



دانشگاه تهران
پردیس دانشکده‌های فنی
دانشکده برق و کامپیوتر



درس

تحلیل داده و مصورسازی

دکتر محمدامین صادقی

طراح تمرین: یارا محمدی

زمان بارگزاری تمرین: ۱۸ مهر ماه

تمرین شماره ۲

جمع آوری و تمیزسازی داده

نیمسال اول سال تحصیلی ۱۴۰۰ - ۱۴۰۱

بخش اول – جمع آوری داده ها با parse کردن صفحات وب

فایل hands_on2-Data_Analytics.ipynb را بخش به بخش مطالعه کنید و قطعه کدهای درون آن را به ترتیب و با توجه به توضیحات مربوطه اجرا کنید. این فایل دارای توضیحات و کدهای آموزشی برای یادگیری چند کتابخانه کاربردی در خزش وب و همچنین تسک‌هایی در قالب جمع‌آوری و تمیزسازی داده از چند وبسایت فارسی با استفاده از همین کتابخانه‌هاست. در بخش‌هایی از این هندزآن شما باید یک کد ناقص را کامل کنید و در بخش‌هایی دیگر باید خودتان از صفر برای تسک مربوطه کد بنویسید. تمام تسک‌های این هندزآن را درون همین کتابچه انجام دهید.

برای یادگیری بهتر، پیشنهاد میشود که علاوه بر مطالب آموزشی این هندزآن، ۱۴ ویدیو اول [دوره آموزشی web scrapping با پایتون آقای حسین خورنگ](#) را نیز مشاهده کنید. همچنین توضیحات ویدیو جلسه آنلاین دوم درس را مشاهده کنید.

بخش دوم – جمع آوری داده ها با API (امتیازی*)

به دلخواه خود از یک وبسایت که دارای API رایگان و یا حتی کتابخانه در پایتون برای استخراج اطلاعات است، اطلاعاتی استخراج کنید و بعد از تمیزسازی، ۳ نمودار ساده از این داده‌ها رسم کرده و آنها را تحلیل کنید. همراه با فایل کدتان، فایل CSV داده‌ها را نیز آپلود کنید. میتوانید لیستی از API‌های قابل استفاده برای عموم را در [اینجا](#) ببینید.

گزارش شما باید شامل توضیح داده‌ها، نام کتابخانه و یا API استفاده شده، تمیزسازی‌های انجام شده و نمودارهایتان همراه با تحلیل آنها باشد. توجه داشته باشید که نیاز به استخراج داده‌های زیاد و تحلیل‌های پیچیده وجود ندارد و هدف از این تمرین، آشنایی شما با جمع‌آوری داده با استفاده از API‌هاست و همچنین برای خلاقیت شما در انتخاب داده نمره‌ای در نظر گرفته میشود.

نکات پیاده سازی و تحویل

- مهلت ارسال این تمرین تا پایان روز جمعه ۷ آبان ماه خواهد بود.
- انجام این تمرین به صورت یک نفره می باشد.
- خروجی مورد انتظار تمرین و هندزآن، در صورت انجام بخش امتیازی، ۲ فایل اجرا شده کتابچه و جویپتر با فرمت Notebook(.ipynb) و جدول داده‌های بخش امتیازی با فرمت CSV است.



- هر گونه توضیحات و گزارش نویسی را به صورت Markdown داخل کتابچه جویتر انجام دهید.
- لطفا گزارش، فایل کدها و سایر ضmann مورد نیاز را با فرمت زیر در سامانه مدیریت دروس بارگذاری نمایید.

HW2_[Lastname]_[StudentNumber].zip

- در صورت وجود سوال و یا ابهام میتوانید از طریق رایانامه زیر با دستیار آموزشی در ارتباط باشید:
yara.mohamadi@gmail.com