

# Teoretiska frågor

1. Beskriv kort hur en relationsdatabas fungerar.

En relationsdatabas organiserar data i tabeller med rader och kolumner. Varje rad har en unik nyckel för identifiering. Nycklar kan användas för att sammanställa information från flera olika tabeller.

2. Vad menas med "CRUD" flödet?

Create, Read, Update, Delete. De fyra grundläggande funktionerna när man jobbar med en databashanterare. I SQL är de fyra: INSERT, SELECT, UPDATE, DELETE.

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

En left och inner join kan användas för att hämta information från två tabeller. En left join returnerar alla rader från tabell 1 och alla matchande rader från tabell 2. Medans en inner join returnerar endast de matchande rader från båda tabellerna.

4. Beskriv kort vad indexering i SQL innebär.

Indexering gör att man kan hämta information snabbare från databasen. Om det t.ex. är någon sökning som sker ofta så kan man skapa ett index för att det ska gå snabbare varje gång man gör den sökningen i databasen

5. Beskriv kort vad en vy i SQL är.

En vy är en tabell baserat på en query som inte lagras i databasen utan istället så läser den in queryn och hämtar infon från de tabeller den är byggd på varje gång den öppnas. Vyer används för att anpassa hur varje användare ser databasen och kan även fungera som en säkerhetsåtgärd om man vill att en/flera personer ska kunna komma åt en viss del av tabellerna men inte grundtabellen.

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

En lagrad procedur är en sparad SQL query som man kan köra om och om igen när man behöver den. Om man använder en query väldigt ofta så kan man alltså spara den och sedan kalla på den när den behövs.

# Rapport

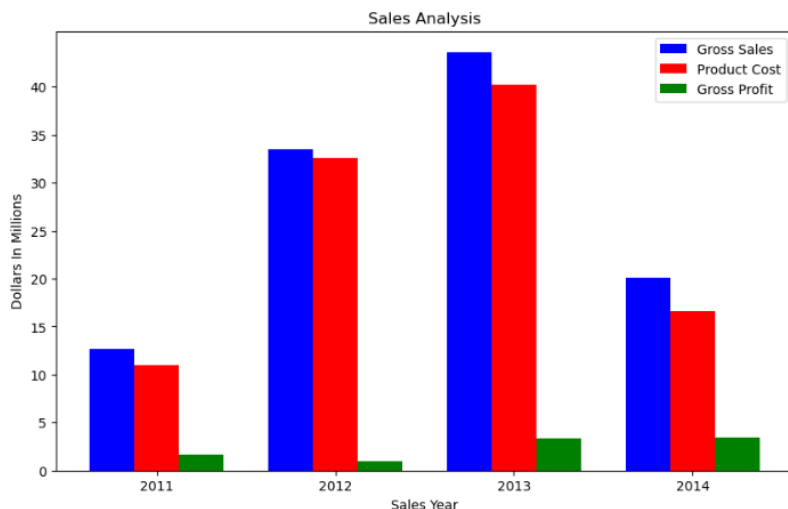
## 1. Deskriptiv sammanfattning

Detta är min deskriptiva sammanfattning av AdventureWorks2022. Jag har valt ut olika dataquerys som jag anser ger en bra överblick vad som finns i databasen och hur man kan använda datan.

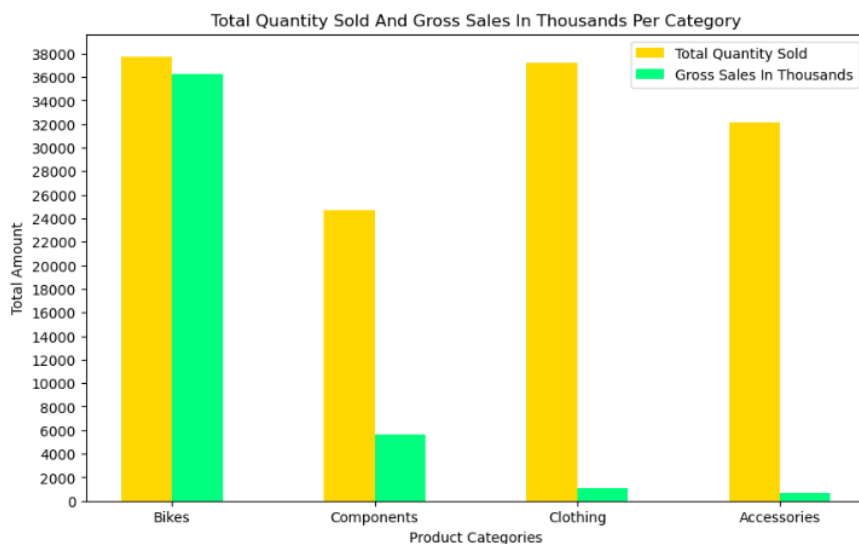
Inledningsvis har jag granskat de tillgängliga scheman och valt en strategi baserad på detta. Med medvetenhet om att försäljning är en nyckelaspekt för företag, har jag börjat genom att fokusera på Sales.

Här ser vi bruttoförsäljningen, produktionskostnad och bruttovinst. Det ger en ganska bra överblick på hur det gått för företaget. Som vi ser var 2013 ett toppen år i volym men vi hade fortfarande nästan samma vinst som 2014. Men det som inte syns här är att vi har ingen komplett data så 2011 startar sista maj medans 2014 slutar sista juni, vilket självklart påverkar resultatet.

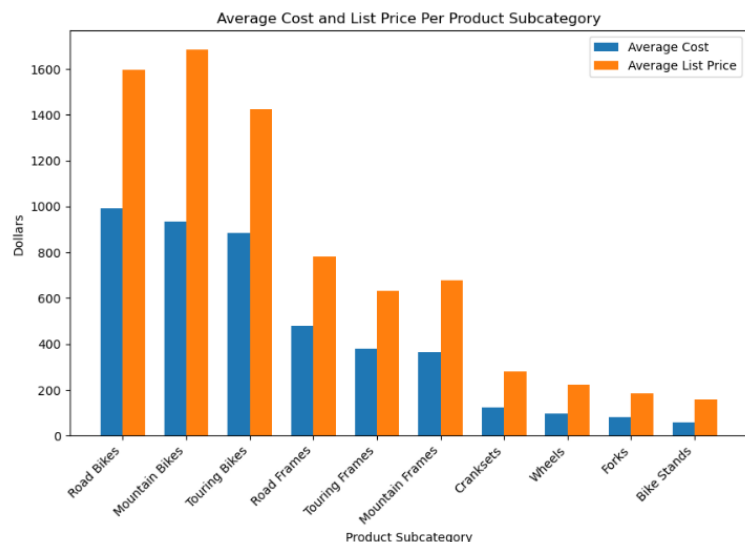
Gross profits är brutto vinsten, alltså endast försäljningen - produktionskostnader, ett vanligt företag brukar ligga med bruttovinst på cirka 30% av totala försäljningen vilket vi ser att dem inte gör här. Jag kan tänka mig att företaget knappt går runt efter alla andra kostnader som löner, lån osv.



Bortsett från bara försäljningssiffror är det viktigt att förstå vilka produkter företaget faktiskt säljer. Här har jag valt ett slumpmässigt år (2013) för att se vilka produktkategorier som finns. Vi kan se att det är cyklar som står för den stora bruttoförsäljningen följt av komponenter, kläder och accessoarer. Trots att kläder och cyklar har liknande sålda volymer, är det uppenbart att cyklarnas bruttoförsäljning är betydligt högre än klädernas.



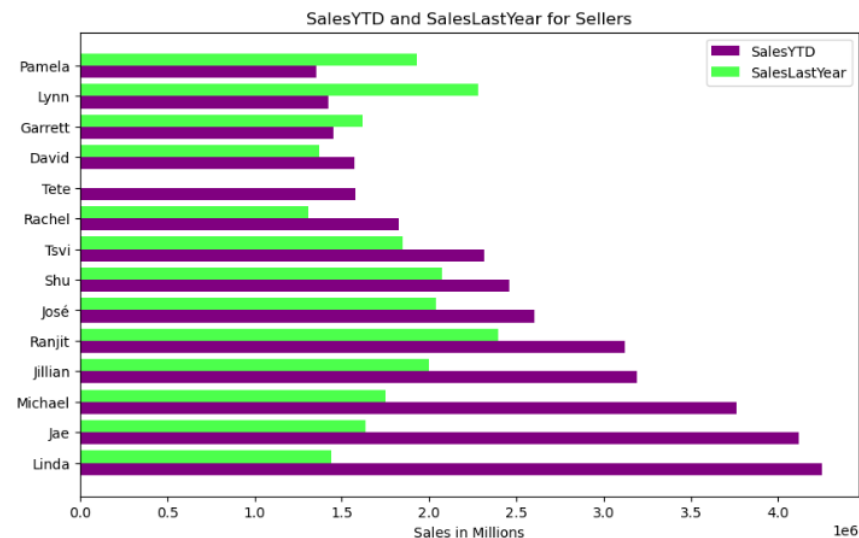
För att få en ännu bättre förståelse över hur sortimentet ser ut dyker vi in i de olika underkategorierna. Vi kan se hur de tre dyraste produkterna är de olika modellerna på cyklar, vilket förklarar deras betydande andel av bruttoförsäljningen. Noterbart är att genomsnittliga marginalen för alla produkter är avsevärt bättre än vad försäljningssiffrorna visade.

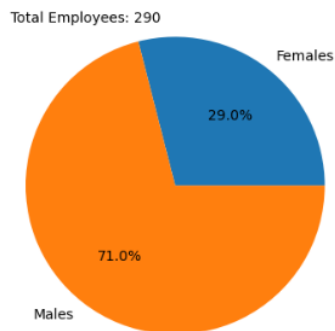


Datan ger en klar överblick över produkterna och deras variationer. Till exempel säljer de 43 olika modeller av "Road Bikes", men endast en modell av cykelställ. En intressant aspekt för produktionsavdelningen att överväga är varför Road Bikes tillverkas för mer pengar trots att deras listpris är lägre jämfört med Mountainbikes. Finns det möjligheter att optimera produktionen och minska tillverkningskostnaderna för Road Bikes? Det är en fråga värd att utforska för att förbättra lönsamheten.

Nu har vi fått en ganska bra överblick över försäljningen, produkterna och en liten del av produktionen. Men vart någonstans i världen tar cyklarna vägen? Databasen har mycket info om kunderna. Vi kan se att AdventureWorks har nästan 20 tusen kunder och hälften kommer från Nordamerika, näst störst andel är europeiska kunder och minst andel kunder kommer från Oceanien. Bland dessa kunder identifierar vi 1 336 företagskunder och 18 484 enskilda kunder

Men för att produkterna ska komma ut till kunder måste vi ha säljare. Genom databasen kan vi identifiera vilka säljare vi har i företaget och deras resultat. Det framgår att Nordamerika är där företaget har flest försäljare, vilket korrelerar till den stora kundbasen som vi tidigare såg finns där. Det är också från Nordamerika som årets bästa säljare Linda kommer ifrån, medans förra årets bästa säljare Lynn gjorde ett kanonår förra året i Australien. (Eftersom vi vet att databasen inte är komplett så vet vi inte vilka årtal som siffrorna är ifrån, men antar 2013 - 2014).

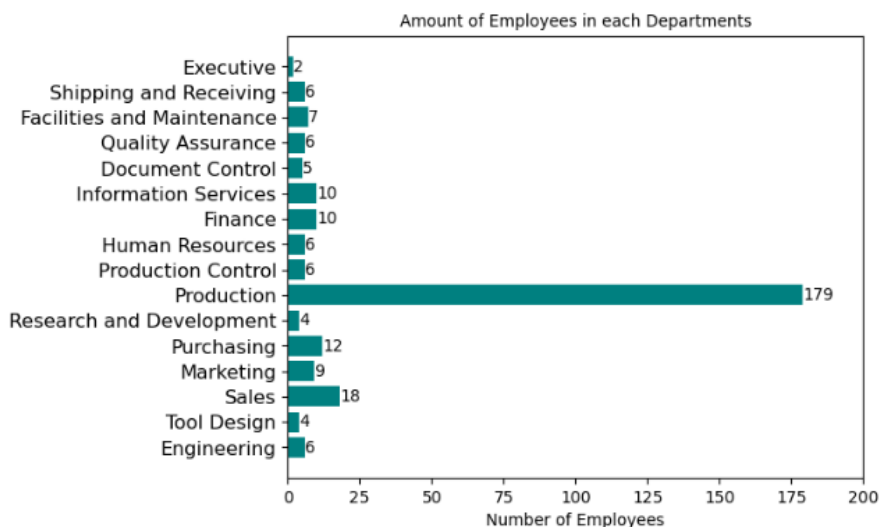




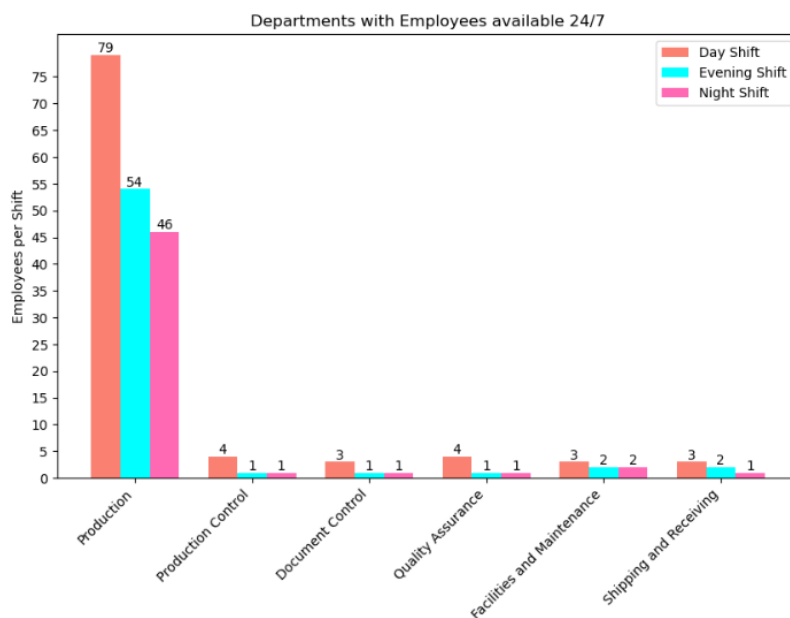
Efter att ha granskat säljarna börjar vi nu komma in mer på företagets personal. Där ser vi att företaget har totalt 290 anställda, varav 71% är män och 29% är kvinnor. Det ger oss en viktig inblick i könsfördelningen bland personalen

För att få en

ännu bättre överblick över personalen så har jag extraherat från databasen alla olika avdelningar och hur många som jobbar på varje avdelning. Som vi ser är produktionen den största med 179 anställda, följt av försäljningen med 18 anställda. Ledningen är minsta avdelningen med 2 anställda.



När man pratar om personal kan det också vara bra att få en insikt i hur arbetstiderna ser ut på arbetsplatsen. Vi kan där se att produktionen, som förväntat, kör dygnet runt. De har mest folk på plats alla tider på dygnet, men vi ser även att fem andra avdelningar har personal på plats dygnet runt. De ser ut som det är fyra stycken produktionsstödjande avdelningar och sedan leverans/ mottagningsavdelningen som alltid har någon på plats.



Denna deskriptiva sammanfattning av AdventureWorks2022 har gett en översikt över företagets försäljning, produkter, personal och arbetsstrukturer baserat på databasen.

## 2. Statistisk analys

I min statistiska analys utgår jag från att databasen är ett stickprov från årtalen 2011-2014. I analysen utgår jag också utifrån att datan är komplett för varje år (vilket vi egentligen kan se på datumen att den inte är) och har inte med den aspekten i min analys.

Jag har fått i uppdrag av chefen att se om man under den tiden kunde se någon skillnad i kasserade ordrar.

Först och främst började jag med att ta ut den informationen som behövdes ur databasen. I detta fall kunde all data hämtas ur Production Schemat. Här är datan som användes för analysen:

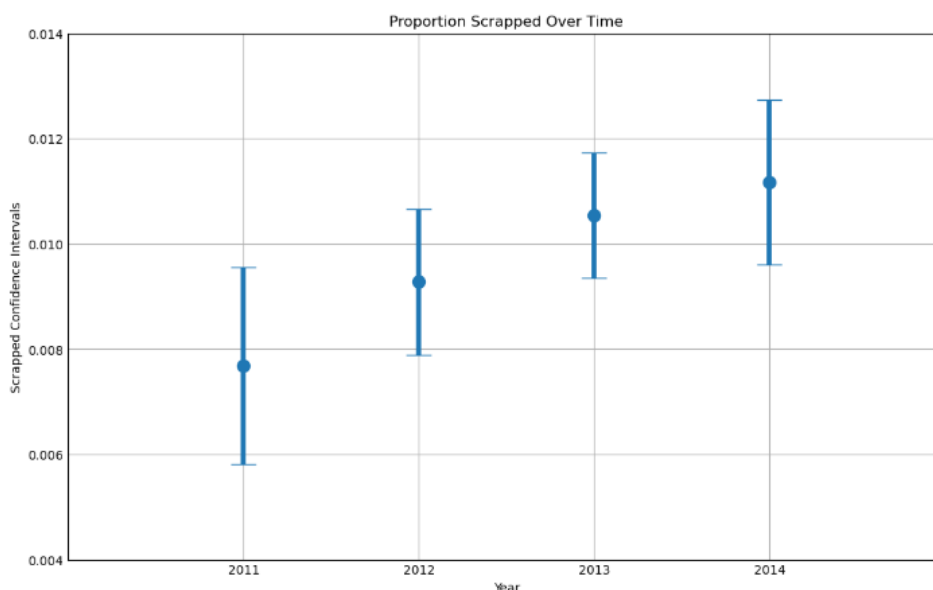
	ScrapYear	TotalWorkOrders	ScrappedWorkOrders
0	2011	8328	64
1	2012	18545	172
2	2013	28264	298
3	2014	17454	195

$p \pm z \sqrt{\frac{p(1-p)}{n}}$  Jag började med är att titta på om det fanns någon skillnad från år till år i andel kasserade ordrar. Jag använde mig av datan och räknade ut ett 95% konfidensintervall för varje år i stickprovet. Jag använde mig av formeln för enskilda stickprov med andelar.

Detta är resultatet som jag fick fram:

ScrapYear	TotalWorkOrders	ScrappedWorkOrders	ProportionScrapped	MarginOfError	LowerCI	UpperCI
2011	8328	64	0.007685	0.001876	0.005809	0.009560
2012	18545	172	0.009275	0.001380	0.007895	0.010654
2013	28264	298	0.010543	0.001191	0.009353	0.011734
2014	17454	195	0.011172	0.001559	0.009613	0.012732

För att få en klarare bild på de olika intervallen skapade jag denna graf:



Efter att ha sett resultatet från första delen av analysen valde jag att också titta på första och sista året, då deras konfidensintervall inte överlappade. Detta är också ett bredare tidsintervall och ger oss ett resultat om något skett över

längre tid. Jag tog då 2011:s kassaktionsandel och jämförde med 2014:s kassaktionsandel med hjälp av ett 95% konfidensintervall. Jag använde mig av formeln för två oberoende stickprov med andelar.

$$(p_1 - p_2) \pm z \cdot \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

Här är resultatet:

Proportion 2011: 0.00768  
 Proportion 2014: 0.01117  
 95% Confidence Interval for Difference: (0.00105, 0.00593)

Sedan för att få en inblick i vad som kan förbättras tog jag fram de olika kasseringsanledningarna från 2011 och 2014:

ScrapReasonID	ScrapYear	ScrapReasonName	TotalScrappedPerYear	ScrappedCount	ProportionOfScrappedReasons	ProportionScrapped	
0	1	2011	Brake assembly not as ordered	64	8	0.1250	0.125000
1	13	2011	Thermoform temperature too low	64	6	0.0938	0.093750
2	11	2011	Stress test failed	64	6	0.0938	0.093750
3	14	2011	Trim length too long	64	4	0.0625	0.062500
4	16	2011	Wheel misaligned	64	4	0.0625	0.062500
5	4	2011	Drill pattern incorrect	64	4	0.0625	0.062500
6	9	2011	Primer process failed	64	4	0.0625	0.062500
7	3	2011	Gouge in metal	64	4	0.0625	0.062500
8	12	2011	Thermoform temperature too high	64	3	0.0469	0.046875
9	6	2011	Drill size too small	64	3	0.0469	0.046875
10	7	2011	Handling damage	64	3	0.0469	0.046875
11	15	2011	Trim length too short	64	3	0.0469	0.046875
12	10	2011	Seat assembly not as ordered	64	3	0.0469	0.046875
13	2	2011	Color incorrect	64	3	0.0469	0.046875
14	5	2011	Drill size too large	64	3	0.0469	0.046875
15	8	2011	Paint process failed	64	3	0.0469	0.046875

ScrapReasonID	ScrapYear	ScrapReasonName	TotalScrappedPerYear	ScrappedCount	ProportionOfScrappedReasons	
0	13	2014	Thermoform temperature too low	195	19	0.0974
1	16	2014	Wheel misaligned	195	19	0.0974
2	15	2014	Trim length too short	195	17	0.0872
3	14	2014	Trim length too long	195	16	0.0821
4	11	2014	Stress test failed	195	15	0.0769
5	9	2014	Primer process failed	195	13	0.0667
6	3	2014	Gouge in metal	195	13	0.0667
7	6	2014	Drill size too small	195	12	0.0615
8	1	2014	Brake assembly not as ordered	195	10	0.0513
9	10	2014	Seat assembly not as ordered	195	10	0.0513
10	12	2014	Thermoform temperature too high	195	10	0.0513
11	2	2014	Color incorrect	195	10	0.0513
12	8	2014	Paint process failed	195	9	0.0462
13	7	2014	Handling damage	195	8	0.0410
14	4	2014	Drill pattern incorrect	195	7	0.0359
15	5	2014	Drill size too large	195	7	0.0359

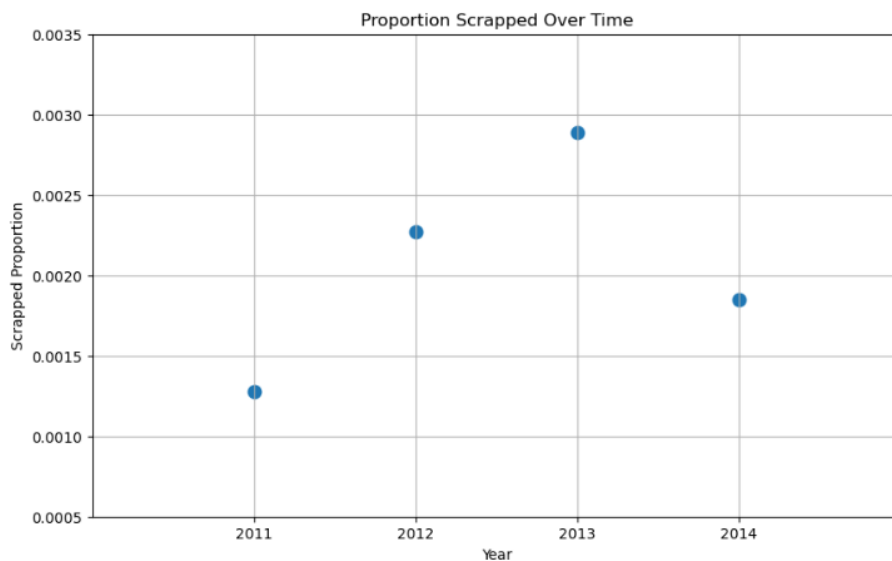
Jag gjorde sedan en jämförelse mellan 2011 och 2014 om man från 2014 skulle tagit bort "trim length" kasseringarna som stod för en stor del av kasseringarna:

```
Proportion 2011: 0.00768
Proportion 2014: 0.00928
95% Confidence Interval for Difference: (-0.00076, 0.00395)
```

För att se om den positiva trenden och ökning i kasserade ordrar också finns bland kasserade produkter valde använda jag mig av ännu en query och hämtade ut denna data från databasen:

	ScrapYear	TotalProducts	ScrappedProducts	Proportion Scrapped
0	2011	376034	483	0.00128
1	2012	1303470	2961	0.00227
2	2013	1900212	5486	0.00289
3	2014	928005	1721	0.00185

Till sist för att bättre visualisera tabellen gjorde jag en graf med produkt kasseringsdatan:



### 3. Slutsatser och rekommendationer

Det vi kan se i analysen är att varje år från 2011 – 2014 ökar andelen kasserade ordrar. Men om man tittar på konfidensintervallen så överlappar dem alltid varandra från år till år vilket tyder på att det inte är en signifikant skillnad mellan åren. Konfidensintervallen för 2011 till 2013 överlappar allihopa varandra, samma sak om man tittar på 2012 till 2014 så överlappar alla konfidensintervall varandra. Dock om man kolla på starten och slutet, det vill säga 2011 och 2014 så överlappar inte de två konfidensintervallen varandra. Detta tyder på en signifikant skillnad i andelen kasserade ordrar från 2011 till 2014. I detta fall är det så att andelen kasserade ordrar har ökat signifikant.

I resultatet ser vi också grafen där man tydligt kan se ett positivt mönster som stärker resultatet för att kasseringsandelen har ökat under åren.

Som vi ser på konfidensintervalls jämförelsen mellan år 2011 och 2014 och så blev resultatet (0,00105, 0,00593). Eftersom intervallet inte inkluderar noll kan vi säga att skillnaden är signifikant vilket stärker vårt antagande från tidigare resultat.

Självklart brukar större produktion och fler produkter betyda större kassering, men om man tittar på siffrorna så var 2013 det största året med mest produktion och ändå ökade man kasserade ordrar när man gick ner i produktion 2014. För att fördjupa analysen har jag därför även valt att tittat på olika kasseringsanledningar för att se vad som skiljer sig. Jag har valt ut tre observationer:

Första observationen är att "Thermoform temperature too low" är en den största kasseringsanledningen 2014 och även den kasseringsanledning som alltid dyker upp i topp fem kasseringsanledningar varje år från 2011 – 2014. Det kan vara något att lägga extra resurser på att förbättra för att få bort dessa temperaturproblem.

Andra observationen som sticker ut i kasseringsanledningarna är "Brake assembly not as ordered" och "Seat assembly not as ordered". Dessa två anledningar verkar relatera till leverantörsproblem, alltså inget som man själv kan påverka i fabriken. Slår man ihop dessa två kasseringsanledningarna år 2014 så blir det den största kasseringsanledningen och slår temperatursproblemet med en kassering. Alltså har man några av de största kasseringsanledningarna inte hos sig själva, utan hos leverantörer. Därför skulle rekommendationen vara att se över leverantörerna för både bromsar och säten.

Sista observationen från kasseringsanledningarna är "Trim length too short" och "Trim length too long" år 2014. Detta verkar vara någon sorts komponent eller "trim" som kapas i fabriken. Om detta görs manuellt kan det vara värt att köpa en maskin för detta eller liknande. Skulle man eliminera dessa två kasseringsanledningar från den totala kasseringsandelen för 2014 så skulle man som vi ser från resultatet (-0.00076, 0.00395) få ett intervall som innehåller noll vilket betyder att vi inte skulle kunna säga att det är en signifikant skillnad mellan 2011 och 2014.

Slutsatsen är att det finns en skillnad i kasserade ordrar i stickprovet. Skillnaden är en ökning i kasserade ordrar, även om man inte kan säga att det finns en signifikant skillnad från år till år, så är det en signifikant ökning om man jämför från början till och slut. Rekommendationen är att noga granska de olika kasseringsanledningarna, särskilt eftersom det finns flera framträdande faktorer som sticker ut. Eftersom uppdraget är fokuserat på kasserade ordrar och inte kasserade produkter har jag inte lagt mycket vikt på just individuella produkter. När man ser över kasseringsanledningar kan det därför vara bra att titta närmare på om det finns vissa kasseringsanledningar som påverkar fler produkter än andra. Som vi kan se i det sista i resultatet, en plot för att kassering av produkter genom åren, visar den inte samma positiva trend som kassering av ordrar visade.



#### **4. Executive summary - Kasserade Ordrar**

Analysen visar en ökning av kasserade ordrar från 2011 till 2014. Trots att konfidensintervallen mellan enskilda år inte är signifikanta, visar jämförelsen mellan 2011 och 2014 en betydande ökning. Grafiska representationen bekräftar en positiv trend. Rekommenderas att ta en närmare titt på de olika kassaktionsanledningarna mellan åren och baserat på min analys skulle jag lägga fokus på temperaturproblem, se över leverantörerna för bromsar / säten och se över trimkapningen som är stora kassaktionsanledningar. Samma positiva trend som fanns bland kasserade ordrar visades inte i kasserade produkter.

#### **5. Datum för muntlig presentation: 10:00 4/1 - 24**

### **Självutvärdering**

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

Jag skulle säga att största utmaningen har varit att välja vad man ska ha med och inte ha med. Hur man ska gå tillväga när man inte är helt kunnig inom området. Men var intressant och väldigt lärorikt.

När det kommer till analysen så tycker jag samma där att det var svårt att välja något. Eftersom databasen är väldigt begränsad var det många analyser som inte gav några bra resultat eftersom det inte fanns tillräckligt med data. Till slut valde jag något som jag tyckte företaget kunde dra nytta av.

Jag förstod inte riktigt Executive summary och exakt vad den skulle summera. Om chefen vill ha den så ser jag ingen anledning att ha med summering av hela databasen, då chefen förmodligen redan vet om det mesta som finns där. Därför gjorde jag min executive summary på analysen då jag anser det är förmodligen vad chefen skulle vilja ha info om i denna situation.

Blev lite fler sidor än tänkt men jag använde mig också av väldigt mycket visualiseringar.

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

Jag anser att jag ska ha VG då jag tycker jag har visat att jag kan ta fram resultat och komma med förklaringar till varför datan är relevant. Jag har gjort en fördjupad analys där jag tittar på olika delar av datan och drar slutsatser därifrån.

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

Välj bara nånting och fortsätt på det. Tänk inte så mycket nu när det är en så fri uppgift. Tror jag gjort analys på minst 5 olika saker under denna uppgift men tyckte aldrig att jag kunde få tillräckligt med bra resultat.