In this project, I will be using the dataset ‘Shawn\_Expenses\_in\_2022’. I will also be using what I have learned in Google Data Analytics course by applying the data analysis process or case study roadmap I called ‘APPASA’ whereby the acronym stands for:

A: Ask

P: Prepare

P: Process

A: Analyze

S: Share

A: Act

I will proceed with the first step, as it is very important to ask questions before you even touch the data. Defining a problem and focusing on it throughout the whole data analysis process is very important, as one might easily stray away.

Let’s move on to the first step:

**Case Study Roadmap 1: ASK**

1. **What topic am I exploring?**

Personal Finance. This is a very common topic that I have plan and track my personal data since the moment I took an interest in data analysis.

1. **What is the main problem?**

I personally notice that I could not save money as per my goal, and I need to figure out my purchasing trends and determine the key factors that is causing the problem.

1. **What type of data will be appropriate for my analysis and how to obtain it?**

Quantitative and continuous data. Since I have been using MoneyManager (MM) application for quite some time (I also paid for premium version), I will be extracting the dataset from my personal MM account. These data are in numerical form and has an infinite number of probable values that can be selected (I am not rich but am still bold enough to use keyword ‘infinite’ in my statement!)

1. **Who are my audience/stakeholders?**

As this is a personal project, I will be publishing this for everyone to see and learn (yes, you are reading it right now!). I will also be asking questions to myself and make sure to identify the questions that are crucial to the proposed problem.

1. **How can my insights help me make decisions?**

By determining the trend of purchase through categories, I can identify the key point that took up a portion of my expenses, and from there I will propose plans that will help me save money.

1. **What metrics am I using?**

Some of the basic metrics that I am using are shown as below:

Income – How much money am I paid? Full time, part time, bank interest?

Expenses – Total money I spent, every cent of it.

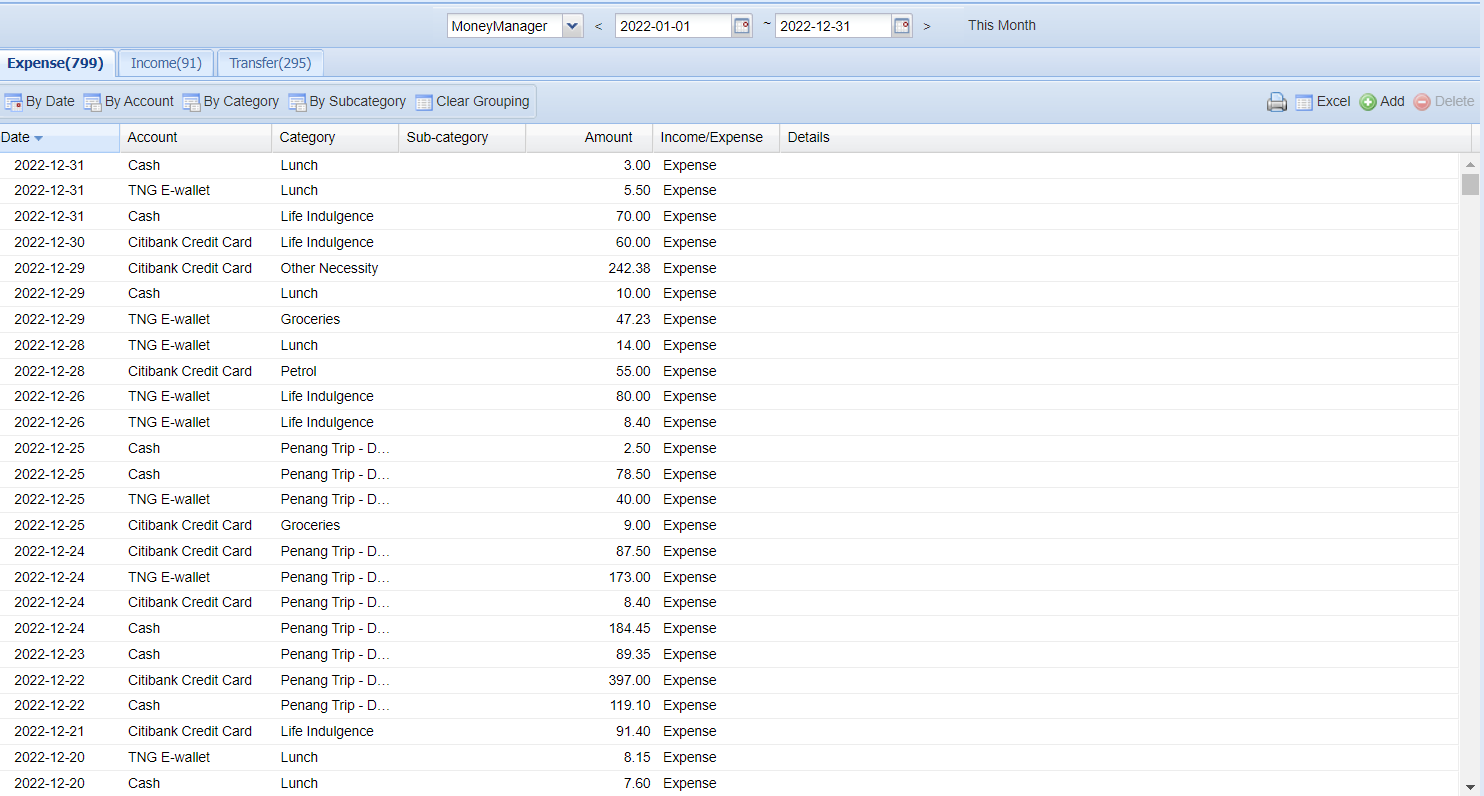
Savings – If you deduct your expenses from income, you will get savings, which is a main metric for today’s analysis.

**Case Study Roadmap 2: PREPARE**

Since my data is located in MM application, I would just require to extract the data from its desktop version.

There are certain ways the data had been organized, by default is by date, however we will extract first the data without tinkering much on the tools provided.

Now, it is important that the data has to be ROCCC: Reliable, Original, Comprehensive, Current and Cited. My data is original and reliable because the information is uploaded by me myself base on my personal expenses, incomes and transfer of money between accounts. Since the expenses are all categorized, it is also comprehensive enough to determine the key factors. Finally, it is current data that is recorded periodically throughout 2022.

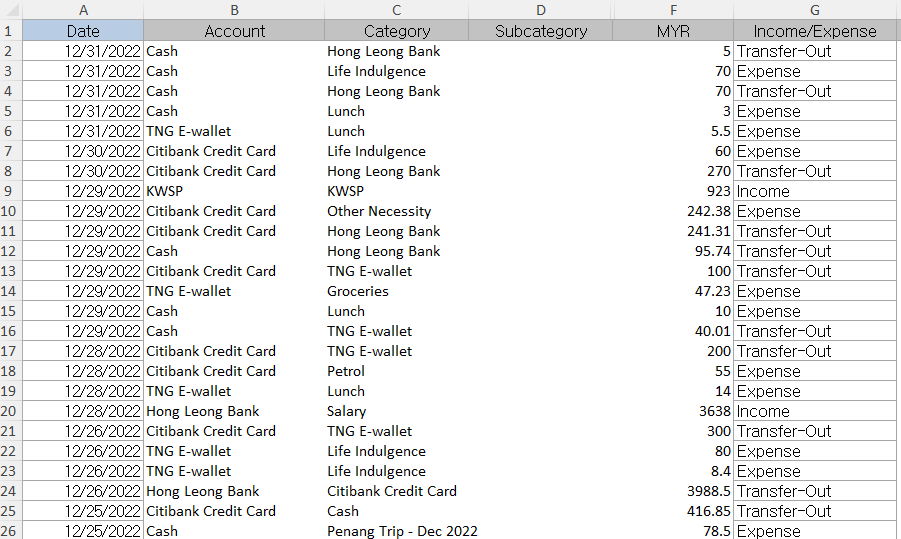


**Case Study Roadmap 3: PROCESS**

Since the data contains 1185 rows (meaning there are 1185 piece of transaction recoded throughout year 2022, whether be it expenses, incomes or transfer), the data size is considered small and can be process through spreadsheet.

The downloaded file is in .XLS format, hence in this case, I will be using Microsoft Excel. For visualization part, I will use Tableau.

Let’s start with the first and most important objective when looking at the excel file: To make it simple for analyzing. By looking at the dataset below, there is one pain point that can be identify and resolve to make the whole analyzing process easier: Too many type of accounts



Firstly, by using =SUM(1/COUNTIF(B2:B1186, B2:B1186)), I identify that there are a total of 20 distinct account categories (Cash, TNG E-wallet, Hong Leong Bank…etc) which can be further simplify by combining some of the categories into one single account. I also use =UNIQUE(B2:B1186,FALSE) to identify the distinct accounts.



Next, I categorize into 4 types of account: Cash, Loan, Credit, Savings. I use the Find & Select tools to replace all the accounts into the 4 accounts as mentioned above.

After completing the cleaning process of column B: Account, now I move on to column C: Category. I apply the same method as how I clean column B. By identify the distinct categories as 41 Accounts, I categorized the accounts into a few categories. Before I take a further step, I firstly researched on what is the best personal budget rule that has to be followed, and pick out the categories after catering to my goals (purchasing an apartment etc). Below are the 7 categories that I considered in expenses: Rent & Utilities, Transportation, Parents, Food (Lunch, Groceries), Misc (Leisure), Savings, Travel. For income, I will only categorize it as income.

I also clean the data by removing ‘Transfer-Out’ from Income/Expense column since it is negligible to the overall analysis. I firstly filter Income/Expense column and unselect ‘Income’ and ‘Expense’, that way I will only have data that is visible which are ‘Transfer-Out’ data. I also unselect the header row which is crucial. Then, I applied Find & Select -> Go to Special and select ‘Visible Cells only’. This enables me to select all the ‘Transfer-Out’ rows only and delete those without accidentally deleting the hidden rows. ‘Modifed Bal.’ are also chosen to be deleted as it recorded transactions that are unclear/forgotten. The amount is not judge to be not too significant to the overall analysis (approx. 1.5k loss throughout whole year)