

teht_4.1.py X teht_4.2.py

C:\> D > Suomi > 01_Tradenomi > 16_Python_Data > Teht_04 > teht_4.1.py > ...

```

1  import pandas as pd
2
3  # Путь к файлу
4  file_path = "Opinnäytetyökysely.xlsx"
5
6  # Загружаем данные
7  df = pd.read_excel(file_path, sheet_name="Kysely")
8
9  # Выбираем нужную переменную
10 time_col = "Opinnäytetyön tekemiseen kulunut aika ensimmäisistä aihekaavailuista työn valmistumiseen:kuukautta"
11
12 # Удаляем строки, где нет данных о времени выполнения
13 df_clean = df.dropna(subset=[time_col])
14
15 # Функция для расчёта статистики
16 def calculate_statistics(df, group_col):
17     stats = df.groupby(group_col)[time_col].agg([
18         ('Lukumäärä', 'count'),
19         ('Keskiarvo', 'mean'),
20         ('Keskihajonta', 'std'),
21         ('Pienin', 'min'),
22         ('Alaneljännes', lambda x: x.quantile(0.25)),
23         ('Mediaani', 'median'),
24         ('Yläneljännes', lambda x: x.quantile(0.75)),
25         ('Suurin', 'max')
26     ])
27     return stats.round(1) # Округляем до 1 знака после запятой
28
29 # 1. Статистика по Opiskelualala
30 stats_opiskelualala = calculate_statistics(df_clean, "Opiskelualala")
31 print("\nTilastot Opiskelualala:\n", stats_opiskelualala)
32
33 # 2. Статистика по Sukupuoli
34 stats_sukupuoli = calculate_statistics(df_clean, "Sukupuoli")
35 print("\nTilastot Sukupuoli:\n", stats_sukupuoli)
36
37

```

C:\WINDOWS\system32\cmd.exe

C:\D\Suomi\01_Tradenomi\16_Python_Data\Teht_04>python teht_4.1.py

Tilastot Opiskelualala:

	Lukumäärä	Keskiarvo	Keski-hajonta	Pienin	Alaneljännes	Mediaani	Ylä-neljännes	Suurin
Opiskelualala								
Kulttuuri	120	10.8	6.5	2.0	6.8	10.0	12.0	30.0
Liiketalous	60	8.2	3.6	2.0	6.0	8.0	11.0	16.0
Tekniikka	54	5.5	2.2	3.0	4.0	5.0	6.0	12.0

Tilastot Sukupuoli:

	Lukumäärä	Keskiarvo	Keski-hajonta	Pienin	Alaneljännes	Mediaani	Ylä-neljännes	Suurin
Sukupuoli								
Mies	84	6.3	3.1	2.0	4.0	6.0	8.0	14.0
Nainen	150	10.4	6.0	2.0	6.0	9.0	12.0	30.0

```
teht_4.1.py  teht_4.2.py X
C:\> D > Suomi > 01_Tradenomi > 16_Python_Data > Teht_04 > teht_4.2.py > ...
1  import pandas as pd
2
3  # Путь к файлу
4  file_path = "Opinnäytetyökysely.xlsx"
5
6  # Загружаем данные
7  df = pd.read_excel(file_path, sheet_name="Kysely")
8
9  # 1. Pivot-таблица для "Sain riittävästi ohjausta"
10 pivot_1 = df.pivot_table(
11     index="Opiskelualue",
12     columns="Sukupuoli",
13     values="Sain riittävästi ohjausta",
14     aggfunc="mean"
15 )
16
17 # Добавляем "Kaikki" (среднее по всем в столбцах)
18 pivot_1["Kaikki"] = pivot_1.mean(axis=1)
19
20 # Добавляем строку "Kaikki" (среднее по всем строкам)
21 pivot_1.loc["Kaikki"] = pivot_1.mean(axis=0)
22
23 # 2. Pivot-таблица для двух переменных (без строки "Kaikki")
24 pivot_2 = df.pivot_table(
25     index="Opiskelualue",
26     columns="Sukupuoli",
27     values=["Luotin ohjaajani neuvoihin", "Ohjaajaani oli helppo lähestyä"],
28     aggfunc="mean"
29 )
30
31 # Вывод таблиц в консоль
32 print("\nPivot-taulukko 1: Sain riittävästi ohjausta\n")
33 print(pivot_1.round(1))
34
35 print("\nPivot-taulukko 2: Ohjaajaani oli helppo lähestyä & Luotin ohjaajani neuvoihin\n")
36 print(pivot_2.round(1))
37
38
```

cmd C:\WINDOWS\system32\cmd.exe

C:\D\Suomi\01_Tradenomi\16_Python_Data\Teht_04>python teht_4.2.py

Pivot-taulukko 1: Sain riittävästi ohjausta

Sukupuoli	Mies	Nainen	Kaikki
Opiskelualue			
Kulttuuri	3.6	3.3	3.4
Liiketalous	3.6	3.0	3.3
Tekniikka	3.8	4.0	3.9
Kaikki	3.6	3.5	3.6

Pivot-taulukko 2: Ohjaajaani oli helppo lähestyä & Luotin ohjaajani neuvoihin

Sukupuoli	Luotin ohjaajani neuvoihin		Ohjaajaani oli helppo lähestyä	
	Mies	Nainen	Mies	Nainen
Opiskelualue				
Kulttuuri	4.1	4.2	4.6	4.5
Liiketalous	4.5	3.6	4.5	3.6
Tekniikka	4.4	5.0	4.7	5.0