

1. Teoretiska Frågor

1.1 Beskriv kort hur en relationsdatabas fungerar.

En relationsdatabas är ett system för att organisera och hantera stora datamängder. Den grundar sig på tabeller för att samla data. Varje rad i tabellen företräder en datamängd och varje kolumn företräder olika attribut. Tabellerna är kopplade via relationer, oftast genom nycklar (primär/foreign). Genom dessa relationer kan en relationsdatabas utföra komplexa queries och analysera över flera tabeller. En annan aspekt är att det tillåter effektiv dataåtkomst och integritet genom att undvika redundans och försäkra konsekvens i datahanteringen.

1.2 Vad menas med "CRUD" flödet?

CRUD är en akronym som beskriver fundamentala operationerna i databashantering och utveckling. C står för **Create** och handlar om att skapa nya dataposter; t.ex. lägga till nya rader i en tabell. R står för **Read** som innebär att det läser eller hämtar befintliga data. Alltså att hämta specifika rader eller kolumner från en tabell eller utföra komplexa queries för att hämta data baserat på vissa kriterier. U står för **Update** som innebär att befintliga dataposter uppdateras. Det innefattar att ändra data i en databas; t.ex. uppdatera en användares information. D står för **Delete** och handlar om att ta bort dataposter; t.ex. radera rader från en tabell.

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

Left join och inner join är ett koncept i relationsdatabaser som används för att sammanställa och analysera relaterad data från olika tabeller. **Left join** returnerar alla rader från vänster tabell och gemensamma rader från högra tabellen. Man använder left join i situationer när man behöver alla poster från en primär tabell även om det finns ett samband med datan i den andra tabellen. **Inner join** returnerar endast de rader som finns i båda tabellerna och används i situationer där man extraherar endast de datarader som har en tydlig relation i båda tabellerna, vilket resulterar till en exakt och begränsad datamängd.

1.4 Beskriv kort vad indexering i SQL innebär.

Indexering är en teknik som hjälper att förbättra hastigheten för dataåtkomst. Att skapa ett index på en eller flera kolumner i en tabell leder det till att separata datatrunkter skapas som sedan snabbt kan lokalisera rader. Indexering är effektivt för att optimera sökningar och frågor men det ökar lagringsanvändningen och kan göra datainsättning och uppdatering långsammare, eftersom indexet måste uppdateras. Om indexeringen implementeras på ett genomtänkt sätt kan det möjliggöra en optimal balans mellan snabb dataåtkomst samt effektiv databashantering.

1.5 Beskriv kort vad en vy i SQL är.

En vy är en virtuell tabell som skapas genom en förfrågan på en eller flera befintliga tabeller i databasen. När man skapar en vy i en databas, abstraherar man de komplexa query och underliggande databasstrukturer till en enklare och mer lättförståelig form. Detta ger flexibilitet och effektivitet i databashantering. De används även för att förbättra säkerhet, begränsa åtkomst till specifika data. Det är bra t.ex. för att ge en personalavdelning enkel och säker åtkomst till anställdas kontaktinformation utan att utsätta känslig information som löne- eller prestanda data.

1.6 Beskriv kort vad en lagrad procedur i SQL är.

En lagrad procedur i SQL är som ett förinställt kommando i en databas. Det omfattar en samling av SQL kommandon som sparas under ett namn och kan köras med bara ett anrop. Denna procedur möjliggör automatisering och återanvändning av komplexa databas operationer. En lagrad procedur i SQL används oftast för att effektivisera återkommande uppgifter, transaktionshantering samt datavalidering. Man behöver inte direkt åtkomst till databasen och detta medför att säkerheten ökar. Det förbättrar prestandan genom att minska nätverkstrafiken mellan klient och server. Det förenklar underhåll genom att centralisera logiken.

2. Rapport

2.1 Personal analysen

Syftet med denna analys är att få en överblick över anställda i olika avdelningar. Jag använde mig av en SQL query för att samla data om antal anställda och deras genomsnittsålder per avdelning. Resultatet visar på att produktionsavdelningen har det högsta antalet med över 175 anställda. Detta är ett tecken på att företaget fortfarande litar sig mycket på manuell arbetskraft trots teknologiska framsteg. Det understryker också företagets fokus på tillverkning och produktutveckling.

Företaget är en cykeltillverkare och därför är produktionen kärnan i deras verksamhet. Jag skulle föreslå att företaget överväger att tilldela resurser för att upprätthålla och förbättra effektiviteten i produktionsavdelningen. Detta innebär till exempel investeringar i avancerad utrustning och teknologi för att öka produktiviteten och effektiviteten. Detta kan innefatta att implementera lean manufacturing principer eller kvalitetssäkringssystem.

När vi kollar på genomsnittsåldern ser vi att det är högre inom ingenjers- och underhålls avdelningarna. Detta kan tyda på en hög erfarenhetsnivå men också signalera en kommande pensions-våg, vilket kan kräva strategiska åtgärder såsom rekrytering eller kompetensutveckling.

2.2 Leverantör analysen

I syfte att granska leverantör prestationen inom företaget har jag skrivit en query som samlar data om både orderstorlek och ordervärde. Jag prioriterade leverantörer baserat på deras totala försäljningsvärde samt genomsnittlig orderstorlek. Resultatet jag fick fram var att Superior Bicycles var ledande i termer av order kvantitet. Det här resultatet kan tyda på konkurrenskraftiga priser eller starka relationer. En annan aspekt jag fram är att stora order ofta korrelerar med mindre enhetskostnader. Denna insikt skulle kunna vara bra för framtida förhandlingar.

Vidare fick jag fram att det finns sammanlagt 104 leverantörer i databasen. Däremot upptäckte jag att min query fick endast upp 86 leverantörer som aktivt bidrog till leveranskedjan. För att förstå gapet mellan det totala antalet och min query som fick upp endast 86 leverantörer så genomförde jag en kontroll av status för samtliga. Denna uppföljning avslöjade att 18 leverantörer inte hade några aktiva transaktioner. Detta kan bero på flera möjliga saker som; att de har upphört sin verksamhet, avslutat sitt samarbete eller aldrig initierat något samarbete. Jag skulle föreslå att företaget undersöker detta mer och eventuellt uppdatera listan för att säkerställa att inköpsorder endast placeras hos aktiva och tillitbara partners.

2.3 Tillverknings Analys

Inom ramen för tillverkning analysen ville jag undersöka efterfrågan och effektiviteten i produktionsprocessen för olika produkter. Detta gjorde jag med hjälp av en query som fick upp de tio mest efterfrågade produkterna och urskilja baserat på antalet arbetsordrar och en bedömning av den genomsnittliga tiden för att genomföra dessa produkter. Resultatet visar att produkter som Front Derailleur, HL Fork och Rear Derailleur dominerar arbetsvolymen, vilket kan peka på en stark marknadsnärvaro. Vissa komponenter såsom Top Tube och Fork Crown produceras i konsekventa volymer, vilket kan betyda att det finns en standarisering inom tillverkningsprocesserna.

Att förstå produktionsdynamiken och planering kan leda till kostnadsbesparingar, marknadsanpassning och kapacitetsplanering. Företaget kan dra nytta av att optimera tillverkningsprocesserna för de mest efterfrågade produkterna. Att fokusera på att göra de vanligaste produkterna mer effektiva innebär ofta att man använder material och arbetskraft på ett smartare sätt, vilket effektiviserar användningen av resurser och minskar kostnader.

2.4 Kund analys

I denna analys indetifierade jag de top tio städer med högsta antalet kunder med hjälp av en query som sammanställde data för att räkna ut antalet kunder per stad. De städer som toppar listan är London, Paris, Burien, Concord, Bellingham, Beaverton, Chula Vista, Berklet, Burlingame och Bellflower. Sedan gjorde jag en separat analys för att bestämma vilka städer som genererade mest försäljning. Denna analys avslöjade att de städer som hade mest försäljning var Toronot, London, Paris, Seattle, Buraby, Richmond, Montreal, Garland, Memphis och Nashua.

Med den här analysen vilje jag se om det finns en korrelation mellan antalet kunder i en stad och den totala försäljningen i samma stad. Denna fråga är central för att förstå hur kundtäthet påverkar försäljningsprestanda och kan ge instiker för strategisk planering och resurs tilldelning. Resultatet från datan avslöjar att endast London och Paris förekommer i både topp tio listorna för antal kunder och total försäljning. Detta tyder på att det inte nödvändigt finns en stark korrelation. För Adventure Works innebär detta att marknadsstrategier inte enbart kan baseras på kundantal utan måste ta hänsyn till andra faktorer som påverkar konsumentbeteenden och köpkraft.

Denna slutsatsen leder till flera frågeställningar: Vad karakteriserar de städer som har hög försäljning men inte nödvändigtvis med stort antal kunder? Vilka faktorer bidrar till att en stad med många kunder inte alltid toppar försäljningslistorna? För att utforska dessa frågor är det viktigt att analysera de socioekonomiska förhållande i varje stad. I städer som Torono, där det säljs mycket, kan ha högre genomsnittlig inkomst per capita, vilket kan leda till att kunderna spenderar mer per köptillfälle. Å andra sidan kan städer som har många kunder men lägre

totalförsäljning ha en befolkning med annorlunda köpbeteende som prioriterar lägre priser eller mindre kvantiteter. Det kan också indikera på att staden har mindre köpkraft.

Om vi tittar på storlek och befolkningsmängd mellan London och Toronto så ser vi att London är betydligt större än Toronto både i term av befolkningsantal och geografisk utsträckning. London har en befolkning på ca 9,5 miljoner med en yta på ca 1570 kvadratkilometer medan Toronto har en befolkning på ca 2,93 miljoner och täcker en yta på ca 630 kvadratkilometer.

Teoretiskt kan London erbjuda en bredare marknad för olika typer av produkter och tjänster eftersom de har en stor befolkning. En stad med en stor befolkning kan antyda en variation i konsumentbehov och köpkraft. En sån mångfald kan leda till att Adventure Works kan öka potentialen för högre försäljning per kund. För att utnyttja den stora kundbasen som finns i London kan Adventure Works anpassa sitt utbud och sina marknadsföringsstrategier för att tillgodose de olika behoven och preferenserna hos stadens invånare. Trots att London har en betydligt högre orderfrekvens än Toronto finns det en möjlighet att anpassa strategier för att nå kunder i London som liknar de i Toronto, det vill säga kunder som är benägna att konsumera dyrare och mer exklusiva produkter.

2.5 Försäljnings analys

I den här analysen fokuserar jag på att identifiera vilka produktkategorier är mest framgångsrika i termer av försäljningen och antalet sålda enheter för varje produktkategori. Resultatet från analysen visar att bland de olika kategorierna utmärker sig cyklar med total försäljning på cirka 4,6 miljoner och 7 525 sålda. Cyklar är kärnprodukten i företagets sortiment och även den mest lönsamma kategorin. Å andra sidan har vi kläder som trots en betydligt lägre försäljningssiffra på 479 781 visar på en högre försäljnings volym med 18 627 sålda. Denna kontrasten mellan de två kategorierna belyser skillnaderna i försäljningsstrategi och kundpreferenser.

Med tanke på att Adventure Works är ett företag inom cykeltillverkning är det naturligtvis att cyklar leder i försäljning. Detta understryker vikten av fortsatta investeringar i denna kategori, vilket kan innebära allt från utveckling av nya modeller till förbättrad marknadsstrategier. När det gäller resultatet på kläder visar den på en stark efterfrågan. Detta är tecken på att även om priset per enhet är lägre per enhet så uppskattas den av en bred kundbas. Detta är en möjlighet att ytterligare utveckla och diversifiera klädsortimentet för att maximera intäkterna från denna kategori. Det kan innebära att införa nya stilar, förbättra kvaliteten eller kanske till och med höja priserna försiktigt för att öka totalförsäljningen.

2.6 Statistisk analys (Konfidensintervall)

I denna delen analyserar jag försäljningsprestanda för cyklar med fokus på att statistiskt validera genomsnittspriset för kategorin cyklar med konfidensgrad 95%. Jag extraherade genomsnittspriset: 1586.74, standaravvikelse: 996.27 och urvalsstorleken: 97 för cyklar sålda mellan 2011 och 2014. Resultat jag fick: Nedre gräns: 1385,94 & Övregräns: 1787,52. Detta betyder att vi kan vara 95% säkra på att det verkliga genomsnittspriset för cyklar ligger inom detta intervall. Konfidensintervallet hjälper oss att förstå osäkerheten i uppskattningar av genomsnittspriset och ger en ram för att fatta informerade beslut om prissättning. Vi kan bättre förstå marknaden och hur de kan anpassas för att möta marknadens förväntningar och behov.

Executive Summary

1. Personal analys: Jag identifierade att produktionsanvändning har flest anställda vilket visar företagets fokus på manuell arbetskraft och produktutveckling. Högre genomsnittsålder inom ingenjörs och underhållsavdelningarna visar på erfaren personal men också en potentiell pensionväg som kräver återgårdar som rekrytering eller kompetensutveckling.

2. Leverantörs analys: Resultaten visar att ledande leverantör i orderkvanitet är Superioror Bicycles vilket kan signalera konkurrenskraftiga priser eller starka relationer. Jag upptäckte även att 18 av 104 leverantörer inte hade några aktiva transaktioner. Detta kräver en granskning och uppdatering av leverantörslistan.

3. Tillverkning analys: De mest efterfrågade produkterna inkluderar Front Derailleur, HL Fork och Rear Derailleur. Denna informationen kan hjälpa optimera tillverkningsdynamiken som kan leda till kostnadsbesparingar och effektivare planering.

4. Kund analys: Kundtäthet och försäljningsprestanda i olika städer visade att endast London och Paris rankas högt i både antal kunder och total försäljning. Detta understryker behovet av skräddarsydda marknadsstrategier.

5. Försäljnings analys: Cyklar är den mest framgångsrika kategorin medan kläder visar en hög försäljningsvolym trots lägre intäker.

6. Statistiskt analys: Fastställa att det genomsnittliga priset för cyklar ligger mellan 1386,94 och 1787.52, vilket ger oss en stabil grund för prissättningsbeslut och en djupare förståelse för marknadstrender.

Muntlig presentation: <https://youtu.be/RrSp-oF0WXk>

3. Reflektion på eget arbete

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

Det har varit riktigt kul att jobba med SQL i det här projektet för det är ett visuellt språk och själva syntaxen är mycket enkelt att greppa. Jag upplevde en utmaningen med logiken i SQL eftersom det är bakvänt än t.ex Python. I Python känns det ofta mer intuitivt att skriva kod sekventiellt, från topp till botten, och bygga upp logiken steg för steg.

Men jag kom till insikt att SQL är ett deklarativt språk. Istället för att detaljerat beskriva varje steg i processen, anger man i princip sitt önskade resultat. Denna förändring i tillvägagångssätt vara ganska omtumlande till en början. Nyckeln med SQL ligger i att förstå dess ordning och hierarki av operationer.

En annan utmaning för mig var att använda rätt JOIN statements för att matcha olika data men det är mer en utmaning med själva datan och min uppfattning, inte språket. Vi hade redan i början av lässåret mjukstartat med mängdlärans teori samt gått in i det djupare i statistik kursen och detta gjorde det enklare att djupdyka i SQL.. till början i alla fall.

2. Vilket betyg du anser att du skall ha och varför.

Betyg är inte det viktigaste för mig så jag väljer att inte svara på denna frågan.

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

Jag känner att mina queries har kanske varit lite för enkla. Jag hade nog tipsat mig själv att utmana mig själv mer och våga använda t.ex CTE eller subqueries för att öva mer på det.