

ISTA 1003 - Statistikk

Betaler vi ekstra for LEGO som assosieres med filmer eller spill?

Author:
Group 8:
10073
10066
10060

November 2023

Innholdsliste

Introduksjon.....	3
Data Preprosessering.....	3
Metodikk	3
Resultater.....	4
Diskusjon.....	5
Konklusjon	5
Liste av Figurer	6
Referanser	6
Vedlegg.....	6

Introduksjon

LEGO har spilt en betydelig rolle i å fremme kreativitet og læring blant både barn og voksne. Deres posisjon i bransjen er veletablert, med et godt rykte for både innovasjon og fantasi.

I oppgave 1 tar vi for oss problemstillingen: «Betaler vi ekstra for LEGO som assosieres med filmer eller spill?» Denne problemstillingen er viktig for å forstå LEGOs merkevarestrategi, der assosiasjoner for populære filmer og spill knyttet til LEGO-produkter kan påvirke prisene deres. Årsaken til at gruppen valgte denne problemstillingen er fordi ettersom filmer og spill kan ha større påvirkning på brukerne sine følelser og tanker, så kan popularitet og pris øke proporsjonelt med denne typen markedsføring for LEGO-settene. Vår hypotese er at LEGO-settene som assosieres med film eller spill har en høyere pris enn dem andre LEGO-settene.

Datasettet som er blitt brukt i oppgave 1 er hentet fra studien til Peterson og Ziegler (2021), som gir oss en grundig oversikt over LEGO-sett fra 1 januar 2018 til 11. september 2020. Disse dataene vil være grunnlaget for vår multipl lineær regresjonsanalyse i denne oppgaven.

Data Preprosessering

Datasettet som ble brukt i Peterson og Ziegler (2021), består av 1304 observasjoner og inneholder 7 primære variabler som vi skal sette søkelys på [1]. En sentral variabel i vår analyse er «Merchandise», som er en forklaringsvariabel. Denne variabelen skal identifisere om et LEGO-sett er assosiert med filmer eller spill. I denne analysen fokuserer vi primært på disse 8 variablene, mens de resterende 8 variablene i datasettet ikke er inkludert ettersom vi ikke skulle fokusere på dem. Attributtene og datatypene er listet i Tabell 1, den inneholder også eksemplverdier for å gi en litt enklere forståelse.

Variabelnavn	Datatype	Beskrivelse	Eksemplverdier
Set_Name	Objekt	Navnet på LEGO-settet.	Extra Dots – Series 2
Theme	Objekt	Temakategorien til LEGO-settet	DOTS
Pieces	Float64	Antall deler i LEGO-settet.	109.0
Price	Objekt	Butikkpris for settet (streng som representerer valuta).	\$3.99
Pages	Float64	Antall sider i intruksjonsmanualen (NaN for sett uten manualer).	3.0
Minifigures	Float64	Antall minifigurer inkludert i settet (NaN for sett uten minifigurer).	7.0
Unqie_Pieces	Float64	Antall unike deler i settet.	6.0
Merchandise	String	Kategori på om det er film, spill eller annet.	Movie

Tabell 1: Data attributter som representerer datasettet vi bruker i denne oppgave 1.

For å sikre at datasettet er klar for å analyseres, så har vi utført rensing og behandling av dataene. Dette inkluderer samt håndtering av manglete data, fjerning av uteliggere, opprettelse og transformering av relevante variabler for å tilpasse dem til problemstillingen vår.

Metodikk

I denne oppgaven så har vi brukt multipl lineær regresjon for å undersøke nærmere på sammenhengen mellom prisen på LEGO-sett og forskjellige forklaringsvariabler. Denne metoden er brukt fordi den kan effektivt identifisere og måle styrken av forholdet mellom en avhengig variabel og en eller flere uavhengige variabler.

Som vår responsvariabel har vi valgt pris (Price), siden hovedmålet med denne analysen er å forstå pris forskjellene knyttet til LEGO-produkter assosiert med filmer og spill. Vi har også tatt med flere forklaringsvariabler i modellen, som antall brikker (Pieces) og tema (Theme). I tillegg har vi laget en egen kategorisk variabel kalt "Merchandise" for å avgjøre om et LEGO-sett er en film- eller spillrelatert. Denne variabelen er avgjørende for analysen vår og gjør oss i stand til å forstå prisstrukturen ved å skille mellom LEGO-sett som er assosiert med spill og filmer.

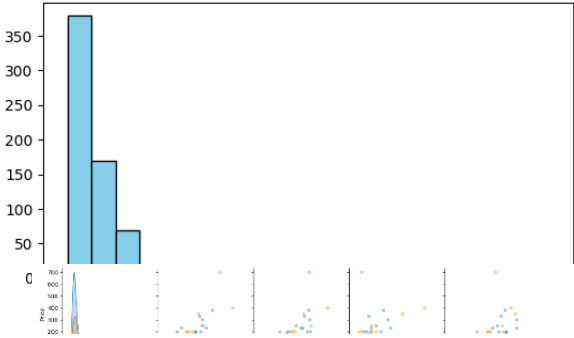
Resultater

Analysen av prisfordelingen av LEGO-sett indikerer en høyre-skjevhet, hvor de fleste settene er priset i det lavere siktet, med en medianpris godt under 100 dollar. Dette kan du se i Figur 1 til høyre, hvor skjevfordelingen er veldig markant.

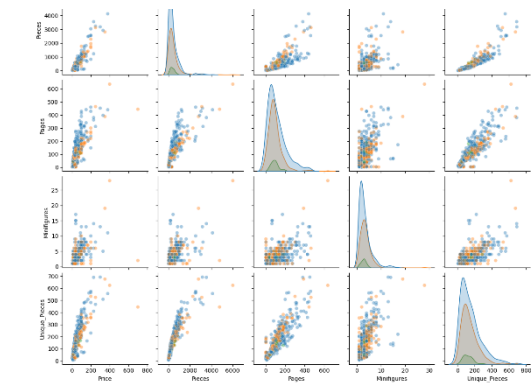
Videre utførte vi en plotanalyse av datasettet vårt, for å visualisere forholdet mellom ulike variablene og deres innvirkning på LEGO-settenes priser. Denne analysen av fordelingen av pris og andre variabler viser en høyre-skjevhet i prisfordelingen, noe som sier at et flertall av LEGO-settet er priset i det lavere segmentet, noe som allerede er vist i Figur 1 opp til høyre.

Plottet viser til en generell trend der prisene øker med antall brikker, sider i intruksjonsmanualen, antall minifigurer og unike brikker. Disse trendene forteller oss at større og mer komplekse sett har en tendens til å være dyrere.

For å støtte opp disse grafene som er gjort, har vi ved bruk av OLS-regresjonsanalyse avdekket betydelig sammenhenger mellom prisen på LEGO-sett og flere variabler. Regresjonsmodellen, som forklarer 84.5% av variansen i pris ($R^2 = 0.845$), indikerer at «Pieces» er sterkt positivt korrelert med pris, med koeffisient på 0.0841 ($p < 0.001$). Videre viser «Merchandise»-variabelen, som klassifiserer settene basert på tilknytting til filmer eller spill, en signifikant effekt på pris, med en koeffisient på 1.7412 ($p = 0.042$). Dette antyder at LEGO-settet assosiert med filmer eller spill har en høyere gjennomsnittspris sammenlignet med andre sett. Dette kan du også legge merke til ved å se på Figur 3 til høyre.



Figur 1: Pris Distribusjon

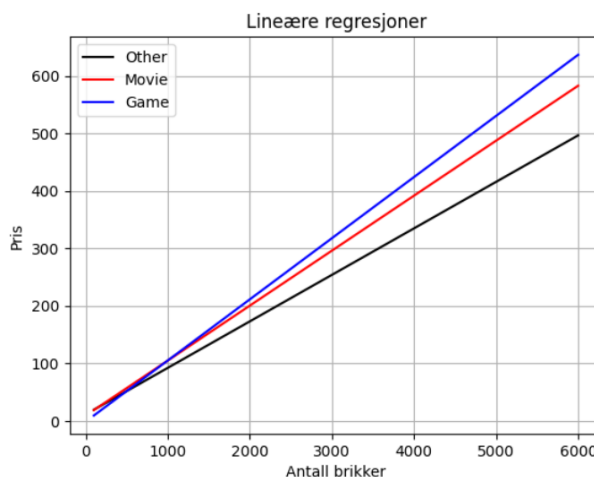


Figur 2: Par Plot Analyse

OLS Regression Results						
Dep. Variable:	Price	R-squared:	0.845			
Model:	OLS	Adj. R-squared:	0.844			
Method:	Least Squares	F-statistic:	893.4			
Date:	Mon, 20 Nov 2023	Prob (F-statistic):	5.49e-264			
Time:	02:36:07	Log-Likelihood:	-2962.8			
No. Observations:	662	AIC:	5938.			
DF Residuals:	657	BIC:	5960.			
DF Model:	4					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	8.0910	1.592	5.081	0.000	4.964	11.218
Merchandise	1.7412	0.856	2.033	0.042	0.059	3.423
Pieces	0.0841	0.003	27.614	0.000	0.078	0.090
Minifigures	0.4936	0.453	1.089	0.277	-0.397	1.384
Unique_Pieces	0.0130	0.017	0.776	0.438	-0.020	0.046
=====						
Omnibus:	671.198	Durbin-Watson:	1.822			
Prob(Omnibus):	0.000	Jarque-Bera (JB)	80433.760			
Skew:	4.235	Prob(JB):	0.00			
Kurtosis:	56.332	Cond. No.	1.40e+03			
=====						
0.844705571471581						
The coefficient for Movie_Game_Associated is: 1.7412312650981532						
Is the coefficient statistically significant? Yes						

Figur 3: OLS Regresjon

Etter å ha observert ett generelt overblikk over prisstrukturen til datasettet og fått vite hvilke kategorier som påvirker hvem. Nede i Figur 4 til høyre kan vi observere tre regresjonslinjer som representerer de ulike kategoriene innen «Merchandise» av LEGO-settet: «Other», «Movie», og «Game». Alle de 3 ulike kategoriene viser en positiv trend, noe som indikerer at prisen øker med antall brikker. Dette er en forventet observasjon, ettersom større sett vanligvis inneholder mer materiale og kan være mer komplekse å produsere. Det uansett en merkbar forskjell mellom stigningstallene til de tre linjene. «Game»-kategorien viser den bratteste økningen, noe som kan tyde på at spillrelaterte LEGO-sett har en høyere pris per prikke sammenlignet med de andre kategoriene. «Movie»-kategoriene ligger også over «Other»-kategorien, noe som støtter hypotesen om at filmrelaterte sett bærer en forskjell i pris. «Other»-kategorien har det laveste stigningstallet, noe som antyder at disse settene har det laveste prisøkningen per ekstra brikke.



Figur 4: Regresjon for Merchandise

Diskusjon

Regresjonsanalysen vi har gjort gir en dypere forståelse av prissettingsmekanismen som LEGO benytter seg av. Den høyre-skjeve prisfordelingen bekrefter at selskapet fokuserer på rimelige produkter som er tilgjengelige for et bredt marked. Samtidig avslører den positive korrelasjonen mellom antall brikker og pris at LEGO også retter seg mot et marked for mer omfattende og detaljerte sett.

Etter undersøkelser med vår egen forklaringsvariabel «Merchandise» har vi merket at settene assosiert med filmer og spill har en høyere pris. Dette kan tyde på at LEGO bruker sin merkevareassosiasjon til å markedsføre og sette høyere verdi på disse typer sett. Brukerne vil gjerne være villig til å betale mer for ulike LEGO-sett som representerer deres favorittspill eller filmer, noe som gir LEGO muligheten til å kapitalisere på populariteten til disse settene.

De ulike stigningstallene fra Figur 3 på «Merchandise»-kategoriene viser hvor LEGO differensierer sine produkter basert på deres merkevareverdi. Siden «Game»-kategorien har høyere stigningstall antyder at spillrelaterte produkter er spesielt mer verdifulle i forhold til stigningstallet til «Movie» og «Other».

Konklusjon

Analysen som er gjort bekrefter vår hypotese og at det er en prisforskjell på LEGO-sett knyttet til filmer og spill. Denne variasjonen innen pris er observert til å ha større utslag hos LEGO-sett som har et større antall brikker, ettersom at i intervallet fra [0, 2000] er det relativt lik pris mellom de ulike kategoriene. Dette er en indikasjon på at LEGO har en vellykket merkevarestrategi. Ved å bruke populariteten til kjente filmer og spill, har LEGO skapt en produktportefølje som ikke bare appellerer til forbrukernes følelser og preferanser, men som også maksimerer deres vilje til å betale mer for produktet.

Denne er en signifikant faktor for hvordan LEGO kan fortsette å vokse og tilpasse seg i et konkurranseutsatt marked. LEGO må balansere å tilby billigere sett for å tiltrekke og beholde et bredt kundesegment, samtidig som de utnytter merketilknyttede sett som kan kreve høyere

priser, for å opprettholde sin markedsposisjon.

Liste av Figurer

Figur 1: Pris Distribusjon	4
Figur 2: Par Plot Analyse	4
Figur 3: OLS Regresjon	4
Figur 4: Regresjon for Merchandise	5

Referanser

[1] Peterson, A. D. and Ziegler, L. (2021). Building a Multiple Linear Regression Model With LEGO Brick Data. Journal of Statistics and Data Science Education, 29(3):297–303

Vedlegg

[1] «Oppgave_1.ipynb», 20 november 2023, «Python kode for analysering av datasettet og problemstillingen», Link til koden: https://github.com/ejhasler/cs-ntnu/blob/main/ntnu-dtu/3-semester-ntnu/ista-1003/project/exam-project-delivery/oppgave_1.ipynb