Kunskapskontroll: SQL

Nike Sandberg Egbalewon

DS2023 Göteborg, EC Utbildning

1. Beskriv kort hur en relationsdatabas fungerar.

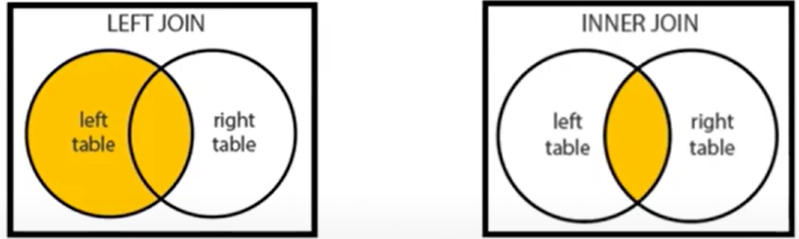
En relationsdatabas är uppbyggd av tabeller som har relationer till varandra. Tabellerna består av rader och kolumner, och innehåller data (information) som är intressant att analysera och hålla koll på. Till exempel kan vi ha tre separata tabeller med information om kunder, beställningar respektive produkter. Tabellen med information om kunderna har en koppling till tabellen med information om beställningar, som i sin tur har en koppling till tabellen med information om produkten. På så sätt kan vi hämta information om vilka kunder företaget har, vad och hur mycket de köpt, samt vilka produkter företaget säljer.

2. Vad menas med ”CRUD” flödet?

CRUD står för Create, Read, Update och Delete och sammanfattar vad vi (bland annat) kan använda relationsdatabaser till. Skapa, läsa in, uppdatera och radera. Vi kan alltså skapa tabeller och rader i databasen som är användbart för ett företag exempelvis, läsa och hämta data för att ta fram information som företaget behöver, uppdatera tabellerna/raderna med ny eller ändrad information, samt radera information i tabeller (eller till och med radera hela tabeller) som är inaktuella eller felaktiga.

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

Om man vill kombinera data från olika tabeller som har en koppling till varandra kan man göra olika typer av ”joins”. I en ”left join” används alla rader från ”vänstra” tabellens (valda) kolumner kombinerat med matchande kolumner i ”högra” tabellen. Se beskrivande bild nedan. Finns det ett rader i den ”vänstra” tabellen som inte matchar med raderna i den högra returneras ”NULL”. I en ”inner join” är det endast de matchande raderna båda tabellerna som returneras.



4. Beskriv kort vad indexering i SQL innebär.

Indexering i SQL innebär att vi kan göra sökningar i vår databas snabbare. Vi kan likna det vid en väldigt tjock lärobok. Om vi vill läsa om något specifikt skulle vi utan indexering behöva börja på sida 1 och leta oss fram sida för sida tills vi hittar det vi vill läsa om. Har vi istället ett register, ett index, i slutet av boken kan vi kolla upp vilka sidor som handlar om det vi vill läsa om. Indexering i SQL fungerar likadant. Vi skapar en typ av register för att göra sökningar snabbare.

5. Beskriv kort vad en vy i SQL är.

Man kan skapa en vy av en query, vilket innebär att det sparas en tabell över vår query som vi kan läsa och analysera om och om igen utan att behöva skriva hela kodsträngen för queryn. En vy är även användbart om vi t ex vill att en användare med begränsad behörighet kunna se viss information utan att då behöva ha tillgång till hela databasen.

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

En lagrad procedur i SQL är en kod som utför en önskad procedur i SQL, som vi har sparat för att lätt kunna anropa proceduren utan att behöva skriva hela kodsträngen gång på gång.

Rapport

**En deskriptiv sammanfattning över databasen AdventureWorks2022**

AdventureWorks är ett globalt företag med affärsverksamhet i Europa, Nordamerika och Stillahavsområdet. Den störa marknaden är i Nordamerika som har försäljningssiffror på drygt 33 miljoner USD, medan Europa och Stillahavsområdet har 13,6 respektive 6 miljoner USD. Försäljningssiffrorna utgörs till största del av cykelförsäljning, men företaget säljer även kläder, accessoarer och cykeldelar.

Cyklarna som företaget säljer kan delas in i tre olika kategorier – mountain bikes, road bikes och touring bikes. Touring bikes är den nyaste typen av cykel som företaget säljer, medan mountain bikes är den mest populära cykeln. Touring bikes började säljas 2013-05-30, men har hunnit bli den näst mest sålda cykeln i företaget. Priset är den mest förekommande anledningen till inköp av cykeln.

Företaget har 290 anställda fördelade över 16 olika enheter. ”Production” utgör den största enheten med 179 anställda.

Verksamheten utgår från USA, där 284 av de anställda bor. Övriga 6 personer är samtliga säljare som bor och arbetar från de andra länder som företaget säljer till – Frankrike, Tyskland, Australien, Kanada och Storbritannien.

En majoritet av företagets anställda är män. 206 av 290 anställda är män. Vi kan se att könsfördelningen skiljer sig åt mellan de olika enheterna.

Medelåldern bland männen i företaget är 44 år och bland kvinnorna är medelåldern 45. Bland cheferna är medelåldern 47 år för både män och kvinnor.

Vi undersöker om det finns en statistiskt säkerställd skillnad på könsfördelningen bland det totala antalet anställda och de anställda i chefsposition. I verkligheten vet vi att könsfördelningen bland olika yrken på arbetsmarknaden är en komplex fråga och har en rad olika förklaringsmodeller. Låt oss åsidosätta dessa möjliga förklaringar och anta att det endast är slumpen som avgjort om det råkar vara en manlig eller kvinnlig chef på företaget när vi undersöker om det finns en statistiskt säkerställd skillnad mellan könsfördelningen bland alla anställda, och könsfördelningen bland chefer. Jag har valt att inte inkludera ”Executives” och jobbtitlar med mer informella ledarpositioner, utan fokuserar enbart på enhetscheferna.

För att göra denna analys har jag undersökt hur många anställda med chefstitel som är kvinnor respektive män. Bland anställda med chefstitel är 35 % kvinnor och 65 % män. I hela populationen är fördelningen 29 % kvinnor och 71 % män.

Med ett 95 % konfidensintervall är mellan 12,5-58 % av de anställda kvinnor om vi antar att könsfördelningen bland cheferna är representativ för alla anställda. Eftersom vi känner till hela populationen, ser vi att andelen kvinnor bland de anställda faller inom detta intervall.

Med hjälp av en hypotesprövning kan vi undersöka om det finns en statistiskt säkerställd skillnad i könsfördelningen bland chefer och totala antalet anställda.

Vi sätter nollhypotesen att det är samma könsfördelning bland chefer som totala antalet anställda, och räknar ut p-värdet i Python med en gräns på 5 %.

P-värdet vi får ut är 0,186 (18,6 %) vilket är betydligt högre än den 5 %-gräns vi satt. Vi kan alltså inte förkasta nollhypotesen och kan inte påvisa en statistiskt säkerställd skillnad i könsfördelningen. Vi kan alltså inte påstå att det finns en bättre könsfördelning bland cheferna.

Värt att notera är att vi har ett litet stickprov då det endast finns 17 chefer på företaget vilket har en påverkan på p-värdet. Och vi får naturligtvis inte glömma att det finns andra parametrar än kön som avgör vem som arbetar som chef på ett företag.

**Executive Summary**

AdventureWorks har sin största inkomstkälla från försäljningen i Nordamerika, som står för drygt 33 miljoner USD av företagets totala försäljningssiffror på knappt 53 miljoner USD under 2014. Största delen av försäljningen kommer från cyklar. AdventureWorks har 290 anställda och verkar från USA, men har säljare i övriga länder som företaget säljer till. En majoritet av företagets anställda är män. 71 % anställda är män bland de anställda. Bland enhetscheferna är könsfördelningen något jämnare, 64,7% män respektive 35,3 %, men vi kan inte påstå att det finns en statistiskt säkerställd skillnad på könsfördelningen bland chefer och totala antalet anställda.

**Utmaningar**

De främsta utmaningarna jag haft under arbetet var att få en bild över datan och vad exakt den mäter. Jag har fått undersöka de olika tabellerna flera gånger för att förstå kolumnnamnen, samt hur tabellerna är kopplade till varandra. Jag har även använt mig av Microsofts hemsida för förklaringar på vissa tabeller.

En ytterligare utmaning var att säkerställa att jag utfört rätt typ av join. Där har jag fått rannsaka vad för information jag egentligen är ute efter, och för att inte tappa bort mig har jag använt mig av kommentarer i koden och även använt mig av nya deskriptiva kolumnnamn.

Muntlig presentation har spelats in 2023-12-27 och skickats till Mark över Teams.

Jag anser att jag ska ha G i betyg då jag genomfört alla steg för godkänt.

Tips som jag hade gett till mig själv i början av kursen är att undersöka och spara ner gjorda querys för att bättre organisera arbetet. Jag hade även gett rådet att experimentera mer med databasen och göra ändringar utan att vara rädd för att göra misstag som inte går att återställa.