

The 4Ws of Problem Scoping - Template

Nama : Eka Setiawan

Kelas : Persevere

Kasus : Efek Pandemi COVID-19 pada Kesehatan Mental Anak

WHO - Siapa yang memiliki masalah tersebut?

1. Orang yang sekolah atau melakukan pembelajaran melalui online

2. Anak – anak dan remaja

3. Beredarnya kabar covid dan meningkatnya kasus covid yang semakin tinggi mengakibatkan semua pembelajaran dilakukan dirumah atau secara online

WHAT - Apa permasalahannya sebenarnya?

1. Pandemi COVID-19 berdampak besar pada peningkatan gangguan masalah kesehatan akibat Depresi dan kecemasan pada anak di Indonesia.

2. Kurangnya interaksi sosial yang mempengaruhi gangguan kesehatan fisik dan mental, biasanya memiliki gejala stress karena jarang bertemu, malas merawat diri, menjadi sensitif dan agresif. Kecanduan Handphone, karena saat pandemi pembelajaran melalui online.

WHERE - Dimana/pada saat apa permasalahan ini muncul?

1. Ditempat dilakukkannya pembelajaran online, biasanya dirumah siswa melakukan pembelajaran online menggunakan handphone.

2. Pada saat pandemi Covid terus meningkat

WHY - Mengapa kamu yakin masalah ini sangat penting dibahas?

1. Pandemi Covid berdampak pada kesehatan mental anak

2. Dengan cara mengawasi atau menemani anak dalam pembelajaran, mengajak bermain atau mengobrol, melakukan aktivitas.

Ringkasan Subjek : Orang yang melakukan pembelajaran online Who Punya masalah berupa : Pandemi berdampak pada kesehatan mental anak What pada saat: Dirumah Where Solusi idealnya: Menemani dan mengajak ngobrol bersama Why

```
import matplotlib.pyplot as plt
import pandas as pd
import seaborn as sns
import numpy as np
from scipy import stats
```

Uji Beda 2 kelompok independent

Misalkan seorang CEO kafe ingin mengetahui apakah ada perbedaan yang signifikan antara jumlah pengunjung kafe di wilayah A dan B selama 2 minggu terakhir

A = [27, 29, 27, 32, 34, 38, 39, 27, 30, 32, 34, 36, 38, 40]

B = [29, 30, 29, 27, 30, 40, 39, 28, 30, 26, 27, 28, 39, 39]

```
A = np.array([27, 29, 27, 32, 34, 38, 39, 27, 30, 32, 34, 36, 38, 40])
B = np.array([28, 30, 29, 35, 34, 37, 39, 28, 30, 32, 33, 35, 39, 39])
```

Latihan 1 Coba lihat lagi cara uji normalitas di atas, coba lakukan pada data A dan B

Uji Normalitas H_0 = data berdistribusi normal

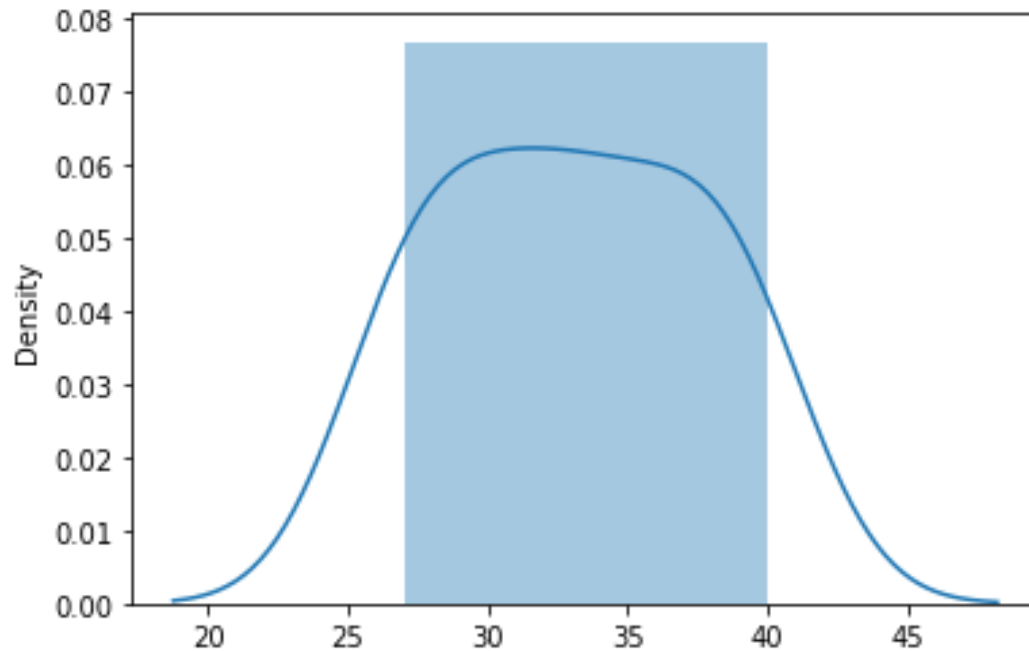
H_a = data tidak berdistribusi normal

```
#Cek Normalitas A Dengan histogram
sns.distplot(A)
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619:
FutureWarning: `distplot` is a deprecated function and will be removed in a
future version. Please adapt your code to use either `displot` (a figure-
level function with similar flexibility) or `histplot` (an axes-level
function for histograms).
```

```
warnings.warn(msg, FutureWarning)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fd4c39e7c90>
```



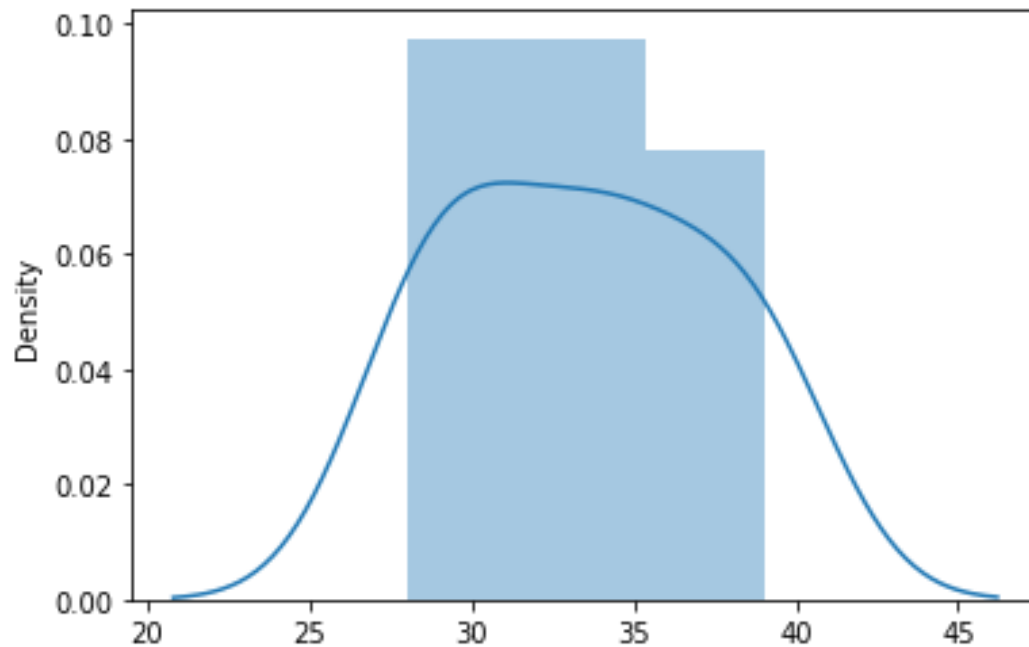
#Cek Normalitas B Dengan histogram

```
sns.distplot(B)
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619:  
FutureWarning: `distplot` is a deprecated function and will be removed in a  
future version. Please adapt your code to use either `displot` (a figure-  
level function with similar flexibility) or `histplot` (an axes-level  
function for histograms).
```

```
warnings.warn(msg, FutureWarning)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fd4c190f790>
```



#Uji Kolmogorov-Smirnov

```
z_score_A=stats.zscore(A)
```

```
z_score_B=stats.zscore(B)
```

```
print('Uji KS data A adalah ', stats.kstest(z_score_A,'norm'))
```

```
print('Uji KS data B adalah ', stats.kstest(z_score_B,'norm'))
```

```
Uji KS data A adalah  KstestResult(statistic=0.15010032999139777,
pvalue=0.9106696943287013)
```

```
Uji KS data B adalah  KstestResult(statistic=0.16609147661220497,
pvalue=0.8346609194732189)
```

Apakah keduanya berdistribusi normal? bagaimana kamu mengetahuinya?

Data berdistribusi normal, karena kedua p-value > 0,05 maka H_0 diterima

Independent Sample T-Test (Parametrik)

H_0 = Tidak ada perbedaan yang signifikan (keduanya sama)

H_a = ada perbedaan yang signifikan (artinya lokasi kafe berpengaruh signifikan terhadap jumlah pengunjung)

```
t_value = stats.ttest_ind(A, B)
```

```
print('hasil uji t: ', t_value)
```

```
hasil uji t:  Ttest_indResult(statistic=-0.21626215417063352,
pvalue=0.8304712072353047)
```

Uji Mann-Whitney

Misalkan datanya tidak berdistribusi normal, uji beda 2 kelompok independent dapat dilakukan dengan uji Mann-whitney. Caranya

```
stats.mannwhitneyu(A, B)
```

```
MannwhitneyuResult(statistic=91.0, pvalue=0.38212917802691737)
```

Latihan 2

Tulis kesimpulanmu disini apakah H_0 diterima atau ditolak dan apa artinya

Karena p-value $0,38 > 0,05$ maka H_0 diterima, tidak ada perbedaan yang signifikan antara jumlah pengunjung kafe wilayah A dan B, berarti lokasi kafe tidak berpengaruh signifikan terhadap jumlah pengunjung

Uji Beda lebih dari 2 kelompok

Dalam statistik parametrik, nama metodenya adalah ANOVA. ANOVA memiliki 2 asumsi :

1. Data Berdistribusi Normal
2. Varians data homogen

Kasus

Misalkan saat ini kamu adalah menteri pendidikan yang ingin mengetahui apakah kualitas pendidikan di negara ini sudah merata atau belum. Daerah dibagi menjadi kota besar di wilayah barat, kota kecil di barat, ibukota, kota besar di wilayah timur, dan kota kecil di timur. Analisa ini dilakukan berdasarkan nilai ujian akhir SMA

```
url='https://drive.google.com/file/d/1FLxWueFCFJGqDBDJmxDESfxQqJcZQ5ew/view'
```

```
url='https://drive.google.com/uc?id=' + url.split('/')[2]
```

```
final_exam = pd.read_csv(url)
```

```
final_exam
```

```
#https://drive.google.com/file/d/1FLxWueFCFJGqDBDJmxDESfxQqJcZQ5ew/view
```

	big_west	small_west	capital_city	big_east	small_east
0	87	77	86	71	70
1	75	83	90	82	76
2	80	73	87	85	81
3	75	82	85	89	85
4	91	71	85	72	75
5	89	87	96	79	75
6	91	77	95	79	75
7	87	84	89	85	72
8	90	86	93	84	72
9	94	93	91	86	75
10	82	83	96	85	90
11	91	94	94	76	83
12	83	79	91	71	80
13	86	92	85	82	85

14	84	72	77	72	80
15	82	86	96	78	80
16	88	76	84	78	78
17	89	84	87	76	86
18	94	85	91	74	83
19	87	93	83	77	77
20	89	87	90	74	76
21	88	89	81	70	75
22	75	89	79	80	81
23	85	77	95	89	74
24	81	92	93	80	84
25	78	73	95	89	75
26	93	86	90	77	75
27	88	87	91	88	80
28	84	93	78	75	80
29	90	95	91	78	75
30	91	90	88	82	70
31	88	84	80	72	80
32	82	82	95	77	90
33	94	82	94	81	86
34	83	93	90	75	70
35	91	72	78	78	85
36	76	85	89	75	80
37	80	90	90	87	72
38	76	81	90	76	90
39	90	90	95	81	90
40	88	78	81	89	63
41	84	83	90	82	82
42	70	70	70	70	70

```
from google.colab import drive
drive.mount('/content/drive')
```

Mounted at /content/drive

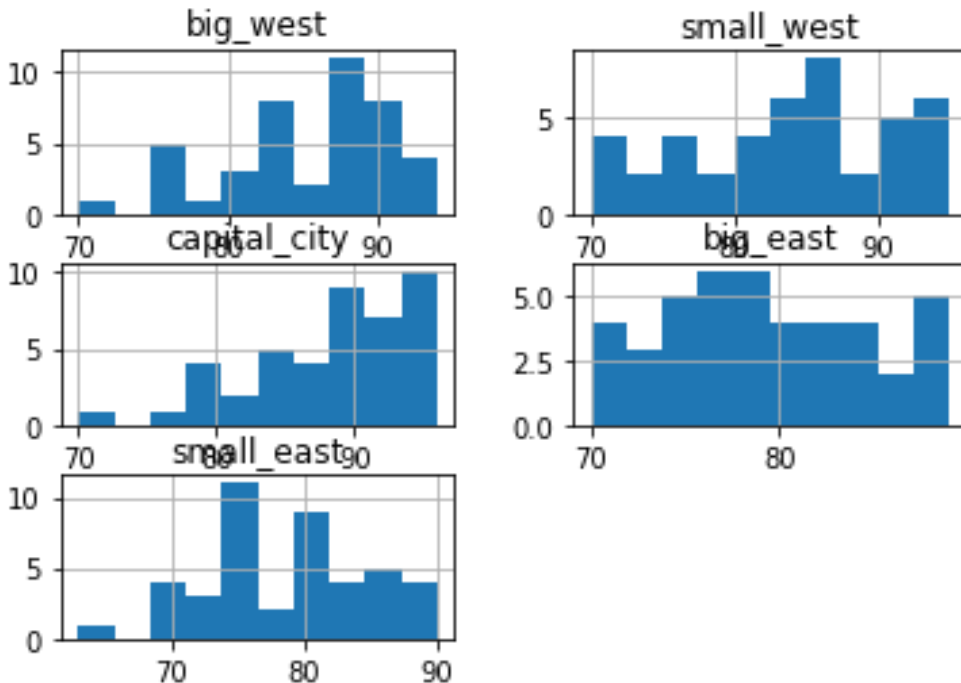
Uji Normalitas

Ho = data berdistribusi normal

Ha = data tidak berdistribusi normal

#Membuat histogram dari pandas dan matplotlib

```
final_exam.hist()
plt.show()
```



Terlihat beberapa datanya seperti tidak normal, mari kita coba uji Kolmogorov-Smirnov untuk mengeceknya

#Check normality using Kolmogorov-Smirnov Test

```
area = ['big_west', 'small_west', 'capital_city', 'small_east', 'big_east']
for a in area:
    z_score = stats.zscore(final_exam[a])
    normality = stats.kstest(z_score, 'norm', )
    print('normality for ', a, '=', normality)
```

```
normality for big_west = KstestResult(statistic=0.1463623852887705,
pvalue=0.2875537193189711)
normality for small_west = KstestResult(statistic=0.09305748870693209,
pvalue=0.8504695155077621)
normality for capital_city = KstestResult(statistic=0.1723544820281414,
pvalue=0.13793715190373557)
normality for small_east = KstestResult(statistic=0.11522951440351409,
pvalue=0.5926673369890815)
normality for big_east = KstestResult(statistic=0.09709346617838455,
pvalue=0.8123161041508566)
```

Latihan 3

Coba simpulkan apakah artinya data berdistribusi normal?

Karena setiap area memiliki nilai P Value > 0,05 Maka data berdistribusi normal

Uji Homogenitas

Ho = varians semua kelompok cenderung sama (homogen)

Ha = varians semua kelompok cenderung tidak sama (homogen)

#Uji Homogenitas dengan Levene

```
stats.levene(final_exam['big_west'], final_exam['small_west'],  
final_exam['capital_city'],  
            , final_exam['small_east'], final_exam['big_east'])
```

```
LeveneResult(statistic=0.5104303607286808, pvalue=0.728125042716542)
```

Karena p-value > 0.05, maka Ho diterima sehingga asumsi homogenitas dari ANOVA sudah terpenuhi

Uji ANOVA (parametrik)

Ho = kualitas pendidikan merata (tidak ada perbedaan kualitas pendidikan yang signifikan antar wilayah)

Ha = kualitas pendidikan tidak merata (ada perbedaan kualitas pendidikan yang signifikan antar wilayah)

```
stats.f_oneway(final_exam['big_west'], final_exam['small_west'],  
final_exam['capital_city'],  
            , final_exam['small_east'], final_exam['big_east'])
```

```
F_onewayResult(statistic=18.48895253915249, pvalue=4.945434101915241e-13)
```

Karena p-value < 0.05, maka Ho ditolak, ternyata kualitas pendidikan di negara ini belum merata. Untuk itu, kamu perlu menyelidiki wilayah dengan rata-rata terendah agar kualitas sarana pendidikan dan gurunya bisa ditingkatkan

```
final_exam.describe()
```

	big_west	small_west	capital_city	big_east	small_east
count	43.000000	43.000000	43.000000	43.000000	43.000000
mean	85.325581	83.837209	88.232558	79.209302	78.627907
std	5.982995	7.006562	6.159921	5.667611	6.302015
min	70.000000	70.000000	70.000000	70.000000	63.000000
25%	82.000000	78.500000	85.000000	75.000000	75.000000
50%	87.000000	84.000000	90.000000	78.000000	80.000000
75%	90.000000	89.500000	93.000000	83.000000	83.000000
max	94.000000	95.000000	96.000000	89.000000	90.000000

Ternyata nilai rata-rata di ibukota jauh diatas wilayah lain, sementara kota kecil di timur mendapat nilai rata-rata terkecil. Sebagai menteri pendidikan berarti kamu harus membuat program untuk membangun sarana pendidikan yang lebih baik dan mengirim guru-guru terbaik ke kota kecil di wilayah timur.

Uji Kruskal-Wallis (nonparametrik)

Jika datanya ternyata tidak berdistribusi normal atau tidak homogen, kamu bisa menggunakan uji Kruskal Wallis sebagai berikut


```
stats.kruskal(final_exam['big_west'], final_exam['small_west'],
             final_exam['capital_city'],
             final_exam['small_east'], final_exam['big_east'])

KruskalResult(statistic=57.244594069662135, pvalue=1.0992920086269013e-11)
```

Latihan 4

Coba bikin kesimpulan dari hasil p-value tersebut

Karena nilai P Value $1,09e-11 < 0,05$ maka H_0 ditolak, berarti bahwa kualitas pendidikan tidak merata, ada perbedaan kualitas pendidikan yang signifikan antar wilayah

Latihan tambahan

Misalkan kamu adalah data scientist di gojek. Kamu ingin mengetahui apakah kemacetan mempengaruhi tingkat kepuasan pelanggan. Jadi kamu ingin memebandingkan nilai "bintang" dari customer saat dia order di jam macet dan tidak macet. (#hint = data dari customer yang sama)

```
macet = [2, 3, 5, 4, 5, 4, 4, 3, 4, 3, 4, 5, 4, 5, 4, 3, 4, 3, 3, 5, 4, 3, 4, 3, 4, 3, 4, 4, 5]
```

```
tidak_macet = [2, 2, 4, 5, 4, 3, 3, 3, 3, 3, 3, 3, 4, 4, 5, 4, 3, 3, 2, 3, 4, 4, 3, 3, 3, 4, 3, 4, 3, 4]
```

1. Buat hipotesisnya
2. Lakukan uji asumsi normalitasnya
3. Jika datanya normal lakukan uji parametrik, jika tidak berdistribusi normal lakukan uji nonparametrik
4. Apa kesimpulan uji hipotesisnya?

```
macet = np.array([2, 3, 5, 4, 5, 4, 4, 3, 4, 3, 4, 5, 4, 5, 4, 3, 4, 3, 3, 5, 4, 3, 4, 3, 4, 4, 5,
4, 3, 4, 3, 4, 3, 4, 4, 5])
tidak_macet = np.array([2, 2, 4, 5, 4, 3, 3, 3, 3, 3, 3, 3, 4, 4, 5, 4, 3, 3, 2, 3, 4, 4, 3, 3, 3, 4, 3, 4, 3, 4,
3, 4, 4, 3, 3, 3, 4, 3, 4, 3, 4])
```

Uji Normalitas

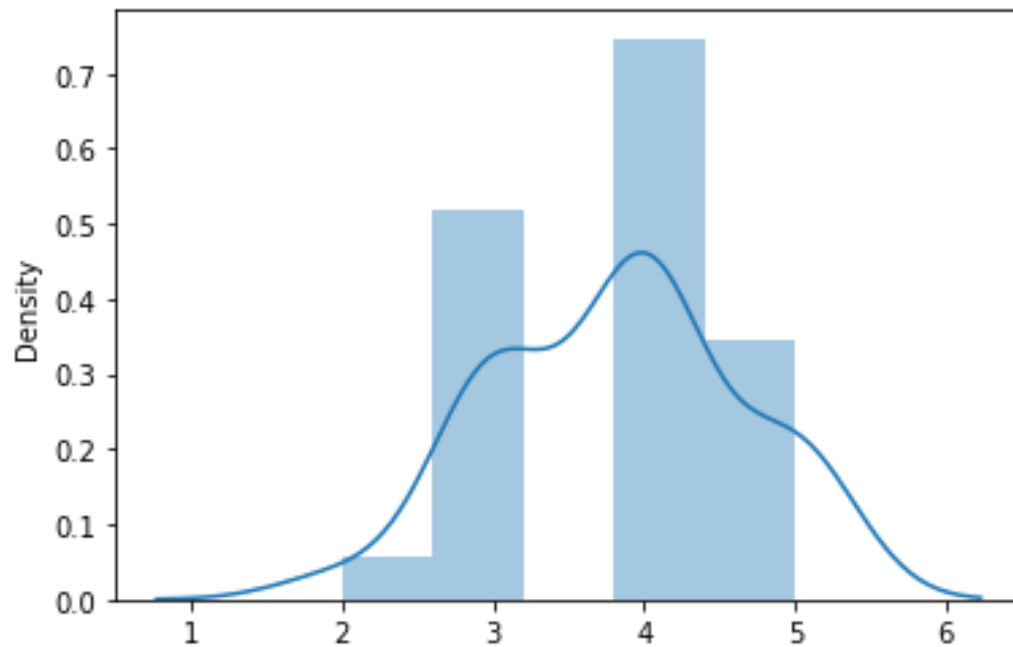
H_0 = Data berdistribusi normal

H_a = Data tidak berdistribusi normal

```
#Cek normalitas data macet dengan histogram
sns.distplot(macet)
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619:
FutureWarning: `distplot` is a deprecated function and will be removed in a
future version. Please adapt your code to use either `displot` (a figure-
level function with similar flexibility) or `histplot` (an axes-level
function for histograms).
  warnings.warn(msg, FutureWarning)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fd4c0bc7590>
```



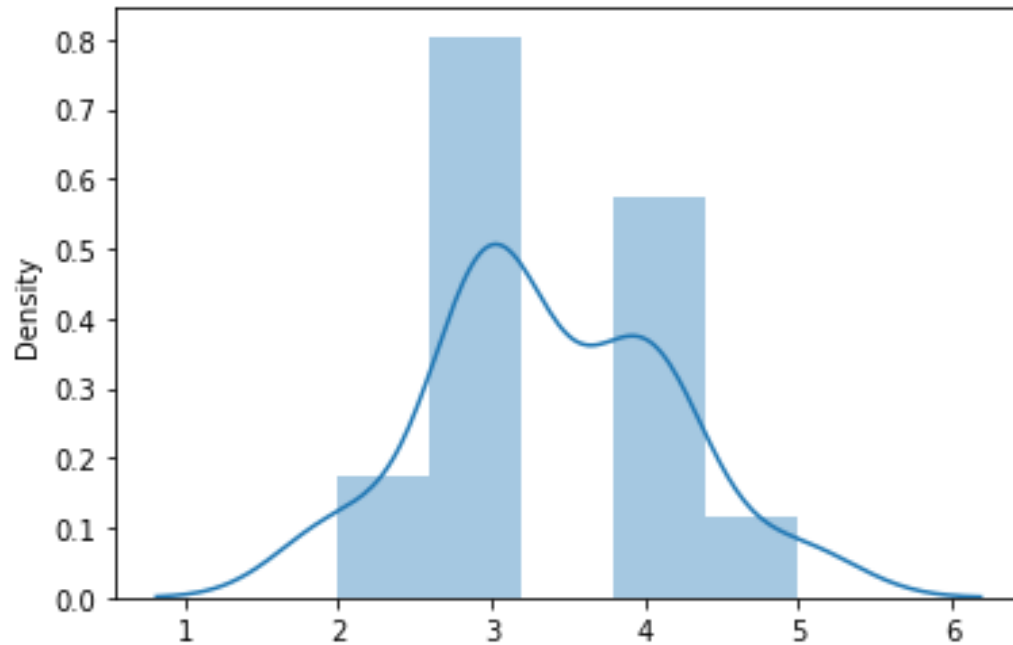
#Cek normalitas data tidak_macet dengan histogram

```
sns.distplot(tidak_macet)
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619:  
FutureWarning: `distplot` is a deprecated function and will be removed in a  
future version. Please adapt your code to use either `displot` (a figure-  
level function with similar flexibility) or `histplot` (an axes-level  
function for histograms).
```

```
warnings.warn(msg, FutureWarning)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fd4c0b6b750>
```



#Uji normalitas data dengan Kolmogorov-Smirnov

```
z_score_macet = stats.zscore(macet)
z_score_tidak_macet = stats.zscore(tidak_macet)
print('Uji KS data macet adalah ', stats.kstest(z_score_macet, 'norm'))
print('Uji KS data tidak_macet adalah ',
stats.kstest(z_score_tidak_macet, 'norm'))
```

Uji KS data macet adalah KstestResult(statistic=0.2414618941043637,
pvalue=0.05668593129743599)

Uji KS data tidak_macet adalah KstestResult(statistic=0.2769460572882385,
pvalue=0.01855050750663906)

Data macet berdistribusi normal, karena nilai P Value = 0,056 > 0,05 maka Ho diterima

Data tidak_macet tidak berdistribusi normal, karena nilai P Value = 0,018 < 0,05 maka Ho ditolak

Karena nilai P Value data macet hanya sedikit diatas 0,05 maka kedua data dianggap tidak berdistribusi normal

Uji Wilcoxon (nonparametrik)

Ho = Tidak ada perbedaan yang signifikan (keduanya sama)

Ha = ada perbedaan yang signifikan (artinya tingkat kemacetan berpengaruh signifikan terhadap tingkat kepuasan pelanggan)

```
d = macet - tidak_macet
stats.wilcoxon(d)
```

`WilcoxonResult(statistic=8.0, pvalue=0.0007891129890156299)`

Nilai P Value = $0,000789 < 0,05$ maka H_0 ditolak, berarti ada perbedaan yang signifikan pada tingkat kepuasan pelanggan, tingkat kemacetan berpengaruh signifikan terhadap tingkat kepuasan pelanggan.

Kesimpulan

Data dianggap tidak berdistribusi normal sehingga dilakukan Uji Wilcoxon (nonparametrik) dan hasilnya H_0 ditolak. Hal ini berarti tingkat kepuasan pelanggan dipengaruhi secara signifikan oleh tingkat kemacetan.