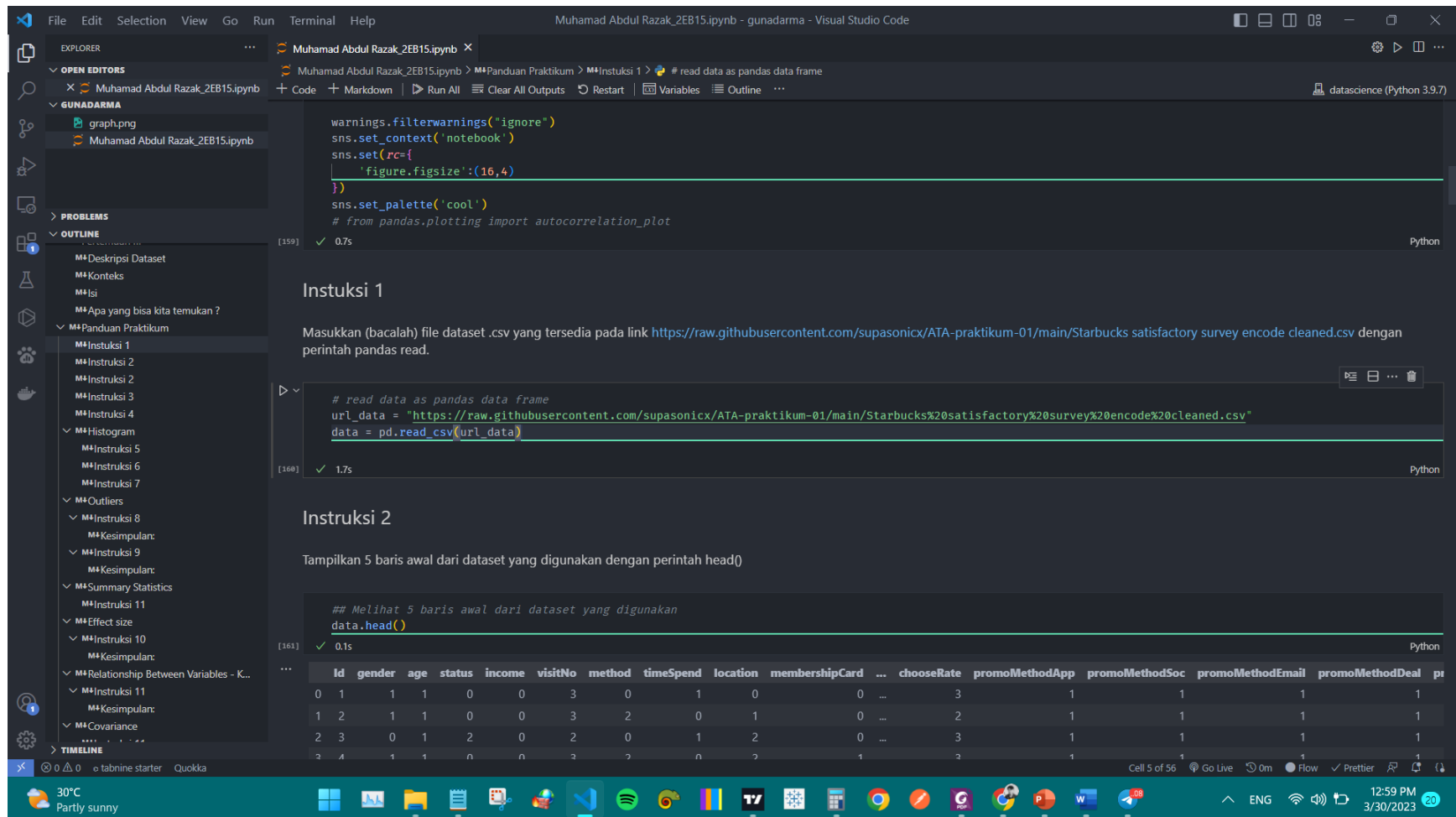


# INSTRUKSI 1



The screenshot shows a Visual Studio Code editor with a Jupyter Notebook open. The notebook has two cells. The first cell contains code to read a CSV file from a GitHub link. The second cell contains code to display the first 5 rows of the dataset using the `head()` method. The output of the second cell is a table with 17 columns and 4 rows of data.

**Instuksi 1**

Masukkan (bacalah) file dataset .csv yang tersedia pada link [https://raw.githubusercontent.com/supasonix/ATA-praktikum-01/main/Starbucks satisfactory survey encode cleaned.csv](https://raw.githubusercontent.com/supasonix/ATA-praktikum-01/main/Starbucks%20satisfactory%20survey%20encode%20cleaned.csv) dengan perintah pandas read.

```
# read data as pandas data frame
url_data = "https://raw.githubusercontent.com/supasonix/ATA-praktikum-01/main/Starbucks%20satisfactory%20survey%20encode%20cleaned.csv"
data = pd.read_csv(url_data)
```

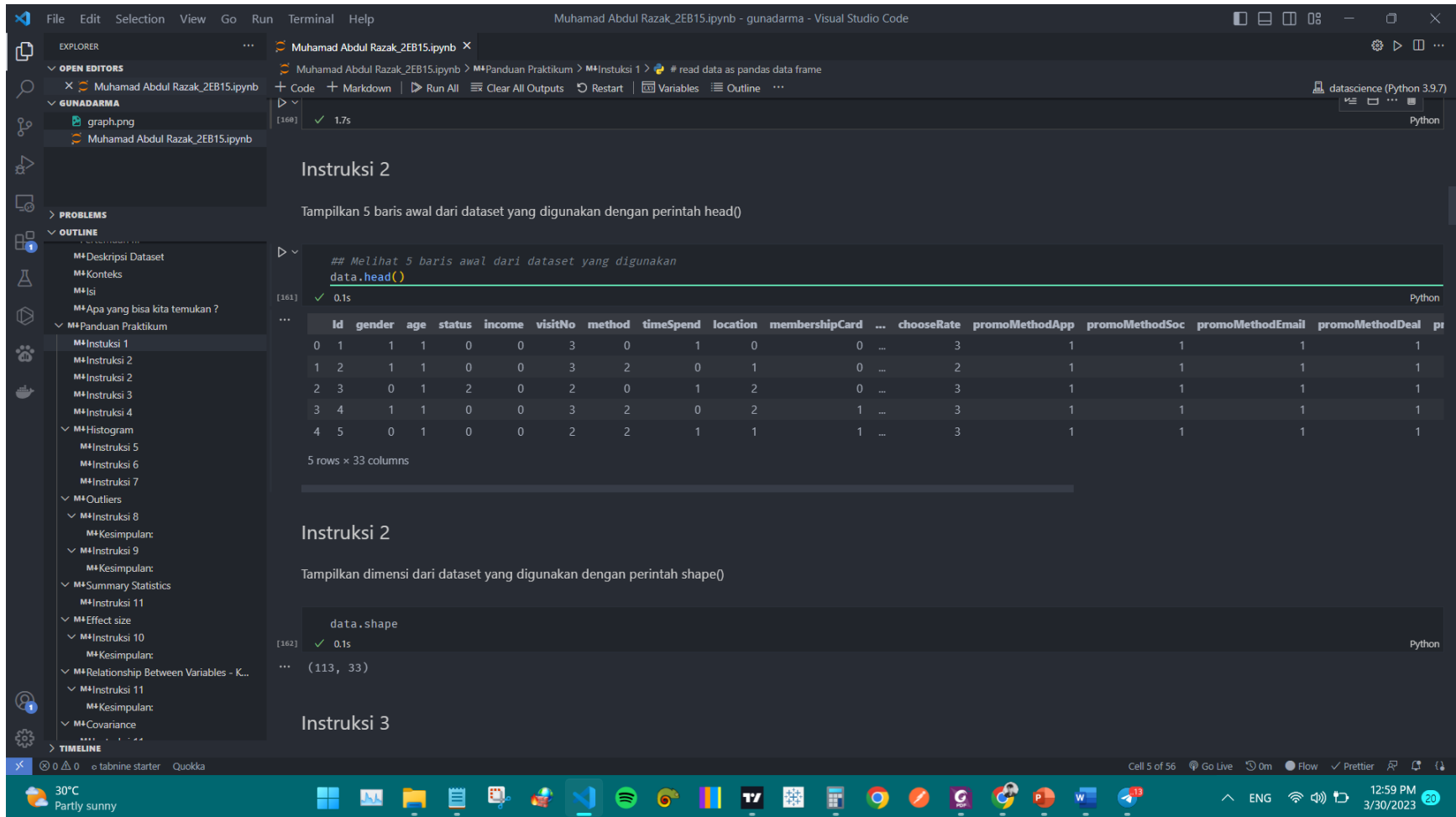
**Instuksi 2**

Tampilkan 5 baris awal dari dataset yang digunakan dengan perintah `head()`

```
## Melihat 5 baris awal dari dataset yang digunakan
data.head()
```

	Id	gender	age	status	income	visitNo	method	timeSpend	location	membershipCard	...	chooseRate	promoMethodApp	promoMethodSoc	promoMethodEmail	promoMethodDeal	ps
0	1	1	1	0	0	3	0	1	0	0	...	3	1	1	1	1	
1	2	1	1	0	0	3	2	0	1	0	...	2	1	1	1	1	
2	3	0	1	2	0	2	0	1	2	0	...	3	1	1	1	1	
3	4	1	1	0	0	3	2	0	2	1	...	2	1	1	1	1	

## INSTRUKSI 2



The screenshot shows a Visual Studio Code editor with a Jupyter Notebook open. The notebook has two instructions, both of which have been executed successfully.

**Instruksi 2**

Tampilkan 5 baris awal dari dataset yang digunakan dengan perintah head()

```
## Melihat 5 baris awal dari dataset yang digunakan
data.head()
```

Output:

	Id	gender	age	status	income	visitNo	method	timeSpent	location	membershipCard	...	chooseRate	promoMethodApp	promoMethodSoc	promoMethodEmail	promoMethodDeal	pr
0	1	1	1	0	0	3	0	1	0	0	...	3	1	1	1	1	1
1	2	1	1	0	0	3	2	0	1	0	...	2	1	1	1	1	1
2	3	0	1	2	0	2	0	1	2	0	...	3	1	1	1	1	1
3	4	1	1	0	0	3	2	0	2	1	...	3	1	1	1	1	1
4	5	0	1	0	0	2	2	1	1	1	...	3	1	1	1	1	1

5 rows x 33 columns

**Instruksi 2**

Tampilkan dimensi dari dataset yang digunakan dengan perintah shape()

```
data.shape
```

Output:

```
(113, 33)
```

**Instruksi 3**

## INSTRUKSI 3

The screenshot shows a Visual Studio Code editor window with a Jupyter Notebook open. The notebook has two cells. The first cell contains the code `data.shape` and its output is `(113, 33)`. The second cell contains the code `data.columns` and its output is a list of 33 column names: `Index(['Id', 'gender', 'age', 'status', 'income', 'visitNo', 'method', 'timeSpend', 'location', 'membershipCard', 'itemPurchaseCoffee', 'itemPurchaseCold', 'itemPurchasePastries', 'itemPurchaseJuices', 'itemPurchaseSandwiches', 'itemPurchaseOthers', 'spendPurchase', 'productRate', 'priceRate', 'promoRate', 'ambianceRate', 'wifiRate', 'serviceRate', 'chooseRate', 'promoMethodApp', 'promoMethodSoc', 'promoMethodEmail', 'promoMethodDeal', 'promoMethodFriend', 'promoMethodDisplay', 'promoMethodBillboard', 'promoMethodOthers', 'loyal'], dtype='object')`. The left sidebar shows a file explorer with a folder named 'GUNADARMA' and a list of files including 'Instruksi 1' through 'Instruksi 11'. The bottom status bar shows the temperature as 30°C and the date as 3/30/2023.

```
data.shape
[162] ✓ 0.1s
... (113, 33)
```

### Instruksi 3

Tampilkan kolom apa saja yang dimiliki oleh dataset yang digunakan dengan perintah `columns()`

```
data.columns
[163] ✓ 0.1s
... Index(['Id', 'gender', 'age', 'status', 'income', 'visitNo', 'method',
' timeSpend', 'location', 'membershipCard', 'itemPurchaseCoffee',
' itemPurchaseCold', 'itemPurchasePastries', 'itemPurchaseJuices',
' itemPurchaseSandwiches', 'itemPurchaseOthers', 'spendPurchase',
' productRate', 'priceRate', 'promoRate', 'ambianceRate', 'wifiRate',
' serviceRate', 'chooseRate', 'promoMethodApp', 'promoMethodSoc',
' promoMethodEmail', 'promoMethodDeal', 'promoMethodFriend',
' promoMethodDisplay', 'promoMethodBillboard', 'promoMethodOthers',
' loyal'],
dtype='object')
```

```
### menampilkan nilai apa saja yang ada pada sebuah kolom
data['gender'].unique()
[164] ✓ 0.1s
... array([1, 0], dtype=int64)
```

### Instruksi 4

Tampilkan nilai berapa saja yang dimiliki oleh kolom dimiliki oleh `timeSpend`

## INSTRUKSI 4

The screenshot shows a Visual Studio Code window with a Jupyter Notebook open. The notebook has two code cells. The first cell contains the following code:

```
## menampilkan nilai apa saja yang ada pada sebuah kolom
data['gender'].unique()
```

The output of the first cell is:

```
[164] ✓ 0.1s
... array([1, 0], dtype=int64)
```

The second cell contains the following code:

```
data.timeSpend.unique()
```

The output of the second cell is:

```
[165] ✓ 0.1s
... array([1, 0, 4, 2, 3], dtype=int64)
```

The left sidebar shows a file explorer with a folder named 'GUNADARMA' and a list of instructions. The bottom status bar shows the system clock and weather.

### Instruksi 4

Tampilkan nilai berapa saja yang dimiliki oleh kolom dimiliki oleh timeSpend.

```
data.timeSpend.unique()
```

```
[165] ✓ 0.1s
... array([1, 0, 4, 2, 3], dtype=int64)
```

### Histogram

- Menampilkan Berapa kali (frekuensi) setiap nilai muncul dalam kumpulan data.
- Jenis deskripsi ini disebut distribusi variabel
- Cara paling umum untuk merepresentasikan distribusi variabel adalah histogram yaitu grafik yang menunjukkan frekuensi dari setiap nilai.
- Frequency = berapa kali setiap nilai muncul
- Contoh: [1,1,1,1,2,2,2]. Frequency dari 1 adalah empat dan frequency dari 2 adalah tiga.

### Instruksi 5

Tampilkan histogram dari kolom timeSpend

```
def histBoxPlot(feature, figsize=(16,4), title='', show value=False):
```

# INSTRUKSI 5

Muhamad Abdul Razak\_2EB15.ipynb - gunadarma - Visual Studio Code

File Edit Selection View Go Run Terminal Help

EXPLORER

Muhamad Abdul Razak\_2EB15.ipynb

File-Praktikum Soshum-Week-3.ipynb

OPEN EDITORS

Muhamad Abdul Razak\_2EB15.ipynb

File-Praktikum Soshum-Week-3.ipynb

GUNADARMA

graph.png

Muhamad Abdul Razak\_2EB15.ipynb

PROBLEMS

OUTLINE

MA Pertemuan III

MA Deskripsi Dataset

MA Konteks

MA Misi

MA Apa yang bisa kita temukan ?

MA Panduan Praktikum

MA Instruksi 1

MA Instruksi 2

MA Instruksi 3

MA Instruksi 4

MA Histogram

MA Instruksi 5

MA Instruksi 6

MA Instruksi 7

MA Outliers

MA Instruksi 8

MA Kesimpulan:

MA Instruksi 9

MA Kesimpulan:

MA Summary Statistics

MA Instruksi 11

MA Effect size

MA Instruksi 10

MA Kesimpulan:

MA Relationship Between Variables - K...

MA Instruksi 11

MA Kesimpulan:

MA Covariance

MA Instruksi 11

MA Pearson Correlation

MA Instruksi 12

MA Kesimpulan:

MA Spearman's Rank Correlation

MA Instruksi 13

TIMELINE

o tabnine starter Quokka

Cell 16 of 56 Go Live 1m Flow Prettier

30°C Partly sunny

1:04 PM 3/30/2023

Muhamad Abdul Razak\_2EB15.ipynb

MA Instruksi 5

Contoh: [1,1,1,1,2,2,2]. Frequency dari 1 adalah empat dan frequency dari 2 adalah tiga.

Instruksi 5

Tampilkan histogram dari kolom timeSpend

```
def histBoxPlot(feature, figsize=(16,4), title='', show_value=False):
    import scipy.stats as stats
    f,(ax_box,ax_hist)=plt.subplots(nrows=2,sharex=True, gridspec_kw={'height_ratios':(0.1,0.9)},figsize=figsize)
    sns.histplot(feature,kde=True,ax=ax_hist,color='violet',label='stdev %.4f\nmode %i' % (np.std(feature),int(''.join(map(str, stats.mode(feature).mode)))).set(xlabel='')
    sns.boxplot(x=feature,ax=ax_box,color='violet').set(xlabel=None)
    ax_hist.axvline(np.mean(feature),color='purple',linestyle=':',label=f'mean {np.mean(feature):.4f}')
    ax_hist.axvline(np.median(feature),color='purple',linestyle='--',label=f'median {np.median(feature):.4f}')
    if show_value:
        ax_hist.bar_label(ax_hist.containers[0])
    plt.legend()
    plt.suptitle(title)
    plt.tight_layout()
```

```
## Buatlah diagram histogram dari kolom Strength
histBoxPlot(data.timeSpend, title='Histogram of timeSpend', figsize=(16,4))
```

Histogram of timeSpend

Count

mean 0.6106  
median 0.0000  
stdev 0.8460  
mode 0

## INSTRUKSI 6

Visual Studio Code interface showing a Jupyter Notebook titled "Instruksi 6". The notebook content includes a Python code snippet for plotting a histogram and a corresponding bar chart.

**Code Snippet:**

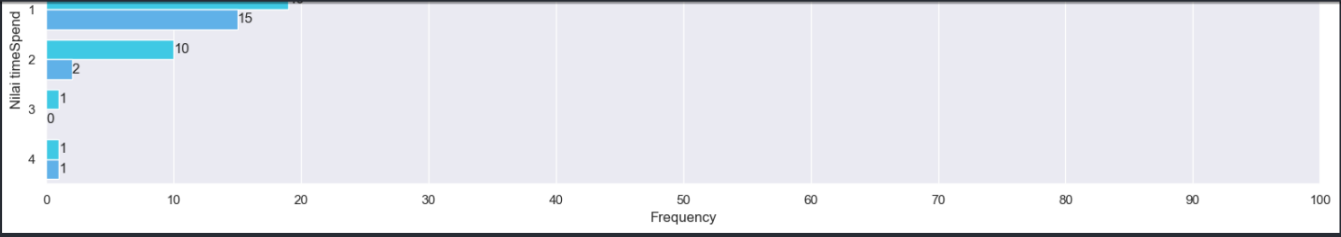
```
## plotting untuk data produk pastry
f, ax = plt.subplots()
sns.countplot(data=data, y='timeSpent', hue='gender').set(xticks=[i for i in np.arange(0,101,10)])
ax.bar_label(ax.containers[0])
ax.bar_label(ax.containers[1])
plt.ylabel("Nilai timeSpent")
plt.xlabel("Frequency")
plt.title("Histogram waktu yang dihabiskan di toko berdasarkan demografi gender")
plt.legend(['Female', 'Male'])
plt.tight_layout()
```

**Bar Chart Data:**

Nilai timeSpent	Female	Male
0	23	41
1	19	15
2	10	2
3	1	0
4	1	1

The chart is titled "Histogram waktu yang dihabiskan di toko berdasarkan demografi gender". The x-axis is labeled "Frequency" and ranges from 0 to 100. The y-axis is labeled "Nilai timeSpent" and ranges from 0 to 4. The legend indicates that blue bars represent "Female" and orange bars represent "Male".

## INSTRUKSI 7



Nilai timeSpend	Frequency
1	15
2	10
3	1
4	1

### Instruksi 7

Buatlah kesimpulan dari informasi yang Anda dapatkan di atas.

- Feature timeSpend terdistribusi positive skew dengan nilai modus 0 dan median 0. Berdasarkan histogram dapat dilihat bahwa waktu yang dihabiskan customer kebanyakan kurang dari 2 jam.
- Product Pastries cenderung dibeli oleh customer pria dan perempuan dengan jumlah customer pria sebanyak 59 dan customer perempuan sebanyak 54.
- Kebanyakan customer menghabiskan waktu kurang dari 1 jam dengan didominasi pria sebanyak 41 orang.

### Outliers

- Nilai outliers bisa diartikan sebagai kesalahan atau kejadian langka.
- Menghitung outlier:
  - pertama kita perlu menghitung kuartil pertama (Q1)(25%)
  - kemudian temukan IQR(inter quartile range) =  $Q3 - Q1$
  - terakhir menghitung  $Q1 - 1.5 \cdot IQR$  dan  $Q3 + 1.5 \cdot IQR$
  - Apa pun di luar kisaran ini adalah outlier
  - mari kita tulis kode untuk distribusi tumor jinak untuk rata-rata fitur radius

### Instruksi 8

## INSTRUKSI 8

File Edit Selection View Go Run Terminal Help

Muhamad Abdul Razak\_2EB15.ipynb - gunadarma - Visual Studio Code

EXPLORER

- OPEN EDITORS
  - Muhamad Abdul Razak\_2EB15.ipynb
  - File-Praktikum Soshum-Week-3.ipynb...
- GUNADARMA
  - graph.png
  - Muhamad Abdul Razak\_2EB15.ipynb
- PROBLEMS
- OUTLINE
  - M+Pertemuan III
    - M+Deskripsi Dataset
    - M+Konteks
    - M+Isi
    - M+Apa yang bisa kita temukan?
  - M+Panduan Praktikum
    - M+Instruksi 1
    - M+Instruksi 2
    - M+Instruksi 3
    - M+Instruksi 4
  - M+Histogram
    - M+Instruksi 5
    - M+Instruksi 6
    - M+Instruksi 7
  - M+Outliers
    - M+Instruksi 8
    - M+Kesimpulan:
    - M+Instruksi 9
    - M+Kesimpulan:
  - M+Summary Statistics
    - M+Instruksi 11
  - M+Effect size
    - M+Instruksi 10
  - M+Relationship Between Variables - K...
  - M+Instruksi 11
  - M+Kesimpulan:
- TIMELINE

Muhamad Abdul Razak\_2EB15.ipynb

Code + Markdown Run All Clear All Outputs Restart Variables Outline

datascience (Python 3.9.7)

mari kita tulis kode untuk distribusi tumor jinak untuk rata-rata fitur radius

### Instruksi 8

Buatlah kesimpulan dari informasi yang Anda dapatkan dari perhitungan nilai outliers dari variabel timeSpend pada data customer dengan gender laki-laki

```
## Menghitung nilai outliers dari variabel timeSpend pada data customer dengan gender laki-laki
male = data[data["gender"] == 1]
female = data[data["gender"] == 0]
desc = male.timeSpend.describe()
Q1 = desc[4]
Q3 = desc[6]
IQR = Q3-Q1
lower_bound = Q1 - 1.5*IQR
upper_bound = Q3 + 1.5*IQR
print("Apa pun di luar kisaran ini adalah outlier timeSpend: (", lower_bound, ",", upper_bound,")")
male[male.timeSpend < lower_bound].timeSpend
print("Outliers: ",male[(male.timeSpend < lower_bound) | (male.timeSpend > upper_bound)].timeSpend.values)
```

[179] ✓ 0.1s

... Apa pun di luar kisaran ini adalah outlier timeSpend: ( -1.5 , 2.5 )

Outliers: [4]

Kesimpulan:

Berdasarkan hasil perhitungan diketahui bahwa jumlah customer pria yang merupakan outliers dari feature timeSpend hanya 1 orang dimana batas bawah outliers dan batas atas outliers adalah -1,5 dan 2,5.

### Instruksi 9

Hitung nilai outliers dari variabel visitNo (jumlah kedatangan) pada data customer dengan gender perempuan

```
## Menghitung nilai outliers dari variabel visitNo (jumlah kedatangan) pada data customer dengan gender perempuan
male = data[data["gender"] == 1]
```

Cell 24 of 56 Go Live 1m Flow Prettier

30°C Partly sunny

1:05 PM 3/30/2023



## INSTRUKSI 9

Kesimpulan:

Berdasarkan hasil perhitungan diketahui bahwa jumlah customer pria yang merupakan outliers dari feature timeSpend hanya 1 orang dimana batas bawah outliers dan batas atas outliers adalah -1,5 dan 2,5.

### Instruksi 9

Hitung nilai outliers dari variabel visitNo (jumlah kedatangan) pada data customer dengan gender perempuan

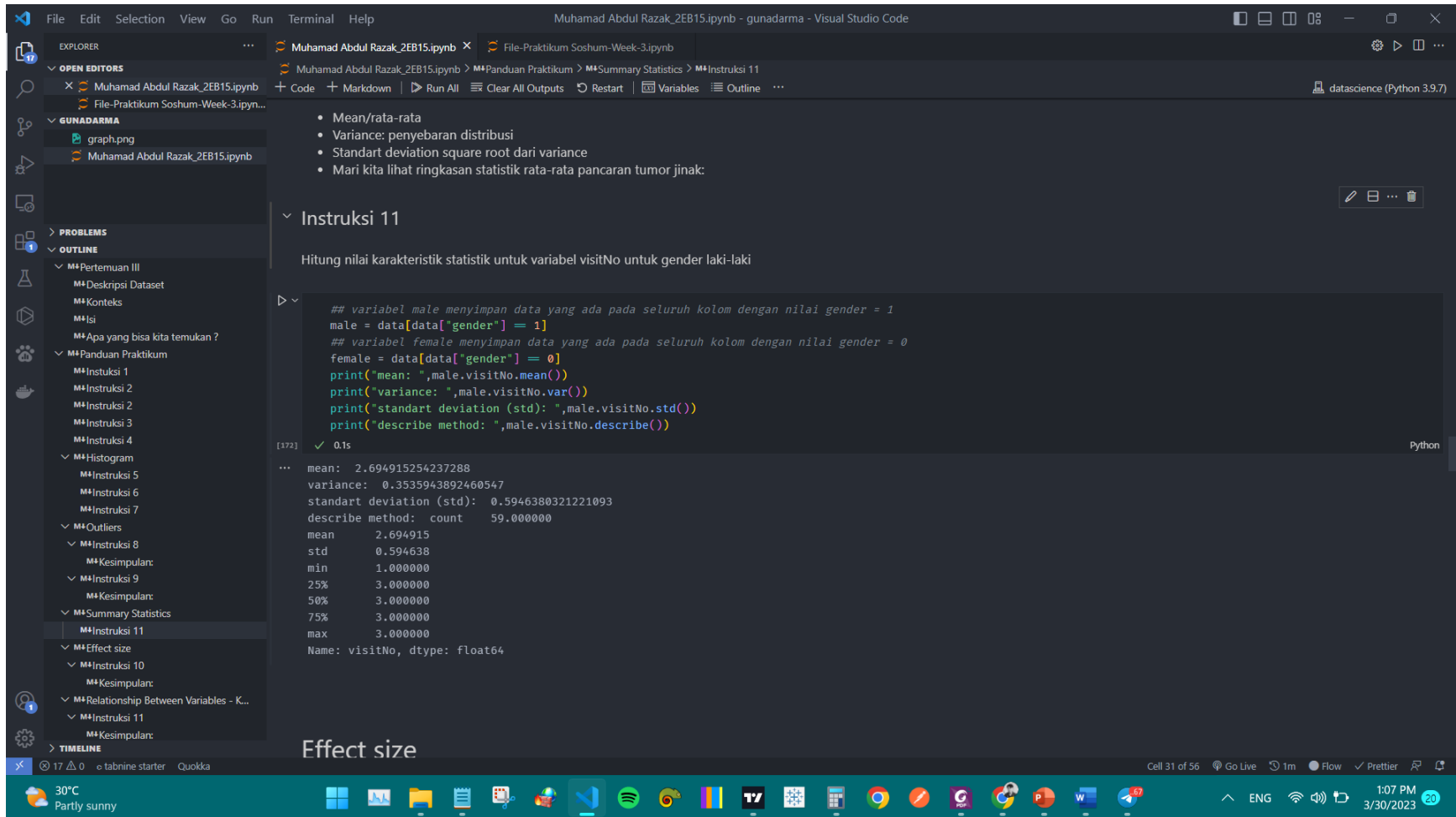
```
## Menghitung nilai outliers dari variabel visitNo (jumlah kedatangan) pada data customer dengan gender perempuan
male = data[data["gender"] == 1]
female = data[data["gender"] == 0]
desc = data.visitNo.describe()
Q1 = desc[4]
Q3 = desc[6]
IQR = Q3 - Q1
lower_bound = Q1 - 1.5 * IQR
upper_bound = Q3 + 1.5 * IQR
print("Apa pun di luar kisaran ini adalah outlier visitNo: (", lower_bound, ",", upper_bound, ")")
data[data.visitNo < lower_bound].visitNo
print("Outliers: ", data[(data.visitNo < lower_bound) | (data.visitNo > upper_bound)].visitNo.values)
```

... Apa pun di luar kisaran ini adalah outlier visitNo: ( 0.5 , 4.5 )  
Outliers: [0 0]

Kesimpulan:

Berdasarkan hasil perhitungan diketahui bahwa jumlah customer perempuan yang merupakan outliers dari feature visitNo (jumlah kedatangan) terdapat 2 orang dimana batas bawah outliers dan batas atas outliers adalah 0,5 dan 4,5.

## INSTRUKSI 11A



The screenshot displays a Visual Studio Code window with a Jupyter Notebook titled 'Muhamad Abdul Razak\_2EB15.ipynb'. The notebook is open to 'Instruksi 11', which contains the following Python code:

```
## variabel male menyimpan data yang ada pada seluruh kolom dengan nilai gender = 1
male = data[data["gender"] == 1]
## variabel female menyimpan data yang ada pada seluruh kolom dengan nilai gender = 0
female = data[data["gender"] == 0]
print("mean: ", male.visitNo.mean())
print("variance: ", male.visitNo.var())
print("standart deviation (std): ", male.visitNo.std())
print("describe method: ", male.visitNo.describe())
```

The output of the code is as follows:

```
[172] ✓ 0.1s
... mean: 2.694915254237288
variance: 0.3535943892460547
standart deviation (std): 0.5946380321221093
describe method: count    59.000000
mean      2.694915
std       0.594638
min       1.000000
25%       3.000000
50%       3.000000
75%       3.000000
max       3.000000
Name: visitNo, dtype: float64
```

The left sidebar shows the Explorer view with the file 'Muhamad Abdul Razak\_2EB15.ipynb' selected. The Outline view shows the notebook's structure, including sections like 'Pertemuan III', 'Deskripsi Dataset', 'Konteks', 'Isi', 'Apa yang bisa kita temukan?', 'Panduan Praktikum', 'Instruksi 1' through 'Instruksi 11', 'Effect size', 'Relationship Between Variables - K...', and 'Kesimpulan:'. The bottom status bar shows the temperature as 30°C, partly sunny, and the time as 1:07 PM on 3/30/2023.

## INSTRUKSI 10

File Edit Selection View Go Run Terminal Help

Muhamad Abdul Razak\_2EB15.ipynb - gunadarma - Visual Studio Code

EXPLORER

OPEN EDITORS

Muhamad Abdul Razak\_2EB15.ipynb

File-Praktikum Soshum-Week-3.ipynb...

GUNADARMA

graph.png

Muhamad Abdul Razak\_2EB15.ipynb

PROBLEMS

OUTLINE

M1Pertemuan III

M1Deskripsi Dataset

M1Konteks

M1Isi

M1Apa yang bisa kita temukan?

M1Panduan Praktikum

M1Instruksi 1

M1Instruksi 2

M1Instruksi 2

M1Instruksi 3

M1Instruksi 4

M1Histogram

M1Instruksi 5

M1Instruksi 6

M1Instruksi 7

M1Outliers

M1Instruksi 8

M1Kesimpulan:

M1Instruksi 9

M1Kesimpulan:

M1Summary Statistics

M1Instruksi 11

M1Effect size

M1Instruksi 10

M1Kesimpulan:

M1Relationship Between Variables - K...

M1Instruksi 11

M1Kesimpulan:

TIMELINE

17:00 tabnine starter Quokka

30°C Partly sunny

Cell 34 of 56 Go Live 1m Flow Prettier

1:07 PM 3/30/2023

Muhamad Abdul Razak\_2EB15.ipynb

File-Praktikum Soshum-Week-3.ipynb

Muhamad Abdul Razak\_2EB15.ipynb

M1Panduan Praktikum

M1Effect size

M1Instruksi 10

Ini menggambarkan ukuran efek/effect size. Ini adalah cara sederhana untuk mengukur perbedaan antara dua kelompok.

- Dengan kata lain, ukuran efek menekankan ukuran perbedaan
- Gunakan ukuran efek kohen/cohen
- Cohen menyarankan bahwa jika  $d(\text{ukuran efek}) = 0,2$ , itu adalah ukuran efek kecil,  $d = 0,5$  ukuran efek sedang,  $d = 0,8$  ukuran efek besar.
- mari kita bandingkan ukuran efek antara mean radius jinak dan mean radius ganas
- Ukuran efek adalah 2.2 dan itu terlalu besar dan dikatakan bahwa dua kelompok itu berbeda satu sama lain seperti yang kita harapkan. Karena kelompok kita adalah mean radius jinak dan mean radius ganas yang berbeda satu sama lain

Instruksi 10

Jalankan perhitungan dibawah ini. Buatlah kesimpulan dari informasi terdapat perbedaan antara jumlah kedatangan dari pelanggan

```
(method) def mean(  
    axis: AxisIndex | None = ...,  
    skipna: _bool = ...,  
    level: None = ...,  
    numeric_only: _bool = ...,  
    **kwargs: Any  
) -> float
```

```
mean_diff = male.visitNo.mean() - female.visitNo.mean()  
var_male = male.visitNo.var()  
var_female = female.visitNo.var()  
var_pooled = (len(male)*var_male + len(female)*var_female) / float(len(male) + len(female))  
effect_size = mean_diff/np.sqrt(var_pooled)  
print("Effect size: ", effect_size)
```

[173] ✓ 0.2s

Effect size: 0.40638112114512326

Kesimpulan:

Berdasarkan hasil perhitungan diketahui bahwa ukuran effect adalah 0.4. Menurut efek cohen, dikatakan bahwa dua kelompok itu yaitu jumlah kedatangan customer pria dan jumlah kedatangan customer wanita adalah KECIL. Oleh karena itu, dua kelompok tersebut memiliki kemiripan satu sama lain.

Relationship Between Variables - Korelasi

- Kita dapat mengatakan bahwa dua variabel terkait satu sama lain, jika salah satunya memberikan informasi tentang yang lain
- Misalnya, harga dan jarak. Jika Anda pergi jarak jauh dengan taksi Anda akan membayar lebih. Oleh karena itu kita dapat mengatakan bahwa harga dan jarak berhubungan positif satu sama

## INSTRUKSI 11B

File Edit Selection View Go Run Terminal Help • Muhamad Abdul Razak\_2EB15.ipynb - gunadarma - Visual Studio Code

EXPLORER

OPEN EDITORS 1 unsaved

Muhamad Abdul Razak\_2EB15.ipynb

File-Praktikum Soshum-Week-3.ipynb...

GUNADARMA

graph.png

Muhamad Abdul Razak\_2EB15.ipynb

PROBLEMS

OUTLINE

MuPertemuan III

MuDeskripsi Dataset

MuKonteks

MuIsi

MuApa yang bisa kita temukan?

MuPanduan Praktikum

MuInstruksi 1

MuInstruksi 2

MuInstruksi 2

MuInstruksi 3

MuInstruksi 4

MuHistogram

MuInstruksi 5

MuInstruksi 6

MuInstruksi 7

MuOutliers

MuInstruksi 8

MuKesimpulan:

MuInstruksi 9

MuKesimpulan:

MuSummary Statistics

MuInstruksi 11

MuKesimpulan:

TIMELINE

17 0 tabnine starter Quokka

30°C Partly sunny

Cell 39 of 56 Go Live 13m Flow Prettier

1:14 PM 3/30/2023

Muhamad Abdul Razak\_2EB15.ipynb • File-Praktikum Soshum-Week-3.ipynb

Muhamad Abdul Razak\_2EB15.ipynb > MuPanduan Praktikum > MuRelationship Between Variables - Korelasi > MuInstruksi 11 > f,ax = plt.subplots(figsize = (16, 4))

+ Code + Markdown | ▶ Run All | Clear All Outputs | Restart | Variables | Outline

data science (Python 3.9.7)

tanda korelasi negatif berarti ada adalah korelasi negatif.

### Instruksi 11

Tuliskan kesimpulan dari hasil diagram heatmap korelasi yang ada. Variabel mana saja yang saling berkorelasi? Jelaskan pendapat Anda.

```
f,ax = plt.subplots(figsize = (16, 4))
dfs = data.loc[:,['productRate','priceRate','serviceRate','visitNo','timeSpend','wifiRate','income','promoRate','membershipCard']]
sns.heatmap(dfs.corr(),annot= True,linewidths=0.5,fmt = ".1f",ax=ax)
plt.title('Correlation Map')
plt.savefig('graph.png')
plt.tight_layout()
```

1.1s Python

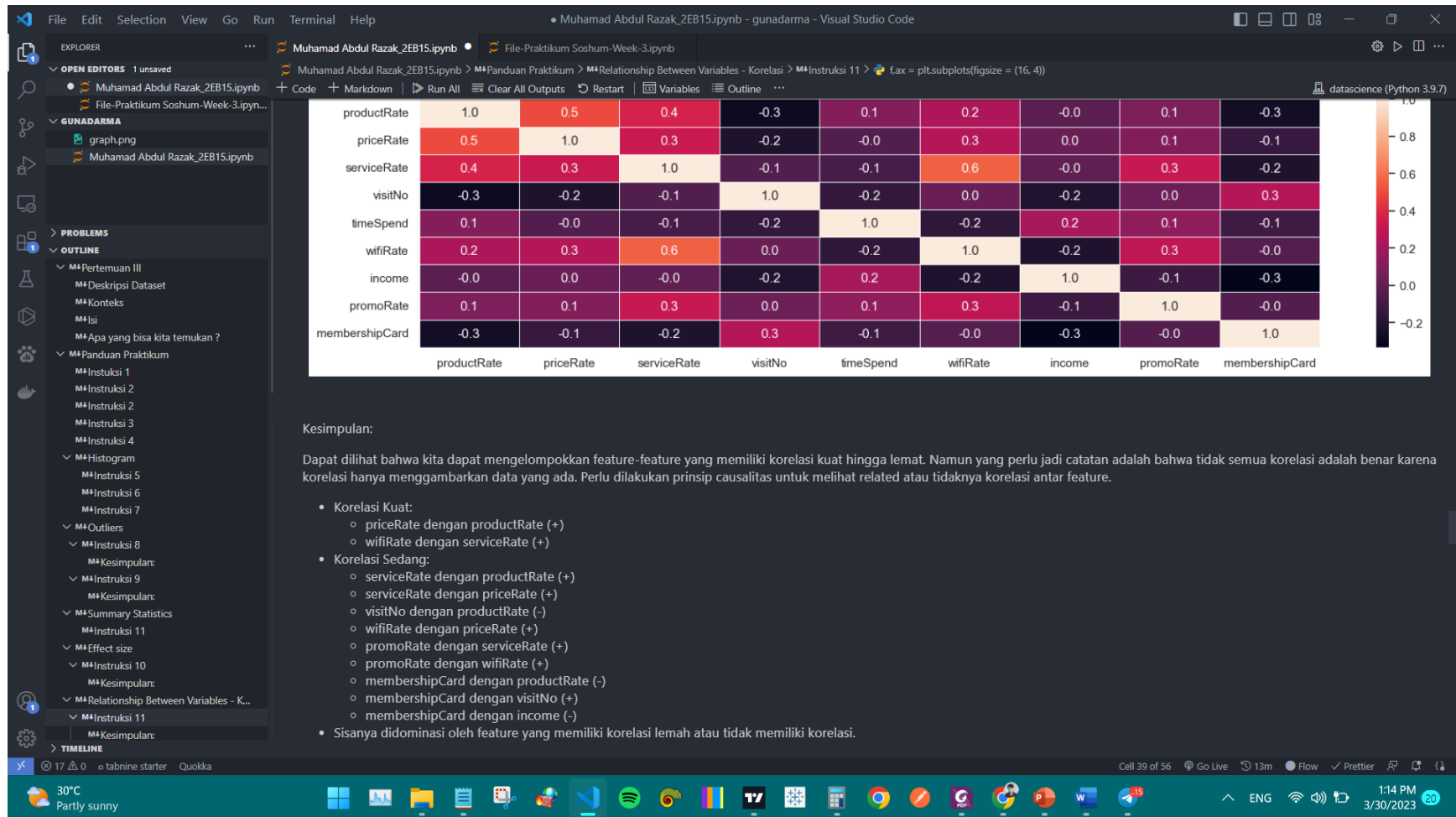
Correlation Map

productRate	1.0	0.5	0.4	-0.3	0.1	0.2	-0.0	0.1	-0.3
priceRate	0.5	1.0	0.3	-0.2	-0.0	0.3	0.0	0.1	-0.1
serviceRate	0.4	0.3	1.0	-0.1	-0.1	0.6	-0.0	0.3	-0.2
visitNo	-0.3	-0.2	-0.1	1.0	-0.2	0.0	-0.2	0.0	0.3
timeSpend	0.1	-0.0	-0.1	-0.2	1.0	-0.2	0.2	0.1	-0.1
wifiRate	0.2	0.3	0.6	0.0	-0.2	1.0	-0.2	0.3	-0.0
income	-0.0	0.0	-0.0	-0.2	0.2	-0.2	1.0	-0.1	-0.3
promoRate	0.1	0.1	0.3	0.0	0.1	0.3	-0.1	1.0	-0.0
membershipCard	-0.3	-0.1	-0.2	0.3	-0.1	-0.0	-0.3	-0.0	1.0

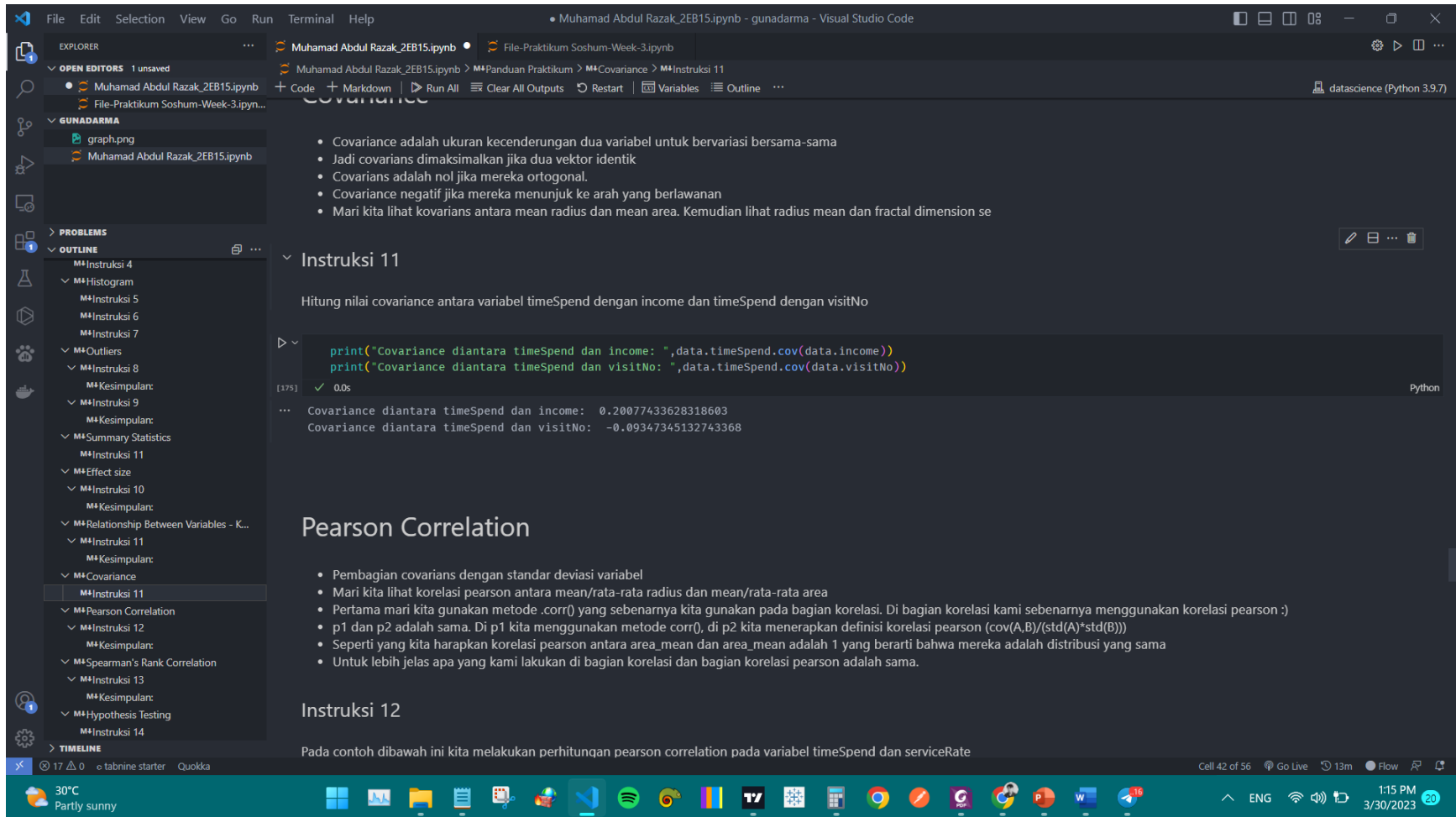
productRate priceRate serviceRate visitNo timeSpend wifiRate income promoRate membershipCard

Kesimpulan:

Dapat dilihat bahwa kita dapat mengelompokkan feature-feature yang memiliki korelasi kuat hingga lemah. Namun yang perlu jadi catatan adalah bahwa tidak semua korelasi adalah benar karena



## INSTRUKSI 11C



File Edit Selection View Go Run Terminal Help • Muhamad Abdul Razak\_2EB15.ipynb - gunadarma - Visual Studio Code

EXPLORER

OPEN EDITORS 1 unsaved

- Muhamad Abdul Razak\_2EB15.ipynb
- File-Praktikum Soshum-Week-3.ipynb...

GUNADARMA

- graph.png
- Muhamad Abdul Razak\_2EB15.ipynb

PROBLEMS

OUTLINE

- M4Instruksi 4
- M4Histogram
- M4Instruksi 5
- M4Instruksi 6
- M4Instruksi 7
- M4Outliers
- M4Instruksi 8
- M4Kesimpulan:
- M4Instruksi 9
- M4Kesimpulan:
- M4Summary Statistics
- M4Instruksi 11
- M4Effect size
- M4Instruksi 10
- M4Kesimpulan:
- M4Relationship Between Variables - K...
- M4Instruksi 11
- M4Kesimpulan:
- M4Covariance
- M4Instruksi 11
- M4Pearson Correlation
- M4Instruksi 12
- M4Kesimpulan:
- M4Spearman's Rank Correlation
- M4Instruksi 13
- M4Kesimpulan:
- M4Hypothesis Testing
- M4Instruksi 14

TIMELINE

Cell 42 of 56 Go Live 13m Flow

30°C Partly sunny

1:15 PM 3/30/2023

Muhamad Abdul Razak\_2EB15.ipynb • File-Praktikum Soshum-Week-3.ipynb

Muhamad Abdul Razak\_2EB15.ipynb > M4Panduan Praktikum > M4Covariance > M4Instruksi 11

+ Code + Markdown ▶ Run All Clear All Outputs Restart Variables Outline ...

datascience (Python 3.9.7)

Covariance

- Covariance adalah ukuran kecenderungan dua variabel untuk bervariasi bersama-sama
- Jadi covarians dimaksimalkan jika dua vektor identik
- Covarians adalah nol jika mereka ortogonal.
- Covariance negatif jika mereka menunjuk ke arah yang berlawanan
- Mari kita lihat kovarians antara mean radius dan mean area. Kemudian lihat radius mean dan fractal dimension se

Instruksi 11

Hitung nilai covariance antara variabel timeSpend dengan income dan timeSpend dengan visitNo

```
print("Covariance diantara timeSpend dan income: ",data.timeSpend.cov(data.income))
print("Covariance diantara timeSpend dan visitNo: ",data.timeSpend.cov(data.visitNo))
```

[179] ✓ 0.0s Python

... Covariance diantara timeSpend dan income: 0.20077433628318603  
Covariance diantara timeSpend dan visitNo: -0.09347345132743368

Pearson Correlation

- Pembagian covarians dengan standar deviasi variabel
- Mari kita lihat korelasi pearson antara mean/rata-rata radius dan mean/rata-rata area
- Pertama mari kita gunakan metode .corr() yang sebenarnya kita gunakan pada bagian korelasi. Di bagian korelasi kami sebenarnya menggunakan korelasi pearson :)
- p1 dan p2 adalah sama. Di p1 kita menggunakan metode corr(), di p2 kita menerapkan definisi korelasi pearson ( $\text{cov}(A,B)/(\text{std}(A)*\text{std}(B))$ )
- Seperti yang kita harapkan korelasi pearson antara area\_mean dan area\_mean adalah 1 yang berarti bahwa mereka adalah distribusi yang sama
- Untuk lebih jelas apa yang kami lakukan di bagian korelasi dan bagian korelasi pearson adalah sama.

Instruksi 12

Pada contoh dibawah ini kita melakukan perhitungan pearson correlation pada variabel timeSpend dan serviceRate

## INSTRUKSI 12

File Edit Selection View Go Run Terminal Help

Muhamad Abdul Razak\_2EB15.ipynb - gunadarma - Visual Studio Code

EXPLORED Muhamad Abdul Razak\_2EB15.ipynb File-Praktikum Soshum-Week-3.ipynb

OPEN EDITORS Muhamad Abdul Razak\_2EB15.ipynb File-Praktikum Soshum-Week-3.ipynb

GUNADARMA graph.png Muhamad Abdul Razak\_2EB15.ipynb

PROBLEMS

OUTLINE M4Instruksi 4 M4Histogram M4Instruksi 5 M4Instruksi 6 M4Instruksi 7 M4Outliers M4Instruksi 8 M4Kesimpulan: M4Instruksi 9 M4Kesimpulan: M4Summary Statistics M4Instruksi 11 M4Effect size M4Instruksi 10 M4Kesimpulan: M4Relationship Between Variables - K... M4Instruksi 11 M4Kesimpulan: M4Covariance M4Instruksi 11 M4Pearson Correlation M4Instruksi 12 M4Kesimpulan: M4Spearman's Rank Correlation M4Instruksi 13 M4Kesimpulan: M4Hypothesis Testing M4Instruksi 14

TIMELINE

Muhamad Abdul Razak\_2EB15.ipynb File-Praktikum Soshum-Week-3.ipynb

Muhamad Abdul Razak\_2EB15.ipynb > M4Panduan Praktikum > M4Pearson Correlation > M4Instruksi 12 > p1 = data.loc[:,["wifiRate","serviceRate"]].corr(method="pearson")

+ Code + Markdown Run All Clear All Outputs Restart Variables Outline

data science (Python 3.9.7)

## Instruksi 12

Pada contoh dibawah ini kita melakukan perhitungan pearson correlation pada variabel timeSpend dan serviceRate

- Sesuaikan perintah untuk melakukan perhitungan pearson correlation pada variabel wifiRate dan serviceRate.
- Kesimpulan apa yang dapat Anda tarik dari dua percobaan diatas.

```
p1 = data.loc[:,["wifiRate","serviceRate"]].corr(method="pearson")
p2 = data.serviceRate.cov(data.wifiRate)/(data.serviceRate.std()*data.wifiRate.std())
print('Pearson correlation: ')
print(p1)
print('Pearson correlation: ',p2)
f, ax = plt.subplots(figsize=(4,2))
sns.heatmap(p1, cmap='coolwarm_r', linewidths=0.5, annot=True)
plt.xticks(rotation=0)
plt.tight_layout()
```

[212] ✓ 0.4s Python

... Pearson correlation:

	wifiRate	serviceRate
wifiRate	1.000000	0.555095
serviceRate	0.555095	1.000000

Pearson correlation: 0.5550950620398712

Kesimpulan:

Feature wifiRate dengan serviceRate berkorelasi kuat positif dengan nilai 0.56. Artinya semakin tinggi wifiRate maka semakin tinggi juga serviceRate

Cell 46 of 56 Go Live 16m Flow Prettier

30°C Partly sunny 1:22 PM 3/30/2023

## INSTRUKSI 13

Visual Studio Code interface showing a Jupyter Notebook titled "Instruksi 13". The notebook content includes a text block, a code cell, and a text block.

**Text Block:**

Pada contoh dibawah ini kita melakukan perhitungan spearman correlation pada variabel timeSpend dan serviceRate

1. Sesuaikan perintah untuk melakukan perhitungan pearson correlation pada variabel priceRate dan productRate.
2. Kesimpulan apa yang dapat Anda tarik dari dua percobaan diatas.

**Code Cell:**

```
ranked_data = data.rank()
spearman_corr = ranked_data.loc[:,["priceRate","productRate"]].corr(method= "pearson")
print("Spearman's correlation: ")
print(spearman_corr)
f, ax = plt.subplots(figsize=(4,2))
sns.heatmap(spearman_corr, cmap='coolwarm_r', linewidths=0.5, annot=True)
plt.xticks(rotation=0)
plt.tight_layout()
```

**Output:**

Spearman's correlation:

	priceRate	productRate
priceRate	1.000000	0.452944
productRate	0.452944	1.000000

**Figure:**

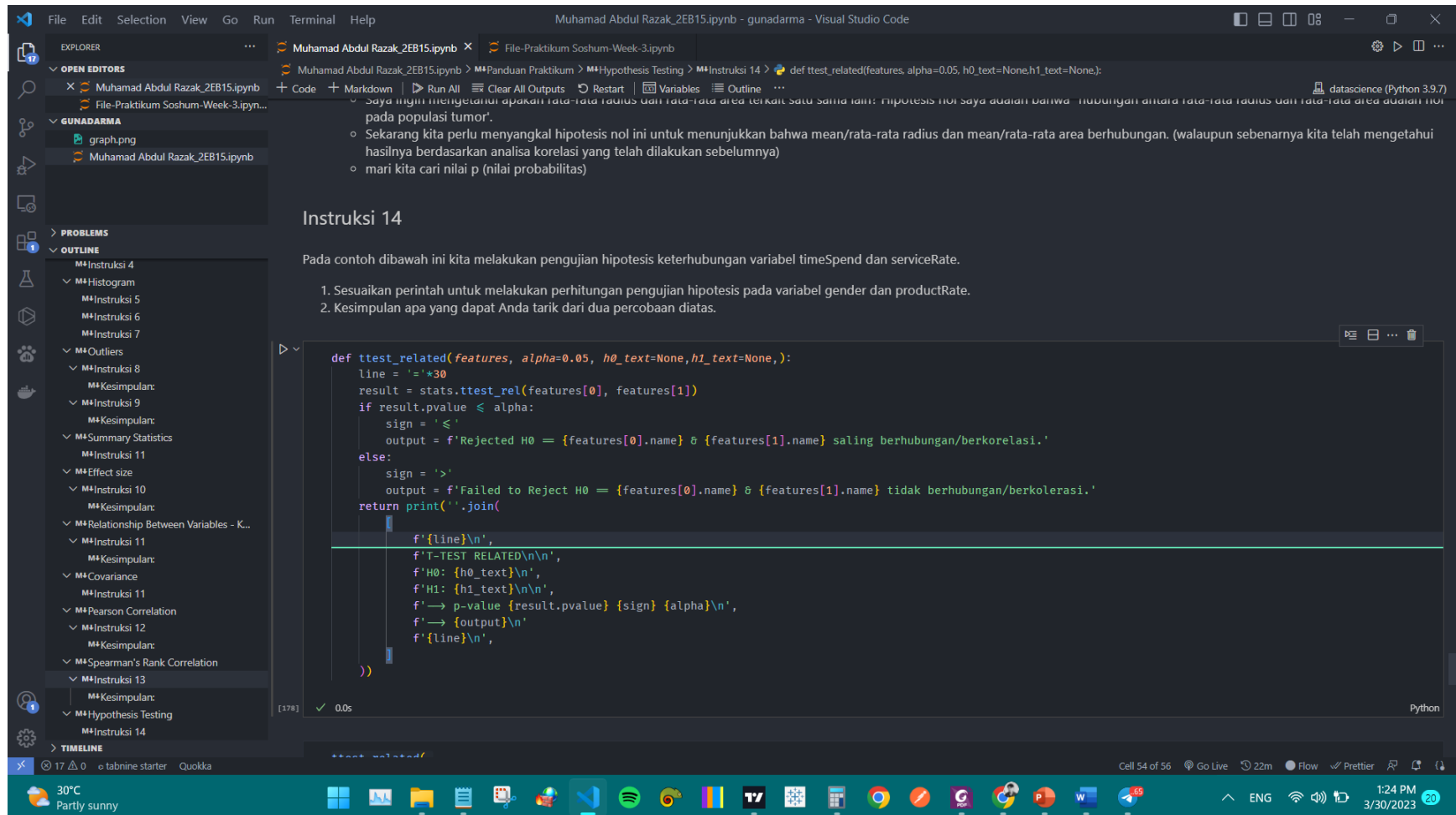
The heatmap displays the Spearman correlation matrix for the variables priceRate and productRate. The color scale ranges from 0.6 (blue) to 1.0 (red). The diagonal cells (priceRate vs priceRate and productRate vs productRate) are blue, indicating a correlation of 1.0. The off-diagonal cells (priceRate vs productRate and productRate vs priceRate) are red, indicating a correlation of 0.45.

**Kesimpulan:**

- Feature priceRate dengan productRate berkorelasi kuat positif dengan nilai 0.56. Artinya semakin tinggi priceRate maka semakin tinggi juga productRate.
- Korelasi Spearman sedikit lebih tinggi dari korelasi pearson



## INSTRUKSI 14



Visual Studio Code interface showing a Jupyter Notebook file named `Muhamad Abdul Razak_2EB15.ipynb` in the Explorer panel. The file is open in the Editor panel, displaying a Python cell with a `ttest_related` function call. The function call is as follows:

```
ttest_related(
    features=[data.gender, data.productRate],
    h0_text='Gender dan Produt Rate tidak memiliki hubungan.',
    h1_text='Gender dan Produt Rate memiliki hubungan.')
```

The output of the cell shows the results of the t-test:

```
T-TEST RELATED

H0: Gender dan Produt Rate tidak memiliki hubungan.
H1: Gender dan Produt Rate memiliki hubungan.

→ p-value 9.83197741893934e-60 ≤ 0.05
→ Rejected H0 = gender & productRate saling berhubungan/berkorelasi.
```

The Explorer panel also shows a list of files and folders, including `graph.png` and `Muhamad Abdul Razak_2EB15.ipynb`. The Timeline panel at the bottom shows the file's history, including `Muhamad Abdul Razak_2EB15.ipynb` and `Muhamad Abdul Razak_2EB15.ipynb`.

The status bar at the bottom indicates the current file is `Cell 56 of 56`, the language is `Go Live`, and the file is `Flow`. The system tray shows the temperature is `26°C` with `Rain showers`, and the date is `3/30/2023` at `1:25 PM`.