

№8 Дәріс. Python бағдарламалау тілі және Pandas кітапханасы

Пәннің мақсаты:

Студенттерді Python бағдарламалау тілі негіздерімен және Pandas кітапханасын қолданып деректерді өңдеу, талдау дағдыларымен таныстыру. Курсты меңгеру нәтижесінде студенттер деректерді жүктеу, тазалау, түрлендіру, агрегаттау, талдау және визуализациялау операцияларын орындай алады.

Кілт сөздер:

Python, Pandas, NumPy, DataFrame, Series, деректерді талдау, деректерді жүктеу, индекстеу, фильтрация, агрегаттау, визуализация, Jupyter Notebook.

Дәріс жоспары

1. Кіріспе
 - Python тілінің деректер ғылымындағы рөлі
 - Pandas кітапханасының маңыздылығы
2. Python бағдарламалау тілі туралы жалпы түсінік
 - Негізгі синтаксис
 - Айнымалылар, деректер типтері
 - Тізімдер, кортеждер, сөздіктер
 - Циклдер және шарттар
3. NumPy кітапханасына қысқаша шолу
 - NumPy массивтері
 - Векторлық операциялар
4. Pandas кітапханасы
 - Series және DataFrame құрылымы
 - Деректерді импорттау және экспорттау (CSV, Excel)
 - Индекстеу және фильтрация
 - Деректерді тазалау (жоғалған мәндер, дубльдер)
 - Топтау және агрегаттау
5. Pandas көмегімен деректерді визуализациялау
 - Matplotlib/Seaborn пайдалану
6. Қорытынды

Кіріспе

Қазіргі цифрлық дәуірде ақпараттың көлемі бұрын-соңды болмаған деңгейге жетті. Күн сайын миллиардтаған құрылғылар мен қызметтер жаңа деректер ағындарын қалыптастырып жатыр. Осындай үлкен көлемді деректерден қажетті ақпаратты бөліп алу, талдау және дұрыс шешімдер қабылдау тек заманауи құралдардың көмегімен жүзеге асырылады. Солардың ішінде Python бағдарламалау тілі мен Pandas кітапханасы деректерді өңдеудің ең қуатты және тиімді құралдары ретінде кеңінен қолданылады.

Python қарапайым синтаксисімен, модульдік құрылымымен және ғылыми бағдарламалау саласында кеңінен таралған кітапханаларымен ерекшеленеді. NumPy, Pandas, Matplotlib, SciPy, Scikit-learn секілді кітапханалардың болуы Python тілін деректер ғылымы, жасанды интеллект, машиналық оқыту, статистикалық талдау және веб-разработка салаларында жетекші құралға айналдырды.

Pandas кітапханасы деректерді құрылымдық форматта сақтауға, өңдеуге және талдауға арналған. Ол деректерді жүктеуді, тазалауды, фильтрациялауды, топтастыруды, статистикалық талдауды және визуализациялауды айтарлықтай жеңілдетеді. Pandas көмегімен күрделі аналитикалық есептерді қысқа және түсінікті код арқылы шешуге болады.

Бұл дәрісте біз Python тілінің негізгі мүмкіндіктерін қарап, Pandas кітапханасының негізгі құрылымдарын — **Series** және **DataFrame** форматтарын, сондай-ақ деректерді жүктеу, өңдеу, агрегаттау және визуализациялау әдістерін меңгереміз. Курсты меңгеру нәтижесінде студенттер нақты деректермен жұмыс істеуге, аналитикалық есептерді орындауға және алынған нәтижелер бойынша тұжырым жасауға қабілетті болады.

1. Python бағдарламалау тілі

Python ерекшеліктері

- Ашық бастапқы код
- Оқу жеңілдігі
- Ғылыми есептеулерге қолайлы
- Көп кітапханалар (NumPy, Pandas, Matplotlib, TensorFlow, Scikit-learn)

Негізгі синтаксис мысалы

```
python

x = 10
y = 5
print("Сума =", x + y)
```

Тізімдер және цикл

```
python

a = [1, 2, 3, 4]
for i in a:
    print(i)
```

2. NumPy кітапханасы

NumPy — көпөлшемді массивтермен жылдам жұмыс істеуге арналған Python кітапханасы.

```
python

import numpy as np

arr = np.array([1, 2, 3, 4])
print(arr.mean())
```

3. Pandas кітапханасы Series және DataFrame

```
python

import pandas as pd

data = {"Аты": ["Айжан", "Ержан"], "Жасы": [20, 22]}
df = pd.DataFrame(data)
print(df)
```

Деректерді оқу

```
python

df = pd.read_csv("data.csv")
```

Фильтрация

```
python

df[df["Жасы"] > 20]
```

Жоғалған деректерді толтыру

```
python

df.fillna(0, inplace=True)
```

Топтау

```
python

df.groupby("Категория").mean()
```

Қорытынды

Сонымен, Python бағдарламалау тілі мен Pandas кітапханасы қазіргі деректер талдау экожүйесінің негізгі құралдары болып табылады. Python-ның икемділігі, қарапайымдылығы және кең экожүйесі деректерді өңдеу процесін

оңтайландырады. Ал Pandas деректерді кестелік форматта ұйымдастыруға, оларды тиімді өңдеуге және талдауға мүмкіндік береді.

Pandas арқылы біз:

- ✓ деректерді жүктеуді, түрлендіруді және тазалауды,
- ✓ топтастыру және агрегаттау операцияларын орындауды,
- ✓ статистикалық талдаулар жүргізуді,
- ✓ визуализация жасауды үйренеміз.

Бұл мүмкіндіктер Pandas-ты бизнес аналитикасы, қаржы, медицина, маркетинг, ғылым, өндіріс, банк жүйесі және жасанды интеллект жүйелерінде әмбебап құралға айналдырады.

Деректермен жұмыс істеу дағдылары — бүгінгі еңбек нарығындағы ең қажетті қабілеттердің бірі. Сондықтан Python мен Pandas-ты терең меңгеру болашақта кәсіби өсуге және бәсекеге қабілетті маман атануға зор мүмкіндік береді.

Болашақта бұл білімді:

- Big Data платформалары (Spark, Hadoop),
 - Машиналық оқыту кітапханалары (Sklearn, TensorFlow, PyTorch),
 - SQL және NoSQL деректер базалары,
 - Stream-processing жүйелері (Kafka, Flink)
- сияқты құралдармен бірге қолдану арқылы кеңейтуге болады.

Осылайша Python және Pandas — деректер саласына алғашқы және өте маңызды қадам болып табылады. Деректерді талдау мәдениетін қалыптастыру, ақпаратты дұрыс түсініп, саналы шешім қабылдай білу — заманауи IT-мамандардың басты дағдыларының бірі.

Бақылау сұрақтары

1. Python тілінің негізгі артықшылықтары қандай?
2. Pandas кітапханасы қандай міндеттерді орындайды?
3. DataFrame және Series дегеніміз не?
4. Pandas кітапханасында деректерді жүктеу қандай функциямен орындалады?
5. Жоғалған деректерді өңдеу әдістерін атаңыз.
6. NumPy мен Pandas кітапханасының айырмашылығы неде?
7. Деректерді топтау қалай жүзеге асады?

Пайдаланылған әдебиеттер

1. Wes McKinney — *Python for Data Analysis*, O'Reilly, 2022
2. Jake VanderPlas — *Python Data Science Handbook*, 2nd Ed., 2023
3. Aurélien Géron — *Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow*, 2022
4. Allen B. Downey — *Think Python*, O'Reilly, 2016
5. Орысша/Қазақша материалдар:
 - <https://pythonworld.ru>

- <https://python.org>
- Pandas ресми құжаты: <https://pandas.pydata.org>