# Creating & Parsing Data

---

## H4 Project Structure

KB A.I. Challenge

```
[KB A.I. Challenge]
    <customerCreater>
        -- Postgre.py
        -- config.json
        -- customerCreater.py
        -- menNames.txt
        -- womenNames.txt
        -- tblCreate.txt

    <consumptionCreater>
        -- Postgre.py
        -- config.json
        -- consumptionData.py
        -- tbl_creation.txt

    <chatBotProto>
        -- Postgre.py
        -- config.json
        -- chatbotProto.py

    <firstModelTotal>
        -- Postgre.py
        -- config.json
        -- firstMachine.py

    <logAnalysis>
        -- Postgre.py
        -- config.json
        -- logCreater.py
        -- logParser.py
        -- tblCreate.txt

    <chatbotLogParse>
        -- Postgre.py
        -- config.json
        -- chatbotLog.py

    <trainDataCreater>
        -- Postgre.py
        -- config.json
        -- trainDataCreater.py
```

## Execute Command & Parameters

### 1. customer data creation

```
$ python ./customerCreater.py 20000
```

> Create 20000 customers' data

| userID | name | gender | age | phone | email | job | hasCar | getMarried | children | incomeLevel | assestsLevel |
|---|---|---|---|---|---|---|---|---|---|---|---|

### 2. consumption data creation

```
$ python ./consumptionData.py 20000
```

> Create 20000 customers' consumption data [90 days]

| userID | a | b | c | d | e | f | g | h | i | j | k | l | m | n | o | consumeDate |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|

### 3. interim numpy matrix creation [interests.npy]

```
$ python ./firstMachine.py 20000 interests.npy
```

> interests.npy is used as parameter for logCreater.py, chatbotLog.py's executing time

### 4. log data creation to text file

```
$ python ./logCreater.py 20000 ../firstModelTotal/interests.npy logs_20000.txt
```

> logs_20000.txt is raw log data

### 5. parsing raw log data to DB

```
$ python ./logParser.py ./logs_20000.txt
```

> after parsing logs_20000.txt, insert into DB parsed data

| IP | accessTime | method | URL | statusCode | referer | userID |
|---|---|---|---|---|---|---|

### 6. analyze log data using chatbot

```
$ python ./chatbotLog.py 20000 ../firstModelTotal/interests.npy
```

> after analyzing log data, create interim numpy matrices. These matrices are used as parameters for trainDataCreater.py 's executing time

- Created File List -

```
trainDataForMachine_0.npy
trainDataForMachine_1.npy
trainDataForMachine_2.npy
trainDataForMachine_3.npy
trainDataForMachine_4.npy
trainDataForMachine_5.npy
trainDataForMachine_6.npy
trainDataForMachine_7.npy
trainDataForMachine_8.npy
trainDataForMachine_9.npy
trainDataForMachine_10.npy
trainDataForMachine_11.npy
trainDataForMachine_12.npy
trainDataForMachine_13.npy
trainDataForMachine_14.npy
```

### 7. parse data, extract train dataset

```
$ python ./trainDataCreater.py ../chatbotLogParse/trainDataForMachine_0.npy
```

> parsing data, extract train dataset for each Deep Learning Model
>
> [ trainDataForMachine_N_trainX , trainDataForMachine_N_trainY ]

- Created File List -

```
trainDataForMachine_0_trainX.npy
trainDataForMachine_0_trainY.npy
```

### Execution

```
$ python ./trainDataCreater.py ../chatbotLogParse/trainDataForMachine_1.npy
$ python ./trainDataCreater.py ../chatbotLogParse/trainDataForMachine_2.npy
$ python ./trainDataCreater.py ../chatbotLogParse/trainDataForMachine_3.npy
$ python ./trainDataCreater.py ../chatbotLogParse/trainDataForMachine_4.npy
$ python ./trainDataCreater.py ../chatbotLogParse/trainDataForMachine_5.npy
$ python ./trainDataCreater.py ../chatbotLogParse/trainDataForMachine_6.npy
$ python ./trainDataCreater.py ../chatbotLogParse/trainDataForMachine_7.npy
$ python ./trainDataCreater.py ../chatbotLogParse/trainDataForMachine_8.npy
$ python ./trainDataCreater.py ../chatbotLogParse/trainDataForMachine_9.npy
$ python ./trainDataCreater.py ../chatbotLogParse/trainDataForMachine_10.npy
```

```
$ python ./trainDataCreater.py ../chatbotLogParse/trainDataForMachine_11.npy
$ python ./trainDataCreater.py ../chatbotLogParse/trainDataForMachine_12.npy
$ python ./trainDataCreater.py ../chatbotLogParse/trainDataForMachine_13.npy
$ python ./trainDataCreater.py ../chatbotLogParse/trainDataForMachine_14.npy
```

### H6 Extracted matrices

```
trainDataForMachine_1_trainX.npy
trainDataForMachine_1_trainY.npy
trainDataForMachine_2_trainX.npy
trainDataForMachine_2_trainY.npy
trainDataForMachine_3_trainX.npy
trainDataForMachine_3_trainY.npy
trainDataForMachine_4_trainX.npy
trainDataForMachine_4_trainY.npy
trainDataForMachine_5_trainX.npy
trainDataForMachine_5_trainY.npy
trainDataForMachine_6_trainX.npy
trainDataForMachine_6_trainY.npy
trainDataForMachine_7_trainX.npy
trainDataForMachine_7_trainY.npy
trainDataForMachine_8_trainX.npy
trainDataForMachine_8_trainY.npy
trainDataForMachine_9_trainX.npy
trainDataForMachine_9_trainY.npy
trainDataForMachine_10_trainX.npy
trainDataForMachine_10_trainY.npy
trainDataForMachine_11_trainX.npy
trainDataForMachine_11_trainY.npy
trainDataForMachine_12_trainX.npy
trainDataForMachine_12_trainY.npy
trainDataForMachine_13_trainX.npy
trainDataForMachine_13_trainY.npy
trainDataForMachine_14_trainX.npy
trainDataForMachine_14_trainY.npy
```