นายศิรวิทย์ เตชประสพชัย 63070501220
นางสาวฉันท์สินี เมืองหนู 63070501221

Homework1: Decriptive Statistics

1) Select a manufacturing dataset (from Kaggle, etc.) and briefly descrbie about the dataset (e.g., the number of features, feature description)

- Reference: https://www.kaggle.com/datasets/ishadss/productivity-prediction-of-garment-employees

```
[101] import pandas as pd
      import numpy as np
      import matplotlib as plt
      %matplotlib inline
      import seaborn as sns
      file = pd.read_csv('/content/garments_worker_productivity.csv', parse_dates=['date'])

[2]   file.head()
```

	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	2015-01-01	Quarter1	sweing	Thursday	8	0.80	26.16	1108.0	7080	98	0.0	0	0	59.0	0.940725
1	2015-01-01	Quarter1	finishing	Thursday	1	0.75	3.94	NaN	960	0	0.0	0	0	8.0	0.886500
2	2015-01-01	Quarter1	sweing	Thursday	11	0.80	11.41	968.0	3660	50	0.0	0	0	30.5	0.800570
3	2015-01-01	Quarter1	sweing	Thursday	12	0.80	11.41	968.0	3660	50	0.0	0	0	30.5	0.800570
4	2015-01-01	Quarter1	sweing	Thursday	6	0.80	25.90	1170.0	1920	50	0.0	0	0	56.0	0.800382

2) Calculate mean, sd, variance, covariance, correlation, max, min

- Calculate mean, sd, variance, max, and min

```
[37] def summary_statistics(df):
       out = {}
       out['mean'] = df.mean()
       out['sd'] = df.std()
       out['variance'] = df.var()
       out['min'] = df.min()
       out['max'] = df.max()
       return pd.DataFrame(out)

[44] summary_statistics(file[['targeted_productivity','smv','wip', 'over_time','incentive', 'idle_time', 'idle_men', 'no_of_style_change', 'no_of_workers', 'actual_productivity']]).T
```

	targeted_productivity	smv	wip	over_time	incentive	idle_time	idle_men	no_of_style_change	no_of_workers	actual_productivity
mean	0.729632	15.062172	6.872281e+02	4.567460e+03	38.210526	0.730159	0.369256	0.150376	34.609858	0.735091
sd	0.097891	10.943219	1.514582e+03	3.348824e+03	160.182643	12.709757	3.268987	0.427848	22.197687	0.174488
variance	0.009583	119.754046	2.293960e+06	1.121462e+07	25658.479053	161.537911	10.686278	0.183054	492.737294	0.030446
min	0.070000	2.900000	0.000000e+00	0.000000e+00	0.000000	0.000000	0.000000	0.000000	2.000000	0.233705
max	0.800000	54.560000	2.312200e+04	2.592000e+04	3600.000000	300.000000	45.000000	2.000000	89.000000	1.120437

- Calculate covariance

```
[42] file.cov()
```

	targeted_productivity	smv	wip	over_time	incentive	idle_time	idle_men	no_of_style_change	no_of_workers	actual_productivity
targeted_productivity	0.009583	-0.074439	2.822141e+00	-2.903062e+01	0.513815	-0.069899	-0.017222	-0.008766	-0.183154	0.007201
smv	-0.074439	119.754046	5.348625e+03	2.473254e+04	57.195571	7.908797	3.788411	1.476655	221.580535	-0.233124
wip	2.822141	5348.624972	2.293960e+06	1.402572e+06	9206.092413	-98.190914	-35.247330	34.534237	12570.880860	12.523736
over_time	-29.030617	24732.539468	1.402572e+06	1.121462e+07	-2571.212375	1321.044885	-196.101555	85.666507	54574.928067	-31.674054
incentive	0.513815	57.195571	9.206092e+03	-2.571212e+03	25658.479053	-24.478679	-11.069442	-1.823491	175.018659	2.139222
idle_time	-0.069899	7.908797	-9.819091e+01	1.321045e+03	-24.478679	161.537911	23.231413	-0.063067	16.377286	-0.179303
idle_men	-0.017222	3.788411	-3.524733e+01	-1.961016e+02	-11.069442	23.231413	10.686278	0.186901	7.760404	-0.103661
no_of_style_change	-0.008766	1.476655	3.453424e+01	8.566651e+01	-1.823491	-0.063067	0.186901	0.183054	3.113065	-0.015481
no_of_workers	-0.183154	221.580535	1.257088e+04	5.457493e+04	175.018659	16.377286	7.760404	3.113065	492.737294	-0.224611
actual_productivity	0.007201	-0.233124	1.252374e+01	-3.167405e+01	2.139222	-0.179303	-0.103661	-0.015481	-0.224611	0.030446

นายศิรวิทย์ เตชประสพชัย 63070501220
นางสาวฉันท์สินี เมืองหนู 63070501221

➢ Calculate correlation

```
[43] file.corr()
```

	targeted_productivity	smv	wip	over_time	incentive	idle_time	idle_men	no_of_style_change	no_of_workers	actual_productivity
targeted_productivity	1.000000	-0.069489	0.019035	-0.088557	0.032768	-0.056181	-0.053818	-0.209294	-0.084288	0.421594
smv	-0.069489	1.000000	0.322704	0.674887	0.032629	0.056863	0.105901	0.315388	0.912176	-0.122089
wip	0.019035	0.322704	1.000000	0.276529	0.037946	-0.005101	-0.007119	0.053293	0.373908	0.047389
over_time	-0.088557	0.674887	0.276529	1.000000	-0.004793	0.031038	-0.017913	0.059790	0.734164	-0.054206
incentive	0.032768	0.032629	0.037946	-0.004793	1.000000	-0.012024	-0.021140	-0.026607	0.049222	0.076538
idle_time	-0.056181	0.056863	-0.005101	0.031038	-0.012024	1.000000	0.559146	-0.011598	0.058049	-0.080851
idle_men	-0.053818	0.105901	-0.007119	-0.017913	-0.021140	0.559146	1.000000	0.133632	0.106946	-0.181734
no_of_style_change	-0.209294	0.315388	0.053293	0.059790	-0.026607	-0.011598	0.133632	1.000000	0.327787	-0.207366
no_of_workers	-0.084288	0.912176	0.373908	0.734164	0.049222	0.058049	0.106946	0.327787	1.000000	-0.057991
actual_productivity	0.421594	-0.122089	0.047389	-0.054206	0.076538	-0.080851	-0.181734	-0.207366	-0.057991	1.000000

3) Visualize and interpret the dataset using box plot, scatter plot, time-series plot (if possible), heatmap, histogram

➢ Box plot

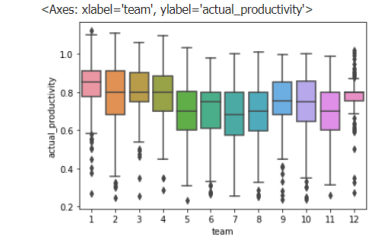
1. เปรียบเทียบ productivity ของแต่ละทีม

จากการพล็อต Boxplot เพื่อวิเคราะห์ความสัมพันธ์ระหว่าง team ทั้ง 12 ทีม และ targeted_productivity หรือประสิทธิภาพเป้าหมายจะพบว่าประสิทธิภาพเป้าหมายจะอยู่ในช่วง 0.7 – 0.8 ซึ่งพบว่า targeted_productivity จะอยู่ในเกณฑ์ที่ค่อนข้างสูงแต่จะพบว่าบางช่วงจะมี targeted_productivity ในช่วง 0.4 – 0.6 อาจเป็นเพราะในวันนั้นมีจำนวนของผลิตภัณฑ์ที่ต้องผลิตน้อย จึงมี targeted_productivity ที่ค่อนข้างต่ำ

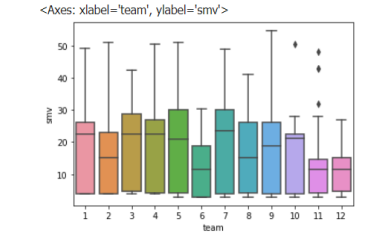
จากการพล็อต Boxplot เพื่อวิเคราะห์ความสัมพันธ์ระหว่าง team กับ smv หรือช่วงเวลาในการผลิตผลิตภัณฑ์ 1 ผลิตภัณฑ์ในหน่วยนาที ซึ่งจะพบว่า โดยส่วนใหญ่จะมี smv ของแต่ละทีมอยู่ที่ต่ำกว่า 10 นาที แต่จะมีช่วงเวลาที่เกินกว่า 10 นาที และอยู่ในช่วง 10 – 30 นาทีอยู่บ้าง ซึ่งอาจเกิดจากการที่ผลิตภัณฑ์ในแต่ละการผลิตมีความซับซ้อนในการผลิตที่ไม่เท่ากัน ดังนั้นจึงเกิดกราฟในลักษณะเช่นนี้

จากการพล็อต Boxplot เพื่อวิเคราะห์ความสัมพันธ์ระหว่าง team กับ wip หรืองานที่ยังเหลืออยู่ในแต่ละวันที่ทำงาน จะพบว่าโดยส่วนใหญ่แล้วจะไม่มี wip หรือ work in progress หลงเหลืออยู่ในแต่ละวัน และหากมีจะมีในปริมาณที่ค่อนข้างน้อย

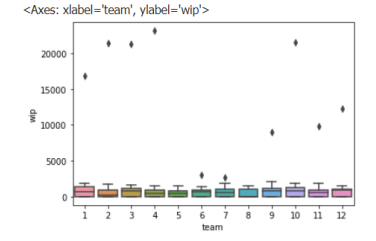
```
[ ] sns.boxplot(x="team",y="actual_productivity",data=file)
```




```
[ ] sns.boxplot(x="team",y="smv",data=file)
```



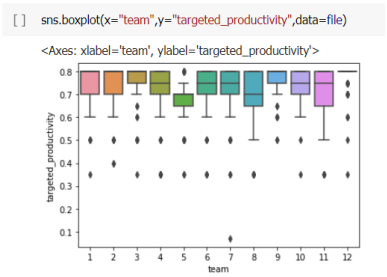

```
[ ] sns.boxplot(x="team",y="wip",data=file)
```

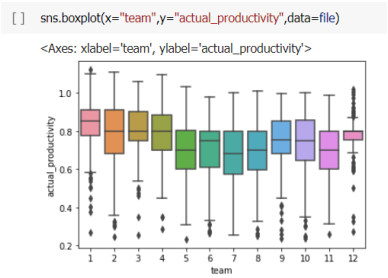


นายศิรวิทย์ เตชประสพชัย 63070501220
นางสาวฉันท์สินี เมืองหนู 63070501221

2. เปรียบเทียบระหว่าง targeted productivity และ actual productivity ที่เกิดขึ้นของแต่ละทีม

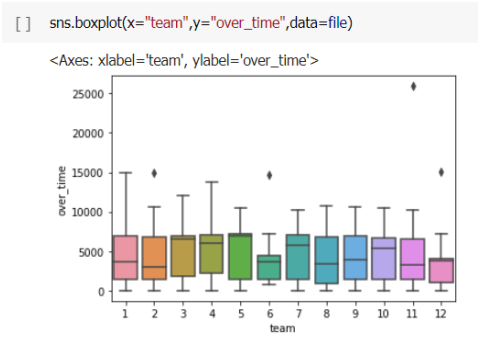
จากการพล็อต Boxplot เพื่อวิเคราะห์ความสัมพันธ์ระหว่าง team ทั้ง 12 ทีม และ targeted_productivity หรือประสิทธิภาพเป้าหมายจะพบว่าประสิทธิภาพเป้าหมายจะอยู่ในช่วง 0.7 – 0.8 ซึ่งพบว่า targeted_productivity จะอยู่ในเกณฑ์ที่ค่อนข้างสูงแต่จะพบว่าบางช่วงจะมี targeted_productivity ในช่วง 0.4 – 0.6 อาจเป็นเพราะในวันนั้นมีจำนวนของผลิตภัณฑ์ที่ต้องผลิตน้อย จึงมี targeted_productivity ที่ค่อนข้างต่ำ





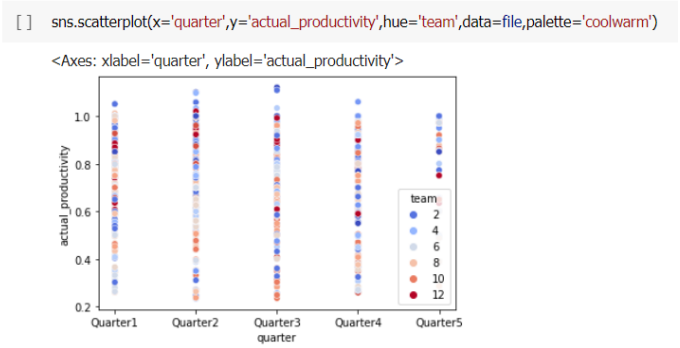
3. เปรียบเทียบการทำ over time (OT) ของแต่ละทีม

จากการพล็อต Boxplot เพื่อวิเคราะห์ความสัมพันธ์ระหว่าง team กับ over_time หรือช่วงเวลาที่ทำงานล่วงเวลาในหน่วยนาที จะพบว่าส่วนใหญ่ในแต่ละทีมจะทำงานล่วงเวลาในช่วง 2500 – 7000 นาที ตลอดทั้ง 4 quarter ซึ่งจะเห็นได้ชัดว่า Team 1, 5 มีช่วงเวลาการทำงานล่วงเวลาที่โดดเด่นซึ่งถือว่าเป็นทีมที่ทำงานล่วงเวลามากที่สุดเมื่อเทียบกับอีก 10 ทีมที่เหลือ



➢ Scatter plot

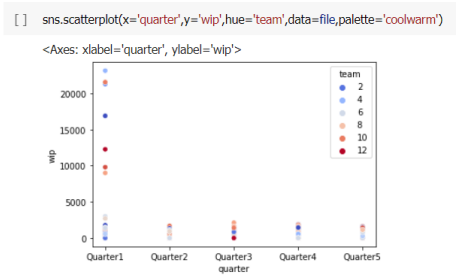
1. เปรียบเทียบ actual productivity ของแต่ละทีมในแต่ละช่วง quarter



นายศิรวิทย์ เตชประสพชัย 63070501220
นางสาวฉันท์สินี เมืองหนู 63070501221

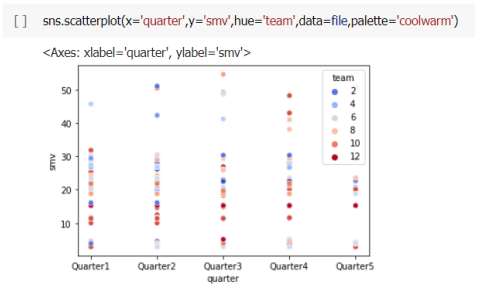
2. เปรียบเทียบ wip (งานที่เหลือตกค้างในแต่ละวัน) ของแต่ละทีมในแต่ละช่วง quarter

จากการพล็อต Scatterplot เพื่อวิเคราะห์ความสัมพันธ์ระหว่าง team กับ wip หรืองานที่ยังเหลืออยู่ในแต่ละวันที่ทำงาน เมื่อเทียบกันในช่วง quarter ต่างๆ จะพบว่าในช่วง quarter ที่ 1 จะเป็นช่วงที่มีการเกิด wip เกิดขึ้นสูงที่สุดเมื่อเทียบกับ quarter อื่นๆและทีมที่มี wip สูงที่สุดคือ team 2,4,10 โดยในส่วนใหญ่ Quarter อื่นๆจะมีจำนวน wip เกิดขึ้นค่อนข้างน้อยสำหรับแต่ละทีม



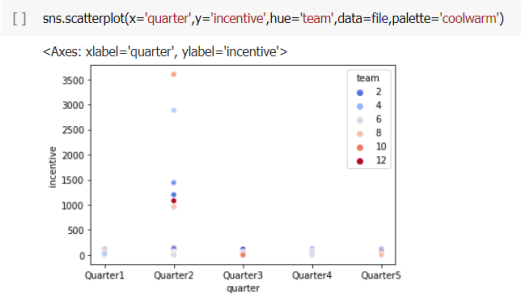
3. เปรียบเทียบการทำ smv (ระยะเวลาในการผลิตสินค้า 1 ชิ้น/นาที) ของแต่ละทีม

จากการพล็อต Scatterplot เพื่อวิเคราะห์ความสัมพันธ์ระหว่าง team กับ smv หรือช่วงเวลาในการผลิตผลิตภัณฑ์ 1 ผลิตภัณฑ์ในหน่วยนาที โดยเทียบกันในช่วง quarter ต่างๆ จะพบว่าโดยส่วนใหญ่แต่ละทีมจะมี smv อยู่ในช่วง 10 – 30 นาที ในทุก quarter ซึ่งหากพิจารณาในช่วง 10 นาทีจะเป็นช่วงที่มีการบันทึก smv ถี่ที่สุดและจึงพอสรุปได้ว่าส่วนใหญ่ team ทุกทีมจะมี smv อยู่ในช่วง 10 นาที



4. เปรียบเทียบการได้รับ incentive ของแต่ละทีมในแต่ละช่วง quarter

จากการพล็อต Scatterplot เพื่อวิเคราะห์ความสัมพันธ์ระหว่าง team กับ incentive หรือค่าตอบแทนพิเศษในการทำงาน ในแต่ละช่วง quarter จะพบว่า ในช่วง quarter ที่ 2 มีการจ่าย incentive ให้กับแรงงานสูงที่สุดเมื่อเทียบกับ quarter อื่นๆ และทีมที่ได้ incentive มากที่สุดคือ team ที่ 8 ซึ่งได้รับ incentive สูงถึง 3500 BDT

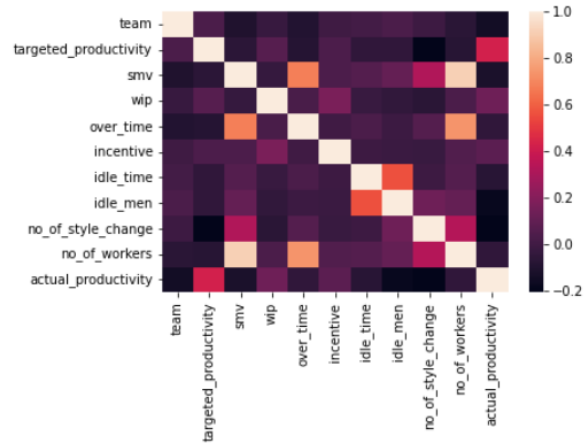


นายศิรวิทย์ เตชประสพชัย 63070501220
นางสาวฉันท์สินี เมืองหนู 63070501221

➢ Heatmap

จากการพล็อต Heatmap และเทียบข้อมูลกันระหว่าง Column จะพบว่าระหว่าง smv กับ actual_productivity มีความสัมพันธ์แปรผกผันกัน ซึ่งตรงกับหลักความเป็นจริงที่ว่าหากใช้เวลาในการผลิตผลิตภัณฑ์นานเท่าใด ประสิทธิภาพก็จะลดลงเท่านั้น

```
[95] sns.heatmap(file.corr())
```




```
[43] file.corr()
```

	targeted_productivity	smv	wip	over_time	incentive	idle_time	idle_men	no_of_style_change	no_of_workers	actual_productivity
targeted_productivity	1.000000	-0.069489	0.019035	-0.088557	0.032768	-0.056181	-0.053818	-0.209294	-0.084288	0.421594
smv	-0.069489	1.000000	0.322704	0.674887	0.032629	0.056863	0.105901	0.315388	0.912176	-0.122089
wip	0.019035	0.322704	1.000000	0.276529	0.037946	-0.005101	-0.007119	0.053293	0.373908	0.047389
over_time	-0.088557	0.674887	0.276529	1.000000	-0.004793	0.031038	-0.017913	0.059790	0.734164	-0.054206
incentive	0.032768	0.032629	0.037946	-0.004793	1.000000	-0.012024	-0.021140	-0.026607	0.049222	0.076538
idle_time	-0.056181	0.056863	-0.005101	0.031038	-0.012024	1.000000	0.559146	-0.011598	0.058049	-0.080851
idle_men	-0.053818	0.105901	-0.007119	-0.017913	-0.021140	0.559146	1.000000	0.133632	0.106946	-0.181734
no_of_style_change	-0.209294	0.315388	0.053293	0.059790	-0.026607	-0.011598	0.133632	1.000000	0.327787	-0.207366
no_of_workers	-0.084288	0.912176	0.373908	0.734164	0.049222	0.058049	0.106946	0.327787	1.000000	-0.057991
actual_productivity	0.421594	-0.122089	0.047389	-0.054206	0.076538	-0.080851	-0.181734	-0.207366	-0.057991	1.000000