
# Winning Space Race with Data Science

Dmitriy Sizionov
10/04/2022

# Outline

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

# Executive Summary

SpaceX has reached incredible results in launching rockets. While, it didn't happen in one day. There were a lot of failures in launching and landing before the Launch success rate starts to grow. This research uses data from SpaceX's API to find the reasons for Launch success rate growth.

Research results suggest that the most successful combination of variable elements for launches are the following:

1. Launch site: KSC LC-39A

2. Booster version: FT

3.Payload Range: 2000 – 5500 kg

# Introduction

SpaceX spends years finding the most optimal and successful combination of factors to launch rockets. Taking knowledge from their issues would be helpful for any new company that wants to launch rockets. This research aims as a target to find the most successful combination of elements for launching rockets. Follow the questions going to be answered:

1. Most convenient Launching site
2. Most successful booster version
3. Most secure payload range

Section 1

# Methodology

# Methodology

## Executive Summary

- Data collection methodology:
  - Data collection from API
- Perform data wrangling
  - Issues with missing values
- 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

- To predict the Falcon 9 first stage successful rate. This information can be used if an alternate company wants to bid against SpaceX for a rocket launch.
- Data collection is possible from SpaceX API. Via this API there is access to the following information: BoosterVersion, PayloadMass, Orbit, LaunchSite, Outcome, Flights, GridFins, Reused , Legs, LandingPad, Block, ReusedCount, Serial, Longitude, Latitude

# Data Collection – SpaceX API

1. Get the data from data API (source)

```
response = requests.get(spacex_url)
```

2. Normalize data and put it in Pandas Dataframe

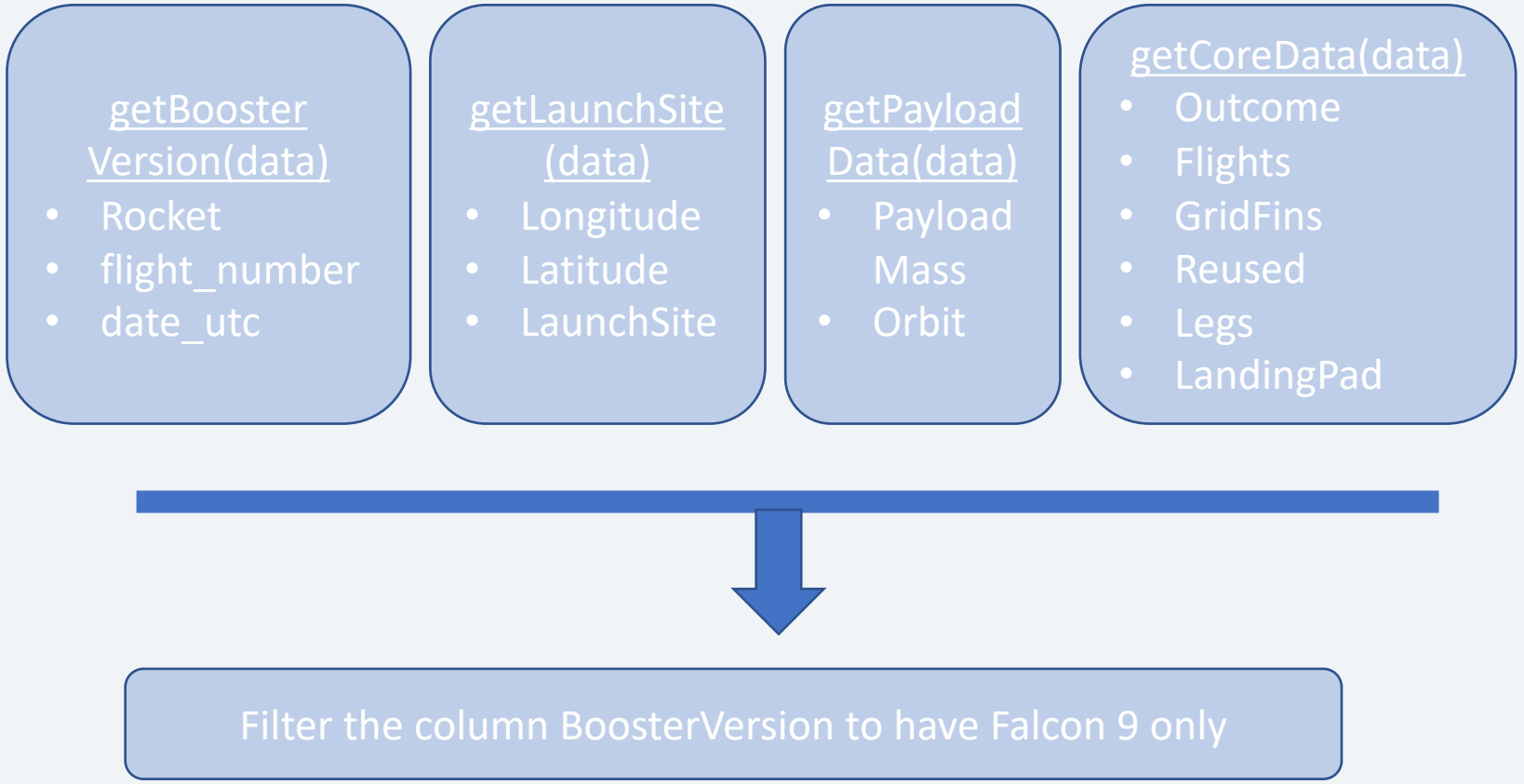
```
Data normalization for Pandas Dataframe
```

- GitHub_URL

# Data Collection - Scraping

1. Execute four ‘def’ commands to fill columns in dataframe
2. Use filter in column “BoosterVersion” to leave Falcon 9 only

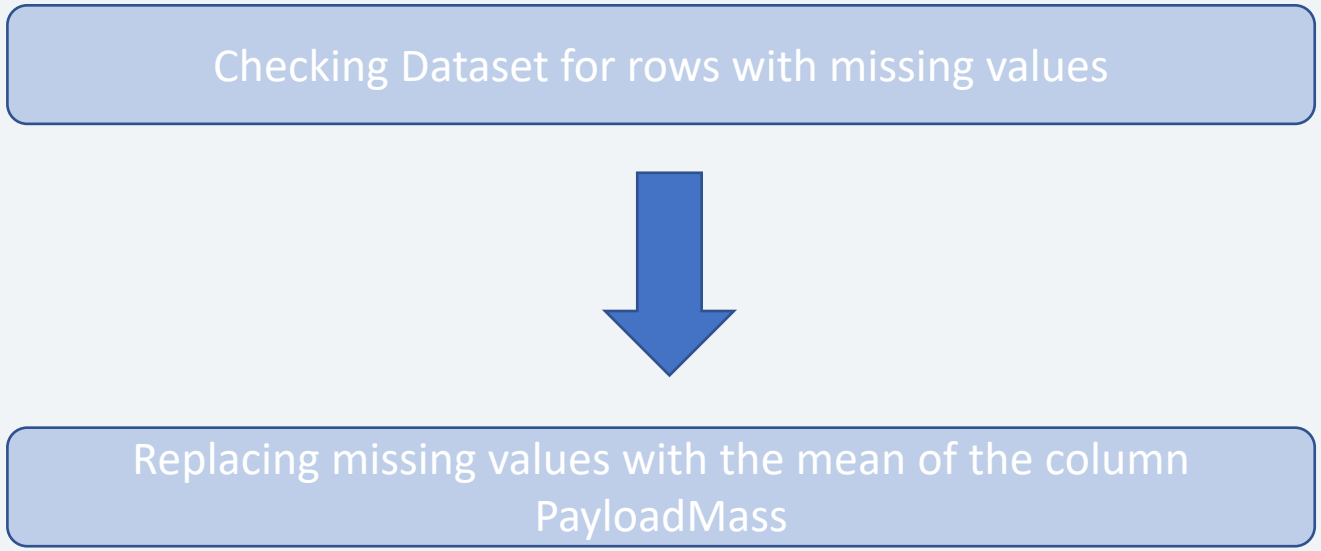
- GitHub_URL



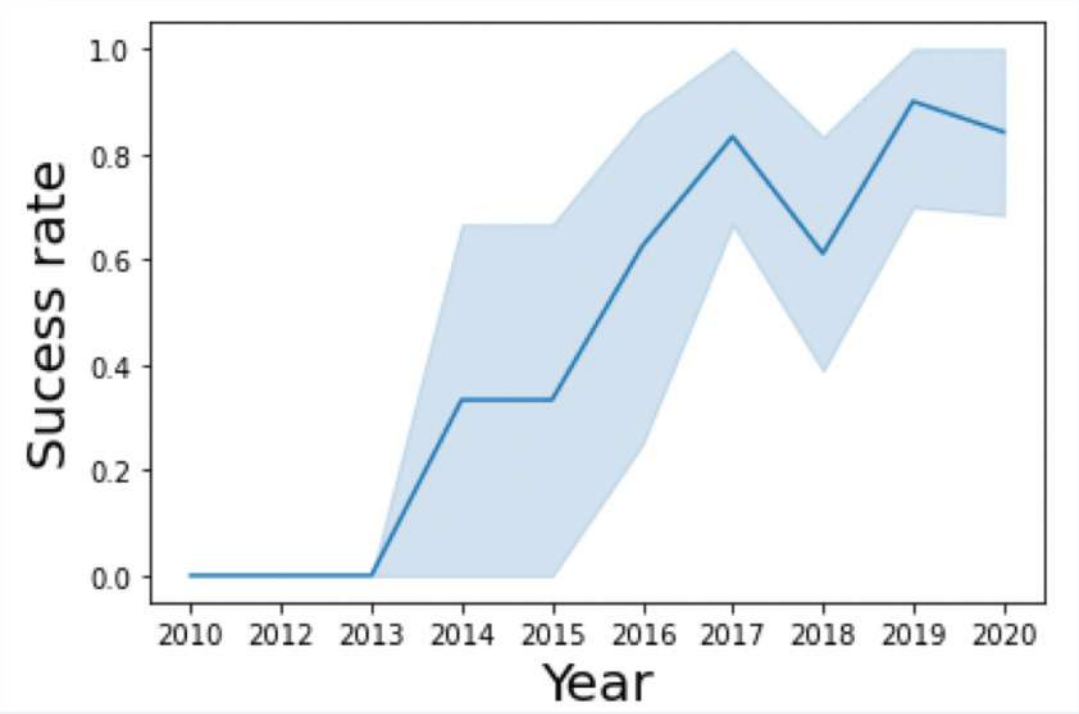
# Data Wrangling

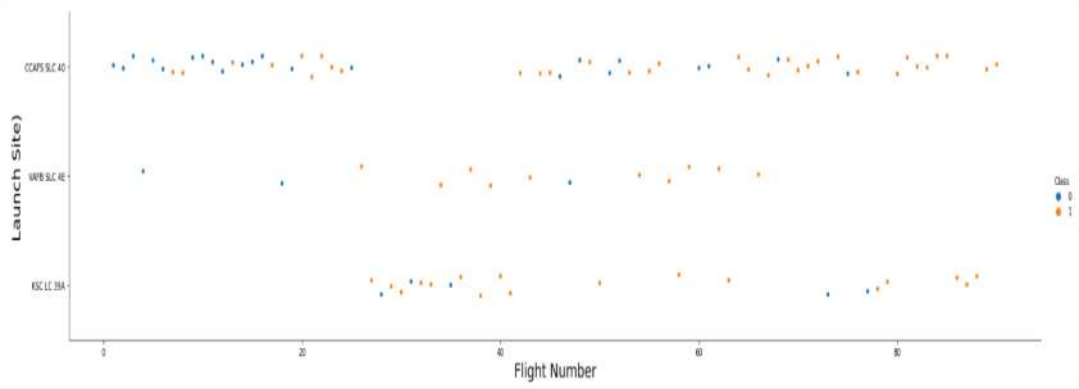
1. Dataset check for rows with missing values
2. Calculation mean in PayloadMass column
3. Replacing missing values with the mean

- GitHub_URL



# EDA with Data Visualization




- Success rate has a positive trend since 2013 grew from scratch to 0.8
- Company started with its site and then add 2 additional sites.
- Most commonly used site CCAFS LC-40 improved success rate during the time

- GitHub URL

# EDA with SQL

- Data source was connected via DB2 magic SQL
- Calculated maximum and average payload mass 15600 and 2534 respectively
- First date of success landing 2010-06-04

- GitHub URL

# Build an Interactive Map with Folium

The Interactive map with Folium includes:

- Circles around Launch sites with names
- Numbers of launches
- Distance between the launch site and roads, railways, cities (green lines)

- GitHub URL

# Build a Dashboard with Plotly Dash

Dashboard includes:

- Pie chart and Scatter plot with possibility chose particular Launch Site and Payload range

Those graphs were added to investigate the following data features:

- Site with largest successful launches – KSC LC-39A
- Site with largest highest launch success rate - KSC LC-39A
- Payload range with highest launch success rate – 3000-4000kg
- Payload range with lowest launch success rate – 6000-7000kg
- F9 Booster version with highest launch success rate - FT
- GitHub_URL

# Predictive Analysis (Classification)

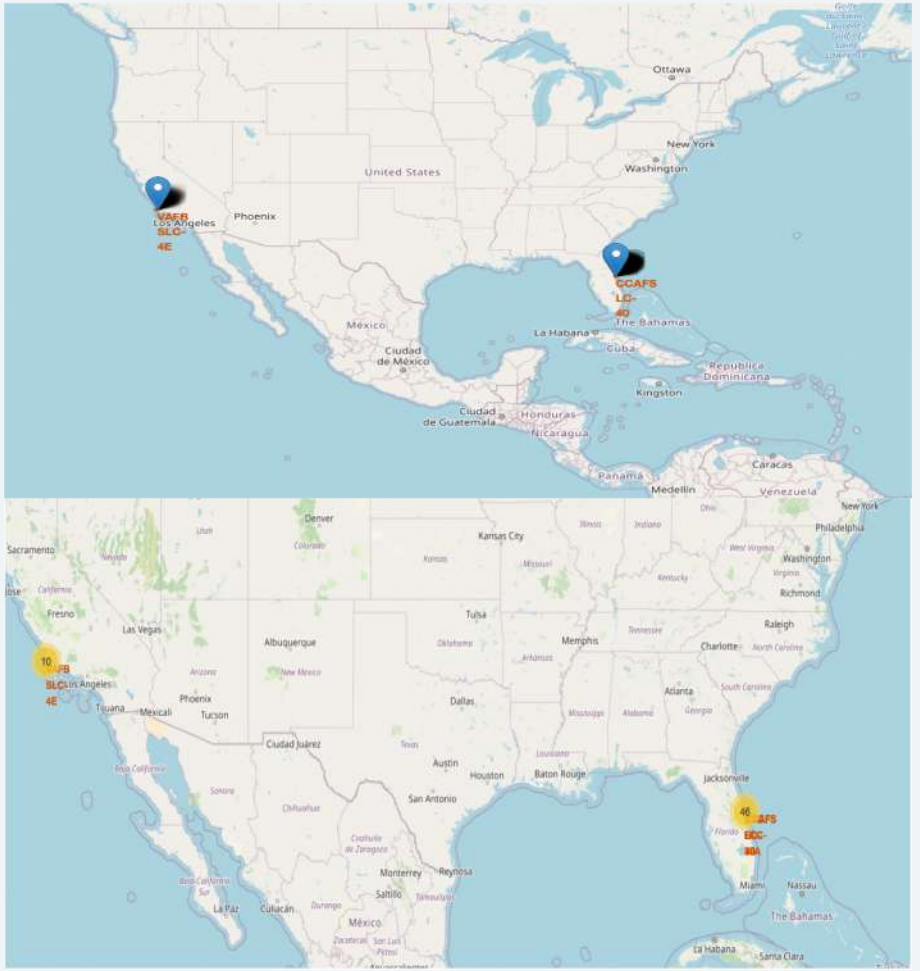
- Dataset was separated for train and test datasets.
- Dependent variable – Landing (yes or no)
- Several models were executed and tested for representativeness of the data and accuracy of prediction
- Regression shows higher representativeness (0.90) and higher accuracy (0.82) vs other models

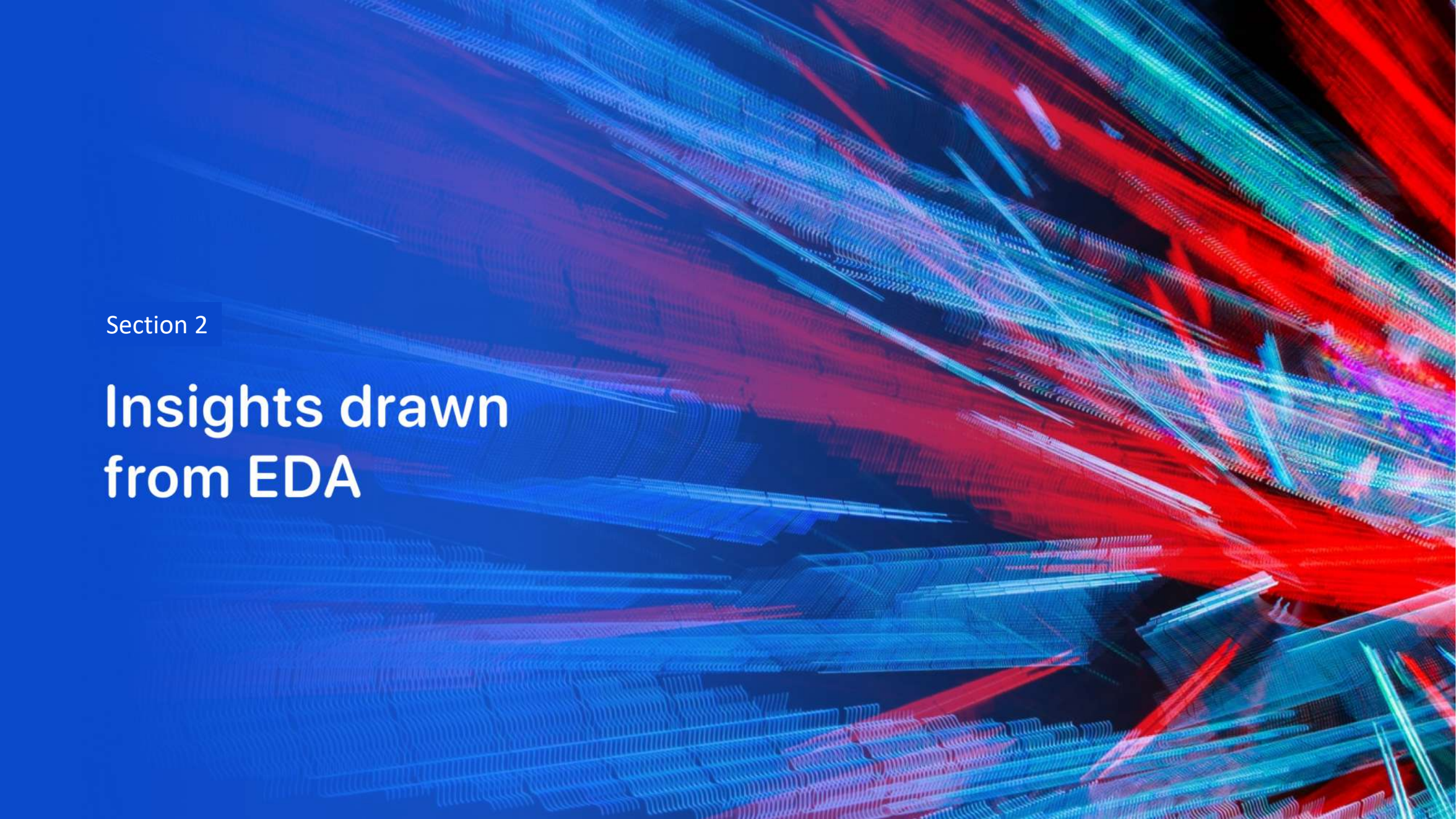
| | Algorithm | Score | Accuracy |
|---|---|---|---|
| 0 | KNN | 0.611111 | 0.664286 |
| 1 | Decision Tree | 0.666667 | 0.875000 |
| 2 | SVM | 0.617614 | 0.667857 |
| 3 | LogisticRegression | 0.902778 | 0.819643 |

- GitHub URL

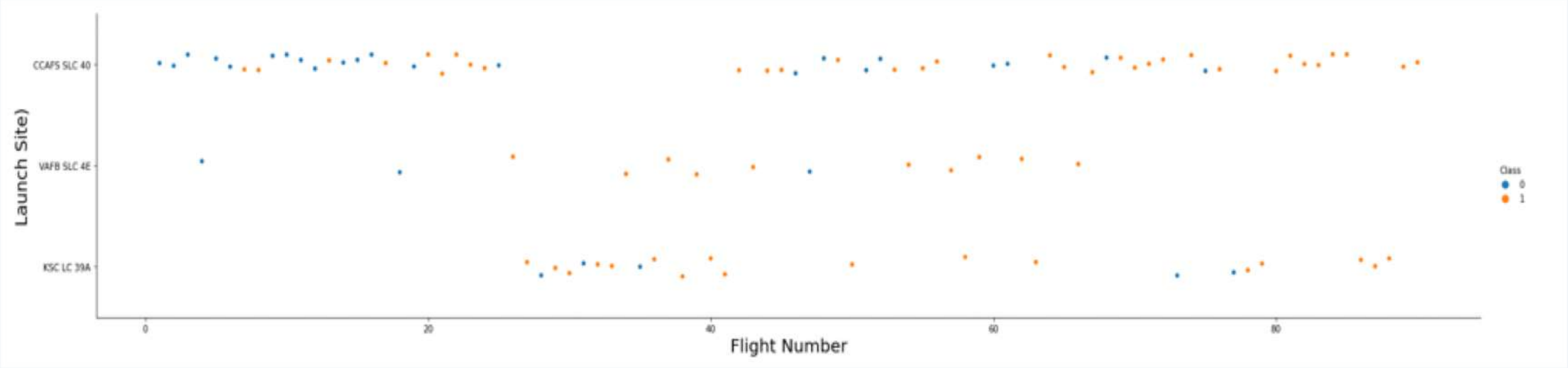
# Results

- Exploratory data analysis results
  - Success landing rate growths over time
  - The most common successful Launch site is KSC LC-39A
  - The most common successful Booster version is FT
- Interactive analytics demo in screenshots
  - Launch site with highest success Landing rate and most commonly used is situated on the coast of Atlantic sea
- Predictive analysis results
  - The regression method shows the highest representativeness of the model and the highest accuracy of prediction




Section 2
Insights drawn
from EDA

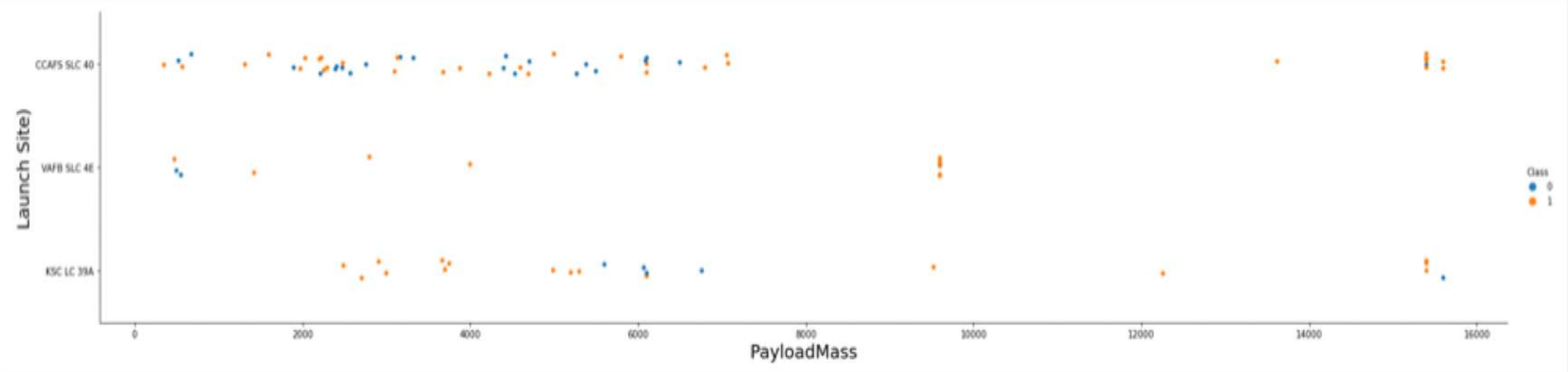
# Flight Number vs. Launch Site




Launch site exploration goes in 3 phases:

1. Launches start on CCAFS LC-40 with a low rate of success launches
2. Launches started on another two Launch sites with a higher success rate
3. All three sites start to be used with improvement in success rate

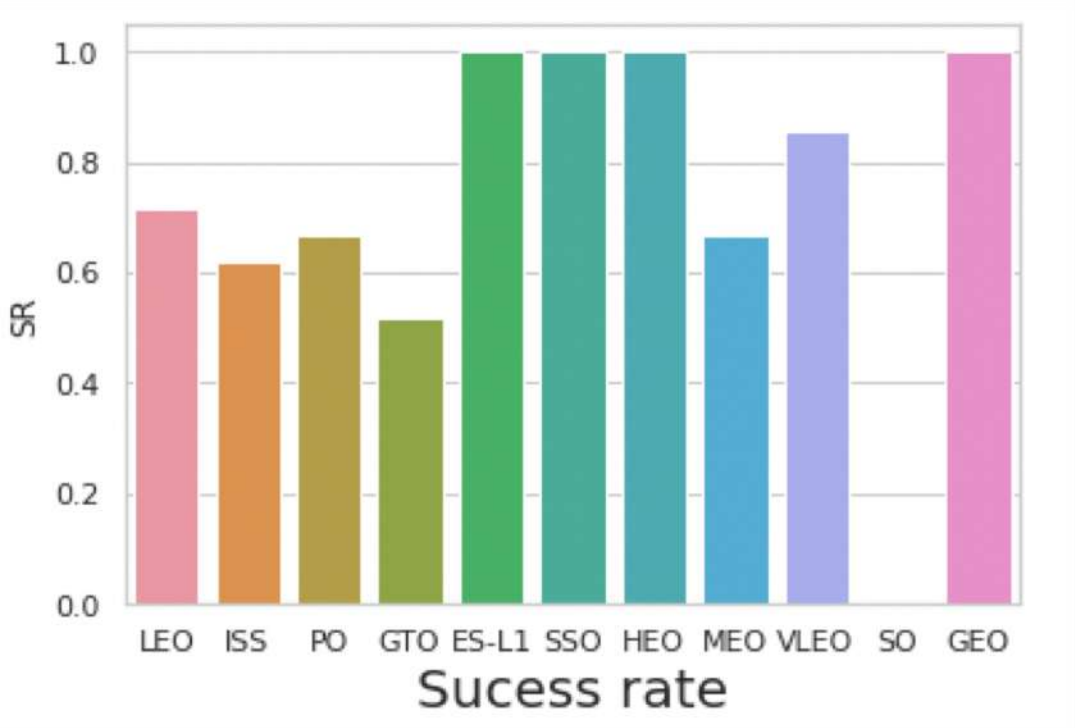
# Payload vs. Launch Site



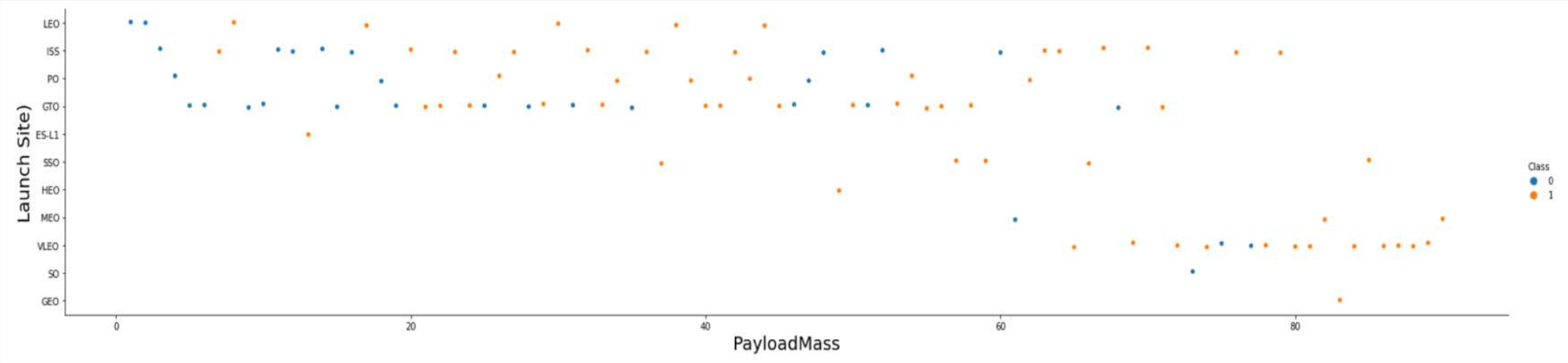

- There are launches on all 3 sites for rockets under 10 000 kg
- Heavy payload rockets were launched only on LC-40 and LC-39A

# Success Rate vs. Orbit Type

- The most Successful rate was achieved with ES-L1, SSO, HEO, GEO
- There is the lowest rate connected to SO orbit type

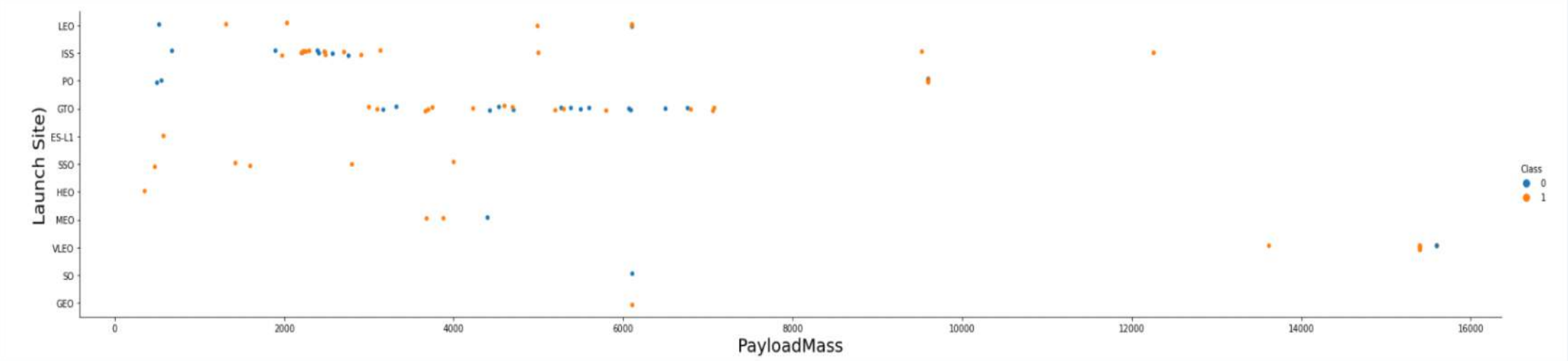


# Flight Number vs. Orbit Type




- Launches started with LEO, ISS, PO, GTO with a low success rate
- After 20 launches success rate started to grow
- After 40 Launches SSO was additionally used
- After 60 Launches VLEO was additionally used

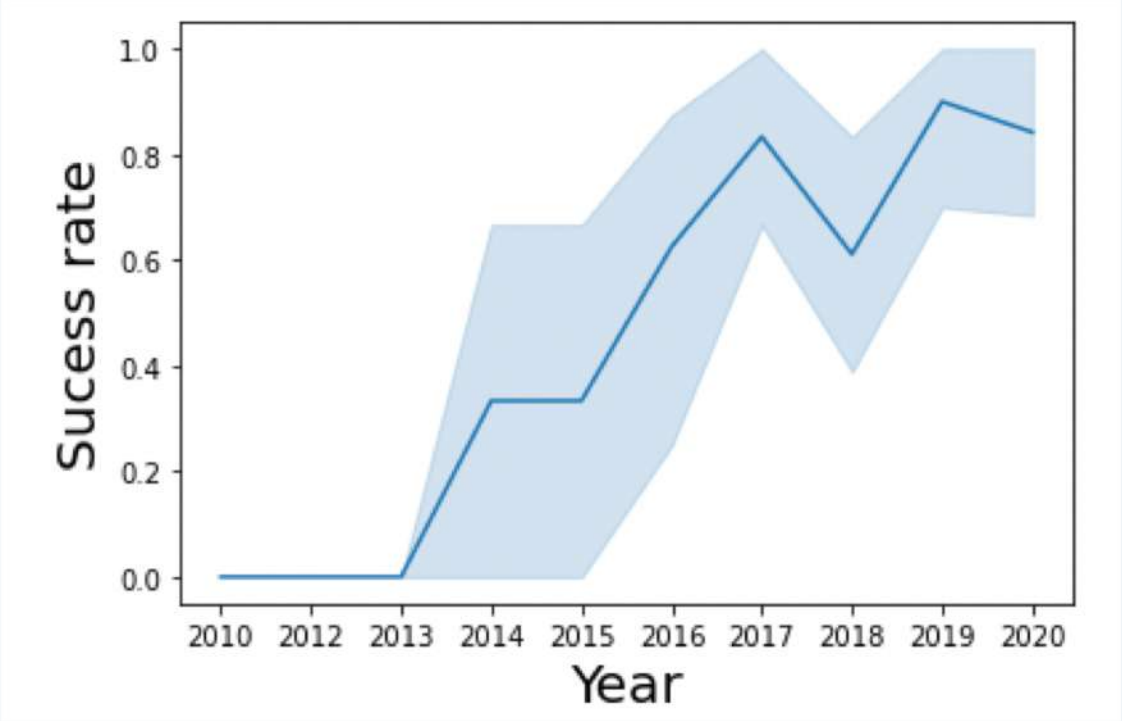
# Payload vs. Orbit Type




- Most launches were in the Payload range of 2000 – 6000 kg
- Most commonly used ISS and GTO orbit types
- Heavy payload rockets used VLEO orbit type

# Launch Success Yearly Trend

- Success rate started to grow in 2013
- Success rate overreached 0.8 success rate in 2017 for the first time
- Success rate is maintained in 2019 and 2020 at over 0.8



# All Launch Site Names

```
%sql select DISTINCT launch_site From SPACEX_30_03
```

```
 * ibm_db_sa://dyd78272:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32536
/bludb
Done.
```

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

- There are 4 Launch sites are used

# Launch Site Names Begin with 'CCA'

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

```
%sql select * From SPACEX_30_03 where launch_site like 'CCA%' limit 5
```

```
 * ibm_db_sa://dyd78272:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqblod8lcg.databases.appdomain.cloud:32536
/bludb
Done.
```

| 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 | 07:44:00 | F9 v1.0 B0005 | CCAFS LC-40 | Dragon demo flight C2 | 525 | LEO (ISS) | NASA (COTS) | Success | No attempt |
| 2012-10-08 | 00:35:00 | 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 |

- 5 Launches from CCAAFS LC-40
- F9 v1.0 Booster participated in all those launches

# Total Payload Mass

```
%sql select Sum (payload_mass__kg_) From SPACEX_30_03 where customer like 'NASA (CRS)'
 * ibm_db_sa://dyd78272:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32536/bludb
Done.
```

3]:

| 1 |
|---|
| 45596 |

- Total payload carried by boosters from NASA - 45 596 kg

# Average Payload Mass by F9 v1.1

```
%sql select AVG (payload_mass__kg_) From SPACEX_30_03 where booster_version like 'F9 v1.1%'
```

```
 * ibm_db_sa://dyd78272:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32536
/bludb
Done.
```

| 1 |
|---|
| 2534 |

- Booster version F9 v1.1 carried an average payload mass of 2534 kg

# First Successful Ground Landing Date

```
%sql select MIN (DATE) From SPACEX_30_03 where mission_outcome like 'Success'
```

```
 * ibm_db_sa://dyd78272:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32536
/bludb
Done.
```

| 1 |
| --- |
| 2010-06-04 |

- First successful landing outcome on the ground pad was at 2010-06-04

# Successful Drone Ship Landing with Payload between 4000 and 6000

```
%sql select booster_version From SPACEX_30_03 where mission_outcome like 'Success' and payload_mass__kg_ BETWEEN 4000 AND 6000 and landing__outcome like 'Success (drone ship)'
```

```
 * ibm_db_sa://dyd78272:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32536/bludb
Done.
```

]:

| booster_version |
|---|
| F9 FT B1022 |
| F9 FT B1026 |
| F9 FT B1021.2 |
| F9 FT B1031.2 |

- List the names of boosters that have successfully landed on a drone ship and had payload mass greater than 4000 but less than 6000

# Total Number of Successful and Failure Mission Outcomes

```
%sql select mission_outcome, count (booster_version) From SPACEX_30_03 GROUP BY mission_outcome
```

```
 * ibm_db_sa://dyd78272:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32536
/bludb
Done.
```

| mission_outcome | 2 |
|---|---|
| Failure (in flight) | 1 |
| Success | 99 |
| Success (payload status unclear) | 1 |

- Successful mission outcomes: 99 + 1 (with payload status unclear)
- Failure mission outcome: 1 (in flight)

# Boosters Carried Maximum Payload

```
%sql select booster_version, payload_mass__kg_ From SPACEX_30_03 ORDER BY payload_mass__kg_ DESC limit 10
```

```
 * ibm_db_sa://dyd78272:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32536
/bludb
Done.
```

| booster_version | payload_mass__kg_ |
|---|---|
| F9 B5 B1048.4 | 15600 |
| F9 B5 B1051.6 | 15600 |
| F9 B5 B1058.3 | 15600 |
| F9 B5 B1060.2 | 15600 |
| F9 B5 B1049.5 | 15600 |
| F9 B5 B1051.4 | 15600 |
| F9 B5 B1048.5 | 15600 |
| F9 B5 B1056.4 | 15600 |
| F9 B5 B1051.3 | 15600 |
| F9 B5 B1049.4 | 15600 |

- Payload mass of 15600 kg is maximal for all booster versions
- All booster versions are variations of F9 B5 B1

# 2015 Launch Records

```
%sql select booster_version, launch_site, date, landing__outcome from SPACEX_30_03  where landing__outcome like 'Failu
re%' and date like '2015%'
```

```
 * ibm_db_sa://dyd78272:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32536
/bludb
Done.
```

| booster_version | launch_site | DATE | landing__outcome |
|---|---|---|---|
| F9 v1.1 B1012 | CCAFS LC-40 | 2015-01-10 | Failure (drone ship) |
| F9 v1.1 B1015 | CCAFS LC-40 | 2015-04-14 | Failure (drone ship) |

- In 2015 only 2 launches finished with landing failure on the CCAFS LC-40 launch site with booster version F9 v1.1

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

```
%sql select landing__outcome, count (landing__outcome) from SPACEX_30_03 where DATE between '2010-06-04' and '2017-03-20' GROUP BY landing__outcome ORDER BY count (landing__outcome) DESC
```

```
 * ibm_db_sa://dyd78272:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32536/bludb
Done.
```

| landing__outcome | 2 |
|---|---|
| No attempt | 10 |
| Failure (drone ship) | 5 |
| Success (drone ship) | 5 |
| Controlled (ocean) | 3 |
| Success (ground pad) | 3 |
| Failure (parachute) | 2 |
| Uncontrolled (ocean) | 2 |
| Precluded (drone ship) | 1 |

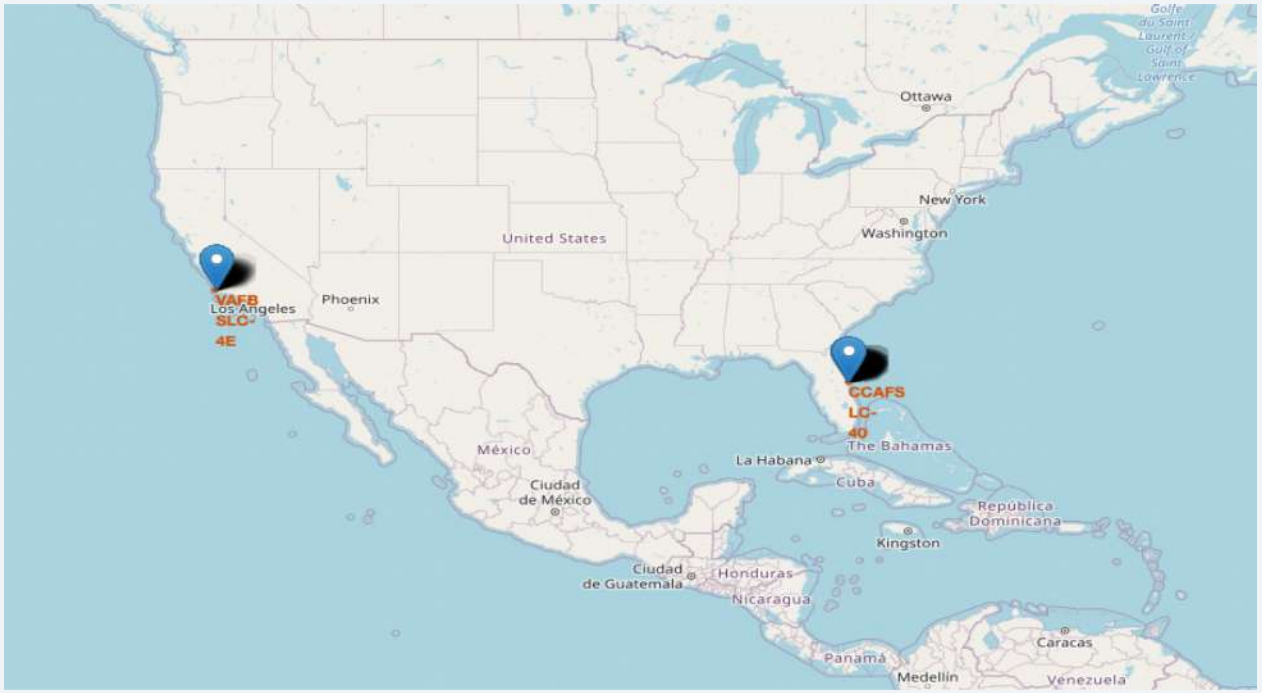
- The most common landing outcome is ”No attempt”

Section 3

# Launch Sites Proximities Analysis

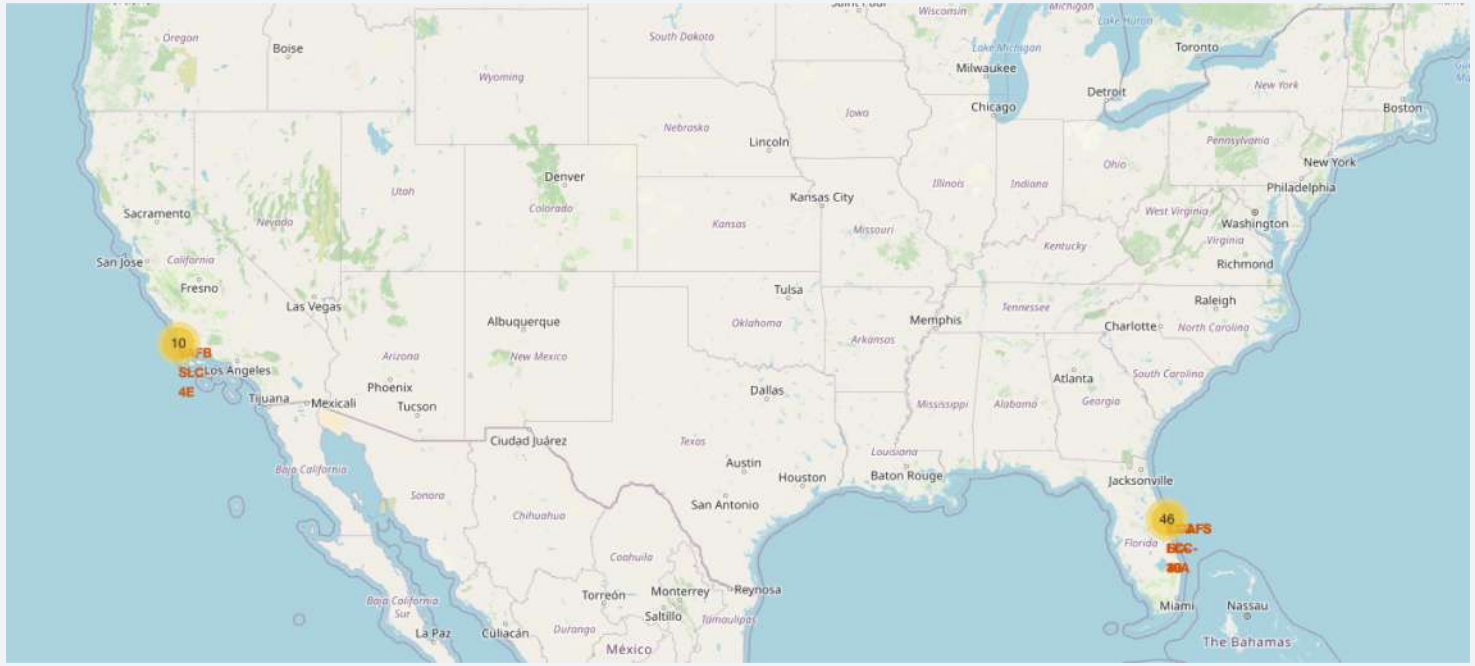
# Map with Launch site location

- There are 4 launch sites
  1 on the West coast of the US
  3 on the East coast of the US

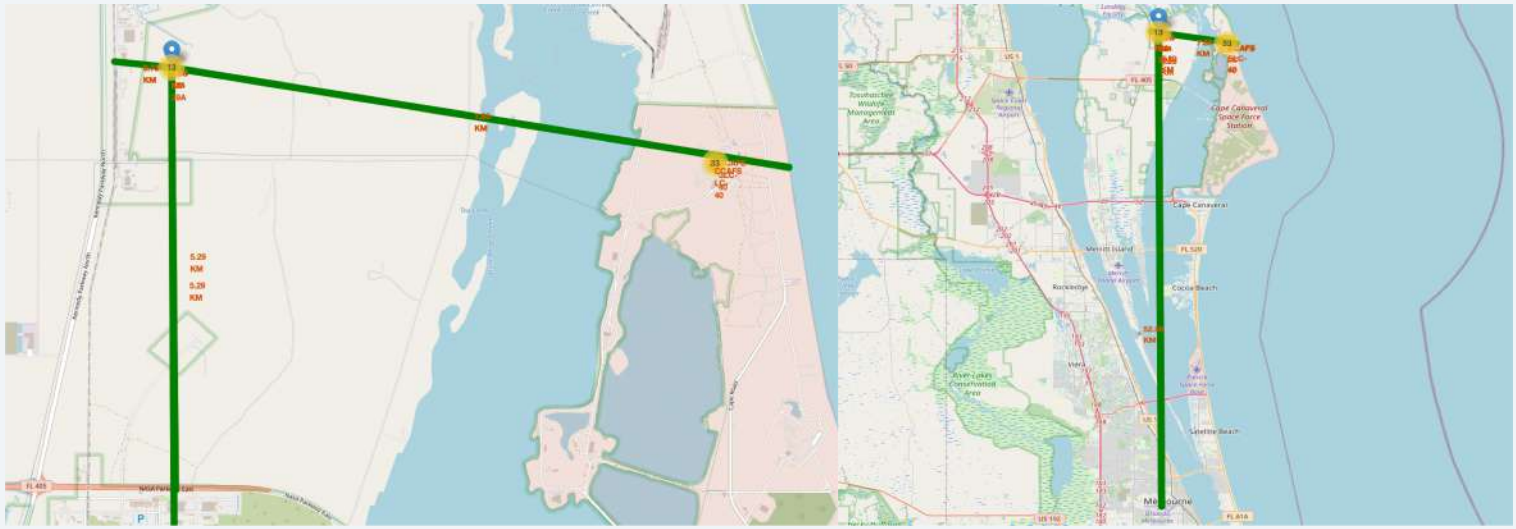


# Map with launch site locations and number of launches

- East coast sites have almost 5 times more launches versus West coast location



# Map with distance to initial logistic points



East coast locations is situated close to:

1.coastline

2.railroad

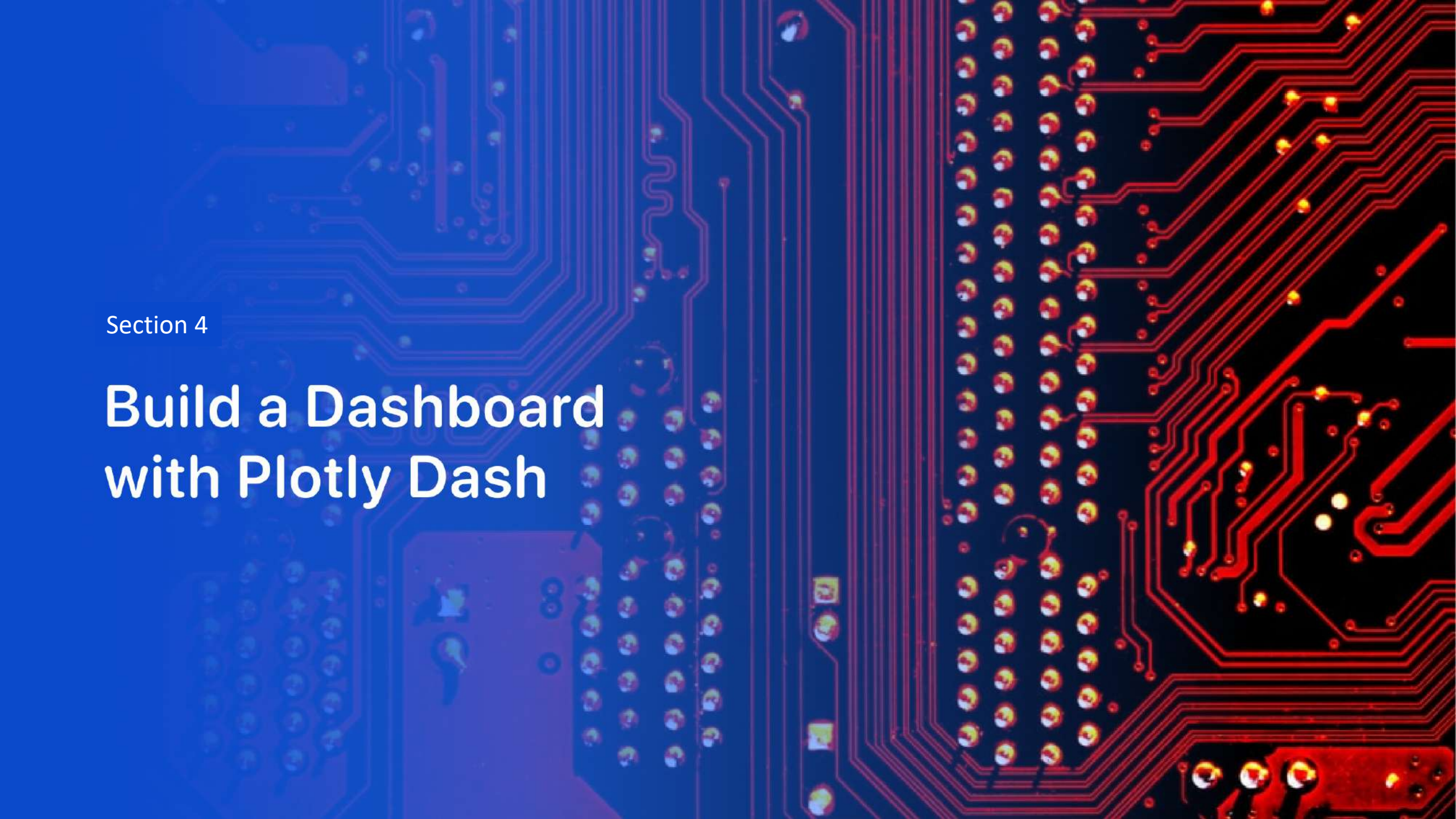
3.road

4.city

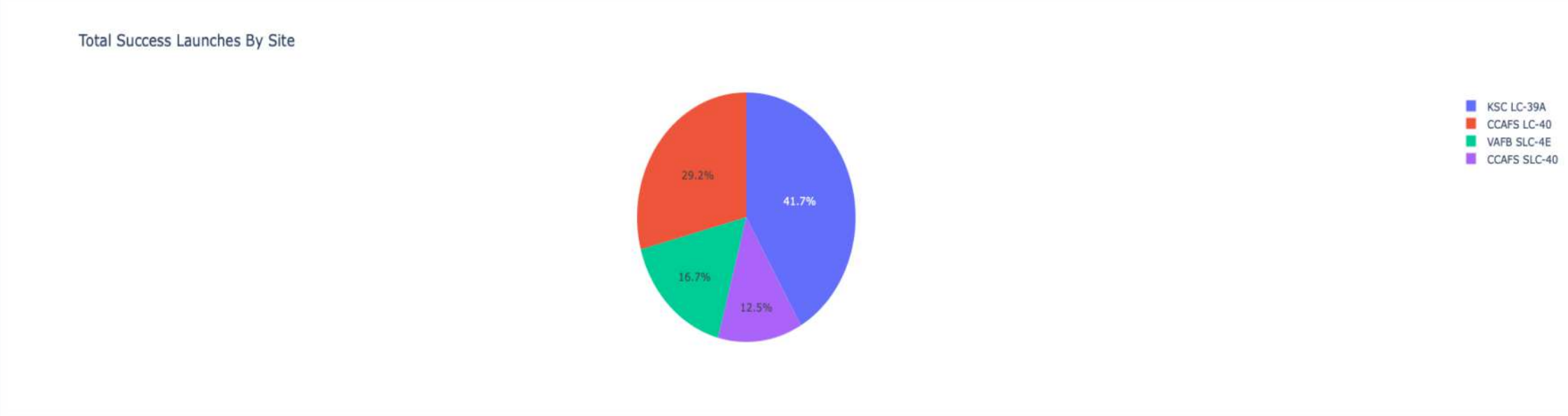
```
distance_coastline : 7.829177938657839
distance_railroad : 0.7167190349513248
distance_road : 5.288139317514574
distance_city : 52.55708317707729
```

Section 4

# Build a Dashboard with Plotly Dash

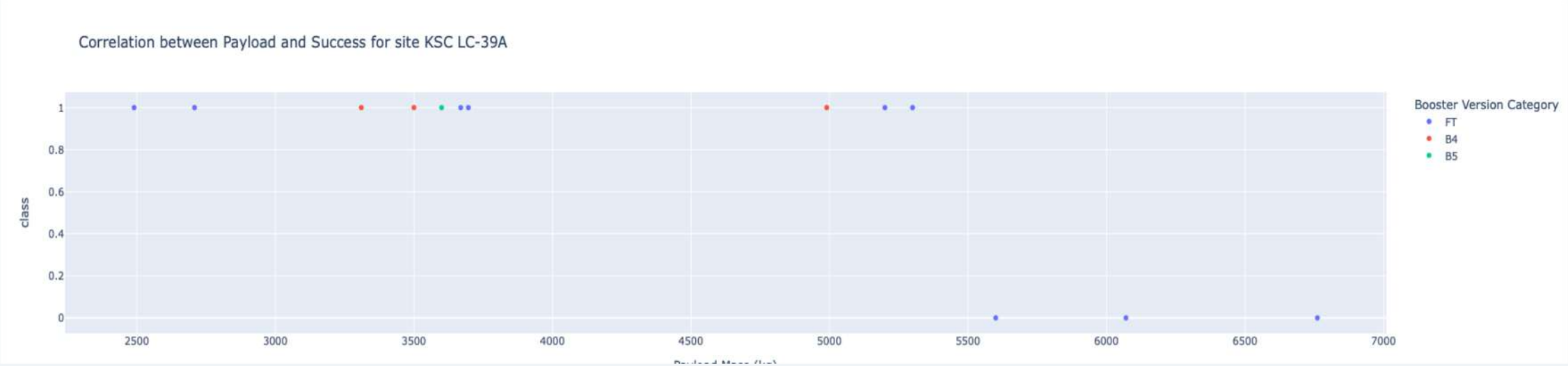


# Success Launches by Site



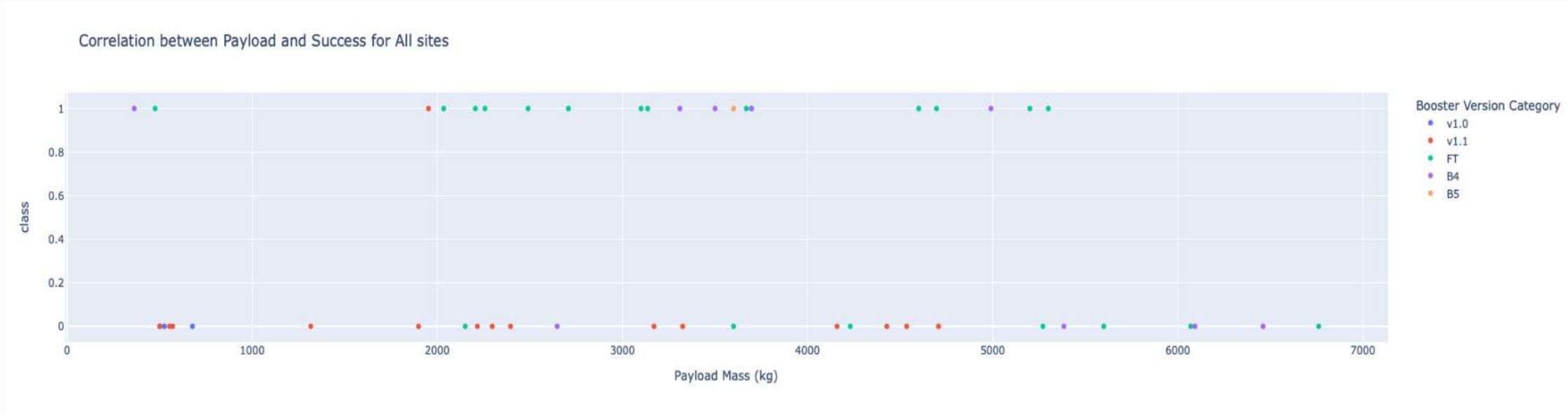

The Highest number of successful launches executed from KSC LC-39A

# KSC LC-39A




- Most successful Booster version on this site is FT
- Most successful Payload range for this site is (2500 – 5500) kg

# Success rate based on Payload and Booster Version



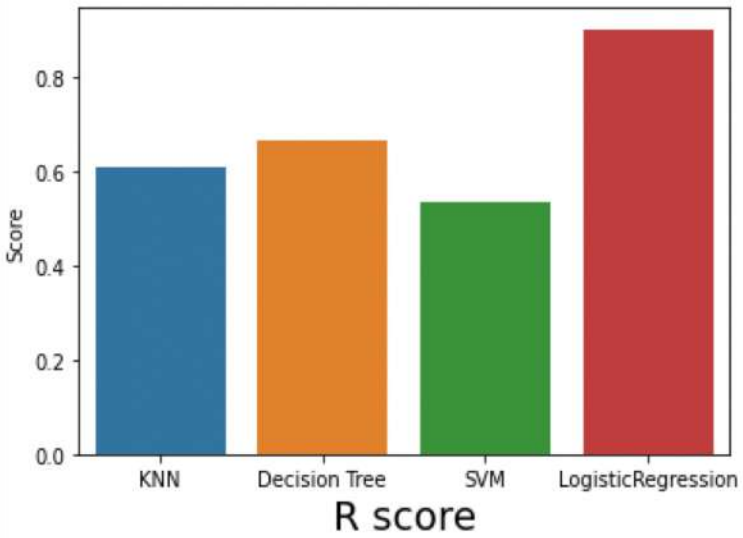

- Most successful Booster version is FT
- Most successful Payload range is (2000 – 5500) kg

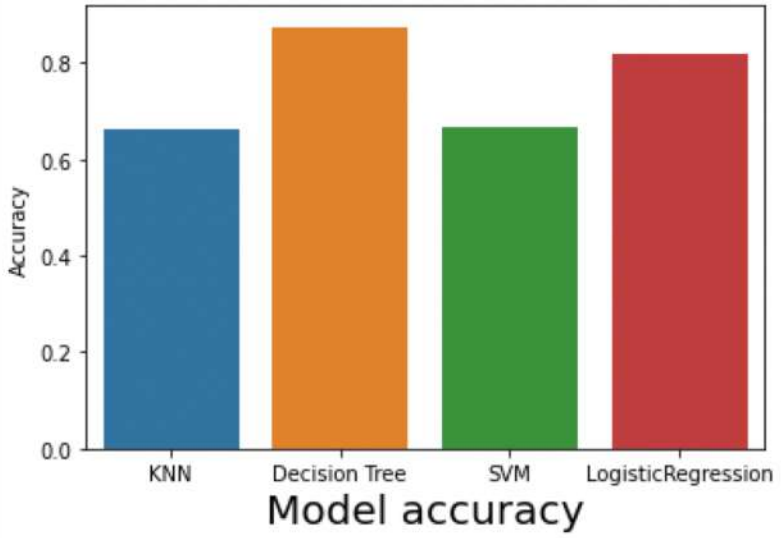
Section 5

# Predictive Analysis (Classification)

# Classification Accuracy

| | Algorithm | Score | Accuracy |
|---|---|---|---|
| 0 | KNN | 0.611111 | 0.664286 |
| 1 | Decision Tree | 0.666667 | 0.875000 |
| 2 | SVM | 0.617614 | 0.667857 |
| 3 | LogisticRegression | 0.902778 | 0.819643 |



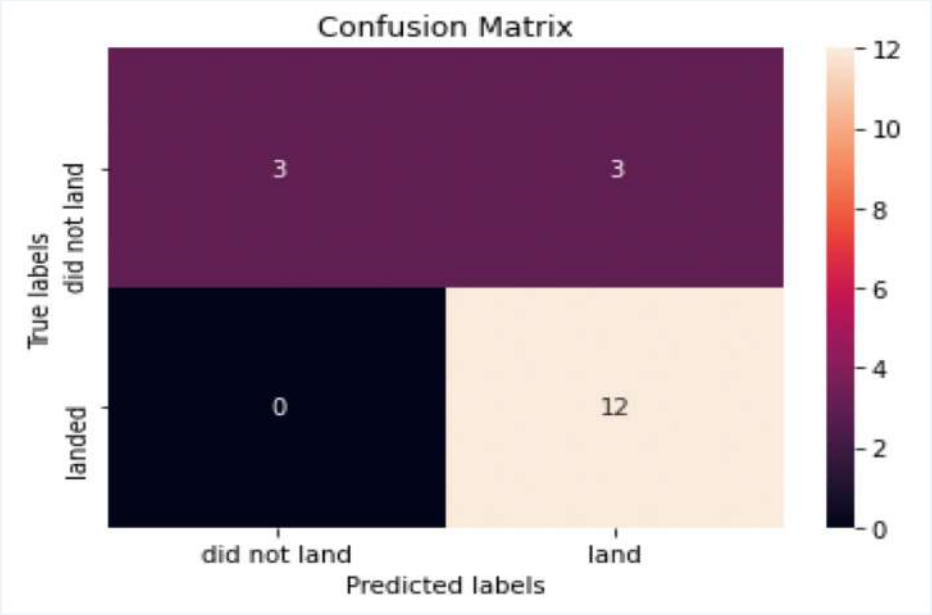

- Logistic regression has the highest R-score : 0.90
- Decision Tree has the highest model accuracy: 0.87

Due to the significant difference in R-score and slight difference in Accuracy, the most optimal choice of model would be Logistic regression with R-score of 0.9 and Accuracy of 0.8

# Confusion Matrix

Confusion matrix Logistic regression shows:

- Model predicts ‘successful landing’ with high accuracy
- Prediction of ‘did not land’ has a low accuracy of prediction



# Conclusions

Logistic Regression provides high predictive power and estimators with high accuracy.

Additionally, the best parameters for a successful launch are the following:

1. Launch site: KSC LC-39A
2. Booster version: FT
3. Payload Range: 2000 – 5500 kg

# Appendix

- Folium Lab link – Git URL

# Thank you!