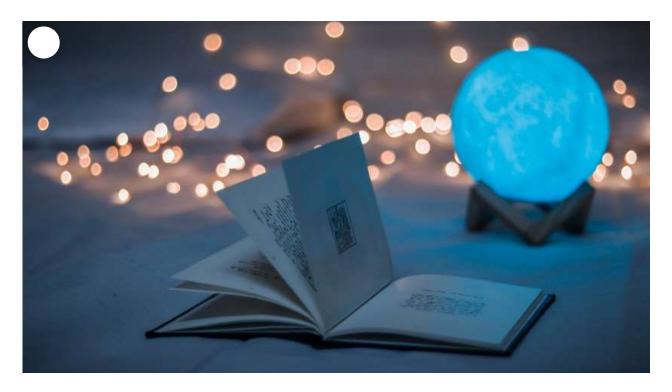
تحلیل داده اکتشافی (قسمت دوم)

در این سوال عملیات EDA روی یک مجموعهداده سوالات انگلیسی انجام می شود.



مجموعه داده

مجموعه داده سوال را میتوانید از این لینک دانلود کنید. دقت کنید این مجموعه سوال همان فایل qoura_questions.csv

صورت مسئله

۱. تعداد کلماتی که با m شروع میشوند، با t تمام می شوند و طول آنها بیشتر از ۴ کاراکتر است را بهدست آورید.

▼ توجه

دقت کنید وجود علامتهای نگارشی در پایان کلمات ایرادی ندارد و کلمات در این حالت نیز شمارش میشوند. همچنین بزرگی و کوچکی حروف مهم نیست.

43

۲. مجموع تعداد ایموجی های متن را بهدست آورید.

66

۳. ۵ کلمه پر تکرار و تعداد تکرار آنها را با با ترتیب نزولی بهدست آورید.

the:14715 of:6742 and:6517 a:4805 to:4707

۴. تعداد کلماتی که فقط یک بار در کل متن تکرار شدهاند را بهدست آورید.

2342

خروجي

برای ارسال پاسخ، جواب هر سوال را مشابه مثال، در یک سطر از یک فایل text بنویسید. در نهایت یک فایل 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 ا

43

66

the:14715 of:6742 and:6517 a:4805 to:4707

2342

▼ توضیحات

در خط i ام جواب سوال i ام نوشته شدهاست.

نمونه خروجی ۲

3256

1000

-1

-1

▼ توضیحات

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

▼ 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 jupter notebook استفاده می کنید بایستی همانند توضیحات قسمت بارگذاری کُد، خروجی pyter را دریافت و برای ارسال در نظر بگیرید. ارسال فایلهای jupyter همانند خروجی و برای ارسال در نظر بگیرید. ارسال فایلهای ipyter همانند بازگذاری کُد، خروجی و برای ارسال در نظر بگیرید. ارسال فایلهای ipyter همانند بازگذاری کُد، خروجی و برای ارسال در نظر بگیرید. ارسال فایلهای ipyter همانند بازگذاری کُد، خروجی و برای ارسال در نظر بگیرید. ارسال فایلهای ipyter و برای ارسال در نظر بگیرید. ارسال فایلهای ipyter و برای ارسال در نظر بگیرید.