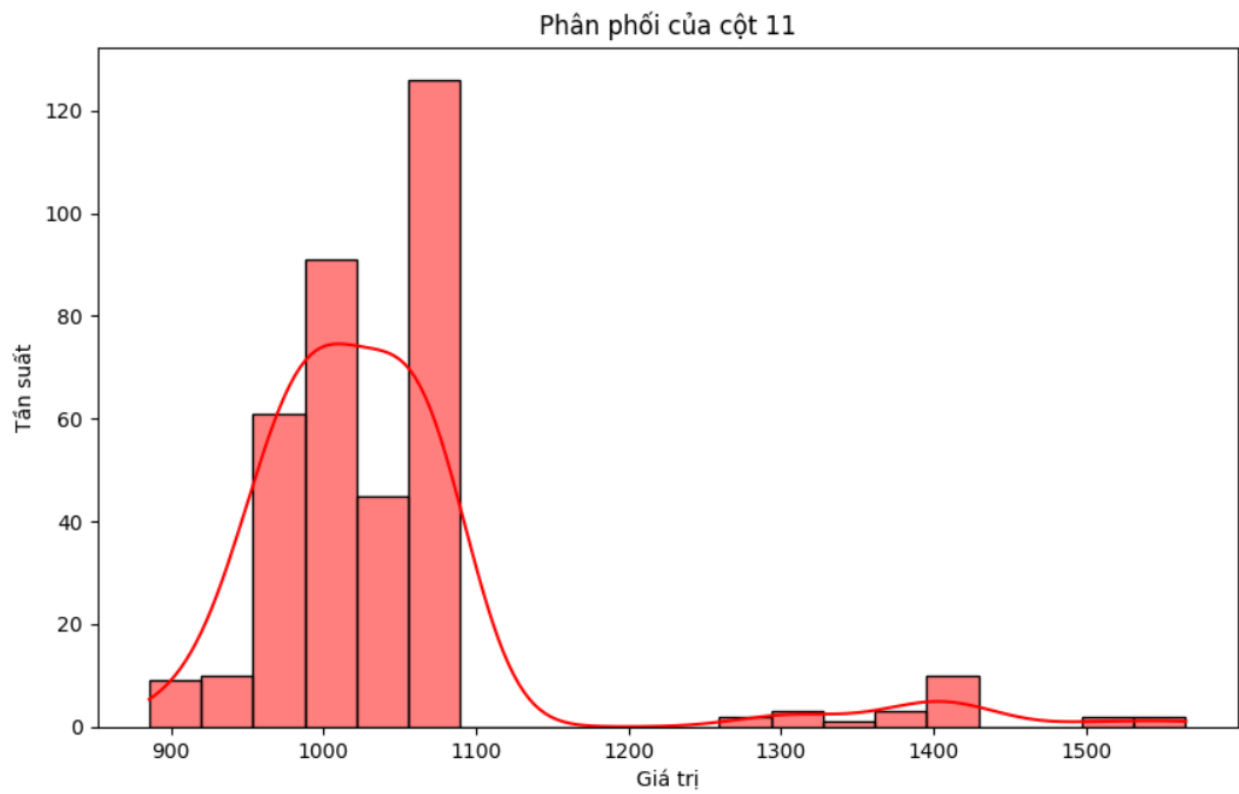
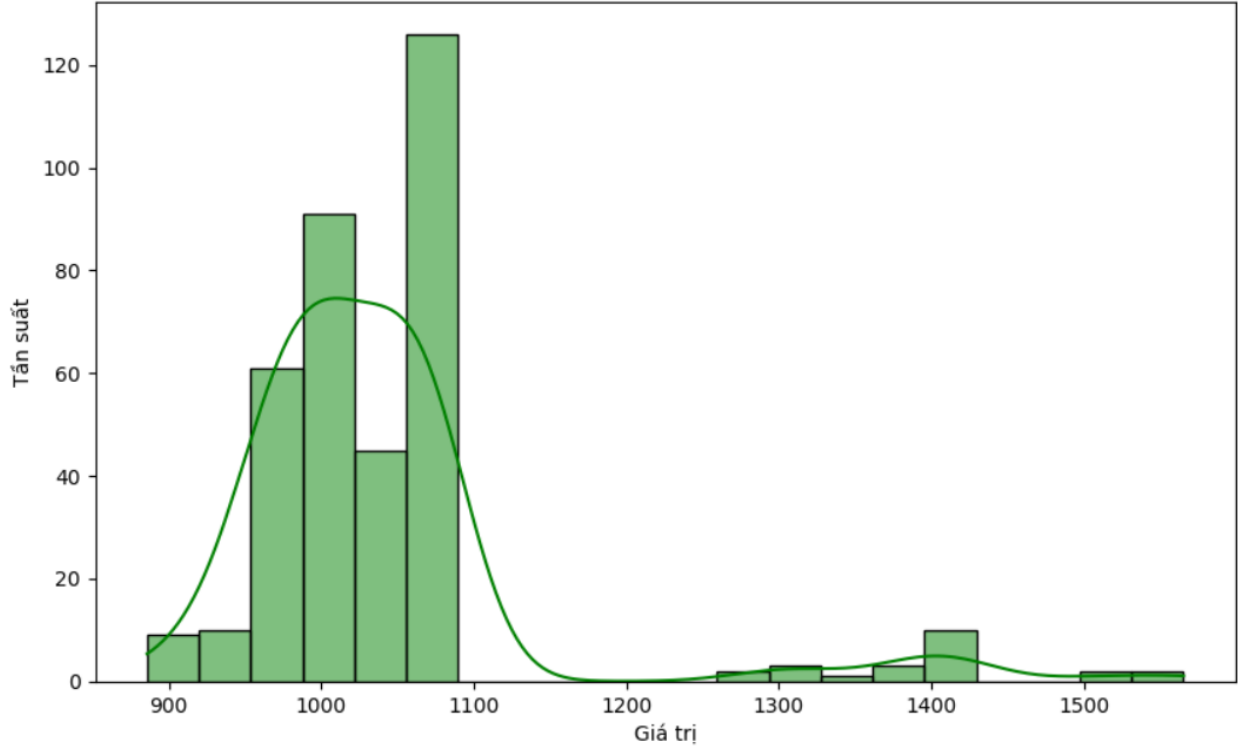


# EDA

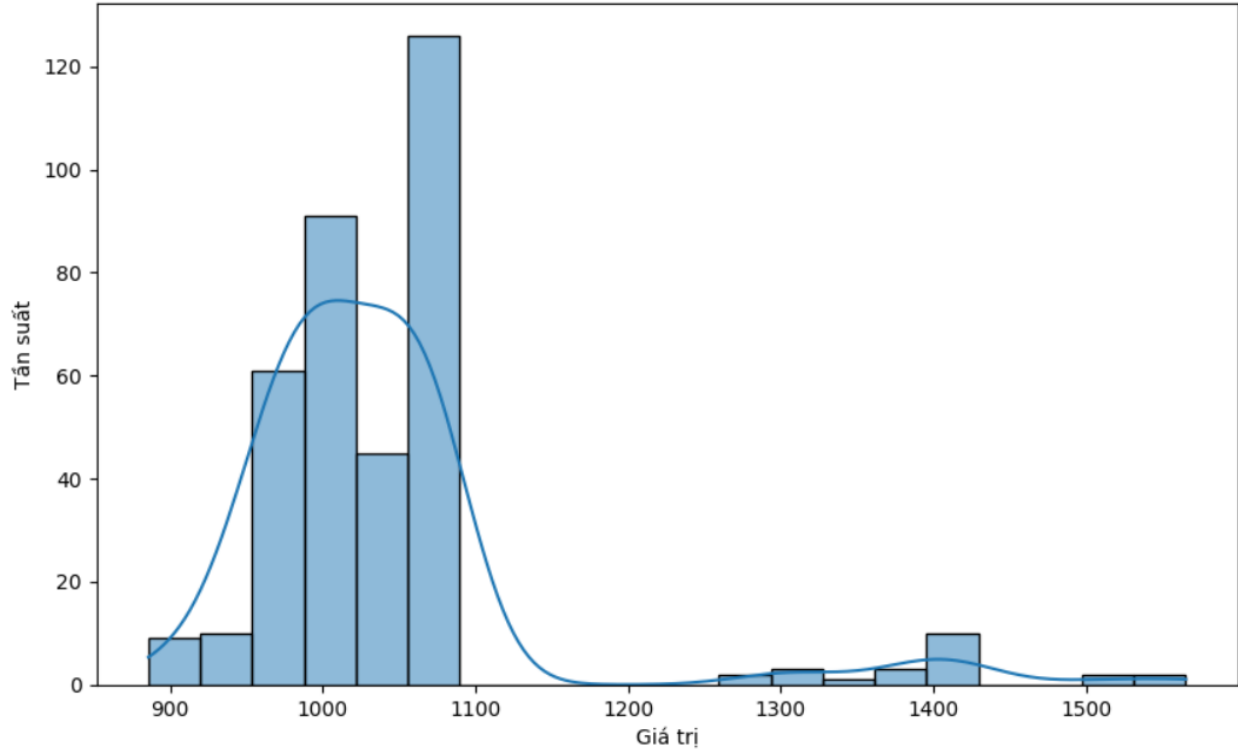
## 1. Biểu đồ phân phối



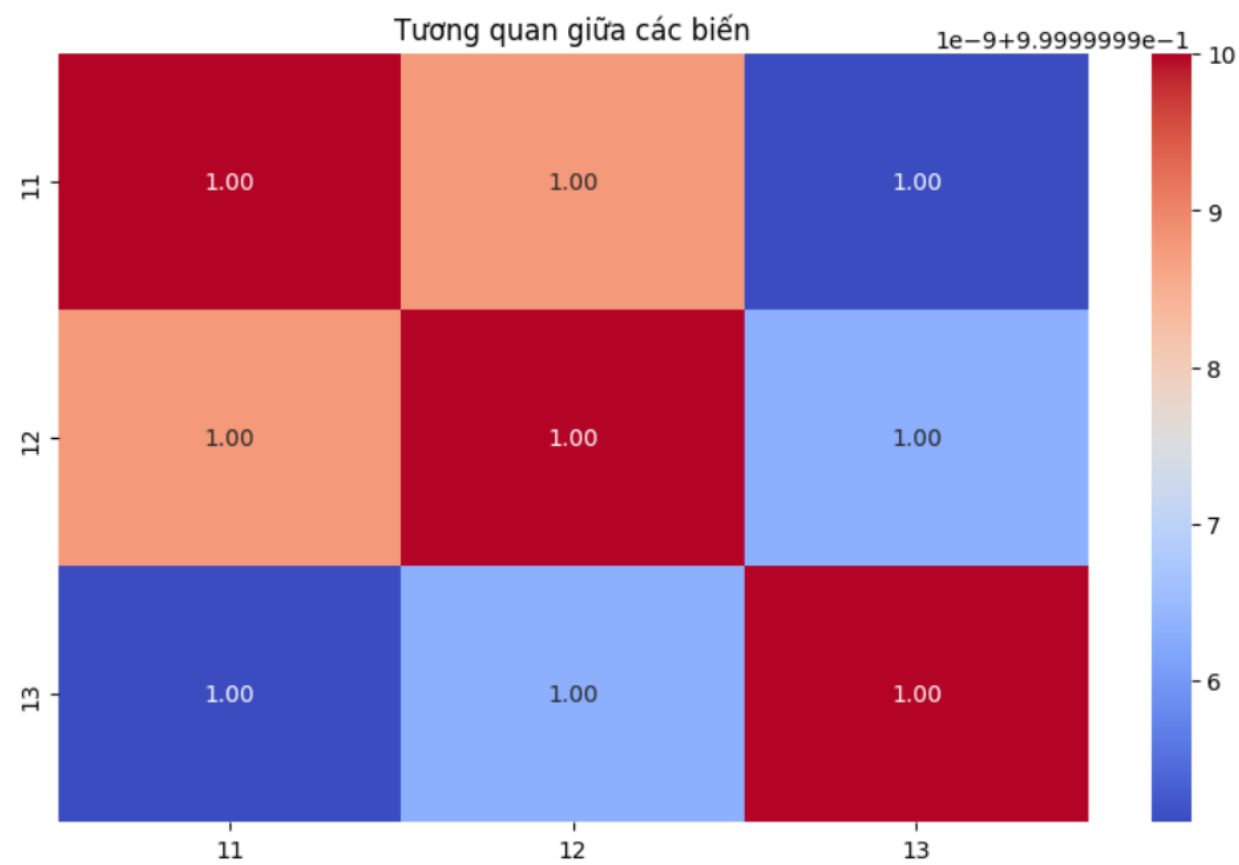
Phân phối của cột 12



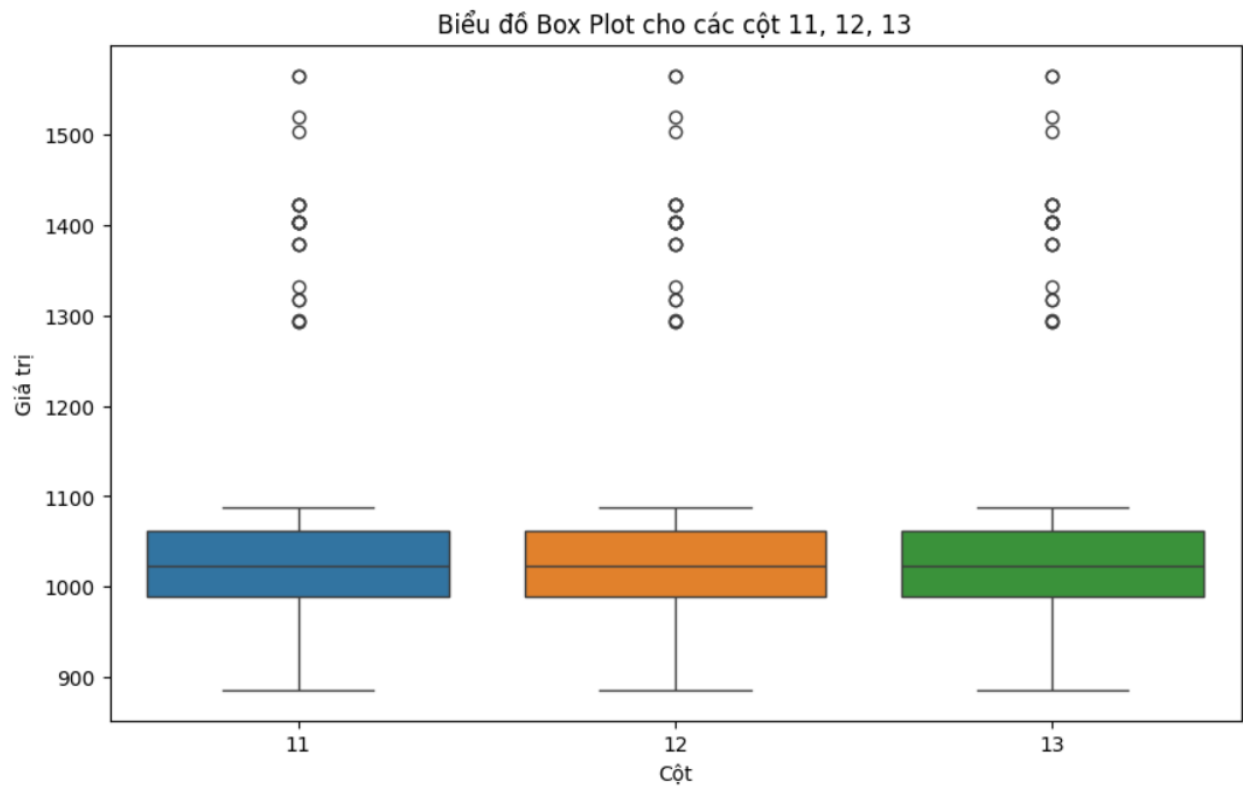
Phân phối của cột 13



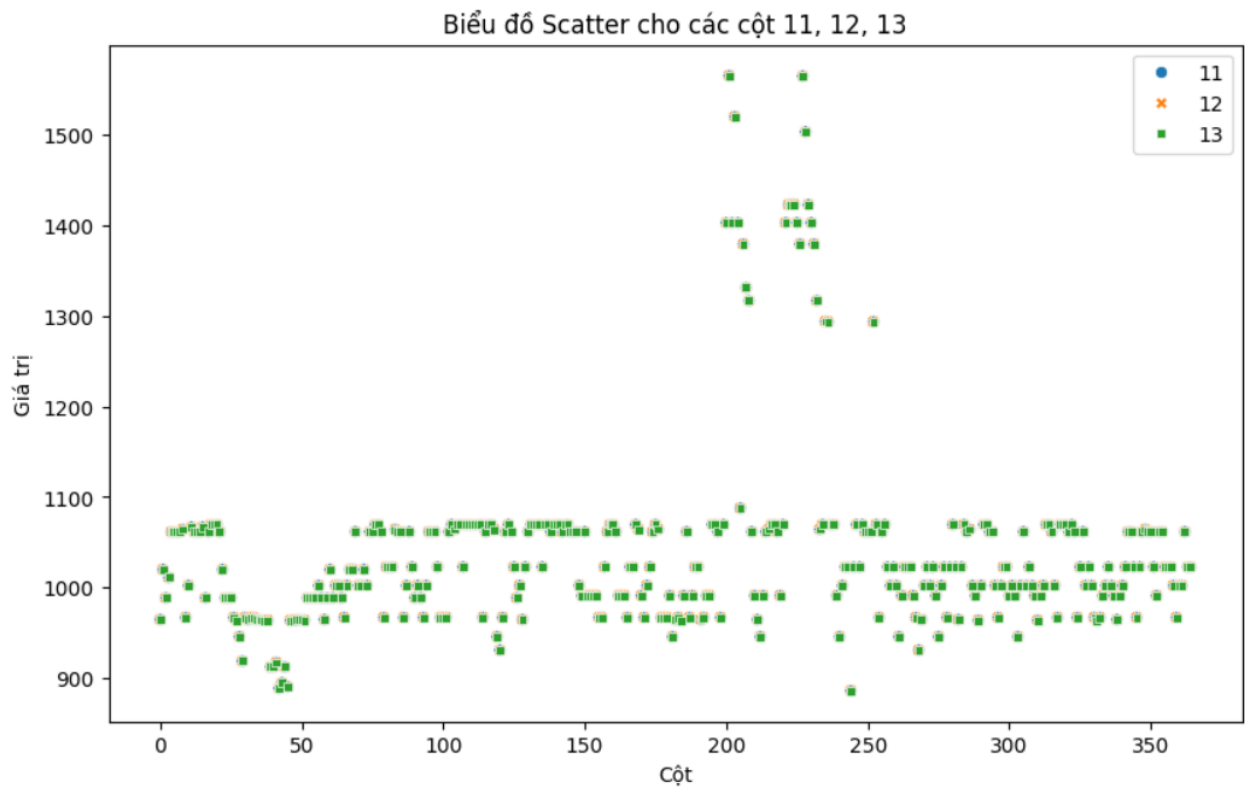
2. Heatmap



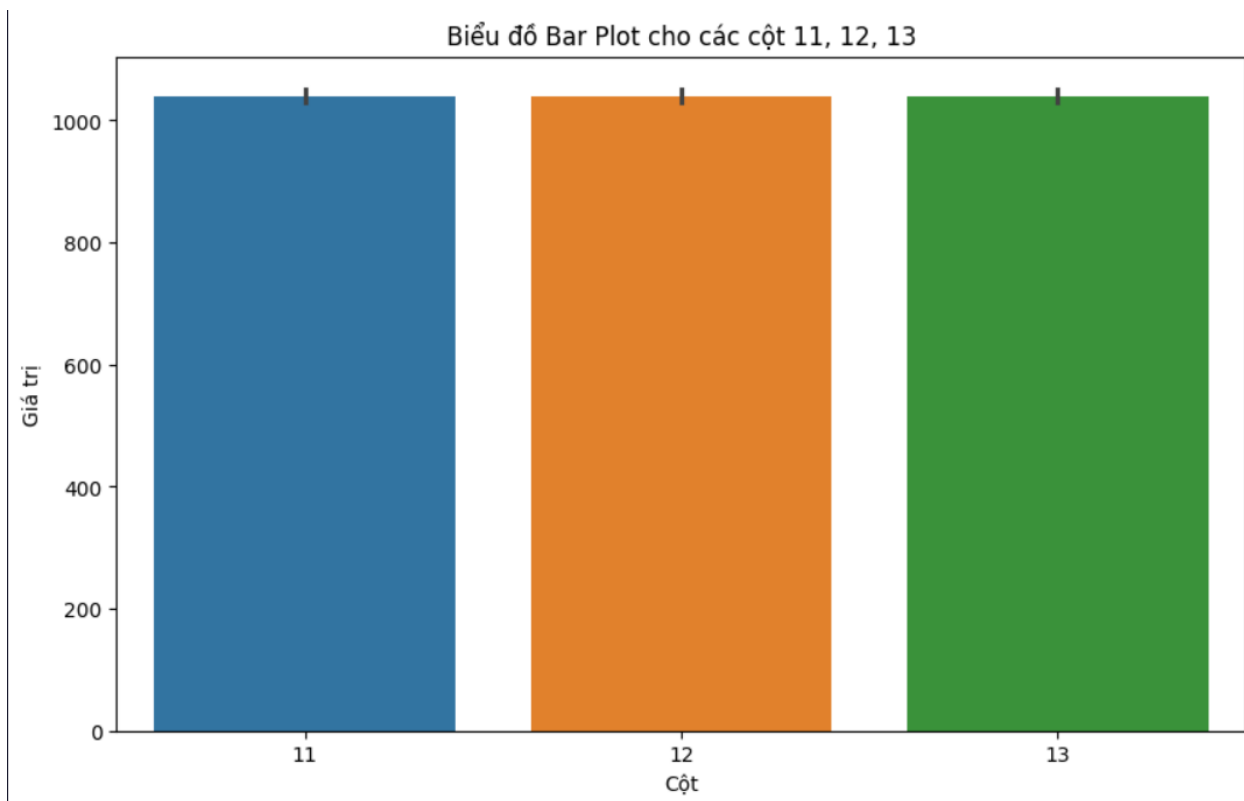
3. Box plot



#### 4. Scatter plot



## 5. Bar plot

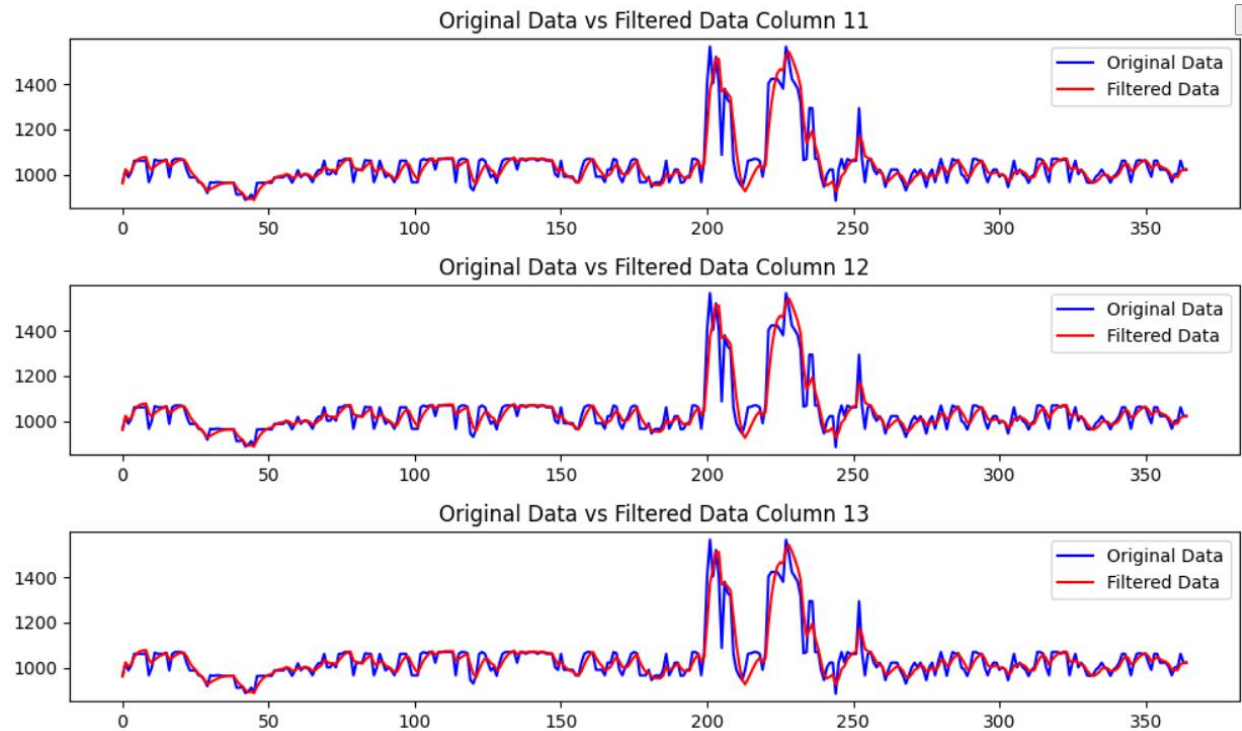


# KALMAN

```
# Function to apply Kalman Filter
def apply_kalman_filter(data):
    kf = KalmanFilter(dim_x=2, dim_z=1)
    kf.x = np.array([0., 0.])
    kf.F = np.array([[1., 1.], [0., 1.]])
    kf.H = np.array([[1., 0.]])
    kf.P *= 1000.
    kf.R = 5
    kf.Q = np.array([[0.1, 0.1], [0.1, 0.1]])

    filtered_data = []
    for z in data:
        kf.predict()
        kf.update(z)
        filtered_data.append(kf.x[0])
    return filtered_data

# Apply Kalman Filter to each column
filtered_data_11 = apply_kalman_filter(selected_data['11'].values)
filtered_data_12 = apply_kalman_filter(selected_data['12'].values)
filtered_data_13 = apply_kalman_filter(selected_data['13'].values)
```



## KMEANS

```

• ✓ from sklearn.cluster import KMeans
  from sklearn.model_selection import train_test_split
  import matplotlib.pyplot as plt
  from mpl_toolkits.mplot3d import Axes3D

# Chia dữ liệu thành tập huấn luyện và tập kiểm tra
X_train, X_test, y_train, y_test = train_test_split(selected_data[['11', '12', '13']], selected_data['Cluster'], test_size=0.2, random_state=42)

# Tạo một đối tượng KMeans với số cụm mong muốn
kmeans = KMeans(n_clusters=3)

# Phân nhóm dữ liệu huấn luyện
clusters_train = kmeans.fit_predict(X_train)

# Thêm cột 'Cluster' vào tập dữ liệu huấn luyện để lưu trữ nhóm của mỗi điểm dữ liệu
X_train['Cluster'] = clusters_train

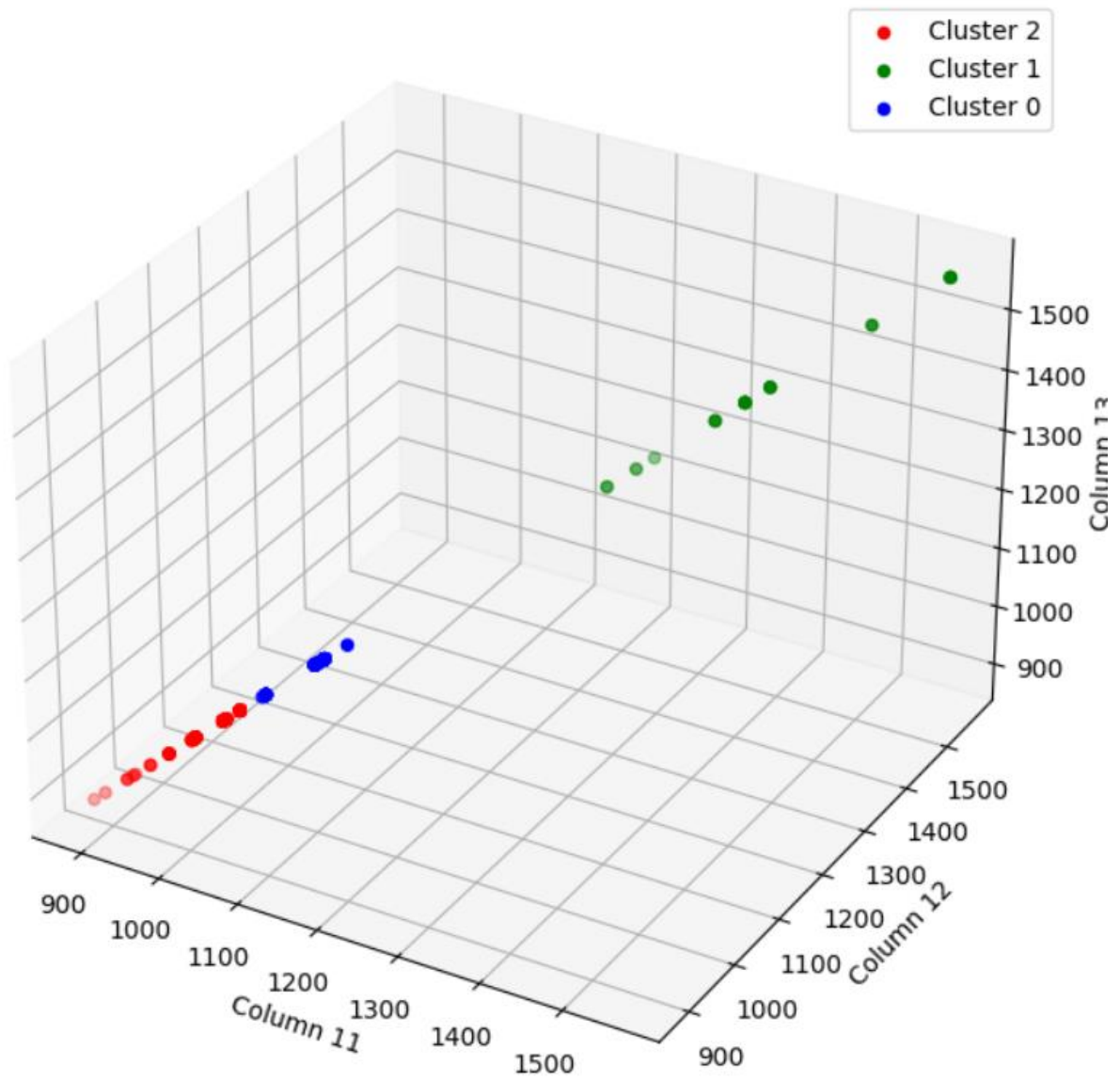
# Vẽ biểu đồ 3D của các nhóm trên tập huấn luyện
fig = plt.figure(figsize=(10, 8))
ax = fig.add_subplot(111, projection='3d')

colors = ['r', 'g', 'b']
✓ for cluster, color in zip(X_train['Cluster'].unique(), colors):
    cluster_data = X_train[X_train['Cluster'] == cluster]
    ax.scatter(cluster_data['11'], cluster_data['12'], cluster_data['13'], c=color, label=f'Cluster {cluster}')

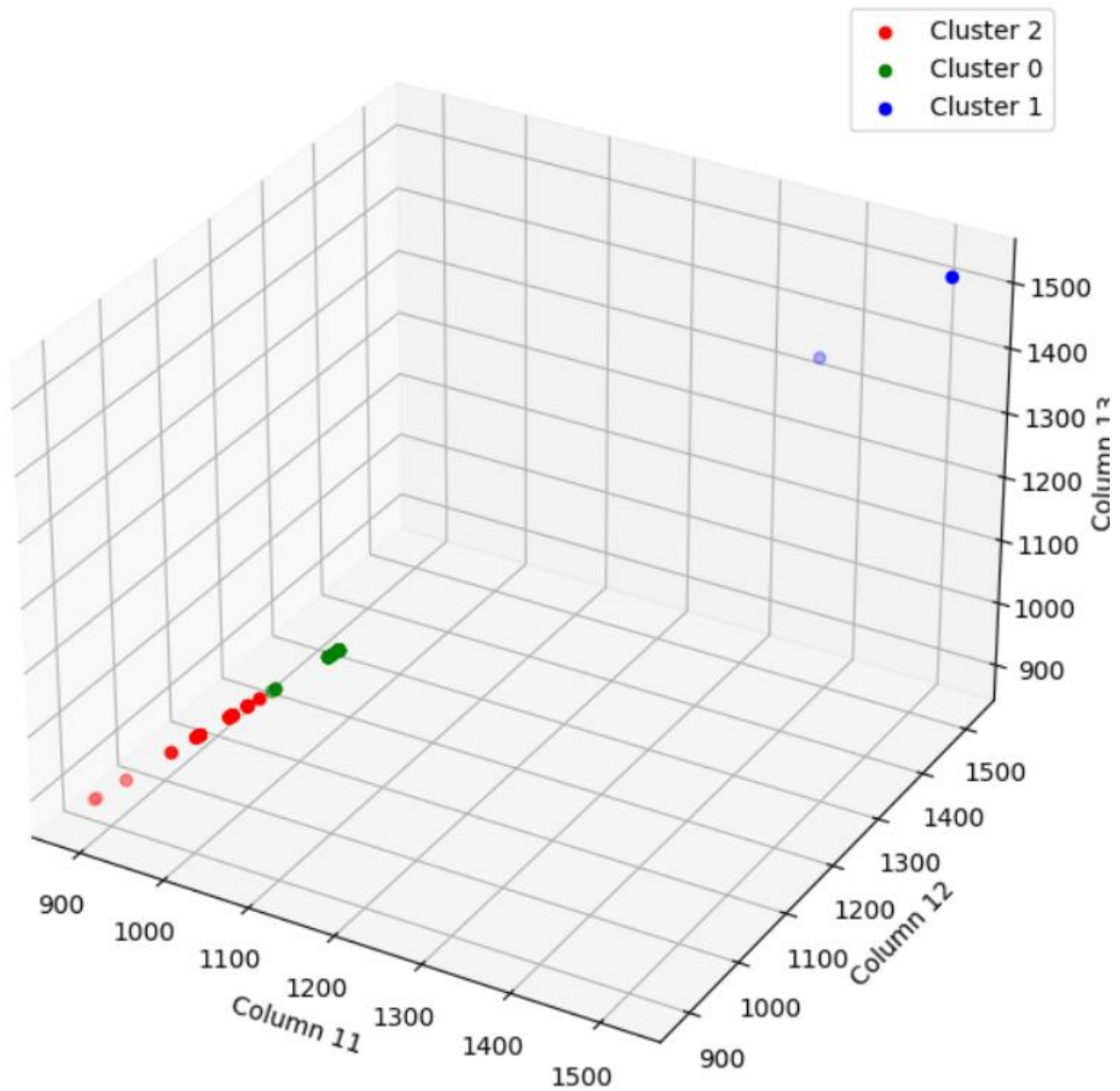
```



KMeans Clustering of Columns 11, 12, and 13 (Training Data)



KMeans Clustering of Columns 11, 12, and 13 (Test Data)



[https://github.com/ducjr/TH3\\_PhanzTichChuoiThoiGian](https://github.com/ducjr/TH3_PhanzTichChuoiThoiGian)