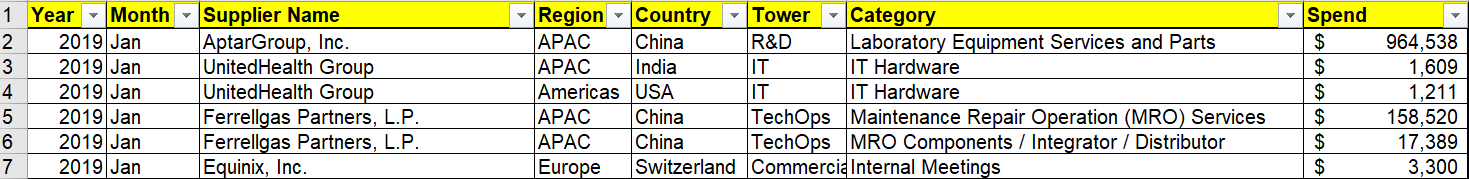
# Studying Data:



We can notice below points:

1. Data is not uniform across month i.e. number of data points per month is varying.
2. There might be interdependency of feature on each other. For example: ‘Spend’ might depend on Region, Americas having higher Spend than APAC.

**Assumption**:

1. Dollar Value of Spend is not converted on purchasing power parity basic. For example: May be because of high dollar exchange rate Spend is comparatively lower in APAC region.

**Created New Features**:

1. Monthly average of expenditure per Vendor, Region, Country, Tower and Category. Thus created 5 additional columns.
2. Scaled Spending wrt. Monthly average per row. This is same as Monthly\_Avg/Spend per row. This also created 5 columns.

# Different Approaches & Libraries

**General Libraries**: Pandas, Numpy, Matplotlib, Seaborn, yellowbrick

1. **Time Series Modeling**: Libraries used are Statsmodels
2. **Panel Data Modeling**: Libraries used are linearmodels
3. **Machine Learning Modeling**: Libraries used are:
   * + - 1. Scikit-learn for feature engineering & model building
         2. Feature-engine for feature engineering

# Algorithms Used

Tried creating a pipeline which will be suitable for both Linear Model and Ensembled Models. Since there are very few linear data available it would be advisable to use tree-based models.

Regressor of below algorithms are tried with the model:

1. Linear Models
2. Random Forest
3. Gradient Boosting
4. eXtreme Gradient Boosting

# Technical Stack & Flow Diagram



# Future Scope

1. A detailed discussion on data is required to study the possibility of converting it in Panel Data Format or Pooled Regression.
2. Few more features could be brought in.
3. Micro/Macro Economic factors could be brought in.