A TECHNICAL REPORT ON

STUDENT INDUSTRIAL WORK EXPERIENCE SCHEME (SIWES)

(SEPTEMBER 2023 – FEBUARY 2024)

UNDERTAKEN AT

BABTECH COMPUTERS
ANISHERE BUSTOP, GOVERNOR ROAD, IKOTUN, LAGOS.

BY

OLAYODE, WASILAT AMOKE

MATRIC NO: EES/19/20/0441

SUBMITTED TO

THE DEPARTMENT OF COMPUTER ENGINEERING

COLLEGE OF ENGINEERING AND ENVIRONMENTAL STUDIES

OLABISI ONABANJO UNIVERSITY (IBOGUN CAMPUS), OGUN STATE.

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF

BACHELOR OF ENGINEERING (B.Eng) DEGREE IN COMPUTER
ENGINEERING

CERTIFICATION

This is to certify that this technical report is a detailed account of Student Industrial Work Experience Scheme (SIWES) Training undertaken by **OLAYODE WASILAT AMOKE** with Matriculation Number EES/19/20/0441 from COMPUTER ENGINEERING Department, OLABISI ONABANJO UNIVERSITY (OOU), IBOGUN CAMPUS, OGUN STATE, NIGERIA, for a period of five months and has been prepared in accordance to regulations guiding the preparation of reports in Computer Engineering Department.

_____   _____
   ENGR OYEDEJI   DATE
SIWES CO-ORDINATOR

_____   _____
   ENGR. ABOLADE   DATE
HEAD OF DEPARTMENT

_____

STUDENT'S SIGNATURE AND DATE

DEDICATION

I extend my heartfelt dedication to my beloved parents and siblings for their unwavering support and encouragement in every aspect of my life. Their love and belief in me have served as a constant source of motivation throughout my academic journey.

Furthermore, I express my gratitude to my mentor and supervisor, whose guidance and support were indispensable during my SIWES program. Their expertise and encouragement played a pivotal role in shaping my professional growth.

Lastly, I dedicate this report to all fellow SIWES trainees, wishing that this experience serves as a significant milestone toward the attainment of your academic and professional aspirations.

ACKNOWLEDGEMENT

My deepest and sincerest gratitude is extended to the Almighty Allah, the compassionate and merciful, who has bestowed upon me the precious gift of life and guided me throughout my SIWES training. May His name be forever exalted.

I extend my appreciation to all who played a role, big or small, in the success of my industrial training. A special acknowledgment is reserved for Mr. Olayemi Idris Babatunde, the CEO of BABTECH COMPUTERS, for granting me the opportunity to undergo training in his establishment and for providing unwavering support throughout the entire program.

Additionally, I would like to express gratitude to my family and friends for their steadfast support and encouragement throughout my academic journey.

Lastly, I want to convey my thanks to my fellow SIWES trainees for their camaraderie and support, which greatly contributed to the positive experience during this program.

ABSTRACT

This report presents a comprehensive overview of the Student Industrial Work Experience Scheme (SIWES) program undertaken by the author at **BABTECH COMPUTERS**. The primary objective of this program was to offer practical experience and facilitate the transition from theoretical knowledge to practical application. Throughout the training period, the author miss Olayode Wasilat actively engaged in hands-on activities related to Data Science, collaborating with students and seasoned industry professionals. The was also able to actively share her knowledge by participating in production, teaching and guiding other student on Data Science, Product Design and Data Processing courses.

This document aims to illuminate the author's firsthand experiences, practice work, address encountered challenges, and outline the valuable lessons assimilated during the program, ultimately equipping the author with the necessary skills for a promising Tech carrier.

TABLE OF CONTENT

Chapter Five

Chapter Six

LIST OF FIGURES

CHAPTER ONE

1.1 **SIWES INTRODUCTION**

Student Industrial Work Experience Scheme (SIWES) is a mandatory skills training program designed by the Nigerian government to help students in higher education gain practical knowledge and experience in their respective fields of study by working in industries and organizations related to their course of study. The scheme was designed to expose and prepare the undergraduates of Nigerian Universities, Polytechnics, and Colleges for the industrial work situation they are likely to meet after graduation. SIWES is a collaborative effort between the government, universities, and industries. It is aimed at fostering the relationship between the academic and industrial sectors by providing students with the opportunity to gain hands-on experience in their chosen fields of study.

1.2 **BRIEF HISTORY OF SIWES**

The Student Industrial Work Experience Scheme (SIWES) is a special, yet mandatory program initiated in October 8, 1973 and is coordinated by the National Universities Commission (NUC), the National Board for Technical Education (NBTE), and the National Commission for Colleges of Education (NCCE) in collaboration with the Industrial Training Fund (ITF) and various industrial organizations. This program was created to bridge the gap between employers' expectations and the actual performance of Nigerian graduates (i.e., the gap between the theoretical knowledge acquired by students in higher institutions and the practical knowledge needed for them to acquire). Hence, the need to allow undergraduates to have real-life working experience related to their field of study.

Before the establishment of the scheme, there was a great concern for students by industrialists because they had insufficient practical knowledge/experience for employment; and they could not find jobs easily because of this.

It was in the year 1973 that the Industrial Training Funds (ITF) decided to help all Nigerian students by introducing the SIWES program and it was officially approved and presented by the Federal Government in 1974. The scheme was solely funded by the ITF in the early years of its establishment; but due to the financial responsibilities becoming unbearable for them, the ITT withdrew from the scheme in 1978. In 1979, the Federal government handed over the scheme to the National Universities Commission (NUC), a regulatory board for universities, and its counterpart, the National Board for Technical Education (NBTE), for colleges of education.In November 1984 management and implementation of the Scheme was again reverted to the ITF with the funding to be solely borne by the Federal Government, and it has been like that till date

Undoubtedly, partaking in this program has become a pre-requisite for the award of certificates in some fields of study in most Nigerian universities, polytechnics, and colleges of education.

1.3 **OBJECTIVES OF SIWES**

Objectives of the Students' Industrial Work Experience Scheme include:

➢ Skills Development
- Provide an avenue for students to acquire industrial skills for experience during their course of study.

➢ Practical Knowledge
- It exposes students to work methods and techniques in handling equipment and machines that may not be available in their institutions.
- Bridging the gap between theory and practice by providing a platform to apply knowledge learnt in school to real work situations.

➢ Field Experience
- Provide Students with an opportunity to apply their knowledge in real work situations, thereby bridging the gap between theory and practice.
- It prepares the students for the industrial work situation they are likely to meet after graduation.

1.4 **ITF IN RELATION TO SIWES**

The Industrial Training Fund (ITF) is a Nigerian government agency established in 1971 with the aim of promoting and encouraging the acquisition of skills in the industrial and commercial sectors of Nigeria. Operating under the Federal Ministry of Industry, Trade, and Investment, the ITF's mandate is to foster human resource development by providing training grants, research, and consultancy services to both public and private sectors.

The ITF is entrusted with the management and disbursement of training funds contributed by employers of labor in Nigeria. These funds play a crucial role in supporting the training and development of Nigerian citizens across various fields, such as engineering, technology, business management, and vocational skills.

Moreover, the ITF oversees a network of vocational training centers across the country, offering practical skills training to individuals, thereby preparing them for the workforce.

CHAPTER TWO

2.1 COMPANY OVERVIEW

Babtech Computers is a technology company established in 2014 and has grown to become a well-regarded player in this industry providing reliable, efficient, and trustworthy services both in computer education and computer application solutions.

The company was established recognizing the potential of the increasing market of companies & individuals that are or would be looking to make their business processes more efficient by automating them through the use of advanced software solutions and also in response to the current market forces that dictate the necessity of online presence to stay in business and increase profit.

Since it was founded in the year 2010, the company has been consistent in delivering first-class computer education and producing highly skilled professionals. The company has also been involved in releasing numerous projects for diverse business spheres ranging from basic websites to highly sophisticated business automation systems. Our proficiency in a wide range of modern computer technologies ensures that the solutions we develop perfectly suit our clients' needs.

2.2 ORGANIZATION'S VISION AND MISSION

2.2.1 Vision Statement

Our Vision is to be the foremost computer education provider and be the first choice of our clients in all our business arms.

2.2.2 Mission Statement

To deliver first-class computer education, build talents, to provide our clients with the most optimal, cost-effective and robust solutions and create entrepreneurs and shakers of the business world.

2.3 Organization's Area of Specialization

- Computer Education
  We are devoted to providing first-class computer education in all the courses that we offer. We are dynamic in our approach and we stay abreast with industry trends and emerging technological tools and features. We are bringing quality computer education to you at affordable prices you can't get anywhere.
- Consultancy

We spend quality time working with individuals, new businesses and startups, so whether you are working towards delivering IT projects or trying to change the world, we are your partners!

2.4 Organization's Core Values

- EXCELLENCE: We display excellence in everything we do.
- GROWTH: We have tireless hunger for new skills.
- SATISFACTION: We make imparts that translate to better businesses for our clients and rewarding careers for our students and interns
- INNOVATIVE: We do things differently.
- COLLABORATION: We work as a team and value everyone.

CHAPTER THREE

3.1 OVERALL WORK DONE

During my SIWES program at Babtech Computers, I had the privilege of being supervised and mentored by an exceptional instructor who guided my training and projects in the field of data science. During this period, I actively contributed in the production environment by teaching data processing courses that covered essential skills, including internet usage, system navigation, and comprehensive lessons on Microsoft packages such as Word, Excel, PowerPoint, and Access. Additionally, I extended my teaching responsibilities to product design students, providing valuable insights based on Figma lessons.

Furthermore, I got the opportunity to educate data students on Power BI, crafting and delivering lessons on its features and functionality. I provided hands-on guidance to assist students in creating their own data visualizations, offering feedback and constructive advice to enhance their work and achieve optimal results.

In addition to my instructional role, I also served as a proficient Personal Assistant (PA), actively participating in strategic decision-making processes and contributing to the overall organizational efficiency.

3.2 DATA SCIENCE

Data science is an interdisciplinary domain that employs statistical and computational techniques to derive insights and knowledge from data. Its primary objective is to discern patterns, relationships, and trends within extensive and intricate data sets, utilizing this information to make informed decisions and predictions.

The field of data science encompasses a diverse set of skills and techniques, including data cleaning and preprocessing, data visualization, statistical analysis, machine learning, and deep learning. Data scientists engage with large and varied data sets, often employing programming languages such as Python, R, and SQL for data analysis and modeling.

The applications of data science span across various sectors such as finance, healthcare, marketing, and social media. For instance, in finance, data science is instrumental in identifying patterns within financial data to inform investment decisions, while in healthcare, it is utilized to analyze patient data for enhanced health outcomes.

In essence, data science is a dynamically expanding field propelled by the increasing availability of data and the imperative for businesses and organizations to base decisions on data-driven insights. Consequently, a background in data science positions individuals favorably for diverse career opportunities across industries.

3.2.1 Key Components

Data science is a multidisciplinary field that incorporates various components to extract valuable insights from data. Here are the some key components of data science:

1. Data Collection and Acquisition:
   - Involves gathering raw data from various sources, such as databases, APIs, sensors, or external datasets.
   - Data collection methods should ensure accuracy, completeness, and relevance to the analysis goals.
2. Data Cleaning and Preprocessing:
   - Addresses issues like missing data, outliers, and inconsistencies to ensure the dataset's quality.
   - Includes tasks like imputation, normalization, and transformation to make the data suitable for analysis.
3. Exploratory Data Analysis (EDA):
   - Involves the use of statistical and graphical techniques to understand the characteristics of the data.
   - Helps identify patterns, trends, and outliers, providing insights into potential relationships.
4. Data Visualization:
   - Utilizes graphical representations to present complex data in an understandable format.
   - Helps communicate findings, trends, and patterns effectively to both technical and non-technical audiences.
5. Statistical Analysis:
   - Involves the application of statistical methods to test hypotheses, make inferences, and validate findings.
   - Provides a foundation for understanding the reliability and significance of results.
6. Machine Learning:
   - Encompasses a set of algorithms and techniques that allow computers to learn patterns from data.
   - Includes supervised learning for prediction, unsupervised learning for pattern discovery, and reinforcement learning for decision-making.
7. Model Evaluation and Validation:

- Assesses the performance of machine learning models using metrics like accuracy, precision, recall, and F1 score.
- Involves techniques such as cross-validation to ensure the model's generalizability.

8. Deployment and Implementation:
   - The process of integrating models into real-world applications and systems.
   - Requires considerations for scalability, efficiency, and ongoing monitoring.

9. Domain Knowledge:
   - Understanding the specific industry or domain for which data science is applied.
   - Helps in framing meaningful questions, interpreting results in context, and making informed decisions.

10. Ethics and Privacy:
    - Addresses the responsible and ethical use of data, ensuring privacy and compliance with regulations.
    - Data scientists must be aware of the potential impact of their work on individuals and society.

3.2.2 Data Analysis

Data analysis is a dynamic process, encompassing the meticulous examination, cleaning, transformation, and interpretation of data, all aimed at uncovering patterns, trends, and relationships that contribute to a comprehensive understanding of the underlying information in order to derive meaningful insights. As an integral component of data science, it synergizes with statistical techniques, exploring datasets to uncover patterns and trends. Data analysis forms the bedrock for strategic decision-making, translating raw data into actionable insights. Through tools like Ms Excel, Power BI, Python, etc.

It aligns with the broader goals of data science, facilitating the extraction of valuable knowledge from complex datasets. In essence, data analysis is the bridge that connects raw data to informed decision-making, playing a pivotal role within the overarching framework of data science.

Moreover, data analysis serves as a means of translating raw data into actionable insights, thereby supporting strategic decision-making processes. Analysts employ tools like PowerBI and various methodologies to distill complex datasets into meaningful visualizations, making the information accessible and comprehensible to

diverse stakeholders. This communicative aspect is vital for conveying findings effectively and driving data-driven decision-making across organizations.

3.2.3 Statistics

Statistics forms the cornerstone of data science, furnishing fundamental knowledge and tools crucial for comprehending, analyzing, and interpreting data. In the realm of data science, statistical methods serve to unveil patterns, relationships, and trends, enabling predictions and informed decision-making.

Key statistical approaches include

- descriptive statistics; summarizing data with measures like mean and standard deviation.
- inferential statistics; making predictions based on a sample.
- hypothesis testing; validating claims about populations
- regression analysis; modeling relationships between variables.
- time series analysis, unveiling patterns over time.

In essence, statistics is indispensable throughout the data science workflow, guiding activities from data cleaning and preparation to modeling and analysis, and is paramount for sound decision-making grounded in data.

3.2.4 Machine learning

Machine learning, a subset of artificial intelligence, revolves around crafting algorithms and models that empower computer systems to learn from data and enhance their performance progressively. It encompasses diverse techniques, including supervised learning, unsupervised learning, and reinforcement learning, finding applications in image recognition, natural language processing, predictive analytics, and more.

The three primary types of machine learning are as follows:

- **Supervised Learning:** In this approach, algorithms are trained on labeled datasets, where each data point has a known output. The objective is for the algorithm to discern the relationship between input variables and the output variable, enabling predictions for new, unseen data.
- **Unsupervised Learning:** Unlike supervised learning, this method involves training algorithms on unlabeled datasets without known output variables. The

goal is for the algorithm to autonomously discover patterns and relationships within the data, devoid of external guidance.

- **Reinforcement Learning:** Here, the algorithm learns through trial and error via interaction with an environment. It receives feedback in the form of rewards or penalties for its actions, with the ultimate aim of maximizing accumulated rewards over time.

3.3 DATA SCIENCE PROGRAMMING LANGUAGES

Here are some programming languages commonly used in data science:

- Python: Python is one of the most popular programming languages for data science due to its simplicity, readability, and vast range of libraries and frameworks, such as NumPy, Pandas, and Scikit-learn.
- R: R is another popular programming language for data science that offers a wide range of statistical and graphical techniques for data analysis, visualisation, and modelling.
- SQL: SQL (Structured Query Language) is a language used for managing relational databases, which are commonly used in data science for storing and retrieving large amounts of structured data.
- Java: Java is a general-purpose programming language that is widely used in data science for building large-scale applications that require high-performance and scalability.
- MATLAB: MATLAB is a programming language that is widely used in engineering, science, and mathematics for numerical analysis and data visualisation.

Overall, the choice of programming language depends on the specific needs and requirements of the data science project, as well as the programmer's skills and expertise.

During the course of my SIWES program I was able to learn and also work with two major data science programming languages (Python and SQL)

3.3.1 Python Language

Python is a high-level, interpreted programming language that is widely used in data science due to its simplicity, readability, and extensive range of libraries and frameworks.

Here are some of the features that make Python a popular choice for data science:

- **User-Friendly:** Python has a simple syntax that is easy to read and write, making it an ideal language for beginners to learn.
- **Abundant Libraries and Frameworks:** The language offers an extensive array of dedicated data science libraries and frameworks, such as NumPy, Pandas, Scikit-learn, TensorFlow, matplotlib, etc. Thereby simplifying intricate tasks in data analysis and machine learning.
- **Open-Source Nature:** Python's open-source status ensures that its code is freely accessible, modifiable, and distributable by the global community.
- **Cross-Platform Compatibility:** Python's versatility shines through its ability to operate seamlessly on diverse operating systems, including Windows, macOS, and Linux, providing flexibility for data science projects.
- **Vibrant Community Support:** Python benefits from a dynamic and expansive community of developers and users actively contributing to the creation of new libraries, frameworks, and tools. This vibrant community also offers substantial support and resources for individuals at varying skill levels, especially those just beginning their journey in Python.

3.3.1.1 Terminologies in Python

Here are some common terminologies used in Python programming:

- Data Types:
    - Integer: An integer is a whole number without a decimal point.
      Example include: 12, -20, 4.
    - Float: A float is a number with a decimal point.
      Example include: 5.22, -6.7, 3.5,0.0.
    - String: A string is a sequence of characters enclosed in quotes.
      Example include: "Computer", 'Babtech', "32".
    - Boolean: A Boolean is a value that is either True or False.
      For example: True, False.
    - List: A list is a collection of values that are ordered and changeable.
      In Python, lists are defined using square brackets [] and can contain any type of value separated by commas.

For example: [4,7,9], ["Book", "Pen", "Food"].

- Tuple: A tuple is a collection of values that are ordered and unchangeable. In Python, tuples are defined using parentheses () and can contain any type of value separated by commas.

  For example: (1, 2, 3), ("Ball", "Cat", "Dog").

- Dictionary: A dictionary is a collection of key-value pairs that are unordered and changeable.
  In Python, dictionaries are defined using curly braces {} and are accessed using keys separated by commas.

  For example: {'Name': 'Bola', 'Age': 30, 'Country': 'Nigeria}.

- Variable: A variable is a named location in memory that stores a value. In Python, variables are created automatically when a value is assigned to them.
  Company_name = "Babtech Computer"

  In this example, company_name is the variable and "Babtech Computer" is the value assigned to it. The equal sign (=) is used to assign a value to a variable.

- Function: A function is a block of reusable code that performs a specific task. In Python, functions are defined using the "def" keyword and can be called with arguments.

  Example:

  ```
  def product(item):
      print ("Hy, your " + item + " is ready to be delivered")
  product("Chair")
  ```

  In this example, product is the function that takes a parameter named item. When the function is called with the argument "Chair", it prints out the message " Hy, your Chair is ready to be delivered".

- Conditional Statement: Conditional statements are used to control the flow of a program based on certain conditions.

They allow you to execute different blocks of code depending on the value of a variable or the result of a comparison.

Example:
```
z = 4
 if z > 0:
        print ("z is positive")
 elif z < 0:
        print ("z is negative")
 else:
        print ("z is zero")
```
In this example, we use the conditional statement to check if z is positive, negative, or zero. If z is greater than 0, we print the message "z is positive". If z is less than 0, we print the message "z is negative". If z is equal to 0, we print the message "z is zero".

The elif statement allows us to check for multiple conditions. If the first condition is not true, we check the second condition. If none of it is true, we execute the else block.

- Loop: A loop is a control structure in Python that repeats a block of code until a condition is met.

  There are two types of loops in Python: "**FOR**" loops and "**WHILE**" loops.

  - For loop:

```
fruits = ["Pineapple", "Banana", "Orange"]

for fruit in fruits:

    print (fruit)
```

In this example, fruits is a list of strings. We use a for loop to iterate over the elements of the list and print each element on a new line.

The loop variable fruit takes on the value of each element in the list, one at a time and print them out. The loop continues until all elements have been processed.

  - While loop:

```
i = 0

while i < 7:

    print(i)
```

$$i += 1$$

In this example, we use a while loop to print the numbers from 0 to 6. The loop continues as long as the condition i < 7 is true.

The variable i is initialized to 0 before the loop. Inside the loop, we print the value of i and increment it by 1 using the += operator. The loop continues until i is no longer less than 7.

Overall, these are just some of the many terminologies used in Python programming. It's important to have a good understanding of these terms in order to write effective and efficient Python code.

3.3.1.2 Snippet of my Python Code

```
In [8]: sales = pd.read_csv("sales.csv")
        sales.head(10)
```

Out[8]:

	Order ID	Region	Sales Channel	Order Priority	Order Date	Ship Date	Shipping duration	Units Sold	Unit Price	Unit Cost	Total Revenue	Total Cost price
0	686900706	Middle East and North Africa	Offline	M	10/18/2014	10/31/2014	13	8446	437.20	263.33	3692591.20	2224085.18
1	185941302	North America	Online	M	11/7/2011	12/8/2011	31	3018	154.06	90.93	464953.08	274426.74
2	246222341	Middle East and North Africa	Offline	C	10/31/2016	12/9/2016	39	1517	255.28	159.42	387259.76	241840.14
3	161442649	Asia	Offline	C	4/10/2010	5/12/2010	32	3322	205.70	117.11	683335.40	389039.42
4	645713555	Sub-Saharan Africa	Offline	H	8/16/2011	8/31/2011	15	3423	9.33	6.92	31936.59	23687.16
5	683458888	Europe	Online	H	11/24/2014	12/28/2014	34	9528	205.70	117.11	1959909.60	1115824.08
6	686800706	Sub-Saharan Africa	Online	H	3/4/2015	4/17/2015	44	2844	205.70	117.11	585010.80	333060.84
7	208630645	Europe	Offline	M	5/17/2012	6/28/2012	42	7299	109.28	35.84	797634.72	261596.16
8	232323233	Central America and the Caribbean	Online	H	1/29/2015	3/7/2015	37	2428	154.06	90.93	374057.68	220778.04
9	118598544	Australia and Oceania	Offline	H	12/24/2013	1/19/2014	26	4800	154.06	90.93	739488.00	436464.00

Fig 3.1 Loading a CSV Dataset with Pandas library

```
In [9]: sales.info()

        <class 'pandas.core.frame.DataFrame'>
        RangeIndex: 1000 entries, 0 to 999
        Data columns (total 12 columns):
         #   Column             Non-Null Count   Dtype
        ---  ------             --------------   -----
         0   Order ID           1000 non-null    int64
         1   Region             1000 non-null    object
         2   Sales Channel      1000 non-null    object
         3   Order Priority     1000 non-null    object
         4   Order Date         1000 non-null    object
         5   Ship Date          1000 non-null    object
         6   Shipping duration  1000 non-null    int64
         7   Units Sold         1000 non-null    int64
         8   Unit Price         1000 non-null    float64
         9   Unit Cost          1000 non-null    float64
         10  Total Revenue      1000 non-null    float64
         11  Total Cost price   1000 non-null    float64
        dtypes: float64(4), int64(3), object(5)
        memory usage: 93.9+ KB
```

Fig 3.2 Using info function to view the datatype and empty cells.

```
In [10]: sales.describe()
Out[10]:
```

	Order ID	Shipping duration	Units Sold	Unit Price	Unit Cost	Total Revenue	Total Cost price
count	1.000000e+03	1000.00000	1000.000000	1000.00000	1000.000000	1.000000e+03	1.000000e+03
mean	5.492576e+08	25.00100	5047.566000	262.10684	184.965110	1.327262e+06	9.360748e+05
std	2.570604e+08	14.60309	2897.865328	216.02106	175.289311	1.486566e+06	1.162605e+06
min	1.029280e+08	0.00000	13.000000	9.33000	6.920000	2.043250e+03	1.416750e+03
25%	3.280740e+08	13.00000	2420.250000	81.73000	56.670000	2.811919e+05	1.649319e+05
50%	5.558113e+08	25.00000	5165.000000	154.06000	97.440000	7.549392e+05	4.647261e+05
75%	7.695114e+08	38.00000	7536.000000	421.89000	263.330000	1.733503e+06	1.141750e+06
max	9.955298e+08	50.00000	9998.000000	668.27000	524.960000	6.617210e+06	5.204978e+06

```
In [11]: sales["Order Date"] = pd.to_datetime(sales["Order Date"])

In [12]: sales["Ship Date"] = pd.to_datetime(sales["Ship Date"])

In [13]: sales["profit"] = sales["Total Revenue"] - sales["Total Cost price"]

In [14]: sales.info()

        <class 'pandas.core.frame.DataFrame'>
        RangeIndex: 1000 entries, 0 to 999
        Data columns (total 13 columns):
         #   Column             Non-Null Count   Dtype
        ---  ------             --------------   -----
```

Fig 3.3 performing simple analysis with Pandas library

```
In [15]: revenue = sales.groupby("Sales Channel")["Total Revenue"].sum()
         revenue

Out[15]: Sales Channel
         Offline    6.988638e+08
         Online     6.283981e+08
         Name: Total Revenue, dtype: float64


In [16]: cost = sales.groupby("Sales Channel")["Total Cost price"].sum()
         cost

Out[16]: Sales Channel
         Offline    4.909920e+08
         Online     4.450828e+08
         Name: Total Cost price, dtype: float64


In [17]: profit = sales.groupby("Sales Channel")["profit"].sum()
         profit

Out[17]: Sales Channel
         Offline    2.078718e+08
         Online     1.833153e+08
         Name: profit, dtype: float64
```

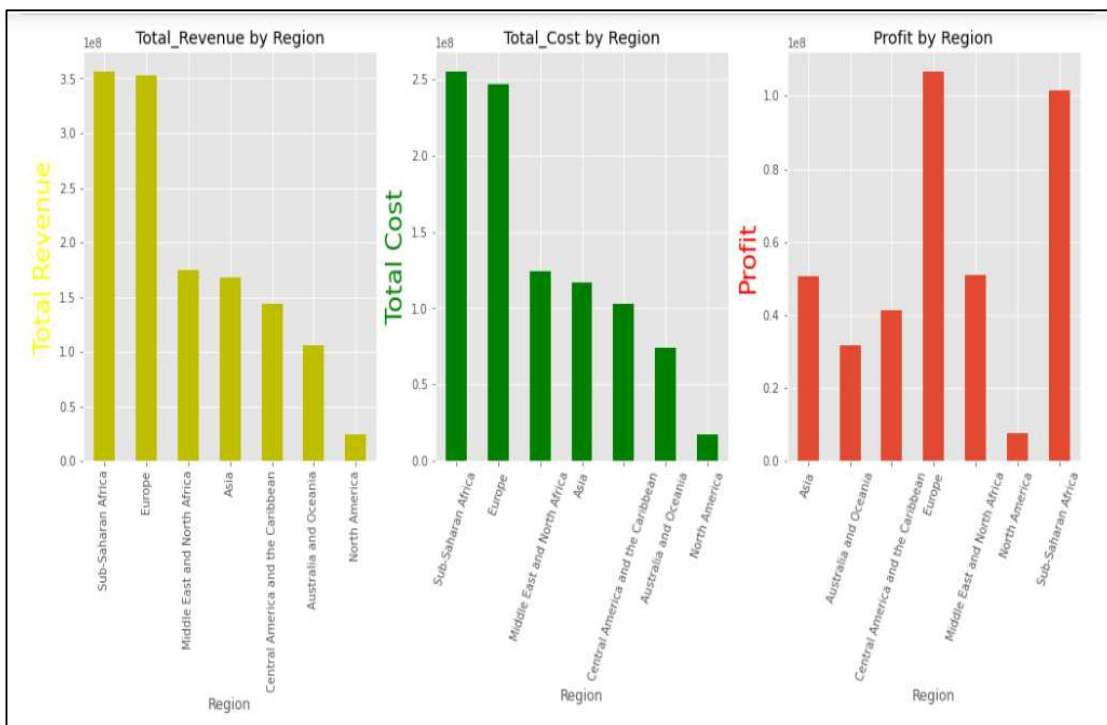Fig 3.4 Performing simple statistical analysis on my sales dataset.

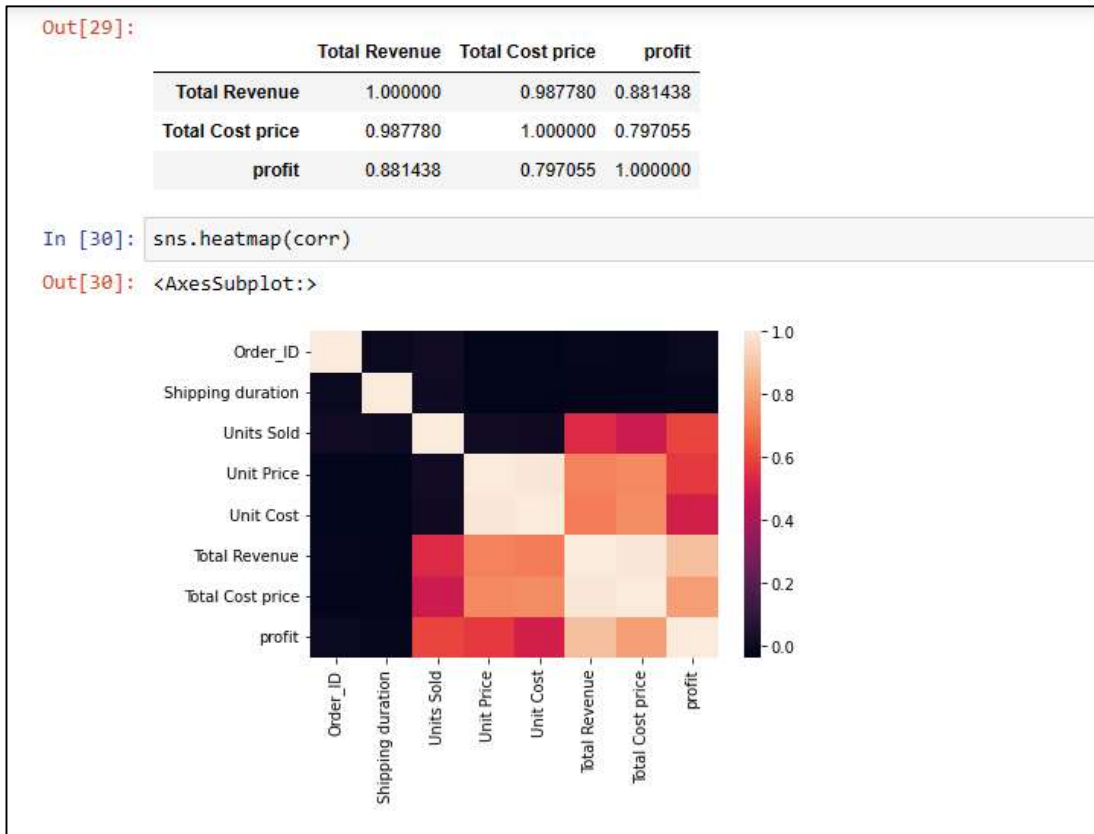

Fig 3.5 Using matplotlib to create Subplot

15

Fig 3.6 Using Seaborn Library to view the correlation between the dataset.

3.3.2 SQL

SQL (Structured Query Language), plays a pivotal role in data science by efficiently managing and organizing large datasets within databases. As an industry-standard language, SQL simplifies the intricate tasks of handling and retrieving data across diverse database systems. Its robust capabilities make it a fundamental tool, enabling data scientists to streamline operations, perform intricate queries, and extract valuable insights from vast repositories of information. The simplicity and versatility of SQL contribute significantly to its widespread adoption, making it an indispensable asset for professionals engaged in data science and database management.

In data science, SQL is commonly used to:

- **Data Retrieval:** SQL's SELECT statement empowers users to fetch specific columns or rows based on defined conditions.
- **Data Cleaning and Transformation**: SQL facilitates data cleaning and transformation by modifying information within the database, covering tasks like updating, deleting, and inserting data.
- **Aggregation and Summarization:** Users leverage SQL for aggregate functions like SUM, COUNT, AVG, MIN, and MAX to derive summary statistics from extensive datasets.
- **Table Joining:** SQL enables the amalgamation of data from multiple sources by joining tables based on a shared column.

3.3.2.1 Terminologies Used In SQL

- Database: A database is a collection of related data that is organised and stored in a structured way. It can be created using the CREATE DATABASE statement.

  Example:

  ```
  -- Create a database named 'mydatabase'
       CREATE DATABASE mydatabase;
  ```

- Table: A table is a collection of related data organised into rows and columns. It can be created using the CREATE TABLE statement.

  Example:

  ```
      -- Create a table named 'customers' with columns for name, email, and
  phone number

           CREATE TABLE customers (
           Name VARCHAR(50),
           email VARCHAR(50),
           phone VARCHAR(15) );
  ```

- Column: A column is a vertical entity in a table that represents a specific attribute of the data being stored. It can be created using the ALTER TABLE statement.

  Example:
  ```
       -- Add a column named 'age' to the 'customers' table
           ALTER TABLE customers ADD age INT;
  ```

- Row: A row is a horizontal entity in a table that represents a single record or instance of the data being stored.

Example:

```
-- Insert a new row into the 'customers' table

INSERT INTO customers (name, email, phone, age)
VALUES ('John Smith', 'john@example.com', '555-1234', 30);
```

- Primary key: A primary key is a column or set of columns in a table that uniquely identifies each row in the table. It can be defined using the PRIMARY KEY constraint.

Example:

```
-- Add a primary key constraint to the 'customers' table on the 'id' column
ALTER TABLE customers ADD CONSTRAINT pk_customers
PRIMARY KEY (id);
```

- Foreign key: A foreign key is a column or set of columns in one table that refers to the primary key of another table. It can be defined using the FOREIGN KEY constraint.

Example:

```
-- Create a foreign key constraint on the 'orders' table that references the 'customers' table

ALTER      TABLE      orders      ADD      CONSTRAINT
fk_orders_customers      FOREIGN      KEY      (customer_id)
REFERENCES customers(id);
```

- Index: An index is a database object that improves the speed of data retrieval by providing quick access to specific rows in a table. It can be created using the CREATE INDEX statement.

Example:
```
-- Create an index on the 'name' column of the 'customers' table

CREATE INDEX idx_customers_name ON customers (name);
```

- **Joins**: joins are used to combine data from two or more tables based on a common column or set of columns, and are an important tool for working with relational databases.

Here are some types of joins in SQL:

- Inner join: An inner join returns only the rows that have matching values in both tables being joined.

  Code Example:

  ```
  SELECT *
  FROM table1
  INNER JOIN table2
  ON table1.column = table2.column;
  ```

- Left join: A left join returns all the rows from the left table and the matching rows from the right table. If there are no matching rows in the right table, NULL values are returned.

  Code Example:

  ```
  SELECT *
  FROM table1
  LEFT JOIN table2
  ON table1.column = table2.column;
  ```

- Right join: A right join returns all the rows from the right table and the matching rows from the left table. If there are no matching rows in the left table, NULL values are returned.
  Code Example:

  ```
  SELECT *
  FROM table1
  RIGHT JOIN table2
  ON table1.column = table2.column;
  ```
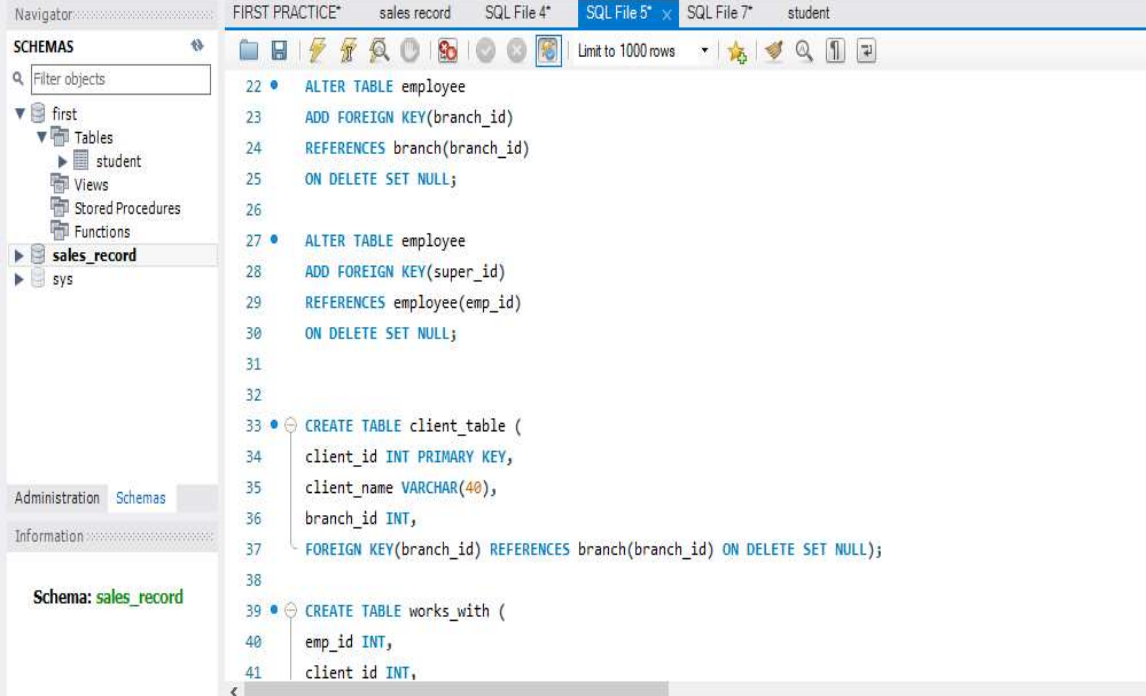
- Full outer join: A full outer join returns all the rows from both tables, along with NULL values for any non-matching rows.

  Code Example:

  ```
  SELECT *
  FROM table1
  FULL OUTER JOIN table2
  ON table1.column = table2.column;
  ```

In summary, SQL is a key tool in data science that enables data scientists to efficiently manage and manipulate large sets of data stored in databases, and perform complex analyses on that data.

3.3.2.2 Snippet of my SQL Code


```
22 •   ALTER TABLE employee
23     ADD FOREIGN KEY(branch_id)
24     REFERENCES branch(branch_id)
25     ON DELETE SET NULL;
26
27 •   ALTER TABLE employee
28     ADD FOREIGN KEY(super_id)
29     REFERENCES employee(emp_id)
30     ON DELETE SET NULL;
31
32
33 •   CREATE TABLE client_table (
34     client_id INT PRIMARY KEY,
35     client_name VARCHAR(40),
36     branch_id INT,
37     FOREIGN KEY(branch_id) REFERENCES branch(branch_id) ON DELETE SET NULL);
38
39 •   CREATE TABLE works_with (
40     emp_id INT,
41     client id INT,
```

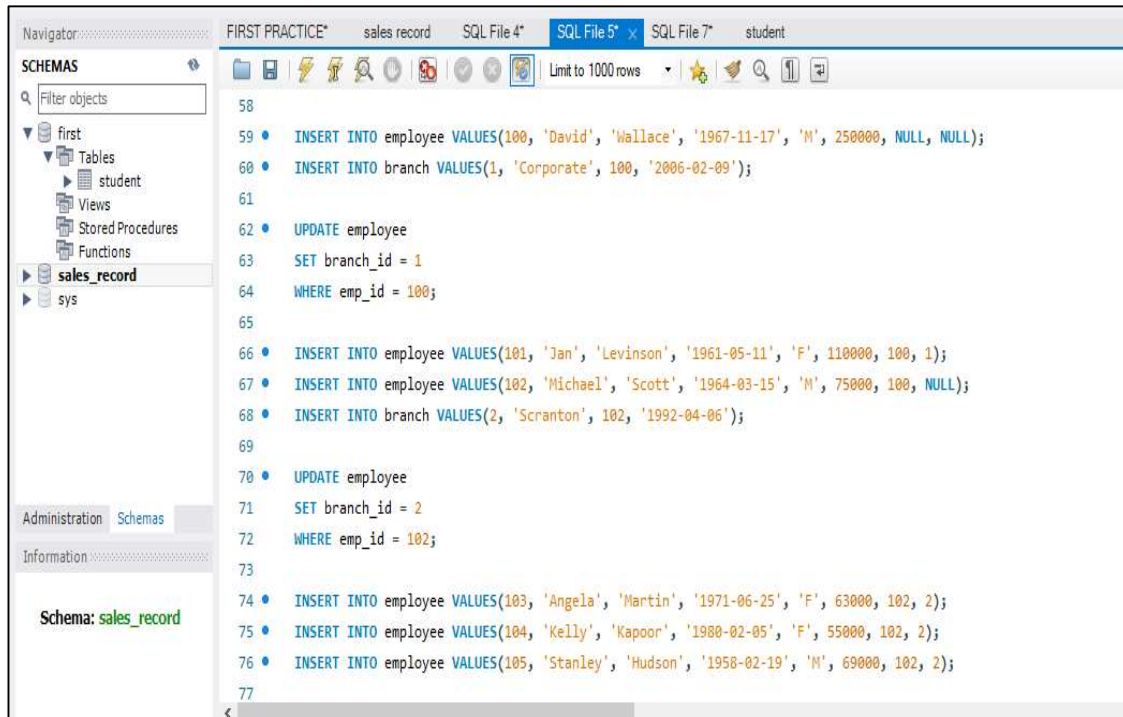Fig 3.7 Creating a table using the create statement on MySQL
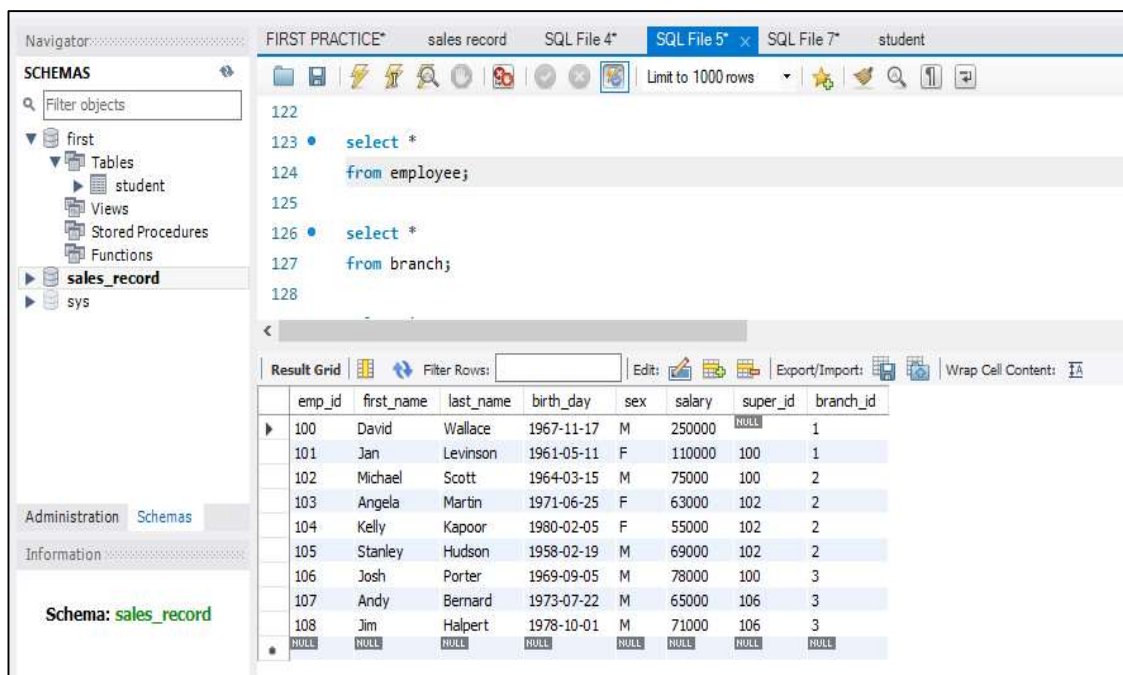
Fig 3.8 Inserting values into my tables on MySQL

Fig 3.9 Using the select all function in other to view the content of my table
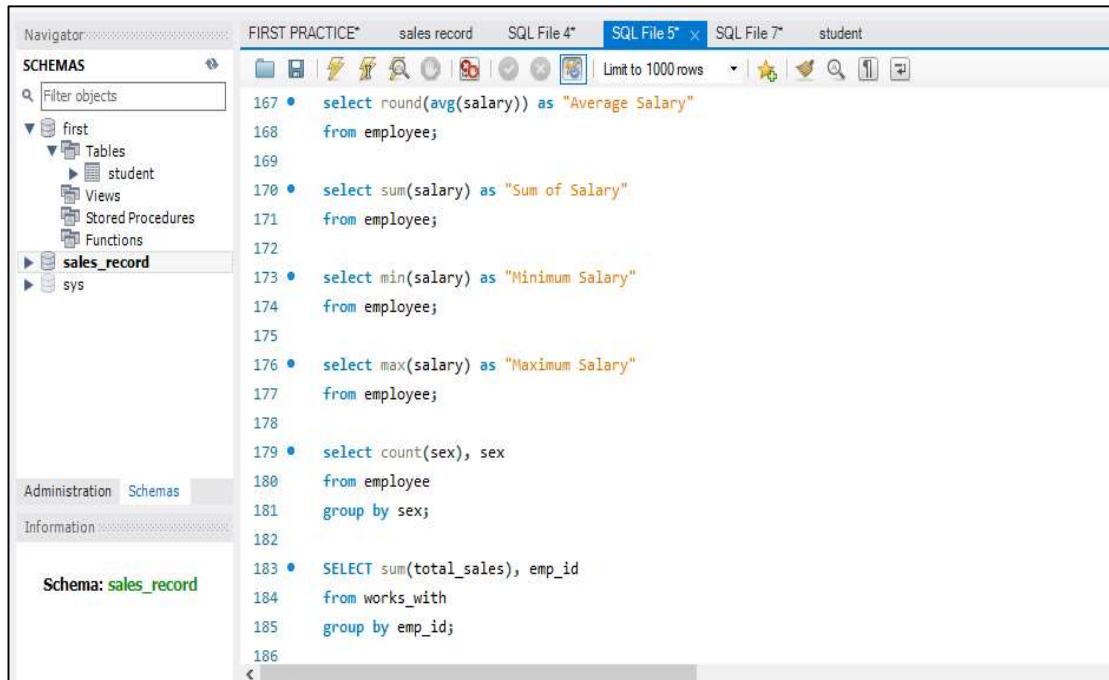
Fig 3.10 Simple arithmetic performance with my tables data.

3.4 APPLICATION USED IN DATA SCIENCE

3.4.1 Ms Excel

Microsoft Excel is a powerful spreadsheet software widely utilized for diverse purposes in business, education, and personal data management. Its primary purpose is to organize, analyze, and visualize data in a tabular format, offering users a versatile tool for various tasks.

Excel serves as a fundamental tool in data science, providing a user-friendly platform for organizing, analyzing, and visualizing data. Its purpose extends to diverse applications in business, education, and personal data management. In data science, Excel is often used for exploratory data analysis, data cleaning, and basic statistical calculations. The software's spreadsheet format enables the creation of structured

datasets, facilitating the initial stages of data preparation. Excel's functions and formulas allow data scientists to perform calculations and manipulate data efficiently.

Excel's functions range from basic arithmetic operations to advanced statistical analysis and complex financial modeling. Its formulae, pivot tables, and charting capabilities make it an indispensable tool for professionals across various industries, offering efficiency and accuracy in data manipulation and representation. Excel's widespread use and adaptability make it a cornerstone in data management and analysis, contributing to its enduring popularity worldwide.

Excel serves as a stepping stone for beginners in data analysis, providing a familiar environment for foundational skills development. Its ubiquity and accessibility make it a valuable tool for quick insights and initial data exploration in the broader context of data science workflows.
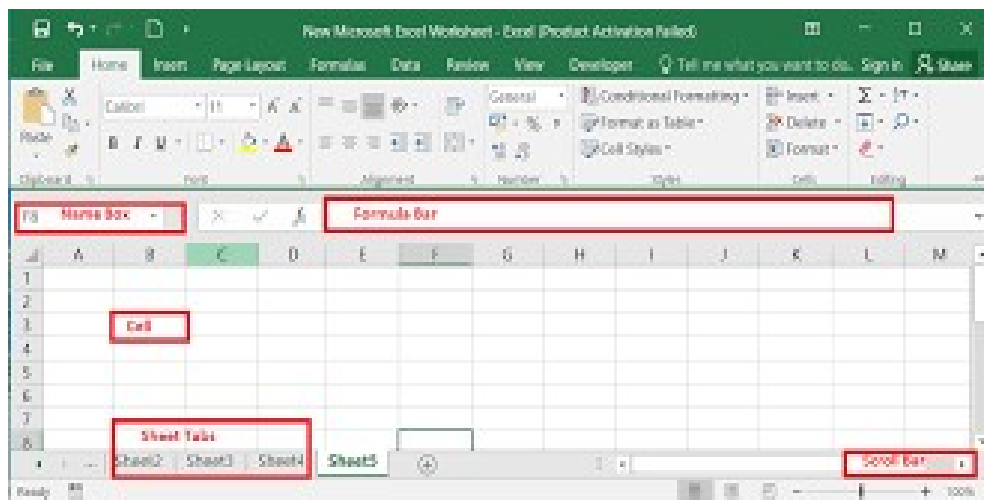

Fig 3.11 Excel workbook view

3.4.2 Power BI Desktop

Power BI, a business analytics service from Microsoft, empowers users to visualize and analyze data from diverse sources. It facilitates the creation of interactive

reports and dashboards, incorporating visualizations like charts and graphs. Integration with Excel, cloud-based, on-premises, and third-party sources enhances its versatility.

Power BI Desktop also includes powerful data modelling capabilities, allowing users to create relationships between tables and perform calculations using DAX formulas. With features encompassing data modeling, visualization, exploration, and collaborative capabilities, users can create, share, and collaborate on interactive dashboards and reports. Available in free and paid editions with advanced features.

Power BI is widely adopted by businesses to extract insights and inform strategic decisions.

Key things to know about in Microsoft Power Bi Desktop Application include:

- Data source: A data source is a location where data is stored, such as a database, Excel file, or web service. In Power BI Desktop, data sources can be connected to and used to create reports and visualizations.
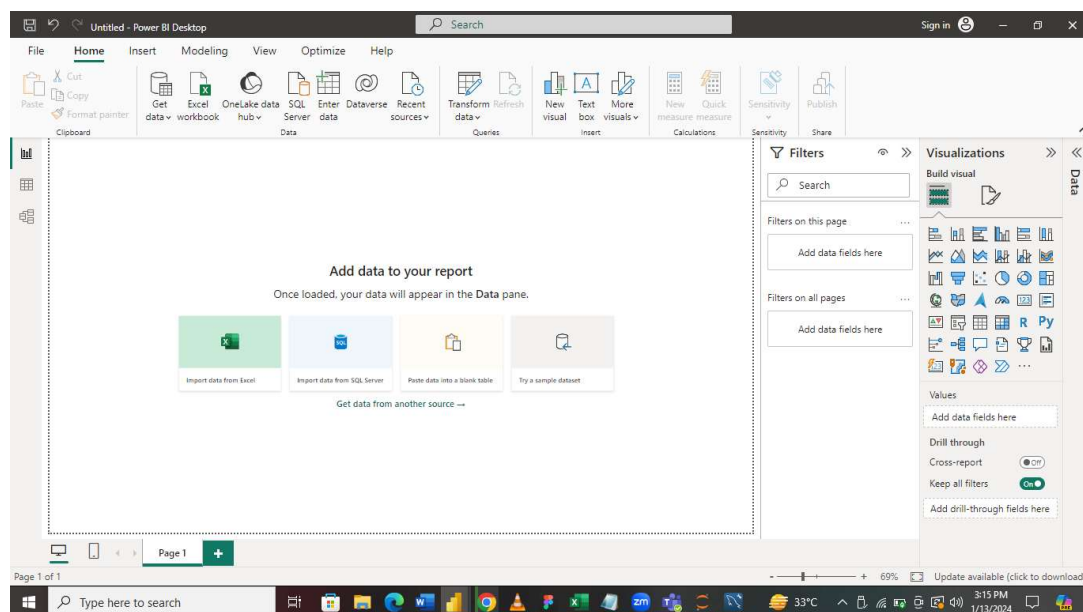

Fig 3.12 Power BI workbook view.

- Query Editor: The Query Editor is a tool in Power BI Desktop that allows users to transform and shape data before it is loaded into a report or visualization. It provides a graphical interface for performing tasks such as filtering, sorting, and grouping data.
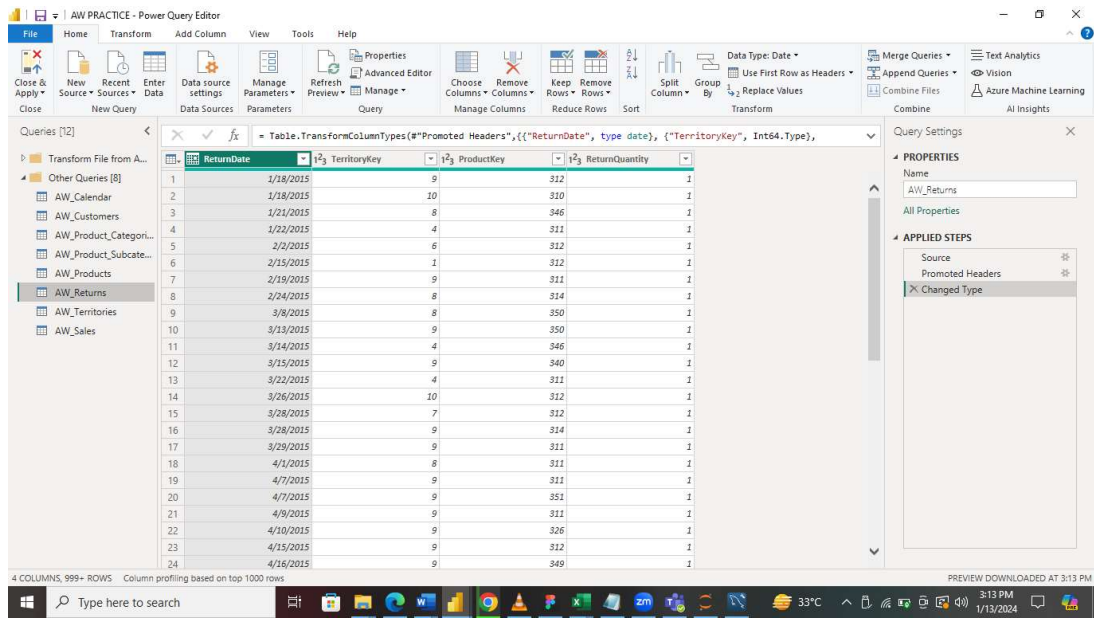
Fig 3.13 Power Query workbook view.

- **Visualizations:** Graphical representations, like charts and tables, showcasing data. Power BI offers built-in visualizations and the ability to create custom ones using Power BI Developer Tools.
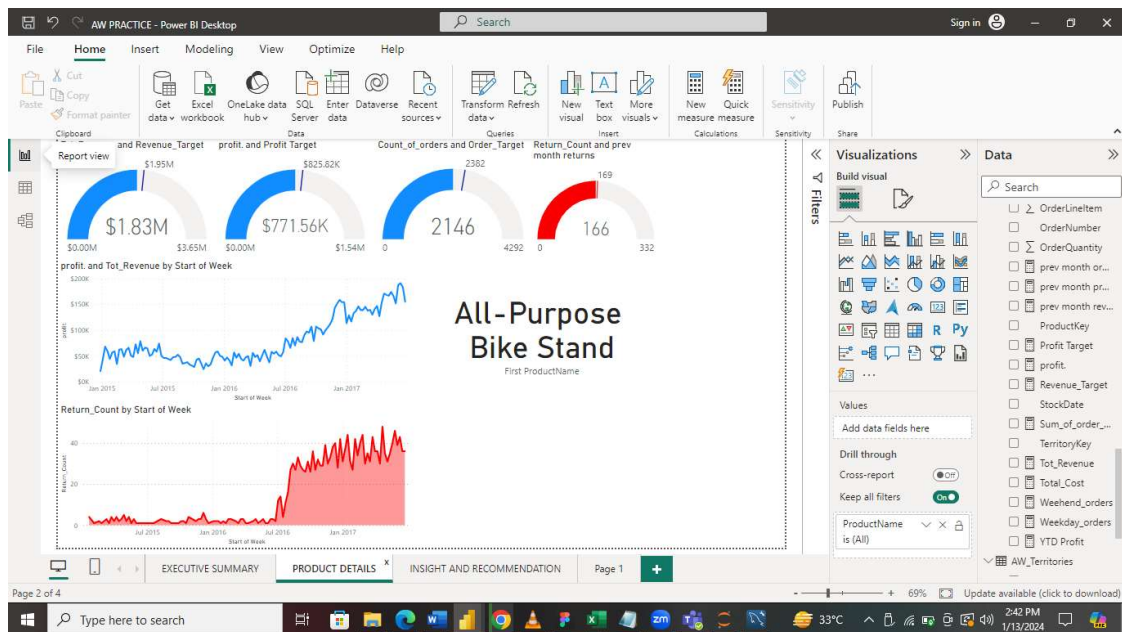

Fig 3.14 Power BI report view showing my Analysis.

- Data Model: Represents relationships between data tables in a database. In Power BI, it establishes table relationships and defines measures and calculated columns.
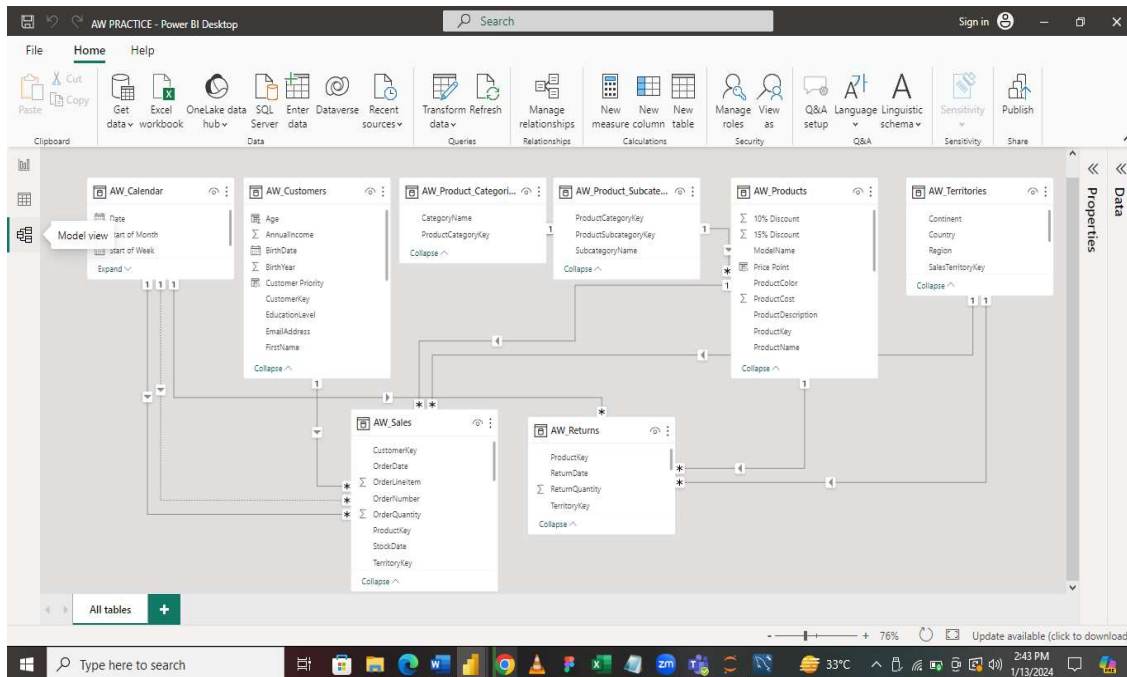
Fig 3.15 Power BI model view showing relationship connection.

3.4.3 MySQL

MySQL is an open-source relational database management system (RDBMS) that is widely used for web applications and other data-driven applications. MySQL is a popular choice for many organizations because it is free to use, easy to install, and can be used on a wide range of operating systems.

MySQL is designed to store and manage structured data in tables, which are organized into rows and columns. Each column in a table is defined with a specific data type, such as integer, string, or date. Rows in a table represent individual records, and each row contains data values for each column in the table.
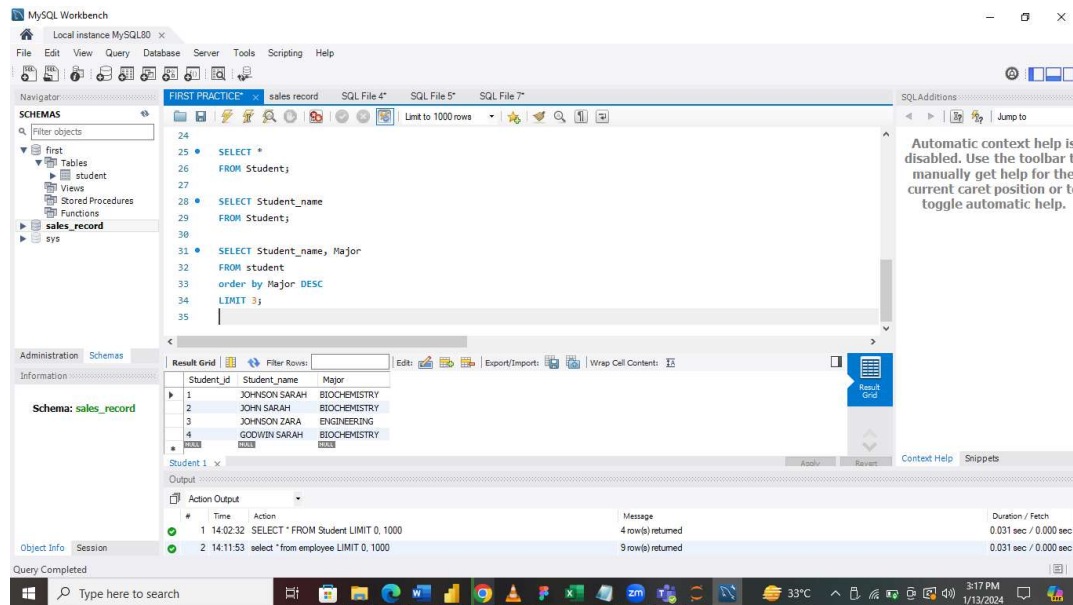
Fig 3.16 MySQL workbook view

MySQL stands out as a robust and versatile relational database management system (RDBMS) extensively employed in web applications, data-driven applications, and various scenarios. Its popularity stems from a combination of user-friendly features, scalability, and its open-source framework. This makes MySQL a favored selection among numerous organizations seeking a reliable database solution.

3.4.4 Anaconda

Anaconda is a popular and free bundle of Python and R programming languages. It's widely used for scientific computing, data science, and machine learning. With over 1,500 tools for tasks like data analysis and machine learning, Anaconda is a handy tool for data scientists and developers. It comes with Conda, its own easy-to-use package manager, allowing users to install, manage, and update packages effortlessly. Anaconda also includes Jupyter Notebook, a web-based tool for interactive coding, equations, and visualizations.

A big advantage of Anaconda is its ability to create a consistent and reproducible environment for data projects, ensuring reliability. It saves time by having many packages pre-installed, making it convenient for new projects. In a nutshell, Anaconda simplifies the setup and management of tools needed for scientific computing and data-related work.
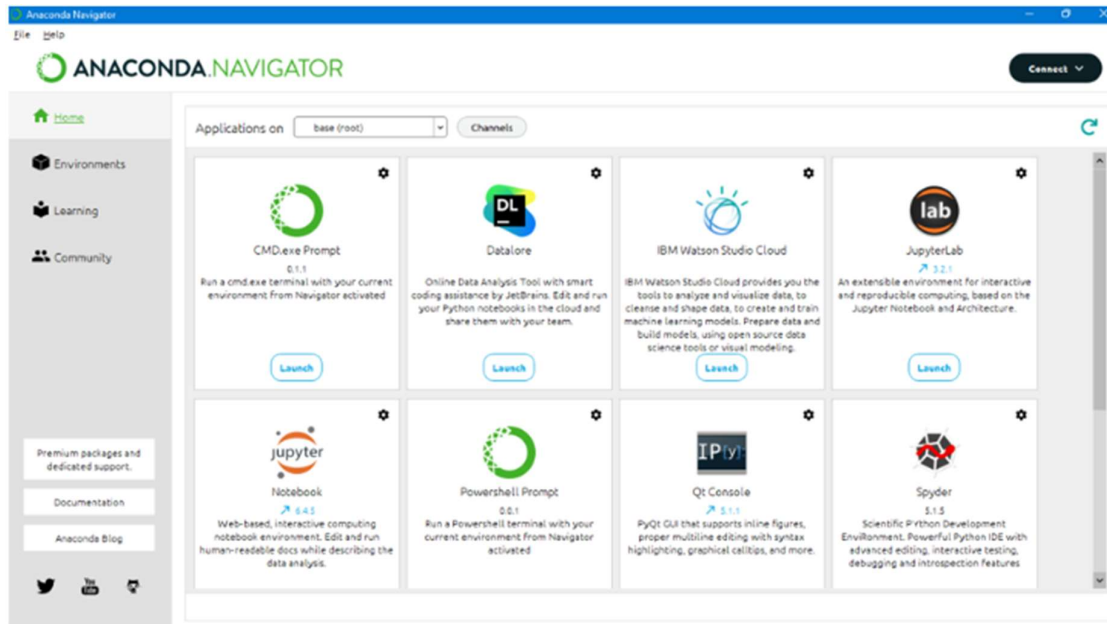

Fig 3.17 Anaconda page view.

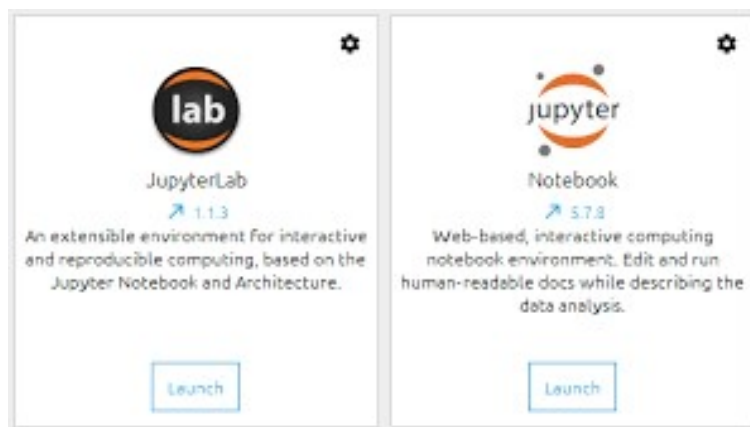### 3.4.4.1 Jupyter Lab And Jupyter Notebook

Fig 3.18 View of Jupyter Lab/Notebook before launch.

JupyterLab and Jupyter Notebook are both included in the Anaconda distribution of Python, which is a popular package manager for data science and machine learning.

When you install Anaconda, it comes with both JupyterLab and Jupyter Notebook pre-installed, along with a range of other popular Python packages and libraries. You can launch JupyterLab or Jupyter Notebook directly from the Anaconda Navigator graphical user interface, or you can launch them from the command line using the "Jupyter lab" or "jupyter notebook" commands.

JupyterLab and Jupyter Notebook both provide an interactive computational environment for data analysis, scientific computing, and machine learning. However, JupyterLab is a more advanced and flexible user interface than Jupyter Notebook, allowing users to work with multiple notebooks, code consoles, text editors, and other tools in a single, integrated environment. Jupyter Notebook, on the other hand, provides a simpler and more focused environment for working with individual notebooks.

CHAPTER FOUR

4.1 Nigeria Economy Growth Data Analysis Project (1999 – 2023)

Nigeria has the greatest economy in Africa and studying Nigeria's economic growth from 1999 to 2023 helps us to:

- Understand Development
- Make Informed Decisions
- Address Challenges
- Plan for the Future.

4.2 Data Collection and Analytics Tools

Data source

kaggle

Analytics tools used

Power BI & Python

4.3 Area Focus

The economic focus areas of this project include;

- Government Debt
- Agriculture and industry
- Unemployment
- Gross domestic product

4.4 Dataset Description

- This dataset contains details about four previous presidents of Nigeria between1999 and 2023.

- It outlines their respective performances across various sectors of Nigeria's economy.

Ex-President Olusegun Obasanjo (2007 - 2010)

Ex-President Musa Yar'adua (2007 - 2010)

Ex-President Goodluck Ebele Johnathan (2010 - 2015)

Ex-President Muhammadu Buhari (2015 - 2023)

4.5 Code Snippets of Python Analysis

This is a short code snippet on the analysis done using **python programming language** on a Jupiter Notebook.

Fig 4.1 Statistical calculation on Nigeria Economic Growth Dataset.

```
In [46]: sns.heatmap(sub_corr, annot=True)

Out[46]: <AxesSubplot:>
```


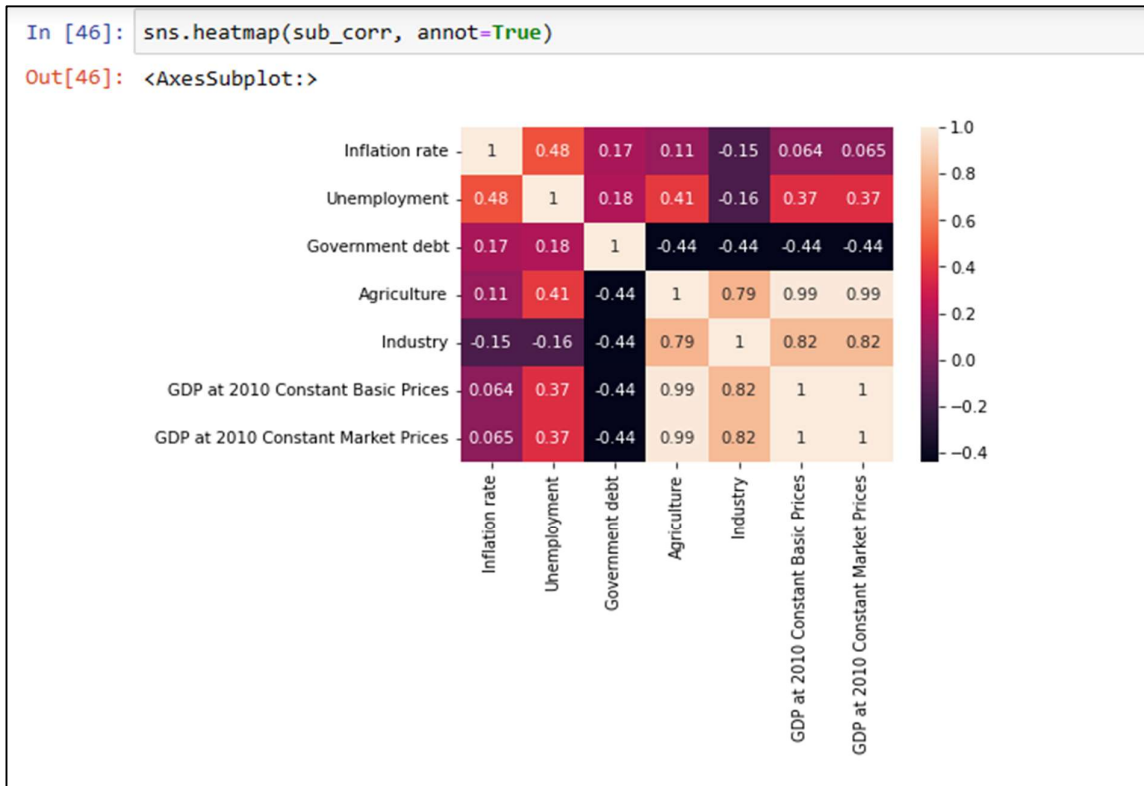|	Inflation rate	Unemployment	Government debt	Agriculture	Industry	GDP at 2010 Constant Basic Prices	GDP at 2010 Constant Market Prices
Inflation rate	1	0.48	0.17	0.11	-0.15	0.064	0.065
Unemployment	0.48	1	0.18	0.41	-0.16	0.37	0.37
Government debt	0.17	0.18	1	-0.44	-0.44	-0.44	-0.44
Agriculture	0.11	0.41	-0.44	1	0.79	0.99	0.99
Industry	-0.15	-0.16	-0.44	0.79	1	0.82	0.82
GDP at 2010 Constant Basic Prices	0.064	0.37	-0.44	0.99	0.82	1	1
GDP at 2010 Constant Market Prices	0.065	0.37	-0.44	0.99	0.82	1	1

Fig 4.2 Using Seaborn Lib. To view the correlation of the Dataset.

4.6 Analysis with Power BI
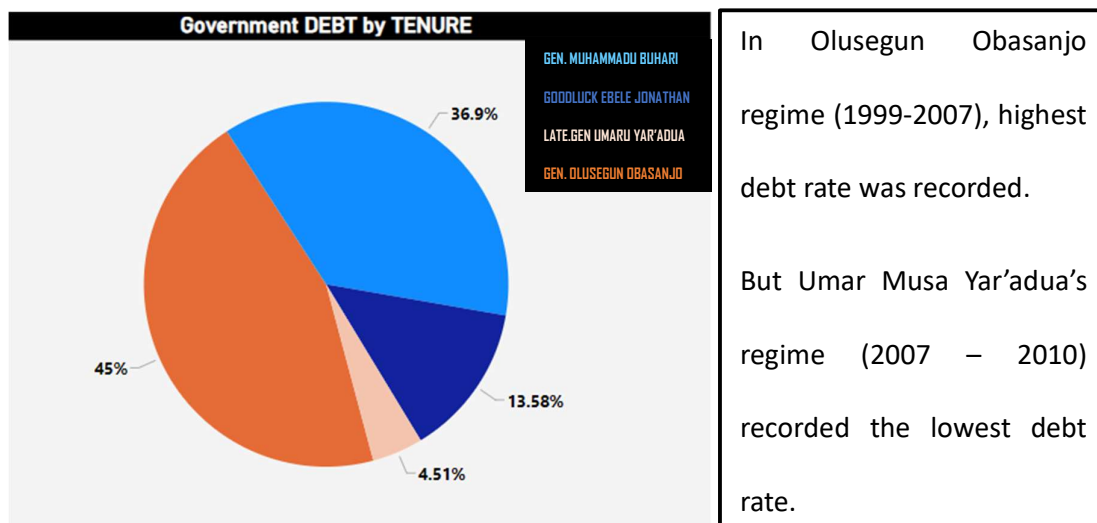
4.6.1 Analysis on Government Debt By Previous Presidents

In Olusegun Obasanjo regime (1999-2007), highest debt rate was recorded.

But Umar Musa Yar'adua's regime (2007 – 2010) recorded the lowest debt rate.

Fig 4.3 Debt Analysis of Nigeria Economic Growth Dataset.

32

4.6.2 **Analysis on Agriculture By Previous Presidents**

Between 1999 and 2022, Nigeria's economy recorded consistent increase in agriculture.

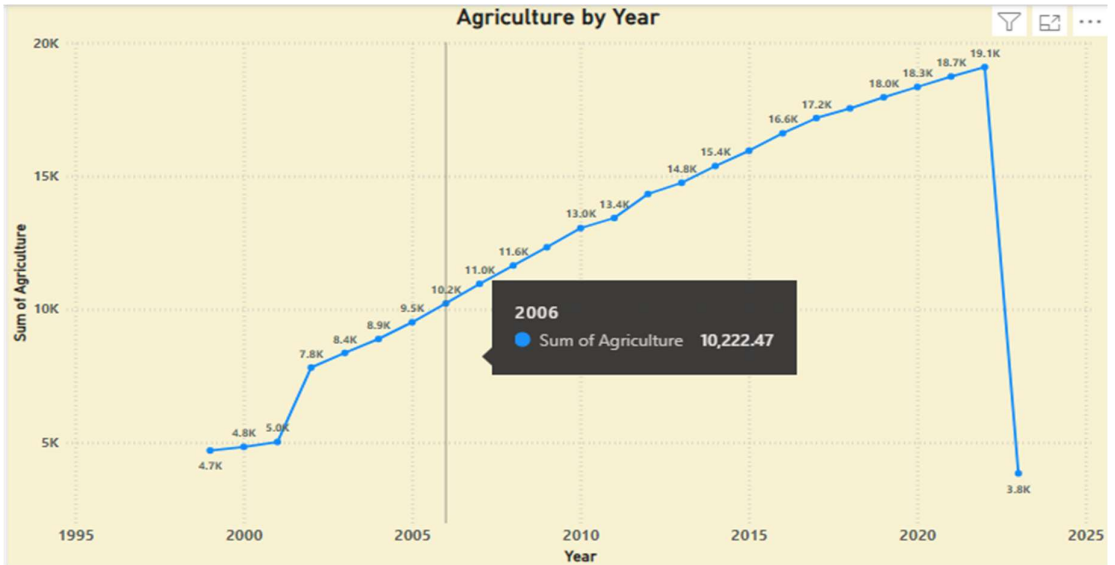Before there was a significant decline between 2022 and 2023.

Fig 4.4 Agriculture Analysis of Nigeria Economic Growth Dataset.

4.6.3 **ANALYSIS ON INDUSTRY BY PREVIOUS PRESIDENTS**

- Between 1999 and 2022, A fluctuating increase in industrial growth was observed.
- There was a significant decline in industrial expansion between 2022 and 2023.
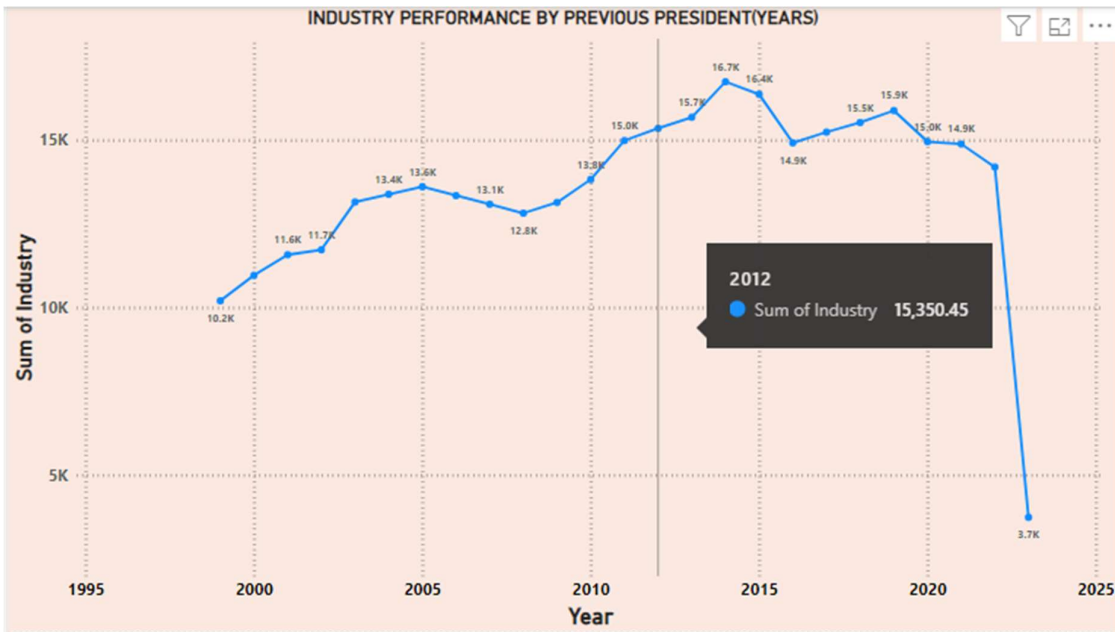

Fig 4.5 Industry Analysis of Nigeria Economic Growth Dataset.

4.6.4 Unemployment Rate Forecast By 2024

- In year 2023, we recorded the highest rate of unemployment (0.41%).
- And as seen in the forecast. By 2024, we're most likely to observe a 6% increase in unemployment rate.
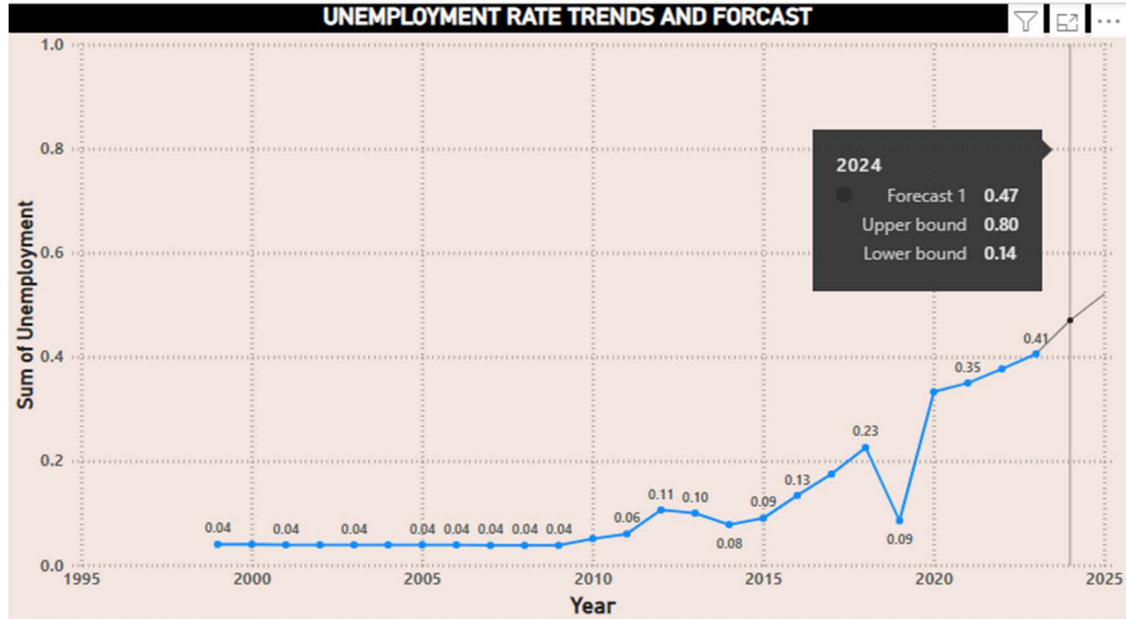

Fig 4.6 Unemployment rate Analysis of Nigeria Economic Growth Dataset.

4.6.5 Analysis on Gross Domestic Product

- There was a slow but steady increase in Gross Domestic Product from the year (1999–2022).
- There was a significant decline in Gross Domestic Product between 2022 and 2023.
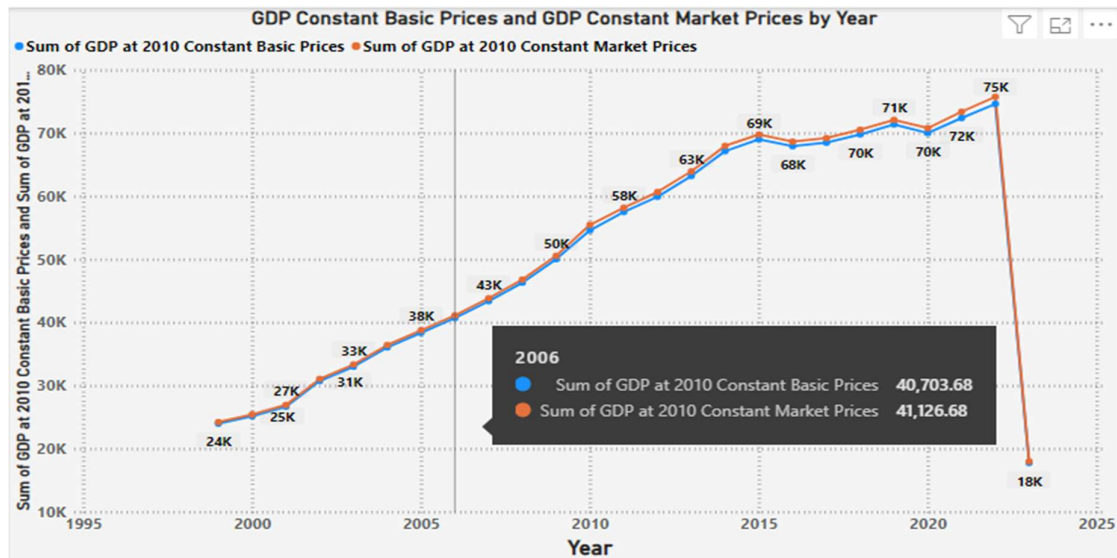

Fig 4.7 GDP Analysis of Nigeria Economic Growth Dataset.

34

4.7 SUMMARY

Focus Areas	Ex-President Olusegun Obasanjo (1999 – 2007)	Ex-President Musa Yar'adua (2007 – 2010)	Ex-President Goodluck Jonathan (2010 – 2015)	Ex-President Muhammadu Buhari (2015 – 2023)
Government Debt (%)	45%	4.5%	13.6%	37%
Agriculture (K)	4.7k – 11k	11k – 13k	13k – 15k	15k – 18k
Industry (K)	10k – 13k	13k – 13.9k	13.9k – 16k	16k – 3k
Gross Domestic Product (K)	24k – 46k	46k – 55k	55k – 69k	69k – 18k

4.8 Strategic Solutions for Nigeria's Economic Challenges

Upon conclusion of the analysis, I was able to suggest possible solutions which include:

- Promoting entrepreneurship.
- Developing initiatives to reduce dependence on a single sector so as to boost our country GDP
- Providing farmers with better access to credit facilities to boost agricultural development
- Offer financial support to industries to increase industrial growth

CHAPTER FIVE

5.1 EXPERIENCE GAINED

I gained valuable experience in various aspects of data analysis/Data Science. Through practical application, I mastered statistical and computational techniques, utilizing tools like Power BI and Python programming languages for extracting insights and knowledge from data. I also gained hands-on experience in data cleaning and preprocessing, data visualization, statistical analysis, and machine learning. Additionally, I was able to work with large and diverse data sets, which helped me to improve my problem-solving, time management and critical thinking skills.

Overall, my experience at Babtech Computers not only readied me for a data science career but also fortified my foundation for future growth. Additionally, it contributed to the refinement of my teaching skills and strategic thinking, presenting a great learning experience.

5.2 CHALLENGES ENCOUNTERED

Prior to the commencement of my SIWES I had difficulties in being accepted in some organizations for Industrial Training due to certain requirements and also some was saying no vacancy for data science SIWES students.

Collaborating and communicating with people in a professional setting also wasn't also as easy as I think it'll be and I as well had issues with transporting myself to my place of work due to Nigerian economy.

CHAPTER SIX

6.1 Conclusion:

Overall, my SIWES experience at Babtech Computers was a great opportunity to gain practical experience in the field of Technology. Through my work which includes PA to the CEO, teaching Data Courses/UI, and providing support to data students, I developed strong communication skills and an ability to work collaboratively with others. Additionally, my project work with analyzing and visualizing data allowed me to gain practical experience with Python, MySQL, Power BI, and Excel.

Beyond gaining industrial experience, I got adapted to the industrial work environment, acquired both working and professional ethics, socialization and also realized that I learnt lot of new things that one would never be able to learn at the university's academic environment.

6.2 Recommendation:

I highly recommend considering Babtech Computers for students looking to gain practical experience in all tech field like software Engineering, Data Science, etc. during their SIWES program. They offer a supportive learning environment, hands-on experience working on diverse projects, and opportunity to teach others, which is also one of the best ways of retentive learning. Working with Babtech Computers will provide you with valuable skills and knowledge that can help you stand out in the job market and prepare you for a career in whatever field you might have choose.

References:

- Grus, Joel. Data Science from Scratch: First Principles with Python. O'Reilly Media, 2019. https://www.oreilly.com/library/view/data-science-from/9781492041122/

- python Data Analytics: Data Analysis and Science Using Pandas, matplotlib, and the Python: Python Data Analytics: Data Analysis and Science Using Pandas, matplotlib, and the Python Programming Language by Fabio Nelli (auth.) - PDF Drive

- SIWES (ITF): https://www.itf.gov.ng/

- Babtech Computers: https://babtechcomputers.com/

- Introduction to My SQL." Lifewire: https://www.pdfdrive.com/learning-mysql-e17258694.html

- Kaggle Dataset: https://www.kaggle.com/datasets/ogbuzurukelechi/nigeria-economy-growth-1990-2023