NAME:Aditya Somani CLASS:TE-A ROLL NO:T1851061

**ASSIGNMENT-1**

**Part B: Assignments based on R and Python Aim:**

**Perform the following operations using R/Python on the Amazon book review and facebook metrics data sets**

1. **Create data subsets**
2. **Merge Data**
3. **Sort Data**
4. **Transposing Data**
5. **Melting Data to long format**
6. **Casting data to wide format**

**-------------------------------------------------------------------------------------------------------------------**

# Introduction

**What is R?**

* R is a programming language and software environment for statistical analysis, graphics representation and reporting.
* R was created by Ross Ihaka and Robert Gentleman at the University of Auckland, New Zealand, and is currently developed by the R Development Core Team.
* R is freely available under the GNU General Public License, and pre-compiled binary versions are provided for various operating systems like Linux, Windows and Mac.
* This programming language was named R, based on the first letter of first name of the two R authors (Robert Gentleman and Ross Ihaka), and partly a play on the name of the Bell Labs Language S.

The core of R is an interpreted computer language which allows branching and looping as well as modular programming using functions.

* R allows integration with the procedures written in the C, C++, .Net, Python or FORTRAN languages for efficiency.
* R is free software distributed under a GNU-style copy left, and an official part of the GNU project called GNU S.

# Evolution of R

* R was initially written by Ross Ihaka and Robert Gentleman at the Department of Statistics of the University of Auckland in Auckland, New Zealand. R made its first appearance in 1993. – A large group of individuals has contributed to R by sending code and bug reports. – Since mid1997 there has been a core group (the "R Core Team") who can modify the R source code archive.

Features of R

* R is a well-developed, simple and effective programming language which includes conditionals, loops, user defined recursive functions and input and output facilities. • R has an effective data handling and storage facility, • R provides a suite of operators for calculations on arrays, lists, vectors and matrices.
* R provides a large, coherent and integrated collection of tools for data analysis.
* R provides graphical facilities for data analysis and display either directly at the computer or printing at the papers.

# R Studio

* RStudio is a free and open-source integrated development environment (IDE) for R, a programming language for statistical computing and graphics.
* RStudio was founded by JJ Allaire, creator of the programming language ColdFusion. Hadley Wickham is the Chief Scientist at RStudio.
* RStudio is available in two editions: RStudio Desktop, where the program is run locally as a regular desktop application; and RStudio Server, which allows accessing RStudio using a web browser while it is running on a remote Linux server.
* Prepackaged distributions of RStudio Desktop are available for Windows, OS X, and Linux.

# Download R Studio

* Windows: –<https://download1.rstudio.org/RStudio-0.99.893.exe>
* Ubuntu: –<https://download1.rstudio.org/rstudio-0.99.893-i386.deb>
* Fedora: –<https://download1.rstudio.org/rstudio-0.99.893-i686.rpm>
* Linux flavors differentiates 32bit and 64bit as well as .deb and .rpm packages.

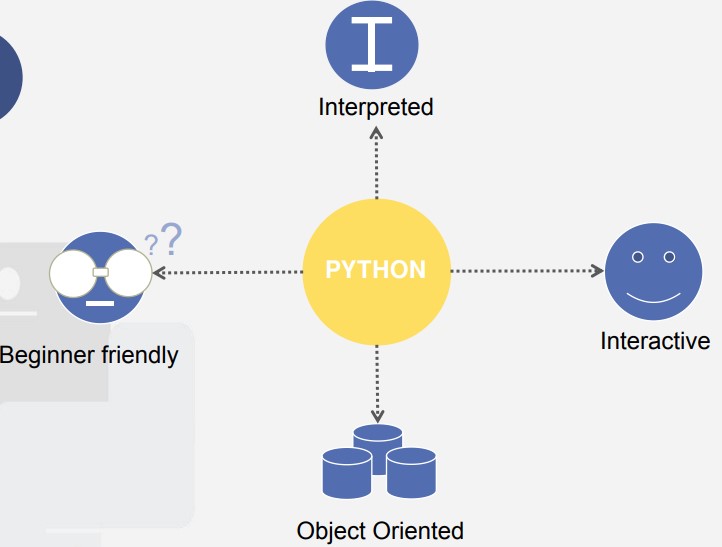
# Python

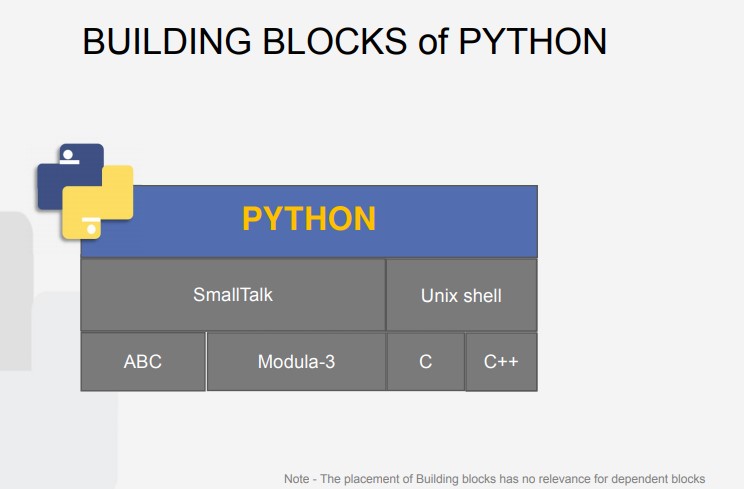
**Python** is an[interpreted](https://en.wikipedia.org/wiki/Interpreted_language)[high-level programming language](https://en.wikipedia.org/wiki/High-level_programming_language)for[general-purpose programming.](https://en.wikipedia.org/wiki/General-purpose_programming_language) Created by[Guido van Rossum](https://en.wikipedia.org/wiki/Guido_van_Rossum)and first released in 1991, Python has a design philosophy that emphasizes[code readability,](https://en.wikipedia.org/wiki/Code_readability) and a[syntax](https://en.wikipedia.org/wiki/Syntax_(programming_languages))that allows programmers to express concepts in fewer[lines of code](https://en.wikipedia.org/wiki/Source_lines_of_code)[,[25][26]](https://en.wikipedia.org/wiki/Python_(programming_language)#cite_note-Summerfield-25)notably using[significant whitespace.](https://en.wikipedia.org/wiki/Significant_whitespace) It provides constructs that enable clear programming on both small and large scales.

## Python Features

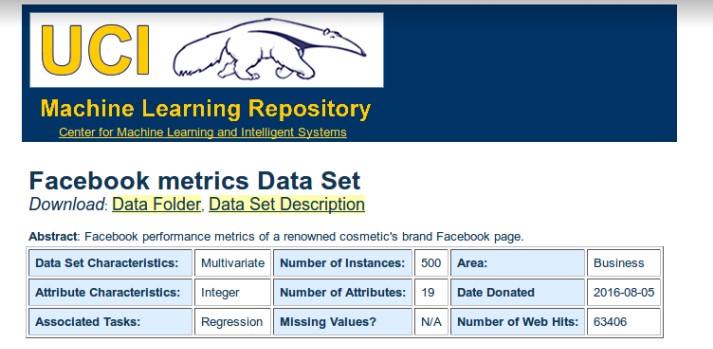
Python's features include −

* **Easy-to-learn** − Python has few keywords, simple structure, and a clearly defined syntax. This allows the student to pick up the language quickly.
* **Easy-to-read** − Python code is more clearly defined and visible to the eyes.
* **Easy-to-maintain** − Python's source code is fairly easy-to-maintain.
* **A broad standard library** − Python's bulk of the library is very portable and crossplatform compatible on UNIX, Windows, and Macintosh.
* **Interactive Mode** − Python has support for an interactive mode which allows interactive testing and debugging of snippets of code.
* **Portable** − Python can run on a wide variety of hardware platforms and has the same interface on all platforms.
* **Extendable** − You can add low-level modules to the Python interpreter. These modules enable programmers to add to or customize their tools to be more efficient.
* **Databases**− Python provides interfaces to all major commercial databases.
* **GUI Programming** − Python supports GUI applications that can be created and ported to many system calls, libraries and windows systems, such as Windows MFC, Macintosh, and the X Window system of Unix.
* **Scalable** − Python provides a better structure and support for large programs than shell scripting.

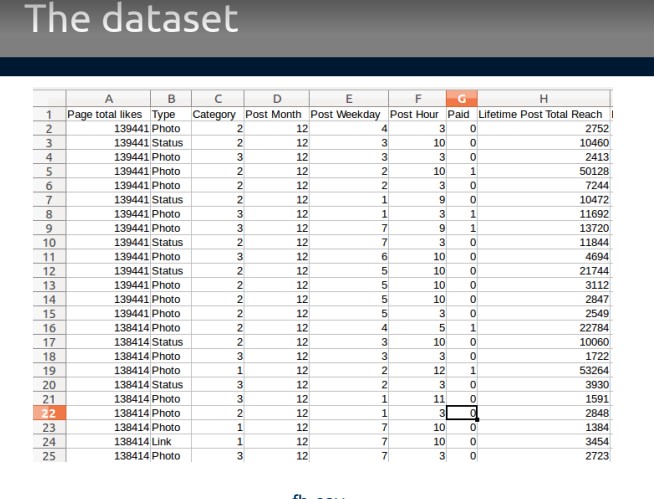




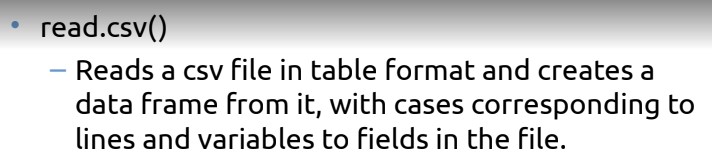
### Assignment Details 1.Download Datasets

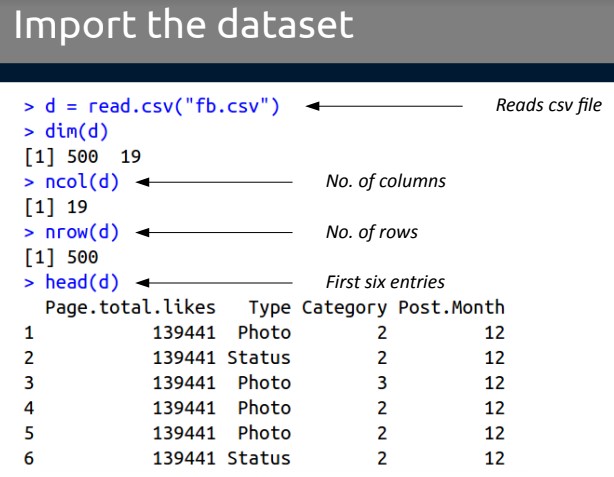


**2.The Dataset:**

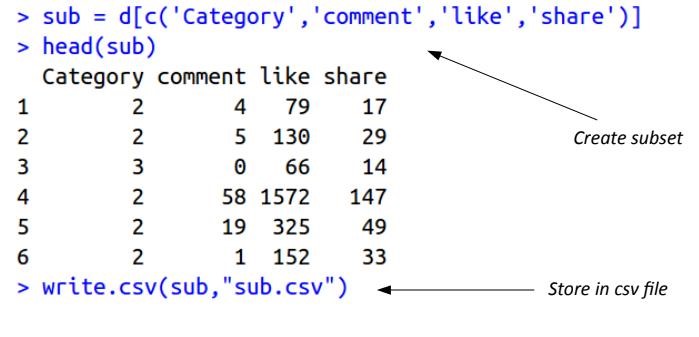


### 3. Read the Downloaded CSV File

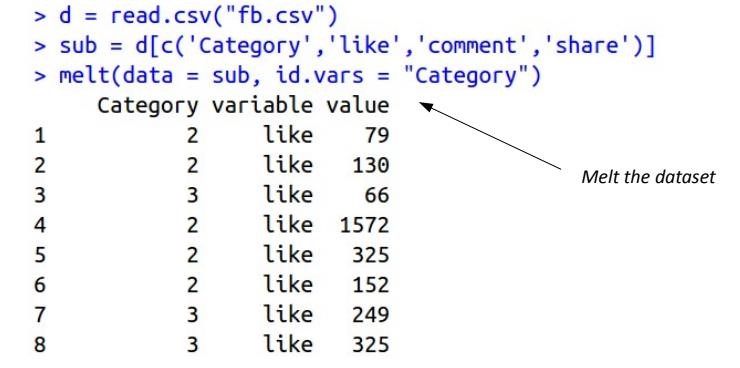




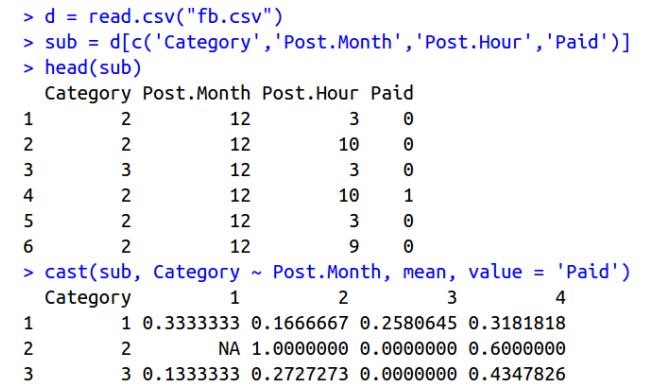
### 3.Create Subset



### 4.Melt Dataset



### 5.Casting Dataset



**Conclusion:** Thus we have learnt various operations of ( Creating data subsets, Merge Data, Sort Data, Transposing Data, Melting Data to long format, Casting data to wide format)with **R Language in RStudio**.