

Analys av den historiska OS-datan med fokus på Frankrike och övriga deltagande länder

Gruppmedlemmar

Ruhullah Nazary, Filip Karlsson, Arijan Berisha

Frågeställningar:

- Antal länder och vilka länder som har deltagit i OS genom åren?
- Vilka sporter har flest medaljer och hur fördelas de mellan olika länder?
- Hur ser åldersfördelningen ut bland spelarna?
- Hur har vårt tilldelade land presterat i OS över åren?
- Hur ser könsfördelningen ut genom åren?
- Vilka är de 10 länder som har tagit flest medaljer?
- Vilka diagram är mest lämpliga för att visualisera analysen?

Tekniker och verktyg som har använts:

- Hämtning av data:
 - Datan är hämtad från Kaggle
En online-plattform med fokus på datavetenskap, maskininlärning och dataanalys
- Användning av verktyg:
 - Python
 - Pandas
 - Matplotlib.pyplot
 - Hashlib
 - Plotly Express
 - Plotly Dash
 - Tabulate

Analysmetoder:

- **describe()**
Ger en statistisk sammanfattning av DataFrame eller Series
- **dropna()**
Tar bort rader eller kolumner som innehåller NaN-värden
- **info()**
Visar information om DataFrame, som antal icke-null värden, datatyper och minnesanvändning
- **nunique()**
Räknar antalet unika värden i en kolumn eller rad
- **unique()**
Returnerar en array med unika värden i en kolumn eller rad
- **value_count()**
Räknar förekomsten av varje unikt värde i en kolumn och returnerar en frekvenstabell
- **groupby()**
Grupperar data baserat på en eller flera kolumner
- **Std()**
Beräknar standard avvikelsen, vilket visar spridningen av data

Användning av olika diagram och graf:

➤ Histogram

Används för att visa fördelningen av kontinuerliga variabeln
Exempel: Åldersfördelningen bland medaljvinnare

➤ Barplot

Används för att jämföra mängder mellan olika kategorier
Exempel: Topp 10 länder som tagit flest medaljer

➤ Linjär

Används för att visa förändringarna över tid
Exempel: Antal deltagare per år

➤ Interaktiv visualisering

Användaren kan filtrera på olika parametrar
Exempel: Möjlighet att utforska datan själva

Design av dashboardapplikationen:

- **Användarvänlighet och enkel navigering**

Exempel: Lätt att navigera, enkelt att använda, passar till olika användare

- **Interaktivitet**

Exempel: Användaren kan anpassa dashboarden efter sina intresse

- **Attraktivitet**

Exempel: Anpassade färger, professionell design

- **Tydliga och informativa visualisering**

Exempel: Användning av diagram som passar bäst

Beskrivning av dataanalys och dashboardapplikation

➤ Dataanalys (1896-2016)

- Antalet deltagande länder
- Länder med flest vinnare av medaljer
- Sporter som har spelats
- Medaljtyper
- Statistik om spelarnas ålder
- Topp 10 länder med flest medaljer
- Medaljfördelning per kön
- Antalmedaljer i Sommar-OS och Vinter-OS
- Könsfördelning
- Antal medaljer per år

➤ Dashboardapplikation

- Val av sport (fotboll, basket, judo) och relaterad statistik för vald sport
- Medaljfördelning mellan länder för vald sport
- Medaljer per sport i OS
- Antal medaljer per år
- Histogram över åldrarna
- Antal medaljer per stad för Frankrike i vald stad
- Antal manliga och kvinnliga spelare för Frankrike
- Spelarnas längd i vald sport
- Val av spelarnas längd i vald sport (fotboll, basket, judo)
- Val av åldersfördelning mellan sporter (fotboll, basket, judo)
- Val av sport (fotboll, basket, judo) från olika kontinenter