

TASK 2 - What is the relationship between annual leave taken and bonus

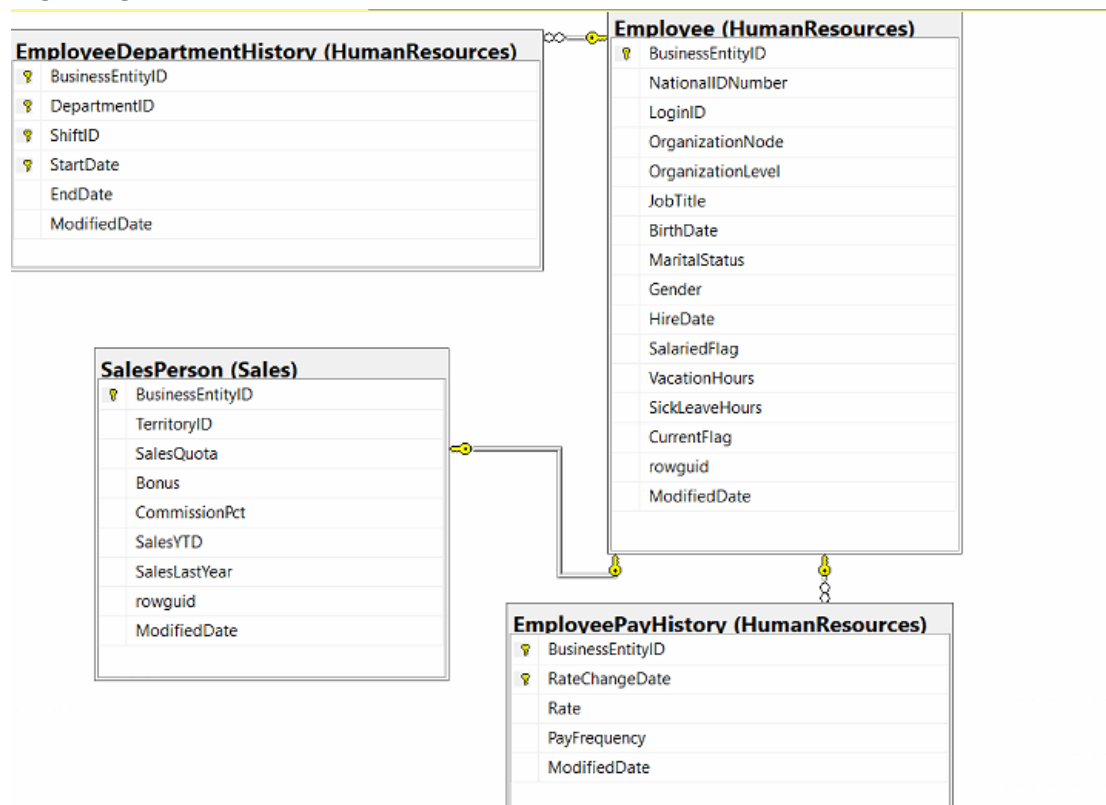
Step by Step

1.Import the database Adventure 2019 , to get access to the dataset.

2.Use the Microsoft SQL SERVER explore data . So we can start exploring the Object and to observe the metadata, tables, key words, columns, rows values ect, we can use the function. We can also check the data on [AdventureWorks Dictionary](#).

So, we started observing tables such as HumanResources.Employee, HumanResources.EmployeeDepartmentHistory, HumanResources.EmployeeDepartmentHistory, HumanResources.EmployeePayHistory. And also checking (using the [AdventureWorks Dictionary](#)), to find in which table we could find Bonus and VacationHours variables. And after did a Diagram to check the key value for both variables. We placed in a SQL query called - [Question2Bonus](#).

***fig. Diagram Question 2**



4. Execute the code. We've chosen the function INNER JOIN to join the columns from Sales.SalesPerson, when we've got the column values for Bonus variable, and HumanResources.Employee, when we've got the VacationHours. The Key value for both is (BusinessEntityID).

```

SELECT Employee.BusinessEntityID,
       VacationHours,
       Bonus
FROM HumanResources.Employee
INNER JOIN Sales.SalesPerson
ON Sales.SalesPerson.BusinessEntityID = HumanResources.Employee.BusinessEntityID
ORDER BY Bonus DESC

```

**We executed the Sales.Sale.Person table to confirm it is just 17 values.

fig. Question 2- Query and Results

```

SELECT Employee.BusinessEntityID,
       VacationHours,
       Bonus
FROM HumanResources.Employee
INNER JOIN Sales.SalesPerson
ON Sales.SalesPerson.BusinessEntityID = HumanResources.Employee.BusinessEntityID
ORDER BY Bonus DESC

```

```

SELECT *
FROM Sales.SalesPerson

```

100 %

Results Messages

	BusinessEntityID	TerritoryID	SalesQuota	Bonus	CommissionPct	SalesYTD	SalesLastYear	rowguid	ModifiedDate
1	274	NULL	NULL	0.00	0.00	559697.5639	0.00	48754992-9EE0-4C0E-9C94-9451604E3E02	2010-12-28 00:00:00.000
2	275	2	300000.00	4100.00	0.012	3763178.1787	1750406.4785	1E0A7274-3064-4F58-88EE-4C6586C87189	2011-05-24 00:00:00.000
3	276	4	250000.00	2000.00	0.015	4251368.5497	1439156.0291	4DD9EEE4-8E81-4FBC-AF97-683394C1F7C0	2011-05-24 00:00:00.000
4	277	3	250000.00	2500.00	0.015	3109410.3662	1997196.2037	39012920-BFEC-4242-874D-423162C3F567	2011-05-24 00:00:00.000
5	278	6	250000.00	500.00	0.01	1453719.4653	1620276.8986	7A0AE1AB-B283-40F9-91D1-167ABF06D720	2011-05-24 00:00:00.000
6	279	5	300000.00	6700.00	0.01	2315185.611	1849640.9418	52A5179D-3239-4157-AE29-17E868296DC0	2011-05-24 00:00:00.000
7	280	1	250000.00	5000.00	0.01	1352577.1325	1927059.178	BE941A4A-FB50-4947-BD44-BB8972365B08	2011-05-24 00:00:00.000
8	281	4	250000.00	3550.00	0.01	2450535.6189	2073505.9999	35326DD8-7278-4FEF-B3BA-EA137B89094E	2011-05-24 00:00:00.000

BusinessEntityID	VacationHours	Bonus
279	29	6700
286	36	5650
289	37	5150
280	22	5000
282	31	5000
275	38	4100
284	39	3900
281	26	3550
283	23	3500
277	24	2500
276	27	2000
290	34	985
278	33	500
288	35	75
274	14	0
285	20	0
287	21	0

5. Download the Results and save as .csv [Question2Bonus.csv](#)

6. Open Python and import the methods pandas, numpy and matplotlib to use functions

```
import pandas as pd
import numpy as np
from matplotlib import pyplot as plt
```

7. Use the function `pd.read_csv()` to run the code.

8. Use the functions `plt.plot()`, `plt.scatter()`, `plt.show()` to present some chartes. During the process to get answers we started thinking in a hypothetical thoughts

There is a correlation (not strong) between Annual Leave Taken and Bonus, considering as the same time that annual leave taken increases the bonus(money) decreases.

To test it we used the charts : Scatterplot, line plot, and dotplot, because they are all used to compare continuous variables and check the correlation between theirs.

We saved the code on [Question2_Python](#)

