Finance Data Analysis – My Report

# Introduction

For this assignment, I worked on a financial dataset named 'Finance\_data.csv'. The main goal was to understand the structure of the data, clean it up, analyze key financial metrics like Revenue, Cost, and Profit, and finally draw insights using Python and some cool visualizations.

# About the Dataset

The dataset had multiple columns, but the most important ones were:  
- Revenue – The amount of money generated.  
- Cost – The amount spent or invested.  
- Profit – Calculated as Revenue minus Cost.  
- Category – Type or classification of product or service.  
- Region – Area or location where the sale occurred.  
- Product – The actual product being sold.  
  
I used pandas to load and explore the data. Here's what I did step-by-step:

# Data Cleaning

When I first opened the dataset, I noticed some missing values. So, to make sure the analysis was accurate:  
- I replaced missing numeric values (like Revenue or Cost) with the median.  
- For missing text values (like Category or Region), I used the mode.  
- I also checked for any weird or unexpected values, but the dataset looked clean after handling the nulls.  
  
This step was important because it made sure the dataset was ready for analysis without any errors popping up during calculations or plotting.

# Analysis & Key Metrics

I calculated the mean, median, and standard deviation for numeric columns like Revenue, Cost, and Profit.  
This helped me understand the average performance and how much the values were spread out.  
- I also created a new column called Net\_Profit (Revenue - Cost).  
- Then I explored which products and regions were the most profitable.  
- Which categories brought in the most revenue.  
- How Revenue and Cost were related.

# Data Filtering & Sorting

I filtered out all the rows where the Profit was greater than ₹10,000. That gave me a quick look at the best-performing transactions. I also sorted the data by Revenue to find which entries brought in the most money.

# Grouping & Pivot Tables

Using groupby, I grouped data by Category and calculated average Revenue. Then I used pivot tables to summarize data in a cleaner way:  
- Revenue by each product.  
- Revenue by product and region.  
  
It helped me break down the performance even further and compare between groups.

# Visualizations

To make sense of the numbers, I created a few visualizations:  
1. Histogram of Profit – Helped me see the distribution.  
2. Bar Plot – Average Revenue by Category.  
3. Scatter Plot – Revenue vs Cost – Showed a strong positive relationship.  
4. Box Plot – Profit by Region – Some regions had really high profit variability.

# Final Thoughts

Overall, this assignment helped me understand how to clean data, analyze it, and extract meaningful insights using Python. I enjoyed seeing how numbers told a story, especially through the visualizations. I also realized how important it is to clean the data properly before jumping into any analysis. If I had more time, I’d try adding machine learning to predict future profits based on category, cost, and region.