Ingin memvisualisasikan data tips, dimana dataset ini berasal dari library seaborn. Library seaborn salah satu library yang dipakai ketika ingin memvisualisasikan suatu data. Data tips mengandung variabel time, sex, smoker, tip, total\_bill, and day. Visualisasi pertama yang saya lakukan yaitu,

Sebelumnya cara memanggil data adalah dengan code berikut ini

tips**=**sns.load\_dataset('tips')

untuk melihat dataset yang ada di library seaborn dapat running code berikut,

sns.get\_dataset\_names()

**Scatterplot**

Masuk kedalam kelompok chart, dimana scatterplot digunakan untuk melihat hubungan positif atau negatif dari suatu data, korelasi suatu data kuat atau lemah, dan linear dan nonlinear suatu data. Pada visualisasi scatterplot saya menggunakan library seaborn, berikut adalah code nya.

Ingin melihat hubungan antara total bill dengan tip yang diperoleh.

scatterplot**=**sns.scatterplot(**data=**tips, **x=**"total\_bill", **y=**"tip", **hue=**"time").set(**title=**'Tips by Time')

outputnya sebagai berikut,

A graph with blue and orange dots

Description automatically generated with low confidence

Dari visualisasi tersebut diperoleh informasi bahwa…

Kemudian saya ingin melihat kelompok data berdasarkan time yaitu dinner dan lunch, berikut code nya.

sns.relplot(**data=**tips, **x=**'total\_bill', **y=**'tip', **col=**'time', **hue=**'day', **style=**'day', **kind=**'scatter').set(**title=**'Tips by Day and divide by Time ')

outputnya berikut ini,

A picture containing screenshot, diagram, text, line

Description automatically generated

Interpretasi…

Selanjutnya visualisasi kedua yaitu bar chart. Bar chart digunakan untuk membandingkan data kategorik satu dengan kategorik yang lainnya dan digunakan untuk data kategorik.

Saya ingin melihat perbandingan time dinner dan lunch berikut code nya.

plt.barh(tips['time'],tips['total\_bill'],**color=**'orange')

plt.ylabel('Time') *#kategorik data untuk yticks*

plt.xlabel('Total bill ($)')

plt.title('Total bill by Time')

plt.show()

outputnya

A picture containing text, screenshot, rectangle, line

Description automatically generated

Berikutnya visualisasi data vertikal bar chart. Saya ingin membandingkan proporsi laki-laki dan perempuan berdasarkan jumlah tip. Di bawah ini adalah codingnya untuk memilih sex dan menjumlahkan tip berdasarkan sex dinamai dengan objek df2.

print('------- Select Data -------')

df2**=**tips.groupby('sex', **as\_index=**False)['tip'].sum()

print(df2)

output df2

A screenshot of a computer

Description automatically generated with medium confidence

setelah itu definisikan data sumbu x dan y nya sebagai berikut.

x**=**df2['sex']

y**=**df2['tip']

print('------- Bar Chart -------')

plt.bar(x,y,**color=**('red','blue'))

plt.xlabel**=**('Jenis Kelamin')

plt.ylabel**=**('Total Tip')

plt.title('Jumlah Tip Berdasarkan Jenis Kelamin')

diperoleh output bar chart di bawah ini.

A red and blue rectangles

Description automatically generated with low confidence

Masih visualisasi dengan bar chart namun stacked bar chart, saya ingin melihat perbandingan jumlah pengunjung laki-laki dan perempuan berdasarkan hari (day). Pertama select total bill dan jumlahkan berdasarkan hari dan sex seperti di bawah ini.

*# long format datframe*

df3**=**tips.groupby(['day','sex'],**as\_index=**False)['total\_bill'].sum()

print(df3)

output

A screenshot of a computer screen

Description automatically generated with low confidence

kemudian diubah menjadi wide forma dataframe.

newdf3**=**pd.pivot\_table(**data=**df3, **index=**['day'],

**columns=**['sex'],**values=**'total\_bill')

print(newdf3)

output

A screenshot of a computer screen

Description automatically generated with low confidence

Setelah data sudah di select dan dibentuk sesuai format, maka selanjutnya code untuk visualisasi.

print('------ Stacked Bar Chart ------')

sb**=**newdf3.plot.bar(**stacked=**True, **color=**['blue','red'])

sb.set\_title('Total bill')

sb.set\_xlabel('Hari')

sb.set\_ylabel('Bill ($)')

sb.set\_xticklabels(['Thursday','Friday','Saturday','Sunday'], **rotation=**0)

Output visualiasi

A picture containing text, screenshot, diagram, plot

Description automatically generated

Visualisasi selanjutnya adalah boxplot. Boxplot digunakan untuk melihat keragaman dan karakteristik distribusi suatu data, melihat kesimetrisan data (dapat dilihat dari median dan whisker), melihat penyebaran data, dan ketika ingin mengidentifikasi nilai rata-rata, sebaran data, dan skewness data dapat menggunakan visualisasi Boxplot.

sns.boxplot(**x=**tips['day'], **y=**tips['tip'],**color=**'green').set(**title=**'Tip Berdasarkan Hari')

output visualisasi

A picture containing diagram, plan, rectangle, technical drawing

Description automatically generated

Pie chart

rokok**=**tips.groupby(['smoker','sex'], **as\_index=**False).size()

display(rokok)

# delete baris dengan value tertentu

rokok2**=**rokok.drop(**labels=**[2,3], **axis=**0)

output visualisasi

A picture containing text, screenshot, diagram, circle

Description automatically generated