# Assignment1 : Data Preprocessing

**Dataset: Telecom Churn Dataset (Telecom\_churn\_data.csv)**

Load Data and Write script to perform following operations.

1. Load the data and display the structure.
2. Check the data types are appropriate and convert char to factors wherever required.
3. Check for missing values and duplicates if any.
4. Find out the churn based on contract and paperlessbilling attribute.
5. Plot histograms for Tenure, TotalCharges and Monthlycharges .
6. Display density plot for TotalCharges vs Churn(Yes/No) and MonthlyCharges vs churn(Yes/No)

# Dataset: IoT Devices Data (IOT\_Temprature.csv)

Load Data and Write script to perform following operations.

1. Load the data and Display structure,summary report.
2. Check the attributes data types and Drop out unwanted columns like id , room\_id.
3. Check for missing values in any of the columns.
4. Separate out date and time values of noted\_date column.
5. Further separate date,month,year from date and store as separate columns. Display the complete data with new columns.
6. Prepare boxplot of temperature.
7. Prepare barplot for in/out vs temprature.

# Dataset: NYC AirBnB Dataset (AB\_NYC\_2019.csv)

Load Data and Write script to perform following operations.

1. Display structure and summary report.
2. Check data types of each column and ensure whether the data types for each of the attributes are appropriate, if not make appropriate corrections.
3. Display NULL values for each of the columns
4. Find out missing values and replace them for the applicable columns.
5. Check which hosts (IDs) have the most listings on Airbnb platform and taking advantage of this service. Display in ascending order.
6. Plot a bar graph for Host\_ID vs Count of Listings.

Based on Room\_Type, prepare a bar graph for neighbourhood vs Count of Listings