



## درس کاوش دادگان انبوه

مدرس: دکتر سامان هراتی زاده

نیم سال دوم ۹۹-۱۳۹۸

فصل چهارم کتاب

تحویل حضوری: اعلام خواهد شد

تمرین شماره ی سه

مهلت ارسال تمرین: ۱۳۹۸/۰۹/۱۶ تا

ساعت ۲۳

هدف از این تمرین، آشنایی با تحلیل داده‌های جریانی<sup>۱</sup> است.

داده مورد استفاده در این تمرین، شامل اطلاعات ثبت شده از تعداد فروش محصولات مختلف توسط فروشندگانهای یک فروشگاه در چند محدوده است. با استفاده از مجموعه داده‌ی ارسالی به همراه تمرین، موارد خواسته شده‌ی زیر را انجام دهید.

(۱) با استفاده از الگوریتم Flajolet\_Martin، تعداد فروشندگانهای منحصر به فرد مشخص شده در لیست را

تخمین بزنید.

(۲) با استفاده از الگوریتم‌های ارائه شده در کتاب مرجع<sup>۲</sup>، مقدار گشتاور دوم<sup>۳</sup> را در هر یک از مناطق، محاسبه

کنید. (بر اساس ستون Goods)

(۳) با استفاده از الگوریتم ارائه شده در کتاب مرجع، مقدار گشتاور دوم را در هر ماه و در هر یک از مناطق،

محاسبه کنید. (بر اساس ستون Goods)

نکته: مجموعه داده بصورت یک فایل با قالب CSV است که سطر اول شامل عنوان ستون‌هاست.

---

<sup>1</sup> Mining Data Streams

<sup>2</sup> Mining of Massive Datasets

<sup>3</sup> Second Moment

## گزارش کار:

گزارش کارهای خود را در قالب یک فایل PDF بنویسید. در صورتیکه از مرجعی برای نوشتن تمرین‌ها استفاده کرده‌اید، حتما در فایل گزارش به آن ارجاع دهید. فایل گزارش باید شامل نام، شماره دانشجویی و متن گزارش به تفکیک هر سؤال باشد. گزارش کار بخشی از نمره نهایی این تمرین خواهد بود. لذا در نگارش آن نهایت دقت را داشته باشید.

## نحوه ارائه:

حتما به تمام بخش‌های تمرین خود مسلط باشید. در زمان ارائه ممکن است از شما توضیحاتی تکمیلی در مورد تمرین و یا تغییراتی در کد خواسته شود. نمره‌ی اختصاص داده شده برای افرادی که مسلط به کار خود باشند قابل تغییر است. در زمان مشخص شده جهت ارائه حتما حضور داشته باشید. غیبت در روز ارائه به منزله عدم ارسال تمرین است و نمره‌ای به آن تعلق نخواهد گرفت.

## نکات تکمیلی:

در صورت تأخیر تا ۲۴ ساعت، ۲۵٪ و تا ۴۸ ساعت ۵۰٪ از نمره تمرین کسر خواهد شد. به تمرین با تأخیر بیشتر از ۲ روز نمره‌ای تعلق نخواهد گرفت.

آدرس ایمیل جهت ارسال تمرین: [mh.qanbari@ut.ac.ir](mailto:mh.qanbari@ut.ac.ir)

فایل ارسالی یک فایل فشرده شده با قالب نامگذاری زیر باشد. (بجای NAME نام خود و به جای STDID شماره دانشجویی خود را قرار دهید).

hw3\_NAME\_STDID.zip

فایل فشرده شده باید شامل موارد زیر باشد:

۱- گزارش کار در قالب یک فایل PDF

۲- کدهای پیاده‌سازی شده

موفق باشید. ☺