

EXCEL DATA CLEANING STEPS

Data Cleaning Step	Dataset Name	Description	Reason	Method	Result
1	Events	Deleted "sport_url" column	Had problems importing entire table into PgAdmin4	Manual deletion	1 column deleted
2	Athletes	Deleted the following columns: "current", "name_short", "name_tv", "function", "country_code", "country", "nationality", "nationality_code", "height", "weight", "disciplines", "events", "birth_place", "birth_country", "residence_place", "residence_country", "nickname", "hobbies", "occupation", "education", "family", "lang", "coach", "reason", "hero", "influence", "philosophy", "sporting_relatives", "ritual", "other sports".	No need for these in the data analysis stage.	Manual deletion	30 columns deleted

POSTGRESQL DATA CLEANING AND TRANSFORMATION STEPS

Data Cleaning Step	Dataset Name	Description	Reason	Method	Result
1	Events	Removed Tag column	Same as "Sport" column, but less presentable than "Sport" column.	Used ALTER TABLE statement and DROP COLUMN command.	1 column removed.
2	Events	Removed "Events URL" column	Contains abbreviations of sports that are not necessary.	Used ALTER TABLE statement and DROP COLUMN command.	1 column removed.
3	Medallists	Removed the following columns: "medal_code", "country_code", "country", "nationality_code", "nationality", "team", "team_gender", "event_type", "url_event", "code_team", "is_medallist".	Columns are not necessary for analysis.	Used ALTER TABLE statement and DROP COLUMN command.	11 columns removed.
4	Medals	Removed the following columns:	Columns are not necessary for analysis.	Used ALTER TABLE statement and DROP COLUMN command.	7 columns removed.

		"medal_code", "event_type", "url_event", "code", "country", "gender" and "country_long".			
5	Medals	Removed "medal" from all medals, so that "Bronze", "Silver" and "Gold" only appear in this column.	For better presentation.	Used UPDATE statement.	All medals appear as either "Bronze", "Silver" or "Gold".
6	Athletes	Inserted the appropriate nationality for those with such data missing.	Null values can distort the analysis.	Used the UPDATE statement.	3 records updated.
7	Athletes	Included correct name for athlete nicknamed "671."	To ensure all names appear in alphabetic format.	Used the UPDATE statement.	1 record updated.
8	Athletes	Renamed "country_long" column to "country".	To have column names that are simpler to read.	Used ALTER TABLE statement and RENAME COLUMN command.	1 column name updated.
9	Athletes	Renamed "nationality_long" column to "nationality".	To have column names that are simpler to read.	Used ALTER TABLE statement and RENAME COLUMN command.	1 column name updated.
10	Medals Table	Removed "country_code" and "country" columns.	These columns were not necessary to keep.	Used ALTER TABLE statement and DROP COLUMN command.	2 columns removed.