Question 1.3: Does the dataset contain any NaN values? Explain what you did to reach your conclusion.

To check for any NaN values, I sum the NaN values in each column with `unemployment_df.isna().sum()`.

```
In [7]: print(unemployment_df.isna().sum())
```

```
Date                    0
Male Unemployment       0
Female Unemployment     0
Overall Unemployment    0
dtype: int64
```

As seen from the printout – there are *no* NaN values in the dataset.

Question 2.2: What kind of trend in unemployment rate can you find from the graph above? Please provide potential reasonings of the trend you find from the graph above.

The graph shows a clear cyclical trend with peaks and troughs most likely corresponding with the economic recessions and expansions. We see a clear increase in unemployment following the 1979 energy crisis, the 1990 oil price shock, the 2008 burst of the housing bubble, and the COVID-19 crisis — all of which events triggered a recession in the following years.

Question 2.4: What kind of differences between the sexes do you see in the plot above? Why may they have occured?

Interestingly, before 1980, women generally seemed to have a higher unemployment rate than men. However, this changed after the 1980s, when females had a similar unemployment rate, and during recessions, their unemployment rate was lower than that of men.

There are several potential reasons for the higher level of unemployment among women before 1980. Women's labour force participation increased in this period, making it possible that their comparative lack of education, experience, and discrimination made their employment less secure.

A reason for the lower unemployment rate during recessions in recent years is likely that women working in sectors less affected by economic fluctuations. At the same time, there is a higher concentration of men in industries like construction and manufacturing, which are usually hit harder during economic downturns.

Question 2.8: We briefly spoke about the Phillip's Curve in lab 2; read more here. Does the graph above roughly match what you would expect from the curve? Why or why not?

The Phillip's Curve generally predicts an inverse relationship between unemployment and inflation – or, more often, inflation expectations. This seems to be partly true, e.g., around 2010, when inflation was relatively high until the housing bubble burst, and inflation and unemployment moved in opposite directions.

However, during the stagflation in the 1970s, we do not have this relationship; instead, both inflation and unemployment rose.

Question 2.13: We defined an `alpha` parameter for you in the function call above. What is that parameter, what does it do and which problem does it help avoid?

Hint: Try removing/changing the alpha value and see how that changes the plot.

The `alpha` parameter adjusts the opacity of the individual point. It takes in values `alpha`$\in [0, 1]$ with a default of 1. Points are entirely transparent when `alpha`$= 0$ and completely opaque with `alpha`$= 1$. When creating scatterplots, setting `alpha`$\neq 1$ helps show overlapping points, e.g. are there in this scatterplot two overlapping points from 1990 around $(7.25\%, 3\%)$ that would not have been as visible if `alpha`$= 1$.

Question 3.3: What do you notice from the violin plot above? (Distribution, skewness, etc.) Please give the reasoning to your answer.

The unemployment data shows a wider distribution, indicating that unemployment rates are centred around $5\% - 6\%$. The shape is somewhat symmetrical but slightly right-skewed, implying that there are fewer instances of extremely low unemployment rates.

The inflation data shows that, most often, inflation is just above the FED target of 2%; however, there are many more examples of low and high inflation. Again, the distribution is right-skewed, indicating occasional periods of very high inflation.

Question 3.5: What does a kernel density estimate try to do? How does it work?

The Kernel Density Estimate (KDE) tries to estimate the underlying smooth the distribution of the data using a continuous probability density curve. Rather than using discrete bins as the histogram, the KDE smooths the observations with a Gaussian kernel, producing a continuous density estimate