

# 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 så diskutera med en klasskamrat, du kan även t.ex. se följande video: [https://www.youtube.com/watch?v=SWYqp7iY\\_Tc&t=1291s](https://www.youtube.com/watch?v=SWYqp7iY_Tc&t=1291s)

## 1. Teoretiska Frågor

1. Beskriv kort hur en relationsdatabas fungerar.
2. Vad menas med "CRUD" flödet?
3. Beskriv kort vad en "left join" och "inner join" är. Varför använder man det?
4. Beskriv kort vad indexering i SQL innebär.
5. Beskriv kort vad en vy i SQL är.
6. Beskriv kort vad en lagrad procedur i SQL är.

## 2. Programmeringsuppgift och Rapport

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

Databasen som kommer användas är: **AdventureWorks2022**.<sup>1</sup>

Scenario: Du har precis börjat arbeta som dataanalytiker på ett företag där ni har en databas som heter **AdventureWorks2022**. När man börjar på ett nytt företag så behöver man bekanta sig med datan som finns. Din uppgift är nu att i Python, genom att använda SQLAlchemy:

1. Ge en deskriptiv sammanfattning över databasen **AdventureWorks2022**. Syftet är att läsaren skall få en god överblick över datan som finns. Du måste här själv bestämma vilka SQL queries du skall göra genom att undersöka och besluta om vad som är relevant. Självklart så kan du efter att du läst in datan i Python använda t.ex. Pandas och Matplotlib för din analys.  
- Använd några visualiseringar för att demonstrera intressanta aspekter av datan. Exempelvis hade man kunnat visualisera följande tabell (som tagits fram med en query) med ett

---

<sup>1</sup> <https://learn.microsoft.com/en-us/sql/samples/adventureworks-install-configure?view=sql-server-ver16&tabs=ssms>

Behöver du ladda in AdventureWorks på nytt så kolla videon vid 51:41:  
[https://www.youtube.com/watch?v=Xi0S\\_q-5t00&list=PLgzaMbMPEHEygEkBD-K6YAxyfJU2VgEu0&index=1&t=3158s](https://www.youtube.com/watch?v=Xi0S_q-5t00&list=PLgzaMbMPEHEygEkBD-K6YAxyfJU2VgEu0&index=1&t=3158s)

stapeldiagram:

	Results	Messages
	JobTitle	nbr
7	Production Technician - WC45	15
8	Sales Representative	14
9	Buyer	9
10	Marketing Specialist	5
11	Scheduling Assistant	4
12	Quality Assurance Technician	4

2. Göra en statistisk analys. Du skall i din statistiska analys även göra ett konfidensintervall för något som du själv bestämmer. Hur tolkar du resultaten?
3. Analysera datan och resultaten du tagit fram. Vilka slutsatser och rekommendationer kan du ge?

## Rapport

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

1. En presentation på fråga 1 ovan där du gjorde *"en deskriptiv sammanfattning över databasen AdventureWorks2022"*.
2. Presentera resultatet av din statistiska analys från fråga 2 ovan.
3. Skriv ned dina svar från fråga 3 ovan.
4. "Executive summary". Cirka 10 meningar som sammanfattar det viktigaste (detta brukar t.ex. chefer vara intresserade av).
5. Skriv vilket datum du gjorde din muntliga presentation av koden.

**Rapporten skall vara ca 2-5 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.

## Muntlig presentation

Sista kursveckan (se veckoplaneringen i studiehandledningen) så skall du muntligt presentera **koden** du skrivit, detta måste göras för att bli godkänd i kursen.

## Betygssättning

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

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 henne 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. För att fördjupa analysen så har jag även undersökt vilka jobbroller de olika personerna har och ...

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.

Lycka till, ha kul och kör hårt!

Antonio & Márk.