Covid 19 Data Analysis

Ignatius S. Condro Atmawan

Data Processing

1. **Open** 2 data sets: clean & chaos
2. **Select** the variables of interest:
   - for clean data: "New Cases", "New Deaths", "New Recovered", "New Active Cases"
   - for chaos data: "New Cases.1", "New Deaths.1", "New Recovered.1", "New Active Cases.1"
   - for both data: "Date", "Location ISO Code", "Location"
3. **Cleaning** the data
4. **Select** the location of interest, e.g. "Jawa Barat", "DKI Jakarta", "Papua"
5. Calculate **statistics** based on our selection
6. **Visualize** the results

CovidData Class

- CovidData Class contain functions for the data processing
- This class is written in the script “covid_data_process.py”
- The processing is done in the script “covid19_main.py”

covid_data_process.py × covid19_main.py

```
22
                                                    ⚠1 ✓9 ^ v
      6 usages
23    class CovidData:
24        """Class to process Covid 19 Data"""
25        def __init__(self):
26            """create class object: Covid Data"""
27            self.about = "Process Covid Data"
28            self.file_name = ""  # the name of the file to be analyzed
29            self.current_directory = ""
30            self.output_directory = ""
31            self.df = None  # pandas data frame
32            self.columns = []  # list of columns of the data frame
33            self.df_oi_dict = {}  # data frame of interest
34            self.df_oi_statistics = {}  # descriptive statistics from data frame of interest
35            self.assign_directory()
```

covid_data_process.py covid19_main.py ×

```
22  import os
23
24  import helper_func.covid_data_process as cvd
25
26
27  # create object: Covid Data
28  myCovidData_clean = cvd.CovidData()
29  myCovidData_chaos = cvd.CovidData()
30  print("\n")
31
```

Open the file to get data frame

```python
def read_csv_data(self, file_name: str):
    """
    Read data from csv file

    :param file_name: (str) file name
    :return: Covid Data object itself
    """
    self.file_name = file_name
    self.df = pd.read_csv(self.file_name)

    self.columns = list(self.df.columns)
    return self
```

```python
# read the CSV file
myCovidData_clean = myCovidData_clean.read_csv_data(FileName_clean)

myCovidData_chaos = myCovidData_chaos.read_csv_data(FileName_chaos)
```

Data cleaning

- Clean numeric data from string/texts
- Drop empty data, e.g. NaN and/or None

```
# chaos data sets
mySelectedColumns_chaos = ["Date", "Location ISO Code", "Location", "New Cases.1", "New Deaths.1
columns_with_expected_integer_chaos = ["New Cases.1", "New Deaths.1", "New Recovered.1", "New Ac

myCovidData_chaos = myCovidData_chaos.select_column(mySelectedColumns_chaos)
myCovidData_chaos = myCovidData_chaos.clean_numeric_data(columns_with_expected_integer_chaos)
myCovidData_chaos = myCovidData_chaos.omit_empty_data()
```

Descriptive statistics

```
mylist = ["Jawa Barat", "DKI Jakarta", "Papua"]
myCovidData_chaos = myCovidData_chaos.get_dataframe_of_interest_based_on_string("Location", mylist)
print(myCovidData_chaos.df_oi_dict.keys())
print("\n")

print("Statistics:  ")
mySelectedStats = ["New Cases.1", "New Deaths.1", "New Recovered.1", "New Active Cases.1"]
print("before :  ", myCovidData_chaos.df_oi_statistics)
myCovidData_chaos = myCovidData_chaos.calculate_descriptive_statistics(mySelectedStats)
print("after :  ", myCovidData_chaos.df_oi_statistics)
print("\n")
myCovidData_chaos.save_descriptive_statistics()
```

Descriptive statistics: results

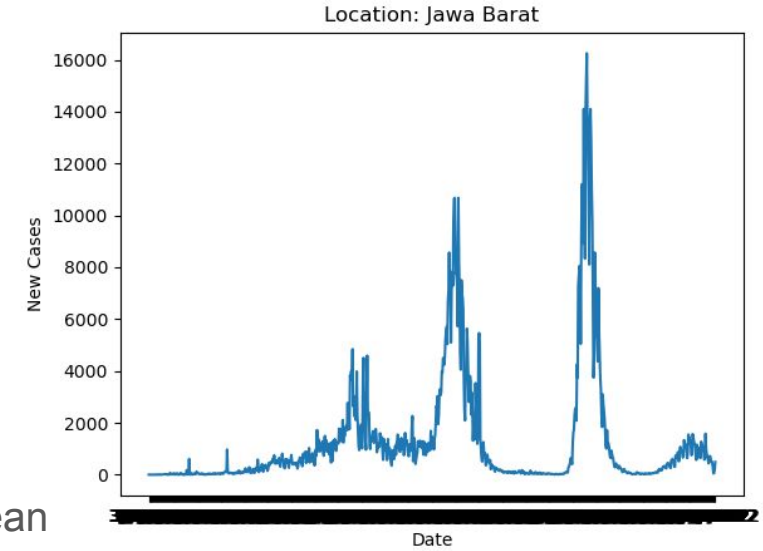
covid19_statistics_2023_10_05_18_3

Datei Bearbeiten Ansicht

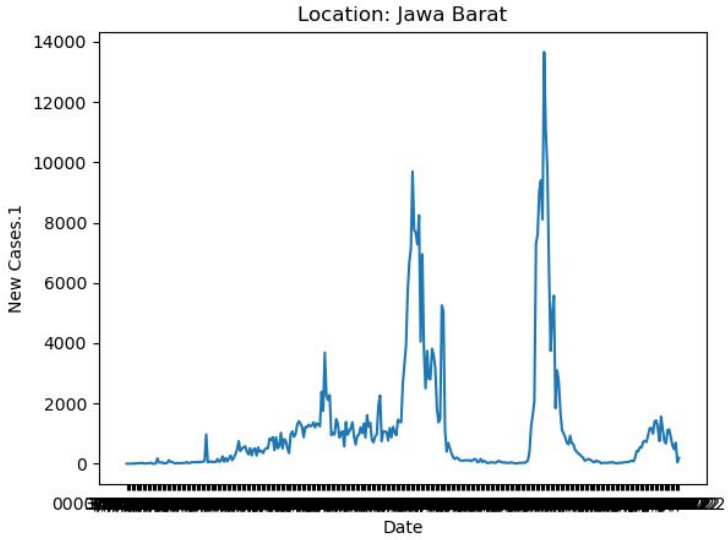
```
Location: Jawa Barat
data sets:  C:\data\belajar_Python\alfatraining2023\working_project\data_sets\covid_19_indonesia_time_series_all.csv

New Cases
sum =  1173731
mean =  1266.160733549083
median =  481.0

New Deaths
sum =  15937
mean =  17.192017259978424
median =  4.0

New Recovered
sum =  1144298
mean =  1234.4099244875945
median =  431.0

New Active Cases
sum =  13496
mean =  14.558791801510248
median =  8.0
```

Data Visualization

```
# plot figures
myCovidData_chaos.plot_from_saved_dict("Date", "New Cases.1", "Location: Jawa Barat", True)
myCovidData_chaos.plot_from_saved_dict("Date", "New Cases.1", "Location: Papua", False)
myCovidData_chaos.plot_from_saved_dict("Date", "New Cases.1", "Location: DKI Jakarta", False)
```



clean



chaos