# **Teoretiska Frågor**

Beskriv kort hur en relationsdatabas fungerar.

* En relationsdatabas fungerar som en vanlig databas där data är organiserat i tabeller och kolumner. Det som gör relationsdatabaser unika är att de även specificerar relationer mellan olika typer av tabeller med hjälp av ”keys” unika identifierare. Detta gör att man snabbt kan koppla samman data från olika tabeller på ett korrekt sätt.

Vad menas med ”CRUD” flödet?

* Uttrycket CRUD är en förkortning av Create, Read, Update och Delete. Create står för skapandet av en databas och dess komponenter. Read står för att läsa och använda sig av datan i databasen. Update innebär att man kan uppdatera data som finns i databasen för att hålla den aktuell. Delete står för att man kan radera delar av datan eller hela databasen. En effektiv och användarvänlig databas bör bygga på CRUD.

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

* Uttrycket ”join” innebär att man för samman data från olika tabeller (ibland även från kolumner inom samma tabell (self join)). En ”left join” innebär att man för samman all data från en tabell med all data från en annan tabell, även NULL-värden kommer med. En ”inner join” innebär att man endast för samman data som finns i båda tabellerna.

Beskriv kort vad indexering i SQL innebär.

* Indexering används i SQL för att enklare sortera olika kolumner. Om vi gör en query på en kolumn utan indexering kommer sql behöva gå igenom hela kolumnen för att hitta värdet vi söker efter. Om vi istället gör en query på en kolumn som har ett index kan SQL snabbare hitta det värdet vi söker efter. Indexet fungerar lite som en vägvisare för SQL när programmet letar efter data.

Beskriv kort vad en vy i SQL är.

* När vi gör en query i SQL får vi fram en tabell som motsvarar det vi skrivit i queryn. Istället för att skriva samma query varje gång vi vill få fram just den tabellen så kan vi spara tabellen som en vy i sql. Vi kan nu enkelt kalla på vyns sparade namn istället för att skriva en hel query som kan innehålla många constraints.

Beskriv kort vad en lagrad procedur i SQL är.

* En lagrad procedur kan ses som en typ av funktion. Vi skapar funktionen och definierar vilken vår variabel kommer vara. När vi sedan vill använda funktionen/lagrade proceduren behöver vi bara specificera variabelns värde för att få fram ett resultat från den lagrade proceduren.

# **Rapport – programmeringsuppgift**

# Deskriptiv sammanfattning över AdventureWorks2022

## En bild som visar text, skärmbild, Teckensnitt, nummer Automatiskt genererad beskrivningÖversiktligt

AdventureWorks är ett företag som inriktar sig på försäljning av sportartiklar.

I tabellen till höger kan vi se information om de högst uppsatta i företaget.

Företagets CEO är Ken Sanchéz som jobbat på företaget sedan januari 2009.

En bild som visar text, skärmbild, skärm, diagram

Automatiskt genererad beskrivning

## Anställda och könsfördelning

Om vi kollar på totala antalet anställda hos företaget ser vi att det rör sig om 290 personer varav 84 är kvinnor och 206 är män.

En bild som visar text, skärmbild, diagram, Teckensnitt

Automatiskt genererad beskrivning

Detta ger en könsfördelning på 29% kvinnor och 71% män i företaget.

## En bild som visar text, skärmbild, diagram, Graf Automatiskt genererad beskrivningTotal årlig försäljning

Här ser vi hur mycket företaget säljer för varje år (åren 2011–2014). Viktigt att notera är att försäljningstotalen på y-axeln visas som tiotals miljoner. (Exempel: under 2011 uppgick den totala försäljningen till 12,641,672 USD.)

Vi kan tydligt se att försäljningen ökade under åren 2011–2013 för att sedan drastiskt minska 2014. Det vi inte ser i figuren är att både år 2011 och 2014 saknar försäljningssiffror för vissa månader.

En bild som visar text, skärmbild, Teckensnitt, Graf

Automatiskt genererad beskrivning

## Försäljning per månad

När vi kollar närmre på försäljningen per månad så ser vi istället att försäljningen svängt rejält under samtliga år. Vi kan se en uppåtgående trend mellan maj 2011 och mars 2014 för att sedan sjunka mellan mars-juni 2014.

En bild som visar text, skärmbild, nummer, Teckensnitt

Automatiskt genererad beskrivning

## Mest sålda varor

Om vi kollar närmare på vilka varor AdventureWorks säljer så ser vi att företagets mest sålda artiklar är kläder, vattenflaskor och hjälmar.

# Statistisk analys

Företagsledningen hos AdventureWorks vill stärka sin konkurrenskraft globalt. För att göra detta behöver de följa med i utvecklingen som sker inom branschen. Det blir vanligare bland konkurrenterna att locka kunder med realisationer av olika slag. AdventureWorks ledning har därför gett i uppgift till sin försäljningsavdelning att titta närmare på hur företaget genom realisation och andra erbjudanden skulle kunna öka försäljningen av varor. Försäljningschefen är dock en klok kvinna som inte vill göra något förhastat utan att ha belägg för att det kommer funka. Därför har man anställt en konsult inom Data Science (jag) för att med hjälp av statistik se hur olika typer av erbjudanden påverkar försäljningen.

Uppdraget som jag har fått är följande:

”Kan vi se att en större realisation verkligen leder till en större försäljning?”

Detta är viktigt att undersöka för företaget. För om en större realisation inte påverkar försäljningen finns ingen anledning till att genomföra den.

## Metodik och teori

Jag kommer att ta mig itu med frågan genom att kolla på redan genomförda realisationer och jämföra deras slagkraft hos kunderna. Jag har valt att jämföra realisationen av hjälmar som först skedde under 2012 30 juni-28 juli (30 dagar) och jämföra den med realisationen av hjälmar som skedde under samma period 2013.

Mitt urval motiverar jag med att reorna är jämförbara då de var lika långa och skedde under samma tid på året. Även samma varor ingår båda åren i reorna. Det rör sig om tre olika typer av cykelhjälmar med produktID 707, 708 och 711. Dessa produkter säljs även i stora kvantiteter vilket gör att de lämpar sig för statistisk analys.

Det som skiljer de olika reaomgångarna åt är endast själva rea-nivån. Under reaperioden 2012 låg rean på 10 procent av ursprungspris men 2013 ökade man rean till 15 procent. Eftersom de andra variablerna hålls statiska mellan perioderna så kan vi med större säkerhet säga att en förändring endast beror på rea-nivån.

I de inledande samtalen med företaget är den populäraste teorin att ”såklart en bättre rea gör att vi säljer fler varor”. Nu ska vi dock testa detta genom att ta datan från de utvalda reaperioderna och använda de i ett konfidensintervall. Detta för att kunna svara på frågan om det är statistiskt säkerställt att det finns en skillnad i försäljning beroende på rea-nivå.

## Analys

Totalt under 2012 såldes 4184 stycken cykelhjälmar med produktID 707, 708 och 711. Under reaperioden (2012-06-30 00:00 till 2012-07-29 00:00) såldes totalt 492 stycken.

Under 2013 såldes 9116 stycken cykelhjälmar med produktID 707, 708 och 711. Under reaperioden (2013-06-30 00:00 till 2013-07-29 00:00) såldes totalt 815 stycken.

Om vi skulle jämföra antal sålda varor direkt mellan reaperioderna skulle vi dock stöta på problem. Som vi sett i den deskriptiva sammanfattningen ovan så såldes överlag fler varor under 2013. Detta innebär att det även såldes fler under reaperioden 2013 oberoende av realisationen. Istället för att direkt jämföra antal sålda varor under reaperioderna så bör vi istället jämföra andelen av varorna som sålts under reaperioderna det valda åren.

Andel för 2012:

Andel för 2013:

Vi kan alltså se att under 2012 skedde 11,8 procent av försäljningen under reaperioden. Under 2013 skedde 8,9 procent av försäljningen under reaperioden. Dessa siffror visar redan på att vår hypotes om att en högre rea leder till fler sålda varor inte verkar stämma. Men låt oss inte stanna här. För att verkligen testa vår hypotes bör vi även göra ett konfidensintervall för att se om dessa andelar är skilda och om skillnaden är statistiskt säkerställd.

För att kunna räkna ut ett konfidensintervall använder vi följande ekvation. Vi kollar om det är statistiskt säkerställt att det såldes en högre andel varor under rean 2012 än under rean 2013. Vi har två oberoende stickprov och vill räkna ut ett konfidensintervall för skillnaden mellan andelar:

Där:

p1 = 0,118

p2 = 0,089

z = 1,96 (95-procentigt konfidensintervall)

n1 = 4184

n2 = 9116

Om vi sätter in våra värden i funktionen för konfidensintervallet för vi följande konfidensintervall som resultat:

Eftersom värdet 0 inte är inkluderat i konfidensintervallet kan vi säga att det finns en statistisk skillnad mellan andelarna sålda varor under reaperioderna 2012 och 2013. Skillnaden är dock inte den vi hade förväntat oss. Vi har precis visat på att rean under 2012 när rea-nivån var 10 procent hade en högre andel av den årliga försäljningen under 2012 än vad rean under 2013 på 15 procent hade av den årliga försäljningen 2013. Vi kan alltså säga att den 10 procentiga rean hade bättre genomslag i försäljningen under den 15 procentiga rean det följande året. Men varför?

## Slutsatser och rekommendationer

Vår statistiska analys har visat på att en högre rea-nivå inte leder till en ökad försäljning. Detta går mot logiken i att en högre rea-nivå är fördelaktig för kunden och att kunden kommer köpa mer ju högre rea-nivå det är.

Denna slutsats lämnar oss med fler frågor än vad vi började med. Vi kan dock nu jobba vidare genom att kolla på vad som kan ha påverkat försäljningen i övrigt. Företaget vet nu att bara höja rea-nivån inte alltid leder till ett bättre resultat.

Framåt hade jag rekommenderat att man djupdyker i utformningen av rean och framförallt marketing-delen. Hur marknadsfördes rean under 2012 jämfört med rean 2013? Det kan finnas skillnader i förmågan att nå ut till kund med information om rean under 2013. Här skulle vi kunna skicka ut reklambladen från 2012 och 2013 till en slumpmässigt vald grupp av kunder och kolla hur de svarar på de olika reklambladen. Även mer djupgående intervjuer med ett fåtal slumpmässigt utvalda kunder kan vara en bra idé för att få mer djupgående svar om hur de uppfattat informationen om rean.

Ytterligare en punkt som är viktig att kolla vidare på är konkurrensen från andra företag och hur den skilde sig från reaperioden 2012 och 2013. Om ett annat eller flera andra företag hade förmånliga reor samtidigt under 2013 kan detta ha påverkat företagets försäljning. I framtiden måste man vara uppmärksam på vad konkurrenter gör och försöka matcha deras försäljningsstrategier för att kunna fortsätta vara en relevant aktör på marknaden.

I en global och konkurrensutsatt bransch kan man inte förlita sig enbart på att höja rea-nivån utan man måste i detalj utforma den så att den slår igenom, framförallt digitalt. Man måste även hålla ett öga på sina konkurrenter för att inte hamna bakom i utvecklingen.

## Executive summary

AdventureWorks är ett företag som under åren 2012–2013 erfor en drastisk expansion. För att företaget även i framtiden ska fortsätta att utvecklas på ett konkurrenskraftigt sätt så krävs det att man satsar på att utveckla sina avdelningar inom sälj och marknadsföring. Rean som genfördes 2013 var inte lika genomslagskraftig som den som genomfördes 2012 trots att företaget expanderade överlag. Om man inte lägger mer fokus på att spetsa till sin försäljning och marknadsföring riskerar man att falla bakom sina konkurrenter och långsiktigt försvaga företagets ställning på marknaden. Jag rekommenderar därför att man i detta läge satsar på utveckling inom just försäljning och marknadsföring för att kunna förbli en relevant aktör på marknaden.

## Datum för muntlig presentation: 2 januari 2024

# **Reflektion**

1. Utmaningar du haft i arbetet samt hur du har hanterat dem.

* Den största utmaningen med uppgiften har varit att själv välja ämne och begränsa det. Databasen är stor och det finns många intressanta saker att undersöka och skapa frågeställningar utifrån. I början la jag ner mycket tid på att kolla runt i databasen och testa olika frågeställningar. Till slut var jag dock tvungen att avgränsa mig i mitt arbete. Efter att jag hade gjort det så gick det snabbare att komma framåt med uppgiften. Jag har även personligen haft det svårt med pandas och matplotlib i Python. Jag behövde repetera biblioteken från grunden igen. Detta har dock gjort att jag lärt mig mycket samt motiverat mig att fortsätta repetera dem för att känna mig mer bekväm när jag använder dem. Det har även varit tufft att göra uppgiften under julhelgerna då det händer mycket annat då som gör att man lätt tappar fokus och motivation att sätta sig ner med uppgiften. Jag har försökt att lösa detta genom att tydligt bestämma vilka dagar jag är ledig och vilka dagar jag studerar.

1. Vilket betyg du anser att du ska ha och varför.

* Jag anser att jag ska ha betyget väl godkänt i kursen. Detta eftersom jag har svarat på samtliga teoretiska frågor och genomfört hela rapporten. Jag har i min deskriptiva sammanfattning visat upp datan visuellt samt även grävt djupare för att visa att vi inte kan kolla på ’yearly sales’ direkt utan måste bryta ner datan och kolla på ’monthly sales’ för att den inte ska vara missvisande. I min rapport har jag tydligt motiverat mitt case. Jag har sedan beskrivit vilken metod jag kommer använda mig av för att besvara frågeställningen. I slutsatsen har jag, trots att jag fick fram ett oväntat resultat, kommit med rekommendation kring vad företaget bör undersöka vidare för att få klarhet i frågan kring realisationer. I helhet har jag visat på att jag vet vad en relationsdatabas är samt hur man kan använda SQL och Python för att få fram intressant data från en sådan databas.

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

* Mitt viktigaste tips till mig själv är nog att inte vara så stressad. Överlag så angriper jag frågan på ett systematiskt sätt som fungerar för mig så det finns ingen anledning till att reagera med stress på en sådan uppgift. Jag hade tipsat mig själv att se uppgiften som en bra repetition av Python och dess bibliotek. Jag hade även tipsat mig själv om att tidigare i processen ta kontakt med kursare för att diskutera uppgiften då det finns mycket att lära av varandra.