

Bruttokansantuotteen muutoksista OECD-maissa vuosina 2019-2022

Mikko Vaarala

Perustieteiden korkeakoulu

Kandidaatintyö
Espoo 08.12.2023

Vastuuopettaja

Prof. Pauliina Ilmonen

Työn ohjaaja

Prof. Pauliina Ilmonen

Copyright © 2023 Mikko Vaarala

The document can be stored and made available to the public on the open internet pages of Aalto University.
All other rights are reserved.



Tekijä Mikko Vaarala

Työn nimi Bruttokansantuotteen muutoksista OECD-maissa vuosina 2019-2022

Koulutusohjelma Teknistieteellinen kandidaattiohjelma

Pääaine Matematiikka ja systeemitieteet

Pääaineen koodi SCI3029

Vastuupettaja Prof. Pauliina Ilmonen

Työn ohjaaja Prof. Pauliina Ilmonen

Päivämäärä 08.12.2023

Sivumäärä 25

Kieli Suomi

Tiivistelmä

Vuonna 2019 alkanut koronaviruspandemia ja vuonna 2022 alkanut Venäjän-Ukrainan sota ovat vaikuttaneet maailmantalouteen ja valtioiden taloudellisiin suorituskykyihin. Tämän työn tavoitteena on tutkia klusterianalyysin avulla näiden tapahtumien vaikutuksia OECD-maissa tarkastelemalla bruttokansantuotteen (BKT) neljännesvuosittaisia muutoksia vuosina 2019-2022.

Työssä käytetään hierarkista klusterointia, jossa etäisyysmittarina käytetään euklidista etäisyyttä ja klustereiden yhdistämiskriteerinä Wardin kriteeriä. Hierarkisen klusteroinnin tuloksena saadaan dendrogrammi, joka havainnollistaa klusteroitujen maiden välisiä suhteita ja hierarkista rakennetta. Tämän lisäksi klusterointi toistetaan valitsemalla klustereiden lukumääräksi kaksi, kolme ja neljä klusteria. Muodostuneet klusterit visualisoidaan aikasarjaan, helpottaen muodostuneiden klustereiden tulkintaa. Muodostuneiden klustereiden maantieteellistä sijaintia tutkitaan visualisoimalla klusterit maailmankartalle.

OECD-maista muodostettua kahta klusteria verrataan EuroMOMO:n standardeituihin ylikuolleisuuslukuihin. Erityisesti klusterissa, jonka maille ominaista oli BKT:n merkittävä lasku vuonna 2020, standardoitu ylikuolleisuus oli korkeampi kuin toisen klusterin, jossa BKT pysyi vakaampana. Tämän lisäksi työssä hyödynnetään Oxford Coronavirus Government Response Tracker (OxCGRT) -hankkeen kehittämää tiukkuusindeksiä klusterirakenteiden tulkitsemisessa. Tuloksena havaittiin, että maat jotka kuuluivat klusteriin, jossa oli korkeammat kuolleisuusluvut, myös toteuttivat tiukempia rajoitustoimenpiteitä koronaviruspandemian aikana.

Työn tulokset viittaavat mahdollisiin korrelaatioihin BKT:n muutoksien, koronastrategioiden ja kuolleisuuden välillä. Työn tulokset ovat kuitenkin suuntaa antavia, sillä työn tuloksia tulkitaan pitkälti kuvailevan analyysin avulla.

Avainsanat Klusterianalyysi, Hierarkinen klusterointi, OECD, BKT, Tiukkuusindeksi, Ylikuolleisuus

Author Mikko Vaarala

Title On GDP changes in OECD countries during 2019-2022

Degree programme Bachelor's Programme in Science and Technology

Major Mathematics and Systems Sciences **Code of major** SCI3029

Teacher in charge Assoc. Prof. **Pauliina Ilmonen**

Advisor Assoc. Prof. **Pauliina Ilmonen**

Date 08.12.2023 **Number of pages** 25 **Language** Finnish

Abstract

The coronavirus pandemic that started in 2019 and the Russian-Ukrainian war that started in 2022 have affected the global economy and the economic performance of countries. The aim of this thesis is to examine the impact of these events in OECD countries by examining quarterly changes in gross domestic product (GDP) between 2019 and 2022 using cluster analysis.

The work uses hierarchical clustering with Euclidean distance as the distance measure and Ward's criterion as the criterion for cluster aggregation. The hierarchical clustering results in a dendrogram that illustrates the relationships and hierarchical structures between the clustered countries. In addition, the clustering is repeated by selecting two, three, and four clusters. The clusters formed are visualized in a time series, facilitating the interpretation of the clusters formed. Furthermore, the geographical location of the clusters is studied by visualizing the clusters on a world map.

The two clusters of OECD countries are compared with the EuroMOMO standardized excess mortality rates. In particular, the cluster characterized by a significant fall in GDP in 2020 had a higher standardized excess mortality rate than the second cluster, where GDP remained more stable. In addition, the work makes use of the stringency index developed by the Oxford Coronavirus Government Response Tracker (OxCGRT) project to interpret the cluster structures. As a result, it was found that countries belonging to clusters with higher mortality rates also had strict restrictions during the coronavirus pandemic.

The results suggest possible correlations between changes in GDP, coronavirus pandemic strategies, and mortality rates. However, the results of this thesis are indicative, as the thesis relies heavily on descriptive analysis.

Keywords Cluster Analysis, Hierarchical Clustering, OECD, GDP, Stringency Index, Excess Mortality

Sisällys

Tiivistelmä	3
Tiivistelmä (englanniksi)	4
Sisällys	5
1 Johdanto	6
2 Kriisit	7
2.1 Koronaviruspandemia	7
2.2 Venäjän-Ukrainan sota	8
3 BKT	8
4 Klusterianalyysi	9
4.1 Hierarkinen klusterointi	9
5 Data	11
6 Tulokset	12
7 Yhteenveto	20
8 Liitteet	25
A Maat ja lyhenteet	25

1 Johdanto

Kuluneet vuodet 2019–2022 ovat olleet merkittävän muutoksen aikaa maailmantaloudessa. Vuonna 2019 koronaviruspandemia ja sen tuomat terveydelliset, sosiaaliset ja taloudelliset kriisit ovat vaikuttaneet merkittävästi valtioiden taloudelliseen suorituskyykyyn. Vuonna 2022 kiristyneet geopoliittiset tilanteet, erityisesti Venäjän tilanne, ovat tuoneet lisää haasteita ja epävarmuustekijöitä maailmantalouteen. Näinä vuosina OECD-maat (Organisation for Economic Co-operation and Development) ovat reagoineet näihin kriiseihin hyvin erilaisilla tavoilla ja strategioilla. Erilaiset strategiat koronapandemian hallinnassa ja riippuvuus Venäjästä ovat vaikuttaneet OECD-maiden taloudelliseen suorituskyykyyn. Kysymys onkin, miksi toiset maat pärjäsivät taloudellisesti kriisien aikana paremmin kuin toiset, ja minkälaiset strategiat tämän mahdollistivat?

Tässä tutkielmassa tarkastellaan OECD-maiden taloudellista suorituskysyä vuosien 2019–2022 aikana keskittyen bruttokansantuotteen (BKT) muutoksiin. Tutkielman tavoitteena on klusterianalyysin avulla selvittää, voiko OECD-maita luokitella klustereihin niiden taloudellisen suorituskyyvyn perusteella kriisin aikana. Aineisto, johon klusterianalyysiä sovelletaan, koostuu OECD-maiden BKT:n neljännesvuositaisista muutoksista vuosien 2019–2022 aikana.

Koronapandemian vaikutusta maailmantalouteen on tutkittu aikaisemmin, ja esimerkiksi [Fernandes \(2020\)](#) on arvioinut tutkimuksessaan pandemian taloudellisia kustannuksia. Hän ennusti tutkimuksessaan maailmanlaajuisia taantumaa, esittäen sen vakavuuden riippuvan muun muassa siitä, kuinka tehokkaasti pandemian leviäminen onnistutaan estämään. Myös maiden klusterirakenteita on tutkittu aikaisemmin, ja [Aydin ja Yurdakul \(2020\)](#) työssään klusteroi maita useita koronapandemiaan liittyviä muuttujia hyödyntäen. Työssä käytettyjä muuttujia olivat esimerkiksi kuolleisuus, tiukkuusindeksi, köyhyys ja BKT. Ensimmäiseen klusteriin erottuivat maat, joissa oli vähän tartuntoja ja kuolemantapauksia sekä alhainen BKT. Toiseen klusteriin kuuluivat maat, joissa sairaalapaikkojen määrä oli alhainen ja joissa oli korkea köyhyysindeksi. Kolmannessa klusterissa korostuivat maat, joissa oli eniten tartuntatapauksia ja suurin kuolleisuus. Aydin ja Yurdakul kuitenkin toteavat tutkimuksessaan, että sairaalapaikkojen määrän ja BKT:n ei voida suoraan päätellä vaikuttavan tartunta- ja kuolemantapausten määrään. Tässä tutkielmassa BKT:n muutosta peilataan Euroopan ylikuolleisuustilastoihin, etsien mahdollisia yhteneväisyyksiä OECD-maiden välillä.

BKT on laajasti käytetty mittari maiden taloudellisen suorituskyyvyn mittaamiseen. BKT ottaa huomioon maan kaikki tuotetut tavarat ja palvelut tietyllä ajanjaksolla, tarjoten yhden vertailukelpoisen luvun, joka kuvastaa maan taloudellista tuotantoa laajasti.

Klusterianalyysin avulla voidaan tunnistaa OECD-maiden välistä samankaltaisuutta ja eroavaisuutta niiden taloudellisen suorituskyyvyn perusteella. Analyysi antaa mahdollisuuden erottaa maita ryhmiin ja tunnistaa mahdollisia yhteisiä tekijöitä näiden ryhmien sisällä. Tämä auttaa ymmärtämään, mitkä ulkoiset tekijät, kuten maan vastaus koronaviruspandemiaan tai suhde Venäjään, ovat voineet vaikuttaa maan talouteen. Työn tavoitteena on klusterianalyysin avulla tuottaa tietoa siitä,

minkälaiset strategiat ovat olleet suotuisia maiden taloudelle kriisien aikana.

2 Kriisit

2.1 Koronaviruspandemia

Vuonna 2019 maailma kohtasi uudenlaisen haasteen, kun koronavirustauti (COVID-19) alkoi levitä globaalisti. Alun perin Kiinan Wuhanista lähtenyt virus levisi nopeasti vaikuttaen maailmanlaajuiseen kansanterveyteen ja aiheuttaen laajamittaisia häiriöitä ihmisten arkeen. (WHO, 2020a) Vuoden 2020 maaliskuussa koronavirus oli tartuttanut ainakin 118 000 ihmistä 114 maassa, ja viruksen aiheuttama hengitystieinfektio oli vaatinut 4291 ihmishenkeä. Samassa kuussa maailman terveystieteiden WHO julisti COVID-19:n pandemiaksi. (WHO, 2020b; Padhan ja Prabheesh, 2021)

Pandemian vuoksi monet maat alkoivat ottaa käyttöön tiukkoja rajoitustoimenpiteitä yrittäessään estää viruksen leviämistä. Näihin rajoitustoimenpiteisiin lukeutuivat muun muassa matkustuskiellot, erilaiset elinkeinon ja liikkumisvapauden rajoitukset sekä karanteenit. Pandemia ja valtioiden asettamat rajoitustoimenpiteet johtivat nopeasti globaaleihin talouden häiriöihin. Pandemia vaikutti moniin talouden sektoreihin, kuten kaupankäyntiin, matkailuun, kulutukseen ja työllisyyteen. Rajoitustoimenpiteet, tartunnan pelko sekä tulonmenetykset vaikuttivat kulujattajien kulutuskäyttäytymiseen vähentäen tavaroiden ja palveluiden ostamista. (Padhan ja Prabheesh, 2021)

Maailman matkailujärjestön (UNWTO) tekemän raportin mukaan kansainvälisten turistien määrä romahti 74 % vuonna 2020. Tästä aiheutui jopa 1.3 biljoonan dollarin tulonmenetykset matkailusektorille. (UNWTO, 2021) Vuoden 2020 toisessa vuosineljänneksessä maailmanlaajuinen BKT laski yli 4.9 %, ja on arvioitu, että palveluiden ja hyödykkeiden kaupankäynti laski enemmän kuin 2007–2008 finanssikriisin aikana. Kysynnän äkillisen laskun ja epävarmuuden vuoksi yritykset vähensivät investointejaan. Ensimmäisellä vuosineljänneksellä menetettiin 130 miljoonaa työpaikkaa maailmanlaajuisesti ja menetykset kasvoivat 300 miljoonaan toisella vuosineljänneksellä. (Padhan ja Prabheesh, 2021)

Koronarokotteen kehitys käynnistyi nopeasti, ja vuoden 2020 lopulla alkoi ensimmäisten rokotusten antaminen virusta vastaan. Heinäkuuhun 2022 mennessä noin kaksi kolmasosaa maailman väestöstä oli saanut ensimmäisen annoksen koronarokotetta. (Mathieu et al., 2020) Ioannidisin (2022) mukaan väestön rokotekattavuuden kasvaessa useat maat alkoivat purkamaan rajoitustoimenpiteitä vuoden 2021 lopulla. Monet maat kuitenkin jatkoivat rajoitustoimenpiteitä vielä vuonna 2022. Rajoitusten luonne ja tiukkuus vaihtelivat maittain, ja monet hallitukset ottivat käyttöön uusia rajoitustoimenpiteitä pandemian eri vaiheissa. (Ioannidis, 2022)

Maailmanlaajuisesti eri maat purkivat rajoitustoimenpiteitään eri aikoihin, mihin on todennäköisesti vaikuttaneet useat eri tekijät. Mahdollisia syitä tähän voivat esimerkiksi olla pandemian alkuvaiheen rajoitusten tiukkuus ja tehokkuus, rokotusten aloittamisen ajankohta ja rokotuskattavuus sekä maan taloudelliset voimavarat. Köyhissä maissa pandemia saattoi olla erityisen haasteellinen resurssien puutteen vuoksi. Toisaalta maat, jotka onnistuivat hillitsemään viruksen leviämisen tehokkaasti

ja aloittivat ihmisten rokottamisen ajoissa, pystyivät poistamaan rajoitustoimenpiteitä nopeammin.

2.2 Venäjän-Ukrainan sota

Vuonna 2014 alkanut konflikti Venäjän ja Ukrainan välillä kärjistyi helmikuussa 2022 Venäjän hyökätessä Ukraina. Tämän seurauksena länsimaat asettivat Venäjälle laajamittaisia taloudellisia pakotteita tavoitteenaan lamauttaa Venäjän taloutta, ja sitä kautta estää mahdolliset lisähyökkäykset Ukraina.

Sodan ja pakotteiden vaikutukset ovat kuitenkin heijastuneet myöskin maailmantalouteen. Pakotteiden johdosta tärkeiden raaka-aineiden vienti Venäjältä ulkomaille romahti, mikä johti esimerkiksi energian ja elintarvikkeiden hintojen nousuun. Venäjä on toimittanut noin 40% Euroopan maakaasusta ja merkittävän osan öljystä ennen sotaa, minkä vuoksi se on ollut Euroopan energiamarkkinoiden keskeinen tekijä. Maakaasun hinta Euroopassa oli noussut maaliskuussa 2022 kuusinkertaiseksi verrattuna vuoden 2021 tammikuuhun, nostaten huomattavasti energiakustannuksia ja kiihdyttäen inflaatiota. (Mbah ja Wasum, 2022; Wiseman ja Mchugh, 2022)

Venäjän ja Ukrainan sodan taloudelliset seuraukset näkyvät energiatoimitusten häiriöinä, elintarvikkeiden ja energian hintojen nousuna, inflaatiopaineina ja rahoitusmarkkinoiden lisääntyneenä epävakautena. Seuraukset heijastuvat myös selkeästi OECD-maihin, erityisesti Euroopan energiariippuvuuden vuoksi. Tämä tekee OECD-maista alttiita hintojen nousulle, inflaatiolle ja taloudelliselle epävakaudelle.

3 BKT

Bruttokansantuote (BKT) on yksi keskeisimmistä makrotalouden käsitteistä ja se mittaa maan tai alueen kokonaistuotannon arvoa tietyllä aikajaksolla, yleensä vuodessa tai kvartaaleittain. Bruttokansantuotteen avulla voidaan arvioida kansantalouden kokoa ja sen kehitystä ajan mittaan. Bruttokansantuotteen kasvu puolestaan viittaa talouden kasvuun, kun taas lasku viittaa taantumaan tai lamaan. (Callen, 2008)

BKT:n laskemiseen on useita tapoja, mutta yleisin ja tunnetuin tapa on arvonalisäysmenetelmä, joka on myös Suomessa päämenetelmä BKT:n laskennassa. Arvonalisäysmenetelmässä tuotoksesta vähennetään välituotekäyttö, josta saadaan arvonalisäys. Arvonalisäykset lasketaan toimialakohtaisesti ja näiden summasta saadaan bruttokansantuote. Toinen tapa laskea bruttokansantuote on tulomenetelmä. Tulomenetelmässä tarkastellaan tuotannon tuloeria, joita ovat palkat, sosiaaliturvamaksut, voitot ja pääomakorvaukset. Nämä lasketaan yhteen ja kun summasta vähennetään tuotantoon liittyvät verot, saadaan bruttokansantuote. Kolmas tapa laskea bruttokansantuote on lopputuotemenetelmä. Lopputuotteita ovat tuotteet, jotka menevät suoraan yksityiseen tai julkiseen kulutukseen, investoinneiksi tai vientiin ulkomaille. Lopputuotemenetelmässä otetaan kulutusmenojen, investointien ja viennin summa, josta vähennetään tuonti ja tästä saadaan bruttokansantuote. (Tilastokeskus, 2013)

Nimellinen bruttokansantuote on taloudellisen toiminnan mittari, joka mitataan nykyisellä hintatasolla, joka kuvaa kaikkien lopputuotteiden ja palveluiden markkina-

arvoa tiettyinä ajanjaksona ilman inflaation vaikutusta. Reaalinen bruttokansantuote mittaa kansantalouden tuotantoa tiettyinä ajanjaksona, jossa on huomioitu ja korjattu inflaation vaikutukset. ([Investopedia, 2022](#)) Tämä tarkoittaa, että reaalinen BKT mitataan jollain perushinnalla, joka on yleensä vuoden tietty hintataso. Inflaation vaikutuksen korjaus mahdollistaa vertailun eri aikakausien välillä ilman, että hintojen muutokset vaikuttavat tulokseen. Erot nimellisen ja reaalisen BKT:n välillä ovat merkittäviä. Nimellinen BKT antaa kuvan taloudellisesta toiminnasta markkina-arvojen näkökulmasta, mutta se ei ota huomioon hintojen nousua tai laskua. Reaalinen BKT puolestaan kuvaa kansantalouden todellista kasvua tai laskua, koska se ottaa huomioon hintatason muutokset. Tämä on erityisen tärkeää taloudellisen kehityksen seuraamisessa, koska inflaatio voi vaikuttaa talouden kasvulukuihin merkittävästi.

4 Klusterianalyysi

Klusterointi on analyysimenetelmä, jonka tarkoituksena on ryhmitellä joukko objekteja niiden samankaltaisuuksien perusteella. Ryhmittely tapahtuu siten, että samankaltaiset objektit sijoitetaan samaan osajoukkoon eli klusteriin, ja vastaavasti erilaiset objektit sijoitetaan eri klustereihin. Klusterointialgoritmit pyrkivät jakamaan tarkasteltavan aineiston osiin siten, että samankaltaisuus klusterin sisällä olisi mahdollisimman suuri. Klusterointi on tehokas tapa etsiä tarkasteltavasta datasta en-tuudestaan tuntemattomia rakenteita ja piirteitä tehden siitä helpommin tulkittavaa. ([Frades ja Matthiesen, 2010](#))

Erilaisia klusterointimenetelmiä on monia ja ne voidaan jakaa neljään eri pääkategoriaan niiden ominaisuuksien perusteella. Kategorioita ovat osittavat menetelmät, hierarkiset menetelmät, tiheyteen perustuvat menetelmät ja ruudustopohjaiset menetelmät. Jokaisella klusterointimenetelmällä on omat vahvuutensa ja heikkoutensa, ja menetelmän valinta riippuu analysoitavasta aineistosta sekä klusterianalyysin tavoitteista. Osittavat menetelmät olettavat, että klustereiden lukumäärä tiedetään ennalta. Yksi yleisesti käytetty osittava menetelmä on k :n kerkiarvon klusterointi, joka jakaa datan iteratiivisen algoritmin avulla k :hon klusteriin minimoimalla klusterin pisteiden etäisyyttä klusterin keskiarvovektoriin. K :n keskiarvon klusteroinnin etu on tehokkuus, mutta tulokset ovat vahvasti riippuvaisia alustavista parametreista. Hierarkisissa menetelmissä klusterit muodostetaan hierarkisesti, joka tarkoittaa, että klusterit voivat olla osana suurempia klustereita. Tiheyteen perustuvat menetelmät taas pystyvät tunnistamaan klustereita alueilta, joissa datapisteitä on tiheästi. Niiden etu on epäsäännöllisen muotoisten klusterien tunnistaminen. Ruudustopohjaiset menetelmät sopivat hyvin moniulotteiselle datalle. Näissä menetelmissä data jaetaan ruudukkoihin ja klusterointia sovelletaan ruutujen ositukseen. ([Han et al., 2012](#))

4.1 Hierarkinen klusterointi

Tässä työssä analysoitavan aineiston ryhmittelyyn käytetään kokoavaa hierarkista klusterointia. Hierarkinen klusterointi on ohjaamaton menetelmä, mikä tarkoittaa sitä, että tavoitteena ei ole ennustaa tiettyä ennakkoon määritettyä piirrettä, vaan

sen sijaan tavoitteena on löytää aineistosta rakenteita ja tunnistaa ryhmät sen sisältä ilman ulkopuolelta tulevaa ohjausta. Kokoavassa hierarkisessa klusteroinnissa klusterointi aloitetaan siten, että jokainen aineiston objekti on aluksi omassa klusterissaan. Tämän jälkeen klustereita yhdistetään vaiheittain, kunnes saavutetaan ennalta määrätty klustereiden määrä tai jäljellä on yksi suuri klusteri, johon kaikki aineiston objektit kuuluvat. (Dayan et al., 1999; Nielsen, 2016)

Hierarkisen klusteroinnin tuloksia esitetään yleensä graafisesti dendrogrammin avulla, joka havainnollistaa klusteroitujen pisteiden suhteita ja hierarkkista rakennetta. Dendrogrammi on puumainen kaavio, joka havainnollistaa, miten klusterit yhdistyvät klusterointiprosessin aikana. Puun solmukohdat kuvaavat kahden klusterin yhdistymistä kokoavassa hierarkisessa klusteroinnissa, kun taas solmukohtien korkeudet kertovat yhdistyvien klustereiden erilaisuudesta. (Zhao ja Karypis, 2002) Dendrogrammi, jossa jokainen OECD-maa on aluksi omassa klusterissaan, on esitetty Kuvassa 2.

Jotta klusteroidessa voidaan mitata datapisteiden samankaltaisuutta, tulee ensin määrittää käytettävä etäisyysmittari. Etäisyysmittarin valinta riippuu analysoitavasta aineistosta, ja vaikuttaa vahvasti klusteroinnin lopputuloksiin. Yleisiä vaihtoehtoja etäisyysmittarille ovat euklidinen etäisyys, Manhattan-etäisyys ja Minkowskin etäisyys. Samankaltaisuuden mittaamiseen tähän työhön valitaan euklidinen etäisyys, joka on hyvin yleisesti käytetty mittari ja sopii hyvin numeeriselle datalle. (Han et al., 2012)

Olkoot $i = (x_{i1}, x_{i2}, \dots, x_{ip})$ ja $j = (x_{j1}, x_{j2}, \dots, x_{jp})$ kaksi objektiä p-dimensioisessa avaruudessa. Euklidinen etäisyys objektien i ja j välillä on määritelty seuraavasti:

$$d(i, j) = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \dots + (x_{ip} - x_{jp})^2}.$$

Klustereiden yhdistämiskriteerinä käytetään Wardin kriteeriä. Wardin kriteeri on yksi yleisesti käytetyistä yhdistämiskriteereistä ja se pyrkii minimoimaan klusterien välistä erilaisuutta yhdistämällä klustereita siten, että niiden sisäinen varianssi kasvaa mahdollisimman vähän.

Olkoot A ja B klustereita. Määritellään klustereiden välinen etäisyys $\Delta(A, B)$ seuraavasti:

$$\Delta(A, B) = \frac{n_A n_B}{n_A + n_B} d(c(A), c(B))^2,$$

missä $n_A = |A|$ klusterin A alkioden lukumäärä, $n_B = |B|$ klusterin B alkioden lukumäärä ja $c(A)$ ja $c(B)$ ovat klustereiden A ja B keskipisteet (keskiarvovektorit).

Wardin yhdistämiskriteeri nojaa etäisyyden $\Delta(A, B)$ minimoimiseen. Klusteroinnissa jokaisella kierroksella yhdistetään klusterit, joilla etäisyys $\Delta(A, B)$ on pienin. (Nielsen, 2016)

Algoritmi 1 esittää hierarkisen klusterointialgoritmin pseudokoodina.

Algoritmi 1: Kokoava hierarkkinen klusterointi (Nielsen, 2016)

```

1 Alusta jokaiselle dataelementille  $x_i \in X$  klusteri  $G_i = \{x_i\}$  listaan;
2 while Listassa on enemmän kuin kaksi elementtiä do
3   | Valitse  $G_i$  ja  $G_j$  niin, että  $\Delta(G_i, G_j)$  on pienin kaikkien parien joukossa;
4   | Yhdistä  $G_{i,j} = G_i \cup G_j$ ;
5   | Lisää  $G_{i,j}$  listaan;
6   | Poista  $G_i$  ja  $G_j$  listalta;
7 end

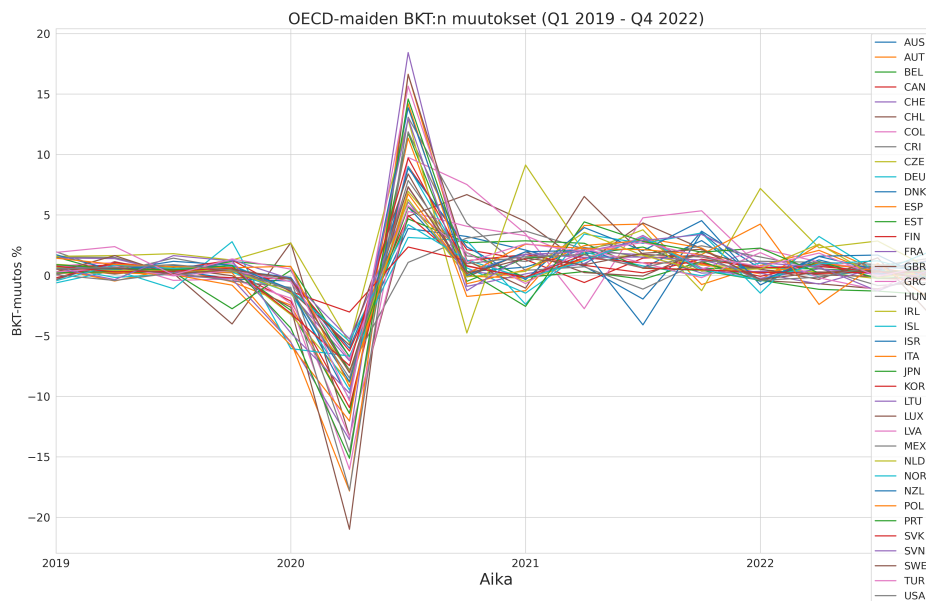
```

5 Data

Tämän työn analyysissä käytetty aineisto on kerätty OECD:n virallisilta verkkosivuilta. "Quaterly GDP" sisältää 38:n OECD maan BKT:n suhteelliset muutokset verrattuna edelliseen vuosineljänneeseen. Tämä indikaattori perustuu lopputuotemenetelmällä laskettuun BKT:hen, ja kyseessä on reaalin BKT, sillä se ottaa huomioon inflaation vaikutuksen. (OECD, 2018) BKT:n erilaiset laskutavat on kuvattu Luvussa 3. Työssä mukana olevat OECD-maat ja niistä käytetyt lyhenteet ovat esitettyinä Taulukossa A. Tutkimuksessa tarkasteltava ajanjakso alkaa vuoden 2019 ensimmäisestä vuosineljänneksestä ja loppuu vuoden 2022 viimeiseen vuosineljänneeseen. Tämän ajanjakson aikana jokaiselle OECD-maalle on saatavilla yhteensä 16 datapistettä, jotka kuvaavat BKT:n muutoksia suhteessa edelliseen vuosineljänneeseen. Valittu ajanjakso tarjoaa kattavan käsityksen OECD-maiden BKT:n kehityksestä koronapandemian ja Ukrainan sodan alun ajalta.

Kuvassa 1 on esitetty visuaalisesti 38:n OECD-maan suhteelliset BKT:n muutokset edelliseen vuosineljänneeseen verrattuna vuoden 2019 alusta vuoden 2022 loppuun. Visuaalinen esitys havainnollistaa hyvin eri kriisien vaikutuksia maiden BKT:hen. Ennen koronapandemian alkua maiden BKT:n muutokset pysyivät hyvin lähellä nollaa tai hieman sen yläpuolella. Kuitenkin vuoden 2020 ensimmäisessä vuosineljänneksessä jokaisen OECD-maan BKT koki merkittävän romahduksen koronapandemian seurauksena. Tämän jälkeen vuoden 2020 toisessa vuosineljänneksessä nähtiin piikki ylöspäin, kun BKT kasvoi suhteessa edellisen vuosineljänneksen romahdukseen. Tämän jälkeen eri OECD-maiden BKT:n muutoksissa on havaittavissa voimakasta vaihtelua sekä positiiviseen että negatiiviseen suuntaan.

Vuoden 2020 ensimmäisen vuosineljänneksen aikana viisi suurinta BKT:n romahdusta kokivat seuraavat maat suuruusjärjestyksessä: Iso-Britannia (-20.99%), Espanja (-17.83%), Meksiko (-17.76%), Kolumbia (-16.04%) ja Portugali (-15.12%). Viisi suurinta BKT:n kasvua toisessa vuosineljänneksessä saavuttanutta maata olivat suuruusjärjestyksessä Ranska (18.43%), Espanja (16.64%), Iso-Britannia (16.60%), Turkki (15.65%) ja Portugali (14.58%).



Kuva 1: Aikasarja OECD-maiden BKT:n muutoksista vuoden 2019 ensimmäisestä vuosineljänneksestä vuoden 2020 viimeiseen vuosineljännekseen.

6 Tulokset

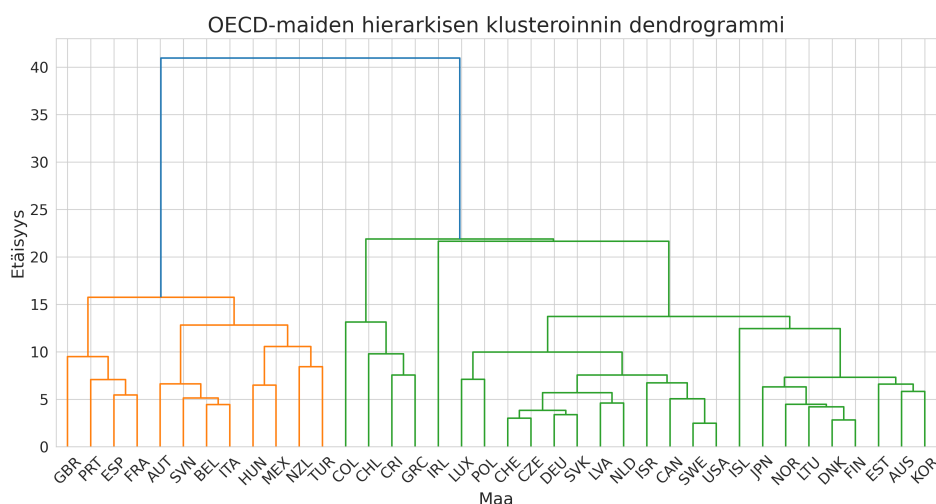
Tässä luvussa esitellään työssä tehdyn klusterianalyysin tulokset. Klusterianalyysia sovellettiin aineistoon, joka on esitelty Luvussa 5. Menetelmänä käytettiin kokoavaa hierarkista klusterointia, joka on tarkemmin esitelty Luvussa 4.1. Etäisyysmittarina klusteroidessa käytettiin euklidista etäisyyttä ja yhdistämiskriteerinä Wardin kriteeriä. Klusterointi toteutettiin Python-ohjelmointikielen avulla hyödyntäen *SciPy*- ja *Scikit-learn*-kirjastoja (Virtanen et al., 2020; Pedregosa et al., 2011).

OECD-maiden BKT:n suhteellisia muutoksia ei standardoitu klusteroidessa, sillä työssä haluttiin korostaa maita, joissa kriisit vaikuttivat eniten. Hierarkisen klusteroinnin tuottama dendrogrammi on esitetty Kuvassa 2. Tämän lisäksi samalle aineistolle toistettiin kolme erilaista klusterointiratkaisua, joissa etukäteen määritettiin haluttujen klustereiden lukumäärä. Klusterointi toistettiin klustereiden määrillä $n = 2$, $n = 3$ ja $n = 4$.

Tulokset visualisoitiin alkuperäiseen aikasarjaan, värittäen samaan klusteriin kuuluvat maat klusteria vastaavalla värillä. Visualisointi auttaa klustereiden muodostumisprosessin ja muodostuneiden klustereiden tulkinnassa. Nämä kuvaajat ovat esitettynä Kuvissa 3, 4 ja 5.

Geopandas-kirjaston (Jordahl et al., 2020) avulla klustereista muodostettiin lisäksi maailmankartta, jonka avulla voidaan helposti tulkita muodostuneiden klustereiden maantieteellistä sijaintia. Maailmankartat klustereiden lukumäärillä 2, 3 ja 4 ovat esitettynä Kuvissa 6, 7 ja 8.

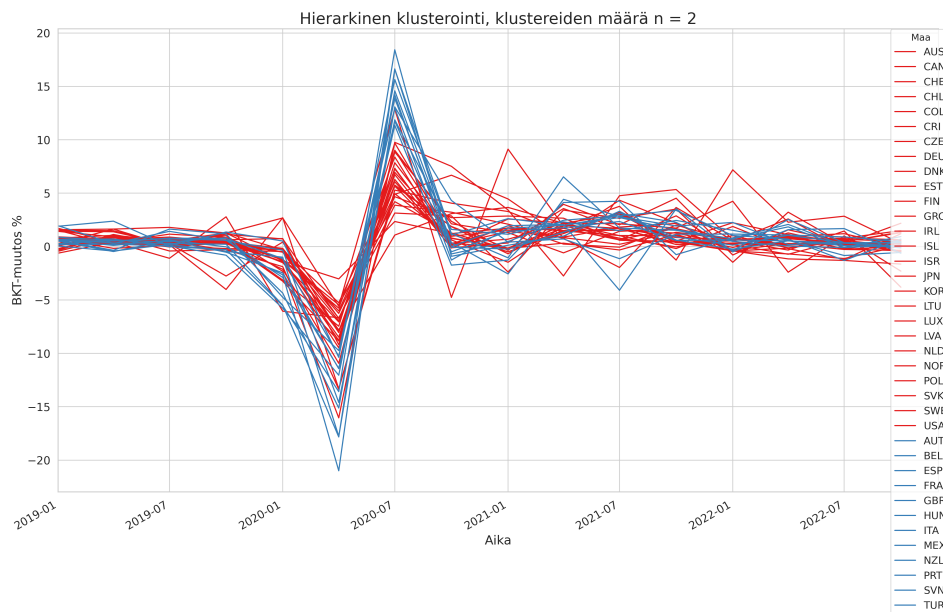
Luvun lopussa muodostuneita klustereita vertaillaan keskenään käyttäen apuna standarsoitua ylikuolleisuutta ja tiukkuusindeksiä. Tavoitteena vertailussa on selvittää korostuuko klusterien maiden keskuudessa suurempi ylikuolleisuus kuin toisissa, sekä ovatko klusterin maat ottaneet käyttöön tiukempia rajoitustoimenpiteitä kuin toisen klusterin maat.



Kuva 2: Hierarkkisen klusteroinnin avulla tuotettu dendrogrammi OECD-maista. Vaaka-akselilla on esitetty maasta käytetty lyhenne ja pystyakselilla muodostuneiden klustereiden välinen etäisyys.

Kuvassa 2 on esitetty kokoavalla hierarkisella klusteroinnilla tuotettu OECD-maiden dendrogrammi. Vaaka-akselilla on esitetty maasta käytetty lyhenne, ja pystyakselin korkeus kuvaa yhdistyneiden klustereiden erilaisuutta. Aluksi jokainen maa on omassa klusterissaan, kunnes ne yhdistyvät kuljettaessa dendrogrammia ylöspäin. Dendrogrammi voidaan katkaista, jos halutaan tunnistaa jokin tietty määrä klustereita. Esimerkiksi, jos katkaisu tehdään korkeudella ≈ 23 , dendrogrammista erottuu selkeästi kaksi toisistaan eroavaa klusteria: oranssi ja vihreä. Jos klustereiden välistä sallittua etäisyyttä pienennetään, huomataan, että Kreikka ja Etelä-Amerikan maat Kolumbia, Chile ja Costa Rica klusteroituvat kolmanteen klusteriin pienellä erolla. Tärkeää on myös huomioida se, kuinka erilainen Irlanti on muihin maihin verrattuna, sillä se klusteroituu yksin neljälteen klusteriin. Tämän voi huomata Irlannin solmukohdan suuresta korkeudesta.

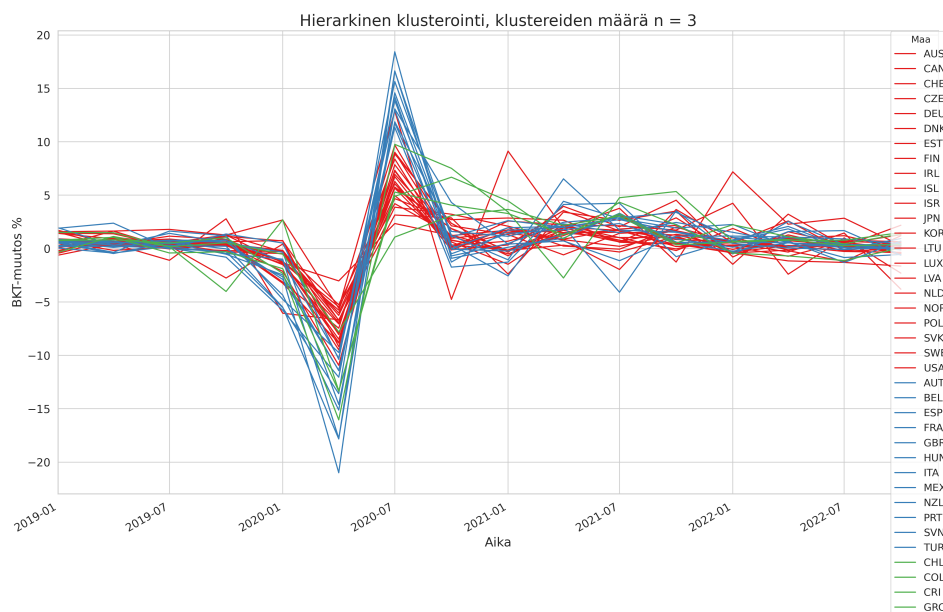
Kuvassa 3 on visualisoituna OECD-maat alkuperäisessä aikasarjassa klusteroituna kahteen klusteriin. Klustereissa erottuvat selkeästi siniset maat, joiden BKT tippui eniten vuoden 2020 ensimmäisessä vuosineljänneksessä. Tämän lisäksi sinisten maiden BKT:n heilahtelu on pienempää vuoden 2020 toisen vuosineljänneksen jälkeen. Punaisten maiden BKT:n muutos oli puolestaan pienempää kuin sinisten ensimmäisessä ja toisessa vuosineljänneksessä vuonna 2020. Punaissa klusterissa kuitenkin korostuvat ne maat, joiden BKT:n muutos jää heilahtelemaan voimakkaasti positiiviseen ja negatiiviseen suuntaan koronapandemian aikana.



Kuva 3: Aikasarja BKT:n muutoksista, jossa havainnollistetaan muodostuneita klustereita, kun OECD-maat klusteroitiin kahteen klusteriin.

Kuvassa 4 on esitetty OECD-maat alkuperäisessä aikasarjassa klusteroituna kolmeen klusteriin. Kuten Kuvasta 2 huomattiin, Kreikka ja Etelä-Amerikan maat klusteroituvat kolmanteen klusteriin. Näiden maiden BKT:n muutos vuoden 2020 ensimmäisessä vuosineljänneksessä on pääosin ollut hillitympää kuin sinisten maiden, mutta suurempaa kuin punaisten. Vuoden 2020 toisessa vuosineljänneksessä kaikkien siniseen klusteriin kuuluvien maiden BKT:t ovat kasvaneet enemmän kuin vihreään klusteriin kuuluvien maiden BKT:t. Tämä voi selittyä osittain sillä, että sinisten maiden BKT:t romahtivat enemmän ensimmäisessä vuosineljänneksessä, joten taloudellinen elpyminen on voinut olla voimakkaampaa toisessa vuosineljänneksessä.

Kuvassa 5 on esitetty OECD-maat alkuperäisessä aikasarjassa klusteroituna neljään klusteriin. Erona kahden ja kolmen klusterin klusterointiin on ainoastaan se, että Irlanti klusteroituu yksin neljänteen klusteriin. Kuten myös Kuvan 2 dendrogrammista huomattiin, on Irlannin BKT:n vaihtelu ollut hyvin erilainen muihin maihin verrattuna. Irlannin BKT romahti vuoden 2020 ensimmäisessä vuosineljänneksessä 5.59 %, kun taas toisessa vuosineljänneksessä se kasvoi 12.88 %. Siniseen ja punaisiin klusteriin kuuluvien maiden BKT:n kasvu toisessa vuosineljänneksessä taas on ollut samaa luokkaa kuin ensimmäisen vuosineljänneksen romahdus. Tämän lisäksi Irlannin BKT:n muutos jää kaikista maista eniten heilahtelemaan koronapandemian aikana.

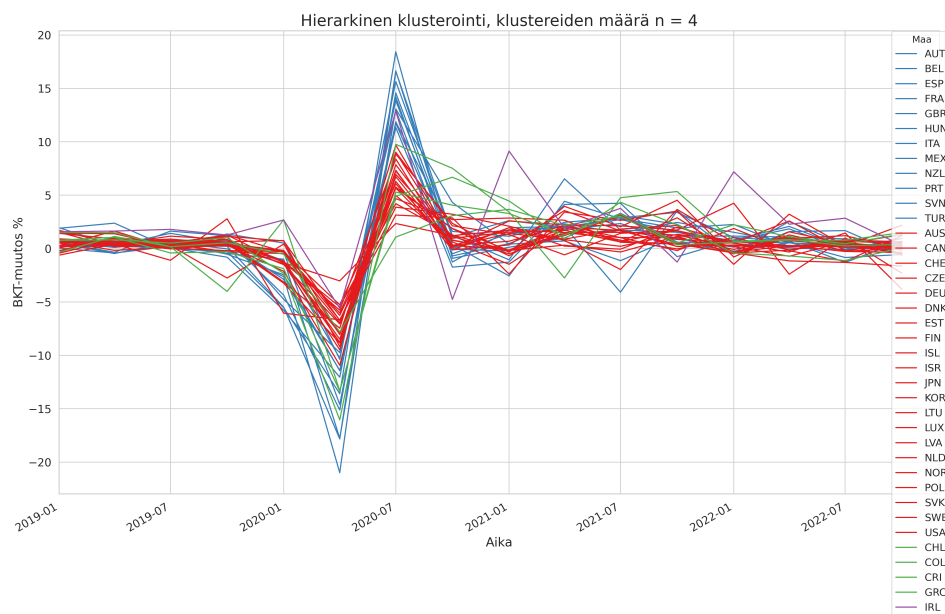


Kuva 4: Aikasarja BKT:n muutoksista, jossa havainnollistetaan muodostuneita klustereita, kun OECD-maat klusteroitiin kolmeen klusteriin.

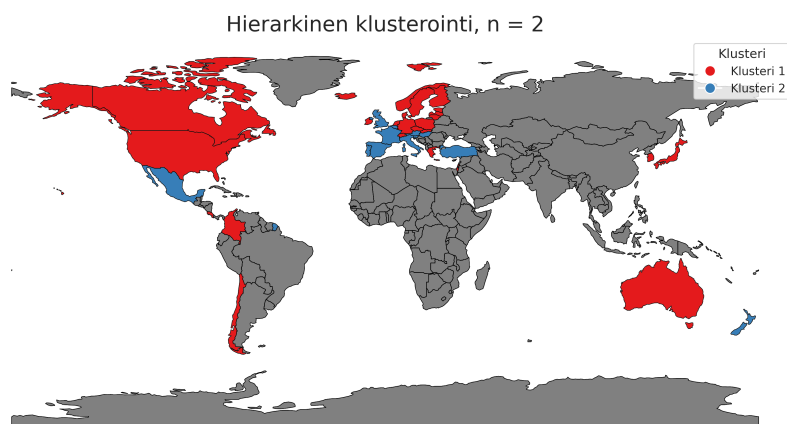
Kuvassa 6 on visualisoituna Kuvan 3 kaksi klusteria maailmankartalla. Samaan klusteriin kuuluvat maat on väritetty klusteria vastaavalla värillä. Harmaalla väritetyt maat eivät ole mukana tarkastelussa. Maantieteellisesti punaiseen klusteriin erottuvat Pohjois-Amerikan maista Yhdysvallat ja Kanada, Keski- ja Etelä-Amerikasta Costa Rica, Columbia ja Chile, Aasiasta Japani ja Korea sekä Australia. Pääosin Euroopassa Pohjois- ja Keski-Euroopan maat kuuluvat punaiseen klusteriin, kun taas Etelä-Euroopan maat siniseen klusteriin.

Kuvassa 7 on visualisoituna Kuvan 4 kolme klusteria maailmankartalla. Kolmella klusterilla alkuperäisen punaisen klusterin maat Kolumbia, Costa Rica, Chile ja Kreikka klusteroituvat kolmanteen vihreään klusteriin. Kuten Kuvasta 2 huomataan, sinisen klusterin maat klusteroituvat erilleen vihreän klusterin maista hyvin pienellä erolla. Maantieteellisesti kolme vihreän klusterin maista sijaitsee Keski- ja Etelä-Amerikassa, ja yksi maa Euroopassa. Muiden maiden osilta punaisessa ja sinisessä klustereissa klusterointi pysyy samana kuin Kuvassa 6.

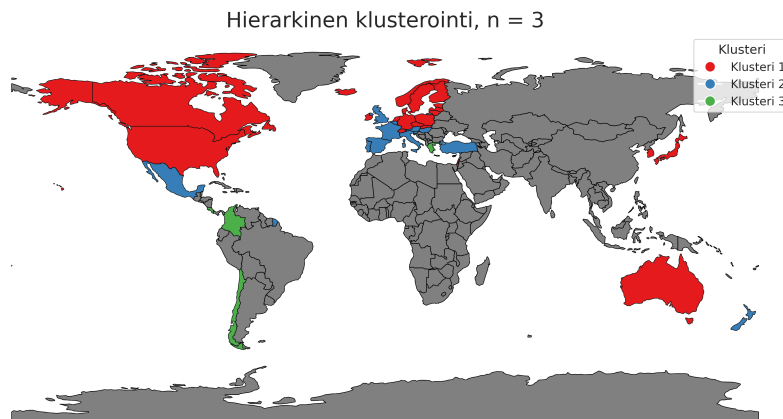
Kuvassa 8 on visualisoituna Kuvan 5 neljä klusteria maailmankartalla. Vastaavasti, kuten kolmen klusterin tapauksessa, sininen ja vihreä klusteri pysyvät ennallaan, kun taas Irlanti siirtyy punaisesta klusterista omaan klusteriinsa. Kuten Kuvasta 5 huomattiin, on Irlannin BKT:n muutos ollut luonteeltaan hyvin erilaista kuin muiden maiden.



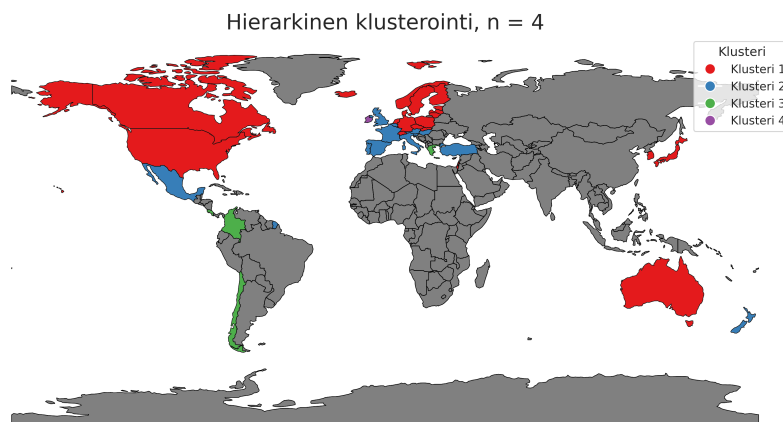
Kuva 5: Aikasarja BKT:n muutoksista, jossa havainnollistetaan muodostuneita klustereita, kun OECD-maat klusteroitiin neljään klusteriin.



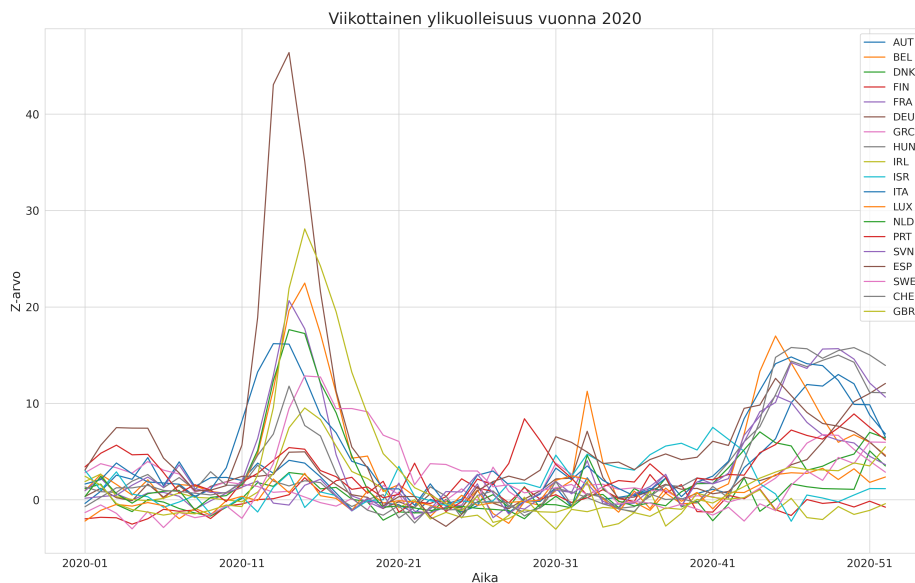
Kuva 6: Muodostuneita klustereita havainnollistava maailmankartta, jossa OECD-maat on klusteroituna kahteen klusteriin. Harmaat maat eivät ole mukana tarkastelussa.



Kuva 7: Muodostuneita klustereita havainnollistava maailmankartta, jossa OECD-maat on klusteroituna kolmeen klusteriin. Harmaat maat eivät ole mukana tarkastelussa.



Kuva 8: Muodostuneita klustereita havainnollistava maailmankartta, jossa OECD-maat on klusteroituna neljään klusteriin. Harmaat maat eivät ole mukana tarkastelussa.



Kuva 9: Ylikuolleisuuden aikasarja vuonna 2020 19:lle työssä mukana olevalle maalle. Vaaka-akselilla on esitettynä aika, ja pystyakseli kuvaa standardoitua ylikuolleisuutta (z-arvo). (EuroMOMO, 2020)

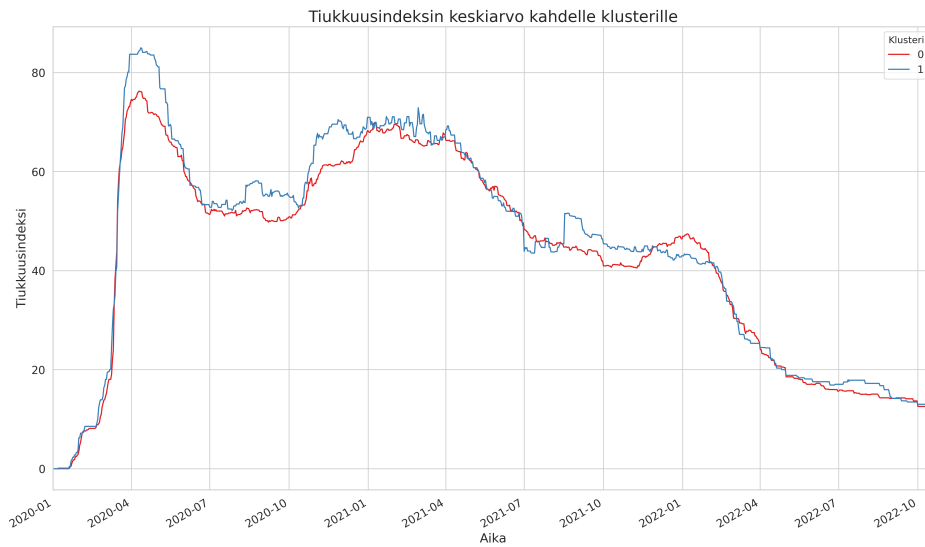
Kuvassa 9 on esitetty 19:lle työssä mukana olevalle maalle vuoden 2020 standardoitu ylikuolleisuus viikottain. Aineisto on kerätty EuroMOMO:n virallisilta verkkosivuilta (EuroMOMO, 2020). Ylikuolleisuus on havaitun kuolleisuuden ja odotusarvoisen peruskuolleisuuden erotuksena saatu luku. Ylikuolleisuus standardoidaan z-arvoksi jakamalla erotuksena saatu luku peruskuolleisuuden keskihajonnalla. Kuvaajasta voidaan havaita koronapandemian eri aaltojen aiheuttamat piikit ylikuolleisuudessa vuoden 2020 aikana. Suurimmat ylikuolleisuusluvut ovat maaliskuussa 2020, mikä näkyy kuvaajassa jyrkkänä piikkinä.

Taulukossa 1 on esitettyinä maittain standardoidun ylikuolleisuuden keskiarvo vuonna 2020 suurimmasta pienimpään. Taulukossa maiden nimet väritettiin niitä vastaavan klusterin värillä, kun aineisto klusteroitiin kahteen klusteriin. Kuten taulukosta huomataan, on ylikuolleisuuden keskiarvo ollut suurin siniseen klusteriin kuuluvissa maissa, joissa myös BKT romahti voimakkaimmin. Alhaisin ylikuolleisuuden keskiarvo taas puolestaan on ollut punaiseen klusteriin kuuluvissa maissa, joissa BKT:n muutos oli hillitympää vuonna 2020.

Taulukko 1: Ylikuolleisuuden keskiarvo maittäin vuonna 2020. Taulukossa on korostettuna maiden klusterit väreittäin.

Maa	Ylikuolleisuuden keskiarvo (z-arvo)
ESP	7.286731
ITA	4.994615
BEL	4.210769
CHE	3.329808
PRT	3.253462
FRA	3.091731
HUN	2.976923
SVN	2.976731
GBR	2.809615
NLD	2.659038
AUT	2.539615
DEU	2.147500
SWE	1.996923
GRC	1.675577
ISR	1.536154
LUX	0.451346
DNK	0.300385
IRL	0.294615
FIN	-0.030192

Oxford Coronavirus Government Response Tracker (OxCGRT) - hanke tarjoaa keinon vertailla eri valtioiden asettamien rajoitusten tiukkuutta koronaviruspandemian aikana (Mathieu et al., 2020). Hankkeessa muodostettu tiukkuusindeksi on yhdeksän erilaisen rajoitustoimenpiteitä kuvaavan mittarin keskiarvo päivätasolla. Tiukkuusindeksi saa arvoja 0:n ja 100:n väliltä, jossa korkeampi arvo tarkoittaa tiukempia rajoitustoimenpiteitä. Kuvassa 10 on esitetty tiukkuusindeksin keskiarvo klustereittain vuosien 2020 ja 2022 välillä. Nämä klusterit vastaavat tapausta, jossa aineisto klusteroitiin kahteen klusteriin. Kuvaajasta voidaan havaita, että sinisen klusterin maiden rajoitustoimenpiteet ovat pääsääntöisesti olleet pandemian aikana huomattavasti tiukempia kuin punaisen klusterin maiden rajoitustoimenpiteet. Molempien klusterien maat ovat reagoineet pandemian alussa yhtä voimakkaasti, mutta sinisen klusterin maat ovat joutuneet ottamaan käyttöön tiukempia rajoitustoimenpiteitä pian pandemian alkamisen jälkeen.



Kuva 10: Aikasarja, joka tiukkuusindeksin avulla havainnollistaa rajoitustoimenpiteiden tiukkuutta klustereittain vuosien 2020 ja 2022 välillä. (Mathieu et al., 2020)

7 Yhteenveto

Kappaleessa 6 esiteltiin kokoavan hierarkisen klusteroinnin tulokset, jossa aineistona käytettiin OECD-maiden neljännesvuosittaista BKT:n muutosprosenttia. Klusteroinnista tuloksena luotiin dendrogrammi, sekä klusterointi toteutettiin kahdella, kolmella ja neljällä klusterilla. Tulokset visualisoitiin aikasarjaan, jonka avulla muodostuneiden klustereiden BKT:n käyttäytymistä pystyttiin havainnollistamaan. Tämän lisäksi tulokset esiteltiin maantieteellisestä näkökulmasta maailmankartalla värittäen maat klustereita vastaavalla värillä.

Klustereista korostui esiin erilaisia rakenteita niin BKT:n muutoksissa, kuin maantieteellisessä sijainnissa. Sinisessä klusterissa korostuivat maat, joiden BKT oli romahtanut eniten, kun taas punaiseen klusteriin kuuluivat maat, joiden BKT:n muutos oli hillitympää. Maantieteellisesti punaisen klusterin maat sijoittuivat pääpiirteittäin Pohjois-Amerikkaan sekä Pohjois- ja Keski-Eurooppaan. Sinisen klusterin maat taas puolestaan edustivat Keski- ja Etelä-Amerikan maita sekä Etelä-Euroopan maita. Kun klusterointi toistettiin kolmella klusterilla Keski- ja Etelä-Amerikan maat klusteroituivat omaan klusteriinsa. Neljällä klusterilla muodostui muista maista hyvin erilainen klusteri, johon ainoastaan Irlanti kuului.

Irlannin klusteroitumista omaan klusteriinsa voisi selittää ongelmat BKT:n käytössä Irlannin taloudellisen suorituskyvyn mittarina. Honohan (2021) artikkelissaan kuvaa, kuinka Irlannin BKT-lukuja vääristää maassa toimivien monikansallisten

teknologia- ja lääkeyritysten läsnäolo, joita Irlannin alhainen yritysverotus houkuttelee. Artikkelissa ehdotetaanakin bruttokansantuloa (BKTL) paremmaksi taloudellisen suorituskyvyn mittariksi Irlannissa.

Vertaillen muodostuneiden klustereiden maiden ylikuolleisuutta sinisessä klusterissa korostuivat maat, joissa ylikuolleisuus oli ollut korkeampi kuin punaisen klusterin maissa koronaviruspandemian aikana. Tämän lisäksi sinisen klusterin maat olivat joutuneet ottamaan käyttöön tiukempia rajoitustoimenpiteitä pandemian hallinnassa. Tämä voi kertoa siitä, että koronaviruspandemiaa ei saatu sinisen klusterin maissa yhtä nopeasti ja tehokkaasti hallintaan, mikä on voinut aiheuttaa huomattavaa laskua valtion taloudellisessa suorituskyvyyssä. Tiukkojen rajoitustoimenpiteiden tehokkuus kuolleisuuden hallinnassa pandemian aikana riippui vahvasti niiden oikea-aikaisesta käyttöönotosta. Mikäli rajoitustoimenpiteet otettiin käyttöön tarpeeksi nopeasti, onnistuttiin pandemian leviäminen hillitsemään. Toisaalta, jos pandemiaa ei saatu hallintaan riittävän ajoissa olivat tiukat rajoitustoimenpiteet välttämättömiä kuolleisuuden hillitsemisessä.

Työssä klustereista tunnistettiin erilaisia rakenteita, kun OECD-maiden BKT:n muutoksia tarkasteltiin vuosina 2019-2022 hierarkisen klusteroinnin avulla. Tulevaisuuden tutkimuksissa voitaisiin harkita erilaisten minimointikriteerien käyttöä klusteroinnissa sekä kehittää parempia menetelmiä klusterien lukumäärän määrittämiseksi. Tulosten tulkinta oli visuaalista, ja tulevat tutkimukset voisivat hyötyä tilastollisemmasta lähestymistavasta menetelmien soveltuvuuden arvioimiseksi.

Viitteet

- Nezir Aydin ja Gökhan Yurdakul. Assessing countries' performances against COVID-19 via WSIDEA and machine learning algorithms. *Applied Soft Computing*, 97: 106792, 2020.
- Tim Callen. What is gross domestic product. *Finance & Development*, 45(4):48–49, 2008.
- Peter Dayan, Maneesh Sahani, ja Grégoire Deback. Unsupervised learning. *The MIT encyclopedia of the cognitive sciences*, pages 857–859, 1999.
- EuroMOMO. Euromomo verkkosivut, 2020. Päivitetty 09.10.2023. Viitattu 18.10.2023. Saatavissa: <https://www.euromomo.eu/>.
- Nuno Fernandes. Economic effects of coronavirus outbreak (COVID-19) on the world economy. 2020.
- Itziar Frades ja Rune Matthiesen. Overview on techniques in cluster analysis. *Bioinformatics methods in clinical research*, pages 81–107, 2010.
- Jiawei Han, Micheline Kamber, ja Jian Pei. Data mining concepts and techniques third edition. *University of Illinois at Urbana-Champaign Micheline Kamber Jian Pei Simon Fraser University*, 2012.
- Patrick Honohan. Is Ireland really the most prosperous country in europe? *Economic Letters*, 1, 2021.
- Investopedia. Nominal gross domestic product: Definition and how to calculate, 2022. Päivitetty 28.10.2022. Viitattu 06.08.2023. Saatavissa: <https://www.investopedia.com/terms/n/nominalgdp.asp>.
- John PA Ioannidis. The end of the COVID-19 pandemic. *European journal of clinical investigation*, 52(6):e13782, 2022.
- Kelsey Jordahl, Joris Van den Bossche, Martin Fleischmann, Jacob Wasserman, James McBride, Jeffrey Gerard, Jeff Tratner, Matthew Perry, Adrian Garcia Badaracco, Carson Farmer, Geir Arne Hjelle, Alan D. Snow, Micah Cochran, Sean Gillies, Lucas Culbertson, Matt Bartos, Nick Eubank, maxalbert, Aleksey Bilogur, Sergio Rey, Christopher Ren, Dani Arribas-Bel, Leah Wasser, Levi John Wolf, Martin Journois, Joshua Wilson, Adam Greenhall, Chris Holdgraf, Filipe, ja François Leblanc. geopandas/geopandas: v0.8.1, 2020. <https://doi.org/10.5281/zenodo.3946761>.
- Edouard Mathieu, Hannah Ritchie, Lucas Rodés-Guirao, Cameron Appel, Charlie Giattino, Joe Hasell, Bobbie Macdonald, Saloni Dattani, Diana Beltekian, Esteban Ortiz-Ospina, ja Max Roser. Coronavirus pandemic (COVID-19). *Our World in Data*, 2020. Päivitetty 20.10.2021. Viitattu 20.10.2023. Saatavissa: <https://ourworldindata.org/coronavirus>.

- Ruth Endam Mbah ja Divine Forcha Wasum. Russian-ukraine 2022 war: A review of the economic impact of Russian-Ukraine crisis on the USA, UK, Canada, and Europe. *Advances in Social Sciences Research Journal*, 9(3):144–153, 2022.
- Frank Nielsen. Hierarchical clustering. *Introduction to HPC with MPI for Data Science*, pages 195–211, 2016.
- OECD. Quarterly GDP, 2018. Viitattu 06.08.2023. Saatavissa: <https://www.oecd-ilibrary.org/content/data/b86d1fc8-en>.
- Rakesh Padhan ja KP Prabheesh. The economics of COVID-19 pandemic: A survey. *Economic analysis and policy*, 70:220–237, 2021.
- Fabian. Pedregosa, Gael Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Oliver Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, ja Edouard Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
- Tilastokeskus. Tilastokoulu, 2013. Päivitetty 07.03.2013. Viitattu 06.08.2023. Saatavissa: https://tilastokoulu.stat.fi/verkkokoulu_v2.xql?course_id=tkoulu_ktal&lesson_id=3&page_type=sisalto&subject_id=6.
- UNWTO. COVID-19 and tourism | 2020: A year in review, 2021. Päivitetty 01.01.2021. Viitattu 05.09.2023. Saatavissa: <https://www.unwto.org/covid-19-and-tourism-2020>.
- Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, ja SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020.
- WHO. Pneumonia of unknown cause - China, 2020a. Päivitetty 05.01.2020. Viitattu 06.08.2023. Saatavissa: <https://www.who.int/emergencies/disease-outbreak-news/item/2020-DON229>.
- WHO. WHO director-general’s opening remarks at the media briefing on COVID-19, 2020b. Päivitetty 11.03.2020. Viitattu 06.08.2023. Saatavissa: <https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020>.
- Paul Wiseman ja David Mchugh. Economic dangers from russia’s invasion ripple across globe, 2022. Päivitetty 02.03.2022. Viitattu 06.08.2023. Saatavissa: <https://www.bbc.com/news/health-61444444>.

[//apnews.com/article/russia-ukraine-vladimir-putin-coronavirus-pandemic-business-health-9478a9825c9abfde5f6505bd34b2998c](https://apnews.com/article/russia-ukraine-vladimir-putin-coronavirus-pandemic-business-health-9478a9825c9abfde5f6505bd34b2998c).

Ying Zhao ja George Karypis. Evaluation of hierarchical clustering algorithms for document datasets. Teoksessa *Proceedings of the eleventh international conference on Information and knowledge management*, pages 515–524, 2002.

8 Liitteet

A Maat ja lyhenteet

Taulukko 2: OECD-maat ja niistä käytetyt lyhenteet

Maa	Lyhenne
Australia	AUS
Itävalta	AUT
Belgia	BEL
Kanada	CAN
Sveitsi	CHE
Chile	CHL
Kolumbia	COL
Costa Rica	CRI
Tšekki	CZE
Saksa	DEU
Tanska	DNK
Espanja	ESP
Viro	EST
Suomi	FIN
Ranska	FRA
Iso-Britannia	GBR
Kreikka	GRC
Unkari	HUN
Irlanti	IRL
Islanti	ISL
Israel	ISR
Italia	ITA
Japani	JPN
Etelä-Korea	KOR
Liettua	LTU
Luxemburg	LUX
Latvia	LVA
Meksiko	MEX
Alankomaat	NLD
Norja	NOR
Uusi-Seelanti	NZL
Puola	POL
Portugali	PRT
Slovakia	SVK
Slovenia	SVN
Ruotsi	SWE
Turkki	TUR
Yhdysvallat	USA