IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Shaheera
13.06.2022

Outline

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

Executive Summary

- **Summary of methodologies**
  - Data collection
  - Data wrangling
  - EDA with data visualization
  - EDA with SQL
  - Building an interactive map with Folium
  - Building a Dashboard with Plotly Dash
  - Predictive analysis (Classification)
- **Summary of all results**
  - Exploratory data analysis results
  - Interactive analytics demo in screenshots
  - Predictive analysis results

Introduction

- SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upwards of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage. The cost of a launch can be determined, if we can determine if the first stage will land.

- To compete with SpaceX, the aim of this project is to determine the price of each launch (for SpaceX). By gathering information about Space X and using a machine learning model and public information to predict if SpaceX will reuse the first stage.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
- ✓ SpaceX launch data is gathered from the SpaceX REST API.
- ✓ The data of launches, including information of rocket used, payload delivered, launch specifications, landing specifications, and landing outcome.
- Perform data wrangling.
- ✓ Calculate the number of launches on each site, the number and occurrence of each orbit, the number and occurrence of mission outcome per orbit type and Create a landing outcome label from Outcome column.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
- ✓ build a machine learning pipeline to predict if the first stage of Falcon 9 lands successfully.
- ✓ Include: Preprocessing, and Train test split, train the model and perform Grid Search, to find the hyperparameters.
- ✓ Using best hyperparameter values, to determine the model with the best accuracy.
- ✓ Test Logistic Regression, Support Vector Machines, Decision Tree Classifier, and K-nearest neighbors.
- ✓ Finally, output the confusion matrix.

Data Collection

- Data sets were gathered from an API, specifically the SpaceX REST API.
- This API gives data about launches, including information about the rocket used,
- payload delivered, launch specifications, landing specifications, and landing outcome.
- Request and parse the SpaceX launch data.
- Filter the data-frame to only include Falcon 9 launches.
- Dealing with Missing Values.

Data Collection – SpaceX API

Data Collection - Scraping

Web scraping Falcon 9 Launch records

2020 [edit]

In late 2019, Gwynne Shotwell stated that SpaceX hoped for as many as 24 launches for Starlink satellites in 2020,[440] in addition to 14 or 15 non-Starlink launches. At 26 launches, 13 of which for Starlink satellites, Falcon 9 had its most prolific year, and Falcon rockets were second most prolific rocket family of 2020, only behind China's Long March rocket family.[79]

[hide] Flight No.	Date and time (UTC)	Version, Booster[b]	Launch site	Payload[c]	Payload mass	Orbit	Customer	Launch outcome	Booster landing
78	7 January 2020, 02:19:21[440]	F9 B5 B1049.4	CCAFS, SLC-40	Starlink 2 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Success (drone ship)
	Third large batch and second operational flight of Starlink constellation. One of the 60 satellites included a test coating to make the satellite less reflective, and thus less likely to interfere with ground-based astronomical observations.[440]								
79	19 January 2020, 15:30[444]	F9 B5 B1046.4	KSC, LC-39A	Crew Dragon in-flight abort test[445] (Dragon C205.1)	12,050 kg (26,570 lb)	Sub-orbital[446]	NASA (CTS)[447]	Success	No attempt
	An atmospheric test of the Dragon 2 abort system after Max Q. The capsule fired its SuperDraco engines, reached an apogee of 40 km (25 mi), deployed parachutes after reentry, and splashed down in the ocean 31 km (19 mi) downrange from the launch site. The test was previously slated to be accomplished with the Crew Dragon Demo-1 capsule;[448] but that test article exploded during a ground test of SuperDraco engines on 20 April 2019.[449] The abort test used the capsule originally intended for the first crewed flight.[450] As expected, the booster was destroyed by aerodynamic forces after the capsule aborted.[451] First flig of a Falcon 9 with only one functional stage — the second stage had a mass simulator in place of its engine.								
80	29 January 2020, 14:07[452]	F9 B5 B1051.3	CCAFS, SLC-40	Starlink 3 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Success (drone ship)
	Third operational and fourth large batch of Starlink satellites, deployed in a circular 290 km (180 mi) orbit. One of the fairing halves was caught, while the other was fished out of the ocean.[452]								
81	17 February 2020, 15:05[453]	F9 B5 B1056.4	CCAFS, SLC-40	Starlink 4 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Failure (drone ship)
	Fourth operational and fifth large batch of Starlink satellites. Used a new flight profile which deployed into a 212 km x 386 km (132 mi x 240 mi) elliptical orbit instead of launching into a circular orbit and firing the second stage engine twice. The first stage booster failed to land on the drone ship[454] due to incorrect wind data.[455] This was the first time a flight proven booster failed to land.								
82	7 March 2020, 04:50[456]	F9 B5 B1059.2	CCAFS, SLC-40	SpaceX CRS-20 (Dragon C112.3 ♺)	1,977 kg (4,359 lb)[457]	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
	Last launch of phase 1 of the CRS contract. Carries Bartolomeo, an ESA platform for hosting external payloads onto ISS.[458] Originally scheduled to launch on 2 March 2020, the launch date was pushed back due to a second stage engine failure. SpaceX decided to swap out the second stage instead of replacing the faulty part.[459] It was SpaceX's 50th successful la[illegible] booster, the third flight of the Dragon C112 and the last launch of the cargo Dragon spacecraft.								
	18 March 2020, 12:16[460]	F9 B5 B1048.5	KSC, LC-39A	Starlink 5 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Failure (drone ship)

Web scraping with BeautifulSoup

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude
0	1	2006-03-24	Falcon 1	20.0	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	Merlin1A	167.743129	9.047721
1	2	2007-03-21	Falcon 1	NaN	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	Merlin2A	167.743129	9.047721
2	4	2008-09-28	Falcon 1	165.0	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	Merlin2C	167.743129	9.047721
3	5	2009-07-13	Falcon 1	200.0	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	Merlin3C	167.743129	9.047721
4	6	2010-06-04	Falcon 9	NaN	LEO	CCAFS SLC 40	None None	1	False	False	False	None	1.0	0	B0003	-80.577366	28.561857

Data Wrangling

- 1: Calculate the number of launches on each site.
- 2: Calculate the number and occurrence of each orbit.
- 3: Calculate the number and occurence of mission outcome per orbit type.
- 4: Create a landing outcome label from Outcome column.

EDA with Data Visualization

EDA with Data Visualization

EDA with SQL

- Store the dataset in database table
- Connect to the database
- Display 5 records where launch sites begin with the string 'CCA'
- Display the names of the unique launch sites in the space mission.
- Display the total payload mass carried by boosters launched by NASA (CRS)
- Display average payload mass carried by booster version F9 v1.1

EDA with SQL

- List the date when the first successful landing outcome in ground pad was achieved.
- List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000.
- List the total number of successful and failure mission outcomes.
- List the names of the booster_versions which have carried the maximum payload mass. Use a subquery.
- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015
- 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

Build an Interactive Map with Folium

1: Mark all launch sites on a map

To answer the following questions:

- Are all launch sites in proximity to the Equator line?
- Are all launch sites in very close proximity to the coast?

2: Mark the success/failed launches for each site on the map

To see which sites have high success rates of success.

3: Calculate the distances between a launch site to its proximities

To explore and analyze the proximities of launch sites.

- After you plot distance lines to the proximities, you can answer the following questions easily:
- Are launch sites in close proximity to railways?
- Are launch sites in close proximity to highways?
- Are launch sites in close proximity to coastline?
- Do launch sites keep certain distance away from cities?

Build a Dashboard with Plotly Dash

- The dashboard contains:
- Launch Site Drop-down menu: to let us select different launch sites.
- Range slider: to easily select different payload range and see if we can identify some visual patterns.
- Pie chart for all sites: to visualizing launch success counts of the selected launch site from site-drop downand.
- Scatter chart: to visually observe how payload may be correlated with mission outcomes for selected site.

Predictive Analysis (Classification)

- 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 performs best using test data.

Results

- The results of EDA with visualization, EDA with SQL, Interactive Map with Folium, and Interactive Dashboard will be shown in the next slides.
- Comparing the accuracy of the four methods, all return the same accuracy of about *83%* for test data.

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

- The results of EDA with visualization, EDA with SQL, Interactive Map with Folium, and Interactive Dashboard will be shown in the next slides.
- Comparing the accuracy of the four methods, all return the same accuracy of about *83%* for test data.

Payload vs. Launch Site

- `Class 0` (blue) represents unsuccessful launch, and `Class 1` (orange) represents successful launch.
- At first glance, the larger pay load mass, the higher the rocket's success rate, but it seems difficult to make decisions based on this figure because **no clear pattern can be found between successful launch and Pay Load Mass**.

Success Rate vs. Orbit Type

- Orbit types **SSO**, **HEO**, **GEO**, and **ES-L1** have **the highest success rates** (*100%*).
- On the other hand, the success rate of orbit type **GTO** is only *50%*, and it is the **lowest** except for type SO, which recorded failure in a single attempt.

Flight Number vs. Orbit Type

- `Class 0` (blue) represents unsuccessful launch, and `Class 1` (orange) represents successful launch.
- In most cases, the launch outcome seems to be correlated with the flight number.
- On the other hand, in **GTO** orbit, there seems to be **no** relationship between flight numbers and success rate.
- SpaceX starts with LEO with a moderate success rate, and it seems that VLEO, which has a high success rate, is used the most in recent launches.

Payload vs. Orbit Type

- `Class 0` (blue) represents unsuccessful launch, and `Class 1` (orange) represents successful launch.
- With heavy payloads the successful landing or positive landing rate are more for LEO and ISS.
- However, in the case of GTO, it is hard to distinguish between the positive landing rate and the negative landing because they are all gathered together.

Launch Success Yearly Trend

- Since 2013, the success rate has continued to **increase** until 2017.
- The rate decreased slightly in 2018.
- Recently, it has shown a success rate of about *80%*.

All Launch Site Names

- When the SQL `DISTINCT` clause is used in the query, only unique values are displayed in the `Launch_Site` column from the `SpaceX` table.
- There are four unique launch sites: CCAFS LC-40, CCAFS SLC-40, KSC LC-39A, VAFB SLC-4E

- Query

```
SELECT DISTINCT LAUNCH_SITE
FROM SPACEXTBL
```

- Result

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

Launch Site Names Begin with 'CCA'

- Only five records of the SpaceX table were displayed using `LIMIT 5` clause in the query.
- Using the `LIKE` operator and the percent sign (%) together, the `Launch_Site` name starting with CAA could be called.

- Query

```
SELECT * FROM SPACEXTBL
WHERE LAUNCH_SITE LIKE 'CCA%'
LIMIT 5
```

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

Total Payload Mass

Query:

```
%%sql
SELECT SUM(PAYLOAD_MASS__KG_) AS total_payload_mass_kg
FROM SPACEXTBL
WHERE CUSTOMER = 'NASA (CRS)'
```

Resoult:

total_payload_mass_kg
45596

Average Payload Mass by F9 v1.1

Query:

```
%%sql
SELECT AVG(PAYLOAD_MASS__KG_) AS avg_payload_mass_kg
FROM SPACEXTBL
WHERE BOOSTER_VERSION = 'F9 v1.1'
```

Resoult:

avg_payload_mass_kg
2928

First Successful Ground Landing Date

Query:

```
%%sql
SELECT MIN(DATE) AS first_successful_landing_date
FROM SPACEXTBL
WHERE LANDING__OUTCOME = 'Success (ground pad)'
```

Resoult:

first_successful_landing_date
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

Query:

```
%%sql
SELECT BOOSTER_VERSION
FROM SPACEXTBL
WHERE LANDING__OUTCOME = 'Success (drone ship)'
    AND (PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000)
```

Resoult:

booster_version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

Query:

```
%%sql
SELECT MISSION_OUTCOME, COUNT(*) AS total_number
FROM SPACEXTBL
GROUP BY MISSION_OUTCOME
```

Resoult:

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

Boosters Carried Maximum Payload

Query:

```
%%sql
SELECT DISTINCT BOOSTER_VERSION, PAYLOAD_MASS__KG_
FROM SPACEXTBL
WHERE PAYLOAD_MASS__KG_ = (
    SELECT MAX(PAYLOAD_MASS__KG_)
    FROM SPACEXTBL);
```

Resoult:

booster_version	payload_mass__kg_
F9 B5 B1048.4	15600
F9 B5 B1048.5	15600
F9 B5 B1049.4	15600
F9 B5 B1049.5	15600
F9 B5 B1049.7	15600
F9 B5 B1051.3	15600
F9 B5 B1051.4	15600
F9 B5 B1051.6	15600
F9 B5 B1056.4	15600
F9 B5 B1058.3	15600
F9 B5 B1060.2	15600
F9 B5 B1060.3	15600

2015 Launch Records

Query:

```
%%sql
SELECT LANDING__OUTCOME, BOOSTER_VERSION, LAUNCH_SITE
FROM SPACEXTBL
WHERE LANDING__OUTCOME = 'Failure (drone ship)' AND YEAR(DATE) = '2015'
```

Resoult:

landing__outcome	booster_version	launch_site
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

Query:

```
%%sql
SELECT LANDING__OUTCOME, COUNT(LANDING__OUTCOME) AS total_number
FROM SPACEXTBL
WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20'
GROUP BY LANDING__OUTCOME
ORDER BY total_number DESC
```

Resoult:

landing__outcome	total_number
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 sites on a map

- The left map shows all *SpaceX* launch sites, and the right map also shows that all launch sites are in the United States.
- As can be seen on the map, all launch sites are near the coast.

the success/failed launches for each site map

Launch sites in Florida

Launch site in California

- By clicking on the marker clusters, successful landing (green) or failed landing (red) are displayed.

the distances between a launch site to its proximities

Are launch sites in close proximity to railways? ***Yes***
Are launch sites in close proximity to highways? ***Yes***
Are launch sites in close proximity to coastline? ***Yes***

Do launch sites keep certain distance away from cities? ***Yes***

Section 4

Build a Dashboard with Plotly Dash

Pie chart of success Launches for all sites

- KSLC-39A records the most launch success among all sites.
- The VAFB SLC-4E has the fewest launch success

Highest Success ratio

- KSLC-39A has the highest success rate with 10 landing successes *(76.9%)* and 3 landing failures *(23.1%)*.

Scatter chart: payload Vs. Luanches outcomes.

Section 5

Predictive Analysis (Classification)

Classification Accuracy

- In the test set, **the accuracy of all models** was virtually the **same** at **83.33%**.

Confusion Matrix

- The models predicted 12 successful landings when the true label was successful and 3 failed landings when the true label was failure. But there were also 3 predictions that said successful landings when the true label was failure *(false positive)*.

Conclusions

- Orbital types SSO, HEO, GEO, and ES-L1 have the highest success rate (100%).
- The launch site is close to railways, highways, and coastline, but far from cities.
- KSLC-39A has the highest number of launch successes and the highest success rate among all sites.
- The launch success rate of low weighted payloads is higher than that of heavy weighted payloads.

Appendix

- GitHub
- Dashboard Python Code

Thank you!