

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

نام ستون	توضیحات ستون
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.

مثال ۱: جمله Are there any good horror movies in 2016 دارای ۴ رقم است.

412 532

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

3423

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

3453

412 532

3423

▼ توضیحات

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

نمونه خروجی ۲

3256

-1

-1

▼ توضیحات

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

▼ هشدار

فراموش نکنید که قبل از پایان زمان مسابقه، بایستی تمامی کدهای این مسابقه را از قسمت بارگذاری کُد برای ما ارسال کنید. در غیر این صورت، شما از این مسابقه، امتیازی کسب نمی کنید.

توجه داشته باشید که اگر از jupyter notebook استفاده می کنید بایستی همانند توضیحات قسمت بارگذاری کُد، خروجی py را دریافت و برای ارسال در نظر بگیرید. ارسال فایل های jupyter همانند ipynb. مورد قبول واقع نخواهند شد.