

داده‌کاوی پیشرفته

تمرین اول

پاسخها را در قالب یک فایل pdf به ایمیل rezanzg@iasbs.ac.ir ارسال فرمایید.

ایمیل خود را حتما با موضوع (subject) زیر ارسال فرمایید.

DataMining_HW1_#####

بجای علامت هشتگ شماره دانشجویی خود را وارد کنید.

آخرین مهلت ارسال تمرین: روز یکشنبه مورخه ۱ آبان‌ماه تا ساعت ۲۳:۵۹

- Answer the following questions.

به سوالات زیر پاسخ دهید.

- 1- What is **Data Mining**?

داده‌کاوی چیست؟

- 2- Why **Data Mining** is so important?

چرا داده‌کاوی بسیار مهم است؟

- 3- Explain the **KDD** process and highlight the steps that show preprocess.

فرآیند اکتشاف دانش از داده‌ها را توضیح دهید و مراحل مربوط به پیش پردازش را برجسته کنید.

- 4- What is the difference between **descriptive** and **predictive** tasks in data mining functionalities?

تفاوت بین وظایف توصیفی و پیش‌بینی در عملکردهای داده کاوی چیست؟

- 5- What are **clustering** and **classification** differences?

تفاوت های خوشه بندی و طبقه بندی چیست؟

- 6- What are **outliers**? Are they useful?

داده پرت چیست؟ آیا داده‌های پرت قابل استفاده هستند؟

- 7- What are the major challenges of mining huge amounts of data?

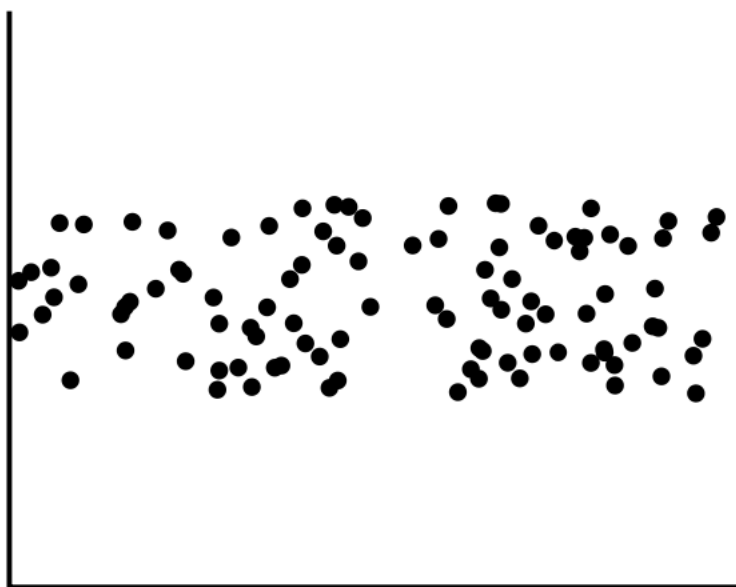
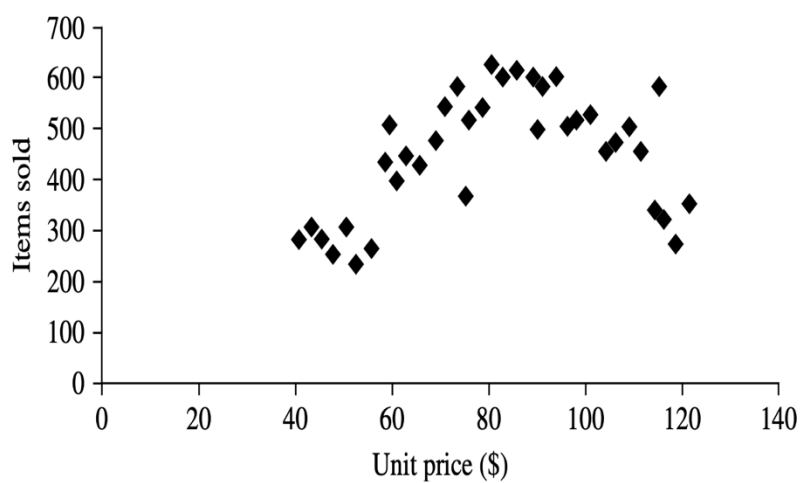
چالش های اصلی در کاوش داده‌های عظیم چیست؟

- 8- Name **types of attributes** with examples.

انواع ویژگی‌ها را با ذکر مثال نام ببرید.

9- Is there any relation between the data points in the following pictures? Why?

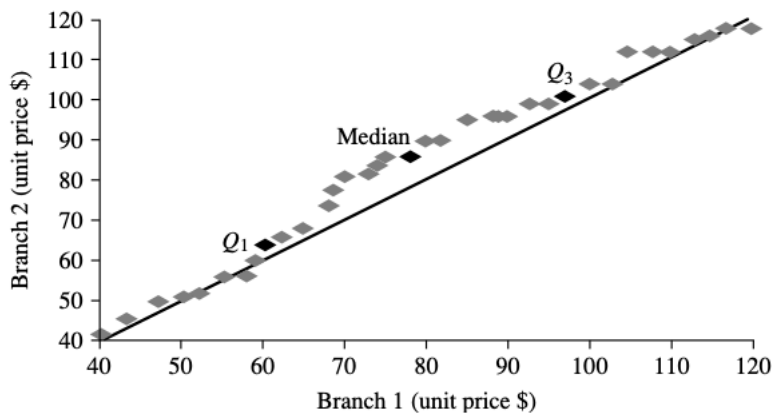
آیا ارتباطی بین داده‌های تصاویر زیر وجود دارد؟ چرا؟



10- Due to the plot which branch would you choose to purchase different items from?

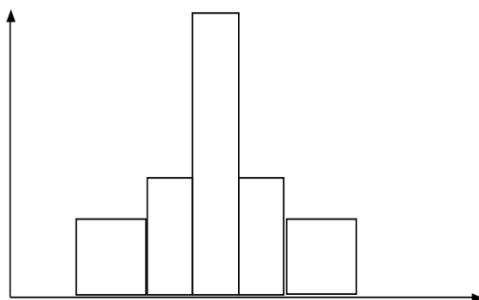
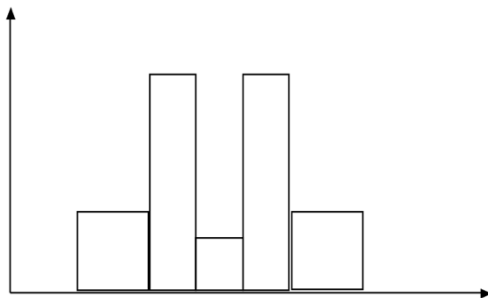
Why?

با توجه به نمودار کدام شعبه را برای خرید انتخاب می‌کنید؟ چرا؟



11- Due to same **boxplots** for the following **histograms**. What extra information we can get from histograms?

با توجه به نمودارهای جعبه‌ای مشابه برای هیستوگرام های زیر. چه اطلاعات اضافی را می توانیم از هیستوگرام به دست آوریم؟



12- Define the following terms:

اصطلاحات زیر را تعریف کنید:

- Mean
- Mode
- Variance
- Standard deviation