

ANÁLISE DE QUEDAS DE METEORITOS NA TERRA: AQUISIÇÃO E PREPARAÇÃO DE DADOS

Sumário

RESUMO.....	2
1. INTRODUÇÃO.....	4
2. DESCRIÇÃO DA BASE DE DADOS.....	5
2.1 Fonte e Características Gerais.....	5
2.2 Estrutura e Atributos.....	5
2.3 Análise Inicial dos Dados.....	6
3. METODOLOGIA.....	7
3.1 Pipeline de Preparação de Dados.....	7
3.2 Implementação do Pipeline.....	7
3.2.1 Limpeza Inicial e Remoção de Valores Ausentes.....	7
3.2.2 Remoção de Duplicatas.....	8
3.2.3 Filtragem Temporal.....	8
3.2.4 Geocodificação Reversa.....	8
3.2.5 Categorização e Criação de Novas Features.....	9
4. RESULTADOS E DISCUSSÃO.....	10
4.1 Distribuição Geográfica dos Meteoritos.....	10
4.2 Distribuição por País.....	10
4.3 Média de Massa por País.....	11
4.4 Distribuição Temporal.....	12
5. CONCLUSÕES.....	13
5.1 Principais Resultados.....	13
5.2 Limitações do Estudo.....	13
5.3 Implicações e Trabalhos Futuros.....	13

RESUMO

Este artigo apresenta uma análise abrangente de dados sobre meteoritos que caíram na Terra no período de 1963 a 2013. Utilizando técnicas de aquisição, limpeza e preparação de dados, foram processados registros da base de dados sobre quedas de meteoritos. O estudo identificou padrões geográficos, temporais e de distribuição de massa dos meteoritos, revelando concentrações significativas em regiões específicas do planeta e variações notáveis na frequência de quedas ao longo do tempo. A análise resultou em uma base de dados limpa contendo 36.188 registros com localização geográfica identificada, proporcionando insights valiosos sobre a distribuição global de meteoritos. Os resultados destacam a importância da preparação adequada dos dados para análises científicas confiáveis e podem contribuir para estudos futuros em astronomia, geologia e ciências planetárias.

Palavras-chave: meteoritos, análise de dados, geocodificação, distribuição geográfica, preparação de dados

1. INTRODUÇÃO

Os meteoritos representam uma das poucas oportunidades que temos de estudar material extraterrestre sem a necessidade de missões espaciais. Estes fragmentos de asteroides, cometas ou até mesmo de outros planetas fornecem informações valiosas sobre a formação e evolução do sistema solar. Registrar e analisar as quedas de meteoritos na Terra é, portanto, uma atividade científica de grande importância.

A NASA mantém uma base de dados abrangente sobre quedas de meteoritos, registrando informações como localização, massa, ano de queda ou descoberta, e classificação. No entanto, como muitas bases de dados do mundo real, esta apresenta desafios significativos para análise, incluindo valores ausentes, outliers e inconsistências que precisam ser tratados adequadamente antes que qualquer análise significativa possa ser realizada.

Este estudo tem como objetivo aplicar técnicas de aquisição e preparação de dados para analisar os padrões de quedas de meteoritos na Terra entre 1963 e 2013. Através de um pipeline estruturado de limpeza e transformação de dados, buscamos responder questões sobre a distribuição geográfica, temporal e de massa dos meteoritos, fornecendo insights que podem ser valiosos para pesquisadores em diversas áreas.

2. DESCRIÇÃO DA BASE DE DADOS

2.1 Fonte e Características Gerais

A base de dados utilizada neste estudo foi obtida do repositório público da NASA sobre quedas de meteoritos, disponível no Kaggle (<https://www.kaggle.com/datasets/nasa/meteorite-landings>). O conjunto de dados original contém registros detalhados de meteoritos que caíram na Terra ao longo da história.

Características da base original:

- **Número de registros:** 45.716
- **Número de variáveis:** 10
- **Período temporal:** 301 a 2501 (com alguns valores possivelmente errôneos)
- **Formato:** CSV (Comma-Separated Values)

2.2 Estrutura e Atributos

A base de dados contém os seguintes atributos:

Atributo	Descrição	Tipo de Dado
name	Nome do meteorito	Texto
id	Identificador único	Inteiro
nametype	Tipo de classificação do nome (Valid/Relict)	Texto
recclass	Classe de classificação do meteorito	Texto
mass	Massa em gramas	Numérico
fall	Se foi observado caindo (Fell) ou encontrado (Found)	Texto
year	Ano da queda/descoberta	Numérico
reclat	Latitude da recuperação	Numérico
reclong	Longitude da recuperação	Numérico
GeoLocation	Coordenadas geográficas (latitude, longitude)	Texto

2.3 Análise Inicial dos Dados

Uma análise exploratória inicial revelou algumas características importantes da base de dados:

- **Valores ausentes:** Foram identificados valores ausentes em várias colunas, principalmente nas coordenadas geográficas (16% dos registros sem latitude/longitude).
- **Distribuição temporal:** Os registros abrangem um período muito amplo, com alguns anos potencialmente errôneos (301 a 2501).
- **Observação vs. Descoberta:** A grande maioria dos meteoritos foi encontrada após a queda (Found), enquanto apenas uma pequena parte foi observada caindo (Fell).

A análise estatística descritiva revelou uma grande variação na massa dos meteoritos, desde fragmentos muito pequenos até exemplares extremamente grandes, com uma média de 13.278,08 gramas e um desvio padrão de 574.988,9 gramas, indicando uma distribuição altamente assimétrica.

3. METODOLOGIA

3.1 Pipeline de Preparação de Dados

Para transformar os dados brutos em uma base analítica de qualidade, foi desenvolvido um pipeline estruturado de preparação de dados, composto pelas seguintes fases:

1. Limpeza inicial e remoção de valores ausentes
2. Remoção de duplicatas
3. Filtragem temporal - 50 Anos de amostra (1963-2013)
4. Geocodificação reversa - Conversão de lat/long para países
5. Categorização e criação de novas features

**OBS Não foi realizada a remoção de outliers, natureza física dos meteoritos, que variam naturalmente de fragmentos microscópicos a massas de várias toneladas, os valores extremos observados representam variação natural do fenômeno e foram mantidos na análise.*

O diagrama abaixo ilustra o fluxo do pipeline implementado:

Unset

```
graph TD
    A[Dados Brutos  
↓ (45.5716 registros)] --> B[Remoção de Valores Ausentes  
↓ (38.116 registros)]
    B --> C[Remoção de Duplicatas  
↓ (38.116 registros)]
    C --> D[Filtragem Temporal  
↓ (36.188 registros)]
    D --> E[Geocodificação Reversa  
↓ (36.188 registros)]
    E --> F[Categorização de Massa  
↓ (36.188 registros)]
    F --> G[Base Final  
(36.188 registros)]
```

3.2 Implementação do Pipeline

3.2.1 Limpeza Inicial e Remoção de Valores Ausentes

O primeiro passo consistiu na remoção de registros com valores ausentes nas colunas essenciais para a análise: coordenadas geográficas (reclat, reclong), ano (year) e massa (mass). Esta etapa é crucial para garantir que apenas registros completos sejam considerados nas análises subsequentes.

Python

```
df_clean = df.dropna(subset=['reclat', 'reclong', 'year',  
'mass'])
```

Após esta etapa, o número de registros foi reduzido para 38.116.

3.2.2 Remoção de Duplicatas

Embora a análise inicial não tenha identificado registros completamente duplicados, esta etapa foi mantida como boa prática para garantir a integridade dos dados.

Python

```
df_clean = df_clean.drop_duplicates()
```

O número de registros permaneceu em 38.116, confirmando a ausência de duplicatas.

3.2.3 Filtragem Temporal

Para focar a análise em um período mais recente e confiável, foram selecionados apenas os registros dos últimos 50 anos até 2013, ou seja, de 1963 a 2013.

Python

```
min_year = 1963  
max_year = 2013  
df_clean = df_clean[(df_clean['year'] >= min_year) &  
(df_clean['year'] <= max_year)]
```

Após esta filtragem, o dataset foi reduzido para 36.188 registros.

3.2.4 Geocodificação Reversa

Um dos desafios mais significativos foi identificar o país correspondente a cada coordenada geográfica. Para isso, foi utilizada a biblioteca `reverse_geocoder`, que permite realizar geocodificação reversa de forma eficiente e offline.

Python

```
import reverse_geocoder as rg
coordinates = list(zip(df_clean['reclat'],
df_clean['reclong']))
results = rg.search(coordinates, mode=1, verbose=False)
```

Esta etapa permitiu identificar o país para todos os 36.188 registros, que foram mantidos na base final.

3.2.5 Categorização e Criação de Novas Features

Para facilitar a análise, foi criada uma categorização da massa dos meteoritos:

Python

```
def categorize_mass(mass):
    if mass < 10:
        return 'Pequeno'
    elif mass < 100:
        return 'Médio'
    elif mass < 1000:
        return 'Grande'
    else:
        return 'Muito Grande'

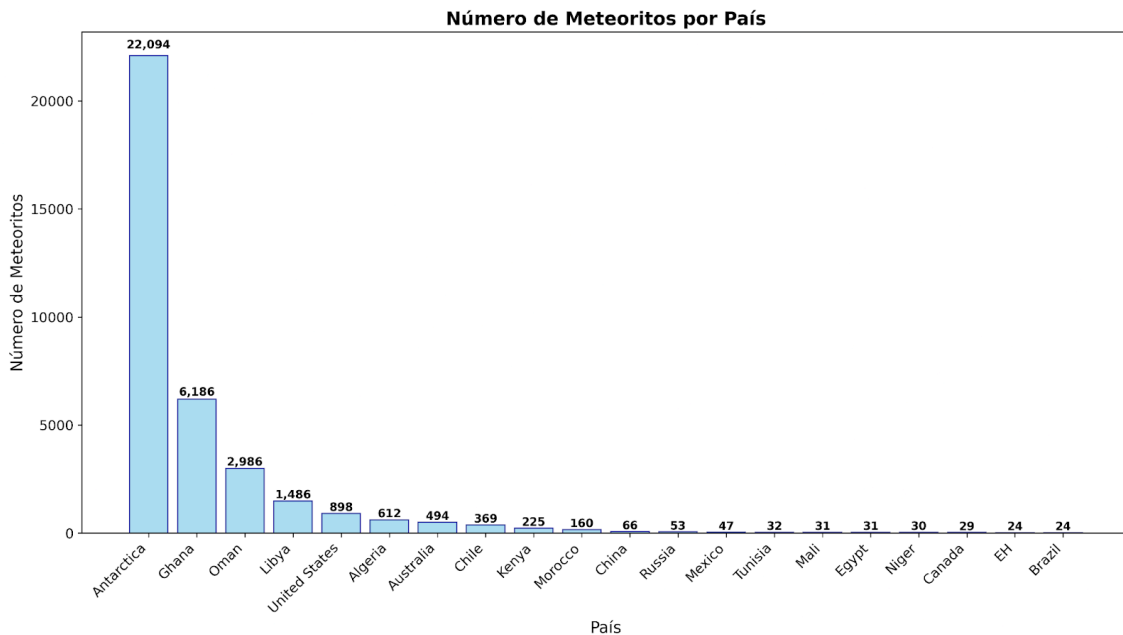
df_final['mass_category'] =
df_final['mass'].apply(categorize_mass)
```

Esta categorização permite uma análise mais intuitiva da distribuição de tamanhos dos meteoritos.

4. RESULTADOS E DISCUSSÃO

4.1 Distribuição Geográfica dos Meteoritos

A análise da distribuição geográfica dos meteoritos revelou padrões interessantes, como pode ser observado na visualização dos dados por país.



Observa-se uma concentração significativa de meteoritos em determinadas regiões do planeta, particularmente na Antártida, que apresenta um número extraordinariamente alto de meteoritos registrados (22.094). Esta distribuição não é uniforme e pode ser influenciada por diversos fatores, incluindo:

1. **Fatores geológicos e climáticos:** Regiões como a Antártida, com sua superfície de gelo, facilitam a identificação visual de meteoritos, que se destacam contra o fundo branco.
2. **Fatores humanos:** Áreas com maior atividade de pesquisa e busca por meteoritos tendem a apresentar mais registros. A Antártida é alvo de expedições científicas específicas para coleta de meteoritos.
3. **Fatores de preservação:** Condições climáticas que favorecem a preservação dos meteoritos ao longo do tempo, como o frio extremo da Antártida, contribuem para um maior número de descobertas.

4.2 Distribuição por País

A análise da distribuição de meteoritos por país revelou que algumas regiões concentram um número significativamente maior de descobertas. Os dez países com maior número de meteoritos registrados são:

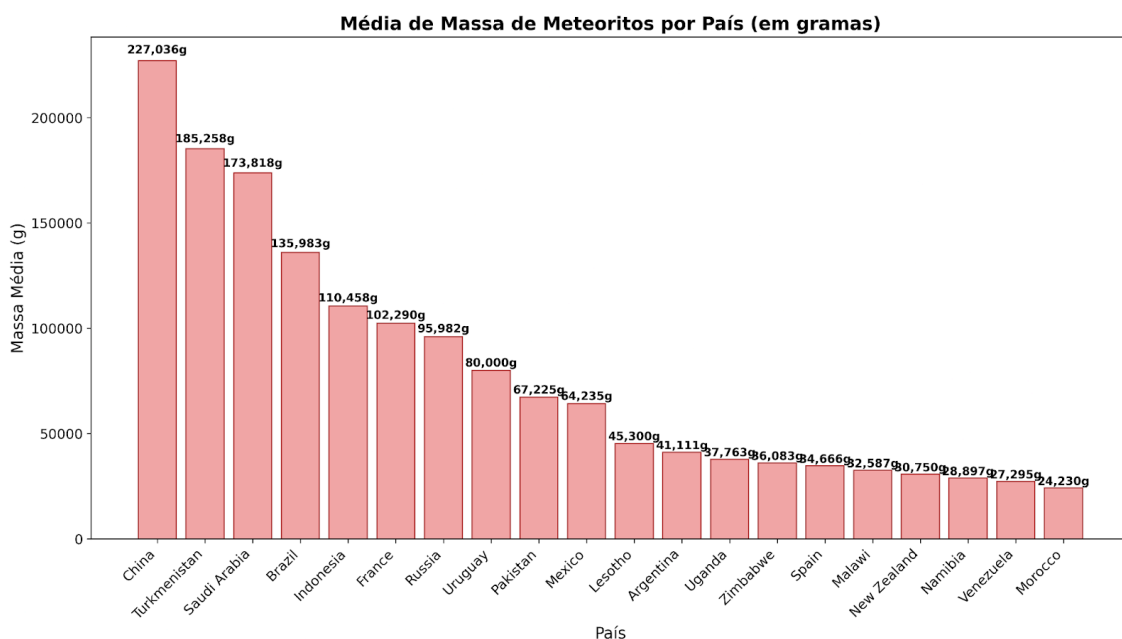
1. **Antártida:** 22.094 meteoritos

2. **Gana:** 6.186 meteoritos
3. **Omã:** 2.986 meteoritos
4. **Líbia:** 1.486 meteoritos
5. **Estados Unidos:** 898 meteoritos
6. **Argélia:** 612 meteoritos
7. **Austrália:** 494 meteoritos
8. **Chile:** 369 meteoritos
9. **Quênia:** 225 meteoritos
10. **Marrocos:** 160 meteoritos

Esta distribuição desigual pode ser explicada por diversos fatores, incluindo condições geográficas favoráveis à preservação e identificação de meteoritos, bem como a presença de programas ativos de busca e catalogação em determinadas regiões.

4.3 Média de Massa por País

A análise da média de massa dos meteoritos por país revela informações interessantes sobre a distribuição de tamanhos em diferentes regiões.



Os países com maior massa média de meteoritos são:

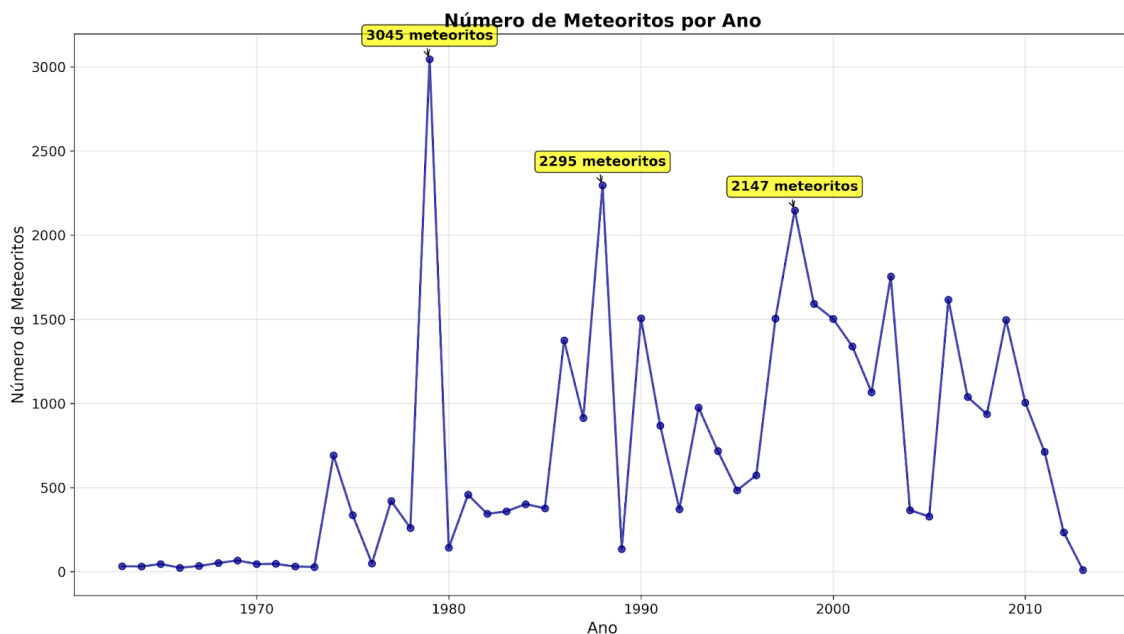
1. **China:** 227,036g
2. **Turcomenistão:** 185,258g
3. **Arábia Saudita:** 173,818g
4. **Brasil:** 135,983g
5. **Indonésia:** 110,458g

É interessante notar que os países com maior número de meteoritos não são necessariamente os mesmos que apresentam as maiores massas médias. Isso sugere

que diferentes regiões podem ter características distintas em termos do tipo e tamanho dos meteoritos encontrados.

4.4 Distribuição Temporal

A análise da distribuição temporal dos meteoritos revela variações significativas no número de registros ao longo dos anos.



Observam-se picos notáveis em determinados anos:

- **1979:** 3.045 meteoritos
- **1988:** 2.295 meteoritos
- **1998:** 2.147 meteoritos
- **2003:** 1.754 meteoritos
- **2006:** 1.616 meteoritos

Estes picos podem estar relacionados a:

1. **Eventos astronômicos específicos:** Como chuvas de meteoros excepcionais.
2. **Intensificação de esforços de pesquisa:** Expedições organizadas ou programas de busca em anos específicos.
3. **Melhorias nas técnicas de detecção:** Avanços tecnológicos que facilitaram a identificação de meteoritos.
4. **Fatores de registro:** Possíveis atualizações em massa no banco de dados em anos específicos.

É interessante notar também o aumento geral na frequência de registros a partir da década de 1970, o que pode refletir um interesse crescente no estudo de meteoritos e melhorias nas técnicas de detecção e catalogação.

5. CONCLUSÕES

5.1 Principais Resultados

Este estudo demonstrou a importância da preparação adequada de dados para análises científicas confiáveis. Através de um pipeline estruturado de limpeza e transformação, foi possível converter uma base de dados bruta com diversos problemas em uma base analítica de qualidade, contendo 36.188 registros de meteoritos com localização geográfica identificada.

Os principais resultados obtidos foram:

1. **Distribuição geográfica não uniforme:** Os meteoritos apresentam concentrações significativas em determinadas regiões do planeta, com destaque para a Antártida, que concentra mais de 60% dos registros.
2. **Concentração por país:** Algumas regiões, como Antártida, Gana e Omã, concentram um número desproporcionalmente alto de registros, o que pode estar relacionado tanto a fatores naturais quanto a esforços de pesquisa.
3. **Variação na massa média:** Países como China, Turcomenistão e Arábia Saudita apresentam meteoritos com massa média significativamente maior que outras regiões, sugerindo diferenças nos tipos de meteoritos encontrados.
4. **Variação temporal significativa:** O número de registros apresenta picos notáveis em determinados anos (1979, 1988, 1998), com um aumento geral na frequência a partir da década de 1970.

5.2 Limitações do Estudo

É importante reconhecer algumas limitações deste estudo:

1. **Viés de detecção:** Meteoritos maiores e em regiões de fácil acesso têm maior probabilidade de serem encontrados e registrados.
2. **Qualidade da geocodificação:** A identificação do país a partir das coordenadas pode conter imprecisões, especialmente em regiões fronteiriças ou remotas.
3. **Dados históricos:** Registros mais antigos podem ser menos precisos ou completos que os mais recentes.
4. **Representatividade:** A base de dados representa apenas os meteoritos que foram encontrados e registrados, não necessariamente todos os que caíram na Terra.

5.3 Implicações e Trabalhos Futuros

Os resultados deste estudo têm implicações importantes para diversas áreas:

1. **Astronomia e geologia:** A distribuição de meteoritos pode fornecer insights sobre padrões de queda de objetos celestes e sua interação com a atmosfera terrestre.

2. **Ciência de dados:** O pipeline desenvolvido pode servir como modelo para o tratamento de outras bases de dados geoespaciais com desafios semelhantes.
3. **Exploração espacial:** O estudo de meteoritos contribui para o entendimento da composição de corpos celestes e da história do sistema solar.

Trabalhos futuros poderiam explorar:

1. **Análise da composição química:** Relacionar a distribuição geográfica com os tipos e composições dos meteoritos.
2. **Modelagem preditiva:** Desenvolver modelos para prever