۱) به سراغ داده‌های غلات صبحانه می‌رویم. بر اساس داده‌های مزبور به پرسش‌های زیر جواب دهید.

الف) کدامیک از متغیرها کمّی/عددی هستند؟ کدامیک ترتیبی هستند؟ کدامیک اسمی[1] هستند؟

ب) مقدار میانگین، میانه، کمینه، بیشینه و انحراف استاندارد را برای هر یک از متغیرهای عددی محاسبه کنید. این‌کار می‌تواند با کمک ***pandas*** انجام شود.

ج) برای هر یک از متغیرهای عددی یک هیستوگرام رسم کنید. بر اساس هیستوگرام‌ها و همچنین برخی از آماره‌های خلاصه، به پرسش‌های زیر پاسخ دهید.

I. کدام متغیر دارای بالاترین مقدار تغییرپذیری[2] هستند؟

II. کدام متغیرها به نظر می‌رسد دارای چولگی هستند؟

III. آیا مقدار کرانگین[3] در داده‌ها وجود دارد؟

د) با رسم نمودارهای جعبه‌ای در کنار یکدیگر، کالری‌های غلات سرد و گرم را مقایسه کنید. این نمودار چه چیزی را به ما نشان می‌دهد؟

ه) یک نمودار جعبه‌ای از رتبه‌بندی مصرف‌کننده به عنوان تابعی از ارتفاع قفسه رسم کنید. اگر مایل به پیش‌بینی مصرف‌کننده از ارتفاع قفسه بودید، آیا لازم بود تا تمامی سه طبقه مربوط به ارتفاع قفسه را نگهداری کنید؟

و) جدول همبستگی متغیرهای عددی را محاسبه کنید (متد ***corr()***). یک نمودار ماتریسی برای این متغیرها تولید کنید (به کمک کتابخانه ***seaborn*** انجام شود).

I. کدامیک از زوج متغیرها دارای بیشترین مقدار همبستگی هستند؟

II. چگونه می‌توان بر اساس این همبستگی‌ها، تعداد متغیرها را کاهش داد؟

III. اگر در ابتدا داده‌ها نرمال‌سازی شوند آنگاه چه تغییراتی در همبستگی‌ها دیده می‌شود؟

ز) مؤلفه اصلی اول حاصل از تحلیل ۱۳ متغیر عددی موجود در اسلاید درس وجود دارد. به طور خلاصه بیان کنید که این مؤلفه اصلی به چه چیزی اشاره می‌کند.

۲) نتیجه اجرای یک الگوریتم داده‌کاوی بر روی یک مجموعه داده‌های تراکنشی بدین شرح است: ۸۸ رکورد به عنوان کلاهبردار رده‌بندی شده‌اند (که ۳۰ رکورد آن به درستی رده‌بندی شده است) و ۹۵۲ رکورد نیز به

---

[1] *Nominal*

[2] *Variability*

[3] *Extreme*

عنوان غیرکلاهبردار رده‌بندی شده‌اند (که ۹۲۰ رکورد آن به درستی رده‌بندی شده است). ماتریس درهم‌ریختگی را ساخته و نرخ خطای کل را محاسبه کنید.

۳) جدول ۱ یک مجموعه کوچکی از نتایج یک مدل رده‌بندی را همراه با مقادیر واقعی آن نشان می‌دهد.

الف) با مقادیر بُرش ۰/۲۵، ۰/۵ و ۰/۷۵، نرخ‌های خطا، حساسیت و وضوح را محاسبه کنید.

ب) یک نمودار خیزش دهکی برای این مجموعه رسم کنید.

جدول ۱: مقادیر واقعی عضویت رده و تمایل‌ها برای یک مجموعه داده‌های اعتبارسنجی.

رده واقعی	تمایل به رده یک
0	***0.03***
0	***0.52***
0	***0.38***
1	***0.82***
0	***0.33***
0	***0.42***
1	***0.55***
0	***0.59***
0	***0.09***
0	***0.21***
0	***0.43***
0	***0.04***
0	***0.08***
0	***0.13***
0	***0.01***
1	***0.79***
0	***0.42***
0	***0.29***
0	***0.08***
0	***0.02***