Course: Introduction to Big Data Analytic INSY 8413
Instructor: Eric Maniraguha | eric.maniraguha@auca.ac.rw | LinkedIn Profile
Assignment Date: 20 July 2025
Groups: A, B & E

Assignment I: Uber Fares Dataset Analysis using Power BI

Assignment Type: Data Analysis Project
Tool: Power BI Desktop
Dataset: Uber Fares Dataset (Kaggle)
Deadline: Friday before Sabbath begins
Submission Method: eric.maniraguha@auca.ac.rw

Objective:

Analyze the Uber Fares Dataset to gain comprehensive insights into fare patterns, ride durations, and key operational metrics. Students will develop an interactive Power BI dashboard and present findings through a detailed analytical report.

Instructions:

Assignment Requirements

1. Data Understanding and Preparation

a. Download the Uber Fares Dataset from Kaggle
b. Load the dataset into a Pandas DataFrame using Python
c. Perform comprehensive exploratory data analysis (EDA) to understand:

- Dataset structure and dimensions
- Data types and variable descriptions
- Initial data quality assessment

d. Handle missing values and clean the data for analysis
e. Export the cleaned dataset as a CSV file for Power BI import

2. Exploratory Data Analysis (EDA)

a. Generate descriptive statistics including:

- Mean, median, mode, standard deviation

- Quartiles and data ranges
- Outlier identification

b. Create visualizations showing fare distribution patterns
c. Analyze relationships between key variables:

- Fare amount vs. distance traveled
- Fare amount vs. time of day
- Additional relevant correlations

3. Feature Engineering

a. Create new analytical features such as:

- Hour, day, month extracted from timestamps
- Day of week categorization
- Peak/off-peak time indicators

b. Identify and properly encode categorical variables
c. Save the enhanced dataset with new features for Power BI import

4. Data Analysis in Power BI

a. Import the cleaned and enhanced dataset into Power BI Desktop
b. Create comprehensive visualizations analyzing:

- Fare patterns across different time intervals
- Hourly, daily, and monthly ride patterns
- Seasonal trends and variations

c. Identify and highlight the busiest periods for Uber rides
d. If available, investigate weather impact on fare patterns

5. Dashboard Creation in Power BI

a. Design an interactive and professional dashboard including:

- **Distribution of fares:** Histograms and box plots
- **Ride durations:** Time-based analysis
- **Time series analysis:** Temporal patterns and trends
- **Geographic distribution:** Spatial analysis of rides

b. Ensure dashboard interactivity with filters and drill-down capabilities

c. Apply consistent formatting and professional design principles

6. Report Writing

a. Prepare a comprehensive analytical report with the following sections:

- **Introduction:** Project overview and objectives
- **Methodology:** Data collection and analysis approach
- **Analysis:** Detailed findings and statistical insights
- **Results:** Key discoveries and pattern identification
- **Conclusion:** Summary of main findings
- **Recommendations:** Data-driven business suggestions

Submission Requirements

Deliverables

1. **Power BI Dashboard File (.pbix)**
    - Interactive dashboard with all required visualizations
    - Professional formatting and user-friendly design
2. **GitHub Repository** (Public Access Required)
    - Cleaned datasets (CSV files)
    - Screenshots documenting your analysis process
    - README file explaining your project structure
3. **Documentation Screenshots**
    - Data loading process
    - Data cleaning steps
    - DAX formulas (if used)
    - Dashboard development stages
4. **Final Report** (Choose one format)
    - **Option A:** Comprehensive GitHub report (Markdown format)
    - **Option B:** PowerPoint presentation including:
        - Dataset description and sources
        - Data cleaning methodology
        - Key visualizations

- Insights and outcomes

Submission Instructions

1. Create a **public GitHub repository** for your project
2. Upload all deliverables to the repository
3. Email your GitHub repository link to the instructor
4. **Deadline:** Friday 25 July 25 5h00 pm, before Sabbath begins

Academic Integrity Notice

Important: While seeking help and collaboration is encouraged, all submissions must demonstrate original work and innovation. Similar or plagiarized submissions will result in grade reduction. Students must add unique insights and analytical approaches to distinguish their work.

Success Tips

- Start early to allow time for thorough analysis
- Document your process with clear screenshots
- Focus on actionable business insights
- Ensure your dashboard tells a compelling data story
- Test all interactive elements before submission

Wishing you all the best in this course!