

Palautettava tehtävä 4

June 1, 2025

1 Tehtävä 4: Kuvaileva tilastotiede ja pivot-taulukot

Tässä notebookissa: 1. Luetaan sama Excel-tiedosto (Opinnäytetyökysely.xlsx) pandas-DataFrameen, sheet "Kysely". 2. Lasketaan kuvailevat tunnusluvut ("Opinnäytetyön tekemiseen kulunut aika..."). 3. Luodaan pivot-taulukko keskiarvoraporteista kolmesta ohjausta mittaavasta kysymyksestä ("Sain riittävästi ohjausta", "Ohjaajaani oli helppo lähestyä", "Luotin ohjaajani neuvoihin") kenttä-tason (Opiskeluala) mukaan.

1.1 1) Kirjastojen tuonti ja Excel-datan lataus

Tuodaan pandas ja luetaan sheet "Kysely" DataFrameen.

Varmista, että Opinnäytetyökysely.xlsx on samassa kansiossa kuin tämä notebook (tai päivitä polku).

```
[10]: import pandas as pd

def load_data(path: str) -> pd.DataFrame:
    """
    Lukee kyselydatan annetuista Excel-tiedostosta.
    """
    return pd.read_excel(path, sheet_name="Kysely", engine="openpyxl")

# Ladataan data
file_path = r"D:\GitHub\PythonDataAnalytics\doc\Opinnäytetyökysely.xlsx"
df = load_data(file_path)

# Tarkistetaan, että DataFrame latautui oikein (rivit x sarakkeet)
df.shape
```

[10]: (242, 42)

1.2 2) Kuvailevat tunnusluvut: "Opinnäytetyön tekemiseen kulunut aika...:kuukautta"

Lasketaan kuvailevat tunnusluvut (lukumäärä, keskiarvo, keskihajonta, minimi, kvartiilit, maksimi) sarakkeesta "Opinnäytetyön tekemiseen kulunut aika ensimmäisistä aihekaavailuista työn valmistumiseen:kuukautta". Poistetaan rivit, joilla arvo puuttuu.

```
[11]: def descriptive_stats(df: pd.DataFrame, col: str) -> pd.DataFrame:
      """
      Laskee kuvailevat tunnusluvut numeeriselle sarakkeelle ja nimeää rivit_
      ↪suomeksi.
      """
      df_time = df.dropna(subset=[col])
      desc = (
          df_time[col]
          .describe()
          .rename({
              "count": "Lukumäärä",
              "mean": "Keskiarvo",
              "std": "Keskihajonta",
              "min": "Minimi",
              "25%": "25 % kvartiili",
              "50%": "Mediaani",
              "75%": "75 % kvartiili",
              "max": "Maksimi"
          })
      )
      return desc.to_frame(name="Arvo")

time_var = (
    "Opinnäytetyön tekemiseen kulunut aika ensimmäisistä "
    "aihekaavailuista työn valmistumiseen:kuukautta"
)

stats_df = descriptive_stats(df, time_var)
stats_df
```

```
[11]:
```

	Arvo
Lukumäärä	234.000000
Keskiarvo	8.918803
Keskihajonta	5.516638
Minimi	2.000000
25 % kvartiili	5.000000
Mediaani	8.000000
75 % kvartiili	12.000000
Maksimi	30.000000

1.3 3) Pivot-taulukko: “Ohjauskysymykset” keskiarvoina kenttä-tasoinen

Kysymykset: - “Sain riittävästi ohjausta” - “Ohjaajaani oli helppo lähestyä” - “Luotin ohjaajani neuvoihin”

Tehdään pivot-taulukko, jossa rivi-indeksinä on **Opiskeluala** ja arvona kunkin kysymyksen keskiarvo.

```
[12]: def guidance_pivot(df: pd.DataFrame, index_col: str, questions: list) -> pd.
      ↪ DataFrame:
      """
      Luo pivot-taulukon, jossa riveinä on index_col (esim. 'Opiskeluala')
      ja sarakkeina esitettynä kunkin kysymyksen keskiarvo.
      """

      pivot = pd.pivot_table(
          df,
          index=index_col,
          values=questions,
          aggfunc="mean"
      )
      return pivot

questions = [
    "Sain riittävästi ohjausta",
    "Ohjaajaani oli helppo lähestyä",
    "Luotin ohjaajani neuvoihin"
]

pivot_df = guidance_pivot(df, "Opiskeluala", questions)
pivot_df
```

```
[12]:          Luotin ohjaajani neuvoihin  Ohjaajaani oli helppo lähestyä \
Opiskeluala
Kulttuuri          4.209677          4.467742
Liiketalous        3.906250          3.875000
Tekniikka          4.407407          4.703704

          Sain riittävästi ohjausta
Opiskeluala
Kulttuuri          3.354839
Liiketalous        3.218750
Tekniikka          3.777778
```

2 Yhteenveto

Jupyter-notebookissa: - Tuodaan `Opinnäytetyökysely.xlsx` pandas-DataFrameksi (sheet: "Kysely"). - Lasketaan kuvailevat tunnusluvut sarakkeesta "**Opinnäytetyön tekemiseen kulunut aika...**" ja näytetään tulos DataFrame-taulukkona. - Luodaan pivot-taulukko, jossa lasketaan kenttä-tasoiset (Opiskeluala) keskiarvot kolmesta ohjausta mittaavasta kysymyksestä ("Sain riittävästi ohjausta", "Ohjaajaani oli helppo lähestyä", "Luotin ohjaajani neuvoihin").