



۱۳۰۷

دانشگاه صنعتی خواجه نصیرالدین طوسی

تمرینات داده های حجیم سری اول

استاد: جناب آقای دکتر خواسته

دستیاران:

جناب آقای دکتر آذرکسب

مبینا صفاری

حمیده سلطانی مهر

پاییز 1402

تمرینات تئوری:

سوال ۱:

در مورد مفاهیم row store و column store توضیح دهید و مزایا و معایب هر یک را بیان کنید.

سوال ۲:

frequent items را با در نظر گرفتن $\text{minsupport} = 3$ برای آیتم ست زیر با استفاده از الگوریتم fp_growth استخراج نمایید.

(با رسم شکل مراحل را توضیح بفرمایید)

ID	Itemsets
T1	{E,K,M,N,O,Y}
T2	{D,E,K,N,O,Y}
T3	{A,E,K,M}
T4	{C,K,M,U,Y}
T5	{C,E,K,I,O,O}

تمرین پیاده سازی:

راهنمایی: برای این تمرین ابتدا mongo را از طریق لینک [نصب ویندوز](#) و [لینک](#) [نصب اوبونتو](#) نصب کنید. داده‌های مورد استفاده را از [لینک](#) دانلود کنید، سپس با استفاده از [آموزش](#) و نمونه کدهایی که در اختیار تون قرار گرفته شده سوالات زیر را در قالب کد ارایه دهید.

سوال:

الف) لیست از وضعیت کلی برای روز چهارم هفته باشگاه به ترتیب تعداد افراد.

ب) مجموع افرادی که در ماه هشتم به باشگاه می‌ایند.

پ) مجموع افرادی که در تعطیلات به باشگاه می‌ایند.

ت) پنج روز از شلوغ‌ترین روزهای باشگاه به ترتیب.

ث) میانگین دما و وضعیت تعطیلی در روزه‌های خلوت باشگاه به ترتیب تاریخ.

ج) مجموعه افرادی که در روز دوم و ماه دهم از ساعت ۱۷ تا ۲۲ به باشگاه می‌ایند.

با آرزوی تندرستی و موفقیت برای تمام دوستان