

# Study on Likelihood of Successful Space X Mission

Ashley Beard August 18, 2023

#### OUTLINE



- Executive Summary
- Introduction
- Methodology
  - Data Collection & Wrangling
  - EDA & Interactive Visual Analytics
  - Predictive Analysis & ML Modelling
- Results
  - EDA with SQL
  - EDA with Visualization
  - Interactive Map with Folium
  - Dashboard
  - Predictive Analysis & ML Modelling
- Discussion
- Conclusion
- References & Acknowledgement
- Appendix
  - Raw Data samples





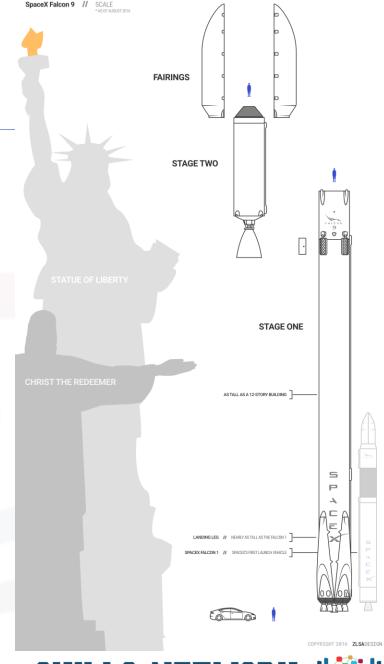
#### **EXECUTIVE SUMMARY**



- Out of many commercial space technology companies, Space X can save a lot of money for their customers by landing and reusing their expensive Stage 1 Booster rockets.
- Space Y wants to estimate the cost of a launch based on the outcome of the booster. This was done with the following methods:
  - Collection of free access Space X launch data
  - Exploratory Data Analysis to identify the relationships between various launch and payload parameters
  - Predictive model training to predict when a booster will land successfully
- There was shown to be a relationship between launch site, flight number, payload mass, and orbit type of each launch.
- Four different predictive model types were trained and compared to reveal the Decision Tree Classifier was the best candidate to predict the landing status to a precision of 86%.

#### INTRODUCTION

- SpaceX Falson 9 missions cost ~ \$62 million
  - Due to landing & reuse of expensive Stage 1 booster
- Other commercial space companies cost ~ \$165 million
  - Blue Origin, Virgin Galactic, Space Y, etc.
- Question: How do we determine the price of each launch?
  - Hypothesis: **Space Y** launch prices will vary depending on factors such as payload mass, location, and target orbit.
- Question: Can we predict when will SpaceX reuse the first stage?
  - a.k.a When will Space X successfully land their Stage 1 boosters?
  - Hypothesis: We can design a machine learning algorithm to do this







# Methodology



### Data Collection & Wrangling



Food Name	Categorical #	Calories
Apple	1	95
Chicken	2	231
Broccoli	3	50

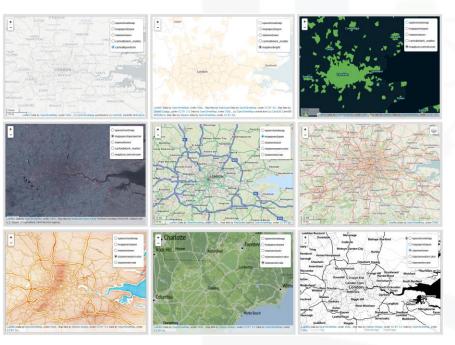


Apple	Chicken	Broccoli	Calories
1	0	0	95
0	1	0	231
0	0	1	50

- 1. Collect already compiled json data from the API: <a href="https://api.spacexdata.com/v4/launches/past">https://api.spacexdata.com/v4/launches/past</a> using HTTP protocol GET request.
  - Keep only Falcon 9 launches (where they reused boosters).
  - Replace missing values with means.
- 2. Collect Falcon 9 launch records from Wikipedia HTML tables. Parsing ("scraping") web data.
- 3. One-hot encoding for the mission outcome to categorize successful or failed booster landing. Calculate unique result counts in independent variables.
- \* All data used for this project is free access data



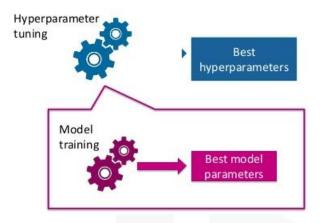
### EDA & Interactive Visual Analytics



- 1. Exploratory Data Analysis (showing deep traits/comparisons out of the data) using SQL queries on SpaceX.sql.
- Exploratory Data Analysis and feature engineering using data visualization.
  - 1. Plots: (Payload vs Flight #), (Launch site vs Flight #), (Launch Site vs Payload), (Orbit vs Outcome), (Flight # vs Orbit), (Orbit vs Payload) (Launch Success over time)
  - 2. One hot encoding for categories such as orbit, launch site, landing pad, and booster version
- Visualize launch site success rate (red vs green) with Folium plots. Group launches by launch site on map. Show proximity to railroads, highways, coastline, and cities.

### Predictive Analysis & ML Modelling

Hyperparameter tuning vs. model training



- One-hot encoding was useful in allowing the ML model to use the categorical data as input
- Perform exploratory Data Analysis and determine Training Labels
  - Create a column for the class
  - Standardize the data
  - Split into training data and test data
- Find best Hyperparameter for SVM, Classification Trees and Logistic Regression
- Find the method that performs best using test data
- We only have 18 test samples for this analysis.

## Results



### EDA with SQL

#### Findings:

- Average overall success rate: 66.66%
- 2010-2017:
  - Majority result: no attempt to land
  - Equal rate success/failure for drone ship target
- First successful ground landing: Dec 2015
- 12 launches w/ max payload (B5 booster)

#### **Implications:**

- Landing target does not affect landing result
- B5 booster was used for larger payloads, and v1.1 was used for smaller payloads
- Larger payloads later in timeline

### SQL Results

1. SELECT DISTINCT LAUNCH\_SITE FROM SPACEX\_DATA;

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

2. SELECT \* FROM SPACEX DATA WHERE LAUNCH SITE LIKE 'CCA%' LIMIT 5;

DATE	Time (UTC)	BOOSTER_VERSION	LAUNCH_SITE	PAYLOAD	PAYLOAD_MASSKG_	ORBIT	CUSTOMER	MISSION_OUTCOME	LANDING_OUTCOME
2010-04-06	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-08-12	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
2012-08-10	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-01-03	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

3. SELECT SUM(PAYLOAD\_MASS\_\_KG\_) FROM SPACEX\_DATA WHERE CUSTOMER = 'NASA (CRS)';

45596

4. SELECT AVG(PAYLOAD\_MASS\_\_KG\_) FROM SPACEX\_DATA WHERE BOOSTER\_VERSION = 'F9 v1.1';

1

2928

### SQL Results (cont.)

5. SELECT MIN(DATE) FROM SPACEX DATA WHERE LANDING OUTCOME = 'Success (ground pad)';

6. SELECT BOOSTER VERSION FROM SPACEX DATA WHERE LANDING OUTCOME = 'Success (drone ship)' AND PAYLOAD MASS KG BETWEEN 4000 AND 6000;

7. SELECT COUNT(MISSION OUTCOME) FROM SPACEX DATA GROUP BY MISSION OUTCOME;

8. SELECT DISTINCT BOOSTER VERSION FROM SPACEX DATA WHERE PAYLOAD MASS KG = (SELECT MAX(PAYLOAD MASS KG ) FROM SPACEX DATA);

1 2015-12-22

MISSION_OUTCOME	2
Failure (in flight)	1
Success	99
Success (payload status unclear)	1
	0

**BOOSTER VERSION** 

OOSTER_VERSION	
9 FT B1022	
9 FT B1026	
9 FT B1021.2	
9 FT B1031.2	

BOOSTER_VEROION
F9 B5 B1048.4
F9 B5 B1048.5
F9 B5 B1049.4
F9 B5 B1049.5
F9 B5 B1049.7
F9 B5 B1051.3
F9 B5 B1051.4
F9 B5 B1051.6
F9 B5 B1056.4
F9 B5 B1058.3
F9 B5 B1060.2
F9 B5 B1060.3

### SQL Results (cont.)

9. SELECT LANDING\_OUTCOME, BOOSTER\_VERSION, LAUNCH\_SITE, YEAR(DATE) FROM SPACEX\_DATA WHERE YEAR(DATE) = '2015' AND LANDING OUTCOME = 'Failure (drone ship)';

LANDING_OUTCOME	BOOSTER_VERSION	LAUNCH_SITE	4
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40	2015
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40	2015

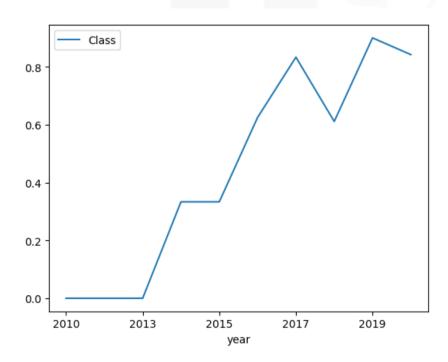
10. SELECT COUNT(LANDING\_OUTCOME) FROM SPACEX\_DATA WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY LANDING\_OUTCOME

ORDER BY COUNT(LANDING OUTCOME) DESC;

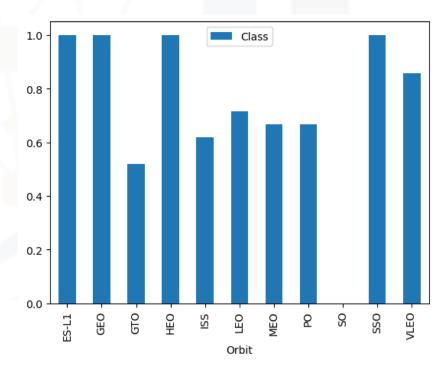
LANDING_OUTCOME	2
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Success (ground pad)	5
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	1
Precluded (drone ship)	1

### EDA with Visualization

 Success rate increased over time, as did max payload mass



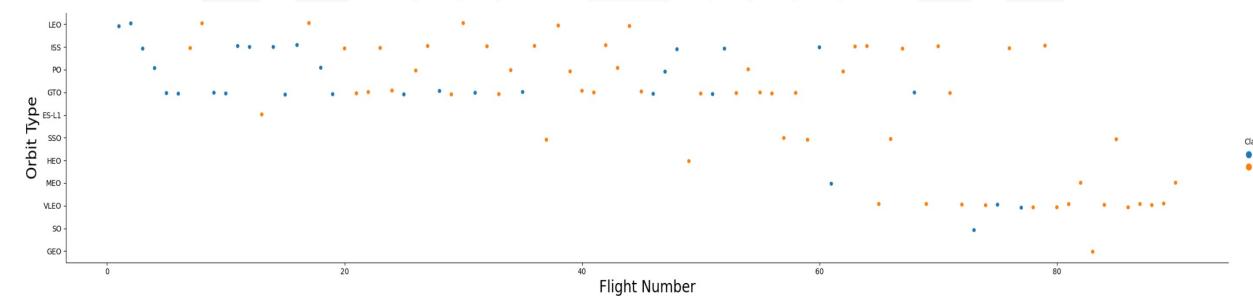
- 3 most target orbit types:
  - GTO, ISS, VLEO
  - Highest success rates (ES-L1, GEO, HEO, and SSO) have only 1-5 launches





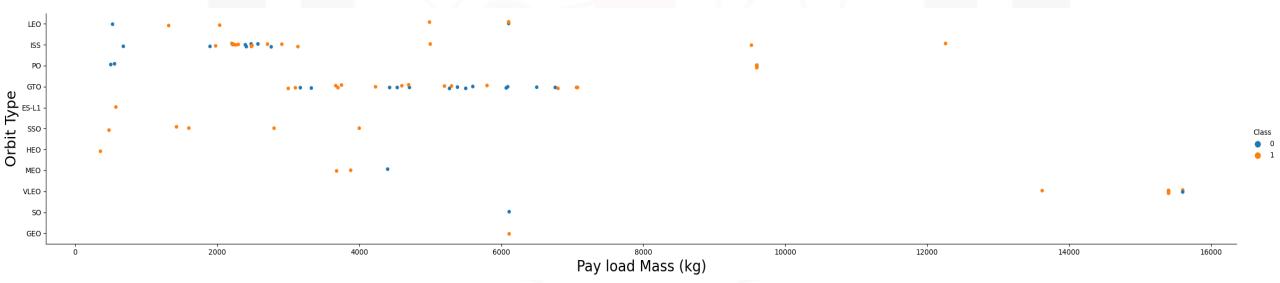
#### Orbit vs Flight Number

- there seems to be no relationship between orbit & flight number when in GTO orbit.
- Some orbits (SSO, VLEO, etc.) not attempted until later flight numbers --> But have higher success rate



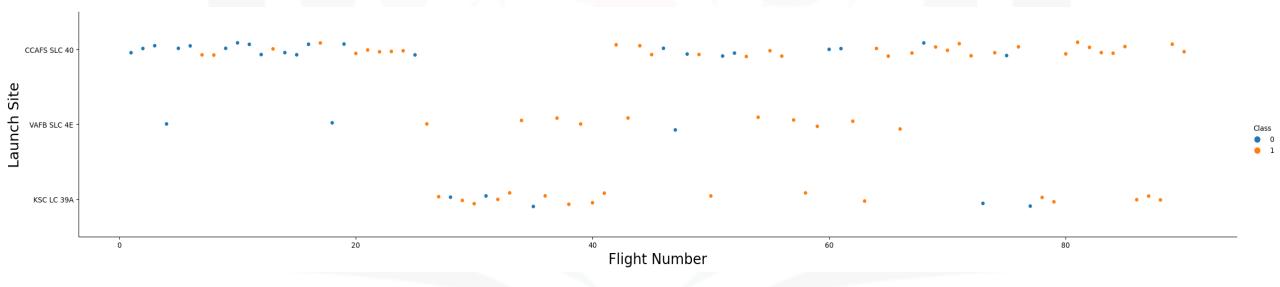
#### Orbit vs Payload Mass

- Orbits ES-L1, SSO, and HEO cannot accept Payloads more than 4000kg
- VLEO and ISS orbits have high success with "heavy" payloads



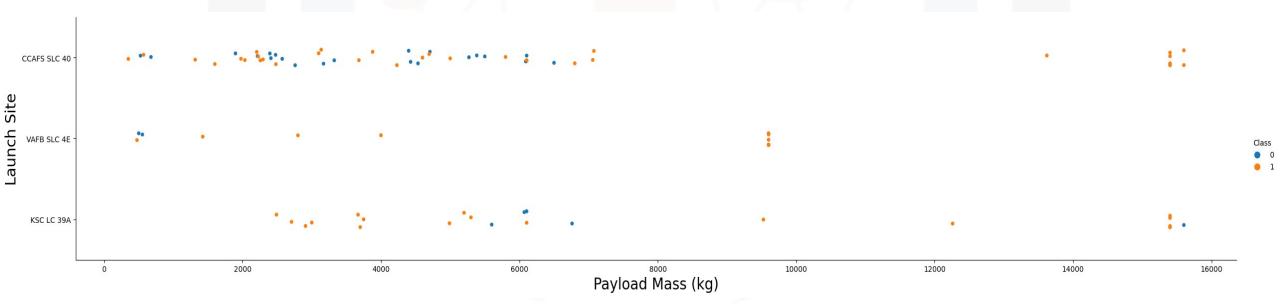
#### Launch Site vs Flight Number

- CCAFS SLC-40 has the most launches across the timespan
- VAFB SLC-4E has the least launches and no launces after ~2019



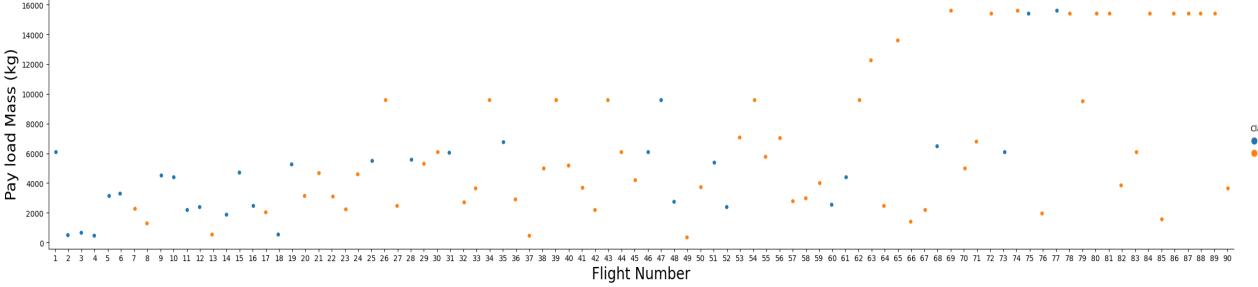
#### Launch Site vs Payload Mass

• VAFB-SLC launch site there are no rockets launched for heavy payload mass (greater than 10000 kg).



#### Payload Mass vs Flight Number

- We see that different launch sites have different success rates:
  - CCAFS LC-40, has a success rate of 60%
  - KSC LC-39A and VAFB SLC 4E has a success rate of 77%.

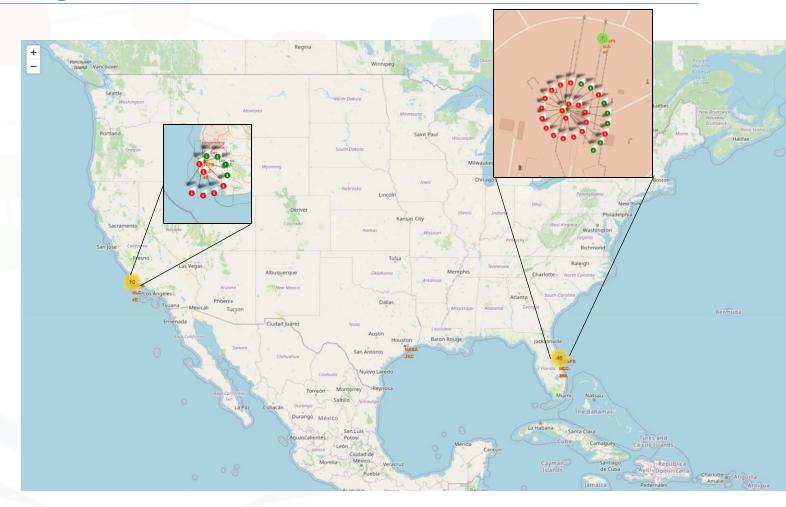




### Interactive Map with Folium

#### • Findings:

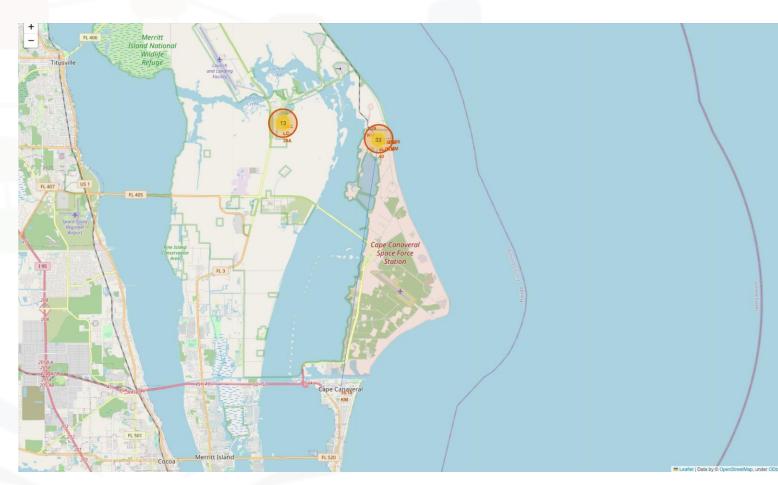
- Four launch sites (including NASA JSC)
  - VAFB SLC-4E
  - CCAFS SLC-40
  - KSC LC-39A
- Red markers added to indicate failed landings
- Green markers indicate success
- Marker groups are used to simplify map. They expand into individual markers automatically when the user zooms into the map.



### Interactive Map with Folium (cont.)

#### • Findings:

- Proximities near Kennedy Launch Complex were calculated as an example but it is true for all launch sites in this data set that they are all located in proximity (within 14 kilometers) to railways, highways, and coastline.
- Kennedy Space Center Launch Complex:
  - .59km to highway
  - .86km to coast
  - 1.29km to rail
  - 18.18km to city
- All launch sites with the exception of the JSC maintain between 14-23 kilometers from the nearest city.
- All launch sites are located between 3176km -3851 km of the Equator.



#### **DASHBOARD**



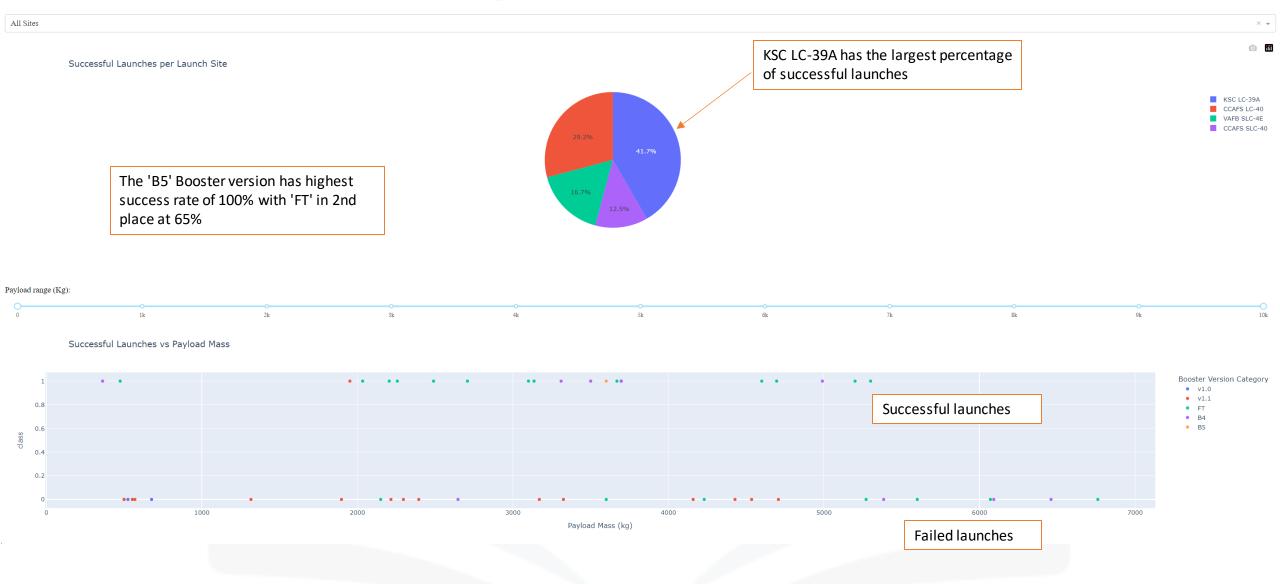
#### **Interactive Dashboard to visualize Space X launch data:**

#### To run:

- Have Cognos Skills Network Labs window open
- Navigate to <a href="https://github.com/dinosrule90/Data-">https://github.com/dinosrule90/Data-</a> Science-Capstone.git
- Download file "spacex\_dash\_app.py"
- Import above Python file to SNL and run according to instructions.

\* These are optional instructions for the curious individual to explore the script themselves.

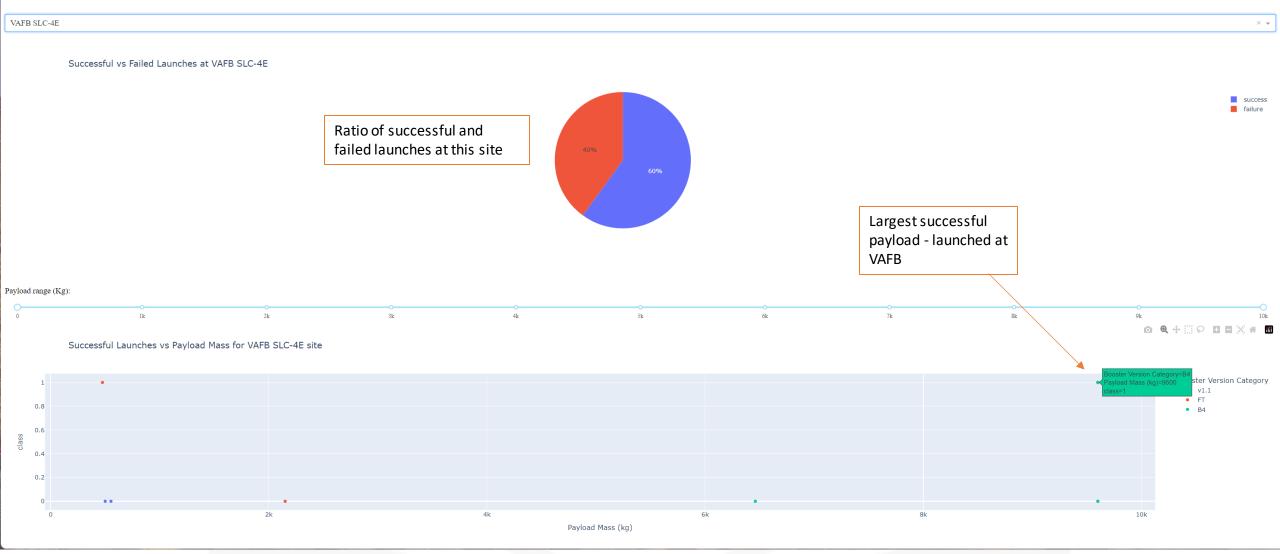
#### **SpaceX Launch Records Dashboard**







#### **SpaceX Launch Records Dashboard**







#### **SpaceX Launch Records Dashboard**







### Predictive Analysis & ML modelling

#### **Findings:**

- Logistic regression w/ parameters:
  - 'C': 0.01, 'penalty': '12', 'solver': 'lbfgs'
  - accuracy : 0.8464285714285713
- Support Vector Machine w/ parameters:
  - 'C': 1.0, 'gamma': 0.03162277660168379, 'kernel': 'sigmoid'
  - accuracy : 0.8482142857142856
- **Decision Tree classifier** w/ parameters:
  - 'criterion': 'entropy', 'max\_depth': 10,
     'max\_features': 'sqrt', 'min\_samples\_leaf': 4,
     'min\_samples\_split': 10, 'splitter': 'random'
  - accuracy : 0.8625
- K-Nearest Neighbors w/ parameters:
  - 'algorithm': 'auto', 'n\_neighbors': 10, 'p': 1
  - accuracy: 0.8482142857142858

#### Implications:

- Decision Tree classifier was the most accurate at predicting whether booster will land successfully
- Sample size is too small --> causes over-fitting and not enough generalizing for new scenarios
- Accuracy could be improved with more launch data

#### **DISCUSSION**



- As *Space Y*, we are able to predict the likelihood of a mission reusing its Stage 1 Booster to 86% precision.
- Potential marketing approach with this study:
  - If your target orbit is a Sun-synchronous Orbit (SO) SpaceX will likely not save the Stage 1
    - go with Space Y
  - If you plan to launch from CCAFS LC-40
    - Risk of failed landing is too high
    - Go with Space Y

### **CONCLUSION**



- We showed certain launch parameters can affect a launch price by using:
  - 1. Data collection
  - Data wrangling
  - Exploratory Data Analysis (EDA)
  - 4. Interactive visualization
- Space Y can extrapolate from the machine learning model a risk factor involved in the price of each Space X mission and create a more accurate price estimate to bid against them.
- **Future Work:** 
  - Improve the accuracy of machine learning model with more Space X and other commercial space companies' data.
  - Check the data for a relationship between weather and landing class.

### References & Acknowledgement

- https://gps-coordinates.org/distance-between-coordinates.php
- https://zlsadesign.com/infographic/
- https://api.spacexdata.com/v4/launches/past
- <a href="https://en.wikipedia.org/wiki/List of Falcon 9 and Falcon Heavy launches">https://en.wikipedia.org/wiki/List of Falcon 9 and Falcon Heavy launches</a>
- I would like to acknowledge the authors of the Applied Data Science Capstone course, Yan Luo and Joseph Santarcangelo for the helpful lessons and Python code frameworks to complete this project.

# Appendix



#### Raw Data from API

#### Sample of Space X data from API:

> See Github for full data file

```
FlightNumber, Date, BoosterVersion, PayloadMass, Orbit, LaunchSite, Outcome, Flights, GridFins, Reused, Legs, LandingPad, Block, ReusedCount, Serial, Longitude, Latitude
1,2010-06-04, Falcon 9,6104.959411764706, LEO, CCAFS SLC 40, None None, 1, False, False, False, 1.0,0, B0003, -80.577366, 28.5618571
2,2012-05-22, Falcon 9,525.0, LEO, CCAFS SLC 40, None None, 1, False, False, False, 1.0,0, B0005, -80.577366, 28.5618571
3,2013-03-01, Falcon 9,677.0, ISS, CCAFS SLC 40, None None, 1, False, False, False, 1.0,0, B0007, -80.577366, 28.5618571
4,2013-09-29,Falcon 9,500.0,PO,VAFB SLC 4E,False Ocean,1,False,False,False,1.0,0,B1003,-120.610829,34.632093
5,2013-12-03, Falcon 9,3170.0,GTO,CCAFS SLC 40, None None, 1, False, False, False, 1.0,0, B1004, -80.577366, 28.5618571
6,2014-01-06,Falcon 9,3325.0,GTO,CCAFS SLC 40,None None,1,False,False,False,,1.0,0,B1005,-80.577366,28.5618571
7,2014-04-18,Falcon 9,2296.0,ISS,CCAFS SLC 40,True Ocean,1,False,False,True,,1.0,0,B1006,-80.577366,28.5618571
8,2014-07-14,Falcon 9,1316.0,LEO,CCAFS SLC 40,True Ocean,1,False,False,True,,1.0,0,B1007,-80.577366,28.5618571
9,2014-08-05, Falcon 9,4535.0, GTO, CCAFS SLC 40, None None, 1, False, False, False, 1.0,0, B1008, -80.577366, 28.5618571
10,2014-09-07, Falcon 9,4428.0,GTO,CCAFS SLC 40,None None,1,False,False,False,1.0,0,B1011,-80.577366,28.5618571
11,2014-09-21, Falcon 9,2216.0, ISS, CCAFS SLC 40, False Ocean, 1, False, False, False, 1.0,0, B1010, -80.577366, 28.5618571
12,2015-01-10,Falcon 9,2395.0,ISS,CCAFS SLC 40,False ASDS,1,True,False,True,5e9e3032383ecb761634e7cb,1.0,0,B1012,-80.577366,28.5618571
13,2015-02-11, Falcon 9,570.0, ES-L1, CCAFS SLC 40, True Ocean, 1, True, False, True, , 1.0,0, B1013, -80.577366, 28.5618571
14,2015-04-14,Falcon 9,1898.0,ISS,CCAFS SLC 40,False ASDS,1,True,False,True,5e9e3032383ecb761634e7cb,1.0,0,B1015,-80.577366,28.5618571
15,2015-04-27, Falcon 9,4707.0,GTO,CCAFS SLC 40,None None,1,False,False,False,1.0,0,B1016,-80.577366,28.5618571
16,2015-06-28, Falcon 9,2477.0, ISS, CCAFS SLC 40, None ASDS, 1, True, False, True, 5e9e3032383ecb6bb234e7ca, 1.0,0, B1018, -80.577366, 28.5618571
17,2015-12-22,Falcon 9,2034.0,LEO,CCAFS SLC 40,True RTLS,1,True,False,True,5e9e3032383ecb267a34e7c7,1.0,0,B1019,-80.577366,28.5618571
18,2016-01-17, Falcon 9,553.0, PO, VAFB SLC 4E, False ASDS, 1, True, False, True, 5e9e3033383ecbb9e534e7cc, 1.0,0, B1017, -120.610829, 34.632093
19,2016-03-04, Falcon 9,5271.0, GTO, CCAFS SLC 40, False ASDS, 1, True, False, True, 5e9e3032383ecb6bb234e7ca, 1.0,0, B1020, -80.577366, 28.5618571
20,2016-04-08, Falcon 9,3136.0, ISS, CCAFS SLC 40, True ASDS, 1, True, False, True, 5e9e3032383ecb6bb234e7ca, 2.0, 1, B1021, -80.577366, 28.5618571
21,2016-05-06, Falcon 9,4696.0, GTO, CCAFS SLC 40, True ASDS, 1, True, False, True, 5e9e3032383ecb6bb234e7ca, 2.0,0, B1022, -80.577366, 28.5618571
22,2016-05-27,Falcon 9,3100.0,GTO,CCAFS SLC 40,True ASDS,1,True,False,True,5e9e3032383ecb6bb234e7ca,2.0,1,B1023,-80.577366,28.5618571
23,2016-07-18, Falcon 9,2257.0, ISS, CCAFS SLC 40, True RTLS, 1, True, False, True, 5e9e3032383ecb267a34e7c7, 2.0, 1, B1025, -80.577366, 28.5618571
24,2016-08-14, Falcon 9,4600.0, GTO, CCAFS SLC 40, True ASDS, 1, True, False, True, 5e9e3032383ecb6bb234e7ca, 2.0,0, B1026, -80.577366, 28.5618571
25,2016-09-01,Falcon 9,5500.0,GTO,CCAFS SLC 40,None ASDS,1,True,False,True,5e9e3032383ecb6bb234e7ca,3.0,0,B1028,-80.577366,28.5618571
```

### Raw Data from Web Scraping

#### Sample of Space X data from Web Scraping:

#### > See Github for full data file

```
Date, Time (UTC), Booster Version, Launch Site, Payload, PAYLOAD MASS KG, Orbit, Customer, Mission Outcome, Landing Outcome
2010-04-06,18:45:00,F9 v1.0 B0003,CCAFS LC-40,Dragon Spacecraft Qualification Unit,0,LEO,SpaceX,Success,Failure (parachute)
2010-08-12,15:43:00,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
2012-08-10,00:35:00,F9 v1.0 B0006,CCAFS LC-40,SpaceX CRS-1,500,LEO (ISS),NASA (CRS),Success,No attempt
2013-01-03,15:10:00,F9 v1.0 B0007,CCAFS LC-40,SpaceX CRS-2,677,LEO (ISS),NASA (CRS),Success,No attempt
2013-09-29,16:00:00,F9 v1.1 B1003,VAFB SLC-4E,CASSIOPE,500,Polar LEO,MDA,Success,Uncontrolled (ocean)
2013-03-12,22:41:00,F9 v1.1,CCAFS LC-40,SES-8,3170,GT0,SES,Success,No attempt
2014-06-01,22:06:00,F9 v1.1,CCAFS LC-40,Thaicom 6,3325,GTO,Thaicom,Success,No attempt
2014-04-18,19:25:00,F9 v1.1,CCAFS LC-40,SpaceX CRS-3,2296,LEO (ISS),NASA (CRS),Success,Controlled (ocean)
2014-07-14,15:15:00,F9 v1.1,CCAFS LC-40,OG2 Mission 1 6 Orbcomm-OG2 satellites,1316,LEO,Orbcomm,Success,Controlled (ocean)
2014-05-08,08:00:00,F9 v1.1,CCAFS LC-40,AsiaSat 8,4535,GTO,AsiaSat,Success,No attempt
2014-07-09,05:00:00,F9 v1.1 B1011,CCAFS LC-40,AsiaSat 6,4428,GTO,AsiaSat,Success,No attempt
2014-09-21,05:52:00,F9 v1.1 B1010,CCAFS LC-40,SpaceX CRS-4,2216,LEO (ISS),NASA (CRS),Success,Uncontrolled (ocean)
2015-10-01,09:47:00,F9 v1.1 B1012,CCAFS LC-40,SpaceX CRS-5,2395,LEO (ISS),NASA (CRS),Success,Failure (drone ship)
2015-11-02,23:03:00,F9 v1.1 B1013,CCAFS LC-40,DSCOVR,570,HEO,U.S. Air Force NASA NOAA,Success,Controlled (ocean)
2015-02-03,03:50:00,F9 v1.1 B1014,CCAFS LC-40,ABS-3A Eutelsat 115 West B,4159,GTO,ABS Eutelsat,Success,No attempt
2015-04-14,20:10:00,F9 v1.1 B1015,CCAFS LC-40,SpaceX CRS-6,1898,LEO (ISS),NASA (CRS),Success,Failure (drone ship)
2015-04-27,23:03:00,F9 v1.1 B1016,CCAFS LC-40,Turkmen 52 / MonacoSAT,4707,GTO,Turkmenistan National Space Agency,Success,No attempt
2015-06-28,14:21:00,F9 v1.1 B1018,CCAFS LC-40,SpaceX CRS-7,1952,LEO (ISS),NASA (CRS),Failure (in flight),Precluded (drone ship)
2015-12-22,01:29:00,F9 FT B1019,CCAFS LC-40,0G2 Mission 2 11 Orbcomm-OG2 satellites,2034,LE0,Orbcomm,Success,Success (ground pad)
2016-01-17,18:42:00,F9 v1.1 B1017,VAFB SLC-4E, Jason-3,553, LEO, NASA (LSP) NOAA CNES, Success, Failure (drone ship)
2016-04-03,23:35:00,F9 FT B1020,CCAFS LC-40,SES-9,5271,GT0,SES,Success,Failure (drone ship)
2016-08-04,20:43:00,F9 FT B1021.1,CCAFS LC-40,SpaceX CRS-8,3136,LEO (ISS),NASA (CRS),Success,Success (drone ship)
2016-06-05,05:21:00,F9 FT B1022,CCAFS LC-40,JCSAT-14,4696,GTO,SKY Perfect JSAT Group,Success,Success (drone ship)
2016-05-27,21:39:00,F9 FT B1023.1,CCAFS LC-40,Thaicom 8,3100,GTO,Thaicom,Success,Success (drone ship)
2016-06-15,14:29:00,F9 FT B1024,CCAFS LC-40,ABS-2A Eutelsat 117 West B,3600,GTO,ABS Eutelsat,Success,Failure (drone ship)
```