

په نام خدا

فصل چهارم

ابزارها و تکنیک‌های دیداری سازی داده

مقدمه

در این فصل، به موضوعات زیر خواهیم پرداخت:

مقدمه

انواع مختلف تصویرسازی

ابزارها و تکنیک‌های محبوب تصویرسازی داده

ابزارها و تکنیک‌های دیگر در دسترس

مورد استفاده از ابزارهای رایج

هدف اصلی این فصل، آشنایی شما با روش‌های مختلف تصویرسازی داده است، مانند نمودار میله‌ای، نقشه درختی، نمودار دایره‌ای، و نمودار دونات به عنوان نمونه‌ها. همچنین، شما در این فصل به مورد استفاده واقعی تکنیک‌های تصویرسازی داده در پلتفرم‌های (ابزارهای) معروفی مانند فیسبوک، گوگل و توییتر خواهید پرداخت.

تصویر سازی داده ها چیست؟

تصویر سازی داده، فرآیندی است که در آن اطلاعات را در یک چارچوب تصویری مانند نقشه یا نمودار قرار می‌دهد تا برای مغز انسان بهتر قابل فهم باشد و از داده‌ها استنباط‌هایی بکشد. در طول ارائه داده، این گرافیک‌ها تصویری واضح را ایجاد کرده و درک مخاطب را بهبود می‌بخشند. هدف اصلی آن، شناسایی نقاط پرت، روندها و الگوها در مجموعه‌های داده بزرگ است. علاوه بر این، تصویر سازی داده مطالعه‌ای است برای اعطای معنا به داده‌ها با قرار دادن آنها در یک چارچوب تصویری تا الگوها، روندها و همبستگی‌هایی که در غیر این صورت قابل مشاهده نباشند، نمایش داده شوند

کتابخانه های مهم برای رسم نمودار در پایتون :

Matplotlib : سطح پایین، امکانات بسیار زیادی را فراهم می کند

Pandas Visualization : رابط کاربری آسان بر پایه Matplotlib

Seaborn : سطح بالا، استایل های پیش فرض عالی

Plotline : بر اساس ggplot2 در R، از دستور زبان گرافیکی استفاده می کند

Plotly : قادر به ایجاد نمودارهای تعاملی است

انواع مختلف دیداری سازی :

Bar chart (نمودار میله ای)

یکی از محبوب‌ترین تجسم‌سازی‌های داده در لیست، نمودار میله‌ای یا نمودار ستونی است. از نمودارهای میله‌ای برای مقایسه داده‌ها در دو محور استفاده می‌شود. یکی از محورها محور عددی است و محور بصری دسته‌ای یا موضوعاتی است که قرار است اندازه‌گیری شوند. یک نمودار میله‌ای می‌تواند میله‌هایی افقی یا عمودی داشته باشد. مقادیر عددی در محور (Y محور عمودی) در نمودارهای میله‌ای عمودی و در محور (X محور افقی) در نمودارهای میله‌ای افقی ظاهر می‌شوند. اگر نام‌های داده‌های کیفی شما طولانی و تفصیلی هستند، استفاده از یک نمودار میله‌ای افقی بهترین گزینه است

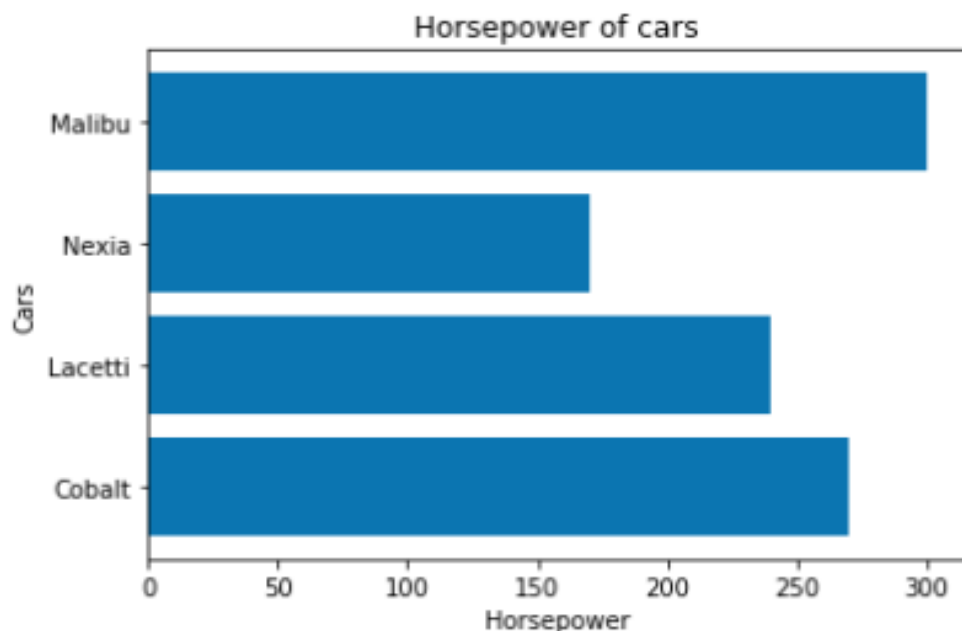


Figure 4.1: Bar Chart

Pie chart (نمودار دایره ای)

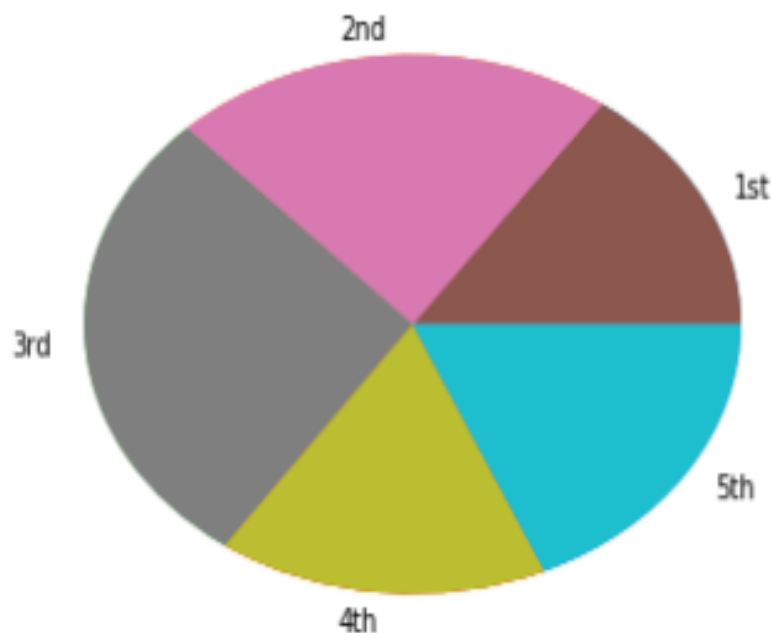


Figure 4.2: Pie Chart

دومین نوع محبوب دیداری سازی داده در این لیست نمودار پای است. داده‌ها در نمودار پای بخش‌های یک به کل را نشان می‌دهند. کل دایره نمایانگر کل است و هر قسمت نمایانگر یک بخش داده است.

نوع مناسب داده برای نمودار پای حداکثر پنج یا شش بخش است. بیشتر از این تعداد باعث می‌شود قطاع‌ها در مرکز نمودار بسیار نازک شوند. اگر بیش از سه مقدار به یکدیگر شبیه باشند، تشخیص بین آن‌ها دشوار خواهد بود.

Donut chart (نمودار دونات)

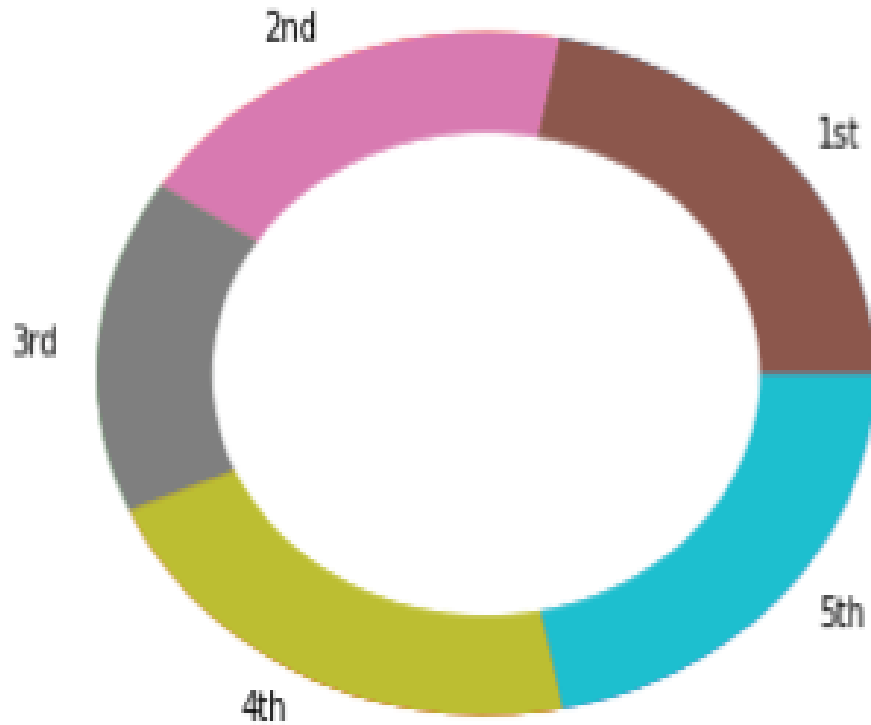


Figure 4.3: Donut Chart

یک نمودار دونات بسیار شبیه به یک نمودار دایره ای است، با این تفاوت که قسمت مرکزی آن حذف شده است. تفاوت اساسی بین آن‌ها اساساً در بُعد بصری است. می‌توانید در یک نمودار دونات بیشتر از تعداد بخش‌های یک نمودار دایره ای داشته باشید و باز هم قابل فهم خواهد بود

Half donut chart and gauge chart (نمودار نصف دونات و نمودار گیج (پیمانه))

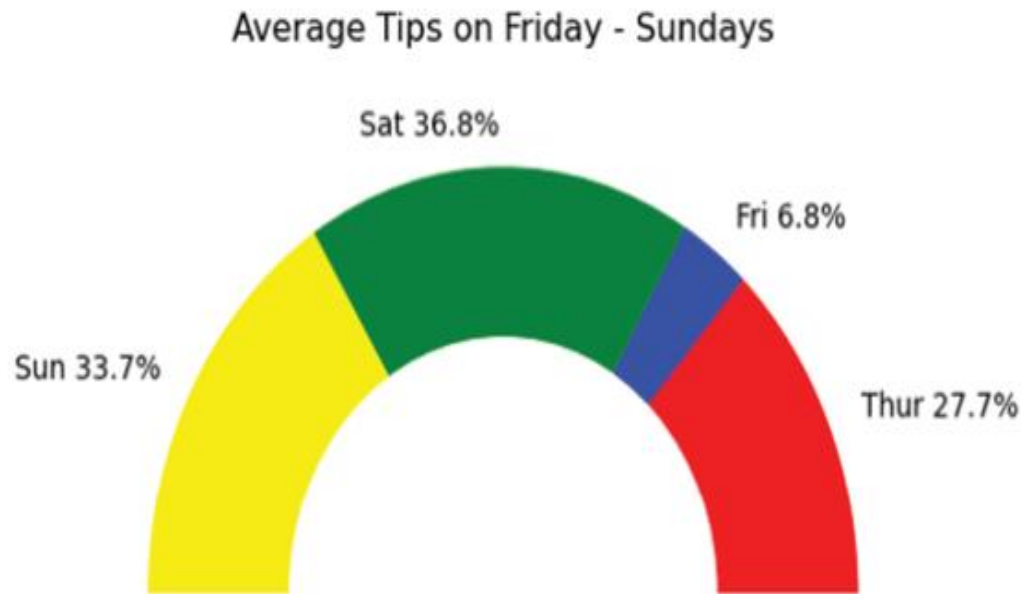


Figure 4.4: Half Donut Chart

نمودار نیمه دونات دقیقاً همان چیزی است که اسمش را نشان می‌دهد، نیمی از یک نمودار دونات است. این نوع دیداری سازی داده مناسب است وقتی نیاز به نمایش مجموعه‌های کوچک داده دارید. از این رو، در نمودار نیمه دونات بهتر است حداکثر سه قطاع استفاده شود تا روشنایی و خوانایی بهتری داشته باشد.

Nested pie chart (نمودار دایره ای تو در تو)

از نمودارهای دایره ای و دونات به صورت ترکیبی استفاده می شود تا یک نمودار دایره ای تو در تو ایجاد شود. این نمودارها برای نمایش داده های پیچیده بسیار مناسب هستند.

در زیر می توانید یک نمودار دایره ای تو در تو را مشاهده کنید که درصد دانش آموزان مرد و زن را در سه بخش کلاس دوازدهم نشان می دهد. نمودار دونات بیرونی، دسته بندی سطح بالا را نمایش می دهد. بخش های کلاس دوازدهم به صورت زیر مشخص شده است: A- پزشکی، B- علوم تجربی، C- هنر. در لایه دوم، بخش های توصیفی در مورد درصد زنان و مردان که در هر دسته اصلی قرار می گیرند، قرار دارند.

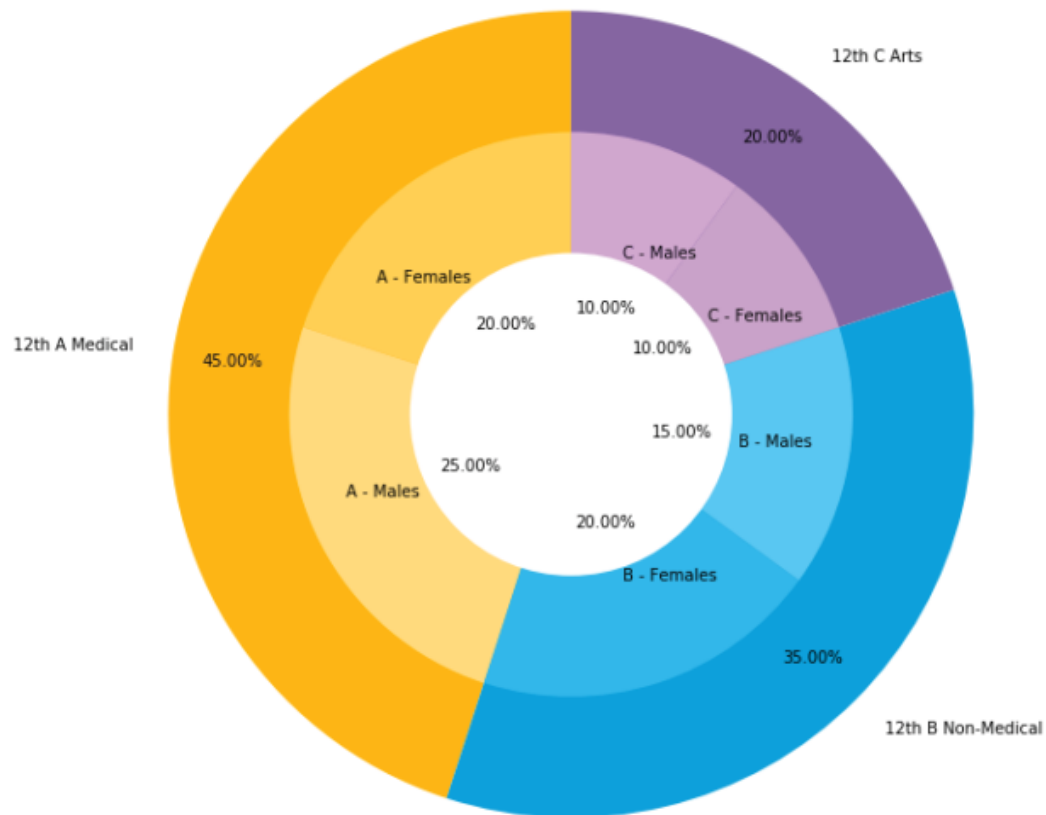


Figure 4.5: Nested Pie Chart

Line chart (نمودار خطی)

نمودار خطی یا نمودار خط، یک نوع دیداری سازی داده است که تغییرات داده را در طول زمان نشان می‌دهد. مانند نمودار میله‌ای، نمودار خطی دارای محور X و Y است. اما تفاوت آن این است که هر دو محور شامل مقادیر عددی نماینده داده هستند.

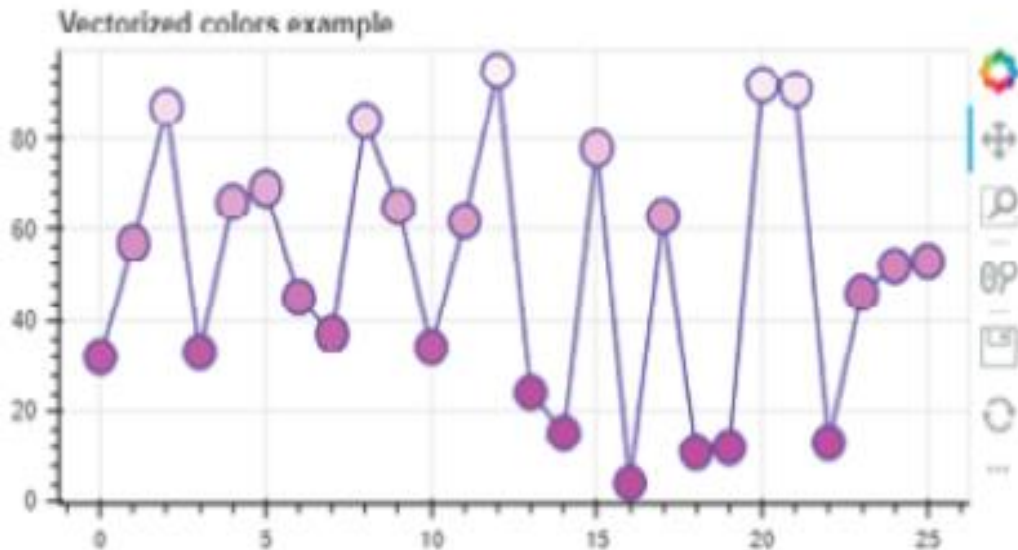


Figure 4.6: Line Chart

برای ایجاد یک نمودار خطی، باید محدوده زمانی مربوطه را در محور X و اندازه‌گیری کمیت را در محور Y وارد کنید. داده‌ها را در نمودار با اتصال مقدار زمان به مقدار عددی بر روی خطوط نمودار نشان دهید. پس از نشان دادن تمام نقاط داده، آن‌ها را با یک خط جدا کنید

Scatter plot (نمودار پراکندگی)

نمودار پراکندگی یک نوع دیداری سازی داده است که برای تحلیل همبستگی بین متغیرها استفاده می‌شود. در نمودار پراکندگی، داده‌ها به صورت نقاط روی نمودار نشان داده می‌شوند و هر نقطه بر اساس دو مقدار متناظرش موقعیت داده می‌شود. وقتی نقاطی خارج از محدوده مورد انتظار قرار می‌گیرند، آنها را نقاط پرت یا نقاط خارج از محدوده می‌نامند و باید در تحلیل داده‌ها مدنظر قرار گیرند. نمودار پراکندگی به کار رفته است که متغیرهای شما به یکدیگر مرتبط هستند و درباره یک گروه از موضوعات تست موجود است.

بعضی از این روابط می‌توانند شامل رابطه بین وزن و قد در کودکان زیر ۱۸ سال، فروش وابسته به دما در یک فروشگاه بستنی، نرخ دیابت و چاقی و غیره باشند.



Figure 4.7: Scatter Plot

Cone chart (نمودار مخروطی)

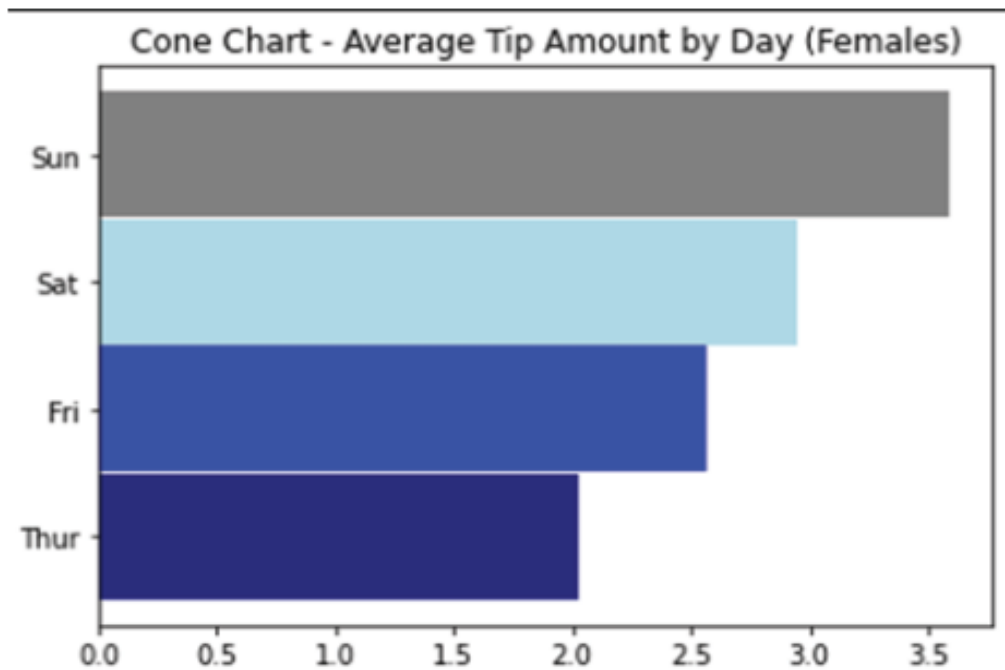


Figure 4.8: Cone Chart

نمودار مخروطی یک نوع دیداری سازی داده است که بخش‌های یک به کل را نشان می‌دهد، مشابه نمودارهای دایره ای. تفاوت آن این است که نمودار مخروطی همچنین سلسله مراتب را نشان می‌دهد. داده با بیشترین مقدار در بالای مخروط قرار می‌گیرد که همچنین بزرگترین منطقه را دارد. سایر مقادیر به ترتیب نزولی به سمت نوک پایینی مخروط جریان دارند.

تعداد مقادیر نمایش داده شده در نمودار مخروطی را محدود کنید و حداکثر به هفت یا هشت مقدار بسنده کنید.

Pyramid chart (نمودار هرمی)

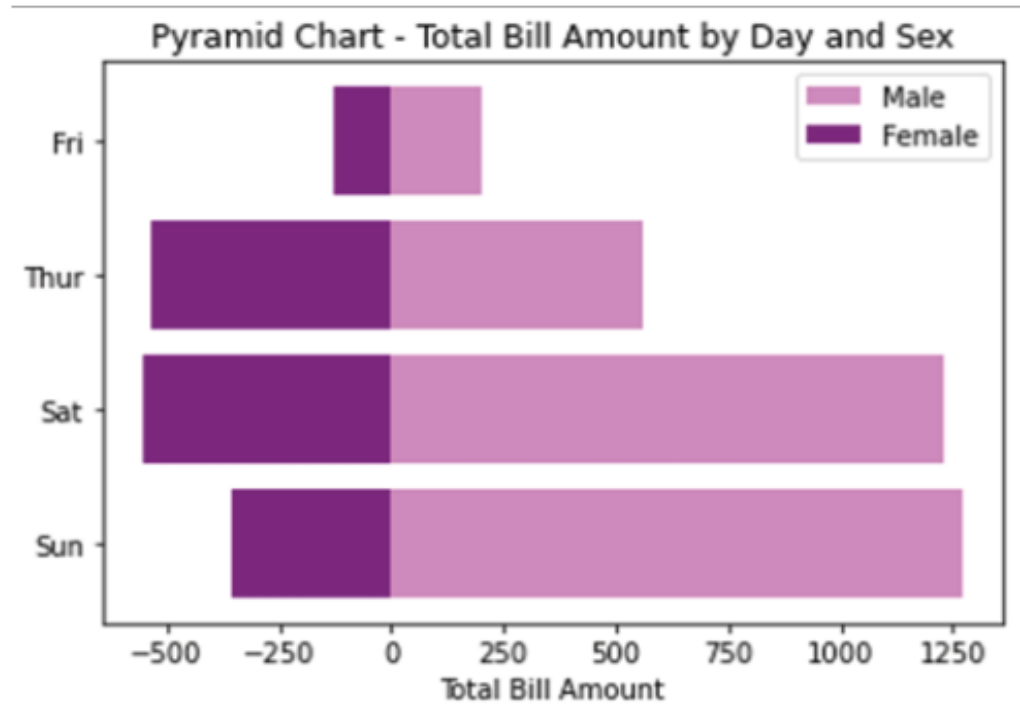


Figure 4.9: Pyramid Chart

هرم (Pyramid) بسیار شبیه به نمودار مخروطی است، اما در جهتی متفاوت قرار می‌گیرد. کوچکترین مجموعه داده در بالا قرار دارد در حالی که بزرگترین مجموعه در پایین قرار می‌گیرد. تصمیم‌گیری درباره استفاده از نمودار مخروطی یا نمودار هرم بستگی به نحوه ارائه داده‌ها دارد؛ آیا به ترتیب صعودی یا نزولی.

نمودار هرم همچنین می‌تواند بدون داده‌های عددی ایجاد شود. بخش‌ها به اندازه‌های مساوی تقسیم می‌شوند و سلسله مراتب گام‌ها یا اجزای یک کل که فقط به صورت بصری سلسله مراتبی هستند را نشان می‌دهند. همانند مثال روبه رو که نمودار هرم با رنگ‌های بنفش آورده شده است

Funnel chart (نمودار قیف)

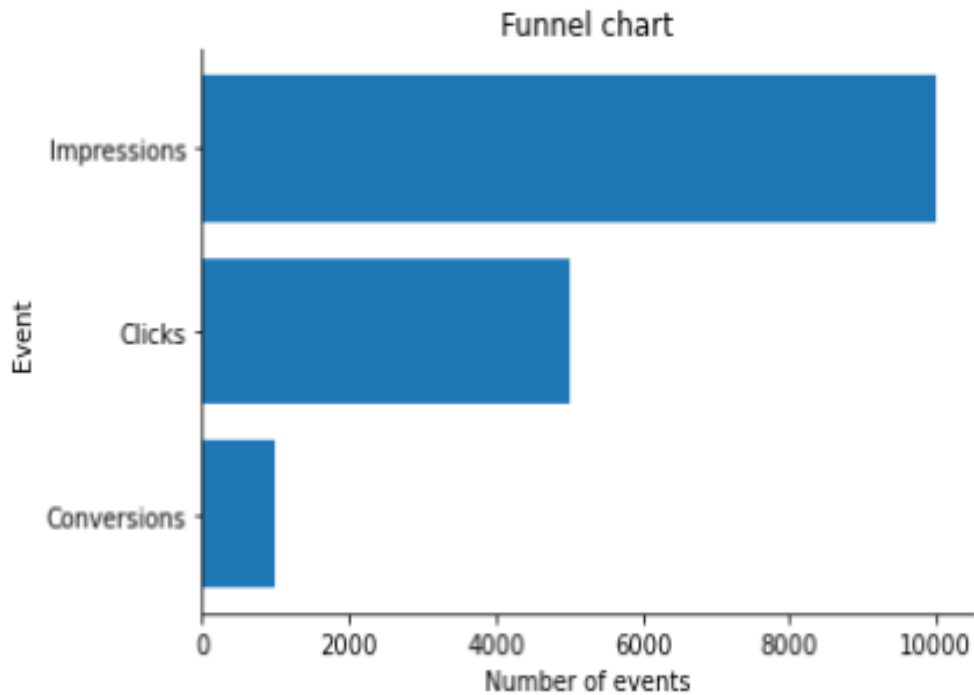


Figure 4.10: Funnel Chart

نمودار قیف به شکل نمودار مخروطی شبیه است، اما هدف کمی متفاوتی دارد. ایده اصلی نمودار قیف، دیداری سازی یک فرآیند متوالی از بالا به پایین است. به طور کلی، مجموعه داده در بالای فرآیند بزرگتر از پایین است زیرا در حین جریان به سمت پایین، تعداد به تدریج کاهش می‌یابد.

نمودارهای قیف در حوزه‌ها و برنامه‌های مختلفی مفید هستند که در آن‌ها یک ترتیب روشنی از مراحل یا گام‌های یک فرآیند وجود دارد و نیاز به دیداری سازی نرخ تبدیل یا پیشرفت از یک مرحله به مرحله بعدی وجود دارد. یکی از کاربردهای رایج نمودارهای قیف در فروش و بازاریابی است که برای پیگیری پیشرفت مشتریان پتانسیل از طریق قیف فروش، از تولید سرآمد تا تبدیل نهایی استفاده می‌شود.

Radar triangle (نمودار راداری)

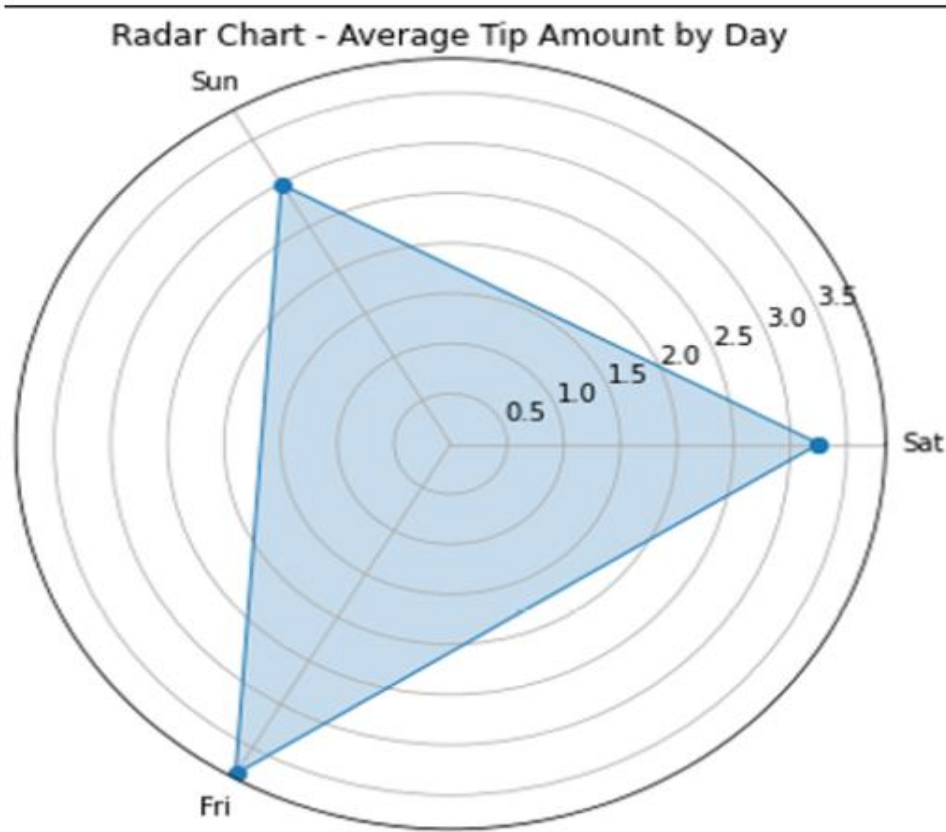


Figure 4.11: Radar Triangle Chart

نمودارهای راداری نوعی دیداری سازی داده هستند که به تجزیه و تحلیل موارد یا دسته‌بندی‌ها بر اساس تعداد خاصی ویژگی کمک می‌کنند. طرح نمودار راداری یک دایره با دایره‌های هم‌مرکز است که داده‌ها به صورت نقاط روی آنها قرار می‌گیرند. سپس نقاط با یکدیگر به صورت خطوط اتصال داده می‌شوند و یک شکل تشکیل می‌شود. هر مورد یا دسته‌بندی یک شکل است.

مثلاً راداری یک نمودار راداری است که موارد یا دسته‌بندی‌ها را بر اساس سه ویژگی مقایسه می‌کند. هر نقطه یک گوشه از مثلث است. مثلث می‌تواند تنها از خطوط تشکیل شده باشد یا با رنگ پر شود.

Radar Polygon (چندضلعی راداری)

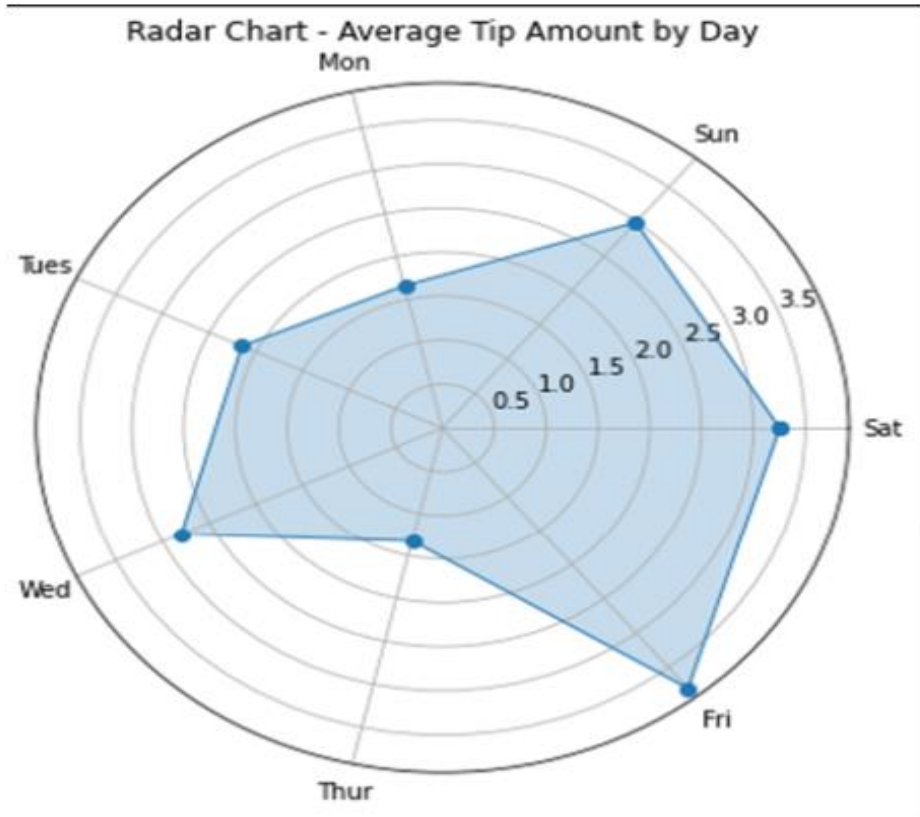
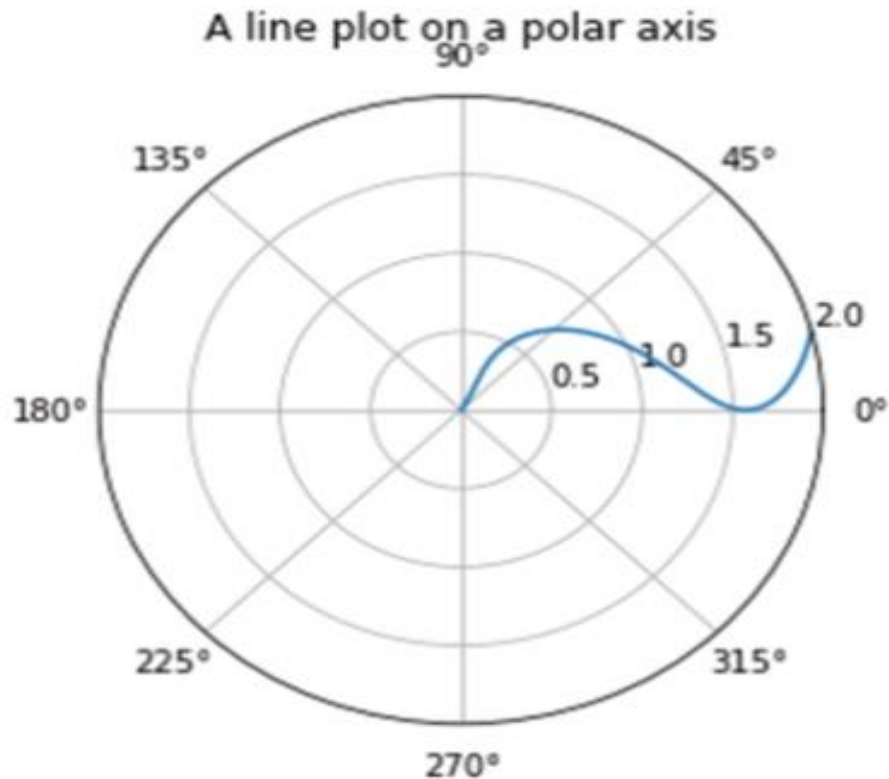


Figure 4.12: Radar Polygon Chart

یک چندضلعی راداری با یک مثلث راداری یکسان است، اما شکل نتیجه گرفته شده متفاوت است. یک مثلث راداری سه نقطه برای داده‌های ویژگی دارد، در حالی که یک چندضلعی راداری چهار یا بیشتر نقطه دارد. حداکثر تعداد نقاط ۹ یا ۱۰ است و حداکثر تعداد لایه‌های موارد ۴ یا ۵ است.

Polar chart (نمودار قطبی)

نمودار قطبی با نمودار راداری یک پایه دایره‌ای مشابه دارد، اما نقاط داده به صورت متفاوتی قرار می‌گیرند. به جای اتصال نقاط به یکدیگر، بخش‌های مخروطی از مرکز به سمت خارج گسترش می‌یابند.



تفاوت اصلی به طور اساسی بصری است. اگر مقادیر داده‌ها از هم بسیار متفاوت باشند، نمودار قطبی را انتخاب کنید. در غیر این صورت، خواندن آن به نگاه اولیه ممکن است چالش‌برانگیز باشد. یک نمودار قطبی به سادگی یک نمودار دکارتی است که محور X آن روی حاشیه دور می‌چرخد. این نمودار می‌تواند انواع متداول سری‌های دکارتی مانند خط، ستون، ناحیه یا بازه ناحیه را نشان دهد.

Area chart (نمودار مساحتی یا ناحیه ای)

نمودار مساحتی تغییری از نمودار خطی است. تفاوت این است که ناحیه بین خط مبنا و مقادیر ترسیم شده روی خط رنگی است. پررنگ نیمه شفاف است به طوری که مناطق همپوشانی به راحتی قابل خواندن هستند. حتی اگر می توانید هر نمودار خطی را به نمودار منطقه ای تغییر دهید، همیشه بهترین روش نیست. نمودار ناحیه نمی تواند بیش از چهار یا پنج مجموعه داده به طور همزمان داشته باشد. احتمال انسداد خیلی زیاد است. نمودارهای منطقه گاهی روی هم چیده می شوند و داده ها را به بخش هایی به عنوان بخشی از کل روابط یا به عنوان داده های تجمعی جدا می کنند

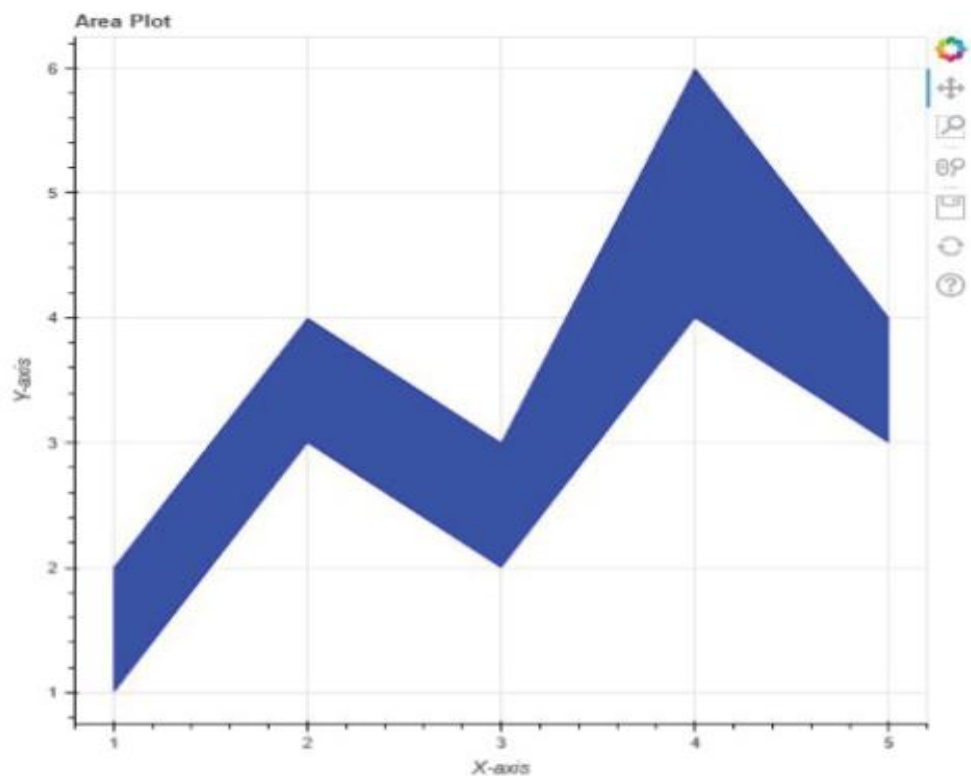


Figure 4.14: Area Chart

Tree chart (نمودار درختی)

یک نمودار درختی بیشتر یک دیداری سازی داده بصری است تا یک دیداری سازی برای داده‌های عددی دقیق. ایده اصلی نمودار درختی این است که داده‌ها را به عنوان اجزای یک به کل داخل یک دسته‌بندی تصویرسازی کند. برای یک نمودار درختی پیچیده‌تر، دسته‌بندی‌های مختلف را کنار یکدیگر قرار دهید.

نمودار درختی را انتخاب کنید زمانی که تصویرسازی شما به داده‌های عددی جزئی بستگی نداشته باشد. بهتر است اگر داده‌ها سلسله‌مراتبی باشند، نمودار درختی عملکرد خوبی خواهد داشت.

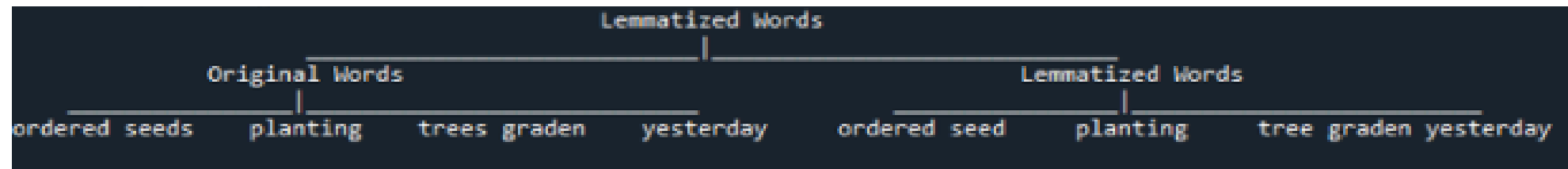


Figure 4.15: Tree Chart

Flowchart (فلوچارت)

یک فلوچارت نوع چند منظوره‌ای از دیداری سازی داده است. از یک فلوچارت برای توصیف بصری یک فرآیند، داده‌های سلسله مراتبی از آیتم‌ها یا افراد و حتی یک نقشه ذهنی برای استراتژی‌های تفکر خلاق استفاده کنید.

بهترین قسمت درباره فلوچارت‌ها این است که بسیار آسان است که آنها را برای هر پروژه سفارشی کنید.

از نظر بصری، فلوچارت‌ها با یک شکل سربزرگ شروع می‌شوند که به یک سری اشکال و خطوطی که به هم وصل می‌شوند، شاخه می‌زنند.

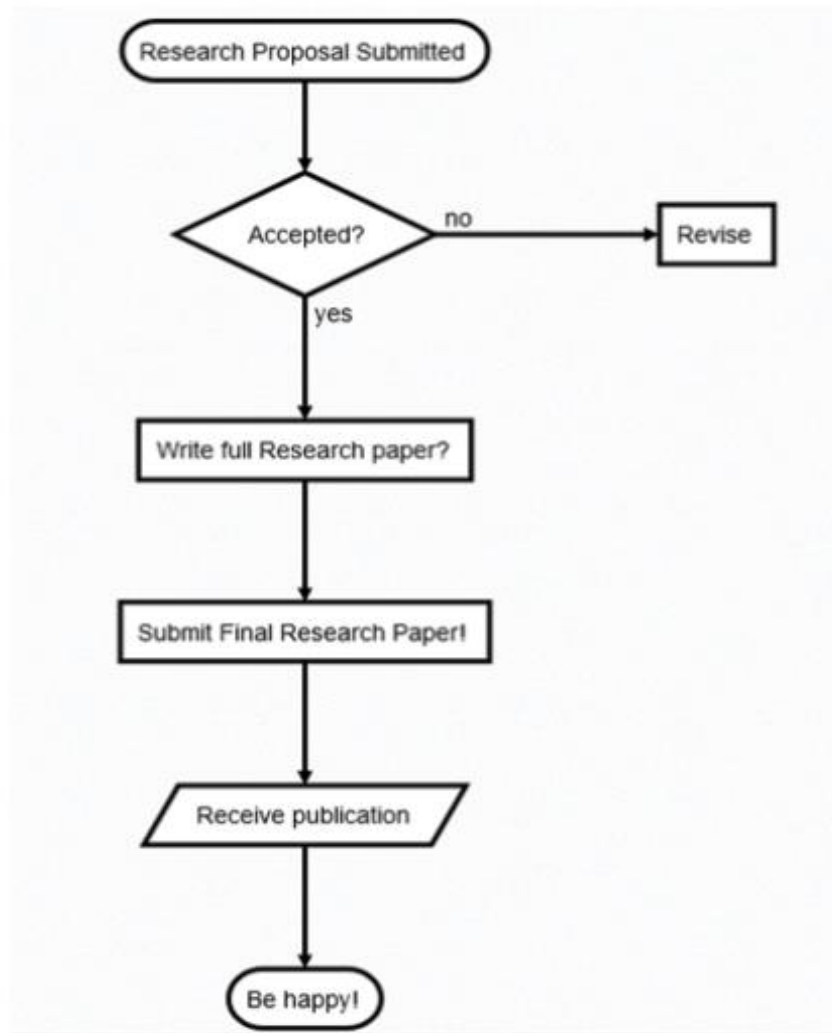


Figure 4.16: Flowchart

Table (جدول)

جداول مانند صفحه‌های ریزرسانه‌ای هستند و داده‌ها را در سطرها و ستون‌ها نشان می‌دهند. از یک جدول برای نمایش قیمت‌گذاری برای یک سرویس، ویژگی‌های مقایسه‌ای یک محصول، گزارش‌های مدرسه و موارد دیگر استفاده کنید.

این نوع دیداری سازی داده به خوبی در سندهای تصویری مانند گزارش‌ها، پیشنهادات و کتابچه‌های آموزشی جا می‌شود. برای استفاده منحصر به فرد از جدول به عنوان یک دیداری سازی داده، از نقاط یا آیکون‌ها برای نمایش داده‌های بله یا خیر درباره یک دسته خاص استفاده کنید.

Chapter Title	Learning Objective
Understanding Data	Familiarize with Data, Types and attributes
Data Visualization: Importance	Understand the importance of Visualization
Data Visualization: Use cases	To explore various use cases on visualization
Data Visualization: Tools and Techniques	Explore and understand various techniques and tools

Figure 4.17: Table

Geospatial map and choropleth map (نقشه جغرافیایی و نقشه کروپلث)

نقشه‌ها بهترین دیداری سازی داده برای هر داده‌ای هستند که ارتباطی با موقعیت جغرافیایی دارد. نقشه داده‌ها کاربردهای متعددی دارد، از اطلاعات کشور به کشور تا تحلیل دقیق مناطق. نقشه کروپلث بر اساس نقشه جغرافیایی است اما هدف خاصی دارد. نقشه کروپلث نمایش جغرافیایی مقادیر آماری بر اساس منطقه است. به عنوان مثال، چگالی جمعیت در یک کشور توسط ایالت‌ها نمایش داده می‌شود.

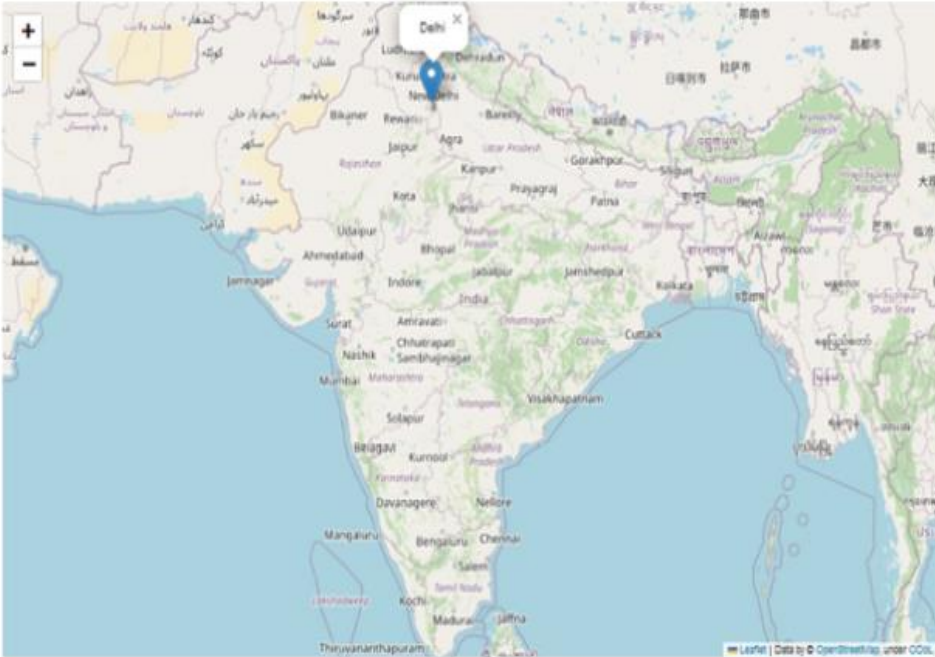


Figure 4.18: Choropleth Map

مقادیر به بخش‌های مساوی تقسیم می‌شوند و برای هر بخش رنگ اختصاص داده می‌شود. سپس مناطق مرتبط با نقشه به رنگ‌ها متناسب با مقادیر آنها رنگ می‌شوند. این تجسم‌ها برای سازمان‌های غیرانتفاعی، کسب و کارهای در صنعت بهداشت و درمان یا هر کسی که می‌خواهد داده‌های آماری یک منطقه جغرافیایی خاص را دیداری سازی کند، ایده‌آل هستند.

Percentage bar (نوار پیشرفت یا درصدی)

نوعی دیداری سازی ساده داده به نام نوار پیشرفت یا درصدی برای نمایش مقدار درصد استفاده می‌شود. نمودار نوار درصدی یک نوع خاص از نمودار نوار است که نسبت یا درصدهای نسبی دسته‌ها یا گروه‌های مختلف را نشان می‌دهد. در یک نمودار نوار درصدی، ارتفاع هر نوار نسبت به کل، مقدار درصد یک دسته را نشان می‌دهد.

این نوع نمودارها در ایجاد یک اطلاعاتی اینفوگرافیک یا گزارش پیشرفت مفید است.

از آنجا که نوارهای درصدی به اندازه کوچک هستند، به خوبی به صورت گروهی کار می‌کنند. تفاوت اصلی بین نمودار نوار درصدی و نمودار نوار معمولی این است که محور Y در نمودار نوار درصدی درصدها را نشان می‌دهد (از ۰٪ تا ۱۰۰٪) در حالی که نمودار نوار معمولی معمولاً مقادیر واقعی داده‌های نمایش داده شده را نشان می‌دهد.

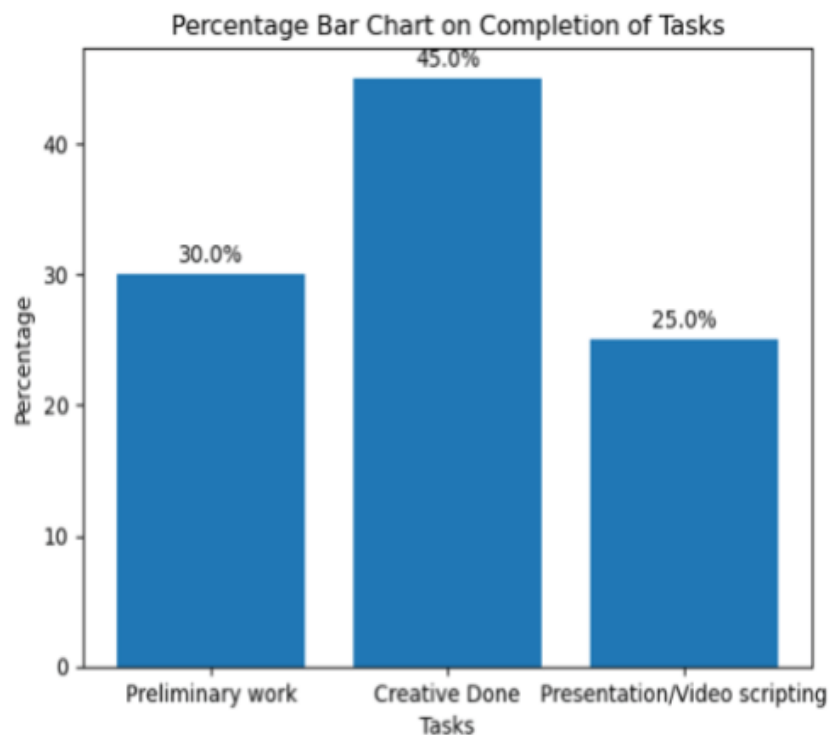


Figure 4.19: Percentage Bar

Radial wheel (نمودار چرخشی)

نمودار چرخشی (همچنین با نام نمودار نواری یا نمودار دایره‌ای) داده‌ها را به صورت نوارهایی نمایش می‌دهد که به صورت شعاعی از مرکز یک دایره بیرون می‌روند.

یک تصویر سازی داده با استفاده از دایره‌های هم‌مرکز مشابه یک نمودار خطی است که بر روی یک محور دایره‌ای قرار دارد. هر دسته یا مورد داده به صورت یک دایره در نمودار نمایش داده می‌شود و هر دایره به تفکیک رنگی دارد و بر اساس داده‌ها روی محور دایره‌ای قرار می‌گیرد. همچنین، دایره‌ها به صورت هم‌مرکز قرار می‌گیرند. این نمایش برای نمایش ساختارهای سلسله‌مراتبی یا ترتیبی بسیار ارزشمند است، مانند سازماندهی افراد در یک محله، شهر، ایالت یا کشور. با استفاده از دایره‌های هم‌مرکز، موقعیت نسبی و روابط بین موجودیت‌ها به راحتی قابل نمایش است.

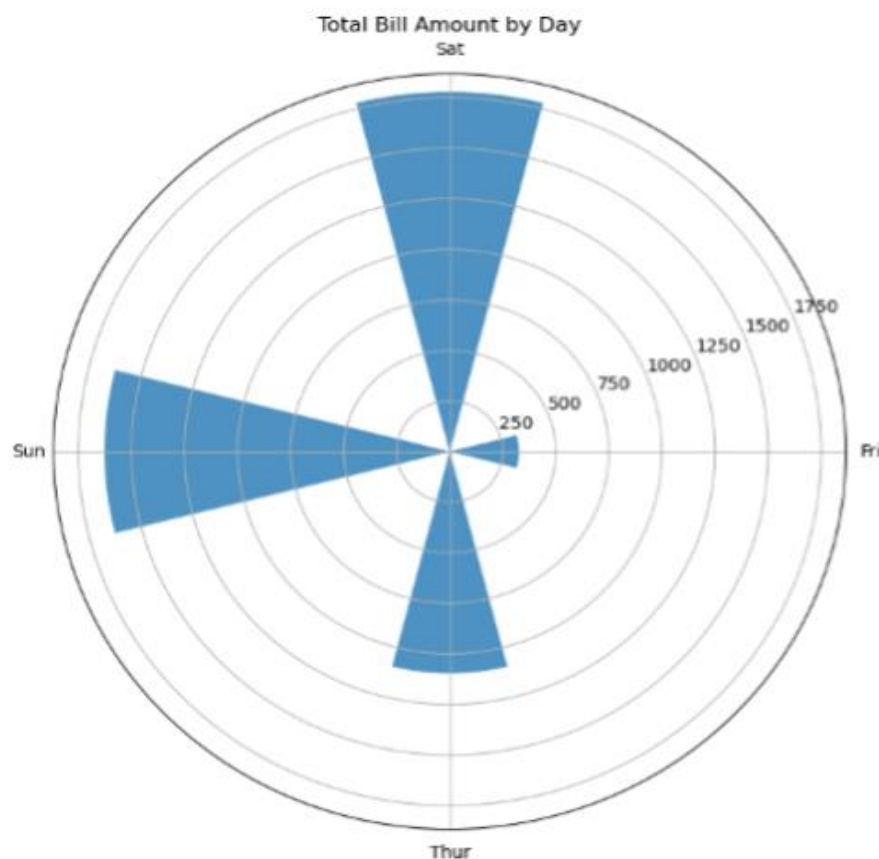


Figure 4.20: Radial Wheel/Circular bar plot

Concentric Circles (نمودار دایره ای متحدالمركز)

نمودار داده‌ای با دایره‌های متحدالمركز شبیه به یک نمودار خطی روی یک محور دایروی است. هر دسته یا داده در نمودار یک دایره است و هر دایره رنگ خاص خود را دارد و بر اساس داده روی محور دایروی رسم می‌شود. همچنین، این دایره‌ها به صورت متحدالمركز چیده شده‌اند. این نوع از نمودار به خصوص برای نمایش سلسله مراتب یا ترتیب، مانند سازماندهی افراد در یک محله، شهر، استان یا کشور، ارزشمند است. با استفاده از دایره‌های متحدالمركز، موقعیت‌های نسبی و روابط بین نهادها به راحتی قابل نمایش است. به عنوان مثال، دایره‌های بزرگ‌تر می‌توانند نمایانگر نهادهای سطح بالاتر مانند یک کشور باشند، در حالی که دایره‌های کوچک‌تر که درون آن‌ها قرار گرفته‌اند می‌توانند نشانگر نهادهای سطح پایین‌تر مانند استان‌ها یا شهرها باشند. این چیدمان سلسله مراتبی به تجسم سازماندهی و نظم سطوح مختلف در یک سیستم کمک می‌کند و درک و تحلیل ساختارهای پیچیده را تسهیل می‌بخشد. برای اینکه خواندن نمودار آسان باشد، نباید بیش از شش دایره متحدالمركز وجود داشته باشد.

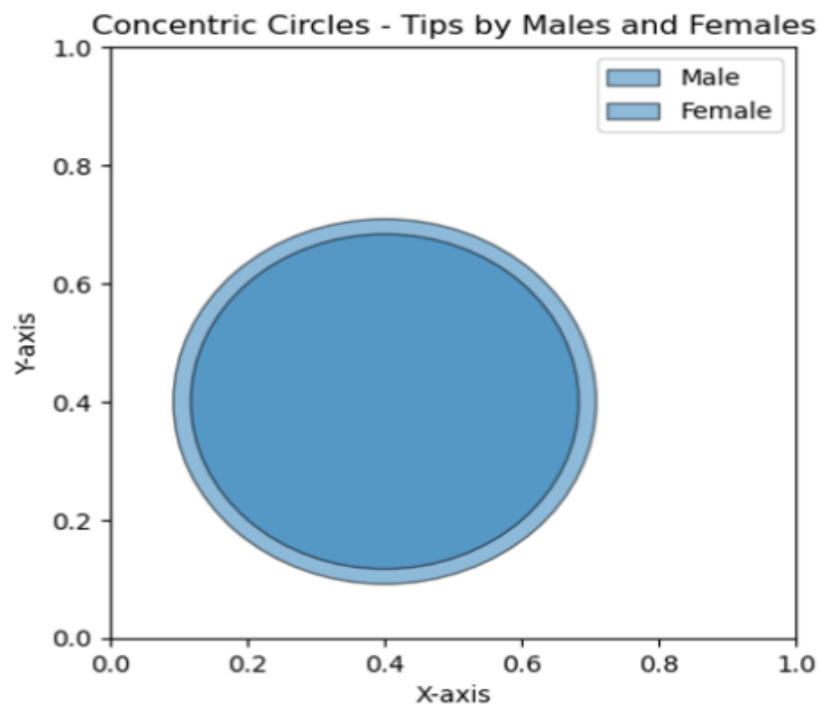


Figure 4.21: Concentric Circles

Gantt chart (نمودار گانت)

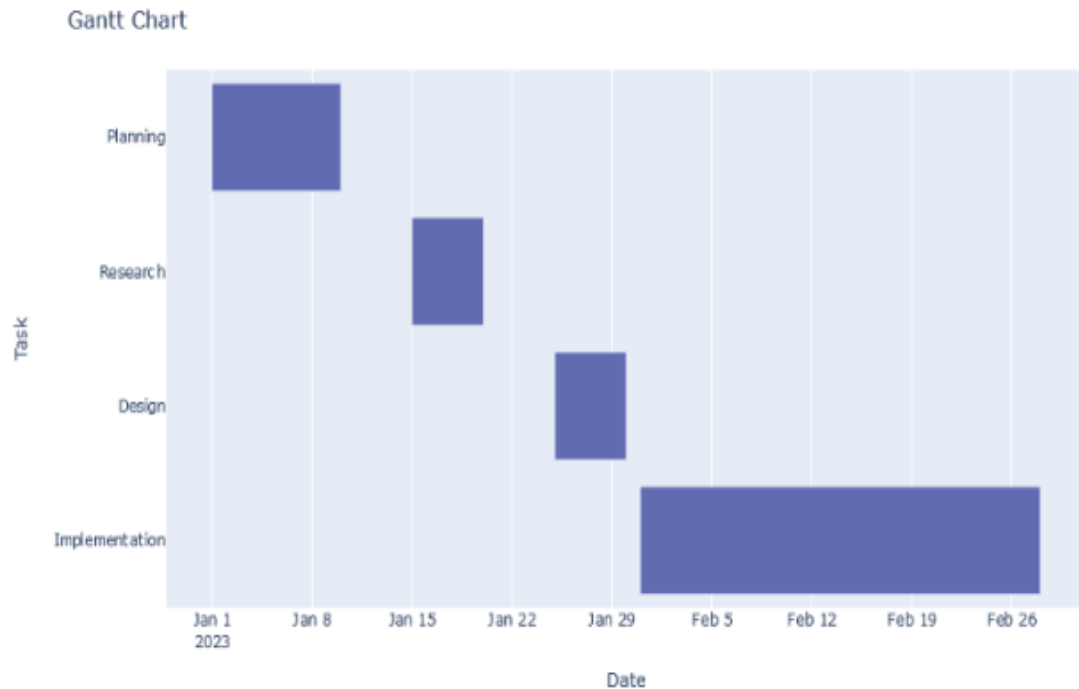


Figure 4.22: Gantt Chart

دیاگرام گانت بر اساس نمودارهای نواری افقی است، اما در یک جنبه کلیدی با آن‌ها تفاوت دارد. در یک نمودار گانت، مهم نیست که داده‌ها در طول زمان چگونه تغییر می‌کنند، بلکه مهم این است که هر فعالیت چه مدت طول می‌کشد تا در یک بازه زمانی خاص تکمیل شود. هر مورد در نمودار توسط یک مستطیل نمایش داده می‌شود که از چپ به راست دراز شده است. هر مستطیل اندازه‌ای متفاوت دارد که بسته به مدت زمانی که هر فعالیت برای تکمیل آن طول می‌کشد، تعیین می‌شود.

Network diagram (نمودار شبکه ای)

این نمودارها به طور کلی به صورت نمودارهای ساده و بدون تزئین زیاد طراحی می‌شوند. آنها اتصال بین یک مجموعه از موجودیت‌ها را نشان می‌دهند. هر موجودیت توسط رئوس/گره‌ها نمایش داده می‌شود. اتصال بین گره‌ها از طریق لبه‌ها نمایش داده می‌شود.

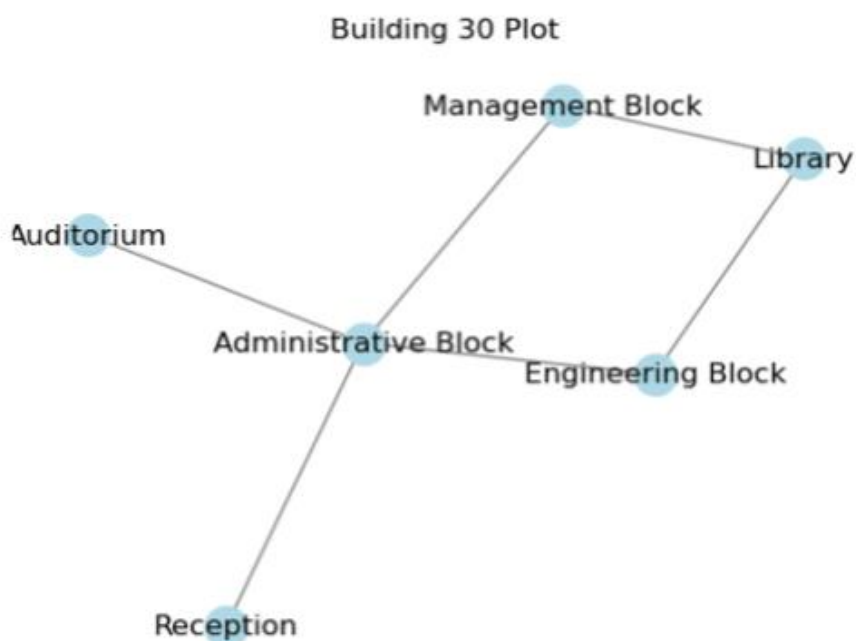


Figure 4.23: Network Diagram

نمودارهای شبکه ابزارهای چند منظوره هستند که در طیف گسترده‌ای از حوزه‌ها استفاده می‌شوند تا روابط و اتصالات بین موجودیت‌ها را به صورت تصویری نمایش دهند. در شبکه‌های اجتماعی، آنها به تصویر کشیدن دوستی‌ها، همکاری‌ها و تعاملات بین افراد کمک می‌کنند و امکان شناسایی تأثیرگذاران کلیدی و جوامع را فراهم می‌کنند. در شبکه‌های ارتباطی، نمودارهای شبکه جریان اطلاعات در سازمان‌ها یا شبکه‌های کامپیوتری را تصویر می‌کنند و در بهینه‌سازی مسیرهای ارتباطی و شناسایی نقاط مهم کمک می‌کنند.

Timeline (جدول زمانی)

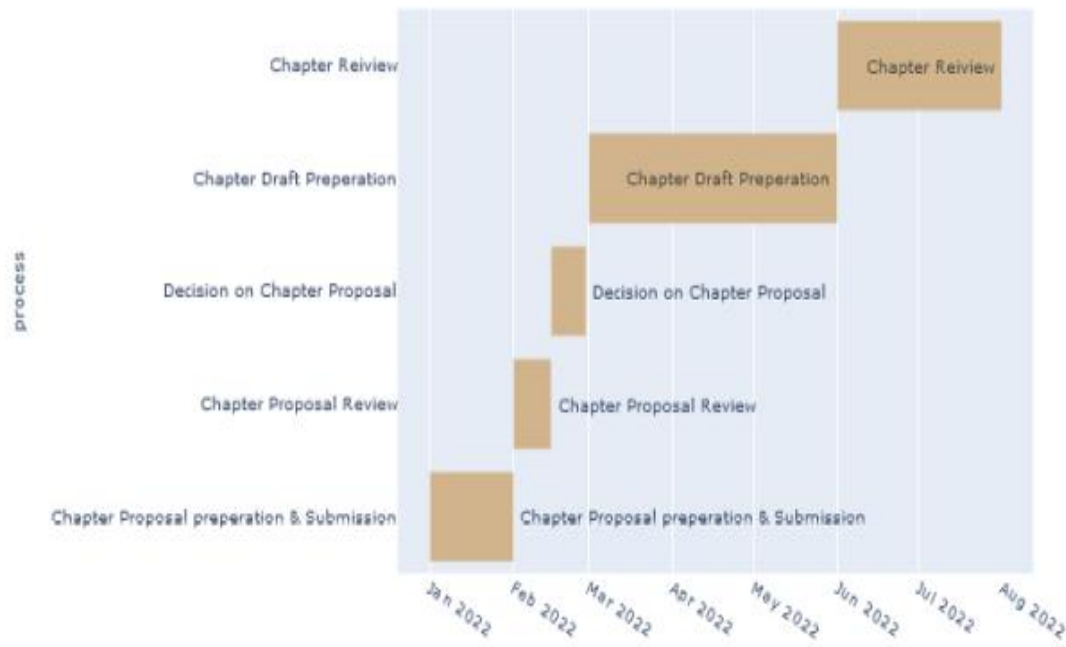


Figure 4.24: Timeline

جدول زمانی‌ها تصویرسازی‌هایی هستند که رویدادهایی که در یک دوره زمانی خاص رخ داده یا خواهند داد، را نشان می‌دهند. جدول زمانی‌ها در طرح‌های عمودی و افقی بر روی یک اسلاید ترتیبی یا چندین اسلاید پی در پی بسیار عالی عمل می‌کنند. از این نوع تصویرسازی داده برای گزارش‌های اطلاعاتی درباره موضوعاتی با داستان پشت صحنه یا برای تصویرسازی داستان رشد یک شرکت استفاده کنید.

Venn Diagram (نمودار ون)

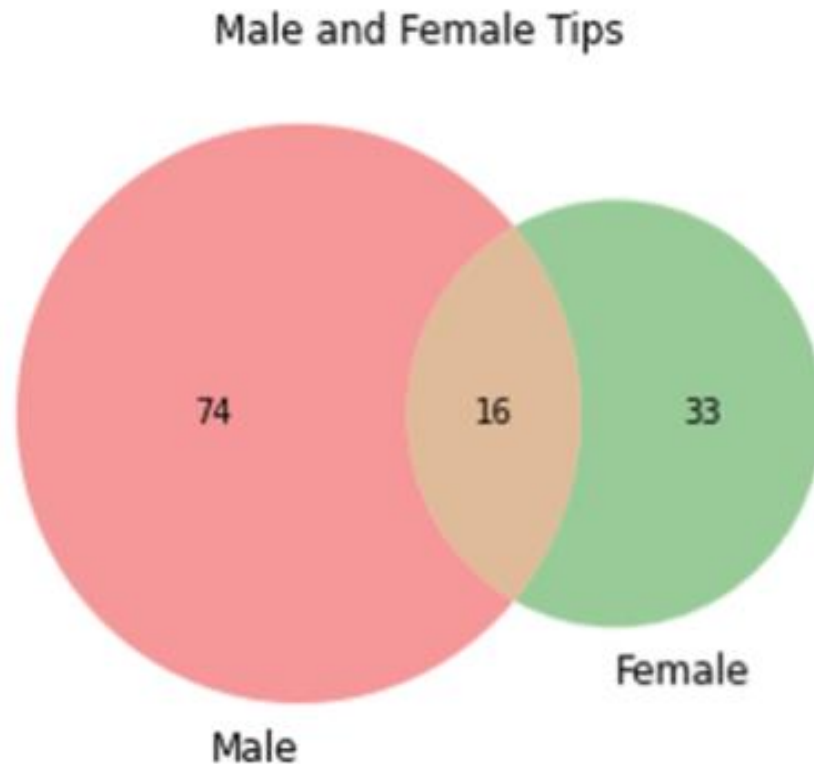


Figure 4.25: Venn Diagram

نمودار ون نوعی تصویرسازی داده است که هدف آن مقایسه دو یا چند مورد با برجسته کردن آنچه که آنها مشترک دارند است. سبک رایج‌تر برای نمودار ون دو دایره است که به هم متداخل می‌شوند. هر دایره نمایانگر یک مفهوم است و ناحیه‌ای که آنها را به هم وصل می‌کند، آنچه که دو مورد مشترک دارند را نشان می‌دهد. نمودارهای ون می‌توانند تا چهار یا پنج دایره مفهوم داشته باشند که ناحیه‌های ترکیبی نشان می‌دهند که چه چیزی در میان آنها مشترک است.

استفاده از بیش از سه یا چهار دایره یا شکل در یک نمودار ون بسیار پیچیده می‌شود.

Histogram (هیستوگرام)

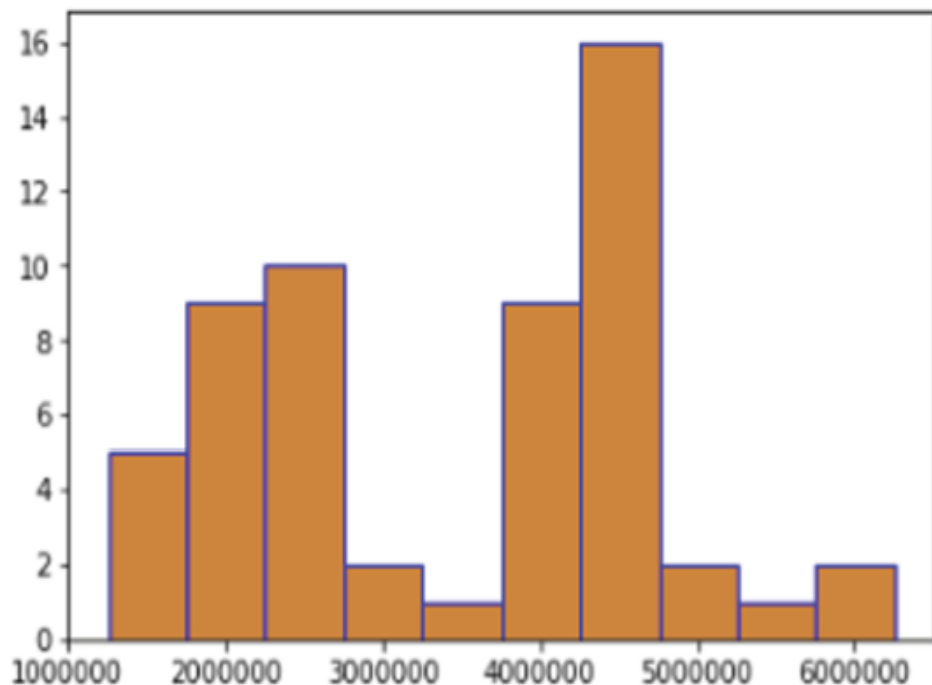


Figure 4.26: Histogram

یک نمودار ستونی مانند نمودار میله‌ای است، اما سیستم نمایش متفاوتی دارد. نمودارهای ستونی بهترین نوع تصویرسازی داده برای تحلیل دامنه‌های داده بر اساس یک فرکانس خاص هستند. آنها مانند یک نمودار میله‌ای ساده هستند، اما به طور خاص برای تصویرسازی داده‌های فراوانی در یک دوره زمانی خاص استفاده می‌شوند.

نمودارهای ستونی تنها به صورت عمودی قابل نمایش هستند، متفاوت از نمودارهای میله‌ای که به صورت عمودی و افقی هر دو قابل نمایش هستند.

Mind map (نمودار ذهنی)

نمودار ذهنی نوع دیگری از تصویرسازی داده است که به فرد در تولید ایده‌ها و سازماندهی آنها کمک می‌کند. از نظر ظاهری، نمودار ذهنی یک شبکه از شکل‌ها است که بر اساس مفهوم مرتب شده‌اند و به ترتیب سلسله‌مراتبی به یکدیگر وصل شده‌اند. یک نمودار ذهنی می‌تواند کوچک باشد و تنها شامل چند شکل مرتبط باشد یا بسیار بزرگ باشد و شکل‌های بسیاری از ایده‌های اصلی بیرون زده شوند. آنها نمایش‌های گرافیکی هستند که به افراد اجازه می‌دهند ایده‌ها، مفاهیم و ارتباطات را به صورت بصری ثبت، کاوش و ارائه کنند. نمودارهای ذهنی اغلب برای برگزاری جلسات ایده‌پردازی، حل مسئله، برنامه‌ریزی، گرفتن یادداشت و سازماندهی دانش استفاده می‌شوند.

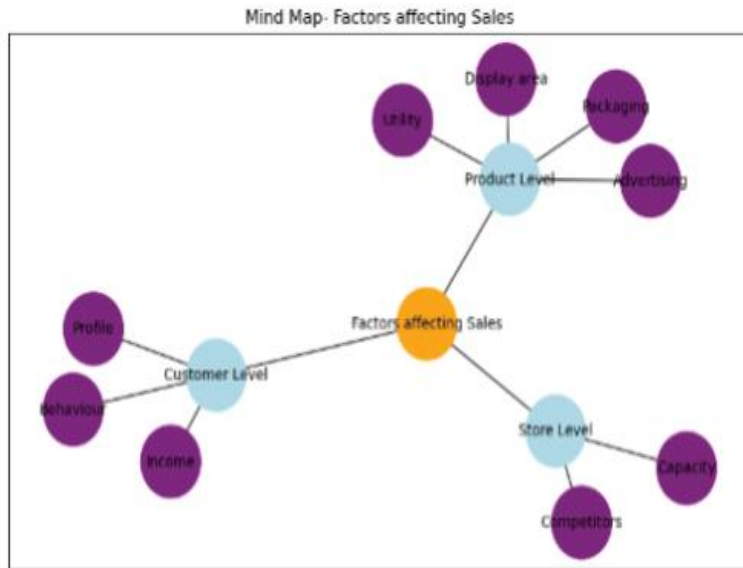


Figure 4.27: Mind Map

اجزای اصلی یک نمودار ذهنی شامل یک موضوع یا ایده مرکزی است که در مرکز نمودار قرار می‌گیرد و شاخه‌ها یا پره‌ها از مرکز به سمت خارج پخش می‌شوند. این شاخه‌ها زیرموضوع‌ها یا ایده‌های مرتبط را نمایندگی می‌کنند و می‌توانند به شاخه‌های فرعی یا جزئیات بیشتری گسترش یابند.

Dichotomous key (کلید دوتایی)

یک کلید دوتایی نوع دیگری از تصویرسازی نموداری است که هدف آن کمک در تصمیم‌گیری است. این نوع تصویرسازی معمولاً به عنوان یک سری سوال با دو پاسخ ممکن (اما می‌تواند سه یا چهار پاسخ داشته باشد وابسته به طول و پیچیدگی کلید) ارائه می‌شود. هر سوال با حذف یکی از احتمالات، انتخاب‌ها را محدود می‌کند. این نوع تصویرسازی معمولاً به صورت خطی نمایش داده می‌شود، به طوری که هر سوال به سوال بعدی منتقل می‌شود تا به تشخیص نهایی برسیم. با پاسخ به هر سوال، در جریان نمودار به پاسخ مناسب حرکت می‌کنید.

کلیدهای دوتایی به طور اصلی در حوزه طبقه‌بندی و زیست‌شناسی برای شناسایی و طبقه‌بندی موجودیت‌ها بر اساس ویژگی‌هایشان استفاده می‌شوند. با این حال، مفهوم تصمیم‌گیری دوتایی می‌تواند در علم داده نیز استفاده شود، به خصوص در زمینه طبقه‌بندی داده و انتخاب ویژگی. در علم داده، تصمیم‌گیری دوتایی می‌تواند برای ساخت درخت تصمیم (Decision Tree) استفاده شود که یک الگوریتم یادگیری ماشین محبوب است.

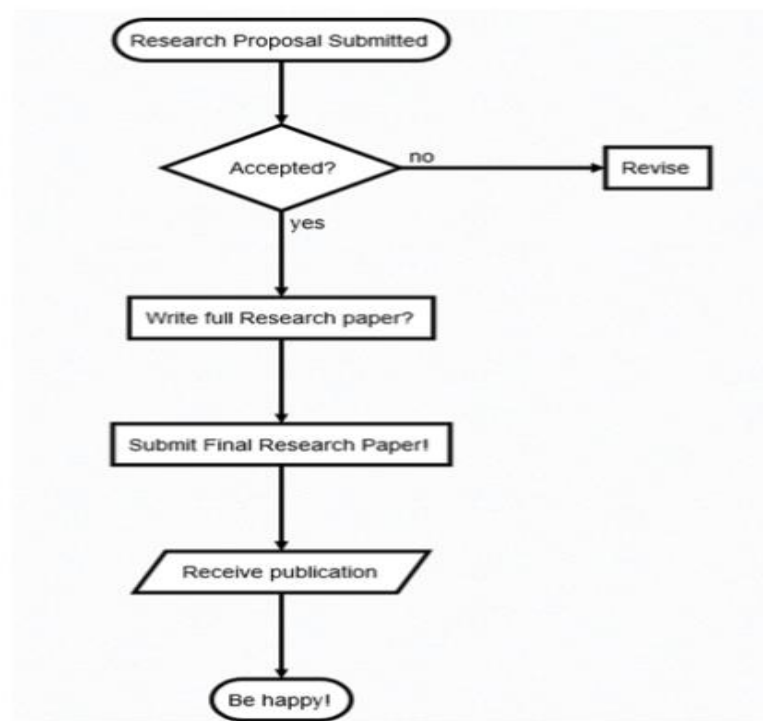


Figure 4.16: Flowchart

PERT chart (نمودار جریان)

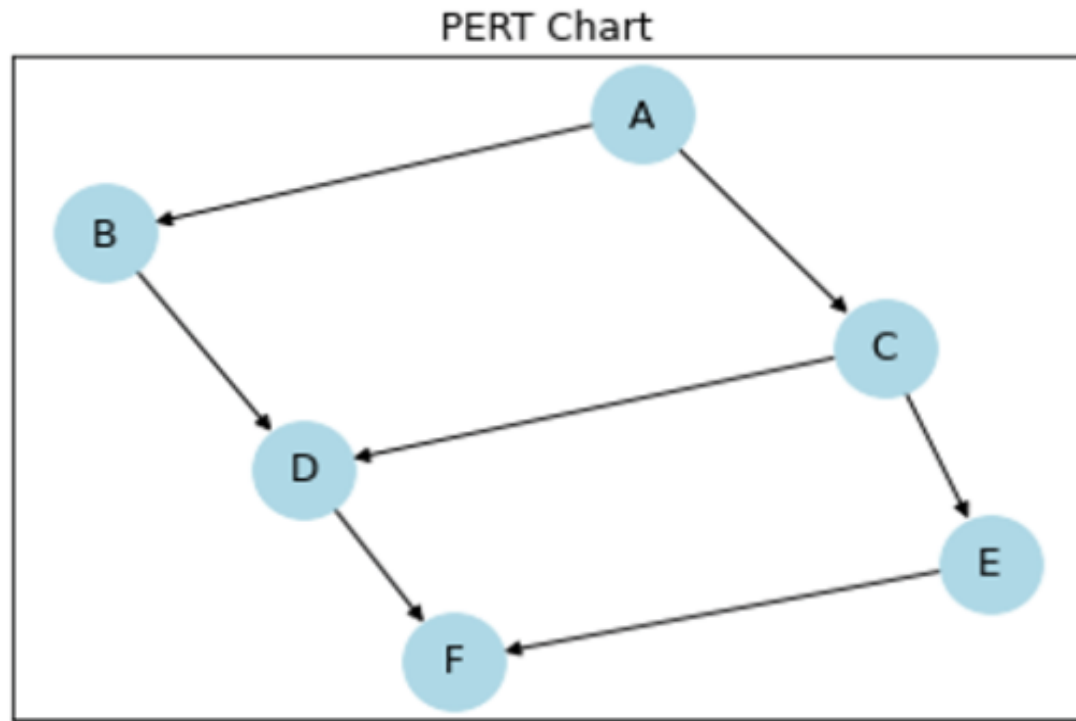


Figure 4.28: Pert Chart

نوع دیگری از دیداری سازی داده بر اساس نمودار جریان معروف، در نمودارهای PERT برنامه ارزیابی و بررسی فنی وجود دارد. در این نوع نمودار، نمودارهای مدار و نقشه‌های فرآیند با یکدیگر ترکیب می‌شوند. هر مورد به عنوان یک فرآیند در نظر گرفته می‌شود و شکل قبلی باید قبل از شروع شکل بعدی تکمیل شود، مگر اینکه در نمودار دیگری مشخص شده باشد.

Box plot (نمودار جعبه ای)

نمودارهای جعبه‌ای برای نمایش توزیع مقادیر عددی داده‌ها استفاده می‌شوند، به ویژه زمانی که می‌خواهید آن‌ها را بین چند گروه مقایسه کنید. این نمودارها برای ارائه اطلاعات سطح بالا در لحظه ایجاد شده‌اند و اطلاعات کلی درباره تقارن، انحراف، وجود پراکندگی و نقاط پرت یک گروه داده را ارائه می‌دهند.

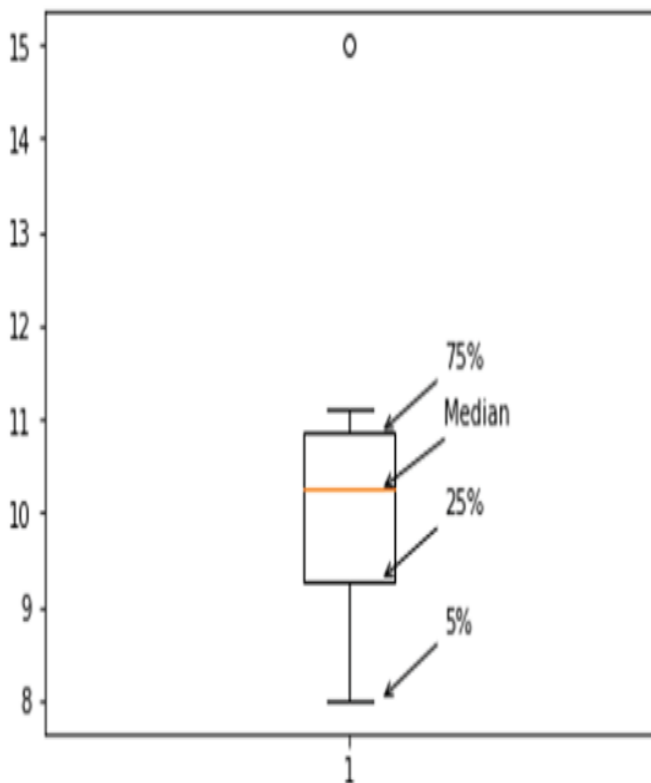


Figure 4.29: Box Plot

میان (پنجاهمین درصدی / $Q2$) : مقدار وسطی مجموعه داده‌ها

چارک اول (بیست و پنجمین درصدی / $Q1$) : میانگین بین کوچکترین عدد و میان مجموعه داده‌ها

چارک سوم (هفتاد و پنجمین درصدی / $Q3$) : میانگین بین میان و بزرگترین مقدار مجموعه داده است.

محدوده بین ربعی (IQR) : 25 تا صدک 75، فاصله بین $Q1$ و $Q3$ به عنوان محدوده بین چارکی شناخته می‌شود

دماغه‌ها IQR : (Whiskers) نقش اصلی خود را در طول دماغه‌ها بازی می‌کند. هر دماغه به نقطه داده‌ای در بالاترین

و پایین‌ترین قسمت هر بال با فاصله‌ای کمتر از 1.5 برابر IQR می‌رسد.

اگر داده‌ها به انتهای دماغه‌ها نرسند، دماغه‌ها به حداقل و حداکثر مقادیر داده‌ها می‌رسند.

نقاط پرت (به صورت حبابی خارج از دماغه نمایش داده می‌شوند): نقطه‌ای که خارج از دماغه‌های نمودار جعبه‌ای

قرار دارد، به عنوان نقطه پرت تعریف می‌شود.

"حداکثر" : $Q3 + 1.5 \times IQR$

"حداقل" : $Q1 - 1.5 \times IQR$

Heatmap (نمودار حرارتی)

نمودار حرارتی یک نوع دیداری سازی داده است که نقاط داده را با استفاده از رنگ‌ها به عنوان نماینده‌های مقادیر داده در قالب جدولی نشان می‌دهد. اغلب از آن برای نشان دادن همبستگی بین دو یا بیشتر متغیر استفاده می‌شود، جایی که هر سلول در جدول نشان دهنده ترکیبی از مقادیر متغیرهاست. رنگ سلول بر اساس بزرگی مقدار داده تعیین می‌شود، به طوری که رنگ‌های تیره یا روشن مقادیر بالا یا پایین را نشان می‌دهند.

نمودارهای حرارتی در زمینه‌هایی مانند زیست‌شناسی، امور مالی، بازاریابی و علوم اجتماعی به طور گسترده استفاده می‌شوند.

در امور مالی، نمودارهای حرارتی برای نمایش عملکرد سهام یا سایر ابزارهای مالی استفاده می‌شوند، جایی که هر ردیف نمایانگر یک سهم است و هر ستون نمایانگر یک زمان است و رنگ هر سلول درصد تغییر در ارزش سهم در طی آن زمان را نشان می‌دهد.

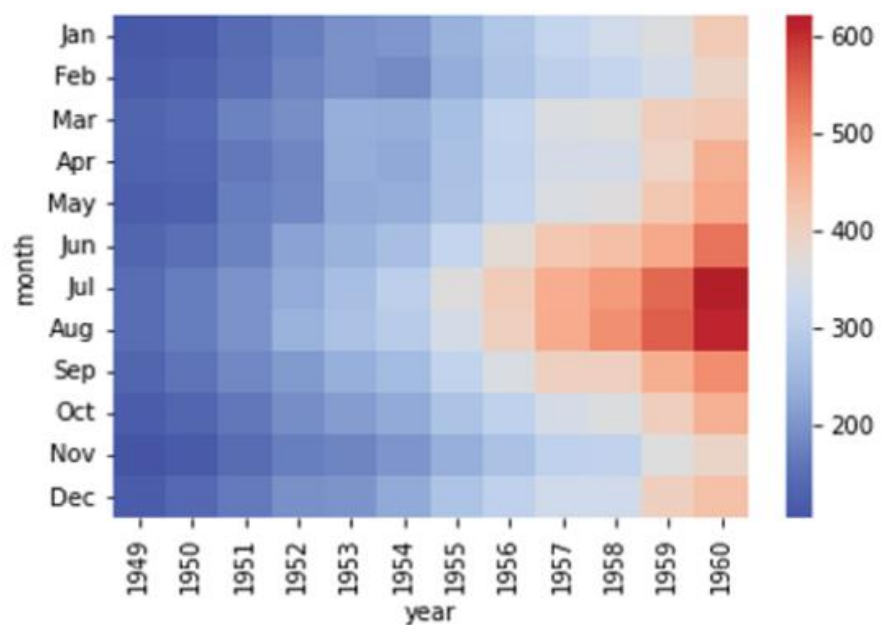


Figure 4.30: Heatmap

ابزارها و تکنیک‌های محبوب تصویرسازی داده :



Tableau

نرم افزار های مختلفی با ویژگی‌های متفاوت برای منبع داده‌های مختلف و تولید سریع تصاویر بصری در دسترس است. نسخه اولیه این ابزار، Tableau Desktop بود. این نسخه قادر به تولید نقشه‌های پویا نمی‌باشد و به جای آن، برای تولید تصاویر ثابتی طراحی شده است که می‌توان آن‌ها را در یک یا چند صفحه وب منتشر کرد.

Tableau Public نسخه رایگانی از نرم‌افزار دسکتاپ Tableau است که با محدودیت‌هایی عرضه شده است.



QlikView

این ابزار یک کیت توسعه نرم‌افزاری کاربرپسند است که سال‌هاست در سازمان‌ها در سراسر جهان استفاده می‌شود. آن قادر است انواع مختلفی از منابع داده را با بصری‌سازی در جداول رنگی، نمودارهای میله‌ای، نمودارهای خطی، نمودارهای دایره‌ای و لغزنده‌ها ترکیب کند.

این ابزار بر روی یک رابط بصری "کشیدن و رها کردن" توسعه یافته است که به کاربران امکان می‌دهد به راحتی داده‌ها را از منابع مختلفی مانند پایگاه داده‌ها یا صفحه‌های محاسباتی اضافه کنند، بدون نیاز به نوشتن هرگونه کد. این ویژگی‌ها همچنین آن را به یک ابزار نسبتاً آسان برای یادگیری و درک می‌کنند.





Microsoft Power BI

Microsoft Power BI یک ابزار بصری سازی داده است که برای نوعی داده های هوش تجاری استفاده می شود. این ابزار برای گزارش دهی، تجزیه و تحلیل خودکار و تجزیه و تحلیل پیش بینی ای استفاده می شود.

علاوه بر این، Power BI یک پلتفرم کاربر نهایی را برای ایجاد گزارش ها و به اشتراک گذاری بینش ها با سایر اعضای سازمان فراهم می کند. این به عنوان یک مخزن متمرکز برای تمام داده های تجاری شما عمل می کند که توسط تمام کاربران تجاری شما قابل دسترسی است.

علاوه بر این، Power BI همچنین امکان ادغام با سایر محصولات SaaS مانند Google Analytics، Mail Chimp، Office 365 و غیره را فراهم می کند. از طریق این ادغام ها، گزارش های ایجاد شده در سازمان به اشتراک گذاشته می شوند، که این امر آن را به یک ابزار بسیار مهم برای سازمان هایی می کند که به دنبال یک سیستم مرکزی گزارش دهی داده می باشند.





Data wrapper

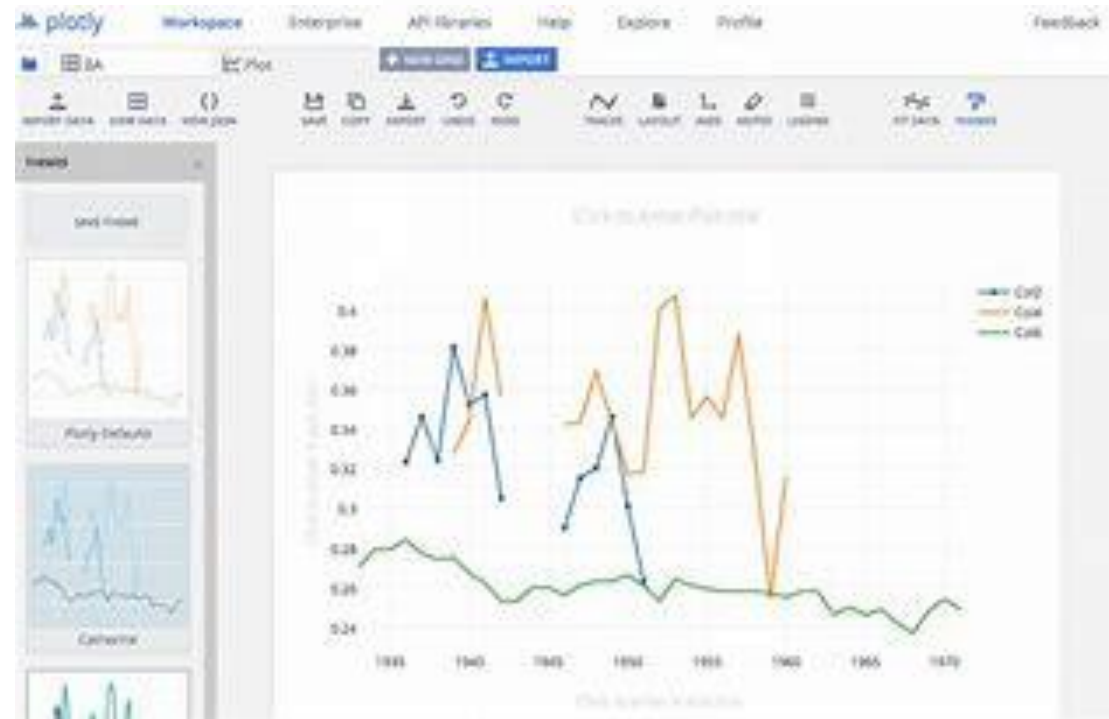
یک Datawrapper یک ابزار بصری سازی داده آنلاین است که در زمینه های مختلف قابل استفاده است. استفاده از آن بسیار آسان است و رابط کاربری آن ساده و قابل فهم است.

Datawrapper به کاربران امکان می دهد نمودارها و نقشه ها را مستقیماً در مرورگر خود ایجاد کنند با آپلود فایل های داده خود.

Datawrapper برای همه کاربران رایگان است؛ با این حال، نسخه رایگان آن محدودیت هایی دارد. به عنوان مثال، فقط به شما امکان می دهد تا ۵۰۰ سطر داده و یک برگه (یا ۵ مگابایت) داده را در یک زمان بارگذاری کنید. انواع نمودارهای قابل استفاده شامل خطی، میله ای، مساحی، ستونی، دایره ای و پراکنده می باشند. فایل های داده قابل بارگذاری شامل فایل های .csv، .tsv و .txt می باشند.



Plotly



Plotly یک ابزار بصری سازی داده است که برای ایجاد نمودارها، نمودارها و نقشه های تعاملی استفاده می شود. همچنین می توانید از Plotly برای ایجاد یک دیداری سازی داده و سپس به اشتراک گذاری لینک آن دیداری سازی با خوانندگان در رسانه های اجتماعی یا وبلاگ خود استفاده کنید. نمودارهای Plotly به راحتی قابل به اشتراک گذاری هستند زیرا تعاملی هستند و دارای یک URL ویژه هستند. با حرکت نشانگر موس بر روی نقاط داده و بررسی اطلاعات مربوط به آنها، خوانندگان می توانند بررسی کنند که چگونه شما آنها را ایجاد کرده اید.



Sisense

Sisense از تکنولوژی چندبُعدی حافظه درونی استفاده می‌کند که برای داده‌های بزرگ طراحی شده است. همچنین این ابزار دارای یک موتور هوش مصنوعی جاسازی شده با تجزیه و تحلیل پیش‌بینی است که به شما امکان می‌دهد روندها و الگوهای پنهان در داده‌هایتان را به راحتی دیداری سازی کنید و کشف کنید.



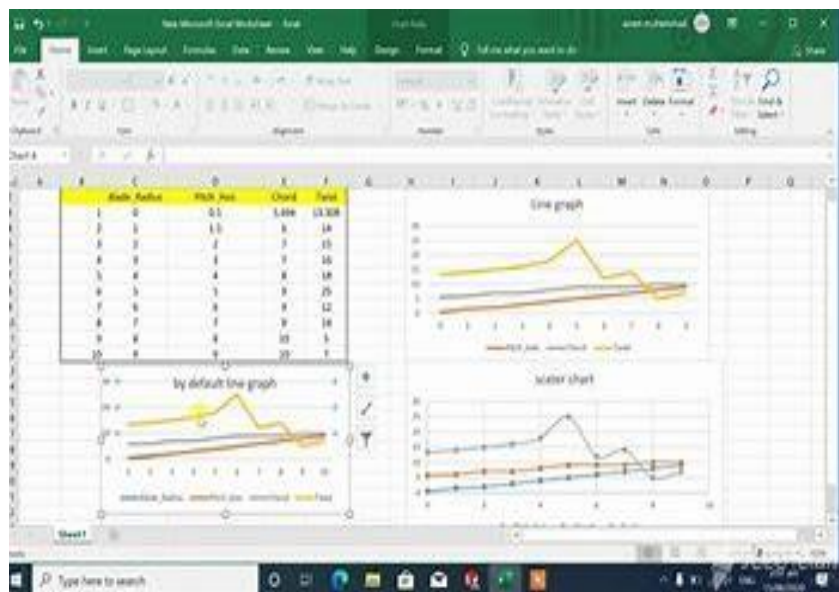
Excel



در اکسل، داده‌ها می‌توانند به چندین روش نشان داده شوند. یکی از این روش‌ها استفاده از نمودارهای پراکندگی است که ارتباط بین دو مجموعه داده را که می‌خواهید مقایسه کنید، نمایش می‌دهد. علاوه بر این، می‌توانید روابط بین متغیرهای مختلف را بررسی کنید تا تصمیم بگیرید که آیا آن‌ها با یکدیگر مرتبط هستند یا خیر.

برای اهدافی مانند تحقیقات بازار یا برنامه‌ریزی مالی، بسیاری از تحلیلگران داده از نمودارهای پراکندگی برای تحلیل داده‌های آماری، علمی، پزشکی و اقتصادی استفاده می‌کنند.

علاوه بر نمودارهای پراکندگی، اکسل مجموعه‌ای وسیع از انواع نمودارها و گزینه‌های بصری‌سازی داده را ارائه می‌دهد، مانند نمودارهای خطی، نمودارهای میله‌ای، نمودارهای دایره‌ای و غیره. این چنین چندوجهی باعث می‌شود اکسل یک انتخاب محبوب برای وظایف تجزیه و تحلیل و بصری‌سازی داده در صنایع و حوزه‌های مختلف باشد.



Zoho Analytics

شما می‌توانید با استفاده از ابزار بصری‌سازی و گزارش‌دهی Zoho Analytics به راحتی گزارشات و داشبوردهای منحصر به فردی تولید کنید.

Zoho Analytics به شما امکان می‌دهد:

- به سرعت گزارشات و داشبوردهای سفارشی را با استفاده از قابلیت کشیدن و رها کردن (drag-and-drop) ایجاد کنید.

- با استفاده از نمودارها و گراف‌های تعاملی، بر روی داده‌های خود تحلیل‌هایی بدست آورید.

- گزارشات و داشبوردها را با همکاران یا مشتریان خود با چند کلیک به اشتراک بگذارید.

- علاوه بر این، Zoho Analytics جزء مجموعه نرم‌افزاری Zoho Office Suite است که شامل Zoho Writer، Zoho Sheet و Zoho Show نیز می‌شود. شما می‌توانید از Zoho Analytics برای گزارش داده‌هایی از هر یک از این برنامه‌ها و همچنین منابع خارجی مانند MySQL، SQL Server، Oracle و Google Sheets استفاده کنید.



ابزارها و تکنیک‌های دیگر در دسترس :

infogram

Infogram

از طریق Infogram نیز می‌توان نمودارها، گزارش‌ها و نقشه‌ها تولید کرد. Infogram یک ابزار محبوب دیگر است که قابلیت تولید اینفوگرافیک را دارد، که به همین دلیل نامش اینفوگرام شده است، و این قابلیت آن را به خصوص در بین کارگران خلاق محبوب کرده است. این برنامه همچنین دارای ویرایشگری با قابلیت کشیدن و رها کردن است که برای مبتدیان مفید است.

برای استفاده آنلاین، تصویربرداری‌ها به عنوان فایل‌های HTML یا تصویر و همچنین به صورت فایل‌های GIF قابل ذخیره سازی هستند تا در گزارش‌ها و سندهای دیگر یکپارچه شوند. Infogram قیمت‌گذاری در سه سطح دارد که از نسخه رایگان تا نسخه سازمانی متغیر است، همانند بیشتر ابزارهای دیگر در این لیست.





Fusion charts

Fusion Charts یک برنامه بصری سازی داده مبتنی بر JavaScript است که بیش از ۱۰۰۰ نوع نقشه و ۱۵۰ نوع نمودار و گراف مختلف را برای نمایش داده ها بکار می برد. این برنامه در عوض از محدود کردن بصری سازی داده به تصاویر ساده خودداری کرده و سعی می کند داشبوردهایی را ارائه دهد که توسط تحلیلگران تجاری مورد ترجیح قرار می گیرد.

با استفاده از قالب های پیش فرض برای ارائه داده ها و تعبیه تصاویر، استفاده از بهترین بصری سازی ها در وب سایت ها آسان می شود. همچنین نسخه های موبایل iOS و Android را نیز بررسی کنید.



D3.js



D3.js (Data-Driven Documents) استفاده از یک کتابخانه جاوااسکریپت داشته و به منظور کار با داده‌ها استفاده می‌شود. یکی از ابزارهای بصری‌سازی داده متن‌باز می‌تواند به طور کامل برای برآورده کردن نیازهای شرکت شما سفارشی شود. با این حال، برای استفاده از این نرم‌افزار و تولید بصری‌سازی‌های تعاملی، باید با زبان‌های برنامه‌نویسی آشنا باشید.

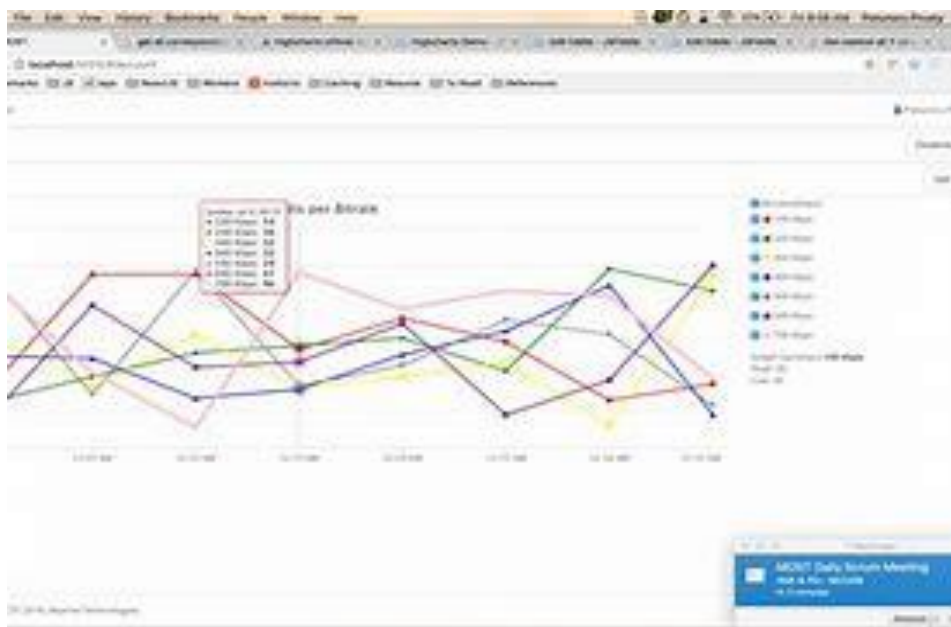


وقتی مجموعه‌های داده بزرگ به اشیاء SVG و توابع نرم‌افزارهای مختلف مرتبط می‌شوند، نمودارهای دایره‌ای، نمودارهای میله‌ای، گراف‌ها و نقشه‌های پویا تولید می‌شوند. تمرکز بر استانداردهای وب و ویژگی‌هایی است که در ساخت و طراحی نقشه‌های پاسخگو و تعاملی کمک می‌کند. این نرم‌افزار تصاویر عالی را تولید می‌کند و کارآمد و قدرتمند است. همچنین با مرورگرهای مختلف سازگار است و استفاده از آن رایگان است.



High charts

Highcharts نیز یکی دیگر از ابزارهای بصری سازی داده و کتابخانه نمودار سازی مبتنی بر جاوااسکریپت است که برای برنامه نویسان و توسعه دهندگان طراحی شده است. این ابزار امکان اضافه کردن فراگرافیک های تعاملی به وبسایت ها را فراهم می کند



برای استفاده از نسخه پرداختی Highcharts، باید یک لایسنس خریداری کنید. مدارس، افراد و سازمان های غیردولتی نیز می توانند از نسخه رایگان برای تعبیه نمودارها در وبسایت خود استفاده کنند. اما این همچنین یکی از محبوب ترین ابزارهای آنلاین برای بصری سازی داده است.



Fine Report

Fine Report یک ابزار سطح سازمانی است که برای استفاده در سطح شرکت ساخته شده است. این ابزار ورود داده و بصری سازی داده را ترکیب می کند و به طور کامل با استفاده از جاوا توسعه یافته است.

این ابزار به راحتی قابل سفارشی سازی است. همچنین، حتی مبتدیان هم به راحتی می توانند از Fine Report استفاده کنند به دلیل شباهت رابط کاربری آن به اکسل. این شرکت نسخه شخصی را به صورت رایگان ارائه می دهد. کسب و کارهای کوچک می توانند با طرح های قیمتی مناسبتر شروع کنند.



مورد استفاده از ابزارهای رایج :

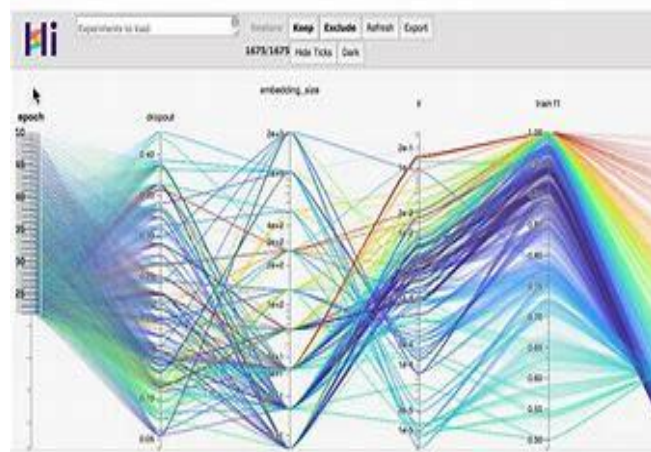
Used tools on Facebook

(ابزارهای مورد استفاده در فیس بوک)



تکنیک‌های بصری‌سازی داده برای کاوش مجموعه‌داده‌ها بسیار مفید هستند. اکوسیستم علم داده از انواع گسترده‌ای از نمودارها استفاده می‌کند. صفات داده و متغیرها تعیین‌کننده آن هستند که کدام نوع برای یک وظیفه خاص مناسبتر است.

در ژانویه، فیسبوک Hi-Plot را منتشر کرد که یک بسته پایتونی است برای امکان بصری‌سازی یک جدول داده بعد بالا. به ویژه، Hi-Plot به خاطر نمودار موازی تعاملی و هوشمندانه‌اش شناخته شده است. Hi-Plot در بصری‌سازی داده‌های بعد بالا تخصص دارد. این بسته قادر است با مجموعه‌داده‌هایی با تعداد زیادی متغیر کار کند و این امر را آسان‌تر می‌کند تا روابط و الگوهای پیچیده در داده را درک کنید. با بصری‌سازی همزمان چندین بعد، Hi-Plot به کاربران کمک می‌کند تا ارتباطات پنهان را کشف کرده و ویژگی‌های مهم را شناسایی کنند. این ابزار انواع نمودارها، طرح‌های رنگ، برچسب‌ها و نشانگرها را ارائه می‌دهد و به کاربران امکان می‌دهد نمودارهای جذاب و مفیدی را ایجاد کنند.



Used tools in Google (ابزارهای مورد استفاده در گوگل)



گوگل ابزار رایگانی به نام گوگل دیتا استودیو ارائه می‌دهد که برای کسانی که تازه با بصری‌سازی داده آشنا هستند بسیار مفید است. شما می‌توانید به سرعت نمودارها، نمودارها و گزارش‌ها را در دیتا استودیو سفارشی کنید و این ابزار گزینه‌های گسترده‌ای برای ساخت گزارش‌های مشترک ارائه می‌دهد که همگی آماده قابل دسترسی هستند.

مزایای استفاده از دیتا استودیو عبارتند از:

اتصالات داخلی به بیش از ۲۰۰ منبع داده که نیاز به برنامه‌ریزی دوره‌ای برای تازه‌سازی داده‌ها در گزارش‌ها را از بین می‌برد.

عدم نیاز به لایسنس و رایگان برای همه.

دانلودهای لازم نیست زیرا همه چیز آنلاین است.



Used tools on Twitter (ابزارهای مورد استفاده در توییتر)



برای دیدن چگونگی ارتباط دو عبارت در یک گروه توییت، می‌توان از تکنیکی به نام نمودار ون توییت استفاده کرد. Tweepsmaps یک ابزار اضافی است که شبکه حساب‌های توییت را تجزیه و نمایش می‌دهد. این ابزار توزیع سهمی براساس درصد دنبال‌کنندگان را در نقشه نمایش می‌دهد. علاوه بر این، این برنامه کاربران غیرفعال که در سه تا شش ماه گذشته توییت نکرده‌اند را لیست می‌کند.

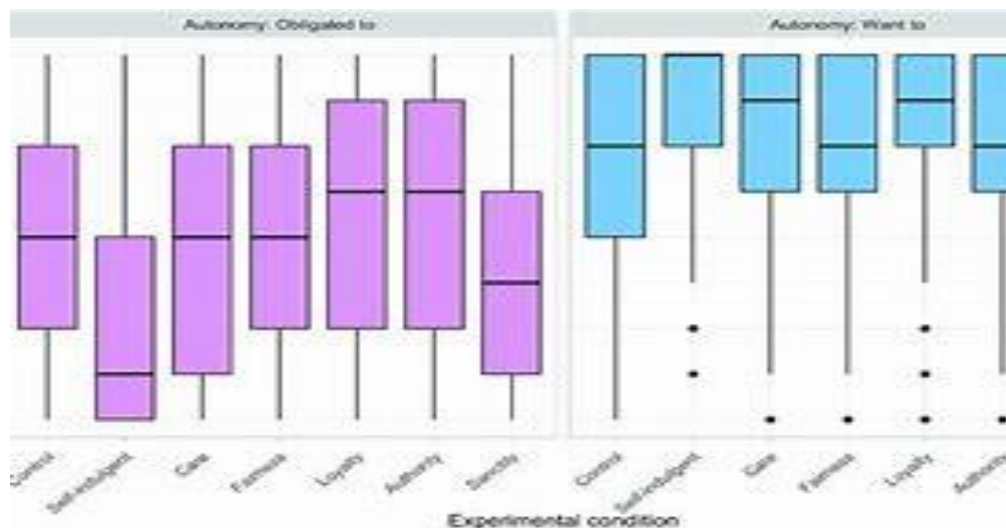
TweepMaps به شما امکان می‌دهد با بصری‌سازی توییت‌ها و فعالیت کاربران روی نقشه، توزیع جغرافیایی داده‌های توییت را بررسی کنید. الگوهای جغرافیایی، نقاط داغ فعالیت و روندهای منطقه‌ای در ارتباط با موضوعات یا رویدادها را کشف کنید.



Autonomy یک پلتفرم قوی است که به تجزیه و تحلیل و بصری‌سازی به روشی منحصربه‌فرد می‌پردازد. قابلیت‌های آن شامل جمع‌آوری اطلاعات جذاب، پیگیری رشد دنبال‌کنندگان، پشتیبان‌گیری/خروجی و تحلیل توییت‌ها و غیره می‌باشد.



AUTONOMY



Follow Me یک سرویس رایگان است که به کاربران امکان می‌دهد اطلاعات کلیدی درباره کاربران توییتر را بررسی کنند. این سرویس یک رابط کاربری کاربرپسند دارد. آن برچسب‌ها، اشاره‌ها و موضوعات را بررسی می‌کند.

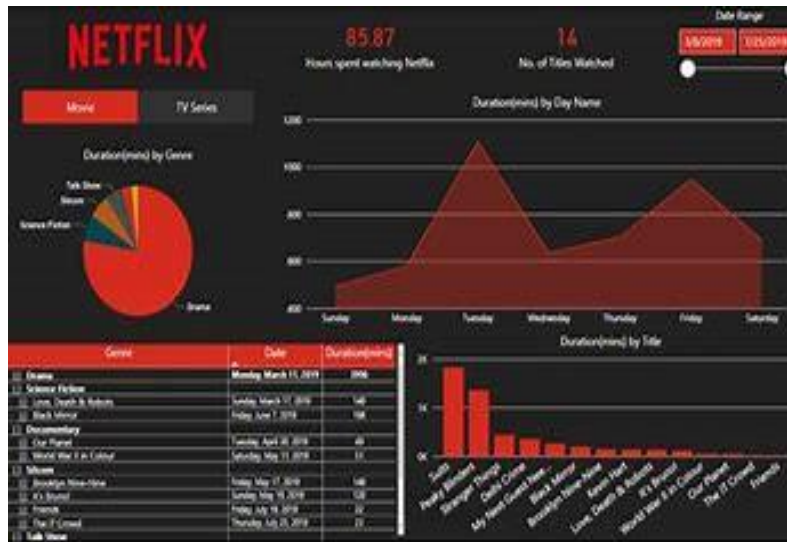


Used tools on Netflix (ابزارهای مورد استفاده در نتفلیکس)



نتفلیکس از DataViz برای پیگیری تجربه تماشای مشتریان خود در یک منطقه جغرافیایی خاص استفاده می‌کند. DataViz اطلاعات برنامه‌های تماشایی به صورت لحظه‌ای را ارائه می‌دهد. این ابزار طرح‌بندی پویاست که نمایی از عملکرد یک برنامه خاص، مشکلات مرتبط با اتصال در صورت وجود، فراوانی ورود کاربران و موارد مشابه را نشان می‌دهد.

با تحلیل ترجیحات مخاطبان، نتفلیکس می‌تواند برنامه‌های تلویزیونی را با توجه به سلیقه تماشاگران ایجاد کند. به عنوان مثال، نتفلیکس از DataViz برای تحلیل مخاطبان برنامه‌های بریتانیایی استفاده کرد و به نتیجه رسید که تماشاگران از فیلم‌های کارگردانی دیوید فینچر لذت می‌برند. این شناخت مفید به نتفلیکس امکان می‌دهد تا سری‌های جدیدی را که توسط دیوید فینچر کارگردانی می‌شوند، بازتولید کند.



Used tools by Intel (ابزارهای مورد استفاده اینتل)



این صحیح است که اینتل از ابزارهای دیداری سازی داده تولید شده توسط کنجویا اینک استفاده می‌کند (که بعدها توسط یوتیمیت سافتویر، یک شرکت فناوری چندملیتی آمریکایی که به مشتریان خود راهکارهای مبتنی بر هوش مصنوعی در حوزه HCM و مدیریت نیروی کار ارائه می‌دهد، به خریداری شد). اینتل از این ابزارها برای تجزیه و تحلیل درونی احساسات کارکنان خود استفاده می‌کند. تجزیه و تحلیل احساسات به معنای شناسایی روندها و تشخیص مثبت یا منفی بودن داده‌ها است. کنجویا از یک الگوریتم یادگیری ماشین برای تفسیر احساسات کارکنان استفاده می‌کند.

در اینتل، شرکت تجزیه و تحلیل پست‌های رسانه‌های اجتماعی کارکنان خود را بررسی می‌کند تا الگوها را شناسایی کرده و نظریه‌های خود را درباره رضایت یا نارضایتی کارکنان از فعالیت‌های کاری یا محیط کار بدست آورد. احساسات می‌توانند مردد کننده باشند. بنابراین، تجزیه و تحلیل احساسات کمک می‌کند تا به دقت بفهمیم افراد چگونه احساس می‌کنند و به آن رسیدگی شود.

Used tools by retails companies

(ابزارهای مورد استفاده شرکت های خرده فروشی)



شرکت های خرده فروشی از تحلیل سبد خرید بازار برای تعیین رفتار خرید مشتریان خود استفاده می کنند. به جای وابستگی به مشتریان جدید، شرکت های خرده فروشی سعی می کنند ترجیحات مشتریان موجود را بدانند. سوابق خرید مشتریان در طول یک بازه زمانی (پایگاه داده تاریخیچه خرید) بررسی می شود. با کشف رفتار خرید مشتری، شرکت های خرده فروشی می توانند محصولات ترجیحی را در بالای قفسه قرار داده و در این راه آن را برای همه مشتریان قابل مشاهده کنند. تحلیل سبد خرید بازار دارای پشتوانه ریاضی است. به همین دلیل، محصولات قابل خوردن و غیرقابل خوردن در یک قسمت از فروشگاه خرده فروشی به هم نمی ریزند.

یک نمونه از این کار از طرف آمازون انجام می شود. در بیشتر موارد، بعد از انجام یک خرید، محصولات مشابه با عبارت "مشتریان این محصولات را همزمان خریداری می کنند" پیشنهاد می شوند. احتمال اینکه مشتری بلافاصله خرید محصول پیشنهادی را انجام دهد بالاست و این باعث افزایش فروش می شود. برای تحلیل داده ها از هادوپ استفاده می شود و داده ها به صورت رسیده ها جمع آوری می شوند.