

Report

- The dataset consists of 148,654 rows and 13 columns.
- Column names: 'Id', 'EmployeeName', 'JobTitle', 'BasePay', 'OvertimePay', 'OtherPay', 'Benefits', 'TotalPay', 'TotalPayBenefits', 'Year', 'Notes', 'Agency', 'Status'.
- There are 4 duplicated rows (with indices 148646, 148650, 148651, 148652), all containing no data.
- The average salary is \$74,768.32, with a standard deviation of \$50,517.01. A higher standard deviation indicates more spread-out data from the mean, while a lower standard deviation indicates data more concentrated around the mean.
- The salary range is from -\$618.13 to \$567,595.43.
- Columns with missing data: 'BasePay' (609 missing), 'OvertimePay' (4 missing), 'OtherPay' (4 missing), 'Benefits' (36,163 missing), 'Notes' (all 148,654 missing), 'Status' (all 148,654 missing).
- Columns with no data: 'Notes', 'Status'.
- The 'Agency' column has the same data, which is 'San Francisco'.
- There is a strong positive correlation (0.9545) between 'TotalPay' and 'BasePay', indicating that as BasePay increases, TotalPay also tends to increase.
- There are different departments such like: 'Police Department', 'Fire Department', 'Svcs Dept', and 'Trnsp Dept'.