## Teoretiska frågor

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

En relationsdatabas baseras på att skapa effektiva länkar mellan datablock som förbinder dem genom tabeller och kolumner baserat på relaterade egenskaper.

2. **Vad menas med ”CRUD” flödet?**

CRUD står för Create, Read, Update och Delete och är ett flödessystem för att modifiera data i tabellerna i en databas, där du kan inserta nya värden (i rader), t.ex. kunddata, och sedan läsa, uppdatera, eller ta bort data.

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

En left join lägger till alla kolumner från den första tabellen till alla matchande kolumner i den andra tabellen, vilket ger null-värden till poster som inte matchar i den andra tabellen. En inner join returnerar bara matchande värden från båda tabellerna. Man använder joins för att hämta relaterade data som inte finns i samma tabell.

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

Indexering är ett sätt för SQL att markera och ordna kolumner i tabellerna så att de kan hittas snabbare.

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

En vy i SQL är en lagrad fråga som är tillgänglig som en virtuell tabell som du skapar för att visa ett begränsat urval av avsiktligt valda kolumner som enkelt och säkert kan delas med andra.

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

En lagrad procedur i SQL är en process för att spara komplicerad kod som är användbar för att återanvändas i ett senare skede som en del av en ny query.

¨

# 2. Programmeringsuppgift och Rapport

1. **Deskriptiv sammanfattning**

AdventureWorks är ett globalt tillverkningsföretag med säte i USA, som tillhandahåller cyklar och produkter till både cykelbutiker och enskilda kunder. A pie chart with a blue and orange circle

Description automatically generated

Företaget består av närmare 300 anställda, fördelade på 6 huvudavdelningar, nämligen tillverkning, R&D, försäljning och marknadsföring, lager, kvalitetssäkring samt verkställande och allmän administration.

Knappt 30% av de anställda är kvinnor, och nära hälften av männen är över 45 år gamla.

Den huvudsakliga verksamheten bedrivs på tre kontinenter - Nordamerika, Europa och Australien - och de flesta av de närmare 20.000 kunderna finns på hemmamarknaden i USA.

A graph with lines and dots

Description automatically generated

Deras produktsortiment består av cyklar, cykeldelar, kläder och tillbehör, men deras kärnmarknad är tillverkning och försäljning av cyklar, som står för nästan 40% av all försäljning. Landsvägscyklar har varit deras bästsäljare med en total försäljning på nära 20 miljoner USD under 2012.

A graph of sales

Description automatically generated

Företaget har sålt denna produktlinje från 2011 till 2014. Närmare 94% av denna försäljning är direkt B2C-försäljning, medan resterande 6% utgörs av B2B-försäljning. Sammantaget uppgick omsättningen för 2013 till drygt 40 miljoner USD.

1. Statistisk analys (konfidensintervall)

Vi har beslutat att fokusera vår statistiska analys på försäljningsmått och ska titta på försäljning och i synnerhet försäljningsomsättning som ett avgörande KPI-mått.

A graph of sales

Description automatically generated

Vi ser här att något är konstigt med stapeldiagrammet, där det märks att försäljningsintäkterna fluktuerar mycket under de 4 år som verksamheten pågått. Vi kommer att behöva undersöka detta närmare med hjälp av statistisk analys

När vi gräver djupare i data ser vi att det saknas data för månader från både 2011 och 2014, vilket har förvrängt ovanstående stapeldiagram och således är de enda jämförbara uppgifterna mellan H1 2012 och H1 2013 samt mellan H1 2013 och H1 2014.

Detta visar en nedgång i omsättningen på drygt 1,5% mellan H1 2013 och H1 2014. Här måste vi alltså undersöka saken närmare. Även om det rör sig om en relativt liten minskning, så händer det något märkligt mellan 2013 och 2014.A graph showing a line of sales

Description automatically generated

Genom att använda oss av historisk försäljningsdata, så att vi kan se övergripande försäljningstrender över tid och på så sätt ta fram sannolika försäljningssiffror för Q3 och Q4 2014.

Efter detta har vi också skapat ett konfidensintervall med hjälp av medelvärde, standardavvikelse och observerade värden (månader) för att få ett intervall för historiska försäljningssiffror. Med hjälp av detta, den tidigare försäljningar linjen och den årliga minskningen på 1,5 % mellan 2013 och 2014 kan vi göra en preliminär analys för budgetering av nästa kvartal.

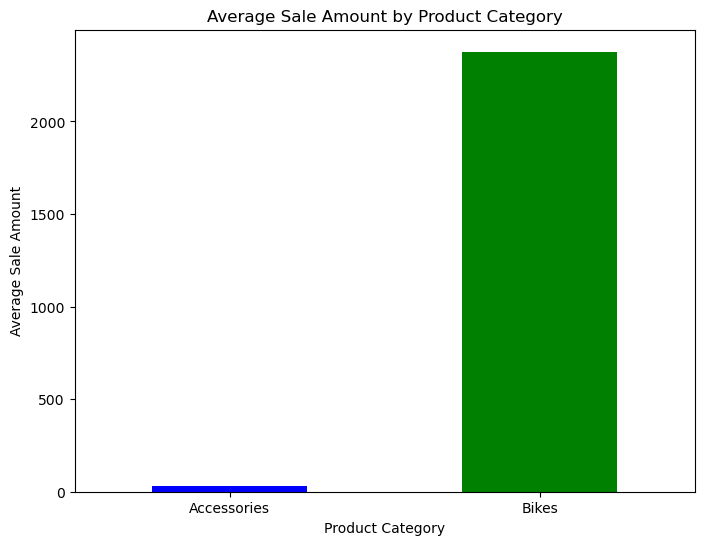
Baserat på våra historiska försäljningsdata är vi 95% säkra på att den verkliga genomsnittliga försäljningen ligger mellan ca 27 och 38 miljoner USD. Det finns en relativ normal fördelning i grafen fram till slutet av 2013 och sedan är variationen i försäljningen härifrån till slutet av Q2 2014 extremt varierad.

Vi ser här att försäljningen ökade dramatiskt både upp och ner under H1 2014, särskilt mellan februari och mars - över 8 miljoner USD - och sedan rasade i april och nådde en rekordlåg nivå nära 2 miljoner USD.

Med beaktande av de första resultaten från vår första översikt av uppgifterna, bestämde vi oss för att göra en preliminär granskning av lagersiffror (inklusive ledtider). Vi behöver mer dataanalys för 2014 för att förstå vad som händer, men med en nedgång på 1,5% jämfört med föregående år skulle jag rekommendera en billig preliminär granskning av lagersiffror (inklusive ledtider) samt en försäljningsanalys av produkter/kategorier med minskad försäljning och en kontroll av eventuella säsongsvariationer eller mönster i försäljnings minskningen.

Vi bestämde att kontrollera ledtider för lager som ett billigt sätt för företaget att snabbt kontrollera om det finns ett problem och också kunna åtgärda det snabbt utan att behöva genomgå en större kostsam och tidskrävande undersökning.

När vi genomförde denna preliminära granskning såg vi att ledtiderna här verkar vara ganska stabila från 2011 till slutet av 2013. En liten variation sker under 2014 från 7 dagar till 8 dagar. Därför rekommenderar vi att kontrollera om det finns en märkbar skillnad i försäljningsvärden mellan kategorierna i de fyra produkt linjerna för att avgöra om det skulle vara bättre att fokusera på några utvalda toppsäljande produkter.

Med bakgrund av detta har vi genomfört ett hypotestest för att kontrollera om de fyra produktlinjerna har en betydande skillnad i försäljningspris som skulle innebära att det skulle vara mer meningsfullt att titta på specifika och avgörande produkter som har en proportionellt större inverkan på ledtiderna.

Följande stapeldiagram på högra sidan visar var vårt resultat.

Hypothestestet bekräftar att P-värdet är 0 så vi kan förkasta nollhypotesen och vi konstaterar att det finns en stor skillnad i försäljning mellan kategorierna

En positiv T-statistik antyder att den första gruppen har ett högre medelvärde, så vi kommer att fokusera på ledtider för underkategorier under denna kategori, nämligen "bikes".

Baserat på detta hypotestest, om vi väljer ut underkategorier med högvärdiga produkter under Bikes, kan vi se att ”road frames” och ”road bikes” har störst påverkan på ledtiderna, även om skillnaden är ganska liten, dvs. mellan 0,2 och 1 dag över genomsnittet. A screenshot of a computer

Description automatically generated

Tillsammans med det faktum att vi vet att försäljningen av ”road bikes” har minskat kraftigt sedan mitten av 2013 innebär dock att det krävs en djupare undersökning mellan avdelningarna för att ta reda på om detta beror på interna problem inom produktion, försäljning och marknadsföring eller externa marknadsvillkor, eller en blandning av båda.

1. **Slutsatser och rekommendationer**

* Den prognostiserade försäljningen för det kommande kvartalet förutsätter att den genomsnittliga historiska tillväxttakten fortsätter in i nästa år på drygt 3,2 miljoner USD.
* Från vår övergripande dataanalys kan vi se i ovanstående försäljningstrender för toppkategorierna att det verkar finnas en oroande nedåtgående försäljningsspiral under de senaste två kvartalen, och det är möjligt att AW har gått in i en företagsövergripande "avgrund", med problem som ökar efter de inledande höjderna under 2011 och 2012.
* Även om dessa historiska försäljningsdata är användbara är det också viktigt att ta hänsyn till alla kända framtida förändringar, såsom nya produktlanseringar, marknadsföringskampanjer eller marknadsvillkor.
* Dessa ungefärliga prognoser gör att vi kan basera våra lagerbehov på försäljningstrender, vilket bidrar till att optimera lagernivåerna och minska risken för lagerbrist eller överskottslager.
* Vi rekommenderar starkt att man genomför en avdelningsövergripande undersökning av produkt-, försäljnings- och marknadsföringsstrategier och att man vid behov tar in externa konsulter för att komplettera den interna undersökningen.
* Med dessa detaljer i åtanke skulle jag vilja hävda att vi bör sätta en konservativ budget för Q3 2014, med försäljningsmål på max 27 miljoner USD.

4. **Executive summary**

1. AdventureWorks är ett globalt tillverkningsföretag baserat i USA, specialiserat på cyklar och tillbehör, med en bred kundbas främst i USA.
2. Företaget har en mångsidig arbetsstyrka på nästan 300 anställda spridda över sex huvudavdelningar.
3. En betydande del av arbetskraften består av män över 45 år, medan kvinnor utgör knappt 30% av de anställda.
4. AdventureWorks verksamhet sträcker sig över tre kontinenter, med huvudfokus på den nordamerikanska marknaden.
5. ”Road Bikes” är företagets mest framgångsrika produkt, och utgör en stor del av försäljningen som nådde 20 miljoner USD under 2012.
6. En statistisk analys avslöjar fluktuationer i försäljningsintäkterna under en fyraårsperiod, vilket kräver vidare undersökning.
7. Begränsad data från 2011 och 2014 förvränger analysen, men en jämförelse av tillgängliga data visar en omsättningsminskning på 1,5% mellan 2013 och 2014.
8. Genom att skapa ett konfidensintervall baserat på historisk försäljning kan vi förutse att genomsnittsförsäljningen ligger mellan 27 och 38 miljoner USD.
9. Med tanke på försäljningstrenderna och arbetsstyrkans sammansättning bör företaget överväga att ändra sina anställningspraxis för att rikta in sig på att anställa kvinnlig personal under 30 år, i syfte att vitalisera produktionen, försäljningen och marknadsföringen under de kommande 24 månaderna.
10. Slutligen rekommenderas en konservativ budgetering för det kommande kvartalet med hänsyn till de identifierade försäljningstrenderna och behovet av en djupgående tväravdelningsundersökning.

5. **Datum för muntlig presentation**

12-29-2023 inspelning

# Reflektion på eget arbete

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

Att utföra en första SQL-rapport på en okänd databas utan vägledande frågor kan innebära en rad utmaningar. För mig är det första steget att börja enkelt genom att utforska databasens schema, tabeller och kolumner. Denna inledande orientering är nödvändig för att bygga en förståelse för databasens struktur och innehåll. När man arbetar med okända databaser är en vanlig utmaning att identifiera relevanta datakällor och förstå hur olika tabeller och kolumner relaterar till varandra. Utan specifika frågeställningar kan det vara svårt att veta var man ska börja eller vilka frågor som är meningsfulla att ställa till databasen.

En annan utmaning är att skriva korrekt SQL-kod. Jag hade ofta stött på problem som indexering av data eller syntaktiska fel i koden. För att lösa dessa problem kan det vara mycket hjälpsamt att använda resurser som Stack Overflow, där man kan hitta lösningar på liknande problem eller ställa egna frågor. Att lära sig att effektivt navigera och utnyttja dessa resurser är en viktig del av inlärningsprocessen.

Att använda verktyg som MSSqlStudio kan också vara en utmaning i början, men det blir gradvis enklare när man vänjer sig vid verktygets funktioner och användargränssnitt. Att använda ett professionellt databasverktyg ger en mer direkt återkoppling på fel i koden jämfört med att köra SQL-skript via Python, vilket kan påskynda inlärningsprocessen.

Kollegor i klassen och datakunniga vänner kan vara ovärderliga källor till idéer och feedback. Att diskutera problem och lösningar med andra kan ge nya perspektiv och insikter som är svåra att uppnå på egen hand.

Sammanfattningsvis innebär att utforska och rapportera från en okänd databas för första gången en serie utmaningar, från att orientera sig i databasens struktur till att hantera kodningsproblem och dra nytta av tillgängliga resurser och feedback. Genom att ta ett steg i taget, testa grundligt och dra nytta av tillgängliga hjälpmedel och nätverk, kan man dock gradvis bygga upp sin kompetens och förtrogenhet med SQL och databasverktyg.

**Vilket betyg du anser att du skall ha och varför.**Jag tror att jag bör få mellan G och VG, eftersom jag verkligen kämpade ibland men tror att jag lyckades sammanställa en rapport som är insiktsfull och koncis.

**Tips du hade ”gett till dig själv” i början av kursen nu när du slutfört den.**Ett råd jag skulle ge mig själv om jag började om igen skulle vara att först fokusera på domänen och affärslogiken bakom databasen och sedan fokusera på koden för att göra detta, med inriktning på en kortare rad av upprepade gånger använd kod i SQL.

Att förstå databasens affärslogik och domän är avgörande för att kunna ställa relevanta frågor och extrahera meningsfull information. Att veta vad informationen representerar och hur den används i verksamheten kan ge vägledning i hur man navigerar i databasen och vilka frågor som är viktiga att ställa.

När det gäller kodningen är det användbart att fokusera på en mindre uppsättning av ofta använda SQL-kommandon och mönster. Detta gör det lättare att bygga och förstå frågor utan att överväldigas av SQL:s fullständiga komplexitet. Det kan också vara effektivt att skapa en personlig "cheat sheet" eller referensguide med vanliga kommandon och syntax för snabb åtkomst under kodningsprocessen.