

تحلیل داده اکتشافی (قسمت اول)

در این سوال عملیات EDA روی یک مجموعه داده فارسی و انگلیسی انجام می شود.



مجموعه داده

مجموعه داده سوال را می توانید از [این لینک](#) دانلود کنید.

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

فایل `shereno.csv` دارای ۴۴۰۰ سطر و ۴ ستون می باشد و جزییات آن در جدول زیر آمده است:

نام ستون	توضیحات ستون
Poem	متن شعر
Poet	شاعر
Title	نام شعر
Book	کتاب

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

نام ستون	توضیحات ستون
question	سوال

صورت مسئله

۱. تعداد کلمات منحصر به فرد (یکتا) در هر سوال (ستون question) در qoura_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.

مثال۱: جمله Are there any good horror movies in 2016 دارای ۴ رقم است.

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

خروجی

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 نمی دانید در سطر مربوطه ۱- وارد کنید.

نمونه خروجی ۱

▼ توضیحات

در خط i ام جواب سوال i ام نوشته شده است.

نمونه خروجی ۲

▼ توضیحات

چون پاسخ سوال دوم و سوم را نمی دانستید، به جای آن عدد ۱- نوشته اید!

