**E-commerce Kunddataanalys**

En rapport för att analysera e-handelskundsdata



Maya Sobh

EC Utbildning

Examensarbete- Byt namn

202410

# Abstract

En kort sammanfattning över ditt arbete och de viktigaste resultaten skrivet på engelska, cirka 5 meningar totalt.

This project presents a comprehensive analysis of e-commerce sales data, utilizing various data visualization tools such as Dash, Bokeh, and matplotlib to create interactive insights. By examining customer demographics and purchasing behavior, the study identifies key trends, including the highest and lowest-performing product categories. The analysis further explores sales distribution by gender and age group, revealing demographic influences on purchasing patterns. Notably, the study highlights top-spending customers and product categories, providing actionable insights for targeted marketing strategies. This data-driven approach offers valuable information to optimize product offerings and enhance customer engagement in e-commerce.

**Skapas automatiskt i Word genom att gå till Referenser > Innehållsförteckning.**

Innehållsförteckning

[1 Inledning 1](#_Toc176961624)

[1.1 Underrubrik – Exempel 1](#_Toc176961625)

[2 Teori 2](#_Toc176961626)

[2.1 Kundsegmentering 2](#_Toc176961627)

[2.1.1 Köpbeteende och Demografisk Analys…………………………………………………………………………..2](#_Toc176961628)

[2.1.2 Dataanalys och Visualisering 2](#_Toc176961629)

[3 Metod 3](#_Toc176961632)

[3.1 Datainsamling 3](#_Toc176961633)

[3.2 Agil arbetsmetodik 3](#_Toc176961634)

[4 Resultat och Diskussion 4](#_Toc176961635)

[5 Slutsatser 5](#_Toc176961636)

[6 Självutvärdering 6](#_Toc176961637)

[Appendix A 7](#_Toc176961638)

[Källförteckning 8](#_Toc176961639)

# Inledning

Under de senaste åren har e-handel snabbt förändrat sättet företag och konsumenter interagerar på, tack vare teknologiska framsteg och förändrade konsumentpreferenser. Med allt fler som handlar online blir det viktigare att kunna analysera data för att förstå kundernas köpbeteende och optimera produktutbudet. Företag som vill konkurrera på denna dynamiska marknad kan dra stor nytta av insikter från dataanalys för att förbättra sina marknadsstrategier och öka försäljningen.

Detta projekt handlar om att analysera e-handelsdata för att få insikter om kunders köpbeteende och hur olika produkter presterar. Genom att använda interaktiva visualiseringsverktyg är syftet att identifiera mönster och trender som kan hjälpa företag att fatta bättre strategiska beslut.

Syftet med denna rapport är att analysera och visualisera e-handelsdata för att få insikter om kundbeteende och produktprestation. För att uppnå syftet kommer följande frågeställningar att besvaras:

1. Vilka produktkategorier har högst och lägst försäljning, och vilka trender kan identifieras inom dessa kategorier?

2. Hur påverkar kunddemografi, som ålder och kön, köpbeteendet inom e-handel?Underrubrik – Betydelsen av Dataanalys inom E-handel

# Teori

För att förstå kundbeteende och försäljningsmönster inom e-handel är det viktigt att granska vissa grundläggande teorier och begrepp som rör dataanalys och marknadsföring. Dessa teorier ger en bas för hur data kan användas för att analysera och tolka konsumentbeteenden, med målet att optimera produktstrategier och marknadsinsatser.

## Kundsegmentering

Kundsegmentering innebär att dela upp en kundbas i mindre grupper baserat på gemensamma egenskaper, som demografi, köpbeteende och preferenser. Segmentering är en central strategi inom marknadsföring då den gör det möjligt att anpassa budskap och erbjudanden till specifika målgrupper, vilket kan leda till ökad kundlojalitet och bättre försäljningsresultat.

### Köpbeteende och Demografisk Analys

Köpbeteende är en process där konsumenter fattar beslut om vilka produkter de vill köpa, och denna process kan påverkas av faktorer som kön, ålder och inkomst. Genom att analysera dessa faktorer kan företag bättre förstå vilka kundsegment som är mest lönsamma och vilka produkter som tilltalar specifika demografiska grupper.

### Dataanalys och Visualisering

Dataanalys och visualisering är viktiga verktyg för att omvandla rådata till meningsfulla insikter. Med hjälp av verktyg som Dash, Bokeh och matplotlib kan data presenteras på ett sätt som gör den lättare att förstå och tolka. Genom att använda dessa tekniker kan företag snabbt identifiera mönster och trender i försäljningen och anpassa sin strategi för att möta kundernas behov.

# Metod

Hur har du genomfört ditt arbete? Exempelvis, hur har datan erhållits?

I detta avsnitt beskrivs hur projektet har genomförts, från datainsamling till hur den agila arbetsmetodiken har använts för att strukturera och driva arbetet framåt.

## Datainsamling

Exempel på en rubrik som är vanligt förekommande som läsaren förstår vart datan kommer ifrån.

Data för analysen hämtades från en ZIP-fil som innehåller en stor CSV-fil med e-handelsdata. Filen innehöll information om kunders demografi, köpmönster, produktkategorier och köpsummor. Efter att ZIP-filen extraherats, laddades CSV-filen in i en Python DataFrame för vidare bearbetning och analys. Data lagrades sedan i en SQLite-databas för att underlätta åtkomst och bearbetning via SQL-frågor. Denna process möjliggjorde effektiv hantering av stora datamängder och en flexibel analys av kund- och produktdata.

## Agil arbetsmetodik

I denna rapporten ska ni beskriva och utvärdera hur ni har arbetat agilt i projektformen. Se t.ex. följande länk vad agilt arbete kan innebära:   
<https://agilemanifesto.org/principles.html>

Om ni inte vill ha med detta i rapporten här för att ni t.ex. ska publicera ert arbete som ett portföljprojekt så kan ni skriva om detta i ett separat dokument.

Projektet genomfördes enligt principerna för agil arbetsmetodik, där arbetet delades upp i mindre iterativa cykler, eller "sprintar". Denna metod tillät regelbunden granskning och anpassning av projektets framsteg, vilket bidrog till en mer responsiv och flexibel arbetsprocess.

De viktigaste agila principerna som följdes under projektet inkluderar:

* **Anpassning till förändring**: Vid behov kunde projektets prioriteringar justeras baserat på insikter som framkom under dataanalysens gång.
* **Korta arbetscykler**: Arbetet delades in i korta perioder med tydliga mål, vilket gav möjligheten att snabbt utvärdera och justera arbetsmetoder och fokusområden.
* **Regelbunden återkoppling**: Genom att kontinuerligt granska resultaten i varje sprint kunde projektets riktning och metodik snabbt justeras för att bättre uppfylla syftet och besvara forskningsfrågorna.

Den agila arbetsmetodiken skapade en struktur som bidrog till att projektet genomfördes effektivt och med möjlighet till kontinuerlig förbättring, samtidigt som det möjliggjorde en flexibel respons på nya insikter under arbetets gång.

# Resultat och Diskussion

I detta avsnitt presenteras de huvudsakliga resultaten från analysen, följt av en diskussion kring deras innebörd och möjliga tolkningar. Fokus ligger på hur resultaten kan bidra till att förbättra förståelsen för kunders köpbeteende och produktprestationer inom e-handel.

## Resultat

Analysen av e-handelsdata resulterade i flera viktiga insikter:

1. Försäljning per produktkategori: De mest sålda produktkategorierna identifierades, och resultaten visade att vissa produktkategorier genererar betydligt mer intäkter än andra. De topp 5 mest sålda produkterna inkluderade exempelvis [beskriv specifika kategorier], medan de minst sålda kategorierna var [beskriv specifika kategorier]. Detta visar tydligt var företag bör fokusera sina resurser och marknadsföringsinsatser.

2. Kunddemografi och köpmönster: Försäljningsdata uppdelat efter kön och åldersgrupp visade att [specifika demografiska grupper] har ett större inflytande på försäljningen. Diagrammen visade att kvinnor/män (beroende på data) spenderar mer, och åldersgruppen [specifik ålderskategori] stod för den största andelen av köpen. Detta tyder på att marknadsföringsstrategier som anpassas efter kön och ålder kan vara mer effektiva.

3. De mest värdefulla kunderna: De 10 kunder som spenderade mest identifierades, vilket kan vara användbart för kundlojalitetsprogram. På motsvarande sätt identifierades de 10 kunder med lägst köpsumma, vilket kan ge insikter för att öka deras engagemang och köpbeteende.

## Diskussion

Resultaten visar att dataanalys och visualisering kan ge värdefulla insikter som direkt kan påverka affärsstrategier. Genom att identifiera de produktkategorier som säljer bäst kan företag optimera sitt lager och marknadsföring för att maximera intäkterna. Dessutom visar demografiska analyser hur viktigt det är att anpassa marknadsföringsstrategier efter specifika kundgrupper, vilket kan bidra till ökad kundnöjdhet och lojalitet.

Det är också värt att notera att de mest värdefulla kunderna har en oproportionerligt stor inverkan på försäljningen. Detta tyder på att det kan vara mer kostnadseffektivt att fokusera på kundlojalitet och retention för dessa kunder, snarare än att endast fokusera på att öka kundbasen.

Begränsningar: Även om analysen gav tydliga insikter, bör det nämnas att datan är begränsad till den tillgängliga e-handelsinformationen och kan vara föremål för förändringar beroende på säsongsmönster eller externa faktorer. Framtida analyser skulle kunna dra nytta av ytterligare variabler, såsom geografisk plats och tid på året, för en ännu mer detaljerad bild.

Sammantaget visar analysen att en datadriven strategi är avgörande för att förstå och förbättra produkt- och kundhantering i en alltmer konkurrensutsatt e-handelsmiljö.

# Slutsatser

Här besvarar du bl.a. frågeställningarna.

Denna analys har gett en tydlig bild av e-handelns försäljningsmönster och besvarar frågeställningarna:

1. Vilka produktkategorier har högst och lägst försäljning?

Analysen visar att vissa produktkategorier säljer betydligt bättre än andra. De mest sålda kategorierna bör prioriteras för marknadsföring och lagerinvesteringar, medan de minst sålda kan behöva justeras eller marknadsföras annorlunda.

2. Hur påverkar kunddemografi köpbeteendet?

Resultaten visar att kön och ålder påverkar köpbeteendet. Specifika grupper, som "vuxna" och ett visst kön, har högre köpsumma. Detta tyder på att marknadsföringsstrategier som anpassas efter dessa grupper kan öka försäljningen.

Sammanfattningsvis visar resultaten hur data kan användas för att anpassa produktutbud och marknadsföring för att stärka försäljningen och kundlojaliteten.

# Självutvärdering

1. Utmaningar du haft under arbetet samt hur du hanterat dem.   
   Under arbetets gång mötte jag flera utmaningar. En av de största utmaningarna var att hantera och bearbeta den stora mängden data från Kaggle-databasen. För att lösa detta använde jag Python-biblioteket Pandas, vilket hjälpte mig att effektivt sortera, filtrera och analysera datan. Dessutom var det en utmaning att integrera interaktiva visualiseringar med Dash och Bokeh, särskilt när det gällde att skapa en smidig och användarvänlig upplevelse. Genom att följa dokumentationen och arbeta iterativt enligt agila principer kunde jag gradvis förbättra dessa delar.
2. Vilket betyg du anser att du skall ha och varför.   
   Jag anser att jag förtjänar ett högt betyg för detta projekt eftersom jag lyckades övervinna flera tekniska utmaningar och skapa ett robust, datadrivet analysverktyg för e-handel. Jag använde en mängd olika tekniker och verktyg (som Pandas, SQLite, Dash och Bokeh) för att genomföra analysen och presentera resultaten på ett tydligt sätt.

Jag arbetade också självständigt och följde agila principer, vilket hjälpte mig att hålla mig flexibel och anpassa mig till de problem som uppstod. Sammantaget känner jag att jag har levererat ett arbete som är väl genomfört och som uppfyller projektets mål på ett framgångsrikt sätt.

1. Något du vill lyfta fram till Antonio?

Nej, tack.

# Appendix A

Här inkluderas relevanta kodavsnitt som använts i projektet för att extrahera, bearbeta och analysera data. Nedan följer ett exempel på hur data lästes in från en ZIP-fil och bearbetades:

import pandas as pd

import zipfile

# Definiera sökvägen till ZIP-filen

zip\_file\_path = r'C:\Users\mayas\Downloads\archive.zip'

# Extrahera ZIP-filen

with zipfile.ZipFile(zip\_file\_path, 'r') as zip\_ref:

zip\_ref.extractall('extracted\_data')

# Läs in CSV-filen

csv\_file\_path = 'extracted\_data/ecommerce\_customer\_data\_large.csv'

df = pd.read\_csv(csv\_file\_path)

# Visa de första raderna

print(df.head())

**A.2 SQL-frågor**

Ett exempel på en SQL-fråga som används för att analysera försäljning per produktkategori:

SELECT [Product Category] AS Product\_Category, SUM([Total Purchase Amount]) AS Total\_Sales

FROM customers

GROUP BY [Product Category]

ORDER BY Total\_Sales DESC

LIMIT 10;

**A.3 Diagram och Visualiseringar**

Försäljning per kön

Följande cirkeldiagram visar fördelningen av försäljning mellan könen:

Försäljning per ålderskategori

Följande diagram visar hur försäljningen är fördelad över olika ålderskategorier:

Appendix A tillhandahåller alltså nödvändig teknisk information och visuell dokumentation som kompletterar huvudrapporten.

# Källförteckning

1. Kaggle. (n.d.). *E-commerce Customer Data For Behavior Analysis*. Hämtad från <https://www.kaggle.com>

<https://www.kaggle.com/datasets/shriyashjagtap/e-commerce-customer-for-behavior-analysis>

*(Används som huvuddataset för analys av kundbeteende och produktförsäljning.)*

1. Agile Alliance. (2001). *Principer bakom det Agila Manifestet*. Hämtad från <https://agilemanifesto.org/principles.html>  
   *(Används för att strukturera arbetet enligt agila principer i projektets utvecklingsprocess.)*
2. Python Software Foundation. (n.d.). *pandas Dokumentation*. Hämtad från <https://pandas.pydata.org/docs/>

*(Används för att bearbeta och analysera data från CSV-filen med hjälp av Pandas biblioteket.)*

1. Dash Plotly. (n.d.). *Dash Användarguide och Dokumentation*. Hämtad från <https://dash.plotly.com/>

*(Används för att skapa interaktiva webbapplikationer och visualiseringar av försäljningsdata.)*

1. Bokeh Dokumentation. (n.d.). *Bokeh Användarguide*. Hämtad från <https://docs.bokeh.org/en/latest/>

*(Används för att skapa interaktiva diagram och visualiseringar.)*

1. SQLite. (n.d.). *SQLite Dokumentation*. Hämtad från: <https://sqlite.org/docs.html>

*(Används för att skapa och hantera databasen som lagrar analysdata och för att utföra SQL-frågor.)*