

پروژه Descriptive Analysis

به فرایند تمیزکردن داده، کاوش در داده و فهمیدن داده تحلیل داده گویند. در تحلیل داده اطلاعات سودمند و مفید برای تصمیم‌گیری استخراج می‌شود. بنابراین جمع‌آوری داده و تحلیل و بررسی آن از اهمیت زیادی برخوردار است. امروزه به دلیل اهمیت تحلیل‌ها و تصمیمات مبتنی بر داده، نیاز به شناخت و توجه بیشتر به این حوزه حس می‌شود.

در این پروژه قصد داریم تا به صورت عملی با دیتا کار کنید و از روش‌های مختلف مصورسازی داده و تکنیک‌های یادگرفته شده به منظور تحلیل و بررسی داده استفاده کنید. بنابراین در این پروژه پس از انتخاب دیتاست و تمیز کردن آن، یک سری سوال مطرح می‌شود که این سوالات با کمک مصورسازی، گرفتن بینش و کشفیات جدید از داده پاسخ داده می‌شود.

نکات مهمی که باید در انجام پروژه دقت شود:

- ساختار دیتا^۱ توضیح داده شود و منبع دیتا ذکر شود.
- در صورت نیاز تمیزسازی داده انجام شود و دلایل آن توضیح داده شود. به طور مثال اگر قسمتهایی از داده از دست رفته است و شامل مقادیر nan هست با روش مناسب مقادیر nan جایگزین شود و دلیل انتخاب روش بیان گردد.
- در رسم نمودارها به فرمت، سایزها، برچسب محورها، تمیزی و مواردی که برای رسم نمودار آموختید، توجه شود.
- برای انتخاب انواع مختلف نمودار و سایر فرض‌های انجام شده توضیحات لازم ارائه نمایید.
- برای هر نمودار توضیح و تحلیل کافی ارائه شود و آوردن نمودار بدون تحلیل هیچ ارزشی ندارد.
- در صورت لزوم با پارامترهای آماری مانند P_value اعتبارسنجی انجام شود.
- اگر دیتا مربوط به موقعیت جغرافیا است، حتماً از کتابخانه‌ی geopandas برای مصورسازی استفاده شود.

¹ Structure of Data

• معیارهای تصحیح:

موارد زیر در تصحیح اهمیت ویژه‌ای دارد.

- اهمیت موضوع
- بیان واضح مساله و فرضیات انجام شده
- انتخاب روش مناسب برای جبران داده از دست رفته
- کیفیت و درست تمیز کردن دیتا
- ارتباط نمودارها و موثر بودن در انتقال مطلب
- کیفیت اسلایدها و زیبایی آنها
- کیفیت داستان‌سازی از دیتا
- کیفیت صدا ، تصویر و ارائه
- رعایت محدودیت زمانی برای ارائه

مراحل انجام:

۱. لود کردن دیتاست (تایید شده باشد) در Jupiter notebook.
۲. بررسی و فهمیدن اشکالات داده با استفاده از مصورسازی آن
۳. تمیز کردن دیتا
۴. اجرای EDA و یافتن الگو و فرضیه از دیتا
۵. رسم نمودارهای لازم و تحلیل آنها
۶. بیان داستان خود در مورد دیتا
۷. اعتبارسنجی با پارامترهای آماری
۸. نتیجه‌گیری
۹. آماده کردن فایل ارائه
۱۰. آماده کردن گزارش به صورت یک فایل ویدیویی



امتیازی

❖ دیتاست مورد استفاده توسط فرد جمع آوری شود. می‌توانید از API ها یا کتابخانه های Web scraping که یاد گرفتید استفاده کنید.

نکات تحویل

- مهلت ارسال این تمرین تا پایان روز جمعه ۱۰ دی ماه خواهد بود.
- گزارش پروژه را به صورت یک فایل ویدیو به مدت ۵ دقیقه تهیه کنید.
- در صورتی که مدت زمان ویدیو از ۵ دقیقه بیشتر و یا از ۴ دقیقه و ۴۰ ثانیه کمتر باشد نمره منفی لحاظ خواهد داشت.
- ویدیوی شما در فرآیند تصحیح اهمیت ویژه‌ای دارد. لذا تمامی کارهای انجام شده مانند تمیز کردن داده و را بیان کنید.
- انجام این تمرین به صورت یک نفره می‌باشد.
- لطفا هر گونه فرض در حل سوالات را در ویدیوی خود ذکر کنید.
- لطفا ویدیو، فایل کدها، پاورپوینت (Prezi) و سایر ضمائم مورد نیاز را با فرمت زیر در سامانه مدیریت دروس بارگذاری نمایید.

HW6_[Lastname]_[StudentNumber].zip

- در صورت وجود سوال و یا ابهام می‌توانید از طریق رایانامه زیر با دستیار آموزشی در ارتباط باشید:
mojtaba.amiri@ut.ac.ir