VISUALIZATION OF EXPLORATORY ANALYSIS ON FITNESS DATASET

MINI PROJECT REPORT

Submitted in partial fulfillment of the requirement for the award of the Degree of

MASTER OF SCIENCE IN

DATA ANALYTICS

Submitted by

YOGESKUMARAN A B

22MDA015

Under the Guidance of

D.SUMATHI, Associate Professor

Since 1988

DEPARTMENT OF COMPUTER SCIENCE

CMS COLLEGE OF SCIENCE & COMMERCE

(AUTONOMOUS)

Re-Accredited at the 'A+' level by NAAC, with CGPA of 3.38 out of 4 (III Cycle)

College with Potential for Excellence – Conferred by UGC, New Delhi

Recognized under Star College of DBT, Ministry of Science and Technology, Govt. of India.

(Affiliated to Bharathiar University)

Coimbatore -640149

November -2023

CERTIFICATE

This is to certify that the Mini Project work entitled

VISUALIZATION OF EXPLORATORY DATA ANALYSIS ON FITNESS DATASET

is a bonafide original record of work done by

YOGESKUMARAN A B

22MDA015

in partial fulfillment of the requirement for the award of the degree of

MASTER OF DATA ANALYTICS

Submitted for viva-voce Examination held on ______________

Countersigned

Project Guide	**Head of the Department**
Internal Examiner	**External Examiner**

DECLARATION

I **"YOGESKUMARAN A B[22MDA015]"** hereby declare that the project work entitled **"VISUALIZATION OF EXPLORATORY DATA ANALYSIS ON FITNESS DATASET"** developed at CMS College of Science & Commerce and submitted to Bharathiar University, Coimbatore in partialfulfillment of the requirements for the award of degree of Master of Data Analytics is a recordof original project work done by me during the period of [2023 – 2024], under the supervisionand guidance of D.SUMATHI MCA,M.Phil.,SET, Associate Professor Department of Computer Science, CMS College of Science & Commerce, Coimbatore.

Place: Coimbatore-49 Signature of the Student

Date : YOGESKUMARAN A B

22MDA015

ACKNOWLEDGEMENT

First of all I express my immense gratitude to Almighty God for the blessings bestowed on us through my project work. grateful to **Dr . H .Balakrishnan**, Principal, CMS College for giving me permission to conduct this study. I wish to express my profound gratitude to **V.Saraswathi** , Head of the Department , Department of Computer Science. I extend my sincere gratitude to **D.Sumathi**, Associate Professor, Department of Computer Science who encouraged and guided me in every stage of my study. I extend to her my deepest sense of gratitude for all her sincere effort to guide me in the best possible way. Last but not the least, I am very thankful to all other faculty members of the Department of Computer Science, CMS college of science and commerce. I am also thankful to my friends and all other well wishers who helped me in one way or other to complete my study.

CONTENTS

ABSTRACT

In today's data driven world,the availability of vast datasets offers valuable insights and opportunities for analysis.The "Exploratory data analysis on fitness dataset" project aims to leverage the power of data to gain a deeper understanding of fitness related information. This project explores the potential of the dataset by employing the descriptive statistics,data visualization using different graphs. Through these analysis we strive to uncover hidden patterns, correlations and predictive insights that can be harnessed to make informed decisions and recommendations in the fitness domain. This project revolves around the various fitness metrics like include calories burn,types of exercise,average weight and the weight they want to achieve etc.

In this collection,there are 12 attributes and 3865 records.
The attributes used in the dataset are;
ID,EXERCISE,CALORIES BURN,DREAM WEIGHT,ACTUAL WEIGHT, AGE,GENDER,DURATION,HEARTRATE,BMI,WEATHER CONDITIONS,EXERCISE INTENSITY.

Moving forward, our objective is to utilize the data visualization tools in excel to create various visual representations in that facilitate them interpretation in the dataset. Through this multidimensional approach,we aim to contribute valuable knowledge to the fitness domain,enabling data driven decision making and improved fitness outcomes.

INTRODUCTION TO DATA ANALYTICS

Data has been the buzzword for ages now. Either the data being generated from large-scale enterprises or the data generated from an individual, each and every aspect of data needs to be analysed to benefit yourself from it. We use Data Analytics for this. Data Analytics has a key role in improving your business as it is used to gather hiddeninsights, generate reports, perform market analysis, and improve business requirements.

Data analytics is the science of raw data analysis to draw conclusions about it. Data Analytics refers to the techniques for analysing data for improving productivity and the profit of the business. Data is extracted and cleaned from different sources to analyse various patterns. Many data analytics techniques and processes are automated into mechanical processes and algorithms which handleraw data for human consumption.

Data analytics is a broad term that encompasses many diverse types of data analysis. Any type of information can be subjected to data analytics techniques to get insight that can be used to improve things. Data analytics techniques can reveal trends and metrics that would otherwise be lost in the mass of information. This information can then be used to optimize processes to increase the overall efficiency of a business or system. Data analytics is important because it helpsbusinesses optimize their performances.

Implementing it into the business model means companiescan help reduce costs by identifying more efficient ways of doing business and by storing large amounts of data. A company can also use data analytics to make better business decisions and help analyse customer trends and satisfaction, which can leadto new—and better—products and services. Data analytics is important because it helps businesses optimize their performances. Implementing it into the business model means companies can help reduce costs by identifying more efficient ways of doing business. A company can also use data analytics to make better business decisions and help analyse customer trends and satisfaction, which can lead to new—and better—products and services.

Data analytics is broken down into four basic types. Descriptive analytics describes what has happened over a given period. Diagnostic analytics focuses more on why something happened. Predictive analytics moves to what is likely going to happen in the near term. Finally, prescriptive analytics suggestsa course of action. Data analytics can do much more than point out bottlenecks in production. Gaming companies use data analytics to set reward schedules for playersthat keep the majority of players active in the game. Content companies use many of the same data analytics to keep you clicking, watching, or re-organizing content to getanother view or another click. Data analytics is the science of raw data analysis to draw conclusions about it. data analytics refers to the techniques for analysing data for improving productivity and the profit of the business.

Data is extracted and cleaned from different sources to analyse various patterns. many data analytics techniques and processes are automated into mechanical 2processes and algorithms which handle raw data for human consumption. As data is becoming more prominent by the minute, organizations are becoming data-driven, which means adopting methods to collect more data. This data is then sorted, stored, and then analysed to derive logical and valuable information. Data analytics makes the process possible.

TYPES OF DATA:

STRUCTURED DATA

Structured data is the data which conforms to a data model, has a well define structure, follows a consistent order and can be easily accessed and used by a personor a computer program. Structured data is usually stored in well-defined schemas such as Databases. It is generally tabular with column and rows that clearly define itsattributes. SQL (Structured Query language) is often used to manage structured datastored in databases.

SEMI-STRUCTURED DATA

Semi-structured data is data that does not conform to a data model but has some structure. It lacks a fixed or rigid schema. It is the data that does not reside in a rationaldatabase but that have some organizational properties that make it easier to analyze .With some processes,

we can store them in the relational database.

- Data does not conform to a data model but has some structure.
- Data cannot be stored in the form of rows and columns as in Databases
- Semi-structured data contains tags and elements (Metadata) which is used togroup

data and describe how the data is stored

- Similar entities are grouped together and organized in a hierarchy
- Entities in the same group may or may not have the same attributes orproperties
- Does not contain sufficient metadata which makes automation and management of
- data difficult
- Size and type of the same attributes in a group may differ
- Due to lack of a well-defined structure, it cannot used by computer programseasily

UNSTRUCTURED DATA

Unstructured data is the data which does not conforms to a data model and has no easily identifiable structure such that it cannot be used by a computer program easily.Unstructured data is not organized in a pre-defined manner or does not have a pre- defined data model, thus it is not a good fit for a mainstream relational database.

- Data neither conforms to a data model nor has any structure.
- Data cannot be stored in the form of rows and columns as in Databases
- Data does not follow any semantic or rules
- Data lacks any particular format or sequence
- Data has no easily identifiable structure
- Due to lack of identifiable structure, it cannot used by computer programseasily

MACHINE LEARNING

Machine Learning is the field of study that gives computers the capability to learn without being explicitly programmed. ML is one of the most exciting technologies thatone would have ever come across. As it is evident from the name, it gives the computerthat makes it more similar to humans: The ability to learn.

Machine learning is actively being used today, perhaps inmany more places than one would expect. Compared to what can be done today, thisfeat seems trivial, but it's considered a major milestone in the field of artificialintelligence. Over the last couple of decades, the technological advances in storage and processing power have enabled some innovative products based on machine learning, such as Netflix's recommendation engine and self-driving cars.

Machine learning is an important component of the growing field of data science. Through the use of statistical methods, algorithms are trained to make classifications or predictions, and to uncover key insights in data mining projects.

These insights subsequently drive decision making within applications and businesses, ideally impacting key growth metrics. As big data continues to expand and grow, the market demand for data scientists will increase. They will be required to help identify the most relevant business questions and the data to answer them. Machine learning algorithms are typically created using frameworks that accelerate solution development, such as TensorFlow and PyTorch.

FITNESS DATASET

In an age where health and well-being have taken center stage in our lives, understanding the intricate relationship between fitness, exercise, and various health metrics has become more critical than ever. The rise of wearable fitness technology, fitness tracking apps, and health-conscious lifestyles has generated a wealth of data related to individual fitness journeys and outcomes. The Fitness Dataset, which serves as the focal point of this research and analysis, represents a treasure trove of valuable information, shedding light on the diverse facets of personal fitness and its impact on our overall health.

As we delve into this dataset, we embark on a journey to uncover the patterns, correlations, and insights hidden within the numbers. This exploration can help us better comprehend the nuanced interplay between factors like exercise routines, dietary choices, physiological measurements, and even socio-demographic variables. Such insights have far-reaching implications for individuals, healthcare professionals, fitness enthusiasts, and policymakers alike, as they offer the promise of enhancing our ability to make informed decisions about our health and well-being.

The Fitness Dataset encompasses a wide array of variables, capturing a holistic view of fitness progress and outcomes. It includes measurements related to physical activity, such as step counts, exercise duration, and calorie expenditure. Additionally, physiological parameters like heart rate, blood pressure, and body mass index (BMI) are included, providing a comprehensive snapshot of individuals' health status. The dataset also comprises data related to dietary habits, sleep patterns, and self-reported well-being, enabling a multifaceted analysis of an individual's lifestyle choices and their impact on fitness and health.

While the dataset itself is a valuable resource, its significance lies in its potential applications. Researchers can harness this data to advance our understanding of fitness trends, predictors of health outcomes, and the efficacy of different exercise and dietary regimes. Healthcare practitioners can utilize these insights to provide more personalized recommendations, fostering a healthier patient population. Fitness enthusiasts can gain a deeper understanding of their own progress and tailor their fitness routines to maximize their well-being.

As we embark on our exploration of the Fitness Dataset, it is essential to approach the analysis with a multidisciplinary perspective. Data science, statistics, and domain knowledge in fitness and healthcare will converge to unearth the underlying stories within the numbers. This dataset represents an opportunity to empower individuals to take control of their fitness journeys, while also contributing to the broader conversation about health and well-being on a societal level.

In the pages that follow, we will navigate the intricate pathways of the Fitness Dataset, peeling back the layers of data to reveal the insights and revelations it holds. Whether you are a data scientist, a healthcare provider, or simply someone passionate about fitness, this dataset has the potential to shape the way we think about our health, providing a roadmap to a healthier, more informed future.

DATASET OF EXPLORATORY ANALYSIS

ID	Exercise	Calories Burn	Dream Weight	Actual Weight	Age	Gender	Duration	Heart Rate	BMI	Weather Conditions	Exercise Intensity
1	Exercise 2	286.9598505	91.89253067	96	45	Male	37	170	29.4	Rainy	5
2	Exercise 7	343.4530361	64.16509681	61	25	Male	43	142	21.3	Rainy	5
3	Exercise 4	261.2234649	70.84622352	72	20	Male	20	148	27.9	Cloudy	4
4	Exercise 5	127.1838584	79.47700756	83	33	Male	39	170	33.7	Sunny	10
5	Exercise 10	416.3183735	89.96022608	86	29	Female	34	118	23.3	Cloudy	3
6	Exercise 1	479.7226901	78.88757763	81	60	Female	41	169	34.7	Rainy	10
7	Exercise 9	457.6313614	65.68112526	62	18	Male	53	103	34.6	Cloudy	10
8	Exercise 4	272.9569819	64.92956278	63	42	Male	25	104	22.1	Cloudy	2
9	Exercise 10	195.0322726	52.73106505	55	49	Male	37	161	30.9	Sunny	1
10	Exercise 8	259.5311447	95.164097	97	41	Male	55	103	31.2	Cloudy	10
11	Exercise 5	248.5361176	56.8297759	54	41	Male	52	151	34.0	Cloudy	3
12	Exercise 1	376.5526493	95.19628309	97	35	Female	28	158	34.6	Sunny	6
13	Exercise 1	311.1060239	83.67915156	84	54	Female	48	173	34.5	Rainy	4
14	Exercise 8	497.549588	59.42012239	60	41	Male	43	121	28.9	Sunny	1
15	Exercise 10	451.8532602	65.27981192	61	19	Male	59	174	26.3	Cloudy	4
16	Exercise 7	325.056217	89.07299305	86	54	Male	42	105	29.0	Sunny	9
17	Exercise 2	216.8663933	84.63807523	88	38	Female	46	119	33.7	Rainy	7
18	Exercise 10	377.1229305	79.64940102	84	42	Male	26	126	23.9	Sunny	9
19	Exercise 1	457.6203214	70.2601051	71	40	Female	23	129	33.2	Sunny	2
20	Exercise 9	103.3354905	88.75086513	84	49	Male	59	149	20.4	Rainy	8
21	Exercise 7	335.1512709	84.92312857	81	38	Male	55	114	24.9	Sunny	7
22	Exercise 9	131.4575318	54.97493556	52	29	Male	37	135	20.5	Sunny	2
23	Exercise 1	187.2145211	70.6800884	70	54	Female	26	117	30.6	Rainy	3
24	Exercise 7	129.0192376	58.10413954	55	30	Male	26	145	33.4	Sunny	9
25	Exercise 1	293.464795	60.08372848	60	45	Female	36	135	18.7	Sunny	7
26	Exercise 10	344.1352887	75.58965322	72	24	Male	57	159	31.6	Cloudy	2
27	Exercise 9	297.6813903	79.36560717	82	56	Male	39	133	20.0	Sunny	1
28	Exercise 10	126.4961165	93.80578537	99	54	Male	33	118	18.9	Cloudy	6

27	Exercise 9	297.6813903	79.36560717	82	56	Male	39	133	20.0	Sunny	1
28	Exercise 10	126.4961165	93.80578537	99	54	Male	33	118	18.9	Cloudy	6
29	Exercise 8	279.7251573	76.76067757	79	21	Male	21	164	21.3	Rainy	10
30	Exercise 6	214.2193561	50.81220347	49	55	Female	48	155	29.2	Sunny	1
31	Exercise 4	115.3368677	87.03452476	86	47	Male	26	154	22.2	Sunny	7
32	Exercise 5	444.6904555	82.72059543	86	24	Male	37	168	24.4	Cloudy	2
33	Exercise 1	261.0427782	87.93811297	91	57	Female	47	111	30.4	Rainy	8
34	Exercise 1	343.2067297	80.10790122	78	30	Female	31	170	24.9	Sunny	4
35	Exercise 8	483.4970362	75.61763231	78	52	Female	49	121	26.3	Sunny	9
36	Exercise 8	364.4688538	96.88634887	96	20	Male	43	103	24.8	Sunny	7
37	Exercise 9	353.2051257	54.24162087	55	37	Male	51	155	34.4	Rainy	10
38	Exercise 7	182.3335938	96.92153758	96	47	Female	25	106	28.2	Sunny	5
39	Exercise 8	316.0366931	76.44365295	80	28	Female	31	157	21.8	Sunny	1
40	Exercise 3	280.5911131	99.79883357	99	57	Female	27	150	32.3	Cloudy	4
41	Exercise 3	240.9247871	98.84875202	101	41	Female	53	162	27.4	Cloudy	5
42	Exercise 4	455.6056505	90.00990748	85	19	Male	60	153	22.2	Cloudy	1
43	Exercise 1	119.4725546	53.20619588	51	32	Female	37	168	19.4	Rainy	9
44	Exercise 9	121.0886855	60.46042111	58	55	Male	36	159	31.8	Cloudy	7
45	Exercise 9	207.9759501	67.72806592	71	39	Female	59	177	30.7	Sunny	2
46	Exercise 2	355.6478253	78.55073371	77	32	Female	28	126	23.0	Rainy	2
47	Exercise 10	176.4886037	76.97485323	75	58	Male	36	151	26.2	Rainy	7
48	Exercise 7	453.4353571	62.04763518	67	40	Male	52	158	20.1	Cloudy	6
49	Exercise 6	120.0669036	86.6758351	83	48	Female	34	174	25.8	Sunny	2
50	Exercise 3	379.3709879	84.14863746	89	36	Male	54	137	30.0	Cloudy	2
51	Exercise 2	244.9790358	82.43364998	78	35	Male	21	129	20.1	Rainy	7
52	Exercise 10	447.7851241	57.90317356	57	28	Male	39	124	24.5	Rainy	3
53	Exercise 9	117.819875	74.29894484	71	37	Male	32	146	19.0	Rainy	3
54	Exercise 1	217.7570965	77.63226847	82	25	Female	50	136	33.6	Cloudy	7
55	Exercise 6	371.1216261	76.78404005	82	51	Male	52	146	34.3	Sunny	3

54	Exercise 1	217.7570965	77.63226847	82	25	Female	50	136	33.6	Cloudy	7
55	Exercise 6	371.1216261	76.78404005	82	51	Male	52	146	34.3	Sunny	3
56	Exercise 9	441.770923	74.27839097	73	28	Female	20	108	21.1	Rainy	1
57	Exercise 5	277.7292176	93.72924375	91	27	Female	53	103	29.9	Sunny	5
58	Exercise 9	168.6552217	96.99129988	92	52	Male	38	136	20.2	Cloudy	8
59	Exercise 8	126.6017619	76.93070857	76	49	Female	26	112	22.8	Sunny	1
60	Exercise 2	283.2182582	65.65555298	63	48	Male	21	101	34.1	Cloudy	8
61	Exercise 9	301.4196335	57.83633874	61	40	Male	48	119	18.6	Rainy	6
62	Exercise 7	318.4410057	76.49824545	80	47	Female	21	164	19.2	Cloudy	4
63	Exercise 6	141.0364555	87.40573433	83	40	Female	25	156	28.2	Cloudy	10
64	Exercise 1	170.5211824	65.45832237	68	26	Male	39	174	34.3	Cloudy	2
65	Exercise 8	151.3656289	85.21196044	84	60	Female	35	121	26.4	Cloudy	7
66	Exercise 9	365.6369757	76.6660257	78	20	Male	30	151	31.1	Cloudy	4
67	Exercise 3	127.9066818	78.8011706	79	52	Male	52	172	21.7	Cloudy	2
68	Exercise 1	111.4136934	71.71853757	75	50	Male	23	147	21.9	Rainy	4
69	Exercise 6	455.756085	81.1732508	82	36	Male	40	118	22.5	Cloudy	1
70	Exercise 4	324.9529154	60.94100863	62	40	Female	41	171	23.8	Sunny	10
71	Exercise 9	224.7566667	53.58722902	54	51	Male	26	139	28.6	Rainy	7
72	Exercise 4	251.9707701	79.66824254	76	19	Female	34	107	25.4	Sunny	2
73	Exercise 9	346.4365542	99.07569469	96	32	Female	49	127	27.4	Sunny	8
74	Exercise 3	233.3392462	55.8392404	52	58	Male	30	153	24.9	Cloudy	6
75	Exercise 6	222.7406397	97.72148267	100	25	Male	53	166	19.6	Sunny	4
76	Exercise 8	254.9823965	78.1162371	77	60	Male	27	128	29.8	Sunny	2
77	Exercise 10	387.9547256	81.0726531	82	42	Female	34	118	27.8	Rainy	4
78	Exercise 1	129.4225026	92.55242138	95	59	Female	51	176	33.0	Sunny	9
79	Exercise 10	211.9936538	98.59539471	96	52	Female	36	125	32.4	Cloudy	3
80	Exercise 6	187.1540026	95.02492746	96	43	Female	28	146	19.1	Rainy	2
81	Exercise 4	362.8855991	65.77778426	61	52	Male	44	145	32.3	Rainy	9
82	Exercise 1	145.4125641	69.8404642	73	27	Male	42	110	26.7	Rainy	10

80	Exercise 6	187.1540026	95.02492746	96	43	Female	28	146	19.1	Rainy	2
81	Exercise 4	362.8855991	65.77778426	61	52	Male	44	145	32.3	Rainy	9
82	Exercise 1	145.4125641	69.8404642	73	27	Male	42	110	26.7	Rainy	10
83	Exercise 1	305.3738804	97.58691919	102	56	Female	46	103	22.1	Rainy	2
84	Exercise 4	472.5085347	61.69187834	63	33	Female	45	143	31.6	Cloudy	10
85	Exercise 1	364.7305563	65.50027596	64	57	Male	24	176	18.9	Cloudy	1
86	Exercise 3	218.439501	91.59064132	89	25	Female	55	156	21.6	Rainy	2
87	Exercise 7	120.509324	66.37023873	62	55	Male	46	110	34.0	Cloudy	7
88	Exercise 4	194.3341766	55.45791564	57	18	Female	42	121	28.7	Sunny	4
89	Exercise 9	169.705561	53.7830683	55	40	Male	51	152	21.0	Sunny	6
90	Exercise 6	178.0728415	80.1558737	83	43	Female	52	108	27.2	Rainy	7
91	Exercise 9	299.8187299	67.94857152	65	28	Male	32	137	24.3	Sunny	8
92	Exercise 10	443.5856836	53.55212315	50	57	Male	40	139	33.3	Rainy	5
93	Exercise 4	271.0956442	86.20290906	84	36	Female	22	180	24.7	Cloudy	3
94	Exercise 6	413.8854243	66.92549295	69	35	Male	36	109	22.1	Sunny	5
95	Exercise 9	176.433223	54.41506976	59	35	Male	27	166	23.5	Cloudy	8
96	Exercise 10	147.1726527	69.72905757	71	40	Female	56	115	28.8	Rainy	6
97	Exercise 3	408.4146909	87.41431198	88	34	Female	23	154	21.5	Sunny	7
98	Exercise 9	168.0208161	97.42627824	99	57	Female	52	107	30.6	Cloudy	5
99	Exercise 2	267.576432	90.12892254	86	41	Male	46	163	32.0	Cloudy	7
100	Exercise 3	366.0372045	79.75710513	75	37	Male	38	115	34.3	Cloudy	10
101	Exercise 5	309.7996685	76.27469891	79	42	Male	57	128	30.2	Sunny	1
102	Exercise 3	377.2902571	50.70619304	51	27	Male	50	166	21.6	Sunny	8
103	Exercise 6	306.2712633	54.14344982	57	53	Female	20	152	25.6	Rainy	3
104	Exercise 9	234.0254065	71.31793923	68	56	Female	30	106	33.2	Rainy	1
105	Exercise 4	329.4466802	56.68192287	61	38	Male	46	108	24.5	Cloudy	9
106	Exercise 6	442.5630513	52.05940184	56	22	Male	51	176	32.7	Sunny	7
107	Exercise 3	399.4897422	81.84193836	81	28	Male	28	151	27.6	Rainy	7
108	Exercise 10	483.969752	94.31314947	96	54	Male	51	127	21.4	Cloudy	4

106	Exercise 6	442.5630513	52.05940184	56	22	Male	51	176	32.7	Sunny	7
107	Exercise 3	399.4897422	81.84193836	81	28	Male	28	151	27.6	Rainy	7
108	Exercise 10	483.969752	94.31314947	96	54	Male	51	127	21.4	Cloudy	4
109	Exercise 5	420.1160418	78.95308602	80	42	Female	58	125	27.3	Sunny	5
110	Exercise 4	209.8023499	54.2371572	59	18	Male	39	123	34.1	Cloudy	8
111	Exercise 4	124.8124867	57.00863719	58	57	Female	60	109	23.2	Cloudy	8
112	Exercise 7	118.3385206	84.5444396	83	47	Female	38	141	31.4	Sunny	6
113	Exercise 8	408.4920534	93.12616649	94	58	Female	22	145	32.1	Rainy	5
114	Exercise 7	118.1771261	99.5113162	103	43	Male	22	151	28.7	Cloudy	5
115	Exercise 1	422.7135013	60.69999543	63	58	Female	22	108	20.2	Cloudy	2
116	Exercise 7	240.0177788	53.32739946	56	23	Male	42	135	20.1	Cloudy	7
117	Exercise 9	419.3066769	85.90666573	89	49	Female	41	101	26.4	Rainy	5
118	Exercise 10	483.7782894	56.35911615	57	39	Female	26	163	21.4	Cloudy	4
119	Exercise 7	266.54832	91.46747422	93	34	Male	52	144	22.3	Cloudy	1
120	Exercise 2	104.8550273	63.53499848	67	19	Male	39	122	20.3	Cloudy	2
121	Exercise 2	302.9548246	78.63395151	83	32	Female	37	103	27.5	Sunny	4
122	Exercise 5	283.4472034	88.55364482	90	43	Female	54	118	24.9	Rainy	7
123	Exercise 10	239.5529794	92.31470194	91	57	Female	33	173	31.4	Rainy	7
124	Exercise 5	410.136624	56.9317965	56	44	Female	29	105	28.3	Sunny	2
125	Exercise 1	472.2571261	73.32088288	70	34	Female	50	146	22.9	Rainy	10
126	Exercise 9	440.3696002	57.83001371	53	40	Female	44	169	28.1	Cloudy	4
127	Exercise 6	475.4189059	52.11344316	51	57	Female	36	161	22.2	Sunny	2
128	Exercise 10	142.9766065	97.24472774	93	59	Female	25	116	22.3	Rainy	9
129	Exercise 3	115.8819824	95.96800941	99	18	Female	50	144	32.0	Sunny	4
130	Exercise 5	443.3890037	80.83581477	84	39	Male	41	177	24.1	Rainy	9
131	Exercise 5	187.0619427	61.66302264	62	58	Female	27	130	29.3	Sunny	7
132	Exercise 6	310.3643974	82.73871276	85	43	Female	23	124	29.7	Sunny	3
133	Exercise 7	129.611367	59.36482569	55	30	Male	28	117	30.5	Cloudy	2
134	Exercise 6	383.2635156	56.26033451	61	25	Female	28	104	21.6	Cloudy	10

172	Exercise 5	271.7984607	96.23437792	92	18	Male	55	159	19.6	Cloudy	3
173	Exercise 5	218.1117723	83.12915796	87	35	Female	54	131	32.3	Sunny	6
174	Exercise 5	372.581496	97.81003857	94	54	Male	31	158	33.3	Sunny	1
175	Exercise 6	479.8325592	91.36767018	96	49	Female	54	107	27.2	Cloudy	2
176	Exercise 9	394.1783809	51.52025874	48	56	Male	31	123	22.7	Sunny	10
177	Exercise 2	444.2549306	96.93743153	96	24	Female	21	152	31.0	Sunny	3
178	Exercise 7	317.3206578	50.68907642	47	37	Female	36	166	23.0	Cloudy	1
179	Exercise 4	294.440263	94.2268042	96	20	Female	33	172	32.3	Rainy	10
180	Exercise 9	336.160237	81.85502061	78	51	Female	40	145	25.6	Rainy	5
181	Exercise 8	179.1368748	96.11824221	94	21	Male	24	126	33.4	Sunny	6
182	Exercise 8	241.2112347	59.20834713	56	31	Female	35	166	20.5	Rainy	5
183	Exercise 10	468.6631805	68.2300733	65	38	Female	30	114	28.0	Rainy	8
184	Exercise 3	370.6934327	81.27285503	78	25	Female	31	144	31.5	Rainy	9
185	Exercise 5	235.4808427	62.75129513	64	18	Male	29	130	19.2	Cloudy	3
186	Exercise 7	497.6911217	71.99043663	69	27	Female	34	155	25.6	Sunny	7
187	Exercise 7	461.5139356	71.32622876	67	52	Female	45	102	33.8	Rainy	10
188	Exercise 2	358.1260719	97.29009016	101	49	Male	27	165	31.1	Sunny	9
189	Exercise 2	420.185107	78.51290543	79	48	Male	35	119	34.2	Rainy	2
190	Exercise 10	247.3177278	66.85790055	67	58	Female	43	136	27.7	Sunny	4
191	Exercise 1	304.6376978	59.58972611	63	60	Male	37	171	22.6	Cloudy	10
192	Exercise 5	485.9618233	76.38874557	74	42	Male	25	113	20.3	Rainy	4
193	Exercise 3	452.3517684	94.86783143	93	28	Female	23	141	28.7	Rainy	4
194	Exercise 5	430.4024765	63.48585996	60	48	Male	47	123	25.2	Sunny	7
195	Exercise 1	283.3606445	75.1904096	78	51	Female	41	171	20.5	Sunny	4
196	Exercise 2	212.742105	79.76207885	83	53	Female	34	117	33.3	Rainy	5
197	Exercise 7	497.8515018	68.93369316	66	25	Female	47	110	34.4	Sunny	2
198	Exercise 1	390.1075923	64.75229773	68	31	Male	45	160	26.1	Sunny	4
199	Exercise 1	214.5143714	63.83401715	68	34	Male	21	159	26.6	Cloudy	10
200	Exercise 7	188.6152102	96.73989768	100	43	Male	55	173	28.2	Rainy	9

134	Exercise 6	383.2635156	56.26033451	61	25	Female	28	104	21.6	Cloudy	10
135	Exercise 4	144.7993694	76.41797628	79	52	Female	53	177	21.3	Rainy	1
136	Exercise 5	102.0959253	91.53400426	92	55	Female	50	106	33.4	Sunny	6
137	Exercise 10	344.712567	61.98804603	59	43	Male	59	108	19.1	Sunny	1
138	Exercise 4	237.6189713	82.4925542	87	29	Male	23	137	24.9	Sunny	1
139	Exercise 10	192.364994	95.81706922	100	40	Male	31	170	20.0	Rainy	2
140	Exercise 9	471.6148446	79.65763617	76	29	Male	58	160	21.5	Cloudy	5
141	Exercise 5	280.455236	75.32336965	75	36	Female	58	131	33.7	Rainy	7
142	Exercise 1	474.2340328	65.58163045	66	26	Female	59	107	24.4	Sunny	10
143	Exercise 8	184.4730934	64.73910381	68	46	Female	45	141	28.9	Cloudy	4
144	Exercise 5	137.7745595	51.05534575	54	22	Male	42	109	26.3	Sunny	7
145	Exercise 3	414.4423123	96.5655895	98	44	Female	31	121	27.5	Cloudy	3
146	Exercise 4	439.2877189	95.0383254	99	44	Male	40	106	28.4	Rainy	8
147	Exercise 1	152.3731861	70.97200523	74	28	Male	28	100	28.6	Cloudy	3
148	Exercise 5	313.1435338	82.27331216	85	52	Male	48	126	27.6	Cloudy	5
149	Exercise 9	376.4953334	85.24942967	83	26	Female	32	109	19.4	Rainy	5
150	Exercise 9	448.9890478	92.22488557	95	35	Male	30	141	22.9	Sunny	6
151	Exercise 9	293.0344535	55.50903185	53	56	Male	20	176	29.8	Cloudy	1
152	Exercise 3	233.122857	94.29167377	90	38	Female	22	105	21.3	Cloudy	8
153	Exercise 9	151.5991295	54.17946007	57	23	Female	43	167	31.3	Sunny	5
154	Exercise 1	246.0202278	62.12980391	62	27	Female	23	104	22.5	Cloudy	7
155	Exercise 4	295.1513244	98.9389131	96	50	Female	36	178	24.5	Cloudy	1
156	Exercise 9	119.6067548	67.81632892	71	23	Male	25	170	18.9	Cloudy	5
157	Exercise 5	418.2618341	69.30636197	69	54	Male	23	177	23.6	Cloudy	2
158	Exercise 9	419.7180341	74.24094884	76	38	Female	51	151	26.5	Sunny	6
159	Exercise 7	237.1214052	70.27505826	75	45	Male	26	154	24.5	Cloudy	9
160	Exercise 4	488.9969371	88.25870267	84	19	Female	55	175	19.5	Cloudy	3
161	Exercise 3	485.0860445	72.31580813	69	60	Male	37	133	30.2	Sunny	3
162	Exercise 9	185.5255732	80.02737334	81	60	Male	56	111	25.1	Cloudy	7

DATA NORMALIZATION

Normalization is the process of reorganizing data in a database so that it meets twobasic requirements:

1. There is no redundancy of data, all data is stored in only one place.
2. Data dependencies are logical ,all related data items are stored together.

Normalization is important for many reasons, but chiefly because it allows databasesto take up as little disk space as possible, resulting in increased performance.

Normalization is also known as data normalization.

DATA NORMALIZATION TECHNIQUES:

If you're new to data science and machine learning, you've certainly questioned a lot about what feature normalization in machine learning is and how it works.The

most widely used types of normalization in machine learning are:

Min-Max Scaling – Subtract the minimum value from each column's highest value and divide by the range. Each new column has a minimum value of 0 and a maximumvalue of 1.

Standardization Scaling – The term standardization refers to the process of centering a variable at zero and standardizing the variance at one. Subtracting the mean of each observation and then dividing by the standard deviation is the procedure:

The features will be rescaled so that they have the attributes of a typical normal distribution with standard deviations.

DATA PREPROCESSING

Data preprocessing, a component of data preparation, describes any type of processing performed on raw data to prepare it for another data processing procedure. It has traditionally been an important preliminary step for the data mining process. More recently, data preprocessing techniques have been adapted for training machine learning models and AI models and for running inferences against them.

Data preprocessing transforms the data into a format that is more easily and effectively processed in data mining, machine learning and other data science tasks. The techniques are generally used at the earliest stages of the machine learning and AI development pipeline to ensure accurate results.

There are several different tools and methods used for preprocessing data, including the following:

- sampling, which selects a representative subset from a large population of data;
- transformation, which manipulates raw data to produce a single input;
- denoising, which removes noise from data;
- imputation, which synthesizes statistically relevant data for missing values;
- normalization, which organizes data for more efficient access; and
- feature extraction, which pulls out a relevant feature subset that is significant in a particular context.

These tools and methods can be used on a variety of data sources, including data stored in files or databases and streaming data.

DATA PRE PROCESSING STEPS

The steps used in data pre processing include the following:

1. Data profiling. Data profiling is the process of examining, analysing and reviewingdata to collect statistics about its quality. It starts with a survey of existing data and itscharacteristics. Data scientists identify data sets that are pertinent to the problem at hand, inventory its significant attributes, and form a hypothesis of features that mightbe relevant for the proposed analytics or machine learning task. They also relate datasources to the relevant business concepts and consider which pre processing librariescould be used.

2. Data cleansing. The aim here is to find the easiest way to rectify quality issues, suchas eliminating bad data, filling in missing data or otherwise ensuring the raw data is suitable for feature engineering.

3. Data reduction. Raw data sets often include redundant data that arise from characterizing phenomena in different ways or data that is not relevant to a particularML, AI or analytics task. Data reduction uses techniques like principal component analysis to transform the raw data into a simpler form suitable for particular use cases.

4. Data transformation. Here, data scientists think about how different aspects of thedata need to be organized to make the most sense for the goal. This could include things like structuring unstructured data, combining salient variables when it makes sense or identifying important ranges to focus on.

5. Data enrichment. In this step, data scientists apply the various feature engineeringlibraries to the data to effect the desired transformations. The result should be a data set organized to achieve the optimal balance between the training time for a new model and the required compute.

6. Data validation. At this stage, the data is split into two sets. The first set is used to train a machine learning or deep learning model. The second set is the testing data that is used to gauge the accuracy and robustness of the resulting model. This second step helps identify any problems in the hypothesis used in the cleaning and feature engineering of the data. If the data scientists are satisfied with the results, they can push pre processing task to a data engineer who figures out how to scale it for production. If not, the data scientists can go back and make changes to the way they implemented the data cleansing and feature engineering steps.

The above data containing global suicide scale is already pre processed. It does not contain any null values and duplicated values. All the dataset are pre processed.
Hence there is no need for further pre processing the data

CLASSIFICATION

The Classification algorithm is a Supervised Learning technique that is used toidentify the category of new observations on the basis of training data. In Classification, a program learns from the given dataset or observations and then classifies new observation into a number of classes or groups. Such as, Yes or No, 0 or1, Spam or Not Spam, cat or dog**,** etc. Classes can be called as targets/labels or categories.

Unlike regression, the output variable of Classification is a category, not a value, suchas "Green or Blue", "fruit or animal", etc. Since the Classification algorithm is a Supervised learning technique, hence it takes labelled input data, which means it contains input with the corresponding output.

Once a data -classification scheme has been designed, the security standards that stipulate proper approaching practices for each division and the storage criteria that determines the data's lifecycle demands should be discussed.

Once you identify the different types of data in your network, you can separate yoursensitive data from general data. In turn, this allows you to:

- Prioritize your security measures
- Adjust your security controls based on data sensitivity
- Find out who can access, modify, or delete data on your network
- Assess all risks and threats, such as the business impact of a breach orransomware attack, and so on.

TYPES OF CLASSIFICATION

Logistic Regression

It is a very basic yet important classification algorithm in machine learning that uses one or more independent variables to determine an outcome. Logistic regression triesto find a best-fitting relationship between the dependent variable and a set of independent variables. The best-fitting line in this algorithm looks like S-shape as shown in the figure.

Naive Bayes

Naive Bayes is based on **Bayes's theorem** which gives an assumption of independenceamong predictors. This classifier assumes that the presence of a particular feature in aclass is not related to the presence of any other feature/variable. Naive Bayes Classifier are of three types: Multinomial Naive Bayes, Bernoulli Naive Bayes, Gaussian Naive Bayes.

K-Nearest Neighbor Algorithm

You must have heard of a popular saying: "Birds of a feather flock together. " KNN works on the very same principle. It classifies the new data points depending upon the class of the majority of data points amongst the K neighbor, where K is the numberof neighbors to be considered. KNN captures the idea of similarity (sometimes calleddistance ,proximity, or closeness) with some basic mathematical distance formulas like euclidean distance, Manhattan distance, etc.

SUPPORT VECTOR MACHINE

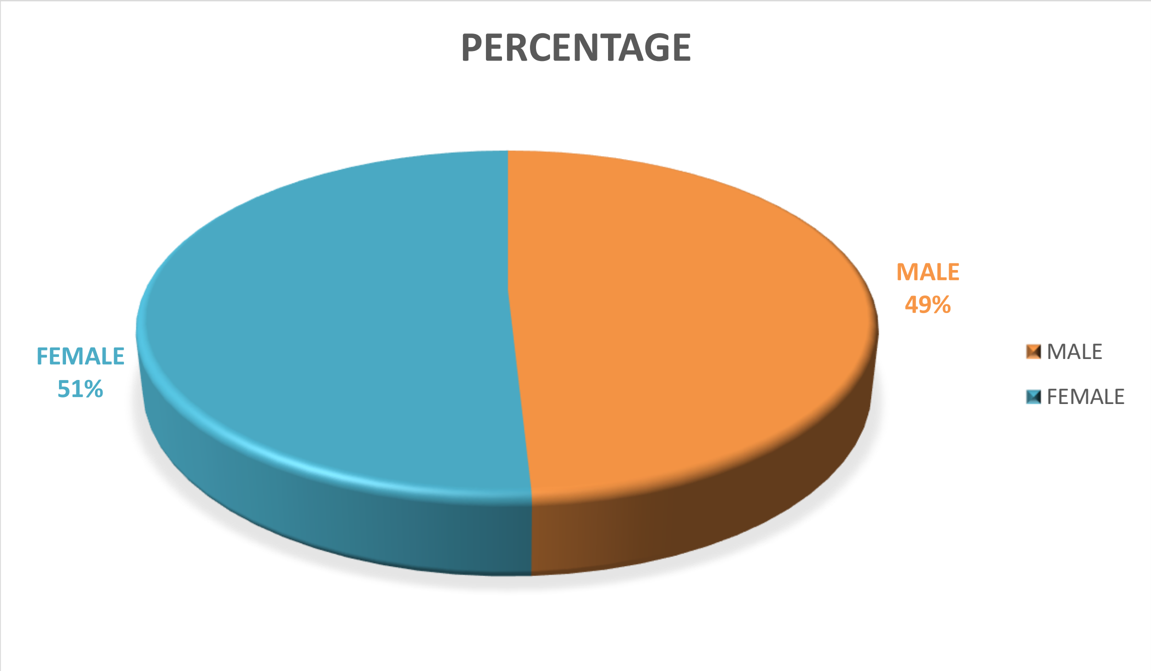
SVM stands for Support Vector Machine. This is a supervised machine learning algorithm that is very often used for both classification and regression challenges. However, it is mostly used in classification problems. The basic concept of the SupportVector Machine and how it works can be best understood by this simple example. So,just imagine you have two tags: green and *blue*, and our data has two features: *x* and *y*.We want a classifier that, given a pair of *(x,y)* coordinates, outputs if it's either *green* or *blue*. Plot labeled training data on a plane and then try to find a plane (hyperplane of dimensions increases) that segregates data points of both colors very clearly.

Decision Tree

The decision tree is one of the most popular machine learning algorithms used. Theyare used for both classification and regression problems. Decision trees mimic human-level thinking so it’s so simple to understand the data and make some good intuitions and interpretations. They actually make you see the logic for the data to interpret. Decision trees are not like black-box algorithms like SVM, Neural Networks, etc.

The above data undergoes classification so there is no need to further classifying the dataset

VISUALIZATION OF FITNESS DATASET

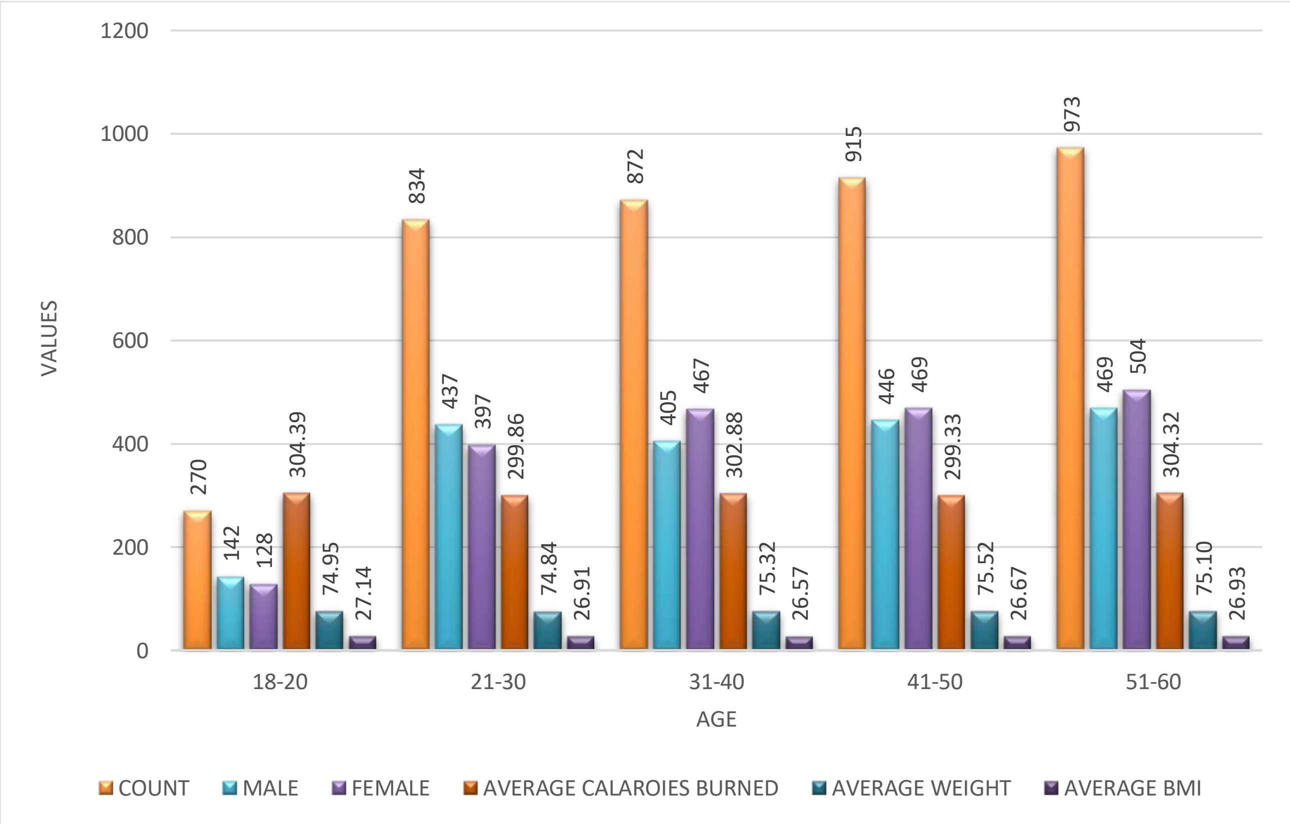


GENDER	TOTAL
MALE	1899
FEMALE	1965

INTERPRETATION:

The above table and chart shows that,now a days Female are more involved in fitness with 1965 members(51%) compared to Male with 1899(49%).

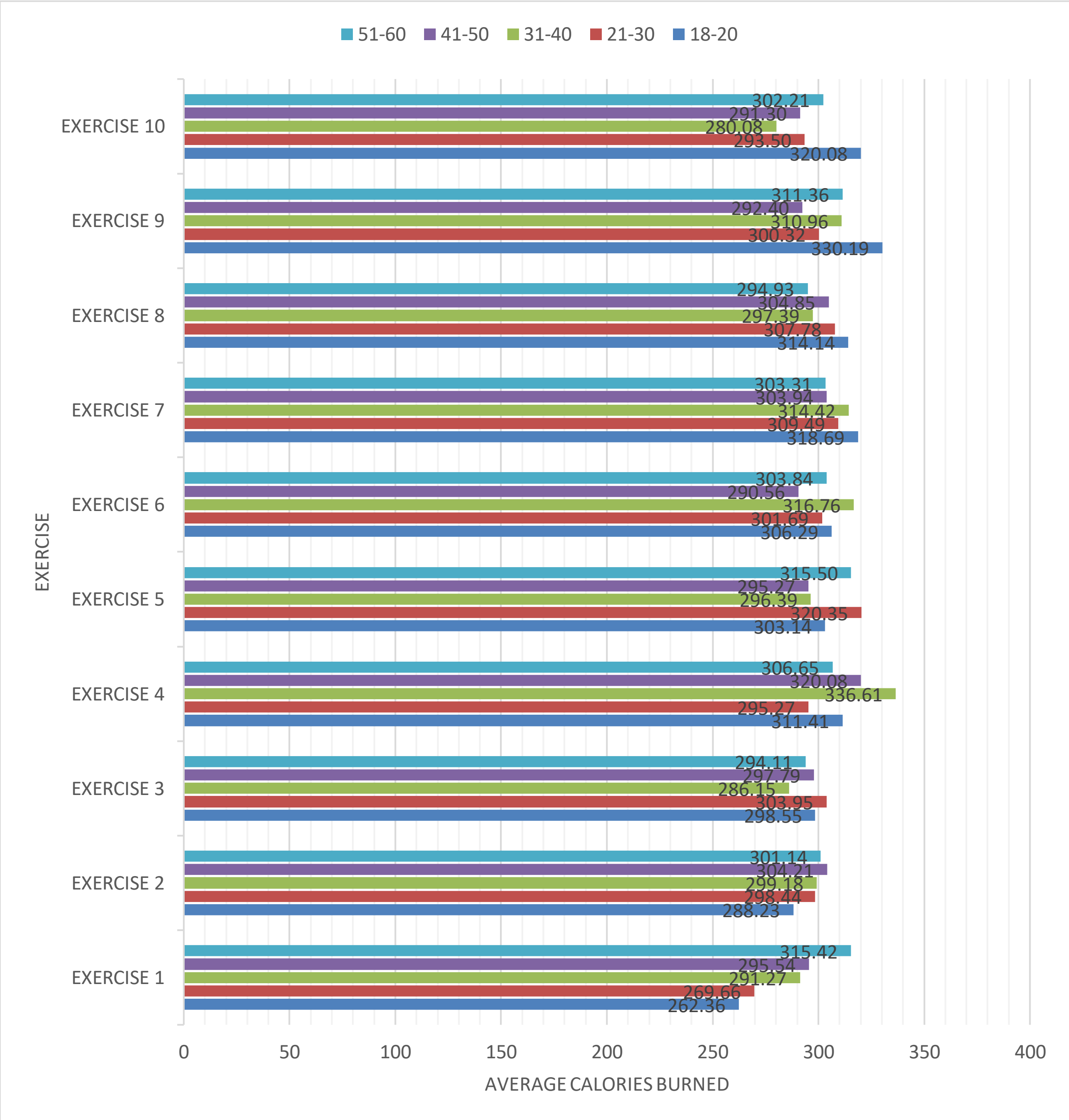
AGE	COUNT	MALE	FEMALE	AVERAGE CALAROIES BURNED	AVERAGE WEIGHT	AVERAGE BMI
18-20	270	142	128	304.39	74.95	27.14
21-30	834	437	397	299.86	74.84	26.91
31-40	872	405	467	302.88	75.32	26.57
41-50	915	446	469	299.33	75.52	26.67
51-60	973	469	504	304.32	75.10	26.93



INTERPRETATION:

From the above table and the clustered barchart we came to that the people from the age group 51-60 are burned more number of calories compared to the other age group and also we can see that the how many male and female in the various kinds of age group.

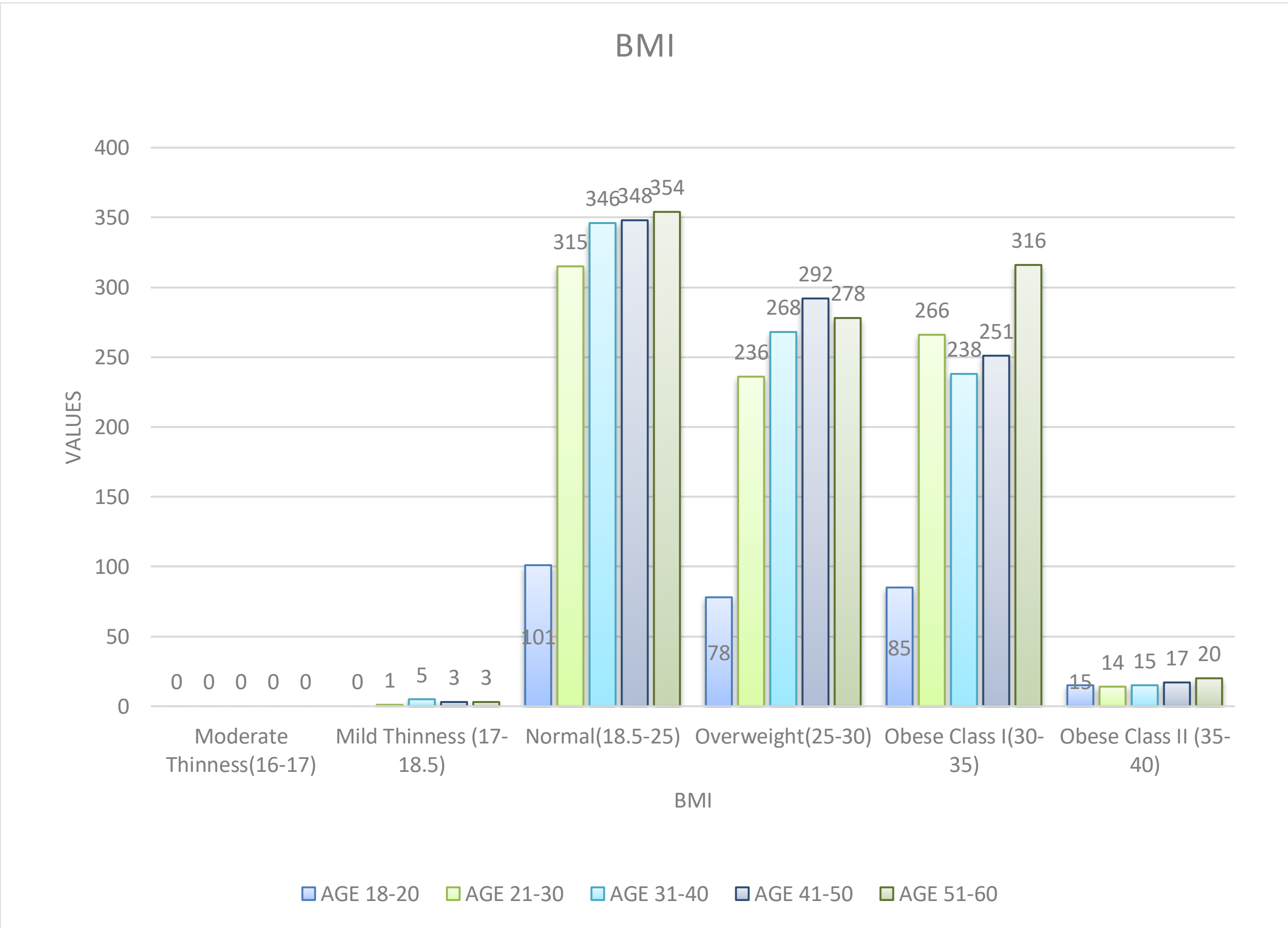
CALORIES BURNED BASED ON EXERCISE										
AGE	**EXERCISE 1**	**EXERCISE 2**	**EXERCISE 3**	**EXERCISE 4**	**EXERCISE 5**	**EXERCISE 6**	**EXERCISE 7**	**EXERCISE 8**	**EXERCISE 9**	**EXERCISE 10**
18-20	262.36	288.23	298.55	311.41	303.14	306.29	318.69	314.14	330.19	320.08
21-30	269.66	298.44	303.95	295.27	320.35	301.69	309.49	307.78	300.32	293.50
31-40	291.27	299.18	286.15	336.61	296.39	316.76	314.42	297.39	310.96	280.08
41-50	295.54	304.21	297.79	320.08	295.27	290.56	303.94	304.85	292.40	291.30
51-60	315.42	301.14	294.11	306.65	315.50	303.84	303.31	294.93	311.36	302.21



INTERPRETATION:

From the above table and chart,we came to know that each age group has a certain efficiency in particular exercise. By doing the particular exercise they can burn more calories.

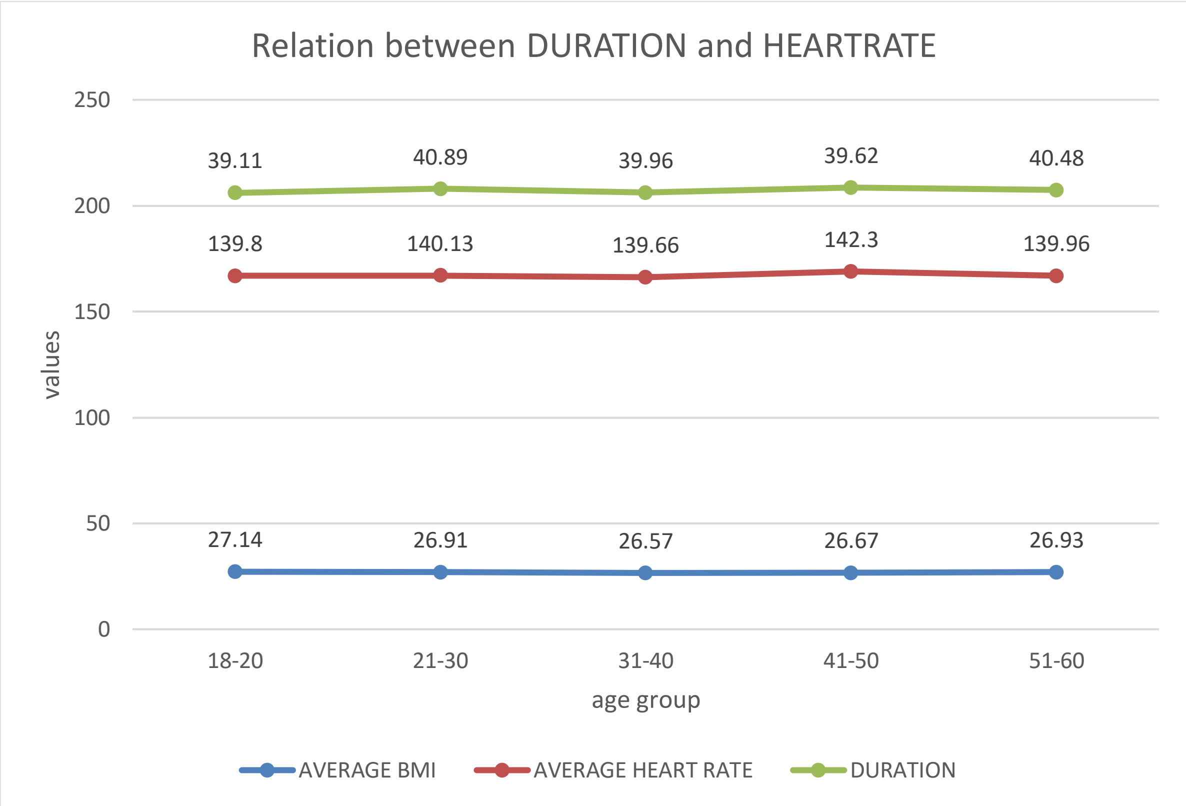
Classification	AGE 18-20	AGE 21-30	AGE 31-40	AGE 41-50	AGE 51-60
Severe Thinness (<16)	0	0	0	0	0
Moderate Thinness (16-17)	0	0	0	0	0
Mild Thinness (17-18.5)	0	1	5	3	3
Normal (18.5-25)	101	315	346	348	354
Overweight (25-30)	78	236	268	292	278
Obese Class I (30-35)	85	266	238	251	316
Obese Class II (35-40)	15	14	15	17	20



INTERPRETATION:

From the above table and chart, we came to know that based on BMI(body mass index)how many number of people are come under normal, underweight, overweight in each age group.

S.NO	AGE	AVERAGE BMI	AVERAGE HEART RATE	DURATION
1	18-20	27.14	139.8	39.11
2	21-30	26.91	140.13	40.89
3	31-40	26.57	139.66	39.96
4	41-50	26.67	142.3	39.62
5	51-60	26.93	139.96	40.48



INTERPRETATION:

From the above table and chart,we came to know that the average range of BMI(body mass index)for each group and the average heart rate of each group based on their workout duration. The average workout duration of each age group is lies between 39-40 mins and the average heartrate of each age group is lies between 139-142.

SCOPE FOR FURTHER ENHANCEMENT

There are several potential scope options for fitness dataset project, depending on the specific goals and resources available. Some potential scopes for such a project might include:

Data collection and cleaning: Gathering and preparing data on fitness dataset, including information on exercise type, calories burned, age, dream weight, and other relevant factors. This could involve scraping data from online sources, collecting data from gym owners other sources, or using a combination of methods.

Feature engineering: Identifying and creating relevant features to use in the prediction model, such as the weight of the person,gender and duration,heart rate and BMI.

Model development: Building and training a machine learning model to predict usedhealthy lifestyle of a person based on the collected and engineered features. This could involve using a variety of algorithms and techniques, such as linear regression, decision trees, or neural networks.

Model evaluation: Testing the performance of the model using a variety of metrics, such as mean absolute error or root mean squared error, to determine how well it predicts the healthy person.

Model deployment: Implementing the model in a production environment, such as by creating a web application or API that allows users to input information about a person's weight,BMI and receive a prediction of height of the person.

Model maintenance: Ongoing monitoring and maintenance of the model to ensure that it continues to perform well and remains accurate over time. This could involve retraining the model on new data as it becomes available, or fine-tuning the model toimprove its performance

CONCLUSION

The conclusion of a exploratory analysis on fitness dataset project will depend on the specific goals and objectives of the project. However, some potential conclusions that could be drawn from such a project might include:

The accuracy of the model: The conclusion could include an evaluation of theperformance of the model, including how well it predicts the healthy person and any limitations or biases that it may have.

The usefulness of the model: The conclusion could discuss how the model can be used in practice, including any potential applications or benefits it could provide.

Recommendations for future work: The conclusion could include suggestions for further research or development that could be pursued to improve the model or expand its capabilities.

Limitations of the project: The conclusion could also discuss any limitations or challenges that were encountered during the project, and how they might be addressed in future work.

Overall, the conclusion of a exploratory analysis on fitness dataset project should provide a clear summary of the key findings and outcomes of the project, as well as any recommendations for future work or application of the model.

REFERANCES

BIBLIOGRAPHY

1.Shonda Kuiper (2008) Introduction to Multiple Regression: Journal of Statistics Education, 16:3, DOI: 10.1080/10691898.2008.11889579

2.Robert T. (1996) Regression Shrinkage and Selection Via the Lasso.In: Journal of the Royal Statistical Society: Series B (Methodological)

3.Geurts P. (2009) Bias vs Variance Decomposition for Regression and Classification. In: Maimon O., Rokach L. (eds) Data Mining and Knowledge Discovery Handbook. Springer, Boston, MA

4.Stephen few(2006) Information dashboard Design:Displaying data for at a glance Monitoring

5.Julie Steele, Noah lliinsky(2010) Beautiful visualization, looking at data through the eyes of experts.

6.Ben fry(2007) Visualizing Data, publishers: O'reilly Media,Inc.

7.Jack dougherty,Ilya Ilyankou(2021)Hands-On Data Visualization

8.Kevin P.Murphy(2012) Machine learning: A Probabilistic Perspective The MIT Press Cambridge,Massachusetts London,England.

9.Christopher M.Bishop(2006) Pattern recognition and machine learning,springer.

10.Andreas C.MULLER and Sarah Gudio(2016) Introduction t**o** machine learning with python,o'Reilly media.

11**.**Ian Goodfellow,Yoshua bengio,and Aaron Courville(2016) DEEP LEARNING,MIT PRESS.

12.aurelien geron,hands on MACHINE LEARNING WITH SCIKIT-LEARN,KERAS,ANDtensorflow(2019),oreilly media**.**

13.Andrew Ng(2018) Machine learning,deeplearning.ai(online book)

14.Shai Shalev-Shwartz and Shai Ben-David(2014) Understanding Machine learning:from theory to algorithms,Cambridge university Press.

15.Peter flach(2012)Machine learning:The Art and Science of Algorithms that Make Sense of Data,Cambridge University Press

WEBLIOGRAPHY

https://www.kaggle.com/

https://www.javatpoint.com/

https://www.tutorialspoint.com/index.htm

YouTube.com

exercise data.csv