

Assignment: Understanding Pandas DataFrames

Objective

The goal of this assignment is to learn how to **create, inspect, manipulate, and analyze Pandas DataFrames**. You will practice using **various DataFrame operations** like accessing, adding, modifying, and deleting columns using real-world datasets.

Dataset Information

We'll use the **Toyota Sales Dataset** and **Sales Reps Dataset** for this assignment.

📌 **GitHub Link:** [\[Here\]](#)

Files Provided

- **toyota_sales_with_commission.csv** – Contains Toyota car sales data with commission amounts.
 - **sales_reps_data.csv** – Contains information about Toyota sales representatives.
-

Tasks

Task 1: DataFrame Creation

1. Import **Pandas** and load the **sales_reps_data.csv** into a DataFrame.
2. Display the **first five rows** of the DataFrame.
3. Print the **data type** of the DataFrame.

Task 2: DataFrame Inspection

1. Use the `.shape` attribute to determine the **number of rows and columns**.
 2. Retrieve the **list of column names**.
 3. Use the `.info()` method to get a **summary of the DataFrame**.
-

Task 3: Accessing Data

1. Access the `region` column as a **Pandas Series**.
 2. Access **two columns** (your choice) and display the **first five rows**.
 3. Retrieve the **first 10 rows** using `.head(10)`.
-

Task 4: Handling Missing Data

1. Check for **missing values** across all columns.
 2. For columns with missing values:
 - Replace missing numerical values with the **mean**.
 - Replace missing text fields with **"Unknown"**.
-

Task 5: DataFrame Analysis

1. Use the `.describe()` method to get summary statistics of the DataFrame.
 2. Identify the **column with the highest variance**.
 3. Find the **top 3 most common regions**.
-

Task 6: DataFrame Manipulations

1. Add a new column called `full_name` by concatenating `first_name` and `last_name`.
2. Drop the `first_name` and `last_name` columns.
3. Rename the column `full_name` to `sales_rep_name`.

Task 7: DataFrame Operations

1. Filter the DataFrame to include **only active sales representatives**.
 2. Sort the filtered DataFrame by **hire_date** in **ascending order**.
 3. Reset the index of the sorted DataFrame without adding it as a new column.
-

Bonus Task (Optional)

1. Calculate the **total sales amount** for each region using the **toyota_sales_with_commission.csv** dataset.
2. Merge the **sales reps data** with the **sales data** based on the **sales rep ID**.
3. Identify the **top 5 sales reps** based on **total sales amount**.

