**Python ja data-analytiikka**

**Lopputehtävä**

**TEHTÄVÄNANTO**

Lopputehtävässä analysoit ja visualisoit haluamaasi dataa Pythonilla. Data voi olla avointa dataa tai esim. työpaikkasi dataa.

Lopputehtävä koostuu kahdesta osasta:

1. Suunnittelu
2. Toteutus.

Molemmista osista palautetaan ohjeen mukainen raportti. Toteutusta ei saa aloittaa ennen, kuin suunnittelu on hyväksytty. Hyväksyttyyn suoritukseen vaaditaan suunnittelusta 5 pistettä ja toteutuksesta 15 pistettä.

Huom! Mikäli sinulla on **todellinen** data-analytiikkaprojekti, joka ei sovi näihin vaatimuksiin ja arviointikriteereihin, ole yhteydessä Petteriin. Soveltavan lopputehtävän tekeminen on mahdollista.

**VAATIMUKSET JA ARVIONTIKRITEERIT**

1. SUUNNITTELU (max 15 p)

Vapaamuotoinen suoraan Moodleen kirjoitettava teksti, josta käy ilmi seuraavat asiat:

* Mitä dataa analysoidaan? Dataa on oltava riittävästi, jotta se hyväksytään. Kiinnitä **halutessasi** datan valinnassa huomiota siihen, että siihen on mahdollista soveltaa korrelaation laskemista ja tilastollista testausta.
* Mitä datasta halutaan saada selville? Ts. miksi projekti toteutetaan.
* Minkälaisia toimenpiteitä (esim. visualisoinnit, yhdistäminen, siivoaminen) dataan tarvitsee tehdä.

Suunnitteluosa arvioidaan asteikolla 5 / 10 / 15 pistettä. Arviointikriteerit ovat:

* 5 pistettä: Analysoitava data esitellään selkeästi. Suunnitelmasta käy suunnilleen selkeästi ilmi se, mitä datalle tehdään ja miksi. Toimenpiteitä ei perustella, vaan ne ainoastaan esitellään.
* 10 pistettä: Analysoitava data esitellään selkeästi. Suunnitelmasta käy selkeästi ilmi se, mitä datalle tehdään ja miksi. Toimenpiteitä ei perustella, vaan ne ainoastaan esitellään.
* 15 pistettä: Analysoitava data esitellään selkeästi. Suunnitelmasta käy todella selkeästi ilmi se, mitä datalle tehdään ja miksi. Tehtävät toimenpiteet ja niiden tarpeellisuus perustellaan selkeästi.

1. TOTEUTUS (max 45 p)

Toteutuksessa palautetaan Savonian raportointipohjalla oleva raportti. Raportista on käytävä ilmi koko työn kulku. Raportti on kirjoitettava siten, että kuka tahansa, joka raportin lukee, saa käsityksen analysointityön tavoitteista, toteutuksesta ja tuloksista.

Pythonilla tehdyt taulukot, kaaviot, yms. on lisättävä osaksi raporttia

Lähteitä ei ole pakko käyttää. Mikäli niitä käytetään, niin lähdeviitteet ja lähdeluettelo on tehtävä Savonian ohjeiden mukaisesti.

**ARVIOINTIKRITEERIT**

Raportti (max 20 p)

* Raportti on toteutettu Savonian raportointiohjeiden mukaisesti (0–5 p).
* Työn tavoitteet esitellään raportissa (0–5 p).
* Työn kulku esitellään raportissa (0–5 p).
* Työn tulokset esitellään raportissa (0–5 p). Kaaviot, taulukot, yms. on liitettävä osaksi raporttia.
* Huom! Erityistä huomiota raportin arvioinnissa kiinnitetään raportin luettavuuteen ja johdonmukaisuuteen.
* **Analyysissa käytetty Python-koodi on liitettävä raporttiin liitteeksi.** Kiinnitä huomiota koodin kommentointiin.

Analysointi (max 0–22 p)

Tässä harjoituksessa ei edellytetä käytettävän mitään tiettyjä kaavioita, funktioita, tjms. Tärkeintä on jalostaa data informaatioksi siten, että dataa on mahdollista käyttää päätöksenteon tukena.

Kiinnitä huomiota seuraaviin asioihin:

* Analyysi on selkeä ja johdonmukainen
* Analyysi on tarkka ja oikea
* Datan visualisointi tukee analyysia
* Tulosten tulkinta ja johtopäätökset ovat perusteellisia ja pohjautuvat analyysiin.

Lisäksi pisteitä saa seuraavista:

* Korrelaation tutkiminen oikein (0–3)
* Jonkin tilastollisen testin käyttö oikein sisältäen testin tulkinnan ja johtopäätökset (0–5).