AZURE ML HANDS ON :-

Q . 1.

A – data collection -Gather and collect relevant data from various sources. This can include databases, files, APIs, or streaming data.

B – Data Ingestion – import the collected data into Azure Machine Learning. Azure provides services like Azure Data Factory or Azure Databricks for data ingestion.

C – Data Exploration - Perform initial data exploration to understand its characteristics. Examine data statistics, distributions, and visualize it to identify patterns

D – Data Cleaning – Cleanse the data by handling missing values, outliers, and inconsistencies .

E – Data Transformation – Transform the data into a format suitable for model training

F – Data splitting - Divide the dataset into training, validation, and test sets

G – Feature Engineering - Create new features or modify existing ones to improve model performance.

H – Data Pipeline - Build a data pipeline to automate data preprocessing steps. Azure offers tools like Azure Data Factory and Azure Databricks to create and manage pipelines.

I - Validation –Continuously monitor and validate the quality of your data

Q. 2.

Splitting is important because :-

--- it helps in model evaluatiuon

--- it helps in preventing overfitting

--- it helps in hyperparameter tuning

--- it helps in benchmarking

Q. 3.

I have used boosted decision tree model as the data set is in numeric and it is in regression.

Q. 4.

Hyperparameters are settings or configurations that are not learned from the training data but are set before training begins. These settings control aspects of the model's training process and architecture and significantly impact the model's performance and generalization ability.

