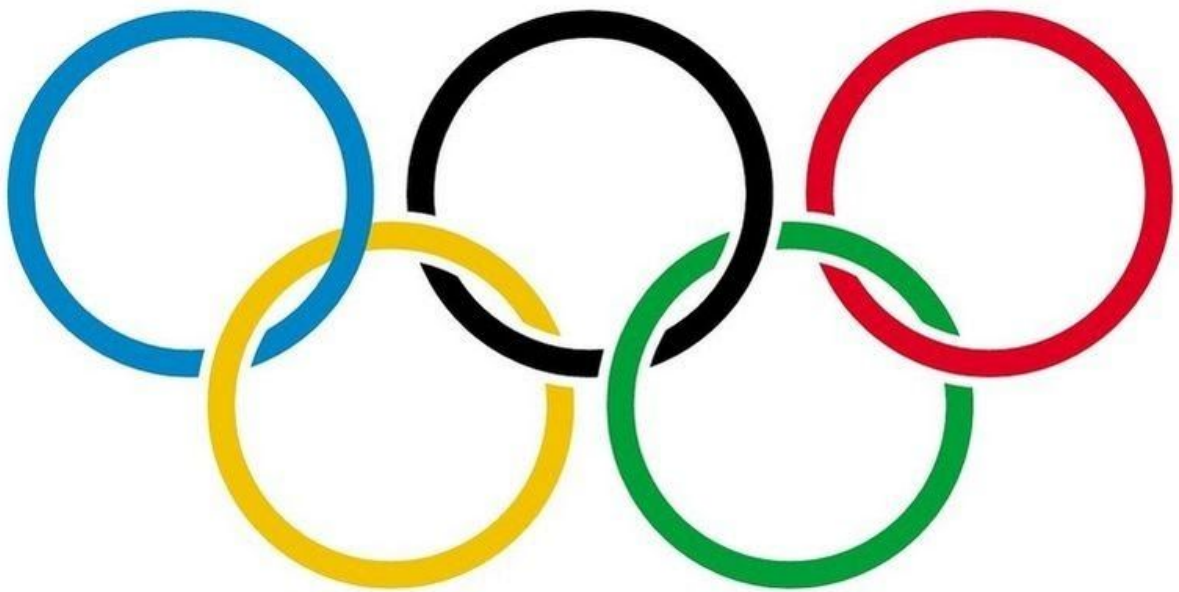


Graph Algorithm



1. Choix du dataset

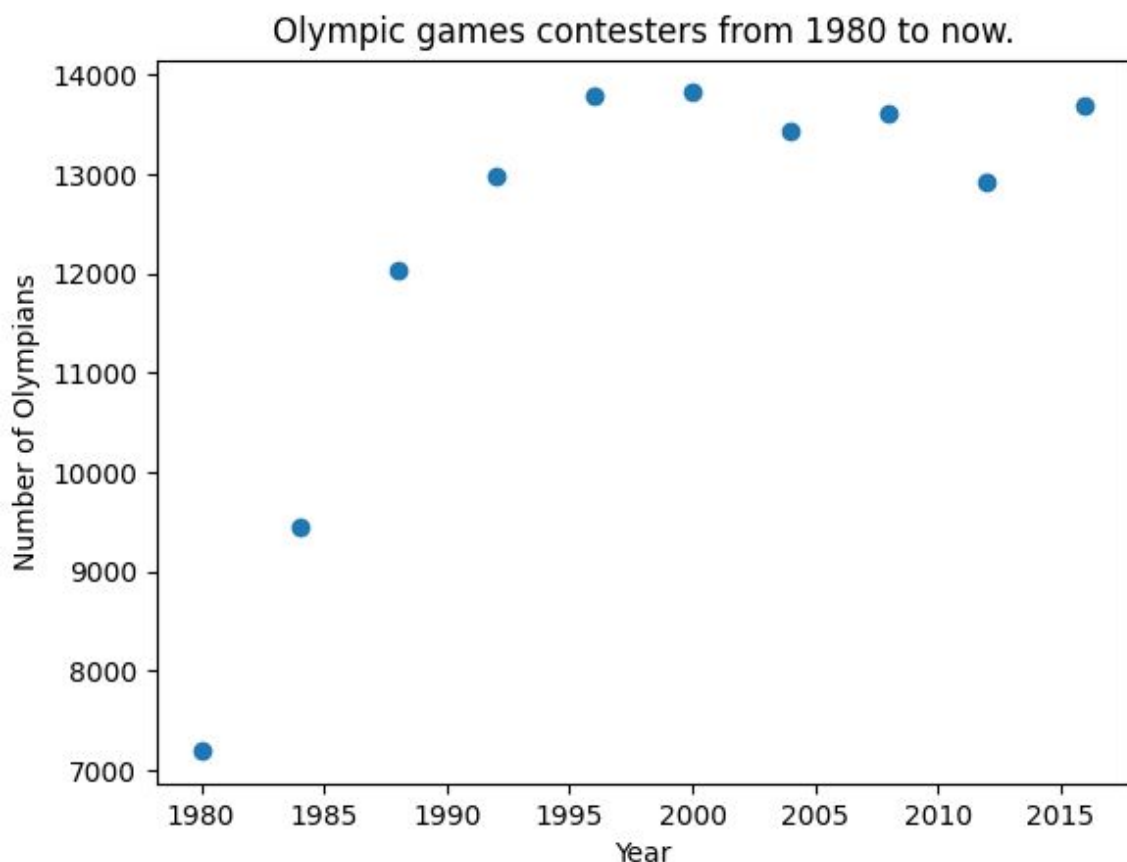
<https://www.kaggle.com/heesoo37/120-years-of-olympic-history-athletes-and-results>

Ce dataset est intéressant car celui-là répertorie tous les athlètes depuis la réorganisation des Jeux Olympiques, avec plusieurs données qui peuvent être analysées. Nous pouvons retrouver le nom de l'athlète, sa taille, son poids, sa nationalité, s'il a gagné une médaille et si oui laquelle...

2. Build metric

Pour le metric, on croise plusieurs colonnes du dataset afin d'en faire ressortir une évolution, une tendance. Dans ce cas là, nous avons quantifié le nombre de participants aux jeux d'été chaque édition depuis 1980 afin d'en juger de l'évolution depuis cette année.

Voici une représentation graphique de cette évolution générée avec la librairie Matplotlib.



On y remarque une nette augmentation du nombre de participants les cinq premières éditions étudiées (1980, 1984, 1988, 1992 et 1996) qui double presque, allant de 7191 à 13780 olympiens. Ensuite leur nombre semble se stabiliser autour des 13500.

3. Clustering

Pour le clustering, nous avons décidé d'analyser les tailles des athlètes en basketball à partir des années 1950. Bien évidemment, il peut y avoir une certaine différence entre les femmes et les hommes, alors nous avons décidé de réaliser un clustering pour les femmes et un clustering pour les hommes.

Le nombre de clusters représente les différentes tailles, c'est-à-dire taille "petite", taille moyenne et taille grande.

La première chose que l'on peut observer, même si ce n'est pas une surprise, est que les hommes sont plus grands que les femmes. Il y a peu d'hommes qui sont inférieurs à 1m80 alors que chez les femmes, la taille moyenne est à 1m85. De plus, la grande taille est prépondérante chez les hommes alors que nous avons plus de femmes de taille moyenne.