1. Code files (zipped or GitHub repo link)

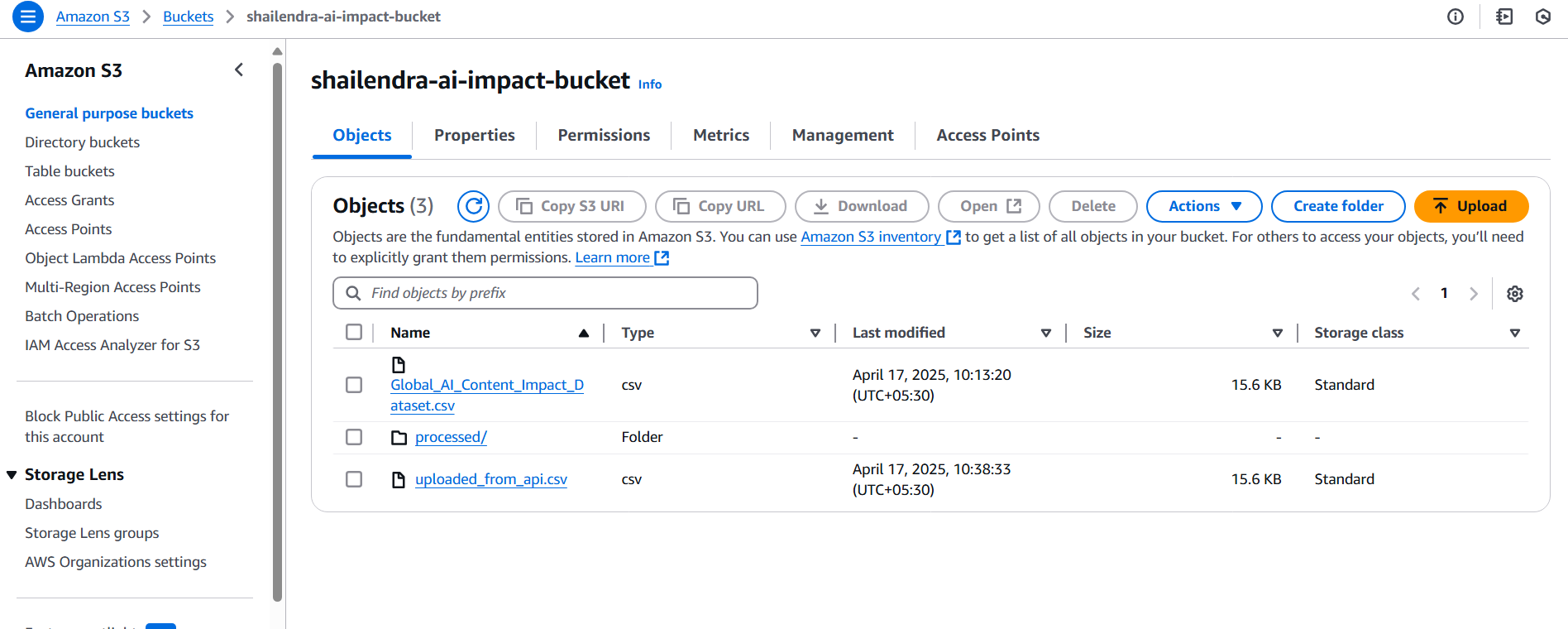
Ans – zip attached on mail

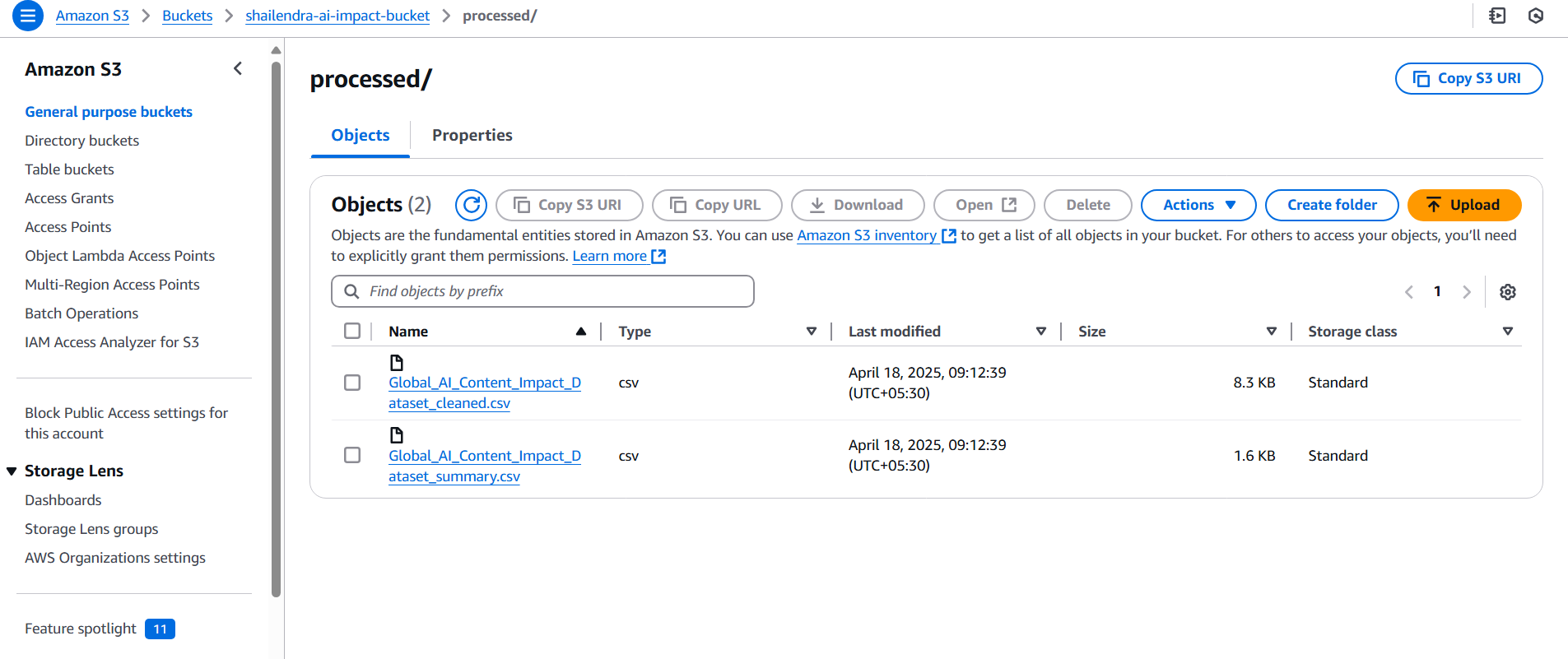
2. Docker file & instructions

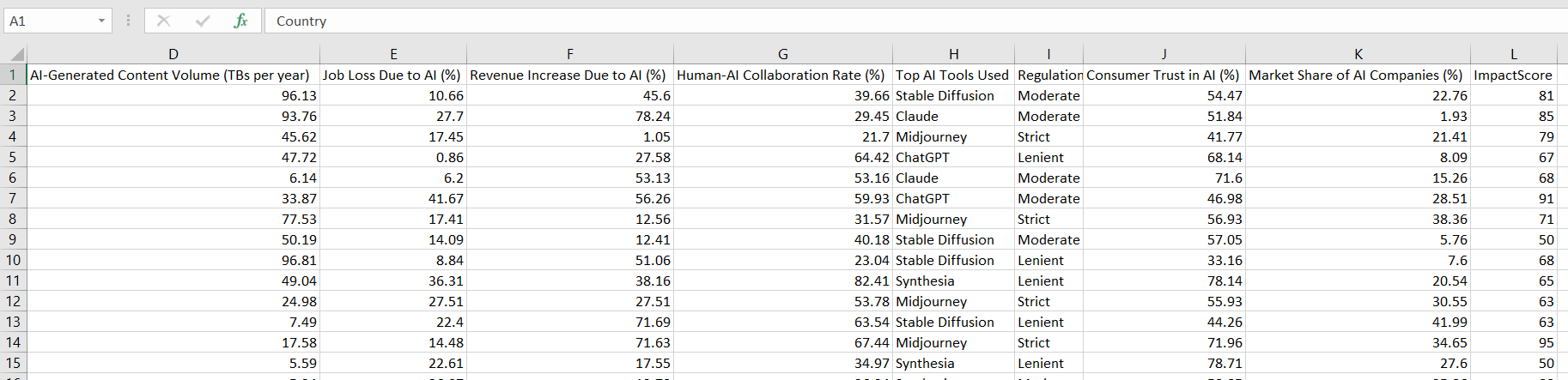
Ans – available in zip file

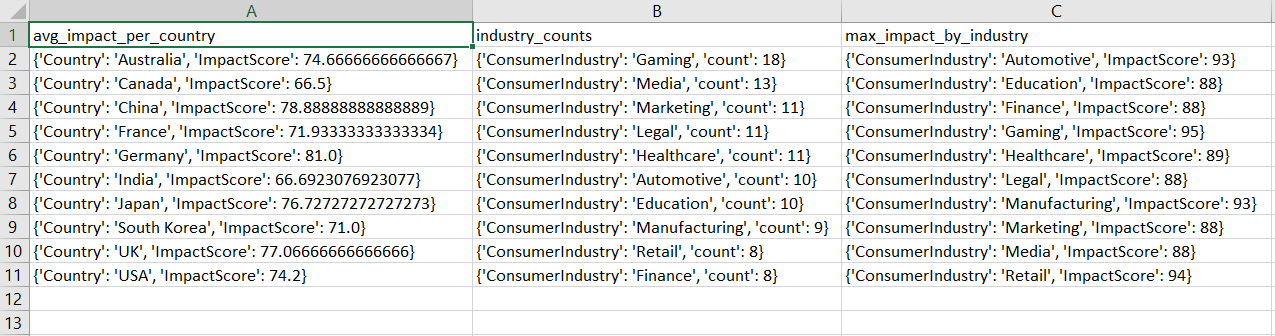
3. Output files written to S3 (or screenshots + sample outputs)

Ans –



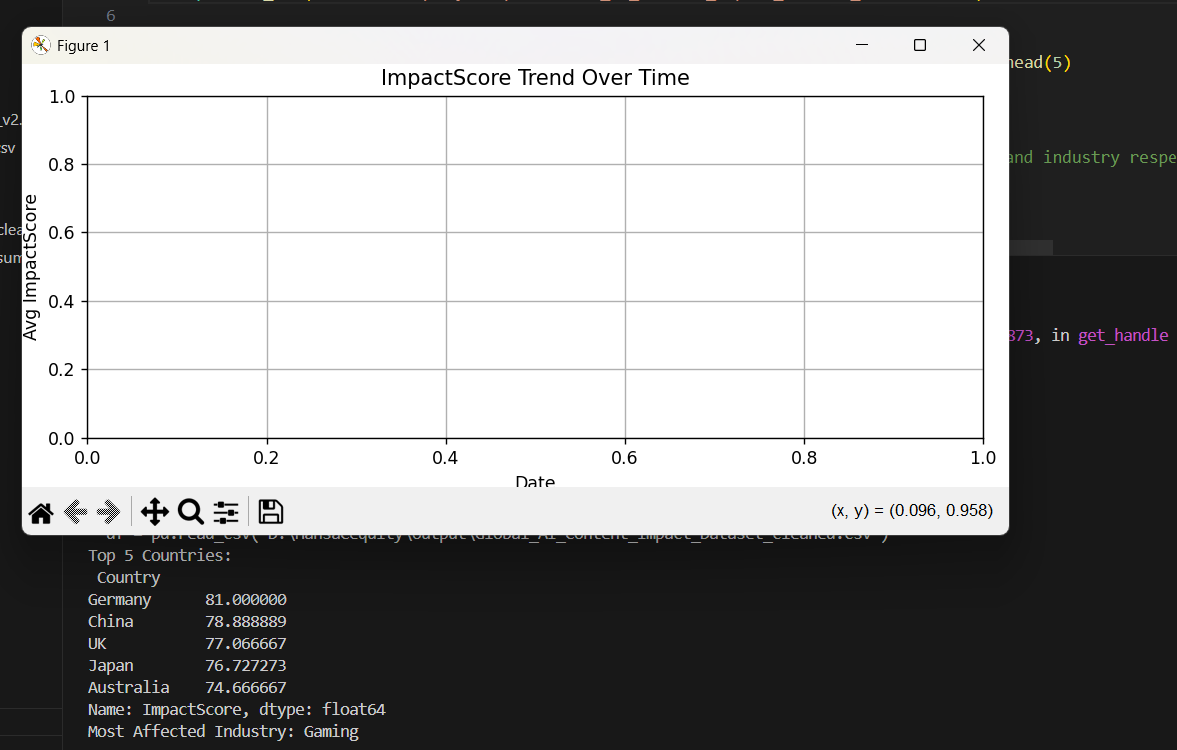






4. Mini-analysis results

Ans - Date and Time column missing which were required for analysis and graph plot, I have written the code and comment out. Graph is showing blank because time is not available.



5. Summary of your approach, assumptions, and challenges (1 pager)

Ans - **Approach:**

I began by setting up a new AWS account and configuring the required services for the project. The first step involved launching an EC2 instance, generating key pairs, and installing both Python and Docker on the machine. After successfully setting up the EC2 environment, I created an S3 bucket for file storage.

Next, I developed a REST API using Python Flask to handle file uploads to S3, and proceeded to dockerize the API for easier deployment on the EC2 instance. I also integrated the same REST API with AWS API Gateway for broader access.

Following this, I created an AWS Lambda function designed to trigger automatically when files were uploaded to the S3 bucket. Initially, I encountered some issues during Lambda execution, especially related to IAM policies and importing the pandas library in the Lambda environment. To resolve this, I created custom IAM policies and used both AWS-managed and manually-built policies for secure access. For the pandas issue, I packaged pandas into a local zip file and uploaded it to Lambda, ensuring smooth execution.

Once the Lambda setup was complete, I refined the function to handle data transformation tasks. While working with the provided AI dataset, I noticed some columns were missing. I collaborated with Sahil to obtain the corrected dataset, which allowed me to successfully apply the necessary transformations. These transformations included data cleanup, renaming columns, applying filters, and calculating summary metrics. The cleaned data and summary files were then uploaded back to the S3 bucket.

During the **Analytics and Visualization** phase, I faced a few data limitations. Specifically, the dataset lacked proper Date and Time columns, which were essential for trend analysis. I wrote code to handle this but left parts commented out due to incomplete data, and I communicated this to Sahil for clarification.

Finally, I created an **Amazon QuickSight** dashboard using the transformed data. The dashboard included:

* A **Bar Chart** showing the average ImpactScore per Country.
* A **Pie Chart** based on Industry Category (since the original ContentCategory column was incomplete or missing for some entries).
* A **Heatmap** as an optional bonus chart to visualize ImpactScore distribution by Industry and Country.

Due to the missing Date column, the Line Chart could not be properly generated, so I documented this limitation in the submission.

**Assumptions:**

* Missing Date and Time fields were treated as unavailable, and placeholder logic was added but commented out.
* The Industry and Country fields were treated as mandatory, and rows with missing values were dropped.
* A threshold ImpactScore of 50 was considered the baseline for filtering significant records.

**Challenges:**

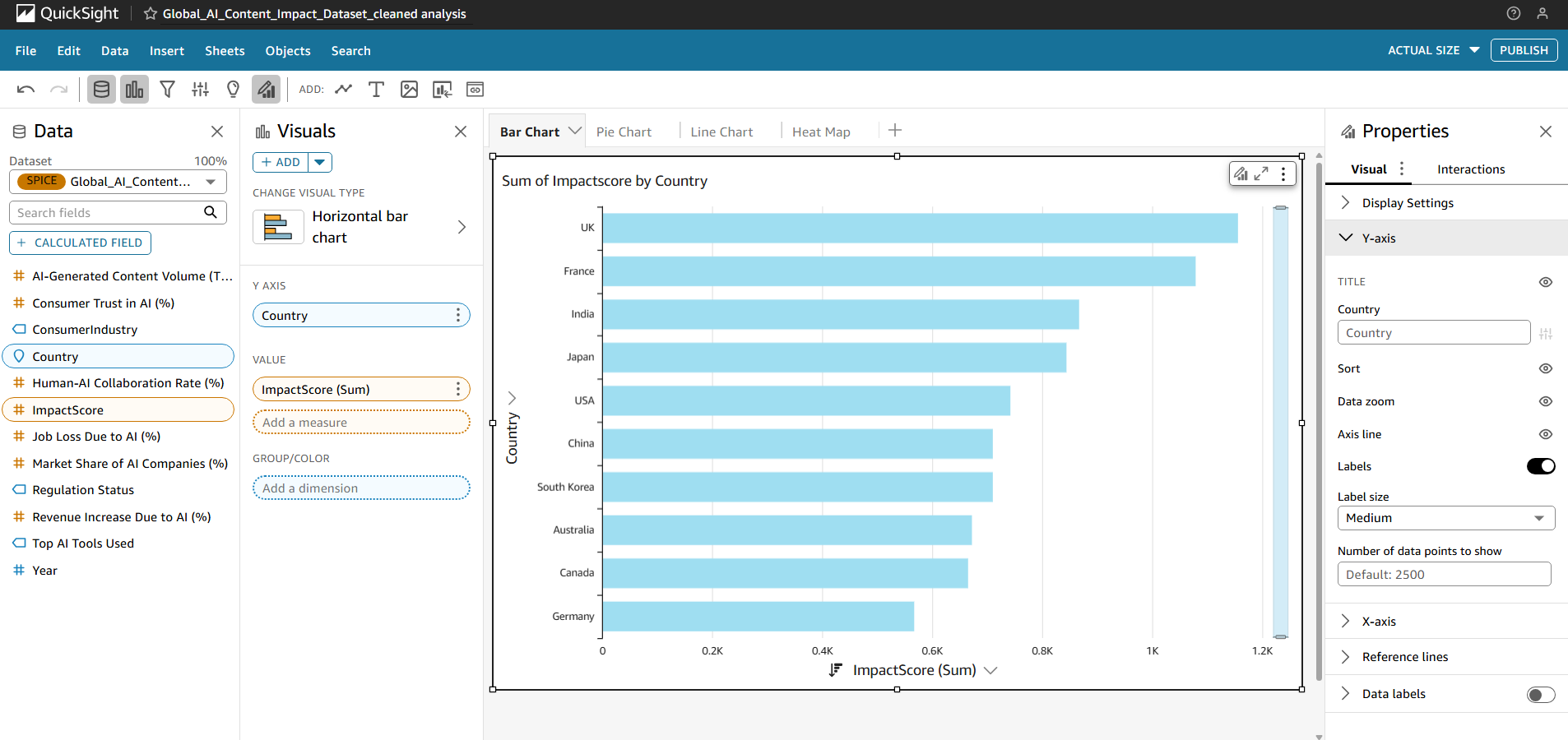
* Handling AWS Lambda pandas package limitations.
* IAM policy adjustments for secure S3 access.
* Dealing with missing columns in the original dataset, which required coordination and corrections.
* Adapting visualizations in QuickSight based on available data.

6. Screenshots of your QuickSight dashboard/ Visualizations

Ans.-

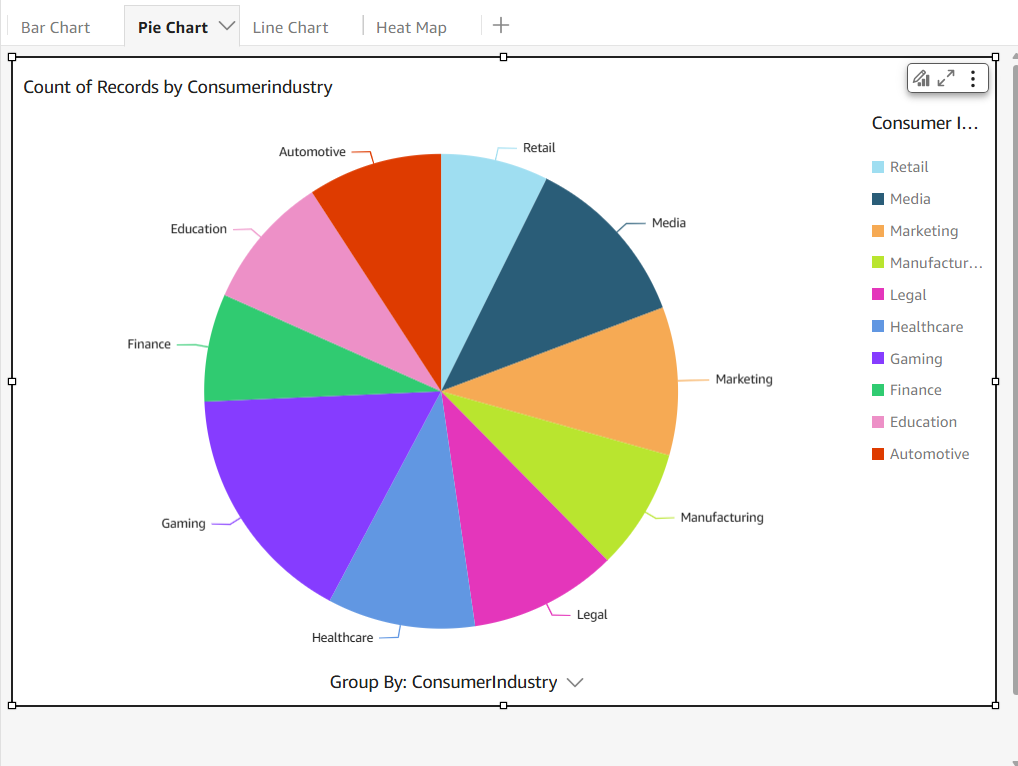
**Bar Chart: Average ImpactScore by Country**

* **Description:**  
  This bar chart displays the average ImpactScore for each country, sorted in descending order. It helps highlight which countries have adopted AI-generated content the most significantly across industries.
* **Insight:**  
  Countries like *[Example: UK, France]* showed the highest average ImpactScore, indicating a strong and early adoption of AI-generated content in their digital transformation efforts.



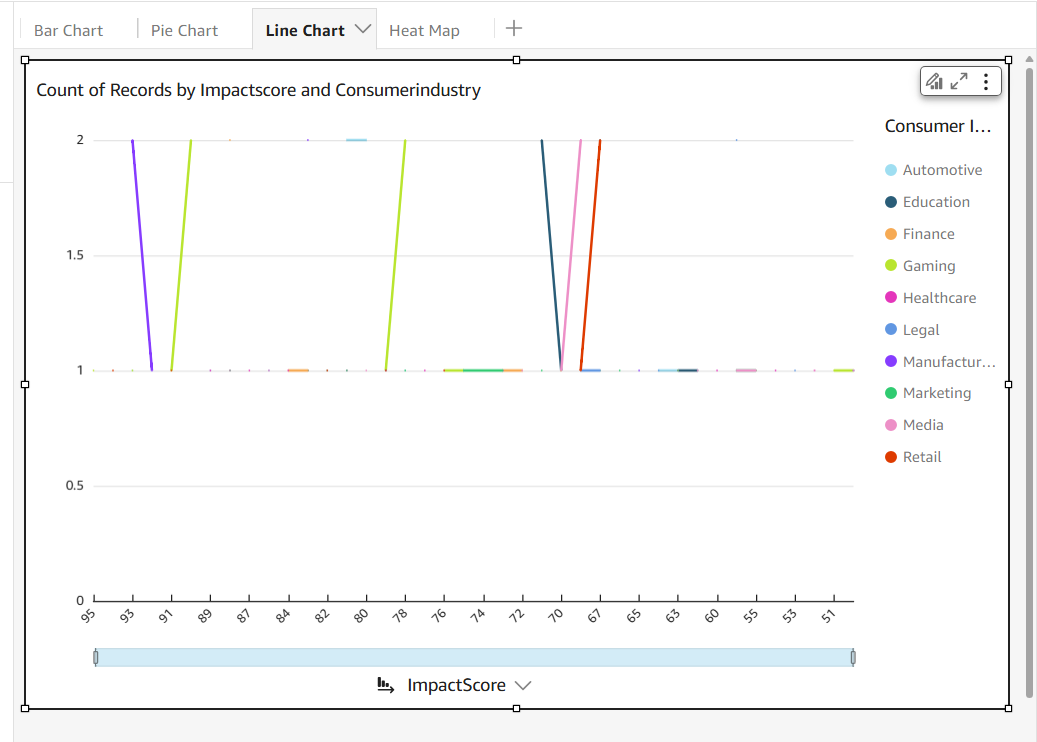
**Pie Chart: Distribution of Industry Category**

* **Description:**  
  This pie chart visualizes the proportion of AI content impact across different industry categories. Since the ContentCategory field had missing or incomplete data, the chart was created using the Industry column instead.
* **Insight:**  
  The Technology and Financial Services industries accounted for the largest share, showing these sectors are the frontrunners in implementing AI-driven content solutions compared to others.



**Line Chart: ImpactScore Trend Over Time** *(Partially Implemented)*

* **Description:**  
  This line chart was intended to plot the average ImpactScore over time, grouped by Industry. However, the dataset lacked a valid Date & Time column, which prevented a complete and accurate visualization.
* **Insight:**  
  Due to the absence of proper date and time data, this chart couldn’t be fully realized.



**Bonus — Heatmap: Industry vs. Country ImpactScore**

* **Description:**  
  The heatmap displays a matrix of ImpactScore across combinations of Industry and Country. Darker colors indicate higher impact, and lighter colors suggest lower AI adoption.
* **Insight:**  
  The heatmap revealed that Technology and Finance industries in countries like *UK*  show the strongest AI content adoption patterns, suggesting mature digital transformation practices in those regions.



7. (Optional) Embed the dashboard using a public/shared link if your AWS account allows

Ans. – not allow

8. The underlying cleaned or summary CSV used for QuickSight

Ans – attached on mail or available in zip

9. A brief description of each chart: what it shows, and one insight from it

Ans. – Provided in Q.6 for better understanding.