تحلیل داده‌محور مشتریان در تجارت الکترونیک

مطالعه موردی برای پشتیبانی از تصمیم‌گیری بازاریابی، نگهداشت و ارزش مشتری

تهیه‌کننده: Sahar Baratian

رشته: مهندسی صنایع

سال: 2026

محیط اجرا: VS Code و Jupyter Notebook

فهرست مطالب

1. مطالعه موردی برای پشتیبانی از تصمیم‌گیری بازاریابی، نگهداشت و ارزش مشتری
2. چکیده
3. 1. مقدمه و بیان مسئله
4. 2. اهداف پژوهش
5. 3. داده‌ها و متغیرهای پژوهش
6. 4. محیط اجرا و کتابخانه‌ها
7. 5. آماده‌سازی و کنترل کیفیت داده
8. 5.1. کنترل ساختار فایل
9. 5.2. تبدیل تاریخ و متغیرهای عددی
10. 5.3. استانداردسازی متغیرهای متنی
11. 5.4. تفسیر کیفیت داده
12. 6. روش‌شناسی کلی
13. 7. سؤال اول: مشتریان اصلی فروشگاه چه کسانی هستند؟
14. 7.1. هدف تحلیل
15. 7.2. منطق کد
16. 7.3. پاسخ تحلیلی و مدیریتی
17. 7.4. نتیجه مدیریتی
18. 8. سؤال دوم: کدام استان و شهر برای سرمایه‌گذاری تبلیغاتی مناسب‌تر است؟
19. 8.1. هدف تحلیل
20. 8.2. تجمیع جغرافیایی
21. 8.3. نتایج
22. 8.4. پاسخ به سؤال مدیریتی
23. 8.5. نتیجه مدیریتی

24. 9. سؤال سوم: کدام مشتریان برای برنامه وفاداری اولویت دارند؟

25. 9.1. هدف تحلیل

26. 9.2. نرمال‌سازی و فرمول

27. 9.3. نتایج

28. 9.4. پاسخ به سؤال مدیریتی

29. 9.5. محدودیت امتیاز

30. 10. سؤال چهارم: کدام مشتریان باارزش در معرض غیرفعال‌شدن هستند؟

31. 10.1. هدف تحلیل

32. 10.2. ساخت آستانه‌ها

33. 10.3. نتایج

34. 10.4. پاسخ دقیق به سؤال مدیریتی

35. 10.5. تفاوت غیرفعال‌بودن و Churn

36. 11. سؤال پنجم: تخفیف، دستگاه و روش پرداخت چه ارتباطی با رفتار مشتری دارند؟

37. 11.1. هدف تحلیل

38. 11.2. روش محاسبه

39. 11.3. نتایج تخفیف

40. 11.4. پاسخ به سؤال مدیریتی

41. 11.5. روش پرداخت

42. 12. سؤال ششم: گزارش پنج‌دقیقه‌ای مدیرعامل چه محتوایی دارد؟

43. 12.1. هدف گزارش مدیریتی

44. 12.2. پاسخ Evidence → Action → KPI

45. 13. تحلیل یکپارچه مشتریان

46. 14. پیشنهادهای مدیریتی

47. 14.1. اولویت‌دادن به مشتریان باارزش در معرض خطر

48. 14.2. تبدیل تخفیف از سیاست عمومی به مداخله قابل آزمایش

49. 14.3. تفکیک شهر مقیاس از شهر کیفیت

50. 14.4. قرار دادن تجربه مشتری در کنار درآمد

51. 15. محدودیت‌های پژوهش

52. 16. مسیر توسعه آینده

53. 17. نتیجه‌گیری نهایی

تحلیل داده‌محور مشتریان در تجارت الکترونیک

مطالعه موردی برای پشتیبانی از تصمیم‌گیری بازاریابی، نگهداشت و ارزش مشتری

تهیه‌کننده: Sahar Baratian

رشته: مهندسی صنایع

سال: 2026

نوع مطالعه: مطالعه موردی تحلیل کسب‌وکار

محیط اجرا: VS Code و Jupyter Notebook

چکیده

در این پروژه، رفتار مشتریان یک مجموعه تجارت الکترونیک را با استفاده از یک چارچوب تحلیل داده‌محور بررسی کردم. هدف من این بود که داده‌های سطح مشتری را از حالت خام و توصیفی خارج کنم و آن‌ها را به شواهدی برای تصمیم‌گیری در زمینه ارزش مشتری، وفاداری، نگهداشت، تجربه مشتری، عملکرد جغرافیایی و محرک‌های بازاریابی تبدیل کنم.

داده مورد استفاده شامل 60 رکورد مشتری و 17 متغیر جمعیت‌شناختی، جغرافیایی، تراکنشی، رفتاری، پرداخت و رضایت است. تحلیل را در محیط VS Code و با استفاده از Jupyter Notebook انجام دادم. کتابخانه‌های اصلی مورد استفاده من Matplotlib، NumPy، Pandas و Pathlib بودند. روند کار از کنترل ساختار فایل و بررسی کیفیت داده شروع شد و سپس به تحلیل ویژگی‌های مشتری، مقایسه شهرها، طراحی امتیاز وفاداری، شناسایی مشتریان باارزش در معرض غیرفعال‌شدن، مقایسه تخفیف و مقایسه روش‌های پرداخت رسید.

نتایج ثبت‌شده نشان داد که مجموع هزینه‌کرد مشتریان 201,985.73، میانگین هزینه‌کرد 3,366.43، میانگین تعداد خرید 17.38، میانگین رضایت 2.98 از 5 و مجموع کالاهای مرجوع‌شده 228 بوده است. این نتایج نشان می‌دهد که پایگاه مشتریان از نظر ارزش اقتصادی و رفتاری ناهمگن است و میانگین به‌تنهایی تصویر مناسبی از همه مشتریان ارائه نمی‌کند.

در تحلیل جغرافیایی، مشهد با مجموع هزینه‌کرد 43,197.58 و میانگین تعداد خرید 22.64، از نظر مقیاس درآمد و تکرار خرید جایگاه بالاتری داشت. در مقابل، اصفهان با میانگین هزینه‌کرد 4,395.06 و میانگین رضایت 4.40، از نظر

ارزش متوسط مشتری و کیفیت تجربه عملکرد بهتری نشان داد. بنابراین به این نتیجه رسیدم که انتخاب شهر برتر به هدف تصمیم بستگی دارد.

برای سنجش وفاداری، یک امتیاز ترکیبی اختصاصی بر اساس هزینه‌کرد، تعداد خرید، تازگی خرید و رضایت طراحی کردم. مشتریان 1057، 1044، 1059، 1009 و 1004 در میان مشتریان دارای امتیاز بالاتر قرار گرفتند. برای شناسایی ریسک نگهداشت نیز سه آستانه بر اساس چارک سوم هزینه‌کرد، میانه تعداد خرید و چارک سوم تعداد روز از آخرین خرید محاسبه کردم. مشتریان 1010، 1035 و 1053 هر سه شرط را داشتند. با توجه به رضایت بسیار پایین مشتریان 1010 و 1053، برای آن‌ها بازیابی تجربه مشتری را پیش از پیشنهاد فروش ضروری دانستم.

در تحلیل تخفیف، مشتریان دارای تخفیف میانگین 17.50 خرید و مشتریان بدون تخفیف میانگین 17.29 خرید داشتند؛ اما میانگین هزینه‌کرد و میانگین ارزش سفارش گروه دارای تخفیف پایین‌تر بود. این نتیجه یک الگوی توصیفی است و رابطه علت و معلولی را ثابت نمی‌کند. همچنین کاربران Card میانگین هزینه‌کرد بالاتری داشتند، اما این نتیجه را به‌عنوان association و شاخص پروفایل مشتری تفسیر کردم، نه اثر علّی.

منطق نهایی این پروژه به شکل زیر خلاصه می‌شود:

> **داده → شواهد → بینش → تصمیم → سنجش اثر**

---

1. مقدمه و بیان مسئله

مشتریان یک فروشگاه اینترنتی از نظر مقدار هزینه‌کرد، تعداد خرید، تازگی تعامل، سطح رضایت، سطح عضویت، روش پرداخت و حساسیت به تخفیف یکسان نیستند. در چنین شرایطی، اجرای یک سیاست بازاریابی یکسان برای همه مشتریان می‌تواند باعث شود منابع سازمان به مشتریانی اختصاص پیدا کند که الزاماً بیشترین ارزش یا بیشترین احتمال بازگشت را ندارند.

برای مثال، مشتری‌ای که در گذشته هزینه‌کرد زیادی داشته اما مدت طولانی از آخرین خرید او گذشته است، از نظر مدیریتی با مشتری‌ای که به‌طور منظم خرید می‌کند اما ارزش هر سفارش او پایین است، وضعیت متفاوتی دارد. مشتری اول ممکن است در معرض از دست رفتن درآمد باشد، درحالی‌که مشتری دوم می‌تواند فرصت مناسبی برای Upsell، Cross-sell یا افزایش ارزش سبد خرید باشد.

در این پروژه، مسئله اصلی را این‌گونه تعریف کردم:

> چگونه می‌توان از داده‌های سطح مشتری برای شناسایی ارزش، وفاداری، ریسک غیرفعال‌شدن و فرصت‌های بازاریابی استفاده کرد و سپس این یافته‌ها را به اقدامات مدیریتی قابل سنجش تبدیل کرد؟

راهنمای اولیه پروژه نیز بر همین موضوع تأکید داشت که هدف فقط تولید نمودار نیست، بلکه باید از یک درخواست مدیریتی شروع کرد، متغیرهای مناسب را انتخاب کرد، تحلیل را اجرا کرد و نتیجه را تفسیر کرد.

2. اهداف پژوهش

هدف اصلی من طراحی یک چارچوب تحلیل مشتری بود که چهار حوزه اصلی تصمیم‌گیری را پوشش دهد.

حوزه	هدفی که در پروژه دنبال کردم
ارزش مشتری	اندازه‌گیری تفاوت میان مشتریان از نظر مشارکت اقتصادی
وفاداری مشتری	ترکیب هزینه‌کرد، تعداد خرید، تازگی و رضایت
نگهداشت مشتری	شناسایی مشتریان باارزشی که نشانه‌های غیرفعال‌شدن دارند
اثربخشی بازاریابی	توصیف تفاوت گروه‌های تخفیف، دستگاه و روش پرداخت بدون ادعای علیت

برای اجرای این هدف، تحلیل را بر اساس شش سؤال مدیریتی تنظیم کردم:

شماره	سؤال مدیریتی	پاسخ تحلیلی پروژه
1	مشتریان اصلی فروشگاه چه ویژگی‌هایی دارند؟	پروفایل سنی، جنسیتی، عضویت، دستگاه و روش پرداخت را محاسبه کردم.
2	کدام استان و شهر برای سرمایه‌گذاری مناسب‌تر است؟	شهرها را بر اساس تعداد مشتری، هزینه‌کرد، خرید، ارزش سفارش و رضایت مقایسه کردم.
3	کدام مشتریان برای برنامه وفاداری اولویت دارند؟	یک امتیاز ترکیبی وفاداری با چهار مؤلفه ساختم.
4	کدام مشتریان باارزش در معرض غیرفعال‌شدن هستند؟	سه شرط ارزش، تکرار خرید و عدم فعالیت را هم‌زمان اعمال کردم.
5	تخفیف، دستگاه و روش پرداخت با رفتار مشتری چه ارتباطی دارند؟	مقایسه توصیفی هزینه‌کرد، خرید، سفارش، مرجوعی و رضایت را انجام دادم.
6	مدیرعامل در پنج دقیقه چه چیزی باید ببیند؟	شش KPI، سه نمودار و سه اقدام Evidence → Action → KPI طراحی کردم.

این ترتیب باعث شد تحلیل از شناخت عمومی مشتری شروع شود و سپس به اولویت‌بندی، تشخیص ریسک و پیشنهاد اقدام برسد.

3. داده‌ها و متغیرهای پژوهش

داده‌ها در سطح مشتری ثبت شده‌اند و شامل 60 رکورد و 17 متغیر هستند.

متغیر	مفهوم	کاربرد در تحلیل
customer_id	شناسه یکتای مشتری	شمارش و شناسایی مشتری
first_name	نام مشتری	گزارش سطح مشتری؛ در انتشار عمومی باید ناشناس‌سازی شود
gender	جنسیت	تحلیل پروفایل
age	سن	گروه‌بندی سنی
city	شهر	تحلیل جغرافیایی
province	استان	تجمیع جغرافیایی
signup_date	تاریخ ثبت‌نام	کنترل نوع داده و توسعه Cohort
membership_tier	سطح عضویت	پروفایل و تفسیر وفاداری
purchase_count	تعداد خرید	مؤلفه Frequency
avg_order_value	میانگین ارزش سفارش	تحلیل اندازه سبد خرید
total_spending	مجموع هزینه‌کرد	مؤلفه Monetary
last_purchase_days	تعداد روز از آخرین خرید	مؤلفه Recency و شاخص غیرفعال‌شدن
payment_method	روش پرداخت	مقایسه توصیفی پروفایل مشتری
device	دستگاه مورد استفاده	مقایسه کانال و پروفایل
discount_used	استفاده از تخفیف	مقایسه گروهی تخفیف
returned_items	تعداد کالاهای مرجوعی	شاخص اصطکاک پس از خرید
satisfaction_score	امتیاز رضایت از 1 تا 5	شاخص تجربه مشتری

این متغیرها امکان بررسی چهار بُعد Recency، Monetary، Frequency و Satisfaction را فراهم کردند. بااین‌حال، چون داده‌ها در سطح مشتری تجمیع شده‌اند، تحلیل محصول، سفارش، کمپین، Cohort، سود و مسیر دقیق خرید در این نسخه امکان‌پذیر نبود.

4. محیط اجرا و کتابخانه‌ها

پروژه را در محیط VS Code و با استفاده از Jupyter Notebook انجام دادم. این محیط به من اجازه داد کد، توضیح، جدول، نمودار و خروجی را در یک جریان تحلیلی منسجم قرار دهم.

ابزار یا کتابخانه	کاربردی که در پروژه داشت
VS Code	مدیریت فایل‌ها، ویرایش Notebook و اجرای تحلیل
Jupyter Notebook	ترکیب کد اجرایی با توضیح و خروجی
Pandas	خواندن Excel، پاکسازی، گروه‌بندی، تجمیع و ذخیره خروجی
NumPy	محاسبات عددی و تعریف کران‌های گروه‌های سنی
Matplotlib	رسم نمودارهای میله‌ای، افقی و پراکندگی
Pathlib	مدیریت مسیر فایل و بررسی وجود فایل ورودی

5. آماده‌سازی و کنترل کیفیت داده

قبل از انجام تحلیل اصلی، داده‌ها را از نظر ساختار، نوع، کامل‌بودن و تکراری‌بودن بررسی کردم. این مرحله برای من یک مرحله مستقل بود؛ زیرا اگر داده‌ها نادرست یا ناسازگار باشند، خروجی‌های آماری و نمودارها نیز قابل اعتماد نخواهند بود.

5.1. کنترل ساختار فایل

ابتدا فهرستی از ستون‌های مورد نیاز تعریف کردم و آن را با ستون‌های موجود در فایل مقایسه کردم. اگر ستونی وجود نداشت، اجرای تحلیل متوقف می‌شد. این تصمیم از آن جهت مهم است که اجرای تحلیل روی یک فایل ناقص ممکن است بدون خطای واضح، نتایج ناقص تولید کند.

5.2. تبدیل تاریخ و متغیرهای عددی

ستون `signup_date` را با `pd.to_datetime` تبدیل کردم. برای متغیرهای عددی مانند سن، تعداد خرید، میانگین ارزش سفارش، مجموع هزینه‌کرد، روزهای گذشته از آخرین خرید، تعداد مرجوعی و رضایت نیز از تبدیل عددی استفاده کردم. مقادیر نامعتبر به missing تبدیل شدند تا تعداد آن‌ها قابل بررسی باشد.

5.3. استانداردسازی متغیرهای متنی

برای متغیرهای متنی، نوع داده را به string تبدیل و فاصله‌های اضافی ابتدا و انتهای متن را حذف کردم. این کار برای جلوگیری از ایجاد گروه‌های ظاهراً متفاوت اما در واقع یکسان ضروری بود؛ برای مثال، یک نام شهر با فاصله اضافی نباید به‌عنوان شهر جداگانه محاسبه شود.

5.4. تفسیر کیفیت داده

کنترل کیفیت داده صرفاً یک مرحله فنی نبود. این کنترل مستقیماً بر اعتبار تصمیم اثر می‌گذارد. اگر تعداد مشتریان یک شهر به دلیل رکورد تکراری بیشتر از واقعیت باشد، آن شهر در رتبه‌بندی جغرافیایی بیش‌ازحد مهم دیده می‌شود. اگر مقادیر عددی به‌صورت متن ذخیره شده باشند، محاسبه میانگین یا مجموع می‌تواند خطا داشته باشد.

به همین دلیل، در گزارش نتایج، اندازه نمونه، محدودیت داده‌ها و تفاوت میان «داده مشاهده‌شده» و «نتیجه قابل تعمیم» را از یکدیگر جدا کردم.

6. روش‌شناسی کلی

روش تحلیل من در چهار سطح انجام شد.

سطح	پرسش	ابزار تحلیلی
توصیفی	در داده چه چیزی مشاهده می‌شود؟	تعداد، مجموع، میانگین، میانه و توزیع
مقایسه‌ای	گروه‌ها چه تفاوتی دارند؟	مقایسه شهر، تخفیف، دستگاه و روش پرداخت
امتیازدهی مشتری	کدام مشتری ارزش یا ریسک بیشتری دارد؟	Loyalty Score و At-Risk Flag
تصمیم‌محور	بر اساس یافته‌ها چه اقدامی پیشنهاد می‌شود؟	Segmentation و KPI → Action → Evidence

این ساختار به من کمک کرد تحلیل را از یک مجموعه کد پراکنده به یک مطالعه موردی منسجم تبدیل کنم.

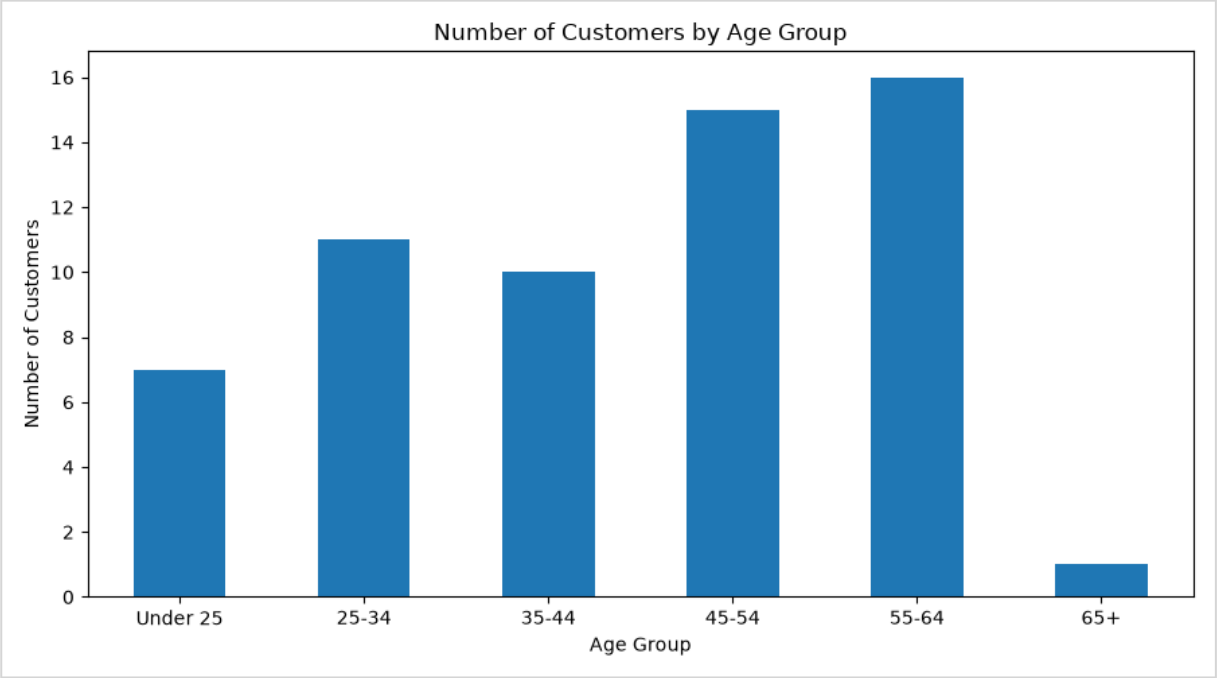
7. سؤال اول: مشتریان اصلی فروشگاه چه کسانی هستند؟

7.1. هدف تحلیل

قبل از طراحی کمپین، لازم بود بدانم پایگاه مشتریان فعلی چه ویژگی‌هایی دارد. برای این کار سن، جنسیت، سطح عضویت، دستگاه و روش پرداخت را بررسی کردم.

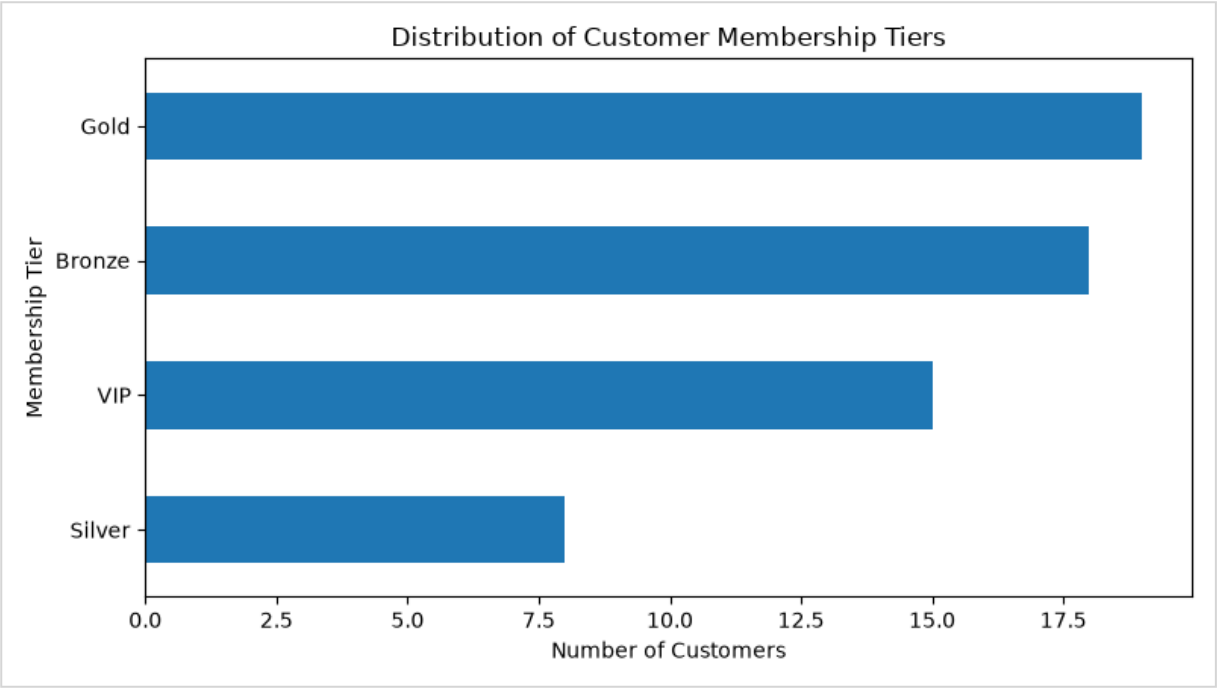
7.2. منطق کد

برای سن، گروه‌های سنی ایجاد کردم تا متغیر پیوسته به دسته‌های قابل تفسیر مدیریتی تبدیل شود. سپس تعداد مشتریان یکتا، میانگین و میانه سن و پرتکرارترین دسته‌های جنسیت، عضویت، دستگاه و روش پرداخت را محاسبه کردم.



شکل 1. تعداد مشتریان بر اساس گروه سنی.

شکل 1. این نمودار میله‌ای توزیع مشتریان را در شش گروه سنی نشان می‌دهد. گروه سنی 55 تا 64 سال بزرگ‌ترین گروه و گروه 65 سال به بالا کوچک‌ترین گروه مشاهده‌شده است. این نمودار به بخش جمعیت‌شناختی سؤال اول پاسخ می‌دهد و نشان می‌دهد که تعریف مشتری غالب باید بر اساس توزیع واقعی سن انجام شود.



شکل 2. توزیع سطح عضویت مشتریان.

شکل 2. این نمودار افقی تعداد مشتریان در سطوح Gold، Bronze، VIP و Silver را مقایسه می‌کند. در داده مشاهده‌شده، Gold و Bronze بزرگ‌ترین گروه‌ها هستند و Silver کوچک‌ترین گروه است. این نمودار ترکیب عضویت را نشان می‌دهد، اما به‌تنهایی ثابت نمی‌کند که بزرگ‌ترین گروه بیشترین ارزش اقتصادی را ایجاد می‌کند.

7.3. پاسخ تحلیلی و مدیریتی

خروجی این بخش، جدول خلاصه ویژگی‌های غالب مشتریان و نمودارهای توزیع گروه‌هاست. این تحلیل به من اجازه داد مشتری غالب را بر اساس داده تعریف کنم، نه بر اساس حدس.

بااین‌حال، پرتعدادترین گروه الزاماً باارزش‌ترین گروه نیست. بنابراین از این تحلیل برای شناخت اولیه و تصمیم درباره نوع پیام و کانال استفاده کردم، اما اولویت اقتصادی را به تحلیل‌های ارزش و نگهداشت سپردم.

7.4. نتیجه مدیریتی

اگر گروه غالب در یک دستگاه یا روش پرداخت مشخص باشد، می‌توان تجربه ارتباطی و طراحی پیام را با آن سازگار کرد. اما برای تصمیم نهایی باید بررسی شود که گروه‌های کم‌تعدادتر چه سهمی از هزینه‌کرد و رضایت دارند. به همین دلیل، پروفایل جمعیتی را در کنار بخش‌بندی ارزش تفسیر کردم.

8. سؤال دوم: کدام استان و شهر برای سرمایه‌گذاری تبلیغاتی مناسب‌تر است؟

8.1. هدف تحلیل

در این بخش، هدفم این بود که شهرها را فقط بر اساس مجموع فروش رتبه‌بندی نکنم. تعداد مشتری، میانگین هزینه‌کرد، تعداد خرید، میانگین ارزش سفارش و رضایت را نیز در تصمیم وارد کردم.

8.2. تجمیع جغرافیایی

داده‌ها را بر اساس ترکیب استان و شهر گروه‌بندی کردم تا شهرهای همنام اشتباه با هم ادغام نشوند. برای هر شهر، شش شاخص محاسبه کردم. سپس شهرهایی با کمتر از سه مشتری را از رتبه‌بندی اصلی کنار گذاشتم؛ زیرا یک یا دو مشاهده نمی‌تواند مبنای مطمئنی برای تصمیم سرمایه‌گذاری باشد.

برای تحلیل هم‌زمان مقیاس مالی و تجربه مشتری، نمودار پراکندگی نیز طراحی کردم.

شکل 3. درآمد شهرها در برابر رضایت مشتری.

شکل 3. این نمودار حبابی، مجموع درآمد شهر را در محور افقی و میانگین رضایت را در محور عمودی نشان می‌دهد و اندازه حباب بر اساس تعداد مشتریان است. مشهد به دلیل مجموع درآمد بالا در سمت راست نمودار قرار دارد، درحالی‌که اصفهان به دلیل رضایت بالاتر در قسمت بالایی دیده می‌شود. این شکل تفاوت میان مقیاس درآمد و کیفیت تجربه مشتری را به‌صورت بصری نشان می‌دهد.

8.3. نتایج

شهر	مشتری	مجموع هزینه‌کرد	میانگین هزینه‌کرد	میانگین خرید	رضایت
Karaj	7	26,189.81	3,741.40	16.00	2.86
Tabriz	12	36,192.49	3,016.04	15.75	3.00
Shiraz	6	20,715.92	3,452.65	16.33	2.00
Rasht	5	15,168.54	3,033.71	16.40	2.60
Isfahan	5	21,975.32	4,395.06	19.80	4.40
Mashhad	11	43,197.58	3,927.05	22.64	2.91
Ahvaz	8	23,384.78	2,923.10	13.88	3.50
Tehran	6	15,161.29	2,526.88	17.17	2.67

8.4. پاسخ به سؤال مدیریتی

مشهد با 43,197.58 بالاترین مجموع هزینه‌کرد مشاهده‌شده را دارد و با میانگین 22.64 خرید، از نظر تکرار خرید نیز در جایگاه اول قرار دارد. بنابراین اگر هدف سازمان توسعه مقیاس درآمد و افزایش حجم بازار باشد، مشهد گزینه مهم‌تری است.

اصفهان با میانگین هزینه‌کرد 4,395.06 و رضایت 4.40، از نظر کیفیت متوسط مشتری و تجربه مشتری بهتر عمل کرده است. در نتیجه، اگر هدف سازمان تمرکز بر مشتریان باارزش و راضی باشد، اصفهان سیگنال مناسب‌تری دارد.

هدف تصمیم	گزینه قوی‌تر در داده	دلیل
افزایش مقیاس درآمد	مشهد	بالاترین مجموع هزینه‌کرد
افزایش تکرار خرید	مشهد	بالاترین میانگین تعداد خرید
تمرکز بر ارزش متوسط	اصفهان	بالاترین میانگین هزینه‌کرد
تمرکز بر تجربه مشتری	اصفهان	بالاترین میانگین رضایت

8.5. نتیجه مدیریتی

به نظر من، تصمیم نهایی نباید فقط بر اساس مجموع فروش باشد. مشهد را برای استراتژی Scale و اصفهان را برای استراتژی Quality و Customer Experience در نظر می‌گیرم. برای تصمیم قطعی، داده‌های هزینه جذب مشتری و حاشیه سود نیز لازم است؛ زیرا درآمد بالا لزوماً به معنی سودآوری بالا نیست.

9. سؤال سوم: کدام مشتریان برای برنامه وفاداری اولویت دارند؟

9.1. هدف تحلیل

وفاداری را یک مفهوم چندبعدی در نظر گرفتم. اگر فقط Total Spending را ملاک قرار می‌دادم، مشتریانی که مدت زیادی خرید نکرده‌اند نیز ممکن بود بدون توجه به Recency در گروه وفادار قرار بگیرند. به همین دلیل، چهار مؤلفه را ترکیب کردم: Recency، Frequency، Monetary و Satisfaction.

9.2. نرمال‌سازی و فرمول

برای هزینه‌کرد، تعداد خرید و رضایت از نرمال‌سازی Min–Max استفاده کردم:

```
(x - minimum) / (maximum - minimum)
```

برای جلوگیری از تقسیم بر صفر، اگر حداقل و حداکثر یک متغیر برابر بودند، امتیاز آن متغیر را 1.0 قرار دادم. برای Recency، چون روز کمتر بهتر است، امتیاز را معکوس کردم:

```
recency_score = 1 - normalized(last_purchase_days)
```

امتیاز نهایی به شکل زیر محاسبه شد:

```
loyalty_score =
0.35 × spending_score
+ 0.30 × purchase_count_score
+ 0.20 × recency_score
+ 0.15 × satisfaction_score
```

وزن 0.35 به ارزش مالی، 0.30 به تکرار خرید، 0.20 به تازگی و 0.15 به رضایت اختصاص داده شد. این امتیاز با منطق RFM هم‌راستا است، اما RFM استاندارد نیست و باید به‌عنوان امتیاز اختصاصی همین پروژه معرفی شود.

9.3. نتایج

مشتری	خرید	هزینه‌کرد	روز از آخرین خرید	رضایت	امتیاز وفاداری
1057	31	13,532.74	82	4	0.8645
1044	33	14,354.34	165	2	0.7809
1059	31	8,898.55	224	5	0.7106
1009	34	11,615.42	232	2	0.6856
1004	23	6,121.45	40	4	0.6385

9.4. پاسخ به سؤال مدیریتی

مشتری 1057 نمونه‌ای از مشتری متعادل و باارزش است؛ زیرا هم تعداد خرید و هزینه‌کرد بالایی دارد، هم فاصله از آخرین خرید او کمتر است و رضایت مناسبی دارد. مشتری 1044 نشان می‌دهد که امتیاز وفاداری بالا می‌تواند با رضایت پایین همراه باشد. بنابراین این مشتری را فقط برای پاداش در نظر نمی‌گیرم، بلکه تجربه او را نیز بررسی می‌کنم.

مشتری 1059 رضایت بسیار خوبی دارد اما فاصله بیشتری از آخرین خرید او گذشته است. این مشتری می‌تواند برای پایش زودهنگام و جلوگیری از غیرفعال‌شدن هدف‌گذاری شود. بنابراین امتیاز وفاداری علاوه بر رتبه‌بندی، برای تشخیص تفاوت نوع مداخله نیز مفید است.

9.5. محدودیت امتیاز

وزن‌های انتخاب‌شده فرض‌های تحلیلی هستند. اگر سازمان Recency را مهم‌تر بداند، باید وزن آن افزایش یابد. اگر هدف تمرکز بر ارزش مالی باشد، وزن Monetary می‌تواند بیشتر شود. بنابراین در نسخه بعدی، تحلیل حساسیت وزن‌ها را اضافه خواهم کرد.

10. سؤال چهارم: کدام مشتریان باارزش در معرض غیرفعال‌شدن هستند؟

10.1. هدف تحلیل

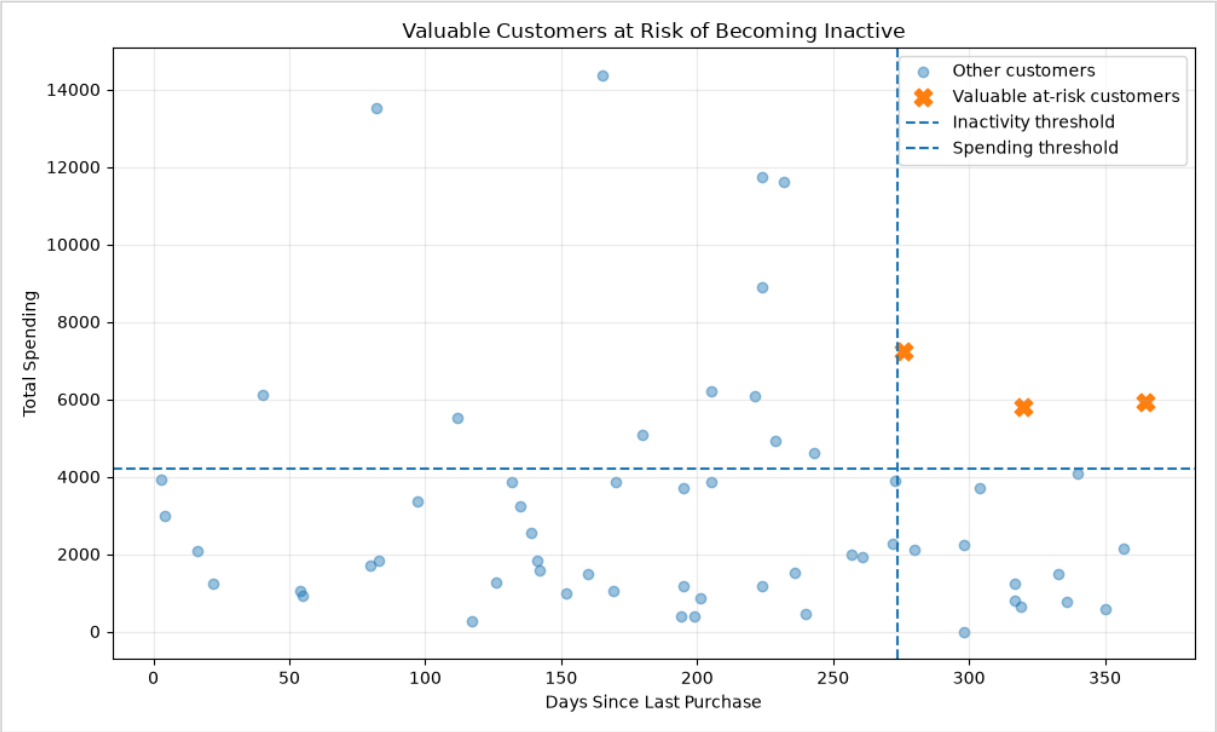
هدف این بخش شناسایی مشتریانی بود که قبلاً برای سازمان ارزش اقتصادی ایجاد کرده‌اند، تعداد خرید مناسبی دارند، اما مدت زیادی از آخرین خریدشان گذشته است.

10.2. ساخت آستانه‌ها

سه آستانه را از خود داده‌ها به دست آوردم:

شاخص	قاعده	مقدار
هزینه‌کرد	چارک سوم	4,213.12
تعداد خرید	میانه	17
تعداد روز از آخرین خرید	چارک سوم	273.75

مشتری فقط زمانی در گروه هدف قرار گرفت که هر سه شرط را هم‌زمان داشته باشد.



شکل 4. مشتریان باارزش در معرض غیرفعال‌شدن.

شکل 4. این نمودار پراکندگی تعداد روز از آخرین خرید را در برابر مجموع هزینه‌کرد نشان می‌دهد. خطوط نقطه‌چین آستانه عدم فعالیت و آستانه هزینه‌کرد هستند و علامت‌های X نارنجی سه مشتری را نشان می‌دهند که هر سه شرط ریسک را دارند. ناحیه بالای سمت راست در این شکل نشان‌دهنده درآمد در معرض خطر است، نه همه مشتریان غیرفعال.

10.3. نتایج

مشتری	تعداد خرید	هزینه‌کرد	روز غیرفعال‌بودن	رضایت
1010	19	7,241.47	276	1
1035	17	5,918.21	365	3
1053	18	5,805.54	320	1

10.4. پاسخ دقیق به سؤال مدیریتی

مشتری 1010 هم از نظر هزینه‌کرد و هم از نظر تعداد خرید باارزش است، اما 276 روز خرید نکرده و رضایت او 1 است. بنابراین برای او ابتدا Service Recovery و Root Cause Investigation را در نظر می‌گیرم و بعد به سراغ Reactivation می‌روم.

مشتری 1035 با 365 روز، طولانی‌ترین فاصله از آخرین خرید را دارد. رضایت او 3 است و نشانه اصلی، عدم فعالیت طولانی است. برای او Personalized Reactivation مناسب‌تر است.

مشتری 1053 نیز ارزش اقتصادی و تعداد خرید مناسبی دارد، اما رضایت او 1 و مدت غیرفعال‌بودن او 320 روز است. این مشتری از مهم‌ترین کاندیداهای بازیابی تجربه و نگهداشت است.

10.5. تفاوت غیرفعال‌بودن و Churn

من در این پروژه از واژه «در معرض غیرفعال‌شدن» استفاده کردم و از عبارت «ریزش‌کرده» استفاده نکردم. دلیل آن این است که داده‌ها outcome آینده مشتری را نشان نمی‌دهند. برای ساخت مدل واقعی Churn باید مشخص شود مشتری در یک بازه آینده، مانند 90 یا 180 روز، دوباره خرید کرده است یا خیر.

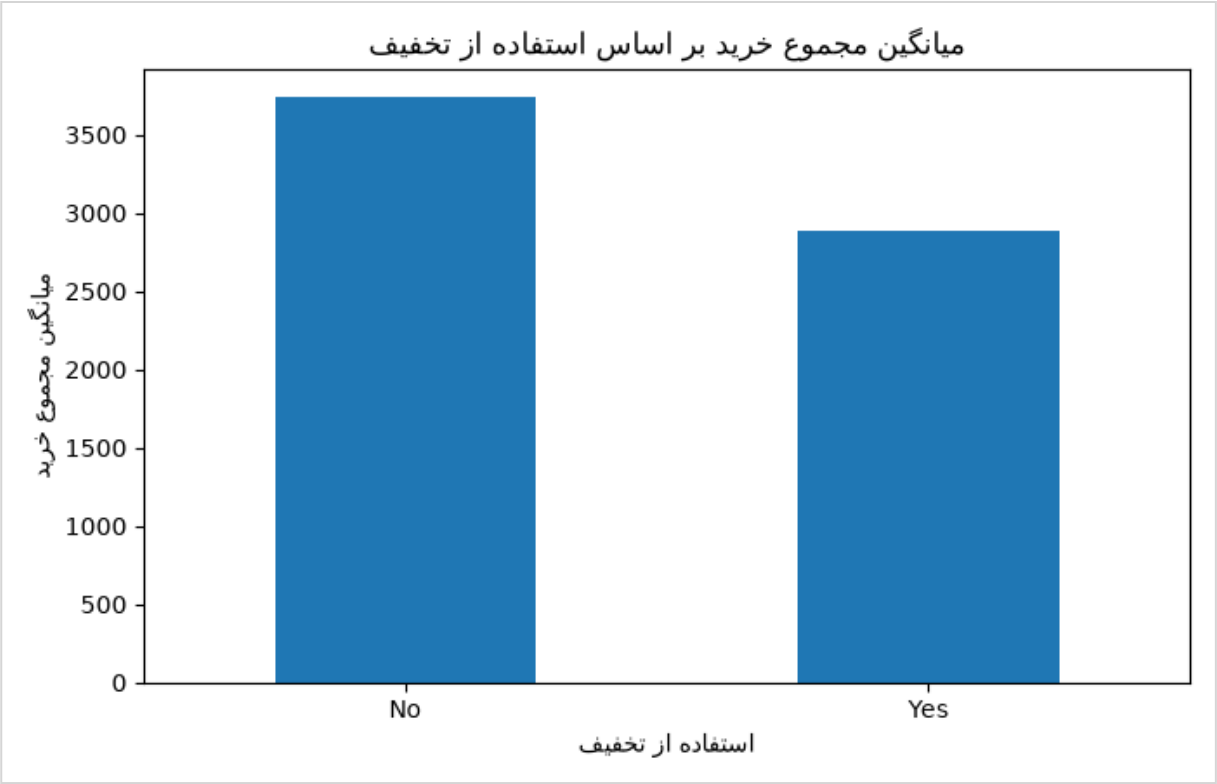
11. سؤال پنجم: تخفیف، دستگاه و روش پرداخت چه ارتباطی با رفتار مشتری دارند؟

11.1. هدف تحلیل

در این بخش، قصد داشتم ببینم گروه‌های دارای تخفیف، دستگاه و روش پرداخت چه تفاوت‌های مشاهده‌شده‌ای در ارزش، تعداد خرید، ارزش سفارش، مرجوعی و رضایت دارند. از ابتدا این فرض روش‌شناختی را رعایت کردم که تفاوت گروهی به‌تنهایی رابطه علّی را ثابت نمی‌کند.

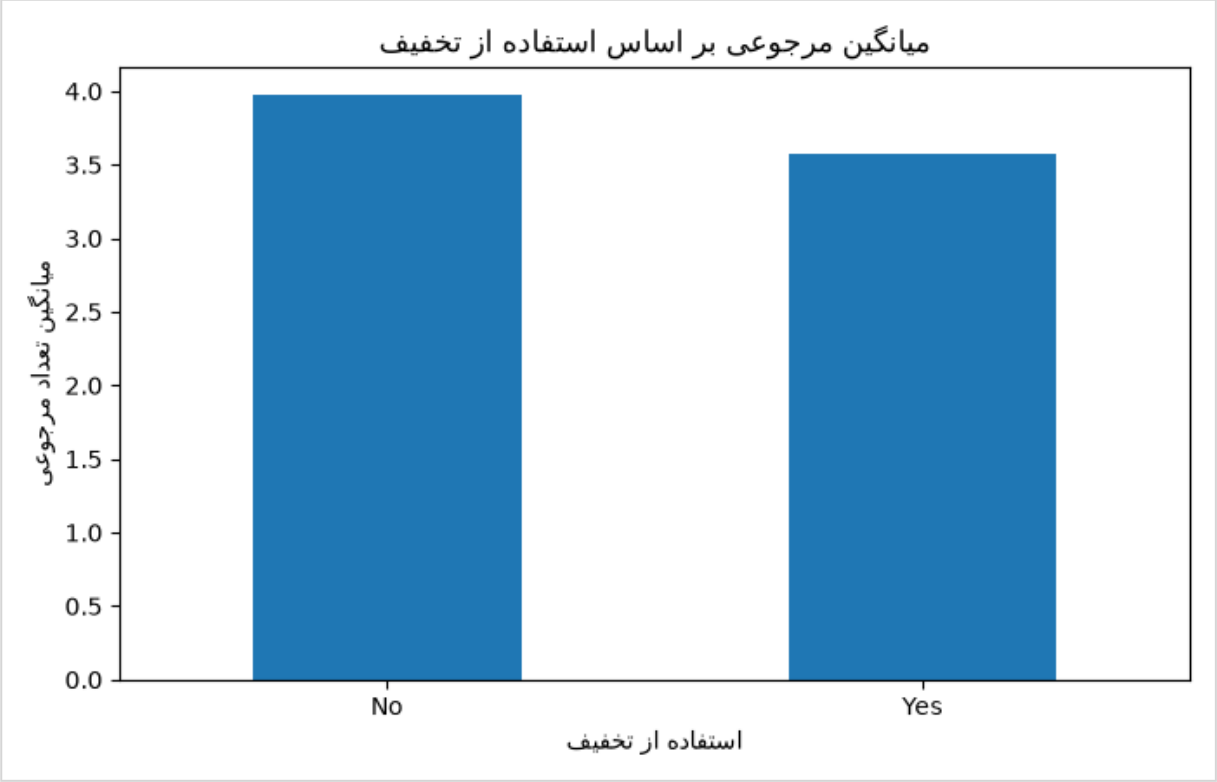
11.2. روش محاسبه

داده‌های دارای missing در متغیرهای مورد نیاز را کنار گذاشتم و بر اساس `discount_used` گروه‌بندی کردم. برای هر گروه، تعداد مشتری، میانگین هزینه‌کرد، تعداد خرید، ارزش سفارش، مرجوعی و رضایت را محاسبه کردم.



شکل 5. میانگین مجموع خرید بر اساس استفاده از تخفیف.

شکل 5. این نمودار میانگین مجموع هزینه‌کرد مشتریان بدون تخفیف و دارای تخفیف را مقایسه می‌کند. میانگین گروه بدون تخفیف بالاتر است. این تفاوت توصیفی است و ممکن است به تفاوت ترکیب مشتریان یا نحوه هدف‌گذاری تخفیف مربوط باشد، نه اثر علّی تخفیف.



شکل 6. میانگین کالاهای مرجوعی بر اساس استفاده از تخفیف.

شکل 6. این نمودار میانگین مرجوعی دو گروه را مقایسه می‌کند. در داده مشاهده‌شده، میانگین مرجوعی گروه دارای تخفیف پایین‌تر است. این نتیجه به‌تنهایی ثابت نمی‌کند که تخفیف باعث کاهش مرجوعی شده است و باید همراه با حجم نمونه و سایر شاخص‌ها تفسیر شود.

11.3. نتایج تخفیف

شاخص	بدون تخفیف	با تخفیف
تعداد مشتری	34	26
میانگین تعداد خرید	17.29	17.50
میانگین هزینه‌کرد	3,736.74	2,882.17
میانگین ارزش سفارش	220.70	203.29
میانگین مرجوعی	3.97	3.58
میانگین رضایت	2.76	3.27

11.4. پاسخ به سؤال مدیریتی

تعداد خرید مشتریان دارای تخفیف فقط 0.21 بیشتر است. این اختلاف برای پشتیبانی از ادعای «تخفیف باعث افزایش خرید شده است» کافی نیست. در مقابل، میانگین هزینه‌کرد و ارزش سفارش در گروه دارای تخفیف پایین‌تر مشاهده شده است.

بااین‌حال، نتیجه‌گیری نمی‌کنم که تخفیف باعث کاهش هزینه‌کرد شده است. ممکن است مشتریان کم‌ارزش‌تر یا حساس‌تر به قیمت، بیشتر تخفیف دریافت کرده باشند. در نتیجه، Selection Bias و Confounding می‌توانند اختلاف را توضیح دهند.

پیشنهاد تحلیلی من این است که تخفیف به‌عنوان Intervention بررسی شود. برای این کار، مشتریان مشابه باید به گروه Treatment و Control تقسیم شوند و Conversion، Incremental Profit، Incremental Revenue، Retention، AOV و Returns سنجیده شود.

11.5. روش پرداخت

روش پرداخت	تعداد مشتری	میانگین هزینه‌کرد
Card	18	3,929.01
Cash	19	3,150.58
Online Wallet	23	3,104.46

کاربران Card بیشترین میانگین هزینه‌کرد را دارند. این نتیجه برای شناخت پروفایل مشتری مفید است، اما نشان نمی‌دهد که Card باعث افزایش هزینه‌کرد شده است. ممکن است مشتریان باارزش‌تر به‌طور طبیعی بیشتر از این روش استفاده کنند.

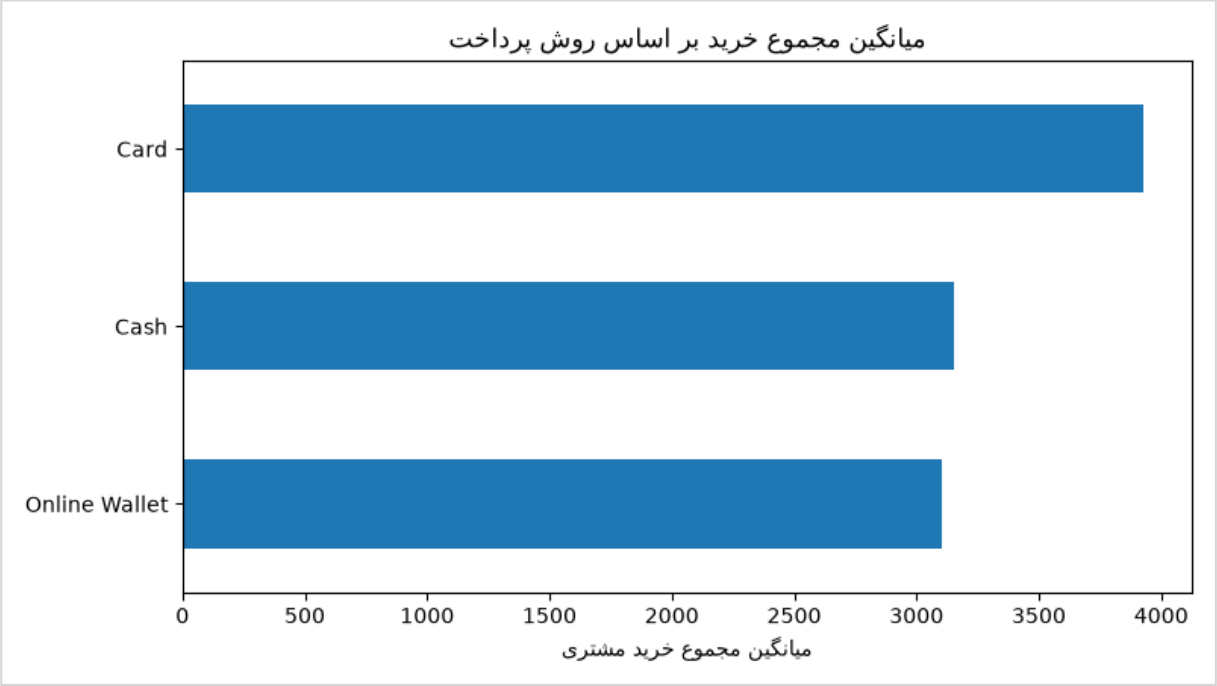
12. سؤال ششم: گزارش پنج‌دقیقه‌ای مدیرعامل چه محتوایی دارد؟

12.1. هدف گزارش مدیریتی

برای مدیرعامل، گزارشی طراحی کردم که در زمان کوتاه قابل مشاهده باشد. به جای قرار دادن تمام نمودارها، شش KPI و سه نمودار را انتخاب کردم.

شش KPI انتخاب‌شده عبارت‌اند از تعداد مشتریان یکتا، مجموع هزینه‌کرد، میانگین هزینه‌کرد، میانگین تعداد خرید، میانگین رضایت و مجموع کالاهای مرجوعی.

سه نمودار منتخب نیز عبارت‌اند از رتبه‌بندی شهرها، نمودار مشتریان باارزش در معرض غیرفعال‌شدن و مقایسه هزینه‌کرد و مرجوعی بر اساس تخفیف.



شکل 7. میانگین مجموع خرید بر اساس روش پرداخت.

شکل 7. این نمودار افقی میانگین هزینه‌کرد را بر اساس روش پرداخت مقایسه می‌کند. کاربران Card بالاترین میانگین هزینه‌کرد را دارند و پس از آن Cash و Online Wallet قرار می‌گیرند. این شکل برای شناخت پروفایل مشتری مفید است، اما نشان نمی‌دهد که روش پرداخت باعث افزایش هزینه‌کرد شده است.

12.2. پاسخ Evidence → Action → KPI

شواهد	اقدام	شاخص سنجش
سه مشتری باارزش، غیرفعال و دارای ارزش تاریخی هستند	برنامه نگهداشت و بازیابی تجربه	Reactivation Rate، Retained Revenue، Satisfaction Recovery
تخفیف تعداد خرید را تقریباً تغییر نداده اما Spending و AOV پایین‌تر است	اجرای آزمایش کنترل‌شده تخفیف	Incremental Profit، Incremental Revenue، AOV
مشهد رهبر Scale و اصفهان رهبر Quality است	تفکیک استراتژی توسعه بازار از کیفیت مشتری	Revenue، CAC، Satisfaction، Repeat Rate، Margin

13. تحلیل یکپارچه مشتریان

با ترکیب یافته‌های بخش‌های مختلف، گروه‌های تصمیم‌گیری زیر به دست آمدند.

وضعیت مشتری	تفسیر	اقدام
ارزش بالا + غیرفعال‌بودن بالا	درآمد در معرض خطر	اولویت نگهداشت
ارزش بالا + رضایت پایین	ریسک تجربه مشتری	بازیابی خدمت و بررسی علت
ارزش بالا + رضایت مناسب + غیرفعال‌بودن	فرصت بازفعال‌سازی	کمپین شخصی‌سازی‌شده
تعداد خرید بالا + ارزش سفارش پایین	فرصت رشد سبد	Upsell، Cross-sell و Bundling
استفاده از تخفیف + ارزش سفارش پایین	احتمال حساسیت به قیمت	آزمایش Treatment و Control
درآمد شهری بالا + رضایت پایین	تنش مقیاس و تجربه	بهبود تجربه مشتری
شهر با ارزش متوسط بالا	پایگاه مشتری باکیفیت	بررسی بهترین تجربه‌ها

این جدول نشان می‌دهد که ارزش مشتری به‌تنهایی برای تعیین اقدام کافی نیست. رضایت و تازگی تعامل مشخص می‌کنند که آیا اقدام باید پاداش، بازیابی خدمت، بازفعال‌سازی یا آزمایش بازاریابی باشد.

14. پیشنهادهای مدیریتی

14.1. اولویت‌دادن به مشتریان باارزش در معرض خطر

مشتریان 1010، 1035 و 1053 باید در اولویت بررسی قرار گیرند، اما پیام یکسانی نباید برای آن‌ها ارسال شود. مشتریان 1010 و 1053 به مسیر بازیابی تجربه نیاز دارند و مشتری 1035 به مسیر بازفعال‌سازی.

14.2. تبدیل تخفیف از سیاست عمومی به مداخله قابل آزمایش

بر اساس داده‌های مشاهده‌شده، تخفیف عمومی برای همه مشتریان توجیه کافی ندارد. تخفیف باید برای گروه‌های هدف مختلف آزمایش شود و معیار اصلی موفقیت آن Incremental Profit باشد، نه فقط تعداد خرید.

14.3. تفکیک شهر مقیاس از شهر کیفیت

مشهد برای توسعه درآمد و تکرار خرید مناسب‌تر به نظر می‌رسد. اصفهان برای کیفیت متوسط مشتری و رضایت مناسب‌تر است. تصمیم نهایی باید با CAC، سود و ظرفیت عملیاتی تکمیل شود.

14.4. قرار دادن تجربه مشتری در کنار درآمد

مشتری باارزش و ناراضی یک فرصت و یک ریسک هم‌زمان است. اگر علت نارضایتی حل نشود، تخفیف ممکن است فقط مشکل اصلی را پنهان کند. بنابراین در تحلیل نگهداشت، Satisfaction را به‌عنوان ابزار تشخیص علت وارد کردم.

15. محدودیت‌های پژوهش

نخستین محدودیت، حجم نمونه 60 مشتری است. این حجم برای تحلیل اکتشافی مناسب است، اما برای تعمیم نتایج به کل جامعه مشتریان کافی نیست.

دومین محدودیت، تجمیع داده در سطح مشتری است. داده‌های محصول، سفارش، کمپین، حاشیه سود، تاریخ تراکنش، مسیر خرید و تماس خدمات مشتری در اختیار نبودند.

سومین محدودیت، مشاهده‌ای‌بودن تحلیل تخفیف است. بدون Randomization نمی‌توان اثر واقعی تخفیف را بر spending یا profit تعیین کرد.

چهارمین محدودیت، نبود outcome آینده برای Churn است. به همین دلیل فقط از مفهوم «در معرض غیرفعال‌شدن» استفاده کردم.

پنجمین محدودیت، فرضی‌بودن وزن‌های Loyalty Score است. تغییر وزن‌ها می‌تواند رتبه مشتریان را تغییر دهد؛ بنابراین تحلیل حساسیت برای نسخه بعدی ضروری است.

ششمین محدودیت، مسئله حریم خصوصی است. در صورت انتشار عمومی، نام‌ها و شناسه‌های قابل شناسایی باید ناشناس‌سازی شوند.

16. مسیر توسعه آینده

برای توسعه پروژه، داده‌های Transaction ID، تاریخ تراکنش، دسته محصول، تعداد کالا، درآمد ناخالص، مبلغ تخفیف، درآمد خالص، حاشیه سود، مواجهه با کمپین، نشست وب‌سایت، رهاکردن سبد خرید و تماس خدمات مشتری را اضافه خواهم کرد.

با این داده‌ها می‌توانم RFM استاندارد، Churn Prediction، Cohort Analysis، Customer Lifetime Value، A/B Testing، Uplift Modeling، Propensity Modeling و بهینه‌سازی مبتنی بر سود را توسعه دهم.

17. نتیجه‌گیری نهایی

در این پروژه، یک جریان کامل تحلیل مشتری را از ورود داده و کنترل کیفیت تا تفسیر مدیریتی اجرا کردم. ابتدا ساختار فایل و نوع داده‌ها را بررسی کردم. سپس پروفایل مشتریان، عملکرد شهرها، وفاداری، ریسک غیرفعال‌شدن، تخفیف و روش پرداخت را تحلیل کردم و در پایان یافته‌ها را به یک گزارش پنج‌دقیقه‌ای مدیریتی تبدیل کردم.

مهم‌ترین نتیجه این بود که ارزش مشتری مفهومی چندبعدی است. مشهد و اصفهان دو نوع موفقیت متفاوت را نشان دادند؛ مشهد در مقیاس درآمد و تکرار خرید برجسته بود و اصفهان در ارزش متوسط و رضایت. سه مشتری 1010، 1035 و 1053 در معرض غیرفعال‌شدن قرار داشتند، اما تفاوت رضایت آن‌ها نشان داد که اقدام نگهداشت باید شخصی‌سازی شود.

تحلیل تخفیف نیز نشان داد که اختلاف کوچک در تعداد خرید، به‌تنهایی برای اثبات موفقیت تخفیف کافی نیست. روش پرداخت نیز بیشتر برای شناخت پروفایل مشتری مفید است و نه برای اثبات اثر علّی.

در نهایت، چارچوب این پروژه را این‌گونه جمع‌بندی می‌کنم:

> **شناسایی ارزش → شناسایی ریسک → تشخیص علت → شخصی‌سازی مداخله → سنجش اثر واقعی**