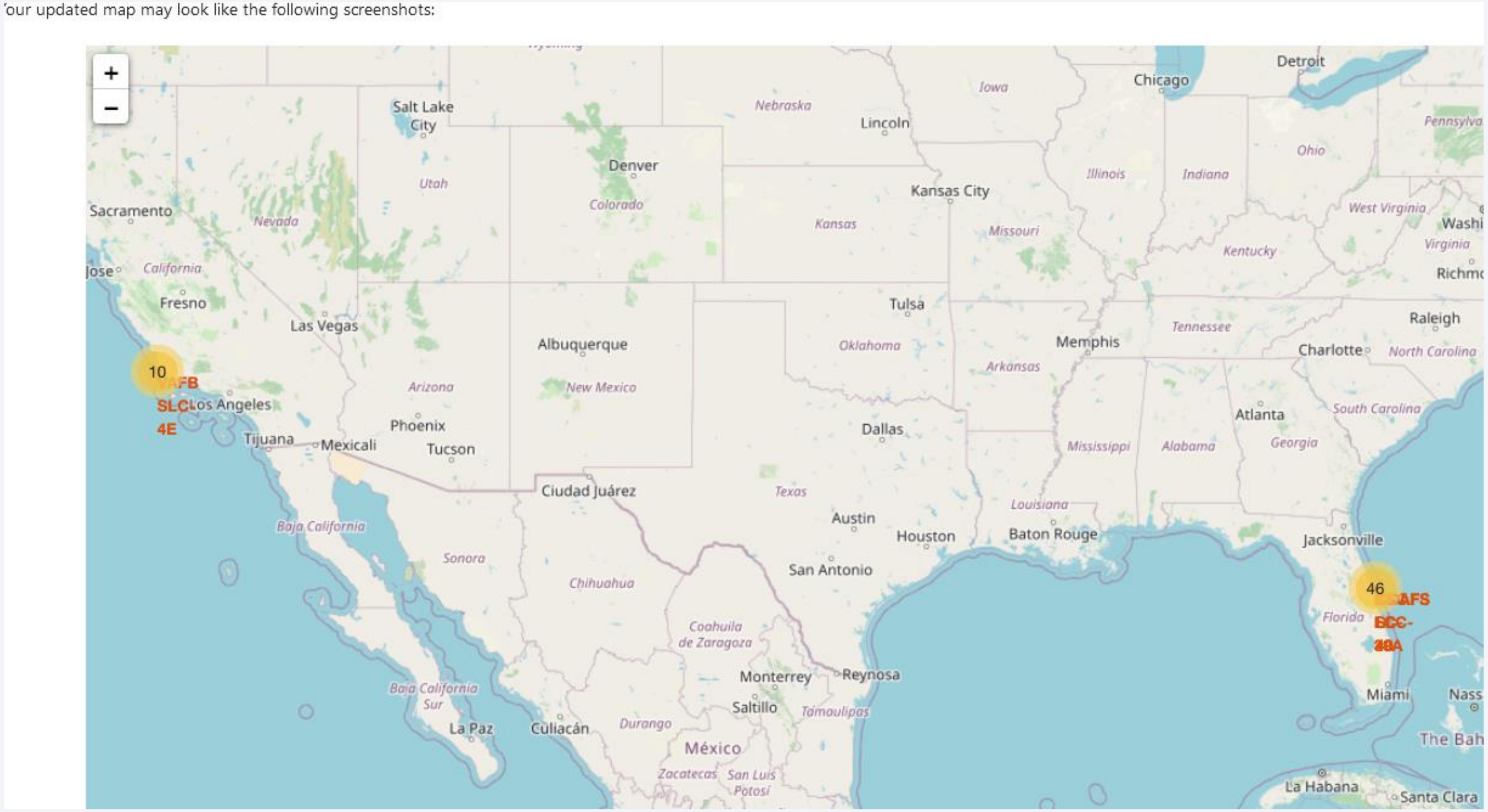
Launch Sites Proximities Analysis

Folium Map



- Launch sites occur in 2 main places, Southern California and The Atlantic side of Florida

Folium Map Successful Missions

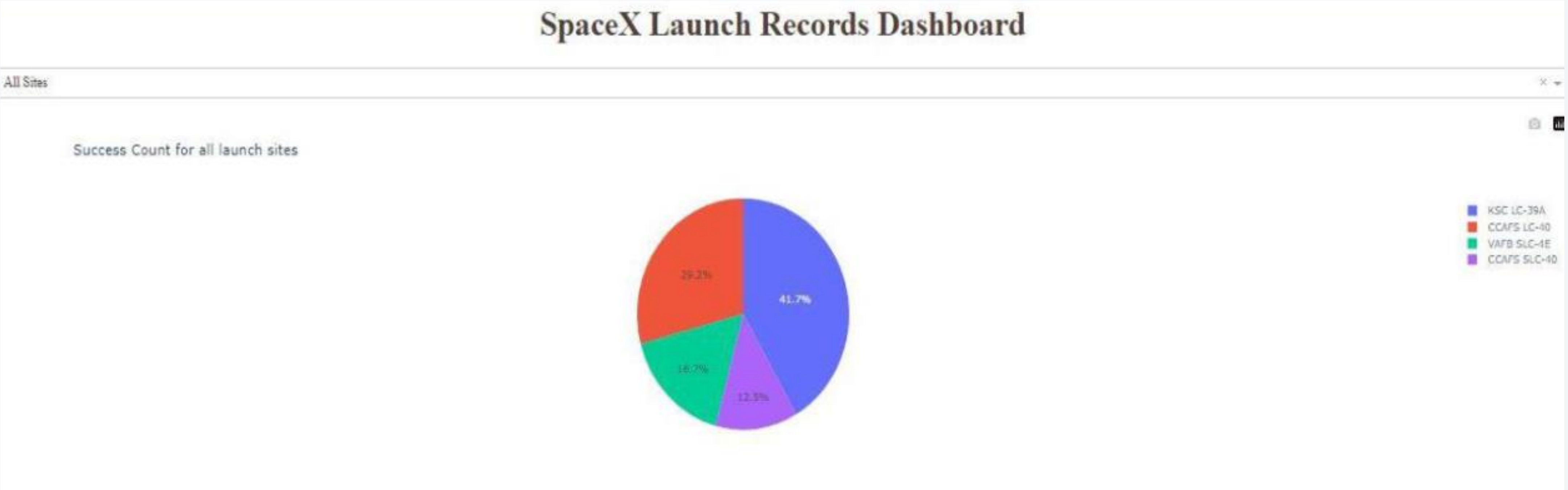


- The majority of successful launches occurred in the Florida locations

Section 4

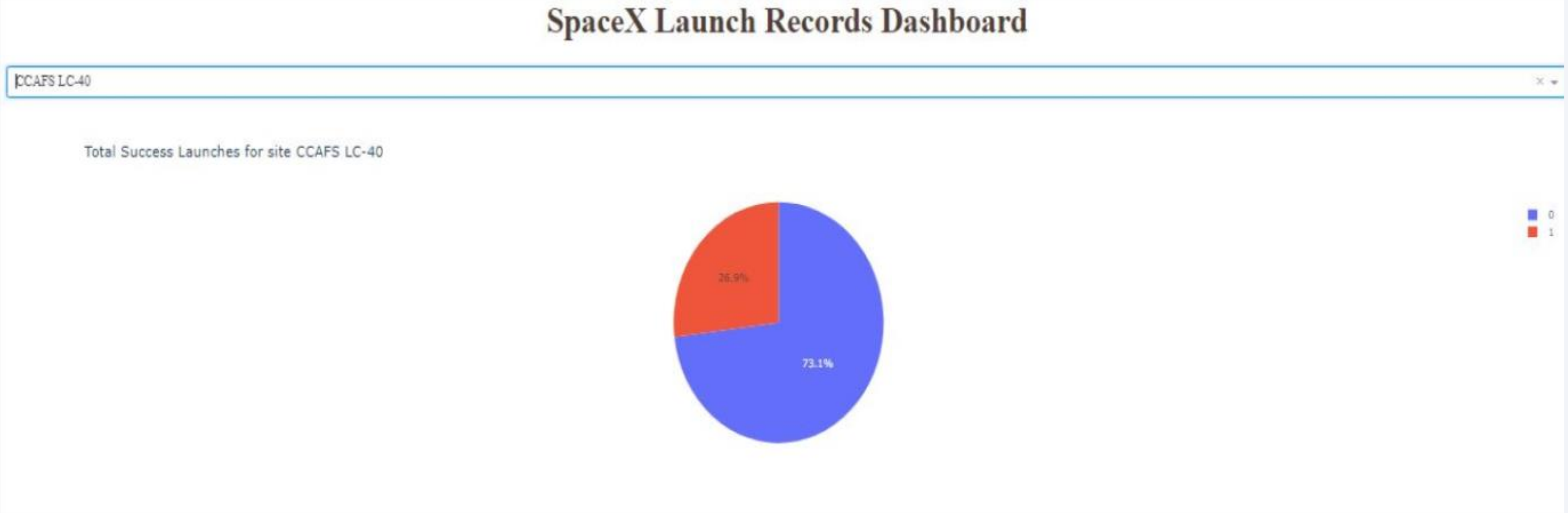
Build a Dashboard with Plotly Dash

Dashboard Screenshot 1



- Site KSC LC-39A has the highest launch success rate, 42%,
- CCAFS LC-40 has 29%
- VAFB SLC-4E has 17%
- CCAFS SLC-40 has13%

Dashboard Screenshot 2



- Site CCAFS LC-40 had a success ratio of 73%

<Dashboard Screenshot 3>

- Replace <Dashboard screenshot 3> title with an appropriate title

- Show screenshots of Payload vs. Launch Outcome scatter plot for all sites, with different payload selected in the range slider

- Explain the important elements and findings on the screenshot, such as which payload range or booster version have the largest success rate, etc.

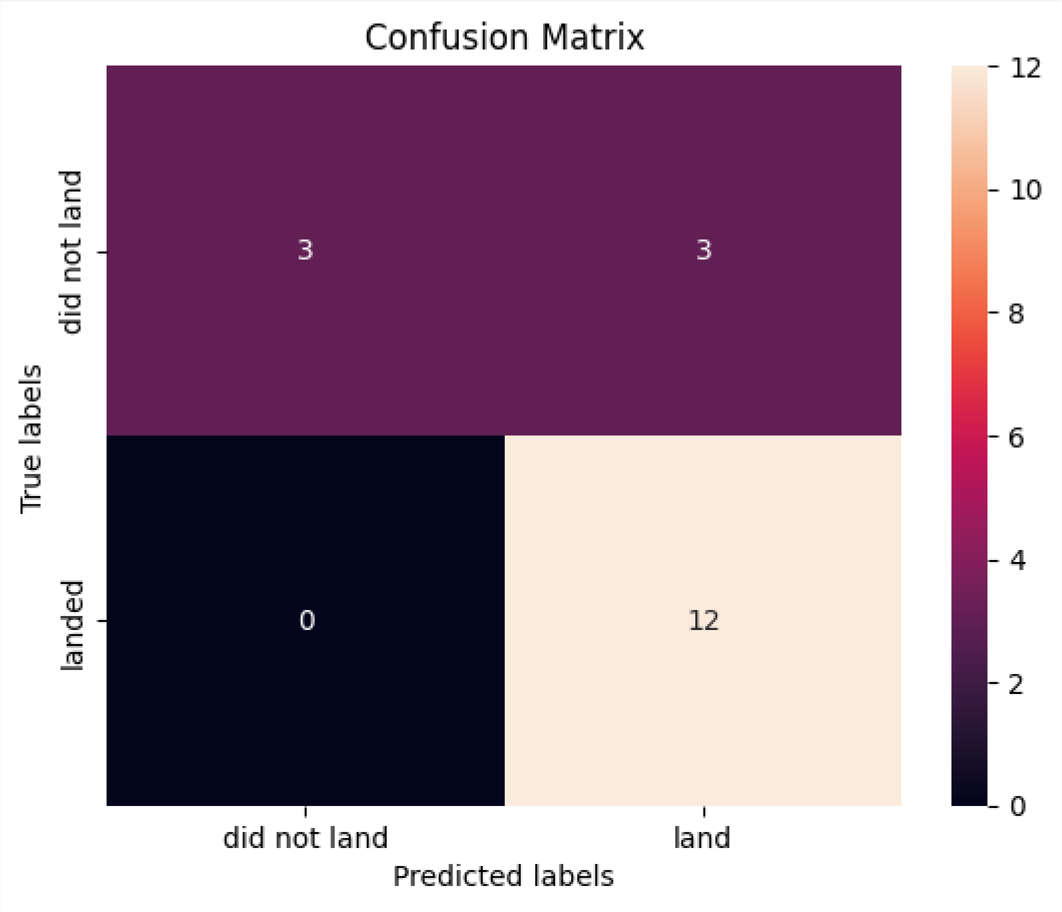
Section 5

Predictive Analysis (Classification)

Classification Accuracy

Best Score for Decision Tree Classifier: 0.8732142857142857 Best Score for Support Vector Machine: 0.8482142857142856 Best Score for KNN: 0.8482142857142858 Best Score for Log Reg: 0.8464285714285713

Confusion Matrix



Conclusions

- The Decision Tree produced the highest accuracy in predicting whether SpaceX will reuse the 1st stage of the rocket.
- Success rate overall increased from 2013 to 2020
- Success was mixed between launch sites and orbits, results were inconclusive here

Thank you!