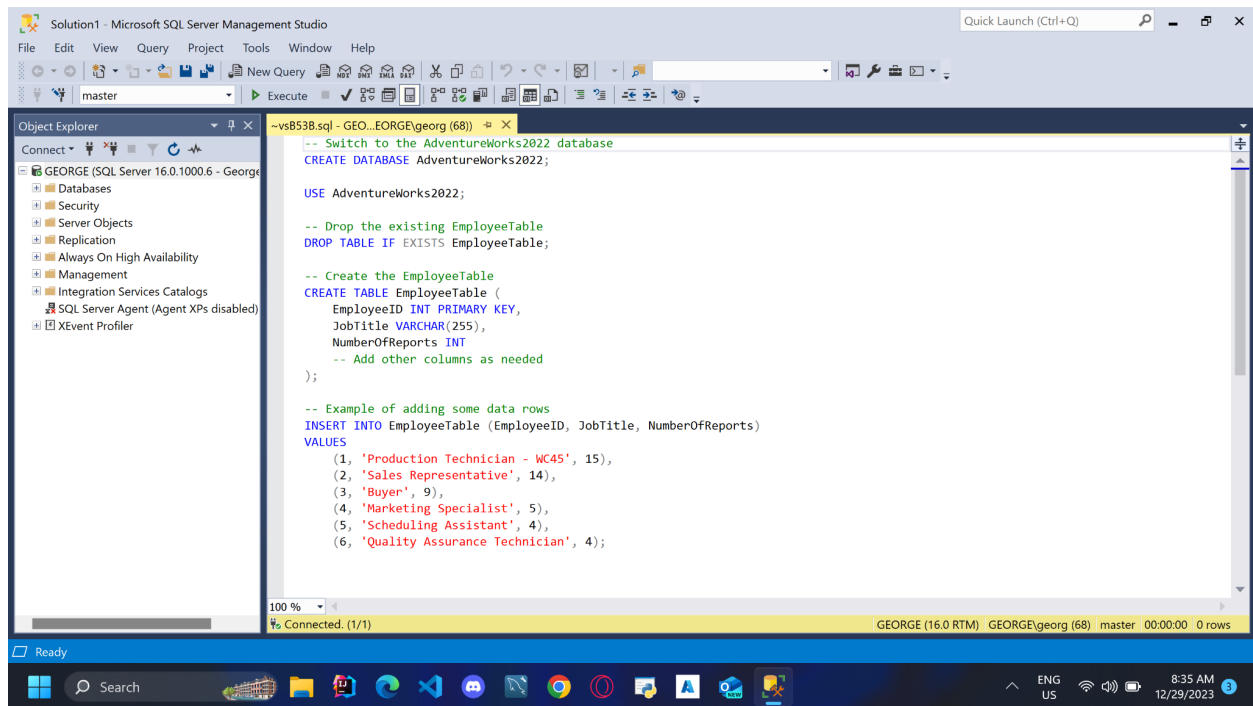


## Rapport: Analytisk Exploration av AdventureWorks2022-databasen för Skolan

Jag fick en uppgift från skolan för att undersöka AdventureWorks2022-databasen, Målet var att använda python och SQL-Alchemy för att skapa en deskriptiv sammanfattning av datan och göra mer tillgänglig för framtida analyser.

För att hämta relevanta uppgifter från AdventureWorks2022-databasen som ser ut så:



```
SELECT * FROM INFORMATION_SCHEMA.TABLES;
```

för att hämta information om alla tabeller i databasen.

Resultatet av denna fråga gav mig en översikt över tabellerna i databasen och deras tillhörande kolumner.

*Visualisering av antal anställda per jobbtitel:*

```
SELECT JobTitle, COUNT(*) as NumberOfEmployees
FROM EmployeeTable
GROUP BY JobTitle
ORDER BY NumberOfEmployees DESC;
```

För att göra datan mer begriplig använde jag Python-biblioteken Pandas och Matplotlib för att skapa visualiseringar. Nedan är ett stapeldiagram som visar antalet anställda per jobbtitel:

```
Visualisering med Matplotlib
*Skapa table 10 bred och 6 hög
plt.figure(figsize=(10, 6))
*skapa stapeldiagram med jobbtider på x-axeln och antal anställda på y-axeln
plt.bar(df_results['JobTitle'], df_results['NumberOfEmployees'])
*Ange label för x-axlen 'jobbtitel'
plt.xlabel('JobTitle')
*Ange label för y-axlen 'Antal Anställda'
plt.ylabel('Antal Anställda')
*Ange titel på antal anställda per jobbtitel
plt.title('Antal Anställda per Jobbtitel')
*Rota x axeln 45 grader och justera de till höger
plt.xticks(rotation=45, ha='right')
*det är för att visa diagrammet
plt.show()
```

För den statistiska analysen beräknade jag medelantalet anställda och skapade ett konfidensintervall:

```
# Statistisk analys
*Beräkna genomsnittet av antalet of anställda
mean_employees = df_results['NumberOfEmployees'].mean()
*Beräkna standardavvikelsen av antalet anställda
std_dev_employees = df_results['NumberOfEmployees'].std()
```

\* Beräkna konfidensintervallet för antalet anställda med en konfidensnivå på 95%  
confidence\_interval = (mean\_employees-1.96\*std\_dev\_employees, mean\_employees+1.96  
\* std\_dev\_employees)

\*Print funktion för att visa anställda, antal anställda och Konfidensintervall för antalet anställda  
print(f"Medelantalet anställda: {mean\_employees}")  
print(f"Standardavvikelse av antal anställda: {std\_dev\_employees}")  
print(f"Konfidensintervall för antalet anställda: {confidence\_interval}")

## Resultat och Slutsatser

- Medelantalet anställda: Det är 1.0. I varje jobbtittlar har endast 1 anställda per jobbtittle.
- Standardavvikelse av antal anställda: Det finns ingen variation eftersom varje jobbtittle har 1 anställd..
- Konfidensintervall för antalet anställda: Konfidensintervallet är (1.0, 1.0). För att det finns ingen variation.

## Förklaring:

Det finns bara en anställd för varje jobbtitel i den givna datan, det finns ingen variation i antal anställda mellan olika jobbtitlar, Konfidensintervallet bekräftar detta genom att visa att osäkerhet kring det genomsnittliga anställda, eftersom intervallet sträcker sig från 1.0 till 1.0.

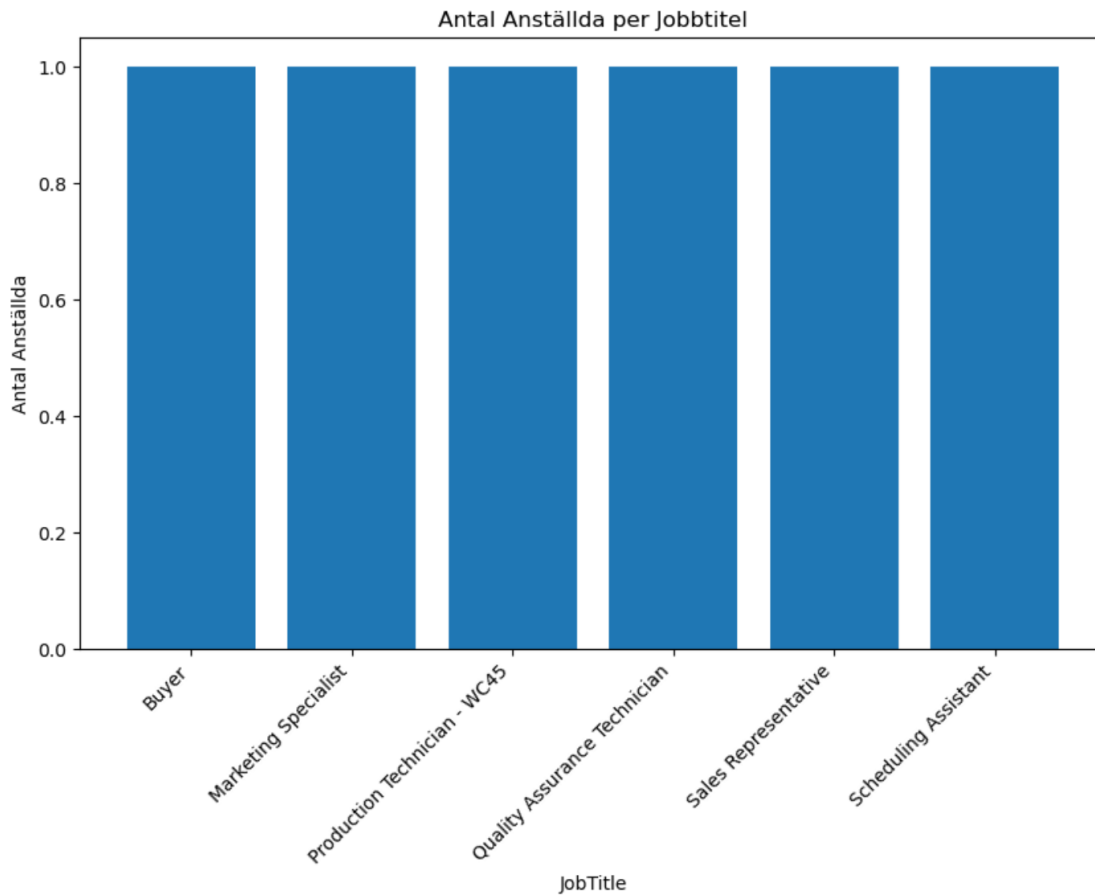
---

Deskriptiv sammanfattning av AdventureWorks2022-databasen:

	TABLE_CATALOG	TABLE_SCHEMA	TABLE_NAME	TABLE_TYPE
0	AdventureWorks2022	dbo	EmployeeTable	BASE TABLE

Resultat av SQL-frågan:

	JobTitle	NumberOfEmployees
0	Buyer	1
1	Marketing Specialist	1
2	Production Technician - WC45	1
3	Quality Assurance Technician	1
4	Sales Representative	1
5	Scheduling Assistant	1



Medelantalet anställda: 1.0  
Standardavvikelse av antal anställda: 0.0  
Konfidsensintervall för antalet anställda: (1.0, 1.0)

---

Genom att kombinera SQL och Python har jag skapat en användbar och visuell presentation av AdventureWorks2022-datan. Den utförda analysen ger underlag för framtida beslut och strategier inom skolan.

ationen av koden genomfördes den [datum].

Min utmaning var att lära mig SQL server, eftersom jag har jobbat på min andra utbildning på Mysql program annars var allt annat lätt.