Project 1 – Milestone 1

Topic –

- This project will analyze productivity patterns among remote workers to identify key factors that influence work performance in remote settings.

Business Problem –

- As remote work becomes more prevalent, organizations may struggle to maintain and enhance employee productivity outside of traditional office environments. This project will help to answer the following questions:
    - What factors most significantly impact the productivity of remote workers?
    - How do work habits, environments, and personal characteristics correlate with productivity outcomes?
- Understanding these patterns can help businesses develop better remote working policies, optimize employee performance, and enhance job satisfaction.

Datasets –

- The data set for this project comes from a simulated survey-based dataset titled 'remote_worker_productivity_1000'
- Some of the variables/ attributes in this set include:
    - Demographic Information
        - Age
        - Gender
        - Education
    - Work Environment Characteristics
        - Workspace Setup
        - Internet Reliability
    - Work Habits
        - Hours Worked
        - Breaks Taken
        - Meeting Frequency
    - Self-Reported Product Scores
- This dataset allows for exploratory and predictive analysis of remote worker performance based on diverse personal and environmental factors

Methods –

- I plan to use a few different types of analysis/ models to answer these questions
    - Exploratory Data Analysis (EDA) –
        - Exploring the data will help to understand data distributions and detect patterns
    - Correlation Analysis –
        - To help identify any relationships between different variables
    - Regression Modeling –
        - Linear Regression
        - Random Forests Modeling
        - These will help to predict productivity
    - Clustering –
        - This will help segment workers into groups based on productivity drivers
    - Visualization –
        - Bar Plots
        - Scatterplots
        - Heatmaps
        - This will help to clearly communicate findings to the audience
- These may change or be altered as the analysis is developed

Ethical Considerations –

- Privacy Concerns
    - This analysis must still ensure employee privacy and anonymity
- Bias in Analysis –
    - Predictive models could perpetuate biases if demographic variables like gender or age are overemphasized without considering fairness
- Interpretation Risks –
    - Misinterpretation of productivity metrics could lead to unjustified assumptions about worker capabilities or needs

Challenges/Issues –

- Data Quality –
    - Handling missing, inconsistent, or biased data could present challenges
- Causality vs Correlation –
    - Establishing true causality between factors and productivity may not be possible with observational data

- Generalization –
    - The dataset may not represent all industries or job types, limiting broader applicability