



LUNDS UNIVERSITET
Ekonomihögskolan

De kom, de klickade, de övergav

Varför överges kundkorgen på nätet? En analys av webbstatistik från en detaljhandels e-handel.

Kandidatuppsats i statistik (15 hp) VT 21

Författare Kim Thurow

Handledare Björn Holmquist

Abstract

This thesis analyses web statistics belonging to a national B2C e-commerce in retail. The purpose of the thesis is to investigate abandoned carts amongst logged in customers with at least one product added in to their cart and at least one visit to the payment page. Since the response variable "cart" is binary and we are studying the probability of a cart being abandoned, the suitable method of choice is the logistic regression model.

The findings conclude that the majority of the customers are women between 30-60. A linear expected probability was found when investigating a correlation between age and abandoned cart. Those under 30 years of age had a 70% probability of abandoning the cart whereas those over 70 years of age had a 30% probability of leaving the cart. Only a small portion of the customers were new to the company. Which proved to be of more significance to the outcome of the cart, than if the customer had been a frequently recurring customer in the past. That is - without including the customers with extreme values in Orderhistory ("Orderhistorik") where a exceedingly high conversion rate could be seen.

After considering age, the only channel proven to have a significant influence on the customer cart, was the "SMS" channel. Both the amount of time spent on the site and the amount of pages scrolled through had a significant positive effect on the probability of a cart being abandoned. The day that the customer entered the web page also had a significant impact. "Day 5" had the highest probability of abandoned carts. However, it seemed like the variation correlated with different behaviors in the age groups.

The most frequently added products were from the categories "Ladies" (Dam) and "Home interior" (Heminredning). However, products from the category "Dam" were expected to be less abandoned than "Heminredning". The number of products also had a significant impact on the cart, but since most customers added between 1 and 3 items to their carts, the relevance of the finding is perhaps low.

When investigating the "Checkout" page, where customers process the purchase, a few findings were made. The number of times a customer visited the "Checkout" page was significant to the outcome of the cart. When leaving the page, it seemed like the customer most frequently went back to look for more products - since the product pages for "Dam" and "Rea" correlated with "Checkout".

When considering the size of the cart, the probability increased. However, there were no interactions between the two of them.

The payment method was found to be of moderate importance, where the least used method also was the most abandoned one. An error message related to a discount was shown to 20% of the customers. When this message occurred the cart had a 20% probability of being abandoned.

Inledning

Företaget, vars webbstatistik ligger till grund för analyserna i uppsatsen, verkar inom retail. Produk-tsortimentet innehåller mode för dam, barn och herr samt inkluderar även produkter inom kategori-erna möbler, heminredning och elektronik. Under de diskussioner som föranlett uppsatsen uttrycker företagaren en önskan att få veta när det är tryggt för dem att störa en kund i kassan, exempelvis med en pop-up eller liknande. Går det att hitta indikationer på att vissa kunder har en större sannolikhet än andra att slutföra eller avsluta ett köp?

Företagare inom detaljhandeln som bedriver e-handel verkar i en starkt konkurrensutsatt bransch, där de inte bara konkurrerar mot lokala butiker i kundens närområde, utan också på en global marknad. Att lära känna kunderna och optimera sin verksamhet är därför av yttersta vikt. Genom att imple-mentera ett analysprogram som registrerar besökare samt dess aktivitet på webbplatsen får företagen en möjlighet att optimera affären och åtgärda eventuella handelshinder. Företagen kan också få insikt i hur väl marknadsföringens fungerar, följa försäljningen i realtid och få en djupare insikt om målgruppen.

Kanske är optimeringen av kundens digitala upplevelse viktigare än någonsin, i en tid när mycket har och kommer att ändras kring hur företag får insikt om sina kunder.¹ När företagen inte längre kan följa kundens köpresa utanför den egna webbplatsen, blir insikter baserade på den egna webbstatistiken allt viktigare.

Men vad är det egentligen som får en kund att konvertera, det vill säga köpa eller överge sin kundkorg? Spelar den tid som kunden lägger på webbplatsen någon roll? Matchar kundens behov inte med de produkter som finns? Eller är den övergivna kundkorgen en del av ett köpbeteende, exempelvis att kunden letar efter inspiration snarare än att köpa? Ja, det är just det som denna uppsats ämnar un-dersöka.

Uppsatsen tar avstamp från det att kunden loggat in på webbplatsen, lagt till en produkt i sin kundkorg och besökt kassan minst en gång. Därefter följer vi kunderna genom deras köpresa på webbplatsen. Eftersom det finns svart på vitt om kundkorgen köptes eller de facto övergavs, kan vi börja nysta i vad som kunnat tänkas hjälpt eller stjälpt. Förhoppningsvis kan insikter genererade från uppsatsen appliceras på den större mängd besökare som varje dag besöker företags webbplats.

¹Författaren syftar här på åtgärder specifikt infört för att begränsa åtkomsten till statistik kring kunder, exempelvis GDPR som infördes 2018 och Apples App Tracking Transparency (ATT) policy som infördes 2021.

Data

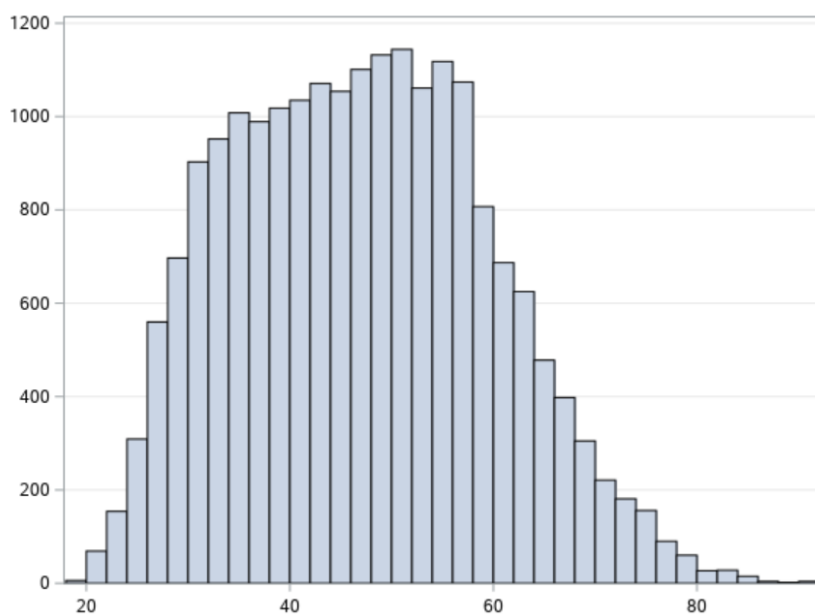
Populationen består av inloggade kunder som adderat minst en produkt till sin kundkorg samt besökt "Checkout" (sidan för betalning) minst en gång. Datamaterialet innehåller strax över 20 000 observationer. Saknade värden samt extremvärden visar sig stå för approximativt 2% av observationerna. Efter gallringen återstår 16 435 observationer. Varje observation motsvarar en kund tillika en kundkorg.

Tabell 1: Responsvariabel			
Variabel	Medelvärde	Min	Max
Ordrar	49,7% (övergivna)	0 st	4 st

För ett litet antal av observationerna har flera köp registrerats av en och samma kund. För att kunna erhålla sannolikheten att en kund överger sin kundkorg har variabeln därför transformerats till en binär variabel; 0 köp till värdet "0" och 1-4 köp till värdet "1". Variabeln är uppsatsens responsvariabel och kommer hädanefter att refereras till som "Kundkorg", där kundkorgar med värdet "0" antingen benämns som "Övergiven" eller "Inget köp" i grafer och löpande text.

Tabell 2: Kontinuerliga variabler			
Variabel	Medelvärde	Min	Max
Ålder	47 år	19	91
Orderhistorik	9 st	0	67
Besökta sidor	18 st	2	57
Tid	21 min	0	158
Checkout	2 st	1	9
Produkter i kundkorg	3 st	1	18

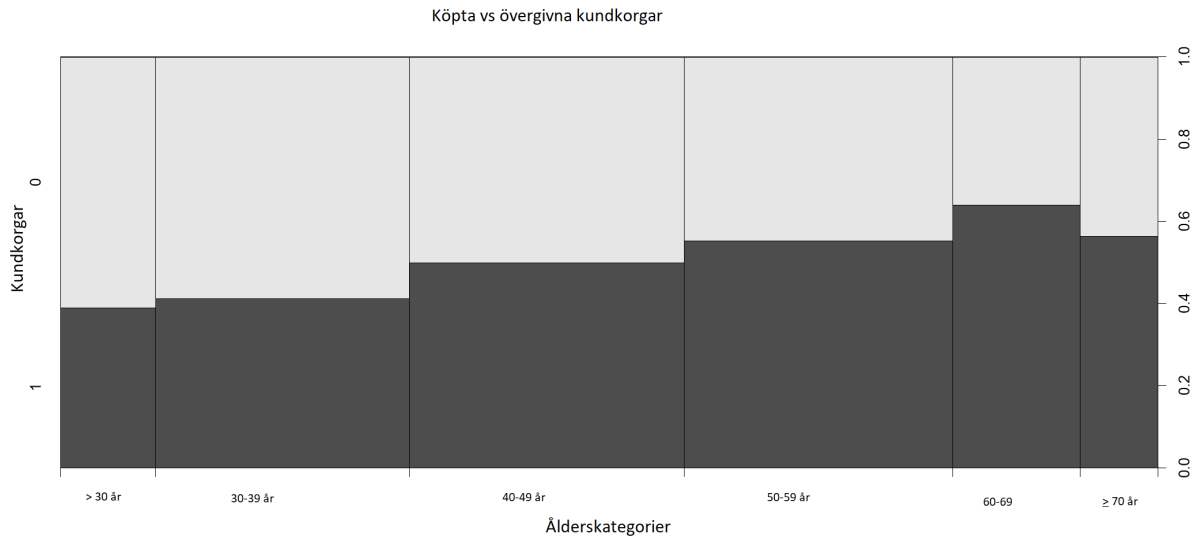
I tabell 2 listas sex kontinuerliga förklarande variabler samt medelvärden och minsta och största värden.



Figur 1: Histogram ålder

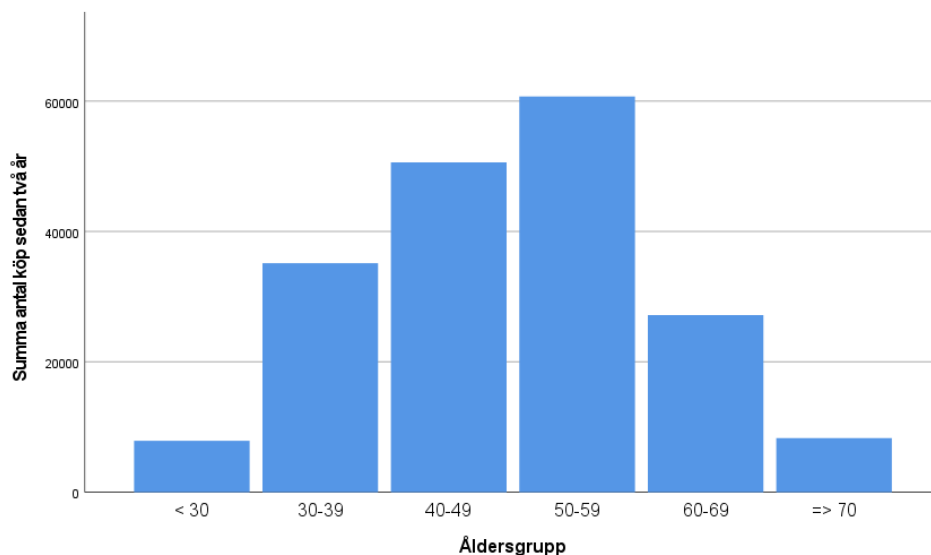
Webbplatsen besöks under observerad tid av kunder mellan 19 och 91 år, med genomsnittet 47 år.

I figur 1 visas ett histogram över åldersfördelningen, där det verkar som att majoriteten är mellan 30-60 år. För att kunna undersöka eventuella skillnader mellan åldrarna delas kunderna in i sex ålderskategorier. I kategorin "< 30" räknas alla kunder mellan 19-29 år in. I kategorin "=>70" räknas alla kunder som är 70 år eller äldre.



Figur 2: Graf som visar andel köpta respektive övergivna kundkorgar uppdelat per ålderskategori

Densitetsgrafen i figur 2 visar ålderskategorier på x-axeln samt antal kundkorgar på y-axeln. Massan delas upp i ljus-, samt mörkgrått där den ljusgråa färgen representerar andelen övergivna kundkorgar och den mörkgråa färgen köpta kundkorgar. Det verkar som att unga kunder är särskilt benägna att överge sin kundkorg jämfört med äldre - där kunder mellan 60-69 år sticker ut med högst andel köpta kundkorgar.



Figur 3: Stapeldiagram som visar summa av antal köp sedan två år, uppdelat per ålderskategori

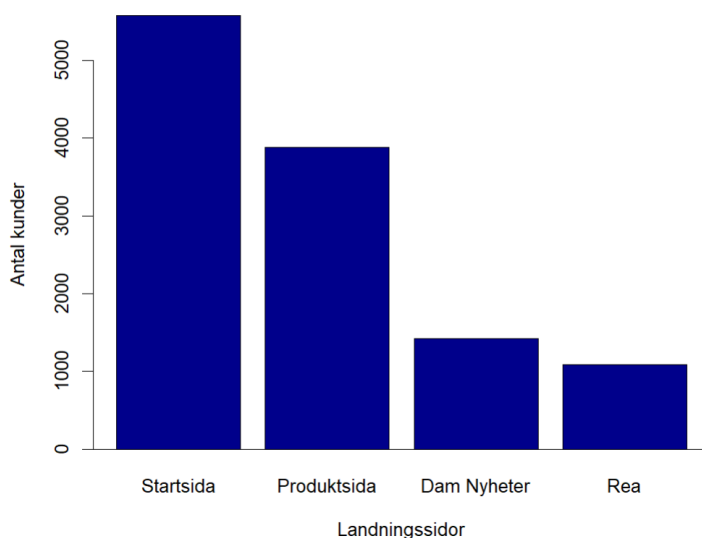
Genom att granska kundernas orderhistorik kan vi veta ytterligare något om deras köpbeteende. Kanske är frekvent återkommande kunder mindre benägna att överge sin kundkorg? I figur 3 visar x-axeln

ålderskategorier och på y-axeln ser vi summa av antal tidigare ordrar (här benämnda som köp). Genomsnittligt antal tidigare ordrar är 6 st. Tittar vi närmare på stapeldiagrammet i figur 3, ser vi att den ålderskategori med flest antal tidigare åldrar är dem mellan 50 och 59 år. Dem med minst antal tidigare ordrar är kunder yngre än 30 år och dem som är 70 år eller äldre.

En liten del av kunderna har ingen tidigare order. För att undersöka ifall det finns några skillnader mellan dem som inte gjort någon tidigare order och övriga, transformeras variabeln "Orderhistorik" om till en ny binär variabel som döps till "Kundtyp". Kunder med minst en tidigare order benämns som en "återkommande kund" och variabeln "Kundtyp" får värdet "1", och kunder med inga tidigare ordrar benämns som "ny kund" och variabeln "Kundtyp" får värdet "0".

När kunden anländer till en webbplats registreras inte bara vilken sida kunden kommer till, utan också vilken sida kunden kom ifrån. I praktiken registreras alla sidor kunden besöker. Det är så vi vet att kunden granskat sin kundkorg, velat betala sin kundkorg, osv.

Den första sida som kunden kommer till kallas landningssida. Genom att skapa ett nytt dataset som begränsas till de mest frekvent besökta landningssidorna kan en djupare analys genomföras om en korrelation finns med kundkorgens utfall. Fyra sidor innehar cirka 70% av alla besök, och har tagits med i det nya datasetet: webbplatsens startsida, en översiktlig produktsida (alla produkter), sidan för nyheter inom produktkategorin "Dam", samt en sida för produkter på rea.

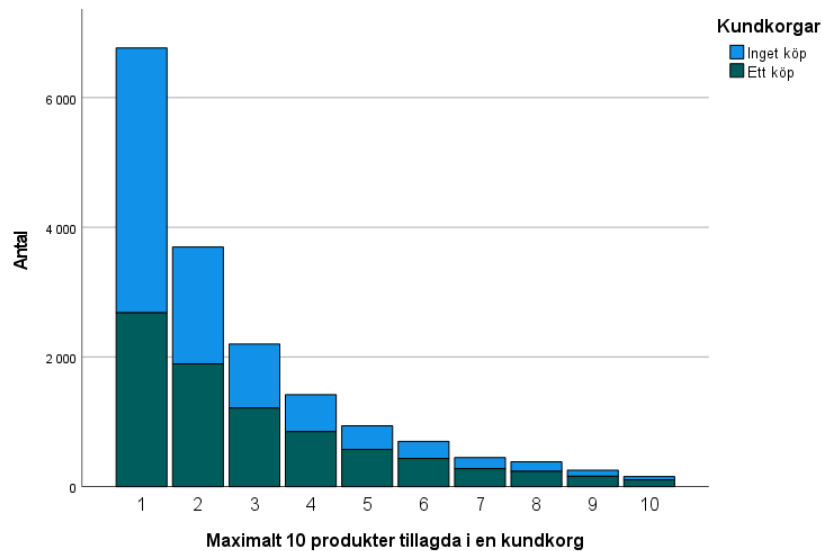


Figur 4: Stapeldiagram som visar de mest frekvent besökta landningssidorna

Sidorna som kunder besöker registreras och listas i datamaterialet. Sidorna aggregeras per unik kund och sammanställs till en numerisk variabel, därmed får vi veta att i genomsnitt klickar sig kunderna igenom drygt 20 sidor per besök.

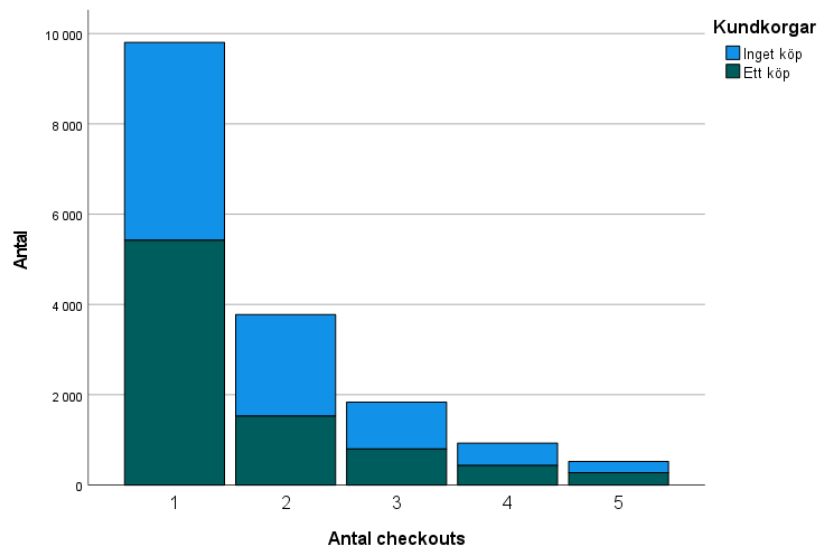
Även tiden en kund befinner sig på varje sida registreras. Tiden grupperas in i sex nivåer, där A är 0-15 sek, B är 15-30 sek och så vidare och F är total tid, det vill säga 30 min. För att erhålla approximativt totalt spenderad tid kunden besöker webbplatsen transformeras variabeln från en kategorisk till en numerisk genom att omvandla grupp A till 15 sek, B till 30 sek, och så vidare. Därefter aggregeras tiden per unikt KundID. Genom detta fås ett medelvärde, vilket är att en kund spenderar drygt 20 minuter på webbplatsen.

Under de 20 minuter som kunden spenderar på webbplatsen adderas i genomsnitt 3 produkter till kundkorgen. Stapeldiagrammet i figur 5 visar antal produkter tillagda i en kundkorg på x-axeln och antal kunder på y-axeln. Stapeldiagrammet är begränsad till maximalt 10 produkter. Kundkorgens utfall är märkt med blått och grönt, där blått representerar andel övergivna kundkorgar. De allra flesta verkar lägga till 1-2 produkter. Andelen övergivna kundkorgar verkar minska med antalet produkter.



Figur 5: Stapeldiagram antal tillagda produkter i kundkorgen

När kunden går till sidan för betalning, kallas det för en "Checkout". Vi påminner oss om att alla kunder i materialet besökt "Checkout" minst en gång. Från stapeldiagrammet i figur 6 ser vi antal checkouts begränsade till maximalt 5 stycken på x-axeln samt frekvens av kundkorgar på y-axeln. Staplarnas färg är definierade utefter utfallet av kundkorgen.



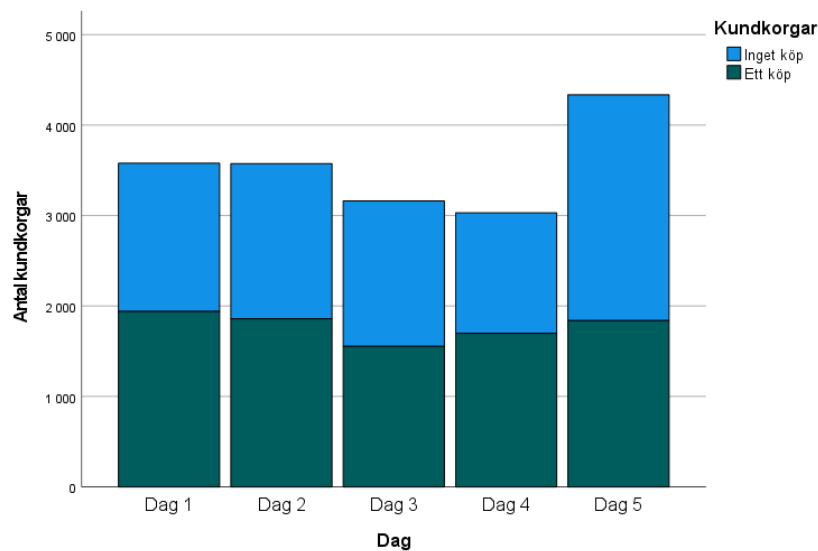
Figur 6: Stapeldiagram antal checkouts

Det genomsnittliga antalet gånger som kunden går till kassan är två gånger. Vi kan i figur 6 se att de allra flesta går till kassan 1 gång och det verkar som att andelen övergivna kundkorgar i någon mån ökar med antal checkouts.

Tabell 3: Kategorivariabler			
Variabel	Nivåer	Variabel	Nivåer
Dag	5	Produktkategorier	5
Kundtyp	2	Betalningsmetod	3
Kanal	5	Felmeddelande	2
Landningssida	4		

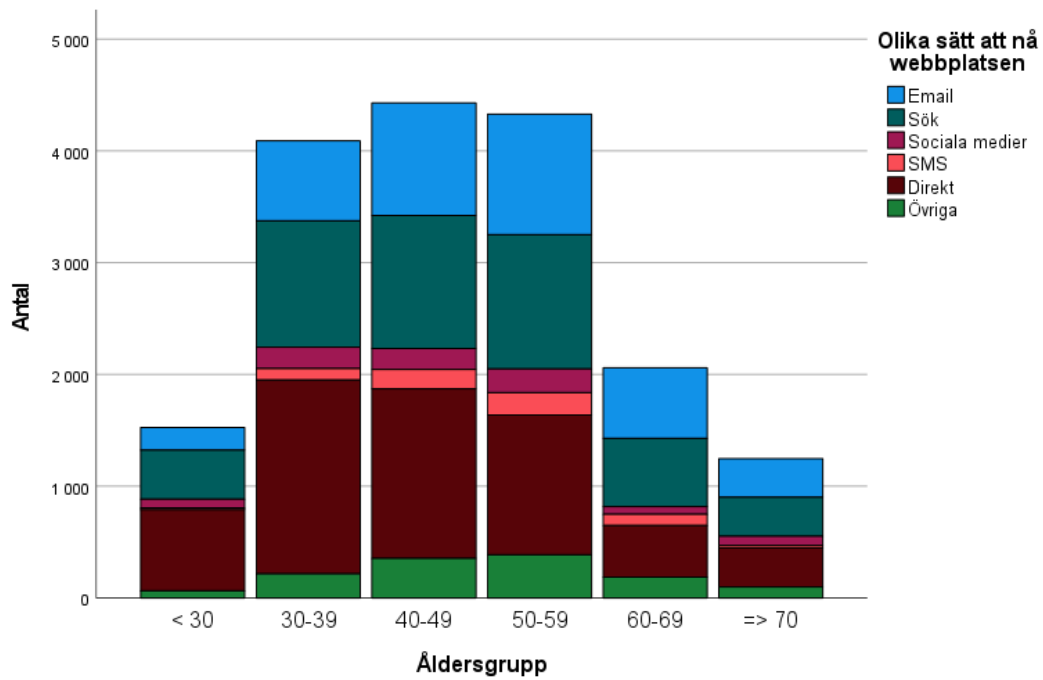
I tabell 3 listas 7 förklarande kategorivariabler.

Datamaterialet innehåller observationer registrerade under fem olika dagar, från ett tidigare år. Det är inte känt för författanden vilket år det gäller, ej heller om dagarna är sammanhängande eller vilka veckodagar de härrör från. I figur 7 ser vi ett stapeldiagram där observerade dagar återfinns på x-axeln samt antal kundkorgar återfinns på y-axeln. Färgerna blått och grönt representerar övergivna respektive köpta kundkorgar. Antalet besökande kunder varierar över de olika dagarna, där Dag 5 verkar innehålla fler kunder än övriga dagar. Andelen kundkorgar som överges representeras av den blåa färgen i staplarna.



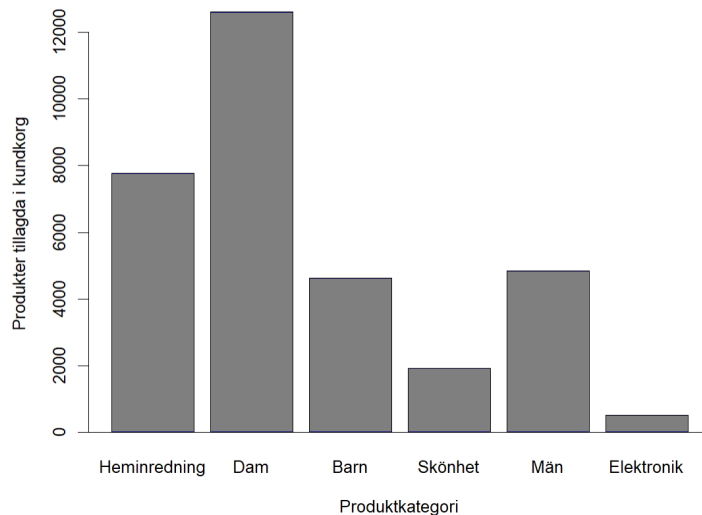
Figur 7: Stapeldiagram antal besök per dag

Kunderna i stickprovet anländer till webbplatsen genom olika kanaler, det vill säga platsen som de var på precis innan de landade på webbplatsen, som exempelvis en sökmotor eller ett social media. I datamaterialet kommer de flesta av kunderna till webbplatsen genom att skriva in namnet i webbläsaren, söka fram det i en sökmotor eller genom att klicka på en länk i ett email skickat av företaget. Figur 8 visar åldersgrupper på x-axeln och antal kundkorgar på y-axeln. Färgerna på staplarna representerar olika kanaler. Vi ser att färgerna som dominerar är mörkgrönt, vinrött och blått. Det vill säga tidigare nämnda kanaler. Den betalda marknadsföringen från sökmotorer har sorterats in under "Övriga". Från stapeldiagrammet ser vi att det verkar finnas ålderskillnader i hur kunderna anländer till webbplatsen.



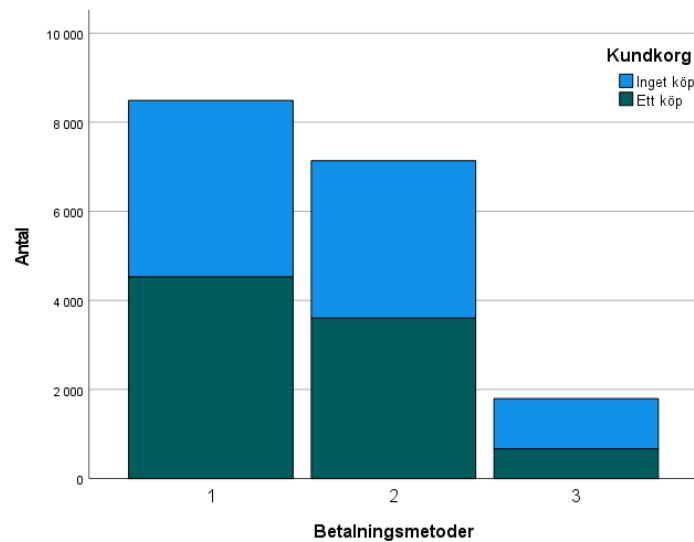
Figur 8: Stapeldiagram antal kundbesök per dag

Det finns sex produktkategorier där produkterna rimligtvis borde skilja sig i många avseenden, exempelvis prisbild och behov, både inom kategorin men även mellan de olika kategorierna. Stapeldiagrammet i figur 9 visar kundernas mest tillagda produkter med produktkategorier på x-axeln samt summa produkter på y-axeln. Här sticker kategorin Dam ut med mest tillagda produkter, med heminredning tätt följt därefter.



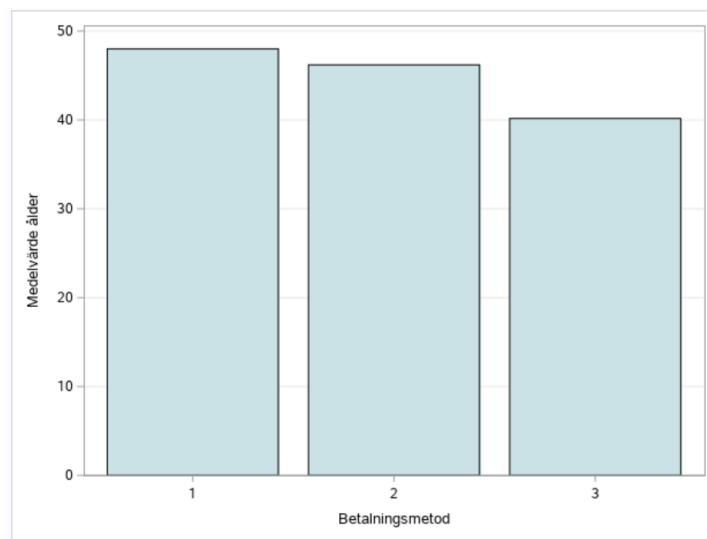
Figur 9: Stapeldiagram som summerar tillagda produkter från respektive produktkategori

Väl i kassan har kunden tre möjliga betalningsmetoder, benämnda som 1, 2 och 3. Betalningsmetod 1 verkar mest frekvent använt, följt av betalningsmetod 2 och sist metod 3 som få verkar välja (se figur 10).



Figur 10: Stapeldiagram som visar betalningsmetoder

Stapeldiagrammet figur 10 visar betalningsmetod på x-axeln samt antal kundkorgar på y-axeln. Diagrammet verkar visa att fler kundkorgar blir köpta med betalningsmetod 1 än övriga metoder. Andelen kundkorgar som överges representeras av blått, samt andelen kundkorgar som köps representeras av grönt. Här kan vi se att överlag verkar det inte finnas några större skillnader mellan betalningsmetod 1 och 2. Däremot sticker betalningsmetod 3 ut med en högre andel övergivna kundkorgar.



Figur 11: Stapeldiagram som visar medelvärdet av ålder i olika betalningsmetoder

I figur 11 ses ett stapeldiagram där betalningsmetod syns på x-axeln och medelvärde i ålder ses på y-axeln. Vi kan se att medelåldern sjunker från betalningsmetod 1 till 3.

I kassan kan ett felmeddelande visas för kunden om ett fel uppstått när kunden angivit en rabattkod eller dylikt. Felmeddelandet visas för en femtedel av kunderna i stickprovet.

Metodbeskrivning

För att testa oberoende mellan kategoriska variabler används Pearson χ^2 test applicerat på utfallet i en korstabell baserad på de två kategoriska variablerna. Testet togs fram år 1900 av Karl Pearson och jämför de observerade frekvenserna i de olika "cellerna" med de förväntade antalen vid oberoende och de givna observerade marginalfrekvenserna för de två kategorivariablerna. Resultatet som fås är en sammanvägning av skillnaderna (i kvadrat) mellan observerade frekvenser och förväntade frekvenser i de olika cellerna. En hypotes om oberoende mellan de kategoriska variablerna förkastas om χ^2 -värdet är onormalt stort i jämförelse med om oberoende föreligger. Det så kallade p-värdet (dvs sannolikheten att erhålla ett så stort χ^2 -värde eller mer, vid oberoende) är då mindre än $\alpha = 0.05$, den så kallade nivån på testet. Denna nivå har använts genomgående i alla tester. Testet har främst använts för att avgöra om den binära responsvariabeln "Kundkorg", är beroende av en annan kategorisk variabel.

För att undersöka inverkan på responsvariabeln, som beskriver om en kundkorg blir köpt eller övergiven, från flera andra kategoriska variabler och även från kontinuerliga eller diskreta kvantitativa variabler, används en logistisk regressionsmodell. En logistisk regressionsmodell är ett specialfall av en generaliserad linjär modell och kallas ibland för logitmodell.

I en logistisk regression beskrivs sannolikheten $\pi(x)$ för att den binära variabeln antar värdet 1 givet att den förklarande variabeln antar värdet x . Denna relation beskrivs i termer av log-oddset som beskrivs som en linjär funktion av x :

$$\log\left(\frac{\pi(x)}{1 - \pi(x)}\right) = \beta_0 + \beta_1 x.$$

Sannolikheten $\pi(x)$ kan då ekvivalent uttryckas som

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$$

vilken som en funktion av x har en form som ofta benämns som en sigmoid funktion.

Om $\beta_1 > 0$, ökar $\pi(x)$ när x ökar. Om istället $\beta_1 < 0$ minskar $\pi(x)$ när x ökar. Parametern β_1 beskriver hur, och hur starkt beroendet av x ser ut. Vid flera förklarande variabler ser motsvarande relation ut som

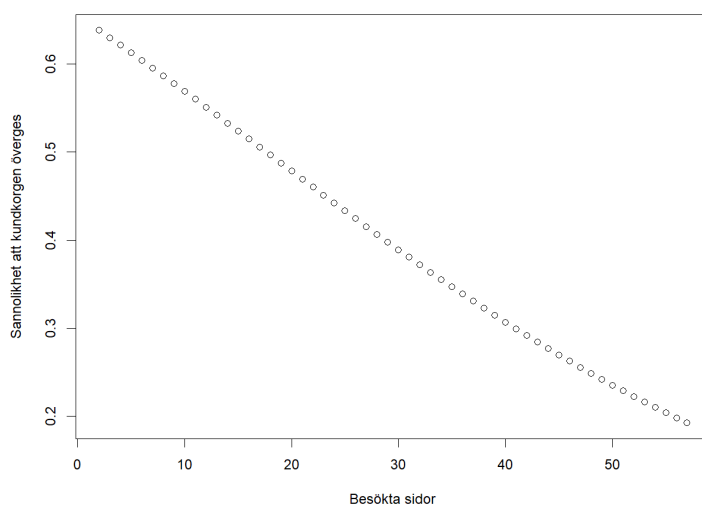
$$\pi(x_1, x_2, x_3) = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3}}$$

Baserat på data om utfall på den binära variabeln och motsvarande värden på de förklarande variablerna, uppskattas värden på parametrarna $\beta_0, \beta_1, \beta_2, \beta_3$. I denna undersökning har programvaran R använts, som ger maximum-likelihood-skattningar av parametrarna. Den deskriptiva statistiken utförs i programvarorna SPSS, R och SAS. En närmare beskrivning av programvarorna ges i listningen av referenser på sidan 17.

Köpresan börjar redan innan kunden anlänt till webbplatsen

Analysen av den övergivna kundkorgen är egentligen en resa i kundens fotspår. En resa som tar sin början redan på sidan innan ankomsten till webbplatsen och som sedan snirklar sig fram genom webbplatsens alla vrår, för att till slut resultera i en köpt eller övergiven kundkorg. På vägen mot kassan analyseras vad som verkar ha en inverkan på kundkorgens utfall.

Kunden anländer och vi kan se att kunderna verkar ha företaget färskt i minnet, vilket märks genom att en tredjedel av kunderna kommer in på webbplatsen genom att skriva in webbadressen i webbläsaren. Genom en logistisk regression framkommer det dock att enbart sökmotorer har en signifikant inverkan på om kundkorgen blir övergiven.



Figur 12: Prediktion besökta sidor

De som anländer genom en organisk länk har en lägre sannolikhet att överge sin kundkorg än de som anländer genom en sponsrad länk. Tas hänsyn till ålder faller sökmotorer bort och SMS blir istället enda kanalen med en signifikant inverkan på kundkorgens utfall. Bland dem som anländer genom ett SMS förväntas fler överge sin kundkorg, denna grupp är dock liten i relation till antal observationer.

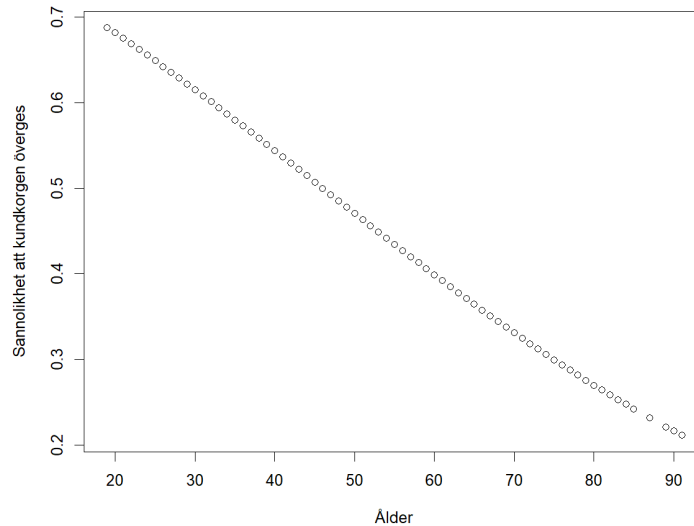
En logistisk regressionsmodell gjord på de mest frekvent besökta landningssidorna visar att det är främst underkategorin "produktsida" samt nyheter för dam, en rea-sida, kundkorgssidan (som är den föregående sidan till checkout) samt kategorisidan för dam som har en inverkan på kundkorgens utfall.

I figur 12 ses en prediktion från en logistisk regressionsmodell som visar att antal sidor kunden går igenom har en inverkan på kundkorgen, och att ju fler sidor kunden går igenom, desto lägre blir sannolikheten för att kunden ska överge kundkorgen. Det verkar som att sannolikheten sjunker först efter att kunden överstigit genomsnittet på 20 sidor.

Ökar sannolikheten för att kundkorgen ska överges, när tiden på webbplatsen ökar? En logistisk regressionsmodell visar att den uppskattade sannolikheten att kunden överger sin kundkorg minskar med drygt 2% per ytterligare minut. Ju längre tid som går, desto lägre blir sannolikheten för att kundkorgen ska överges.

Kunden loggar in

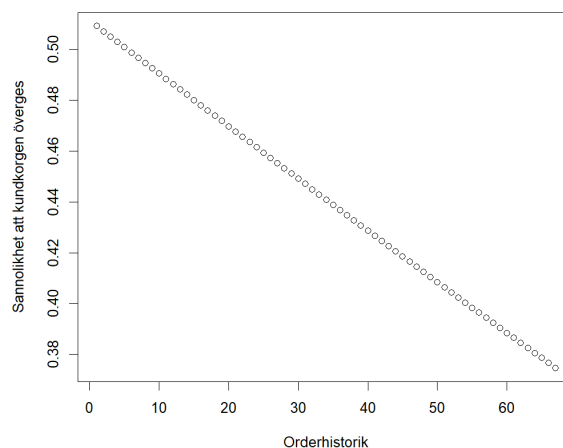
Kunden loggar in och knyts till ett KundID. Vi vet därför, som tidigare beskrivits, bland annat kundens ålder och orderhistorik. Att det finns ett samband mellan ålder och övergivna kundkorgar blir tydligt i en prediktion som bygger på en logistisk regressionsmodell. I figur 13 ses ålder på x-axeln och skattad



Figur 13: Graf uppskattad sannolikhet ålder

sannolikhet för en kundkorg att bli övergiven på y-axeln. Vi kan se att kunder runt 20 år förväntas i 70% av fallen överge sin kundkorg. Ju yngre kunden är, desto mer troligt är det alltså att kundkorgen överges.

Har kundens tidigare orderhistorik någon betydelse för dagens besök? Ett χ^2 -test visar att det finns en signifikant skillnad mellan dem som inte gjort ett tidigare köp samt dem som gjort en eller flera tidigare köp. En logistisk regression visar att nya kunder har en högre sannolikhet att överge sin kundkorg.



Figur 14: Prediktion Orderhistorik

I figur 14 ser vi antal ordrar på x-axeln och sannolikheten att överge sin kundkorg på y-axeln. Det skiljer bara drygt 10 procentenheter i sannolikhet mellan 10 ordrar och 50 ordrar. Det verkar som att antalet

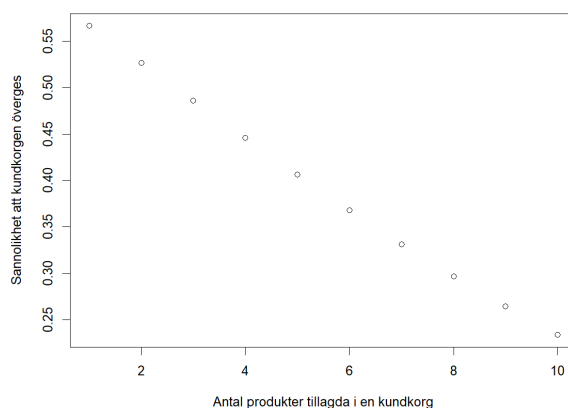
ordrar inte verkar vara av särskilt stor vikt, utan det viktigaste momentet för att sannolikheten för en övergiven kundkorg ska minska är att en order överhuvudtaget lagts.

Har det någon betydelse om vilken dag kunden loggat in? Om det är en tisdag eller en lördag? Eller en dag i början på månaden, jämfört med en dag i slutet på månaden? Det verkar finnas skillnader mellan dagarna, genom ett signifikant resultat i ett χ^2 test. I en fördjupad granskning med en logistisk regression framkommer att enbart dag 3 och 5 kan sägas ha en signifikant inverkan på att kundkorgen ska överges, där det även verkade vara så att det under dag 5 förväntades fler övergivna kundkorgar.

Eftersom ålder har spelat in tidigare, granskas samspelseffekter med ålderskategorin 40-49 som referensgrupp. Det visar sig att när hänsyn tas till ålder, har alla dagarna förutom dag 4 en signifikant inverkan på övergivna kundkorgar. Det verkar också finnas ett samspel mellan dag 2 och 4 och kunder under 30 år, samt mellan dag 3 och dag 4 och kunder mellan 50-59 år.

En produkt läggs till i kundkorgen

Har antal produkter i kundkorgen en inverkan på om den ska bli övergiven eller ej? En logistisk regressionsmodell visar att det är så. I figur 15 ser vi antal produkter tillagda i en kundkorg på x-axeln och sannolikheten för att kundkorgen ska överges på y-axeln. Vi ser att sannolikheten minskas för varje ny produkt som läggs till. Exempelvis ser vi att en kund som lägger till 4 produkter i sin kundkorg har en uppskattad sannolikhet på 45% att överge sin kundkorg.



Figur 15: Prediktion sannolikhet för olika storlekar på kundkorgen

Vi har tidigare sett att kategorin "Dam" har flest tillagda produkter. Ett χ^2 -test visar ett signifikant resultat i kategorierna "Dam", "Heminredning" samt "Skönhet". När inverkan granskas i separata logistiska regressionsmodeller framkommer det dock att endast kategorierna "Dam" och "Heminredning" har en signifikant inverkan på kundkorgens utfall. Det framkommer att kundkorgar som innehåller en produkt från kategorin:

- "Heminredning", har en uppskattad sannolikhet att överges på 65%.
- "Dam", har en uppskattad sannolikheten att överges på 36%.

Det är alltså mer troligt att en kundkorg överges om den innehåller en produkt från kategorin "Heminredning" än från "Dam".

Eftersom vi vet sedan tidigare att kunder lägger till flera produkter i kundkorgen, testas också modeller där hänsyn tas till att antal produkter i kundkorgen varierar.

Kunden går till kassan

Har antalet checkouts en inverkan på utfallet av kundkorgen? En logistisk regressionsmodell visar att för varje checkout en kund gör, ökar risken för en övergiven kundkorg med strax över 5% per checkout. När hänsyn tas till ålder, minskar sannolikheten minimalt. Det visar sig dock att det finns en samspelseffekt mellan checkout och åldersgrupperna 30-39, 60-69 samt 70+.

Man kanske också kan tänka sig att det finns olika effekter av att storleken på kundkorgen varierar. Tas hänsyn till detta, dubblas sannolikheten att kundkorgen överges. Det finns dock inget samspel mellan antal checkouts och antal produkter i kundkorgen.

I ett försök att utreda varför kunden lämnar checkout, undersöks vilka andra besökta sidor som korrelerar med sidan för checkout. Genom en logistisk regressionsmodell gjord på det begränsade datasetet med de mest frekvent besökta sidorna återfinns korrelationer mellan checkouts och sidorna: Nyheter för dam, Rea, samt kategorisidan för Dam.

Har betalningsmetoden en inverkan på om en kundkorg överges? Ett χ^2 -test bekräftar att betalningsmetoderna signifikant skiljer sig åt. För att vidare utreda hur den antydda inverkan ser ut görs en logistisk regression, med betalningsmetod 3 som referensgrupp. Det framkommer att en kundkorg förväntas överges oftare med betalningsmetod 1 än med metod 2. När hänsyn tas till variationer bland ålderskategorierna, sjunker skillnaderna mellan betalningsmetoderna. Enbart dem som är 60 år eller äldre har en minskad sannolikhet att överge kundkorgen.

För en femtedel av kunderna i kassan visades ett felmeddelande relaterat till en rabattkod. Fick kunden detta felmeddelande, höjdes sannolikheten att kunden skulle lämna kundkorgen med 20%.

Slutsats

Inför arbetet med uppsatsen uppstod diskussioner kring hur företagets intresse av att bevara företags-hemligheter skulle kunna jämkas med författarens intresse av att producera en läsvärdig rapport. Diskussionen resulterade i att företaget kommer hållas anonym samt att vissa namn byts ut. Vissa av variablerna samt dess nivåer är av samma anledning ersatta med generiska namn.

För att undvika påverkan från större kampanjer då konverteringsgrad köp inte kan antas vara representativt har dagarna inte valts ut slumpmässigt. I datamaterialet är konverteringsgraden 50,3%. Den höga konverteringsgraden förklaras förmodligen av populationen som varit i kassan minst en gång.

I sektionen data kan förekomsten av extremvärden ses i datamaterialet. Dock förändrades inte medelvärdena av rensningen av observationer. Ett fåtal kunder som förekommit under flera av de observerade dagarna har för enkelhetens skull rensats ut.

Det verkar som att webbplatsen ligger högt i top of mind hos kunderna, merparten skriver in webbadressen direkt i webbläsaren eller i ett sökfält hos en sökmotor. Dem som anländer genom en organisk länk har en lägre sannolikhet att överge sin kundkorg än bland dem som anländer genom en sponsrad länk. Tas hänsyn till ålder faller sökmotorer bort och SMS blir istället enda signifikanta kanalen som har en signifikant inverkan på om en kundkorg ska bli övergiven, där fler förväntas överge sin kund-

korg. Bland dem som anlände genom SMS verkar flest vara mellan 40-59 år.

Ju fler sidor kunden går igenom, desto lägre blir sannolikheten för att kunden ska överge kundkorgen. Men det verkar som att sannolikheten sjunker först efter att kunden överstigit genomsnittet på 20 sidor.

Det verkar logiskt att ju längre tid en kund spenderar på webbplatsen, desto fler aktiviteter hinner kunden genomföra. Men innebär det att konverteringsgraden ökar, när tiden ökar? Vi ser att genomsnittstiden för en kunds besök på webbplatsen är drygt 20 minuter. De som spenderade mindre än en kvart på webbplatsen verkar överge sin kundkorg oftare, än de som lägger uppemot en halvtimme och längre. Dock är det tveksamt hur stort prediktivt värde tiden har, eftersom majoriteten av kunderna sågs lämna webbplatsen efter senast 30 minuter och sannolikheten för att kundkorgen ska överges minskar först efter det.

Den vanligaste kunden att besöka webbplatsen i stickprovet är kvinnor strax under 50 år och den vanligaste köparen är lite äldre än så - mellan 50 och 60 år. De som överger flest kundkorgar är 30-39 år men procentuellt sett är den yngsta kundgruppen som förväntas överge flest kundkorgar. Dem under 30 år förväntas i 70% av fallen överge sin kundkorg. Ju yngre kunden är, desto mer troligt är det alltså att kundkorgen övergivs. Åldersvariationen i köpbeteende följer med hela köpresan, från vilken kanal kunden kommer ifrån till vilken betalningsmetod som väljs.

Det viktigaste är inte hur många tidigare orders en kund lagt, utan om en order överhuvudtaget har lagts. Det är alltså bland dem som inte har en tidigare order som en negativ inverkan på sannolikheten att kundkorgen överges hittas.

Dag 3 och 5 verkar ha en signifikant inverkan på att kundkorgen ska överges, där det även verkar vara så att det under dag 5 förväntas fler övergivna kundkorgar. När hänsyn tagits till eventuella variationer bland åldrarna har alla dagar (förutom dag 4) en signifikant inverkan på kundkorgens utfall.

Det verkar finnas ett samspel mellan dag 2 och 4 samt med kunder under 30 år. Det verkar också finnas ett samspel mellan dag 3 och 4 samt dem mellan 50-59 år.

Majoriteten av kunderna lägger till 1-3 produkter i kundkorgen. För varje ytterligare produkt som läggs till i kundkorgen, minskar sannolikheten att kundkorgen överges. Kunder som exempelvis adderat fyra produkter till sin kundkorg, förväntas strax under 45% överge kundkorgen. Med tanke på att de flesta adderar 1-3 produkter är denna prediktion rimligtvis inte av större värde. När hänsyn tas till kategorierna "Dam" samt "Heminredning", påverkar det inte logoddsen i någon större grad.

Flest produkter läggs till från kategorierna "Heminredning" eller "Dam". Vi kan se att det är mer troligt att en kundkorg överges om den innehåller en produkt från kategorin "Heminredning" än om den innehåller en produkt från "Dam".

Eftersom kunder lägger mer än en produkt i kundkorgen, borde det rimligtvis innebära förändringar i sannolikhet när kundkorgen blandas med produkter från olika kategorier. Det visar sig dock att kategoriernas korrelation inte är särskilt stor.

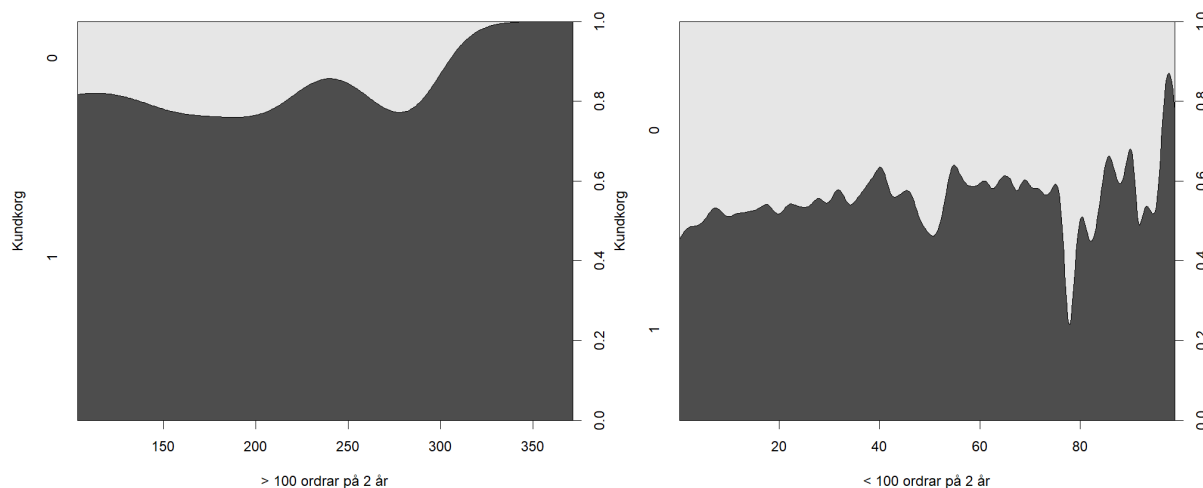
När kunden ska betala sin kundkorg och går till "Checkout", visar en logistisk regressionsmodell att varje gång kunden lämnar sidan för "Checkout", ökar sannolikheten att kundkorgen ska överges med strax över 5%. Tas hänsyn till antal produkter tillagda i kundkorgen, ökar sannolikheten att kundkorgen

ska överges drastiskt. Det verkar som att kunderna lämnar "Checkout" för att besöka sidor med produkter från kategorin Dam samt sidor som innehåller ordet "Rea". Det verkar inte finnas några variationer bland åldrarna.

Det framkommer att en kundkorg förväntas överges oftare med betalningsmetod 1 än med metod 2. Tittar vi närmare på hur åldersfördelningen ser ut kan vi i figur 17 se att medelvärdet av åldern är högst i betalningsmetod 1 och lägst i betalningsmetod 3. Sedan tidigare vet vi att ju yngre en kund är, desto högre är benägenheten att överge en kundkorg. När hänsyn tas till variationer bland ålderskategorierna, sjunker skillnaderna mellan betalningsmetoderna. Enbart dem som är 60 år eller äldre har nu en minskad sannolikhet att överge kundkorgen.

För en femtedel av kunderna i kassan visades ett felmeddelande relaterat till en rabattkod. Fick kunden detta felmeddelande, hade kunden en uppskattad sannolikhet att lämna kundkorgen på 20%.

Det hade varit intressant att se om åldern har något samband med tiden som spenderas på webbplatsen. Vi vet sedan tidigare att äldre kunder har en lägre sannolikhet att överge kundkorgen, är det även dem som spenderar mer tid på webbplatsen?



Figur 16: Densitetsgrafer som visar utfallet för kundkorgar uppdelat i över eller under 100 tidigare ordrar

Under arbetets gång upptäcktes en tendens till ett extremt beteende. Ett exempel är variabeln "Orderhistorik". Innan materialet rensades på extremvärden gjordes två densitetsgrafer, som visas i figur 16. Dem illustrerar hur andelen genomförda köp skiljer sig åt beroende på antal tidigare ordrar. Grafen till höger visar majoriteten: kunder med mindre än 100 ordrar. Grafen till vänster visar kunder med över 100 ordrar. Den mörkgrå färgen representerar massan av köpta kundkorgar, den ljusgrå massan övergivna kundkorgar. Det verkar som att när antal ordrar går mot ett extremt värde, ökar andelen köpta kundkorgar kraftigt. Innebörden av detta hade varit intressant att undersöka vidare.

Referenser

Agresti. An Introduction to Categorical Data Analysis, Wiley, 2007

Everitt, Hothorn. A handbook of statistical analyses using R, Taylor and Francis Group, 2010.

R Core Team. The R Stats Package, Version 4.1.1, 2021.

SAS Studio, OnDemand for Academics, Version 9.4 for Windows.

IBM SPSS Statistics, Version 27.0.1.0.