What is the relationship between annual leave taken and bonus?

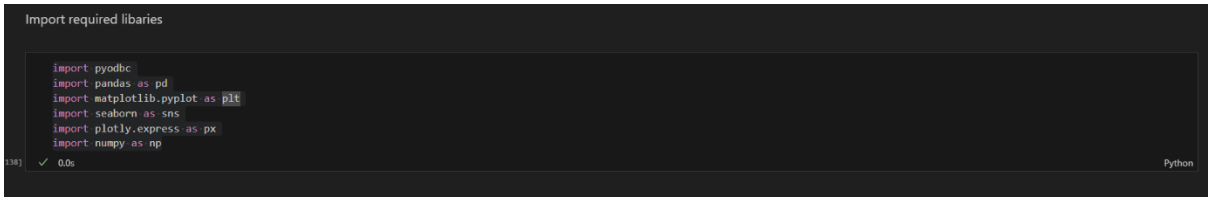
SQL SSMS Source Code:

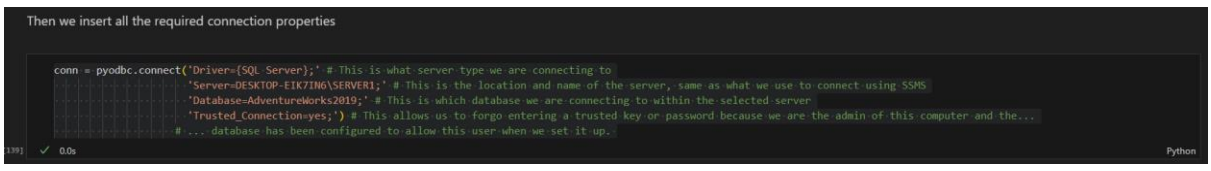
```
CREATE VIEW v_sales_vacationhours_bonus AS
SELECT E.OrganizationLevel, E.JobTitle, E.VacationHours, S.Bonus
FROM HumanResources.Employee AS E
INNER JOIN Sales.SalesPerson AS S
ON E.BusinessEntityID = S.BusinessEntityID ;
```

Result:

OrganizationalLevel	JobTitle	VacationHours	Bonus
2	North American Sales Manager	14	0
3	Sales Representative	38	4100
3	Sales Representative	27	2000
3	Sales Representative	24	2500
3	Sales Representative	33	500
3	Sales Representative	29	6700
3	Sales Representative	22	5000
3	Sales Representative	26	3550
3	Sales Representative	31	5000
3	Sales Representative	23	3500
3	Sales Representative	39	3900
2	Pacific Sales Manager	20	0
3	Sales Representative	36	5650
2	European Sales Manager	21	0
3	Sales Representative	35	75
3	Sales Representative	37	5150
3	Sales Representative	34	985

Python Source Code:





Now we create the cursor

```
cursor = conn.cursor()
```

[140] ✓ 0.0s Python

Here we define what query we want executed in the database

```
query = 'SELECT E.OrganizationLevel, E.JobTitle, E.VacationHours, S.Bonus FROM HumanResources.Employee E INNER JOIN Sales.SalesPerson S ON E.BusinessEntityID = S.BusinessEntityID'
```

[141] ✓ 0.0s Python

Here we assign the resulting table returned from the database as a variable 'salesdata'. This is also where our connection actually runs using the defined properties and query above

```
salesdata = pd.read_sql(query, conn)
```

[142] ✓ 0.0s Python

Let's get a preview of what our data looks like:

+ Code + Markdown

```
print(salesdata.head())
```

[143] ✓ 0.0s Python

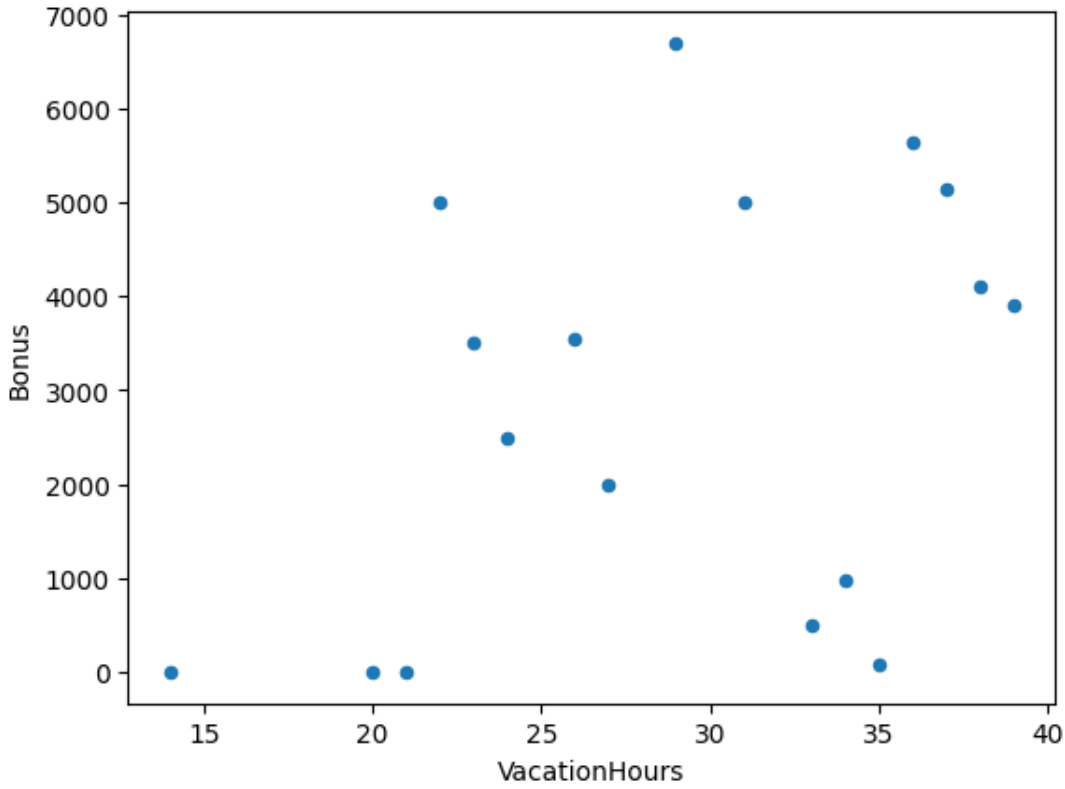
```
   OrganizationLevel                  JobTitle  VacationHours   Bonus
0                  2  North American Sales Manager            14     0.0
1                  3         Sales Representative            38  4100.0
2                  3         Sales Representative            27  2000.0
3                  3         Sales Representative            24  2500.0
4                  3         Sales Representative            33   500.0
```

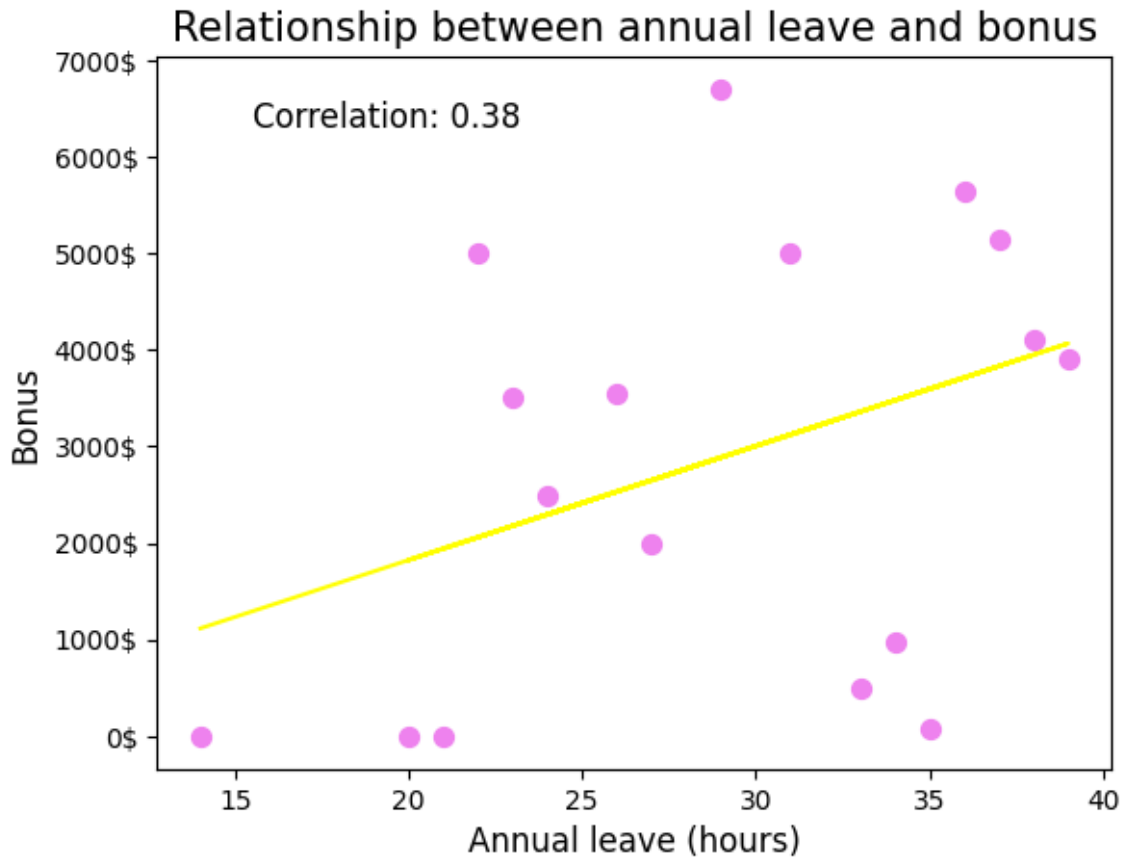
Lastly i create a scatter plot using two of the fields returned from the database

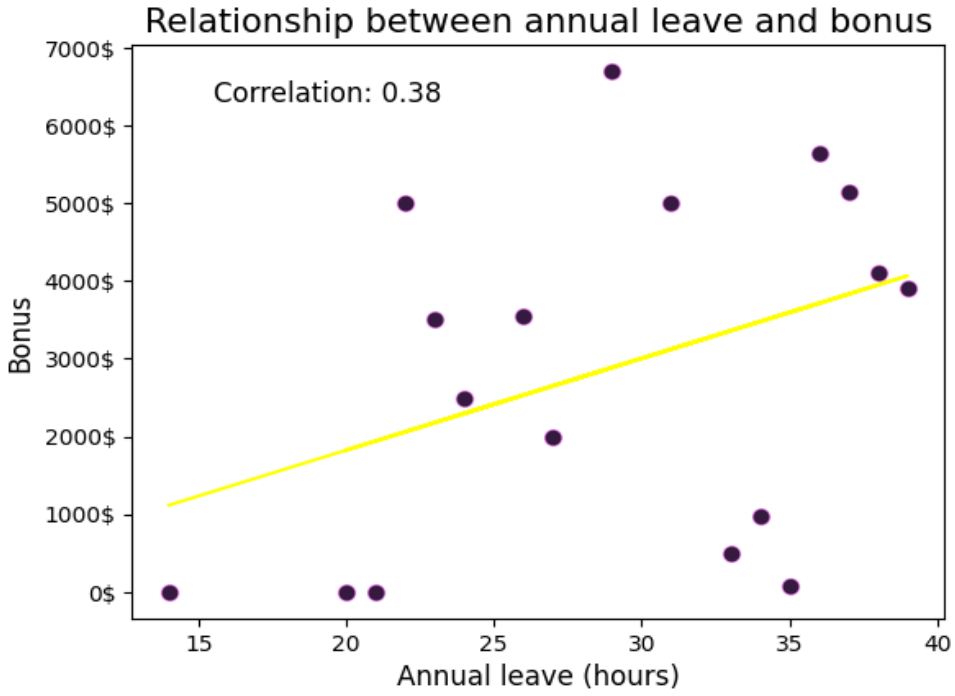
```
# Create the scatter plot
salesdata.plot.scatter(x='VacationHours', y='Bonus', color='violet', s=50)

# Add labels and title
plt.title('Relationship between annual leave and bonus', fontsize=15)
plt.xlabel('Annual leave (hours)', fontsize=12)
plt.ylabel('Bonus', fontsize=12)

# Customize y-axis ticks and labels
plt.yticks([0, 1000, 2000, 3000, 4000, 5000, 6000, 7000],
           ['0$', '1000$', '2000$', '3000$', '4000$', '5000$', '6000$', '7000$'])

# Calculate correlation coefficient
correlation = salesdata['VacationHours'].corr(salesdata['Bonus'])

# Annotate the scatter plot with the correlation coefficient
plt.annotate('Correlation: {:.2f}'.format(correlation), xy=(0.1, 0.9),
             xycoords='axes fraction', fontsize=12)

# Set the 'rocket' color palette
sns.set_palette("rocket")

# Create the scatter plot
plt.scatter(x='VacationHours', y='Bonus', data=salesdata)

# Add a line of best fit
m, b = np.polyfit(salesdata['VacationHours'], salesdata['Bonus'], 1)
plt.plot(salesdata['VacationHours'], m * salesdata['VacationHours'] + b, color='yellow')

# Get the 'rocket' color palette as a list of RGB tuples
colors = sns.color_palette("rocket", len(salesdata))

# Assign colors to specific elements
plt.scatter(x='VacationHours', y='Bonus', data=salesdata, c=colors)
```

```
# Show Plot

plt.show()
```

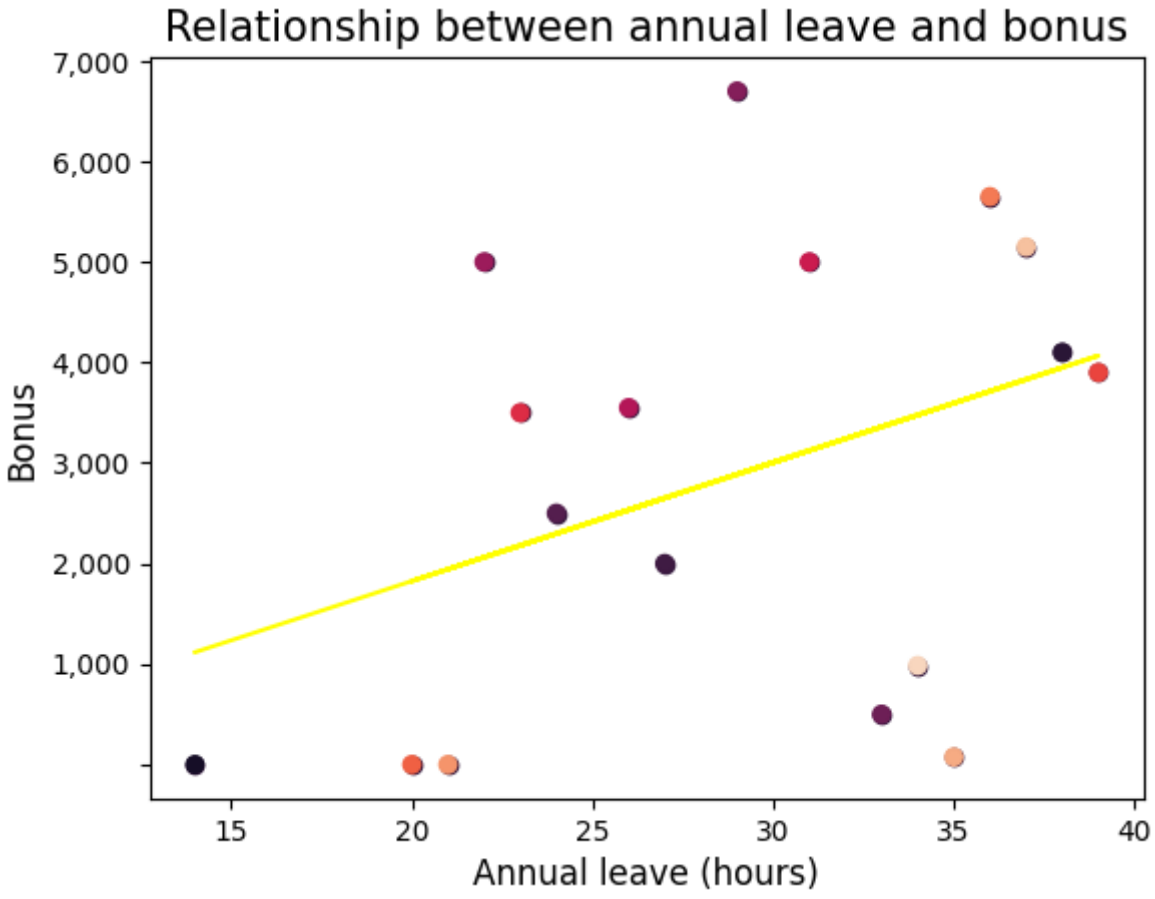
Graph Result Progress:






Relationship between annual leave and bonus
Correlation: 0.38
7000$
6000$
5000$
4000$
3000$
2000$
1000$
0$
Bonus
15
20
25
30
35
40
Annual leave (hours)

Final Graph:


Relationship between annual leave and bonus
7,000
6,000
5,000
4,000
3,000
2,000
1,000
Bonus
15
20
25
30
35
40
Annual leave (hours)

The purpose of this report is to analyse the relationship between annual leave and bonuses for employees, using data from the AdventureWorks2019 database. The objective is to provide stakeholders with insights into how annual leave allocation impacts bonus payments.

Analysis:
This analysis involved querying the AdventureWorks2019 database to retrieve data on annual leave and bonuses.

Results:
The analysis found a strong correlation between vacation hours and bonuses that employees receive. From this we can conclude that vacation hours have a positive effect on the performance of employees, as well as on their bonuses paid for work. Employees who have minimum vacation hours either receive no bonuses at all or receive very few bonuses. There is also an inverse correlation - people who rest more are more motivated to work and to do it in less time, bringing profit to the company and receiving bonuses.

Conclusion:
Based on the analysis, it can be concluded that there is a positive relationship between annual leave and bonus payments. The analysis clearly advises the company to evaluate its vacation hours policy, since the relationship between performance and bonuses is quite strong. This will also help improve employee motivation and the quality of their work. Further investigations and adjustments based on these findings can contribute to the overall success and well-being of the workforce.

My Analysis:
The scatter plot illustrates a **weak positive correlation** between the number of annual leave hours taken and the bonus amount. This means that as the number of annual leave hours increases, there's a slight tendency for the bonus amount to also increase. However, the relationship is not strong, as indicated by the scattered distribution of data points.

Observations:

- **Data Points:** The data points are spread out, suggesting a weak linear relationship. There's no clear, well-defined pattern.
- **Outliers:** There are a few data points that deviate significantly from the general trend. These outliers might be influencing the overall correlation.
- **Correlation Coefficient:** The correlation coefficient of 0.38 confirms the weak positive relationship. A value closer to 1 would indicate a stronger positive correlation.
-

Analysis:

- **Limited Impact:** The weak correlation suggests that annual leave taken is not a major determinant of bonus amounts. Other factors, such as individual performance, job role, or company-wide performance likely play a more significant role.
- **Potential Reasons for Weak Correlation:**
    - **Different Departments:** Employees in different departments might have varying leave policies and bonus structures.

    - **Individual Performance:** Individual performance metrics, such as sales targets or project completion, might be more influential in determining bonuses.
    - **Company Performance:** Overall company performance could significantly impact bonus distribution, regardless of individual leave hours.

Colour Analysis in regard to the graph:

The colour palette used in the scatter plot enhances the visual appeal and might help identify potential clusters or patterns within the data. Here's a breakdown of how the colours might be interpreted:

Colour Interpretation:

- Darker Colours (Purple, Dark Red): These colours likely represent data points that deviate from the general trend. They could be outliers or data points that have a relatively higher bonus compared to their annual leave hours.
- Lighter Colours (Orange, Yellow): These colours might represent data points that are closer to the line of best fit. They might have a more typical relationship between annual leave and bonus.

Relationship to Annual Leave and Bonus:

- Higher Bonus, Lower Leave: Data points in darker colours with lower annual leave hours but higher bonuses might indicate employees with exceptional performance or those in critical roles who receive higher bonuses despite fewer leave hours.
- Lower Bonus, Higher Leave: Data points in lighter colours with higher annual leave hours but lower bonuses might indicate employees who have taken more leave but whose performance or role might not warrant a higher bonus.

Overall, the colour palette helps visually differentiate between data points with varying relationships between annual leave and bonus. It adds context to the scatter plot, making it easier to identify potential trends and outliers.

In conclusion, while there is a slight positive correlation between annual leave and bonus, it is not a strong relationship. Further investigation is needed to uncover the underlying factors driving bonus distribution.