Assignment 1

# Q1

1. Join the two DataFrames based on the 'Store' column to associate sales with store types. Then, group the data by 'Type' and calculate the total sales for each store type.

The result shows the total sales for each store type:

* Store Type 'A' had total sales of $4,331,014,722.75.
* Store Type 'B' had total sales of $2,000,700,736.82.
* Store Type 'C' had total sales of $405,503,527.54.

1. The average sales for holidays and non-holidays were calculated using the 'IsHoliday' column and aggregated.

The result shows the average sales for holidays and non-holidays:

* On holidays, the average sales are approximately $17,035.82.
* On non-holidays, the average sales are approximately $15,901.45.

Comparing sales performance between holidays and non-holidays, this indicates that, on average, sales are slightly higher during holidays.

# Q2

My approach involves:

**Data Loading**: Initiated the process by importing the social network's friendship data via Google Drive into Spark.

**Data Structuring**: Structured the data into RDD to effectively represent the relationships between users and manipulating of data via Transformations (map, groupByKey, filter, and reduceByKey etc).

**User Relationship Analysis**: My approach involved in-depth analysis of user relationships to differentiate between direct friends and potential new connections.

**Recommendation Generation**: Calculate connections between users by analysing both direct friends and pairs of friends, while tracking mutual friend connections and generated personalized recommendations based on their strength.

Rec Generator

Generate Connections

Group Connections

Data Ingestion

User Analysis

Filter & Sum Connections

Initialize Spark

Data Output

Sort Rec

Transform Data

Recommendations for the users with following user IDs: 10, 152, 288, 603, 714, 1525, 2434, 2681.

* User ID 10: 2,3,4,5,6,7,8,9,11,12
* User ID 152: 2,3,4,5,6,7,8,9,10,11
* User ID 288: 71,1525,69,90,217,2348,2351,2352,2354,2356
* User ID 603: 1,289,290,291,292,293,294,295,296,297
* User ID 714: 1,712,713,715,717,718,1525,90,217,247
* User ID 1525: 288,1,710,714,603
* User ID 2434: 71,288,711,716,719,720,2348,2351,2352,2354
* User ID 2681: 71,288,710,711,716,719,720,721,722,2348