

Analys av Supermarket försäljning



ECUTBILDNING

Natalie Dobrovolska

EC Utbildning

Abstract

This report analyzes supermarket sales data to uncover trends in product performance, seasonality, and regional differences. Key findings show that Snacks and Eggs, Meat & Fish are top-selling categories, with sales peaking in November and December. The West and East regions lead in both sales and profit. Additionally, there is a positive correlation between sales and profit, suggesting that higher sales generally boost profitability. These insights can guide inventory management and targeted promotions.

Innehållsförteckning

Abstract	2
1 Inledning	1
1.1 Syfte och Frågeställningar	1
2 Metod	2
2.1 Datainsamling	2
2.2 Agil arbetsmetodik	2
3 Resultat och Diskussion	3
3.1 Dataintroduktion och preliminär bearbetning	3
3.2 Statistisk sammanfattning av nyckelvariabler	3
3.3 Tidskonvertering och utökning av datumkolumner	4
3.4 Produktkategorier och subkategorier	4
3.5 Försäljning (antal sålda) per Produktkategori och Underkategori	4
3.6 Försäljningstrender (antal sålda) över Tid	5
3.6.1 Försäljning (antal sålda) på årsbasis.....	6
3.6.2 Försäljning (antal sålda) på månadsbasis.....	6
3.6.3 Försäljning (antal sålda) på veckodagsbasis.....	6
3.6.4 Dagliga försäljningstrender (antal sålda)	7
3.7 Försäljning och vinst per kategori och subkategori.....	7
3.8 Månadsvis och årlig försäljningstrend.....	8
3.9 Vinst per år, månad och dag.....	8
3.10 Samband mellan Försäljning och Vinst.....	9
3.11 Samband mellan Rabatt och Vinst.....	9
3.12 Försäljning per Region och Kategori.....	10
3.12.1 Antal Sålda artiklar per Region	10
3.12.2 Försäljning (omsättning) per Region och Kategori.....	10
3.12.3 Antal sålda artiklar per Region och Underkategori.....	10
3.12.4 Vinst per Region.....	11
3.12.5 Vinst per Region och Kategori.....	11
4 Slutsatser	12
5 Självutvärdering.....	13

1 Inledning

Idag är det viktigt för företag inom detaljhandeln att använda data för att fatta bra beslut. Genom att analysera data från en stormarknad kan man få värdefull kunskap om vad kunderna köper, vilka produkter som säljer bäst och hur efterfrågan varierar under året. Detta hjälper företaget att planera sitt lager, göra effektiva marknadsföringskampanjer och förbättra sin lönsamhet.

Syftet med denna rapport är att undersöka försäljningsdata från en stormarknad för att identifiera viktiga mönster och insikter. Rapporten fokuserar på att analysera olika produktkategorier, försäljningsvolym, vinstmarginaler och regionala skillnader, vilket kan hjälpa företaget att fatta bättre strategiska beslut.

Det har även gjorts en rapport i Power Bi som möjliggör kartläggning av alla resultat för slutanvändaren.

1.1 Syfte och Frågeställningar

Syftet med denna rapport är att analysera stormarknadens försäljningsdata för att förstå trender inom försäljning och lönsamhet. Följande frågeställningar kommer att besvaras:

1. Vilka produktkategorier och underkategorier säljer mest och bidrar mest till vinsten?
2. Hur varierar försäljning och lönsamhet mellan olika regioner och säsonger, och vilka mönster kan användas för att fatta bättre beslut?

2 Metod

2.1 Datainsamling

Datan som används i denna rapport är hämtad från Kaggle, en plattform för öppna datamängder och maskininlärningsprojekt. Datasetet innehåller försäljningsdata från en stormarknad med information om produktkategorier, försäljning, vinst, rabatter, samt information om regioner och datum för varje köp. Inköp som analyserats har gjorts av kunder som bor i delstaten Tamil Nadu, Indien.

Detta dataset valdes eftersom det ger en bred insikt i stormarknadsförsäljning och möjliggör analyser av olika aspekter, sådana som produktprestation, säsongsvariationer och regionala skillnader.

2.2 Agil arbetsmetodik

Vi arbetade agilt genom att följa Scrum-metoden, vilket hjälpte oss att ha en flexibel och strukturerad arbetsprocess.

Som första steg diskuterade vi olika idéer och lade en plan för projektet. Vi bestämde oss för att ha korta möten två gånger i veckan för att följa upp hur arbetet gick, vilket gjorde det lätt att snabbt hantera problem och göra ändringar vid behov. Vi hade också ett längre möte där vi presenterade våra framsteg och fick feedback, det hjälpte oss att förbättra resultatet.

Genom att dela upp arbetet i mindre delar kunde vi leverera resultat snabbt. Den här strukturen gjorde att vi kunde justera planen när nya idéer dök upp. Vi kommunicerade tydligt, både via möten och även skriftligt, genom en kort sammanfattning av mötet, så att alla kunde följa projektet även om någon missade ett möte. Denna arbetsmetod gjorde att vi kunde samarbeta bra och leverera ett slutresultat som vi önskat.

3 Resultat och Diskussion

Denna rapport syftar till att analysera försäljningsdata för en stormarknad med målet att identifiera insikter som kan vara värdefulla för beställaren av analysen. Datasetet omfattar totalt 9 994 beställningar, där information om kundnamn, produktkategori, subkategori, stad, region, försäljningsvärde, rabatt, vinst och orderdatum ingår.

3.1 Dataintroduktion och preliminär bearbetning

Datasetet har följande kolumner: Customer Name, Category, Sub Category, City, Order Date, Region, Sales, Discount, Profit, och State.

Order ID togs bort från datasetet då det ansågs överflödigt för analysen.

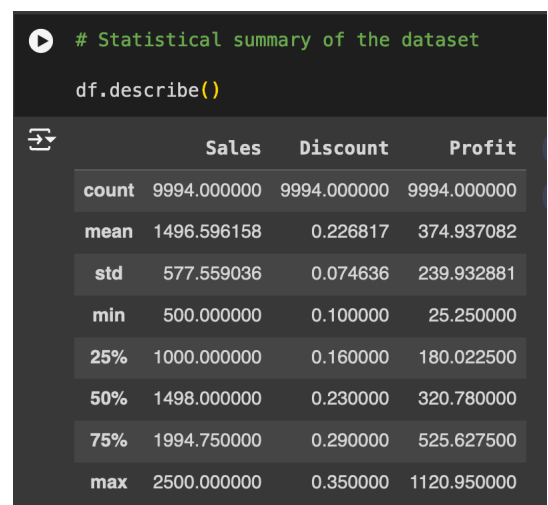
Inga saknade värden förekommer, vilket innebär att datakvaliteten är tillfredsställande.

Orderdatumet (Order Date) har inkonsekventa format, vilket kommer att hanteras i senare steg för att säkerställa att tidsanalyser är möjliga.

3.2 Statistisk sammanfattning av nyckelvariabler

En beskrivande analys utfördes för att undersöka spridningen av försäljning (Sales), rabatt (Discount), och vinst (Profit):

- Medelvärdet för försäljning ligger på ca 1 497 med ett spann mellan 500 och 2 500, vilket visar på en bred variation i beställningsvärde. (Figur 1)
- Rabatten varierar från 10 % till 35 %, där 23 % är medelvärdet, vilket tyder på att rabatter används i olika omfattning beroende på produkt och beställning.
- Vinstvärdet har en stor variation, med en medelvinst på 374,9, men vissa beställningar genererar mycket högre vinster (upp till 1 121).



```
# Statistical summary of the dataset
df.describe()
```

	Sales	Discount	Profit
count	9994.000000	9994.000000	9994.000000
mean	1496.596158	0.226817	374.937082
std	577.559036	0.074636	239.932881
min	500.000000	0.100000	25.250000
25%	1000.000000	0.160000	180.022500
50%	1498.000000	0.230000	320.780000
75%	1994.750000	0.290000	525.627500
max	2500.000000	0.350000	1120.950000

Figur 1: Statistisk sammanfattning av nyckelvariabler

3.3 Tidskonvertering och utökning av datumkolumner

Efter en första granskning av datasetet upptäcktes att Order Date har en blandning av datumformat, vilket kräver att data standardiseras för vidare analys.

För att möjliggöra tidsbaserade analyser har kolumnen Order Date konverterats till en standardiserad Pandas DateTime-format. Vi extraherade även ytterligare tidskomponenter, såsom:

- År (Year): För att se trender i försäljning och vinst på årsbasis.
- Månad (Month): Detta underlättar för säsongsanalys, som kan vara värdefullt för att identifiera hög- och lågsäsonger för olika produktkategorier.
- Datum (Date) och Veckodag (Weekday): För att analysera veckovisa och månatliga variationer i försäljningen.

Därefter togs den ursprungliga kolumnen Order Date bort från datasetet, då den var överflödig.

3.4 Produktkategorier och subkategorier

Analysen av produktkategorier och deras underkategorier visade att datasetet omfattar en rad olika produkter. Totalt finns det 7 huvudkategorier, var och en med flera tillhörande underkategorier.

Nedan följer en sammanfattning av kategorier och deras respektive underkategorierna:

- Bakery: Biscuits, Cakes, Breads & Buns
- Beverages: Health Drinks, Soft Drinks
- Eggs, Meat & Fish: Eggs, Chicken, Mutton, Fish
- Food Grains: Atta & Flour, Organic Staples, Dals & Pulses, Rice
- Fruits & Veggies: Fresh Vegetables, Fresh Fruits, Organic Vegetables, Organic Fruits
- Oil & Masala: Masalas, Edible Oil & Ghee, Spices
- Snacks: Chocolates, Cookies, Noodles

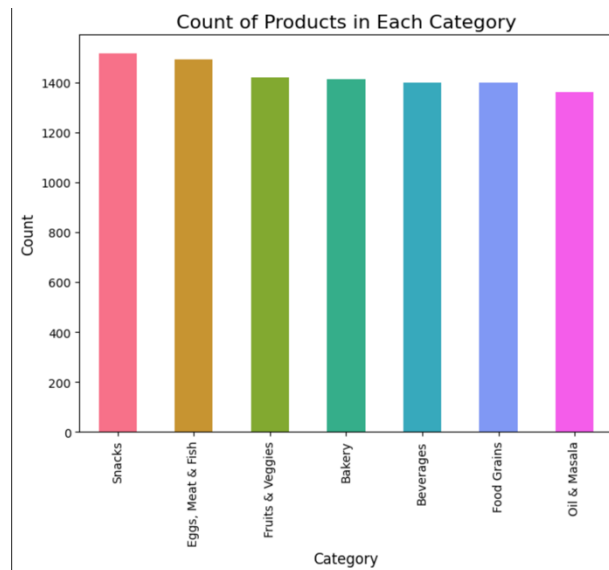
3.5 Försäljning (antal sålda) per Produktkategori och Underkategori

En analys av antalet sålda produkter i varje kategori visar fördelningen av försäljningen inom olika produktkategorier: (Figur 2)

Snacks är den mest sålda kategorin med 1 514 sålda enheter, följt av Eggs, Meat & Fish och Fruits & Veggies. Dessa kategorier står för en betydande andel av den totala försäljningen.

Kategorierna har en jämn fördelning, där varje kategori bidrar relativt likvärdigt till den totala försäljningen, vilket tyder på att stormarknaden har ett varierat och välbalanserat sortiment.

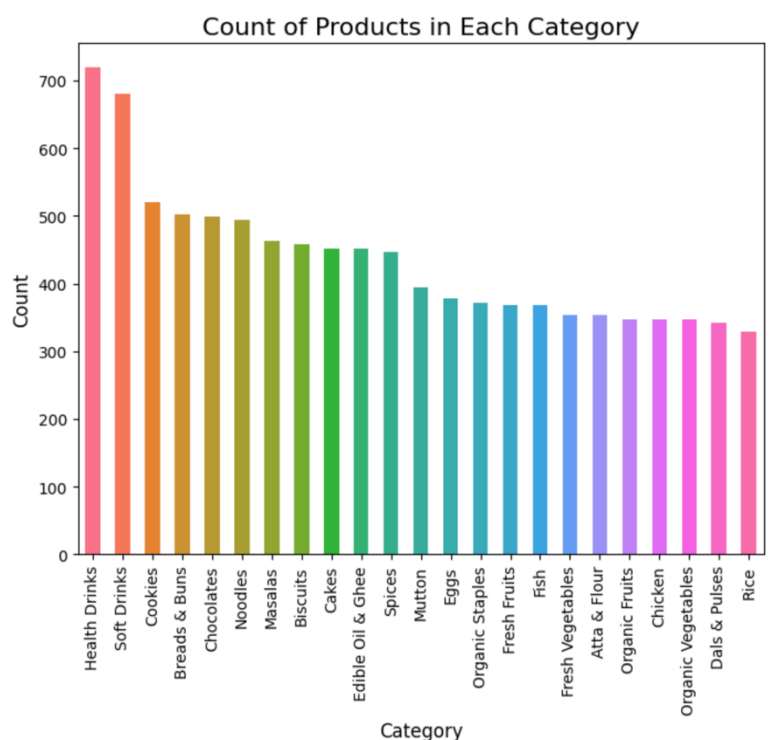
Category	count
Snacks	1514
Eggs, Meat & Fish	1490
Fruits & Veggies	1418
Bakery	1413
Beverages	1400
Food Grains	1398
Oil & Masala	1361



Figur 2: Försäljning (antal sålda) per Kategori

Kollar man närmare på Underkategori försäljningen så kan man se att några av de mest populära underkategorierna är Health Drinks (719 enheter), Soft Drinks (681 enheter), och Cookies (520 enheter). (Figur 3)

Detta indikerar att de spelar en viktig roll i försäljningen och kan vara intressanta att prioritera vid framtida kampanjer.



Figur 3: Försäljning (antal sålda) per Underkategori

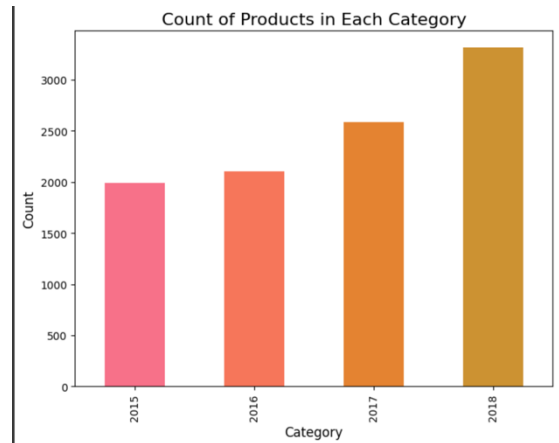
3.6 Försäljningstrender (antal sålda) över Tid

Försäljningen har analyserats på årsbasis, månadsbasis, veckodagsbasis och dag på månaden för att identifiera säsongsmönster.

3.6.1 Försäljning (antal sålda) på årsbasis

Försäljningen har ökat stadigt varje år från 2015 till 2018, med en tydlig topp under 2018 då totalt 3 312 produkter såldes. (Figur 4)

Denna positiva trend över åren kan indikera en ökad efterfrågan eller expansion av verksamheten och dess kundbas.

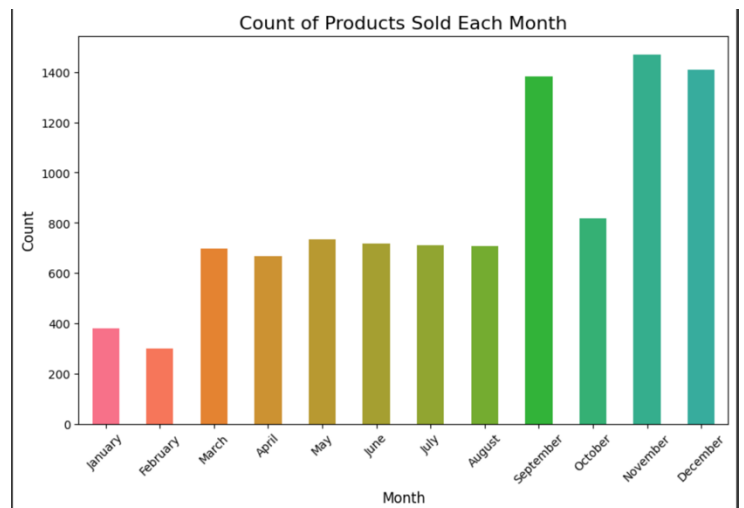


Figur 4: Försäljning (antal sålda) på årsbasis

3.6.2 Försäljning (antal sålda) på månadsbasis

November har den högsta försäljningen (1 471), följt av December (1 408), vilket tyder på att försäljningen ökar markant mot årets slut. Detta kan bero på ökade inköp inför högtider och helger. (Figur 5)

Säsongstrenderna antyder att planering av lager och kampanjer kan vara särskilt viktig under årets sista månader.

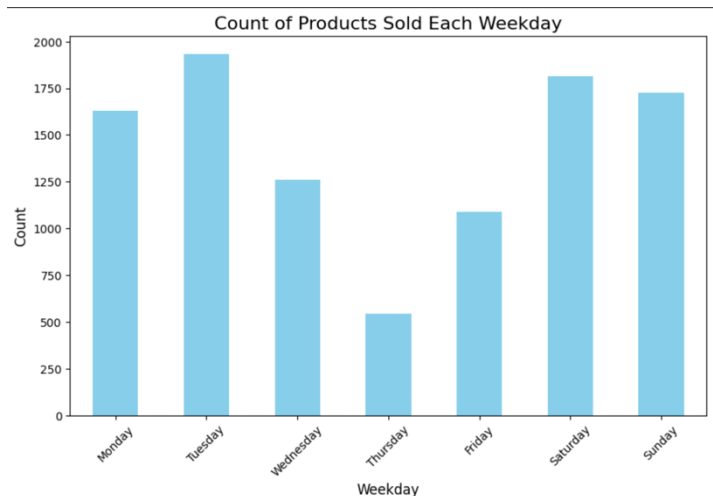


Figur 5: Försäljning (antal sålda) på månadsbasis

3.6.3 Försäljning (antal sålda) på veckodagsbasis

Tisdagar har den högsta försäljningen följt av Lördag och Söndag, vilket indikerar att försäljningen tenderar att vara högre i början av veckan och under helgen. (Figur 6)

Torsdagar har den lägsta försäljningen, vilket kan vara ett bra tillfälle för att optimera genom att lansera kampanjer eller specialerbjudanden.



Figur 6: Försäljning (antal sålda) på veckodagsbasis

3.6.4 Dagliga försäljningstrender (antal sålda)

Försäljningen per datum (1-31) visar variationer, där försäljningen är högst den 20 och 21 i varje månad. Detta kan indikera att försäljningen är högre mot mitten eller slutet av månaden, vilket kan bero på löneutbetalningar eller specifika köpmönster hos kunderna.

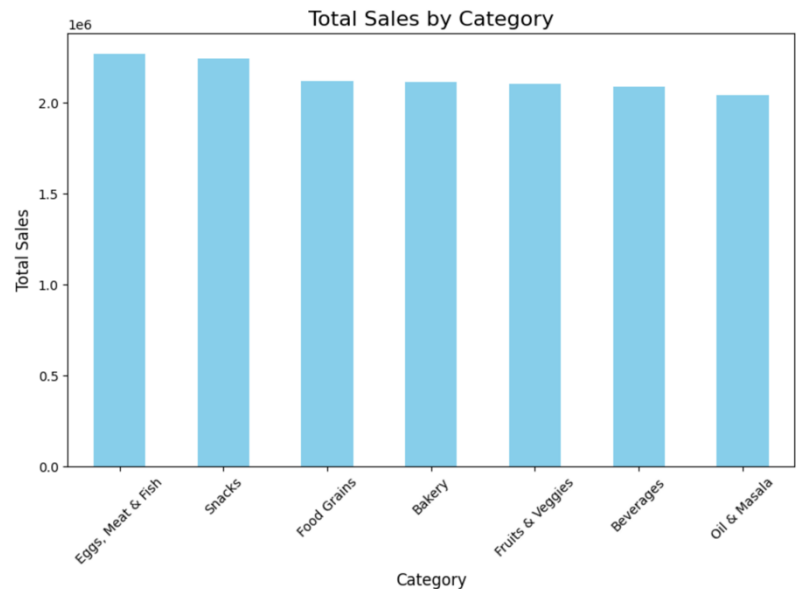
3.7 Försäljning och vinst per kategori och subkategori

Analysen av total försäljning per kategori visar vilka produktområden som genererar mest intäkter:

Kategorierna Eggs, Meat & Fish och Snacks står för de högsta försäljningsvärdena, följt tätt av Food Grains och Bakery. (Figur 7)

Detta överensstämmer med tidigare analys av antal sålda enheter, vilket indikerar att dessa produktområden är centrala för verksamhetens intäkter.

En ytterligare analys av subkategorier bekräftar några av de mest lönsamma produkterna: Health Drinks och Soft Drinks visar höga försäljningsvärden och är de mest lönsamma subkategorierna. Dessa produkter kan potentiellt prioriteras för marknadsföring eller kampanjer för att ytterligare öka intäkterna.



Figur 7: Försäljning per Kategori

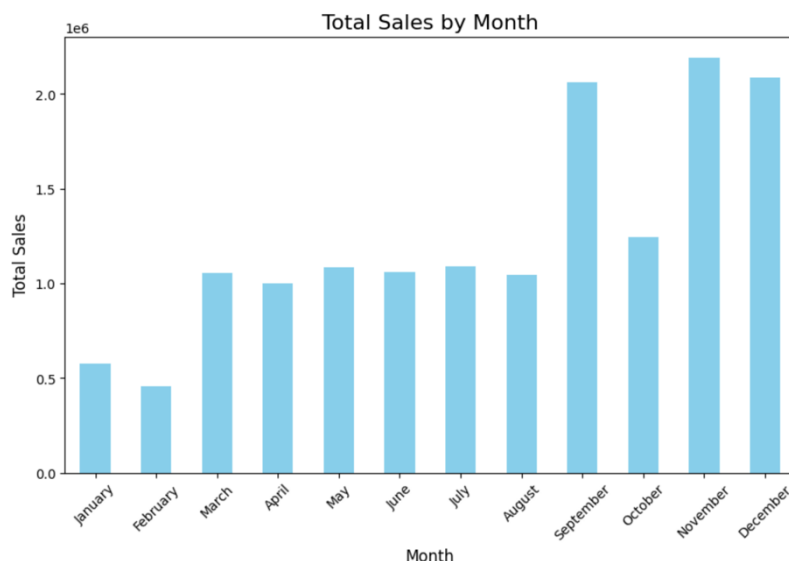
3.8 Månadsvis och årlig försäljningstrend

November, följt av December, har den högsta försäljningen, vilket antyder en ökad efterfrågan mot slutet av året. (Figur 8)

Detta kan bero på högtidshandel som indikerar att dessa månader är viktiga för affärens resultat.

Våren och sommaren, med månader som Mars, Maj, och Juli, visar också höga försäljningsvärden och kan vara en period för att överväga extra marknadsinsatser.

Kollar man på År från 2015 till 2018 så har försäljningen ökat med en tydlig topp under 2018. Detta visar på en stadig tillväxt och potentiellt ökande efterfrågan.



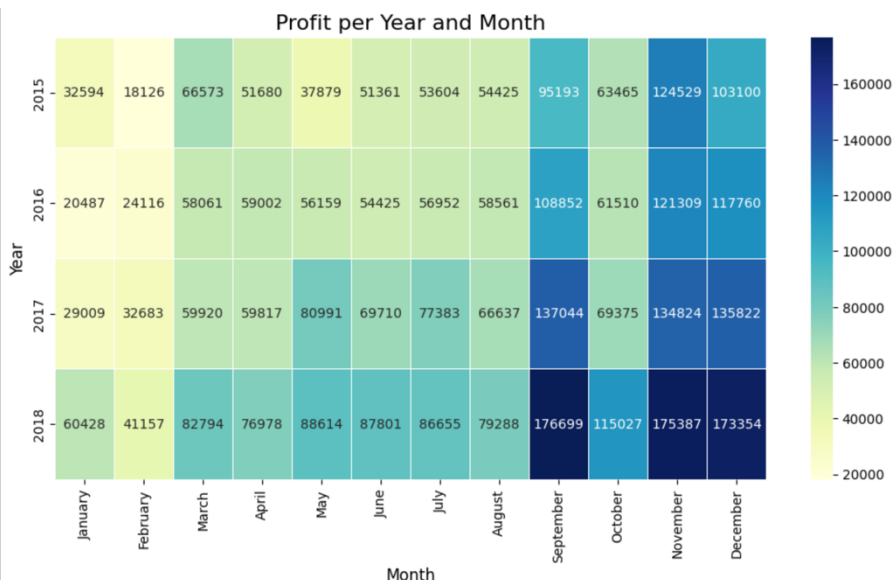
Figur 8: Försäljning per Månad

3.9 Vinst per år, månad och dag

Vinsten har ökat varje år med en tydlig topp 2018. (Figur 9)

Denna tillväxt parallellt med försäljningen visar på en gynnsam utveckling för lönsamheten över tid.

Detaljerad analys av vinst per månad och år visar variationer som tydligt följer säsongstrender. November och September är generellt sett de mest lönsamma månaderna, särskilt under senare år. Detta bekräftar tidigare observationer att slutet av året är en lönsam period för stormarknaden.



Figur 9: Vinst per År och Månad

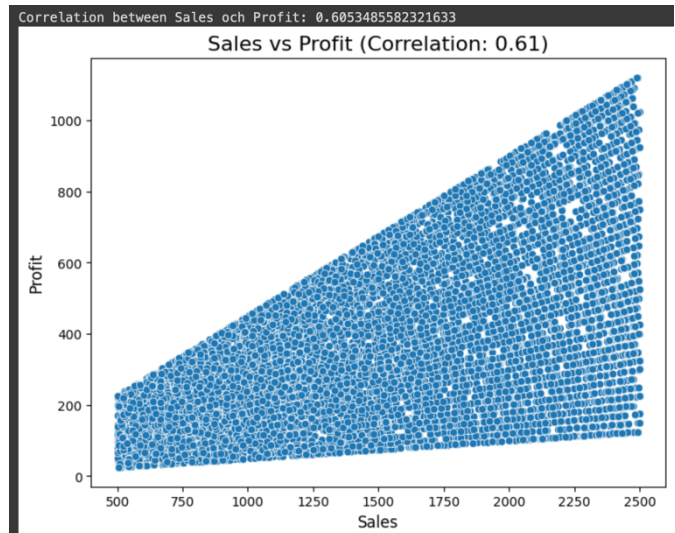
En analys av vinst per datum (1-31) visar att 20 och 21 i varje månad har de högsta vinsterna, med ett tydligt mönster av högre vinst under mitten till slutet av månaden.

Detta kan spegla kundbeteenden kopplade till löneutbetalningar, där konsumenter handlar mer efter att de har fått lön.

3.10 Samband mellan Försäljning och Vinst

En analys av korrelationen mellan försäljningsvärde (Sales) och vinst (Profit) utfördes för att undersöka hur väl försäljningen påverkar lönsamheten. (Figur 10)

Korrelationskoefficienten beräknades till 0.61, vilket indikerar ett positivt samband mellan försäljning och vinst. Detta innebär att när försäljningen ökar, tenderar även vinsten att öka, om än inte i perfekt linje.

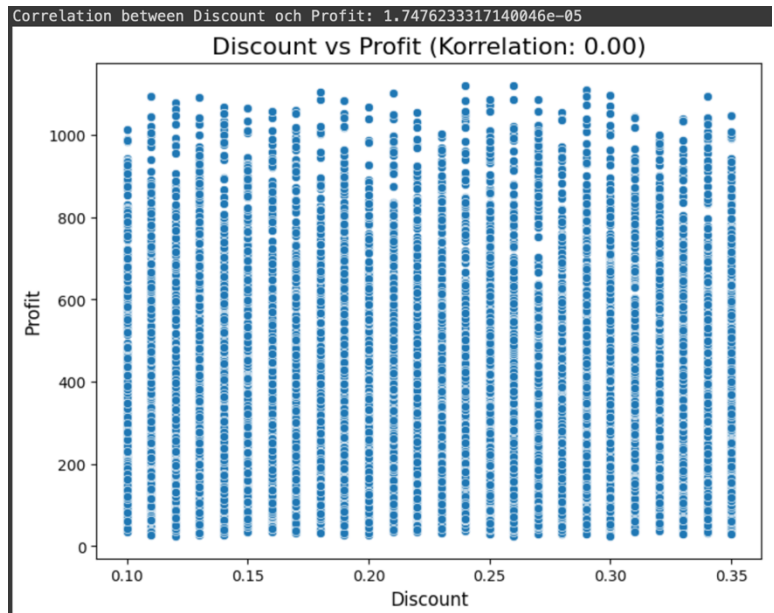


Figur 10: Samband mellan Försäljning och Vinst

3.11 Samband mellan Rabatt och Vinst

En korrelationsanalys mellan rabatt (Discount) och vinst (Profit) genomfördes för att undersöka om rabatter har någon direkt inverkan på lönsamheten. (Figur 11)

Resultatet visar att korrelationen mellan rabatt och vinst är 0.00, vilket indikerar att det inte finns något linjärt samband mellan dessa två variabler i datasetet.



Figur 11: Samband mellan Rabatt och Vinst

3.12 Försäljning per Region och Kategori

En detaljerad analys av försäljningskvantiteten utfördes för att förstå hur många artiklar som sålts i varje region och produktkategori. Detta ger insikt i hur olika produktkategorier presterar över olika geografiska områden.

3.12.1 Antal Sålda artiklar per Region

West och East är de regioner med högst antal försäljningar, medan North endast har en försäljning registrerad. Detta tyder på att verksamheten är mest etablerad i West och East.

Dessa siffror kan vara en indikation på marknadspotential eller etableringsgrad i varje region.

3.12.2 Försäljning (omsättning) per Region och Kategori

I varje region är produktkategorierna relativt balanserade i försäljningsantal, men vissa kategorier, såsom Eggs, Meat & Fish och Snacks, visar högre efterfrågan i specifika regioner.

Regionerna West och East har högre försäljning i nästan alla kategorier, vilket kan bero på större marknadsnärvaro eller bättre anpassat sortiment för lokala preferenser.

3.12.3 Antal sålda artiklar per Region och Underkategori

Analysen på underkategorinivå avslöjar ytterligare detaljer om produktprestanda per region. Några underkategorier med hög försäljning i West inkluderar Soft Drinks (206 sålda enheter), Health Drinks (239 sålda enheter), och Cookies (170 sålda enheter).

Denna information är viktig för att förstå kundpreferenser på detaljnivå och kan användas för att skraddarsy sortimentet efter varje regions efterfrågan.

3.12.4 Vinst per Region

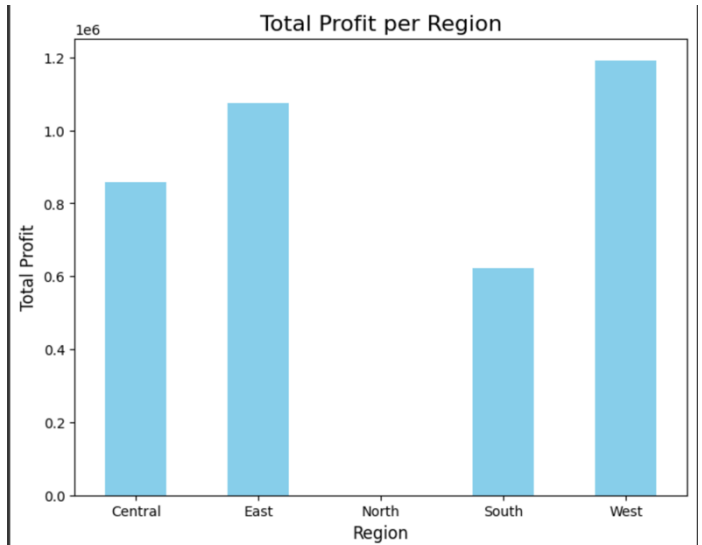
West och East genererar de högsta vinsterna, vilket stämmer överens med de högre försäljningsantalen i dessa regioner. (Figur 12)

North har minimal vinst, vilket är en naturlig följd av den låga försäljningskvantiteten i regionen.

3.12.5 Vinst per Region och Kategori

Snacks och Eggs, Meat & Fish genererar hög vinst i de flesta regioner, med West och East som toppregioner för dessa kategorier.

Den höga vinsten inom Bakery och Beverages i West visar på dessa produkters lönsamhet i regionen, vilket kan vara användbart för att prioritera marknadsföring och kampanjer.



Figur 12: Vints per Region

4 Slutsatser

Denna rapport har analyserat stormarknadens försäljningsdata för att identifiera trender inom försäljning och lönsamhet. Genom analysen framkom att Snacks, Eggs, Meat & Fish, och Food Grains är de mest sålda produktkategorierna, och att dessa även bidrar signifikant till den totala vinsten. Bland subkategorierna är Health Drinks och Soft Drinks särskilt lönsamma, vilket visar att vissa underkategorier har en avgörande roll i företagets intäkter. Dessa produkter kan därför vara prioriterade i framtida lagerhantering och marknadsstrategier.

Analysen visar tydliga variationer i försäljning och lönsamhet mellan regioner och säsonger. West och East framstår som de mest lönsamma regionerna, med högst försäljningsvolym och vinst, vilket gör dem till viktiga områden för framtida satsningar. Säsongsmässigt når försäljningen sin topp i november och december, vilket tyder på en stark efterfrågan under högtider. Denna säsongsvariation kan ligga till grund för att planera högre lagerkapacitet och riktade kampanjer under årets sista månader för att maximera lönsamheten.

För att fortsätta utvecklas bör företaget regelbundet uppdatera och granska sin försäljningsdata för att snabbt upptäcka förändringar i kundbeteenden och anpassa sig efter marknadens behov. Genom att använda de insikter som framkommit i denna rapport kan stormarknaden stärka sin ställning och skapa bättre möjligheter för långsiktig lönsamhet och tillväxt.

5 Självutvärdering

1. Utmaningar du haft under arbetet samt hur du hanterat dem.

Inga utmaningar. Önskar dock att det skulle finnas fler kolumner i datasetet med information om t.ex. vilka varor som kunder köpt, märkningar (eko, fairtrade), varumärken. Även större regionalt område skulle vara mycket intressant att analysera. Detta skulle göra analysen ännu mer spännande.

2. Vilket betyg du anser att du skall ha och varför.

Jag satsade på en VG och tycker att arbetet som gjorts levererar på de mål som var satta för betyget VG.

3. Något du vill lyfta fram till Antonio?

Det var en fin kurs.