

UNIVERSIDADE FEDERAL DE SANTA CATARINA
DEPARTAMENTO DE INFORMÁTICA E ESTATÍSTICA - INE
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

CRIAÇÃO DE BASE DE DADOS DE MARATONA E ARREMESSO DE PESO

BOLSISTA: Henrique Pereira Ramos
ORIENTADOR: Carina Friedrich Dorneles

Relatório Final PIBIC.

FLORIANÓPOLIS
SANTA CATARINA - BRASIL
SETEMBRO 2023

UNIVERSIDADE FEDERAL DE SANTA CATARINA
DEPARTAMENTO DE INFORMÁTICA E ESTATÍSTICA - INE
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

RESUMO

CRIAÇÃO DE BASE DE DADOS DE MARATONA E ARREMESSO DE PESO

O trabalho realizado teve como objetivo a extração, tratamento e armazenamento de dados de competições de atletismo, mais especificamente da maratona e arremesso de peso que estavam acessíveis no site da *World Athletics* e trouxe como produto final uma base de dados contendo as competições dessas modalidades. Para cumprir com esse objetivo, são desenvolvidos *scripts* para a extração dos dados de competições do Website. Após, é feito o tratamento dos dados coletados e por fim a criação da base de dados.

Palavras-chave: *Web Scraping*, extração de dados, *Scrapy*, base de dados, atletismo.

SUMÁRIO

1	INTRODUÇÃO	3
2	CONCEITOS BÁSICOS	3
3	VISÃO GERAL DAS ATIVIDADES EXECUTADAS	3
4	DESENVOLVIMENTO DAS ATIVIDADES	5
4.1	Estudo e preparação do ambiente para desenvolvimento	5
4.2	Extração dos dados	5
4.3	Criação do <i>script SQL</i>	6
4.4	Limpeza dos dados	7
4.5	Criação da base de dados	8
4.6	Uso e características da base de dados	9
5	CONSIDERAÇÕES FINAIS	11
6	REFERENCIAS	12

1 INTRODUÇÃO

No mundo tecnológico, as organizações e empresas estão se tornando cada vez mais movidas por dados. A área do *Web Scraping* é parte primordial na construção de relatórios e *Dashboards*, que são ferramentas auxiliaadoras na tomada de decisão, *Web Scraping* é usado amplamente em muitos campos da ciência da computação, como em *Business Intelligence*, Inteligência Artificial, Ciência de dados, *Big Data*, Computação em Nuvem e Cibersegurança [6].

Já no mundo esportivo, o estudo da performance de atletas em competições é peça chave para a análise da progressão do desempenho com o passar dos anos. Portanto, o intuito desse projeto é desenvolver uma base de dados de duas modalidades, passando pela extração, onde será preciso encontrar sites com informações coerentes às desejadas para o projeto, lidar com problemas nas estruturas dos sites e criar códigos para a extração dos dados. Passará também pela limpeza e construção da base de dados, tendo que tratar as inconsistências nos dados coletados, realizar a criação da base de dados e a inserção dos dados já tratados. E ao fim trazer algumas relações nos dados utilizando ferramentas de Ciência de Dados, a fim de possibilitar análises de performance esportivas futuras [8].

2 CONCEITOS BÁSICOS

Para compreensão total do que será exposto neste relatório, é preciso conceituar alguns termos que aparecerão com frequência no texto. Primeiramente é importante trazer o conceito de *Web Scraping*, visto que essa técnica foi utilizada no projeto. *Web Scraping* é uma técnica de extração de dados não estruturados de *websites* e transformação desses dados em dados estruturados para que ele possa ser armazenado e analisado em uma base de dados [1]. Para fazer a raspagem dos dados da Web, foi utilizado o *Scrapy*, um *framework* para a raspagem de *Websites* e extração de dados estruturados [2].

Sobre as modalidades de atletismo que foram escolhidas, a maratona é uma corrida na qual os atletas correm a distância de 42 quilômetros e 195 metros e seu nome é derivado da história da corrida da cidade de Maratona até Atenas pelo soldado grego Fidípides [4].

Já o Arremesso de Peso é um esporte que consiste em que os atletas arremessam com uma mão uma bola de metal de 7,26kg para o masculino e 4kg para o feminino de dentro de um setor de diâmetro de 2,135m [5].

3 VISÃO GERAL DAS ATIVIDADES EXECUTADAS

Tendo como objetivo final a criação de uma base de dados com os resultados de competições de maratona e arremesso de peso, fez-se necessário a fragmentação das tarefas para alcançar tal objetivo.

Antes de começar a desenvolver o projeto, foi necessário separar um tempo para o estudo da ferramenta que seria utilizada para a extração, houve a preparação do ambiente para

que os trabalhos começassem.

Como pontapé inicial, foi estabelecido que o primeiro passo seria o reconhecimento de sites em que a colheita dos dados poderia ser feita. Após encontrar alguns sites compatíveis, foi constatado que o site da *World Athletics* seria o ideal para a extração dos dados desejados pois contém os dados da maioria das competições que ocorreram nos últimos 10 anos; visto que a *World Athletics* é a organização que cuida do Atletismo no âmbito internacional. Como o objetivo final do projeto era criar um *dataset* das competições de cada modalidade, foi decidido que as tabelas teriam as seguintes colunas: nome, posição, país, marca, competição, esporte e gênero.

Depois de estabelecer de onde seriam extraídos os dados das competições, foi decidido começar com as maratonas, para a extração de dados, foi utilizado o *Scrapy*, um *framework* em *python* para *Web Scraping*, foi desenvolvido uma *Spider(Bot) template* que seria utilizado para extrair os dados das páginas das competições, já que a estrutura do *HTML* era igual para a maioria das competições. Após aprimorar essa *Spider* e coletar todos os *links* das competições, foi feito o *crawling* das páginas, os dados que foram raspados foram primeiramente armazenados em arquivos CSV. Após a raspagem de todos os dados, cada competição de maratona tinha um .csv com uma tabela com os dados da competição. Esse mesmo processo foi feito para as competições de arremesso de peso, porém com adaptações no *template* para que se encaixasse nas páginas de competições de arremesso de peso.

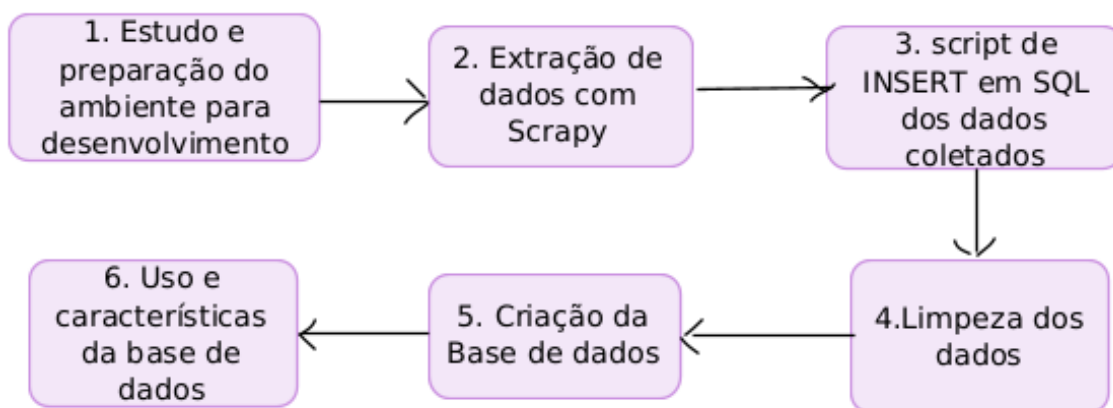
Ao final dessa etapa, foi evidenciado que foram coletadas um total de 74 competições de maratona, sendo 37 masculinos e 37 femininos, essas competições variaram entre Olimpíadas, Campeonatos mundiais de Atletismo e as *Majors*, e as competições ocorreram entre os anos 2000 e 2023. Já na de arremesso de peso, foi notado que as competições foram menos constantes, tendo um total de 60 competições de arremesso de peso coletadas, sendo 33 delas do masculino e 27 do feminino, esses 60 resultados foram espalhados entre 45 competições, com um intervalo de 2011 à 2023.

Na próxima fase, foi criado uma base de dados utilizando o *MySQL*, lá foram criadas duas tabelas, uma de maratona e outra de arremesso de peso, as colunas das tabelas consistem em: nome, posição, país, marca, competição, esporte e gênero. para pegar os dados de todos os arquivos .csv, foi desenvolvido um *script* em *Python* que tinha como intuito pegar cada dado da tabela do .csv e transformar em uma *string* de *INSERT* em *SQL*, como por exemplo: “*INSERT INTO shotput (competition, country, gender, link, mark, name, position, sport) VALUES ('Memorial Van Damme - 01–02 SEP 2022', 'NZL', 'M', '21.60', 'Tomas WALSH', '2', 'Shot Put')*”;”. Esse *script* passou por todos os arquivos .csv e armazenava as *strings* de *INSERT* em um único arquivo .sql. Depois de feito, foi realizada uma limpeza manual nos dados, para adequá-los para a inserção. Quando todos os dados foram devidamente limpos, foi feita a inserção dos dados na base de dados.

4 DESENVOLVIMENTO DAS ATIVIDADES

O Capítulo 4 traz um aprofundamento das atividades mostradas em cada item do *Pipeline* da Figura 1, respectivamente. Nele será descrito o passo a passo para o desenvolvimento do projeto, desde a preparação do ambiente de trabalho até o uso e interpretação da base de dados.

Figura 1 – *Pipeline* das tarefas realizadas



4.1 ESTUDO E PREPARAÇÃO DO AMBIENTE PARA DESENVOLVIMENTO

Para a execução do projeto, primeiramente foi necessário realizar um estudo teórico em banco de dados, estrutura de sites em *HTML*, ambientes virtuais e *Web Scraping*. Após, foi feito um ambiente virtual para a utilização do *Scrapy*, por recomendação da documentação do *Scrapy*, nesse ambiente foi feito um treinamento nesse *framework*. Depois, foi criado um novo projeto, onde foi desenvolvido o trabalho.

4.2 EXTRAÇÃO DOS DADOS

Para a extração dos dados, foi desenvolvido uma *Spider* parecida com a mostrada na Figura 2.

A lógica de funcionamento dessa *Spider*, é uma iteração que passa por todas as linhas de uma tabela na estrutura *HTML* do *Website*, e em cada linha ela extrai os dados contidos nela. quando rodamos essa *Spider* utilizando o comando `scrapy crawl dlmen14_shangai -o dlmen14_shangai.csv`, os dados da competição são armazenados em um arquivo `.csv`, como mostrado na Figura 3.

Após realizar esse processo com todas as competições selecionadas, parte-se para a terceira parte.

Figura 2 – Exemplo de *Spider*

```

import scrapy
from shotputdb.items import ShotputdbItem

...

class DLMenSpider14_shanghai(scrapy.Spider):
    name = 'dlmen14_shanghai'
    start_urls = ['https://www.worldathletics.org/competitions/diamond-league/calendar-results/7065890/result?eventId=10229619&gender=M']

    def parse(self, response):
        N = 0
        for athletes in response.css('table.styles_table__yjrQa > tbody > tr'):
            item = ShotputdbItem()
            item['name'] = athletes.css('td.EventResults_name__3UzJp> a::text').get(),
            aux = athletes.css('span.ResultsLOC_top3Val__2Dxi-::text').get()
            item['position'] = N + 1
            item['country'] = athletes.css('span.Flags_name__28uFw::text').get(),
            item['mark'] = str(response.css('tbody tr td:nth-child(5)::text')[N].getall()),
            item['competition'] = "Shanghai Golden Grand Prix - 18 MAY 2014",
            item['sport'] = "Shot Put",
            item['gender'] = "M"
            yield item
            N+=1

```

Figura 3 – Dados extraídos no CSV

```

competition,country,gender,link,mark,name,position,sport
Shanghai Golden Grand Prix - 18 MAY 2014,USA,M,,['21.73'],Christian CANTWELL,1,Shot Put
Shanghai Golden Grand Prix - 18 MAY 2014,USA,M,,['21.52'],Joe KOVACS,2,Shot Put
Shanghai Golden Grand Prix - 18 MAY 2014,USA,M,,['21.31'],Ryan WHITING,3,Shot Put
Shanghai Golden Grand Prix - 18 MAY 2014,GER,M,,['21.09'],David STORL,4,Shot Put
Shanghai Golden Grand Prix - 18 MAY 2014,POL,M,,['20.93'],Tomasz MAJEWSKI,5,Shot Put
Shanghai Golden Grand Prix - 18 MAY 2014,USA,M,,['20.60'],Kurt ROBERTS,6,Shot Put
Shanghai Golden Grand Prix - 18 MAY 2014,USA,M,,['20.11'],Cory MARTIN,7,Shot Put
Shanghai Golden Grand Prix - 18 MAY 2014,BUL,M,,['20.06'],Georgi IVANOV,8,Shot Put
Shanghai Golden Grand Prix - 18 MAY 2014,JAM,M,,['19.78'],O'Dayne RICHARDS,9,Shot Put
Shanghai Golden Grand Prix - 18 MAY 2014,CHN,M,,['19.18'],Guangfu WANG,10,Shot Put
Shanghai Golden Grand Prix - 18 MAY 2014,TPE,M,,['19.18'],Minghuang CHANG,11,Shot Put
Shanghai Golden Grand Prix - 18 MAY 2014,CHN,M,,['18.33'],Zizhong TIAN,12,Shot Put

```

4.3 CRIAÇÃO DO *SCRIPT SQL*

Depois que os dados de todas as competições foram coletados e armazenados, foi preciso criar um *script* em *Python* que cria uma *string* *sql_script* vazia e lê o arquivo CSV linha por linha, transformando os dados dessa tabela em *INSERTS* em SQL, que são incrementadas na *string* *sql_script*, no fim ocorre a escrita dessa *string* em um arquivo SQL. Esse processo ocorreu de forma recursiva afetando todos os arquivos CSV contidos na pasta. A lógica do algoritmo fica mais clara observando um trecho do *script* exposto a na Figura 4.

Figura 4 – Script que transforma os dados em *INSERT*

```

table_name = 'shotput'
field_names = ['competition', 'country', 'link', 'mark',
               'name', 'position', 'sport']
sql_script = ''
# Open the CSV file and read its contents.
for filename in os.listdir(directory):
    if filename.endswith('.csv'):
        file_path = os.path.join(directory, filename)
        with open(file_path, 'r') as csv_file:
            reader = csv.reader(csv_file)
            # Skip the header row.
            next(reader)
            for row in reader:
                sql_script += f"INSERT INTO {table_name}
                (competition, country,gender, link, mark,
                name, position, sport) VALUES ('{row[0]}', '{
                {row[1]}', '{row[2]}', '{row[3]}', '{row[4]}', '{row
                [5]}', '{row[6]}', '{row[7]}');\n"
                # Write the SQL statement to the SQL file.
with open('insert_data.sql', 'w') as sql_file:
    sql_file.write(sql_script)
print("script gerado")

```

4.4 LIMPEZA DOS DADOS

A limpeza dos dados foi feita de forma manual, visto que poucas inconsistências foram encontradas. Uma das inconsistências eram nomes que continham apóstrofe no nome. Para lidar com esse problema, foi preciso inserir uma barra inversa antes do apóstrofe de cada nome em que esse problema se enquadrava, como mostrado na Figura 5 no nome "Lydia O'DONNELL".

Figura 5 – Trecho do Script em SQL com inconsistência tratada

```

105 INSERT INTO marathon (competition, country,gender, link, mark, name, position, sport)
VALUES ('New York Marathon - 04 NOV 2018', 'NZL', 'F','',CONCAT('0', '2:40:54'),'Lydia
O'DONNELL','27', 'Marathon');
106 INSERT INTO marathon (competition, country,gender, link, mark, name, position, sport)
VALUES ('New York Marathon - 04 NOV 2018', 'USA', 'F','',CONCAT('0', '2:48:24'),'Bridie
MCCAREY','28', 'Marathon');
107 INSERT INTO marathon (competition, country,gender, link, mark, name, position, sport)
VALUES ('New York Marathon - 04 NOV 2018', 'USA', 'F','',CONCAT('0', '2:49:09'),'Anoush
SHEHADEH','29', 'Marathon');

```


4.5 CRIAÇÃO DA BASE DE DADOS

A base de dados foi criada em *MySQL*, foi criado uma base de dados chamada *projetoic*, e nela foi criada duas tabelas, uma chamada *shotput*(arremesso de peso) e outra chamada *marathon*(maratona), a Figura 6 mostra o processo de criação da tabela *marathon*.

Figura 6 – Criação da tabela *marathon*

```
CREATE TABLE marathon (  
    competition varchar(200),  
    country varchar(10),  
    gender varchar(10),  
    link varchar(200),  
    mark time,  
    name varchar(100),  
    position varchar(100),  
    sport varchar(100)
```

Esse mesmo processo foi feito para a tabela *shotput*, com apenas uma mudança no tipo da variável "*mark*", visto que o "*mark*" da maratona é igual o tempo para a conclusão da maratona e o "*mark*" do arremesso de peso equivale a distância do arremesso(sendo *VARCHAR*).

Após a criação das tabelas, foi feito a inserção dos dados nelas como exposto na Figura 7, todo o código SQL gerado pelo *script* exposto na Figura 4 foi rodado dentro da base de dados *projetoic*.

Figura 7 – Trecho da inserção de todos os dados nas tabelas

```
File Edit View Search Terminal Help  
  
mysql> INSERT INTO marathon (competition, country, gender, link, mark, name, position, sport) VALUES ('New York Marathon - 04 NOV 2018', 'USA', 'F', '', CONCAT('0', '3:01:46'), 'Juliet HERSHEY-BEATT', '90', 'Marathon');  
Query OK, 1 row affected (0,01 sec)  
  
mysql> INSERT INTO marathon (competition, country, gender, link, mark, name, position, sport) VALUES ('New York Marathon - 04 NOV 2018', 'SUI', 'F', '', CONCAT('0', '3:01:55'), 'Kathrin KNUCHEL', '91', 'Marathon');  
Query OK, 1 row affected (0,01 sec)  
  
mysql> INSERT INTO marathon (competition, country, gender, link, mark, name, position, sport) VALUES ('New York Marathon - 04 NOV 2018', 'USA', 'F', '', CONCAT('0', '3:01:55'), 'Jessica SPAR', '92', 'Marathon');  
Query OK, 1 row affected (0,00 sec)  
  
mysql> INSERT INTO marathon (competition, country, gender, link, mark, name, position, sport) VALUES ('New York Marathon - 04 NOV 2018', 'AUS', 'F', '', CONCAT('0', '3:01:57'), 'Ingrid CLELAND', '93', 'Marathon');  
Query OK, 1 row affected (0,01 sec)  
  
mysql> INSERT INTO marathon (competition, country, gender, link, mark, name, position, sport) VALUES ('New York Marathon - 04 NOV 2018', 'USA', 'F', '', CONCAT('0', '3:02:20'), 'Sarah MELLEA', '94', 'Marathon');  
Query OK, 1 row affected (0,00 sec)
```


4.6 USO E CARACTERÍSTICAS DA BASE DE DADOS

Com a base de dados preenchida, é possível visualizar os dados como um todo fazendo a consulta nessa base e com uso de ferramentas de Ciência de dados, nos casos dos gráficos gerados nesse relatório, foram utilizadas as bibliotecas *Matplotlib* e *Pandas* do *Python*. As Figuras 8 e 9 são relacionadas a distribuição de vitórias nas competições por país e modalidade.

Figura 8 – Países com mais medalha de ouro no Arremesso de Peso Geral

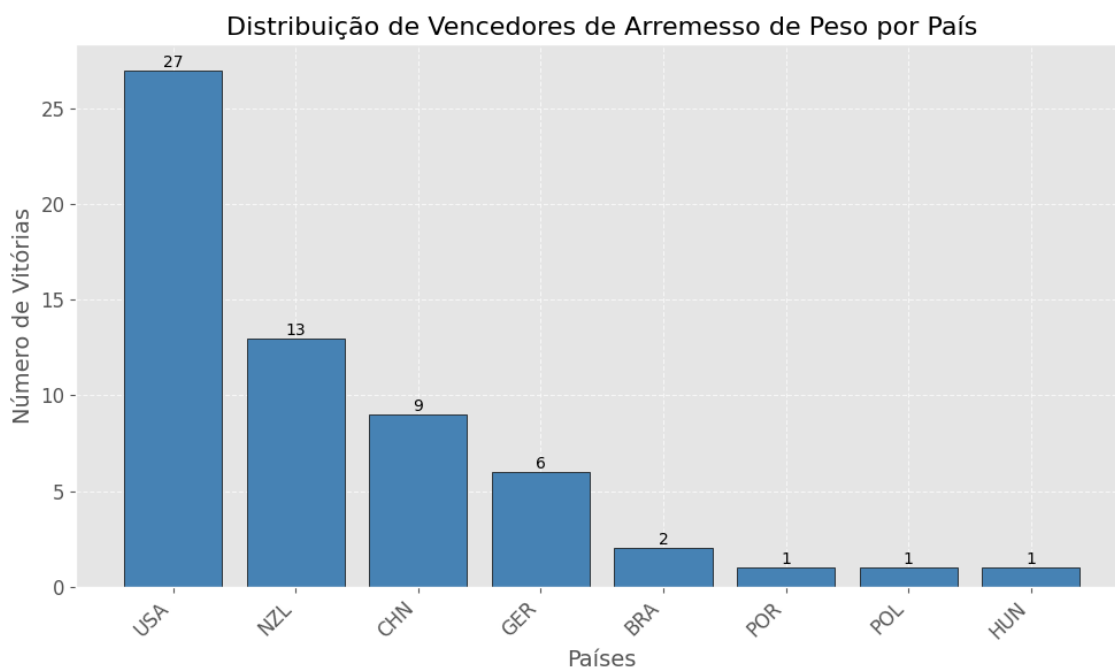
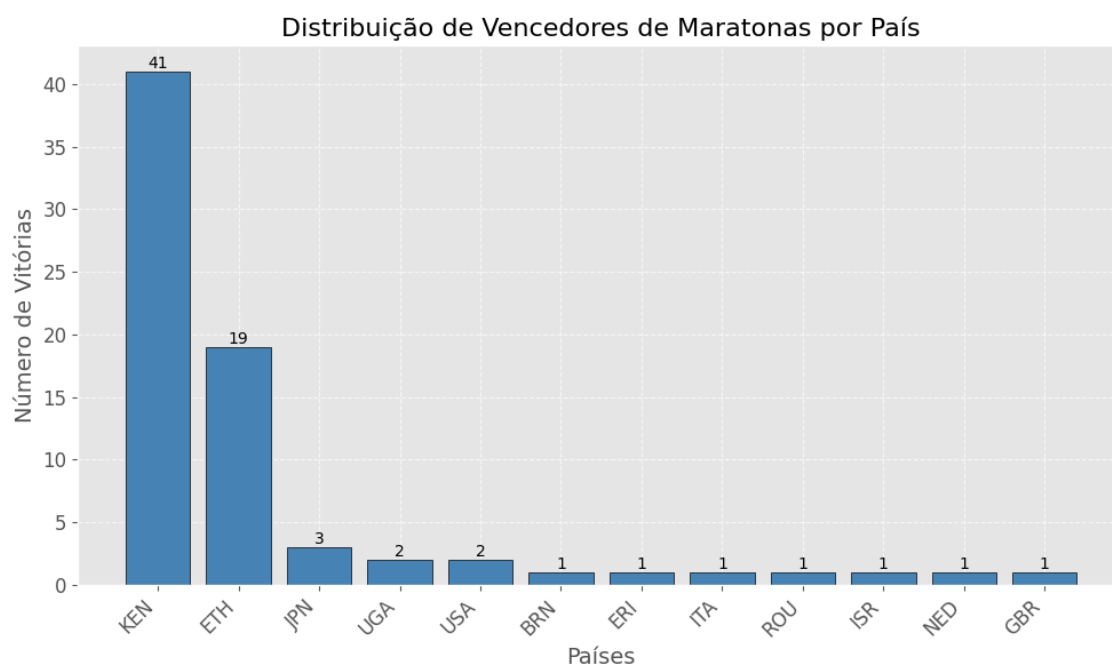


Figura 9 – Países com mais medalha de ouro na Maratona Geral



Também pode-se consultar os atletas que mais ganharam a medalha de ouro e seu respectivo país por gênero, como mostram as Figuras 10 a 13.

count	name	country
9	Lijiao GONG	CHN
6	Valerie ADAMS	NZL
4	Michelle CARTER	USA
3	Chase EALEY	USA
2	Christina SCHWANITZ	GER
1	Auriol DONGMO	POR
1	Tia BROOKS-WANNEMACHER	USA
1	Anita MÁRTON	HUN

Figura 10 – Atletas com mais medalhas de ouro no Arremesso de Peso Feminino

count	name	country
8	Ryan CROUSER	USA
7	Tomas WALSH	NZL
5	Joe KOVACS	USA
4	David STORL	GER
2	Ryan WHITING	USA
2	Darlan ROMANI	BRA
1	Kurt ROBERTS	USA
1	Tomasz MAJEWSKI	POL
1	Darrell HILL	USA
1	Christian CANTWELL	USA
1	Reese HOFFA	USA

Figura 11 – Atletas com mais medalhas de ouro no Arremesso de Peso Masculino

count	name	country
3	Edna Ngeringwony KIPLAGAT	KEN
3	Brigid KOSGEI	KEN
3	Peres JEPCHIRCHIR	KEN
3	Ruth CHEPNGETICH	KEN
2	Gotytom GEBRESLASE	ETH
1	Rosemary WANJIRU	KEN
1	Mary Jepkosgei KEITANY	KEN
1	Tigst ASSEFA	ETH
1	Ashete BEKERE	ETH
1	Gladys CHERONO	KEN
1	Desiree LINDEN	USA
1	Hellen OBIRI	KEN
1	Mia BEHM	USA
1	Vivian Jepkemoi CHERUIYOT	KEN
1	Yalemzerf YEHUALAW	ETH
1	Sifan HASSAN	NED
1	Joyciline JEPKOSGEI	KEN
1	Sharon LOKEDI	KEN
1	Rose CHELIMO	BRN
1	Mizuki NOGUCHI	JPN
1	Lonah Chemtai SALPETER	ISR
1	Birhane DIBABA	ETH
1	Ruti AGA	ETH
1	Constantina DITA	ROU
1	Jemima Jelagat SUMGONG	KEN
1	Tiki GELANA	ETH
1	Naoko TAKAHASHI	JPN
1	Mare DIBABA	ETH

Figura 12 – Atletas com mais medalhas de ouro na Maratona Feminino

count	name	country
7	Eliud KIPCHOGE	KEN
3	Evans CHEBET	KEN
2	Stephen KIPROTICH	UGA
2	Benson KIPRUTO	KEN
2	Lelisa DESISA	ETH
2	Birhanu LEGESE	ETH
1	Dickson CHUMBA	KEN
1	Kenenisa BEKELE	ETH
1	Guye Idemo ADOLA	ETH
1	Lawrence CHERONO	KEN
1	Yuki KAWAUCHI	JPN
1	Mo FARAH	GBR
1	Seifu TURA	ETH
1	Kelvin KIPTUM	KEN
1	Amos KIPRUTO	KEN
1	Albert KORIR	KEN
1	Geoffrey KAMWOROR	KEN
1	Chalu DESO	ETH
1	Tamirat TOLA	ETH
1	Geoffrey KIRUI	KEN
1	Stefano BALDINI	ITA
1	Gezahegne ABERA	ETH
1	Ghirmay GHEBRESLASSIE	ERI
1	Abel KIRUI	KEN
1	Samuel Kamau WANJIRU	KEN

Figura 13 – Atletas com mais medalhas de ouro na Maratona Masculino

Como resultado, a tabela da maratona criada nessa pesquisa é constituída por um total de 9501 linhas de dados com desempenho dos atletas(6350 atletas distintos) nas competições, dentro dessa tabela, há os resultados finais de 74 competições de maratonas, sendo 37 do masculino e 37 do feminino como mostrado na Figura 15, são competições que ocorreram no intervalo dos anos 2000 ao ano 2023. As competições que constituem a tabela são as mais famosas no nicho da maratona, tendo a presença de maratonas olímpicas, campeonatos mundiais de atletismo e as *Majors*, que são as maratonas de Tóquio, *Chicago*, *Boston*, Londres, *Berlin*

e *New York City*. Já na tabela de arremesso de peso, foram armazenadas 644 linhas com os desempenhos individuais dos atletas (161 atletas distintos) dentro de competições, de um total de 60 competições, 33 são do masculino e 27 do feminino, como mostra a Figura 14.

Número de Competições coletadas por gênero no arremesso de peso

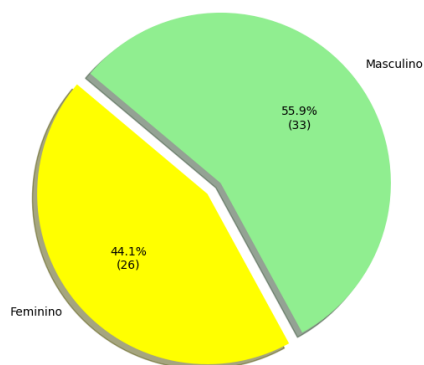


Figura 14 – Distribuição de competições por gênero em Arremesso de Peso

Número de Competições coletadas por gênero na maratona

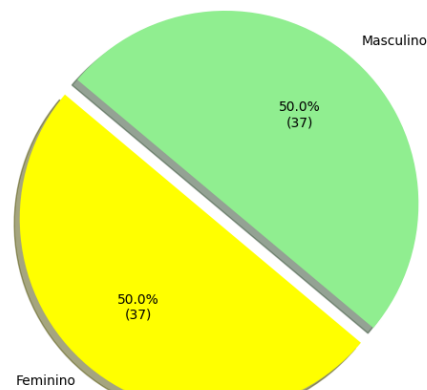


Figura 15 – Distribuição de competições por gênero em Maratona

5 CONSIDERAÇÕES FINAIS

Esse relatório final traz uma descrição dos trabalhos realizados durante o período de 2022/2023 na bolsa da PIBIC, apresentando de forma sucinta todas as etapas no desenvolvimento do projeto de criação da base de dados das competições de maratona e arremesso de peso.

Em conclusão, o produto final desta pesquisa cumpre com as expectativas, visto que a base de dados foi criada e contém dezenas de competições em ambos os esportes. Há espaço para aperfeiçoamento em trabalhos futuros, sendo possível expandir as informações dos atletas, como por exemplo adicionar a idade do atleta quando a competição foi realizada. Qualquer incremento dentro dessa base de dados trará mais profundidade nos *insights* que poderão ser retirados desse projeto.

No ponto de vista da bolsa de iniciação científica, vejo que foi uma oportunidade de me aprofundar em temas que me interessavam, onde desenvolvi minhas habilidades na área de banco de dados e extração de dados na *Web*. Além de desenvolver as habilidades de realização de trabalhos a longo prazo, cumprir metas, realizar escrita e pesquisa científica.

6 REFERENCIAS

- [1] Sirisuriya, D. S., (2015). A comparative study on web scraping. Proceedings of 8th International Research Conference, KDU
- [2] SCRAPY. Scrapy 2.10 documentation. Disponível em: <https://docs.scrapy.org/en/latest/>. Acesso em: 20 jul. 2023.
- [3] MYSQL. MySQL Documentation. Disponível em: <https://dev.mysql.com/doc/>. Acesso em: 20 jul. 2023.
- [4] World Athletics. Marathon. Disponível em: <https://worldathletics.org/disciplines/road-running/marathon>. Acesso em: 20 jul. 2023.
- [5] World Athletics. Shot Put. Disponível em: <https://worldathletics.org/disciplines/throws/shot-put>. Acesso em: 20 jul. 2023.
- [6] Khder, Moaiad. (2021). Web Scraping or Web Crawling: State of Art, Techniques, Approaches and Application. International Journal of Advances in Soft Computing and its Applications. 13. 145-168. 10.15849/IJASCA.211128.11.
- [7] NASCIMENTO, Frederico Queiroz. ANÁLISE DE DESEMPENHO ESPORTIVO BASEADA EM DADOS ESTATÍSTICOS. 2017. 33 f. TCC (Graduação) - Curso de Sistemas de Informação, Departamento de Ciência da Computação, Universidade Federal Fluminense, Niterói, 2017.