

Avaliação de atletas usando dados do Garmin

Projeto da Disciplina de BI

Prof. Anderson Nascimento
prof.anderson@ica.ele.puc-rio.br

Componentes do Projeto:
Gustavo Araújo - Gustavo.g3a@gmail.com

Histórico de Versões

Data	Versão	Descrição	Autor	Aprovado por
04/11/2021	1.0	Primeira versão	Gustavo Araújo	Anderson Nascimento

Sumário

1	INTRODUÇÃO.....	4
2	ESTUDO DE CASO.....	5
2.1	DESCRIÇÃO DO ESTUDO DE CASO	5
	DESCRIÇÃO DO MODELO TRANSACIONAL	6
2.2	FONTE 1 – BANCO DE DADOS TRANSACIONAL.....	6
3	PROPOSTA DE PROCESSO DE BI	8
	MODELO MULTIDIMENSIONAL.....	9
	ELABORAÇÃO DO DATA WAREHOUSE.....	10
3.1	DEFINIÇÃO DO DW	10
4	PROJETO DE ETL	11
4.1	DESCRIÇÃO DO PROJETO DE ETL.....	11
	DASHBOARD	13
4.2	DESCRIÇÃO DA ELABORAÇÃO	13
4.3	TELAS DO DASHBOARD	13
	CONCLUSÃO	15
5	ARQUIVOS	16

1 Introdução

Este documento tem por finalidade coletar, analisar e definir as principais necessidades do projeto do estudo de caso ***Avaliação de atletas***. O documento procura demonstrar a comparação de diversas métricas de corrida e natação entre atletas.

2 Estudo de Caso

2.1 Descrição do Estudo de Caso

Durante uma conversa entre três amigos que praticam esporte, surgiu um debate de qual deles seria o mais completo atleta.

Cada atleta levantou uma métrica para a classificação, por exemplo, Luiz que era o melhor na modalidade corrida, disse que por ser o mais veloz, seria o melhor atleta. Porém, Ronaldo não concordou e disse que o próprio seria o melhor atleta já que ultimamente teria percorrido grandes distâncias correndo e que tem sido mais veloz que Luiz, além de praticar a natação. Já o Gustavo admitiu que não poderia competir com os outros dois na modalidade corrida, mas que sua corrida não estava tão aquém da velocidade e nem da distância de ambos e finalizou sua argumentação com a justificativa que ele deveria ser considerado o melhor atleta, pois acreditava que sua natação seria melhor que a dos outros dois e que seria o esporte mais completo.

A conversa não teve uma conclusão, pois em épocas diferentes um atleta dizia que era o melhor em tal modalidade devido a tal métrica de performance. Então, Gustavo que estava estudando banco de dados e business intelligence, resolveu propor uma solução. Propôs que os outros dois atletas enviassem a ele seus dados coletados pelo relógio Garmin que todos tinham em comum e usavam para medir todas as atividades. Os atletas deveriam enviar todos os seus dados disponíveis em um intervalo de 1 ano (Setembro-2020 a Agosto-2021). Com os dados em mãos, Gustavo explicou que prepararia no SQL um sistema transacional e que a partir desse sistema poderia retirar as informações necessárias para a análise da performance de cada atleta e compará-las usando um “Dashboard” com gráficos para a melhor visualização de todos.

Definido as informações relevantes para a realização do “Dashboard”, Gustavo começou seu projeto que será detalhado nas próximas seções.

Descrição do Modelo Transacional

O primeiro passo foi criar um modelo transacional, primeiramente cada atleta baixou seus respectivos dados no site da Garmin. Foi acordado que os atletas deveriam baixar todos os dados em arquivo .csv.

Os arquivos enviados foram referente as medias mensais das métricas em um intervalo de 1 ano, por exemplo, frequência cardíaca media durante as corridas no mês de fevereiro.

Para a inserção dos arquivos .csv de cada atleta, foi criado um modelo transacional usando a plataforma Postgres. O arquivo “garmin_dados_DDL” na folha de anexos contem a criação das tabelas do modelo, já o arquivo “garmin_dados_DML” contem a rotina de carga de dos dados dos arquivos .csv para o modelo.

2.2 Fonte 1 – Banco de dados transacional

Abaixo a representação UML do modelo transacional utilizado para o projeto.

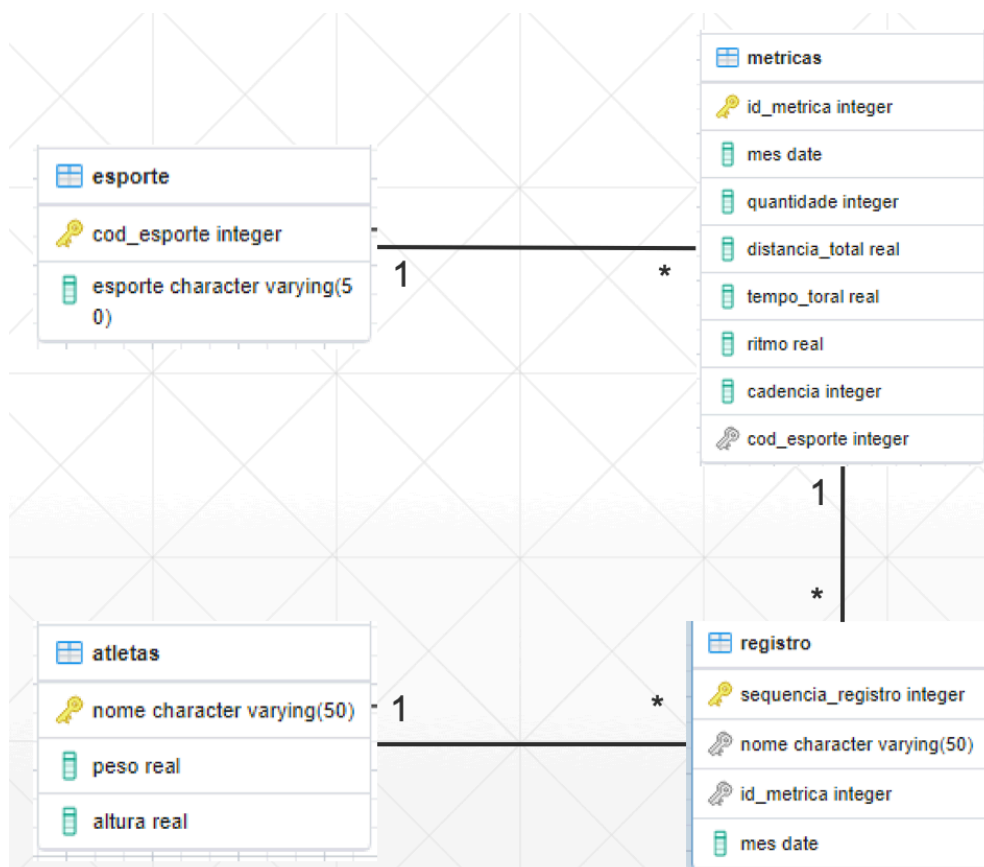


Figure 1 - Modelo Transacional

Quatro tabelas foram criadas em linguagem DDL:

- Tabela “Atletas” -> Com o nome de cada atleta e seus dados como peso e altura;
- Tabela “Esporte” -> Com os tipos de esportes e um código para cada modalidade;
- Tabela “Métricas” -> Nessa tabela todos as métricas foram listadas por modalidade. Cada grupo de métricas referente a um atleta, a um mês e a uma modalidade tem seu identificador único;
- Tabela “Registro” -> Nessa tabela ocorrem todas as conexões entre as tabelas, relacionando atleta a sua métrica e adicionando a data da realização da atividade.

A carga das quatro tabelas foram realizadas em linguagem DML, como detalhadas a seguir:

- Tabela “Atletas” -> Carga inserida através do comando “INSERT INTO” do SQL, foram adicionados os atletas e suas características;
- Tabela “Esporte” -> Carga inserida através do comando “INSERT INTO” do SQL, foram adicionadas as modalidades e seus códigos de identificação;
- Tabela “Métricas” -> Carga inserida através do comando “COPY” dos arquivos .csv de métricas de cada atleta;
- Tabela “Registro” -> Carga inserida através do comando “INSERT INTO” do SQL, foram adicionados cada realização em determinado mês por parte de cada atleta;

3 Proposta de Processo de BI

Com o modelo transacional pronto, o próximo passo foi a preparação do “Data Warehouse” considerando apenas as informações essenciais acordadas durante a discussão do projeto.

Primeiramente, foi selecionada o Pentaho (PDI) como ferramenta para a extração de dados do modelo transacional. Através do PDI, as informações necessárias são extraídas do banco de dados, que se localiza no Postgres como descrito anteriormente, e tratadas para serem alocadas no “Data Warehouse” (DW), processo denominado como “ETL” (Extraction, Transformation and Load).

O DW foi armazenado também no Postgres em uma “Database” separada da “Database” do transacional. A partir do DW os dados já customizados estão prontos para serem acessados através de uma ferramenta de análise gráfica de visualização de dados.

O Power BI, usado extensivamente no mercado, foi escolhido como aplicação para a visualização de dados. Através do acesso ao DW, através do Database no Postgres, o Power BI coleta as informações para serem plotadas de forma direta e simples para o proposito do usuário ter liberdade de filtrar ou adicionar dados em sua tela de visualização, também conhecida como “Dashboard”.

Abaixo segue uma ilustração do processo descrito acima:

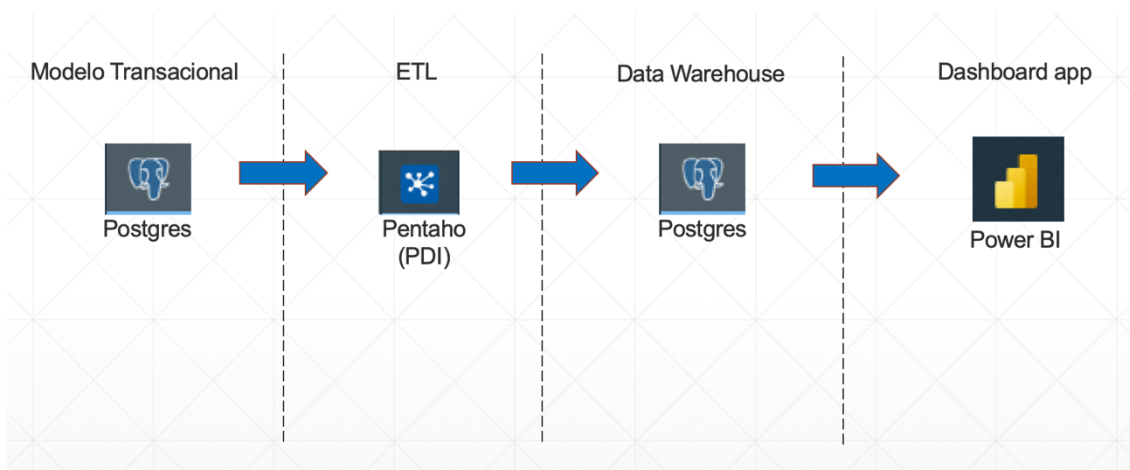


Figure 2 - Processo de BI

Modelo Multidimensional

Esta seção apresenta o modelo estrela (star) do estudo de caso Avaliação de atletas usando o Garmin.

Para a criação do DW, foi utilizado o modelo Star, onde uma tabela chamada de tabela fato eh conectada a outras tabelas chamadas de tabelas dimensão.

A tabela Fato foi nomeada de “FT_Registro” e nesta estão inclusas todas as chaves estrangeiras (FK) relacionadas as chaves primarias de cada tabela dimensão.

As tabelas Dimensão são formadas por três tabelas , são essas:

- DIM_tempo -> Contem os meses onde foram realizadas atividades por cada atleta;
- DIM_atletas -> Contem os nomes de cada atleta;
- DIM_metricas -> Tabela originada de uma desnormalizacao, conhecida também como “JOIN” das tabelas “métricas” e “esporte” originadas no modelo transacional.

Abaixo segue uma ilustração do processo multidimensional descrito acima:

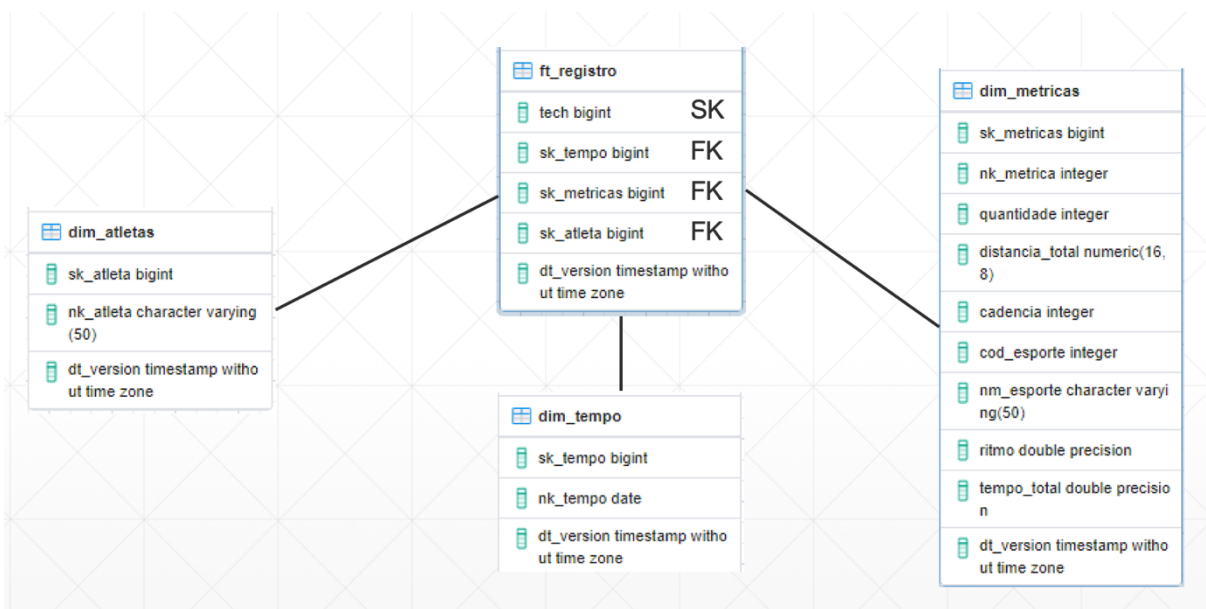


Figure 3 - Modelo Multidimensional

Elaboração do Data Warehouse

O Data Warehouse será a fonte integradora de informações da empresa, a tecnologia será utilizada com o intuito de servir de base para a camada de aplicação que será responsável por fornecer dados para a tomada de decisão na organização.

Os atletas envolvidos no projeto decidiram quais as mais informações mais importantes para a análise final de cada modalidade e as mesmas foram as escolhidas para permanecer no DW. São elas:

- A cadencia de passos na corrida e braçadas na natação. Pois essa métrica permite avaliar melhor técnica já que em ambos esportes quanto maior a cadencia, melhor o nível técnico do atleta;
- O ritmo do atleta, pois o ritmo define a velocidade do atleta, pois eh medido em distancia percorrida por tempo;
- A quantidade de atividades que o atleta realiza, pois atletas que realizam mais atividades ficam mais cansados e tendem a ter o ritmo e a cadencia um pouco comprometidas;
- As horas investidas em cada modalidade, pois algumas atividades podem ser muito longas e isso implica em uma capacidade mental e paciência que também deveria ser levado em consideração;
- E os quilômetros percorridos, já que o atleta precisa ser também resistente além de veloz, pois um atleta completo precisa ter um corpo capaz de aguentar longas jornadas;

3.1 Definição do DW

3.1.1 Arquitetura

A arquitetura do DW escolhida foi a independente devido a uma rápida implementação, sem necessidades de ter redundância ou Data Marts em outros locais.

3.1.2 Abordagem de Construção

A abordagem de construção escolhida foi a “bottom up”, pois a fonte de cada dado vem de um atleta diferente e são todas agrupadas em um DW final.

3.1.3 Arquitetura Física

Como não houve investimento financeiro para a elaboração do projeto, a arquitetura física do DW escolhida foi a nuvem devido ao custo zero para pequenas quantidades de dados.

4 Projeto de ETL

4.1 Descrição do Projeto de ETL

O projeto ETL realizado no PDI para a carga de dados e criação das tabelas discutidas no modelo multidimensional foi composto por 4 arquivos:

- **Tempo.ktr** -> Arquivo que extrai a informação dos meses (DATE) da tabela “metricas” do transacional e cria/carrega no DW na tabela “Dim_tempo”;
- **Atletas.ktr** -> Arquivo que extrai a informação dos nomes dos atletas da tabela “atletas” do transacional e cria/carrega no DW na tabela “Dim_atletas”;
- **Metrica_esporte.ktr** -> Arquivo que extrai a informação das metricas da tabela “metricas” do transacional e informação dos esportes da tabela “esporte” do transacional e faz um “join” de ambas. Também nesse mesmo arquivo de transformação, foi realizado para as metricas de “ritmo” e “tempo_total”, a conversão de valores de unidades para o melhor entendimento do usuário final. Por exemplo, para ritmo, o valor 5.88 de corrida fora convertido em 5:53 min/km e para o tempo total de corrida, o valor de 405.1 fora convertido em 6:45 horas. Os dados são todos carregados no DW na tabela “Dim_metricas”.
- **Ft_registro.ktr** -> Arquivo que extrai a informação das realizações da tabela “registro” do transacional, consulta o DW em cada dimensão em serie e cria/carrega as conexões no DW na tabela “ft_registro”;

Abaixo a ilustração dos processos descritos acima no PDI:

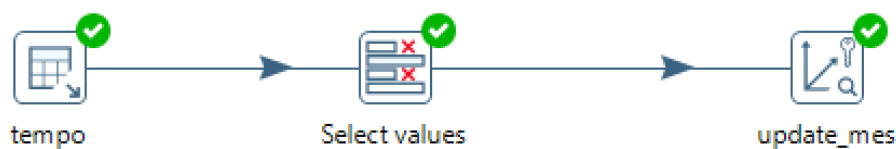


Figure 4 - Transformação tempo



Figure 5 - Transformação Atletas

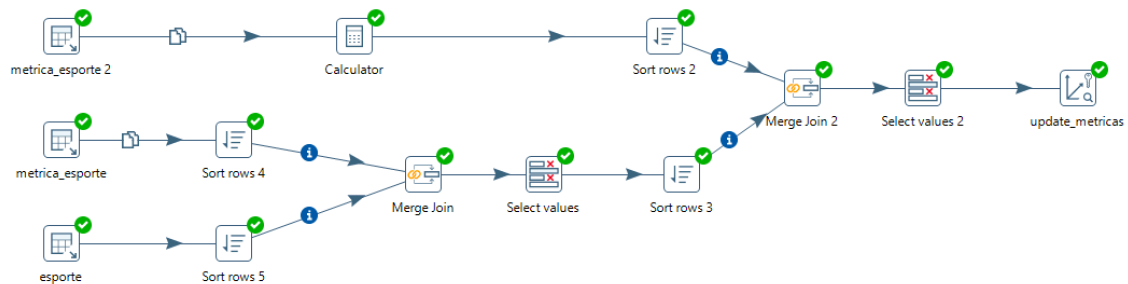


Figure 6 - transformação Metricas_esporte

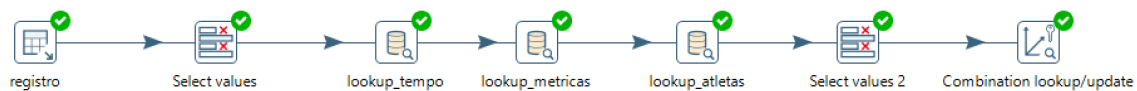


Figure 7 - Transformação Registros

Dashboard

4.2 Descrição da Elaboração

Como o cerne do projeto era a comparação das métricas entre atletas, a melhor solução foi implementar um único dashboard para que os usuários possam comparar e visualizar rapidamente e integrado os gráficos de cada métrica.

4.3 Telas do Dashboard

Na tela de Dashboard criada, o atleta consegue filtrar o mês ou período de vários meses, a modalidade e com qual outro atleta ele pretende comparar seu desempenho. Feita a filtragem, um gráfico de barras demonstra visualmente a comparação das métricas da modalidade entre os atletas naquele mês.

Para os critérios chaves discutidos ao longo do projeto, foram feitas medidas para dar o resultado de qual foi o atleta mais bem avaliado no determinado quesito.

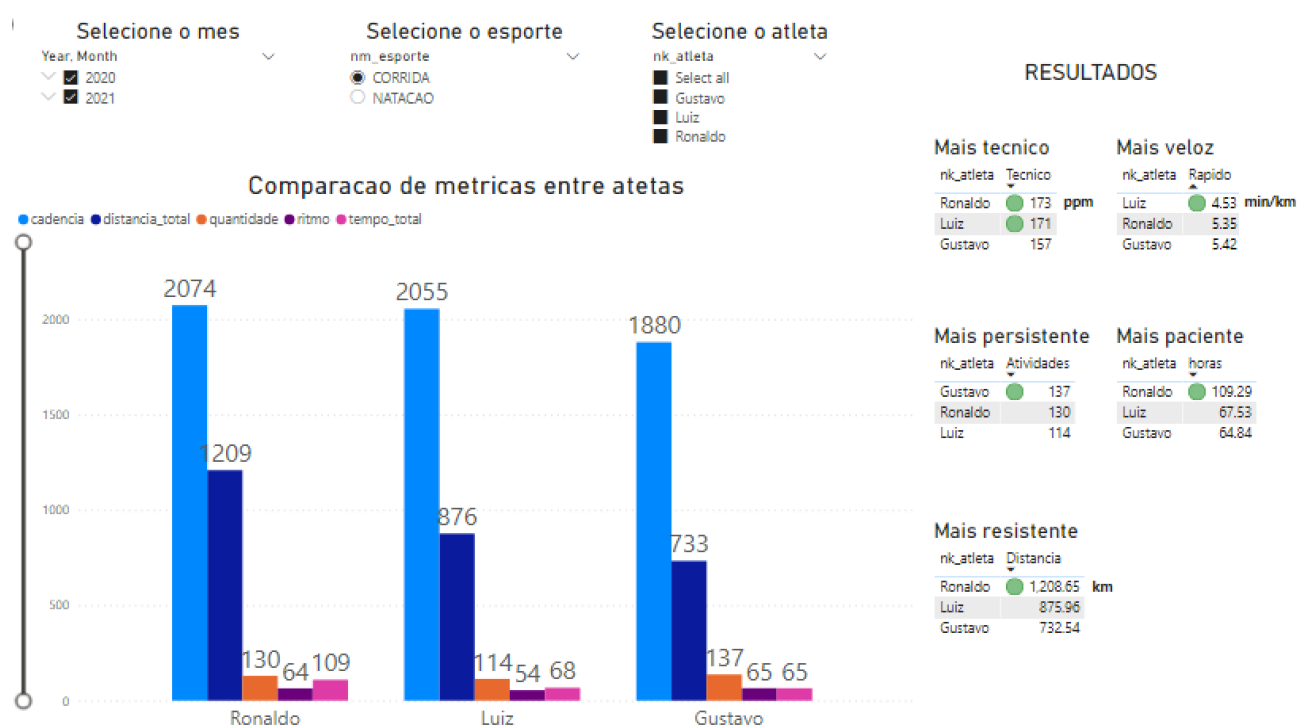


Figure 8 - Tela de Dashboard

Os resultados finais calculados por medidas no Power BI são:

- Atleta mais técnico: Atleta que possui melhor media de passos por minuto, caso o filtro seja a modalidade corrida, e melhor media de braçadas por minuto, caso natação.
- Atleta mais veloz: Melhor ritmo, minutos por quilometro, tanto para a modalidade corrida quanto para a natação.
- Atleta mais persistente: Mais atividades realizadas durante o mês ou período selecionado.
- Atleta mais paciente: Atleta que mais investiu horas correndo ou nadando com sua mente sendo também trabalhada além do corpo.
- Atleta mais resistente: Atleta que mais percorreu quilômetros no mar ou na terra, prova de um bom preparo físico e resistência do corpo para longas distancias.

Conclusão

O desenvolvimento deste projeto, apesar de ser de objetivo recreativo e não convencional ou corporativo, foi importante para o aprendizado do processo desde a coleta de dados , o armazenamento em um banco de dados, o tratamento, a escolha e filtragem dos dados e ate a ilustração gráfica para tomada de decisões.

Especificamente para o projeto em questão, a ideia foi mostrar que nada eh binário, ou seja, não há como dizer que um atleta seja melhor que o outro sem um critério bem definido de avaliação. A base de dados mostra que há diversos métodos e critérios de avaliação e com esses dados transparentes ao ver do julgador, fica mais fácil obter uma avaliação mais realística.

Na modalidade corrida, Luiz foi o mais veloz, porem não o mais técnico muito menos o mais resistente, já o Gustavo nessa mesma modalidade foi o mais lento e menos resistente , porem na modalidade natação foi o melhor em todos os critérios. Com o dashboard rapidamente consegue obter novas conclusões anteriormente ocultas pela dificuldade de cruzamento de dados.

5 Arquivos

Postgres:

- Database: **Garmin_dados**:
 - “**garmin_dados_DDL**” – Criação de tabelas para o modelo transacional;
 - “**garmin_dados_DML**” – Carga de dados para as tabelas para o modelo transacional;
- Database: **Garmin_dw**:
 - “**garmin_ERD.pgerd**” - Tabelas geradas a partir do multidimensional;

PDI:

- **Tempo.ktr** -> Arquivo que extrai a informação dos meses (DATE) da tabela “metricas” do transacional e cria/carrega no DW na tabela “Dim_tempo”;
- **Atletas.ktr** -> Arquivo que extrai a informação dos nomes dos atletas da tabela “atletas” do transacional e cria/carrega no DW na tabela “Dim_atletas”;
- **Metrica_esporte.ktr** -> Arquivo que extrai a informação das metricas da tabela “metricas” do transacional e informação dos esportes da tabela “esporte” do transacional e faz um “join” de ambas. Também nesse mesmo arquivo de transformação, foi realizado para as metricas de “ritmo” e “tempo_total”, a conversão de valores de unidades para o melhor entendimento do usuário final. Por exemplo, para ritmo, o valor 5.88 de corrida fora convertido em 5:53 min/km e para o tempo total de corrida, o valor de 405.1 fora convertido em 6:45 horas. Os dados são todos carregados no DW na tabela “Dim_metricas”.
- **Ft_registro.ktr** -> Arquivo que extrai a informação das realizações da tabela “registro” do transacional, consulta o DW em cada dimensão em serie e cria/carrega as conexões no DW na tabela “ft_registro”;

POWER BI:

- **ProjetoBI_garmin_dashboard.pbix** -> Dashboard de visualização dos dados do projeto de BI;