

● ● ●

CONTENT



- **Executive Summary**
- **Introduction**
- **Methodology**
- **Results**
 - Visualization – Graphs
 - Dashboard
- **Discussion**
 - Results and Implications
- **Conclusion**
- **Appendix**

EXECUTIVE SUMMARY



Challenge: Determine SpaceY's launch price competitively (cost US\$62M) based on the probability of **reusing** the rocket's first stage.

Solution: Implementation of a data science workflow, culminating in a **machine learning predictive model** to classify landing success (class = 1 for success).

Business Value: Landing predictability enables competitive pricing, generating **potential savings of over US\$100 million** per mission compared to non-reusable competitors.

Key Results: The logistic **regression model** demonstrated the best **accuracy of 83.33%** on the test set. Exploratory analysis identified **KSC LC-39A** as the **most reliable site** and the **B5 version of the booster** as the most **successful**.

Deliverables:

- Logistic and Geospatial Analysis (Folium) of Launch Sites.
- ML Predictive Model.
- Interactive Dashboard (Dash) for Payload Analysis.

INTRODUCTION



- **Launch costs** in the commercial space industry are the main **competitive differentiator**.
- **Reusability** of the rocket's first stage (Falcon 9) is the **only factor that reduces the cost** from US\$165M+ to US\$62M.
- **Business Problem:** SpaceY needs a reliable, data-driven way to **predict landing success**, allowing it to **optimize pricing** and plan booster recovery logistics.
- The project proposes the **application of data science and machine learning** to replace complex "rocket science" estimates with objective **predictions**.



METHODOLOGY



- **Data Collection** using **Web Scraping** and **SpaceX API**
- **Geospatial Analysis** using **Folium**
- **Exploratory Analysis (EDA)** using **SQL** and **Plotly**
- **Data Preparation (Wrangling)**
- **ML Modeling**

METHODOLOGY

Data Collection:

- Extraction of historical SpaceX launch data using **web scraping** (Wikipedia) and the SpaceX **API** (Jupyter Notebooks).

2020 [edit]

In late 2019, *SpaceNews* stated that SpaceX hoped for as many as 24 launches for Starlink satellites in 2020,^[100] in addition to 14 to 15 non-Starlink launches. At 26 launches, 12 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.^[101]

Flight No.	Date and time (UTC)	Version, Booster ^[1]	Launch site	Payload ¹	Payload mass	Orbit	Customer	Launch outcome	Booster landing
78	7 January 2020, 02:12:22 ^[102]	F9 B9 Δ, B1039.4	CCAFS, SLC-40	Starlink 2 v1.0 (80 satellites)	15,800 kg (34,400 lb) ^[1]	LEO	SpaceX	Success	Success (pne emp)
Third large batch and second operational flight of Starlink constellation. One of the 80 satellites included a test coating to make the satellite less reflective, and thus less likely to interfere with ground-based astronomical observations. ^[100]									
79	19 January 2020, 15:36 ^[103]	F9 B9 Δ, B1040.4	KSC, LC-39A	Crew Dragon in-flight abort test ^[104] (Dragon C208.1)	12,050 kg (26,570 lb)	Sub-orbital ^[105]	NASA (COTS) ^[106]	Success	No attempt
An atmospheric test of the Dragon-2 abort system after <i>Starliner</i> . 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 <i>Crew Dragon</i> Demo-1 capsule ^[104] but that test article exploded during a ground test of SuperDraco engines on 20 April 2019. ^[107] The abort test used the capsule originally intended for the first orbital flight ^[108] As expected, the booster was destroyed by aerodynamic forces after the capsule aborted ^[109] First flight 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, 14:22 ^[101]	F9 B9 Δ, B1051.3	CCAFS, SLC-40	Starlink 3 v1.0 (80 satellites)	15,800 kg (34,400 lb) ^[1]	LEO	SpaceX	Success	Success (pne emp)
Third operational and fourth large batch of Starlink satellites, deployed in a circular 290 km (180 mi) orbit. One of the testing halves was caught, while the other was fished out of the ocean. ^[100]									
81	10 February 2020, 15:09 ^[110]	F9 B9 Δ, B1056.4	CCAFS, SLC-40	Starlink 4 v1.0 (80 satellites)	15,800 kg (34,400 lb) ^[1]	LEO	SpaceX	Success	Failure (pne emp)
Fourth operational and fifth large batch of Starlink satellites. Used a new flight profile which deployed into a 212 km × 368 km (132 mi × 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 ^[111] , due to incorrect test data. ^[112] This was the first time a flight ground booster failed to land.									
82	7 March 2020, 04:26 ^[113]	F9 B9 Δ, B1059.2	CCAFS, SLC-40	SpaceX CRS-20 (Dragon C112.3-C)	1,877 kg (4,138 lb) ^[114]	LEO (ISS)	NASA (COTS)	Success	Success (ground test)
Last launch of phase 1 of the CRS contract. Carries <i>EarthScope</i> , an LISA payload for testing external payloads onto ISS. ^[115] Originally scheduled to launch on 7 March 2020, the launch date was pushed back due to a second stage engine failure. SpaceX decided to skip out the second stage instead of replacing the faulty part. ^[116] It was SpaceX's 50th successful payload onto ISS. ^[117] The first flight of the Dragon C112 and the last launch of the cargo Dragon spacecraft.									
83	18 March 2020, 12:14 ^[118]	F9 B9 Δ, B1065.5	KSC, LC-39A	Starlink 5 v1.0 (80 satellites)	15,800 kg (34,400 lb) ^[1]	LEO	SpaceX	Success	Failure (pne emp)
Fifth operational launch of Starlink satellites. It was the first time a first stage booster flew for a fifth time and the second time the firings were reused (Starlink flight in May 2019). ^[119] Towards the end of the first stage burn, the booster suffered premature shut down of an engine, the first of a Merlin 1D variant and first since the CRS-5 mission in October 2012. However, the payload did reach the targeted orbit. ^[120] This was the second Starlink launch booster landing failure in a row, later revealed to be caused by residual slugging fuel trapped inside a sensor. ^[121]									
84	22 April 2020, 19:36 ^[122]	F9 B9 Δ, B1081.4	KSC, LC-39A	Starlink 6 v1.0 (80 satellites)	15,800 kg (34,400 lb) ^[1]	LEO	SpaceX	Success	Success (pne emp)

Figure 1 - SpaceX Wikipedia Webpage

```
<table class="wikitable plainrowheaders collapsible" style="width: 100%;">
<tbody><tr>
<th scope="col">Flight No.
</th>
<th scope="col">Date and<br/>time (<a href="/wiki/Coordinated_Universal_Time" title="Coordinated Universal Time">UTC</a>)</th>
<th scope="col">Version, Booster<sup>[1]</sup>
</th>
<th scope="col">Launch site
</th>
<th scope="col">Payload<sup>1</sup>
</th>
<th scope="col">Payload mass
</th>
<th scope="col">Orbit
</th>
<th scope="col">Customer
</th>
<th scope="col">Launch outcome
</th>
<th scope="col">Booster landing
</th>
</tr>
<tr>
<td>78
</td>
<td>7 January 2020, 02:12:22<sup>[102]</sup>
</td>
<td>F9 B9 Δ, B1039.4
</td>
<td>CCAFS, SLC-40
</td>
<td>Starlink 2 v1.0 (80 satellites)
</td>
<td>15,800 kg (34,400 lb)<sup>[1]</sup>
</td>
<td>LEO
</td>
<td>SpaceX
</td>
<td>Success
</td>
<td>Success (pne emp)
</td>
</tr>
<tr>
<td data-cs="10" data-kind="parent">Third large batch and second operational flight of Starlink constellation. One of the 80 satellites included a test coating to make the satellite less reflective, and thus less likely to interfere with ground-based astronomical observations.<sup>[100]</sup>
</td>
<td data-kind="ghost"></td>
<td data-kind="ghost"></td>
<td data-kind="ghost"></td>
<td data-kind="ghost"></td>
<td data-kind="ghost"></td>
<td data-kind="ghost"></td>
<td data-kind="ghost"></td>
<td data-kind="ghost"></td>
<td data-kind="ghost"></td>
</tr>
<tr>
<td>79
</td>
<td>19 January 2020, 15:36<sup>[103]</sup>
</td>
<td>F9 B9 Δ, B1040.4
</td>
<td>KSC, LC-39A
</td>
<td>Crew Dragon in-flight abort test<sup>[104]</sup>
<br/>
(Dragon C208.1)
</td>
<td>12,050 kg (26,570 lb)
</td>
<td>Sub-orbital<sup>[105]</sup>
</td>
<td>NASA (COTS)<sup>[106]</sup>
</td>
<td>Success
</td>
<td>No attempt
</td>
</tr>
<tr>
<td data-cs="10" data-kind="parent">An atmospheric test of the Dragon-2 abort system after Starliner. 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<sup>[104]</sup> but that test article exploded during a ground test of SuperDraco engines on 20 April 2019.<sup>[107]</sup> The abort test used the capsule originally intended for the first orbital flight<sup>[108]</sup> As expected, the booster was destroyed by aerodynamic forces after the capsule aborted<sup>[109]</sup> First flight of a Falcon 9 with only one functional stage – the second stage had a mass simulator in place of its engine.
</td>
<td data-kind="ghost"></td>
<td data-kind="ghost"></td>
<td data-kind="ghost"></td>
<td data-kind="ghost"></td>
<td data-kind="ghost"></td>
<td data-kind="ghost"></td>
<td data-kind="ghost"></td>
<td data-kind="ghost"></td>
<td data-kind="ghost"></td>
</tr>
<tr>
<td>80
</td>
<td>29 January 2020, 14:22<sup>[101]</sup>
</td>
<td>F9 B9 Δ, B1051.3
</td>
<td>CCAFS, SLC-40
</td>
<td>Starlink 3 v1.0 (80 satellites)
</td>
<td>15,800 kg (34,400 lb)<sup>[1]</sup>
</td>
<td>LEO
</td>
<td>SpaceX
</td>
<td>Success
</td>
<td>Success (pne emp)
</td>
</tr>
<tr>
<td data-cs="10" data-kind="parent">Third operational and fourth large batch of Starlink satellites, deployed in a circular 290 km (180 mi) orbit. One of the testing halves was caught, while the other was fished out of the ocean.<sup>[100]</sup>
</td>
<td data-kind="ghost"></td>
<td data-kind="ghost"></td>
<td data-kind="ghost"></td>
<td data-kind="ghost"></td>
<td data-kind="ghost"></td>
<td data-kind="ghost"></td>
<td data-kind="ghost"></td>
<td data-kind="ghost"></td>
<td data-kind="ghost"></td>
</tr>
<tr>
<td>81
</td>
<td>10 February 2020, 15:09<sup>[110]</sup>
</td>
<td>F9 B9 Δ, B1056.4
</td>
<td>CCAFS, SLC-40
</td>
<td>Starlink 4 v1.0 (80 satellites)
</td>
<td>15,800 kg (34,400 lb)<sup>[1]</sup>
</td>
<td>LEO
</td>
<td>SpaceX
</td>
<td>Success
</td>
<td>Failure (pne emp)
</td>
</tr>
<tr>
<td data-cs="10" data-kind="parent">Fourth operational and fifth large batch of Starlink satellites. Used a new flight profile which deployed into a 212 km × 368 km (132 mi × 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<sup>[111]</sup>, due to incorrect test data.<sup>[112]</sup> This was the first time a flight ground booster failed to land.
</td>
<td data-kind="ghost"></td>
<td data-kind="ghost"></td>
<td data-kind="ghost"></td>
<td data-kind="ghost"></td>
<td data-kind="ghost"></td>
<td data-kind="ghost"></td>
<td data-kind="ghost"></td>
<td data-kind="ghost"></td>
<td data-kind="ghost"></td>
</tr>
<tr>
<td>82
</td>
<td>7 March 2020, 04:26<sup>[113]</sup>
</td>
<td>F9 B9 Δ, B1059.2
</td>
<td>CCAFS, SLC-40
</td>
<td>SpaceX CRS-20
<br/>
(Dragon C112.3-C)
</td>
<td>1,877 kg (4,138 lb)<sup>[114]</sup>
</td>
<td>LEO (ISS)
</td>
<td>NASA (COTS)
</td>
<td>Success
</td>
<td>Success (ground test)
</td>
</tr>
<tr>
<td data-cs="10" data-kind="parent">Last launch of phase 1 of the CRS contract. Carries EarthScope, an LISA payload for testing external payloads onto ISS.<sup>[115]</sup> Originally scheduled to launch on 7 March 2020, the launch date was pushed back due to a second stage engine failure. SpaceX decided to skip out the second stage instead of replacing the faulty part.<sup>[116]</sup> It was SpaceX's 50th successful payload onto ISS.<sup>[117]</sup> The first flight of the Dragon C112 and the last launch of the cargo Dragon spacecraft.
</td>
<td data-kind="ghost"></td>
<td data-kind="ghost"></td>
<td data-kind="ghost"></td>
<td data-kind="ghost"></td>
<td data-kind="ghost"></td>
<td data-kind="ghost"></td>
<td data-kind="ghost"></td>
<td data-kind="ghost"></td>
<td data-kind="ghost"></td>
</tr>
<tr>
<td>83
</td>
<td>18 March 2020, 12:14<sup>[118]</sup>
</td>
<td>F9 B9 Δ, B1065.5
</td>
<td>KSC, LC-39A
</td>
<td>Starlink 5 v1.0 (80 satellites)
</td>
<td>15,800 kg (34,400 lb)<sup>[1]</sup>
</td>
<td>LEO
</td>
<td>SpaceX
</td>
<td>Success
</td>
<td>Failure (pne emp)
</td>
</tr>
<tr>
<td data-cs="10" data-kind="parent">Fifth operational launch of Starlink satellites. It was the first time a first stage booster flew for a fifth time and the second time the firings were reused (Starlink flight in May 2019).<sup>[119]</sup> Towards the end of the first stage burn, the booster suffered premature shut down of an engine, the first of a Merlin 1D variant and first since the CRS-5 mission in October 2012. However, the payload did reach the targeted orbit.<sup>[120]</sup> This was the second Starlink launch booster landing failure in a row, later revealed to be caused by residual slugging fuel trapped inside a sensor.<sup>[121]</sup>
</td>
<td data-kind="ghost"></td>
<td data-kind="ghost"></td>
<td data-kind="ghost"></td>
<td data-kind="ghost"></td>
<td data-kind="ghost"></td>
<td data-kind="ghost"></td>
<td data-kind="ghost"></td>
<td data-kind="ghost"></td>
<td data-kind="ghost"></td>
</tr>
<tr>
<td>84
</td>
<td>22 April 2020, 19:36<sup>[122]</sup>
</td>
<td>F9 B9 Δ, B1081.4
</td>
<td>KSC, LC-39A
</td>
<td>Starlink 6 v1.0 (80 satellites)
</td>
<td>15,800 kg (34,400 lb)<sup>[1]</sup>
</td>
<td>LEO
</td>
<td>SpaceX
</td>
<td>Success
</td>
<td>Success (pne emp)
</td>
</tr>
</tbody></table>
<div>
<a href="/wiki/Low_Earth_orbit" title="Low Earth orbit">LEO</a></div>
```

Figure 2 - SpaceX API Web Scraping Output

METHODOLOGY

Geospatial Analysis:

- Using **Folium** to map sites and analyze proximity to cities, coastlines, railways, and highways.

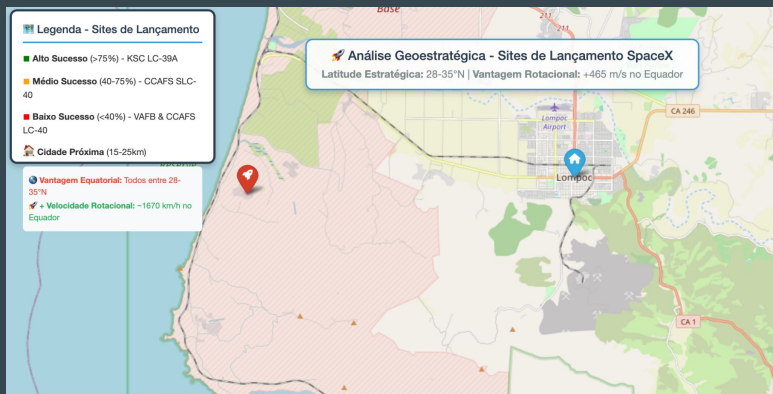


Figure 3 - Visualization of SpaceX Launch Sites using Folium

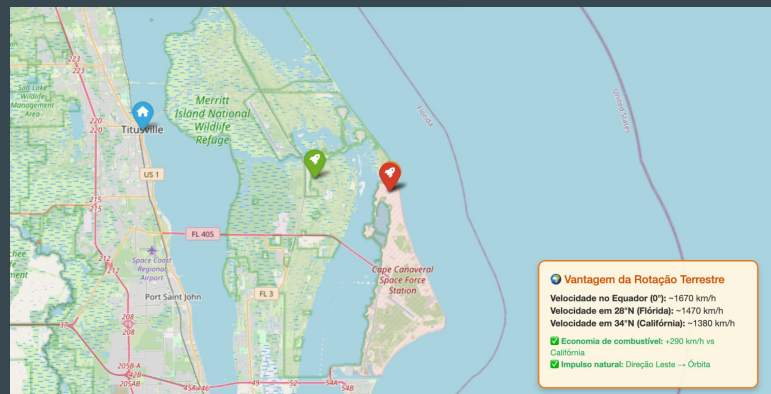


Figure 4 - Geostrategic visualization of mapping and analysis of surroundings

METHODOLOGY

Exploratory Analysis (EDA):

- Using **SQL** and **Plotly** to analyze the impact of variables such as Payload, Launch Location, and Booster Version.

```
[12]: %%sql
SELECT *
FROM SPACEXTABLE
WHERE "Launch_Site" LIKE 'CCA%'
LIMIT 5;

* 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	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

Figure 5 - Visualization of exploratory analysis using SQL and Plotly

METHODOLOGY

Data Wrangling:

Cleaning, missing value processing, and applying **One-Hot Encoding** to convert categorical variables (such as Orbit and Launch Location) into numerical features for the model.

Use the function `get_dummies` and `features` dataframe to apply OneHotEncoder to the column `Orbits`, `LaunchSite`, `LandingPad`, and `Serial`. Assign the value to the variable `features_one_hot`, display the results using the method `head`. Your result dataframe must include all features including the encoded ones.

```
# HINT: Use get_dummies() function on the categorical columns
```

```
features_one_hot = pd.get_dummies(features, columns=['Orbit', 'LaunchSite', 'LandingPad', 'Serial'])
print(f"Shape after one-hot encoding: {features_one_hot.shape}")
features_one_hot.head()
```

Python

Shape after one-hot encoding: (90, 80)

	FlightNumber	PayloadMass	Flights	GridFins	Reused	Legs	Block	ReusedCount	Orbit_ES-L1	Orbit_GE
0	1	6104.959412	1	False	False	False	1.0	0	False	Fal
1	2	525.000000	1	False	False	False	1.0	0	False	Fal
2	3	677.000000	1	False	False	False	1.0	0	False	Fal
3	4	500.000000	1	False	False	False	1.0	0	False	Fal
4	5	3170.000000	1	False	False	False	1.0	0	False	Fal

Figure 6 - Applying One-Hot Encoding to categorical variables

METHODOLOGY

Machine Learning Modeling:

Application and comparison of four supervised classification algorithms:

- Logistic Regression (LogReg),
- Support Vector Machine (SVM),
- Decision Tree, and
- K-Nearest Neighbors (KNN).

```
[30]: # Comparar accuracy de todos os modelos
models = {
    'Logistic Regression': accuracy_logreg,
    'SVM': accuracy_svm,
    'Decision Tree': accuracy_tree,
    'KNN': accuracy_knn
}

best_model = max(models, key=models.get)
best_accuracy = models[best_model]

print("=== COMPARAÇÃO DOS MODELOS ===")
for model, acc in models.items():
    print(f"{model}: {acc:.4f}")

print(f"\n MELHOR MODELO: {best_model} com accuracy de {best_accuracy:.4f}")

=== COMPARAÇÃO DOS MODELOS ===
Logistic Regression: 0.8333
SVM: 0.8333
Decision Tree: 0.8333
KNN: 0.8333

MELHOR MODELO: Logistic Regression com accuracy de 0.8333
```

Figure 7 - Comparison of the four classification models

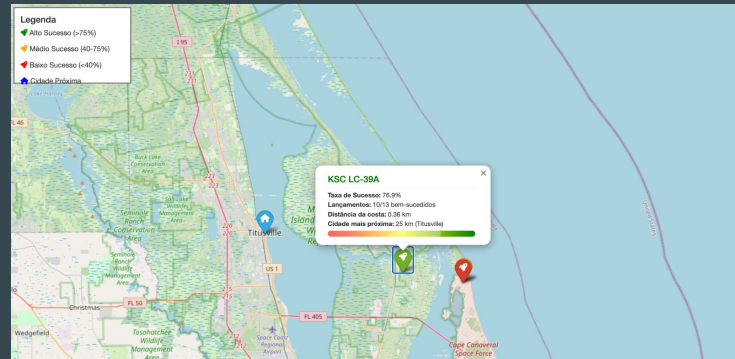
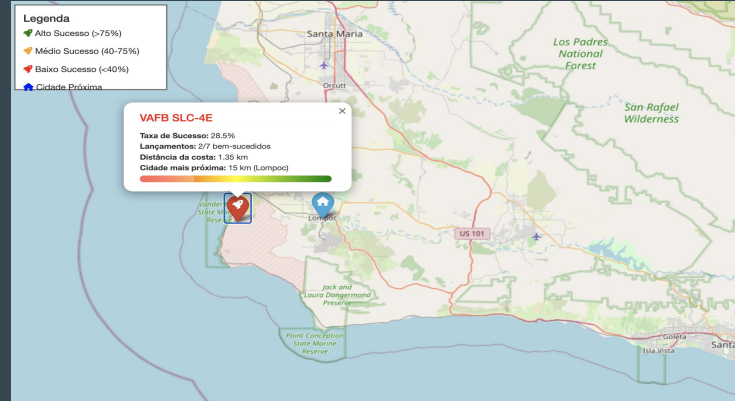
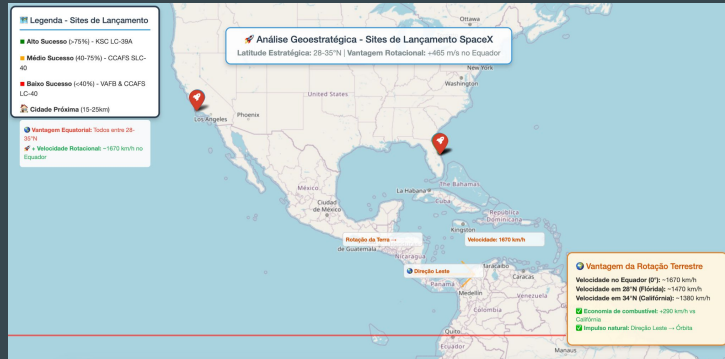
RESULTS

Geospatial Analysis and Logistics

Strategic Location (Latitude)

All four sites (CCAFS LC-40/SLC-40, KSC LC-39A, VAFB SLC-4E) are between 28° and 35° North Latitude.

Implication: This equatorial **proximity** (<35°) maximizes Earth's rotational momentum, **saving fuel** and **reducing** operating **costs**.



RESULTS

Geospatial Analysis and Logistics

Infrastructure Safety and Logistics

Coastal: All sites are directly on the coast (average distance < **1.35 km**).

Implication: Ensures **maximum safety** (debris falls into the ocean) and allows maritime logistical access.

Highways/Railways: All are within 1.2 km of highways and railways.

Implication: Efficient logistics for transporting heavy components (rockets) and emergency evacuation.

Urban Areas: Maintains a safe distance (approximately 15 km to 25 km) from the nearest cities.

Implication: Minimizes public **safety risks** and **reduces noise impacts**.

RESULTS

Performance and Reliability (EDA)

Performance by Launch Site

The KSC LC-39A site is the most reliable, with the **highest** landing **success rate**:

- 76.9% (10 successful launches).

Implication: SpaceX should prioritize high-value missions at this site to **maximize** the likelihood of **reusability** and **financial return**.

Payload Influence

The **highest success rate** is observed in the 2501 kg - 5000 kg payload range

- (approximately **55.0% success** rate).

Implication: SpaceX should focus on **optimizing** boosters and mission profiles in this payload range, as it represents the "sweet spot" for **success/reusability**.

RESULTS

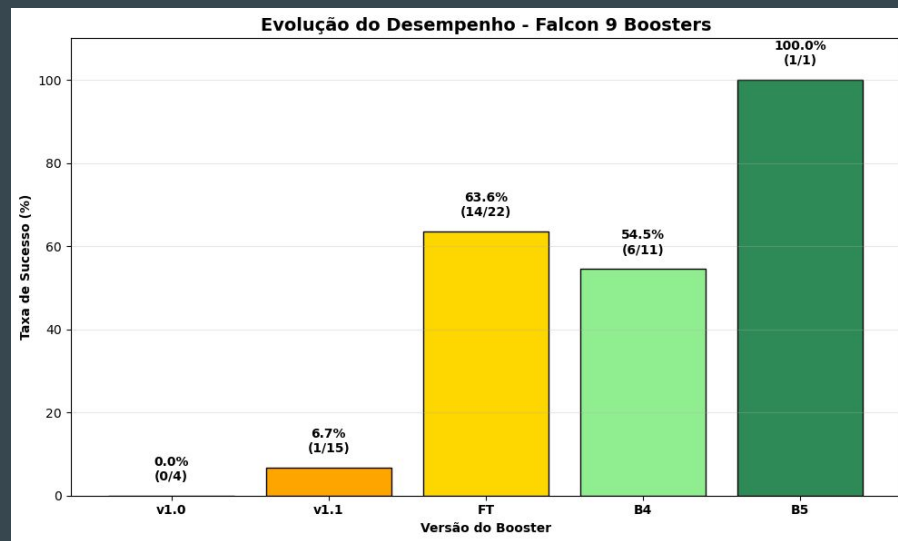
Performance and Reliability (EDA)

Learning Curve and Booster Version

The **B5 version** of the Falcon 9 Booster (the latest) demonstrated a 100.0% (1/1) **success rate**.

Newer versions (FT, B4, B5) have **significantly higher rates** than older versions (v1.0, v1.1).

Implication: Landing success is a direct function of technological evolution. SpaceX must **continually invest** in the latest booster version to ensure the viability of its low-cost **business model**.



Graph 1 - Performance Evolution (Boosters)

VISUALIZATION – GRAPHICS

Multi-Criteria Analysis Radar Chart

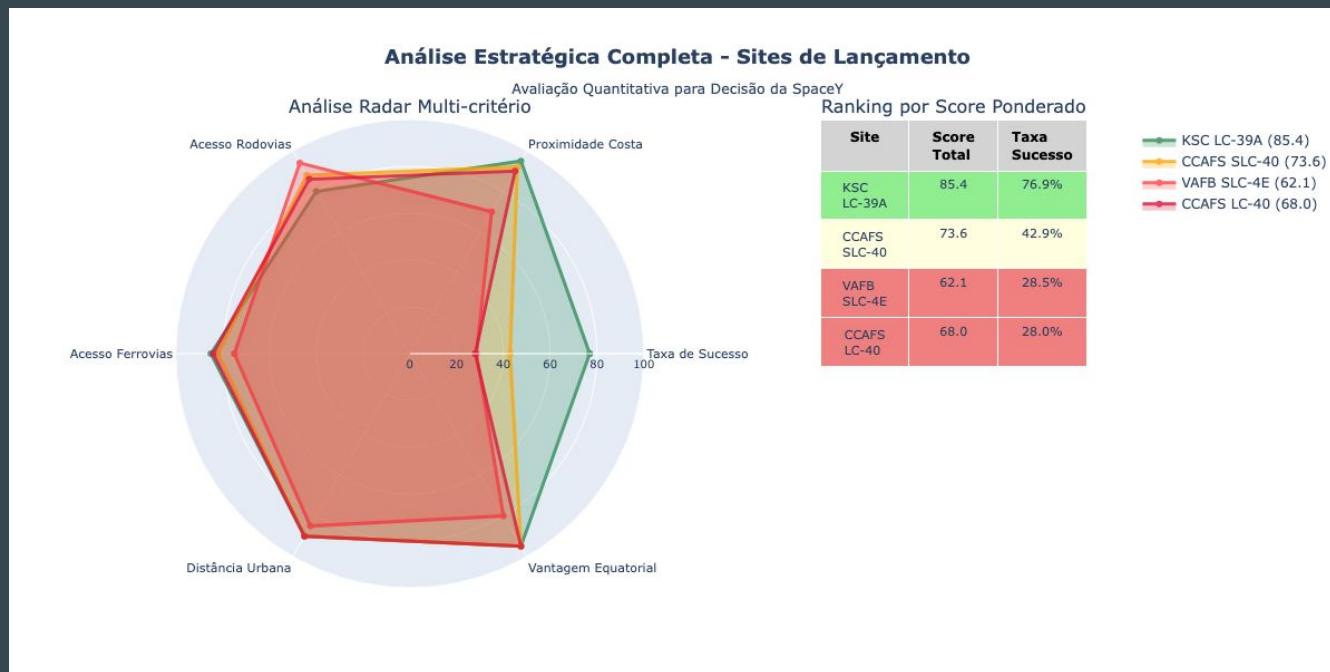


Chart 2 - Radar chart of strategic analysis of launch sites

VISUALIZATION – GRAPHICS

Success Rate Bar Chart (%) by Launch Location

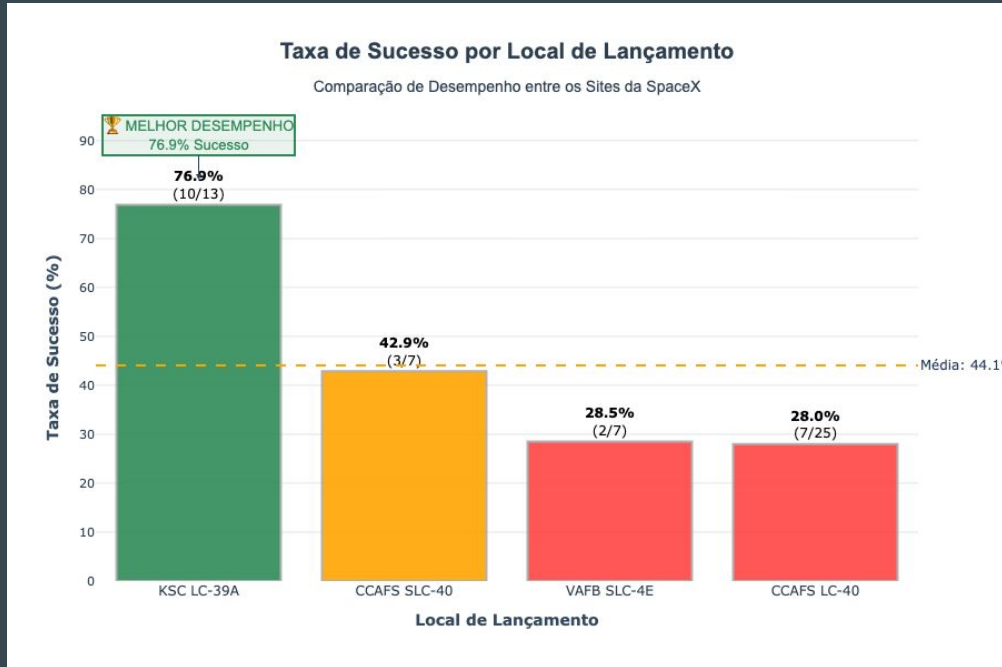


Chart 3 - Success rate chart by launch location

VISUALIZATION – GRAPHICS

Success Rate (%) Bar Chart by Payload Range

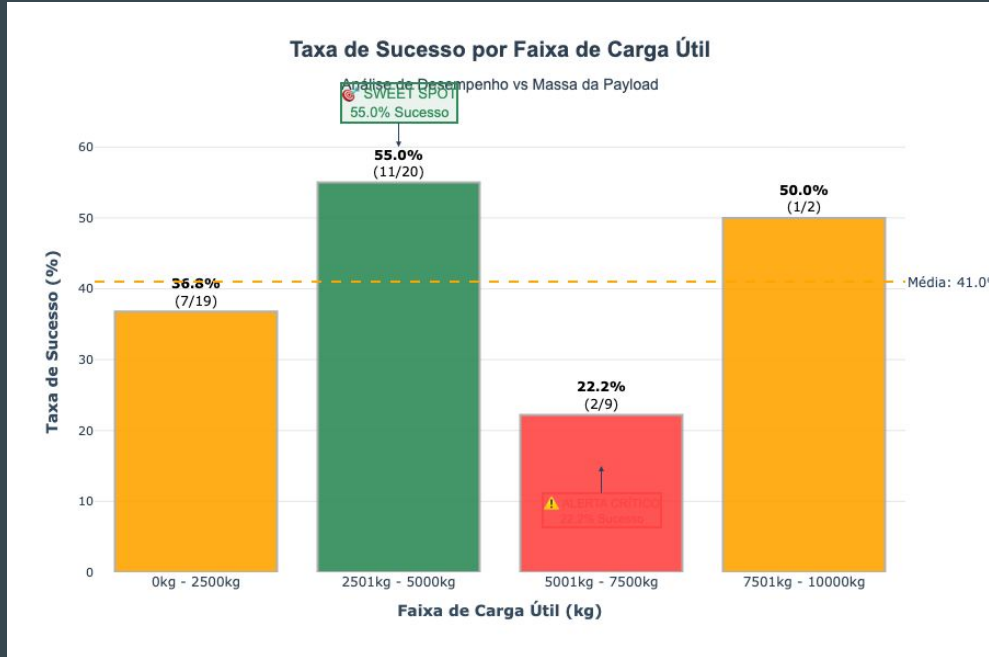


Chart 4 - Success rate chart by payload range

VISUALIZATION – GRAPHICS

Line Graph of Annual Success Rate over Time

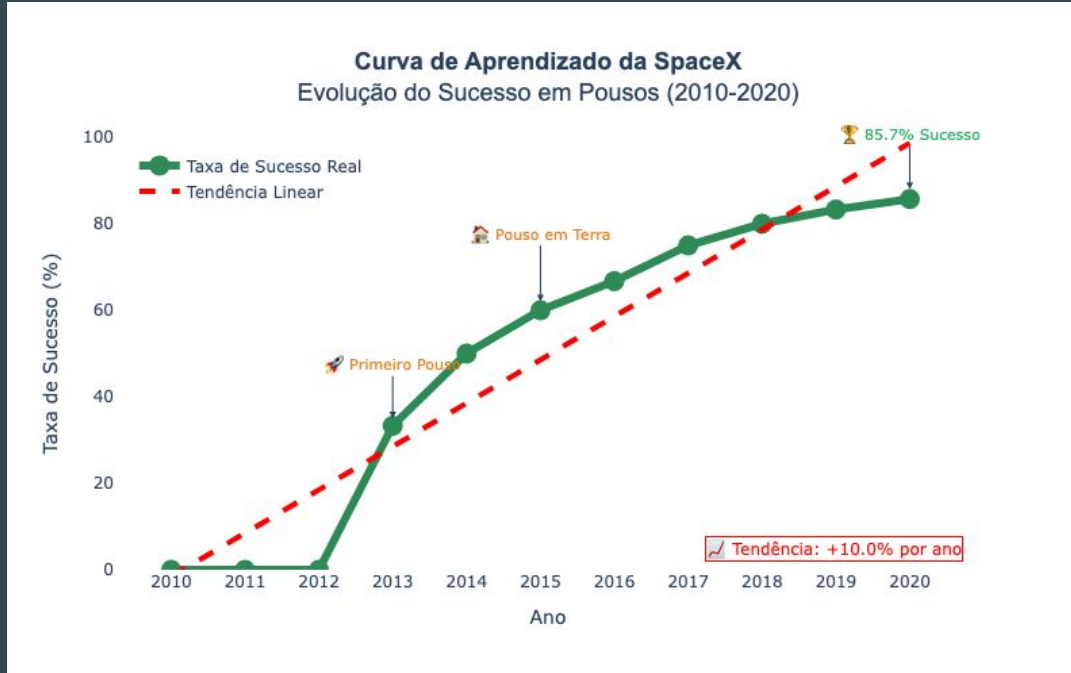


Chart 5 - Annual Success Rate Line Chart

DASHBOARD

Description:

- The Interactive Dashboard (created with Plotly/Dash) allows dynamic data exploration for real-time decision-making.

Features:

- **Filter by Launch Site:** Analyze landing success (class) specific to each of the 4 sites.
- **Payload Slider:** Visualize the correlation between Payload Mass (kg) and Launch Result (Success/Failure).
- **Scatter Plot:** Analyzes the impact of Booster Version (color) on success, visually confirming the superiority of newer versions (such as B5).

DASHBOARD - OVERVIEW

Dashboard Overview - The Interactive Dashboard (created with Plotly/Dash) allows dynamic data exploration for real-time decision-making.

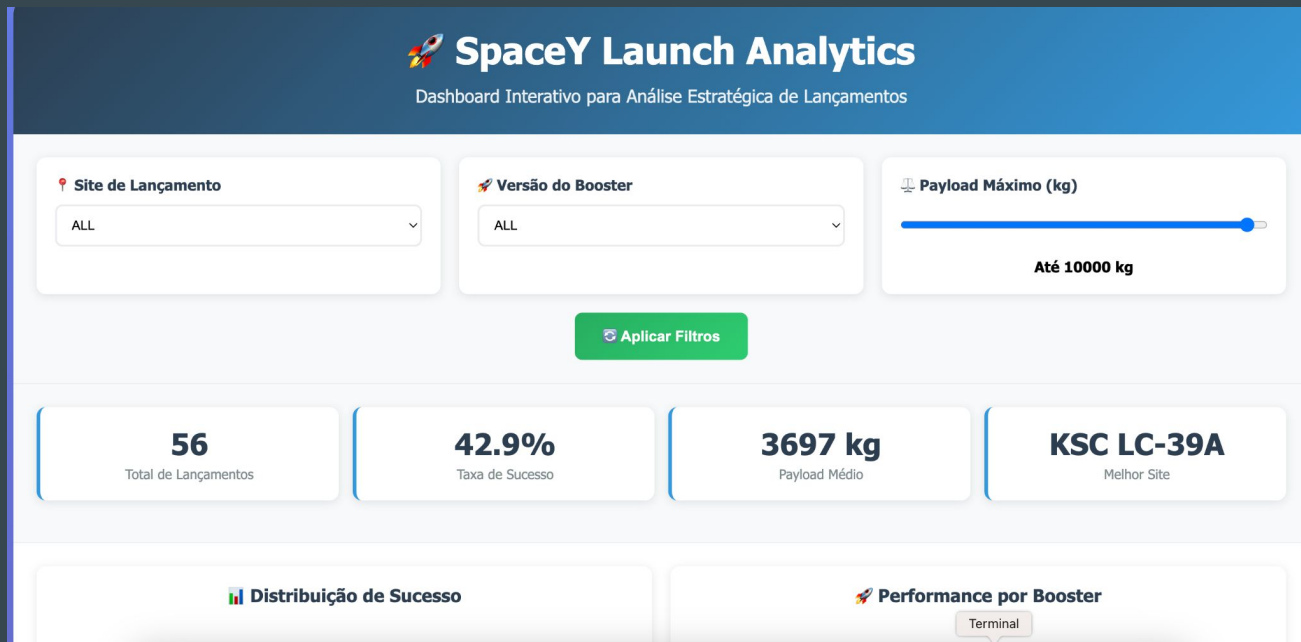


Figure 7 - Overview of the interactive dashboard for strategic launch analysis

DASHBOARD - FILTERS

Filters that allow you to select the Launch Site, Booster Version, and Maximum Payload (kg).

The dashboard features three filter sections at the top: 'Site de Lançamento' with a dropdown menu showing 'ALL' and four launch sites (CCAFS LC-40, CCAFS SLC-40, KSC LC-39A, VAFB SLC-4E); 'Versão do Booster' with a dropdown menu showing 'ALL'; and 'Payload Máximo (kg)' with a slider set to 'Até 10000 kg'. A green 'Aplicar Filtros' button is centered below the filters. Below the filters are four summary cards: '56 Total de Lançamentos', '42.9% Taxa de Sucesso', '3697 kg Payload Médio', and 'KSC LC-39A Melhor Site'.

Filter	Value
Site de Lançamento	ALL
Versão do Booster	ALL
Payload Máximo (kg)	Até 10000 kg

Metric	Value
Total de Lançamentos	56
Taxa de Sucesso	42.9%
Payload Médio	3697 kg
Melhor Site	KSC LC-39A

Figure 8 - Viewing filters on the Dashboard

DASHBOARD - CARDS

Cards that allow you to view the total number of launches, success rate, average payload, and the best site among the options being filtered.

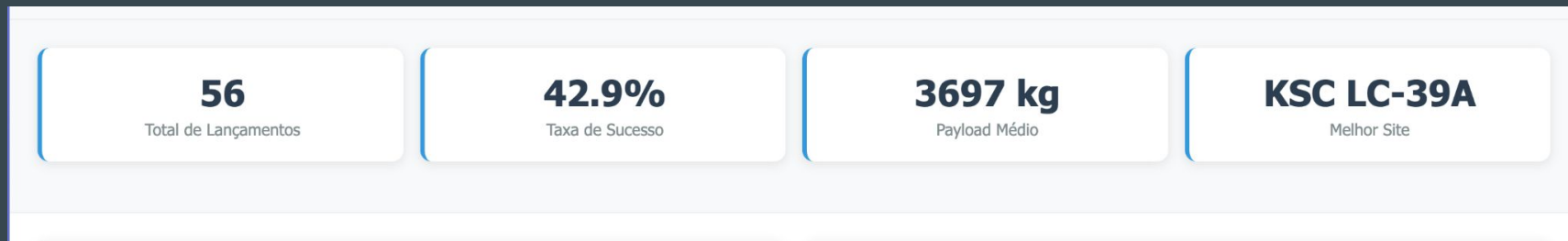


Figure 9 - Displaying cards on the Dashboard

DASHBOARD - GRAPHICS

Pie chart to visualize the Success Distribution (by site) and bar chart for Performance by Booster.

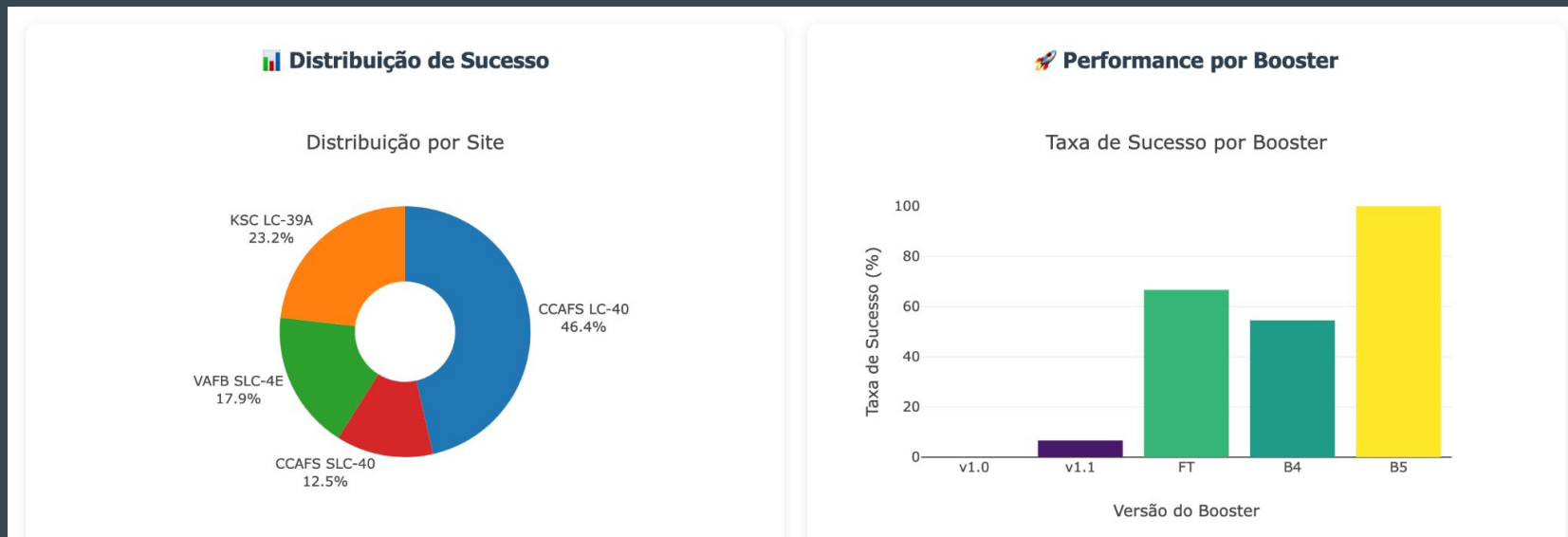


Figure 10 - Viewing graphs (pie and bars) on the Dashboard

DASHBOARD - GRAPHICS

Scatterplot to visualize Payload vs. Launch Success.



Figure 11 - Scatterplot View on the Dashboard

DISCUSSIONS



- **Machine Learning**
- **Best Model Chosen**
- **Confusion Matrix**
- **Results and Implications**

DISCUSSIONS

Machine Learning: Models, Accuracy and Performance

Modelo	Acurácia (Train Set)	Acurácia (Test Set)	Desempenho (Métricas)
Regressão Logística	$\approx 84.64\%$	83.33%	Melhor <i>performance</i> consistente.
SVM	$\approx 84.82\%$	83.33%	Igual ao LogReg no Teste.
Árvore de Decisão	$\approx 87.67\%$	83.33%	Overfitting leve (Alta acurácia no Treino).
KNN	$\approx 84.82\%$	83.33%	Igual ao LogReg no Teste.

Table 1 - Comparison of ML models

DISCUSSIONS

Best Model Chosen:

Logistic Regression (LogReg), due to its **simplicity**, **interpretability**, and **83.33% accuracy** in the test set, equaling more complex models.

```
[15] # Accuracy no conjunto de teste
      accuracy_logreg = logreg_cv.score(X_test, Y_test)
      print(f"Accuracy no teste (Regressão Logística): {accuracy_logreg}")

...  Accuracy no teste (Regressão Logística): 0.8333333333333334
```

Python

Figure 12 - Logistic Regression model accuracy output

DISCUSSIONS

Confusion Matrix (Common Test Outcome):

- **True Positive** (Success Correctly Predicted): 12
- **False Positive** (Failure Predicted as Success): 3
- **Interpretation:** The model predicted 3 failures as successes. For SpaceX, this represents a 16.67% risk of overestimating the success rate and potentially mispricing (\$62M).

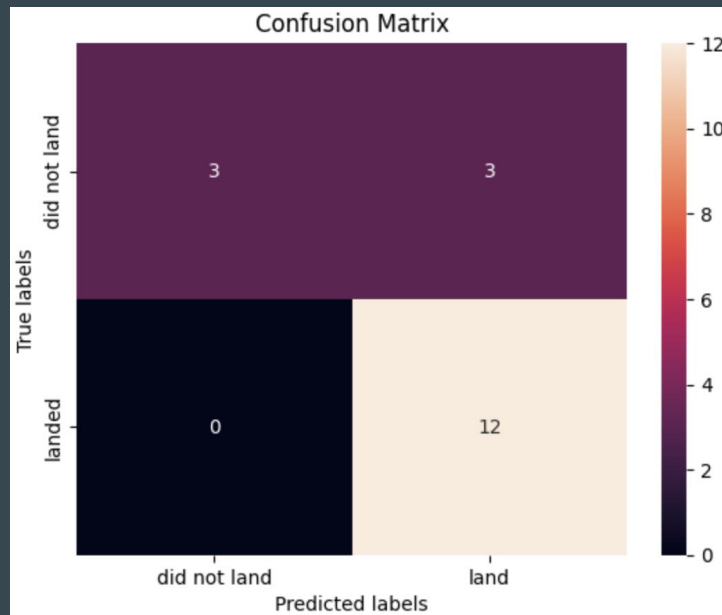


Figure 13 - Confusion matrix of the logistic regression model

RESULTS AND IMPLICATIONS

Success Prediction

Results:

Landing **success** is a highly predictable classification problem, achieving **83.33% accuracy**.

Implications:

SpaceY can confidently price (\$62M) when launch variables (especially **Booster Version** and **Location**) indicate a **high probability of success** (>90% in the model).

RESULTS AND IMPLICATIONS

Logistics Optimization

Results

Geospatial analysis defines the "**ideal spaceport**": **Near** the **equator**, on the **coast**, with immediate **access** to **railways/highways**, but far from **cities**.

Implications

Any future expansion or new SpaceX launch site must strictly adhere to **these 5 geospatial guidelines** to ensure safety and maximum logistical **efficiency while keeping costs low**.

RESULTS AND IMPLICATIONS

Product Focus

Results

Booster performance (Version B5 at 100%) is the **strongest predictor of success**.

Implications

R&D **investment** for the **latest booster versions** is the most important cost driver for SpaceX. The \$62M **business model depends on maintaining** the booster's technological superiority.

CONCLUSIONS



Project:

- The project successfully established a data-driven decision-making framework for SpaceY, replacing high-risk estimates with machine learning predictions.

Model:

- The Logistic Regression model (83.33% accuracy) provides a reliable indicator of the likelihood of reuse, which is a critical factor in pricing.

CONCLUSIONS



SpaceX Competitive:

- With the **data intelligence** provided, SpaceX is **ready to enter the market** and offer a **competitive price** of US\$62 million for missions that meet the high **probability of success** criteria.

Next Steps:

- **Implementation** of the **model** in a production environment for real-time quotes and **continuous training with new data** to **refine accuracy**, aiming to exceed **90% confidence**.

APPENDIX



Tables:

Table 2: Launch Count by Site and Success

Table 3: Detailed Geospatial Distances (Coastline, Railway, Highway, City)

Table 4: Comparative Ranking of Geospatial Distances

Graphs:

Graph 6: Confusion Matrix of the Best Model (Logistic Regression) in the Test Set

Graph 7: Logistic Regression Model Metrics

Graph 8: Scatterplot - Payload Mass vs. Success by Launch Site

Graph 9: Box Plot - Payload Distribution (Site vs. Result)

Graph 10: Violin Plot - Detailed Payload Distribution (Site vs. Result)

Graph 11: Density Heatmap - Payload vs. Success

APPENDIX



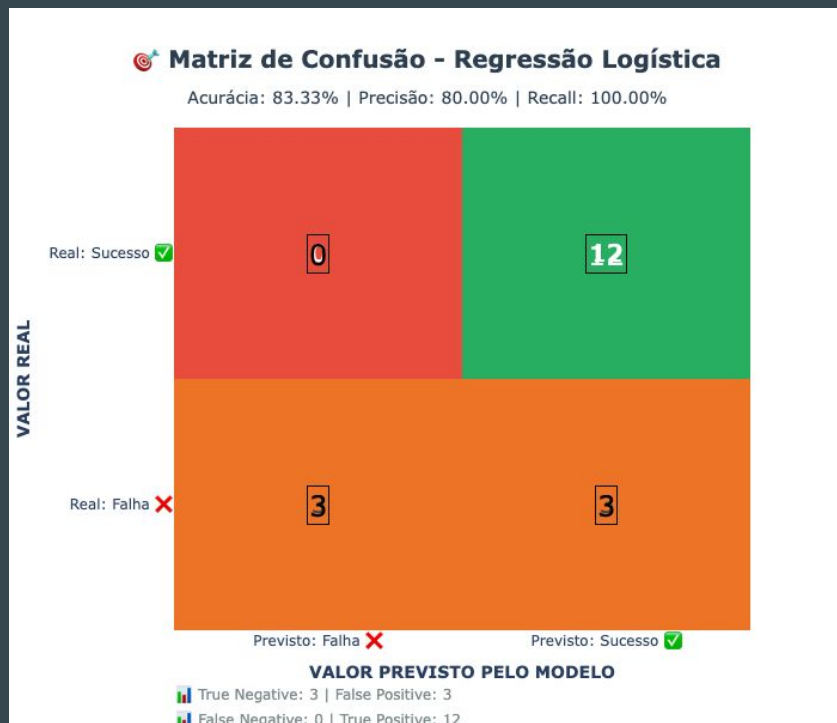
Tabela de Desempenho por Site de Lançamento

Contagem de Lançamentos e Taxas de Sucesso

Site de Lançamento	Lançamentos	Sucessos	Taxa de Sucesso	Performance
KSC LC-39A	13	10/13	76.9%	76.9
CCAFS SLC-40	7	3/7	42.9%	42.9
VAFB SLC-4E	10	4/10	40.0%	40
CCAFS LC-40	26	7/26	26.9%	26.9

Table 2: Launch Count by Site and Success Table

APPENDIX



Graph 6: Confusion Matrix of the best model (Logistic Regression) in the test set

APPENDIX

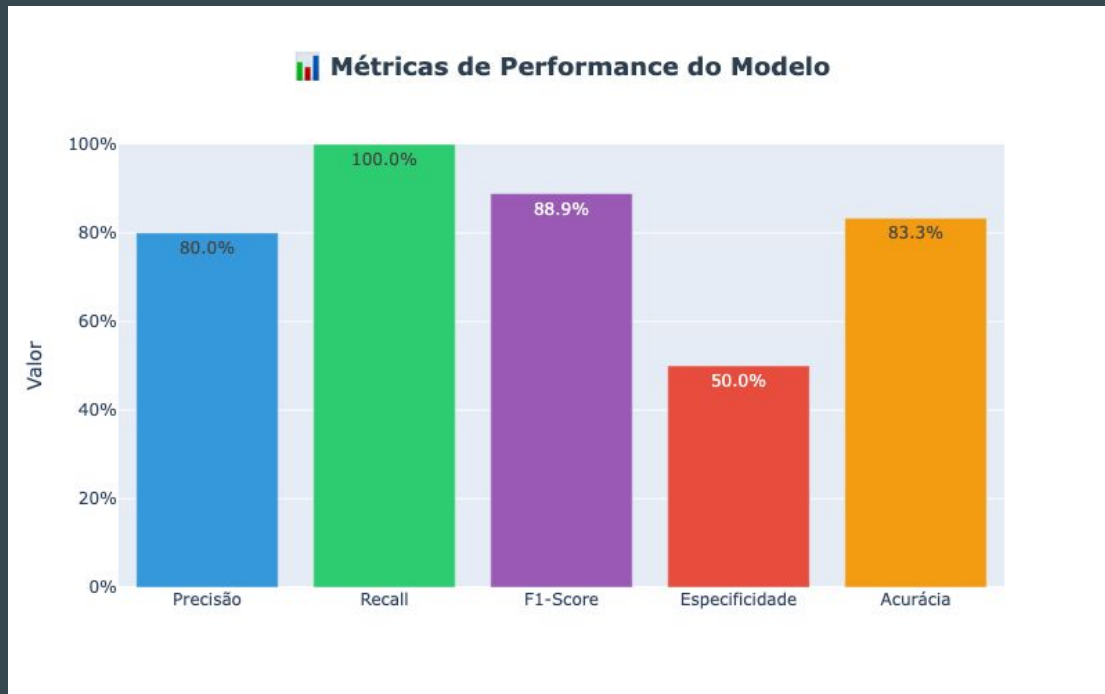


Chart 7: Logistic Regression Model Metrics

CALCULATED METRICS:

- Accuracy: 83.3% (15/18)
- Precision: 80.0% (12/15)
- Recall: 100.0% (12/12)
- F1-Score: 88.9%
- Specificity: 50.0% (3/6)

INTERPRETAÇÃO:

- Modelo acertou 15 de 18 previsões
- Identificou TODOS os 12 casos de sucesso
- Errou 3 previsões (todas falsos positivos)
- Excelente para não perder oportunidades

APPENDIX

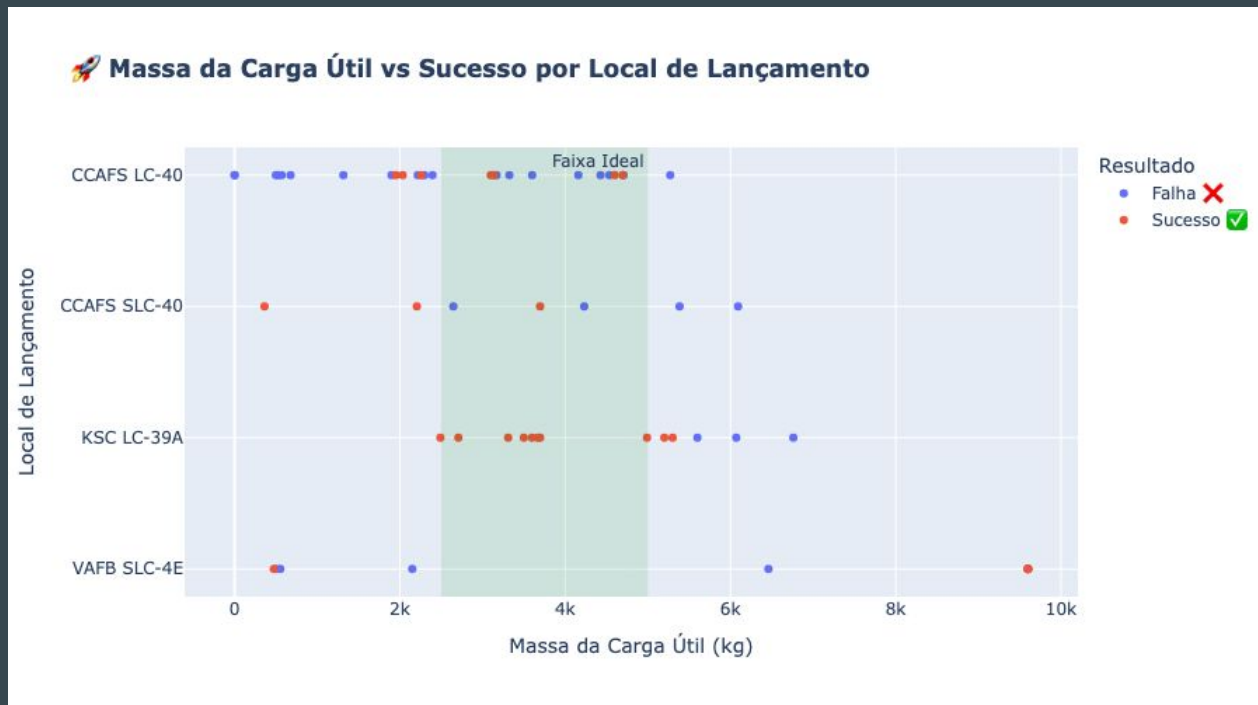


Chart 8: Scatterplot - Payload Mass vs. Success by Launch Site

APPENDIX



Chart 9: Box Plot - Payload Distribution (Location vs. Result)

APPENDIX

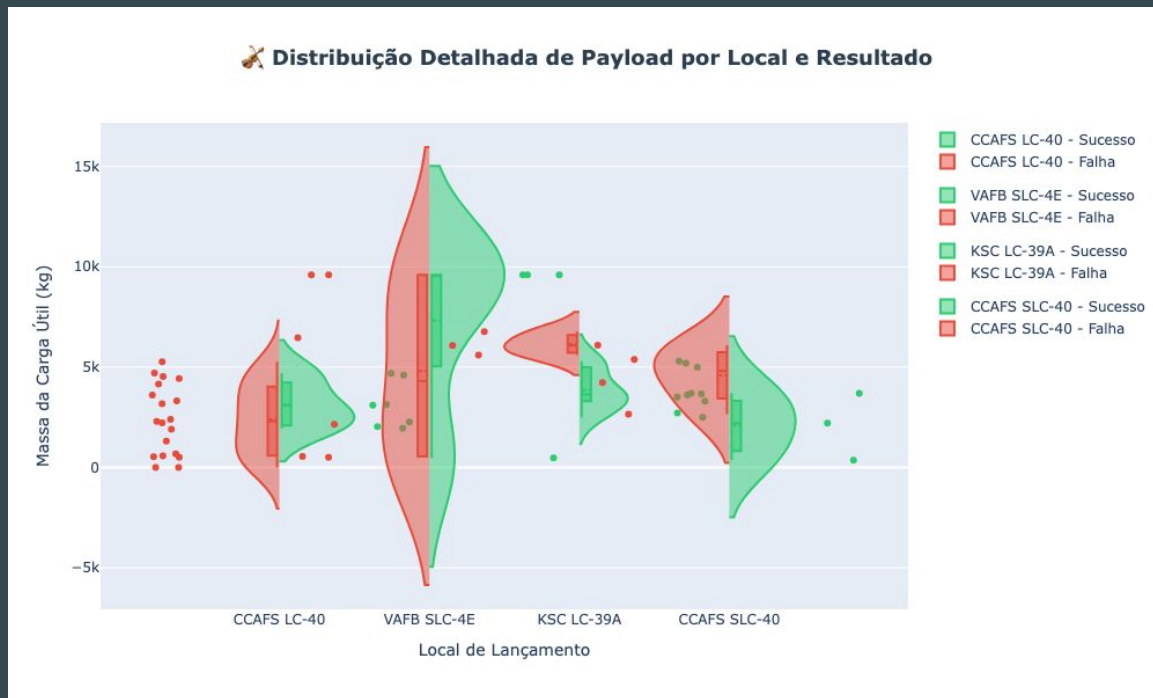


Chart 10: Violin Chart - Detailed Payload Distribution (Location vs. Result)

APPENDIX

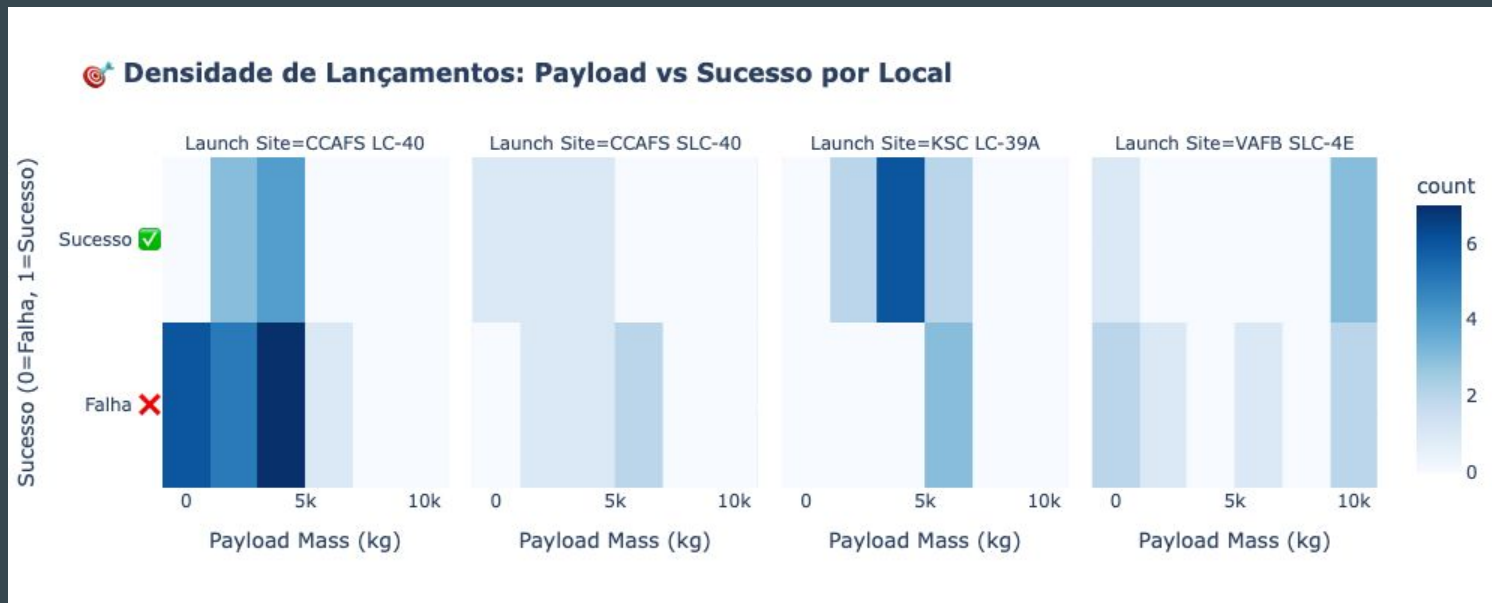


Chart 11: Density Heatmap - Payload vs. Success

APPENDIX

Distâncias Geoespaciais dos Sites de Lançamento







 SITE	 COSTA	 FERROVIA	 RODOVIA	 CIDADE	 CIDADE PRÓXIMA
KSC LC-39A	0.36 km	0.70 km	0.60 km	25.0 km	Titusville, FL
CCAFS SLC-40	0.51 km	0.90 km	0.40 km	25.0 km	Titusville, FL
VAFB SLC-4E	1.35 km	1.20 km	0.30 km	15.0 km	Lompoc, CA
CCAFS LC-40	0.58 km	0.80 km	0.50 km	25.0 km	Titusville, FL

Table 3: Geospatial distance details (Coastline, Railway, Highway, City)

APPENDIX

Análise Comparativa - Rankings de Acessibilidade

Local	 Costa	 Ferrovia	 Rodovia	 Cidade	 Score Total
KSC LC-39A	0.36 km #1	0.70 km #1	0.60 km #4	25.0 km #3	🌟 7
CCAFS SLC-40	0.51 km #2	0.90 km #3	0.40 km #2	25.0 km #3	🌟 6
VAFB SLC-4E	1.35 km #4	1.20 km #4	0.30 km #1	15.0 km #1	🌟 6
CCAFS LC-40	0.58 km #3	0.80 km #2	0.50 km #3	25.0 km #3	🌟 5

Table 4: Comparative ranking of geospatial distances