

Kunskapskontroll SQL

Kunskapskontroll SQL

Kunskapskontrollen består av att du skall

1. besvara några teoretiska frågeställningar, och
2. göra en programmeringsuppgift där du använder SQL och Python.

I Omniway kommer du skicka in en länk till en GitHub-repository. Den skall innehålla en Jupyter Notebook med:

- svaren på de teoretiska frågorna nedan,
- kod där du utför programmeringsuppgifterna nedan och presenterar resultaten, samt
- en självvärdering (se frågor nedan).

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.

Teoretiska Frågor

1. Beskriv kort hur en relationsdatabas fungerar.
2. Vad menas med “CRUD”-flödet?
3. Beskriv kort vad en “left join” och en “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.

Programmeringsuppgift

I programmeringsuppgiften kommer du använda Python och biblioteket SQLAlchemy för att interagera med SQL.

Databasen som kommer användas är AdventureWorks2022, som du kan ladda ned här (men antagligen inte behöver eftersom vi använt den genom hela kursen redan): <https://learn.microsoft.com/en-us/sql/samples/adventureworks-install-configure?view=sql-server-ver16&tabs=ssms>.

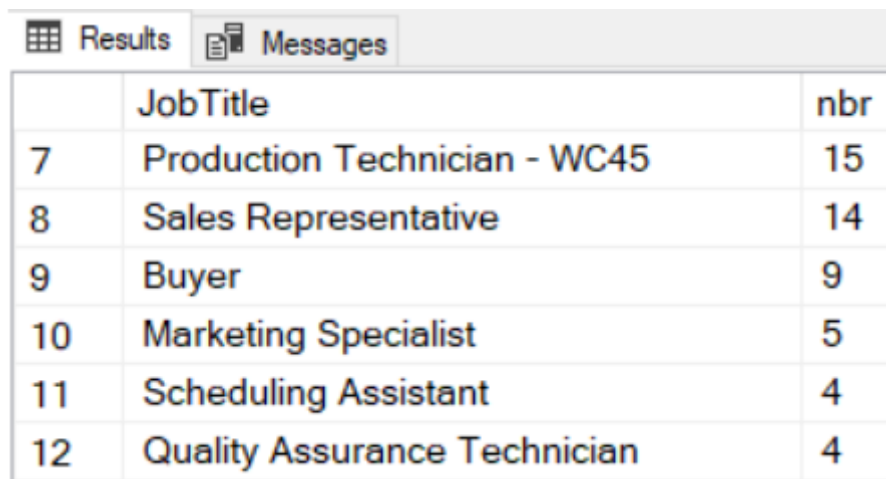
Behöver du ladda in AdventureWorks på nytt så kolla videon vid 51:41: https://www.youtube.com/watch?v=Xi0S_q-5t00&list=PLgzaMbMPEHEygEkBDK6YAxyfJU2VgEu0&index=1&t=3158s

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. Du har tre uppgifter:

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 kan du, efter att du läst in datan i Python, använda exempelvis 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 stapeldiagram:



| | 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 |

Figur 1: Exempeltabell

2. Gör en statistisk analys av valfri del av datan. Den skall innehålla åtminstone ett konfidensintervall. Hur tolkar du resultaten?
3. Analysera datan och resultaten du tagit fram. Vilka slutsatser och rekommendationer kan du ge?

Du ska lägga upp din Notebook så att den innehåller följande:

1. En presentation av svaren på fråga 1 ovan där du gjorde en deskriptiv sammanfattning över databasen **AdventureWorks2022**.
2. En presentation av resultatet av din statistiska analys från fråga 2 ovan.

3. Resultaten av analysen från fråga 3 ovan.
4. En så kallad *executive summary*, det vill säga runt 10 meningar som sammanfattar det viktigaste du kommit fram till. Detta brukar till exempel chefer vara intresserade av.
5. Skriv vilket datum du gjorde din muntliga presentation av koden.

Självvärdering

Efter att du är klar skall du även skriva en kort redogörelse i slutet av rapporten där du beskriver:

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) skall du muntligt presentera koden du skrivit. Detta måste göras för att du skall bli godkänd på kursen.

Betygssättning

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

VG-kriterium: "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 ålderskillnaden.

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

Här ovan ser vi att på G-nivå har resultat tagits fram. På VG-nivå har man reflekterat över valen man gjort, och dessutom har man med hjälp av programmering kunnat ta fram intressanta resultat, till exempel att personen med högst lön är kvinna. Resultaten är även väl underbyggda.

Lycka till, ha kul och kör hårt!

Antonio & Linus.