



AI

pandas & matplotlib

1. فایل تمرین را در پنل خود آپلود کنید. i

2. title فایل تمرین به صورت (نام تمرین+نام و نام خانوادگی) به انگلیسی باشد.

3. فایل های مربوط به تمرین را از این [لينك](#) دانلود کنید.

4. در صورتی که سوال و یا ابهامی دارید در گروه چت تلگرامی بپرسید.

۱. آماده سازی و پردازش داده ها با Pandas

هدف: تسلط بر ادغام دیتاستها (Merging)، کار با سری های زمانی (Time Series) و جداول محوری

.(Pivot Tables)

فایل های داده:

•: سوابق تمام تراکنش ها، شامل میزان فروش (Sales) و سود (Profit).orders.csv

•: جزئیات کالاهای فروخته شده (مانند دسته بندی و نام کالا).products.csv

•: اطلاعات دموگرافیک مشتریان (مانند منطقه و نوع مشتری).customers.csv

مراحل انجام کار:

1. بارگذاری و بررسی اولیه:

- هر سه فایل products.csv، customers.csv و orders.csv را در سه DataFrame جداگانه بارگذاری کنید.
- با استفاده از .head() و .shape ساختار هر دیتافریم را بررسی کنید تا متوجه شوید چه ستون‌هایی دارند.

2. ادغام:

- جدول orders فقط شامل شناسه (ID) است و جزئیات ندارد. ما باید اطلاعات را به آن اضافه کنیم.
- گام اول: دیتافریم orders را با products بر اساس ستون ProductID ادغام کنید.
- گام دوم: نتیجه را با دیتافریم customers بر اساس ستون CustomerID ادغام کنید.
- نتیجه نهایی را در متغیری به نام df_master ذخیره کنید. ابعاد آن را چک کنید تا مطمئن شوید داده‌ای از دست نرفته است.

3. مهندسی ویژگی‌ها:

- قیمت واحد: در دیتاست، ستون Sales نشان‌دهنده درآمد کل آن تراکنش است. ستون جدیدی به نام UnitPrice بسازید که حاصل تقسیم Sales بر Quantity باشد.
- حاشیه سود: برای محاسبه کارایی فروش، ستون جدیدی به نام Margin بسازید که حاصل تقسیم Profit بر Sales باشد.

4. تحلیل سری‌های زمانی:

- ستون OrderDate را به فرمت datetime تبدیل کنید.
- ستون OrderDate را به عنوان Index دیتافریم df_master تنظیم کنید.

- با استفاده از متدها (resample('ME')), دادهها را به صورت ماهانه جمع‌بندی کنید و مجموع Sales و Profit را محاسبه کنید.
- سؤال: دقیقاً کدام ماه (سال-ماه) بیشترین میزان فروش را داشته است؟

5. بخش‌بندی مشتریان (Pivot Tables):
- یک Pivot Table بسازید تا سودآوری را بر اساس نوع مشتری تحلیل کنید.
 - ستون Segment (مانند Consumer, Corporate)
 - ستون Category (مانند Furniture, Technology)
 - ستون Values (مانند Profit)
 - تابع Aggfunc (مانند .sum)

سؤال: کدام نوع مشتری (Segment) بیشترین سود را در دسته تکنولوژی (Technology) ایجاد کرده است؟

۲. مصورسازی داده‌ها با Seaborn و Matplotlib

- هدف: کشف روندها و داده‌های پرت با استفاده از نمودارها.
- پیش‌نیاز: از دیتافریم df_master که در بخش قبل ساختید استفاده کنید.
- مراحل انجام کار:
1. روند فروش در برابر سود:
 - روی داده‌های df_master، مجدداً نمونه‌برداری ماهانه (resample) انجام دهید تا مجموع فروش و سود ماهانه را داشته باشید.

- با استفاده از plt.plot() یک نمودار خطی بکشید.
- فروش (Sales) را با رنگ آبی و سود (Profit) را با رنگ سبز روی همان نمودار رسم کنید.
- برای نمودار خود عنوان ("Monthly Sales vs Profit")، برچسب محورها و راهنمای قرار دهید.
- آیا فروش و سود همیشه با هم حرکت می‌کنند؟ یا ماههایی وجود دارد که فروش بالاست اما سود پایین است؟

2. بررسی عملکرد دسته‌بندی‌ها:

- با استفاده از sns.boxplot() توزیع Profit را برای Category‌های مختلف رسم کنید.
- چالش: ممکن است داده‌های پرت نمودار را فشرده کنند. با استفاده از plt.ylim(-50, 100) روی بخش اصلی نمودار زوم کنید.
- کدام دسته‌بندی بیشترین تراکنش‌های با سود منفی را دارد؟

3. تأثیر تخفیف‌ها (Correlation Heatmap):

- ستون‌های عددی زیر را انتخاب کنید: Sales, Quantity, Profit, Margin
- ماتریس همبستگی (Correlation Matrix) را با استفاده از متدها corr() و .corr() محاسبه کنید.
- این ماتریس را با استفاده از sns.heatmap() رسم کنید (آرگومان annot=True را فعال کنید تا اعداد نمایش داده شوند).
- به همبستگی بین Discount (تخفیف) و Profit (سود) دقت کنید. آیا این رابطه مثبت است یا منفی؟ این چه چیزی را در مورد استراتژی تخفیف‌دهی فروشگاه نشان می‌دهد؟