| **Activity** | |
| --- | --- |
| **Launching the Activity**   * **Open** the data\_cleaning\_manipulation project based on the guidelines mentioned in the TIFC local programming lab guide. * **Examine** the example code snippet provided. | |
| **Virtual\_Environment\_and\_Pandas\_Setup.ipynb**   | ## Set Up a Virtual Environment  In the terminal, navigate to your project folder and run the following command to create a virtual environment:  ```bash  python -m venv .venv  ```  Activate the virtual environment by running:  - \*\*Windows\*\*: `.venv\Scripts\activate`  - \*\*macOS/Linux\*\*: `source .venv/bin/activate`  Once activated, you should see your virtual environment’s name in the terminal prompt.  ## Install Libraries  With the virtual environment activated, install the required libraries by running:  ```bash  pip install pandas openpyxl ipykernel  ``` | | --- |   **Data\_Cleaning\_and\_Manipulation\_Activity.ipynb**   | # Import necessary libraries  import pandas as pd # Pandas library is essential for data manipulation  # Task 1: Import the datasets  # TODO Import the sales data (../data/sales\_data.csv)  sales\_data = ... # Load the sales data CSV file  # TODO Import the product inventory data (../data/product\_inventory.xlsx)  inventory\_data = ... # Load the inventory data Excel file  # Inspect the sales data  print("Sales Data Info:")  print(sales\_data.info()) # Inspect structure and data types of sales data  print(sales\_data.head()) # View the first few rows of sales data  # print(sales\_data) # if you want to see the whole data set  # Inspect the inventory data  print("\nInventory Data Info:")  print(inventory\_data.info()) # Inspect structure and data types of inventory data  print(inventory\_data.head()) # View the first few rows of inventory data  # print(inventory\_data) # if you want to see the whole data set  # Task 2: Data Cleaning  # TODO: Clean the data  # Verify cleaning steps  print("\nCleaned Sales Data:")  print(sales\_data.head())  print("\nCleaned Inventory Data:")  print(inventory\_data.head())  # Task 3: Data Merging  # Merge sales and inventory data  # TODO: Merge the datasets. Replace ColumnName with the key column name  merged\_data = pd.merge(sales\_data, inventory\_data, on='ColumnName', how='left')  # Inspect merged data  print("\nMerged Data Info:")  print(merged\_data.info())  print(merged\_data.head())  # Task 4: Group and Aggregate Data  # Group and aggregate data  # TODO: Group by products names and calculate total quantities sold and prices for each product.  grouped\_data = merged\_data.groupby('ColumnName').agg({'ColumnName': 'aggFunctionName', 'ColumnName': 'aggFunctionName'})  print("\nGrouped Data (Total Sales per Product):")  print(grouped\_data)  # Task 5: Pivot Table  # Create a pivot table  # TODO: Create a pivot table to summarise the quantity of products sold on each date, with product names as rows and dates as columns.  pivot\_table = merged\_data.pivot\_table(index='ColumnName', columns='ColumnName', values='ColumnName', aggfunc='sum') # Define index, columns, values, and aggfunc  # Replace NaN values with empty strings  pivot\_table = pivot\_table.fillna('')  print("\nPivot Table (Sales by Product and Date):")  print(pivot\_table) | | --- |   The first notebook, **Virtual\_Environment\_and\_Pandas\_Setup.ipynb**, offers step-by-step instructions for setting up a virtual environment, installing the required libraries, and running the project.  The second notebook, **Data\_Cleaning\_and\_Manipulation\_Activity.ipynb**, features programming exercises where you will use the Pandas library to clean and prepare two datasets, Product Inventory and Sales Data, for analysis.  Each task section includes instructions and a TODO comment where you are required to write code. | |
| * **Task 0**: Set up the Virtual Environment and Install Libraries   + 1. Open **notebooks > Virtual\_Environment\_and\_Pandas\_Setup.ipynb**.   + 2. Set up the virtual environment   + 3. Activate the virtual environment   + 4. Install the required libraries. * **Task 1:** Data Import   + 1. Open **notebooks > Data\_Cleaning\_and\_Manipulation\_Activity.ipynb**.   + 2. Import the datasets.     - For the CSV file: Use read\_csv.     - For the Excel file: Use read\_excel.   + 3. Inspect the structure of the datasets. * **Task 2:** Data Cleaning   + 1. When cleaning data, look for inconsistent format issues and consider the appropriate methods to address them.   + 2. Refer to the **PY-04 - Data Cleaning Guide - Learner Instructions** handout for assistance identifying and resolving inconsistent formats. * **Task 3:** Data Merging   + 1. Merge the sales and inventory data. * **Task 4:** Group and Aggregate Data   + 1. Summarise the sales data by grouping it based on the product names and calculating the total quantities sold and total prices for each product.     - Identify the column that will act as the group.     - Decide on the aggregation functions to apply to other columns. * **Task 5**: Pivot Table   + 1. Create a table that summarises the quantity of products sold on each date, with product names as rows and dates as columns.     - Think about which columns will serve as the rows (index) and columns in your table.     - Identify the values you want to summarise and how they should be aggregated (e.g., summed up). | |