# Data Analyst Portfolio Project – Covid 19 Analysis

## Business Problem

The challenge for this data analyst project is outlined below. This has been used continuously to ensure that the right data has been selected, transformed and used in the data visualization which is meant to be passed on to the business users.

“As a data analyst working at a news company you are asked to visualize data that will help readers understand how countries have performed historically in the summer Olympic Games.

You also know that there is an interest in details about the competitors, so if you find anything interesting then don’t hesitate to bring that in also.

The main task is still to show historical performance for different countries, with the possibility to select your own country.”

## Data Collection & Table Structures

## The necessary data was first put into a SQL database and afterwards transformed using the transformations that you can see below.

## **Olympic Games View**

## SELECT

## [ID]

## ,[Name] AS 'Competitor Name' -- Renamed Column

## ,CASE WHEN SEX = 'M' THEN 'Male' ELSE 'Female' END AS Sex -- Better name for filters and visualizations

## ,[Age]

## ,CASE WHEN [Age] < 18 THEN 'Under 18'

## WHEN [Age] BETWEEN 18 AND 25 THEN '18-25'

## WHEN [Age] BETWEEN 25 AND 30 THEN '25-30'

## WHEN [Age] > 30 THEN 'Over 30'

## END AS [Age Grouping]

## ,[Height]

## ,[Weight]

## ,[NOC] AS 'Nation Code' -- Explained abbreviation

## -- ,CHARINDEX(' ', Games)-1 AS 'Example 1'

## -- ,CHARINDEX(' ', REVERSE(Games))-1 AS 'Example 2'

## ,LEFT(Games, CHARINDEX(' ', Games) - 1) AS 'Year' -- Split column to isolate Year, based on space

## -- ,RIGHT(Games,CHARINDEX(' ', REVERSE(Games))-1) AS 'Season' -- Split column to isolate Season, based on space

## -- ,[Games]

## -- ,[City] -- Commented out as it is not necessary for the analysis

## ,[Sport]

## ,[Event]

## ,CASE WHEN Medal = 'NA' THEN 'Not Registered' ELSE Medal END AS Medal -- Replaced NA with Not Registered

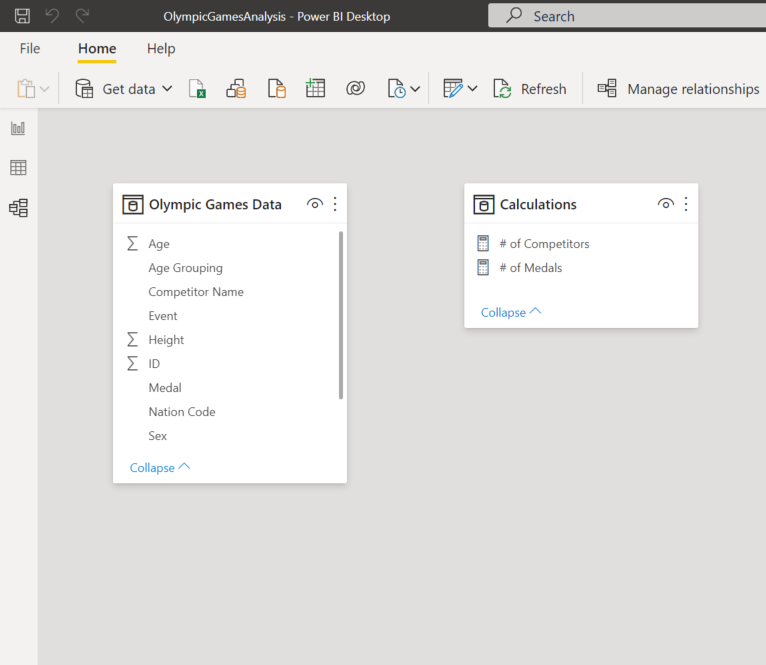
## FROM [olympic\_games].[dbo].[athletes\_event\_results]

## WHERE RIGHT(Games,CHARINDEX(' ', REVERSE(Games))-1) = 'Summer' -- Where Clause to isolate Summer Season

## Data Model

As this is a view where dimensions and facts have been combined, the data model that is created in Power BI is one table. The query from previous step was loaded in directly.

Afterwards I created a new table to isolate the calculations created in Power BI that I will explain in the next section.



## Calculations

The following calculations were created in the Power BI reports using DAX (Data Analysis Expressions). To lessen the extent of coding, the re-use of measures (measure branching) was emphasized:

# of Competitors = DISTINCTCOUNT('Olympic Games Data'[ID])

# of Medals =

CALCULATE (

COUNTROWS('Olympic Games Data'),

FILTER (

'Olympic Games Data',

'Olympic Games Data'[Medal] = "Bronze"

|| 'Olympic Games Data'[Medal] = "Silver"

|| 'Olympic Games Data'[Medal] = "Gold"

)

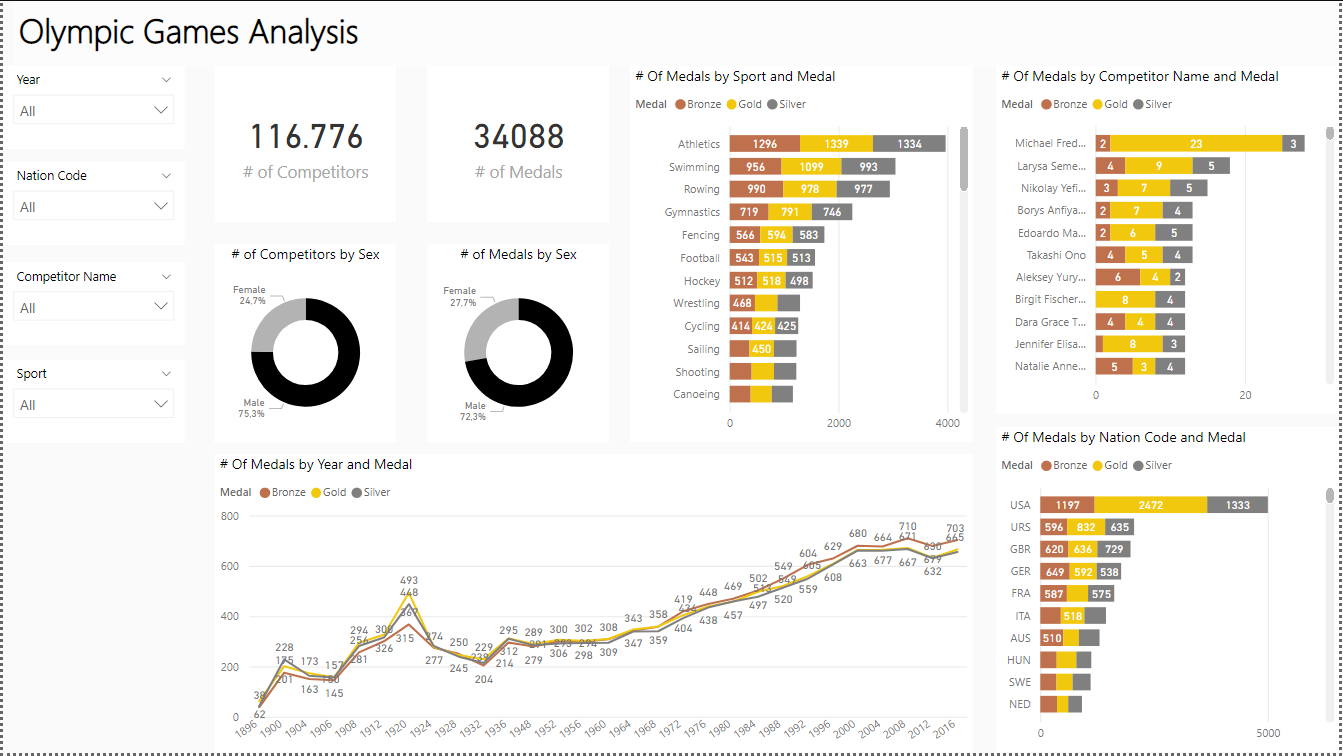
)

## Olympic Games Analysis

The finished dashboard consists of visualizations and filters that gives an easy option for the end users to navigate the summer games through history. Some possibilities are to filter by period using year, nation code to focus on one country or search for either a competitor or specific sports over time.

All the dashboards are interactive with each other so if you click on a specific sport, competitor, country or gender in their respective dashboards all the others adjust to focus on that information.

**Click the picture to open the dashboard and try it out!**

[](https://app.powerbi.com/view?r=eyJrIjoiMTUzZTZjYTMtMjNjMi00NzQyLWI5ZjgtNTQxMzQ3NjM0YTA0IiwidCI6IjBiZmE4NTAwLWIxZjItNDU2Ni1iYWYxLTZmNTkzNzA4OTNlNyIsImMiOjh9)