**STATISTICAL ANALYSIS OF DIET, EXERCISE AND FITNESS VIA PEOPLE’S OWN VISUALISATION OF PARAMETERS**

A project report submitted for the partial fulfilment of Semester-I, Programing for Data Science – Big Data Analytics



Department of Computer Science

Ramakrishna Mission Vivekananda Educational and Research Institute



Submitted By-

Saikat Patra

Ujjwal Chowdhury

Krishnakanta Maity

**Student’s Declaration**

We, hereby declare that, the project work entitled ***“Statistical Analysis of Diet, Exercise and Fitness via people's own visualisation of parameters”*** is a record of an original work done by us under the guidance of Dr. Sudeep Mallick and this project work is submitted in the partial fulfilment of the requirements for the paper Programming for Data Science of Big Data Analytics.

We have collected the data from primary sources as per the requirement of my dissertation, which are appropriately referred in the report. All the computations involved in this dissertation are result of our own calculations on the data that we collected.

The formulae that are used in the dissertation are acknowledged providing appropriate reference of the source from which they are obtained. No part of the dissertation is been submitted to any other institution for the purpose of any degree/diploma.

**Saikat Patra**

**Ujjwal Chowdhury**

**Krishnakanta Maity**

**Date: December 1, 2021**

**Acknowledgement**

It is great pleasure for us to undertake this project. I am grateful to my project guide Dr. Sudeep Mallick.

This project would not have completed without his enormous help and worthy experience. Whenever we were in need, he was there behind us.

Although, this project report has been prepared with utmost care and deep routed interest I accept responsibility for any imperfection.

**Table of Content**

|  |  |
| --- | --- |
|  | Page No |
| Abstract |  |
| Introduction |  |
| Objectives |  |
| Data Set |  |
| Problem Statement |  |
| Questionnaire Design |  |
| Data Collection Methodology |  |
| Data Processing |  |
| Exploratory Data Analysis |  |
| Limitations |  |
| Future Scope |  |
| Reference |  |
| Appendix |  |

**Abstract**

Choices about nutrition and exercise are very much linked with individual’s health. Decisions about food intake also have great impact on an individual. Diet is the sum of food consumed by a person or other organism.

The word diet often implies the use of specific intake of nutrition for health or weight-management reasons. Although humans are omnivores, each person holds some food preferences due to personal tastes. Individual dietary choices may be more or less healthy. A healthy diet can improve and maintain optimal health. Exercise involves engaging in physical activity and increasing the heart rate beyond resting levels. It is an important part of preserving physical and mental health. Whether people engage in light exercise, such as going for a walk, or high intensity activities, for example, weight training. Regular exercise provides a huge range of benefits for the body and mind.

Food preferences are the evaluative attitudes that people express toward foods. Food preferences include the qualitative evaluation of foods, and also how much people like and dislike them.

Height and weight measurements are used to calculate measure of healthy versus unhealthy weights. Food preference and age can also contribute for health differences.

**Introduction**

In our day to day life, especially amidst the Covid situation, our diet and our exercise effect our fitness to a very large level which is very important to live a safe & healthy life. Now our main objective behind this study is to analysis how much our diet and exercise effects our fitness and write a report on it.

**Objectives**

The primary goal of our project is to understand the insight of food consumption, exercise and fitness. The main purpose of EDA is to look at data before making any assumptions. It can help identify obvious errors, as well as better understand patterns within the data, detect outliers or anomalous events, find interesting relations among the variables.

Our approach to measure food habits, fast food composition, physical activity, has greatly motivated our project for understanding of the key determinants and the health problems we currently facing.

**Data Set**

* **Data Source**: All data are collected through google [forms](https://forms.gle/KNmjsd5iNB4cQrwUA).
* **About Data**: This dataset contains 21 variables with timestamps. Description of variables are given below.

**Data Description**

Our data consists of 259 observations with the information of the following variables:

ts: time of the responses

age: age of the respondent

sex: gender of the respondent

work: work preference of the respondent

phy\_ff: rating of liking fast food (in 10 scale)

phy\_health: rating of healthiness (in 10 scale)

phy\_bw: rating of one's preference to maintain body weight

phy\_ex: rating of importance of exercise

meal: number of meals in a day

weight: weight of the respondent

exercise: type of exercises

fruit: number of meals contain fruits

veg: number of meals contain vegetables

cook: number of cooked plates

spend: expenditure on fitness

income: monthly family income

gymtime: time spent in gym

disease: whether suffers from any regular disease

review: review of the survey

rate: rate the project topics and question

**Data Collection Methodology**

We collected the primary data for our project using the self-assignment survey method using google forms. Our targeted audience are mostly UG and PG students.

Our questioner is created considering the dietary recommendations of individuals and their behaviour about exercise and fitness. We connected our targeted audience using Social Media and Instant messaging apps. The time frame for which we collect our data is 3 weeks.

**Data Processing**

**Exploratory Data Analysis**

**Limitations**

Our data set involves a smaller sample, hence the results cannot be accurately interpreted for a generalized population. Most of the inferences are focus on quantitative data that leads to some outliers. Our targeted audiences are mostly around the age 23, so there are some judgmental bias.

Not giving respondents enough options to respond properly as per their choice of others. We should have given a textual input to users so that they could have mentioned if their preferred options was not one of the choices provided by us.

**Future Scope**

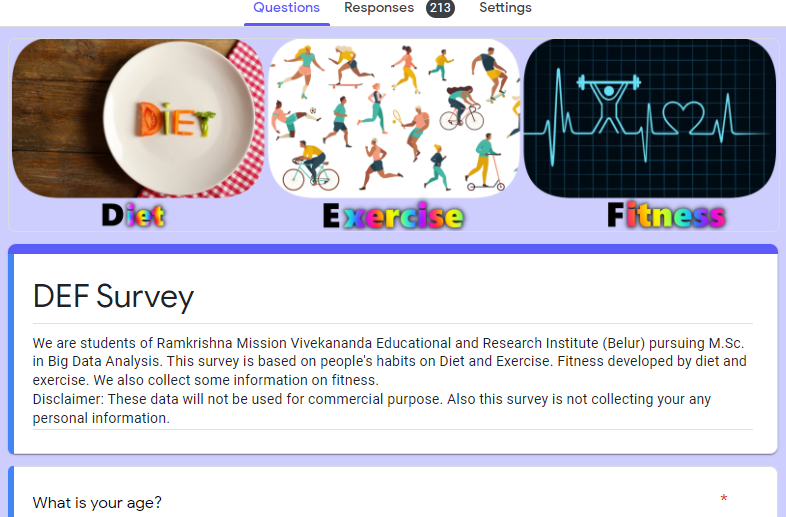
With an ample amount of time and data-set we can create a model which give a clear picture of people’s mind set about diet and fitness. With our analysis, people’s fondness for fast food consumption also can be verified.

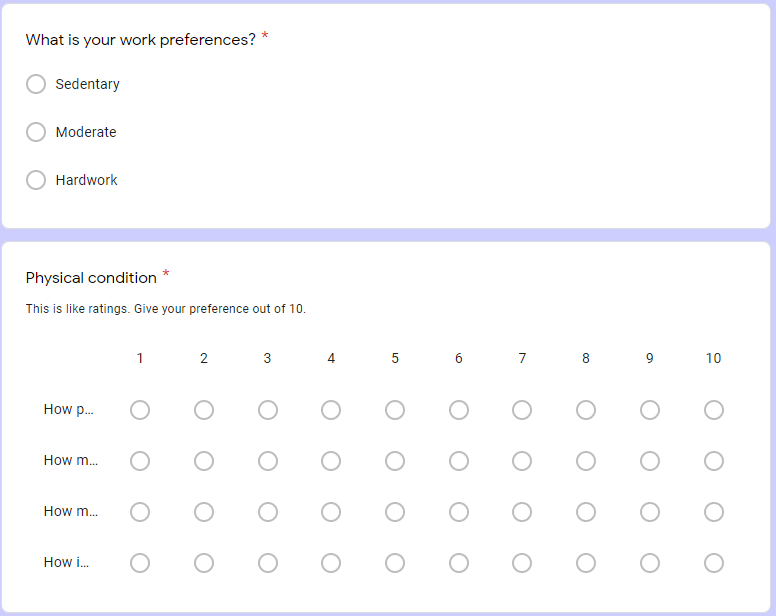
The changing health scenario globally has increased the challenges for public nutrition and people’s behaviour about fitness. With the help of our project we can tackle some of those problems.

**References**

**Appendix**

Google form was created for data collection. Snapshot of questionnaire is attached below:





R-programming is used for exploratory data analysis. All codes are given below: (Codes can be found here)