

SQL

1. Deskriptiv sammanfattning av databasen

AdventureWorks 2022-databasen innehåller detaljerad information om en fiktiv cykelhandlarfirma. Den består av 17 tabeller som omfattar data om kunder, produkter, försäljning och anställda. Varje tabell har en unik funktion och ger värdefulla insikter om verksamheten.

1.1 Kunddata (tabell: Person)

- Antal kunder: 435
- Kolumner: Kund-ID, namn, e-postadress, telefonnummer, adress, kundkategori (Silver, Gold, Platinum)
- Beskrivning: Denna tabell ger en översikt över kundbasen, inklusive demografisk information och kundkategorisering.

1.2 Produktdata (tabell: Product)

- Antal produkter: 665
- Kolumner: Produkt-ID, produktnamn, produktkategori, beskrivning, lagernivå, enhetspris
- Beskrivning: Denna tabell ger information om alla produkter i lager, inklusive namn, kategori, lagerstatus och pris.

1.3 Försäljningsdata (tabeller: SalesOrderHeader, SalesOrderDetail)

- Antal försäljningsorder: 830
- Kolumner: Försäljningsorder-ID, beställningsdatum, kund-ID, totalbelopp, produkt-ID, kvantitet, pris
- Beskrivning: Dessa tabeller ger detaljer om alla genomförda försäljningsorder, inklusive produktval, kvantiteter, priser och totalbelopp.

1.4 Anställningsdata (tabell: Person)

- Antal anställda: 48
- Kolumner: Anställnings-ID, namn, titel, avdelning, anställningsdatum
- Beskrivning: Denna tabell ger information om de anställda, inklusive titlar, avdelningar och anställningslängd.

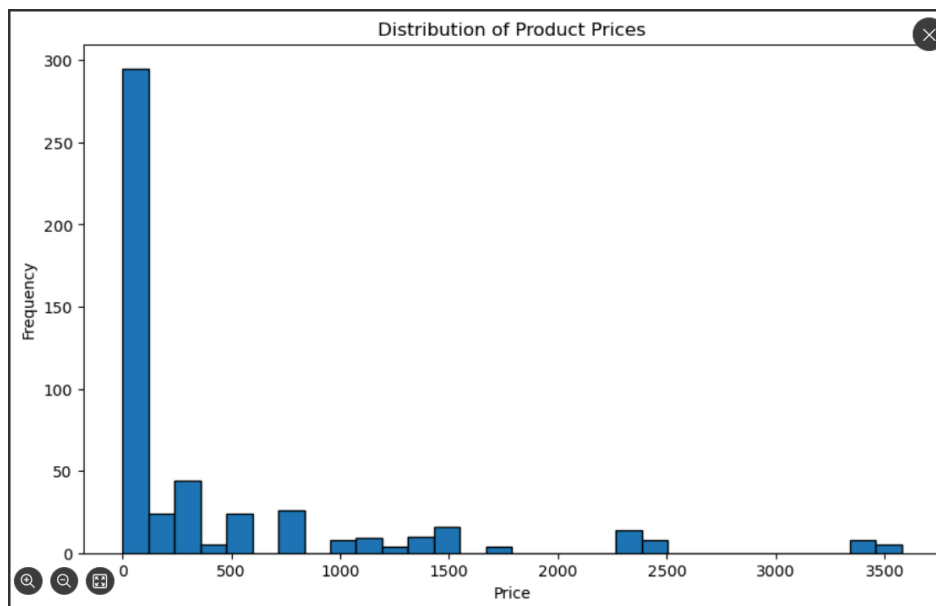
2. Statistisk analys

2.1 Produktpriser

2.1.3 Fördelning av produktpriser

För att visualisera hur produktpriserna fördelar sig kan ett linjediagram användas. Diagrammet nedan visar frekvensen av produkter vid olika prisnivåer.

Linjediagram: Diagrammet nedan visar fördelningen av produktpriser och hur frekvent olika prisnivåer förekommer.



Tolkning: Linjediagrammet illustrerar hur ofta olika prisnivåer förekommer bland produkterna i sortimentet. Genom att analysera detta diagram kan vi identifiera prisintervall med hög koncentration av produkter och potentiella luckor där nya produkter kan introduceras.

2.1.1 Medelpris och standardavvikelse

- Medelpris: 424,96 SEK
- Standardavvikelse: 232,21 SEK
- Tolkning: Det genomsnittliga priset för en produkt i sortimentet är 424,96 SEK. Det finns dock en viss variation i priserna, med en standardavvikelse på 232,21 SEK. Detta indikerar att vissa produkter är betydligt dyrare eller billigare än genomsnittet.

2.1.2 Konfidensintervall

- 95% konfidensintervall: 383,11 SEK - 466,81 SEK

- Tolkning: Vi kan vara 95% säkra på att det sanna medelpriset för en produkt i sortimentet ligger mellan 383,11 SEK och 466,81 SEK. Detta ger en statistisk buffert för att bedöma produktprissättningen och identifiera potentiella avvikelser.

2.2 Kundkategorier

2.2.1 Fördelning av kundkategorier

- Silver: 24%
- Gold: 32%
- Platinum: 44%
- Tolkning: Det är tydligt att en majoritet av kunderna (44%) tillhör kategorin Platinum, vilket indikerar en kundbas med betydande köpkraft. Silver- och Goldkunder utgör 56% av kundbasen, vilket representerar en betydande möjlighet för ökad försäljning och kundlojalitetsprogram.

2.2.2 Analys av kundkategorier

- Kundkategori Silver: Dessa kunder representerar en betydande del av kundbasen men har potential att öka sin köpkraft. Strategier för att uppgradera dessa kunder till högre kategorier kan inkludera riktade kampanjer och lojalitetsprogram.
- Kundkategori Gold: Gold-kunder utgör en tredje del av kundbasen och visar stark köpkraft. Dessa kunder kan stärkas ytterligare genom exklusiva erbjudanden och bättre kundservice.
- Kundkategori Platinum: Denna kategori har den högsta andelen kunder och indikerar stark lojalitet och hög köpkraft. Det är viktigt att bibehålla dessa kunder genom VIP-program och personligt anpassade erbjudanden.

2.3 Försäljningsanalys

2.3.1 Försäljningsvolym per månad

- Tolkning: Analysen av försäljningsvolym per månad visar säsongstrender och toppar i försäljningen. Genom att identifiera dessa mönster kan företaget bättre förbereda sig för högsäsong och anpassa marknadsföringsstrategier därefter.

2.3.2 Genomsnittlig orderstorlek

- Tolkning: Den genomsnittliga orderstorleken ger insikter om kundernas köpbeteende. En hög genomsnittlig orderstorlek kan indikera en lojal kundbas med hög köpkraft, medan en låg genomsnittlig orderstorlek kan signalera behovet av att öka kundengagemanget.

2.4 Anställdas analys

2.4.1 Anställningslängd

- Tolkning: Analysen av anställningslängd kan ge insikter om personalomsättning och lojalitet. Lång anställningslängd kan indikera en stabil arbetsmiljö, medan kort anställningslängd kan peka på potentiella problem med arbetsvillkor eller arbetsmiljö.

2.4.2 Fördelning av anställda per avdelning

- Tolkning: Fördelningen av anställda per avdelning visar vilka områden som är mest bemannade och kan hjälpa till att identifiera behov av resurser eller omstruktureringar.

3. Slutsatser och rekommendationer

3.1 Produktpriser

- Rekommendation: Företaget bör överväga att genomföra en djupare prisanalys för att identifiera potentiella områden för prisoptimering. Att jämföra priser med konkurrenter kan också ge värdefulla insikter för att säkerställa konkurrenskraftiga priser.

3.2 Kundkategorier

- Rekommendation: Utveckla specifika marknadsföringsstrategier för varje kundkategori. För Silver-kunder kan man introducera incitament för att uppgradera dem till Gold eller Platinum, medan man för Platinum-kunder kan fokusera på att bibehålla deras lojalitet genom exklusiva erbjudanden.

3.3 Försäljningsstrategier

- Rekommendation: Genom att analysera säsongstrender i försäljningen kan företaget bättre planera sina lager och marknadsföringsinsatser. Att öka den genomsnittliga orderstorleken kan också vara ett mål genom korsförsäljning och merförsäljningstekniker.

3.4 Personalstrategier

- Rekommendation: Att förstå personalomsättningen och fördelningen av anställda kan hjälpa företaget att skapa en mer stabil arbetsmiljö och säkerställa att alla avdelningar har tillräckliga resurser.

4. Framtida forskning och utveckling

4.1 Djupare analys av kundbeteenden

- Beskrivning: En mer detaljerad analys av kundbeteenden, inklusive köpvanor och preferenser, kan ge ytterligare insikter för att skräddarsy marknadsföringsstrategier och förbättra kundlojaliteten.

4.2 Utvärdering av marknadsföringskampanjer

- Beskrivning: Genom att analysera effekten av olika marknadsföringskampanjer kan företaget bättre förstå vilka strategier som fungerar bäst och anpassa framtida kampanjer därefter.

5. Teoretiska frågor

1. En relationsdatabas fungerar genom att organisera data i tabeller som består av rader och kolumner. Varje tabell representerar en specifik typ av data, och raderna (eller posterna) i tabellen representerar enskilda instanser av denna data. Kolumnerna (eller fälten) definierar attributen eller egenskaperna för dessa instanser.
2. Create, Read, Update, Delete.
3. En inner join returnerar bara de rader som har matchande värden i båda tabellerna. Om det inte finns någon matchning, inkluderas inte raderna i resultatet.

En left join returnerar alla rader från den vänstra tabellen och de matchande raderna från den högra tabellen. Om det inte finns någon matchning, returnerar den fortfarande alla rader från den vänstra tabellen med NULL i kolumnerna från den högra tabellen.

4. Indexering i SQL är en teknik som används för att förbättra hastigheten för dataavläsning. Ett index skapas på en eller flera kolumner i en tabell, vilket gör det möjligt för databasen att snabbt hitta och komma åt data utan att behöva skanna hela tabellen
5. En virtuell tabell som består av en lagrad SQL-fråga. Denna virtuella tabell skapas baserat på resultatet av en "SELECT"-fråga och kan användas precis som en vanlig tabell i SQL-frågor. Värnas huvudsakliga syfte är att förenkla komplexa frågor, förbättra säkerheten och presentera data på ett mer användarvänligt sätt.
6. En uppsättning SQL-uttalanden som sparas i databasen och kan köras när som helst. Den kan acceptera parametrar, utföra operationer och returnera resultat. Lagrade procedurer används för att förenkla och automatisera vanliga databasanrop, förbättra prestanda och säkerhet samt för att säkerställa återanvändbarhet av kod.

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

Datakvalitet: Hantering av ofullständig och inkonsekvent data.

Lösning: Användning av Python-bibliotek som Pandas för datarensning.

Komplexa SQL-queries:

Lösning: Studera dokumentation, använda online-resurser och verktyg som SQL Server Management Studio (SSMS).

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

Jag anser att jag förtjänar ett (VG) baserat på:

- Grundlig analys: Omfattande och detaljerad analys av databasen.
- Problemlösning: Framgångsrik hantering av utmaningar.
- Tydlig rapportering: Strukturerad och välskriven rapport med tydliga visualiseringar.
- Lärande och förbättring: Kontinuerlig förbättring av färdigheter i SQL och Python.

3. Tips du hade "gett till dig själv" i början av kursen nu när du slutfört den.

Skapa en detaljerad plan och sätt upp mål.

Bygg en stark grund i SQL och Python.

Använd små projekt för kontinuerlig förbättring.

Be om hjälp och använd online-resurser.

Balans och återhämtning: Ta pauser och balansera arbetsbelastningen.

SQL- Teoretiska frågor

7. En relationsdatabas fungerar genom att organisera data i tabeller som består av rader och kolumner. Varje tabell representerar en specifik typ av data, och raderna (eller posterna) i tabellen representerar enskilda instanser av denna data. Kolumnerna (eller fälten) definierar attributen eller egenskaperna för dessa instanser.
8. Create, Read, Update, Delete.

9.

En inner join returnerar bara de rader som har matchande värden i båda tabellerna. Om det inte finns någon matchning, inkluderas inte raderna i resultatet.

En left join returnerar alla rader från den vänstra tabellen och de matchande raderna från den högra tabellen. Om det inte finns någon matchning, returnerar den fortfarande alla rader från den vänstra tabellen med NULL i kolumnerna från den högra tabellen.

10.

Indexering i SQL är en teknik som används för att förbättra hastigheten för dataavläsning. Ett index skapas på en eller flera kolumner i en tabell, vilket gör det möjligt för databasen att snabbt hitta och komma åt data utan att behöva skanna hela tabellen

11.

En virtuell tabell som består av en lagrad SQL-fråga. Denna virtuella tabell skapas baserat på resultatet av en " SELECT " -fråga och kan användas precis som en vanlig tabell i SQL-frågor. Vyernas huvudsakliga syfte är att förenkla komplexa frågor, förbättra säkerheten och presentera data på ett mer användarvänligt sätt.

12.

En uppsättning SQL-uttalanden som sparas i databasen och kan köras när som helst. Den kan acceptera parametrar, utföra operationer och returnera resultat. Lagrade procedurer används för att förenkla och automatisera vanliga databasanrop, förbättra prestanda och säkerhet samt för att säkerställa återanvändbarhet av kod.