

Teoretiska frågor

Beskriv kort hur en relationsdatabas fungerar

En relationsdatabas består av en mängd tabeller med data som relaterar till varandra genom *keys*, värden i tabellerna som är gemensamma.

Vad menas med "CRUD"-flödet?

CRUD står för "Create, Read, Update, Delete" och är de fyra grundläggande aktiviteter som utgör arbetet med en databas.

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

En left join väljer all data i den "vänstra" tabellen, och sammanfogar den med den data i den "högra" tabellen som har värden som överensstämmer med värden i den vänstra.

En inner join väljer ut den data som har värden som finns i båda tabellerna.

Joins används i relationsdatabaser för att sammanfoga olika tabeller.

Beskriv kort vad indexering i SQL innebär.

En indexerad kolumn går snabbare att söka i eftersom alla värden sparats i ordning i ett separat index. Detta gör att det som sagt går snabbare att söka, men i gengäld går det långsammare att spara ny data i tabellen. En avvägning behöver göras gällande vilka kolumner som ska indexeras.

Beskriv kort vad en vy i SQL är.

En vy är en virtuell tabell som definieras av en förfrågan. En vy är bra att använda om man till exempel finner sig själv göra samma komplexa förfrågningar ofta, eller vill ha kontroll över hur olika användare ser datan i databasen.

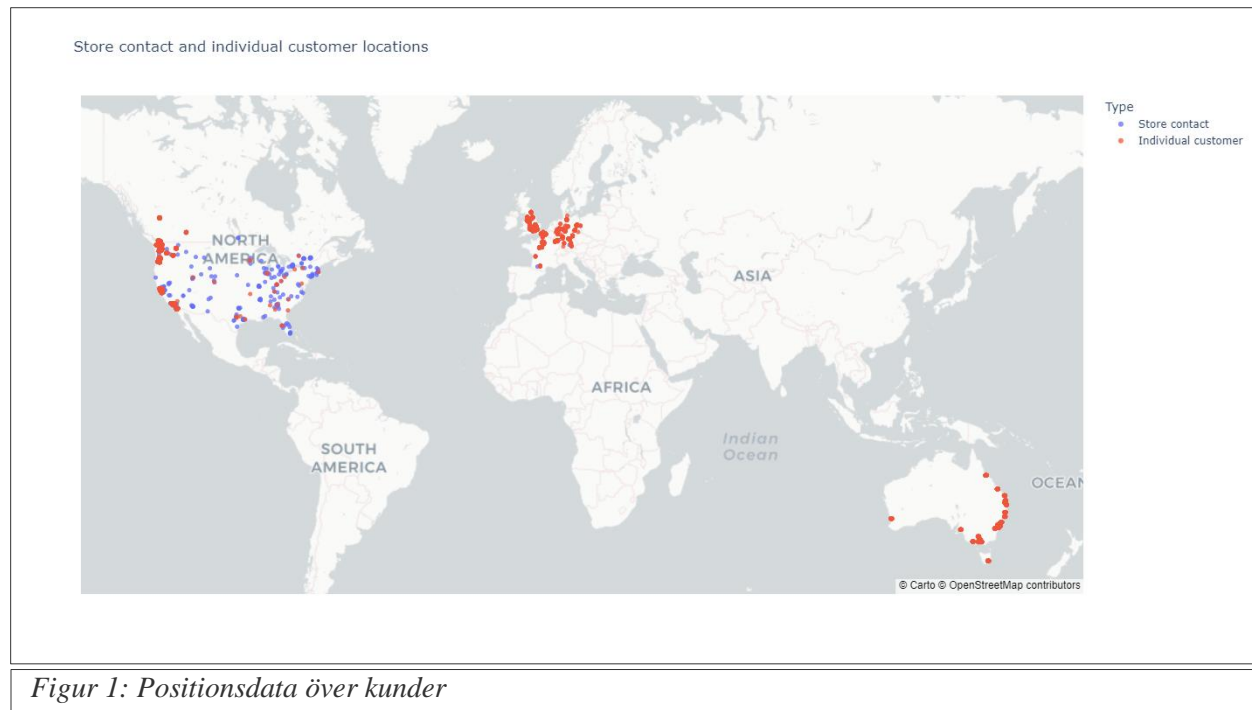
Beskriv kort vad en lagrad procedur i SQL är.

En lagrad procedur är ungefär som en funktion. Den kan ta parametrar som input och ge tillbaka andra parametrar som output. En lagrad procedur kan också göra förändringar i databasen. Bland fördelarna med lagrade procedurer kan nämnas sparad dataanvändning och ökad säkerhet.

Programmering och rapport

AdventureWorks

AdventureWorks är ett företag som tillverkar och handlar med framförallt cyklar och cykeltillbehör. Företaget har 290 anställda och är baserat i USA. Företaget säljer både till andra företag och privatpersoner, där företagen också är USA-baserade medan privatkunderna även finns i Europa och i Australien.

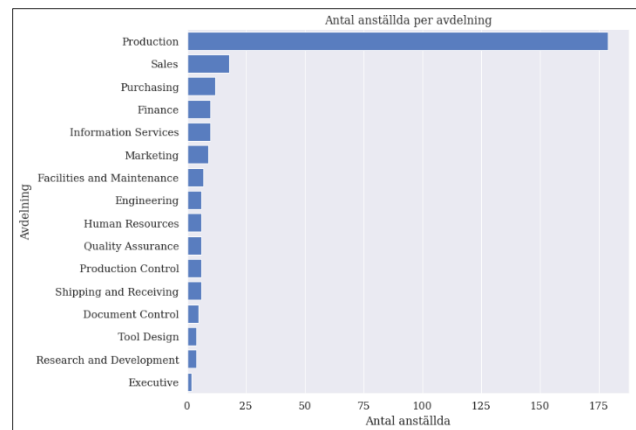


Figur 1: Positionsdata över kunder

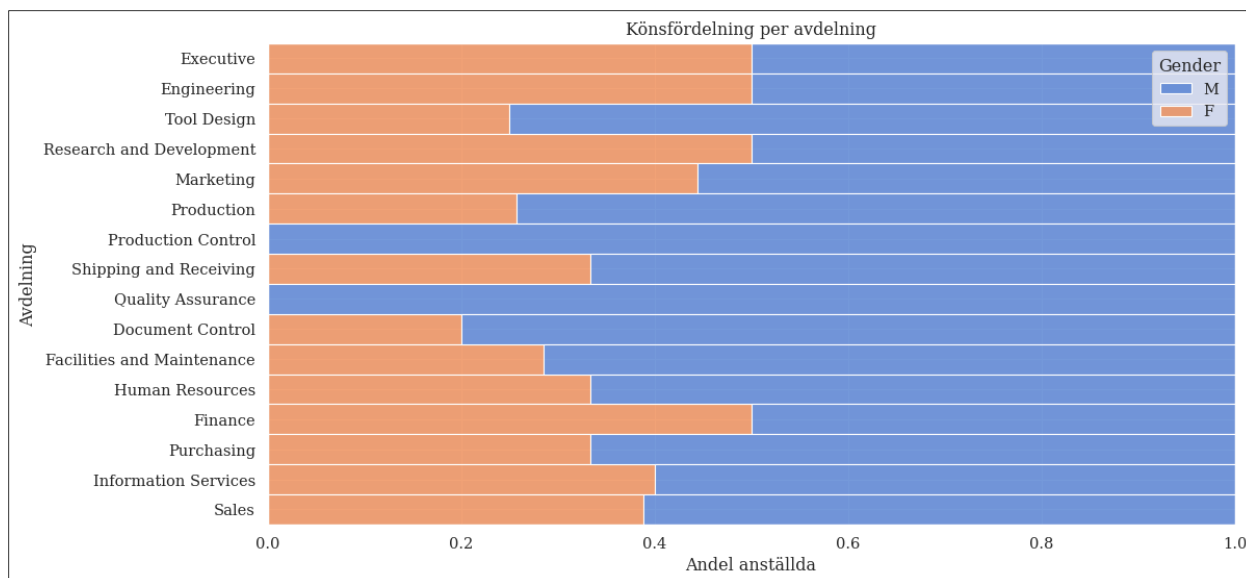
Anställda

Av de anställda är 206 män och 84 kvinnor, fördelade över 16 avdelningar. Avdelningen med störst andel anställda är produktionsavdelningen med 179 av de 290 anställda.

Med tanke på den låga andelen kvinnor bland de anställda kanske det inte är så förvånande att ingen avdelning har en majoritet kvinnor. Fördelningen per avdelning ser ut som i Figur 3.



Figur 2: Antal anställda per avdelning



Figur 3: Könsfördelning per avdelning

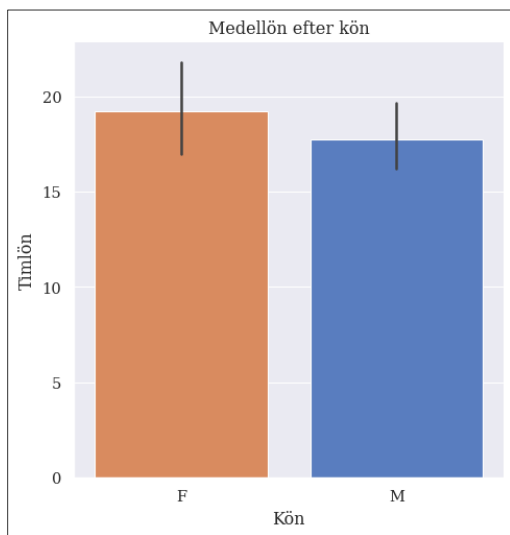
Analys

Löneskillnader mellan könen

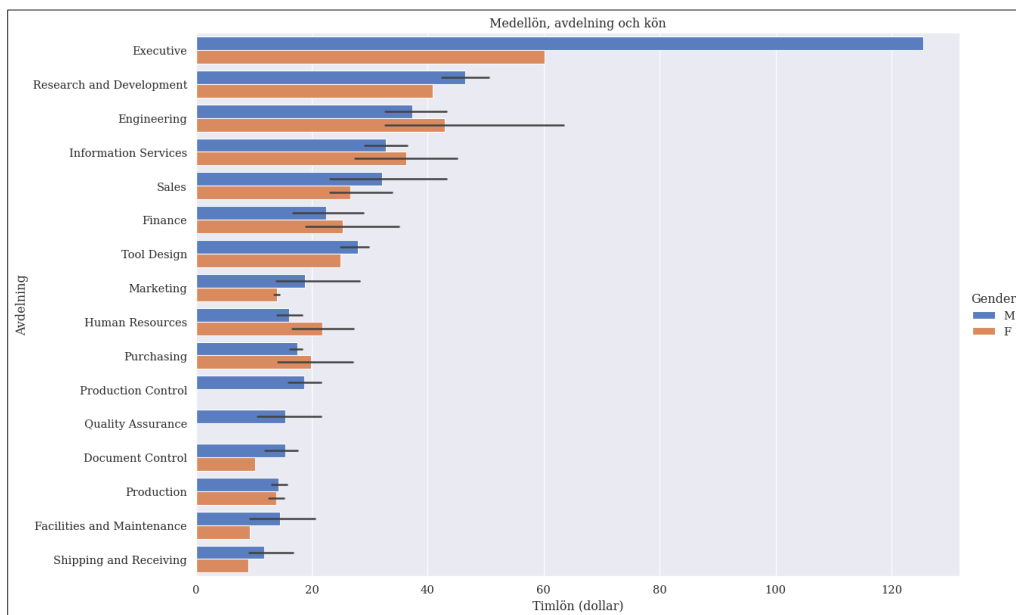
I genomsnitt verkar kvinnliga anställda tjäna något mer i timmen än manliga, men skillnaden är inte statistiskt säkerställd. (Se kod för t-test.)

Löneskillnader mellan avdelningar

Skillnaderna är större mellan avdelningarna (se Figur 5). Mest tjänar chefer, forskare och ingenjörer. I botten hittar vi lagerarbetare, städare och vaktmästare.



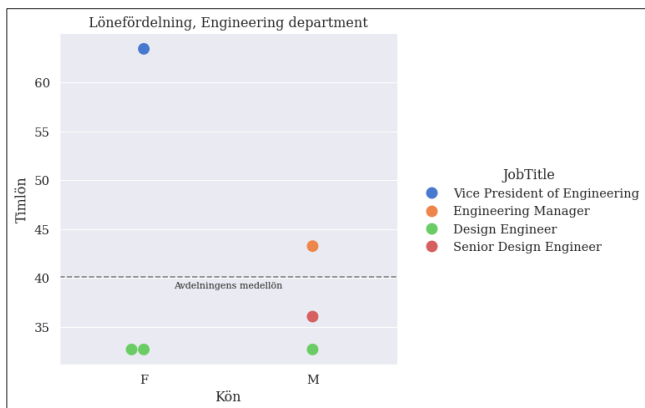
Figur 4: Medellön efter kön



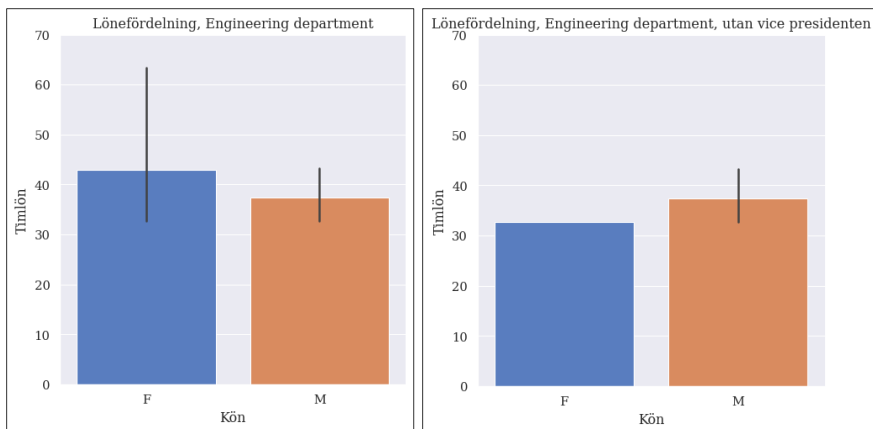
Figur 5: Medellön per avdelning och kön

Eftersom alla avdelningar förutom produktionsavdelningen är så pass små (mellan 2 och 18 personer) kan enskilda personers löner göra stor skillnad för medellönen på avdelningen. Det stora konfidensintervallet för *Engineering*-avdelningen i Figur 5 kan till exempel undersökas vidare (se figur 6-7).

I figur 6 ser vi hur avdelningens vice president, som är kvinna, tjänar långt mer än någon annan på avdelningen. Om vi visualiserar avdelningens medellöner utan vice presidentens inräknad (figur 7) ser vi att det blir en stor skillnad och att de två kvarvarande kvinnornas medellön är lägre än de tre männens.



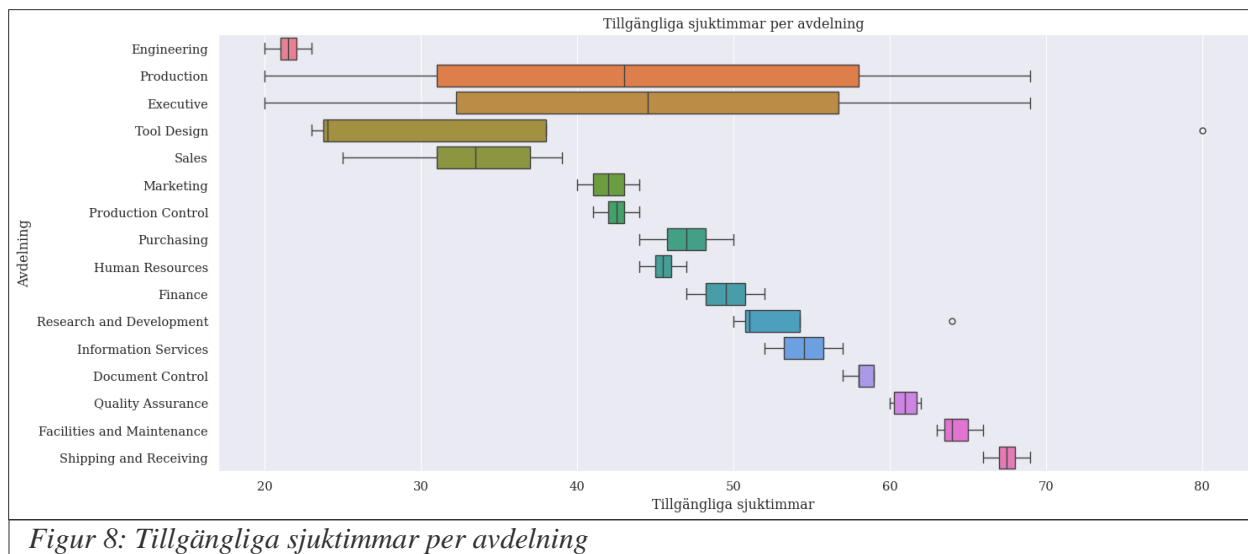
Figur 6: Lönefördelning, Engineering-avdelningen



Figur 7: Lönefördelning Engineering-avdelningen med och utan VP

Sjuktimmar

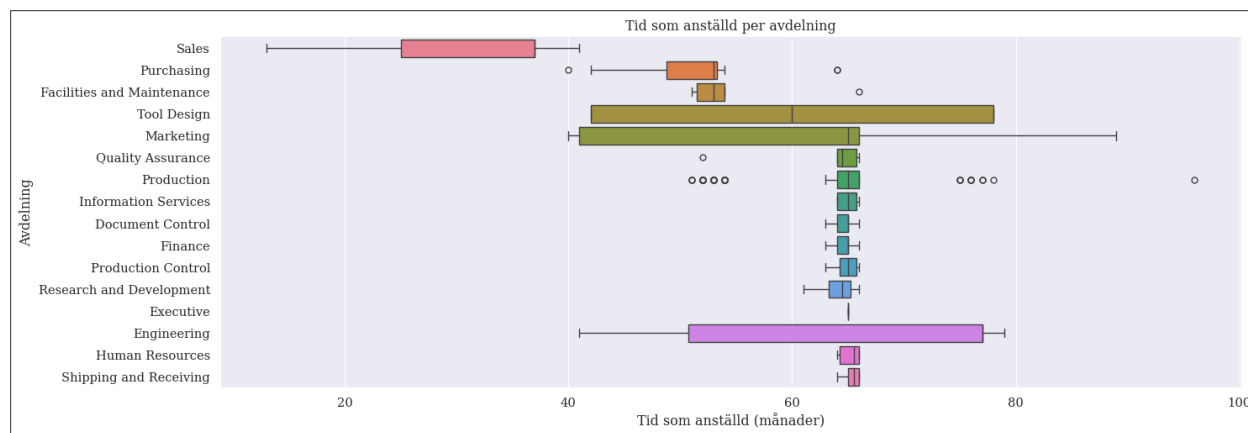
Jag har också tittat på mängden använda sjuktimmar för respektive avdelning. Enligt amerikansk lagstiftning får en anställd en sjuktimme per 30:e arbetad timme. Jag testade att göra en uppskattning på hur många sjuktimmar varje anställd bör ha samlat på sig genom att räkna antalet veckor hen har arbetat på företaget * 1.33 (för 40 timmars arbetsvecka) men de siffrorna blev alldeles för stora. Kanske är det en artefakt av att databasen är påhittad? Istället nöjer jag mig med att titta på mängden tillgängliga sjuktimmar per avdelning (figur 8), kombinerat med en snabb koll på hur länge personerna på avdelningen har arbetat där.



Figur 8: Tillgängliga sjuktimmar per avdelning

Även här sticker *Engineering*-avdelningen ut med väldigt få tillgängliga sjuktimmar. Spridningen på anställningstid (figur 9) är relativt stor på avdelningen, men motsvaras ungefär av *Marketing* i det

avseendet. Skillnaderna i sjuktimmar mellan *Engineering* och *Marketing* är dock avsevärd. Det är svårt att veta exakt varför det är så, men skulle kunna vara värt att undersöka vidare, till exempel genom att intervjua de anställda på avdelningen.

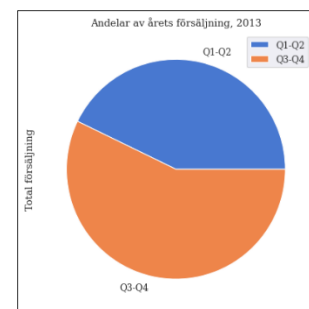


Figur 9: Anställningstid per avdelning

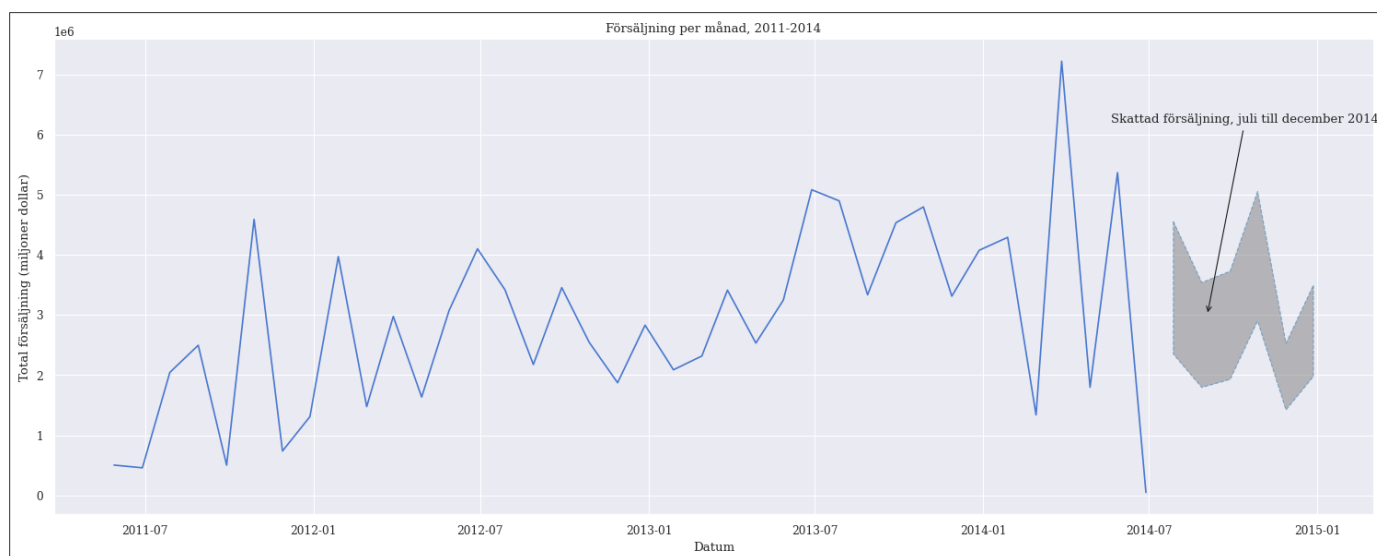
Ekonomi

AdventureWorks omsatte som mest ca 44 miljoner dollar 2013. Kompletta försäljningssiffror finns inte för 2014, men siffrorna för första halvåret pekar på en ökning jämför med 2013 då andra halvåret är starkare (figur 10).

Jag har gjort en skattning av försäljningen under andra halvan av 2014 (figur 11). Eftersom den är baserad på försäljningen under motsvarande månader 2011-2013 kan man se att den lägre delen av intervallet ligger på ungefär 2011 års nivåer och den högre på 2013 års nivåer. Här skulle nog skattningen kunna viktas något, med större vikt på de senare årens försäljning.



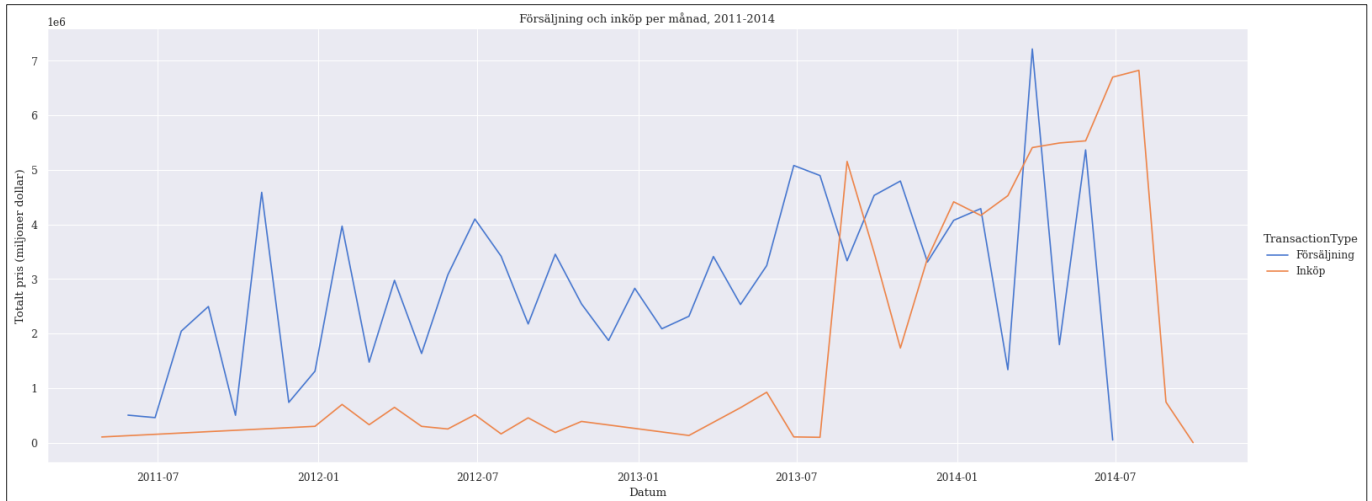
Figur 10: Andelar av försäljning 2013



Figur 11: Månadsvis försäljning 2011-2014 med skattning av sista halvåret 2014

Försäljning och inköp

Om man jämför försäljningssiffrorna med summan av inköpta varor över samma tid (figur 12) verkar det som att kostnaderna för inköp ökar kraftigt i mitten av 2013, sjunker något för att sedan fortsätta hålla sig på en nivå långt över tidigare.



Figur 12: Försäljning och inköp månadsvis 2011-2014

Det verkar som att det är ett antal komponenter som börjar köpas in i väldigt mycket större antal i augusti 2013. Om det skriver jag kanske mer imorgon!

Rekommendationer

Antalet tillgängliga sjuktimmar hos de anställda i *Engineering*-avdelningen är avsevärt lägre än för andra avdelningar. Koppla in HR.

Kostnaderna för inköp har ökat kraftigt och riskerar att passera intäkterna från försäljningen. Ett antal komponenter har börjar köpas in allt oftare. Kolla med produktionsavdelningen hur detta kommer sig.

Executive summary

Datum för muntlig presentation

2/1 kl. 10.00

Egen reflektion

Utmaningar under arbetet

Den absolut största utmaningen har varit att avgränsa. Det finns helt enkelt för mycket i databasen att undersöka och varje gång jag lyckats svara på en fråga jag ställt har den väckt flera nya frågor. Jag vill nog göra det lite svårt för mig själv också, vilket till stor del handlar om att utmana mig själv och utvecklas, men det gäller för mig att hitta en rimlig ribba att lägga för mig själv.

Det har också varit svårt att sätta mig in i hur jag ska skriva en sådan här rapport. Jag har väldigt lite erfarenhet av den ekonomiska delen av världen och det har varit svårt att veta vad jag ska fokusera på.

Det har också varit svårt att göra rapporten snygg rent grafiskt med alla diagram...

Vilket betyg anser jag att jag ska ha?

Jag tycker att jag fyller kriteriet för VG i kursen.

Tips till mig själv i början av kursen

Jag skulle kunna tipsa mig själv om att välja ett spår och hålla mig till det, snarare än att hoppa mellan, som det blev, Human resources-schemat och de mer ekonomiska delarna. Nu blev arbetet lite spretigt och jag känner att jag kanske mest skrapat på ytan.