IBM Developer
SKILLS NETWORK

WINNING SPACE RACE WITH DATA SCIENCE

Seoyoung Jeon

17 March 2024


USAF
ULA
DELTA

EXECUTIVE SUMMARY

Analysis data

- Data set on success or failure of Falcon 9 by launch conditions

Objective

- Exploring the relationship between launch conditions and success rate
- Prediction model for Falcon 9 launch outcome

Target launch conditions for analysis

- Launch site, Pay load mass, booster version, targeted orbit

Machine learning technique applied for prediction model

- Logistic regression, Support vector machine, Decision tree classifier, K nearest neighbors

Summary of results

- Launch site and pay load mass have low impact
- Booster version and target orbit have high impact
- All machine learning technique have same accuracy (88%)

INTRODUCTION

- With the advent of the space era, a number of attemps to send spacecraft into space.
- However, launching a spacecraft requires a large budget, and hundreds of tons waste generated.
- The SpaceX, founded by Elon Musk, aims to complete the spacecraft from production to launch for $30 million.
- To this goal, SpaceX has planned a project to recover and reuse fairings and launch vericles, etc.
- However, recovery of projectiles and launch projectile were not always successful.
- Therefore, this analysis was conducted to analyze the factors affecting landing and mission outcome of SpaceX.

Brief description of Falcon 9 – Launch site -

[California]
VAFB SLC-4E

[Florida]
KSC LC-39A
CCAFS SLC-40
CCAFS LC-40

Closest Coast & City

- Considered for safety reasons (Temporary closure etc.)
- Considered for noise and pollution while firing

Closest Rail & Highway

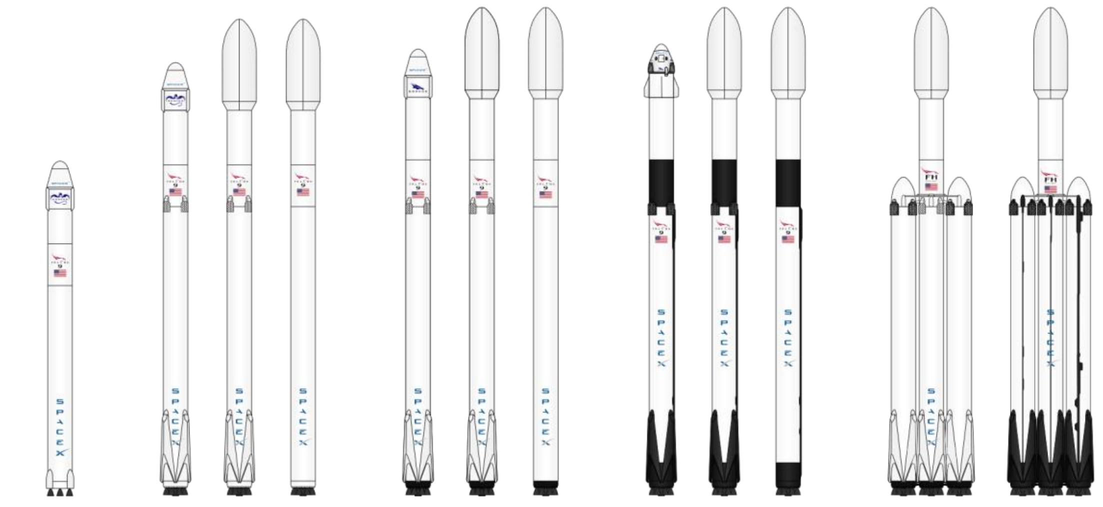
- Considered to reduce transportation costs

Brief description of Falcon 9 - Orbit -

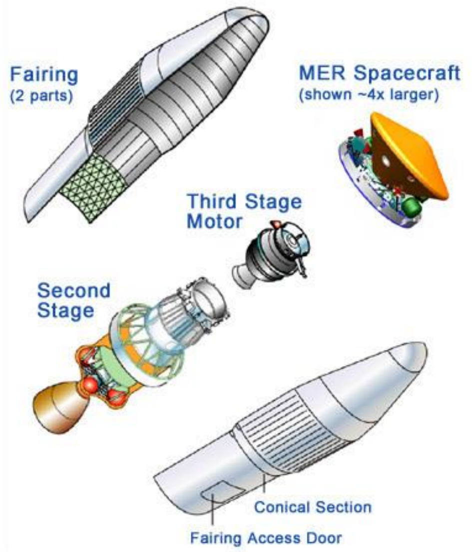
Table 1. Description of the orbit targeted by Falcon 9

Orbit	Description	Satellite usage
LEO (Low earth orbit)	Altitude 2,000 km	Most of Satellite
VLEO (Very low earth orbit)	Lower than LEO	Communications, observation, military
ISS (International space station orbit)	Located LEO	ISS
GEO (Geostationary orbit)	Altitude 35,786 km	Communication & weather
MEO (Medium earth orbit)	Between LEO and GEO	Navigation & telecommunication
PO (Polar orbit)	Passes above both polar	Monitoring polar
GTO (Geostationary transfer orbit)	Elliptical orbit with the Earth's altitude	Communication satellite
ES-L1 (Earth-Sun Lagrange point)	Between Earth and Sun (L1 Lagrange point)	Solar astronomy research
SSO (Sun-Synchronous orbit)	Constant solar illumination conditions at all times	Scientific, environmental monitoring
HEO (Highly elliptical orbit)	Elliptical orbit, closest and farthest points are significantly different from Earth	Communication, astronomical observation
SO (Synchronous orbit)	Passing through a designated area at a certain time	Communication, exploration

Brief description of Falcon 9 - Booster & Pay load -



Version	Period	Pay load mass (max, t)
V1.0	Jun/2010 – Mar/2013	LEO (9), GTO(3.4)
V1.1	Sep/2013 – Jan/2016	LEO (1.3), GTO (4.8)
FT	Dec/2015 – Jun/2018	LEO (22.8), GTO (8.3), Mars (4)
B4	Aug/2017 – 2018	
B5	Mar/2018 -	



Booster

- Starts with v1.0, and Block 5 (B5) is currently in operation

Pay load mass

- How much weight a projectile can carry
- Depends on the version of the booster and the target orbit
- Also, depends on whether it is new or recycled (recycled one have low capacity)

MATERIALS & METHODS

01

Data collection

- SpeceX API
- Wikipedia

02

Data wrangling

- Data filtering
- Data grouping

03

Data analysis

- SQL
- Data visualisation

04

Visualisation

- Folium
- Plotly dash board

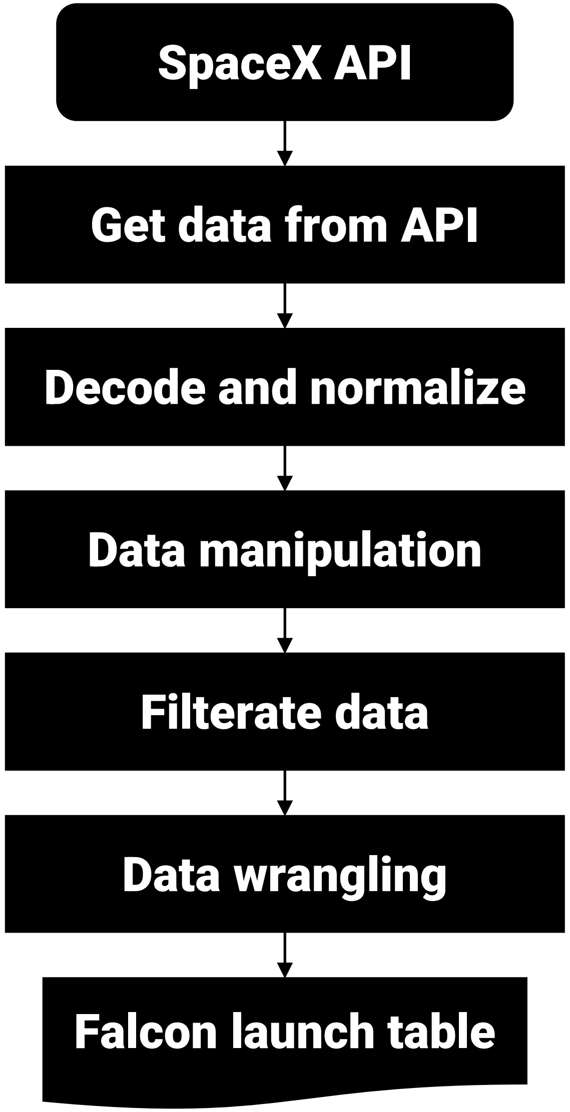
05

Modeling

- Classification models
- Tune and evaluation

DATA COLLECTION

DATA COLLECTION - SpaceX API -



Get data from API
https://api.spacexdata.com/v4/launches/past

Decode and normalize
Decode the table obtained from the API and retrieve information about 43 items

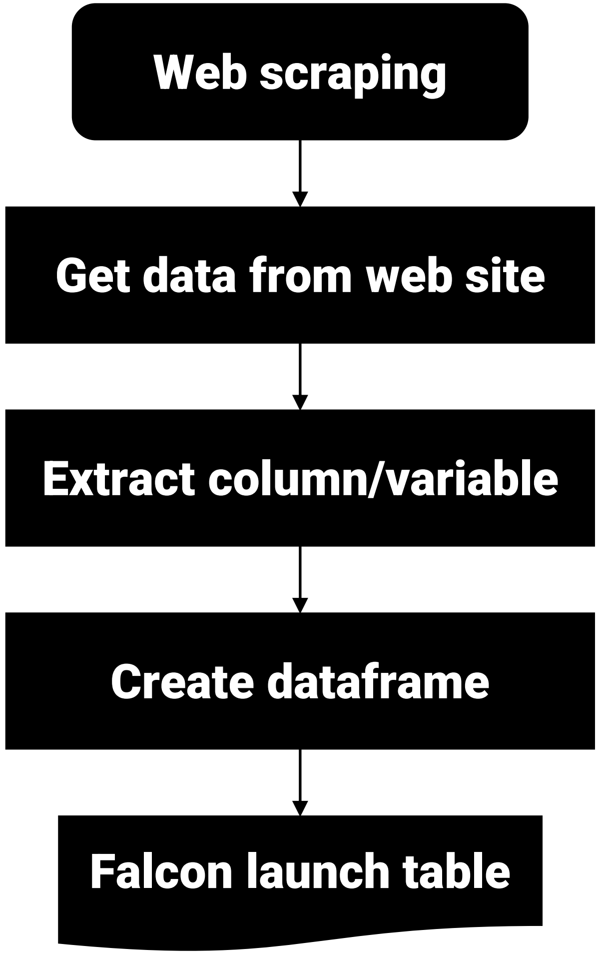
Data manipulation
Select only the interesting attributes from the retrieved table

Filterate data
Create data table containing only information about 'Falcon 9'

Data wrangling
Replace missing value (Payload mass) with the average value

Data collection script: https://github.com/SeoyoungJeon/IBM_CHP10/blob/main/Week1_Data_collection_API.ipynb

DATA COLLECTION - Web Scraping -



Get data from web site (wikipedia)
https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922
Create BeautifulSoup object from HTML response

Extract all column & variable names
From the HTML table header

Create a data frame
By parsing the launch HTML tables

Data collection script: https://github.com/SeoyoungJeon/IBM_CHP10/blob/main/Week1_webscraping.ipynb

DATA WRANGLING

DATA WRANGLING

Data set from calculate

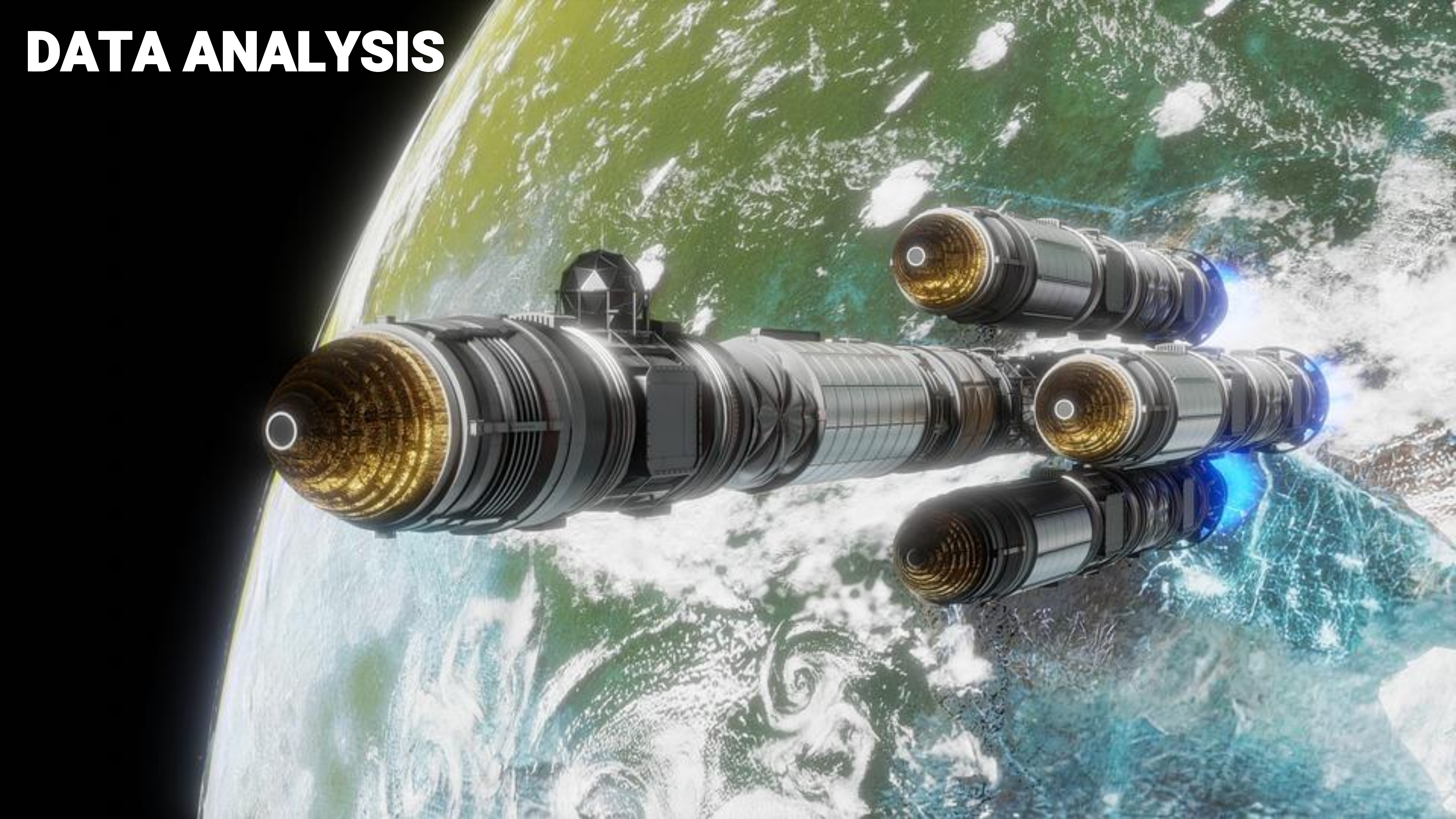
- The number of launches on each site
- The number of occurrence of each orbit
- The number and occurrence of mission outcome of the orbits

Create a landing outcome column

- To facilitate analysis, the landing outcome is indicated into 1 (success) and 0 (failure)
- In case of success, indicate 1: True ASDS, True RTLS, True Ocean
- In case of failure, indicate 0: None None, False ASDA, False Ocean, None ASDS, Fals RTLS

Data wrangling script: https://github.com/SeoyoungJeon/IBM_CHP10/blob/main/Week1_Data_Wrangling.ipynb

DATA ANALYSIS



DATA ANALYSIS - EDA with SQL -

Table 1. SpaceX data list requested through SQL

Data set	Condition
Launch site	All
Launch site	Launch site name starts with ‘CCA’
Total Pay load mass (kg)	Booster launched by NASA
Average Pay load mass (kg)	Booster version is F9 v1.1
All	First successful landing date
Booster version	Successful landing, Pay load mass 4 ~ 6 tons
Number of data per mission outcome	-
Booster version	When Pay load mass is max
Month, Booster version, Launch site	Failed landing in 2015
Ranking of the number of data by landing outcome	04/June/2010 – 20/March/2017

SQL script: https://github.com/SeoyoungJeon/IBM_CHP10/blob/main/Week2_EDA_SQL.ipynb

DATA ANALYSIS - EDA with data visualization -

Table 2. Dependent (Y) and independent (X) variables and graph type of each graph

Y	X	Graph type	Note
Pay load mass (kg)	Flight number	Scatter plot	
Launch site	Flight number	Scatter plot	Class
Launch site	Pay load mass (kg)	Scatter plot	
Orbit	Filgith number	Scatter plot	
Orbit	Pay load mass (kg)	Scatter plot	Class
Class (average)	Orbit	Bar chart	
Class (average)	Year	Line chart	

Visualization script: https://github.com/SeoyoungJeon/IBM_CHP10/blob/main/Week2_EDA_visualization.ipynb

VISUALISATION

VISUALISATION - Interactive map with Folium -

- **Mark all launch sites on a map**
  - CCAFS LC-40
  - CCAFS SLC-40
  - KSC LC-39A
  - VAFB SLC-4E
- **Mark the success/failed launches for each site on the map**
  - Success: Green mark
  - Fail: Red mark
- **Calculate the distances between a launch site to its proximities**
  - Closest coast line, railway, highway, and city

Folium map script: https://github.com/SeoyoungJeon/IBM_CHP10/blob/main/Week3_map.jupyterlite.ipynb

VISUALISATION - Dashboard with Plotly Dash -

- **Select launch site**
  - User allow to select a launch site (ALL vs. specific launch site) using the drop down function
- **Select Pay load mass**
  - User allow to select pay load mass (kg) by slider from 0 kg to 10,000 kg (step = 1,000 kg)
- **Pie chart**
  - When launch site is « ALL »: Pie chart of success rates by launch site
  - When specific launch site selected: Pie chart of ratio between success : fail
- **Scatter chart**
  - When launch site is « ALL »: Scatter chart of outcome by pay load mass (kg)
  - When specific launch site selected: Scatter chart of success possibility by pay load mass (kg)

Plotly dashboard script: https://github.com/SeoyoungJeon/IBM_CHP10/blob/main/Week3_dashboard.py


MODELING

MODELING - Classification -

- **Objective**
  - Find best hyperparameter to predict Launch success, using machine learning
- **Training & Test set**
  - Independent parameter dataset was normalized
  - Train & Test set split: Test size = 0.2, random state = 2
- **Machine learning**
  - CV = 10 was assumed in all approaches
  - Logistic regression, Support vector machine, Decision tree classifier, and K nearest neighbors
- **Evaluation**
  - Accuracy test using test set

Machine learning script: https://github.com/SeoyoungJeon/IBM_CHP10/blob/main/Week4_Machine_Learning_Pred.ipynb

RESULTS & CONCLUSION

01

Visualisation

- Success rate
- Yearly trend

02

SpaceX data

- Launch site
- Pay load
- Booster version
- Outcome

03

Launch site

- On the map

04

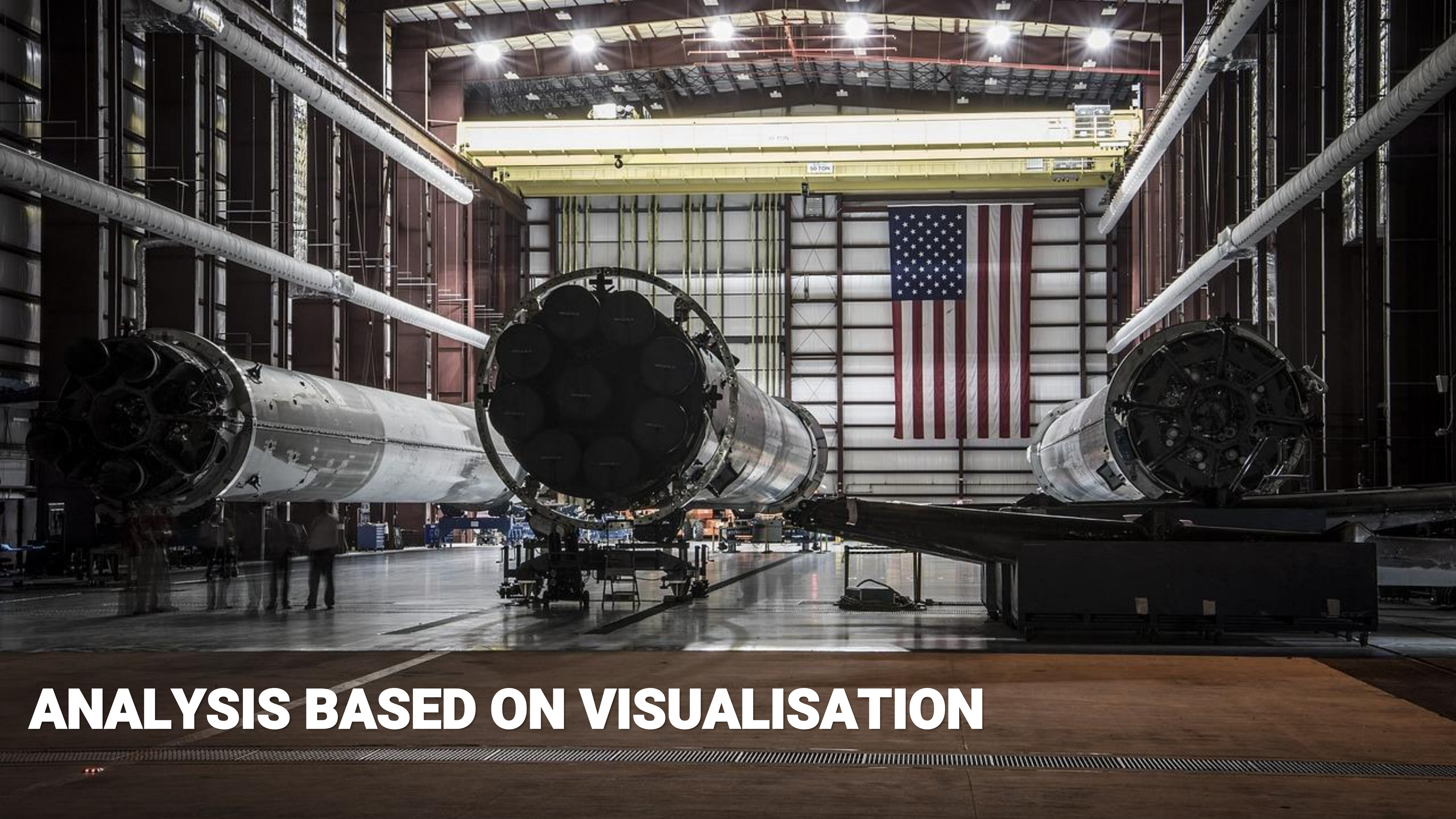
Dash board

- Success rate

05

Prediction

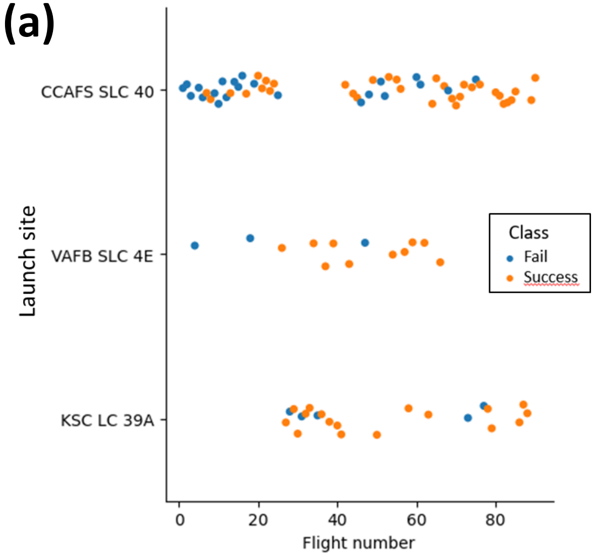
- Machine learning



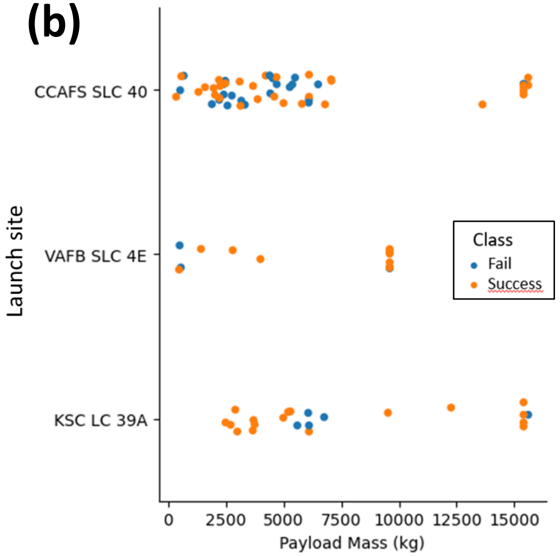
ANALYSIS BASED ON VISUALISATION

Success rate: Launch site, Orbit, Flight number, Pay load mass

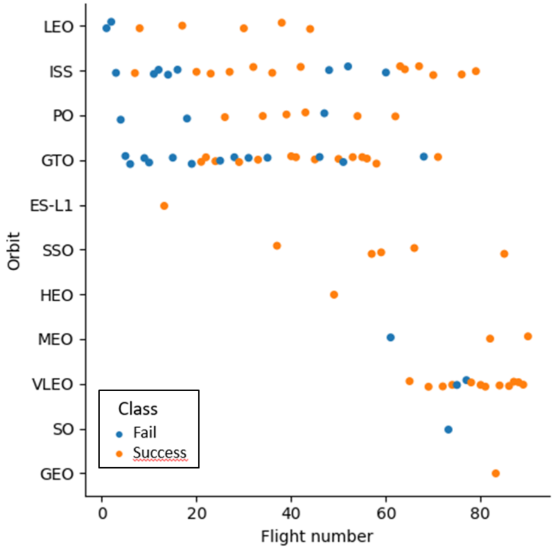
(a)



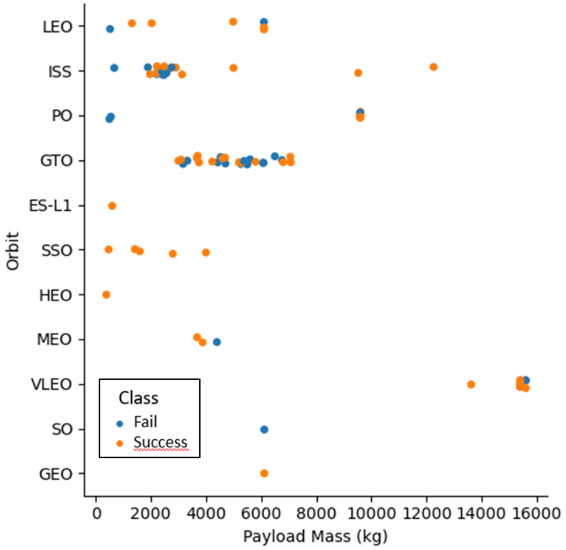
(b)



(c)



(d)



Launch site

There were the most attmeps at CCAFS SLC 40 site.

Flight number

The larger the flight number, the higher the success rate.

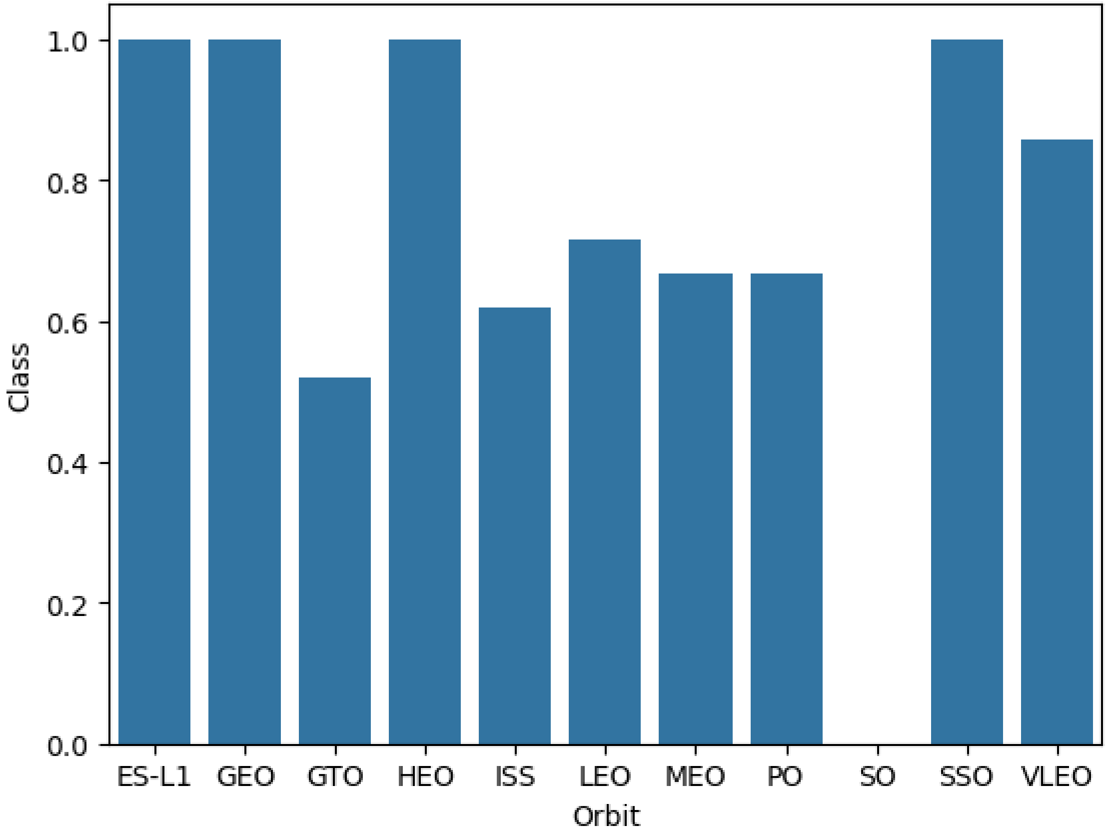
Pay load mass

- (< 8 tons) The difference in success rate according to pay load mass was not significant
- (≥ 8 tons) The heavier the pay load mass, the higher the success rate

Orbit

Success rate varied depending on the orbit type

Success rate: Orbit Type



Best success rate (= 100%)
ES-L1, GEO, HEO, SSO

High success rate (> 80%)
VLEO

Middle success rate (50 – 80%)
GTO, ISS, LEO, MEO, PO

Worst success rate (= 0%)
SO

Launch success yearly trend

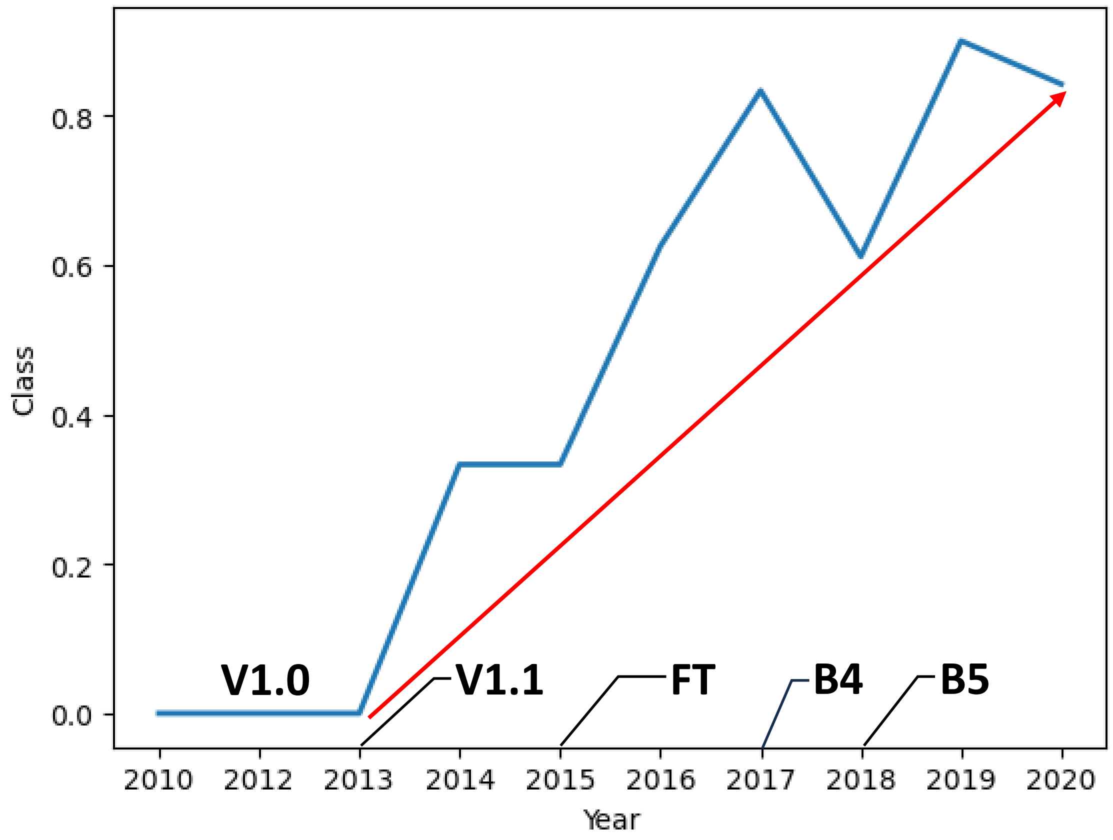


Figure 6. Yearly trend of launch success rate and booster version

2010 – 2020: Trend of **increasing** success rate

2010 – 2013: 0% success rate

2013 – 2014: **Increased** success rate

2014 – 2015: **Maintained** success rate

2015 – 2017: Success rate **increased** to 80%

2017 – 2018: Success rate **decreased** to 60%

2018 – 2019: Success rate **increased** to 90%

2019 – 2020: Success rate **decreased** to 80%

DATA LIST REQUESTED WITH SQL

Requested SpaceX data list

- EDA with SQL -
Data set about launch site

There are 4 launch sites

: CCAFS LC-40, VAFB SLC-4E, KSC LC-39A, CCAFS SLC-40

```
%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 data sets starting with CCA

: include data from CCAFS LS40 and SLC-40

```
%sql select * from SPACEXTABLE 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
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

Requested SpaceX data list

- EDA with SQL -
Pay load, Pay load mass (kg)

20 pay load launched by NASA (CRS)
: SpaceX CRS-1 ~ 18, 20 ~ 21

```
%sql select Payload,PAYLOAD_MASS__KG_, Customer from SPACEXTABLE where Customer == "NASA (CRS)"
```

* sqlite:///my_data1.db
Done.

Payload	PAYLOAD_MASS__KG_	Customer
SpaceX CRS-1	500	NASA (CRS)
SpaceX CRS-2	677	NASA (CRS)
SpaceX CRS-3	2296	NASA (CRS)
SpaceX CRS-4	2216	NASA (CRS)
SpaceX CRS-5	2395	NASA (CRS)
SpaceX CRS-6	1898	NASA (CRS)
SpaceX CRS-7	1952	NASA (CRS)
SpaceX CRS-8	3136	NASA (CRS)
SpaceX CRS-9	2257	NASA (CRS)
SpaceX CRS-10	2490	NASA (CRS)
SpaceX CRS-11	2708	NASA (CRS)
SpaceX CRS-12	3310	NASA (CRS)
SpaceX CRS-13	2205	NASA (CRS)
SpaceX CRS-14	2647	NASA (CRS)
SpaceX CRS-15	2697	NASA (CRS)
SpaceX CRS-16	2500	NASA (CRS)
SpaceX CRS-17, Starlink v0.9	2495	NASA (CRS)
SpaceX CRS-18, AMOS-17	2268	NASA (CRS)
SpaceX CRS-20, Starlink 5 v1.0	1977	NASA (CRS)
SpaceX CRS-21	2972	NASA (CRS)

Average pay load mass when booster version F9 v1.1
: 2,928 kg

```
%sql select avg(PAYLOAD_MASS__KG_) from SPACEXTABLE where Booster_Version == "F9 v1.1"
```

* sqlite:///my_data1.db
Done.

avg(PAYLOAD_MASS__KG_)
2928.4

Requested SpaceX data list

- EDA with SQL -
Booster version

There are 7 Boosters when Pay loass mass is 4 – 6 tons
: F9 B5 B1046.2, B1047.2, B1048.3, B1051.2, B1060.1, B1058.2, B1062.1

```
%sql select Booster_Version from SPACEXTABLE ₩
where Landing_Outcome == "Success" AND PAYLOAD_MASS__KG_ > 4000 AND PAYLOAD_MASS__KG_ < 6000
```

* sqlite:///my_data1.db
Done.

Booster_Version
F9 B5 B1046.2
F9 B5 B1047.2
F9 B5 B1048.3
F9 B5 B1051.2
F9 B5B1060.1
F9 B5 B1058.2
F9 B5B1062.1

There are 12 Boosters when Pay loass mass is max
: F9 B5 B1048.4, B1048.5, B1049.4, B1049.5, B1049.7, B1051.3, B1051.4, B1051.6, B1056.4, B1058.3, B1060.2, B1060.3

```
%sql select Booster_Version from SPACEXTABLE ₩
where PAYLOAD_MASS__KG_ == (select max(PAYLOAD_MASS__KG_) from SPACEXTABLE)
```

* sqlite:///my_data1.db
Done.

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

Requested SpaceX data list

- EDA with SQL - Mission / Landing outcome

There are 100 mission success and 1 failure

```
%sql select Mission_Outcome, count(*) from SPACEXTABLE GROUP BY Mission_Outcome
```

* sqlite:///my_data1.db
Done.

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

There were 2 landing failure in 2015

```
%sql select substr(Date,6,2) as month, Date, Landing_Outcome, Booster_Version, Launch_Site from SPACEXTABLE ₩
where substr(Date,0,5) == "2015" AND Landing_Outcome like "Failure%"
```

* sqlite:///my_data1.db
Done.

month	Date	Landing_Outcome	Booster_Version	Launch_Site
01	2015-01-10	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	2015-04-14	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

The first landing success was in 22/July/2018

```
%sql select * from SPACEXTABLE where Date = (select min(Date) from SPACEXTABLE where Landing_Outcome == "Success")
```

* sqlite:///my_data1.db
Done.

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2018-07-22	5:50:00	F9 B5B1047.1	CCAFS SLC-40	Telstar 19V	7075	GTO	Telesat	Success	Success

Ranking landing outcomes: 04/June/2010 – 20/March/2017

```
%sql select count(Landing_Outcome) as count_outcomes, Date, Landing_Outcome from SPACEXTABLE ₩
where Date between "2010-06-04" and "2017-03-20" group by Landing_Outcome order by count_outcomes
```

* sqlite:///my_data1.db
Done.

count_outcomes	Date	Landing_Outcome
1	2015-06-28	Precluded (drone ship)
2	2010-06-04	Failure (parachute)
2	2013-09-29	Uncontrolled (ocean)
3	2014-04-18	Controlled (ocean)
3	2015-12-22	Success (ground pad)
5	2015-01-10	Failure (drone ship)
5	2016-04-08	Success (drone ship)
10	2012-05-22	No attempt

Requested SpaceX data list

- EDA with SQL -
Mission / Landing outcome

There are 100 mission success and 1 failure

```
%sql select Mission_Outcome, count(*) from SPACEXTABLE GROUP BY Mission_Outcome
```

* sqlite:///my_data1.db
Done.

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

There were 2 landing failure in 2015

```
%sql select substr(Date,6,2) as month, Date, Landing_Outcome, Booster_Version, Launch_Site from SPACEXTABLE ₩
where substr(Date,0,5) == "2015" AND Landing_Outcome like "Failure%"
```

* sqlite:///my_data1.db
Done.

month	Date	Landing_Outcome	Booster_Version	Launch_Site
01	2015-01-10	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	2015-04-14	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

The first landing success was in 22/July/2018

```
%sql select * from SPACEXTABLE where Date = (select min(Date) from SPACEXTABLE where Landing_Outcome == "Success")
```

* sqlite:///my_data1.db
Done.

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2018-07-22	5:50:00	F9 B5B1047.1	CCAFS SLC-40	Telstar 19V	7075	GTO	Telesat	Success	Success

Ranking landing outcomes: 04/June/2010 – 20/March/2017

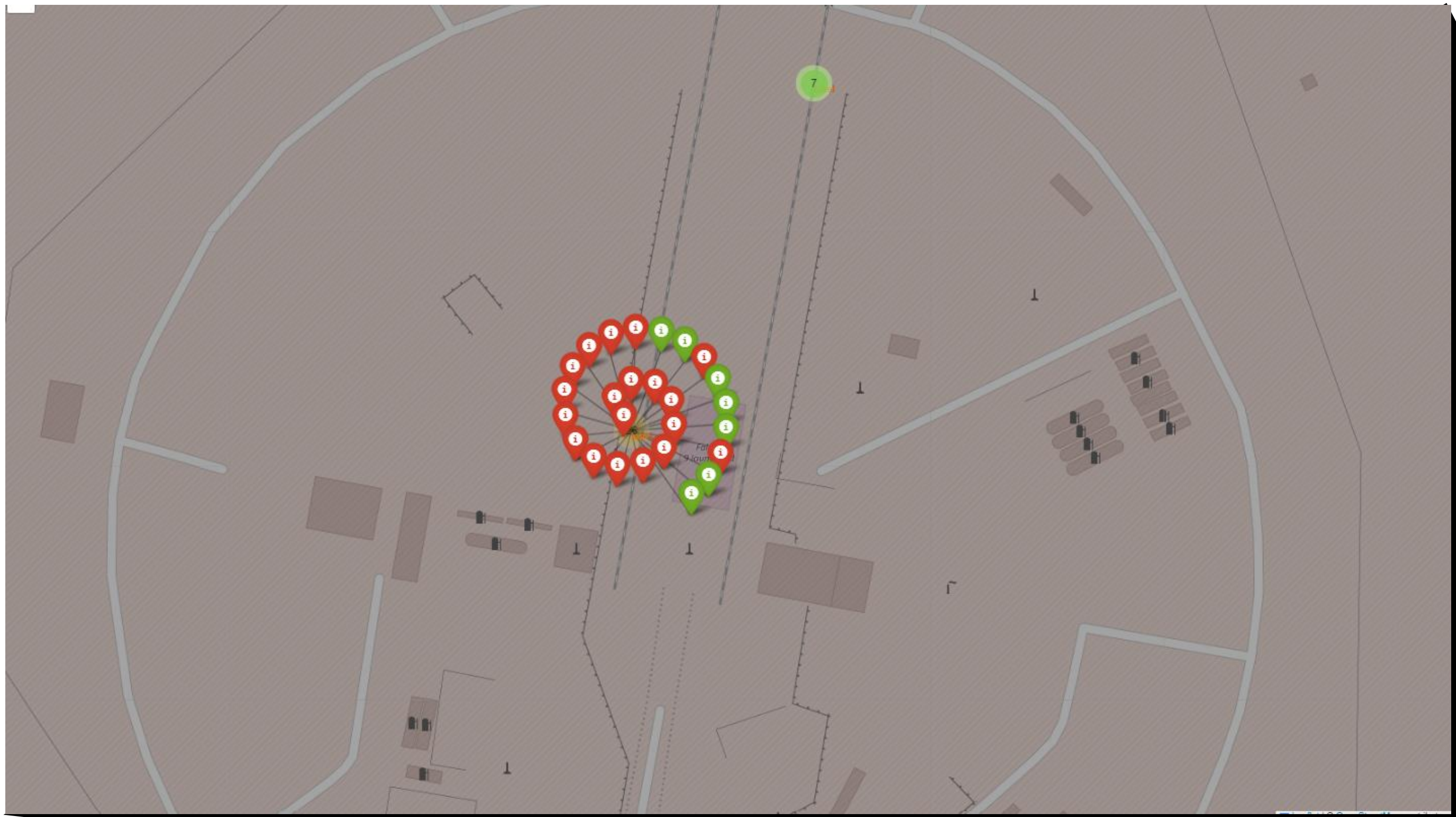
```
%sql select count(Landing_Outcome) as count_outcomes, Date, Landing_Outcome from SPACEXTABLE ₩
where Date between "2010-06-04" and "2017-03-20" group by Landing_Outcome order by count_outcomes
```

* sqlite:///my_data1.db
Done.

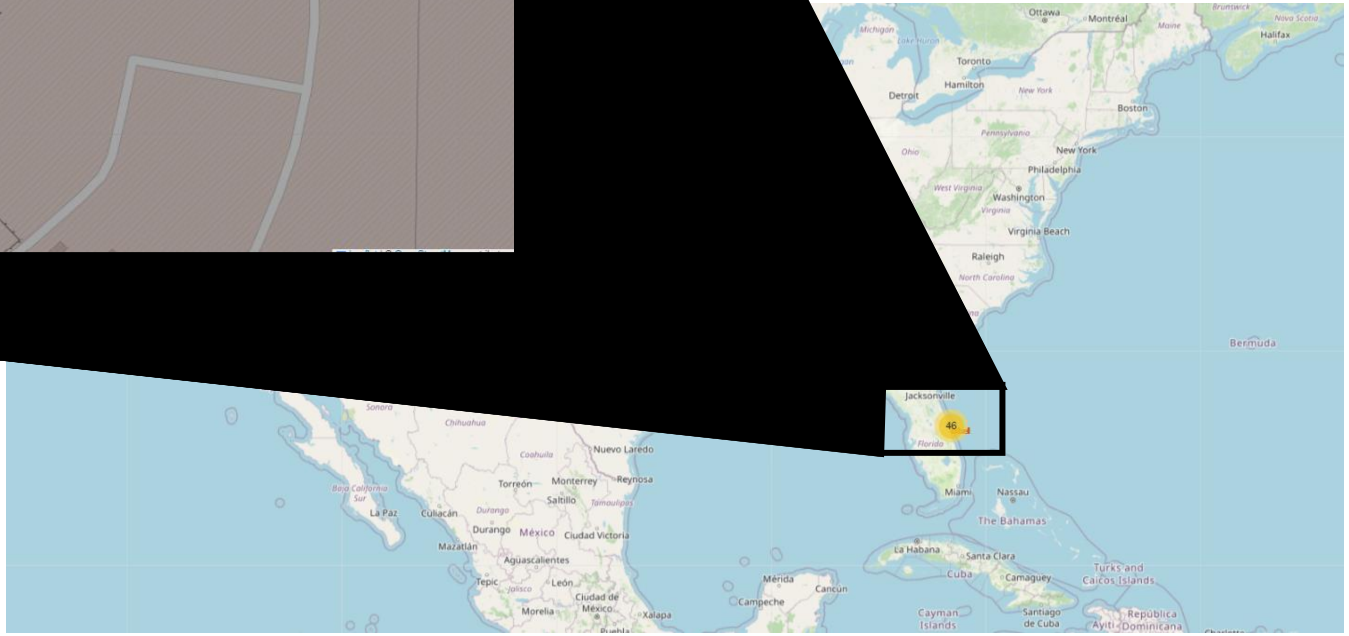
count_outcomes	Date	Landing_Outcome
1	2015-06-28	Precluded (drone ship)
2	2010-06-04	Failure (parachute)
2	2013-09-29	Uncontrolled (ocean)
3	2014-04-18	Controlled (ocean)
3	2015-12-22	Success (ground pad)
5	2015-01-10	Failure (drone ship)
5	2016-04-08	Success (drone ship)
10	2012-05-22	No attempt

LAUNCH SITES ON THE MAP

Launch sites

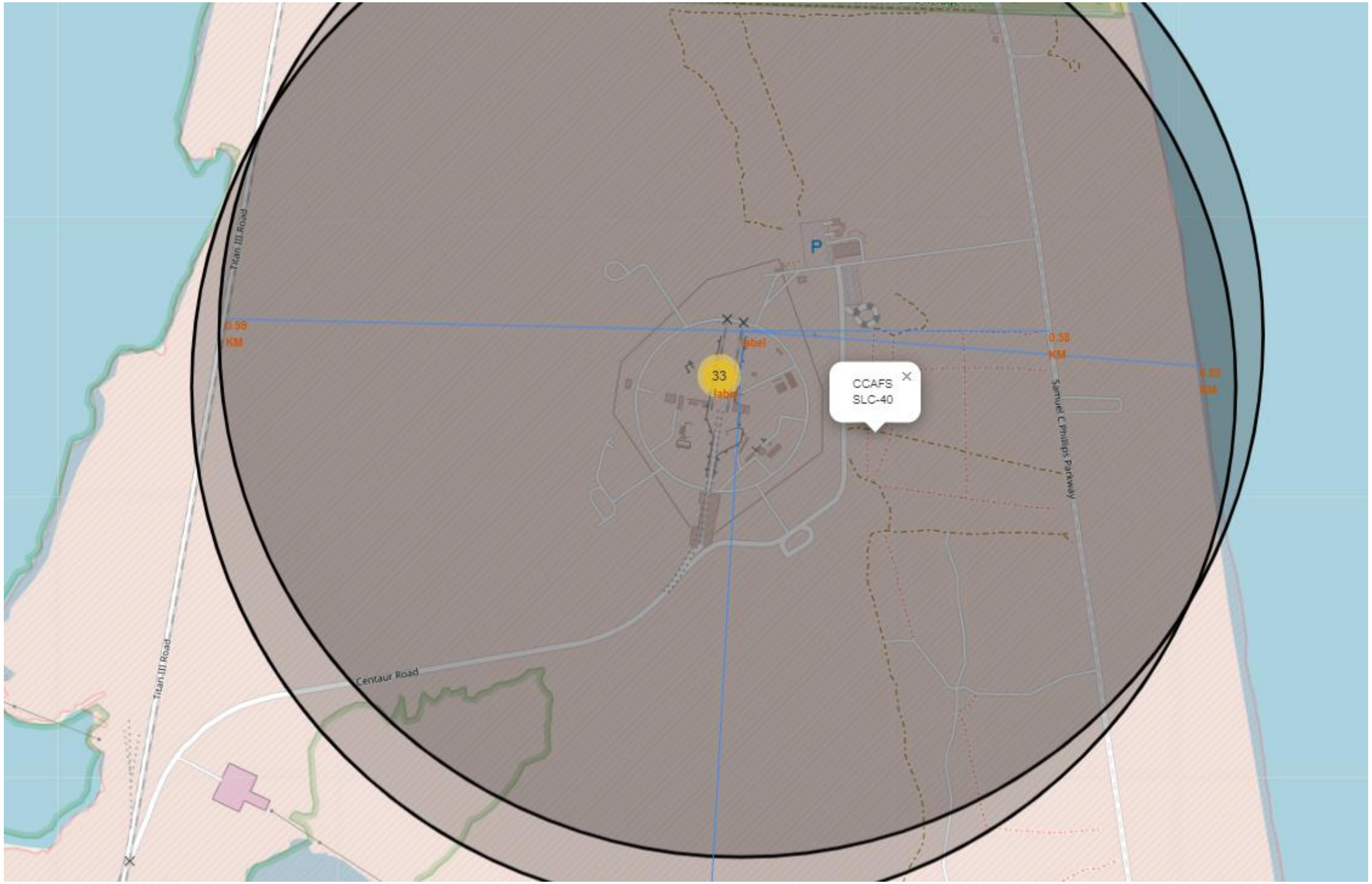


Green marker: Success
Red marker: Fail



More launch were attemped in Florida than LA

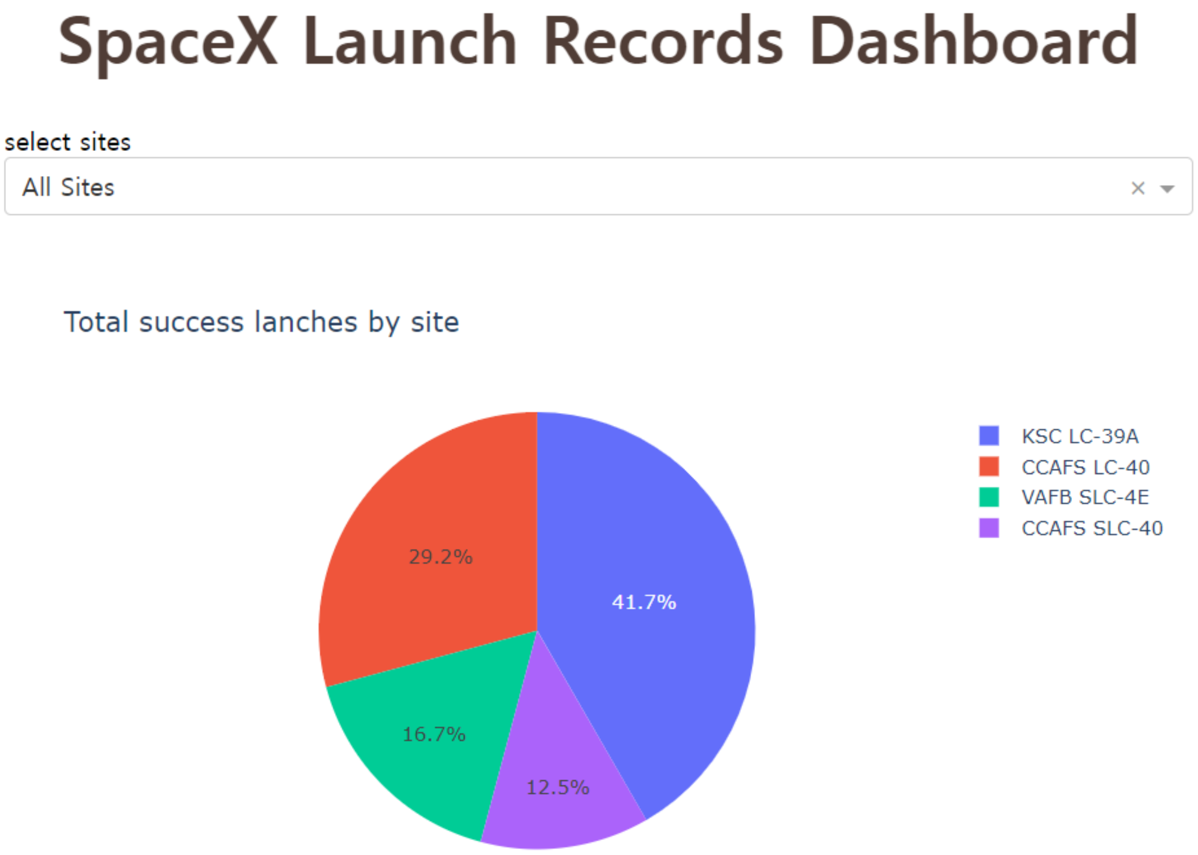
Launch sites

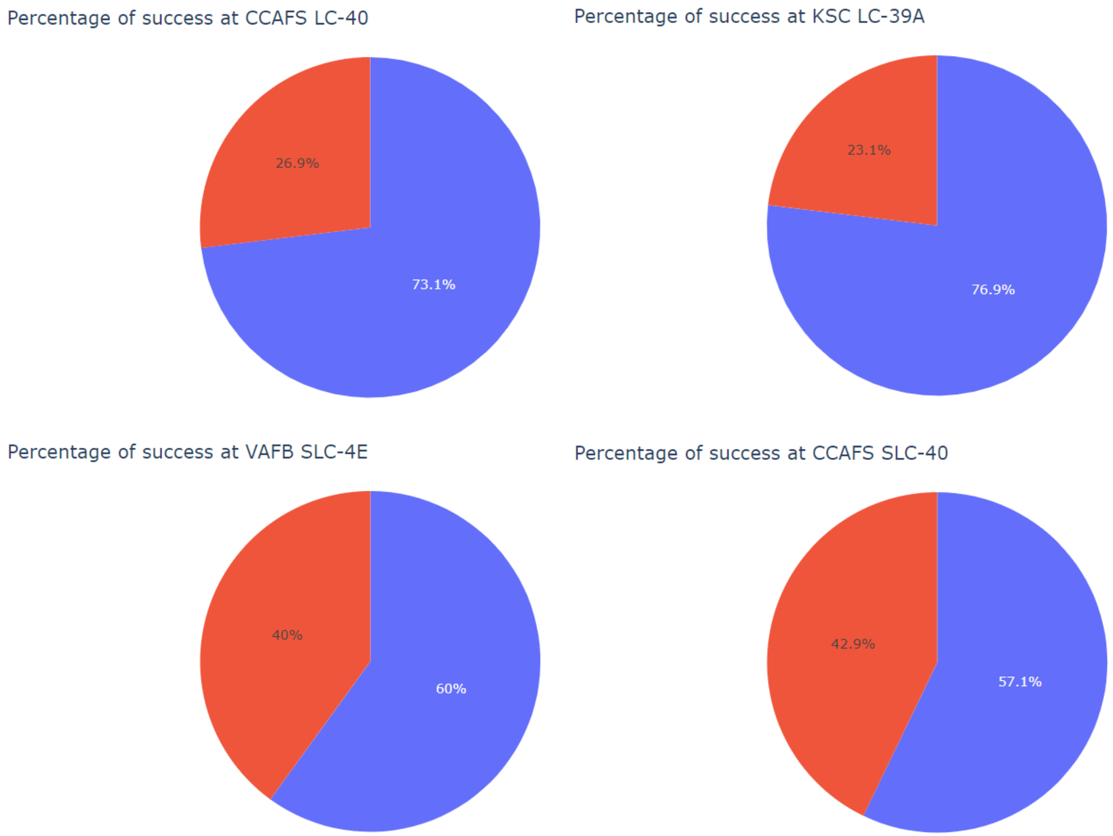


Closest coast line: 0.88 km
Closest railway: 0.99 km
Closest highway: 0.58 km
Melbourne: 54 km

DASHBOARD WITH PLOTLY

Launch success rate by Site





KSC LC-39A: The most successful launch site (41.7%)

CCAFS SLC-40: The least successful launch site (12.5%)

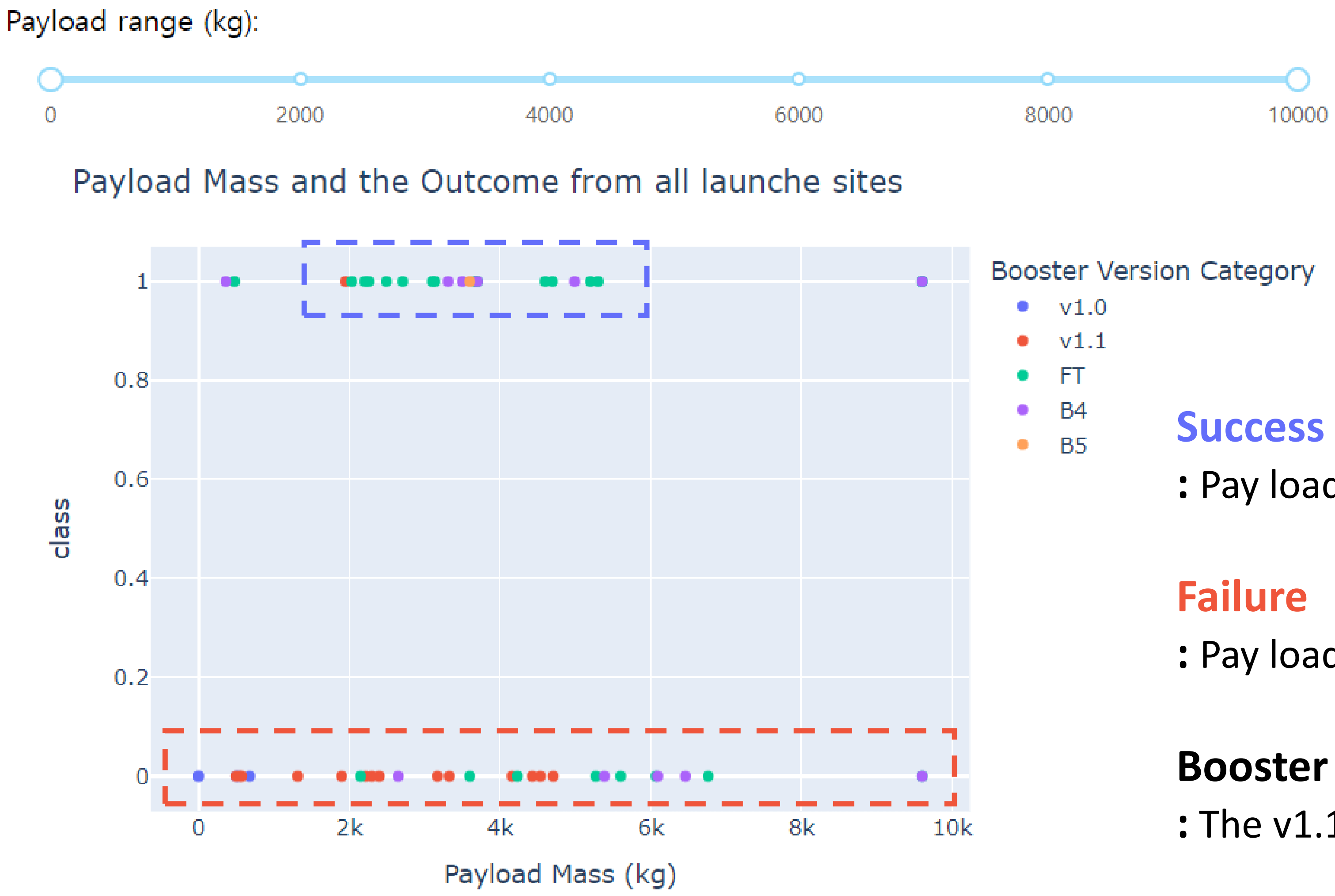
CCAFS LC-40: 73.1% success rate

KSC LC-39A [Best]: 76.9% success rate

VAFB SLC-4E: 60% success rate

CCAFS SLC-40 [Worst]: 57.1% success rate

Pay load mass and Success - By booster version -



Success

: Pay load mass is widely distributed between 2K and 6K

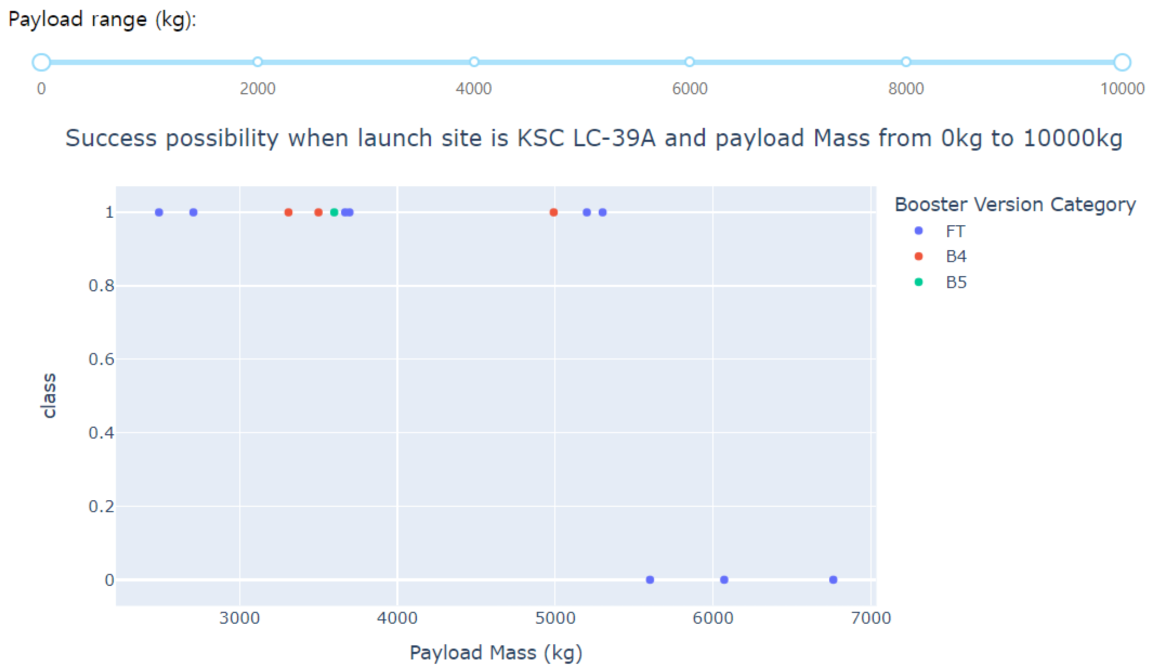
Failure

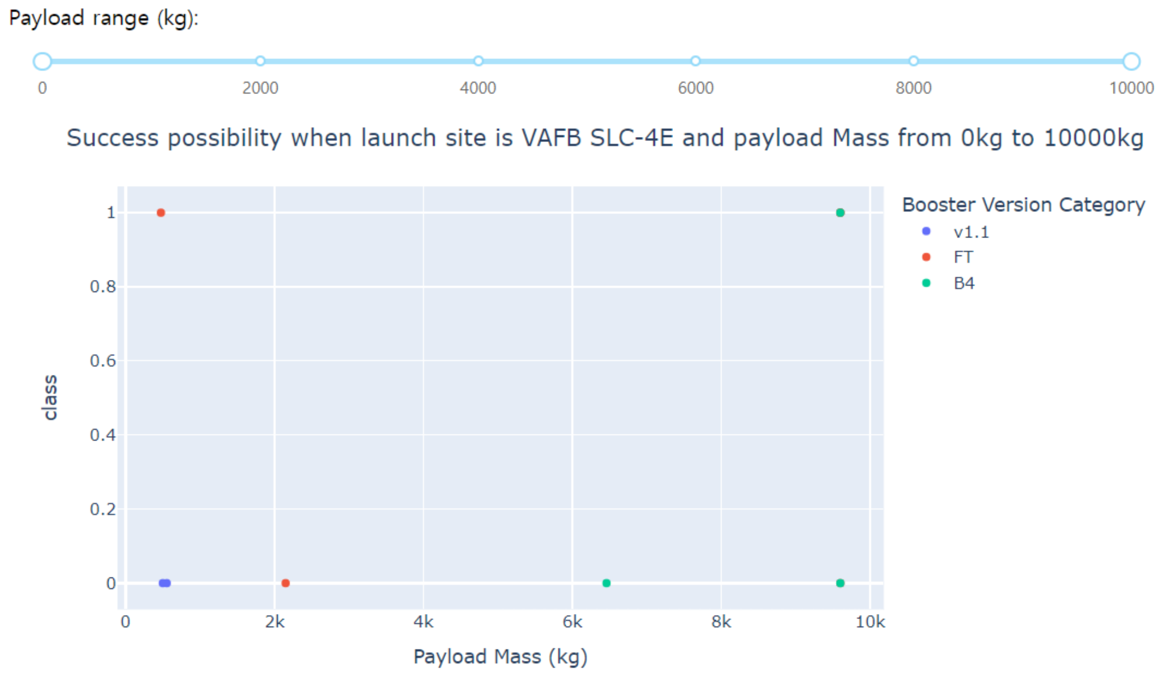
: Pay load mass is widely distributed across all sections

Booster version

: The v1.1 booster had more failures than successes

Pay load mass and Success - By booster version -





KSC LC-39A had highest sucess rate (76.9%)

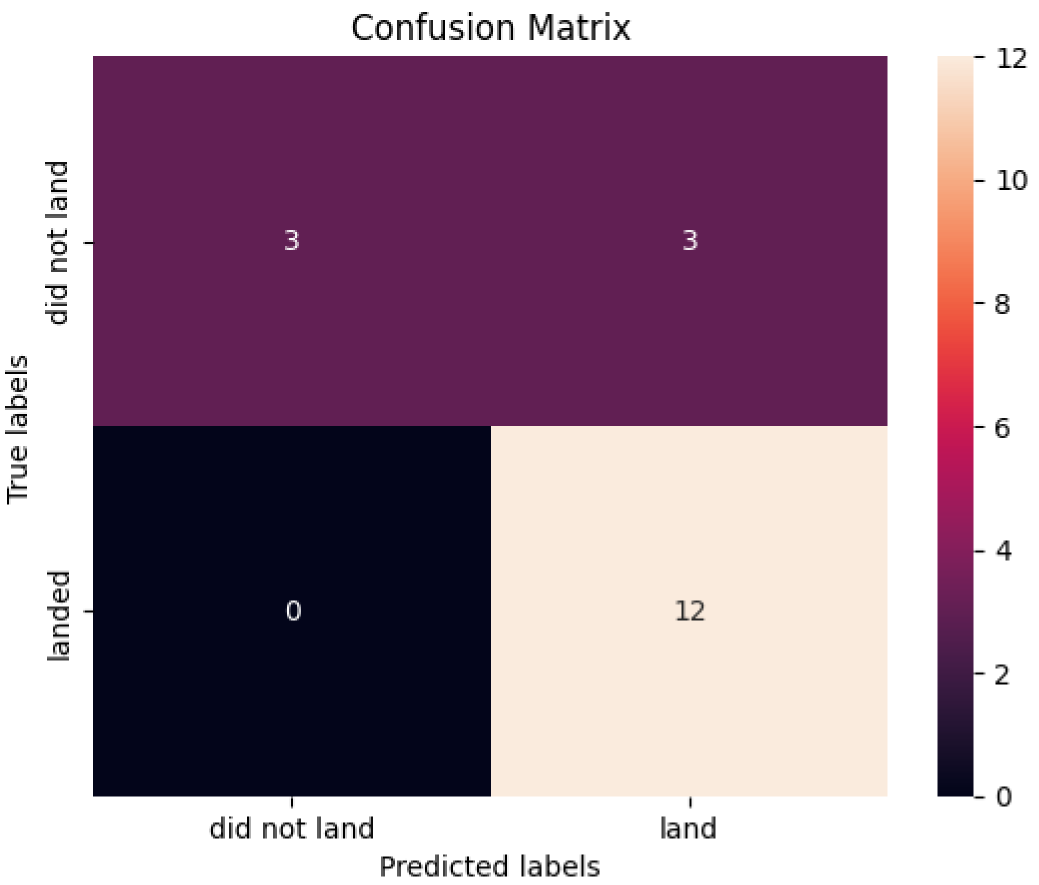
- All launches are successful if pay load mass is about 5.5K or less
- Booster FT with pay load mass greater than 5.5K fails to launch on every attempt

CCAFS LC-40 had lowest success rate (57.1%)

- Booster FT had a higher success rate
- Booster v1.1 had lower success rate

PREDICTIVE ANALYSIS
- MACHINE LEARNING -

Predictive analysis - Machine learning -




```
print("Logistic: ", acc_log, "₩n", "SCV: ", acc_scv, "₩n", "Tree: ", acc_tree, "₩n", "KNN: ", acc_knn)
```

```
Logistic:  0.8333333333333334
 SCV:  0.8333333333333334
 Tree:  0.8333333333333334
 KNN:  0.8333333333333334
```

Table 0. Best hyperparameters of each classifier

Logistic regression	Support vector machine
C: 0.01 Penalty: l2 Solver: lbfgs	C: 1.0 Gamma: 0.03 Kernal: Sigmoid
Decision tree classifier	**K nearest neighbors**
Criterion: gini Max depth: 6 Max features: sqrt Min samples leaf: 2 Min samples split : 5 Splitter: random	Algorithm: auto N neighbors: 10 P: 1

Accuracy was the same for all classifiers

CONCLUSION

Summary - Impact of each factor on launch success rate & Prediction -

Lauch site

- KSC LC-39A (77% success) vs. CCAFS LC-40 (57% success)
- Both locations are located in Florida ➔ Longitude and latitude seems not affected the launch outcome

Pay load mass

- All booster versions are designed to handle pay loads of up to 10 tons.
- However, from the FT version, it has been developed to be able to handle more weight than the previous version.
- In cases where the launce was successful, the pay load was more often less than 5.5 tons.
- But in the case of failure, the range of the pay load was evenly distributed.
- Thus, pay load weight itself does not seem to be a factor that determines the success or failure of the launch.

Booster version

- The initial versions (v1.0 and v1.1) had a higher failure probability than latest versions (FT, B4, and B5).
- The increase in success rate over time is thought to be related to the development of booster versions.

Orbit

- ES-L1, GEO, HEO, and SSO show a 100% success rate.

Prediction with machine learning classifier

- All classifier based machine learning technique applied to predict landing outcomes showed an accuracy of about 83%

Conclusion & Future study

Conclusion

- The influential factor for launch success rate of Falcon 9
  → Booster version & target orbit have high impact
  → Launch site & pay load mass have low impact

- Any classifier can be applied if the best hyperparameter is applied

Future study plan

: The following additional analysis is required to confirm the influence of each factor

1) Launch site: success rate for each launch site under the follow conditions
- Pay load mass < 5.5 tons
- Latest booster version (FT, B4, or B5)
- Targeted orbit with 100% success rate (ES-L1, GEO, HEO, and SSO)

2) Pay load mass
- Influence of pay load mass on the same target orbit (ES-L1, GEO, HEO, or SSO) with the same booster (FT, B4 or B5)
- Analysis of whether there is a limit to the appropriate pay load mass depending on the target orbit