

سوال ۱

مهمترین تفاوت این دو نوع پایگاه داده در نوع داده های قابل ذخیره در آنهاست. در پایگاه SQL داده ساختارمند است و صفات مشخصی را داراست. این پایگاه داده عمدتاً متشکل از یک یا چند جدول است. مثلاً جدولی را در نظر بگیرید که به منظور ذخیره سازی سفارشات مشتریان ایجاد شده است؛ این داده ها دارای صفات مشخصی مثل تاریخ، قیمت سفارش، تعداد و... می باشند و همیشه از این فورمت پیروی می کنند. در نقطه مقابل داده هایی غیرساختارمند را داریم که در پایگاه های مبتنی بر NoSQL ذخیره می شوند. به عنوان نمونه اطلاعات ترافیکی (مبتنی بر گراف)، اطلاعات ژنتیکی و پزشکی (توالی آمینواسیدها در ساختار یک پروتئین) یا حتی مثال ساده تر، یک لغتنامه که به صورت کلید-مقدار ذخیره سازی می شود.

لازم به ذکر است که NoSQL را میتوان برای داده های ساختارمند نیز استفاده کرد که البته بهینه نیست.

سوال ۲

کوئری مورد نظر در دیالکت PostgreSQL و MySQL (کامنت شده) به صورت زیر است:

```
SELECT
    products.name AS Product_Name,
    sum(num) AS Total_Ordered,
    sum(products.price * num) AS Total_Sold
FROM
    orders LEFT JOIN products ON orders.product_id = products.id
WHERE order_date >= (NOW() - INTERVAL '1 month') - MYSQL: WHERE order_date
    >= DATEADD(month, -1, GETDATE()) AND order_date < GETDATE()
GROUP BY name;
```

نکته ۱: اگر به جای تعداد کل سفارشات، تعداد کل اردرها مد نظر بود، SUM(num) با COUNT(num) جایگزین میکنیم.

نکته ۲: میزان فروش = تعداد سفارش * قیمت (به صورت جمع کل روی جدول)

سوال ۳

با فرض بر تمیز، valid و تکمیل بودن داده ها:

در گام صفر، باید دید آیا در زنجیره تامین محصول مشکلی وجود دارد یا خیر. دلایلی مثل تحریم، کمبود تولید و... در گام اول و قبل از شروع کار با داده ها، دلایل فصلی، اخبار و یا اتفاقات ناگهانی را بررسی میکنم. مثلاً در بازه های زمانی خاصی (مثلاً ماه محرم، صفر و یا رمضان) الگو و میزان فروش دسته های مشخصی از محصولات (آرایشی بهداشتی یا مواد خوراکی) دستخوش تغییر می شود. پس بررسی پارامترهای فصلی/محیطی از این قبیل بسیار مهم است.

در گام دوم، میزان فروش محصول را در پنجره ها و بازه های زمانی دیگر نیز بررسی میکنم. مثلاً ماهانه یا هفتگی (چند هفته قبل). شاید ترند بزرگتری در این کاهش ناگهانی فروش وجود داشته باشد. شاید افزایش قیمت کالا در چندماه اخیر باعث کاهش فروش آن شده است. به طور کلی در این مرحله چند پارامتر را باید در این پنجره ها بررسی کرد. ممکن است کاهش فروش به دلیل کاهش عرضه محصول برند A باشد که بین مشتریان محبوب تر از برند B است. تعداد کالاهای مرجوعی این محصول چقدر است؟ مهم است که میزان فروش کل خانواده این محصول را بررسی کنیم. مثلاً ممکن است کاهش ناگهانی در فروش سیگار به این علت باشد که مشتریان اقبال بیشتری به خرید دستگاه های Vape داشته اند. هر دو محصول در خانواده محصولات دخانی هستند.

در گام سوم اطلاعات را به سمت مشخصات مشتریان می بریم: شهر محل سکونت، جنسیت، گروه سنی. هدف یافتن این پاسخ است که آیا این کاهش فروش به دلیل روی گردانی گروه خاصی از مشتریان است یا اینکه کل فروش به صورت همگن کاهش یافته است. در صورتی که مورد دوم صحیح باشد، یک اشکال جدی در مجموعه ما وجود دارد.

لازم به ذکر است که در مراحل ۲ و ۳ خروجی به صورت Visual از کوئری ها باید تهیه شود.

در گام آخر باید به طور دقیق داده های اجتماعی (نظرات روی سایت، اینستاگرام و...) را با کمک مدلهای یادگیری عمیق بررسی کنیم. آیا گروهی به طور سازمان یافته در حال انجام تبلیغات منفی هستند؟ یا اینکه مشکل جدی در کیفیت این محصول وجود دارد؟ شاید ناوگان ترابری محصولات در ارسال آنها دقت کافی نداشته و محصولات با آثار ضربات جزیی و... به دست مشتری میرسند؟ پاسخ این قبیل سوالات را میتوان در نظرات یافت.

در گام بعدی در صورت در دسترس بودن اطلاعات اجتماعی بازدهی تبلیغات خود و رقبا را نیز باید تحت بررسی قرار داد. مثلاً شاید به این علت که فروشگاه رقیب جشنواره X را برگزار کرده و روی همه محصولات 20% تخفیف میدهد، تعداد ویوها و لایکهای آن در شبکه های اجتماعی به صورت Spiking زیاد شده و بالتبع فروش آن نیز بالا رفته است؛ که باعث کاهش فروش ما شده. نکته مهم: تمام فرضیات احتمالی باید تحت تستهای آماری (مثل t-test) مورد ارزیابی قرار گیرند.