# WEEK 3 REPORT

## Topic: Exploratory Data Analysis (EDA) – Understanding & Analyzing the Dataset

**Intern Name:** Harika Vavilapalli
**Internship Domain:** Data Analytics / Data Science
**Tool Used:** VS Code (for analysis)
**Focus:** Exploratory Data Analysis, Data Understanding, Finding Patterns & Insights

# 1. Introduction

In the third week of my internship, I worked on **Exploratory Data Analysis (EDA)**. This week was mainly focused on understanding the dataset clearly before moving into any advanced steps like live data integration or further processing.

EDA helped me explore what the dataset contains, how it is structured, and whether it has any issues such as missing values, duplicates, or inconsistent formats. Instead of randomly checking things, I followed a structured EDA approach shared by my mentor. This method made the analysis more organized and easier to understand.

Overall, Week 3 was a very important week because it helped me build confidence in analyzing real-world datasets and extracting meaningful insights from them.

# 2. Objectives of Week 3

The main objectives of Week 3 were:

- To understand the dataset structure and features clearly
- To check the volume and quality of the data
- To analyze the distribution of values in the dataset
- To compare different categories/values to identify patterns
- To study relationships between columns for better understanding
- To prepare the dataset for the next stage of the project

# 3. Work Done / Tasks Completed

### 3.1 Understanding Power BI Basics

Even though this week was focused on EDA using code, I understood how this step fits into the overall workflow. Power BI can be used later for dashboards and reporting, but EDA is necessary before that because it tells us what the data actually contains and what insights we can highlight.

This week helped me understand that analysis becomes much easier when we explore the dataset properly first.

### 3.2 Importing Data into Power BI

This week, I did not work directly on importing data into Power BI or building any reports. The focus was completely on understanding the dataset through EDA. However, I understood that once EDA is completed, the insights can later be used for reports and dashboards if required.

### 3.3 Data Cleaning Using Power Query

In Week 3, I did not perform cleaning using Power Query because most cleaning was already handled earlier through automation in VS Code. However, during EDA, I still verified the dataset quality by checking if:

- missing values are handled properly
- duplicates are removed
- data types are correct
- values are consistent and meaningful

This helped ensure the dataset is reliable for analysis.

## EDA Method Followed

For EDA, I followed the structured approach shared by my mentor. This method helped me explore the dataset in a clear and step-by-step way.

### 1. Volume / Structure of the Data

In this step, I focused on understanding the overall size and structure of the dataset. This included checking:

- number of rows and columns
- column names and what they represent
- data types (numeric, categorical, date, etc.)
- missing values and duplicates
- general data quality

This gave me a clear idea of what kind of dataset I'm working with and what issues might affect analysis.

### 2. Distribution of the Data

After understanding the structure, I explored how the data values are spread across different columns.
This included:

- checking how frequently certain values occur
- understanding whether numeric values are evenly distributed or skewed
- identifying unusual values or outliers
- checking whether the data looks balanced or not

This step helped me understand the behavior of the dataset and the patterns in the values.

### 3. Comparison

In this stage, I compared different parts of the dataset to find meaningful differences. For example:

- comparing values across categories
- checking differences between groups
- identifying which category has higher or lower values

This step helped me understand variations in the dataset and which factors stand out.

### 4. Relationships

Finally, I explored relationships between different columns to understand how they are connected.
This included:

- identifying correlation between numeric features
- checking how one column may influence another
- understanding connections between multiple features

This step gave me deeper insights into the dataset and helped me prepare for the next steps of the project.

### 4. Key Learnings

By the end of Week 3, I learned:

- EDA is one of the most important steps in a data project
- A structured method makes analysis clearer and easier
- Checking volume and structure helps understand the dataset quickly
- Distribution analysis helps identify outliers and common patterns
- Comparison helps find differences between categories or groups
- Relationship analysis helps understand how features are connected
- Good EDA helps in better decision-making for future steps

**5. Challenges Faced**

Some challenges I faced during Week 3 were:

- Understanding what each column represents and how it impacts analysis
- Selecting which insights are most meaningful and relevant
- Identifying inconsistent values and verifying whether they are valid
- Making sure the dataset is ready for future stages without missing important details

However, following the mentor's step-by-step EDA method helped me stay organized and complete the analysis smoothly.

**6. Outcome of Week 3**

At the end of Week 3, I was able to:

✅ Understand the dataset structure clearly
✅ Perform EDA using a structured mentor-guided method
✅ Analyze volume, distribution, comparisons, and relationships
✅ Identify important patterns and insights from the dataset
✅ Ensure the dataset is ready for the next stage (Live Data)

**7. Plan for Week 4**

In Week 4, I will focus on **Live Data**, including:

- Understanding how live/real-time data works
- Exploring how data can be updated continuously
- Learning how APIs help in fetching real-time data
- Understanding automation pipelines for handling live data

**Conclusion**

Week 3 was focused on **Exploratory Data Analysis (EDA)**, where I explored the dataset in detail using a structured method shared by my mentor. I analyzed the dataset's structure, distribution, comparisons, and relationships to gain meaningful insights. This week helped me understand the dataset deeply and prepared me for the next stage of the internship, which is working with live data.