

Palautettava tehtävä 1

June 1, 2025

1 Tehtävä 1: Excel-tiedoston tuonti ja perusrakenne

Tässä Jupyter-notebookissa suoritetaan seuraavat vaiheet: 1. Luetaan Excel-tiedosto (Opinnäytetyökysely.xlsx) pandas-DataFrameen. 2. Tarkastellaan aineiston rakennetta eri näkökulmista (alku, loppu, sarakkeiden määrä, puuttuvat arvot, tietotyypit). 3. Laaditaan frekvenssitaulukko yhdestä kategorisesta sarakkeesta (Opiskeluala).

```
[32]: import pandas as pd

file_path = r"D:\GitHub\PythonDataAnalytics\doc\Opinnäytetyökysely.xlsx"
df = pd.read_excel(file_path)
```

1.1 1) Tarkastellaan aineiston alkua ja loppua

Näytetään DataFrame-olion ensimmäiset ja viimeiset rivit: - `df.head()` näyttää 5 ensimmäistä riviä. - `df.tail()` näyttää 5 viimeistä riviä.

```
[33]: # Ensimmäiset 5 riviä
print("First 5 rows (head)")
print(df.head(), end="\n\n")

# Viimeiset 5 riviä
print("Last 5 rows (tail)")
print(df.tail(), end="\n\n")
```

First 5 rows (head)

	Aikaleima	Kuinka löysit aiheesi? \
0	26.10.2015 21:11:28	Itse
1	26.10.2015 21:11:28	Itse
2	26.10.2015 21:11:28	Itse
3	26.10.2015 21:11:28	Itse
4	26.10.2015 21:11:28	Työharjoittelupaikasta

	Opinnäytetyöni oli hankkeistettu \
0	Kyllä
1	Ei
2	Kyllä
3	Ei
4	Kyllä

Oliko työsi teoreettinen vai käytännöllinen? Teoreettinen (1) -
Käytännöllinen (5) \

0	3
1	1
2	3
3	4
4	2

Pystyin itse vaikuttamaan aiheen valintaan \

0	5
1	5
2	5
3	5
4	4

Olin innostunut opinnäytetyötä tehdessäni \

0	5
1	5
2	4
3	5
4	2

Pystyin itse vaikuttamaan opinnäytetyöni ohjaajan valintaan \

0	5
1	3
2	1
3	3
4	4

Sain riittävästi ohjausta Hankin itse aktiivisesti tietoa työni aiheesta \

0	1	5
1	1	4
2	5	4
3	2	4
4	4	4

Tutkimusaiheeni kiinnosti minua ... \

0	5	...
1	4	...
2	5	...
3	5	...
4	3	...

Ohjaajan tuki opinnäytetyön eri vaiheissa:aineiston analyysissa \

0	1
1	2
2	4

3	4
4	3

Ohjaajan tuki opinnäytetyön eri vaiheissa:johtopäätösten ja yhteenvedon tekemisessä \

0	2
1	1
2	4
3	4
4	4

Ohjaajan tuki opinnäytetyön eri vaiheissa:raportoinnissa	Thesis grade	\
0	4	NaN
1	4	NaN
2	5	NaN
3	4	5.0
4	4	NaN

Opinnäytetyön tekemiseen kulunut aika ensimmäisistä aihekaavailuista työn valmistumiseen:kuukautta \

0	24.0
1	11.0
2	7.0
3	5.0
4	14.0

Opinnäytetyön tekemisaika työviikkoina (40 h) aihekuvauksen tekemisestä työn valmistumiseen:työviikkoa \

0	20.0
1	7.0
2	28.0
3	17.0
4	10.0

Oliko työ parityö?	Opiskeluala	Sukupuoli	Ikä
0	0 Kulttuuri	Nainen	23.0
1	0 Kulttuuri	Nainen	27.0
2	0 Kulttuuri	Nainen	22.0
3	0 Kulttuuri	Nainen	26.0
4	0 Kulttuuri	Mies	36.0

[5 rows x 42 columns]

Last 5 rows (tail)

	Aikaleima	Kuinka löysit aiheesi? \
237	26.10.2015 21:11:40	Työharjoittelupaikasta
238	26.10.2015 21:11:40	Työharjoittelupaikasta
239	26.10.2015 21:11:40	Itse

240 26.10.2015 21:11:40 Itse
241 26.10.2015 21:11:40 Työelämästä (ei työharjoittelupaikka)

Opinnäytetyöni oli hankkeistettu \

237	Kyllä
238	Kyllä
239	Kyllä
240	Kyllä
241	Kyllä

Oliko työsi teoreettinen vai käytännöllinen? Teoreettinen (1) -
Käytännöllinen (5) \

237	4
238	3
239	4
240	3
241	3

Pystyin itse vaikuttamaan aiheen valintaan \

237	4
238	4
239	5
240	5
241	2

Olin innostunut opinnäytetyötä tehdessäni \

237	5
238	5
239	5
240	5
241	5

Pystyin itse vaikuttamaan opinnäytetyöni ohjaajan valintaan \

237	5
238	5
239	1
240	1
241	5

Sain riittävästi ohjausta \

237	4
238	5
239	4
240	5
241	2

Hankin itse aktiivisesti tietoa työni aiheesta \

237	5
-----	---

238	5
239	5
240	5
241	4

	Tutkimusaiheeni kiinnosti minua ... \
237	5 ...
238	5 ...
239	5 ...
240	5 ...
241	4 ...

	Ohjaajan tuki opinnäytetyön eri vaiheissa:aineiston analyysissa \
237	2
238	4
239	2
240	5
241	3

	Ohjaajan tuki opinnäytetyön eri vaiheissa:johtopäätösten ja yhteenvedon tekemisessä \
237	3
238	3
239	2
240	4
241	4

	Ohjaajan tuki opinnäytetyön eri vaiheissa:raportoinnissa	Thesis grade \
237	3	NaN
238	3	4.0
239	5	NaN
240	5	4.0
241	5	4.0

	Opinnäytetyön tekemiseen kulunut aika ensimmäisistä aihekaavailuista työn valmistumiseen:kuukautta \
237	9.0
238	6.0
239	8.0
240	3.0
241	6.0

	Opinnäytetyön tekemisaika työviikkoina (40 h) aihekuvauksen tekemisestä työn valmistumiseen:työviikkoa \
237	16.0
238	21.0
239	32.0
240	15.0

241 11.0

	Oliko työ parityö?	Opiskeluala	Sukupuoli	Ikä
237	0	Tekniikka	Mies	24.0
238	0	Tekniikka	Mies	23.0
239	0	Tekniikka	Mies	33.0
240	0	Tekniikka	Mies	25.0
241	0	Tekniikka	Mies	27.0

[5 rows x 42 columns]

1.2 2) Sarakkeiden otsikot ja niiden alla olevien arvojen lukumäärä

`df.count()` antaa kullekin sarakkeelle (kolonnille) sen ei-tyhjien solujen lukumäärän. Näin näemme, kuinka monta havaintoa kussakin sarakkeessa on.

```
[34]: print("Column names with non-null counts")
      print(df.count(), end="\n\n")
```

```
Column names with non-null counts
Aikaleima
242
Kuinka löysit aiheesi?
242
Opinnäytetyöni oli hankkeistettu
242
Oliko työsi teoreettinen vai käytännöllinen? Teoreettinen (1) - Käytännöllinen
(5) 242
Pystyin itse vaikuttamaan aiheen valintaan
242
Olin innostunut opinnäytetyötä tehdessäni
242
Pystyin itse vaikuttamaan opinnäytetyöni ohjaajan valintaan
242
Sain riittävästi ohjausta
242
Hankin itse aktiivisesti tietoa työni aiheesta
242
Tutkimusaiheeni kiinnosti minua
242
Sain muilta opiskelijoilta tukea työni tekemisessä
242
Työn toimeksiantaja oli kiinnostunut työstäni
154
Työn toimeksiantaja oli kiinnostunut ohjaamaan työtäni
154
Ohjaajani panos tuki työtäni
```

242
 Saamani ohjaus oli asiantuntevaa
 242
 Valmistauduin ohjauspalaveriin
 242
 Saamani ohjaus oli motivoivaa
 242
 Työni ohjaaja vastasi nopeasti tiedusteluihini
 242
 Ohjaustilanteet eivät tuntuneet minusta pelottavilta
 242
 Ohjaajaani oli helppo lähestyä
 242
 Luotin ohjaajani neuvoihin
 242
 Eri ohjaustapojen käyttäminen ja niiden hyödyllisyys: Seminaarit
 242
 Eri ohjaustapojen käyttäminen ja niiden hyödyllisyys: Henkilökohtaiset
 tapaamiset ohjaajan kanssa 242
 Eri ohjaustapojen käyttäminen ja niiden hyödyllisyys: Sähköiset
 yhteydet ohjaajan kanssa 242
 Eri ohjaustapojen käyttäminen ja niiden hyödyllisyys: Keskustelut
 toimeksiantajan kanssa 242
 Hyödyllisyys: Seminaarit
 214
 Hyödyllisyys:Henkilökohtaiset tapaamiset ohjaajan kanssa
 220
 Hyödyllisyys: Sähköiset yhteydet ohjaajan kanssa
 226
 Hyödyllisyys: Keskustelut toimeksiantajan kanssa
 180
 Ohjaajan tuki opinnäytetyön eri vaiheissa:aiheen valinnassa
 242
 Ohjaajan tuki opinnäytetyön eri vaiheissa:tutkimuskysymysten tai
 kehittämistehtävän rajauksessa 242
 Ohjaajan tuki opinnäytetyön eri vaiheissa:menetelmien valinnassa
 242
 Ohjaajan tuki opinnäytetyön eri vaiheissa:aineiston analyysissa
 242
 Ohjaajan tuki opinnäytetyön eri vaiheissa:johtopäätösten ja yhteenvedon
 tekemisessä 242
 Ohjaajan tuki opinnäytetyön eri vaiheissa:raportoinnissa
 242
 Thesis grade
 78
 Opinnäytetyön tekemiseen kulunut aika ensimmäisistä aihekaavailuista työn
 valmistumiseen:kuukautta 234
 Opinnäytetyön tekemisaika työviikkoina (40 h) aihekuvauksen tekemisestä työn

```
valmistumiseen:työviikkoa      204
Oliko työ parityö?
242
Opiskeluala
242
Sukupuoli
242
Ikä
240
dtype: int64
```

1.3 3) Puuttuvien arvojen (NaN) lukumäärät sarakkeittain

`df.isnull().sum()` laskee jokaisen sarakkeen NaN-arvojen määrän. Kaikissa sarakkeissa, joissa on puuttuvia havaintoja, tämä näyttää tarkalleen montako riviä on tyhjänä.

```
[35]: print("Missing values per column")
      print(df.isnull().sum(), end="\n\n")
```

```
Missing values per column
Aikaleima
0
Kuinka löysit aiheesi?
0
Opinnäytetyöni oli hankkeistettu
0
Oliko työsi teoreettinen vai käytännöllinen? Teoreettinen (1) - Käytännöllinen
(5)                                0
Pystyin itse vaikuttamaan aiheen valintaan
0
Olin innostunut opinnäytetyötä tehdessäni
0
Pystyin itse vaikuttamaan opinnäytetyöni ohjaajan valintaan
0
Sain riittävästi ohjausta
0
Hankin itse aktiivisesti tietoa työni aiheesta
0
Tutkimusaiheeni kiinnosti minua
0
Sain muilta opiskelijoilta tukea työni tekemisessä
0
Työn toimeksiantaja oli kiinnostunut työstäni
88
Työn toimeksiantaja oli kiinnostunut ohjaamaan työtäni
88
Ohjaajani panos tuki työtäni
```


0
Saamani ohjaus oli asiantuntevaa
0
Valmistauduin ohjauspalaveriini
0
Saamani ohjaus oli motivoivaa
0
Työni ohjaaja vastasi nopeasti tiedusteluihini
0
Ohjaustilanteet eivät tuntuneet minusta pelottavilta
0
Ohjaajaani oli helppo lähestyä
0
Luotin ohjaajani neuvoihin
0
Eri ohjaustapojen käyttäminen ja niiden hyödyllisyys: Seminaarit
0
Eri ohjaustapojen käyttäminen ja niiden hyödyllisyys: Henkilökohtaiset
tapaamiset ohjaajan kanssa 0
Eri ohjaustapojen käyttäminen ja niiden hyödyllisyys: Sähköiset
yhteydet ohjaajan kanssa 0
Eri ohjaustapojen käyttäminen ja niiden hyödyllisyys: Keskustelut
toimeksiantajan kanssa 0
Hyödyllisyys: Seminaarit
28
Hyödyllisyys:Henkilökohtaiset tapaamiset ohjaajan kanssa
22
Hyödyllisyys: Sähköiset yhteydet ohjaajan kanssa
16
Hyödyllisyys: Keskustelut toimeksiantajan kanssa
62
Ohjaajan tuki opinnäytetyön eri vaiheissa:aiheen valinnassa
0
Ohjaajan tuki opinnäytetyön eri vaiheissa:tutkimuskysymysten tai
kehittämistehtävän rajauksessa 0
Ohjaajan tuki opinnäytetyön eri vaiheissa:menetelmien valinnassa
0
Ohjaajan tuki opinnäytetyön eri vaiheissa:aineiston analyysissa
0
Ohjaajan tuki opinnäytetyön eri vaiheissa:johtopäätösten ja yhteenvedon
tekemisessä 0
Ohjaajan tuki opinnäytetyön eri vaiheissa:raportoinnissa
0
Thesis grade
164
Opinnäytetyön tekemiseen kulunut aika ensimmäisistä aihekaavailuista työn
valmistumiseen:kuukautta 8
Opinnäytetyön tekemisaika työviikkoina (40 h) aihekuvauksen tekemisestä työn

```

valmistumiseen:työviikkoa      38
Oliko työ parityö?
0
Opiskeluala
0
Sukupuoli
0
Ikä
2
dtype: int64

```

1.4 4) Sarakkeiden tietotyypit

df.dtypes näyttää jokaisen sarakkeen (kolonin) tietotyypin. Tyypillisesti: - object = merkkijonoja (kategoria/data-kysymykset) - float64 / int64 = numeroarvot - datetime64[ns] = päivämäärä-aikatyypit

```

[36]: print("Data types of each column")
      print(df.dtypes, end="\n\n")

```

```

Data types of each column
Aikaleima
object
Kuinka löysit aiheesi?
object
Opinnäytetyöni oli hankkeistettu
object
Oliko työsi teoreettinen vai käytännöllinen? Teoreettinen (1) - Käytännöllinen
(5)                                     int64
Pystyin itse vaikuttamaan aiheen valintaan
int64
Olin innostunut opinnäytetyötä tehdessäni
int64
Pystyin itse vaikuttamaan opinnäytetyöni ohjaajan valintaan
int64
Sain riittävästi ohjausta
int64
Hankin itse aktiivisesti tietoa työni aiheesta
int64
Tutkimusaiheeni kiinnosti minua
int64
Sain muilta opiskelijoilta tukea työni tekemisessä
int64
Työn toimeksiantaja oli kiinnostunut työstäni
float64
Työn toimeksiantaja oli kiinnostunut ohjaamaan työtäni
float64

```

Ohjaajani panos tuki työtäni
int64

Saamani ohjaus oli asiantuntevaa
int64

Valmistauduin ohjauspalaveriini
int64

Saamani ohjaus oli motivoivaa
int64

Työni ohjaaja vastasi nopeasti tiedusteluihini
int64

Ohjaustilanteet eivät tuntuneet minusta pelottavilta
int64

Ohjaajaani oli helppo lähestyä
int64

Luotin ohjaajani neuvoihin
int64

Eri ohjaustapojen käyttäminen ja niiden hyödyllisyys: Seminaarit
int64

Eri ohjaustapojen käyttäminen ja niiden hyödyllisyys: Henkilökohtaiset
tapaamiset ohjaajan kanssa int64

Eri ohjaustapojen käyttäminen ja niiden hyödyllisyys: Sähköiset
yhteydet ohjaajan kanssa int64

Eri ohjaustapojen käyttäminen ja niiden hyödyllisyys: Keskustelut
toimeksiantajan kanssa int64

Hyödyllisyys: Seminaarit
float64

Hyödyllisyys:Henkilökohtaiset tapaamiset ohjaajan kanssa
float64

Hyödyllisyys: Sähköiset yhteydet ohjaajan kanssa
float64

Hyödyllisyys: Keskustelut toimeksiantajan kanssa
float64

Ohjaajan tuki opinnäytetyön eri vaiheissa:aiheen valinnassa
int64

Ohjaajan tuki opinnäytetyön eri vaiheissa:tutkimuskysymysten tai
kehittämistehtävän rajauksessa int64

Ohjaajan tuki opinnäytetyön eri vaiheissa:menetelmien valinnassa
int64

Ohjaajan tuki opinnäytetyön eri vaiheissa:aineiston analyysissa
int64

Ohjaajan tuki opinnäytetyön eri vaiheissa:johtopäätösten ja yhteenvedon
tekemisessä int64

Ohjaajan tuki opinnäytetyön eri vaiheissa:raportoinnissa
int64

Thesis grade
float64

Opinnäytetyön tekemiseen kulunut aika ensimmäisistä aihekaavailuista työn
valmistumiseen:kuukautta float64

```
Opinnäytetyön tekemisaika työviikkoina (40 h) aihekuvauksen tekemisestä työn valmistumiseen:työviikkoa    float64
Oliko työ parityö?
int64
Opiskeluala
object
Sukupuoli
object
Ikä
float64
dtype: object
```

1.5 5) Frekvenssitaulukko esimerkkisarakkeesta

Tehtävänannossa pyydettiin “teknisesti 5 pisteeseen vaadittavaa frekvenssitaulukkoa”. Yleensä se tarkoittaa jonkin kategorisen (tekstimuotoisen) sarakkeen arvojen esiintymistiheyksiä. Tässä tehdään frekvenssitaulukko sarakkeesta **“Opiskeluala”**.

Voit halutessasi korvata “Opiskeluala” jollain muulla kategorisella sarakkeella, esim. “Sukupuoli” tai “Olin innostunut opinnäytetyötä tehdessäni” tms.

```
[37]: # Frequency table: montako kappaletta kutakin "Opiskeluala"-arvoa
print("Frequency table for 'Opiskeluala'")
freq = df["Opiskeluala"].value_counts(dropna=False)
print(freq)
```

```
Frequency table for 'Opiskeluala'
Opiskeluala
Kulttuuri      124
Liiketalous     64
Tekniikka       54
Name: count, dtype: int64
```

1.6 Yhteenveto

Jupyter-notebookin:

- Ensimmäiset ja viimeiset rivit (`\.head()`, `\.tail()`)
- Sarakkeiden nimet ja ei-tyhjien solujen määrä (`\.count()`)
- Puuttuvien arvojen määrä (`\.isnull().sum()`)
- Sarakkeiden tietotyyppit (`\.dtypes`)
- Frekvenssitaulukko (`\.value_counts()`)