

An Internship Report

On

**“Data Analytics”**

Batch Details

|  |  |  |
| --- | --- | --- |
| Sl. No. | Roll Number | Student Name |
| 1 | 20211CSE0648 | NAGARATNA |

**School of Computer Science,**

**Presidency University, Bengaluru.**

Under the guidance of,

Ms. Shweta Singh

School of Computer Science,

Presidency University, Bengaluru

**CONTENTS**

1. Abstract
2. Review of literature
3. Objectives
4. Problem identification and formulation of problem statement.
5. Road map, impact and advantages of the proposed work.
6. (Time Line by Gantt chart).
7. References

## **ABSTRACT**

Data analytics enables organizations to extract meaningful insights from raw data, improving decision-making and strategic planning. It involves data mining, cleaning, transformation, statistical analysis, and visualization to enhance operational efficiency and detect trends.

This internship at UptoSkills focused on **data mining using Excel**, followed by **data analysis** to organize and analyze structured datasets. The project involved data extraction, cleaning, and transformation to ensure accuracy and usability. Exploratory analysis was conducted to identify patterns, followed by visualization techniques such as charts and pivot tables for better interpretation.

The report provides an overview of the literature review, problem identification, methodologies, and project roadmap. It highlights the effectiveness of Excel for data mining while evaluating its scalability and automation compared to advanced analytics tools.

**2. LITERATURE REVIEW**

.

|  |  |  |
| --- | --- | --- |
| Method | Advantages | Limitations |
| Excel for Data Mining | Easy to use, widely available, supports basic analysis | Limited for large datasets, lacks automation |
| SQL-based Data Mining | Efficient for structured data handling | Requires SQL knowledge, not ideal for unstructured data |
| Python (Pandas, NumPy) | Flexible, automates workflows | Requires programming knowledge |
| R for Data Analytics | Powerful for statistical analysis | Steep learning curve |
| Power BI/Tableau | Excellent visualization tools | Limited data transformation features |
| Google Sheets | Cloud-based collaboration | Slower for large datasets |
| Apache Spark for Big Data | Handles massive datasets | Requires technical expertise |
| Machine Learning for Data Mining | Automates pattern recognition | High computational requirements |

## **OBJECTIVES**

* **Efficient Data Mining** – Extract, clean, and structure data using Excel to ensure high data quality.
* **Data Analysis & Insights** – Apply statistical methods to identify trends and patterns in the dataset.
* **Visualization & Reporting** – Utilize graphs, pivot tables, and dashboards for effective data representation.
* **Evaluation of Methods** – Compare Excel-based analysis with advanced tools like Python or Power BI to assess efficiency and scalability.

### **PROBLEM IDENTIFICATION AND FORMULATION OF PROBLEM STATEMENT**

#### ****Problem Identification****

In the competitive job market, effective communication between **Training and Placement Officers (TPOs) of colleges** and **corporate companies** is crucial for successful campus placements. However, there is a lack of a centralized and structured database that contains verified information about TPOs across different colleges in India and corporate company details such as **contact numbers, job vacancies, contact persons, and email addresses**.

Currently, organizations and colleges struggle with **unstructured data, missing or outdated contact details, and inefficient ways of accessing relevant corporate hiring information**. While advanced customer relationship management (CRM) tools exist, many institutions and recruiters still rely on Excel for organizing and managing placement-related data due to its simplicity and accessibility.

#### ****Problem Statement****

"Develop an efficient methodology for collecting, organizing, and analyzing TPO and corporate hiring data using Excel to create a structured database that improves accessibility, ensures data accuracy, and enhances communication between colleges and recruiters while addressing scalability and automation challenges."

### **ROAD MAP (METHODOLOGY)**

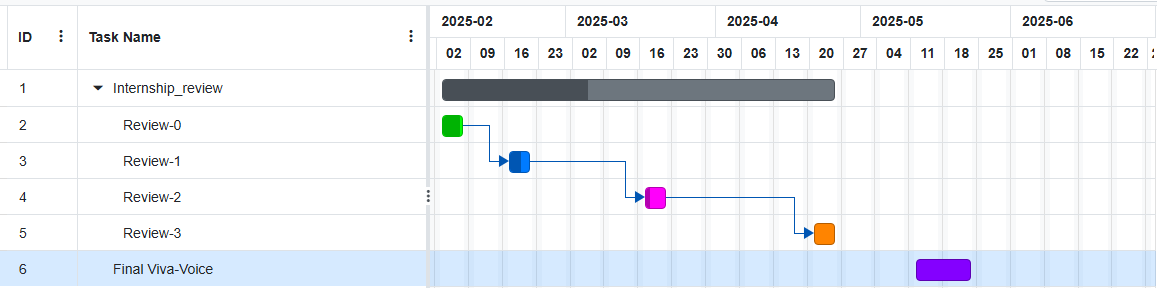
1. **Data Collection** – Importing raw data into Excel.
2. **Data Cleaning & Preprocessing** – Handling missing values and duplicates.
3. **Data Transformation** – Using formulas, pivot tables, and Power Query.
4. **Exploratory Data Analysis (EDA)** – Applying statistical methods.
5. **Visualization & Reporting** – Generating graphs and charts.
6. **Comparison with Advanced Tools –** Evaluating Python/Power BI for insights.

### **IMPACT OF THE PROPOSED WORK**

* **Enhanced Data Analysis Efficiency** – The proposed work significantly improves the ability to process and analyze data using Excel. With built-in functions and tools like PivotTables, data validation, and conditional formatting, users can extract meaningful insights quickly and efficiently.
* **Improved Decision-Making** – By structuring raw data into understandable formats such as tables, charts, and dashboards, the work facilitates better decision-making. Businesses and individuals can identify trends, detect anomalies, and make data-driven choices with ease.
* **Minimized Dependence on Advanced Programming** – Many data analysis tasks that typically require coding in Python, R, or SQL can be accomplished using Excel’s features. This reduces the need for specialized programming skills, making analytics more accessible to a broader audience.
* **Scalability for Small to Medium Businesses** – Excel provides an efficient solution for small and medium enterprises (SMEs) that may not have the resources to invest in sophisticated data analytics software. This enables them to analyze business performance, manage finances, and optimize operations with minimal investment.
* **Enhanced Collaboration and Reporting** – With Excel’s integration capabilities (such as Power Query, Power Pivot, and cloud-based collaboration through OneDrive and SharePoint), teams can work together efficiently, share real-time insights, and generate automated reports.

### **ADVANTAGES**

* **User-Friendly** – Excel’s intuitive interface allows users with little to no programming knowledge to perform complex data analysis. Functions like drag-and-drop, formula-based calculations, and graphical data representation make it accessible to a wider audience.
* **Quick Implementation** – With pre-built functions and templates, users can analyze data faster than developing custom scripts in programming languages. Features like built-in statistical analysis, conditional formatting, and lookup functions (e.g., VLOOKUP, XLOOKUP) enable swift insights.
* **Cost-Effective** – Since Excel is widely available and often included in Microsoft Office packages, no additional investment in software or training is required. This makes it an economical choice for individuals and businesses looking to leverage data analytics without significant financial overhead.
* **Versatile and Customizable** – Excel offers extensive customization options, allowing users to tailor their analysis with formulas, macros, and automation tools like VBA (Visual Basic for Applications). It also supports data import/export from various sources, making it a flexible solution for multiple use cases.
* **Integration with Other Tools** – Excel seamlessly integrates with databases, cloud platforms, and third-party applications like Power BI, Google Sheets, and SQL databases. This expands its functionality and allows for more advanced analytics when needed.

**TIMELINE (GANTT CHART)**

## **REFERENCES**

[1] Han, J., Kamber, M., & Pei, J. (2011). Data Mining: Concepts and Techniques.  
[2] Provost, F., & Fawcett, T. (2013). Data Science for Business.  
[3] McKinney, W. (2017). Python for Data Analysis.