**Data Visualization**

**Dataset 1: (Titanic)**

I am using the “Titanic”.

I have tried to visualize the dataset to know the number of male and female that were present when the accident occurred, then I plotted the fares comparison.

The dataset has all the passenger details, their names, ages, the fares they paid, etc.

**Pie Chart:**

Pie charts are often used **in business**. Examples include showing percentages of types of customers, percentage of revenue from different products, and profits from different countries. Pie charts can be helpful for showing the relationship of parts to the whole when there are a small number of levels.

The pie chart is useless when there are many pieces of data that is variables, because when we plot many variables using pie chart, the chart will not be so clear, it will become crowded. Pie chart is useful if there is no crowded data.I want to plot male and female count, so the best suit here is Pie chart, because it will give me clear results.

df = pd.read\_csv('titanic.csv')

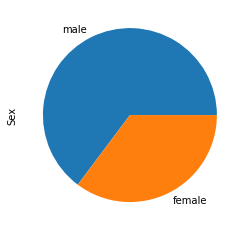
df.set\_index('PassengerId', inplace=True)

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

df.Sex.value\_counts().plot(kind = 'pie')



As you can see, the male and female comparison is shown beautifully. Any one can now easily understand that how many were male and female.

**Scatter Plot:**

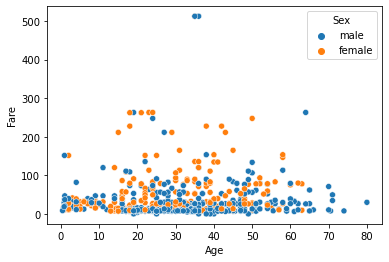
I am not happy with the results, as the above chart only show the number of male and female, our dataset also has the detail of fares, is it possible to plot something which will give me the result that how much fare each gender had paid? Like how much fare did males paid and same for female.

The Answer is Scatter plot, why? Because this plot gives the best relationship between two or more variables, now as the fare is related to the passenger, so scatter is the best option.

Normally, scatter plots are used to know how much each variable are related to each other.

Add this line to the above given code.

sns.scatterplot(x = 'Age', y = 'Fare', hue = 'Sex', data = df)



And here we are, a beautiful scatter plot is being plotted which show how much fares are paid by male and female.

For Example, the highest fares are paid by two males, that is above 500 and the age of those two was between 30 to 40.

This is the beauty of scatter plots.

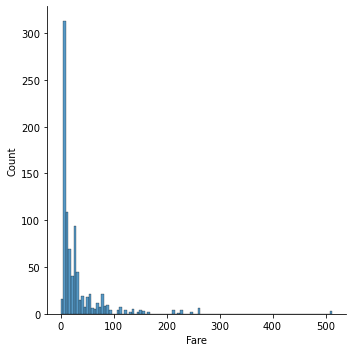
**Bar Graph:**

As you can see in the above graph, 90 percent (just an estimation) of the people paid low fares, this means that they didn’t pay for luxurious rooms. Now I want to confirm the results by using the Bar graph, the graph will show me as a whole result.

Bar graphs are used to compare different data and to track changes over time.

Just add the below line

sns.displot(df.Fare)



And here it is, the bar graph showed me that only some paid 500, as the scatter plot showed me that only two males paid 500, scatter plots are more detailed than bar charts. The 90 percent estimation is almost correct.

Using the above charts and graphs, we have now a detailed information shown graphically, now if you give this charts and graph to anyone, they will immediately understand the deep-down insight of the data. The Purpose of data visualization is to being the data into graphical form.

**Dataset 1: (Advertising)**

I have another dataset called advertising, the dataset contains TV, Newspaper and Radio advertising data and their increase in sales by paying for advertisement is given.

I want to know that how much sales are increased by paying for advertisement.

I will use Scatter plot again because it will give me the best comparison.

**Pair Plot:**

The pairplot function creates a grid of Axes such that each variable in data will by shared in the y-axis across a single row and in the x-axis across a single column.

I want to compare the advertisement of TV, Radio and Newspaper with the sales, so pare plot will help me to do that.

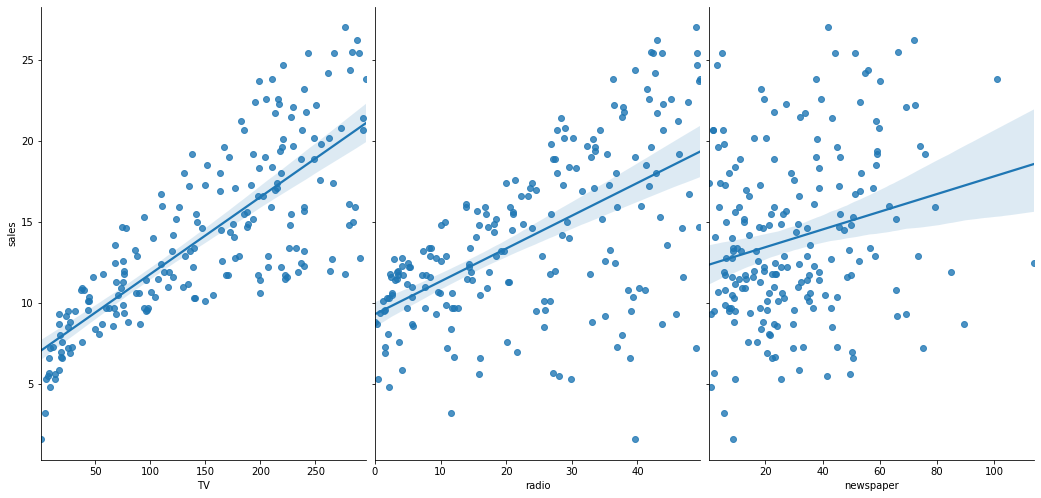
import pandas as pd

import seaborn as sns

%matplotlib inline

df = pd.read\_csv('Advertising.csv')

sns.pairplot(df,x\_vars = ['TV','radio','newspaper'], y\_vars = 'sales', size = 7, aspect = 0.7, kind = 'reg')



As you can see that the pair plot has compared sales with the advertisement of TV, Radio and Newspaper.

I can see that the higher they paid for advertisement, their sales are going higher.

**Dataset 1: (Financial Sample)**

I have another dataset called Financial Sample in which products details are given, the products are made by government and private companies, the data is collected of different countries, Data features are date, sale price, profit, etc.

Now I want to know the profit made by a single product over time, I will take a random product and the country where it is actually made up and a segment, means the product is manufactured by government or a private company.

**Line Graph:**

The usage of Line Graph is actually to know the trends, means to compare the rate of change. In our problem, the Line Graph is the best suit, because we want to see the profit made by a product over time.

import pandas as pd

import matplotlib.pyplot as plt

excel\_file\_path = "Financial Sample.xlsx"

df = pd.read\_excel(excel\_file\_path)

df.head()

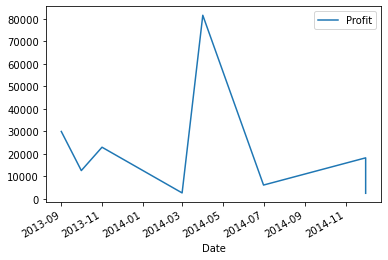
# Time series

df\_vtt\_canada = df.loc[(df['Country'] == 'Canada') & (df['Product'] == 'VTT') & (df['Segmentent'] == 'Government')]

df\_vtt\_canada = df\_vtt\_canada.sort\_values(by=['Date'])

df\_vtt\_canada.plot(x='Date', y='Profit')

plt.show()



Look at the Line Graph, I selected a product called VTT, segmentate by the government of Canada and tried to plot the profit made by that product over time, and look at the results, the product was touching the sky between 2014-03 to 2014-07. The product suddenly hit the ground after 2014-11.

**Pie Chart:**

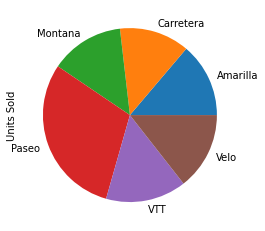
There are total of six products in the dataset, I want to know that which product is sold more then the other, well for this comparison, Pie chart is the best option.

Just add the below code in the code I have given above.

df\_products = df.groupby(['Product']).sum()

df\_products['Units Sold'].plot.pie()

plt.show()



And here it is, the chart gave me the best comparison of all six products and their sales, from the chart, I can that the most sold item is called Paseo and the least one is Carretera.

This is the Power of Data Visualization, to bring out the hidden pattern from the data and show it in a pictorial form, It is said that a Picture worth thousands words, so human can understand picture more easily than giving them dataset to check the row and columns.

Thank you.