



سوال ۱: تحلیل سری‌های زمانی فروش و پیش‌بینی روند با رولینگ میانگین

- مرحله ۱: خواندن فایل و آماده‌سازی داده‌ها

- ۱- فایل [daily_sales.csv](#) را با استفاده از تابع `pandas.read_csv()` بخوانید و آن را در یک `DataFrame` ذخیره کنید.
- ۲- بررسی کنید که فایل شامل دو ستون زیر باشد:

`Date`: تاریخ فروش (به فرمت YYYY-MM-DD)

`Sales`: مقدار فروش روزانه (عدد صحیح)

- ۳- ستون `Date` را به نوع تاریخ (`datetime`) تبدیل کنید.
- ۴- ستون `Date` را به عنوان ایندکس (`index`) دیتافریم تنظیم کنید.

- مرحله ۲: محاسبه میانگین متحرک ۷ روزه

- ۱- با استفاده از متد `rolling()`، میانگین متحرک ۷ روزه برای ستون `Sales` را محاسبه کنید.
- ۲- نتیجه را به عنوان یک ستون جدید با نام `RollingMean` به دیتافریم اضافه کنید.

- مرحله ۳: ترسیم نمودار

- ۱- با استفاده از کتابخانه `matplotlib` یک نمودار خطی رسم کنید.
 - ۲- در این نمودار:
- فروش اصلی (`Sales`) را با یک خط آبی یا مشکی نمایش دهید.
- فروش هموارشده با میانگین متحرک (`RollingMean`) را با یک خط قرمز نمایش دهید.
- ۳- عنوان نمودار را تنظیم کنید: `Daily Sales vs 7-Day Moving Average`
 - ۴- محورهای `X` و `Y` را برچسب‌گذاری کرده و `legend` اضافه کنید.

- مرحله ۴: تحلیل داده‌ها (به صورت نوشتاری یا گفتاری پاسخ دهید)

- ۱- در کدام بازه‌های زمانی فروش افزایش یا کاهش ناگهانی داشته است؟ (راهنما: به نقاطی نگاه کنید که نمودار فروش از میانگین متحرک فاصله زیادی می‌گیرد)
- ۲- آیا روند کلی فروش در بازه‌ی زمانی داده‌شده صعودی است یا نزولی؟ (راهنما: به شیب کلی نمودار میانگین متحرک نگاه کنید)
- ۳- آیا نوسانات فروش زیاد است یا نسبتاً پایدار؟

سوال ۲: تحلیل چندبعدی داده‌های مسکن با نمودارهای جفتی (Pairplot)

- مرحله ۱: خواندن فایل و بررسی داده‌ها

۱. فایل [housing_data.csv](#) را با استفاده از تابع `pandas.read_csv` بخوانید و آن را در یک `DataFrame` ذخیره کنید.

۲. بررسی کنید که فایل شامل ستون‌های زیر باشد:

- Price: قیمت خانه (عدد)
- Area: مساحت خانه (متر مربع، عدد)
- Bedrooms: تعداد اتاق خواب (عدد صحیح)
- Bathrooms: تعداد سرویس بهداشتی (عدد صحیح)
- LocationScore: امتیاز مکان (عدد بین ۱ تا ۱۰)

مطمئن شوید که تمام این ستون‌ها از نوع داده‌ی عددی (`float` یا `int`) هستند.

- مرحله ۲: ترسیم نمودار جفتی (Pairplot)

۱. از کتابخانه‌ی `seaborn` برای رسم نمودار جفتی استفاده کنید. (این نمودار به شما کمک می‌کند تا روابط بین جفت‌های مختلف از ویژگی‌های

عددی را به صورت تصویری بررسی کنید)

- مرحله ۳: تحلیل داده‌ها (پاسخ نوشتاری یا گفتاری)

پس از بررسی نمودار `pairplot`، به سوالات زیر به صورت توصیفی پاسخ دهید:

۱. آیا بین قیمت (`Price`) و مساحت (`Area`) رابطه‌ی خطی دیده می‌شود؟

۲. آیا تعداد اتاق‌ها (`Bedrooms` و `Bathrooms`) با قیمت رابطه مشخصی دارند؟

۳. کدام ویژگی‌ها بیشترین همبستگی با قیمت دارند؟

(راهنما: به چگالی نقطه‌ها در نمودارهای `scatter` و الگوهای خطی نگاه کنید)

سوال ۳: تحلیل ناهنجاری (Outlier Detection) در داده‌های فروش

شناسایی محصولات با فروش غیرعادی (بسیار بالا یا بسیار پایین) با استفاده از تحلیل آماری و نمایش تصویری با **Box Plot**

- مرحله ۱: خواندن و بررسی داده‌ها

۱. فایل [product_sales.csv](#) را با استفاده از pandas بخوانید.

این فایل باید شامل دو ستون زیر باشد:

- Product: نام یا شناسه‌ی محصول

- Sales: مقدار فروش محصول (عددی)

۲. بررسی کنید که ستون Sales از نوع عددی باشد.

- مرحله ۲: محاسبه میانگین، انحراف معیار و شناسایی نقاط پرت

۱. میانگین و انحراف معیار فروش را محاسبه کنید.

۲. تعریف ناهنجاری: محصولاتی که فروش آن‌ها بیش از ۲ انحراف معیار بالاتر یا پایین‌تر از میانگین هستند.

۳. فیلتر کردن نقاط پرت.

۴. چاپ یا نمایش محصولات ناهنجار

- مرحله ۳: نمایش بصری با **Box Plot**

۱. از Seaborn یا Matplotlib برای رسم نمودار Box Plot استفاده کنید.

- مرحله ۴: تحلیل و نتیجه‌گیری (پاسخ نوشتاری یا گفتاری)

۱. چند محصول فروش ناهنجار (Outlier) داشته‌اند؟

۲. آیا این ناهنجاری‌ها ممکن است به دلیل خطای ثبت داده باشند؟

۳. یا این ناهنجاری‌ها نشان‌دهنده پرفروش‌ترین یا کم‌فروش‌ترین محصولات هستند؟

۴. چه اقداماتی برای بررسی دقیق‌تر این موارد می‌توان انجام داد؟

سؤال ۴: مصورسازی ویژگی‌ها (Feature Visualization)

۱. در این تمرین می‌خواهیم ویژگی‌های چند میوه را به صورت بصری روی نمودار دوبعدی نمایش دهیم، به شکلی که اطلاعات عددی و کیفی (رنگ،

وزن، شیرینی) در کنار هم تحلیل پذیر باشند.

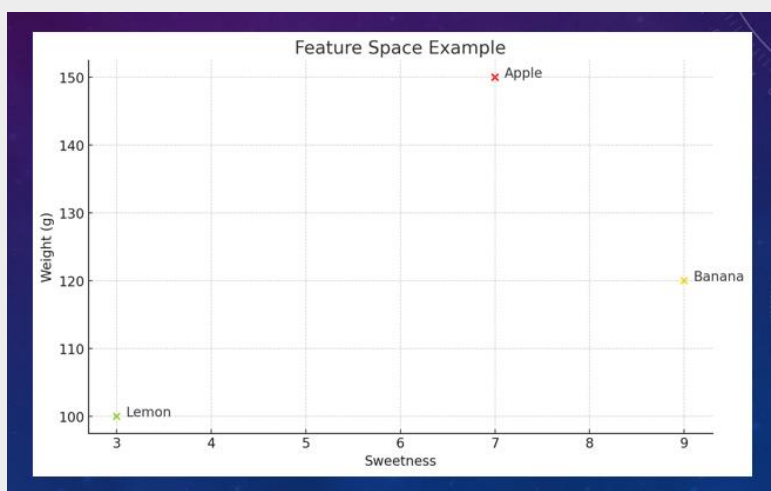
جدول زیر اطلاعات سه میوه مختلف را نشان میدهد:

Fruit	Color	Weight	Sweetness
Apple	Red	150g	7
Banana	Yellow	120g	9
Lemon	Yellow	100g	3

۱- این جدول را به صورت یک لیست از دیکشنری‌ها یا دیتافریم در پایتون تعریف کنید

۲- مقدار ستون **Weight** را به عدد تبدیل کنید (بدون g).

۳- با استفاده از **matplotlib** و **pandas** مصورسازی‌ای انجام دهید تا خروجی زیر حاصل شود.



(راهنمایی:

- محور افقی (X): درجه شیرینی (Sweetness)

محور عمودی (Y): وزن میوه (Weight)

رنگ نقطه‌ها بر اساس رنگ واقعی میوه‌ها باشد

برای هر میوه، یک نقطه روی نمودار رسم کنید که:

- موقعیت آن توسط ویژگی‌های عددی (Sweetness و Weight) مشخص می‌شود.

- رنگ آن از ستون Color گرفته شود.

نام میوه‌ها (Apple, Banana, Lemon) را به عنوان برچسب (label) کنار یا روی نقطه‌ها نمایش دهید.

سوال ۵: گسترش داده و رسم چندنمونه‌ای بر اساس ویژگی‌ها

-مرحله ۱: تولید داده‌های جدید برای هر میوه

با توجه به جدول اولیه در سؤال قبل، برای هر میوه (Lemon, Banana, Apple) ۳ نمونه‌ی جدید به صورت دستی ایجاد کنید.

نکته: مقادیر باید در محدوده‌ای منطقی و نزدیک به داده‌ی اصلی میوه باشند. (مثلاً Apple حدود ۱۵۰ g و شیرینی حدود ۷)

- مرحله ۲: اضافه کردن داده‌های جدید به دیتافریم قبلی

۱. داده‌های جدید را به دیتافریم قبلی اضافه کنید (می‌توانید لیست دیکشنری‌ها را به DataFrame تبدیل کرده و به قبلی **append** یا **concat** کنید).

۲. مطمئن شوید که تمام ستون‌ها (Fruit, Color, Weight, Sweetness) در دیتافریم نهایی وجود دارند.

-مرحله ۳: رسم نمودار نهایی

۱. همانند سوال قبل، یک نمودار Scatter رسم کنید که:

- محور X: درجه شیرینی (Sweetness)
- محور Y: وزن میوه (Weight)
- رنگ نقطه‌ها بر اساس رنگ میوه (Color)
- نام میوه‌ها به عنوان برچسب (label یا text)

۲. تفاوت جدید:

- از شکل (marker) متفاوت برای داده‌های اصلی و داده‌های جدید استفاده کنید:

- مثلاً داده‌های اولیه با 'x'
- داده‌های جدید با مثلث '^' یا ستاره '*' (دلخواه)

۳. یک **legend** اضافه کنید که مشخص کند کدام نقاط داده‌ی اولیه هستند و کدام داده‌ی جدید.

سوال ۶: افزودن یک ویژگی دلخواه و رسم نمودار سه بعدی

در این تمرین می‌خواهیم یک ویژگی (Feature) جدید به داده‌های میوه‌ها اضافه کنیم و با استفاده از آن، داده‌ها را در یک فضای سه بعدی مصور کنیم.

مرحله ۱: افزودن ویژگی جدید (Feature Engineering)

۱. یک ویژگی عددی جدید به دلخواه خودتان طراحی و به جدول میوه‌ها اضافه کنید.

۲. این ویژگی می‌تواند یکی از موارد زیر باشد (یا هر ویژگی خلاقانه‌ی دیگر):

- **Juiciness** (آبدار بودن) → عددی بین ۱ تا ۱۰

- **Crunchiness** (ترد بودن)

- **Ripeness** (میزان رسیدگی)

- یا هر ویژگی دیگر که با میوه‌ها مرتبط باشد.

برای هر نمونه (چه اولیه و چه نمونه‌های جدید)، مقدار این ویژگی را تعیین کنید.

توجه: ویژگی انتخابی باید عددی (**float** یا **int**) باشد تا امکان نمایش در نمودار سه بعدی وجود داشته باشد.

مرحله ۲: رسم نمودار سه بعدی

مراحل زیر را انجام دهید:

۱. یک نمودار سه بعدی رسم کنید که:

- محور **X**: ویژگی **Sweetness**

- محور **Y**: ویژگی **Weight**

- محور **Z**: ویژگی جدیدی که خودتان اضافه کرده‌اید.

- رنگ نقاط بر اساس ستون **Color** باشد (رنگ واقعی میوه)

۲. نمودار را طوری تنظیم کنید که:

- عنوان مناسب داشته باشد.

- محورهای **X**، **Y** و **Z** برچسب‌گذاری شده باشند.

سوال ۷: تبدیل داده‌های متنی به داده‌های عددی قابل تحلیل

HasAwards	Rating (1-10)	Duration (min)	Genre	Movie
Yes	8.8	148	Sci-Fi	Inception
Yes	8.3	81	Animation	Toy Story
No	6.9	130	Action	Fast & Furious

- ۱- این جدول را به صورت یک لیست از دیکشنری‌ها یا دیتافریم در پایتون تعریف کنید
- ۲- اطلاعات ۷ فیلم دلخواه را از طریق سرچ در اینترنت استخراج کنید و به دیتا فریم خود اضافه کنید.
- ۳- ستون HasAwards را به عدد تبدیل کنید (Yes \rightarrow 1, No \rightarrow 0).
- ۴- ستون Genre را به یکی از روش‌های زیر تبدیل کنید:

Label Encoding (Sci-Fi \rightarrow 0, Animation \rightarrow 1, Action \rightarrow 2, ...) یا One-Hot Encoding

سوال ۸: (فیچر شرطی - ساخت ستون جدید با if/else)

ساخت ویژگی جدید با شرط ساده

با توجه به ستون‌های Rating و Duration، یک ستون جدید به نام LikelyToWinAward بسازید.
این ستون با استفاده از قانون زیر مقادری می‌شود:
اگر Rating بیشتر از ۸ و Duration بیشتر از ۱۰۰ باشد، مقدار ستون برابر با ۱ باشد.
در غیر این صورت، مقدار ستون برابر با ۰ قرار گیرد.

سوال ۹: تحلیل پیش‌بینی جایزه با ستون LikelyToWinAward

با استفاده از ستون LikelyToWinAward که در سوال قبل ساختید، موارد زیر را بررسی کنید:

۱. چند فیلم از نظر این قانون، "احتمال برنده‌شدن جایزه" را دارند؟
۲. چند مورد از این پیش‌بینی‌ها با مقدار واقعی HasAwards مطابقت دارد؟

سوال ۱۰: مصورسازی پیش‌بینی جایزه

با استفاده از کتابخانه‌ی matplotlib یا seaborn، یک نمودار برای مقایسه‌ی پیش‌بینی جایزه با مقدار واقعی HasAwards رسم کنید.

۱. یک نمودار میله‌ای یا countplot رسم کنید که در آن:

- مقدار ستون LikelyToWinAward (پیش‌بینی)
- در برابر مقدار واقعی HasAwards

۲- تحلیل کنید:

- آیا پیش‌بینی‌ها با واقعیت هماهنگ هستند؟
- چه فیلم‌هایی اشتباه پیش‌بینی شده‌اند؟
- آیا تغییر در شرط پیش‌بینی می‌تواند دقت را افزایش دهد؟