

# تحليل سناريو كسب و کار موقعيت شغلي تحليلگر ديتا مهر 1404

نام كاندیدا: محمدصادق ميرشمسي  
تاريخ ارسال سناريو: 04/07/27  
مهلت انجام سناريو: 04/07/29

## هدف و شرح سناریو

این تسک برای ارزیابی توانایی‌ها در تحلیل داده، مدلسازی پیش‌بینی و ارائه گزارش طراحی شده. سناریو شبیه‌سازی شده‌ای از داده‌های مرکز تماس در اختیار قرار گرفته و انتظار داریم با بررسی داده‌ها، بینش‌های کلیدی استخراج کنی، یک مدل پیش‌بینی طراحی کنی و در نهایت نتایج رو در قالب یک داشبورد ساده و کاربردی نمایش بدی. هدف ما اینه که بفهمیم چطور از داده برای تصمیم‌سازی استفاده می‌کنی و تواناییت در برقراری ارتباط بین داده، تحلیل و ارائه نتیجه چطوره.

## داده‌ها و منابع

شما یک فایل داده با عنوان زیر در اختیار دارید:

### • [Customer Interactions](#)

شامل داده‌های شبیه‌سازی شده از تعاملات مشتریان با مرکز تماس، شامل نوع تماس (تماس، ایمیل، چت)، زمان پاسخ‌گویی، نوع مشکل، امتیاز رضایت مشتری، و یک ستون متنی (transcript\_excerpt) برای بررسی محتوای مکالمه.

## پرسش‌های کلیدی

### چه بینش‌های مهمی از تحلیل داده به دست آوردی؟

حداقل سه insight استخراج کن. برای مثال می‌تونی بررسی کنی که:

- کدام نوع مشکلات با رضایت پایین‌تر همراه هست.

🔗 **تحويل:** خلاصه‌ای کوتاه از بینش‌ها + نمودار یا گراف پشتیبان

### یک مدل پیش‌بینی ساده برای امتیاز رضایت بساز.

این مسئله رو می‌تونی به عنوان classification یا regression ببینی (بسته به رویکردت). از هر الگوریتمی که باهات راحت‌تری استفاده کن (مثلاً XGBoost، Logistic Regression، Random Forest و ...). توضیح بده که چه مراحل برای آماده‌سازی داده انجام دادی، چه فیچرهایی استفاده کردی، چطور مدل رو ارزیابی کردی و نتایج چی بودن.

🔗 **تحويل:** فایل کد (Python یا R ترجیحاً) + خلاصه ۱ صفحه‌ای از مدل و نتایج

### یک داشبورد ساده و تمیز طراحی کن.

داشبوردت باید اطلاعات کلیدی رو به‌وضوح نمایش بده. لطفا علاوه بر موارد زیر چند مورد دیگه هم خودت اضافه کن:


- میانگین رضایت بر اساس نوع تعامل و نوع مشکل
- میانگین زمان پاسخ‌گویی
- روند امتیاز رضایت در طول زمان

ابزار مورد استفاده می‌تونه Excel، Power BI، Tableau یا هر ابزار آشنای دیگه‌ای باشه.

🔗 **تحويل:** فایل داشبورد (یا اسکرین‌شات در صورت نیاز)

---

### (اختیاری - امتیاز اضافه) تحلیل متنی ساده با NLP انجام بده.

با استفاده از ستون transcript\_excerpt بررسی کن که چه کلمات یا عبارتهایی بیشتر در تعاملاتی دیده می‌شن که امتیاز رضایت پایین (۱ یا ۲) داشتن. می‌تونی فقط یک لیست از کلمات پرتکرار بدی یا به word cloud ساده طراحی کنی.  **تحويل:** لیست کلمات یا تصویر word cloud

### نحوه پاسخ‌دهی

---

لطفاً پاسخ خود را به صورت یک گزارش تحلیلی در صفحه بعدی این Google Doc ثبت نمایید، به شکلی که پاسخی تحلیلی، دقیق و متمرکز بر اقدامات کلیدی باشد. هدف ما دریافت تحلیلی است که به روشنی نشان دهد رویکرد شما چگونه می‌تواند در حل مسئله نقش‌آفرین باشد؛ نه یک گزارش طولانی یا ارائه پرزنتیشن. همچنین اقداماتی که روی دیتا انجام دادین را به شکل یک فایل جداگانه برای ما ارسال کنید.

با توجه به حجم و ماهیت این تسک، برآورد می‌شود تکمیل آن بین 3 تا 5 ساعت زمان نیاز داشته باشد. مهلت ارسال پاسخ تا **29 مهر** تعیین شده است. پس از تکمیل، لطفاً انجام تسک را از طریق ایمیل به **ماهور محمدی** به آدرس [mahour.mohammadi@zoodfood.com](mailto:mahour.mohammadi@zoodfood.com) اطلاع دهید.

در صورت وجود هرگونه پرسش یا نیاز به توضیحات بیشتر، می‌توانید از طریق Microsoft Teams با **ماهور محمدی** در ارتباط باشید.

### محل پاسخ و تحلیل شما

---

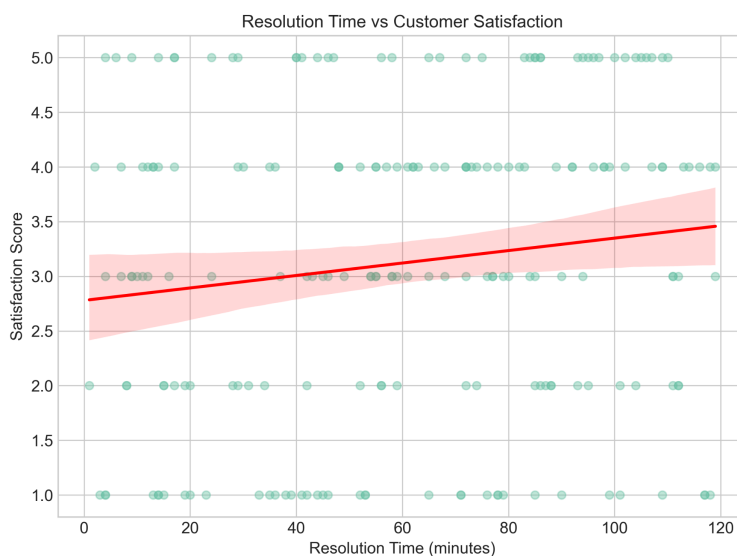
لطفاً در این بخش، تحلیل و راهکارهای پیشنهادی خود را به شکلی شفاف، ساختار یافته و متمرکز بر اقدامات کلیدی ارائه دهید. در صورت نیاز می‌توانید از داده‌ها، جداول یا نمودارها استفاده کنید. لطفاً در طول پاسخ‌گویی به داده‌ها و شواهد ارائه‌شده استناد کنید و در صورت استفاده از منابع خارجی، آن‌ها را ذکر نمایید.

## قسمت اول

در این تحلیل، داده‌های تعامل مشتریان با تیم پشتیبانی بررسی شد تا الگوهای رضایت، نوع مشکل و نحوه‌ی تعامل مشخص شوند. هدف این بود که بفهمیم چه عواملی بیشترین تأثیر را بر رضایت مشتری دارند و در چه حوزه‌هایی می‌توان فرآیند پاسخ‌گویی را بهبود داد.

### بینش 1 – رابطه‌ی زمان حل مسئله و رضایت مشتری:

تحلیل رابطه‌ی میان زمان حل درخواست‌ها و نمره‌ی رضایت مشتریان نشان می‌دهد که برخلاف انتظار، ارتباط میان این دو متغیر ضعیف اما اندکی مثبت است؛ یعنی در برخی موارد، تعاملاتی که زمان بیشتری برای رسیدگی نیاز داشته‌اند، رضایت بالاتری نیز کسب کرده‌اند. این موضوع می‌تواند بیانگر آن باشد که مشتریان، در کنار سرعت پاسخ‌گویی، به کیفیت و دقت در حل مسئله نیز اهمیت می‌دهند. بنابراین صرفاً کاهش زمان پاسخ‌گویی همیشه به افزایش رضایت منجر نمی‌شود و لازم است تمرکز شرکت بر ارائه‌ی تجربه‌ای شفاف، دقیق و با ارتباط مؤثر در طول فرآیند حل مسئله باشد. این بینش تأکید می‌کند که «کیفیت تعامل» در بسیاری از موارد مهم‌تر از «سرعت پاسخ» است.



### اعداد کلیدی:

- Pearson  $r = 0.138$
- Spearman  $\rho = 0.141$
- p-value = **0.0464** (در آستانه معنی‌داری آماری)

ضریب همبستگی پیرسون برابر با **0.138** و اسپیرمن برابر با **0.141** است که هر دو نشان‌دهنده‌ی یک **رابطه‌ی مثبت ضعیف** میان زمان حل مسئله و میزان رضایت مشتری هستند. به بیان ساده‌تر، هرچه زمان رسیدگی کمی افزایش یابد، رضایت مشتری نیز اندکی بیشتر می‌شود، هرچند این اثر بسیار جزئی است.

مقدار **p-value = 0.0464** نشان می‌دهد که این رابطه در سطح اطمینان ۹۵٪ به‌طور مرزی **معنی‌دار آماری** است؛ یعنی احتمال اینکه این الگو به‌صورت تصادفی رخ داده باشد کمتر از ۵٪ است. با این حال، شدت رابطه پایین است و از نظر عملی باید محتاطانه تفسیر شود.

## بینش 2: نوع مشکل و تأثیر آن بر رضایت مشتری

این تحلیل بررسی می‌کند که آیا نوع مشکلی که مشتری تجربه کرده، بر میزان رضایت او تأثیر دارد یا خیر. نتایج نشان می‌دهد که مشتریانی که با **مشکلات مالی (billing)** روبه‌رو شده‌اند، به‌طور میانگین **کمترین رضایت** را داشته‌اند، در حالی که رضایت در میان افرادی که با **مشکلات لغو (cancellation)** یا **سؤالات عمومی (general inquiry)** مواجه بوده‌اند، اندکی بالاتر است. به‌طور کلی، تفاوت‌ها میان گروه‌ها زیاد چشمگیر نیست اما الگوی کلی نشان می‌دهد که **مشکلات مالی و فنی بیش از سایر حوزه‌ها منجر به ناراضایتی می‌شوند** – موضوعی که می‌تواند نشانه‌ای از فرآیندهای پیچیده یا تجربه کاربری ضعیف در آن بخش‌ها باشد.



برای بررسی معناداری تفاوت رضایت بین انواع مشکلات، از آزمون **Kruskal-Wallis** استفاده شد. نتایج زیر به‌دست آمد:

- **H = 7.703**
- **p-value = 0.1031**

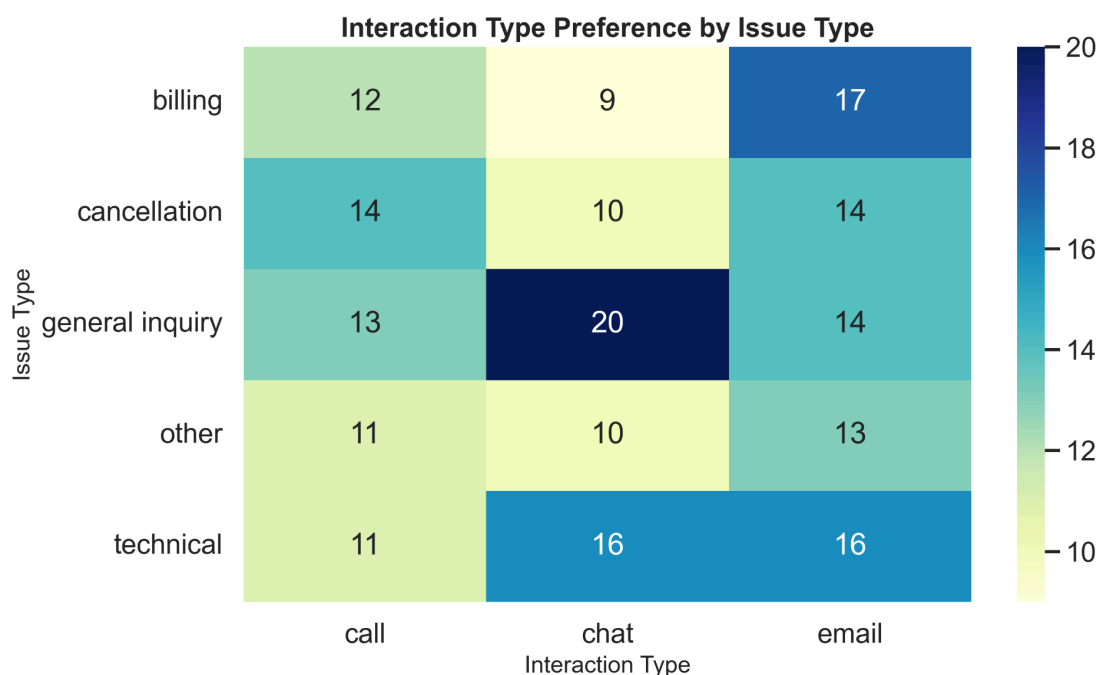
از آنجا که مقدار  $p$  بالاتر از سطح معنی‌داری 0.05 است، نمی‌توان با اطمینان گفت تفاوت رضایت بین انواع مشکلات از نظر آماری معنی‌دار است. به بیان ساده‌تر، اگرچه اختلاف‌هایی در میانگین رضایت مشاهده می‌شود، اما این تفاوت‌ها ممکن است تصادفی باشند.

### بینش 3 : ترجیحات نوع ارتباط بر اساس نوع مشکل

این تحلیل بررسی می‌کند که مشتریان با توجه به نوع مشکلی که دارند، کدام کانال ارتباطی (تماس، چت، ایمیل) را ترجیح می‌دهند.

نتایج نشان می‌دهد که در موارد **سؤالات عمومی**، **چت** محبوب‌ترین گزینه است، زیرا پاسخ‌ها سریع‌تر و تعامل سبک‌تر است. در مقابل، برای **صورتحساب و لغو خدمات**، استفاده از **ایمیل و تماس تلفنی** رایج‌تر است که نشان می‌دهد مشتریان برای مسائل رسمی یا حساس‌تر، کانال‌های قابل پیگیری‌تر را ترجیح می‌دهند.

اما نکته‌ی قابل توجه این است که در **مشکلات فنی (technical issues)**، میزان استفاده از **چت و ایمیل** تقریباً برابر و در سطح بالایی قرار دارد. این موضوع می‌تواند نشان دهد که کاربران برای مسائل فنی، به دنبال پاسخ سریع از طریق چت هستند، ولی در عین حال تمایل دارند جزئیات فنی و راهکارهای دقیق‌تر را از طریق ایمیل دریافت کنند. در نتیجه، ترکیب این دو کانال برای مشکلات فنی می‌تواند مؤثرترین رویکرد پشتیبانی باشد.

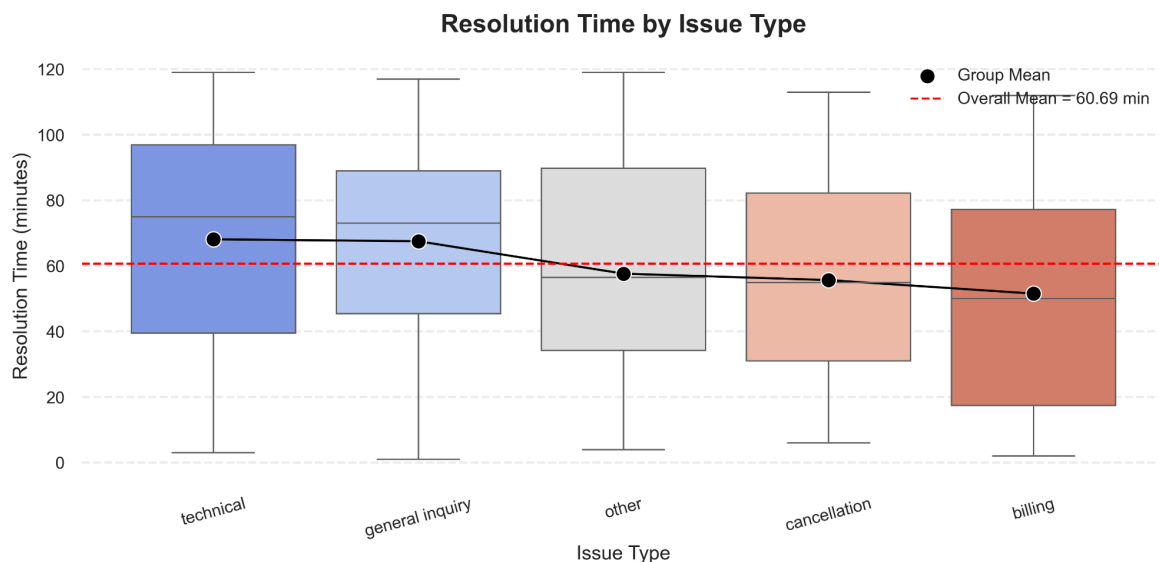


اگرچه آزمون کای-دو ( $p\text{-value} = 0.70$ ) نشان می‌دهد که تفاوت‌ها از نظر آماری معنادار نیستند، اما این الگوها از نظر رفتاری اهمیت دارند و می‌توانند در بهینه‌سازی توزیع نیروی پشتیبانی در کانال‌های مختلف مورد استفاده قرار گیرند.

#### بینش 4 : زمان حل مسئله بر اساس نوع مشکل

این تحلیل با هدف مقایسه‌ی میانگین زمان رسیدگی به مشکلات مختلف انجام شده است تا مشخص شود کدام نوع از درخواست‌ها به‌طور معمول زمان بیشتری برای حل شدن نیاز دارند. نتایج نشان داد که هرچند بین انواع مشکلات (مانند صورتحساب، لغو خدمات، مشکلات فنی و...) تفاوت‌هایی در میانگین زمان حل وجود دارد، اما طبق آزمون **Kruskal-Wallis** با مقدار آماره 7.63 و  $p\text{-value} = 0.106$ ، این تفاوت‌ها از نظر آماری معنادار نیستند.

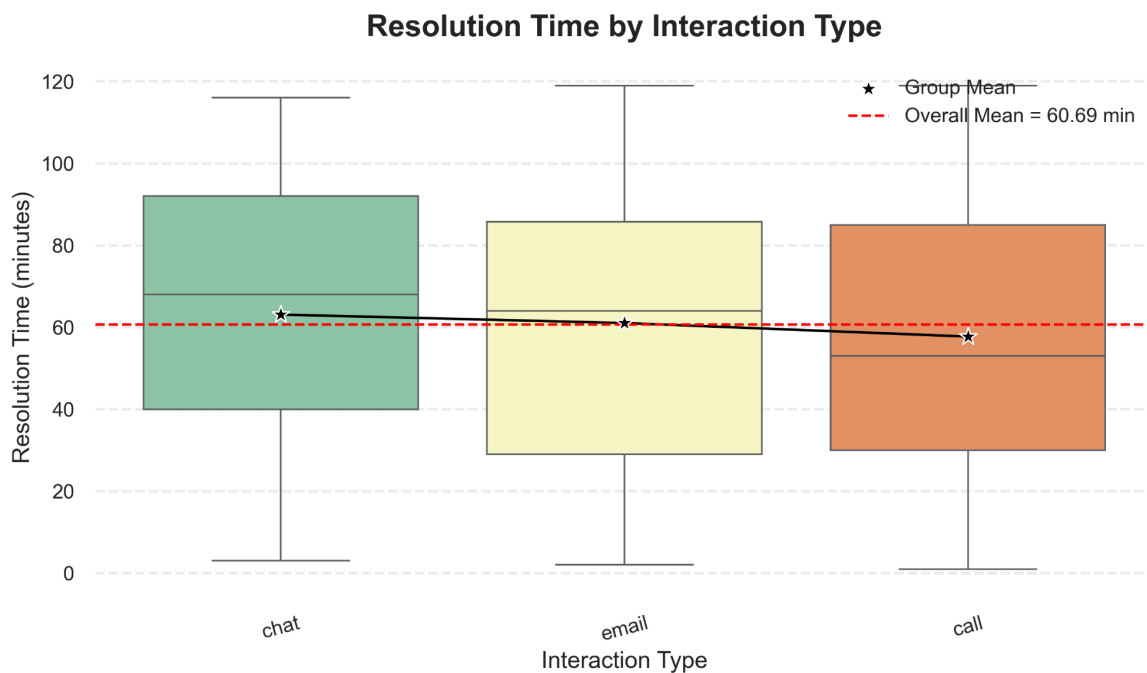
به بیان ساده‌تر، می‌توان گفت که سیستم پشتیبانی در حل مشکلات مختلف، عملکردی نسبتاً یکنواخت دارد و هیچ گروه خاصی از مشکلات به شکل چشمگیری زمان‌برتر از بقیه نیست. این نتیجه نشان‌دهنده‌ی کارایی نسبی فرآیندهای پشتیبانی در مدیریت تنوع موضوعات مشتریان است. با این حال، بررسی دقیق‌تر توزیع زمان حل در هر نوع مشکل می‌تواند به شناسایی نقاطی که گاهی تأخیر ایجاد می‌شود (مثلاً در موارد فنی یا مالی پیچیده‌تر) کمک کند.



## بینش 5: زمان حل مسئله بر اساس نوع کانال ارتباطی

این تحلیل با هدف بررسی ارتباط بین نوع کانال ارتباطی (چت، ایمیل و تماس تلفنی) و مدت زمان حل مشکلات مشتریان انجام شد. نتایج نشان داد که میانگین زمان حل در سه کانال تفاوت قابل توجهی ندارد – به طور کلی، میانگین کل حدود ۶۰ دقیقه بوده و تمامی کانال ها حول همین مقدار نوسان دارند. آزمون آماری **Kruskal-Wallis** با مقدار آماره 0.84 و  $p\text{-value} = 0.6567$  نیز تأیید می کند که این تفاوت از نظر آماری معنادار نیست.

به زبان ساده، عملکرد تیم پشتیبانی در سه کانال تقریباً هم سطح و پایدار است؛ یعنی فرقی نمی کند مشتری از طریق چت، تماس یا ایمیل ارتباط بگیرد، زمان رسیدگی مشابهی تجربه می کند. این موضوع نشانه ی مثبتی از استانداردسازی و هماهنگی فرآیندهای پاسخ گویی در کانال های مختلف است. با این حال، می توان بررسی کرد که آیا برخی کانال ها (مثل چت) در نوع خاصی از مشکلات سریع تر عمل می کنند یا خیر تا فرصت های بهینه سازی بیشتری شناسایی شود.





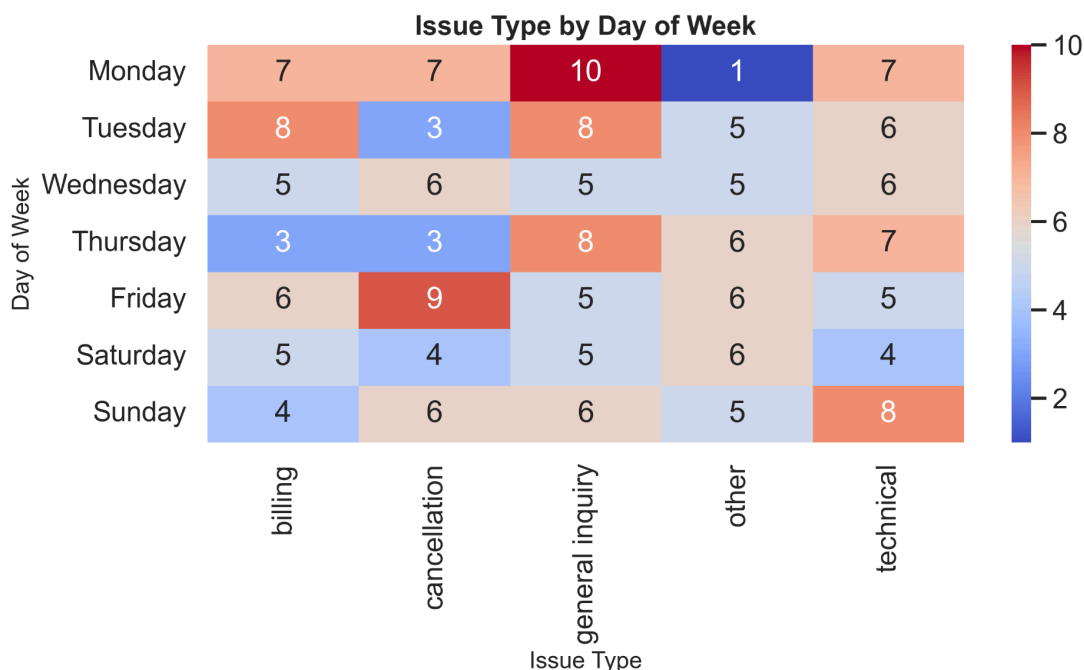
## بینش 6: توزیع نوع مشکلات در روزهای مختلف هفته

در این تحلیل، توزیع نوع مشکلات کاربران در روزهای مختلف هفته بررسی شد. همان‌طور که در نمودار دیده می‌شود، الگوی کلی نشان می‌دهد که اگرچه تفاوت چشمگیری میان روزهای هفته وجود ندارد، اما برخی روندهای رفتاری قابل توجه هستند.

در روز جمعه که تعطیل رسمی و پایان هفته کاری محسوب می‌شود، تعداد درخواست‌های لغو خدمات (cancellation) افزایش محسوسی دارد. این موضوع احتمالاً به این دلیل است که کاربران در پایان هفته زمان بیشتری برای بررسی سرویس‌های خود دارند و تصمیم می‌گیرند پیش از شروع هفته‌ی جدید، نسبت به لغو یا تنظیم مجدد خدمات اقدام کنند.

در مقابل، در روزهای میانی هفته (سه‌شنبه و چهارشنبه) سهم بیشتری از سؤالات عمومی دیده می‌شود؛ همچنین در روزهای آغازین هفته (یکشنبه و دوشنبه) سهم بیشتری از مشکلات فنی دیده می‌شود که می‌تواند ناشی از اوج استفاده از خدمات در روزهای کاری باشد.

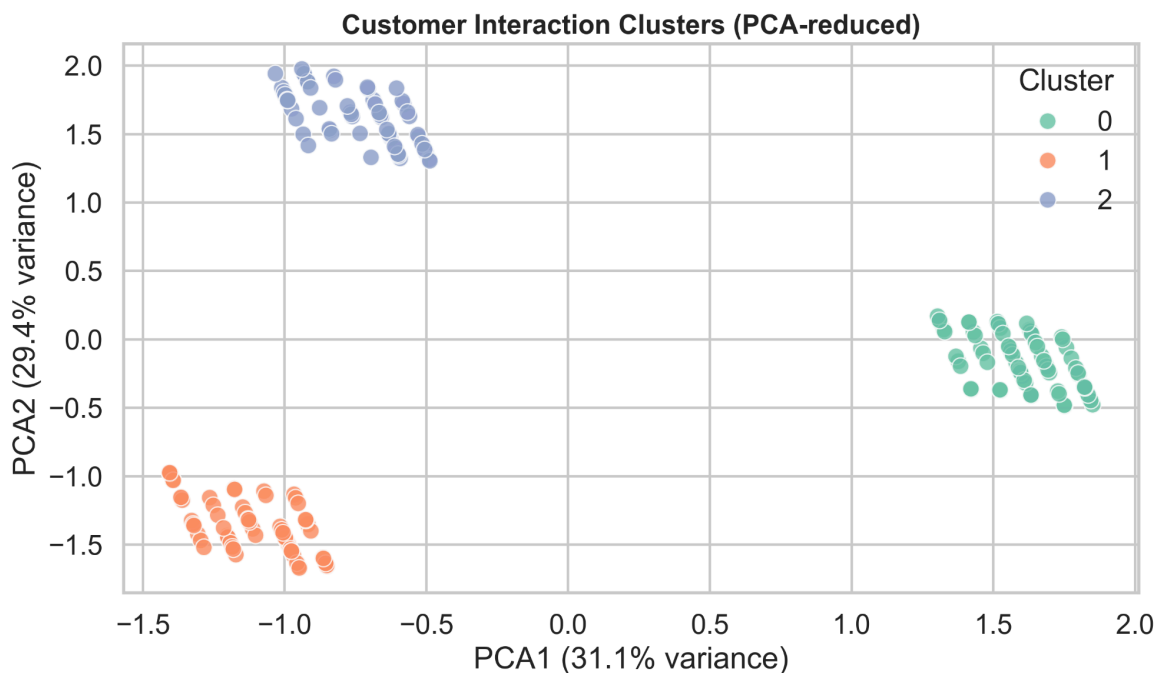
به طور کلی، این الگو نشان می‌دهد که نوع مشکل کاربران ارتباط قوی با روز هفته ندارد، اما برخی نوسانات طبیعی در ابتدای و انتهای هفته وجود دارد. این بینش می‌تواند برای برنامه‌ریزی نیروی انسانی مفید باشد، به‌ویژه در تخصیص پشتیبان‌های فنی در روزهای پایانی هفته و نیروهای مرتبط با لغو و سؤالات عمومی در ابتدای هفته.



## بینش 7: تفکیک گروه‌های مشتری بر اساس رضایت و سرعت پاسخ‌گویی

نتایج نشان می‌دهد که هر سه خوشه از نظر تعداد تعامل تقریباً مشابه‌اند (میانگین ۱ تعامل در هر خوشه)، اما از نظر میانگین رضایت و زمان متوسط رفع مشکل تفاوت‌های ظریفی وجود دارد. خوشه ۰ با میانگین رضایت ۳.۲۴ و زمان پاسخگویی حدود ۶۱ دقیقه بهترین وضعیت را دارد، در حالی که خوشه ۱ پایین‌ترین میزان رضایت (۳.۰۴) و طولانی‌ترین زمان پاسخ (۶۳ دقیقه) را نشان می‌دهد. خوشه ۲ نیز رضایتی نزدیک به خوشه ۱ دارد اما با زمان پاسخگویی کوتاه‌تر (۵۷ دقیقه) عملکرد متعادل‌تری دارد.

به طور کلی، می‌توان گفت که رضایت مشتری ارتباط مستقیمی با سرعت پاسخ‌گویی دارد، هرچند تفاوت‌ها جزئی هستند. این تحلیل می‌تواند مبنایی برای شخصی‌سازی تجربه مشتریان باشد؛ به‌طور مثال، خوشه‌هایی با رضایت پایین‌تر می‌توانند هدف برنامه‌های بهبود خدمات یا پشتیبانی سریع‌تر قرار گیرند.



خوشه	تعداد مشتریان	میانگین تعامل	میانگین رضایت	میانگین زمان رفع مشکل
0	74	1	3.24	61.03
1	65	1	3.05	63.09
2	61	1	3.07	57.072

قسمت دوم

## مدل پیش‌بینی امتیاز رضایت.

در این پروژه هدف اصلی، پیش‌بینی رضایت مشتریان بر اساس داده‌های مربوط به تعاملات آن‌ها با مرکز پشتیبانی است.

داده‌ها شامل اطلاعات ساختاری (زمان، نوع تعامل، نوع مشکل) و متنی (بخشی از گفت‌وگوی مشتری با پشتیبان) هستند. این مسئله به صورت دسته‌بندی دودویی (Binary Classification) در نظر گرفته شد تا مشتریان به دو گروه تقسیم شوند:

- ۰ → ناراضی یا رضایت پایین
- ۱ → رضایت متوسط یا بالا

برای مدل‌سازی از الگوریتم رگرسیون لجستیک (Logistic Regression) به همراه جست‌وجوی شبکه‌ای (GridSearchCV) جهت تنظیم ابرپارامترها استفاده شد.

بهترین تنظیمات به دست آمده:

Best Params: {

'clf\_\_C': 0.1, 'clf\_\_penalty': 'l2', 'clf\_\_solver': 'liblinear', 'clf\_\_fit\_intercept': False

}

Best CV F1 Score: 0.61

## مراحل آماده‌سازی داده

### ۱. پاک‌سازی داده‌ها

- حذف نویزها شامل لینک‌ها، منشن‌ها، اعداد و کاراکترهای غیرضروری با استفاده از عبارات منظم (Regex).
- نرمال‌سازی متون فارسی با تبدیل حروف عربی به فارسی (مثلاً «ی» → «ی»، «ک» → «ک»).
- توکنیزه کردن جملات با استفاده از کتابخانه Hazm.
- حذف کلمات توقف (Stopwords) مانند «از»، «در»، «که»، «اما» برای تمرکز مدل بر کلمات معنی‌دار.

### ۲. استخراج ویژگی‌های متنی

از هر گفت‌وگو چند ویژگی متنی استخراج شد:

- word\_count: تعداد کل واژه‌ها.
- avg\_word\_length: میانگین طول واژه‌ها.
- exclamation\_count: تعداد علامت‌های تعجب (نشانه‌ی هیجان یا عصبانیت).
- question\_count: تعداد علامت سؤال (نشانه‌ی تردید یا ناراضی).

### ۳. استخراج ویژگی‌های زمانی

از ستون تاریخ تعامل (`interaction_date`) ویژگی‌های زیر استخراج شد:

- روز (`day`)، ماه (`month`)، سال (`year`)، فصل (`quarter`)
- متغیر باینری `is_holiday` برای تشخیص تعطیلی رسمی یا روز جمعه (با استفاده از کتابخانه `holidays`).

### ۴. تبدیل ویژگی‌های متنی به بردار عددی

- متون پاک‌شده به کمک `TF-IDF` به بردار عددی تبدیل شدند.
- سپس برای کاهش ابعاد از `SVD (TruncatedSVD)` با ۲۰ مؤلفه استفاده شد تا نویز کاهش یابد و آموزش مدل سریع‌تر شود.

### ۵. ویژگی‌های ساختاری

ویژگی‌های عددی و طبقه‌ای شامل:

- زمان حل مشکل (`resolution_time_minutes`)
  - نوع تعامل (`interaction_type`)
  - نوع مسئله (`issue_type`)
- به‌صورت `One-Hot Encoding` تبدیل و در مدل لحاظ شدند.

### نتایج مدل

	recall	f1-score	precision
0	0.77	0.73	0.70
1	0.61	0.65	0.69
accuracy		0.69	

مدل توانست به دقت کلی  $\approx 69\%$  برسد، که برای داده‌های متنی و دودویی سطح مناسبی محسوب می‌شود. عملکرد کلاس ۰ (نارضایتی) اندکی بهتر از کلاس ۱ است، که نشان می‌دهد مدل در تشخیص نشانه‌های نارضایتی در متن مؤثرتر عمل کرده است.

## ابرکمات

