

Tehtävä 1 – Ongelman kuvailu

Tehtävänä on pohtia millä tavalla data-analytiikka voisi hyödyttää vesiyhtiötä. Voitaneen pitää selvänä, että seuraamalla yksinkertaista aikasarjaa veden kulutuksesta voidaan nähdä esimerkiksi lomakausien ja juhlapyhien vaikutusta vedenkulutukseen. Tällaisella tiedolla on merkitystä esimerkiksi raakavedenoton tarpeen arviointiin. Yksityiskohtaisemmalla tuntitason seurannalla saadaan tutkittua vaikkapa sitä, suosiiko tietty alue peseytymistä aamuisin vai iltaisin. Tuntitason datasta voidaan kätevästi myös tarkkailla kotona vietetyn ajan jakautumista vuorokauteen. Tästä voitaisiin vaikkapa tehdä arvioita siitä, käydäänkö alueella tyypillisesti päivätyössä tai viettääkö väestö viikonloppuisin aikaa kotonaan.

Tehtävässä on ilmeisesti kyseessä pelkästään puhtaasta vedestä, mutta jätevesiseulonalla voidaan seurata erilaisten kemikaalien, lääkeaineiden ja eräiden taudinaiheuttajien esiintymistä väestössä. Tällaisesta aikasarjasta voidaan seurata esimerkiksi jälkiä huumausaineiden ja tiettyjen lääkeainoiden käytöstä. Mikrobiseulonasta saadaan tietoa väestön tautitaakasta.

Datan rikastamisella tarkoitetaan alkuperäisen, itse kerätyn datan yhdistämistä ulkopuoliseen dataan. Tämä voisi tuottaa hyvinkin herkullista materiaalia mitä erilaisimpiin tutkimuksiin. Vertaamalla vedenkulutusta vaikkapa kaupunginosan rakennuskannan ikään voitaisiin saada tietoa vesijohtoverkoston kunnosta ja pyrkiä ennustamaan putkiston kulumisesta aiheutuvaa huoltotarvetta.

CRISP-DM -lähestymistavan onnistunut soveltaminen vaatii täsmällistä ongelmanmäärittelyä, joka on tyypillisesti liiketoimintalähtöinen. Vesihuollon tapauksessa liiketoiminnallista etua voisi ajatella syntyvän vaikkapa seurannasta, jolla pyritään havaitsemaan verkoston vuotoja. Riittävän pitkästä aikasarjasta voi havaita varsin pieniäkin signaaleja, jotka saattavat paljastaa alkavan vuodon ennen kuin putki on laajemmin rikki. Samalla asiakastyytyväisyyden voi ajatella paranevan, kun huoltokatkokset ovat ennakoitavissa.