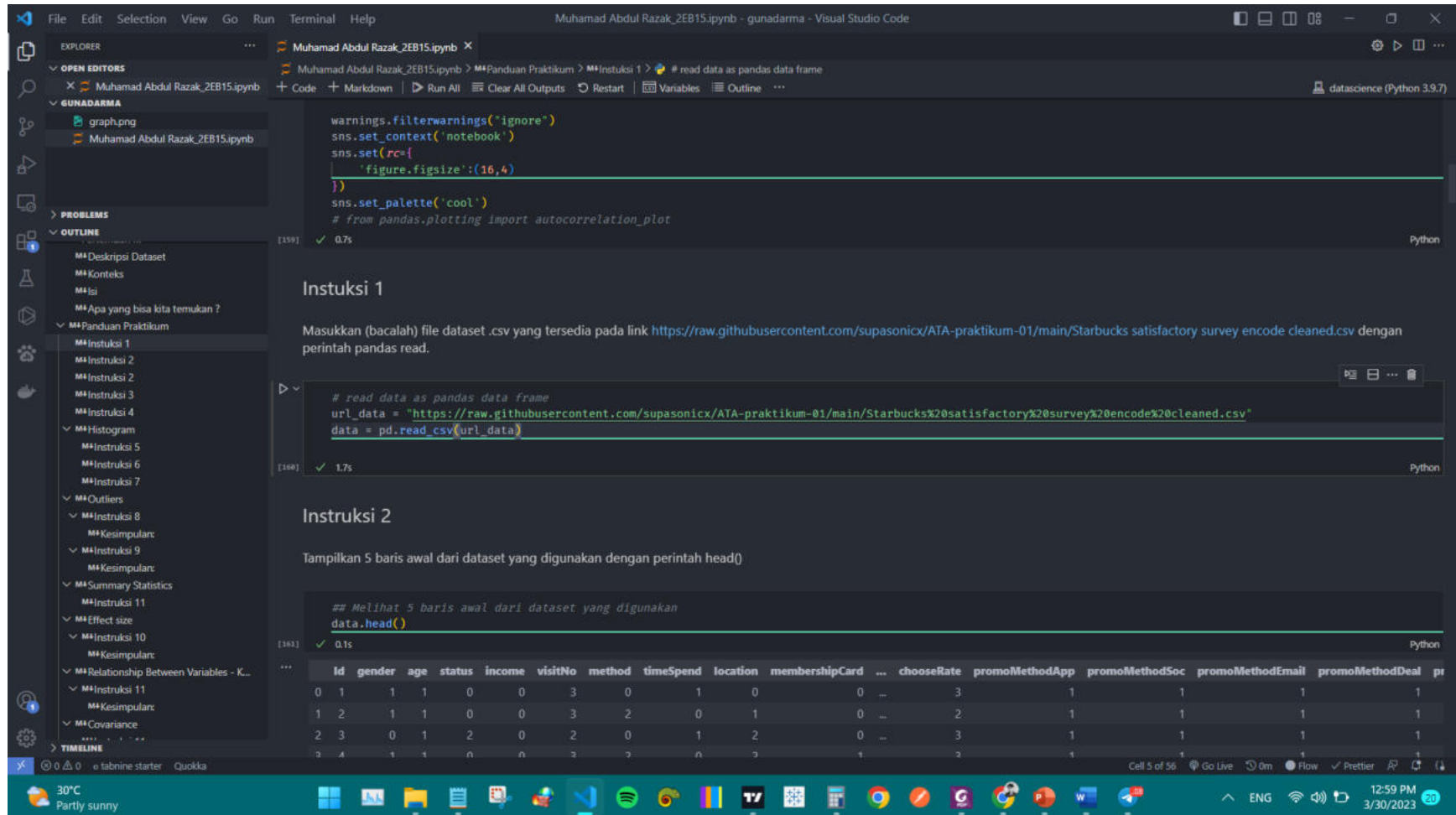


INSTRUKSI 1



The screenshot shows a Visual Studio Code editor with a Jupyter Notebook open. The notebook has two cells, each containing a Python instruction. The first cell, labeled 'Instruksi 1', contains code to read a CSV file from a GitHub repository. The second cell, labeled 'Instruksi 2', 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.

Instruksi 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)
```

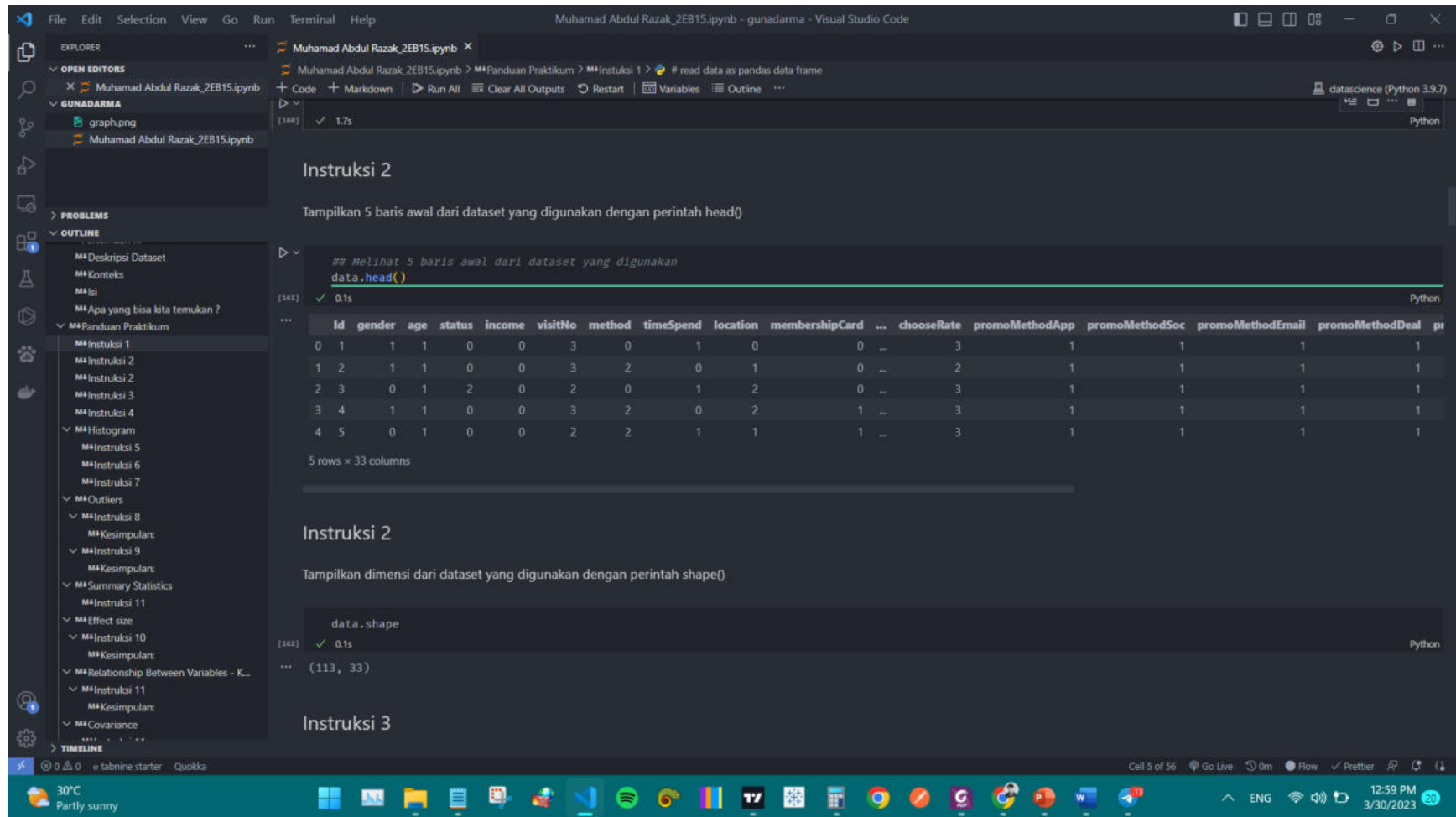
Instruksi 2

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	pr
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	1	0	...	2	1	1	1	1	

INSTRUKSI 2



The screenshot shows a Visual Studio Code window with a Jupyter Notebook open. The notebook has two cells, both of which have been executed successfully. The first cell contains a comment in Indonesian and a code snippet to display the first 5 rows of a dataset using the `head()` function. The output is a table with 5 rows and 33 columns. The second cell contains a code snippet to display the dimensions of the dataset using the `shape()` function. The output is a tuple `(113, 33)`.

Instruksi 2

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

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

5 rows x 33 columns

	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

Instruksi 2

Tampilkan dimensi dari dataset yang digunakan dengan perintah `shape()`

```
data.shape
```

`(113, 33)`

Instruksi 3

INSTRUKSI 3

The screenshot shows a Visual Studio Code window with a Jupyter Notebook titled 'Muhamad Abdul Razak_2EB15.ipynb'. The left sidebar contains an Explorer panel with a file tree and an Outline panel with a list of instructions. The main editor area displays two code cells.

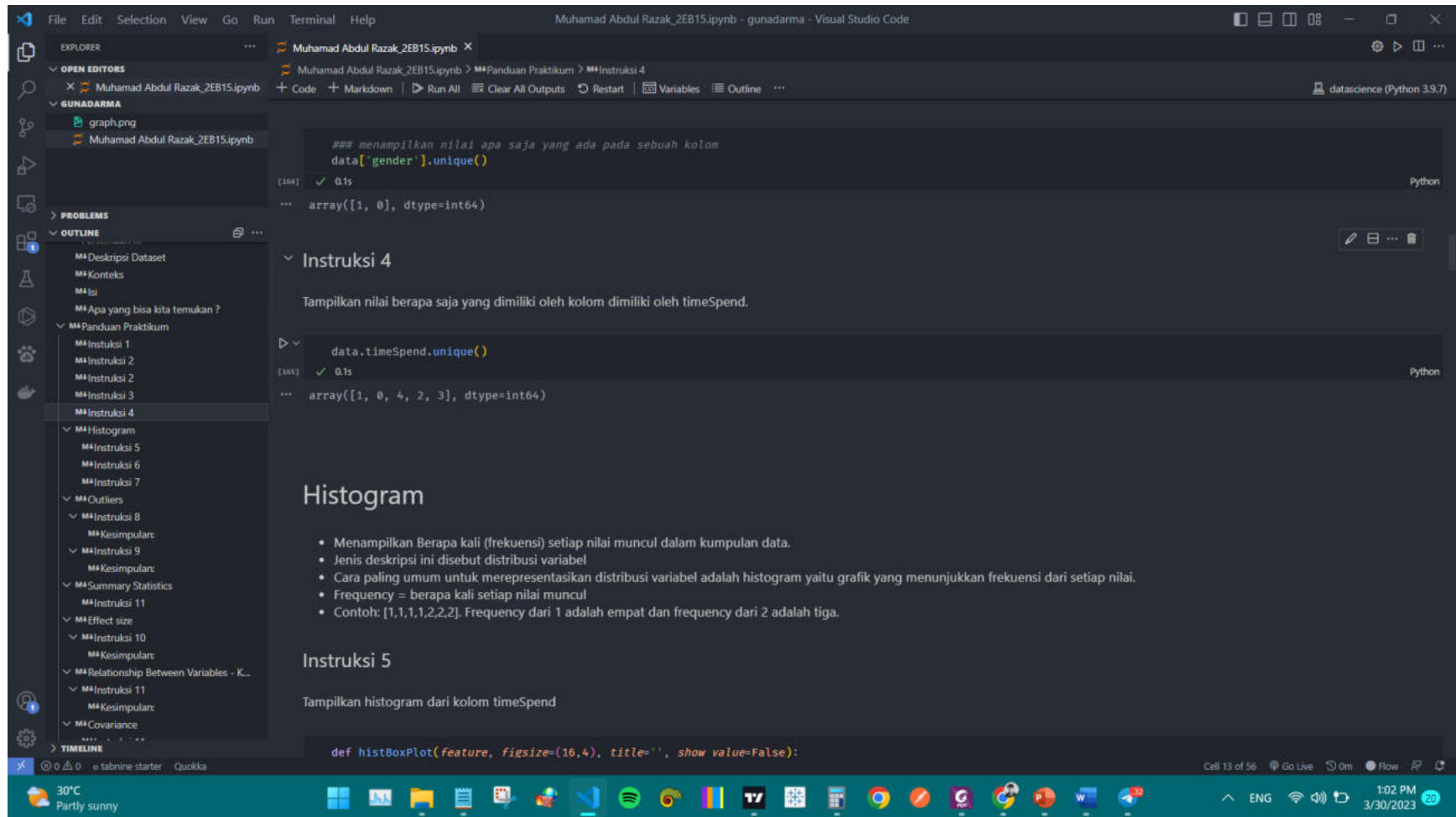
Cell 1: The instruction is 'Tampilkan dimensi dari dataset yang digunakan dengan perintah shape()'. The code is `data.shape`, and the output is `(113, 33)`.

Cell 2: The instruction is 'Tampilkan kolom apa saja yang dimiliki oleh dataset yang digunakan dengan perintah columns()'. The code is `data.columns`, and the output is a list of 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')`.

Cell 3: The instruction is 'Tampilkan nilai apa saja yang dimiliki oleh dataset yang digunakan dengan perintah unique()'. The code is `data['gender'].unique()`, and the output is `array([1, 0], dtype=int64)`.

The bottom status bar shows 'Cell 11 of 56', 'Go Live', '0m', 'Flow', 'Prettier', and the system clock '1:01 PM 3/30/2023'.

INSTRUKSI 4



Visual Studio Code interface showing a Jupyter Notebook with two cells. The first cell displays the unique values of the 'gender' column. The second cell displays the unique values of the 'timeSpend' column.

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

[104] ✓ 0.1s
... array([1, 0], dtype=int64)
```

Instruksi 4

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

```
data.timeSpend.unique()

[105] ✓ 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):
```

Visual Studio Code interface showing a Jupyter Notebook with two cells. The first cell displays the unique values of the 'gender' column. The second cell displays the unique values of the 'timeSpend' column.

INSTRUKSI 5

File Edit Selection View Go Run Terminal Help

Muhamad Abdul Razak_2EB15.ipynb - gunadarma - Visual Studio Code

EXPLORER

Muhamad Abdul Razak_2EB15.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

MAI

MA Apa yang bisa kita temukan ?

MA Panduan Praktikum

MA Instruksi 1

MA Instruksi 2

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 Kesimpulanc

MA Instruksi 9

MA Kesimpulanc

MA Summary Statistics

MA Instruksi 11

MA Effect size

MA Instruksi 10

MA Kesimpulanc

MA Relationship Between Variables - K...

MA Instruksi 11

MA Kesimpulanc

MA Covariance

MA Instruksi 11

MA Pearson Correlation

MA Instruksi 12

MA Kesimpulanc

MA Spearman's Rank Correlation

MA Instruksi 13

TIMELINE

17 0 0 tabline starter Quokka

30°C Partly sunny

1:04 PM 3/30/2023

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 contains a Python code cell that plots a histogram of timeSpent by gender.

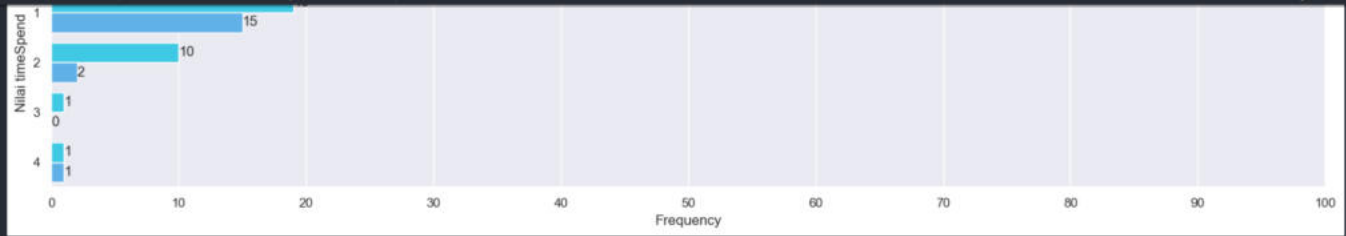
```
## plotting untuk data produk pastry
f, ax = plt.subplots()
sns.countplot(data=data, y='timeSpent', hue='gender').set(xticks=[i for i in np.arange(0,101,10)])
ax.bar_label(ax.containers[0])
ax.bar_label(ax.containers[1])
plt.ylabel("Nilai timeSpent")
plt.xlabel("Frequency")
plt.title("Histogram waktu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

The output shows a horizontal bar chart titled "Histogram waktu yang dihabiskan di toko berdasarkan demografi gender". The y-axis is labeled "Nilai timeSpent" and ranges from 0 to 4. The x-axis is labeled "Frequency" and ranges from 0 to 100. The chart displays the frequency of timeSpent values for Female (light blue) and Male (dark blue) customers.

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

The bottom of the image shows the Windows taskbar with the date and time: 1:04 PM, 3/30/2023.

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 $Q1 - 1.5 \times IQR$ dan $Q3 + 1.5 \times IQR$
 - 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

Muhamad Abdul Razak_2EB15.ipynb - gunadarma - Visual Studio Code

File Edit Selection View Go Run Terminal Help

EXPLORER

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 Isi

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

TIMELINE

17 0 tabnine starter Quokka

Cell 24 of 56 Go Live 1m Flow Prettier

30°C Partly sunny

1:05 PM 3/30/2023

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)
```

[178] ✓ Q1s

... 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]
```


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)
```

[171] ✓ 0.1s

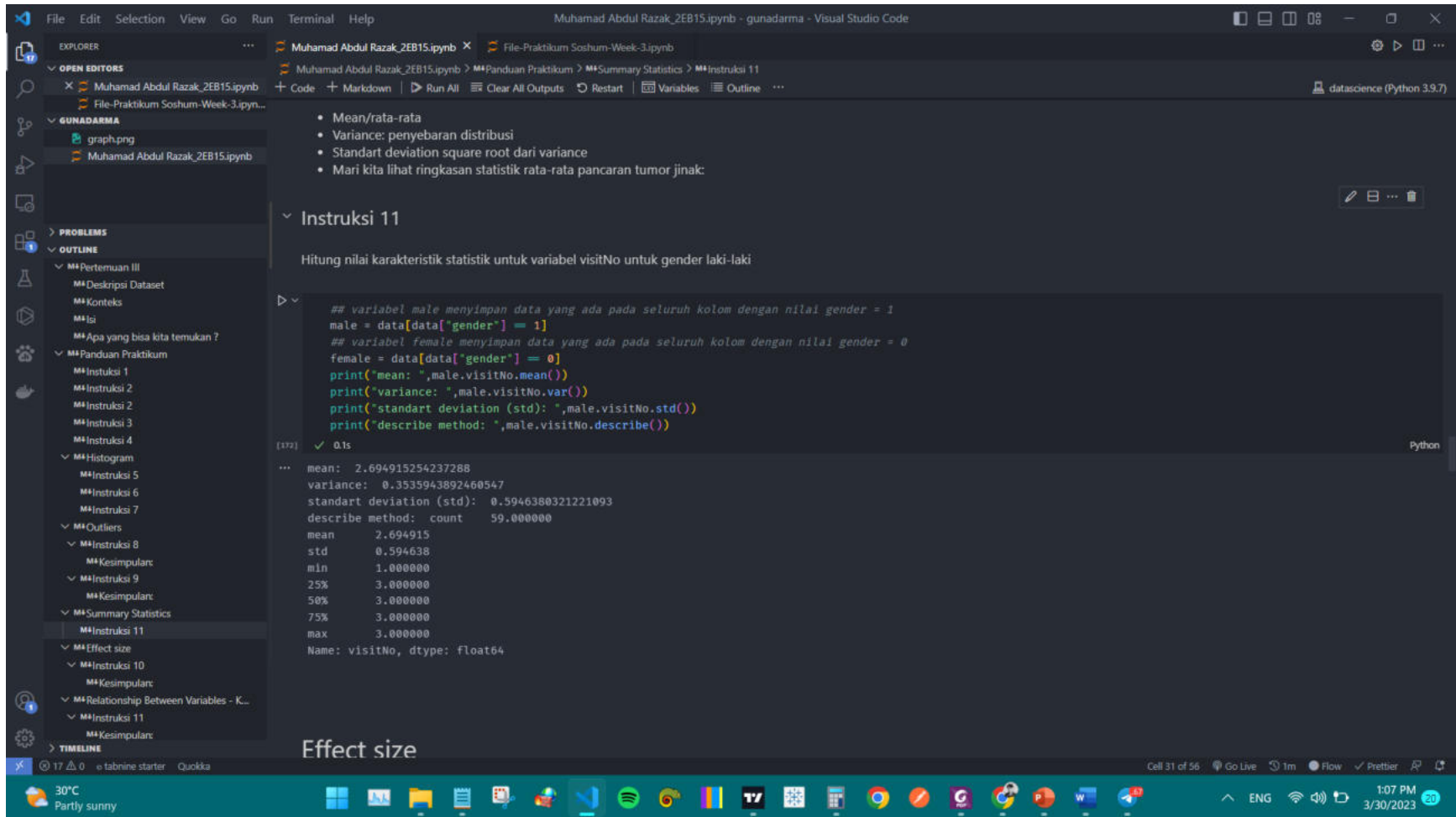
... Apa pun di luar kisaran ini adalah outlier visitNo: (0.5 , 4.5)

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 shows a Visual Studio Code window with a Jupyter Notebook titled 'Muhamad Abdul Razak_2EB15.ipynb'. The notebook is open to 'Instruksi 11'. The code in the cell calculates statistics for the 'visitNo' variable where 'gender' is 1. The output shows the mean, variance, standard deviation, and a describe method result.

```
## 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())
```

Output:

```
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 bottom status bar shows the temperature as 30°C and the date as 3/30/2023.

INSTRUKSI 10

The screenshot shows a Visual Studio Code window with a Jupyter Notebook titled "Muhamad Abdul Razak_2EB15.ipynb". The notebook is open to a cell containing instructions and code for calculating effect size.

Instructions:

- 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

Code:

```
(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)
```

Output:

```
[179] ✓ 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

MA Pertemuan III

MA Deskripsi Dataset

MA Konteks

MA Isi

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:

TIMELINE

17 0 tabnine starter Quokka

30°C Partly sunny

1:14 PM 3/30/2023

Muhamad Abdul Razak_2EB15.ipynb • File-Praktikum Soshum-Week-3.ipynb

Muhamad Abdul Razak_2EB15.ipynb > MA Panduan Praktikum > MA Relationship Between Variables - Korelasi > MA Instruksi 11 > Lax = 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()
```

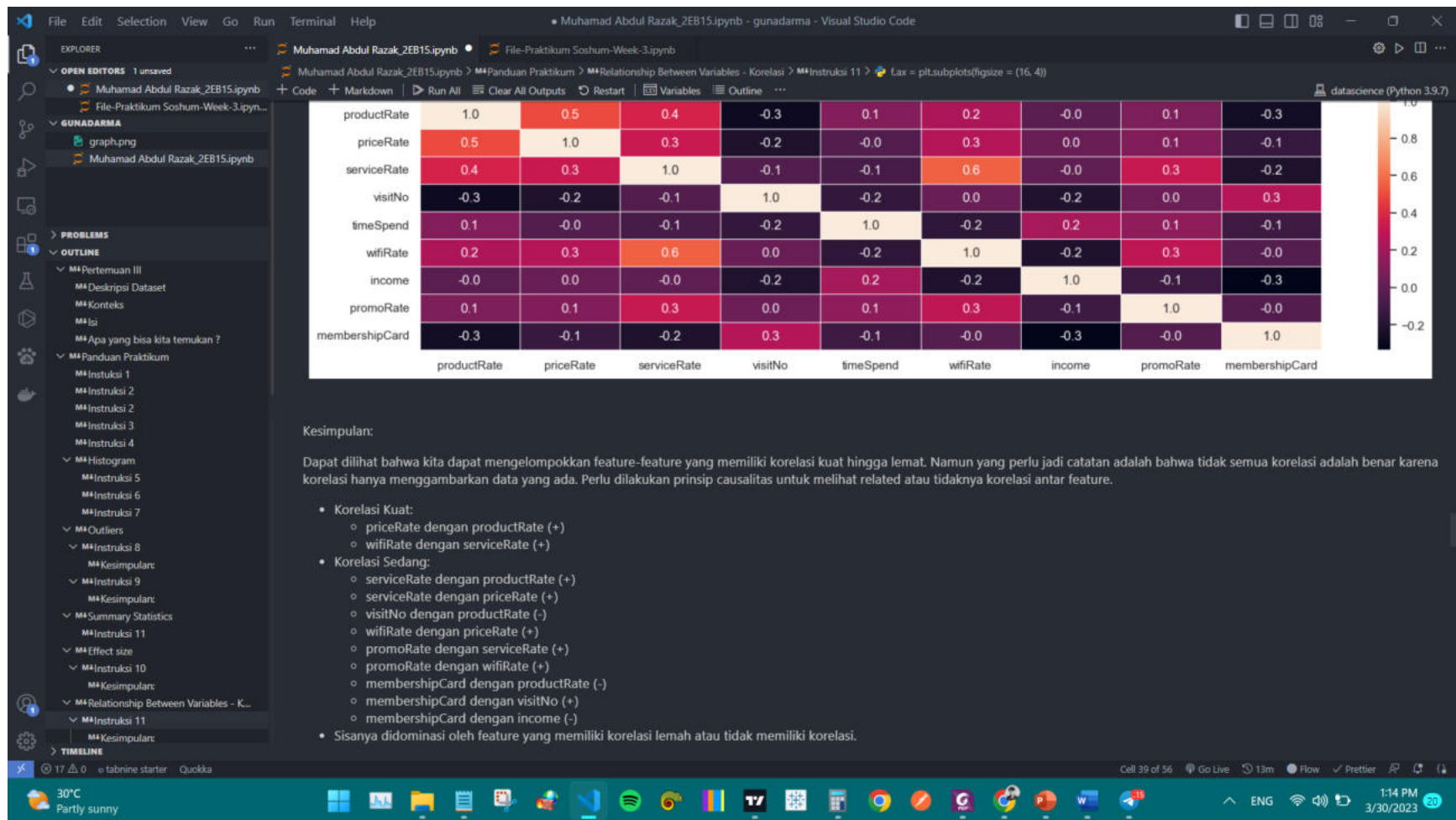
1.1s Python

Correlation Map

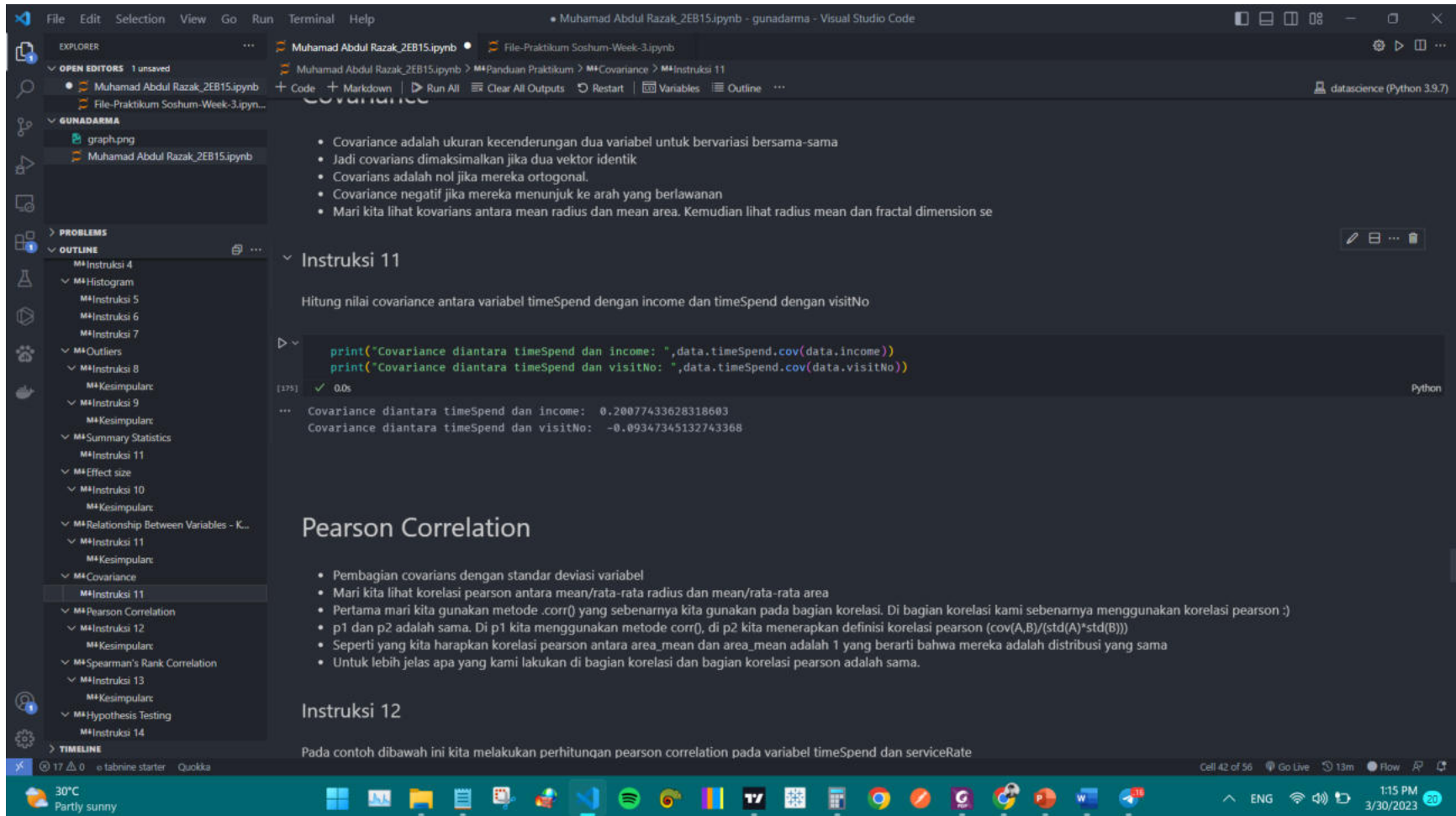
	productRate	priceRate	serviceRate	visitNo	timeSpend	wifiRate	income	promoRate	membershipCard
productRate	1.0	0.5	0.4	-0.3	0.1	0.2	-0.0	0.1	-0.3
priceRate	0.5	1.0	0.3	-0.2	-0.0	0.3	0.0	0.1	-0.1
serviceRate	0.4	0.3	1.0	-0.1	-0.1	0.6	-0.0	0.3	-0.2
visitNo	-0.3	-0.2	-0.1	1.0	-0.2	0.0	-0.2	0.0	0.3
timeSpend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

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

- MAInstruksi 4
- MAHistogram
- MAInstruksi 5
- MAInstruksi 6
- MAInstruksi 7
- MAOutliers
- MAInstruksi 8
- MAKesimpulan:
- MAInstruksi 9
- MAKesimpulan:
- MASummary Statistics
- MAInstruksi 11
- MAEffect size
- MAInstruksi 10
- MAKesimpulan:
- MARelationship Between Variables - K...
- MAInstruksi 11
- MAKesimpulan:
- MACovariance
- MAInstruksi 11
- MAPearson Correlation
- MAInstruksi 12
- MAKesimpulan:
- MASpearman's Rank Correlation
- MAInstruksi 13
- MAKesimpulan:
- MAHypothesis Testing
- MAInstruksi 14

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

Cell 42 of 56 Go Live 13m Flow

30°C Partly sunny

1:15 PM 3/30/2023

INSTRUKSI 12

File Edit Selection View Go Run Terminal Help

Muhamad Abdul Razak_2EB15.ipynb - gunadarma - Visual Studio Code

EXPLORE

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

M4Kesimpulanc

M4Instruksi 9

M4Kesimpulanc

M4Summary Statistics

M4Instruksi 11

M4Effect size

M4Instruksi 10

M4Kesimpulanc

M4Relationship Between Variables - K...

M4Instruksi 11

M4Kesimpulanc

M4Covariance

M4Instruksi 11

M4Pearson Correlation

M4Instruksi 12

M4Kesimpulanc

M4Spearman's Rank Correlation

M4Instruksi 13

M4Kesimpulanc

M4Hypothesis Testing

M4Instruksi 14

TIMELINE

tabnine starter Quokka

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

1. Sesuaikan perintah untuk melakukan perhitungan pearson correlation pada variabel wifiRate dan serviceRate.

2. 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()
```


Python

0.4s

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

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:

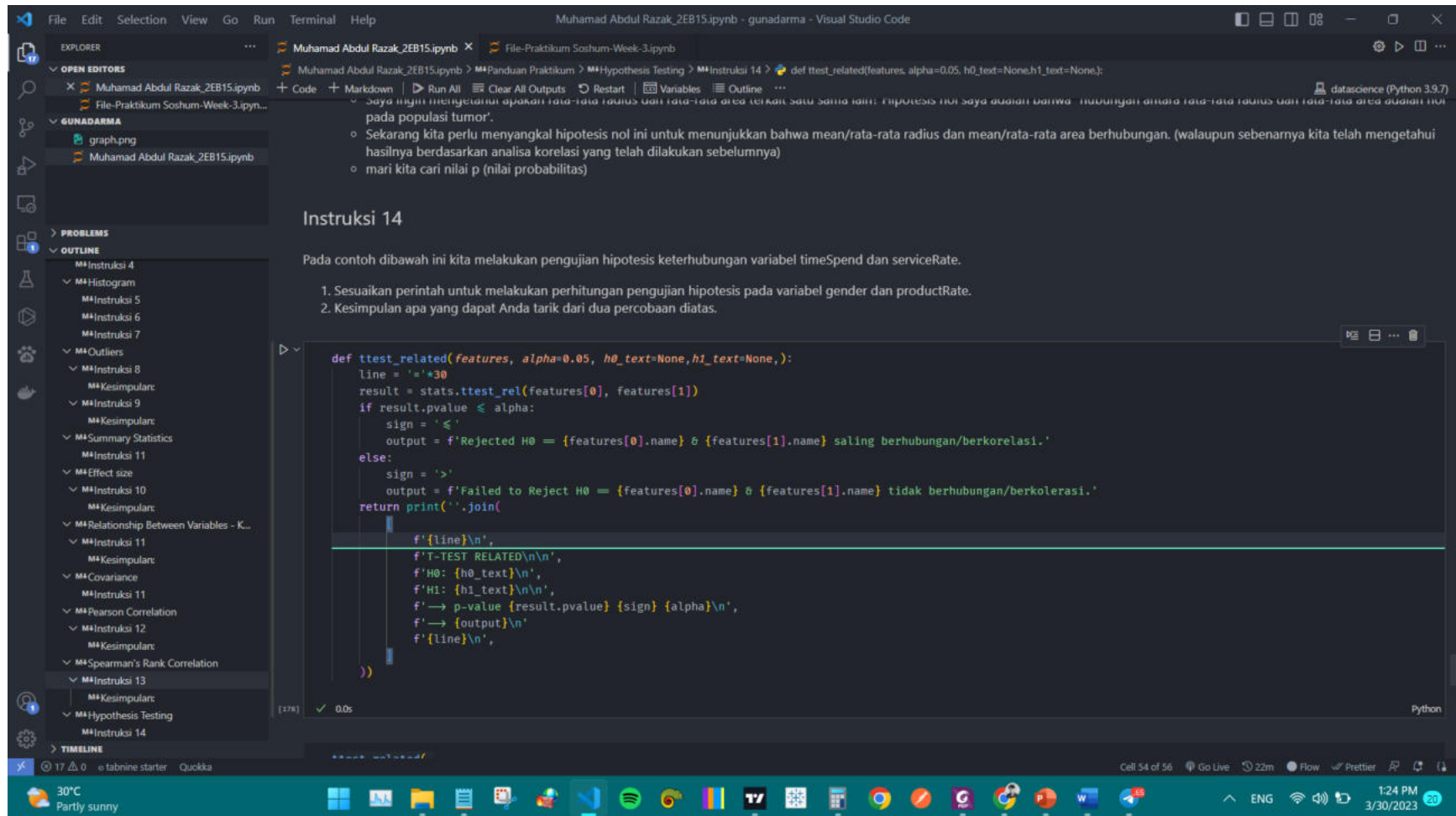
	priceRate	productRate
priceRate	1.000000	0.452944
productRate	0.452944	1.000000

Figure:

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 `gunadarma` workspace.

The Explorer sidebar on the left shows the file structure, including `graph.png` and the notebook file.

The Outline sidebar on the left lists the notebook's sections, including `M4Instruksi 4` through `M4Instruksi 14`, `M4Hypothesis Testing`, and `M4Kesimpulan`.

The main editor displays two code cells from the notebook:

Cell 179:

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

Cell 179 Output:

```
...  
T-TEST RELATED  
  
H0: Gender dan Produkt Rate tidak memiliki hubungan.  
H1: Gender dan Produkt Rate memiliki hubungan.  
  
→ p-value 9.83197741893934e-60 ≤ 0.05  
→ Rejected H0 = gender & productRate saling berhubungan/berkorelasi.  
...
```

Cell 180:

```
ttest_related(  
    features=[data.timeSpend, data.serviceRate],  
    h0_text='Time Spend dan Service Rate tidak memiliki hubungan.',  
    h1_text='Time Spend dan Service Rate memiliki hubungan.')  
)
```

Cell 180 Output:

```
...  
T-TEST RELATED  
  
H0: Time Spend dan Service Rate tidak memiliki hubungan.  
H1: Time Spend dan Service Rate memiliki hubungan.  
  
→ p-value 1.4365649807632511e-52 ≤ 0.05  
→ Rejected H0 = timeSpend & serviceRate saling berhubungan/berkorelasi.  
...
```

The bottom status bar shows the system clock as 1:25 PM on 3/30/2023, with a temperature of 26°C and rain showers.