

Projekt OS



Syftet med det här projektet är att använda verktygen ni lärt er i Python, databehandling och datavisualisering för att rensa och filtrera ut användbar information och skapa en dashboard.

I den här projektet ska ni hämta historisk data på OS från [Kaggle](#). Ni ska jobba i grupper på 3-4 personer.

Uppgifter

Uppgift 0 kan ni göra i Jupyter Notebook eller en/flera Pythonskript. I uppgift 1-2 ska ni använda Plotly Dash och skapa en dashboard. Uppgift 3 ska ni presentera tillsammans och uppgift 4 är individuell video. För inlämning ska varje person skicka in länk till:

- gemensamma github-repot
 - länk till individuella videon
 - länk till er deployade dash app
-

Uppgift 0 - Uppvärmning

Börja med att göra explorativ dataanalys över hela datasettet. Generella frågor att besvara med hjälp av datan:

- a. hur många länder som är med i datan?
- b. vilka länder är med? (förkortningarna räcker)
- c. vilka sporter är med?
- d. vilka medaljtyper finns det?
- e. ta reda på statistik för åldern: medelvärde, median, min, max, standardavvikelse,
- f. utforska datan vidare med egna frågor

Se även till att plotta några intressanta features, exempelvis:

- g. diagram över könsfördelningen
 - h. diagram över topp 10 länder som tagit flest medaljer
 - i. plotta gärna fler saker som är intressant.
-

Uppgift 1 - Landstatistik

Ni får ett land tilldelat till er grupp. Börja med att **anonymisera** kolumnen med idrottarnas namn med hashfunktionen SHA-256.

Undersök därefter hur det gått för landet i OS genom tiderna. Visualisera exempelvis:

- de sporter landet fått flest medaljer i
- antal medaljer per OS
- histogram över åldrar

Skapa fler plots för att visualisera flera aspekter kring ert land och dess sportprestationer i OS.

Uppgift 2 - Sportstatistik

Välj 2-4 sporter och skapa lämpliga grafer/diagram för att visualisera exempelvis:

- medaljfördelning mellan länder i sporterna
- åldersfördelning i sporterna

Skapa fler plots för att visualisera olika aspekter kring sporterna.

Uppgift 3 - Presentation

Skapa en snygg dashboardapplikation som sammanfattar det ni undersökt i uppgift 1 och 2. Deploya därefter dashapplikationen i [render](#).

Presentera därefter i grupp på ca 10-15 min/grupp där ni bland annat förklarar:

- vilka frågeställningar ni haft.
- hur ni gått tillväga för att besvara dessa frågeställningar.
- varför ni valt just de olika diagrammen/graferna
- varför ni designat dashboardapplikationen som ni gjort
- förslag är att kort beskriva er dataanalys, följt av dashboardapplikationen

Notera att **alla** i gruppen ska presentera.

Uppgift 4 - Video

Använd [OBS](#) eller teams för att spela in din skärm och dig själv där du förklarar koden som ni gjort. Notera att denna uppgift är individuell. Videon ska vara ca 5-10 min lång.

Bedömning

Om ni har fått någon kodsnuitt från någon annan eller hittat i någon sida är det **viktigt** att ni källhänvisar, annars räknas det som plagiat. Skriv en kommentar bredvid koden som ni har tagit.

Godkänt

- löst uppgift 0-4 på ett korrekt sätt
- koden är kommenterad med relevanta kommentarer
- variabelnamnen är bra valda

- gjort flera relevanta git commits
 - presentationen är tydlig och lättförståelig
-

Väl Godkänt

Uppfyllt allt för godkänt samt:

- koden är effektiv och enkel att följa
- koden är välstrukturerad med funktioner och/eller OOP
- du motiverar koden väl i videon med datavetenskapligt korrekt språk
- dashboarden är användarvänlig med väl valda visualiseringar och bra motivation
- dashboarden är enhetlig, det ska visuellt inte se ut att ni är flera som jobbat på den
- presentationen är välgenomtänkt med tydlig storytelling