

شماره دانشجویی شما رقم یکان، دهگان و صدگان دارد. سؤال ۱ طبق فرد و زوج بودن رقم یکان آن مشخص می‌شود، سؤال ۲ طبق فرد و زوج بودن دهگان و سؤال ۳ طبق فرد و زوج بودن صدگان. بر اساس فرد یا زوج بودن رقم شماره دانشجویی خود تنها به یک نسخه از دو نسخه‌ی هر سؤال پاسخ دهید. لطفاً با دقت سؤال مربوطه را پیدا کنید و در پاسخنامه پاسخ دهید. تنها در صورتی نمره درست دریافت می‌کنید که به سؤال صحیح پاسخ دهید.

برای هر سؤال یک کد کامل پایتون را ارسال کنید که مستقلاً اجرا شود و تمام خروجی‌های لازم را تولید کند. می‌توانید قبل از ارسال، کد خود را در محیط jupyter notebook تست کنید.

با شرکت در این امتحانات تعهد می‌دهید که از هیچ فرد دیگری برای پاسخ به سؤالات کمک نمی‌گیرید و همچنین پاسخ‌ها را برای هیچ کسی ارسال نمی‌کنید. البته سرچ در اینترنت و مطالعه‌ی documentation آزاد است.

سؤال ۱ برای یکان فرد

با دستورهای رو به رو ۱۰ نقطه‌ی رندوم در فضای سه بعدی ایجاد کنید:

```
import numpy as np
np.random.seed(0)
a=np.random.randn(10,3)
```

کدی بنویسید که یک آرایه numpy به نام nn به طول ۱۰ را در خروجی چاپ کند. در این آرایه برای هر کدام از ۱۰ نقطه‌ی ورودی، اندیس نزدیک‌ترین نقطه به آن نوشته شود (فاصله اقلیدسی). توجه بفرمایید که اندیس‌ها از ۰ تا ۹ هستند.

برای کسب نمره‌ی کامل از حلقه (مثلاً for) و کتابخانه‌های دیگر استفاده نکنید و تمام محاسبات را به صورت ماتریسی با numpy انجام دهید. در صورتی که لازم دارید از حلقه یا کتابخانه‌های دیگر استفاده کنید، می‌توانید، اما نمره‌ی شما از ۰.۵ محاسبه می‌شود.

سؤال ۱ برای یکان زوج

تابعی بنویسید به نام date_difference که دو تاریخ را به شکل شمسی (جلالی) در رشته‌های s و t را دریافت کند و فاصله‌ی دو تاریخ را به شکل یک عدد حسابی خروجی بدهد. برای مثال به شکل زیر می‌توان این تابع را صدا زد:

```
date_difference('1399/10/12', '1399/7/5')
```

97

فرض بفرمایید که تاریخ‌های ورودی بین سالهای ۹۶ تا ۹۹ هستند. یعنی غیر از اسفند سال ۹۹، مابقی استفاده‌ها ۲۹ روزه هستند. همچنین خروجی می‌تواند منفی یا مثبت یا صفر باشد. در ورودی صفر پشت عدد نداریم.

سؤال ۲ برای دهگان فرد

الف) جدول income.csv را با استفاده از pandas لود کنید.

ب) در ستون workclass و occupation مقادیر "?" را با nan جایگذاری کنید.

ج) برای هر کدام از ستون‌های `workclass` و `occupation` بنویسید که چند سطر `nan` وجود دارد. همچنین بنویسید که در چند سطر حداقل یکی از این ستونها `nan` هست.

د) تمام سطرهایی که `nan` دارند را در یک جدول جدا نشان دهید.

ه) ۱۰ فردی که بیشترین تعداد ساعت در هفته کار می‌کنند را در یک جدول نمایش دهید.

و) چاپ کنید که طبق این جدول ایرانی‌ها به صورت میانگین چند ساعت در هفته کار می‌کنند؟

ز) چاپ کنید که چند درصد ایرانی‌های این جدول زن هستند.

سؤال ۲ برای دهگان زوج

جدول `income.csv` را با استفاده از `pandas` لود کنید و جدولی را نمایش دهید که برای هر کشور موارد زیر را نشان دهد:

الف) نام کشور،

ب) تعداد افرادی که در این جدول هستند،

ج) میانگین تعداد ساعت کار،

د) درصد شاغلین زن.

سؤال ۳ برای صدگان فرد:

الف) جدول `income.csv` را با استفاده از `pandas` لود کنید.

ب) نمودار ستونی (`bar plot`) تعداد مردان و زنان را بر حسب متغیر `income` در یک نمودار رسم کنید. توجه فرمایید که فرمت نمودار و جزئیات آن کامل و صحیح باشد.

ج) با استفاده از `barplot` توضیح `marital-status` را رسم کنید. توجه بفرمایید که فرمت نمودار و جزئیات آن کامل و صحیح باشد.

د) برای هر کدام از دسته‌های `marital-status` هیستوگرام سن را به شکل خوانا رسم کنید.

سؤال ۳ برای صدگان زوج

الف) جدول `income.csv` را با استفاده از `pandas` لود کنید.

ب) نمودار `scatterplot` سن و تعداد ساعت کاری را رسم کنید. برای موارد مربوط به ایران نقاط بزرگ استفاده کنید. برای سایر کشورها نقاط کوچک و آبی استفاده کنید. توجه بفرمایید که فرمت نمودار و جزئیات آن کامل و صحیح باشد.

ج) برای هر کشور به دست بیاورید که چند درصد افراد سیاه‌پوست هستند و برای ۵ کشوری که بیشترین سیاه‌پوست را دارند، یک نمودار `bar plot` رسم کنید که درصد سیاه‌پوست‌ها را به صورت سورت شده و نزولی نشان دهد. توجه بفرمایید که فرمت نمودار و جزئیات آن کامل و صحیح باشد.

موفق باشید!