**PROJECT REPORT**

on

***Data Analytics Using R Programming***

*Submitted in partial fulfillment of requirements for the award of*

**Masters of Computer Application**

In the department of

**Information Technology**

****

*Submitted by*:

**Name of the student (Student ID):**

**Albert Tungoe (CS17MCAGN005)**

**Chandan Sharma (CS17MCAGN001)**

*Under the Supervision of*

**MR. PRANAV KUMAR**

ASSISTANT PROFESSOR

**School of Computing Sciences**

**The Assam Kaziranga University, Jorhat, Assam**

**May 2020**

**SCHOOL OF COMPUTING SCIENCES**

**THE ASSAM KAZIRANGA UNIVERSITY**

**JORHAT-785006 :: ASSAM :: INDIA**

**CERTIFICATE**

This is to certify that the project report entitled ***“Data Analytics Using R Programming”****,* submitted to the School of Computing Sciences (SCS), **THE ASSAM KAZIRANGA UNIVERSITY, JORHAT, ASSAM,** in partialfulfillment for the completion of **Semester – VI** of the degree of **Masters in Computer Application** in the department of **Information Technology**, is a record of bona fide work carried out by **Mr. Albert Tungoe** **Roll No.** **CS17MCAGN005, Mr. Chandan Sharma,** **Roll No**. **CS17MCAGN001** under my supervision and guidance.

All the help received by us from various sources have been duly acknowledged.

No part of this report has been submitted elsewhere for award of any other degree.

----------------------------- -----------------------------

Dean, SCS, KU **Faculty Supervisor**

**Dept. of Information Technology**

**Mr. PRANAV KUMAR**

**ASSITANT PROFESSOR**

**ACKNOWLEDGEMENT**

We would like to express my special thanks of gratitude to my guide Mr. Pranav Kumar who gave us the opportunity to do this research on the topic “*Data Analytics using R Programming*” which helped us a lot in understanding the about R and also many new things we came to know while researching.

This research would not have been done without the invaluable guidance of our guide. He taught us the methodology to carry out the research and to present the research works as clearly as possible.

We would also like to thank our classmates for their encouragement and more over for their timely support and guidance till the completion of our research project work.

We are also extremely grateful to our parents for their love, prayers, caring and sacrifice for educating and preparing us for the future.

**DECLARATION**

We Mr. Albert Tungoe and Mr. Chandan Sharma, declare that all the materials embodied in the project report entitled ***Data Analytics Using R Programming*** submitted to the School of Computing Science (SCS), The Assam Kaziranga University, Jorhat, Assam, by me as a VI semester project in the partial fulfillment of the Masters of Computer Application in the department of Information Technology(IT) is original and no part of the report submitted to any other institution or organization for the award of any degree or diploma. I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact from any other sources in my project report. For any violation of the above facts I shall remain solely responsible.

-------------------------------

**Signature of Student**

**Full Name: ALBERT TUNGOE**

**MCA VI SEMESTER**

**Enrollment No. : CS17MCAGN005**

-------------------------------

**Signature of Student**

**Full Name: CHANDAN SHARMA**

**MCA VI SEMESTER**

**Enrollment No. : CS17MCAGN005**

**Department: Information Technology**

**School: School of Computing Sciences**

**CONTENTS**

Certificate

Acknowledgement i

Declaration ii

Abstract iii

List of Figures iv

1. Introduction 1

2. Literature Review 3

3. Methods/Approaches 7

4. Conclusion 23

Bibliography 26

**ABSTRACT**

Analytics companies has developed the ability to support their decisions through analytic reasoning using a variety of statistical and mathematical techniques. Thomas Davenport in his book titled, “Completing on analytics: The new science of winning”, claims that a significant proportion of high performance companies have high analytical skills among their employees. On the other hand, a recent study has also revealed that that more than 59% of the organizations do not have information required for decision making. Learning “Data analytics using R programming” not only adds to existing analytics knowledge and methodology, but also equips with exposure into latest analytics techniques including forecasting, social media analytics, text mining, etc. It gives an opportunity to work on real time data from social networking sites like twitter, Facebook, Instagram, etc. Here R programming language and environment is used in Statistical computing, data analytics and scientific research.

Keywords: R, Big data, Analytics

.

**List of Figures**

Figure Name Figure Number Page Number

1. R’s Growth Fig 1.1 2

2. Overview of the stages in data Analysis. Fig 3.1 9

3. Processes involved in data collection Fig 3.2 10

4. Steps to extract data from WhatsApp Fig 3.3 10

Using email chat feature

5. Process involved in data transformation. Fig 3.4 12

6. Form design Fig 3.5 13

7. Process involved in data loading. Fig 3.6 14

8. Exploratory Data Analysis and Fig 3.7 15

Visualization.

9. Most frequently used communication Fig 3.8 15

medium.

10. Graphical Representation of Frequently Fig 3.9 16

used Communication medium by Males and Females.

11. Most active day of week Fig 3.10 17

12. Active age groups Fig 3.11 18

13. Frequency of using WhatsApp Fig 3.12 19

group by males and females

14. Total Number of Messages send as Fig 3.13 20

per timestamp.