Validation Report for rds - customer_churn_dataset_rds_20250314234741.csv

Issues:

missing_data - customerID: 0 missing (0.00%)
missing_data - gender: 1000 missing (10.00%)
missing_data - SeniorCitizen: 1000 missing (10.00%)
missing_data - Partner: 1000 missing (10.00%)
missing_data - Dependents: 1000 missing (10.00%)
missing_data - tenure: 1000 missing (10.00%)
missing_data - PhoneService: 1000 missing (10.00%)
missing_data - MultipleLines: 1000 missing (10.00%)
missing_data - InternetService: 1000 missing (10.00%)
missing_data - OnlineSecurity: 1000 missing (10.00%)
missing_data - OnlineBackup: 1000 missing (10.00%)
missing_data - Churn: 0 missing (0.00%)
Churn: Binary column; outlier check not applicable.
tenure: 0 out of 10000 rows (0.00%) are outliers.
PhoneService: Unique values: ['Yes', nan, 'No']
OnlineSecurity: Unique values: ['No', 'Yes', 'No internet service', nan]
gender: Unique values: ['Male', 'Female', nan]
Dependents: Unique values: ['No', 'Yes', nan]
OnlineBackup: Unique values: ['No internet service', 'Yes', nan, 'No']
Partner: Unique values: ['Yes', 'No', nan]
InternetService: Unique values: [nan, 'Fiber optic', 'DSL', 'No']
MultipleLines: Unique values: ['No', 'Yes', 'No phone service', nan]
duplicates: 0 duplicate rows found out of 10000 rows (0.00%).
outliers - Churn: Binary column; outlier check not applicable.
outliers - tenure: 0 out of 10000 rows (0.00%) are outliers.

Resolutions:

missing_data - gender: Consider imputing missing values (mean/median/mode) or dropping rows if appropriate.
missing_data - SeniorCitizen: Consider imputing missing values (mean/median/mode) or dropping rows if appropriate.
missing_data - Partner: Consider imputing missing values (mean/median/mode) or dropping rows if appropriate.
missing_data - Dependents: Consider imputing missing values (mean/median/mode) or dropping rows if appropriate.
missing_data - tenure: Consider imputing missing values (mean/median/mode) or dropping rows if appropriate.
missing_data - PhoneService: Consider imputing missing values (mean/median/mode) or dropping rows if appropriate.
missing_data - MultipleLines: Consider imputing missing values (mean/median/mode) or dropping rows if appropriate.
missing_data - InternetService: Consider imputing missing values (mean/median/mode) or dropping rows if appropriate.
missing_data - OnlineSecurity: Consider imputing missing values (mean/median/mode) or dropping rows if appropriate.
missing_data - OnlineBackup: Consider imputing missing values (mean/median/mode) or dropping rows if appropriate.
missing_data - customerID: No action needed.
missing_data - Churn: No action needed.
Churn: No specific resolution suggested.
tenure: Review and correct data entries or adjust the expected range if appropriate.
PhoneService: No specific resolution suggested.
OnlineSecurity: No specific resolution suggested.

gender: No specific resolution suggested.

Dependents: No specific resolution suggested.

OnlineBackup: No specific resolution suggested.

Partner: No specific resolution suggested.

InternetService: No specific resolution suggested.

MultipleLines: No specific resolution suggested.

duplicates: Consider removing duplicate rows based on 'customerID'.

outliers - Churn: Review outlier values; consider transformation or removal if they are errors.

outliers - tenure: Review outlier values; consider transformation or removal if they are errors.