IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Ronald de Jong
November 2021

Outline

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

Executive Summary

Summary of methodologies

Based upon several open datasets on the Falcon-9 launches of Spacex, retrieved via API, web-scraping, databases the data is explored with different exploratory data analysis techniques. Relevant Features are used to create a model that predicts the successful landing of the first stage.

Summary of all results

a. Via analysis was found that it takes several years to build up the experience to lauch rockets including a successful landing of the first stage.

b. For Spacex it took 6 years (2010-2016) before the success rate was above 60%.

c. The launches of KSC were the most successful with 77%.

d. The ocean landing pads were successful so launch pad near ocean is advised.

e. The payload mass gave the best results in the range of 2000-6000 kg.

f. No clear relation could be found between the successful landing and the orbits of the payload.

g. The DecisionTreeClassifier model gave an accuracy of 89% for predicting a successful landing.

Introduction

Project background and context

In this presentation we try to predict under what conditions an expensive part of a rocket can be reused by a successful landing of the first stage.

The cost of a launch can be reduced by multiple times reusing the first stage.

Problems you want to find answers

- How much time is needed before rockets can be launched successfully
- How much time is needed for successful landings
- What are successful landings conditions
- Which launch sites to use
- How much payload can be launched successful with successful landings

Section 1

Methodology

Methodology

Executive Summary

In this sheet a step wise approach is presented from data collection from public sources, prepping to use this data for further analysis. Analysis based upon visualization, grouping information out of a database and use the data to predict the outcome of a landing based upon several critical features.

In this sheet only the steps and the reference to further explanation of the procedure and outcome.

- Data collection methodology:
  - Data Collection – SpaceX API (sheet 8)
  - Web Scraping (sheet 9)
- Perform data wrangling
  - Create a digital label outcome (sheet 10)
- Perform exploratory data analysis (EDA) using visualization and SQL (sheets 11, 12)
- Perform interactive visual analytics using Folium and Plotly Dash (sheets 13, 14)
- Perform predictive analysis using classification models (15)
  - How to build, tune, evaluate classification models

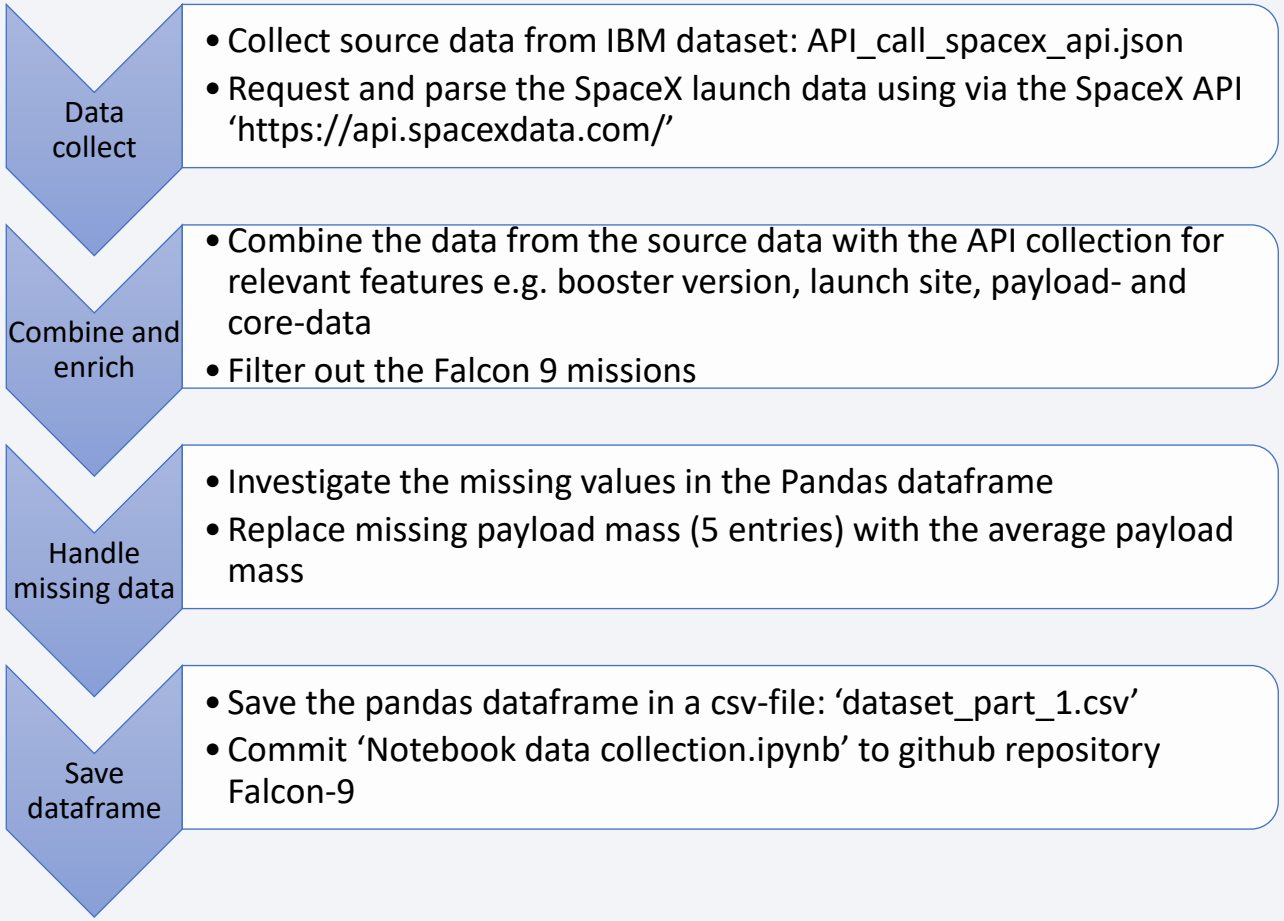
Data Collection

- Describe how data sets were collected.
  - In the next sheets (8, 9, 10) the steps are described how the data is collected via different data sources e.g. web scraping, sql collection on public tables and via the spacex API.
- You need to present your data collection process use key phrases and flowcharts
  - The process steps of the data collection, the handling, cleaning and wrangling of data is described in the following sheets om sheets

Data Collection – SpaceX API

- The data collection and enrichment are depicted in the flow.
- The GitHub URL of the completed SpaceX API calls notebook is:

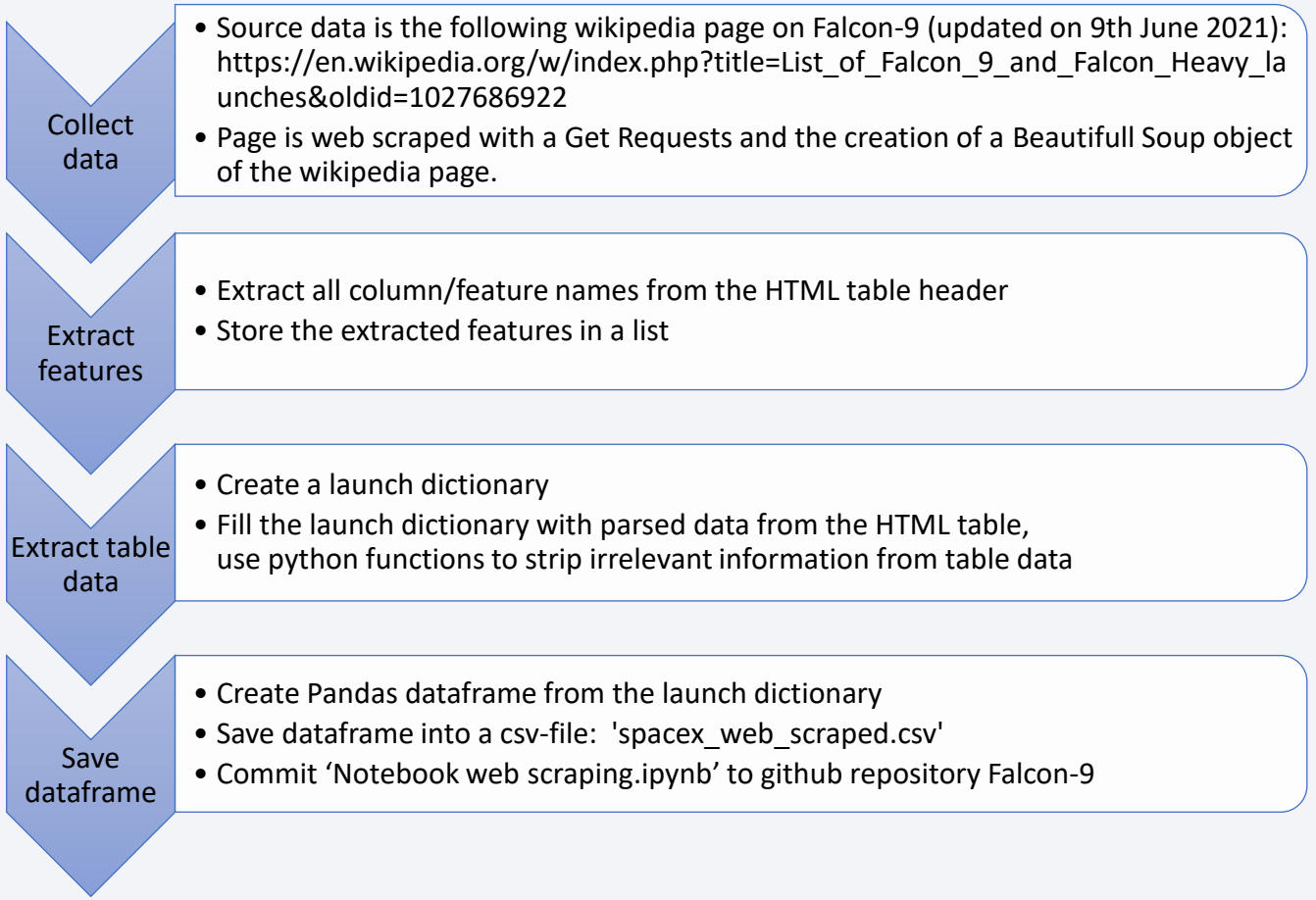
  Falcon-9/Notebook data collection.ipynb at master · RMdeJong/Falcon-9 · GitHub



Data Collection - Scraping

- Collect data via web scraping of tables on a Wikipedia page. Parse the data from tables into dataframe.
- The GitHub URL of the completed web scraping notebook IS:

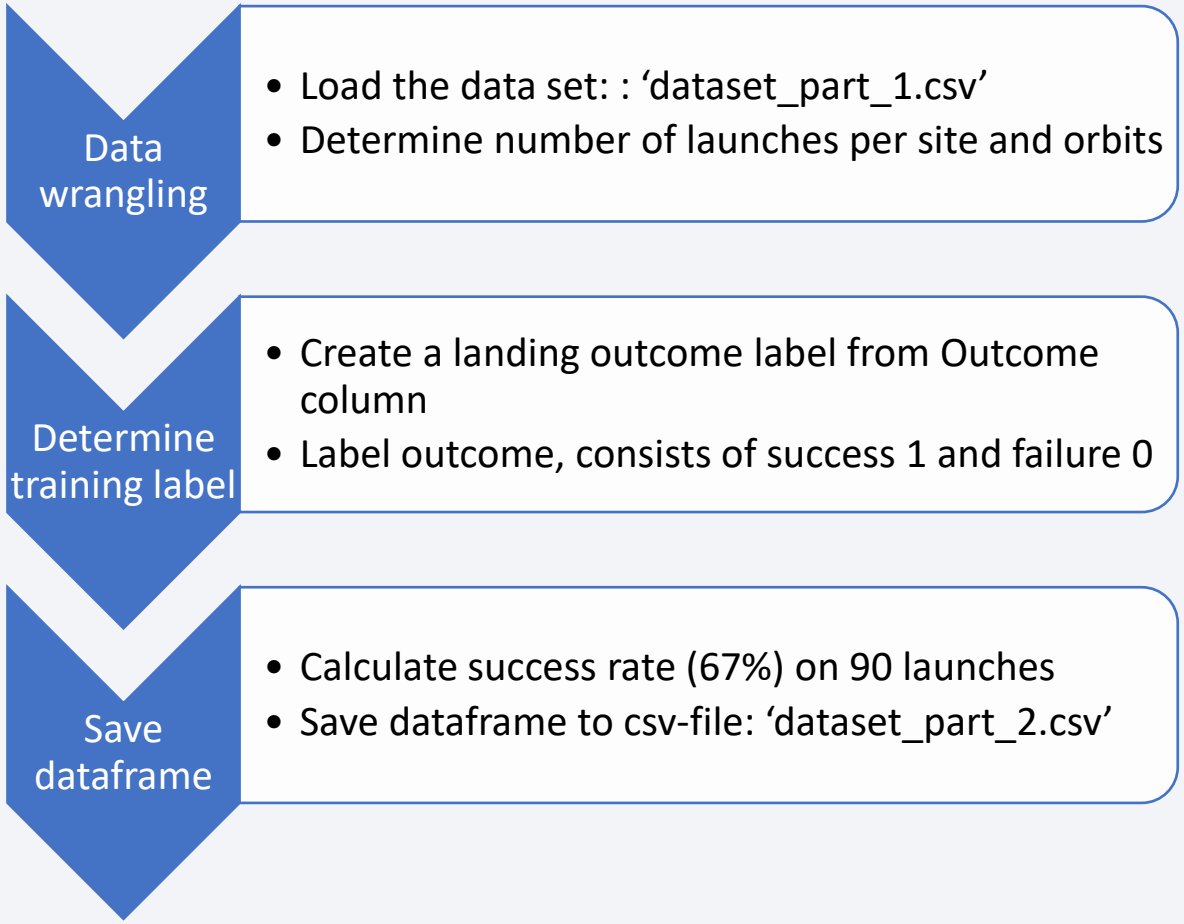
https://github.com/RMdeJong/Falcon-9/blob/master/Notebook%20web%20scraping.ipynb



Data Wrangling

- Load earlier collected data set
- Create from text object on landing outcome a digital training label with ‘1’ and ‘0’ for success and failure on landing first stage.
- Add the GitHub URL of your completed data wrangling related notebooks:

https://github.com/RMdeJong/Falcon-9/blob/master/EDA_1.ipynb



EDA with Data Visualization

- On this sheet the visualizations are described that are included in the github url Notebook.
- Add the GitHub URL of your completed EDA with data visualization notebook, as an external reference and peer-review purpose

https://github.com/RMdeJong/Falcon-9/blob/master/EDA%20Visualization.ipynb

- Charts plotted with Flight-number vs Payload mass and launch site with label Class to find relation between different features and outcome plotted in time.
- Bar chart with relation feature Orbid and label Class to see relation between Orbid and success rate. Also plot with Orbid vs Flight-number to see development in time.
- Plot between features Orbid and Payload Mass to find correlation between these features.
- Yearly plot on success rate indicating that there is a learning effect over the years resulting in an increase of success rate.

EDA with SQL

- SQL queries on SPACEXTBL table that was loaded as csv-file from: Spacex.csv
- In the GitHub URL are included SQL queries to retrieve the following information:

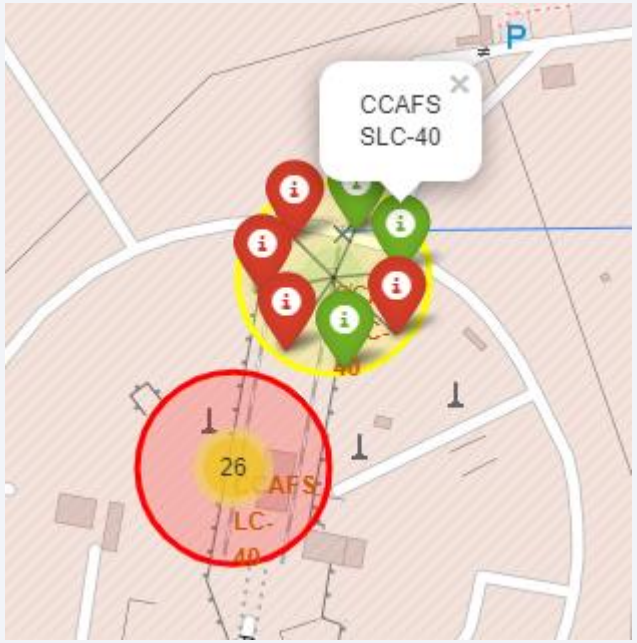
1. names of the unique launch sites in the space mission
2. launch sites begin with the string 'CCA'
3. total payload mass carried by boosters launched by NASA (CRS)
4. average payload mass carried by booster version F9 v1.1
5. date when the first successful landing outcome in ground pad was achieved
6. names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000 kg
7. the total number of successful and failure mission outcomes
8. names of the booster_versions which have carried the maximum payload mass
9. the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015
10. the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

https://github.com/RMdeJong/Falcon-9/blob/master/EDA-SQL.ipynb

Build an Interactive Map with Folium

- On different Folium Maps, the location of Launch sites are depicted.
- With Circle Object in different colors the Launch sites are shown with the site name as label.
- At the sites additional information is presented on the number and success of the launches.
- With lines several distances are shown with distance marker to relevant details e.g. coastline and highways.
- GitHub URL:

https://github.com/RMdeJong/Falcon-9/blob/master/Folium%20.ipynb



Build a Dashboard with Plotly Dash

- In a single dashboard the options are available to determine the success of launches for all sites and per site the success rate.
- There is also the opportunity to visualize the relation between the payload mass and success of the landing grouped per booster version.
- The python file is included in the GitHub repository:

https://github.com/RMdeJong/Falcon-9/blob/master/spacex_dash_app.py

Predictive Analysis (Classification)

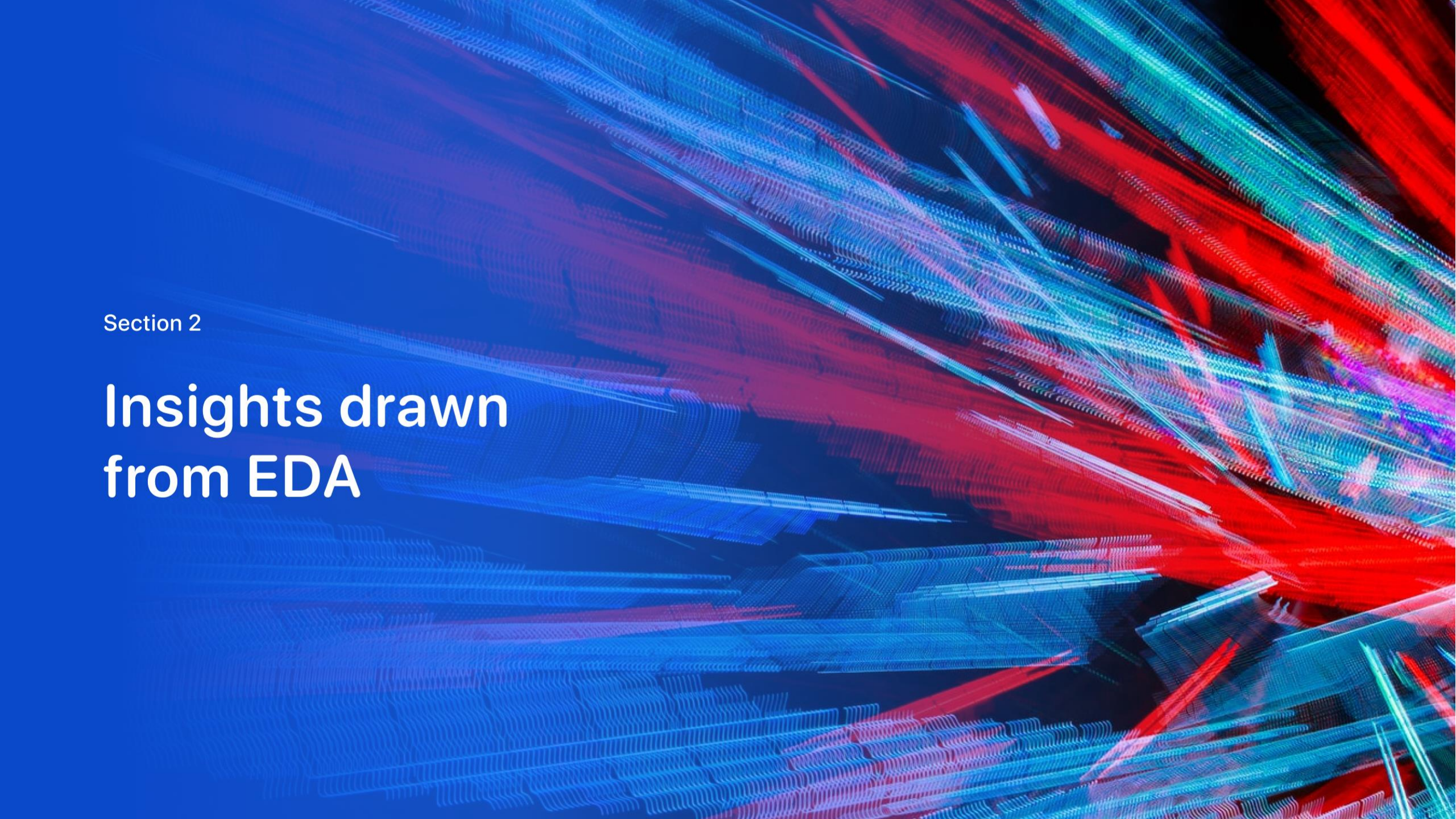
- Input datasets (dataset_part_2.csv and dataset_part_3.csv) that are cleaned. Hot-endcoded, with a selection of essential features and where an outcome label 'Class' is included as result of successful landing.
- Feature data is first transformed via a standard scalar. Next step is to split the data into training and testing set for features and label. Test size is 20% of the dataset of 90 records.
- Based on a range of relevant parameters and with GridSearchCV where the test set is folded with cv=10 the best set of parameters is found on the training data. With the best parameters the accuracy is calculated.
- Four models are tested with accuracies between 0.84 and 0.89 for LogisticRegression, Support Vector Machine, DecisionTreeClassifier and KNeighborsClassifier.
- The GitHub URL of your completed predictive analysis lab: https://github.com/RMdeJong/Falcon-9/blob/master/Machine%20Learning%20Prediction%20lab.ipynb

Results

- Exploratory data analysis results
  - Results are shown on sheets 18-23
- Interactive analytics demo in screenshots
  - Results are shown on sheets 35-43
- Predictive analysis results
  - Results are shown on sheets 45-47

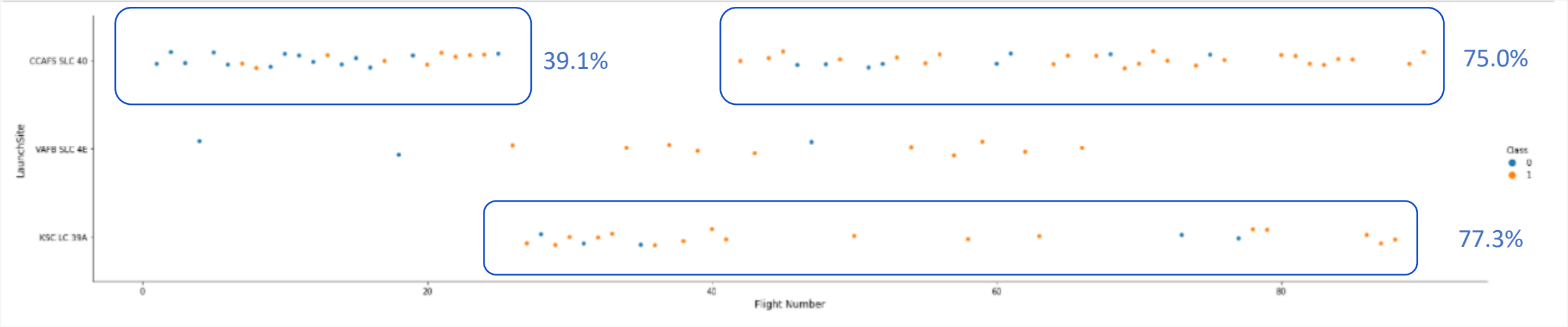
Section 2

Insights drawn from EDA



Flight Number vs. Launch Site

- The number of flights from the different launch sites with the outcome of the successful landing of the first stage.
- The first batch of 23 launches from CCAFS SLC 40 have a success rate of 39.1%. The second batch of 32 flights are improved to 75%. In total CCAFS SLC 40 launches 55 of the 90 flights.
  KSC LC 39A has a success rate of 77.3% in 22 flights.
- After flight 20 the successful landings increases independent of the Launch site.

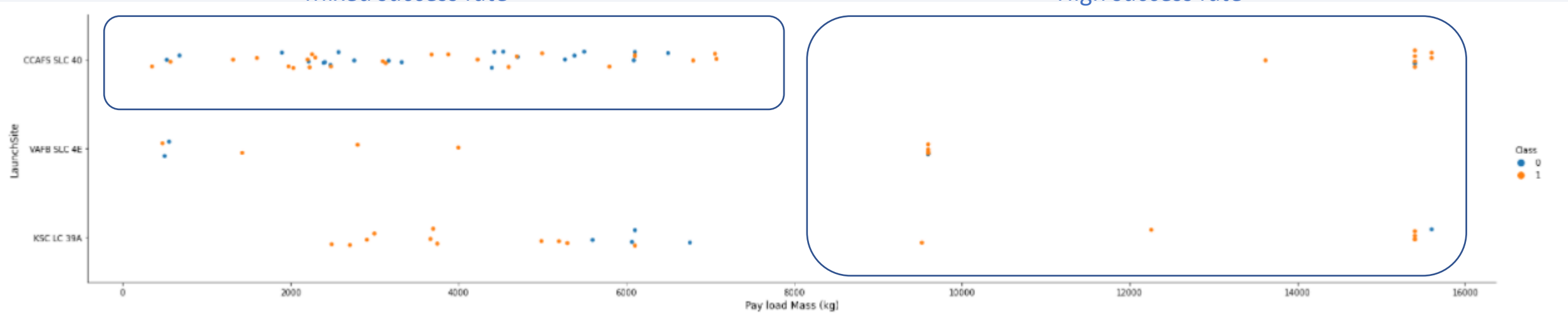


Payload vs. Launch Site

- Large part of the flights have a payload below 8000 kg.
- Large payloads (> 8000 kg) seem to be more successful. However, those flights were launched in the later flights withs higher success rates.
- The success rate of CCAFS SLC 40 with payload below 8000 kg have a low success rates again in combination with earlier flights.

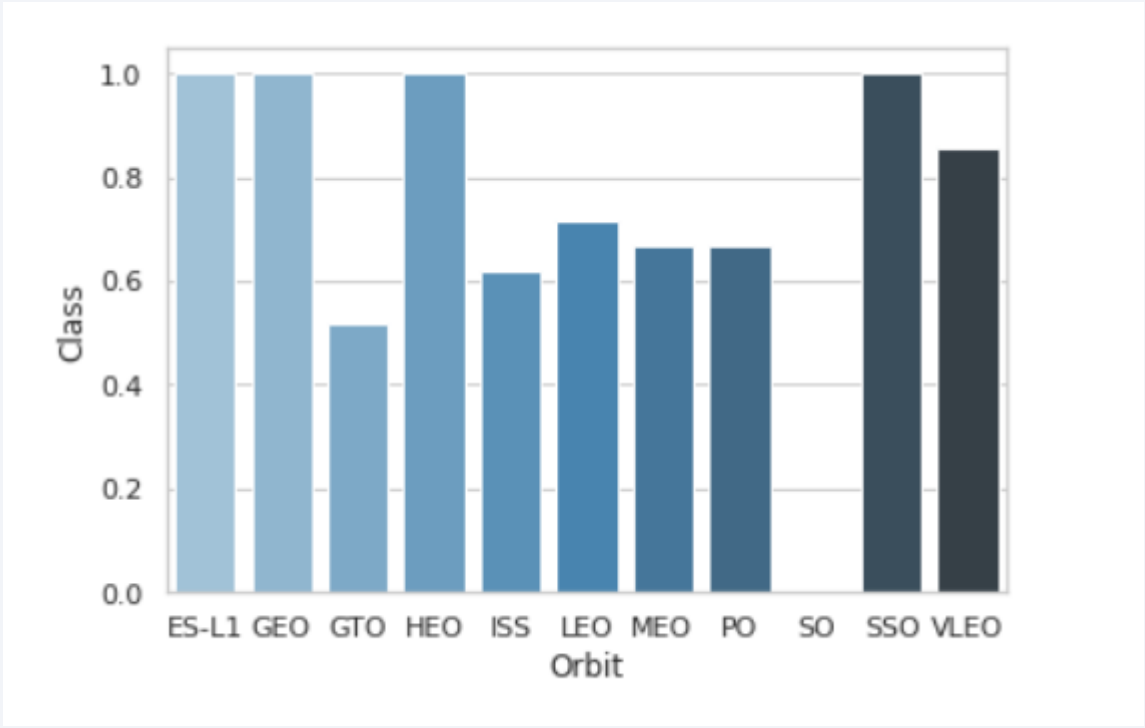
Mixed success rate

High success rate



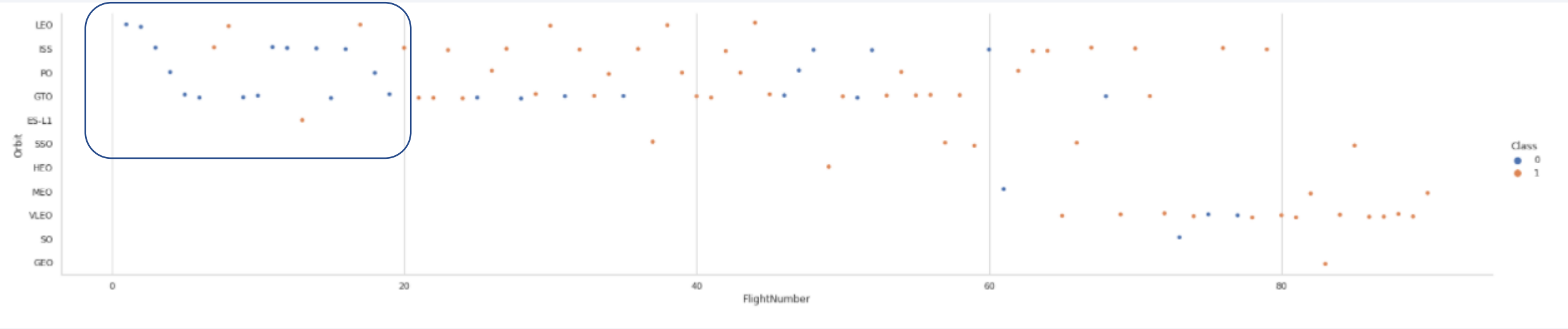
Success Rate vs. Orbit Type

- There are several launches with a success rate of 100% e.g. ES-L1, GEO and HEO but there was only 1 flight per Orbid.
- On the other hand success rate of 0% for SO but also for 1 flight.
- In the scatterplot Orbit vs flight number (next sheet) can be seen that that in the first 20 flights the success rate was low, independent of the orbit.



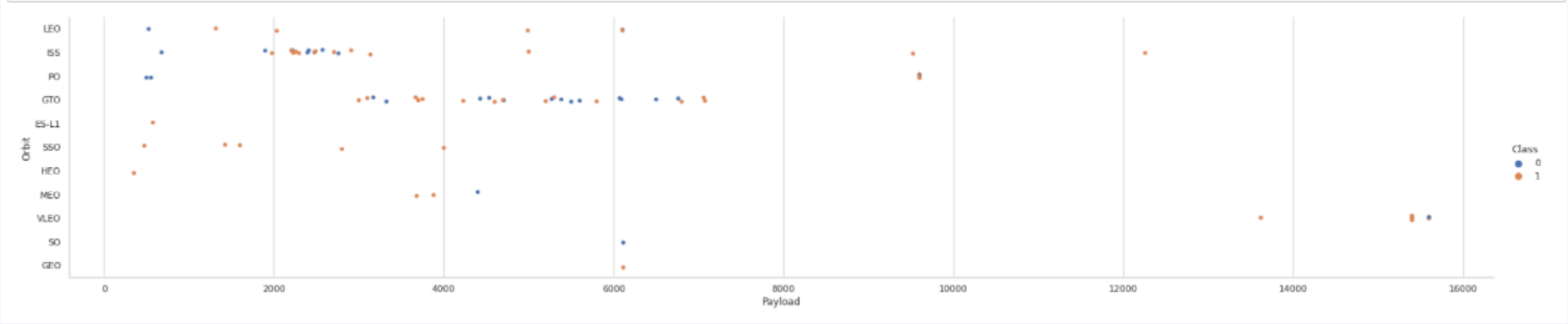
Flight Number vs. Orbit Type

- As mentioned with the bar chart (previous sheet) the first 20 launches were not that successful independent of the Orbid
- After 20 launches the success rate increases more or less for all Orbids. After 60 launches, VLEO Orbids are added. That seem to replace the LEO Orbids.



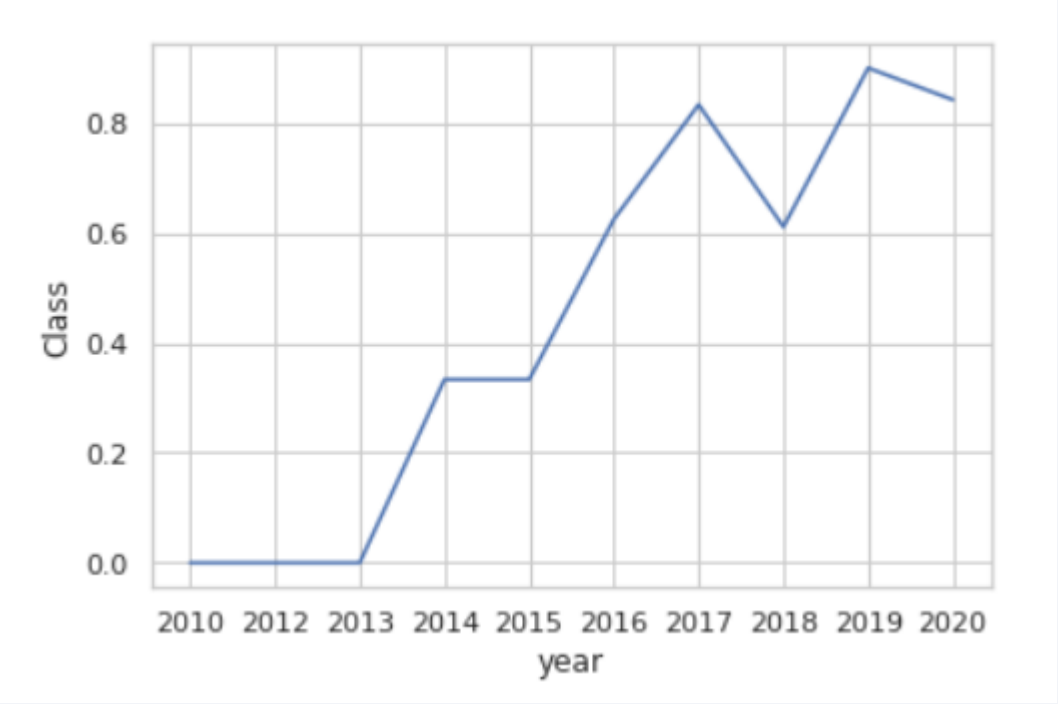
Payload vs. Orbit Type

- The majority of the payloads is well below the 8000 kg.
- Loads above 8000 kg are launched for ISS, Polar (PO) and VLEO Orbids.



Launch Success Yearly Trend

- The first three years there were no successful launches.
- Between 2013 and 2017 the success rates keeps increasing to 83.3 %.
- In 2018 there is a dip in the success rate with a decrease to 61.1 % for 18 launches.
- After 2018 the success rate is well above 80 %, with a top year in 2019 with 90% (10 launches).



All Launch Site Names

- The names of the launch sites are in the SPACEXTBL table.
- By adding the distinct parameter to the lauch_site column only the unique launch sites will be shown.

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

```
]: %sql select distinct launch_site from SPACEXTBL;
```

```
 * ibm_db_sa://hdb98214:***@764264db-9824-4b7c-82df-40d1b13
Done.
```

[6]:

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

Launch Site Names Begin with 'CCA'

- Select only the names that start with ‘CCA’ s done with the LIKE command starting with ‘CCA%’ followed by everything.
- The records are limited by the command LIMIT 5

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

```
7]: %sql select * from SPACEXTBL \
    where launch_site LIKE 'CCA%' \
    limit 5;
```

```
 * ibm_db_sa://hdb98214:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32536/bludb
Done.
```

ut[7]:

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

Task 3

Total Payload Mass

- The total payload carried by boosters from NASA are retrieved by calculating the SUM of the column payload_mass_kg where the customers is equal to ‘NASA (CRS)’.
- The outcome is 45,596 kg.

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

```
]: %sql select sum(payload_mass__kg_) as total_payload_mass from SPACEXTBL \
   where customer = 'NASA (CRS)';
```

```
 * ibm_db_sa://hdb98214:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08l
Done.
```

:[8]:

total_payload_mass
45596

Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1 is calculated by the command avg on the column payload_mass_kg where the booster_version = ‘F9 v1.1’.
- The outcome of the average payload is 2928 kg.

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

```
: %sql select avg(payload_mass__kg_) as avg_payload_mass from SPACEXTBL \
  where booster_version = 'F9 v1.1';
```

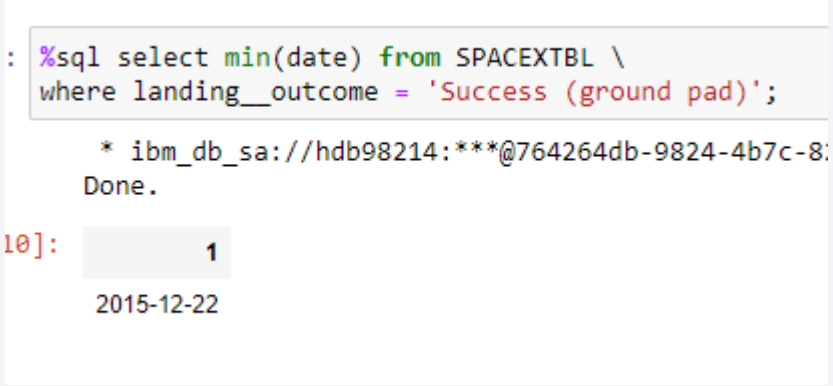
```
 * ibm_db_sa://hdb98214:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90
Done.
```

[9]:

avg_payload_mass
2928

First Successful Ground Landing Date

- The first successful landing outcome on ground pad can be retrieved by the command min on the column date where the landing_outcome is equal to ‘Success (ground pad)’
- The result date is: 2015-12-22



Successful Drone Ship Landing with Payload between 4000 and 6000

- The booster_version landed successful can be limited by a landing on a drone ship by the WHICH command that makes the landing_outcome equal to ‘Success (drone ship)’ AND also by the payload_mass_kg BETWEEN 4000 AND 6000
- The outcome results in four booster versions with associated payload_mass.

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

```
[11]: %sql select booster_version, payload_mass__kg_ from SPACEXTBL \
      where landing__outcome = 'Success (drone ship)'\
      and payload_mass__kg_ between 4000 and 6000 ;
```

```
 * ibm_db_sa://hdb98214:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:325
Done.
```

t[11]:

booster_version	payload_mass__kg_
F9 FT B1022	4696
F9 FT B1026	4600
F9 FT B1021.2	5300
F9 FT B1031.2	5200

Total Number of Successful and Failure Mission Outcomes

- The total number of successful and failure mission outcomes are found with the cound command on the column ‘mission_outcome’ where the mission_outcome are grouped by unique item.
- The outcome shows 1 Failure in flight and 100 successful.

List the total number of successful and failure mission outcomes

```
%sql select mission_outcome,count(mission_outcome) from SPACEXTBL \
group by mission_outcome;
```

```
 * ibm_db_sa://hdb98214:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2
Done.
```

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

Boosters Carried Maximum Payload

- A subquery finds the max value of the payload_mass_kg column.
- The booster_version slection uses this value to compare it with the payload_mass_kg.
- The outcome show a list of booster-version that carried the maximum payload.

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);
```

```
 * ibm_db_sa://hdb98214:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1od8lcg.databa
Done.
```

29]:

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

2015 Launch Records

- The date, booster_version and lauch site are queried where the landing_outcome is limited till ‘Failure (drone ship)’ AND
  the data contains 2015.
- The result is two rows in January and April both launched from CCAFS LC-40 with two different booster versions.

List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

```
%sql select date,booster_version, launch_site from SPACEXTBL \
where landing__outcome = 'Failure (drone ship)'\
and date like '2015%' ;
```

```
 * ibm_db_sa://hdb98214:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1od8lcg.databases.app
Done.
```

43]:

DATE	booster_version	launch_site
2015-01-10	F9 v1.1 B1012	CCAFS LC-40
2015-04-14	F9 v1.1 B1015	CCAFS LC-40

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

- The landing_outcome and count(landing_outcome) are limited between the two given dates and grouped by the landing_outcome.
- The count(landing_outcome) is presented by the command ORDER BY in DESC (descending) order.
- The result is given in the screen shot.

```
%sql select landing__outcome,count(landing__outcome) from SPACEXTBL \
where date between '2010-06-04' and '2017-03-20' \
group by landing__outcome \
order by count(landing__outcome) desc;
```

```
 * ibm_db_sa://hdb98214:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io
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

Section 4

Launch Sites Proximities Analysis

All launch sites for Falcon-9

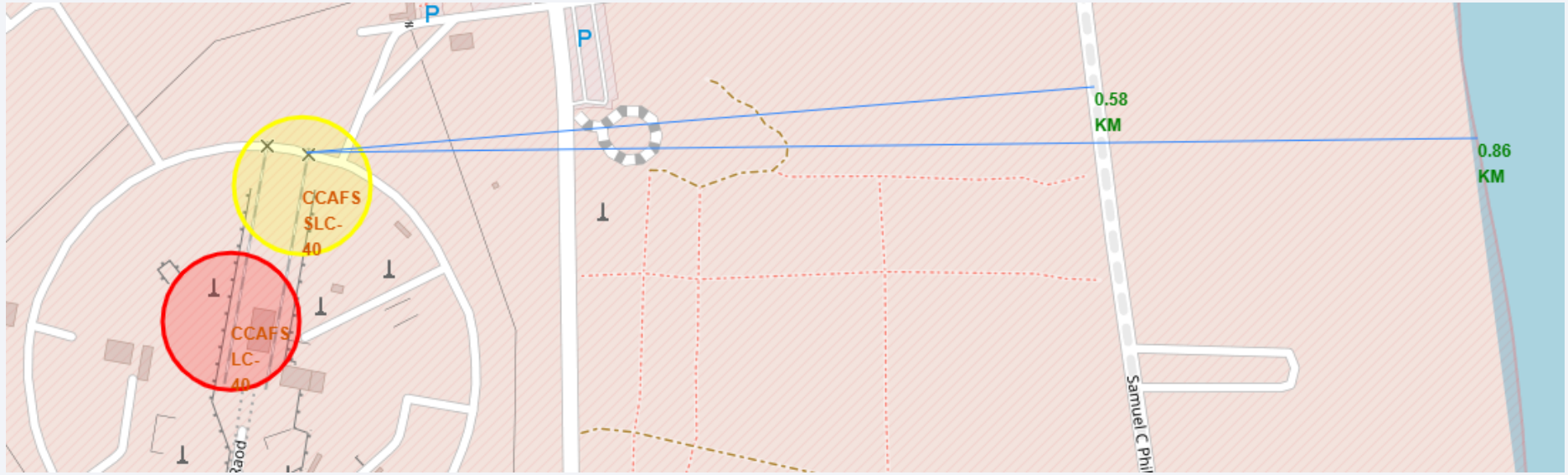
- Map shows the four launch sites with on the west coast: Vandenberg Space Launch Complex 4 and on the east coast: Kennedy Space Center Launch Complex 39A and two platforms on Cape Canaveral Space Launch Complex 40.
- The American launch sites are not near the equator line. Launches near the equator get an extra boost in speed due to the rotation of the Earth.
- The Launch sites are near the coastline so when something goes wrong no populated areas are in risk.

Launch outcomes for KSC LC-39A

- Map shows the total amount of Falcon-9 launches from Kennedy Space Center Launch Complex 39A.
- The color of the info-icons indicate the success of the in total 13 launches:
  - Green is successful (10)
  - Red is unsuccessful (3)

Cape Canaveral Space Launch Complex 40 proximities

- In the map below you can find the distance to import feature of the launch site such as the coastline (0.86 km) and highway (0.58 km)
- The Launch sites are near the coastline so when something goes wrong no populated areas are in risk. From the highway material for rockets can be transported easily via trucks on highways.

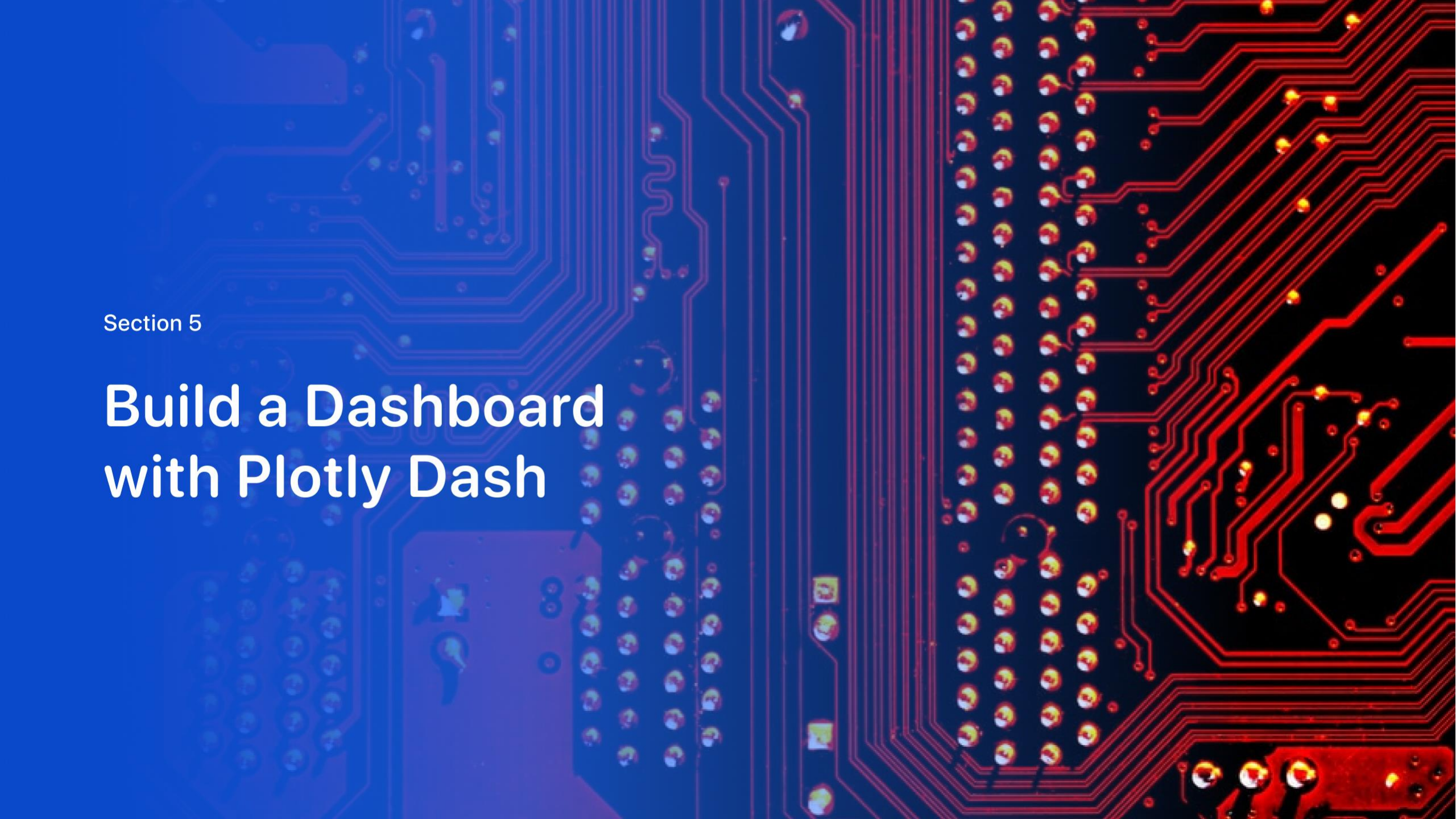


Benefit Launch site close to coast and port

The first stage of the SpaceX Falcon 9 rocket that launched the Demo-2 mission on May 30, 2020, arrives in Florida's Port Canaveral on June 2, 2020. (Image credit: SpaceX via Twitter)

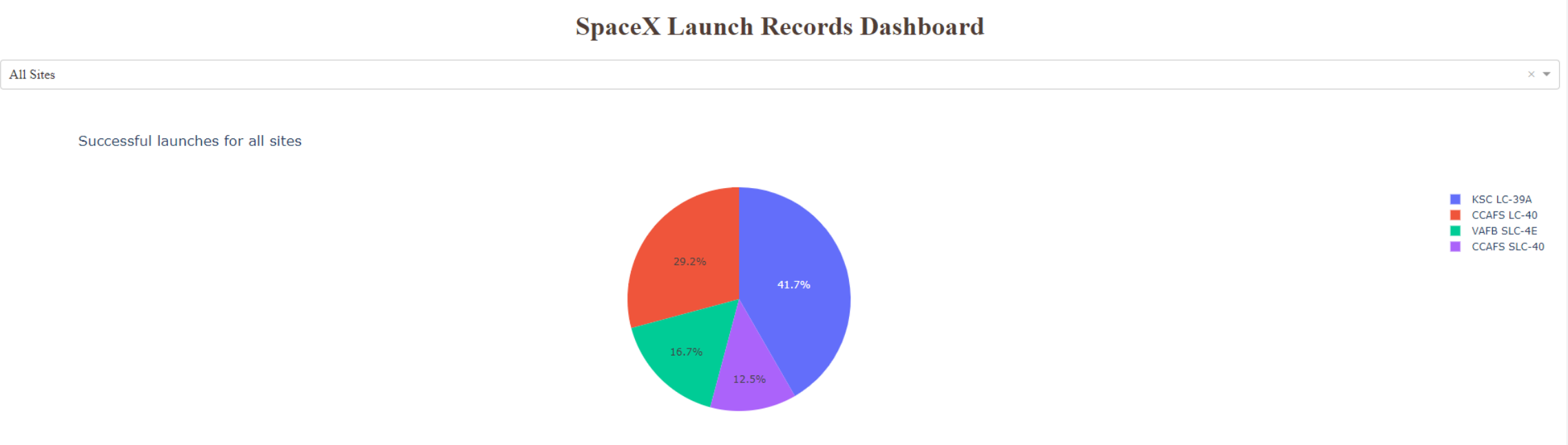
Section 5

Build a Dashboard with Plotly Dash



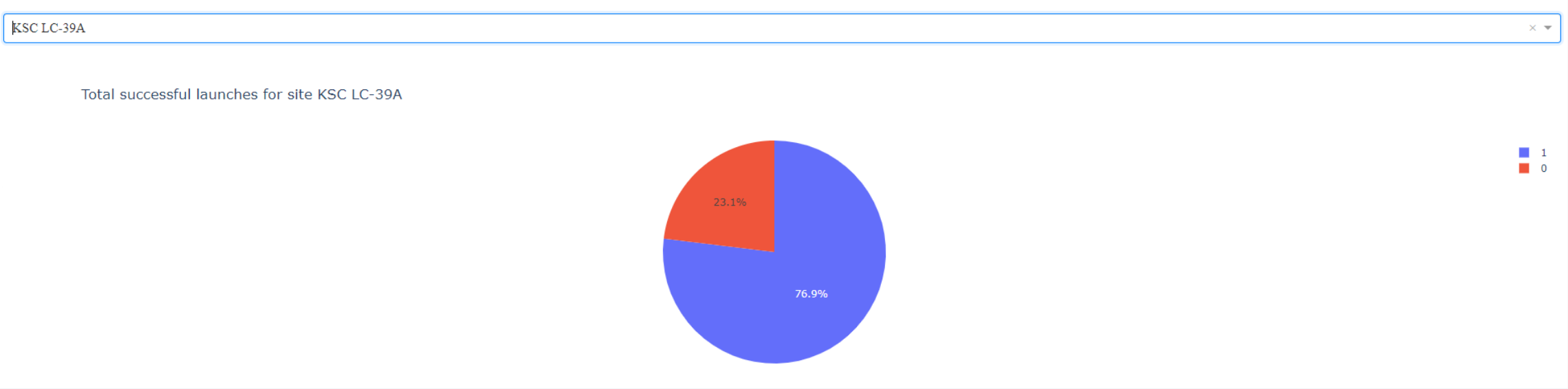
Dashboard successful launches per site

- Interactive dashboard where (in this selection) for all launch sites the percentage of successful launches are presented in a pei chart.
- For KSC LC-39A the share of successful launches was 41.7%



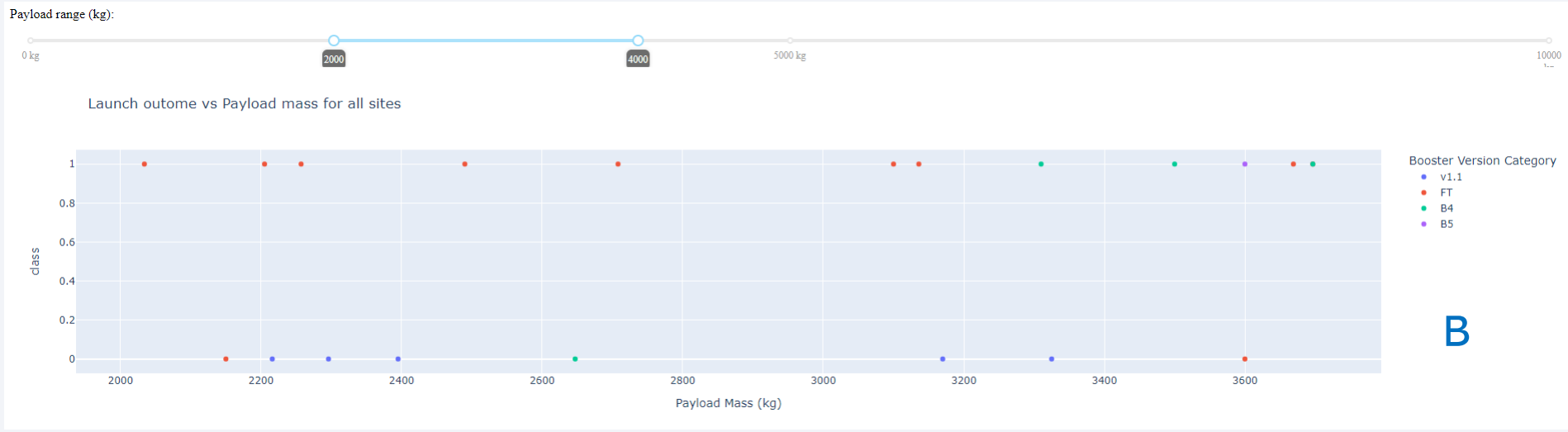
Success rate launches from KSC LC-29A

- Selected in the interactive dashboard launch site KSC LC-39A with the best success rate of 76.9%.
- In the dashboard via a pop-up can be seen that the number of successful launches was 10 vs unsuccessful 3.



Successful launches vs payload mass per booster version

- Figure A: launch outcome for all sites and all payloads, showing a low success rate on payloads higher than 6000 kg.
- Figure B: launch outcome for all sites with payload between 2000 and 4000 kg, showing high success rate for booster version FT (red dots)



Successful launches vs payload mass per booster version

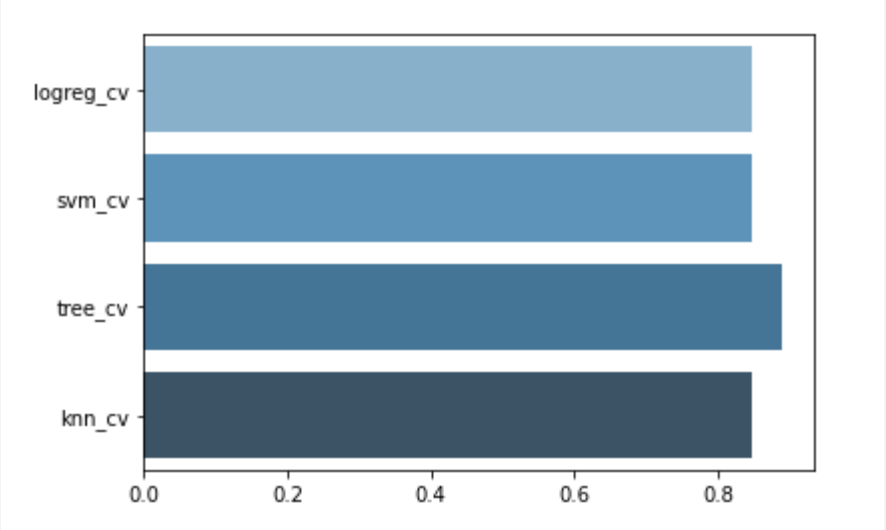
- Which site has the largest successful launches?
  - KSC – 10 launches
- Which site has the highest launch success rate?
  - KSC – 76.9%
- Which payload range(s) has the highest launch success rate?
  - Payloads between 2000-4000 kg
- Which payload range(s) has the lowest launch success rate?
  - Payload between 6000-8000 kg
- Which F9 Booster version (v1.0, v1.1, FT, B4, B5, etc.) has the highest launch success rate?
  - Booster version FT has the highest success rate

Section 6

Predictive Analysis (Classification)

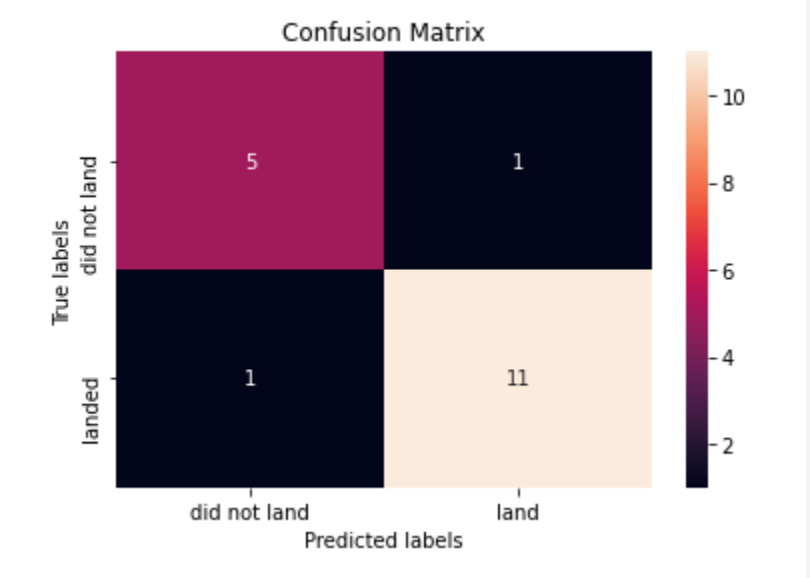
Classification Accuracy

- Bar chart with the four classification models.
- Model DecisionTreeClassifier has the highest classification accuracy: 0.889285714285714 5
- DecisionTreeClassifier(criterion='entropy',max_depth=4,max_features='auto',min_samples_leaf=4, min_samples_split=10,splitter='random')



Confusion Matrix

- In the confusion matrix is shown that of the 18 launches in the test set that 6 did not have a successful landing and 12 did.
- From the 12 successful landings 11 were predicted correct. 1 false negative
- Of the 6 not successful landings 5 were predicted correct and 1 was not (false positives)
- DecisionTreeClassifier(criterion='entropy',max_depth=4,max_features='auto',min_samples_leaf=4, min_samples_split=10,splitter='random')



Conclusions

- The models are very accurate to predict the outcome of successful landings of the first stage
- The outcome of the model DecisionTreeClassifier the best with highest accuracy.
- Relevant features in the model:
    - Block, ReusedCound, LandingPad , GridFins True, Boosterversion and LaunchSite
- The ocean landings pad has the best outcome.
- KSC gives the best result for launches: 76.9%
- F9 Booster version FT gives the best results
- Best success rates for payloads between 2000-4000 kg

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!