

kelompok 4

September 25, 2023

1 Kelompok 4

MUHAMAD SABIL FAUSTA

MUHAMMAD ALFI FAIZ

MUHAMMAD RAFIE AL HABSYI SETIAWAN

RIFA SANIA

SETYAWAN HUMAY SENJA

```
[ ]: %matplotlib inline
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

2 Praproses

mengisi data pekerjaan NaN menjadi 0

mengubah spasi menjadi _

mengubah type data jumlah pekerjaan menjadi int

mengubah type kelurahan menjadi category

mengisi data NaN

2.1 Load dataset

```
[ ]: df = pd.read_csv("data-masyarakat-miskin-tahun-2020.csv")
df.head()
```

```
[ ]: Kode Kecamatan Kecamatan Kode Kelurahan Kelurahan RW \
0      32.73.20 Antapani 32.73.20.1001 Antapani Kulon 1
1      32.73.20 Antapani 32.73.20.1001 Antapani Kulon 2
2      32.73.20 Antapani 32.73.20.1001 Antapani Kulon 3
3      32.73.20 Antapani 32.73.20.1001 Antapani Kulon 4
4      32.73.20 Antapani 32.73.20.1001 Antapani Kulon 5
```

	Jumlah Kepala Keluarga	Jumlah Anggota Keluarga
0	98	255
1	7	21
2	26	56
3	6	17
4	161	379

```
[ ]: df1 = pd.read_csv("data-usaha-kecil-dan-menengah-tahun-2020.csv")
df1.head()
```

```
[ ]: Kode Kecamatan Kecamatan Kode Kelurahan Kelurahan Nama Perusahaan \
0      32.73.20 Antapani 32.73.20.1001 Antapani Kulon Abdullah Syuaib
1      32.73.20 Antapani 32.73.20.1006 Antapani Wetan Achmad Sutisna
2      32.73.20 Antapani 32.73.20.1001 Antapani Kulon Ade
3      32.73.20 Antapani 32.73.20.1001 Antapani Kulon Agus
4      32.73.20 Antapani 32.73.20.1006 Antapani Wetan Agus Koswara
```

	Jenis Usaha	Alamat Usaha	Jumlah Pekerja
0	Craft	Jl. Kalijati No. 42 RT 07/04 Antapani Kulon	20.0
1	Kuliner	Jl. Malangbong V No. 38 Antapani Bandung	0.0
2	Kuliner	Jl. Tarumasari RT 04/05 Antapani Kulon	3.0
3	Kuliner	Jl. Sukapura No. 2 RT 01/06 Antapani Kulon	1.0
4	Jasa	Jl. Sindangsari III RT 01/09 Antapani Wetan	2.0

2.2 Mengeksplorasi dataset Jumlah kemiskinan

```
[ ]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 46 entries, 0 to 45
Data columns (total 7 columns):
#   Column              Non-Null Count  Dtype
---  -
0   Kode Kecamatan      46 non-null    object
1   Kecamatan           46 non-null    object
2   Kode Kelurahan      46 non-null    object
3   Kelurahan           46 non-null    object
4   RW                  46 non-null    object
5   Jumlah Kepala Keluarga 46 non-null    int64
6   Jumlah Anggota Keluarga 46 non-null    int64
dtypes: int64(2), object(5)
memory usage: 2.6+ KB
```

3 Mengganti datatype atribut

```
[ ]: df.Kelurahan = pd.Categorical(df.Kelurahan,ordered=True,categories=['Antapani_Kulon', 'Antapani Tengah', 'Antapani Wetan', 'Antapani Kidul'])
df.head()
```

```
[ ]: Kode Kecamatan Kecamatan Kode Kelurahan Kelurahan RW \
0      32.73.20 Antapani 32.73.20.1001 Antapani Kulon 1
1      32.73.20 Antapani 32.73.20.1001 Antapani Kulon 2
2      32.73.20 Antapani 32.73.20.1001 Antapani Kulon 3
3      32.73.20 Antapani 32.73.20.1001 Antapani Kulon 4
4      32.73.20 Antapani 32.73.20.1001 Antapani Kulon 5

Jumlah Kepala Keluarga Jumlah Anggota Keluarga
0                      98                    255
1                       7                     21
2                      26                     56
3                       6                     17
4                     161                    379
```

Mengecek data NaN pada dataframe kemiskinan

```
[ ]: df.isna().sum()
```

```
[ ]: Kode Kecamatan      0
Kecamatan                0
Kode Kelurahan           0
Kelurahan                1
RW                        0
Jumlah Kepala Keluarga   0
Jumlah Anggota Keluarga   0
dtype: int64
```

Memperbaiki nama atribut di tiap dataframe

```
[ ]: df.columns = df.columns.str.replace(' ','_', regex=True)
```

```
[ ]: df.isna().sum()
```

```
[ ]: Kode_Kecamatan      0
Kecamatan                0
Kode_Kelurahan           0
Kelurahan                1
RW                        0
Jumlah_Kepala_Keluarga   0
Jumlah_Anggota_Keluarga   0
dtype: int64
```

mengisi data NaN kelurahan

```
[ ]: df["Kelurahan"].fillna("Antapani Wetan", inplace = True)
```

menampilkan data yang telah dikelompokkan berdasarkan kelurahan

```
[ ]: df2 = df.groupby(['Kelurahan']).sum()
df2 = df2.drop(["Kode_Kecamatan", "Kecamatan", "Kode_Kelurahan", "RW"], axis=1)
df2.head()
```

```
[ ]:
```

	Jumlah_Kepala_Keluarga	Jumlah_Anggota_Keluarga
Kelurahan		
Antapani Kulon	481	1238
Antapani Tengah	866	1913
Antapani Wetan	2547	27
Antapani Kidul	388	942

4 Mengeksplorasi dataset Umkm

```
[ ]: df1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 143 entries, 0 to 142
Data columns (total 8 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Kode Kecamatan        143 non-null    object
1   Kecamatan              143 non-null    object
2   Kode Kelurahan        143 non-null    object
3   Kelurahan              143 non-null    object
4   Nama Perusahaan       143 non-null    object
5   Jenis Usaha           143 non-null    object
6   Alamat Usaha          143 non-null    object
7   Jumlah Pekerja        142 non-null    float64
dtypes: float64(1), object(7)
memory usage: 9.1+ KB
```

Memperbaiki nama atribut di tiap dataframe

```
[ ]: df1.columns = df1.columns.str.replace(' ', '_', regex=True)
```

Mengecek data NaN pada dataframe Umkm

```
[ ]: df1.isna().sum()
```

```
[ ]:
```

Kode_Kecamatan	0
Kecamatan	0
Kode_Kelurahan	0
Kelurahan	0
Nama_Perusahaan	0
Jenis_Usaha	0

```

Alamat_Usaha      0
Jumlah_Pekerja    1
dtype: int64

```

mengisi data NaN pada data umkm jumlah pekerja dengan 0

```

[ ]: import warnings
      warnings.filterwarnings('ignore')
      df1["Jumlah_Pekerja"].fillna("0", inplace = True)
      df1["Jumlah_Pekerja"] = df1["Jumlah_Pekerja"].astype("int64")

```

Mengecek data NaN pada dataframe kemiskinan

```

[ ]: df1.isna().sum()

```

```

[ ]: Kode_Kecamatan      0
      Kecamatan          0
      Kode_Kelurahan     0
      Kelurahan          0
      Nama_Perusahaan    0
      Jenis_Usaha        0
      Alamat_Usaha       0
      Jumlah_Pekerja     0
      dtype: int64

```

Mengganti datatype atribut

```

[ ]: df1.Jenis_Usaha = pd.Categorical(df1.
      ↪Jenis_Usaha,ordered=True,categories=['Fashion','Craft','Jasa','lainnya'])
      df1.head()

```

```

[ ]:  Kode_Kecamatan  Kecamatan  Kode_Kelurahan      Kelurahan  Nama_Perusahaan \
0      32.73.20      Antapani   32.73.20.1001  Antapani Kulon  Abdullah Syuaib
1      32.73.20      Antapani   32.73.20.1006  Antapani Wetan   Achmad Sutisna
2      32.73.20      Antapani   32.73.20.1001  Antapani Kulon      Ade
3      32.73.20      Antapani   32.73.20.1001  Antapani Kulon      Agus
4      32.73.20      Antapani   32.73.20.1006  Antapani Wetan   Agus Koswara

      Jenis_Usaha      Alamat_Usaha  Jumlah_Pekerja
0      Craft  Jl. Kalijati No. 42 RT 07/04 Antapani Kulon      20
1      NaN    Jl. Malangbong V No. 38 Antapani Bandung      0
2      NaN    Jl. Tarumasari RT 04/05 Antapani Kulon      3
3      NaN    Jl. Sukapura No. 2 RT 01/06 Antapani Kulon      1
4      Jasa   Jl. Sindangsari III RT 01/09 Antapani Wetan      2

```

Mengecek data NaN pada dataframe umkm

```

[ ]: df1.isna().sum()

```

```
[ ]: Kode_Kecamatan      0
      Kecamatan          0
      Kode_Kelurahan     0
      Kelurahan          0
      Nama_Perusahaan     0
      Jenis_Usaha        99
      Alamat_Usaha       0
      Jumlah_Pekerja     0
      dtype: int64
```

mengisi data NaN pada dataframe umkm menjadi lainnya

```
[ ]: df1["Jenis_Usaha"].fillna("lainnya", inplace = True)
```

mengelompokan data berdasarkan jenis usaha

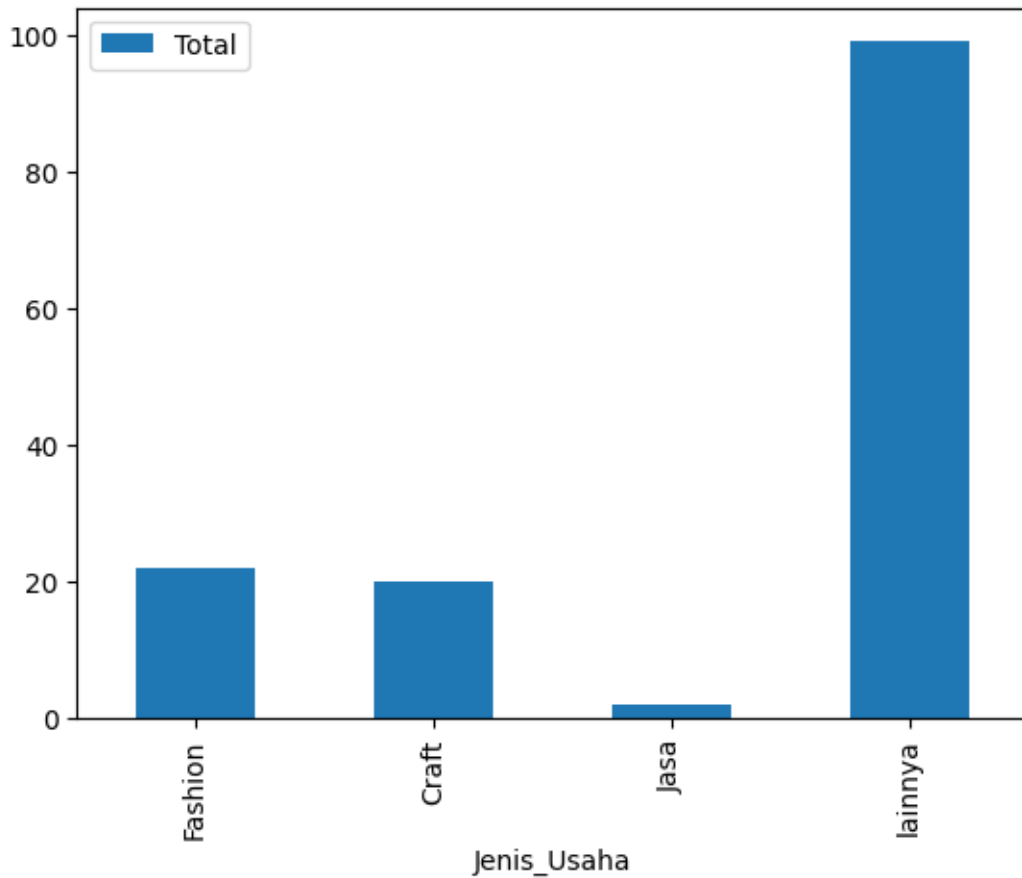
```
[ ]: import warnings
      warnings.filterwarnings('ignore')
      df3 = df1.groupby(['Jenis_Usaha']).count()
      df3 = df3.drop(["Kode_Kecamatan", "Kecamatan", "Kode_Kelurahan", "Kelurahan", "
      ↪Alamat_Usaha", "Nama_Perusahaan"], axis=1)
      df3.columns = df3.columns.str.replace('Jumlah_Pekerja', 'Total', regex=True)
      df3.head()
```

```
[ ]:
      Total
Jenis_Usaha
Fashion      22
Craft        20
Jasa         2
lainnya      99
```

menampilkan data dengan diagram batang berdasarkan jenis usaha

```
[ ]: import warnings
      warnings.filterwarnings('ignore')
      df_groups1 = df3.groupby(['Jenis_Usaha'])[['Total']].mean()
      df_groups1.plot(kind='bar')
```

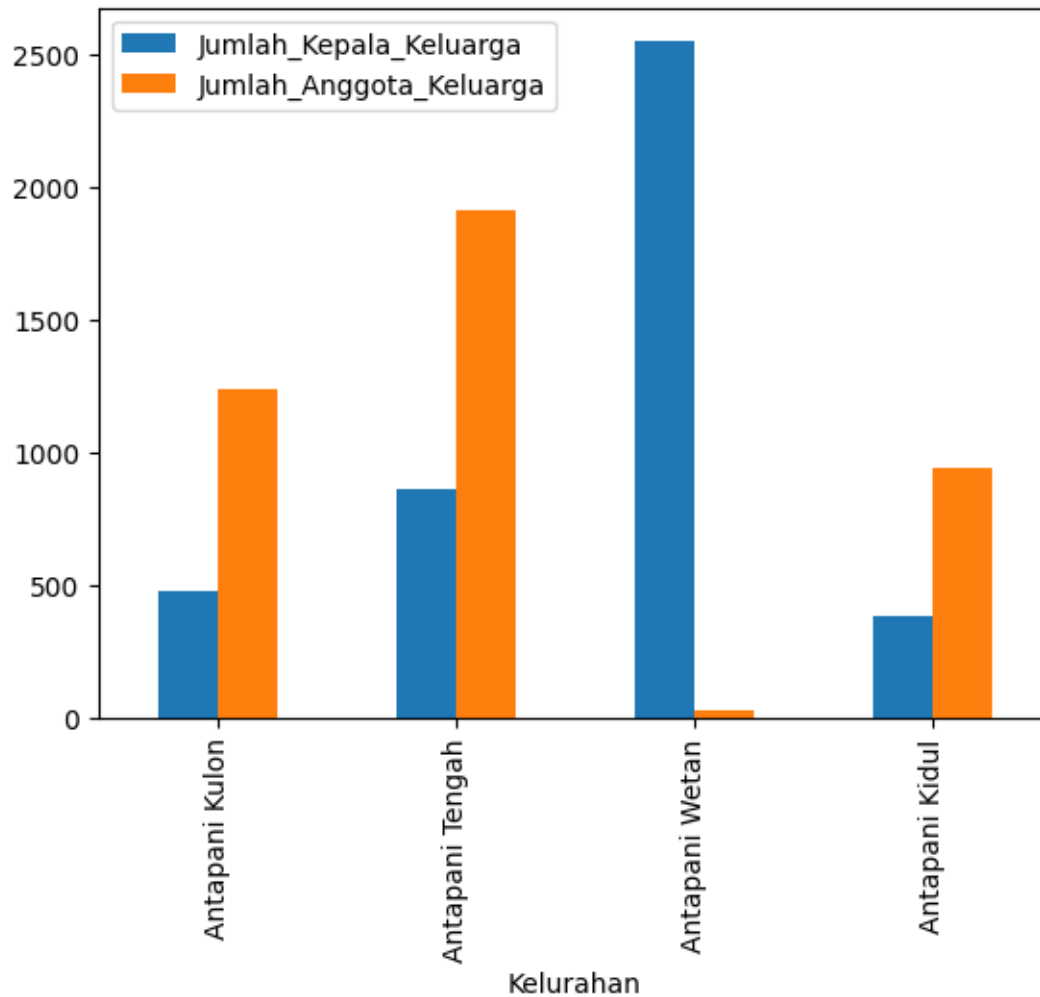
```
[ ]: <Axes: xlabel='Jenis_Usaha'>
```



menampilkan data dengan diagram batang data kemiskinan jumlah kepala keluarga dan anggota keluarga

```
[ ]: import warnings
warnings.filterwarnings('ignore')
df_groups2 = df2.groupby(['Kelurahan'])[['Jumlah_Kepala_Keluarga',
    'Jumlah_Anggota_Keluarga']].mean()
df_groups2.plot(kind='bar')
```

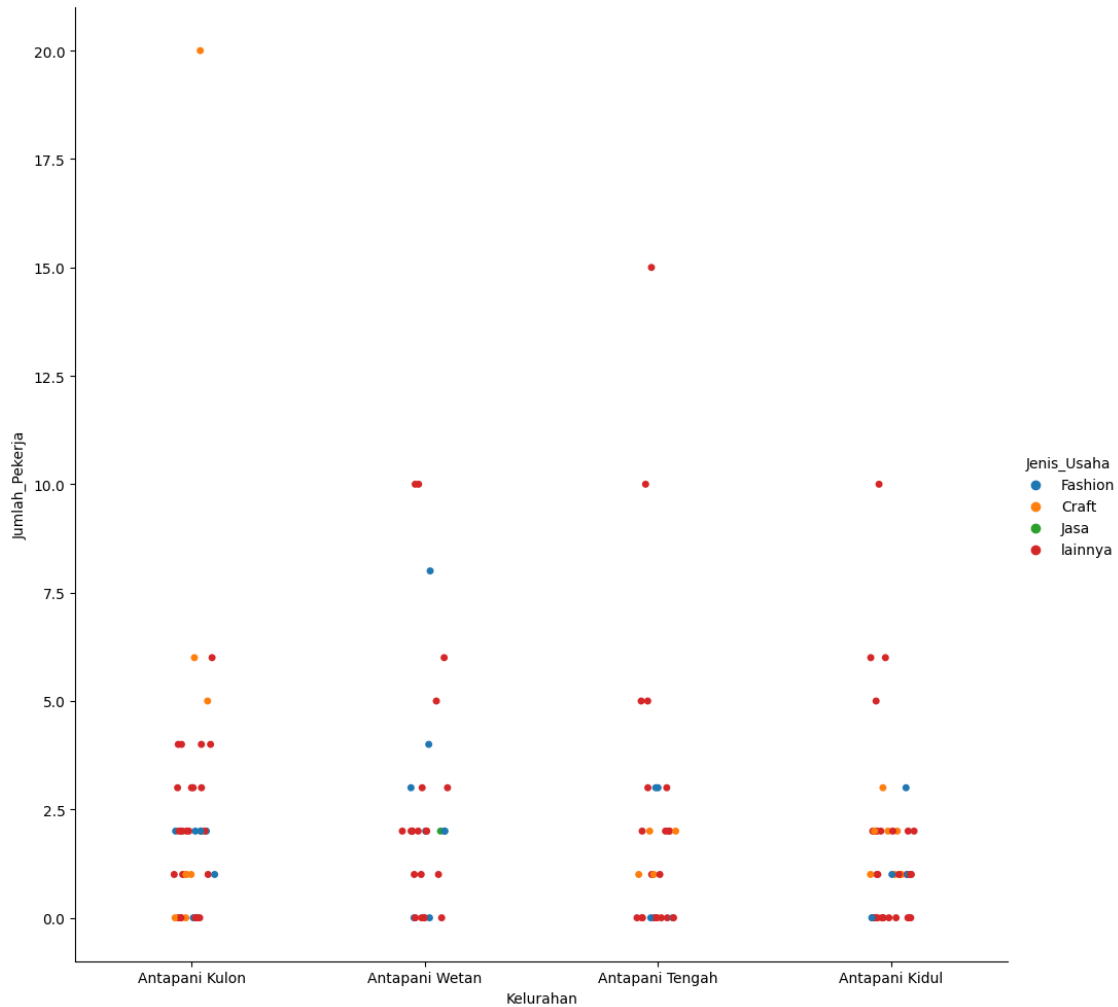
```
[ ]: <Axes: xlabel='Kelurahan'>
```



mnampilkan data pesebaran jumlah usaha [ada setiap kelurahan]

```
[ ]: import warnings
warnings.filterwarnings('ignore')
sns.catplot(x="Kelurahan", y="Jumlah_Pekerja", hue="Jenis_Usaha", data=df1,
            height=10)
```

```
[ ]: <seaborn.axisgrid.FacetGrid at 0x1e2634308d0>
```

```
[ ]: df1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 143 entries, 0 to 142
Data columns (total 8 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Kode_Kecamatan  143 non-null    object
1   Kecamatan       143 non-null    object
2   Kode_Kelurahan  143 non-null    object
3   Kelurahan       143 non-null    object
4   Nama_Perusahaan 143 non-null    object
5   Jenis_Usaha     143 non-null    category
6   Alamat_Usaha    143 non-null    object
7   Jumlah_Pekerja  143 non-null    int64
dtypes: category(1), int64(1), object(6)
```

memory usage: 8.3+ KB

mengelompokkan jumlah umkm berdasarkan kelurahan

```
[ ]: df4 = df1.groupby(['Kelurahan']).count()
df4 = df4.drop(["Kode_Kecamatan", "Kecamatan", "Kode_Kelurahan",
↳ "Alamat_Usaha", "Nama_Perusahaan", "Jenis_Usaha"], axis=1)
df4.columns = df4.columns.str.replace('Jumlah_Pekerja', 'Total_Umkm', regex=True)
df4.head()
```

```
[ ]:
Total_Umkm
Kelurahan
Antapani Kidul      43
Antapani Kulon      41
Antapani Tengah     29
Antapani Wetan      30
```

menggabungkan data jumlah umkm dan data kemiskinan di kecamatan antapani berdasarkan kelurahan

```
[ ]: df5 = pd.merge(df2, df4, how="left", on=["Kelurahan"])
df5.columns = df5.columns.str.
↳ replace('Jumlah_Kepala_Keluarga', 'Jumlah_Kepala_Keluarga_miskin', regex=True)
df5.columns = df5.columns.str.
↳ replace('Jumlah_Anggota_Keluarga', 'Jumlah_Anggota_Keluarga_miskin',
↳ regex=True)
df5.head()
```

```
[ ]:
Jumlah_Kepala_Keluarga_miskin \
Kelurahan
Antapani Kulon                481
Antapani Tengah              866
Antapani Wetan             2547
Antapani Kidul               388

Jumlah_Anggota_Keluarga_miskin  Total_Umkm
Kelurahan
Antapani Kulon                1238         41
Antapani Tengah              1913         29
Antapani Wetan                 27         30
Antapani Kidul                942         43
```

```
[ ]: df5.isna().sum()
```

```
[ ]: Jumlah_Kepala_Keluarga_miskin    0
Jumlah_Anggota_Keluarga_miskin    0
Total_Umkm                        0
dtype: int64
```

```
[ ]: df5.head()
```

```
[ ]:          Jumlah_Kepala_Keluarga_miskin  \
Kelurahan
Antapani Kulon                        481
Antapani Tengah                       866
Antapani Wetan                       2547
Antapani Kidul                        388

          Jumlah_Anggota_Keluarga_miskin  Total_Umkm
Kelurahan
Antapani Kulon                        1238          41
Antapani Tengah                       1913          29
Antapani Wetan                         27          30
Antapani Kidul                        942          43
```

menampilkan data dan membandingkan jumlah umkm di kecamatan antapani per kelurahan dengan jumlah umkm

```
[ ]: df_groups3 = df5.groupby(['Kelurahan'])[['Jumlah_Kepala_Keluarga_miskin',
↪ 'Jumlah_Anggota_Keluarga_miskin', 'Total_Umkm']].mean()
df_groups3.plot(kind='bar')
```

```
[ ]: <Axes: xlabel='Kelurahan'>
```

