Exploratory Data Analysis (EDA) – Milestone 1

Project: Azure Demand Forecasting
Phase: Data Collection & Preprocessing
Team: Backend – B
Date: 30-Aug-2025

1. Data Sources

Primary Dataset: Azure Usage Data
- Daily consumption details for compute and storage resources
- Regional usage patterns across multiple Azure regions
- Metrics for resource utilization and user activity

Secondary Dataset: External Influencers
- Economic performance indicators
- Cloud market demand metrics
- Holiday and weekend markers impacting usage

2. Key Insights from Data Analysis

Usage Behavior
- Data examined from the consolidated file: Merged_Data.csv
- Coverage: East US, West US, North Europe, Southeast Asia

CPU Usage:
- Significant variations in daily averages across regions
- Clear time-series patterns for different resource categories

Storage Utilization:
- Regional comparison indicates wide distribution differences
- Boxplot analysis highlights variation by resource category

User Engagement:
- Active users show distinct regional trends
- Usage spikes align with business activity patterns

External Factors:
- Strong correlation between economic indicators and demand
- Holiday periods impact system utilization noticeably

3. Data Quality Review

- Missing Values: None
- Duplicate Records: Eliminated
- Consistency: Verified
- Outliers: Identified and treated

4. Visualizations Generated

Resource Usage Trends
- CPU utilization trends over time
- Average CPU usage by region
- Storage usage distribution by region and resource type

User Behavior
- Regional comparison of active user counts

Impact of External Indicators
- Economic index vs demand scatter plots
- Holiday impact analysis

Correlation Heatmap
- Strong relationships between CPU, storage, and user activity

5. Data Merging & Processing

- Successfully integrated usage data with external factors
- Temporal alignment maintained
- No information loss during processing

6. Final Deliverables

- Cleaned dataset: data/processed/cleaned_merged.csv
- Visual reports saved in visualizations/ directory
- Jupyter/Colab notebook with full EDA workflow

Technical Summary

- Preprocessing completed for modeling phase
- Baseline usage trends and relationships documented