

## TEORISKA FRÅGOR

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

En relationsdatabas är ett system för att organisera och lagra data i en structured form där information organiseras i tabeller. Tabellen består av kolumner och rader. Kolumner definierar de typer av data som lagras i tabellen såsom CustomersID, FirstName, LastName, etc. Rader representerar ett unikt och individuellt objekt såsom namn, adress, produkter, pris, etc. I en relationsdatabas används SQL-språk för att definiera och hantera data. Relationer mellan tabeller sker när man använder "nycklar", som primärnyckel och främmande nycklar. Man kan ha en unik identifierare som CustomersID och använda den för att koppla samman information mellan olika tabeller. Det finns olika typer av relationer:

**-En-till-en (one-to-one):** när det finns en rad i den första tabellen som matchar till en rad i den andra tabellen.

**-En-till-många (one-to-many):** En rad i en tabell kan vara kopplad till flera rader i en annan tabell. Detta kan man se när en kund kan ha flera order.

**-Många-till-många (many-to-many):** Det är när många rader i en tabell är kopplad till många rader i en annan tabell. Till exempel en kund har beställt flera produkter, och en produkt har beställts av flera kunder.

### 2. Vad menas med "CRUD"-flödet?

CRUD är en akronym för Create, Read, Update, Delete. De är grundläggande operationer som kan utföras på databashanterare.

**CREATE:** används för att skapa strukturer och olika objekt i en databas såsom tabeller, databaser, index, etc.

**READ:** används för att läsa ut eller hämta data från databasen med hjälp av SELECT-kommandot.

**UPDATE:** används för att uppdatera existerande data i en tabell. Med hjälp av UPDATE-kommandot kan man modifiera data i olika kolumner och rader.

**DELETE:** Kommandot DELETE används för att ta bort en eller flera rader från en tabell baserat på ett villkor.

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

Joins används för att kombinera kolumner från en eller flera tabeller till en ny tabell.

LEFT JOIN är den som behåller alla rader från den vänstra tabellen, dvs inkluderar alla rader från den första tabellen och matchar dessa med den andra tabellen.

INNER JOIN används för att kombinera rader från två eller flera tabeller som har matchande värden i tabellerna.

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

Indexering i SQL används för att titta rader som uppfyller ett visst kriterium i en specifik kolumn. Detta gör det enklare att hitta ett visst värde i data. Utan indexering måste systemet gå igenom tabellen rad för rad för att hitta matchande rader, vilket gör att undersökning långsammare.

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

En vy är en virtuell tabell vars innehåll definieras av en Query, vilket gör data lättillgänglig i tabellformat. En vy innehåller rader och kolumner precis som en riktigt tabell. Men detta tillåter oss att kunna dela viss utvald information med användare som inte har direkt tillgång till databasen.

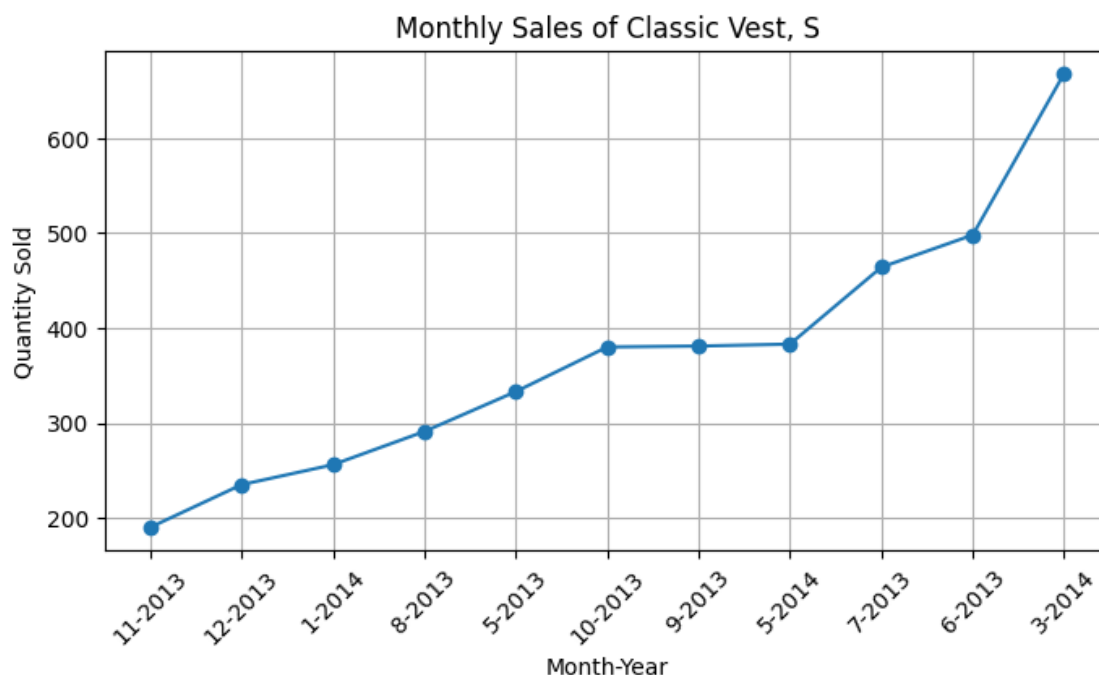
### **6. Beskriv kort vad en lagrad procedur i SQL är.**

En lagrad procedur i SQL är sparad kod som kan återanvändas, om man har en SQL-kod som man skriver flera gånger då är det bättre att man sparar den som en lagrad procedur och sedan anropa den för att köra den.

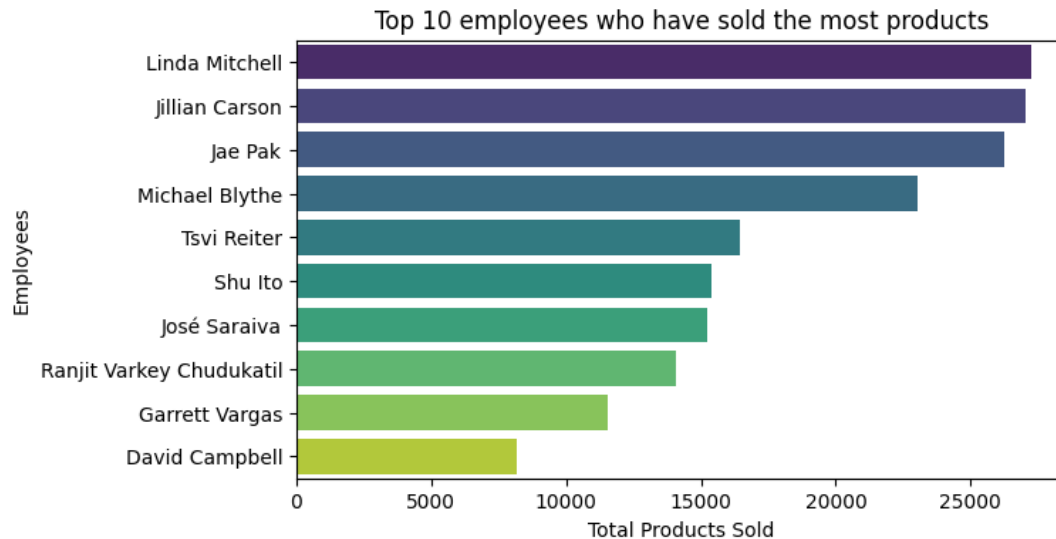
## En deskriptiv sammanfattning över databasen AdventureWorks2022

Den första tabellen jag visar i VS-code är "All Tables" från databasen AdventureWorks2022. Det är viktigt att visa alla tabeller som finns i databasen innan man granskar den. Den visar att det finns tabeller för Production.Product, Person, Sales.SalesOrderHeader, HumanResources.Employee, etc. Totalt finns det 75 tabeller i databasen. Den andra tabellen jag visar är "Product Category" som dataanalytiker på ett företag behöver man veta vad företaget handlar om, därför jag bestämde mig för att använda SQL-queries för att undersöka vilken typ av produktkategori företaget säljer. Den visar att företaget säljer fyra typer av produktkategori: cyklar, komponenter, kläder och tillbehör. Den tredje tabellen visar de försäljningsorder med det högsta antalet sålda enheter. Det är intressant att se hur många produkter finns i varje order, vad är priset per enhet, vilken beställning som har varit med fler produkter och se det totala priset. I det här fallet har jag förenklat slutpriset, men det inkluderar delsumma, skatter och frakt. Följande tabell handlar om arbetarnas information, där namn, efternamn, ID-nummer, adress och anställningsdatum visas.

Det finns även en tabell där man kan se antalet sålda produkter under en period av 3 år (2011-01-01 och 2014-01-01). Tabellen visar att produkt "Long-Sleeve Logo Jersey, L" är den bästsäljande produkten under den perioden. I min presentation har jag gjort visualiseringar av några data som jag tyckte var intressanta, såsom: månadsförsäljning av produkter, i detta fall har jag valt den mest sålda produkten varje månad.



Det var också intressant att se att det finns arbetare som har sålt fler produkter än andra. I det här fallet kan det vara så Jag har också gjort visualiseringar av de 10 bästa arbetarna som har sålt flest produkter. I det här fallet kanske en bonus kan övervägas för den arbetare som säljer mest produkter under året. Tabellen visar att det var Linda Mitchell som har sålt fler produkter, totalt: 27 229.



Jag valde att göra en statistisk analys av total försäljning. Eftersom det är intressant att se om de 10 arbetare som har sålt flest produkter har fler prestationer eller om majoriteten av arbetarna (population) säljer lika mycket i genomsnitt. I VS-code kan man se hur man beräknar ett konfidensintervall för medelvärdet av Total Sales. Resultatet var medelvärde av Total Sales är 7,002,034.54, det innebär att det genomsnittliga försäljningsbeloppet för de 10 säljarna är 7 miljoner. Konfidensintervall för medelvärde där konfidensintervall är 0,95 (95%) är: (5 178 904.07, 8 825 165.01). Detta innebär att man är 95% säker att det rätta medelvärdet för alla säljare ligger mellan 5,178,904.07 och 8,825,165.01. Det kan vi dock se att bredden på intervallen är ca 3 miljoner. Detta för att vi har valt ett väldigt litet stickprov med bara 10 observationer (säljare). För att få ett bättre och exakt konfidensintervall vore det att göra ett större stickprov med fler säljare. På så sätt skulle vi få mer exakta intervall. Företaget skulle kunna använda detta intervall (med större stickprov) för att veta vilken säljare som producerar mest för företaget och på så sätt få en bonus eller en löneökning.

Muntlig presentation: 2025-01-04.