SI 201 Project 1

The name of the dataset: Sample Superstore Dataset
Collaborators: Xiwen Mark, Chih-Hsiang Chang
GenAI: We might need to use generative AI for later implementation process
Columns: Sales, profits, discount, subcategory, category

Questions and Calculations:
Load_csv():

a. Output: cat_dict: dict{category: {subcat:[sales, discount, profits]}}

Get function: total sales, profits and discount

a. **Function name:** get_total_sales_discount_profits(cat_dict)
b. Input: dict{category: {subcat:[sales, discount, profits]}} *[cat_dict]*
c. Output: dict{category: {subcat:[total_sales, total_discount, total_profits]} *[cat_total_dict]*

Get function: total entries of category

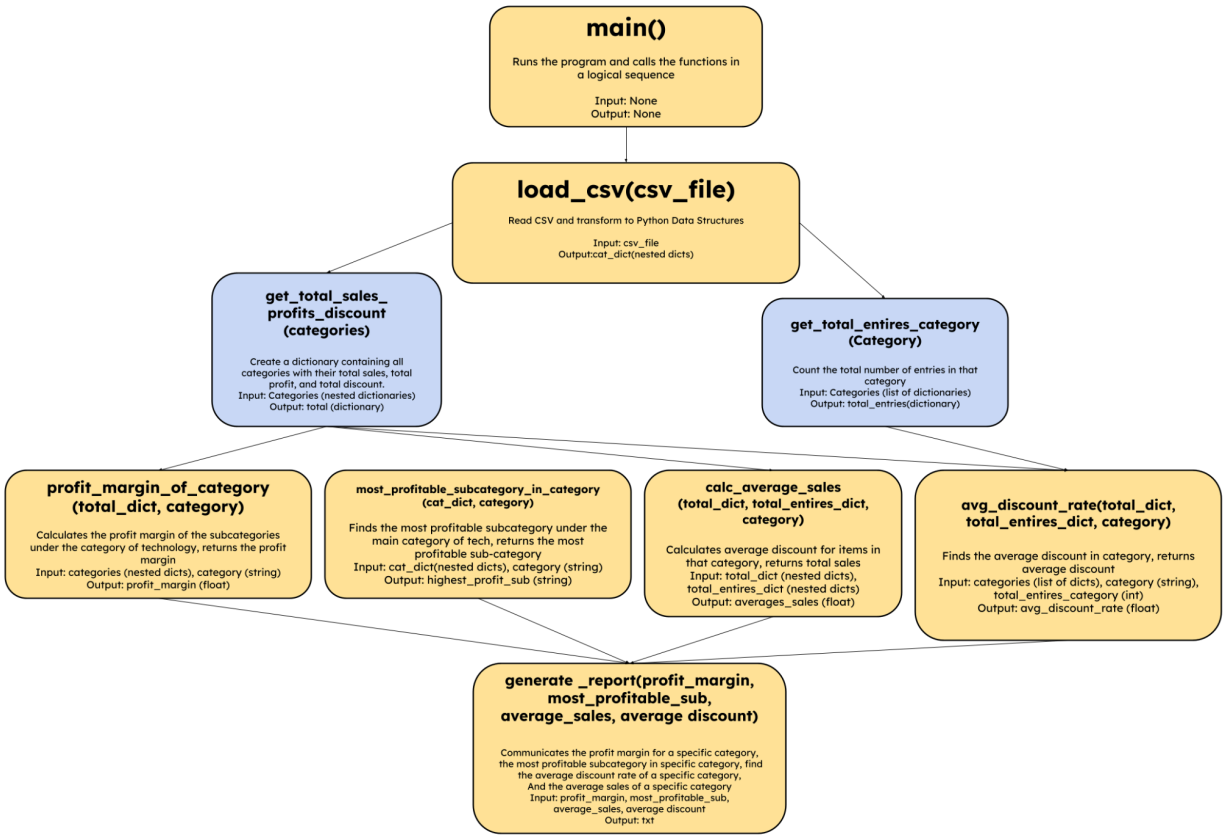
a. **Function name:** get_total_entires(cat_dict)
b. Input: dict{category: {subcat:[sales, discount, profits]}}*[cat_dict]*
c. Output: dict{category: total_entires(int)]}

1. **What is the profit margin of Technology? (Chih-Hsiang Chang)**
    a. **Function name**: profit_margin(cat_dict, category)
    b. Input: Input: dict{category: {subcat:[total_profits, total_sales, total_discount]} *[get function 1]*, category (string)
    c. Output: profit_margin (float)

2. **What is the most profitable subcategory in Technology? (Chih-Hsiang Chang)**
    a. **Function name**: most_profitable_subcategory_in_category(cat_dict, category)
    b. Input: dict{category: {sub:[profits, sales, discount]}}, category(string)
    c. Output: subcategory (string)
        i. Sum up profits in each sub category, and compute the highest profit

3. **What is the average sales for Technology? (Xiwen Mark)**
    a. **Function name**: calc_average_sales(cat_dict, total_entires_category, category)
    b. Input: dict{category: {subcat:[total_profits, total_sales, total_discount]} *[get function 1]*, total entries of category *[get function 2],* category(string)
    c. Output: average sales (float)
        i. total sales / n

4. **What is the average discount rate for Technology? (Xiwen Mark)**
    a. **Function name**: avg_discount(cat_dict, total_entires_category, category)
    b. Input: dict{category: {subcat:[total_profits, total_sales, total_discount]} *[get function1],* dict{category: total_entires(int)]} *[get function 2],* category(string)
    c. Output: average_discount (float)
        i. total discount / n


main()
Runs the program and calls the functions in a logical sequence
Input: None
Output: None
load_csv(csv_file)
Read CSV and transform to Python Data Structures
Input: csv_file
Output:cat_dict(nested dicts)
get_total_sales_ profits_discount (categories)
Create a dictionary containing all categories with their total sales, total profit, and total discount.
Input: Categories (nested dictionaries)
Output: total (dictionary)
get_total_entires_category (Category)
Count the total number of entries in that category
Input: Categories (list of dictionaries)
Output: total_entries(dictionary)
profit_margin_of_category (total_dict, category)
Calculates the profit margin of the subcategories under the category of technology, returns the profit margin
Input: categories (nested dicts), category (string)
Output: profit_margin (float)
most_profitable_subcategory_in_category (cat_dict, category)
Finds the most profitable subcategory under the main category of tech, returns the most profitable sub-category
Input: cat_dict(nested dicts), category (string)
Output: highest_profit_sub (string)
calc_average_sales (total_dict, total_entires_dict, category)
Calculates average discount for items in that category, returns total sales
Input: total_dict (nested dicts), total_entires_dict (nested dicts)
Output: averages_sales (float)
avg_discount_rate(total_dict, total_entires_dict, category)
Finds the average discount in category, returns average discount
Input: categories (list of dicts), category (string), total_entires_category (int)
Output: avg_discount_rate (float)
generate _report(profit_margin, most_profitable_sub, average_sales, average discount)
Communicates the profit margin for a specific category, the most profitable subcategory in specific category, find the average discount rate of a specific category, And the average sales of a specific category
Input: profit_margin, most_profitable_sub, average_sales, average discount
Output: txt