

Teoretiska Frågor

1. Beskriv kort hur en relationsdatabas fungerar.
2. Vad menas med "CRUD" flödet?
3. Beskriv kort vad en "left join" och "inner join" är. Varför använder man det?
4. Beskriv kort vad indexering i SQL innebär.
5. Beskriv kort vad en vy i SQL är.
6. Beskriv kort vad en lagrad procedur i SQL är.

===== Svaren är på nästa sida! =====

Svar

1. En **relationsdatabas** organiserar data i tabeller där varje tabell innehåller rader och kolumner. Dessa tabeller är sammankopplade genom gemensamma nycklar, vilket möjliggör relationer mellan olika datamängder. Genom att använda SQL (Structured Query Language) kan du skapa, läsa, uppdatera och ta bort data inom dessa tabeller.
2. **CRUD-flödet** står för **Create**, **Read**, **Update** och **Delete** och representerar de grundläggande operationerna för att hantera data: skapa, läsa, uppdatera och radera.
3. **"Left join"** kombinerar alla rader från den vänstra tabellen med matchande rader från den högra tabellen. Om det inte finns någon matchande rad i den högra tabellen, fylls den med NULL-värden. **"Inner join"** returnerar endast rader som har matchande värden i båda tabellerna. Det exkluderar rader där det inte finns någon matchning.
De används för att kombinera data från olika tabeller i en databas och extrahera relevanta resultat baserat på relationer mellan tabellerna.
4. **Indexering** i SQL innebär att skapa en struktur för att snabbt hitta data i en databastabell. Det ökar sökhastigheten genom att skapa en sorterad referens till data, vilket gör frågor snabbare genom att minska söktiden.
5. En **vy** i SQL är en virtuell tabell som inte lagrar data fysiskt utan skapar en sparad fråga eller filter. Den låter användare, beroende på deras behov, se en specifik delmängd eller bearbetning av data från en eller flera tabeller som om det vore en separat tabell. Det ger en enkel och flexibel metod att visa data utan att ändra den underliggande strukturen.
6. **En lagrad procedur** i SQL är en fördefinierad uppsättning SQL-instruktioner som sparas i databasen och kan köras när det behövs. Det tillåter återanvändning av kod och kan inkludera logik och affärsregler för att utföra specifika uppgifter eller operationer på databasen.

DESKRIPTIV SAMMANFATTNING

Databasen AdventureWorks 2022 har en kombination av inköpsorderdata som relaterar till produktleverantörer, produkter och inköpsdetaljer. Den insamlade informationen inkluderar leverantörsprestanda relaterat till genomsnittliga enhetspriser, skatter, frakt, leveranstid och leverantörens kreditvärdering. Informationen är organiserad enligt inköpsorder-ID, leverantörs-ID, produkt-ID och andra relevanta attribut. Enligt den utförda SQL-frågan har varje inköpsorder-ID relaterade inköpsdetaljer och produkter associerade med specifika leverantörer och deras prestanda. Data visar genomsnittliga enhetspriser, skatter, frakt och leveranstider för inköp som gjorts från olika leverantörer för olika produkter. Det finns ett fokus på leverantörens prestanda och produkter, med betoning på genomsnittliga kostnader och prestandaattribut. Uppgifterna ger en översikt över hur olika leverantörer presterar i termer av pris, leveranstid och kreditvärderingar i samband med inköpta produkter.

ANALYS: Average Unit Price by Vendor Performance Rating & Active Flag Status

Resultaten från den statistiska analysen visar tydliga skillnader i genomsnittliga enhetspriser baserat på leverantörernas prestation och aktiva status:

1. För leverantörer med prestationen **1 och 2** visar de aktiva leverantörerna betydligt högre genomsnittliga enhetspriser jämfört med de inaktiva.
2. För prestation **3** har aktiva leverantörer också visat högre genomsnittliga enhetspriser.
3. Vid prestation **4** finns en betydande skillnad mellan inaktiva och aktiva leverantörer.
4. För prestation **5** är skillnaden i genomsnittliga enhetspriser extremt stor mellan aktiva och inaktiva leverantörer.

Den beräknade **T-statistiken på (2,7951)** med ett **P-värde på (0,0053)** visar på statistisk signifikans, vilket innebär en betydande skillnad i genomsnittliga enhetspriser mellan aktiva och inaktiva leverantörer. Detta leder till avvisande av noll hypotesen (H_0), vilket bekräftar att skillnaderna i enhetspriser är signifikanta mellan de aktiva och inaktiva leverantörerna.

ANALYS: Average Unit Price by Vendor Performance Rating & Preferred Vendor Status

Resultaten av den statistiska analysen visar signifikanta skillnader i genomsnittliga enhetspriser mellan 'Önskade' och 'Inte önskade' leverantörer baserat på deras 'Leverantörsprestation'. För de leverantörer som rankats som **'1'** visar intervallet för **'Önskade'** (**36.64 - 38.63**) lägre enhetspriser jämfört med **'Inte önskade'** (**39.57 - 41.96**). Likaså för rankningen **'2'** ligger intervallet för **'Önskade'** (**23.00 - 37.12**) betydligt lägre än **'Inte önskade'** (**41.77 - 44.89**).

Baserat på **T-statistiken (-3,9173)** och det låga **P-värdet (0,0001)** avvisas noll hypotesen, vilket indikerar en betydande skillnad i genomsnittliga enhetspriser mellan de två grupperna.

SLUTSATSER

Leverantörsprestation och aktiv status påverkar enhetspriser: Aktiva leverantörer, särskilt vid prestation 1 till 4, visar högre genomsnittliga enhetspriser jämfört med inaktiva. För prestation 5 är skillnaden extremt stor.

Önskad status indikerar lägre enhetspriser: 'Önskade' leverantörer tenderar att erbjuda lägre enhetspriser vid prestation 1 och 2 jämfört med 'Inte önskade'. En liknande trend observeras för prestation 2, men prisintervallet för 'Inte önskade' är högre.

Förbättra relationer med 'Inte önskade' vid hög prestation: Trots högre priser, om prestationen är hög hos 'Inte önskade', överväg att bygga starkare relationer för att minska prisgapet.

REKOMMENDATIONER

Fokusera på att främja aktivt samarbete med högpresterande 'Inte önskade' leverantörer för att möjliggöra förbättrade relationer och potentiellt minska prisgapet mellan 'Önskade' och 'Inte önskade'. Detta kan bidra till att förbättra övergripande prestation och kostnadsbesparingar för företaget.

”EXECUTIVE SUMMARY”

Den utförda analysen tydliggör betydande skillnader i genomsnittliga enhetspriser baserat på både leverantörernas prestation och deras aktivitetsstatus. Aktiva leverantörer visar i allmänhet högre enhetspriser vid lägre prestationer, medan 'Önskade' leverantörer tenderar att erbjuda lägre priser vid vissa prestationsnivåer. Dessa resultat indikerar att förbättrade relationer med högpresterande 'Inte önskade' leverantörer kan minska prisgapet och potentiellt öka kostnadsbesparingar för företaget. Detta ger en öppning för att optimera leverantörsrelationer och potentiellt förbättra övergripande prestanda och kostnadsstrukturer för företaget, vilket kan vara av strategisk betydelse för långsiktig framgång.

DATUM FÖR MUNTLLIG PRESENTATION AV KODEN

Januari 5, 2024

REDOGÖRELSE

Under analysarbetet stötte jag på utmaningar med datakvalitet och ojämnheter i datasetet. För att tackla detta använde jag noggrann filtrering och rensning av datamängden för att säkerställa endast relevanta och pålitliga data användes. Jag tillämpade olika statistiska metoder för att hantera variationen och säkerställa tillförlitliga resultat trots datans ojämnheter. För att hantera okända termer i AdventureWorks-databasen genomförde jag grundlig forskning online för att förstå dess struktur och relevans för analysen. Genom att aktivt söka förklaringar kunde jag bättre relatera termer till analysen och kontextualisera dem för att förstå deras påverkan på resultaten. Detta tillvägagångssätt bidrog till en mer robust och relevant analys med trovärdiga insikter och slutsatser.

BETYG

G. Eftersom genom noggrann datafiltrering och användning av varierande statistiska metoder lyckades jag hantera utmaningar med datakvalitet och säkerställa en robust analys. Min aktiva sökning och research om okända termer visar min förmåga att förstå och kontextualisera relevanta begrepp för att bidra till en mer fullständig och korrekt analys.

TIPS

Att sätta av tid i början för att förstå databasstrukturen och dess termer kan underlätta senare för analysen. En stark grund inom statistik och datarensning ger en bättre förståelse och ökad komfort i analysprocessen.