**Big Data Analysis Design and Innovation**

|  |  |
| --- | --- |
| **Date** | **26-10-2023** |
| **Team ID** | **673** |
| **Project Name** | **BIG DATA ANALYSIS WITH IBM CLOUD DATABASES** |

CODE:

import pandas as pd

# Load the CSV file into a pandas DataFrame

csv\_file = r'C:\Users\Admin\Pictures\IBM\rainfall.csv' # Update the file path

df = pd.read\_csv(csv\_file)

# Data Exploration

# Display the first few rows of the DataFrame

print("First few rows of the dataset:")

print(df.head())

# Display summary statistics of numeric columns

print("\n\nSummary statistics of numeric columns:")

print(df.describe())

# Data Cleaning

# Remove rows with missing values

df.dropna(inplace=True)

print("\n\nDataset is cleaned !! and free of null values")

# Remove unwanted columns

unwanted\_columns = ['YEAR']

df.drop(columns=unwanted\_columns, inplace=True)

# Save the cleaned DataFrame to a new CSV file

cleaned\_csv\_file = r'C:\Users\Admin\Pictures\IBM\cleaned\_dataset.csv' # Update the file path

df.to\_csv(cleaned\_csv\_file, index=False)

print("\n\nFirst few rows of the dataset after dropping a column:")

print(df.head())

print("\n\nData analysis, small operations, and data cleaning completed.")



