# Teoretiska frågor

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

En relationsdatabas är en databas som förvarar data i tabeller, rader och kolumner där vissa tabeller har en relation till varandra. Detta genom unika kolumner som även kan kallas keys. De keys är likadana i olika tabeller och dessa gör det enkelt att sammankoppla data från flera olika tabeller.

1. **Vadmenas med ”CRUD” flödet?**

CRUD är en förkortning som betyder (C:Create, R:Read, U:Update och D: Delete) och de är dom fyra mest grundläggande och viktiga operationerna för att skapa och underhålla en relationsdatabas.

1. **Beskriv kort vad en ”left join” och ”inner join” är. Varför använder man det?**

En join kombinerar data från olika tabeller som har en gemensam key. INNER JOIN returnerar rader som matchar i båda tabellerna och LEFT JOIN returnerar alla rader från den vänstra tabellen och de som matchar från den högra. Så man använder det för att kombinera data från olika tabeller.

1. **Beskriv kort vad indexering i SQL innebär.**

Indexering är ett sätt att optimera sökningar i SQL. Man gör detta genom att ha en data struktur (keys) som gör att SQL programmet inte behöver läsa igenom alla rader för att hitta det man sökte efter, utan man kan optimera sökningen baserat på olika värden.

1. **Beskriv kort vad en vy i SQL är.**

En vy är en virtuell tabell som innehåller data från en fråga (Query), vyer är väldigt användbara då man kan skapa och namnge en vy av en komplex fråga som man vet att man kommer använda många gånger. På så sätt sparar man mycket tid då man bara refererar till vyns namn när man vill använda data som ens komplexa fråga genererar. Vyn kan alltså behandlas som en vanlig tabell men man kan till exempel inte lägga in (Insert) eller uppdatera (Update) data som går att göra i en vanlig tabell.

Vyn sparar ingen data utan det är bara själva frågan som sparas, så varje gång man refererar till namnet på vyn så körs frågan och returnerar data.

1. **Beskriv kort vad en lagrad procedur i SQL är.**

En lagrad procedur är en eller flera frågor (Querys) som tillsammans gör en specifik sak. Om det finns en fråga eller en funktion man använder mycket så kan man skapa en lagrad procedur för att spara tid. Så en lagrad procedur är en kod bit sparat med ett namn som man kallar på om man vill använda koden/funktionen.

# Programmeringsuppgift och Rapport

**1. Deskriptiv sammanfattning**

AdventureWorks2022 är en databas tillhörande företaget AdventureWorks.

AdventureWorks är ett internationellt tillverkningsföretag med cykelinriktning. Företaget tillverkar och säljer huvudsakligen cyklar i olika material, men också tillbehör och varor relaterade till cyklar.

AdventureWorks har 290 anställda varav 206 är män och 84 är kvinnor vilket betyder att det

är cirka 71% män och 29% kvinnor som man kan se på diagrammet nedan.

En bild som visar text, skärmbild, diagram, Rektangel

Automatiskt genererad beskrivning

Företaget säljer mestadels av sina produkter på Nordamerikanska, Asiatiska och Europeiska marknaden.

Databasen är relativt stor och innehåller väldigt mycket information som kan hittas i databasens tabeller.

Företaget har fyra olika produktkategorier (Bikes, Components, Clothing, Accessories) och 504 olika produkter som de säljer.

Förra året sålde företaget för 23 685 963,6244 och för 36 277 591,9034 det här året, vilket är en ökning på cirka 53% som man kan se på bilden nedan.

En bild som visar text, skärmbild, Rektangel, nummer

Automatiskt genererad beskrivning

1. **Statistisk analys (konfidensintervall)**

Jag har valt att kolla närmare på ledighet, specifikt på vanlig ledighet och sjukledighet. För att se om man kan se något direkt samband mellan hur mycket ledigt män och kvinnor tar och om det påverkar deras lön.

Jag använde tabellen HumanResources.Employee som grund för mina frågar då det är här man kan se ledigheten.

Min första fråga var att sen hur mycket totalen och genomsnittet var för vanlig ledighet och sjukledigt grupperat efter kön.

En bild som visar text, skärmbild, diagram, Rektangel

Automatiskt genererad beskrivning

Som man kan se ovan så är det stor skillnad på totalen men nästan ingen på genomsnittet. Det betyder att det är betydligt mer män än kvinnor på företaget vilket vi redan vet.

En bild som visar text, skärmbild, diagram, Rektangel

Automatiskt genererad beskrivning

Som vi nu vet så är det nästintill ingen skillnad på genomsnittet ledighet män och kvinnor tar. Så nu ska jag kolla på om det finns någon skillnad på lön mellan män och kvinnor.

För att göra det så kommer jag använda data från 2011–2014 som ett stickprov och göra ett konfidens intervall.

Confidence Interval for Male Average Salery: (21.190798150989806, 40.035151849010205)

Confidence Interval for Female Average Salery: (18.47988759803251, 28.98745086350596)

Som vi kan se så är det med 95% Säkerhet att männens genomsnittliga timlön är någonstans mellan 21–40 och 18–29 för kvinnorna. Konfidens intervallen överlappar så än så länge kan vi inte dra någon slutsats. Därför gjorde jag också en hypotesprövning där noll hypotesen är att det inte är någon löneskillnad och mothypotesen är att det är en statistiks säkerställd skillnad på lönen mellan män och kvinnor.

Från Hypotesprövningen blev resultatet:

Test statistic: 1.276015016680297

p-value: 0.21468622919384675

Fail to reject the null hypothesis.

Resultatet blev att inte förkasta noll hypotesen, vilket betyder att det inte finns någon statistisk säkerställd skillnad på lönen mellan män och kvinnor.

Nu när jag kommit fram till att det inte finns någon statistisk säkerställd skillnad på lönen mellan män och kvinnor och inte heller någon skillnad på ledigheten på män och kvinnor tar, så har jag kollat på vilka som tar mest ledigt på företaget.

Jag valde ut de personer som fanns i både topp 10 mest vanlig ledighet och topp 10 mest sjukledigt och fick detta resultat nedan. Jag la även in data som kan vara intressant att kolla närmare på, plus mejl och telefonnummer om man vill följa upp och se om det finns en

anledning till att personen varit borta så mycket och om det är något företaget kan hjälpa

till med.

En bild som visar text, skärmbild, nummer, Teckensnitt

Automatiskt genererad beskrivning

Som jämförelse så har jag även genomsnittet på ledigheten här:

En bild som visar text, skärmbild, Teckensnitt, linje

Automatiskt genererad beskrivning

1. **Slutsatser och rekommendationer**

Slutsatserna är att det inte finns någon eller väldigt lite skillnad på båda sortens ledighet mellan män och kvinnor. Det finns inte heller någon statistiskt säkerställd skillnad på lönen mellan män och kvinnor enligt hypotesprövningen då p-värdet var större än 5%.

Denna information är vad jag kunde ha förväntat mig innan började min undersökning och det är inget som ser konstigt ut.

Jag tycker därför att det inte behövs ändras på något ifrån den aspekten.

Jag kollade också på de anställda som tar mest ledigt och där fans det några individer som stack ut. Därför skulle rekommendera att kolla närmare på det och ringa eller mejla de

anställda för att följa upp om det finns något företaget kan hjälpa till med för att sänka

speciellt sjukledigheten.

1. **Executive summary**

Resultaten från undersökningen är vad som förväntas. Ledigheten som män och kvinnor tog var nästan exakt samma och jag kunde inte heller säkerställa at det fanns någon skillnad på lönen mellan dem. Det finns alltså inget specifikt att rapportera där.

Det finns dock en grupp med anställda som tog betydligt mer ledigt än andra.

Jag skulle föreslå att kolla lite närmare på det.

1. **Datum för muntlig presentation**

28:e december

# Reflektion på eget arbete

1. **Utmaningar du haft under arbetet samt hur du hanterat dem.**

Det finns ju alltid lite generella utmaningar man har när man ska lära sig något.

Men några specifika utmaningar var att det kännes som att det var väldigt mycket att göra i början av kursen och det var lite små stressigt. Jag gjorde så mycket jag han och sen så

minskade det man skulle göra, så då löste sig det problemet av sig självt.

En annan utmaning som jag har haft var denna kunskapskontroll. Jag känner mig väldigt

säker på själva SQL kod delen, men jag är mer osäker på den skriftliga delen.

Vi ”lärde” oss aldrig hur man skulle skriva en rapport utan bara SQL i själva kursen.

Jag löste det till slut ändå men mestadels av tiden gick till att förstå hur man skulle skriva och inte själva SQL kod delen. Kod delen kändes som den enkla delen medan rapporten var det svåra. Men jag löste ju det till slut ändå.

I framtiden kanske man ska visa någon färdig exempel rapport, då jag tror att det hade

underlättat ganska mycket.

1. **Vilket betyg du anser att du skall ha och varför.**

Jag anser att jag skall ha VG då jag känner mig väldigt säker på SQL.

Jag var inte lika säker på rapporten i början men jag tycker att jag löste det bra till slut.

Då detta är en kurs i SQL och jag är väldigt säker på kod delen plus att jag kan reflektera och resonera varför jag gjort det val jag gjort så täcker jag kraven för väl godkänt och därför anser jag att jag skall ha VG.

**Tips du hade ”gett till dig själv” i början av kursen nu när du slutfört den.**

Det enda tipset jag skulle ha gett till mig själv i början av kursen om jag kunde, hade varit att bara börja direkt även om jag inte var helt säker på vad det var jag skulle göra. Det var först när jag började kolla runt i databasen som jag hittade intressant data som jag ville kolla närmare och bygga min uppgift på.

Så det är det enda tipset jag hade velat gett till mig själv.