



**ACADGILD**

# SESSION 10: Correlations

## Assignment 1

---

## Table of Contents

1.Introduction .....	3
2.Objective .....	3
3.Prerequisites .....	3
4.Associated Data Files .....	3
5.Problem Statement .....	3
6.Expected Output .....	3
7.Approximate Time to Complete Task .....	3

## 1. Introduction

This assignment will help you understand the concepts learnt in the session.

## 2. Objective

This assignment will test your skills on **Correlations** in R.

## 3. Prerequisites

Not applicable.

## 4. Associated Data Files

Not applicable.

## 5. Problem Statement

1. Import dataset from the following link:

<https://archive.ics.uci.edu/ml/machine-learning-databases/00360/>

Perform the below written operations:

- a. Read the file in Zip format and get it into R
- b. Create Univariate for all the columns.
- c. Check for missing values in all columns.
- d. Impute the missing values using appropriate methods
- e. Create bi-variate analysis for all relationships
- f. Test relevant hypothesis for valid relations
- g. Create cross tabulations with derived variables
- h. check for trends and patterns in time series
- i. Find out the most polluted time of the day and the name of the chemical compound.

## 6. Expected Format

1. R file should be submitted where applicable.
2. R file should be in PDF or in .r format
3. Proper screenshots of the outputs should be submitted as well
4. The r codes, if submitted in any other format, will be subjected to deduction in marks

Note: Your solution will not be entertained if it is any other format, e.g., .zip, .doc, .rtf etc.

## **7. Approximate Time to Complete Task**

40 mins.