

Analiza danych o klockach LEGO

AUTHOR

Arkadiusz Urbaniak, Hubert Zawerbny

PUBLISHED

December 9, 2024

1 Wstęp

1.1 Motywacje

LEGO to jedna z najbardziej rozpoznawalnych marek zabawek na świecie, oferująca rozmaite zestawy skierowane do różnych grup wiekowych. Analiza danych związanych z zestawami LEGO udzieli nam odpowiedzi na ciekawe pytania, które mogą zainteresować zarówno producentów, sprzedawców, jak i samych kupujących. Po przeprowadzeniu tej analizy będziemy mogli lepiej zrozumieć zależności pomiędzy kluczowymi cechami zestawów, a także poznać strukturę rynku zabawek.

1.2 Pytania badawcze

Motywnym przewodnikiem naszej pracy jest zbadanie czynników wpływających na cenę klocków LEGO. Aby zrealizować ten cel, postawiliśmy przed sobą następujące pytania badawcze:

- Czy istnieje korelacja pomiędzy liczbą elementów a ceną detaliczną?
- Jak kształtują się ceny zestawów na przestrzeni miesięcy?
- Jak ilość minifigurek w zestawie wpływa na jego cenę?
- Czy rodzaj opakowania jest związany z ceną?
- Jak cena zestawu zmienia się w zależności od serii?
- Która seria klocków jest najdroższa?

1.3 Charakterystyka danych

Dane, z których zamierzamy korzystać, pochodzą ze strony: <https://mostwiedzy.pl/en/open-research-data/data-on-lego-sets-release-dates-and-retail-prices-combined-with-aftermarket-transaction-prices-betwe,10210741381038465-0>. Są to dane zebrane ze 3 źródeł. Ceny detaliczne zestawów LEGO, daty premiery i identyfikatory pobrano z Brickset.com, gdzie można znaleźć numer identyfikacyjny każdego zestawu wydanego przez Lego i jego cenę detaliczną. Aktualny status zestawów pobrano z Lego.com, a ceny detaliczne dla Polski i ceny z transakcji na rynku wtórnym pobrano z promoklocki.pl

Dane posiadają następujące kolumny:

- setID – wewnętrzny numer identyfikacyjny zestawu LEGO Brickset.com,
- number – oficjalny identyfikator zestawu LEGO,
- numberVariant – oficjalny podwariant zestawu LEGO (np. inna minifigurka ukryta w losowej torbie),

- name – oficjalna nazwa zestawu LEGO,
- year – ustalony rok wydania,
- theme – oficjalna nazwa tematu zestawu,
- themeGroup – oficjalna nazwa grupy motywów zestawów (jeśli dostępna),
- subtheme – oficjalna nazwa podtematu zestawu (jeśli jest dostępna),
- category – typ zestawu wewnętrznego brickset.com,
- released – wskazuje, czy zestaw został oficjalnie wydany (1) czy nie (0),
- pieces – ilość części w zestawie,
- minifigs – liczba minifigurek w zestawie,
- ownedBy – liczba użytkowników brickset.com, którzy twierdzą, że są właścicielami zestawu,
- wantedBy – liczba użytkowników brickset.com, którzy twierdzą, że chcą kupić zestaw,
- rating – średnia ocena zestawu według użytkowników brickset.com,
- reviewCount – liczba recenzji zestawów napisanych przez użytkowników brickset.com,
- packagingType – rodzaj opakowania zestawu (jeśli podano),
- availability – wskazuje, czy zestaw był dostępny w sklepach detalicznych, czy tylko na oficjalnej stronie sklepu LEGO,
- instructionsCount – liczba książek z instrukcjami budowania dodanych do zestawu,
- minAge – zalecany przez LEGO minimalny wiek użytkownika zestawu,
- maxAge – zalecany przez LEGO maksymalny wiek użytkownika zestawu (nieokreślony lub 99),
- tags – lista tagów przypisanych do zestawu brickset.com,
- LastUpdated – data i godzina ostatniej aktualizacji danych w brickset.com w formacie ISO 8601,
- urlRetailPriceCheckPLN – adres URL, z którego pobrano cenę detaliczną w PLN,
- US_retailPrice – cena detaliczna w Stanach Zjednoczonych w dolarach amerykańskich,
- US_dateFirstAvailable – data i godzina, kiedy zestaw stał się dostępny w Stanach Zjednoczonych w formacie ISO 8601,
- US_dateLastAvailable – data i godzina, kiedy zestaw przestał być oficjalnie dostępny w Stanach Zjednoczonych w formacie ISO 8601,
- UK_retailPrice – cena detaliczna w Wielkiej Brytanii w GBP,
- UK_dateFirstAvailable – data i godzina, kiedy zestaw stał się dostępny w Wielkiej Brytanii w formacie ISO 8601,

- UK_dateLastAvailable – data i godzina udostępnienia zestawu w Wielkiej Brytanii w formacie ISO 8601,
- CA_retailPrice – cena detaliczna w Kanadzie w dolarach kanadyjskich,
- CA_dateFirstAvailable – data i godzina, kiedy zestaw stał się dostępny w Kanadzie w formacie ISO 8601,
- CA_dateLastAvailable – data i godzina, kiedy zestaw stał się dostępny w Kanadzie w formacie ISO 8601,
- DE_retailPrice – cena detaliczna w Niemczech w EUR,
- DE_dateFirstAvailable – data i godzina udostępnienia zestawu w Niemczech w formacie ISO 8601,
- DE_dateLastAvailable – data i godzina udostępnienia zestawu w Niemczech w formacie ISO 8601,
- PL_retailPrice – cena detaliczna w Polsce w PLN,
- Date – rok i miesiąc, dla którego podawana jest cena PriceMonthPLN,
- priceMonthPLN – cena w PLN odczytana z promoklocki.pl za rok i miesiąc podane w kolumnie Data,
- status – oficjalny status zestawu (jeśli jest dostępny) w sklepie internetowym LEGO,
- urlRetailPriceHistoryPLN – adres URL zawierający zmiany cen detalicznych i rynkowych od dnia premiery zestawu, w PLN

1.4 Wczytanie danych

Ze względu na niekompletność danych ze wcześniejszych okresów zdecydowaliśmy się analizować te, które pochodzą najwcześniej z 2018 roku. Dane w latach 2018 - 2023 posiadają dodatkową zmienną, jaką jest cena na rynku wtórnym na przestrzeni miesięcy.

Ze względu na specyfikę zapisu danych (tzn. cena danego zestawu na przestrzeni miesięcy zapisana jest w osobnych wierszach, przez co dochodzi do powtarzania się danych, w nierównomiernym stopniu, ponieważ nie każdy zestaw posiada identyczną ilość obserwacji) zdecydowaliśmy się w pierwszym kroku, dla każdej wartości "number" usunąć powtarzające się zestawy.

Jednak przed usunięciem danych, które posiadają ten sam numer seryjny zestawu, ale inny miesiąc obserwacji, zapisaliśmy je jako osobne dane, których użyjemy do analizy cen na przestrzeni miesięcy. Następnie ze względu na to, że interesują nas tylko zestawy z oficjalnej dystrybucji, postanowiliśmy także usunąć te oznaczone cyfrą 0 w parametrze "released", który to informuje nas o tym, czy jest to zestaw oficjalnie wydany.

Kolejną nietypową kategorią, którą decydujemy się odrzucić podczas analizy zbiorczej, są zestawy niedostępne do masowej sprzedaży. Dla przeciętnego konsumenta są one trudno dostępne, dlatego nie nadają się do zbiorczej analizy z zestawami szeroko powszechnymi w sprzedaży. Uważamy także, że stanowią one za dużą część całości danych, aby móc analizować je wspólnie z resztą zestawów.

Dlatego, podjęliśmy decyzję na ustawienie dostępności klocków LEGO tylko na te, które nie są w żaden sposób limitowane. Tym samym, do dalszej analizy uzyskaliśmy tylko te klocki, które uważamy za najbardziej klasyczne i dla tych zestawów przeprowadzimy badania.

Po starannym wyselekcjonowaniu danych, na których mamy zamiar opierać nasz raport, przejdźmy do rozważania pojawiających się braków. Aby nie tracić cennych obserwacji, nie usuwamy całych wierszy, w których pojawił się brak danych. Skupiamy się natomiast na tym, by przed każdym sporządzonym przez nas wykresem usunąć tylko te wiersze, w których znajdują się braki danych w interesujących nas kolumnach. Tym samym, przed narysowaniem każdego wykresu odwołujemy się do pierwotnych danych, które nie zostały nadpisane.

1.5 Statystyki opisowe

Zanim przystąpimy do właściwej analizy, skierujmy naszą uwagę w stronę podstawowych statystyk opisowych, których zdefiniowanie pozwoli nam na pełniejsze zrozumienie charakterystyk danych.

Statystyki opisowe dla wybranych zmiennych LEGO

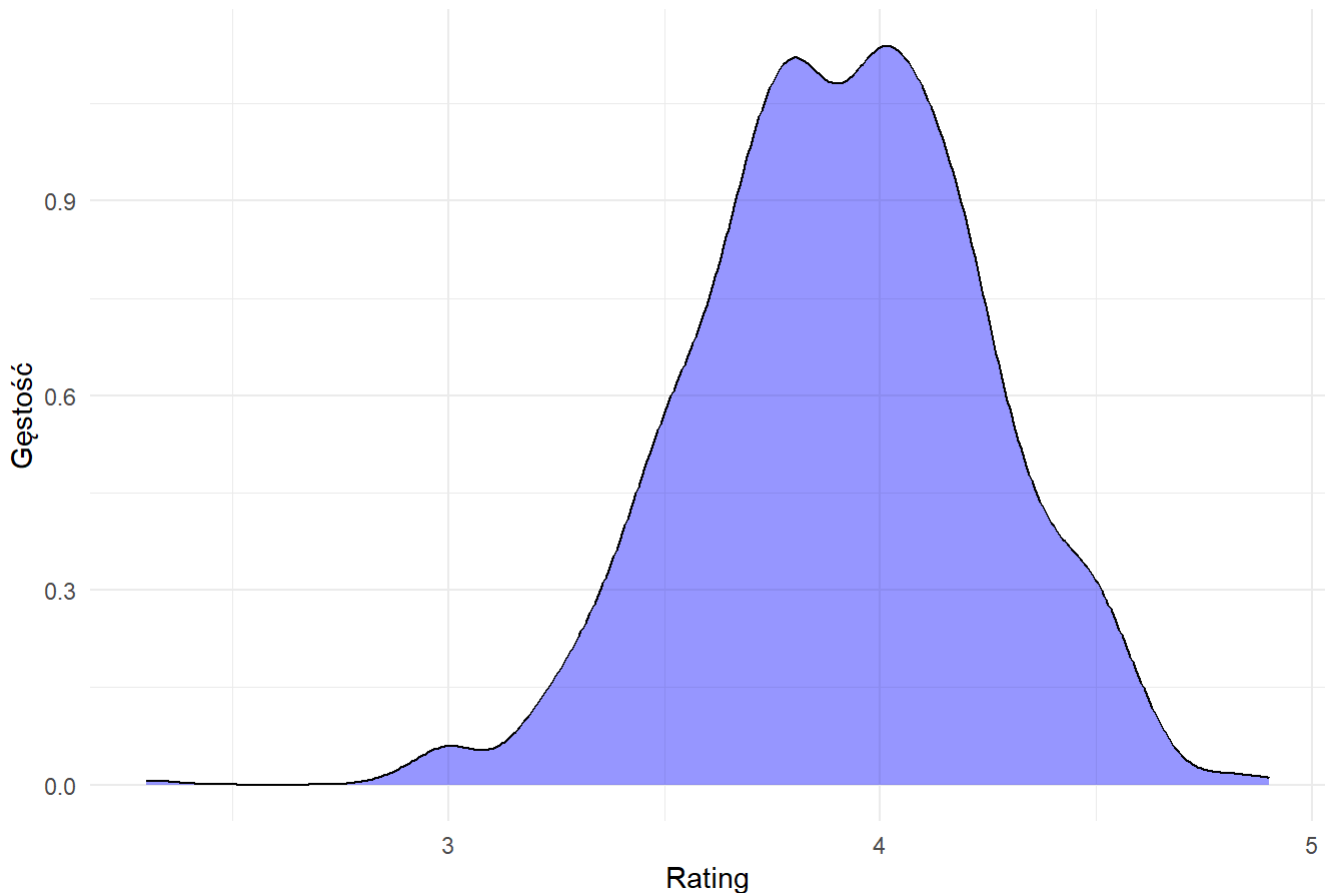
	Średnia	Mediana	Min	Max	SD	CV	Skośność	Kurtoza
pieces	382.45	279.50	5.0	2708.00	340.07	88.92	1.75	4.53
minifigs	3.32	3.00	1.0	20.00	2.27	68.29	1.98	7.05
rating	3.91	3.90	2.3	4.90	0.33	8.57	-0.23	0.32
PL_retailPrice	195.21	139.99	12.0	1199.99	160.32	82.13	1.60	3.11

Liczba elementów oraz **cena detaliczna** mają znacząco wyższą średnią od mediany, co oznacza, że rozkład jest prawostronnie skośny, większość zestawów ma niewielką liczbę klocków/cenę, ale istnieje mała grupa bardzo dużych/drogich zestawów, które podnoszą średnią. Wysokie odchylenie standardowe (*SD*) oraz współczynnik zmienności (*CV*) wskazują na dużą rozpiętość wartości w tych zmiennych, dane są mocno rozproszone. Dodatkowo dane te mają dodatnią asymetrię oraz są ciężkoogonowe.

W przypadku **liczby minifigurek** mamy do czynienia z danymi o mniejszym zróżnicowaniu, aczkolwiek też posiadają one ciężki ogon.

Oceny charakteryzują się małą rozpiętością, a ich rozkład jest prawie symetryczny. Niskie odchylenie standardowe oraz współczynnik zmienności wskazują na minimalną zmienność w ocenach. Rozkład ten jest zbliżony do normalnego.

Wykres gęstości dla zmiennej rating



Patrząc na powstały wykres gęstości dla zmiennej rating, możemy potwierdzić nasze wcześniejsze przypuszczenia, że rozkład ten jest zbliżony do normalnego z tą różnicą, że jest on lekko leptokurtyczny.

Zdecydujmy teraz, czy oceny zestawów rzeczywiście mają rozkład normalny. W tym celu przetestujemy hipotezę zerową, mówiącą, że rozkład jest normalny, przeciwko hipotezie alternatywnej, która twierdzi, że tak nie jest.

Shapiro-Wilk normality test

data: ratings

W = 0.98771, p-value = 4.838e-07

Po wykonaniu testu Shapiro-Wilka na normalność dla zmiennej rating odrzucamy hipotezę zerową i przyjmujemy hipotezę alternatywną - zmienna rating nie pochodzi z populacji o rozkładzie normalnym.

2 Analiza

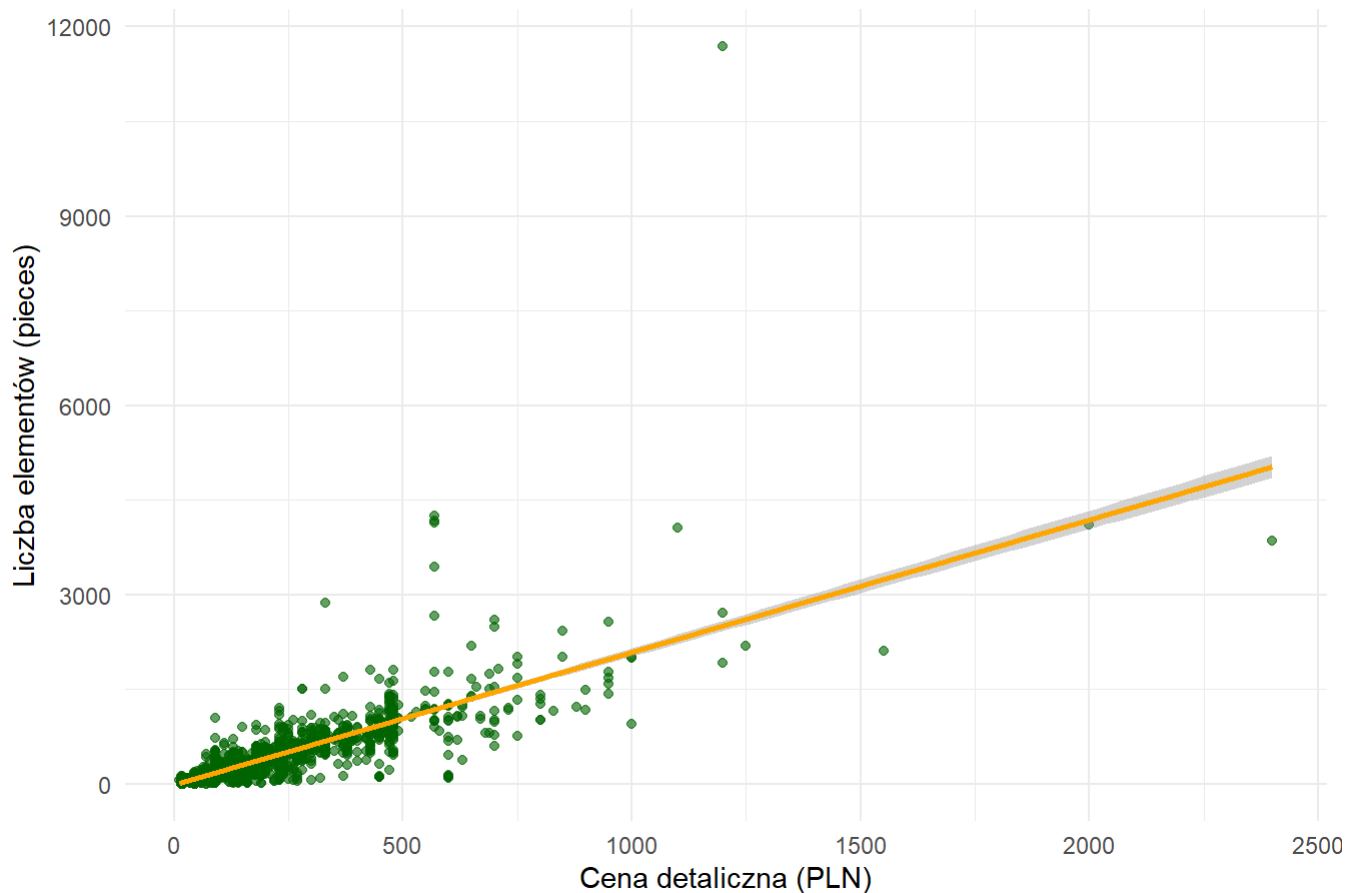
2.1 Analiza zbiorcza

W pierwszej części analizy zajmiemy się analizą zbiorczą na podstawie wszystkich wyselekcjonowanych przez nas danych.

2.1.1 Zależność ceny od ilości elementów

Naszą przygodę z klockami LEGO zaczniemy od sprawdzenia najbardziej oczywistej, ale też najistotniejszej zależności. Będzie to korelacja ilości elementów w zestawie z jego ceną. Zgodnie z intuicją, spodziewamy się sporej zależności między tymi dwoma zmiennymi.

Zależność między ceną a liczbą elementów w zestawach LEGO

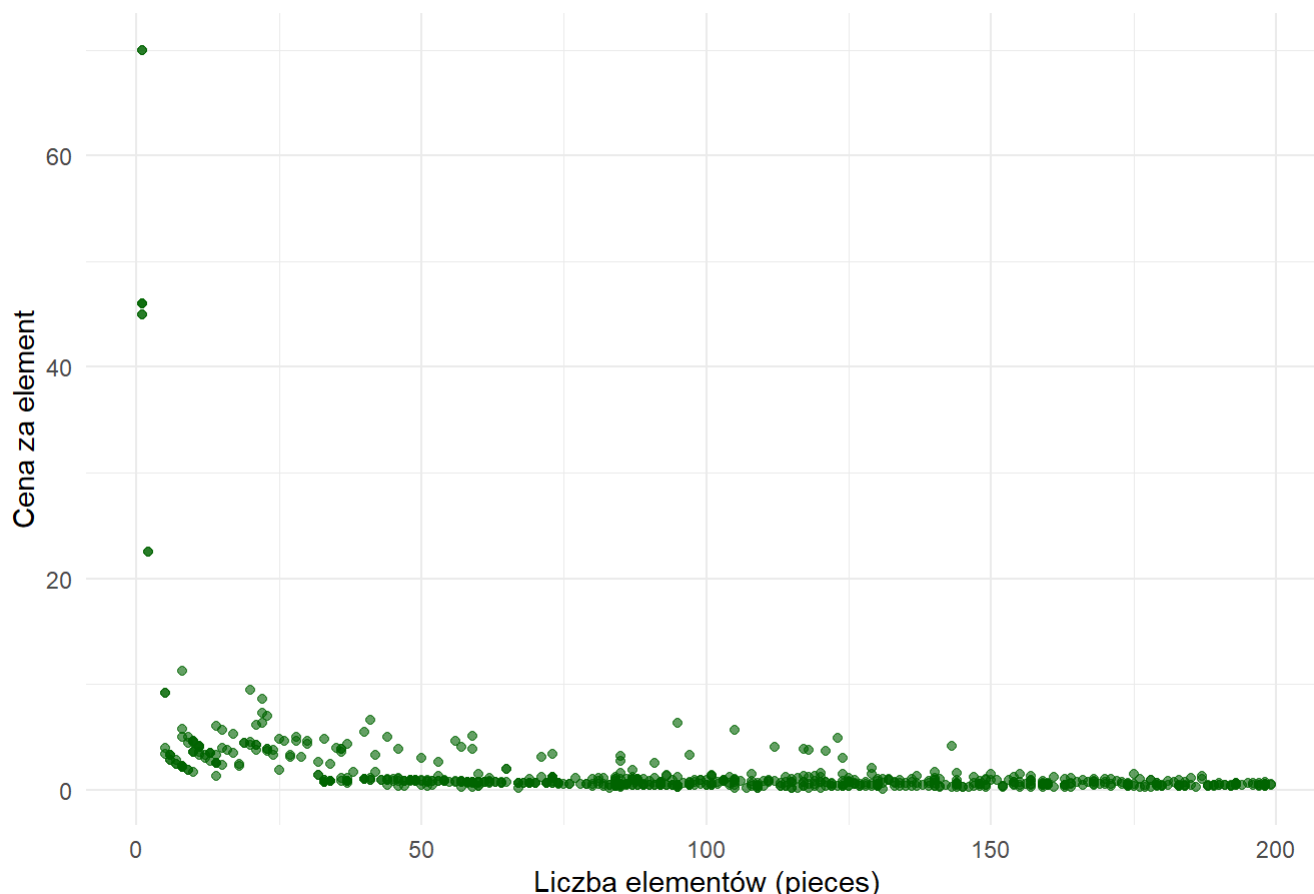


Do stworzenia tego wykresu posłużyliśmy się ceną detaliczną, czyli oficjalną ceną, jaką podała marka LEGO w dniu wydania zestawu. Zgodnie z przypuszczeniami, zaobserwowaliśmy zależność liniową między tymi dwoma zmiennymi. Oznacza to, że istnieje korelacja pomiędzy ilością elementów a ceną, która wynosi około 0,77. Z tego powodu możemy przypuszczać, że wielkość zestawu determinuje jego cenę.

Warto także przyjrzeć się znacząco odstającej obserwacji od głównego trendu. Punktem na wykresie o ilości elementów około 12000 i cenie w okolicach 1200 zł jest zestaw mapy świata, który zbudowany jest z dużej ilości mniejszych od standardowych rozmiarów klocków. Z tego powodu nie zachowuje się on podobnie do reszty zestawów.

Badając zależność ceny od liczby elementów, warto spojrzeć również, jak kształtują się ceny za element w zależności od wielkości zestawów.

Zależność między ceną a liczbą elementów w zestawach LEGO



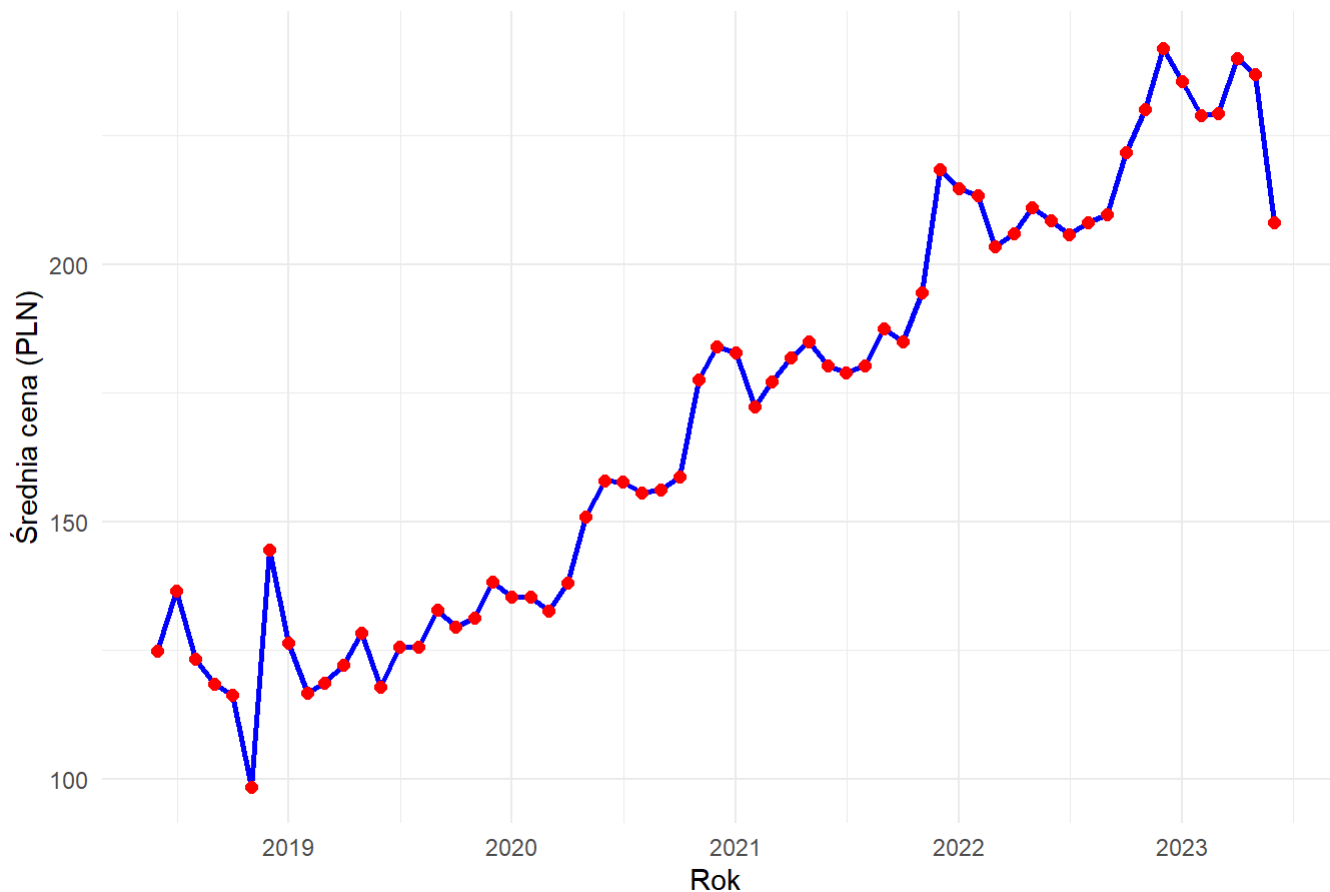
Na wykresie zauważyć możemy, że ceny zestawów z małą ilością klocków posiadają niekorzystny dla konsumenta przelicznik ceny za klocek. Dopiero na poziomie około 50-100 klocków w zestawie cena zaczyna się stabilizować. Małe zestawy posiadają tak wysoką cenę z kilku powodów. Jednym z nich jest to, że są to zazwyczaj zestawy posiadające bardzo duże elementy (np. tory kolejowe LEGO, jednolite płyty służące jako podstawki). Innym powodem może być występowanie zjawiska, które polega na tym, iż kupując więcej sztuk, w tym przypadku klocków, zapłacimy mniej za jedną sztukę (klocek).

2.1.2 Zmiany ceny na przestrzeni czasu

Dane, które pozyskaliśmy, przedstawiają także zmiany ceny na przestrzeni miesięcy w latach 2018 - 2023. Zaczniemy zatem od przeanalizowania, jak zmieniała się średnia cena za zestaw w tym okresie.

Analizowane zestawy pozostają bez zmian, zatem badamy te ogólnodostępne. Warto jednak podkreślić, że teraz będziemy korzystać nie z ceny detalicznej, a z ceny, po której najtaniej jesteśmy w stanie kupić zestaw według strony promoklocki.pl. Zdecydowaliśmy się na ten krok, ponieważ cena detaliczna zestawu w danych, do których mamy dostęp, nie zmienia się w czasie, zatem nie byłibyśmy w stanie sprawdzić, w jakiej cenie ten sam zestaw dostępny jest na rynku.

Średnia cena na przestrzeni miesięcy



Jak możemy zauważyć, tendencja ceny za zestaw na przestrzeni miesięcy jest rosnąca. Warto jednak odnotować, że jest to średnia cena za zestaw, a nie za element. Przyczyn możemy szukać w tym, że do oficjalnej dystrybucji doszły nowe zestawy, posiadające większą ilość klocków, co powoduje wzrost średniej ceny zestawu.

Dlatego, następnym etapem badań, będzie wykres zależności ceny za element od czasu.

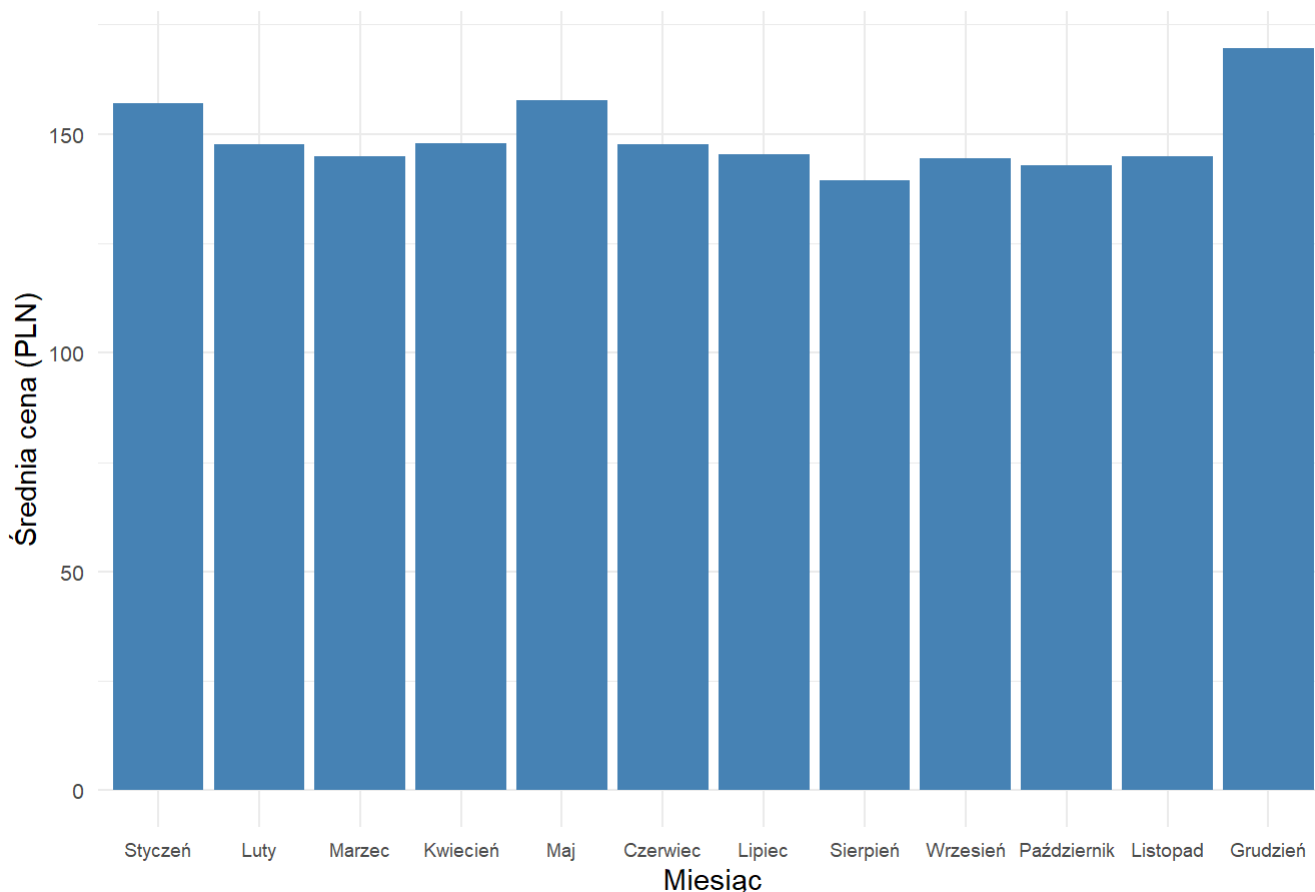
The graph displays the monthly evolution of COVID-19 cases in the Val d'Aosta region from 2019 to 2023. The x-axis represents the year (Rok) and the y-axis represents the number of cases (Nº). The data shows a period of low activity from 2019 through early 2022, followed by a sharp increase in cases starting in early 2022, peaking in late 2022 and early 2023, and then a sharp decline.

Rok	Nº
2019-01-01	10
2019-02-01	10
2019-03-01	5
2019-04-01	2
2019-05-01	1
2019-06-01	15
2019-07-01	10
2019-08-01	5
2019-09-01	5
2019-10-01	2
2019-11-01	5
2019-12-01	2
2020-01-01	1
2020-02-01	5
2020-03-01	8
2020-04-01	5
2020-05-01	5
2020-06-01	8
2020-07-01	7
2020-08-01	9
2020-09-01	7
2020-10-01	5
2020-11-01	7
2020-12-01	15
2021-01-01	20
2021-02-01	18
2021-03-01	18
2021-04-01	18
2021-05-01	19
2021-06-01	18
2021-07-01	15
2021-08-01	12
2021-09-01	15
2021-10-01	18
2021-11-01	20
2021-12-01	25
2022-01-01	35
2022-02-01	45
2022-03-01	60
2022-04-01	80
2022-05-01	100
2022-06-01	120
2022-07-01	150
2022-08-01	180
2022-09-01	160
2022-10-01	155
2022-11-01	155
2022-12-01	145
2023-01-01	120
2023-02-01	150
2023-03-01	170
2023-04-01	190
2023-05-01	210
2023-06-01	190
2023-07-01	170
2023-08-01	150
2023-09-01	120

2.1.3 Jaki miesiąc jest najdroższy?

Jak powszechnie wiadomo, sprzedawcy przez cały rok zasypują nas różnego rodzaju promocjami, próbując tym samym sprzedać konsumentom jak najwięcej produktów. Jednakże warto momentami zatrzymać się i na spokojnie przeanalizować, czy istnieją miesiące w roku, w których to średnio jesteśmy w stanie zakupić produkty taniej. Potencjalnie oferty podczas "Black Friday", a więc ceny w listopadzie powinny okazać się tymi, gdzie zestawy LEGO możemy kupić w naprawdę okazjnych cenach. Jednak czy takie zjawisko rzeczywiście ma miejsce? Z pomocą przychodzą nam zatem dwa wykresy poniżej.

Średnia cena zestawów LEGO w podziale na miesiące



Na zaproponowanym przez nas diagramie słupkowym możemy wskazać 3 miesiące, dla których ceny są największe. Zaczniemy od tych, gdzie odpowiedzi wydają się najbardziej oczywiste. Grudzień jest miesiącem zakupowych szaleństw, podczas których szukamy prezentów pod choinkę dla naszych bliskich. Sprzedawcy doskonale zdają sobie z tego sprawę. Wiedzą, że mogą podnieść cenę produktu, ponieważ jeśli nie zaopatrzyliśmy się w prezent szybciej, będziemy zmuszeni na skorzystanie z ich usług. Skutkuje to zwiększeniem zapotrzebowania, co implikuje wyższe ceny.

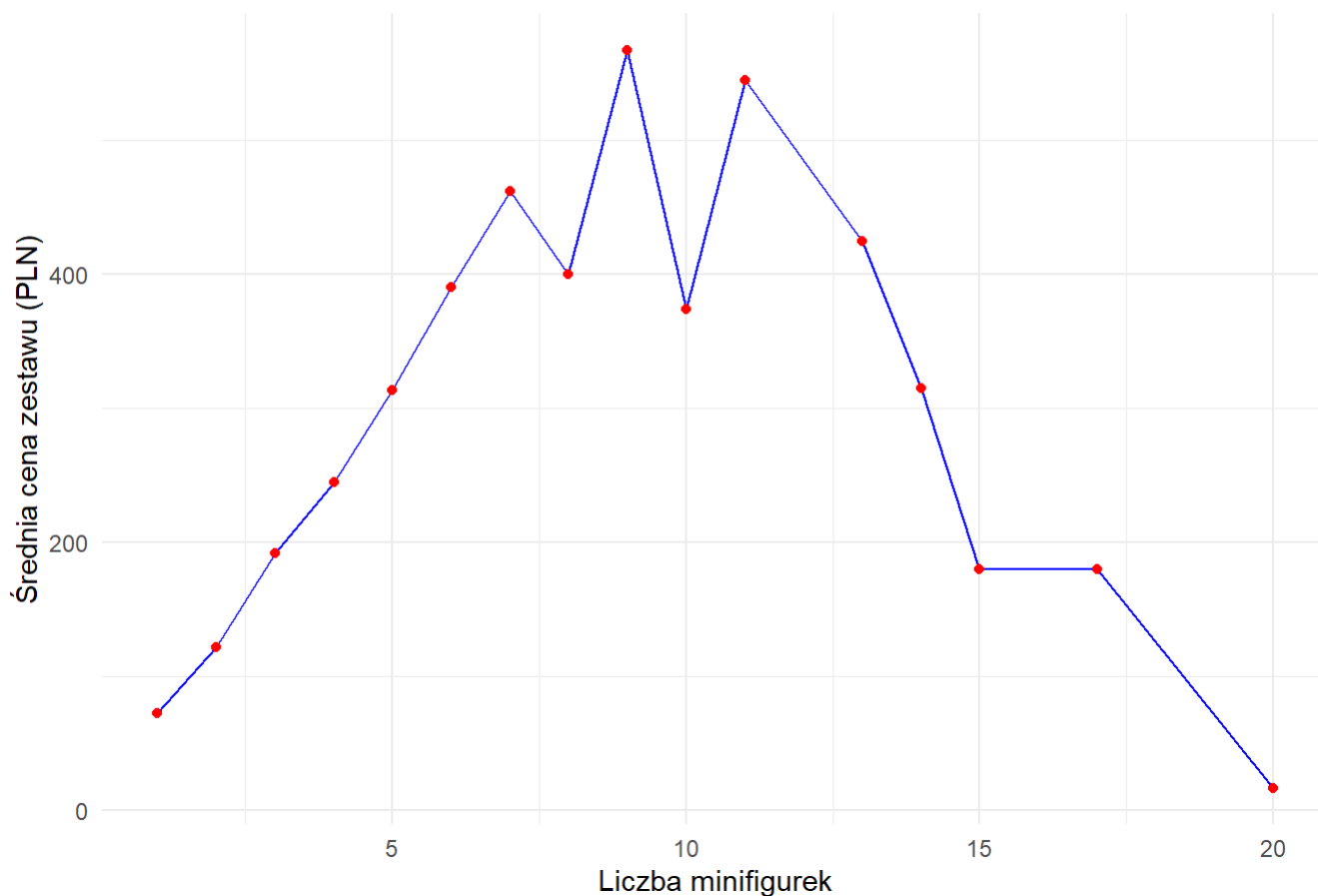
Wysoką cenę zaobserwować możemy także w styczniu, co wydaje się trochę nieintuicyjne. W naszej teorii wynika to z nieunormowanej sytuacji na rynku po świętach Bożego Narodzenia, na półkach sklepowych widnieją jeszcze stare ceny. Nie wiemy niestety, jakiego dnia miesiąca ceny były zbierane. Przypuszczamy, że działa się to pierwszego dnia miesiąca, co potwierdzałoby tę teorię.

Ostatnim z miesięcy, w którym zaobserwować możemy wzrost cen, jest maj. Trudno wnioskować, dlaczego tak się dzieje, ale najbliższą nam teorią, na którą nie mamy jednak żadnego potwierdzenia, jest zbliżający się pierwszego czerwca Dzień Dziecka. Rodzice w tym miesiącu mogą kupować dla swoich pociek więcej zestawów klocków LEGO, niż ma to miejsce w innych miesiącach, a jak wiemy, dynamika cen silnie skorelowana jest z popytem.

2.1.4 Jak ilość minifigurek wpływa na kształtowanie się cen?

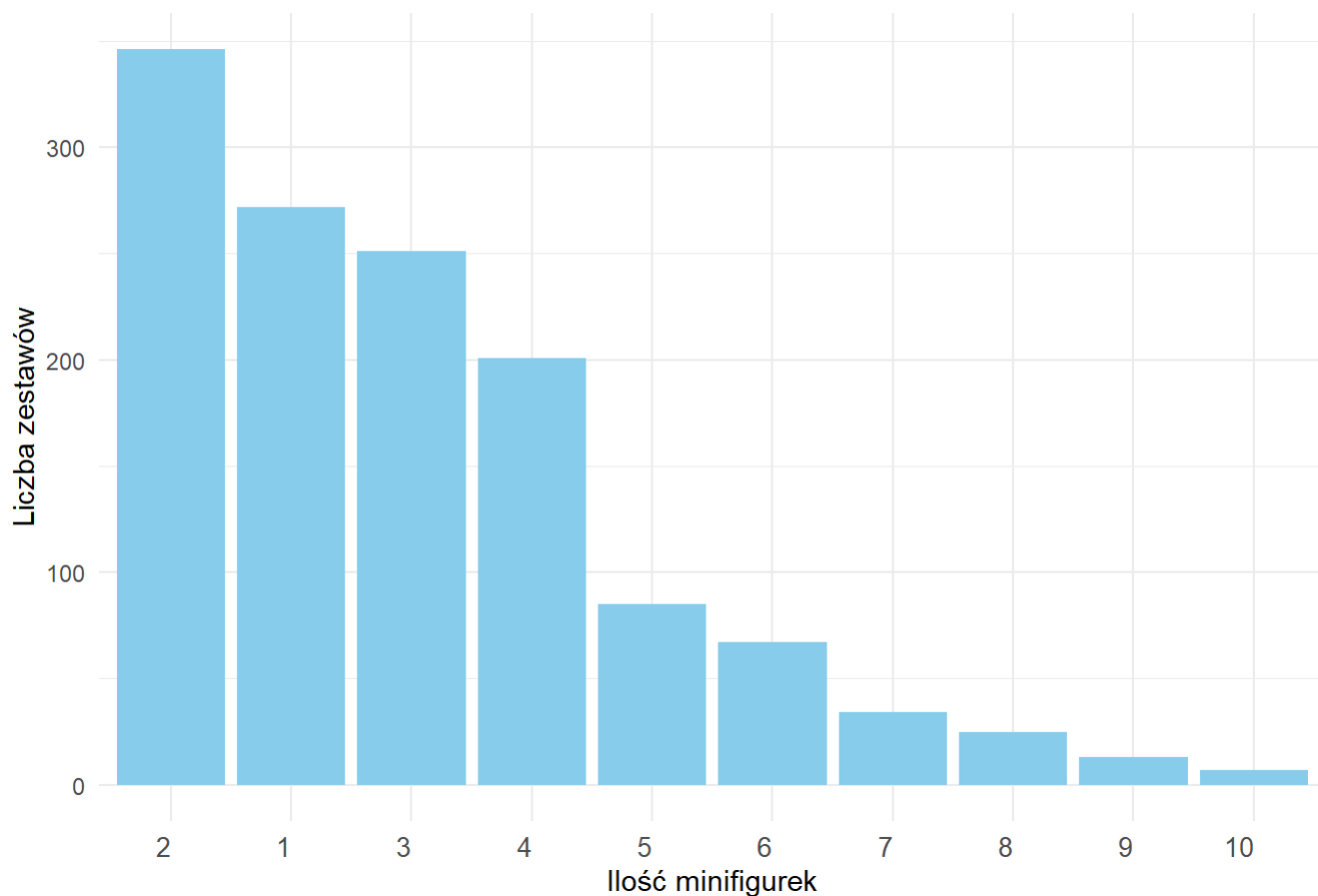
Następnym istotnym czynnikiem, który może mieć wpływ na cenę zestawu, jest ilość jego minifigurek. Oczywiście, potencjalnie warto wziąć pod uwagę, że zestaw, który posiada więcej minifigurek, jest także zestawem większym, posiadającym więcej klocków, co znacząco może odbić się na cenie produktu.

Średnia cena zestawu LEGO w zależności od liczby minifigurek



Z wykresu odczytać można, że początkowo cena jest wprost zależna od ilości minifigurek. Jednakże zauważyć możemy, że na przedziale od około 7 do 12 minifigurek cena zestawu oscyluje wokół 450 złotych. Następnie, cena zaczyna zachowywać się inaczej, co związane może być z kilkoma kwestiami. Pierwsza z nich, dość naturalna, jest taka, że istnieje znacznie mniej zestawów posiadających taką ilość minifigurek, przez co obserwacje odstające znacząco wpływają na wykres. Drugą kwestią, jaką zdołaliśmy ustalić, jest to, że istnieją zestawy, w których minifigurki są mniejsze od tych standardowych. Produkty te, to w większości takie, które przypominają gry planszowe, co w naszej opinii, znacząco odbiega od klasycznych zestawów LEGO.

Liczność 10 najczęściej występującej ilości minifigurek w zestawie

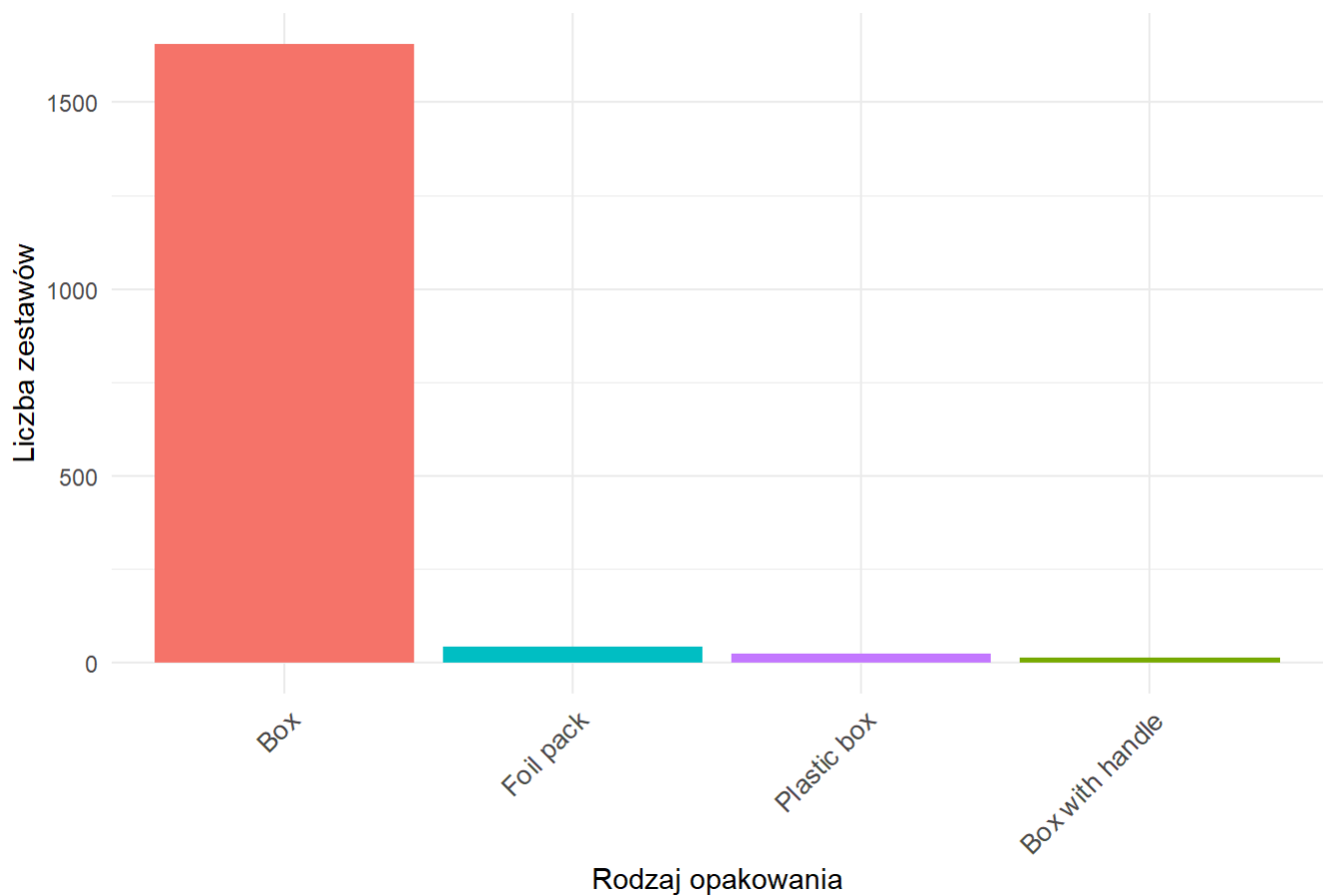


Jak zauważyć można na diagramie licznosci, najczęściej do czynienia mamy z 2 lub 1 minifigurkami w zestawie, a od około 7 nasze dane są podatne na spore odchylenia, ze względu na małą liczbę zestawów posiadających taką ilość minifigurek. Z tego powodu, uznajemy je za mało istotne obserwacje.

2.1.5 Czy po rodzaju opakowania możemy się spodziewać ceny?

Klocki LEGO sprzedawane są w różnych rodzajach opakowań, ale czy na ich podstawie możemy spodziewać się, że cena zestawu będzie inna? Aby odpowiedzieć na to pytanie, zacznijmy od sprawdzenia, jakie są najczęściej występujące rodzaje opakowań klocków na półkach sklepowych.

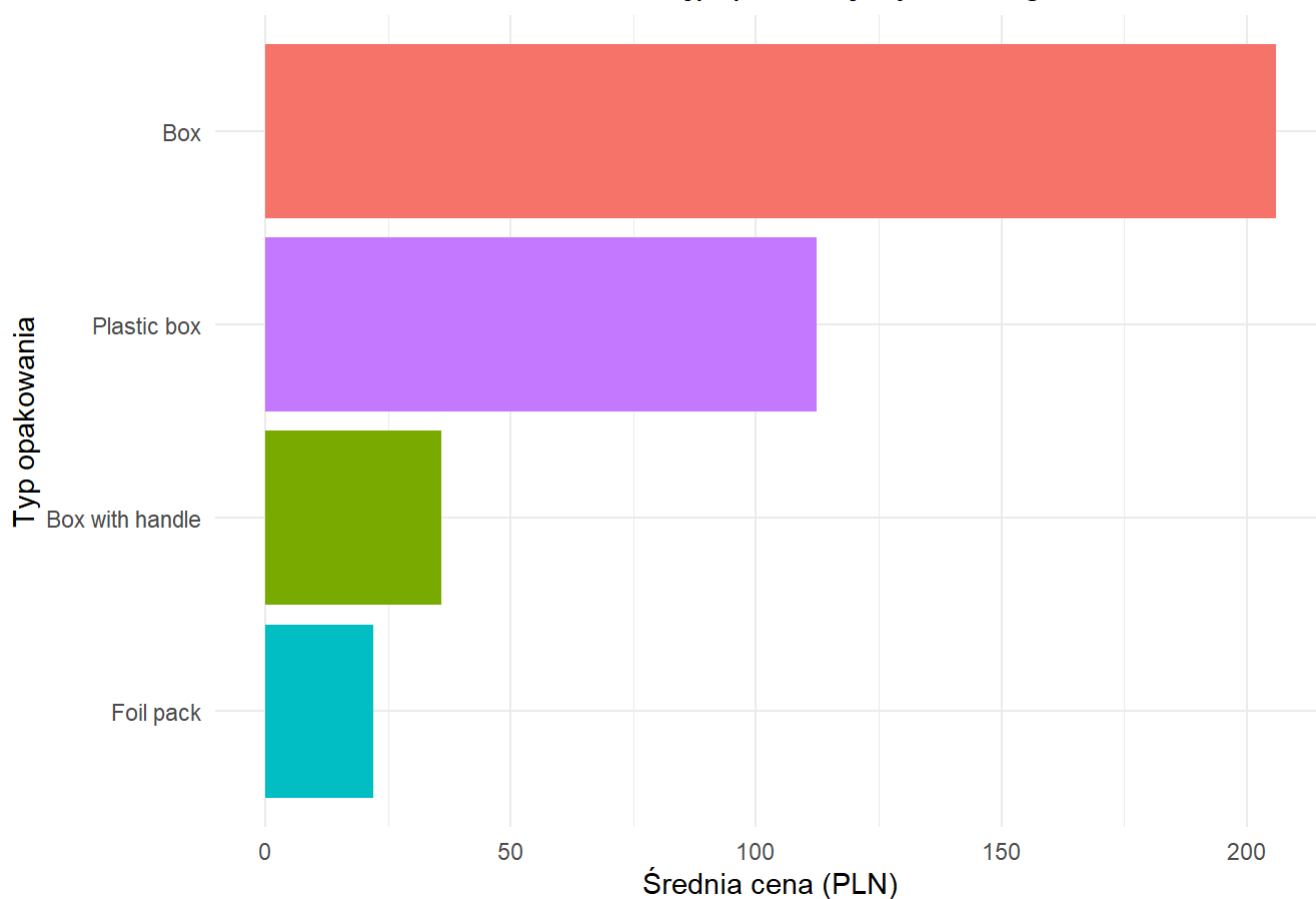
Liczność 4 najpopularniejszych opakowań zestawów Lego



Jak zauważyć można, klasyczne pudełka dominują tę kategorię. Mimo to spróbujmy przeanalizować, jak kształtują się ceny dla tych 4 najpopularniejszych kategorii.

Zacznijmy od zbadania średnich arytmetycznych cen za zestaw w danym opakowaniu.

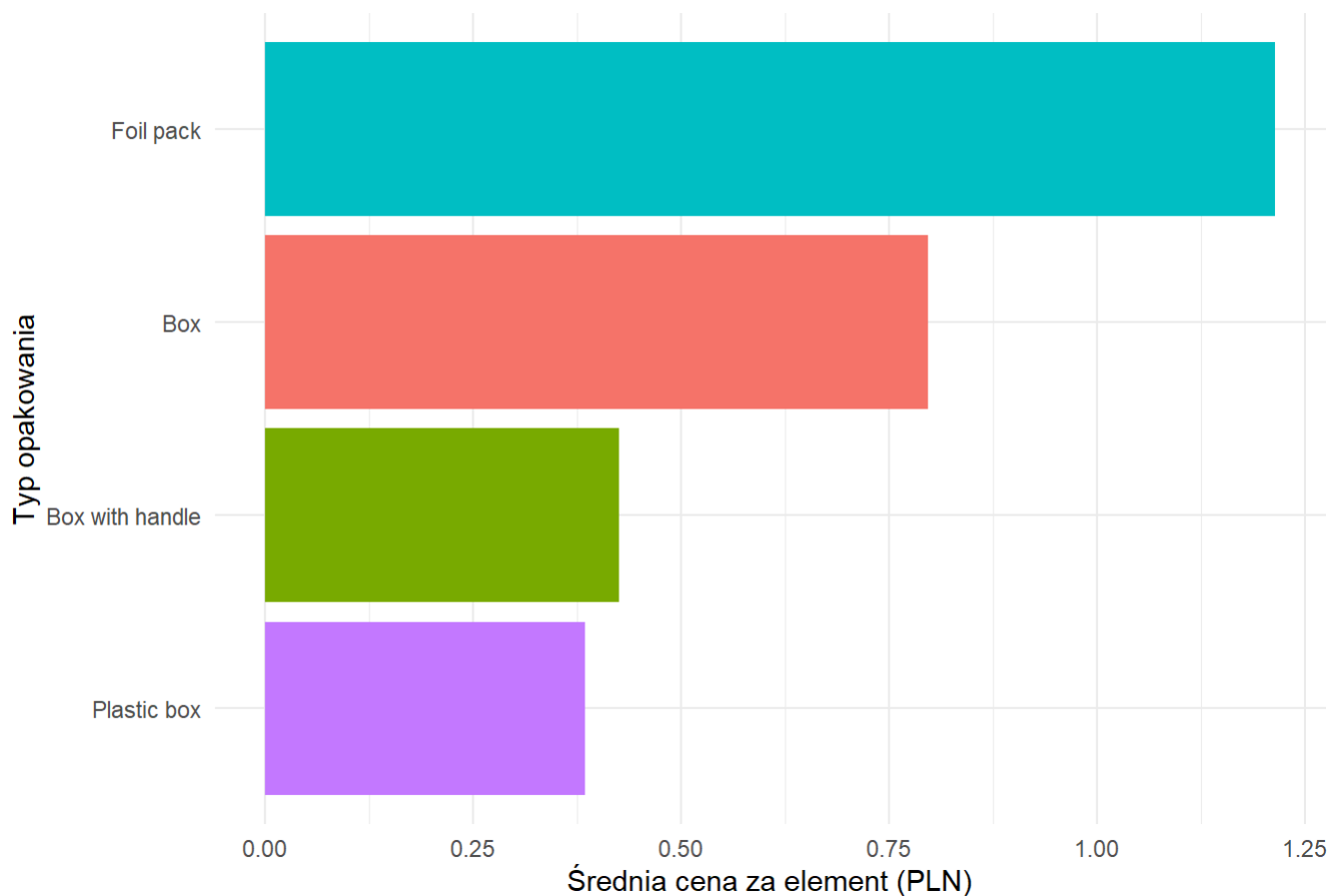
Średnia cena za zestaw w 4 najpopularniejszych kategoriach



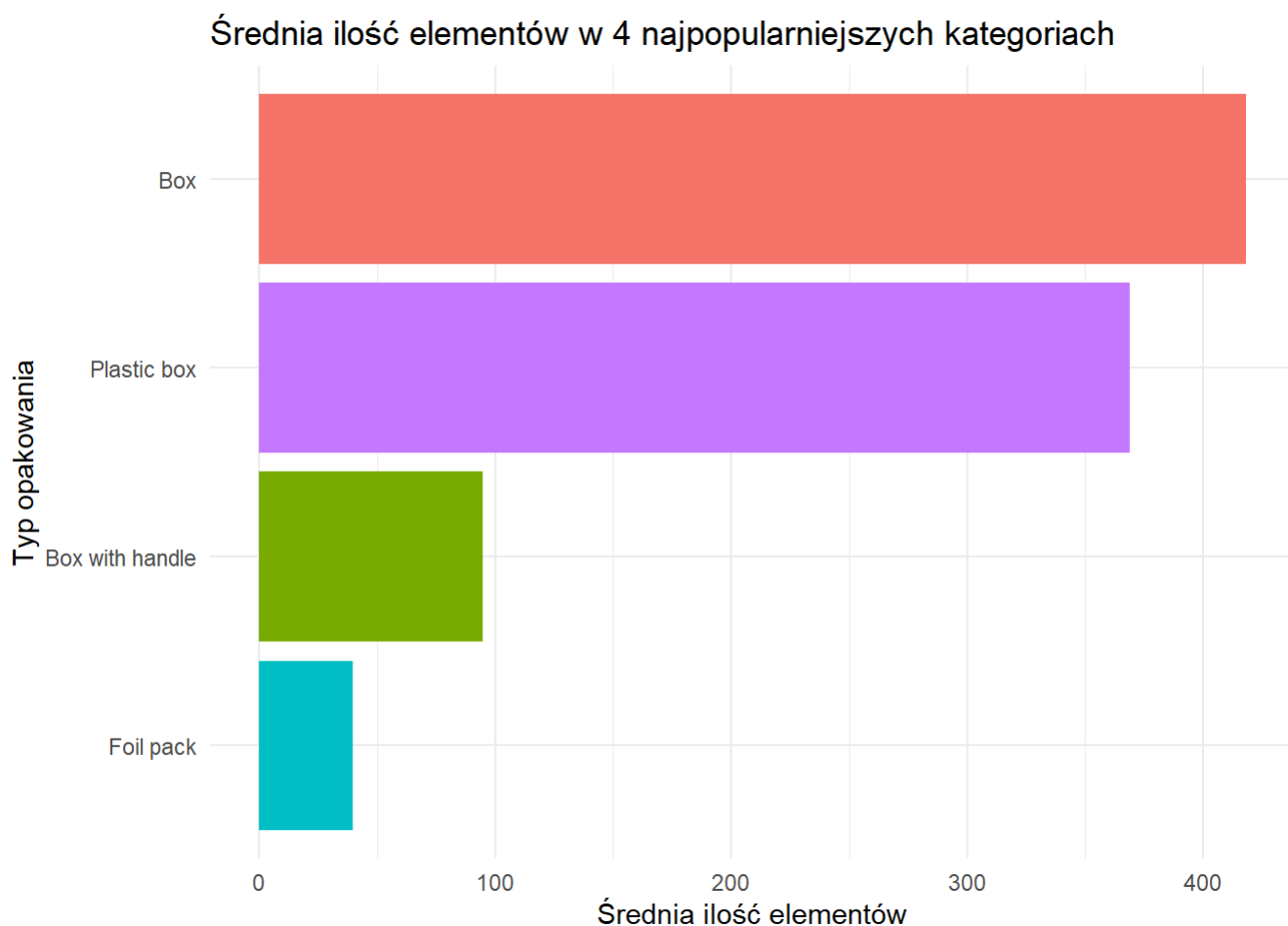
Jak widać największa średnia cena za zestaw, pojawia się przy zestawach pakowanych w pudełkach. Nie jest to jednak zaskoczenie, ponieważ zestawy pakowane w pudełka są statystycznie największe (posiadają największą ilość klocków), a to jest silnie skorelowane z ceną.

Z tego samego powodu, ponownie lepszym współczynnikiem do określania zależności między rodzajem opakowania a ceną będzie średnia cena za element.

Średnia cena za element w 4 najpopularniejszych kategoriach



Po zbadaniu średniej ceny za element, w zależności od rodzaju opakowania, widzimy, że ceny kształtują się inaczej, niż mogłoby to wynikać z poprzedniego wykresu. W przypadku opakowań typu "Foil pack" możemy zaobserwować najwyższą cenę za element. Dzieje się tak, ponieważ zestawy ze stosunkowo małą ilością klocków, posiadają większą cenę za element. Jednakże, żeby upewnić się, że nasze przypuszczenia są prawidłowe, sprawdźmy, ile średnio elementów posiadają zestawy w podanych opakowaniach.



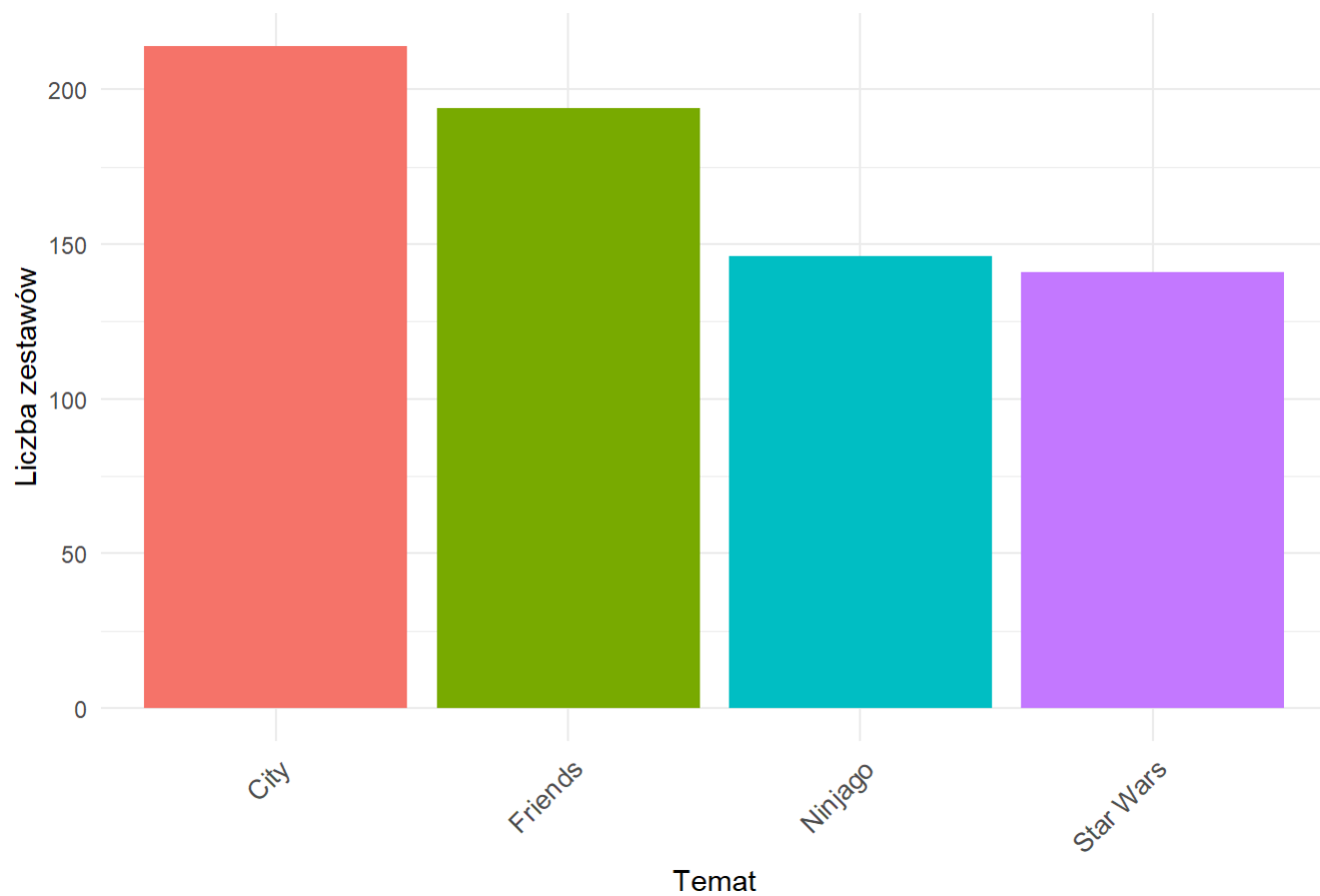
Przypuszczenia, że klocki zapakowane w foliowe opakowania posiadają mniej elementów, okazały się trafne.

Na podstawie badań przeprowadzonych w tej sekcji, trudno wskazać jednoznacznie czy opakowanie posiada realny wpływ na cenę zestawów. Istotnie zauważyć możemy, że istnieją różnice pomiędzy zestawami pakowanymi w inny sposób. Niektóre z tych różnic możemy tłumaczyć tym, że zestawy o dużej ilości klocków posiadają mniejszą cenę za element. Mimo to wciąż istnieje wyraźna różnica cenowa pomiędzy zestawami pakowanymi w pudełka a tymi zapakowanymi w plastikowe odpowiedniki, pomimo zbliżonej średniej elementów w zestawie.

2.2 Analiza porównawcza wybranych serii klocków

W tej części skupimy się na analizie porównawczej wybranych kategorii klocków LEGO. W tym celu spośród wszystkich dostępnych danych wybraliśmy 4 najbardziej popularne serie zestawów.

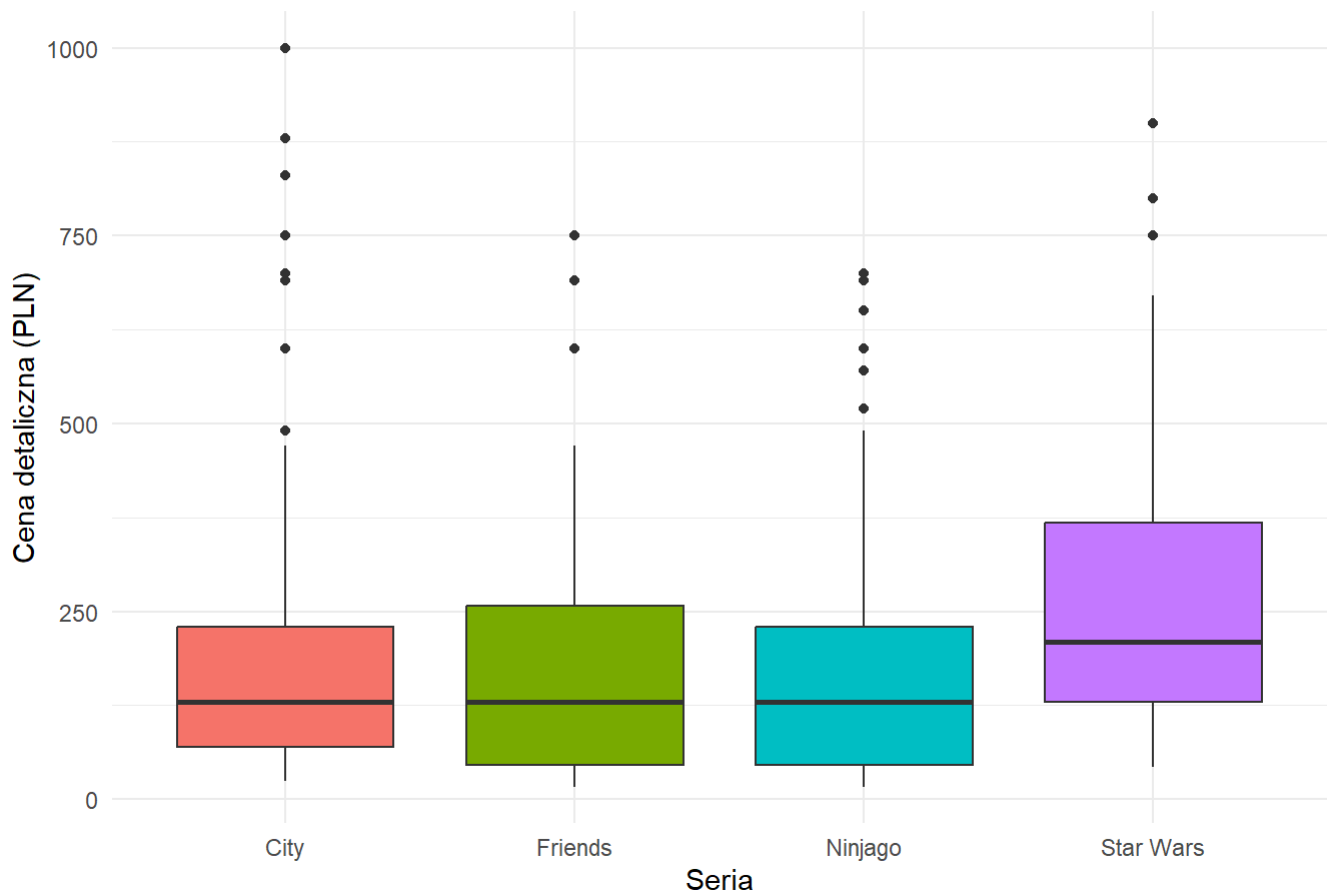
Liczność 4 najliczniejszych zestawów LEGO według tematów



City jest najliczniejszą serią (214 zestawów), następne w kolejności są Friends, Ninjago oraz Star Wars.

W pierwszej kolejności przeanalizujemy rozkłady cen powyższych serii klocków.

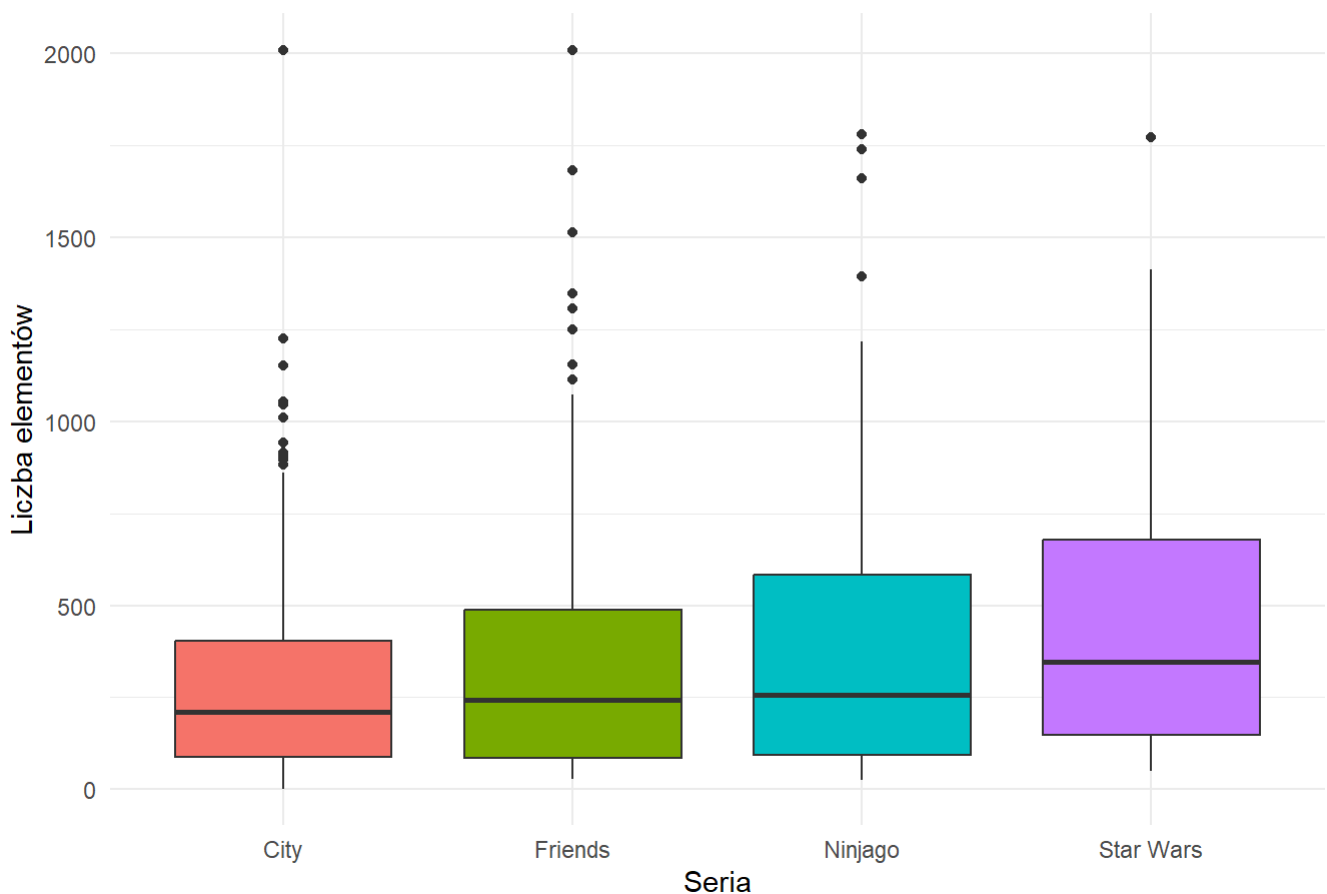
Rozkład cen zestawów LEGO w 4 najliczniejszych seriach



Star Wars charakteryzuje się znacząco wyższą medianą od pozostałych serii, razem z City są to najbardziej zróżnicowane rodzaje zestawów pod względem cenowym. Możemy także zauważyć, że większość cen zestawów Star Wars mieści się pomiędzy 2 a 3 kwartyłem (w pozostałych seriach ta cecha nie występuje lub jest bardziej marginalna), możemy zatem wywnioskować, że seria ta jest średnio droższa od innych. Natomiast z serii City pochodzą najbardziej odstające obserwacje. LEGO Friends oraz Ninjago mają podobne rozkłady cen.

Teraz przyjrzyjmy się temu, jak rozkłada się liczba elementów w analizowanych seriach.

Rozkład liczby elementów zestawów LEGO w 4 najliczniejszych seriach

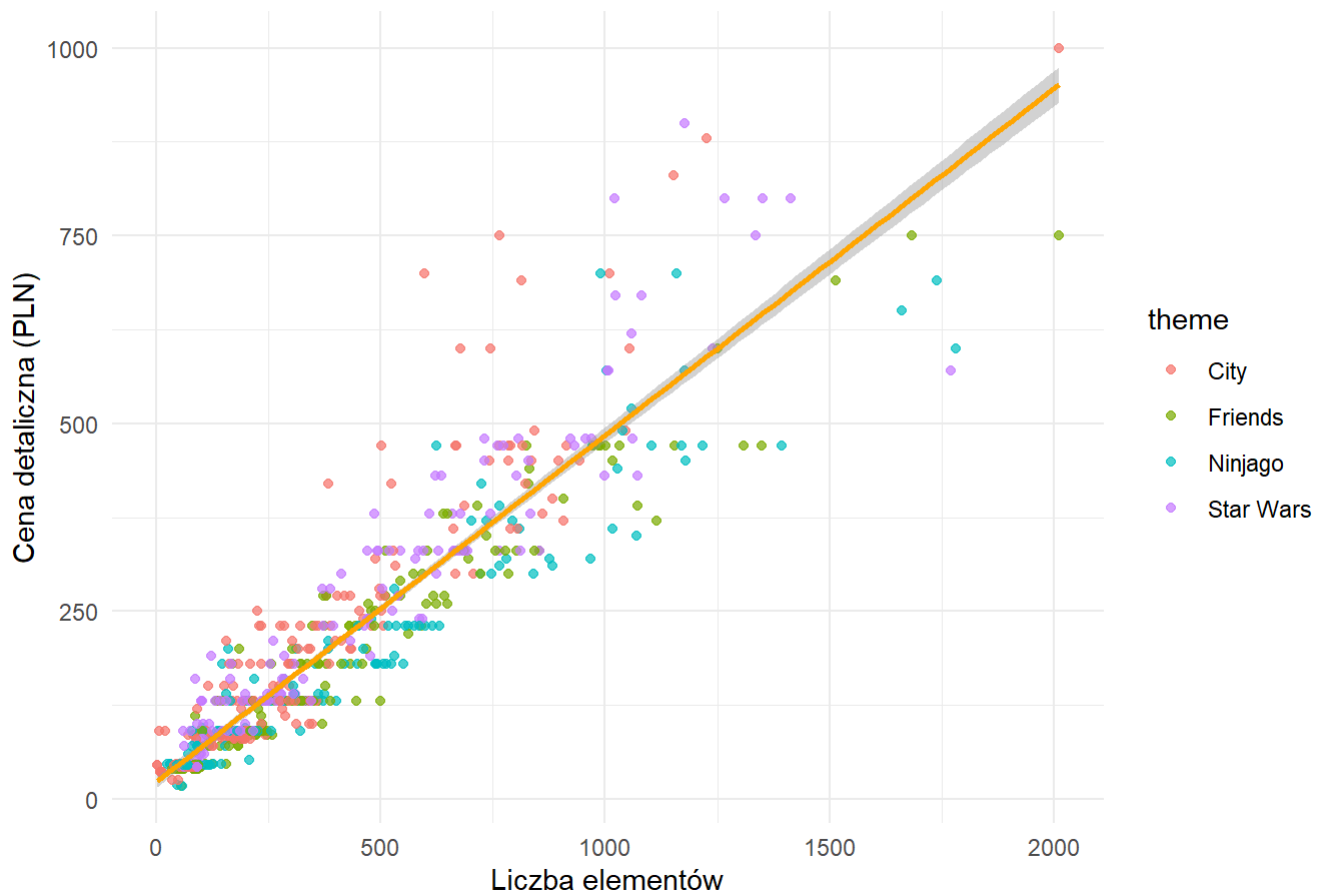


Możemy zauważyć, że każda z serii jest dosyć mocno zróżnicowana pod względem liczby elementów w zestawie, na czele z LEGO City. Powyższy wykres ukazuje także zależności pomiędzy średnimi liczbami elementów w każdej serii. Grupując zestawy rosnąco, według średniej liczby elementów, na pierwszym miejscu plasuje się City, następnie Friends, Ninjago i stawkę zamyka Star Wars.

Czy to oznacza, że Star Wars jest średnio najdroższą serią ze względu na to, że statystycznie w jej zestawach znajdziemy najwięcej klocków?

Zanim odpowiemy na to pytanie, przyjrzyjmy się zależności pomiędzy ceną, a liczbą elementów w 4 najpopularniejszych rodzajach zestawów.

Liczba elementów vs Cena w 4 najliczniejszych seriach LEGO



Wniosek jest jasny - dla 4 najliczniejszych serii występuje bardzo silna dodatnia korelacja pomiędzy ceną zestawu a liczbą jego elementów. Jest ona silniejsza od korelacji, którą zbadaliśmy podczas analizy zbiorczej, co dodatkowo uwypukla różnice pomiędzy seriami.

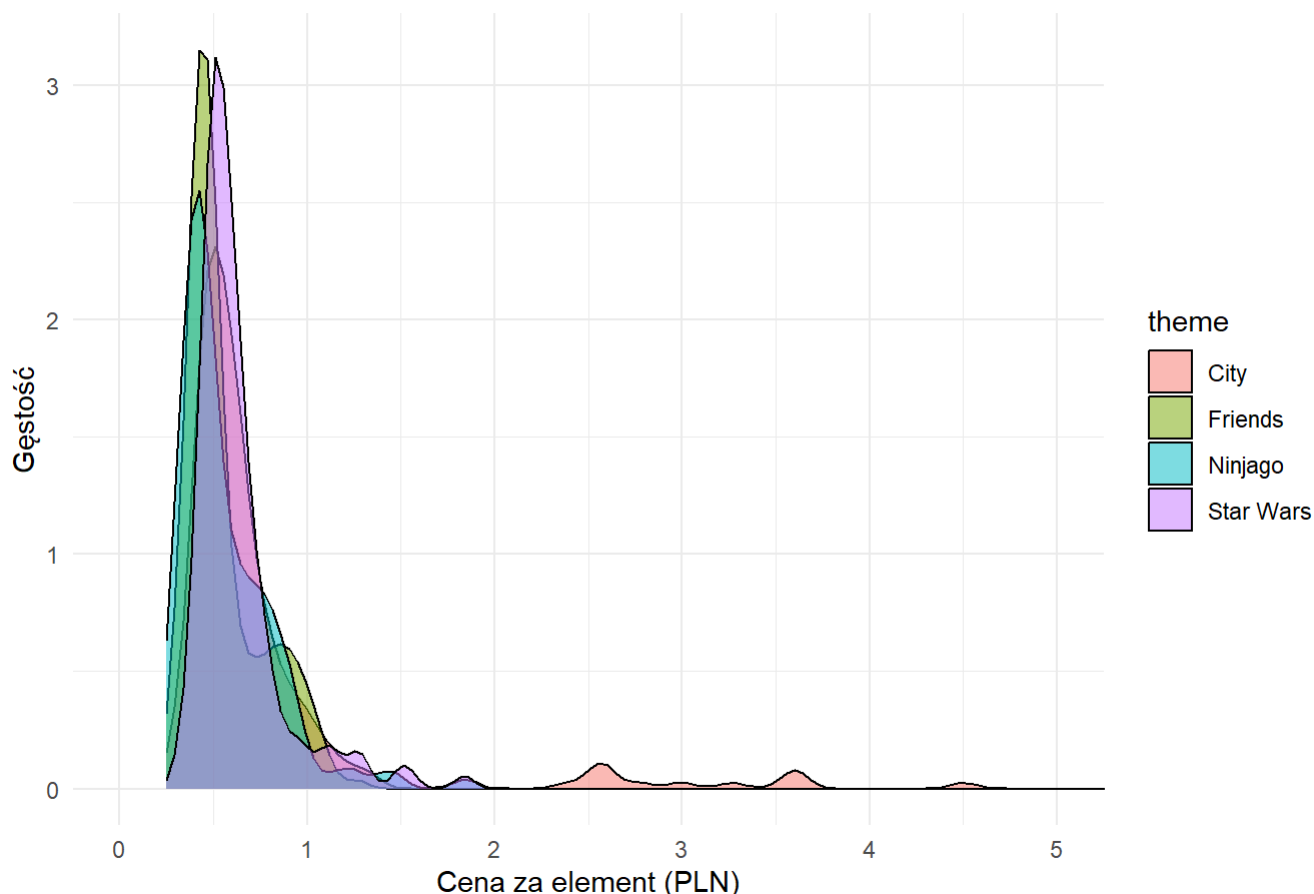
Korelacja pomiędzy ceną zestawu a liczbą jego elementów dla 4 najliczniejszych serii LEGO

	V1	V2	V3	V4
theme	City	Friends	Ninjago	Star Wars
correlation	0.93	0.97	0.95	0.94

Jak wynika z powyższej tabeli, korelacja pomiędzy ceną zestawu a jego liczbą elementów jest bardzo silna dla każdej serii z osobna.

Powróćmy więc, do wcześniej zadanego pytania odnoszącego się do serii Star Wars i je uogólnijmy - czy dla tych rodzajów zestawów cena zależy wyłącznie od liczby klocków w środku? Aby móc odpowiedzieć na to pytanie przeanalizujemy, jak rozkłada się cena za pojedynczy element.

Porównanie gęstości ceny za element



Cena za element jest w ogólności zbliżona dla każdej z badanych serii, jednak można się tu doszukać kilku subtelnych różnic. LEGO City jest serią o najbardziej odstających obserwacjach, co oznacza, że możemy znaleźć tam zestawy, które mają bardzo wysoką cenę za pojedynczy element, jednak są to obserwacje wysoce rzadkie, które nie grają istotnej roli w analizie ogólnych trendów.

Naszym zdaniem, najbardziej wyróżniającą się serią na powyższym wykresie jest Star Wars. Gęstość ceny za element dla tej kategorii jest przesunięta w prawo w porównaniu do innych rodzajów zestawów. Oznacza to, że wysoka liczba elementów w produktach z tej serii nie jest jedynym czynnikiem wpływającym na ich cenę. Przypuszczamy, że wyższe ceny zestawów z serii Star Wars mogą wynikać z jej licencjonowanego charakteru.

3 Podsumowanie

W przeprowadzonej analizie skupiliśmy się na kilku kluczowych pytaniach badawczych, starając się zidentyfikować czynniki wpływające na kształtowanie się cen zestawów, odkryć istotne trendy w czasie, a także uwypuklić różnice pomiędzy poszczególnymi seriami LEGO.

Najważniejszym czynnikiem determinującym cenę zestawu okazała się liczba elementów. W analizie zbiorczej udało nam się wykryć silną korelację łączącą tę cechę z ceną. Mimo że był to najistotniejszy czynnik, nie był on jedynym. Po wnikliwej analizie doszliśmy do wniosku, że cena zestawu zmienia swoją wartość w czasie, a najdroższym miesiącem okazał się grudzień.

Zauważyliśmy także istotną zależność, mówiącą o tym, że zestawy posiadające do około 50-100 elementów są droższe co do ceny za element, w porównaniu z większymi zestawami. Odnaleźliśmy również korelację z ilością minifigurek w zestawie. Przypuszczamy jednak, że zależność ta wynika z pierwszego aspektu (korelacja pomiędzy liczbą elementów w zestawie i jego ceną), ponieważ większe

zestawy posiadają statystycznie więcej minifigurek. Ustaliliśmy także, że dla dużej ilości minifigurek (więcej niż 7) danych jest za mało, żeby móc je porównywać z pozostałą częścią. Badanie rodzajów opakowań pozwoliło natomiast na zaobserwowanie, że istnieją inne czynniki wpływające na koszt klocków LEGO, niż tylko ich ilość. Trudno jednak jest wskazać, dlaczego tak się dzieje, ale klocki LEGO pakowane w plastikowych pudełkach, okazały się statystycznie tańsze za element od tych, pakowanych w klasyczne pudełka, mimo ich podobnej średniej ilości w zestawie.

W ramach analizy porównawczej skupiliśmy się na czterech najpopularniejszych seriach: City, Friends, Ninjago i Star Wars.

Udało nam się ustalić, że Star Wars ma najwyższą średnią oraz medianę cen, a razem z City, także ponadprzeciętne zróżnicowanie tej zmiennej. Samo City wyróżnia się ekstremalnymi obserwacjami o wysokich wartościach zestawów, natomiast Friends i Ninjago mają podobne rozkłady cen.

Doszliśmy także do wniosku, że średnio najdroższą serią LEGO (spośród 4 przeanalizowanych), zarówno pod względem ceny zestawów, jak i ceny za pojedynczy element, jest Star Wars. Wyższe koszty produktów z tej grupy wynikają nie tylko z liczby elementów, ale także z dodatkowych czynników, za jakie podejrzewamy licencjonowanie.

Podsumowując, cena zestawów LEGO skorelowana jest z ilością elementów, czasem, rodzajem opakowania i serią.