



مقدمه

پیش از شروع به انجام تمرین، کتابچه هندزآن را مطالعه و Task های آن را در داخل آن انجام دهید.

بخش اول – خواندن انواع فایل

در بخش قبل به راه اندازی و نصب پکیج های مختلف پرداختیم. در این قسمت قصد داریم تا با خواندن انواع فایل در بخش های جداگانه آشنا بشویم.

فایل txt

از سایت [Gutenberg](https://www.gutenberg.org/)¹ به دلخواه یک کتاب را انتخاب کرده و مراحل زیر را انجام دهید .
(برای شروع میتوانید کتاب های با حجم کم را دانلود کرده و یا کتاب مورد نظر را خودتان خلاصه کنید، بدین صورت که حداقل 50 سطر از کتاب را انتخاب کنید و سایر بخش ها را دستی پاک کنید).

- تنها 5 خط اول کتاب را چاپ کنید.
- کلمات هر خط را جدا کرده و در یک متغیر ذخیره کنید.

فایل csv

داده های مربوط به واکسن کرونا به تفکیک هر کشور (با نام country_vaccination.csv ضمیمه شده است) را بخوانید و مراحل زیر را انجام دهید .

- تمامی داده ها را خوانده و داده ها را تمیز کنید (حذف داده های Nan و تکراری)

فایل log

از لینک [پیوست](https://www.kaggle.com/arpitdw/exploring-the-evolution-of-linux/data)² شده داده های مربوط به log پیدایش لینوکس را دانلود کرده و به سوالات زیر پاسخ دهید دقت شود که در پاسخ به سوالات تنها مجاز به استفاده از فایل با پسوند log هستیم. (در صورت مشکل در دانلود، فایل لازم ضمیمه شده است)

¹ <https://www.gutenberg.org/>

² <https://www.kaggle.com/arpitdw/exploring-the-evolution-of-linux/data>



- با توجه به نوع و حجم داده ها ، به بهترین روش داده ها را بخوانید.

بخش دوم – گزارش آماره ها

با استفاده از فایل های خوانده شده در بخش قبل، آماره های خواسته شده را گزارش کنید.

فایل txt

- تعداد کل کلمات را گزارش کنید.
- در صورت امکان تعداد کلمات منحصر به فرد را گزارش کنید.

فایل csv

- آماره های کلی تمامی ستون ها را استخراج کنید.
- ستون جدید (first_vac) به عنوان اولین واکسن را از روی ستون vaccines استخراج کنید.

فایل log

- تعداد مشارکت به ازای هر author را گزارش کنید.
- فعال ترین author ها (5 فرد برتر) را مشخص کنید

بخش سوم – دستکاری انواع داده ها (امتیازی*)

در این بخش قصد داریم که کمی بیشتر با انواع داده ها کار کنیم.

فایل csv

داده های میزان ابتلا به کرونا و مرگ و میر ناشی از کرونا را به ازای یک کشور دلخواه (داده های مربوط به سال 2021) دانلود کنید و به سوالات زیر پاسخ دهید.

- داده های دو جدول واکسن و ابتلا به کرونا را با توجه به تاریخ ادغام کنید.
 - به ازای یک کشور دلخواه ، نمودار تزریق / ابتلا به کرونا را رسم کنید.
- (محور افقی تاریخ و محور عمودی تعداد ، توجه شود نیاز به مقایسه ی نرخ ابتلا و تزریق داریم.)



نکات تحویل

- مهلت ارسال این تمرین تا پایان روز یکشنبه 30 آبان ماه خواهد بود.
- انجام این تمرین به صورت یک نفره می باشد.
- لطفا هر گونه فرض در حل سوالات را در گزارش خود ذکر کنید.
- لطفا گزارش ، فایل کدها و سایر ضمائم مورد نیاز را با فرمت زیر در سامانه مدیریت دروس بارگذاری نمائید.

HW1_[Lastname]_[StudentNumber].zip

- در صورت وجود سوال و یا ابهام می توانید از طریق رایانامه زیر با دستیار آموزشی در ارتباط باشید:

mohammad.nili@ut.ac.ir