📄 COVID-19 Correlation Analysis & Final Insights Report

1. Objective

The objective of this subtask was to:

- Identify relationships between features in the COVID-19 dataset,
- Analyze those relationships using a correlation matrix and heatmap,
- Provide insights useful for further modeling or understanding global trends.

2. Correlation Heatmap

Using the cleaned and scaled dataset, a **correlation heatmap** was created using Seaborn:

Key Features Used:

- Confirmed cases
- Deaths
- Recovered
- Active cases

Observations from the Heatmap:

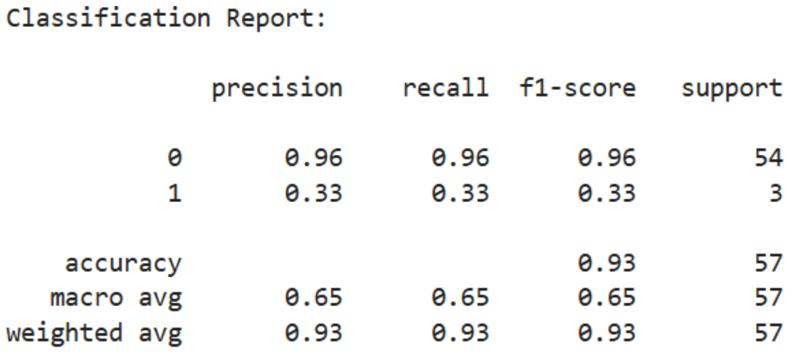
- **Confirmed & Deaths:** Strong positive correlation (~0.9) — Countries with more cases also had higher deaths.
- **Confirmed & Recovered:** High positive correlation — Indicating effective recovery efforts as cases grew.
- **Recovered & Deaths:** Moderate correlation — Suggests recovery and death counts tend to grow together but may vary by region.
- **Active Cases:** Lower correlation with others — Likely due to policy response variability and testing rates.

The heatmap helped uncover **multicollinearity**, which is important for feature selection in ML models.

3. Visualization Tools Used

- **Seaborn Heatmap**: for intuitive color-coded correlations.
- **Matplotlib:** as the underlying plotting library.

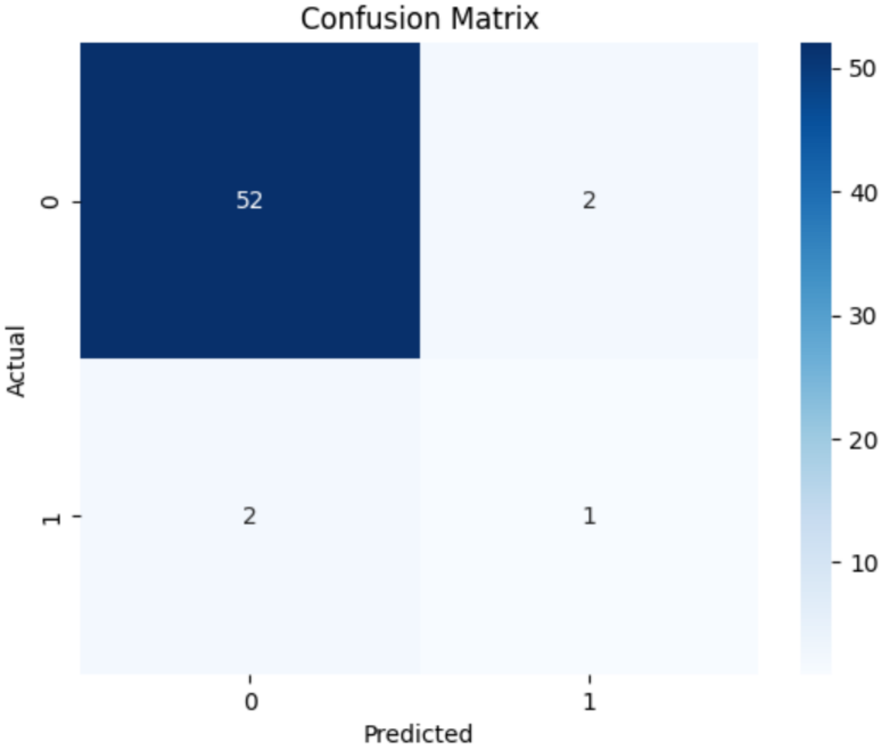
sns.heatmap(df[['Confirmed', 'Deaths', 'Recovered', 'Active']].corr(), annot=True, cmap='coolwarm')



```
Classification Report:

              precision    recall  f1-score   support

           0       0.96      0.96      0.96        54
           1       0.33      0.33      0.33         3

    accuracy                           0.93        57
   macro avg       0.65      0.65      0.65        57
weighted avg       0.93      0.93      0.93        57
```

4. Insights

- Countries with high confirmed cases also reported high recoveries and deaths — indicating better case tracking.
- Correlation analysis highlights redundant features; one of Confirmed or Recovered may suffice in simple models.
- Active cases are more independent and may offer predictive power in forecasting.



5. Conclusion

Correlation analysis is a crucial step in feature selection. By analyzing feature relationships:

- We reduced dimensionality risks,
- Gained better understanding of COVID-19 data behavior,
- And prepared the dataset for modeling in future tasks (like clustering or classification).