

تمرین اول: محاسبه CLV و تحلیل RFM

معرفی مجموعه داده

فایل اکسل پیوست حاوی داده‌های فروشگاه زنجیره‌ای مواد غذایی Mondrian است. این داده‌ها مربوط به تراکنش‌های خرید (۳۲۵۹۱ رکورد) مربوط به ۱۰۲۳ مشتری در بازه زمانی ابتدای سال ۱۹۹۷ تا انتهای سال ۱۹۹۸ است. همان‌طور که مشاهده می‌کنید این فایل از ۵ جدول تشکیل شده است. این جداول عبارتند از:

Sales Data, Customer information, Product information, Time information و Store Information

جدول اصلی که اطلاعات خرید مشتریان در آن ثبت شده است، جدول "Sales Data" می‌باشد که مربوط به ثبت تراکنش‌های خرید مشتریان و شامل ۳۲۹۵۰ رکورد است. سایر جداول بیش‌تر جنبه توضیحی داشته و اطلاعات مربوط به متغیرها و مشخصه‌های این جدول را شامل می‌شوند. مشخصه‌ها و متغیرهایی که در این جدول ثبت شده است عبارتند از customer_id, product_id, time_id, store_id, store_sales, store_cost و unit_sales که توضیحات مربوط به آنها در جدول ۱ آمده است.

جدول ۱: اطلاعات مشخصه‌های جدول Sales data

ردیف	نام مشخصه (متغیر)	تعریف مشخصه (متغیر)	توضیحات
۱	customer_id	شناسه مشتری	با استفاده از این مشخصه و اطلاعات جدول Customer Information می‌توان به اطلاعات مربوط به هر یک از این مشتریان از جمله اطلاعات دموگرافیک، جغرافیایی و... دست یافت.
۲	product_id	شناسه محصول	با استفاده از این مشخصه و اطلاعات جدول Product Information می‌توان به اطلاعات مربوط به هر یک از محصولات که توسط مشتریان خریداری شده است، دست یافت. این جدول به دلیل ایجاد دیدگاه کامل‌تری نسبت به محصولات این فروشگاه مواد غذایی ارائه شده است.
۳	time_id	شناسه زمان	با استفاده از این مشخصه و اطلاعات جدول Time Information می‌توان به اطلاعات دقیق مربوط به زمانی که هر یک از تراکنش‌ها انجام شده، دست یافت.

۴	store_id	شناسه شعبه فروشگاه	با استفاده از این مشخصه و اطلاعات جدول Store Information می توان به اطلاعات دقیق مربوط به شعب مختلف که خرید از آن ها صورت گرفته است، دست یافت.
۵	store_sales	میزان فروش حاصل از خرید مشتری در تراکنش مربوطه	
۶	store_cost	هزینه تحمیلی در تراکنش مربوطه	
۷	unit_sales	تعداد محصول خریداری شده	

مبتنی بر این داده تحلیل های ذیل را انجام دهید.

الف) با استفاده از فرمول زیر که یکی از پایه ترین مدل های ارائه شده برای محاسبه CLV است، مقدار ارزش در حال حاضر (Current value) را برای هر یک از مشتریان محاسبه کنید. سپس مقدار CE (Customer Equity) را نیز محاسبه نمایید.

$$CLV = \sum_{i=1}^n \frac{(R_i - C_i)}{(1 + d)^{i-0.5}}$$

در این فرمول i دوره جریان مالی از تراکنش های مشتری را نشان می دهد. R_i درآمد حاصل از مشتری در دوره i ، C_i هزینه کلی تولید درآمد R_i در دوره i و n تعداد دوره های بررسی شده در طول عمر مشتری است. بازه های زمانی را به صورت ماهانه و نرخ تنزیل را برابر با ۲۰٪ در نظر بگیرید.

ب) اصل پارتو (Pareto principle) را بررسی نمایید. به عبارتی ۲۰٪ مشتریانی که ۸۰٪ CE را حاصل می کنند، شناسایی کنید. آیا الگویی در زمینه ویژگی های غالب این بخش از مشتریان وجود دارد؟ می توانید از تحلیل های آمار توصیفی برای مشخصه های مختلف فراهم شده از مشتریان در داده از جمله مشخصه های جمعیت شناختی و جغرافیایی استفاده کنید.

ج) در ابتدا مقادیر R ، F و M را برای هر یک از مشتریان محاسبه کنید. در محاسبه مشخصه های فوق، انتهای سال ۱۹۹۸ را به عنوان زمان تحلیل در نظر بگیرید. مشتریان را بر اساس این سه مشخصه به صورت زیر بخش بندی نمایید:

مقدار میانگین هر یک از این سه مشخصه را محاسبه کرده و بر اساس مقدار هر یک از این مشخصه ها برای هر یک از مشتریان، مقدار High و یا Low را به آن اختصاص دهید. اگر مقدار مشخصه مورد نظر برای یک مشتری از میانگین بیشتر باشد، مقدار High و اگر کمتر باشد، مقدار Low را اختصاص دهید. بر این اساس ۸ گروه مختلف از مشتریان به صورت زیر امکان پذیر است.

ردیف	بخش های مختلف ممکن
۱	$R \downarrow F \uparrow M \uparrow$
۲	$R \downarrow F \uparrow M \downarrow$
۳	$R \downarrow F \downarrow M \uparrow$
۴	$R \downarrow F \downarrow M \downarrow$
۵	$R \uparrow F \uparrow M \uparrow$
۶	$R \uparrow F \downarrow M \uparrow$
۷	$R \uparrow F \uparrow M \downarrow$
۸	$R \uparrow F \downarrow M \downarrow$

High با علامت \uparrow و Low با علامت \downarrow نشان داده شده است.

بر این اساس بخش های موجود در مشتریان را شناسایی کرده، تعداد مشتریان هر یک از آنها را استخراج کرده و متوسط CLV را برای هر یک از این بخش ها محاسبه کنید. توجه کنید که مقدار CLV برای انتهای سال ۹۸ محاسبه شود. می توانید وزن های سه متغیر R، F و M را به ترتیب برابر با ۰,۱۵، ۰,۳۵ و ۰,۵۰ در نظر بگیرید.

(۵) هر یک از بخش های به دست آمده در قسمت ج را توصیف کرده و تحلیلی از هر یک از آنها بر مبنای مشخصه های مشتریان آنها ارائه دهید. می توانید از تحلیل های آمار توصیفی برای مشخصه های مختلف فراهم شده از مشتریان در داده از جمله مشخصه های جمعیت شناختی و جغرافیایی استفاده کنید. در واقع هدف اصلی این بخش از تمرین، توصیف و استخراج ویژگی های غالب هر یک از بخش های مشتریان است.