**Report on Project Finalization**

**Project Title:** Data Collection, Compilation, and Analysis

The project will run from [3-12-2023 until 23-12-2023].

**Project Synopsis:**

Gathering, cleaning, and analyzing several datasets were among the project's goals in order to facilitate intelligent analysis and wise decision-making. The jobs were divided into phases for data collection, wrangling, and analysis.

**First Task: Data Collection**

After a quick glance, download the data set to Jupiter Environment using the given Google link:The first, secondary, and tertiary datasets

**Second Task 2 is Data Wrangling (involving Sets 1, 2, and 3).**

1. **Merging Datasets**: Imported both datasets by importing the Python libraries pandas and numpy. After that, they were placed in a variable for further analysis (in the form of attributes, tables, etc.) to ensure a comprehensive look.

2. **Category-Based Investigation**: The task of categorical investigation was successfully completed, yielding unique values for categorical variables. This contributed to a comprehensive understanding of the numerous categories included in the dataset.

3. **Simplifying the Data**: The dataset was made more efficient and easy to analyze by removing superfluous columns. This optimization promises faster and less complex data processing along with a more focused and effective way to extract meaningful information from the dataset.

4. **Overview of the Dataset:** Looked at the dataset's dimensions to make sure you can quickly determine how big and how organized it is. Effective data management and analysis are made feasible by knowing the scope and depth of the available data, which is made possible by the helpful context of this preliminary assessment.

5. **Consistency Check:** The dataset information types were confirmed in order to ensure consistency and uniformity. By preventing inconsistencies and facilitating accurate analysis, this review enhances data integrity. Consistent datatypes guarantee the reliability of the data, facilitate informed decision-making using a standardized dataset, and offer a solid foundation for further processing.

6. **Datatype Summary:** An overview of the dataset datatypes was produced in order to have a comprehensive understanding. By identifying the type of data, this summary enhances comprehension and facilitates additional analysis. The summary is a helpful tool for efficient data management and interpretation in a range of analytical processes.

7. **Handling Missing Values**: The dataset's missing values were handled with the proper methods. imputation and deletion techniques were applied to increase the data's completeness and dependability. As a result, the data is more correctly represented, enabling in-depth examination and well-rounded dataset-based decision-making.

8. **Data Validation:** Assured main level data accuracy and checked entries for accuracy. Through validation of the dataset integrity, this preliminary check reduces the likelihood of errors and offers a strong foundation for additional research. The overall quality and reliability of the information are enhanced by early data correctness validation.

9. **Central Tendency Exploration:** Examined the central tendency measurements of the combined dataset by calculating its mean, median, and mode. This exploration provides a full understanding of the underlying values of the data, which helps uncover trends and make informed decisions based on the typical properties of the dataset.

10. **Enhancing Data Quality:** To improve the overall degree of data quality, outliers and missing values were corrected. To increase accuracy and completeness, pertinent techniques like imputation or removal required to be applied to this assignment. Fixing these issues improved the dataset's integrity, which increased the reliability of analysis and decision-making processes.

**Phase of Analysis:**

1. **Understanding Data Distribution:** To understand the distribution features of the merged dataset, skewness was evaluated. Understanding the asymmetry of the data is essential for comprehending its form and possible outliers, and this research sheds light on it. By ensuring a nuanced interpretation of the dataset and assisting in the selection of relevant statistical methods, skewness analysis improves the accuracy of subsequent analyses and judgments based on distribution features.

2. **Identification of Variable Relationships:** To identify relationships within the dataset, a comprehensive analysis of variable correlations was carried out. This analysis helps to clarify dependencies by revealing the direction and degree of relationships between variables. Investigating correlations provided insightful information about the relationships and patterns between various components, which aided in the formulation of well-informed decisions and deepened comprehension of the interdependencies within the data.

In conclusion, the systematic process of collecting, organizing, and assessing data ensured a systematic approach, providing a solid foundation for drawing important conclusions and supporting informed decision-making. The project was successfully completed, meeting the original objectives. The dataset is now ready for use and interpretation after being cleaned and analyzed.