



دانشگاه تهران  
پردیس دانشکده‌های فنی  
دانشکده مهندسی برق و کامپیوتر

به نام خدا

## درس تحلیل داده

سال تحصیلی ۹۸-۹۹  
نیم‌سال اول

## تمرین شماره ۱

طراحان:

محمدامین صادقی

محمد یزدان‌پناه

## مقدمه

هدف از این تمرین، آشنایی با مفاهیم ضروری کار با داده‌ها در زبان برنامه‌نویسی پایتون می‌باشد. با توجه به صورت هر یک از مسائل، مورد یا موارد خواسته شده را انجام داده و نتیجه بدست آمده را به صورت شفاف مشخص کنید و در فایل خروجی قرار دهید.

برای حل هر کدام از صورت مساله‌ها، قدم به قدم کدها را در فایل Jupyter Notebook اجرا کنید و پس از رسیدن به نتیجه مطلوب قدم دیگر را انجام دهید. شکستن مساله به زیرمساله‌های کوچک‌تر بخشی از حل مساله می‌باشد و به صلاح دید خود این کار را انجام دهید.

همچنین از کتابخانه‌های numpy و pandas با توجه به شرایط پروژه خود در صورت لزوم استفاده کنید و سعی کنید بهینه‌ترین حالت برای رسیدن به جواب را پیدا کنید.

---

## مسئله شماره ۱

بر اساس لینک منبع ۱، نتیجه هر بازی در جریان رقابت‌های لالیگا از فصل ۱۹۲۹-۱۹۲۸ تا فصل ۲۰۱۹-۲۰۱۸ را می‌توانید در غالب فایل CSV دریافت کنید. تصور کنید تمام این بازی‌ها در یک فصل اتفاق افتاده‌است، مجموع امتیازات تیم‌هایی که در تمام این فصل‌ها در این سطح از رقابت‌ها حاضر بودند را بیابید. ضمناً، نسبت مجموع امتیازات هر تیم به ازای مجموع تعداد بازی‌هایی که انجام داده‌است را پیدا کنید و به صورت نزولی تیم‌ها را بر این اساس مرتب کنید.

---

## مسئله شماره ۲

در این پروژه با استفاده از لینک منبع ۲، سه فایل `name.basics.tsv.gz` و `title.crew.tsv.gz` و `title.rating.tsv.gz` را دریافت کنید و با توجه به امتیاز تخصیص یافته توسط مخاطبین برای هر فیلم، بیان کنید کدام کارگردان میانگین بهتری در ارائه آثار سینمایی دارد؟

---

## مسئله شماره ۳

با توجه به ۲ فایل قرار گرفته در Att1.zip قیمت طلا و نفت را پس از سال ۱۹۸۸ به صورت ماهیانه در یک DataFrame ذخیره کنید و سپس نمودار این دو قیمت را بر حسب ماه در یک نمودار رسم کنید. قیمت‌ها بر حسب دلار آمریکا در نظر گرفته می‌شود.

## نکات

- برای انجام این تمرین به هیچ عنوان مجاز به دستکاری در فایل‌های دریافتی (که ورودی کدهای شما خواهند بود) نیستید.
- به دیگر دوستان خود رو نیندازید، پاسخ تمرین‌ها در اینترنت قابل جستجو بوده و احتمالاً با کمی زحمت شما را به جواب می‌رسانند.
- در صورت دیدن تشابه غیر منطقی بین گزارش‌های ارسالی نمره تمرین صفر در نظر گرفته خواهد شد.
- تمامی کدهای شما اجرا خواهند شد. از صحت کدها اطمینان حاصل کنید. در صورتی که کدها اجرا نشوند نمره بخش مربوطه را از دست خواهید داد. همچنین بکارگیری محیط‌های مجازی<sup>۱</sup> و قراردادن یک فایل `requirement.txt` در کنار دیگر فایل‌ها به شدت توصیه می‌شود و احتمال از بین رفتن حق شما را کاهش خواهد داد.
- لطفاً فایل Jupyter Notebook (با پسوند `.ipynb`) و خروجی PDF شده آن را (به صورتی که خروجی‌های شما کاملاً روشن باشد) با فرمت نامگذاری `PRJ01_[LastName]_[StudentNumber].zip` در قالب یک فایل در سیستم مدیریت دروس بارگذاری کنید.
- در صورت وجود سوال یا ابهام می‌توانید با ایمیل [moh.yazdanpanah@ut.ac.ir](mailto:moh.yazdanpanah@ut.ac.ir) مکاتبه کنید.

موفق باشید  
صادقی

## مراجع

[1] <https://github.com/footballcsv/espana>

[2] <https://datasets.imdbws.com/>

---

<sup>1</sup> Virtual Environments