



دانشگاه صنعتی امیرکبیر  
(پلی تکنیک تهران)  
دانشکده مهندسی صنایع

به نام خدا

نام استاد: دکتر مریم اشرفی

## کنترل کیفیت آماری

تدریسار: پدram پیرو اصفیا

یک دیتاست واقعی که شامل ۱۵ کشور میشود، در اختیار شما قرار گرفته است. در این دیتاست، ۷ ستون اول راجع به تعداد مقالاتی است که کشور های ذکر شده، منتشر کرده اند و آن کشور ها را بر اساس این تعداد، مرتب کرده است و به آن ها رنک داده است.

سه ستون بعد، در رابطه با انرژی مورد نیاز هر کشور میباشد، به ترتیب ستون های Energy Supply، Energy Supply per Capita، %Renewable کل انرژی مورد نیاز کشور، انرژی مورد نیاز به ازای هر نفر از جمعیت (کل انرژی تقسیم بر کل جمعیت کشور) و درصد انرژی تجدید پذیر را نشان میدهد.

ستون های ۲۰۰۶ تا ۲۰۱۵ نیز در رابطه با مقدار GDP این کشور ها میباشد.

با توجه به اطلاعات داده شده راجع به دیتاست، به سوالات زیر پاسخ دهید و موارد خواسته شده را، در یک فایل گزارش وورد بیاورید. (اگر از jupyter notebook استفاده میکنید، میتوانید گزارش خود را به همراه کد در آنجا بنویسید و ارسال نمایید).

۱. آیا دیتاست، شامل NaN Value میباشد؟ اگر بله، این مقدار را با میانگین داده های قبلی مربوط به آن زمینه در رابطه با آن کشور پر کنید. همینطور مقدار پر شده، کشور و ستونی که NaN Value داشت را به صورت یک tuple ۳ عضوی چاپ کنید.

۲. ۱۵ کشور برتر برای میانگین GDP طی ۱۰ سال گذشته کدامند؟ خروجی شما باید یک DataFrame باشد که ایندکس آن اسم کشور، و ستون روبروی آن میانگین GDP کشور ها باشد، همینطور این مقادیر از زیاد به کم باید مرتب شده باشند.

۳. روند GDP ۳ کشور برتر از لحاظ میانگین GDP (که در سوال قبل محاسبه کردید) را طی ۱۰ سال گذشته در یک figure رسم کنید.

نکته: برای این کار حتما مشخص کنید که کدام خط، مربوط به کدام کشور است. برای محور افقی و عمودی نیز لیبل مناسب قرار دهید.

۴. با استفاده از ستون های "کل انرژی مورد نیاز کشور" و "انرژی مورد نیاز به ازای هر نفر از جمعیت"، مقدار جمعیت هر کشور را بدست آورید و در ستون Population ذخیره کنید. سپس جمعیت ۵ کشور با کمترین جمعیت را در قالب یک PieChart رسم کنید و درصد مربوطه را داخل پای چارت نمایش دهید.

- امتیازی: ابعاد پای چارت را به صورت مربعی قرار دهید. (در یک فریم ۱:۱ و بدون مشخص کردن سائز فیگور) کد مربوطه را در متن گزارش خود بیاورید.

۵. نسبت تعداد مقالاتی که خود نویسنده به خودش رفرنس زده را به کل مقالات قابل رفرنس در هر کشور بدست آورید. سپس BarChart مربوطه را با مرتب کردن داده ها از بزرگ به کوچک رسم کنید. (به طوری که bar اول از همه بزرگتر باشد و رفته رفته کوچک شود)

- امتیازی: اسم کشور ها را به عنوان لیبل در زیر هر میله به صورتی که معلوم باشند، بیاورید. همچنین مقدار دقیق تا دو رقم اعشار را بالای هر bar بیاورید.

به صورت کلی تحلیل کوچکی (در حد مینیمم یک خط) راجع به هر بخش بنویسید.