

# Data Science

DATA24HEL\_DSE40

## Kunskapskontroll 3: Projektuppgift

### Beskrivning

I grupper om cirka 1-3 personer ska ni genomföra ett projekt inom data science. Projektet består av att:

- Läs in data från en SQL-databas.
- Transformera datan i Pandas, om det behövs.
- Genomföra en analys av datan i Pandas.
- Presentera datan i en webbapp med hjälp av Streamlit.

Om ni vill kan ni testa att publicera webbappen på exempelvis Streamlit Community Cloud eller annan molntjänst så att den är publikt tillgänglig. Det är inte nödvändigt för att få godkänt men är bra att ha testat på.

Ni ska presentera era resultat, och er process, i klassen på lektionen den 30/1. Presentationen ska ta max 20 minuter per grupp.

Om man inte har möjlighet att presentera inför klassen ska man spela in sin presentation och lämna in tillsammans med sin kod.

Koden ska lämnas in i en välstrukturerad Github-repository med en informativ README-fil.

Alla i gruppen kan lämna in en länk till samma repository.

## Vidare instruktioner

Det är upp till er i gruppen att välja vilken data ni vill använda. Om den kommer i ett annat format än en SQL-databas från början (csv-filer eller liknande) behöver ni först skapa en databas, till exempel genom att läsa in dem som Pandas `DataFrames` och exportera dem till en databas.

För de grupper som eventuellt inte hittar någon data att arbeta med finns det en SQLite-databas på kursens Github, `data/Köksglädje.db`, som ni kan använda. Databasen finns också att ladda ned från Learnpoint.

## Tips

- Välj ut ett antal saker i datan ni vill fokusera på. I exemplet med Köksglädje-databasen kan det vara analysera skillnader i försäljning utifrån butik, produkt, och över tid.
- Utför först analysen i en notebook innan ni gör om den till en webapp.
- Utgå från er nyfikenhet på datan!

## Bedömning

Ni kommer att bedömas på både er slutprodukt och på arbetsprocessen. Er presentation ska alltså handla både om era resultat och hur ni har arbetat tillsammans i gruppen.

## Godkänt

För betyget **Godkänt** krävs att ni genomfört de fyra punkterna ovan.

## Väl godkänt (VG)

I kursplanen står det

Den studerande har genomfört kursen och nått samtliga läranderesultat för kursen.  
Den studerande kan dessutom med säkerhet självständigt planera och genomföra enklare databehandlings- och datavisualiseringsprojekt.

För betyget **Väl godkänt** ska alltså arbetet i gruppen varit självständigt. Det innebär att ni själva försöker lösa problem som uppstår. “Med säkerhet” innebär att er kod är korrekt och följer konventioner, att er analys stämmer och att era visualiseringar är väl anpassade till det ni vill visa.

## **Självvärdering**

Förutom kod ska ni också lämna in en självvärdering där ni svarar på följande frågor:

- Vilka utmaningar har du stött på under arbetet med kursen? Hur har du löst dem?
- Vilket betyg tycker du själv att du ska ha på kursen? Motivera!
- Har du något du vill hälsa dig själv i början av kursen?

Självvärderingen ska lämnas in som ett separat dokument på Learnpoint.

## **Deadline**

Sista dag för inlämning av kod, självvärdering och eventuell inspelning av presentation är 31/1 klockan 23.59.