

# Retail Sales Data Cleaning — Azure DevOps CI/CD Pipeline Setup

## Part 1 - Project Overview

This project demonstrates how to set up an **Azure DevOps pipeline** to automate the following:

- Read the **raw sales data** (raw\_sales\_data.csv)
- Clean the data using **data\_processing.py**
- Generate clean\_sales\_data.csv
- Publish both CSVs as **Azure DevOps artifacts**

## Part 2 -Folder Structure

```
SalesPipelineProject/  
|  
├─ data/  
|   ├─ raw_sales_data.csv  
|   └─ clean_sales_data.csv  
|  
├─ data_processing.py  
├─ requirements.txt  
└─ azure-pipelines.yml
```

## Part 3 -Step-by-Step Azure DevOps Setup

### Step 1 — Create a New Project

1. Go to <https://dev.azure.com>
2. Click **New Project**
3. Enter:
  - **Project Name:** RetailSalesPipeline
  - **Visibility:** Private or Public

#### 4. Click **Create**

### **Step 2 -Push Code to Azure Repo**

git init

git remote add origin <your\_repo\_url>

git add .

git commit -m "Added data cleaning pipeline project"

git push origin main

### **Step 3 -Create Pipeline**

1. Navigate to **Pipelines → Create Pipeline**
2. Select **Azure Repos Git**
3. Choose your repository
4. Select **Existing Azure Pipelines YAML file**
5. Choose azure-pipelines.yml
6. Click **Run**

### **Step 4 - View Pipeline Run**

- Go to **Pipelines → Runs**
- see:
  - Python installed
  - Dependencies installed
  - Data cleaned
  - Artifacts published

### **Step 5 -Download CSV Artifacts**

- Go to **Pipelines → Last Successful Run → Artifacts → SalesData**
- Download raw\_sales\_data.csv & clean\_sales\_data.csv

## Final Outcome

After successful pipeline execution:

- **Input:** raw\_sales\_data.csv
- **Output:** clean\_sales\_data.csv
- Both files stored as **Azure artifacts** automatically.