

MULTI-DOMAIN DATA ANALYSIS PORTFOLIO

❖ Project Overview

The Multi-Domain Data Analysis Portfolio is a comprehensive data analytics project consisting of five real-world data analysis case studies across different domains such as Retail, Education, Weather, Healthcare, and Finance.

The objective of this project is to demonstrate end-to-end data analysis skills

It includes:

- Data collection and cleaning
 - Exploratory Data Analysis (EDA)
 - Statistical analysis
 - Data visualization
 - Business insight generation
 - Professional reporting and portfolio presentation
-

❖ Project Objectives

- Build 5 complete data analysis projects from different domains
- Apply data cleaning, transformation, and validation techniques
- Perform statistical analysis (mean, median, correlation, trends)
- Create multiple data visualizations for insights
- Generate actionable business recommendations
- Develop professional reports and GitHub documentation

❖ **Tools & Technologies Used**

- **Programming Language: Python**
 - **Libraries: Pandas, NumPy, Matplotlib, Seaborn**
 - **Data Formats: CSV (real/simulated datasets)**
 - **Version Control: Git & GitHub**
 - **Documentation: Markdown, PDF Reports**
 - **Environment: VS Code**
-

❖ **Technical Requirements Fulfilled**

- **Five distinct data analysis projects**
- **Pandas for data manipulation and cleaning**
- **Minimum 3 visualizations per project**
- **Business insights and recommendations**
- **Professional PDF reports with executive summaries**
- **Organized GitHub portfolio**
- **Data validation and quality checks**
- **Statistical analysis methods applied**

Week-Wise Project Breakdown

Project 1: Healthcare Data Analysis (COVID-19 Trends)

Week 1

❖ Objective:

Analyze COVID-19 case data to understand infection and recovery trends.

❖ Key Activities:

- Trend analysis of daily cases and recoveries
- Mortality rate calculation
- Regional comparison

❖ Visualizations:

- Line charts for case trends
- Bar charts for region-wise analysis

❖ Insights:

- Identified peak infection periods
- Highlighted regions needing stronger healthcare response

❖ Objective:

Analyze stock market data to observe price movements and volatility.

❖ Key Activities:

- Price trend analysis
- Daily return calculations
- Volatility measurement

❖ Visualizations:

- Line chart for stock price movement
- Histogram of daily returns
- Rolling average plots

❖ Insights:

- Identified high-risk and stable periods
- Provided investment risk observations

❖ Objective:

Evaluate student academic performance and understand factors affecting results.

❖ Key Activities:

- Data preprocessing and validation
- Calculation of pass/fail rates
- Subject-wise performance analysis
- Attendance vs performance correlation

❖ Visualizations Created:

- Bar chart for subject-wise scores
- Heatmap for correlation analysis
- Histogram of score distribution

❖ Insights & Outcomes:

- Strong correlation between attendance and performance
- Identified subjects requiring academic intervention
- Suggested strategies for improving student outcomes

Project 4 – Supermarket Sales Analysis (Retail Domain)

❖ Objective:

Analyze supermarket sales data to understand customer behavior and sales performance.

❖ Key Activities:

- Data loading and cleaning (handling missing values, date formatting)
- Daily and monthly sales trend analysis
- Identification of best-selling products
- Customer purchase behavior analysis

❖ Visualizations Created:

- Line chart for daily/monthly sales trends
- Bar chart for top-selling products
- Pie chart for product category contribution

❖ Insights & Outcomes:

- Identified peak sales periods
- Highlighted high-performing products
- Recommended inventory optimization strategies

Project 5: Weather Data Analysis (Weather Domain)

❖ Objective:

Analyze historical weather data to identify trends and seasonal patterns.

❖ Key Activities:

- Time-series analysis of temperature data
- Seasonal trend identification
- Rainfall distribution analysis
- Detection of extreme weather events

❖ Visualizations Created:

- Line plot for temperature trends
- Box plot for seasonal variations
- Histogram for rainfall distribution

❖ Insights & Outcomes:

- Clear evidence of seasonal temperature fluctuations
- Identified extreme weather anomalies
- Useful insights for agriculture and climate planning

Week 5: Portfolio Integration

❖ Objective:

Integrate all projects into a single professional portfolio.

❖ Key Activities:

- Organized GitHub repository structure
- Unified documentation format
- Summary report creation
- Code optimization and refactoring

❖ Deliverables:

- Well-structured GitHub repository
- Consolidated portfolio overview
- Clean and reusable code

Week 6: Professional Packaging & Submission

❖ Objective:

Prepare the portfolio for professional and academic presentation.

❖ Key Activities:

- Creation of presentation slides
- Executive summary for each project
- Final PDF reports

❖ Final Outputs:

- Complete professional portfolio
 - Ready for Submission
-

Setup and Installation Instructions

System Requirements

- Operating System: Windows / Linux / macOS
 - Python Version: Python 3.8 or above
 - IDE/Editor: VS Code
 - Internet connection (for library installation)
-

❖ Software & Library Requirements

- The following Python libraries are required to run the project:

- pandas
- numpy
- matplotlib
- seaborn

These libraries are listed in the requirements.txt file

❖ Installation Steps

1. Install Python

- Download Python from the official website
- Ensure Python is added to system PATH

2. Create a Virtual Environment

```
python -m venv venv
```

Activate the environment:

- Windows: `venv\Scripts\activate.ps1`

3. Install Required Libraries

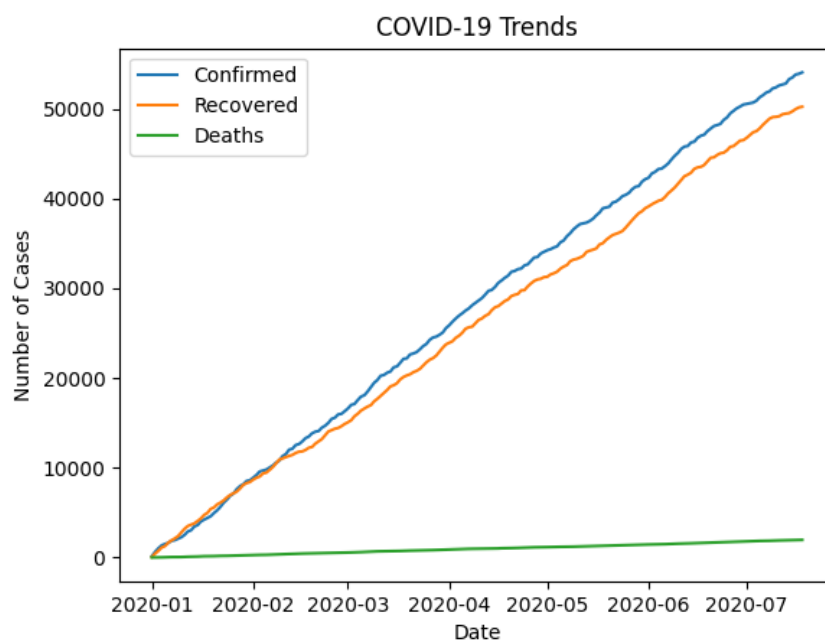
```
pip install -r requirements.txt
```

4. Run the Project

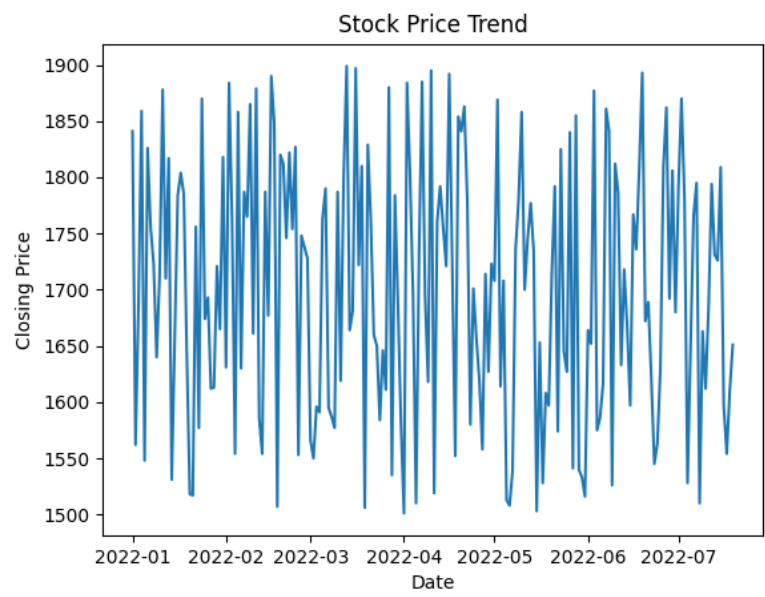
- Open the project folder in VS Code
- Execute the Python scripts or notebooks sequentially

❖ Screenshots of Working Application

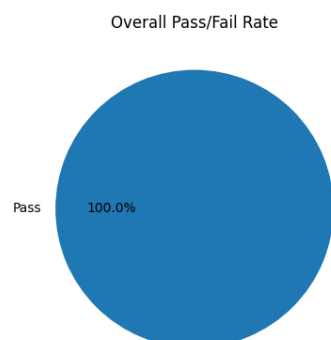
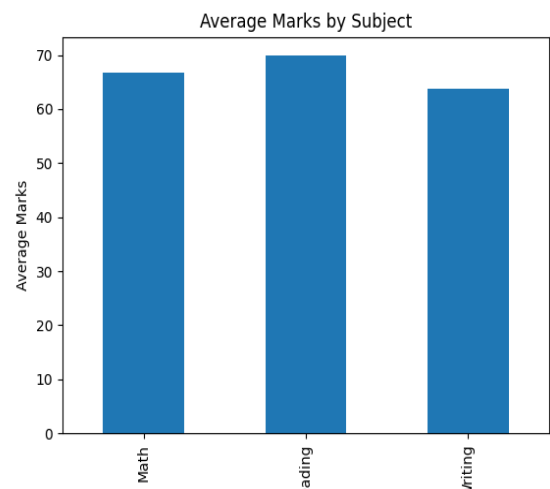
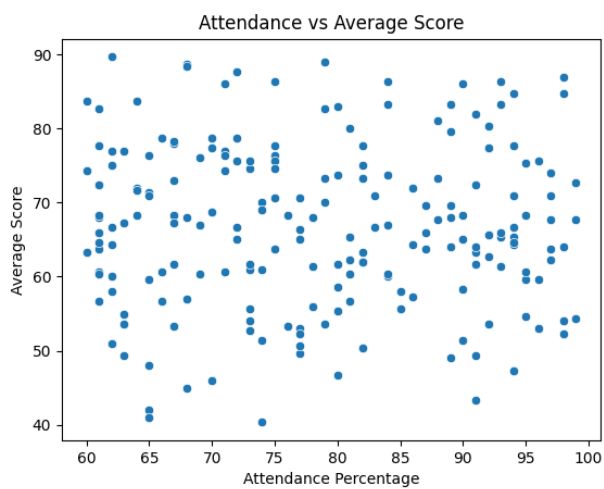
Week 1: Covid-19 Trends



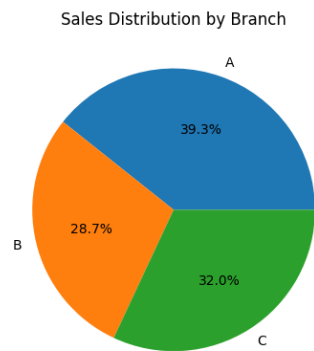
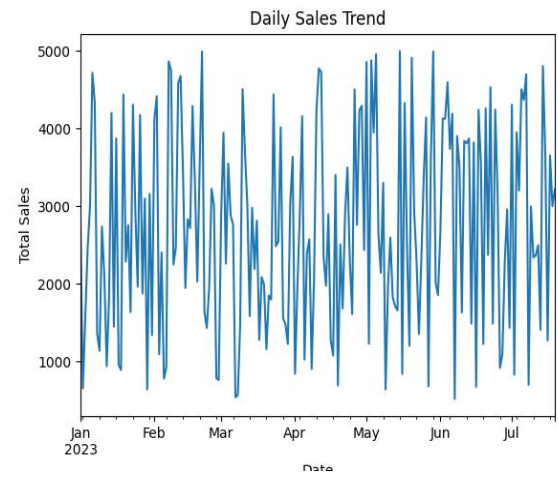
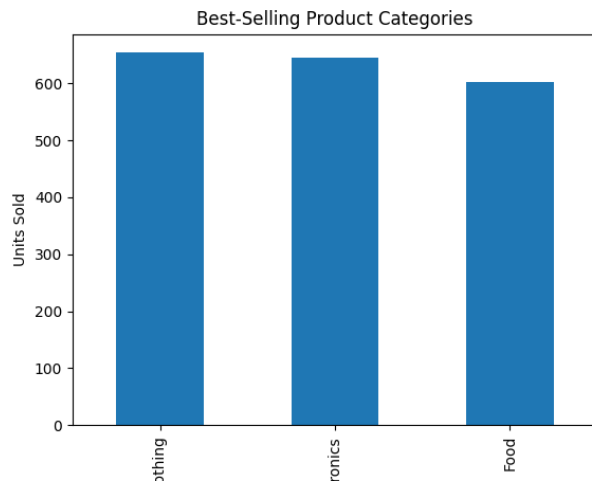
Week 2: Stock Price Trend



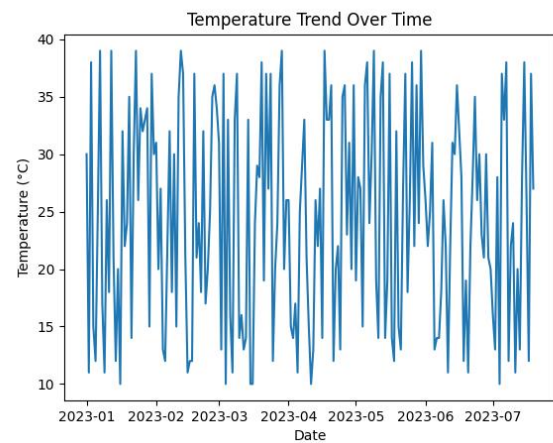
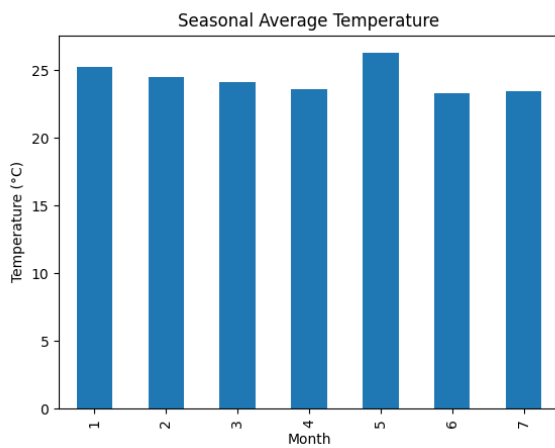
Week 3: Student Analysis



Week 4 (a): Super Market Sales Analysis



Week 4 (b): Super Market Sales Analysis



❖ Explanations

1. Use of Programming Language

- Python was used as the primary programming language for all data analysis tasks.
-

2. Data Handling and Processing

- Pandas was used for:
 - Reading CSV datasets
 - Cleaning missing and inconsistent data
 - Data transformation and aggregation
-

3. Data Analysis Techniques

- Exploratory Data Analysis (EDA) was performed using:
 - Descriptive statistics (mean, median, trends)
 - Correlation analysis
 - Group-based aggregations
-

4. Data Visualization

- Matplotlib and Seaborn were used to create:
 - Line charts for trend analysis
 - Bar charts for comparison
 - Histograms for distribution analysis
 - Heatmaps for correlation analysis
-

5. Multi-Domain Implementation

- The project includes five different domains, proving versatility:

- **Retail**
 - **Education**
 - **Weather**
 - **Healthcare**
 - **Finance**
-

6. Code Quality and Organization

- **Modular folder structure followed**
 - **Meaningful variable names and comments used**
 - **Reusable and readable code practices implemented**
-

7. Documentation and Reporting

- **Detailed documentation provided**
- **Week-wise project breakdown included**
- **Screenshots and visual evidence attached**
- **Professional PDF is provided**
-

❖ Conclusion

This Multi-Domain Data Analysis Portfolio demonstrates practical, real-world data analytics skills across multiple industries. The project highlights the ability to transform raw data into meaningful insights, support data-driven decision-making, and present results in a professional and structured manner.

Author:

Ajit Kumar Bishwkarma

Data Science Intern, The Developers Arena