

تمرین سری اول

هدف سری اول تمرین‌ها، تقویت توانایی شما در استفاده از تحلیل‌ها و استنتاج‌های آماری برای تجزیه و تحلیل داده‌ها و همچنین پاسخ آماری به سوالات مختلف درباره داده‌ها است. در ادامه دو مجموعه داده مختلف در اختیار شما قرار گرفته است. برای هر مجموعه داده، سوالاتی در اختیار شما قرار گرفته است، که با استفاده از روش‌های استنتاج آماری که در درس فراگرفته‌اید به آن‌ها پاسخ می‌دهید. بعلاوه شما می‌بایست بر اساس تشخیص خودتان سوالات ابتکاری و مرتبط دیگری را در مورد داده‌ها مطرح و سعی کنید با استفاده از روش‌های تحلیل و استنتاج آماری به آن‌ها پاسخ دهید.

برای هر مجموعه داده، یک فایل گزارش متنی تهیه خواهید کرد و در آن نتایج مربوط به سوالات و آزمایش‌های انجام شده را که در غالب نمودارها، جداول و یا تصاویر مختلف تهیه نموده‌اید را گزارش خواهید کرد. همچنین، کدهای مربوط به تحلیل‌های انجام شده را نیز به همراه گزارش ارسال خواهید کرد. پیشنهاد می‌گردد که به این منظور از زبان پایتون و کتابخانه‌های مرتبط با تحلیل‌های آماری استفاده نمایید.

مجموعه داده اول

توصیف: مجموعه داده اول مربوط به داده‌های جمع‌آوری شده از خانه‌ای اجاره‌ای برای اقامت کوتاه مدت (Airbnb) در شهر نیویورک آمریکا است و در آن اطلاعاتی در مورد میزبان‌ها، میهمان‌ها، مکان اقامت‌گاه‌ها، زمان و مدت اجاره، قیمت اجاره و ... وجود دارد.

دریافت: برای دریافت این مجموعه داده از لینک زیر استفاده نمایید:

<https://www.kaggle.com/dgomonov/new-york-city-airbnb-open-data>

سوالات:

- What can we learn about different hosts and areas?
- What can we learn from predictions? (ex: locations, prices, reviews, etc)
- Which hosts are the busiest and why?
- Is there any noticeable difference of traffic among different areas and what could be the reason for it?

*شما می‌بایست علاوه بر این سوالات، سوالات ابتکاری متفاوتی را مطرح و استنتاج‌های آماری و آزمون‌های آماری متناسب را انجام و نتایج را ارائه دهید.

مجموعه داده دوم

توصیف: این مجموعه داده مربوط به مسابقات فوتبال بین‌المللی است که در قابل مسابقات جام جهانی، جام‌های قاره‌ای، تورنومنت‌ها، بازی‌های دوستانه و ... انجام شده است. به این منظور اطلاعات مختلفی از قبیل، نام تیم‌ها، محل انجام مسابقه، میزبان و میهمان، زمان بازی، تعداد گل‌ها، نتیجه بازی و ... ذخیره شده‌اند.

دریافت: برای دریافت این مجموعه داده از لینک زیر استفاده نمایید:

<https://www.kaggle.com/martj42/international-football-results-from-1872-to-2017>

سوالات:

- Who is the best team of all time
- Which teams dominated different eras of football

- What trends have there been in international football throughout the ages - home advantage, total goals scored, distribution of teams' strength etc
- Can we say anything about geopolitics from football fixtures - how has the number of countries changed, which teams like to play each other
- Which countries host the most matches where they themselves are not participating in
- How much, if at all, does hosting a major tournament help a country's chances in the tournament
- Which teams are the most active in playing friendlies and friendly tournaments - does it help or hurt them

*شما می‌بایست علاوه بر این سوالات، سوالات ابتکاری متفاوتی را مطرح و استنتاج‌های آماری و آزمون‌های آماری متناسب را انجام و نتایج را ارائه دهید.