**Python ja data-analytiikka**

**Palautettava tehtävä 5**

Tässä harjoituksessa tutustutaan korrelaatiokertoimen laskemiseen sekä uusien muuttujien laskemiseen olemassa olevista. Muista, että korrelaatio ei välttämättä tarkoita kausaliteetin (syy-seuraus-suhde) olemassaoloa.

Palauta kaikki tehtävät Moodleen yhtenä pdf-tiedostona. Liitä tiedostoon myös käytetty Python-koodi.

TEHTÄVÄ 1

Laske opinnäytetyökyselystä seuraavien muuttujien väliset korrelaatiokertoimet ja niiden selitysasteet.

* Hankin itse aktiivisesti tietoa työni aiheesta / Tutkimusaiheeni kiinnosti minua
* Sain riittävästi ohjausta / Opinnäytetyön tekemiseen kulunut aika ensimmäisistä aihekaavailuista työn valmistumiseen:kuukautta

Muuttujan selitysaste on korrelaatiokerroin korotettuna toiseen potenssiin. Selitysaste kuvaa, että kuinka monta prosenttia muuttujan y vaihtelusta voidaan selittää muuttujan y vaihtelulla.   
  
Kahden tietyn sarakkeen korrelaatiokerroin lasketaan Dataframe-oliosta komennolla *korrelaatio = df['sarake1'].corr(df['sarake2'])*. Oletuksena lasketaan Pearsonin korrelaatiokerroin.

Tulosta tulokset kahden desimaalin tarkkuudella. Huomaa, että et voi suoraan liittää float-tyyppistä muuttujaa string-tulostukseen.

Kuva, joka sisältää kohteen teksti, Fontti, valkoinen

Tekoälyn generoima sisältö voi olla virheellistä.

TEHTÄVÄ 2

Yleinen tapa kyselyjen analysoinnissa on yhdistää samaan aiheeseen liittyviä kysymyksiä analysointia varten. Tämä tapahtuu laskemalla niiden keskiarvo.

Tutki onko sillä, että opiskelija sai itse vaikuttaa ohjaajan valintaan yhteyttä siihen, kuinka tyytyväinen hän oli saamaansa ohjaukseen. Tyytyväisyyttä ohjaukseen mitataan ainakin seuraavilla muuttujilla

* Ohjaajani panos tuki työtäni
* Saamani ohjaus oli asiantuntevaa
* Saamani ohjaus oli motivoivaa
* Työni ohjaaja vastasi nopeasti tiedusteluihini
* Ohjaustilanteet eivät tuntuneet minusta pelottavilta
* Ohjaajaani oli helppo lähestyä
* Luotin ohjaajani neuvoihin

Laske yllä mainituista sarakkeista keskiarvo ja muodosta keskiarvosta uusi sarake ”Kokonaistyytyväisyys ohjaukseen” Dataframe-olioon. Tämä tehdään komennolla *df["uusi\_keskiarvo"] = df[["sarake1", "sarake2"]].mean(axis=1)*. Parametrilla *axis = 1* kerrotaan, että keskiarvo lasketaan rivi kerrallaan (ts. sarakkeiden arvoista samalla rivillä).

Laske tämän jälkeen korrelaatiokerroin ja selitysaste muuttujille Pystyin itse vaikuttamaan opinnäytetyöni ohjaajan valintaan ja Kokonaistyytyväisyys ohjaukseen.

Huomaa, että pystyt eri funktioita soveltamalla laskemaan myös muunlaisia uusia muuttujia, kuten jollakin tavalla painotettuja.

TEHTÄVÄ 3

Muodosta opinnäytetyökyselyn seuraavista sarakkeista uusi Dataframe-olio ja tutki korreloiko opiskelijan innokkuus sen kanssa, että millaiseksi hän koki ohjaajan tuen opinnäytetyöprosessin aikana.

* Olin innostunut opinnäytetyötä tehdessäni
* Ohjaajan tuki opinnäytetyön eri vaiheissa:aiheen valinnassa
* Ohjaajan tuki opinnäytetyön eri vaiheissa:tutkimuskysymysten tai kehittämistehtävän rajauksessa
* Ohjaajan tuki opinnäytetyön eri vaiheissa:menetelmien valinnassa
* Ohjaajan tuki opinnäytetyön eri vaiheissa:aineiston analyysissa
* Ohjaajan tuki opinnäytetyön eri vaiheissa:johtopäätösten ja yhteenvedon tekemisessa
* Ohjaajan tuki opinnäytetyön eri vaiheissa:raportoinnissa

Tulosta kaikki korrelaatiot Taanilan matskun mukaisena taulukkona yhdellä Python-käskyllä.

Hyväksyttyyn suoritukseen (3 p) vaaditaan tehtävä 1 ja jompikumpi tehtävistä 2 tai 3 oikein tehtynä. Viiteen pisteeseen vaaditaan kaikki tehtävät oikein tehtynä.