

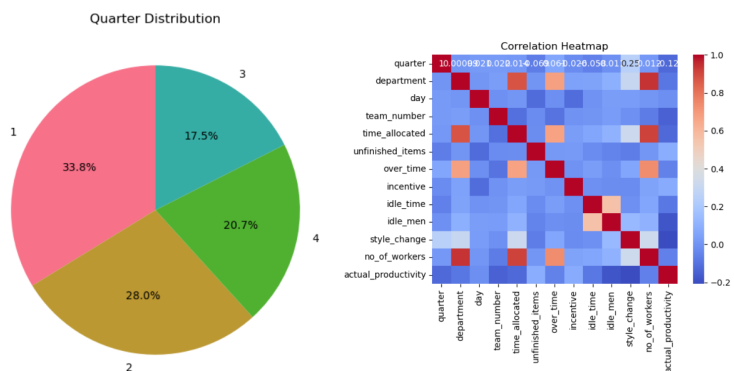
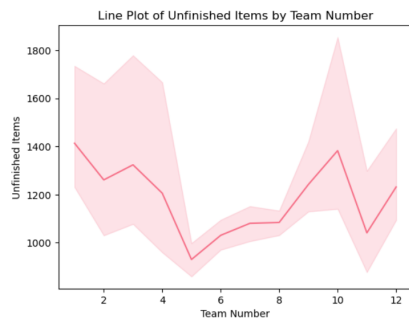
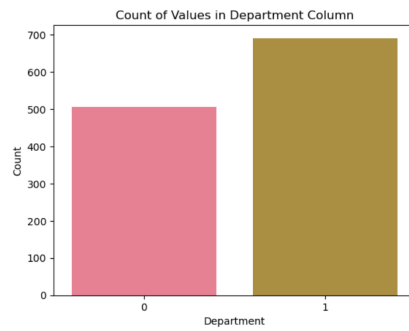
Data Collection and Preprocessing Phase

Date	15 March 2024
Team ID	SWTID1720673861
Project Title	Garment Worker efficiency Calculator
Maximum Marks	6 Marks

Data Exploration and Preprocessing Report

Dataset variables will be statistically analyzed to identify patterns and outliers, with Python employed for preprocessing tasks like normalization and feature engineering. Data cleaning will address missing values and outliers, ensuring quality for subsequent analysis and modeling, and forming a strong foundation for insights and predictions.

Section	Description																																																																																																																																								
Data Overview	<div><div><div>Dimension:</div><div>1197 rows × 15 columns</div></div><div><div></div><div>Descriptive statistics:</div></div></div>																																																																																																																																								
	<table><tr><th></th><th>A</th><th>B</th><th>C</th><th>D</th><th>E</th><th>F</th><th>G</th><th>H</th><th>I</th><th>J</th><th>K</th><th>L</th><th>M</th><th>N</th><th>O</th><th>P</th></tr><tr><td>1</td><td>date</td><td>quarter</td><td>departmer</td><td>day</td><td>team</td><td>targeted_p</td><td>smv</td><td>wip</td><td>over_time</td><td>incentive</td><td>idle_time</td><td>idle_men</td><td>no_of_style</td><td>no_of_worl</td><td>actual</td><td>productivity</td></tr><tr><td>2</td><td>1/1/2015</td><td>Quarter1</td><td>sweing</td><td>Thursday</td><td></td><td>8</td><td>0.8</td><td>26.16</td><td>1108</td><td>7080</td><td>98</td><td>0</td><td>0</td><td>0</td><td>59</td><td>0.940725</td></tr><tr><td>3</td><td>1/1/2015</td><td>Quarter1</td><td>finishing</td><td>Thursday</td><td></td><td>1</td><td>0.75</td><td>3.94</td><td></td><td>960</td><td>0</td><td>0</td><td>0</td><td>0</td><td>8</td><td>0.8865</td></tr><tr><td>4</td><td>1/1/2015</td><td>Quarter1</td><td>sweing</td><td>Thursday</td><td></td><td>11</td><td>0.8</td><td>11.41</td><td>968</td><td>3660</td><td>50</td><td>0</td><td>0</td><td>0</td><td>30.5</td><td>0.80057</td></tr><tr><td>5</td><td>1/1/2015</td><td>Quarter1</td><td>sweing</td><td>Thursday</td><td></td><td>12</td><td>0.8</td><td>11.41</td><td>968</td><td>3660</td><td>50</td><td>0</td><td>0</td><td>0</td><td>30.5</td><td>0.80057</td></tr><tr><td>6</td><td>1/1/2015</td><td>Quarter1</td><td>sweing</td><td>Thursday</td><td></td><td>6</td><td>0.8</td><td>25.9</td><td>1170</td><td>1920</td><td>50</td><td>0</td><td>0</td><td>0</td><td>56</td><td>0.800382</td></tr><tr><td>7</td><td>1/1/2015</td><td>Quarter1</td><td>sweing</td><td>Thursday</td><td></td><td>7</td><td>0.8</td><td>25.9</td><td>984</td><td>6720</td><td>38</td><td>0</td><td>0</td><td>0</td><td>56</td><td>0.800125</td></tr></table>		A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	1	date	quarter	departmer	day	team	targeted_p	smv	wip	over_time	incentive	idle_time	idle_men	no_of_style	no_of_worl	actual	productivity	2	1/1/2015	Quarter1	sweing	Thursday		8	0.8	26.16	1108	7080	98	0	0	0	59	0.940725	3	1/1/2015	Quarter1	finishing	Thursday		1	0.75	3.94		960	0	0	0	0	8	0.8865	4	1/1/2015	Quarter1	sweing	Thursday		11	0.8	11.41	968	3660	50	0	0	0	30.5	0.80057	5	1/1/2015	Quarter1	sweing	Thursday		12	0.8	11.41	968	3660	50	0	0	0	30.5	0.80057	6	1/1/2015	Quarter1	sweing	Thursday		6	0.8	25.9	1170	1920	50	0	0	0	56	0.800382	7	1/1/2015	Quarter1	sweing	Thursday		7	0.8	25.9	984	6720	38	0	0	0	56	0.800125
		A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P																																																																																																																								
	1	date	quarter	departmer	day	team	targeted_p	smv	wip	over_time	incentive	idle_time	idle_men	no_of_style	no_of_worl	actual	productivity																																																																																																																								
2	1/1/2015	Quarter1	sweing	Thursday		8	0.8	26.16	1108	7080	98	0	0	0	59	0.940725																																																																																																																									
3	1/1/2015	Quarter1	finishing	Thursday		1	0.75	3.94		960	0	0	0	0	8	0.8865																																																																																																																									
4	1/1/2015	Quarter1	sweing	Thursday		11	0.8	11.41	968	3660	50	0	0	0	30.5	0.80057																																																																																																																									
5	1/1/2015	Quarter1	sweing	Thursday		12	0.8	11.41	968	3660	50	0	0	0	30.5	0.80057																																																																																																																									
6	1/1/2015	Quarter1	sweing	Thursday		6	0.8	25.9	1170	1920	50	0	0	0	56	0.800382																																																																																																																									
7	1/1/2015	Quarter1	sweing	Thursday		7	0.8	25.9	984	6720	38	0	0	0	56	0.800125																																																																																																																									
Univariate Analysis																																																																																																																																									



Outliers and Anomalies	-																																																																																				
Data Preprocessing Code Screenshots																																																																																					
Loading Data	<pre>In [31]: # Load the dataset file_path = r"C:\Users\grshr\OneDrive\Desktop\AIwithTensorflow Project\garments_worker_productivity.csv" df = pd.read_csv(file_path) In [32]: df.head()</pre> <table><thead><tr><th></th><th>date</th><th>quarter</th><th>department</th><th>day</th><th>team</th><th>targeted_productivity</th><th>smv</th><th>wip</th><th>over_time</th><th>incentive</th><th>idle_time</th><th>idle_men</th><th>no_o</th></tr></thead><tbody><tr><td>0</td><td>1/1/2015</td><td>Quarter1</td><td>sewing</td><td>Thursday</td><td>8</td><td>0.80</td><td>26.16</td><td>1108.0</td><td>7080</td><td>98</td><td>0.0</td><td>0</td><td></td></tr><tr><td>1</td><td>1/1/2015</td><td>Quarter1</td><td>finishing</td><td>Thursday</td><td>1</td><td>0.75</td><td>3.94</td><td>NaN</td><td>960</td><td>0</td><td>0.0</td><td>0</td><td></td></tr><tr><td>2</td><td>1/1/2015</td><td>Quarter1</td><td>sewing</td><td>Thursday</td><td>11</td><td>0.80</td><td>11.41</td><td>968.0</td><td>3660</td><td>50</td><td>0.0</td><td>0</td><td></td></tr><tr><td>3</td><td>1/1/2015</td><td>Quarter1</td><td>sewing</td><td>Thursday</td><td>12</td><td>0.80</td><td>11.41</td><td>968.0</td><td>3660</td><td>50</td><td>0.0</td><td>0</td><td></td></tr><tr><td>4</td><td>1/1/2015</td><td>Quarter1</td><td>sewing</td><td>Thursday</td><td>6</td><td>0.80</td><td>25.90</td><td>1170.0</td><td>1920</td><td>50</td><td>0.0</td><td>0</td><td></td></tr></tbody></table>		date	quarter	department	day	team	targeted_productivity	smv	wip	over_time	incentive	idle_time	idle_men	no_o	0	1/1/2015	Quarter1	sewing	Thursday	8	0.80	26.16	1108.0	7080	98	0.0	0		1	1/1/2015	Quarter1	finishing	Thursday	1	0.75	3.94	NaN	960	0	0.0	0		2	1/1/2015	Quarter1	sewing	Thursday	11	0.80	11.41	968.0	3660	50	0.0	0		3	1/1/2015	Quarter1	sewing	Thursday	12	0.80	11.41	968.0	3660	50	0.0	0		4	1/1/2015	Quarter1	sewing	Thursday	6	0.80	25.90	1170.0	1920	50	0.0	0	
	date	quarter	department	day	team	targeted_productivity	smv	wip	over_time	incentive	idle_time	idle_men	no_o																																																																								
0	1/1/2015	Quarter1	sewing	Thursday	8	0.80	26.16	1108.0	7080	98	0.0	0																																																																									
1	1/1/2015	Quarter1	finishing	Thursday	1	0.75	3.94	NaN	960	0	0.0	0																																																																									
2	1/1/2015	Quarter1	sewing	Thursday	11	0.80	11.41	968.0	3660	50	0.0	0																																																																									
3	1/1/2015	Quarter1	sewing	Thursday	12	0.80	11.41	968.0	3660	50	0.0	0																																																																									
4	1/1/2015	Quarter1	sewing	Thursday	6	0.80	25.90	1170.0	1920	50	0.0	0																																																																									
Handling Missing Data	<pre>In [84]: df['team_number'] = df['team_number'].astype(int) df['over_time'] = df['over_time'].astype(int) df['incentive'] = df['incentive'].astype(int) df['idle_time'] = df['idle_time'].astype(int) df['idle_men'] = df['idle_men'].astype(int) df['style_change'] = df['style_change'].astype(int) df['time_allocated'] = df['time_allocated'].astype(int) df['unfinished_items'] = df['unfinished_items'].astype(int) df['idle_time'] = df['idle_time'].astype(int) df['no_of_workers'] = df['no_of_workers'].astype(int) x_test['quarter'] = x_test['quarter'].astype('category') In [43]: lc = LabelEncoder()</pre>																																																																																				
Data Transformation	<pre>In [44]: print('Before encoding:', df['department'].unique()) df['department'] = lc.fit_transform(df['department']) print('After encoding:', df['department'].unique()) Before encoding: ['sewing' 'finishing'] After encoding: [1 0] In [45]: print('Before encoding:', df['day'].unique()) df['day'] = lc.fit_transform(df['day']) print('After encoding:', df['day'].unique()) Before encoding: ['Thursday' 'Saturday' 'Sunday' 'Monday' 'Tuesday' 'Wednesday'] After encoding: [3 1 2 0 4 5] In [54]: sc = StandardScaler() X_scaled = sc.fit_transform(x) In [55]: col = x.columns In [56]: X_scaled = pd.DataFrame(X_scaled, columns = col) X_scaled</pre>																																																																																				
Feature Engineering	Attached the codes in final submission.																																																																																				
Save Processed Data	-																																																																																				