# Kunskapskontroll

Kunskapskontrollen består av att du skall (1) besvara några teoretiska frågeställningar och (2) göra en programmeringsuppgift där du använder SQL och Python, du kommer skriva en rapport om detta.

I Omniway kommer du skicka in en GitHub länk som skall innehålla:

1. Ditt Python skript.
2. Din skriftliga rapport.

Behöver du repetera Git och Github, se t.ex. följande video: <https://www.youtube.com/watch?v=SWYqp7iY_Tc&t=1291s>

## 1. Teoretiska Frågor

1. Beskriv kort hur en relationsdatabas fungerar.
2. Beskriv kort vad en ”left join” och ”inner join” är.
3. Beskriv kort vad indexering i SQL innebär.

## 2. Programmeringsuppgift och Rapport (Mark utformar den)

I programmeringsuppgiften kommer du använda Python för att interagera med SQL. Biblioteket som kommer användas heter SQLAlchemy.

Datan som kommer användas är: Hitta ett roligt dataset

Din uppgift består av att:

1. Göra en deskriptiv analys av datasetet. Syftet med detta är att man skall få en god överblick över datan som finns.

* Här skall du använda visualiseringar för att underlätta.

1. Göra en statistisk analys. Du skall göra ett konfidensintervall för något samt en Hypotesprövning.
2. Analysera datan. Vilka slutsatser kan vi dra?

## Rapport

När du är klar med koden skall du skriva en rapport som innehåller följande:

1. Introduktion som innehåller underrubrikerna:  
   - Bakgrund  
   - Syfte och Frågeställning
2. Databeskrivning / EDA (Exploratory Data Analysis)
3. Projekt Resultat och Analys  
   - Resultat är deskriptiva i sin natur, t.ex. att man presenterar medelålder för kvinnor respektive män. Ofta kan tabeller vara användbara vid redogörelse av resultaten.
4. Slutsats.
5. ”Executive summary”. Cirka 5-10 meningar som sammanfattar det viktigaste (detta brukar t.ex. de högsta cheferna vara intresserade av).

**Rapporten skall vara ca 3-5 sidor**. Skriv koncist och fundera på vad du vill lyfta fram.

**Efter att du är klar skall du skriva en kort redogörelse i slutet av rapporten för:**

1. **Utmaningar du haft under arbetet samt hur du hanterat dem.**
2. **Vilket betyg du anser att du skall ha och varför.**
3. **Tips du hade ”gett till dig själv” i början av kursen nu när du slutfört den.**

## Betygssättning

Kunskapskontrollens fokus är på programmering, därmed kommer ni göra flertalet programmeringstekniska val kopplat till datan. Mer specifikt:

1. Beroende på hur datan ser ut så kommer ni göra olika programmeringstekniska val, t.ex. är det lämpligt att visualisera den? Är det lämpligt att göra sammanfattande tabeller?
2. Ni kommer genomföra viss statistisk analys.
3. Ni skall visa att ni från programmeringen ni gjort kan reflektera och motivera t.ex. slutsatser på åtgärder som kanske vore intressanta för en HR avdelning eller en försäljningsavdelning.

Ni kommer skriva en slutrapport.

VG kriterierum: *”Skriftligen, självständigt reflektera över och motivera de programmeringstekniska val, som gjorts i inlämningsuppgiften, med väl underbyggda resonemang.”*

## Exempel på G respektive VG nivå

* **G nivå:** Vi ser att männen i företaget är i snitt 45 år gamla och tjänar 34 400 kr medan kvinnorna i snitt är 37 år gamla och har en medellön på 40 200 kr
* **VG nivå:** Vi ser att männen i företaget är i snitt 45 år gamla och har en medellön på 34 400 kr. Detta kan jämföras med kvinnorna som i snitt är 37 år gamla och har en medellön på 40 200 kr. Personen med högst lön i företaget är VD:n som är en kvinna. Tar vi bort henns från lönestatistiken så ser vi att männen har en medellön på 34 400 kr och kvinnorna på 32 700 kr. Denna skillnad kan tänkas bero på åldersskillnaden.

Här ser vi:   
På G nivå så har resultat tagits fram. Medan på VG nivå så har man reflekterat över valen man gjort och dessutom så har man med hjälp av programmering (t.ex. sett att personen med högst lön är kvinna) kunnat ta fram intressanta resultat som är väl underbyggda.