

بسم الله الرحمن الرحيم

پروژه بخش بندی کاربران خرده فروشی

رضا رحمتی

زمستان 1403

فهرست مطالب

مقدمه.....	۵
پاکسازی داده‌ها.....	۵
تبدیل داده‌ها.....	۹
پیاده‌سازی مدل RFM.....	۱۰
پیاده‌سازی الگوریتم k-means.....	۱۲
تحلیل خوشه‌های مشتریان.....	۱۴
نتیجه‌گیری.....	۱۷

فهرست نمودارها

- نمودار ۱ بررسی توزیع قیمت و مبلغ واحد در تراکنش‌ها ۷
- نمودار ۲ تعداد تراکنش‌ها در طول زمان ۸
- نمودار ۳ تعداد تراکنش‌ها بر حسب کشور ۸
- نمودار ۴ روش Elbow برای مشخص کردن مقدار k بهینه ۱۳
- نمودار ۵ استفاده از Silhouette Score برای مشخص کردن مقدار k بهینه ۱۴
- نمودار ۶ مقدار Recency برای هر خوشه ۱۵
- نمودار ۷ مقدار Frequency برای هر خوشه ۱۵
- نمودار ۸ مقدار Monetary برای هر خوشه ۱۶

فهرست جداول

جدول ۱ داده‌های آماری تراکنش‌ها.....	۶
جدول ۲ محاسبات Cohort.....	۹
جدول ۳ درصد مشتریان موجود در هر گروه.....	۱۱

مقدمه

در دنیای رقابتی امروز، تحلیل رفتار مشتریان یکی از مهم‌ترین استراتژی‌های کسب‌وکارها برای بهبود عملکرد و افزایش سودآوری است. مدل RFM یکی از روش‌های محبوب و اثبات‌شده در حوزه تقسیم‌بندی مشتریان (Customer Segmentation) است که بر اساس سه معیار کلیدی مشتریان را ارزیابی می‌کند:

- تازگی (Recency): تعداد روزهایی که از آخرین خرید مشتری گذشته است.
- تکرار خرید (Frequency): دفعاتی که مشتری در یک بازه زمانی مشخص خرید انجام داده است.
- مقدار پول خرج شده (Monetization): مجموع مبلغی که مشتری در بازه مشخص خرج کرده است.

هدف از اجرای این تحلیل، شناسایی مشتریان ارزشمند، وفادار، در معرض خطر، و مشتریان از دست‌رفته است تا کسب‌وکار بتواند تصمیمات استراتژیک بهتری برای بازاریابی، حفظ مشتریان و افزایش فروش اتخاذ کند.

در این پروژه، از مجموعه داده‌ای شامل تراکنش‌های یک فروشگاه آنلاین مستقر در انگلستان که بین دسامبر ۲۰۱۰ تا دسامبر ۲۰۱۱ ثبت شده است، استفاده شده است. مراحل انجام تحلیل شامل پاک‌سازی داده‌ها، ایجاد مدل RFM، خوشه‌بندی مشتریان با الگوریتم K-Means و تحلیل نتایج است.

در ادامه، ابتدا داده‌ها پردازش و استانداردسازی شده، سپس تحلیل‌های آماری و مدل RFM اجرا گردیده و در نهایت تقسیم‌بندی مشتریان با الگوریتم K-Means انجام خواهد شد تا الگوهای رفتاری مشتریان مشخص شود.

پاکسازی داده‌ها

برای شروع داده‌های از نظر نوع و معتبر بودن به صورت مختصر بررسی می‌کنیم. داده‌ها تا حد خوبی مرتب و قابل استفاده هستند. اما این اطلاعات خام برای تراکنش‌ها هستند و موارد زیر نیاز به اصلاح دارند:

1. تعدادی از تراکنش‌ها به هیچ کاربری مرتبط نیستند. با توجه به اینکه مدل‌های ما حول رفتار کاربران ایجاد می‌شوند، حذف این رکوردها بهترین اقدام برای آن‌ها است.
2. تراکنش‌های کنسل شده، یا مواردی که مربوط به بازگشت وجه به دلایل مختلف هستند و مقدار یا هزینه صفر یا منفی دارند به عنوان خرید کاربر در نظر گرفته نمی‌شوند. بنابراین این دسته از اطلاعات را هم برای تحلیل‌ها در نظر نمی‌گیریم.
3. داده‌های تکراری که احتمالاً به علت خطا در عملکرد برنامه‌ها ایجاد شده‌اند باید فقط یکبار دیده شوند و فقط یکبار در نظر گرفته شوند.
4. بهتر است برای تحلیل مبلغ کل خرید مشتریان در هر تراکنش یک ستون به داده‌ها اضافه کنیم که از ضرب تعداد در مبلغ واحد در هر تراکنش استفاده کنیم.

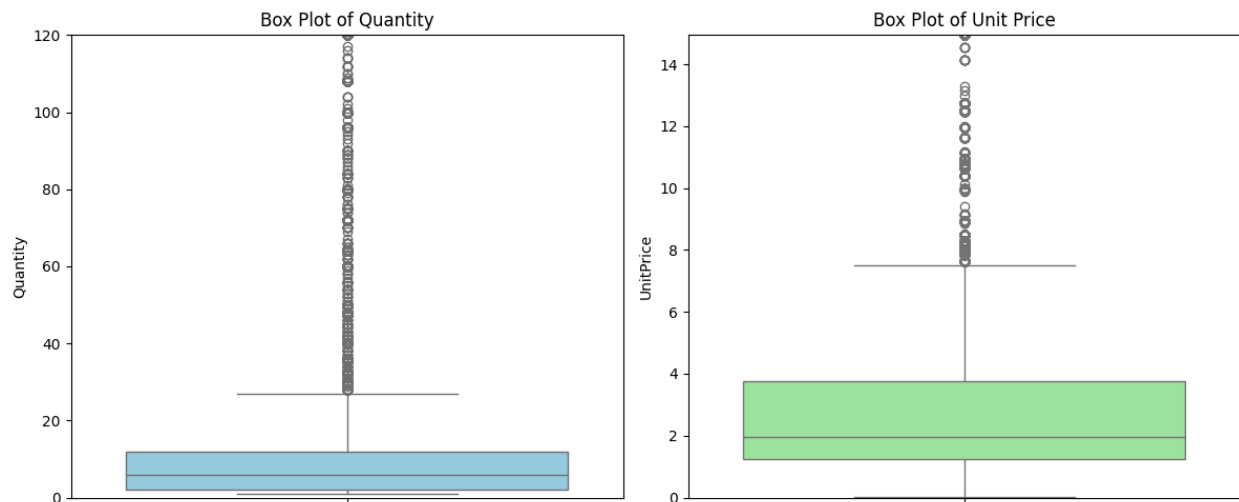
برای آشنایی کیفی با داده‌ها، کمیت‌های آماری مربوط به هر ستون از جدول داده‌ها را بدست می‌آوریم:

جدول 1 داده‌های آماری تراکنش‌ها

	Quantity		InvoiceDate	UnitPrice	CustomerID	TotalSum
count	276190.000000		276190	276190.000000	276190.000000	276190.000000
mean	13.202719	2011-07-10 13:51:30.167928064		3.117113	15290.353670	22.764697
min	1.000000	2010-12-01 08:26:00		0.001000	12346.000000	0.001000
25%	2.000000	2011-04-06 15:47:00		1.250000	13969.000000	4.950000
50%	6.000000	2011-07-29 16:24:00		1.950000	15153.000000	11.900000
75%	12.000000	2011-10-20 12:33:15		3.750000	16791.000000	19.800000
max	80995.000000	2011-12-09 12:50:00		4161.060000	18287.000000	168469.600000
std	213.415620	NaN		19.736248	1712.161155	367.237241

با توجه به جدول 1، میانگین تعداد خرید در هر سفارش حدود 13 کالا و میانگین قیمت هر واحد محصول 3.11 واحد پولی است. برخی سفارش‌ها بسیار بزرگ هستند، مانند سفارشی با 80,995 کالا، و برخی محصولات قیمت‌های بسیار بالایی دارند که نشان‌دهنده وجود مقادیر پرت است. انحراف معیار بالا در تعداد خرید و قیمت نشان می‌دهد که توزیع داده‌ها پراکندگی زیادی دارد و نیاز به بررسی مقادیر پرت دارد. بعد از بررسی داده‌هایی که قیمت یا تعداد خیلی بزرگ دارند به نظر می‌رسد که داده‌ها درست باشند و به علت تنوع بالای محصولات این خرده‌فروشی، دیدن داده‌هایی با مقادیر متفاوت طبیعی است اما هنگام استفاده از الگوریتم k-mean باید با روش‌هایی مانند تبدیل لگاریتمی یا حذف داده‌های پرت عملکرد الگوریتم را بالا ببریم.

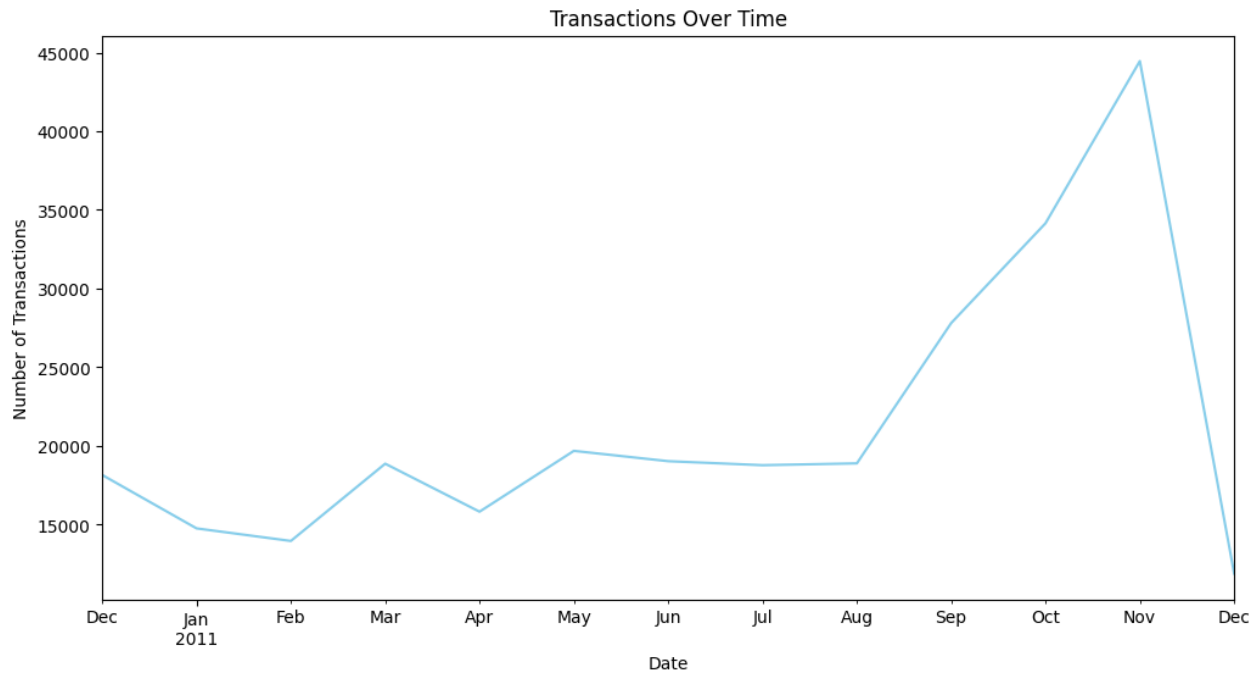
برای توصیف بهتر توزیع قیمت و تعداد در تراکنش‌ها از نمودار 1 استفاده می‌کنیم. همانگونه که مشاهده می‌شود داده‌ها چولگی بالایی دارند و با توزیع نرمال تفاوت دارند.



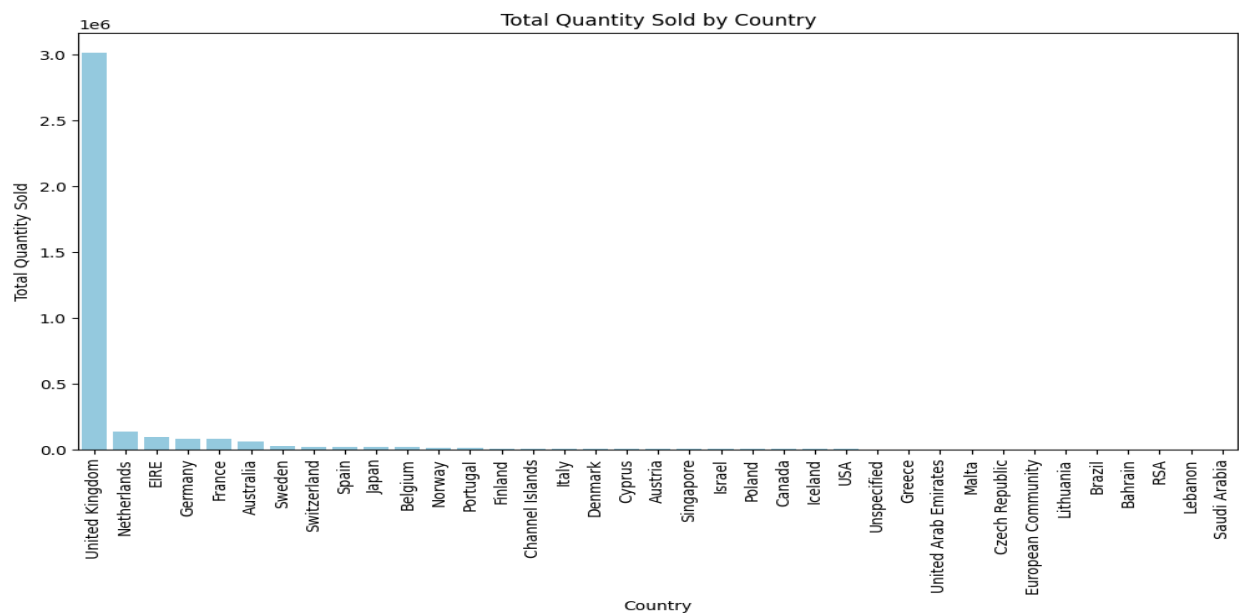
نمودار ۱ بررسی توزیع قیمت و مبلغ واحد در تراکنش‌ها

بررسی تعداد تراکنش‌ها در طول زمان می‌تواند برای تحلیل کسب و کار جالب باشد که در نمودار ۲ آورده شده است. این نمودار نشان می‌دهد که تعداد تراکنش‌ها در اوایل سال ۲۰۱۱ نسبتاً پایین بوده و در ماه‌های ابتدایی یک روند نزولی داشته است. از ماه مارس تا آگوست تغییرات زیادی در تعداد تراکنش‌ها مشاهده نمی‌شود و میزان فروش تقریباً ثابت باقی مانده است. از ماه سپتامبر به بعد، تعداد تراکنش‌ها افزایش چشمگیری داشته و در نوامبر به اوج خود رسیده است. این افزایش احتمالاً به دلیل فصل خرید یا تخفیفات ویژه بوده است. با این حال، در دسامبر تعداد تراکنش‌ها به شدت کاهش یافته که ممکن است ناشی از تعطیلات پایان سال یا کاهش تقاضا باشد.

از سوی دیگر، با بررسی مکان هر تراکنش در نمودار ۳ مشاهده می‌شود که بیشتر تراکنش‌ها در انگلستان انجام شده‌اند. با توجه به اینکه شرکت نیز انگلیسی است، این موضوع قابل پیش‌بینی بود.



نمودار ۲ تعداد تراکنش‌ها در طول زمان

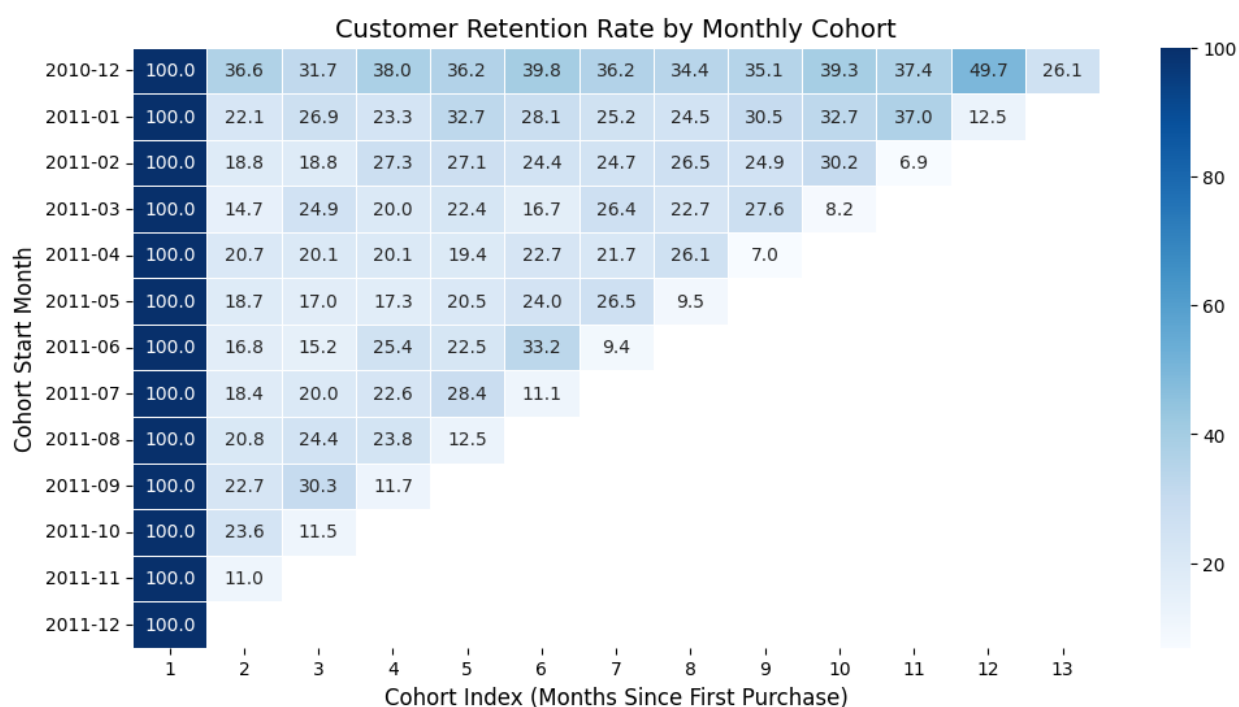


نمودار ۳ تعداد تراکنش‌ها بر حسب کشور

تبدیل داده‌ها

به صورت کلی در داده‌های کوهورت هر گروه مشتریان با گذشت زمان به تدریج کاهش پیدا می‌کند، که این یک الگوی طبیعی در تحلیل رفتار مشتریان است. در ماه اول، هر کوهورت با تعداد مشخصی از مشتریان آغاز می‌شود، اما در ماه‌های بعدی، بخشی از مشتریان دیگر خریدی انجام نمی‌دهند و نرخ ماندگاری کاهش می‌یابد. این کاهش برای برخی گروه‌ها سریع‌تر از بقیه رخ می‌دهد، که می‌تواند به عوامل مختلفی مانند تغییرات فصلی، استراتژی‌های بازاریابی، یا کیفیت تجربه مشتری مرتبط باشد. جدول 2 تحلیل کوهورت را نشان می‌دهد.

جدول ۲ محاسبات Cohort



نرخ دسامبر ۲۰۱۰ نسبت به سایر گروه‌ها نرخ ماندگاری بالاتری دارد و مشتریان بیشتری در طول زمان فعال باقی مانده‌اند. در ماه‌های ششم تا دوازدهم این گروه، هنوز تعداد قابل توجهی از مشتریان به خرید ادامه داده‌اند که نشان‌دهنده موفقیت در حفظ این مشتریان است. در مقابل، کوهورت‌های جدیدتر مانند سپتامبر ۲۰۱۱ به بعد، افت سریع‌تری را نشان می‌دهند، به طوری که تعداد مشتریان فعال پس از چند ماه بسیار کاهش می‌یابد. این موضوع ممکن است نشان‌دهنده کاهش اثربخشی در استراتژی‌های بازاریابی، تغییر در ترکیب مشتریان، یا افزایش رقابت در بازار باشد.

همچنین مشاهده می‌شود که برخی از کوهورت‌ها پس از ماه ششم مجدداً افزایش نسبی در تعداد مشتریان فعال دارند. این پدیده می‌تواند ناشی از کمپین‌های بازاریابی مجدد، تخفیف‌های فصلی، یا تغییرات رفتاری در الگوی خرید مشتریان باشد. اما این افزایش در همه کوهورت‌ها دیده نمی‌شود، که نشان می‌دهد تأثیرگذاری چنین سیاست‌هایی در گروه‌های مختلف مشتریان متفاوت است.

یکی از نکات مهم این تحلیل، کاهش شدید تعداد مشتریان در داده‌های پایانی مانند نوامبر و دسامبر ۲۰۱۱ است. از آنجا که این بخش‌ها هنوز ماه‌های زیادی را طی نکرده‌اند، این کاهش می‌تواند به دلیل کوتاه بودن دوره مشاهده باشد. بنابراین، برای نتیجه‌گیری دقیق‌تر باید داده‌های بیشتری از ماه‌های آینده در دسترس باشد.

به‌طور کلی، این تحلیل نشان می‌دهد که در برخی از قسمت‌ها مشتریان در بازه‌های طولانی‌تری حفظ شده‌اند، درحالی‌که برخی دیگر کاهش سریع‌تری در نرخ ماندگاری دارند. این اطلاعات می‌تواند به شرکت کمک کند تا استراتژی‌های بازاریابی، برنامه‌های وفاداری، و سیاست‌های تبلیغاتی خود را بر اساس رفتار واقعی مشتریان تنظیم کند.

پیاده‌سازی مدل RFM

برای پیاده‌سازی مدل RFM، ابتدا آخرین تاریخ خرید (snapshot date) در مجموعه داده مشخص می‌شود، سپس برای هر مشتری سه مقدار کلیدی محاسبه می‌شود. تازگی (Recency) تعداد روزهایی است که از آخرین خرید مشتری گذشته و هرچه مقدار آن کمتر باشد، نشان می‌دهد که مشتری اخیراً خرید کرده و فعال‌تر است. تکرار خرید (Frequency) تعداد دفعات خرید مشتری در کل دوره را نشان می‌دهد و هرچه مقدار آن بیشتر باشد، مشتری وفادارتر است. مقدار پول خرج‌شده (Monetary) مجموع مبلغی است که مشتری در کل خریدهایش پرداخت کرده و نشان‌دهنده ارزش مالی او برای کسب‌وکار است.

پس از محاسبه این مقادیر، مشتریان به ۶ گروه رتبه‌بندی می‌شوند. مشتریانی که اخیراً خرید کرده‌اند، امتیاز ۵ دریافت می‌کنند، درحالی‌که مشتریانی که مدت زیادی از آخرین خریدشان گذشته است، امتیاز ۱ خواهند داشت. از طرفی، مشتریانی که تعداد دفعات خرید زیادی داشته‌اند یا مقدار زیادی هزینه کرده‌اند، امتیاز ۵ دریافت کرده و مشتریانی که خریدهای کمی داشته‌اند، امتیاز ۱ می‌گیرند. سپس، این سه امتیاز ترکیب شده و یک کد سه‌رقمی RFM برای هر مشتری ایجاد می‌شود که نشان‌دهنده رفتار خرید او است.

با توجه به این امتیازها، مشتریان در گروه‌های مختلفی دسته‌بندی شده‌اند:

مشتریان VIP شامل افرادی هستند که هم به‌تازگی خرید کرده‌اند، هم به‌صورت مکرر خرید دارند و هم مبلغ بالایی هزینه کرده‌اند. این مشتریان ارزشمندترین گروه برای کسب‌وکار هستند و باید به دنبال افزایش حجم و مبلغ فروش به این گروه باشیم و با طرح‌های متنوع هزینه سوئیچینگ را برای آنان بالا ببریم.

مشتریان وفادار افرادی هستند که اخیراً خرید کرده‌اند و یا به‌صورت دوره‌ای خرید انجام می‌دهند، اما ممکن است مقدار هزینه‌ی آن‌ها نسبت به مشتریان VIP کمتر باشد. این مشتریان نیاز به حفظ تعامل دارند و می‌توان با ارائه تخفیف‌های خاص، ارسال پیشنهادهای شخصی‌سازی‌شده و برنامه‌های امتیازی، آن‌ها را تشویق به خرید بیشتر کرد.

مشتریان معمولی دسته‌ای از مشتریان هستند که میزان تازگی، تکرار خرید و ارزش خرید آن‌ها در حد متوسط قرار دارد. این افراد مشتریان فعال محسوب می‌شوند اما هنوز خریدهای پرحجم یا مکرر ندارند. برای تشویق این گروه به افزایش خرید، می‌توان پیشنهادهای محدود به زمان، بسته‌های خرید ویژه و تبلیغات هدفمند ارائه کرد.

مشتریان باارزش اما در حال ریزش گروهی هستند که قبلاً خریدهای زیادی داشته‌اند اما به‌تازگی خرید آن‌ها کم شده است. این مشتریان ممکن است به دلایل مختلف مانند تغییر نیازها، پیشنهادهای رقبا یا کاهش ارتباط با برند، خریدهای خود را متوقف کرده باشند. برای بازگرداندن آن‌ها می‌توان از ایمیل‌های یادآوری، تخفیف‌های ویژه و پیشنهادهای انحصاری برای خرید مجدد استفاده کرد.

مشتریان از دست رفته شامل کسانی هستند که مدت زیادی است خرید نکرده‌اند و تعامل کمی با کسب‌وکار دارند. احتمال بازگشت این مشتریان پایین است، اما می‌توان از تبلیغات هدفمند، پیشنهادهای شگفت‌انگیز و استراتژی‌های بازاریابی خاص برای جذب مجدد آن‌ها استفاده کرد.

مشتریان در معرض ریزش گروهی هستند که هنوز کاملاً غیرفعال نشده‌اند، اما میزان خرید آن‌ها رو به کاهش است. این گروه نیاز به توجه ویژه دارد و می‌توان با پیشنهادهای تشویقی و پیام‌های یادآوری آن‌ها را دوباره به خرید ترغیب کرد.

جدول زیر درصد مشتریان در هر گروه را مشخص می‌کند.

جدول ۳ درصد مشتریان موجود در هر گروه

Customer_Segment	
VIP Customers	21.990741
At Risk Customers	19.629630
Lost Customers	19.166667
Loyal Customers	13.495370
Other	11.967593
Regular Customers	9.907407
High Value but Churning	3.842593

این مدل RFM باعث می‌شود که کسب‌وکار بتواند برای هر گروه از مشتریان استراتژی‌های بازاریابی مخصوص به خود را تدوین کند. به جای ارائه پیشنهادهای یکسان برای همه، این رویکرد به شرکت کمک می‌کند تا مشتریان VIP را حفظ کند، مشتریان معمولی را به خرید بیشتر تشویق کند و مشتریان در معرض ریزش را بازگرداند.

پیاده‌سازی الگوریتم k-means

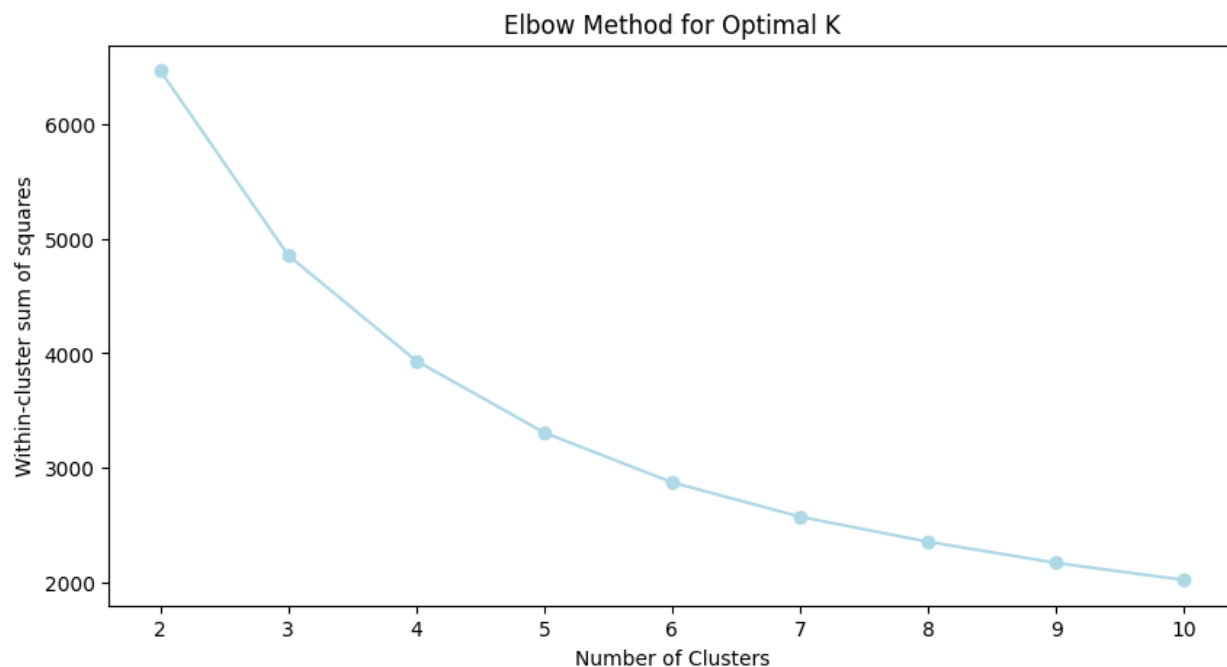
برای انجام تحلیل و خوشه‌بندی مشتریان با الگوریتم K-Means، ابتدا باید داده‌ها را پاک‌سازی، تبدیل و استانداردسازی کنیم. برخی از متغیرهای اصلی مانند تعداد خرید (Quantity) و قیمت واحد (Unit Price) توزیع نامتقارنی دارند که می‌تواند بر عملکرد مدل تأثیر بگذارد. برای حل این مشکل، از تبدیل لگاریتمی (Log Transformation) استفاده شده تا داده‌ها به توزیع نرمال نزدیک‌تر شوند. همچنین، برای یکنواخت‌سازی مقیاس یا اندازه داده‌ها، از استانداردسازی (StandardScaler) استفاده شده است.

مشابه با پیاده‌سازی مدل RFM، ابتدا داده‌های مشتریانی که شناسه ندارند (CustomerID) حذف شده‌اند، زیرا بدون شناسه امکان ردیابی رفتار خرید آن‌ها وجود ندارد. سپس، فاکتورهایی که با "C" شروع می‌شوند حذف شده‌اند، چرا که این کد نشان‌دهنده‌ی لغو تراکنش‌ها است. علاوه بر این، تراکنش‌هایی که شامل مقادیر منفی یا صفر برای تعداد و قیمت واحد هستند، نیز حذف شده‌اند، چرا که این داده‌ها در تحلیل نامعتبر خواهند بود. در نهایت، یک متغیر جدید به نام TotalPrice ایجاد شده که حاصل ضرب تعداد در قیمت واحد است و نشان‌دهنده‌ی کل مبلغ خرید در هر تراکنش است.

پس از حذف داده‌های نامعتبر، با استفاده از نمودار Box Plot توزیع مقادیر تعداد و مبلغ واحد بررسی شده است. این نمودار نشان می‌دهد که هر دو ویژگی دارای مقادیر پرت (Outliers) هستند که می‌توانند بر تحلیل تأثیر بگذارند. برای مقابله با این مسئله، تبدیل لگاریتمی به کار گرفته شده تا توزیع داده‌ها متقارن‌تر شود.

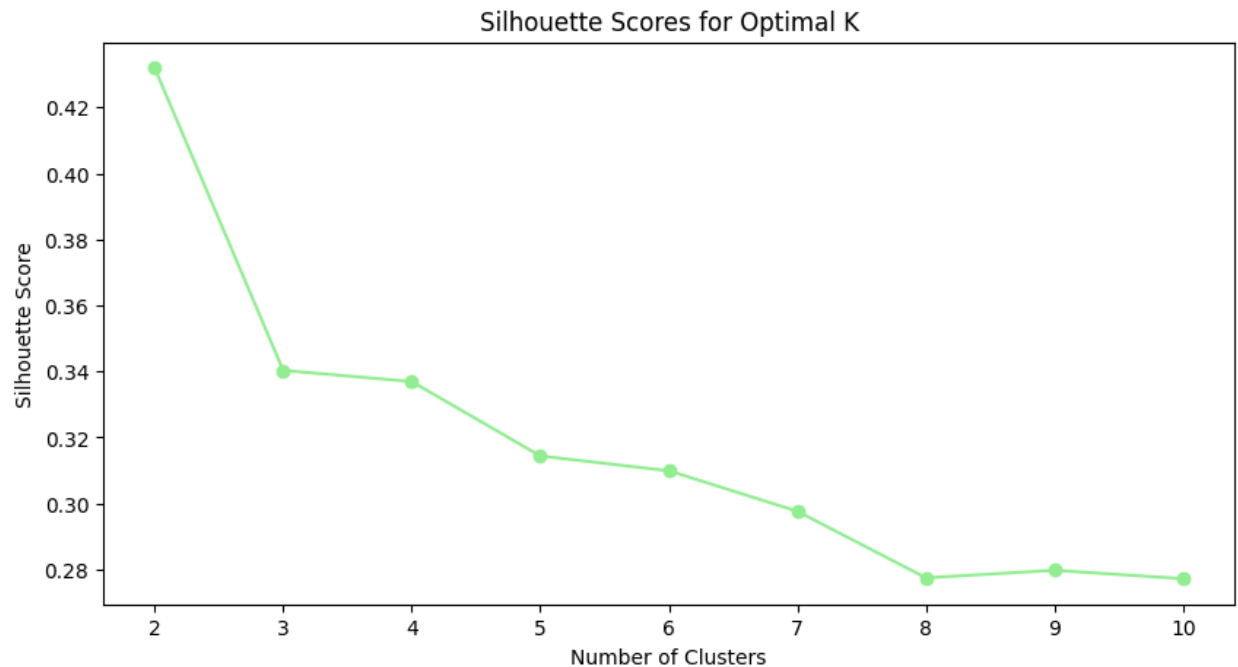
برای تعیین تعداد بهینه خوشه‌ها در الگوریتم K-Means، از دو روش اصلی استفاده شده است :

در روش Elbow، مقدار WCSS یا همان مجموع مربعات فاصله داده‌ها از مرکز خوشه برای تعداد مختلفی از خوشه‌ها محاسبه شده است. سپس نقطه‌ای که در آن کاهش WCSS دیگر به شدت کند می‌شود به عنوان تعداد بهینه خوشه‌ها در نظر گرفته می‌شود. این نقطه نشان می‌دهد که افزودن خوشه‌های بیشتر تأثیر قابل توجهی در کاهش واریانس داخل خوشه‌ها ندارد. نمودار 4 نتیجه‌ی این روش را نشان می‌دهد.



نمودار ۴ روش *Elbow* برای مشخص کردن مقدار k بهینه

در روش *Silhouette Score*، کیفیت خوشه‌بندی بر اساس میزان جداسازی و فشردگی داده‌ها در هر خوشه اندازه‌گیری شده است. امتیاز بالاتر نشان می‌دهد که داده‌های داخل یک خوشه به هم نزدیک‌تر و از خوشه‌های دیگر متمایزتر هستند. نمودار 5 نتیجه‌ی این روش را نشان می‌دهد. بر اساس این تحلیل، مقدار بهینه‌ی 4 انتخاب شده است، زیرا در این حالت هم *WCSS* کاهش مناسبی دارد و هم *Silhouette Score* مقدار قابل قبولی را نشان می‌دهد.

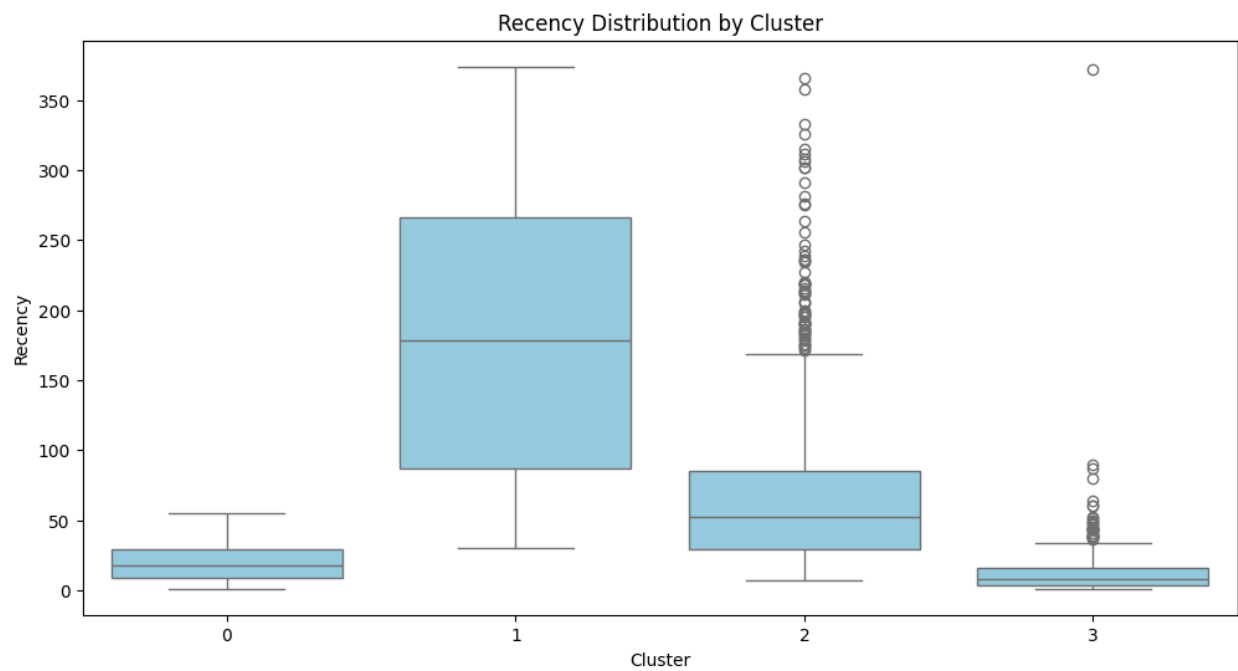


نمودار ۵ استفاده از Silhouette Score برای مشخص کردن مقدار k بهینه

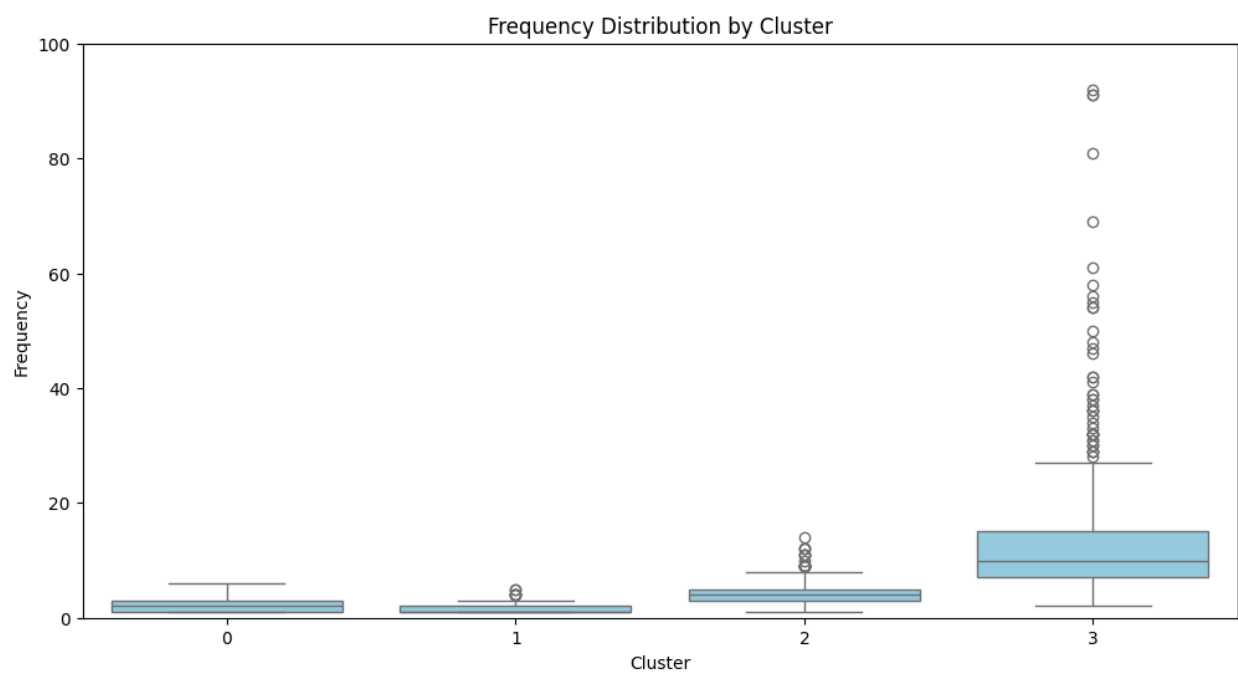
با اجرای الگوریتم **K-Means** و دسته‌بندی مشتریان بر اساس معیارهای **RFM**، چهار خوشه به دست آمده است. نمودارهای جعبه‌ای (Box Plot) نشان می‌دهند که این خوشه‌ها از نظر رفتار خرید تفاوت‌های قابل توجهی دارند. در ادامه، هر خوشه را تحلیل می‌کنیم.

تحلیل خوشه‌های مشتریان

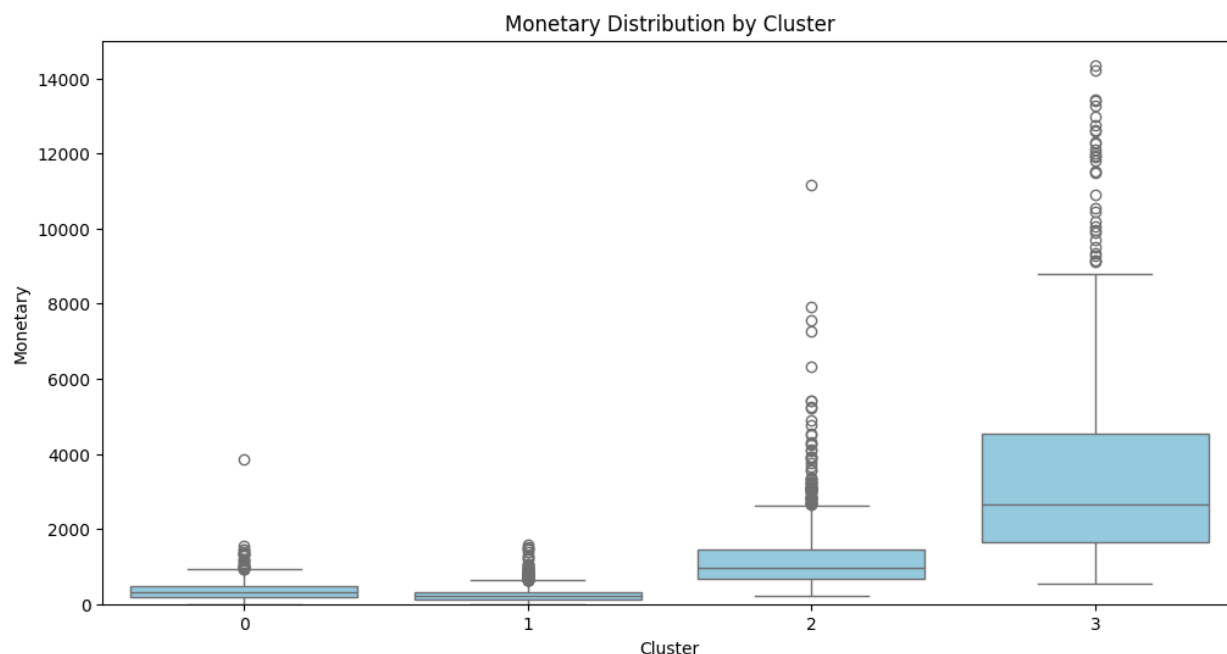
برای تحلیل مشتریان، از مدل **RFM (Recency, Frequency, Monetary)** و الگوریتم **K-Means** استفاده شده است. در این روش، مشتریان بر اساس میزان تازگی خرید، تعداد دفعات خرید و مقدار هزینه‌کرد دسته‌بندی شده‌اند. نتایج این خوشه‌بندی به چهار گروه تقسیم شده است که هر کدام ویژگی‌ها و رفتارهای متفاوتی دارند. نمودارهای زیر ویژگی‌های خوشه‌ها را نشان می‌دهد.



نمودار مقدار *Recency* برای هر خوشه



نمودار مقدار *Frequency* برای هر خوشه



نمودار ۸ مقدار Monetary برای هر خوشه

با توجه به نمودارها، خوشه 0 شامل مشتریانی است که خریدهای بسیار کمی داشته‌اند و مدت زیادی از آخرین خریدشان گذشته است. این مشتریان تعامل محدودی با کسب‌وکار داشته‌اند و احتمال خرید مجدد آن‌ها بسیار کم است. برای بازگرداندن این گروه می‌توان از ایمیل‌های بازاریابی و پیشنهادهای تخفیفی ویژه استفاده کرد. همچنین تحلیل دلایل عدم خرید و بهبود تجربه مشتری می‌تواند به افزایش نرخ بازگشت آن‌ها کمک کند.

خوشه 1 شامل مشتریانی است که خریدهای پراکنده اما کم‌ارزشی دارند. برخی از آن‌ها اخیراً خرید کرده‌اند، اما بیشترشان در فواصل زمانی طولانی خرید می‌کنند. این گروه مشتریان بالقوه‌ای هستند که با ارائه پیشنهادهای شخصی‌سازی‌شده و بسته‌های تخفیفی می‌توان آن‌ها را به خرید بیشتر تشویق کرد. اجرای کمپین‌های تبلیغاتی و ایمیل‌های یادآوری نیز می‌تواند باعث افزایش تعامل آن‌ها با برند شود.

خوشه 2 مشتریانی را شامل می‌شود که تازگی خرید کرده‌اند و ارزش مالی متوسطی دارند. آن‌ها ارتباط تقریباً منظمی با برند دارند، اما هنوز به سطح مشتریان VIP نرسیده‌اند. برای تقویت این مشتریان می‌توان برنامه‌های وفاداری و امتیازدهی به خریدهای بیشتر اجرا کرد. همچنین ارسال پیشنهادهای ویژه برای مشتریان وفادار می‌تواند آن‌ها را به خرید بیشتر ترغیب کند.

خوشه 3 شامل مشتریان VIP و پرارزش است که خریدهای مکرر و ارزش مالی بسیار بالایی دارند. این مشتریان ارتباط قوی با برند دارند و از مشتریان اصلی کسب‌وکار محسوب می‌شوند. برای حفظ این مشتریان، ارائه برنامه‌های ویژه VIP مانند تخفیف‌های اختصاصی

و خدمات مشتریان برتر ضروری است اما باید برای فروش محصولات مرتب و افزایش خریدشان هم برنامه‌هایی در نظر گرفته شود. همچنین، ایجاد تعامل مداوم از طریق پیشنهادهای ویژه و دسترسی‌های انحصاری می‌تواند این مشتریان را به وفاداری بیشتر تشویق کند.

این تحلیل نشان می‌دهد که مشتریان در چهار خوشه‌ی مختلف قرار دارند. هر گروه رفتار خرید خاص خود را دارد و نتایج خوشه‌بندی با الگوریتم k-mean معنی‌دار و قابل اتکا شدند.

نتیجه‌گیری

در این پروژه، تحلیل داده‌های مشتریان یک فروشگاه آنلاین بر اساس مدل RFM و الگوریتم K-Means انجام شد تا الگوهای رفتاری مشتریان شناسایی شده و گروه‌بندی مناسبی برای بهبود تصمیم‌گیری‌های کسب‌وکار ارائه گردد. نتایج این تحلیل نشان داد که مشتریان به چهار دسته اصلی تقسیم می‌شوند که هر کدام رفتار خرید و ارزش متفاوتی برای کسب‌وکار دارند.

استفاده از این نتایج در استراتژی‌های بازاریابی و فروش می‌تواند به کسب‌وکار کمک کند تا روی مشتریان ارزشمند تمرکز کرده، نرخ حفظ مشتریان را افزایش دهد و بهینه‌ترین پیشنهادها را برای هر گروه ارائه کند. همچنین، این تحلیل نشان داد که خوشه‌بندی مشتریان با الگوریتم K-Means، الگوهای قابل اعتمادی را ارائه می‌دهد که می‌توانند در تصمیم‌گیری‌های آتی مورد استفاده قرار گیرند.