

## به نام خدا

### ساخت یک دیتاست از وب

به کمک مطالب آموخته شده از فصل 9 کتاب، دیتاستی را تهیه کنید که داده‌های وب را استخراج کرده و بصورت یک فایل CSV ذخیره نماید. و همراه با آن، گزارشی تهیه کنید که آدرس وبسایت و خصوصیات فیلدها و نحوه سازماندهی آن‌ها را شرح می‌دهد. همچنین در این گزارش، حداقل 5 ایده از آنچه که می‌توان از دیتاست استخراج کرد را بنویسید. در این راستا موارد زیر را رعایت کنید:

- اگر از داده غیرساخت‌یافته (مانند توضیحات متنی) استفاده می‌کنید سعی کنید آن‌ها را بنحوی ذخیره کنید که بعداً قابل جستجو و استفاده باشد.
- استفاده از کتابخانه‌ها آزاد است.
- استفاده از API هم آزاد است.
- سعی کنید داده‌هایی که جمع‌آوری می‌کنید حداقل 5 فیلد داشته باشند و تا حد امکان فیلدهای عددی داشته باشند.
- در مواردی که فیلدی خالی است را با قانونی هماهنگ (و استاندارد) پر کنید.
- در صورت علاقه، می‌توانید داده‌های تصویری را نیز بعنوان یکی از فیلدها جمع‌آوری نمایید اما شرایط لازم را برای فرمت و سایز و کلاس تصاویر رعایت نمایید.

بعنوان پیشنهاد می‌توانید دیتاستی از موضوعات زیر را انتخاب کنید:

- داده‌های بیماری‌ها (مثلاً در بیمارستانی که در مورد کوید 19 یا بیماری‌های دیگری صحبت شده است)
- وبسایتهای ورزشی
- داده‌های پژوهشی دانشگاهی (مثلاً از اساتید دانشگاه یزد مواردی مانند رشته، علایق پژوهشی و ... را از وبسایت دانشگاه و همچنین از سرویس اسکالر گوگل دریافت کنید)
- داده‌های مالی (مانند اطلاعات شرکت‌های عرضه شده در بورس که پس از ورود به وبسایت آن‌ها برخی اطلاعات جذاب آن‌ها را استخراج کنید)
- داده‌های بازار (استخراج داده‌هایی از قبیل نظر کاربران سن و جنسیت (در صورت وجود) و ... از فروشگاههای دیجیتال)
- موضوعات داغ توییتر
- نظرات کاربرای درباره فیلم‌ها
- وبسایت مجلس و استخراج داده نمایندگان
- اطلاعات مربوط به دانشگاههای دنیا یا کشور
- و ...