**Summary: Data Preparation Notebook (Jupyter notebook 01)**

1. **Workflow:**

• The data preparation process involved importing the dataset, inspecting its structure, checking for missing values, and performing initial data cleaning to ensure data quality. This step set the foundation for reliable analysis in later stages of the project.

1. **Aim of the Project/Analysis:**

• The primary aim of this project is to analyze various factors affecting mental health by preparing the data for exploratory and in-depth analysis. The goal is to identify trends that could help understand mental health issues across different demographics.

1. **Survey Period:**

• The survey includes a “Timestamp” column indicating the date and time when each response was collected. However, the exact length of the survey period isn’t explicitly stated in the notebook, which could be revisited if it’s crucial for the analysis.

1. **Aim of the Survey:**

• The survey aimed to collect data on mental health, focusing on factors like occupation, self-employment status, family history, and individual habits to understand their impact on mental well-being.

1. **Countries Involved:**

• The dataset contains responses from 35 countries, with the highest representation from the United States, providing a diverse set of data points to analyze global mental health trends.

1. **Respondents Involved:**

• The analysis involved a substantial dataset with 292,364 respondents, providing a comprehensive view of various factors influencing mental health.

1. **Main Factors Affecting Mental Health:**

• Key factors identified include gender, occupation, family history of mental illness, self-employment, and coping mechanisms, which are consistent with risk factors highlighted by the CDC.

1. **Conclusion:**

• The data preparation phase successfully cleaned and organized the dataset, laying a strong foundation for detailed analysis. It identified potential areas of interest that align with CDC guidelines on mental health, such as self-employment and family history being crucial factors.

1. **Recommendations:**

• Based on CDC guidelines, recommendations could include focusing on individuals with a family history of mental illness and those in high-stress occupations. Further analysis should explore targeted interventions for these groups.

1. **Further Investigation and Considerations:**

• Future steps might involve validating these findings against CDC data to ensure consistency and reliability. Further investigation into the exact survey period and additional demographic details would also strengthen the analysis.

**Summary: Exploratory Data Analysis Notebook (Jupyter Notebook 02)**

1. **Workflow:**

• This phase involved loading the cleaned dataset, exploring the data distributions, checking for missing or undefined values, and generating visualizations to identify trends. Tools like Matplotlib and Bokeh were used to create interactive plots for a detailed understanding of mental health patterns.

1. **Aim of the Project/Analysis:**

• The goal of this exploratory analysis is to discover underlying patterns and correlations within the data that may influence mental health, serving as a basis for more detailed analysis and actionable insights.

1. **Survey Period:**

• The exact survey period is not explicitly stated in the notebook. If the dates or the length of the survey are critical, it could be mentioned or considered for further investigation to understand the context of data collection.

1. **Aim of the Survey:**

• The survey aimed to collect data on a variety of mental health factors, including individual behaviors, occupation, stress levels, and social interactions, to understand their impact on mental well-being.

1. **Countries Involved:**

• The dataset includes responses from a wide range of countries, with a significant focus on the United States, which aligns with the CDC’s focus on analyzing mental health trends within the country.

1. **Respondents Involved:**

• The analysis covered a large number of participants, totaling 290,051 respondents, providing a robust dataset for examining trends and factors influencing mental health.

1. **Main Factors Affecting Mental Health:**

• The key factors identified include gender, occupation, self-employment status, family history of mental illness, and changes in personal habits, which are consistent with known risk factors highlighted by CDC data.

1. **Data Encoding for Machine Learning:**

• To prepare the dataset for potential machine learning applications, categorical variables should be converted into numerical formats using techniques like label encoding or one-hot encoding. This conversion will make the data more compatible with machine learning models, enabling more precise analysis and predictive capabilities.

1. **Conclusion:**

• The exploratory data analysis reveals significant insights into the factors impacting mental health. Many of these findings align with CDC guidelines on risk factors, reinforcing the importance of targeting specific demographic groups in mental health interventions.

1. **Recommendations:**

• Recommendations based on this analysis might include developing targeted strategies for at-risk groups, such as those with a family history of mental health issues or individuals in high-stress occupations. Aligning these strategies with CDC guidelines would strengthen their effectiveness.

1. **Further Investigation and Considerations:**

• To enhance the validity of the analysis, comparing these findings against data provided by the CDC would be beneficial. Future research could also focus on analyzing specific trends over time to understand how mental health conditions evolve.

**Summary: Data Transformation Notebook (Jupyter notebook 03)**

1. **Workflow:**

• The data transformation process involved cleaning and standardizing the dataset, creating unique identifiers for each respondent, and structuring the data to facilitate further analysis. These steps were crucial to ensure data consistency and usability in downstream processes.

1. **Aim of the Project/Analysis:**

• The goal of the data transformation phase was to refine the dataset into a structured format that supports detailed analysis. This transformation aimed to eliminate inconsistencies and prepare the data for integration into various analysis and visualization tools.

1. **Survey Period:**

• While the survey period specifics were not addressed in this notebook, the dataset was prepared to focus on key variables, enabling efficient time-based and demographic analysis in later stages.

1. **Aim of the Survey:**

• The transformed data aimed to streamline analysis by highlighting critical factors influencing mental health, such as gender, self-employment status, and country of residence. These factors are crucial for understanding mental health trends and conditions.

1. **Countries Involved:**

• The dataset contains information from numerous countries, prominently featuring the United States, which aligns with CDC’s focus on understanding mental health patterns within specific demographics.

1. **Respondents Involved:**

• With over 290,000 respondents, the transformed data ensures that each individual’s details are properly organized and uniquely identifiable, facilitating precise analysis of mental health trends.

1. **Main Factors Affecting Mental Health:**

• During this phase, significant emphasis was placed on factors like self-employment, gender, and geographic location. These variables are known to influence mental health and are consistent with factors identified by the CDC.

1. **Conclusion:**

• The data transformation successfully organized the dataset into a clear structure that supports robust analysis. By standardizing and cleaning the data, the project can now proceed with a focus on generating insights that align with CDC guidelines on mental health.

1. **Recommendations:**

• Recommendations from this phase involve leveraging the structured dataset to focus on specific at-risk groups identified by CDC research. Emphasis should be on exploring how factors like employment status and country of origin impact mental health outcomes.

1. **Further Investigation and Considerations:**

• Further work could involve validating this dataset against external sources such as CDC data to reinforce its reliability. Future analyses might also explore correlations between demographic factors and mental health conditions to align with public health objectives.