

## مرحله اول: فراخوانی و تمیزکاری اولیه داده‌ها

در گام نخست، داده‌ها را با استفاده از پایتون فراخوانی کرده و با دستور `info()` ساختار کلی دیتاست را بررسی کردیم. در این مرحله از بین ۴۴۷۳ ردیف موجود، مشخص شد که برخی ستون‌ها دارای داده‌های گمشده (Missing) هستند و نیاز است این مقادیر با مقادیر استاندارد جایگزین شوند.

از میان ۳۵ ستون دیتاست، تنها ستون‌های زیر دارای مقادیر خالی بودند:

`Grade, DepartmentGrade, Nezam, SkillName, SkillFullText, DoubleUserID, DoubleUserName, VisitDesc, AddressLAT, AddressLANG.`

در ادامه:

- ابتدا نام ستون‌ها را به فرمت استاندارد تغییر دادیم (`rename`) و داده‌ها را به فرمت صحیح تبدیل کردیم.
- برای ستون‌های متنی (`String`) مقادیر خالی با مقدار "No-Data" پر شد.
- برای ستون‌های عددی، مقادیر خالی با رشته خالی ("") جایگزین شدند.

---

## تمیزکاری داده‌های تاریخ و زمان

- سپس به تمیزکاری داده‌های مربوط به تاریخ پرداختیم. تاریخ‌ها باید در قالب استاندارد `YYYY-MM-DD` باشند:
- ابتدا سال با چهار رقم درج شد.
- سپس ماه به صورت دو رقمی و پس از آن روز نیز به صورت دو رقمی نوشته شد.

برای داده‌های مکانی: (`Location`)

- مقدار `Latitude` باید در بازه ۹۰- تا ۹۰ باشد.
- مقدار `Longitude` باید در بازه ۱۸۰- تا ۱۸۰ قرار گیرد.

در بررسی داده‌ها مشاهده شد که تمامی رکوردها مقداردهی شده‌اند، اما پس از بررسی محدوده‌ها مشخص شد ۴۱۵ رکورد دارای مختصات نامعتبر هستند. این رکوردها حذف شدند و ادامه تحلیل روی داده‌های معتبر انجام شد.

برای محاسبه مدت زمان حضور در هر ویزیت، ستون‌های مربوط به تاریخ و ساعت شروع (`Start`) و پایان (`End`) را به یکدیگر متصل کرده و اختلاف آن‌ها محاسبه شد. نتایج نشان داد تمامی زمان‌ها کمتر از یک ساعت بوده‌اند.

---

## آماده‌سازی داده‌ها برای تحلیل

در گام بعد، به منظور تحلیل داده‌ها، تعداد کاربران یکتا در ستون `user_name` بررسی شد. سپس برای سهولت در مصورسازی و گروه‌بندی، به هر کاربر یک کد حرفی انگلیسی از `a` تا `q` اختصاص داده شد. این کار از طریق:

1. ساخت یک لیست از حروف الفبای انگلیسی،
2. شناسایی نام‌های یکتا در ستون `user_name`،
3. و نگاشت هر کاربر به یک حرف، انجام شد.

سوال اول: میانگین فروش ماهانه براساس نام گروه و تنها فروش ها موفق

اسم فرد	نام گروه
d	مهلا بخشی
c	سحر پسوه ای
b	تینا حمزه نژاد
a	مهرداد صالحی
h	نگار قاصدی
g	غزاله مهرجو
f	شیوا شجاعی
e	زهرا حقیقی
i	فاطمه کریمی

### تحلیل میانگین ویزیت‌های ماهانه (موفق)

برای بررسی میانگین ویزیت‌های ماهانه، ابتدا ویزیت‌های موفق شناسایی شدند. سپس بر اساس کدهای نماینده‌ها (که به هر فرد اختصاص داده شده بود)، بررسی شد که هر نماینده در هر ماه و به صورت روزانه چه تعداد ویزیت موفق داشته است. این تحلیل هم در فایل کد مشخص است و هم در نمودارهای ترسیم‌شده به وضوح دیده می‌شود.

نکات برجسته بر اساس ماه‌ها:

- **خرداد ماه:** نماینده **d** (مهلا بخشی) بهترین عملکرد را داشته است.
- **تیر ماه:** عملکرد نمایندگان به هم نزدیک بوده، اما **d** (مهلا بخشی) و **c** (سحر پسوه‌ای) اندکی بهتر از بقیه بودند.
- **مرداد ماه:** فروش موفق به طور کلی افزایش یافته و نماینده **a** (مهرداد صالحی) بیشترین عملکرد را داشته است.
- **شهریور ماه:** نمایندگان **e** و **d** بهترین نتایج را ثبت کرده‌اند.

جمع‌بندی:

همان‌طور که از داده‌ها و نمودارها مشاهده می‌شود، نماینده **d** (مهلا بخشی) در بیشتر ماه‌ها عملکرد برتری نسبت به سایر نماینده‌ها داشته است.

### تحلیل میانگین ویزیت روزانه (بر اساس روزهای کاری)

برای بررسی میانگین ویزیت روزانه، ابتدا مجموع کل ویزیت‌ها از داده‌های ثبت‌شده محاسبه شد. سپس با استفاده از تابع unique، تعداد روزهای کاری یکتا به دست آمد. مطابق صورت مسئله، معیار این است که در هر روز کاری حداقل یک ویزیت انجام شود؛ بررسی‌ها نشان داد که تمام نمایندگان بیش از این حداقل عمل کرده‌اند.

نتایج شاخص:

- **بهترین عملکرد:** نماینده **d** (مهلا بخشی) (بیشترین میانگین ویزیت روزانه را ثبت کرده است).
- **ضعیف‌ترین عملکرد:** نماینده **i** (فاطمه کریمی) کمترین تعداد ویزیت روزانه را داشته است.

### سوال 3

در این بخش هدف، محاسبه میانگین مدتزمان هر ویزیت از لحظه شروع تا پایان است. برای این کار، پیشتر زمان شروع و زمان پایان هر ویزیت به تاریخ مربوطه متصل شد تا اختلاف آن‌ها دقیق‌تر محاسبه شود. بررسی‌ها نشان داد که:

اکثر ویزیت‌ها کمتر از یک ساعت به طول انجامیده‌اند.

تنها استثنا، نماینده (hنگار قاصدی) بوده که برخی ویزیت‌های او طولانی‌تر از یک ساعت ثبت شده است.

نکته تحلیلی:

در صورت داشتن یک سیستم امتیازدهی عملکرد، می‌توانستیم بررسی کنیم که آیا کوتاه‌تر بودن زمان ویزیت نشانه بهبود کارایی است یا خیر؛ زیرا ممکن است ویزیت‌های سریع‌تر کیفیت پایین‌تری داشته باشند و این عامل در نهایت به کاهش رضایت یا عملکرد منجر شود.

### سوال 4

هدف این بخش، بررسی تعداد ویزیت‌ها به تفکیک نوع مرکز (داروخانه یا پزشک) بود تا مشخص شود کدام گروه سهم بیشتری از ویزیت‌ها را به خود اختصاص داده است.

نتیجه تحلیل نشان داد که:

- بیشتر ویزیت‌ها مربوط به داروخانه‌ها بوده است.
- تعداد ویزیت‌های مربوط به پزشکان کمتر از داروخانه‌ها ثبت شده است.

این یافته نشان‌دهنده تمرکز بیشتر فعالیت‌ها بر داروخانه‌ها نسبت به پزشکان است

### سوال 5

در این بخش، هدف بررسی این بود که در دسته‌بندی‌های انجام‌شده، هر نماینده در کدام داروخانه یا پزشک عملکرد بهتری داشته است.

به عنوان نمونه: مهرداد صالحی در مراکز هم‌چون بیمارستان پیامبران، علیرضا احمدی، و بیمارستان امام خمینی بازدیدهای موفق‌تر و بیشتری داشته است.

این تحلیل به شناسایی مراکز کلیدی برای هر نماینده کمک می‌کند و می‌تواند مبنایی برای تمرکز استراتژی‌های بازاریابی و بهبود روابط در مراکز پرتردد باشد.

### سوال 6

#### تحلیل پوشش استان‌ها و شهرها در ویزیت‌ها

هدف این بخش بررسی میزان ویزیت در استان‌ها و شهرهای مختلف بود. نتایج نشان داد:

- استان تهران با اختلاف چشمگیر، بیشترین تعداد ویزیت را دارد و پس از آن استان‌های خراسان رضوی و اصفهان قرار دارند.

- در میان شهرها، تهران، مشهد، اصفهان و تبریز بیشترین بازدیدها را داشته‌اند و به‌عنوان شهرهای کلیدی شناسایی می‌شوند.

همچنین:

- شهرهایی وجود دارند که کمتر از ۱۰ ویزیت ثبت کرده‌اند. این شهرها باید بررسی شوند تا مشخص شود آیا این تعداد کم به دلیل کوچک بودن شهر و محدودیت تعداد داروخانه‌ها و مطب‌هاست یا علت دیگری دارد.
- در مقابل، برای شهرهای بزرگ نیز باید بررسی شود که آیا امکان افزایش بازدید و فروش وجود دارد یا خیر، چرا که پتانسیل بالاتری برای جذب مشتری دارند

توضیحات مربوط به نمودار ها

توزیع وزیت ها براساس نوع ویزیت انجام شد (یعنی برای داروخانه یا دکتر بوده است)

## تحلیل عملکرد نمایندگان در ویزیت داروخانه‌ها و پزشکان

بر اساس تحلیل این نمودار:

- مهلا بخشی و مهرداد صالحی در هر دو بخش ویزیت داروخانه‌ها و ویزیت پزشکان عملکردی خوب و متعادل داشته‌اند.
- زهرا نجفی در ویزیت داروخانه‌ها از همه نمایندگان بهتر عمل کرده است.
- در مقابل، فاطمه کریمی، زهرا حقیقی و نگار قاصدی در هر دو بخش (داروخانه و پزشک) عملکرد ضعیف‌تری داشته‌اند.

این نتایج به وضوح نشان می‌دهد که برخی نمایندگان در هر دو حوزه کارآمد هستند، درحالی‌که برخی دیگر نیاز به آموزش یا بازنگری در برنامه‌ریزی بازدیدها دارند.

## تحلیل نمایندگان برتر و شهرهای کلیدی بازدید

هدف این نمودار، شناسایی نمایندگان برتر و سپس بررسی عملکرد آن‌ها در شهرهای مختلف است. در این تحلیل:

- ابتدا دو نماینده برتر انتخاب شدند:
  - مهلا بخشی
  - مهرداد صالحی
- نتایج نشان داد:
  - مهرداد صالحی بیشترین بازدید را در شهر تهران داشته است.
  - مهلا بخشی بیشترین بازدید را در استان خراسان رضوی انجام داده است.

این تحلیل به خوبی نشان می‌دهد که مهلا بخشی و مهرداد صالحی کارشناسان برتر تیم هستند و در مناطق کلیدی نیز بیشترین فعالیت را داشته‌اند.

