EXECUTIVE REPORT

“Predicting the outcome of SpaceX Falcon9 first stage landing”

Juan Francisco Velázquez Vadillo
SPACE Y
Data Scientist
November 2022

Executive Summary

- ❖ **The race to make space travel affordable for every one is here**.
- ❖ Some providers advertise a cost upward of **165 million dollars**.
- ❖ **SpaceX** advertises **Falcon 9** rocket launches with a cost of **62 million dollars.**
- ❖ This **reduced cost** is mainly due to the fact that **SpaceX reuses the first stage** of Falcon 9 rocket.

- Through the use of **Data Science** and **Machine Learning** tools we developed **models that predict the outcome (success vs failure) of Falcon9 first stage landing.**

- This information will allow us to **estimate the cost** of a Falcon9 launch.

Knowing the price of each Falcon9 launch will **guide our strategies to bid against SpaceX**

Table of Contents

Slide

Table of Contents

Slide

Table of Contents | Slide

Introduction

Scenario

- Our company, **SpaceY**, wants to **lead the race to make space travel affordable for every one** .
- **Virgin Galactic**, **Rocket Lab**, **Blue Origin** and **SpaceX** are heavily investing in making space travel affordable for every one.
- By reusing the “first stage” of the Falcon9 rocket **SpaceX** can offer the **“cheapest” ticket to space**.

Introduction

Goal

Accurately predict SpaceX Falcon9 first stage landing outcome.

Strategy

Using **Data Science** and **Machine Learning tools** we developed models that accurately predict the outcome (success vs failure) of the Falcon9 first stage landing.

Falcon 9 overview / Successful first stage landing

https://www.youtube.com/watch?v=Z4TXCZG_NEY

Capstone Project overview

Collect data about SpaceX past launches

SpaceX REST API

Webscrape data from a Wikipage

Data cleaning and wrangling

Interactive visual analytics

Exploratory Data Analysis and Feature Engineering using visualization tools
(Identify variables relevant to launch outcome)

Exploratory Data Analysis using SQL
(Get info about launch sites, rocket´s characteristics and launch outcome)

Build a Plotly Dashboard app
(See SpaceX launch data analysis in real-time)

Launch sites location analysis with Folium
(Importance of the location and proximities to a launch site to launch outcome)

Machine Learning Prediction of SpaceX Falcon9 first stage landing

Methodology

Here you can find the information used to create this report:

Data Collection from SpaceX REST API

https://nbviewer.org/github/VVJF/Coursera-IBM-Capstone-Project-2022/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

Data Collection from Wikipedia

https://nbviewer.org/github/VVJF/Coursera-IBM-Capstone-Project-2022/blob/main/jupyter-labs-webscraping.ipynb

Data Wrangling

https://nbviewer.org/github/VVJF/Coursera-IBM-Capstone-Project-2022/blob/main/IBM-DS0321EN-SkillsNetwork_labs_module_1_L3_labs-jupyter-spacex-data_wrangling_jupyterlite.jupyterlite.ipynb

Exploratory Data Analysis using SQL

https://nbviewer.org/github/VVJF/Coursera-IBM-Capstone-Project-2022/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Exploring and Preparing Data for EDA viz

https://nbviewer.org/github/VVJF/Coursera-IBM-Capstone-Project-2022/blob/main/IBM-DS0321EN-SkillsNetwork_labs_module_2_jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb

Launch Sites Locations Analysis with Folium

https://nbviewer.org/github/VVJF/Coursera-IBM-Capstone-Project-2022/blob/main/IBM-DS0321EN-SkillsNetwork_labs_module_3_lab_jupyter_launch_site_location.jupyterlite%20%281%29.ipynb

Make a Dashboard with PlotlyDash

https://nbviewer.org/github/VVJF/Coursera-IBM-Capstone-Project-2022/blob/main/spacex_dash.py

Machine Learning Prediction lab

Assignment: Machine Learning Prediction

https://nbviewer.org/github/VVJF/Coursera-IBM-Capstone-Project-2022/blob/main/IBM-DS0321EN-SkillsNetwork_labs_module_4_SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

Briefly:
Data collection about SpaceX Falcon9 past launches

Open Source SpaceX REST API

(https://api.spacexdata.com/v4/launches/past)

Wikipedia

page titled "**List of Falcon 9 and Falcon Heavy launches**"

https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches

Data cleaning and wrangling

As described in the jupyter notebooks “Data Collection from SpaceX REST API”, “Data Collection from Wikipedia”, and “Data Wrangling”:

- We went through the process of fixing or eliminating incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within the datasets downloaded from SpaceX REST API and Wikipedia page “List of Falcon 9 and Falcon Heavy launches”.

- We found patterns in the data that helped determine what would be the label and relevant variables for training Machine Learning supervised models.

Predictive Analysis:

We went through the process of building a machine learning pipeline to predict if the first stage of the Falcon 9 lands successfully.

Data was **standardized** and **split** into **training and testing data**.

Grid Search was performed on the trained data to find the **hyperparameters** that allow a given Machine Learning model to perform best.

Predictive Analysis:

Machine Learning supervised learning techniques used:

Logistic Regression

Support Vector Machine

Decision Trees

K Nearest Neighbors

K nearest neighbors

We **output the confusion matrix** and determined which model best predicts the outcome of each Falcon9 launch.

RESULTS
Data wrangling

Dataset was imported from
URL = 'https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/dataset_part_1.csv‘.

We identified and calculated the percentage of the missing values in each attribute.

FlightNumber	0.000
Date	0.000
BoosterVersion	0.000
PayloadMass	0.000
Orbit	0.000
LaunchSite	0.000
Outcome	0.000
Flights	0.000
GridFins	0.000
Reused	0.000
Legs	0.000
LandingPad	40.625
Block	0.000
ReusedCount	0.000
Serial	0.000
Longitude	0.000
Latitude	0.000

Data wrangling

We calculated the number of launches on each site.

```
CCAFS SLC 40     55
KSC LC 39A       22
VAFB SLC 4E      13
```

We calculated the number and occurrence of each orbit.

```
Number of each orbit:  GTO      27
ISS       21
VLEO      14
PO         9
LEO        7
SSO        5
MEO        3
ES-L1      1
HEO        1
SO         1
GEO        1
Name: Orbit, dtype: int64
Occurrence of each orbit:
 GTO       30.000000
ISS       23.333333
VLEO      15.555556
PO        10.000000
LEO        7.777778
SSO        5.555556
MEO        3.333333
ES-L1      1.111111
HEO        1.111111
SO         1.111111
GEO        1.111111
Name: Orbit, dtype: float64
```

RESULTS

Data wrangling

We calculated the number and occurrence of mission outcome per orbit type.

True ASDS	41
None None	19
True RTLS	14
False ASDS	6
True Ocean	5
False Ocean	2
None ASDS	2
False RTLS	1

We created a landing outcome label from Outcome column.

Successful landing: 0
Failed landing: 1

	Class
0	0
1	0
2	0
3	0
4	0
5	0
6	1
7	1

We calculated the successful launch rate: 0.666

RESULTS

Dataframe generated from Data wrangling

Dataframe generated in the Data Wrangling laboratory.

Dataframe used in Machine Learning Prediction

```
df.head(5)
```

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude	Class
0	1	2010-06-04	Falcon 9	6104.959412	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0003	-80.577366	28.561857	0
1	2	2012-05-22	Falcon 9	525.000000	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0005	-80.577366	28.561857	0
2	3	2013-03-01	Falcon 9	677.000000	ISS	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0007	-80.577366	28.561857	0
3	4	2013-09-29	Falcon 9	500.000000	PO	VAFB SLC 4E	False Ocean	1	False	False	False	NaN	1.0	0	B1003	-120.610829	34.632093	0
4	5	2013-12-03	Falcon 9	3170.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B1004	-80.577366	28.561857	0

Data Collection from Wikipedia

Clean dataframe

Dataframe used in Exploratory Data Analysis using SQL

In [20]:
```
df.head()
```

Out[20]:

	Flight No.	Launch site	Payload	Payload mass	Orbit	Customer	Launch outcome	Version Booster	Booster landing	Date	Time
0	1	CCAFS	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success\n	F9 v1.0B0003.1	Failure	4 June 2010	18:45
1	2	CCAFS	Dragon	0	LEO	NASA (COTS)\nNRO	Success	F9 v1.0B0004.1	Failure	8 December 2010	15:43
2	3	CCAFS	Dragon	525 kg	LEO	NASA (COTS)	Success	F9 v1.0B0005.1	No attempt\n	22 May 2012	07:44
3	4	CCAFS	SpaceX CRS-1	4,700 kg	LEO	NASA (CRS)	Success\n	F9 v1.0B0006.1	No attempt	8 October 2012	00:35
4	5	CCAFS	SpaceX CRS-2	4,877 kg	LEO	NASA (CRS)	Success\n	F9 v1.0B0007.1	No attempt\n	1 March 2013	15:10

Exploratory Data Analysis using SQL

Examples of information gathered:

- Names of the unique launch sites in the space mission.
- Display the number of successful and failed missions.
- List the date when the first successful landing outcome in ground pad was achieved.
- Rank the count of successful landing_outcomes between the date 04-06-2010 and 20-03-2017 in descending order.

Exploratory Data Analysis using SQL

1.- Display the names of the unique launch sites in the space mission:

```
[10]: %sql select distinct "Launch_Site" from SPACEXTBL;

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

[10]:

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

2.- Display the number of successful and failed missions

```
%sql select count("Mission_Outcome") as "Number_of_successful_missions" from SPACEXTBL where "Mission_Outcome" like "Succ%";

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

Number_of_successful_missions
100

```
%sql select count("Mission_Outcome") as "Number_of_failed_missions" from SPACEXTBL where "Mission_Outcome" like "Fail%";

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

Number_of_failed_missions
1

Exploratory Data Analysis using SQL

3.- Display the date of the first successful landing outcome in ground pad.

```
%sql select min("Date") as "First_successful_landing_outcome", "Landing _Outcome" as "Ground-pad" from SPACEXTBL where "Landing
```

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

First_successful_landing_outcome	Ground-pad
01-05-2017	Success (ground pad)

Exploratory Data Analysis using SQL

4.- Display 5 records where launch sites begin with the string 'CCA':

```
[11]: %sql select "Launch_Site" from SPACEXTBL where "Launch_Site" like 'CCA%' limit 5;
 * sqlite:///my_data1.db
Done.
```

[11]:

Launch_Site
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40

5.- Display the total payload mass carried by boosters launched by NASA (CRS):

```
[12]: %sql select  "PAYLOAD_MASS__KG_", "Customer" from SPACEXTBL where "Customer" = "NASA (CRS)" limit 5;
 * sqlite:///my_data1.db
Done.
```

[12]:

PAYLOAD_MASS__KG_	Customer
500	NASA (CRS)
677	NASA (CRS)
2296	NASA (CRS)
2216	NASA (CRS)
2395	NASA (CRS)

Exploratory Data Analysis using SQL

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

```
[13]: %sql select  AVG("PAYLOAD_MASS__KG_") as "AVERAGE_PAYLOADMASS_KG", "Booster_Version" from SPACEXTBL where "Booster_Version" = "

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

[13]:

AVERAGE_PAYLOADMASS_KG	Booster_Version
2928.4	F9 v1.1

7.- List the date when the first successful landing outcome in ground pad was achieved.

```
[15]: %sql select min("Date") as "First_successful_landing_outcome", "Landing _Outcome" as "Ground-pad" from SPACEXTBL where "Landing

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

[15]:

First_successful_landing_outcome	Ground-pad
01-05-2017	Success (ground pad)

Exploratory Data Analysis using SQL

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

```
[16]: %sql select "Booster_Version", "PAYLOAD_MASS__KG_", "Landing _Outcome" from SPACEXTBL where ("PAYLOAD_MASS__KG_" BETWEEN 4000 A
```

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

[16]:

Booster_Version	PAYLOAD_MASS__KG_	Landing _Outcome
F9 FT B1022	4696	Success (drone ship)
F9 FT B1026	4600	Success (drone ship)
F9 FT B1021.2	5300	Success (drone ship)
F9 FT B1031.2	5200	Success (drone ship)

9.- List the total number of successful and failure mission outcomes.

```
[18]: %sql select count("Mission_Outcome") as "Number_of_successful_missions" from SPACEXTBL where "Mission_Outcome" like "Succ%";
```

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

[18]:

Number_of_successful_missions
100

```
[19]: %sql select count("Mission_Outcome") as "Number_of_failed_missions" from SPACEXTBL where "Mission_Outcome" like "Fail%";
```

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

[19]:

Number_of_failed_missions
1

Exploratory Data Analysis using SQL

10.- Names of the booster_versions which have carried the maximum payload mass.

```
[20]: %sql select "Booster_Version", "PAYLOAD_MASS__KG_" as "Carrying_maximum_payload_mass_kg" from SPACEXTBL where "PAYLOAD_MASS__KG

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

[20]:

Booster_Version	Carrying_maximum_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

Exploratory Data Analysis using SQL

11.- Records which will display the month names, failure landing_outcomes in drone ship, booster versions, launch_site for the months in year 2015.

```
[128]: %sql select  substr("Date", 4, 2) as "Month_of_2015", "Landing _Outcome", "Booster_Version", "Launch_Site"  from SPACEXTBL wher

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

[128]:

Month_of_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

12.- Rank the count of successful landing_outcomes between the date 04-06-2010 and 20-03-2017 in descending order.

```
[150]: %sql select  "Landing _Outcome", count(*) from SPACEXTBL where ("Date" between "04-06-2010" and "20-03-2017") AND "Landing _Out

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

[150]:

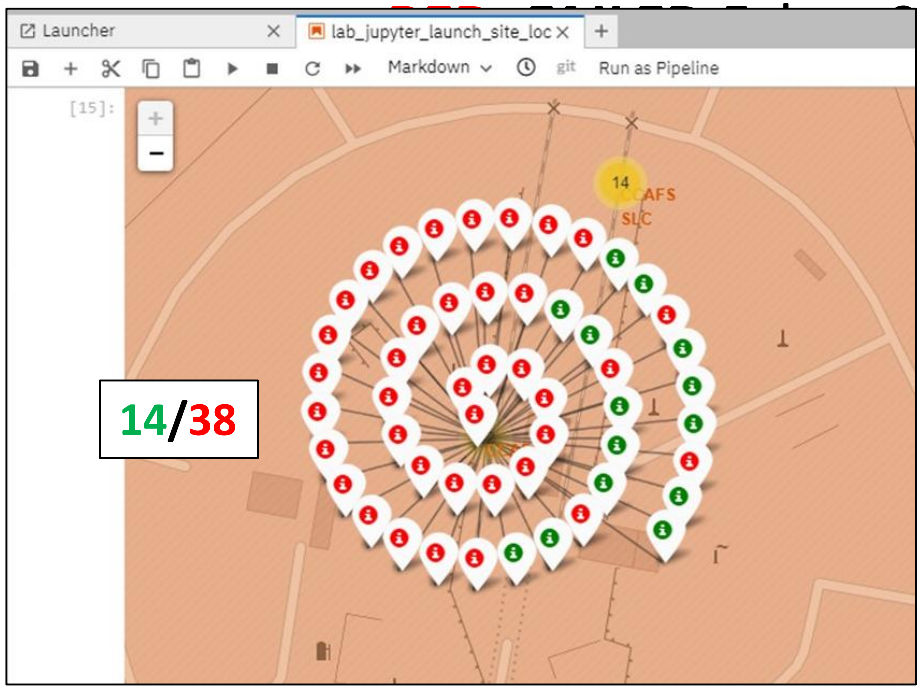
Landing _Outcome	count(*)
Success	20
Success (drone ship)	8
Success (ground pad)	6

Interactive map: SpaceX launch sites

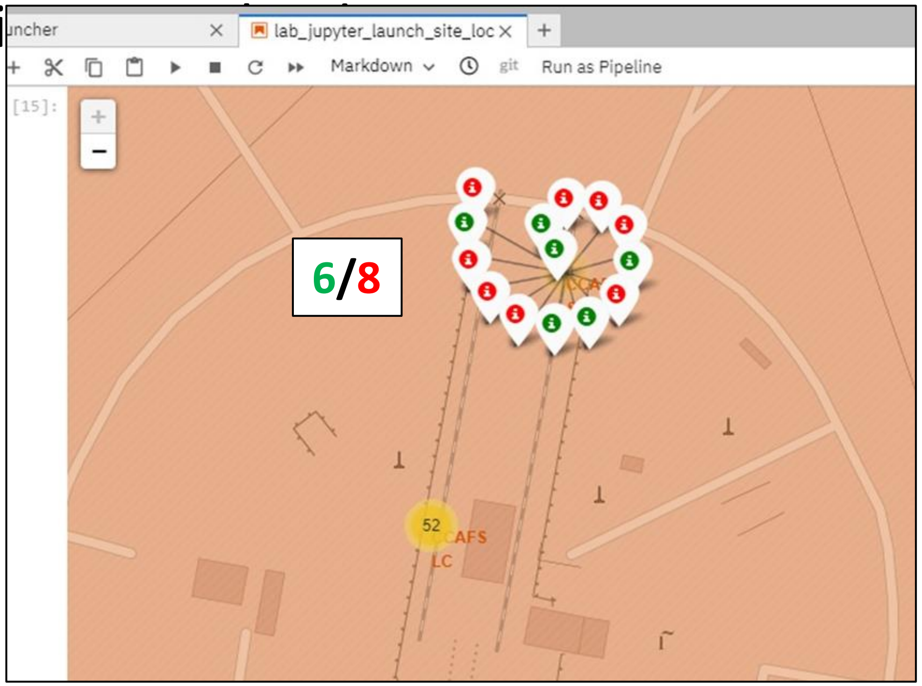
Launch sites:
VAFB SLC-4E
KSC LC-39A
CCAFS LC-40
CCAFS SLC-40

Interactive map: Falcon9 first stage landing outcome per launch site

GREEN: **SUCCESSFUL** Falcon9 first stage landing



CCAFS LC-40



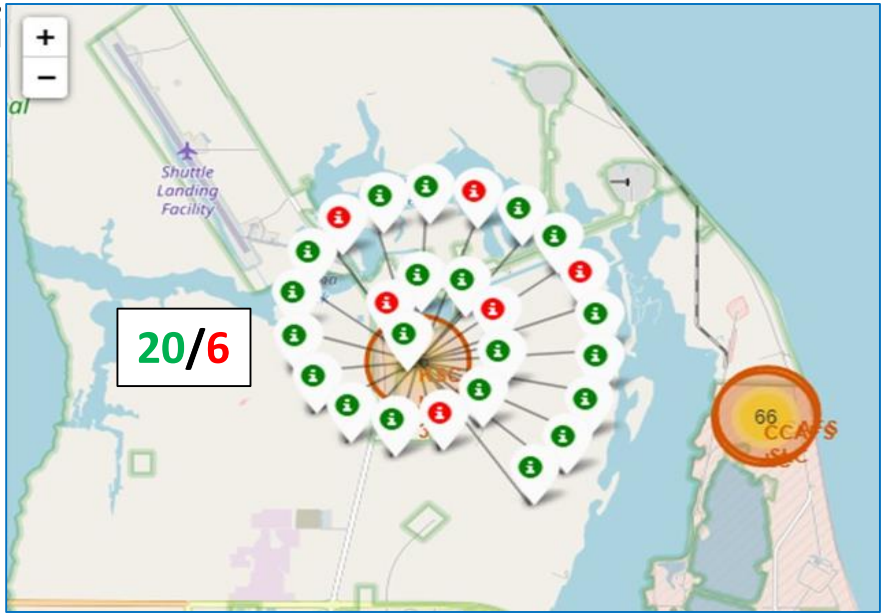
CCAFS SLC-40

Interactive map: Falcon9 first stage landing outcome per launch site

GREEN: **SUCCESSFUL** Falcon9 first stage landing

9 fi

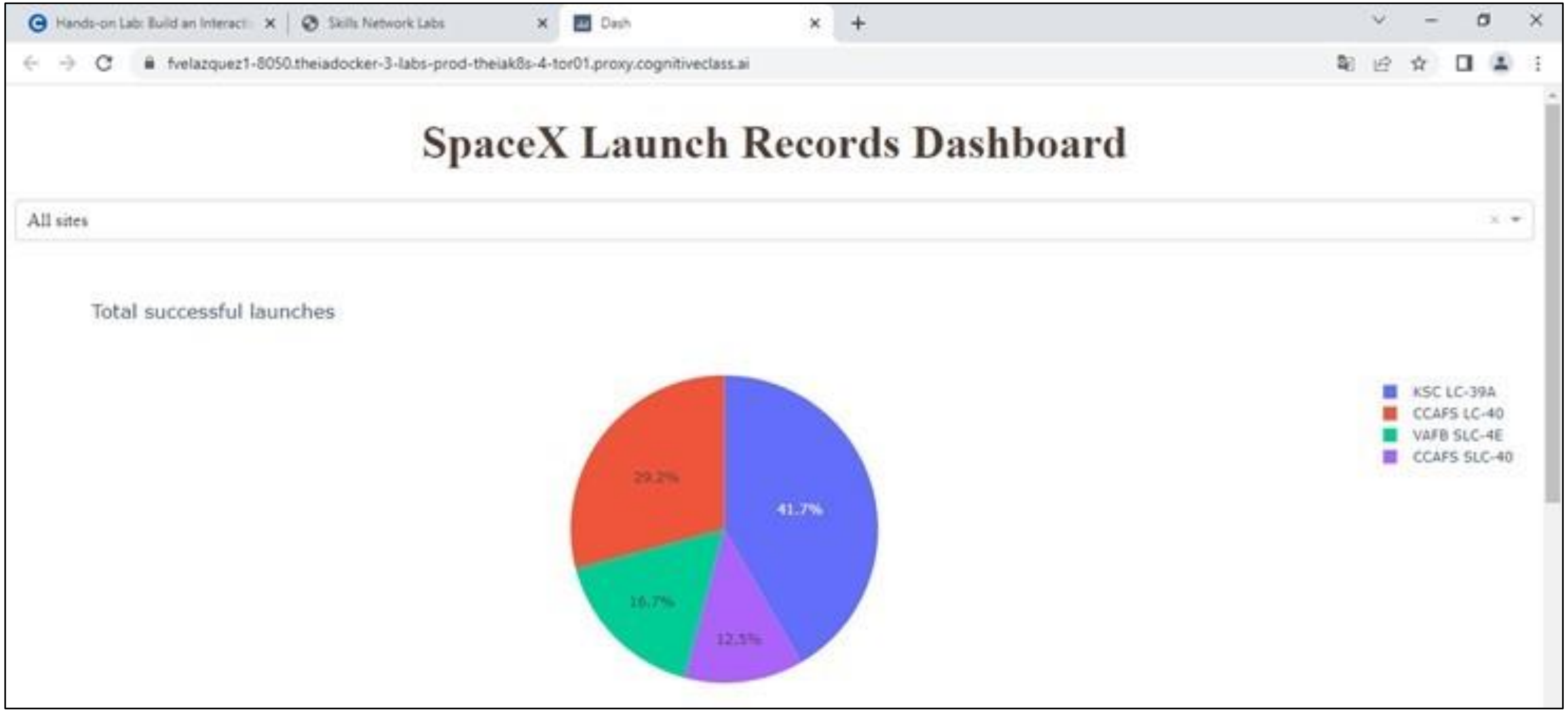
VAFB SLC 4E



KSC LC-39A

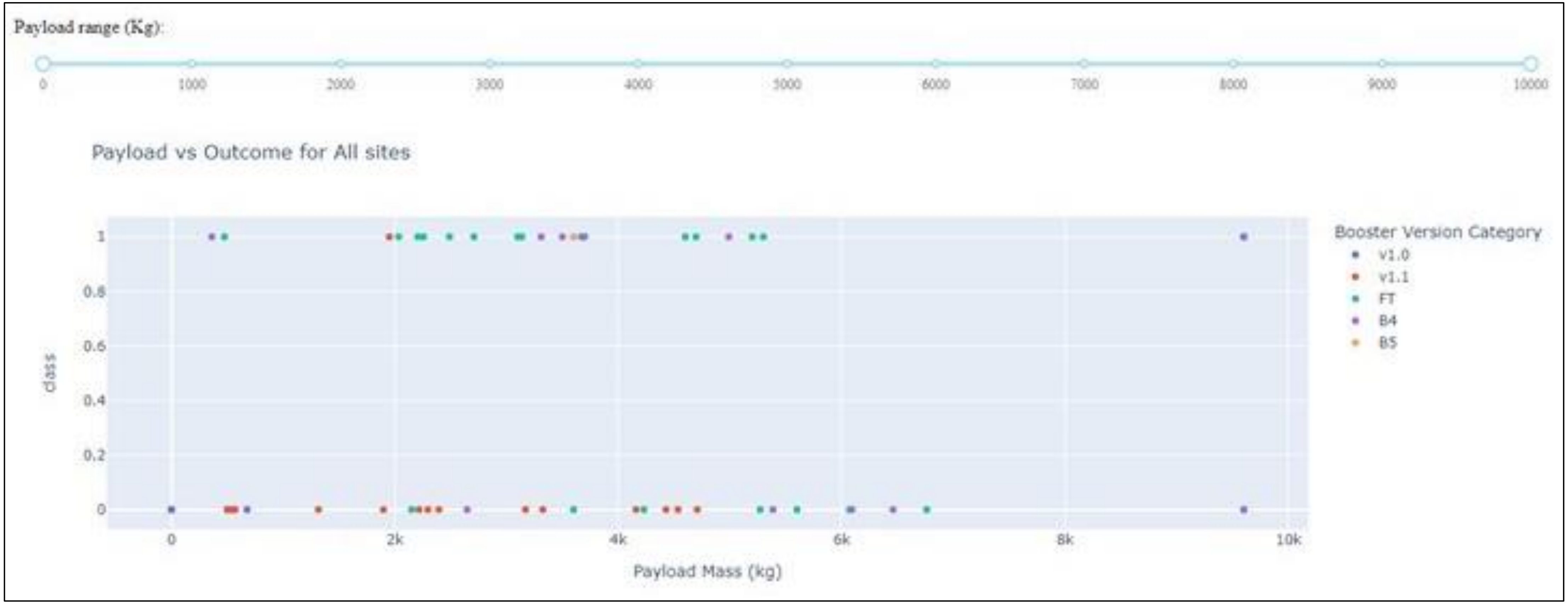
Interactive Dashboard

KSC LC-39A has the **highest number of successful** Falcon9 first stage landings among all four launch sites: **41.7%.**



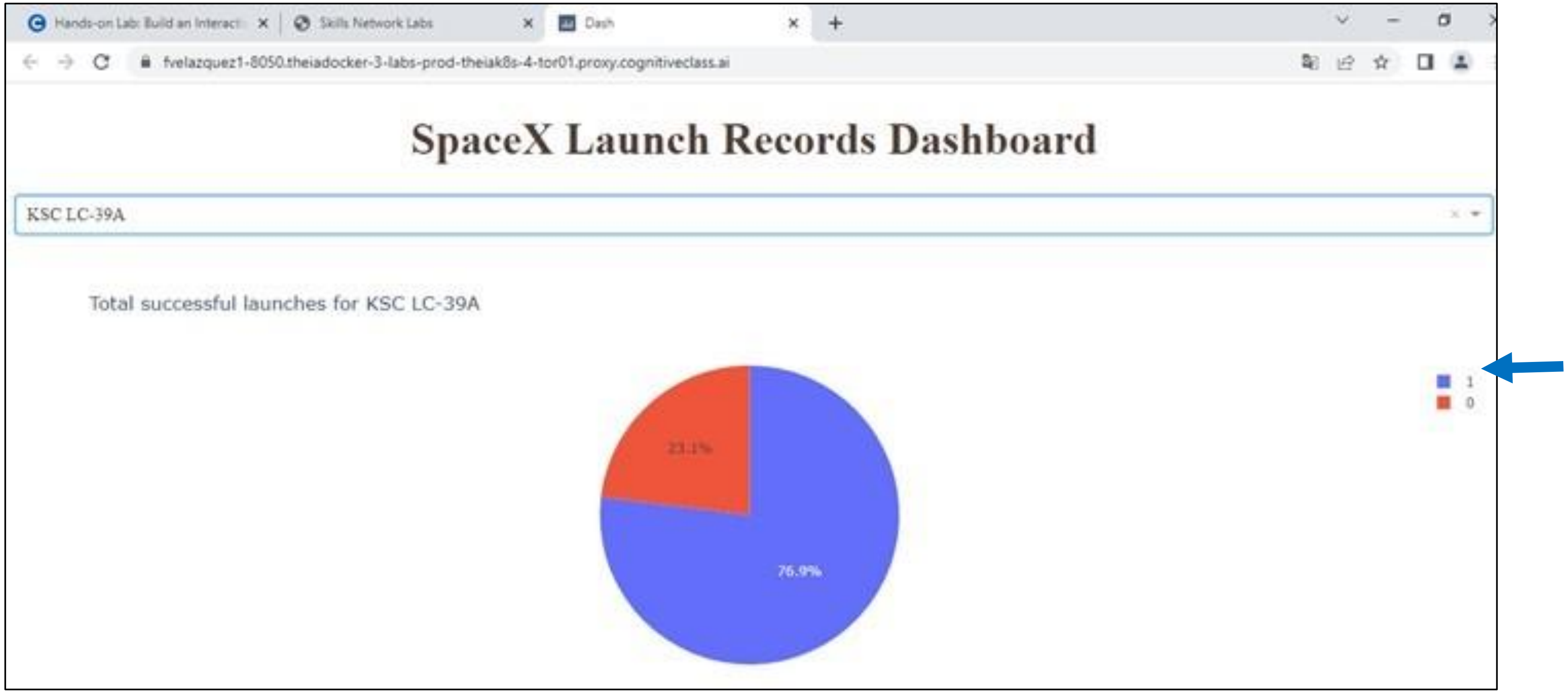
Interactive Dashboard

FT Booster version has the **highest number of successful** Falcon9 first stage landing outcomes.



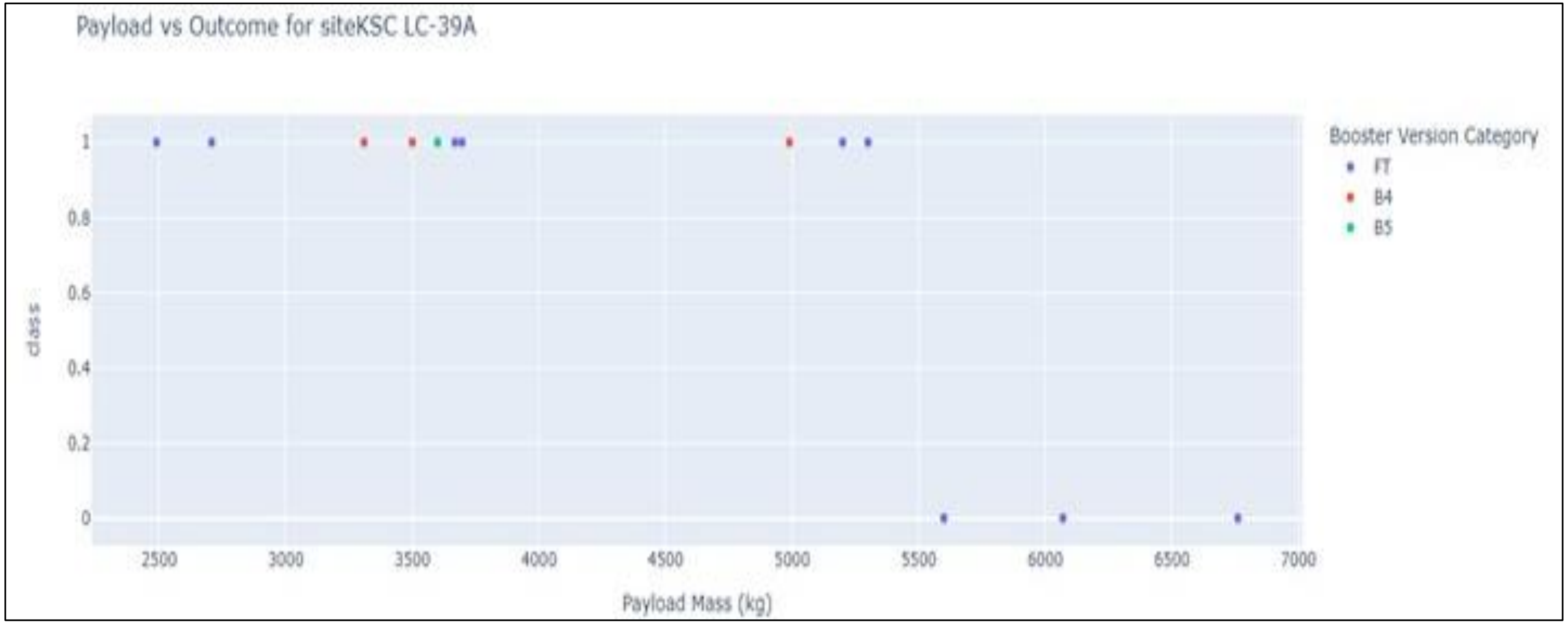
Interactive Dashboard

KSC LC-39A launch site: **76.9% rate of successful launches** (Falcon9 first stage successful landing).



Interactive Dashboard

KSC LC-39A launch site: relationship between **Payload Mass** and **Booster version category** with **launch success/failure** (Falcon9 first stage landing outcome).



Successful launch rate per launch site

*** KSC LS-39A has the highest successful launch rate**

Launch site	Number of launches	Successful launches	Successful launch rate
KSC LC-39A	**26**	**20**	**76.9%**
VAFB SLC-4E	20	8	40.0%
CCAFS LC-40	52	14	26.9%
CCAFS SLC-40	14	6	42.9%

Predictive Analysis:
Machine Learning model´s predictive accuracy

Machine Learning model	TP	TN	FP	FN	Total	Accuracy
Logistic Regression	3	12	3	0	18	0.833333333
Support Vector Machine	**5**	**12**	**1**	**0**	18	**0.944444444**
Decision Tree	2	11	4	1	18	0.722222222
K Nearest Neighbors	**5**	**12**	**1**	**0**	18	**0.944444444**

TP: True positives
FP: False positives
TN: True negatives
FN: False negatives

DISCUSSION

“Predicting the outcome of SpaceX Falcon9 first stage landing”

Scenario: Our company SpaceY wants to compete with SpaceX in the race to make space travel affordable for every one.

Goal: To **accurately predict** SpaceX Falcon9 first stage landing outcome.

Strategy: Develope **Machine Learning models** that **accurately predict the outcome** (success vs failure) of the Falcon9 first stage landing.

CONCLUSION

“Predicting the outcome of SpaceX Falcon9 first stage landing”

- We successfully **built a machine learning pipeline to predict the outcome of Falcon 9 first stage landing.**

- **The Machine Learning models that performs best are:**

  Support Vector Machine and K Nearest Neighbor