

AI Data Insights Report

Key Insights: 1. **Dataset Overview**: - The dataset contains 8,851 rows and 6 columns, with 4 numeric columns and 2 categorical columns. - There are no missing values in the dataset, indicating data completeness. 2. **Numeric Column Analysis**: - **Column `8.72`**: - Mean: 228.74, Median: 54.5, indicating a right-skewed distribution with a significant number of higher values. - The maximum value (22,638.48) is significantly higher than the 75th percentile (209.97), suggesting the presence of extreme outliers. - **Column `4`**: - Mean: 3.79, Median: 3.0, suggesting a relatively balanced distribution. - The range is small (1 to 14), with most values clustered between 2 and 5. - **Column `0`**: - Mean: 0.155, Median: 0.2, indicating a left-skewed distribution with most values near 0. - The maximum value is 0.8, with the majority of values concentrated at or below 0.2. - **Column `2.8776`**: - Mean: 29.45, Median: 8.64, indicating a right-skewed distribution. - The minimum value is negative (-6,599.978), which could indicate errors or specific business scenarios (e.g., losses or refunds). - The maximum value (8,399.976) is significantly higher than the 75th percentile (29.68), suggesting the presence of extreme outliers. 3. **Outliers**: - A total of 442 outliers were detected across the dataset, primarily in columns `8.72` and `2.8776`. - These outliers could significantly impact statistical analyses and should be investigated further. --- ### Trends: 1. **Skewed Distributions**: - Columns `8.72` and `2.8776` exhibit highly skewed distributions, with a small number of very high values driving up the mean. - This could indicate a small subset of transactions or records contributing disproportionately to the overall metrics. 2. **Concentration of Values**: - Columns `4` and `0` show relatively narrow distributions, with most values concentrated around the median. This suggests consistency in these variables. 3. **Negative Values in `2.8776`**: - The presence of negative values in `2.8776` could indicate refunds, losses, or errors in the data. This trend warrants further investigation. --- ### Risks: 1. **Impact of Outliers**: - The presence of 442 outliers could distort averages, trends, and other statistical analyses. If not addressed, these could lead to incorrect business decisions. - Outliers in `8.72` and `2.8776` may represent exceptional cases (e.g., high-value transactions or errors) that require further scrutiny. 2. **Negative Values in `2.8776`**: - Negative values may indicate operational issues such as refunds, losses, or data entry errors. If these are errors, they could compromise the integrity of the dataset. 3. **Skewed Data**: - The skewed distributions in `8.72` and `2.8776` may make it challenging to derive meaningful insights without appropriate transformations or adjustments. --- ### Business Recommendations: 1. **Investigate Outliers**: - Perform a detailed analysis of the 442 outliers to determine their nature (e.g., errors, exceptional cases, or legitimate high-value transactions). - If outliers are errors, consider removing or correcting them. If legitimate, segment them for separate analysis. 2. **Address Negative Values**: - Investigate the negative values in `2.8776` to understand their origin. If they represent refunds or losses, ensure they are correctly categorized and accounted for in analyses. 3. **Data Transformation**: - Apply transformations (e.g., log transformation) to skewed columns (`8.72` and `2.8776`) to normalize their distributions and improve the accuracy of statistical analyses. 4. **Segmentation**: - Segment the data based on categorical columns (`rs` and `Stockwell Push Pins`) to identify patterns or trends specific to different groups. 5. **Monitor High-Value Transactions**: - For columns like `8.72`, monitor high-value transactions to identify potential opportunities (e.g., high-value customers) or risks (e.g., fraud). 6. **Enhance Data Quality**: - Implement data validation checks to prevent negative or erroneous values in future data collection processes. - Regularly audit the dataset to ensure data integrity and accuracy. 7. **Further Analysis**: - Conduct a deeper analysis of categorical variables (`rs` and `Stockwell Push Pins`) to uncover potential relationships with numeric variables. - Explore time-based trends (if applicable) to identify seasonality or patterns over time. By addressing these insights, trends, and risks, the business can improve decision-making, enhance operational efficiency, and mitigate potential risks.





