

## Executive summary

- 1) Denna rapport svarar på teoretiska aspekter av i relation till databas och SQL.
- 2) Fokus ligger sedan på att analysera lönsamheten för kategorin "Kläder", förutsatt att alla produkter säljs till listade priser.
- 3) Utmaningar uppstår när det gäller att länka valutadata, vilket skapar tvivel om resultatens riktighet.
- 4) Trots en betydande vinstskillnad mellan "Kläder" och andra kategorier, tyder snabba produktionstider för kläder på behovet av att överväga effektivitet.
- 5) Rapporten avslutas med en uppmaning till förbättring av databasen, inkluderande produktionstiddata, och betonar vikten av att överväga försäljning och produktionsplats för korrekta ekonomiska insikter.

## Teoretiska Frågor

1. Beskriv kort hur en relationsdatabas fungerar.

En relationsdatabas är en databas där ett samspel mellan rå data och dess relationer. Exempelvis har man en tabell som listar kolumner där flera eller en utgör endast en "primary key". Medan "foreign key" utgör också av kolumner men är hänvisat till en annan relations / tabells primary key. Översiktligt, kan dessa tabeller ha one-to-one, one-to-many, och many-to-many relationer, exempelvis en författare tabell har one-to-many relation med en bok tabell (och relationen kan förmodligen inte vara null pga definition av en författare).

2. Vad menas med "CRUD" flödet?

Det är operationerna som är kända i SQL-världen

-Create: Skapa ny data.

-Read: Läs och hämta befintlig data.

-Update: Uppdatera befintlig data.

-Delete: Ta bort data.

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

Inner join returnerar endast rows som matchar, medan left join gör samma men också returnerar rader där vänstra tabellen inte matchade, och det står då null på den högra delen.

4. Beskriv kort vad indexering i SQL innebär.

Syftet med Indexering är att öka hastigheten på sökning inom en databas. Man kan skapa indexering med en eller flera kolumner och därmed snabbt lokalisera de rader som matchar. Det finns någon sorts algoritm bakom det, men man behöver inte veta om det för att använda.

5. Beskriv kort vad en vy i SQL är.

En vy är en virtuell tabell baserat på en query, alltså den lagras inte i databasen. Man kan använda en vy på samma sätt som en vanlig tabell, men datan som visas i vyn är fortfarande resultatet av den ursprungliga query.

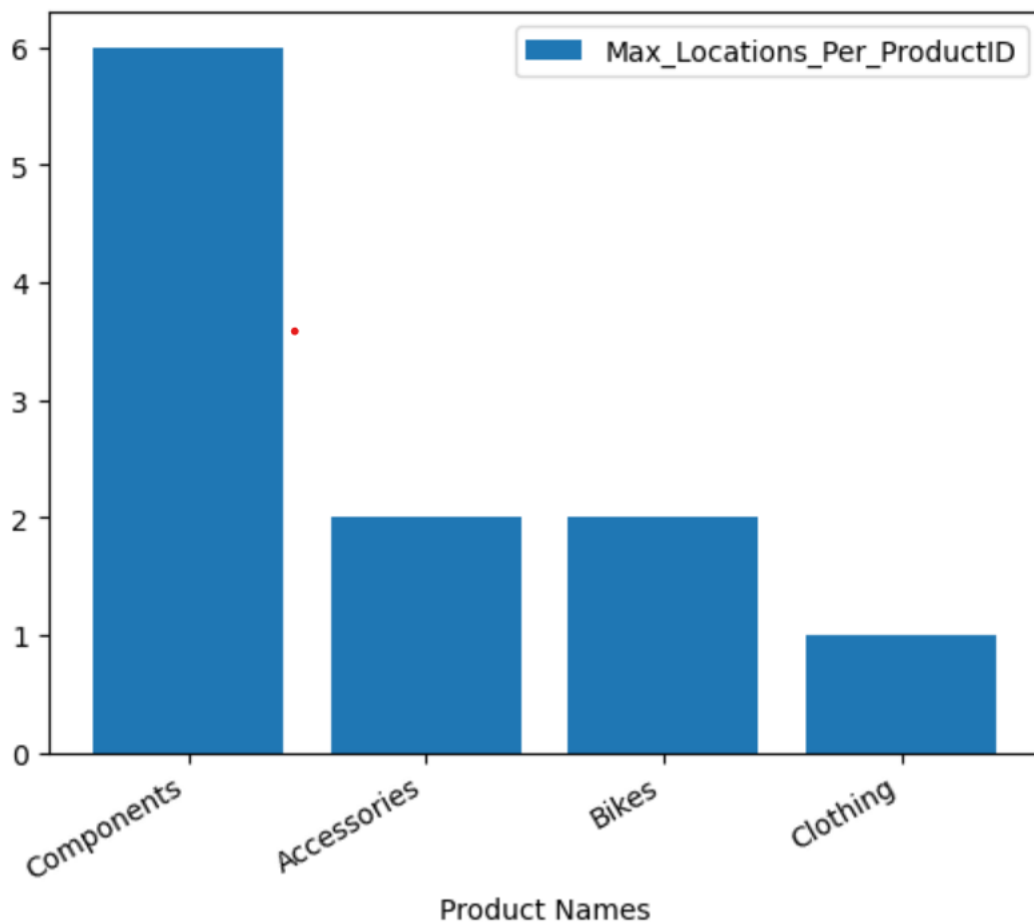
6. Beskriv kort vad en lagrad procedur i SQL är.

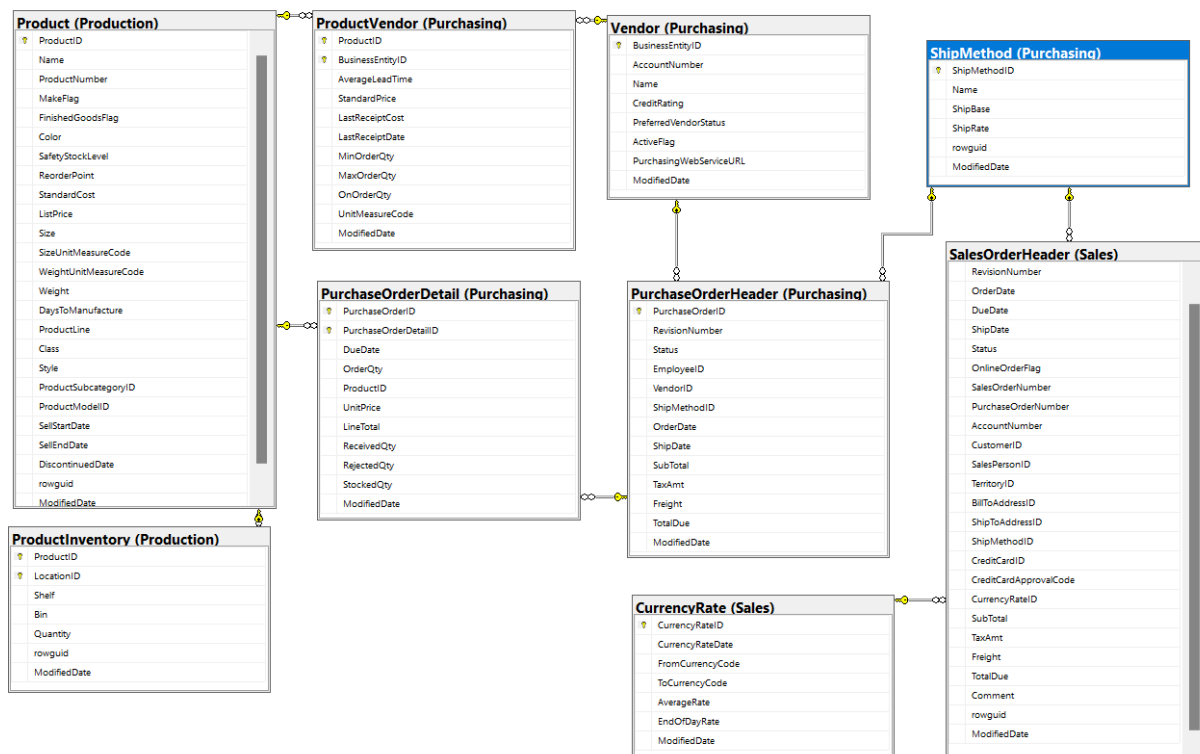
En lagrad procedur är rader av SQL-instruktioner som är lagrad i databasen och kan återanvändas.

## Motivering

I min kod försöker jag analysera hur värda produktkategorierna är, men jag förutsätter att alla skapade produkter säljs med respektive pris (enligt [ListPrice] kolumnen).

I koden ser man att det finns kolumnen [Locations\_Per\_ProductID] extraherad från [Production.Location] vilket kan tyda på att valutan, men själva tabellen är kopplad till [Production.Product] via many-to-many. Medan [Production.Product] har en lång väg innan den kopplas till [Sales.Currency] som visas i bilderna nedan. På grund av osäkerheten kommer rapporten om endast produktkategorin kläder vara säker när det gäller kostnad och försäljningspris. Jag kommer ändå anta att det osäkerheten inte finns med de andra produktkategorierna.





Ett försök att nå Sales.Currency för varje Production.Product.ProductID på två sätt gjordes, men antalet stämde inte båda (se bilden nedan). Antalet ProductID är mycket fler än 23 st.

	ProductID	Product_Name	FromCurrencyCode_Count	ToCurrencyCode_Count
0	907	Components	1	4
1	912	Components	1	4
2	913	Components	1	4
3	914	Components	1	4
4	915	Components	1	4
5	916	Components	1	4
6	922	Accessories	1	5
7	923	Accessories	1	5
8	928	Accessories	1	4
9	929	Accessories	1	4

Real size of table above: 23

## Skillnad mellan vinst i endast "Clothing" och i alla kategorier

I bilden nedan ser man att det finns en signifikant skillnad då konfidensintervallet inte innehåller 0.

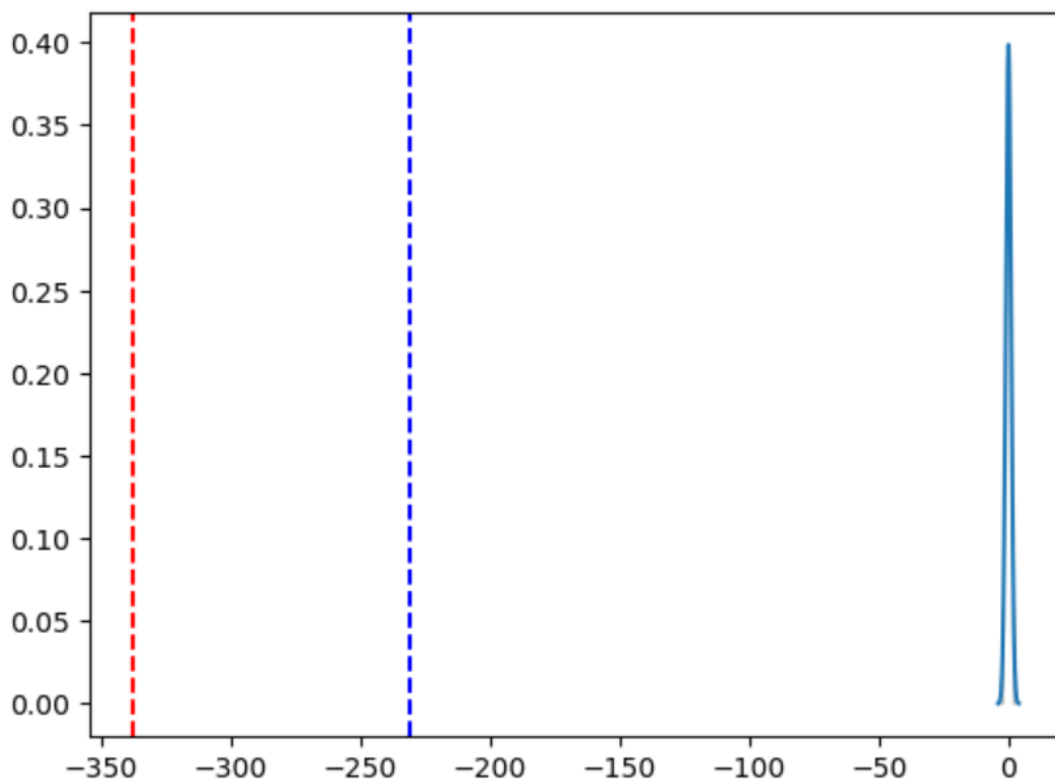
Group 1 är endast kläder och Group 2 är allt oavsett kategori. Intervallet är  $-337 < \mu < -231$  och är 95% konfidensintervall för differensen av medelvärdet mellan grupp 1 och grupp 2.

```

Mean of Group 1: [26.18826]
Mean of Group 2: [310.25800673]
STD of Group 1: [16.4134699]
STD of Group 2: [402.13042273]
Sample size of Group 1: 35
Sample size of Group 2: 223
95% Confidence Interval for the Difference: [-337.38032658], [-230.75916688]

```

```
<matplotlib.lines.Line2D at 0x1ad0ff53670>
```



## Kategorier med hänsyn till antal dagar varje för produktion

Kläder har en stor skillnad i vinst gentemot alla andra kategorier, men kläder produceras snabbt! (se bilden nedan). Komponenter tar 1–2 dagar att produceras och även cyklar tar 4 dagar att produceras. Tid att produceras måste tas hänsyn för att optimalt använda den kapacitet som fabriker har. Det vill säga, tidsmässigt kanske skillnaden i vinst inte är signifikant. Men eftersom produktionen inte visar antal timmar i stället för antal dagar, så är det omöjligt att estimeras konfidensintervall mellan hur mycket vinst per dag för kläder och vinst per dag för alla kategorier.

	Total_ProductIDs	ProductName	DaysToManufacture
0	25200	Accessories	0
1	29148	Clothing	0
2	48346	Components	1
3	5960	Components	2
4	83962	Bikes	4

Real size of table above: 5\*

## Slutsats

Medelvärde respektive stickprovets standardavvikelse av vinst på alla produkter är 310 resp. 402 enheter och medelvärdet för kläder är 26 resp. 16 enheter. Skillnaden är statistisk signifikant, men beroende hur lång tid ett klädesplagg tar att producera, så kan det vara mer lönsamt att producera kläder. Databasen hade behövt förbättra kolumnen "DaysToManufacture" till "HoursToManufacture" för att kunna räkna ut en bättre helbild av vinsten. En annan miss är att rapporten inte har tagit hänsyn till om produkterna har sålts eller inte. Det medför en felaktig bild när bara försäljningspriset räknas som omsättning per produkt. En annan osäkerhet till rapporten är att de sålda varorna var i olika platser, vilket kan tyda på annan valuta om några av dessa platser var mellan länder. Däremot var kläder bara på en plats och medelvärdet samt stickprovets standardavvikelse har med säkerhet korrekta proportioner, kontra mot om valutan skulle vara olika under uträkningarna.

## Slut

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

Jag har haft svårt att hitta syntaxfel vid körningsproblem, främst med kommatecken. En annan utmaning är att jag vill koppla olika tabeller som är långt ifrån varandra, men det är lätt när man väl förstår hur man ska göra. Den första stora utmaningen som fick mig att vara att det inte fanns riktiga uppdrag på saker man skulle extrahera från databasen, utan man fick i stället hitta på egna uppdrag. Det som var svårt att attackera från något håll, men jag valde till slut produktion-delen och blev sedan intresserad av att man kunde räkna ut vinst där.

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

Ett G åtminstone.

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

Bara kör på och välj något. Det spelar inte roll vad du väljer i slutändan.