
# SpaceX Rocket Recapture Modeling

Jamahl Stokes

2/5/2023

# OUTLINE

- Executive Summary
- Introduction
- Methodology
- Results
    - Visualization – Charts
    - Dashboard
- Discussion
    - Findings & Implications
- Conclusion
- Appendix

# EXECUTIVE SUMMARY

- We try to find features that predict the success, or failure, in landing the 1st stage of a Falcon 9 rocket.
- Captured relevant data for analysis.
  - Data collected from SpaceX API and Wikipedia.
  - Used SQL and Visualization to explore data.
  - Plotted geographic data to garner additional insight.
- Built dashboard to explore data granularly.
- Developed predictive models that accurately predict SpaceX Falcon 9 rocket landing failures, and (with less accuracy) successes.

# INTRODUCTION

- We will use past data to predict whether the Falcon 9 rocket will land successfully.
- The Falcon 9 rocket has a stage that can be reused, upon successful landing.
- Reusing rocket parts allows for significant savings, a competitive advantage in the aeronautics industry.
- What factors are important to the successful recapture of stage 1 of a rocket?
  - Can these factors be used to successfully predict the success of rocket launches?

# METHODOLOGY (Data Wrangling)

1. Import Libraries
2. Define Auxilliary Functions
3. Collect Data
4. Format Data

- Retrieved and parsed data from SpaceX API.
  - Imported libraries, defined helper functions, retrieved json response, parsed json into dataframe, dropped undesirable data, and imputed missing values.
- Scraped and parsed additional launch data from Wikipedia.
  - Imported libraries, defined helper functions, retrieved html data, parsed html data into dataframe columns.
- Wrangled data and uploaded files here.
  - Imported libraries, defined auxillary functions, calculated launches by site/orbit, and developed training labels.

Using the `Outcome`, create a list where the element is zero if the corresponding row in `Outcome` is in the set `bad_outcome`; otherwise, it's one. Then assign it to the variable `landing_class`:

In [19]:
```
# landing_class = 0 if bad_outcome
# landing_class = 1 otherwise
landing_class = []
for var in df['Outcome']:
    if var in bad_outcomes:
        landing_class.append(0)
    else:
        landing_class.append(1)
len(landing_class)
```

# METHODOLOGY (EDA w/ SQL & Visuals)

- Developed SQL queries to explore data further.
- Used feature engineering to convert categorical features to numeric values & uploaded files here.
- Displayed geographic data to identify insights not easily captured from tabular analyses.

Display the names of the unique launch sites in the space mission

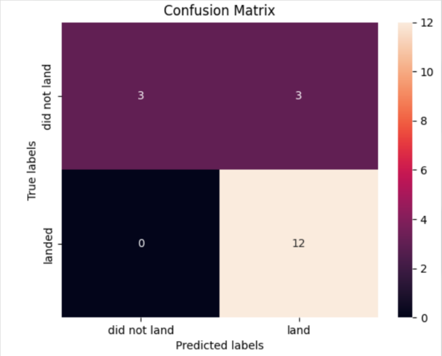
```
%sql select distinct "Launch_Site" from SPACEXTBL
```

```
 * sqlite:///my_data1.db
Done.
```

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

# METHODOLOGY (Predictive Analysis)

- Standardized numeric values for consistency.
- Transformed data into format for model input.
- Split data into training and test sets.
- Used GridSearchCV to identify parameters that improve model performance.
  - Compared logistic regression, support vector machine, decision tree classifier, and k-nearest neighbors.
  - Calculated accuracy scores (R2) and plotted confusion matrices.




1. Gather Model Input Data
2. Transform Data as Needed
3. Split Data into Test/Training Sets
4. Search for Optimal Parameters

# RESULTS

# EDA WITH VISUALIZATION - FINDINGS & IMPLICATIONS

## Findings

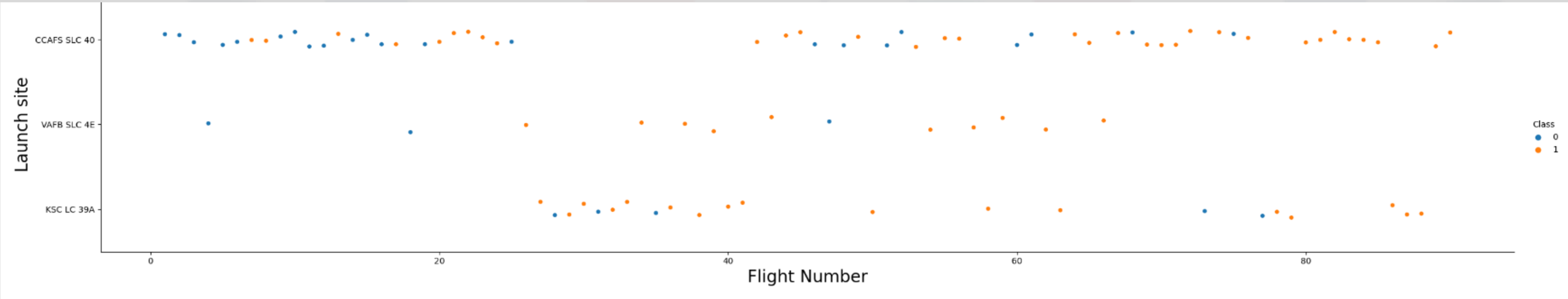
- Success varied with orbit
- Later launches showed greater success
- Launch sites varied in the payloads of rockets launched

## Implications

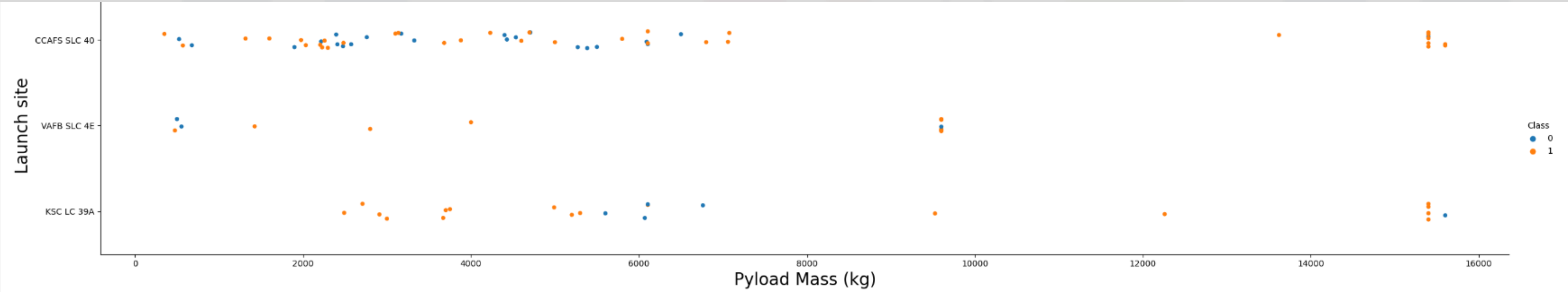
- Some orbits inherently difficult
- Over time, effective improvements were made
- Some sites are better suited for launches of large rockets

# EDA VISUALIZATIONS

## Flight Number and Launch Site Scatter

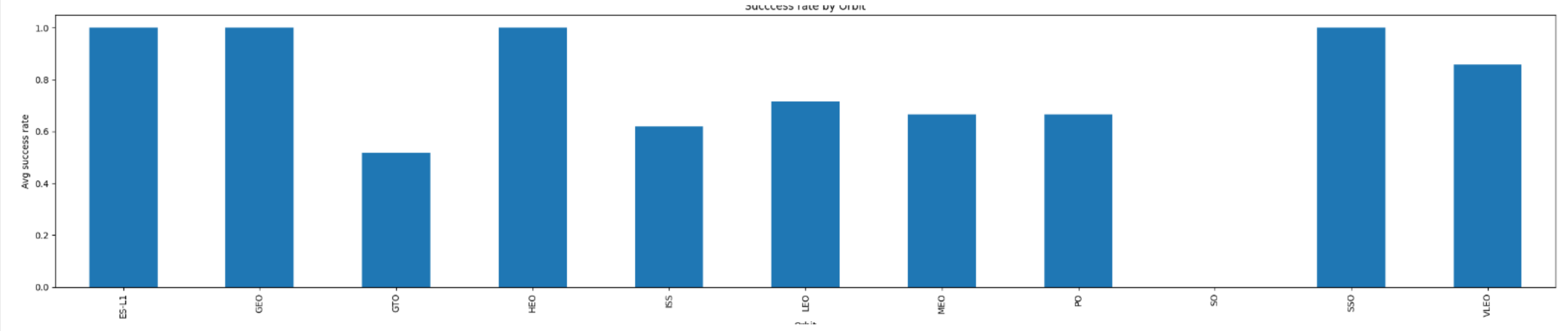



## Payload and Launch Site Scatter

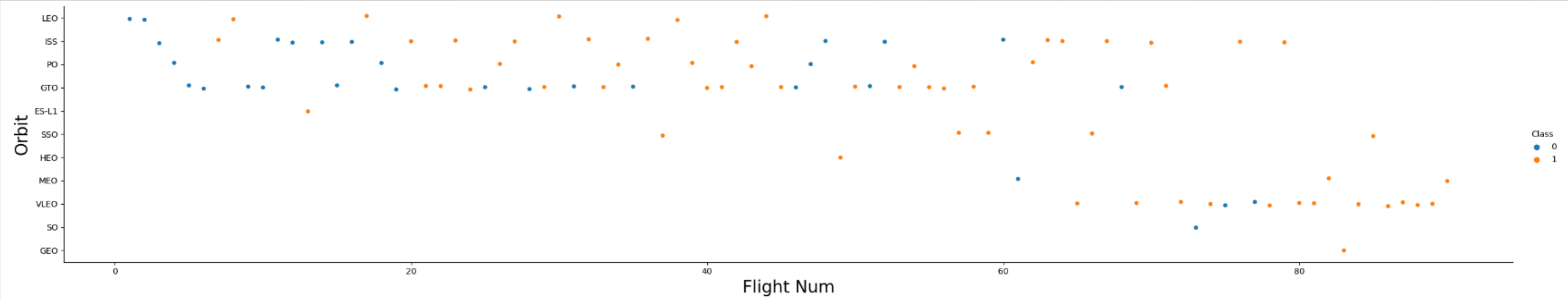


# EDA VISUALIZATIONS

## Success rate vs Orbit bar chart

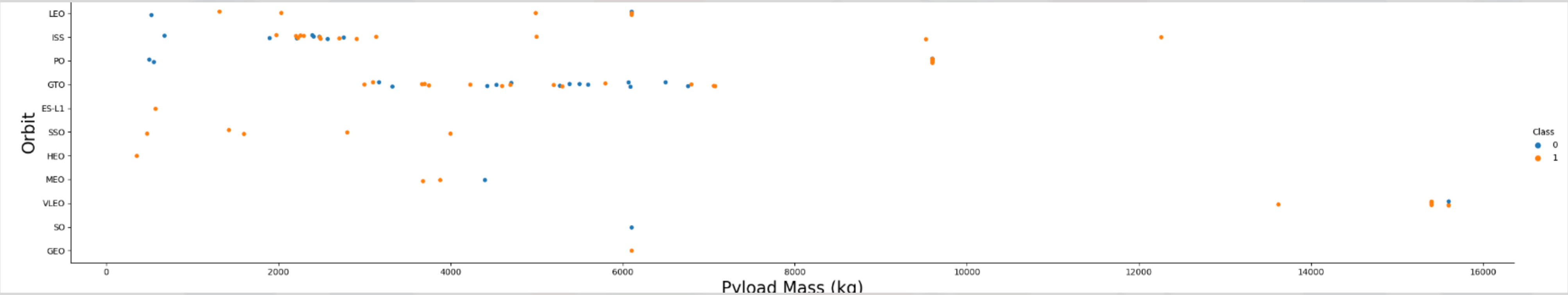



## Flight Number vs Orbit Scatter

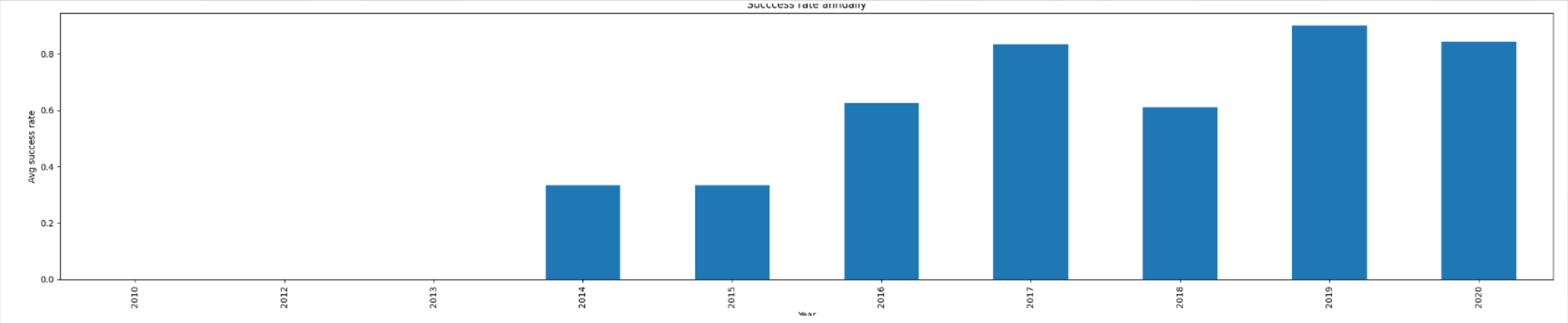


# EDA VISUALIZATIONS

## Payload vs Orbit scatter




## Launch Success Yearly Trend



# EDA WITH SQL - FINDINGS & IMPLICATIONS

## Findings

- 4 launch sites, though some are similarly named
- Calculated average mass of payload carried by booster
- Calculated number of success and failure outcomes

## Implications

- Different sites used for different launch types
- Payload may be correlated with other features (e.g. orbit)
- Variations in data can illuminate impactful features

# SQL EXPLORATORY DATA ANALYSIS

## Find distinct launch sites

Display the names of the unique launch sites in the space mission

```
%sql select distinct "Launch_Site" from SPACEXTBL
```

```
 * sqlite:///my_data1.db
Done.
```

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

## 5 sites that begin with 'CCA'

Display 5 records where launch sites begin with the string 'CCA'

```
%sql select * from SPACEXTBL where "Launch_Site" like "CCA%" limit 5
```

```
 * sqlite:///my_data1.db
Done.
```

| Date | Time (UTC) | Booster_Version | Launch_Site | Payload | PAYLOAD_MASS__KG_ | Orbit | Customer | Mission_Outcome | Landing _Outcome |
|---|---|---|---|---|---|---|---|---|---|
| 04-06-2010 | 18:45:00 | F9 v1.0 B0003 | CCAFS LC-40 | Dragon Spacecraft Qualification Unit | 0 | LEO | SpaceX | Success | Failure (parachute) |
| 08-12-2010 | 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) |
| 22-05-2012 | 07:44:00 | F9 v1.0 B0005 | CCAFS LC-40 | Dragon demo flight C2 | 525 | LEO (ISS) | NASA (COTS) | Success | No attempt |
| 08-10-2012 | 00:35:00 | F9 v1.0 B0006 | CCAFS LC-40 | SpaceX CRS-1 | 500 | LEO (ISS) | NASA (CRS) | Success | No attempt |
| 01-03-2013 | 15:10:00 | F9 v1.0 B0007 | CCAFS LC-40 | SpaceX CRS-2 | 677 | LEO (ISS) | NASA (CRS) | Success | No attempt |

# SQL EXPLORATORY DATA ANALYSIS

## Total Payload Mass

Display the total payload mass carried by boosters launched by NASA (CRS)

```
%sql select sum("PAYLOAD_MASS__KG_") from SPACEXTBL where "Customer" = "NASA (CRS)"
```

```
 * sqlite:///my_data1.db
Done.
```

| sum("PAYLOAD_MASS__KG_") |
|---|
| 45596 |

## Average Payload Mass by F9 v1.1

Display average payload mass carried by booster version F9 v1.1

```
%sql select sum("PAYLOAD_MASS__KG_")/count("PAYLOAD_MASS__KG_") as avgMassf9 from SPACEXTBL where "Booster_Version" = "F9 v1.1"
```

```
 * sqlite:///my_data1.db
Done.
```

| avgMassf9 |
|---|
| 2928 |

# SQL EXPLORATORY DATA ANALYSIS

## First Successful Ground Pad Landing

### Task 5

List the date when the first succesful landing outcome in ground pad was acheived.

*Hint:Use min function*

```
%sql select min("Date") from SPACEXTBL where "Mission_Outcome" = "Success" and lower("Landing _Outcome") like "%ground pad%"
```

```
 * sqlite:///my_data1.db
Done.
```

| min("Date") |
|---|
| 01-05-2017 |

## Boosters Landed on Drone Ships with Payload From 4-6k

List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

```
%sql select distinct "Booster_Version" from SPACEXTBL where "Mission_Outcome" = "Success" and lower("Landing _Outcome") like "%drone ship%" and "PAYLOAD_MASS__KG_" between 4000 and 6000
```

```
 * sqlite:///my_data1.db
Done.
```

| Booster_Version |
|---|
| F9 FT B1020 |
| F9 FT B1022 |
| F9 FT B1026 |
| F9 FT B1021.2 |
| F9 FT B1031.2 |

# SQL EXPLORATORY DATA ANALYSIS

## Total Number of Success/Failure Outcomes

List the total number of successful and failure mission outcomes

```
%sql select "Mission_Outcome", count(*) from SPACEXTBL group by "Mission_Outcome"
```

```
 * sqlite:///my_data1.db
Done.
```

| Mission_Outcome | count(*) |
|---|---|
| Failure (in flight) | 1 |
| Success | 98 |
| Success | 1 |
| Success (payload status unclear) | 1 |

## Which Boosters Carried Maximum Payloads

List the names of the booster_versions which have carried the maximum payload mass. Use a subquery

```
%sql select "Booster_Version" from SPACEXTBL where "PAYLOAD_MASS__KG_" = (select max("PAYLOAD_MASS__KG_") from SPACEXTBL)
```

```
 * sqlite:///my_data1.db
Done.
```

| Booster_Version |
|---|
| F9 B5 B1048.4 |
| F9 B5 B1049.4 |
| F9 B5 B1051.3 |
| F9 B5 B1056.4 |
| F9 B5 B1048.5 |
| F9 B5 B1051.4 |
| F9 B5 B1049.5 |
| F9 B5 B1060.2 |
| F9 B5 B1058.3 |
| F9 B5 B1051.6 |
| F9 B5 B1060.3 |
| F9 B5 B1049.7 |

# SQL EXPLORATORY DATA ANALYSIS

## 2015 Launch Records

List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.

**Note: SQLLite does not support monthnames. So you need to use substr(Date, 4, 2) as month to get the months and substr(Date,7,4)='2015' for year.**

```
%sql select substr("Date",4,2) as month, "Landing _Outcome", "Booster_Version", "Launch_Site" from SPACEXTBL where substr("Date",7,4)="2015" and lower("Landing _Outcome") like "%Failure%drone
```

```
 * sqlite:///my_data1.db
Done.
```

| month | Landing _Outcome | Booster_Version | Launch_Site |
|---|---|---|---|
| 01 | Failure (drone ship) | F9 v1.1 B1012 | CCAFS LC-40 |
| 04 | Failure (drone ship) | F9 v1.1 B1015 | CCAFS LC-40 |

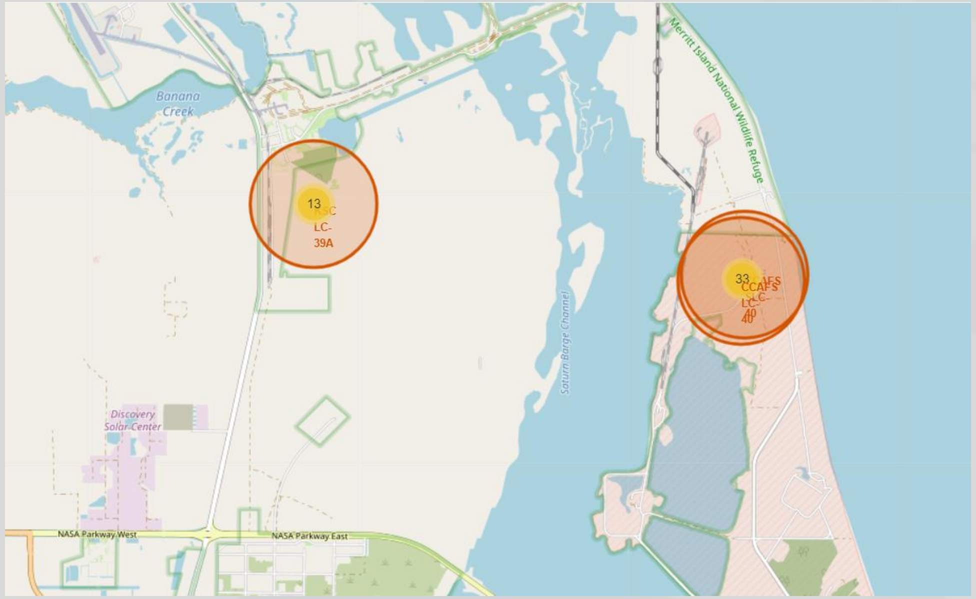
## Rank Success Count Between Dates

Rank the count of successful landing_outcomes between the date 04-06-2010 and 20-03-2017 in descending order.

```
%sql select substr("Date",4,2) as month, "Landing _Outcome", "Booster_Version", "Launch_Site" from SPACEXTBL where "Date" > '04-06-2010' and "Date" < '20-03-2017'
```

# FOLIUM RESULTS

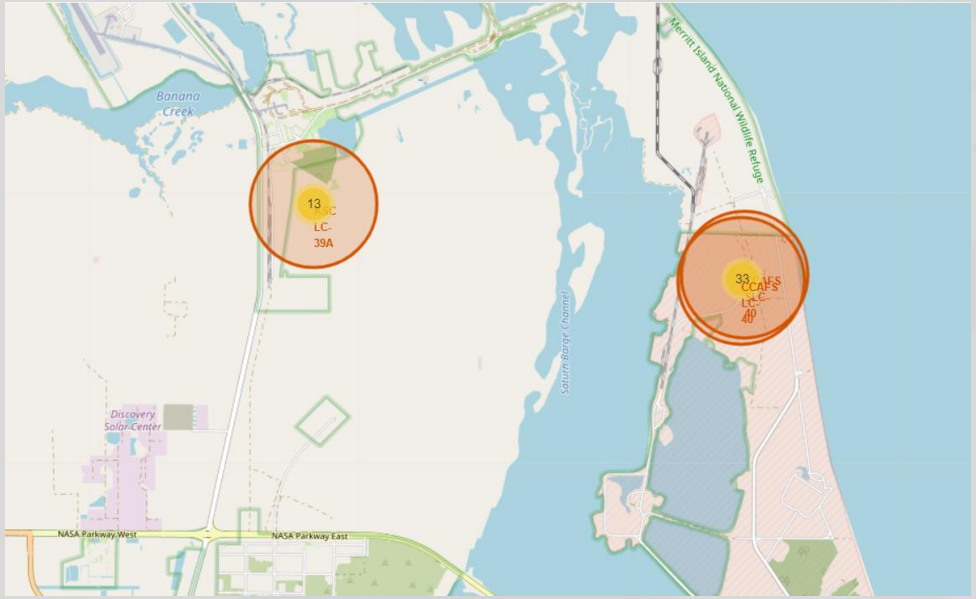
All launch records are associated with a site




Markers show multiple launches per each site

# FOLIUM RESULTS

Success rates vary by launch site



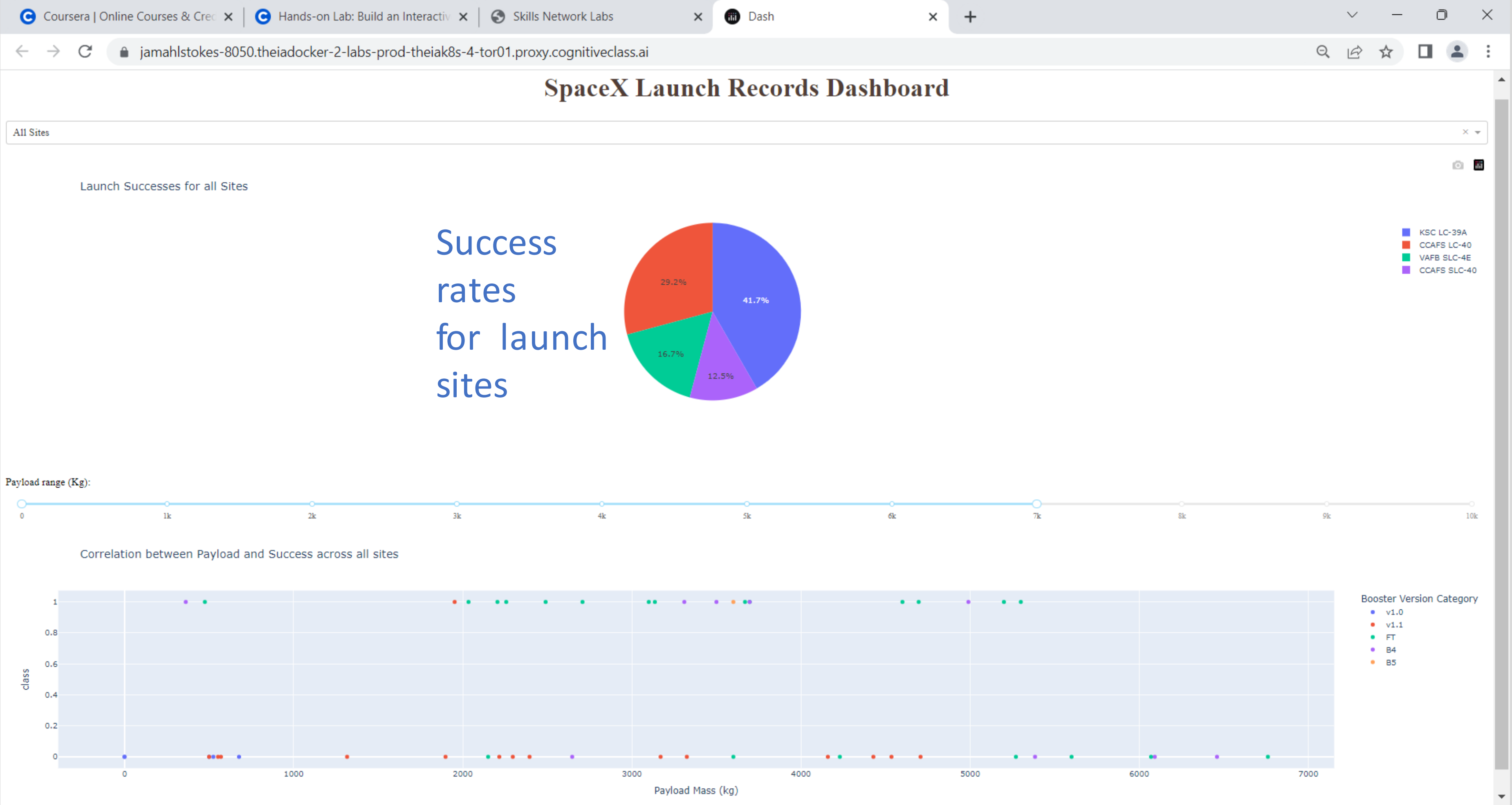

Launch sites close to coasts, highways, and rail

# DASHBOARD

The dashboard uploaded here was developed to examine the success rates across launch sites and to explore how launch sites varied in the payload mass that rockets caried.

Results showed that payload mass and launch site were correlated with launch and re-landing success.

# DASHBOARD – Dropdown & Slider




Success rates for launch sites

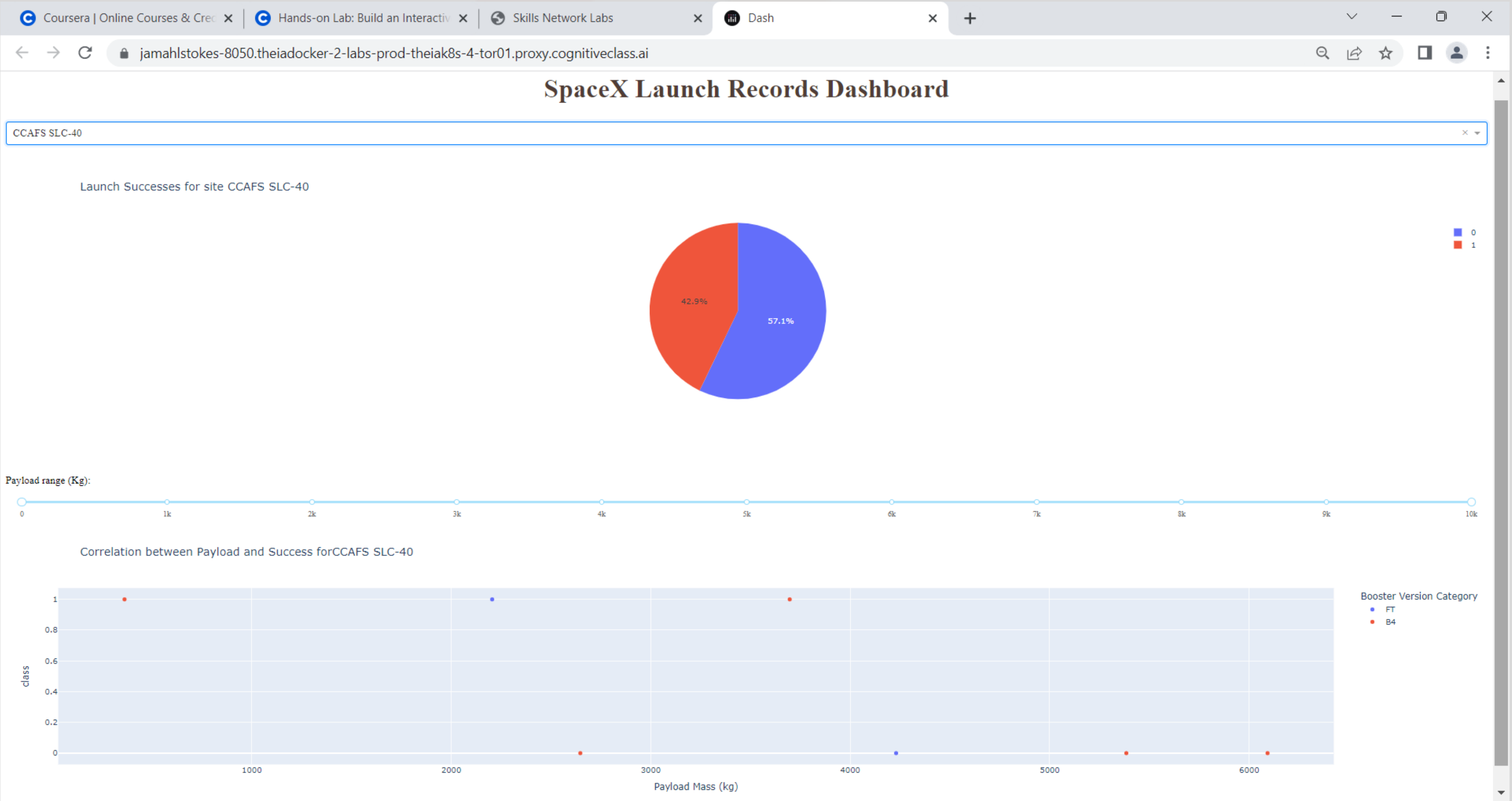
Interactive Slider that updates Scatter Plot

IBM Developer

SKILLS NETWORK

# DASHBOARD

Dashboard filters to specific site upon selection



# PREDICTIVE ANALYSIS

- Various classification models were evaluated on training data and had accuracy scores as follows.

| | |
|---|---|
| Logistic Regression | .846 |
| Support Vector Machine | .848 |
| Decision Tree Classification | .888 |
| K-Nearest Neighbors | .848 |

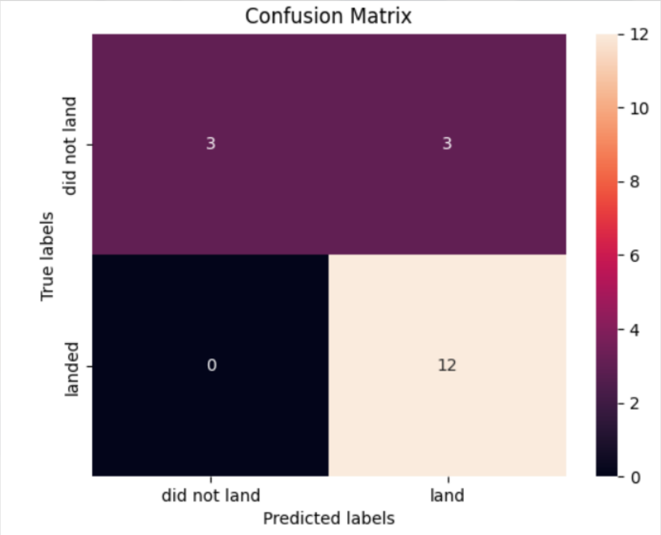
# PREDICTIVE ANALYSIS

- When these models were evaluated on test data their accuracy scored as follows.

| | |
|---|---|
| Logistic Regression | .833 |
| Support Vector Machine | .833 |
| Decision Tree Classification | .666 |
| K-Nearest Neighbors | .833 |

# PREDICTIVE ANALYSIS

- SVM, and KNN show equal performance here, superior to that of decision tree and logistic regression.



# CONCLUSION

Overall Findings

- Geography may influence selection of launch sites
- Sites varied in the profiles of payloads they launched
- Several booster versions carry the maximum payload
- Predictive models accurately classified failures, with lower scores classifying successes

# CONCLUSION

- Re-capture of stage 1 of rockets saves significant costs.
- Factors such as mission orbit, launch site, launch date, and payload impact success rates.
- Predictive models, informed by past launches, can be developed to help forecast future launch successes.

# APPENDIX

# SUPPLEMENTAL FILE LOCATIONS

Supporting files for all analyses presented have been uploaded to my personal Github account <https://github.com/jamahlstokes/IBM-DataScience-Public>.

Enjoy!