City_of_Chicago_Payroll_Data PROBLEM Solutions

Concept	Description	Example
NaN (Not a Number)	Represents missing or undefined data in a DataFrame.	pay['Typical Hours'] contains NaN values for some rows.
DataFrame Info	Provides an overview of the DataFrame, including column names, non-null counts, and data types.	pay.info() shows the column names and data types for the DataFrame, as well as how many non-null values there are.
DataFrame isnull()	Checks for missing values (NaNs) in the DataFrame. Returns a boolean mask.	pay.isnull().sum() shows the number of missing values in each column of the DataFrame.
describe() Method	Provides summary statistics for numerical columns, including count, mean, min, max, etc.	pay.describe() shows summary statistics of the Typical Hours column.
describe(include='all')	Provides summary statistics for both numerical and categorical columns.	pay.describe(include='all') shows summary statistics for all columns, including categorical data.
max() Function	Returns the maximum value in a column.	pay['Typical Hours'].max() returns the maximum value of Typical Hours (40.0).
min() Function	Returns the minimum value in a column.	pay['Typical Hours'].min() returns the minimum value of Typical Hours (10.0).
mean() Function	Returns the mean (average) value of a column.	pay['Typical Hours'].mean() returns the mean of Typical Hours (34.7).
GroupBy	Groups the DataFrame by a specific column and performs aggregate operations on other columns.	pay.groupby('Salary or Hourly').count() shows the count of each category in the 'Salary or Hourly' column.
value_counts()	Returns a count of unique values in a column, showing their frequencies.	pay['Salary or Hourly'].value_counts() shows the count of employees on Salary vs Hourly.
idxmax()	Returns the index of the first occurrence of the maximum value in a column.	pay['Salary'].idxmax() returns the index of the employee with the highest salary.

Concept	Description	Example
idxmin()	Returns the index of the first occurrence of the minimum value in a column.	pay['Salary'].idxmin() returns the index of the employee with the lowest salary.
iloc[]	Used for integer-location based indexing for selection by position.	pay.iloc[15387] returns the row at position 15387 in the DataFrame.
str.replace()	Replaces specified characters or substrings in a string column.	pay['Salary'] = pay['Annual Salary'].str.replace('$', '') removes the dollar sign from the 'Annual Salary'.