

داستان روژه

ما میخواهیم از روی دیتای آگهی‌های فروش خانه در تهران تصمیم بگیریم با توجه به بودجه‌ای که داریم. چه آپارتمانی بخریم تا سود بیشتری داشته باشیم. دیتایی که ما داریم دیتای ۱ ماه در سال ۲۰۲۱ و سال ۲۰۲۳ است و ما بر اساس رشد در این دو سال تصمیم خواهیم گرفت.

دیتاست

برای این بررسی ما از دو دیتاست که از سایت دیوار کراال شده و در kaggle موجود هست استفاده می‌کنیم.

که هر کدام فرمت متفاوت و تفاوت‌هایی در ساختار دیتا دارند. اما چون یکی از دیتاست‌ها [متعلق به سال ۲۰۲۱](#) و یکی [متعلق به سال ۲۰۲۳](#) است، امکان کراال مجدد نداریم و باید نهایت استفاده از این دو دیتاست را داشته باشیم.

خروجی روژه

- یکسان سازی فرمت هر دو دیتاست به صورت CSV و یکسان کردن مقادیر معتبر برای هر فیچر یا ستون.
- مقایسه تغییر قیمتی مناطق مختلف تهران و یافتن مناطق با بیشترین رشد قیمتی
- با بودجه‌های مختلف بهتر است در کدام مناطق خرید انجام دهیم تا سود بیشتری کسب کنیم.
- مناطق با بیشترین تعداد آگهی در سال‌های مختلف
- یافتن مناطق که برای خرید خانه قیمت مناسبی دارند.
- ارائه مراحل پروژه و پیاده سازی و نمودارها با storytelling مناسب

امتیاز بیشتر

بر اساس آگهی‌های اجاره آپارتمان و خرید آپارتمان در دیتاست ۲۰۲۱ ، به صرفه بودن اجاره در محله‌های مختلف را بررسی کنید.

تفاوت دیتاست ها

در دیتاست ۲۰۲۳ ستونها به شرح زیر است:

Area : Apartment area (square meters) / مساحت ملک (متر مربع)

Construction: Year of construction / سال ساخت

Room: Number of Room / تعداد اتاق خواب

Warehouse: Binary Variables: This is True if the property has Warehouse / انباری
دارد یا خیر؟

Parking: Binary Variables: This is True if the property has Parking / پارکینگ دارد یا
خیر؟

Elevator (lift): This is True if the property has an Elevator / آسانسور دارد یا خیر؟

Address : Location / آدرس محله

Price : in Toman (Iranian Currency) / قیمت

به تومان

در دیتاست ۲۰۲۱ اما ما همه آگهی‌های ملک رو داریم (مغازه، خانه، آپارتمان و خرید یا اجاره هر کدام) که شما باید فقط دی ای " apartment-sell " رو جدا کنید.

- خروجی هر فیچر ممکن است متفاوت باشد که باید یکسان کنید مثل قیمت که به عدد فارسی یا انگلیسی ذخیره شده.(به خصوص در مورد محله ها در دیتاست ۲۰۲۳ که باید ایده بزنید)
- اسم فیچرها هم ممکنه متفاوت باشه و خودتون باید تشخیص بدید.
- دیتاست ۲۰۲۱ به فرمت json هستش که باید تبدیل به CSV بشود.