



**CHALMERS**

# Utvinning av sambandsregler som beslutsunderlag mot utvecklad artikelplacering

Examensarbete inom högskoleingenjörsprogrammet

Ekonomi och produktionsteknik

Hampus Aderhjelm

Julia Lassagård

**INSTITUTIONEN FÖR TEKNIKENS EKONOMI OCH ORGANISATION  
AVDELNING FÖR SUPPLY AND OPERATIONS MANAGEMENT**

---

CHALMERS TEKNISKA HÖGSKOLA  
Göteborg, Sverige 2022  
[www.chalmers.se](http://www.chalmers.se)



Rapportnummer E2022:026

# Utvinning av sambandsregler som beslutsunderlag mot utvecklade artikelplacering

HAMPUS ADERHJELM  
JULIA LASSAGÅRD

TEKNIKENS EKONOMI OCH ORGANISATION  
Avdelning för Supply and Operations Management  
CHALMERS TEKNISKA HÖGSKOLA  
Göteborg, Sverige 2022

Utvinning av sambandsregler som beslutsunderlag mot utvecklad artikelplacering

HAMPUS ADERHJELM  
JULIA LASSAGÅRD

© HAMPUS ADERHJELM, 2022  
© JULIA LASSAGÅRD, 2022

Rapportnummer E2022:026  
Teknikens ekonomi och organisation  
Chalmers tekniska högskola  
412 96 Göteborg  
Sverige  
Telefon + 46 (0)31-772 1000

Göteborg, Sverige 2022

## **Förord**

Examensarbetet har utförts under våren 2022 på programmet Ekonomi och produktionsteknik på Chalmers Tekniska Högskola. Det har varit en utmanande och givande resa inom området datautvinning. Först vill vi tacka vår handledare på grossistverksamheten för att ha hjälpt oss få tillgång till den information som behövts för att genomföra arbetet.

Vi vill vidare tacka Ola Hultkrantz, lektor vid avdelningen Supply and Operations Management, för att ha granskat slutresultatet och visat intresse för vårt examensarbete. Vi vill också tacka våra opponenter Frida Johansson och Johanna Tobiasson för utmärkt återkoppling som hjälpt oss att öka tydligheten och kvaliteten på arbetet.

Slutligen vill vi tacka vår handledare Hafez Shurrab, doktor vid avdelningen Supply and Operations Management, för hans hjälp och utmärkta vägledning under hela arbetet.

Hampus Aderhjelm och Julia Lassagård, Göteborg, 2022

Göteborg, Sverige 2022

Utvinning av sambandsregler som beslutsunderlag mot utvecklad artikelplacering

HAMPUS ADERHJELM  
JULIA LASSAGÅRD

Institutionen för Teknikens ekonomi och organisation  
Chalmers tekniska högskola

## SAMMANFATTNING

För många företag med en lager- och distributionsverksamhet är den största kostnaden orderplockning, som är processen att paketera varor för utleverans till kund. För att effektivisera orderplockningsprocessen har artikelplacering visat sig ha en betydlig påverkan. Ett sätt att utveckla artikelplacering är att anpassa den efter samband mellan artiklar. En teknik för att hitta samband mellan artiklar är utvinning av sambandsregler. Trots den snabba utvecklingen och en ökad uppmärksamhet för datautvinning inom logistikbranschen, ger befintlig litteratur begränsade insikter om vilka förutsättningar som krävs och hur utvinning av sambandsregler hos en lagerverksamhet fungerar. Syftet med studien är att föreslå hur en lagerverksamhet kan utvinna sambandsregler som ska användas som beslutsunderlag för att utveckla artikelplacering, samt identifiera förutsättningar för att utvinna användbara sambandsregler.

Arbetet utförs på ett fallföretag som är en grossistverksamhet med lager- och logistikverksamhet. Datainsamling gjordes i form av litteraturstudie, intervjuer, observationer och dokumentstudie hos fallföretaget. Den insamlade data analyserades och resultatet visade flera förutsättningar som krävdes för att utvinna sambandsregler på en lagerverksamhet. Dessa är behovet av tillgänglig kundorderdata, verktyg som kan hantera datamängden, ren och relevant data, mänsklig interaktion med datamängden, mänsklig analys av resultatet och tillgängliga resurser. Vidare presenterade studien ett förslag på en modell i Microsoft Excel för att utvinna sambandsregler. Resultaten från modellen redovisar visuellt de samband som identifieras från en grossistverksamhets kundorderdata.

På grund av begränsningar av studien anses den inte kunna generera generella svar för alla lagerverksamheter. För att uppnå en högre generaliserbarhet hade flera fallföretag behövts undersökas och jämföras.

Nyckelord: Utvinning av sambandsregler, datautvinning, artikelplacering

Göteborg, Sverige 2022

Utvinning av sambandsregler som beslutsunderlag mot utvecklade artikelpacering

HAMPUS ADERHJELM  
JULIA LASSAGÅRD

Institutionen för Teknikens ekonomi och organisation  
Chalmers tekniska högskola

## ABSTRACT

For many companies with a warehousing and distribution business, the biggest cost is order picking, which is the process of packaging goods for delivery to the customer. To streamline the order picking process, article placement has proven to have a significant impact. One way to develop article placement is to customize it according to the associations between articles. One technique for finding associations between articles is association-rule mining. Despite the rapid development and increased attention for the data mining industry, existing literature provides limited insights into what prerequisites are required and how the mining of association rules in a warehouse could be performed. The purpose of the study is to propose how a warehousing company can extract association rules that are to be used as a decision basis for developing article placement, and to identify prerequisites for extracting useful association rules.

The work is carried out at a company which is a wholesale business with warehousing and logistics operations. Data collection is done in the form of literature studies, interviews, observations, and document studies at the single case company. The data collected was analyzed, and the results showed several prerequisites required to mine association rules on a warehouse business. The prerequisites found were *available sales order data, tool that can handle the amount of data, clean and relevant data, human interaction with data, human analysis of the result and available resources*. Furthermore, a proposal for a model is presented in Microsoft Excel to extract association rules. The result from the model visually presents the associations identified from a wholesale business sales order data.

Due to limitations of the study, the results are not considered to be general for all warehousing operations. To achieve a higher generalizability, several case companies must be investigated and compared.

Key words: association-rule mining, data mining, article placement





# INNEHÅLLSFÖRTECKNING

<b>1. Inledning</b>	<b>1</b>
1.1 Bakgrund	1
1.2 Syfte	2
1.3 Frågeställningar	2
<b>2. Teoretisk referensram</b>	<b>4</b>
2.1 Lager och dess funktioner	4
2.1.2 Fast och flytande artikelplacering	4
2.2 Mängder och sannolikhetslära	5
2.3 Datautvinning	5
2.4 Tekniker inom datautvinning	6
2.5 Utvinning av sambandsregler	8
2.5.1 Dataformat för orderdata	8
2.5.2 A-priori algoritmen	9
2.5.3 Stöd och tillförlitlighet	9
2.5.4 Process för att generera sambandsregler	12
2.6 Arbetsmetodik vid datautvinning	12
2.7 Förutsättningar för effektiv datautvinning	15
2.8 Visualisering av data	16
<b>3. Metod</b>	<b>18</b>
3.1 Datainsamling	18
3.1.1 Litteraturstudie	19
3.1.2 Intervjuer	19
3.1.3 Observation	20
3.1.4 Dokumentstudie	20
3.2 Dataanalys	21
<b>4. Resultat</b>	<b>23</b>
4.1 Förstå verksamheten	23
4.1.1 Orderplockningsprocess	23
4.1.2 Artikelplacering	24
4.1.3 Definiera målen för datautvinningen	25
4.2 Förstå datamängden	26
4.2.1 Insamling av initial datamängden	26

4.2.2 Begränsningar av insamling av datamängden	27
4.3 Förbereda datamängden	28
4.4 Skapa en modell	29
4.4.1 Dela upp modellsamlingen	29
4.4.2 Skapande av modellen i Microsoft Excel	30
4.4.2 Visualisering av sambandsreglerna	30
4.5 Utvärdera modellen och dess resultat	32
4.5.1 Resultatet kontra företagsmål	32
4.5.2 Användbara sambandsregler	32
<b>5. Diskussion</b>	<b>33</b>
5.1 Svar på första frågeställningen	33
5.2 Svar på andra frågeställningen	36
5.3 Metoddiskussion	37
5.4 Diskussion utifrån hållbarhet och etik	38
<b>6. Slutsats</b>	<b>40</b>
<b>Referenser</b>	<b>41</b>
<b>Bilagor</b>	<b>43</b>
Bilaga 1: Intervjufrågor	43
Bilaga 2: Makro 1 kodat i Visual Basic Editor	44
Bilaga 3: Makro 2 kodat i Visual Basic Editor	47

# 1. Inledning

I detta kapitel ges en bakgrund till varför datautvinning är intressant för en lagerverksamhet, följt av gapet i teorin som det här arbetet försöker uppfylla. Vidare definieras syftet och frågeställningar för arbetet.

## 1.1 Bakgrund

När en lagerverksamhet utformas och styrs är syftet att minimera lagerhållningskostnaderna och hanteringskostnaderna (Jonsson & Mattsson, 2016). För många företag med en lager- och distributionsverksamhet är den största kostnaden orderplockning, som är processen att paketera varor för utleverans till kund. Enligt Chiang et al. (2011) har det i tidigare studier undersökts flera sätt att minska kostnaden i orderplockningsprocessen. Metoder som undersökts handlar dels om påverkan av plocksekvensen, dels om lagerlayout. Den metod som visat sig ha en betydlig påverkan på effektiviteten av orderplockningsprocessen är placeringen av artiklar.

Berry och Linoff (2011) skriver att den snabba utvecklingen inom informationsteknik gör det möjligt för företag att snabbt samla in och lagra kunddata, orderdata samt produktdata i ett informationssystem. Jukic och Nestorov (2005) menar att många företag och organisationer inte utnyttjar den informationen som de har tillgång till från deras datalager fullt ut. Chiang et al. (2011) påstår att informationen som hämtas genom datautvinning kan användas för att förbättra effektiviteten genom att skapa regler för artikelplacering. De menar att hitta samband mellan artiklar som ofta köps tillsammans, och sedan anpassa artikelplaceringen utefter dessa samband, har visat sig vara en effektiv metod för att effektivisera orderplockningsprocessen. Ett samband kan uttryckas som; en kundorder som innehåller mjölk innehåller troligtvis bröd. För att hitta dessa samband anser Jukic och Nestorov (2005) att en av de mest framgångsrika teknikerna är utvinning av sambandsregler (association-rule mining).

Trots den snabba utvecklingen och en ökad uppmärksamhet för datautvinning inom logistikbranchen, ger befintlig litteratur begränsade insikter om vilka förutsättningar som krävs för att utvinna sambandsregler hos en lagerverksamhet. Det är också intressant att undersöka hur sambandsregler kan utvinnas för att en lagerverksamhet ska kunna använda sig av sambandsregler som beslutsunderlag för utvecklad artikelplacering.

## 1.2 Syfte

Syftet med studien är att föreslå hur en lagerverksamhet kan utvinna sambandsregler som ska användas som beslutsunderlag för att utveckla artikelplacering, samt identifiera förutsättningar för att utvinna användbara sambandsregler.

## 1.3 Frågeställningar

För att minska de vanligaste felen i en datautvinningsprocess har ett antal organisationer utvecklat en standardprocess för datautvinning (Jackson, 2002). Ett viktigt steg i standardprocessen är att identifiera förutsättningarna för datautvinningen innan analysen påbörjas. Det finns därför ett behov att identifiera förutsättningarna för utvinning av sambandsregler som ska användas som beslutsunderlag för artikelplacering hos en lagerverksamhet. Detta leder till den första frågeställningen:

1. Vilka är förutsättningarna som krävs för att utvinna användbara sambandsregler för artikelplacering?

Ett svar på frågeställningen kommer ge en ökad förståelse om de förutsättningarna som krävs för att utvinna sambandsregler som ska användas för artikelplacering. Den ökade förståelsen gör det sedan möjligt att diskutera på vilket sätt förutsättningarna kommer underlätta utvinningen av sambandsregler, samt hur sambandsreglerna blir användbara för artikelplaceringen.

Enligt Jukic och Nestorov (2005) måste vanligtvis företag använda sig av ett separat verktyg för datautvinning. Dels för att data först måste extraheras från datalagret, dels transformeras för att fungera i ett verktyg för datautvinning. Därför finns det ett behov att undersöka hur en lagerverksamhet kan utforma ett verktyg för att ta reda på sambandsregler som sedan kan hjälpa dem vid beslut om artikelplacering. Detta leder till den andra frågeställningen:

2. Hur kan sambandsregler utvinnas för att stödja beslutsfattare mot en utvecklad artikelplacering?

Ett svar på frågeställningen gör det möjligt att undersöka och föreslå hur en lagerverksamhet skulle kunna utvinna samband mellan artiklar från befintliga kundorder. Frågeställningen ger möjlighet att presentera hur ett verktyg skulle kunna utformas enligt de identifierade

förutsättningarna från den första frågeställningen. Verktuget skulle i sin tur kunna användas för att förmedla information om samband till beslutsfattare för artikelplaceringen.

## **2. Teoretisk referensram**

I detta kapitel redogörs det för hur en lagerverksamhet fungerar. Vidare beskrivs definitioner, tekniker och förutsättningar för datautvinning. Vid utvinning av sambandsregler används statistiska formler och därav beskrivs relevanta begrepp inom sannolikhetslära. Till sist beskrivs CRISP-DM som är en metod för datautvinning.

### **2.1 Lager och dess funktioner**

Ett lager är en byggnad som hanterar gods för lagerhållning, sortering och förberedelse av vidare transport (Lumsden et al., 2019). Anledningen till att lager är något som företag behöver är för att uppnå transporter med högre fyllnadsgrad, möta en fluktuerande efterfrågan samt överkomma avstånd och tidsskillnader mellan producenter och kunder (Koster et al, 2006).

De huvudsakliga aktiviteterna i ett lager är att ta emot gods, placera gods på lager, plock av order, sortering och transport (Koster et al, 2006). Att ta emot gods innebär att avlasta gods från inkommande transportbärare och inspektera om det förekommer några felaktigheter. Placera gods på lager innebär att förflytta godset till en specifik lagringsplats för lagerhållning. Orderplockning, som är den mest resurskrävande aktiviteten, är att plocka ihop rätt antal artiklar som ska vidare med transport. Efter orderplockning kan varorna behöva paketeras och sorteras till rätt transport.

#### **2.1.2 Fast och flytande artikelplacering**

Det finns två typer av placeringsprinciper för artiklar på ett lager; fast och flytande placering (Lumsden et al., 2019). Om en artikel på lagret en förutbestämd lagringsplats används fast placering. Genom att använda flytande placering kan antalet pallplatser utnyttjas bättre och på så sätt minimera utrymmesbehovet (Lumsden et al., 2019). Vid flytande placering kan artikeln placeras på valfri ledig plats. Det är vanligt att verksamheter använder lagersystem för att hitta den lediga platsen som är optimal för lagring av den inkommande artikeln. Genom att tilldela artiklar fördelaktiga platser på lagret kan plocksträckan minimeras (Chiang et al., 2011). Det finns flera metoder för att tilldela artiklar platser, där metoderna består av en uppsättning regler som kan användas för att tilldela produkter till lagringsplats.

## 2.2 Mängder och sannolikhetslära

En mängd är en samling av element vilket kan vara exempelvis namn, bokstäver eller tal. En mängd  $A$  som innehåller elementen 1 och 2 kan skrivas som  $A = \{1, 2\}$ . Det finns flera olika mängdoperationer att tillämpa på mängder (Dahlblom, 2003).

**Tabell 2.1:** Beskrivning av mängdoperationerna union och snitt.

Mängdoperation	Beskrivning
Union	<p>Unionen av två mängder <math>A</math> och <math>B</math> är en mängd som innehåller alla element som finns i <math>A</math>, <math>B</math> eller i båda. Unionen av mängderna <math>A</math> och <math>B</math> skrivs som <math>A \cup B</math>.</p> <p><i>Exempel:</i> Om <math>A = \{1, 2\}</math> och <math>B = \{2, 3\}</math> så blir <math>A \cup B = \{1, 2, 3\}</math>, då 1 finns i <math>A</math>, 2 finns i både <math>A</math> och <math>B</math>, 3 finns i <math>B</math>.</p>
Snitt	<p>Snittet av två mängder <math>A</math> och <math>B</math> är en mängd som innehåller alla element som finns i både <math>A</math> och <math>B</math>. Snittet av mängderna <math>A</math> och <math>B</math> skrivs som <math>A \cap B</math>.</p> <p><i>Exempel:</i> Om <math>A = \{1, 2\}</math> och <math>B = \{2, 3\}</math> så blir <math>A \cap B = \{2\}</math>, då 2 finns i både <math>A</math> och <math>B</math>.</p>

Det går även att tillämpa sannolikhet på mängder. Sannolikhet är hur troligt det är att en specifik händelse inträffar. Sannolikhet av en händelse  $H$ , brukar betecknas enligt  $P(H)$ . Det kan också vara intressant att beräkna hur troligt det är att en händelse  $J$  inträffar, förutsatt att en händelse  $H$  redan har inträffat. Detta skrivs som  $P(J|H)$  och kallas *betingad sannolikhet* (Dahlblom, 2003). Den betingade sannolikheten kan också beräknas enligt  $P(J|H) = \frac{P(H \cap J)}{P(H)}$ .

## 2.3 Datautvinning

För många verksamheter genereras mängder med data från bland annat transaktioner, försäljningar, produktbeskrivningar och kunder (Han et al., 2011). Kunddata kan exempelvis vara kundens adress, mail och telefonnummer. För att omvandla den mängden data till

värdefull information finns det ett behov av att använda sig av kraftfulla tekniker och verktyg. Det är från det behovet som Han et al. (2011) menar att termen datautvinning har formats.

Enligt D. Larose och C. Larose (2015) är datautvinning en process för att hitta användbara mönster och trender i stora datamängder. Berry och Linoff (2011) definierar datautvinning som en verksamhetsprocess för att analysera stora mängder data för att hitta användbara mönster och regler. Båda författarna påpekar att datautvinning inte är en teknik utan en samling verktyg för att extrahera information ur data.

Datautvinning kan tillämpas i flera olika branschområden som till exempel sjukvården, marknadsföring och logistikbranschen (Berry & Linoff, 2011). Datautvinning kan användas överallt där data samlas in och lagras. Informationen som utvinns kan användas för att ge en ökad förståelse för företags dagliga verksamhet och kunder. Det kan hjälpa företag att identifiera försäljningstrender, utveckla marknadsföringsstrategier och ligga till grund för olika verksamhetsbeslut.

## **2.4 Tekniker inom datautvinning**

Det finns ett flertal olika tekniker inom datautvinning. Teknikerna består av olika algoritmer och metoder för att analysera datamängder. De fem vanligaste beskrivs i tabell 2.1. Det som är gemensamt för teknikerna är att de ska användas till att bygga en modell som löser konkreta problem i verkliga livet (Gorunescu, 2011).



**Tabell 2.2:** Beskrivning av några vanliga tekniker inom datautvinning

---

Namn på teknik	Beskrivning
Klassificering	Klassificering är en teknik inom datautvinning som innebär att klassificera dataobjekt baserat på egenskaper till en eller flera fördefinierade klasser (Gorunescu, 2011). Ett exempel på användning av klassificering är att identifiera mail som anses vara skräppost.
Klustering	Klustering är en teknik för att dela upp data i kluster med förutbestämda likheter (Gorunescu, 2011). Ett exempel på när klustering är användbart är inom sjukvården för att placera sjukdomar i grupper efter liknande symptom och behandlingar.
Anomalidektering	Anomalidektering är en teknik för att identifiera händelser eller observationer som avviker avsevärt från definitionen av normalt beteende hos den stora datamängden (Gorunescu, 2011). Ett exempel är att upptäcka om ett kort är stulet genom att identifiera om köpesumman är större än vanligt, eller om köp genomförs på andra orter än var kunden brukar genomföra sina köp.
Regression	Regression är en teknik för datautvinning som används för att estimeras ett värde för en datauppsättning (Gorunescu, 2011). Tekniken kan användas för att estimeras vindhastigheter och riktningar som en funktion av temperatur, luftfuktighet och lufttryck.
Utvinning av sambandsregler	Tekniken innebär att hitta samband mellan artiklar i beställningar som ofta köps tillsammans i en kundorder (Jukić & Nestorov, 2005). Ett exempel på samband är en order som innehåller mjölk, innehåller troligtvis också smör.

---

## 2.5 Utvinning av sambandsregler

En sambandsregel mellan artiklar som ofta köps tillsammans på en order kan uttryckas som;  $X \Rightarrow Y$ , det vill säga att en order som innehåller X innehåller troligtvis Y, där X och Y representerar olika artiklar som köps. Enligt Gorunescu (2011) kan beslutsfattare baserat på sambanden förändra artikelplaceringen, lagerlayouten, marknadsföringsstrategier och identifiera olika kundsegment. D. Larose och C. Larose, (2015) trycker på att X och Y måste vara unika artiklar för att ge intressanta samband. Det gör att ointressanta samband som till exempel “en order som innehåller mjölk innehåller troligtvis mjölk” utesluts från de intressanta sambanden. Anledningen till att de är ointressanta är att de alltid kommer ge ett 100 procentigt samband. I följande avsnitt kommer det redogöras för de dataformat som används för att formatera orderdata. Sedan beskrivs a-priori algoritmen som används för att hitta de mest frekventa sambanden samt värderingsfaktorerna stöd och tillförlitlighet. Till sist beskrivs en process i tre steg för att generera sambandsregler.

### 2.5.1 Dataformat för orderdata

När orderdata insamlas finns det två vanliga dataformat som är fördelaktiga att formatera datamängden till (D. Larose & C. Larose, 2015). Den första är *transaktions dataformat* (*transactional data format*) som består av två kolumner; ID och innehåll. Varje rad av data består av en ensam artikel. D. Larose & C. Larose (2015) lyfter att det inte är relevant i analysen att veta hur stora antal av varje enskild artikel som är beställd. Det som är av intresse är huruvida en artikel finns på ordern eller inte. Exempelvis kan orderdata i figur 2.1a, presenteras i transaktions dataformat som figur 2.1b.

Det andra dataformatet är *tabell dataformat* (*tabular data format*) som innebär att varje rad representerar en order och vilka artiklar som finns på ordern (D. Larose & C. Larose, 2015). Den första kolumnen representerar order ID, och resterande kolumner representerar unika artiklar. Innehåller ordern en specifik artikel visas “1” i kolumnen för artikeln, om inte visas “0”. Orderdata från figur 2.1a, kan presenteras i tabell dataformat som figur 2.1c.

a)	Order ID	Innehåll
	1	A, B, C
	2	A, C
	3	B, C, D, E

b)	Order ID	Artikel
	1	A
	1	B
	1	C
	2	A
	2	C
	3	B
	3	C
	3	D
	3	E

c)	Order ID	A	B	C	D	E
	1	1	1	1	0	0
	2	1	0	1	0	0
	3	0	1	1	1	1

Figur 2.1: a) Orderdata b) Transaktionellt dataformat av orderdata c) Tabell dataformat av orderdata.

### 2.5.2 A-priori algoritmen

Antalet samband som kan genereras från en datamängd blir snabbt många (D. Larose & C. Larose, 2015). En analys av ordrar med tillsammans 2000 unika artiklar genererar  $\frac{2000!}{(2000-2)!} \approx 3,9$  miljoner möjliga samband med två artiklar. För att minimera analysen till en mer hanterbar storlek används a-priori algoritmen. D. Larose och C. Larose (2015) konstaterar att om en mängd  $Z$  inte är frekvent, kommer  $Z \cup X$  inte vara frekvent, oavsett vilken artikel  $X$  är. Algoritmen framför alltså att om en mängd med endast en artikel inte är frekvent, kommer inte mängden bli mer frekvent om fler artiklar adderas till den. Detta gäller även om de adderade artiklarna är högfrekventa. En frekvent mängd är en mängd som har frekvensen  $\geq \phi$ .

Låt exempelvis  $\phi = 2$ , så att en frekvent mängd är en mängd som uppkommer två eller fler gånger, alltså i minst  $\frac{2}{3} = 67\%$  av orderarna i figur 2.1a som analyseras. Baserat på orderarna i figur 2.1a, som har mängden  $F_1 = \{A, B, C, D, E\}$ , kommer den nya mängden baserat på att  $\phi = 2$  vara  $F_2 = \{A, B, C\}$ , då artikel D och E inte uppkommer två eller fler gånger. Definieras då mängden  $Z$  till de icke frekventa artiklarna,  $Z = \{D, E\}$ , bevisar a-priori algoritmen att  $Z \cup X$  inte kommer vara frekvent, även om artikel  $X$  är frekvent.

### 2.5.3 Stöd och tillförlitlighet

Sambanden som utvinns värderas genom att beräkna *stöd* (*support*) och *tillförlitlighet* (*confidence*). Stöd mäts i hur många order av det totala antalet order som sambandet förekommer, vilket tolkas som hur viktigt ett samband är (Jukić & Nestorov, 2005). Stödet för en sambandsregel  $X \Rightarrow Y$  beräknas enligt formeln:

$$\text{Stöd} = P(X \cap Y) = \frac{\text{Antalet order som innehåller både artikel X och Y}}{\text{Totala antalet order}}$$

Baserat på orderarna i figur 2.1a, kan exempelvis stödet för sambandet  $A \Rightarrow C$  beräknas som  $\frac{2}{3} = 67\%$ , då två av totalt tre orderar innehåller både A och C.

Tillförlitlighet mäter andelen order som innehåller både X och Y av alla order som innehåller X. Detta tolkas som styrkan av sambandet (Jukić & Nestorov, 2005). Tillförlitlighet för en sambandsregel  $X \Rightarrow Y$  beräknas enligt formeln:

$$\text{Tillförlitlighet} = P(Y|X) = \frac{P(X \cap Y)}{P(X)} = \frac{\text{Antalet order som innehåller både artikel X och Y}}{\text{Antalet order som innehåller X}}$$

Tillförlitligheten för sambandet  $A \Rightarrow C$  beräknas enligt  $\frac{2}{2} = 100\%$ , då det är två order som innehåller både A och C, och totalt två order som innehåller A. Tabell 2.2 presenterar stöd och tillförlitlighet för resterande samband från orderarna i figur 2.1a.

Vid datautvinning kan analytiker föredra samband med högt stöd eller hög tillförlitlighet, och mer vanligt att båda faktorerna ska vara höga (D. Larose & C. Larose, 2015). För att samband ska värderas starkt ska både stöd och tillförlitlighet vara inom de bestämda gränser som analysen baseras på. För att utvinna starka samband behöver dock inte alltid stöd och tillförlitlighet sättas högt. Till exempel kan en analytiker som försöker upptäcka bedrägeri sätta stöd på en procent eller mindre på grund av att relativt få transaktioner är bedrägeri. Jukic och Nestorov (2005) menar också att resultatet av utvinning av sambandsregler kan antingen ge för många sambandsregler eller för få sambandsregler. I båda fallen måste användaren justera stöd och tillförlitlighet och köra utvinningsprocessen igen.

**Tabell 2.3:** Beräknad stöd och tillförlitlighet baserat på order från figur 2.1a.

<b>Samband, <math>X \rightarrow Y</math></b>	<b>Stöd</b>	<b>Tillförlitlighet</b>
A $\rightarrow$ B	1/3 = 33%	1/2 = 50%
A $\rightarrow$ C	2/3 = 67%	2/2 = 100%
A $\rightarrow$ D	0/3 = 0%	0/2 = 0%
A $\rightarrow$ E	0/3 = 0%	0/2 = 0%
B $\rightarrow$ A	1/3 = 33%	1/2 = 50%
B $\rightarrow$ C	2/3 = 67%	2/2 = 100%
B $\rightarrow$ D	1/3 = 33%	1/2 = 50%
B $\rightarrow$ E	1/3 = 33%	1/2 = 50%
C $\rightarrow$ A	2/3 = 67%	2/3 = 67%
C $\rightarrow$ B	2/3 = 67%	2/3 = 67%
C $\rightarrow$ D	1/3 = 33%	1/3 = 33%
C $\rightarrow$ E	1/3 = 33%	1/3 = 33%
D $\rightarrow$ A	0/3 = 0%	0/1 = 0%
D $\rightarrow$ B	1/3 = 33%	1/1 = 100%
D $\rightarrow$ C	1/3 = 33%	1/1 = 100%
D $\rightarrow$ E	1/3 = 33%	1/1 = 100%
E $\rightarrow$ A	0/3 = 0%	0/1 = 0%
E $\rightarrow$ B	1/3 = 33%	1/1 = 100%
E $\rightarrow$ C	1/3 = 33%	1/1 = 100%
E $\rightarrow$ D	1/3 = 33%	1/1 = 100%

## 2.5.4 Process för att generera sambandsregler

D. Larose och C. Larose (2015) beskriver att processen för att generera sambandsreglerna delas upp i tre steg. Först används a-priori algoritmen för att identifiera de artiklar som har frekvensen  $\geq \phi$ . Frekvensen sätts då till den minsta gränsen av stöd som man vill uppnå för sambanden. Sedan tas alla möjliga kombinationer mellan artiklarna i den frekventa mängden fram. Det sista steget är att beräkna stöd och tillförlitlighet för de identifierade sambanden.

Baserat på orderna från figur 2.1a, var den frekventa mängden  $F_2 = \{A, B, C\}$  då  $\phi = 2$ . De möjliga sambanden identifieras då enligt figur 2.2a. Frekvensen  $\phi = 2$  ger dock en gräns på att stödet av sambanden minst ska vara  $\frac{2}{3} = 67\%$ . Sambandet  $A \Rightarrow B$  och  $B \Rightarrow A$  ska då sorteras bort från figur 2.2a, då de inte uppfyller kravet för stöd. Om den minimala tillförlitligheten är satt till 80% kommer de slutgiltiga reglerna att presenteras enligt figur 2.2b.

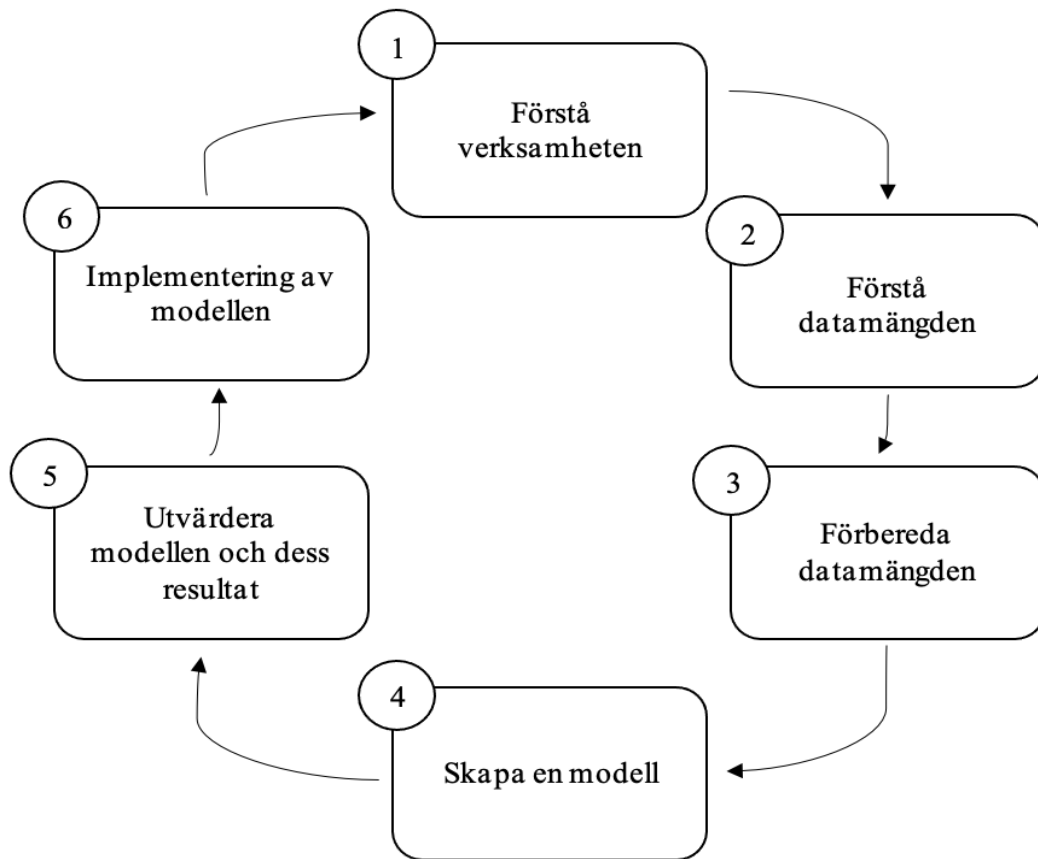
a) Samband, $X \rightarrow Y$	Stöd	Tillförlitlighet
$A \rightarrow B$	$1/3 = 33\%$	$1/2 = 50\%$
$A \rightarrow C$	$2/3 = 67\%$	$2/2 = 100\%$
$B \rightarrow A$	$1/3 = 33\%$	$1/2 = 50\%$
$B \rightarrow C$	$2/3 = 67\%$	$2/2 = 100\%$
$C \rightarrow A$	$2/3 = 67\%$	$2/3 = 67\%$
$C \rightarrow B$	$2/3 = 67\%$	$2/3 = 67\%$

b) Samband, $X \rightarrow Y$	Stöd	Tillförlitlighet
$A \rightarrow C$	$2/3 = 67\%$	$2/2 = 100\%$
$B \rightarrow C$	$2/3 = 67\%$	$2/2 = 100\%$

Figur 2.2: a) Identifierade samband från mängden  $F_2$  då  $\phi = 2$ . b) Identifierade samband från mängden  $F_2$  då  $\phi = 2$ , stöd  $\geq 67\%$  och tillförlitlighet  $\geq 80\%$

## 2.6 Arbetsmetodik vid datautvinning

En metod för datautvinning som finansierades av Europeiska unionen och utvecklades av ett antal organisationer är Cross-Industry standardprocess för datautvinning, även förkortat CRISP-DM (Jackson, 2002). Målet med metoden var att utveckla en metod som kan appliceras i många industriområden och som kan göra datautvinningsprojekt snabbare, billigare, pålitligare och mer hanterbart. Som framgår av figur 2.3 består metoden av sex faser:



Figur 2.3: Metodologi för datautvinning

### 1. Förstå verksamheten

Den första fasen är att förstå verksamheten och förstå vad verksamheten vill uppnå (Jackson, 2002). Verksamhetens problem ska översättas till ett datautvinningsproblem. Detta innebär att ett bredare verksamhetsproblem ska brytas ner i ett eller fler mindre specifika mål som är lätta att mäta. I fasen ska också begränsningar, resurser och antaganden identifieras.

### 2. Förstå datamängden

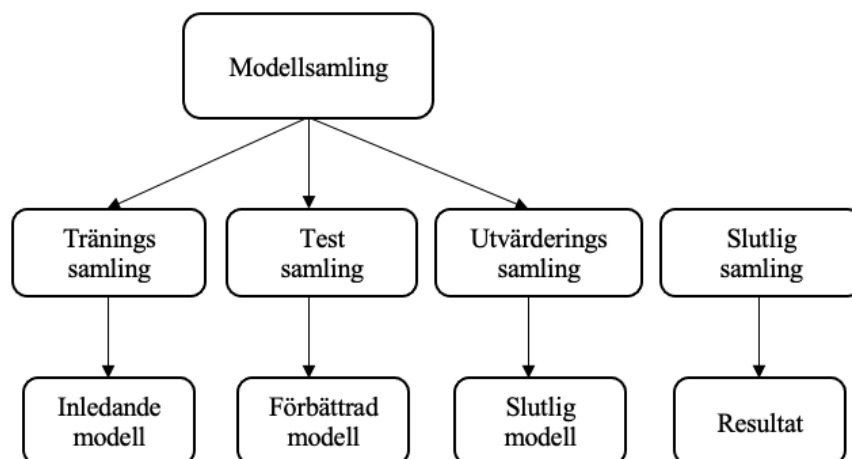
Den andra fasen startar med en initial datainsamling (Jackson, 2002). I detta stadié är det bara en idé och tanke om vilken data som faktiskt kommer att behövas. Det är viktigt att få en uppfattning om hur datamängden beter sig genom att studera den. Det bör också reflekteras över hur mycket data som är tillräcklig och hur mycket historiska data som kan behövas.

### 3. Förbereda datamängden

Den tredje fasen är att förbereda en modellsamling som innehåller all data som används i modelleringsprocessen (Jackson, 2002). För att skapa en modellsamling krävs det att data från samtliga källor är sammanställda. Samlingen består av en eller flera tabeller med en rad per objekt som ska studeras, och kolumner för allt som är känt och relevant om objektet. Relevant data ska väljas ut för analysen och rensas för att lösa problem. Efter att data är insamlad och större problem är lösta, måste datamängden förberedas för att analyseras. Detta innebär exempelvis att addera fält för att göra informationen synlig, integrera data, ta bort extremvärden, gruppera klasser och formatera data för att passa modelleringsverktyget.

### 4. Skapa en modell

Den fjärde fasen är att skapa en modell. Olika datautvinningstekniker, till exempel utvinning av sambandsregler, tillämpas i denna fas. Enligt Jackson (2002) ska modellsamlingen delas in i tre komponenter; tränings-, test- och utvärderingssamling, se figur 2.4. Träningsamlingen används för att bygga den inledande modellen. Testsamlingen används för att förbättra modellen och säkerställa så att den fungerar på data som inte behandlats tidigare. Utvärderingssamlingen skiljer sig helt från tränings- och testsamlingen eftersom den används för att utvärdera noggrannheten på modellen. Den slutliga samlingen, som är helt separat från de tre andra samlingarna, används för att flytta modellen till från modelleringsmiljön till den verkliga miljön.



Figur 2.4: Indelning av modellsamlingen



### *5. Utvärdera modellen och dess resultat*

Den femte fasen är att utvärdera modellen och dess resultat. Detta görs för att säkerställa att modellen uppfyller de målen som bestämdes i början av projektet (Jackson, 2002). En utvärdering bör svara på frågor som hur exakt modellen är, hur väl modellen beskriver den observerade data, hur mycket tillit kan placeras i modellen och hur begriplig är modellen.

### *6. Implementering av modellen*

Den sjätte fasen innebär att implementera resultatet i verksamheten Jackson (2002) skriver att skapandet av modellen är i inte slutet på projektet. Även om syftet med modellen är att öka kunskapen om datamängden, kommer den kunskap som erhålls att behöva organiseras. I det här steget är bevakning och underhåll viktigt om resultaten av datautvinningen blir en del av den dagliga verksamheten. Syftet är att undvika att felaktiga data används under en längre period.

## **2.7 Förutsättningar för effektiv datautvinning**

En uppfattning inom datautvinning är att det sker automatiskt, och kräver ingen eller lite mänsklig översikt (D. Larose och C. Larose, 2015). Det finns också en uppfattning om att datautvinning kommer lösa problemen automatiskt. I verkligheten finns det inga automatiska datautvinningsverktyg som löser problemen. Vidare är datautvinning en process som kräver kontinuerlig kvalitetsövervakning och andra utvärderande processer.

Enligt D. Larose och C. Larose (2015) finns det en uppfattning om att datautvinningsverktyg är enkla att använda, men i verkligheten kräver datautvinningsverktyg stor mänsklig interaktion. Det finns också en uppfattning om att datautvinning kommer automatiskt att organisera en rörig databas. I verkligheten är en stor fas av datautvinningsprocessen att förbereda data. Det krävs att data behöver omvandlas till rätt format för att passa verktyget.

Många menar att datautvinning snabbt blir lönsamt (D. Larose och C. Larose, 2015). I verkligheten kan resultatet av en datautvinningsprocess variera. Faktorer som uppstartskostnader, kostnader för analyser och förberedelser inför datautvinningen har en stor

påverkan om resultatet är lönsamt. Det finns också en uppfattning om att ett resultat av datautvinning alltid är bra. I verkligheten finns det ingen garanti för att det löser det definierade verksamhetsproblemet. Dock om datautvinning används korrekt av människor som förstår modellen, och dess syfte kan den ligga till grund för ett lönsamt resultat.

D. Larose och C. Larose (2015) menar att de som arbetar med datautvinning aldrig bör lita på resultat från datamodeller utan att ha förståelse för hur modellen fungerar och systemet bakom modellen. För att nå användbara och lönsamma resultat måste den mänskliga expertisen användas. Att göra dåliga analyser utan att förstå det underliggande faktorerna för resultatet menar D. Larose och C. Larose (2015) kan kosta flera miljoner kronor.

Vid skapandet av olika modeller menar Gorunescu (2011) att det är nödvändigt att utföra en preliminär undersökning av datamängden som ska användas för att få en ökad förståelse av egenskaperna hos datamängden. Han menar att en sådan undersökning dels använder människans förmåga att känna igen mönster i data från tidigare erfarenheter, som inte alltid kan betonas med klassiska undersökningsmetoder. Dels menar Gorunescu (2011) även att all den erfarenheten som skapas under den preliminära undersökningen av datamängden, avsevärt kan hjälpa till att effektivt kunna välja den optimala datautvinningsmetoden för det analyserade datamängden.

Vid praktiska tillämpningar av utvinning av sambandsregler menar Gorunescu (2011) att det lätt blir komplext. Han påstår detta då utvinningen av sambandsregler kräver hundra- eller helst tusentals transaktioner innan de kan anses vara statistiskt signifikanta. Vidare menar han att datasamlingar ofta innehåller miljontals transaktioner och att då hitta alla möjliga samband är en svår uppgift.

## **2.8 Visualisering av data**

För en effektiv datautvinning menar Keim (2002) att det är viktigt att inkludera människan. Visualisering av data är att presentera data i någon visuell form så att människan kan tolka och dra slutsatser. Enligt Mohd Ali et al. (2016) kämpar många företag med att hantera stora mängder data och att analysera, tolka och presentera data på ett meningsfullt sätt. Från en datamängd kan mönster och korrelationer identifieras, men resultatet är ofta bara råa siffror. Mohd Ali et al. (2016) menar att datavisualisering är ett enkelt och snabbt sätt att presentera komplexa datamängder för att förmedla ett budskap. Detta för att råa siffror kan vara svåra att

tolka och att om siffrorna i stället representeras visuellt blir det lättare att hitta meningsfulla mönster och fatta beslut därefter.

Det är vanligt att visualisering av data följer tre steg; överblick, zooma och filtrera och sen detaljer. Keim (2002) menar att användaren först måste få en överblick av datamängden. I överblicken identifierar användaren intressanta mönster eller korrelationer och fokuserar på någon eller några av dem. För att sedan analysera och kunna dra slutsatser måste användaren kunna få tillgång till detaljer om den specifika data. Mohd Ali et al. (2016) skriver att det finns flera verktyg för att visualisera data som Tableau, Microsoft Excel, Microsoft Power BI, Plotly och Gephi. Det viktigaste är att verktygen är interaktiva. Användaren ska tillåtas att interagera med datamängden.

### **3. Metod**

Syftet med studien är att öka förståelsen för utvinning av sambandsregler, vilket kräver en fördjupning av ämnet och därav har metoddesignen för arbetet valts till en enfallsstudie. En enfallsstudie öppnar upp för en djupgående analys och resultaten blir empiriskt rika, dock innebär det att resultatet inte blir lika generaliserat som vid en flerfallsstudie. Flyvbjerg (2006) menar dock att man ofta kan generalisera från en enfallsstudie och att studien kan vara central för den vetenskapliga utvecklingen som komplement till andra metoder.

Fallföretaget är en grossistverksamhet med en lager- och logistikverksamhet som lagerhåller, förpackar och levererar artiklar till olika kunder inom hotell-, storkök- och restaurangbranschen. Företaget har flera lager runt om i Sverige och arbetet utförs på ett av dessa lager. Fallföretagets önskan är att driva verksamheten med höga mål gällande resurseffektivitet och hållbarhet. För deras orderplockning innebär det en process med en hög grad värdeskapande tid. Vidare är fallföretagets ledord ansvarsfull och effektiv hantering. Fallföretaget har önskat att vara anonyma i studien och kommer i rapporten benämnas som grossistverksamheten.

Yin (2017) menar att ett skäl för att välja ett fallföretag är att det representerar ett vanligt fall. Syftet med att undersöka ett sådant fallföretag är att det fångar omständigheterna och förutsättningarna för en vardaglig situation. Lärdomarna från ett vanligt fall antas vara informativa för den generella branschen. Grossistverksamheten valdes för att det ansågs representera en typisk lagerverksamhet i logistikbranschen. Resultatet från arbetet ansågs därför kunna appliceras på andra vanliga lagerverksamheter i logistikbranschen.

#### **3.1 Datainsamling**

Datainsamling vid en fallstudie kan komma från flera olika insamlingsmetoder såsom intervjuer, dokumentstudier och observationer. Yin (2017) menar att vid datainsamling för en fallstudie ska data insamlas från flera datakällor för att öka validiteten och reliabiliteten av fallstudiedata. Detta kapitel redogör för litteraturstudien, intervjuerna, observationerna, och dokumentstudierna.

### 3.1.1 Litteraturstudie

En litteraturstudie genomfördes för att identifiera vad som redan fanns inom ämnesområdet och vilka koncept och teorier som är relevanta. För studien genomfördes litteraturstudien i syftet att få en djupare förståelse för lagerverksamhetens funktion, artikelplacering, databearbetning och datautvinning. Ett annat syfte var att genom litteraturstudien förstå vilka förutsättningar som är identifierade för att utvinna samband och hur en modell kan skapas för utvinning av sambandsregler.

Chalmers bibliotek och Google Scholar användes för att söka fram relevanta artiklar och annan litteratur inom ämnesområdet. Sökord som användes var delvis *family grouping*, *product placement warehouse*, *frequently bought together*, *market basket analysis*, *association analysis*, *data visualization*, *order picking warehouse*, *warehouse management* och *data mining-based storage assignment*.

För att försäkra sig om att informationen som hittades i litteraturstudien kunde anses pålitlig används flera metoder. En metod som användes var att söka efter artiklar och tidskrifter i respekterade utgivare. En annan var att söka efter fackgranskad litteratur då den informationen redan har granskats och erkänts av andra forskare. För att ytterligare försäkra sig om pålitligheten utvärderades också litteraturen utifrån antalet citationer. Litteraturstudien uppdaterades och analyserades kontinuerligt under examensarbetet för att bli mer relevant i takt med att arbetet fortskred då ny information och nya insikter hittades.

### 3.1.2 Intervjuer

För att erhålla kunskaper, uppfattningar, åsikter och tyckande i ett forskningssammanhang kan intervjuer användas (Ejvegård, 2012). Intervjuer genomfördes med en logistikutvecklare och med en planerings- och utvecklingschef på grossistverksamheten. Intervjuerna genomfördes i syfte att få kunskaper om grossistverksamheten och hur de arbetar med data som beslutsunderlag för artikelplacering i nuläget. Det genomfördes även intervjuer i syfte att samla feedback från företaget som sedan användes som grund för hur en modell till datautvinning kan utformas och justeras. Anledningen till att intervjuer genomfördes med logistikutvecklare var för att det är dem som planerar och förbättrar artikelplaceringen. Intervjuer genomfördes med planerings- och utvecklingschefen för att förstå hur grossistverksamheten vill utveckla sin

lagerverksamhet. Planerings- och utvecklingschefen arbetar med att ta ut datarapporter ur grossistverksamhetens datasystem så intervjuer genomfördes för att få en förståelse för hur filer med data kunde hämtas och struktureras.

Intervjuerna genomfördes digitalt. Strukturen på intervjuerna var relativt öppna, det vill säga få färdiga frågor utformades för att tillåta en mer fri konversation. Frågorna (bilaga 1) som utformades skickades i förväg genom mail för att ge tid för respondenterna att förbereda sig med syfte att få mer utvecklade svar. Enligt Bryman (2012) är extern reliabilitet hur väl en studie kan upprepas. Risken med ostrukturerade intervjuer är att samtalet kan övergå till andra diskussioner än vad som var tänkt från början vilket kan påverka den externa reliabiliteten för studien. Kontinuerlig kommunikation med företaget gjordes för att kunna ställa frågor som uppkom under arbetets gång, samt stämma av eventuella antaganden och oklarheter som uppstod.

Bryman (2012) skriver att intern reliabilitet är hur väl tolkningarna av data stämmer överens mellan medlemmarna i en forskningsgrupp. Vid intervjuerna var båda gruppmedlemmarna närvarande och förde anteckningarna. Direkt efter intervjutillfällena diskuterades och sammanfattades resultatet för att undvika skilda tolkningar. Oklarheter och eventuella antaganden kompletterades genom mail. En av personerna som intervjuades fick också möjlighet att bekräfta tolkningen för att ytterligare förbättra den interna reliabiliteten av den empiriska datainsamlingen och dataanalysen.

### **3.1.3 Observation**

För att samla in information och förstå hur grossistverksamheten arbetar med data genomfördes observationer på grossistverksamhetens affärssystem. Enligt Flick (2009) hjälper observation av system till att förstå de underliggande strukturerna för ett fenomen. Observationerna gav en mer visuell förståelse för grossistverksamhetens affärssystem och hur data kan hämtas och sorteras.

### **3.1.4 Dokumentstudie**

Dokumentstudier genomfördes i form av granskning av interna dokument från grossistverksamheten. Enligt Yin (2017) är fördelen med dokumentstudier att dokument inte förändras vilket gav möjligheter att undersöka samma dokument flera gånger. Dokumentation tillför också exakta detaljer om en händelse. Dokument som ansågs relevanta bestod av

kundorder från grossistverksamhetens dataredovisningssystem och två artikelplaceringsverktyg i Microsoft Excel. Syftet med dokumentstudier var att få en djupare kunskap och förståelse för hur grossistverksamheten arbetar med affärssystem och hur deras information hanteras, samt att få en bild av vilken information och data som grossistverksamheten kunde tillhandahålla till datautvinningen.

### **3.2 Dataanalys**

För att säkerställa reliabiliteten och validiteten på dataanalysen följde den specifika procedurer. Den data som erhöles från intervjuerna och observationerna samt insikterna från dokumentstudierna sammanställdes. När skilda tolkningar, oklarheter eller eventuella antaganden uppstod i dataanalysen om den empiriska data som samlats från grossistverksamheten klargjordes det med en av personerna från intervjuerna. En av personerna från fallföretaget fick också möjlighet att granska den tolkning för att ytterligare förbättra den interna reliabiliteten av den empiriska datainsamlingen och dataanalysen.

Den empiriska data som erhöles från datainsamlingen låg till grund för genomförandet av datautvinningsprojektet. Arbetet fokuserade på datautvinning och metoden CRISP-DM följdes för att på ett strukturerat sätt genomföra ett datautvinningsprojekt. Metoden valdes för att den har utvecklats att fungera som ett neutralt verktyg i olika branscher och minska de vanligaste felen vid ett datautvinningsprojekt. Tabell 3.1 presenterar de mer specifika tillvägagångssätt som genomfördes i varje steg vid CRISP-DM metoden. Från metoden kommer insikterna att ligga till grund för vilka förutsättningar som krävs för att hitta användbara sambandsregler. För att identifiera de förutsättningar som krävs för att utvinna användbara sambandsregler, jämfördes insikterna med den teoretiska referensramen.

**Tabell 3.1:** Beskrivning av varje steg som genomfördes baserat på metodiken CRISP-DM.

Steg	Beskrivning
1. Förstå verksamheten	I steget var det viktigt att försäkra sig om att det definierade målet för projektet skulle bidra med värdeskapande information till en lagerverksamhet. För att säkerställa detta genomfördes intervjuer med logistikutvecklare samt planerings- och utvecklingschef på en grossistverksamhet.
2. Förstå datamängden	För att utforska och verifiera kvaliteten visualiserades grossistverksamhetens kundorderdata. Visualiseringsprogrammet som användes i detta projekt var Microsoft Excel. För att fastställa kvaliteten på den inhämtade datamängden utvärderades den utifrån om det saknas data, om attributen var relevanta för att uppnå målet och om storleken på datamängden var tillräcklig.
3. Förbereda datamängden	För att förbereda datamängden och förbättra kvaliteten användes Excel som verktyg för att eliminera data som ansågs överflödiga och transformera data till ett lämpligt format som ska användas för att utveckla modellen.
4. Skapa en modell	Projektet fortsatte i Excel för att skapa en modell som genererar sambandsregler. Processen bestod av att utvärdera olika funktioner i Excel, implementera vald funktionstyp och utvärdera vald funktionstyp. Modellen baserades på processen för utvinning av sambandsregler som beskrevs i kapitel 2.3.4.
5. Utvärdera modellen och dess resultat	Modellen utvärderades genom att testa på nya datamängder. Detta gjordes genom att samtliga rapporter i utvärderingssamlingen kördes genom modellen. Detta säkerställer att modellen fungerade för olika datauppsättningar. Steget genomfördes också för att utvärdera att modellen uppfyller dem målen och resultat som grossistverksamheten "önskade".
6. Implementering av modellen	Implementeringen av modellen består av att övervaka modellen i en verklig miljö under en längre tid, men detta arbete utslöt fasen då det ligger utanför arbetets omfattning och tidsram.



## **4. Resultat**

Det här kapitlet är uppdelat i fem avsnitt som illustrerar de resultat och insikter som erhållits från olika faserna av CRISP-DM-metoden.

### **4.1 Förstå verksamheten**

För att kunna definiera målen för datautvinningen var en förståelse för det valda fallföretaget nödvändig. Nedan beskrivs grossistverksamhetens orderplockningsprocess samt nuvarande artikelplacering. Därefter definieras målen för datautvinningen.

#### **4.1.1 Orderplockningsprocess**

Grossistverksamheten använder sig av en beställningssida där kunden registrerar sin order. Kunden lägger själv in vilka artiklar de önskar att beställa och vilken dag de vill ha sina varor levererade. Kunden har möjlighet att ändra eller komplettera sin order fram till en viss tidpunkt. Grossistverksamheten säger att det är vanligt att kunden placerar en ny order med samma leveransdag, i stället för att ändra en redan befintlig beställning. Om en kund har lagt flera order med samma leveransdatum, slås dem vid orderplockningen ihop till en plockorder.

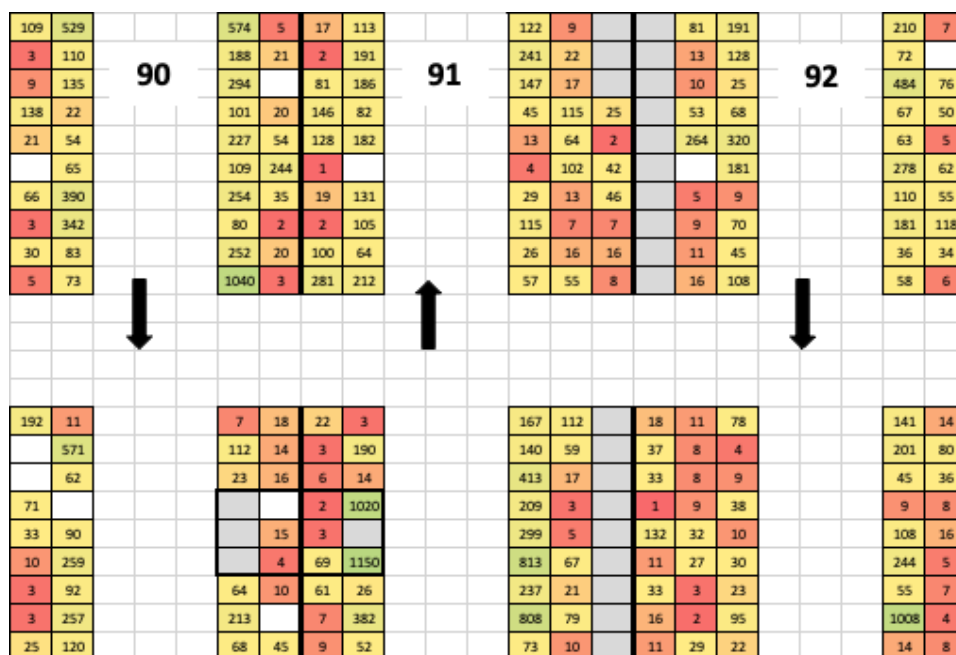
Artiklarna som hanteras av grossistverksamheten är fördelade i olika zoner på lagret beroende på dess egenskaper. Artiklar som kräver lägre temperatur placeras exempelvis i kyl- eller fryszonen. Vid plockning av order använder sig grossistverksamheten av zonplockning. Zonplockning innebär att en plockorder delas upp i olika order beroende på vilken zon som artiklarna tillhör. En plockare börjar plocka artiklar som hör till en zon, och sedan fortsätter samma eller en annan plockare att plocka artiklarna som finns i en annan zon. För att underlätta orderplockningsprocessen använder grossistverksamheten det tekniska hjälpmedlet Pick by Voice som är ett röststyrt plocksystem. Plockarna förflyttar sig med plocktruckar och under en dag kan det samtidigt röra sig upp emot 90 plocktruckar inom samma zon.

### 4.1.2 Artikelplacering

Grossistverksamheten använder data som samlats in från orderplockningsprocessen för att genomföra en dubbel ABC-analys och till ett frekvensläggningsverktyg. Den dubbla ABC-analysen och frekvensläggningsverktyget används sedan som beslutsunderlag för artikelplacering. Den data som används är plockfrekvens, det vill säga hur ofta en plockare besöker platsen för den artikel som ska plockas, och försäljningsvolym, det vill säga hur mycket av en artikel som ska plockas.

Grossistverksamheten genomför en dubbel ABC-analys som baseras på plockfrekvens och försäljningsvolymen för att identifiera vilka artiklar som är felplacerade och kan placeras på ett mer fördelaktigt sätt. Ett fördelaktigt sätt innebär att placeringen bidrar till att minska körsträcka. Den dubbla ABC-analysen genomförs månadsvis med en tidsram på en månad. Den dubbla ABC-analysen genomförs främst på varm- och frysagret.

Frekvensläggningsverktyget baseras på plockfrekvensen och används av grossistverksamheten som för att se hur ofta en plockare besöker en plockplats. Detta visualiseras som en ”heatmap”, se figur 4.1. Gröna rutor representerar högfrekventa artiklar, gul till orange representerar medelfrekventa artiklar och röda rutor representerar lågfrekventa artiklar. Pilarna representerar körriktning och siffrorna representerar gångnummer. Kartan används för att få en överblick över frekvensen på artiklarna och hur dem är placerade på lagret.



Figur 4.1: Utdrag ur frekvensläggningsverktyget från grossistverksamheten.

Resultatet från ABC-analysen och frekvensläggningsverktyget används för att placera artiklarna på ett fördelaktigt sätt. Grossistverksamheten strävar efter att placera artiklar som plockas frekvent så nära tillsammans som möjligt för att minska plocksträckan. Dock kan högfrekventa artiklar inte placeras för nära varandra, då det kan leda till köer mellan plockarna. Om det bildas köer får plockarna vänta och det uppstår kötid som är en form av icke värdeskapande tid. Grossistverksamheten menar därför att det är viktigt att de hittar en balans mellan hur tätt de placerar de högfrekventa artiklarna för att ha en så hög del värdeskapande tid som möjligt. Artikelplaceringen baseras inte strikt på analysen utan lagerläggarna baserar också artikelplaceringen på egen känsla. Artiklar placeras också utefter ett säkerhetsperspektiv och artiklarnas olika egenskaper. Till exempel ska högfrekventa och tyngre artiklar placeras på ergonomiska plockplatser. Kyl- och frysartiklar placeras i zoner med lägre temperaturer.

#### **4.1.3 Definiera målen för datautvinningen**

Grossistverksamheten använder sig av affärssystemet Servo och lagerstyrningssystemet Astro. Det är från lagerstyrningssystemet som data för dubbla ABC-analysen och frekvenslagerläggningsverktyget hämtas. Grossistverksamheten menar att den data som tas ut från affärssystemen är av enklare karaktär. De önskar att kunna nyttja mer data i deras datalager för att kunna göra komplexare analyser. Grossistverksamheten vill att den data som finns ska visa hur dem på ett bra sätt bör placera artiklarna, men menar också att det är nödvändigt att göra en analys av vilka egenskaper som artiklarna har. Vidare menar grossistverksamheten att deras verksamhet är beroende av säsong och det är därför intressant att veta hur det såg ut vid samma tidpunkt tidigare år. Exempelvis är det inför julen 2022, intressant att veta vilka samband som fanns julen 2021, 2020 och så vidare.

Grossistverksamhetens önskan är att driva verksamheten med höga mål gällande resurseffektivitet och hållbarhet. Den önskade effekten är att öka den värdeskapande tiden för orderplockningsprocessen, det vill säga minska den tiden det tar för en plockare att plocka en order till kund. Grossistverksamheten använder sig av flera artikelplaceringsprinciper, men ingen som tar hänsyn till ett samband mellan artiklar. Verksamhetens problem översattes till ett datautvinningsmål: "målet med projektet är att hitta par av artiklar som ofta köps tillsammans under en given tidpunkt för att sedan användas som beslutsunderlag vid artikelplacering".

## 4.2 Förstå datamängden

I detta kapitel presenteras insamlingen av datamängden samt identifierade begränsningar av insamlingen. Sedan presenteras en analys av den insamlade datamängden.

### 4.2.1 Insamling av initial datamängden

För att få tillgång till data som ska uppnå målet med datautvinning och användas till modellen hämtades datamängder från grossistverksamhetens dataredovisningssystem. Dataredovisningssystemet låter användaren bestämma vilken typ av data (attribut) som ska extraheras. Data som ansågs nödvändig för att skapa modellen var ordernummer, leveransdatum och artikelnummer. För att hämta datamängder från dataredovisningssystemet behövdes minst ett attribut och ett mätvärde anges. Då utvinning av sambandsregler inte är beroende av några mätvärden valdes i samråd med grossistverksamheten plockrader som mätvärde. Filerna med datamängden exporterades till Microsoft Excel i ett transaktionsdataformat, se figur 4.2.

Kund	Lev datum	Artikel	Måttenheter
41245	Företag A	2022-03-01 010694	ÄPPLE AROMA SE FGX Antal plockrader logistik 1
41245	Företag A	2022-03-01 072637	TOMAT COCKTAIL RÖD 250G FGX Antal plockrader logistik 1
41245	Företag A	2022-03-01 103127	MOROT 12KG SRS SGS Antal plockrader logistik 1
41245	Företag A	2022-03-01 109223	SVAMP PORTABELLA FGX Antal plockrader logistik 1
41245	Företag A	2022-03-01 118513	ÄRTER GRÖNA PETIT POIS2,5KGFIN Antal plockrader logistik 1
58923	Företag B	2022-03-01 124578	FLÄSKKARRE BF DE KY 2,5/10KSER Antal plockrader logistik 1
58923	Företag B	2022-03-01 126003	MYNTA KNIPPE 100G FGX Antal plockrader logistik 1
58923	Företag B	2022-03-01 126292	POTATIS 48-52mm 10KG GÄR Antal plockrader logistik 1
58923	Företag B	2022-03-01 139527	PAPRIKA GRÖN FGX Antal plockrader logistik 1
58923	Företag B	2022-03-01 142943	CREME FRAICHE 32% 2L M/S Antal plockrader logistik 1
58923	Företag B	2022-03-01 148908	POTATIS KUL FARMARKL 10KG MMP Antal plockrader logistik 1
34576	Företag C	2022-03-01 175992	RIS LÅNGK PARB 10KG M/S Antal plockrader logistik 1
34576	Företag C	2022-03-01 179226	PAPRIKA RÖD FGX Antal plockrader logistik 1
34576	Företag C	2022-03-01 190298	BASILIKA PÅSE 100G FGX Antal plockrader logistik 1
34576	Företag C	2022-03-01 244657	CHILI GRÖN HG FGX Antal plockrader logistik 1
34576	Företag C	2022-03-01 244665	CHILI RÖD HG FGX Antal plockrader logistik 1
34576	Företag C	2022-03-01 298091	PERSILJA BLAD 250G FGX Antal plockrader logistik 1
34576	Företag C	2022-03-01 299289	LÖK RÖD 5KG FGX Antal plockrader logistik 1

Figur 4.2: Format på filerna som hämtades från grossistverksamhetens dataredovisningssystem i ett transaktionsdataformat.

#### **4.2.2 Begränsningar av insamling av datamängden**

I samråd med grossistverksamheten bestämdes det först att hämta ut filer med data från en månad eftersom deras andra artikelplaceringsverktyg genomförs med en tidsram på en månad. Dataredovisningssystemet som genererade filerna med data hade en begränsning för antalet rader som kan tas ut i samma fil och datamängden under en månad var för stor för dataredovisningssystemet att hantera. Tidsramen begränsades därför till en vecka vilket dataredovisningssystemet kunde hantera. När filen importerades till Microsoft Excel uppstod processproblem, då mängden data var för stor för Microsoft Excel att hantera. När filen manipulerades slutade Microsoft Excel att svara eller kraschade. För att få en hanterbar datamängd för Microsoft Excel begränsades tidsramen till en dag.

#### **4.2.3 Analysera datamängden**

En inledande analys av datamängden utfördes för att fastställa kvaliteten och några insikter identifierades. Några av kolumnerna (attribut) i datamängden ansågs inte nödvändiga eller kommer ha en påverkan på resultatet av datautvinningen. Vid val av attributen kundnummer och artikelnummer följde attributen kundnamn och artikelnamn automatiskt med. Kundnamn och artikelnamn ansågs inte nödvändiga eftersom tillräcklig information fås av artikel- respektive kundnummer. Vid vidare analys av datamängden identifierades det att artiklar som inte är fysiska förekom i datamängden, exempelvis broavgifter och leveransavgifter. Artiklarna har ingen fysisk placering på lagret, plockas inte av plockare och ansågs därför kunna ge felaktiga samband som ska kunna användas till artikelplacering. Det identifierades också att returbackar, tomma backar som kunderna skickar tillbaka, fanns med på en majoritet av kundorderna. Returback är ingen artikel som har en plats på lagret, utan den används för att förpacka varorna. Då returbacken förekom i nästan alla order skulle den påverka resultatet eftersom den skulle utgöra samband med många artiklar. Det identifierades inga saknade värden i datamängderna.

### 4.3 Förbereda datamängden

En modellsamling skapades genom att ta ut data från sju olika datum från olika månader under samma år med attributen kundnummer, leveransdatum, artikelnummer och plockrader. När filerna hämtades från grossistverksamhetens dataredovisningssystemet följde attributen kund- och artikelnamn automatiskt med attributen kund- och artikelnummer. Baserat på analysen av datamängden ansågs de överflödiga och därför togs kolumnerna kund- och artikelnamn bort. Attributet och mätvärdet som behövde väljas för att hämta filer togs också bort.

Artikeln returback, som inte har någon specifik placering på lagret, kunde grossistverksamheten filtrera bort innan filerna togs ut. Andra artiklar som tillkom på ordena som inte har någon fysisk placering på lagret valdes att ha kvar i datamängden. Det gjordes ett antagande att baserat på de order som plockades under en dag var andelen icke fysiska artiklar relativt få och skulle därför försvinna när a-priori algoritmen senare tillämpades. Att manuellt ta bort dessa artiklar från varje fil bedömdes vara en för tidskrävande process.

Efter rensningen resulterade datamängden i en modellsamling bestående av sju filer med attributen kundnummer, leveransdatum och artikelnummer. Formatet var ett transaktions dataformat för att sedan kunna användas i skapandet av modellen, se figur 4.3. Den rensade datamängden har de nödvändiga egenskaperna för att skapa en modell för att hitta sambandsregler mellan artiklar under en tidsperiod.

Kundnummer	Leveransdatum	Artikelnummer
41245	2022-03-01	010694
41245	2022-03-01	072637
41245	2022-03-01	103127
41245	2022-03-01	109223
41245	2022-03-01	118513
58923	2022-03-01	124578
58923	2022-03-01	126003
58923	2022-03-01	126292
58923	2022-03-01	139527
58923	2022-03-01	142943
58923	2022-03-01	148908
34576	2022-03-01	175992
34576	2022-03-01	179226
34576	2022-03-01	190298
34576	2022-03-01	244657
34576	2022-03-01	244665
34576	2022-03-01	298091
34576	2022-03-01	299289

Figur 4.3: Formatet på filerna i modellsamlingen efter rensning.

Efter rensningen transformerades filerna till ett tabellformat (figur 4.4) där den vågräta raden representerar artikelnummer och den lodrätta kolumnen representerar kundnummer. Ettorna representerar om artikeln förekommer på kundordern.

	438939	528141	528208	528240	537779	573659
34543					1	
43513						
34534						
64566	1	1	1	1		
65775					1	1
86785						
36457				1		
75675	1				1	1
23496			1			
34563	1					
75567						1
87567		1				
56763				1		
35345						
64577			1			1
56747	1	1			1	
85678						
56756				1		

Figur 4.4: Kundorderdata i ett tabellformat.

## 4.4 Skapa en modell

I detta kapitel presenteras först uppdelningen av modellsamlingen. Sedan beskrivs skapandet av modellen som används för att utvinna sambandsregler.

### 4.4.1 Dela upp modellsamlingen

De sju filerna som hämtades till modellsamlingen delades upp i tre olika samlingar. Träningssamlingen bestod av en fil som användes för att skapa modellen. Testsamlingen bestod av tre filer för att testa modellens kvalitet och funktion. Utvärderingssamlingen bestod av tre filer för att utvärdera resultatet som modellen ger. Uppdelningen av data görs för att utvärdera modellen vilket bestämmer kvaliteten på modellen och resultaten som genereras. Modellen byggdes på att använda träningssamlingen medan modellens funktioner och kvalitet testades med hjälp av testsamlingen och utvärderingssamlingen.

#### 4.4.2 Skapande av modellen i Microsoft Excel

Efter att modellsamlingen delats in i en träningsamling, testsamling och utvärderingssamling skapades modellen från träningsamlingen. För att automatisera utvinningen av sambandsregler så mycket som möjligt består modellen av två makron som kodades i Visual Basic Editor, som är en inbyggd applikation i Microsoft Excel för att skriva makron. Det första makrot (bilaga 2) räknar ut frekvensen av varje unik artikel. Detta görs genom att verktyget undersöker varje order i datamängden och beräknar i hur många ordrar som en artikel förekommer.

När frekvensen har identifierats för varje unik artikel kodades det så att användaren av modellen manuellt väljer en frekvens a-priori algoritmen ska utgå från, se figur 4.5. Verktyget kodades så att när frekvensen är bestämd kan användaren se hur många unika artiklar som den genererar. Modellen sorterar sedan bort artiklar som förekommer mer sällan än den valda frekvensen. Det andra makrot (bilaga 3) hittar sedan alla möjliga kombinationer mellan artiklarna i den frekventa mängden, samt hur ofta paren av artiklar förekommer tillsammans under en dag.

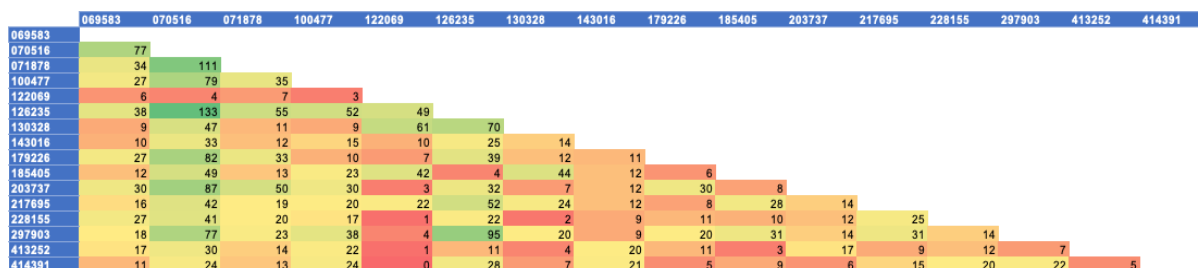
<b>Antal order:</b>	<b>2359</b>	
<b>Antal unika artiklar:</b>	<b>7589</b>	
<b>Frekvens:</b>		Ange minimal frekvens för artikel
<b>Antal unika artiklar genererade:</b>	<b>7589</b>	

Figur 4.5: Instrumentpanel där användaren väljer en frekvens.

#### 4.4.2 Visualisering av sambandsreglerna

För att visualisera sambandsreglerna som genereras från modellen delades resultat upp i två delar. Den första delen är en sambandskarta där användaren visuellt kan se alla möjliga samband mellan par av artiklar och antalet gånger sambanden förekommer, se figur 4.6. För att öka visualiseringen av sambanden kodades resultatet som en heatmap där färgskalan beskriver frekvensen av sambandet. Färgskalan går från grön till röd färg där grön representerar högfrekventa samband och röd representerar lågfrekventa samband. Syftet var att användaren av verktyget ska få en snabb överblick mellan vilka två artiklar som det förekommer ett frekvent samband.





Figur 4.6: Sambandskarta med färgskala som redovisar antalet gånger sambandet mellan två artiklar förekommer i vald datamängd. Färgskalan går från grön till röd färg där grön representerar högfrekventa samband och röd representerar lågfrekventa samband.

Den andra delen av resultatet presenterar stöd och tillförlitlighet för sambanden mellan artiklarna. Verktøget kodades så att användaren av modellen väljer en artikel att undersöka och sedan beräknas stöd och tillförlitlighet för artikeln tillsammans med andra artiklar, se figur 4.7. Resultatet sorterades på tillförlitlighet, så att den artikeln som har högst tillförlitlighet hamnar överst i resultatet. Det resultatet presenterar är exempelvis att artikel 130328 och artikel 126235 förekommer 70 gånger tillsammans som utgör 3,05 % av alla ordrar med en tillförlitlighet på 50 %.

Artikel 1	Artikel 2	Antal samband	Stöd	Tillförlitlighet
130328	130328	140	6,09%	100,00%
	126235	70	3,05%	50,00%
	122069	61	2,66%	43,57%
	070516	47	2,05%	33,57%
	185405	44	1,92%	31,43%
	573659	43	1,87%	30,71%
	217695	24	1,04%	17,14%
	297903	20	0,87%	14,29%
	532002	18	0,78%	12,86%
	435917	15	0,65%	10,71%
	143016	14	0,61%	10,00%
	179226	12	0,52%	8,57%
	528133	12	0,52%	8,57%
	537779	11	0,48%	7,86%
	071878	11	0,48%	7,86%
	100477	9	0,39%	6,43%

Figur 4.7: Resultat som redovisar antal samband, stöd och tillförlitlighet för samband med den valda artikeln 130328. Den röda raden visar ett ointressant samband då artikel 130328 representerar artikel 1 och artikel 2.

## **4.5 Utvärdera modellen och dess resultat**

Den slutliga modellen utvärderas för att kontrollera om modellen överensstämmer med målen för detta projekt. Följande kapitel presenterar resultatet kontra företagsmålet och en utvärdering om modellen genererar användbara sambandsregler.

### **4.5.1 Resultatet kontra företagsmål**

Målet med datautvinningsprojektet som definierades i första fasen av CRISP-DM metoden var “målet med projektet är att hitta par av artiklar som ofta köps tillsammans under en given tidpunkt för att sedan användas som beslutsunderlag vid artikelplacering”. Modellen hittar par av artiklar som ofta köps tillsammans under en given tidpunkt, men har en begränsad tidsram på en dag. Det beror på att Microsoft Excel som modellen är utvecklad i inte kan hantera den datamängd som grossistverksamheten genererar för längre tidsperioder. Grossistverksamheten ansåg för att ta få ut så bra samband som möjligt ska modellen kunna kolla på längre tidsperioder för att identifiera vilka artiklar som köps tillsammans under olika säsonger. Tidsramen för modellen har begränsats till en dag, då det i nuläget genererar en hanterbar datamängd uppemot 60 000 rader. Modellen kan alltså inte identifiera artiklar som köps tillsammans under olika säsonger.

### **4.5.2 Användbara sambandsregler**

Sambanden som modellen genererar anses vara korrekta med stöd och tillförlitligheten som avrundas uppåt till procent med två decimaler. Alla samband som genereras är dock inte användbara för artikelplacering. På grund av att alla icke fysiska artiklar inte kommer rensas bort ur datamängden finns det en risk att icke intressanta samband genereras i resultatet. Detta minskar kvaliteten på modellen och dess resultat. Modellen tar inte heller hänsyn till artikelspecifika egenskaper, exempelvis att tunga artiklar måste placeras på specifika platser. Detta gör att resultatet kan presentera starka samband som anses vara starka enligt värderingsfaktorerna stöd och tillförlitlighet. Modellen genererar dock samband som inte anses användbara då det inte är möjligt att fysiskt anpassa artikelplaceringen efter sambandet.

## 5. Diskussion

Detta kapitel tar upp en diskussion av analysen relaterade till forskningsfrågorna, följt av en diskussion gällande begränsningar vad gäller arbetets metod. Slutligen diskuteras arbetet utifrån hållbarhet och etik.

### 5.1 Svar på första frågeställningen

Den första frågeställningen är: *Vilka är förutsättningarna som krävs för att utvinna användbara sambandsregler för artikelplacering?*. För att svara på denna fråga jämförs det teoretiska ramverket med insikterna som erhållits under metodologin CRISP-DM.

Om inte grossistverksamheten hade haft ett fungerande dataredovisningssystem för att hämta ut filer med datamängder hade inte sambandsregler kunnat utvinnas. För att utvinna sambandsregler måste det finnas historiska data på vilka artiklar kunder köper i en order och därför identifieras den första förutsättningen till att det måste finnas *tillgänglig kundorderdata*.

Vidare menar Gorunescu (2011) att för att få statistiskt bra resultat bör den historiska datamängden innehålla hundratal eller tusentals transaktioner. På samma sätt menade grossistverksamheten att det var av hög vikt för dem att modellen kan kolla på olika tidsperioder för att få ut ett så användbart resultat som möjligt. Dock var mängden data som grossistverksamheten genererade för de längre tidsperioder som de ville undersöka ohanterbart både för deras egna dataredovisningssystem och Microsoft Excel. Därför begränsades tidsramen till en mindre period för att dataredovisningssystemet och Microsoft Excel skulle kunna hantera datamängden. Utifrån detta definierades en förutsättning att det krävs ett *verktyg som kan hantera datamängden* som sambanden ska utvinnas från. Detta efterliknar även förutsättningen som Jukic och Nestorov (2005) presenterar; att det ofta krävs ett separat verktyg med stor processorkraft för att hitta sambandsregler.

Grossistverksamhetens datamängd innehöll artiklar som inte har en fysisk placering på lagret, vilket kommer ge felaktiga sambandsregler om resultatet ska användas för artikelplacering. Själva utvinningen av sambandsregler fungerar även fast "felaktiga" artiklar används, men kvaliteten på resultatet av utvinningen försämras. För att utvinna användbara sambandsregler med resultat som motsvarar målet med utvinningen behövs data som är relevant för syftet. Baserat på detta identifierades en förutsättning att det krävs *ren och relevant data* för att utvinna användbara sambandsregler. I litteraturen påminner förutsättningen om det som Brown och

Kros (2003) menar; att korrupt data kan lätt leda till oanvändbara regler eller att det inte går att identifiera sambandsregler som normalt finns i datamängden.

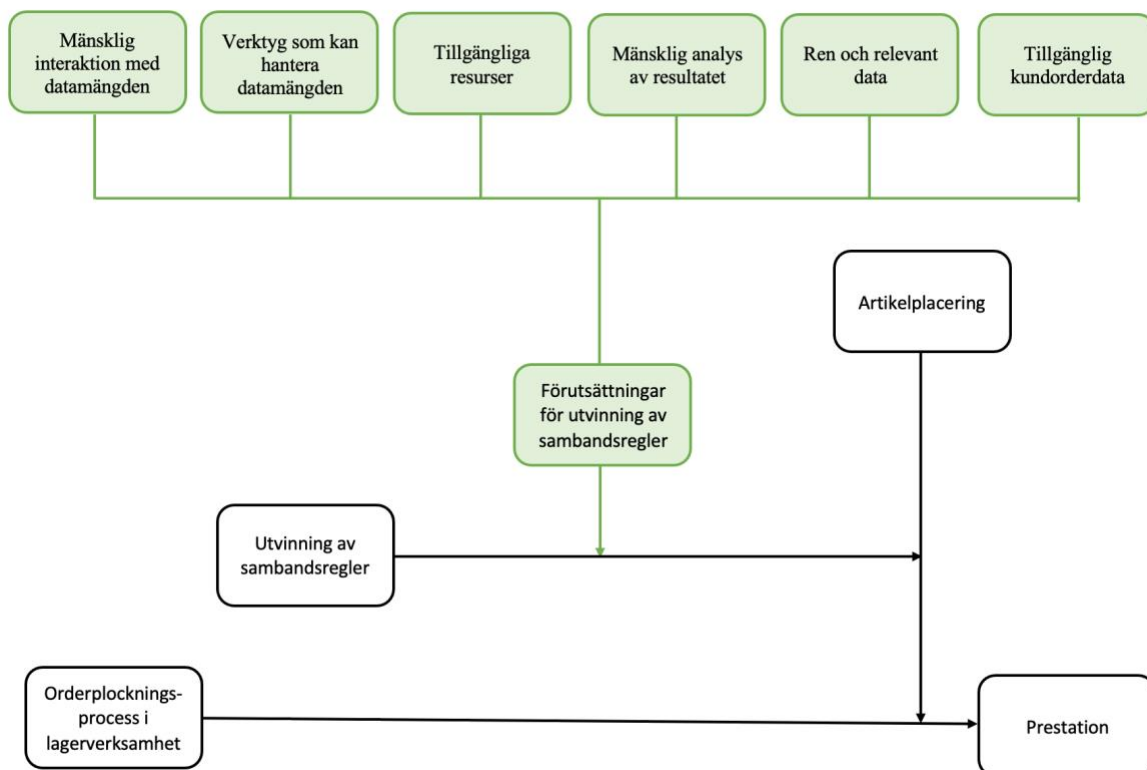
Under analysen hämtades, organiserades och förbereddes datamängden manuellt innan den importerades till modellen. För att göra detta krävdes mänsklig interaktion med datamängden. Eftersom modellen baseras på a-priori algoritmen måste användaren själv bestämma en lämplig frekvens som modellen ska utgå ifrån. Utifrån dessa två aspekter definierades en förutsättning till att det krävs *mänsklig interaktion med datamängden* för att utvinna sambandsregler. Från litteraturen påstår D. Larose och C. Larose (2015) att processen för datautvinning inte kan vara helt automatiskt, vilket analysen visar även gäller för utvinning av sambandsregler. Förutsättningen om mänsklig interaktion skulle även kunna vara användbar för att uppnå den tidigare nämnda förutsättningen, att det krävs ren och relevant data. Dels kan det krävas mänsklig undersökning av datamängden för att upptäcka att datamängden inte är ren, dels kan användaren rensa datamängden så att den blir ren innan den appliceras i verktyget.

Vidare beskrivs det i litteraturen att personer som arbetar med datautvinning aldrig bör lita på dess resultat (D. Larose & C. Larose, 2015). De sambandsregler som genererats från modellen i analysen tar hänsyn till relationer mellan olika artiklar, men inte vilka egenskaper artiklarna har. Detta innebär att samband som genereras inte behöver vara användbara ur en artikelplaceringssynvinkel för en lagerverksamhet. Ett exempel är om en lätt artikel har ett starkt samband med en tung artikel. Det kan då vara fördelaktigt att placera dessa nära varandra för att effektivisera orderplockningsprocessen, men grossistverksamheten strävar efter att placera tunga artiklar i början av plocksträckan och lätta artiklar i slutet. Anledningen är om tyngre artiklar placeras ovan på lättare artiklar finns risken att de förstörs. Utifrån detta resonemang identifieras en förutsättning, för att få ett användbart resultat av utvinningen, att det krävs en *mänsklig analys av resultatet* som genereras av modellen.

Grossistverksamheten menar att orsaken till att de inte har utvunnit sambandsregler är för att det saknats kompetens och tid. Som om en följd av att det vid utvinning av sambandsregler krävs mänsklig interaktion, måste det finnas tillgänglig tid för att förbereda utvinningen, genomföra utvinningen och analysera resultatet. Även Jukic och Nestorov (2005) menar att utvinning av sambandsregler är en tidskrävande process. Utöver tillgänglig tid måste det finnas tillräcklig kompetens för att kunna genomföra utvinningen och kunna analysera resultatet. Den mänskliga interaktionen på både datamängden och resultatet måste göras av någon med kompetens om modellens uppbyggnad och hur sambandsregler värderas. Om inte denna

kompetens finns kommer troligtvis inte ett användbart resultat att uppnås. D. Larose och C. Larose (2015) menar, generellt för datautvinning, att den mänskliga expertisen måste användas för att nå användbara resultat. Att göra analyser utan att förstå de underliggande faktorerna för resultatet menar dem resulterar i något som inte är användbart. Utifrån arbetets analys, med stöd av teorin, definieras en förutsättning att det måste finnas *tillgängliga resurser* för att utvinna sambandsregler med användbara resultat för artikelplacering.

Utifrån diskussionen har sex förutsättningar som behövs för att utvinna användbara sambandsregler för att stödja artikelplacering identifierats. Dessa är behovet av *tillgänglig kundorderdata*, *verktyg som kan hantera datamängden*, *ren och relevant data*, *mänsklig interaktion med datamängden*, *mänsklig analys av resultatet* och *tillgängliga resurser*. I figur 5.1 visualiseras detta bidrag till teorin och hur det hänger samman med prestationen av orderplockningsprocessen.



Figur 5.1: Studiens bidrag (gröna rutor) till utveckling av artikelplacering i orderplockningsprocessen i en lagerverksamhet.

## 5.2 Svar på andra frågeställningen

Den andra frågeställningen är: *Hur kan sambandsregler utvinnas för att stödja beslutsfattare mot en utvecklad artikelplacering?*

För att utvinna sambandsregler som kan stödja beslutsfattare mot en utvecklad artikelplacering, identifierades det att en förutsättning är att det måste finnas ett *verktyg som kan hantera datamängden*. Som redovisas i analysen kan sambandsregler utvinnas genom Microsoft Excel. Detta är dock enligt resultatet från analysen förutsatt att den valda datamängden som ska undersökas är en hanterbar mängd för Microsoft Excel. Mohd Ali et al. (2016) menar att Microsoft Excel är ett bra verktyg för dataanalys och visualisering av data, men tar inte upp något om begränsningarna av mängden data som kan hanteras. Microsoft Excel anses inte från analysen kunna vara en programvara som kan användas i samtliga fall av utvinning av sambandsregler och visualisering. Programvaran hade exempelvis inte klarat av att undersöka grossistverksamhetens datamängd från en tidsperiod på en månad då för mycket data genereras.

Verktygen som skapades i Microsoft Excel består först av ett steg där a-priori algoritmen appliceras för att antalet sambandsregler som ska genereras blir en hanterbar mängd. A-priori algoritmen sorterar bort artiklar som inte är frekventa enligt en vald frekvens. En lämplig frekvens väljs av användaren. Sedan räknar det kodade programmet ut alla möjliga kombinationer mellan artiklarna i den frekventa mängden och hur ofta paren av artiklar förekommer. Till sist beräknar verktyget stöd och tillförlitlighet för de genererade sambanden. Analysen visar att trestegsprocessen som D. Larose och C. Larose (2015) beskriver för att generera sambandsregler är applicerbar på datamängder från en lagerverksamhet där syftet är att använda sambandsregler för att utveckla artikelplacering.

För att kunna utvinna sambandsregler som ska stödja beslutsfattare mot en utvecklad artikelplacering är det viktigt att resultaten som modellen genererar presenteras på ett lättförståeligt sätt. Det är viktigt att viktiga detaljer som stöd, tillförlitlighet och vilka artiklar som sambanden förekommer mellan är tydligt visuellt presenterade, så att det blir enkelt för en beslutsfattare att analysera resultatet. Detta stöds även av Mohd Ali et al. (2016) som menar att det är av stor vikt att resultatet från en datautvinning presenteras på ett visuellt tydligt sätt för att det ska bli lättare att ta beslut utifrån resultatet.

Resultatet som modellen genererar kan därför presenteras i två delar. En del är en sambandskarta för att få en översikt över intressanta samband mellan artiklar. Sambanden i sambandskartan färgkodas för att snabbt ge en visuell överblick om vilka samband som är högfrekventa respektive lågfrekventa. Detta styrks även Mohd Ali et al. (2016) som menar att visualisering av data, exempelvis genom färgkodning, gör att det blir snabbt och enkelt att förmedla budskap till beslutsfattare. Den andra delen är ett mer detaljerat resultat där alla samband utvärderas utifrån stöd, tillförlitlighet och antalet gånger sambandet uppkommer under den insatta tidsperioden. För att underlätta sorteras resultatet automatiskt så att samband med högt stöd och tillförlitlighet presenteras överst. Det räcker dock inte att sambandsreglerna som utvinns presenteras på ett visuellt sätt som är lätt för en beslutsfattare att förstå. För att sambandsreglerna som utvinns ska vara användbara krävs att förutsättningarna identifierade från första frågeställningen uppfylls.

### **5.3 Metoddiskussion**

En begränsning av denna studie rör den begränsade generaliserbarheten av resultatet på grund av den antagna designen för en fallstudie. En enfallsstudie påverkar reliabiliteten av studien då resultatet inte testas på andra företag. Om fler fallföretag hade studerats, hade dessa företag kunnat verifiera de resultat som framkommit i denna studie. Dock ansågs det nödvändigt att börja med en enfallsstudie eftersom studiens fokus och frågorna är explorativa och krävde en fördjupning. Framtida forskning hade kunnat dra nytta av en flerfallsstudie som täcker fler lagerverksamheter för att ytterligare utforska olika förutsättningar i andra miljöer. En flerfallsstudie hade kunnat påvisa fler eller andra förutsättningar än vad som identifierats i detta arbete.

De intervjuer som genomfördes i datainsamlingssyfte gjordes med två personer på grossistverksamheten. Personerna hade heller ingen direkt erfarenhet av datautvinning mot artikelplacering. Om fler personer hade intervjuats hade insamlad data kunnat berikas samt jämföras mellan fler personer vilket hade kunnat öka validiteten av datainsamlingen. Det hade varit intressant att intervjua personer som arbetar med datautvinning och som har genomfört ett datautvinningsprojekt för att erhålla deras erfarenheter kring utvinning av sambandsregler. Intervjuerna som genomfördes var också relativt ostrukturerade. Bryman (2012) menar att strukturerade intervjuer resulterar i en högre reliabilitet, då strukturen på den insamlade data från intervjun blir bättre.

Verktyget för att utvinna sambandsregler som skapades under dataanalysen hade kunnat skapas med andra programvaror än Microsoft Excel. Exempelvis hade starkare och mer anpassade dataanalysverktyg för ändamålet kunnat användas, såsom Microsoft Power BI eller Tableau Software. Anledningen till att Microsoft Excel valdes var att fallföretaget använde detta system för andra artikelplaceringsverktyg, samt att det var ett verktyg som fanns tillgängligt för arbetet. Ett annat program hade kunnat ge ett annat svar på andra frågeställningen hur sambandsregler kan utvinnas för beslutsfattare, men också kunnat påvisa andra förutsättningar som krävs.

#### **5.4 Diskussion utifrån hållbarhet och etik**

Hållbar utveckling definierar Hedenus et al. (2014) som utveckling som möter dagens behov utan att kompromissa framtida generationers förmåga att möta deras behov. Hållbar utveckling kan delas upp i tre dimensioner; ekonomisk, ekologisk och social hållbarhet. För att förstå vilken påverkan sambandsregler i artikelplacering kan ha för en lagerverksamhet analyseras det utifrån de tre hållbarhetsdimensionerna.

Om sambandsreglerna används som beslutsunderlag för artikelplacering menar Chiang et al. (2011) att det kan förbättra effektiviteten för orderplockningsprocessen. Från ett ekonomiskt perspektiv kan en effektivare artikelplacering leda till minskade plocksträckor som i sin tur leder till minskade lagerkostnader. En plockare lägger mycket tid på att hämta varor för ordrar, och dessa arbetskostnader kan vara höga om det inte görs effektivt.

Miljöfördelarna av en effektivare orderplockningsprocess kan ge en minskad energiförbrukning då truckarna förväntas köra kortare sträckor. Från ett socialt perspektiv kan en implementering av sambandsregler i artikelplacering som effektiviserar orderplockningsprocessen förbättra arbetsförhållanden. En effektivare orderplockningsprocess utnyttjar plockarnas tid bättre, vilket med största sannolikhet minskar behovet av overtid.

När resultatet från detta arbete analyseras genom de tre hållbarhetsdimensionerna ovan dras slutsatsen att implementering av sambandsregler i artikelplacering kan leda till flera hållbara fördelar: effektivare orderplockningsprocess, minskade lagerkostnader, minskad energiförbrukning och ökad effektivitet av daglig drift.

Vid intervjuer som genomförts är det enligt Bryman och Bell (2011) viktigt att informera respondenterna om syftet med intervjun. Det är också viktigt att respondenterna ger sitt medgivande till att deras svar används i studien. Före varje intervju förklarades syftet med den



och vad svaren skulle användas till. Namnen på respondenterna hölls anonyma dels för att säkerställa grossistverksamhetens anonymitet, dels för att respektera respondenterna. Eventuell kritik som tas upp i arbetet riktas mot verksamhetens processer, inte enskilda individer, grupper eller avdelningar på företaget.

## 6. Slutsats

Syftet med arbetet var att föreslå hur en lagerverksamhet kan utvinna sambandsregler som ska användas som beslutsunderlag för att utveckla artikelplacering, samt identifiera förutsättningar för att utvinna användbara sambandsregler.

Arbetet föreslår sex förutsättningar som behövs för att utvinna användbara sambandsregler för att stödja artikelplacering. Dessa är behovet av *tillgänglig kundorderdata, verktyg som kan hantera datamängden, ren och relevant data, mänsklig interaktion med datamängden, mänsklig analys av resultatet och tillgängliga resurser.*

Vidare visar studien ett exempel på hur sambandsregler kan utvinnas genom en modell som utvecklats i Microsoft Excel. Modellen baseras på två makron som är kodade i Visual Basic Editor för att automatisera utvinningen så mycket som möjligt. Resultaten från modellen redovisar visuellt de samband som identifieras från en grossistverksamhets kundorderdata. Sambanden utvärderas utifrån faktorerna stöd och tillförlitlighet.

En lagerverksamhet kan använda modellen för att generera användbara sambandsregler, genom att först uppfylla de förutsättningar som identifierats. Sambandsreglerna kan sedan användas som beslutsunderlag för placering av artiklar. Att placera artiklar utefter deras samband leder till en utvecklad artikelplacering som kan bidra till ökad prestation för orderplockningsprocessen. En förbättrad orderplockningsprocess kan i sin tur leda till minskade lagerkostnader, minskad energiförbrukning och ökad effektivitet av daglig drift.

Utvinning av sambandsregler för att utveckla artikelplacering för en lagerverksamhet är ett intressant undersökningsområde. På grund av begränsningar av studien anses den inte kunna generera generella svar för alla lagerverksamheter. För att uppnå en högre generaliserbarhet hade flera fallföretag behövts undersökas och jämföras. Ett ämne för vidare forskning skulle kunna vara vilka effekter implementeringen av sambandsregler har på orderplockningsprocessen. Att undersöka hur orderplockningens effektivitet förändras skulle kunna ge en ökad förståelse för betydelsen av att utvinna och använda sambandsregler vid artikelplacering.

## Referenser

Berry, M.J.A., & Linoff, G.S. (2011). *Data mining techniques: For Marketing, Sales, and Customer Relationship Management*. John Wiley & Sons.

Brown, M. L., & Kros, J. F. (2003). Data mining and the impact of missing data. *Industrial Management & Data Systems*, 103(8), 611–621. <https://doi.org/10.1108/02635570310497657>

Bryman, A. (2012). *Social research methods*. Oxford University Press.

Bryman, A. & Bell, E. (2011). *Business research methods*. Oxford University Press.

Chiang, D.M.H., Lin, C.P., & Chen, M.C. (2011) The adaptive approach for storage assignment by mining data of warehouse management system for distribution centres. *Enterprise Information Systems*, 5(2), 219–234. <https://doi.org/10.1080/17517575.2010.537784>

Dahlblom, U. (2003). *Matematisk statistik*. HB Matematiklitteratur.

Ejvegård, R. (2012). *Vetenskaplig metod*. Studentlitteratur AB.

Flyvbjerg, B. (2006). Five Misunderstanding About Case-Study Research. *Qualitative Inquiry*, 12(2), 219-245. <https://doi.org/10.1177/1077800405284363>

Flick, U. (2009). *An Introduction to Qualitative Research* (4. uppl.). SAGE Publications Ltd.

Gorunescu, F. (2011). *Data Mining*. Springer Publishing. <https://doi.org/10.1007/978-3-642-19721-5>

Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques* (3. uppl.). Morgan Kaufmann.

Hedenus, F., Persson, M., & Sprei, F. (2014). *Hållbar Utveckling. Historia, definition & ingenjörens roll*. Studentlitteratur AB. [http://hagesjo.se/static/books/Hållbar\\_utveckling.pdf](http://hagesjo.se/static/books/Hållbar_utveckling.pdf)

Jackson, J. (2002). Data Mining; A Conceptual Overview. *Communications of the Association for Information Systems*, 5(19), 267-296. DOI: 10.17705/1CAIS.00819

Jonsson, P., & Mattsson, S.A. (2016). *Logistik: läran om effektiva materialflöden*. Studentlitteratur AB.

Jukic, N., & Nestorov, S. (2005). Comprehensive data warehouse exploration with qualified association-rule mining. *Decision Support Systems*, 42(2), 859–878. <https://doi.org/10.1016/j.dss.2005.07.009>

Keim, D. A. (2002). Information visualization and visual data mining. *IEEE Transactions on Visualization and Computer Graphics*, 8(1), 1–8. <https://doi.org/10.1109/2945.981847>

Larose, D. T., & Larose C. D. (2015). *Data Mining and Predictive Analytics*. Wiley.

Mohd Ali, S., Gupta, N., Krishna Nayak, G. & Kumar Lenka, R. (2021). Big Data Visualization: Tools and Challenges. *2016 2nd International Conference on Contemporary Computing and Informatics (IC3I)*, 656-660. DOI: 10.1109/IC3I.2016.7918044.

Yin, R.K. (2017). *Case Study Research: Design and Methods*. SAGE Publications Ltd

## **Bilagor**

### **Bilaga 1: Intervjufrågor**

Hur används ABC-analysen?

Hur ofta gör ni analysen?

Under vilken tidsram görs analysen? (data från en månad, en vecka?)

Använder ni frekvensen från frekvensläggningsverktyget till ABC-analysen? Eller vad kommer frekvensfaktorn ifrån till ABC?

Hur används frekvensläggningsverktyget?

Hur ofta gör ni analysen?

Utvärderar ni er artikelplacering? Isåfall hur? Vilka mätvärden?

Hur vill ni att artiklarna ska vara placerade i en optimal situation? Vad vill ni undvika?

Hur skulle du bedöma företagets nuvarande artikelplacering?

Hur mycket kunskap anser du att ni har inom datautvinning idag?

Vilka affärssystem används idag? Används några till artikelplacering?

Vilka fördelar tror ni utvinning av sambandsregler kan ha inom artikelplacering?

Vilka förutsättningar tror ni krävs för att utvinna data för att hitta sambandsregler?

Vilka svårigheter tror ni finns att använda datautvinning för att hitta sambandsregler?

Vad är anledningarna till att ni inte använder datautvinning för att hitta sambandsregler?

Systemet där ni tar ut rapporterna, vad heter det? Vill du ge en liten kort beskrivning av det?

Måttenheter-kolumnen i rapporterna; varför behövde den alltid vara med? Gick den inte att ta bort? Vad betyder den?

## Bilaga 2: Makro 1 kodat i Visual Basic Editor

```
Sub Steg_1()  
  
Application.ScreenUpdating = False  
Application.DisplayStatusBar = False  
Application.EnableEvents = False  
  
'Byter namn på första fliken  
Sheets(1).Name = "Data"  
  
'Rensar bort tomma rader  
Range("1:1,2:2").Select  
Range("A1").Activate  
Selection.Delete Shift:=xlUp  
  
'Separerar rubrikerna  
Rows("1:1").Select  
With Selection  
    .HorizontalAlignment = xlGeneral  
    .VerticalAlignment = xlCenter  
    .WrapText = True  
    .Orientation = 0  
    .AddIndent = False  
    .IndentLevel = 0  
    .ShrinkToFit = False  
    .ReadingOrder = xlContext  
    .MergeCells = False  
End With  
Selection.UnMerge  
  
'Tar bort onödiga attribut  
Range("B:B,E:E,F:F,G:G").Select  
Range("G1").Activate  
Selection.Delete Shift:=xlToLeft  
  
'Räknar ut frekvensen för varje artikel  
lr = Cells.Find("*", Cells(1, 1), xlFormulas, xlPart, xlByRows,  
xlPrevious, False).Row  
Range("D1").Select  
ActiveCell.FormulaR1C1 = "Antal"  
Range("D2").Select  
ActiveCell.FormulaR1C1 = "=COUNTIF(C[-1],RC[-1])"  
Range("D2").Select  
Selection.Autofill Destination:=Range("D2:D" & lr)  
  
'Gör rådata till en tabell  
lr = Cells.Find("*", Cells(1, 1), xlFormulas, xlPart, xlByRows,  
xlPrevious, False).Row  
Range("A1:E1").Select  
Range(Selection, Selection.End(xlDown)).Select  
Application.CutCopyMode = False  
ActiveSheet.ListObjects.Add(xlSrcRange, Range("$A$1:$D$" & lr), ,  
xlYes).Name =  
    "Tabell1"  
  
'Formatering  
Sheets("Data").Select  
Columns("C:C").Select  
Selection.Copy
```

```

Columns("D:E").Select
Selection.PasteSpecial Paste:=xlPasteFormats, Operation:=xlNone, _
    SkipBlanks:=False, Transpose:=False
Application.CutCopyMode = False

'Apriorifliken
Sheets.Add After:=Sheets(1)
Sheets(2).Name = "Apriori"

'Instrumentpanelen
ActiveCell.FormulaR1C1 = "Antal order:"
Range("A2").Select
ActiveCell.FormulaR1C1 = "Antal unika artiklar:"
Range("A3").Select
ActiveCell.FormulaR1C1 = "Frekvens:"
Range("A4").Select
ActiveCell.FormulaR1C1 = "Antal unika artiklar genererade:"
Range("A5").Select
Columns("A:A").EntireColumn.AutoFit
Range("B1").Select
ActiveCell.FormulaR1C1 = "=COUNTA(UNIQUE(Tabell1[#Data],[Kund]))"
Range("B2").Select
ActiveCell.FormulaR1C1 = "=COUNTA(UNIQUE(Tabell1[#Data],[Artikel]))"

'Kolumn med falskt och sant i rådata
Sheets("Data").Select
Range("E1").Select
ActiveCell.FormulaR1C1 = "Artikel 2"
Range("E2").Select
ActiveCell.FormulaR1C1 = "=IF([@Antal]>(Apriori!R3C2-1),[@Artikel],0)"

'Antal unika
Sheets("Apriori").Select
Range("B4").Select
ActiveCell.FormulaR1C1 = "=IF(R[-
1]C>1,COUNTA(UNIQUE(Tabell1[#Data],[Artikel 2]))-1,Apriori!R[-2]C)"

'Skapar knappen "Räkna ut samband"
Sheets("Apriori").Select
ActiveSheet.Buttons.Add(10, 90, 120, 30).Select
Selection.OnAction = "PERSONAL.XLSB!Steg_2"
Selection.Characters.Text = "Räkna ut samband"
With Selection.Characters(Start:=1, Length:=13).Font
    .Name = "Calibri"
    .FontStyle = "Normal"
    .Size = 12
    .Strikethrough = False
    .Superscript = False
    .Subscript = False
    .OutlineFont = False
    .Shadow = False
    .Underline = xlUnderlineStyleNone

End With

'Formatering
Sheets("Data").Select
Range("Tabell1[#Headers],[Kund]").Select
Selection.Copy
Sheets("Apriori").Select
Range("A1:A4").Select

```

```
Selection.PasteSpecial Paste:=xlPasteFormats, Operation:=xlNone, _  
    SkipBlanks:=False, Transpose:=False  
Application.CutCopyMode = False  
Columns("A:A").ColumnWidth = 21  
Range("B1:B2").Select  
Application.CutCopyMode = False  
Selection.Style = "Beräkning"  
Range("B3").Select  
Selection.Style = "Indata"  
Range("B4").Select  
Selection.Style = "Utdata"  
Range("C3").Select  
Selection.Style = "Förklarande text"  
ActiveCell.FormulaR1C1 = "Ange minimal frekvens för artikel"  
Range("C4").Select  
  
Application.ScreenUpdating = True  
Application.DisplayStatusBar = True  
Application.EnableEvents = True
```

End Sub



## Bilaga 3: Makro 2 kodat i Visual Basic Editor

```
Sub Steg_2()  
  
Application.ScreenUpdating = False  
Application.DisplayStatusBar = False  
Application.EnableEvents = False  
  
'Tar bort flikar om användaren kör modellen igen  
Dim xWs As Worksheet  
Application.DisplayAlerts = False  
    For Each xWs In Application.ActiveWorkbook.Worksheets  
        If xWs.Name <> "Data" And xWs.Name <> "Apriori" Then  
            xWs.Delete  
        End If  
    Next  
    Application.DisplayAlerts = True  
  
'Filtrerar rådata beroende på frekvensen som användaren angivit  
Dim lo As ListObject  
Set lo = Sheets("Data").ListObjects("Tabell1")  
lo.Range.AutoFilter Field:=5  
Sheets("Data").ListObjects("Tabell1").Range.AutoFilter Field:=5,  
Criteria1:="<>0"  
  
'Kopierar den filtrerade data  
    Sheets("Data").Select  
    Range("Tabell1[#Headers]").Select  
    Range(Selection, Selection.End(xlDown)).Select  
    Selection.SpecialCells(xlCellTypeVisible).Select  
    Selection.Copy  
    Sheets.Add After:=Sheets("Apriori")  
    Sheets(3).Select  
    Sheets(3).Name = "Filtrerad data"  
    Range("A1").Select  
    ActiveSheet.Paste  
  
'Filtrerad data till en tabell  
    lr = Cells.Find("*", Cells(1, 1), xlFormulas, xlPart, xlByRows,  
xlPrevious, False).Row  
    Sheets("Filtrerad data").Select  
    Range("A1:E1").Select  
    Range(Selection, Selection.End(xlDown)).Select  
    Application.CutCopyMode = False  
    ActiveSheet.ListObjects.Add(xlSrcRange, Range("$A$1:$E$" & lr), ,  
xlYes).Name =  
        "Tabell2"  
  
'Pivot-tabell av rådata  
    Sheets("Filtrerad data").Select  
    Application.CutCopyMode = False  
    Sheets.Add After:=Sheets("Filtrerad Data")  
    Sheets(4).Name = "Pivot"  
    ActiveWorkbook.PivotCaches.Create(SourceType:=xlDatabase, SourceData:=  
    "Tabell2", Version:=7).CreatePivotTable  
TableDestination:="Pivot!R3C1",  
    TableName:="Pivottabell1", DefaultVersion:=7  
    Sheets("Pivot").Select  
    With ActiveSheet.PivotTables("Pivottabell1").PivotFields("Kund")
```

```

        .Orientation = xlRowField
        .Position = 1
    End With
    With ActiveSheet.PivotTables("Pivottabell1").PivotFields("Artikel")
        .Orientation = xlColumnField
        .Position = 1
    End With
    ActiveSheet.PivotTables("Pivottabell1").AddDataField
ActiveSheet.PivotTables( _
    "Pivottabell1").PivotFields("Artikel 2"), "Antal av Artikel 2",
xlCount
    With ActiveSheet.PivotTables("Pivottabell1")
        .ColumnGrand = False
        .RowGrand = False
    End With

'Kopierar data i Pivot-tabell och formaterar till en "vanlig" tabell
Application.DisplayAlerts = False
Sheets.Add After:=Sheets(4)
With Sheets(4).PivotTables(1)
With .PivotFields("Antal av Artikel 2").DataRange
.Offset(-1).Resize(.Rows.Count + 2).Copy Sheets(5).Range("A1")
End With
End With
Sheets(5).Delete
Sheets.Add After:=Sheets(4)
Sheets(5).Name = "Pivot-kopia"
With Sheets(4).PivotTables(1)
With .PivotFields("Antal av Artikel 2").DataRange
.Offset(-1).Resize(.Rows.Count + 2).Copy Sheets(5).Range("A1")
End With
End With
Sheets(5).ListObjects.Add(xlSrcRange, Range("A1", Cells(1,
Columns.Count).End(xlToLeft)).EntireColumn, , xlYes).Name = _
    "Tabell3"

'Resultat del 1
Range("A1").Select
Range(Selection, Selection.End(xlToRight)).Select
Selection.Copy
Sheets.Add After:=Sheets(5)
Sheets(6).Name = "Resultatkarta"
Range("B1").Select
Sheets(6).Paste
Range("A2").Select
Selection.PasteSpecial Paste:=xlPasteAll, Operation:=xlNone,
SkipBlanks:= _
    False, Transpose:=True

'Formatering
Cells.Select
Selection.FormatConditions.AddColorScale ColorScaleType:=3

Selection.FormatConditions(Selection.FormatConditions.Count).SetFirstPrioriti
ty
Selection.FormatConditions(1).ColorScaleCriteria(1).Type = _
    xlConditionValueLowestValue
With Selection.FormatConditions(1).ColorScaleCriteria(1).FormatColor
    .Color = 7039480
    .TintAndShade = 0
End With

```

```

Selection.FormatConditions(1).ColorScaleCriteria(2).Type = _
    xlConditionValuePercentile
Selection.FormatConditions(1).ColorScaleCriteria(2).Value = 50
With Selection.FormatConditions(1).ColorScaleCriteria(2).FormatColor
    .Color = 8711167
    .TintAndShade = 0
End With
Selection.FormatConditions(1).ColorScaleCriteria(3).Type = _
    xlConditionValueHighestValue
With Selection.FormatConditions(1).ColorScaleCriteria(3).FormatColor
    .Color = 8109667
    .TintAndShade = 0
End With

'Räknar ut antal samband
Dim lastRow As Long
lastRow = Range("A" & Rows.Count).End(xlUp).Row
Dim lc As Long, I As Long, j As Long
lc = Cells(1, Columns.Count).End(xlToLeft).Column
j = 3
For I = 2 To lc
    Cells(j, I).Formula2 =
"=SUM(INDIRECT("'"&R1C&'"')*INDIRECT("'"&R1&'"'))"
    Cells(j, I).Select
    Selection.Autofill Destination:=Range(Cells(j, I), Cells(lastRow,
I))
    On Error GoTo ErrorHandler1
    j = j + 1
Next I

ErrorHandler1:

'Resultat del 2
Sheets.Add After:=Sheets(6)
Sheets(7).Select
Sheets(7).Name = "Stöd och tillförlitlighet"

Cells.Select
With Selection
    .HorizontalAlignment = xlGeneral
    .VerticalAlignment = xlBottom
    .WrapText = False
    .Orientation = 0
    .AddIndent = False
    .IndentLevel = 0
    .ShrinkToFit = False
    .ReadingOrder = xlContext
    .MergeCells = False
End With
With Selection
    .HorizontalAlignment = xlCenter
    .VerticalAlignment = xlBottom
    .WrapText = False
    .Orientation = 0
    .AddIndent = False
    .IndentLevel = 0
    .ShrinkToFit = False
    .ReadingOrder = xlContext
    .MergeCells = False
End With

```

### 'Artikel 1

```
Dim lr2 As Long
lr2 = Range("A" & Rows.Count).End(xlDown).Row
Range("A1").Select
ActiveCell.FormulaR1C1 = "Artikel 1"
Range("A1").Select
Selection.Font.Bold = True
Range("A2").Select
Application.CutCopyMode = False
With Selection.Validation
    .Delete
    .Add Type:=xlValidateList, AlertStyle:=xlValidAlertStop, Operator:=
```

-

```
xlBetween, Formula1:="'Resultatkarta'!$A$2:A" & lr2
    .IgnoreBlank = True
    .InCellDropdown = True
    .InputTitle = ""
    .ErrorTitle = ""
    .InputMessage = ""
    .ErrorMessage = ""
    .ShowInput = True
    .ShowError = True
End With
```

### 'Artikel 2

```
Range("B1").Select
ActiveCell.FormulaR1C1 = "Artikel 2"
Range("B1").Select
Selection.Font.Bold = True
Range("B2").Select
ActiveCell.Formula2R1C1 = "=UNIQUE(Tabell2[ [#Data], [Artikel] ])"
Range("B3").Select
```

### 'Antal samband

```
Dim lrow As Long
lrow = Cells(Rows.Count, 2).End(xlUp).Row
Range("C1").Select
ActiveCell.FormulaR1C1 = "Antal samband"
Range("C1").Select
Selection.Font.Bold = True
Columns("C:C").EntireColumn.AutoFit
Range("C2").Select
ActiveCell.Formula2R1C1 = _
    "=SUM(INDIRECT("'"&R2C1&""') * INDIRECT("'"&RC[-1] & ""'""))"
Range("C2").Select
Selection.Autofill Destination:=Range("C2:C" & lrow)
```

### 'Stöd

```
Range("D1").Select
ActiveCell.FormulaR1C1 = "Stöd"
Selection.Font.Bold = True
Range("D2").Select
ActiveCell.FormulaR1C1 = "=RC[-1]/Apriori!R1C2"
Range("D2").Select
Selection.Autofill Destination:=Range("D2:D" & lrow)
Range("D2:D" & lrow).Select
Selection.Style = "Percent"
Selection.NumberFormat = "0.00%"
```

```

'Tillförlitlighet
  Range("E1").Select
  ActiveCell.FormulaR1C1 = "Tillförlitlighet"
  Selection.Font.Bold = True
  Range("E2").Select
  Columns("E:E").EntireColumn.AutoFit
  ActiveCell.FormulaR1C1 = _
    "=RC[-2]/XLOOKUP(R2C1,'Filtrerad data'!C[-2],'Filtrerad data'!C[-
1])"
  Range("E2").Select
  Selection.Style = "Percent"
  Selection.NumberFormat = "0.00%"
  Selection.AutoFill Destination:=Range("E2:E" & lrow)

Application.ScreenUpdating = True
Application.DisplayStatusBar = True
Application.EnableEvents = True

End Sub

```

INSTITUTIONEN FÖR TEKNIKENS EKONOMI OCH ORGANISATION  
AVDELNINGEN FÖR SUPPLY AND OPERATIONS MANAGEMENT  
CHALMERS TEKNISKA HÖGSKOLA  
Göteborg, Sverige 2022



**CHALMERS**