```
### Network Slicing Recognition
```

The telecom industry is going through a massive digital transformation with the adoption of ML, AI, feedback-based automation and advanced analytics to handle the next generation applications and services. AI concepts are not new; the algorithms used by Machine Learning and Deep Learning are being currently implemented in various industries and technology verticals. With growing data and immense volume

of information over 5G, the ability to predict data proactively, swiftly and with accuracy, is critically important. Data-driven decision making will be vital in future communication networks due to the traffic explosion and Artificial Intelligence (AI) will accelerate the 5G network performance.

Mobile operators are looking for a programmable solution that will allow them to accommodate multiple independent tenants on the same physical infrastructure and 5G networks allow for end-to-end network resource allocation using the concept of Network Slicing (NS).

Network Slicing will play a vital role in enabling a multitude of 5G applications, use cases, and services. Network slicing functions will provide an end-to-end isolation between slices with an ability to customize each slice based on the service demands (bandwidth, coverage, security, latency, reliability, etc).

Your Task is to build a Machine Learning model that will be able to to proactively detect and eliminate threats based on incoming connections thereby selecting the most appropriate network slice, even in case of a network failure.

LTE/5g - User Equipment categories or classes to define the performance specifications
Packet Loss Rate - number of packets not received divided by the total number of packets sent.
Packet Delay - The time for a packet to be received.
Slice type - network configuration that allows multiple networks (virtualized and independent)
GBR - Guaranteed Bit Rate
Healthcare - Usage in Healthcare (1 or 0)
Industry 4.0 - Usage in Digital Enterprises(1 or 0)
IoT Devices - Usage
Public Safety - Usage for public welfare and safety purposes (1 or 0)
Smart City & Home - usage in daily household chores
Smart Transportation - usage in public transportation
Smartphone - whether used for smartphone cellular data

```
###! pip install neattext
```

```
import pandas as pd
import numpy as np
import neattext.functions as nfx
import seaborn as sn

from sklearn.feature_extraction.text import TfidfVectorizer,CountVectorizer
from sklearn.metrics.pairwise import cosine_similarity,linear_kernel
```

```
##! pip uninstall numpy
###!pip install numpy==1.20
```

```
###!mkdir ~/.kaggle
```

```
###!cp /kaggle.json ~/.kaggle/
```

```
###! pip install kaggle
###!pip install keras-tuner
```

```
###!kaggle datasets download -d gauravduttakiit/network-slicing-recognition
```

```
###!unzip /content/network-slicing-recognition.zip
```

```
train_dataset = pd.read_csv("/content/train_dataset.csv")
test_dataset = pd.read_csv("/content/test_dataset.csv")
```

```
print(train_dataset.shape, test_dataset.shape)
```

```
(31583, 17) (31584, 16)
```

```
test_dataset['slice Type'] = 0
```

```
train_dataset = train_dataset.reset_index()
test_dataset = test_dataset.reset_index()
train_dataset.rename(columns = { "index" : "ID"}, inplace = True)
test_dataset.rename(columns = { "index" : "ID"}, inplace = True)
```

```
train_dataset.columns
```

```
Index(['ID', 'LTE/5g Category', 'Time', 'Packet Loss Rate', 'Packet delay',
       'IoT', 'LTE/5G', 'GBR', 'Non-GBR', 'AR/VR/Gaming', 'Healthcare',
       'Industry 4.0', 'IoT Devices', 'Public Safety', 'Smart City & Home',
       'Smart Transportation', 'Smartphone', 'slice Type'],
      dtype='object')
```

```
train_dataset.shape
```

```
(31583, 18)
```

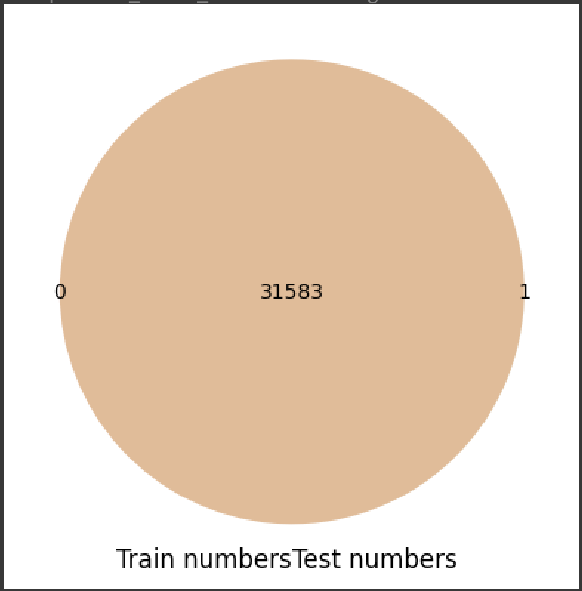
```
train_dataset['slice Type'].value_counts()
```

```
1    16799
3     7392
2     7392
Name: slice Type, dtype: int64
```

```
from matplotlib_venn import venn2, venn2_circles, venn2_unweighted
from matplotlib_venn import venn3, venn3_circles
```

```
set_numbers_train = set(train_dataset[['ID']].drop_duplicates().sort_values(by = 'ID')['ID'].tolist())
set_numbers_test = set(test_dataset[['ID']].drop_duplicates().sort_values(by = 'ID')['ID'].tolist())
venn2((set_numbers_train, set_numbers_test), set_labels = ('Train numbers', 'Test numbers'))
```

```
<matplotlib_venn._common.VennDiagram at 0x7a6abca9dab0>
```




```
train_dataset.columns
```

```
Index(['ID', 'LTE/5g Category', 'Time', 'Packet Loss Rate', 'Packet delay',
       'IoT', 'LTE/5G', 'GBR', 'Non-GBR', 'AR/VR/Gaming', 'Healthcare',
       'Industry 4.0', 'IoT Devices', 'Public Safety', 'Smart City & Home',
```

```
       'Smart Transportation', 'Smartphone', 'slice Type'],
      dtype='object')
```

```
####! pip install klib
```

```
import klib
```

```
train_dataset = klib.clean_column_names(train_dataset)
test_dataset = klib.clean_column_names(test_dataset)
```

```
train_dataset = klib.convert_datatypes(train_dataset)
test_dataset = klib.convert_datatypes(test_dataset)
```

```
train_dataset.columns
```

```
Index(['id', 'lte_5g_category', 'time', 'packet_loss_rate', 'packet_delay',
       'io_t', 'lte_5g', 'gbr', 'non_gbr', 'ar_vr_gaming', 'healthcare',
       'industry_4_0', 'io_t_devices', 'public_safety', 'smart_city_and_home',
       'smart_transportation', 'smartphone', 'slice_type'],
      dtype='object')
```

Anomaly Detection Using One-Class SVM

```
from sklearn import svm
```

```
clf = svm.OneClassSVM(nu=0.05, kernel="rbf", gamma=0.1)
clf.fit(train_dataset)
```

```
▾          OneClassSVM
OneClassSVM(gamma=0.1, nu=0.05)
```

```
pred = clf.predict(train_dataset)

# inliers are labeled 1, outliers are labeled -1
normal = train_dataset[pred == 1]
abnormal = train_dataset[pred == -1]
```

```
print(normal.shape, abnormal.shape)
```

```
(18373, 18) (13210, 18)
```

```
normal.columns
```

```
Index(['id', 'lte_5g_category', 'time', 'packet_loss_rate', 'packet_delay',
       'io_t', 'lte_5g', 'gbr', 'non_gbr', 'ar_vr_gaming', 'healthcare',
       'industry_4_0', 'io_t_devices', 'public_safety', 'smart_city_and_home',
       'smart_transportation', 'smartphone', 'slice_type'],
      dtype='object')
```

```
normal['slice_type'].value_counts()
```

```
1    9839
2    4294
3    4240
Name: slice_type, dtype: int64
```

```
train_dataset = normal
```

```
print(train_dataset.shape)
print(train_dataset.columns)
```

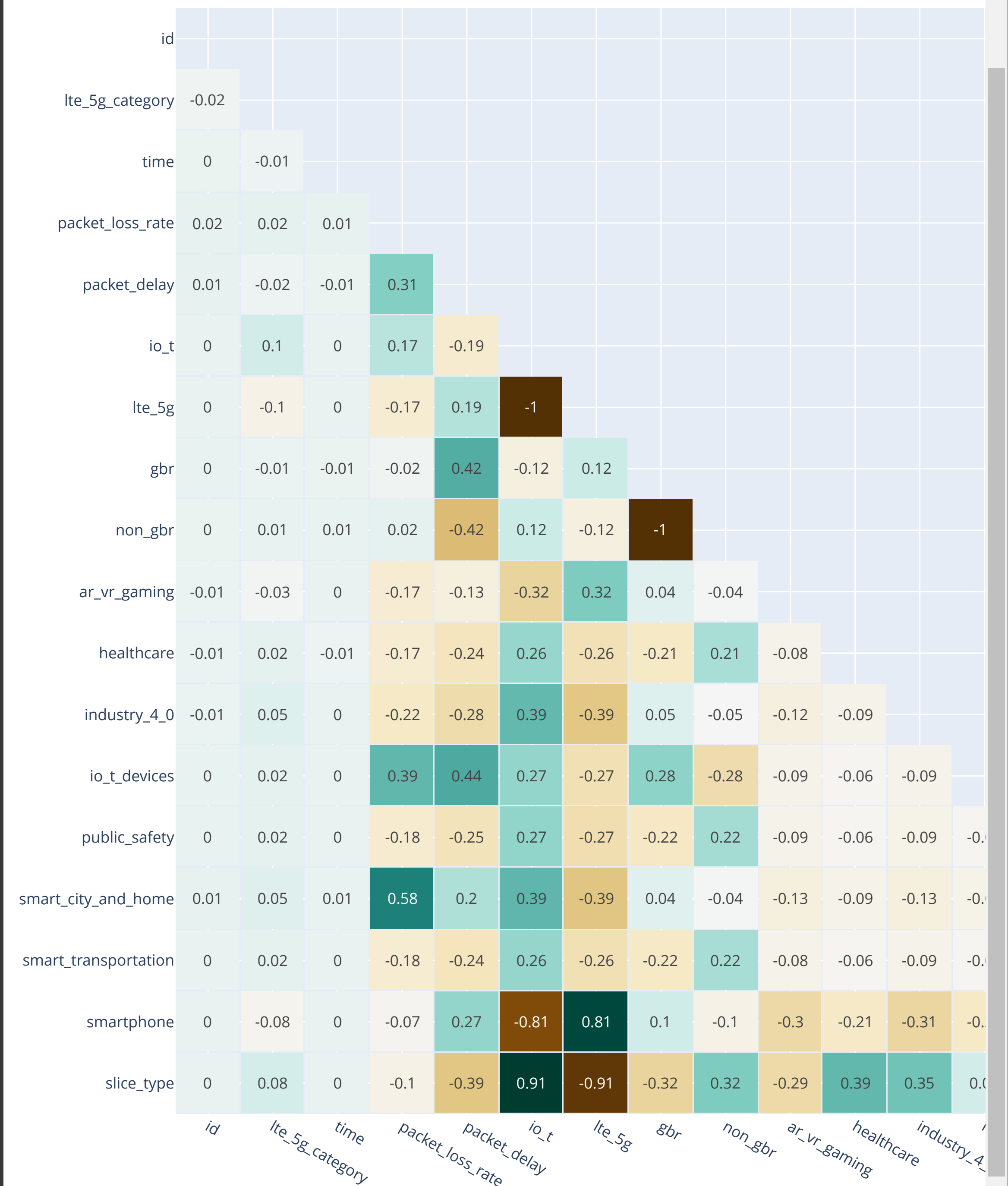
```
(18373, 18)
Index(['id', 'lte_5g_category', 'time', 'packet_loss_rate', 'packet_delay',
       'io_t', 'lte_5g', 'gbr', 'non_gbr', 'ar_vr_gaming', 'healthcare',
       'industry_4_0', 'io_t_devices', 'public_safety', 'smart_city_and_home',
       'smart_transportation', 'smartphone', 'slice_type'],
      dtype='object')
```

```
test_dataset['slice_type'] = 0
```

```
klib.cat_plot(train_dataset)
```

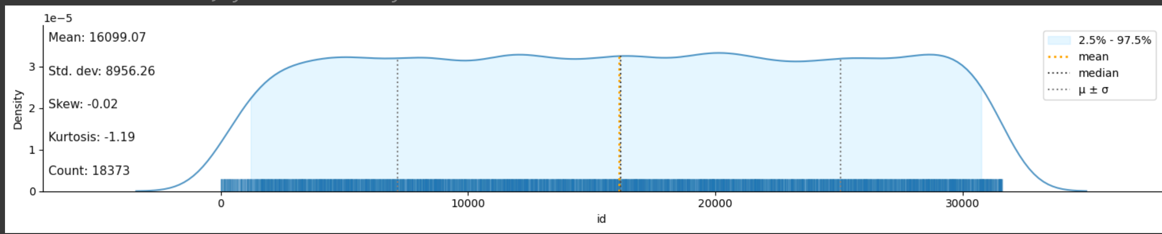
```
No columns with categorical data were detected.
```

```
klib.corr_interactive_plot(train_dataset)
```




```
klib.dist_plot(train_dataset)
```

```
Large dataset detected, using 10000 random samples for the plots. Summary statistics are still based or
<Axes: xlabel='id', ylabel='Density'>
```




```python
klib.missingval_plot(train_dataset)
```

```
No missing values found in the dataset.
```

```python
klib.corr_mat(train_dataset)
```

	id	lte_5g_category	time	packet_loss_rate	packet_delay	io_t	lte_5g	gbr	no
id	1.00	-0.02	-0.00	0.02	0.01	0.00	-0.00	0.00	
lte_5g_category	-0.02	1.00	-0.01	0.02	-0.02	0.10	-0.10	-0.01	
time	-0.00	-0.01	1.00	0.01	-0.01	0.00	-0.00	-0.01	
packet_loss_rate	0.02	0.02	0.01	1.00	0.31	0.17	-0.17	-0.02	
packet_delay	0.01	-0.02	-0.01	0.31	1.00	-0.19	0.19	0.42	
io_t	0.00	0.10	0.00	0.17	-0.19	1.00	-1.00	-0.12	
lte_5g	-0.00	-0.10	-0.00	-0.17	0.19	-1.00	1.00	0.12	
gbr	0.00	-0.01	-0.01	-0.02	0.42	-0.12	0.12	1.00	
non_gbr	-0.00	0.01	0.01	0.02	-0.42	0.12	-0.12	-1.00	
ar_vr_gaming	-0.01	-0.03	-0.00	-0.17	-0.13	-0.32	0.32	0.04	
healthcare	-0.01	0.02	-0.01	-0.17	-0.24	0.26	-0.26	-0.21	
industry_4_0	-0.01	0.05	0.00	-0.22	-0.28	0.39	-0.39	0.05	
io_t_devices	0.00	0.02	-0.00	0.39	0.44	0.27	-0.27	0.28	
public_safety	0.00	0.02	0.00	-0.18	-0.25	0.27	-0.27	-0.22	
smart_city_and_home	0.01	0.05	0.01	0.58	0.20	0.39	-0.39	0.04	
smart_transportation	-0.00	0.02	-0.00	-0.18	-0.24	0.26	-0.26	-0.22	
smartphone	0.00	-0.08	0.00	-0.07	0.27	-0.81	0.81	0.10	
slice_type	-0.00	0.08	-0.00	-0.10	-0.39	0.91	-0.91	-0.32	

```python
train_dataset.columns
```

```
Index(['id', 'lte_5g_category', 'time', 'packet_loss_rate', 'packet_delay',
       'io_t', 'lte_5g', 'gbr', 'non_gbr', 'ar_vr_gaming', 'healthcare',
       'industry_4_0', 'io_t_devices', 'public_safety', 'smart_city_and_home',
       'smart_transportation', 'smartphone', 'slice_type'],
      dtype='object')
```

```python
# Checking for outliers in the continuous variables
num_train_dataset = train_dataset[['id', 'lte_5g_category', 'time', 'packet_loss_rate', 'packet_delay',
       'io_t', 'lte_5g', 'gbr', 'non_gbr', 'ar_vr_gaming', 'healthcare',
       'industry_4_0', 'io_t_devices', 'public_safety', 'smart_city_and_home',
       'smart_transportation', 'smartphone', 'slice_type']]
```

```python
train_dataset['slice_type'].value_counts()
```

```
1    9839
2    4294
3    4240
Name: slice_type, dtype: int64
```

```
train_dataset.columns
```

```
Index(['id', 'lte_5g_category', 'time', 'packet_loss_rate', 'packet_delay',
       'io_t', 'lte_5g', 'gbr', 'non_gbr', 'ar_vr_gaming', 'healthcare',
       'industry_4_0', 'io_t_devices', 'public_safety', 'smart_city_and_home',
       'smart_transportation', 'smartphone', 'slice_type'],
      dtype='object')
```

```
y_train = train_dataset['slice_type']
x_train = train_dataset.drop('slice_type', axis = 1)
y_test = test_dataset['slice_type']
x_test = test_dataset.drop('slice_type', axis = 1)
```

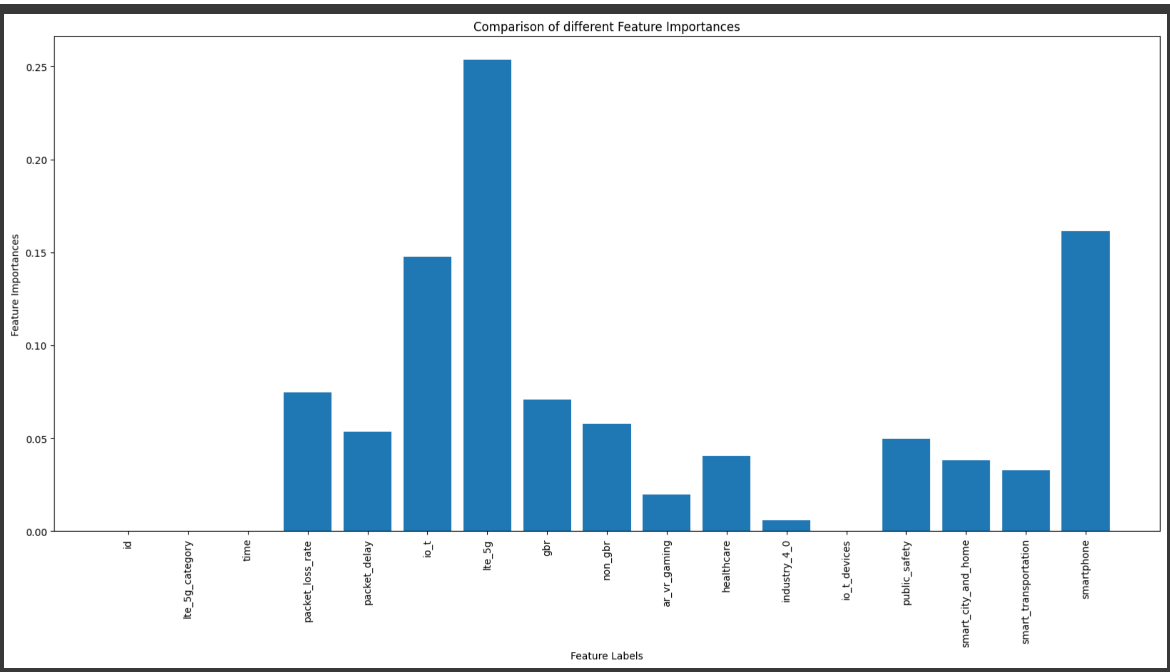
```
from sklearn.ensemble import ExtraTreesClassifier
extra_tree_forest = ExtraTreesClassifier(n_estimators = 5,
                                        criterion ='entropy', max_features = 2)
extra_tree_forest.fit(x_train, y_train)
feature_importance = extra_tree_forest.feature_importances_
feature_importance_normalized = np.std([tree.feature_importances_ for tree in
                                        extra_tree_forest.estimators_],
                                        axis = 0)
```

```
import matplotlib.pyplot as plt
```

```
feature_importance_normalized
```

```
array([9.95850150e-05, 1.82899866e-05, 1.45823303e-05, 7.45023463e-02,
       5.36511123e-02, 1.47469251e-01, 2.53387050e-01, 7.08798855e-02,
       5.78137139e-02, 1.97852518e-02, 4.03864792e-02, 5.94845917e-03,
       0.00000000e+00, 4.98605017e-02, 3.80180060e-02, 3.29054755e-02,
       1.61403769e-01])
```

```
plt.figure(figsize = [20,9])
plt.bar(x_train.columns, feature_importance_normalized)
plt.xlabel('Feature Labels')
plt.ylabel('Feature Importances')
plt.xticks(rotation = 90)
plt.title('Comparison of different Feature Importances')
plt.show()
```




```
x_train.columns
```

```
Index(['id', 'lte_5g_category', 'time', 'packet_loss_rate', 'packet_delay',
       'io_t', 'lte_5g', 'gbr', 'non_gbr', 'ar_vr_gaming', 'healthcare',
       'industry_4_0', 'io_t_devices', 'public_safety', 'smart_city_and_home',
       'smart_transportation', 'smartphone'],
      dtype='object')
```

```
x_train2 = x_train[['lte_5g_category', 'packet_loss_rate', 'packet_delay',
       'io_t', 'lte_5g', 'gbr', 'non_gbr', 'ar_vr_gaming', 'healthcare',
       'industry_4_0', 'io_t_devices', 'public_safety', 'smart_city_and_home',
       'smart_transportation', 'smartphone']]
```

```
x_test2 = x_test[['lte_5g_category', 'packet_loss_rate', 'packet_delay',
       'io_t', 'lte_5g', 'gbr', 'non_gbr', 'ar_vr_gaming', 'healthcare',
       'industry_4_0', 'io_t_devices', 'public_safety', 'smart_city_and_home',
       'smart_transportation', 'smartphone']]
```

```
print(x_train2.shape, x_test2.shape)
```

```
(18373, 15) (31584, 15)
```

```
y_train.value_counts()
```

```
1    9839
2    4294
3    4240
Name: slice_type, dtype: int64
```

```
x_train2 = pd.DataFrame(x_train2)
```

```
x_train2.head(4)
```

	lte_5g_category	packet_loss_rate	packet_delay	io_t	lte_5g	gbr	non_gbr	ar_vr_gaming	healthcare	industry_4_0	io_t_devices	pu
0	14	0.000001	10	1	0	0	1	0	0	0	0	
4	9	0.010000	50	1	0	0	1	0	0	0	0	
5	19	0.000001	10	1	0	0	1	0	0	1	0	
8	8	0.001000	150	0	1	0	1	0	0	0	0	

Next steps: Generate code with `x_train2` View recommended plots

```
x_train2 = pd.DataFrame(x_train2)
x_test2 = pd.DataFrame(x_test2)
```

Pearson Correlation

```
x_train2.astype(float).corr()
```

	lte_5g_category	packet_loss_rate	packet_delay	io_t	lte_5g	gbr	no
lte_5g_category	1.000000	0.022254	-0.015085	0.095605	-0.095605	-0.007156	0.00
packet_loss_rate	0.022254	1.000000	0.306436	0.170858	-0.170858	-0.022315	0.02
packet_delay	-0.015085	0.306436	1.000000	-0.187910	0.187910	0.424124	-0.42
io_t	0.095605	0.170858	-0.187910	1.000000	-1.000000	-0.122402	0.12
lte_5g	-0.095605	-0.170858	0.187910	-1.000000	1.000000	0.122402	-0.12
gbr	-0.007156	-0.022315	0.424124	-0.122402	0.122402	1.000000	-1.00
non_gbr	0.007156	0.022315	-0.424124	0.122402	-0.122402	-1.000000	1.00
ar_vr_gaming	-0.034010	-0.168172	-0.127614	-0.321284	0.321284	0.037476	-0.03
healthcare	0.018408	-0.174050	-0.238384	0.260863	-0.260863	-0.213782	0.21
industry_4_0	0.045106	-0.216202	-0.284384	0.385515	-0.385515	0.045353	-0.04
io_t_devices	0.019458	0.389878	0.437355	0.265817	-0.265817	0.281337	-0.28
public_safety	0.019485	-0.182358	-0.249763	0.273314	-0.273314	-0.223986	0.22
smart_city_and_home	0.047347	0.578743	0.200818	0.394585	-0.394585	0.035940	-0.03
smart_transportation	0.019787	-0.176111	-0.241206	0.263951	-0.263951	-0.216313	0.21
smartphone	-0.075152	-0.067416	0.268831	-0.807524	0.807524	0.099995	-0.09

```
def correlation(dataset, threshold):
    col_corr = set() # Set of all the names of deleted columns
    corr_matrix = dataset.corr()
    for i in range(len(corr_matrix.columns)):
        for j in range(i):
            if (corr_matrix.iloc[i, j] >= threshold) and (corr_matrix.columns[j] not in col_corr):
                colname = corr_matrix.columns[i] # getting the name of column
                col_corr.add(colname)
                if colname in dataset.columns:
                    del dataset[colname] # deleting the column from the dataset

    print(dataset.columns)
    print(dataset.shape)
```

```
correlation(x_train2, 0.95)
```

```
Index(['lte_5g_category', 'packet_loss_rate', 'packet_delay', 'io_t', 'lte_5g',
       'gbr', 'non_gbr', 'ar_vr_gaming', 'healthcare', 'industry_4_0',
       'io_t_devices', 'public_safety', 'smart_city_and_home',
```

```
       'smart_transportation', 'smartphone'],
      dtype='object')
(18373, 15)
```

```
x_test2.columns
x_test2.shape
```

```
(31584, 15)
```

Standard Scaler

```
from sklearn.preprocessing import StandardScaler
scaler=StandardScaler()
x_train2=pd.DataFrame(scaler.fit_transform(x_train2),columns=x_train2.columns)
x_train2.head()
```

	lte_5g_category	packet_loss_rate	packet_delay	io_t	lte_5g	gbr	non_gbr	ar_vr_gaming	healthcare	industry_4_0	io_t_
0	0.496569	-0.716409	-0.981216	1.073740	-1.073740	-0.87995	0.87995	-0.344976	-0.242948	-0.359040	-(
1	-0.319527	1.574868	-0.602200	1.073740	-1.073740	-0.87995	0.87995	-0.344976	-0.242948	-0.359040	-(
2	1.312664	-0.716409	-0.981216	1.073740	-1.073740	-0.87995	0.87995	-0.344976	-0.242948	2.785209	-(
3	-0.482746	-0.487488	0.345339	-0.931324	0.931324	-0.87995	0.87995	-0.344976	-0.242948	-0.359040	-(
4	-0.645965	-0.716409	-0.981216	1.073740	-1.073740	-0.87995	0.87995	-0.344976	4.116113	-0.359040	-(

Next steps: Generate code with x_train2 View recommended plots

```
x_test2=pd.DataFrame(scaler.fit_transform(x_test2),columns=x_test2.columns)
x_test2.head()
```

	lte_5g_category	packet_loss_rate	packet_delay	io_t	lte_5g	gbr	non_gbr	ar_vr_gaming	healthcare	industry_4_0	io_
0	0.664781	-0.482940	-0.136290	-0.938083	0.938083	1.124027	-1.124027	2.896828	-0.25058	-0.365662	
1	0.500168	-0.712444	-0.982767	1.066004	-1.066004	-0.889658	0.889658	-0.345205	-0.25058	-0.365662	
2	0.006327	-0.482940	-0.606555	1.066004	-1.066004	1.124027	-1.124027	-0.345205	-0.25058	2.734762	