IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

By Oleh Lykho
27.04.2024

Outline

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

Introduction

In this capstone, we will predict if the Falcon 9 first stage will land successfully. SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage. Therefore if we can determine if the first stage will land, we can determine the cost of a launch. This information can be used if an alternate company wants to bid against SpaceX for a rocket launch. In this lab, you will collect and make sure the data is in the correct format from an API.

Section
1

Methodology

Methodology

Executive Summary

Data collection methodology:

- Data collection involves gathering information from specified sources to address a given problem or achieve certain objectives. This could be done via surveys, web scraping, databases, APIs, or other means. Careful planning is needed to ensure data integrity and relevance.

Perform data wrangling

- This stage focuses on converting raw data into a format suitable for analysis. It includes data normalization, transformation, and feature engineering. Effective processing ensures that subsequent analyses are efficient and reliable.

Exploratory Data Analysis (EDA) Using Visualization and SQL:

- EDA involves analyzing datasets to summarize their main characteristics, often using statistical graphics and other data visualization methods. SQL can be used to query and aggregate data, which is then visualized to uncover patterns, spot anomalies, and test hypotheses.

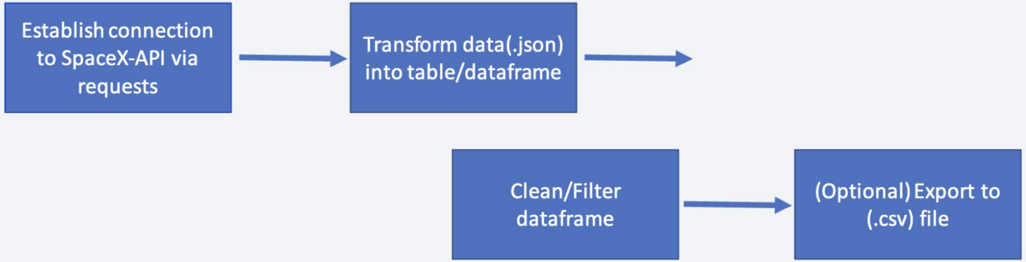
Building, Tuning, and Evaluating Classification Models:

Building: Models are constructed based on the problem and data type.
Tuning: Model parameters are optimized (e.g., using grid search) to improve performance.
Evaluating: Models are evaluated using metrics like accuracy, precision, recall, and F1-score to determine their effectiveness.

Data Collection

- How data sets were collected?
    - Gather and measure information on targeted variables
- Data collection process:

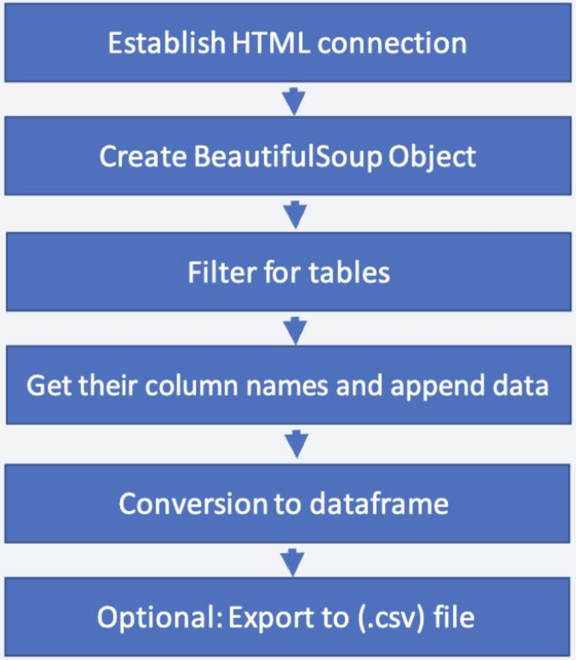
Data Collection – SpaceX API



https://github.com/OlehLy/Applied-Data-Science-Capstone-Final-Exam/blob/main/Complete%20the%20EDA%20with%20SQL.ipynb

Data Collection - Scraping

https://github.com/OlehLy/Applied-Data-Science-Capstone-Final-Exam/blob/main/Complete%20the%20Data%20Collection%20with%20Web%20Scraping%20lab.ipynb

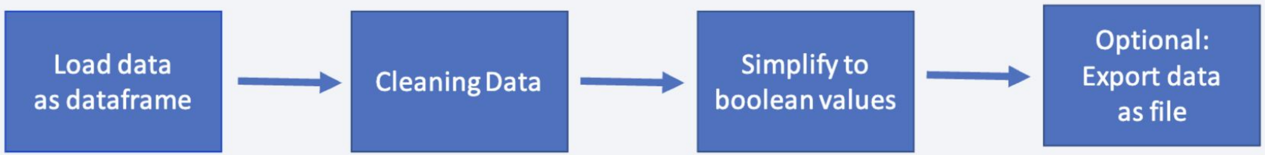


Data Wrangling

How data were processed?

- Cleaning and polishing possible messy/complex data sets for better handling

https://github.com/OlehLy/Applied-Data-Science-Capstone-Final-Exam/blob/main/Complete%20the%20Machine%20Learning%20Prediction%20lab.ipynb



EDA with Data Visualization

What needs to be done?

• Develop visuals and compile optical insights.

• Here's how:

- Generate scatterplots correlating Payload Mass/Flight Number with Launch Site/Orbit Type/Flight Number to illustrate dependencies.
- Construct bar charts comparing Success Rate with Orbit Type to evaluate its impact.
- Plot Launch Success against Year to observe trends over time.

https://github.com/OlehLy/Applied-Data-Science-Capstone-Final-Exam/blob/main/Complete%20the%20EDA%20with%20SQL.ipynb

EDA with SQL

What needs to be done?

• Employ heuristic methods to explore the database and make educated guesses about potential occurrences.

• Our inquiries are as follows:

• Provide a general overview of available landslides, particularly the first five entries starting with 'CCA'.

• Determine the number of successful and failed mission outcomes, listing the failed ones specifically in 2015.

• Identify the initial successful landing outcome on a drone ship.

• Specify, count, and rank outcomes between approximately 2010 and 2017.

• Calculate the average payload per booster for the 'F9v1.1' version and the total for boosters utilized by NASA (CRS).

• Determine the booster version with the maximum payload capacity.

• List the names of boosters that achieved successful ground pad landings with specific payloads.

https://github.com/OlehLy/Applied-Data-Science-Capstone-Final-Exam/blob/main/Complete%20the%20EDA%20with%20SQL.ipynb

Build an Interactive Map with Folium

What has been accomplished?

- Implemented Map(Simple) with Icons, Circles markers, and Lines.

Why? To improve and consolidate visual representations of success and failure at each landing site.

https://github.com/OlehLy/Applied-Data-Science-Capstone-Final-Exam/blob/main/Interactive%20Visual%20Analytics%20with%20Folium%20lab.ipynb

Build a Dashboard with Plotly Dash

What has been accomplished?

• Chosen Launch Site.

• Created Pie charts illustrating the percentage relation based on launch site.

• Developed Scatter graphs depicting the correlation between payload and success according to launch site.

• Why? To visually represent the efficacy of launch sites and payload mass.

https://github.com/OlehLy/Applied-Data-Science-Capstone-Final-Exam/blob/main/Interactive%20Visual%20Analytics%20with%20Folium%20lab.ipynb

Predictive Analysis (Classification)

Standardizing/Transforming the data.

Dividing into test and training sets.

Initiating various Machine Learning algorithms using the training set.

Evaluate accuracy through:
- R-score
- Confusion matrix (distinguishing true/false positives and negatives for landings)

=> Seek the optimal:
- Score

-> Parameters for Machine Learning.

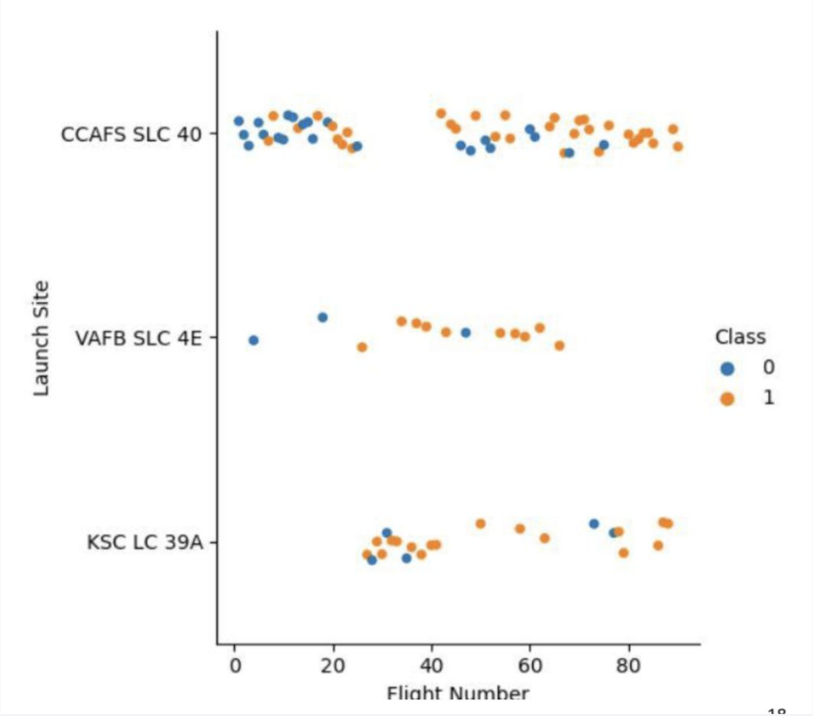
Based on accuracy:
Select the most suitable model.

Section
2

Insights drawn from EDA

Flight Number vs. Launch Site

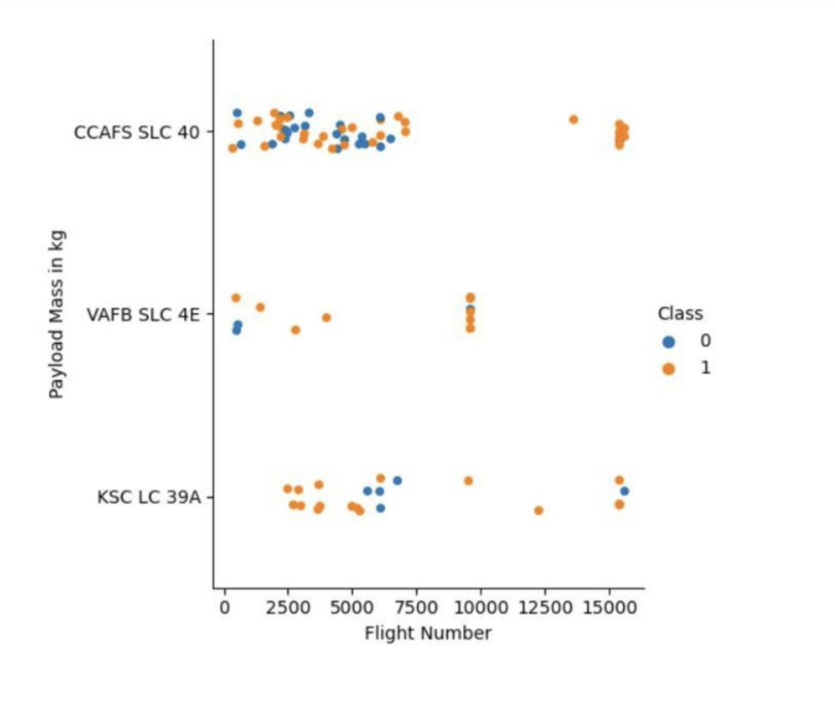
- First 25 Starts, in particular. at CCAFS SLC 40 were mostly failures
- Starts at KSC LC 39A gave them then information for success
- VAFB SLC 4E with most successful launches



Payload vs. Launch Site

Following the successful launch of a payload weighing approximately 7500kg,

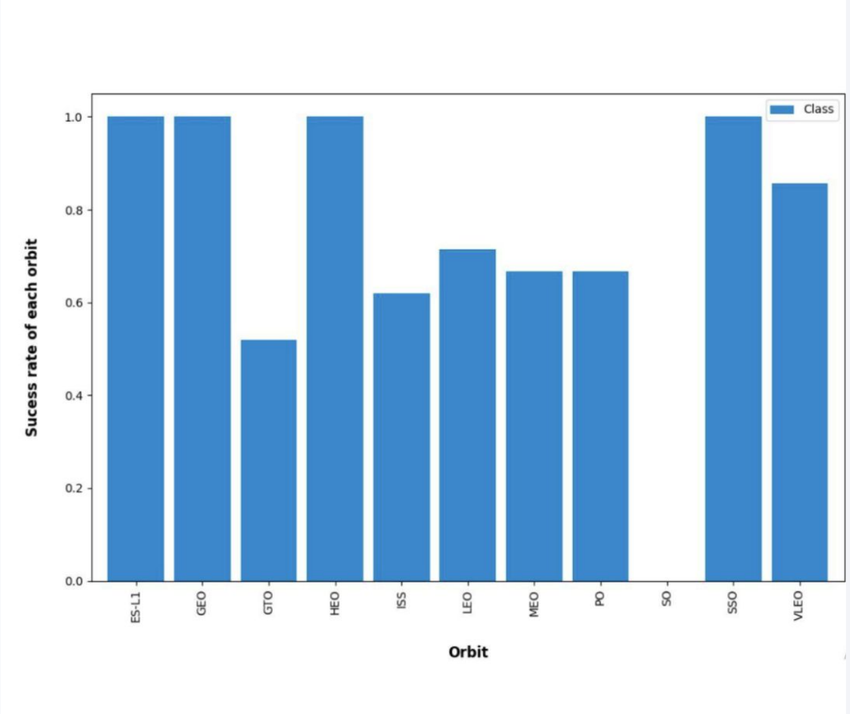
Launches carrying heavier payloads consistently achieved successful outcomes.



Success Rate vs. Orbit Type

Launches to Sun-synchronous orbit (SSO) ended disastrously, with only one attempted launch.

However, launches to Earth-Sun Lagrange point 1 (ES-L1), Geostationary Earth Orbit (GEO), High Earth Orbit (HEO), and SSO were successful.

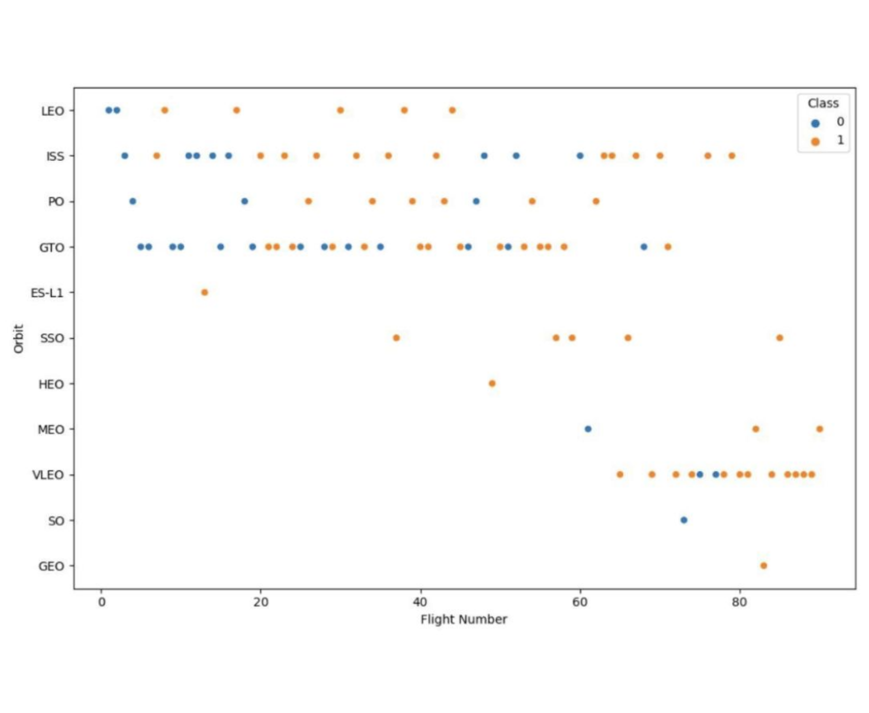


Flight Number vs. Orbit Type

Initially, testing was conducted on four selected orbits.

Significant breakthroughs occurred after approximately the 20th flight, particularly with the International Space Station (ISS) and Geostationary Transfer Orbit (GTO).

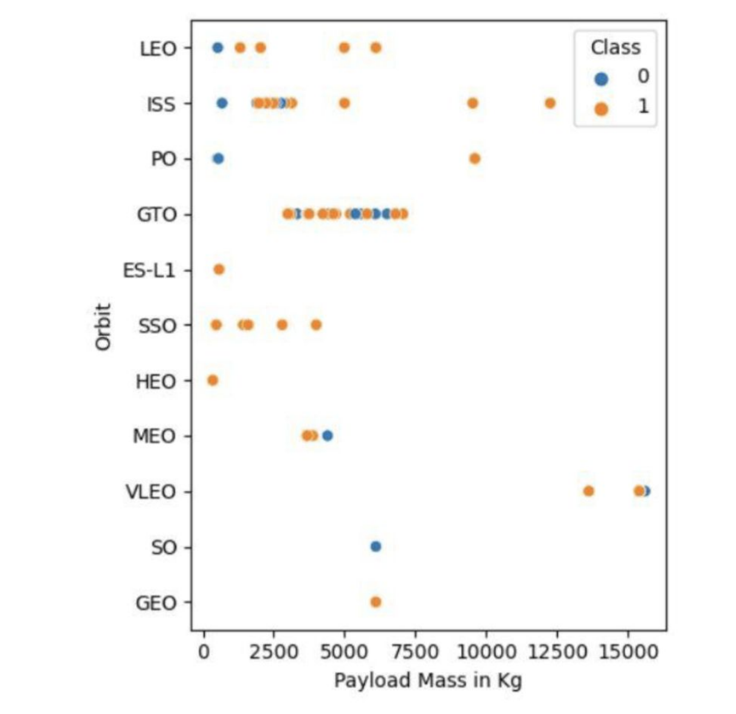
Subsequently, after approximately the 60th flight, attention was also directed towards other orbits, notably Very Low Earth Orbit (VLEO).



Payload vs. Orbit Type

Started mostly with 2500-7500kg

Launches with higher payload > 7500kg rather successful

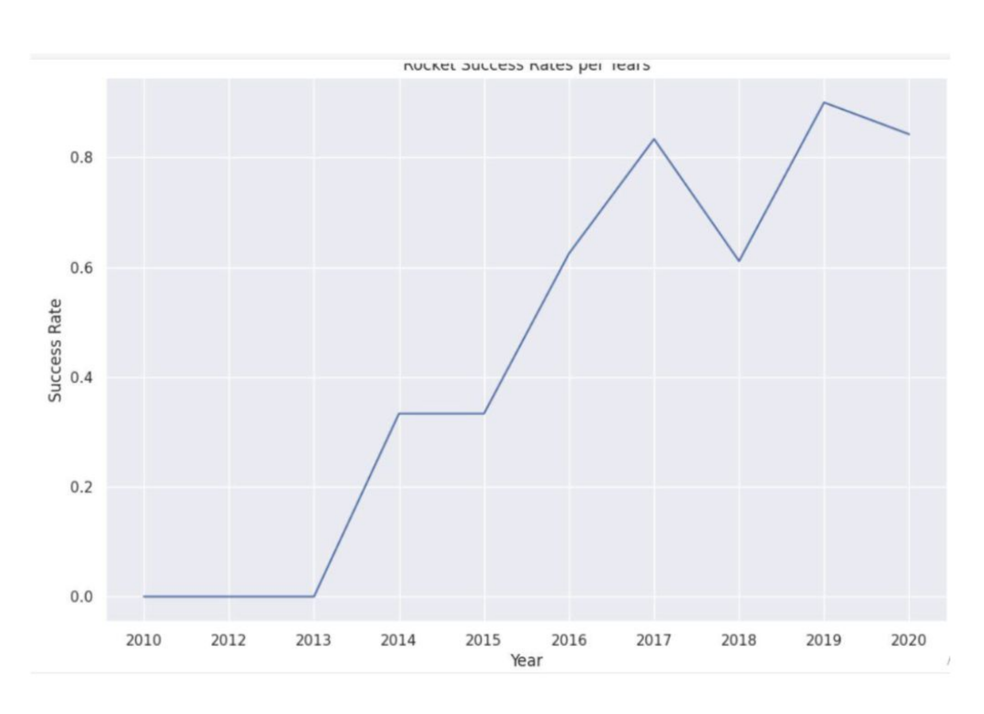


Launch Success Yearly Trend

Since 2013 mostly improving

Since 2017 mostly >80% except for

Dip in Success in 2018



All Launch Site Names

- Find the names of the unique launch sites
- Present your query result with a short explanation here

```
%sql SELECT DISTINCT launch_site FROM SPACEX
```

```
 * ibm_db_sa://vzv80836:***@764264db-9824-4b7c
   sqlite:///my_data1.db
Done.
```

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

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`
- Present your query result with a short explanation here

```
%sql SELECT * FROM SPACEX WHERE launch_site LIKE 'CCA%' LIMIT 5

 * ibm_db_sa://vzv80836:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32536/BLUDB
   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	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

26

Total Payload Mass

- Calculate the total payload carried by boosters from NASA
- Present your query result with a short explanation here

```
%sql SELECT SUM(payload_mass__kg_) as "Total payload mass" FROM SPACEX WHERE customer LIKE 'NASA (CRS)'
```

```
 * ibm_db_sa://vzv80836:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32536/BLUDB
   sqlite:///my_data1.db
Done.
```

Total payload mass
45596

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1
- Present your query result with a short explanation here

```
%sql SELECT AVG(payload_mass__kg_) as "Total payload mass" FROM SPACEX WHERE booster_version LIKE 'F9 v1.1'
```

```
 * ibm_db_sa://vzv80836:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32536/BLUDB
   sqlite:///my_data1.db
Done.
```

Total payload mass
2928

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad
- Present your query result with a short explanation here

```
%sql SELECT * FROM SPACEX WHERE landing__outcome LIKE 'Success%' ORDER BY DATE LIMIT 1
```

```
 * ibm_db_sa://vzv80836:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32536/BLUDB
   sqlite:///my_data1.db
Done.
```

DATE	time__utc_	booster_version	launch_site	payload	payload_mass__kg_	orbit	customer	mission_outcome	landing__outcome
2015-12-22	01:29:00	F9 FT B1019	CCAFS LC-40	OG2 Mission 2 11 Orbcomm-OG2 satellites	2034	LEO	Orbcomm	Success	Success (ground pad)

Successful Drone Ship Landing with Payload between 4000 and 6000

- List the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

- Present your query result with a short explanation here

```
%sql SELECT * FROM SPACEX WHERE landing__outcome LIKE 'Success (drone ship)' AND payload_mass__kg_ > 4000 AND payload_mass__kg_ < 6000
```

```
 * ibm_db_sa://vzv80836:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32536/BLUDB
   sqlite:///my_data1.db
Done.
```

DATE	time__utc_	booster_version	launch_site	payload	payload_mass__kg_	orbit	customer	mission_outcome	landing__outcome
2016-05-06	05:21:00	F9 FT B1022	CCAFS LC-40	JCSAT-14	4696	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
2016-08-14	05:26:00	F9 FT B1026	CCAFS LC-40	JCSAT-16	4600	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
2017-03-30	22:27:00	F9 FT B1021.2	KSC LC-39A	SES-10	5300	GTO	SES	Success	Success (drone ship)
2017-10-11	22:53:00	F9 FT B1031.2	KSC LC-39A	SES-11 / EchoStar 105	5200	GTO	SES EchoStar	Success	Success (drone ship)

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes
- Present your query result with a short explanation here

```
%sql SELECT mission_outcome, COUNT(mission_outcome) as count FROM SPACEX GROUP BY mission_outcome
```

```
 * ibm_db_sa://vzv80836:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32536/BLUDB
   sqlite:///my_data1.db
Done.
```

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

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass
- Present your query result with a short explanation here

```
%sql SELECT DISTINCT booster_version FROM SPACEX WHERE payload_mass__kg_ = (SELECT MAX(payload_mass__kg_) FROM SPACEX)
```

```
 * ibm_db_sa://vzv80836:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32536/BLUDB
   sqlite:///my_data1.db
Done.
```

booster_version
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

2015 Launch Records

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

- Present your query result with a short explanation here

```
%%sql SELECT substr(Date,6,2) as Month, landing__outcome, booster_version, launch_site FROM SPACEX WHERE substr(Date,1,4) = '2015'
AND landing__outcome LIKE 'Failure (drone ship)'
```

```
 * ibm_db_sa://vzv80836:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32536/BLUDB
Done.
```

MONTH	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

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

- Rank 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
- Present your query result with a short explanation here

```
%%sql SELECT landing__outcome, count(landing__outcome) as COUNT FROM SPACEX WHERE date BETWEEN '2010-06-04' AND '2017-03-20'
GROUP BY landing__outcome ORDER BY COUNT(landing__outcome) DESC
```

```
 * ibm_db_sa://vzv80836:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32536/BLUDB
Done.
```

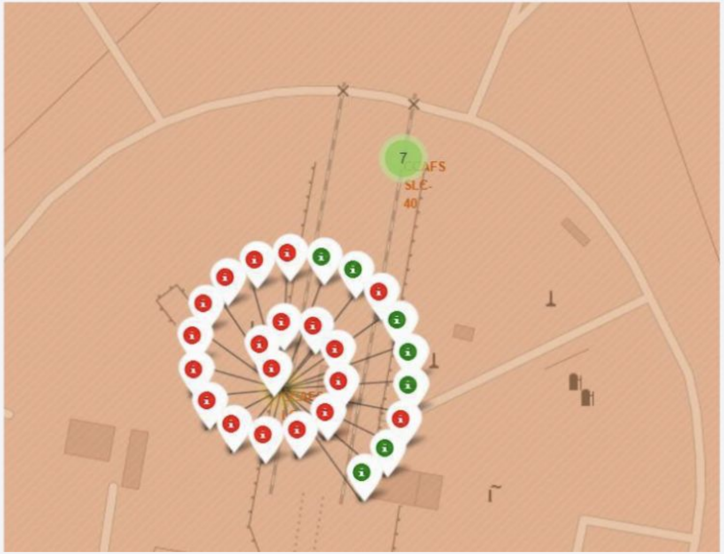
landing__outcome	COUNT
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

Section
3

Launch Sites Proximities Analysis

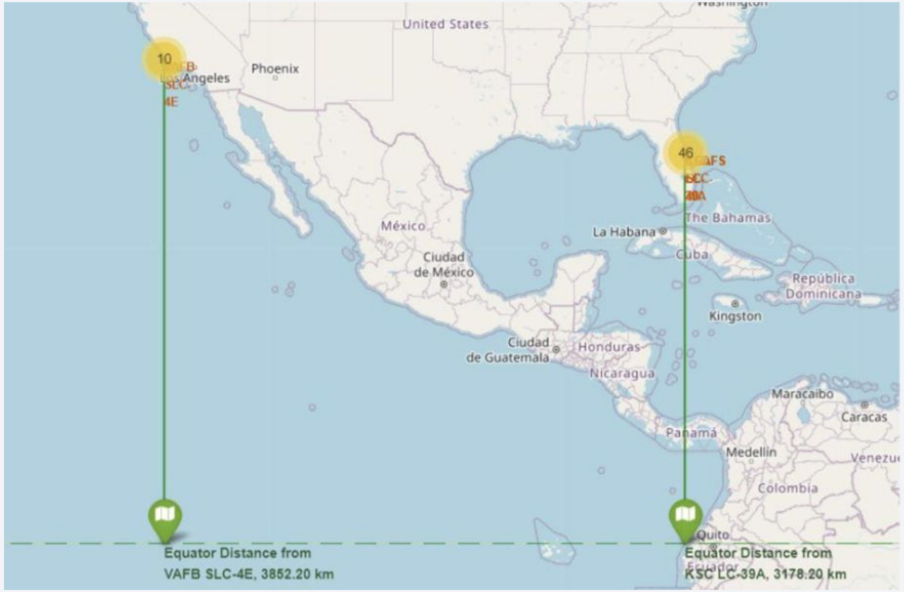
launch_site
CCAFS LC-40
CCAFS SLC-40
KSC LC-39A
VAFB SLC-4E

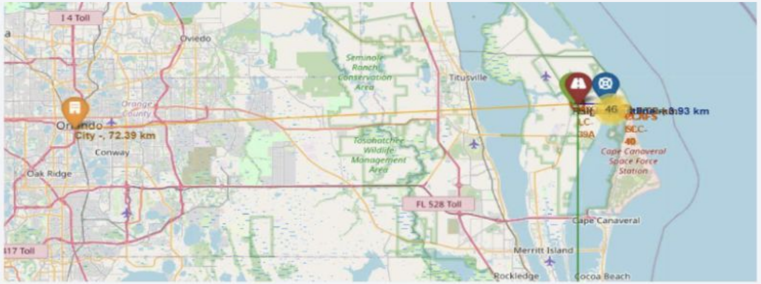
Launch Sites are situated on both the east and west coasts of the United States
Specifically in California and Florida.



Markers have been added to denote success and failure at each launch site.

To accommodate their size, they are consolidated in larger zooming views.





Launch Sites are directly at the coasts but ~ 10-15km away from the next big city and next highway, safe enough in case of an accident

Roughly 3000-4000km from the equator where the highest earth rotational speed of the earth is and thus, better launch conditions

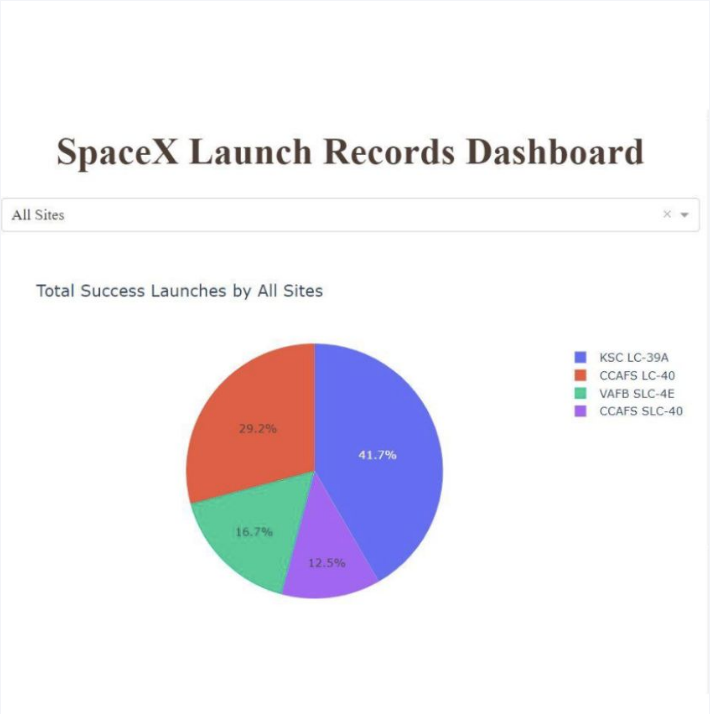
Section
4

Build a Dashboard with Plotly Dash

Launch success count for all sites

KSC-LC-39A has highest share of Success

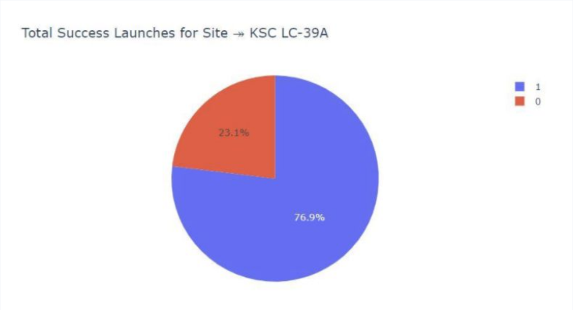
CCAFS.SLC-40 has lowest share of success



The launch site with highest launch success ratio

Mixed success for FT boosters while overall best

B4 and B5 version are well



Payload vs. Launch Outcome

In general, Booster category FT performs well for lower payload masses.

However, for payloads exceeding 5000kg, it tends to result in failure.

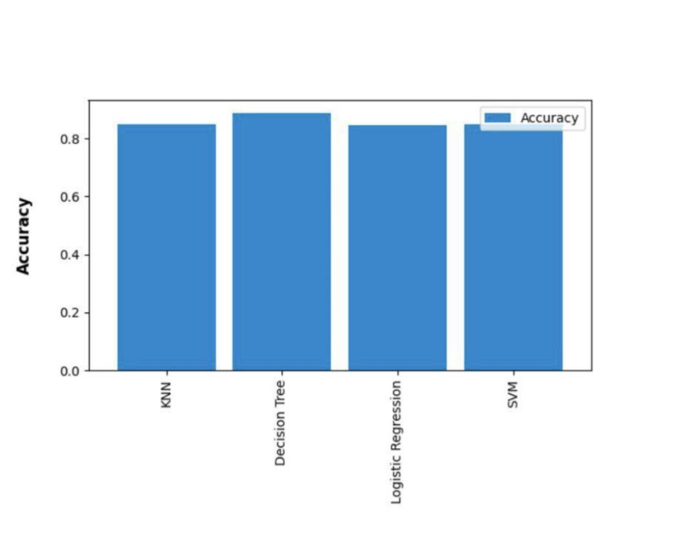
Booster version v1.1 performs poorly for mid-size payloads ranging from 2000kg to 7000kg.

Section 5

Predictive Analysis (Classification)

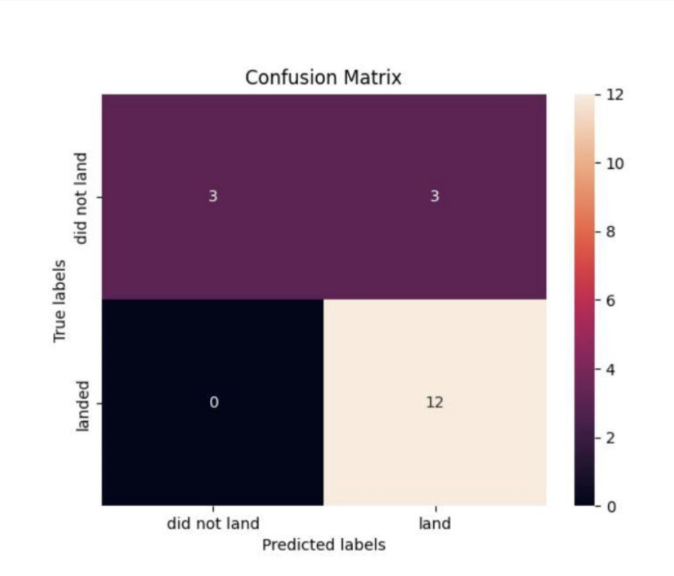
Classification Accuracy

Decision Tree Model is the best with an accuracy with arround 91%
Others perform only minorly worse with accuracy rates > 84%



Confusion Matrix

True/false versus Land/Not Land Matrix: Predicted landing outcomes for the test data (a subset of the original data) were assessed.
Regrettably, there were instances of True/Not-Land outcomes.
Nevertheless, the model achieved an overall accuracy of 15 out of 18 correct predictions.



Conclusions

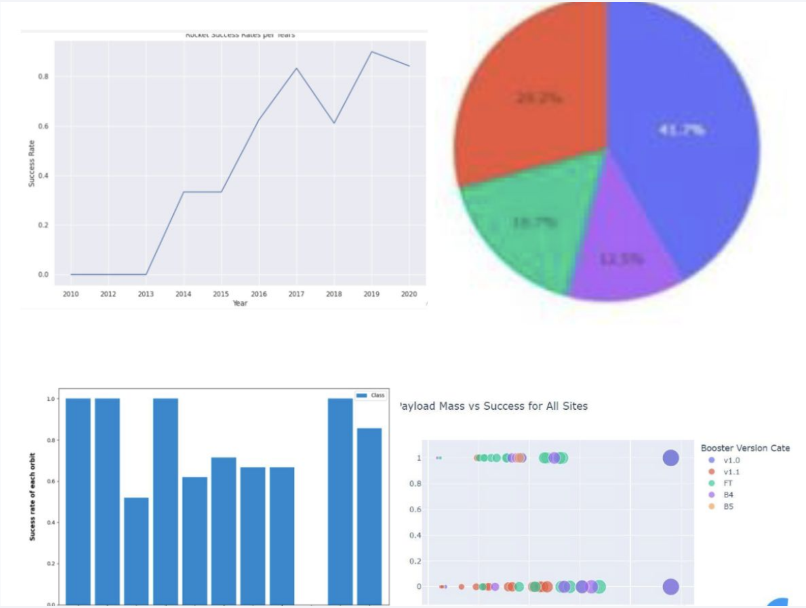
Orbits ES-L1, GEO, HEO, SSO have highest success rates

Success rate for launches increased over time

KSC LC-39A had the most successful launches

For higher payloads (> 5000kg) rather failure

Decision Tree is the best predictive Model



Thank you!