№8 ЗЕРТХАНАЛЫҚ ЖҰМЫС
Pandas құралдарымен кестелік деректерді сүзу және талдау

1. Жұмыстың мақсаты
Pandas кітапханасының көмегімен CSV форматындағы көлік деректерін оқу, DataFrame құрылымында сақтау, бағандар мен жолдарды таңдау, шарттар бойынша сүзу, статистикалық талдау жүргізу және өңделген нәтижелерді жаңа CSV файлына жазу.

2. Теориялық мәліметтер
Pandas — Python тіліндегі кестелік деректерді өңдеуге және талдауға арналған кітапхана.

Pandas кітапханасының негізгі құрылымдары:
- Series — бірөлшемді деректер құрылымы;
- DataFrame — жолдар мен бағандардан тұратын екіөлшемді кестелік құрылым.

NumPy массиві көбінесе біртекті сандық деректермен жұмыс істеуге бағытталған болса, Pandas кестесінде бірнеше түрлі деректер типіндегі бағандар болуы мүмкін.

Мысалы:

ID	**Temperature**	**Voltage**	**SOC**
1	30	380	85
2	35	375	80
3	42	370	75
4	48	365	70
5	55	360	65

3. Pandas кітапханасын импорттау

```
import pandas as pd
```

Google Colab ортасында Pandas әдетте алдын ала орнатылған.

Тапсырма №1. DataFrame құру

Электрмобиль батареясының өлшеу деректерінен кесте құрыңыз.

```
import pandas as pd

data = {
   "ID": [1, 2, 3, 4, 5, 6, 7, 8, 9, 10],
   "Temperature": [30, 35, 42, 48, 55, 38, 44, 60, 33, 50],
   "Voltage": [380, 375, 370, 365, 360, 378, 372, 355, 382, 368],
   "SOC": [85, 80, 75, 70, 65, 82, 74, 60, 88, 68]
}

df = pd.DataFrame(data)

print(df)
```

Орындалатын тапсырмалар:

1. DataFrame құру.
2. Жолдар мен бағандар санын анықтау.
3. Баған атауларын шығару.
4. Әр бағанның деректер типін анықтау.

Қажетті командалар:

```
print(df.shape)
print(df.columns)
print(df.dtypes)
```

Тапсырма №2. DataFrame құрылымын зерттеу

Pandas кітапханасында деректерді бастапқы тексеруге арналған әдістер бар.

```
print(df.head())
print(df.tail())
print(df.info())
print(df.describe())
```

Команда	**Қызметі**
head()	Алғашқы 5 жолды көрсетеді
tail()	Соңғы 5 жолды көрсетеді
info()	Бағандар мен деректер типтері туралы ақпарат береді
describe()	Сандық бағандардың статистикасын көрсетеді
shape	Жолдар мен бағандар санын береді

Тапсырмалар:

1. Алғашқы 3 жолды шығару.
2. Соңғы 3 жолды шығару.
3. Температураның орташа мәнін анықтау.
4. Кернеудің минималды мәнін анықтау.
5. SOC көрсеткішінің максималды мәнін анықтау.

Тапсырма №3. Бағандарды таңдау

DataFrame құрылымынан жеке бағандарды таңдаңыз.

```
print(df["Temperature"])
```

Екі бағанды таңдау:

```
print(df[["Temperature", "Voltage"]])
```

Тапсырмалар:

1. Тек Temperature бағанын шығару.
2. Voltage және SOC бағандарын шығару.
3. Температураның орташа мәнін есептеу.

4. Кернеудің минималды және максималды мәндерін анықтау.

```
average_temp = df["Temperature"].mean()

print("Average temperature:", average_temp)
```

Тапсырма №4. Деректерді шарт бойынша сүзу

Температурасы 45°C-тан жоғары өлшеулерді анықтаңыз.

```
high_temp = df[df["Temperature"] > 45]

print(high_temp)
```

Екі шартты біріктіру:

```
filtered = df[
    (df["Temperature"] > 45) &
    (df["SOC"] < 75)
]

print(filtered)
```

Тапсырмалар:

1. Температурасы 45°C-тан жоғары өлшеулерді табу.
2. SOC деңгейі 70%-дан төмен өлшеулерді табу.
3. Кернеуі 370 В-тан төмен өлшеулерді табу.
4. Температурасы 45°C-тан жоғары және SOC деңгейі 75%-дан төмен өлшеулерді табу.
5. Әр шартқа сәйкес келетін жолдар санын анықтау.

Тапсырма №5. Жаңа есептелетін баған қосу

Температураны классификациялайтын жаңа Status бағанын құрыңыз.

Оқу диапазондары:

Температура	Күйі
<15°C	LOW
15–45°C	NORMAL
>45°C	HIGH

```
def check_temperature(temp):

    if temp < 15:
        return "LOW"

    elif temp <= 45:
        return "NORMAL"

    else:
        return "HIGH"


df["Status"] = df["Temperature"].apply(check_temperature)

print(df)
```

apply() функциясы берілген функцияны бағанның әр элементіне қолданады.

Тапсырмалар:

1. Status бағанын құру.
2. Әр жолдың күйін анықтау.
3. HIGH күйіндегі өлшеулерді сүзу.
4. Әр күйдегі өлшеулер санын есептеу.

Тапсырма №6. Санаттар/Категориялар бойынша деректерді талдау

Pandas кітапханасында санаттардың жиілігін есептеу үшін value_counts() әдісі қолданылады.

```
print(df["Status"].value_counts())
```

Тапсырмалар:

1. LOW, NORMAL, HIGH санаттарының санын есептеу.
2. Әр санаттағы орташа температураны анықтау.
3. Ең жоғары температура тіркелген жолды табу.

Санаттар бойынша орташа мән:

```
print(df.groupby("Status")["Temperature"].mean())
```

Ең жоғары температура тіркелген жол:

```
print(df.loc[df["Temperature"].idxmax()])
```

Тапсырма №7. DataFrame деректерін CSV файлына сақтау

Өңделген кестені CSV файлына сақтаңыз.

```
df.to_csv("battery_result.csv", index=False)

print("File saved successfully")
```

Мұндағы:

- to_csv() — DataFrame деректерін CSV файлына жазады;
- index=False — Pandas индексін жеке баған ретінде жазбайды.

Енді файлды қайта оқыңыз:

```
result = pd.read_csv("battery_result.csv")

print(result)
```

Файлдағы деректердің дұрыс сақталғанын тексеріңіз.

Тапсырма №8. CSV файлын Pandas арқылы оқу

№6 зертханалық жұмыста құрылған temperature.csv файлын пайдаланыңыз.

```
import pandas as pd

df = pd.read_csv("temperature.csv")

print(df)
```

Жеке тапсырмалар

Әр студент кемінде 15 жолдан және үш бағаннан тұратын CSV файлын құруы тиіс. Бірінші баған — ID, қалған екі баған — өлшеу параметрлері.
Барлық нұсқалар үшін ортақ талаптар:

1. CSV файлын pd.read_csv() арқылы оқу.
2. DataFrame құрылымын зерттеу.
3. Әр сандық бағанның минималды, максималды және орташа мәндерін есептеу.
4. Екі параметр бойынша сүзу орындау.
5. Жаңа есептелетін немесе классификациялық баған қосу.
6. Әр санаттағы жазбалар санын есептеу.
7. Өңделген DataFrame құрылымын жаңа CSV файлына сақтау.
8. Нәтижелер бойынша қорытынды жасау.
9.

№	Тақырып	CSV бағандары	Негізгі талдау
1	EV батарея мониторингі	ID,Voltage,Temperature	Кернеу <300 V, температура >45°C
2	Қозғалтқыш мониторингі	ID,Temperature,RPM	Температура >100°C, RPM >4000
3	Шина параметрлерін талдау	ID,TirePressure,TireTemperature	Қысым санаттары бойынша орташа температураны есептеу

4	Электрмобиль энергиясын талдау	ID,Voltage,Current	Қуатты есептеу және ең жоғары мәндерді сүзу
5	Батарея заряды	ID,SOC,Temperature	SOC <20%, температура >45°C
6	Көлік қозғалысының тиімділігі	ID,Speed,FuelConsumption	Жылдамдық топтары бойынша орташа отын шығынын есептеу
7	Электр қозғалтқышының мониторингі	ID,MotorTemp,Current	Температура мен токтың жоғары мәндерін анықтау
8	Екі температура датчигін салыстыру	ID,Sensor1,Sensor2	Difference = abs(Sensor1 - Sensor2)
9	Жылдамдық пен кедергі арақашықтығы	ID,Speed,Distance	Қашықтық <50 cm болатын өлшеулерді анықтау
10	Көлік жүйесінің кешенді мониторингі	ID,Temperature,Voltage,Speed	Үш параметр бойынша классификация және статистикалық талдау