SUMMER TRAINING REPORT

On

"Data Science and Analytics"

At

ZIDIO
DEVELOPMENT

Spanning

20th July - 20th September

for partial fulfillment of the requirement for the award of

BACHELOR OF TECHNOLOGY

In

COMPUTER SCIENCE AND ENGINEERING

From

School of Studies Engineering and Technology,

Guru Ghasidas Vishwavidyalaya

Submitted By:

Lanka Aravind

Roll No.: 21027136

Batch: 2021 – 2025

Institute: Guru Ghasidas Vishwavidyalaya

Submitted To:

Mr. Vaibhav Kant Singh

Assistant Professor.

Department of CSE, SOSET,

Guru Ghasidas Vishwavidyalaya

ACKNOWLEDGEMENT

I would like to extend my heartfelt gratitude to Zidio Development Company for providing me the opportunity to undertake this internship and work on the Data Science and Analytics project. This experience has been invaluable to my professional and personal growth.

I also want to express my appreciation to the entire team at Zidio Development Companyfor their cooperation and assistance. The collaborative environment and the knowledge shared by them greatly contributed to the learning experience.

Finally, I would like to acknowledge my academic institution, **Guru Ghasidas Vishwavidyalaya, Bilaspur**, for supporting and encouraging me to take on this internship, which has been a significant step in my career development.

Thank you all for your invaluable support and encouragement.

warm regards,
Lanka Aravind

DECLARATION

I **Lanka Aravind** Roll No**. 21027136**, Enrollment No.**GGV/21/01136** a full time student of **B.Tech**. Programme, **Batch (2021- 2025)** of Department of Computer Science and Engineering, School of Studies of Engineering and Technology, **Guru GhasidasVishwavidyalaya, Bilaspur**, hereby certify that this summer training carried out by me at **Zidio Development** and report submitted in partial fulfillment ofthe requirements of the programme is an original work of mine under the guidance of Zidio Development organization.

(Student's Signature)   **(Tacher's Signature)**
Name: Lanka Aravind   **Mr. Vaibhav Kant Singh**
Roll No : 21027136   **Assistant Professor**
Enrollment No : GGV/21/01136

INTERNSHIP OFFER LETTER

Issue Dt : 13-07-2024

ZIDIO
DEVELOPMENT

Zidio Internship
OFFER LETTER

This Internship is offered to

Lanka Aravind

....

on _____ *Data Science & Analytics* _____ internship program

We are excited to offer you an internship at Zidio Development, a rapidly growing product-based startup dedicated to advancing in the age of AI through learning and development. As an intern at Zidio Development, you will receive training in the latest industry-required technologies and be assigned real-world projects. This is a unique opportunity for you to engage with and contribute to solving real-world problems. Your key responsibilities in this internship will include:

1. **Assisting in the development and implementation of the solutions.**
2. **Working with team members to design, code, and test new features.**
3. **Participating is project planning and review meetings.**
4. **Conducting Research and Analysis to support ongoing projects.**

We believe this internship will significantly enhance your professional development, providing hands-on experience with the latest technologies and product development.

Please note that this offer letter is confidential. Unauthorized copying may result in legal action. During your internship, you will have access to the company's clients and confidential information. You agree to keep all company and client information strictly confidential and not share it with anyone outside the company. Additionally, you agree not to use this information for personal gain.

We look forward to onboard and welcome you to Zidio development internship program and explore the latest technologies while working on real world projects.

20-07-2024	20-09-2024	*Rashmitaga Samy*
Internship start date	Internship end date	HR signature

ZIDIO
DEVELOPMENT

for more information mail us **support@zidio.in**

📞 **+918455807965**   🌐 **www.zidio.in**   📍 **Bengaluru,Karnataka,India**

CERTIFIED **ISO 9001:2015** COMPANY

Table of Contents

ABOUT THE COMPANY

Zidio Development is a dynamic company specializing in leveraging data science and analytics to drive business growth and innovation. With a focus on harnessing the power of big data, Zidio develops cutting-edge solutions that optimize operations and enhance decision-making processes. The company operates across various industries, offering tailored analytics services to meet the unique needs of each sector. Their team of experts is dedicated to turning complex data into actionable insights, empowering clients to stay ahead in a competitive market. Zidio Development is committed to delivering impactful results through data-driven strategies.

Product and Service Offerings:

Zidio Development offers a diverse range of products and services designed to meet the needs of businesses across various industries:

1. **Data Analytics Solutions**: Zidio provides comprehensive data analytics services, including data mining, predictive modeling, and statistical analysis, helping businesses uncover hidden patterns and trends to make informed decisions.

2. **Business Intelligence (BI) Tools**: The company develops custom BI tools that allow clients to visualize their data through interactive dashboards and reports, enabling real-time insights and strategic planning.

3. **Machine Learning and AI Solutions**: Zidio specializes in building machine learning models and AI-driven applications that automate processes, enhance customer experiences, and drive innovation.

4. **Data Management Services**: Zidio offers end-to-end data management solutions, including data cleaning, integration, and governance, ensuring that businesses have access to accurate and reliable data.

Training and Development

Zidio Development is committed to the continuous growth of its employees through comprehensive training and development programs. The company offers a variety of learning opportunities, including workshops, online courses, and mentorship programs, designed to enhance technical skills and industry knowledge.

Company Focus

Zidio Development is focused on empowering businesses through data-driven innovation. The company prioritizes leveraging advanced analytics, machine learning, and AI technologies to solve complex challenges and optimize operations. With a client-centric approach, Zidio tailors its solutions to meet the unique needs of each industry, ensuring impactful outcomes. The company is dedicated to staying ahead of technological advancements.

ABOUT THE INTERNSHIP

---

1. Introduction

The Data Science and Analytics internship at Zidio Development offers a unique opportunity for aspiring professionals to gain hands-on experience in the rapidly evolving field of data science.

As an intern, you'll work alongside experienced data scientists and analysts, contributing to real-world projects that involve data analysis, machine learning, and business intelligence.

This internship is designed to enhance your technical skills, providing you with practical knowledge in data manipulation, predictive modeling, and data visualization. You'll have the chance to collaborate with cross-functional teams, apply your learning to solve business challenges, and gain insights into how data-driven decisions shape the future of industries.

Through this program, Zidio aims to cultivate the next generation of data science leaders.

1.1   Internship Objectives

The objectives of the Data Science and Analytics internship at Zidio Development are designed to provide a comprehensive learning experience:

1. **Skill Development**: Enhance technical skills in data analysis, machine learning, and data visualization using tools like Python, R, and SQL.

2. **Practical Application**: Apply theoretical knowledge to real-world projects, gaining hands-on experience in solving business problems through data-driven approaches.

3. **Collaboration**: Work closely with experienced professionals and cross-functional teams to understand the dynamics of data science in a business environment.

4. **Project Ownership**: Take ownership of specific tasks or projects, contributing to the development of actionable insights that impact business decisions.

5. **Professional Growth**: Build a strong foundation for a career in data science by learning industry best practices and gaining exposure to the latest trends and technologies in the field.

2. Software Stack Used

The following are the tech stacks which I used to complete the project during the internship

2.1 Python

Fig : Python

Business analysis using Python makes task easier since Python Programming language has many advantages over any other programming language. It has prominent features like being a high-level programming language (the codes are in human readable form) it is easy to understand and use by any programmer or user. Many libraries and functions for statistical, numerical analysis are available in Python. Moreover, the source code is freely available to anyone (free and open source).

Libraries of Python:

1. **Numpy:**

   o Definition: NumPy is a Python library that provides support for large, multi-dimensional arrays and matrices, along with a collection of mathematical functions to operate on these arrays.

   o Key Features:
       ▪ Efficient numerical computations.
       ▪ Core data structure: N-dimensional array (ndarray).
       ▪ Broadcasting for element-wise operations.
       ▪ Linear algebra, Fourier transforms, and more.

   o Use Case: NumPy is the backbone for scientific computing and data manipulation.

2. Pandas:

o Definition: Pandas is a powerful data manipulation library for Python. It provides data structures (such as DataFrames and Series) and tools for cleaning, transforming, and analyzing data.

o Key Features:

- Tabular data representation (like spreadsheets).

- Data cleaning, filtering, and aggregation.

- Integration with databases and CSV files.

- Time series analysis.

o Use Case: Pandas simplifies data wrangling and exploration.

3. Matplotlib:

o Definition: Matplotlib is a versatile data visualization library built on NumPy arrays. It allows you to create various types of plots, including line charts, scatter plots, histograms, and more.

o Key Features:

- Customizable plots and charts.

- Supports 2D and 3D visualizations.

- Publication-quality graphics.

o Use Case: Matplotlib helps you convey insights through visual representations.

Visualization Tools

Tableau and Power BI: These are powerful business intelligence tools used for creating interactive dashboards and visualizations. Tableau and Power BI enable you to transform complex data into visually appealing and easy-to-understand reports, allowing stakeholders to explore data trends, patterns, and insights. They are essential for making data-driven decisions, as they provide a user-friendly interface to analyze and present data in real-time, supporting strategic business planning and performance tracking.

MAIN PHASES IN DATA SCIENCE ANDDATA ANALYTICS

The main phases in Data Science and Data Analysis typically involve the following steps:

1. **Problem Definition**: Clearly defining the business problem or research question that needs to be addressed. This phase involves understanding the objectives, identifying the key stakeholders, and outlining the expected outcomes.

2. **Data Collection**: Gathering relevant data from various sources, such as databases, APIs, surveys, or web scraping. This phase is crucial as the quality and relevance of the data directly impact the analysis.

3. **Data Cleaning and Preprocessing**: Preparing the data for analysis by handling missing values, removing duplicates, normalizing data, and converting data into the appropriate format. This phase ensures the data is accurate, consistent, and usable.

4. **Exploratory Data Analysis (EDA)**: Analyzing the data to uncover patterns, correlations, and trends. EDA involves the use of statistical methods and visualizations to gain insights and understand the data's underlying structure.

5. **Modeling**: Developing predictive or descriptive models using machine learning algorithms or statistical techniques. This phase involves selecting the appropriate model, training it on the data, and validating its performance.

6. **Evaluation**: Assessing the performance of the model using metrics such as accuracy, precision, recall, and F1-score. This phase ensures that the model meets the desired objectives and can generalize well to new data.

7. **Interpretation and Communication**: Translating the findings into actionable insights and communicating the results to stakeholders through reports, presentations, or dashboards. This phase is crucial for decision-making and ensuring that the analysis leads to practical outcomes.

8. **Deployment and Monitoring**: Implementing the model or analysis into a production environment where it can be used in real-time decision-making. Continuous monitoring is essential to ensure the model remains effective and accurate over time.

9. **Iteration and Improvement**: Based on feedback and new data, revisiting and refining the model or analysis to improve its performance or adapt to changing conditions. This phase is part of a continuous improvement process.

DATA SCIENCE AND ANALYTICS USING PYTHON

Why Use Python?

Python is a high-level, interpreted programming language known for its versatility and ease of use. It supports various programming paradigms, including procedural and object-oriented programming, making it suitable for a wide range of applications. Python excels in statistical computing due to its extensive libraries and is accessible to those with minimal programming experience. Key features include:

- **Open Source and Free**: Python is freely available with a large community.
- **Interpreted Language**: Code is executed line-by-line, simplifying debugging.
- **Dynamic Typing**: Variables do not require explicit type definitions.
- **Portable**: Code can run on different operating systems with minimal adjustments.
- **Numerous IDEs**: Includes Jupyter Notebook, PyCharm, and VS Code.

Packages Used:

- **Numpy**: For numerical operations and handling multi-dimensional arrays.
- **Pandas**: For data manipulation and analysis with data frames.
- **Seaborn**: For creating attractive and informative statistical graphics.

Dataset Used:

- **Diwali Sales Dataset**: Contains sales data related to Diwali, including product, customer, and performance attributes.

Working with Dataset:

The main process involves importing libraries, loading and preparing data analyzing and visualizing.

CHALLENGES FACED

Over the course of my internship, I encountered a diverse range of challenges, each offering unique insights and opportunities for growth. Some of the notable challenges I confronted included:

1. **Data Quality Issues**: Handling missing values, inconsistencies, and errors in the dataset can impact the accuracy of analysis.

2. **Data Integration**: Combining data from multiple sources or formats can be complex and require significant preprocessing.

3. **Scalability**: Managing and processing large datasets efficiently can be challenging, especially with limited computational resources.

4. **Model Complexity**: Selecting and tuning appropriate models can be difficult, requiring expertise in machine learning and statistical methods.

5. **Interpretation and Communication**: Effectively translating complex data findings into actionable insights and clear reports for stakeholders can be challenging.

6. **Data Security and Privacy**: Ensuring that sensitive data is protected and complies with privacy regulations requires robust security measures and careful handling.

7. **Algorithm Bias**: Identifying and mitigating biases in algorithms and models is crucial to ensure fair and accurate results.

8. **Data Volume and Variety**: Managing the sheer volume and diverse types of data can be overwhelming, requiring sophisticated tools and techniques to handle effectively.

LEARNING AND OUTCOMES

My internship proved to be invaluable, providing me with extensive hands-on experience while collaborating on a real-world project with my team. This practical exposure has significantly enhanced my project development skills and allowed me to apply my newfound knowledge effectively. Among the key takeaways from this experience are:

1. **Technical Skills Enhancement**: Gained proficiency in using Python libraries such as Pandas, Numpy, and Seaborn for data analysis and visualization. Improved ability to handle data preprocessing, transformation, and modeling.

2. **Data Cleaning Techniques**: Developed skills in identifying and addressing data quality issues, such as missing values and inconsistencies, ensuring the dataset is accurate and reliable.

3. **Exploratory Data Analysis (EDA)**: Learned to perform EDA to uncover patterns and insights, utilizing statistical methods and visualizations to understand data better.

4. **Model Building and Evaluation**: Acquired experience in building and evaluating predictive models, including selecting appropriate algorithms and assessing their performance using various metrics.

5. **Data Visualization**: Enhanced ability to create effective visualizations using tools like Matplotlib and Seaborn, helping to communicate data insights clearly to stakeholders.

6. **Problem-Solving Skills**: Improved problem-solving abilities by tackling real-world data challenges, developing solutions, and making data-driven decisions.

7. **Communication of Findings**: Gained experience in presenting complex data insights in a clear and actionable format, creating reports and presentations that effectively convey findings to non-technical audiences.

PROJECT RESULT

TASK 1 : Created A visual Dashboard of employees and their complete info working in the organisation.
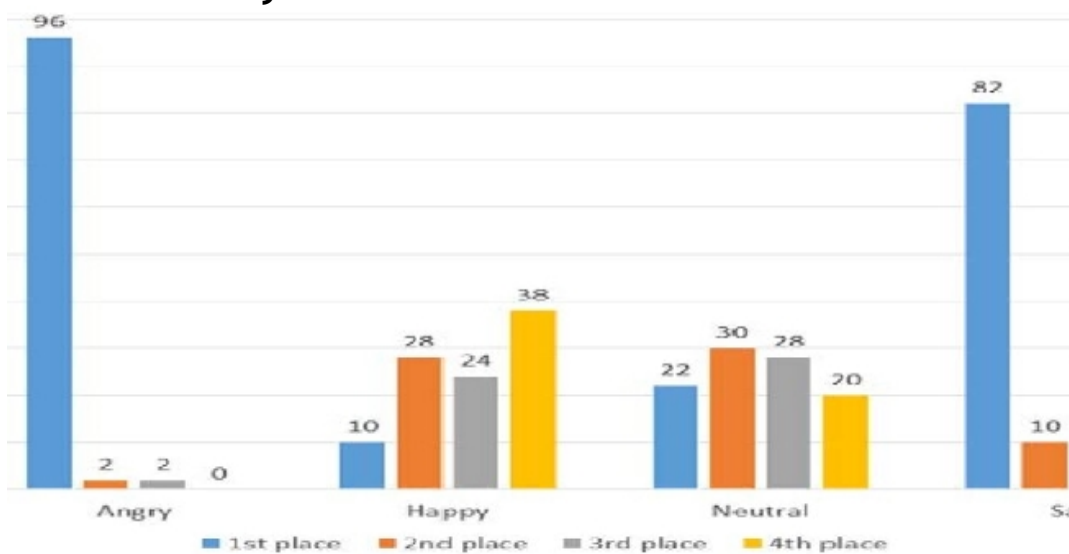Result :

Tools : Tableau, M s-Excel.
Git hub Link :
https://github.com/Aravind1727/HRDasboardCraetion.git

Task 2 : By using Toronto Dataset worked on Speech emotion recognition Project.

Analysis of different emotions

Legend: 1st place, 2nd place, 3rd place, 4th place

Emotion	1st place	2nd place	3rd place	4th place
Angry	96	2	2	0
Happy	10	28	24	38
Neutral	22	30	28	20
Sa...	82	10		

TASK 3: Worked on Creating a visual representation of Famous Ecommerce Site BLINKIT on various analysis.

Tools : Tableau, SQL, Data.

Githu Link : https://github.com/Aravind1727/Data-Analytics Dashboard-Creation.git

CONCLUSION

My internship at Zidio Development was an invaluable experience that significantly contributed to my growth as a data analyst. The challenges I encountered, from data cleaning and handling large datasets to selecting the right visualization techniques, were pivotal in shaping my problem-solving abilities and technical skills. Working with real-world datasets provided practical insights into the complexities of data analytics, bridging the gap between academic knowledge and industry application.

Throughout the internship, I enhanced my proficiency in Data Science and data analysis and essential python libraries such as NumPy, Pandas, Seaborn, and Matplotlib. The experience of communicating analytical insights to non-technical stakeholders taught me the importance of clarity and simplicity in reporting. Additionally, balancing multiple tasks and deadlines helped me develop strong time management and project planning skills, while collaboration with team members improved my ability to work effectively in a team setting.

Zidio Development provided a supportive environment that encouraged learning and experimentation, enabling me to gain hands-on experience with the latest tools and techniques in the field. The skills and knowledge I acquired during this internship have laid a solid foundation for my future endeavors in data analytics,and I am confident that the experience I gained at Zidio Development will be instrumental in advancing my career.