



# LUNDS UNIVERSITET

Ekonomihögskolan

*Institutionen för informatik*

---

## Användarbeteendedata i skymundan

En kvantitativ studie om unga vuxnas medvetenhet om och värdering av datainsamling på YouTube

Kandidatuppsats 15 hp, kurs SYSK16 i Informationssystem

Författare: Axel Nilsson  
Simon Jobs

Handledare: **Osama Mansour**

Rättande lärare: Björn Svensson  
Paul Pierce

# Användarbeteendedata i skymundan: En kvantitativ studie om unga vuxnas medvetenhet om och värdering av datainsamling på YouTube

ENGELSK TITEL: User-Behavioral Data in the Dark: A quantitative study on young adults awareness and valuation of data collection on YouTube.

FÖRFATTARE: Axel Nilsson, Simon Jobs

UTGIVARE: Institutionen för informatik, Ekonomihögskolan, Lunds universitet

EXAMINATOR: Christina Keller, Professor

FRAMLAGD: maj, 2021

DOKUMENTTYP: Kandidatuppsats

ANTAL SIDOR: 30

NYCKELORD: persondata, användarbeteendedata, datainsamling, YouTube

SAMMANFATTNING (MAX. 200 ORD):

Användandet av big data bland näringsidkare fortsätter öka exponentiellt vilket medför möjligheter men även flera risker. Datainsamling har infunnit sig centralt i gemene affärsmodell och data har beskrivits som den nya oljan. Andra sidan av myntet är att konsumenter har svårt att närvara online utan att lämna efter sig ett spår av data. Data som näringsidkare kan använda till att profilera användaren och sälja annonser till andra näringsidkare utan vetskap av användaren. I den här uppsatsen har vi valt att kolla närmare på två kategorier av data, persondata och användarbeteendedata genom att fråga konsumenter av tjänsten YouTube. De två faktorerna vi har tittat på är medvetenhet och värdering. Vi kommer fram till att medvetenheten om insamling av användarbeteendedata är väsentligt lägre än för persondata trots att konsumenter tenderar att värdera den högre. Vi hittar även ett samband mellan en högre medvetenhet och en högre värdering. Våra resultat pekar på att det existerar ett kunskapsgap gällande vilken data som samlas in och vi anser att det är problematiskt ur både moraletiska och monetära skäl.

## Innehåll

1	Inledning .....	1
1.1	Bakgrund .....	1
1.2	Problemformulering .....	2
1.3	Syfte .....	3
1.4	Forskningsfrågor .....	3
1.5	Avgränsningar .....	4
2	Litteraturgenomgång .....	5
2.1	Social media .....	5
2.1.1	YouTube .....	5
2.2	Digital profil .....	6
2.3	GDPR om data .....	6
2.4	Informationsdelning .....	7
2.5	Individen och medvetenhet .....	8
3	Metod .....	10
3.1	Val av metod .....	10
3.2	Val av plattform .....	10
3.3	Datainsamling .....	10
3.3.1	Spridning av formulär .....	10
3.3.2	Utformning av formulär .....	11
3.4	Analys av data .....	13
3.5	Validitet och Reliabilitet .....	14
3.6	Etiska överväganden .....	15
4	Resultat .....	16
4.1	Beskrivande statistik .....	16
4.2	Medvetenhet .....	18
4.3	Värdering .....	19
4.4	Sammanfattning av medelvärden .....	20
5	Diskussion .....	21
5.1	Analys och hypotesbesvarande .....	21
5.1.1	Medvetenheten för användarbeteendedata är lägre än vad den är för persondata. .....	21

---

5.1.2	Medvetenheten blir högre för användarbeteendedata och persondata efter vår informationstillförel.....	21
5.1.3	Medvetenheten påverkar hur individer värderar persondata och användarbeteendedata.....	21
5.2	Medvetenheten hos konsumenten .....	21
5.3	Skillnaden mellan persondata och användarbeteendedata .....	22
5.4	Medvetenhetens inverkan på värderingen.....	23
5.5	Begränsningar och framtida studier .....	24
6	Slutsats .....	25
	Appendix I: Frågeformulär.....	26
	Appendix II: Medvetenhet alla punkter .....	28
	Appendix III: Samtliga värderingar .....	29
	Referenser.....	30

## Figurer

Figur 1: Åldersfördelning för deltagarna i studien.....	16
Figur 2: Andel deltagare per genomsnittlig tid spenderad på YouTube per vecka.....	17
Figur 3: Fördelning över oplanerat till planerat tittande på YouTube .....	17
Figur 4: Andel av deltagarna som svarat att de använder YouTube på respektive sätt .....	18
Figur 5: Hur många procent av deltagarna som har angett x antal punkter per kategori innan och efter informationstillförel.....	18
Figur 6: Medianen av deltagarnas fem värderingar per kategori fördelat över kategorierna...	19

## Tabeller

Tabell 1: Informationspunkter inkluderade i studien uppdelat per kategori .....	12
Tabell 2: Medelvärde och standardavvikelse över medvetenhet .....	19
Tabell 3: Medelvärde och standardavvikelse över värdering.....	19
Tabell 4: Sammanfattning över medelvärden .....	20

# 1 Inledning

## 1.1 Bakgrund

År 2018 omsatte big data cirka 42 miljarder dollar, 2027 förväntas branschen omsätta cirka 107 miljarder, mer än en dubbling under en tioårsperiod. Anledningen till detta är främst den ökade användningen av mobiltelefoner, molnlösningar, utvecklingen av artificiell intelligens samt en ökad användning av uppkopplade produkter (Statista, 2018). Internationella teknikbolag som Apple, Google, Netflix och Amazon är också en starkt bidragande till den ökade användningen och branschens tillväxt. Med hjälp av olika tjänster som företagen tillhandahåller möjliggör de samtidigt för insamling av stora datamängder, ett exempel på det här är Googles videoströmningstjänst YouTube. Konsumenterna och användarna är också påverkade av denna utveckling, i rapporten *The Trust Advantage: How to Win with Big Data* från Boston Consulting Group så konstateras det att dataintegritet är ett högprioriterat område för 75% av konsumenterna. Samtidigt visar trenden att den yngre generationen inte är lika oroad och delar sin privata information på internet till en högre grad än vad tidigare skett (Rose et al. 2013).

Ett område som påverkas starkt av ökad tillgång och användning av big data är marknadsföring. Företag upplever att en av deras största utmaningar är att förstå kunder och användares beteenden (Forrester, 2014). Data kan användas för att hjälpa till att förstå beteenden och utnyttja dem till företagets fördel. Forrester listar ett antal insikter i rapporten, *Marketing's Big Leap Forward*, där data kan användas till att utvinna kunders beteende och därmed genomföra mer framgångsrik marknadsföring. Bland annat kan kunder få personliga erbjudanden och information, förstå vilka plattformar och system kunden interagerar på samt förstå kunders framtida beteende (Forrester, 2014).

Dagens digitala landskap öppnar upp flertal sätt för intressenter att samla in data. I internets begynnelse var applikationer och hemsidor de stora källorna för data (Morey et al. 2015). Dessa metoder används även idag men kompletteras med allt mer avancerade metoder, till exempel så går det att spåra en persons plats och dess användning av olika tjänster mer precist idag. Den mer personliga anpassningen av data och dess insamling öppnar upp både för nya möjligheter men även utmaningar. Möjligheterna ligger till stor del som tidigare nämnt i en mer riktad marknadsföring, men även tjänster där konsumentens egna data används för att förbättra den elektronik som konsumenten använder (Morey et al. 2015). Utmaningarna ligger främst i hur data kan skyddas från att hamna i fel händer och för att inte göra användaren eller konsumenten någon skada. En stor risk är läckor där betaluppgifter kommer ut till grupper eller individer som vill utnyttja konsumenterna är ett exempel på en aktuell utmaning. Även att data kan användas på ett sätt som inte användare vill eller har godkänt den för. Den amerikanska kedjan Target förutspådde en kunds graviditet utifrån den data de hade samlat in, innan kunden själv var medveten om det vilket väckte uppmärksamhet och oro (Morey et al. 2015).

Eftersom datainsamling sker överallt lämnar användare efter sig ett digitalt fotspår. I rapporten *The Value of Our Digital Identity* från 2012 av Boston Consulting Group (BCG) så presenteras begreppet "digital identitet", vilket är den profil som individer idag har online parallellt med sitt verkliga liv. I denna profil finns all den data som finns om en specifik individ. Denna profil blir allt mer komplett och precis, drivet av den digitala och tekniska utvecklingen (Rose et al. 2012).

Data kan delas in i kategorier, två vanliga kategorier berörs i den här studien. Den första, persondata kan beskrivas som information som användaren delar med sig till en tjänst redan innan användning, namn, ålder, e-mailadress är exempel på dessa (Larsson & Ledendal, 2017). Den andra typen av data är användarbeteendedata, till exempel kan detta innefatta kommentarer, gillningar, visningar och klick. Den skiljer sig från persondata genom att den genereras av beteenden och användning. Med hjälp av dessa typer av data samlas en bild av en person ihop och en digital identitet skapas (Yang, 2010). Med tillräckligt mycket data av båda karaktärer, persondata och användarbeteendedata kan man skapa målgrupper för bland flera andra: ålder, kön och karaktärsdrag av den föreställda personen. All data är relevanta i ett marknads-perspektiv och är med dagens beräkningskraft, bandbredd och lagringsmöjligheter lönsamma att samla in ur en näringsidkares perspektiv (Larsson & Ledendal, 2017).

Att samla in data från konsumenter är lagligt och som konsument godkänner du det varje gång du godkänner cookies på en hemsida eller blir medlem på en sida. Inget är gratis är något många får lära sig i tidig ålder och samma sak gäller till stor grad i dagens digitala samhälle. Är en tjänst gratis är risken väldigt hög att du betalar med din data (Larsson & Ledendal, 2017). Via plattformar som inte syns för gemene konsument sker tusentals transaktioner av data mellan näringsidkare varje minut. Vilken reklam som visas var och för vem har utvecklats och det finns mer avancerade metoder än att välja vilka busskurer ens plakat ska sitta på. Komplexa algoritmer räknar ut hur mycket en person är värd beroende på dess digitala identitet och sätter därefter ett pris på dess data. På andra sidan finns en annan algoritm som beräknar ifall samma person är värd den summan för sin näringsidkare och om svaret är ja, slår den till. En transaktion sker och en annons visas för en intet ont anande konsument på andra sidan skärmen (Google Adsense, 2021).

En undersökning gjord på ett mindre företag i Australien visade att genom datadriven marknadsföring kunde företaget öka försäljningen 23 procent (Walsh, 2017). Detta genom att kombinera den egna kunddatan med tredjepartsdata. Undersökningen visar att små insatser med rätt data kan öka ett företags försäljning. Priset för detta är kunder och individers data, både den som generas av individen själv, persondata, men även den som uppstår i samband med att en individ helt enkelt använder sig om olika digitala tjänster, användarbeteendedata.

Vi vill i den här uppsatsen undersöka vilka faktorer som påverkar konsumentens värdering av deras data för att bidra till en ökad medvetenhet kring datahandel och samtidigt undersöka om konsumenter gör en skillnad på persondata och användarbeteendedata.

## 1.2 Problemformulering

Datainsamling har under en längre tid blivit normaliserat och frodas i konsumenters okunskap. Näringsidkare har insett värdet som data kan generera via insikter. Wal-Mart, en amerikansk jätte inom livsmedelsförsäljning och detaljhandel rapporterades under 2012 att samla in 2.5 petabytes av data varje timme. Den insamlade datan används för att öka försäljning inom verksamheten eller säljs vidare. Det har byggts upp hela marknader på att sälja information om och kring personer utan deras vetskap, eller med minimal vetskap (McAfee & Brynjolfs-son, 2012). Det finns dels en monetär aspekt av det hela där konsumenten utan full förståelse går med på en transaktion och är ovetande om dess egentliga värde men även en moralisk aspekt där det saknas en medvetenhet om vilken data som finns om sin person och vem som har tillgång till den (Smith et al. 2012; Godel, Litchfield & Mantovani, 2014).



I en studie av Spiekermann och Korunovska (2016) undersöker de olika faktorer som kan påverka konsumentens värdering av sin persondata. Mer specifikt undersökte de persondata som aktivt delats på sociala medieplattformen Facebook och sätter persondata i kontext av att vara en tillgång. De konstaterar att marknadsmedvetenhet har en markant påverkan på värdering av persondata. De noterar även att psykologiska teorier som ägandeskap och tillgångsmedvetenhet kan ligga till grund för en högre värdering. Studien pekar som tidigare nämnt på en icke existerande till otillräcklig förståelse för närvaro av marknader där persondata är varan som köps och säljs.

Larsson och Ledendal (2017) tar upp en del av problematiken kring insamlingen av konsumentdata till affärsmodeller som är uppbyggd kring att samla in, lagra och sälja data till bland andra, reklambranschen, snarare än att sälja en specifik produkt. Det kan uppfattas som missvisande då datainsamling ofta sker via en annan tjänst i fokus. De nämner även flera globala företag inom den digitala ekonomin såsom Amazon, Spotify, och Google vilkas affärsmodell bygger på att samla in användarbeteendedata som primärt kapital snarare än persondata. Med hjälp av datautvinningsmetoder kan man uttyda mönster och trender långt mer värdeskapande än rå persondata (Ghani et al. 2019; Acharya et al. 2018). Insamling av användarbeteendedata sker till större del passivt kontra det aktiva valet av att dela persondata i enlighet med GDPR. Det här leder till ytterligare påbyggnad på problematiken kring konsumentens bristande kännedom kring försäljning av konsumentdata.

Tidigare forskning av konsumenters värdering av data och vilka faktorer som påverkar värderingen har fokuserat på persondata. Vår uppfattning är att det speglar sig i den allmänna medvetenheten kring användarbeteendedata och att det finns en bristande medvetenhet om insamlingen av användarbeteendedata. Då vi personligen anser att det finns både moraliska samt monetära incitament till att vara medveten om olika sorters datainsamling är detta problematiskt. Vi vill därför i den här uppsatsen fokusera på medvetenhet och värdering som konsumenter har om sin användarbeteendedata i relation till tidigare utforskade persondata.

### 1.3 Syfte

Denna studie har till syfte att jämföra de två datakategorierna persondata och användarbeteendedata och identifiera hur de skiljer sig åt utifrån konsumenters medvetenhet och värdering, samt att undersöka om medvetenhet om datainsamling är relaterad till dess värdering.

### 1.4 Forskningsfrågor

Forskningsfrågan har delats upp i tre stycken frågor för att belysa perspektiven: initial kunskap, skillnad mellan kategorier och samband mellan kategorier.

1. Hur ser konsumentens medvetenhet och värdering av data som samlas in ut?
2. Hur skiljer sig medvetenheten och värderingen mellan persondata och användarbeteendedata?
3. Hur påverkar medvetenhet om insamlade data dess värdering om densamma?

## 1.5 Avgränsningar

Data kan kategoriseras och benämnas ur många olika perspektiv. I den här uppsatsen har vi valt att fokusera på två kategorier av data, persondata och användarbeteendedata. I studien avser vi att mäta medvetenhet samt värdering för de två datakategorierna. Måttet på medvetenhet är en uppskattning av ett satt urval av informationspunkter och deltagarna i studiens medvetenhet om de specifika punkterna. Värdering i den här studien avser deltagarens uppfattning om olika datapunkter som samlas in på YouTube vilka blir presenterade under metodkapitlet.

Vår målgrupp i studien är konsumenter av den sociala media och videoströmningstjänsten YouTube. YouTube erbjuder en tjänst där användare själva kan lägga upp videor som tittas på av andra användare men majoriteten av användare tittar endast på videor via tjänsten. Anledningen till att vi väljer YouTube framför annan social media är att de har stort fokus på användarbeteendedata och mängden persondata som samlas in via tjänsten i övrigt är förhållandevis låg. Åldersgruppen som vi har valt att undersöka i den här studien är 18–29 år (födelseår mellan 1992 och 2002) och fångas i den demografiska kohorten Generation Z med födelseår mellan 1993 och 2005. Gen Z karakteriseras bland annat av en ökad användning av och familjaritet till internet, sociala medier och smarta enheter. De har blivit kallade för “digitala infödingar”, en term för att beteckna att de vuxit upp med internet och digitala medel som en väsentlig del av samhället (Turner, 2015). Grunden till vår åldersavgränsning är att vi vill fokusera på konsumenter som har varit med längs utvecklingen av YouTube och samtidigt begränsa oss till unga vuxna. Det är även därför vi har valt att göra en mindre justering till kohorten Gen Z.

YouTube ingår i ett bredare sortiment av ihopkopplade tjänster som erbjuds av Google. Det innebär att det finns en bred överlapp i konsumenter av tjänsterna och att data om dessa delas fritt mellan tjänsterna (Google, 2021). Uppsatsen är formad utefter plattformen YouTube för att kunna bygga upp realistiska scenario och relaterbara informationspunkter för deltagare i undersökningen. Vidare benämns data som delas mellan Googles tjänster i uppsatsen som tillhörande YouTube.

## 2 Litteraturgenomgång

### 2.1 Social media

År 2004 kom de första sociala medierna och vad som benämns Web 2.0 föddes. I samband med detta började en ny typ av innehåll dyka upp, den som skapas av användaren. Tre stycken faktorer kan användas för att definiera vad som är ett socialt media. Först ska användaren kunna skapa en personlig profil som är öppen. Användaren skall sedan kunna hitta och kopplas ihop med andra användare. Slutligen skall användaren kunna interagera med andra användare genom exempelvis gillningar och kommentar (Ghani et al. 2017). Exempel på sociala medier är Facebook, Instagram, Twitter, LinkedIn och YouTube.

#### 2.1.1 YouTube

YouTube är en tjänst där den huvudsakliga idén är att ladda upp och dela videoklipp med andra användare. Företaget är ett dotterbolag till söktjänsten Google som i sin tur ägs av Alphabet Inc. Det konto som en individ skapar på YouTube är gemensamt för alla tjänster som Google erbjuder, det finns dock möjlighet att använda YouTube utan ett konto även om möjligheterna till att ladda upp egna filmer eller att interagera då inte är tillgängliga. Övriga funktioner förutom att ladda upp och titta på videoklipp är att föra diskussioner och driva så kallade kanaler, där användare kan lägga upp flera videoklipp i ett flöde. Bland svenska internetanvändare ska 87 procent av dessa tittat på YouTube under 2019 enligt en undersökning gjord av Internetstiftelsen. Cirka var tredje användare tittade varje dag på YouTube och de vanligaste kategorierna att titta på var underhållning, följt av instruktionsvideos och musik (Svenskarna och Internet, 2019).

Den primära inkomstkällan för YouTube är genom annonser vilka kan visas både innan, under och efter ett videoklipp spelas upp. Annonser kan även visas runt om videospelaren. Det finns möjlighet för användare att betala en månadsprenumeration, vilket gör att användaren inte behöver titta på någon reklam (YouTube, 2021a). Enligt YouTube's egen beskrivning av annonsinställningar så är deras mål att visa relevanta och användbara annonser baserat på användarens aktivitet. Vilka annonser som visas baseras dels på vad för information som en användare har angett i sitt Google-konto, dels på vad YouTube gissar om användarens intressen och interaktioner med vad de kallar andra annonsörer. Ett utdrag från YouTube's hemsida beskriver deras metoder för att visa annonser:

Vi kanske till exempel tror att du är fotbollsfantast eftersom du nyligen har tittat på höjdpunkter från en match på YouTube eller sökt på ”fotbollsplaner nära mig” på Google Sök. Och om du har tillbringat tid på en annons-samarbetspartners webbplats kan vi föreslå annonser baserat på besöket (YouTube, 2021a).

På samma sätt som YouTube samlar in information och gissar vad för typer av intressen en konsument har och visar annonser som den kan tänkas vara intresserad av, ger YouTube också förslag på vilka videor som konsumenten kan tänkas vilja se. Syftet med att samla in information om konsumentens aktivitet på YouTube och andra Google-tjänster blir därför dubbelt (YouTube, 2021a).

## 2.2 Digital profil

I rapporten som nämndes i inledningen av Boston Consulting Group har idag i princip alla som använder internet en digital profil. Den digitala profilen fylls på av data eller information som är en hörnsten i hur dagens digitala samhälle har utvecklats och kommer att utvecklas. Data finns i princip överallt och när vi spenderar allt mer tid online så lämnar vi allt fler spår efter oss. Mobiltelefoner har drivit på utveckling av digitala spår, bland annat genom geo-lokalisering samt det faktum att tillgängligheten för digital konsumtion ökat (Loukides 2011). All data behöver dock sparas på något sätt. Den tekniska utvecklingen har gjort att det blivit 40000 gånger billigare att köpa 1MB lagringskapacitet idag än vad det var i början av 80-talet, priset har sjunkit exponentiellt (Rosenthal et al. 2010). En utveckling som bidragit till att det överhuvudtaget är realistiskt att samla in stora mängder data.

Insamlad data kan sedan användas för profilering, processen av att hitta samband i data som kan användas för att identifiera eller representera en fysisk eller icke-fysisk person, individ som grupp (Shermer, 2011). Med hjälp av datautvinningsmetoder automatiseras processen och flera samband kan kartläggas för att hitta mönster i databaser. Tekniken kan ha stora konsekvenser gentemot integritet när den används för att gissa, eller snarare hitta sig till personlig information som karaktärsdrag eller intressen (King & Forder, 2016). Konsumentprofilering, en subgenre med fokus på att kartlägga köp och beteendemönster i syfte att rikta allt mer specifika annonser och därmed höja inkomsten per spenderad krona på marknadsföring leder till ytterligare dilemma (King & Forder, 2016).

Det finns två huvudsakliga typer av data som Google och andra företag samlar in, persondata och användarbeteendedata (Google, 2021). För att jämföra de olika typerna av data och hur insamlingen av den har förändrats kan vi säga att Y vill veta vilka webbplatser X besöker, i syfte att förstå sin kund bättre. Tidigare har datan samlats in genom att fråga X vad för webbplatser den besöker. Med hjälp av användarbeteendedata som samlas in, kan nu Y direkt se vilka webbplatser X besökt genom att Y kollar på X webbhistorik istället för att fråga X direkt. Genom den utökade datainsamlingen har företag alltså möjligheten att spåra potentiella kunder genom deras beteende och vanor utan transparens för att ett informationsutbyte har skett (King & Forder, 2016).

## 2.3 GDPR om data

GDPR är en förkortning på General Data Protection Regulation, på svenska dataskyddsförordningen. Förordningen gäller inom hela EU och syftar till att skapa en likvärdig och enhetlig nivå för skyddet av personuppgifter (Dataskyddsförordningen, 2016). Personuppgifter definieras i artikel 4 som:

varje upplysning som avser en identifierad eller identifierbar fysisk person (nedan kallad en registrerad), varvid en identifierbar fysisk person är en person som direkt eller indirekt kan identifieras särskilt med hänvisning till en identifierare som ett namn, ett identifikationsnummer, en lokaliseringssuppgift eller onlineidentifikatorer eller en eller flera faktorer som är specifika för den fysiska personens fysiska, fysiologiska, genetiska, psykiska, ekonomiska, kulturella eller sociala identitet,

Integritetsskyddsmyndigheten, som granskar efterlevnaden av förordningen i Sverige, beskriver personuppgifter som all information eller som kan knytas till en levande person. Vilket kan vara namn, adress och personnummer. Det måste också finnas samtycke för att personuppgifter skall kunna samlas in.

Information som inte kan knytas till en specifik individ omfattas dock inte av förordningen. Alltså omfattas inte användarbeteendedata genererad av en individ av dataskyddsförordningen (Dataskyddsförordningen, 2016).

Dock innebär GDPR att de flesta företag som hanterar eller samlar in data behöver förändra sina metoder för detta. Speciellt de företag där riktad reklam spelar en stor roll för deras affärsmodell. Eftersom GDPR gör att all data som skall samlas in måste aktivt godkännas av individen istället för att passivt samla in detta som tidigare, blir det tydligare för individen att dessa data samlas in och bearbetas (Ghosh, 2018). Ghosh (2018) menar i sin artikel *How GDPR Will Transform Digital Marketing* att detta samtycke till datainsamling inte är självklart för individen. Författaren lyfter även fram en ny typ av reklam som inte baseras på användarbeteendedata utan på kontext som en möjlig lösning för företagen.

Det vill säga att reklamen inte baseras på vad företagen vet om individen utan på vad individen faktiskt tittar på just nu. Tittar individen på en film om bilar på YouTube är det alltså stor sannolikhet att det dyker upp reklam om just bilar.

Det finns dock fortfarande en framtid för användarbeteendedata enligt Ghosh (2018), dels eftersom företag fortfarande får samla in data, och dels för att förordningen endast gäller inom Europa. Vidare skriver författaren att i teorin blir GDPR ett medel för att göra datainsamling och riktad reklam mer jämlikt, framförallt för slutanvändare. Det är också troligt att om GDPR anses vara en lyckad förordning, kan det komma fler, men även att det kan komma liknande förordningar i andra delar av världen och inte bara inom Europa (Ghosh, 2018).

## 2.4 Informationsdelning

Företag som samlar in data arbetar aktivt med att få konsumenter att dela med sig av mer av deras data. Ur ett företagsperspektiv handlar det till stor del om att få konsumenter att lita på att deras data inte används på sätt som de inte har godkänt eller finner nödvändigt (Mohan Rao et al. 2018). Riskerna som konsumenterna ser främst kopplat till att deras data används på ett sätt som de inte har godkänt eller känner till. Åtgärderna som presenteras för att bygga upp detta förtroende främst kopplat till att företagen skall vara transparenta med hur de använder konsumenternas data, möjliggöra att kunna få ta bort sin data, samt att presentera översiktliga och lättillgängliga användaravtal (Mohan Rao et al. 2018).

Att individer delar sin information till företag bygger till stor del på att de får tillgång till en tjänst eller en produkt, individen ser en större fördel än vad den ser risk. Lee et al. (2013) tar i en studie upp användares beteenden kopplade till fördelar och risker med att dela sin data med sociala medier. Författarna tar upp fem olika typer av fördelar som kopplas till sociala medier, ett sätt att uttrycka sig, social kontroll, relationsbygge, social bekräftelse samt självbekräftelse (Lee et al. 2013). De tar även upp fem riskfaktorer kopplat till sociala medier, fysisk och ekonomisk risk, stigma, relationer, trovärdighet och roller. Individer tenderar att söka fördelar och undvika risker, varför tre olika risk-reducerade metoder presenterades inför deltagarna i studien. a) individen väljer vem den skall dela informationen med. b) användaren utesluter viss detaljerad information kopplat till vad den väljer att dela för något. c) användaren identifierar vissa tidpunkter på sociala medier när aktiviteten är låg respektive hög, och därmed hur synligt deras information blir i förhållande till övrig information.

Slutsatserna som författarna kunde dra var att förväntade fördelar med att dela sin information överträffade de förväntade riskerna. Även att deltagarna tenderade att fokusera på riskerna över fördelarna när de delade med sig av vad de ansåg vara riskfylld information. Vidare såg deltagarna till att garantera fördelarna med sociala medier i första hand när risken ansågs vara optimal.

Fenomenet med att optimera fördelarna samtidigt som att inte göra det med riskerna kan förklaras med begreppet "Privacy Paradox" som kan översättas till "integritets paradox" och syftar på dikotomin mellan hur individen uppfattar sin ställning till insamling av persondata och dess faktiska beteende vid densamme. I en studie utförd av Norberg et al. (2007) blev deltagare i en första fas ombudda att uppge vilken information de är villiga att dela med sig. När man efter tolv veckor utklädda till tjänstemän på en bank åter frågade deltagarna visade det sig att de i snitt delade med sig av dubbelt så många datapunkter i ett realistiskt scenario än de först uppgett. Studier relaterade till Privacy Paradox har inte bara gjorts på persondata som namn och adress utan även i relation till geodata och bilder. Teorier inom psykologi och ekonomi behöver i många fall kombineras för att förklara beteendemönstret (Spiekermann et al, 2001; Beresford et al, 2012; Zafeiropoulou et al, 2013).

Senare forskning har kritiserat användandet av termen paradox i sammanhanget då det finns faktiska om än komplexa förklaringar till fenomenet. Till exempel så är integritetsbeteende högst kontextuellt och bör därför ej jämföras mellan situationer som skiljer sig åt även om de delar vissa karaktärsdrag. Därför blir det svårt att dra paralleller som i tidigare studier ansetts tillräckliga för att påvisa ett visst beteende. På grund av den kontextuella karaktären blir det i många situationer en tolkningsfråga samt att olika individer värderar olika persondata annorlunda (Kokolakis, 2017).

## 2.5 Individen och medvetenhet

För att ytterligare förklara Privacy Paradox, medvetenhet och hur data samspelar ger artikeln *Privacy Awareness: A Means to Solve the Privacy Paradox?* av Stefanie Pöttsch (2008) en del förklaringar. Artikeln beskriver medvetenhet som en individs uppmärksamhet, eller förmåga att uppfatta information, såväl fysisk som icke-fysisk. När en individ inte blir påmind eller uppmärksam på information tenderar informationen att försvinna och därmed blir medvetenheten lägre. Det finns två sätt som individer eller konsumenter kan bli uppmärksammade på vad för data som samlas in om dem, dels generell information, dels applikationsspecifik information.

Den generella informationen kan beskrivas som saker som riktar sig till den stora massan, det kan vara informationskampanjer, föreläsningar eller guider. Applikationsspecifika är exempelvis att individen utsätts för information om vad för data som individen delar med sig när den använder en hemsida eller applikation, eller godkänner så kallade "cookies" eller web-trackers (Pöttsch, 2008). Artikeln tar också upp att individer glömmar bort eller inte vet vad för information de delar med sig på sociala medier eller webbshoppar. Detta är en del i den paradox som beskrivits tidigare och författaren menar att individer behöver vara minst dubbelt så medvetna om vad för risker de tar när de delar med sig av information på nätet kontra hur medvetna de är om fördelarna med att göra det (Pöttsch, 2008).

För att kunna öka medvetenheten och individer finns det enligt artikeln två alternativ, antingen får individers beteende ändras så att de stämmer överens med deras inställning eller tvärtom. Det första alternativet föreslås som det bästa eftersom det enligt författaren leder till en förbättrad integritet för individen. Hur detta skall göras är genom att påminna individen i en större grad än vad som görs idag. Det största problemet med detta beskrivs i artikeln inte att presentera information till användaren, utan att få den att förstå vad den tittar på. Inte heller vad för information som samlas in, utan vad den faktiskt används till beskrivs som ett stort problem. Eftersom individer generellt inte har en särskilt stor kunskap om hur data fungerar räcker det enligt författaren inte bara informera, individer måste förstå (Pözsch, 2008).

En annan studie som undersökt medvetenhet hos användare av Facebook upptäckte att medvetenhet inte ökade integriteten. Studien syftade till att öka medvetenheten om datainsamling på Facebook och belysa användarna om riskerna med att dela sin data med Facebook. Deltagarna informerades om riskerna och fick svara på en undersökning deras data. Före undersökningen dokumenterades vilka inställningar de svarande hade, det vill säga om hur "öppen" deras profil var. Detsamma gjordes efter de svarande hade genomfört undersökningen, då framgick att deltagarna i studien endast hade gjort minimala förändringar gällande vilken data som delades (Steijn, 2014).

Anledningarna till detta kan enligt författaren av studien vara att deltagarna känner sig bekväma i vilken data de delar samt att de inte ser någon direkt risk med att dela denna, trots att undersökningen informerade om just dessa. Det kan enligt författaren också vara att deltagarna eller någon i deras närhet måste utsättas för någon av dessa risker först, exempelvis identitetsstöld för att de ska vara benägna att se över vad de delar för data (Steijn, 2014).



## 3 Metod

### 3.1 Val av metod

För undersökningen har vi valt att använda oss av en kvantitativ metod. Anledningen att vi valt en kvantitativ metod för datainsamling grundar sig i vårt syfte samt våra forskningsfrågor. Då vi vill jämföra värderingar mellan persondata och användargenererad data vill vi ha en större datamängd att utgå ifrån. Ett större urval ökar även sannolikheten att vi representerar hela populationen och att undersökningen ska gå att upprepa med liknande resultat (Oates, 2006). Vi anser det även som den bäst lämpade metoden då vi vill få bättre insikt i ett redan existerande forskningsområde snarare än att söka efter nya områden.

I mer detalj så syftar vår kvantitativa metod på att samla in fakta och värderingar utifrån svarsalternativ som vi har utformat i förväg. På det sätt försäkras vi oss om att samtliga svar är kategoriserat på ett lämpligt sätt och gör det lättare att bearbeta och finna mönster i svaren. För att undersöka deltagarnas värdering har vi utformat frågor som är designade med svarsalternativ utefter en fempunkts Likert-skala. Vår Likertskala sträcker sig från 1-5, med titlarna "Väldigt lågt" till "Väldigt högt" där en trea ska uppfattas som neutralt, en fempunktskala har tillsammans med 4, 6 och 7 punkts-skala bäst stöd för reliabilitet och validitet. Mer specifikt har en fempunkts skala valts för att koncentrera värdet i det neutrala alternativet samt minska risken för att en deltagare börjar ta frågorna på mindre allvar vid för många alternativ så kallat "form fatigue" (Dawes, 2008). Sista frågan i formuläret är ställd öppet och besvaras i fritext. Där vi ber om deltagarnas kompletterande åsikt om datainsamling och ger de själva valet att svara kort, långt eller inte alls.

### 3.2 Val av plattform

Avgränsningen till en enskild social media kom naturligt då det ofta är stora skillnader både i utformningen av olika sociala medier men specifikt i hur de används (Ghani et al. 2019). Anledningen till att har valt att rikta in oss på just YouTube är att det är en tjänst som fokuserar på att samla in användarbeteendedata då den är utformad med begränsade möjligheter för användargenererade data, till exempel kommentarer, som konsument av tjänsten. Det passar vårt syfte med studien samt våra frågeställningar väl.

### 3.3 Datainsamling

#### 3.3.1 Spridning av formulär

I syfte att samla svar på formuläret så har vi spridit det i författarnas kretsar främst via social media i form av Facebook. Där har formuläret delats via författarnas egna tidslinjer samt i klassgrupper för Systemvetenskapliga programmet vid Lunds universitet med examensår 21, 22 respektive 23. Formuläret låg uppe i 10 dagar, den 9/4/2021 till 19/4/2021 och spreds med en förklaring om att författarna var i färd med att skriva sin kandidatuppsats och sökte svar på ett formulär för att undersöka medvetenhet och värdering av person och



användarbeteendedata på YouTube. Övrig spridning har skett i form av länkar i privata chatt och mailgrupper genom uppmaningar att gå in och titta på publiceringen i endera författares tidslinje.

### **3.3.2 Utformning av formulär**

Vi har utformat formuläret i sex huvudsakliga delar. Varje del har ett eget huvudsakligt syfte med frågor som är anpassade för att ge oss god möjlighet till insikt i deltagarnas svar samt för att kunna göra en analys på de svar som samlas in. Vidare är varje fråga noggrant formulerad enligt rekommendationer från Gustafsson et al. (2006) där varje fråga ska vara, koncis, relevant, entydig, tydlig och objektiv. Del 3, 4 och 5 är centrala till undersökningen och är uppbyggda så att vi ska kunna samla in deltagarnas initiala medvetenhet och värdering. I del 6, efter att ha tillfört information om vilken faktisk information som YouTube har på deltagaren kan vi sedan se om deras medvetenhet har ökat och hur deras värdering har förändrats i relation till det. I del 3 och 5 används en likertskala för att bestämma värdering. Skalan är satt från "Väldigt lågt" till "Väldigt högt" och är subjektiv mot deltagarnas egen tolkning. Då vårt syfte är att titta på hur värderingen ser ut i förhållande till varandra och inte att sätta en exakt värdering anser vi att det inte är något problem.

#### **3.3.2.1 Kvalificering (1, 2)**

Syftet med första delen av undersökningen är att kvalificera deltagarna som giltiga för vår undersökning. Vi frågar efter ålder på deltagaren i form av åldersgrupper, detta är för att kunna analysera resultatet utifrån vår målgrupp samt för att ge oss en uppfattning om spridningen av formuläret och underlätta för återskapning av undersökningen. Då ett konto på YouTube krävs för att delta i undersökningen frågar vi det här och avslutar formuläret ifall deltagaren svarar nekande på frågan, annars går deltagaren vidare till nästa del.

#### **3.3.2.2 Användning (3, 4, 5, 6)**

De fyra frågornas syfte är att ge oss bättre förståelse för hur deltagaren använder sig av plattformen YouTube. Var för sig ger de oss grundläggande information om användarens beteende och omfattningen av användandet. Tillsammans kan vi gissa oss till om YouTube har en mer omfattande bild av användaren eller inte.

#### **3.3.2.3 Medvetande och värdering (7, 8)**

Syftet med den här delen av undersökningen är att få svar på dels vilken medvetenhet deltagaren har om den data som YouTube samlar in och dels hur deltagaren värderar samma data oavsett om vad de tror om insamlingen för data. Den här delen av datainsamlingen lägger grund för den data vi är som mest intresserade av när vi ska analysera resultaten.

De tio olika data eller information som vi har valt som svarsalternativ i undersökning är samtliga utvalda ur ett flertal kriterier och visas i tabell 1. Först och främst så vet vi om att YouTube samlar in samtliga av alternativen om de har möjligheten till det. De ska även vara lätta att förstå vad de innebär samt vara valda utan någon större vikt lagd vid deras innebörd från vår sida.

**Tabell 1:** Informationspunkter inkluderade i studien uppdelat per kategori

Persondata	Användarbeteendedata
Namn	Civilstatus
Kön	Språk
Ålder	Föräldrastatus
Land	Matpreferenser
Telefonnummer	Fritidsintresse

De tio alternativen är även uppdelade i två grupper om fem. Uppdelningen är en direkt följd till vårt syfte med undersökningen då vi vill undersöka skillnaden mellan medvetenhet och värdering av persondata och användarbeteendedata. Vidare så är även indelningen i våra två grupper tagen från hur Google kategoriserar dem (Google, 2021). Vi är i slutändan inte intresserade utav resultaten från de individuella alternativen utan det sammanfattade resultatet för gruppen. Främsta anledningen till det är att individuella Likert-skolor kan uppfattas som svaga och praxis är därför att kombinera flera liknande frågor i grupperingar för att stärka trovärdigheten på resultaten från skalan (Uebersax, 2006).

#### 3.3.2.4 *Informationstillförsel*

Den här delen består av instruktioner för att ta del av vår informationstillförsel samt en fråga för att bekräfta att deltagaren har tagit del av informationen. Frågan tillhör ej undersökningen utan är endast en del av formuläret.

Informationstillförseln är en länk till Googles annonsinställningar. Vi ber deltagaren att följa länken till den officiella sidan där de blir ombedda att logga in om ej redan inloggade. Vi ber användaren att vara inloggad på det konto de främst använder för YouTube. Här blir deltagaren presenterad med olika taggar som Google har användaren.

När användaren klickar på en av taggarna kommer det upp en ruta som anger ifall deltagaren själv har gett Google den här informationen eller hurvida Google har gissat sig till den utifrån användning och beteende. Syftet med informationstillförseln är att ge användaren perspektiv på deras svar om medvetenhet och värdering i föregående del.

#### 3.3.2.5 *Justering av medvetande och värdering (9, 10)*

Frågorna 9 och 10 är identiska 7 och 8 i del 3 av formuläret och är till för att undersöka om deltagaren väljer att justera något av sina svar från del 3 efter informationstillförseln. På det sättet vill undersöka om det existerar en skillnad i värdering efter en ökning i medvetande eller tvärtom.

#### 3.3.2.6 *Kvalitativt avsnitt (11)*

Syftet med den avslutande frågan är att få en kvalitativ aspekt på undersökningen genom att låta användaren ge sin egen åsikt om datainsamling på YouTube i fritext. Svaren ämnas användas för att nyansera vår diskussion om resultaten i studien.

### 3.4 Analys av data

För att tydligare kunna förstå och diskutera de svar som vi får in från vår kvantitativa undersökning har vi valt att formulera ett antal hypoteser. Dessa bygger på det resonemang som Spiekermann et al. (2016) för kring medvetenhet och datainsamling. Främst då att individer tenderar att bli mer intresserade av sin egen data om de vet vad de delat med sig av, samt hur den används (Spiekermann et al. 2016).

Hypoteserna är följande:

H1: Medvetenheten för användarbeteendedata är lägre än vad den är för persondata.

H2: Medvetenheten blir högre för användarbeteendedata och persondata efter vår informationstillförelse.

H3: Medvetenheten påverkar hur individer värderar persondata och användarbeteendedata.

Då frågan för att undersöka medvetenhet och den för värdering är utformade med olika underliggande skalor så kommer det ej vara möjligt att översätta resultaten rakt av mellan varandra men genom att bearbeta resultaten så får vi resultat som går att jämföra och se samband mellan. Vi har valt att titta på medelvärden för medvetenhet respektive värdering för deltagarna i vår studie. Målet är att vi för båda grupperna ska ha ett initialt medelvärde samt ett efter informationstillförelse. Det är mellan våra medelvärden som de fyra jämförelserna ovan kommer att genomföras.

Medelvärde för medvetenhet har räknats ut genom att räkna ut antalet persondata respektive användarbeteendedata var deltagare har svarat i respektive fråga (min=0, max=5), listan summeras och delas därefter på antalet deltagare för att ge oss ett medelvärde.

*räkneexempel:*

*deltagare x har svarat "namn, kön, ålder, civilstatus, fritidsintresse"*

*x = persondata: 3, användarbeteendedata: 2*

*deltagare y har svarat "namn, fritidsintresse"*

*y = persondata: 1, användarbeteendedata: 1*

*Medelvärde xy persondata =  $3+1 / 2 = 2$*

*Medelvärde xy användarbeteendedata =  $1+1 / 2 = 1$*

Medelvärde för värdering tas fram genom att ta genomsnittet av deltagarens fem värderingar (min=1, max=5) per kategori. Deltagarnas medelvärden summeras och delas på antalet totala deltagare för att ge oss vårt slutgiltiga medelvärde för värdering på respektive fråga.

*räkneexempel:*

*deltagare x har angett värderingarna "namn: 1, kön: 1, ålder: 2, land: 3, telefonnummer: 4, civilstatus: 5, språk:*

*2, föräldrastatus: 3, matpreferenser: 3, fritidsintresse: 4"*

*Medelvärde x persondata =  $1 + 1 + 2 + 3 + 4 / 5 = 2.2$*

*Medelvärde x användarbeteendedata =  $5 + 2 + 3 + 3 + 4 / 5 = 3.4$*

### 3.5 Validitet och reliabilitet

För att den kvantitativa undersökningen som vi kommer att genomföra ska ge ett bra och användbart resultat, har vi tagit i beaktning ett antal aspekter. De aspekter vi har valt att belysa är uppdelade i validitet som avser träffsäkerheten av vår studie och reliabilitet som avser tillförlitligheten av densamme. En hög validitet behöver inte betyda en hög reliabilitet och tvärtom, båda punkterna bör provas var för sig (Recker, 2013).

Enligt Bryman och Bell (2017) så bygger reliabilitet på tre stycken områden, stabilitet, intern reliabilitet och konsensus mellan intressenter. Med stabilitet menas att det går att återskapa studien och att resultatet inte kommer att variera på ett sätt som inte är menat. Intern reliabilitet syftar till att våra skalor samt svaren kopplade till dessa inte går att misstolka på ett sätt som gör att resultatet blir missvisande samt att det går att sätta i ett sammanhang. Konsensus mellan intressenter bygger på att information eller kategoriseringen av denna tolkas lika och det finns en idé eller konsensus om vad den betyder. För att uppnå stabilitet har vi gjort våra frågor och svar tillgängliga, för att just kunna ställa samma frågor om samma data i framtiden. Konsensus är inte självklart att uppnå med den begränsade mängd kunskap och tyngd vi som författare har, vi har därför i så stor mån det går tolkat begrepp efter andra artiklar och författare. Kopplat till den interna reliabiliteten anser vi att sammanhanget för de svarande är tydligt och framkommer i undersökningen, vi har därför inte genomfört test som Bryman och Bell föreslår, till exempel ett så kallat split-half test.

Det finns även flera metoder för att utvärdera validiteten i en studie och som presenteras av Bryman och Bell (2017) samt Recker (2013). Ett urval av de som nämns och provas i vår studie är face, content, construct och convergent validity. Den första, face validity avser huruvida undersökningen vid första intryck mäter det som avses och därmed att resultaten har betydelse i sammanhanget. Då vi har valt tydliga och väl underbyggda mätpunkter för medvetenhet och värdering avser vi ha hög face validity i vår studie. Content validity avser mer i detalj huruvida mätinstrumentet, i vårt fall frågeformuläret, på ett representabelt sätt kan mäta det som avses. För att undersöka träffsäkerheten av frågeformuläret så testades formuläret mot ett fåtal utvalda deltagare innan den släpptes i sin helhet. Testet anser vi ej vara tillräckligt för att uppfylla kriterium för en pilotstudie men tillräckliga för att försäkra oss om ett högt content validity. Construct validity syftar till att teorier och hypoteser härleds till relevanta koncept och idéer, samt att detta mäter relevanta aspekter. De artiklar vi bygger vårt resonemang och studie på anser vi vara av hög relevans och bygger den grund vi avser att undersöka. Avslutningsvis har vi uppskattat vår convergent validity som avser huruvida vi kan bekräfta att vårt mätinstrument fungerar väl. Ett exempel på detta i vår studie hade varit genom att bekräfta vilka datapunkter som samlas in på varje person. Något som hade kunnats anses vara integritetskränkande och för med sig fler problem. Istället bekräftas vårt mätinstrument gentemot läsaren själv och med vår informationstillförsel. Vi argumenterar därför att deltagarna rättar till sig själva och att vi därmed kan anse att vårt mätinstrument fungerar som tänkt.

### 3.6 Etiska överväganden

Ett stort fokus under skrivandet av den här uppsatsen har lagts på etiska överväganden och utvärdering av potentiella konsekvenser. Framförallt så har fokus legat på att utforma ett frågeformulär med bästa möjliga etiska behandling.

Oates (2006) lägger fram fem rättigheter för deltagare i en studie. Rätt att inte delta, rätt att avbryta, rätt att ge informerad tillåtelse, rätt till anonymitet och rätt till sekretess. Vissa av de här rättigheterna innefattas per automatik via en digital undersökning. Trots detta har vårt frågeformulär utformats med de här fem rättigheterna i åtanke vid varje steg. Till att börja med så är studien helt frivillig att delta i och har delats via öppna forum för att säkerställa rätten att ej delta. Innan insamling av data påbörjades i frågeformuläret blev deltagaren presenterad med information om studien, författarna bakom den och deras rättigheter som deltagare varav bland annat deras rätt att när som helst avbryta formuläret. De blev även ombedda att bekräfta inskickandet av svar. I och med detta anser vi ha säkrat dels rätten till att avbryta men även rätten till informerad tillåtelse. Slutligen så har vi endast samlat in den information som vi avsett nödvändig för att tilldela deltagaren en demografisk indelning i vår studie tillsammans med deras svar om medvetenhet och värdering. I den här studien är den informationen en ålderskategori och användarvanor på YouTube. De kvarvarande uppgifterna anser vi är omöjligt att använda uteslutande för att identifiera en fysisk person. Utöver det, så har vi även valt att ej dela med oss av de anonymiserade resultaten i sin helhet för att bibehålla yttersta sekretess för deltagaren.

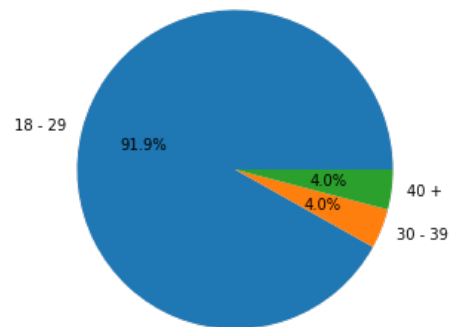
Vidare så bryter Bryman och Bell (2017) ned etiska principer i forskning till fyra områden, risk för skada på deltagaren, avsaknad av informerad tillåtelse, integritetskränkning och huru vidare deltagaren har vilseletts innan eller under studien. Avsaknad för informerad tillåtelse samt risk för integritetskränkning har behandlats i föregående stycke. Risk för skada på deltagaren innefattar enligt författarna fysisk skada, skada på självförtroende, risk för stress m.fl. faktorer. Då vår undersökning innefattar ett moment av informationstillförsel där vi inte kan kontrollera till fullo vilka datapunkter som Google har kopplats till deltagaren, och inte heller vilka som deltar i studien så kan vi inte utesluta att deltagaren far illa utav informationen. Däremot så är vår definitiva uppfattning att informationen som tillförs ej är av skadlig karaktär i allmänhet och bedömer risken som minimal. Vidare så har vi i utformningen av frågeformuläret designat ett inledande kapitel med största fokus för att undvika att vilseleda deltagare. I det inledande kapitlet förklarar vi i stora drag vilken information som kommer samlas in samt de momenten som innefattas i formuläret. Med detta i åtanke bedömer vi även risken för att vilseleda deltagare som minimal.

## 4 Resultat

I det här kapitlet redovisar vi resultaten från vår studie. I de fall att det har behövts så har resultaten avrundats till närmsta hundradel.

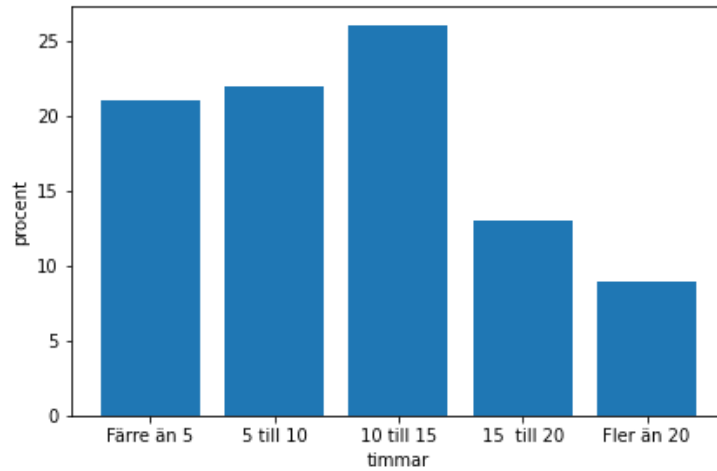
### 4.1 Beskrivande statistik

Totalt hade vi 108 deltagare i studien. Utav de 108 uppgav 8,33% (9 st) att de ej hade ett YouTube konto och studien avslutades efter fråga 2 för dem. Åldersfördelningen är kraftigt viktad mot vår målgrupp av 18–29 år och uppnår 92% av totalen med åldersgrupperna “30–39” och “40+” på 4% vardera. Det lämnar oss med 91 deltagare i vår målgrupp som vi kan använda oss vid analys av medvetenhet och värdering. Åldersfördelningen av deltagarna illustrerad i figur 1.



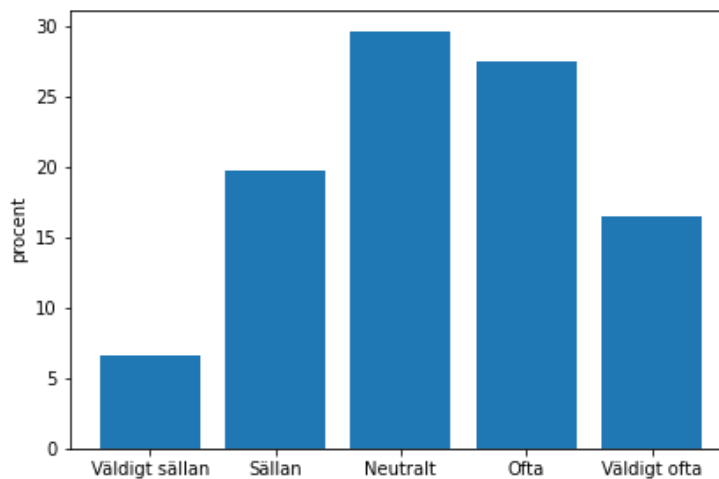
**Figur 1:** Åldersfördelning för deltagarna i studien

Tiden som våra deltagare spenderar på YouTube i genomsnitt varje vecka är fördelad med typvärdet på “10 till 15” (26%) följt av “5 till 10” (22%) och “Färre än 5” (21%) dessutom “15 till 20” (13%) och “Fler än 20” (9%).



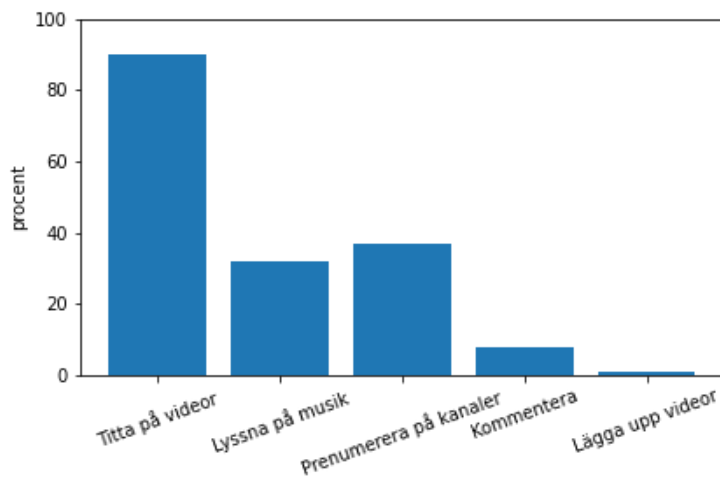
**Figur 2:** Andel deltagare per genomsnittlig tid spenderad på YouTube per vecka

Flest (27%) svarar “Neutralt” gällande om de vet vad de vill titta på innan de går in på YouTube medan 25% ger svarar “Ofta”. “Sällan” svarar 18% följt av “Väldigt ofta” (15%) och “Väldigt sällan” (6%).



**Figur 3:** Fördelning över oplanerat till planerat tittande på YouTube

Det vanligaste användningsområdet för YouTube bland våra deltagare är “Titta på videor” med 99% svarsfrekvens följt av “Prenumerera på kanaler” (40%), “Lyssna på musik” (36%), “Kommentera” (8%) och “Lägga upp videor” (1%).

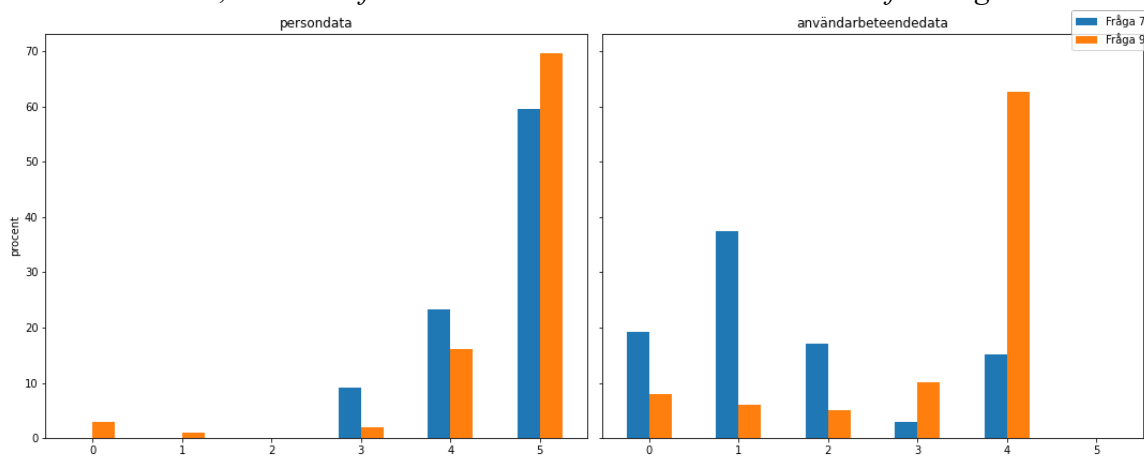


Figur 4: Andel av deltagarna som svarat att de använder YouTube på respektive sätt

Slutligen är det endast 6 deltagare (6,7%) som anger att de betalar för YouTube Premium.

## 4.2 Medvetenhet

I den här delen presenteras resultaten från fråga 7, före informationstillförel samt 9, efter informationstillförel, “*Vilken information tror du att YouTube tar del av från dig?*”.



Figur 5: Hur många procent av deltagarna som har angett x antal punkter per kategori innan och efter informationstillförel

I figur 5 ser vi fördelningen av hur många informationspunkter deltagarna uppger att de är medvetna om samlas in. I fråga 7 uppgav deltagarna i genomsnitt en medvetenhet om 4,55 med en standardavvikelse på 0,67. I fråga 9 var medvetenheten för persondata 4,57 med en standardavvikelse på 1,04. Genomsnittet ökar med 0,4% eller 0,02 enheter samtidigt som standardavvikelsen höjs med 55% eller 0,37 enheter. När det kommer till användarbeteendedata så var genomsnittet 1,54 med en standardavvikelse på 1,31 i fråga 7 och 3,23 med en standardavvikelse på 1,32 i fråga 9. Medvetenheten för användarbeteendedata ökade med 109% och standardavvikelsen ökade 0,7%.

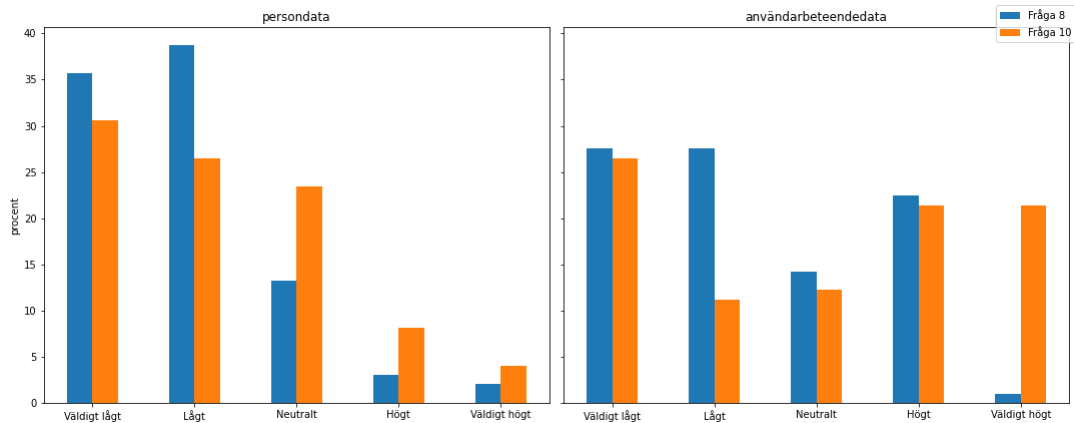


**Tabell 2:** Medelvärde och standardavvikelse över medvetenhet

	Fråga 7	Fråga 9
persondata	4,55 (0,67)	4,57 (1,04)
användarbeteendedata	1,54 (1,31)	3,23 (1,32)

### 4.3 Värdering

I det här delkapitlet presenteras resultaten gällande värdering, främst kommer fråga 8 och 10 att avhandlas.

**Figur 6:** Medianen av deltagarnas fem värderingar per kategori fördelat över kategorierna

Innan informationstillförelse så är genomsnittet för persondata 1,99 med en standardavvikelse på 0,99 jämfört med 2,45 och en standardavvikelse på 1,91 efteråt. Både värderingen och standardavvikelsen ökade med 17%. Tittar vi på användarbeteendedata är genomsnittet innan informationstillförelse 2,45 och standardavvikelsen 1,91 respektive 3,04 och standardavvikelse 1,51 efter. Värderingen ökade med 59% och standardavvikelsen med 15%.

**Tabell 3:** Medelvärde och standardavvikelse över värdering

	Fråga 8	Fråga 10
persondata	1,99 (0,99)	2,45 (1,91)
användarbeteendedata	1,91 (1,31)	3,04 (1,51)

## 4.4 Sammanfattning av medelvärden

En sammanfattning av resultaten i medvetenhet och värdering presenteras i tabell 4. Värdena för medvetenhet är genomsnittet av hur många informationspunkter varje deltagare har valt (min=0, max=5). Värdena för värdering är ett genomsnitt av hur varje deltagare har värderat vardera informationspunkt (min=1, max=5). Därför kan själva värdena inte jämföras mellan kategorierna men det går att identifiera samband i hur de har ändrats.

**Tabell 4:** Sammanfattning över medelvärden

	Fråga 7 & 9		Fråga 8 & 10		Differens	
	medvetenhet	värdering	medvetenhet	värdering	medvetenhet	värdering
persondata	4,55	1,99	4,57	2,45	0,02	0,46
användarbeteendedata	1,54	1,91	3,23	3,04	1,69	1,13
medelvärde	3,045	1,95	3,9	2,745	0,855	0,795

Det första resultatet som vi har identifierat är en stor skillnad i initial medvetenhet för användarbeteendedata på 30% (M=1,54) i kontrast till persondata 91% (M=4,55) en skillnad i medvetenhet på 61%.

Den initiala skillnaden i värdering skiljer sig åt med 4%. Med persondata (M=1,99) något högre än användarbeteendedata (M=1,91). Medelvärdet mellan de två grupperna hamnar på 1.95 alltså något under en låg värdering.

Medvetenhet för persondata efter informationstillförelse gick upp 0,4% (0.02). Användarbeteendedata skiljer sig med 109% (1,69) och medvetenhet ökade jämfört med innan informationstillförelse.

Värdering efter informationstillförelse ökade med 23% (0,46) för persondata och med 59% (1,13) för användarbeteendedata. Genomsnittet för de båda ökade med 41% (0,795) till 2,745 vilket innebär att totala värderingen efter informationstillförelse är 0,245 enheter närmare "neutral" i kontrast till 0,05 under "låg" innan informationstillförelse.

Slutligen så kan vi även se att värderingen har gått upp med 0,795 i samband med att medvetenheten har gått upp 0,825.

## 5 Diskussion

### 5.1 Analys och hypotesbesvarande

#### 5.1.1 *Medvetenheten för användarbeteendedata är lägre än vad den är för persondata.*

Resultatet från vår undersökning visar i att medvetenheten initialt är lägre för användarbeteendedata än vad den är för persondata. I tabell fyra kan vi se att den ligger på 4,55 av 5 för persondata och 1,54 av 5 för användarbeteendedata, en klar skillnad. Detta följer dels vad Spiekermann et al. (2016) presenterar i sin artikel och även vad Kokolakis (2017) tar upp i sin artikel. H1 att medvetenheten för användarbeteendedata är lägre än vad den är för persondata kan alltså bekräftas.

#### 5.1.2 *Medvetenheten blir högre för användarbeteendedata och persondata efter vår informationstillförsel.*

Undersökningen som vi låtit genomföra visar i att medvetenheten för persondata bara ökar med 0,02 efter vår informationstillförsel, vilket vi anser vara inom felmarginalen då standardavvikelsen även ökade. Därmed anser vi att det ej sker en ökning. Däremot ökar medvetenheten för användarbeteendedata efter vår informationstillförsel från 1,54 till 3,23 av 5, en ökning med 1,69 enheter. Därför kan H2 medvetenheten blir högre för användarbeteendedata och persondata efter vår informationstillförsel endast delvis bekräftas.

#### 5.1.3 *Medvetenheten påverkar hur individer värderar persondata och användarbeteendedata.*

Resultatet visar att efter informationstillförsel så ökar medvetenheten i genomsnitt med 0,855 enheter samtidigt som värderingen ökar med 0,795 enheter. Inom ramarna för vår undersökning menar vi att det finns en koppling mellan en ökad medvetenhet och ökad värdering och H3 kan därför bekräftas.

### 5.2 Medvetenheten hos konsumenten

Vår undersökning visade att medvetenheten, det vill säga hur bra koll konsumenten har på vilken data som delas, är lägre för användarbeteendedata än vad den är för persondata. Anledningarna till detta kan vara flera och det är därför initialt intressant att föra en diskussion kring detta. För att börja med själva idén med sociala medier och YouTube. YouTube ägs som tidigare nämnt av Google vars själva affärsidé är data. Denna centrala aspekt är en grundsten till varför data kan hamna i skymundan, och därmed minska medvetenheten kopplat till den. Konsumenten och företaget som erbjuder tjänsten sociala medier har i någon mening en skillnad i vad tjänsten syftar till att göra, kopplat till vad de tjänar på att använda respektive tillhandahålla den (Larsson & Ledendal, 2017).

Skillnaden i perspektiv och kunskap mellan YouTube och konsumenten kan förklaras av att det inte är direkt uppenbart att konsumenten betalar något för att använda tjänsten. Betalningen görs inte i direkta pengar utan i konsumentens data, vilket den godkänner via att ge sitt medgivande innan den använder tjänsten. Vår undersökning visade att endast 6,1 procent av de svarande betalar för tjänsten YouTube premium, en tilläggstjänst som gör att konsumenten inte behöver titta på någon reklam. Vilket går att tolka som att konsumenterna av YouTube förmodligen inte fäster någon särskild vikt vid att behöva titta på reklam när de använder tjänsten. Detta gör dock inte att YouTube slutar att samla in någon data, vilket bekräftar hur viktigt data är för tjänsten (YouTube, 2021b).

Sedan går det att se att rekommendationer baserade på data kan korrelera med, som YouTube själva beskriver det, att visa relevanta annonser. Därför kan det gynna både konsumenten och YouTube. Konsumenten som får relevanta annonser får information som den möjligtvis är intresserad av utan att den aktivt behöver söka upp den. Samtidigt får YouTube en fördel när de ska sälja dessa annonser till något företag. Det finns även en fördel med att YouTube ingår i Googles nätverk av data, då datan kan samlas in på fler plattformar och därmed ge en mer komplett bild. Denna kompletta bild går direkt att koppla till begreppet digital identitet (Larsson & Ledendal, 2017). Vår undersökning kan tolkas såsom att individen kan veta att den har en digital identitet, dock att individen möjligtvis inte kan identifiera omfattningen av den.

Att den generella medvetenheten kring vad för data som delas av konsumenterna är låg kan också kopplas till att individen helt enkelt ser större fördelar med att dela sin data än vad den ser nackdelar. Eftersom det är nära omöjligt eller i alla fall kräver flera aktiva åtgärder för att kunna använda sociala medier och YouTube utan att dela med sig av någon data så ser konsumenten inte fördelen med att genomföra dessa åtgärder. Som togs upp i litteraturgenomgången tenderar konsumenter att optimera fördelarna samtidigt som de inte ser förbi riskerna.

### 5.3 Skillnaden mellan persondata och användarbeteendedata

En intressant skillnad som går att urskilja i vår undersökning är att medvetenheten kring vad deltagarna tror att YouTube tar del av för data före och efter kunskapstillförsel är stor. Före kunskapstillförsel är medvetenheten kring persondata hög, och förblir så även efter. Medvetenheten kring användarbeteendedata är låg innan kunskapstillförselen men ökar efter. Det finns alltså två mönster värda att diskutera. För att börja med persondata, varför konsumenter i vår undersökning i högre grad vet att data av typen persondata samlas in och användas av YouTube kan ha flera anledningar. En av dessa kan vara den relativt nya men mycket uppmärksammade dataskyddsförordningen, även kallad GDPR.

Förordningen reglerar som tidigare nämnt endast persondata, varför det kan falla sig naturligt att uppmärksamheten kring data också fokuserats just till persondata. Förordningen kräver också att YouTube och andra företag aktivt frågar om tillåtelse för att samla in data, i vilket syfte och slutligen presenterar vilken eller vilka datapunkter som samlas in, vilket ytterligare påminner konsumenten om att just den här typen av data samlas in. Att detta påverkar medvetenheten kring persondata är inte säkert, men det kan vara bidragande. Persondata har också karaktären av att vara definitiv på ett sätt som inte användarbeteendedata är. Namn, adress, telefonnummer är data som individer tenderar att veta själva och inte är lika föränderlig, den är dessutom ofta nödvändig att dela med sig för att få tillgång till en tjänst som YouTube (Larsson & Ledendal, 2017). Sammantaget går det att se att ens digitala profil gällande persondata borde vara mycket lik den som personen har i verkligheten.

Den information som ingår i ens digitala profil men som består av användarbeteendedata kan dock vara svårare att överblicka än den som består av persondata. Google beskriver själva att de gissar sig till en konsuments intressen och preferenser, det går då inte heller att utesluta att de ibland helt enkelt gissar fel. Men även om de gissar rätt, behöver det nödvändigtvis inte vara gissningar som konsumenten själv är medveten om. Användarbeteendedata är också precis som namnet avslöjar, beteenden, vilket i sin natur inte nödvändigtvis är något som individer har koll eller kontroll över. Det är därför inte heller särskilt konstigt att individer helt enkelt är mindre medvetna om sin användarbeteendedata eftersom de inte vet vad den består av.

Vi har valt att titta på civilstatus, alltså om man är singel, sambo eller gift m.fl. Språk, föräldrastatus, matpreferenser och fritidsintressen. Kategorierna valdes ut eftersom det är så som Google och YouTube benämner dem i deras annonsinställningar. Språk sticker ut som den enda av datapunkterna i användarbeteendedata där en förändring i medvetenhet är ganska liten. Det kan till exempel bero på är att individer är medvetna om vad de använder för språk och därmed finns det inte några luckor i medvetenhet att fylla.

Civilstatus är något som vi kan anta att de allra flesta individer vet om i sitt verkliga liv, däremot skedde till skillnad från språk en stor ökning i medvetenheten i förhållande till vad individer trodde att deras digitala identitet såg ut, och vad den faktiskt innehöll. Varför medvetenheten av civilstatus är låg initialt kan kopplas till att den inte är något individen själv ger till YouTube. Detta gissas fram precis som alla andra datapunkter som är användarbeteendedata. Föräldrastatus, alltså om individen har några barn, ser ut på samma sätt som civilstatus, en låg medvetenhet före informationstillförelsen för att sedan öka efter den.

Att individer delar med sig av användarbeteendedata till YouTube men att medvetenheten till vad som delas är låg kan ha andra anledningar än att individen som tidigare sagt inte har full koll på vad ens beteenden är. Studien som Lee et al. (2013) genomför kring risker och fördelar med att dela information på sociala medier kan ge en annan vinkel på problemet. Författarna tar upp att individer tenderar att leta fördelarna samtidigt som de försöker att undvika riskera relaterat till sociala medier. Detta kan förklara varför en del i att individer helt enkelt ser fördelarna med YouTube som större än att dela med sig av sin data och därmed inte bryr sig om konsekvenserna av detta. Individerna kan helt enkelt anse att riskerna med att dela med sig av data till YouTube är så låg att de inte är värt att bry sig om. Konsekvensen av detta kan vara som vår undersökning visade, en låg medvetenhet kring användarbeteendedata.

## 5.4 Medvetenhetens inverkan på värderingen

I vår undersökning går det att se en förändring i både medvetenhet och värdering från före informationstillförelsen till efter denna. En låg medvetenhet verkar ge en lägre värdering och en hög medvetenhet verkar ge en högre värdering. För att förstå varför det är så, kan vi börja med att anta att det är svårt för en individ att värdera något som den inte har någon kunskap eller information om. Stefanie Pöttsch beskriver medvetenhet som att bli påmind eller uppmärksammas om information (Pöttsch 2008). I vår undersökning kan vi visa på att just en påminnelse eller ett uppmärksammande ökar medvetenheten, men också värderingen i viss mån. Pöttsch tar också upp att det största problemet är att på användaren att förstå vad det är den tittar på för information. Genom att vi verkligen visar vad för information YouTube och Google har om individen så borde också förståelsen öka.

Till skillnad från om vi hade endast sagt att YouTube eller Google hade sparat och använt sig av persondata och användarbeteendedata. Intressant att lyfta är också den undersökning som Steijn låtit göra om medvetenhet. Visserligen på Facebook men slutsatserna som han drar om att de svarande inte gjorde några åtgärder kring sin data trots att medvetenheten ökade samt att han menar att de behövs en incident eller att användaren faktiskt utsätts för en risk för att några åtgärder skall vara aktuella (Steijn 2014). I någon mån kan detta tolkas som att de svarande inte värderar sin data tillräckligt högt. Kanske spelar inte bara medvetenhet roll för hur någon värderar sin data utan också vilken erfarenhet och hur utsatt någon känner sig vara för eventuella risker som finns. Poängen med ökad medvetenhet är dock att försöka minimera dessa risker varför det fortfarande kan finnas incitament till att öka värderingen av speciellt användarbeteendedata genom medvetenhet.

## 5.5 Begränsningar och framtida studier

Våra resultat är begränsade i den mån att vi inte har en insikt i varför medvetenhet och värdering skiljer sig åt mellan persondata och användarbeteendedata och inte vilka faktorer som har störst inverkan. Vi kan heller inte säkerställa att vår studie representerar hela populationen då vi har ett litet urval (N=91). Vi rekommenderar att framtida forskning siktar på ett större och bredare urval och att data om faktorer som kan påverka medvetenhet och värdering samlas in från deltagarna.

Genom en kovariatanalys skulle man kunna skilja på faktorer som har en påverkan på medvetenhet respektive värdering och fokusera ytterligare på de faktorerna. För att öka validiteten i studien tvärkulturellt bör studien även sträcka sig över flera länder. Slutligen så är ett viktigt inslag för att föra forskning på ämnet vidare att undersöka det monetära värdet av användarbeteendedata, dels för konsumenten men även för näringsidkare och därmed hitta det sanna värdet för data.

En stor del med vår studie har varit att tillföra fler perspektiv på datainsamling. Insamling av data fortsätter öka exponentiellt och vår studie lägger fram viktiga och högt aktuella insikter i unga vuxna konsumenters kunskap om datainsamling. Vi tror att den här studien kommer att ligga som grund till vidare forskning och öka fokus på konsumenters dataintegritet och säkerhet.

## 6 Slutsats

Vi kan dra tre slutsatser utifrån våra två första forskningsfrågor: “*Hur ser medvetenheten ut och värderingen av data som samlas in bland konsumenterna av YouTube?*” och “*Hur skiljer sig medvetenheten och värderingen mellan persondata och användarbeteendedata?*”. Vi kan se att medvetenhet är initialt hög för persondata. Användarbeteendedata skiljer sig markant och medvetenheten är låg. Efter informationstillförseln i vår undersökning så ökar medvetenhet kring användarbeteendedata samtidigt som persondata ligger på en liknande nivå som innan informationstillförsel. Den första slutsatsen vi kan göra utifrån vår undersökning är att det existerar ett gap i medvetenhet om insamling av persondata i relation till användarbeteendedata. För det andra så går det att med små medel tillföra en markant ökning i medvetenhet för insamling av användarbeteendedata. Slutligen så är den initiala värderingen för användarbeteendedata högre än persondata medan medvetenheten är lägre och pekar återigen på ett kunskapsgap.

Gällande vår tredje och sista fråga: “*Hur påverkar medvetenhet om insamlade data dess värdering?*” kan vi se tendenser till att medvetenhetens inverkan på värderingen kan vara sammanlänkad. I vår studie ökade värderingen i samband med att medvetenheten ökade. Det är dock inte uteslutet att andra faktorer påverkar konsumentens värdering av både persondata och användarbeteendedata, varför fler studier i ämnet hade varit både nödvändigt och intressant.

I helhet kan vi konstatera att utifrån vår studie och forskning i ämnet så existerar det en lägre medvetenhet för insamling om användarbeteendedata än för persondata. Vi ser på det här som problematiskt, speciellt då våra resultat även pekar på att en högre medvetenhet är sammankopplat till en högre värdering. Vår mening är att det pekar på en tydlig undervärdering av användarbeteendedata hos konsumenterna idag och att frågan inte längre är uteslutande moralisk utan även monetär.

# Appendix I: Frågeformulär

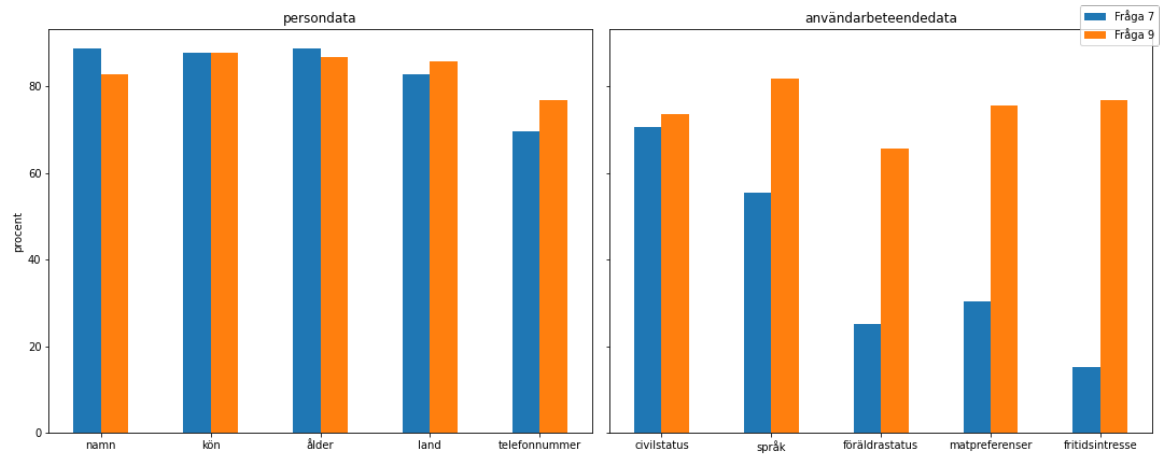
Frågeformuläret som användes till vår studie med samtliga frågor och svarsalternativ.

1. Hur gammal är du?
  - - 17
  - 18-29
  - 30-39
  - 40 +
2. Har du ett Youtubekonto?
  - Ja
  - Nej
3. Hur många timmar per vecka spenderar du på Youtube i genomsnitt?
  - Under 5
  - 5 till 10
  - 10 till 15
  - 15 till 20
  - Över 20)
4. Hur ofta vet du vad du vill titta på innan du går in på Youtube?
  - Väldigt sällan
  - Sällan
  - Ibland
  - Ofta
  - Väldigt ofta)
5. Hur använder du Youtube? (Checklista)
  - Titta på videor
  - Lyssna på musik
  - Kommentera
  - Prenumerera på kanaler
  - Lägga upp videor
6. Betalar du för Youtube Premium?
  - Ja
  - Nej
7. Vilken information tror du att Youtube tar del av från dig? (Checklista)
  - Namn
  - Kön
  - Ålder
  - Land
  - Telefonnummer
  - Civilstatus
  - Språk
  - Föräldrastatus
  - Matpreferenser
  - Fritidsintressen
8. Hur värderar du informationen "x"? (x = en av de tio alternativen i föregående fråga)
  - Väldigt lågt
  - Lågt
  - Neutralt
  - Högt
  - Väldigt högt

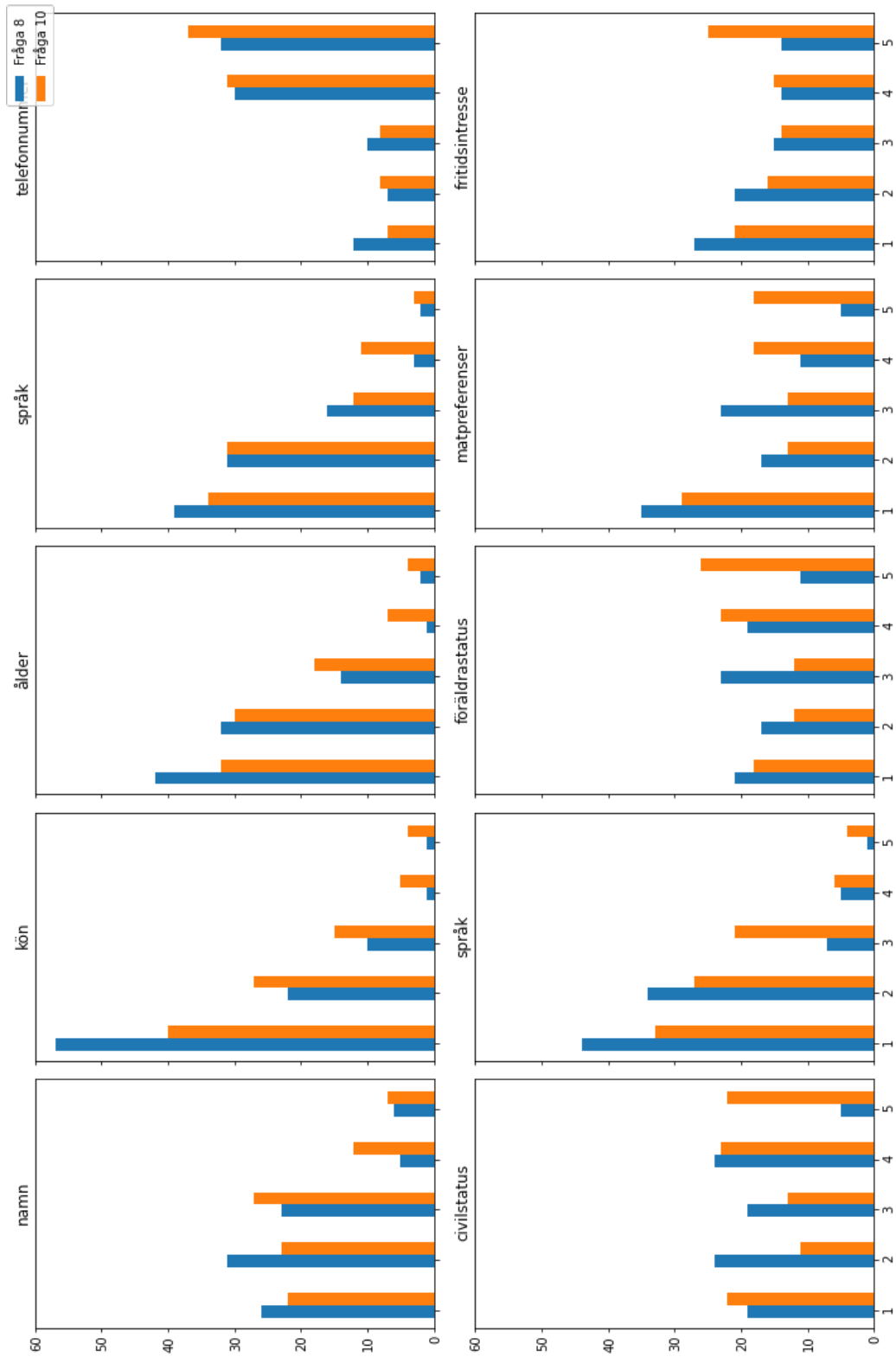


9. Vilken information tror du att Youtube tar del av från dig? (Checklista)
- Namn
  - Kön
  - Ålder
  - Land
  - Telefonnummer
  - Civilstatus
  - Språk
  - Föräldrastatus
  - Matpreferenser
  - Fritidsintressen
10. Hur värderar du informationen “x”? (x = en av de tio alternativen i föregående fråga)
- Väldigt lågt
  - Lågt
  - Neutralt
  - Högt
  - Väldigt högt
11. Kompletterande åsikt om datainsamling på Youtube.
- Fritext

## Appendix II: Medvetenhet alla punkter



# Appendix III: Samtliga värderingar



## Referenser

- Acharya, A. Singh, K. S. Pereira, V. & Singh, P. (2018). Big Data, Knowledge Co-creation and Decision Making in Fashion Industry, *International Journal of Information Management*, vol. 42, no. 1, pp. 90-101
- Beresford A.R. Kübler, D. Preibusch, S. (2012). Unwillingness to pay for privacy: A field experiment, *Economics Letters*, vol. 117, no. 1, pp. 25-27
- Bryman, A., & Bell, E. (2017). *Företagsekonomiska Forskningsmetoder*, 3e uppl, Stockholm: Liber
- Dawes, J. (2008). Do Data Characteristics Change According to the Number of Scale Points Used? An experiment using 5-Point, 7-Point and 10-Point scales, *International Journal of Market Research*, vol. 50, no. 1, pp. 61-104
- Rao, V. and Dwivedi, K. 2021. To share or not to share. *Deloitte Insights*, Tillgänglig online: <<https://www2.deloitte.com/us/en/insights/industry/retail-distribution/sharing-personal-information-consumer-privacy-concerns.html>> [Hämtad 8 maj 2021].
- Forrester. (2014). Marketing's Big Leap Forward: Overcome the urgent challenge to improve customer experience and marketing performance, Tillgänglig online: <https://app.compendium.com/uploads/user/4f91a3ee-6ace-42a7-be93-3b21f3a1635f/7c7092b4-1fc8-48d9-8bd4-bcf49cfc6c97/File/b09103706cbc150358adc5fa67a3dd07/1394484576218.pdf> [Hämtad 12 April 2021]
- Europaparlamentets och rådets förordning (EU) 2016/679 av den 27 april 2016 om skydd för fysiska personer med avseende på behandling av personuppgifter och om det fria flödet av sådana uppgifter och om upphävande av direktiv 95/46/EG (allmän dataskyddsförordning)
- Ghani, A. N. Hamid, S. Hashem, I. A. T. & Ahmed, E. (2019). Social Media Big Analytics: A survey, *Computers in Human Behavior*, vol. 101, pp. 417-428
- Google AdSense. (2021). Ad targeting, Tillgänglig online: <https://support.google.com/adsense/answer/160525?hl=en> [Hämtad 1 April 2021]
- Godel, M. Litchfield, A. Mantovani, I. (2012). The Value of Personal Information: Evidence from Empirical Economic Studies, *Communications & Strategies*, vol. 88, pp. 41-60
- Ghosh, D. (2018). How GDPR Will Transform Digital Marketing, Harvard Scholar, Tillgänglig online: [https://scholar.harvard.edu/files/dipayan/files/how\\_gdpr\\_will\\_transform\\_digital\\_marketing.pdf](https://scholar.harvard.edu/files/dipayan/files/how_gdpr_will_transform_digital_marketing.pdf) [Hämtad 17 April]
- Gustafsson, B. Hermeren, G. Petersson, B. (2006). *Good Research Practice*, Swedish Research Council.
- King, N. J. & Forder, J. (2016). Data Analytics and Consumer Profiling: Finding appropriate privacy principles for discovered data, *Computer Law & Security Review*, vol. 32, no. 5, 696-714
- Kokolakis, S. (2017). Privacy Attitudes and Privacy Behaviour: A review of current research on the privacy paradox phenomenon, *Computers & Security*, vol. 64, pp. 122-134
- Larsson, S. & Ledendal, J. (2017). *Personuppgifter som betalningsmedel*. (4 uppl.) Konsumentverket.  
Tillgänglig online: [https://lup.lub.lu.se/search/ws/files/25410448/Larsson\\_Ledendal\\_2017\\_Personuppgifter\\_som\\_betalningsmedel\\_konsumentverket.pdf](https://lup.lub.lu.se/search/ws/files/25410448/Larsson_Ledendal_2017_Personuppgifter_som_betalningsmedel_konsumentverket.pdf) [Hämtad 20 April 2021]
- Lee, H. Park, H. & Kim, J. (2013). Why Do People Share Their Context Information on Social Network Services? A qualitative study and an experimental study on users'

- behavior of balancing perceived benefit and risk, *International Journal of Human-Computer Studies*, vol. 71, no. 9, pp. 862-877
- Loukides, M. (2018). What Is Data Science?: A Beginner's Guide to Key Concepts and Pitfalls
- McAfee, A. & Brynjolfsson, E. (2012). Big Data: The management revolution, *Harvard Business Review*, Tillgänglig online: <https://hbr.org/2012/10/big-data-the-management-revolution> [Hämtad 20 April 2021]
- Mohan Rao, P. Murali Krishna, S. Siva Kumar, A.P. (2018). Privacy preservation techniques in big data analytics: a survey, *Journal of Big Data*, vol. 5, no. 33
- Morey, T. Forbath, T. & Schoop, A. (2015). Customer Data: Designing for transparency and trust, *Harvard Business Review*, Tillgänglig online: <https://hbr.org/2015/05/customer-data-designing-for-transparency-and-trust> [Hämtad 20 April 2021]
- Norberg, P. A. Horne, D. R. & Horne, D. R. (2007). The Privacy Paradox: Personal information disclosure intentions versus behaviors, *The Journal of Consumer Affairs*, vol. 41, no. 1, pp. 100-126
- Oates, B.J. (2006). *Researching Information Systems and Computing*, SAGE Publications.
- Pötzsch S. (2009) Privacy Awareness: A Means to Solve the Privacy Paradox?, *The Future of Identity in the Information Society. Privacy and Identity 2008. IFIP Advances in Information and Communication Technology*, vol 298.
- Recker, J. (2013). *Scientific Research in Information Systems A Beginner's Guide*, Springer Berlin Heidelberg
- Rose, J. Barton, C. Souza, R. & Platt, J. (2013). The Trust Advantage: How to win with big data, The Boston Consulting Group, Tillgänglig online: [https://image-src.bcg.com/Images/The\\_Trust\\_Advantage\\_Nov\\_2013\\_tcm9-92206.pdf](https://image-src.bcg.com/Images/The_Trust_Advantage_Nov_2013_tcm9-92206.pdf) [Hämtad 12 April 2021]
- Rose, J. Rehse, O. Röber, B. (2012). The Value of Our Digital Identity, The Boston Consulting Group, Tillgänglig online: <https://www.bcg.com/publications/2012/digital-economy-consumer-insight-value-of-our-digital-identity> [Hämtad 13 April 2021]
- Rosenthal, D. S. H. Rosenthal, D. C. Miller, E. L. Adams, I. F. Storer, M. W. & Zadok, E. (2010). The Economics of Long-Term Digital Storage, Stanford University Libraries, Tillgänglig online: [http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/mow/VC\\_Rosenthal\\_et\\_al\\_27\\_B\\_1330.pdf](http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/mow/VC_Rosenthal_et_al_27_B_1330.pdf) [Hämtad 1 April 2021]
- Schermer, B. W. (2011). The Limits of Privacy in Automated Profiling and Data Mining, *Computer Law & Security Review*, vol. 27, no. 1, pp. 45-52
- Smith, M. Szongott, C. Henne, B. & von Voigt, G. (2012). Big Data Privacy Issues in Public Social Media, *IEEE Explore*, Tillgänglig online: <https://ieeexplore.ieee.org/abstract/document/6227909/authors#authors>[Hämtad 10 April 2021]
- Spiekermann, S. & Korunovska, J. (2016). Towards a Value Theory for Personal Data, *Journal of Information Technology*, vol. 32, no. 1, pp. 62-84
- Statista. (2018). Global Big Data Market Size Revenue Forecast Worldwide from 2011 to 2027, Tillgänglig online: <https://www.statista.com/statistics/254266/global-big-data-market-forecast/> [Hämtad 12 April]
- Steijn, W. (2014). A developmental perspective regarding the behaviour of adolescents, young adults, and adults on social network sites. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, vol. 8, no. 2
- Svenskarna och Internet. (2019). Poddar och Digitala Böcker Ökar, Tillgänglig online: <https://svenskarnaochinternet.se/rapporter/svenskarna-och-internet-2019/mediekonsumtion-titta-lyssna-och-lasa-digitalt> [Hämtad 1 April 2021]

- Turner, A. (2015). "Generation Z: Technology And Social Interest", *Journal of Individual Psychology*, vol. 71, no. 2, pp. 103–113.
- Uebersax JS. (2006). Likert scales: dispelling the confusion, *Statistical Methods for Rater Agreement*, Tillgänglig online: <http://john-uebersax.com/stat/likert.htm> [Hämtad: 12 April 2021]
- Walsh, J. W. (2017). How a Small Company Used Big Data to Increase Its Sales, IMD, Tillgänglig online: <https://www.imd.org/research-knowledge/articles/how-a-small-company-used-big-data-to-increase-its-sales/> [Hämtad 10 April 2021]
- Yang, Y. (2010). Web user behavioral profiling for user identification, *Decision Support Systems*, vol. 49, no. 3, pp. 261-271.
- YouTube. (2021a). Hur tjänar YouTube pengar?, Tillgänglig online: <https://www.YouTube.com/howYouTubeworks/our-commitments/sharing-revenue/#revenue-alternatives> [Hämtad 1 April 2021]
- YouTube. (2021b). Annonstinställningar, Tillgänglig online: [https://www.YouTube.com/intl/ALL\\_se/howYouTubeworks/user-settings/ad-settings/](https://www.YouTube.com/intl/ALL_se/howYouTubeworks/user-settings/ad-settings/) [Hämtad 1 April 2021]
- Zafeiropoulou, A.M. Millard, D.E. Webber, C. O'Hara, K. (2013). Unpicking the privacy paradox: can structuration theory help to explain location-based privacy decisions? *In Proceedings of the 5th Annual ACM Web Science Conference (WebSci '13)*, pp. 463–472.