**PR 1: Purchases.txt Dataset**

**1. Instead of breaking the sales down by store, give us a sales breakdown by product category across all of our stores**

**2. What is the value of total sales for the following categories?**

**􀂇 Toys**

**􀂇 Consumer Electronics**

**3.Find the monetary value for the highest individual sale for each separate store**

**4.What are the values for the following stores?**

**􀂇 Reno**

**􀂇 Toledo**

**􀂇 Chandler**

**5.Find the total sales value across all the stores, and the total number of sales.**

To implement the given tasks in a Python Pig script, we need to handle data processing similarly to how Apache Pig handles large datasets. Here, I'll provide a solution using Python with the help of the pandas library, which is well-suited for data manipulation and analysis.

First, let's assume the Purchases.txt dataset has the following columns:

**store,category,product,amount**

### Step 1: Load the Dataset

We'll use pandas to load the dataset from a CSV file.

**import pandas as pd**

**# Load the dataset**

**df = pd.read\_csv('Purchases.txt')**

### Step 2: Sales Breakdown by Product Category

To get a sales breakdown by product category across all stores:

**# Group by 'category' and sum the 'amount'**

**category\_sales = df.groupby('category')['amount'].sum().reset\_index()**

**print(category\_sales)**

### Step 3: Total Sales for Specific Categories

To find the value of total sales for 'Toys' and 'Consumer Electronics':

**# Filter for specific categories**

**toys\_sales = df[df['category'] == 'Toys']['amount'].sum()**

**consumer\_electronics\_sales = df[df['category'] == 'Consumer Electronics']['amount'].sum()**

**print(f'Toys: {toys\_sales}')**

**print(f'Consumer Electronics: {consumer\_electronics\_sales}')**

### Step 4: Highest Individual Sale for Each Store

To find the monetary value for the highest individual sale for each store:

**# Group by 'store' and find the max 'amount'**

**highest\_sales\_per\_store = df.groupby('store')['amount'].max().reset\_index()**

**print(highest\_sales\_per\_store)**

### Step 5: Values for Specific Stores

To get the values for the stores 'Reno', 'Toledo', and 'Chandler':

**# Filter for specific stores and get the required details**

**reno\_sales = df[df['store'] == 'Reno']['amount'].sum()**

**toledo\_sales = df[df['store'] == 'Toledo']['amount'].sum()**

**chandler\_sales = df[df['store'] == 'Chandler']['amount'].sum()**

**print(f'Reno: {reno\_sales}')**

**print(f'Toledo: {toledo\_sales}')**

**print(f'Chandler: {chandler\_sales}')**

### Step 6: Total Sales Value and Total Number of Sales

To find the total sales value across all stores and the total number of sales:

**# Calculate total sales value and total number of sales**

**total\_sales\_value = df['amount'].sum()**

**total\_number\_of\_sales = df.shape[0]**

**print(f'Total Sales Value: {total\_sales\_value}')**

**print(f'Total Number of Sales: {total\_number\_of\_sales}')**

### Complete Script

Here is the complete script in one place:

**import pandas as pd**

**# Load the dataset**

**df = pd.read\_csv('Purchases.txt')**

**# 1. Sales breakdown by product category**

**category\_sales = df.groupby('category')['amount'].sum().reset\_index()**

**print("Sales Breakdown by Product Category:\n", category\_sales)**

**# 2. Total sales for 'Toys' and 'Consumer Electronics'**

**toys\_sales = df[df['category'] == 'Toys']['amount'].sum()**

**consumer\_electronics\_sales = df[df['category'] == 'Consumer Electronics']['amount'].sum()**

**print(f'Total Sales for Toys: {toys\_sales}')**

**print(f'Total Sales for Consumer Electronics: {consumer\_electronics\_sales}')**

**# 3. Highest individual sale for each store**

**highest\_sales\_per\_store = df.groupby('store')['amount'].max().reset\_index()**

**print("Highest Individual Sale for Each Store:\n", highest\_sales\_per\_store)**

**# 4. Values for specific stores**

**reno\_sales = df[df['store'] == 'Reno']['amount'].sum()**

**toledo\_sales = df[df['store'] == 'Toledo']['amount'].sum()**

**chandler\_sales = df[df['store'] == 'Chandler']['amount'].sum()**

**print(f'Total Sales for Reno: {reno\_sales}')**

**print(f'Total Sales for Toledo: {toledo\_sales}')**

**print(f'Total Sales for Chandler: {chandler\_sales}')**

**# 5. Total sales value and total number of sales**

**total\_sales\_value = df['amount'].sum()**

**total\_number\_of\_sales = df.shape[0]**

**print(f'Total Sales Value: {total\_sales\_value}')**

**print(f'Total Number of Sales: {total\_number\_of\_sales}')**

This script uses pandas to perform data analysis tasks similar to what would be done using Apache Pig for a large dataset. The dataset is assumed to be in CSV format and loaded into a pandas DataFrame for processing.