

1. Teoretiska Frågor

1. Beskriv kort hur en relationsdatabas fungerar.

I en relationsdatabas lagras data i olika tabeller. Tabellerna har överlappande information så man kan koppla data från flera tabeller.

2. Vad menas med "CRUD" flödet?

Create Read Update Delete

De är viktiga operationer som används i databaser.

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

Om vi anger "inner join" kommer matchande data från båda tabellerna att returneras.

Om vi anger "left join" kommer all data från den vänstra tabellen och matchande data från den högra att returneras.

Man använder "join" att kombinera data från flera tabeller.

4. Beskriv kort vad indexering i SQL innebär.

Index är ett speciellt värde som man ger till data i databasen. Med hjälp av index kan man snabbare hämta data från databasen.

5. Beskriv kort vad en vy i SQL är.

Vy är en virtuell tabell som skapas från svar på användarförfrågningar, men denna information sparas inte på SQL Server.

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

Det är en procedur som lagras i databasen. Det betyder att man kan spara små programmer i databasen.

2. Programmeringsuppgift och Rapport

Jag började arbeta på ett företag som säljer 504 olika produkter

Results	Messages						
<table><tr><th>ProductID</th><th>Name</th></tr><tr><td>503</td><td>998 Road-750 Black, 48</td></tr><tr><td>504</td><td>999 Road-750 Black, 52</td></tr></table>	ProductID	Name	503	998 Road-750 Black, 48	504	999 Road-750 Black, 52	
ProductID	Name						
503	998 Road-750 Black, 48						
504	999 Road-750 Black, 52						
Query executed successfully. LAPTOP-1T21SEQ (16.0 RTM) LAPTOP-1T21SEQ\lidii ... AdventureWorks2022 00:00:00 504 rows							

i 10 regioner :

Results

Messages

TerritoryID	Name	CountryRegionCode
1	Northwest	US
2	Northeast	US
3	Central	US
4	Southwest	US
5	Southeast	US
6	Canada	CA
7	France	FR
8	Germany	DE
9	Australia	AU
10	United Kingdom	GB

Totalt antal anställda på företaget är 290.

	BusinessEntityID	NationalIDNumber	LoginID	OrganizationNode	OrganizationLevel	JobTitle	BirthDate	MaritalStat
1	1	295847284	adventure-works\ken0	NULL	NULL	Chief Executive Officer	1969-01-29	S
2	2	245797967	adventure-works\terri0	0x58	1	Vice President of Engineering	1971-08-01	S
3	3	500647174	adventure-works\rob0	0x5A00	2	Engineering Manager	1974-11-10	M

Query executed successfully. LAPTOP-1T21SEQ (16.0 RTM) LAPTOP-1T21SEQ\idii ... AdventureWorks2022 00:00:00 290 rows

Tabellen visar att varje "NationalIDNumber" motsvarar ett unikt "BusinessEntityID" så "BusinessEntityID" är unikt för varje anställd

Företagsdatabasen kallas AdventureWorks2022. Den är uppbyggd och uppdelad i 5 huvudsektioner:

1. Human Resources
2. Person
3. Production
4. Purchasing
5. Sales

Jag antar att det på grund av att tabellerna är pedagogiska, är förvirring i datumen. Till exempel:

- i tabellerna Sales.SalesTerritory och Sales.SalesPerson anges inte vilket år som kallas "year to date" och vilket som kallas "last year". Det finns "ModifiedDate" men för "Sales.SalesTerritory" är den 2008-04-30 och Sales.SalesPerson är "ModifiedDate" även tidigare att "StartDate" i tabellen "HumanResources.EmployeeDepartmentHistory"
- i tabellen "Sales.Customer" finns bara det "ModifiedDate" för alla rader samma 2012-09-14 and "Sales.SalesOrderHeader" finns information för varje dag från 2011-05-31 till OrderData 2014-06-30.

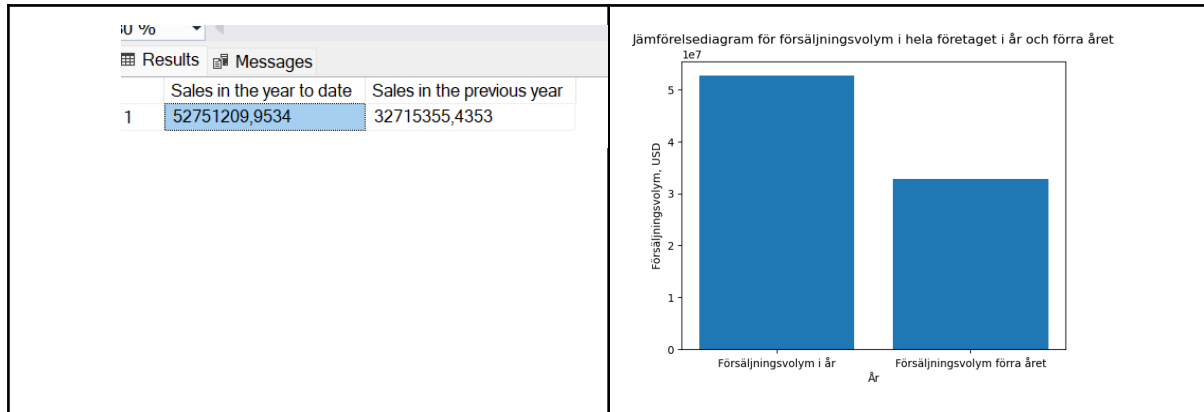
I detta avseende kommer jag att anta att nu är året slut och chefer vill sammanfatta försäljningsresultat. I år (year to date) genomfördes en stor reklamkampanj främst för att öka onlineförsäljningen. Jag behöver utvärdera hur försäljningsvolymen förändrats i hela företaget och per region. Även hur försäljningsvolymen har förändrats i år i förhållandet mellan anställdas försäljning och onlineförsäljning. Jag kommer att anse att onlineförsäljningar är alla försäljningar vilka de anställda inte gjorde.

För att slutföra uppgiften kommer jag att använda följande tabeller:

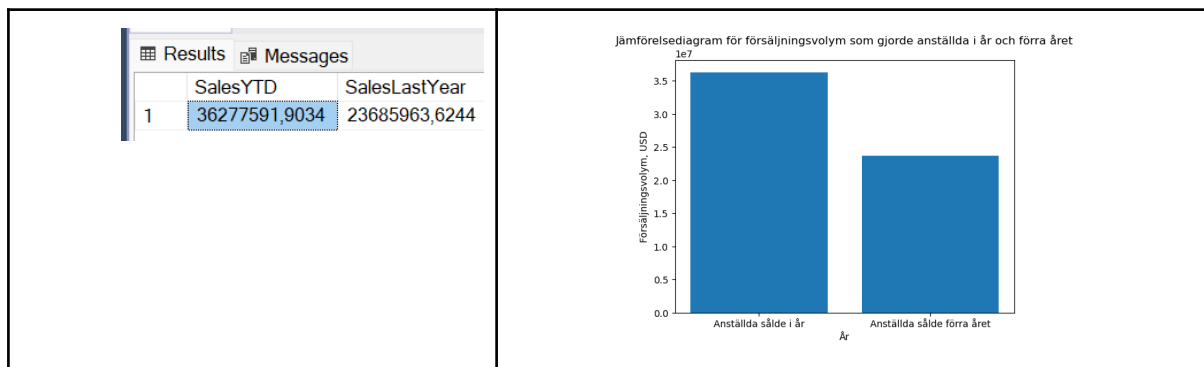
1. Sales.SalesTerritory. I den här tabellen är jag intresserad av följande kolumner: TerritoryID, Name, CountryRegionCode, SalesYTD, SalesLastYear.
2. Sales.SalesPerson. I den här tabellen är jag intresserad av följande kolumner: BusinessEntityID, TerritoryID, SalesYTD, SalesLastYear
3. Sales.Customer. I den här tabellen är jag intresserad av följande kolumner: CustomerID, PersonID, TerritoryID
4. HumanResources.EmployeeDepartmentHistory. I den här tabellen är jag intresserad av följande kolumner: BusinessEntityID, JobTitle
5. HumanResources.Department. I den här tabellen är jag intresserad av följande kolumner: DepartmentID, Name
6. Sales.SalesOrderHeader. I den här tabellen är jag intresserad av följande kolumner: SalesOrderID, OrderDate, Status, OnlineOrderFlag, TerritoryID, SalesPersonID

1.Försäljningsresultat för hela företaget

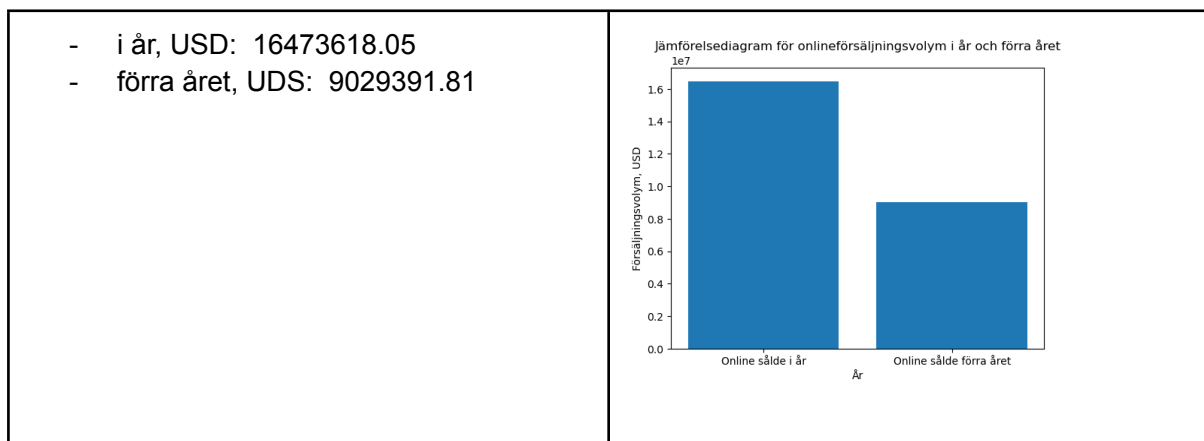
1.1. Den totala försäljningen för hela företaget i år ökade med 20 035 850,0 USD miljoner eller 61,24 % jämfört med förra året.



1.2. Summa försäljningen av anställda ökade med 12 591 630,0 USD eller 53,16% jämfört med föregående år.



1.3. Summa onlineförsäljningen ökade med 7 444 226.24 USD eller 82.44% jämfört med föregående år.



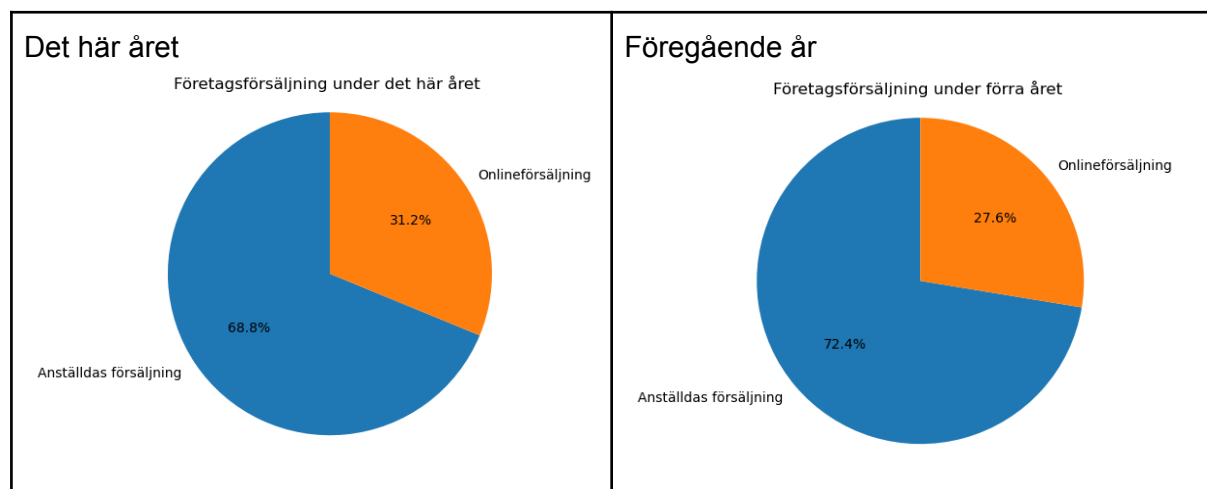
1.4. Slutsatser om försäljningen i företaget som helhet

Pivottabellen: "Försäljning av företaget som helhet, försäljning av anställda och onlineförsäljning"

	Företaget som helhet, USD	Anställda, USD	Online, USD \
I år	52751209.95	36277591.90	16473618.05
Förra året	32715355.44	23685963.62	9029391.81

	Andel av anställdas försäljning, %	Andel av onlineförsäljning, %
I år	68.77	31.23
Förra året	72.40	27.60

Andelar av anställdas försäljning och onlineförsäljning i hela företaget

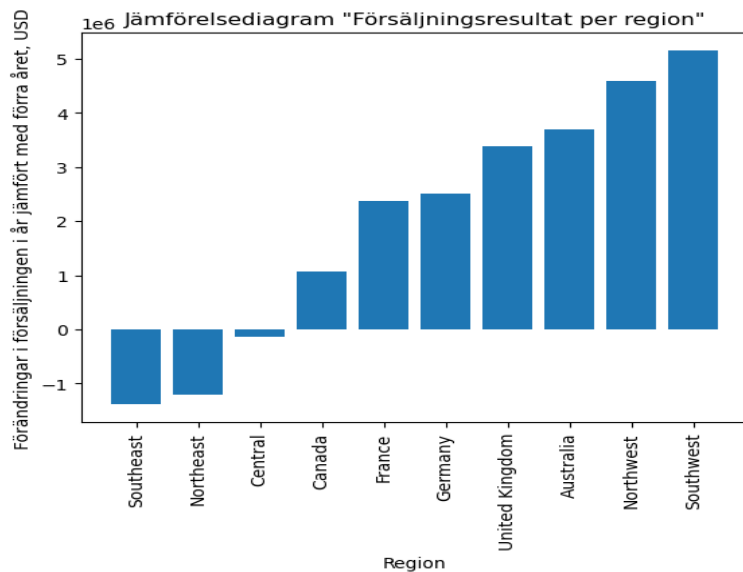


Totalt sett ökade företagets försäljning med 61,24 % i år jämfört med förra året. Anställdas försäljning ökade med 53,16 %. Onlineförsäljning ökade med 82,44%. Därför kan vi dra slutsatsen att reklamkampanj som syftade till att öka onlineförsäljningen gav resultat eftersom den relativa ökningen av onlineförsäljningen (82,44%) är betydligt högre än den relativa ökningen av anställdas försäljningen (53,16%). Dessutom onlineförsäljningens andel av företagets totala försäljning ökade med 3,6 %.

2. Försäljningsresultat i regioner

2.1. Genomsnittlig data av försäljningsvolym per region

Genomsnittlig försäljningsvolym per region förra året, USD: 3271535.54
Genomsnittlig försäljningsvolym per region i det här året, USD: 5275121.0
Genomsnittligt antal kunder per region, stycken: 1982.0
Genomsnittlig försäljningsökning i år jämfört med förra året, USD: 2003585.45
Genomsnittlig försäljningsökning i år jämfört med förra året, %: 83.99



2.2. Det finns 3 regioner som har ett utflöde vad gäller försäljningsvolym Southeast, Northeast och Central

	TerritoryID	Name	CountryRegionCode	SalesYTD	SalesLastYear	ChangeYTD	ChangeYTD_procent	NumberOfCustomer
0	2	Northeast	US	2402176.85	3607148.94	-1204972.09	-33.41	113
1	3	Central	US	3072175.12	3205014.08	-132838.96	-4.14	132
2	5	Southeast	US	2538667.25	3925071.43	-1386404.18	-35.32	176
3	6	Canada	CA	6771829.14	5693988.86	1077840.28	18.93	1791
4	8	Germany	DE	3805202.35	1307949.79	2497252.56	190.93	1852
5	7	France	FR	4772398.31	2396539.76	2375858.55	99.14	1884
6	10	United Kingdom	GB	5012905.37	1635823.40	3377081.97	206.45	1991
7	1	Northwest	US	7887186.79	3298694.49	4588492.29	139.10	3520
8	9	Australia	AU	5977814.92	2278548.98	3699265.94	162.35	3665
9	4	Southwest	US	10510853.87	5366575.71	5144278.16	95.86	4696

Även i dessa regioner kan ett onormalt lågt antal kunder noteras, vilket kan påverka försäljningsvolymerna.

Det bör dock noteras att det finns en felaktighet i uppgifterna om anställdas försäljning i regioner Northeast (TerritoryID=2) och Central (TerritoryID=2). Den totala försäljningen i regionen är mindre än den anställdas försäljning.

	TerritoryID	Name	SalesRegYTD	SalesPersonYTD	OnlinesalesYTD
0	2	Northeast	2402176.85	3763178.18	-1361001.33
1	3	Central	3072175.12	3189418.37	-117243.25

Det antas att orsaken kan vara att de anställda bytt arbetsplats.

	BusinessEntityID	TerritoryID	StartDate	EndDate	rowguid	ModifiedDate
0	275	2	2011-05-31	2012-11-29	8563CE6A-00FF-47D7-BA4D-3C3E1CDEF531	2012-11-22
1	275	3	2012-11-30	NaT	2F44304C-EE87-4C72-813E-CA75C5F61F4C	2012-11-23
2	277	3	2011-05-31	2012-11-29	3E9F893D-5142-46C9-A76A-867D1E3D6F90	2012-11-22
3	277	2	2012-11-30	NaT	132E4721-32DD-4A73-B556-1837F3A2B9AE	2012-11-23

Men inom ramen för denna rapport kommer jag inte att fördjupa mig i sökandet efter orsaken, utan kommer att fokusera på att analysera försäljningen i de återstående 8 regionerna.

2.3. Sammanfattningstabell per region

TerritoryID	Name	SalesRegYTD	SalesPersonYTD	Online_salesYTD	SalesRegLY	SalesPersonLY	Online_salesLY	ChangeReg	ChangePerson	ChangeOnline
5	Southeast	2538667.25	2315185.61	223481.64	3925071.43	1849640.94	2075430.49	-1386404.18	465544.67	223481.64
10	United Kingdom	5012905.37	4116871.23	896034.14	1635823.40	1635823.40	0.00	3377081.97	2481047.83	896034.14
7	France	4772398.31	3121616.32	1650781.99	2396539.76	2396539.76	0.00	2375858.55	725076.56	1650781.99
8	Germany	3805202.35	1827066.71	1978135.64	1307949.79	1307949.79	0.00	2497252.56	519116.92	1978135.64
6	Canada	6771829.14	4058260.18	2713568.96	5693988.86	3658511.55	2035477.31	1077840.28	399748.63	2713568.96
1	Northwest	7887186.79	4502152.27	3385034.52	3298694.49	3298694.49	0.00	4588492.29	1203457.77	3385034.52
4	Southwest	10510853.87	6709904.17	3800949.71	5366575.71	3512662.03	1853913.68	5144278.16	3197242.14	3800949.71
9	Australia	5977814.92	1421810.92	4556003.99	2278548.98	2278548.98	0.00	3699265.94	-856738.05	4556003.99

Den sammanfattande tabellen visar att:

- 1) Förra året hade endast 3 regioner onlineförsäljning, men i år har alla regioner onlineförsäljning
- 2) I Australia (TerritoryID=2) har ett utflöde i försäljning av anställda, men samtidigt en ökning av onlineförsäljningen
- 3) Analyser per region visade också att reklamkampanjen för att öka onlineförsäljningen är framgångsrik
- 4) Det är skillnad i försäljningsvolymen för anställda generellt och per region. Detta på grund av att det är 3 anställda som inte är ansatta i regionen. Läs mer om anställdas försäljning i punkt 3.

3. Försäljningsresultat av anställda på försäljningsavdelningen

3.1. Totalt sysselsätter försäljningsavdelningen 18 personer:

- BusinessEntityID=273 Vice President of Sales;
- de 17 anställda som gjorde försäljningar i den här året ;
-

3.2. De tre anställda är inte tilldelade regioner:

Results Messages					
	BusinessEntityID	JobTitle	TerritoryID	SalesYTD	SalesLastYear
1	274	North American Sales Manager	NULL	559697,5639	0,00
2	285	Pacific Sales Manager	NULL	172524,4512	0,00
3	287	European Sales Manager	NULL	519905,932	0,00

Dessa Managers gjorde inte försäljningarna förra året. Det totala beloppet av deras försäljningar i detta år är 1 252 127.95 USD

3.3. De 14 anställda är tilldelade regioner:

	TerritoryID	Name	BusinessEntityID	SalesYTD	SalesLastYear	SalesGrowthYTD
1	1	Northwest	280	1352577,1325	1927059,178	-574482,0455
2	1	Northwest	283	1573012,9383	1371635,3158	201377,6225
3	1	Northwest	284	1576562,1966	0,00	1576562,1966
4	2	Northeast	275	3763178,1787	1750406,4785	2012771,7002
5	3	Central	277	3189418,3662	1997186,2037	1192232,1625
6	4	Southwest	276	4251368,5497	1439156,0291	2812212,5206
7	4	Southwest	281	2458535,6169	2073505,9999	385029,617
8	5	Southeast	279	2315185,611	1849640,9418	465544,6692
9	6	Canada	278	1453719,4653	1620276,8966	-166557,4313
10	6	Canada	282	2604540,7172	2038234,6549	566306,0623
11	7	France	290	3121616,3202	2396539,7601	725076,5601
12	8	Germany	288	1827066,7118	1307949,7917	519116,9201
13	9	Australia	286	1421810,9242	2278548,9776	-856738,0534
14	10	United Ki...	289	4116871,2277	1635823,3967	2481047,831

Bland de återstående 14 anställda sålde 3 anställda mindre i år än förra året.

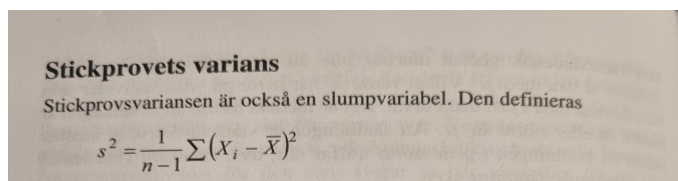
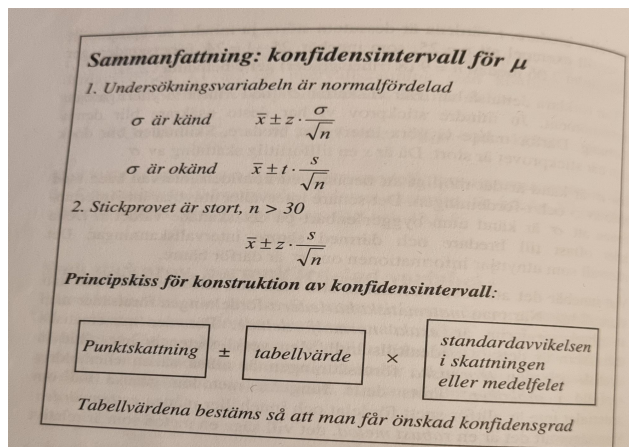
	TerritoryID	Name	BusinessEntityID	SalesYTD	SalesLastYear	SalesGrowthYTD
1	6	Canada	278	1453719,4653	1620276,8966	-166557,4313
2	1	Northwest	280	1352577,1325	1927059,178	-574482,0455
3	9	Australia	286	1421810,9242	2278548,9776	-856738,0534

4. Konfidensintervall

Tabellen "Sales.SalesOrderHeader" har kolumner CustomerID och SubTotal (beställningssumma). Totalt har vi 19119 kunder. Jag hittade att riktigt medelvärde av beställningssumma för 1 kund är 5745.4USD.

	CustomerID	SaleAmount
1	14324	5121,4282
2	22814	4,99
3	11407	53,99
4	28387	583,97
5	19897	596,96
6	15675	7206,38
7	24165	3046,84
8	27036	7,28

För att räkna ut konfidensintervall för medelvärdet använder jag formler:



Jag vill kontrollera om medelvärdet ligger inom 95 % i konfidensintervall:

1) om jag tar ett stickprov på 30 slumpmässigt utvalda kunder använder jag Tabell 4.t-fördelning ($t=2,045$ om Frihetsgrader=29 och Sannolikhet=95%)

```
In [122]: #Tabeller 4 (t-fördelning)
#Frihetsgrader=29
#t(p=95%)=2,045
a = 2.045 * math.sqrt(varians_kvadrat_30/30)
a_rounded = round(a, 2)
print('Konfidensintervall of ordersumma 30 slumpmässigt utvalda kunder: ', average_amount_30_rounded, '+ -', a_rounded)
```

Konfidensintervall of ordersumma 30 slumpmässigt utvalda kunder: 2507.54 + - 998.53

Nej, riktiga medelvärdet (5745.4USD) ligger inte i den här konfidensintervallen som jag fick (2507,54+-997,53). Dessutom, när beräkningen upprepas många gånger ger resultat för 30 slumpmässigt utvalda kunder alltid felaktiga resultat

2) om jag tar ett stickprov på 1000 slumpmässigt utvalda kunder använder jag Tabell 3a. Normalfördelningen ($z=1,96$ om Sannolikhet=95%)

```
# Konfidensintervall för medelvärdet orderbeloppet per kund för slumpmässigt utvalda 1000 kunder

#Tabeller 3b (Normalfördelningen)
#z(p=95%)=1,96
b = 1.96 * math.sqrt(varians_kvadrat_1000/1000)
b_rounded = round(b, 2)
print('Konfidensintervall of ordersumma 1000 slumpmässigt utvalda kunder: ', average_amount_1000_rounded, '+ - ', b_rounded)

Konfidensintervall of ordersumma 1000 slumpmässigt utvalda kunder:  5923.53 + - 2740.62
```

Ja, riktiga medelvärdet (5745.4USD) ligger inom den här konfidensintervallen som jag fick (5183.42+-2080,38).

På grund av att beställningssumma varierar från 1,37 USD till 877 107,19 USD ger ett slumpmässigt urval på 30 kunder inte det korrekta resultatet.

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

Jag hade frågor när jag skrev kod i Python. Till exempel:

- visa antalet decimaler;
- pie chart.

För en lösning letade jag efter svar på Internet.

Det var lite svårt att komma på en fråga som jag skulle svara på med den konfidentiella intervall beräkningen. Jag försökte flera alternativ för att beräkna konfidentialitets intervallet.

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

VG

Jag försökte uppfylla villkoren för VG och förstå SQL i detalj

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

När jag tittar på träningsvideor är det bra att stoppa videon efter varje exempel och försök upprepa exemplet själv.