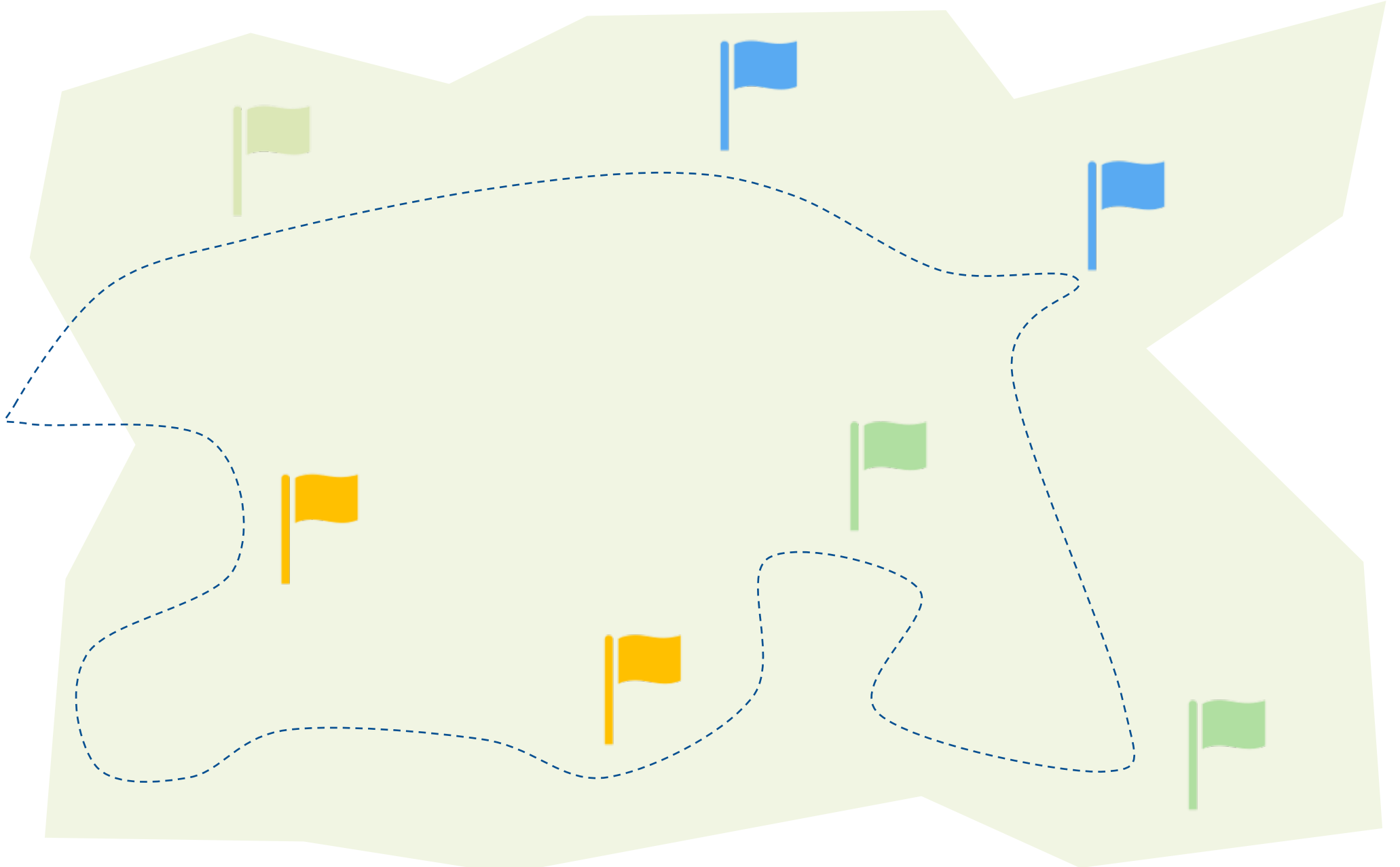




Johdatus data-analytiikkaan

Yhteystiedot

- Mikko Kontio
- Secora Systems Oy / Co-founder / CEO
- mikko.kontio@secora.fi
- www.linkedin.com/in/kontio



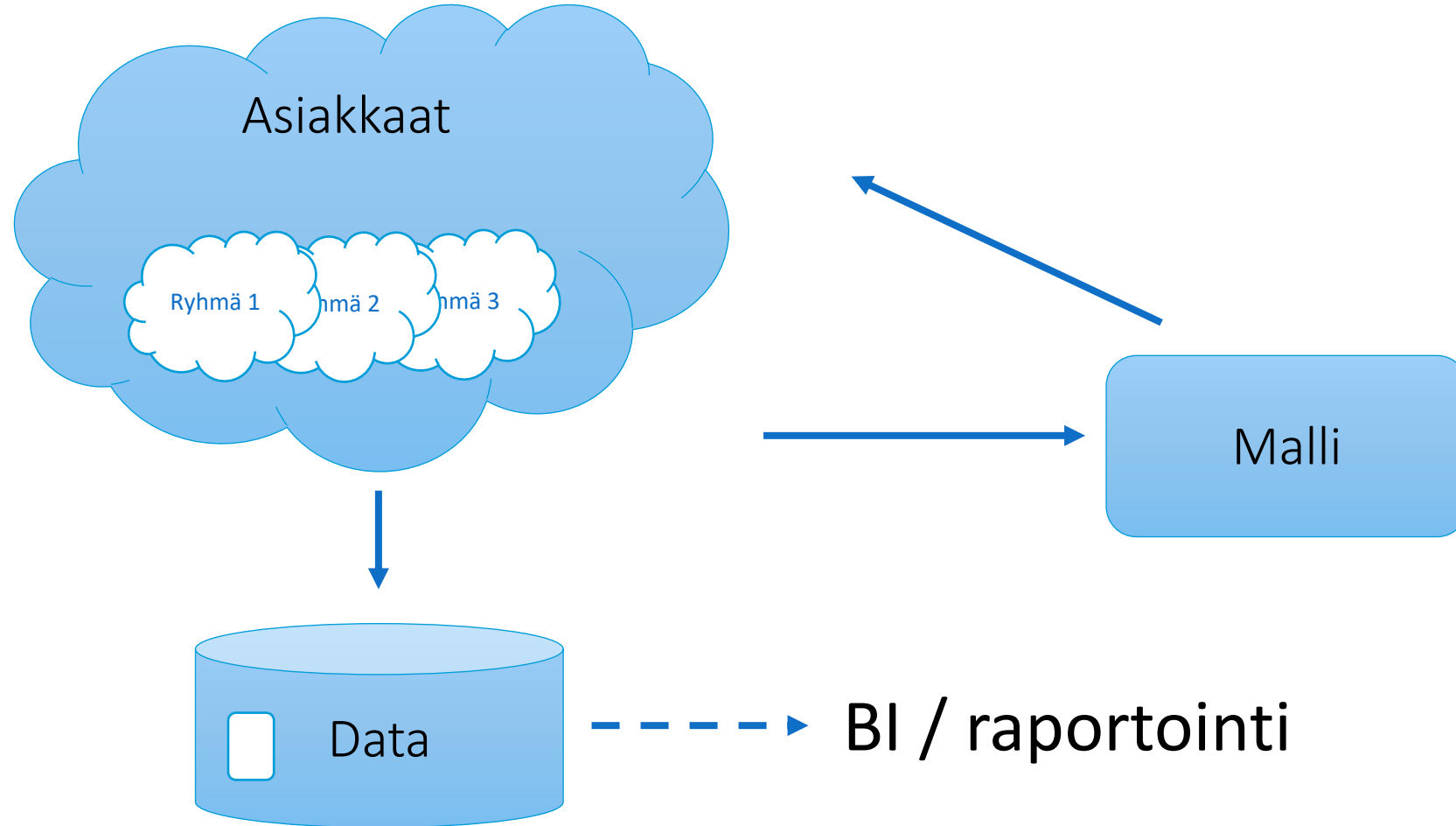


"Target knows before it shows"



<http://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/>

Eli...



Tekoäly voittaisi jo nyt ihmisen oikeassa sodassa – ja sen ymmärtäminen on olennaista, kun Suomi tekee päätöksen historiallisesta hävittäjähankinnasta

Kauppa ennustaa tuhansien työpaikkojen katoavan – "Ei ehkä tarvita samalla tavalla varastotyöntekijöitä, mutta tarvitaan data-analytiikan osaamista"

6.8.2019 04:30 | päivitetty 6.8.2019 08:47

Suomen A-maajoukkuepelaaja on mukana yrityksessä, joka vie data-analytiikkaa nappulakenkään

INFOTILAISUUS: TUNNEANALYYSI JA KASVOJENTUNNISTUS ETÄHOIVAN JA ENNAKOIVAN HOIDON TYÖKALUNA

Liverpoolin salaiset aseet: pallonhallinta ja datatutkimus

Markku Pervilä 25.1.2020 12:12 [DATA JA ANALYTIikka](#) [URHEILU](#) [CIO](#)

Englannin Valioliigassa jyräävä Liverpool on ottanut kaikki keinot käyttöön sekä kentällä että vihreän veran ulkopuolella. Pelaajien ohella mahtiseura menestyy sen ydinfyysikoista ja tohtoreista koostuvan datatutkimuksen tiimin ansiosta.

Kone, Nokia ja YIT perustivat kiinteistödatan ekosysteemin – “Dataa on ennennäkemätön määrä”

Eeva Törmänen 6.4.2020 18:40 [DATA JA ANALYTIikka](#) [ASUMINEN](#) [DIGITALOUS](#) [RAKENTAMINEN](#)

Data on arvokasta, pitäisikö se kirjata taseeseen? ”Voisi heittää joillakin toimialoilla arvostukset ylösalaisin”

Antti Lehmusvirta 9.6.2020 13:55 [SIJOITTAMINEN](#) [DIGITALOUS](#) [DATA](#)

Fortum perustaa data science ja analytiikka -yksikön

Aleksi Kolehmainen 11.2.2020 20:00 [DATA JA ANALYTIikka](#) [TEKOÄLY](#) [DIGITALOUS](#) [ENERGIA](#)

Data science ja analytiikka sijoittuvat samaan toimintoon it:n, innovaatioiden ja digitalisaation kanssa.

Apple apologizes for listening to Siri conversations

Valtio tahkoaa miljoonia euroja myymällä yhteystietoja muun muassa vaippafirmoille ja autokauppiaille – näin estät tietojesi luovuttamisen

Väestörekisterikeskus teki viime vuonna henkilötietojen myynnillä noin 9 miljoonan euron voitot. Etlan asiantuntijan mukaan tietojen pitäisi olla maksuttomia.

Kotimaa 19.11.2019 klo 09.27

Joka viidennessä yrityksessä pelätään tekoälyä - "Vie työpaikat ja tuo ongelmia"

The Cambridge Analytica whistleblower explains how the firm used Facebook data to sway elections

Facebook-testeistä voi olla vakavat seuraukset – myytkö tietosi ulkopuolisille?

Moni käyttäjä suhtautuu epäluuloisesti Fitbitin uuteen omistajaan - "Sairasvakuutusyhtiöt haluaisivat varmasti päästä käsiksi tietoihin"

18.11.2019 19:29 [DIGITALOUS](#) [ÄLYKELLOT](#) [HAKUKONEET](#) [TIETOSUOJA](#) [TIETOTURVA](#)

Yksityisyydensuoja

Kuluttajien dna-testejä tekevä yritys antoi 2 miljoonan ihmisen tietoja FBI:n käyttöön – ei kertonut asiakkailleen

2010

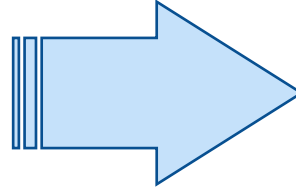
2020 ...

Big Data

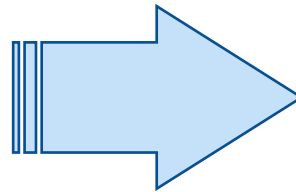
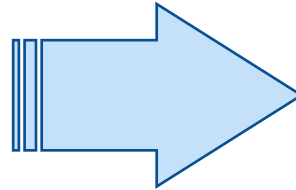
Kehittynyt data-analytiikka

Tekoäly

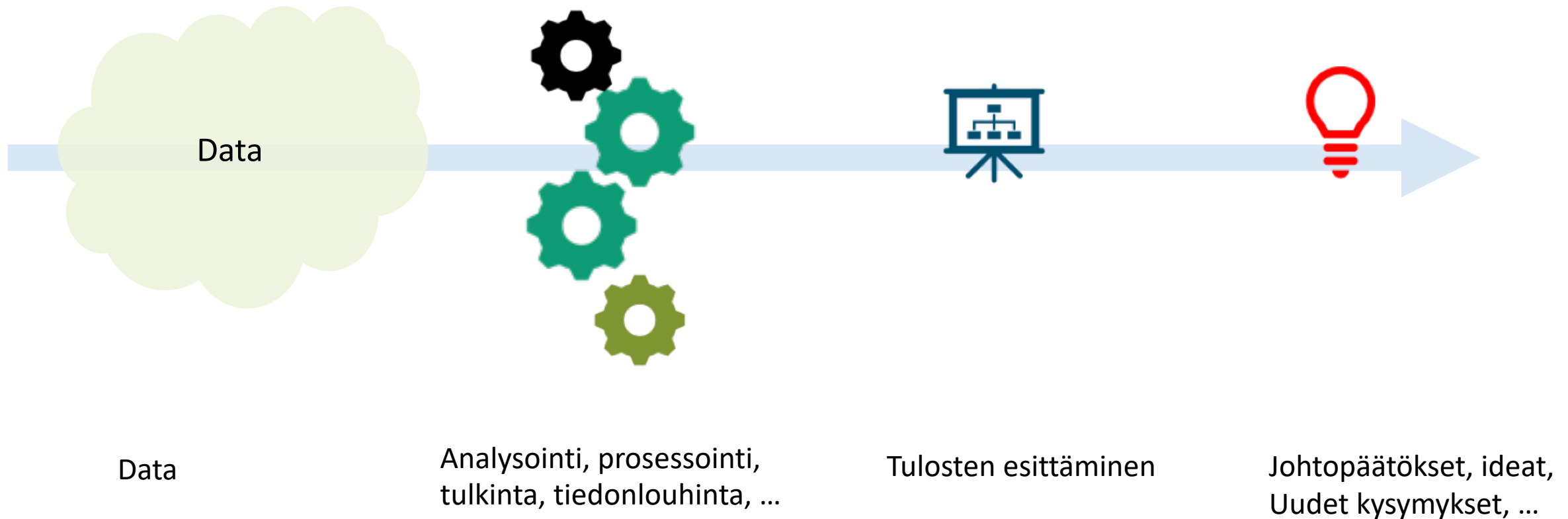
Digitalisaation
vaikutus _____



10110011010011100101000000
11010101000111010010001010
01110100101001010101010001
11001010001111010100101011
10010101000101011101010001
01101010100010110100111000
10110011010011100101000000
11010101000111010010001010
01110100101001010101010001
11001010001111010100101011
10010101000101011101010001
01101010100010110100111000
10110011010011100101000000
11010101000111010010001010
01110100101001010101010001
11001010001111010100101011
10010101000101011101010001
01101010100010110100111000
10110011010011100101000000
11010101000111010010001010
01110100101001010101010001
11001010001111010100101011



Mitä on data-analytiikka?



IoT-laitteista massoittain dataa

Internet of Things, lyhyemmin IoT, myös teollinen **internet**) tarkoitetaan **internet**-verkon laajentumista laitteisiin ja koneisiin, joita voidaan ohjata, mitata ja sensoroida **internet**-verkon yli.

“Enterprises are buying into IoT to save money. They’re doing it to either get predictive maintenance or real-time monitoring of their equipment.”

Dr David Pugh
IoT analyst for IDTechEx

<https://www.iotworldtoday.com/2021/01/07/iot-trends-2021-a-focus-on-fundamentals-not-nice-to-haves/>

<https://elearningindustry.com/iot-internet-of-things-trends-2020>

<https://www.forbes.com/sites/bernardmarr/2018/01/04/the-internet-of-things-iot-will-be-massive-in-2018-here-are-the-4-predictions-from-ibm/>

<https://www.forbes.com/sites/sap/2018/01/10/8-internet-of-things-predictions-for-2018/>

<https://www.networkworld.com/article/3322517/a-critical-look-at-gartners-top-10-iot-trends.html>

<https://www.iotforall.com/6-key-iot-trends-and-predictions-for-2019/>

Helsinki Airport

Helsinki Airport in Finland is using Wi-Fi and iBeacons to track passengers and offer location-based services.

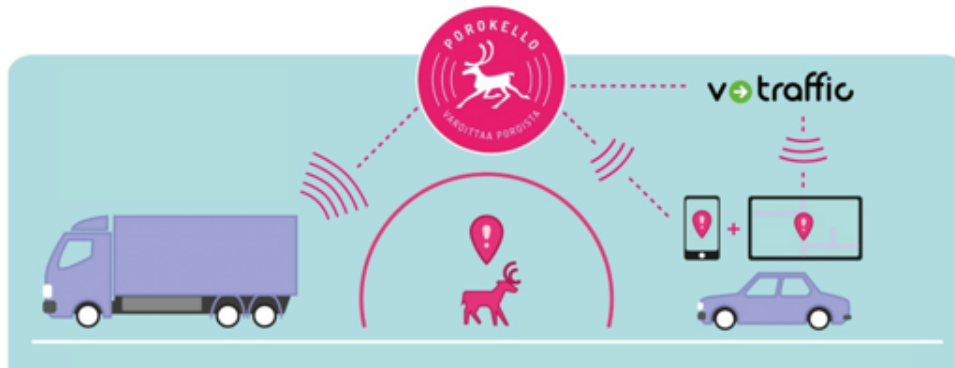


Airport operator Finavia teamed up with local company Walkbase to install dozens of sensors throughout the terminal, and these are able to track all smartphones from the car park to the terminal. By doing this, the airport can prevent queues and bottlenecks where they appear, and also allow retailers to send push notifications to travellers based on their exact location.

<https://internetofbusiness.com/10-real-life-examples-iot-aviation/>



Mimo Monitor:



<https://porokello.fi/>



<https://circuitdigest.com/ten-examples-of-internet-of-things-iot>

<https://medium.com/datadriveninvestor/10-amazing-cases-of-iot-applications-taken-from-the-real-life-a8682cdb48d0>

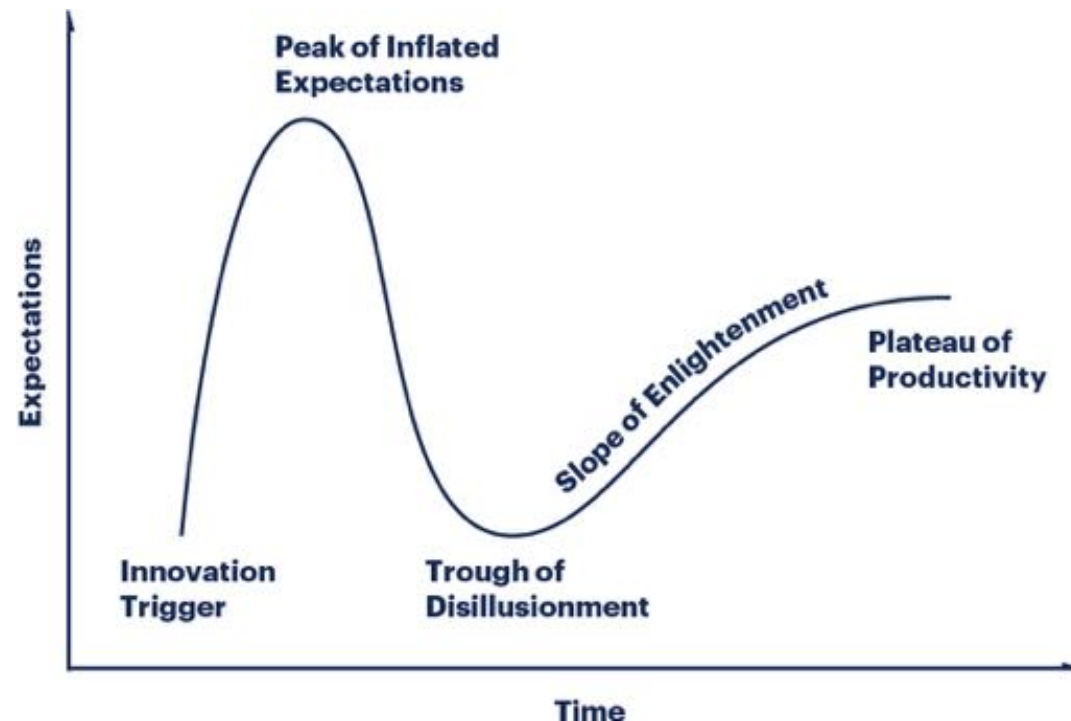
Gartner Hype Cycle

- Gartnerin (<https://www.gartner.com/en>) konsulttitalon vuosittain julkaisema näkemys teknologian nousevista ja laskevista trendeistä
- Trendeissä viime vuosina useita data-analytiikkaan ja tekoälyyn liittyviä teknologiainnovaatioita

Gartner sanoo, että vuonna 2022 sinunkin yritysstrategiasi rakentuu datalla johtamisen varaan – kuinka pääset alkuun datan kesyttämisessä?

Julkaistu 10.3.2020 klo 10.12

<https://studio.tivi.fi/sofigate-the-business-technology-company/gartner-sanoo-etta-vuonna-2022-sinunkin-yritysstrategiasi-rakentuu-datalla-johtamisen-varaan-kuinka-paaset-alkuun-datan-kesyttamisessa>



Gartner Hype Cycle for Emerging Technologies, 2019



[gartner.com/SmarterWithGartner](https://www.gartner.com/SmarterWithGartner)

Source: Gartner
© 2019 Gartner, Inc. and/or its affiliates. All rights reserved.

Gartner

Lähde: <https://www.gartner.com/smarterwithgartner/5-trends-appear-on-the-gartner-hype-cycle-for-emerging-technologies-2019/>

Trendit 2019:

- Self-service BI
- AI / ML
- Increasing number of AI fails!
- Augmented analytics
- Augmented data management
- Continuous intelligence
- In a cloud, on the edge

Hype Cycle for Emerging Technologies, 2020



[gartner.com/SmarterWithGartner](https://www.gartner.com/SmarterWithGartner)

Source: Gartner
© 2020 Gartner, Inc. and/or its affiliates. All rights reserved. Gartner and Hype Cycle are registered trademarks of Gartner, Inc. and its affiliates in the U.S.

Gartner.

Lähde: <https://www.gartner.com/smarterwithgartner/5-trends-drive-the-gartner-hype-cycle-for-emerging-technologies-2020/>

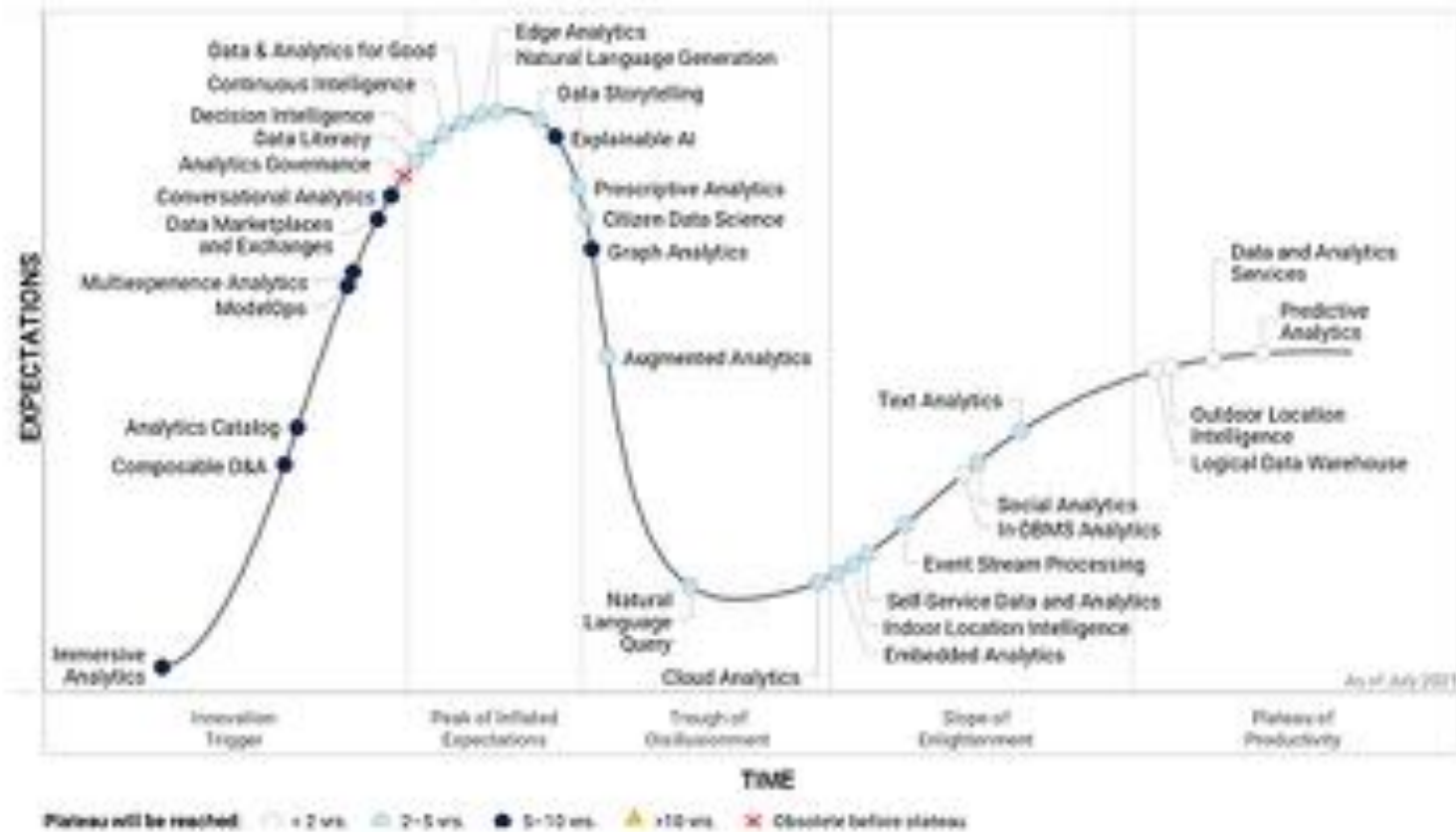
Huomioitavaa

- Explainable AI
- Embeddable AI
- Responsible AI
- (BMI !)

Yleisesti

- Algorithmic trust
- Beyond silicon
- Formative artificial intelligence (AI)
- Digital me

Hype Cycle for Analytics and Business Intelligence, 2021



Source: Gartner (July 2021)

747544

Source: Gartner (July 2021)

<https://www.revealbi.io/whitepapers/gartner-report>

Tehtäviä

- **Pohdi, mitä dataa sinä olet tuottanut/luovuttanut viimeisen viikon aikana**
- Pohdi, mikä tai kuka on saattanut hyötyä tuottamastasi datasta
- Voisiko Gartnerin Hype Cycleä soveltaa jossakin muuallakin kuin IT-maailmassa? Löydätkö esim. omasta elämästäsi tai ympäristöstäsi samankaltaista ”hypetyksen” säännönmukaisuutta?
- Mieti, mitä sensoreita/mittareita sinulla on kodissasi, autossasi ja kenties päälläsi

(Minkälaista dataa olet viimeisen viikon aikana tuottanut/luovuttanut?)

tiedoitusta mm. webinaareista,
sähköposteihin vastaamista,
palavereihin osallistumista
netissä

...mihin klikkailen nettisivuilla,
minkä asioiden kohdalle
pysähdyn, paljonko käytän
sivuilla aikaa

Kirjautumisia mm.
vakuutusyhtiöiden, pankkien,
sähköyhtiöiden,
puhelinyhtiön ja
lehtikonsernin sivuille,
sähköposteja, latauksia,
tykkäyksiä Twitteriin ja
Facebookiin

Viimeisen viikon aikana en ole
mitään dataa
tuottanut/luovuttanut

Tämä kysely on hyvä esimerkki
tuotetusta datasta jonka tiedän
luovuttavani eteenpäin
vapaaehtoisesti

Autolla ajaessani ohitan
liikennelaskureita, jotka
tuottavat dataa
liikennemääristä

Uuden tietokoneen
käyttöönottoon liittyvät
rekisteröinnit ja uusien tilien
luominen. MOOC-kurssille
osallistuminen

Pankkikortin käyttö, bonuskortin käyttö

TV:n kautta käyttäjä
tietoja: yle areena,
Youtube, Netflix

Autolla liikkuminen

Työhakemusten lähettäminen

Olen hyväksynyt
evästeitä
nettisivuilla,
katsonut videoita ja
mainoksia.

Olen myös tuottanut
käyttödataa omalle
ohjelmistolle

...Polar-urheilukello ja Oura-
sormus, joiden käytöstä luovutan
dataa näille firmoille

Data-analytiikka & liiketoiminta

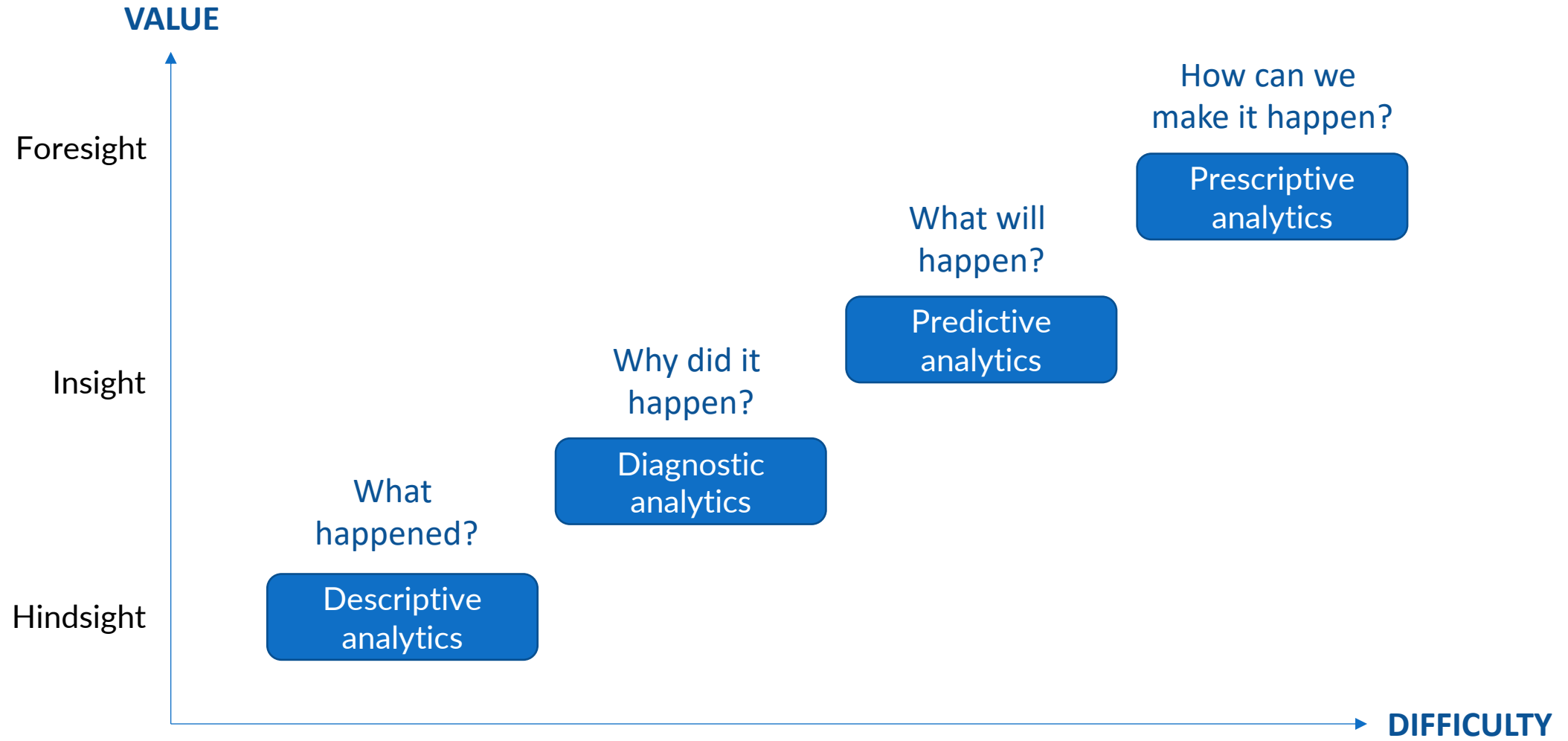
Sisäistä ja ulkoista data-analytiikan hyödyntämistä

Kehitystyö/innovaatiot:
Esim. **sovellukset ja ratkaisut**, joissa
hyödynnetään data-
analytiikkaa

VS

Data-analytiikan
hyödyntäminen
kyseisen yrityksen
liiketoiminnassa
(asiakasanalyysit,
brändiarvo, KPI:t,...)

Gartner: data-analytiikan ”portaati”:



	Business Intelligence	Kehittynyt analytiikka
	Menneisyys	Tuleva
Kysymykset	Mitä tapahtui? Koska? Kenelle? Kuinka monta?	Mitä tulee tapahtumaan? Entä jos muutamme X:ää?
Metodit	Raportit (listat, KPI:t, metriikat) Hälytykset Dashboardit Scorecardit OLAP Ad hoc	Ennustava mallintaminen Tiedonlouhinta Tekstinlouhinta Tilastotiede Simulaatiot ja optimointi
Big data	Kyllä	Kyllä
Data	Pääasiassa rakenteellista	Rakenteellista ja rakenteetonta
Käyttäjät	Liiketoiminta	Kehittäjät, data scientistit, analytiikot, IT
Vaikutus	Reagoiva	Ennakoiva

Tuotteesta palveluksi

Tuote



Arvoa transaktiosta



Palvelu



Arvoa käytöstä

Palvelun hyvä hallinta, proaktiivisuus, asiakastyytyväisyys



Rolls-Royce®



<https://medium.com/swlh/from-product-to-product-as-a-service-37baed471cd6>

<https://www.kone.com/en/products-and-services/maintenance-and-modernization/24-7-connected-services.aspx>

Tehtävä

- Valitse jokin ostohinnaltaan isohko tuote, joko kuluttajatuote (auto, vene, moottoripyörä, tms.) tai yritystuote (tulostin, pakettiauto, sähkömoottori, nostin, tms.)
- Mieti
 - Asiakkaan kannalta käyttöön liittyviä arvoja
 - Asiakkaan riskejä, ja niihin mahdollisesti liittyviä sanktioita
 - Mitä mielenkiintoista käyttöön liittyvää analytiikkaa toimittaja voisi tehdä tai jopa tarjota asiakkaalle
- Millaisilla asiakassegmenteillä tunnistat palvelulle hyvän mahdollisuuden?

Kilpailukyky, kilpailuetu

- **Tunne** asiakkaasi, yrityksesi ja kilpailijasi

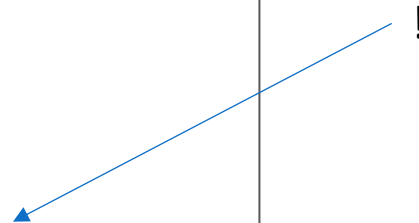
- Kohdennettu markkinointi
- Segmentointi
- Optimointi
- Ennakointi
- Simulointi

- Terveysteknologia, bioteknologia, X-teknologia, ...
- Pankki- ja vakuutusala(riskienhallinta, luottoanalyysit, petosten torjunta...)
- Kaupan ala (tuotesijoittelu, verkkoanalytiikka,...)
- Mainonta
- Tuotekehitys
- Liikenne
- ”Kova” teollisuus
- Yhteiskunta, politiikka, vaikuttaminen
- Kunnat, kaupungit

Kehittynyttä data-analytiikkaa/tekoälyä

“Today it isn’t just online and information firms that can create products and services from analyses of data. It’s every firm in every industry.”

Thomas H. Davenport,
Harvard Business Review, **2016**



The market size for business intelligence and analytics software applications is forecast to increase worldwide over the next few years from **14.9 billion** U.S. dollars in 2019 to **17.6 billion** in 2024.
<https://www.statista.com/statistics/590054/worldwide-business-analytics-software-vendor-market/>

Huomio!

- Maailma on täynnä toinen toistaan mielenkiintoisempia kertomuksia data-analytiikan hyödyntämisestä
- Yleensä julkisesti kerrotaan mihin **tarpeeseen** (resurssien kohdentaminen, markkinointi, hinnoittelu, asiakassegmentointi, asiakaspalvelu, suosittelujärjestelmät,...) analytiikkaa hyödynnettiin ja mitä vaikutuksia (yleensä positiivisia) prosessilla oli
- Se **MITEN** data-analytiikkaa hyödynnettiin on usein **LIIKESALAISUUS**. Harva tahtoo julkaista ratkaisun siihen, **millä keinoin ja menetelmin** yrityksen kilpailukyky parantui.
 - Keinot ja menetelmät = *data + käytetyt algoritmit*

*FYI



Final 2018 version, updated 07/15/2018

© Matt Turck (@mattturck), Demi Obayomi (@demi_obayomi), & FirstMark (@firstmarkcap)

mattturck.com/bigdata2018

FIRSTMARK
EARLY STAGE VENTURE CAPITAL

Edellisen
vuosineljänneksen
tulos oli
positiivinen

Minna myi enemmän
(+55000€) kuin Arto

15% asiakkaista
lopetti
asiakkuutensa
heinäkuussa

Mittaroinnista

Taaksepäin katsomisesta kohti edessä olevia haasteita ja mahdollisuuksia

Suorituskykymittarit



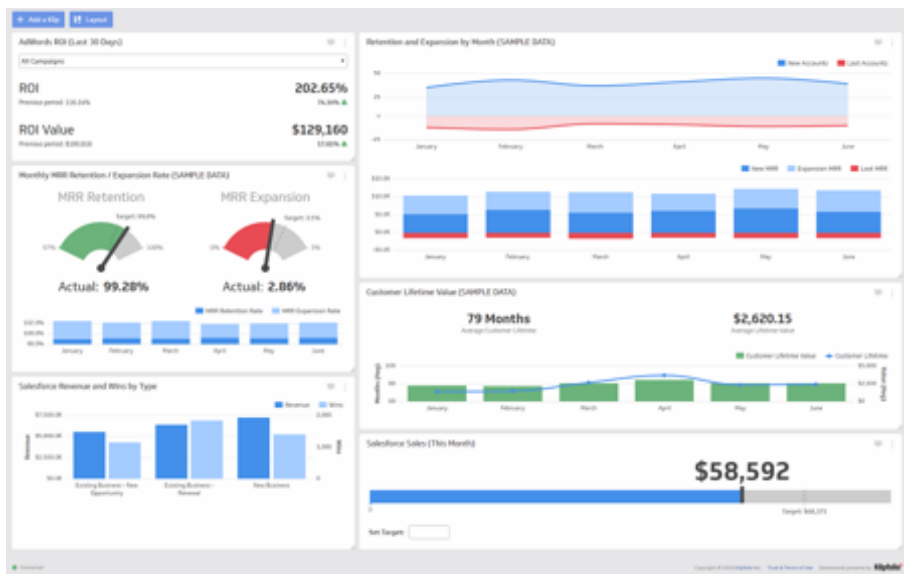
<https://www.intel.co.uk/content/www/uk/en/it-management/cloud-analytic-hub/big-data-powers-athletes.html>

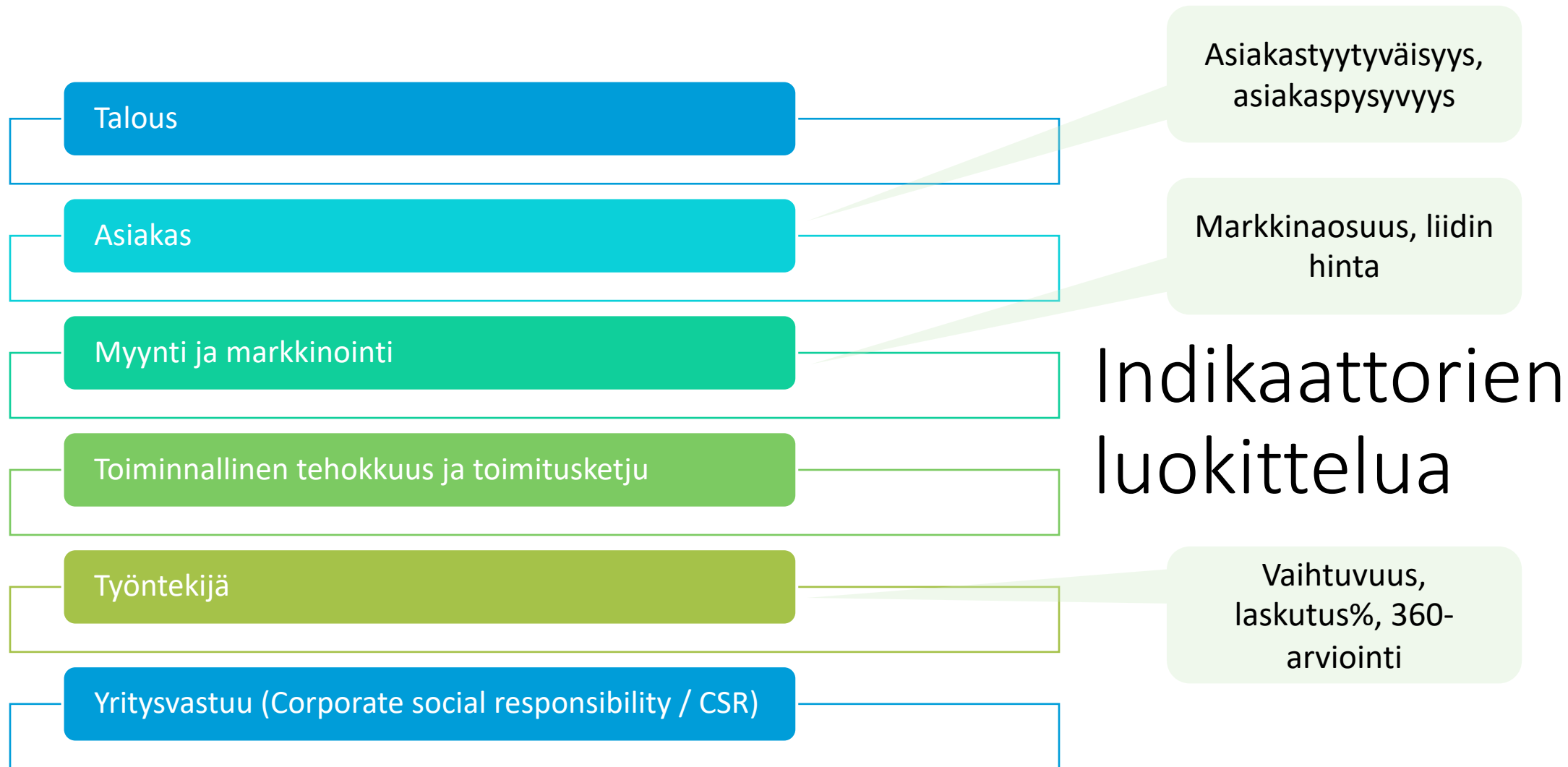


<https://www.geckoboard.com/learn/dashboard-examples/executive/financial-dashboard-example/>

Datan liiketoimintahyödyntäminen: Indikaattorit, tunnusluvut, mittarit

- Indikaattorien avulla organisaatio voi seurata (omaa ja muiden) suoriutumista
- Indikaattorilla on tyypillisesti numeraalinen arvo
- Erilaisia indikaattoreita on tuhansia ja uusia kehitellään koko ajan
- Organisaatio (tai henkilö) valitsee itselleen tärkeimmät indikaattorit - Key Performance Indicators (KPI)
- Visualisointi normaalisti dashboardeilla (“mittaritaulu”)

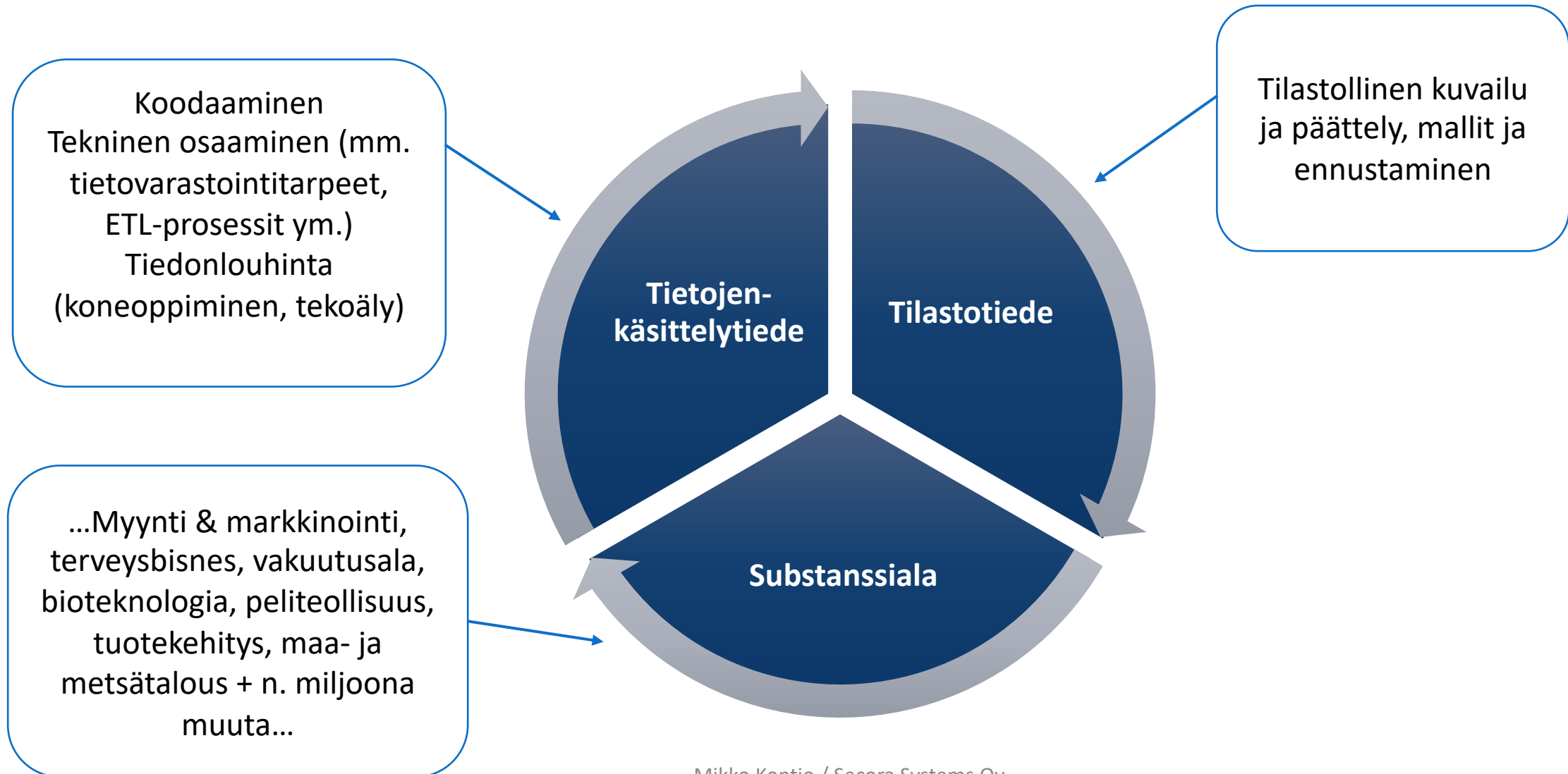




Data-analytiikan osa-alueet ja osaajat

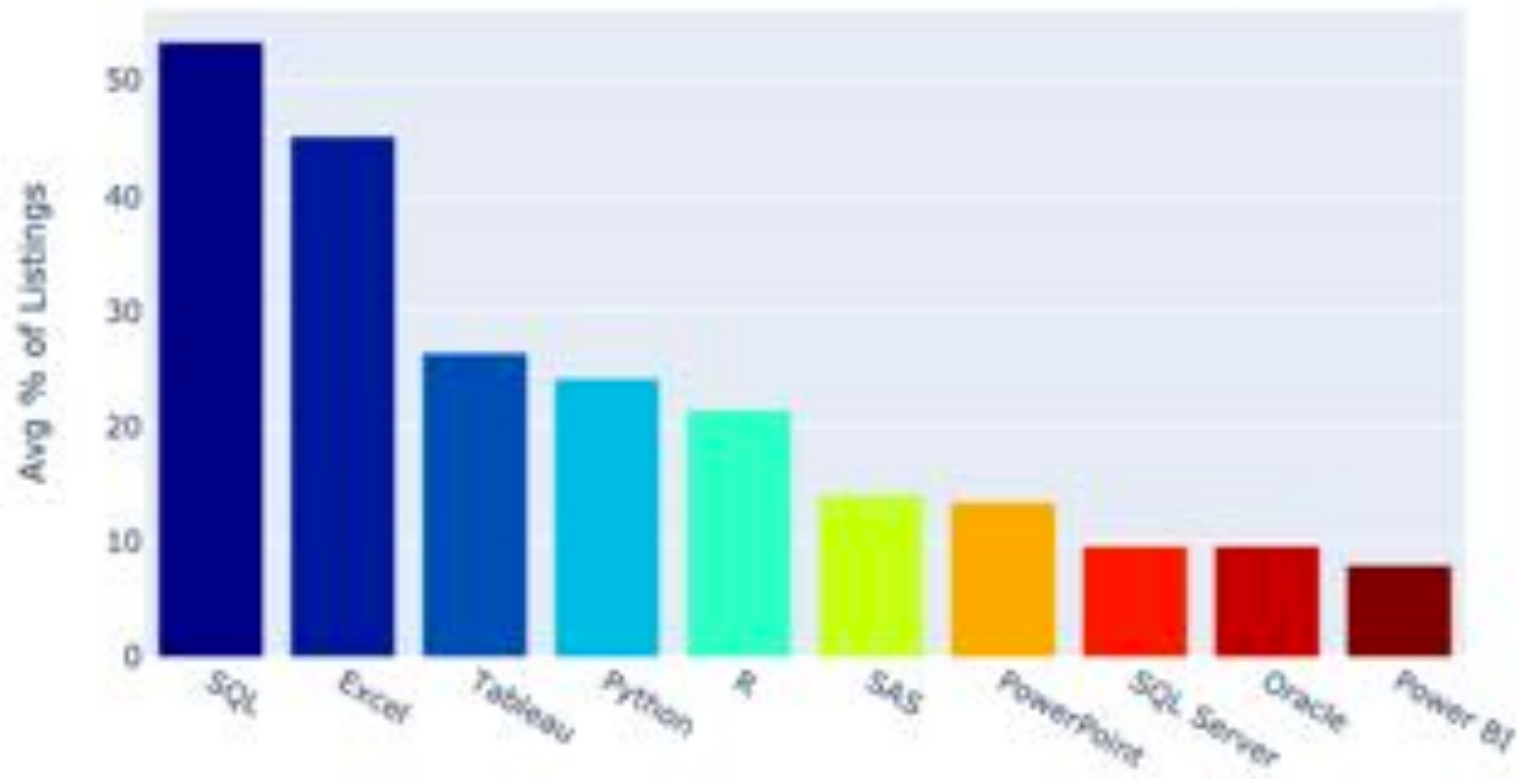
Osaajia tarvitaan laidasta laitaan!

Data-analytiikan osa-alueet



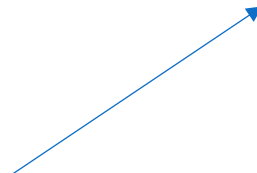
Data-analyttikko

Technologies in Data Analyst Job Listings 2020



<https://towardsdatascience.com/most-in-demand-tech-skills-for-data-analysts-26d4ea4450f8>

Data-analyttikko...?



	Data-analyttikko Varma Julkaistu 13.3.2020 Helsinki
	Data Engineer Varma Julkaistu 20.3.2020 Helsinki
	Data Scientist SOK Julkaistu 10.3.2020 Helsinki
	Erityisasiantuntija (datatieteet), Data Scientist Finnish Institute of Occupational Health Julkaistu 24.3.2020 Helsinki

it/data-analyttikko/1350349

Tehtäviisi kuuluu datan poimintaa, käsittelyä ja analysointia:

- Uusien datalähteiden kartoittamista
- Tiedon jalostamista tutkimusaineistoksi tai tuotannolliseksi palveluksi
- Tekstianalytiikan hyödyntämistä datan rakenteellistamisessa
- Analyysi- ja ennustavien mallien rakentamista
- Hypoteesien muodostamista ja työkyvyttömyyteen johtavien polkujen visualisointia
- Tulosten julkaisua, viestintää ja toimenpide-ehdotusten tuottamista

Arvostamme seuraavia osaamisia:

- Data-analytiikan, datan käsittelyn tai tilastomallinnuksen osaaminen
- Kehittyneiden tilastomenetelmien, koneoppimisen ja big data -analytiikan tuntemus
- Ohjelmistoarkkitehtuurien tuntemus, tietovarastoinnin periaatteet ja tietomallinnuksen käsitteet
- Osaaminen ja kokemus tilastovälineistä, kuten SAS, R ja SPSS sekä ohjelmointikielistä Python.

Lähde:

https://tyopaikat.oikotie.fi/haku?jq=Data&sort_by=score

Tehtävä

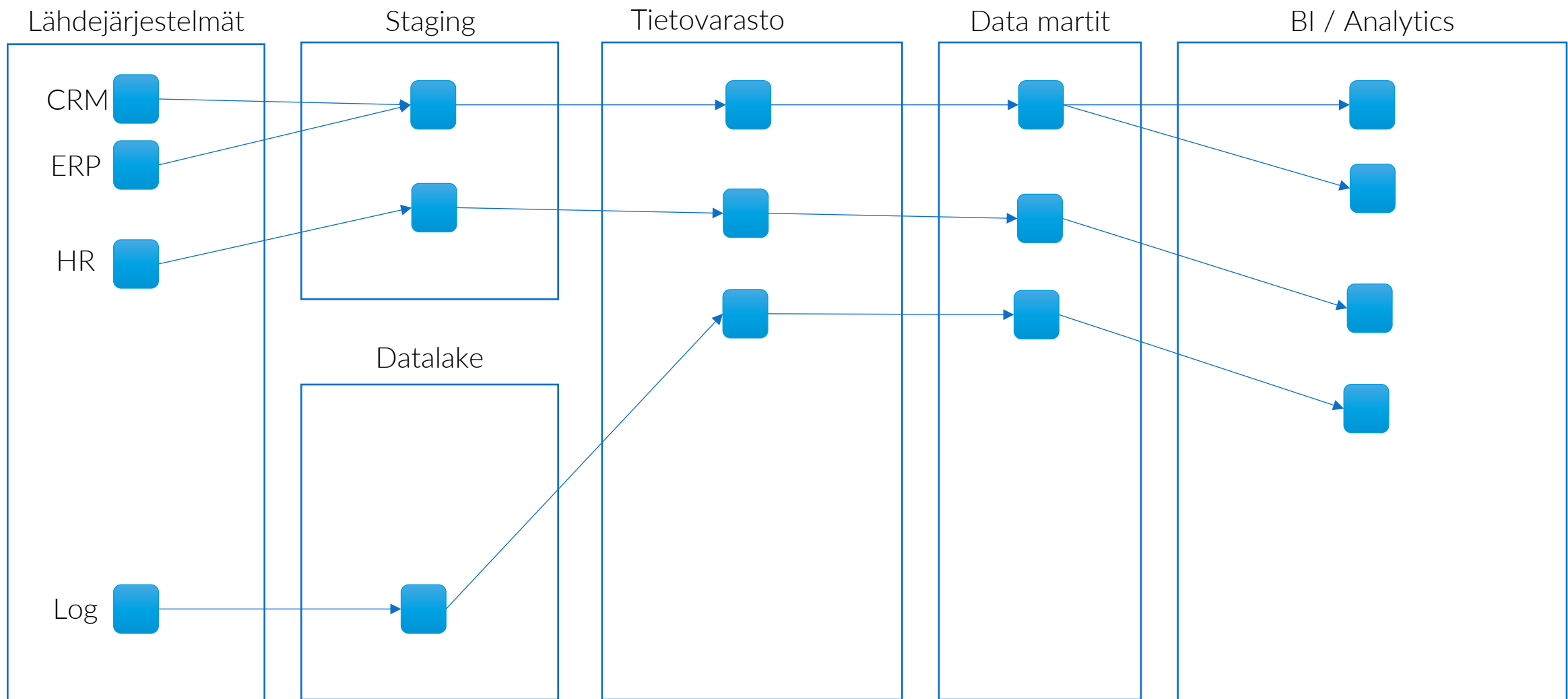
- Tutustu, minkälaisia työtehtäviä ja nimikkeitä löydät data-analytiikkaan liittyen.
 - Onko työnkuvaus mielestäsi realistinen?
- Mitkä ovat sinun kyvykkyytesi ja vahvuutesi työelämässä? Mitä jo osaat?
- Mistä olet kiinnostunut? Mistä oikealla olevista sektoreista haluaisit tietää/oppia lisää?

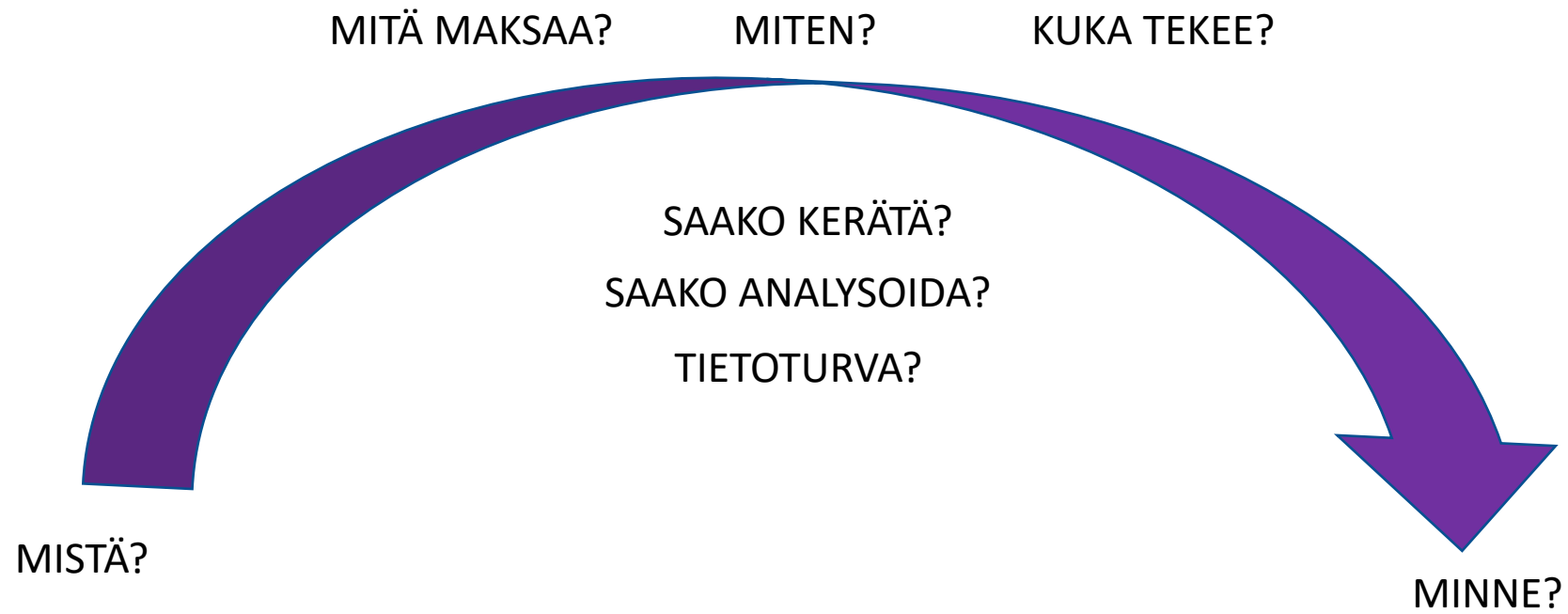
Datasta päätöksentekoon

Raakadatan jalostussykli

Jalostussykli: “Datakaruselli”

- Seuraavalla dialla esitetään datan hyödyntämisen kiertokulku, ystävällisemmin karuselli, jonka alkupisteenä on datalähteet ja loppupisteenä päätöksenteko/reagointi.
- Karuselli sisältää monta välietappia, osa niistä on vaihtoehtoisia (riippuu, mitä haetaan)
- Päätöksenteko ja reagointi synnyttävät (välillisesti ja välittömästi) uutta dataa, joka sekin kannattaa hyödyntää uudelleen alkupisteessä
- Karuselli on hyvin suuri ja sen saaminen ”käyntiin” vaatii/voi vaatia useita osaajia ja useita – kenties aivan uusia resursseja- eri sektoreilleen
- Yrityksen laajuudesta ja analytiikan hyödyntämisen vaatimuksista riippuu, kuinka suuren investoinnin ja resurssoinnin prosessin käynnistäminen vaatii
- Kuvan tarkemmat selitykset jatkuvat seuraavilla dioilla





1. Data

1. Data import/data export?

- Datan lähteet?
 - Yrityksen ulkoiset ja sisäiset?
Rakenteelliset/rakenteettomat?
 - Tietokannat ja API:t (RDBMS, NoSQL, Hadoop, jne...)
 - Kustannukset? (Avoin data, oma data vai maksullinen?)
- Mitä tietoa on oikeus kerätä?
- Datan luotettavuus? (=lähteen luotettavuus)

Jo pelkästään yrityksen sisällä olevan, kenties siiloutuneen, datan "löytäminen" ja eri datojen yhdistäminen voi olla kompleksinen ja aikaa vievä työ.



Jatkuu...

- Datan laatu? (mm. “reikäistä dataa?”)
- Datan määrä?
 - Massadata vai pienempi?
- Datan muoto, formaatit
 - JSON, XML, CSV, Excel, RDF,
- Minkälaiseen tietovarastoon data tallennetaan
 - Riittääkö tallennuskapasiteetti?
 - Voiko viedä pilveen? (Tietoturva)
 - Nopeus/saatavuus?
 - Vaatiiko analysointi järeää laskentaa?



Kuvalähde: <http://www.marketstatsnews.com/proliferation-of-crucial-data-in-healthcare-and-bfsi-sectors-bolster-uptake-of-master-data-management-solutions/>

Datan lähteet

	Rakenteellinen (structured)	Rakenteeton (unstructured)
Sisäinen (internal)	Myynti Tuotanto Palkat Talous Alakohtainen data (esim. CRM, ERP, potilastietojärjestelmät,...)	Muistiot Lokitiedostot Sähköpostit Dokumentit Alakohtainen data (esim. potilastietojärjestelmät)
Ulkoinen (external)	Pörssi Valuutta Yritystiedot	Sosiaalinen med. Foorumit Blogit Media Lainsäädäntö

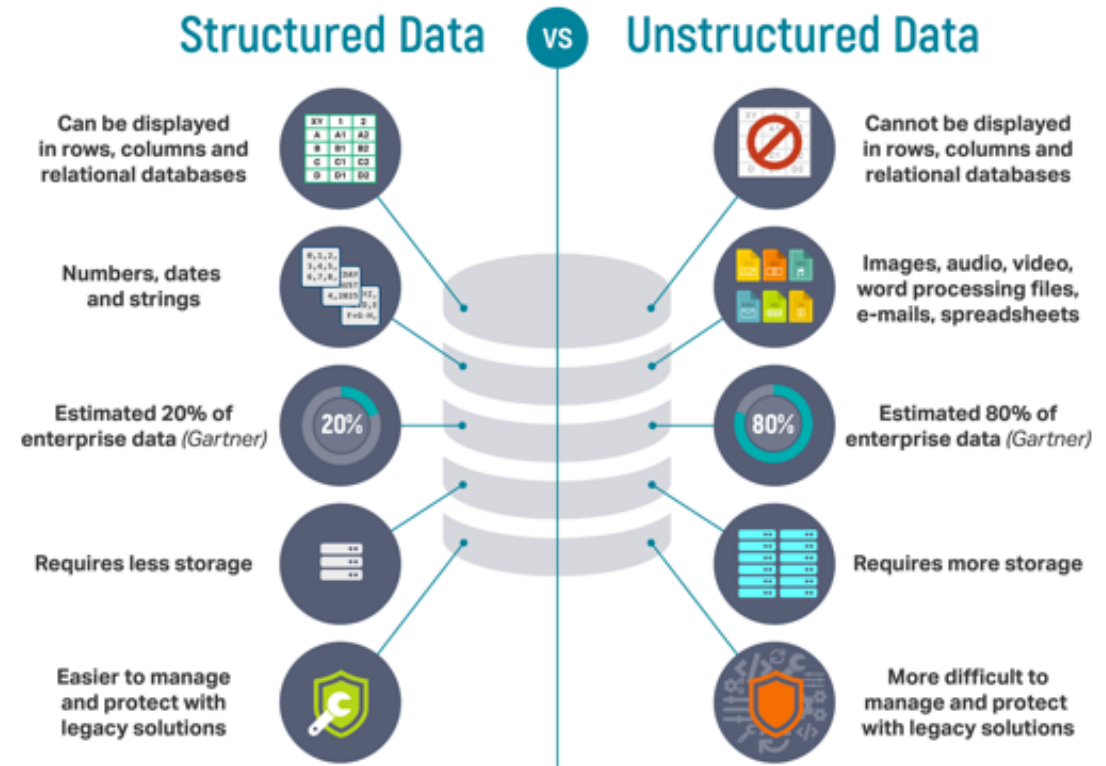
Mitä on rakenteinen data?

- Ennalta määriteltyä
- Tarkasti organisoitua
- Data identifioitavissa
- **Relaatiotietokannat**, taulukot (spreadsheets)
- Data warehouses
- XML
- CRM- ja ERP-data
- Helppo/helpohko käsitellä datamallia noudattamalla
- Metadata saatavilla?
- Rakenteisuuttakin eri tasoista!



Rakenteeton data:

- Rakenteettomalla datalla ei ole tietoa tiedosta eli nk. metadataa
 - Video
 - Audio
 - Sähköposti
 - RSS-syötteet
 - Tekstidokumentit



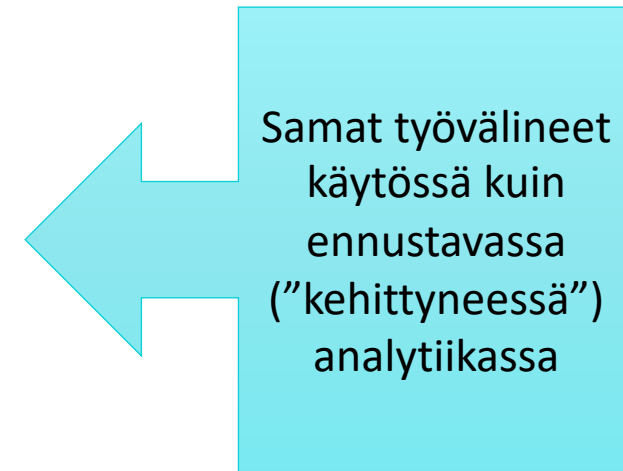
Kuvalähde: <https://lawtomated.com/structured-data-vs-unstructured-data-what-are-they-and-why-care/>

Rakenteeton data...

- Ennakkoon ei tiedetä, mitä sisältää (ei datamallia, ei metadataa)
- Voi sisältää useampaa eri dataformaattia
- Ei metadataa
- Vapaa teksti perinteisin esimerkki
- Rakenteettoman datan hyödyntäminen yksi tämän hetken ”trendeistä” (twitter-, facebook-, blogit, asiakaspalautteet, potilaskertomukset, artikkelit, palvelutiketit...)
- Perinteiset koneet/ohjelmat eivät (vielä) pysty helposti sulattamaan rakenteetonta dataa mutta kehitystyötä tehdään valtavasti

Mikä puree rakenteettomaan dataan?

- Tiedonlouhinta (data mining)
 - Tekoälyä
 - Koneoppimista (ohjattua/ohjaamatonta)
 - Analytiikkaa
 - Statistiikkaa

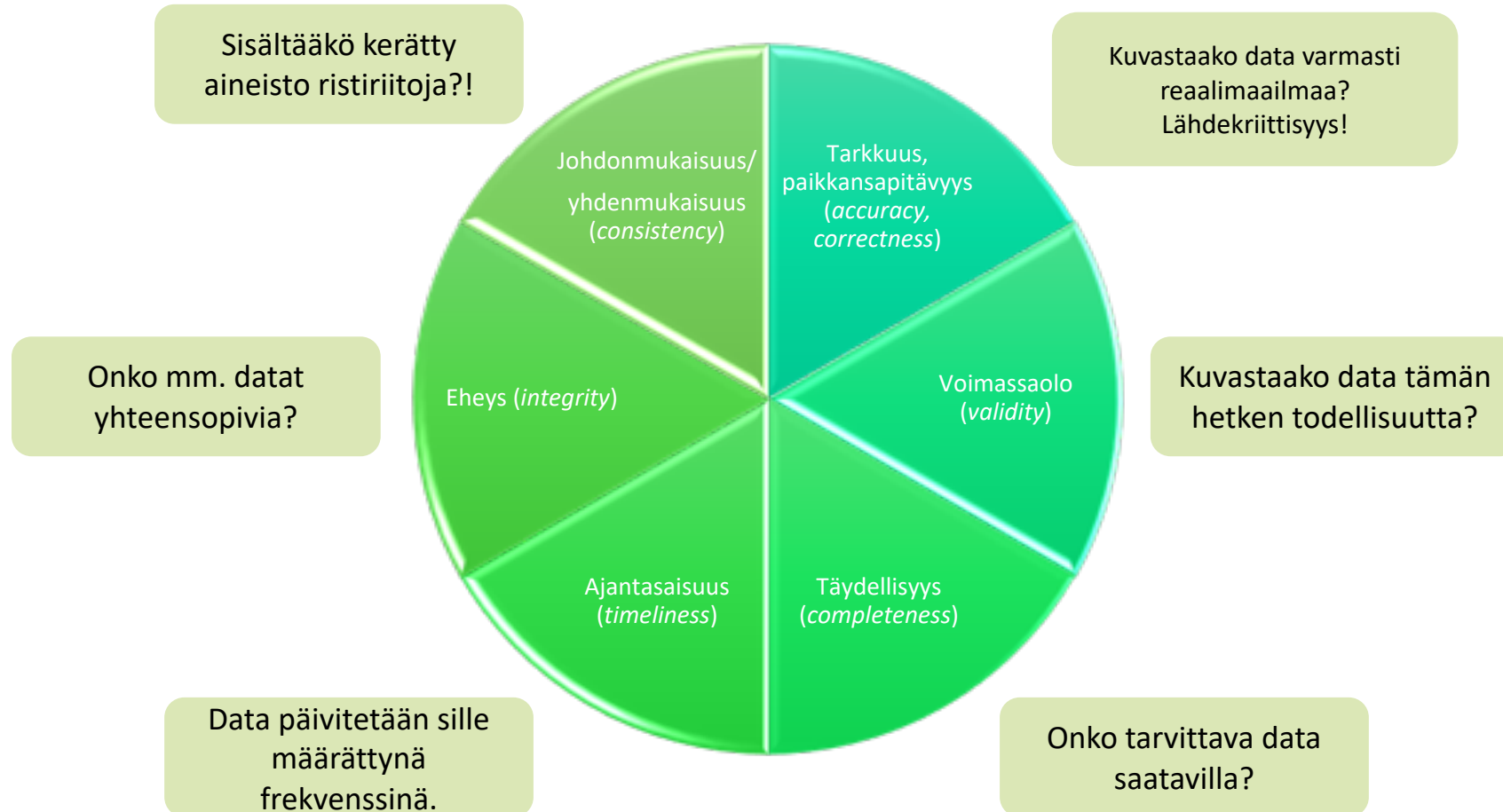


- Mm. yhteyksien (relationship), "lainalaisuuksien", hahmojen etsintää (pattern recognition)

Datan laadusta:

- Korvaako määrä laadun?
- Mitä dataa on kerätty? (Onko kaikki tarvittava data varmasti mukana?)
- Mistä dataa on kerätty? (Lähteen/aineiston luotettavuus?)
- Millä keinoin?
- Miten on putsattu, korvattu ja/tai yhdistelty? (esim. puuttuvat arvot korvattu keskiarvoilla?)
- Milloin on kerätty/päivitetty (ajankohta, ajankohtaisuus, aikaikkuna?)

Laadukkaan datan vaateita:



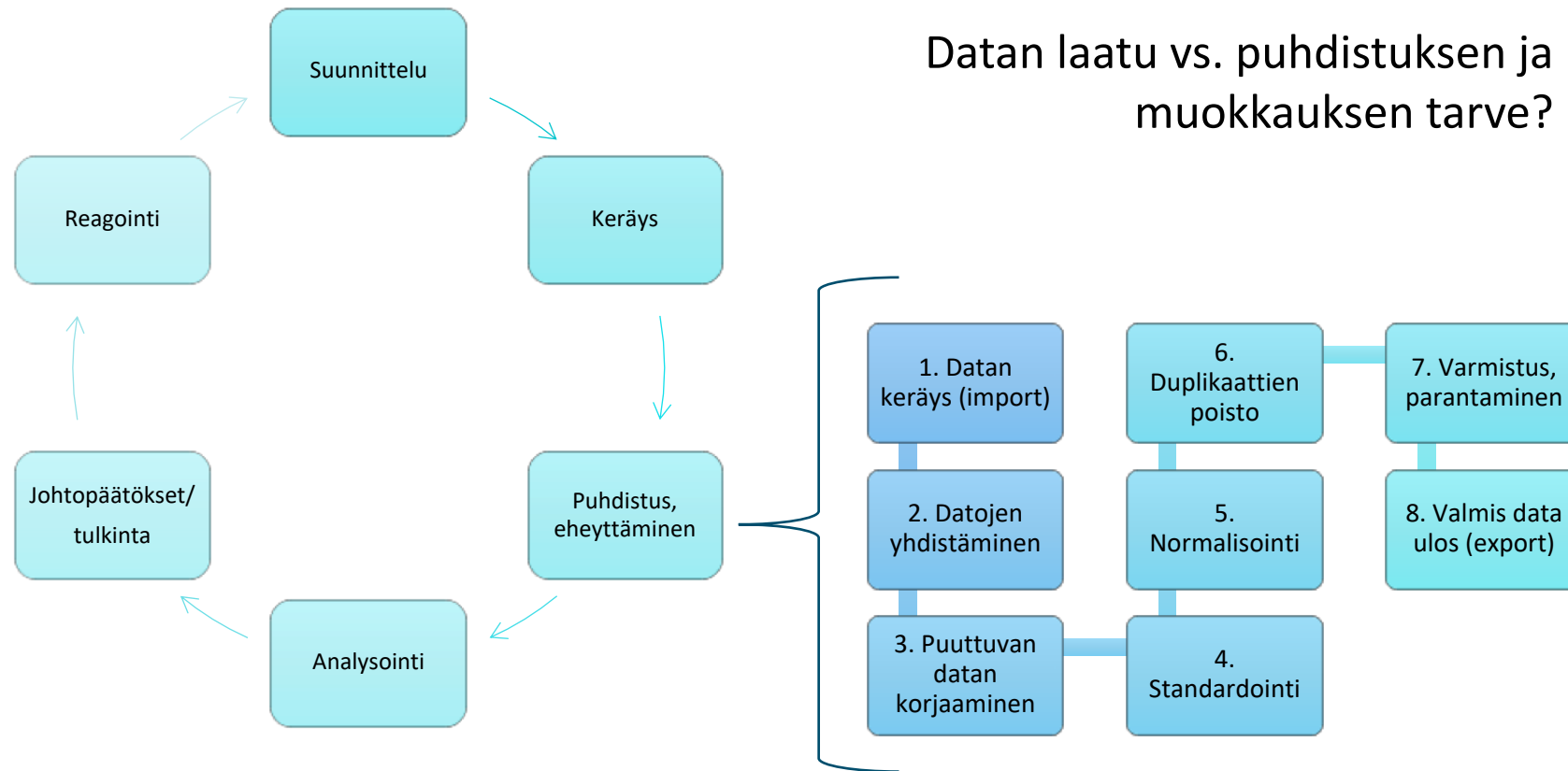
Minne jatkokäsiteltäväksi?

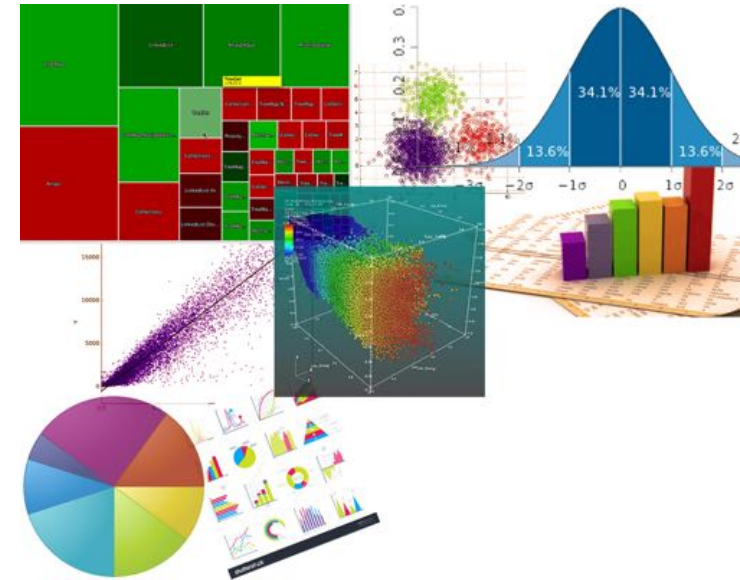
- Tallennuskapasiteetin tarve?
 - Kuinka paljon dataa tulee?
 - Reaaliaikaisuuden tarve?
 - Oma tietokanta/palvelin/private cloud?
 - Tallennuskapasiteetti pilvestä?
 - Hadoop?
 - Relaatiotietokanta vai NoSQL-kanta?
- Laskentaresurssit?
 - Pyörittääkö Excel vai tarvitaanko järeämpi kalusto?
 - Laskennan aikavaativuus?

2. Datat esiprosessointi

70-90% data-analyttikon työstä

Puhdistus ja eheyttäminen



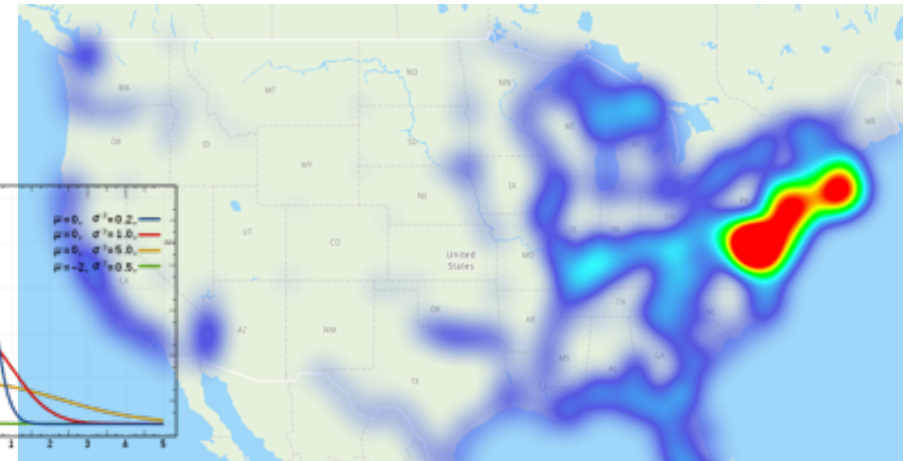
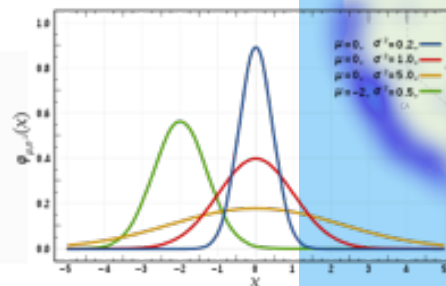
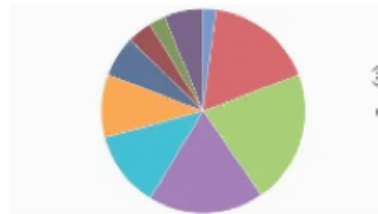
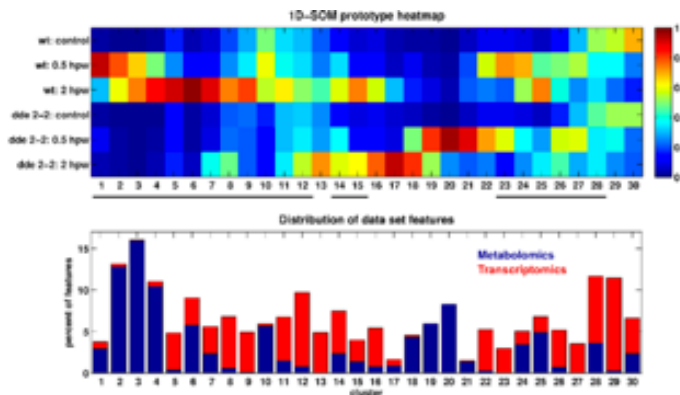


3. Kuvaileva analytiikka

Tilastotiedettä ja datan visualisointia

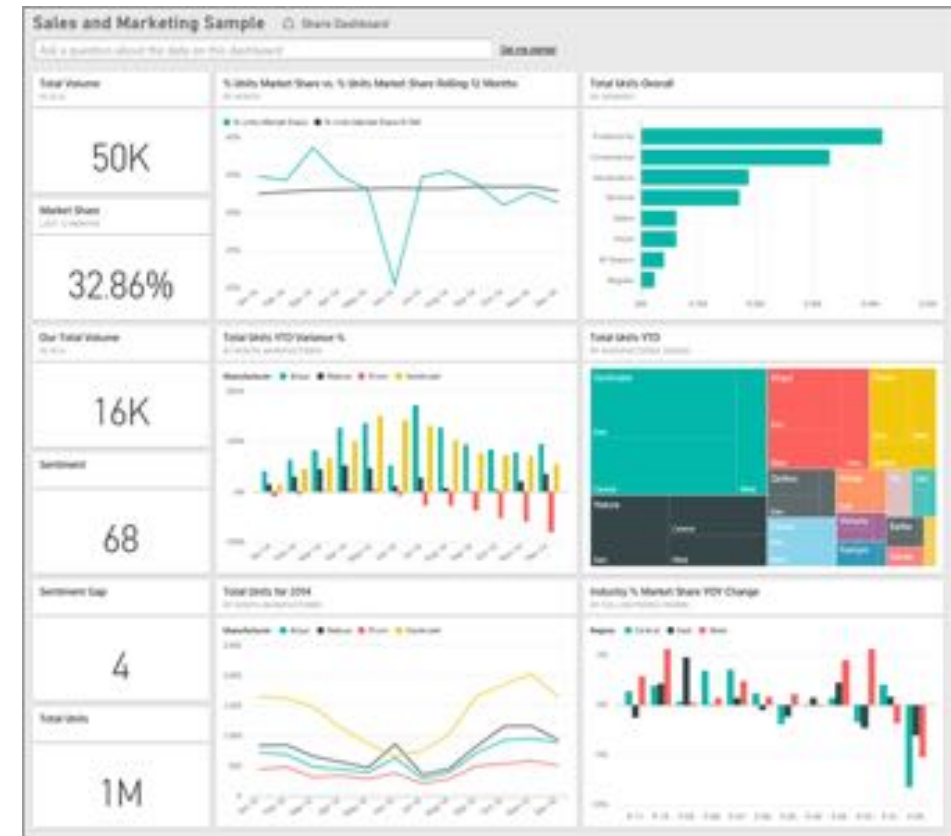
Kuvaileva (deskriptiivinen) analytiikka

- BI:n “lopputuote” tai datan tutkimisen työkalu ennen siirtymistä kehittyneen analytiikan puolelle
- **Kuvaileva analytiikka** on ollut “perus-tilastotieteen” työkalujen käyttöä mutta modernit visualisointityökalut tuovat kenties uusia mahdollisuuksia *tiedon* havainnointiin datasta
 - Piirakkakuviot, pylväsdiagrammit, heatmapit, jne ...



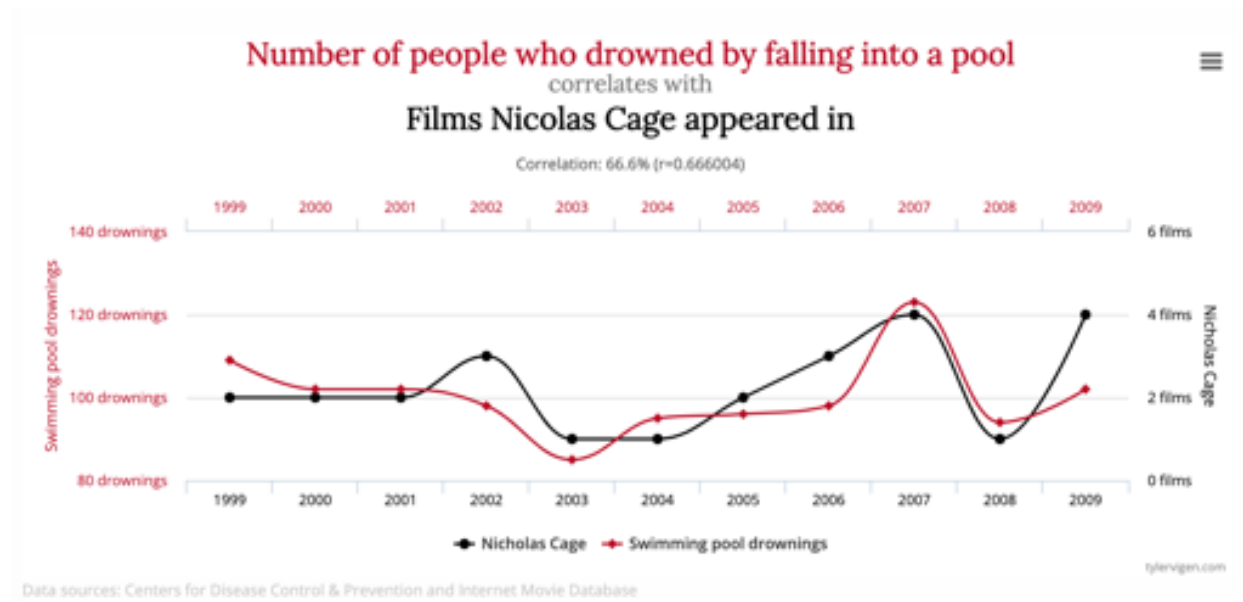
Business Intelligence

- Perusnimitys peruutuspeiliin katsovalle raportoinnille
- Välineitä paljon
 - Power BI
 - Tableau
 - AWS Quicksight
 - Google Datastudio

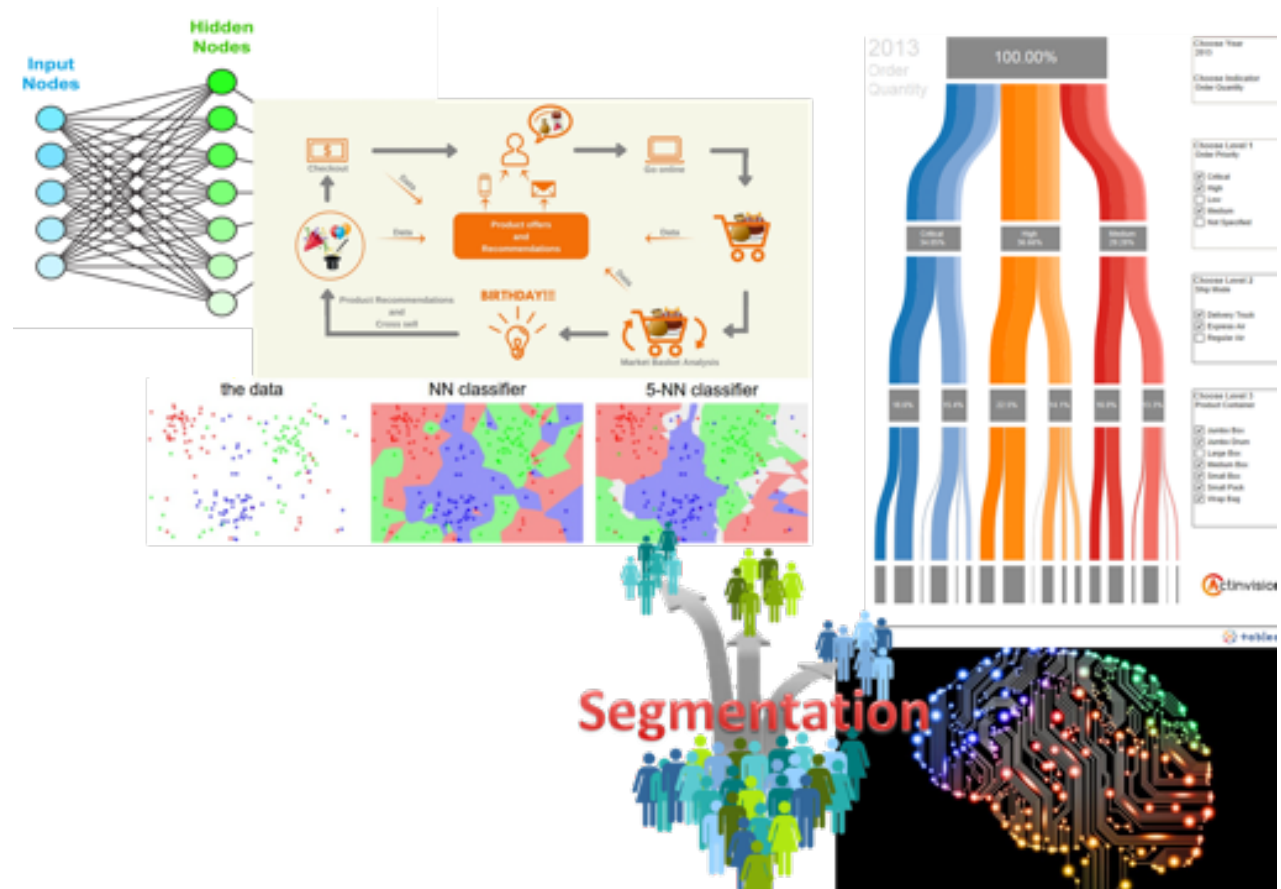


Kuvaileva analytiikka jatkuu

- Ennen siirtymistä kehittyneen analytiikan puolelle: mitä saadaan tehdä ja mitä ei (*“tilastotieteellinen uskottavuus”*)
- Tunnusluvut, jakaumat, hajonta, keskiluvut, korrelaatiot, testit...
 - Korrelaatio vs. kausaaliteetti



Kuvalähde: <http://www.tylervigen.com/spurious-correlations>

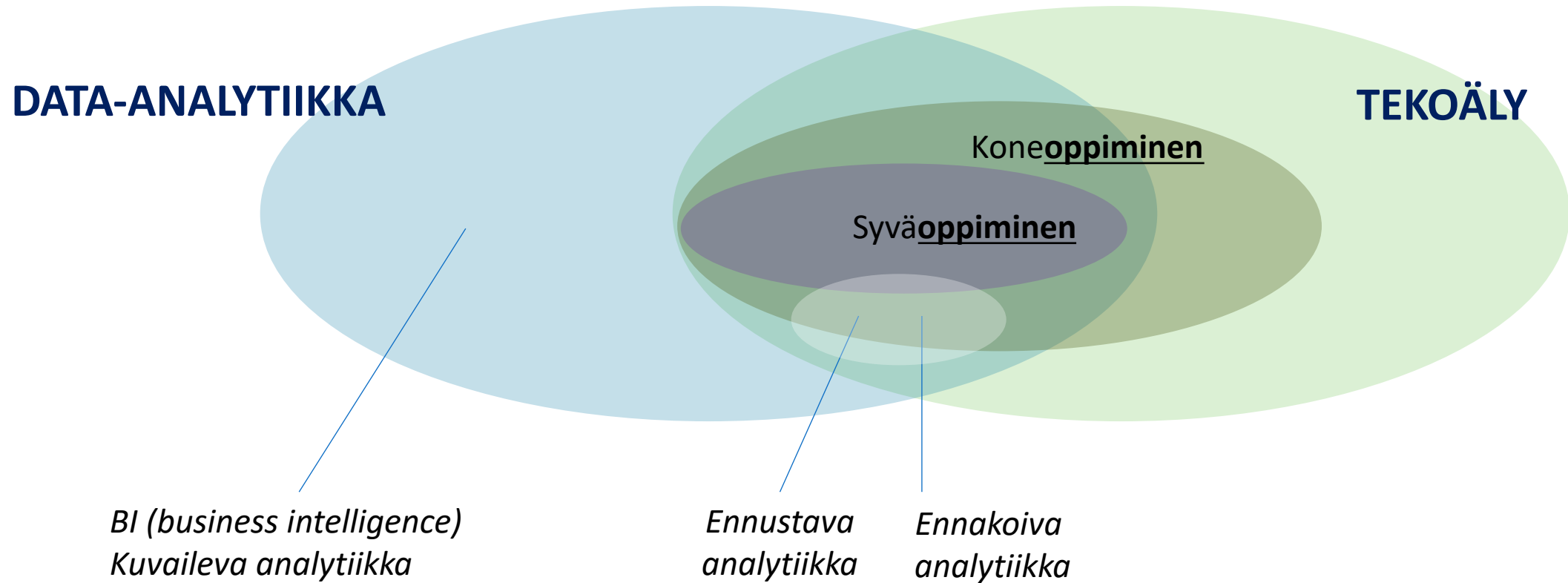


4. Kehittynyt data-analytiikka

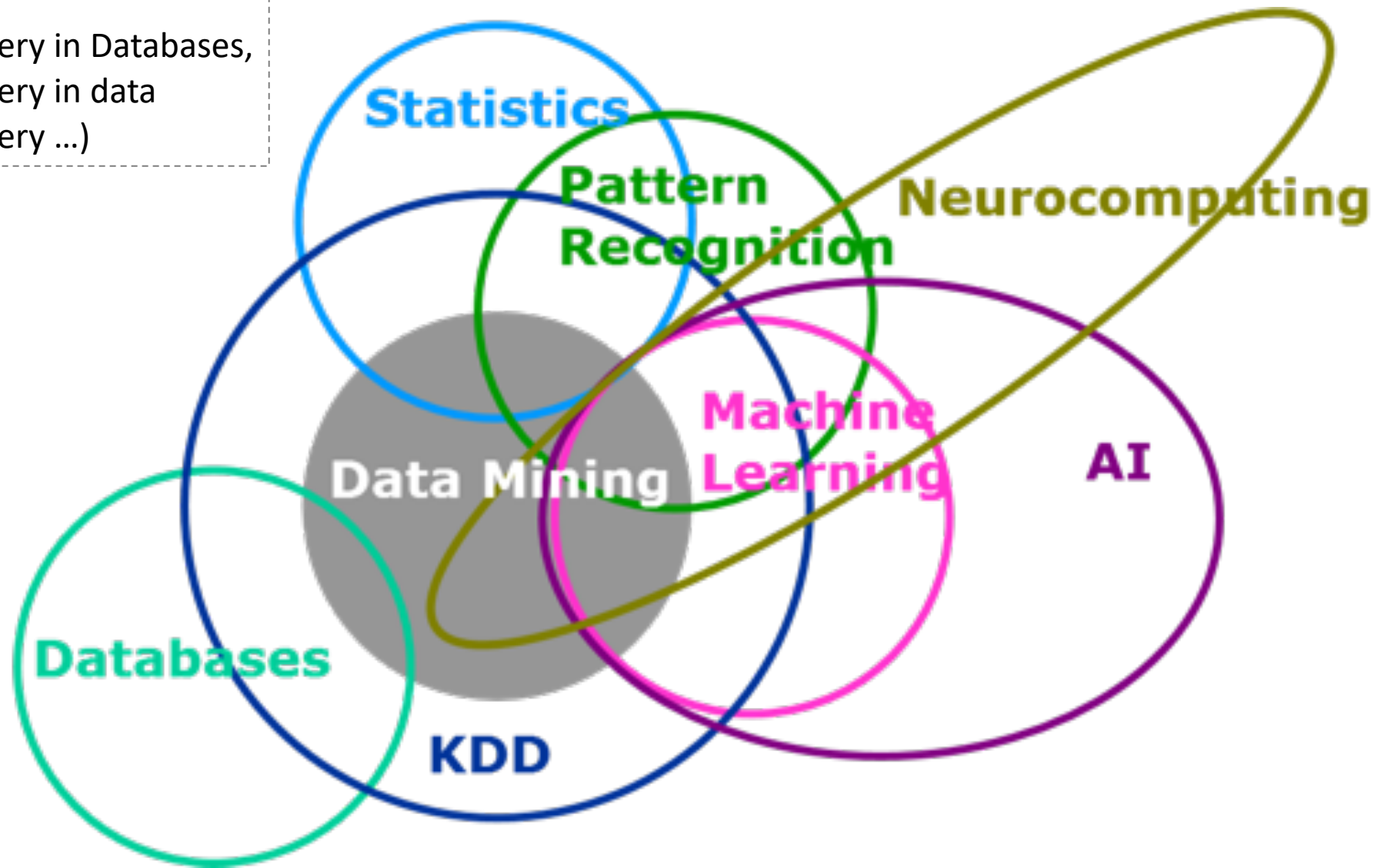
”Ennustava data-analytiikka”, ”moderni data-analytiikka”,...

*Healthtech companies and healthcare organizations are using a technique called Discrete Event Simulation to predict wait times for patients in emergency department waiting rooms. The models use factors such as staffing levels, patient data, emergency department charts, and even the layout of the emergency room itself to predict wait times. (Lähdettä ei tiedossa)

Lisätietoa: <https://www.diva-portal.org/smash/get/diva2:1131383/FULLTEXT01.pdf>

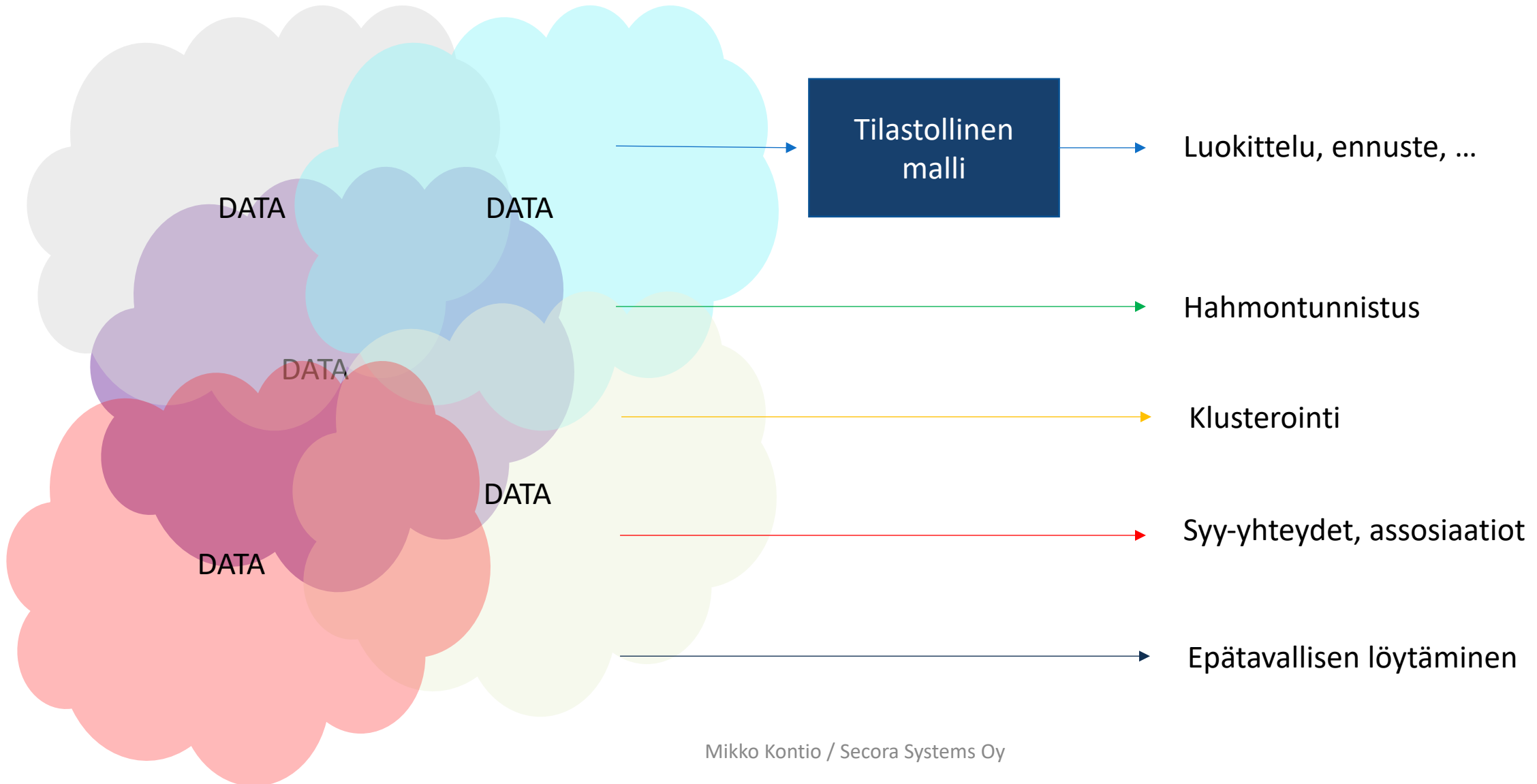


(KDD=
Knowledge Discovery in Databases,
Knowledge Discovery in data
Knowledge Discovery ...)



Lähde: <http://blogs.sas.com/content/subconsciousmusings/2014/08/22/looking-backwards-looking-forwards-sas-data-mining-and-machine-learning/>

Muutakin kuin ennustamista



TIEDONLOUHINTA (data mining)

Olennaisen etsimistä sopivin työkaluin

- Tilastotiedettä, tekoälyä, koneoppimista,...
- Tehdään siis paljon muutakin kuin vain *ennustetaan* tulevaa
- Hahmontunnistus, lainalaisuudet, assosiaatiot, syy-yhteydet, jne...
 - Koiranomistajat suosivat farmariautoja
 - Tofun ja hummuksen ostaminen nostaa banaanin ostamisen mahdollisuutta 87%
 - Yksi metri lisää rantaviivaa nostaan mökkitontin arvoa 10000 €:lla
 - Luottopäätökset, asiakaspysyvyys, petosten havainnointi,...

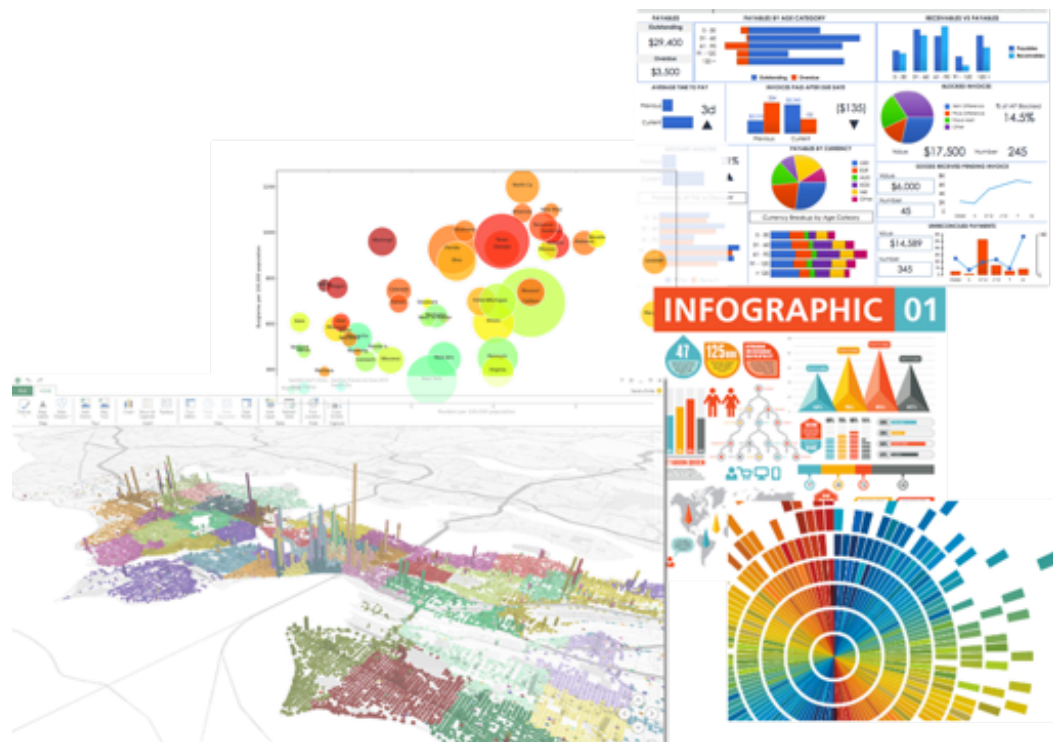
• Työkaluja:

- **Deep Learning**
- **Decision tree learning**
- **Association rule learning**
- **Artificial neural networks**
- **Inductive logic programming**
- **Support vector machines**
- **Clustering**
- **Bayesian networks**
- **Representation learning**
- **Similarity and metric learning**
- **Regression**

*Lisää esimerkkejä

- Detecting fraud at banks
- Identifying which consumers are most likely to respond favorably to: direct mail, Groupon deals, Facebook advertisements
- Trading stocks and derivatives
- Pricing insurance premiums (Google recently funded a startup that helps natural disaster insurance companies determine prices based on weather patterns due to global warming)
- Identifying human genes that make people more likely to develop cancer
- Predicting housing prices for real estate companies
- Predicting wine-tasting ratings
- Programmatically reading text from a random photograph
- Programmatically recognizing faces
- Autonomous, self-driving cars
- Anti-virus software (packet inspection)
- Factory maintenance diagnostics
- Delayed airplane flights
- Determining which voters to canvas during an election
- Developing pharmaceutical drugs (combinatorial chemistry)
- Predicting tastes in music (Pandora)
- Predicting tastes in movies/shows (Netflix)
- Search engines (Google)
- Predicting interests (Facebook)
- Predicting other books you might like (Amazon)

<http://www.quora.com/Machine-Learning/What-are-some-interesting-possible-applications-of-machine-learning>



5. Data storytelling

Tulkki lopputuloksen ja päätöksentekijän välillä

5. Tulosten esittäminen: “data storytelling”



- Visualisoinnin merkitys: *analysoinnin tuloksen pystyy sisäistämään sekunneissa*
- Vaatii todellista osaajaa (data storytelling skills)

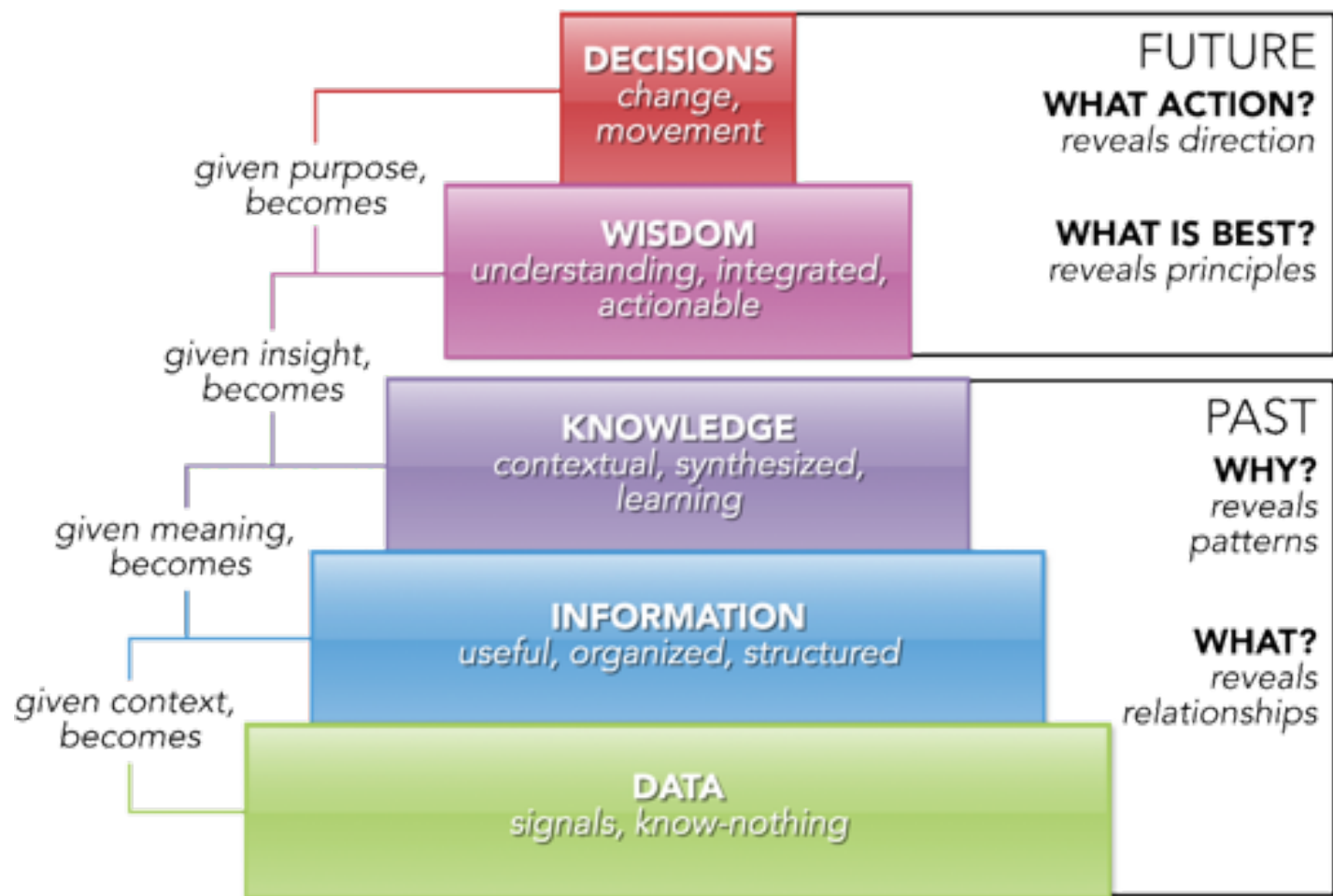
Kuva ja animaatio kertoo enemmän:

- <http://manpopex.us/>
- <http://guns.periscopic.com/?year=2013>
- <http://vimeo.com/28413747>
- <http://www.creativebloq.com/design-tools/data-visualization-712402>
- <http://www.informationisbeautiful.net/>
- <http://flowingdata.com/>
- https://www.ted.com/talks/hans_rosling_the_best_stats_you_ve_ever_seen?language=en#t-13109

Pohdintatehtäviä

1. Mitä dataa aiemmissa työpaikoissasi on kerätty sinusta (työntekijänä) tai asiakkaasta.
2. Katso [Hans Rosling -video](#) (alku – 4:40)
 - Mikä merkitys mielestäsi 'korostetulla/hyvällä' tarinankerronnalla voi olla viestin erottumiseen
3. Mitä uhkakuvia kaiken mittaamisella saattaa olla?
4. Mitä mielestäsi tarkoittaa sanonta: "Sitä saat, mitä mittaat".

Datan jalostuminen päätökseksi



Kypsyysmalli

Organisaation kypsyystaso analytiikan hyödyntämisessä (AMM)

- Ihmiset – miten laajasti analytiikkaa voidaan ja osataan hyödyntää
- Liiketoiminta – strateginen tahtotila sekä sen tuottama arvo
- Teknologia – ajantasaisuus, monipuoliset datan lähteet, laatu
- Kulttuuri – analytiikan käyttö päätöksenteossa ja käyttö yleisesti
- BI-organisaatio – rahoitus, osa-/kokoaikainen tiimi, sponsorit

	1. Started	2. Preliminary	3. Repeatable	4. Managed	5. Optimized
Ihmiset					
Liiketoiminta					
Teknologia					
Kulttuuri					
BI-hanke					

	1. Started	2. Preliminary	3. Repeatable	4. Managed	5. Optimized
Ihmiset					
Liiketoiminta					
Teknologia					
Kulttuuri					
BI-hanke					

	Taso 1	Taso 2	Taso 3	Taso 4	Taso 5
	Aloitettu	Alustava	Toistettava	Johdettu	Optimoitu
Ihmiset	Data avustaa joissakin yksilötason päätöksenteossa	Työntekijät voivat tehdä data-pohjaisia päätöksiä	Päätöksenteko on nopeampaa datan johdosta	BI:llä on positiivinen vaikutus henkilöiden tuottavuuteen	Organisaation tuottavuus on korkeampi suhteessa kilpailijoihin
Liiketoiminta	Hajallaan tapahtuva BI-kehitys	Keskitetty IT:n johtama BI	Näkyvä sponsori, alustava BI-strategia	BI-palvelukatalogi määritelty	Ladukas BI-strategia
Teknologia	Kuukausi-tason dataa, vähän datan lähteitä, exceleitä	Viikko-tason dataa, useita sisäisiä datan lähteitä, BI-palveluita olemassa	Päivä-tason dataa, useita sisäisiä ja ulkoisia datan lähteitä, dashboardeja ja DW	Oikea-aikaista dataa, ennustavaa analytiikkaa	Edelliset + rakenteeton data
Kulttuuri	Joillakin pääsy raportteihin, data ei yleisesti päätöksenteon tukena	Raporttien ja tunnuslukujen käyttö käytännön työn ongelmia ratkottaessa	Laajempi prosessien monitorointia	Dataa käytetään strategian monitorointiin ja tarkentamiseen	Automatisoitua päätöksentekoa
BI-hanke	Ei tiimiä, ei rahoitusta, ei sponsoreja	Yksiköiden omasta budjetista, paikallisia sponsoreja, edistetään pieninä projekteina	Virtuaalinen tiimi, yleinen ylemmän johdon tuki, edistetään projekteina	Yritystason budjetti, täyipäiväinen tiimi, ylin johto sponsoreina	Itsensä rahoittava, BI-osaamiskeskus

Organisaation kypsyystaso analytiikan hyödyntämisessä (AMM)

- Ihmiset – miten laajasti analytiikkaa voidaan ja osataan hyödyntää
- Liiketoiminta – strateginen tahtotila sekä sen tuottama arvo
- Teknologia – ajantasaisuus, monipuoliset datan lähteet, laatu
- Kulttuuri – analytiikan käyttö päätöksen teossa ja käyttö yleisesti
- BI-organisaatio – rahoitus, osa-/kokoaikainen tiimi, sponsorit

	1. Started	2. Preliminary	3. Repeatable	4. Managed	5. Optimized
Ihmiset					
Liiketoiminta					
Teknologia					
Kulttuuri					
BI-hanke					

	1. Started	2. Preliminary	3. Repeatable	4. Managed	5. Optimized
Ihmiset					
Liiketoiminta					
Teknologia					
Kulttuuri					
BI-hanke					