



دانشکده ملی مهارت

شهرستان میناب

تمرین های درس مباحث ویژه(بخش پنجم)

مرتبط به درس:

مباحث ویژه

گردآورنده:

محمدحسین خبیری

استاد:

محمد احمدزاده

ترم تحصیلی بهمن ۱۴۰۳-۱۴۰۴

1. Line Chart چرا برای نمایش رابطه‌های خطی استفاده می‌شود؟

نمودار خطی (**Line Chart**) برای نمایش رابطه‌های خطی استفاده می‌شود زیرا:

1. **نمایش روندها**: این نمودار تغییرات داده‌ها را در طول زمان یا در یک توالی مشخص نشان می‌دهد. برای مثال، افزایش یا کاهش قیمت سهام در یک دوره زمانی.
2. **اتصال نقاط داده**: در نمودار خطی، نقاط داده به وسیله‌ی خطوط به هم متصل می‌شوند، که این کار کمک می‌کند تا رابطه بین متغیرها به شکل بصری قابل درک باشد.
3. **تشخیص الگوها**: تغییرات تدریجی، نوسانات و الگوهای خاص (مثل رشد، کاهش، یا ثبات) در داده‌ها به‌راحتی قابل مشاهده هستند.
4. **سادگی و خوانایی**: نمودارهای خطی ساختار ساده‌ای دارند و می‌توانند روندهای خطی یا غیرخطی را به شکل مؤثری نمایش دهند.
5. **مقایسه چندین مجموعه داده**: می‌توان چندین خط را روی یک نمودار نمایش داد تا مقایسه‌ای بین داده‌های مختلف انجام شود.

به همین دلایل، نمودار خطی ابزاری کارآمد برای نمایش و تحلیل روندها و روابط خطی بین داده‌ها محسوب می‌شود.

1. Bar Chart چرا برای مقایسه داده‌های گروهی کاربرد دارد؟

نمودار میله‌ای (Bar Chart) برای مقایسه داده‌های گروهی کاربرد دارد زیرا:

1. وضوح در مقایسه مقادیر: هر دسته داده با یک میله مشخص نمایش داده می‌شود و ارتفاع (یا طول) میله‌ها نشان‌دهنده مقدار داده است، که مقایسه بین گروه‌ها را آسان می‌کند.
2. امکان نمایش چندین گروه: می‌توان چندین میله را در کنار هم قرار داد تا مقایسه‌ای مستقیم بین چند گروه یا دسته مختلف انجام شود.
3. قابل‌فهم بودن: ساختار ساده‌ای دارد و افراد می‌توانند به‌راحتی تفاوت بین مقادیر را مشاهده و تحلیل کنند.
4. انعطاف‌پذیری بالا: می‌توان آن را به دو صورت افقی و عمودی رسم کرد که بسته به نوع داده، خوانایی و تحلیل را بهبود می‌بخشد.
5. مناسب برای داده‌های گسسته: داده‌هایی که در دسته‌های مشخص (مثل فروش محصولات مختلف، تعداد دانش‌آموزان در کلاس‌های مختلف، یا درآمد سالانه شرکت‌ها) قرار می‌گیرند، به‌خوبی در این نمودار نمایش داده می‌شوند.

به همین دلیل، نمودار میله‌ای یکی از بهترین گزینه‌ها برای مقایسه داده‌های گروه‌بندی‌شده است.

1. Scatter Plot چرا برای نمایش رابطه‌های غیرخطی استفاده می‌شود؟

نمودار پراکندگی (Scatter Plot) برای نمایش رابطه‌های غیرخطی استفاده می‌شود زیرا:

1. نمایش توزیع داده‌ها: این نمودار نقاط داده را بر اساس دو متغیر روی محورهای افقی و عمودی قرار می‌دهد، که به وضوح الگوی توزیع داده‌ها را نشان می‌دهد.
2. آشکارسازی روابط پیچیده: برخلاف نمودار خطی که بیشتر برای داده‌های دارای الگوی خطی مناسب است، نمودار پراکندگی می‌تواند روابط غیرخطی (مانند منحنی‌ها، همبستگی‌های درجه دوم یا سوم) را نشان دهد.
3. تشخیص الگوها و روندها: اگر بین متغیرها همبستگی وجود داشته باشد، نقاط به شکلی خاص (مثلاً منحنی یا خوشه‌ای) توزیع می‌شوند، که نشان‌دهنده یک رابطه غیرخطی است.
4. شناسایی نقاط پرت: این نمودار به راحتی داده‌های نامتعارف (نقاط پرت) را نشان می‌دهد که ممکن است در سایر نمودارها کمتر قابل تشخیص باشند.
5. تحلیل دقیق وابستگی متغیرها: در بسیاری از موارد، داده‌های دنیای واقعی دارای روابط پیچیده هستند که به کمک نمودار پراکندگی می‌توان آن‌ها را بهتر تحلیل کرد.

به همین دلیل، Scatter Plot یک ابزار مؤثر برای بررسی روابط غیرخطی میان متغیرها و تحلیل روندهای پیچیده در داده‌ها است.

1. Bubble Chart چرا برای نمایش سه متغیر استفاده می‌شود؟

نمودار حبابی (Bubble Chart) برای نمایش سه متغیر استفاده می‌شود زیرا:

1. نمایش هم‌زمان سه بعد از داده:

محور X یک متغیر را نشان می‌دهد.

محور Y متغیر دوم را نشان می‌دهد.

اندازه حباب (قطر یا مساحت) متغیر سوم را نمایش می‌دهد.

1. قابلیت مقایسه بین چندین داده: با تغییر اندازه‌ی حباب‌ها، می‌توان میزان اهمیت یا وزن متغیر سوم را در مقایسه با سایر داده‌ها مشاهده کرد.
2. قابل درک بودن: شباهت زیادی به نمودار پراکندگی دارد، اما با افزودن یک بُعد دیگر (اندازه‌ی حباب)، اطلاعات بیشتری را به صورت بصری نمایش می‌دهد.
3. تحلیل روابط پیچیده‌تر: نمودار حبابی امکان بررسی هم‌زمان ارتباط بین سه متغیر را فراهم می‌کند، که در تحلیل‌های اقتصادی، جمعیتی، و علمی بسیار مفید است.
4. جذابیت بصری: از آنجا که حباب‌ها اندازه‌های مختلفی دارند، نمودار از نظر بصری جذاب‌تر و خواناتر از برخی نمودارهای دیگر برای نمایش داده‌های چندبعدی است.
5. Heatmap چرا برای نمایش رابطه‌های بین متغیرها کاربرد دارد؟

نقشه حرارتی (Heatmap) برای نمایش رابطه‌های بین متغیرها کاربرد دارد زیرا:

1. کدگذاری رنگی برای نمایش داده‌ها: در Heatmap، مقادیر مختلف با رنگ‌های متفاوت نمایش داده می‌شوند، که باعث می‌شود الگوها و روندهای کلی در داده‌ها به‌سرعت قابل مشاهده باشند.
2. امکان نمایش حجم زیادی از داده‌ها: این نمودار برای نمایش ماتریس‌های داده‌ای بزرگ مناسب است، زیرا رنگ‌ها به جای اعداد خام به درک سریع‌تر اطلاعات کمک می‌کنند.
3. شناسایی روابط و همبستگی‌ها: از این نمودار می‌توان برای نمایش میزان همبستگی بین متغیرها استفاده کرد، به‌ویژه در تحلیل داده‌های آماری و یادگیری ماشین.
4. برجسته‌سازی نقاط پرت یا الگوهای خاص: تفاوت‌های شدید در داده‌ها از طریق تغییر رنگ مشخص می‌شوند و نقاط پرت یا بخش‌های مهم داده‌ها را نمایان می‌کنند.
5. بهینه‌سازی فضای نمایش: در مقایسه با نمودارهای عددی یا متنی، نقشه حرارتی اطلاعات را به شکلی فشرده و خوانا ارائه می‌دهد.
6. Pairplot چرا برای تحلیل روابط بین متغیرها کاربرد دارد؟

Pairplot برای تحلیل روابط بین متغیرها کاربرد دارد زیرا:

1. نمایش همه‌ی روابط دومتغیره در یک نگاه: این نمودار مجموعه‌ای از Scatter Plot‌ها (نمودارهای پراکندگی) را برای تمام جفت‌های ممکن از متغیرها رسم می‌کند و امکان بررسی روابط بین آن‌ها را فراهم می‌آورد.
2. شناسایی همبستگی‌ها و الگوها: با مشاهده نحوه توزیع نقاط در نمودارهای جفتی، می‌توان روابط خطی، غیرخطی یا عدم ارتباط بین متغیرها را تشخیص داد.
3. نمایش توزیع داده‌ها: قطر اصلی Pairplot معمولاً Histogram یا KDE Plot (چگالی احتمال) را نشان می‌دهد که توزیع هر متغیر را به‌صورت جداگانه بررسی می‌کند.
4. مقایسه چندین متغیر به‌طور هم‌زمان: به‌جای تحلیل جداگانه هر نمودار پراکندگی، Pairplot امکان مقایسه همه متغیرها را در یک مجموعه نمودار واحد فراهم می‌کند.
5. شناسایی نقاط پرت (Outliers): نقاط پرت که ممکن است روی یک متغیر تأثیر بگذارند، در Pairplot به‌راحتی قابل مشاهده هستند.
6. چرا Boxplot برای تشخیص Outliers استفاده می‌شود؟

Boxplot (نمودار جعبه‌ای) برای تشخیص نقاط پرت (Outliers) استفاده می‌شود زیرا:

1. نمایش خلاصه‌ای از توزیع داده‌ها:

میانه (Median): خط وسط جعبه، مقدار میانه داده‌ها را نشان می‌دهد.

چارک اول (Q1) و چارک سوم (Q3): حد پایین و بالای جعبه، محدوده ۲۵٪ تا ۷۵٪ داده‌ها را نشان می‌دهد.

دامنه بین چارکی (IQR): فاصله بین Q1 و Q3 که نشان‌دهنده توزیع مرکزی داده‌ها است.

1. تشخیص نقاط پرت با استفاده از IQR:

مقدارهای پرت خارج از محدوده‌ی (Q1 - 1.5×IQR) تا (Q3 + 1.5×IQR) به عنوان Outlier در نظر گرفته می‌شوند.

این نقاط معمولاً در خارج از خطوط افقی (Whiskers) و به‌صورت نقاط جدا از جعبه نمایش داده می‌شوند.

مقایسه بین گروه‌های مختلف: Boxplot امکان مقایسه بین چند مجموعه داده را فراهم می‌کند و مشخص می‌کند که کدام گروه دارای نقاط پرت یا پراکندگی بیشتر است.

1. نمایش چولگی (Skewness) داده‌ها:

اگر جعبه به سمت چپ یا راست کشیده شده باشد، نشان‌دهنده توزیع چوله است.

این ویژگی به شناسایی عدم تعادل در داده‌ها کمک می‌کند.

سادگی و خوانایی بالا: به دلیل نمایش گرافیکی فشرده، Boxplot یک روش سریع و کارآمد برای تشخیص نقاط پرت و تحلیل توزیع داده‌ها است.

1. Histogram چرا برای نمایش توزیع داده‌ها کاربرد دارد؟

Histogram برای نمایش توزیع داده‌ها کاربرد دارد زیرا:

1. تقسیم داده‌ها به بازه‌های مشخص (Bins):
   * هیستوگرام داده‌ها را به بازه‌های عددی (Bins) تقسیم کرده و تعداد داده‌های موجود در هر بازه را نمایش می‌دهد.
   * این ویژگی کمک می‌کند تا بفهمیم که داده‌ها بیشتر در کدام محدوده‌ها متمرکز شده‌اند.
2. نمایش شکل کلی توزیع داده‌ها:
   * آیا داده‌ها نرمال (Normal Distribution) هستند؟
   * آیا چولگی (Skewness) دارند؟ (مثلاً چوله به چپ یا راست)
   * آیا دارای چندین قله (Multimodal Distribution) هستند؟
3. مقایسه چگالی داده‌ها در بازه‌های مختلف: با مشاهده ارتفاع ستون‌ها، می‌توان دید که داده‌ها در کدام بخش‌ها متراکم‌تر یا پراکنده‌تر هستند.
4. تحلیل رفتار داده‌ها و تصمیم‌گیری بهتر:
   * در آمار و علم داده، از هیستوگرام برای درک بهتر ویژگی‌های داده‌ها قبل از مدل‌سازی استفاده می‌شود.
   * برای مثال، در تحلیل نمرات دانش‌آموزان، هیستوگرام نشان می‌دهد که بیشتر دانش‌آموزان چه نمره‌ای گرفته‌اند.
5. سادگی و خوانایی بالا:
   * برخلاف جداول عددی، هیستوگرام به‌صورت بصری نمایش می‌دهد که چگونه داده‌ها توزیع شده‌اند.
   * به‌راحتی می‌توان نقاط پرت (Outliers) یا داده‌های نادر را شناسایی کرد.
6. چگونه می‌توانید یک 3D Plot را در Python ایجاد کنید؟



1. چرا Seaborn برای تجسم داده‌های پیشرفته استفاده می‌شود؟

Seaborn برای تجسم داده‌های پیشرفته استفاده می‌شود زیرا:

1. سادگی و زیبایی بصری:  
   Seaborn نمودارهایی با طراحی زیبا و پیش‌فرض‌های مناسب ارائه می‌دهد که به راحتی می‌توان از آن‌ها برای تحلیل داده‌ها استفاده کرد.
2. پشتیبانی از انواع مختلف نمودارها:  
   Seaborn از انواع نمودارهای پیشرفته مانند Boxplot، Violin Plot، Pairplot، Heatmap، FacetGrid و بسیاری دیگر پشتیبانی می‌کند که برای تحلیل‌های پیچیده و تجزیه و تحلیل چندمتغیری مناسب است.
3. کار با داده‌های پیچیده:  
   Seaborn به‌خوبی با Pandas DataFrame هماهنگ است و می‌تواند داده‌های پیچیده و گروه‌بندی شده را به‌راحتی تجزیه و تحلیل کند.
4. قابلیت سفارشی‌سازی آسان:  
   Seaborn این امکان را فراهم می‌کند که تنظیمات مختلف مانند رنگ، شکل و سبک نمودارها را به راحتی انجام دهید.
5. عملکرد بالا در تجزیه و تحلیل داده‌های آماری:  
   Seaborn امکان تجزیه و تحلیل آماری مانند رسم توزیع داده‌ها، همبستگی‌ها، روندها و داده‌های گروه‌بندی شده را فراهم می‌کند.