Visualizing and Analysis

Visualization is used to better understand data distribution, patterns, and relationships between features. Libraries like Matplotlib and Seaborn can help plot histograms, boxplots, heatmaps, and scatter plots. These visual insights guide preprocessing and model selection.

Data Collection

The dataset is collected from HR records, employee logs, or company databases. It typically includes information like:

- **quarter**
- **department**
- **day**
- **team**
- **targeted_productivity**
- **smv**
- **over_time**
- **incentive**
- **idle_time**
- **idle_men**
- **no_of_style_change**
- **no_of_workers**
- **month**

Data Collection

Double-click (or enter) to edit

```
dataframe = pd.read_csv('/content/garments_worker_productivity.csv')
dataframe
```

	date	quarter	department	day	team	targeted_productivity	smv	wip	over_time	incentive	idle_time	idle_men	no_of_style_change	no_of_workers	actual_productivity
0	1/1/2015	Quarter1	sweing	Thursday	8	0.80	26.16	1108.0	7080	98	0.0	0	0	59.0	0.940725
1	1/1/2015	Quarter1	finishing	Thursday	1	0.75	3.94	NaN	960	0	0.0	0	0	8.0	0.886500
2	1/1/2015	Quarter1	sweing	Thursday	11	0.80	11.41	968.0	3660	50	0.0	0	0	30.5	0.800570
3	1/1/2015	Quarter1	sweing	Thursday	12	0.80	11.41	968.0	3660	50	0.0	0	0	30.5	0.800570
4	1/1/2015	Quarter1	sweing	Thursday	6	0.80	25.90	1170.0	1920	50	0.0	0	0	56.0	0.800382
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1192	3/11/2015	Quarter2	finishing	Wednesday	10	0.75	2.90	NaN	960	0	0.0	0	0	8.0	0.628333
1193	3/11/2015	Quarter2	finishing	Wednesday	8	0.70	3.90	NaN	960	0	0.0	0	0	8.0	0.625625
1194	3/11/2015	Quarter2	finishing	Wednesday	7	0.65	3.90	NaN	960	0	0.0	0	0	8.0	0.625625
1195	3/11/2015	Quarter2	finishing	Wednesday	9	0.75	2.90	NaN	1800	0	0.0	0	0	15.0	0.505889
1196	3/11/2015	Quarter2	finishing	Wednesday	6	0.70	2.90	NaN	720	0	0.0	0	0	6.0	0.394722