IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Gaxxx Txxx
03. 25. 2025.

Outline

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

Executive Summary

- Summary of methodologies (6th slide)
- Summary of all results (16th slide)

- Project background and context

SpaceX has revolutionized the aerospace industry by pioneering the recovery and reuse of rockets, significantly reducing costs per mission.

Since the mid-2000s, it has dominated the global market, making competition challenging due to its cost-efficiency and innovation in reusable technology.

- Problems we want to find answers

How can SpaceY go out and compete against SpaceX?

If SpaceX's greatest strength is the recovery of parts, how likely is it that it will successfully recover them?

Section 1

Methodology

Executive Summary

- Data collection methodology:
  - Using SpaceX API and Requests Python library.
  - Using BeautifulSoup Python library on a Wikipedia page to fetch a HTML table.
- Perform data wrangling
  - Missing values replacement with the mean value of the column.
  - Feature engineering
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

  Using Train and test subsets, cross validation, accuracy rate evaluation and confusion matrix analysis.

How were the datasets collected?

The data were collected in two different ways:

1. First, we made a Get request to the SpaceX API, to get records (JSON file) of SpaceX launches (all rockets in general).
2. Then, we webscraped a wikipedia page that contained a HTML table with records of the Falcon 9 launches.

- API :

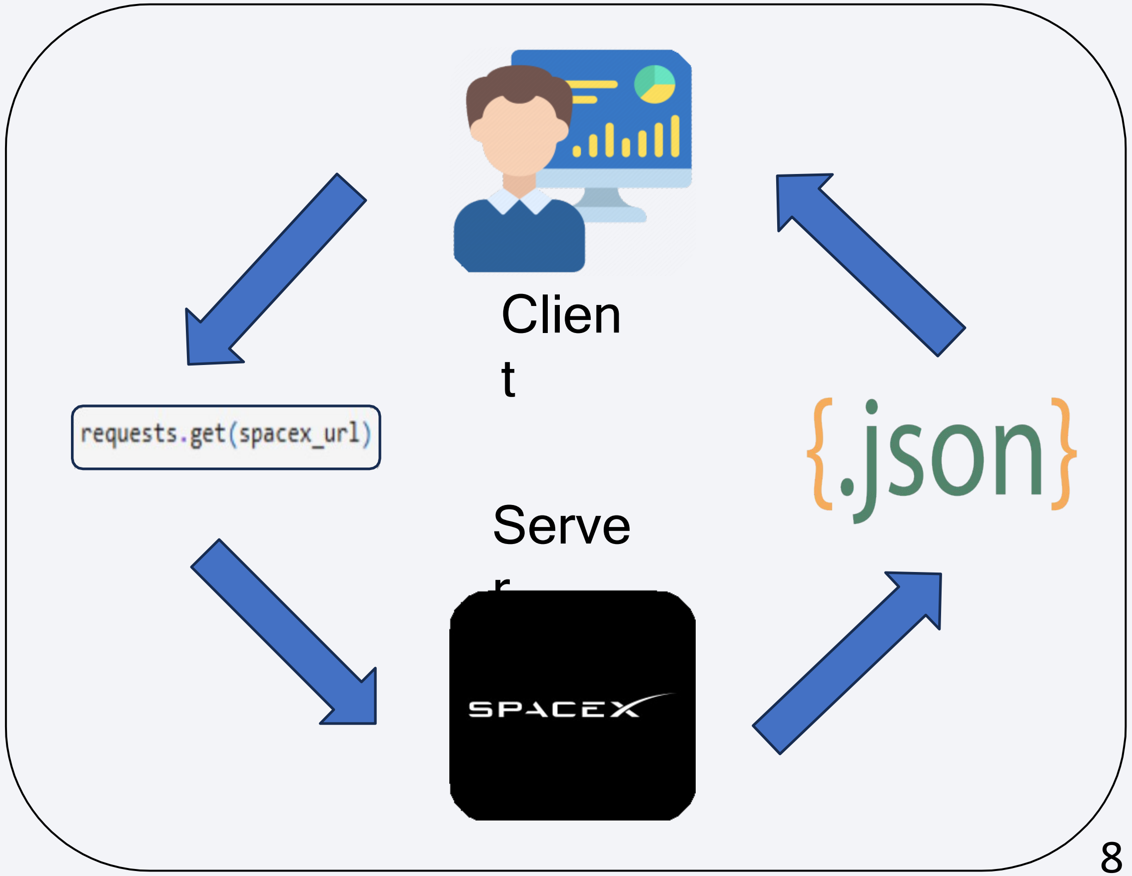
https://api.spacexdata.com/v4

/launches/past

- GitHub URL :

[Applied-Data-Science-Capstone/jupyter-labs-spacex-data-collection-api.ipynb at main · TheBarronInDecline/Applied-Data-Science-Capstone]

HTTP protocol



Data Collection - Scraping

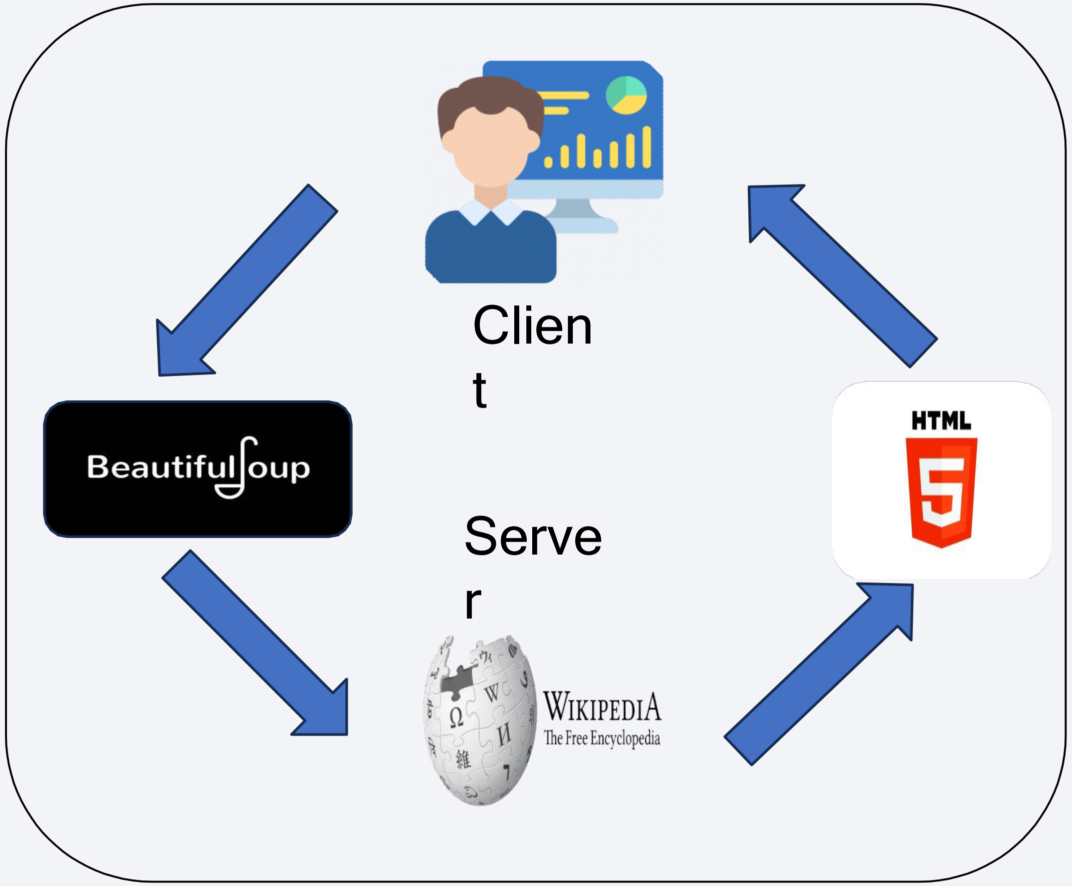
- Source :

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

- GitHub URL :

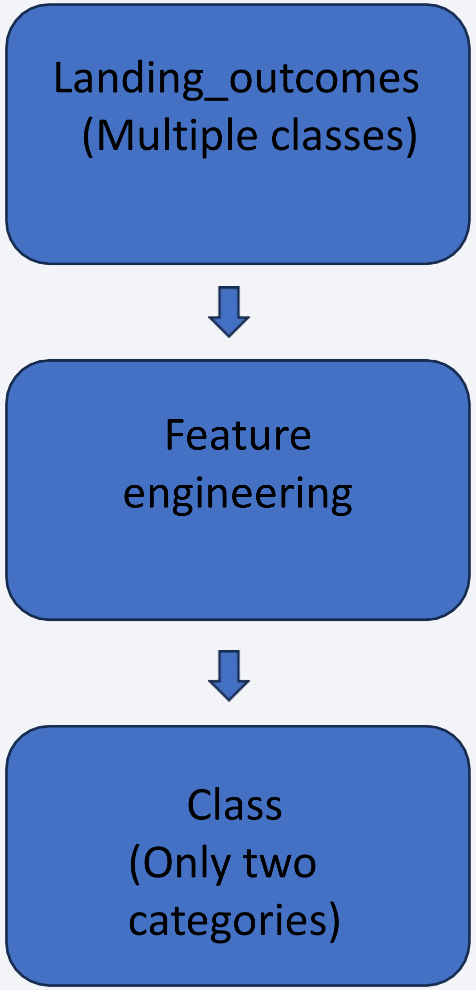
Applied-Data-Science-Capstone/jupyter-labs-webscraping.ipynb at main · TheBarronInDecline/Applied-Data-Science-Capstone

HTTP protocol



Data Wrangling

- In the data obtained from the API we had to convert some missing values in the average value of the column, the data was also filtered only to consider falcon 9 missions.
- In the wikipedia data, an irrelevant column was eliminated. However, the "Landing Pad" column retained its 28 missing values, since these represent that a landing pad was not used in the mission.
- A feature called "class" was created, to be able to summarize the different categorical values of the "landing_outcomes" column, to only two categories: positive (1) or negative (0) result.



Applied-Data-Science-Capstone/labs-jupyter-spacex-Data

EDA with Data Visualization

In total we visualized the data with 7 charts, one line chart, one bar chart and five scatter plots. The purposes of these charts were:

Scatter plots:
The data points have a color mapping depending on whether they belong to the category of successful or failed landings. Plots retract the following correlations:
-Flight number & payload mass
-Flight number & launch site
-Payload mass & launch site
-Flight number & orbit
-Payload mass & orbit

Bar chart:
One chart to visualize de relationship between success rate (the mean of the variable “class”) of each orbit:
-GEO
-HEO
-MEO
-GTO
-etc

Line chart:
One chart to visualize the launch success yearly trend (the mean of variable “class” for each year). The ten years trend, 2010-2020.

[Applied-Data-Science-Capstone/edadataviz.ipynb at main · TheBarronInDecline/Applied-Data-Science-Capstone]()

EDA with SQL

- Display all categorical values of launch site column.
- Display 5 missions which took place in Cape Canaveral.
- Display the total payload mass of the NASA(CRS) missions.
- Display the average payload mass for the v1.1 version of Falcon 9 booster.
- Display the date of the first successful Falcon 9 landing on a ground pad.
- Display a list of al the boosters which have a load between 4000kg and 6000kg and successfully landed in a drone ship.
- Display the total number of successful missions and failed ones.
- Display a list of all boosters which have carried the maximum payload mass.
- Display a table that shows month, landing outcome, booster version and launch site only for missions in the year 2015.
- Display a count of all different landing outcomes between **2010-06-04** and **2017-03-20**, in descending order.

Applied-Data-Science-Capstone/jupyter-labs-eda-sql-coursera_sqllite.ipynb at main ·

Build an Interactive Map with Folium

On a map of the United States, we set the following Map Objects:

- I added popup labels with the names of each launch site and drew circles around them. This allows quick insights into their proximity to the ocean and the equator.
- I used clustering to display all mission results–successful (green markers) and unsuccessful (red markers)–for each launch site. This helps visualize which sites achieve higher success rates.
- I added mouse position tracking to the map for retrieving coordinates. Using these, I created popup labels showing the distances from launch sites to key infrastructures and added polylines for better visualization. This helps assess proximity to coastlines, logistics, urban areas, and reserves.

Applied-Data-Science-Capstone/lab_jupyter_launch_site_location.ipynb at main · TheBarronInDecline/Applied-Data-Science-Capstone

Build a Dashboard with Plotly Dash

In my dashboard I displayed one interactive pie chart and one interactive scatter plot:

- Both change depending on whether the user selects via dropdown to see a

  report for all sites simultaneously or for a particular one.

- The pie chart displays the distribution of successful missions across all four sites when all are selected, highlighting the most and least successful ones. When a specific site is selected via the dropdown, it shows that site's success and failure rates.

- The scatter plot visualizes the correlation between Payload Mass and Mission Outcome, with color mapping for Booster versions. Users can filter by site using a dropdown or adjust the Payload Mass range via a slider for tailored insights.

- [Applied-Data-Science-Capstone/spacex-dash-app.py at main · TheBarronInDecline/Applied-Data-Science-Capstone]

Predictive Analysis (Classification)

Four classification models were developed to predict the landing outcome, logistic regression, decision trees, SVM and K-NN. All models were developed, optimized and evaluated with the following procedure:

Dataset Split:
With the standardized data, I créated training and testing subsets:

-Testing = 20%

-Training = 80%

Cross validation:
With the train set and a parameter dictionary, I run a cross validation grid search (10 folds) in order to find the best hyperparameters for each model

Final test:
I make predictions, Using the test set, I evaluate the quality of the predictions with:

-Accuracy rate

-Confusion matrix

[Applied-Data-Science-Capstone/SpaceX_Machine Learning Prediction_Part_5.ipynb at main · TheBarronInDecline/Applied-Data-Science-Capstone]

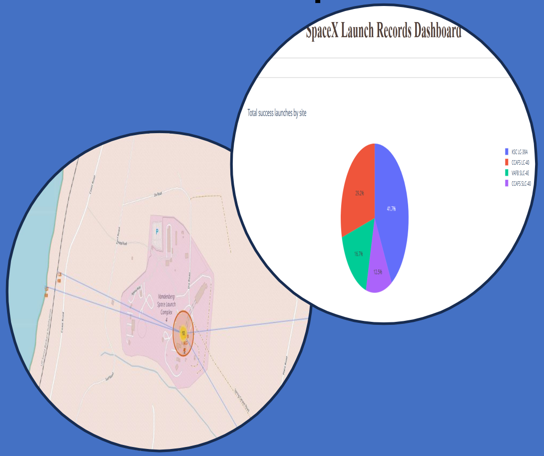
EDA results:

After analysis, it is discovered that the success of the Falcon 9 is not something that happened overnight. An important correlation is noted between the success of landings and:

-The passage of time,
-The increase in the number of missions carried out,
-The development of new booster versions.

Interactive analytics:

Interactive Dashboard and Folium map.



Predictive Analysis:

Accuracy:

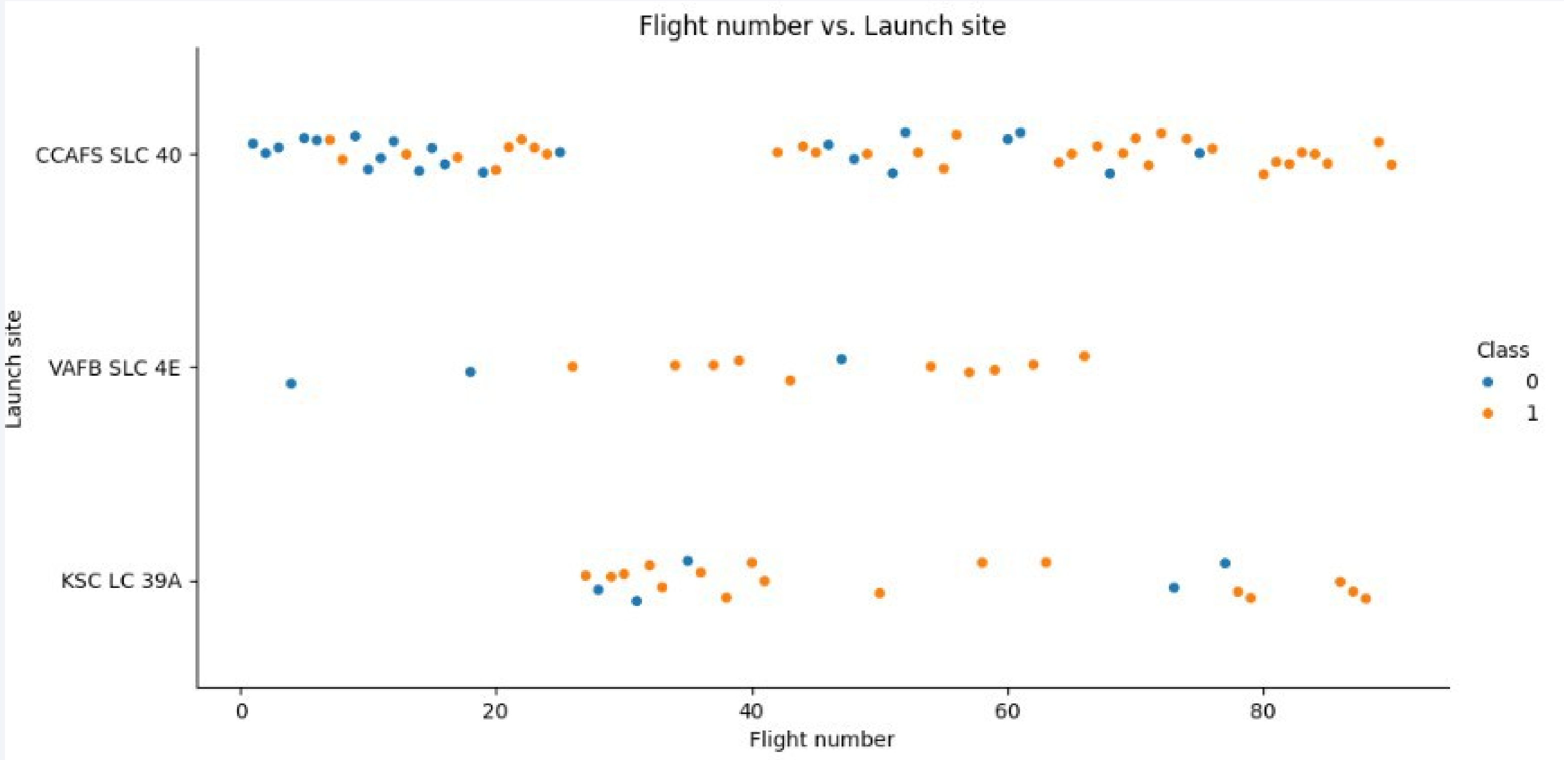
83%

All developed models perform the same.

Section 2

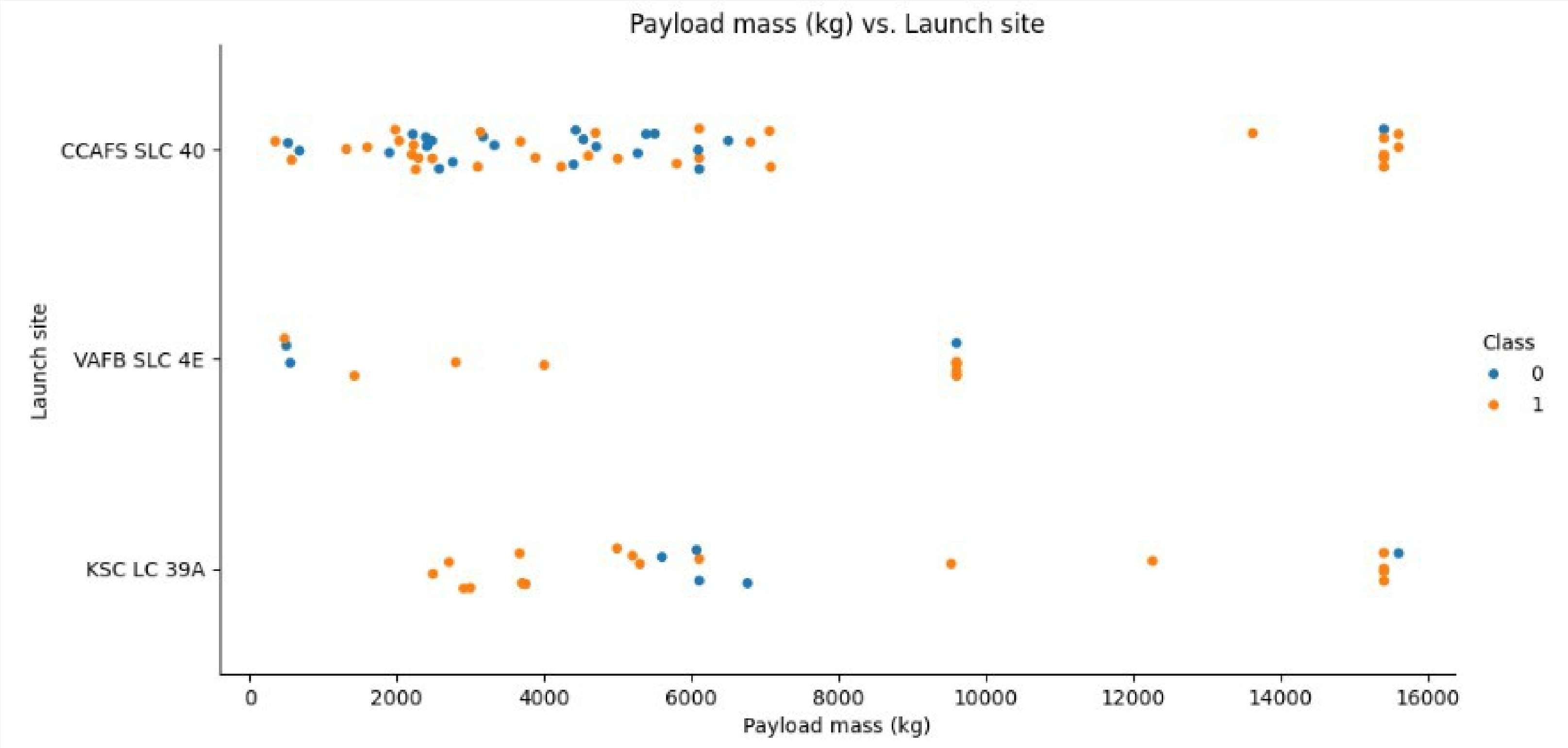
Insights drawn from EDA

Flight Number vs. Launch Site



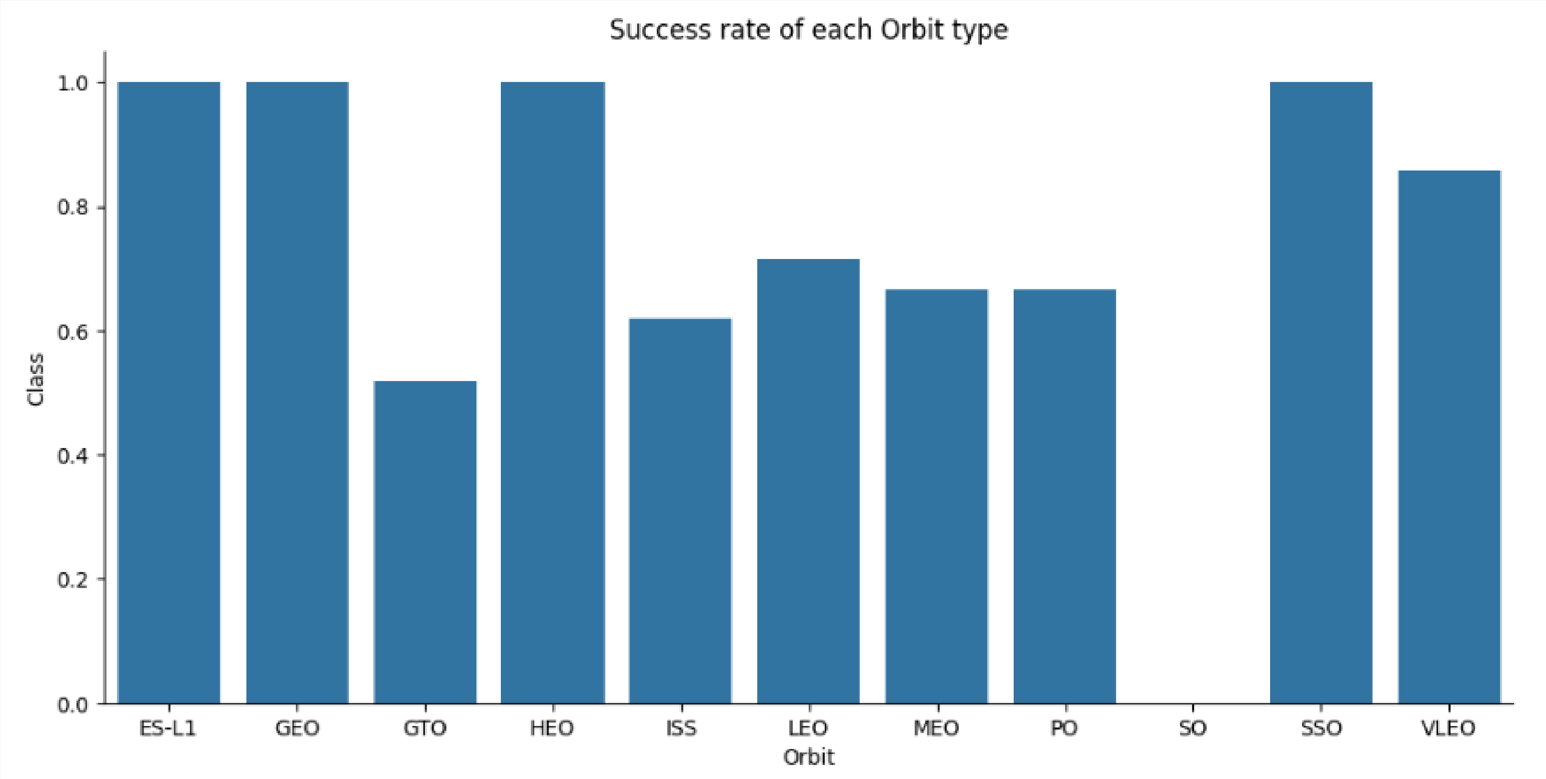
This scatter plot lets us appreciate (among other things) that as the mission's series progressed, the Falcon 9 improved the result of its landings.

Payload vs. Launch Site



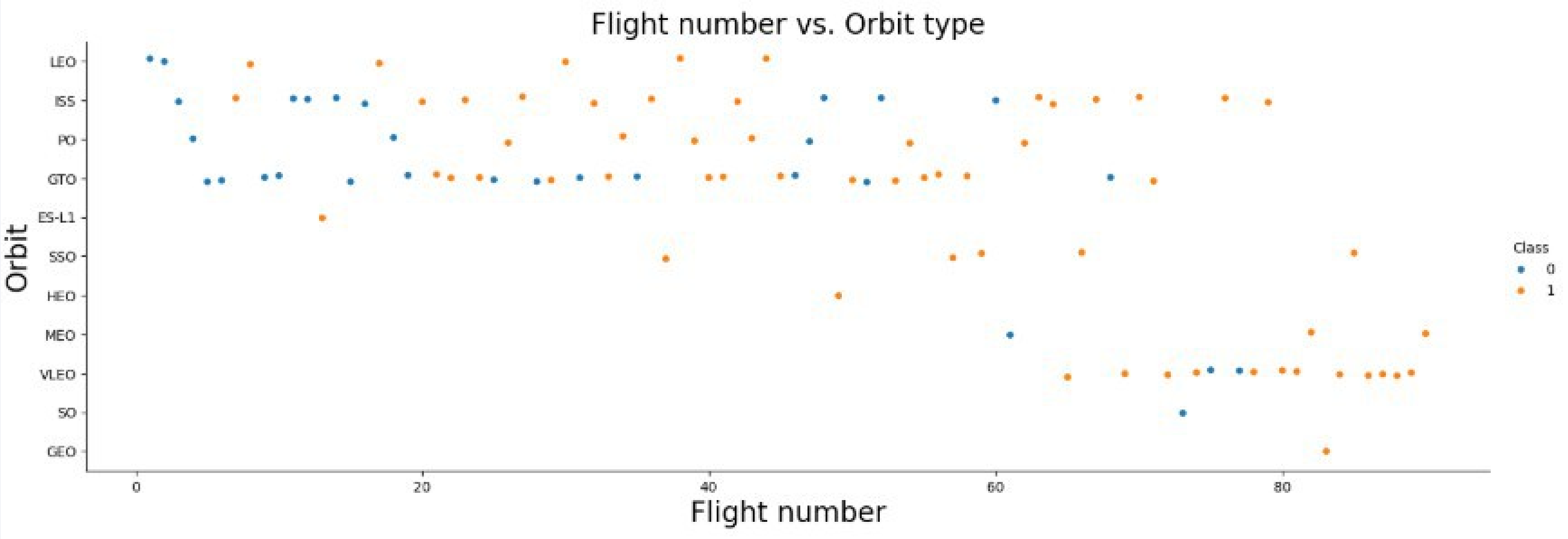
This scatter plot allows us to appreciate that at the Vandenberg (CA) site, rockets with a payload of 10000 kg or more are never launched.

Success Rate vs. Orbit Type



ES_L1, GEO, HEO and SSO orbits have the highest success rate.

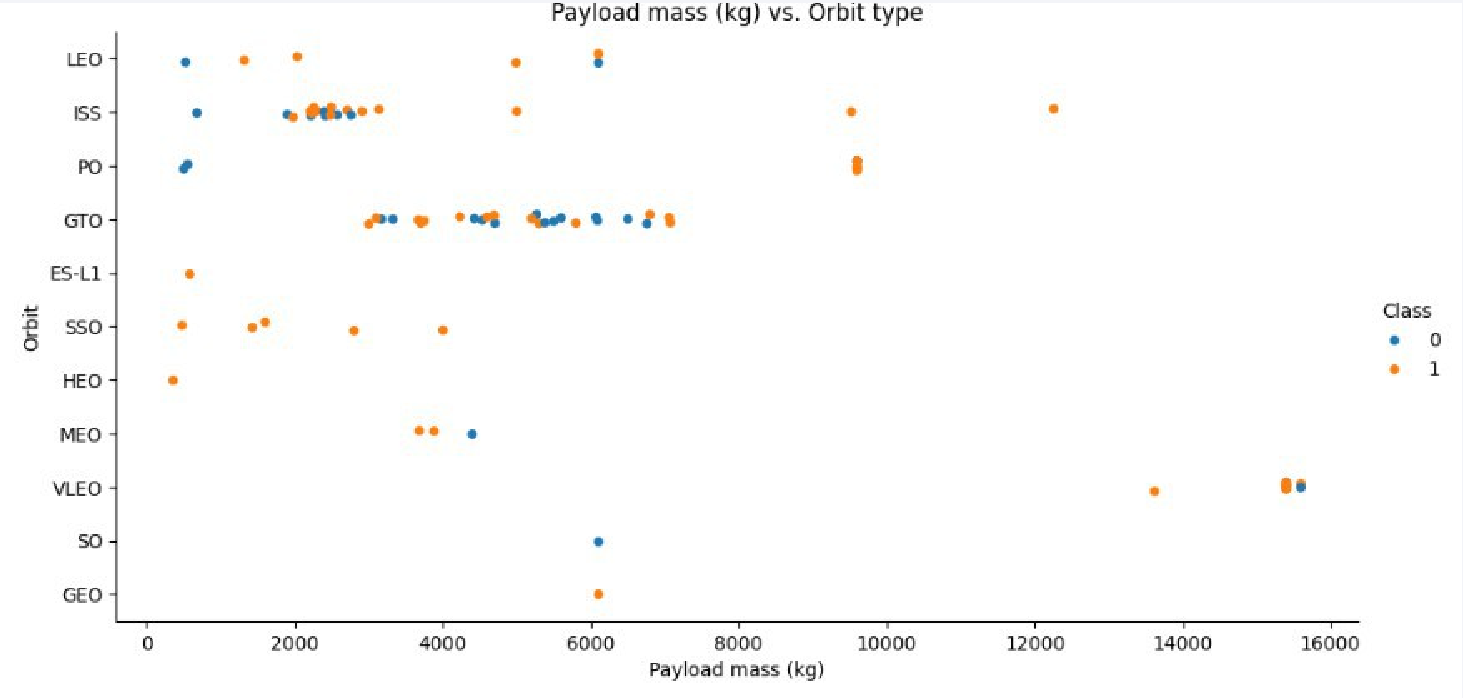
Flight Number vs. Orbit Type



We can see that in LEO orbit the success of the landings improved with the passage of the series of missions.

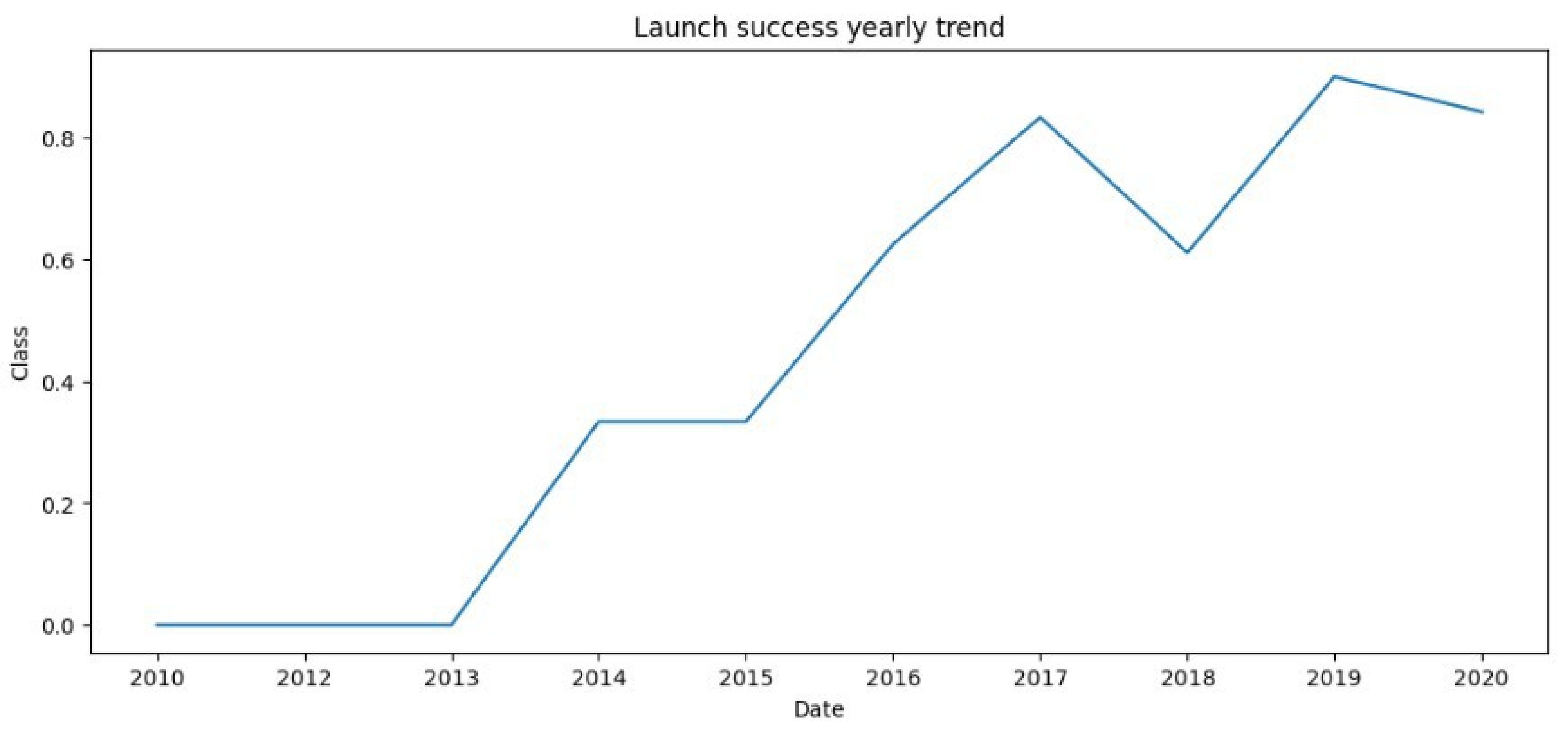
However, other orbits such as GTO do not prove that any
correlation exists.

Payload vs. Orbit Type



In the LEO, ISS and PO orbits we can see that the higher the payload mass, the better the landing outcome.
However, for the GTO orbit it is difficult to define whether the higher payload mass improves the outcome or ruins it.

Launch Success Yearly Trend



It seems that since 2013 the Falcon 9 launches are on the road to success, all the way to 2020 (and beyond stars).

All Launch Site Names

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

- CCAFS LC-40, KSC LC-39A and CCAFS SLC-40 are all launch sites at Cape Canaveral.
- VAFB SLC-4E is the launch site in Vandenberg, California.

Launch Site Names Begin with 'CCA'

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

SpaceX has been experimenting with the Falcon 9 since 2010 at CCAFS site.

Total Payload Mass

TotalPayloadMass_NASA_CRS
45596

A total of 45596 KG were transported for NASA(CRS) on the Falcon 9 missions.

Average Payload Mass by F9 v1.1

AveragePayloadMass_Booster_F9_v11
2534.666666666665

The V1.1 version of the Falcon 9 booster can carry an average payload mass of 2534.66 kg.

First Successful Ground Landing Date

First_SuccesfulLanding_In_GroundPad
2015-12-22

The first successful landing on a Ground pad occurred on December 22, 2015.

Successful Drone Ship Landing with Payload between 4000 and 6000

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

Only these four versions of the Falcon 9 booster are capable of landing on the ship drone with a payload mass of 4000-6000 kg.

Total Number of Successful and Failure Mission Outcomes

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

Boosters Carried Maximum Payload

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

These versions of the Falcon 9 booster are
capable of carrying maximum payload mass.

2015 Launch Records

This is List of the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015.

Months_2015	Landing_Outcome	Booster_Version	Launch_Site
01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
02	Failure (drone ship)	F9 v1.1 B1013	CCAFS LC-40
03	Failure (drone ship)	F9 v1.1 B1014	CCAFS LC-40
04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40
04	Failure (drone ship)	F9 v1.1 B1016	CCAFS LC-40
06	Failure (drone ship)	F9 v1.1 B1018	CCAFS LC-40
12	Failure (drone ship)	F9 FT B1019	CCAFS LC-40

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

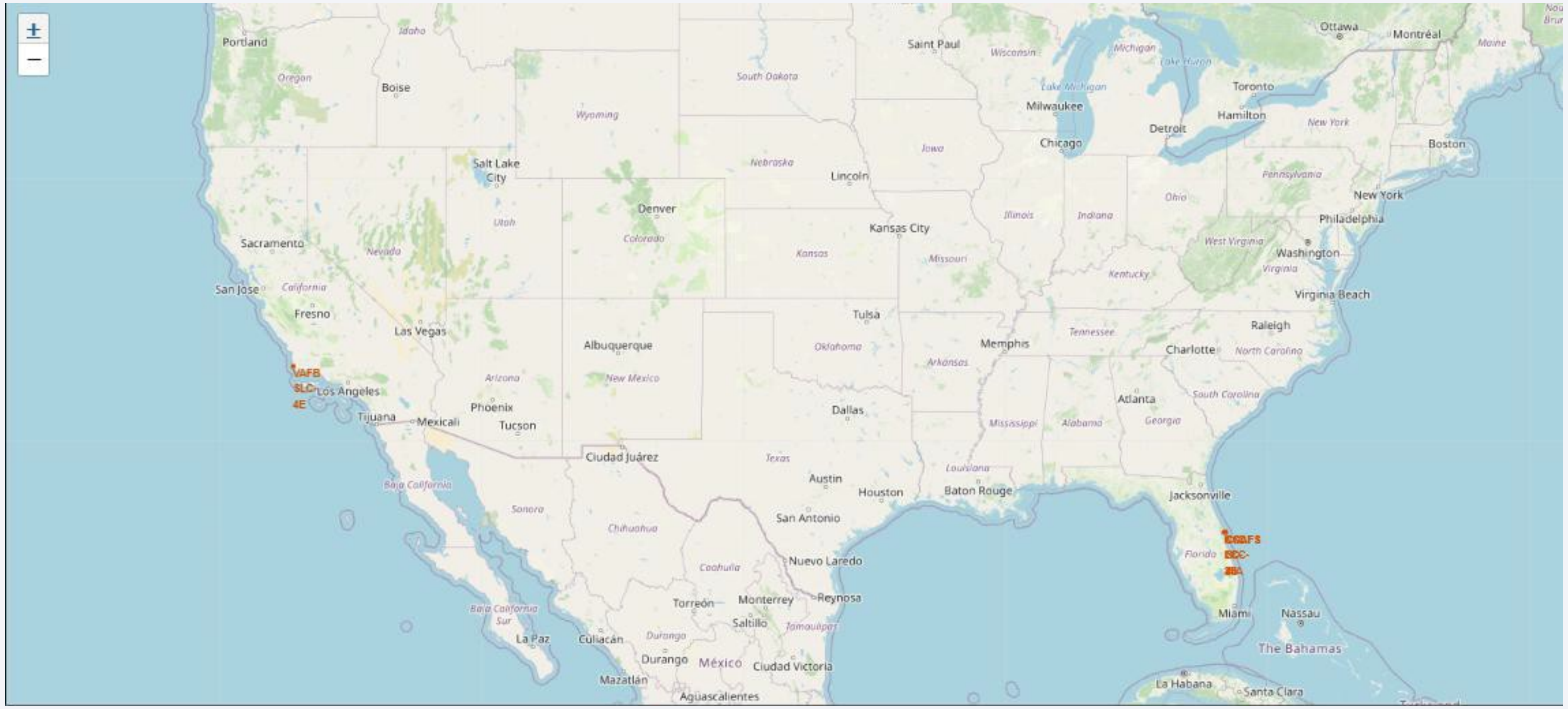
Landing_Outcome	count
Controlled (ocean)	3
Failure (drone ship)	5
Failure (parachute)	2
No attempt	10
Precluded (drone ship)	1
Success (drone ship)	5
Success (ground pad)	3
Uncontrolled (ocean)	2

This is the distribution of all landing outcomes between 2010-06-04 and 2017-03-20, among the different classes of the variable.

Section 3

Launch Sites Proximities Analysis

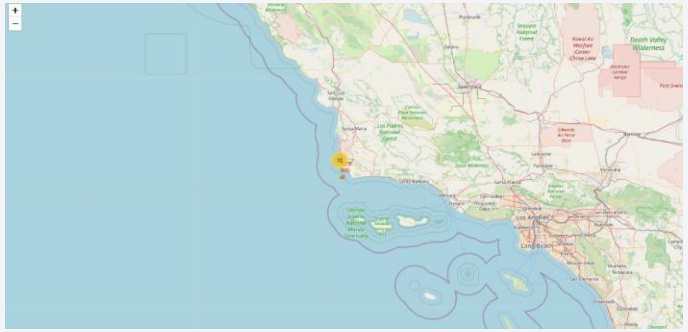
All Launch sites.



Launch sites are located close to the equator to maximize the use of Earth's rotational speed, which helps save fuel during takeoff. Additionally, they are positioned near oceans to reduce risks in case of mission failures, as any debris would fall into water instead of populated areas.

Landing outcomes for each site.

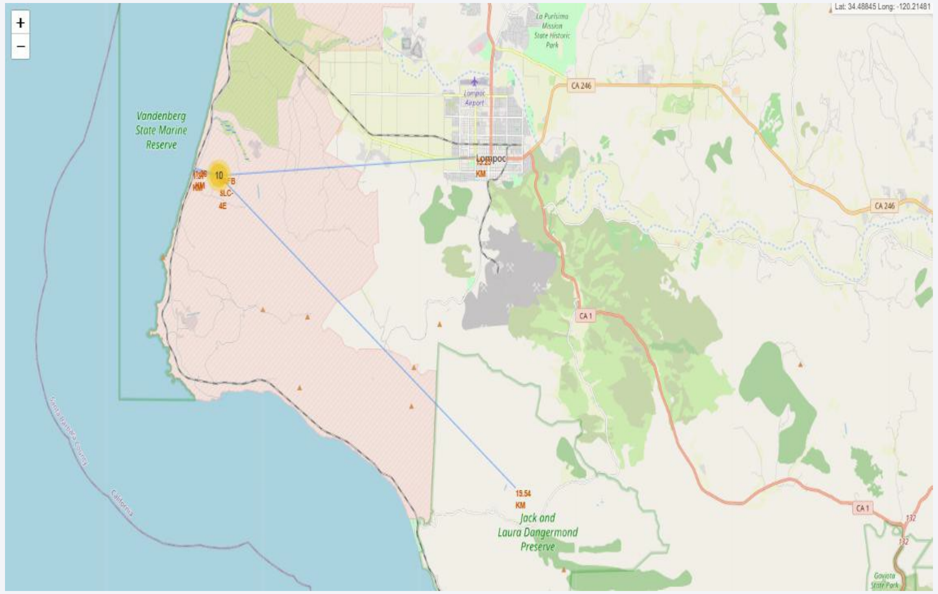
Vandenberg (CA) cluster



Vandenberg (CA) landing outcomes

Clustering all Mission Outcomes by site allows us to appreciate the success rate of each site, for example the Vandenberg site has a success rate of 40%.

Surrounding infrastructure and geography.



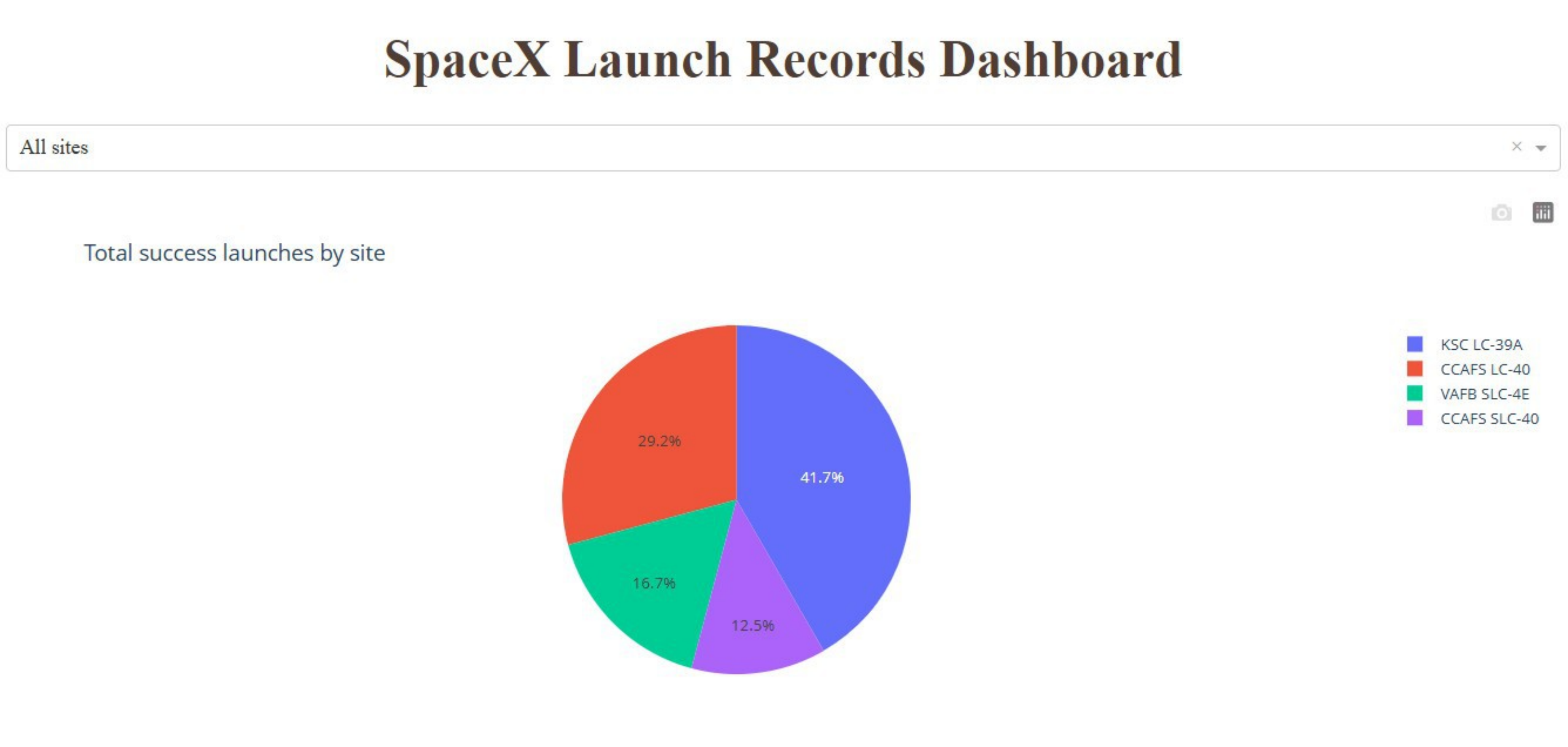
We can appreciate the proximity of the sites with key geographical characteristics (coastline) and logistical infrastructure (train tracks).

We can also see the remoteness of the site from the cities(Lompoc,CA) and nature reserves (Jack and Laura Dangermond preserve).

Section 4

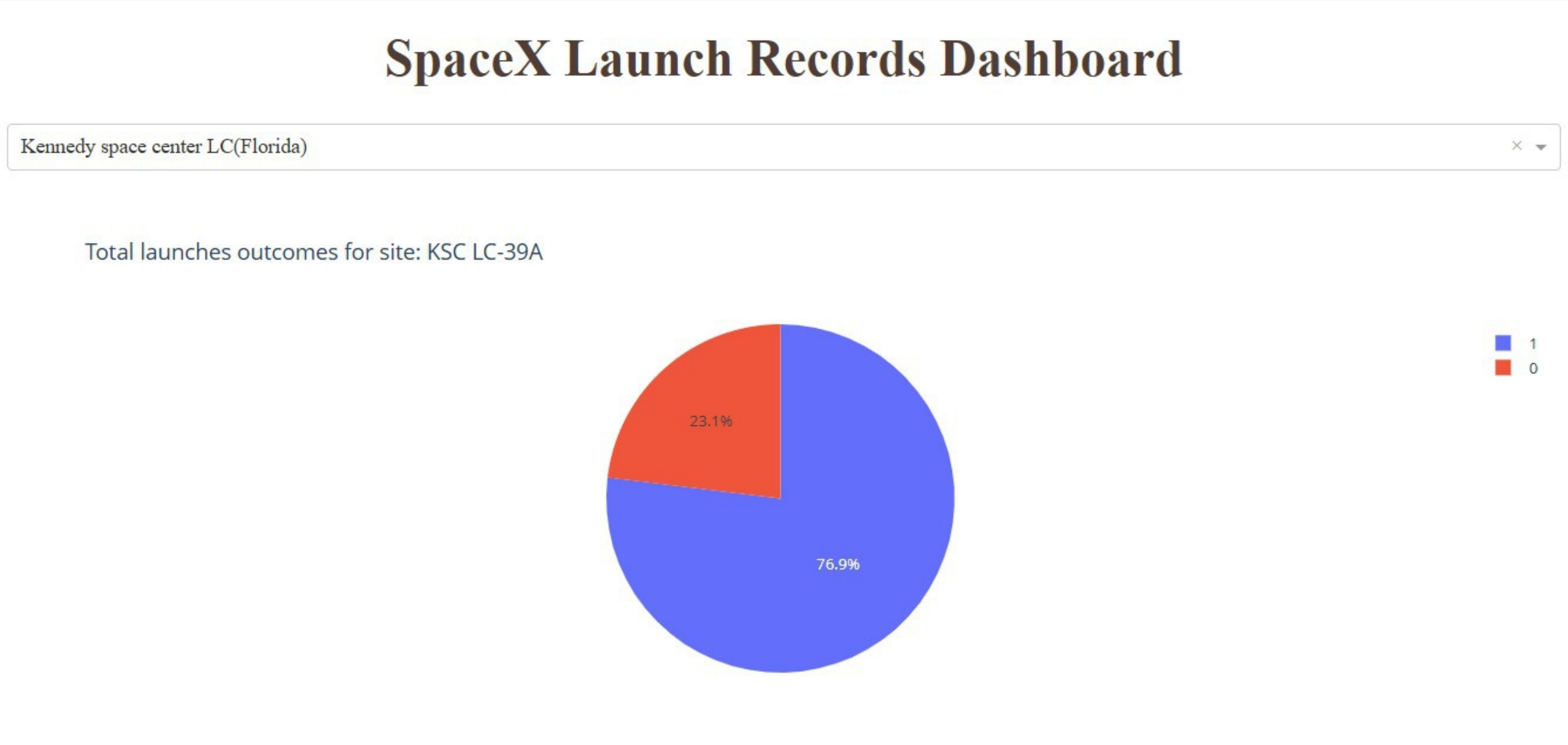
Build a Dashboard with Plotly Dash

Successful landings for all sites.



Most of the successful landings occurred at Kennedy Space Center.

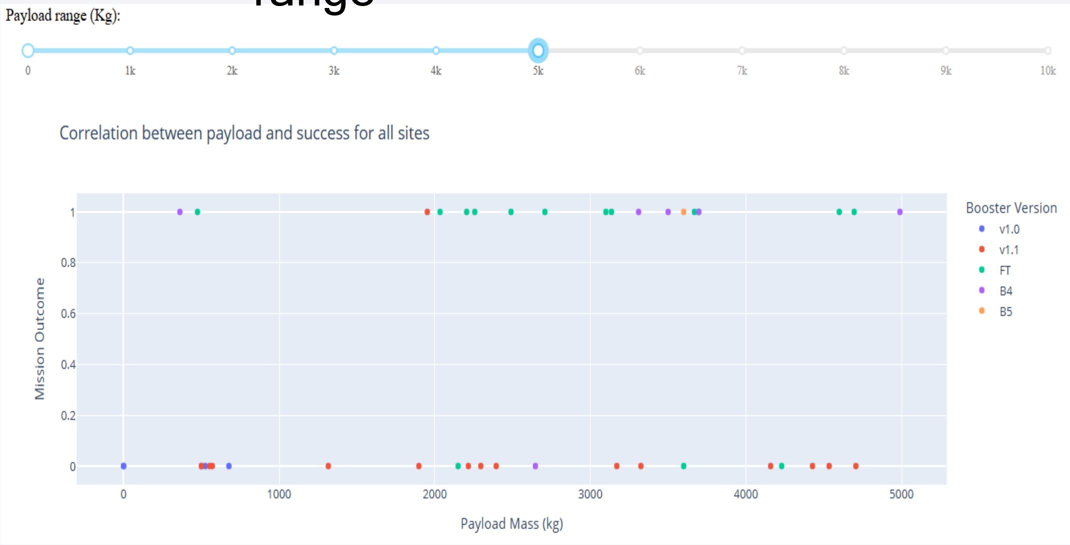
Highest success rate



The Kennedy Space Center has by far the highest success rate. Its success rate is about 77%.

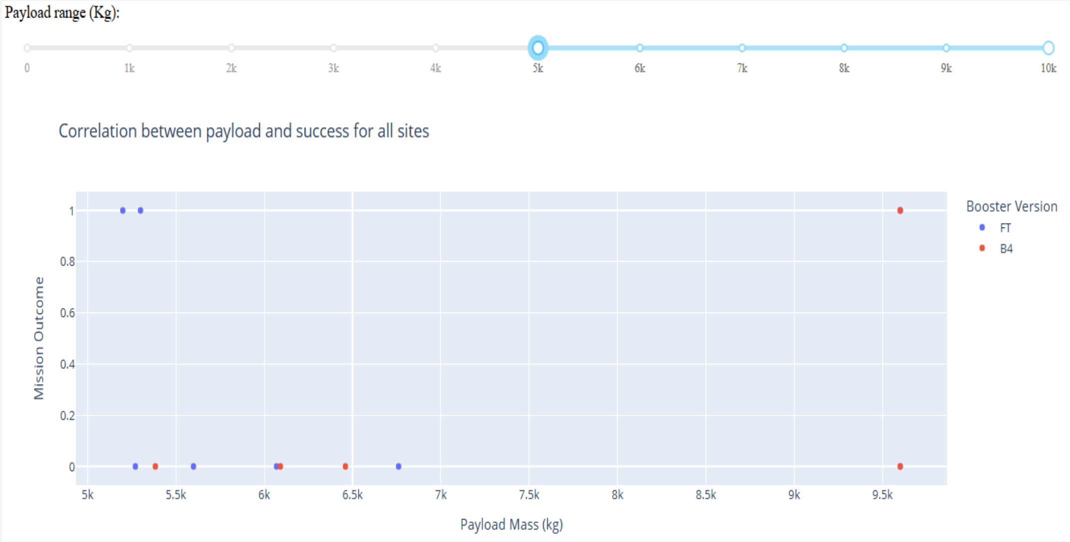
Payload mass VS. Launch Outcome for each booster version

0-5000 KG Payload range



In this range, the B4 booster version of the Falcon 9 is the version with the highest success rate

5000-10000 KG Payload range

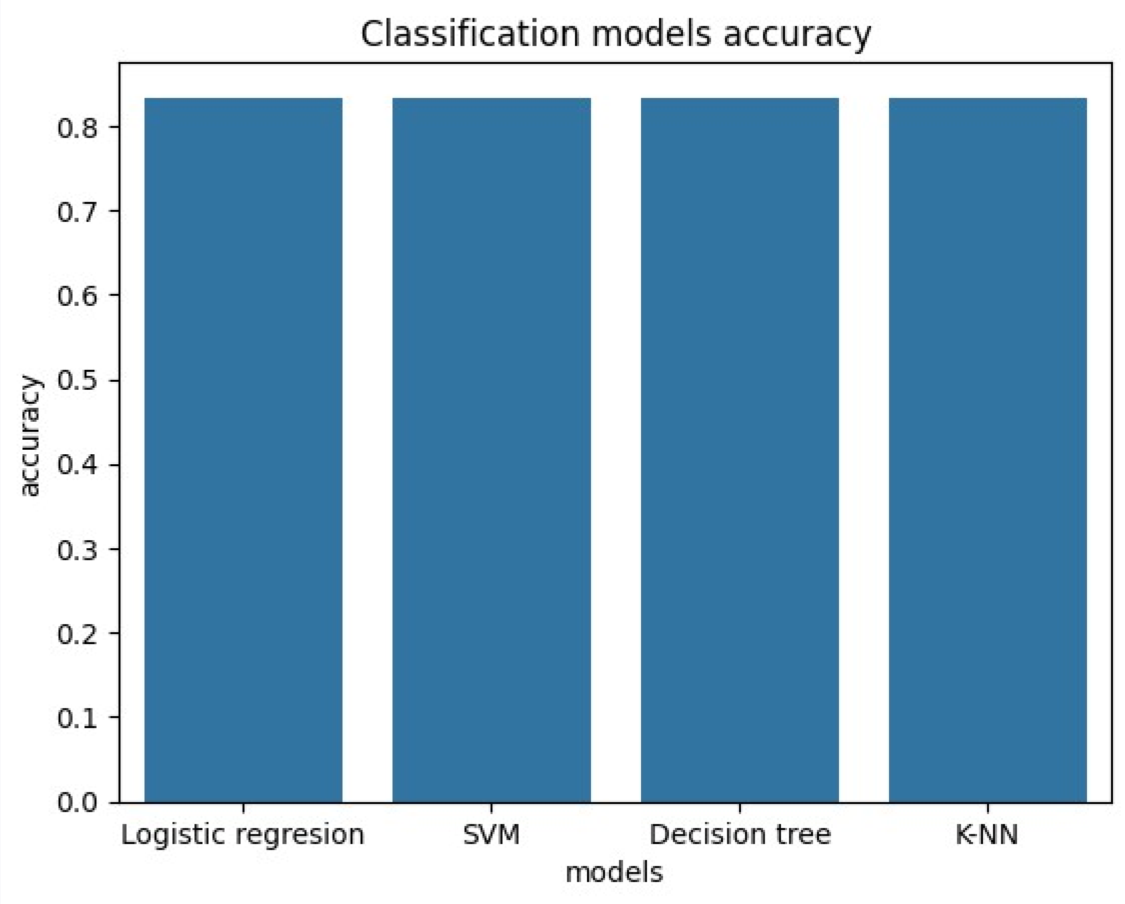


In this range only FT and B4 versions can carry the load, and only B4 could carry payload mass up to 9600 kg.

Section 5

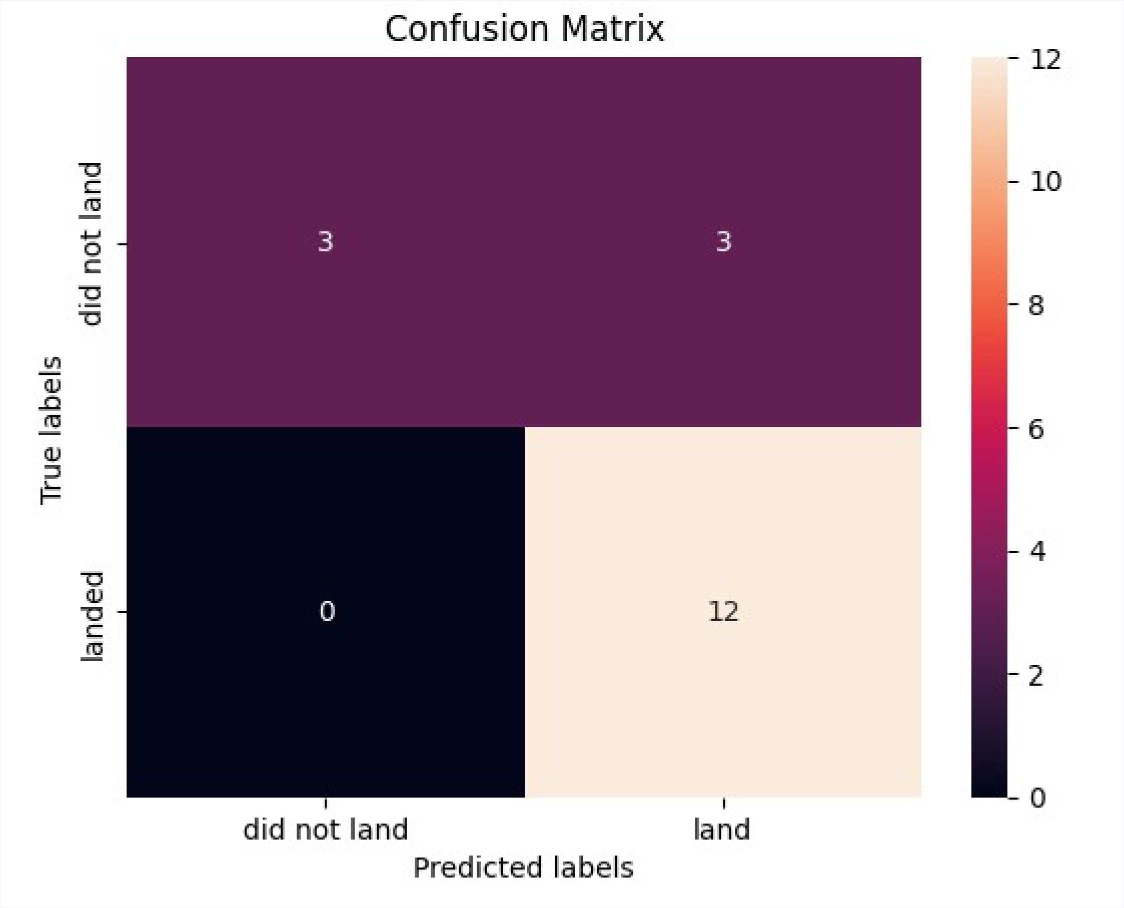
Predictive Analysis (Classification)

Classification Accuracy



All four models have same
accuracy rate, about 83%.

Confusion Matrix



All four models all have the same confusion matrix:

15 correct guesses, 12 true positives and 3 true negatives.

3 errors, all false positives.

Conclusions

- If we wanted to compete against SpaceX we would have a particularly difficult time. Throughout this project, it was shown that SpaceX has a very solid engineering behind its Falcon 9 launch and recovery process. SpaceX took years and many updates to reach the formula that defines its success today.
- Most missions end with a successful outcome. Betting against successful landings is a lost battle, as the success rate may not only be maintained but continue to grow, as SpaceX continues to test and develop new boosters
- It would be quite difficult to reach SpaceX's level of engineering, since as I said before, the company took several years, a lot of tests and a lot of errors to get to where it is today. Basically, the success of the competition depends essentially on SpaceX making mistakes, something that currently seems unlikely.
- I think one mistake I recorded in this project was the lack of data in the stage of developing classification models, it would be worth going back in the data science lifecycle, to ensure a greater abundance of data to train, validate and evaluate the model.

Appendix

All classification reports:

Syntax

```python
from sklearn.metrics import classification_report
trained_models = [logreg_cv, svm_cv, tree_cv, knn_cv]
model_names = ["Logistic Regression", "SVM", "Decision Tree", "KNN"]
for model, name in zip(trained_models, model_names):
    y_hat = model.predict(X_test)
    print(f"\n {name} testing data classification report:")
    print(classification_report(Y_test, y_hat))
    print("-" * 50)
```

```
 Logistic Regression testing data classification report:
              precision    recall  f1-score   support

           0       1.00      0.50      0.67         6
           1       0.80      1.00      0.89        12

    accuracy                           0.83        18
   macro avg       0.90      0.75      0.78        18
weighted avg       0.87      0.83      0.81        18

--------------------------------------------------

 SVM testing data classification report:
              precision    recall  f1-score   support

           0       1.00      0.50      0.67         6
           1       0.80      1.00      0.89        12

    accuracy                           0.83        18
   macro avg       0.90      0.75      0.78        18
weighted avg       0.87      0.83      0.81        18

--------------------------------------------------
```

```
 Decision Tree testing data classification report:
              precision    recall  f1-score   support

           0       1.00      0.50      0.67         6
           1       0.80      1.00      0.89        12

    accuracy                           0.83        18
   macro avg       0.90      0.75      0.78        18
weighted avg       0.87      0.83      0.81        18

--------------------------------------------------

 KNN testing data classification report:
              precision    recall  f1-score   support

           0       1.00      0.50      0.67         6
           1       0.80      1.00      0.89        12

    accuracy                           0.83        18
   macro avg       0.90      0.75      0.78        18
weighted avg       0.87      0.83      0.81        18

--------------------------------------------------
```

Thank you!