نکته مهم درباره محیط اجرا

این تمرین را حتماً در یکی از محیط‌های تعاملی زیر انجام دهید:

گزینه 1: Google Colab

- **مزیت:** نیازی به نصب چیزی ندارید، فقط با مرورگر کار کنید
- **لینک:** colab.research.google.com
- **راهنما:** یک Notebook جدید بسازید و کدها را در سلول‌های جداگانه اجرا کنید
- نمودارها مستقیماً زیر هر سلول نمایش داده می‌شوند

گزینه 2: Jupyter Notebook

سوال اول: تحلیل آب و هوای شهرها

پروژه کاملی بنویسید که:

بخش 1: تولید داده با NumPy

با NumPy داده‌های تصادفی برای **30 روز** و **6 شهر** ایجاد کنید:

- **دمای روزانه:** بین 15 تا 40 درجه
- **رطوبت:** بین 20 تا 80 درصد
- **بارندگی:** بین 0 تا 50 میلی‌متر

بخش 2: ساخت DataFrame با Pandas

با Pandas یک DataFrame بسازید با ستون‌های زیر:

- City: نام شهر (Tehran, Mashhad, Isfahan, Tabriz, Shiraz, Ahvaz)
- Day: شماره روز (1 تا 30)
- Temperature: دما
- Humidity: رطوبت
- Rainfall: بارندگی

بخش 3: تحلیل‌های آماری

محاسبه کنید:

1. میانگین دما، رطوبت و بارندگی هر شهر (با groupby)
2. شهری که گرم‌ترین و سردترین میانگین دما را دارد
3. تعداد روزهایی که بارندگی بیش از 10 میلی‌متر بوده (با فیلتر و groupby)

بخش 4: رسم نمودار

با Matplotlib یک **نمودار پراکندگی** (Scatter Plot) رسم کنید که رابطه بین دما و رطوبت را **فقط برای شهر اصفهان** نشان دهد:

- **محور x: دمای اصفهان**
- **محور y: رطوبت اصفهان**
- **عنوان نمودار: 'رابطه دما و رطوبت - اصفهان (30 روز)'**
- **برچسب محور x: 'دما (C°)'**
- **برچسب محور y: 'رطوبت (٪)'**

بخش 5: تفسیر

آیا بین دما و رطوبت در اصفهان رابطه‌ای مشاهده می‌شود؟ توضیح دهید.

سوال 2: تمیزکاری و تحلیل داده‌های دیابت

شما فایل diabetes.csv را خوانده و مراحل زیر را طی کنید:

بخش 1: تمیزکاری داده

1. مقادیر مفقود را با **میانگین** همان ستون پر کنید.
2. بررسی کنید که آیا هنوز مقدار مفقودی باقی مانده است.
3. 5 سطر اول داده تمیز شده را نمایش دهید.

بخش 2: رسم Heatmap

4. فقط 6 ستون مهم زیر را انتخاب کنید:
   - Outcome ,Insulin ,Age ,BMI ,BloodPressure ,Glucose
5. ماتریس همبستگی این ستون‌ها را محاسبه و چاپ کنید.
6. با استفاده از کتابخانه seaborn یک **Heatmap** رسم کنید که:
   - اعداد همبستگی را روی نمودار نمایش دهد
   - عنوان مناسب داشته باشد
7. از روی **Heatmap** و ماتریس همبستگی به سوال زیر پاسخ دهید:

الف) کدام ویژگی بیشترین همبستگی با Outcome (تشخیص دیابت) دارد و مقدار همبستگی آن چقدر است؟

- **راهنمایی:**
- ابتدا ستون Outcome را از ماتریس همبستگی انتخاب کنید
- سپس خود Outcome را با متد .drop('Outcome') حذف کنید (چون همبستگی خودش با خودش همیشه 1 است)
- با استفاده از .idxmax() نام ویژگی با بیشترین همبستگی را پیدا کنید
- با استفاده از .max() مقدار این همبستگی را نمایش دهید

نحوه تحویل:

تمرین را در Git آپلود کنید در هنگام تحویل تمرین فایل های شما در گیت بررسی میشود.