IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Christian Wussow
15.06.2022

Outline

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

Executive Summary

- API's, Webscraping, Data Wrangling, Data Visualization, Machine Learning algorithms
- The landing outcome of the stage 1 Falcon 9 rocket is dependent on various factors, like launching site, payload mass and booster version
- Our best prediction model has an accuracy of 88% on the training and 83% on the test set

Introduction

- SpaceX advertises Falcon 9 rocket launches on its website, with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage.
- We try to predict if Falcon 9 first stage will land successfully given certain factors to determine the cost of a launch. This might be valuable information for competitors of SpaceX

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - API's and Webscraping
- Perform data wrangling
  - Missing values were replaced by the mean value of the column
- 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 sets were collected using API's and with Webscraping

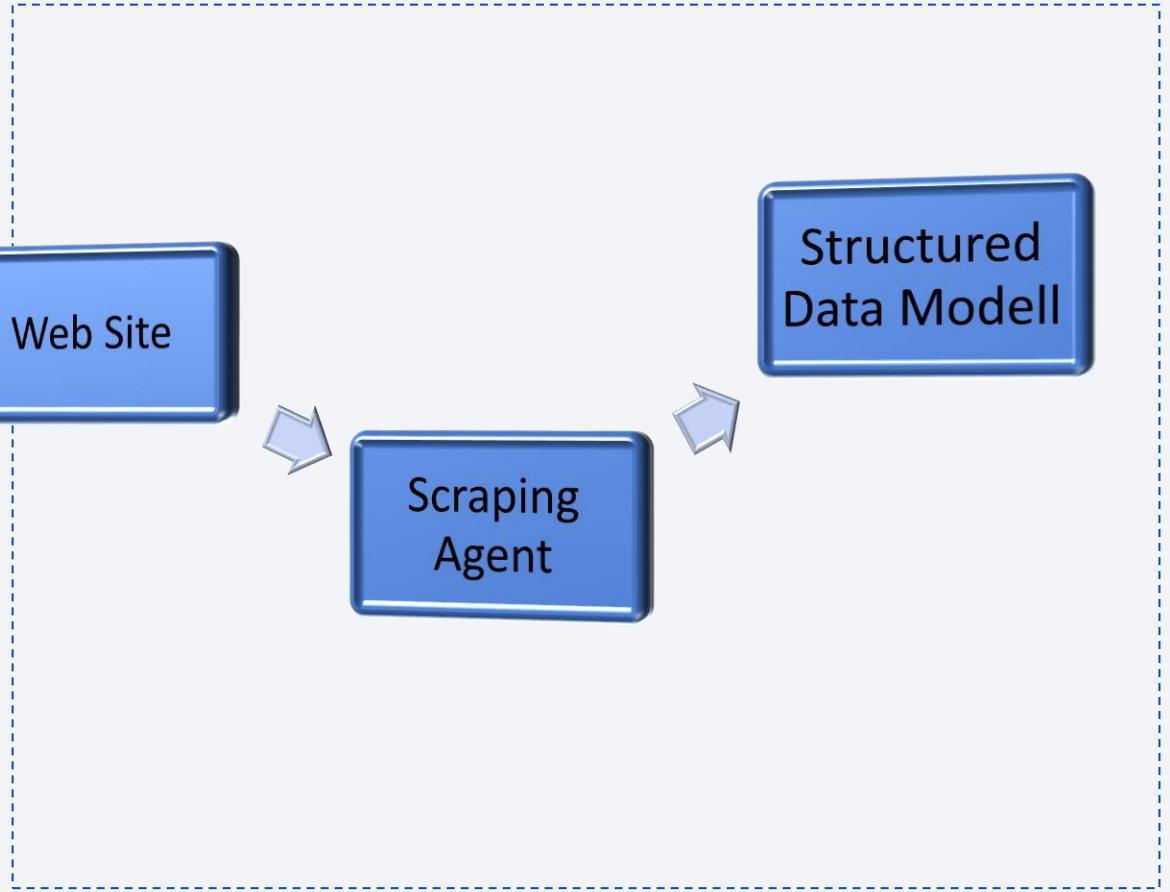
Data Collection – SpaceX API

- Client → Request → API → Server
- Server → Response → Service

- https://github.com/Flexenschnitzel/Projects/blob/main/Capstone_Project_1_API.ipynb

Data Collection - Scraping

- https://github.com/Flexenschnitzel/Projects/blob/main/Capatone_Project_3_webscraping.ipynb



Data Wrangling

- Missing values were replaced by the mean value of the column

- https://github.com/Flexenschnitzel/Projects/blob/main/Capstone_Project_3_Data%20wrangling.ipynb

EDA with Data Visualization

- Pie charts, Line charts, scatter plots
- https://github.com/Flexenschnitzel/Projects/blob/main/Capstone_Project_5_Data_Viz.ipynb

EDA with SQL

- Aggregate functions like GROUP BY and SUM() or MEAN()
- https://github.com/Flexenschnitzel/Projects/blob/main/Capstone_Project_4_SQL.ipynb

Build an Interactive Map with Folium

- Launch sites were marked and every past launch was flagged with a green a or red marker dependent of the success of the launch
- The distance from launch sites to proximities like coast, highways or railways was measured to see if there was connection
- https://github.com/Flexenschnitzel/Projects/blob/main/Capstone_Project_6_Folium.ipynb

Build a Dashboard with Plotly Dash

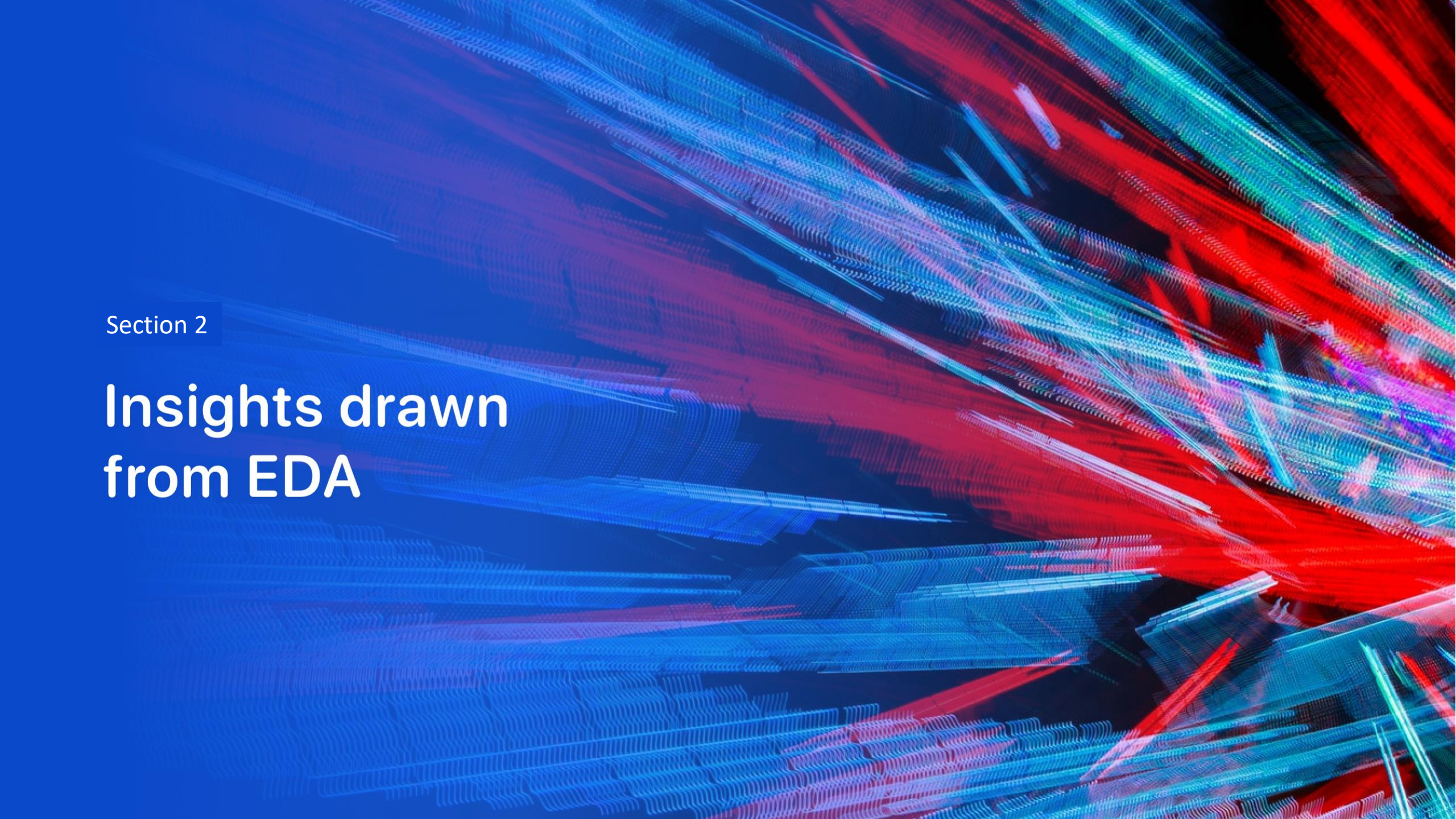
- Pie chart and scatter plot
- The pie chart is a good visualization of the percentages
- The scatterplot can be manipulated be a slider to change the payload range
- https://github.com/Flexenschnitzel/Projects/blob/main/Capstone_Project_7_dashboard.ipynb

Predictive Analysis (Classification)

- Cross validation (train/test set), hyperparameter tuning with Grid Search
- Different classification methods were used to compare prediction accuracy
- Logarithmic Regression, Support Vector Machines, Decision Trees K and nearest neighbor
- https://github.com/Flexenschnitzel/Projects/blob/main/Capstone_Project_9_Machinge_Learning.ipynb

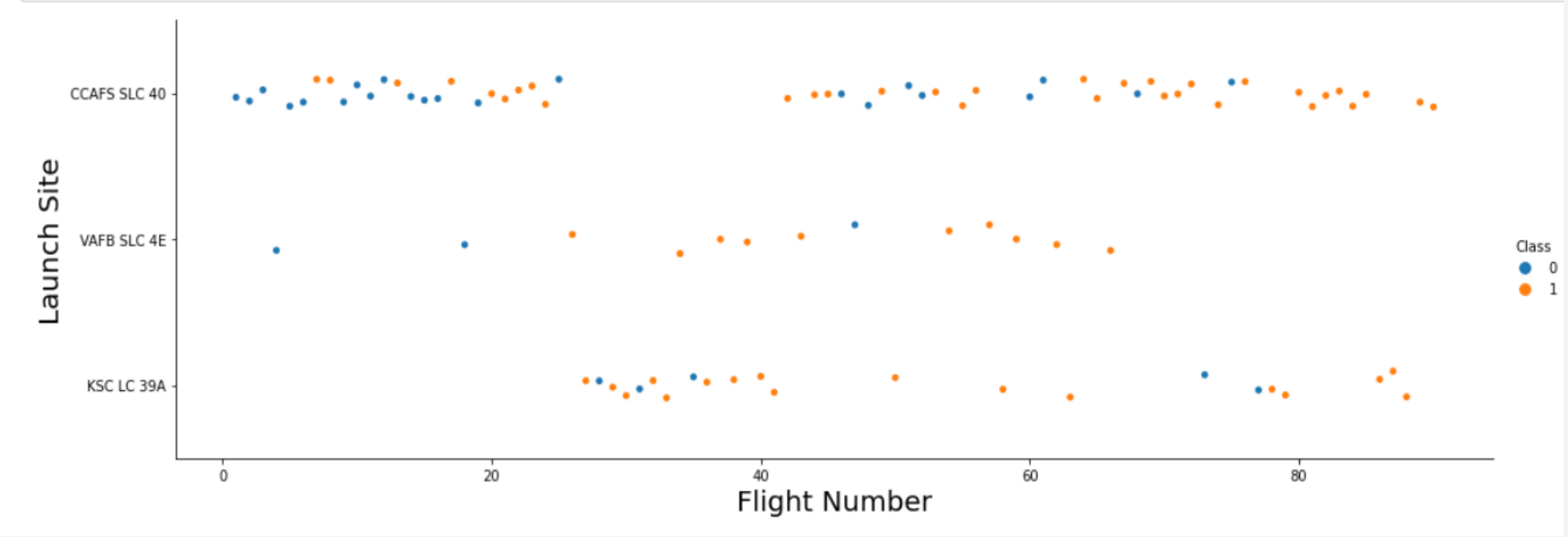
Results

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


Section 2
Insights drawn from EDA

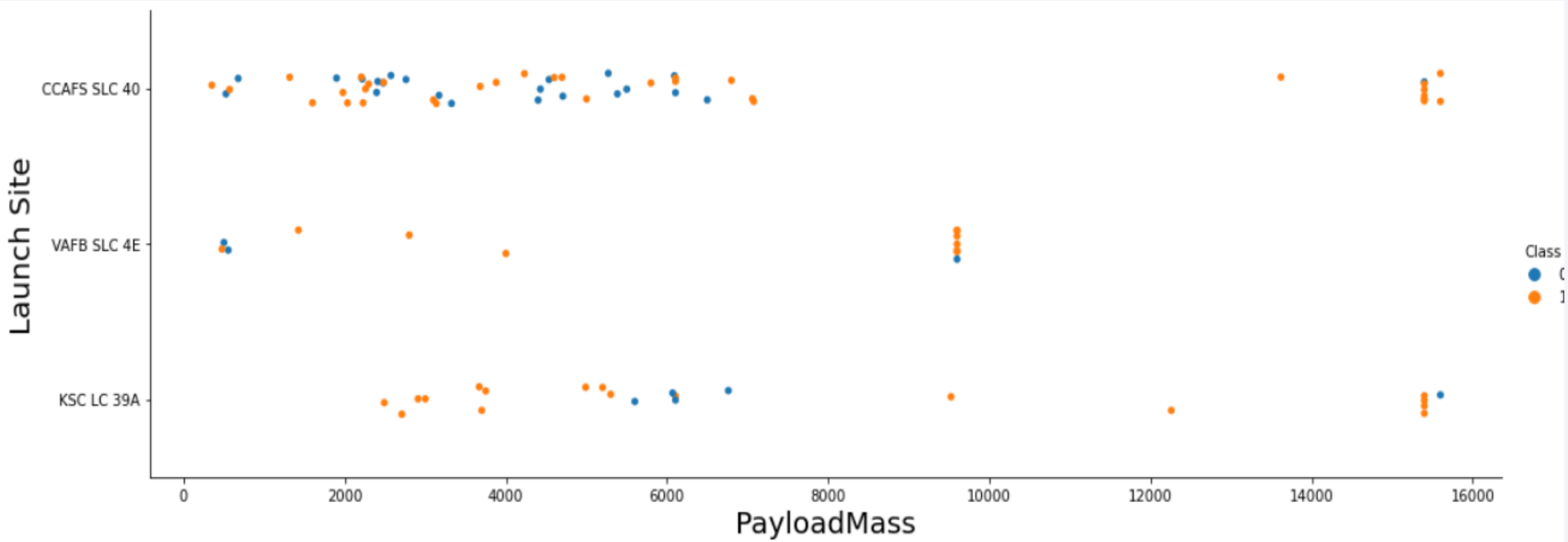
Flight Number vs. Launch Site

- All launch sites have a higher success rate given a higher flight number



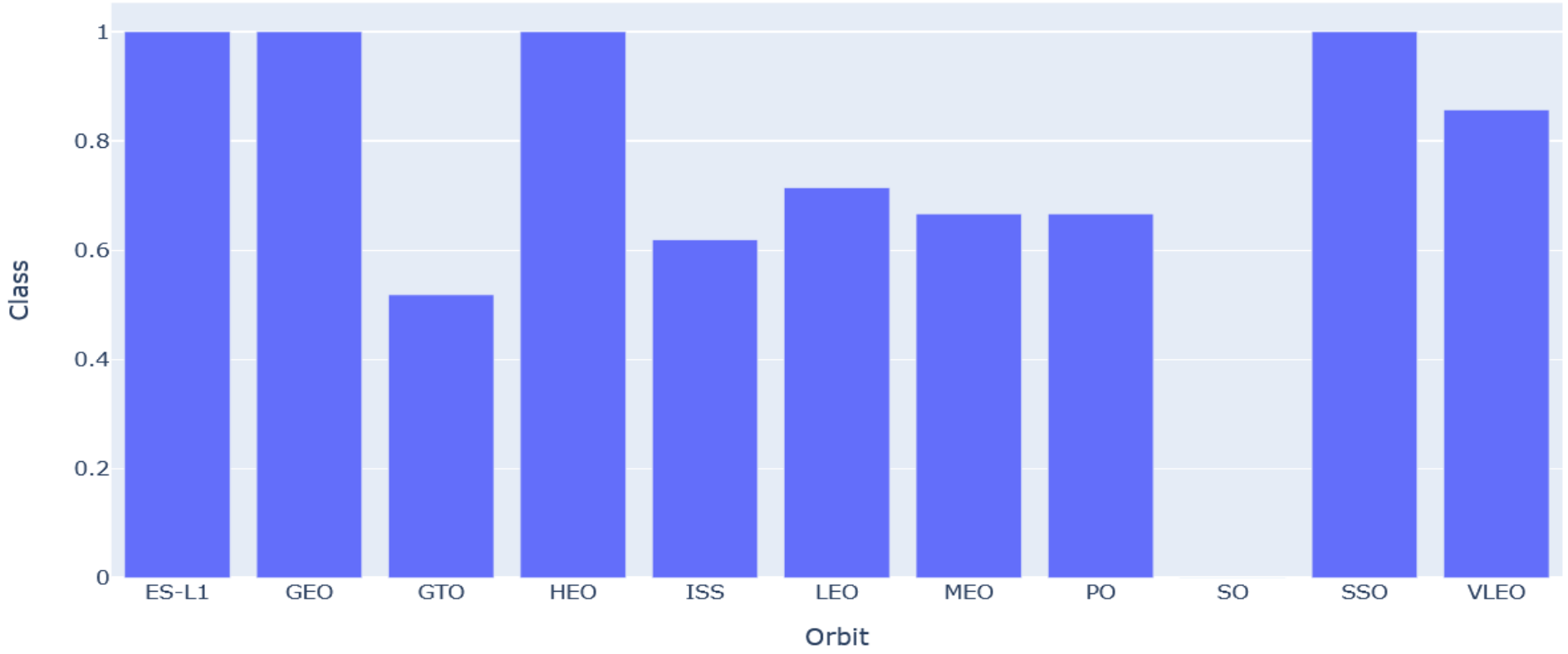
Payload vs. Launch Site

- Higher payloads are connected with successful launches



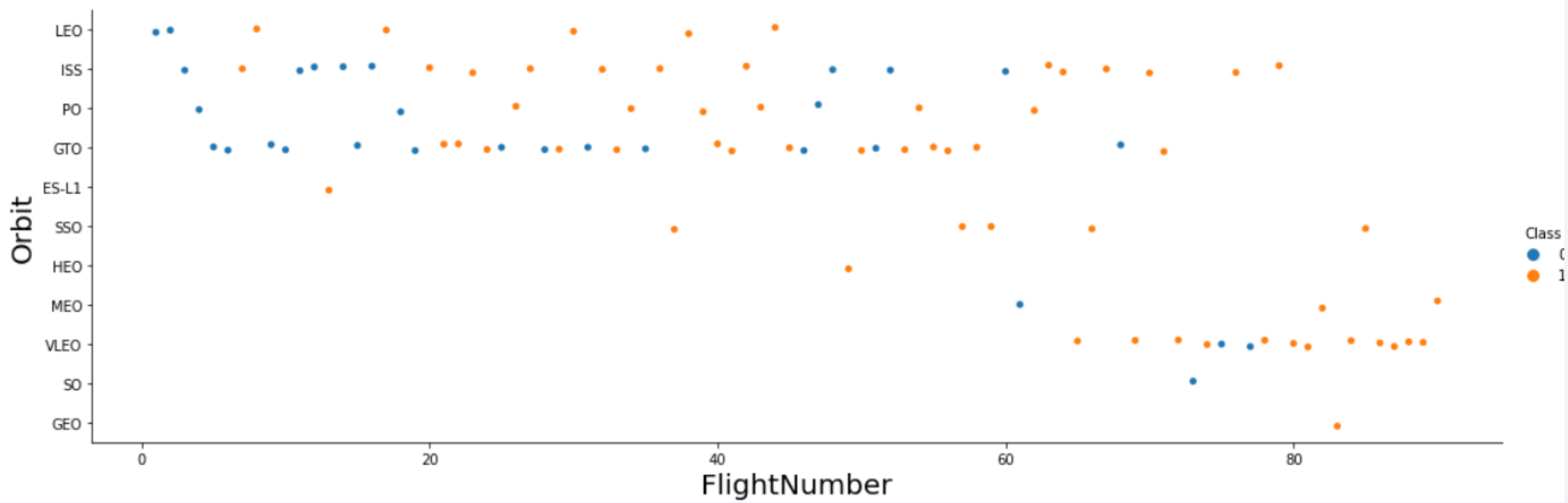
Success Rate vs. Orbit Type

- Success of launch is correlated to the respective orbit



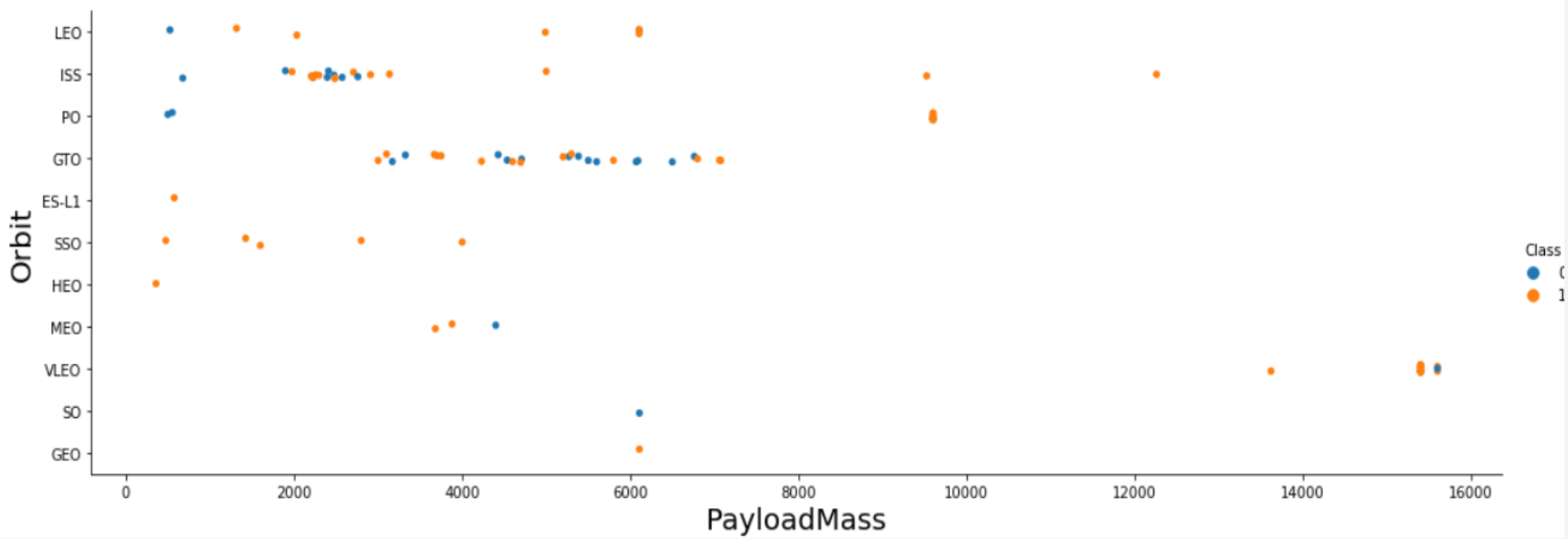
Flight Number vs. Orbit Type

- Flight number and successful launch are positively correlated for the orbit LEO and SSO for example



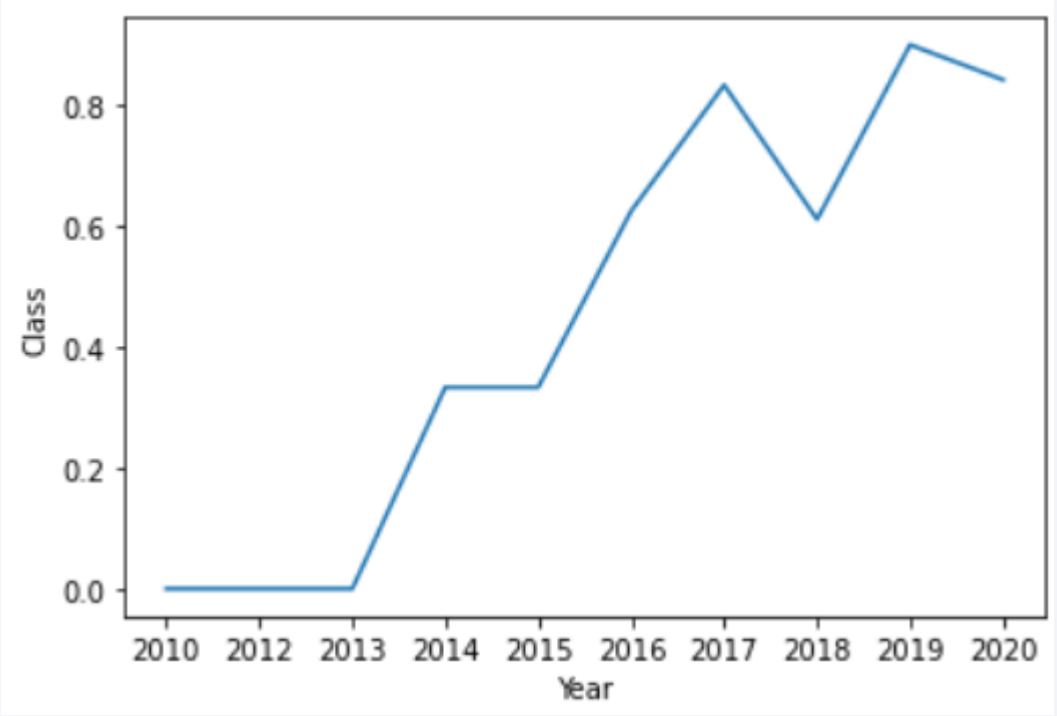
Payload vs. Orbit Type

- Heavy payloads and the successful landing rate are correlated more for Polar, LEO and ISS



Launch Success Yearly Trend

- The successful launch rate keeps increasing since 2013



All Launch Site Names

- We evaluated the data of 4 different launch sites

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

Launch Site Names Begin with 'CCA'

- 5 records where launch sites begin with `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

Total Payload Mass

- Total payload carried by boosters from NASA

```
%%sql
Select SUM(PAYLOAD_MASS__KG_) AS SUM_Payload_Mass_NASA_CRS from SPACEXTBL where "Customer" = "NASA (CRS)";
```

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

SUM_Payload_Mass_NASA_CRS
45596

Average Payload Mass by F9 v1.1

- That's the average payload the booster version F9 v.1.1 was carrying

AVG Payload Mass Booster F9 v1.1
2534.67

First Successful Ground Landing Date

- Date of the first successful landing outcome on ground pad

Date
22-12-2015

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

Total Number of Successful and Failure Mission Outcomes

- Total number of successful and failure mission outcomes

Total successful and failure missions
71

Boosters Carried Maximum Payload

- Names of the booster which have carried the maximum payload mass

Booster_Version maximum payload mass
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

2015 Launch Records

- Failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015

Month 2015	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 Landing Outcomes Between 2010-06-04 and 2017-03-20

- Landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

Landing _Outcome	Count
Success	20
Success (drone ship)	8
Success (ground pad)	6

Section 3

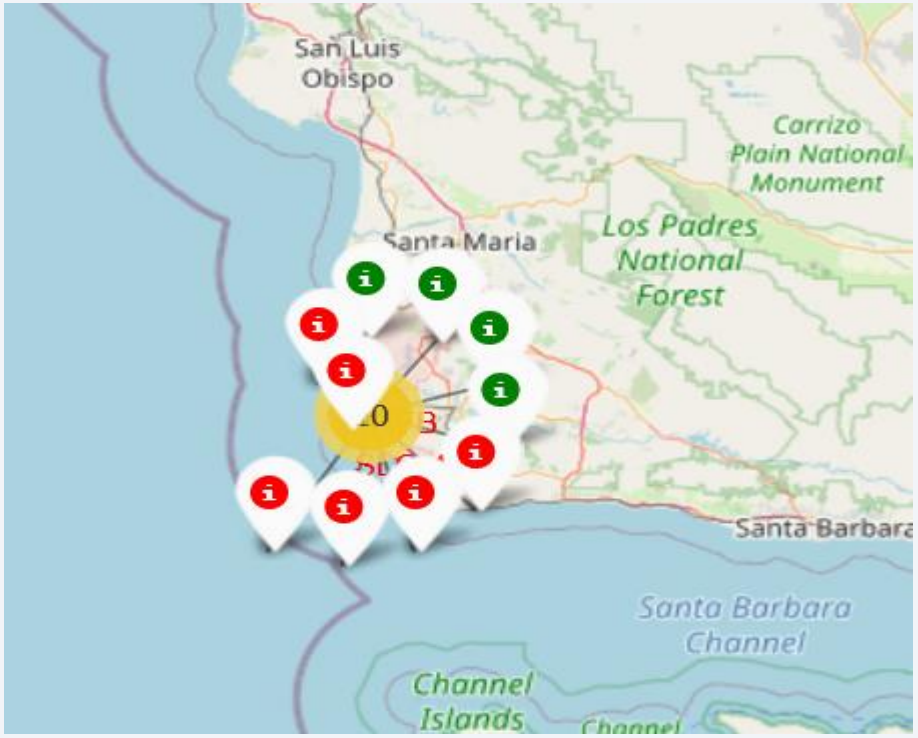
Launch Sites Proximities Analysis

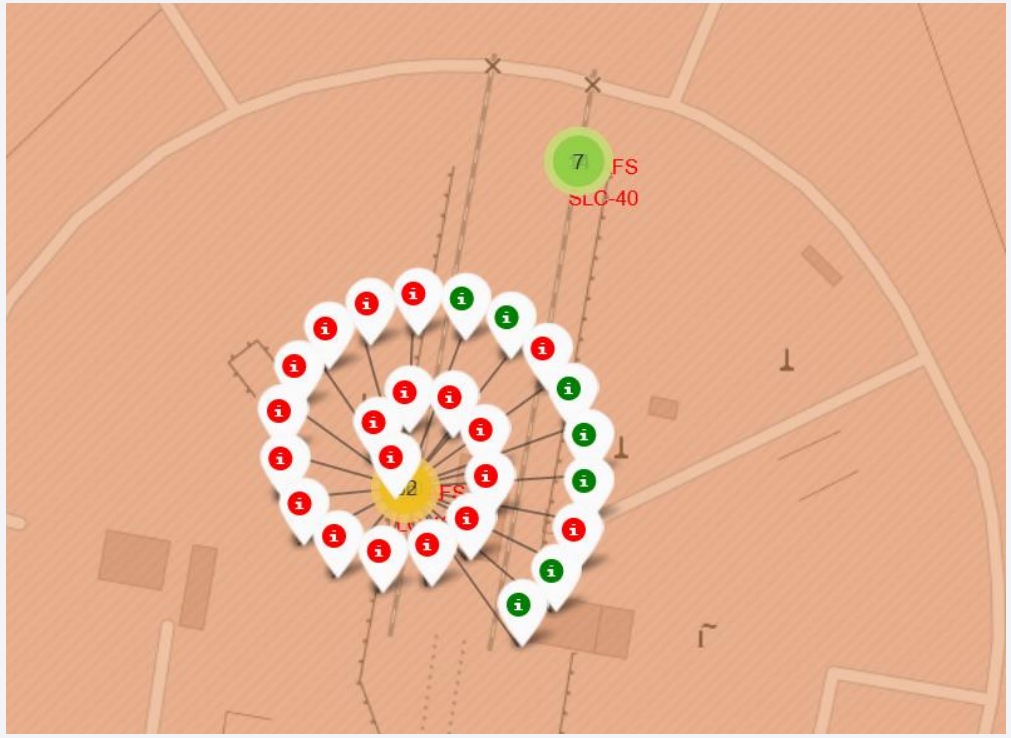
Location of Launching Sites

- All 4 launching sites are on the coast

Launch Markers

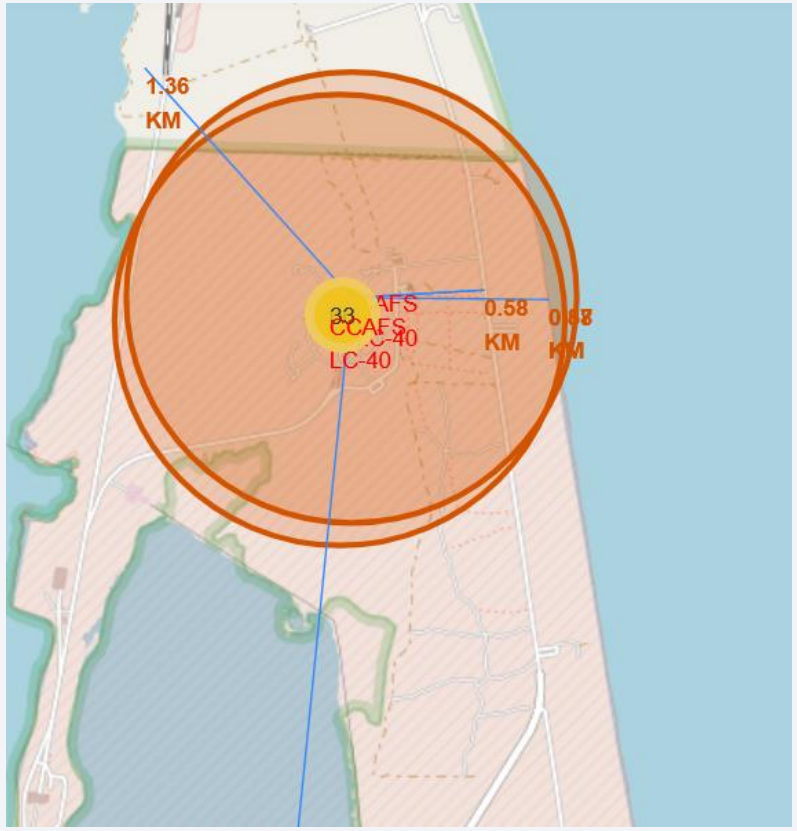
- Green markers → successful launches
- Red markers → failed launches

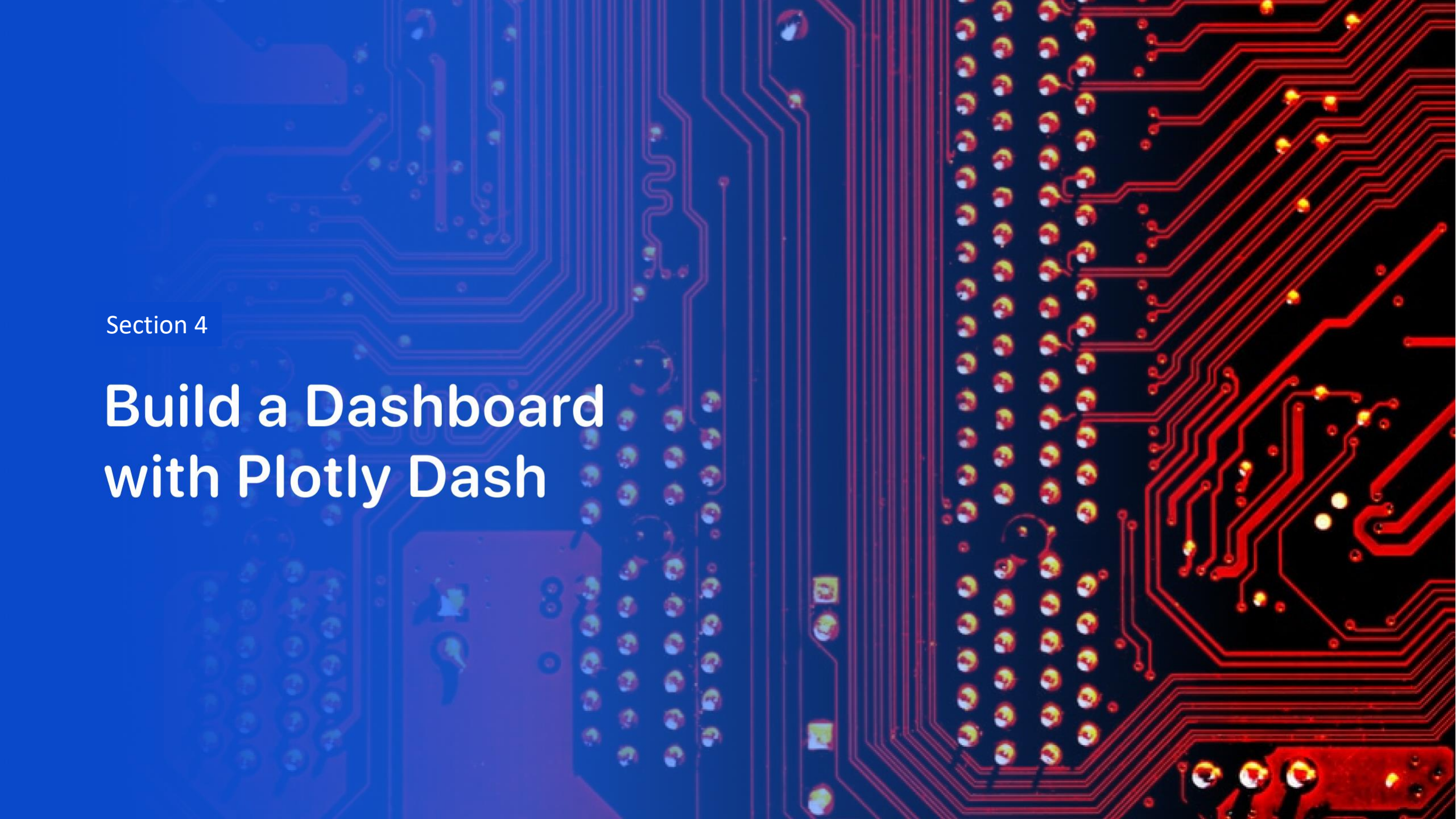




Launch Site Proximities

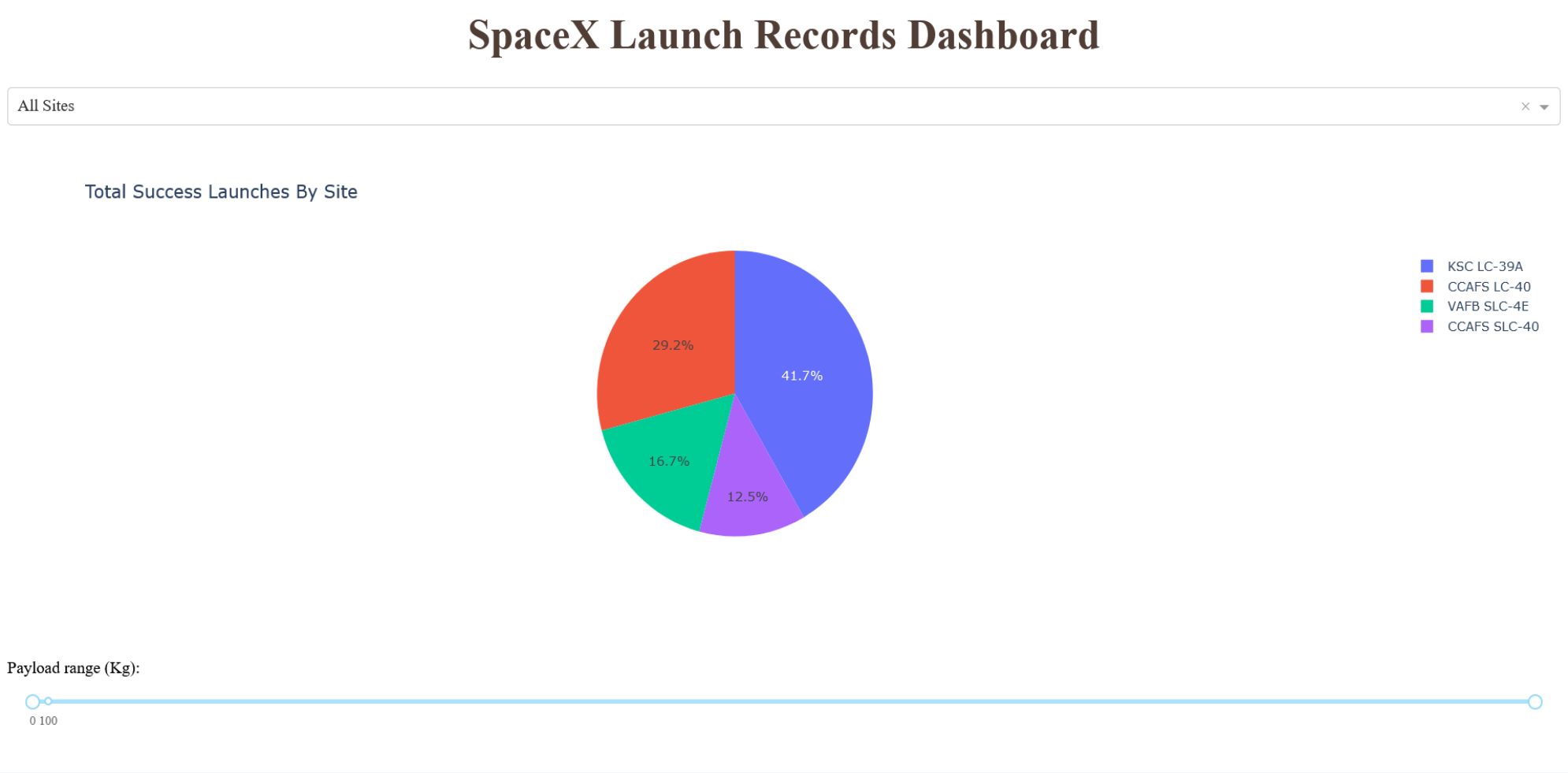
- Distance to coast, railroad and next city




Section 4
Build a Dashboard with Plotly Dash

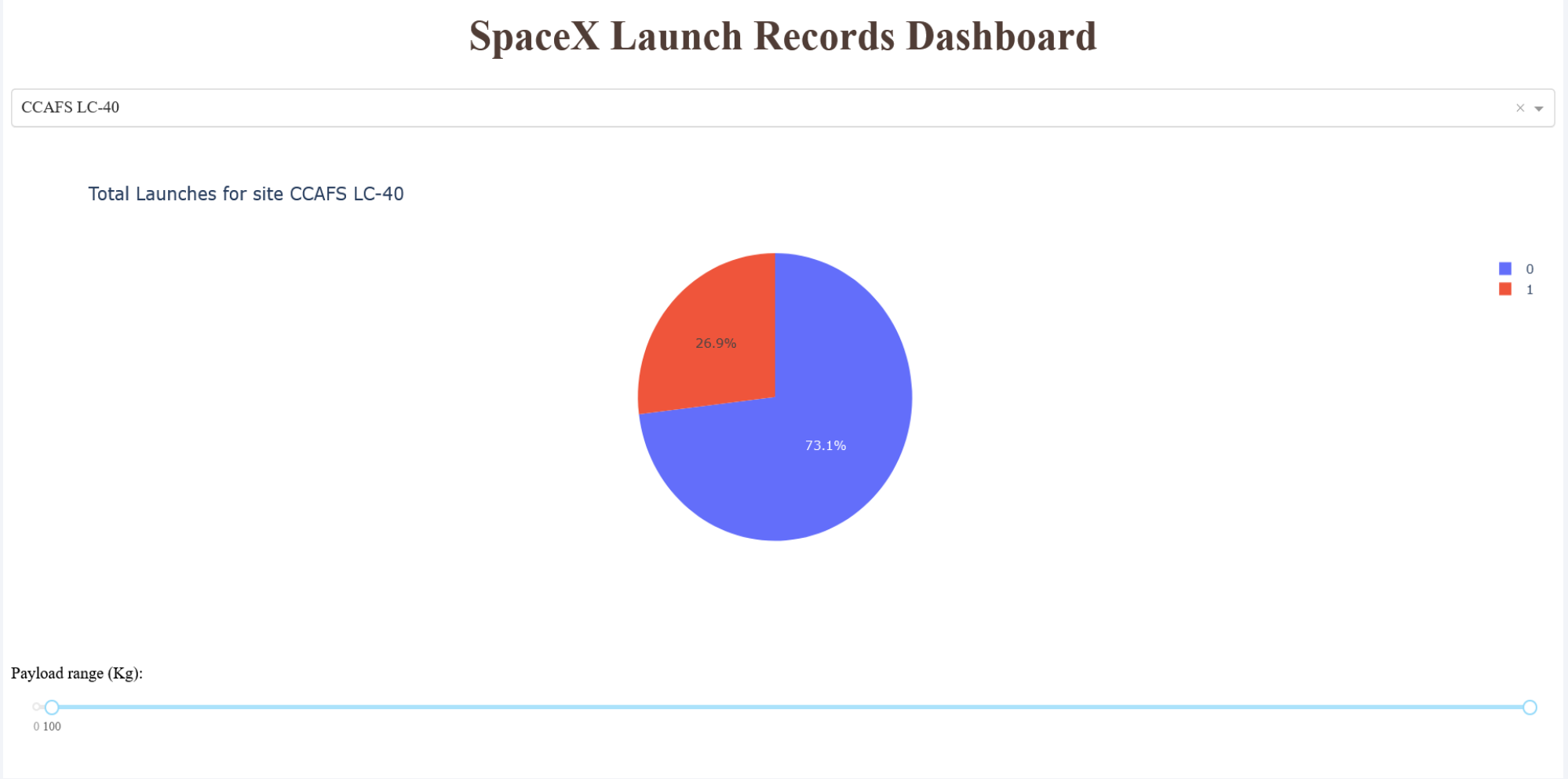
Successful Launches

- All 4 landing sites in percent



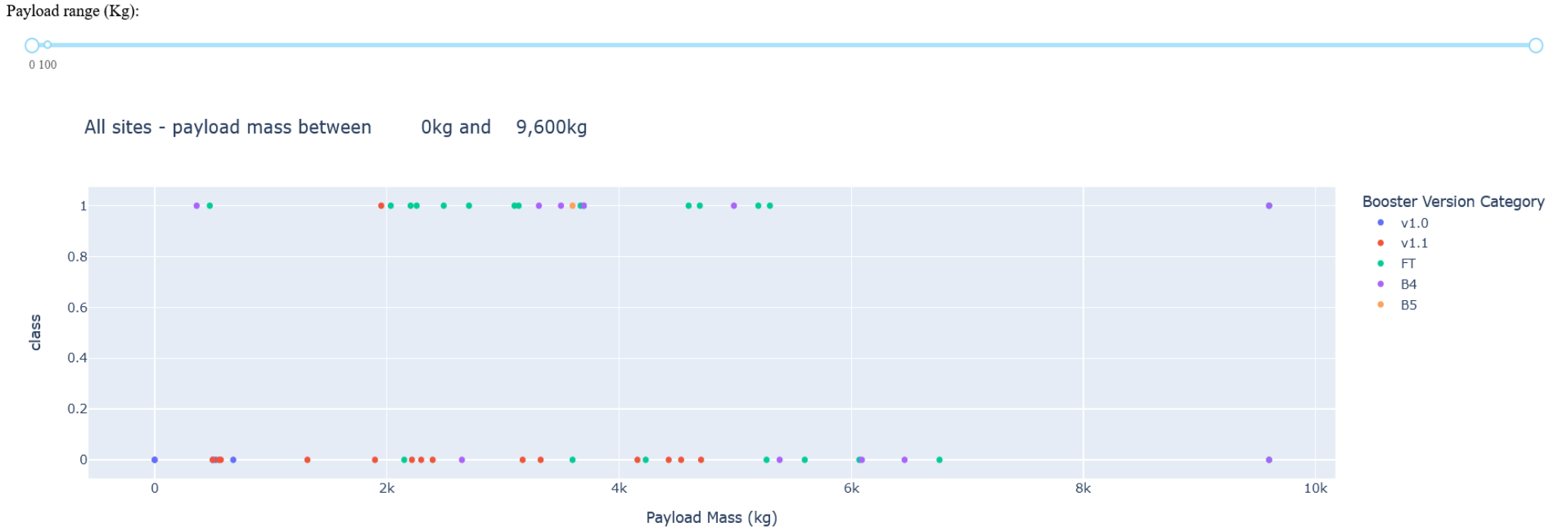
Highest successful Launch ratio

- Launch Site CCAFA LC-40 had the best successful launch percentage (73.1%)



Scatterplot of Launch Outcome and Payload

- Booster Version “FT” had the most successful launches

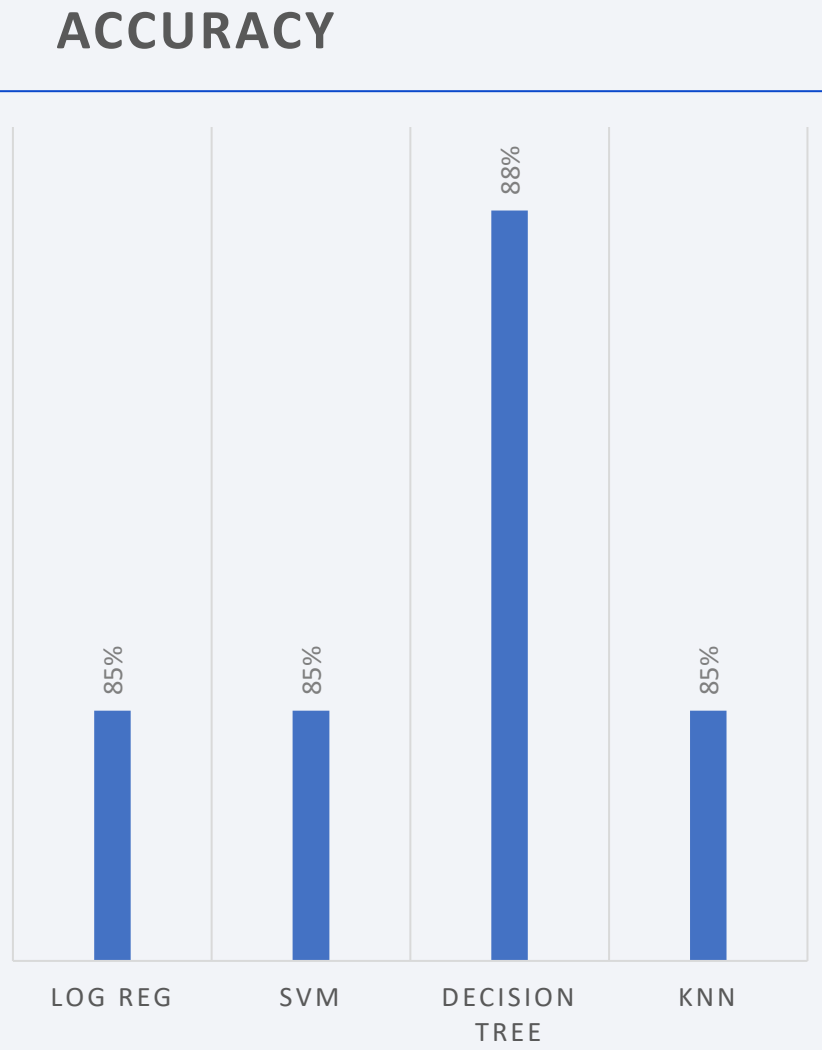


Section 5

Predictive Analysis (Classification)

Classification Accuracy

- The best method on the training data is the decision tree classifier (accuracy = 0.88)
- The accuracy with the 'score' method on the test data is identical for every algorithm (accuracy=0.83)



Confusion Matrix

- The confusion matrix looks identical for all 4 models
- The False Positive categorization seems to be problematic

Conclusions

- The are various factors that influence the landing outcome:
- Orbit
- Launch Site
- Booster Version
- Payload Mass
- Our best prediction model has an accuracy of 88% on the training and 83% on the test set

Appendix

- Thanks a lot!

Thank you!