BONUS QUESTIONS-Task-2

1.Why is data cleaning important in real-time data processing?

Data cleaning ensures that incoming data streams are accurate, consistent, and reliable. In real-time processing, even small errors like missing values, duplicate records, or incorrect formats can quickly propagate and affect dashboards, analytics, or machine learning models. Cleaning improves data quality, reduces noise, and enables faster, more trustworthy decision-making.

2. What are pipeline artifacts and how are they used in DevOps workflows?

Pipeline artifacts are files or outputs (such as logs, reports, or datasets) generated during a CI/CD pipeline run. In DevOps workflows, artifacts are stored and shared across different stages of the pipeline. For example, a build stage may produce an artifact (like a cleaned dataset or a compiled app), which the deployment stage can then use. This ensures reproducibility, traceability, and smooth handoff between pipeline steps.

3. How would you modify the pipeline to store the cleaned data into Azure Blob Storage?

To store the cleaned data in Azure Blob Storage, the pipeline needs to include a step that connects to Azure and uploads the file. This is done by creating a service connection between Azure DevOps and the Azure Storage account. Once authenticated, the cleaned CSV file can be uploaded to a specific container in Blob Storage. This ensures that processed data is securely stored, easily accessible for downstream applications, and can be integrated into analytics or reporting workflows.