Exploratory Data Analysis Data Axis (x,y,z)

referensi:
https://www.kaggle.com/code/estevanamazonas/agressive-behaviour-eda-modeling-finetuning
https://www.kaggle.com/code/pratul007/deep-learning-with-high-graphics
https://www.kaggle.com/code/pratul007/deep-learning-with-high-graphics

Import Packages

In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pylab as plt
import plotly.express as px

import seaborn as sns
from mpl_toolkits.mplot3d import Axes3D
import matplotlib.pyplot as plt
from sklearn.utils import shuffle
from sklearn.preprocessing import LabelEncoder, StandardScaler
from sklearn.model_selection import train_test_split
```

In [2]:
```
# Read CSV
df = pd.read_csv('dataset/merge_data.csv')
```

In [3]:
```
train_data, test_data = train_test_split(df, test_size=0.2, random_state=42)
```

In [4]:
```
train_data
```

Out[4]:

	timestamp	x	y	z	label
11034	2022-07-06 07:54:15.200	-1.0	242.0	43.0	1.0
12555	2022-07-06 07:55:16.040	0.0	247.0	42.0	0.0
21861	2022-07-06 07:49:11.960	1.0	246.0	40.0	0.0
12157	2022-07-06 07:55:00.120	0.0	247.0	42.0	0.0
3737	2022-07-06 07:44:23.800	1.0	245.0	39.0	0.0
...	...	...	...	...	...
21575	2022-07-06 07:49:00.520	1.0	246.0	42.0	0.0
5390	2022-07-06 07:56:02.240	0.0	248.0	50.0	1.0
860	2022-07-06 07:42:28.720	1.0	250.0	34.0	1.0
15795	2022-07-06 07:51:15.400	0.0	245.0	43.0	1.0
23654	2022-07-06 07:46:01.600	11.0	262.0	49.0	1.0

22713 rows × 5 columns

In [5]:
```
train_data.shape, test_data.shape
```

Out[5]:
```
((22713, 5), (5679, 5))
```

Analisis Statistical

- rentang nilai dari sumbu x, -97 sampai 394
- rentang nilai dari sumbu y, 169 sampai 394
- rentang nilai dari sumbu y, -48 sampai 352

seharusnya mencari rata-rata, median, std tidak diperbolehkan terdapat nilai yang negatif. Nilai negatif ini adalah posisi letak sumbu x,y,z yang memungkinkan nilai menjadi negatif. Maka, akan dilakukan perhitungan ulang dengan menambahkan absolute

In [6]:
```
# Calculate the mean, median, and standard deviation for each axis
mean_x = np.mean(abs(train_data['x']))
mean_y = np.mean(abs(train_data['y']))
mean_z = np.mean(abs(train_data['z']))

median_x = np.median(abs(train_data['x']))
median_y = np.median(abs(train_data['y']))
median_z = np.median(abs(train_data['z']))

std_x = np.std(abs(train_data['x']))
std_y = np.std(abs(train_data['y']))
std_z = np.std(abs(train_data['z']))
```

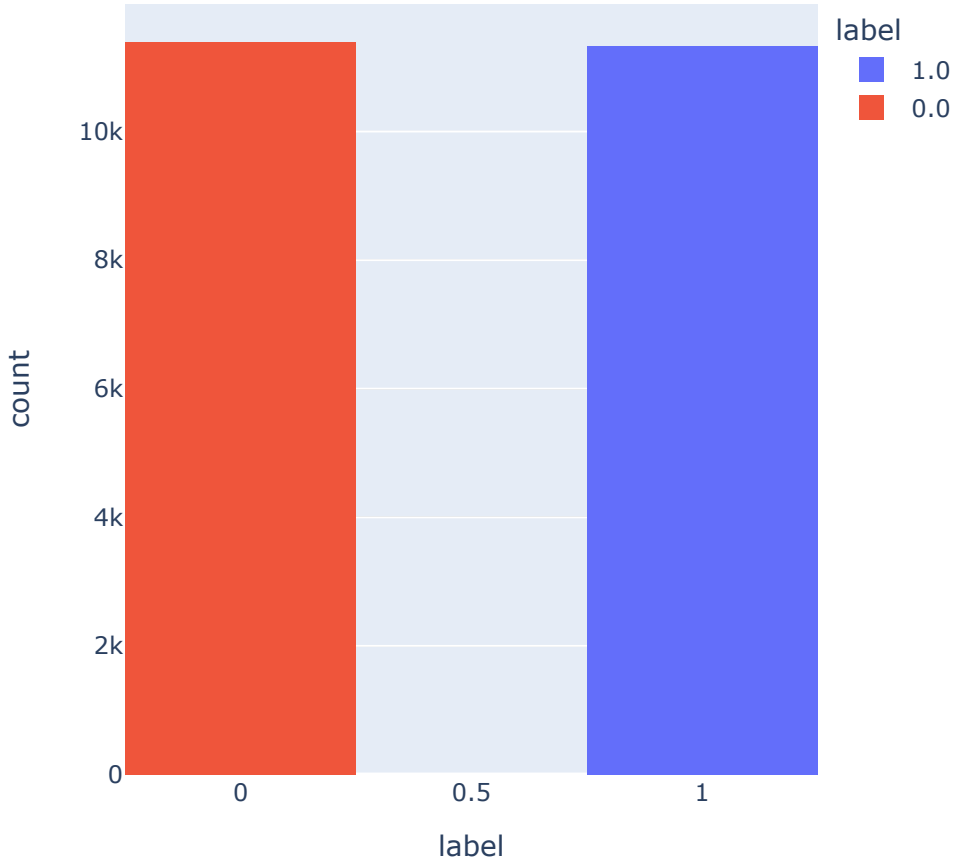
In [7]:
```
# Print the results
print('Axis X: Mean = {:.2f}, Median = {:.2f}, Standard Deviation = {:.2f}'.format(me
print('Axis Y: Mean = {:.2f}, Median = {:.2f}, Standard Deviation = {:.2f}'.format(me
print('Axis Z: Mean = {:.2f}, Median = {:.2f}, Standard Deviation = {:.2f}'.format(me
```

```
Axis X: Mean = 3.87, Median = 2.00, Standard Deviation = 5.84
Axis Y: Mean = 245.60, Median = 246.00, Standard Deviation = 8.22
Axis Z: Mean = 41.91, Median = 42.00, Standard Deviation = 7.89
```

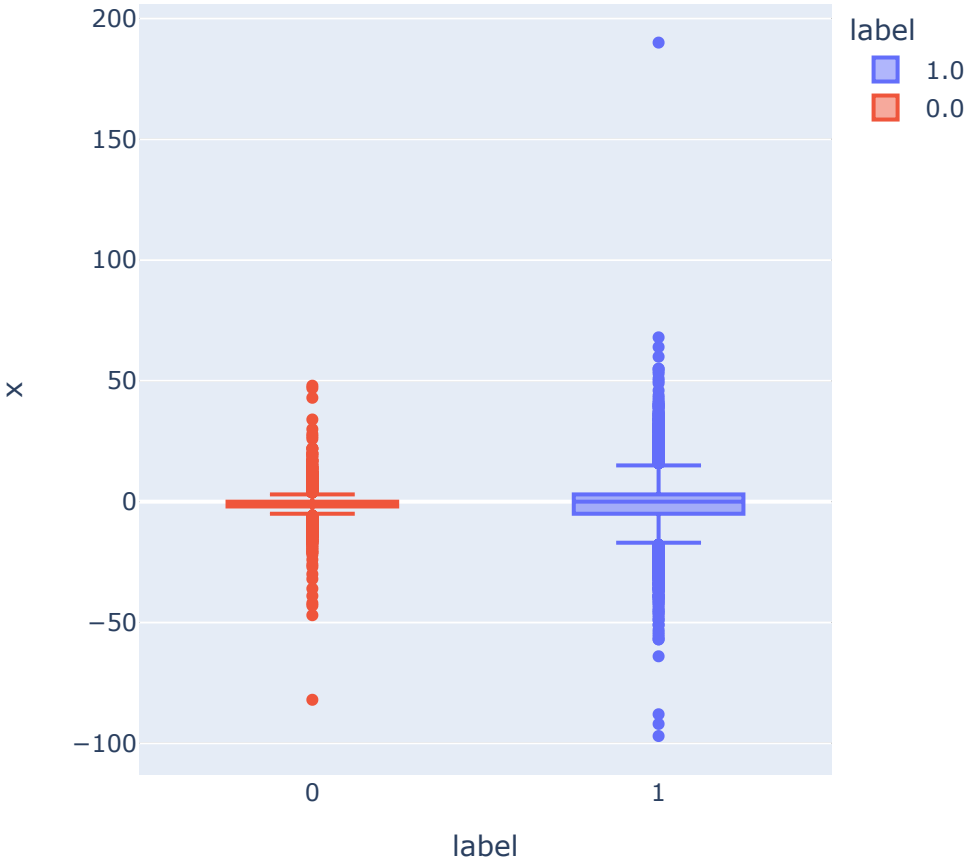
Visualize the Data

Visualize with Histogram and Boxplot

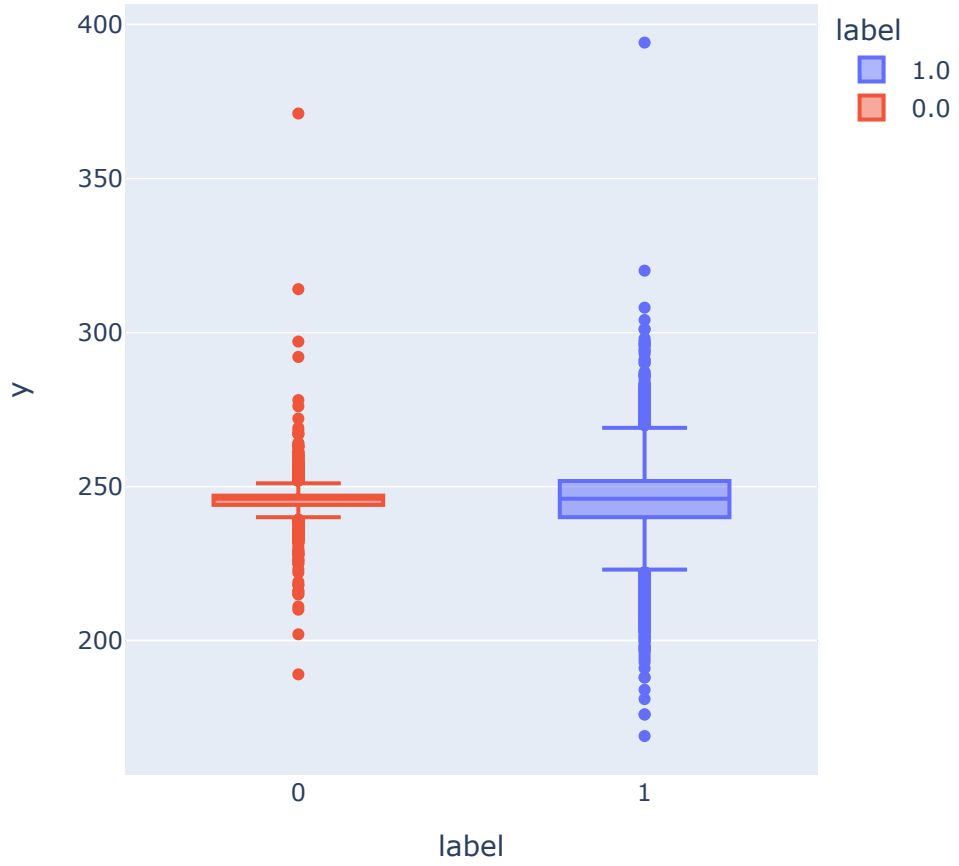
In [8]:
```
# bar plot
fig = px.histogram(train_data, x='label', nbins=3, color='label')
fig.show()
```



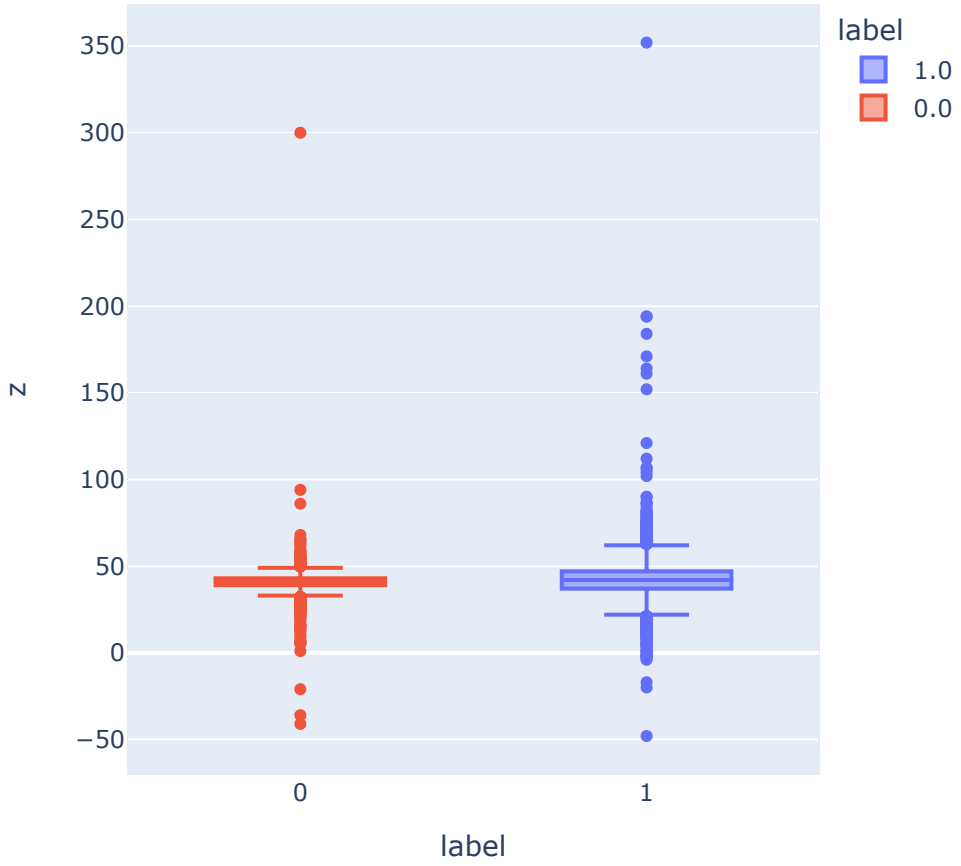
In [9]:
```
# box plots
fig = px.box(train_data, x='label', y='x', color='label')
fig.show()
```



In [10]:
```
fig = px.box(train_data, x='label', y='y', color='label')
fig.show()
```

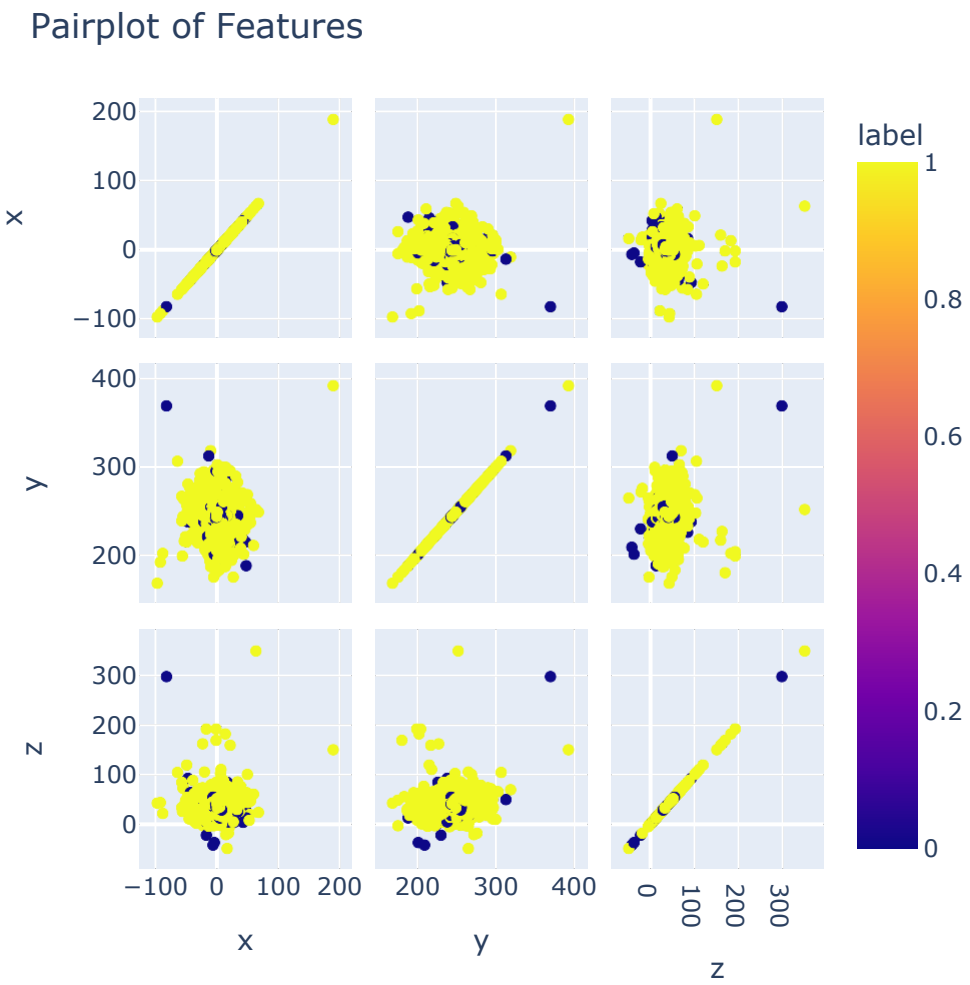


In [11]:
```
fig = px.box(train_data, x='label', y='z', color='label')
fig.show()
```



Visualize Pairplot and Correlation Heatmap

In [12]:
```
fig = px.scatter_matrix(train_data, dimensions=['x', 'y', 'z'], color='label')
fig.update_layout(title='Pairplot of Features')
fig.show()
```



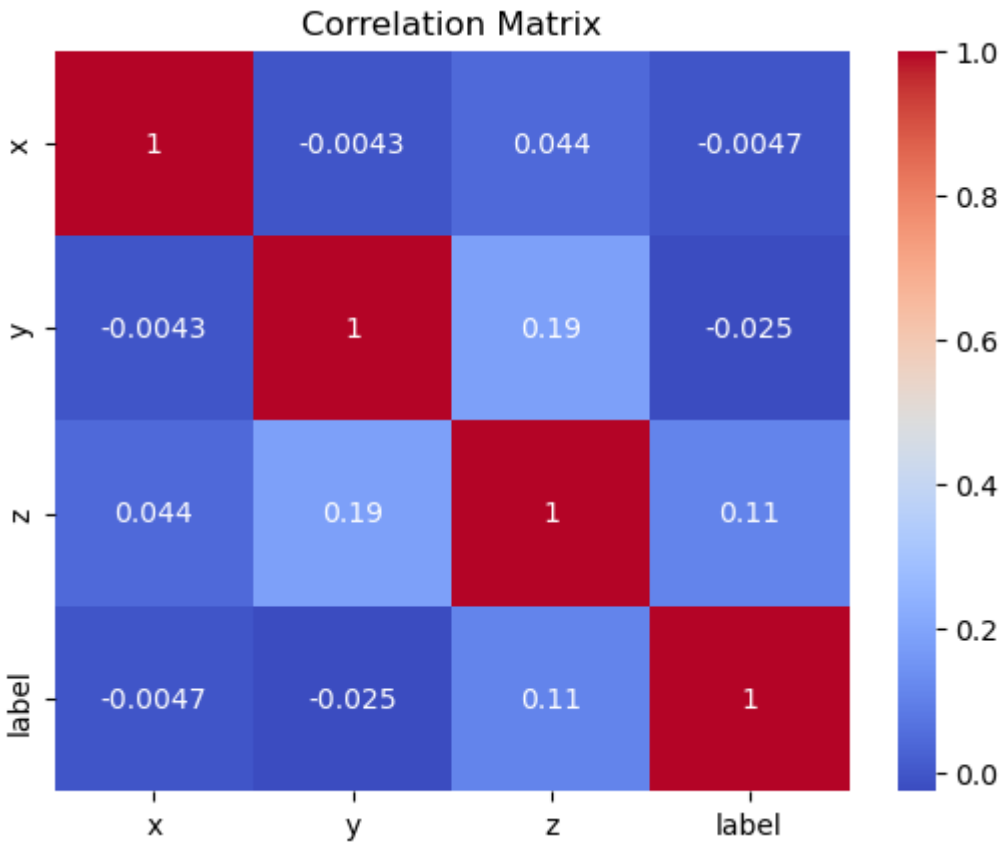
How to read Pairplot

- Korelasi positif: Ketika scatterplot menunjukkan tren naik, itu berarti bahwa ketika satu variabel meningkat, demikian juga variabel lainnya.
- Korelasi negatif: Ketika scatterplot menunjukkan tren menurun, itu berarti bahwa ketika satu variabel meningkat, variabel lainnya menurun.
- Tidak ada korelasi: Ketika scatterplot tidak menunjukkan tren yang jelas, berarti kedua variabel tersebut tidak terkait.

Maka dalam visualisasi pairplot diatas, bisa dikatakan bahwa setiap variable Axis x,y,z tidak mempunya korelasi antar variabel

In [13]:
```
corr_matrix = train_data[['timestamp','x', 'y', 'z', 'label']].corr()

# Print correlation matrix
sns.heatmap(corr_matrix, annot=True, cmap='coolwarm')
plt.title('Correlation Matrix')
plt.show()
```



Correlation Heatmap menunjukkan korelasi antara semua pasangan variabel dalam dataset. Warna yang lebih terang menunjukkan korelasi positif, sedangkan warna yang lebih gelap menunjukkan korelasi negatif. Kita dapat membaca nilai korelasi dari matriks warna di heatmap. Semakin dekat nilai korelasi dengan 1, semakin kuat hubungan antara dua variabel. **How to read Correlation Matrix:**

- Korelasi positif: Ketika warna lebih dekat ke merah, itu berarti bahwa satu variabel meningkat, demikian juga variabel lainnya. Semakin gelap warna merah, semakin kuat korelasinya.
- Korelasi negatif: Ketika warnanya lebih dekat ke biru, itu berarti bahwa ketika satu variabel meningkat, variabel lainnya menurun. Semakin gelap warna biru, semakin kuat korelasinya.
- Tidak ada korelasi: Bila warnanya mendekati putih, berarti kedua variabel tidak berhubungan.

In [ ]: