

فایل "DirectMarketing.csv" را از پورتال درس دانلود کرده و به شکل دیتا فریم در R ذخیره کنید. این فایل حاوی اطلاعات مشتریان یک بازاریاب مستقیم است که محصولاتش را با استفاده از پست مستقیم به فروش می‌رساند. او کاتالوگ محصولاتش را به افراد می‌فرستد و مشتریان با تلفن محصول موردنیاز خود را سفارش می‌دهند. این بازاریاب داده‌های مشتریان خود را که حاوی ۱۰۰۰ رکورد هست جمع‌آوری کرده است و هدفش این است که ببیند چه عواملی باعث می‌شود برخی مشتریان بیشتر از برخی دیگر سفارش دهند.

متغیرهای حاوی این فایل داده شامل موارد زیر است:

- Age: دسته‌بندی سنی
- Gender: دسته‌بندی جنسیتی
- OwnHome: مشتری مالک است یا مستأجر
- Married: مجرد یا متأهل
- Location: میزان نزدیکی به فروشگاه‌های که محصولات مشابه را بفروشد.
- Salary: حقوق سالیانه مشتری برحسب دلار
- Children: تعداد فرزندان
- History: حجم خرید گذشته
- Catalogs: تعداد کاتالوگ‌های ارسالی به مشتری
- AmountSpent: میزان خرید برحسب دلار

به پرسش‌های زیر پاسخ دهید. توجه کنید انتظار این است که پاسخ‌های خود را همراه با کدهای R در قالب یک فایل Word یا PDF ارائه کنید.

- ۱- با استفاده از تابع summary یک خلاصه از کلیه متغیرهای موجود در دیتا فریم تهیه کنید (۴ نمره).
- ۲- چند درصد مشتریان این بازاریاب را زنان تشکیل می‌دهند؟ (۴ نمره)
- ۳- چند درصد مشتریان مرد متأهل، دارای حقوق سالیانه‌ای بالاتر از ۵۰ هزار دلار هستند؟ (۴ نمره)
- ۴- نمودار فراوانی مشتریان را بر اساس رده‌بندی سنی رسم کنید بطوریکه ویژگی‌های زیر را داشته باشد (۴ نمره):
 - دسته‌های سنی بر اساس فراوانی‌شان از کمترین به بیشترین مرتب شده باشند.
 - این نمودار باید دارای عنوان "Age Distribution" باشد.
 - محور افقی دارای عنوان "Age Category" باشد.
 - حدود محور عمودی از ۰ تا ۶۰۰ تنظیم شود.
- ۵- نمودار سؤال ۴ را این بار به صورت انباشته رسم کنید بطوریکه در هر دسته سنی تعداد زنان و مردان هم نمایش داده شود (۴ نمره). نمودار باید ویژگی‌های زیر را داشته باشد:
 - این نمودار باید دارای عنوان "Distribution by Age and Gender" باشد.
 - محور افقی دارای عنوان "Age Category" باشد.
 - حدود محور عمودی از ۰ تا ۶۰۰ تنظیم شود.
 - فراوانی زنان با رنگ سبز و مردان با رنگ قرمز مشخص شده باشد.

نمودار دارای راهنما (Legend) باشد که رنگ مربوط به فراوانی زنان و مردان را مشخص کند.

۶- نمودار چگالی توزیع میزان خرید بر حسب دلار را رسم و بر این داده‌ها توزیع نمایی را برازش کنید (۴ نمره).

راهنمایی: از تابع `dexp` استفاده کنید و پارامتر `rate` را معکوس میانگین داده‌ها قرار دهید.

۷- با استفاده از نمودار جعبه‌ای (Box Plot) رابطه بین تعداد کاتالوگ‌های دریافتی و میزان خرید بر حسب دلار را بررسی کنید. محورهای عمودی و افقی را به شکل صحیح نام‌گذاری و برای نمودار خود یک عنوان انتخاب کنید (۴ نمره).

۸- با استفاده از نمودار پراکندگی رابطه بین سطح درآمد و میزان خرید بر حسب دلار را بررسی کنید. خط رگرسیون عبوری از داده‌ها را رسم کرده و برای نمودار خود یک عنوان انتخاب کنید (خط رگرسیون قرمز باشد و ضخامت آن را معادل ۲ قرار دهید). (۴ نمره)

۹- فرض کنید شما علاقمند هستید تا اثر جنسیت مشتریان را بر روی میزان خرید بررسی کنید (۸ نمره).
الف) با استفاده از نمودار جعبه‌ای توزیع میزان خرید بر حسب دلار را بر حسب جنسیت بسنجید. آیا تفاوتی بین میزان خرید مردان و زنان مشاهده می‌کنید؟ میانگین میزان خرید این دو گروه را نیز مقایسه کنید (۲ نمره).
ب) شخصی ادعا می‌کند تفاوت میزان خرید بین زنان و مردان ربطی به جنسیت آنان ندارد، بلکه ریشه اصلی تفاوت در سطح دستمزدهاست. برای بررسی این ادعا، میانگین دستمزد دو گروه را محاسبه و مقایسه کنید. همچنین با استفاده از نمودار جعبه‌ای توزیع دستمزد را بر حسب جنسیت بسنجید. این ادعا را چطور ارزیابی می‌کنید؟ (۲ نمره)
ج) یک روش برخورد با مشکل بالا این است تا حد ممکن اثر دستمزد را خنثی کنید و بعد بررسی کنید آیا در میزان خرید بین مردان و زنان تفاوتی وجود دارد یا خیر. برای این منظور زنان و مردانی را که دستمزدی بالاتر از میانه دستمزدها دارند با هم و زنان و مردانی را که دستمزدی پایین‌تر از میانه دستمزد دارند، با هم مقایسه کنید (۴ نمره).

راهنمایی: توجه کنید در ابتدا لازم است تا یک متغیر جدید ایجاد کنید تا افراد را در دو گروه با درآمد بالا و درآمد پایین تقسیم کند. همچنین توجه کنید انتظار این است به عنوان پشتوانه استدلال خود، نمودارهایی مانند آنچه در مثال سوم بخش نمایش داده‌ها با استفاده از `ggplot2` ارائه شد، رسم کنید.