

Data Engineering Test_Shell_ENERGY_Output

Input Data Inspection/Validation(For details please see code in Jupyter notebook.ipynb file)

- Inspection: Get csv file name in ConsumptionData folder
- Method: `os.listdir()` to get the file list; loop through the `file_list` and appendix the file names to `name_list`
- Inspection: inspect the data in csv files
- Method: Pandas: `read_csv` and `concat` to combine ConsumptionData csv files
- Inspection: see how many unique values in each column
- Method: `df.nunique()` to check how many unique values in each column
- Inspection: Check the year range
- Method: use dataframe `.min()` `.max()` to get the year range
- Inspection: after merge dataframe together, check dataframe information
- Method: `dataframe.info()` to see null values, datatype and column name...
- Inspection: check the row qty, and other aggregation values of df
- Method: `df.describe()`
- Inspection: check "Quantity" data distribution
- Method: seaborn and matplotlib to generate a plot
- Inspection: Also tried to use python great_expectations to inspect data
- Method: But don't have time to configure the data