تحلیل و مصورسازی داده‌های حوزه سلامت

یک داشبورد تعاملی با پایتون و استریم‌لیت

ارائه‌دهنده: کسری عبدی

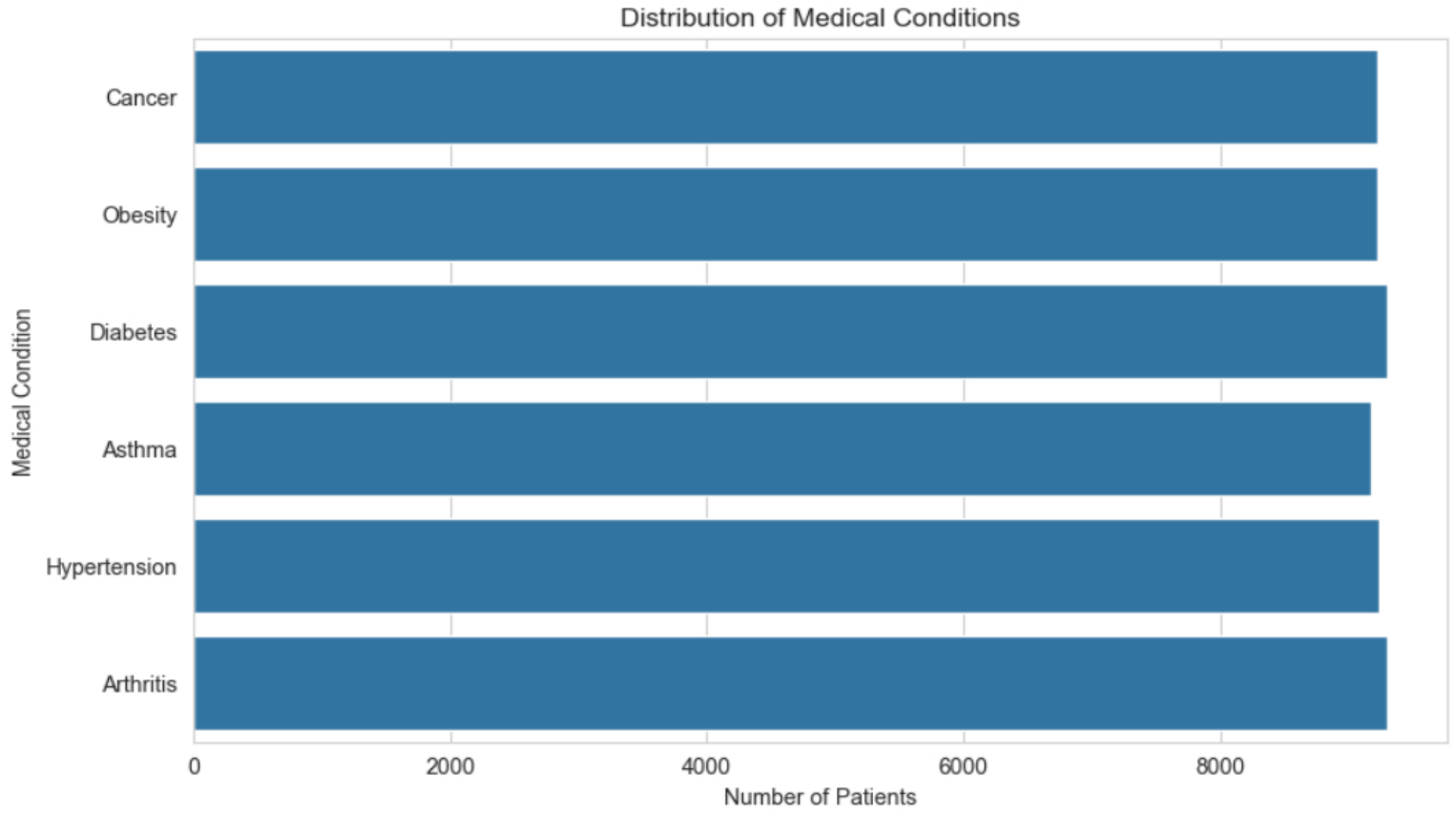
مقدمه:

- هدف اصلی این پروژه، طراحی یک داشبورد تحلیلی تعاملی برای کاوش در داده‌های بیماران است.
- این داشبور به کاربر اجازه می‌دهد تا الگوهای پنهان در داده‌ها را کشف کند.
- در این تحلیل، روابط میان شرایط پزشکی، اطلاعات دموگرافیک بیماران (سن، جنسیت) و هزینه‌های درمانی بررسی شده است.

درباره دیتاست:

- نام فایل : healthcare_dataset
- تعداد رکوردها: ۱۰٬۰۰۰ بیمار
- ویژگی‌های کلیدی:
- اطلاعات دموگرافیک (سن، جنسیت، گروه خونی)
- شرایط پزشکی (آسم، سرطان، دیابت و...)
- اطلاعات پذیرش (نوع و تاریخ پذیرش)
- هزینه‌های درمانی (Billing Amount)

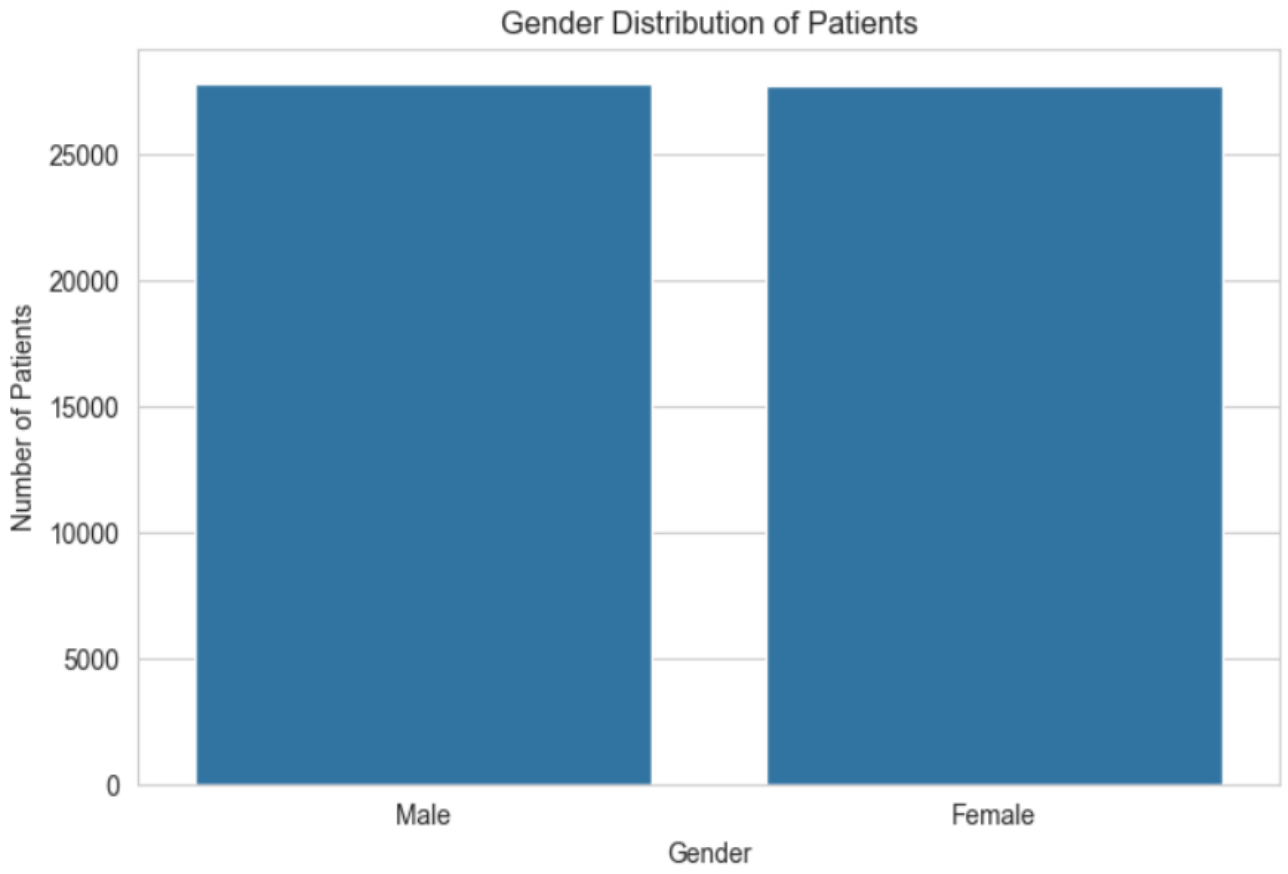
تحلیل توزیع شرایط پزشکی:



یافته کلیدی: توزیع بیماران در میان شش بیماری اصلی (سرطان، دیابت و...) تقریباً یکسان و متوازن است.

اهمیت: این توازن نشان می‌دهد که داده‌ها برای تحلیل مقایسه‌ای مناسب هستند و نتایج ما تحت تأثیر تعداد بیشتر یک بیماری خاص قرار نمی‌گیرد.

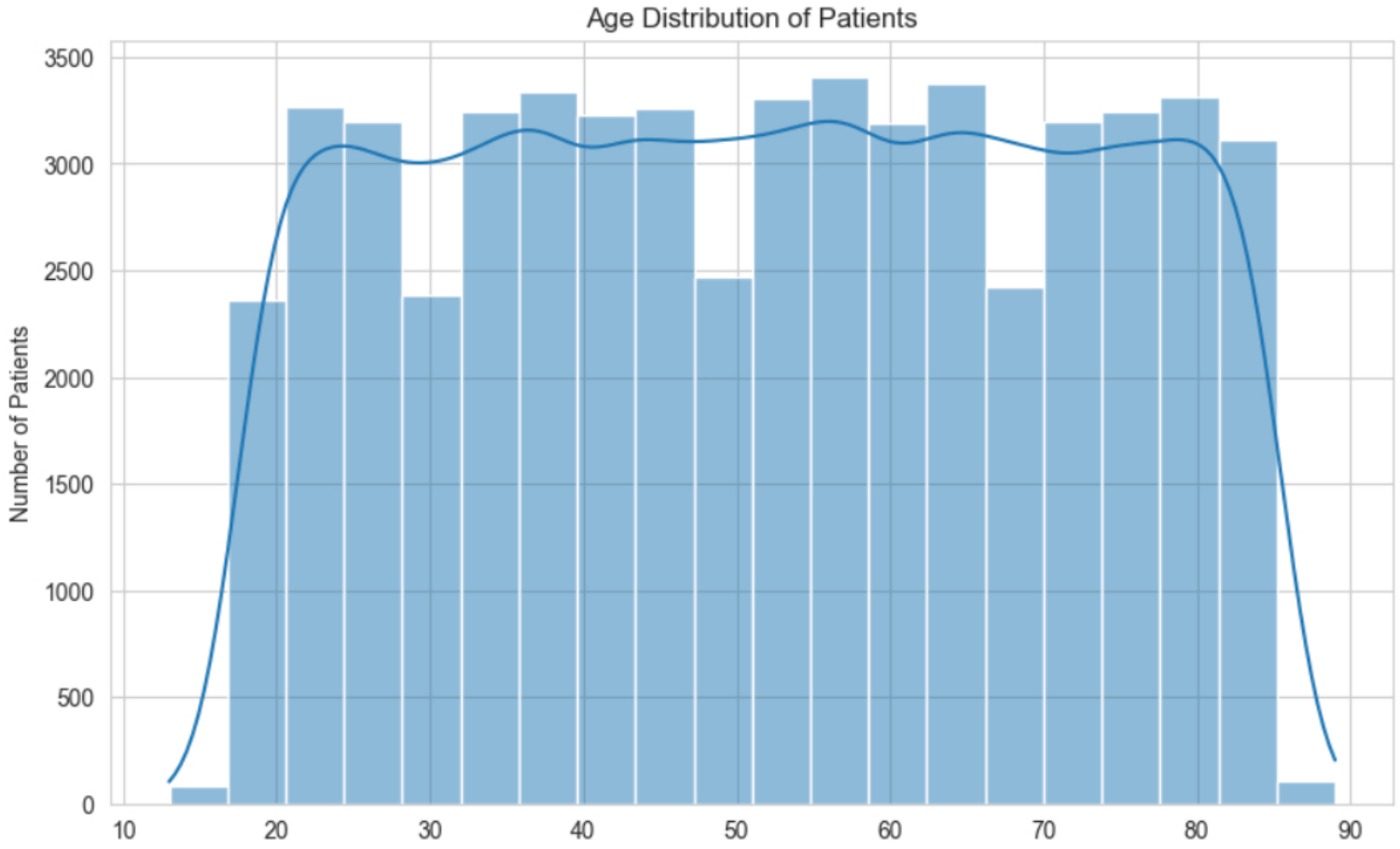
تحلیل توزیع جنسیتی بیماران:



یافته کلیدی: در این مجموعه داده، تعداد بیماران زن (Female) اندکی بیشتر از بیماران مرد (Male) است.

اهمیت: این توزیع جنسیتی نسبتاً متعادل است و به ما اجازه می‌دهد تا در تحلیل‌های بعدی، تأثیر جنسیت بر سایر متغیرها را بدون بایاس (سوگیری) شدید بررسی کنیم.

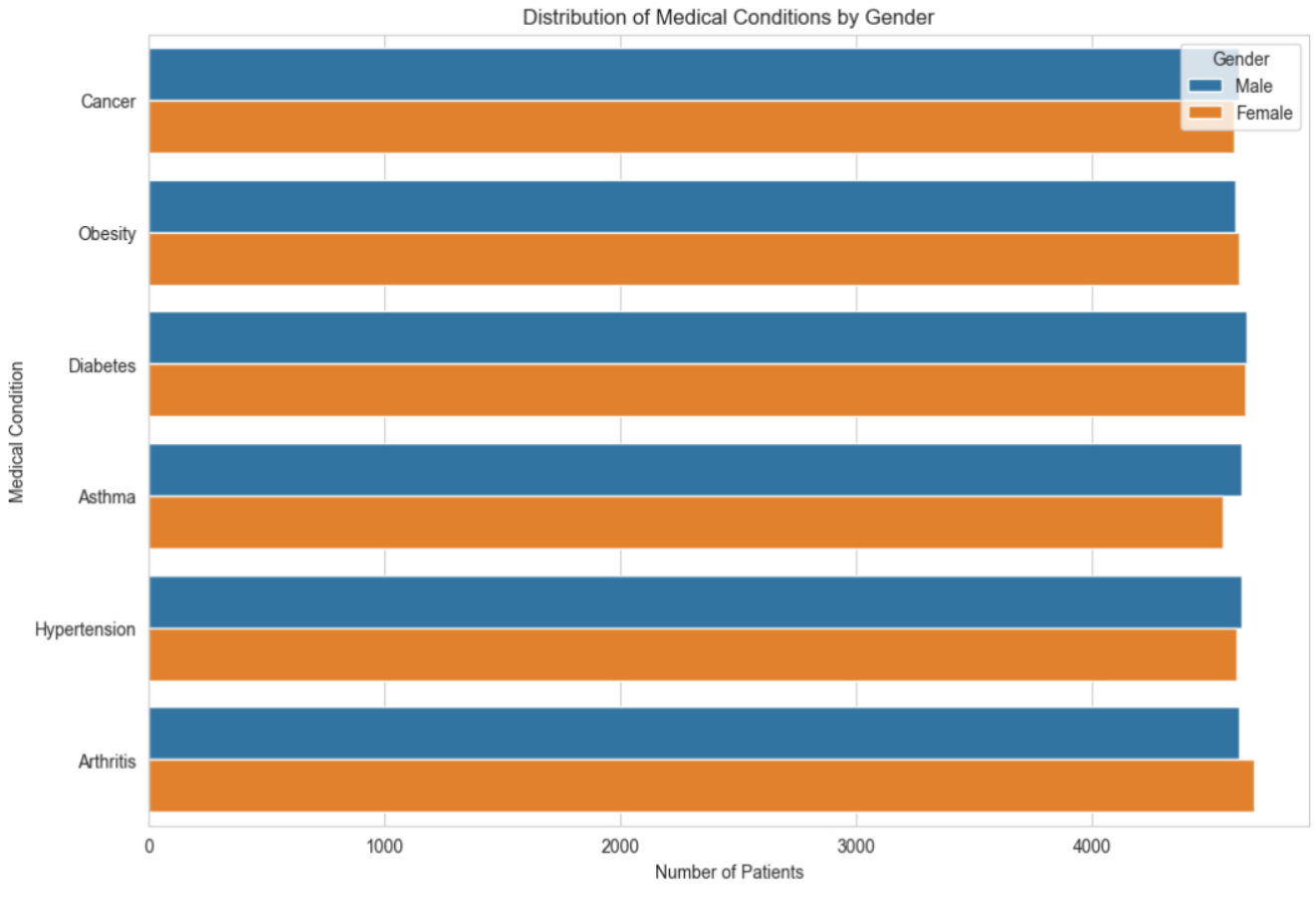
تحلیل توزیع سنی بیماران:



یافته کلیدی: توزیع سنی بیماران گسترده است، اما تراکم اصلی جمعیت بیماران در بازه سنی میانسالی تا کهنسالی قرار دارد. تعداد بیماران جوان و کودک کمتر است.

اهمیت: این الگو منطقی به نظر می‌رسد، زیرا شرایط پزشکی ذکر شده در دیتاست (مانند سرطان، دیابت، آرتروز) بیشتر در سنین بالا شایع هستند. این نمودار تأیید می‌کند که مجموعه داده ما نمایانگر جمعیت واقعی بیماران برای این نوع بیماری‌هاست

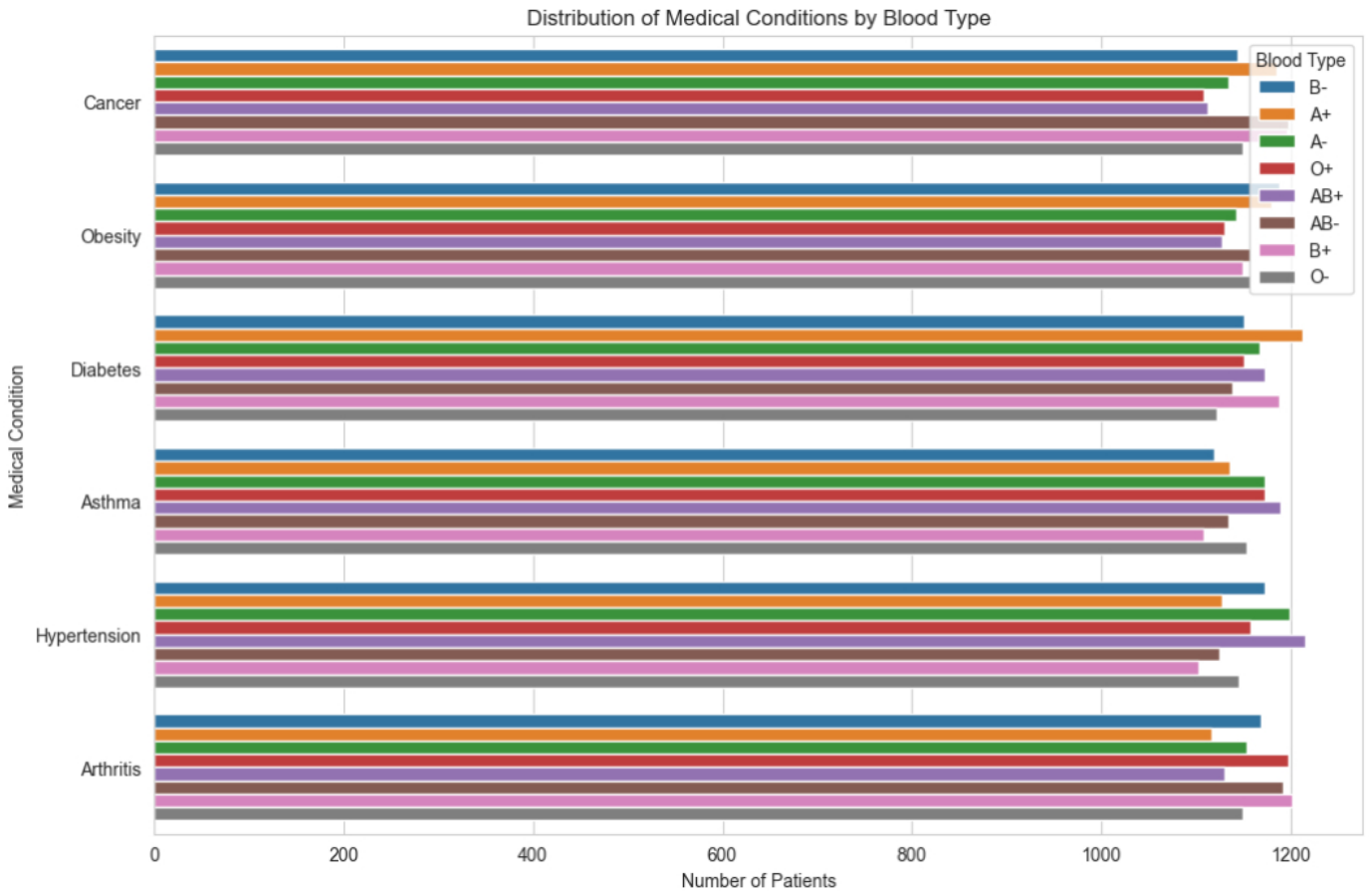
تحلیل شرایط پزشکی بر اساس جنسیت:



یافته کلیدی: این نمودار توزیع هر بیماری را به تفکیک جنسیت نشان می‌دهد. مشاهده می‌شود که برای تمام شرایط پزشکی موجود در این دیتاست، تفاوت چشمگیری بین تعداد بیماران زن و مرد وجود ندارد.

اهمیت: این یافته نشان می‌دهد که در جمعیت مورد مطالعه ما، هیچ‌کدام از این بیماری‌ها به طور خاص در یک جنسیت شایع‌تر از دیگری نیست و تقریباً توزیع برابری دارند.

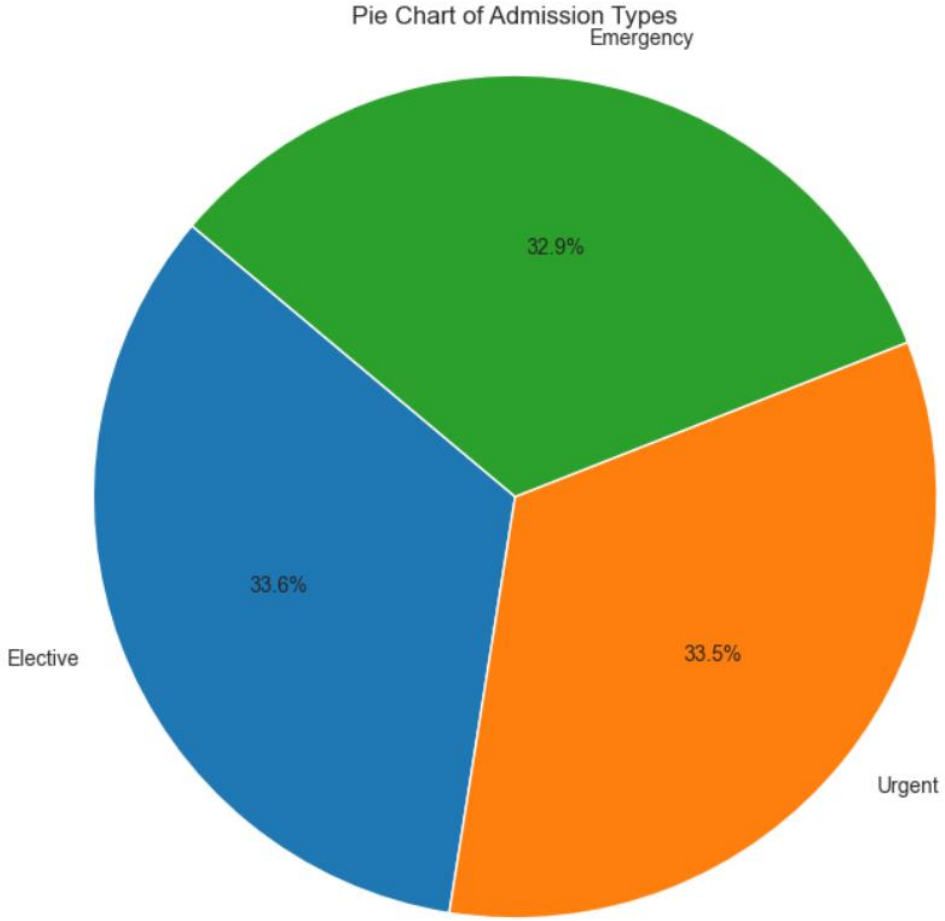
تحلیل شرایط پزشکی بر اساس گروه خونی:



یافته کلیدی: با بررسی توزیع بیماری‌ها در میان گروه‌های خونی مختلف، الگوی خاص یا ارتباط معناداری مشاهده نمی‌شود. هر بیماری تقریباً به صورت یکنواخت در تمام گروه‌های خونی پخش شده است.

اهمیت: این تحلیل نشان می‌دهد که در چارچوب این مجموعه داده، به نظر نمی‌رسد گروه خونی عامل تعیین‌کننده‌ای برای ابتلا به این بیماری‌های خاص باشد.

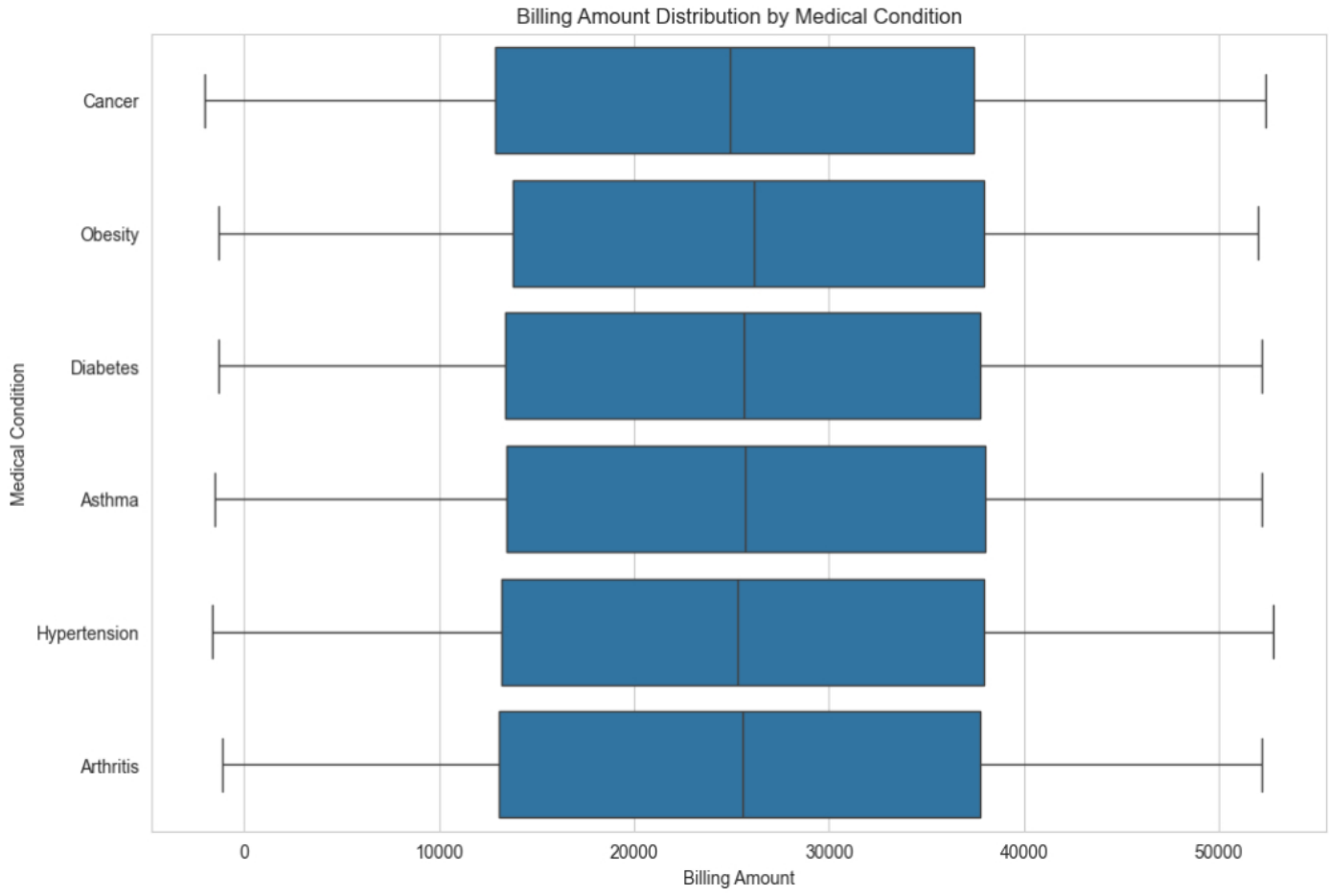
تحلیل توزیع نوع پذیرش بیماران:



یافته کلیدی: این نمودار دایره‌ای نشان می‌دهد که انواع پذیرش (اورژانسی، فوری و اختیاری) تقریباً به نسبت‌های مساوی در میان بیماران تقسیم شده‌اند و هیچ نوع پذیرشی به طور چشمگیر بر دیگری برتری ندارد.

اهمیت: این توزیع متوازن می‌تواند نشان‌دهنده عملکرد یک مرکز درمانی جامع باشد که هم به موارد اورژانسی و هم به جراحی‌ها یا درمان‌های برنامه‌ریزی شده (اختیاری) به یک اندازه خدمات ارائه می‌دهد.

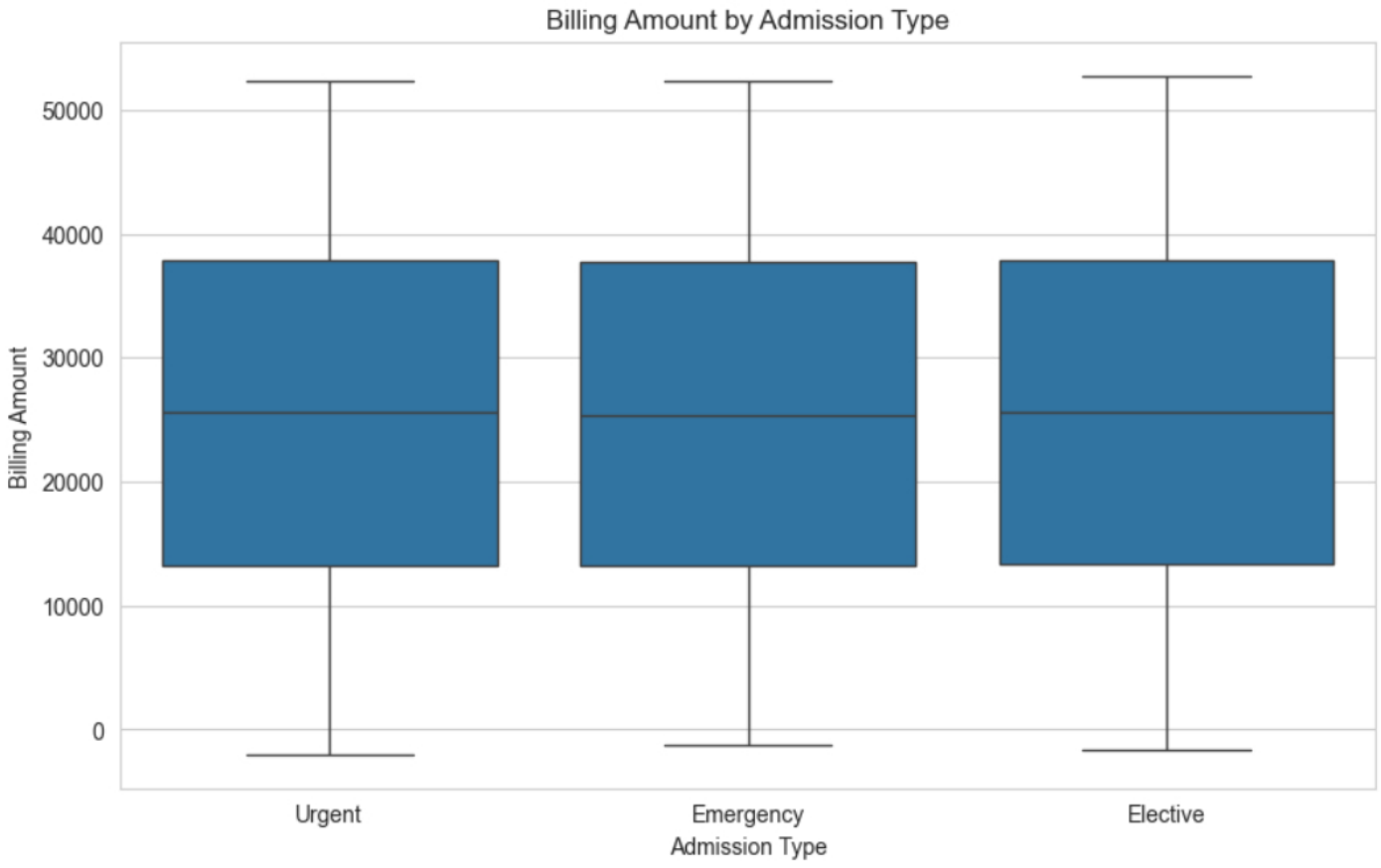
تحلیل هزینه درمانی بر اساس نوع بیماری:



یافته کلیدی: این نمودار نشان می‌دهد که میانگین و پراکندگی هزینه‌های درمانی بین بیماری‌های مختلف، متفاوت است. به طور مشخص، درمان سرطان (Cancer) به طور میانگین بالاترین هزینه و بیشترین پراکندگی را دارد، در حالی که بیماری‌هایی مانند آسم (Asthma) هزینه‌های کمتر و قابل پیش‌بینی‌تری دارند.

اهمیت: این تحلیل برای مدیریت مالی بیمارستان بسیار مهم است، زیرا به شناسایی پرهزینه‌ترین درمان‌ها کمک کرده و می‌تواند مبنایی برای تخصیص منابع و برنامه‌ریزی بودجه باشد.

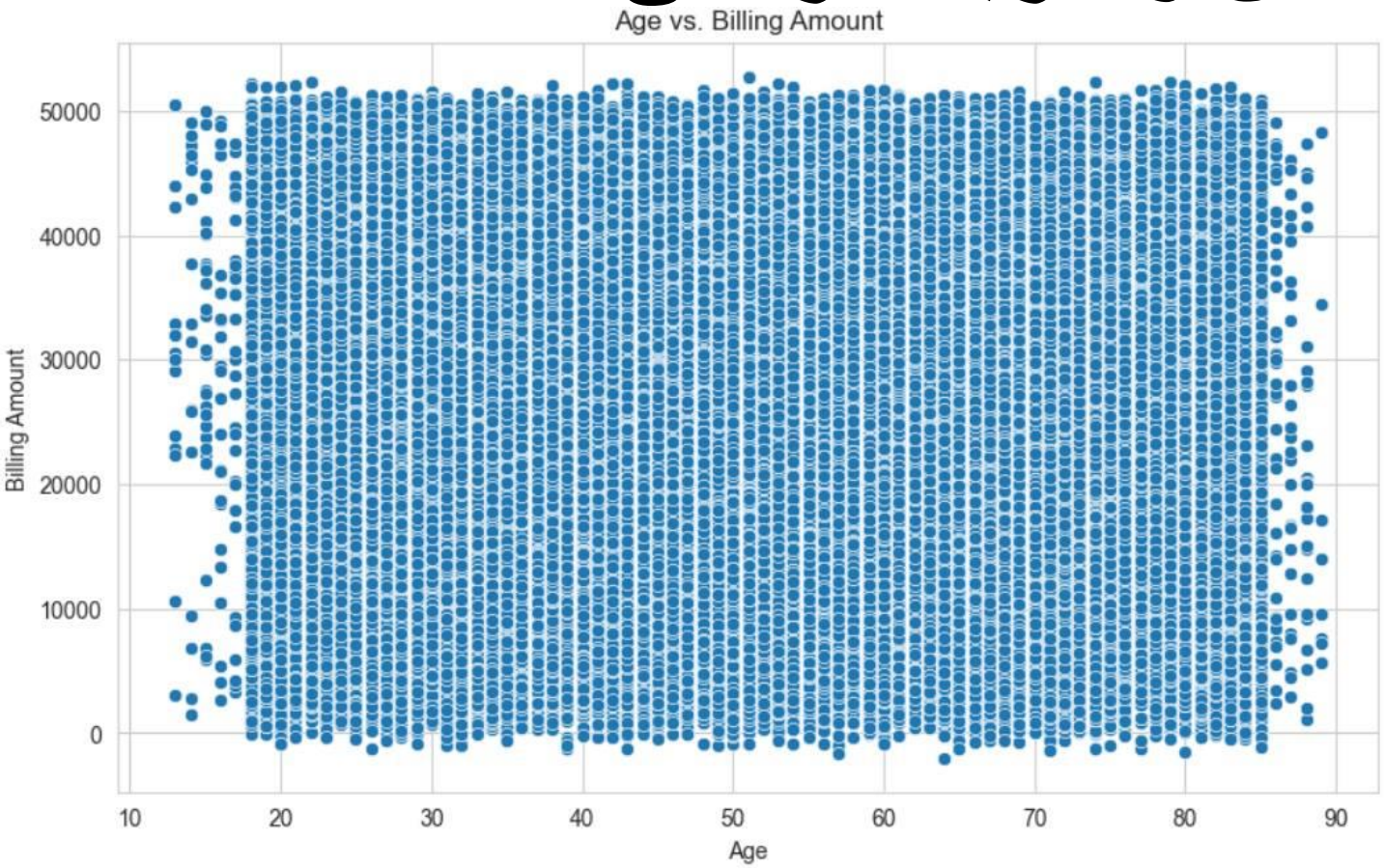
تحلیل هزینه درمانی بر اساس نوع پذیرش:



یافته کلیدی: مقایسه هزینه‌ها بر اساس نوع پذیرش نشان می‌دهد که پذیرش‌های اورژانسی (Emergency) و (Urgent) به طور میانگین هزینه‌های بالاتری نسبت به پذیرش‌های فوری (Elective) اختیاری دارند.

اهمیت: این یافته منطقی است، زیرا موارد اورژانسی اغلب نیازمند اقدامات پزشکی پیچیده‌تر، تجهیزات خاص و مراقبت‌های ویژه هستند. این تحلیل می‌تواند به بهینه‌سازی هزینه‌ها در بخش اورژانس کمک کند.

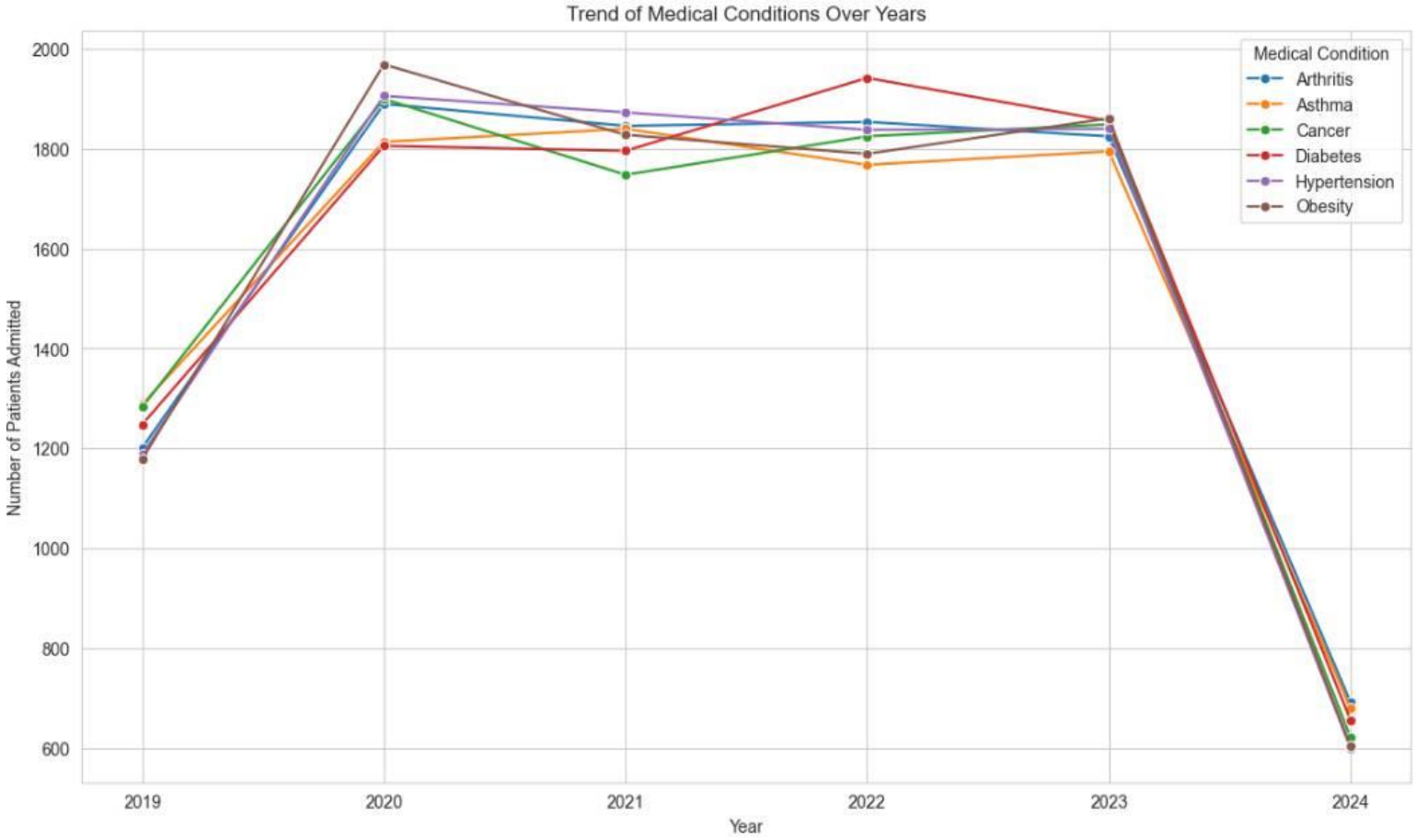
تحلیل رابطه سن و هزینه درمانی :



یافته کلیدی: این نمودار هر بیمار را به صورت یک نقطه نشان می‌دهد. با اینکه نقاط پراکندگی زیادی دارند، اما می‌توان یک روند مثبت ضعیف را مشاهده کرد. این یعنی به طور کلی، با افزایش سن، هزینه‌های درمانی نیز تمایل به افزایش دارند.

اهمیت: این رابطه به شرکت‌های بیمه و مدیران بیمارستان در پیش‌بینی هزینه‌ها و مدیریت ریسک کمک می‌کند. این تحلیل تأیید می‌کند که بیماران مسن‌تر ممکن است به مراقبت‌های پرهزینه‌تری نیاز داشته باشند.

تحلیل روند زمانی بیماری‌ها:



یافته کلیدی: این نمودار خطی، تعداد بیماران پذیرش شده برای هر بیماری را در طول سال‌های مختلف نشان می‌دهد. مشاهده می‌شود که روند پذیرش برای تمام بیماری‌ها تقریباً ثابت و پایدار بوده و افزایش یا کاهش چشمگیری در طول زمان نداشته است.

اهمیت: این پایداری به مدیران بیمارستان اجازه می‌دهد تا با اطمینان بیشتری برای آینده برنامه‌ریزی کنند، زیرا می‌توانند انتظار داشته باشند که بار کاری مربوط به این بیماری‌ها در سال‌های آتی نیز مشابه باقی بماند.

نتیجه‌گیری نهایی:

- توازن داده‌ها: مجموعه داده مورد بررسی از نظر توزیع بیماری‌ها و جنسیت، متوازن بود که اعتبار تحلیل‌های مقایسه‌ای را افزایش داد.
- جمعیت‌شناسی بیماران: تمرکز اصلی جمعیت بیماران بر روی افراد میانسال و مسن بود که با ماهیت بیماری‌های مزمن (مانند دیابت و سرطان) همخوانی دارد.
- عوامل مؤثر بر هزینه: تحلیل هزینه‌ها نشان داد که نوع بیماری (مانند سرطان) و نوع پذیرش (مانند اورژانسی) دو عامل کلیدی در افزایش هزینه‌های درمانی هستند.
- روند زمانی پایدار: روند پذیرش بیماران برای بیماری‌های مختلف در طول زمان ثابت بوده و افزایش یا کاهش چشمگیری نداشته است