# Teoretiska frågor

1. Beskriv kort hur en relationsdatabas fungerar.

En relationsdatabas är en typ av databas som lagrar och hanterar data på ett strukturerat sätt med hjälp av tabeller. RDBMS (Relational Database Management System) är programvaran som möjliggör skapande, underhåll och sökning av relationsdatabaser. I den här kursen är vårt RDBMS MSSQL.

1. Vad menas med ”CRUD” flödet?

CRUD är en akronym som står för Create, Read, Update, and Delete. Det representerar de grundläggande operationerna som kan utföras på data i en databas eller något beständigt lagringssystem. CRUD-operationerna är grundläggande för databashantering och används ofta i samband med applikationsutveckling och datamanipulation.

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

En ”left join” är en typ av joinoperation i relationsdatabaser där alla rader från den vänstra tabellen och de matchande raderna från den högra tabellen returneras. Den används när du vill hämta alla poster från den vänstra tabellen och de matchande posterna från den högra tabellen.

En ”inner join” är en annan typ av joinoperation som endast returnerar de rader där det finns en matchning i både vänster och höger tabell baserat på det angivna joinvillkoret. Den används när du bara vill hämta poster som har matchande värden i båda tabellerna.

1. Beskriv kort vad indexering i SQL innebär.

Indexering i SQL är en databasoptimeringsteknik som innebär att man skapar en datastruktur som kallas ett "index" på en eller flera kolumner i en databastabell. Ett index är en separat struktur som ger ett snabbare sätt att slå upp rader baserat på värdena i de indexerade kolumnerna. Det primära syftet med indexering är att förbättra hastigheten och effektiviteten för datahämtning, särskilt i SELECT queries.

1. Beskriv kort vad en vy i SQL är.

En vy är en virtuell tabell som är baserad på resultatet av en SELECT query. Till skillnad från en fysisk tabell lagrar en vy inte själva data utan representerar en sparad query som kan refereras till och användas som en tabell. Vyer används för att förenkla komplex queries, abstrahera den underliggande datastrukturen och tillhandahålla ett sätt att kontrollera åtkomsten till specifika kolumner eller rader i en tabell.

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

En lagrad procedur är en förkompilerad samling av en eller flera SQL-satser eller kommandon som är lagrade i databasen och kan köras som en enda enhet. Lagrade procedurer används vanligtvis för att kapsla in en serie SQL-satser, utföra specifika uppgifter eller implementera affärslogik. De erbjuder flera fördelar, inklusive kodåteranvändbarhet, förbättrad prestanda och förbättrad säkerhet.

# 2. Programmeringsuppgift och Rapport

1. Deskriptiv sammanfattning

AdventureWorks2022 är en exempeldatabas skapad av Microsoft för Microsoft SQL Server. Den är utformad som ett fiktivt företags databas och fungerar som ett lärande- och testverktyg för utvecklare, databasadministratörer och individer som är intresserade av att arbeta med Microsoft SQL Server. Det tillåter användare att köra queries och utforska funktioner utan att behöva skapa data manuellt.

* Databasen täcker en mängd olika scheman, inklusive mänskliga resurser, person, produktion, inköp och försäljning. Det finns totalt 18 scheman.
* Databasen innehåller en uppsättning tabeller som representerar olika enheter som anställda, produkter, order och transaktioner. Det finns totalt 72 tabeller.
* Den innehåller olika databasobjekt såsom lagrade procedurer, vyer och funktioner för att demonstrera användningen av dessa element i ett praktiskt scenario.

1. Statistisk analys (konfidensintervall)

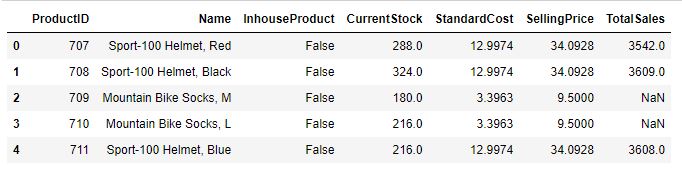


Bild 1

Datauppsättningen hämtad från AdventureWorks2022 kapslar in central information relaterad till produkthantering, som omfattar totalt 293 rader och 7 kolumner. Bland nyckelattributen som tillhandahålls är självkostnadspris och försäljningspris, vilket ger insikter i de ekonomiska aspekterna av varje produkt. Antalet sålda varor ger en heltäckande bild av produkternas marknadsprestanda, vilket hjälper till att förstå konsumenternas efterfrågan och produktens popularitet. Aktuell lagerinformation är avgörande för lagerhanteringen, vilket gör det möjligt för företaget att optimera sina lagernivåer och undvika potentiella brister eller överlagersituationer. Noterbart innehåller datasetet en särskiljande faktor – om en produkt tillverkas på fabriken eller utlagd på entreprenad. Denna distinktion belyser företagets produktionsstrategi och indikerar omfattningen av intern tillverkning kontra extern inköp. Denna datauppsättning fungerar som en värdefull resurs för att genomföra omfattande analyser, allt från kostnads-nyttobedömningar till strategiskt beslutsfattande, vilket främjar en djupare förståelse för dynamiken som styr företagets produktlandskap.

Bild 2

I bild 2 avslöjas avgörande statistiska insikter i varje numerisk kolumn genom implementeringen av metoden describe() i Pandas bibliotek. Denna metod ger en heltäckande översikt, som omfattar väsentliga statistiska mått som räkning, medelvärde (som representerar medelvärdet), standardavvikelse (indikerar spridningen av datapunkter från medelvärdet), minimi- och maximivärden, såväl som olika percentilvärden. Användningen av describe() underlättar en kortfattad och informativ sammanfattning, vilket möjliggör en snabb förståelse av den centrala tendensen, variabiliteten och distributionen av numeriska data inom varje kolumn. Dessa statistiska mått fungerar som grundelement för djupgående datautforskning och analys, och erbjuder en robust grund för informerat beslutsfattande och ytterligare utredningar.

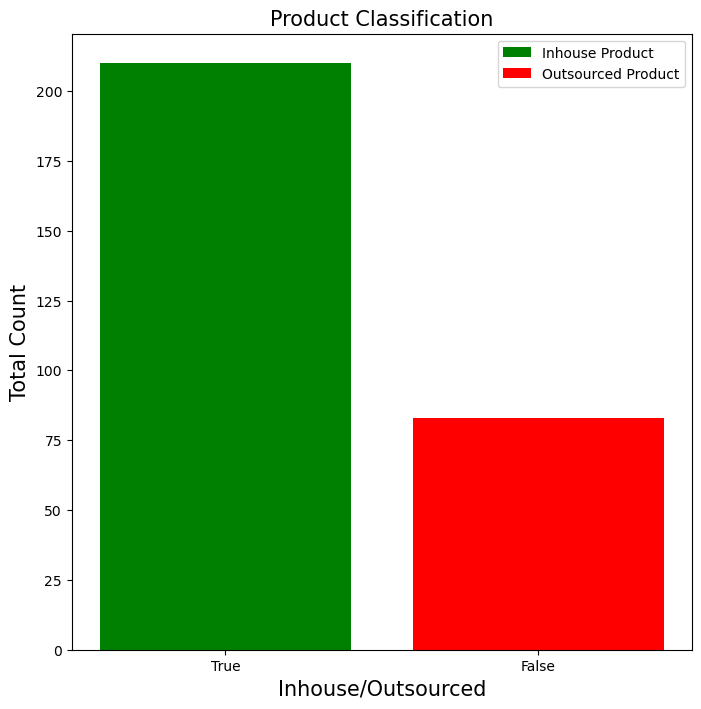


Bild 3

Som tidigare nämnts består datasetet av 293 kolumner, som var och en representerar en distinkt produkt. Efter noggrann analys av kolumnen "InhouseProduct" presenteras resultaten kortfattat i bild 3, och belyser klassificeringen av produkter baserat på deras tillverkningsursprung. Visualiseringen indikerar särskilt att en betydande majoritet, motsvarande 210 produkter, tillverkas i egen regi vid företagets produktionsanläggningar. Däremot identifieras de återstående 83 produkterna som outsourcade, vilket indikerar att företaget förlitar sig på externa källor för produktionen av denna delmängd av sin produktportfölj. Denna klassificeringsinsikt ger en värdefull förståelse för fördelningen mellan egen och outsourcad tillverkning inom datasetet, vilket banar väg för strategiska överväganden och beslutsfattande angående produktionsstrategier och resursallokering.

**Sales Analysis and its Production type**

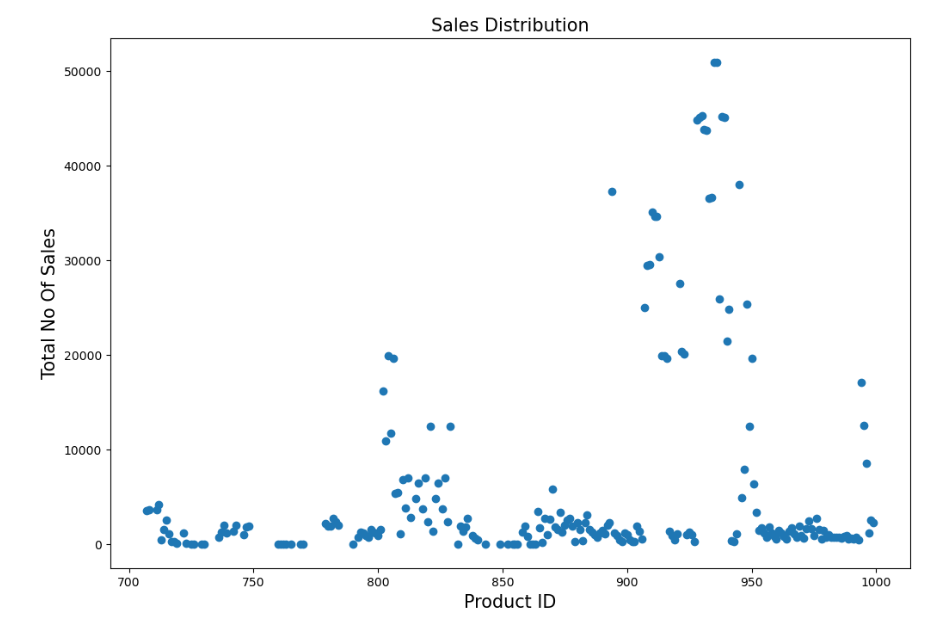


Bild 4

På bild 4 skildras en omfattande försäljningsanalys av hela produktsortimentet. Undersökningen avslöjar en utbredd trend där en betydande del av produkterna registrerar försäljningskvantiteter under 5 000 enheter. Anmärkningsvärda undantag ligger inom produkt-ID:n från 900 till 950, där en delmängd av produkter uppvisar anmärkningsvärt höga försäljningssiffror. Denna anomali föranleder en närmare inspektion, vilket leder till identifieringen av ett distinkt tillstånd som isolerar och lyfter fram 30 specifika produkter. Dessa 30 produkter sticker ut när de överskrider tröskeln på 19 000 sålda kvantiteter, vilket indikerar en undergrupp av högpresterande artiklar inom företagets produktportfölj. Denna nyanserade analys ger värdefulla insikter om försäljningsdistributionsmönster, vilket möjliggör strategiskt beslutsfattande och riktade insatser för att ytterligare förbättra den övergripande försäljningsprestanda.

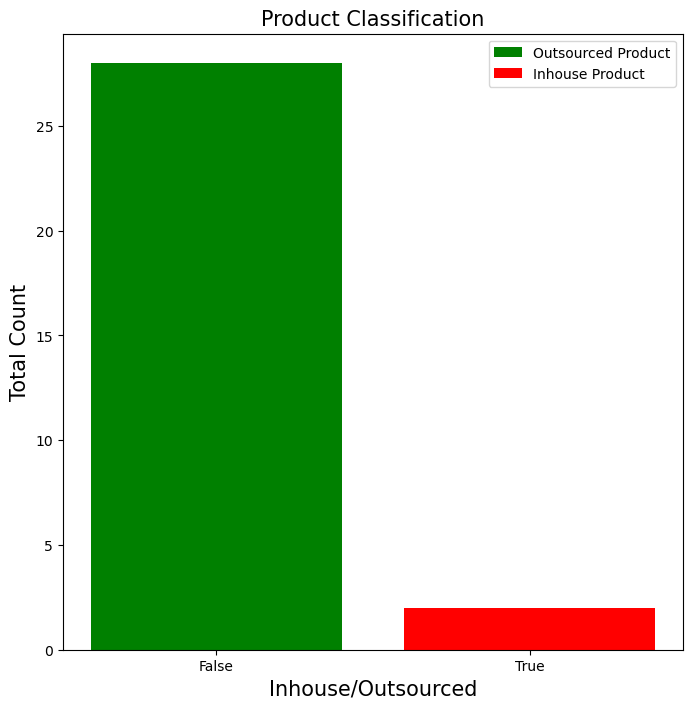


Bild 5

Med utgångspunkt i försäljningsanalysen som presenteras i bild 4 avslöjar en mer detaljerad undersökning i bild 5 ett distinkt mönster inom undergruppen av 30 produkter som överstiger tröskeln på 19 000 sålda kvantiteter. Påfallande nog är det bara 2 av dessa högpresterande produkter som identifieras som egentillverkade vid företagets produktionsanläggningar. Däremot kommer de återstående 28 produkterna inom denna elitkategori från externa leverantörer, vilket visar på ett betydande beroende av outsourcing för produktionen av dessa exceptionellt välsäljande varor. Denna insikt understryker den avgörande skärningspunkten mellan försäljningsprestanda och tillverkningsstrategi, och erbjuder värdefull vägledning för strategiska beslut relaterade till produktion, resursallokering och potentiella möjligheter för att förbättra operativ effektivitet.

**Profit Percentage Analysis and its Confidence Interval**

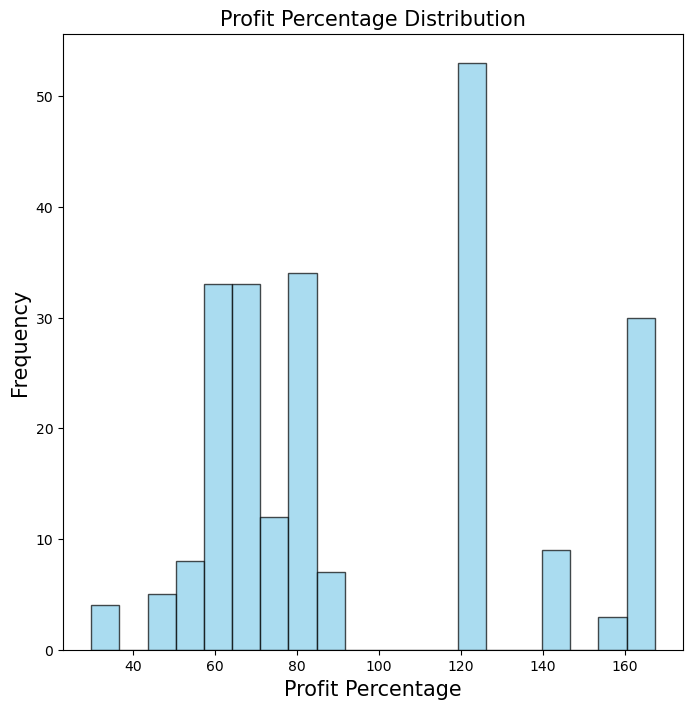


Bild 6

Vinstprocentanalysen, som presenteras visuellt i bild 6, avslöjar distinkta mönster i datasetet. En majoritet av produkterna uppvisar vinstprocent inom intervallet 60 till 100, vilket tyder på en konsekvent prestanda för dessa artiklar. En anmärkningsvärd delmängd av produkter sticker dock ut med vinstprocenter från 120 till 200, vilket indikerar en högre lönsamhetsnivå för denna specifika grupp. Påfallande nog överensstämmer inte fördelningen av vinstprocenter med en typisk normalfördelning, vilket understryker behovet av ett nyanserat förhållningssätt för att förstå den underliggande dynamiken.

Vid närmare granskning blir det uppenbart att konfidensintervallet för hela datasetet verkar ologiskt, vilket föranleder en mer förfinad analys. Datapunkterna i kategorin vinstprocent har strategiskt delats in i två grupper för en mer exakt bedömning. För grupp 1 sträcker sig det beräknade konfidensintervallet från 69,14 till 72,61, vilket belyser det förväntade intervallet av vinstprocentsatser inom denna delmängd. Grupp 2, som representerar produkter med vinstprocent mellan 120 till 200, uppvisar ett konfidensintervall som sträcker sig från 137,11 till 144,75, vilket fångar variationen i lönsamhet för detta distinkta kluster.

Det avgörande är att konfidensintervallet för skillnaden mellan de två grupperna ger djupare insikter. Det här intervallet sträcker sig från 65,79 till 74,29 och beskriver den potentiella skillnaden i vinstprocent mellan de två grupperna. Denna nyanserade analys förfinar inte bara vår förståelse av individuell gruppdynamik utan betonar också vikten av att överväga samspelet mellan distinkta delmängder inom den bredare datamängden för en mer omfattande utvärdering av vinstprocenter.

1. Slutsatser och rekommendationer

Sammanfattningsvis avslöjar försäljningsanalysen en mångsidig produktportfölj med 231 artiklar, där majoriteten upplever måttlig försäljning under 5000 kvantiteter. En distinkt grupp på 30 produkter sticker ut, som visar anmärkningsvärda försäljningskvantiteter som överstiger 19 000. Intressant nog tillverkas endast 2 av dessa högförsäljningsprodukter internt, vilket indikerar ett betydande beroende av outsourcing för majoriteten av de bästa produkterna. Denna outsourcingstrategi väcker oro för vinstmarginalerna, eftersom en betydande del av vinsten från högförsäljningsprodukterna kan gå till externa tillverkningspartners.

**Rekommendationer:** För att optimera lönsamheten och få större kontroll över leveranskedjan bör företaget överväga att omvärdera sin tillverkningsstrategi. En strategisk förändring mot intern produktion för fler av högförsäljningsprodukterna kan leda till kostnadsbesparingar, förbättrade vinstmarginaler och ökad flexibilitet när det gäller att svara på marknadens krav. Dessutom kommer att genomföra en detaljerad kostnads-nyttoanalys för varje produkt i kategorin högförsäljning att ge värdefulla insikter om möjligheten att ta in fler tillverkningsprocesser internt. Detta tillvägagångssätt förbättrar inte bara företagets vinstbehållning utan positionerar det också för större smidighet och hållbarhet i det dynamiska marknadslandskapet.

1. Executive summary

Försäljningsanalysen målar upp en levande bild av vårt produktlandskap och visar upp en rik mångfald av 231 artiklar. Medan majoriteten upplever en stadig försäljning, kräver den enastående prestandan för 30 produkter, över 19 000 kvantiteter, vår uppmärksamhet. Noterbart är att endast två av dessa topppresterande företag tillverkas internt, vilket väcker kritiska frågor om vår nuvarande outsourcingstrategi och dess inverkan på vinstmarginalerna.

Mot bakgrund av detta rekommenderar jag en strategisk omvärdering av vår tillverkningsmetod. Att gå över till intern produktion för varor med hög försäljning kan låsa upp kostnadseffektivitet, höja vinstmarginalerna och förbättra vår anpassningsförmåga till marknadsdynamiken. En grundlig kostnads-nyttoanalys för varje produkt i denna elitkategori kommer att erbjuda värdefulla insikter, som guidar oss mot en mer hållbar och smidig framtid.

Låt oss ta denna möjlighet att optimera vår lönsamhet och stärka vår position på marknaden.

1. Datum för muntlig presentation

02 Jan 2024

# Reflektion på eget arbete

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

Jag hade utmaningar med att beräkna konfidensintervall för icke-normalfördelning men med hjälp av tidigare kurs och använde google för att övervinna det.

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

Jag tror att jag borde klarat mycket bra eftersom jag tror att jag uppfyllde alla krav för kunskapskontrollen

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

Tipsen jag vill ge mig själv är att arbeta mer med dataanalys för att lära mig lite mer kunskap inom programmering.