Summer Internship Report

On

**DATA SCIENCE (USING R)**

Submitted to

Amity University Uttar Pradesh



In partial fulfilment of the requirements for the award of the degree of

Bachelor of Technology

in

Computer Science and Engineering

by

**MRINAL JHAMB**

**A2305214017**

**B.TECH CSE+MBA**

**2014-2019**

Under the guidance of

**MR. ARUN PRAKASH Agrawal**

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

AMITY SCHOOL OF ENGINEERING AND TECHNOLOGY

AMITY UNIVERSITY UTTAR PRADESH

NOIDA (U.P.)

# DECLARATION

I, MRINAL JHAMB student of B.Tech CSE + MBA hereby declare that the project title “Data Science (Using R)” which is submitted by me to Department of Computer Science and Engineering, Amity School of Engineering and Technology, Amity University Uttar Pradesh, Noida, in partial fulfilment of requirement for the award of the degree of Bachelor of Technology in Computer Science and Engineering, has not been previously formed the basis for the award of any degree, diploma or other similar title or recognition.

-Mrinal Jhamb

-A2305214017

ii

**CERTIFICATE**

On the basis of declaration submitted by MRINAL JHAMB student of B.Tech CSE + MBA, I hereby certify that the project titled “Data Science (Using R)” which is submitted to Department of Computer Science and Engineering, Amity School of Engineering and Technology, Amity University, Noida, in partial fulfilment of the requirement for the award of the degree of Bachelor of Technology in Computer Science and Engineering is an original contribution with existing knowledge and faithful record of work carried out by him under my guidance and supervision.

To the best of my knowledge this work has not been submitted in any part or full for any degree or diploma to this university or somewhere else.

Date: Mr. Arun Prakash Agrawal

Department of Computer Science and Technology

Amity School of Engineering and Technology

Amity University, Noida, UP

iii

**ACKNOWLEDGEMENT**

I would like to thank Head of our Department Prof Dr. Abhay Bansal for their direction in my summer internship work as an important aspect of the curriculum and constantly guiding me.

Secondly, I would like my project guide Mr. Arun Prakash Agrawal whose guidance I was able to complete my summer internship report. I am highly thankful to him for giving me his valuable time and attention and or providing me various research papers which helped in my research.

My summer internship report has been successful, thanks to all the support staff of my friends and colleagues with gratitude. I wish to acknowledge all of them.

**iv**

**ABSTRACT**

Data science is an interdisciplinary field about scientific methods, processes and systems. It is used either on structured or unstructured data to get insights. It unifies statistics, data analysis and the methods related to them with the aim to understand and analyze actual facts. This porject majorly focuses on building the basic understanding of the flow of data science problem with major focus on EDA and Modelling. It also contains to the report to know why best and most experienced employees are leaving prematurely by looking at the HR\_Data of certain company.

v

Table of Contents

Declaration ii

Certificate iii

Acknowledgement iv

Abstract v

Table of contents vi-vii

[1.Introduction 1](#_Toc493782594)

[1.1 DATA ANALYSIS 1](#_Toc493782595)

[1.2 DATA SCIENCE 1](#_Toc493782596)

[1.3 DATA SCIENCE VS DATA ANALYSIS 2](#_Toc493782597)

[2.Tools & Technologies 3](#_Toc493782598)

[3.Modus Operandi of Data Science 5](#_Toc493782602)

[3.1 Data Acquisition 5](#_Toc493782603)

[3.2 Data Preparation 5](#_Toc493782604)

[3.3 EDA: Exploratory Data Analysis 7](#_Toc493782605)

[3.4 Modeling 18](#_Toc493782606)

[4.Analysis of HR Data 20](#_Toc493782607)

[4.1 Problem statement 20](#_Toc493782608)

[4.2 Data Acquisition 20](#_Toc493782609)

[4.3 Data Preparation 20](#_Toc493782610)

[4.4 EDA 20](#_Toc493782611)

[4.5 Modeling 37](#_Toc493782612)

[5.Conclusion 57](#_Toc493782613)

[5.1 Learning 57](#_Toc493782614)

[5.2 Conclusion 57](#_Toc493782615)

[6.Future Scope 60](#_Toc493782616)

vi

[REFERENCES 61](#_Toc493782617)

**vii**