



IBM Developer  
SKILLS NETWORK

# Winning Space Race with Data Science

Aline Castro  
Fevereiro de 2024



# Outline

---

- **Executive Summary**

Resumo das principais descobertas e resultados.

- **Introduction**

Introdução ao tópico e contexto.

- **Methodology**

Descrição dos métodos e abordagens utilizados.

- **Results**

Apresentação dos resultados obtidos.

- **Conclusion**

Conclusões tiradas com base nos resultados.

- **Appendix**

Informações adicionais e detalhes técnicos.

# Executive Summary

---

## **Resumo Executivo**

### **Resumo das metodologias**

- Coleta de dados e organização de dados;
- Análise exploratória de dados;
- Análise visual interativa;
- Previsão com aprendizado de máquina.

### **Resumo de todos os resultados**

- Um pipeline de aprendizado de máquina foi construído para prever se a primeira etapa do Falcon 9 pousa com sucesso.

# Introduction

---

- **Objetivo da SpaceX:**

Enviar espaçonaves para a Estação Espacial Internacional.

Starlink, uma constelação de internet via satélite que fornece acesso à internet.

Enviar missões tripuladas para o espaço.

- **Objeto: Lançamento do Falcon 9 da SpaceX**

- **Meta: Determinar o preço de cada lançamento.**

- **Razão: Se pudermos determinar se a primeira etapa pousará, podemos determinar o custo de um lançamento**

- **Método: Coletando informações sobre a SpaceX e criando painéis:**

Para determinar se a SpaceX reutilizará a primeira etapa.

Treinar um modelo de aprendizado de máquina e usar informações públicas para prever se a SpaceX reutilizará a primeira etapa.



Section 1

# Methodology

# Metodologia

---

- **Resumo Executivo**

Resumo das principais descobertas e resultados.

- **Metodologia de coleta de dados:**

Os dados foram coletados de uma API e página Wiki através do método de web scraping.

- **Realizar a preparação de dados:**

Os resultados foram convertidos em Rótulos de Treinamento onde 1 significa que o propulsor pousou com sucesso e 0 significa que foi malsucedido.

- **Realizar análise de dados exploratória (EDA) usando visualização e SQL.**

- **Realizar análise visual interativa usando Folium e Plotly Dash.**

# Metodologia

---

## **Realizar análise preditiva usando modelos de classificação:**

- Pré-processamento (nos permitiu padronizar nossos dados);
- Train\_test\_split (nos permitiu dividir nossos dados em dados de treinamento e teste);
- Treinamento do modelo;
- Realização da Busca em Grade (nos permitiu encontrar os hiperparâmetros que permitem que um determinado algoritmo tenha o melhor desempenho);
- Determinação do modelo com a melhor precisão (Usando os melhores valores de hiperparâmetro e dados de treinamento);
- Teste de Regressão Logística, Máquinas de Vetores de Suporte, Classificador de Árvore de Decisão e K-vizinhos mais próximos.
- Saída da matriz de confusão.

# Data Collection

---

## 1. API

- URL da API: <https://api.spacexdata.com/v4/launches/past>
- Requisição de dados da API da SpaceX
- Limpeza dos dados

## 2. Página Web

- URL da página web: [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)
- Extração de registros de lançamento do Falcon 9 de uma tabela HTML da Wikipedia
- Análise da tabela e conversão em um quadro de dados do Pandas



# Data Collection – SpaceX API

---

- **Data Collection:**

1. Solicitar dados da API da SpaceX
2. Converter o resultado json em um dataframe
3. Filtrar o dataframe para apenas os lançamentos do Falcon 9 e faça a manipulação dos dados
4. Exportar para csv

- **Link:** <https://github.com/Aline-Castro/IBM-Data-Science-Professional-Certificate/blob/main/10%20-%20Applied%20Data%20Science%20Capstone%20/01%20-%20Introduction/01-Spacex-data-collection-api.ipynb>

```
spacex_url="https://api.spacexdata.com/v4/launches/past"
```

```
response = requests.get(spacex_url)
```

# Data Collection - Scraping

---

- **Web scraping process:**

1. Solicite a página Falcon9 Launch Wiki a partir de sua URL

2. Extraia todos os nomes de colunas/variáveis do cabeçalho da tabela HTML

```
static_url = "https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922"
```

3. Crie um dataframe analisando as tabelas de lançamento HTML

```
data = requests.get(static_url).text  
data
```

4. Exporte para csv

- **Link:** <https://github.com/Aline-Castro/IBM-Data-Science-Professional-Certificate/blob/main/10%20-%20Applied%20Data%20Science%20Capstone%2001%20-%20Introduction/02-jupyter-labs-webscraping.ipynb>

# Data Wrangling

---

- **Data Wrangling Process:**

1. Especifique o valor ausente
2. Calcule o número de lançamentos em cada site
3. Calcule o número e a ocorrência de cada órbita
4. Crie um rótulo de resultado de pouso a partir da coluna Outcome

- **Link:** <https://github.com/Aline-Castro/IBM-Data-Science-Professional-Certificate/blob/main/10%20-%20Applied%20Data%20Science%20Capstone%20/01%20-%20Introduction/03-spacex-Data%20wrangling.ipynb>

```
bad_outcomes=set(landing_outcomes.keys()[[1,3,5,6,7]])  
bad_outcomes
```

```
# landing_class = 0 if bad_outcome  
# landing_class = 1 otherwise  
  
df['Class'] = df['Outcome'].apply(lambda x: 0 if x in bad_outcomes else 1)  
df['Class'].value_counts()
```

```
df['Class']=landing_class  
df[['Class']].head(8)
```

# EDA with Data Visualization

---

1. Especifique o valor ausente
2. Calcule o número de lançamentos em cada local
3. Calcule o número e a ocorrência de cada tipo de órbita
4. Crie um rótulo de resultado de pouso a partir da coluna de resultado
5. A relação entre o número do voo e o local de lançamento -> gráfico de dispersão
6. A relação entre a carga útil e o local de lançamento -> gráfico de dispersão
7. A taxa de sucesso de cada tipo de órbita -> gráfico de barras
8. A relação entre o número do voo e o tipo de órbita -> gráfico de dispersão
9. A relação entre a carga útil e o tipo de órbita -> gráfico de dispersão
10. A tendência anual de sucesso de lançamento -> gráfico de linha

[https://github.com/Aline-Castro/IBM-Data-Science-Professional-Certificate/blob/main/10%20-%20Applied%20Data%20Science%20Capstone%20/02%20-%20Exploratory%20Data%20Analysis%20\(EDA\)/Parte%202/02%20-%20module\\_2\\_jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb](https://github.com/Aline-Castro/IBM-Data-Science-Professional-Certificate/blob/main/10%20-%20Applied%20Data%20Science%20Capstone%20/02%20-%20Exploratory%20Data%20Analysis%20(EDA)/Parte%202/02%20-%20module_2_jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb)

# EDA with SQL

---

1. Exiba os nomes dos locais de lançamento únicos na missão espacial
2. Exiba 5 registros onde os locais de lançamento começam com a string 'CCA'
3. Exiba a massa total da carga útil transportada pelos propulsores lançados pela NASA (CRS)
4. Exiba a massa média da carga útil transportada pela versão do propulsor F9 v1.1
5. Liste a data em que foi alcançado o primeiro resultado de pouso bem-sucedido na plataforma terrestre
6. Liste os nomes dos propulsores que tiveram sucesso no navio drone e têm massa de carga útil maior que 4000, mas menor que 6000
7. Liste o número total de resultados de missões bem-sucedidas e fracassadas
8. Liste os nomes das versões dos propulsores que transportaram a maior massa de carga útil
9. Liste os registros que exibirão os nomes dos meses, resultados de pouso fracassados no navio drone, versões dos propulsores, local de lançamento para os meses do ano de 2015
10. Classifique a contagem de resultados de pouso bem-sucedidos entre a data 04-06-2010 e 20-03-2017 em ordem decrescente

**LINK:**

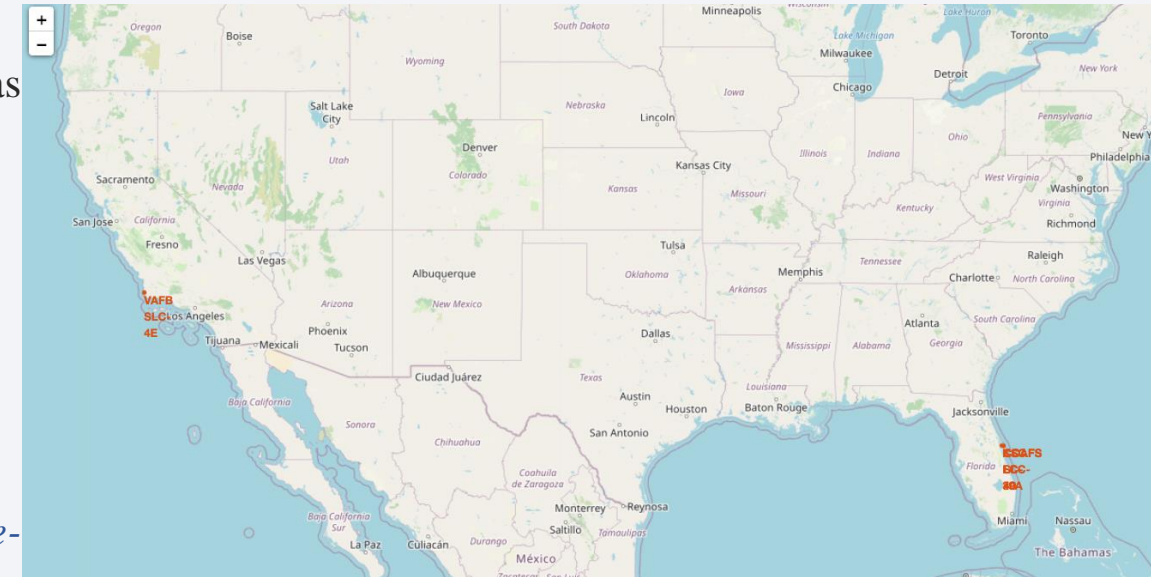
[https://github.com/Aline-Castro/IBM-Data-Science-Professional-Certificate/blob/main/10%20-%20Applied%20Data%20Science%20Capstone%20/02%20-%20Exploratory%20Data%20Analysis%20\(EDA\)/Parte%201%20/%2001%20-%20jupyter-labs-eda-sql-coursera\\_sqllite.ipynb](https://github.com/Aline-Castro/IBM-Data-Science-Professional-Certificate/blob/main/10%20-%20Applied%20Data%20Science%20Capstone%20/02%20-%20Exploratory%20Data%20Analysis%20(EDA)/Parte%201%20/%2001%20-%20jupyter-labs-eda-sql-coursera_sqllite.ipynb)



# Build an Interactive Map with Folium

- Marque todos os locais de lançamento em um mapa
- Marque os lançamentos bem-sucedidos/falhos para cada local no mapa
- Calcule as distâncias entre um local de lançamento e suas proximidades
  1. Se está perto da costa
  2. Se está perto da ferrovia
  3. Se está perto da rodovia
  4. Se está perto da cidade

**LINK:** [https://github.com/Aline-Castro/IBM-Data-Science-Professional-Certificate/blob/main/10%20-%20Applied%20Data%20Science%20Capstone%20/03%20-%20Interactive%20Visual%20Analytics%20and%20Dashboards/01-lab\\_jupyter\\_launch\\_site\\_location.ipynb](https://github.com/Aline-Castro/IBM-Data-Science-Professional-Certificate/blob/main/10%20-%20Applied%20Data%20Science%20Capstone%20/03%20-%20Interactive%20Visual%20Analytics%20and%20Dashboards/01-lab_jupyter_launch_site_location.ipynb)



# Build a Dashboard with Plotly Dash

---

1. Um componente de entrada suspenso do local de lançamento
  2. Um gráfico de pizza de sucesso baseado na lista suspensa do local selecionado
  3. Um cortador de intervalo para selecionar a carga útil
  4. Um gráfico de dispersão de sucesso-carga útil baseado na lista suspensa do local selecionado
- Para inspecionar a relação da taxa de sucesso entre o local de lançamento e a carga útil.

**LINK:** [https://github.com/Aline-Castro/IBM-Data-Science-Professional-Certificate/blob/main/10%20-%20Applied%20Data%20Science%20Capstone%20/03%20-%20Interactive%20Visual%20Analytics%20and%20Dashboards/01-lab\\_jupyter\\_launch\\_site\\_location.ipynb](https://github.com/Aline-Castro/IBM-Data-Science-Professional-Certificate/blob/main/10%20-%20Applied%20Data%20Science%20Capstone%20/03%20-%20Interactive%20Visual%20Analytics%20and%20Dashboards/01-lab_jupyter_launch_site_location.ipynb)

# Predictive Analysis (Classification)

---

- Manipulação de dados
- Padronização de dados
- Divisão em conjuntos de dados de treinamento e teste
- Avaliação do modelo preditivo
- Seleção do modelo preditivo
  1. Regressão logística
  2. Máquina de vetores de suporte
  3. Classificador de árvore de decisão
  4. K vizinhos mais próximos

**LINK:** [https://github.com/Aline-Castro/IBM-Data-Science-Professional-Certificate/blob/main/10%20-%20Applied%20Data%20Science%20Capstone%20/04%20-%20Predictive%20Analysis%20\(Classification\)/SpaceX\\_Machine\\_Learning\\_Prediction\\_Part\\_5\\_jupyterlite.ipynb](https://github.com/Aline-Castro/IBM-Data-Science-Professional-Certificate/blob/main/10%20-%20Applied%20Data%20Science%20Capstone%20/04%20-%20Predictive%20Analysis%20(Classification)/SpaceX_Machine_Learning_Prediction_Part_5_jupyterlite.ipynb)

# Results

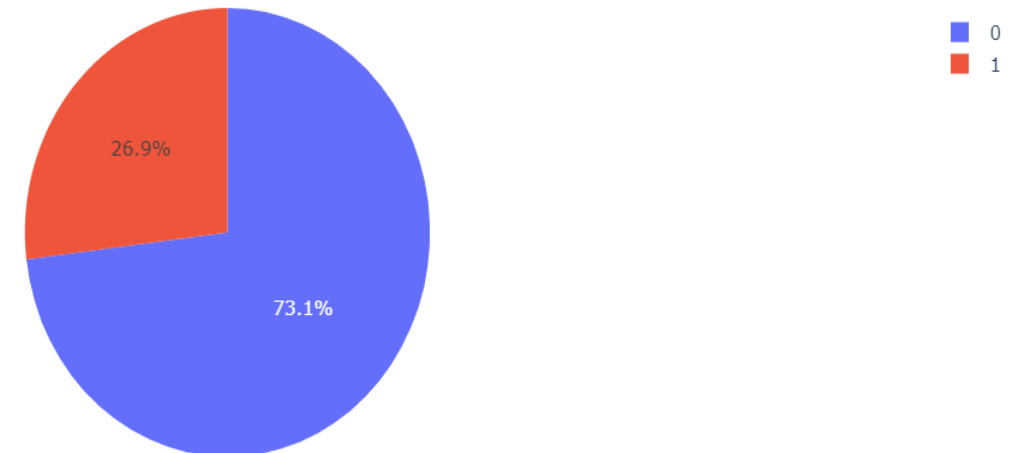
## EDA

1. KSC LC-39A e VAFB SLC 4E têm uma taxa de sucesso de 77%
2. VAFB SLC 4E não tem carga útil acima de 10000 kg
3. Na órbita LEO, o sucesso parece estar relacionado ao número de voos
4. Com cargas úteis pesadas, a taxa de pouso bem-sucedida ou positiva é maior para Polar, LEO e ISS
5. A taxa de sucesso desde 2013 continuou aumentando até 2020

**Predictive Analysis:** De acordo com o modelo de classificador de árvore de decisão, o modelo preditivo nos diz que haverá 4 verdadeiros positivos, 7 verdadeiros negativos, 5 falsos positivos e 2 falsos negativos. A precisão do modelo é de cerca de 89% com os melhores parâmetros.

CCAFS LC-40

Total Success Launches for site CCAFS LC-40





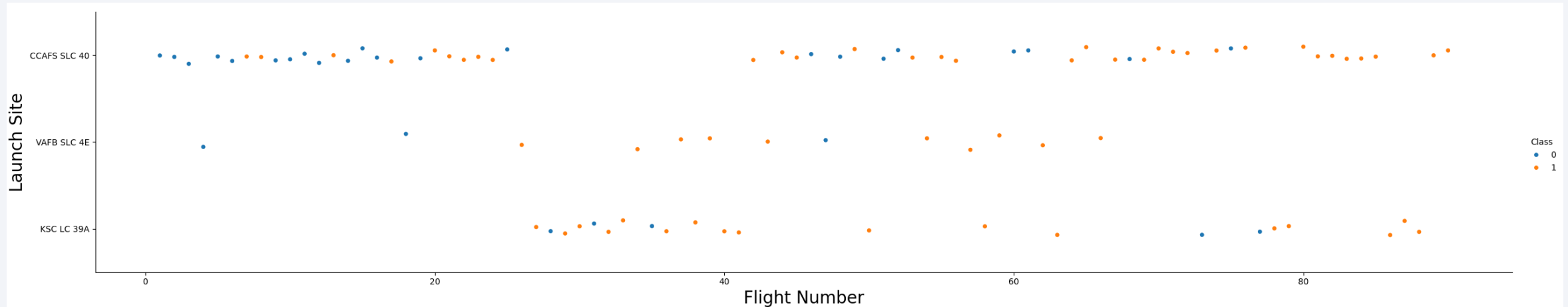
The background of the slide is an abstract composition. It features a dark blue base color. Overlaid on this are numerous diagonal streaks in shades of red and cyan. A faint, light blue grid pattern is also visible, particularly in the lower half of the image. The overall effect is dynamic and technological.

Section 2

# Insights drawn from EDA



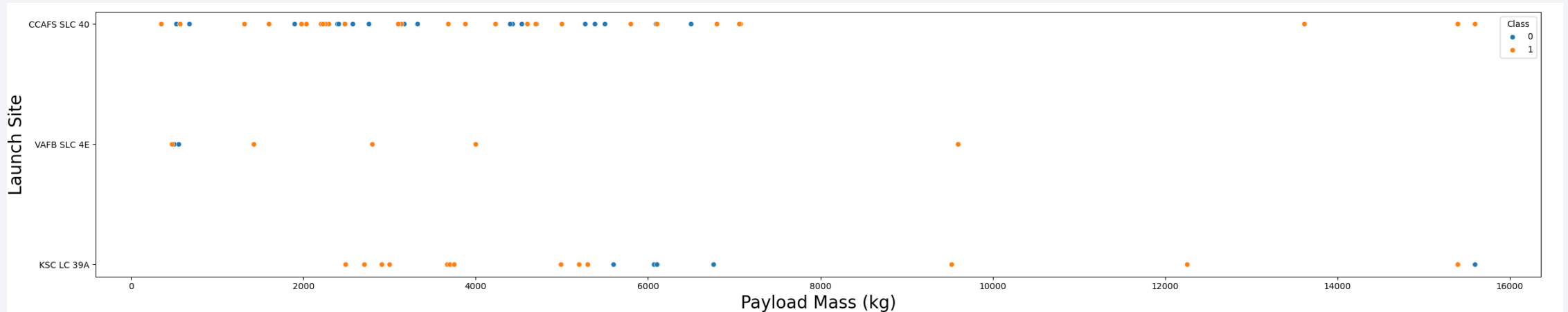
# Flight Number vs. Launch Site



A partir dos gráficos de dispersão do Número do Voo versus Local de Lançamento, podemos observar os seguintes padrões:

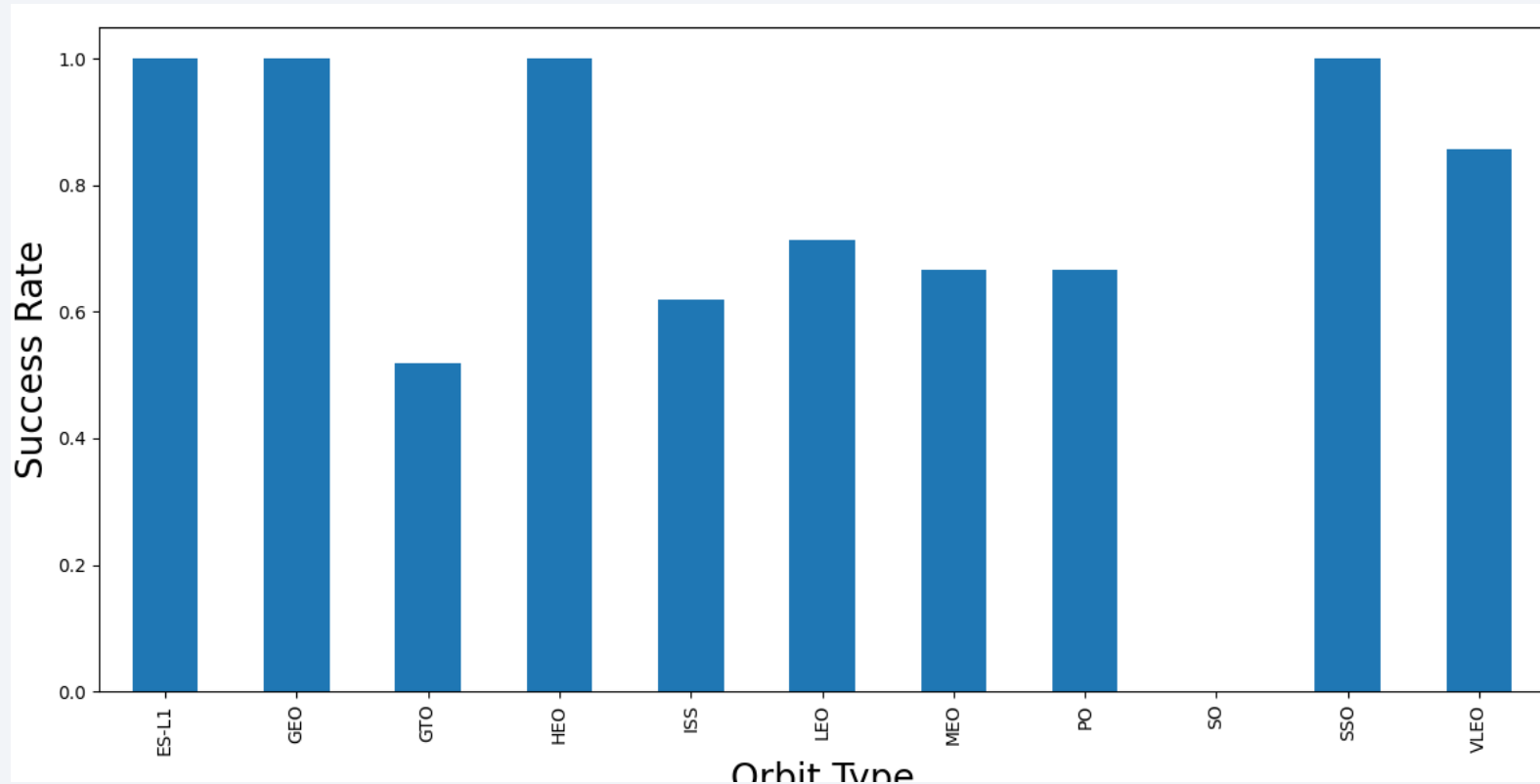
- **Distribuição dos Locais de Lançamento:** Os lançamentos não estão uniformemente distribuídos entre os locais de lançamento. Alguns locais de lançamento foram usados com mais frequência do que outros.
- **Progressão do Número do Voo:** Parece haver uma progressão nos números de voo para cada local de lançamento. Isso pode indicar que certos locais de lançamento foram preferidos ou disponíveis durante diferentes períodos.
- **Distribuição de Classes:** A codificação de cores representa a classe (ou seja, se o lançamento foi bem-sucedido ou não). Parece que a taxa de sucesso não varia significativamente com o número do voo ou o local de lançamento. No entanto, sem uma legenda clara, é desafiador tirar conclusões definitivas sobre a taxa de sucesso.
- **Sobreposição de Pontos de Dados:** Há uma sobreposição significativa de pontos de dados, especialmente para lançamentos do mesmo local. Essa sobreposição pode mascarar quaisquer padrões potenciais relacionados ao sucesso dos lançamentos.

# Payload vs. Launch Site



1. Não há foguetes lançados para a massa de carga pesada (maior que 10000 kg) no local de lançamento VAFB-SLC.
2. Quase todos os foguetes com massa de carga ao redor de 6000 kg têm voos malsucedidos no local de lançamento KSC LC 39A.
3. Todos os foguetes com carga superior a 7000 kg têm todos os voos bem-sucedidos no local de lançamento CCAFS SLC 40 e uma taxa muito alta de voos bem-sucedidos em outros locais de lançamento.

# Success Rate vs. Orbit Type

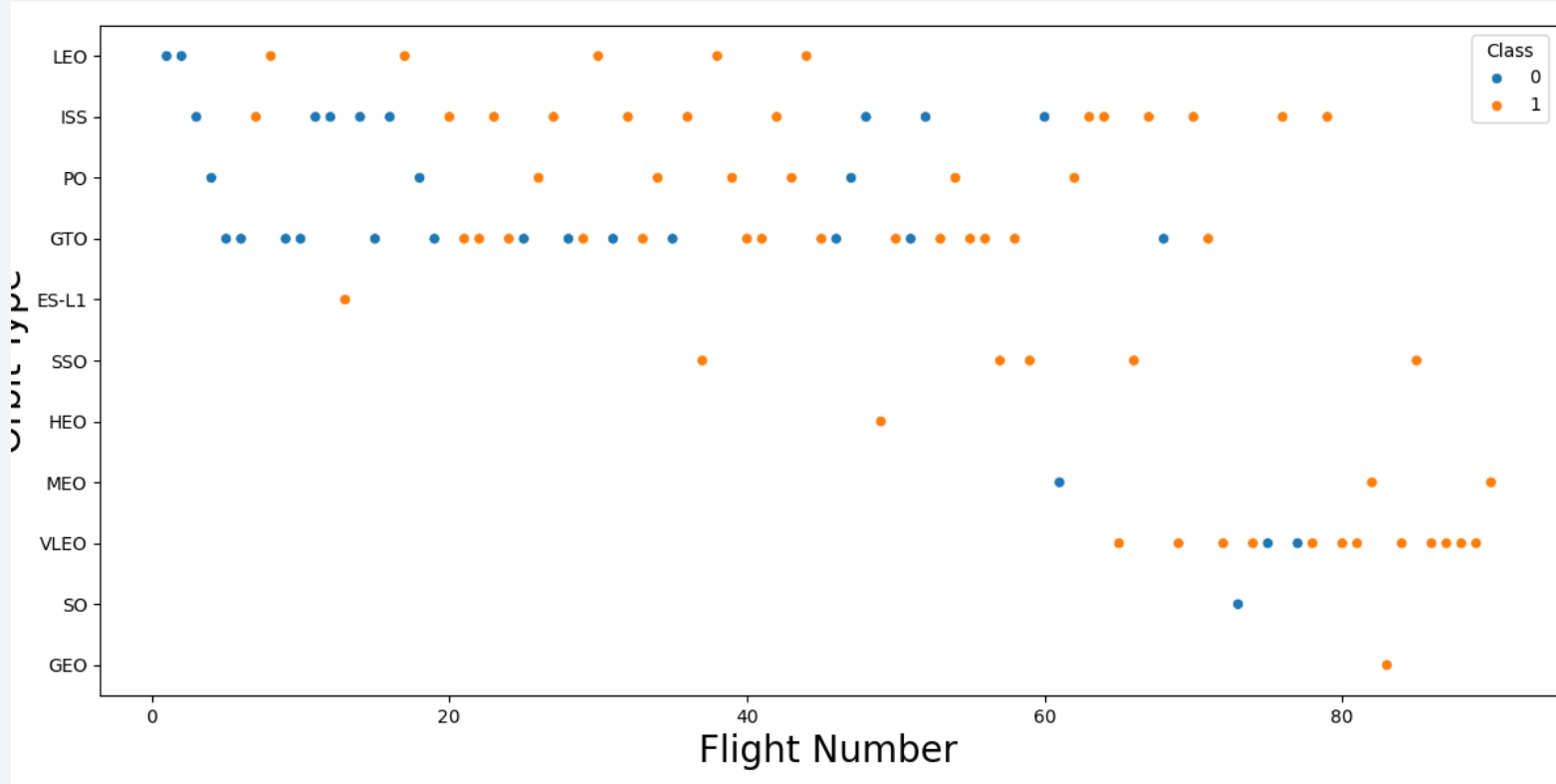


1. Apenas 4 de 11 órbitas (ES-L1, GEO, HEO, SSO) têm 100% de voos bem-sucedidos.

2. A órbita SO não tem voos bem-sucedidos.

3. A taxa média de voos bem-sucedidos para as outras órbitas é superior a 60%.

# Flight Number vs. Orbit Type

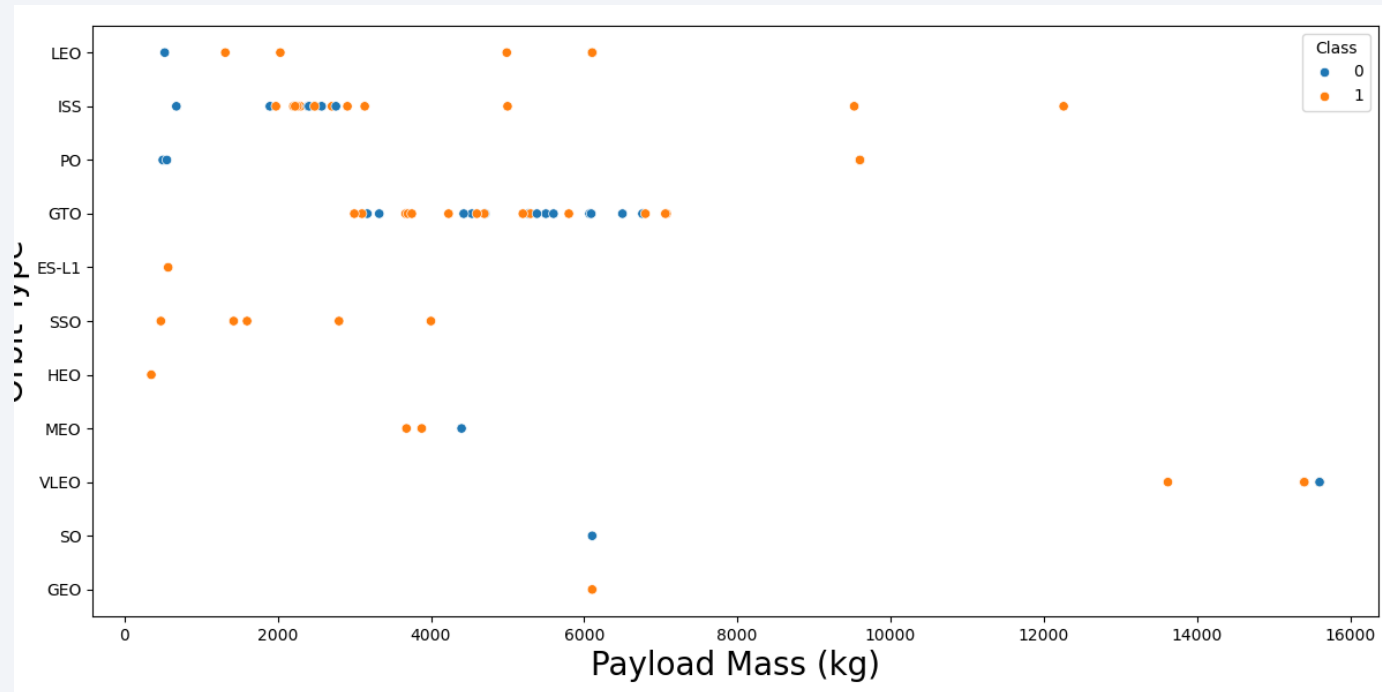


Parece haver uma relação entre o sucesso e o número de voos do LEO (Low Earth Orbit). Os sucessos (Classe 1) estão mais agrupados em torno de números de voos mais baixos.

Isto pode sugerir que, à medida que a SpaceX ganhou mais experiência com voos, a taxa de sucesso das missões LEO melhorou. Por outro lado, para o GTO (Órbita de Transferência Geoestacionária), os sucessos e fracassos (Classe 0 e Classe 1) estão distribuídos de forma mais uniforme entre os números de voo.

Isto indica que não existe uma relação clara entre o número de voos e o sucesso das missões GTO.

# Payload vs. Orbit Type



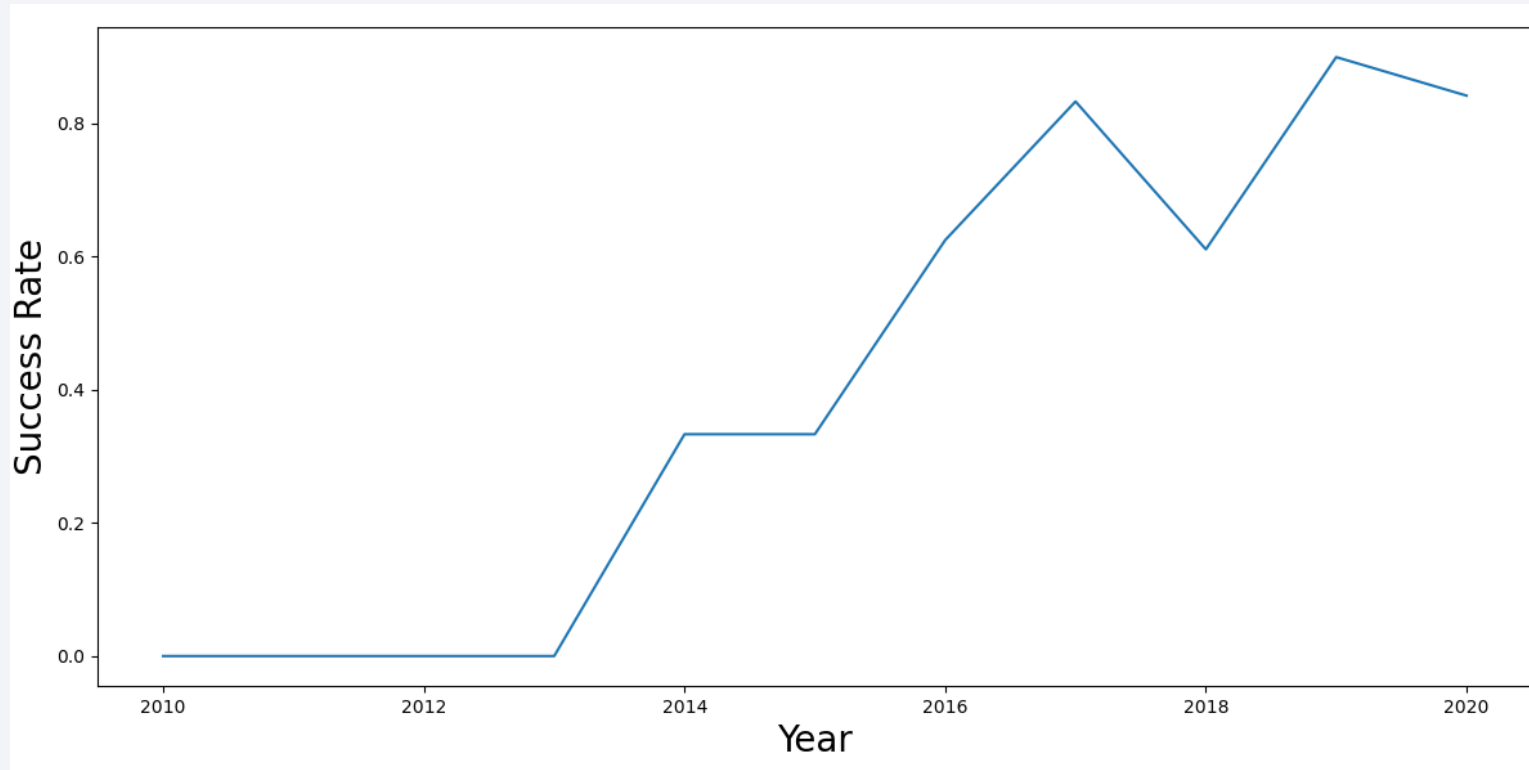
Com cargas pesadas, o pouso bem-sucedido ou a taxa de pouso positiva são maiores para Polar, LEO e ISS.

No entanto, para o GTO, não podemos distinguir isso bem, pois tanto a taxa de pouso positiva quanto a de pouso negativo (missão malsucedida) estão ambas aqui.



# Launch Success Yearly Trend

---



É observável que a taxa de sucesso tem realmente aumentado desde 2013 até 2020.

O gráfico começa com uma taxa de sucesso de quase zero em 2010, com um aumento acentuado por volta de 2013, continuando uma tendência ascendente até atingir o pico de uma taxa de sucesso de aproximadamente 0,7. por volta do final de 2019 ou início de 2020.

Depois de atingir o seu pico, há um ligeiro declínio na taxa de sucesso, retratado pela desaceleração da linha em direção a 2020.

Esta tendência sugere que houve melhorias na taxa de sucesso dos lançamentos ao longo dos anos.

# All Launch Site Names

---

```
%sql SELECT DISTINCT "Launch_Site" FROM SPACEXTABLE;  
  
* sqlite:///my_data1.db  
Done.  
  
Launch_Site  
-----  
CCAFS LC-40  
VAFB SLC-4E  
KSC LC-39A  
CCAFS SLC-40
```

Existem 4 locais de lançamento de acordo com o resultado da consulta SQL.

# Launch Site Names Begin with 'CCA'

```
%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

# Total Payload Mass

---

```
%sql SELECT SUM(PAYLOAD_MASS_KG_) FROM SPACEXTBL WHERE CUSTOMER LIKE "%NASA (CRS)%"
* sqlite:///my_data1.db
Done.
```

<b>SUM(PAYLOAD_MASS_KG_)</b>
48213

A massa total da carga útil transportada pelos boosters da NASA é 48213kg.

# Average Payload Mass by F9 v1.1

---

```
%sql SELECT AVG("Payload_Mass__kg_") FROM SPACEXTABLE WHERE "Booster_Version" = 'F9 v1.1';
```

```
* sqlite:///my_data1.db
```

```
Done.
```

```
AVG("Payload_Mass__kg_")
```

```
2928.4
```

Uma massa média de carga transportada por booster a versão F9 v1.1 pesa 2.928,4 kg.



# First Successful Ground Landing Date

---

```
%sql SELECT MIN(Date) FROM SPACEXTBL WHERE Mission_Outcome = "Success"
* sqlite:///my_data1.db
Done.
MIN(Date)
01-03-2013
```

A data em que o primeiro resultado de pouso bem-sucedido no terreno foi alcançado em 1º de março de 2013.

# Successful Drone Ship Landing with Payload between 4000 and 6000

---

```
%sql SELECT DISTINCT "Booster_Version" FROM SPACEXTABLE WHERE "Landing_Outcome" = 'Success (drone ship)' AND "Payload_Mass_
* sqlite:///my_data1.db
Done.
Booster_Version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2
```

Existem apenas quatro boosters que pousaram com sucesso no drone navio e tinha massa de carga útil superior a 4.000, mas inferior a 6.000, seu os nomes são:

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

# Total Number of Successful and Failure Mission Outcomes

---

```
%sql SELECT Mission_Outcome, COUNT(Mission_Outcome) FROM SPACEXTBL GROUP BY Mission_Outcome
```

```
* sqlite:///my_data1.db
```

```
Done.
```

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

- O número total de resultados de missões bem-sucedidas é 100
- O número total de resultados de missões com falha é 1

# Boosters Carried Maximum Payload

---

```
%sql SELECT "Booster_Version" FROM SPACESTABLE WHERE "Payload_Mass__kg_" = (SELECT MAX("Payload_Mass__kg_") FROM SPACESTABLE)

* sqlite:///my_data1.db
Done.
```

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

Existem 12 boosters que transportaram a massa máxima da carga útil.

# 2015 Launch Records

---

```
%sql SELECT SUBSTR("Date", 6, 2) AS Month, "Booster_Version", "Launch_Site", "Landing_Outcome" FROM SPACEXTABLE WHERE "Landi
```

```
* sqlite:///my_data1.db  
Done.
```

Month	Booster_Version	Launch_Site	Landing_Outcome
01	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
04	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

No ano de 2015 houve apenas duas falhas em drones resultados de desembarque de navios, em janeiro e em abril, ambos do local de lançamento do CCAFS LC-40.

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

---

```
%sql SELECT "Landing _Outcome", COUNT(*) AS COUNT_LAUNCHES FROM SPACEXTBL WHERE "Landing _Outcome" LIKE 'Success'

* sqlite:///my_data1.db
Done.
```

Landing _Outcome	COUNT_LAUNCHES
Success	20
Success (drone ship)	8
Success (ground pad)	6

Houve 34 pousos bem-sucedidos entre 04.06.2010 e 20.03.2017, sendo 8 para drone navio, 6 estão em uma base terrestre, 20 sem qualquer Informações adicionais.

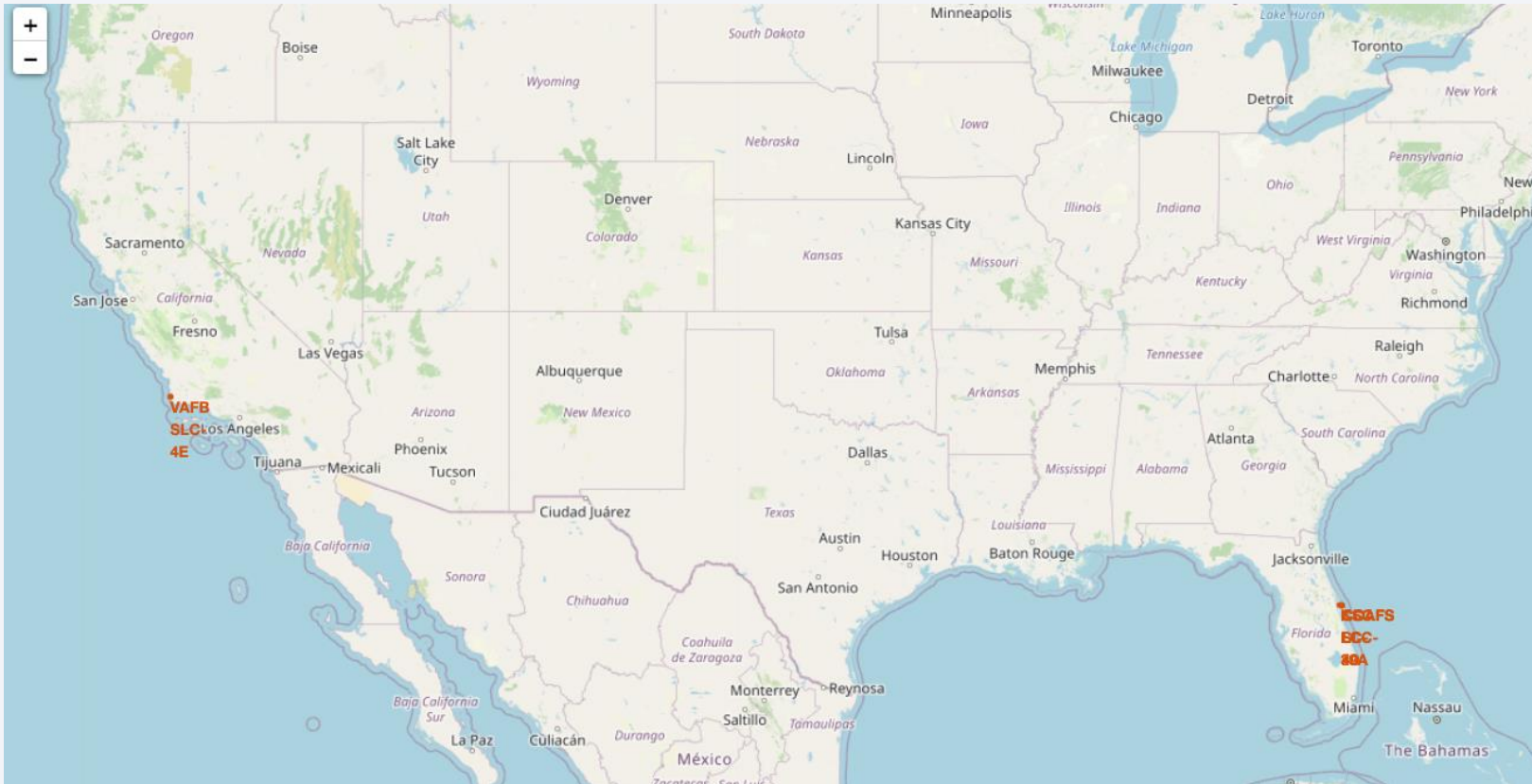
A satellite view of Earth from space, showing the curvature of the planet and city lights at night. The background is a deep blue gradient.

Section 3

# Launch Sites Proximities Analysis

# Map with market launch sites

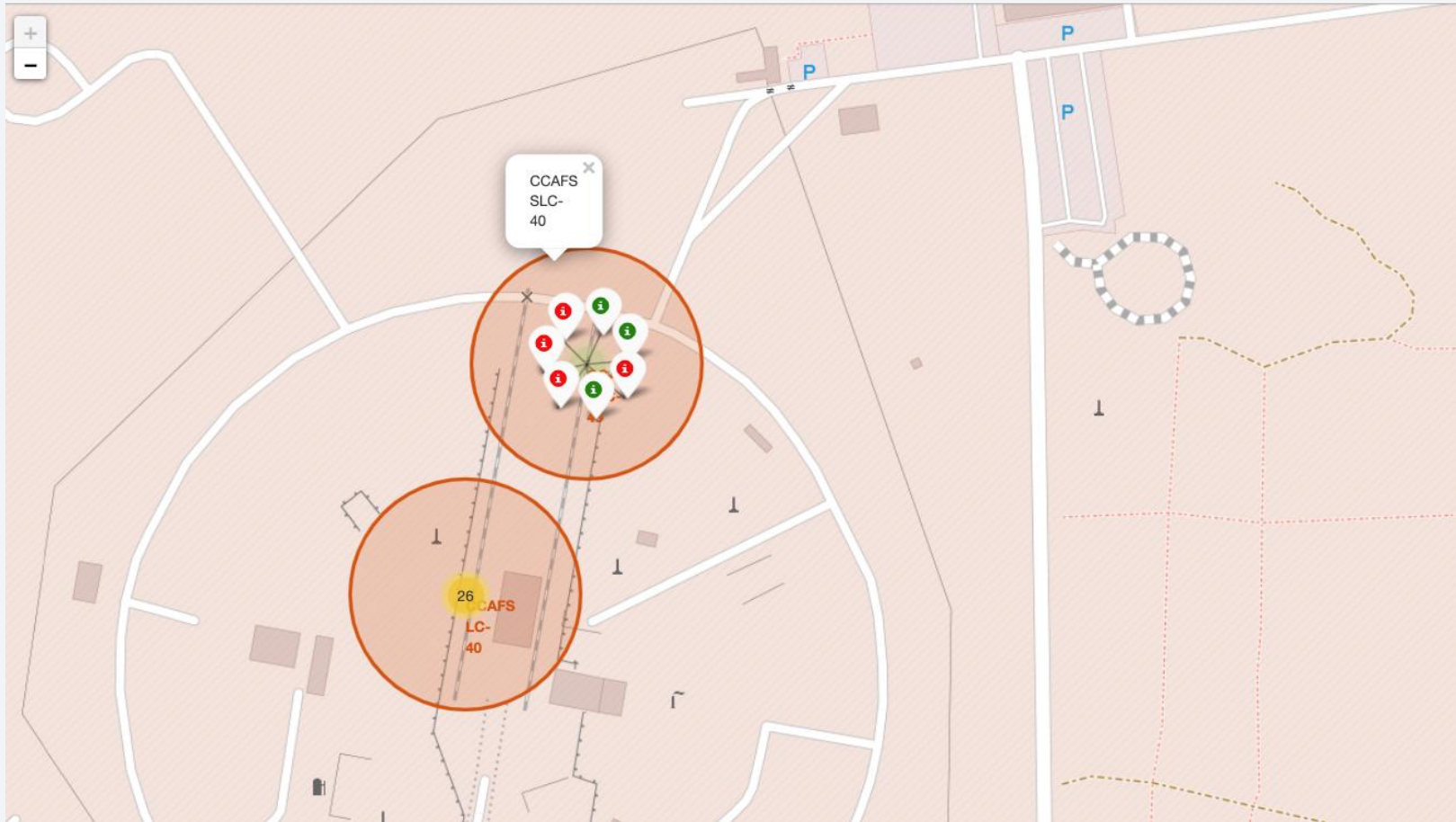
---



- O local de lançamento do VAFB SLC 4E está localizado na costa oeste do continente da América do Norte, enquanto dois outros locais de lançamento (KSC LC 39A e CCAFS SLC 40) estão localizados na costa leste.
- Enquanto isso, todos os três locais estão próximos dos oceanos.



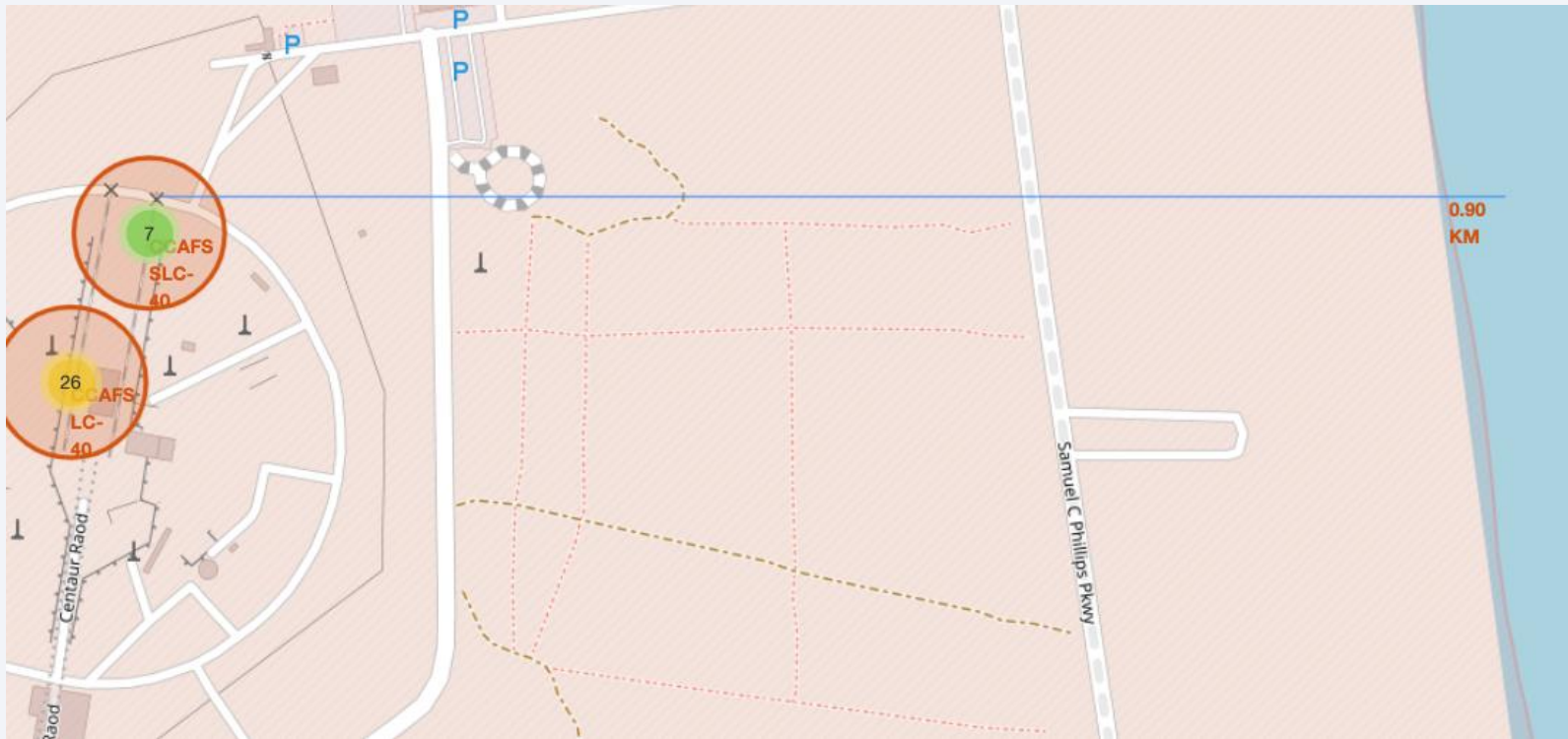
# Map with marked successful and failed launches



Dos marcadores rotulados por cores em clusters de marcadores, somos capazes para identificar facilmente que KSC LC-39<sup>a</sup> O local de lançamento tem relativamente alta taxa de sucesso.

## Map with the Distances between a Launch Site to its proximities

---



O local de lançamento do CCAFS SLC 40 é localizado a menos de 1 km de da costa, poderá ter impacto sobre os resultados do lançamento.





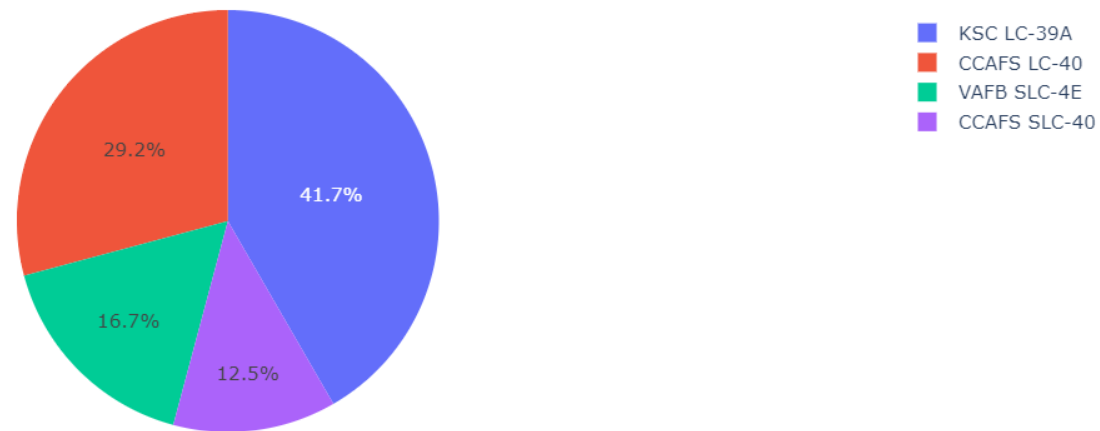
Section 4

# Build a Dashboard with Plotly Dash

# Success Rates

---

Success Count for all launch sites



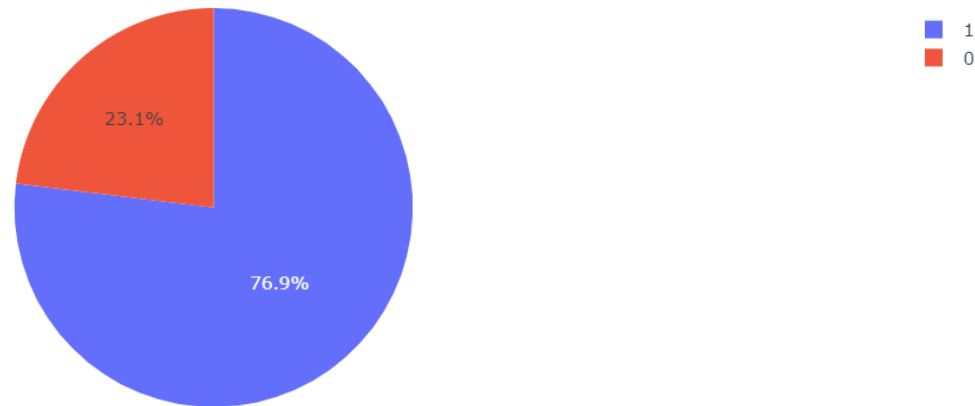
O gráfico de pizza ilustra que o local de lançamento do KSC LC-39A tem o maior número de voos bem-sucedidos entre todos locais de lançamento.

Contudo, isso não significa que seja o melhor site para lançamento, poderia ser apenas uma aprovação que o número de tentativas deste site é maior.

# Launch Site with Highest Launch Success Ratio

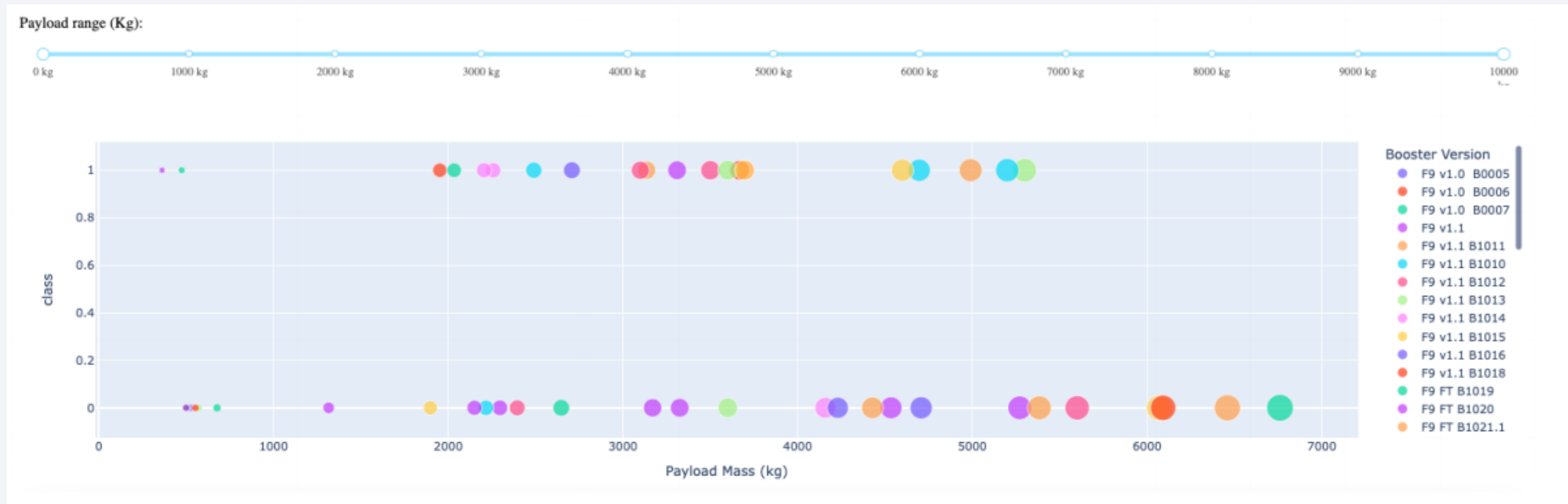
---

Total Success Launches for site KSC LC-39A



O local de lançamento KSC LC-39A tem a maior taxa de sucesso de lançamento, quase 77%.

# Payload vs. Launch Outcome



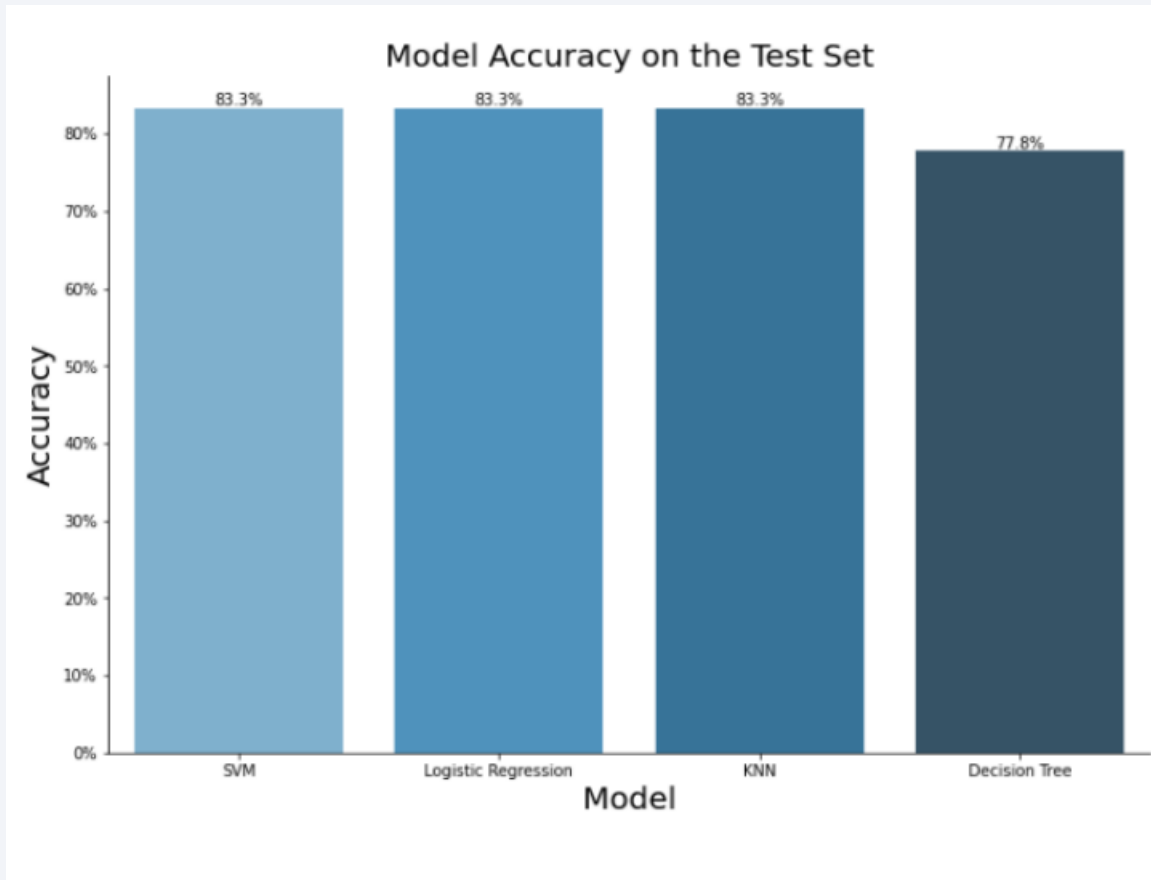
O gráfico de dispersão para todos os locais demonstra que os boosters mais pesados (mais de 5.500 kg) não têm chances de voos bem-sucedidos. O mesmo podemos concluir para os boosters leves (500-1500 kg).

Section 5

# Predictive Analysis (Classification)

# Classification Accuracy

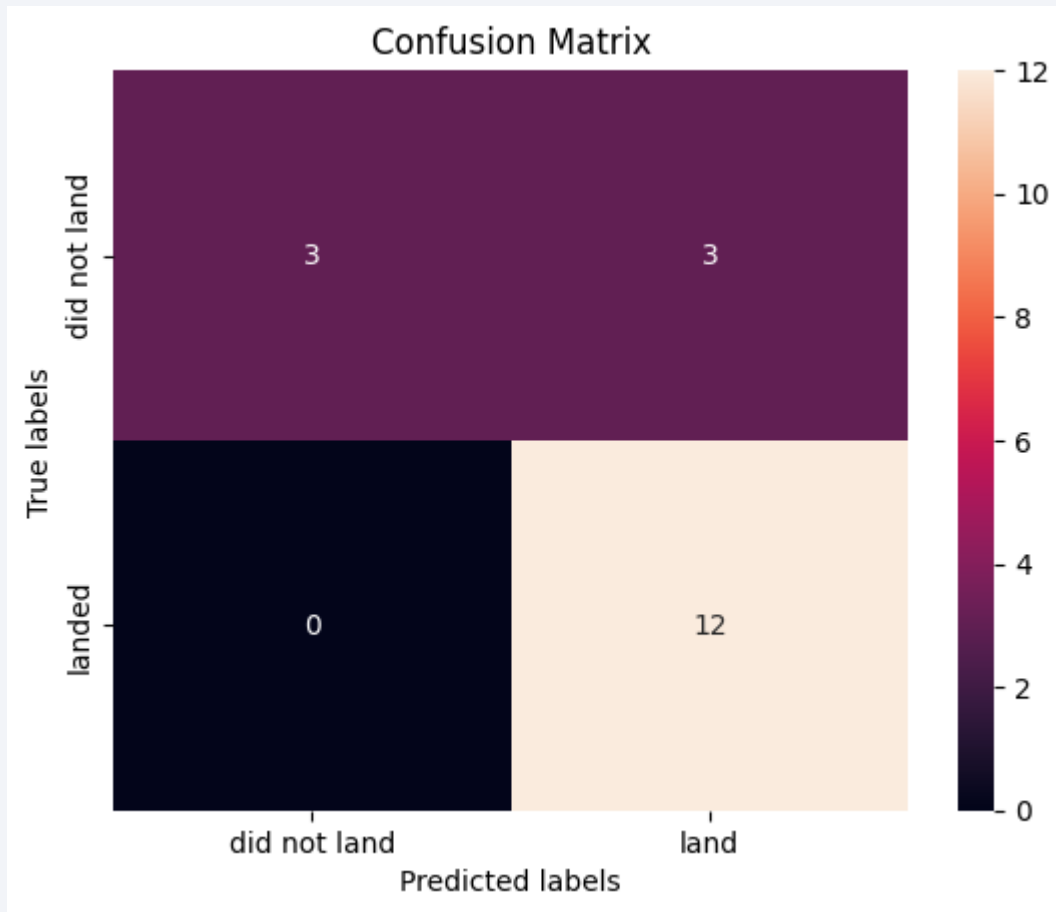
---



Três em cada quatro modelos têm a maior precisão de classificação, 83,3%, o que significa que todos eles podem ser usados como modelos de classificação para aprendizado de máquina supervisionado.



# Confusion Matrix



As matrizes de confusão para todos os três modelos apropriados são as mesmas e ilustram que 83,3% de todas as previsões estarão corretas. Ao mesmo tempo, 16,7% serão erroneamente previstos como bem-sucedidos.

# Conclusions

---

Para aumentar a probabilidade de um lançamento bem-sucedido, é melhor:

- Escolher as órbitas ES-L1, GEO, HEO, SSO, pois elas têm 100% de sucesso voos;
- Realizar um lançamento a partir do local de lançamento KSC LC-39A, pois possui o maior proporção de resultados bem-sucedidos;
- Evitar lançamentos muito leves (até 2000 kg) e muito pesados (de 5.500 kg), pois têm as menores chances de voos bem-sucedidos;
- Tentar mais, porque:
  - maior número de voos, maior possibilidade de sucesso;
  - a taxa de sucesso de lançamento está aumentando permanentemente desde 2013

# Appendix

---

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

Thank you!

