Achieving a successful launch through Data Science

Alejandra Peña Flores

2023-01-25

IBM Developer

INTERNAL USE ONLY

SKILLS NETWORK

OUTLINE

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

EXECUTIVE SUMMARY

- Results summary:
  - Public domain data was collected with Python scripts
  - Data preparation completed to create a viable data source
  - Measurements and analysis performed on data source allowed the success/failure rate of launches
  - Machine learning techniques were implemented to determine the principal factors to ensure a successful launch.
- Analysis Methodology:
  - Web Scrapping and SpaceX API for data collection
  - Exploratory Data Analysis (EDA) for data preparation and analytics
  - Machine Learning for result predictions.

INTRODUCTION

- In this project, the process to successfully retrieve the first stage of a rocket launch will be analyzed.

- Determining factors for success will be defined:
    - Booster version
    - Launch Site
    - Orbit
    - Payload mass

- Successful rocket launches cost can be reduced in 62 million dollars per event, by reusing the first stage.

METHODOLOGY – Data Collection and Preparation

- Data Collected from:
  - SpaceX API (http://api.spacexdata.com/v4/rockets()
  - Public records (List of Falcon 9 and Falcon Heavy launches (2010–2019) – Wikipedia)

- Steps
  - Extraction of SpaceX Launch data
  - Data filtering for Falcon 9 launches
  - Data preparation in case of missing values
  - Extraction of column/variable names from HTML table
  - Creation of data frame

- Complete data prep procedure in appendix.

METHODOLOGY – Data Analysis

- Exploratory Data Analysis performed on dataset
- Results of:
  - Launches per site
  - Occurrences of each orbit
  - Occurrences of mission outcomes per orbit
  - Landing outcome label created

METHODOLOGY – Data Visualization

- Data exploration and graphics created to visualize features relationship. Full list of queries in Appendix.
  - Payload Mass x Flight Number
  - Launch Site x Flight Number
  - Launch Site x Payload Mass
  - Orbit vs Flight Number
  - Payload vs. Orbit

METHODOLOGY – Data Visualization

- Folium Interactive Map created
  - Markers indicate points like launch sites
  - Circles indicate highlighted areas around specific coordinates, like NASA Johnson Space Center
  - Marker clusters indicate groups of events in each coordinate, like launches in a launch site
  - Lines indicate distances between two coordinates

METHODOLOGY – Data Visualization

- Plotly Dashboard
  - Graphs and plots created:
    - Percentage of launches by site
    - Payload range
  - Results indicate relationship between factors like payloads and launch sites. This allows identification of **Best place to launch according to payloads**.

METHODOLOGY – Data Prediction

- Classfication models compared:
  - Logistic regression
  - Support vector machine
  - Decision tree
  - K Nearest neighbors

RESULTS

There are 4 main launch sites for Space X.

Average payload for a F9 booster is 2,928 kg.

The first success landing outcome happened in 2017.

The highest success rate for landing on F9 boosters with payloads above the average is drone ships.

Two booster versions failed at landing in drone ships in 2017.

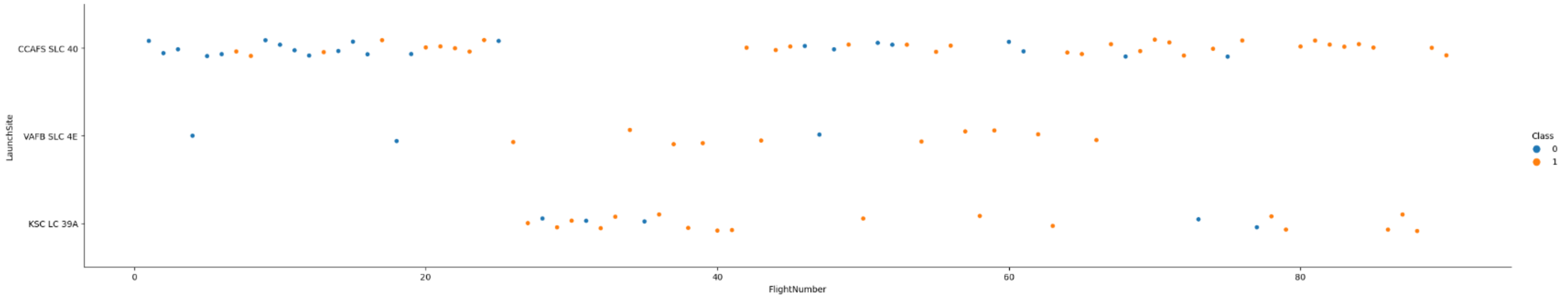
The number of successful landing outcomes improve in time.

RESULTS

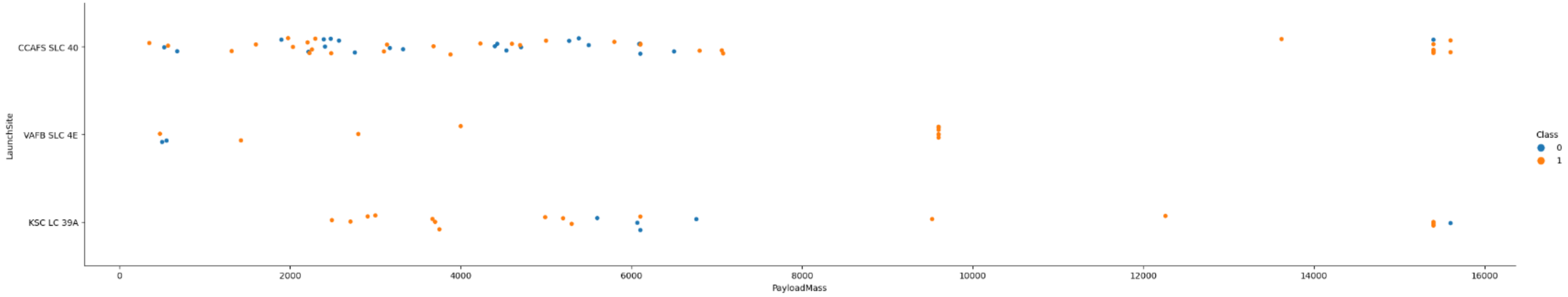
Launch sites analysis. Most launches happens at each cost launch sites.

RESULTS

Launch Site VAFB SLC 4E has the best score of all sites. Although the success rate improves in time

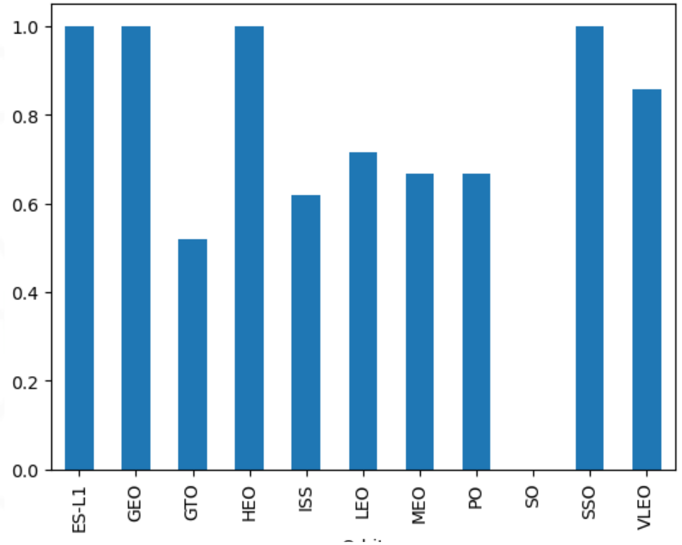


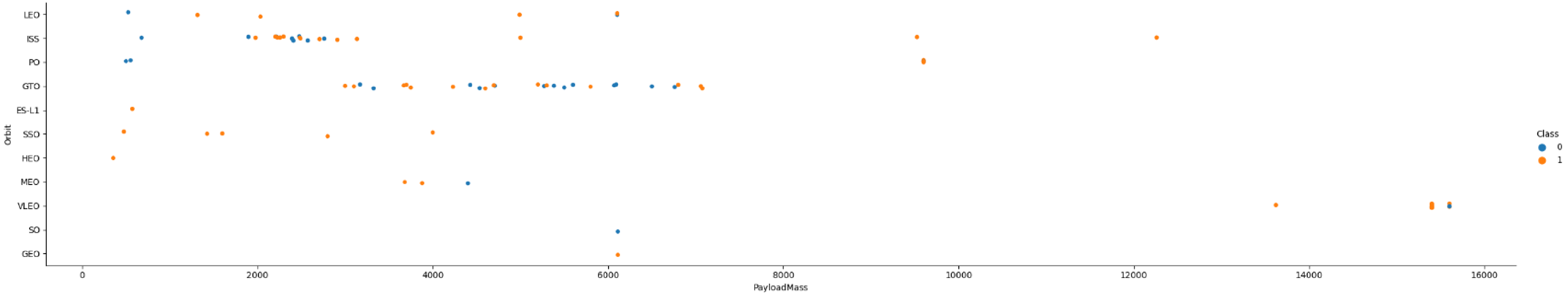
Also the probability of a successful launch is 100% when the Payload mass is over 9000 and under 1400kg.
Payloads over 1200kg are not performed by VAFB SLC 4E launch.



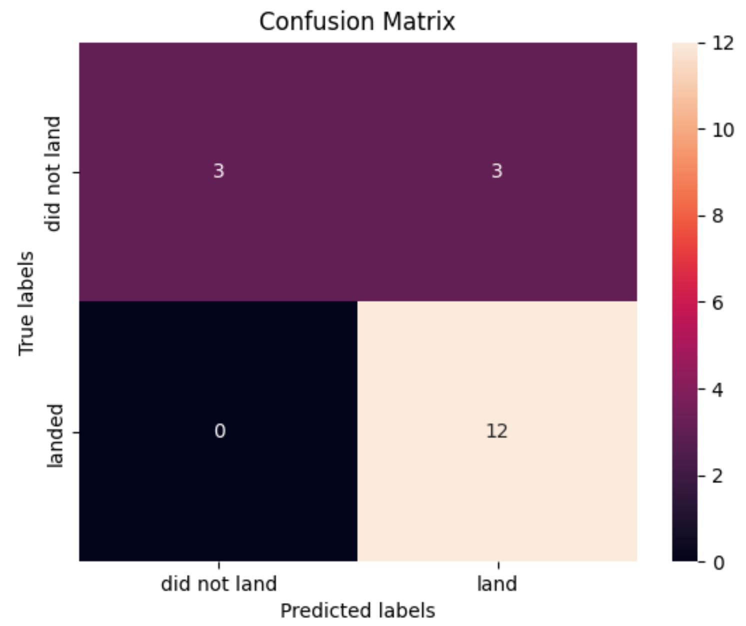
RESULTS

- The largest success rate are presented in orbits:
  - ES-L1
  - GEO
  - HEO
  - SSO
- Success rate improves over time.
- The success rate of the better orbits is higher at payload mass under 4000 kg.
- Orbits HEO and SSO have a low count of launches, so the success rate is not as reliable.



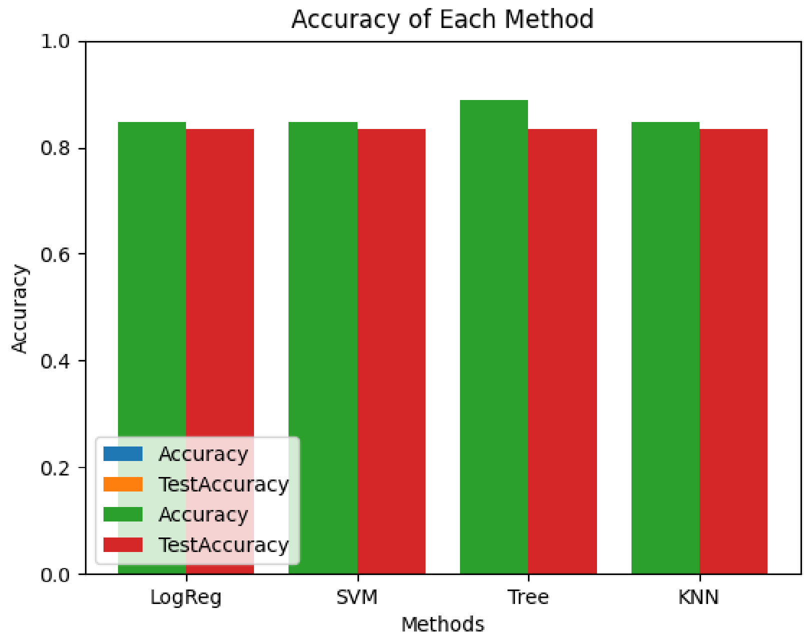


RESULTS



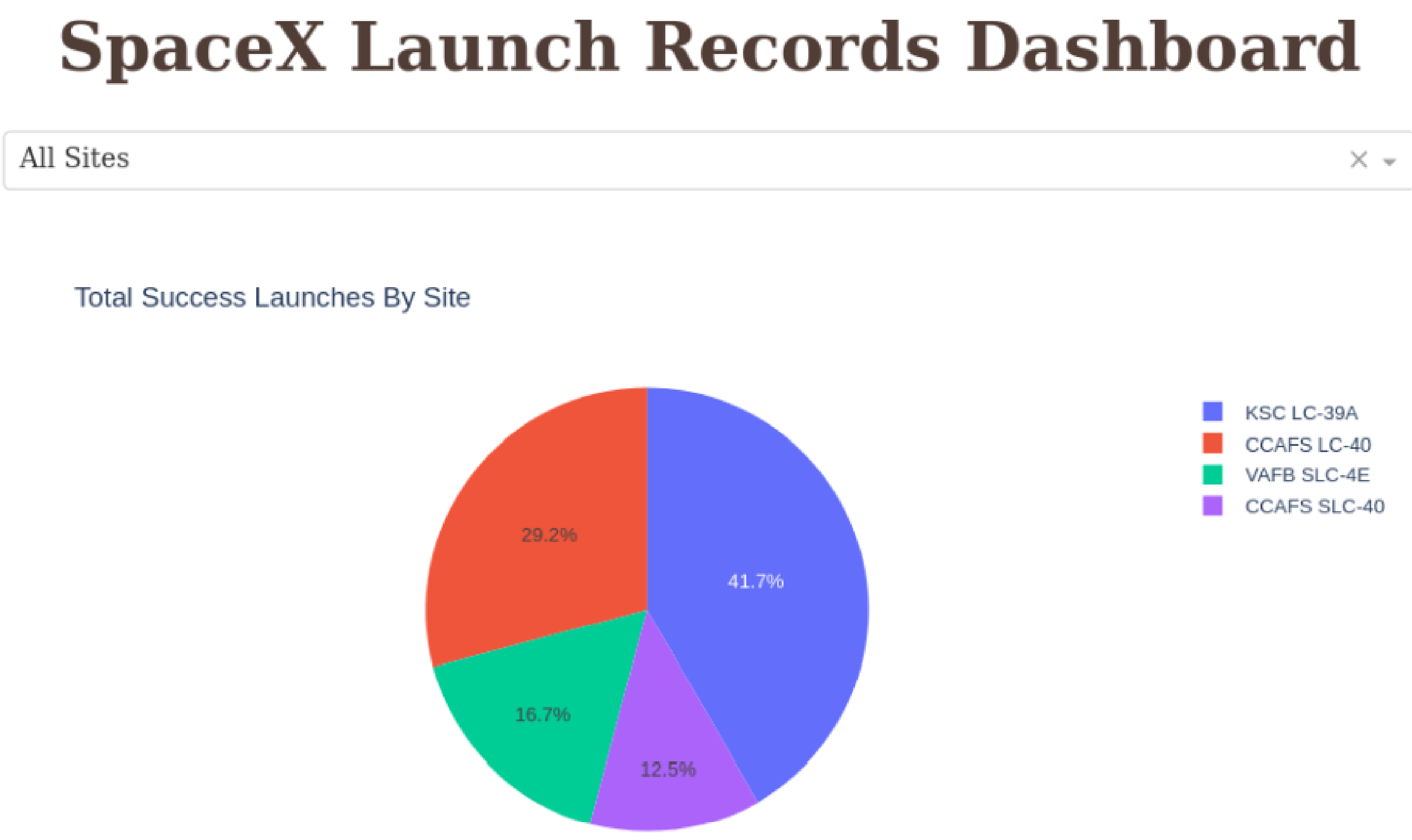
Of all classification models tested, the Decision Tree Classifier is the best to predict successful landings, having an accuracy over 87% in total.

The accuracy is proven in the confusion matrix for the Decision Tree Classifier, by showing the largest numbers of true positives.



PLOTLY DASHBOARD

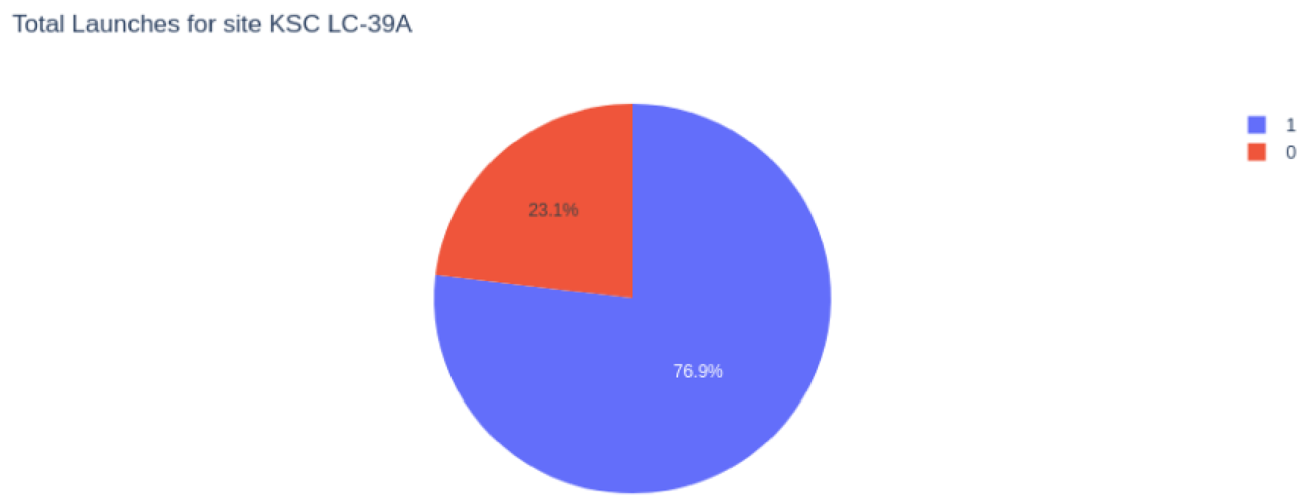
The probability of success shows dependency on the Launch Site



PLOTLY DASHBOARD

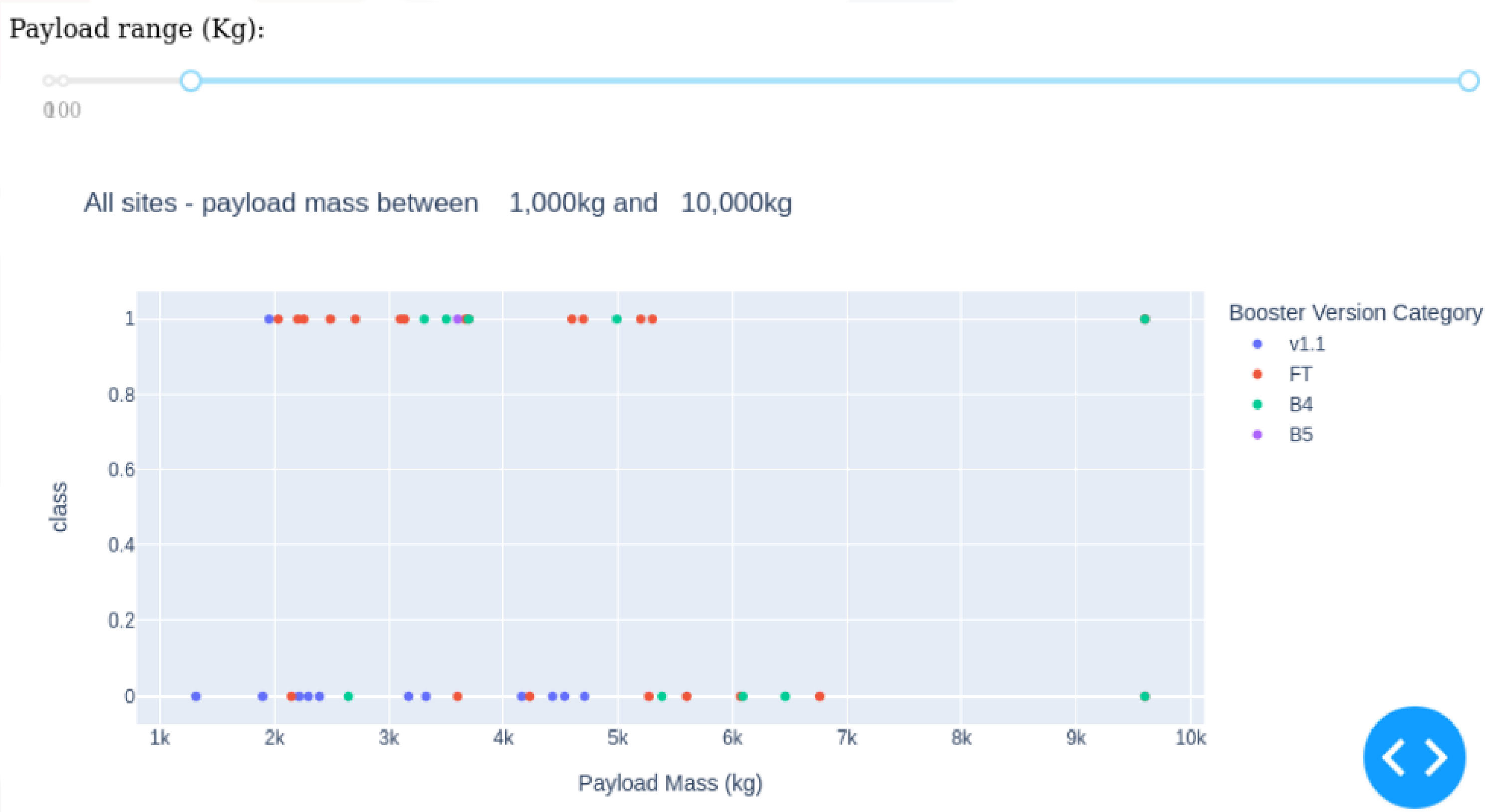
When a specific site is selected, its percentage of success/failure can be observed.

In this case, site KSC LC-39A has a 76.9% rate of success.



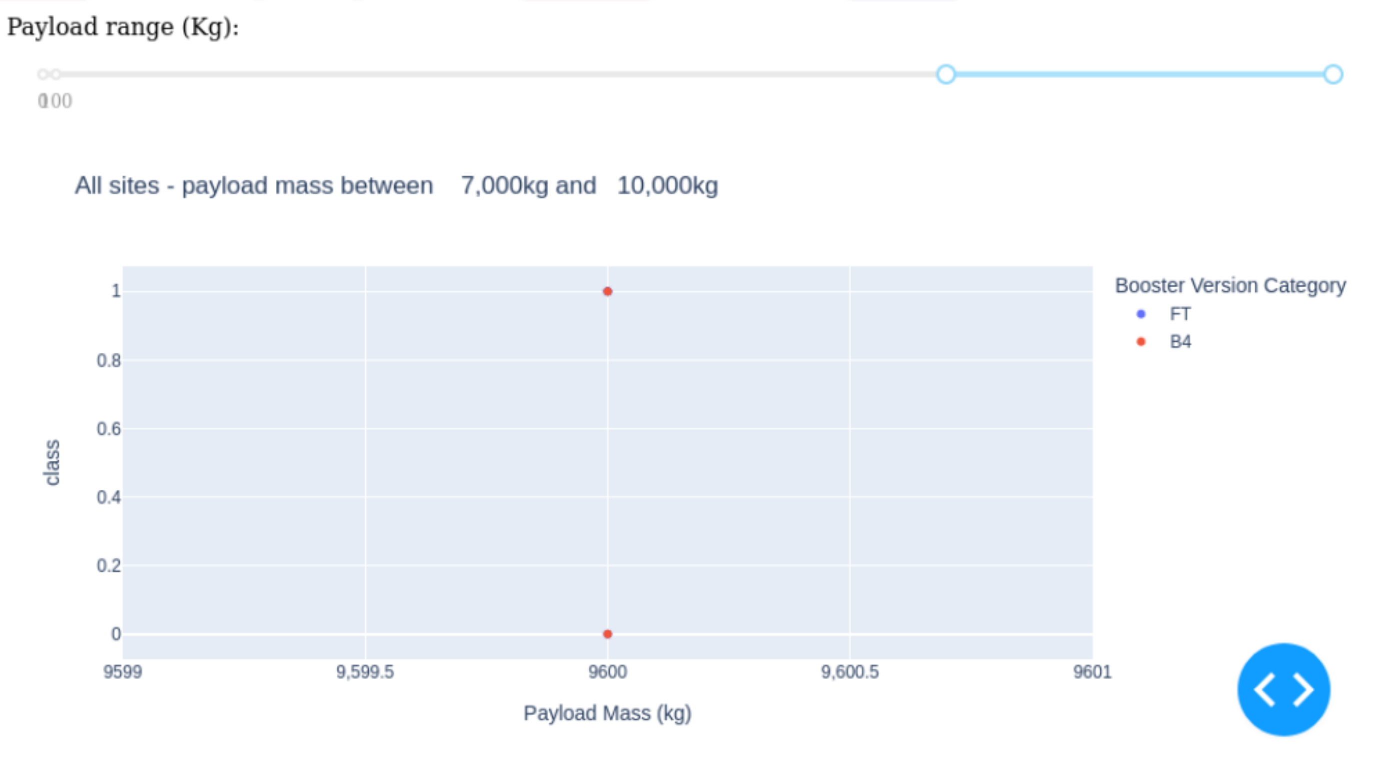
PLOTLY DASHBOARD

It can be observed that the FT boosters have a better success rate in payload masses under 6,000 kgs.



PLOTLY DASHBOARD

There are not enough events to determine the level of success/failure in Launches with a mass over 9,000 kgs.



CONCLUSIONS

- The best launch site is KSC LC-39A
- There are not enough data about launches over 7,000kg to determine success rates.
- Improvement on rocket launch technologies has increased the success rate over time.
- Decision Tree Classifier can be used to predict successful landings to increase profit.

APPENDIX

- Appendix I: SQL Queries Summary
- Appendix II: SQL Queries Executed
- Appendix III: Launch Site Maps

Appendix I: SQL Queries Summary

- Names of the unique launch sites in the space mission
- Launch sites with name starting as ‘CCA’
- Total payload mass carried by bosters launched by NASA
- Average payload mass carried by booster version F9
- Date when the first successful landing is achieved
- Successful booster launches in drone shops and payload mass between 4000 and 6000 kgs.
- Total number of successful and failed mission outcomes
- Booster versions carrying the maximum payload mass
- Failed landing outcomes in drone shop, booster versionas and launch site names
- Rank of landing outcomes

Appendix II: SQL Queries Executed

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

```
sql select distinct launch_site from spacextbl order by 1;
```

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

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

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

```
sql select sum(payload_mass__kg_) as total_payload from spacextbl where payload like '%CRS%';
```

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

total_payload
111268

Appendix II: SQL Queries Executed

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

```
sql select avg(payload_mass__kg_) as avg_payload from spacextbl where booster_version ='F9 v1.1';
```

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

avg_payload
2928.4

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

Hint:Use min function

```
sql select min(date) as first_success_gp from spacextbl where [Landing _Outcome] = 'Success (ground pad)';
```

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

first_success_gp
01-05-2017

Appendix II: SQL Queries Executed

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 payload_mass__kg_ between 4000 and 6000 and [Landing _Outcome] ='Success (drone ship)';
```

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

Booster_Version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

List the total number of successful and failure mission outcomes

```
sql select mission_outcome, count(*) as qty from spacextbl group by mission_outcome order by mission_outcome;
```

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

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

Appendix II: SQL Queries Executed

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

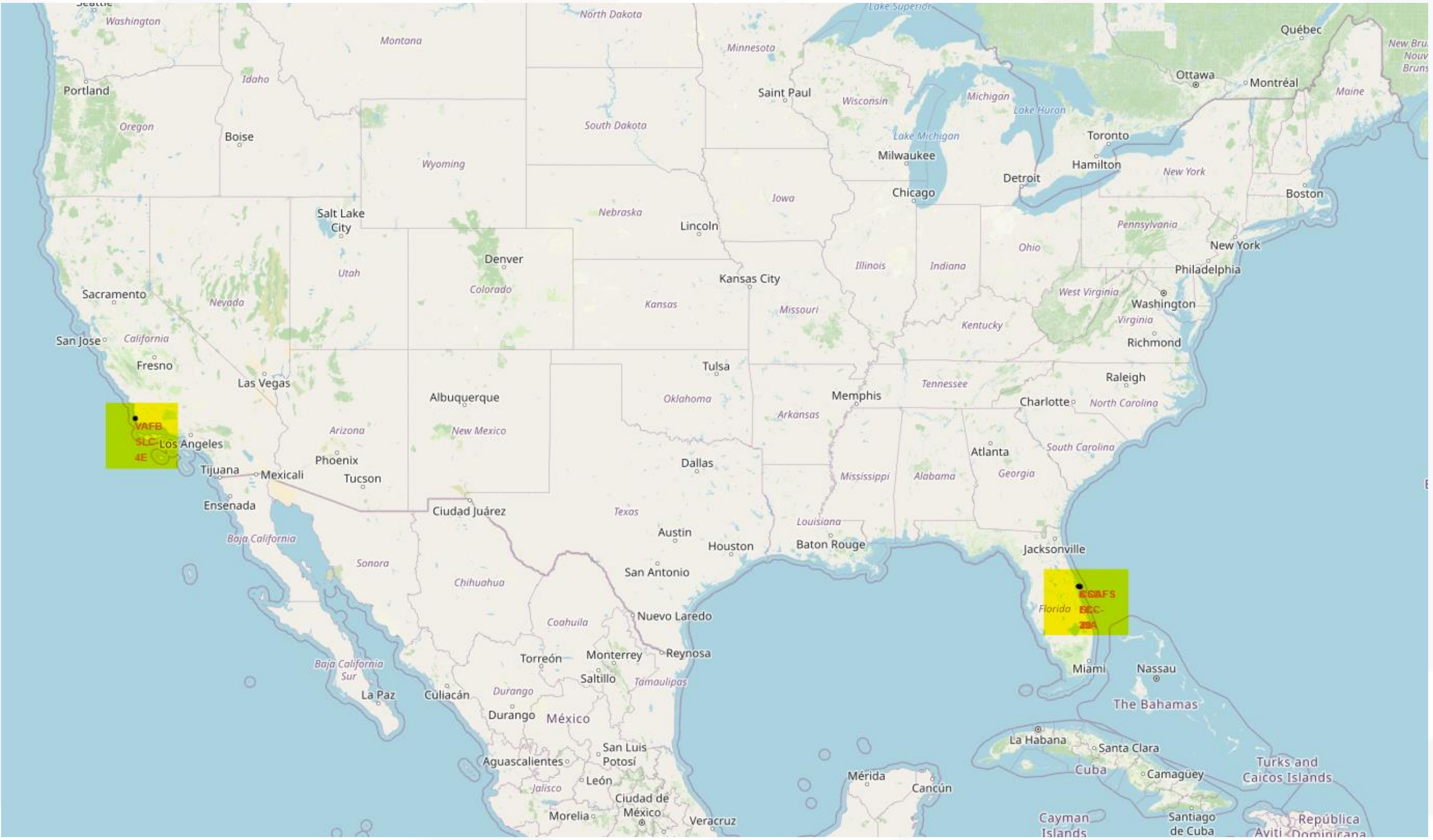
```
sql select [Landing _Outcome], count(*) as qty from spacextbl where date between '04-06-2010' and '20-03-2017' group by [Landing _Outcome] order by qty desc;
```

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

Landing _Outcome	qty
Success	20
No attempt	10
Success (drone ship)	8
Success (ground pad)	6
Failure (drone ship)	4
Failure	3
Controlled (ocean)	3
Failure (parachute)	2
No attempt	1

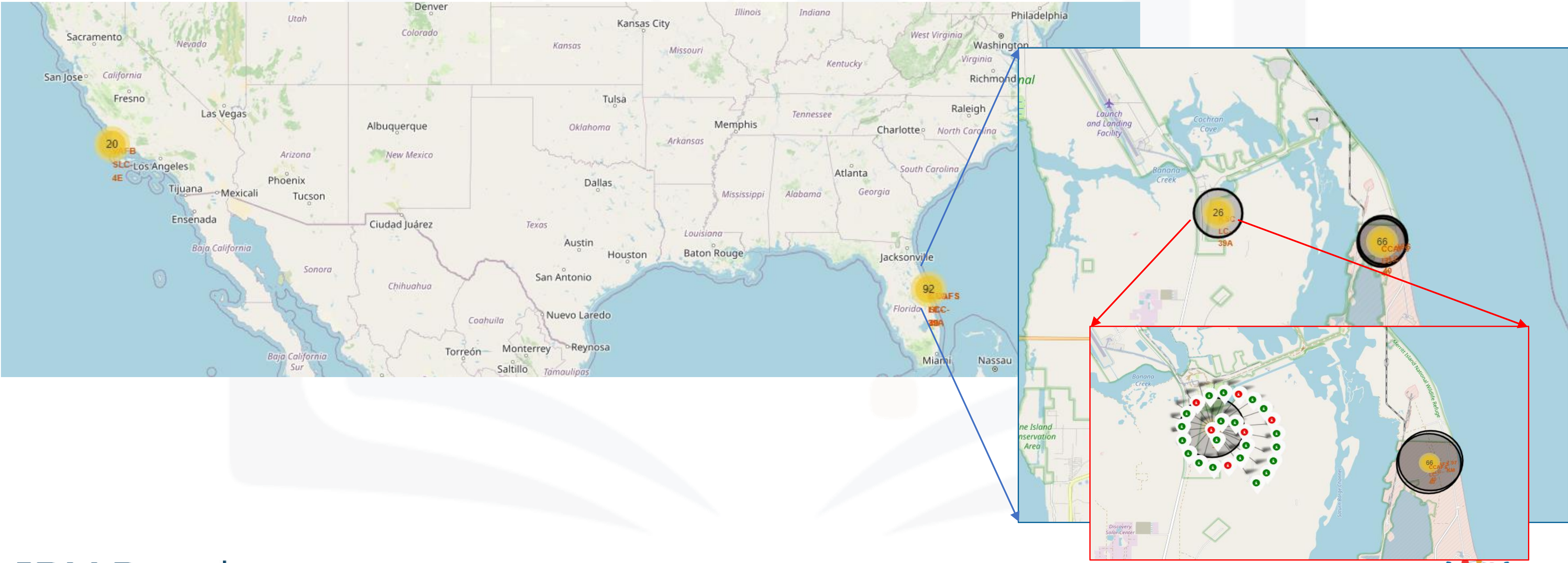
Appendix III: Launch Site Maps

Launch site locations.



Appendix III: Launch Site Maps

Launch counts by site. And success/fail launch result represented by color green/red on each location.



Appendix III: Launch Site Maps

Total distance from habitated zones, rails and roads to reference safety and logistics on launch sites.

Thank you!

https://github.com/AlePenaFlores/AppliedDataScienceCapstone.git