

Administração de Banco de Dados Oracle  
INFNET



Aluno: João Pedro De Sousa Moura

# Joao Pedro De Sousa Moura

## DATAMART Olimpíadas

Dashboard contando a comparação da participação de gênero por ano.

**RESUMO:** O Projeto se baseia na contagem da participação de cada atleta independente da repetição, para indicar o crescimento da participação do gênero feminino nas partidas das olimpíadas de 1896 a 2016.

## SUMARIO

<b>1 Introdução.....</b>	<b>4</b>
1.1 Fonte de dados.....	4
1.2 Periodicidade.....	4
1.3 Fases do projeto.....	4

<b>2 Analise inicial dos dados.....</b>	<b>5</b>
2.1 Origem dos dados.....	5
2.2 Analise dos dados.....	5
<b>3 Modelo dimensional.....</b>	<b>5</b>
3.1 Modelo estrela.....	6
3.2 Dicionário de dados.....	6
<b>4 Carga de dados ETL.....</b>	<b>7</b>
4.1 Artefatos Scrum.....	7
<b>5 Documentação ETL.....</b>	<b>8</b>
5.1 Tabelas.....	8
5.2 Carga de dados.....	9
<b>6 Analise multidimensional.....</b>	<b>10</b>
6.1 Estrutura do cubo.....	10
<b>7 Analise tabular.....</b>	<b>11</b>
7.1 Relacionamentos.....	12
7.2 Tabelas.....	12
<b>8 Dashboards.....</b>	<b>15</b>
<b>9 Conclusões.....</b>	<b>17</b>

## 1 INTRODUÇÃO

1.1 Fonte de dados: Database dos dados contendo informações sobre todos os atletas que participaram das olimpíadas, de inverno e verão.

1.2 Periodicidade: O database abrange todos os eventos envolvendo as olimpíadas, datado de 1896 a 2016

### 1.3 FASES DO PROJETO

#### a) Análise Inicial dos Dados

Nesta primeira fase os dados dos atletas e as datas serão carregados dos arquivos de origem CSV para tabelas, com objetivo de explorar e compreender os dados.

#### b) Modelo Dimensional

Após a análise inicial dos dados, será desenvolvido um modelo dimensional estrela, assim como seus aspectos físicos, específicos para o SQL Server.

#### c) Carga dos Dados (ETL)

Nesta fase serão desenvolvidos pacotes de carga de dados com o Microsoft Integration Services, alimentando as tabelas do modelo estrela.

#### d) Banco de Dados de Análise

Desenvolvimento de banco multidimensional utilizando o Microsoft Analysis Services, para proporcionar livre cruzamentos das informações das Olimpíadas, utilizando Pivot Table.

#### e) Dashboard

Desenvolvimento de Dashboard utilizando o Microsoft Power BI contendo os principais indicadores, com o objetivo de avaliar o aumento da participação do gênero feminino nas partidas.

## 2 Análise Inicial dos Dados

- 2.1 Origem dos dados: A origem dos dados foi obtida no site Kaggle.com, onde disponibilizam inúmeras databases gratuitas.

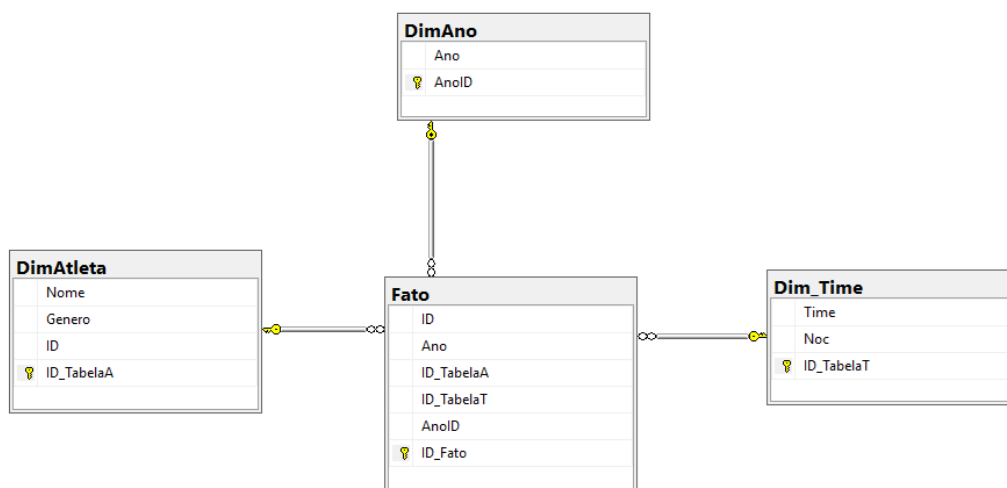
[120 years of Olympic history: athletes and results \(kaggle.com\)](https://www.kaggle.com/120years/olympic-history-athletes-and-results)

### 2.2 ANÁLISE DOS DADOS

Os dados de gênero e ano foram impostados para tabela Stage sem qualquer tratamento, com objetivo de fazer uma análise prévia dos dados.

## 3 MODELO DIMENSIONAL

### 3.1 Modelo estrela:



### 3.2 Dicionário de dados

Dim.Ano: Dimensão trazendo cada ano, de cada evento realizado

Dim\_Time: Dimensao trazendo o time e a nacionalidade da equipe.

Coluna	Tipo de dado	Descrição
Time	VARCHAR(50)	Time no qual o atleta representa
Noc	VARCHAR(50)	A abreviação do país de origem
ID_TabelaT	INT	Coluna

Dim.Atleta: Dimensão trazendo as informações de cada atleta.

Coluna	Tipo de dado	Descrição
Nome	VARCHAR(max)	Nome do atleta
Genero	VARCHAR(50)	Gênero do atleta
ID	VARCHAR(50)	Identificação de cada atleta

ID_TabelaA	INT	Identificador das participações
------------	-----	---------------------------------

### Tabela Fato:

Coluna	Tipo de dado	Descrição
Ano	VARCHAR(50)	Ano do evento
Genero	VARCHAR(50)	Gênero do atleta
ID	VARCHAR(50)	Identificação de cada atleta
NOC	VARCHAR(50)	A abreviação do país de origem
ID_Fato	INT	Identificador das participações

## 4 Carga dos dados ETL

### 4.1 Artefatos scrum

#### Backlog do produto

- 1) Desenvolver Datamart e banco de análise das olimpíadas:  
“como funcionário das olimpíadas, necessito dos dados da participação dos gêneros por data, com o objetivo de analisar o crescimento do sexo feminino nos esportes.”
- 2) Desenvolver Dashboards:  
“como funcionário das olimpíadas, necessito de dashboards para analisar efetivamente os dados recolhidos na etapa anterior.”

#### Backlog Sprint

##### 1.1 Carga Dim.Atleta

A dimensão Dim.Atleta receberá {dados de um arquivo CSV contendo todos os anos das olimpíadas} para que o product owner possa validar e analisar os dados, o que complementa ainda mais a

historia: “como funcionário das olimpiadas, necessito dos dados da participação dos gêneros por data, com o objetivo de analisar o crescimento do sexo feminino nos esportes.”

### 1.2 Carga Dim\_Time

A dimensão Dim\_Time receberá {dados de um arquivo CSV contendo todos os anos das olimpiadas} para que o product owner possa validar e analisar os dados, o que complementa ainda mais a história: “como funcionário das olimpiadas, necessito dos dados da participação dos gêneros por data, com o objetivo de analisar o crescimento do sexo feminino nos esportes.”

### 1.3 carga DimAno

A dimensão DimAno receberá {dados de um arquivo CSV contendo todos os anos das olimpiadas} para que o product owner possa validar e analisar os dados, o que complementa ainda mais a história: “como funcionário das olimpiadas, necessito dos dados da participação dos gêneros por data, com o objetivo de analisar o crescimento do sexo feminino nos esportes.”

## 5) Documentação ETL

### 5.1 Tabelas

Tabela Dim\_Time

Coluna origem	Coluna destino	tipo de dado	Transformação
Team	Time	Varchar(50)	
NOC	Noc	Varchar(50)	

Tabela Dim.Atleta

Coluna origem	Coluna destino	tipo de dado	Transformação
Name	Nome	Varchar(MAX)	
Sex	Genero	Varchar(50)	
ID	ID	Varchar(50)	

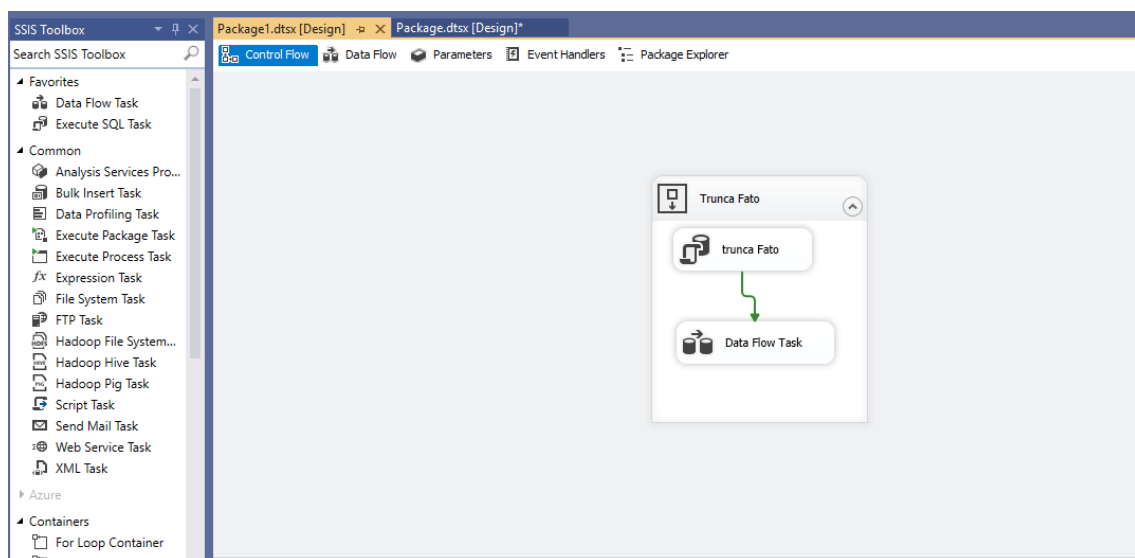
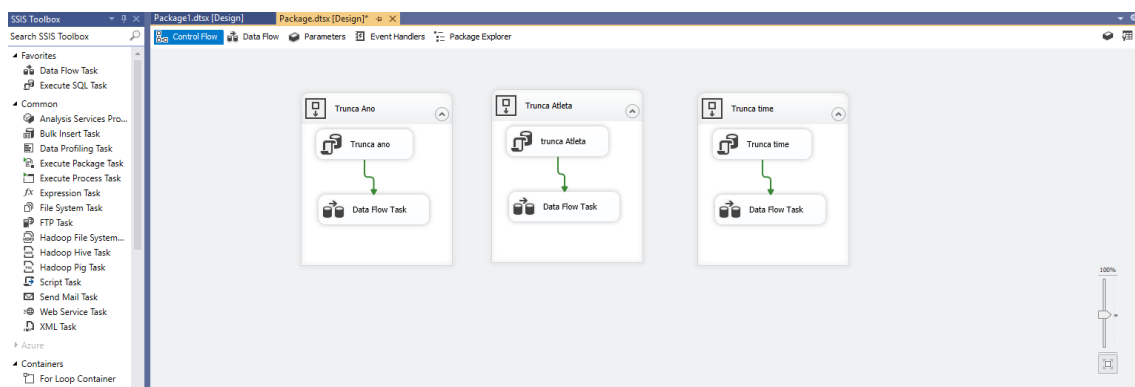


## DimAno

Coluna origem	Coluna destino	tipo de dado	Transformação
Year	Ano	Varchar(50)	

### 5.2 Carga dos dados

O projeto consiste em dois pacotes SSIS, realizando o truncamento das tabelas e a inserção dos dados nas dimensões no primeiro pacote e a carga da tabela fato no segundo, com o objetivo de manter a organização e distinção dos dados.



## 6) Banco de análises Multidimensional

Tabela Dim\_Time

Coluna origem	Coluna destino	tipo de dado	Transformação
Team	Time	Varchar(50)	
NOC	Noc	Varchar(50)	

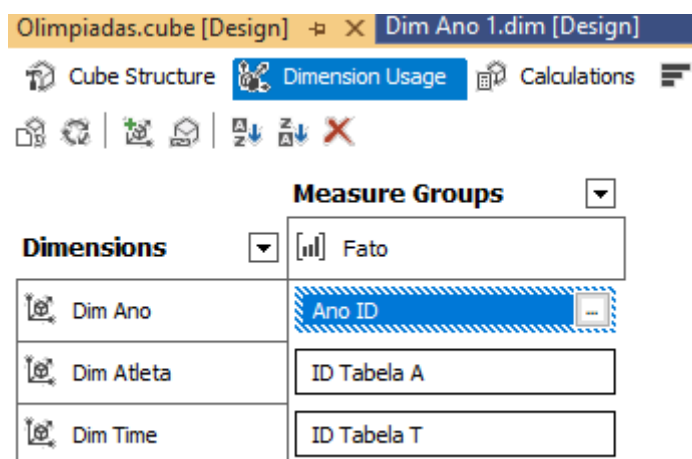
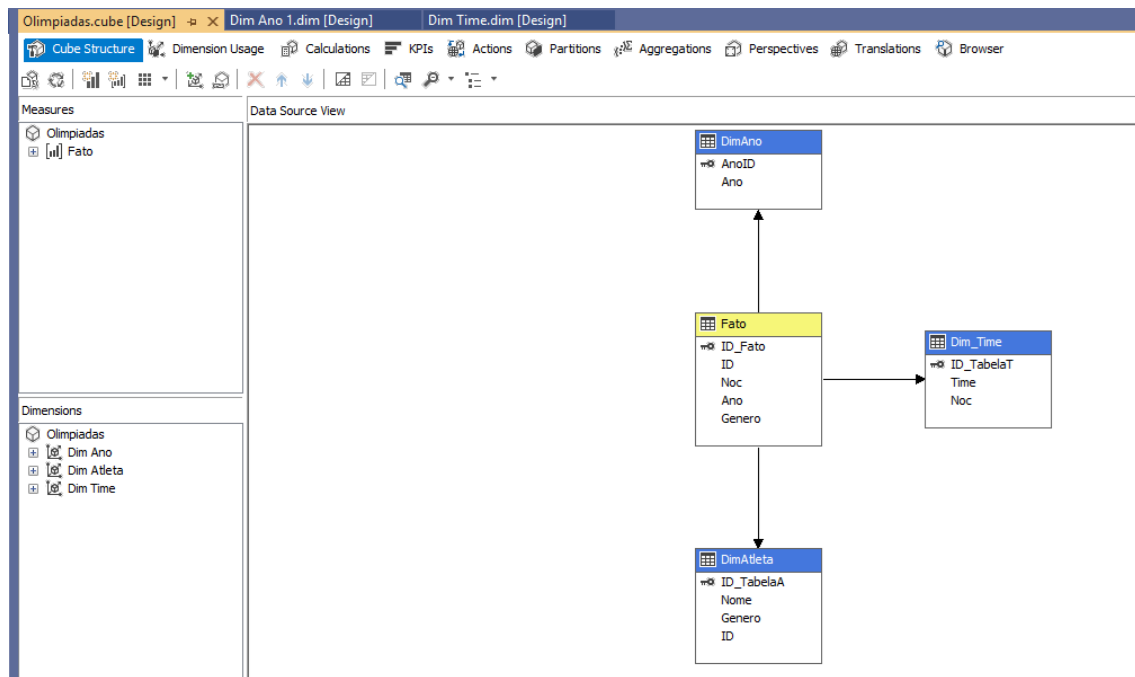
Tabela Dim.Atlela

Coluna origem	Coluna destino	tipo de dado	Transformação
Name	Nome	Varchar(MAX)	
Sex	Genero	Varchar(50)	
ID	ID	Varchar(50)	

DimAno

Coluna origem	Coluna destino	tipo de dado	Transformação
Year	Ano	Varchar(50)	




### 6.1 Estrutura do Cubo



## 7) Banco de análise Tabular

### 7.1 Relacionamentos



Model.bim*  				
[Time]				
	Time	Noc	ID_Ta...	Add Column
1	"Unite...	"USA"		3
2	"Unite...	"USA"		4
3	"Unite...	"USA"		5
4	"Unite...	"USA"		6
5	"Unite...	"USA"		7
6	"Unite...	"USA"		8
7	"Unite...	"USA"		9
8	"Unite...	"USA"		10
9	"Unite...	"USA"		11
10	"Unite...	"USA"		12
11	"Unite...	"USA"		13
12	"Unite...	"USA"		14
13	"Unite...	"USA"		49
14	"Unite...	"USA"		192
15	"Unite...	"USA"		193
16	"Unite...	"USA"		194
17	"Unite...	"USA"		195
18	"Unite...	"USA"		196
19	"Unite...	"USA"		197
20	"Unite...	"USA"		198
21	"Unite...	"USA"		199
Dim_Time		DimAno	DimAtleta	Fato 

Record:   1 of 271,116  

Model.bim\* [icon] X

[Ano] ▼

	Ano ▼	A... 07/10 ▼	Add Column
1	1988	1	
2	1988	15	
3	1988	16	
4	1988	34	
5	1988	97	
6	1988	98	
7	1988	99	
8	1988	132	
9	1988	133	
10	1988	135	
11	1988	176	
12	1988	177	
13	1988	179	
14	1988	250	
15	1988	260	
16	1988	261	
17	1988	283	
18	1988	284	
19	1988	285	

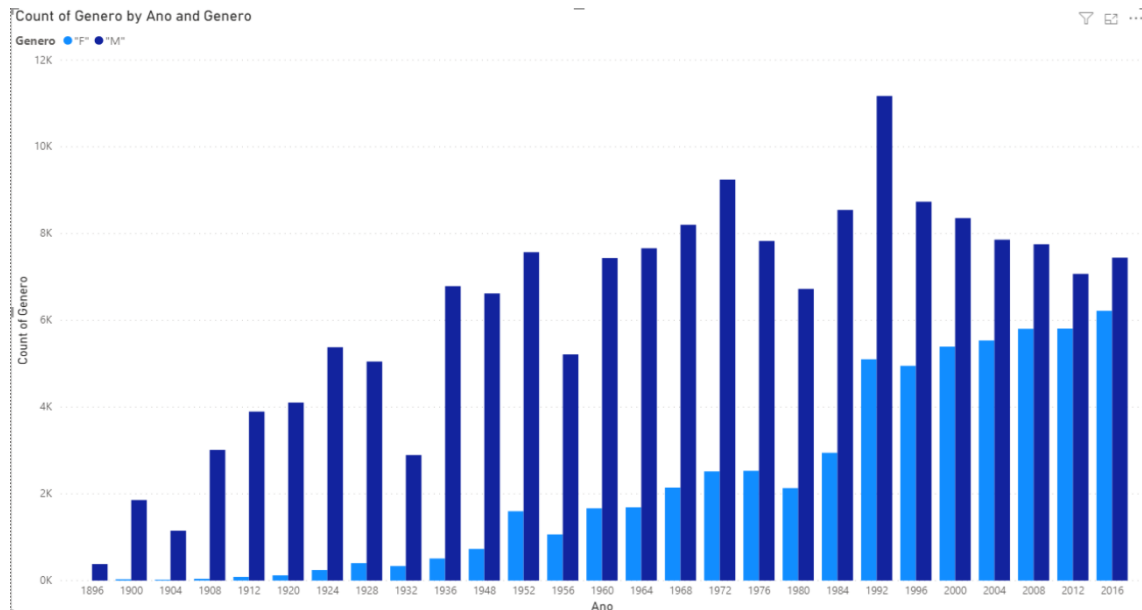
Dim\_Time DimAno DimAtleta Fato [icon]

Record: 1 of 271,116

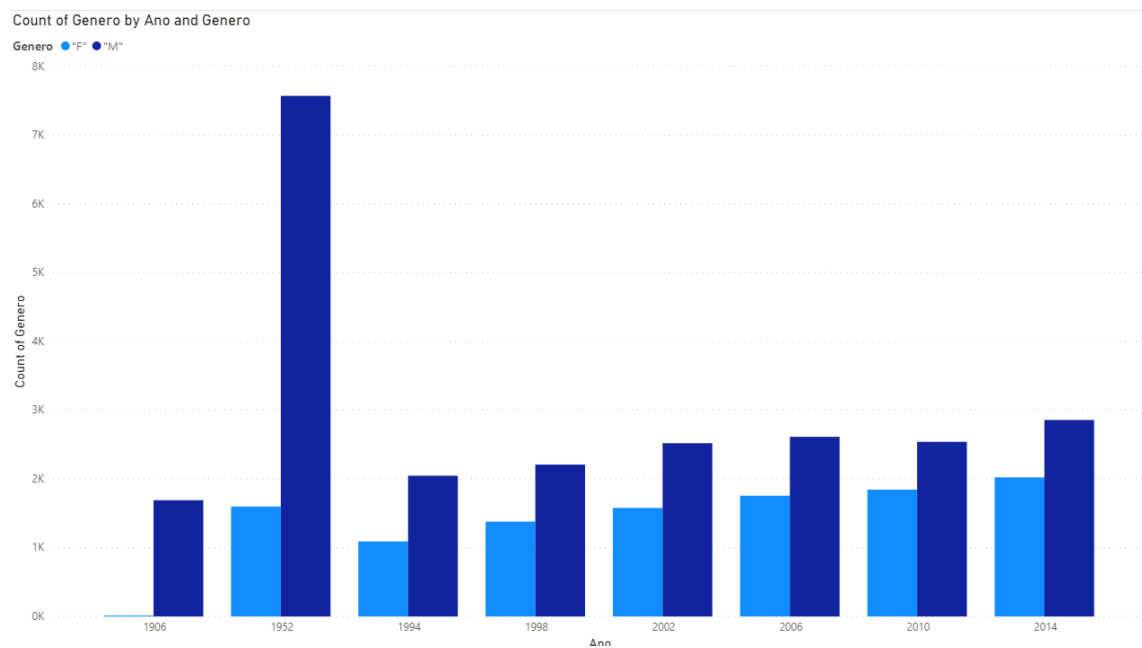


-As linhas em azul escuro representam a quantidade de participações que os homens tiveram no respectivo ano, podendo ou não ter participado de mais de uma partida no mesmo esporte ou em esportes diferentes.

## Olimpiadas de verão



## Olimpiadas de inverno





Link powerBI: [https://app.powerbi.com/links/qevbG\\_qea9?ctid=aa495f2c-d37e-4b58-ab95-bf0874a3adb2&pbi\\_source=linkShare](https://app.powerbi.com/links/qevbG_qea9?ctid=aa495f2c-d37e-4b58-ab95-bf0874a3adb2&pbi_source=linkShare)

## 9) Conclusão

O projeto Olimpíadas desenvolveu um banco de dados para analisar a quantidade de atletas que participaram de toda a trajetória das olimpíadas, podendo chegar a conclusões através dos gráficos e consultas realizadas.

Exemplo: a presença da participação do sexo feminino nos esportes ,apesar de haver um crescimento da participação dos homens, a uma crescente participação das mulheres de forma mais acentuada/acelerada do que a dos homens, indicando visualmente que a diferença se estreita com o passar dos anos.