WEEK 2 REPORT

Topic: VS Code – Environment Setup & Data Cleaning Automation

Intern Name: Harika Vavilapalli
Internship Domain: Data Analytics / Data Science
Tool Used: Visual Studio Code (VS Code)
Focus: IDE Setup, Writing Scripts, Automating Data Cleaning, Running Code

1. Introduction

In the second week of my internship, I worked with **Visual Studio Code (VS Code)** and learned how it can be used for practical data-related tasks. This week was focused on moving from tool-based work (Power BI) to a more coding-based approach, where I started using scripts to handle and clean data.

One of the main things I learned is that when datasets are large or keep changing, manual cleaning becomes difficult. So, instead of repeating the same steps again and again, we can automate the data cleaning process using code. This week helped me understand how automation makes the work faster, more accurate, and reusable.

Overall, Week 2 was about building coding confidence and learning how to use VS Code for data cleaning automation.

2. Objectives of Week 2

The main objectives of Week 2 were:

- To understand and work with **VS Code** as a development environment
- To learn how to write and execute scripts inside VS Code
- To perform **data cleaning using code** instead of manual steps
- To automate repeated cleaning operations for efficiency
- To understand how automation helps in real-time and real-world datasets
- To prepare for Exploratory Data Analysis (EDA) in the next week

3. Work Done / Tasks Completed

3.1 Understanding Power BI Basics

Even though this week was focused on VS Code, I connected the concepts from Week 1 and understood that Power BI is useful for cleaning data visually, while VS Code is useful when we need to clean and process data using scripts.

This helped me understand that both tools are important, and choosing the right tool depends on the project requirement.

3.2 Importing Data into Power BI

I did not directly work on Power BI reports during Week 2, but I understood how datasets can be used across multiple tools. For example, data that is cleaned in Power BI can be used for reporting, and the same data can also be cleaned using scripts for automation.

This made it clear that coding gives more flexibility when working on repeated tasks and large datasets.

3.3 Data Cleaning Using Power Query

Instead of using Power Query this week, I performed similar data cleaning steps through automation in VS Code. I learned how to clean datasets programmatically, which is more efficient when we want to repeat the same process multiple times.

Using VS Code, I worked on automating common cleaning tasks such as:

- Identifying and handling **missing values**

- Removing **duplicate records**

- Correcting **data types and formats**

- Renaming columns for better clarity

- Filtering unnecessary data

- Ensuring the dataset is consistent and ready for analysis

This week helped me understand that automation is very useful because it saves time and avoids manual errors. Once a cleaning script is created, it can be reused whenever the dataset updates.

4. Key Learnings

By the end of Week 2, I learned:

- VS Code is an important tool for real-world coding and data tasks

- Data cleaning can be automated using scripts instead of manual work

- Automation makes cleaning faster, consistent, and reusable

- Repeated tasks like handling null values and duplicates can be easily managed through code

- Writing clean scripts improves workflow and reduces human errors

- This approach will be helpful for future work like EDA and live data handling

5. Challenges Faced

Some challenges I faced during Week 2 were:

- Getting used to the VS Code environment and shortcuts

- Understanding how to structure scripts properly

- Debugging small issues while running the code

- Making sure the automation works correctly for different datasets

However, with practice, I improved my understanding and became more comfortable working with VS Code.

6. Outcome of Week 2

At the end of Week 2, I was able to:

- ✅ Work confidently in VS Code
- ✅ Write and run scripts for data cleaning
- ✅ Automate repeated cleaning tasks
- ✅ Improve accuracy and consistency in the cleaned dataset
- ✅ Build a strong base for EDA in the upcoming week

7. Plan for Week 3

In Week 3, I will focus on **Exploratory Data Analysis (EDA)**, including:

- Understanding the dataset deeply

- Finding patterns, trends, and outliers

- Generating meaningful insights from data

- Preparing data for further processing and model building

Conclusion

Week 2 was an important step in my internship where I learned how to use **VS Code for data cleaning automation**. Instead of manually cleaning data every time, I practiced writing scripts that can automatically clean and prepare datasets. This week improved my coding confidence and prepared me for EDA and live data work in the next weeks.