

Task Report

- The dataset consists of 13 columns and 148,654 rows.
- Columns and Data Types:

Column Name	Data Type	
Id	Integer	
EmployeeName	Object	
JobTitle	Object	
BasePay	Float64	
OvertimePay	Float64	
OtherPay	Float64	
Benefits	Float64	
TotalPay	Float64	
TotalPayBenefits	Float64	
Year	Integer	
Notes	Float64	
Agency	Object	
Status	Float64	

- Missing Values:

Column Name	Missing Values
Notes	148,654
Status	148,654
Benefits	36,163
BasePay	609
OvertimePay	4
OtherPay	4

- Basic Statistical Analysis:

Statistic	Value
Mode of TotalPay	\$0.0
Mean TotalPay	\$74,768.32
Median TotalPay	\$71,426.61
Minimum TotalPay	-\$618.13
Maximum TotalPay	\$567,595.43
Range of Salaries	[\$-618.13, \$567,595.43]
Standard Deviation of TotalPay	\$50,517.01

- Missing data filled with mean due to its common occurrence in the dataset.
- Histogram Analysis: Majority of salaries fall within the range of 0 to \$100,000.
- Grouped Analysis by Year: Statistical analysis grouped by year shows minimal difference in mean and standard deviation of salaries across different years.
 - Low correlation observed between year and salary (0.032).
 - Strong correlation observed between salary and BasePay (0.95).