Gen AI for Startup Founders Data Analysis exercise

To start we're going to be using ChatGPT for Excel analysis and then will look at how it compares to M365 Copilot Chat and Copilot for Excel!



- 1. To start navigate to http://chat.openai.com/ and sign up or log in
- 2. If you have ChatGPT Plus (Paid account) you will have access to multiple models:

DC	ChatGPT 40 V	
	Model	0
	GPT-40 Great for most questions	0
	GPT-40 with scheduled tasks (BETA Ask ChatGPT to follow up later	
	o1 Uses advanced reasoning	What can I help with?
	o3-mini Fast at advanced reasoning	age ChatGPT
	o3-mini-high Great at coding and logic	(Search ····
	More models	>
	Temporary chat	Create image 😑 Summarize text 😤 Surprise me 💿 Analyze images More

- **GPT -4O and GPT4o mini (Free and Plus)** has a memory feature so ChatGPT can understand you better (but a Temporary chat can be used to avoid this). It is an accomplished model that is capable of advanced data analysis. This is the model that is used by default in the free version of ChatGPT and what we will use for this exercise. Note that usage is limited with a free account and it will auto switch back to mini based on usage and demand.
- **OpenAl o1 and Open Al o1 mini**: Intentionally slower that 40 models because they provide advanced reasoning for maths and coding. Less suited to everyday tasks and no web browsing or upload capabilities.
- **OpenAl O3 / O3 mini-high:** Announced in December 2024, o3 builds upon the o1 model's foundation, further enhancing reasoning capabilities. When I asked ChatGPT



🕸 Learn Data Insights

why I might use o3 rather than o1 its response was "Given your background in **financial modelling and Power BI**, you might find **o3** more effective for Analysing large datasets. Structuring financial models, Generating better insights from complex queries".

- 3. Review the ChatGPT interface:
 - 3.1. Chat History: here you can revisit any of the chats you have initiated previously
 - 3.2. Explore GPTs: Here you can explore custom GPTs that people have made available e.g. ChatGPT for Canva,



3.3. Profile: Here you can customise ChatGPT (e.g. to provide some instructions on the kind of work you do, here you can tell it what type of tasks you do (including what kind of data you work with in your business and what kind of things you want to use AI to help with!). You can also provide a style that you want ChatGPT to respond with e.g. if you want it to provide examples and external links.





Customize ChatGPT Introduce yourself to get better, more personalized responses ③
What should ChatGPT call you?
What do you do?
Gastroenterologist
What traits should ChatGPT have? ①
Describe or select traits
+ Chatty + Witty + Straight shooting + Encouraging + Gen Z
+ Skeptical + Traditional + Forward thinking + Poetic 0
Cancel Save

- 3.4. MyGPTs (Paid account only): You can build a new GPT to train it on specific tasks and data that you use regularly!
- 4. Open and review the "Invoice Data_Original.xlsx" file that we will be working with and review it – it contains over a thousand invoices from various customers, with different organisation units and amounts in different currencies (which can be converted to the base currency – USD – using the Exchange Rates tab)

1	Source.Name	Date 💽 Transaction Type 💽	Product/Service	💌 Qty 💌 Rate 💌	Account
2	/*abr/*ah/*am J. Willi/*ams 20201130Inv.xlsx	44164 Invoice			1210 Accounts Receivable
3	/*abr/*ah/*am J. Willi/*ams 20201130Inv.xlsx	44164 Invoice	Case Management	-1 125	4100 Legal Fee Income
4	/*abr/*ah/*am J. Willi/*ams 20201130Inv.xlsx	44164 Invoice		-0.75 125	4100 Legal Fee Income
5	/*abr/*ah/*am J. Willi/*ams 20201231Inv.xlsx	44170 Invoice			1210 Accounts Receivable
6	/*abr/*ah/*am J. Willi/*ams 20201231Inv.xlsx	44170 Invoice	Reimbursable Expenses:Filing & Court fees	150	4700 Reimbursed Exp. (Income)
7	/*abr/*ah/*am J. Willi/*ams 20201231Inv.xlsx	44170 Invoice	Reimbursable Expenses:Overnight Mail	-2 35	4700 Reimbursed Exp. (Income)
8	Brend/*a Bowden Showroom 20200531Inv.xlsx	43980 Invoice			1210 Accounts Receivable
9	Brend/*a Bowden Showroom 20200531Inv.xlsx	43980 Invoice	Markup	-180 0.1	4500 Markup on Reimbursable Expenses
10	Brend/*a Bowden Showroom 20200531Inv.xlsx	43980 Invoice	Reimbursable Expenses:Filing & Court fees	-3 50	4700 Reimbursed Exp. (Income)
11	Brend/*a Bowden Showroom 20200531Inv.xlsx	43980 Invoice	Reimbursable Expenses:Overnight Mail	-1 15	4700 Reimbursed Exp. (Income)
12	Brend/*a Bowden Showroom 20200531Inv.xlsx	43980 Invoice	Reimbursable Expenses:Overnight Mail	-1 15	4700 Reimbursed Exp. (Income)
13	Brend/*a Bowden Showroom 20200531Inv.xlsx	43980 Invoice	Telephone	-4 95	4100 Legal Fee Income
14	Brend/*a Bowden Showroom 20200531Inv.xlsx	43980 Invoice	Legal Writing	-12 95	4100 Legal Fee Income
15	Brend/*a Bowden Showroom 20200531Inv.xlsx	43980 Invoice	Court Appearance	-2 95	4100 Legal Fee Income
16	Brend/*a Bowden Showroom 20200531Inv.xlsx	43980 Invoice	Court Appearance	-8 95	4100 Legal Fee Income
17	Brend/*a Bowden Showroom 20200531Inv.xlsx	43980 Invoice	Meetings	-4 95	4100 Legal Fee Income
18	Brend/*a Bowden Showroom 20200531Inv.xlsx	43980 Invoice	Legal Research	-10 95	4100 Legal Fee Income
19	Brend/*a Bowden Showroom 20200531Inv.xlsx	43980 Invoice	Travel	-2 95	4100 Legal Fee Income
20	Brend/*a Bowden Showroom 20200531Inv.xlsx	43980 Invoice	Legal Writing	-12 95	4100 Legal Fee Income

5. We probably need to clean the data before we can apply analytics on it. There's a screenshot of the data saved in the folder, let's attach that to ChatGPT and ask what cleaning steps we can apply.

"**Prompt**: I have the attached invoice data in Excel which I would like to perform some analysis on, though it needs cleaning first. Please suggest some data cleaning steps that I can apply"

5.1. ChatGPT has come up with some data cleaning steps, some of which are less obvious than others. This is a useful starting point!





Your invoice data needs cleaning before analysis. Here are some key data cleaning steps you can apply:

1. Fix Date Format

• The Date column contains numbers like 44164, which appears to be in Excel's serial date format. Convert it to a proper date format using:

```
excel O Copy & Edit
=TEXT(A2, "YYYY-MM-DD")
```

Or use Power Query to transform it into an actual date.

2. Standardize Source.Name

 Remove special characters (/*abr/*ah/*am ...) from the Source.Name column to make it more readable.

3. Handle Missing or Blank Values

• Identify missing values in critical fields (Transaction Type, Product/Service, Qty, Rate, Account) and decide whether to fill them (e.g., using previous values) or remove rows if necessary.

4. Normalize Qty and Rate Columns

- Ensure Qty values are numeric (convert text numbers to real numbers).
- Check for inconsistencies, like negative values, and ensure they align with business rules.

5. Remove Duplicate Entries

5.2. It suggests that it can automatically apply these cleaning steps in Python or Power Query but perhaps we want to do this in a simpler way. Rather than just taking its suggestion at face value, let's ask for some options on how to fix the date format so we can choose the right one.

Prompt: "what are the ways in which I can fix the date format?"

- 5.3. ChatGPT gives you multiple options including Excel formulas, Power Query, VBA, Python or SQL. You happen to have come across this before and think it can just be changed by changing the number format so you ask Chat GPT: "is this not just as simple as changing the data type of the column?"
- 5.4. And indeed ChatGPT gives us the "quick fix" that in our case works just fine and doesn't require any code or even formulas!





is this not just as simple as changing the data type of the column?

Yes! In Excel, if the Date column is already stored as an Excel serial date, the easiest way to fix it is:

Quick Fix: Change Data Type

- 1. Select the Date column.
- 2. Go to the "Home" tab \rightarrow Click "Number Format" (Dropdown in Number Section).
- 3. Choose "Short Date" or "Long Date".
- 4. Excel will automatically convert the serial numbers to proper date format.

This method works if the numbers are **true Excel serial dates**. However, if the dates still appear as numbers after changing the format, it could mean:

- They are stored as text, not actual numbers.
- The data was imported incorrectly.

If Dates Are Stored as Text (Not Recognized by Excel)

If changing the format doesn't work, you may need to:

- 1. Use "Text to Columns" (Under "Data" tab):
 - Select the column \rightarrow Click Data \rightarrow Text to Columns \rightarrow Finish.
 - This forces Excel to recognize the values as dates.
- 5.5. Let's be a bit more specific this time in the type of solution that we would be looking for, and also upload the excel file so that ChatGPT can suggest an appropriate solution. Upload the Excel file (Choose + and then 'Upload from computer') and provide the following **prompt**:

"The excel file is attached. Please provide me the simplest solution to removing the special characters from the Source.Name column in a way that is repeatable as new data is added."

- 5.6. This has provided us with the cleaned data, which we can copy and paste into Excel but we would like it as an excel formula (as it suggests as an option) to actually make it repeatable rather than us having to revert back to ChatGPT each time! Let's ask for this: "Yes please provide this as an Excel formula"
- 5.7. It's given us quite a complex looking formula (to remove all alphanumeric characters) but let's try it in Excel and see if it fixes it:

R2	2 `	× .	$f_x \sim$ =TEXTJOIN("	',TRUE,IF(ISNUMBE	R(SEARCH(MI	<pre>D(A2,ROW(INDIRECT("1:"&LEN(A2))),1),</pre>				
"ABCDEFGHIJKLMNOPQRSTUVWXYZabcdefghijklmnopqrstuvwxyz0123456789 ")),MID(A2,ROW(IM								"&LEN(A2	2))),1),"	"))
	М	N	0	Р	Q	R	S	Т	U	V
1	Org Unit 🔄	Currency 💌	DR in Local CCY 🛛 💌	CR in Local CCY 👘 💌	Month 🛛 💌	Source cleaned 🗾 🔽				
2	Org 2	GBP	145.17	0.00	30/11/2020	*abr*ah*am J Willi*ams 20201130Invxlsx				
3	Org 2	GBP	0.00	82.96	30/11/2020	*abr*ah*am J Willi*ams 20201130Invxlsx				
4	Org 2	GBP	0.00	62.22	30/11/2020	*abr*ah*am J Willi*ams 20201130Invxlsx				
5	Org 2	GBP	146.00	0.00	31/12/2020	*abr*ah*am J Willi*ams 20201231Invxlsx				
6	Org 2	GBP	0.00	99.55	31/12/2020	*abr*ah*am J Willi*ams 20201231Invxlsx				
7	Org 2	GBP	0.00	46.46	31/12/2020	*abr*ah*am J Willi*ams 20201231Invxlsx				
8	Org 1	USD	18933.00	0.00	31/05/2020	Brend*a Bowden Showroom 20200531Invxlsx				
9	Org 1	USD	0.00	18.00	31/05/2020	Brend*a Bowden Showroom 20200531Invxlsx				
0	Org 1	USD	0.00	150.00	31/05/2020	Brend*a Bowden Showroom 20200531Invxlsx				
11	Org 1	USD	0.00	15.00	31/05/2020	Brend*a Bowden Showroom 20200531Invxlsx				
12	Org 1	USD	0.00	15.00	31/05/2020	Brend*a Bowden Showroom 20200531Invxlsx				
13	Org 1	USD	0.00	380.00	31/05/2020	Brend*a Bowden Showroom 20200531Invxlsx				
4	Org 1	USD	0.00	1140.00	31/05/2020	Brend*a Bowden Showroom 20200531Invxlsx				
15	Org 1	USD	0.00	190.00	31/05/2020	Brend*a Bowden Showroom 20200531Invxlsx				

(N.B: If ChatGPT didn't give you this, try it in Excel anyway):





=TEXTJOIN("", TRUE, IF(ISNUMBER(SEARCH(MID(A2, ROW(INDIRECT("1:" & LEN(A2))), 1), "ABCDEFGHIJKLMNOPQRSTUVWXYZabcdefghijklmnopqrstuvwxyz0123456789")), MID(A2, ROW(INDIRECT("1:" & LEN(A2))), 1), ""))

- 5.8. As you can see it has removed the / character but not the asterix. Let's try again and give it some hints as to what we want it to do (remove both the * and / characters) in a simple excel formula
 Prompt: "This has still got the * character in it. It seems that / and * are the only two special characters. Could you please write a simple excel formula that removes these characters from the text?"
- 5.9. This has now given us a simpler formula that seems to work

Formula (assuming "Source.Name" is in column A)

excel	ට් Copy	🗞 Edit					
=SUBSTITUTE(SUBSTITUTE(A2, "/", ""), "*", "")							
How It Works:							
1. SUBSTITUTE(A2, "/", "") → Removes all / characters.							

2. SUBSTITUTE(..., "*", "") \rightarrow Removes all * characters from the result.

S2	S2 \vee : $\times \checkmark f_x \vee$ =SUBSTITUTE(SUBSTITUTE(A2, "/", ""), "*", "")								
	R	S							
1	irce cleaned	Source cleaned_2							
2	pr*ah*am J Willi*ams 20201130Invxlsx	abraham J. Williams 20201130Inv.xlsx							
3	or*ah*am J Willi*ams 20201130Invxlsx	abraham J. Williams 20201130Inv.xlsx							
4	or*ah*am J Willi*ams 20201130Invxlsx	abraham J. Williams 20201130Inv.xlsx							
5	or*ah*am J Willi*ams 20201231Invxlsx	abraham J. Williams 20201231Inv.xlsx							
6	or*ah*am J Willi*ams 20201231Invxlsx	abraham J. Williams 20201231Inv.xlsx							
7	pr*ah*am J Willi*ams 20201231Invxlsx	abraham J. Williams 20201231Inv.xlsx							
8	nd*a Bowden Showroom 20200531Invxlsx	Brenda Bowden Showroom 20200531Inv.xlsx							
9	nd*a Bowden Showroom 20200531Invxlsx	Brenda Bowden Showroom 20200531Inv.xlsx							
10	nd*a Bowden Showroom 20200531Invxlsx	Brenda Bowden Showroom 20200531Inv.xlsx							
11	nd*a Bowden Showroom 20200531Invxlsx	Brenda Bowden Showroom 20200531Inv.xlsx							
12	nd*a Bowden Showroom 20200531Invxlsx	Brenda Bowden Showroom 20200531Inv.xlsx							
13	nd*a Bowden Showroom 20200531Invxlsx	Brenda Bowden Showroom 20200531Inv.xlsx							
1/	nd*a Bowdon Showroom 20200531 Involey	Branda Bowdan Showroom 20200521 Inv view							

5.10. All we actually want from this column is the customer name (not the date or "Inv.xlsx"). Let's provide a more focused prompt that tells ChatGPT how the text is structured and what we are looking for (despite uploading the file it doesn't properly read the file and understand its contents unfortunately!)



Try the **prompt**: "The Source.Name column contains the Customer name followed by the date in YYYYMMDD format followed by the text "Inv.xlsx". The customer name also contains some special characters - / and * - which need to be removed. Please provide a simple Excel formula which returns the cleaned customer name from this column."

5.11. Now that we've been more specific and given it instructions on what we want, it seems to have produced a simple formula that extracts the customer name correctly. It also provides an explanation of how the formula works so if we think that future data wouldn't conform to this exact pattern we could ask it to adjust it!

To extract the **cleaned customer name** from the **"Source.Name**" column while removing **//** and *****, you can use the following Excel formula:

Formula (assuming "Source.Name" is in column A)

excel	ට් Copy	🗞 Edit
=SUBSTITUTE(SUBSTITUTE(LEFT(A2, FIND("20", A2)-1), "/", ""), "*", "")		

Explanation:

FIND("20", A2)-1 → Identifies the position where the date (YYYYMMDD) starts. This assumes the date always starts with "20".

T2	T2 \rightarrow : $\times \checkmark f_x \rightarrow$ =SUBSTITUTE(SUBSTITUTE(LEFT(A2, FIND("20", A2)-1), "/", ""), "*", "")								
	R	S	Т						
1	ırce cleaned 🛛 🚽	Source cleaned_2	Customer Name 🗾						
2	pr*ah*am J Willi*ams 20201130Invxlsx	abraham J. Williams 20201130Inv.xlsx	abraham J. Williams						
3	or*ah*am J Willi*ams 20201130Invxlsx	abraham J. Williams 20201130Inv.xlsx	abraham J. Williams						
4	or*ah*am J Willi*ams 20201130Invxlsx	abraham J. Williams 20201130Inv.xlsx	abraham J. Williams						
5	pr*ah*am J Willi*ams 20201231Invxlsx	abraham J. Williams 20201231Inv.xlsx	abraham J. Williams						
6	or*ah*am J Willi*ams 20201231Invxlsx	abraham J. Williams 20201231Inv.xlsx	abraham J. Williams						
7	pr*ah*am J Willi*ams 20201231Invxlsx	abraham J. Williams 20201231Inv.xlsx	abraham J. Williams						
8	nd*a Bowden Showroom 20200531Invxlsx	Brenda Bowden Showroom 20200531Inv.xlsx	Brenda Bowden Showroom						
9	nd*a Bowden Showroom 20200531Invxlsx	Brenda Bowden Showroom 20200531Inv.xlsx	Brenda Bowden Showroom						
10	nd*a Bowden Showroom 20200531Invxlsx	Brenda Bowden Showroom 20200531Inv.xlsx	Brenda Bowden Showroom						
11	nd*a Bowden Showroom 20200531Invxlsx	Brenda Bowden Showroom 20200531Inv.xlsx	Brenda Bowden Showroom						
12	nd*a Bowden Showroom 20200531Invxlsx	Brenda Bowden Showroom 20200531Inv.xlsx	Brenda Bowden Showroom						
13	nd*a Bowden Showroom 20200531Invxlsx	Brenda Bowden Showroom 20200531Inv.xlsx	Brenda Bowden Showroom						
14	nd*a Bowden Showroom 20200531Invxlsx	Brenda Bowden Showroom 20200531Inv.xlsx	Brenda Bowden Showroom						
15	nd*a Bowden Showroom 20200531Invxlsx	Brenda Bowden Showroom 20200531Inv.xlsx	Brenda Bowden Showroom						
16	nd*a Bowden Showroom 20200531Invxlsx	Brenda Bowden Showroom 20200531Inv.xlsx	Brenda Bowden Showroom						

This file with the Customer Name column is saved as Invoice Data_Cleaned.xlsx

- 6. Let's try something more complex! Provide it the following prompt: "There are columns "DR in Local CCY" and "CR in Local CCY" which represent the debit and credit amounts in local currency. I need to convert these into a common currency (USD) using the Exchange Rates worksheet. Can you write me a lookup formula to do this?"
 - 6.1. It has made some assumptions about how the data is structured and provided some different formulas based on your Excel version.



M Power Platform Finance



- Looks up the exchange rate for the currency in column ${\bf A}$ of the "Exchange Rates" sheet.
- Fetches the corresponding rate from column B.
- The FALSE ensures an exact match.
- The / divides the local currency amount by the exchange rate to get the value in USD.

If You're Using Excel 365 or Excel 2019 (XLOOKUP Alternative):

For better flexibility, you can use **XLOOKUP** instead of **VLOOKUP**:

⊔∘ Сору	🕫 Edit
🗗 Сору	🕫 Edit

- 6.2. It hasn't picked up the right column references but even when correcting for this when copying the formula, it hasn't performed the currency conversion correctly because our data is structured very differently to how it had assumed!

	Ν	0	Р	Q	R	S
-	Org Unit	Currency	DR in Local CCY 🛛 💌	CR in Local CCY 🗾 💌	Month 🗾	Column1 🗾 💌
	Org 2	GBP	145.1747044	0	30/11/2020	145.1747044
	Org 2	GBP	0	82.95697395	30/11/2020	0
	Org 2	GBP	0	62.21773046	30/11/2020	0
	Org 2	GBP	146.0042742	0	31/12/2020	146.0042742
	Org 2	GBP	0	99.54836874	31/12/2020	0
	Org 2	GBP	0	46.45590541	31/12/2020	0
	Org 1	USD	18933	0	31/05/2020	18933
	Org 1	USD	0	18	31/05/2020	0
	Org 1	USD	0	150	31/05/2020	0
	Org 1	USD	0	15	31/05/2020	0
	Org 1	USD	0	15	31/05/2020	0
	Org 1	USD	0	380	31/05/2020	0
	01	LICD	0	1140	21/05/0000	0

- 6.3. Provide the following prompt: "This hasn't worked because the data is structured differently to what has been assumed. Use the file that was attached to give me a formula in the invoices table that looks up the exchange rate in the Exchange Rates sheet for the corresponding date and currency."
- 6.4. Try the formula it gives you. In my case, it appeared to have correctly analysed the data structure and given us formulas to get the exchange rates but the column references are all wrong so the formula still doesn't work!





L	M	N	0	Р	Q	R	S
ReportBasis 🛛 💌	Mapped Name	💌 Org Unit 📘	Currency 🔽	DR in Local CCY 🛛 💌	CR in Local CCY 🗾 💌	Month 🗾 💌	Columr 💌
TRUE	Abraham J. Williams	Org 2	GBP	145.1747044	0	30/11/ 🕂 20	#VALUE!
TRUE	Abraham J. Williams	Org 2	GBP	0	82.95697395	30/11/2020	#VALUE!
TRUE	Abraham J. Williams	Org 2	GBP	0	62.21773046	30/11/2020	#VALUE!
TRUE	Abraham J. Williams	Org 2	GBP	146.0042742	0	31/12/2020	#VALUE!
TRUE	Abraham J. Williams	Org 2	GBP	0	99.54836874	31/12/2020	#VALUE!
TRUE	Abraham J. Williams	Org 2	GBP	0	46.45590541	31/12/2020	#VALUE!
TRUE	Brenda Bowden Showroom	Org 1	USD	18933	0	31/05/2020	#VALUE!
TRUE	Brenda Bowden Showroom	Org 1	USD	0	18	31/05/2020	#VALUE!
TRUE	Brenda Bowden Showroom	Org 1	USD	0	150	31/05/2020	#VALUE!
TRUE	Brenda Bowden Showroom	Org 1	USD	0	15	31/05/2020	#VALUE!
TRUE	Brenda Bowden Showroom	Org 1	USD	0	15	31/05/2020	#VALUE!
TRUE	Brenda Bowden Showroom	Org 1	USD	0	380	31/05/2020	#VALUE!
TRUE	Brenda Bowden Showroom	Org 1	USD	0	1140	31/05/2020	#VALUE!
TRUE	Brenda Bowden Showroom	Org 1	USD	0	190	31/05/2020	#VALUE!

- 6.5. If the formula doesn't work, let's ask ChatGPT to correct it. Try the following prompt: "This formula isn't working because the column references appear to be wrong. Please provide updated formulas that refer to the correct columns or that use table references (since both the Invoices and Exchange rates data is structured as a table).
- 6.6. It didn't pick up the correct name of the Exchange rates table so we pass this in as a prompt: "The name of the exchange rates table is called ExchangeRates". It then gives us a formula which works but it still returns the incorrect results

=[@[DR in Loca]	L CCY]] / ING	DEX(Exchange	Rates[[EUR]:[AUD]]	, MATCH([@Date],	ExchangeRa	tes[Date],	0), MAT	CH([@Curi	rency],	Exchange	Rates[#H	eaders],	0))
M	N	0	Р	Q	R	S	Т	U	V	W	Х	Y	
	💌 Org Unit	Currency	🖞 DR in Local CCY 🛛 💌 (CR in Local CCY 🗾 💌	Month 🗾	Column1 💌							
liams	Org 2	GBP	145.1747044	0	30/11/2020	208.610122							
liams	Org 2	GBP	0	82.95697395	30/11/2020	0							
liams	Org 2	GBP	0	62.21773046	30/11/2020	0							
liams	Org 2	GBP	146.0042742	0	31/12/2020	209.8021798							
liams	Org 2	GBP	0	99.54836874	31/12/2020	0							
liams	Org 2	GBP	0	46.45590541	31/12/2020	0							
h Showroom	Org 1	USD	18933	0	31/05/2020	#N/A							
1 Showroom	Org 1	USD	0	18	31/05/2020	#N/A							
h Showroom	Org 1	USD	0	150	31/05/2020	#N/A							
1 Showroom	Org 1	USD	0	15	31/05/2020	#N/A							
h Showroom	Org 1	USD	0	15	31/05/2020	#N/A							
1 Showroom	Org 1	USD	0	380	31/05/2020	#N/A							
h Showroom	Org 1	USD	0	1140	31/05/2020	#N/A							
1 Showroom	Org 1	USD	0	190	31/05/2020	#N/A							

6.7. This is partly because it can't find an exchange rate for USD (this is always just 1 obviously!) and it's written the formula wrong. Let's try and teach it how to write an INDEX/MATCH formula! Provide the following prompt:

"This doesn't work for the following reasons:

🎕 Learn Data Insights

1) Some of the invoices are already in USD for which no exchange rate exists

2) It is looking up against the wrong columns - the result returned from MATCH against the headers of the table does not work when INDEX is looking only at the EUR:AUD columns rather than the whole table

6.8. Apart from the fact that the exchange rates are the other way around (it is USD per unit so we need to multiply the DR/CR amounts rather than divide them), the formulas now appear to work but we've likely lost all confidence in the ability of ChatGPT to help here! Let's try an alternative approach. Open a new chat, upload the excel file and we'll use the prompting guide from Microsoft learn to write a detailed prompt.







Try the following prompt:

"You are a financial analyst who has been given the attached set of Invoices in Excel and need to understand the trends in revenue over time by customer and organisation unit. The first challenge it that there are columns "DR in Local CCY" and "CR in Local CCY" which represent the debit and credit amounts in local currency. I need to convert these into a common currency (USD) using the Exchange Rates worksheet. Can you write me a lookup formula to do this?

Please consider the following instructions carefully:

- The data is structured as tables so use Table references rather than static cell references

- The exchange rates table does not contain a currency column – each currency is in a separate column for a given date

-Don't hardcode the currency values in the formula – I want this to work if additional currencies are added into the Exchange Rate table

- Use the date and currency columns in the invoices table to look up the appropriate exchange rate.

- Ensure you reference the names of the tables – Invoices and ExchangeRates – as appropriate

-Some of the invoice values are already in USD so should be taken as is

- Provide formulas that are simple to understand and provide explanations for how the formulas work
 - Validate the results to ensure that the correct formulas are being used
- Note that the exchange rates are USD per unit of currency.
- 6.9. The response provided by ChatGPT suggests that it has now taken the request more seriously and provided more considered answers before just blurting out the "first thing that comes to it's head" (!!!). It has even offered to apply the formulas directly to the



data in the files

Explanation:

- 1. XLOOKUP([@Date], ExchangeRates[Date], ExchangeRates[@[Currency]])
 - Finds the exchange rate for the given invoice date in the correct currency column.
- 2. Multiply by DR in Local CCY or CR in Local CCY
 - Converts the value from local currency to USD.

Validation Steps:

- Check that the formulas pull the correct exchange rate for each date.
- Verify that multiplication correctly converts the values.
- Ensure no missing or incorrect matches.

Would you like me to apply these calculations to the dataset and display the results? [-]

- 6.10. But again it is important to check this in this specific instance of running ChatGPT, it didn't actually take into account the invoice date in the formula and didn't check the structure of the exchange rates table (it assumed there was a currency column without checking) so needed to be re-prompted to do this!
- 6.11. The file it gave to download contained just a single sheet with all values calculated so we ask "These have values already calculated, can you provide me with the original excel file just with the additional columns added as formulas?"

To provide the Excel file with the formulas intact, I will modify the original file by adding the formulabased columns directly into the **Invoices** table.

I'll ensure that:

- DR in USD and CR in USD columns use Excel formulas referencing the ExchangeRates table.
- The formulas use structured table references (XLOOKUP) to dynamically pull exchange rates.

I'll now update the file and provide it for download.

You can download the updated Excel file with formulas for USD conversion from the link below:

Download Invoices_with_Formula.xlsx

This file contains the new **DR in USD** and **CR in USD** columns with **XLOOKUP formulas**, ensuring dynamic exchange rate lookup and conversion.

Let me know if you need any further modifications! [>-]

6.12. For me at least this didn't work (the file it gave had the reporting currency values incorrectly hardcoded!) so perhaps we do need to just add the formulas provided ourselves. (ChatGPT isn't really designed to directly edit/update Excel files – that's why we might want to consider Copilot for Excel!). I also tried regenerating the answer with



🛍 Power Platform Finance

o3 mini. The response provided wasn't necessarily more reasoned and the file output was almost the same for this example.

Next Steps

I'll now add two new columns to the Invoices table:

- DR in USD (Calculated as: DR in Local CCY * Exchange Rate)
- CR in USD (Calculated as: CR in Local CCY * Exchange Rate)

I'll match the currency in **Invoices** with the corresponding column in **Exchange Rates**, using the most recent exchange rate available. [>-]

The updated Excel file with the added **DR in USD** and **CR in USD** columns is ready. You can download it using the link below:

🛓 Download Updated Invoice Data

Let me know if you need any modifications! [>-]

- 7. Maybe we're "insulting" the intelligence of Gen AI by asking it to write lookup formulas for us
 ..let's now pass in a file with the amounts in USD already populated
 (InvoiceData_WithConvertedCurrencies.xlsx) and give it some more advanced challenges of analysing this data!
 - 7.1. This time, let's save the back and forth with clarifying what we want and be a bit more specific with our prompt. We'll use the CREATE framework to do this.

C D E A T E MODEL

G.N.E.A.T.E. MODEL
C for Character
R for Request for Information
E for Examples
A for Adjustments
T for Type of Output
E for Extras

Try the following Prompt:

"c

7.2. This looks promising – it's given us a graph to highlight the key trends and says it has produced a pivot table in Excel with the drill down capability!



🎕 Learn Data Insights

RISHI SAPRA



I have provided a pivot table summarizing revenue by customer and month, along with a line chart visualizing the revenue trends over time.

Next, I will create an interactive Excel file where you can drill down into monthly totals to see a breakdown by day and account. Stay tuned. [--]

I have generated an Excel file containing the revenue analysis, including a pivot table and the raw data for further exploration. You can download it using the link below:

Download Revenue Analysis Excel File

This file allows you to drill down on specific monthly totals to see a breakdown by day and account within Excel. Let me know if you need any modifications! [--]

7.3. Unfortunately it didn't generate a pivot table (ChatGPT was a bit overambitious with this one!) – when prompted it explained why it's unable to generate an actual pivot table and gave me instructions to create one myself.





It looks like xlsxwriter doesn't support creating actual pivot tables programmatically. However, I will create an Excel file with the raw data and a defined pivot table structure using openpyx1, so you can refresh and interact with it directly in Excel. Let me generate that now. [>-]

It seems that <code>openpyx1</code> does not support programmatically creating pivot tables. However, I can generate an Excel file that contains the raw data, and I will add a Pivot Table structure that you can refresh in Excel to interact with it dynamically. Let me create that now. [6-]

Since xlsxwriter does not support creating actual pivot tables, I will take a different approach:

- I will generate an Excel file containing:
 - 1. Raw Data The invoice details, including dates, customers, and credit amounts.
 - 2. Pivot Table Sheet A pre-configured area where you can easily create a pivot table in Excel.

I will provide step-by-step instructions on how to add the pivot table manually once you open the file in Excel. Let me generate the correct file now. [6-]

I have created an Excel file containing the raw invoice data and a sheet with step-by-step instructions on how to create a Pivot Table in Excel.

You can download it here: Download Revenue Analysis with Pivot Instructions

Once you open the file, follow the instructions in the "Instructions" sheet to create an interactive Pivot Table that allows drilling down into monthly totals for further analysis. Let me know if you need any refinements! [-]

7.4. This data looks interesting! Let's ask for some more advanced analysis that may not be obvious how we would do ourselves, even if we are quite familiar with Excel

Again using the C.R.E.A.T.E methodology, try the following prompt:

"OK thanks the revenue (CR in USD Amounts) seems very volatile especially for certain customers. Management have asked for the reasons behind this. For each customer can you please calculate and explain the following:

-The actual volatility score across months/years for the periods where there is a positive revenue

- Whether there is a correlation between the volatility (and total revenue) and specific income accounts

- Whether there is any seasonality overall and/or for specific customers, both within a month across months/within a year

Please produce outputs which are the most suitable for this, though please make any insights clear e.g. by using visuals such as a scatter plot and tables with conditional formatting that indicate trends/patterns in the data"

7.5. This time even with the 40 model it took a long time to think! And returned some quite detailed outputs with charts to support them:

Customer Revenue Volatility





Mapped Name	Total Revenue	Volatility Score
Jimenez Services	142074.46176325783	939.9768186173017
Kirby Freeman Wholesale Distribution	112628.39503566966	472.5521368011436
Peacock, Karen	190334.4	617.5924132605392
Tingey Worldwide Freight Company	92689.91950527289	1333.9220245750091
Weeks, Matt	46435.53463724062	331.91113381060876
Willams, Cindy	86652.9841810836	815.6277444231949

Volatility vs. Total Revenue by Customer



I have provided a table showing the **Total Revenue and Volatility Score** for each customer, along with a scatter plot illustrating the relationship between **revenue and volatility**.

Key Insights:

- **Higher volatility generally corresponds with higher total revenue**, indicating that larger customers experience more fluctuation in revenue.
- Some customers have high revenue but relatively low volatility, suggesting stable, consistent revenue streams.
- Customers with low revenue and high volatility likely represent sporadic or seasonal business relationships.

Next, I will analyze:

- 1. Correlation between volatility, total revenue, and specific income accounts.
- 2. Seasonality trends (monthly and yearly patterns) for revenue overall and by customer.

Stay tuned for further insights.





4100 Legal Fee Income	4200 Settlement Income	4500 Markup on Reimbursable Expenses	4700 Reimbursed Exp. (Income)
0.9919578806874902	0.0893608856097043 3	0.8895409138293561	0.8846466304056385
0.5545384258333163	0.683931721037834	0.5166696957464184	0.4972389250809933
1.0	- 0.0372900067336100 9	0.8815393535130865	0.8776435906904001
- 0.0372900067336100 9	1.0	0.0657302879178868 6	0.05710991983710815 6
0.8815393535130865	0.0657302879178868 6	1.0	0.9971903421441899

Correlation between Revenue, Volatility, and Accounts

Correlation between Revenue, Volatility, and Income Accounts



I have provided a **correlation matrix** showing how revenue volatility and total revenue relate to specific **income accounts**. The heatmap visually highlights strong correlations.

Key Insights:



- Certain income accounts have a strong correlation with volatility, indicating they contribute significantly to revenue fluctuations.
- If some accounts correlate positively with volatility, it may suggest they drive unpredictable revenue streams.
- Conversely, negatively correlated accounts may indicate stable, recurring revenue sources.

Next, I will analyze **seasonality trends**, both across months and within a year, to determine if there are predictable revenue patterns.

Monthly Revenue Seasonality

Month Abraham J. Williams Brenda Bowden Showroom Hamby, Shane

1	0.0	0.0	3468.939131828725
2	0.0	0.0	13740.886864465025
3	0.0	0.0	24108.79676843839
4	0.0	0.0	14432.77891539386
5	0.0	18933.0	0.0
6	0.0	22966.2	0.0

Overall Monthly Seasonality in Revenue



Seasonality Trends by Customer







I have provided:

- 1. A table of monthly revenue seasonality by customer.
- 2. A line chart showing overall revenue trends across months.
- 3. A seasonality trend analysis for individual customers.

Key Insights:

- **Overall revenue trends**: Certain months may show significant peaks and dips, indicating seasonality.
- **Customer-specific trends**: Some customers display clear patterns of increased activity in specific months.
- **Potential cyclicality**: If revenue spikes occur at regular intervals, management can plan accordingly.
- 7.6. Wow! I've been working in data for over a decade and would've struggled to even conceptually come up with this analysis if asked! I'm not sure I'd be able to just present this back to management/my client without being able to explain how the scores are calculated though. Let's ask for this memo that we can present to management.

Example prompt:

"I need to produce a one page memo (no more than 1000 words) that I can use as the basis for my next meeting with management. Please use the following structure for this document:

1) Overview of revenue performance, volatility and seasonality. This should include the key insights around performance over time (including which types of income) explained



to a level of detail that provides credibility to the analysis but should be understandable to a non-technical audience.

2) Customer analysis. Please highlight which customers have given stable revenues over the period, versus those with one off purchases or where purchasing patterns have been volatile. This should be supported with calculations such as volatility scores but they need to be explained in plain English in a way that is understandable by a non-technical audience.

3) Recommendations. Please indicate which customers we should focus our efforts on either to make more regular/consistent purchases or to generate different types of income from in order to reduce volatility.

Use bullet points and formatting (e.g. bold/underline, spacing) to make this digestible "

- 7.7. The first version of this (in both 4o and o3-mini-high) produced a nicely worded and formatted memo, but didn't contain any specific customer names or values (it used placeholders like [insert volatile customers] or \$xxx). Additional prompts:
 - 7.7.1. "This is generic with no actual numbers/calculations from the document to support the narrative. Please modify this"
 - 7.7.2. "Why are there placeholders for the numbers/customer names rather than actual values?"

The output of the memo [40 produced this in ChatGPT Canvas] in both 40 and 03-mini is saved down as a separate document in this folder. It's pretty impressive!!

7.8. Lastly, we'd like it to be repeatable so as we get new data in Excel we can have it updated automatically. It clearly uses python quite heavily to do this data analytics, so let's ask if it can generate the python code that I can insert into Excel

Prompt: "I have access to Python in Excel so want to produce some analysis worksheets with Python based analysis that will automatically update as new data is added. Please produce the python code that I can insert into Excel that includes:

-Overview of revenue performance over time and by customer/account.

-A volatility score across months/years for the periods where there is a positive revenue

- Whether there is a correlation between the volatility (and total revenue) and specific income accounts

- Whether there is any seasonality overall and/or for specific customers, both within a month across months/within a year

Please produce outputs which are the most suitable for this, though please make any insights clear e.g. by using visuals such as a scatter plot and tables with conditional formatting that indicate trends/patterns in the data





🛍 Power Platform Finance



Please also include detailed comments in the code to explain what each line is doing (I'm not at all familiar with Python so need to be able to justify this if asked!) and include a dynamic narrative similar to what was produced in the memo that explains the insights and provides some detail as to how values have been calculated."

7.9. The python code generated (along with that generated by Excel copilot for a comparison) is saved as a separate document in this folder. Just copying and pasting this massive code block exactly as is into an excel Python cell is unlikely to work in the first instance. However, you can copy and paste the code into separate dataframes – e.g. getting it to first just return the invoice data as a dataframe in python and then referencing this data frame in different blocks of analysis (Rather than trying to do everything in one code block) is probably a more sensible approach to take here!



