Olympic Gold Medals and Country Average Height

**Data Sources**

Both csv files with common countries:

* Height of Male and Female by Country 2022
  + <https://www.kaggle.com/majyhain/height-of-male-and-female-by-country-2022>
* Olympic Medals 2022
  + <https://data.world/govind6991/1-bi-project>

**Transformation**

* Join on country name left join using Olympic Medals 2022 at base data
* Cleaning will include setting country to id and renaming some country name values so they match
* It may be wise to aggregate a third column which is average country height so no men/female distinction. Olympic data doesn’t have male/female distinction

**Load into Postgres -> Relational**

**Final production load to database**

* Olympic medals data with “country” “gold” “silver” bronze” “total”
* Height data with “country” “Average male height” “average female height” “average height
  + MERGE datasets on country for final product

**Timeline:**

Thursday **–** complete and finalise code

Saturday – report writing