Data Pre-Processing

Data preprocessing involves cleaning and transforming the data to make it suitable for modeling. This includes:

- Converting dates into useful features like year, month, or days worked
- Encoding categorical variables into numerical form
- Scaling numerical data if required

Checking for Null Values

Before model building, null or missing values are identified and handled. Techniques include:

- Filling missing values with mean/median/mode
- Forward/backward fill
- Dropping rows/columns if missing data is excessive

Data Pre-Processing

Checking for null Values

Double-click (or enter) to edit

```
[ ]  dataframe.shape
```

```
(1197, 15)
```

```
dataframe.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1197 entries, 0 to 1196
Data columns (total 15 columns):
 #   Column                   Non-Null Count  Dtype  
---  ------                   --------------  -----  
 0   date                     1197 non-null   object 
 1   quarter                  1197 non-null   object 
 2   department               1197 non-null   object 
 3   day                      1197 non-null   object 
 4   team                     1197 non-null   int64  
 5   targeted_productivity    1197 non-null   float64
 6   smv                      1197 non-null   float64
 7   wip                      691 non-null    float64
 8   over_time                1197 non-null   int64  
 9   incentive                1197 non-null   int64  
 10  idle_time                1197 non-null   float64
 11  idle_men                 1197 non-null   int64  
 12  no_of_style_change       1197 non-null   int64  
 13  no_of_workers            1197 non-null   float64
 14  actual_productivity      1197 non-null   float64
dtypes: float64(6), int64(5), object(4)
memory usage: 140.4+ KB
```