

## 1. Teoretiska frågor

### 1. Beskriv kort hur en relationsdatabas fungerar.

I en relationsdatabas är data uppdelad i flera tabeller. Detta för att inte spara exempelvis ortsnamn eller personers kontaktuppgifter dubbelt, utan istället referera till tabellen som informationen finns i med hjälp av unikt id. Således är tabellerna i relation till varandra och några få tabeller kan inte ge en helhetsbild. Det innebär att man kan ge olika roller tillträde till olika delar av databasen utan att riskera att känslig information hamnar hos obehöriga personer. Utöver detta krävs enbart uppdatering i en tabell mot om flera tabeller skulle innehålla samma information.

### 2. Vad menas med "CRUD" flödet?

CRUD står för: Skapa (Create), Läs (Read), Uppdatera (Update) och Ta bort (Delete).

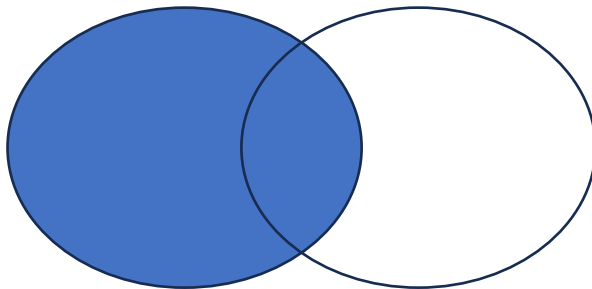
Vilket anses vara nödvändiga funktioner för att skapa en databas applikation på ett bra sätt, uttrycket används förframförallt när man arbetar med relationsdatabaser.

I SQL refererar dessa fyra ord till Insert (create), select (read), update (update) och delete (delete).

Innan data kan hittas i en databas behöver den skapas, vilket ger create som första instans, select eller read som andra instans då data ska kunna identifieras och hittas, uppdatera för att data ska kunna ändras och ta bort för att den ska kunna raderas då den inte längre är av vikt. Detta är även viktiga basfunktioner när man ska göra ett user interface, att användaren kan skapa, hitta, uppdatera och ta bort på ett logiskt sätt.

### 3. Beskriv kort vad en "left join" och "inner join" är. Varför använder man det?

Left join är när man lägger ihop två tabeller på samma unika värde och sparar data från vänstra (eller första) tabellen på de poster som det även finns data för unika värdet i högra (andra) tabellen.



### 4. Beskriv kort vad indexering i SQL innebär.

Indexering används för att snabba upp sökningar i SQL-databaser. Data dupliceras för de kolumner som index läggs på och index fungerar sedan som en pekare eller innehållsförteckning för att snabba på initiala sökandet efter data ex. i numerisk ordning eller bokstavsordning. Index blir då ungefär som en innehållsförteckning i en bok, att om man letar efter något på c startar sökningen på c med hjälp av index, men detta är något som användaren aldrig ser.

### 5. Beskriv kort vad en vy i SQL är.

En VY i SQL är en sparad sökning, en virtuell tabell som hämtar data från en riktig tabell. I själva vyn kan man inte lägga till eller ta bort data utan enbart se data från för vyn valda tabeller. Däremot kan man söka i, sortera i olika ordningar och välja vad man ska visa för delar från vyn.

### 6. Beskriv kort vad en lagrad procedur i SQL är

En lagrad procedur i SQL är när man med hjälp av kod skriver ett litet program eller en liten funktion. Det kan vara en upprepande sökning som för att genomföras behöver kontrollera om viss data redan finns för att sedan avslutas med att lägga till, uppdatera eller ta bort data.

Detta hjälper oss att minimera fel då andra ska uppdatera delar av databasen samt för oss själva då det minskar mängden kod som behöver skrivas in varje gång denna del ska genomföras.

## 2. Rapport och kod

### 2.1 Deskriptiv sammanfattning av AdventureWorks2022

AdventureWorks2022 är en sample databas från Microsoft som innehåller data rörande anställda, köpare, säljare, produkter, order samt frakter för företaget AdventureWorks.

Datamässigt är databasen uppdelad i 5 olika schemas. Dessa är Human Resources, Person, Production, Purchasing och Sales. Relationerna mellan de olika scheman är huvudsakligen BusinessEntityID för koppling avseende personer och ProductID för kopplingar gällande produkter men det finns även andra unika kopplingar mellan tabeller med specifikt innehåll.

Företaget har 290 anställda varav 17 är säljare. Totalt finns det 67 olika befattningar på företaget uppdelat på 16 avdelningar. Medelåldern på företaget är 45 år och de anställda inom produktion arbetar i treskift.

Produkterna som AdventureWorks2022 säljer är cykeldelar i fyra olika huvudkategorier med totalt 37 subkategorier. Det finns sammanlagt 504 produkter varav 248 inte har någon kulör angiven. Av produkterna är det 104 som tar längre tid än en dag att tillverka.

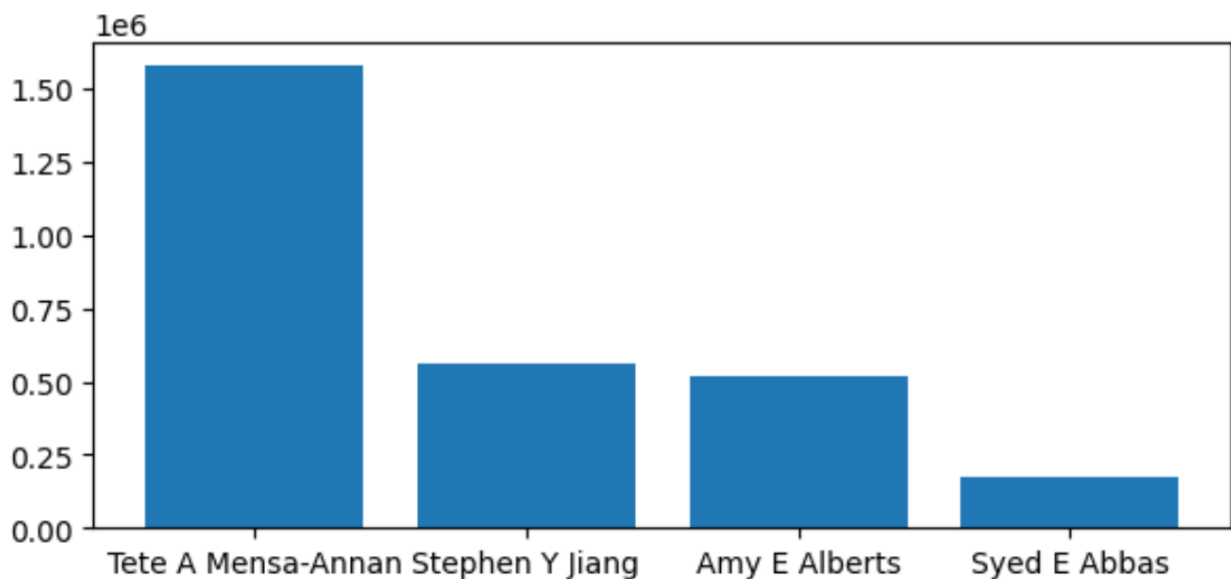
Den totala omsättningen hittills har varit 109 846 381 vad jag antar är dollar före skatt och frakter. De säljer varor till totalt 701 olika affärer och återförsäljare.

All data som presenteras i databasen kommer från åren 2011-2014.

### 2.2 Statistisk analys och konfidensintervall

De produkter som ger företaget mest vinst per produkt är Mountain Silver i olika utförande, dessa ger en vinst på 1488 dollar per såld produkt. Om man ser till vinst per produkt och antal sålda produkter är det i stället Mountain-200 Black 38 som genererat mest vinst historiskt med sina 1252 stycken sålda produkter. Det finns ungefär 200 varor som varken har StandardCost eller ListPrice, möjligen är dessa delar till cyklar eller reservdelar som de inte tar något betalt för.

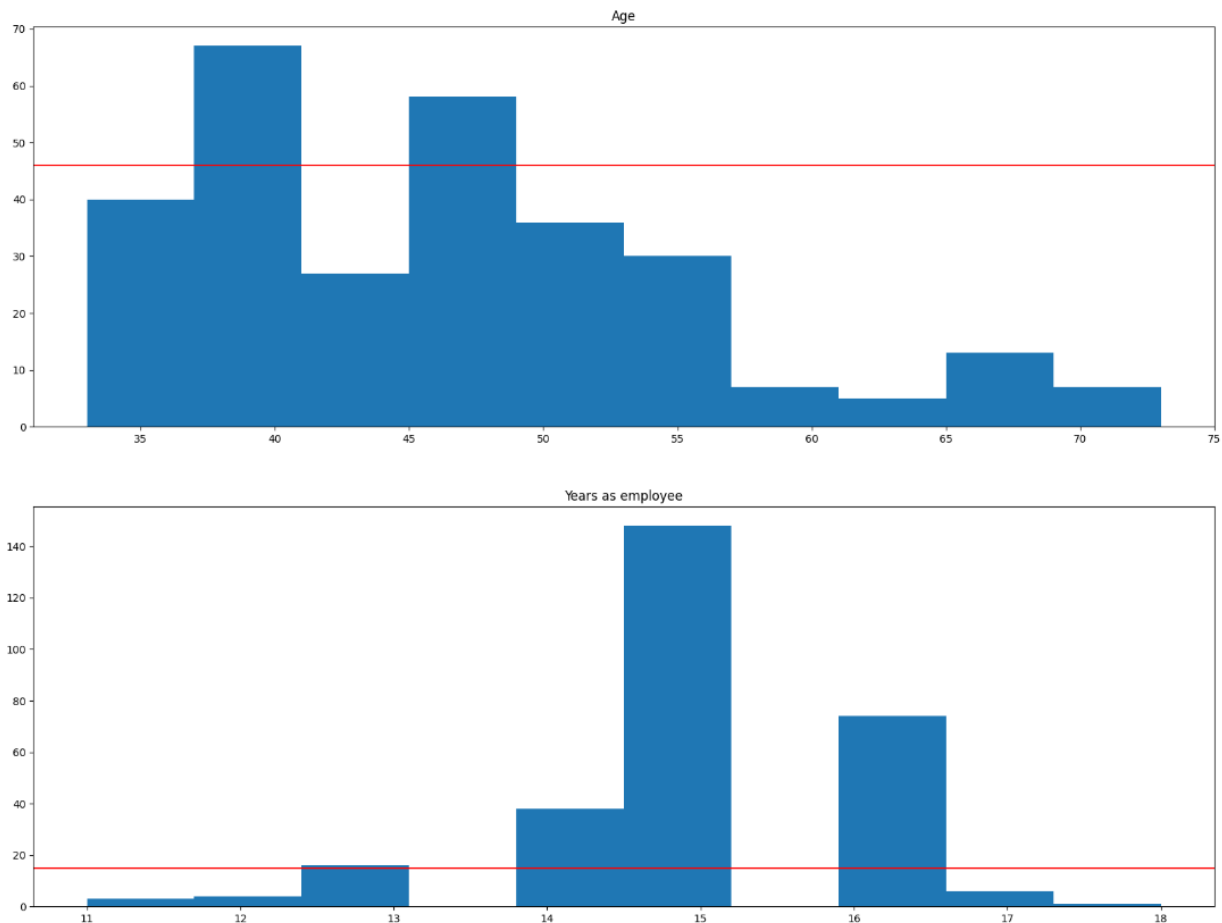
Företaget har tre säljare som hittills under 2014 inte nått upp till samma försäljning som föregående år och 10 försäljare som har uppnått eller överskridit sina försäljningssiffror från 2013. Det finns även fyra nya säljare som inte har några siffror från föregående år att jämföra sig med.



Figur som anger försäljning hittills för de fyra säljare som inte hade någon försäljning föregående år

Företaget har totalt 290 anställda, om man ska få fram en medelperson bland dessa baserat på mest förekommande roll, kön, civilstånd, lön, medelålder och antal år som anställd oberoende av varandra så är detta en:

Man, 45 år, gift, arbetat på företaget i 14 år, arbetar som Production Technician och har en medellön på 18,19 dollar per timme.



Figurer över ålder och år som anställd med inritat streck som anger medel

### Konfidensintervall

Orderstorlek i mått antal produkter per order på företaget är med 95 % säkerhet inom konfidensintervallet 57,12 till 67,08

### 2.3 Slutsats och reflektioner

Företaget kan fundera på att anställa fler kvinnor för att få en jämnare fördelning mellan könen, då det i dagsläget är mer än dubbelt så många män som kvinnor anställda. Åldersmässigt har företaget ett bra spann och samma gäller för gifta eller singlar. Gällande medellön är denna på 18,2 dollar per timme, detta stämmer bra överens med att de har många som anställda som produktionstekniker och färre som är anställda i högre roller med mer i lön. Görs en manuell sökning på anställd med medelålder 45 år och 14 år som anställd på företaget finns en man, men denne är singel och har rollen Senior Design Engineer. Så det finns ingen medelperson. Att det inte finns någon sådan kan bero på att resultaten är framtagna oberoende av varandra. Skulle man istället göra samman frågeställning med frågor beroende av varandra skulle resultatet se annorlunda ut. Det skulle då till exempel vara vanligast att vara en man är singel istället för gift om man frågar om kön och civilstånd tillsammans.

Produktionsmässigt tillverkas mertalet av företags produkter på mindre än en dag, 20 procent av produkterna tar längre tid att tillverka. Detta är cyklar och ramar till cyklar, så vill man tjäna in tid någonstans vore det på att automatisera eller se över flödet för dessa kategorier.

**2.4 Sammanfattning/Executive summary**

- 80 % av företagets produkter tar mindre än en dag att tillverka, de övriga är cyklar eller ramar
- 84 av 290 anställda på företaget är kvinnor
- det skiljer 119 dollar mellan högsta och lägsta timlönen bland de anställda
- medelålder är 45 år och tid på företaget 14 år
- vid oberoende frågeställning är den vanliga anställda man och gift medan vid beroende frågeställning är denne istället man och singel
- tio av 17 säljare har under 2014 överskridit sina försäljningssiffror från 2013
- konfidensintervall avseende orderstorlek är 57,12 till 67,08 stycken

**2.5 Datum för muntlig redovisning**

2024-01-04

### 3. Reflektion på eget arbete

#### 3.1 Utmaningar

Stora utmaningar i den här kursen har varit de privata förutsättningarna. Med barn hemma, vabb, egen förkylning och ett försök att hinna fira jul, nyår och ha lite julleighet så har det varit svårt att hitta tiden att sitta koncentrerat. Detta har jag fått arbeta runt genom att bli störd i arbetet, tappa fokus och jobba på att hitta fokus igen och snabbt komma igång från där jag avslutade arbetet senast. Svårast har varit att få tid till att genomföra LinkedIn-kurserna. En nyckel till att ändå hänga med och lära mig mycket under kursen har varit att prioritera att vara med på lektionerna även om det har varit stökigt hemma.

#### 3.2 Vilket betyg man bör ha

Detta är en kurs som jag är nöjd med om jag klarar av kraven för att få godkänt. Jag hade väldigt gärna skapat många fler och mer komplicerade frågor och presentationer men tiden har helt enkelt inte funnits.

#### 3.3 Tips att ge sig själv om göra om

Det hade varit en god idé att sänka krav på mig själv och de frågor som jag ville skriva i SQL för att istället arbetat vidare igenom hela flödet tidigare. Nu fastnade jag länge på att försöka skapa vettiga och kluriga frågor i SQL istället för att arbeta vidare med Python och rapporten för att under tiden fylla på med frågor då nya tankar dök upp.