**Projektowanie i implementacja hurtowni danych dla analizy efektywności sprzedaży i strategii marketingowych w przedsiębiorstwie Adventure Works**

Uniwersytet Ekonomiczny w Katowicach

Kosarski Ksawery

Inżynieria Hurtowni Danych

Spis treści

[Modelowanie struktury konstelacji faktów dla potrzeb budowy hurtowni danych. 3](#_Toc166239309)

[Cel realizacji projektu 9](#_Toc166239310)

[Opis źródeł danych 10](#_Toc166239311)

[Proces 1 20](#_Toc166239312)

[Proces 2 22](#_Toc166239313)

[Proces 3 25](#_Toc166239314)

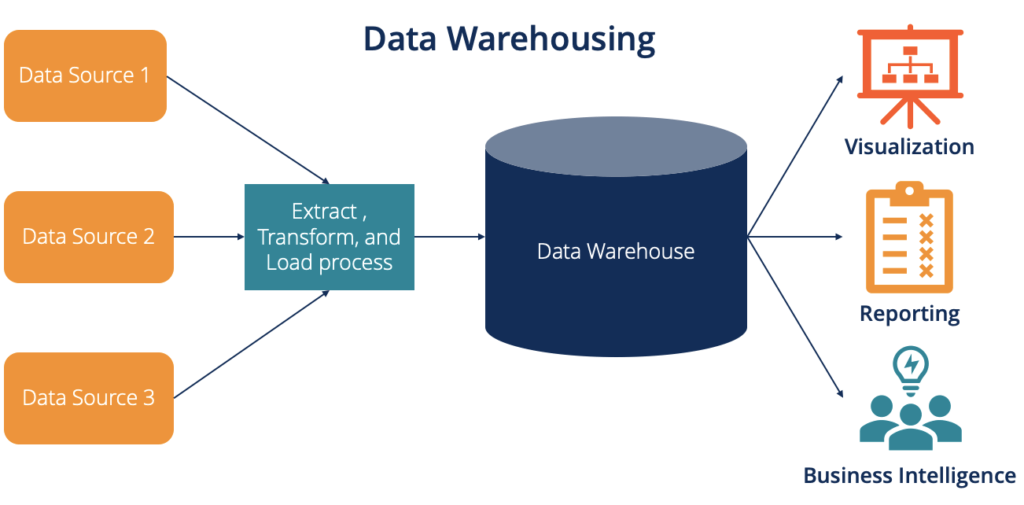
[Proces 4 27](#_Toc166239315)

[Proces 5 29](#_Toc166239316)

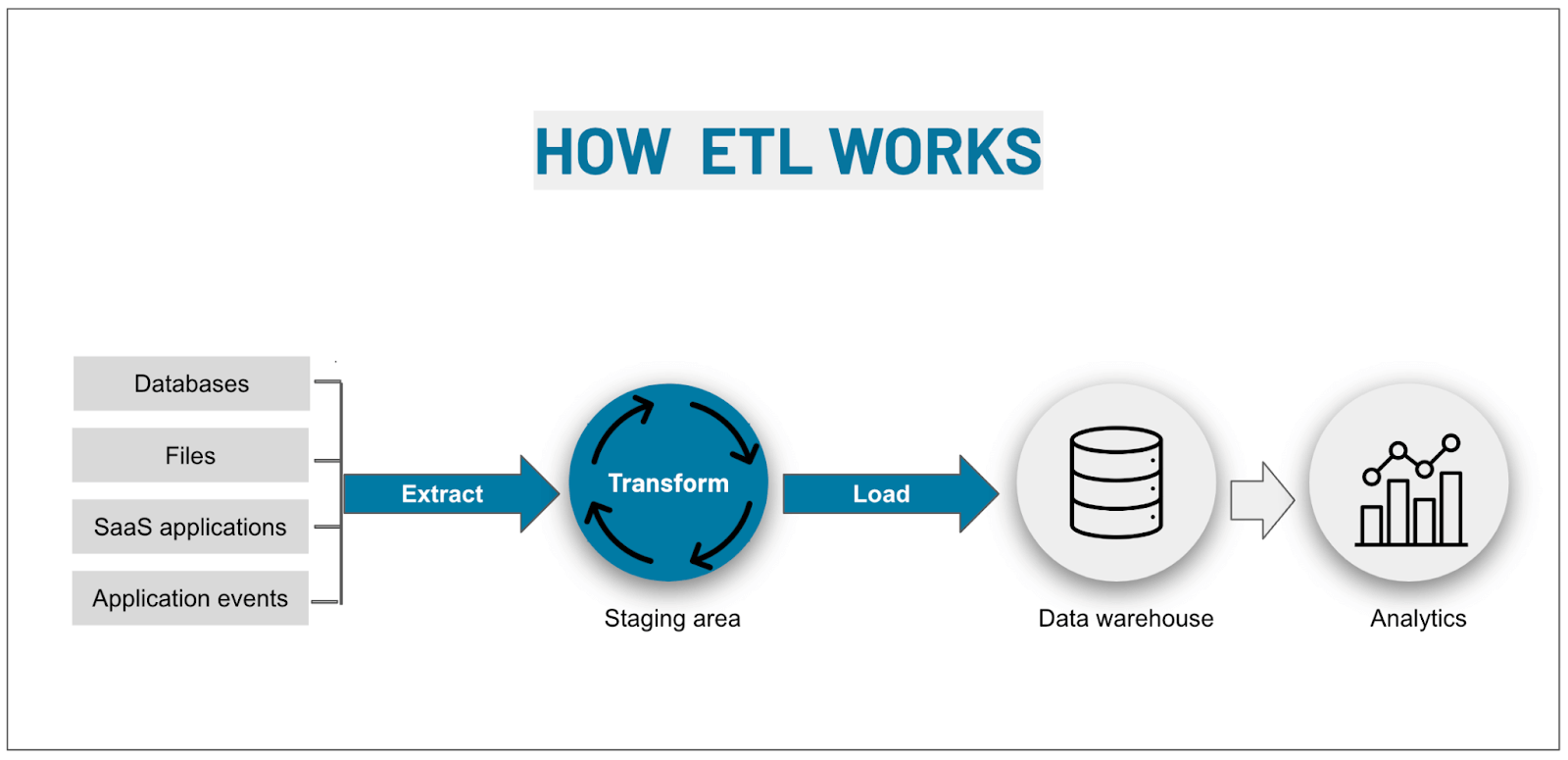
[Wnioski końcowe 32](#_Toc166239317)

# Modelowanie struktury konstelacji faktów dla potrzeb budowy hurtowni danych.

Hurtownia danych (data warehouse) to centralne repozytorium, które agreguje dane z różnych źródeł w celu wydajnego zarządzania i przechowywania. Jej główną funkcją jest wsparcie procesów analitycznych, które pomagają w podejmowaniu decyzji w organizacji.

Hurtownie danych konsolidują informacje z różnych źródeł zapewniając ich spójność i jednolitość, dzięki temu analiza danych staje się łatwiejsza a firma uzyskuje przekrojowy obraz swoich działań. Dane przechowywane w hurtowni są traktowane jako stałe, czyli historyczne dane nie ulegają zmianom co zapewnia wiarygodność i stabilność w retrospektywnych analizach. *Rysunek 1. Tworzenie oraz zastosowanie hurtowni danych  
Źródło:* [*https://corporatefinanceinstitute.com/resources/business-intelligence/data-warehousing/*](https://corporatefinanceinstitute.com/resources/business-intelligence/data-warehousing/)

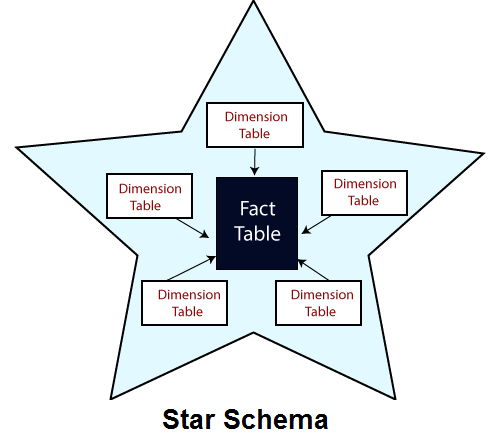
Wdrożenie hurtowni danych ma wiele korzyści, przede wszystkim przyspieszenie procesu decyzyjnego poprzez natychmiastowy dostęp do zintegrowanych i dobrze przetworzonych danych. Automatyzacja zbierania danych i generowania raportów oszczędza czas co pozwala pracownikom na koncentrację na innych zadaniach. Firma dzięki temu staje się bardziej konkurencyjna, mogąc szybko reagować na zmieniające się warunki rynkowe.

Proces wdrożenia hurtowni rozpoczyna się od analizy potrzeb i wymagań danego przedsiębiorstwa. Wymaga to współpracy między analitykami biznesowymi, zespołem IT oraz użytkownikami końcowymi, aby zidentyfikować oczekiwania użytkowników końcowych co do danych, procesy decyzyjne oraz potrzeby raportowe. Taka współpraca pozwala na ustalenie jakie dane są potrzebne i jak będą wykorzystywane. Następnie przechodzi się do procesu modelowania danych który dzieli się na:  
- Modelowanie logiczne które obejmuje tworzenie abstrakcyjnego modelu danych, który odzwierciedla struktury danych, niezależnie od fizycznej implementacji. Na tym etapie definiuje się kluczowe encje, ich atrybuty oraz relacje między nimi. Celem jest utworzenie zrozumiałego modelu, który będzie dostosowany do wymagań biznesowych.  
-Modelowanie fizyczne jest procesem dostosowania modelu logicznego do konkretnego środowiska technologicznego oraz optymalizacja pod kątem wydajności operacji na danych. W tym etapie określa się struktury tabel, klucze oraz partycjonowanie danych oraz inne aspekty które mają wpływ na prędkość i efektywność przechowywania danych.   
Kolejny krok to załadowanie danych do hurtowni, za względu na różne formaty oraz struktury danych to zadanie bywa utrudnione. Przydatne w tym miejscu są procesy ETL lub alternatywnie ELT. *Rysunek 2. Opis działania procesów ETL  
Źródło:* [*https://blog.coupler.io/what-is-etl/*](https://blog.coupler.io/what-is-etl/)

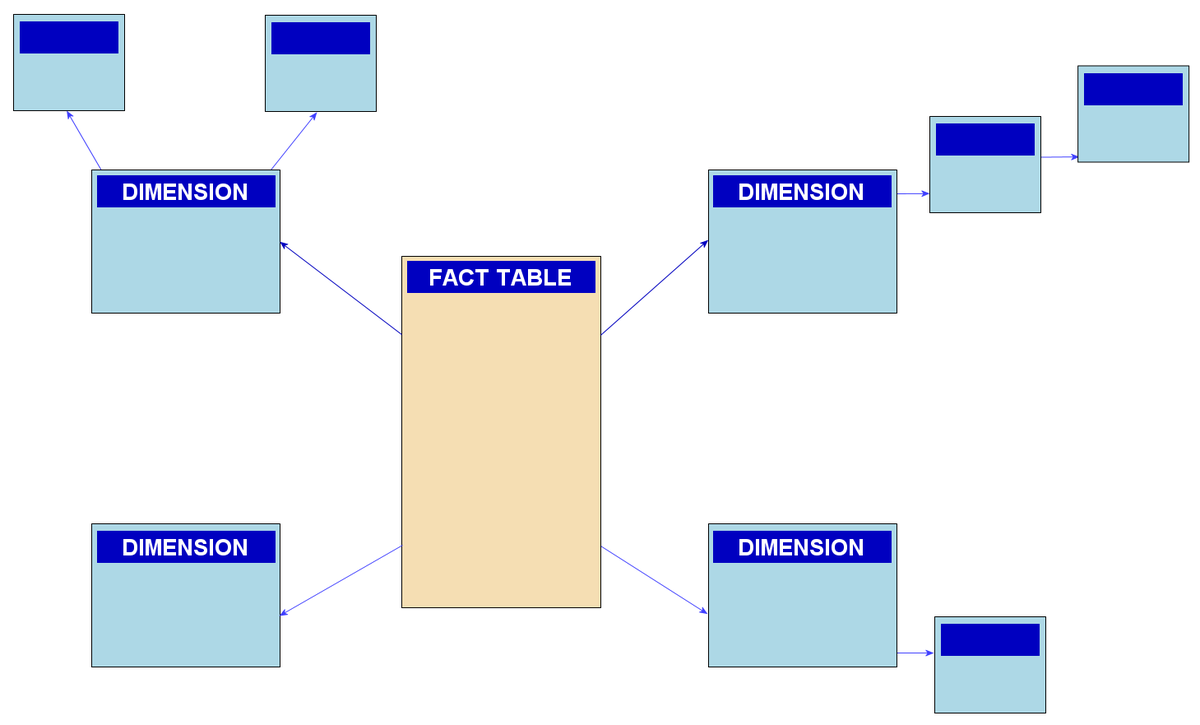
Ekstrakcja – Etap, w którym dane pozyskiwane są z różnych źródeł zewnętrznych, gdzie można natrafić na wiele różnorakich trudności takich jak brak dostępności danych lub problemy ze zgodnością formatów. Dane są importowane do obszaru tymczasowego (staging area), gdzie przeprowadza się transformacje w sposób nie obciążający systemu. Aby optymalizować proces ekstrakcji danych i zminimalizować ryzyko pogorszenia wydajności, zaleca się przed rozpoczęciem etapu pozyskiwania danych zweryfikować ich pochodzenie, sprawdzić typy danych oraz przygotować alternatywne ścieżki postępowania na wypadek różnych rodzajów niepowodzeń.  
Transformacja – Oryginalne, surowe dane rzadko są odpowiednie do bezpośredniego użycia, dlatego niezbędne jest podjęcie działań mających na celu ich adaptację do późniejszego wykorzystania. Transformacja danych często wiąże się z koniecznością integracji danych zewnętrznych z wewnętrznymi danymi firmy, aby stworzyć spójną strukturę. Etap ten obejmuje takie czynności jak oczyszczanie, mapowanie, agregacja, normalizacja i inne przekształcenia danych, żeby były one spójne i gotowe do analiz  
Ładowanie – Ostatnim etapem w procesie ELT jest załadunek przetworzonych danych do hurtowni danych. Jest to proces, w którym dane są umieszczane w odpowiednich tabelach schematu, przygotowane do analizy i raportowania.

Ostatnim etapem budowy hurtowni danych jest testowanie wydajności oraz optymalizacja zapytań. Poprawna optymalizacja zapewnia, że system spełnia wymagania czasu odpowiedzi i jest w stanie obsłużyć oczekiwany wolumen danych. Po tych etapach pozostaje już tylko zarządzanie zmianami w hurtowni danych, ponieważ systemy danych są dynamiczne i muszą ewoluować w odpowiedzi na zmieniające się potrzeby biznesowe. Zarządzanie zmianą obejmuje monitorowanie, przegląd i implementację zmian w strukturze danych i procesach ETL. Zmiany musza być wprowadzane metodycznie, z odpowiednimi testami i zatwierdzeniami, aby zapewnić integralność danych.  
  
  
  
  
  
  
  
  
  
Podczas wdrażania projektu hurtowni danych i modelowania logicznego należy zdecydować się jaką strukturę hurtowni wybrać. Wybór zależy od założeń projektowych i potrzeb biznesowych. W tym kontekście, oprócz popularnych schematów gwiazdy czy płatka śniegu, warto zwrócić uwagę na konstelację faktów. Konstelacja faktów to zaawansowany sposób organizacji tabel w hurtowni danych, który pozwala na jeszcze większą elastyczność poprzez możliwość obsługi wielu tematów biznesowych w ramach jednej hurtowni. Jest to rozbudowanie schematu gwiazdy, które pozwala na bardziej skomplikowane analizy i zestawienia danych z różnych perspektyw biznesowych. Schemat ten jest używany, gdy potrzeba zintegrować więcej niż jedno zagadnienie biznesowe w jednej strukturze hurtowni, zazwyczaj w dużych i złożonych organizacjach.

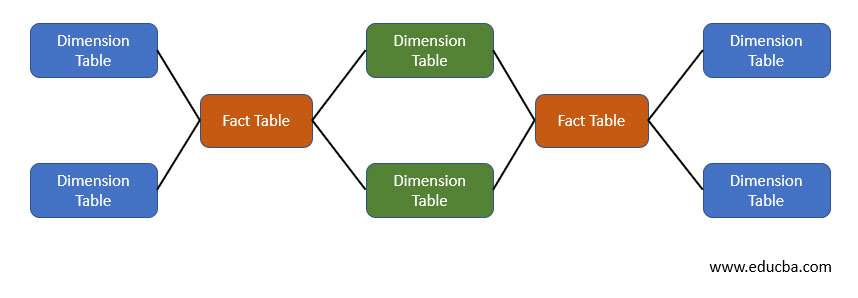
Schemat gwiazdy to struktura projektowa hurtowni danych zorientowana na temat. Charakteryzuje się uproszczoną, intuicyjną strukturą, która umożliwia efektywne przetwarzanie zapytań i analiz danych. W centrum schematu gwiazdy znajduje się tabela faktu, która przechowuje kluczowe miary biznesowe oraz wskaźniki, jak np. sprzedaż, zyski, ilość transakcji i jest powiązana z tabelami wymiarów za pomocą kluczy obcych, które otaczają tabelę faktu i dostarczają kontekstu dla miar przechowywanych w tabeli faktu. Przykłady tabel wymiarów to czas, klient, produkt, lokalizacja, każda z nich opisuje atrybuty powiązane z odpowiednimi wymiarami. Zalety schematu gwiazdy to jej wydajność dzięki uproszczonej strukturze zapewniającej szybkie przetwarzanie zapytań, które jest kluczowe dla analityki biznesowej oraz następną zaletą jest łatwość zrozumienia, ponieważ struktura jest łatwa do przyswojenia nawet dla użytkowników biznesowych, co ułatwia analizę danych, jednak ma też ograniczenia takie jak redundancja danych, która może prowadzić do duplikacji danych w tabelach wymiarów.

  
*Rysunek 3. Schemat gwiazdy  
Źródło:* [*https://www.javatpoint.com/data-warehouse-what-is-star-schema*](https://www.javatpoint.com/data-warehouse-what-is-star-schema)

Płatek śniegu jest rozwinięciem schematu gwiazdy, gdzie tabele wymiarów są dodatkowo normalizowane poprzez dzielenie ich na mniejsze tabele, co pozwala na eliminację redundancji danych. W schemacie płatka śniegu tabele wymiarów są dalej rozbijane na mniejsze tabele, co pozwala na bardziej szczegółowe modelowanie danych, a proces normalizacji redukuje redundancję danych, poprawia integralność i ułatwia zarządzanie danymi. Zalety takiego schematu to zoptymalizowane zużycie przestrzeni dyskowej, ponieważ mniejsza redundancja danych prowadzi do oszczędności w zużyciu miejsca na dysku oraz lepsza integralność danych, gdzie normalizacja zapewnia wyższą integralność przez eliminację niespójności. Ograniczenia to złożoność, która może utrudniać zrozumienie i operowanie modelem, zwłaszcza dla użytkowników biznesowych, oraz wydajność zapytań, gdzie większa ilość złączeń może spowolnić przetwarzanie zapytań, co jest niekorzystne w środowiskach wymagających szybkiej odpowiedzi.

 *Rysunek 4. Schemat płatka śniegu  
Źródło:* [*https://en.wikipedia.org/wiki/Snowflake\_schema*](https://en.wikipedia.org/wiki/Snowflake_schema)

W schemacie konstelacji faktów, różne tabele faktu mogą być połączone poprzez wspólne tabele wymiarów. Każda tabela faktu reprezentuje inne zagadnienie biznesowe, ale dzięki wspólnym wymiarom, możliwe jest tworzenie złożonych zapytań przechodzących przez wiele aspektów biznesowych. Wspólne tabele wymiarów w schemacie konstelacji faktów są kluczowe dla integracji danych. Dzięki temu, że różne tabele faktu odnoszą się do tych samych tabel wymiarów, możliwe jest przeprowadzanie analiz na przecięciu różnych dziedzin, takich jak sprzedaż, logistyka czy zarządzanie zasobami ludzkimi.

  
*Rysunek 5. Schemat konstelacji faktów*

Schemat konstelacji faktów jest wykorzystywany w dużych przedsiębiorstwach, które potrzebują kompleksowej analizy danych z różnych działów. Pozwala na elastyczne badanie zależności między różnymi obszarami biznesowymi, co jest szczególnie przydatne w branżach takich jak finanse, sprzedaż detaliczna oraz telekomunikacja, gdzie działania w jednym segmencie mogą wpływać na wyniki w innym.

Ogromnym plusem zastosowania schematu konstelacji faktów jest wszechstronność analiz która umożliwia przeprowadzanie analiz wielowymiarowych, które są trudne do realizacji w prostszych schematach, a integracja danych sprzyja łączeniu danych z różnych obszarów biznesowych, co poprawia jakość analiz i decyzji strategicznych.   
Schemat konstelacji faktów jest bardziej złożony w implementacji i utrzymaniu niż schematy gwiazdy czy płatka śniegu, a ze względu na złożoność i wielkość danych, schematy konstelacji mogą wymagać bardziej wydajnego sprzętu i optymalizacji zapytań.

# Cel realizacji projektu

Celem realizacji projektu hurtowni danych, jest zbudowanie skutecznego i wydajnego systemu do gromadzenia, przetwarzania i analizowania dużych ilości danych sprzedażowych. Adventure Works Cycles to fikcyjna firma z siedzibą w Bothell, w stanie Waszyngton, działająca na rynkach Ameryki Północnej oraz Europy. Celem firmy jest zwiększenie jej udziału na rynkach światowych, poprawienie wyników sprzedaży z lat ubiegłych, lepsze dostosowanie oferty do preferencji odbiorców oraz zrozumienie trendów wpływających na decyzje zakupowe klientów. W ramach projektu hurtowni danych skoncentrowałem się na analizie sprzedażowej oferowanych produktów.



*Rysunek 6. Logo firmy adventure works cycles*

Początkowo przedstawiam schemat użytej bazy danych Adventure Works, dokładnie opisując struktury wykorzystanych tabel oraz pochodzenie i zakres danych źródłowych. Na tej podstawie stworzę procesy ETL (Extract, Transform, Load), które umożliwiają zbudowanie tabel wymiarów i faktów według schematu gwiazdy.

W części analitycznej projektu uzyskam informacje na temat ogólnej sprzedaży oraz efektów różnych promocji na wyniki sprzedażowe. Wykorzystując narzędzie SAS Data Integration Studio mogę efektywnie zarządzać integracją i przetwarzaniem danych, co pozwoli na szybkie generowanie analiz potrzebnych do strategicznych decyzji biznesowych. Analizy te będą obejmowały między innymi ocenę sprzedaży według produktu i okresu, analizę rentowności produktów, porównanie sprzedaży w okresach promocyjnych, a także badanie trendów zakupowych w różnych dniach tygodnia. Projekt ten pozwala Adventure Works Cycles lepiej rozumieć dynamikę rynku i wpasować się w zmieniające się preferencje klientów.

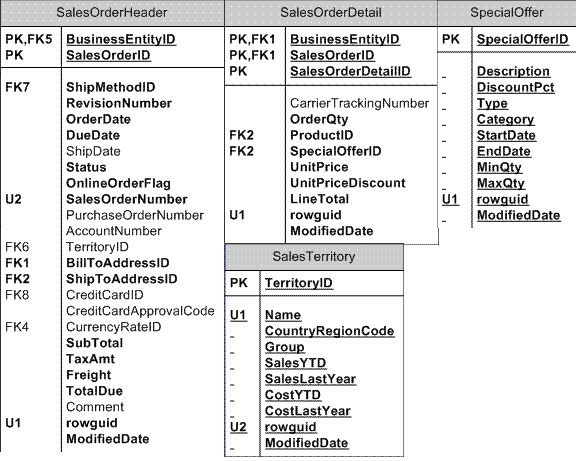
# Opis źródeł danych

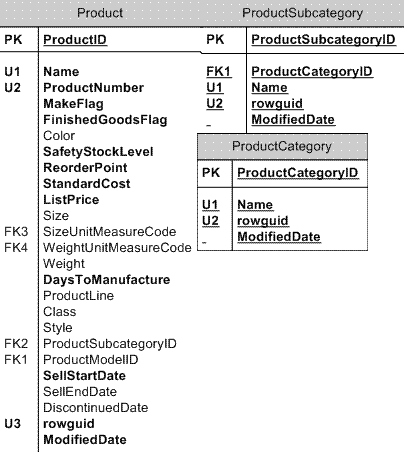
Do realizacji projektu użyłem bazę Adventure Works, przygotowaną przez firmę Microsoft, baza ta symuluje działalność fikcyjnej, międzynarodowej firmy produkującej rowery i akcesoria rowerowe, oraz dodatkowo oferującej usługi serwisowe. Dane w bazie są zorganizowane w kilka głównych schematów które obejmują obszary:

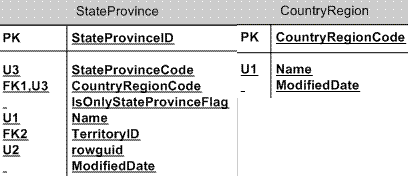
* Sales,
* Purchasing,
* Person,
* Production,
* Human Resources,
* DBO.

Obraz zawierający tekst, diagram, Plan, mapa

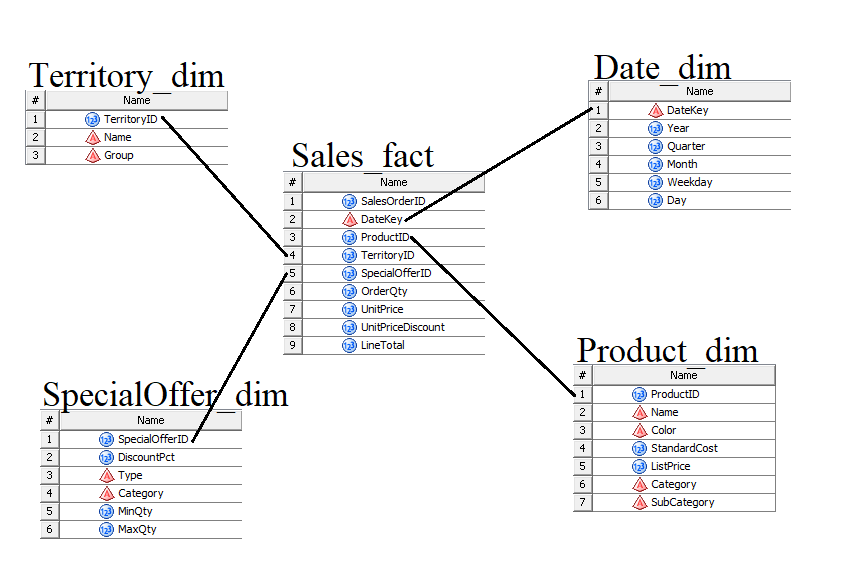
Opis wygenerowany automatycznie  
*Rysunek 7. Schemat bazy danych AdventureWorks*

W projekcie wykorzystałem tabele pochodzące z obszarów Sales, Production oraz Person. Część Sales jest w moim projekcie uniwersalna, ponieważ użyłem jej do wszystkich tabel oprócz wymiaru produktu, skorzystałem z tabel SalesOrderHeader, SalesTerritory, SalesSpecialOffer, SalesOrderDetails (Rysunek 8).  
*Rysunek 8.*  
Do utworzenia wymiaru produktu użyłem tabel Product, ProductSubcategory, ProductCategory(Rysunek 9).

  
*Rysunek 9.*

Z regionu Person wykorzystałem tabele PersonStateProvince oraz PersonCountryRegion, aby w wymiarze regionu uzyskać nazwy państw oraz kontynentów (Rysunek 10).   
*Rysunek 10.*  
Model logiczny hurtowni danych – schemat gwiazdy

W celu stworzenia hurtowni danych zdecydowałem się skorzystać ze schematu gwiazdy. Jest to prosty schemat, który polega na centralnej tabeli faktu oraz połączonych z nią tabel wymiarów, ułatwia to analizę danych, ponieważ można szybko i efektywnie dokonać przekrojów danych według różnych wymiarów takich czas, produkt czy region.

  
*Rysunek 11. Model logiczny hurtowni, schemat gwiazdy*  
Zbudowany przeze mnie schemat gwiazdy składa się tabeli faktów opisującej sprzedaż oraz tabel wymiarów czasu, produktu, regionu oraz rabatu.

Wymiar czasu

Date\_dim to tabela związana z wymiarem czasowym, której dane określają datę złożenia zamówienia, umożliwia przeglądanie i analizowanie informacji według roku, kwartału, miesiąca, dnia tygodnia i dnia.Obraz zawierający zrzut ekranu, Czcionka, linia

Opis wygenerowany automatycznie *Rysunek 12. Proces ETL tworzący wymiar czasu*  
 Do utworzenia tego wymiaru korzystam z tabeli SALES\_SALESORDERHEADER. Znajdują się w niej kluczowe informacje o zamówieniach w tym data złożenia zamówienia. Korzystając z węzła Create Table(Rysunek 13) tworzę kolumny DateKey, Year, Quarter, Month, Weekday oraz Day na podstawie daty zamówienia (OrderDate). Do utworzenia identyfikatora czasu skorzystałem z funkcji:   
compress(substr(put(YEAR(DATEPART(SALES\_SALESORDERHEADER."OrderDate"n )),4.),3,2) ||'0'|| put(MONTH(DATEPART(SALES\_SALESORDERHEADER."OrderDate"n)),2.) ||'0'|| put(DAY(DATEPART(SALES\_SALESORDERHEADER."OrderDate"n )),2.))  
Pozostałe kolumny utworzyłem dzięki funkcjom daty i czasu YEAR(), QTR(), MONTH(), WEEKDAY(), DAY().

Obraz zawierający tekst, Czcionka, numer, linia

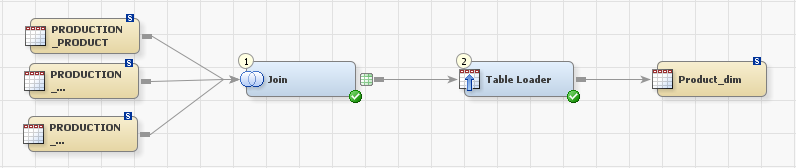
Opis wygenerowany automatycznie

*Rysunek 13.*

Otrzymana tabela wynikowa wygląda następująco(Rysunek 14):Obraz zawierający tekst, numer, zrzut ekranu, Równolegle

Opis wygenerowany automatycznie  
*Rysunek 14.*

Wymiar produktu

Product\_dim to tabela przechowywująca informacje o nazwie, kolorze, koszcie wyprodukowania, kategorii, podkategorii oraz jego cenie wyjściowej.   
*Rysunek 15. Proces ETL tworzący wymiar produktu*

Do utworzenia wymiaru produktu korzystam z tabel PRODUCTION\_PRODUCT, PRODUCTION\_PRODUCTSUBCATEGORY, PRODUCTION\_PRODUCTCATEGORY. Korzystając z węzła JOIN(Rysunek 16) łącze 3 tabele ze sobą w sposób przedstawiony poniżej: Obraz zawierający tekst, zrzut ekranu, linia, Czcionka

Opis wygenerowany automatycznie *Rysunek 16.*

Następnie za pomocą klauzuli SELECT(Rysunek 17) wybieram pola potrzbne do przyszłych analiz. Obraz zawierający tekst, zrzut ekranu, Równolegle, linia

Opis wygenerowany automatycznie  
*Rysunek 17.*

Po użyciu węzła TABLE LOADER tabela wymiaru produktu prezentuje się następująco: Obraz zawierający tekst, zrzut ekranu, numer, Czcionka

Opis wygenerowany automatycznie  
*Rysunek 18.*

Wymiar terytorium

Territory\_dim to tabela zawierająca informacje o regionie sprzedarzy. Dane w niej zawarte pozwalają na przegląd według kraju oraz kontynentu. Obraz zawierający zrzut ekranu, linia, diagram, tekst

Opis wygenerowany automatycznie  
*Rysunek 19. Proces ETL tworzący wymiar regionu*

Do utworzenia tego wymiaru korzystam z tabel PERSON\_STATEPROVINCE, SALES\_SALESTERRITORY, oraz PERSON\_COUNTRYREGION, z ostatniej z nich korzystam żeby przekształcić id w nazwy. Łączę je za pomocą węzłów JOIN(Rysunek 20). Obraz zawierający tekst, zrzut ekranu, linia, numer

Opis wygenerowany automatycznie  
*Rysunek 20.*  
A następnie używam węzła TABLE LOADER do załadowania powyższych danych(Rysunek 21).

Obraz zawierający tekst, zrzut ekranu, Czcionka, numer

Opis wygenerowany automatycznie  
*Rysunek 21.*

Wymiar promocji

SpecialOffer\_dim przechowywuje informajce o rabatach i promocjach na poszczególne zamówienia, pozwala to na ocenę wydajności poszczególnych promocji. Obraz zawierający tekst, zrzut ekranu, Prostokąt, Czcionka

Opis wygenerowany automatycznie  
*Rysunek 22. Proces ETL tworzący wymiar rabatu*

Wymiar tworzę za pomocą węzła CREATE TABLE(Rysunek 23) w którym przeładowywuje interesujące mnie kolumny. Obraz zawierający tekst, zrzut ekranu, numer, Czcionka

Opis wygenerowany automatycznie  
*Rysunek 23.*

Tabela faktów sprzedaży

Sales\_fact to tabela, która przechowuje głównie klucze obce odnoszące się do poszczególnych wymiarów oraz wartość oraz ilość poszczególnych zamówień. Obraz zawierający tekst, zrzut ekranu, linia, Czcionka

Opis wygenerowany automatycznie  
*Rysunek 24. Proces tworzący tabelę faktów sprzedaży*

Korzystam z tabel SALES\_SALESORDERHEADER oraz SALES\_SALESORDERDETAIL które zawierają wszystkie potrzebne mi klucze obce do wymiarów (Rysunek 25). Do stworzenia klucza daty korzystam z tej samej funkcji co w przyadku wymiaru czasu. Mimo poprawności danych oraz ich mappowania kolumna ta zwraca błąd który w żaden sposób nie rzutuje na dane wynikowe.

Obraz zawierający tekst, zrzut ekranu, numer, Równolegle

Opis wygenerowany automatycznie  
*Rysunek 25.*  
Nastepnie używam węzła TABLE LOADER(Rysunek 26) i tworzę tabelę faktów która prezentuje się następująco:

Obraz zawierający tekst, zrzut ekranu, Czcionka, linia

Opis wygenerowany automatycznie  
*Rysunek 26.*

Proces 1   
W pierwszym procesie przeprowadzona została analiza sprzedaży produktów. Analiza obejmuje okres od 2001 do 2004 roku, prezentując dane miesięczne. Dzięki temu możliwe jest badanie sezonowości oraz identyfikacja trendów sprzedażowych. Poszczególne elementy procesu zostały przedstawione poniżej(Rysunek 27). Obraz zawierający linia, zrzut ekranu, diagram, Wykres

Opis wygenerowany automatycznie *Rysunek 27.*  
Pierwszy węzeł JOIN(Rysunek 28) ma na celu połączenie ze sobą faktu sprzedaży i wymiaru daty za pomocą klucza DateKey(Rysunek 29). Obraz zawierający tekst, zrzut ekranu, linia, diagram

Opis wygenerowany automatycznie  
*Rysunek 28.* *Rysunek 29.*Następnie w klauzuli SELECT tworzę nową kolumnę TotalSales(Rysunek 30) która sumuje wszystkie zyski ze sprzedaży osiągnięte w poszczególnych okresach. Używam do tego funkcji SUM oraz w późniejszej klauzuli polecenia grupowania.   
*Rysunek 30.*  
Grupuję oraz sortuję dane ze względu na ID produktu, rok oraz miesiąc(Rysunek 31).   
Obraz zawierający tekst, Czcionka, zrzut ekranu, linia

Opis wygenerowany automatycznie  
*Rysunek 31.*

W następnym węźle JOIN do już istniejącej tabeli skumulowanych zysków dołączam wymiar produktu który pozwoli mi na zamienienie kolumny ProductID na oficialną nazwę produktu(Rysunek 32). Obraz zawierający tekst, numer, linia, Czcionka

Opis wygenerowany automatycznie  
*Rysunek 32.*  
Dzięki temu dostaję dane gotowe do analizy sezonowości sprzedaży które prezentują się następująco(Rysunek 33):  
 Obraz zawierający tekst, zrzut ekranu, numer, Czcionka

Opis wygenerowany automatycznie  
*Rysunek 33.*

# Proces 2

Proces ten ma na celu przygotowanie danych do analizy, która pomoże zrozumieć jak różne kampanie promocyjne wpływają na sprzedaż w poszczególnych regionach. Poprzez porównanie danych sprzedażowych możliwe jest wyodrębnienie efektu promocji od innych czynników wpływających na sprzedaż. W analizie uwzględniono dane sprzedażowe z ostatnich kilku lat, z podziałem na poszczególne regiony. Każdy region analizowany jest osobno, co pozwala na identyfikację regionów, w których kampanie promocyjne są najbardziej efektywne, oraz tych, gdzie promocje mają ograniczony wpływ na wzrost sprzedaży. Obraz zawierający zrzut ekranu, linia, diagram, Wykres

Opis wygenerowany automatycznie  
*Rysunek 34. Proces ETL przygotowujący dane do drugiej analizy*  
Węzeł JOIN zaczyna od połączenia tabeli faktu sprzedaży z wymiarami daty, regionu oraz rabatu(Rysunek 36). Obraz zawierający tekst, diagram, zrzut ekranu, Plan

Opis wygenerowany automatycznie  
*Rysunek 35. Węzeł JOIN*Obraz zawierający tekst, zrzut ekranu, Czcionka, linia

Opis wygenerowany automatycznie *Rysunek 36.*

Następnie w klauzuli SELECT(Rysunek 37) dokonuję wyboru interesujących mnie kolumn którymi są nazwa regionu, rok, miesiąc oraz tworzę kolumny SalesDuringPromotion i SalesWithoutPromotion które sumują łączną wartość sprzedaży w zależności od występowania promocji lub jej braku. Obraz zawierający tekst, linia, Czcionka, numer

Opis wygenerowany automatycznie  
*Rysunek 37.*  
Użyłem do tego formuły: sum(CASE WHEN SpecialOffer\_dim."SpecialOfferID"n <> 1 THEN Sales\_fact."LineTotal"n ELSE 0 END). Oznacza ona że w sytuacjach kiedy ID rabatu jest inne niż 1(jedynka w bazie oznacza brak promocji) funkcja sumuje wartość sprzedaży, w przeciwnym przypadku suma sprzedaży trafia do drugiej kolumny. Następnie klauzula GROUP BY grupuje dane ze względu na nazwę regionu, rok i miesiąc(Rysunek 38). Obraz zawierający tekst, zrzut ekranu, Czcionka, linia

Opis wygenerowany automatycznie  
*Rysunek 38.*  
  
  
  
  
  
  
  
  
  
W wyniku czego po przeładowaniu tabeli za pomocą węzła TABLE LOADER otrzymuję tabele wynikową(Rysunek 39).

Obraz zawierający tekst, zrzut ekranu, numer, Czcionka

Opis wygenerowany automatycznie  
*Rysunek 39.*

# Proces 3

Proces ten ma na celu przygotowanie danych do analizy rentowności produktów oferowanych przez firmę. Skupia się on na obliczeniu marży która jest określana jako stosunek różnicy między ceną sprzedaży a kosztem produkcji.

**Obraz zawierający linia, zrzut ekranu, diagram, Czcionka

Opis wygenerowany automatycznie**  
*Rysunek 40. Proces ETL przygotowujący dane do trzeciej analizy*Pierwszy w kolejności węzeł CREATE TABLE(Rysunek 41) redukuje ilość kolumn wprowadzanych do węzła JOIN oraz modyfikuje nazwe kolumny ID produktu aby zapobiec konfliktowi nazw występującemu w węźle JOIN. Obraz zawierający tekst, zrzut ekranu, Czcionka, numer

Opis wygenerowany automatycznie  
*Rysunek 41.*  
Następnie węzeł JOIN łączy ze soba tabele kluczem ProductID. W klauzuli SELECT wybieram kolumny IDProduktu, jego nazwę oraz kategorię. Tworzę kolumny TotalRevenue która sumuję łączną sprzedaż, TotalCost która sumuję łączne koszty wyprodukowania danych produktów(iloczyn kosztu produkcji i ilości produktów w zamówieniu) oraz GrossMarginPercentage w którym obliczana jest procentowa marża, użyłem do tego formuły:  
(sum(W827AFDQ."LineTotal"n) -sum(Product\_dim."StandardCost"n\*W827AFDQ."OrderQty"n))  
/ sum(W827AFDQ."LineTotal"n ) \* 100  
oblicza ona różnicę całkowitej sprzedaży danego produktu i kosztów jego wyprodukowania a następnie dzieli przez wartośc sprzedaży i mnoży razy sto aby osiągnąć wynik procentowy(Rysunek 42). Obraz zawierający tekst, zrzut ekranu, Czcionka, linia

Opis wygenerowany automatycznie  
*Rysunek 42.*  
Dane te są grupowane ze względu na nazwę i kategorię oraz sortowane malejąco pod względem marży, dzięki czemu otrzymuję tabelę wyglądającą następująco(Rysunek 43): Obraz zawierający tekst, zrzut ekranu, Czcionka, numer

Opis wygenerowany automatycznie  
*Rysunek 43.*

Kolejnym węzłem jest SPLITER który rozdziela produkty ze względu na ich kategorię czyli na: akcesoria, ubrania, rowery oraz komponenty. Umożliwia to analizę poszczególnych grup produktów. Węzeł ten również usuwa kolumnę IDProduktu i kategorię dzięki czemu przykładowo tabela wynikowa dla kategorii rowerów wygląda następująco(Rysunek 44): Obraz zawierający tekst, zrzut ekranu, numer, Czcionka

Opis wygenerowany automatycznie  
*Rysunek 44.*

# Proces 4

Celem tego procesu jest zrozumienie dynamiki sprzedaży oraz zwrotów w różnych regionach, na które firma jest podzielona. Analiza koncentruje się na porównaniu, które terytoria generują najwięcej przychodów oraz które charakteryzują się największym odsetkiem zwrotów. Dzięki tej analizie możliwe jest zidentyfikowanie obszarów, które mogą wymagać działań naprawczych w strategiach sprzedaży i logistyki. Obraz zawierający zrzut ekranu, linia, diagram

Opis wygenerowany automatycznie  
*Rysunek 45. Proces ETL tworzący dane do analizy czwartej*  
Pierwszy węzeł JOIN łączy ze sobą tabelę faktów oraz tabelę wymiaru produktu w celu obliczenia łącznych zysków i kosztów, odbywa się to w klauzuli SELECT gdzie sumowana jest łaczna wartość sprzedaży(TotalSales) oraz sumowany jest iloraz jednostkowego kosztu produkcji i ilości zamówionych towarów(TotalCosts)(Rysunek 46).

Obraz zawierający tekst, zrzut ekranu, linia, Czcionka

Opis wygenerowany automatycznie  
*Rysunek 46.*  
Dane następnie grupuję według terytorium, co pozwala na agregację sprzedaży i kosztów w każdym regionie. Otrzymaną tabele łączę z wymiarem regionu na podstawie klucza TerritoryID aby dołączyć odpowiednie informacje o terytorium sprzedaży(Rysunek 47).  
  
*Rysunek 47.*  
  
  
  
  
  
  
  
  
W klauzuli select tworze dodatkową kolumnę TotalReturns(Rysunek 48) która oblicza różnice między wartością sprzedaży(TotalSales) a łącznym kosztem produkcji(TotalCosts). Jest użyteczna do obliczenia marży z każdego terytorium. Obraz zawierający tekst, zrzut ekranu, linia, Czcionka

Opis wygenerowany automatycznie   
*Rysunek 48.*Otrzymana tabela wynikowa prezentuje się następująco(Rysunek 49):Obraz zawierający tekst, zrzut ekranu, Czcionka, numer

Opis wygenerowany automatycznie  
*Rysunek 49.*

# 

# Proces 5

Proces ten ma na celu przygotowanie danych do analizy wpływu promocji na sprzedaż kategorii produktów w czasie. Analiza opiera się na zliczaniu transakcji oraz sumowaniu sprzedaży produktów objętych promocjami oraz tych sprzedawanych bez promocji. Dzięki temu zapytaniu firma może lepiej zrozumieć, które kategorie produktów są wrażliwe na działania promocyjne. Analiza ta pozwala również na ocenę, jak promocje wpływają na sprzedaż w poszczególnych okresach roku, co jest istotne przy planowaniu przyszłych kampanii marketingowych. Obraz zawierający zrzut ekranu, linia, Czcionka, diagram

Opis wygenerowany automatycznie  
*Rysunek 50. Proces ETL tworzący dane do analizy piątej*

Pierwszy węzeł CREATE TABLE odpowiada za zmniejszenie wolumenu dołączanej tabeli oraz zapobiega konfliktowi nazw. Wybiera on z wymiaru produktu wyłącznie IDproduktu oraz nazwę kategorii(Rysunek 51).  
 Obraz zawierający tekst, zrzut ekranu, Czcionka, linia

Opis wygenerowany automatycznie  
*Rysunek 51.*

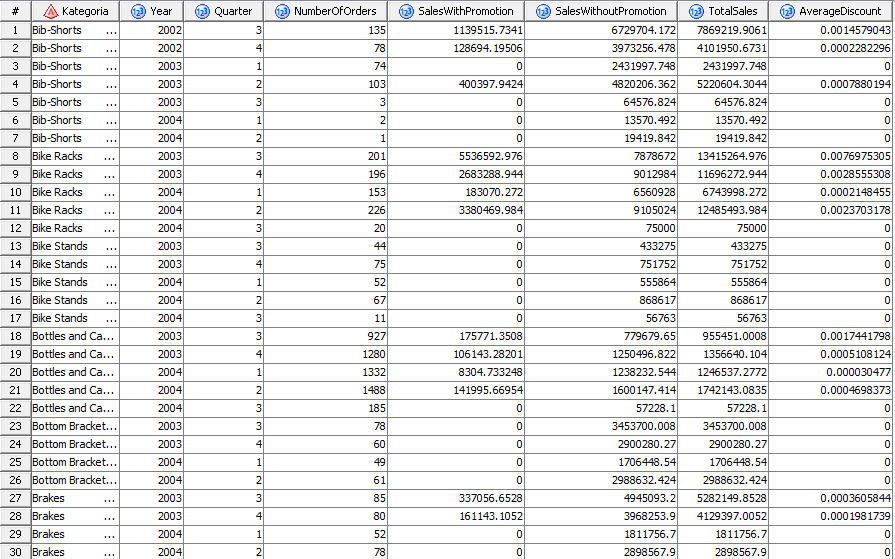
W węźle JOIN(Rysunek 52) następuje złączenie ze sobą 4 tabel za pomocą ich kluczy głównych. Obraz zawierający tekst, diagram, zrzut ekranu, Plan

Opis wygenerowany automatycznie  
*Rysunek 52.*

Klauzula SELECT odpowiada za wybranie kolumn Kategoria, Rok i Kwartał oraz za utworzenie NumberOfOrders która zlicza ilość zamówień na dany produkt, SalesWithPromotion która sumuje wartość sprzedaży produktów objętych promocją, wykonuje to formuła: SUM(CASE WHEN SpecialOffer\_dim."SpecialOfferID"n <> 1 then Sales\_fact."LineTotal"n else 0 end). Tworzone jest również SalesWithoutPromotion które sumuje sprzedaż nieobjętą promocjami oraz TotalSales sumująca całkowitą sprzedaż w danym okresie jak i AverageDiscount które pokazuje średnią zniżkę na daną kategorię produktu w danym czasie(Rysunek 53). Obraz zawierający tekst, linia, Czcionka, numer

Opis wygenerowany automatycznie  
*Rysunek 53.*

Dane są następnie grupowane ze względu na kategorię oraz rok i kwartał, sortowane są w taki sam sposób. Tak przygotowana do przyszłych analiz tabela prezentuje się następująco(Rysunek 54):

  
*Rysunek 54.*

# Raport uzyskanych wyników

Uzyskane rezultaty pozwoliły na wstępną ocenę przez decydentów wyników sprzedaży, przeprowadzonych promocji oraz jak kształtuje się popyt na różnych rynkach. Schemat gwiazdy utworzony dla tych danych daje podstawę do wielu analiz biznesowych.

Pierwszy proces ukazał kształtowanie się całkowitej sprzedaży danego produktu miesiąc po miesiącu, co pozwala na analizę sezonowości. Ponieważ dane obejmują sprzedaż miesięczną produktów przez kilka lat, można zbadać wzorce sezonowe. Przykładowo, możliwa jest weryfikacja, czy w określonych miesiącach sprzedaż wzrasta, co może sugerować większe zapotrzebowanie na produkty w tych okresach.

Dodatkowo, przeprowadzenie analizy trendów pozwoli zobaczyć, czy ogólna sprzedaż produktów rośnie, maleje czy też utrzymuje się na stabilnym poziomie w kolejnych latach. Taka analiza jest wspierana przez raporty porównujące wyniki sprzedaży z jednego roku do innego, pokazujące wzrost lub spadek sprzedaży. Pozwoli to zidentyfikować ewentualne zmiany na rynku lub w preferencjach klientów.

Drugi proces ma na celu przygotowanie danych do analizy, która pomoże zrozumieć, jak różne kampanie promocyjne wpływają na sprzedaż w poszczególnych regionach. Poprzez porównanie danych sprzedażowych możliwe jest wyodrębnienie efektu promocji od innych czynników wpływających na sprzedaż. W analizie uwzględniono dane sprzedażowe z ostatnich kilku lat, z podziałem na poszczególne regiony, co pozwala na identyfikację regionów, w których kampanie promocyjne są najbardziej efektywne, oraz tych, gdzie promocje mają ograniczony wpływ na wzrost sprzedaży. Możliwa do przeprowadzenia jest analiza różnicy w sprzedaży, aby zrozumieć, jaki wpływ mają kampanie promocyjne na sprzedaż

Proces trzeci jest skierowany na przygotowanie danych do analizy rentowności produktów oferowanych przez firmę. Skoncentrowany jest na obliczeniu marży co pozwala na ocenę, które produkty generują największe zyski. W ramach analizy produktów, skupienie się na identyfikacji produktów z najwyższą i najniższą marżą brutto jest kluczowe. Pozwala to zrozumieć, które produkty są najbardziej rentowne oraz które przynoszą najmniejsze korzyści finansowe, co jest niezbędne dla strategicznego planowania przyszłych cen, produkcji, i działań promocyjnych.

Dodatkowo, porównując różne modele i rozmiary, można zidentyfikować, które specyfikacje przynoszą największe korzyści finansowe. Taka analiza może również ujawnić potencjalne obszary do poprawy w projektowaniu lub kosztach produkcji

Celem czwartego procesu jest zrozumienie dynamiki sprzedaży oraz zwrotów w różnych regionach, na które firma jest podzielona. Analiza skupia się na porównaniu, które terytoria generują najwięcej przychodów oraz które charakteryzują się największym odsetkiem zwrotów. Dzięki tej analizie możliwe jest zidentyfikowanie obszarów wymagających działań naprawczych w strategiach sprzedaży i logistyki.

Ostatni proces skupia się na przygotowaniu danych do analizy wpływu promocji na sprzedaż różnych kategorii produktów w czasie. Analiza opiera się na zliczaniu liczby transakcji oraz sumowaniu sprzedaży produktów objętych promocjami i tych sprzedawanych bez promocji. Dzięki temu firma może lepiej zrozumieć, które kategorie produktów są wrażliwe na działania promocyjne. Analiza ta jest również cenna w ocenie, jak promocje wpływają na sprzedaż w poszczególnych okresach roku, co ma kluczowe znaczenie przy planowaniu przyszłych kampanii marketingowych. W ramach analizy sprzedaży z promocją vs. bez promocji, możliwe jest zbadanie dla każdej kategorii produktu i każdego kwartału, jak sprzedaż z promocją porównuje się do sprzedaży bez promocji. To pomoże zrozumieć, które kategorie produktów są najbardziej responsywne na promocje i w jakim stopniu. Dodatkowo, analiza wpływu promocji na ilość zamówień pozwoli sprawdzić, jak promocje wpływają na liczbę zamówień w każdej kategorii produktów.