Mason Hunerkoch

Google Data Analytics Capstone

17 July 2025

OVERVIEW

The purpose of this project is to bring in personal spending data and break down my buying habits. The data collection started in 2015 and is currently still being gathered.

I'm undergoing a career change and will most likely take a pay cut leaving the military. I want to see what trends exist in my spending habits and how significant they are. Last, are there any areas I can work on and cut back?

This project will serve as an example of the data life cycle from start to finish. We will explore the steps: ask, prepare, process, analyze and share through the life of this project.

GOALS

1. Execute all the data life cycle steps
2. The focus of this project is the process and demonstration of abilities
3. Find major budgetary trends and assess where scaling back is an option

PHASES

Ask

First, let's evaluate the task, the audience, and the data.

1. Our task is to use the full data life cycle to facilitate and guide our handling of budgetary data so that we can uncover trends and use this information to reduce spending.
2. This project has two layers: at surface level, I am on my own audience with my own goals. I know that as a stakeholder, I would appreciate a dashboard that is easy to update and allows for limited interaction.

   The primary audience consists of those looking through this case study. I will work to seamlessly integrate a real world feel while also elaborating on my own process and decision making at each step.

3. The data I'll be working with is self generated and therefore does not require any outside entities permission. My data comes from tracking 99% of purchases over the last eight

years. From 2017 through 2023, the data is broken into several categories with no further details. From 2023 to present day, the data consists of each individual transaction. We'll explore ways to integrate all eight years of data and, if able, a more detailed analysis of the last two years of data.

Metrics we'll look at: we'll evaluate spending over time, spending by category, and a combination of the two. Hopefully these metrics can provide further insight when combined with more perspective, such as spending around holidays and life events.

Prepare & Process

The data is currently stored in two locations. I'll detail the process for how I set up and cleaned the data. While objectively not a lot of data, I am going to work with the data in SQL.

Data Source A:

I have archived data starting in July 2017 through August 2023 in a google sheet. This data consists of the sum of my purchases in a given month split into major categories. In the earlier years, the data is broken into 5 categories but later expands to seven.

	A	B	C	D	E	F	G	H	I
1	Year	Month	Gas	Groceries	Meals	Dogs	House/Car	Travel	Other
2	2017	July	$63.00	$411.00	$81.00	$200.00	$0.00	$0.00	$105.00
3	2017	August	$190.56	$298.25	$148.80	$156.99	$0.00	$0.00	$91.91
4	2017	September	$299.55	$179.02	$194.78	$50.00	$0.00	$0.00	$858.92
5	2017	October	$153.25	$160.96	$175.19	$125.00	$0.00	$0.00	$369.21
6	2017	November	$73.26	$189.36	$184.34	$86.97	$0.00	$0.00	$372.45
7	2017	December	$0.00	$204.24	$111.37	$246.01	$0.00	$0.00	$450.05
8	2018	January	$201.07	$291.65	$247.62	$171.43	$0.00	$0.00	$94.62
9	2018	February	$212.82	$101.88	$241.79	$132.61	$0.00	$0.00	$559.61
10	2018	March	$800.00	$100.00	$800.00	$200.00	$0.00	$0.00	$2,200.00
11	2018	April	$301.26	$158.98	$563.65	$0.00	$0.00	$0.00	$2,397.84
12	2018	May	$86.25	$40.16	$84.53	$106.14	$0.00	$0.00	$167.64
13	2018	June	$0.00	$23.28	$0.00	$686.08	$0.00	$0.00	$100.22
14	2018	July	$0.00	$0.00	$19.00	$357.00	$0.00	$0.00	$177.89
15	2018	August	$0.00	$0.00	$28.00	$93.00	$0.00	$0.00	$56.17
16	2018	September	$0.00	$0.00	$116.24	$171.35	$0.00	$0.00	$168.17
17	2018	October	$130.89	$198.74	$548.01	$116.00	$0.00	$0.00	$2,488.05
18	2018	November	$248.00	$148.21	$169.21	$270.94	$474.38	$0.00	$1,040.68
19	2018	December	$211.92	$103.95	$442.12	$237.09	$1,580.75	$0.00	$1,807.56
20	2019	January	$109.98	$145.66	$498.25	$160.00	$285.01	$0.00	$99.23
21	2019	February	$59.98	$124.96	$484.06	$215.66	$495.39	$0.00	$287.40
22	2019	March	$299.89	$152.15	$511.91	$177.00	$456.51	$0.00	$155.33
23	2019	April	$269.93	$227.79	$636.09	$618.02	$811.17	$0.00	$610.11

Data Source B:

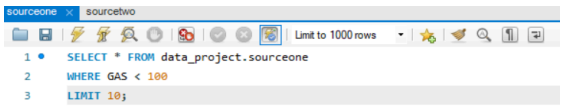
The second data source lives on a mobile application designed for managing transactions manually. I'll start by exporting this data also to a google sheet.

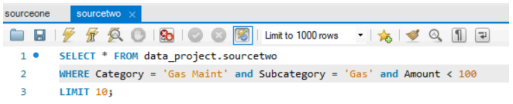
	A	C	D	E	F	G	H	I	J	K
1	Date	Category	Subcategory	Note	USD	Income/Expense	Description	Amount	Currency	Account
2	07/16/2025 17:16:54	Yelena Salary			1706.4	Income		1706.4	USD	1706.4
3	07/16/2025 7:05:32	Bills	Gas		28.77	Expense		28.77	USD	28.77
4	07/16/2025	Subscriptions	Gym		25.05	Expense		25.05	USD	25.05
5	07/15/2025 20:25:02	Passive	Dividends	O	2.69	Income		2.69	USD	2.69
6	07/15/2025 20:24:31	Passive	Dividends	MAIN	5.1	Income		5.1	USD	5.1
7	07/15/2025 20:24:18	Passive	Dividends	OBDC	10.11	Income		10.11	USD	10.11
8	07/14/2025 20:34:19	Bills	Water		85.86	Expense		85.86	USD	85.86
9	07/14/2025 20:30:50	Passive	Dividends	Bonds	3.29	Income		3.29	USD	3.29
10	07/14/2025 8:10:28	Passive	Upside		15	Income		15	USD	15
11	07/14/2025 6:42:10	Groceries	Home		76.83	Expense		76.83	USD	76.83
12	07/14/2025 6:41:55	Home			50.49	Expense		50.49	USD	50.49
13	07/13/2025 13:25:02	Groceries	Home		172.05	Expense		172.05	USD	172.05
14	07/13/2025	Subscriptions	Peacock		7.99	Expense		7.99	USD	7.99
15	07/12/2025 12:35:31	Passive	Upside		1.35	Income		1.35	USD	1.35
16	07/12/2025 12:35:22	Passive	Upside		1.81	Income		1.81	USD	1.81
17	07/12/2025 11:50:13	Groceries	Home		21.89	Expense		21.89	USD	21.89
18	07/12/2025 6:57:28	Savings Interest			0.06	Income		0.06	USD	0.06
19	07/11/2025 21:53:42	Food	Snacks		7.66	Expense		7.66	USD	7.66
20	07/11/2025 21:06:52	Entertainment		Board game	43.78	Expense		43.78	USD	43.78
21	07/11/2025 19:30:52	Gas & Maint.	Gas		32.65	Expense		32.65	USD	32.65

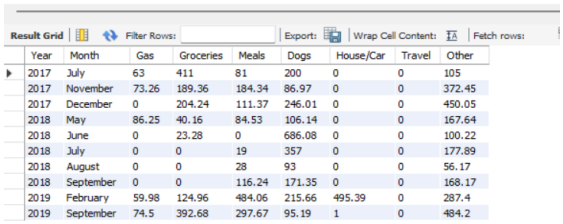
This data, while similar in concept, is very different.

I decided to start cleaning in google sheets prior to migrating the data over to a MySQL database. There were extra columns in the second set of data that I removed. With both sheets I used a few formulas to ensure data types were limited to INT, TEXT, DATETIME, and DECIMAL.

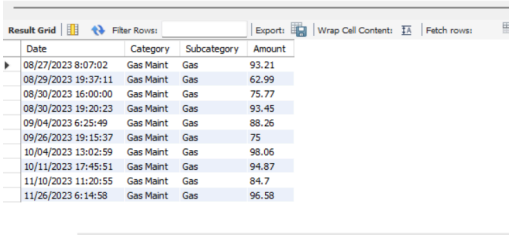
I went ahead and created tables to bring over the data, below are pictures showing the different layouts for each data set. I can achieve similar views of the data but via different statements.





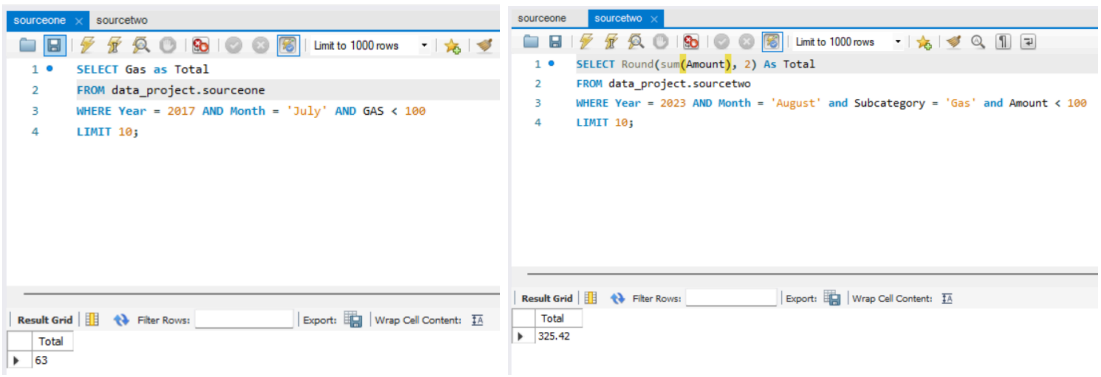


Year	Month	Gas	Groceries	Meals	Dogs	House/Car	Travel	Other
2017	July	63	411	81	200	0	0	105
2017	November	73.26	189.36	184.34	86.97	0	0	372.45
2017	December	0	204.24	111.37	246.01	0	0	450.05
2018	May	86.25	40.16	84.53	106.14	0	0	167.64
2018	June	0	23.28	0	686.08	0	0	100.22
2018	July	0	0	19	357	0	0	177.89
2018	August	0	0	28	93	0	0	56.17
2018	September	0	0	116.24	171.35	0	0	168.17
2019	February	59.98	124.96	484.06	215.66	495.39	0	287.4
2019	September	74.5	392.68	297.67	95.19	1	0	484.2



Date	Category	Subcategory	Amount
08/27/2023 8:07:02	Gas Maint	Gas	93.21
08/29/2023 19:37:11	Gas Maint	Gas	62.99
08/30/2023 16:00:00	Gas Maint	Gas	75.77
08/30/2023 19:20:23	Gas Maint	Gas	93.45
09/04/2023 6:25:49	Gas Maint	Gas	88.26
09/26/2023 19:15:37	Gas Maint	Gas	75
10/04/2023 13:02:59	Gas Maint	Gas	98.06
10/11/2023 17:45:51	Gas Maint	Gas	94.87
11/10/2023 11:20:55	Gas Maint	Gas	84.7
11/26/2023 6:14:58	Gas Maint	Gas	96.58

After creating mirror columns in data set two for month and year and migrating the date column over, we have come one step closer to integrating the data.



I next created a modified version of the data source two so that it mirrors data source one and then created a combined table:

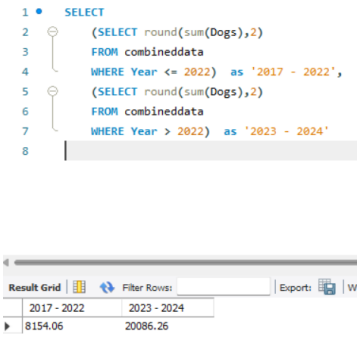
```
CREATE TABLE data_project.combineddata AS
SELECT Year, Month, Gas AS Gas, Groceries AS Groceries, Meals AS Meals, Dogs AS Dogs, `House/Car` AS `House/Car`, Travel AS Travel, Other AS Other
FROM(SELECT Year, Month,
            Gas,
            Groceries,
            Meals,
            Dogs,
            `House/Car`,
            Travel,
            Other
        FROM data_project.sourceone
        UNION ALL
        SELECT Year, Month,
            Gas,
            Groceries,
            Meals,
            Dogs,
            `House/Car`,
            Travel,
            Other
        FROM data_project.modifiedsourcetwo) AS combined_sources
GROUP BY
    Year,
    Month
ORDER BY
    Year,
    Month;
```

This combined data table is in the format of data source one where we have each month's totals in common categories.

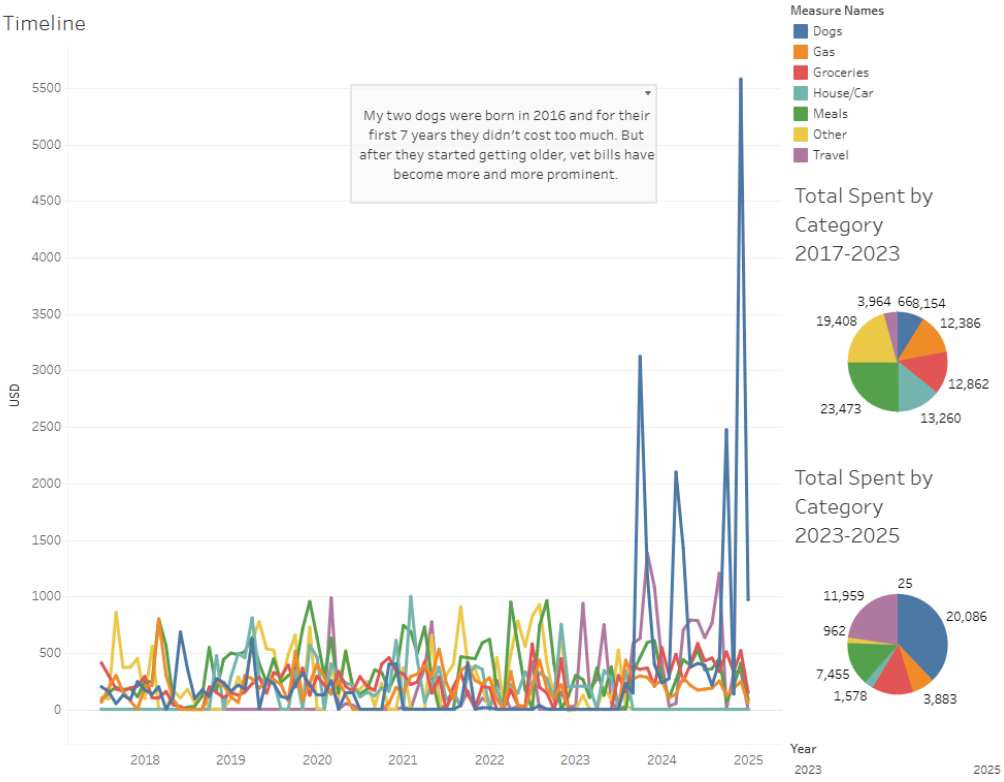
We'll next take this data into the process phase.

Analyze

I'm primarily choosing to work with MySQL for working with the data and Tableau for displaying. While looking through the data I noticed increased totals within the dog category towards the end. Throwing together a quick timeline in Tableau showed me that the totals spent on the dogs in the last couple of years were significantly higher.



SQL statements grouping dog costs by year



This prompted me to add some pie charts and really compare my dog costs in two new ways. I wanted to see what the costs were for the first 6 years vs the last 2 years and how they look

compared to the other categories. Alternatively we could have done a pie chart just comparing the two time periods.

TBC

SHARE

ACT