**Description of Data**

Winter Olympics (1924 - 2018) From https://www.kaggle.com/datasets

This dataset contains the data of all the past Winter Olympic events

le Minh Nguyen

Context

As Winter Olympic 2022 is approaching, I aim to better understand the different disciplines and sports that are competed at the event, so that I can better appreciate the event overall. Thus, I have scraped the following data

Content

There are 3 CSV. The first one olympicevents.csv, include the information about the different Olympic events on a high level, with information such as name of the event, where and when it took place, number of overall participants, etc. eventsmedals.csv dive deeper into which are the countries that have won a medal, how many have they won, etc.

The last one, discipline\_details.csv, dive deeper into all the sport competed at each individual event

Acknowledgements

This data has been collected through https://olympedia.org/

Inspiration

Which countries have improved the most in recent year?

Does the location of the event, whether it is in Asia, Europe, etc. affect individual countries performance?

**Data Set Manipulations**

After loading the dataset into a data frame, I extracted the gender of the participants of each event by parsing the ‘$category’ column. I also fixed some encoding format characters with simple text transitions using string manipulation.