**Summary of Assumptions, Challenges and my Approach -**

The project involved building a complete data processing pipeline on AWS, including:

* Dockerizing a Flask-based API to handle S3 uploads.
* Triggering a Lambda function to transform and summarize data using pandas.
* Storing cleaned and summarized files back to S3.
* Performing analytics and visualizing results via Amazon QuickSight.
* Mini report in jupyter

Approach

Since I already have an account on AWS so I directly started from creating EC2 instance connecting it through cmd and used pre installed python(3.12 since it failed for installing 3.6 on ubuntu free version) and the install docker and move further.

**Flask API with Docker**

Created a simple Flask app (restapi.py) to accept file uploads and directly store them in the S3 bucket using boto3. I wrote a Dockerfile to containerize this app and used docker build and docker run to deploy it on an EC2 instance. API testing was done using curl.

**S3 Trigger & Lambda Integration**

A Lambda function (lambda\_function.py) was triggered automatically on S3 uploads. It performed the following:

Cleaned the dataset (removed nulls, renamed columns, filtered scores).

Calculated average impact scores and counts per group.

Uploaded both cleaned and summarized datasets back to S3.

**Data Exploration**

Downloaded transformed data and used Jupyter Notebook to perform further analysis like:

**Top 5 countries by Impact Score –**

Germany, China, UK, Japan, and Australia lead in AI adoption, showing the highest average Impact score.This indicates strong AI integration across industries in these nations –

**Country most affected by job loss due to AI –**

Japan recorded the highest potential job loss due to AI.  
This reflects greater automation risks in its rapidly growing sectors.

**Year-over-year trend of Impact Score (overall and by industry)**

The impact is start reflect more from 2023 and it continue to grow year on year showing the rapid involment of AI.

**QuickSight Dashboard**

Open 30-days free account then connected the S3 summary file to QuickSight and built various charts: bar chart, pie chart, line chart, and heatmaps also unable to created KPI Card so created it in jupyter file you can refer it there.

**Assumptions -**

The “AI Adoption Rate (%)” column was treated as the impact score.

Assumed that the output files generated by Lambda (cleaned and summarized CSVs) are the final authoritative versions used for downstream analytics (QuickSight) and Mini EDA.

WinSCP is used for transferring code file like restapi, lambdafile(zip file)

**Challenges & Troubleshooting -**

Docker API to Lambda Connection Failure: Initially, I tried triggering Lambda directly from the Flask container, but due to AWS permission complexities and cross-service integration, this route failed. Instead, I shifted to using S3 as a decoupling layer with an event trigger.

Lambda Code Upload Issue: Faced an issue where want to upload lambda zip file directly from EC2(since zip file created and store there) to AWS CLI. Eventually, downloaded the ZIP via WinSCP to my local system for Lambda upload.

After that face an issue when lambda.py file not open in lambda editor due to issue in corrupted zip all the contain and then upload again

QuickSight Complexity: Since I had never used QuickSight before, understanding its interface and getting multi-axis or grouped visualizations working took some exploration. KPI and heatmap configurations were especially non-intuitive.

Overall, I love to solve this project, it provide me overview of how various AWS services integrate and help me to strengthened my existing AWS hands-on. The debugging journey was very real and filled with honest problem-solving, which taught me more than any tutorial.