# Report on Handling Missing Data in the Dataset

## Introduction

This report provides a summary of the steps taken to handle missing data in the provided dataset. Missing data can lead to inaccuracies and biases in analyses, and addressing this issue is essential to ensure data quality.

## Dataset Overview

The dataset contained information regarding transactions with the following key columns: 'Date', 'Category', 'Amount', and other related fields. The main focus of this analysis was on handling missing values in the 'Date', 'Category', and 'Amount' columns.

## Handling Missing Values

### 1. Date Column

The missing values in the 'Date' column were filled using forward-fill (propagation of the last valid value) and backward-fill (propagation of the next valid value). This ensured continuity and consistency in the timeline.

### 2. Category Column

The 'Category' column contained categorical data with missing entries. These missing values were addressed by filling them with the mode (most frequently occurring value) of the column. This method helped maintain the general distribution of categories in the dataset.

### 3. Amount Column

The 'Amount' column had numerical data with missing values. These were filled using the median value of the column, as the median is robust to outliers and provides a central value that is not overly influenced by extreme data points.

## Summary

In summary, the handling of missing data involved techniques tailored to the nature of each column. Forward-fill and backward-fill were applied to the 'Date' column to ensure timeline consistency. The mode was used to impute missing categorical values in the 'Category' column, and the median was used for the 'Amount' column to handle numerical missing values effectively.

## Conclusion

Addressing missing data is a crucial step in data preprocessing. By applying appropriate imputation techniques, the dataset's integrity and usability were significantly improved, ensuring more reliable and accurate analyses.