



# Data Challenge

---

Engineering and Analytics

# Welcome

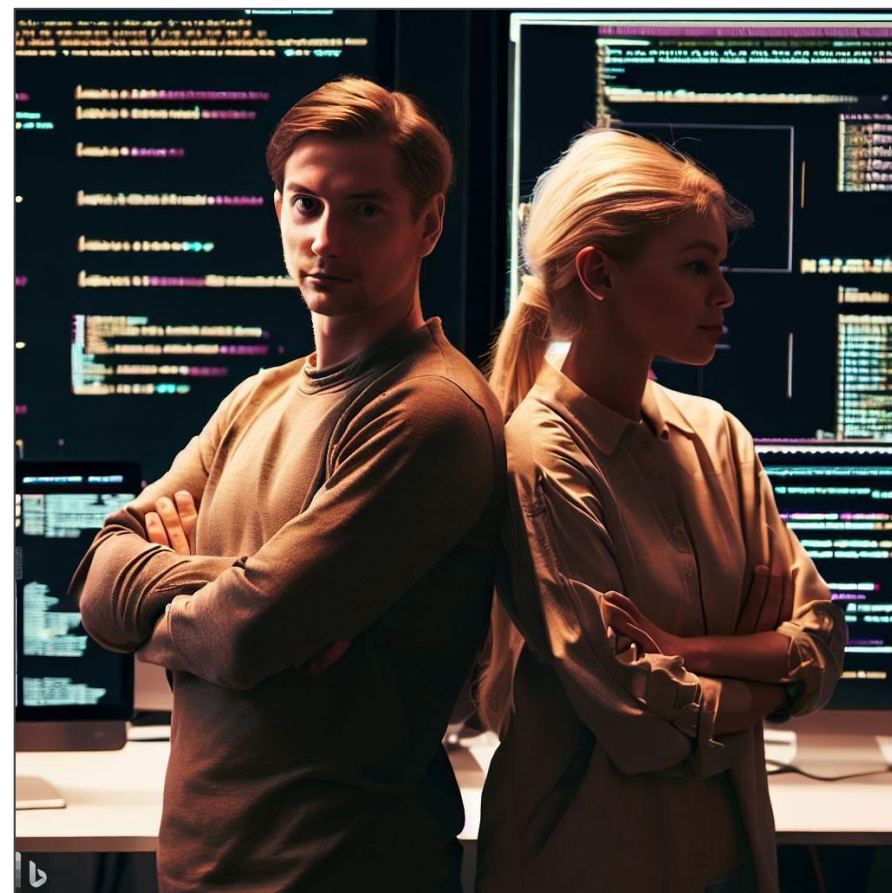
---

Thank you for signing up for the **Argusa Data Challenge!**

Our goal was to create a fun event for you to test and showcase your skills in **data engineering, analytics** and **visualization**.

We are passionate about **data** and want you to experience a bit of the work we do every day.

**We hope you enjoy the Challenge!**

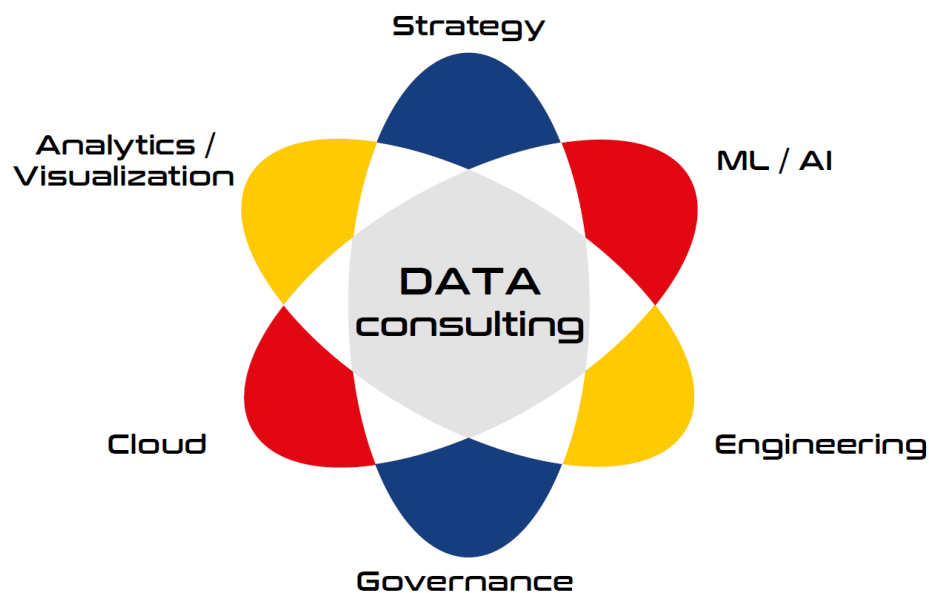


# So... who are we?

**Argusa** was founded in Lausanne in 2018

We are a diverse team of 12 consultants

We help our clients with anything and everything **data**!



Cédric



Fatima



Luca



Joana



Matthieu



Lino



Gil



Tyler



Dalal



Konie



Vincenzo



Dalil



# Code of Conduct

---

- Participants must treat each other with respect and dignity. They must not insult, demean or harass any other participant or organizer. Please do not disturb other people in the mixed-use areas. Keep discussions constructive. Argusa reserves the right to remove anyone behaving in an unacceptable manner.
- Participants agree to be recorded and photographed by the organizers and consent to the publication of the material by Argusa on social media (e.g., LinkedIn).
- Argusa is not responsible for accidents, losses or theft occurring during the Data Challenge.
- Argusa respects your intellectual property. While Argusa makes no claims on the work you create, the data provided should only be used for educational purposes and not for commercial gain.

# Agenda

## MORNING

**08:30** Check-in and Croissants *(1h)*

**09:30** Presenting the challenge *(30m)*

**10:00** Working on the challenge *(3h)*

**13:00** Pizza lunch and pitch  
preparation *(1h30m)*

## AFTERNOON

**14:30** Session 1 Pitch delivery *(1h)*

**15:30** Coffee break *(30m)*

**16:00** Session 2 Pitch delivery *(1h)*

**17:00** Reviewing results *(30m)*

**17:30** Winners and prizes *(30m)*

**18:00** Apéro

# Logistics

## WIFI

Connect to **EIP** - click «*Receive your credentials by SMS*» and fill out the registration page

If EIP does not work, you can try one of the **iSpot\_Connect\_XXXX** networks - passwords will be on the whiteboard

## FOOD & DRINKS

Water, soft drinks, coffee and tea will be available throughout the day

We will have pizzas for lunch, sweet treats in the afternoon and a small apéro at the end of the day

## ROOM

Room Uranus will be available all day for participants to work on the Challenge

You can drink in the room but please use the reserved tables in the hall if you are consuming food

# The Argusa Data Challenge

## SETUP

You are Data Consultants for Argifts, a large online retailer

The CEO of Argifts wants to evaluate the performance of the organization

You need to take Argifts' data and create **compelling visualizations** that answer the CEO's questions

You can use any tools you like to transform the data any way you want and create whichever plots, charts, graphs or dashboards you prefer

You have **3 hours** to complete the challenge

# The Argusa Data Challenge

## RESULTS

Once you have completed the challenge, you must **send us your results**

Email [datachallenge@argusa.ch](mailto:datachallenge@argusa.ch)

**Attach the visualizations** you prepared in image or pdf format

Send us the code you used to produce the results

The results we receive by email are **final** and cannot change!



# The Argusa Data Challenge

## PITCH

You will have **5 minutes** to present your results in front of the 3 members of the jury.

The presentations will happen in the afternoon. The order will be determined randomly.  
You can use the lunch break to prepare your pitch.

The jury will score during the presentation, so you **must pitch** your results to stay in the competition.

You **must answer the questions** you were asked during your pitch. Consult the score sheet to know exactly what you should present.

You will be judged based on whether or not you answer the questions correctly, but also based on your ability to answer the questions using a **graphical representation**.

# The Argusa Data Challenge

## SCORES AND PRIZES

You can score a maximum of 40 points in the Challenge

Consult the score sheet for more details

**Prizes** will be given to the three teams with the highest number of points

In case of a **tie**, we will use the **timestamp** of the e-mail with the results to find the fastest team



# The Argusa Data Challenge

## THE AFTERMATH

Cash prizes and Argusa swag will be given today to the top three teams

The top two teams will be contacted directly to schedule job interviews with Argusa next week

All teams will receive their digital badges by Friday

We will also send an exit survey by Friday and we would appreciate your feedback!

# The Argifs Dataset

- Download the data from [tinyurl.com/argusadata](https://tinyurl.com/argusadata) (or use the QR code below)
- Extract 8 csv files from ArgusaChallenge.zip



ArgusaChallenge2023.zip			
Name	Type	Size	Compressed size
clients.csv	Microsoft Excel Comma Separated Values File	2'399 KB	934 KB
orders_2020.csv	Microsoft Excel Comma Separated Values File	771 KB	266 KB
orders_2021.csv	Microsoft Excel Comma Separated Values File	1'681 KB	620 KB
orders_2022.csv	Microsoft Excel Comma Separated Values File	1'637 KB	611 KB
orders_2023.csv	Microsoft Excel Comma Separated Values File	1'094 KB	389 KB
products.csv	Microsoft Excel Comma Separated Values File	4'315 KB	113 KB
ratings.csv	Microsoft Excel Comma Separated Values File	697 KB	235 KB
returns.csv	Microsoft Excel Comma Separated Values File	3'667 KB	1'041 KB

# orders\_YYYY.csv

## ARGIFTS DATASET

- Orders dataset indicating when the order was placed, the products purchased and the client who made the purchase
- **Notice the two highlighted rows are identical:**
  - This indicates the product was purchased 2 times in the same order by the same client
  - Make sure you calculate the *purchased quantity* for each order and product as the *number of rows*

	A	B	C	D
1	client_id	product_id	order_id	order_date
1772	v4-52229771	47	O-078G10708	14/02/2021
1773	x8-2901393G	47	O-424DE5448	14/02/2021
1774	f1-0505873p	47	O-529Ri3554	17/02/2021
1775	l0-8286945U	47	O-426Dk380	19/02/2021
1776	s4-1071745J	47	O-011OR7295	19/02/2021
1777	s4-1071745J	47	O-011OR7295	19/02/2021
1778	F2-3384321K	47	O-566Vg7114	20/02/2021
1779	H2-1292829g	47	O-553Eb5061	20/02/2021
1780	l4-0067823H	47	O-579IX9524	22/02/2021
1781	V8-2519489X	47	O-473la9769	26/02/2021
1782	j6-861831l	47	O-099TE3901	01/03/2021
1783	b0-3455175B	47	O-560dA4073	02/03/2021
1784	b7-2813284P	47	O-930vE7999	02/03/2021
1785	U3-8689088U	47	O-710CU4966	03/03/2021
1786	t2-1229779q	47	O-471LU8377	04/03/2021
1787	G6-169153O	47	O-637aw1398	05/03/2021

# products.csv

## ARGIFTS DATASET

- Detailed information about the products, including name, categories and base price
- Notice the highlighted rows:**
  - The sale price of a product depends on the order date (in the orders dataset)
    - The order date must be between the start date and the end date
  - An empty end date indicates the current sale price of the product
    - Consider replacing empty end dates by today's date

	A	B	C	D	E	F	G	H	I	J	K	L
1	category_1	category_2	category_3	category_4	category_5	name	description	base_price	start_date	end_date	sale_price	id
2591	Bricolage + jardin	Jardin	Horticulture	Fontaine		Climaqua Liv S., Fontaine couverte	Apporte calme et s'@r@nit@ dans votre maison et c	139	14/01/2023	08/02/2023	161.42	54
2592	Bricolage + jardin	Jardin	Horticulture	Fontaine		Climaqua Liv S., Fontaine couverte	Apporte calme et s'@r@nit@ dans votre maison et c	139	09/02/2023	24/02/2023	160.11	54
2593	Bricolage + jardin	Jardin	Horticulture	Fontaine		Climaqua Liv S., Fontaine couverte	Apporte calme et s'@r@nit@ dans votre maison et c	139	28/03/2023	01/05/2023	169.36	54
2594	Bricolage + jardin	Jardin	Horticulture	Fontaine		Climaqua Liv S., Fontaine couverte	Apporte calme et s'@r@nit@ dans votre maison et c	139	02/05/2023	29/05/2023	176.85	54
2595	Bricolage + jardin	Jardin	Horticulture	Fontaine		Climaqua Liv S., Fontaine couverte	Apporte calme et s'@r@nit@ dans votre maison et c	139	30/05/2023	30/06/2023	172.87	54
2596	Bricolage + jardin	Jardin	Horticulture	Fontaine		Climaqua Liv S., Fontaine couverte	Apporte calme et s'@r@nit@ dans votre maison et c	139	17/07/2023	13/08/2023	180.98	54
2597	Bricolage + jardin	Jardin	Horticulture	Fontaine		Climaqua Liv S., Fontaine couverte	Apporte calme et s'@r@nit@ dans votre maison et c	139	14/08/2023		171.28	54
2598	Bricolage + jardin	Jardin	Horticulture	Fontaine		Infactory Village de montagne, Fontaine couverte	Faites le plein d'@nergie et prenez une pause bienfais	68.95	10/01/2020	13/02/2020	67.97	55
2599	Bricolage + jardin	Jardin	Horticulture	Fontaine		Infactory Village de montagne, Fontaine couverte	Faites le plein d'@nergie et prenez une pause bienfais	68.95	14/02/2020	04/03/2020	67.15	55
2600	Bricolage + jardin	Jardin	Horticulture	Fontaine		Infactory Village de montagne, Fontaine couverte	Faites le plein d'@nergie et prenez une pause bienfais	68.95	25/03/2020	06/04/2020	65.9	55
2601	Bricolage + jardin	Jardin	Horticulture	Fontaine		Infactory Village de montagne, Fontaine couverte	Faites le plein d'@nergie et prenez une pause bienfais	68.95	07/04/2020	10/05/2020	65.32	55

# returns.csv

## ARGIFTS DATASET

- Return dates and return reasons for different products in different orders
- Notice the highlighted rows have the same order ID and product ID:**
  - This indicates two items of the same product were returned
  - Make sure you calculate the *returned quantity* for each order and products as the *number of rows*

	A	B	C	D	E
1	order_id	product_id	return_date	reason	return_received
10877	O-952sQ839	47	21/10/2020	Praesentium illum earum voluptatum.	27/10/2020
10878	O-788Ci1983	47	31/12/2022	Delectus suscipit tenetur illum enim	06/01/2023
10879	O-011OR7295	47	22/03/2021	Architecto error cum quisquam magna	28/03/2021
10880	O-011OR7295	47	06/04/2021	Aut molestias consectetur aut	08/04/2021
10881	O-329Rw5798	47	13/12/2022	Repudiandae optio mollitia ipsa volupt	15/12/2022
10882	O-863kF4072	47	23/08/2023	Id harum sapiente saepe fugit. Rem ut	26/08/2023
10883	O-099OA7471	47	01/03/2022	Pariatur illum alias possimus voluptas	07/03/2022
10884	O-657YR5921	47	11/05/2022	Doloremque eaque voluptatum eius	12/05/2022
10885	O-685eF3059	47	25/04/2022	Expedita error pariatur omnis sed fugia	26/04/2022
10886	O-879UE2571	47	25/01/2022	Quo hic praesentium accusamus vero	30/01/2022
10887	O-355IY9488	47	25/04/2021	Voluptas eum deleniti repudiandae qu	27/04/2021
10888	O-969SY2280	47	04/04/2021	Adipisci laudantium illo amet.	10/04/2021
10889	O-985ae1757	47	01/10/2020	Repudiandae quas quis magni. At quo	05/10/2020
10890	O-059Wd7346	47	30/11/2022	Eius repellat vitae nobis. A error	01/12/2022
10891	O-824ud4381	47	03/10/2022	Suscipit amet consectetur optio enim	06/10/2022
10892	O-038nW9561	47	10/03/2023	Nulla accusantium in occaecati	15/03/2023
10893	O-038nW9561	56	22/01/2023	Porro nulla similique. Vel eum rerum d	26/01/2023

# clients.csv

## ARGIFTS DATASET

- Names, address, contact information and other details for each client

	A	B	C	D	E	F	G	H	I	J	K
1	client_id	first_name	last_name	date_of_birth	phone	email	first_purchase	address_street	address_city	address_postcode	CANTON
2	k3-3096581b	Clara	Bujard	15/03/1995	+41 67 249 07 28	c_bujard962@hispeed.ch	01/01/2020	MÃ¼hleweg 57	Murten	3280	FR
3	A4-819583b	François	DelÃ©ze	05/03/1983	+41 50 739 35 59	f.deleze926@sunrise.ch	01/01/2020	Avenue de ChÃ¢teline 32	GenÃ¨ve	1203	GE
4	J4-9268766O	Lisa	Bochud	09/04/1993	040 227 69 45	lisa_bochud758@vtxnet.ch	01/01/2020	Akazienweg 16	Frauenfeld	8500	TG
5	V8-1786312G	Denise	Tinguely	03/06/1970	063 503 18 68	d.tinguely750@vtxnet.ch	01/01/2020	Untere Landstrasse 4	Ehrikon	8489	ZH
6	e1-6653699a	Jessica	Joly	14/12/1936	+41 (0)33 293 49 05	j_joly@bluewin.ch	01/01/2020	Kisperstrasse 16.3	KÄlliken	5742	AG
7	D7-6803276B	Caroline	Marti	22/11/1936	+41 (0)49 468 94 44	c_marti@hotmail.fr	01/01/2020	Rheinstrasse 93a	Pratteln	4133	BL
8	u6-5156172j	Albert	Marty	01/08/1971	+41 (0)82 931 67 88	a.marty@yahoo.fr	01/01/2020	Meikirchstrasse 57a	Ortschwaben	3042	BE
9	V9-9523491F	Nathalie	Bouvier	16/01/1962	+41 (0)33 679 13 85	nathaliebouvier@vtxnet.ch	02/01/2020	Strickweg 12	NiedergÃ¶sgen	5013	SO
10	t4-7733897B	Sarah	Duroux	06/11/1940	+41 (0)51 631 70 52	sarahduroux964@gmail.com	02/01/2020	Obgardstrasse 18b	Adligenswil	6043	LU
11	Y4-2518220z	Jessica	Chappuis	10/04/1986	065 550 80 65	j_chappuis@hotmail.fr	02/01/2020	Stanserstrasse 24	Kerns	6064	OW
12	R0-3652879d	Jacqueline	Rosselet	19/03/1970	+41 21 322 15 39	jacqueline_rosselet@vtxnet.ch	02/01/2020	BlÃ¤simÃ¼hle 1.4	Madetswil	8322	ZH
13	V0-0674179j	Jonathan	Jacot-Guillarmod	13/03/1956	+41 (0)24 178 90 94	j.jacot-guillarmod@bluewin.ch	02/01/2020	Bockrain 11	Stein am Rhein	8260	SH
14	m2-9866469n	Paul	Uldry	20/06/1955	+41 89 521 03 54	puldry404@yahoo.fr	02/01/2020	Choshof 459	Niederwil SG	9203	SG
15	d4-913663b	Nathalie	Humbert	25/05/1945	+41 43 625 11 63	nhumbert951@vtxnet.ch	02/01/2020	Via Rive di Sopra 11	Besazio	6863	TI
16	P3-6967730h	Luca	Pachoud	02/12/1979	050 269 83 48	lpachoud590@romandie.com	03/01/2020	Glasacker 16a	Dulliken	4657	SO
17	Z1-6771281c	Danielle	DelÃ©ze	16/08/1999	+41 (0)94 557 17 09	danielledeleze979@yahoo.fr	03/01/2020	DÃ¶rfli 8g	Schwendibach	3624	BE



# ratings.csv

## ARGIFTS DATASET

- Customer ratings (1-5) for products in each order

	A	B	C
1	order_id	product_id	rating
2	O-081dZ8150	0	5
3	O-471yz6599	0	3
4	O-884ct2044	0	2
5	O-831zo6118	0	4
6	O-060Gx5763	0	4
7	O-929in3736	0	5
8	O-621sa3999	0	3
9	O-037He8431	0	5
10	O-225nh4456	0	2
11	O-406pn2132	0	3
12	O-754Ap8597	0	2
13	O-275eN1811	0	4
14	O-459yv5387	0	4
15	O-440FN4492	0	3
16	O-378gF1170	0	4
17	O-760ld3087	0	4
18	O-075dz2875	0	4
19	O-042mg4445	0	3
20	O-798rv7978	0	3

# Definitions

## QUANTITY PURCHASED

number of rows per order and product in the orders dataset

## QUANTITY RETURNED

number of rows per order and product in the returns dataset

## PROFIT MARGIN (%)

$(\text{sale\_price} - \text{base\_price}) / \text{sale\_price}$

## SALES

$\text{sale\_price} \times \text{quantity\_purchased}$

## PROFIT

$(\text{sale\_price} - \text{base\_price}) \times \text{quantity\_purchased}$

## PROFIT RATIO (%)

$\text{net profit} / \text{revenue}$

## REVENUE (NET SALES)

$\text{sale\_price} \times (\text{quantity\_purchased} - \text{quantity\_returned})$

## NET PROFIT

$(\text{sale\_price} - \text{base\_price}) \times (\text{quantity\_purchased} - \text{quantity\_returned})$

# Important notes

- The csv files use comma (,) as separator and double quotes (") as text qualifier
- Columns labelled \*\_ID in the different files are unique identifiers
- To prepare the data:
  - Figure out how to combine the different tables to answer the questions
  - Be mindful of the granularity of each table and of identical rows when combining
  - Consider calculating quantities instead of having duplicated rows
  - Be careful when calculating ratios
- If you think you cannot do the necessary data preparation, we can provide you with a **prepared dataset 1 hour after the start of the challenge**
  - You will receive a **10-point penalty** (consult the score sheet for details)

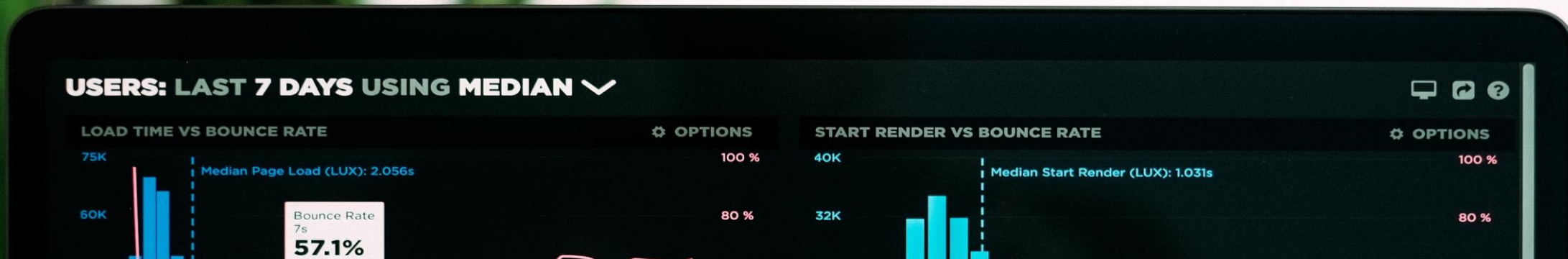
# Questions

- 1. Find the most profitable month and year. What are the profit and revenue (net sales) for that month?** (6 points)
- 2. Which product(s) have the highest sales, profit ratio and average profit margin and what are those values?** (8 points)
- 3. What fraction (%) of returned products are in each category (use category 1)?** (4 points)
- 4. Who are the top 3 clients by revenue and what is the revenue?** (7 points)
- 5. Which client has the most returned products? If multiple indicate the one with the highest return rate. What is their return rate?** (5 points)
- 6. Is there a correlation between the average rating of a product and the return rate? Provide the Pearson correlation coefficient (use 3 decimal places).** (4 points)

# Score sheet

QUESTION	ANSWER	POINTS
1.Find the most profitable month and year. What are the profit and revenue (net sales) for that month?	Month and year (e.g., October 2023)	1
	Profit	2
	Net sales	2
	Graphical representation	1
Total		6
2.Which product(s) have the highest sales, profit ratio and average profit margin and what are those values?	Product 1 ID	0.5
	Product 1 name	0.5
	Sales	1
	Product 2 ID	0.5
	Product 2 name	0.5
	Profit ratio	1
	Product 3 ID	0.5
	Product 3 name	0.5
	Average profit margin	1
	Graphical representation	1
	Excellent representation	1
Total		8
3.What fraction (%) of returned products are in each category (use category 1)?	Correct percent values for each category	1
	Graphical representation	1
	Excellent representation	2
	Total	4

QUESTION	ANSWER	POINTS
4.Who are the top 3 clients by revenue and what is the revenue?	Client 1 ID	0.5
	Client 1 name	0.5
	Revenue	1
	Client 2 ID	0.5
	Client 2 name	0.5
	Revenue	1
	Client 3 ID	0.5
	Client 3 name	0.5
	Revenue	1
	Graphical representation	1
Total		7
5.Which client has the most returned products? If multiple indicate the one with the highest return rate. What is their return rate?	Quantity returned	1
	Client ID	0.5
	Client name	0.5
	Return rate	2
	Graphical representation	1
Total		5
6.Is there a correlation between the average rating of a product and the return rate? Provide the Pearson correlation coefficient (use 3 decimal places).	Coefficient	2
	Graphical representation	1
	Excellent representation	1
	Total	4
Total for questions		34
Bonus		6 points max at the discretion of the jury
GRAND TOTAL		40
Deductions		if pre-prepared dataset was used
		-10



# Thank you for participating!

