

Python for Exploratory Data Analysis (Workshop) Olimpic Games

**UJIAN AKHIR SEMESTER
(UAS)**



Disusun oleh:

**MAGISTER TEKNIK INFORMATIKA
INSTITUT INFORMATIKA & BISNIS DARMAJAYA
BANDAR LAMPUNG
2022**

1. Introduction

Dalam pembahasan ini dilakukan analisis mendalam dari kumpulan data olympic selama 120 tahun yang didalamnya berisikan biodata para atlet dan juga informasi pertandingan apa saja yang diikuti oleh atlet tersebut. Untuk melakukan analisis secara menyeluruh terkait data oimpic ini dilakukan *Exploratory Data Analysis* (EDA) yang tujuannya untuk mendapatkan karakteristik data olympic, pola yang ada yang pada olympic, serta visualisasinya. *Exploratory Data Analysis* (EDA) dapat disinkronkan dengan aplikasi seperti *machine learning* untuk melakukan analisa yang mendalam dari kumpulan data. Aplikasi *machine learning* yang mudah digunakan yaitu phyton, kelebihanannya phyton memiliki *libraries* yang mudah untuk digunakan seperti Pandas, Seaborn, Numpy, dan juga Matplotlib. Pembahasan ini dilakukan dengan menggunakan data yang nyata yaitu data historis kegiatan olympic. Data yang akan dianalisis berisikan data mulai dari tahun 1896 sampai dengan tahun 2016. Data-data ini selanjutnya akan diproses dengan cara *Correlating*, *Converting*, *Completing*, *Correcting*, *Creating* dan *Charting* data. Selain itu *Exploratory Data Analysis* (EDA) juga akan melakukan pembahasan mengenai *Jupyter Notebooks* yang berupa aplikasi *open source* yang bisa digunakan untuk mengolah berbagi dokumen yang bersikan *live code*, persamaan, visualisasi, dan juga teks naratif. Dengan mengolah data secara menyeluruh akan memberikan gambaran terkait olympic, seperti negara mana yang mendominasi penerimaan medali emas, atau bisa juga negara mana yang masih kesulitan untuk memperoleh medali.

2. Problem

Masalah yang ada adalah bagaimana agar data-data yang *missing* atau tidak ada agar bisa dimanipulasi sehingga analisa yang diperoleh menjadi lebih baik. Proses manipulasi tentunya menggunakan metode yang dapat diandalkan sehingga hasil manipulasinya juga representatif. Dari kumpulan data olympic yang ada, data yang paling banyak *missing* adalah data perolehan medali atlet, dimana dari total 271116 data entri hanya terdapat 39783 entri medali, kurang lebih hanya 14% dari total keseluruhan data. Selain memperoleh data yang *missing*, juga perlu dilakukan proses visualisasi data agar lebih mudah untuk dilakukan interpretasi data.

ID	Name	Sex	Age	Height	Weight	Team	NOC	Games	Year	Season	City	Sport	Event	Medal	
0	1	A Dijing	M	24.0	180.0	80.0	China	CHN	1992 Summer	1992	Summer	Barcelona	Basketball	Basketball Men's Basketball	NaN
1	2	A Lamusi	M	23.0	170.0	60.0	China	CHN	2012 Summer	2012	Summer	London	Judo	Judo Men's Extra-Lightweight	NaN
2	3	Gunnar Nielsen Aaby	M	24.0	NaN	NaN	Denmark	DEN	1920 Summer	1920	Summer	Antwerpen	Football	Football Men's Football	NaN
3	4	Edgar Lindenu Aabye	M	34.0	NaN	NaN	Denmark/Sweden	DEN	1900 Summer	1900	Summer	Paris	Tug-Of-War	Tug-Of-War Men's Tug-Of-War	Gold
4	5	Christine Jacoba Aaftink	F	21.0	185.0	82.0	Netherlands	NED	1988 Winter	1988	Winter	Calgary	Speed Skating	Speed Skating Women's 500 metres	NaN

Gambar 1. Data olympic dalam tabel

Terlihat pada Gambar 1 dimana dalam data olympic yang tersedia masih ada kekosongan data seperti pada tinggi dan berat badan atlet. Hal ini akan lebih baik jika dilakukan pelengkapan data sebelum membuatnya data tersebut menjadi sebuah visualisasi data. Semakin baik kumpulan data performanya dapat diukur melalui visualisasi data yang dihasilkan. Pola pola yang dihasilkan akan memberikan gambaran kualitas data yang dimiliki. Untuk melengkapi data-data yang tidak ada digunakan rata-rata data dari objek tertentu, misalnya ada data usia atlet laki-laki yang tidak lengkap maka akan digunakan rata-rata usia dari atlet laki-laki yang ada untuk melengkapi data yang ada.

3. Prerequisites

Dalam pembahasan kali ini saya beranggapan bahwa pembaca sudah mengetahui atau memahami konsep dasar terkait proses pembelajaran *machine learning* klasik dan juga *deep learning*, *data preprocessing*, *exploratory data analysis* (EDA). Jika pembaca sudah memahami konsep dasarnya akan lebih mudah untuk memahami apa yang akan dibahas selanjutnya dalam pembahasan terkait olimpik data. Hal ini tidak menjadi kewajiban hanya saja dapat lebih memudahkan, pembahasan akan lebih banyak menjelaskan visualisasi tampilan data sehingga secara umum lebih mudah untuk dipelajari bersama.

4. Data Source

Sumber data yang digunakan pada pembahasan ini diperoleh dari Github yang berjumlah 2 data, data pertama berisikan data historis olimpik dan data kedua berisikan kode negara atau panitia suatu negara beserta informasinya. Data pertama dan kedua memiliki ukuran sebesar 13,7 MB. Data pertama (*athlete events.csv*) memiliki baris data sebanyak 271116 dengan 15 kolom sedangkan data kedua (*noc_regions.csv*) memiliki baris data sebanyak 230 dengan 3 kolom. Untuk data pertama terdapat banyak kekosongan data seperti pada kolom berat, tinggi badan, dan usia atlet. Data yang diperoleh berekstensi *.csv* sehingga dapat dibuka melalui excel ataupun *machine learning*.

5. Existing Approaches

Pembahasan terkait dataset olimpiade telah dilakukan oleh Marco Giuseppe De Pinto dan Arunsankar Kumarakurubaran melalui python. Marco Giuseppe De Pinto membahas dataset olimpiade dengan judul “Let's discover more about the Olympic Games!” yang telah diriview sebanyak 40771 kali dalam waktu 4 tahun, sedangkan Arunsankar Kumarakurubaran membahas dengan judul “Key Insights from Olympic history data” yang telah diriview sebanyak 6290 kali dalam waktu yang sama yaitu 4 tahun. Keduanya mengulik lebih dalam dataset olimpiade dengan memvisualisasikannya ke dalam bentuk tabel dan juga diagram, ada juga yang divisualisasikan dalam bentuk heat map, scatter plot, box plot, point plot, dan juga bar plot. Pembahasan keduanya dilakukan dengan rinci dan jelas dan disertakan juga coding untuk visualisasi datasetnya sehingga memudahkan penulis berikutnya untuk melakukan modifikasi terkait visualisasi yang diinginkannya.

6. Understanding the Data

Setiap baris data pada file *athlete events.csv* mewakili kolom data yang ada pada dataset. Dataset olimpiade terdiri dari 15 kolom data dan 271116 baris data, 15 kolom data tersebut berupa:

- ID - Unique number for each athlete,

ID sebagai tanda pengenalan masing-masing atlet, setiap atlet memiliki ID yang berbeda. Jika 1 atlet mengikuti berbagai cabang olahraga maka IDnya tetap sama dalam dataset.

- Name - Athlete's name,

Nama atlet sebagai pengenal lainnya untuk para atlet. Setiap kali atlet mengikuti olimpiade pada tahun yang berbeda atau pada tahun yang sama dengan jumlah cabang olahraga yang berbeda maka IDnya akan tetap sama.

- Sex - M or F,

Kolom ini merupakan jenis kelamin atlet, terdiri dari M yaitu Male berarti laki-laki dan F yaitu Female berarti wanita. Setiap ID dan nama atlet yang sama maka kolom Sexnya juga akan sama.

- Age – Integer,

Kolom ini berisikan usia atlet, setiap ID dan nama atlet bisa sama namun untuk usia bisa berbeda karena usia atlet bertambah ketika ia mengikuti olimpiade berikutnya yang berjarak 4 tahun maka usianya di olimpiade berikutnya akan bertambah.

- Height – in centimeters,

Kolom ini berisikan data tinggi badan atlet, setiap ID dan nama atlet akan memiliki kolom height yang sama. Tinggi badan atlet dalam dataset akan sama pada setiap IDnya meskipun berbeda olimpiade.

- Weight – in kilograms,

Kolom ini berisikan berat badan atlet, sama halnya dengan berat badan setiap ID dan nama atlet akan memiliki berat badan yang sama. Meskipun berbeda penyelenggaraan olimpiade berat badan atlet akan tetap sama.

- Team – Team Name

Team berisikan asal negara setiap atlet. ID dan nama atlet yang sama juga akan memiliki Team atau asal negara yang sama. Jika ada atlet yang mewakili 2 negara maka akan ditulis dengan tanda strip (/).

- NOC - National Olympic Committee 3-letter code,

NOC berisikan kode untuk panitia negara tersebut. NOC berisikan 3 huruf yang melambangkan suatu negara misal Argentina: ARG dan France: FRA.

- Games – Year and season,

Kolom Game Year dan Season merupakan 3 kolom data yang berbeda. Kolom Game berisikan gabungan antara Year dan Season, sedangkan Year berisikan tahun terselenggaranya olimpiade dan Season berisikan terselenggaranya olimpiade pada musim panas atau musim dingin.

- City – Host city,

City berisikan tempat berlangsungnya olimpiade. Setiap ID dan nama atlet akan berbeda beda kolom citynya karena olimpiade yang diikuti juga berbeda beda. City

berbeda dengan negara, kolom city berisikan nama kota terselenggaranya olimpiade bukan nama negaranya.

- Sport – Sport,

Kolom sport berisikan cabang induk olahraga yang diikuti oleh atlet. Untuk kolom sport setiap ID dan nama atlet akan memiliki kolom sport yang sama meskipun berbeda tahun berlangsungnya olimpiade.

- Event – Event,

Kolom event berisikan cabang dari induk olahraga yang diikuti oleh masing-masing atlet. Berbeda dengan kolom sport yang sama, untuk kolom event berisikan value yang berbeda beda tergantung atlet mengikuti cabang induk olahraga yang diikutinya, misalnya induk olahraga renang memiliki cabang renang 100m gaya bebas, 100m gaya dada, dan sebagainya.

- Medal – Gold, Silver, Bronze, N/A

Kolom medal berisikan medali yang diperoleh oleh atlet pada masing-masing event olahraga yang ia ikuti. Kolom medal dapat berupa Gold jika atlet meraih emas, Silver jika atlet meraih perak, Bronze jika atlet meraih perunggu, dan N/A jika atlet tidak mendapatkan medali.

Sedangkan untuk file noc regions.csv yang berisikan 230 baris data dengan 3 kolom data diantaranya terdiri dari:

- NOC,

NOC berisikan singkatan 3 huruf dari suatu negara untuk memudahkan penulisan nama negara ketika bertanding tanpa harus menulis namanya secara keseluruhan.

- Region – name of country,

Kolom region berisikan nama lengkap dari suatu negara yang mengikuti olimpiade. Berbeda dengan NOC yang hanya berisikan 3 huruf pada kolom region nama negara ditulis dengan lengkap.

- Notes - String containing any useful information about region and NOC

Kolom notes berisikan informasi tambahan dari suatu nama negara yang terdengar asing. Informasi ini menjelaskan bahwa negara tersebut merupakan jajahan dari suatu negara yang bersangkutan atau menggunakan bendera yang sama.

7. Exploratory Data Analysis

- **Importing the modules**

Hal yang dilakukan pertama adalah mengimport module yang akan digunakan untuk menganalisis dataset. Module yang diimport adalah numpy, pandas, seaborn, dan matplotlib.

```
In [1]:
import numpy as np
import pandas as pd
import seaborn as sns
from matplotlib import pyplot as plt

import os
print(os.listdir("../input"))

['athlete_events.csv', 'noc_regions.csv']
```

- **Data Importing**

Selanjutnya yang dilakukan adalah dengan mengimport dataset ke dalam python. File yang digunakan adalah athlete events.csv dan juga noc regions.csv.

```
data = pd.read_csv('../input/athlete_events.csv')
regions = pd.read_csv('../input/noc_regions.csv')
```

- **Collecting information about the two dataset**

Kemudian langkah berikutnya yang akan dilakukan adalah meriview baris data pertama dan melakukan beberapa fungsi untuk mendapatkan informasi statistik dataset, tipe dataset, nama-nama kolom pada dataset, dan juga informasi lainnya.

ID	Name	Sex	Age	Height	Weight	Team	NOC	Games	Year	Season	City	Sport	Event	Medal
0	A. Djiang	M	24.0	180.0	80.0	China	CHN	1992 Summer	1992	Summer	Barcelona	Basketball	Basketball Men's Fastball	NaN
1	A. Lamusi	M	23.0	170.0	60.0	China	CHN	2012 Summer	2012	Summer	London	Judo	Judo Men's Extra-Lightweight	NaN
2	Gunnar Nielson Astly	M	24.0	NaN	NaN	Denmark	DNK	1920 Summer	1920	Summer	Antwerpen	Football	Football Men's Football	NaN
3	Elger Lindqvist Astly	M	34.0	NaN	NaN	Denmark/Sweden	DNK	1900 Summer	1900	Summer	Paris	Tug-Of-War	Tug-Of-War Men's Tug-Of-War	Gold
4	Christine Jacobs Aalfink	F	21.0	185.0	82.0	Netherlands	NED	1988 Winter	1988	Winter	Calgary	Speed Skating	Speed Skating Women's 500 metres	NaN

Perintah yang dilakukan adalah menampilkan 5 data teratas dan munculnya gambar diatas, ID sudah sesuai 1-5. Penomeran atlet tidak berdasarkan urutan alphahabet namun dilakukan secara random.

	ID	Age	Height	Weight	Year
count	271116.000000	261642.000000	210945.000000	208241.000000	271116.000000
mean	68248.954396	25.556808	175.338970	70.702393	1978.378460
std	39022.286345	6.393561	10.518462	14.348020	29.877632
min	1.000000	10.000000	127.000000	25.000000	1896.000000
25%	34643.000000	21.000000	168.000000	60.000000	1960.000000
50%	68205.000000	24.000000	175.000000	70.000000	1988.000000
75%	102097.250000	28.000000	183.000000	79.000000	2002.000000
max	135571.000000	97.000000	226.000000	214.000000	2016.000000

Selanjutnya dilakukan perintah untuk menampilkan statistik dataset. Terlihat pada gambar diatas bahwa jumlah ID sebanyak 271116, sedangkan untuk Age, Height, dan Weight jumlahnya tidak ada yang seperti ID kecuali tahun dimana jumlahnya sama yaitu 27116 data. Dataset dilakukan perhitungan rata-rata, standar deviasi, nilai minimal dan maksimal, kuartil 25, kuartil 50, dan kuartil 75. Ada hal menarik dalam tampilan statistik dimana usia paling kecil dari peserta yang mengikuti olimpiade adalah 10 tahun dan yang tertua 97 tahun.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 271116 entries, 0 to 271115
Data columns (total 15 columns):
ID          271116 non-null int64
Name        271116 non-null object
Sex         271116 non-null object
Age         261642 non-null float64
Height      210945 non-null float64
Weight      208241 non-null float64
Team        271116 non-null object
NOC         271116 non-null object
Games       271116 non-null object
Year        271116 non-null int64
Season      271116 non-null object
City        271116 non-null object
Sport       271116 non-null object
Event       271116 non-null object
Medal       39783 non-null object
dtypes: float64(3), int64(2), object(10)
memory usage: 31.0+ MB
```

Selanjutnya dilakukan pengecekan kelengkapan data dari masing masing kolom data, entri data dimulai dari 0 sampai 271116. Hampir semua data lengkap kecuali kolom usia, tinggi badan, berat badan, dan perolehan medali. Untuk perolehan medali pasti tidak akan lengkap karena hanya juara 1, 2, dan 3 yang mendapatkan medali, jadi akan lebih banyak jumlah atlet yang tidak memperoleh medali pada setiap gelaran olimpiade.

	NOC	region	notes
0	AFG	Afghanistan	NaN
1	AHO	Curacao	Netherlands Antilles
2	ALB	Albania	NaN
3	ALG	Algeria	NaN
4	AND	Andorra	NaN

Perintah selanjutnya adalah untuk menampilkan data teratas dari file noc region. Hasilnya seperti pada gambar diatas dimana data NOC tersusun berdasarkan alfabet mulai dari a sampai z. Berbeda jika dibandingkan dengan susunan data pada file athlete events yang tersusun secara random. Kolom region berisikan nama lengkap suatu negara peserta olimpiade dan kolom notes berisikan informasi tambahan suatu negara yang bersangkutan. Misalnya negara tersebut merupakan negara lainnya penggunaan benderanya sama dengan negara yang memiliki kepemilikan atas negara tersebut.

- **Joining the dataframes**

ID	Name	Sex	Age	Height	Weight	Team	NOC	Games	Year	Season	City	Sport	Event	Medal	region	notes
1	A Diliang	M	24.0	180.0	80.0	China	CHN	1992 Summer	1992	Summer	Barcelona	Basketball	Basketball Men's Basketball	NaN	China	NaN
2	A Lamusi	M	23.0	170.0	80.0	China	CHN	2012 Summer	2012	Summer	London	Judo	Judo Men's Extra-Lightweight	NaN	China	NaN
3	Gunnar Nielsen Aaby	M	24.0	NaN	NaN	Denmark	DEN	1920 Summer	1920	Summer	Antwerpen	Football	Football Men's Football	NaN	Denmark	NaN
4	Edgar Lindenau Aabye	M	34.0	NaN	NaN	Denmark/Sweden	DEN	1900 Summer	1900	Summer	Paris	Tug-Of-War	Tug-Of-War Men's Tug-Of-War	Gold	Denmark	NaN
5	Christine Jacobs Aaftink	F	21.0	185.0	82.0	Netherlands	NED	1988 Winter	1988	Winter	Calgary	Speed Skating	Speed Skating Women's 500 metres	NaN	Netherlands	NaN

Perintah lainnya yang bisa dilakukan adalah menggabungkan kedua file dataset menjadi satu tabel. Hasilnya seperti pada gambar diatas dimana data NOC, region, dan notes tergabung ke dalam file athlete events. Untuk lebih memudahkan proses analisis akan lebih baik jika kedua data tersebut dilakukan penggabungan seperti agar tidak perlu melakukan panggilan data dua kali.

- **Distribution of the age of gold medalists**

Hal yang pertama dilakukan adalah membuat dataframe baru hanya dengan memasukkan atlet peraih data medali emas dengan menggunakan tabel data yang telah digabungkan sebelumnya.

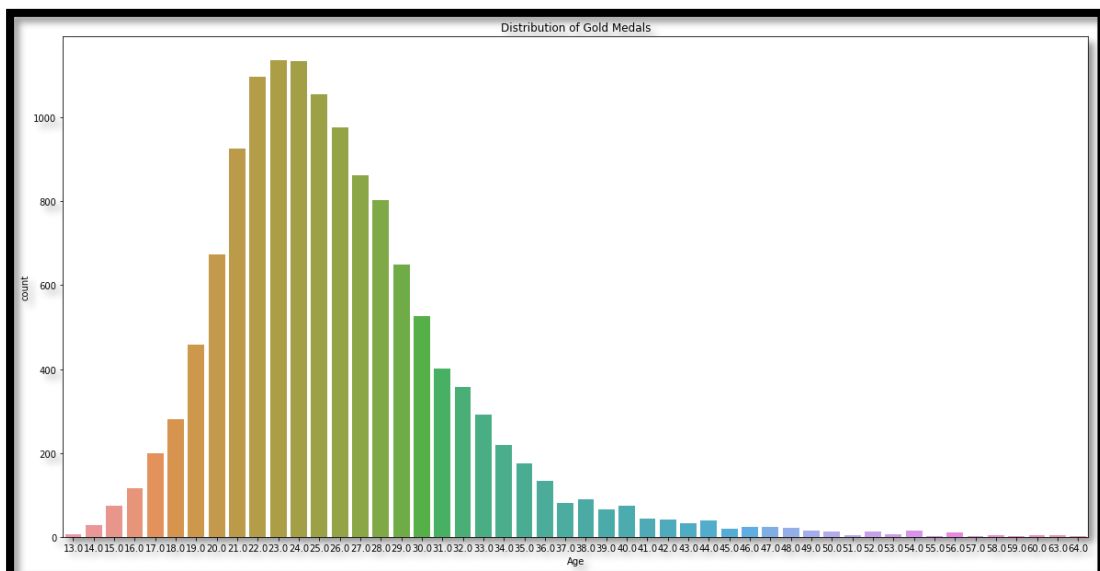
3	4	Edgar Lindenau Aabye	M	34.0	NaN	NaN	Denmark/Sweden	DEN	1900 Summer	1900	Summer	Paris	Tug-Of-War	Tug-Of-War Men's Tug-Of-War	Gold	Denmark	NaN
42	17	Paavo Johannes Aaltonen	M	28.0	175.0	64.0	Finland	FIN	1948 Summer	1948	Summer	London	Gymnastics	Gymnastics Men's Team All-Around	Gold	Finland	NaN
44	17	Paavo Johannes Aaltonen	M	28.0	175.0	64.0	Finland	FIN	1948 Summer	1948	Summer	London	Gymnastics	Gymnastics Men's Horse Vault	Gold	Finland	NaN
48	17	Paavo Johannes Aaltonen	M	28.0	175.0	64.0	Finland	FIN	1948 Summer	1948	Summer	London	Gymnastics	Gymnastics Men's Pommel Horse	Gold	Finland	NaN
60	20	Kjetil Andr Amrodt	M	20.0	178.0	85.0	Norway	NOR	1992 Winter	1992	Winter	Albertville	Alpine Skiing	Alpine Skiing Men's Super G	Gold	Norway	NaN

Tabel barunya akan tampil seperti gambar diatas dimana pada kolom medali hanya akan tampil atlet dengan peraih medali emas saja. Kemudian dilakukan pengecekan kolom usia apakah ada data yang hilang atau missing, karena nantinya dalam hasil plot akan

ditampilkan distribusi data medali dan usia untuk itu data usia tidak boleh ada yang missing.

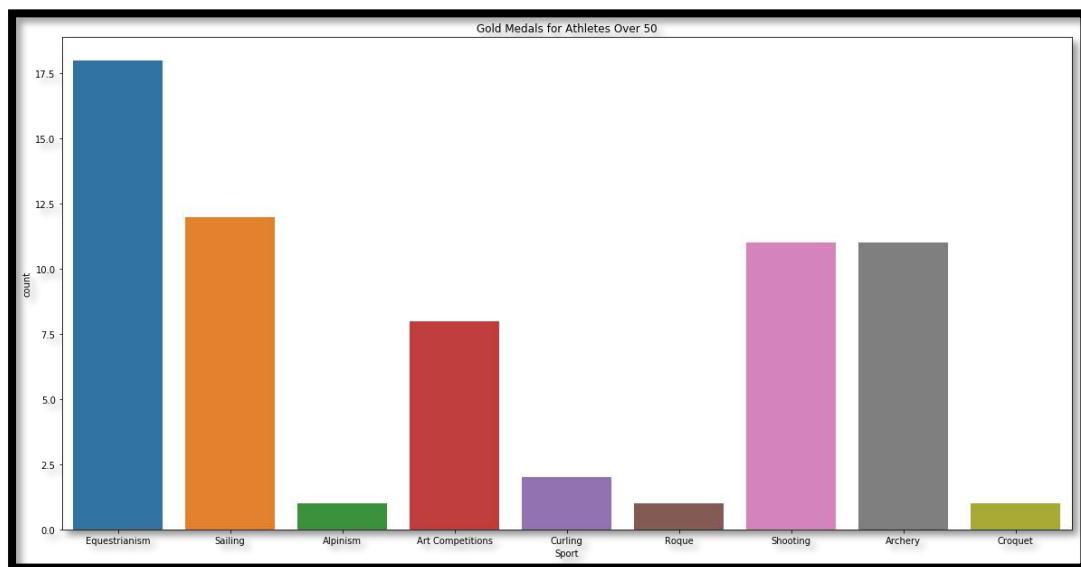
ID	False
Name	False
Sex	False
Age	True
Height	True
Weight	True
Team	False
NOC	False
Games	False
Year	False
Season	False
City	False
Sport	False
Event	False
Medal	False
region	True
notes	True
dtype:	bool

Setelah dilakukan pengecekan kelengkapan data hasilnya untuk kolom usia tidak ada data yang missing atau hilang. Setelah mengetahui jika data usia lengkap selanjutnya ada mengesplot data perolehan medali emas dengan usia atlet.



Visualisasi antara usia atlet dan perolehan medali emas seperti pada gambar diatas dimana terlihat distribusinya berupa eksponensial. Rentang usia pada tabel antara 13 tahun sampai dengan 64 tahun. Distribusi medali emas paling banyak berada pada rentang usia 23 dan 24 tahun. Distribusi medali emas paling banyak dominan berada

pada usia <30 tahun. Sedangkan untuk usia >30 tahun perolehan medali emasnya tidak begitu signifikan. Terlebih di usia >50 tahun sudah sangat minim perolehan medali emas. Hal ini sangat wajar dimana yang namanya olahraga sangat mengandalkan kekuatan fisik sedangkan untuk usia >40 tahun tentunya kekuatan tubuh sudah tidak begitu prima. Atlet yang memperoleh medali emas di usia >50 tahun kebanyakan mengikuti cabang olahraga yang tidak begitu mengandalkan otot namun banyak mengandalkan pikiran. Jumlah atlet yang mendapatkan medali emas di usia lebih dari 50 tahun sebanyak 65 orang. Luar biasa!



Kemudian dilakukan perintah untuk memvisualisasikan distribusi data atlet yang memperoleh medali emas di usia lebih dari 50 tahun ke dalam grafik bar dan diperoleh hasil seperti pada gambar diatas. Cabang olahraga dengan perolehan medali emas terbanyak dari atlet dengan usia lebih dari 50 tahun adalah equestrianism atau berkuda. Olahraga lainnya yang disumbangkan medali emas adalah sailing atau berlayar, alpinism atau panjat tebing gunung, seni, curling sport atau olahraga meluncurkan batu sejauh mungkin, roque sport olahraga sejenis golf, tembak, panah, dan croquet.

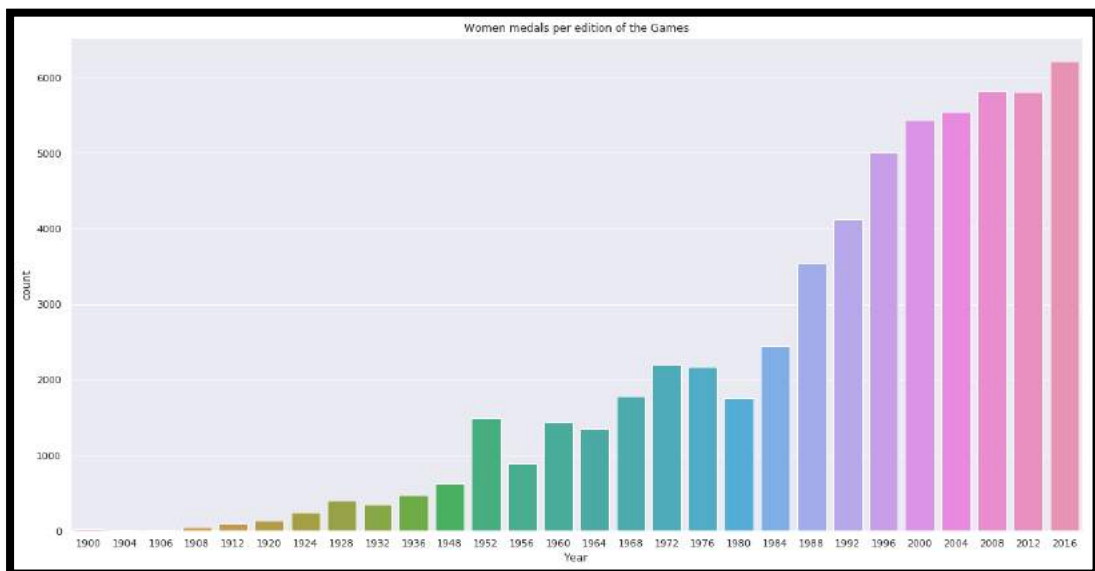
- **Women in Olympics**



Pertama dilakukan pengaturan dataset baru yang hanya berisikan atlet wanita saja dan olimpiade musim panas. Lalu diambil sampel 10 data teratas untuk melihat kondisi datanya.

ID	Name	Sex	Age	Height	Weight	Team	NOC	Games	Year	Season	City	Sport	Event	Medal	region
26	Cornelia "Cor" Aalten (- Strannood)	F	18.0	168.0	NaN	Netherlands	NED	1932 Summer	1932	Summer	Los Angeles	Athletics	Athletics Women's 100 metres	NaN	Netherlands
27	Cornelia "Cor" Aalten (- Strannood)	F	18.0	168.0	NaN	Netherlands	NED	1932 Summer	1932	Summer	Los Angeles	Athletics	Athletics Women's 4 x 100 metres Relay	NaN	Netherlands
32	Minna Maarit Aalto	F	30.0	159.0	55.5	Finland	FIN	1996 Summer	1996	Summer	Atlanta	Sailing	Sailing Women's Windsurfer	NaN	Finland
33	Minna Maarit Aalto	F	34.0	159.0	55.5	Finland	FIN	2000 Summer	2000	Summer	Sydney	Sailing	Sailing Women's Windsurfer	NaN	Finland
79	Ragnhild Margrethe Aamodt	F	27.0	163.0	NaN	Norway	NOR	2008 Summer	2008	Summer	Beijing	Handball	Handball Women's Handball	Gold	Norway
80	Andreea Aanei	F	22.0	170.0	125.0	Romania	ROU	2016 Summer	2016	Summer	Rio de Janeiro	Weightlifting	Weightlifting Women's Super-Heavyweight	NaN	Romania
91	Willemlen Aardenburg	F	22.0	NaN	NaN	Netherlands	NED	1988 Summer	1988	Summer	Seoul	Hockey	Hockey Women's Hockey	Bronze	Netherlands
105	Ann Kristin Aarnes	F	23.0	182.0	64.0	Norway	NOR	1996 Summer	1996	Summer	Atlanta	Football	Football Women's Football	Bronze	Norway
135	Moonika Aava	F	24.0	168.0	65.0	Estonia	EST	2004 Summer	2004	Summer	Athina	Athletics	Athletics Women's Javelin Throw	NaN	Estonia
136	Moonika Aava	F	28.0	168.0	65.0	Estonia	EST	2008 Summer	2008	Summer	Beijing	Athletics	Athletics Women's Javelin Throw	NaN	Estonia

Tampilannya seperti pada gambar diatas dimana yang tampil semua adalah atlet wanita. Selanjutnya dilakukan plot antara tahun penyelenggaraan olimpiade dengan jumlah medali yang diperoleh oleh atlet wanita.



Trend perolehan medali yang didapatkan oleh atlet wanita mengalami peningkatan dari waktu ke waktu. Perolehan medali paling banyak terjadi pada gelaran olimpiade terakhir yaitu tahun 2016 dimana atlet wanita mendapatkan lebih dari 6000 medali. Hal ini begitu signifikan jika dibandingkan dengan kondisi dimana awal-awal partisipan atlet wanita diperbolehkan untuk ikut gelaran olimpiade, periode tahun 1900 sampai 1906 bahkan bisa dikatakan hampir tidak ada medali yang didapatkan oleh para atlet

wanita namun sekarang jumlahnya bisa mencapai 6000. Hal ini membuat wanita tidak bisa dikesampingkan lagi dalam hal olahraga, bukan tidak mungkin kedepannya akan dipertandingkan antara pria dan wanita tidak lagi dipisahkan seperti sekarang.

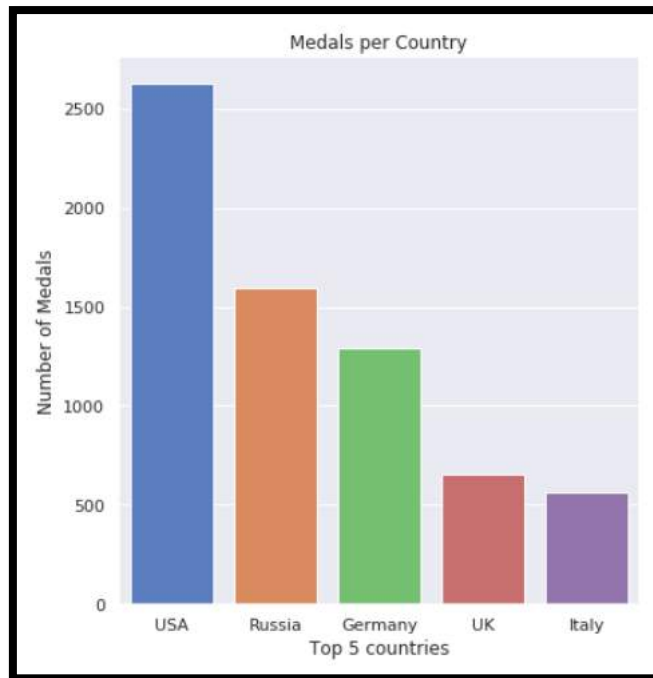
	ID	Name	Sex	Age	Height	Weight	Team	NOC	Games	Year	Season	City	Sport	Event	Medal	region	notes
283	150	Margaret Ives Abbott (-Dunne)	F	23.0	NaN	NaN	United States	USA	1900 Summer	1900	Summer	Paris	Golf	Golf Women's Individual	Gold	USA	NaN
284	151	Mary Perkins Ives Abbott (Perkins-)	F	42.0	NaN	NaN	United States	USA	1900 Summer	1900	Summer	Paris	Golf	Golf Women's Individual	NaN	USA	NaN
30535	15740	A. Brun	F	NaN	NaN	NaN	France	FRA	1900 Summer	1900	Summer	Paris	Golf	Golf Women's Individual	NaN	France	NaN
44448	22925	Charlotte Reinagle Cooper (-Sterry)	F	29.0	NaN	NaN	Great Britain	GBR	1900 Summer	1900	Summer	Paris	Tennis	Tennis Women's Singles	Gold	UK	NaN
44449	22925	Charlotte Reinagle Cooper (-Sterry)	F	29.0	NaN	NaN	Great Britain	GBR	1900 Summer	1900	Summer	Paris	Tennis	Tennis Mixed Doubles	Gold	UK	NaN
51667	26559	Hilne de Pourtais (Barbey-)	F	32.0	NaN	NaN	Lerina	SUI	1900 Summer	1900	Summer	Paris	Sailing	Sailing Mixed Open	NaN	Switzerland	NaN
51668	26559	Hilne de Pourtais (Barbey-)	F	32.0	NaN	NaN	Lerina	SUI	1900 Summer	1900	Summer	Paris	Sailing	Sailing Mixed 1-2 Ton	Gold	Switzerland	NaN
51669	26559	Hilne de Pourtais (Barbey-)	F	32.0	NaN	NaN	Lerina	SUI	1900 Summer	1900	Summer	Paris	Sailing	Sailing Mixed 1-2 Ton	Silver	Switzerland	NaN
54280	27851	Mme. Desprs	F	NaN	NaN	NaN	France	FRA	1900 Summer	1900	Summer	Paris	Croquet	Croquet Mixed Singles, One Ball	NaN	France	NaN
54281	27851	Mme. Desprs	F	NaN	NaN	NaN	France	FRA	1900 Summer	1900	Summer	Paris	Croquet	Croquet Mixed Singles, Two Balls	NaN	France	NaN

Untuk melakukan kroscek data pada gambar sebelumnya dilakukan validasi sampel terhadap dataset atlet perempuan pada olimpiade 1900 musim panas untuk 10 data paling awal, untuk memastikan apakah betul pada tahun tersebut ada atlet perempuan yang mendapatkan medali. Ternyata hasilnya adalah ada, dimana seperti yang ada pada gambar diatas ada 4 orang atlet wanita berhasil memperoleh medali yaitu pada cabang olahraga Golf, Tennis, dan Sailing. Selanjutnya juga dilakukan perhitungan total medali yang didapatkan pada gelaran olimpiade tahun 1900 totalnya ada 33 medali yang didapatkan oleh atlet wanita, hal ini juga telah dikroscek melalui wikipedia dan hasilnya sama.

- **Medals per country**

	index	Medal
0	USA	2627
1	Russia	1599
2	Germany	1293
3	UK	657
4	Italy	567

Pada poin ini dilakukan perintah untuk menampilkan 5 negara dengan jumlah perolehan medali terbanyak. Didapatkan seperti pada gambar diatas dimana USA atau Amerika memimpin dengan total 2627 medali, disusul Russia dengan 1599 medali, Germany 1293 medali, UK atau Inggris 657 medali, dan Italy dengan 567 medali.



Tampilan tabel 5 negara dengan jumlah perolehan medali terbanyak juga dapat dibuatkan ke dalam bentuk grafik bar seperti gambar diatas. Hal ini akan lebih memudahkan ketika melakukan presentasi agar lebih menarik, namun disini tidak begitu jelas nilai pastinya berapa.

- **Disciplines with the greatest number of Gold Medals**

Selanjutnya adalah membuat dataframe untuk memfilter medali emas hanya untuk negara USA. Selanjutnya tampil deretan cabang olahraga penyumbang medali emas terbanyak untuk negara USA, data yang diambil adalah 20 besar.

	index	Medal
0	Basketball Men's Basketball	186
1	Swimming Men's 4 × 200 metres Freestyle Relay	111
2	Swimming Men's 4 × 100 metres Medley Relay	108
3	Rowing Men's Coxed Eights	107
4	Basketball Women's Basketball	95
5	Athletics Men's 4 × 400 metres Relay	81
6	Swimming Women's 4 × 100 metres Medley Relay	79
7	Swimming Women's 4 × 100 metres Freestyle Relay	78
8	Football Women's Football	66
9	Athletics Men's 4 × 100 metres Relay	63
10	Swimming Men's 4 × 100 metres Freestyle Relay	58
11	Athletics Women's 4 × 100 metres Relay	50
12	Softball Women's Softball	45
13	Athletics Women's 4 × 400 metres Relay	38
14	Volleyball Men's Volleyball	36
15	Ice Hockey Men's Ice Hockey	36
16	Rugby Men's Rugby	36
17	Rowing Women's Coxed Eights	36
18	Swimming Women's 4 × 200 metres Freestyle Relay	33
19	Water Polo Women's Water Polo	25

Berdasarkan gambar diatas basket menjadi penyumbang terbanyak medali emas untuk USA selama gelaran olimpiade. 186 medali telah disumbangkan oleh cabang olahraga basket sepanjang keikutsertaannya. Selanjutnya disusul olahraga renang pria 4x200 m gaya bebas 111 medali dan renang pria 4x100 m gaya kombinasi dengan 108 medali di posisi kedua dan ketiga.

ID	Name	Sex	Age	Height	Weight	Team	NOC	Games	Year	Season	City	Sport	Event	Medal	Region	Notes
109529	Ernest Lee Johnson	M	25.0	188.0	70.0	United States	USA	1936	Summer	Berlin	Basketball	Basketball Men's Basketball	Gold	USA	NaN	
143588	Ernie John Lukan	M	26.0	208.0	113.0	United States	USA	1936	Summer	Berlin	Basketball	Basketball Men's Basketball	Gold	USA	NaN	
161770	Arthur Owen "Art" Miller	M	28.0	183.0	73.0	United States	USA	1936	Summer	Berlin	Basketball	Basketball Men's Basketball	Gold	USA	NaN	
78052	John "Hobby" "Foy" Gibson	M	28.0	185.0	79.0	United States	USA	1936	Summer	Berlin	Basketball	Basketball Men's Basketball	Gold	USA	NaN	
188347	Donald Arthur "Doc" Piper	M	25.0	180.0	75.0	United States	USA	1936	Summer	Berlin	Basketball	Basketball Men's Basketball	Gold	USA	NaN	
71407	Joseph "Cotton" "Doc" Farnsworth	M	25.0	203.0	84.0	United States	USA	1936	Summer	Berlin	Basketball	Basketball Men's Basketball	Gold	USA	NaN	
195790	Jack Williamson Redman	M	22.0	183.0	79.0	United States	USA	1936	Summer	Berlin	Basketball	Basketball Men's Basketball	Gold	USA	NaN	
213368	Willard Theodore Schmitt	M	26.0	205.0	86.0	United States	USA	1936	Summer	Berlin	Basketball	Basketball Men's Basketball	Gold	USA	NaN	
219204	Carl Leslie Stry	M	27.0	183.0	77.0	United States	USA	1936	Summer	Berlin	Basketball	Basketball Men's Basketball	Gold	USA	NaN	
22390	Ralph English Bishop	M	26.0	193.0	80.0	United States	USA	1936	Summer	Berlin	Basketball	Basketball Men's Basketball	Gold	USA	NaN	
23447	Duane Alexander Swenson	M	22.0	188.0	70.0	United States	USA	1936	Summer	Berlin	Basketball	Basketball Men's Basketball	Gold	USA	NaN	
127770	Carl Stanley Knoules	M	26.0	188.0	75.0	United States	USA	1936	Summer	Berlin	Basketball	Basketball Men's Basketball	Gold	USA	NaN	
258116	William John "Bill" Winstedley	M	27.0	188.0	70.0	United States	USA	1936	Summer	Berlin	Basketball	Basketball Men's Basketball	Gold	USA	NaN	
13663	Samuel J. "Sam" Beller Jr	M	26.0	178.0	66.0	United States	USA	1936	Summer	Berlin	Basketball	Basketball Men's Basketball	Gold	USA	NaN	
201900	Robert Lloyd Jackson "Jack" Robinson	M	21.0	183.0	82.0	United States	USA	1948	Summer	London	Basketball	Basketball Men's Basketball	Gold	USA	NaN	

Selanjutnya dilakukan penjabaran penerima medali emas dari anggota tim basket USA yang pada gambar sebelumnya berada di puncak penyumbang medali emas untuk negara USA sepanjang gelaran olimpiade. Namun ternyata yang dilakukan adalah menghitung medali emas dari setiap anggota tim sedangkan olahraga basket adalah olahraga tim yang dimainkan oleh 5 orang maka dari itu mengapa jumlahnya bisa sangat banyak karena menghitung jumlah medali emas dari setiap anggota tim.

Year	ID	Name	Sex	Age	Height	Weight	Team	NOC	Games	Season	City	Sport	Event	Medal	Region	Notes
1936	55325	Ernest Lee Johnson	M	25.0	188.0	70.0	United States	USA	1936	Summer	Berlin	Basketball	Basketball Men's Basketball	Gold	USA	NaN
1948	101443	Robert Lloyd Jackson "Jack" Robinson	M	21.0	183.0	82.0	United States	USA	1948	Summer	London	Basketball	Basketball Men's Basketball	Gold	USA	NaN
1952	58738	Robert Earl "Bob" Kennedy	M	21.0	188.0	84.0	United States	USA	1952	Summer	Helsinki	Basketball	Basketball Men's Basketball	Gold	USA	NaN
1956	128271	James Patrick "Jim" Walsh	M	26.0	193.0	88.0	United States	USA	1956	Summer	Melbourne	Basketball	Basketball Men's Basketball	Gold	USA	NaN
1960	133271	Robert Lewis "Bob" Soccer	M	23.0	203.0	89.0	United States	USA	1960	Summer	Rome	Basketball	Basketball Men's Basketball	Gold	USA	NaN
1964	130843	George "Bill" Wilson	M	22.0	203.0	86.0	United States	USA	1964	Summer	Tokyo	Basketball	Basketball Men's Basketball	Gold	USA	NaN
1968	8104	Michael Thomas "Mike" Bennett	M	25.0	188.0	73.0	United States	USA	1968	Summer	Mexico City	Basketball	Basketball Men's Basketball	Gold	USA	NaN
1976	56603	Thomas Joseph "Tom" LaGarde	M	21.0	203.0	87.0	United States	USA	1976	Summer	Montreal	Basketball	Basketball Men's Basketball	Gold	USA	NaN
1984	120501	Wayman Lawrence Tisdale	M	26.0	203.0	118.0	United States	USA	1984	Summer	Los Angeles	Basketball	Basketball Men's Basketball	Gold	USA	NaN
1992	33553	Frederick Anthony Cwing	M	29.0	213.0	109.0	United States	USA	1992	Summer	Barcelona	Basketball	Basketball Men's Basketball	Gold	USA	NaN
1996	116325	John Houston Stockton	M	34.0	185.0	79.0	United States	USA	1996	Summer	Atlanta	Basketball	Basketball Men's Basketball	Gold	USA	NaN
2000	2863	Walter Ray Allen	M	25.0	192.0	88.0	United States	USA	2000	Summer	Sydney	Basketball	Basketball Men's Basketball	Gold	USA	NaN
2008	33739	Christopher Maxwell "Chris" Bosh	M	24.0	208.0	104.0	United States	USA	2008	Summer	Beijing	Basketball	Basketball Men's Basketball	Gold	USA	NaN
2012	53494	LeBron Raymone James	M	27.0	203.0	114.0	United States	USA	2012	Summer	London	Basketball	Basketball Men's Basketball	Gold	USA	NaN
2016	55674	Hyland DeAndre Jordan, Jr.	M	28.0	211.0	120.0	United States	USA	2016	Summer	Rio de Janeiro	Basketball	Basketball Men's Basketball	Gold	USA	NaN

Setelah diketahui penyebab kesalahannya selanjutnya dilakukan pengelompokkan setiap anggota tim berdasarkan tahun mereka juara. Hasilnya seperti pada gambar diatas yang merupakan tabel yang berisikan tim basket USA yang mendapatkan medali emas. Ketika sudah dikelompokkan berdasarkan tahun juaranya maka akan terlihat jumlah medali emas diperoleh negara USA selama gelaran olimpiade berapa kali. Berdasarkan pengelompokkan yang telah dilakukan jumlah jika dihitung berdasarkan tim maka negara USA telah memenangkan medali emas basket sebanyak 15 kali, hal ini sudah dilakukan kroscek dengan wikipedia dan hasilnya cukup baik.

- **What is the median height/weight of an Olympic medalist?**

ID	Name	Sex	Age	Height	Weight	Team	NOC	Games	Year	Season	City	Sport	Event	Medal	region	note
3	Edgar Lindorou Aabye	M	34.0	NaN	NaN	Denmark/Sweden	DEN	1900 Summer	1900	Summer	Paris	Tug-Of-War	Tug-Of-War Men's Tug-Of-War	Gold	Denmark	NaN
42	Paavo Johannes Aaltonen	M	28.0	175.0	64.0	Finland	FIN	1948 Summer	1948	Summer	London	Gymnastics	Gymnastics Men's Team All-Around	Gold	Finland	NaN
44	Paavo Johannes Aaltonen	M	28.0	175.0	64.0	Finland	FIN	1948 Summer	1948	Summer	London	Gymnastics	Gymnastics Men's Horse Vault	Gold	Finland	NaN
48	Paavo Johannes Aaltonen	M	28.0	175.0	64.0	Finland	FIN	1948 Summer	1948	Summer	London	Gymnastics	Gymnastics Men's Pommel Horse	Gold	Finland	NaN
60	Kjetil Andr Aamodt	M	20.0	176.0	85.0	Norway	NOR	1992 Winter	1992	Winter	Albertville	Alpine Skiing	Alpine Skiing Men's Super G	Gold	Norway	NaN

Hal pertama yang dilakukan untuk menentukan nilai median tinggi atau berat badan dari peserta olympic yaitu dengan kembali menampilkan dataset disini digunakan peraih medali emas. Terlihat pada gambar diatas jika ada nilai missing atau NaN pada kolom tinggi dan berat badan. Ada 2 langkah yang bisa dijalani yaitu abaikan peraih medali yang tidak memiliki berat dan tinggi badan atau dengan cara menggunakan rata-rata berat dan tinggi badan berdasarkan data partisipan lainnya. Tentunya akan sangat tidak baik jika menggunakan langkah yang kedua, dimana olimpiade berisikan partisipan dari cabang olahraga yang berbeda dengan kemampuan yang berbeda pula untuk itu akan digunakan langkah pertama dengan mengabaikan peraih medali yang tidak memiliki data tinggi dan berat badan untuk proses selanjutnya.

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 13224 entries, 3 to 271076
Data columns (total 17 columns):
ID          13224 non-null int64
Name       13224 non-null object
Sex        13224 non-null object
Age        13224 non-null float64
Height     10532 non-null float64
Weight     10248 non-null float64
Team       13224 non-null object
NOC        13224 non-null object
Games      13224 non-null object
Year       13224 non-null int64
Season     13224 non-null object
City       13224 non-null object
Sport      13224 non-null object
Event      13224 non-null object
Medal      13224 non-null object
region     13223 non-null object
notes      171 non-null object
dtypes: float64(3), int64(2), object(12)
memory usage: 2.4+ MB
```

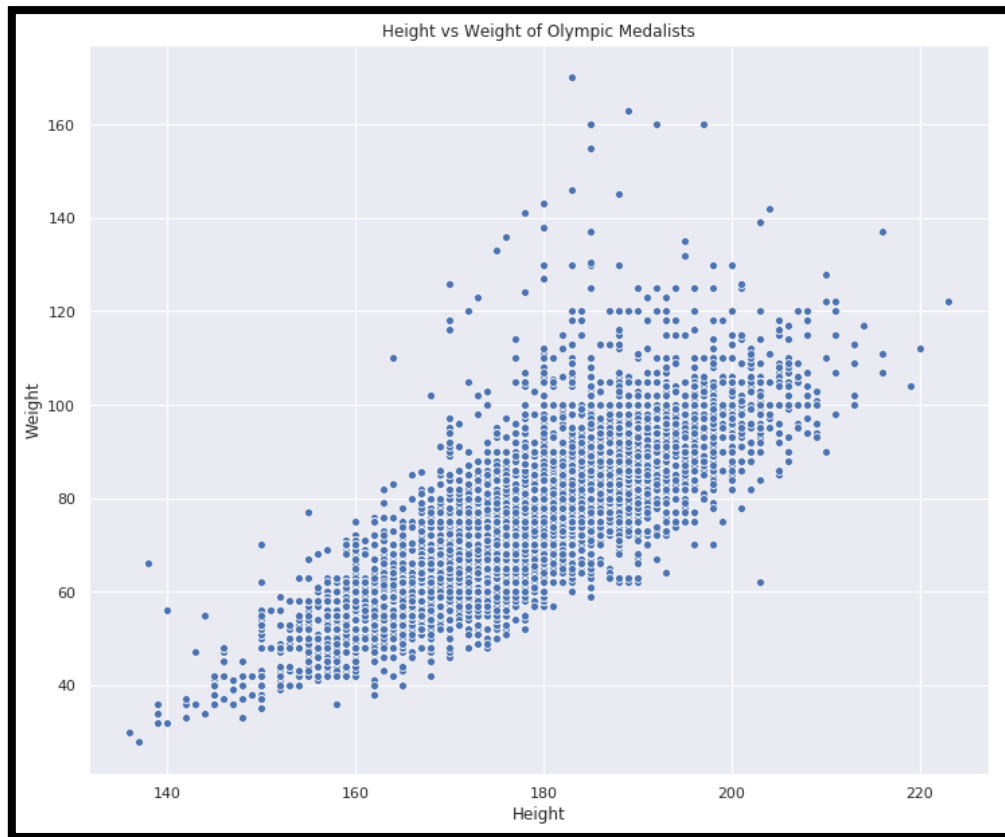
Sebelum melangkah lebih jauh dilakukan pengecekan jumlah data dari peraih medali emas. Disini terlihat jika data tinggi dan berat badan hanya berjumlah 10532 dan 10248 data sedangkan jumlah data lainnya lebih dari 13000 data. Untuk itu disini akan digunakan populasi data yang seluruh datanya lengkap terutama untuk tinggi dan berat badan.

ID	Name	Sex	Age	Height	Weight	Team	NOC	Games	Year	Season	City	Sport	Event	Medal	region	notes
42	Paavo Johannes Aaltonen	M	28.0	175.0	84.0	Finland	FIN	1948 Summer	1948	Summer	London	Gymnastics	Gymnastics Men's Team All-Around	Gold	Finland	NaN
44	Paavo Johannes Aaltonen	M	28.0	175.0	84.0	Finland	FIN	1948 Summer	1948	Summer	London	Gymnastics	Gymnastics Men's Horse Vault	Gold	Finland	NaN
48	Paavo Johannes Aaltonen	M	28.0	175.0	84.0	Finland	FIN	1948 Summer	1948	Summer	London	Gymnastics	Gymnastics Men's Pommel Horse	Gold	Finland	NaN
60	Kjetil Andr Aamodt	M	20.0	176.0	85.0	Norway	NOR	1992 Winter	1992	Winter	Albertville	Alpine Skiing	Alpine Skiing Men's Super G	Gold	Norway	NaN
73	Kjetil Andr Aamodt	M	30.0	176.0	85.0	Norway	NOR	2002 Winter	2002	Winter	Salt Lake City	Alpine Skiing	Alpine Skiing Men's Super G	Gold	Norway	NaN

Gambar diatas merupakan dataset baru yang menggunakan dataframe peraih medali emas yang memiliki data tinggi dan berat badan. Gambar dibawah ini terlihat jumlah data sekarang berada di angka 10167 dan hampir seluruh kolom data bernilai sama kecuali untuk kolom region dan notes dimana kedua kolom itu tidak bergitu signifikan pengaruhnya terhadap visualisasi data.

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 10167 entries, 42 to 271076
Data columns (total 17 columns):
ID          10167 non-null int64
Name        10167 non-null object
Sex         10167 non-null object
Age         10167 non-null float64
Height      10167 non-null float64
Weight      10167 non-null float64
Team        10167 non-null object
NOC         10167 non-null object
Games       10167 non-null object
Year        10167 non-null int64
Season      10167 non-null object
City        10167 non-null object
Sport       10167 non-null object
Event       10167 non-null object
Medal       10167 non-null object
region      10166 non-null object
notes       143 non-null object
dtypes: float64(3), int64(2), object(12)
memory usage: 1.4+ MB
```

Setelah kita mendapatkan dataframe terbaik seperti gambar diatas selanjutnya data ini sudah bisa dilakukan proses lebih lanjut yaitu pembuatan scatterplot antara tinggi badan vs berat badan untuk setiap peraih medali. Hasilnya seperti gambar dibawah ini dimana sebaran data antara tinggi dan berat badan atlet peraih medali cukup erat hubungannya. Dapat terlihat sebaran data memiliki hubungan yang positif dimana jika ditarik garis lurus akan membentuk garis lurus diagonal miring ke kiri.



Gambar diatas merupakan hasil plotting scatterplot dari data peraih medali yang telah dilakukan pengelompokkan agar tidak ada atlet yang tidak memiliki data tinggi dan berat badan. Jika dilihat dari polanya maka hubungan kedua data ini sangat baik atau positif, hubungan antara tinggi badan dan berat badan erat hubungannya. Mudahnya seperti ini jika data berat badan meningkat maka tinggi badan juga meningkat. Memang ada beberapa data outlier atau data yang berada diluar kelompok data, hal ini memang selalu terjadi karena pasti akan ada anomali data, namun jika sudah terlihat jelas seperti gambar diatas, nilai outliernya bisa dibuang untuk proses merapikan data karena data outlier juga tidak dapat untuk dibunyikan sebagai analisis.

	ID	Name	Sex	Age	Height	Weight	Team	NOC	Games	Year	Season	City	Sport	Event	Medal	region	note	
	39181	20144	Andrey Ivanovich Chemerkhin	M	24.0	183.0	170.0	Russia	RUS	1996 Summer	1996	Summer	Atlanta	Weightlifting	Men's Super-Heavyweight	Gold	Russia	NaN
	268659	134407	Leonid Ivanovich Zhabotynskiy	M	26.0	189.0	163.0	Soviet Union	URS	1964 Summer	1964	Summer	Tokyo	Weightlifting	Men's Heavyweight	Gold	Russia	NaN
	268660	134407	Leonid Ivanovich Zhabotynskiy	M	30.0	189.0	163.0	Soviet Union	URS	1968 Summer	1968	Summer	Mexico City	Weightlifting	Men's Heavyweight	Gold	Russia	NaN

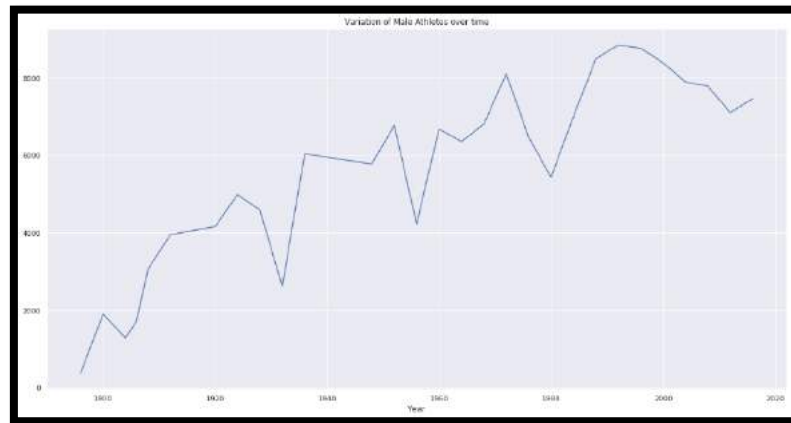
Terakhir dilakukan percobaan untuk melihat apakah ada atlet peraih medali yang beratnya lebih dari 160 kilogram di dalam dataset. Pada gambar diatas terlihat ternyata ada 3 atlet yang memiliki berat badan diatas 160 kilogram dan semuanya berasal dari cabang olahraga angkat beban, sangat tepat sekali. Akan sangat aneh jika dengan berat segitu ternyata bukan dari olahraga angkat beban atau judo mungkin, misalnya berasal dari atletik tentunya hal ini akan sangat aneh. Berat badan ketiga atlet tersebut berturut turut sebesar 170, 163, dan 163 kilogram.

- Evolution of the Olympics over time

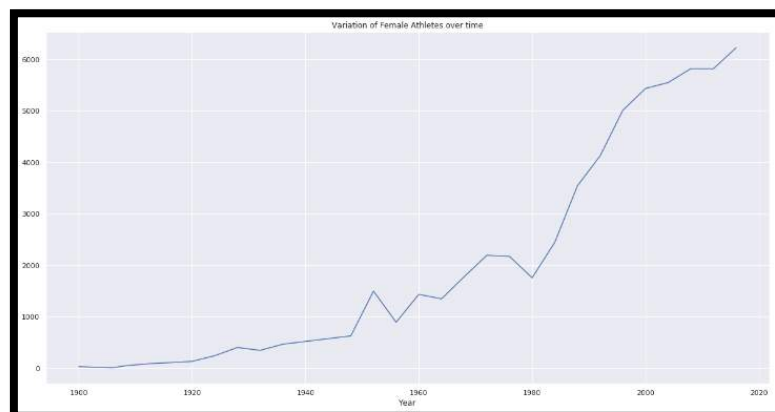
- Variation of male/female athletes over time (Summer Games)

ID	Name	Sex	Age	Height	Weight	Team	NOC	Games	Year	Season	City	Sport	Event	Medal	region	note	
0	1	A Dijiang	M	24.0	180.0	80.0	China	CHN	1992 Summer	1992	Summer	Barcelona	Basketball	Men's Basketball	NaN	China	NaN
1	2	A Lamusi	M	23.0	170.0	60.0	China	CHN	2012 Summer	2012	Summer	London	Judo	Judo Men's Extra-Lightweight	NaN	China	NaN
2	3	Gunnar Nielsen Aaby	M	24.0	NaN	NaN	Denmark	DEN	1920 Summer	1920	Summer	Antwerpen	Football	Football Men's Football	NaN	Denmark	NaN
3	4	Egner Lindqvist Asbyll	M	34.0	NaN	NaN	Denmark/Sweden	DEN	1900 Summer	1900	Summer	Paris	Tag-Of-War	Tag-Of-War Men's Tag-Of-War	Gold	Denmark	NaN
20	10	Einar Ferdinand "Einar" Aalto	M	26.0	NaN	NaN	Finland	FIN	1952 Summer	1952	Summer	Helsinki	Swimming	Swimming Men's 400 meters Freestyle	NaN	Finland	NaN

Dengan menggunakan dataframe diatas selanjutnya akan dilakukan analisis dengan visualisasi terkait jumlah atlet pria dan wanita terhadap penyelenggaraan waktu olimpiade musim panas.

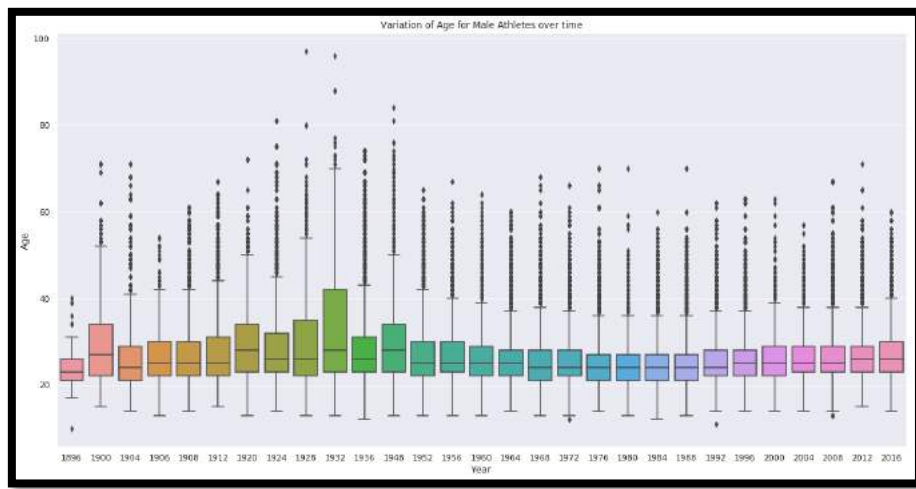


Pertama visualisasi atlet pria, secara garis besar trennya mengalami peningkatan jumlah dari tahun ke tahun. Jumlahnya bahkan sudah mencapai lebih dari 8000 peserta pada gelaran tahun 1990an. Berdasarkan gambar diatas terlihat penurunan jumlah partisipasn pria yang cukup signifikan terjadi 3 kali yaitu pada gelaran tahun 1930an, 1950an, dan 1980 sebelumnya jumlahnya mengalami peningkatan.



Kedua adalah visualisasi atlet wanita, pada awal gelaran olimpiade musim panas sampai periode tahun 1950an jumlah peningkatannya selalu meningkat namun tidak begitu signifikan jumlahnya tidak sampai 1000 orang. Pada tahun 1980an baru jumlahnya mulai mengalami peningkatan yang signifikan, sampai dengan gelaran terakhir yang jumlahnya sudah menyentuh angka 6000an orang. Hal ini menandakan bahwa eksistensi atlet wanita sudah mulai hampir sama dengan atlet pria, dimana jumlahnya sekarang tidak begitu jauh selisihnya. Bukan tidak mungkin dalam waktu 10 tahun mendatang jumlahnya bahkan bisa lebih banyak wanita dalam gelaran olimpiade.

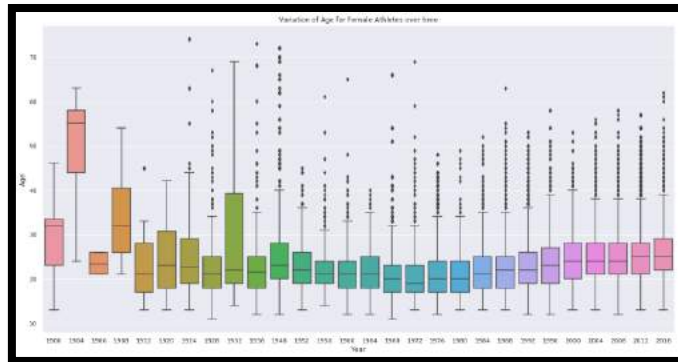
➤ **Variation of age along time**



Analisis selanjutnya adalah melakukan visualisasi terkait distribusi atlet pria dari tahun ke tahun gelaran olimpiade. Nilai median dari keseluruhan data usia atlet pria berada pada rentang usia 20-30 tahun. Namun ada beberapa hal yang cukup menarik terutama pada gelaran olimpiade 1924 sampai dengan 1948 dimana terlihat banyak atlet pria dengan usia lebih dari 80 tahun, luas biasa. Menjadi pertanyaan olahraga apa yang mereka ikuti? Apakah kesalahan data usia? Hal ini akan dilakukan pengecekan lebih lanjut pada pembahasan berikutnya untuk memastikan jika tidak ada kesalahan pada datasetnya.

ID	Name	Sex	Age	Height	Weight	Team	NOC	Games	Year	Season	City	Sport	Event	Medal	Region
3271	Swain Davidson Hansen	M	84.0	NaN	NaN	Great Britain	GBR	1928 Summer	1948	Summer	London	Art Competitions	Art Competitions Mixed Painting, Unknown Event	NaN	UK
10301	Thomas Comberthwait Giblin	M	80.0	NaN	NaN	United States	USA	1922 Summer	1922	Summer	Los Angeles	Art Competitions	Art Competitions Mixed Painting, Unknown Event	NaN	USA
10302	Thomas Comberthwait Giblin	M	80.0	NaN	NaN	United States	USA	1922 Summer	1922	Summer	Los Angeles	Art Competitions	Art Competitions Mixed Painting, Unknown Event	NaN	USA
10482	Thomas Comberthwait Giblin	M	80.0	NaN	NaN	United States	USA	1922 Summer	1922	Summer	Los Angeles	Art Competitions	Art Competitions Mixed Painting, Unknown Event	NaN	USA
18716	Winlow Hovner	M	90.0	NaN	NaN	United States	USA	1922 Summer	1922	Summer	Los Angeles	Art Competitions	Art Competitions Mixed Painting, Unknown Event	NaN	USA
16485	Robert Neil McKenzie	M	81.0	NaN	NaN	Canada	CAN	1948 Summer	1948	Summer	London	Art Competitions	Art Competitions Mixed Sculpturing, Unknown Event	NaN	Canada
238912	Louis Truett	M	81.0	NaN	NaN	France	FIN	1924 Summer	1924	Summer	Paris	Art Competitions	Art Competitions Mixed Sculpturing	NaN	France
20702	John Quincy Adams Ward	M	87.0	NaN	NaN	United States	USA	1928 Summer	1928	Summer	Amsterdam	Art Competitions	Art Competitions Mixed Sculpturing, Statues	NaN	USA

Akhirnya untuk memastikan apakah betul ada atlet pria dengan usia lebih dari 80 tahun yang mengikuti gelaran olimpiade pada periode tahun 1924 – 1948 dilakukan pengecekan data dengan memanggil data usia atlet yang lebih dari 80 tahun. Setelah melakukan analisis ternyata memang betul ada banyak atlet dengan usia lebih dari 80 tahun yang mengikuti gelaran olimpiade yaitu cabang art competition. Ini merupakan perlombaan menggambar atau melukis sehingga jika diikuti atlet dengan usia 80 tahun masih sangat layak karena bukan merupakan olahraga yang menggunakan otot. Perlombaan melukis ini merupakan komitmen dari para pendiri olimpiade untuk melombakan karya seni yang terinspirasi dari olahraga.

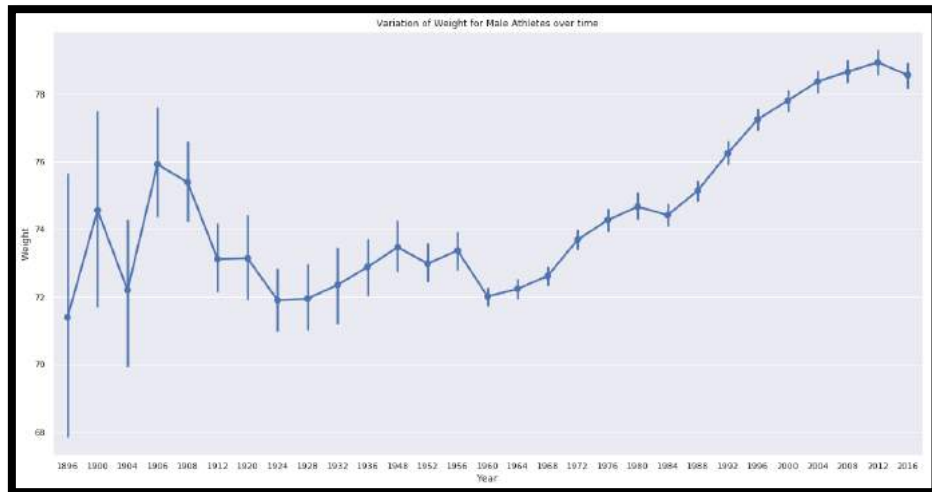


Analisis berikutnya adalah distribusi atlet wanita dari tahun ke tahun gelaran olimpiade. Untuk distribusi usia atlet wanita lebih bervariasi jika dibandingkan distribusi usia atlet pria terutama periode tahun 1900 – 1932, setelah periode tahun tersebut distribusi usia atlet wanita jauh lebih stabil dan memiliki nilai median 20-25 tahun. Pada gambar diatas yang palung mencolok perbedaannya adalah distribusi usia atlet wanita pada tahun 1904, untuk melihat datanya pada tahun tersebut akan dilakukan pembahasan lebih lanjut. Untuk usia atlet wanita tidak ada yang mengikuti gelaran olimpiade sampai dengan usia 80 tahun namun ada yang berusia 70 tahunan pada periode tahun 1924-1948, sama persis dengan kondisi pada atlet pria.

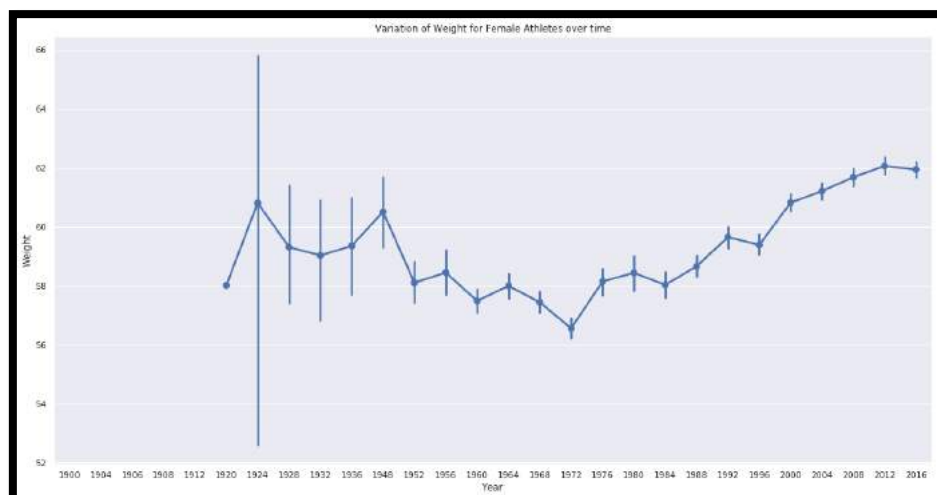
ID	Name	Sex	Year	Height	Weight	Event	IOC	Country	Year	Season	City	Event	Result	Notes
141303	Ernesto S. Frutos	F	1904	1665	5451	100m Backstroke	USA	USA	1904	Summer	St. Louis	Art Competition	70	
141304	Ernesto S. Frutos	F	1904	1665	5451	100m Backstroke	USA	USA	1904	Summer	St. Louis	Art Competition	70	
141305	Ernesto S. Frutos	F	1904	1665	5451	100m Backstroke	USA	USA	1904	Summer	St. Louis	Art Competition	70	
141306	Ernesto S. Frutos	F	1904	1665	5451	100m Backstroke	USA	USA	1904	Summer	St. Louis	Art Competition	70	
141307	Ernesto S. Frutos	F	1904	1665	5451	100m Backstroke	USA	USA	1904	Summer	St. Louis	Art Competition	70	
141308	Ernesto S. Frutos	F	1904	1665	5451	100m Backstroke	USA	USA	1904	Summer	St. Louis	Art Competition	70	
141309	Ernesto S. Frutos	F	1904	1665	5451	100m Backstroke	USA	USA	1904	Summer	St. Louis	Art Competition	70	
141310	Ernesto S. Frutos	F	1904	1665	5451	100m Backstroke	USA	USA	1904	Summer	St. Louis	Art Competition	70	
141311	Ernesto S. Frutos	F	1904	1665	5451	100m Backstroke	USA	USA	1904	Summer	St. Louis	Art Competition	70	
141312	Ernesto S. Frutos	F	1904	1665	5451	100m Backstroke	USA	USA	1904	Summer	St. Louis	Art Competition	70	
141313	Ernesto S. Frutos	F	1904	1665	5451	100m Backstroke	USA	USA	1904	Summer	St. Louis	Art Competition	70	
141314	Ernesto S. Frutos	F	1904	1665	5451	100m Backstroke	USA	USA	1904	Summer	St. Louis	Art Competition	70	
141315	Ernesto S. Frutos	F	1904	1665	5451	100m Backstroke	USA	USA	1904	Summer	St. Louis	Art Competition	70	
141316	Ernesto S. Frutos	F	1904	1665	5451	100m Backstroke	USA	USA	1904	Summer	St. Louis	Art Competition	70	
141317	Ernesto S. Frutos	F	1904	1665	5451	100m Backstroke	USA	USA	1904	Summer	St. Louis	Art Competition	70	
141318	Ernesto S. Frutos	F	1904	1665	5451	100m Backstroke	USA	USA	1904	Summer	St. Louis	Art Competition	70	
141319	Ernesto S. Frutos	F	1904	1665	5451	100m Backstroke	USA	USA	1904	Summer	St. Louis	Art Competition	70	
141320	Ernesto S. Frutos	F	1904	1665	5451	100m Backstroke	USA	USA	1904	Summer	St. Louis	Art Competition	70	
141321	Ernesto S. Frutos	F	1904	1665	5451	100m Backstroke	USA	USA	1904	Summer	St. Louis	Art Competition	70	
141322	Ernesto S. Frutos	F	1904	1665	5451	100m Backstroke	USA	USA	1904	Summer	St. Louis	Art Competition	70	
141323	Ernesto S. Frutos	F	1904	1665	5451	100m Backstroke	USA	USA	1904	Summer	St. Louis	Art Competition	70	
141324	Ernesto S. Frutos	F	1904	1665	5451	100m Backstroke	USA	USA	1904	Summer	St. Louis	Art Competition	70	
141325	Ernesto S. Frutos	F	1904	1665	5451	100m Backstroke	USA	USA	1904	Summer	St. Louis	Art Competition	70	
141326	Ernesto S. Frutos	F	1904	1665	5451	100m Backstroke	USA	USA	1904	Summer	St. Louis	Art Competition	70	
141327	Ernesto S. Frutos	F	1904	1665	5451	100m Backstroke	USA	USA	1904	Summer	St. Louis	Art Competition	70	
141328	Ernesto S. Frutos	F	1904	1665	5451	100m Backstroke	USA	USA	1904	Summer	St. Louis	Art Competition	70	
141329	Ernesto S. Frutos	F	1904	1665	5451	100m Backstroke	USA	USA	1904	Summer	St. Louis	Art Competition	70	
141330	Ernesto S. Frutos	F	1904	1665	5451	100m Backstroke	USA	USA	1904	Summer	St. Louis	Art Competition	70	

Untuk menjawab keanehan distribusi usia atlet wanita pada tahun 1904 dilakukan pengecekan data keseluruhan untuk atlet wanita pada tahun tersebut dan akhirnya terjawab mengapa rata-rata usianya berbeda dengan yang lainnya. Hampir seluruh atlet yang mengikuti olimpiade pada tahun tersebut berusia diatas 40 tahun dan hanya 2 atlet saja yang berusia 24 tahun. Maka dari itulah mengapa mediannya terlihat aneh sendiri pada tahun 1904. Seluruh atlet wanita pada tahun tersebut hanya mengikuti perlombaan memanah.

➤ **Variation of weight along time**



Analisis berikutnya adalah distribusi berat badan atlet pria dan wanita selama gelaran olimpiade. Pertama gambar diatas merupakan visualisasi dari distribusi berat badan atlet pria, di awal-awal gelaran olimpiade rentang berat badan atlet pria cukup besar gapnya namun perlahan semakin kesini jarak gapnya semakin kecil dan semakin stabil. Terutama mulai periode tahun 1960 sampai dengan sekarang gapnya kecil dan begitu stabil. Gap yang besar disini digambarkan dengan tiang-tiang pada gambar yang semakin panjang dari atas ke bawah, sebaliknya gap yang kecil digambarkan dengan pendeknya tiang-tiang pada gambar tersebut. Untuk distribusi gap atlet wanita akan dibahas pada selanjutnya.

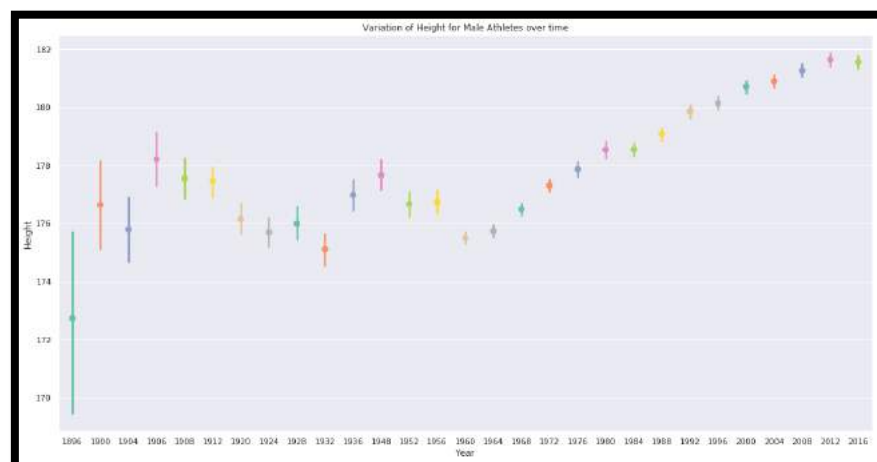


Selanjutnya adalah distribusi berat badan atlet wanita, namun disini ada yang menarik karena ternyata ada kekosongan data pada periode tahun dibawah 1924. Terlihat pada gambar diatas gap yang sangat besar terjadi pada tahun 1924. Gapnya mulai stabil mulai tahun 1960 sampai dengan sekarang. Semakin kesini gapnya semakin kecil, hal ini ditandai dengan tiang tiang vertikal yang semakin pendek.

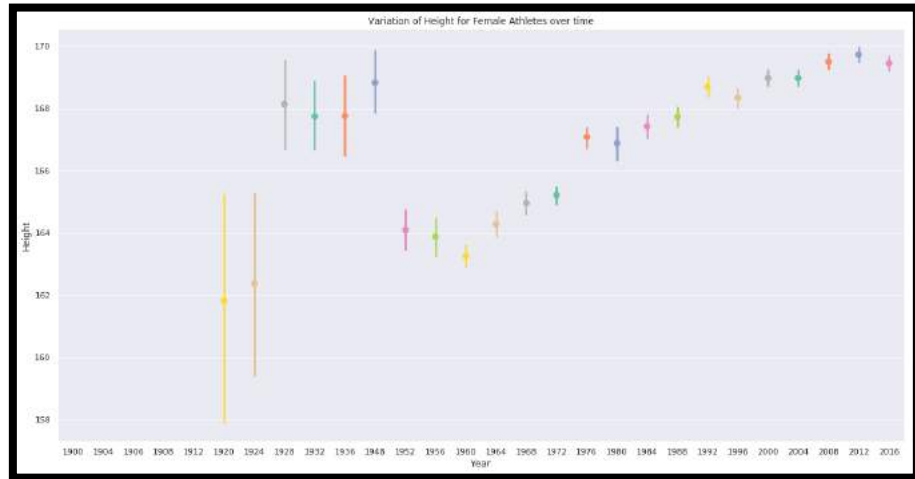
ID	Name	Sex	Age	Height	Weight	Team	IOC	Compe	Year	Season	City	Sport	Event	Medal	Region
183	Margaret Iona Adcock (Ireland)	F	23.0	NaN	NaN	United States	USA	1900 Summer	1908	Summer	Paris	Golf	Golf Women's Individual	Gold	USA
184	Mary Pickens Ross Adcock (Ireland)	F	43.0	NaN	NaN	United States	USA	1900 Summer	1908	Summer	Paris	Golf	Golf Women's Individual	NaN	USA
1738	Marijanna "Grove" Astor	F	18.0	NaN	NaN	Austria	AUT	1912 Summer	1912	Summer	Stockholm	Sporting	Sporting Women's 100 metres Freestyle	NaN	Austria
1739	Marijanna "Grove" Astor	F	18.0	NaN	NaN	Austria	AUT	1912 Summer	1912	Summer	Stockholm	Sporting	Sporting Women's 100 metres Freestyle Relay	Bronze	Austria
1758	Astrid Milla Ahrenberg (Sweden)	F	39.0	NaN	NaN	Sweden	SWE	1908 Summer	1908	Summer	London	Tennis	Tennis Women's Singles, Covered Courts	Bronze	Sweden
1768	Milla Erika Ahrenberg (Sweden)	F	16.0	NaN	NaN	Sweden	SWE	1912 Summer	1912	Summer	Stockholm	Diving	Diving Women's Platform	NaN	Sweden
1769	Milla Erika Ahrenberg (Sweden)	F	23.0	NaN	NaN	Sweden	SWE	1920 Summer	1920	Summer	Antwerpen	Diving	Diving Women's Platform	NaN	Sweden
1749	Florence Helen Adcock (Ireland)	F	30.0	NaN	NaN	Great Britain	GBR	1912 Summer	1912	Summer	Stockholm	Tennis	Tennis Mixed Doubles, Covered Courts	Silver	UK
1750	Florence Helen Adcock (Ireland)	F	30.0	NaN	NaN	Great Britain	GBR	1912 Summer	1912	Summer	Stockholm	Tennis	Tennis Women's Singles, Covered Courts	NaN	UK
1983	Mary Ellen Adams (USA)	F	31.0	167.0	NaN	United States	USA	1920 Summer	1920	Summer	Antwerpen	Diving	Diving Women's Springboard	NaN	USA
1988	Ella Henning Anderson (Sweden)	F	17.0	NaN	NaN	Sweden	SWE	1912 Summer	1912	Summer	Stockholm	Diving	Diving Women's Platform	NaN	Sweden
1129	Selma Augusta Andersson (Sweden)	F	17.0	NaN	NaN	Sweden	SWE	1912 Summer	1912	Summer	Stockholm	Diving	Diving Women's Platform High	NaN	Sweden
1130	Selma Augusta Andersson (Sweden)	F	25.0	NaN	NaN	Sweden	SWE	1920 Summer	1920	Summer	Antwerpen	Diving	Diving Women's Platform High	NaN	Sweden
1030	Gertrude Archibald (England)	F	43.0	NaN	NaN	Great Britain	GBR	1908 Summer	1908	Summer	London	Archery	Archery Women's Double National Round	NaN	UK
1048	Fernanda Anselmi (Belgium)	F	NaN	NaN	NaN	Belgium	BEL	1920 Summer	1920	Summer	Antwerpen	Tennis	Tennis Women's Singles	NaN	Belgium
1049	Fernanda Anselmi (Belgium)	F	NaN	NaN	NaN	Belgium	BEL	1920 Summer	1920	Summer	Antwerpen	Tennis	Tennis Women's Doubles	NaN	Belgium
1055	Ethel Isabel Archibald (Scotland)	F	34.0	NaN	NaN	Great Britain	GBR	1908 Summer	1908	Summer	London	Archery	Archery Women's Double National Round	NaN	UK
1400	Beatrice Ellen Armstrong (Ireland)	F	28.0	NaN	NaN	Great Britain	GBR	1920 Summer	1920	Summer	Antwerpen	Diving	Diving Women's Platform High	Silver	UK
1902	Elin Aronsson (Sweden)	F	28.0	NaN	NaN	Sweden	SWE	1912 Summer	1912	Summer	Stockholm	Tennis	Tennis Women's Singles	NaN	Sweden
1903	Elin Aronsson (Sweden)	F	28.0	NaN	NaN	Sweden	SWE	1912 Summer	1912	Summer	Stockholm	Tennis	Tennis Mixed Doubles	NaN	Sweden

Untuk memastikan bahwa apakah benar tidak ada data berat badan pada atlet wanita pada tahun 1924 maka langsung dilakukan kroscek ke dalam dataset. Ternyata memang betul pada periode tahun dibawah 1924 tidak ada record data berat badan atlet wanita ditulis dengan keterangan NaN. Tidak hanya data berat badan ternyata data tinggi badan juga tidak ada pada periode waktu tersebut.

➤ **Variation of height along time**



Visualisasi gambar diatas menjelaskan distribusi tinggi badan atlet pria dari masa ke masa. Rata-rata tinggi badan atlet pria terus mengalami peningkatan mulai tahun 1960 hingga tahun 2016. Terjadi penurunan beberapa kali pada periode tahun tersebut namun tidak signifikan karena tahun setelahnya kembali mengalami peningkatan, jadi secara umum trennya meningkat. Untuk gap yang paling besar ada di tahun 1896 ditandai dengan tiang vertikal yang sangat tinggi namun setelah itu gap yang ada semakin kecil dan semakin kecil.



Selanjutnya adalah distribusi tinggi badan atlet wanita, polanya cukup stabil mulai tahun 1952 sampai dengan tahun 2016. Namun ada yang aneh pada periode tahun 1928 sampai 1948 dimana polanya berada diluar jalur yang cukup jauh, untuk mengulik lebih dalam nanti akan dilakukan pemanggilan data pada dataset untuk melihat ada apa pada periode waktu tersebut. Gap yang paling besar adalah gelaran 2 olimpiade pertama, yaitu tahun 1920 dan 1924 dimana hal ini ditandai dengan tiang vertikal yang tinggi. Mulai tahun 1952 sampai dengan tahun 2016 gapnya semakin baik dan semakin baik.

ID	Name	Sex	Age	Height	Weight	Team	NOC	Games	Year	Season	City	Sport	Event	Medal	region
26	Cornelia "Cor" Aalten (-Strannood)	F	18.0	168.0	NaN	Netherlands	NED	1932 Summer	1932	Summer	Los Angeles	Athletics	Athletics Women's 100 metres	NaN	Netherlands
27	Cornelia "Cor" Aalten (-Strannood)	F	18.0	168.0	NaN	Netherlands	NED	1932 Summer	1932	Summer	Los Angeles	Athletics	Athletics Women's 4 x 100 metres Relay	NaN	Netherlands
731	Majken berg	F	30.0	170.0	60.0	Sweden	SWE	1948 Summer	1948	Summer	London	Athletics	Athletics Women's Discus Throw	NaN	Sweden
1301	Iloa cs (-Zimmermann)	F	16.0	NaN	NaN	Hungary	HUN	1936 Summer	1936	Summer	Berlin	Swimming	Swimming Women's 100 metres Freestyle	NaN	Hungary
1302	Iloa cs (-Zimmermann)	F	16.0	NaN	NaN	Hungary	HUN	1936 Summer	1936	Summer	Berlin	Swimming	Swimming Women's 4 x 100 metres Freestyle Relay	NaN	Hungary
1484	Lynda Riley Adams (-Hunt)	F	16.0	NaN	NaN	Canada	CAN	1936 Summer	1936	Summer	Berlin	Diving	Diving Women's Springboard	NaN	Canada
1485	Lynda Riley Adams (-Hunt)	F	16.0	NaN	NaN	Canada	CAN	1936 Summer	1936	Summer	Berlin	Diving	Diving Women's Platform	NaN	Canada
1525	Birgitta Ann-Agnes "Bride" Adams-Ray	F	21.0	NaN	NaN	Sweden	SWE	1928 Summer	1928	Summer	Amsterdam	Athletics	Athletics Women's High Jump	NaN	Sweden
1567	Edith Addams de Habelbeinck (-Lutters, -Taylor, ...)	F	20.0	NaN	NaN	Belgium	BEL	1928 Summer	1928	Summer	Amsterdam	Fencing	Fencing Women's Foil, Individual	NaN	Belgium
1568	Jenny Marie Beatrice Addams	F	19.0	NaN	NaN	Belgium	BEL	1928 Summer	1928	Summer	Amsterdam	Fencing	Fencing Women's Foil, Individual	NaN	Belgium

Ternyata setelah dilakukan pemanggilan data, banyak data tinggi badan atlet wanita yang missing atau NaN. Hal inilah yang menyebabkan median datanya berada jauh dilaur jalur data yang lainnya. Dari 10 yang ditampilkan pada frame hanya 3 atlet wanita yang memiliki data tinggi badan selebihnya NaN.

➤ **Variation of age for Italian athletes**

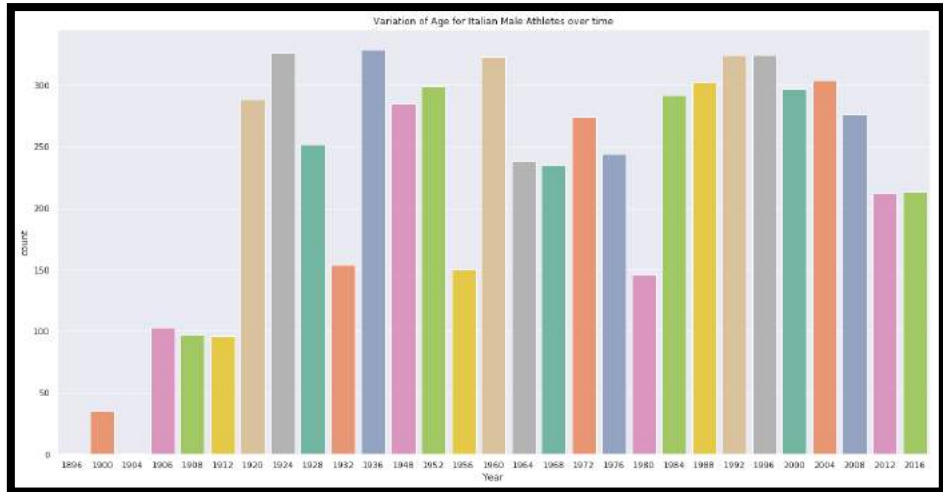


Pada pembahasan kali ini akan dilakukan analisis terkait data atlet dari negara Italia.

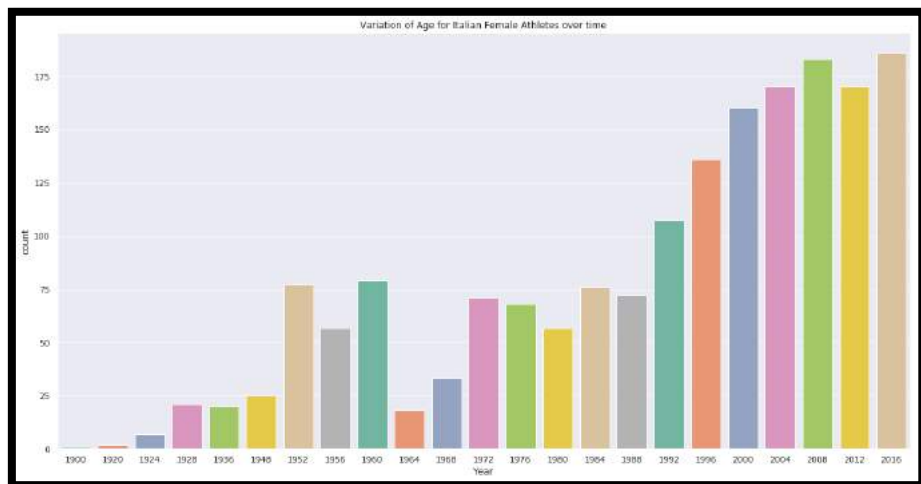
ID	Name	Sex	Age	Height	Weight	Team	NOC	Games	Year	Season	City	Sport	Event	Medal	region	note	
0	1	A Dijiang	M	24.0	180.0	80.0	China	CHN	1992 Summer	1992	Summer	Barcelona	Basketball	Basketball Men's Basketball	NaN	China	NaN
1	2	A Lamesi	M	23.0	170.0	60.0	China	CHN	2012 Summer	2012	Summer	London	Judo	Judo Men's Extra-Lightweight	NaN	China	NaN
2	3	Gunnar Nielsen Asby	M	24.0	NaN	NaN	Denmark	DEN	1920 Summer	1920	Summer	Antwerpen	Football	Football Men's Football	NaN	Denmark	NaN
3	4	Edgar Lindskou Aabye	M	34.0	NaN	NaN	Denmark/Sweden	DEN	1900 Summer	1900	Summer	Paris	Tug-Of-War	Tug-Of-War Men's Tug-Of-War	Gold	Denmark	NaN
29	10	Eitar Ferdinand "Eitar" Aalto	M	26.0	NaN	NaN	Finland	FIN	1952 Summer	1952	Summer	Helsinki	Swimming	Swimming Men's 400 metres Freestyle	NaN	Finland	NaN

ID	Name	Sex	Age	Height	Weight	Team	NOC	Games	Year	Season	City	Sport	Event	Medal	region	notes	
158	62	Giovanni Abagnate	M	21.0	198.0	90.0	Italy	ITA	2016 Summer	2016	Summer	Rio de Janeiro	Rowing	Rowing Men's Coxless Pairs	Bronze	Italy	NaN
197	81	Emanuele Abate	M	27.0	190.0	80.0	Italy	ITA	2012 Summer	2012	Summer	London	Athletics	Athletics Men's 110 metres Hurdles	NaN	Italy	NaN
198	82	Ignazio Abate	M	21.0	180.0	73.0	Italy	ITA	2008 Summer	2008	Summer	Beijing	Football	Football Men's Football	NaN	Italy	NaN
214	103	Silvano Abos	M	25.0	NaN	NaN	Italy	ITA	1936 Summer	1936	Summer	Berlin	Modern Pentathlon	Modern Pentathlon Men's Individual	Bronze	Italy	NaN
218	106	Agostino Abbagnate	M	22.0	188.0	86.0	Italy	ITA	1988 Summer	1988	Summer	Seoul	Rowing	Rowing Men's Quadruple Sculls	Gold	Italy	NaN

Pertama dilakukan pemanggilan data seluruh atlet pria yang ada pada dataset, kemudian dilakukan filter data untuk menampilkan atlet pria yang berasal dari negara Italia saja. Hasilnya seperti pada gambar diatas yang ditampilkan hanya 5 data teratas saja untuk mengecek apakah sudah betul data yang dipanggil hanya berasal dari negara Italia saja.



Selanjutnya adalah pembahasan terkait jumlah atlet pria yang berasal dari Italia dari gelaran olimpiade dari waktu ke waktu. Pada gelaran 1896 dan 1904 tidak ada perwakilan atlet pria dari Italia. Untuk tren jumlah atlet pria tidak begitu jelas polanya, namun dalam 4 gelaran olimpiade terakhir jumlahnya selalu menurun. Jumlah atlet pria paling banyak terjadi pada gelaran olimpiade tahun 1936.



Kemudian untuk jumlah atlet wanita dari Italia selalu ada mulai gelaran olimpiade tahun 1900 sampai dengan tahun 2016. Untuk atlet wanita trennya cukup jelas selalu mengalami peningkatan dari awal gelaran olimpiade sampai dengan tahun terakhir. Terjadi peningkatan yang cukup signifikan di tahun 1992 dimana jumlah atlet lebih dari 100 orang dan terus meningkat sampai dengan tahun 2016 yang sudah mencapai 175 orang atlet lebih.

- **Variation of height/weight along time for particular disciplines**

- **Gymnastics**

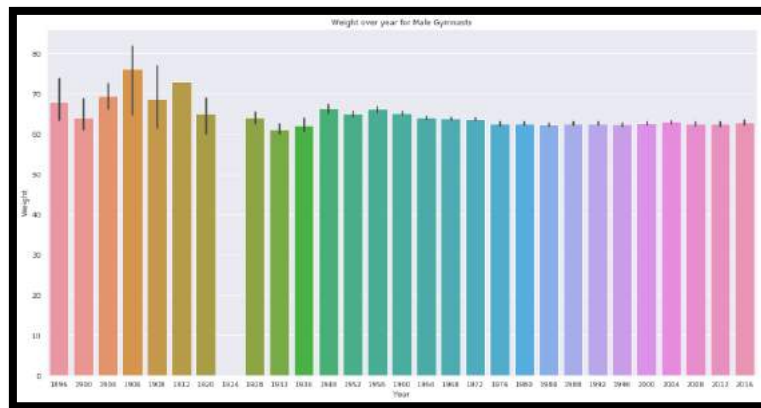
Pertama yang dilakukan adalah memanggil data pria dan wanita untuk olahraga gymnastic. Nantinya akan dilakukan analisa terkait tinggi dan berat badan atlet yang mengikuti olahraga gymnastics baik pria maupun wanita.

	ID	Name	Sex	Age	Height	Weight	Team	NOC	Games	Year	Season	City	Sport	Event	Medal	region	note
0	1	A. Dijiang	M	24.0	180.0	80.0	China	CHN	1992 Summer	1992	Summer	Barcelona	Basketball	Basketball Men's Basketball	NaN	China	NaN
1	2	A Lamusi	M	23.0	170.0	60.0	China	CHN	2012 Summer	2012	Summer	London	Judo	Judo Men's Extra-Lightweight	NaN	China	NaN
2	3	Gunnar Nielsen Aaby	M	24.0	NaN	NaN	Denmark	DEN	1920 Summer	1920	Summer	Antwerpen	Football	Football Men's Football	NaN	Denmark	NaN
3	4	Edgar Lindstrau Aabye	M	34.0	NaN	NaN	Denmark/Sweden	DEN	1900 Summer	1900	Summer	Paris	Tug-Of-War	Tug-Of-War Men's Tug-Of-War	Gold	Denmark	NaN
29	10	Einar Ferdinand "Einar" Aalto	M	26.0	NaN	NaN	Finland	FIN	1952 Summer	1952	Summer	Helsinki	Swimming	Swimming Men's 400 metres Freestyle	NaN	Finland	NaN

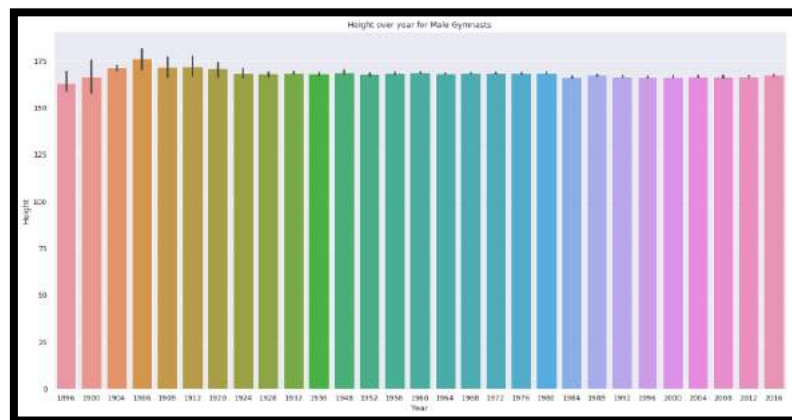
Kemudian dilakukan pengecekan penulisan gymnastics pada dataset untuk memastikan tidak ada kesalahan dalam penulisan gymnastics.

```
[ 'Basketball',
  'Judo',
  'Football',
  'Tug-Of-War',
  'Swimming',
  'Badminton',
  'Gymnastics',
  'Athletics',
  'Art Competitions',
  'Wrestling',
  'Water Polo',
  'Sailing',
  'Rowing',
  'Fencing',
  'Equestrianism',
  'Shooting',
  'Boxing',
  'Taekwondo',
  'Cycling',
  'Weightlifting',
  'Diving',
  'Canoeing',
  'Handball',
  'Tennis',
  'Modern Pentathlon',
  'Hockey',
  'Volleyball',
  'Baseball',
  'Table Tennis',
  'Archery',
  'Trampoline',
  'Beach Volleyball',
  'Golf',
  'Rugby Sevens',
  'Triathlon',
  'Rugby',
  'Lacrosse',
  'Polo',
  'Cricket',
  'Ice Hockey',
  'Racquets',
  'Motorboating',
  'Croquet',
  'Figure Skating',
  'Jeu De Paume',
  'Roque',
  'Basque Pelota',
  'Alpinism',
  'Aeronautics' ]
```

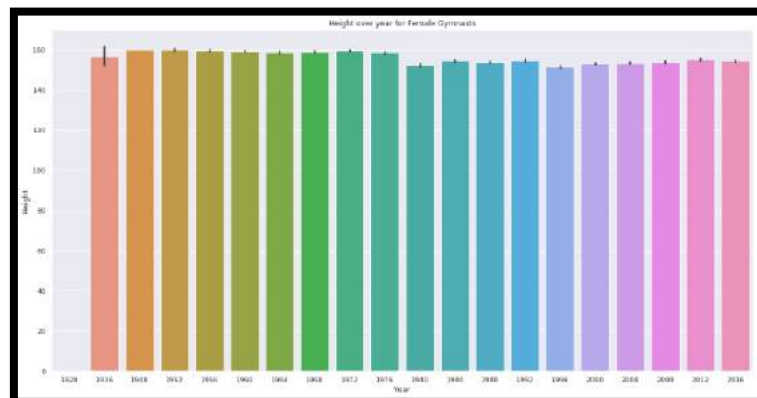
Setelah melakukan pengecekan penulisan penamaan Gymnastics sudah benar selanjutnya adalah melakukan visualisasi data atlet gymnastics untuk mendapat informasi yang ada didalamnya.



Visualisasi diatas merupakan rata-rata berat badan atlet pria untuk olahraga gymnastics selama gelaran olimpiade. Ternyata tidak ada atlet gymnastic yang ikut olimpiade pada tahun 1924 atau ada atlet yang ikut namun data berat badannya missing atau NaN, nanti akan dicoba dilakukan pemanggilan datanya pada tahun tersebut untuk dilakukan pengecekan. Rata-rata berat badan atlet pria cenderung stabil selama gelaran olimpiade, yang tertinggi terjadi pada tahun 1904 dimana rata-rata berat badan atlet hampir 80kg.



Selanjutnya adalah visualisasi tinggi badan atlet pria, dimana selama gelaran olimpiade rata-rata tinggi badan atlet pria tidak pernah kurang dari 150cm. Rata-rata paling tinggi terjadi di tahun 1906 dengan tinggi lebih dari 175cm.

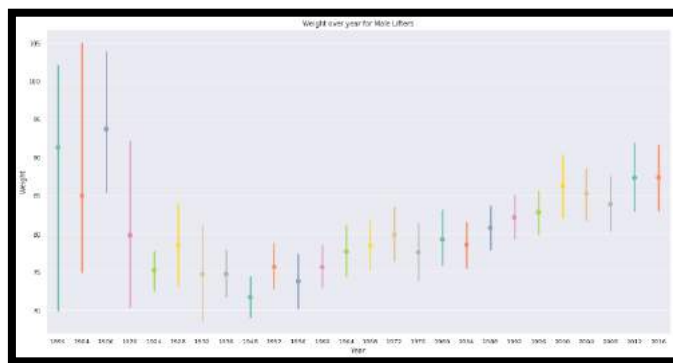


Untuk rata-rata tinggi badan atlet wanita pada olahraga gymnastics cukup stabil tidak pernah berada di bawah 150cm selama gelaran olimpiade. Rata-rata paling tinggi terjadi pada rentang tahun 1948 sampai 1976 dimana tinggi rata-rata sekitar 160cm.

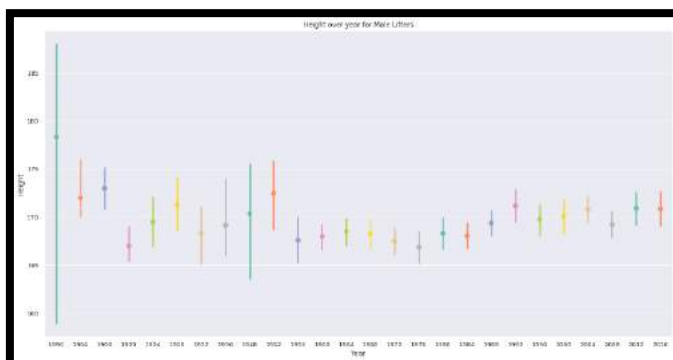
➤ **Weightlifting**



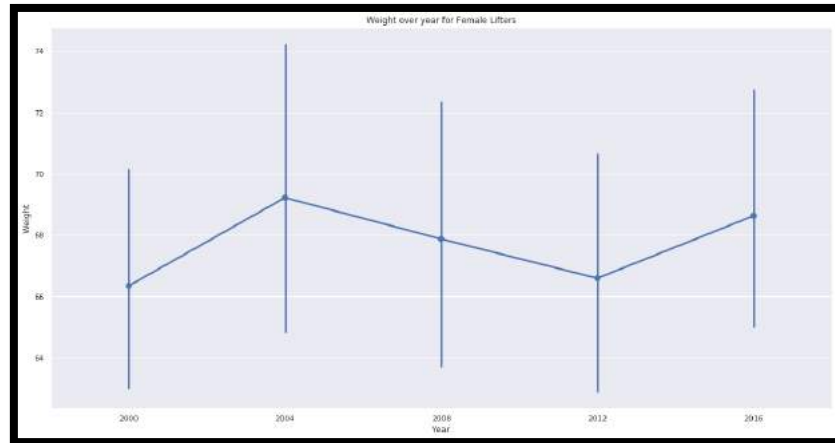
Selanjutnya adalah melakukan analisa untuk atlet angkat berat terutama untuk rata-rata tinggi dan berat badannya selama gelaran olimpiade.



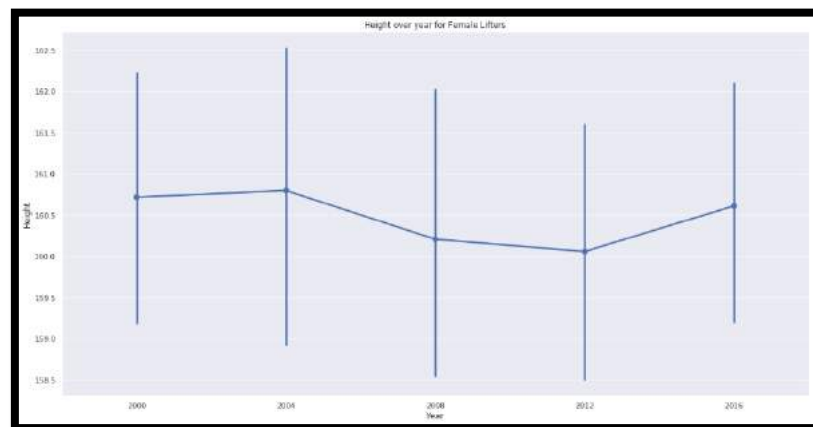
Visualisasi terkait rata-rata berat badan atlet angkat berat selama gelaran olimpiade cukup besar gapnya pada gelaran awal-awal olimpiade tepatnya periode tahun 1896 sampai 1920. Pada periode tersebut terlihat mediannya lebih dari 90kg dengan gap yang sangat besar ditandai dengan tingginya garis vertikal di awal-awal. Namun perlahan rata-ratanya mulai membaik seiring dengan bertambahnya jumlah atlet, terutama pada periode tahun 1948 sampai dengan 2016 terlihat trennya mulai jelas mengalami peningkatan.



Kemudian untuk rata-rata berat badan atlet angkat berat nilai media rata-rata berada di angka 165-170kg. Pada gambar terlihat ada atlet yang memiliki berat badan lebih dari 175kg dalam 4 gelaran olimpiade yaitu tahun 1904, 1906, 1948, dan 1952. Untuk rata-rata berat tertinggi terjadi pada tahun 1906 dan terendah terjadi pada tahun 1920 dan 1976.



Atlet wanita baru mulai mengikuti olimpiade angkat berat pada tahun 2000 di Sydney. Sebelumnya tidak ada atlet wanita yang mengikut gelaran olimpiade pada cabang angkat berat maka dari itu data pada frame tidak begitu banyak. Berat rata-rata tertinggi terjadi pada tahun 2004 dengan berat 69kg dan rata-rata terendah terjadi pada tahun 2000 dengan berat 66kg. Untuk pola trennya belum begitu terlihat karena datanya belum begitu banyak.



Untuk rata-rata tinggi atlet wanita berkisar antara 160-160,5cm selama gelaran olimpiade tahun 2000 sampai tahun 2016. Untuk rata-rata tertinggi terjadi pada tahun 2004 dan rata terendah pada tahun 2012. Tiga gelaran olimpiade terakhir rata-rata tinggi badan atlet wanita mengalami penurunan namun untuk melihat trennya belum jelas terlihat karena datanya belum begitu banyak.

8. Conclusion

Berdasarkan analisa yang telah dilakukan diperoleh beberapa poin penting diantaranya:

- Jumlah data maksimal dalam dataset berjumlah 271116, namun tidak semua kolom memiliki data yang lengkap,
- Visualisasi antara usia atlet dan perolehan medali emas distribusi datanya berupa eksponensial. Rentang usia antara 13 tahun sampai dengan 64 tahun. Distribusi medali emas paling banyak berada pada rentang usia 23 dan 24 tahun,
- Cabang olahraga dengan perolehan medali emas terbanyak dari atlet dengan usia lebih dari 50 tahun adalah equestrianism atau berkuda,
- Trend perolehan medali yang didapatkan oleh atlet wanita mengalami peningkatan dari waktu ke waktu. Perolehan medali paling banyak terjadi pada gelaran olimpiade terakhir yaitu tahun 2016 dimana atlet wanita mendapatkan lebih dari 6000 medali,
- Negara dengan perolehan medali terbanyak adalah USA atau Amerika memimpin dengan total 2627 medali, disusul Russia dengan 1599 medali, Germany 1293 medali, UK atau Inggris 657 medali, dan Italy dengan 567 medali,
- Basket menjadi penyumbang terbanyak medali emas untuk USA selama gelaran olimpiade. 186 medali telah disumbangkan oleh cabang olahraga basket sepanjang keikutsertaannya. Selanjutnya disusul olahraga renang pria 4x200 m gaya bebas 111 medali dan renang pria 4x100 m gaya kombinasi dengan 108 medali di posisi kedua dan ketiga,
- Setelah dilakukan pengelompokkan jika dihitung berdasarkan tim maka negara USA telah memenangkan medali emas basket sebanyak 15 kali, hal ini sudah dilakukan kroscek dengan wikipedia dan hasilnya cukup baik,
- Hubungan antara tinggi badan dan berat badan ini sangat baik atau positif, hubungan antara tinggi badan dan berat badan erat hubungannya. Mudah-mudahan seperti ini jika data berat badan meningkat maka tinggi badan juga meningkat,
- Ada 3 atlet yang memiliki berat badan diatas 160 kilogram dan semuanya berasal dari cabang olahraga angkat beban, berat badan ketiga atlet tersebut berturut turut sebesar 170, 163, dan 163 kilogram,
- Visualisasi terkait distribusi atlet pria dari tahun ke tahun gelaran olimpiade diperoleh nilai median dari keseluruhan data usia atlet pria berada pada rentang usia 20-30 tahun. Namun ada beberapa hal yang cukup menarik terutama pada gelaran olimpiade 1924 sampai dengan 1948 dimana terlihat banyak atlet pria dengan usia lebih dari 80 tahun,
- Ternyata memang betul ada banyak atlet dengan usia lebih dari 80 tahun yang mengikuti gelaran olimpiade yaitu cabang art competition. Ini merupakan perlombaan menggambar atau melukis sehingga jika diikuti atlet dengan usia 80 tahun masih sangat layak karena bukan merupakan olahraga yang menggunakan otot,
- Distribusi usia atlet wanita lebih bervariasi jika dibandingkan distribusi usia atlet pria terutama periode tahun 1900 – 1932, setelah periode tahun tersebut distribusi usia atlet wanita jauh lebih stabil dan memiliki nilai median 20-25 tahun,
- Visualisasi dari distribusi berat badan atlet pria, di awal-awal gelaran olimpiade rentang berat badan atlet pria cukup besar gapnya namun perlahan semakin kesini jarak gapnya semakin kecil dan semakin stabil. Terutama mulai periode tahun 1960 sampai dengan sekarang gapnya kecil dan begitu stabil,
- Rata-rata tinggi badan atlet pria terus mengalami peningkatan mulai tahun 1960 hingga tahun 2016. Terjadi penurunan beberapa kali pada periode tahun tersebut namun tidak signifikan karena tahun setelahnya kembali mengalami peningkatan, jadi secara umum trennya meningkat,
- Distribusi tinggi badan atlet wanita, polanya cukup stabil mulai tahun 1952 sampai dengan tahun 2016. Namun ada yang aneh pada periode tahun 1928 sampai 1948 dimana polanya berada diluar jalur yang cukup jauh,

- Jumlah atlet pria yang berasal dari Italia dari gelaran olimpiade dari waktu ke waktu. Pada gelaran 1896 dan 1904 tidak ada perwakilan atlet pria dari Italia,
- Jumlah atlet wanita dari Italia selalu ada mulai gelaran olimpiade tahun 1900 sampai dengan tahun 2016. Untuk atlet wanita trennya cukup jelas selalu mengalami peningkatan dari awal gelaran olimpiade sampai dengan tahun terakhir,
- Berat badan atlet pria untuk olahraga gymnastics selama gelaran olimpiade ternyata tidak ada atlet gymnastic yang ikut olimpiade pada tahun 1924 atau ada atlet yang ikut namun data berat badannya missing atau NaN,
- Tinggi badan atlet pria selama gelaran olimpiade rata-rata tinggi badan atlet pria tidak pernah kurang dari 150cm. Rata-rata paling tinggi terjadi di tahun 1906 dengan tinggi lebih dari 175cm,
- Rata-rata berat badan atlet angkat berat nilai media rata-rata berada di angka 165-170kg. Pada gambar terlihat ada atlet yang memiliki berat badan lebih dari 175kg dalam 4 gelaran olimpiade yaitu tahun 1904, 1906, 1948, dan 1952,
- Berat rata-rata wanita tertinggi terjadi pada tahun 2004 dengan berat 69kg dan rata-rata terendah terjadi pada tahun 2000 dengan berat 66kg,
- Rata-rata tinggi badan atlet wanita berkisar antara 160-160,5cm selama gelaran olimpiade tahun 2000 sampai tahun 2016. Untuk rata-rata tertinggi terjadi pada tahun 2004 dan rata terendah pada tahun 2012.