

In [1]:

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import OrdinalEncoder
from sklearn.metrics import mean_squared_error
from sklearn.ensemble import RandomForestClassifier
from xgboost import XGBClassifier
```

In [2]:

```
df2019 = pd.read_excel("Financial_Report2019.xlsx", sheet_name="CONSOLIDATED STATEMENTS OF INCO", skiprows=1)
df2016 = pd.read_excel("Financial_Report2016.xlsx", sheet_name="CONSOLIDATED STATEMENTS OF INCO", skiprows=1)
df2013 = pd.read_excel("Financial_Report2013.xlsx", sheet_name="CONSOLIDATED STATEMENTS OF _INC", skiprows=1)
```

In [3]:

```
df2019.shape
```

Out[3]:

```
(20, 4)
```

In [4]:

```
df2016.shape
```

Out[4]:

```
(22, 5)
```

In [5]:

```
df2013.shape
```

Out[5]:

```
(22, 7)
```

In [6]:

```
df2019.head()
```

Out[6]:

	Unnamed: 0	Dec. 31, 2019	Dec. 31, 2018	Dec. 31, 2017
0	Revenues	37266.0	34300.0	36212.0
1	Cost of Goods and Services Sold	14619.0	13067.0	13721.0
2	GROSS PROFIT	22647.0	21233.0	22491.0
3	Selling, General and Administrative Expense	12103.0	11002.0	12834.0
4	Other Cost and Expense, Operating	458.0	1079.0	1902.0

In [7]:

```
df2019[["Dec. 31, 2016", "Dec. 31, 2015", "Dec. 31, 2014"]] = df2016[["Dec. 31, 2016", "Dec. 31, 2015", "Dec. 31, 2014"]]
```

In [8]:

```
df2019[["Dec. 31, 2013", "Dec. 31, 2012", "Dec. 31, 2011"]] = df2013[["Dec. 31, 2013", "Dec. 31, 2012", "Dec. 31, 2011"]]
```

In [9]:

```
df2016.head()
```

Out[9]:

	Unnamed: 0	Unnamed: 1	Dec. 31, 2016	Dec. 31, 2015	Dec. 31, 2014
0	NET OPERATING REVENUES	NaN	41863.0	44294.0	45998.0
1	Cost of goods sold	NaN	16465.0	17482.0	17889.0
2	GROSS PROFIT	NaN	25398.0	26812.0	28109.0
3	Selling, general and administrative expenses	NaN	15262.0	16427.0	17218.0
4	Other operating charges	NaN	1510.0	1657.0	1183.0

In [10]:

```
df2016.drop(columns=["Unnamed: 1"], inplace=True)
```

In [11]:

```
df2016.drop([20, 21], inplace=True)
```

In [12]:

```
df2013.drop([20, 21], inplace=True)
```

In [13]:

```
df2013.head()
```

Out[13]:

	In Millions, except Per Share data, unless otherwise specified	Dec. 31, 2013	Unnamed: 2	Dec. 31, 2012	Unnamed: 4	Dec. 31, 2011	Unnamed: 6
0	NET OPERATING REVENUES	46854	NaN	48017.0	NaN	46542.0	NaN
1	Cost of goods sold	18421	NaN	19053.0	NaN	18215.0	NaN
2	GROSS PROFIT	28433	NaN	28964.0	NaN	28327.0	NaN
3	Selling, general and administrative expenses	17310	NaN	17738.0	NaN	17422.0	NaN
4	Other operating charges	895	NaN	447.0	NaN	732.0	NaN

In [14]:

```
df2013.drop(columns=["Unnamed: 2", "Unnamed: 4", "Unnamed: 6"], inplace=True)
```