**Project Finalization Report**

**Title of Project: Gathering, Compiling and Examining Data**

**Duration of the project: [3\_12\_2023 to 23\_12\_2023]**

**Project Overview:**

To enable insightful analysis and sound decision-making, the project's objectives included gathering, cleaning, and analyzing several datasets. Phases of data collection, wrangling, and analysis separated the jobs.

**Task 1: Gathering Data**

Data set downloaded to Jupiter Environment after a cursory glance using the provided Google link:-

The first, second, and third datasets

**Task 2. Data Wrangling (Datasets 1, 2, and 3) is Task 2.**

1. **Merging Datasets:**

Imported both datasets by importing the pandas and numpy Python libraries, and then stored them in a variable for additional analysis (such as attributes, tables, etc.) for a thorough examination.

**2. Investigation by Category:**

Unique values for categorical variables were found in the successfully completed task of categorical exploration, which added to a thorough comprehension of the many categories present in the dataset.

**3. Simplifying the Data:**

Reduced the number of unnecessary columns in the dataset, which made it easier to analyse and more efficient. This optimisation guarantees a more targeted and efficient method of obtaining insightful information from the dataset while also increasing data processing speed and decreasing complexity.

4. **Overview of the Dataset:**

Examined the dataset's dimensions to ensure that you have a quick understanding of its size and organization. Understanding the breadth and depth of the available data allows for effective data management and analysis, which is made possible by this preliminary assessment's useful context.

**5. Consistency Check:**

To guarantee uniformity and consistency, dataset datatypes were verified. This examination improves data integrity by averting inconsistencies and enabling precise analysis. Consistent datatypes ensure data trustworthiness, support smart decision-making based on a standardized dataset, and provide a dependable basis for additional processing.

**6. Datatype Summary:**

To obtain a thorough knowledge, a summary of the dataset datatypes was generated. This synopsis helps identify the type of data, improving understanding and making further analysis easier. For effective data manipulation and interpretation in a variety of analytical procedures, the summary is a useful resource.

**7. Managing Missing Values:**

Appropriate techniques were used to address missing values in the dataset. Used methods like imputation or deletion to improve the dependability and completeness of the data. As a result, the information is represented more accurately, supporting thorough analysis and decision-making based on a well-rounded dataset.

**8. Data Validation:**

Verified entries for accuracy and ensured data accuracy at the primary level. By confirming the dataset's integrity, this basic check lowers the possibility of errors and provides a solid platform for further investigation. Early data correctness validation improves the information's overall quality and reliability.

**9. Central Tendency Exploration:**

Analyzed the mean, median, and mode of the combined dataset's central tendency measurements. Based on the typical characteristics of the dataset, this exploration helps discover trends and make educated judgements by offering a thorough grasp of the data's fundamental values.

**10. Improving Data Quality:**

Outliers and missing values were fixed to raise the general level of data quality. For this assignment, relevant methods like imputation or removal had to be used in order to improve accuracy and completeness. The integrity of the dataset was strengthened by resolving these problems, which helped to make analyses and decision-making procedures more trustworthy.

**Analysis Phase:**

**1. Data Distribution Understanding:**

Evaluated skewness in the combined dataset to comprehend its distribution characteristics. This analysis provides insights into the asymmetry of the data, crucial for understanding its shape and potential outliers. Examining skewness aids in selecting appropriate statistical methods and ensures a nuanced interpretation of the dataset, enhancing the accuracy of subsequent analyses and decisions based on the distribution properties.

**2. Variable Relationships Identification:**

Conducted a thorough examination of variable correlations to discern relationships within the dataset. This analysis unveils the strength and direction of associations between variables, guiding the understanding of dependencies. By exploring correlations, valuable insights into patterns and connections among different elements were gained, contributing to informed decision-making and a deeper understanding of the data's interdependencies.

**Conclusion:**

In summary, the methodical procedure of gathering, organising, and evaluating data guaranteed a methodical approach, offering a strong base for deriving significant insights and assisting with well-informed decision-making. The project was effectively finished, fulfilling the initial goals set forth. The dataset has been cleaned and analysed and is now prepared for usage and interpretation.