



IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Rafael Serafim Agum
29/02/24



Outline

- **Sumário executivo**
- **Introdução**
- **Metodologia**
- **Resultados**
- **Conclusão**
- **Apêndice**

Sumário Executivo

- **Resumo das metodologias**
- **Resumo de todos os resultados**

Introdução

- **Histórico e contexto do projeto**
- **Problemas para os quais você deseja encontrar respostas**

Seção 1

Methodology

Methodology

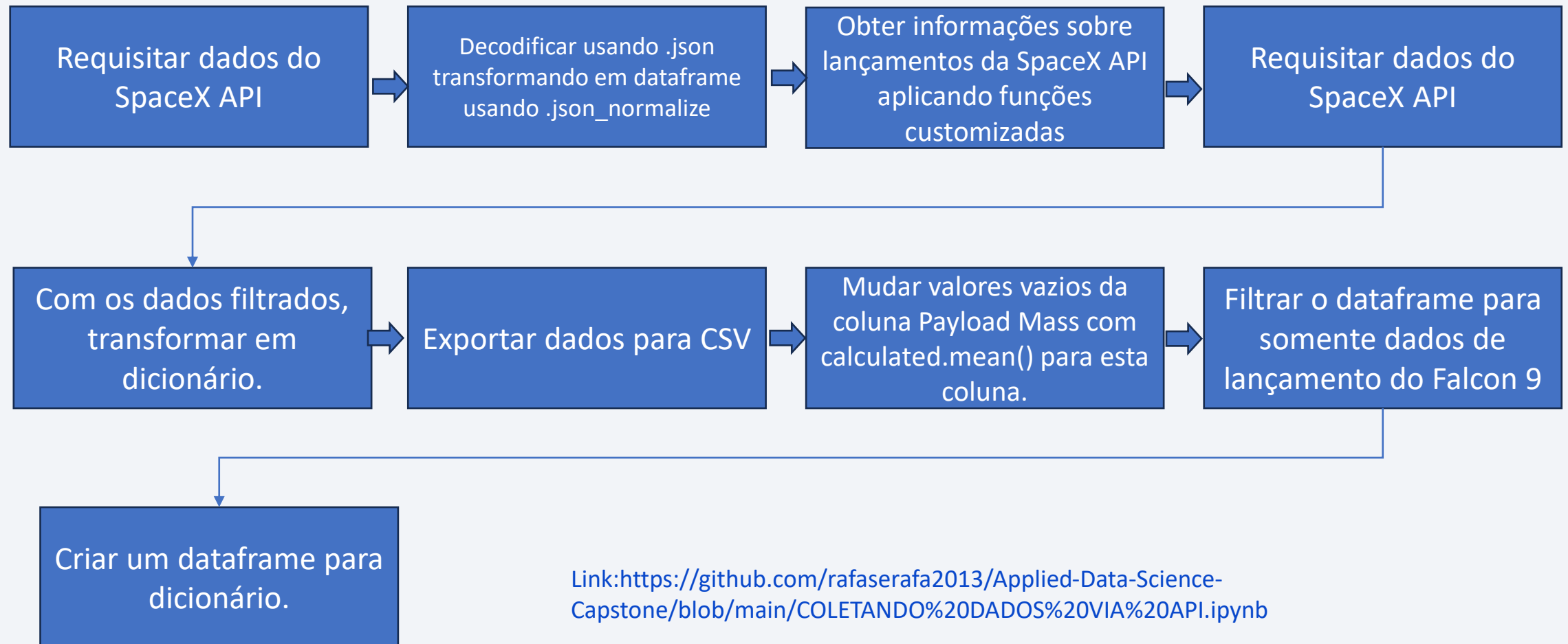
Sumário Executivo

- Metodologia para coleta de dados:
 - Os dados foram coletados usando o REST API da SpaceX e raspagem de dados via Wikipedia.
- Preparação dos dados
 - Os dados foram processados usando codificação one-hot para recursos categóricos.
- Análise exploratória dos dados (EDA) usando visualização e SQL
- Análises visuais interativas usando Folium e Plotly Dash
- Análises preditivas usando modelos de machine learning
 - Construção, ajuste e avaliação de modelos de classificação para garantir os melhores resultados.

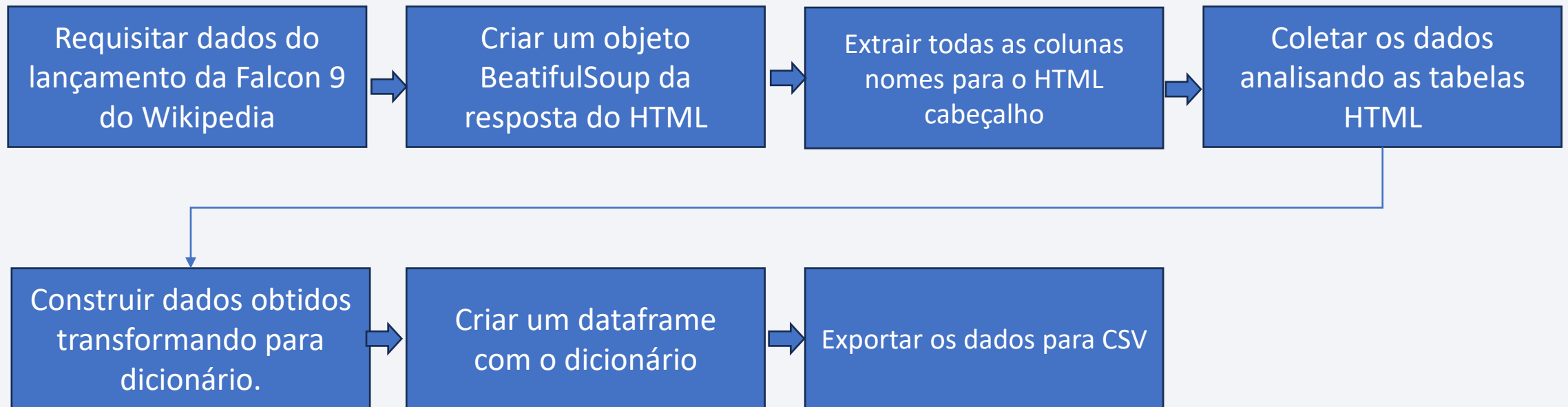
Coleta de dados

- A coleta de dados é o processo de reunir e medir informações sobre variáveis em um sistema estabelecido, que então permite responder perguntas e avaliar resultados. Conforme mencionado, o conjunto de dados foi coletado por RESTAPI e Web Scrapping da Wikipedia.
- Para API REST, é iniciado usando a solicitação get. Então, decodificamos a resposta conteúdo como Json e transformá-lo em um dataframe do pandas usando `json_normalize()`. Nós em seguida, limpei os dados, verifiquei os valores ausentes e preenchi com o que fosse necessário.
- Para web scrapping, usaremos o BeautifulSoup para extrair os registros de lançamento como tabela HTML, analise a tabela e converta-a em um dataframe do pandas para posterior análise.

Coleta de dados via SpaceX API



Coleta de dados via Raspagem de dados

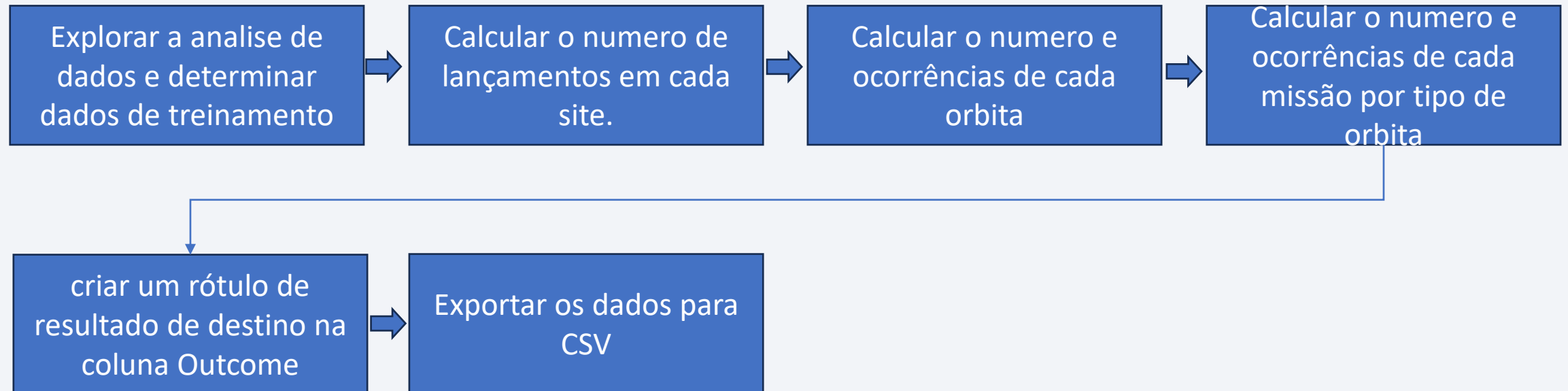


Link: <https://github.com/rafaserafa2013/Applied-Data-Science-Capstone/blob/main/COLETANDO%20DADOS%20VIA%20WEB%20SCRAPPING.ipynb>

Preparação dos dados

- Data Wrangling é o processo de limpeza e unificando conjuntos de dados confusos e complexos para fácil acesso e Análise Exploratória de Dados (EDA).
- Primeiro calcularemos o número de lançamentos em cada local e, em seguida, calcular o número e a ocorrência do resultado da missão por tipo de órbita.
- Em seguida, criamos um rótulo de resultado de destino a partir do coluna de resultado. Isto tornará mais fácil para futuras análises, visualização e machine learning. Por último, exportaremos o resultado para um CSV.

Preparação dos dados



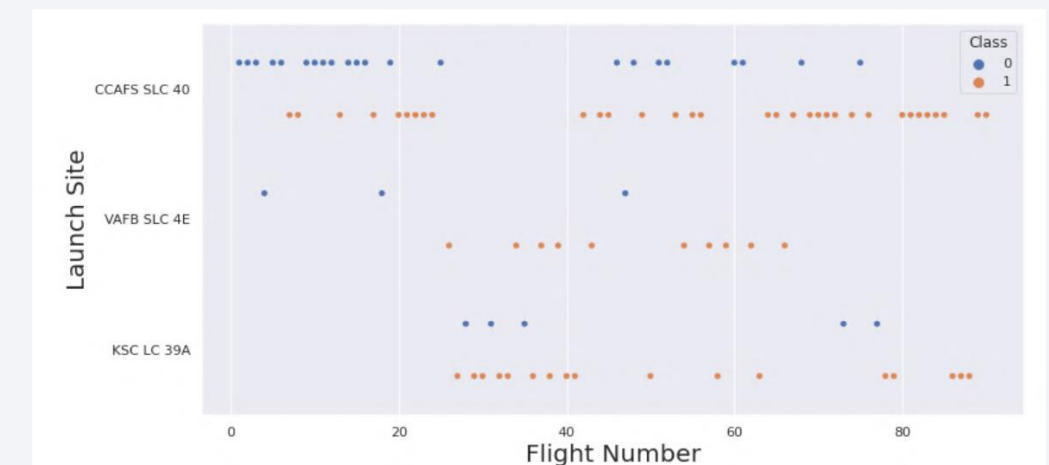
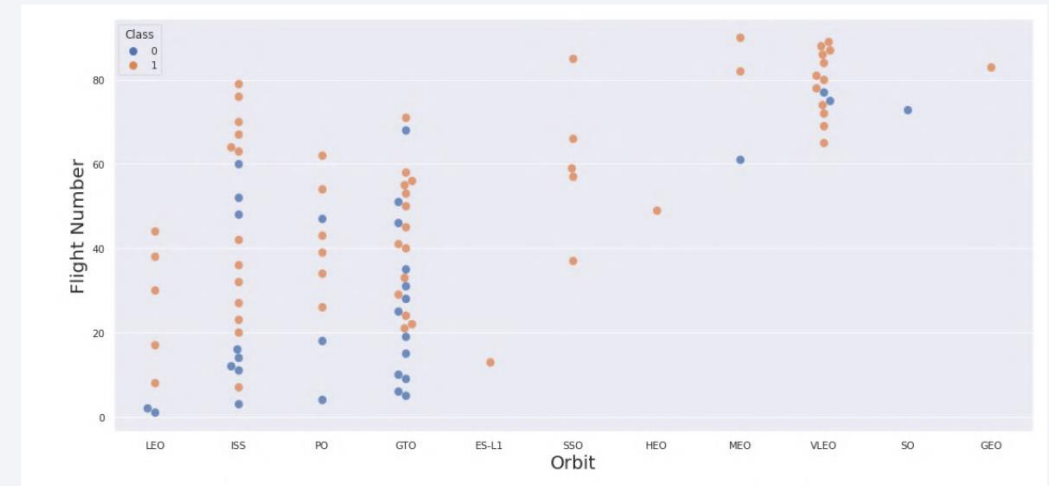
Link: <https://github.com/rafaserafa2013/Applied-Data-Science-Capstone/blob/main/DISPUTA%20DE%20DADOS.ipynb>

EDA com visualização dos dados

Começamos usando o gráfico de dispersão para encontrar a relação entre os atributos, como entre:

- Carga Útil e Número do Voo.
- Número do voo e local de lançamento.
- Carga útil e local de lançamento.
- Número do voo e tipo de órbita.
- Carga útil e tipo de órbita.

Os gráficos de dispersão mostram a dependência dos atributos entre si. Uma vez que um padrão é determinado a partir dos gráficos, é muito fácil. Veja quais fatores afetam mais o sucesso do resultados de pouso.

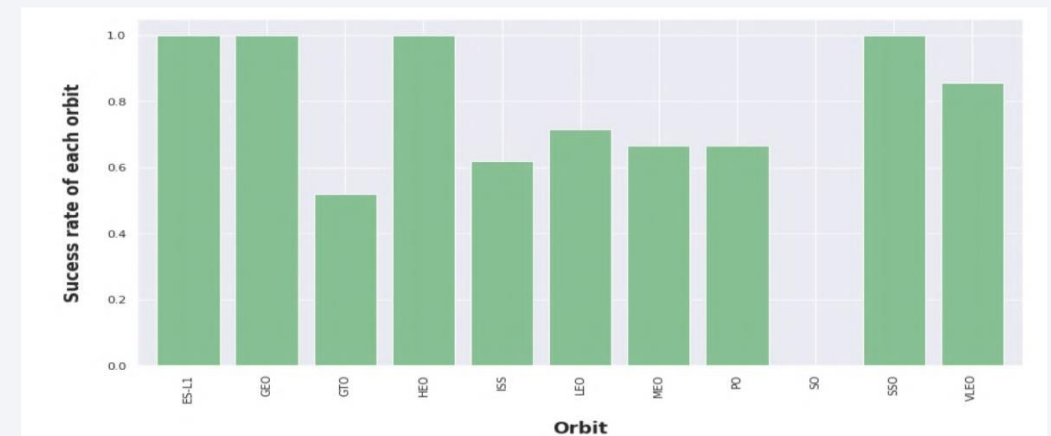
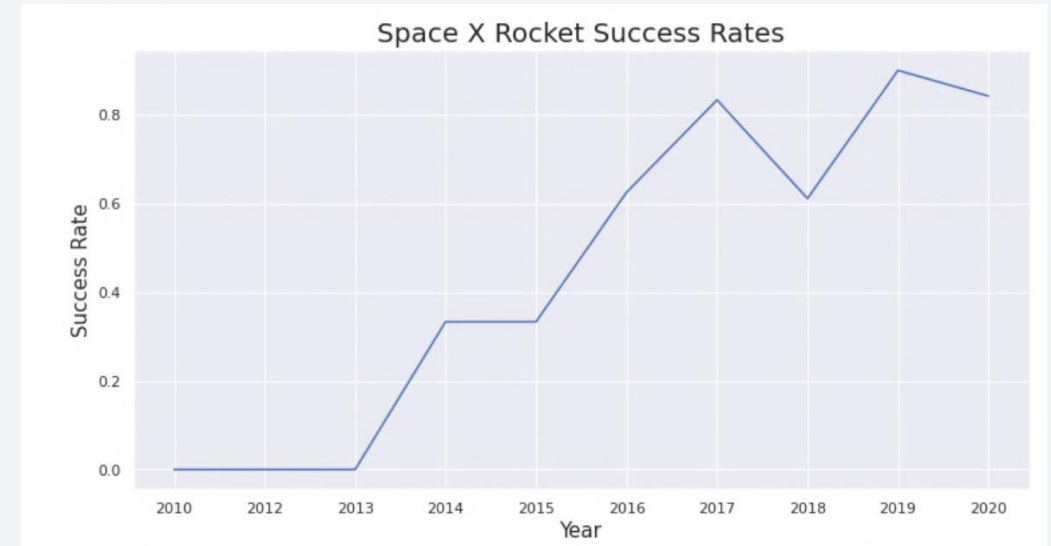


Link: <https://github.com/rafaserafa2013/Applied-Data-Science-Capstone/blob/main/EDA%20COM%20VISUALIZA%C3%87%C3%83O%20DE%20DADOS.ipynb>

EDA com visualização dos dados

Desta forma, conseguimos achar relações usando o gráfico de dispersão. Usaremos então outras ferramentas de visualização, como gráfico de barras e gráfico de plotagens de linha para análise exploratórias. Os gráficos de barras são uma das maneiras mais fáceis de interpretar o relacionamento entre os atributos. Neste caso, usaremos o gráfico de barras para determinar quais órbitas têm a maior probabilidade de sucesso. Em seguida, usamos o gráfico de linha para mostrar tendências ou padrão do atributo ao longo do tempo que, neste caso, é usado para ver a tendência de lançamento anual de sucesso. Em seguida, usamos a engenharia de recursos para ser usada com sucesso na previsão do módulo futuro criando variáveis para colunas categóricas.

Link:<https://github.com/rafaserafa2013/Applied-Data-Science-Capstone/blob/main/EDA%20COM%20VISUALIZA%C3%87%C3%83O%20DE%20ADOS.ipynb>



EDA com SQL

Usando SQL, realizaremos muitas consultas para entender melhor o conjunto de dados, Ex:

- Exibindo os nomes dos locais de lançamento.
- Exibindo 5 registros onde os locais de lançamento começam com a string 'CCA'.
- Exibindo a massa total da carga transportada pelo booster lançado pela NASA (CRS).
- Exibindo a massa média da carga transportada pela versão booster F9 v1.1.
- Listar a data em que o primeiro resultado de pouso bem-sucedido no solo foi alcançado.
- Listando os nomes dos boosters que têm sucesso em navios drones e têm massa de carga útil maior que 4.000, mas menor que 6.000.
- Listando o número total de resultados de missões bem-sucedidas e fracassadas.
- Listando os nomes das booster_versions que transportaram a massa máxima de carga útil.
- Listando os landing_outcomes com falha no navio drone, suas versões de reforço e locais de lançamento nomes para o ano de 2015.
- Classificando a contagem de resultados ou sucessos de pouso entre a data 04/06/2010 e 20/03/2017, em ordem decrescente.

Link:<https://github.com/rafaserafa2013/Applied-Data-Science-Capstone/blob/main/EDA%20COM%20SQL.ipynb>

Construindo um mapa interativo com Folium

Marcadores de todos os locais de lançamento:

- Adicionado marcador com círculo, rótulo pop-up e rótulo de texto do NASA Johnson Space Center usando suas coordenadas de latitude e longitude como local de partida.
- Adicionados marcadores com círculo, rótulo pop-up e rótulo de texto de todos os sites de lançamento usando sua latitude e coordenadas de longitude para mostrar suas localizações geográficas e proximidade com o Equador e costas.

Marcadores coloridos dos resultados do lançamento para cada local de lançamento:

- Adicionados marcadores coloridos de sucesso (verde) e lançamentos com falha (vermelho) usando Marker Cluster para identificar quais locais de lançamento têm taxas de sucesso relativamente altas.

Distâncias entre um Local de Lançamento e suas proximidades:

- Adicionadas linhas coloridas para mostrar as distâncias entre o local de lançamento KSC LC-39A (exemplo) e suas proximidades como Ferrovia, Rodovia, Litoral e Cidade Mais Próxima.

Link: <https://github.com/rafaserafa2013/Applied-Data-Science-Capstone/blob/main/ANALISE%20DE%20INTERA%C3%87%C3%83O%20VISUAL%20COM%20FOLIUM.ipynb>

Construindo um dashboard com Plotly Dash

Construiremos um painel interativo com Plotly dash que permite ao usuário alocar os dados conforme sua necessidade.

Traçaremos gráficos de pizza mostrando o total de lançamentos de determinados sites.

Em seguida, traçaremos um gráfico de dispersão mostrando a relação com o Resultado e a Carga Útil Massa (Kg) para as diferentes versões de reforço.

Link:<https://github.com/rafaserafa2013/Applied-Data-Science-Capstone/blob/main/RAFASERAFA.APP.py>

Análises Preditivas (Classificação)

Construindo o Modelo

- Carregar o conjunto de dados em NumPy e Pandas
- Transformar os dados e então dividir em dados de treinamento e de dados de teste.
- Decidir qual tipo de ML usar.
- definir os parâmetros e algoritmos para GridSearchCV e ajusta-lo ao conjunto de dados.

Avaliando o modelo

- Verificar a precisão de cada modelo.
- Obter hiperparâmetros ajustados para cada tipo de algoritmo.
- plotar a matriz de confusão.

Melhorando o modelo

- Usar engenharia de recursos e ajuste de algoritmo

Encontrar o melhor modelo

- O modelo com o melhor pontuação de precisão será o modelo de melhor desempenho.

Link: <https://github.com/rafaserafa2013/Applied-Data-Science-Capstone/blob/main/RAFASERAFA.APP.py>

Resultados

Os resultados serão categorizados em 3 resultados principais, que serão:

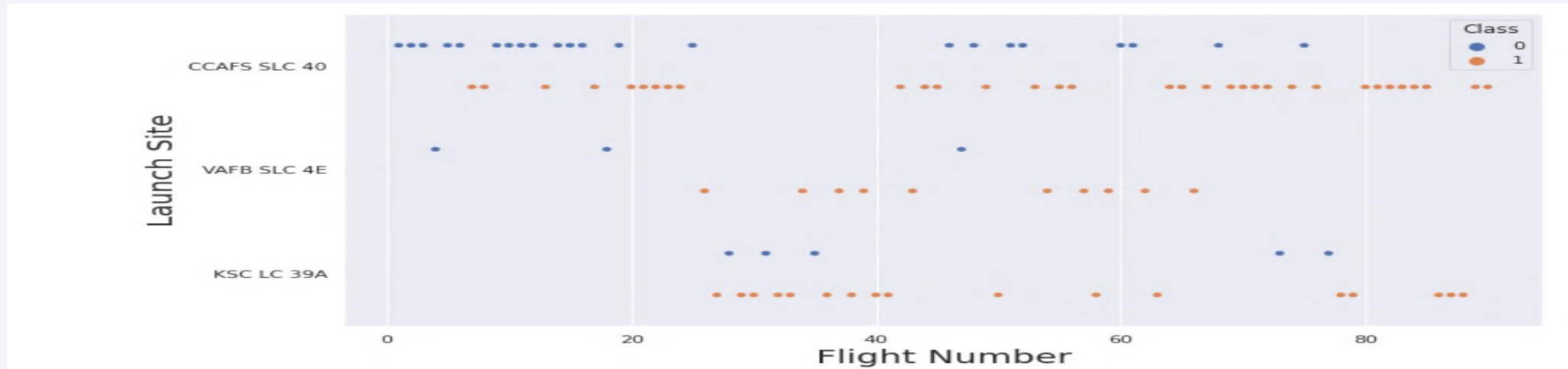
- Resultados de análise exploratória de dados.
- Demonstração de análise interativa em capturas de tela.
- Resultados de análises preditivas.

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-left quadrant. The overall effect is dynamic and technological.

Section 2

Insights drawn from EDA

Números de voos vs Plataforma de lançamentos



Este gráfico de dispersão mostra que quanto maior o valor dos voos do local de lançamento, maior será a taxa de sucesso. No entanto, o site CCAFS SLC40 mostra o mínimo padrão disso.

Pontos a serem observados:

- Todos os primeiros voos fracassaram, enquanto os últimos voos tiveram sucesso.
- O local de lançamento do CCAFS SLC 40 possui cerca de metade de todos os lançamentos.
- VAFB SLC 4E e KSC LC 39A apresentaram taxas de sucesso mais altas.
- Pode-se presumir que a cada novo lançamento tem uma taxa de sucesso mais elevada.

Carga útil vs. Local de lançamento

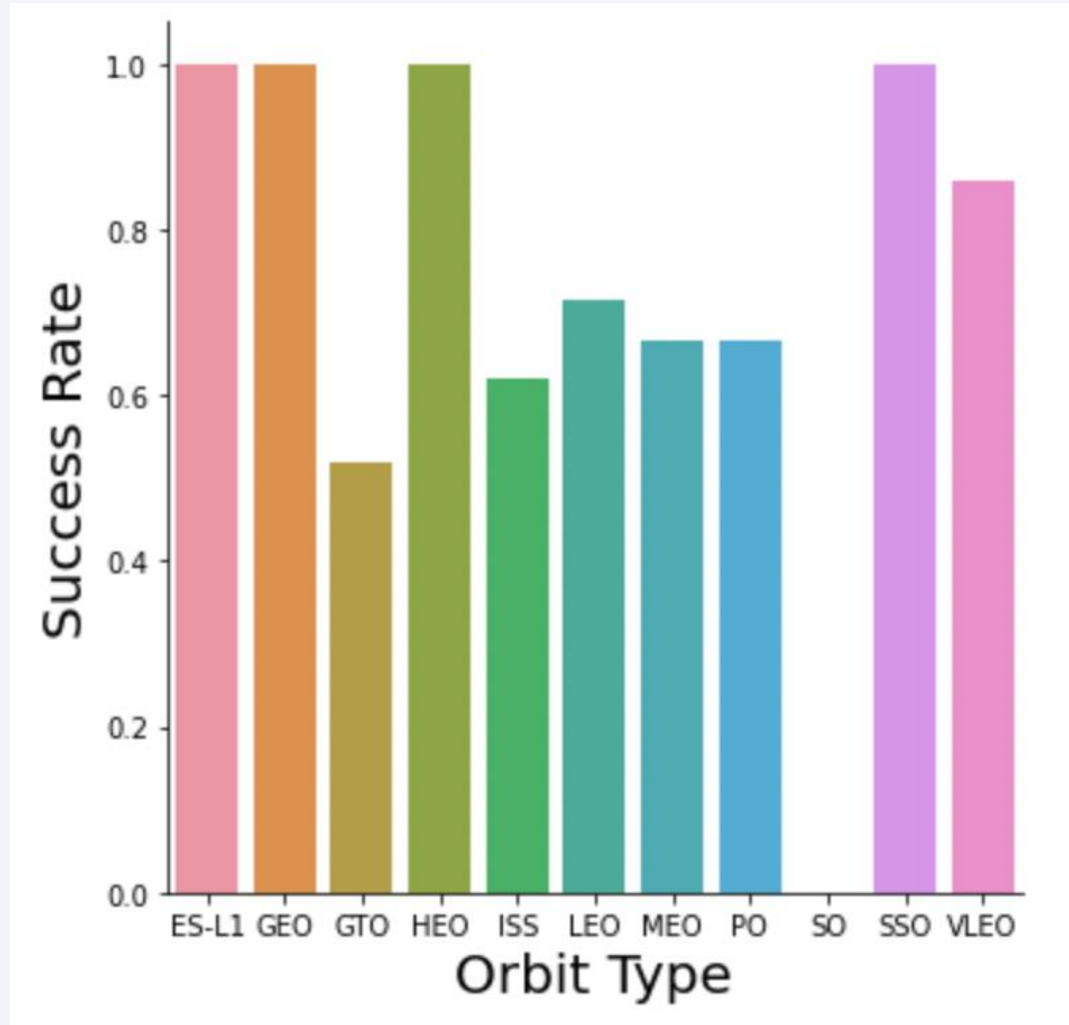


Este gráfico de dispersão mostra que a massa da carga útil é superior a 7.000 kg, e a probabilidade de sucesso da taxa será altamente aumentada. Contudo, não há um claro padrão para dizer se o lançamento site é dependente da massa de carga útil para o sucesso de retorno .

Outros pontos a observar:

- Para cada local de lançamento, quanto maior a massa da carga útil, maior será o sucesso avaliar.
- A maioria dos lançamentos com massa útil superior a 7.000 kg foram bem-sucedidos.
- O KSC LC 39A também tem uma taxa de sucesso de 100% para massa de carga útil inferior a 5.500 kg.

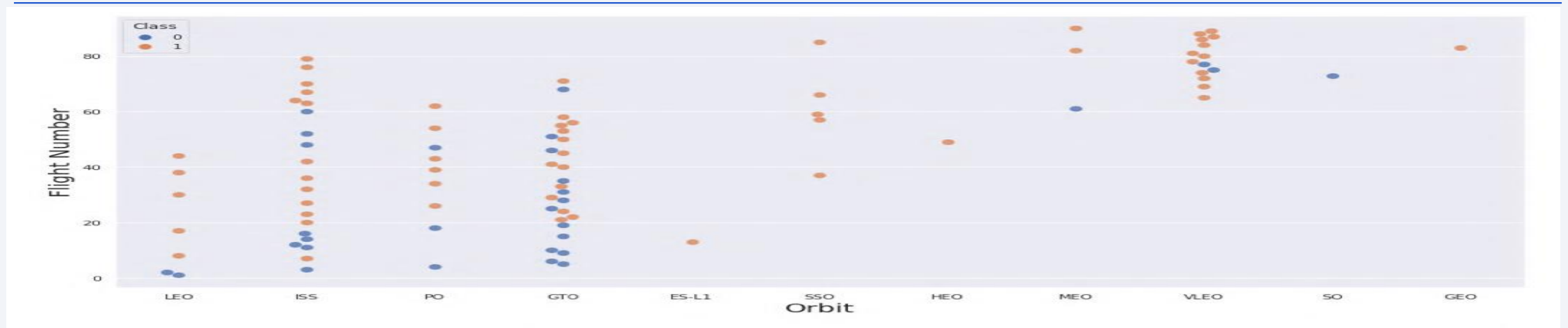
Taxa de sucesso vs Tipo de Órbita



Explicação:

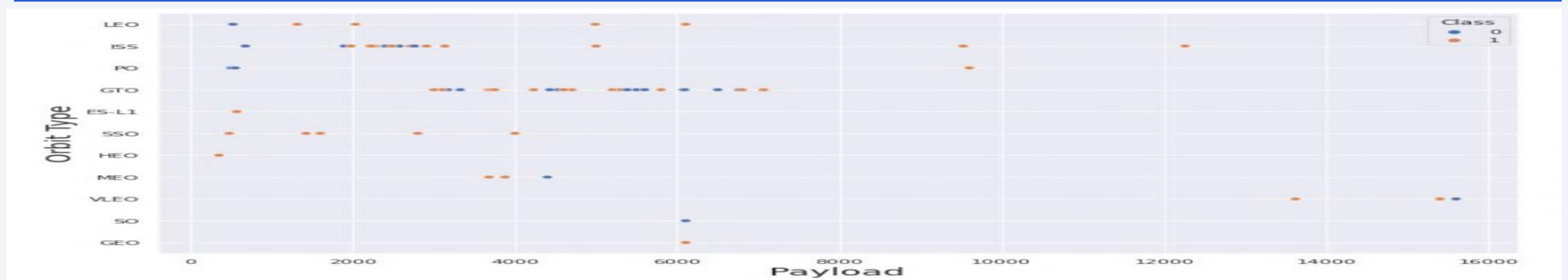
- Órbitas com taxa de sucesso de 100%:
 - ES-L1, GEO, HEO, SSO
- Órbitas com taxa de sucesso de 0%:
 - SO
- Órbitas com taxa de sucesso entre 50% e 85%:
 - GTO, ISS, LEO, MEO, PO

Números de voos vs Tipo de órbitas



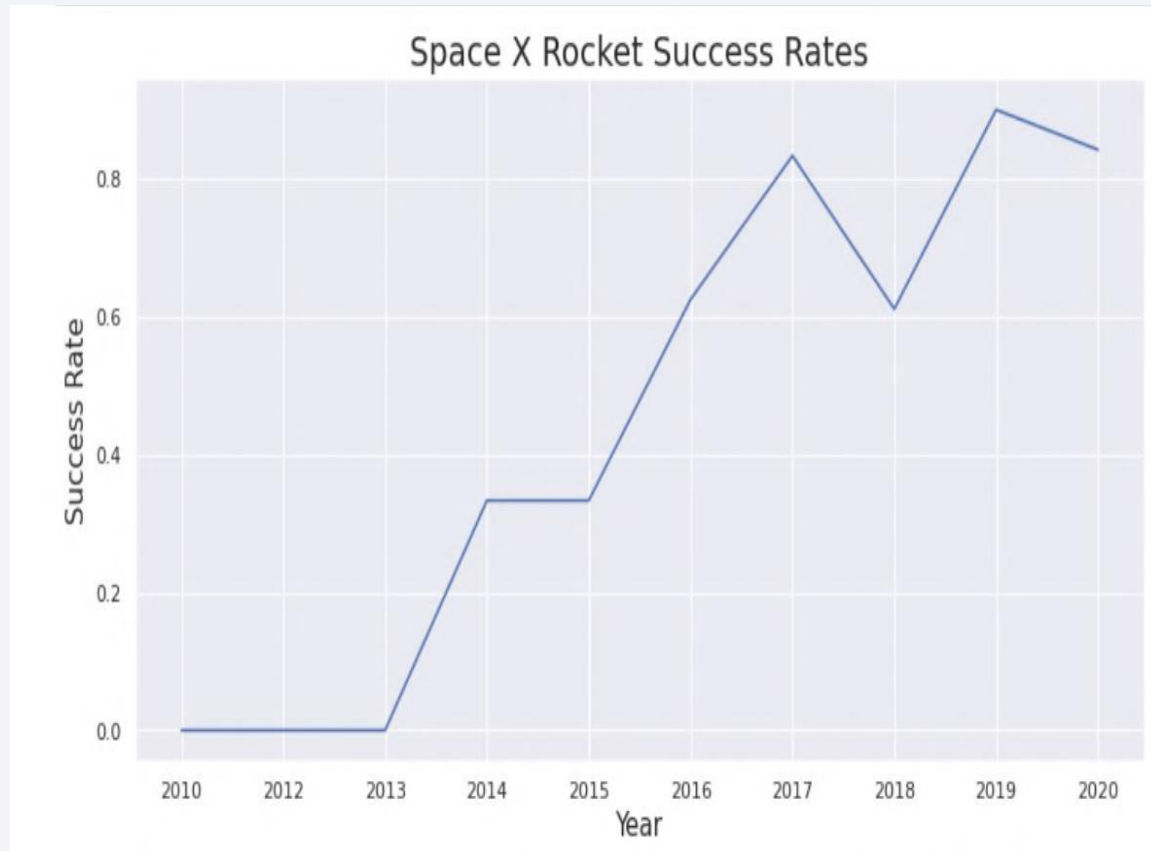
Este gráfico de dispersão mostra que geralmente, quanto maior o número de voos de cada órbita, maior a taxa de sucesso (especialmente órbita LEO), exceto para Órbita GTO que não mostrou relações entre os atributos. A Órbita que tem apenas 1 ocorrência também deveria ser excluído da relação acima pois necessita de mais dados para se chegar a uma conclusão.conforme necessário mais conjunto de dados.

Carga Útil vs Tipo de Órbita



Carga útil maior tem positivo impacto na órbita LEO, ISS e PO. No entanto, tem impacto negativo na órbita MEO e VLEO. A órbita GTO parece não representar relação entre os atributos. Enquanto isso, novamente as órbitas SO, GEO e HEO precisam de mais conjunto de dados para observarmos qualquer padrão ou tendência.

Tendência anual de lançamentos com sucesso



Esta figura claramente representada possui tendência crescente do ano de 2013 até 2020. Se esta tendência continuar durante o ano que vem em diante, a taxa de sucesso será constante e irá aumentar até atingir Taxa de sucesso de 1/100%. Apenas em 2017 ate 2018 que houve uma pequena queda, logo recuperando em seguida.

Todos os nomes das plataformas de lançamento

In [5]:

```
%sql SELECT DISTINCT LAUNCH_SITE as "Launch_Sites" FROM SPACEX;
```

```
* ibm_db_sa://zpw86771:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.clogj3  
sd0tgtu0lqde00.databases.appdomain.cloud:32731/bludb  
Done.
```

Out[5]: **Launch_Sites**

CCAFS LC-40

CCAFS SLC-40

KSC LC-39A

VAFB SLC-4E

Usamos a palavra-chave DISTINCT para mostrar apenas locais de lançamento exclusivos a partir dos dados da SpaceX.

Nomes de plataformas de lançamento que começam com 'CCA'

Usamos a seguinte consulta ao lado para
exibir 5 registros onde os sites de
lançamento comecem com `CCA`

Display 5 records where launch sites begin with the string 'CCA'

```
In [11]: task_2 = '''
        SELECT *
        FROM SpaceX
        WHERE LaunchSite LIKE 'CCA%'
        LIMIT 5
        '''
        create_pandas_df(task_2, database=conn)
```

Out[11]:

| | date | time | boosterversion | launchsite | payload | payloadmasskg | orbit | customer | missionoutcome | landingoutcome |
|---|------------|----------|----------------|-------------|---|---------------|-----------|-----------------|----------------|---------------------|
| 0 | 2010-04-06 | 18:45:00 | F9 v1.0 B0003 | CCAFS LC-40 | Dragon Spacecraft Qualification Unit | 0 | LEO | SpaceX | Success | Failure (parachute) |
| 1 | 2010-08-12 | 15:43:00 | F9 v1.0 B0004 | CCAFS LC-40 | Dragon demo flight C1, two CubeSats, barrel of... | 0 | LEO (ISS) | NASA (COTS) NRO | Success | Failure (parachute) |
| 2 | 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 |
| 3 | 2012-08-10 | 00:35:00 | F9 v1.0 B0006 | CCAFS LC-40 | SpaceX CRS-1 | 500 | LEO (ISS) | NASA (CRS) | Success | No attempt |
| 4 | 2013-01-03 | 15:10:00 | F9 v1.0 B0007 | CCAFS LC-40 | SpaceX CRS-2 | 677 | LEO (ISS) | NASA (CRS) | Success | No attempt |

Massa de carga útil total

Display the total payload mass carried by boosters launched by NASA (CRS)

```
%sql SELECT SUM(PAYLOAD_MASS__KG_) AS "Total Payload Mass by NASA (CRS)"
```

```
* ibm_db_sa://zpw86771:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.clogj3  
sd0tgtu0lqde00.databases.appdomain.cloud:32731/bludb  
Done.
```

Total Payload Mass by NASA (CRS)

45596

Calculamos a carga útil total transportada pelos boosters da NASA com resultado de 45596 usando a consulta acima.

Média de carga útil pelo F9 v1.1

Display average payload mass carried by booster version F9 v1.1

```
%sql SELECT AVG(PAYLOAD_MASS_KG_) AS "Average Payload Mass by Booster  
WHERE BOOSTER_VERSION = 'F9 v1.1';
```

```
* ibm_db_sa://zpw86771:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.clogj3  
sd0tgtu01qde00.databases.appdomain.cloud:32731/bludb
```

Done.

Average Payload Mass by Booster Version F9 v1.1

2928

Calculamos a massa média da carga transportada pela versão booster F9 v1.1 com o resultado de 2928.

Primeira data de pouso terrestre bem-sucedida

```
%sql SELECT MIN(DATE) AS "First Succesful Landing Outcome in Ground Pad  
WHERE LANDING__OUTCOME = 'Success (ground pad)';
```

```
* ibm_db_sa://zpw86771:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.clogj3  
sd0tgtu01qde00.databases.appdomain.cloud:32731/bludb  
Done.
```

First Succesful Landing Outcome in Ground Pad

2015-12-22

Usamos a função min() para encontrar o resultado. Observamos que a data do primeiro resultado de pouso bem-sucedido em terra foi em 22 de dezembro de 2015.

Pouso bem-sucedido de navio drone com carga útil entre 4.000 e 6.000

```
%sql SELECT BOOSTER_VERSION FROM SPACEX WHERE LANDING__OUTCOME = 'Success (drone ship)' \
AND PAYLOAD_MASS__KG_ > 4000 AND PAYLOAD_MASS__KG_ < 6000;
```

```
* ibm_db_sa://zpw86771:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.clogj3sd0tgtu0lqde00.datab
ases.appdomain.cloud:32731/bludb
Done.
```

booster_version

F9 FT B1022

F9 FT B1026

F9 FT B1021.2

F9 FT B1031.2

Foi usado a cláusula WHERE para filtrar boosters que pousaram com sucesso em navio drone e foi aplicado a condição AND para determinar o pouso bem-sucedido com massa da carga útil maior que 4.000, porem menor que 6.000 e encontrado os valores acima.

Número total de resultados de missões bem-sucedidas e fracassadas

List the total number of successful and failure mission outcomes

```
%sql SELECT COUNT(MISSION_OUTCOME) AS "Successful Mission" FROM SPACEX WHERE MISSION_OUTCOME LIKE 'Success%';
```

```
* ibm_db_sa://zpw86771:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.clogj3sd0tgtu0lqde00.databases.appdomain.cloud:32731/bludb
Done.
```

| Successful Mission |
|--------------------|
|--------------------|

| |
|-----|
| 100 |
|-----|

```
%sql SELECT COUNT(MISSION_OUTCOME) AS "Failure Mission" FROM SPACEX WHERE MISSION_OUTCOME LIKE 'Failure%';
```

```
* ibm_db_sa://zpw86771:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.clogj3sd0tgtu0lqde00.databases.appdomain.cloud:32731/bludb
Done.
```

| Failure Mission |
|-----------------|
|-----------------|

| |
|---|
| 1 |
|---|

Foi usado caracteres como '%' para filtrar ONDE MissionOutcome foi um sucesso ou um fracasso.

Transporte de carga útil máxima pelos boosters

```
%sql SELECT DISTINCT BOOSTER_VERSION AS "Booster Versions which carried the Maximum Payload Mass" FROM SPACEX  
WHERE PAYLOAD_MASS_KG_ =(SELECT MAX(PAYLOAD_MASS_KG_) FROM SPACEX);
```

```
* ibm_db_sa://zpw86771:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.clogj3sd0tgtu0lqde00.databases.appdomain.clou  
d:32731/bludb
```

Done.

Booster Versions which carried the Maximum Payload Mass

F9 B5 B1048.4

F9 B5 B1048.5

F9 B5 B1049.4

F9 B5 B1049.5

F9 B5 B1049.7

F9 B5 B1051.3

F9 B5 B1051.4

F9 B5 B1051.6

F9 B5 B1056.4

F9 B5 B1058.3

F9 B5 B1060.2

F9 B5 B1060.3

Foi determinado o reforço que tem levou a máxima carga útil usando um subconsulta no Cláusula WHERE e a função MAX().

Recordes de lançamento de 2015

```
%sql SELECT BOOSTER_VERSION, LAUNCH_SITE FROM SPACEX WHERE DATE LIKE '2015-%' AND \
LANDING__OUTCOME = 'Failure (drone ship)';
```

```
* ibm_db_sa://zpw86771:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.clogj3sd0tgtu0lqde00.
databases.appdomain.cloud:32731/bludb
Done.
```

| booster_version | launch_site |
|-----------------|-------------|
| F9 v1.1 B1012 | CCAFS LC-40 |
| F9 v1.1 B1015 | CCAFS LC-40 |

Foi usado combinações da cláusula WHERE, LIKE, AND e BETWEEN para filtrar os resultados de pouso fracassados em navios drones, sua versões boosters e nomes de sites de lançamento para o ano de 2015.

Resultados de desembarque e classificação entre 04/06/2010 e 20/03/2017

```
%sql SELECT LANDING__OUTCOME as "Landing Outcome", COUNT(LANDING__OUTCOME) AS "Total Count" FROM SPACEX \
WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20' \
GROUP BY LANDING__OUTCOME \
ORDER BY COUNT(LANDING__OUTCOME) DESC ;
```

```
* ibm_db_sa://zpw86771:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.clogj3sd0tgtu01qde00.databases.appdomain.c
loud:32731/blddb
Done.
```

| Landing Outcome | Total Count |
|------------------------|-------------|
| No attempt | 10 |
| Failure (drone ship) | 5 |
| Success (drone ship) | 5 |
| Controlled (ocean) | 3 |
| Success (ground pad) | 3 |
| Failure (parachute) | 2 |
| Uncontrolled (ocean) | 2 |
| Precluded (drone ship) | 1 |

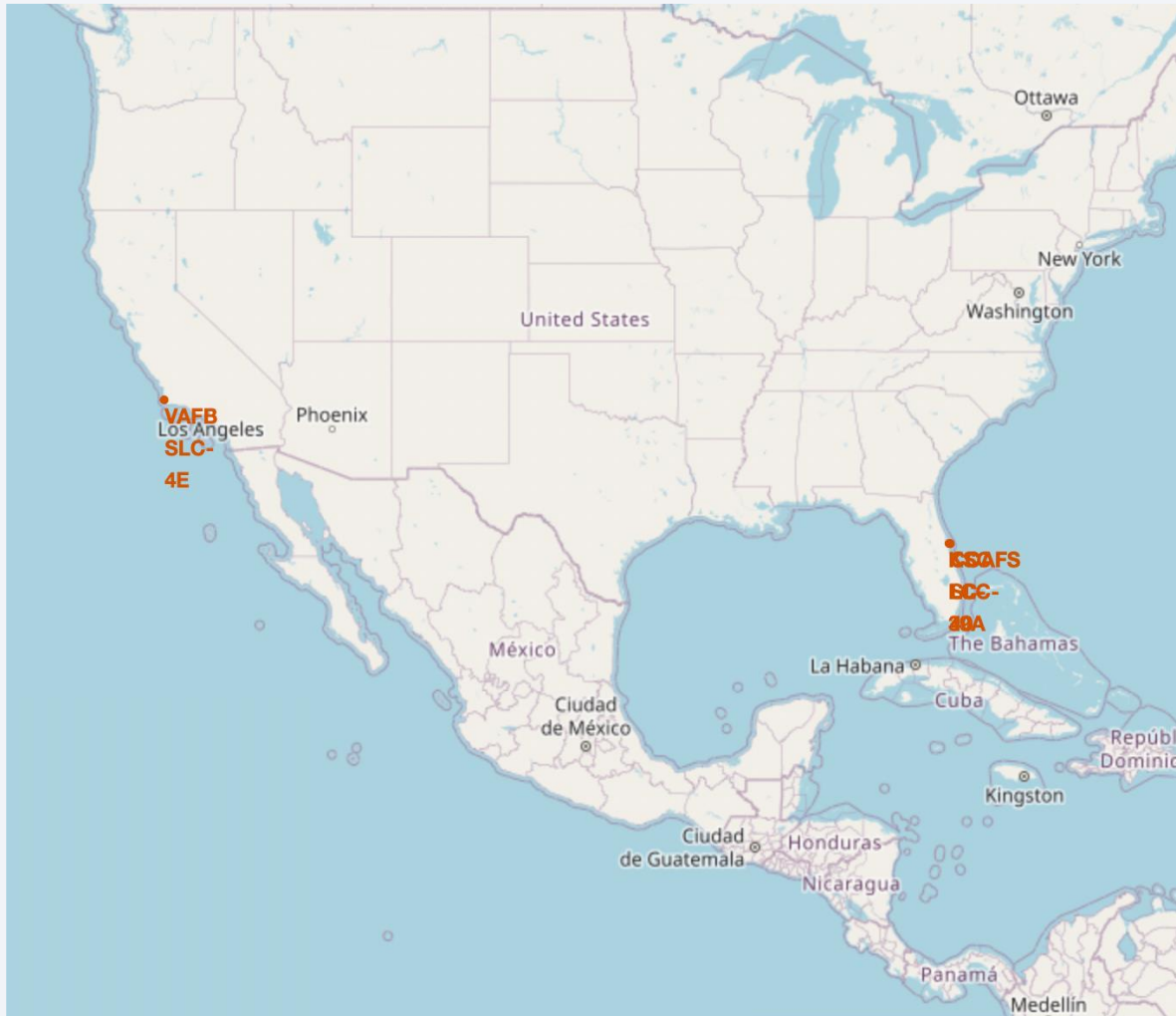
Esta figura mostra como resultou a classificação da contagem dos resultados de pouso como falha de navio drone ou sucesso da base de solo entre as datas 04/06/2010 e 20/03/2017 em ordem decrescente.

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

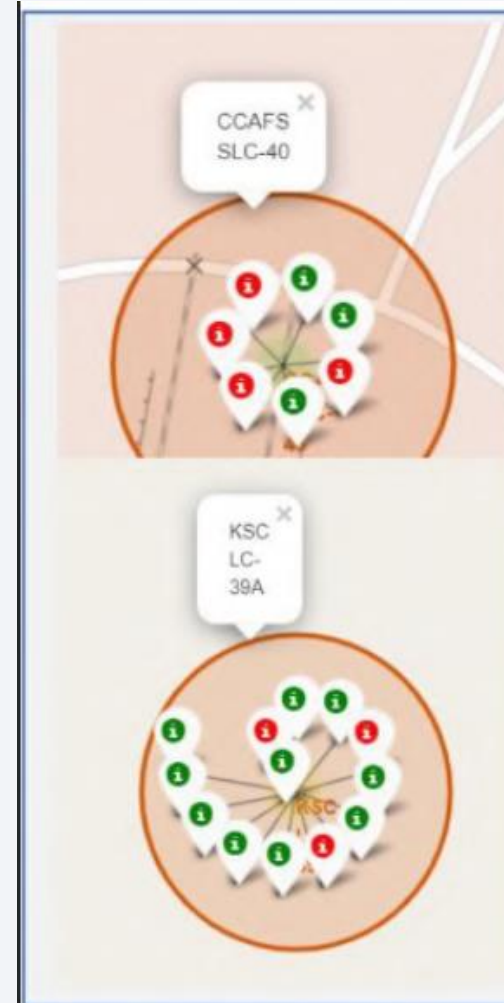
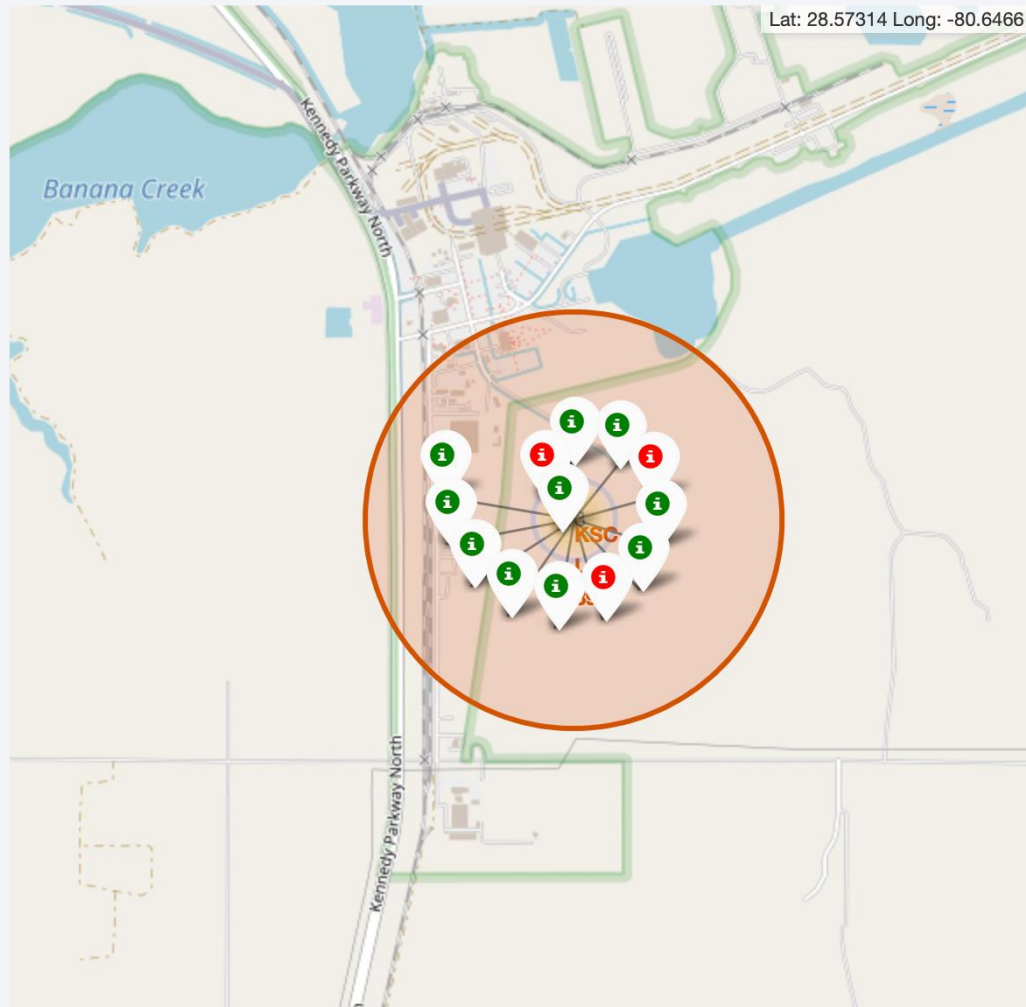
Bases de lançamento marcadas no mapa



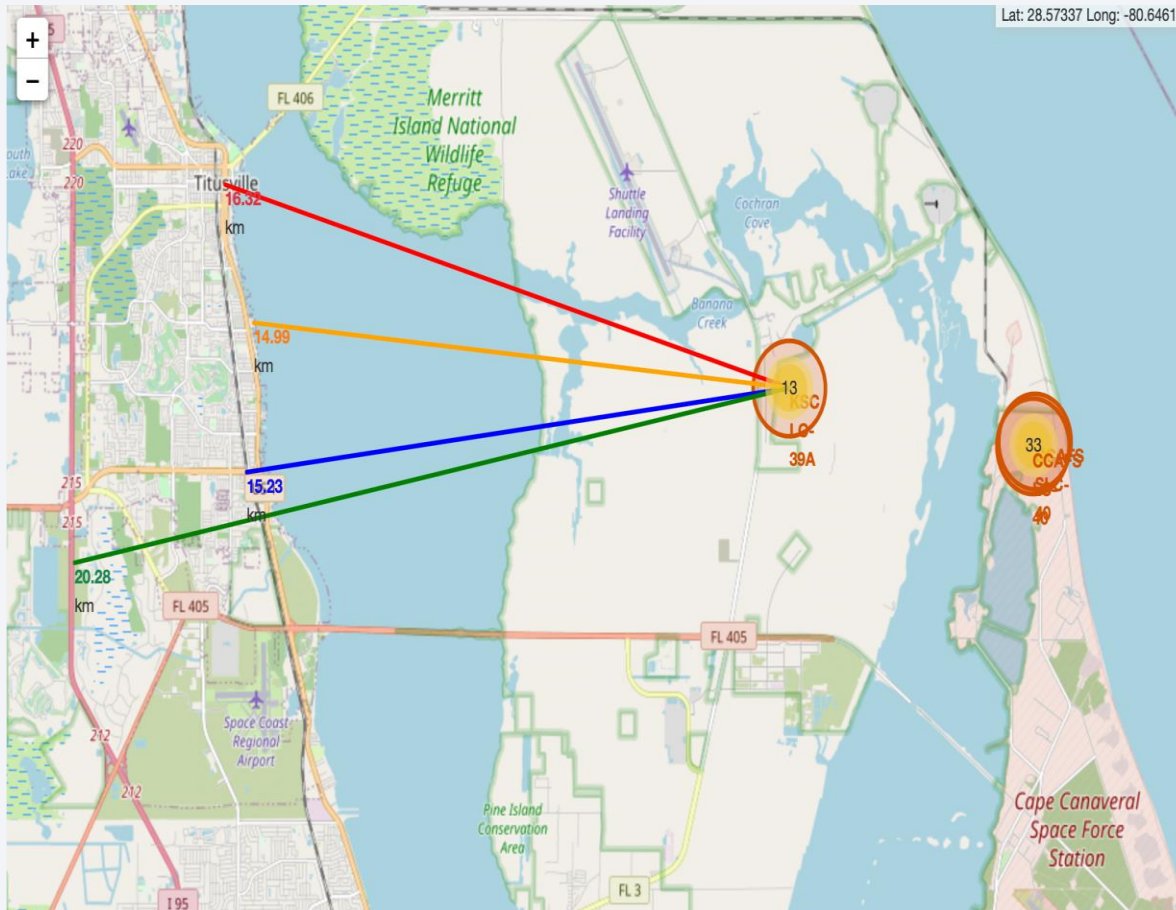
A maioria dos locais de lançamento estão próximos ao Linha do Equador. A terra está se movendo mais rápido em torno do equador do que em qualquer outro lugar da superfície da Terra. Qualquer coisa na superfície da Terra no equador se move a 1670 km/hora.

Essa velocidade ajudará a espaçonave a acompanhar uma velocidade boa o suficiente para permanecer em órbita. Todos os locais de lançamento estão muito próximos da costa, enquanto lançar foguetes em direção ao oceano, minimiza o risco de ter qualquer detrito caindo ou explodindo perto das pessoas.

Marcadores mostrando locais de lançamento com rótulos coloridos



Distância dos locais de lançamento até os pontos de referência



Da análise visual do lançamento site KSC LC-39A podemos ver claramente que isso é:

- relativamente próximo da ferrovia (15,23 km)
- relativamente próximo à rodovia (20,28 km)
- relativamente próximo da costa (14,99 km)
- Além disso, o local de lançamento KSC LC-39A é relativamente perto de sua cidade mais próxima Titusville (16,32 km).
- Foguete falhado com sua lata de alta velocidade cobrir distâncias como 15-20 km em poucos segundos. Poderia ser potencialmente perigoso para áreas povoadas.



Section 4

Build a Dashboard with Plotly Dash

Contagem de sucesso de lançamento para todos os sites

Total Success Launches by Site



O gráfico mostra claramente que de todos os locais, o KSC LC-39A tem o maior aproveitamento de lançamentos com sucesso.

A maior taxa de sucesso de lançamento: KSC LC-39A

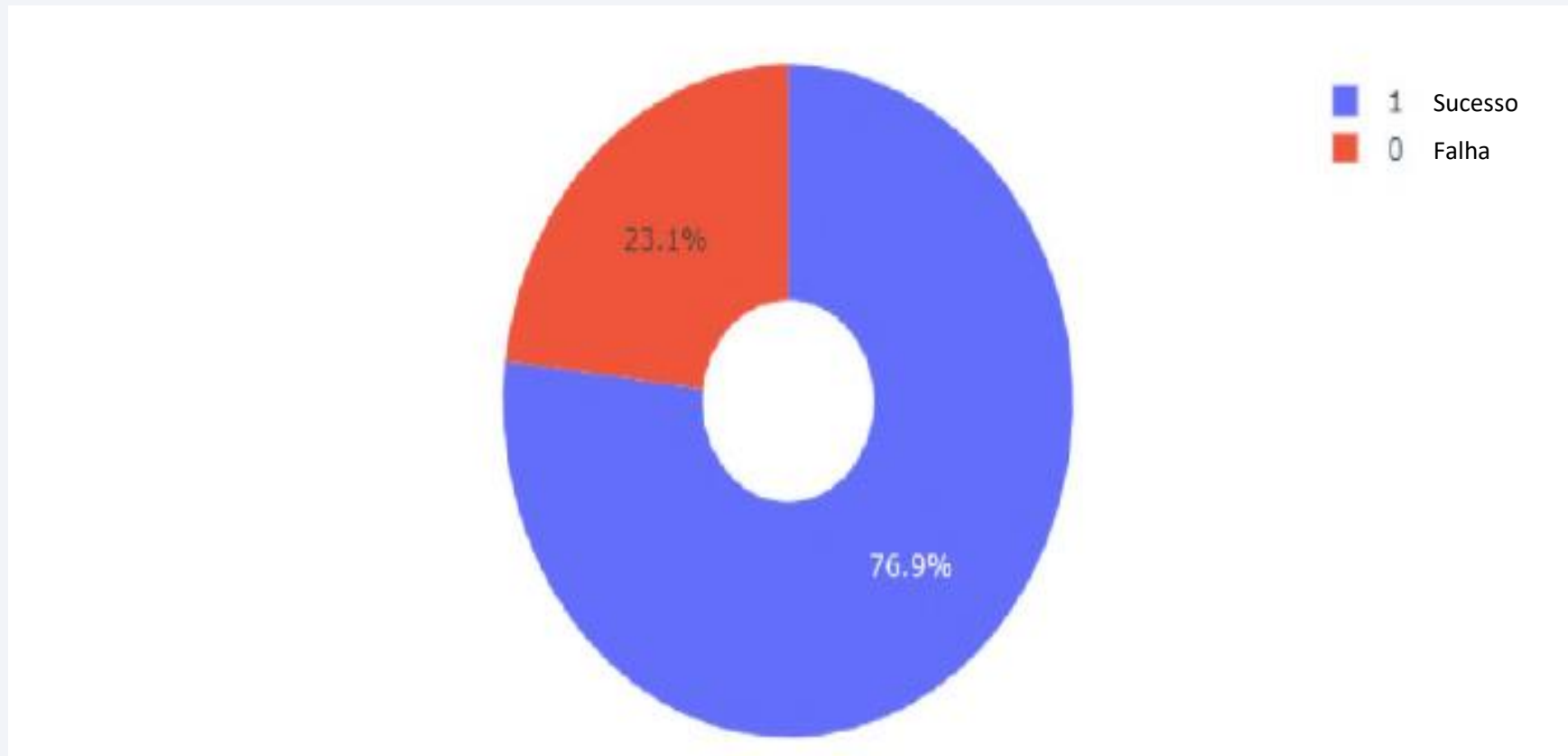
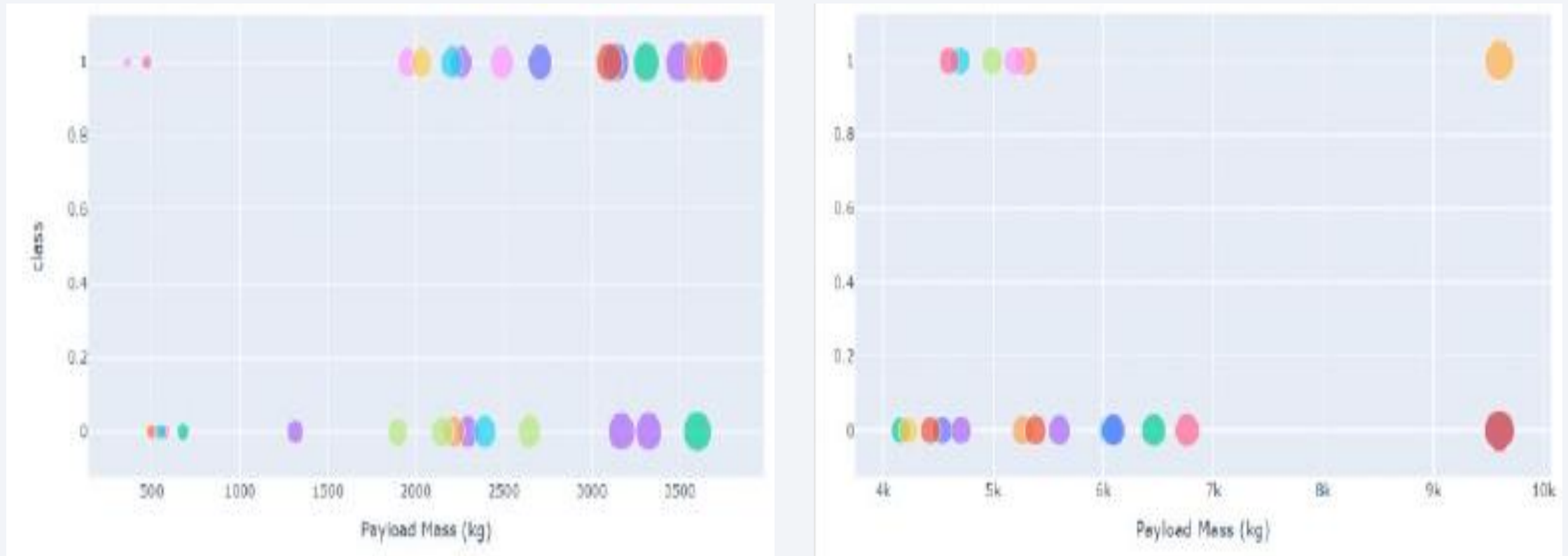


Gráfico de dispersão de carga útil vs resultado de lançamento



Podemos ver que a taxa de sucesso para cargas úteis de baixo peso é maior do que para cargas úteis de alto peso.

Section 5

Predictive Analysis (Classification)

Precisão de Classificação

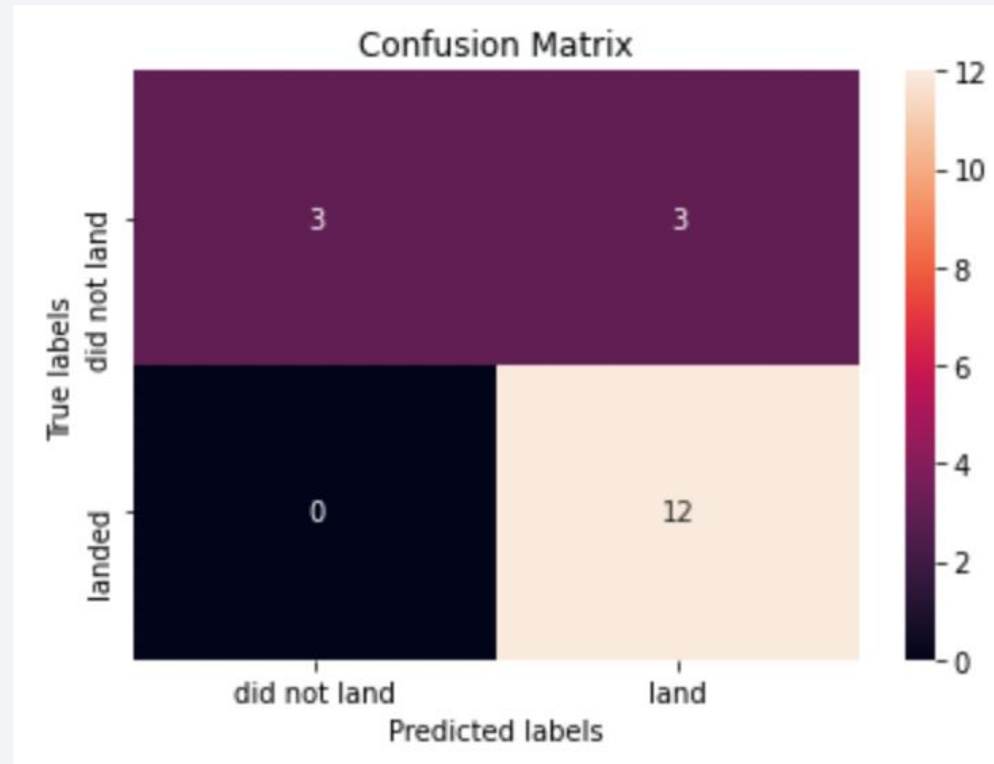
```
algorithms = {'KNN':knn_cv.best_score_, 'Tree':tree_cv.best_score_, 'LogisticRegression':logreg_cv.best_score_}
bestalgorithm = max(algorithms, key=algorithms.get)
print('Best Algorithm is',bestalgorithm,'with a score of',algorithms[bestalgorithm])
if bestalgorithm == 'Tree':
    print('Best Params is :',tree_cv.best_params_)
if bestalgorithm == 'KNN':
    print('Best Params is :',knn_cv.best_params_)
if bestalgorithm == 'LogisticRegression':
    print('Best Params is :',logreg_cv.best_params_)
```

Best Algorithm is Tree with a score of 0.9017857142857142

Best Params is : {'criterion': 'entropy', 'max_depth': 10, 'max_features': 'auto', 'min_samples_leaf': 2, 'min_samples_split': 10, 'splitter': 'random'}

Como podemos ver, usando o código acima podemos identificar que o melhor algoritmo a ser utilizado é o Algoritmo de Árvore , que possui a maior precisão de classificação.

Matriz de confusão



Matriz de confusão

| | | Predicted Values | |
|---------------|----------|------------------|-----------|
| | | Negative | Positive |
| Actual Values | Negative | TN | FP |
| | Positive | FN | TP |

A matriz de confusão para o classificador de árvore de decisão mostra que o classificador pode distinguir entre as diferentes classes. O maior problema são os falsos positivos, ou seja, o pouso malsucedido marcado como pouso bem-sucedido pelo classificador.

Conclusões

- O Modelo de Árvore de Decisão é o melhor algoritmo para este conjunto de dados.
- Lançamentos com baixa massa de carga útil apresentam melhores resultados do que lançamentos com maior massa de carga útil.
- A maioria dos locais de lançamento está próxima da linha do Equador e todos os locais estão muito próximos da costa.
- A taxa de sucesso dos lançamentos aumenta ao longo dos anos.
- KSC LC-39A tem a maior taxa de sucesso dos lançamentos de todos os sites.
- As órbitas ES-L1, GEO, HEO e SSO têm 100% de taxa de sucesso.

Muito obrigado !