

## 1. Teoretiska Frågor

1. Beskriv kort hur en relationsdatabas fungerar.

2. Vad menas med "CRUD" flödet?

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

4. Beskriv kort vad indexering i SQL innebär.

5. Beskriv kort vad en vy i SQL är.

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

1. Relationsdatabas är en typ av databas, som används som en sammansättning som tillåter oss att identifiera och ha tillgång till en data i relation till en annan data i databasen. Oftast är relationsdatabas är upplagt eller organiserade i tabeller.

2. CRUD är en akronym som står för CREATE, READ, UPDATE och DELETE. Detta är fyra grundläggande operationer för att skapa och hantera beständiga dataelement.

3. Left Join – Returnerar alla rader från den första tabellen och matchande data från den andra tabellen. Left join används när man vill ha all data från första tabellen men också det gemensamma data från båda tabellerna till en och samma tabell.

Inner Join – Returnerar gemensamma data från båda tabellerna. Används när man vill använda matchande data från båda tabellerna till och samma tabell.

4. Indexering används för att hämta data från databasen snabbare än vanligtvis. Användare kan inte se indexen, de är bara till för att snabba på sökandet och queries.

5. Vy är en virtuell tabell över faktiska data, olika tabeller och output. Genom en vy ser vi bara resultatet av en query och inte själva koden i databasen, samt att innehållet av vy lagras heller inte i SQL Server databas.

6. Lagrad procedur i SQL är olika påståenden som lagras tillsammans i en databas. Man kan skapa en kod och spara det som en lagrad procedur, och sedan blir koden kompilerad och kan sedan återanvändas i upprepande gånger.

## 2. Programmerings uppgift och Rapport

1. Genom databasen AdventureWorks2022 får man ett översende om ett företag som tillverkar och säljer främst olika cykeldelar men även andra produkter. Men också detaljerad överblick såsom affärsenheter, personal, kunder, produkter, tillverkning, orderhantering, lager, försäljning och administration. Med hjälp av databasen har jag tagit fram lite olika information och data som kan vara intressant för en djupare inblick. Min huvudfokus och det jag tyckte var intressant att ta fram data ifrån var personalen (Human Resources), men också tagit fram lite data från produktion, försäljning och orderhantering.

Först ville jag se hur många anställda som fanns och veta könsfördelningen bland alla anställda.

Gender Employees		
0	F	84
1	M	206

Totala mängden anställda är 290 personer som vi kan se från tabellen om man räknar ihop summan, men främst ser vi att det är flera män (206) än kvinnor (84) som arbetar. Detta tog jag fram genom en aggregat funktion COUNT (\*) från tabellen Human Resources och specifikt gender för att beräkna fram hur anställda det är av varje kön.

Gender		Gender		MaritalStatus		TotalEmployees
0	F	Female		M	Married	49
1	F	Female		S	Single	35
2	M	Male		M	Married	97
3	M	Male		S	Single	109

Sedan ville jag ta reda på status av alla män och kvinnor som arbetar för att se vem som är singlar eller gifta och då gjorde jag förut genom COUNT (\*). Men jag ville också att bokstäverna F, M, S och M skulle stå som en text efteråt och detta gjorde jag genom en CASE Statement för till exempel F skulle stå som texten Female efteråt.



Jag gjorde även en graf så man kunde se skillnaden tydligare och då ser vi att bland kvinnor så är det flera som gifta än singlar, och bland män är det tvärtom då det är flera som är singlar än gifta.

BusinessEntityID	FirstName	LastName	JobTitle
0	1	Ken Sánchez	Chief Executive Officer
1	2	Terri Duffy	Vice President of Engineering
2	3	Roberto Tamburello	Engineering Manager
3	4	Rob Walters	Senior Tool Designer
4	5	Gail Erickson	Design Engineer
...	...	...	...
285	286	Lynn Tsoflias	Sales Representative
286	287	Amy Alberts	European Sales Manager
287	288	Rachel Valdez	Sales Representative
288	289	Jae Pak	Sales Representative
289	290	Ranjit Varkey Chudukatil	Sales Representative

290 rows × 4 columns

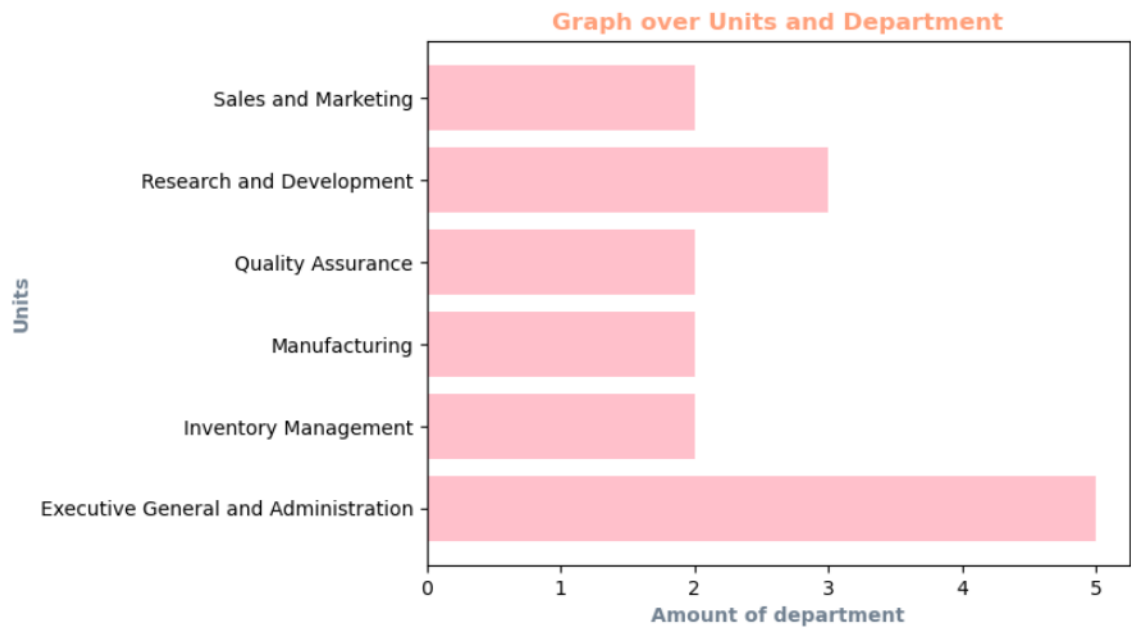
	BusinessEntityID	Department	DepartmentID	ShiftID	WorkHour
0	1	Executive	16	1	Day
1	2	Engineering	1	1	Day
2	3	Engineering	1	1	Day
3	4	Engineering	1	1	Day
4	4	Tool Design	2	1	Day
...	...	...	...	...	...
291	286	Sales	3	1	Day
292	287	Sales	3	1	Day
293	288	Sales	3	1	Day
294	289	Sales	3	1	Day
295	290	Sales	3	1	Day

296 rows × 5 columns

Dessa två tabeller tog jag fram då på första bilden så ville jag veta en mer detaljerad information från varje anställd då man ser deras namn och jobbtitel. Detta gjorde genom att jag gjorde en subquery av två tabeller för att få fram den informationen.

Medan på andra bilden så ville jag få fram vilket skift de arbetade det vill säga när de arbetade under dagen om det var dag, kväll eller natt. Detta tog jag fram genom inner join av tre tabeller för att kunna ta reda vilket skift varje anställd hade.

	DepartmentID	Name	GroupName
0	1	Engineering	Research and Development
1	2	Tool Design	Research and Development
2	3	Sales	Sales and Marketing
3	4	Marketing	Sales and Marketing
4	5	Purchasing	Inventory Management
5	6	Research and Development	Research and Development
6	7	Production	Manufacturing
7	8	Production Control	Manufacturing
8	9	Human Resources	Executive General and Administration
9	10	Finance	Executive General and Administration
10	11	Information Services	Executive General and Administration
11	12	Document Control	Quality Assurance
12	13	Quality Assurance	Quality Assurance
13	14	Facilities and Maintenance	Executive General and Administration
14	15	Shipping and Receiving	Inventory Management
15	16	Executive	Executive General and Administration



Under bild 1 ser vi totalt 16 avdelningar och varje avdelning tillhör en enhet och då ville jag beräkna på hur många avdelningar som fanns i varje enhet. I bild 2 ser man att den enhet som har flest avdelningar är executive general and administration som har fem avdelningar. Tabellen fick jag fram genom en att jag tog en COUNT (\*) av groupname för att kunna beräkna varje avdelning till enhet.

	PersonType	Name of person type	Number of people
0	IN	Individual aka Customers	18484
1	EM	Employee	273
2	SP	Sales Person	17
3	SC	Store Contact	753
4	VC	Vendor Contact	156
5	GC	General Contact	289

Från databasen tog jag även fram till hur många kunder det finns, anställda, säljare och lagerarbetare bland annat. Detta beräknade jag genom en COUNT (\*) och en CASE Statement för att få fram data.

## SQL - Kunskapskontroll

	ProductID	Name	Color	Prices
0	749	Road-150 Red, 62	Red	3578.27
1	750	Road-150 Red, 44	Red	3578.27
2	751	Road-150 Red, 48	Red	3578.27
3	752	Road-150 Red, 52	Red	3578.27
4	753	Road-150 Red, 56	Red	3578.27
...	...	...	...	...
499	509	HL Mountain Rim	None	0.00
500	510	LL Road Rim	None	0.00
501	511	ML Road Rim	None	0.00
502	512	HL Road Rim	None	0.00
503	513	Touring Rim	None	0.00

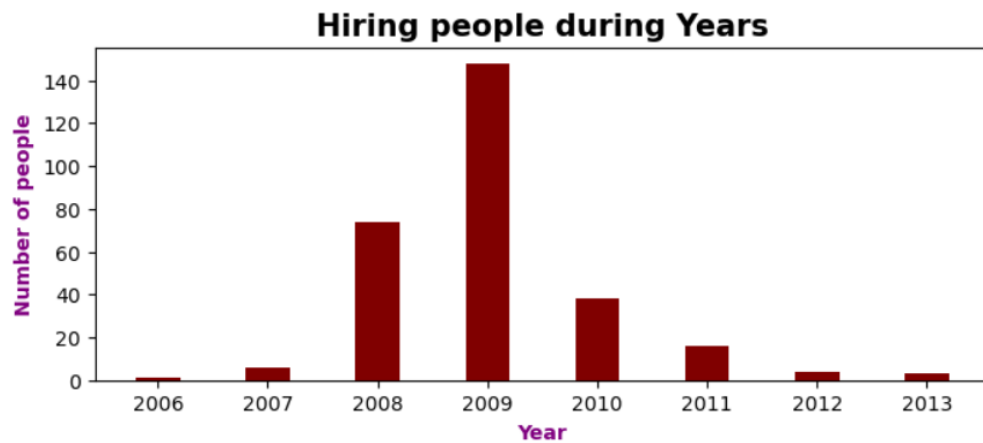
504 rows × 4 columns

	ProductID	Name	Color	Prices
0	749	Road-150 Red, 62	Red	3578.27
1	750	Road-150 Red, 44	Red	3578.27
2	751	Road-150 Red, 48	Red	3578.27
3	752	Road-150 Red, 52	Red	3578.27
4	753	Road-150 Red, 56	Red	3578.27
...	...	...	...	...
299	870	Water Bottle - 30 oz.	None	4.99
300	921	Mountain Tire Tube	None	4.99
301	923	Touring Tire Tube	None	4.99
302	922	Road Tire Tube	None	3.99
303	873	Patch Kit/8 Patches	None	2.29

304 rows × 4 columns

Av dessa liknande tabeller ville jag beräkna de produkter som säljs som den dyraste varan men då ser vi också från första bilden att det är många produkter som inte har något pris eller att det står noll. Därför filtrerade jag tabellen så att man bara ser produkter som har priset större än noll och då minskade raderna från 504 till 304 rader.

Jag beräknade detta genom att jag valde de kolumner jag ville ta reda från tabellen Production Product och sedan en aggregat funktion MAX () för ett maximumpris på varorna och en ORDER BY för att tabellen ska visas från en fallande ordning.



Av dessa grafer visas under vilket år och månad som det var högsta och lägsta personer som var anställda under olika år och månader. Dessa informationer fick jag genom aggregat funktion COUNT (\*) igen av tabellen Human Resources.

Dessa tabeller och grafer är en del av data jag fick fram från databasen. Mina queries har bestått väldigt mycket av COUNT (\*) för jag har velat beräkna på olika saker för att få fram den information som jag är intresserad av. Sammanfattningsvis av min data så visas det detalj hur mycket information jag har fått, främst från de anställda men även olika kategorier som produkter och kunder.

2. I denna tabell beräknade jag summan av medelvärdet av lead time för varje unit measure code. Detta tog jag fram genom en aggregat funktion SUM () på average lead time och en GROUP BY på unit measure code. Vi kan se resultatet att EA har högst average lead time från början till slutet att paketera produkter, och PAK har minst.

	UnitMeasureCode	Sum of average lead time
0	CAN	1192
1	CS	678
2	CTN	2960
3	DZ	122
4	EA	3746
5	GAL	165
6	PAK	84

	BusinessEntityID	JobTitle	BirthDate	Age
0	1	Chief Executive Officer	1969-01-29	55
1	2	Vice President of Engineering	1971-08-01	53
2	3	Engineering Manager	1974-11-12	50
3	4	Senior Tool Designer	1974-12-23	50
4	5	Design Engineer	1952-09-27	72
...	...	...	...	...
285	286	Sales Representative	1977-02-14	47
286	287	European Sales Manager	1957-09-20	67
287	288	Sales Representative	1975-07-09	49
288	289	Sales Representative	1968-03-17	56
289	290	Sales Representative	1975-09-30	49

290 rows × 4 columns

I denna tabell kan vi se när varje anställd har sin födelsedag men vi vet inte exakt hur gamla de är och då la jag till ålder som en kolumn genom en datediff för att få fram detta. Genom en datediff funktion kan vi få fram skillnaden från två datum som visas som år, då jag skrev från anställdas födelsedatum till dagens datum.

Av åldern bland alla så ser vi tydligt att det är mer äldre i medelåldern som arbetar och än yngre som är under 30 år.



Jag har beräknat medelvärdet från tabellen Purchasing Product Vendor av olika kolumner.

	Average standard price	Average of minimum order QTY	Average of maximum order QTY
0	34.6765	145	776

Här ser vi medelvärdet från standardpriset som försäljaren säljer, samt medelvärdet av de maxima och minimumorder kvantitet. Detta gjorde jag från aggregat funktionen AVG () på alla tre kolumner.

Jag ville sedan se om jag kunde få ett konfidensintervall av standardpriset av denna tabell då jag redan hade räknat ut medelvärdet. Jag ville uppskatta ett intervall med konfidensgraden 95 procent och se om standardpriset är högre än 40,0 och kan då i sådant fall kan försäljaren öka produkternas standardpris.

	ProductID	BusinessEntityID	AverageLeadTime	StandardPrice	LastReceiptCost	LastReceiptDate	MinOrderQty	MaxOrderQty	OnOrderQty	UnitMeasureCode
0	1	1580	17	47.87	50.2635	2011-08-29	1	5	3.0	CS
1	2	1688	19	39.92	41.9160	2011-08-29	1	5	3.0	CTN
2	4	1650	17	54.31	57.0255	2011-08-29	1	5	NaN	CTN
3	317	1578	19	28.17	29.5785	2011-08-29	100	1000	300.0	EA
4	317	1678	17	25.77	27.0585	2011-08-25	100	1000	NaN	EA

```
selling_price = df.StandardPrice
```

```
ci = stats.norm.interval(0.95, loc=np.mean(selling_price), scale=stats.sem(selling_price))
```

```
print(f"Confidence Interval for Average of Vendor's selling price: {ci}")
```

```
Confidence Interval for Average of Vendor's selling price: (33.44269460095867, 35.910392355563054)
```

Genom python kunde jag då få fram ett konfidensintervall för standardpriset av försäljaren. Av medelvärdet från standardpriset ser vi att det är 34,68 om man avrundar det, och man kollar på konfidensintervallet så ligger mellan ungefär 33,44 och 35,9 och kan vi konstatera att det inte är en jättestor skillnad från medelvärdet och att intervallet ligger under 40,0 vilket gör att produkternas standardpris inte kommer att ökas.

3. Vi kan först dra slutsatser om alla anställda som arbetar för det är den information och data jag har tagit fram mest ifrån. Det är 290 anställda var av 206 män och 84 kvinnor, vilket gör att majoriteten är män som arbetar på detta företag och sitter även i högre positioner och ligger i medelåldern vilket också medför högre löner.

Det finns 3 skift under en arbetsdag men majoriteten arbetar under dagen som har som rollen från chef till säljare, jämfört med de som arbetar under kvällen och natten som är för det mesta produktionen men också frakt och order.

Under åren från 2006 till 2013 så var mest anställda personer år 2008 med 74 personer och 2009 med 148 personer. Man kan säga att företaget blev större och antal anställda från bara några få till signifikant ökning efter 3 år och mer än dubbelt året efter. Men sedan så minskades med antal anställda efter 2009 och det kan vara till exempel det stabiliserades och det var redan många anställda som arbetade och behövde därefter inte anställa så många personer.

Månaderna som det var blev flest anställda var vintermånaderna december, januari och februari vilket kan vara rimligt då den hektiska perioden människor shoppar är under vintern runt julen till exempel och därav anställde företaget flest under dessa månader. Vissa månader var det ingen som blev anställd till exempel sommarmånader juli och augusti då många har semester och därav inte många som anställer under sommaren.

Företaget säljer produkter från en lägre prisklass till en högre prisklass, då det lägsta priset är från 2,29 till högsta priset 3578,27. Men det var även många produkter värderades som noll och då kan man undra om det var för att produkterna inte kunde få det sålt och därav slutade sälja produkterna, eller att produkterna har tillverkats men inte satt ett pris ännu.

Slutsats om företaget skulle jag säga att det går bra för företaget från att tabellen då antal kunder företaget har och alla anställda som finns i olika avdelningar och enheter. De tillverkar och säljer många och olika produkter i form av färg och kategorier i olika prisklasser, samt försäljarna gör ett bra arbete och många lyckas sälja bra och utvecklas. Av denna databas skulle jag rekommendera folk som vill kolla och fördjupa sig i ett företag som har ett bra framsteg.

4. Ett företag som säljer cykeldelar varför vill man fördjupa sig deras databas? Med AdventureWorks2022 kan vi följa från början till slut av en process från tillverkning av produkter tills produkten fraktas och hamnar utanför en kunds dörr. Få en mer detaljerad information om olika hanteringar såsom personal, kunder, order, lager och försäljning. Från ett lägre till högre prisklass av produkter, majoriteten som arbetar är män än kvinnor och medelvärde kan beräknas på ett standardpris. Detta är bara en liten del av information som kan tas fram från databasen, kreativiteten ökar ju mer man fördjupa sig i databasen och gör därför bara fler queries för att få fram den informationen man vill ha.
5. Jag gjorde min muntliga presentation den 2023-01-06

### 3. Reflektion på eget arbete

1. Hela inlämningsuppfiten har varit en utmaning. När vi har gått igenom kunskapskontrollen på lektioner så känns det som förstår men när man väl har sätter sig in i uppgifterna dyker flera frågetecken upp. Utmaningar har först varit att få fram någon information i databasen och sedan komma på queries för att veta mer och beräkna på grejer. Först tog det tid för mig vad jag ville få fram men ju mer queries man skrev desto fler frågor kom man på som man ville ha svar på och då gick det bättre. Andra var att försöka få min sql kod till jupyter notebook men sedan lyckades jag kopiera och skriva det i python. Tredje var att skriva rapporten, då jag har varit osäker vad som ska skrivas i varje fråga, som skrev jag en deskriptiv sammanfattning nu och hur skriver jag en statisk analys. Google har varit mycket till min hjälp av arbete för att hjälpa mig att skriva rapporten och även om jag fastnade att skriva koder.
2. Känner mig osäker vilket betyg jag ska ha då jag har varit osäker om själva hela kursen då jag tycker att kursen har gått för snabbt och det hade varit bra om kursen var längre för jag känner att jag vill förstå mer inom SQL och timingen över perioden jul och nyår har varit hektisk att hinna med. Jag hoppas på ett VG men jag är glad om jag blir godkänd för denna kurs har varit en utmaning.
3. Börja tidigare. Allt har tagit väldigt lång tid att utföra från kod till rapport och därför hade jag gett mig själv tipsen att börja tidigare med allting. Ett bra tips jag har gjort sedan början är att skriva anteckningar från alla videos från linkedIn för det har hjälpt mig att skriva queries och all innebörd av från varje funktion, clause och statement.