

VoteOn

Mielipidedatan kerääminen ja analyysi

Niklas Tuomola
510313
Kvantitatiiviset aineistot
Essee
nihetu@utu.fi

1. Johdanto

”Demokratia on kriisissä” on sanonta, jota kuulee nykyään paljon. Edustajat ja kansa ovat kaukana toisistaan ja päätöksenteko on eriytynyt kansalaisten mielipiteestä. Päätöksenteon seuranta on vaikeaa, minkä takia edustajien vastuu kansalaisia kohtaa heikkenee. Yksittäiset äänestykset, mitkä vaikuttavat kansalaisten elämään jäävät usein unholaan ja vain vahvoja mielipiteitä herättävät äänestykset saavat tarvittavaa mediahuomiota. Ihmisillä ei myöskään ole riittävää ja helppoa mahdollisuutta osallistua omaan elämäänsä koskevaan päätöksentekoon.

VoteOnin lähtökohtana on korjata nämä asiat ja parantaa demokratiaa. Yritys on perustettu Turun yliopiston startup! -kurssilla vuonna 2016. Kehitimme sovelluksen, jonka avulla käyttäjät saavat antaa mielipiteensä eduskunnassa, Euroopan Parlamentissa sekä valituissa kaupunginvaltuustoissa äänestettäviin esityksiin ja näin saavat samalla tietoa kyseessä olevien tahojen päätettävistä asioista. Sovelluksemme kertoo myös käyttäjälle kuka edustaja/valtuutettu/europarlamentaarikko ja mikä puolue on äänestänyt hänen kanssaan eniten samalla lailla. Näin käyttäjä pystyy asettamaan edustajat vastuuseen seuraavissa vaaleissa tai vaikka vaalien välillä.

Sovellus kehittyi ajan saatossa osallistamistyökaluksi erilaisille organisaatioille aina kaupungeista yrityksiin ja lopetimme eduskunnan sekä muiden yksiköiden aktiivisen päivittämisen vuonna 2020. Tällä hetkellä olemme eriyttämässä organisaatioille suunnattua palvelua omaksi sovellukseksi ja avoimeen dataan perustuvaa ”valvontatyökalua” omaksi, ilmaiseksi palveluksi kansalaisille. Tässä tutkielmassa keskityn juuri jälkimmäisen palvelun datan mahdollisuuksiin sekä analyysiin.

Kansalaisille suunnatun sovelluksen aikana keräsimme dataa myös käyttäjiemme poliittisesta käyttäytymisestä. Tässä esseessä keskityn analysoimaan juuri tuota dataa. Selvitän miten yrityksen tulisi kehittää datankeruuta uudessa versiossa ja miten eri tahot voivat sitä mahdollisesti hyödyntää. Teen myös analyysin, jossa

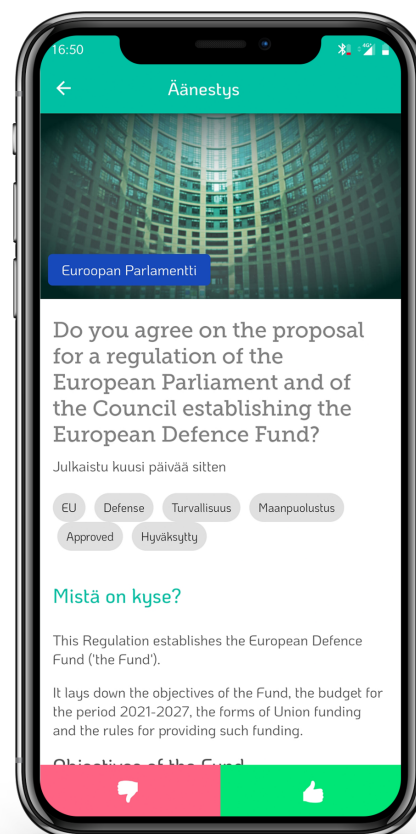
vertailen ristiintaulukoinnin avulla nuorien (18-35 -vuotiaat) ja vanhempien (36-100) suhtautumista yksittäisiin asiakysymyksiin. Teen myös samanlaisen vertailun suurien kaupunkien (Turku, Tampere, Helsinki, Espoo, Vantaa, Oulu) mielipiteistä verrattuna pienempiin paikkakuntiin.

Alkuperäinen, kansalaisille suunnattu VoteOn (silloin nimellä Democratize) -sovellus julkaistiin 9.9.2018. Vuonna 2019 sovellus oli hetken aikaa Play kaupan ladatuimpien sovellusten listalla sekä Apple App Storen uutiset-kategorian kolmas. Palvelu sai myös huomattavaa mediahuomiota ja esiintyi monien lehtien palstoilla. Vuoden 2019 eduskuntavaalien jälkeen palvelusta kehitettiin myös selaimessa toimiva web-sovellus. Kehitystyö toteutettiin yksityisrahoituksen sekä Business Finlandin tuen voimin.

Esseeni ensimmäisessä osassa käsittelen sovellustamme ja kerron mitä kaikkea se mahdollistaa käyttäjille sekä mahdollisista tulevaisuuden suunnitelmista. Tämän jälkeen keskityn sovelluksen tuottamaan dataan ja sen keräämiseen. Kerron, mitä eri muuttujia käytämme sekä millä muuttujilla pystymme käsittelemään yksittäisiä asiakysymyksiä tai asiakysymysnippuja. Osiossa 6 syvennyn palvelun käyttäjiin, analyysin teen osiossa 7 ja johtopäätökset osiossa 8.

2. VoteOn -sovellus

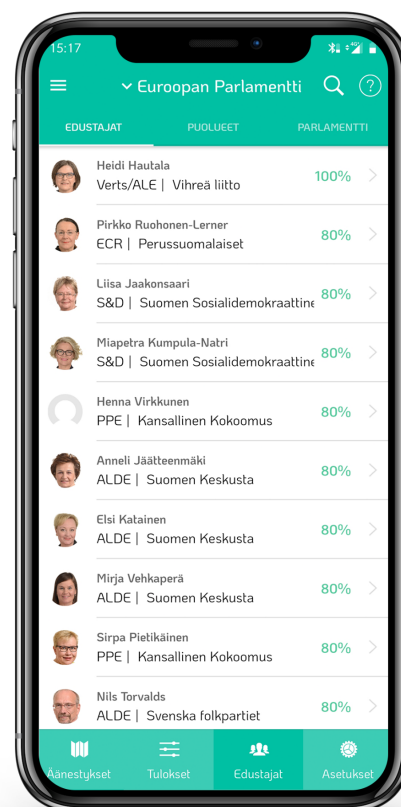
Kuten jo johdannossa toin esille, niin VoteOn on sovellus, joka pyrkii parantamaan demokratiaa. Tarjoamme kansalaisille helpon tavan antaa mielipiteensä eduskunnassa, Euroopan Parlamentissa ja valituissa kaupunginvaltuustoissa tuleviin äänestyksiin. Haemme äänestykset sekä edustajien äänet suoraan avoimen data palveluista.



Tässä esseessä keskityn eduskunnan äänestyksiin ja niistä saatuun käyttäjien vastausdataan. Kysymykset ovat samat, mistä kansanedustajat äänestävät eduskunnassa. Vastausvaihtoehdot ovat lähes samat kuin eduskunnan äänestäessä eli ”Jaa”, ”Ei” ja ”Tyhjä”. ”Poissa” vaihtoehto on jätetty pois, sillä se ei ole käyttäjäkokemuksen ja käyttäjätutkimuksen perusteella tarpeellinen.

Tulevaisuudessa käyttäjä saa valita palvelussa ”suosikkiedustajansa”, jolle lähetämme käyttäjän mielipiteen suoraan (esim. Ville Niinistön asettanut 500 käyttäjää suosikikseen, niin lähetämme hänelle suoraan kootun mielipiteen esim. 20% JAA 80% EI), näin käyttäjän äänelle tulee enemmän merkitystä ja mahdollisuudet vaikuttamiseen lisääntyvät. Samalla edustaja saa mahdollisuuden kysyä kysymyksiä ja kommunikoida vastauksiaan käyttäjien kanssa, jotka ovat valinneet edustajan suosikikseen. Tämä ominaisuus oli toive kansanedustajien puolelta.

Eri tahojen äänestyksen jälkeen käyttäjät saavat heti automaattisesti tietää, ketkä edustajat ja mitkä puolueet ovat äänestäneet hänen kanssaan samalla lailla. Äänestysten karttuessa sovellus pystyy muodostamaan käyttäjälle parhaat edustajat ja puolueet vaalipiireittäin. Ilmoitamme sopivuuden prosenteissa, jotka ilmestyvät parhaimmuusjärjestyksessä käyttäjän ”omalle sivulle” hänen ”suosikkiedustajansa” ollessa ensimmäisenä.



Tarjoamme myös graafisen esityksen äänestettävä olevista kysymyksistä. Kaikkien käyttäjien jakauman käyttäjä näkee annettuaan mielipiteensä ”donitsi” muotoisena ja kun eduskunta on äänestänyt asiasta, niin käyttäjät näkevät myös

eduskunnan äänijakauman, jota he voivat helposti verrata käyttäjien äänijakaumaan.

Erilaiset organisaatiot, kuten kaupungit, yritykset, puolueet ja järjestöt voivat taas hyödyntää palvelumme osallistamislustaa, joka mahdollistaa sovelluksen käyttäjien osallistamisen omaan toimintaansa. Olemme liiketoiminnallisesti keskittyneet tämän alustan kehittämiseen sekä osallistamisen automatisointiin. Tällä hetkellä palvelemme alustan avulla asiakkaita eri sektoreilla. En kuitenkaan tarkastele osallistumislustaa sen tarkemmin, sillä se ei kuulu tämän tutkimuksen aiheisiin, enkä hyödynnä alustan avulla saatavaa dataa.

3. Datan säilytys ja muuttujat

Kuten johdannossa esitin, niin sovellus kerää dataa käyttäjien poliittisista mielipiteistä ja osallistumisen halusta yksittäisiin asiakysymyksiin. Data voidaan jaotella tiettyjen parametrien mukaan, joita kysymme käyttäjiä rekisteröityessä.

Pakolliset tiedot ovat:

- syntymävuosi
- asuinpaikka
- sähköposti (tallenneta dataan)
- salasana (ei tallenneta dataan)

Vapaaehtoinen tieto on:

- sukupuoli: mies, nainen, muu, en halua sanoa

Tiettyjen asiakkaiden käyttäjiltä kysymme myös lisätietoja kuten:

- postinumero
- kaupungin palveluiden käyttäminen
- osasto
- jaosto

(asiakkaiden omia parametreja ei tallenneta tässä tutkimuksessa käytettävään dataan)

Käyttäjien vastausdata sekä käyttäjätiedot tallennetaan Mongo databaseen, josta se voidaan ladata muuhun käyttöön. Koska käsittelemme käyttäjien mielipiteitä, niin panostamme tietoturvaan. Jokainen käyttäjä on meille ainoastaan numerosarja, emmekä yhdistä sähköpostiosoitteita annettuihin äänestyksiin. Alaikäisten käyttäjien datat säilömmme myös eri databaseen, kuin täysi-ikäisten käyttäjien. Asiakkaidemme osallistamisesta syntynyt data säilötään eri paikkaan, kuin avoimen datan yksiköiden vastauksista saatu data. Tämä tarkoittaa, että tässä tutkimuksessa on otantana ainoastaan eduskunnan kysymyksiin perustuva käyttäjien vastauksista koostuva datapaketti.

Vastausdata voidaan tallentaa joko .csv tai .json muodoissa jatkojalostusta varten.

Data sisältää seuraavat tiedot:

- indeksinumero
- käyttäjännumero
- sukupuoli
- ikä
- viimeisin kirjautuminen sovellukseen
- viimeisin kirjautuminen web-sovellukseen
- rekisteröitymispäivämäärä
- asuinpaikka
- äänestysten vastaukset äänestyksittäin

	Index	User	gender	age	last_login	last_login_webapp	regist_	reside_	Vote 1	Vote 2	Vote 3	Vote 4	Vote 5	Vote 6	Vote 7
42	78	User 79	male	23	nan	nan	Fri Oct _	Tampere	nan	jaa	nan	nan	jaa	nan	nan
43	79	User 80	male	46	Wed Oct 16 2019 0_	nan	Fri Oct _	Lieto	jaa	jaa	jaa	jaa	jaa	jaa	ei
44	80	User 81	male	42	Wed Feb 26 2020 1_	nan	Fri Oct _	Turku	nan	nan	nan	nan	nan	nan	nan
45	81	User 82	male	25	nan	nan	Sat Oct _	Tampere	jaa	jaa	jaa	jaa	ei	jaa	jaa
46	82	User 83	male	37	Thu Oct 24 2019 1_	nan	Sat Oct _	Helsinki	nan	nan	nan	nan	nan	nan	nan

Kuva 1. Käyttäjädatta Jupyter Notebookin avulla tarkasteltuna.

4 . Datan kerääminen

Analyysissä käytettävä vastaustdata tulee suoraan VoteOn sovelluksen käyttäjiltä. Tämä tarkoittaa, että vastaajia ei ole karsittu millään tavalla, eikä otanta ole suoraan verrattavissa esimerkiksi Suomen väestörakenteeseen (tarkempi analyysi käyttäjistä **osiossa**). Sovellus on mahdollista ladata sovelluskaupoista ja web-sovellus on avoimena internetissä. VoteOn ei siis pysty tarkistamaan varmaksi käyttäjien "aitoutta" tai tietojen paikkaansapitävyyttä. Molempiin sovelluksiin on sisäänrakennettu "bot" esto eli sovellus tunnistaa tietyissä tapauksissa, mikäli käyttäjä ei ole oikea ihminen. Sovellus estää käyttäjiä äänestämästä kahta kertaa samaan äänestykseen, eli voimme kuitenkin olettaa, että yksi käyttäjä on aina antanut enintään yhden äänen yhteen kysymykseen.

Vahvan varmennuksen puuttuminen tarkoittaa, että sovellusta ei voida käyttää sitovaan päätöksentekoon. Sen tuottama data on ainoastaan suuntaa-antavaa mielipidedataa.

5. Analyysin suunnittelu ja työkalut

Tutkin, miten nuoret ja vanhemmat käyttäjät ovat vastanneet tiettyihin eduskunnan äänestyksiin ja onko mielipiteissä huomattavissa iän mukaan muutosta. Tässä tapauksessa käytössä on siis kaksi eri kategorista muuttujaa, joten tulen käyttämään ristiintaulukointia ja khiin neliö testiä, joista analysoin, ovatko löydökset tilastollisesti merkitseviä ($p < 0,05$). Khiin neliötestille oletuksena on, että on vähintään 5 havaintoa / solu, jonka uskon tässä tapauksessa tulevan täyteen. Solu tarkoittaa tässä siis esimerkiksi nuori + jaa -vastaus kombinaatiota.

Analyysi toteutetaan Python ohjelmointikielellä, joka soveltuu hyvin suurien datamäärien tarkasteluun sekä analysointiin. Käytän apunani myös pandas,

nymphy ja `scipy.stats` -kirjastoja. Piirretyissä kaavioissa käytän apunani Jupyter Notebook -sovellusta sekä `matplotlib` -kirjastoa.

Käytetty Python koodi on tarkasteltavissa erillisenä tiedostona, joka on saavutettavissa GitHub -palvelussa. Linkki koodiin on lähdeluettelossa.

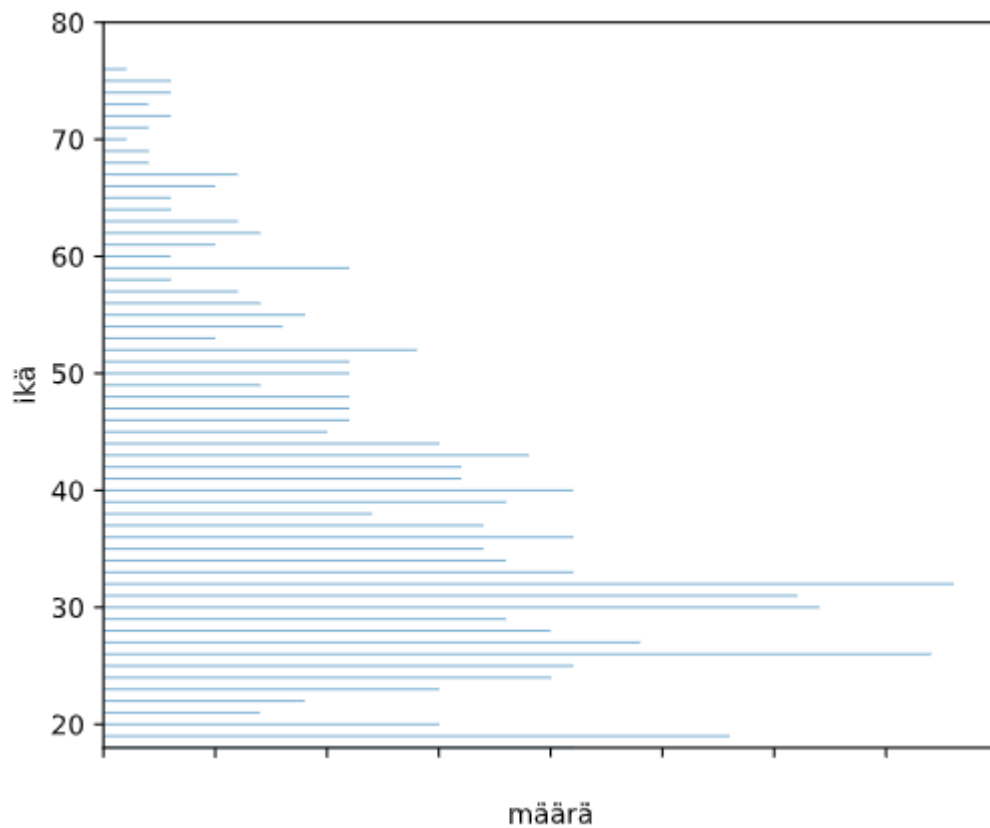
6. Sovelluksen käyttäjät

Pohjustuksena analyysille on tärkeää hieman tarkastella sovelluksen käyttäjiä eli kysymyksiin vastaajia. VoteOn ei tarjoa julkista tietoa käyttäjämääristä, eikä voi antaa käytettyä datasettiä kolmansille osapuolille. Tuon tässä kuitenkin esille käyttäjien eri piirteitä. Työvaiheet tietojen saamiseksi on dokumentoitu täydellisesti GitHubissa tarkasteltavassa koodissa.

Dataa analysoidessa selviää, että palvelun käyttäjien iän keskiarvo on noin 37 vuotta ja mediaani 34. Nuorimmat käyttäjät ovat 18-vuotiaita (alaikäiset eivät ole tässä datasetissä) ja vanhin 76-vuotias. Tilastokeskuksen raportoima suomalaisten keski-ikä on miehillä 41,8 ja naisilla 44,5 (Tilastokeskus 2021). Käyttäjämme ovat siis väestöön nähden nuorempia. Tämä voidaan mahdollisesti selittää sillä, että esimerkiksi Helsingin kaupungin tutkimuksen mukaan älypuhelimien käyttö vähenee, mitä suuremmaksi ikä kasvaa (ikääntyneethelsingissa.fi 2020). Palvelua on myös mainostettu vähäisesti sosiaalisen media kautta, jossa käyttäjäprofiili on Suomen väestörakennetta nuorempi. Tässä kohtaa on hyvä mainita, että kyselytutkimuksemme mukaan alle 18-vuotiaat saattavat rekisteröityessä valita iäkseen 18-vuotta, koska he luulevat, että 18-vuotta täyttäneet saavat palvelun kaikki ominaisuudet käyttöönsä. Tämä voi mahdollisesti selittää 18-vuotiaiden suurta määrää sovelluksessa (Kuva 2 alin palkki).

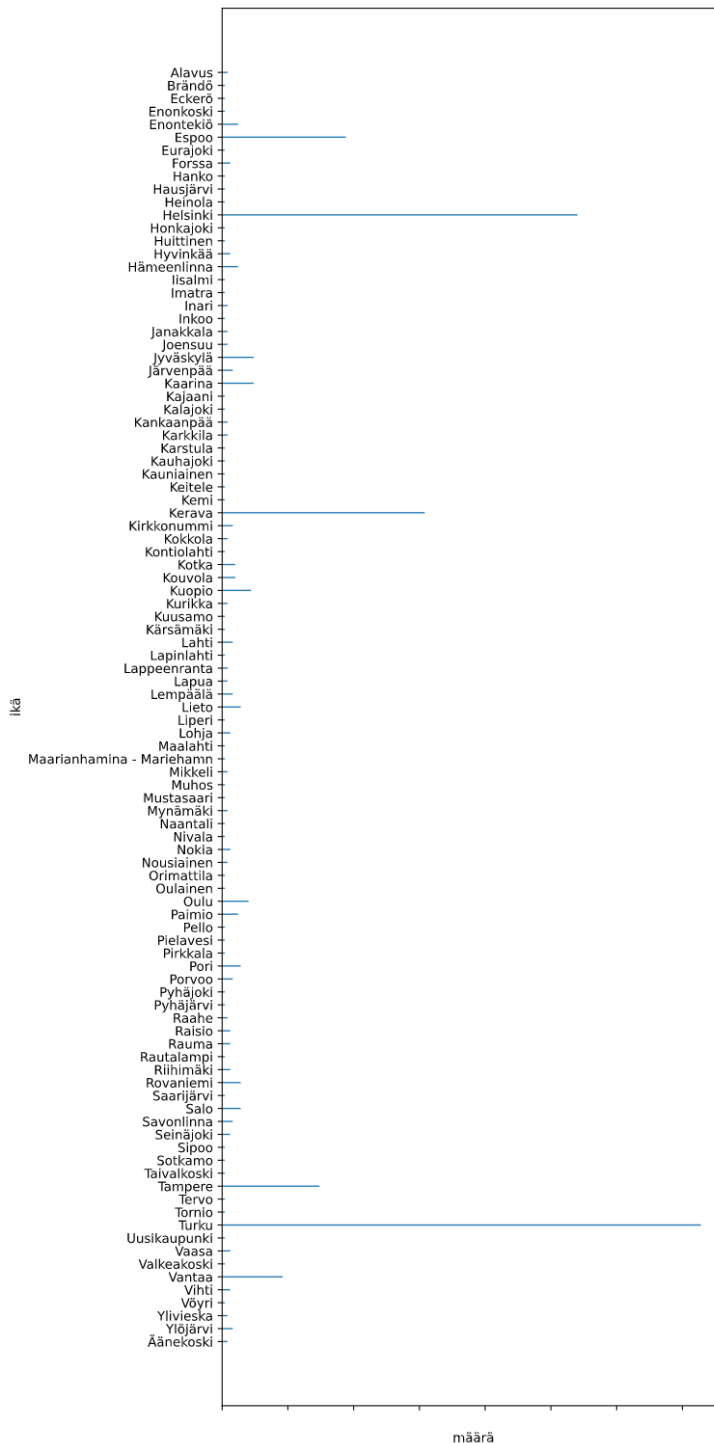
```
mean    37.267930
std      13.331857
min      18.000000
25%      27.000000
```


50% 34.000000
75% 45.000000
max 76.000000



Kuva 2 Käyttäjien ikäjakauma määrän mukaan tarkasteltuna.

Käyttäjiä palvelulla on yhteensä 99 kunnasta, mutta jakauma on hyvin epätasainen. Eniten käyttäjiä on suurista kaupungeista kuten Helsingistä, Espoosta, Tampereelta ja Turusta. Jatkotutkimuksena olisi mahdollista selvittää, kuinka hyvin kaupunkijakauman epätasavertaisuus edustaa Suomen väestön hajaantuneisuutta. Suoralta kädeltä pystytään kuitenkin sanomaan, että Turku on otoksessa yliedustettuna. Tämä voi johtua siitä, että palvelu on turkulainen ja se on saanut Turussa mediahuomioita muuta maata enemmän.



Kuva 3 käyttäjien määrä kaupungeittain.

Analyysissä käyttäjät on siis jaettu “nuoriin” ja “vanhoihin”. “Nuoriin” kuuluvat 18-35-vuotiaat ja “vanhoihin” 36-76-vuotiaat.

Tästä tutkimuksesta on poistettu kaikki ne käyttäjät, jotka eivät ole vastanneet yhteenkään kysymykseen, jotta nämä “haamukäyttäjät” eivät sotkisi analyysiä.

7. Analyysi

Analyysin avulla on siis tarkoitus selvittää, miten nuorempien ja vanhempien käyttäjien mielipiteet eroavat toisistaan poliittisissa kysymyksissä. Nollahypoteesi tähän kysymykseen on, että nuorten ja vanhempien käyttäjien mielipiteet äänestyksissä eriävät toisistaan. Tarkastelen tätä kahden sovelluksesta löytyvän kysymyksen avulla:

1. Välikysymys vanhusten hoidon tilasta. Nauttiiko hallitus luottamustasi?
(Vote 28. Julkaistu 4.2.2019 13:35)
2. Kannatatko hallituksen eroa? (Vote 38. Julkaistu 8.3.2019 11:34)

Molemmat kysymykset ovat Suomen 74. hallituksen eli Sipilän hallituksen kaudelta. Kysymyksessä 1 ”Jaa” vastaus on hallituksen suhteen myönteinen ja kysymyksessä 2 vastainen.

Ensimmäiseen kysymykseen on vastannut 185 käyttäjää poislukien tyhjää äänestäneet. Toiseen kysymykseen on taas vastannut 197 käyttäjää poislukien tyhjää äänestäneet.

Vertailun mahdolliseksi suoritetaan ristiintaulukointi äänestyksittäin sekä ikäryhmittäin (nuoret, vanhat).

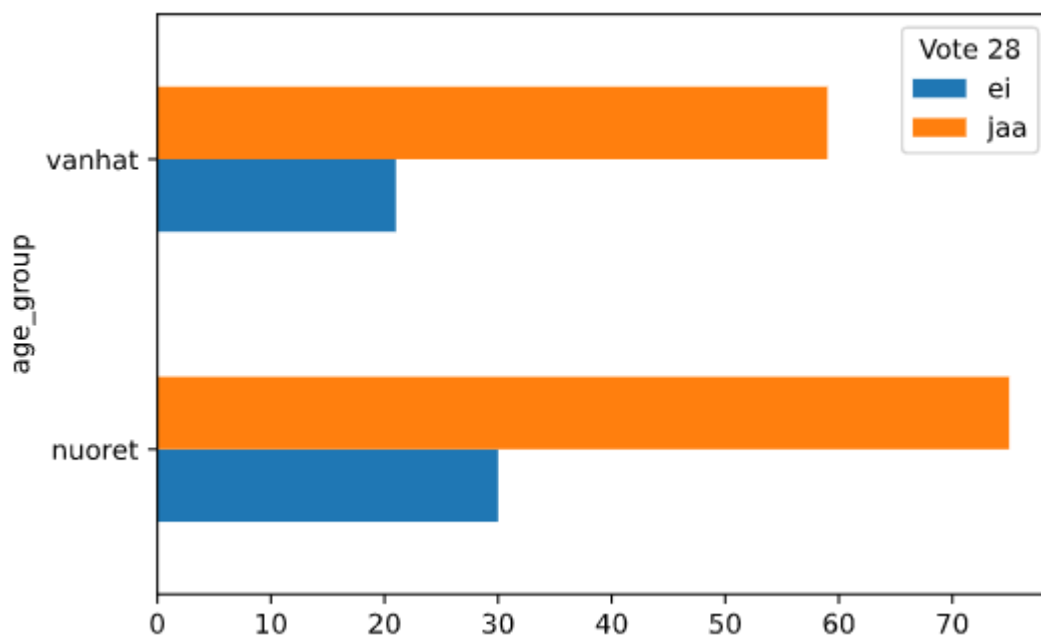
Vote 28	ei	jaa
nuoret	30	75
vanhat	21	59

Taulukko 1. Vote 28 - Välikysymys vanhusten hoidon tilasta. Nauttiiko hallitus luottamustasi? Ristiintaulukointi.

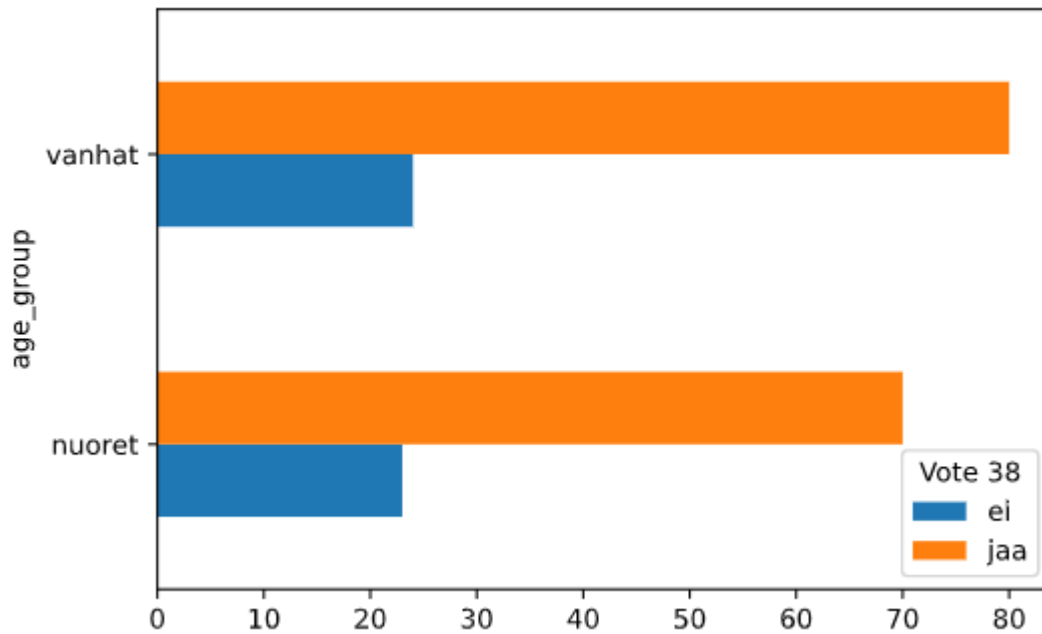
Vote 38	ei	jaa
nuoret	23	70
vanhat	24	80

Taulukko 2. Vote 38 - Kannatatko hallituksen eroa?. Ristiintaulukointi.

Seuraavaksi esitetään ristiintaulukoinnin tulokset myös graafisessa muodossa.



Kuva 4. Vote 28 - Välikysymys vanhusten hoidon tilasta. Nauttiiko hallitus luottamustasi? Ristiintaulukointi.



Kuva 5. Vote 38 - Kannatatko hallituksen eroa?. Ristiintaulukointi.

Kuten voimme ristiintalulukoinnista jo huomata, niin vanhempien ja nuorempien mielipiteet ovat kahdessa asiakysymyksessä hyvin samanlaiset. Ainoa hieman huomattavampi ero on vanhempien käyttäjien suurempi äänestysaktiivisuus kysymyksessä 2. Kysymyksessä 1 taas nuoremmat ovat äänestäneet hieman vanhempia käyttäjiä enemmän. Kysymyksessä 1 80 vanhempaa käyttäjää on antanut äänensä, nuorempia 105. Kysymyksessä 2 taas on äänestänyt 104 vanhempaa käyttäjää ja 93 nuorempaa. Kysymysten ajallinen ero on noin kuukausi eli selittävänä tekijänä tälle ei ole sovelluksen vanhempien käyttäjien määrän kasvu. Jotta nuorten ja vanhempien käyttäjien määrän vaihtelu voisi olla selittävä tekijä, niin äänestysten ajallinen väli tulisi olla pidempi.

Vote 38	ei	jaa
nuoret	25	75
vanhat	23	76

Taulukko 3. Vote 38 - äänien prosentuaalinen osuus

Vote 28	ei	jaa
nuoret	29	71
vanhat	26	74

Taulukko 4. Vote 28 - äänien prosentuaalinen osuus

Äänien prosentuaalinen osuus kertoo samaa tarinaa taulukoiden 1 ja 2 kanssa. Voidaan huomata, että äänet ovat jakautuneet suhteellisen samalla tavalla vanhempien ja nuorempien käyttäjien välillä. Khiin neliötesti tukee myös tätä väitettä kysymyksen 1 p arvo on 0.854 ja kysymyksen 2 on 0.916 eli nuorten ja vanhojen käyttäjien mielipiteillä ei ole tilastollisesti merkitsevää eroa kyseisissä asiakysymyksissä.

8. Johtopäätökset

Kuten analyysistä voidaan huomata, niin nuorten ja vanhojen VoteOn-sovelluksen käyttäjien mielipide-erot tutkituissa asiakysymyksissä ovat hyvin vähäiset. Tulos on hieman yllättävä, sillä esimerkiksi Koivula, Räsänen ja Saarinen (2015) tulevat siihen tulokseen, että erityisesti nuoret opiskelijat ovat äänestäneet muita puolueita enemmän vihreitä sekä vasemmistoliittoa. Puolueet olivat käsiteltävien asiakysymysten aikaan oppositiossa ja näin oltaisiin voitu olettaa, että nuoret olisivat olleet vanhempia äänestäjiä enemmän hallitusta vastaan. Näin ei kuitenkaan ollut, vaan tulokset olivat hyvin samankaltaisia.

Tutkimuksen parantamiseksi olisi hyvä tarkastella kaikki palvelussa olleita äänestyksiä, jotka täyttävät ristiintaulukoinnin ehdot. Muita mielenkiintoisia lisätutkimusaiheita olisi se, miten käyttäjän äänestysaktiivisuus on verrannollinen hänen äänestyksiinsä. Yksi hyvin tärkeä analysointikohde on käyttäjien jakauma puoluepoliittisella kentällä. Mikäli jakauma on hyvin kallellaan oikealle tai vasemmalle, niin se haittaa tulosten verrattavuutta ns. reaali maailmaan. Mielenkiintoista olisi myös tutkia, miten asuinpaikka vaikuttaa

käyttäjän mielipiteisiin. Esimerkiksi on isoissa kaupungeissa asuvien käyttäjien mielipiteillä eroja pienemmissä paikkakunnissa asuviin.

Tutkimus kuitenkin todistaa sen, että on mahdollista rakentaa kaupallinen ja kansalaisia palveleva sovellus, jonka avulla tutkimuksia voidaan ylipäättänsä tehdä. Kehitystyön sekä sisällön päivittäminen on kuitenkin oltava jatkuvaa, jotta käyttäjät vastaavat aktiivisesti kysymyksiin sekä, että palvelu saa lisää käyttäjiä. Tutkimuksen aikan selvisi myös se, että suurin osa palveluun kirjautuneista oli antanut äänensä vähintään kahteen asiakysymykseen. Tämä tarkoittaa, että juuri VoteOnille tai samankaltaiselle palvelulle on tarvetta.

Lähdeluettelo

Äänestykset:

Vote 28 - Välikysymys vanhusten hoidon tilasta. Nauttiiko hallitus luottamustasi?

URL: <https://app.voteon.fi/votes/5c582383dbe23817505b0d95>

Vote 38 - Kannatatko hallituksen eroa?

URL: <https://app.voteon.fi/votes/5c8237be00d1533663318e43>

Analyysin koodi:

<https://github.com/Shamaani/Kvantitatiiviset-aineistot>

Tilastokeskus, Väestörakenne, päivitetty 31.3.2021.

URL: https://www.tilastokeskus.fi/tup/suoluk/suoluk_vaesto.html

Ikaantyneethelsingissä.fi, "Sähköisten palveluiden ja internetin käyttö", 2020.

URL: https://www.ikaantyneethelsingissa.fi/internetin_kaytto

Aki Koivula, Pekka Räsänen & Arttu Saarinen, 2015, "Puolueiden kannatus eri väestöryhmissä ja yhteiskunnallinen vaikutusvalta". *Tieteessä tapahtuu* 5/2015.