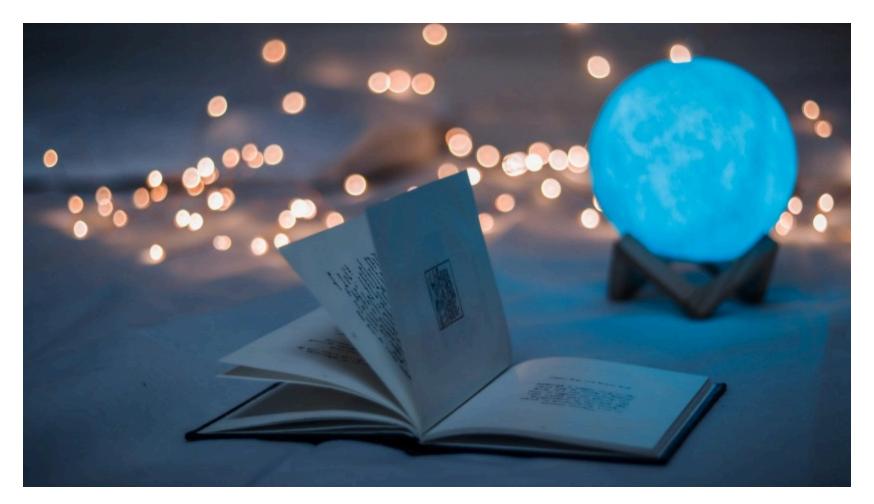
تحلیل داده اکتشافی (قسمت اول)

در این سوال عملیات EDA روی یک مجموعهداده فارسی و انگلیسی انجام میشود.



مجموعهداده

مجموعه داده سوال را میتوانید از این لینک دانلود کنید.

هنگامی که این فایل را از حالت فشرده خارج کنید، فایل goura_questions.csv و shereno.csv در اختیار شما خواهد بود.

فایل shereno.csv دارای ۴۴۰۰ سطر و ۴ ستون میباشد و جزییات آن در جدول زیر آمدهاست:

توضيحات ستون	نام ستون
متن شعر	Poem
شاعر	Poet
نام شعر	Title
كتاب	Book

فایل qoura_questions.csv دارای ۱۹۰۳ سطر و ۱ ستون میباشد و جزییات آن در جدول زیر آمدهاست:

توضيحات ستون	نام ستون
سوال	question

صورت مسئله

۱. تعداد کلمات منحصر به فرد (یکتا) در هر سوال (ستون question) در questions.csv را پیدا کنید سپس مجموع تعداد کلمات منحصر به فرد در هر سوال را ارسال کنید.

مثال ۱: جمله ?Why do I not lose weight when I throw up دارای ۹ کلمه یکتا میباشد.

مثال ۲: جمله Hi, my name is Ali.\nmy cousin's name is Ali too! دارای ۸ کلمه یکتا میباشد.

3453

۲. ابتدا مجموع تعداد رقمهای موجود در ستون سوال (question) در فایل qoura_questions.csv و سپس تعداد رقم موجود در ستون متن شعر (Poem) در فایل shereno.csv را بهدست آورید.

مثالا: جمله Are there any good horror movies in 2016 دارای ۴ رقم است.

۳. با استفاده از فایل stopwords.txt تعداد stopwordsهای استفاده شده در متن شعر (Poem) در فایل shereno.csv را بهدست آورید.

3423

خروجي

برای ارسال پاسخ، جواب هر سوال را مشابه مثال، در یک سطر از یک فایل text بنویسید. در نهایت یک فایل سه خطی با نام output.txt را برای ما بارگذاری کنید.

در صورتی که جواب سوالی را نمیدانید در سطر مربوطه ۱- وارد کنید.

نمونه خروجی ۱

3453 412 532 3423

▼ توضیحات

نمونه خروجی ۲

3256

-1

-1

▼ توضیحات

چون پاسخ سوال دوم و سوم را نمیدانستید، به جای آن عدد ۱- نوشتهاید!