What is the relationship between sick leave and Job Title (PersonType)?

I began by finding the tables in the AdventureWork2019 database that have the columns for sick leave, job title, and person type. After some searching, I was able to find two tables that had those columns, the tables being ***HumanResources.Employee*** and ***Person.Person***.

I then began looking for a relationship between the two tables, to do this I used the Database Diagram feature on SSMS (**Image 1a**). As you can see in Image 1a below, both the ***HumanResources*.Employee** and ***Person.Person*** tables share the same Primary Key: ***BusinessEntityID*.**

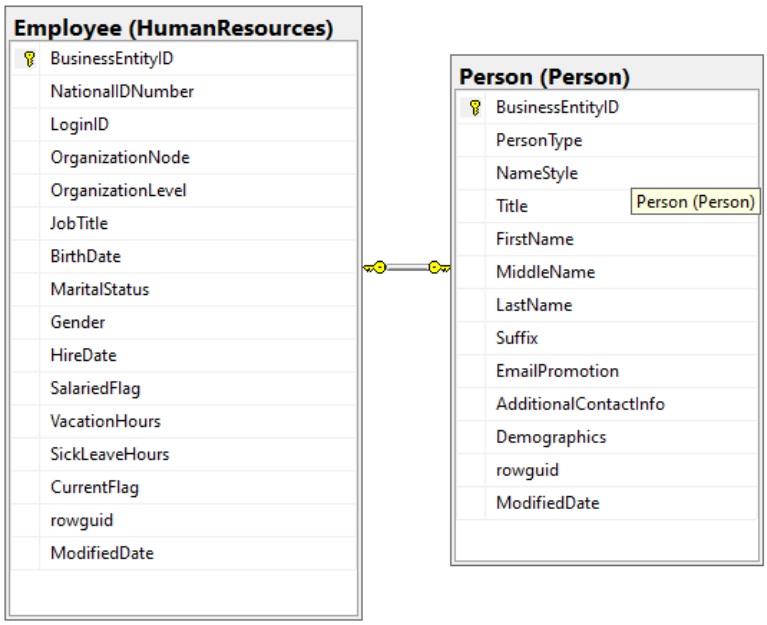


Image 1a

Once I was able to determine a relationship between the two tables, I could then write a query to select the relevant columns I needed.

I began by finding out how many distinct job titles there were in the ***HumanResources.Employee*** table by using the 'COUNT(DISTINCT) query (**Image 1b**). I did this because I wanted to know how many values I would be working with and whether the amount would clearly fit onto a chart. As you can see in the result (**Image 1c**), there were 67 distinct values of ***JobTitle*** in the ***HumanResources.Employee*** table, this would be too many to clearly showcase on a chart, I knew I would need to group the job titles by another variable to fix this.

```
SELECT COUNT(DISTINCT JobTitle) AS CountDistinctJobTitle
FROM HumanResources.Employee
```

Image 1b

	CountDistinctJobTitle
1	67

Image 1c

The variable I chose to group the ***JobTitle*** by is the ***OrganizationLevel***. Each ***JobTitle*** is a part of 4 distinct levels in the organization, with the exception being the Chief Executive Officer, who has no ***OrganizationLeve***l. As I was grouping by the ***OrganizationLevel***, I also had to then average the ***SickLeaveHours*** and alias it as ***AverageSickLeaveHours***. The diagram (**Image 1d**) below shows the query I used.

```
SELECT hre.OrganizationLevel, pp.PersonType, AVG(hre.SickLeaveHours) AS AverageSickLeaveHours
FROM HumanResources.Employee AS hre
INNER JOIN Person.Person AS pp ON pp.BusinessEntityID = hre.BusinessEntityID
GROUP BY hre.OrganizationLevel, pp.PersonType
ORDER BY AVG(hre.SickLeaveHours) DESC;
```

Image 1d

I joined the tables together using an INNER JOIN, to return the records that have matching values in both tables. I then used a GROUP BY statement to group the query by the ***OrganizationLevel*** and ***PersonType***. I finished the query off with an ORDER BY Statement to order the results by the average sick leave hours. The results of the query are in the image (**Image 1e**) below.

	OrganizationLevel	PersonType	AverageSickLeaveHours
1	NULL	EM	69
2	3	EM	49
3	2	EM	47
4	4	EM	44
5	3	SP	35
6	1	EM	34
7	2	SP	29

Image 1e

I saved the results as a CSV file to import in Visual Studio Code, where I could then plot a chart to analyse any relationships between the variables.

I imported the pandas package so python could read my csv file (**Image 2a**) and used the query .Head() to see the first 5 rows, I did this to make sure that the csv file was imported correctly, without errors.

```
#pandas imported, aliased as pd
import pandas as pd

#pandas loads the csv file 'jobtitlesickleave'
JTSL = pd.read_csv(r'C:\Users\User\Documents\CSV Python Files\QuestionFourV2.csv')
print(JTSL.head())
```

```
   OrganizationLevel PersonType  AverageSickLeaveHours
0                NaN         EM                     69
1                3.0         EM                     49
2                2.0         EM                     47
3                4.0         EM                     44
4                3.0         SP                     35
```

Image 2a

Once I was happy that the csv file was working correctly, I began to plot a series of charts to determine which chart showed the relationship between AverageSickLeaveHours and OrganizationLevel the best. The charts I looked at were Bar Charts, Scatter Plots, and Line Plots (**Image 2b**).

I decided to use the bar chart as I felt it displayed the data in the clearest manner, whilst the scatter plots and line plots did not. I added a title for the chart and labels to the x and y axis.

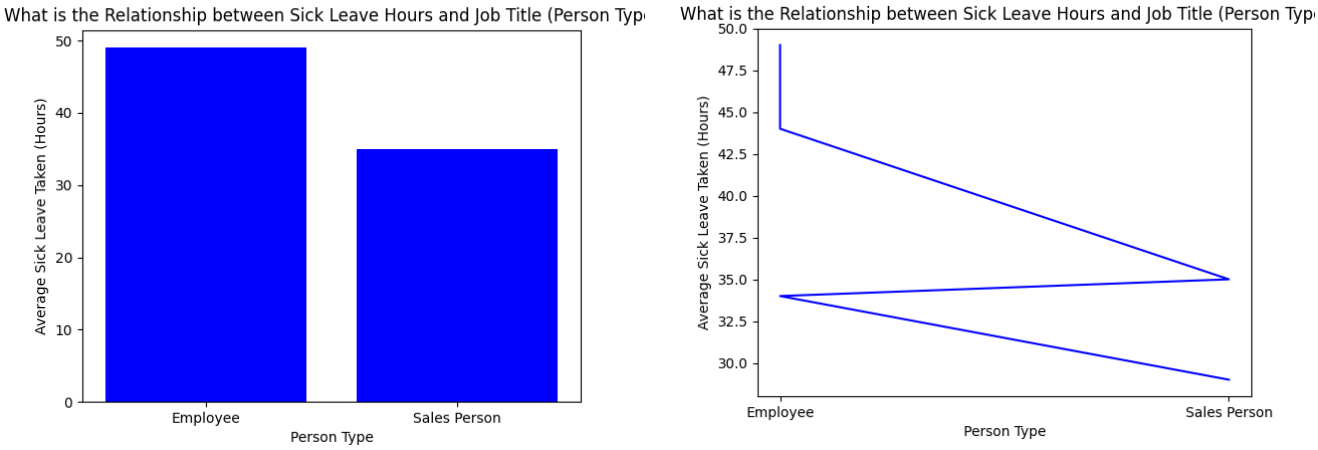
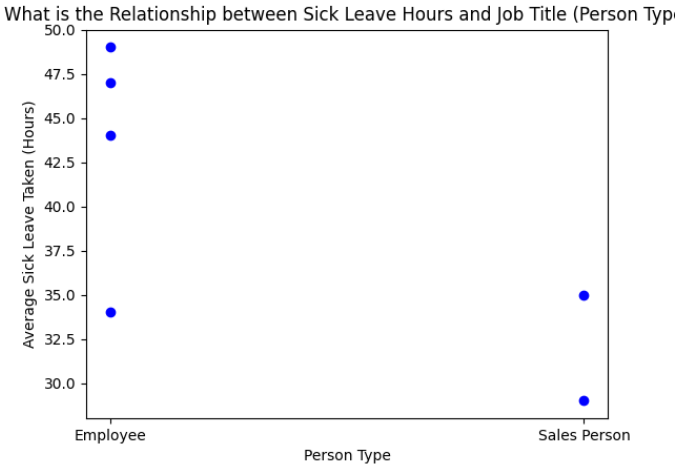


Image 2b



Whilst I was happy with the relationship the bar chart was showing, I felt that I could expand on it further by creating another chart showing the average sick leave taken by each level in the organization. I wrote a query on SSMS that showed me how exactly that (**Image 2c**) and I once again made a csv file from the data and imported it onto Visual Studio Code, where I created a bar chart (**Image 2d**) that showed how the more senior a job title, the less sick leave taken.

```
SELECT DISTINCT hre.OrganizationLevel, hre.JobTitle, AVG(hre.SickLeaveHours) AS AverageSickLeaveHours, pp.PersonType
FROM HumanResources.Employee AS hre
INNER JOIN Person.Person AS pp ON pp.BusinessEntityID = hre.BusinessEntityID
GROUP BY hre.JobTitle, pp.PersonType, hre.OrganizationLevel
HAVING hre.OrganizationLevel IS NOT NULL
ORDER BY hre.OrganizationLevel
```

Image 2c


Average Sick Leave Taken by Organization Level
Average Sick Leave Taken (Hours)
0
10
20
30
40
50
60
70
Level 1
Level 2
Level 3
Level 4
Organization Level

Image 2d