

DATA SCIENCE PROJECT: DASHBOARD

Name - Noordeep Kaur

Group - 3CS6

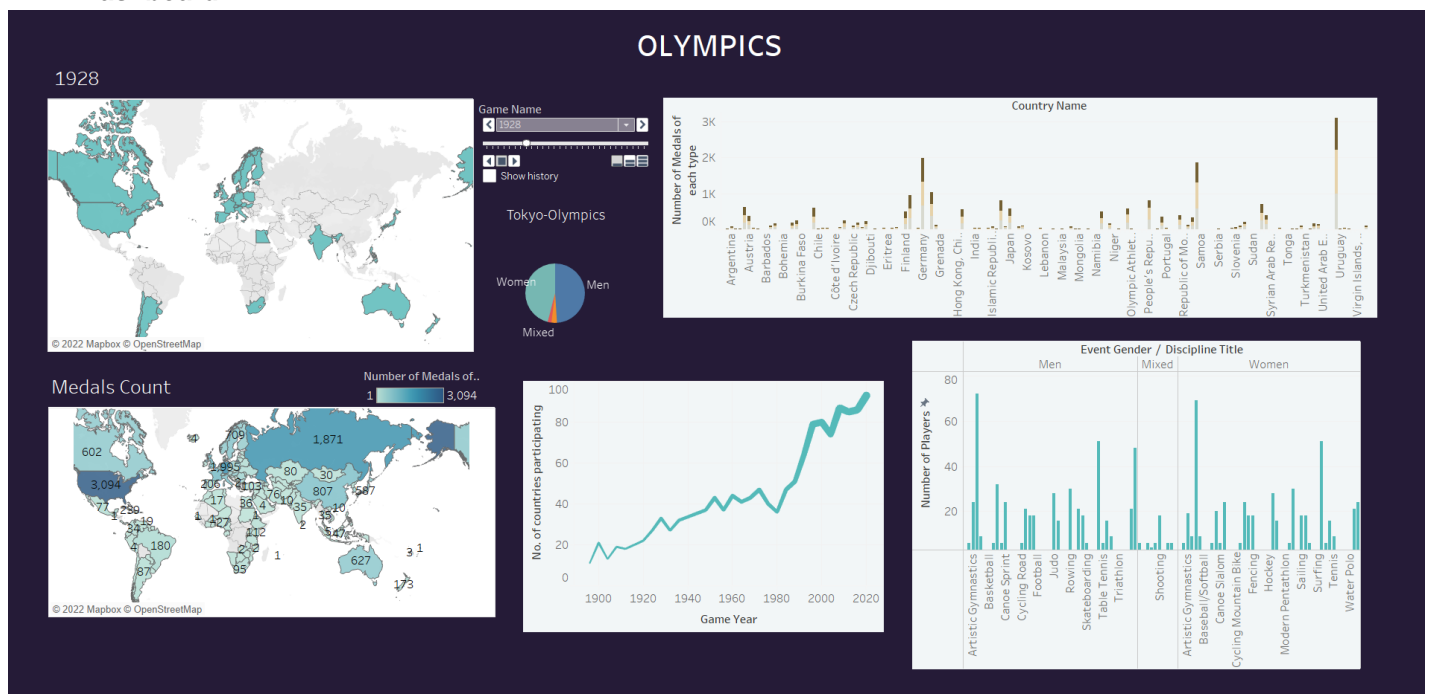
Roll No.- 102017145

I have made a Dashboard on Olympics Data taken from Kaggle. It had 4 CSV files named as follows:

Olympic_athletes, olympic_hosts, olympic_results and olympic_medals, which I have cleaned and Refined using Tableau Prep.

The final dashboard was made from 4 files GamePlayer Count, Hosts, MedalCountry, and Tokyo-2020.

Dashboard:



Initial Datasets:

olympic_medals12 fields | Filter Values...

Clear the check box to remove fields. You can also filter your data or change data types. [Add a clean step](#) to view and clean data.

Fields selected: 10 of 12

<input type="checkbox"/>	Type	Field Name	Changes	Preview
<input checked="" type="checkbox"/>	Abc	discipline_title		Curling
<input checked="" type="checkbox"/>	Abc	slug_game		beijing-2022
<input checked="" type="checkbox"/>	Abc	event_title		Mixed Doubles
<input checked="" type="checkbox"/>	Abc	event_gender		Mixed
<input checked="" type="checkbox"/>	Abc	medal_type		GOLD, SILVER
<input checked="" type="checkbox"/>	Abc	participant_type		GameTeam
<input checked="" type="checkbox"/>	Abc	participant_title		Italy, Norway
<input type="checkbox"/>	Abc	athlete_url	<input type="checkbox"/>	https://olympics.com/en/athletes/stefar
<input checked="" type="checkbox"/>	Abc	athlete_full_name		Stefania CONSTANTINI, Amos MOSANEI
<input checked="" type="checkbox"/>	Abc	country_name		Italy, Norway
<input type="checkbox"/>	Abc	country_code	<input type="checkbox"/>	IT, NO
<input checked="" type="checkbox"/>	Abc	country_3_letter_code		ITA, NOR

olympic_athletes7 fields | Filter Values...

Clear the check box to remove fields. You can also filter your data or change data types. [Add a clean step](#) to view and clean data.

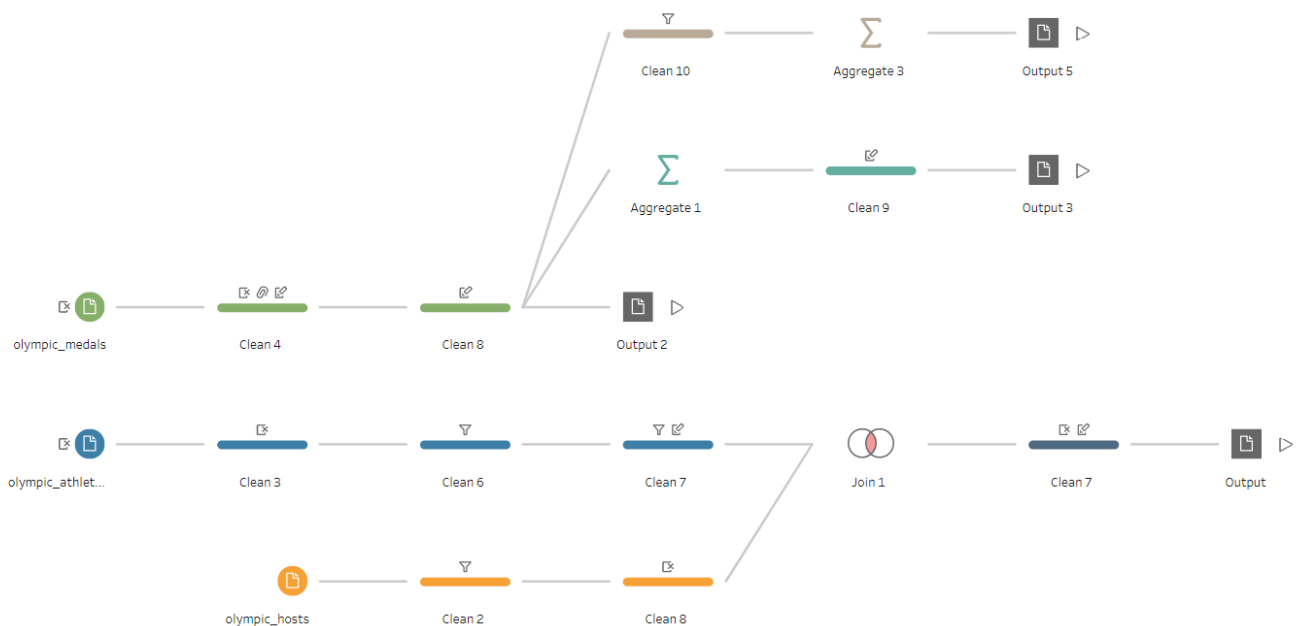
Fields selected: 6 of 7

<input type="checkbox"/>	Type	Field Name	Changes	Preview
<input type="checkbox"/>	Abc	athlete_url	<input type="checkbox"/>	https://olympics.com/en/athletes/coope
<input checked="" type="checkbox"/>	Abc	athlete_full_name		Cooper WOODS-TOPALOVIC, Felix ELOFS
<input checked="" type="checkbox"/>	#	games_participations		1, 2
<input checked="" type="checkbox"/>	Abc	first_game		Beijing 2022, PyeongChang 2018
<input checked="" type="checkbox"/>	#	athlete_year_birth		2,000, 1,995, 1,993
<input checked="" type="checkbox"/>	Abc	athlete_medals		null
<input checked="" type="checkbox"/>	Abc	bio		null

olympic_hosts 7 fields Filter Values...				
Clear the check box to remove fields. You can also filter your data or change data types. Add a clean step to view and clean data.				
Fields selected: 7 of 7				
<input checked="" type="checkbox"/>	Type	Field Name	Changes	Preview
<input checked="" type="checkbox"/>	Abc	game_slug		beijing-2022, tokyo-2020, pyeongchang
<input checked="" type="checkbox"/>	📅	game_end_date		02/20/2022, 12:00:00 PM, 08/08/2021,
<input checked="" type="checkbox"/>	📅	game_start_date		02/04/2022, 03:00:00 PM, 07/23/2021,
<input checked="" type="checkbox"/>	Abc	game_location		China, Japan, Republic of Korea
<input checked="" type="checkbox"/>	Abc	game_name		Beijing 2022, Tokyo 2020, PyeongChang
<input checked="" type="checkbox"/>	Abc	game_season		Winter, Summer
<input checked="" type="checkbox"/>	#	game_year		2,022, 2,020, 2,018

Data Cleaning and Refining Steps: (Using Tableau)

Data cleaning and joining flow:



Steps:

Following are only examples as many steps were involved:

Renaming and removing fields.

Getting data ready to be used.



Clean 4 9 fields 22K rows Filter Values...

Changes (4) <

- Group Values**
[country_3_letter_code]
4 values replaced by 2 values
- Remove Field**
[participant_title]
- Rename Field**
[country_code]
From [country_3_letter_code] to [country_code]
- Group Values**
[country_name]
4 values replaced by 2 values

Filtering Data:

Clean 10 9 fields 1K rows Filter Values.

Changes (1) <

- Filter: Selected Values**
[game_name]
Keep only: "tokyo-2020"

Clean 2 7 fields 29 rows

Changes (1)

Filter: Selected Values

[game_season]

Exclude: "Winter"

Aggregating fields to make a new field:

Aggregate 13 fields 372 rows

Filter Values...

1 Recommendation

Search

Settings

Changes (0)

Additional Fields

Drag fields to aggregate or group them.

Search

Add AllRemove All

athlete_full_name

country_code

discipline_title

event_gender

event_title

game_name

participant_type

Grouped Fields

medal_type3

SILVER

BRONZE

GOLD

country_name151

Afghanistan

Algeria

Argentina

Armenia

Australasia

Australia

Austria

Azerbaijan

Bahamas

Bahrain

Barbados

Belarus

medal_type	country_name
GOLD	Burundi
GOLD	Jordan

Aggregated Fields

Number of Rows (Aggre... 111)

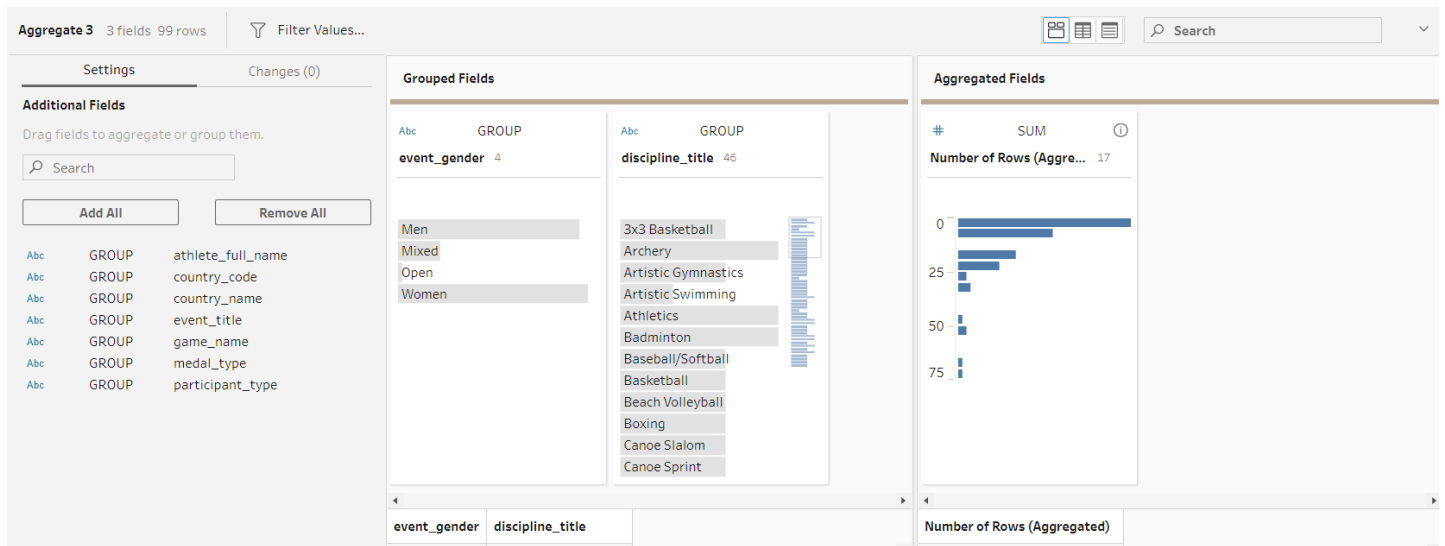
0

400

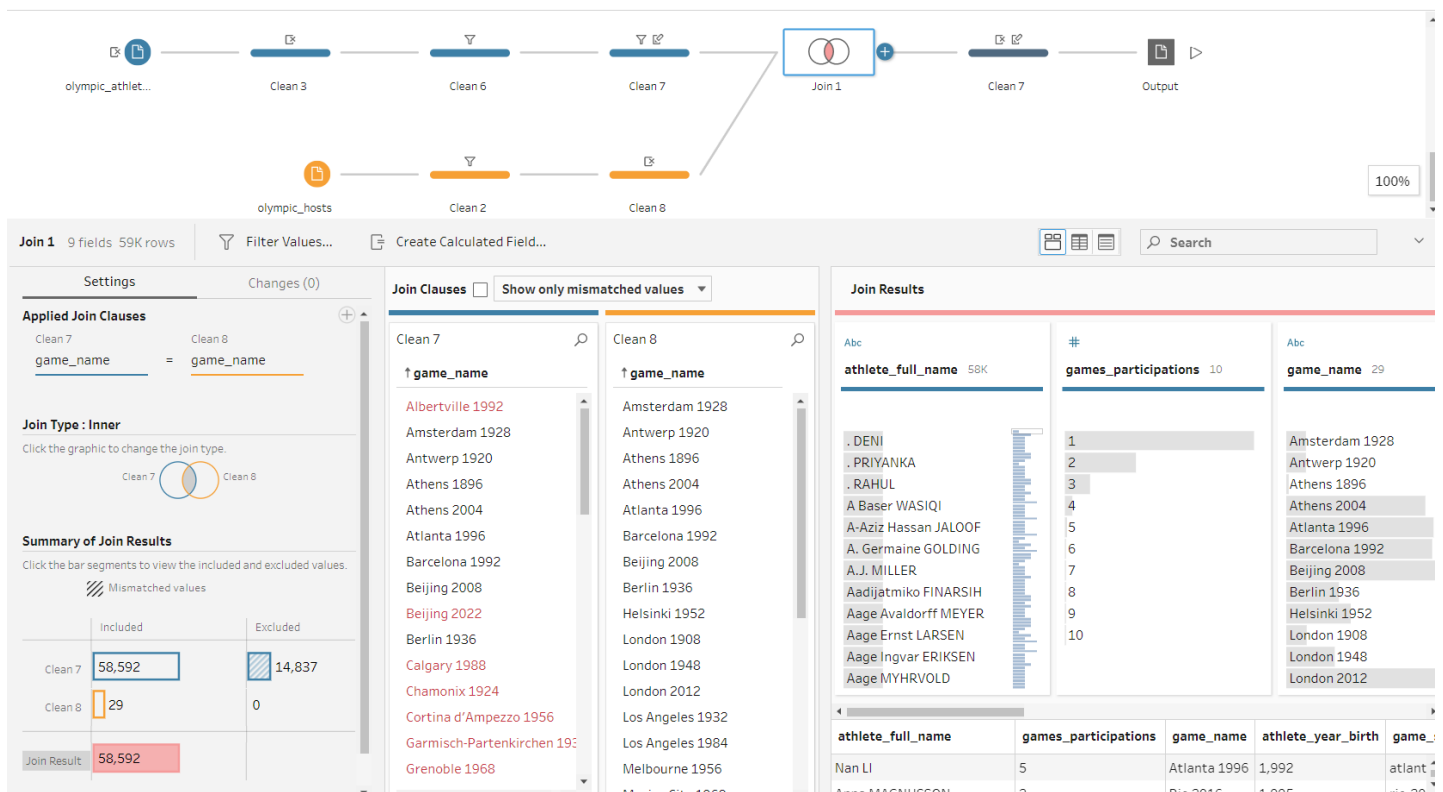
800

1,200

Number of Rows (Aggregated)
1
1



Joining Data:



Getting Output CSV files after many cleaning and refining steps:

Output 5 3 fields

Save output to

File

Browse

Name

no. of players in each game

Location

C:\Users\HOME\Desktop

Output type

Comma Separated Values (.csv)

Write Options

Select an option to create or update your output table.

Full refresh

Create table

Run Flow

Save to no. of players in each game.csv

Number of Rows (Aggregated)	event_gender	discipline_title
6	Women	Archery
3	Women	Rugby Sevens
9	Men	Tennis
3	Men	Rugby Sevens
3	Men	Triathlon
3	Men	Football
24	Men	Canoe Sprint
18	Mixed	Shooting
3	Men	Marathon Swimming
51	Women	Swimming
16	Men	Karate
6	Men	Archery
6	Men	Beach Volleyball
16	Women	Karate
21	Men	Cycling Track
21	Women	Weightlifting

Making Dashboard in Tableau:

Importing Datasets:

Hosts
Text file

Files

☐ Use Data Interpreter
Data Interpreter might be able to clean your Text file workbook.

GamePlayer Count.csv

Hosts.csv

Medals.csv

MedalsCountry.csv

Tokyo-2020.csv

New Union

New Table Extension

Hosts.csv

Medals.csv

Medals.csv

9 fields 21697 rows

100 → rows

Name
Medals.csv

Fields

Type	Field Name	Phys...	Rem...
Abc	Discipline Title	Medal...	discipl...
Abc	game name (Medals.csv)	Medal...	game_...
Abc	Event Title	Medal...	event_...

Abc Medals.csv	Abc Medals.csv	Abc Medals.csv	Abc Medals.csv	Abc Medals.csv	Abc Medals.csv
Discipline Title	game name (Medals.csv)	Event Title	Event Gender	Medal Type	Participant Type
Curling	beijing-2022	Mixed Doubles	Mixed	GOLD	GameTeam
Curling	beijing-2022	Mixed Doubles	Mixed	GOLD	GameTeam
Curling	beijing-2022	Mixed Doubles	Mixed	SILVER	GameTeam
Curling	beijing-2022	Mixed Doubles	Mixed	SILVER	GameTeam
Curling	beijing-2022	Mixed Doubles	Mixed	BRONZE	GameTeam
Curling	beijing-2022	Mixed Doubles	Mixed	BRONZE	GameTeam

Data Source

Participating Countries

Sheet 7

Sheet 4

Sheet 5

Sheet 6

Sheet 8

OLYMPICS

Data Analytics <

- GamePlayer Count
- Hosts+**
- MedalsCountry
- Tokyo- 2020

Data Analytics **Pages** Columns Rows

Game Year **Longitude (generate...)** **Latitude (generated)**

Filters

Latitude (generated) **Longitude (generate...)**

Tables

Hosts.csv

- Athlete Full Na...
- Game Location
- Game Name
- Game Season
- Game Year
- Athlete Year Bi...
- Games Parti...
- Hosts.csv (Co...

Medals.csv

- athlete full na...
- Country Name
- Discipline Title
- Event Gender
- Event Title
- game name (...)
- Medal Type
- Participant Type
- Country Code ...
- Medals.csv (C...

Measure Names

- Latitude (generat...
- Longitude (gener...
- Measure Values

Search **Map** **Color** **Size** **Label** **Detail** **Tooltip** **Country Name**

Game Name

2020

© 2022 Mapbox © OpenStreetMap

