PROFESSIONAL TRAINING REPORT

at

Satyabhama Institute of Science and Technology (Deemed to be University)

Submitted in partial fulfillment of the requirements for the award of

Bachelor of Engineering Degree in Computer Science and Engineering

By

HARIPRIYANKA DAVULURU
REG. NO. 39110374

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

SCHOOL OF COMPUTING

SATHYABAMA INSTITUTE OF SCIENCE AND TECHNOLOGY
JEPPIAAR NAGAR, RAJIV GANDHI SALAI,
CHENNAI – 600119, TAMILNADU

NOVEMBER 2021

SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY

(DEEMED TO BE UNIVERSITY)

Accredited with Grade “A” by NAAC

(Established under Section 3 of UGC Act, 1956)

JEPPIAAR NAGAR, RAJIV GANDHI SALAI

CHENNAI– 600119

www.sathyabama.ac.in

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the Bonafide work of **HARIPRIYANKA DAVULURU (Reg. No: 39110374)** who carried out the project entitled **Abalone using power BI and regression** under my supervision from June 2021 to November 2021.

Internal Guide

Dr.M.Kanipriya M.E., Ph.D

Head of the Department

Dr. S. VIGNESHWARI, M.E., Ph.D.
Dr. LAKSHMANAN L, M.E., Ph.D.,

Submitted for Viva voce Examination held on____________________

Internal Examiner **External Examiner**

DECLARATION

HARIPRIYANKA DAVULURU hereby declare that the project report entitled **Abalone dataset using power BI and regression** done by me under the guidance of Dr.M.Kanipriya M.E., Ph.D. is submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering Degree in Computer Science and Engineering.

DATE:

PLACE: Chennai

Haripriyanka Davuluru

SIGNTURE OF THE CANDIDATE

ACKNOWLEDGEMENT

I am pleased to acknowledge my sincere thanks to **Board of Management** of **SATHYABAMA** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr. T. Sasikala M.E., Ph.D.**, **Dean**, School of Computing, **Dr. S. Vigneshwari, M.E., Ph.D. and Dr. L. Lakshmanan, M.E., Ph.D., Heads of the Department** of **Computer Science and Engineering** for providing me necessary support and details at the right time during the progressive reviews.

I would like to express our sincere and deep sense of gratitude to our Project **Dr.M.Kanipriya M.E., Ph.D.,** for her valuable guidance, suggestions and constant encouragement paved way for the successful completion of my project work.

I wish to express my thanks to all Teaching and Non-teaching staff members of the **Department of Computer Science and Engineering** who were helpful in many ways for the completion of the project.

TRAINING CERTIFICATE

CERTIFICATE

OF TRAINING

2021 AWARD

PROUDLY PRESENTED TO

Haripriyanka Davuluru

39110374

for successfully completing the 40 hours professional training program on Data Analytics and Visualization conducted by Cognibot Labs

Given on November 1, 2021

COGNIBOT
AI meets Industry

Ajar Kumar
Trainer - Cognibot Labs

TABLE OF CONTENTS:

ABSTRACT

There is a large amount of data available about abalone and its features. But there has been little analysis and visualization done to understand features of the abalone to find its rings(ages). This research focuses on how to analyze and how the other features of abalone effects on the rings. This paper suggests using leveraged Power BI framework to filter and visualize the data in a simple and digestible format which can assist to formulate conclusions on the effects that these features may have on the rings(ages). After visualizing the data via Power BI, the framework obtains logistic regressions to determine the relationships. The results indicate that the Power BI framework can effectively and innovatively provide analysis on the abalone features. Through the reformulated data for an individual abalone, we can analyze and predict the rings of an abalone.

In statistical modelling, regression analysis is a statistical process for estimating the relationships among variables. More specifically, regression analysis helps the reader understand how the dependent variable changes when any of the independent variables is varied. Thus, regression analysis estimates the average value of the dependent variable when the independent variables are fixed. Therefore, the estimation target is a function of the independent variables called regression function. In limited circumstances, regression analysis can be used to infer causal relationships between the independent and dependent variables.

LIST OF FIGURES

LIST OF ABBREVIATIONS

SI NO	ABBREVIATION	EXPANSION
1	DF	DATA FRAME
2	LR	LOGISTIC REGRESSION
3	BI	Business intelligence
4	SNS	SEABORN
5	PD	PANDAS
6	CORR	CO RELATION
7	SB	SEABORN
8	SKLEARN	SCIKIT LEARN
9	LE	LABEL ENCODER
10	PLT	MATPLOT LIB.PYPLOT
11	M	Male
12	F	Female
13	I	Infant

CHAPTER 1: INTRODUCTION

DATA ANALYTICS:

Data Analytics refers to the techniques used to analyze data to enhance productivity and business gain. Data is extracted from various sources and is cleaned and categorized to analyze various behavioral patterns. The techniques and the tools used vary according to the organization or individual.

if you understand your Business Administration and have the capability to perform Exploratory Data Analysis, to gather the required information, then you are good to go with a career in Data Analytics.

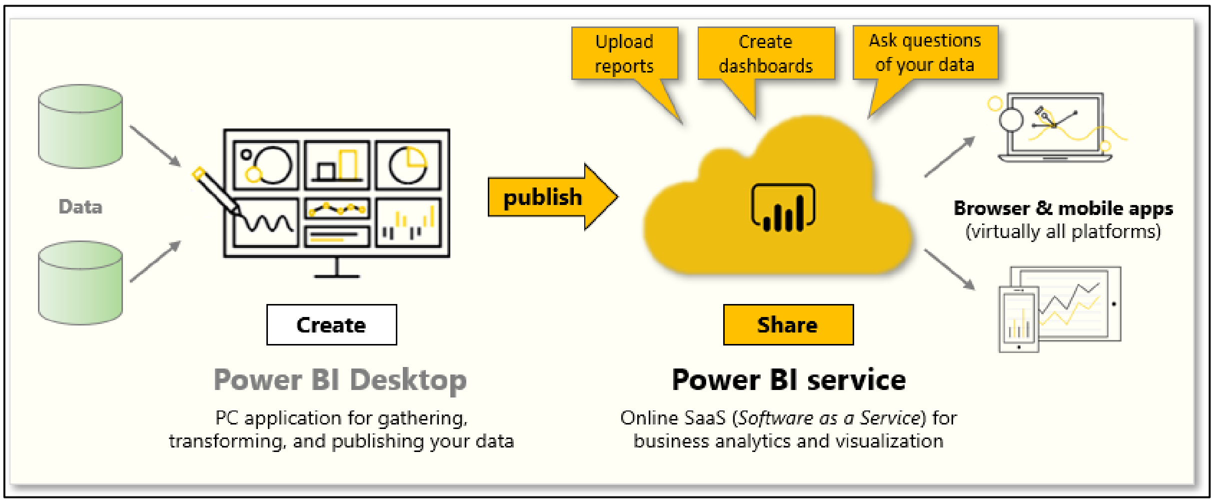
DATA ANALYTICS

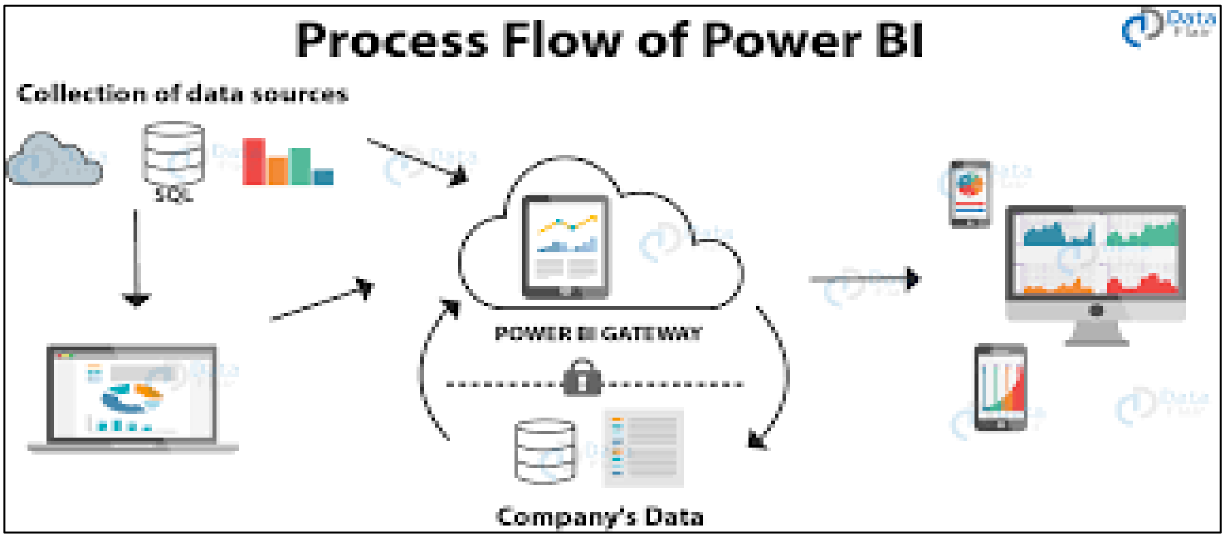
VISUALIZATION IN DATA ANALYTICS

POWER BI:

Microsoft Power BI is a business intelligence platform that provides nontechnical business users with tools for aggregating, analyzing, visualizing, and sharing data. Power BI's user interface is intuitive for users familiar with Excel and its deep integration with other Microsoft products makes it a very versatile self-service tool that requires little upfront training.

A free version of Power BI is intended for small to midsize business owners; a professional version called Power BI Plus is available for a monthly subscription fee. Users can download an application for Windows 10, called Power BI Desktop, and native mobile apps for Windows, Android, and iOS devices.

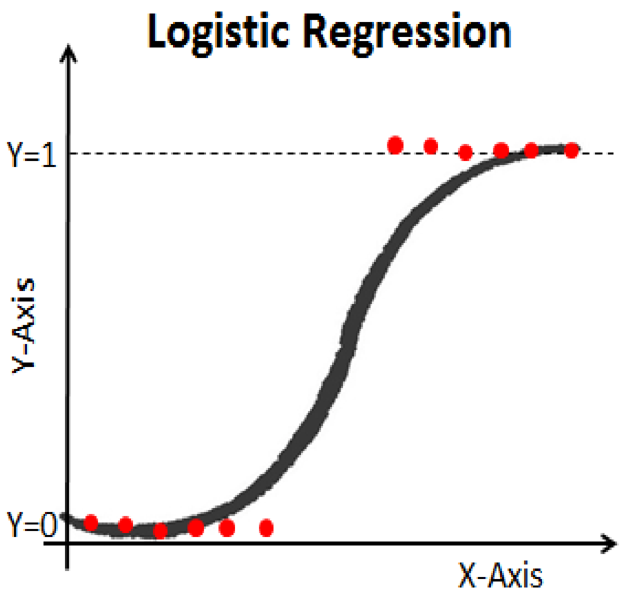




REGRESSION:

Logistic Regression is one of the simplest and commonly used Machine Learning algorithms for two-class classification. It is easy to implement and can be used as the baseline for any binary classification problem. Its fundamental concepts are also constructive in deep learning. Logistic regression describes and estimates the relationship between one dependent binary variable and independent variables.

Predictive models built using this approach can make a positive difference in your business or organization. Because these models help you understand relationships and predict outcomes, you can act to improve decision-making. For example, a manufacturer’s analytics team can use logistic regression analysis as part of a statistics software package to discover a probability between part failures in machines and the length of time those parts are held in inventory. With the information it receives from this analysis, the team can decide to adjust delivery schedules or installation times to eliminate future failures



ABALONE:

Abalone shell is very durable and used by many ancient cultures for its healing properties.

They used shells to burn sage believing that it will carry their message to heaven.
It represents a human connection to the ocean, circle of life and protection on a voyage.

The shell of abalones is convex, rounded to oval in and may be highly arched or very flattened. The shell of the majority of species has a small, flat spire and two to three whorls.

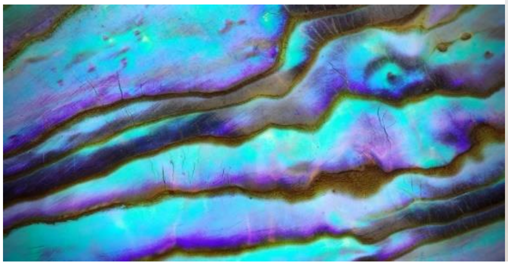
PURPOSE OF ABALONE:

Abalone has been an important staple in native cultures around the world, specifically in Africa and on the North American West coast.

The meat was used as food, and the shell was used as currency for many tribes.
Native use, Abalone has been an important staple in native cultures around the world, specifically in Africa and on the North American West coast. The meat was used as food, and the shell was used as currency for many tribes.

BENEFITS:

Abalone Shell is said to enhance feelings of peace, compassion and love. It has a lovely warm, gentle vibration.
It is great in times of tough emotional issues, soothing the nerves and encouraging a calm demeanor.
It is said to gently help open our psychic and intuitive connections.



CHAPTER 2: Aim and Scope of the present investigation

2.1 AIM OF THE POWER BI:

Power BI is a Business Intelligence (BI) tool that collates and analyses data from a wide range of sources such as Excel workbooks, SQL databases, web sites and cloud services and displays it in user friendly, interactive BI dashboards.

2.2 AIM OF THE REGRESSION:

The purpose of logistic regression is to estimate the probabilities of events, including determining a relationship between features and the probabilities of outcomes.

2.3 SCOPE OF THE DATA ANALYTICS USING POWER BI AND REGRESSION:

Living in the 21st century, you might have often come across the word 'data analytics. Currently, it is one of the most buzzing terminologies.

1. Analytics will play an important role in data security. Analytics are already transforming differential privacy, intrusion detection, digital watermarking and malware countermeasures.
2. The Internet of Things (IOT) will continue to grow rapidly. Analytics tools and methods for dealing with large amounts of structured and unstructured data generated by IOT will continue to gain importance.
3. Companies will voice their need of routinely monetizing their own data for financial gain.
4. Growth of Cognitive Analytics.
5. Relevance of 'Open-Source Solutions' will regain momentum.
6. Boost in demand for Data Scientists- a hunt for people who can balance quantitative analysis skills with an ability to tell the story of their data in compelling, visual ways.
7. Companies would become over-critical and cautious about Data Accuracy.

2.4 PRESENT CONDITION IN DATA ANALYTICS:

The pandemic has created monumental shifts in daily life, making all of us reevaluate almost every aspect of our work and home lives.

Full-time remote work pushed everyone into the cloud. Today, being able to empower employees successfully means leveraging SaaS and the cloud more readily and effectively. Leading organizations were already making this transition before Covid-19, with others now expediting the move out of necessity. As companies experience the increased cost savings and flexibility, along with becoming more comfortable with the inherent security and scalability of cloud and SaaS, there will be no reason to go back.

CHAPTER 3: HOW TO USE THE DATA ANALYTICS

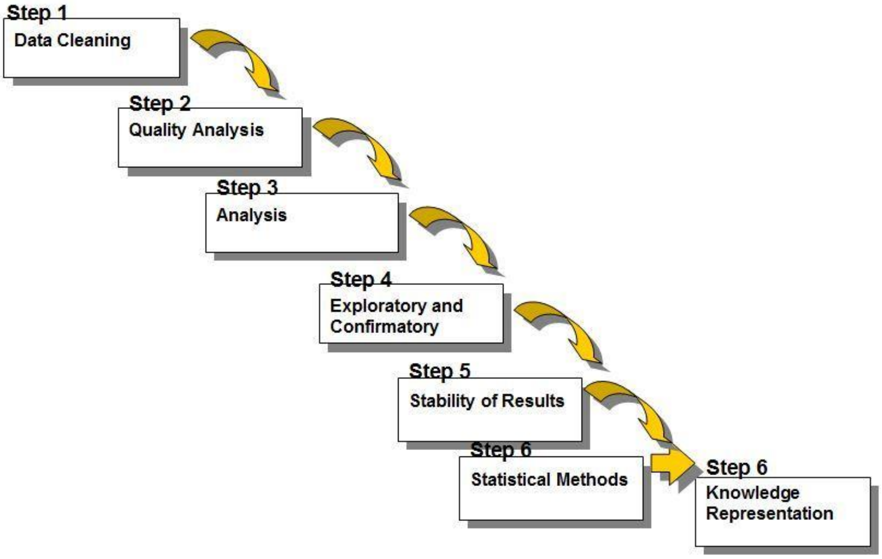


Fig (3.1) STEPS FOR THE DATA ANALYTICS:

Data analysis can help companies better understand their customers, evaluate their ad campaigns, personalize content, create content strategies and develop products.

Analysis techniques give businesses access to insights that can help them to improve their performance. It can help you improve your knowledge of your customers, ad campaigns, budget and more.

3.1.1 THERE ARE 4 WAYS TO USE DATA ANALYTICS:

1.Improved Decision Making:

Data analytics eliminates much of the guesswork from planning marketing campaigns, choosing what content to create, developing products and more. It gives you a 360-degree view of your customers, which means you understand them more fully, enabling you to better meet their needs.

2. More Effective Marketing

Using the Lota, me Campaign Analytics tool, you can gain insights into which audience segments are most likely to interact with a campaign and convert. You can use this information to adjust your targeting criteria either manually or through automation or use it to develop different messaging and creative for different segments. Improving your targeting results in more conversions and less ad waste.

3.Better Customer Service

Your data can reveal information about your customers' communications preferences, their interests, their concerns and more. Having a central location for this data also ensures that your whole customer service team, as well as your sales and marketing teams, are on the same page.

4. More Efficient Operations

Data analytics can help you streamline your processes, save money and boost your bottom line. When you have an improved understanding of what your audience wants, you waste less time on creating ads and content that don't match your audience's interests.

3.1.2 HOW Companies use data analytics?

Companies use Big Data Analytics to Increase Customer Retention. ... And the more data that a company has about its customer base, the more accurately they can observe customer trends and patterns which will ensure that the company can deliver exactly what its customers want.

3.1.3 Why do we need data analysis?

Data analysis is important in business to understand problems facing an organisation, and to explore data in meaningful ways. Data in itself is merely facts and figures. Data analysis organises, interprets, structures and presents the data into useful information that provides context for the data.

3.1.4 What is the purpose of a data analysis?

Data Analysis is a process of inspecting, cleansing, transforming, and modelling data with the goal of discovering useful information, suggesting conclusions, and supporting decision-making.

3.2 STEPS FOR THE POWER BI:

3.2.1 INSTALLATION OF POWER BI TOOL:

Step 1: Download Power BI Desktop.

Step 2: Install Power BI Desktop.

Step 3: Import data to Power BI Dashboard.

Step 4: Format data in Power BI Dashboard.

Step 5: Create Data Visualization in Power BI Dashboard.

3.2.2 MAJOR STEPS FOR THE POWER BI:

STEP1: Import dataset from excel workbook.
STEP 2: understand and represent the given data set.
STEP 3: Visualize the data using power bi tool.
STEP 4: Compare and predict the given dataset.

3.2.3 Types of Charts in Power BI:

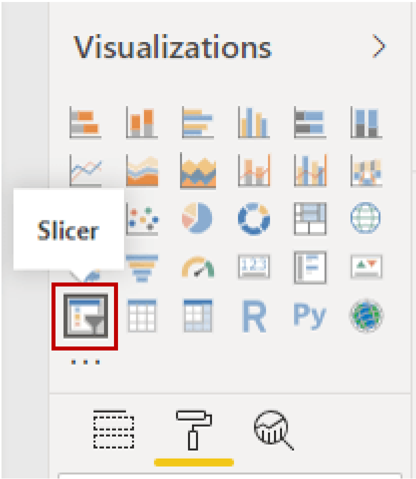


Fig (3.2) Power BI Slicers

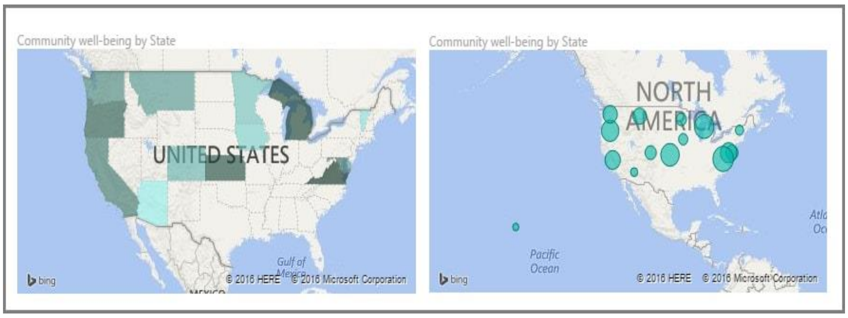
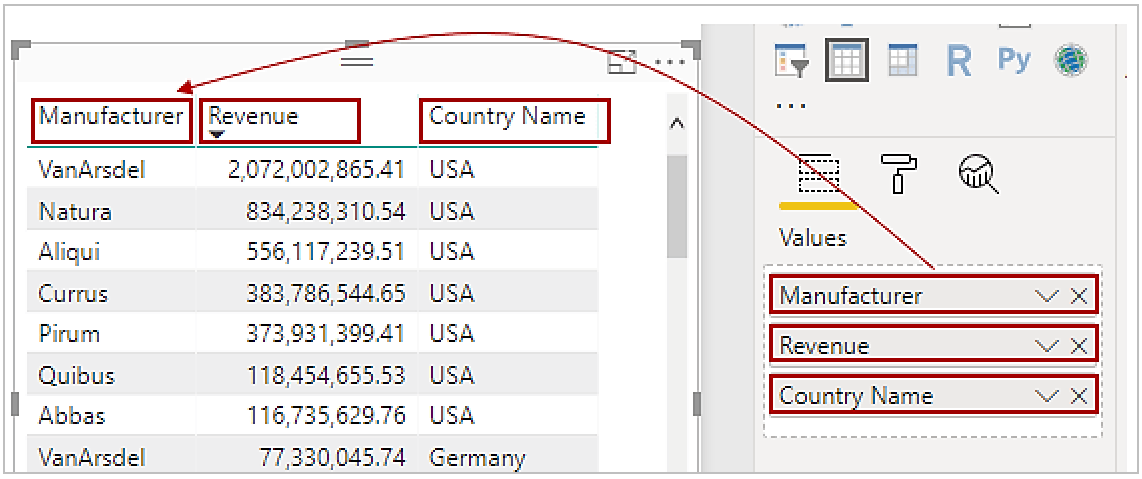


Fig (3.2) Power BI Map Visualizations



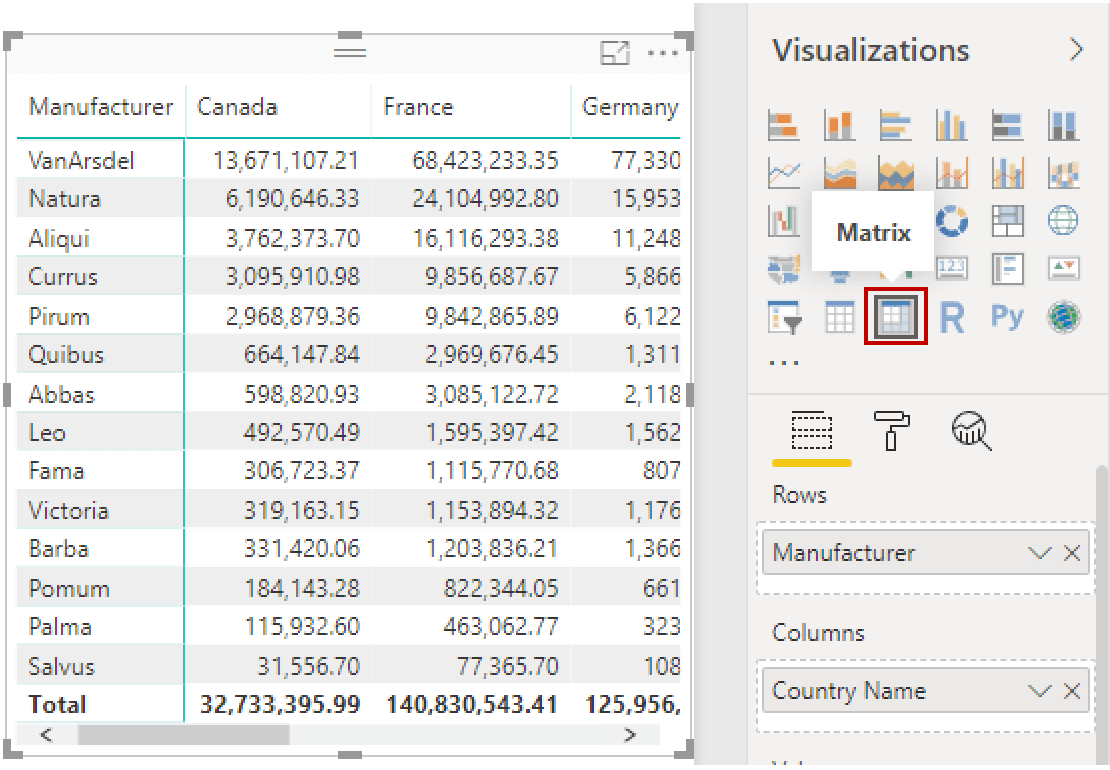


Fig (3.3) Power BI Tables & Matrices

4) Power BI Scatter Chart, Waterfall Chart, and Funnel Charts

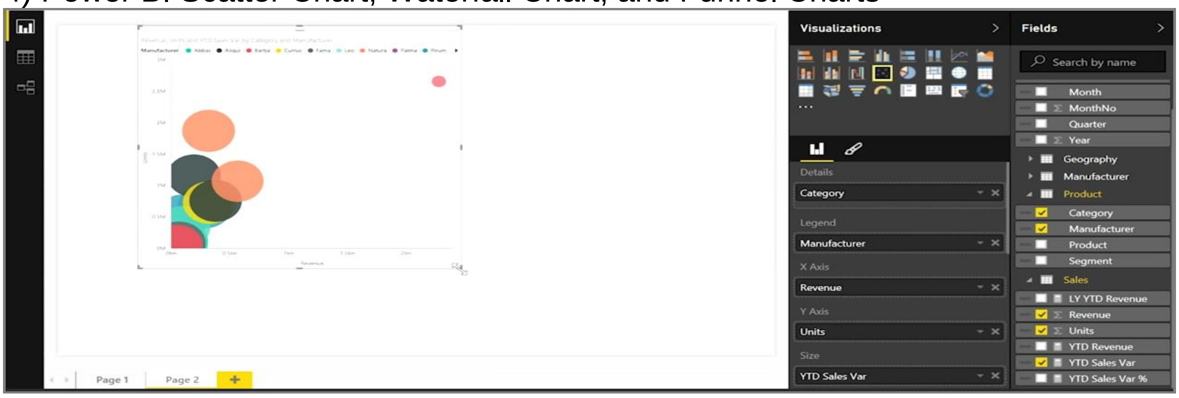


Fig (3.4) Power BI Scatter Chart

Fig (3.4.1) Waterfall chart

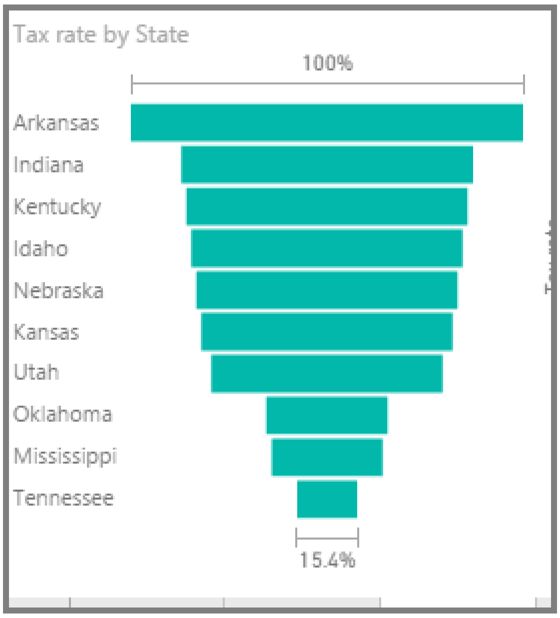
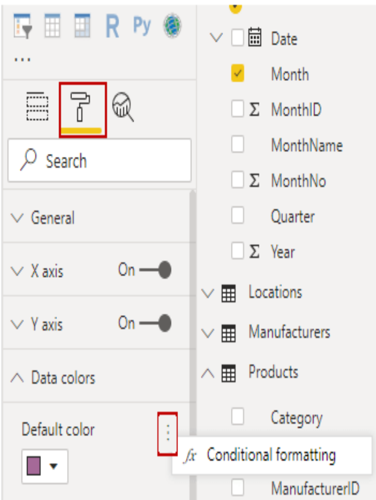


Fig (3.4.2) Funnel chart



Fig(3.5) Modifying Colors in Power BI

3.2.4 Types of Power BI tools:

Power BI Desktop

Power BI desktop is the primary authoring and publishing tool for Power BI. Developers and power users use it to create brand new models and reports from scratch.

Costs: Free

Power BI service

Online Software as a Service (SaaS) where Powe BI data models, reports, dashboards are hosted. Administration, sharing, collaboration happens in the cloud.

Pro license: $10/users/month

Power BI Data Gateway

Power BI Data Gateway works as the bridge between the Power BI Service and on-premises data sources like Direct Query, Import, Live Query. It is Installed by BI Admin.

Power BI Report Server

It can host paginated reports, KPIs, mobile reports, & Power BI Desktop reports. It is updated every 4 months and installed/managed by the IT team. The users can modify Power BI reports other reports created by the development team.

Power BI Mobile Apps

Power BI mobile app is available for iOS, Android, Windows. It can be managed using Microsoft Intune. You can use this tool to view reports and dashboards on the Power BI Service Report Server.

3.3 STEPS FOR THE GRAPHS (ABALONE).

Attribute Information:

Given is the attribute name, attribute type, the measurement unit, and a brief description. The number of rings is the value to predict either as a continuous value or as a classification problem.

Name / Data Type / Measurement Unit / Description

->Sex / nominal / -- / M, F, and I (infant)
->Length / continuous / mm / Longest shell measurement
->Diameter / continuous / mm / perpendicular to length
->Height / continuous / mm / with meat in shell
->Whole weight / continuous / grams / whole abalone
->Shucked weight / continuous / grams / weight of meat
->Viscera weight / continuous / grams / gut weight (after bleeding)
->Shell weight / continuous / grams / after being dried
->Rings / integer / -- / +1.5 gives the age in years

We have used the following graph:

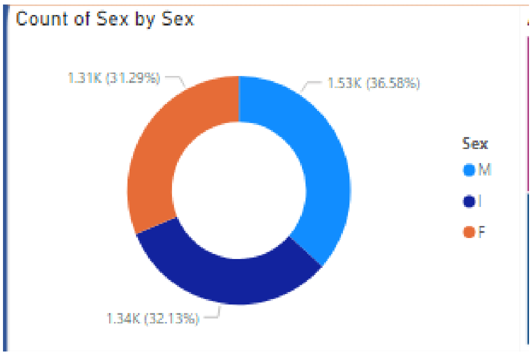


Fig (3.3.1) COUNT BY SEX BY SEX

*In this graph we are representing the total sum of sex by sex.

*Three sexes are

M-male

F-female

I-infant

*The sum of the values for the graph:

M-1528 (36%)

F-1307 (32%)

I-1342 (32%)

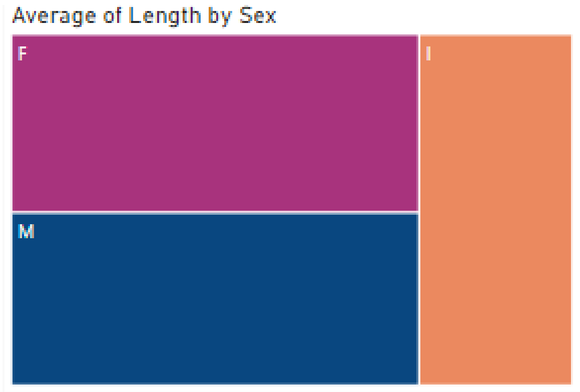


Fig (3.3.2) AVERAGE LENGTH BY SEX

*In this graph we are showing the average length by sex.

*Average for the male, female and infant.

*The average of length by sex values:

M-0.56

F-0.58

I-0.43

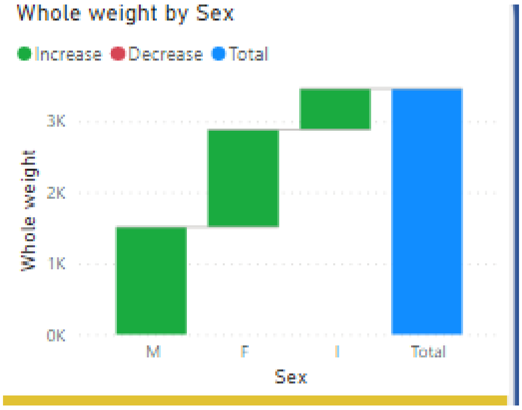


Fig (3.3.3) WHOLE WEIGHT BY SEX:

* In this graph we are representing the total sum of whole weight by sex.

* The values for the genders are

M- 1514

F-1367

I-1578

*The total of the whole weight by sex is 3461.

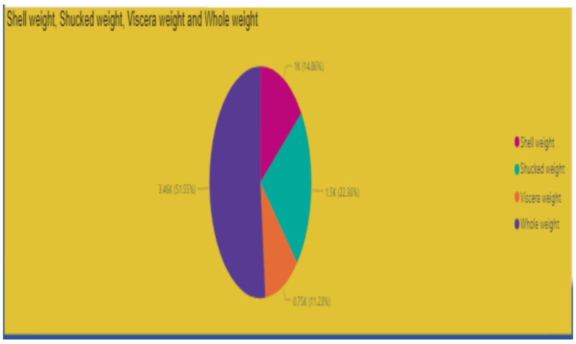


Fig (3.3.4) SHELL WEIGHT, SHUCKED WEIGHT, VISCERA WEIGHT AND WHOLE WEIGHT.

*In this graph we are representing four types of weight are:

1. SHELL WEIGHT- shell after being the dried of the abalone(seashell)
2. SHUCKED WEIGHT – weight of the meat present in the abalone shell.
3. VISCERA WEIGHT – After removing the unnecessary organs. (guts)
4. WHOLE WEIGHT – the total weight all the three which are the shell, shucked, viscera weight.

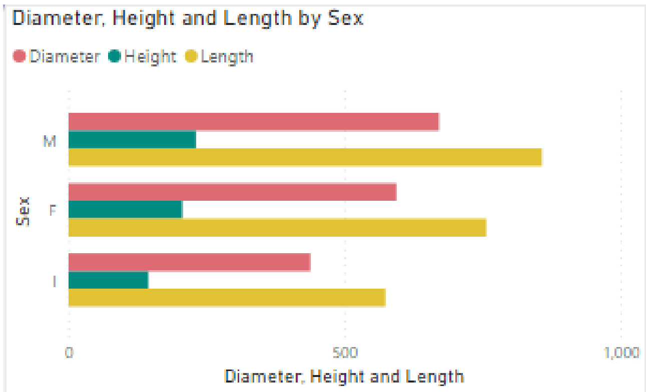


Fig (3.3.5) DIAMETER, HEIGHT, AND LENGTH BY SEX.

*In this graph we are presenting the three criteria are diameter, height and length by sex.

*In this graph the values are:

M- LENGTH -857
HEIGHT- 231
DIAMETER – 671

F- LENGTH -756
HEIGHT- 206
DIAMETER – 594

I- LENGTH -574
HEIGHT- 144
DIAMETER – 438

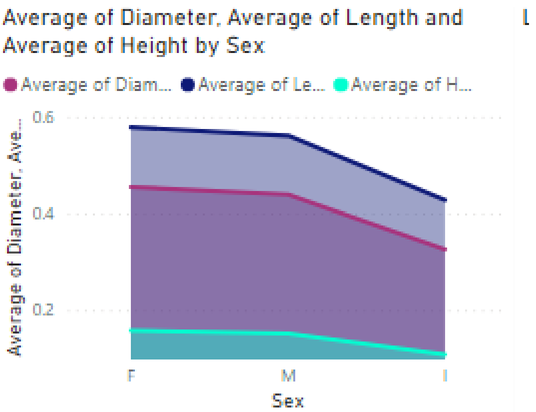


Fig (3.3.6) AVERAGE OF DIAMETER, AVERAGE OF LENGTH AND AVERAGE OF HEIGHT BY SEX.

*In this graph we are showing AVERAGE OF DIAMETER, LENGTH AND HEIGHT BY SEX.

*In Females(F):

AVERAGE LENGTH IS -0.58

AVERAGE HEIGHT IS- 0.16

AVERAGE DIAMETER IS – 0.45

*In Male(M):

AVERAGE LENGTH IS – 0.56

AVERAGE HEIGHT IS – 0.15

AVERAGE DIAMETER IS – 0.44

*In Infant(I):

AVERAGE LENGTH IS -0.43

AVERAGE HEIGHT IS- 0.11

AVERAGE DIAMETER IS – 0.33

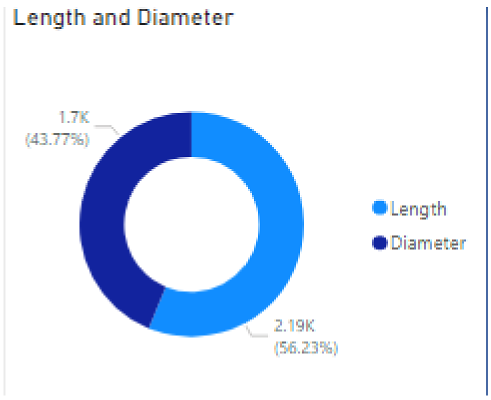


Fig (3.3.7) LENGTH AND DIAMETER.

*In this graph we are representing the length and diameter which is given in the abalone dataset.

*The values are:

LENGTH – 1.703 (43.77%)

DIAMETER – 2.188(56.23%)

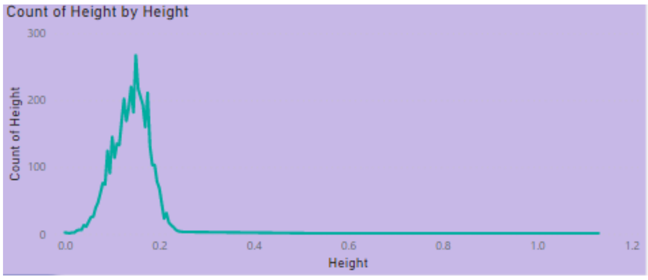


Fig (3.3.8) COUNT OF HEIGHT BY HEIGHT

*In this graph we are showing the total number of counts of weight.

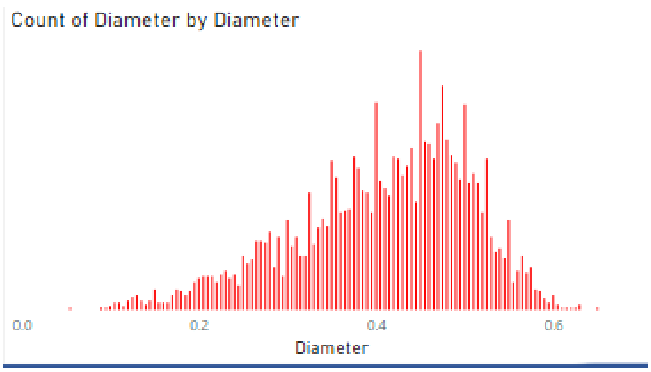


Fig (3.3.9) COUNT OF DIAMETER BY DIAMETER.

*In this graph we are representing the total number of counts of diameter.

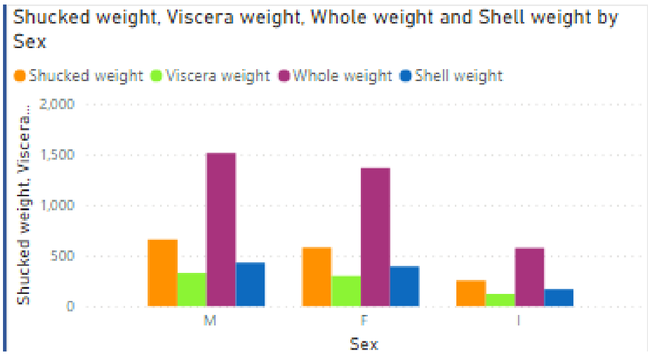


Fig (3.3.10) COUNT OF DIAMETER BY DIAMETER.

*In this graph we are showing the all the weights present in the data set by sex.

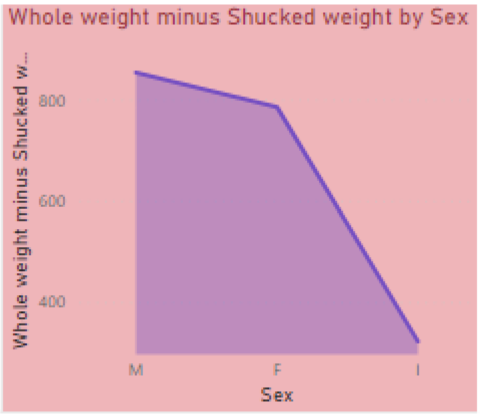


Fig (3.3.11) WHOLE WEIGHT MINUS SHUCKED WEIGHT BY SEX.

*In this graph we are taking out the difference between the whole weight and shucked weight.

*In males the difference of the whole weight by the shucked is 853.4

*In females the difference of the whole weight by the shucked is 784.5

* In infant the difference of the whole weight by the shucked is 332.5

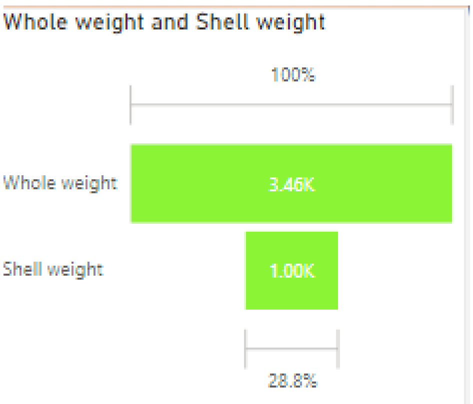


Fig (3.3.12) WHOLE WEIGHT AND SHELL WEIGHT.

*In this graph we are showing the data of shell weight present in whole weight.

*The total percentage of shell weight is 28%

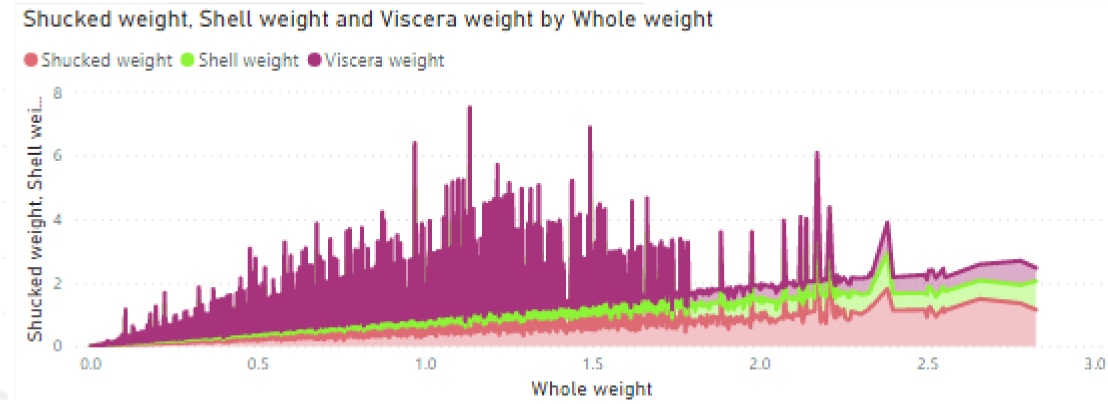


Fig (3.3.13) SHUCKED WEIGHT, SHELL WEIGHT AND VISCERA WEIGHT BY WHOLE WEIGHT.

*In this graph we are representing all three weights by whole weights.

*By this graph we can say that how much contents of weights are present in each weight clearly.

Fig (3.3.14) SHUCKED WEIGHT BY VISCERA WEIGHT:

*From the above figure selecting the particular value i.e. A –2.38 we get the graph of the shucked weight by viscera weight for the given value.

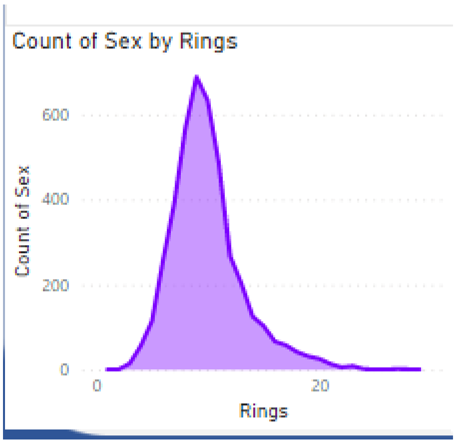


Fig (3.3.15) COUNT OF SEX BY RINGS

*By this graph we can represent the different ages of the abalone shell.

*We represent the count of sex by rings(ages).

*That means Total number of sex present in particular age. (rings).

Ring-9 (max)

Count of sex-689

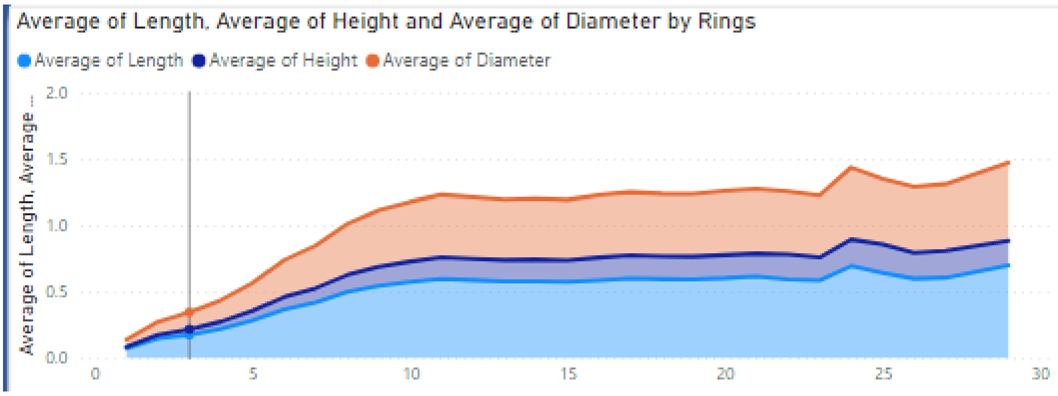


Fig (3.3.16) AVERAGE OF LENGTH, AVERAGE OF HEIGHT AND AVERAGE OF DIAMETER OF RINGS.

*In this graph we show the average of length, height, diameter by rings.

*And can show the values for the particular ring.

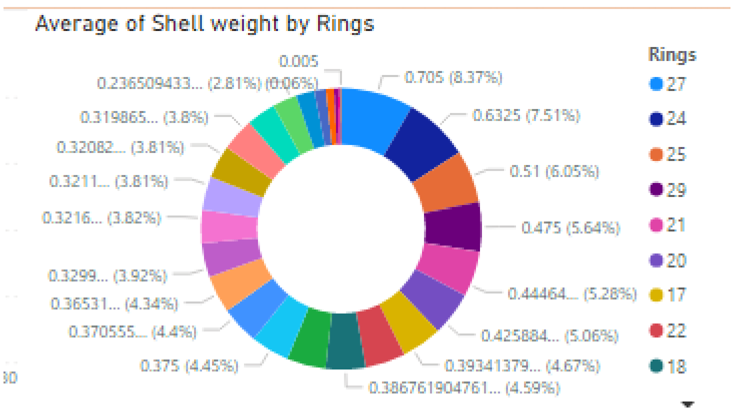


Fig (3.3.17) AVERAGE OF SHELL WEIGHT BY RINGS.

*In this graph we are representing the average of shell weight by rings.

*So, we can show that the ages of different shells.

*For example:

Rings 27

Average of shell weight 0.71(8.37%)

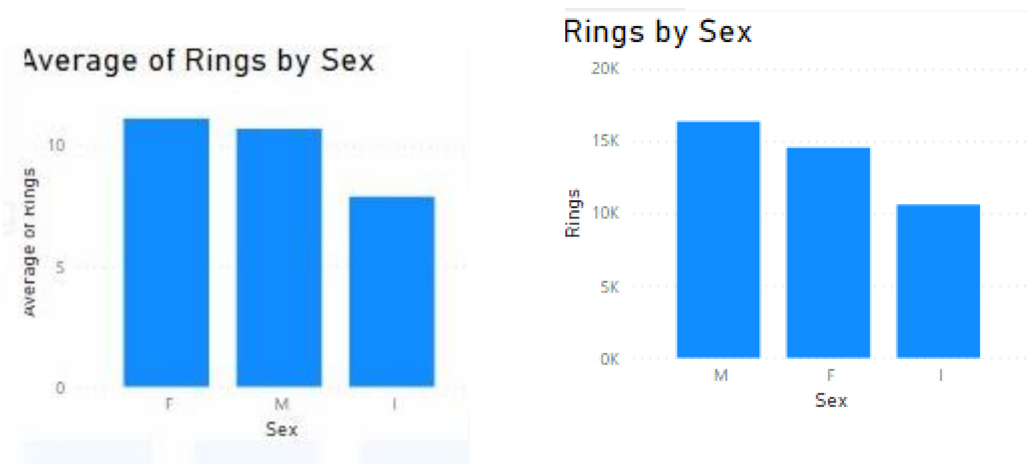


Fig (3.3.18) RINGS BY SEX.

*In this graph we are representing the rings by sex and finding their sum.

*And we can find that:

MALES-16358

FEMALE-14546

INFANT-10589

*Average of rings by sex.

FEMALE – 11.13

MALE – 10.71

INFANT – 7.89

Fig (3.3.19) Average of rings.

*USING CARD, we can represent the average of rings for selected one.

3.4 Steps for the regression:

3.4.1 Logistic regression in data analytics?

Logistic regression is **a statistical analysis method used to predict a data value based on prior observations of a data set**. ... A logistic regression model predicts a dependent data variable by analysing the relationship between one or more existing independent variables.

3.4.2 Steps for the logistic regression:

- Data Pre-processing step.
- Fitting Logistic Regression to the Training set.
- Predicting the test result.
- Test accuracy of the result (Creation of Confusion matrix)
- Visualizing the test set result.

3.4.3 How to Build a logistic regression model in model:

1.Import the required libraries
2. Read and understand the data
3. Exploratory Data Analysis
4. Data Preparation
5. Building Logistic Regression Model
6. Making Predictions on Test Set
7. Assigning Scores as per predicted probability values

CHAPTER 4: RESULTS AND DISCUSSION, PERFORMANCE ANALYSIS

Result:

From this project we were able to find out the Rings(age) of the abalone and we found out from the given dataset that most of the abalone rings were on an average of **9.93** as we can see in the previous figure (**19**).

From the dataset we also found out that most of the abalone/seashells are Male around **1528** then next females around **1307** and last the infants are **1342**.
From this we can say that the dataset contained more of Males when compared with other two sexes. The female and the infants are almost same.

Most of the abalone present in the dataset had height of **0.15** which were around **267**, and most of the seashells had diameter ranged between **0.4** to **0.6**.

On top we found out that the average length for Male were around **0.56**, for females it was around **0.58** and for infant's it is around **0.43**.

We also found out that the weights of the abalone of different sex were almost same except for the infants. The Sum weight for Males were around **1514**, females it's **1,367** and for the infants its **578**.

We also found that the male abalone (seashells) had more shucked weight(meat) when compared to females and infants. The male had average of **853 (0.991)**, female **784 (1.04)** and for infants it was **322 (0.43)**.

As we have shown in fig.14, we have selected a value i.e., 2.38 which is used to get the graph for the shucked weight by viscera weight to find the perfect weight of the abalone meat.

In the dataset the greatest number of the abalones (seashell) were younger …… at the Rings(age) of 9 the total abalones were 689 then followed by 8 at 568,10 at 634 and 11 at 487. And the older abalone from age 24 to 29 were very less when compared.

The average rings(age) for males were **11.3** females **10.71** and infants **7.89**.

REGRESSION:

For the project we used the Python's pandas' concept of regression to find/prediction out the values for rings(age). We used logistics regression to find the values and we train the model and then test the trained predicted model with the original model, and we find the accuracy for the model we have to check if its correct.

CHAPTER 5: Conclusion

Conclusion/Summary

This project is about to Analyzing and Visualizing the given data set. The given data set is about the Abalone. The dataset is used to find out/ predict the values for Rings(age).
In this project we use the software called Power Bi to do the visualization of the data set and we use python- panda's regression to predict/find the age for the abalone.

Importing the dataset to the power bi the data needs to be visualized and then the data present in the file which are the Length, height, diameter, Rings(age) and all the different weights of the abalones and the data can be represented in different forms of graphs which were Bar, pie…… and also using the python with the concept of pandas and regression to predict the values for the rings(age). And then checked withed the original value to make sure that the predicted values were perfect for the data set.

Form this project we were able to find out the Length, height, diameter, Rings (age) and all the different weights of the abalones with help of power bi visualization and pandas' regression to find/ predict the rings(age) of the abalones.

CHAPTER 6: APPENDIX

Screenshots:

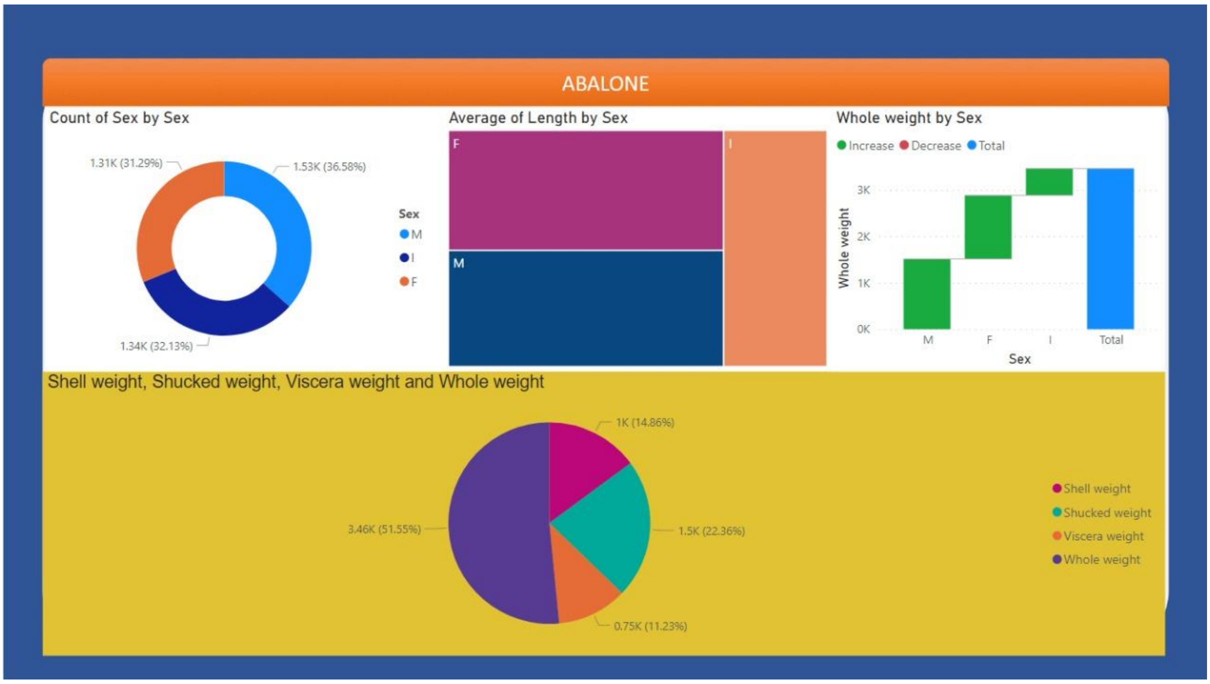


Fig (5.1) ABALONE

Fig (5.2) DIAMETER, HEIGHT AND LENGTH

Fig (5.3) WEIGHTS

Fig (5.4) RINGS(Ages)

LOGISTIC REGRESSION

In [1]:
```
%matplotlib inline
```

In [2]:
```
import pandas as pd
```

In [3]:
```
df=pd.read_csv("C:/Users/ACER/Downloads/abalone.csv")
```

In [4]:
```
df
```

Out[4]:

	Sex	Length	Diameter	Height	Whole weight	Shucked weight	Viscera weight	Shell weight	Rings
0	M	0.455	0.365	0.095	0.5140	0.2245	0.1010	0.1500	15
1	M	0.350	0.265	0.090	0.2255	0.0995	0.0485	0.0700	7
2	F	0.530	0.420	0.135	0.6770	0.2565	0.1415	0.2100	9
3	M	0.440	0.365	0.125	0.5160	0.2155	0.1140	0.1550	10
4	I	0.330	0.255	0.080	0.2050	0.0895	0.0395	0.0550	7
...	...	...	...	...	...	...	...	...	...
4172	F	0.565	0.450	0.165	0.8870	0.3700	0.2390	0.2490	11
4173	M	0.590	0.440	0.135	0.9660	0.4390	0.2145	0.2605	10
4174	M	0.600	0.475	0.205	1.1760	0.5255	0.2875	0.3080	9
4175	F	0.625	0.485	0.150	1.0945	0.5310	0.2610	0.2960	10
4176	M	0.710	0.555	0.195	1.9485	0.9455	0.3765	0.4950	12

4177 rows × 9 columns

In [5]:
```
df.shape
```

Out[5]: (4177, 9)

In [6]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4177 entries, 0 to 4176
Data columns (total 9 columns):
 #   Column          Non-Null Count  Dtype  
---  ------          --------------  -----  
 0   Sex             4177 non-null   object 
 1   Length          4177 non-null   float64
 2   Diameter        4177 non-null   float64
 3   Height          4177 non-null   float64
 4   Whole weight    4177 non-null   float64
 5   Shucked weight  4177 non-null   float64
 6   Viscera weight  4177 non-null   float64
 7   Shell weight    4177 non-null   float64
 8   Rings           4177 non-null   int64  
dtypes: float64(7), int64(1), object(1)
memory usage: 293.8+ KB
```

In [7]:
```
df.describe()
```

Out[7]:

	Length	Diameter	Height	Whole weight	Shucked weight	Viscera weight	Shell weight
count	4177.000000	4177.000000	4177.000000	4177.000000	4177.000000	4177.000000	4177.000000
mean	0.523992	0.407881	0.139516	0.828742	0.359367	0.180594	0.238831
std	0.120093	0.099240	0.041827	0.490389	0.221963	0.109614	0.139203
min	0.075000	0.055000	0.000000	0.002000	0.001000	0.000500	0.001500
25%	0.450000	0.350000	0.115000	0.441500	0.186000	0.093500	0.130000
50%	0.545000	0.425000	0.140000	0.799500	0.336000	0.171000	0.234000
75%	0.615000	0.480000	0.165000	1.153000	0.502000	0.253000	0.329000
max	0.815000	0.650000	1.130000	2.825500	1.488000	0.760000	1.005000

In [8]:
```
df.describe().columns
```

Out[8]:
```
Index(['Length', 'Diameter', 'Height', 'Whole weight', 'Shucked weight',
       'Viscera weight', 'Shell weight', 'Rings'],
      dtype='object')
```

In [9]:
```
df.columns
```

Out[9]:
```
Index(['Sex', 'Length', 'Diameter', 'Height', 'Whole weight', 'Shucked weight',
       'Viscera weight', 'Shell weight', 'Rings'],
      dtype='object')
```

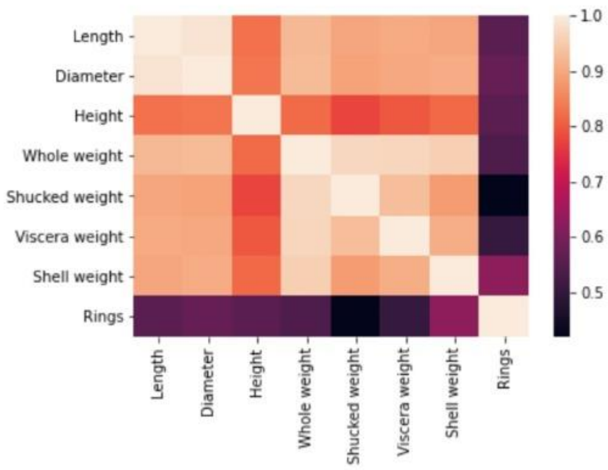
In [10]:
```
df.corr()
```

Out[10]:

	Length	Diameter	Height	Whole weight	Shucked weight	Viscera weight	Shell weight	Rings
Length	1.000000	0.986812	0.827554	0.925261	0.897914	0.903018	0.897706	0.556720
Diameter	0.986812	1.000000	0.833684	0.925452	0.893162	0.899724	0.905330	0.574660
Height	0.827554	0.833684	1.000000	0.819221	0.774972	0.798319	0.817338	0.557467
Whole weight	0.925261	0.925452	0.819221	1.000000	0.969405	0.966375	0.955355	0.540390
Shucked weight	0.897914	0.893162	0.774972	0.969405	1.000000	0.931961	0.882617	0.420884
Viscera weight	0.903018	0.899724	0.798319	0.966375	0.931961	1.000000	0.907656	0.503819
Shell weight	0.897706	0.905330	0.817338	0.955355	0.882617	0.907656	1.000000	0.627574
Rings	0.556720	0.574660	0.557467	0.540390	0.420884	0.503819	0.627574	1.000000

In [11]:
```
import seaborn as sb
sb.heatmap(df.corr())
```

Out[11]: <AxesSubplot:>



In [12]:
```
cat_col=list(set(df.columns)-set(df.describe().columns))
```

In [13]:
```
cat_col
```

Out[13]: ['Sex']

In [14]:
```
from sklearn.preprocessing import LabelEncoder
```

In [15]:
```
for each_cat in cat_col:
    le=LabelEncoder()
    df[each_cat]=le.fit_transform(df[each_cat])
```

In [16]:
```
df
```

Out[16]:

	Sex	Length	Diameter	Height	Whole weight	Shucked weight	Viscera weight	Shell weight	Rings
0	2	0.455	0.365	0.095	0.5140	0.2245	0.1010	0.1500	15
1	2	0.350	0.265	0.090	0.2255	0.0995	0.0485	0.0700	7
2	0	0.530	0.420	0.135	0.6770	0.2565	0.1415	0.2100	9
3	2	0.440	0.365	0.125	0.5160	0.2155	0.1140	0.1550	10
4	1	0.330	0.255	0.080	0.2050	0.0895	0.0395	0.0550	7
...	...	...	...	...	...	...	...	...	...
4172	0	0.565	0.450	0.165	0.8870	0.3700	0.2390	0.2490	11
4173	2	0.590	0.440	0.135	0.9660	0.4390	0.2145	0.2605	10
4174	2	0.600	0.475	0.205	1.1760	0.5255	0.2875	0.3080	9
4175	0	0.625	0.485	0.150	1.0945	0.5310	0.2610	0.2960	10
4176	2	0.710	0.555	0.195	1.9485	0.9455	0.3765	0.4950	12

4177 rows × 9 columns

In [17]:
```
df['Sex'].value_counts()
```

Out[17]:
```
2    1528
1    1342
0    1307
Name: Sex, dtype: int64
```

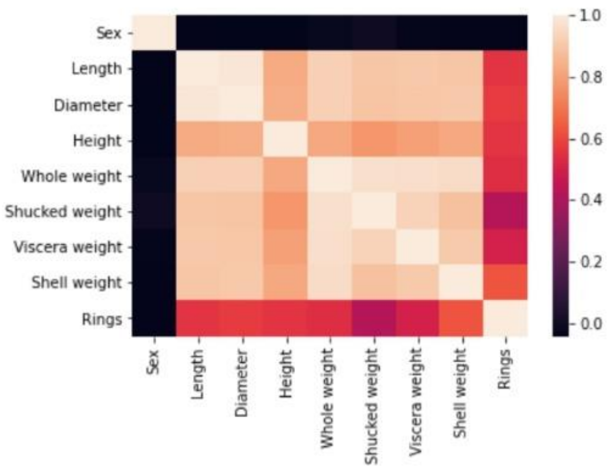
In [18]:
```
df.corr()
```

Out[18]:

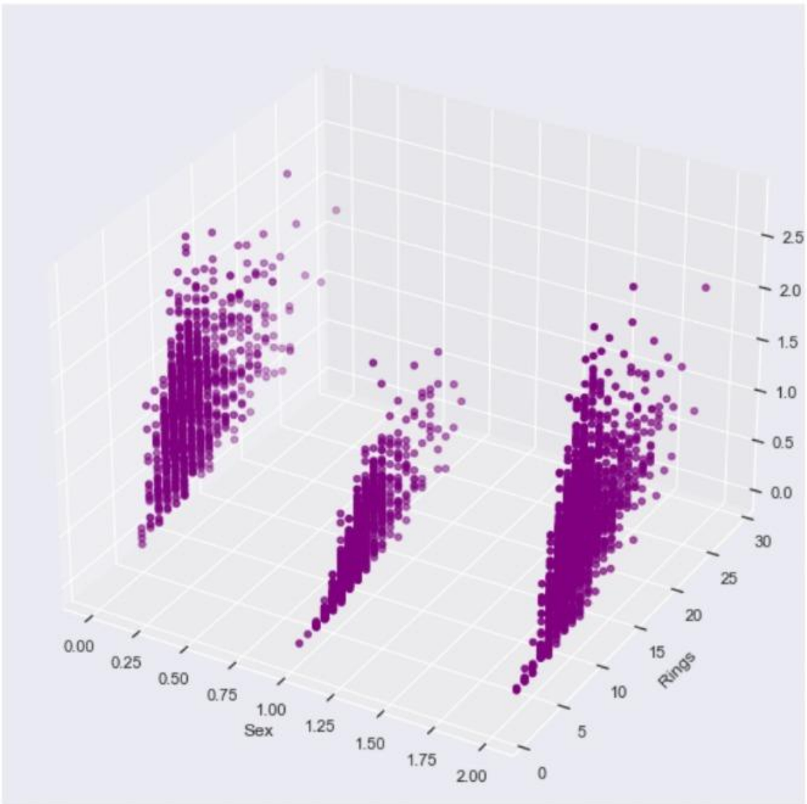
	Sex	Length	Diameter	Height	Whole weight	Shucked weight	Viscera weight	Shell weight	
Sex	1.000000	-0.036066	-0.038874	-0.042077	-0.021391	-0.001373	-0.032067	-0.034854	-0.
Length	-0.036066	1.000000	0.986812	0.827554	0.925261	0.897914	0.903018	0.897706	0.
Diameter	-0.038874	0.986812	1.000000	0.833684	0.925452	0.893162	0.899724	0.905330	0.
Height	-0.042077	0.827554	0.833684	1.000000	0.819221	0.774972	0.798319	0.817338	0.
Whole weight	-0.021391	0.925261	0.925452	0.819221	1.000000	0.969405	0.966375	0.955355	0.
Shucked weight	-0.001373	0.897914	0.893162	0.774972	0.969405	1.000000	0.931961	0.882617	0.
Viscera weight	-0.032067	0.903018	0.899724	0.798319	0.966375	0.931961	1.000000	0.907656	0.
Shell weight	-0.034854	0.897706	0.905330	0.817338	0.955355	0.882617	0.907656	1.000000	0.
Rings	-0.034627	0.556720	0.574660	0.557467	0.540390	0.420884	0.503819	0.627574	1.

In [19]:
```
import seaborn as sb
sb.heatmap(df.corr())
```

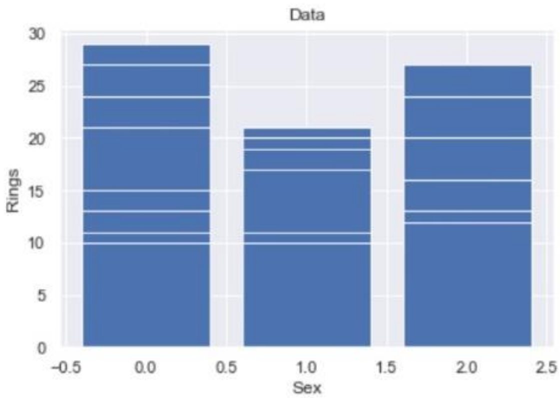
Out[19]: <AxesSubplot:>



In [20]:
```
import matplotlib.pyplot as plt
import pandas as pd
import seaborn as sns
from mpl_toolkits.mplot3d import Axes3D


sns.set(style = "darkgrid")

fig = plt.figure(figsize=(20,10))
ax = fig.add_subplot(111, projection = '3d')

x = df['Sex']
y = df['Rings']
z = df['Whole weight']

ax.set_xlabel("Sex")
ax.set_ylabel("Rings")
ax.set_zlabel("Whole weight")

ax.scatter(x, y, z,c='purple')

plt.show()
```



In [21]:
```
import matplotlib.pyplot as plt
S = df["Sex"]
R = df["Rings"]
x=[]
y=[]
x=list(S)
y=list(R)
plt.bar(x,y)
plt.xlabel('Sex')
plt.ylabel('Rings')
plt.title('Data')
plt.show()
```



In [22]:
```
X=df[['Sex','Length','Diameter','Height','Whole weight','Shucked weight','Viscera
Y=df[['Rings']].values
```

In [23]:
```
from sklearn.model_selection import train_test_split
```

In [24]:
```
from sklearn.datasets import make_blobs
```

In [25]:
```
X, Y = make_blobs(n_samples=1000, centers=3, random_state=1)
```

In [26]:
```
X_train,X_test,Y_train,Y_test= train_test_split(X,Y,train_size=0.7)
```

In [27]:
```
X_train.shape,X_test.shape,Y_train.shape,Y_test.shape
```

Out[27]:
```
((700, 2), (300, 2), (700,), (300,))
```

In [28]:
```
from sklearn.linear_model import LogisticRegression
```

```
In [29]: lr = LogisticRegression()
```

```
In [30]: lr.fit(X_train,Y_train)
```

```
Out[30]: LogisticRegression()
```

```
In [31]: Y_pred=lr.predict(X_test)
```

```
In [32]: Y_pred
```

```
Out[32]: array([1, 2, 1, 0, 0, 0, 2, 0, 1, 1, 2, 0, 0, 2, 2, 0, 1, 2, 1, 2, 2, 2,
                0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 2, 1, 1, 2, 1, 0, 1, 0, 0, 1, 0, 0,
                2, 2, 2, 0, 1, 1, 0, 2, 1, 0, 2, 1, 1, 0, 2, 0, 0, 1, 2, 2, 0, 2,
                2, 2, 1, 2, 2, 1, 2, 2, 1, 2, 2, 2, 2, 2, 0, 1, 1, 0, 1, 1, 1, 0,
                1, 1, 1, 2, 1, 1, 1, 0, 1, 0, 2, 0, 0, 2, 0, 1, 1, 2, 1, 1, 1, 2,
                2, 1, 2, 0, 2, 0, 1, 0, 0, 1, 0, 1, 0, 2, 0, 0, 2, 1, 2, 0, 1, 2,
                0, 1, 1, 2, 1, 0, 0, 0, 0, 2, 2, 2, 1, 2, 2, 0, 1, 1, 2, 1, 2, 0,
                1, 1, 2, 2, 0, 1, 1, 2, 0, 2, 0, 1, 0, 1, 2, 2, 2, 2, 2, 2, 0, 2,
                1, 0, 0, 0, 2, 0, 2, 1, 2, 0, 2, 1, 1, 0, 0, 0, 1, 1, 0, 1, 1, 0,
                0, 2, 2, 1, 2, 1, 1, 2, 0, 2, 1, 2, 0, 0, 1, 2, 1, 1, 0, 1, 0, 2,
                2, 0, 1, 0, 1, 2, 0, 1, 2, 0, 2, 2, 0, 1, 0, 1, 1, 0, 0, 0, 1, 2,
                0, 2, 0, 0, 2, 0, 1, 0, 1, 0, 2, 2, 2, 2, 2, 1, 0, 0, 1, 0, 1, 0,
                0, 0, 1, 2, 2, 0, 0, 1, 2, 0, 2, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 1,
                1, 0, 1, 1, 1, 2, 1, 1, 0, 2, 2, 0, 1, 2])
```

```
In [33]: Y_test
```

```
Out[33]: array([1, 2, 1, 0, 0, 0, 2, 0, 1, 1, 2, 0, 0, 2, 2, 0, 1, 2, 1, 2, 2, 2,
                0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 2, 1, 1, 2, 1, 0, 1, 0, 0, 1, 0, 0,
                2, 2, 2, 0, 1, 1, 0, 2, 1, 0, 2, 1, 1, 0, 2, 0, 0, 1, 2, 2, 0, 2,
                2, 2, 1, 2, 2, 1, 2, 2, 1, 2, 2, 2, 2, 2, 0, 1, 1, 0, 1, 1, 1, 0,
                1, 1, 1, 2, 1, 1, 1, 0, 1, 0, 2, 0, 0, 1, 0, 1, 1, 2, 1, 1, 1, 2,
                2, 1, 2, 0, 2, 0, 1, 0, 0, 1, 0, 1, 0, 2, 0, 0, 2, 1, 2, 0, 1, 2,
                0, 1, 1, 2, 1, 0, 0, 0, 0, 2, 2, 2, 1, 2, 2, 0, 1, 1, 2, 1, 2, 0,
                1, 1, 2, 2, 0, 1, 1, 2, 0, 2, 0, 1, 0, 1, 2, 2, 2, 2, 2, 2, 0, 2,
                1, 0, 0, 0, 2, 0, 2, 1, 2, 0, 2, 1, 1, 0, 0, 0, 1, 1, 0, 1, 1, 0,
                0, 2, 2, 1, 2, 1, 1, 2, 0, 2, 1, 2, 0, 0, 1, 2, 1, 1, 0, 1, 0, 2,
                2, 0, 1, 0, 1, 2, 0, 1, 2, 0, 2, 2, 0, 1, 0, 1, 1, 0, 0, 0, 1, 2,
                0, 2, 0, 0, 2, 0, 1, 0, 1, 0, 2, 2, 2, 2, 2, 1, 0, 0, 1, 0, 1, 0,
                0, 0, 1, 2, 2, 0, 0, 1, 2, 0, 2, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 1,
                1, 0, 1, 1, 1, 2, 1, 1, 0, 2, 2, 0, 1, 2])
```

```
In [34]: from sklearn.preprocessing import MinMaxScaler
```

```
In [35]: norm_data=MinMaxScaler().fit_transform(X)
```

```
In [36]: from sklearn.neural_network import MLPClassifier
```

```
In [37]: mlp=MLPClassifier(hidden_layer_sizes=(20,4), activation = 'relu', solver = 'adam'
```

```
In [38]: mlp.fit(X_train,Y_train)

Iteration 1, loss = 1.38306661
Iteration 2, loss = 1.28375159
Iteration 3, loss = 1.19913381
Iteration 4, loss = 1.12641247
Iteration 5, loss = 1.06509369
Iteration 6, loss = 1.01604734
Iteration 7, loss = 0.97284907
Iteration 8, loss = 0.93325222
Iteration 9, loss = 0.89204576
Iteration 10, loss = 0.85022527
Iteration 11, loss = 0.80730355
Iteration 12, loss = 0.76413490
Iteration 13, loss = 0.72256917
Iteration 14, loss = 0.68152574
Iteration 15, loss = 0.64040605
Iteration 16, loss = 0.60211241
Iteration 17, loss = 0.56402028
Iteration 18, loss = 0.52665627
Iteration 19, loss = 0.49157120

In [39]: Y_pred_nn=mlp.predict(X_test)

In [40]: Y_test

Out[40]: array([1, 2, 1, 0, 0, 0, 2, 0, 1, 1, 2, 0, 0, 2, 2, 0, 1, 2, 1, 2, 2, 2,
       0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 2, 1, 1, 2, 1, 0, 1, 0, 0, 1, 0, 0,
       2, 2, 2, 0, 1, 1, 0, 2, 1, 0, 2, 1, 1, 0, 2, 0, 0, 1, 2, 2, 0, 2,
       2, 2, 1, 2, 2, 1, 2, 2, 1, 2, 2, 2, 2, 2, 0, 1, 1, 0, 1, 1, 1, 0,
       1, 1, 1, 2, 1, 1, 1, 0, 1, 0, 2, 0, 0, 1, 0, 1, 1, 2, 1, 1, 1, 2,
       2, 1, 2, 0, 2, 0, 1, 0, 0, 1, 0, 1, 0, 2, 0, 0, 2, 1, 2, 0, 1, 2,
       0, 1, 1, 2, 1, 0, 0, 0, 0, 2, 2, 2, 1, 2, 2, 0, 1, 1, 2, 1, 2, 0,
       1, 1, 2, 2, 0, 1, 1, 2, 0, 2, 0, 1, 0, 1, 2, 2, 2, 2, 2, 2, 0, 2,
       1, 0, 0, 0, 2, 0, 2, 1, 2, 0, 2, 1, 1, 0, 0, 0, 1, 1, 0, 1, 1, 0,
       0, 2, 2, 1, 2, 1, 1, 2, 0, 2, 1, 2, 0, 0, 1, 2, 1, 1, 0, 1, 0, 2,
       2, 0, 1, 0, 1, 2, 0, 1, 2, 0, 2, 2, 0, 1, 0, 1, 1, 0, 0, 0, 1, 2,
       0, 2, 0, 0, 2, 0, 1, 0, 1, 0, 2, 2, 2, 2, 2, 1, 0, 0, 1, 0, 1, 0,
       0, 0, 1, 2, 2, 0, 0, 1, 2, 0, 2, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 1,
       1, 0, 1, 1, 1, 2, 1, 1, 0, 2, 2, 0, 1, 2])
```

```
       0, 1, 2, 1, 2, 2, 0, 1, 2, 0, 2, 0, 2, 1])

In [42]: from sklearn.metrics import accuracy_score

In [43]: accuracy_score(Y_test,Y_pred_nn)

Out[43]: 0.9933333333333333

In [44]: accuracy_score(Y_test,Y_pred)

Out[44]: 0.9933333333333333
```

CHAPTER 7: Reference

https://archive.ics.uci.edu/ml/datasets/abalone

https://www.kaggle.com/rodolfomendes/abalone-dataset

https://data.world/uci/abalone

https://searchcontentmanagement.techtarget.com/definition/Microsoft-Power-BI