IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Name>
<Date>

Outline

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

Executive Summary

- The following methodologies were used to analyse data
    1. Data Collection using web scraping and SpaceX API;
    2. Exploratory Data Analysis (EDA), including data wrangling, data visualization and interactive visual analytics;
    3. Machine Learning Prediction
- Results:
    1. It was possible to collect valuable data from public sources;
    2. EDA allowed to identify which features are the best to predict success of launchings;
    3. Machine Learning Prediction shows the best model to predict which characteristics are important to drive this opportunity by the best way, using all collected data.

Introduction

- The objective is to evaluate the viability of the new company Space Y to compete with Space X
- Answers:
    1. The best way to estimate the total cost for launches, by predicting successful landings of the first stage of rockets;
    2. Where is the best place to make launches.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data from Space X was obtained from 2 sources:
    Space X API (https://api.spacexdata.com/v4/rockets/)
    WebScraping (https://en.Wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches)
- Perform data wrangling
  - Collected data was enriched by creating a landing outcome label based on outcome data after summarizing and analyzing features
- Perform exploratory data analysis (EDA) using visualization and SQL

Methodology

Executive Summary

- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Data that was collected until this step were normalized, divided in training and test data sets and evaluated by four different classification models, being the accuracy of each model evaluated using different combinations of parameters.

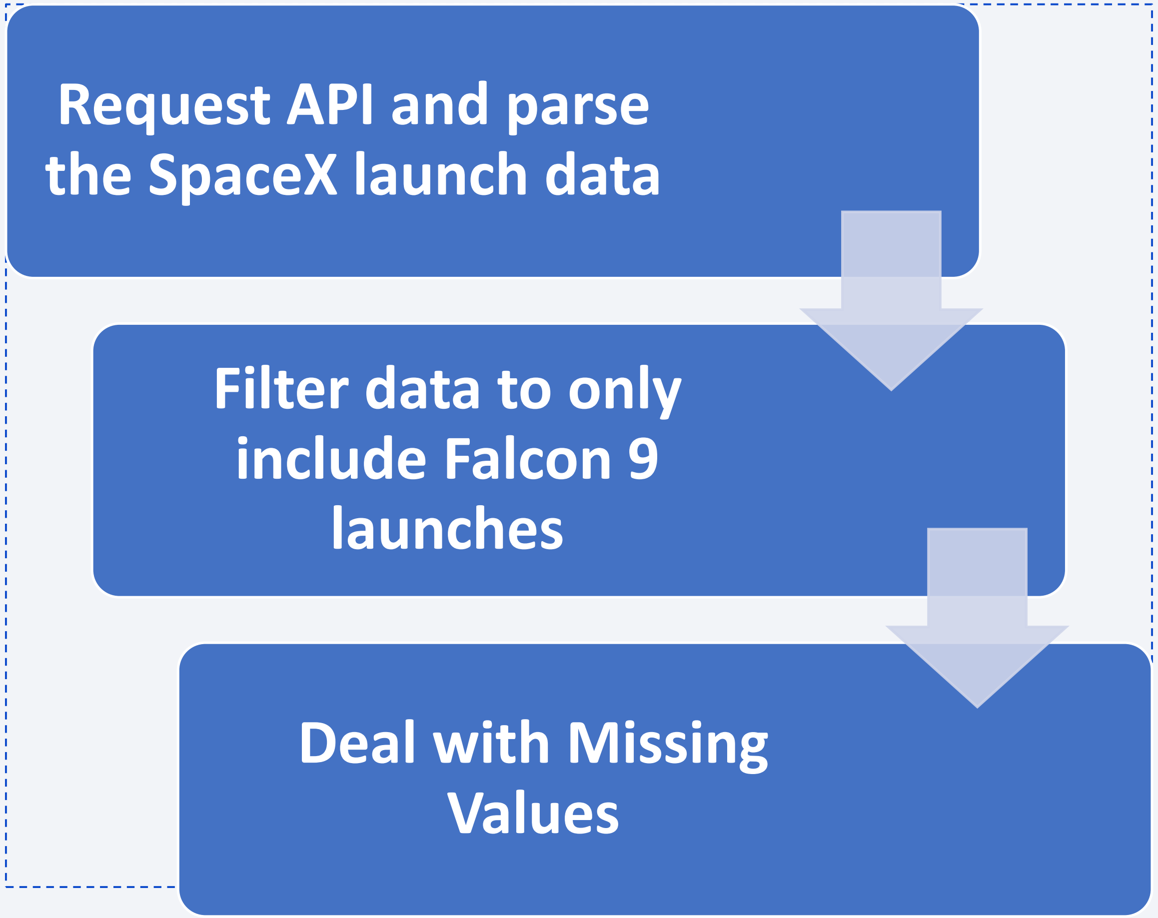
Data Collection

- Data sets were collected from Space X API (https://api.spacexdata.com/v4/rockets/) and from Wikipedia (https://en.Wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches), using web scraping techniques.

Data Collection – SpaceX API

- SpaceX offers a public API from where data can be obtained and then used;
- This API was used according to the flowchart beside and then data is persisted.

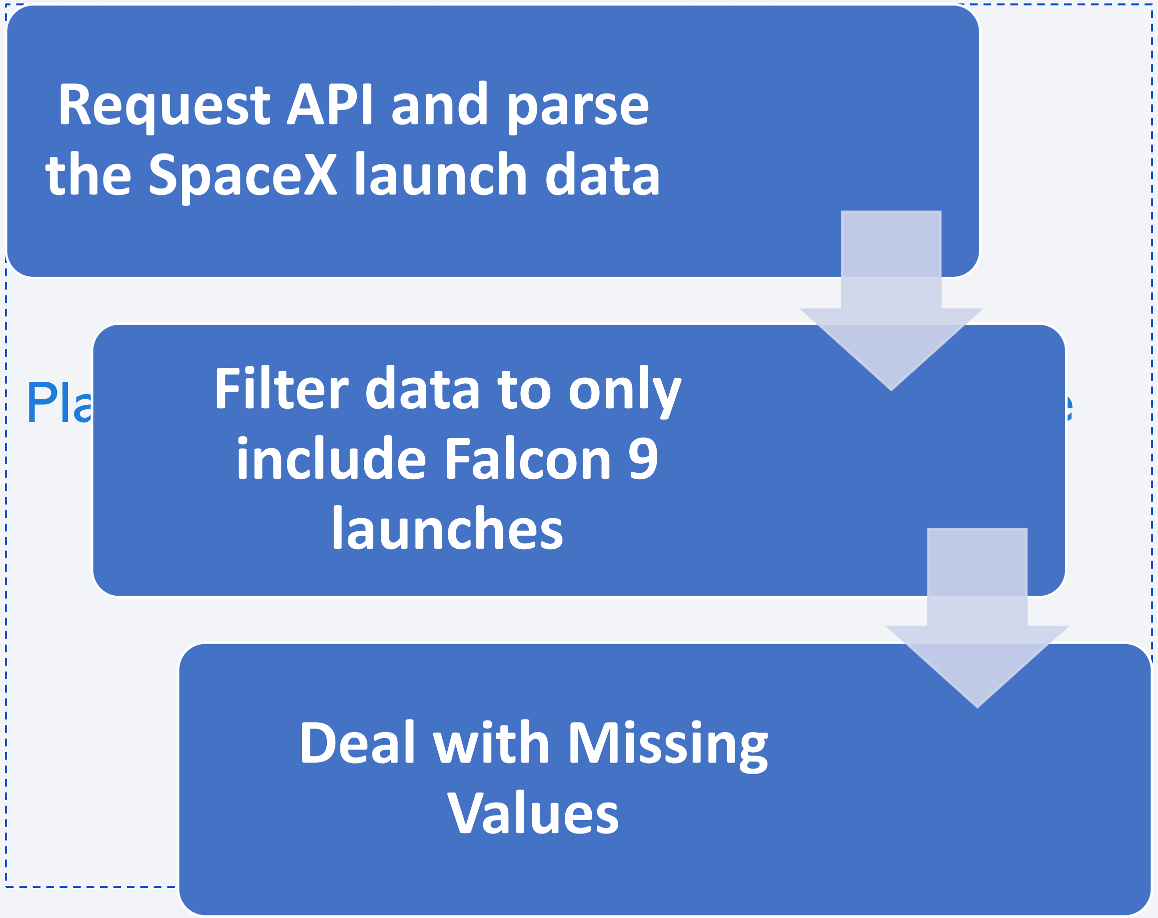
- Source code:[Data-Science-Capstone/jupyter-labs-spacex-data-collection-api.ipynb at main · weiliang1995/Data-Science-Capstone (github.com)]



Data Collection - Scraping

- Data from SpaceX launches can also be obtained from Wikipedia;
- Data are downloaded from Wikipedia according to the flowchart and then persisted.

- Source code:[Data-Science-Capstone/jupyter-labs-webscraping.ipynb at main · weiliang1995/Data-Science-Capstone (github.com)]



Data Wrangling

- Initially some Exploratory Data Analysis (EDA) was performed on the dataset.
- Then the summaries launches per site, occurences of each orbit and occurrences of mission outcome per orbit type were calculated.
- Finally, the landing outcome label was created from Outcome column.

- Source code:[Data-Science-Capstone/labs-jupyter-spacex-Data wrangling.ipynb at main · weiliang1995/Data-Science-Capstone (github.com)]

EDA with Data Visualization

- To explore data, scatterplots and barplots were used to visualize the relationship between pair of features:

    1. Payload Mass x Flight Number
    2. Launch Site x Flight Number
    3. Launch Site x Payload Mass
    4. Orbit and Flight Number
    5. Payload and Orbit

Source code:[Data-Science-Capstone/jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb at main · weiliang1995/Data-Science-Capstone (github.com)]

EDA with SQL

- The following SQL queries were performed:
    - Names of the unique launch sites in the space mission
    - Top 5 launch sites whose name begin with the string ‘CCA’
    - Total payload mass carried by boosters launched by NASA (CRS)
    - Average payload mass carried by booster version F9 v1.1
    - Date when the first successful landing outcome in ground pad was achieved
    - Names of the boosters which have success in drone ship and have payload mass by 4000 and 6000kg
    - Total number of successful and failure mission outcomes
    - Names of the booster versions which have carried the maximum payload mass
    - Failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015
    - Rank of the count of landing outcomes (such as Failure (drone ship) or Success(group pad)) between the date 2010-06-04 and 2017-03-20.

Source code:[Data-Science-Capstone/jupyter-labs-eda-sql-coursera_sqllite.ipynb at main · weiliang1995/Data-Science-Capstone (github.com)]

Build an Interactive Map with Folium

- Marker, circles, lines and marker clusters were used with Folium Maps
  - Markers indicate points like launch sites
  - Circles indicate highlighted areas around specific coordinates, like NASA Johnson Space Center
  - Marker clusters indicates groups of events in each coordinate, like launches in a launch site
  - Lines are used to indicate distances between two coordinates

Source code:Data-Science-Capstone/lab_jupyter_launch_site_location.jupyterlite.ipynb at main · weiliang1995/Data-Science-Capstone (github.com)

Build a Dashboard with Plotly Dash

- The following graphs and plots were used to visualize data
  - Percentage of launches by site
  - Payload range
- This combination allowed to quickly analyze the relation between payloads and launch sites, helping to identify where is best place to launch according to payloads.

Source code: [Data-Science-Capstone/spacex_dash_app.py at main · weiliang1995/Data-Science-Capstone (github.com)]

Predictive Analysis (Classification)

- Four classification models were compared:

1. Logistic Regression
2. Support Vector Machine (SVM)
3. Decision Tree
4. K-nearest Neighbors

Source code: Data-Science-Capstone/SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb at main · weiliang1995/Data-Science-Capstone (github.com)

Results

- Exploratory data analysis results:
  1. Space X uses 4 different launch sites
  2. The first launches were done to Space X itself and NASA
  3. The average payload of F9 v1.1 booster is 2928 kg
  4. The first success landing outcome happened in 2015
  5. Many Falcon 9 booster versions were successful at landing in drone ships having payload above the average
  6. Almost 100% of mission outcomes were successful
  7. Two booster versions failed at landing in drone ships in 2015: F9 v1.1 B1012 and F9 v1.1 B1015
  8. The number of landing outcomes became as better as years passed

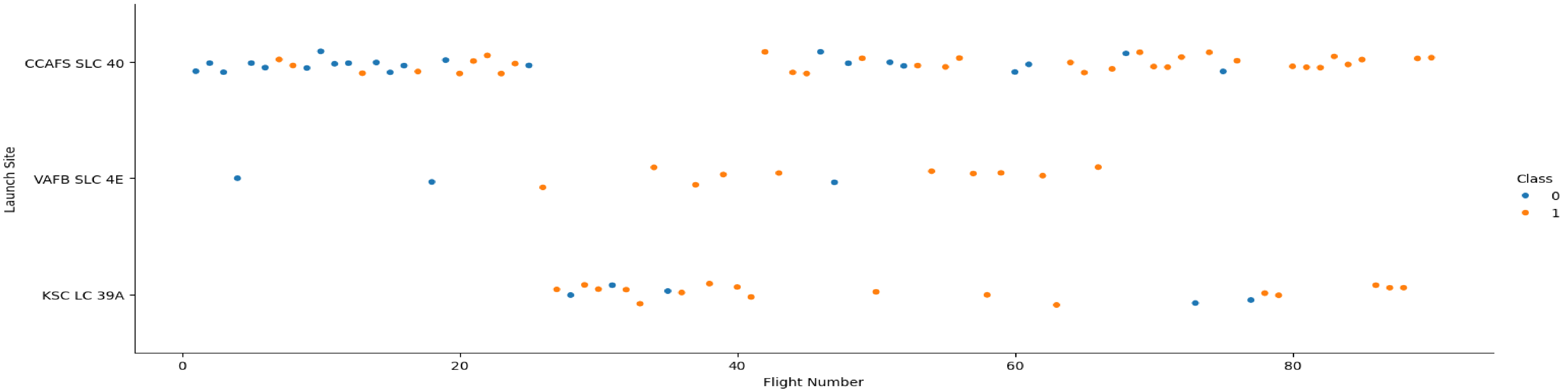
Results

- Using interactive analytics was possible to identify that launch sites use to be in safety places, near sea, for example and have a good logistic infrastructure around.
- Most launches happens at east coast launch sites.
- Predictive Analysis shows that Decision Tree Classifier is the best model to predict successful landings, having accuracy over 87% and accuracy for test data over 94%.

Section 2

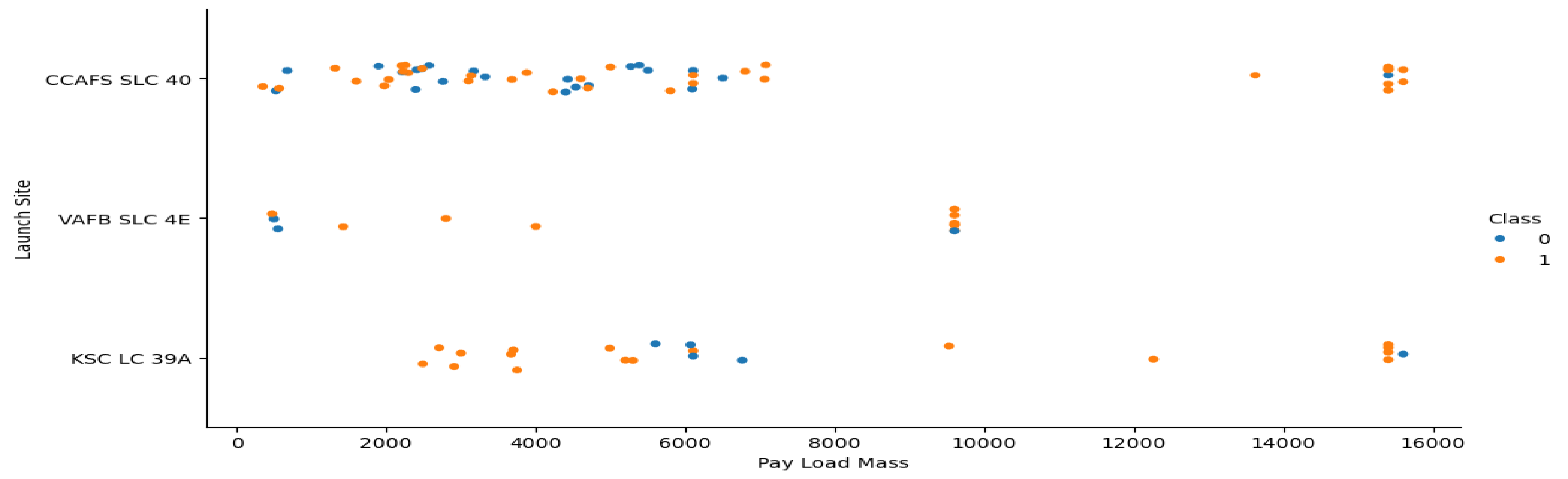
Insights drawn from EDA

Flight Number vs. Launch Site



- According to the plot above, it's possible to verify that the best launch site is CCAF5 SLC 40, where most of recent launches were successful, and followed by VAFB SLC 4E and KSC LC 39A.
- It's also possible to see that success rate improved over time.

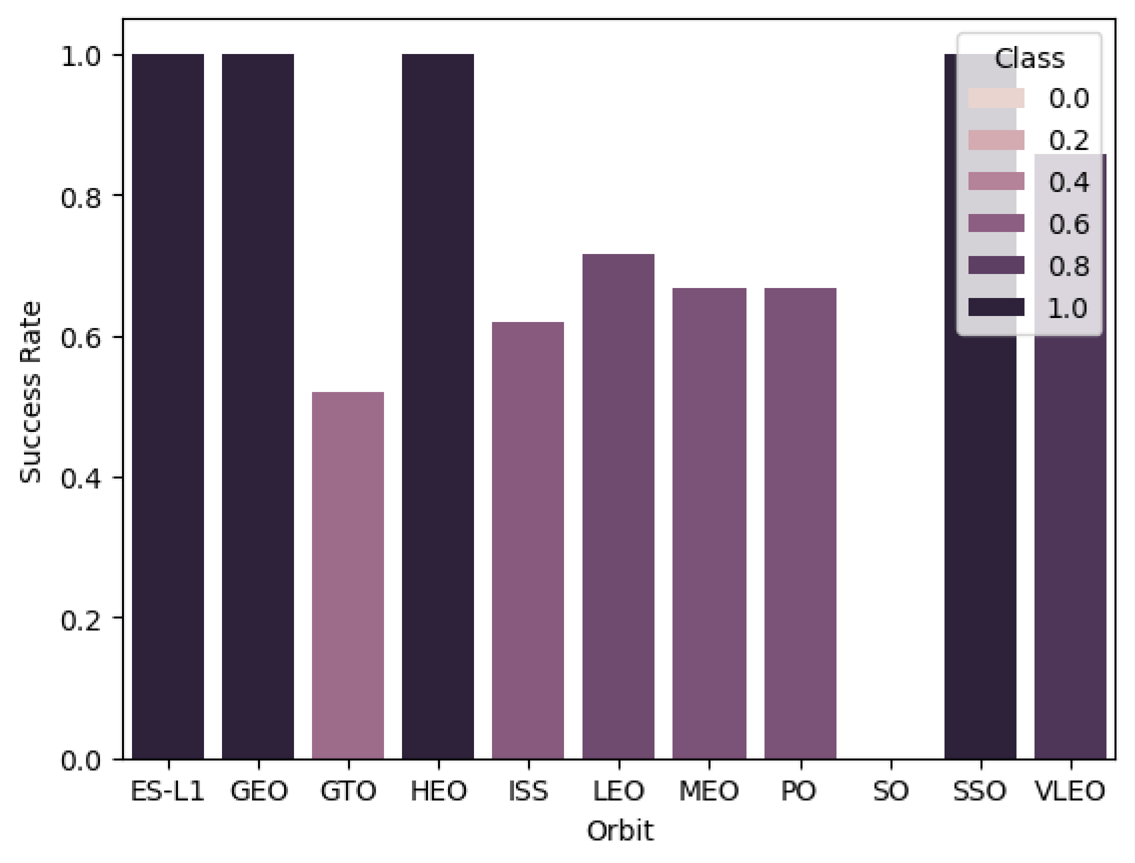
Payload vs. Launch Site



- Payloads over 9000kg (about the weight of a school bus) have excellent success rate
- Payloads over 10000kg seems impossible on VAFB SLC 4E launch site.

Success Rate vs. Orbit Type

- The biggest success rates happens to orbits:
  - ES-L1
  - GEO
  - HEO
  - SSO
- Followed by:
  - VLEO (above 80%)
  - LFO (above 70%)

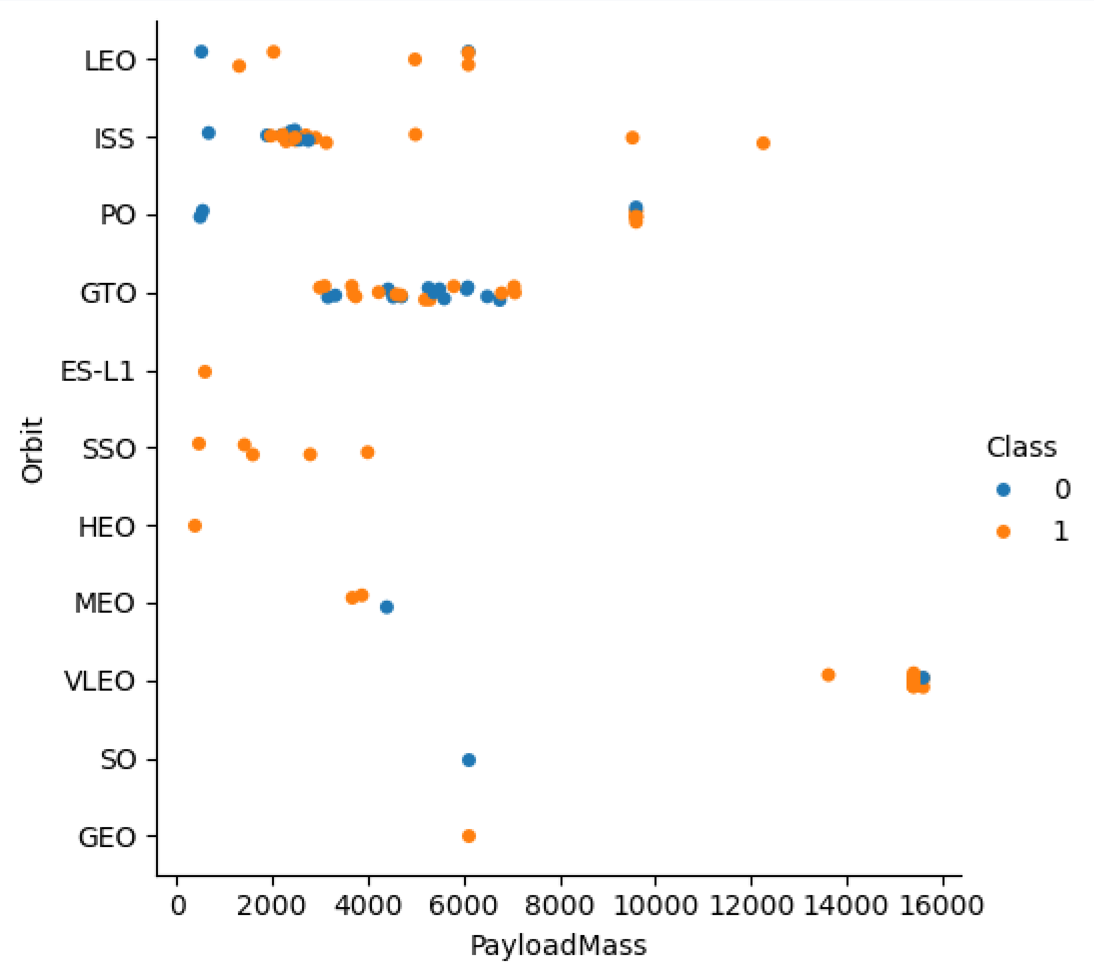


Flight Number vs. Orbit Type

- Success rate improved over time to all orbits
- The number of flight in VLEO increases a lot recently.

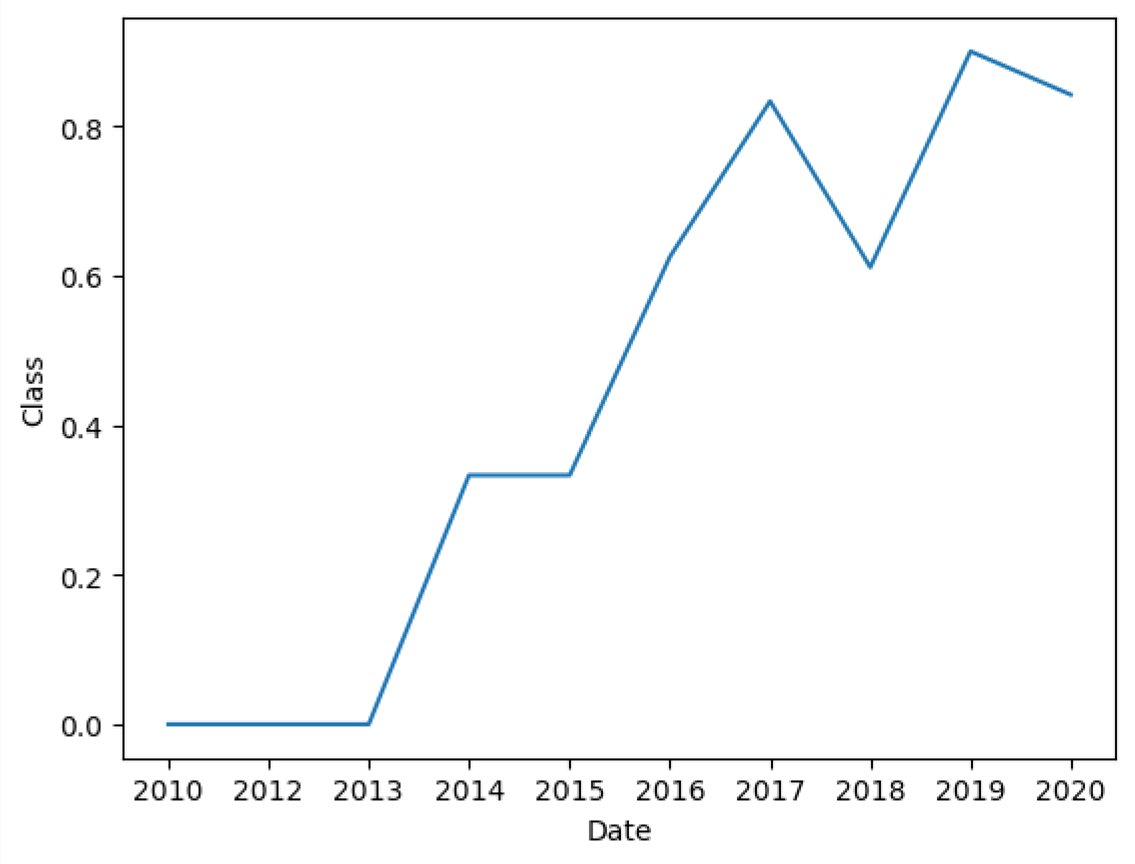
Payload vs. Orbit Type

- There are only a few launches over 10000kg, mainly on VLEO orbit, and a few on ISS and PO.
- The success rate in GTO seems unrelated with its pay load mass.



Launch Success Yearly Trend

- Success rate started increasing in 2013 and kept until 2020.
- It seems that the first three years were a period of adjustment and improvement of its technology.



All Launch Site Names

```
%sql select distinct launch_site from spacextable
```

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

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

Launch Site Names Begin with 'CCA'

In [16]:

```
%sql select * from spacextable where launch_site LIKE 'CCA%' limit 5
```

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

Out[16]:

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

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

In [23]:
```
%sql select sum(PAYLOAD_MASS__KG_) from spacextable where customer like 'NASA%(CRS)'
```

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

Out[23]:

sum(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

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

In [24]:
```
%sql select avg(PAYLOAD_MASS__KG_) from spacextable where booster_version like 'F9 v1.1%'
```

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

Out[24]:

avg(PAYLOAD_MASS__KG_)
2534.6666666666665

First Successful Ground Landing Date

```
In [29]: %sql select min(date) from spacextable where landing_outcome like 'Success%(ground pad)'

 * sqlite:///my_data1.db
Done.

Out[29]:
```

min(date)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

In [31]:

```
%sql select distinct booster_version from spacextable where landing_outcome like 'Success%(drone ship)' and PAYLOAD_MASS__KG
```

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

Out[31]:

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

Total Number of Successful and Failure Mission Outcomes

```
%sql select mission_outcome, count(mission_outcome) from spacextable group by mission_outcome
```

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

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

Boosters Carried Maximum Payload

```
%sql select booster_version,payload_mass__kg_ from spacextable where payload_mass__kg_ is (select max(payload_mass__kg_) fro
```

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

Booster_Version	PAYLOAD_MASS__KG_
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

2015 Launch Records

```
%sql select substr(Date,6,2) as month,booster_version,launch_site from spacextable where landing_outcome l
```

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

month	Booster_Version	Launch_Site
01	F9 v1.1 B1012	CCAFS LC-40
04	F9 v1.1 B1015	CCAFS LC-40

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

Landing Outcome	Occurrences
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

Section 3

Launch Sites Proximities Analysis

Launch Sites

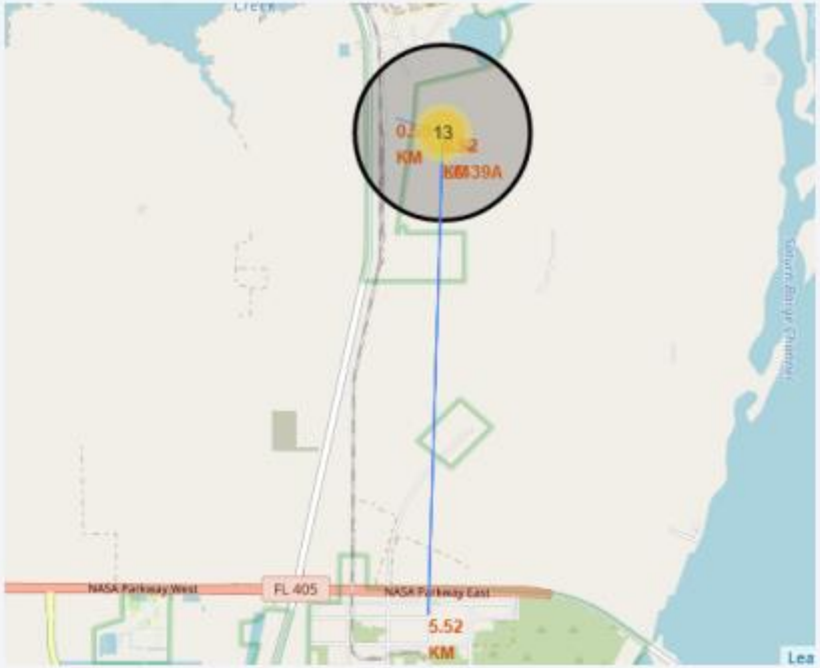
- Launch sites are near sea, probably by safety, but not too far from roads and railroads.

Launch Outcomes by Site

- Example of KSC LC-39A launch site launch outcomes

- Green markers indicate successful and red markers indicate failure.

Logistics and Safety

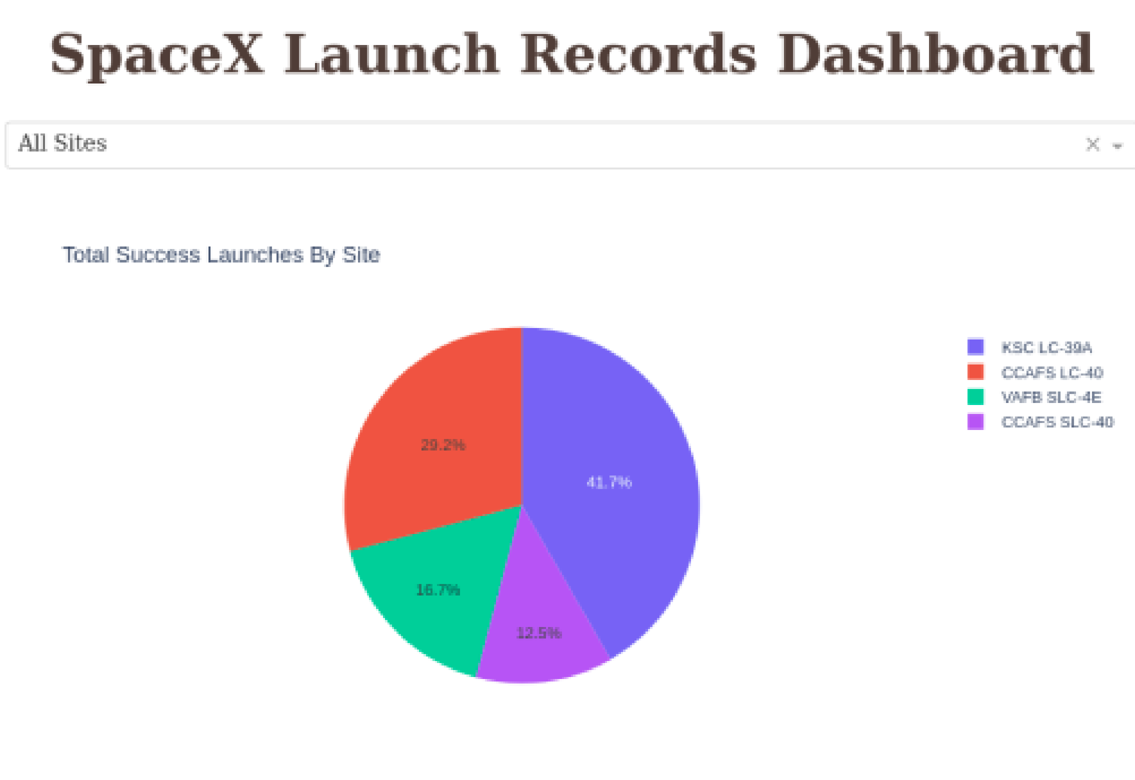


- Launch site KSC LC-39A has good logistics aspects, being near railroad and road and relatively far from inhabited areas.

Section 4

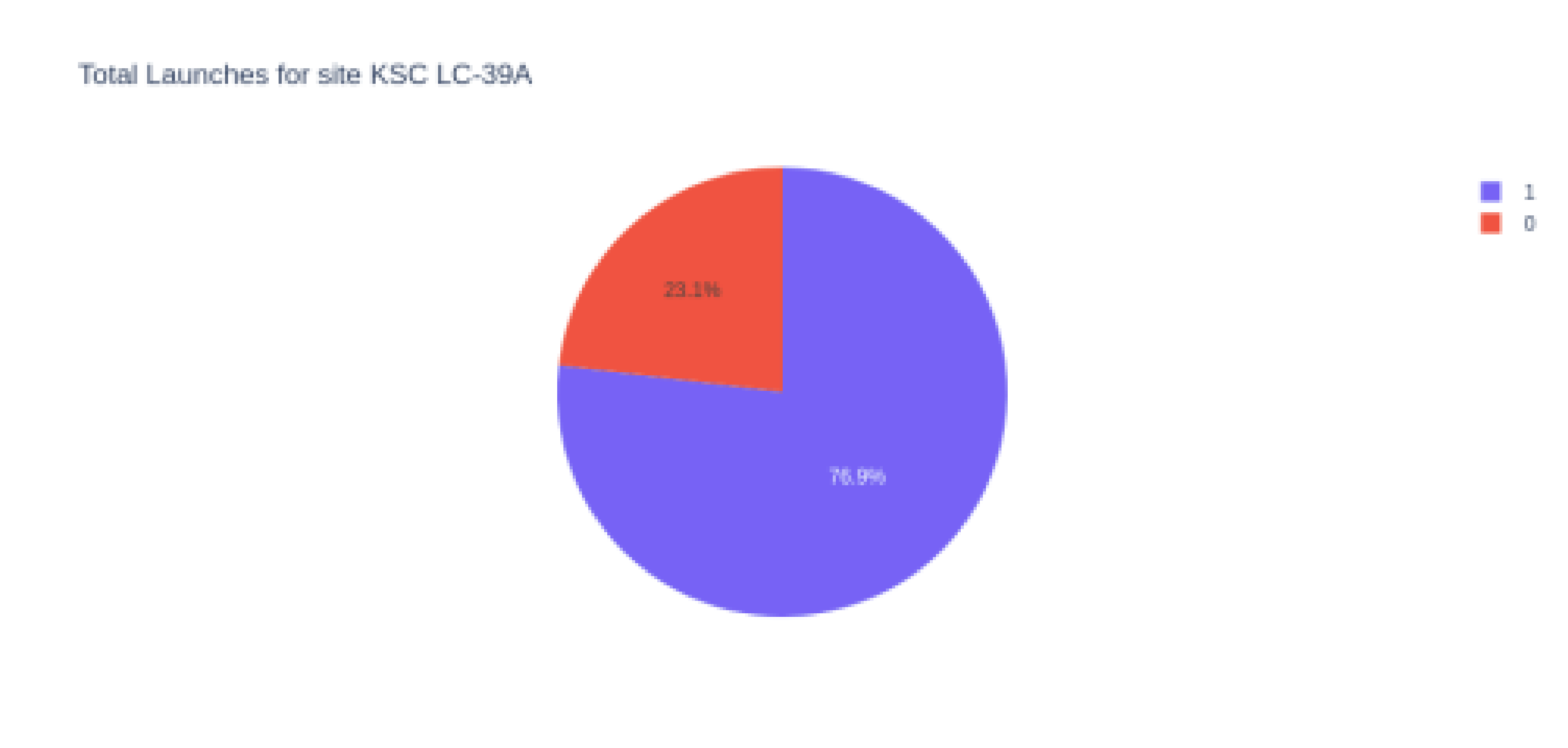
Build a Dashboard with Plotly Dash

Successful Launches by Site



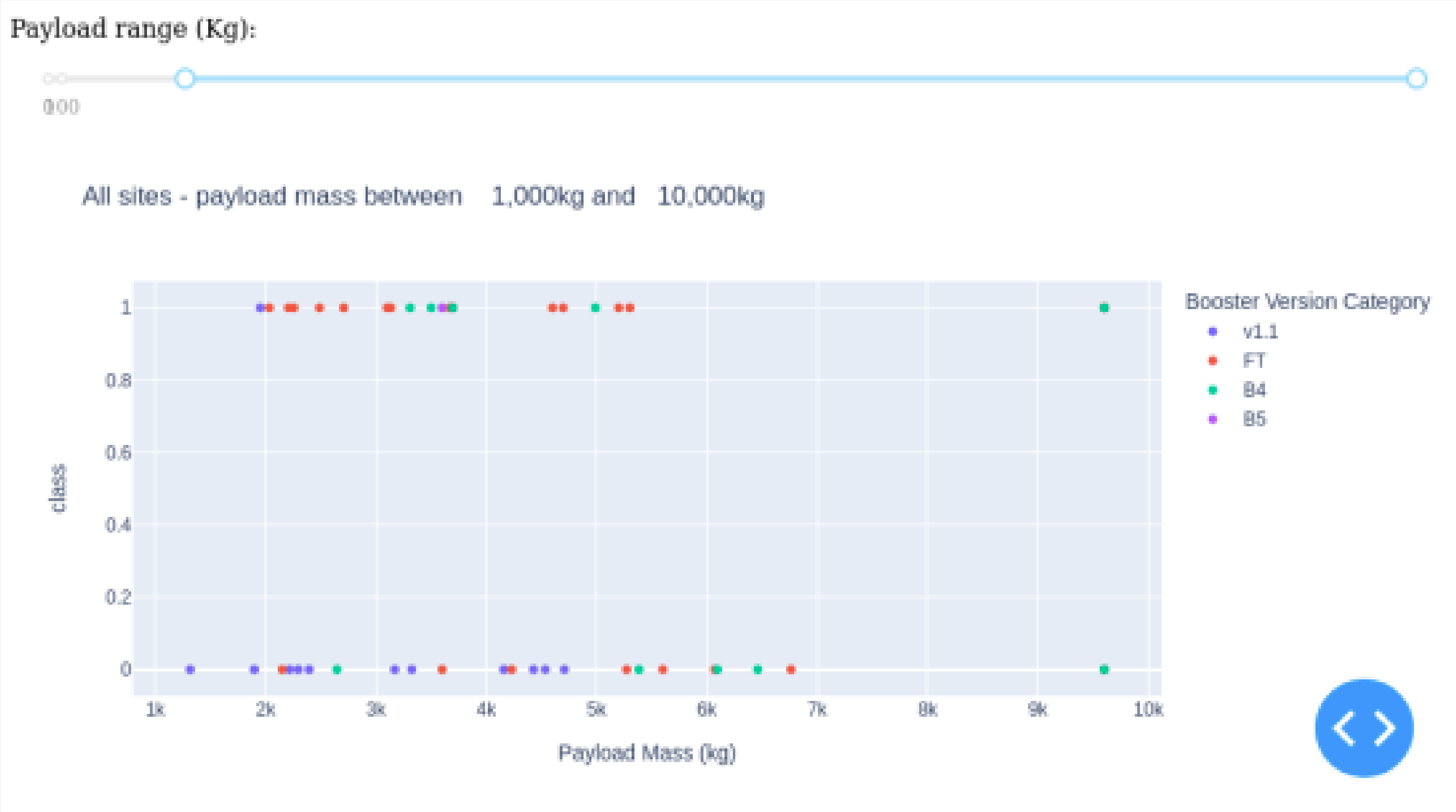
- The place from where launches are done seems to be a very important factor of success of missions.

Launch Success Ratio for KSC LC-39A



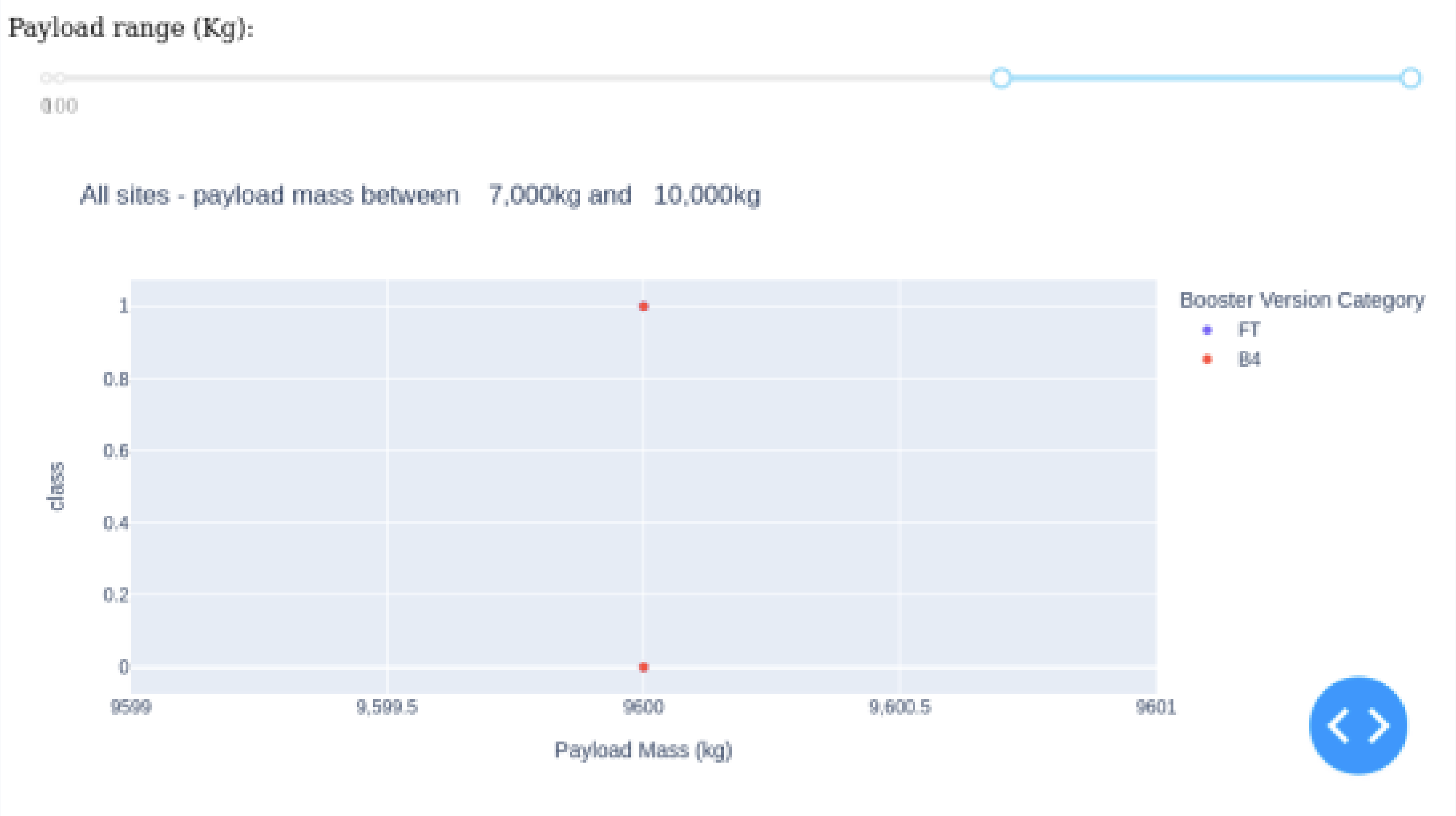
- 76.9% of launches are successful in this site.

Payload vs. Launch Outcome



- Payloads under 6000kg and FT boosters are the most successful combination.

Payload vs. Launch Outcome



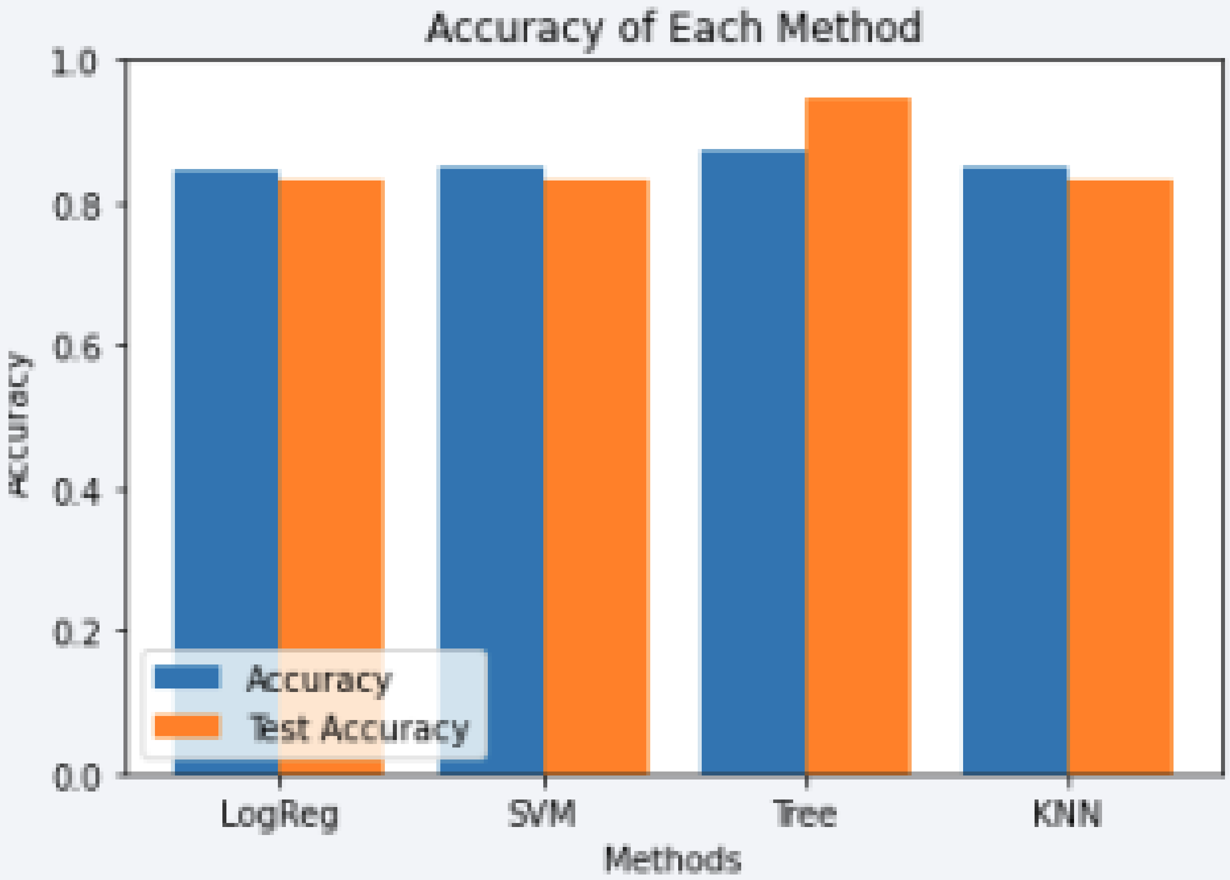
- There is not enough data to estimate risk of launches over 7000kg.

Section 5

Predictive Analysis (Classification)

Classification Accuracy

- Four classification models were tested, and their accuracies are plotted beside.
- The model with the highest classification accuracy is Decision Tree Classifier, which has accuracies over than 87%.



Confusion Matrix

- Confusion matrix of Decision Tree Classifier proves its accuracy by showing the big numbers of true positive and true negative compared to the false ones,

Conclusions

- Different data sources were analyzed, refining conclusions along the process.
- The best launch site is KSC LC-39A.
- Launches above 7000kg are less risky.
- Although most of mission outcomes are successful, successful landing outcomes seem to improve over time, according the evolution of processes and rockets.
- Decision Tree Classifier can be used to predict successful landings and increase profits.

Appendix

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

Thank you!