

Tópicos Avançados em Base de Dados (CC4007)

Construção e análise de uma datawarehouse sobre um competição nacional de natação

Ricardo Ribeiro
Rita Abreu
Tatiana Araújo

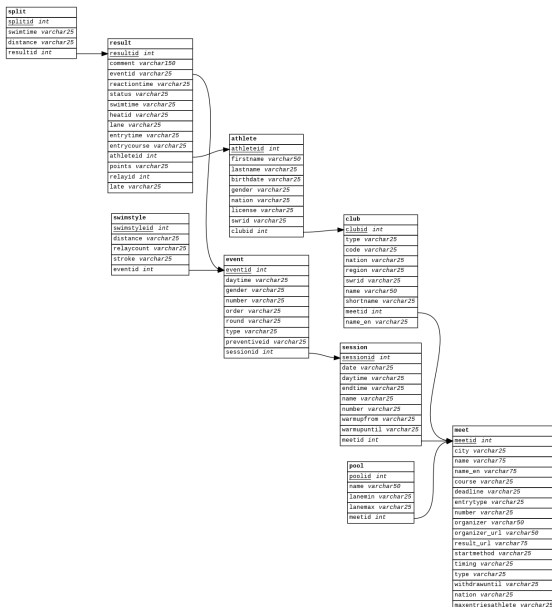
DCC/FCUP

Junho 2022

- 1 Introdução
- 2 Análise da Base de dados Inicial
- 3 Construção da Datawarehouse
- 4 Análise da Base de Dados
- 5 Conclusão

- Esta base de dados é composta por vários dados relacionados com a competição nacional de natação de nadadores no nível *Master*.
- Primeiramente construímos uma nova base de dados a partir da base de dados fornecida pelo professor, mas com o esquema de um *Datawarehouse*.
- De seguida, passamos à análise da base de dados.

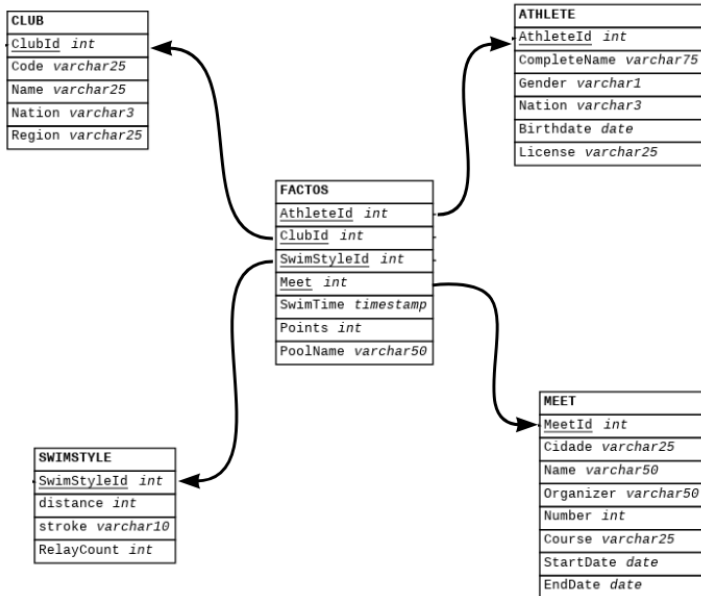
Análise da Base de dados Inicial



Construção da Datawarehouse

- Através da tabela *athlete* criamos o atributo *completeName* através da junção do *firstname* com o *lastname* e retiramos o *swrid*.
- Na tabela *club*, retiramos os atributos *type*, *swrid* *name_en* porque consideramos que não eram essenciais para análise futura.
- Na tabela *swimstyle*, mantivemos todos os atributos.
- Na tabela *meet*, achamos que os atributos *name_en*, *deadline*, *entrytype*, *organizer*, *organizer_url*, *result_url*, *startmethod*, *timing*, *type*, *withdrawuntil*, *maxentriesathlete* e *entrystartdate* não eram revelantes para a análise do meet.
- Por fim, consideramos que as tabelas *split*, *pool*, *event*, *result* e *session* não nos forneciam informação relevante para a nossa análise.

Construção da Datawarehouse - Resultado Final



Construção da Datawarehouse - Explicação do código

- "INSERT IGNORE INTO" por "ON CONFLICT DO NOTHING"
- *order* na tabela *event*
- *organizer.url*, *name.en* na tabela *meet*
- *decode_LXF.py* *meetid*

Análise da Base de Dados

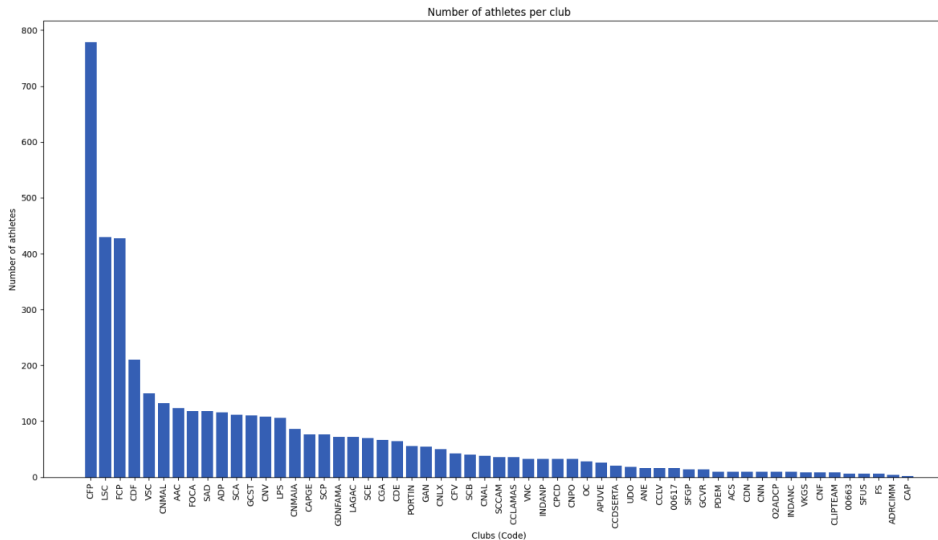
Considerações gerais de cada meet

O *meet* com id 41231234 representa o XXII Campeonato Nacional Masters de Verão - OPEN e o 41231235 Troféu Pescada - José Carlos Freitas e Troféu Master ANNP.

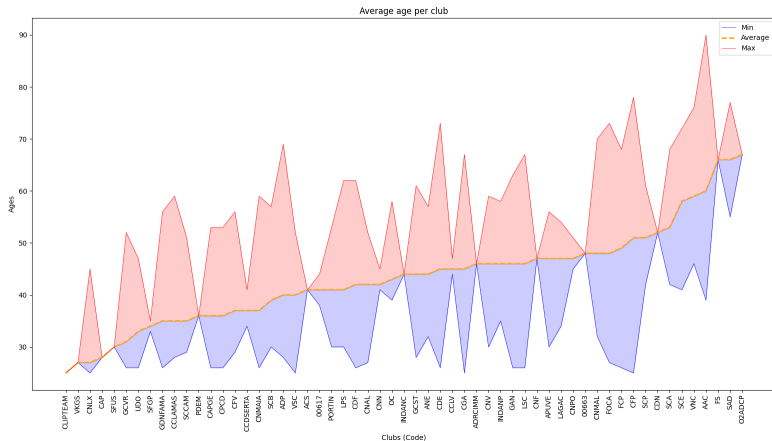
Tabela 1: Tabela com dados de meets

	Nr de clubes	Nr de Atletas	%Femininas	% Masculino
41231234	27	421	35%	65%
41231235	52	225	32%	67%

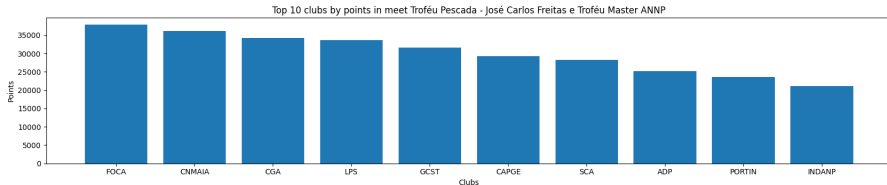
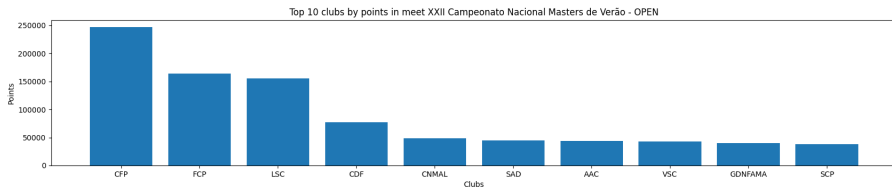
Número de Atletas por Clube



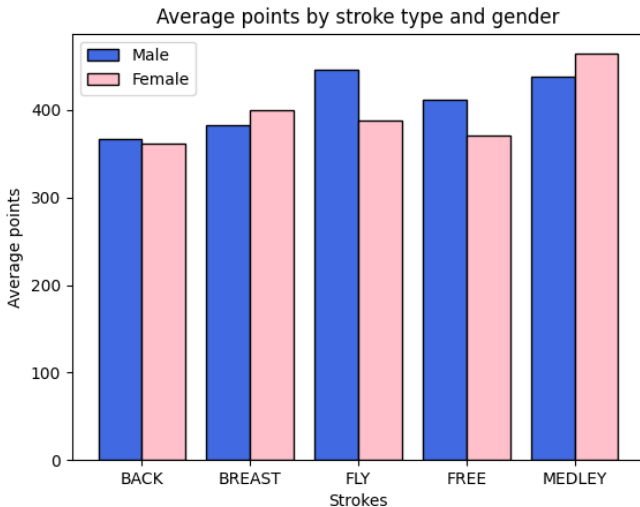
Idade Mínima, Média e Máxima dos atletas por clube



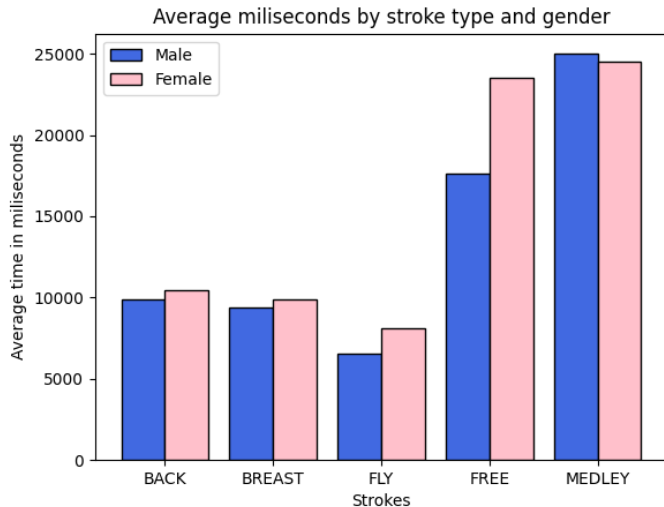
Os 10 melhores clubes da competição



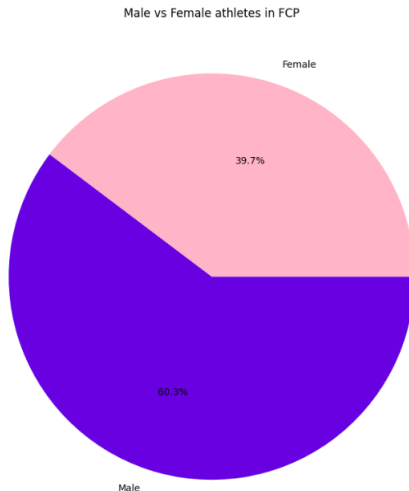
Pontuação média por estilo de natação e género



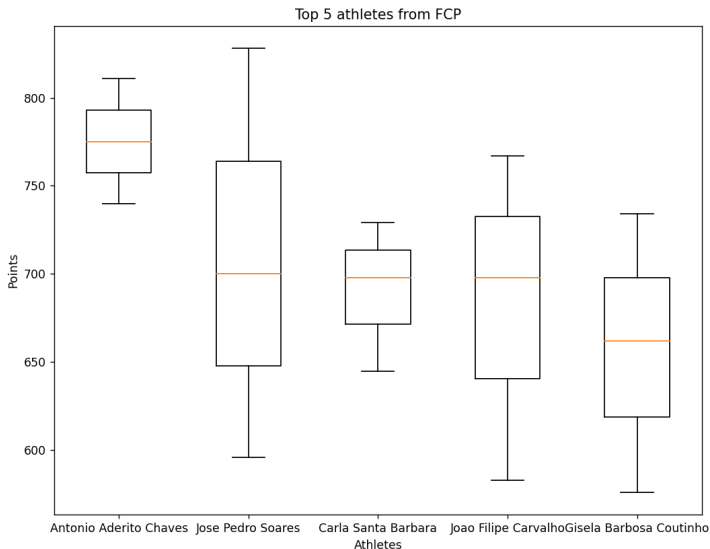
Tempo médio por estilo de natação e género



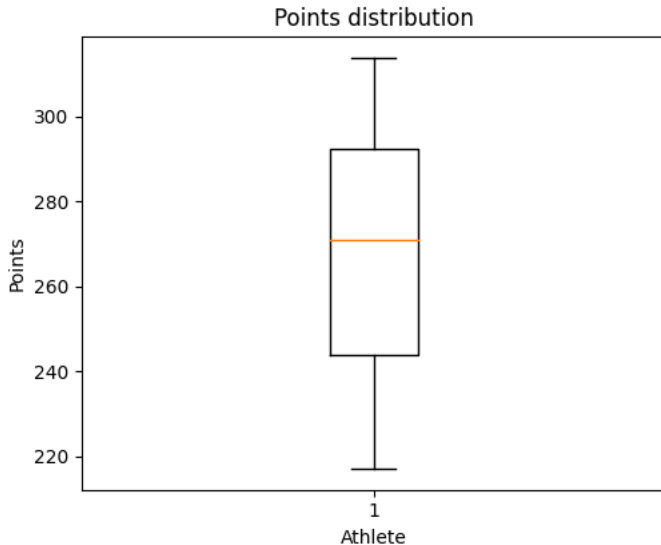
Análise por clube: Distribuição de géneros



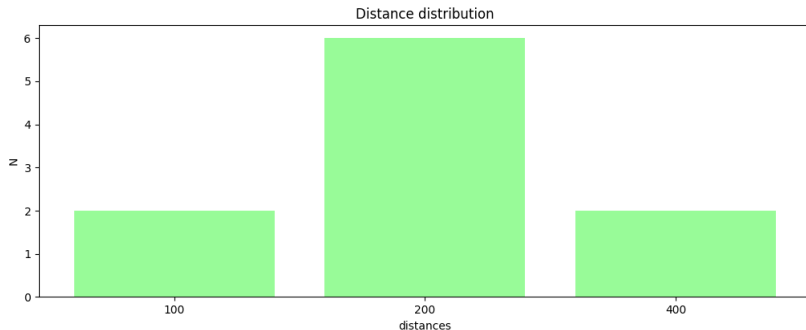
Análise por clube: 5 melhores jogadores



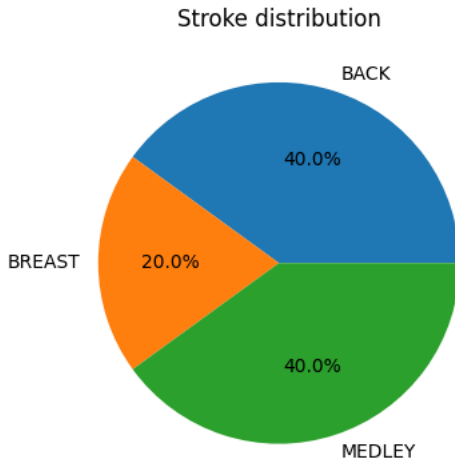
Análise por atleta : Pontos ranking de um jogador



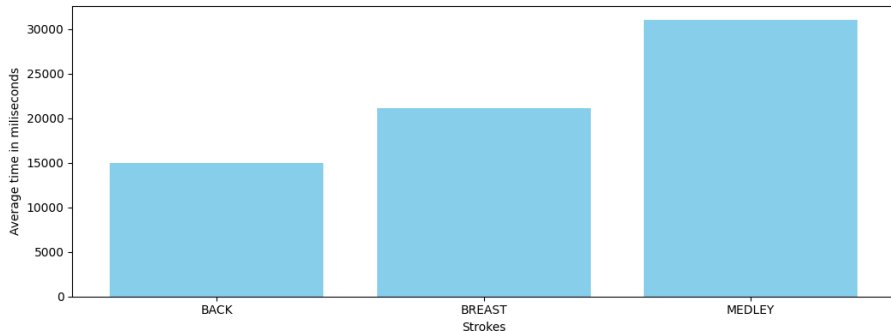
Análise por atleta: Número de vezes que um atleta fez uma prova de determinada distância



Análise por atleta: Distribuição dos estilos de natação utilizados pelo atleta



Análise por atleta: Average time por estilo do atleta



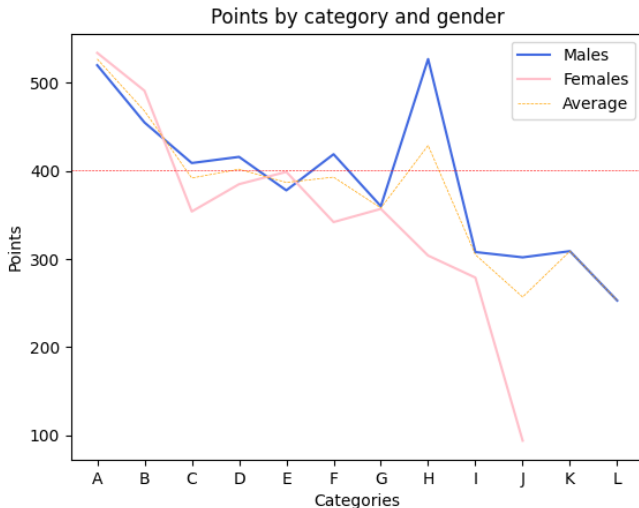
Análise por escalão

- Nesta secção iremos analisar a nossa base de dados por escalão.
- Definimos os escalões com recurso ao regulamento de competições nacionais da categoria *Master* da Federação Portuguesa de Natação.

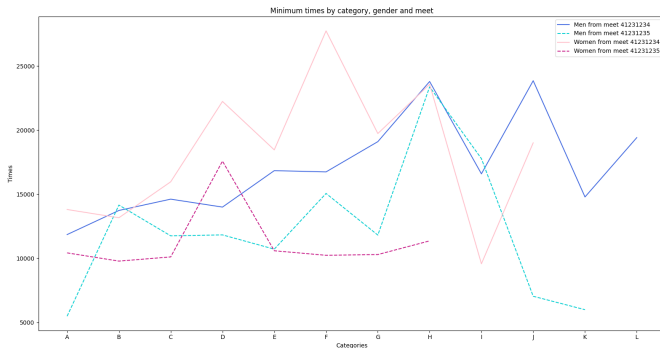
Tabela 2: Tabela de Escalões

Grupo	Escalão Etário	Ano de Nascimento
Master A	25 – 29	94-90
Master B	30 – 34	89-85
Master C	35 – 39	84-80
Master D	40 – 44	79-75
Master E	45 – 49	74-70
Master F	50 – 54	69-65
Master G	55 – 59	64-60
Master H	60 – 64	59-55
Master I	65 – 69	54-50
Master J	70 – 74	49-45
Master K	75 – 79	44-40
Master L	80 – 84	39-35
Master M	85 – 89	34-30
Master N	90 – 94	29-25
Master O	95 - 99	24-20

Análise da pontuação por género e por escalão



Análise dos tempo por género, por escalão e por meet



- Este projeto deu-nos a oportunidade de trabalhar com o *PostgreSQL*, que demonstrou ser uma ferramenta bastante útil, assim como melhorar os nossos conhecimentos de SQL.
- Ao longo do trabalho deparamo-nos com alguns problemas, no entanto, o nosso maior problema prendeu-se com a construção da *datawarehouse*, isto porque inicialmente não estávamos a conseguir perceber como a devíamos estruturar. Depois de alguma persistência e ajuda do professor conseguimos ultrapassar esta dificuldade.
- Infelizmente, apenas pouco tempo antes da data limite de entrega do trabalho conseguimos adicionar mais do que um *meet* à nossa base de dados, o que nos levou a não conseguir fazer uma análise suficientemente completa de vários *meets*. Este problema foi uma grande entrave no desenvolvimento do projeto.

FIM