



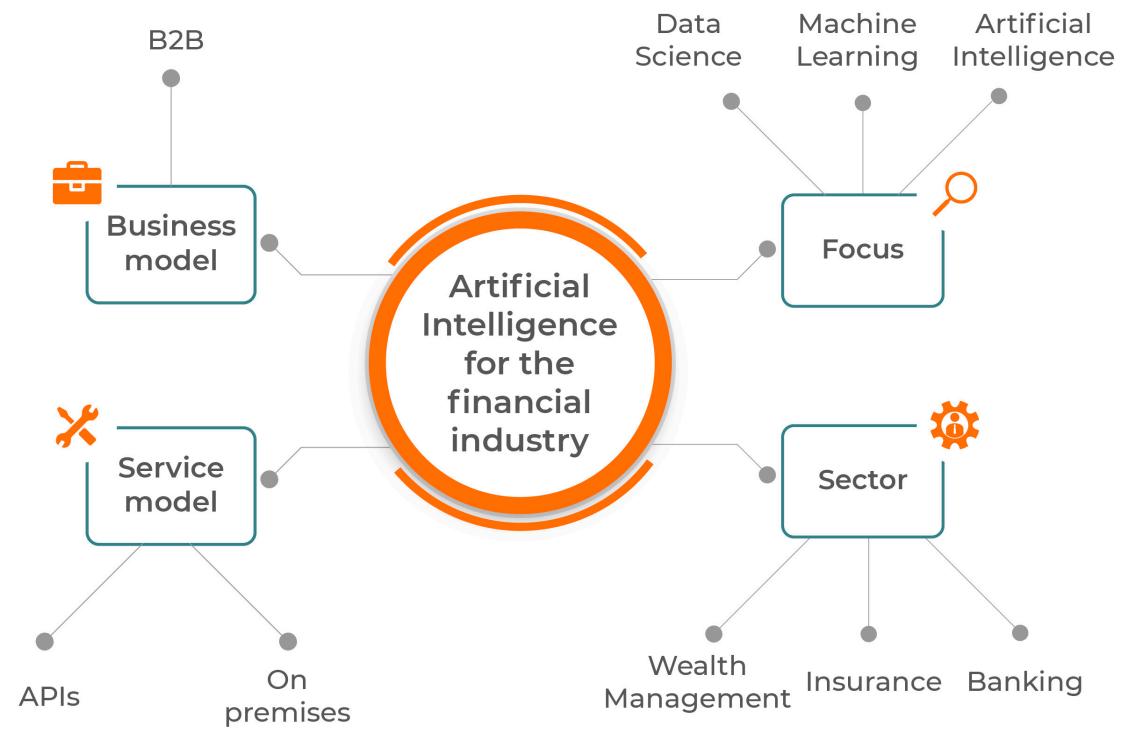
Module #1

Data-driven customer profiling for financial institutions/Fintechs

Raffaele Zenti (raffaele.zenti@wealthype.it)

Co-Founder, COO, Chief Data Scientist,
Wealthype-ai SpA

**Why here?
Because of
Wealthype-ai SpA**



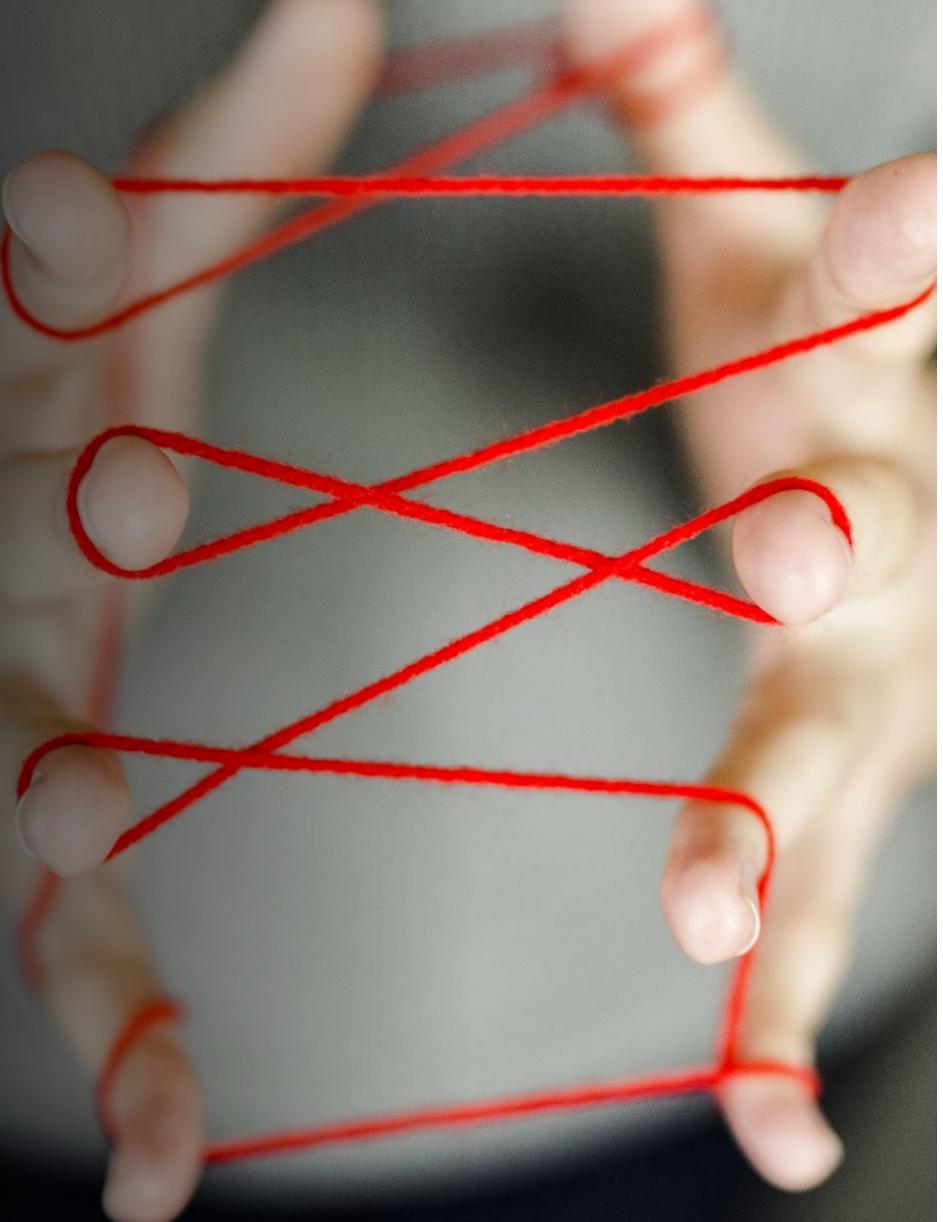
Overall goals

- You will savor the practice of Machine Learning in the Fintech sector with a series of business cases
- We will cover various aspects of the business (e.g., customer intelligence, digital marketing, investing, risk analysis) and a significative selection Data Science techniques (obviously not all)
- I would like to stimulate your ability to solve real problems with Data Science - having modeling vision – not just writing a piece of code, or using that new fancy model...





Organizational aspects



GROUP WORK & POSTER PRESENTATIONS

Group formation & regular assignments

- **Groups** of 4-5 students
- Each group must work on all **4 business cases**
- Output required: **poster format** (2 columns × 4 rows of A4 sheets) **plus supporting materials** (e.g., being able to show the code on a laptop)

Presentation & Scoring

- **Each group presents their poster** during 4 discussion sessions
- **Minimum 2 group members** must be present for each presentation
- Each successful presentation earns **2 points** for ALL group members
- **Mentorship sessions** will be scheduled before poster sessions – they are videocall with me, if you have questions, you need some coaching, brainstorming, etc
- Maximum overall possible poster points: **8 points** (2 points × 4 posters)

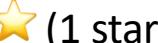
FINAL PROJECT & BONUS OPPORTUNITIES:

max Machine Learning score = $8+10+2 = 20$

Machine Learning final Project Work

- Groups must complete a **Project Work based on one of the 4 business cases** – NOT alternative datasets
- Submission (ppt + code) due at least 10 days before exam date
- This final case WILL be directly evaluated (0-10 points)
- Final score calculation: Poster points (0-8) + Final case points (0-10)

Machine Learning Bonus system (for attending students only)

- I will award **stars to top posters** for each business case:
 - Best poster:  (3 stars)
 - Second best:  (2 stars)
 - Third best:  (1 star)
- Bonus points awarded at end of course based on **total stars**:
 - 1st place group(s): +2 points
 - 2nd place group(s): +1.5 points
 - 3rd place group(s): +1 point

Hands on lessons

A number of different business cases =
problems to solve with Data Science

- I present the case from a **business point of view**
- I present an associated **dataset**
- We will do some **brainstorming**: I want to see you come up with ideas on how to solve the problem
- I **show you some code** that solves the problem:
My code is never “definitive”, it is just a starting point and it requires work on your part - but you will find a lot of hints
- I leave you to **develop your own ideas and code** (the aim of the course), **playing at home, in groups**
- There are **ancillary hours («office hours») of my mentoring** to support you in your work: just drop me an **email** here: **raffaele.zenti@wealthype.it**



Coding

- In my lectures I will use both Python and Matlab - sometimes structured code, sometimes just example/code snippets
- You can use the programming language of your choice, Python, Matlab, or R
- Please use markdowns, comments, formatting and make the code as clear as possible (better if you use notebooks like Colab, Jupyter, etc)

LEVERAGING LLMs FOR CODING SUPPORT – My 2 Cents

- **Encouraged use:** LLMs can significantly enhance your coding efficiency and problem-solving capabilities
- **Develop Cointelligence:** Build a collaborative relationship with AI where
 - YOU maintain strategic direction and control
 - YOU define the approach and goals clearly—the strategy must come from you, not the AI
 - YOU govern the development process and retain oversight of all generated solutions
 - YOUR critical judgment is key
 - The LLM provides suggestions, implementations, and alternatives, becoming a great partner
- **Human-in-the-Loop Value:**
"Use AI as a powerful tool, not as a replacement for your own understanding."

Homework: a note



It is designed to familiarize you with real problems, to let you try different techniques



It is not a competition to see who has the best model accuracy, or MSE, or whatever...



You don't have to get better results than I have shown in class



It's just a little bit of experimenting with the data

CREATING EFFECTIVE ACADEMIC POSTERS: KEY PRINCIPLES



Structure matters: Organize content logically (left-to-right, top-to-bottom)

Clear sections with consistent layout



Visual hierarchy: Direct attention through font sizing and positioning

Headings: 36-48pt

Body text: 24-28pt (readable from 3-4 feet away)



Show, don't tell: Aim for 40% text, 60% visuals

Replace text with diagrams, charts, and images



Simplify & Synthesize: Be ruthlessly concise

Focus only on key messages and findings, time is short



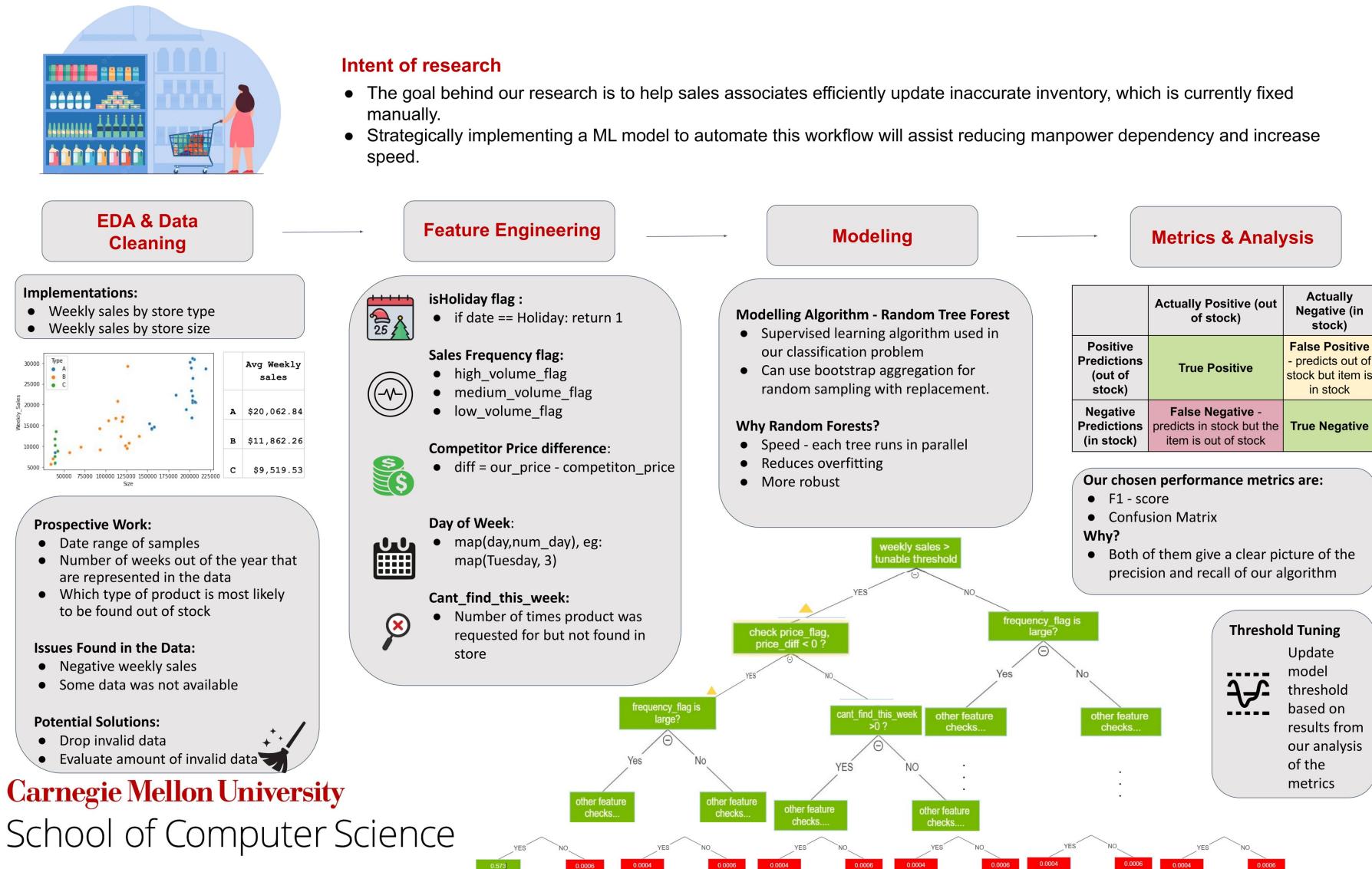
Presentation skills: Be prepared to present your work

Stand beside your poster and be ready to answer questions with supporting materials (code, data)

Practice a 2-minute summary of your key points

Machine Learning to Automate Store Operations

By: Senam Anaglate, Melissa Crozier, Makayla Palmore & Ujvala Pradeep



Carnegie Mellon University

School of Computer Science

Useful links

- <https://www.theonlinescientist.com/poster-guidelines/>
- <https://365datascience.com/trending/how-to-make-a-poster-for-your-data-science-project/>
- <https://guides.nyu.edu/posters>
- <https://www.cmu.edu/cs/ourcs/poster/index.html>

POSTER STAR RATING CRITERIA

How Stars will be awarded

- **Quality of work (★)**
 - Model tuning, optimization, etc
 - Novel approaches (from class "Hints" or your own ideas)
 - Evidence of experimentation with algorithms and data
- **Poster design (★)**
 - Logical structure and flow
 - Effective visual hierarchy
 - Balance of text and visuals
 - Clarity and conciseness
- **Presentation ability (★)**
 - Clear explanation of methodology
 - Ability to defend choices and approaches
 - Understanding of underlying concepts

Important note

- **A poster is NOT the Final Project Work** – depth of analysis will be evaluated accordingly. Focus on hands-on experimentation rather than comprehensive analysis.

FINAL PROJECT WORK EVALUATION CRITERIA

Deliverables

- Concise **presentation** (\approx 12 slides)
- Well-organized **code** (Python, R, Matlab, or mixed)

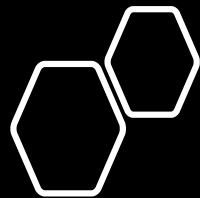
Evaluation Framework

- **Modeling vision & Problem-solving (★★)**
 - Ability to "connect the dots" in practical scenarios
 - Focus on solving the actual business problem
 - Pragmatic approach over unnecessary complexity
- **Code quality (★★)**
 - Clean, readable, and well-structured code
 - Intelligent use of functions and libraries
 - Effective integration if using multiple languages
- **Presentation skills (★★)**
 - Clear, concise communication of methods and results
 - Focus on core findings without excess detail
 - Logical flow and professional delivery
- **Originality & exploration (★★)**
 - Differentiation from provided examples
 - Exploration of alternative approaches
 - Following Occam's Razor: valuing simplicity when equally effective



Business case #1:

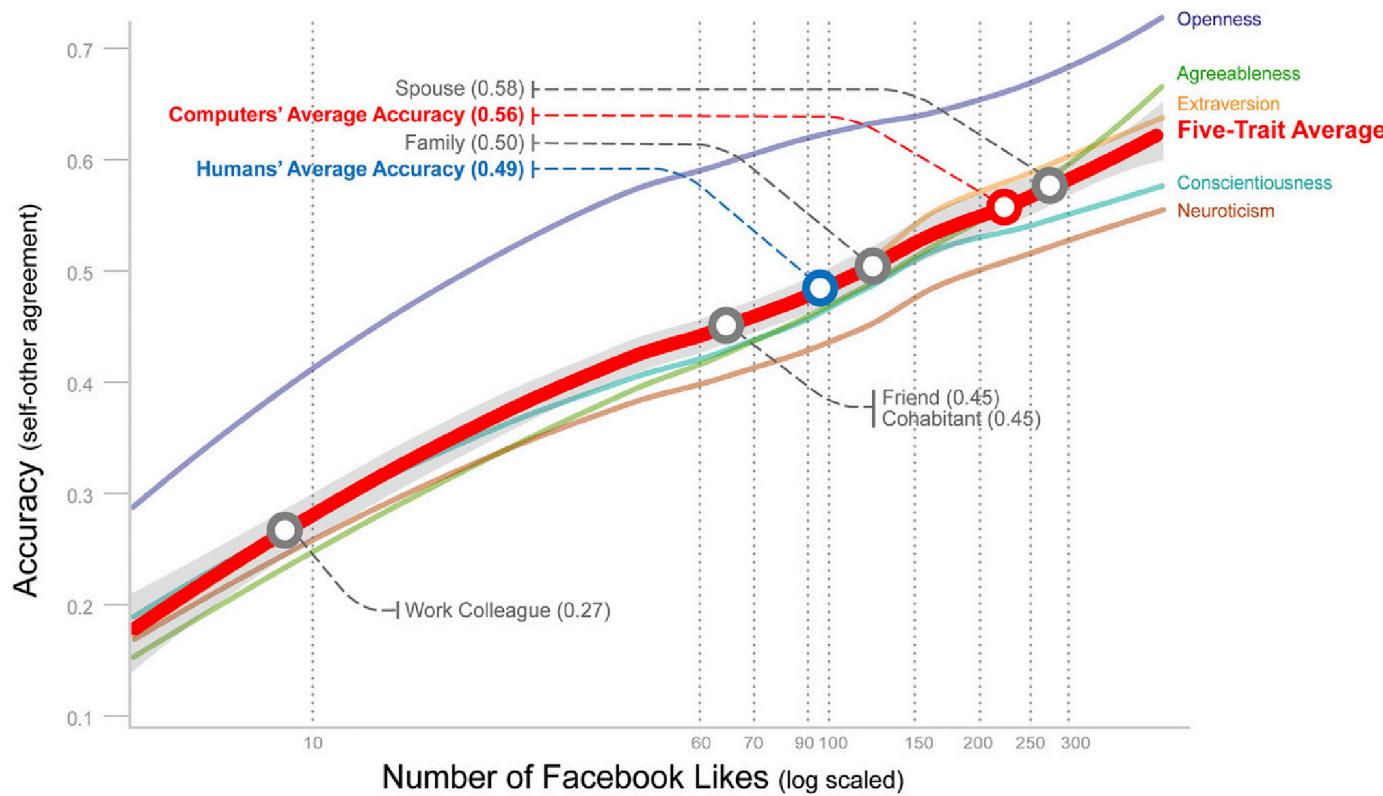
Segmenting Clients



AI/Machine Learning aimed for...?

- Understand or predict:
 1. Markets
 2. People (Customers) →
this is the case

Motivation: AI/ML and people



A Machine Learning algorithm needs 70, 150 and 300 Facebook Likes, respectively, to outperform an average friend, family member, and spouse in assessing your personality*

*Five-Factor Model model

Source: Youyou, W., Kosinski, M., & Stillwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans. PNAS Proceedings of the National Academy of Sciences of the United States of America, 112(4), 1036-1040

Why segment? To efficiently customize any service





Client segmentation: the old school in finance

- Money, e.g
 - AuM
 - insurance premiums
 - number of trades p.a.
 - ...
- Age
- A combination of 1. and 2.

(Yeah. Boring. There is room for some improvement - unless you think that money and/or age fully define a human being)

Personas: general idea

- Premise: we are all different, but with common traits
- Personas (or “Marketing Personas”, or Financial Personas, in Financial Services) = **human prototypes with similar traits**
- Depending on the perspective, human prototypes change – here we take the financial perspective
- Key to summarize information ↔ **dimensionality reduction**
- That is, instead of reasoning on, say, 300k customers, you can reason on 6 TYPES of customers – which is easier

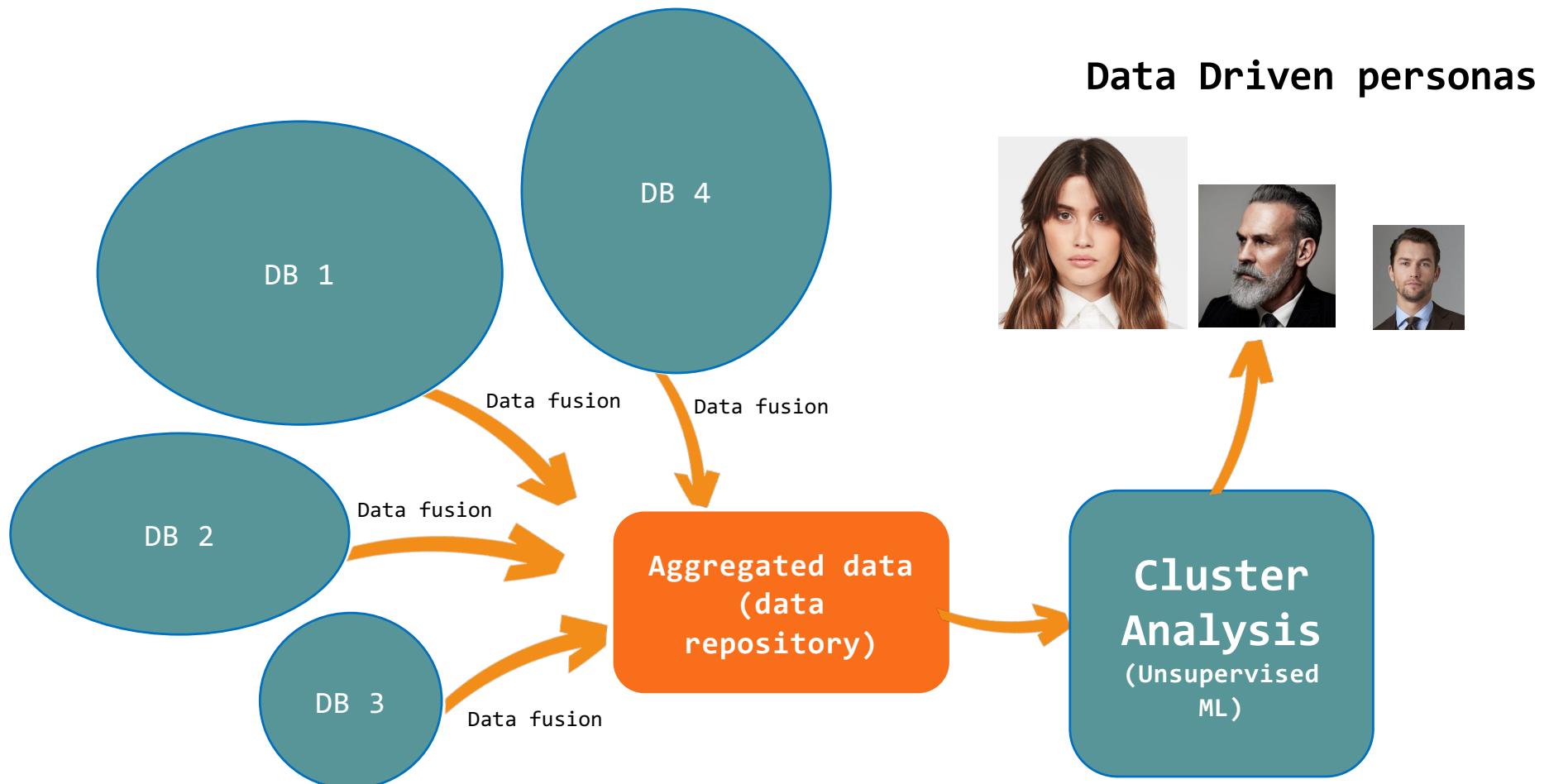


Personas according to traditional marketing

- Put some experts in a room
- Super-brainstorming
- Based on their experience, they will identify «customer prototypes», i.e., personas – they will likely include their biased personal ideas

(Better. But there is still room for some improvement)

Data-driven personas: (simplified) overview



The problem

- **Goal: segment clients based on the information contained in their corresponding vectors of features**
- Spoiler: in real world client data contain heterogeneous data (categorical nominal/ordinal, numerical ordinal/continuous)
- Introductory courses to unsupervised learning quite often discuss ideal use cases, such as tutorials using k-means, which works great, but only apply to numerical features – so we are out of the comfort zone...
- Now: let's inspect data and discuss!



Coding session
starts

At the end of the process...

- Qualitative overlay might change the number of clusters, or might change centroids/medoids

- Each client is a vector:

$z(i), i = 1, 2, 3, \dots n_{Clienti}$

- Each Persona (centroid/medoid) is a vector:

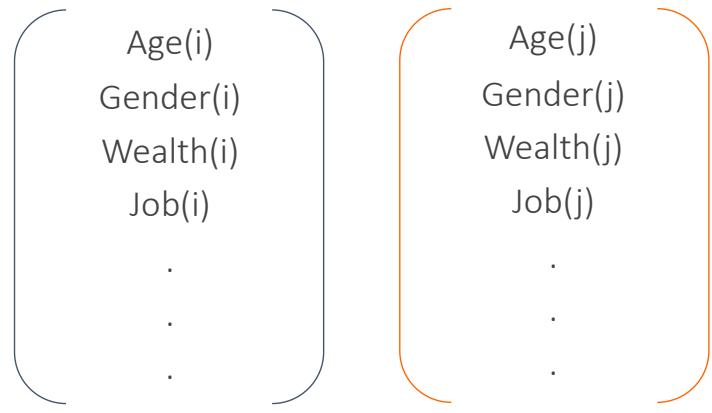
$C(j), j = 1, 2, 3, \dots n_{Personas}$

- Client(i) belongs to the closest Personas(j), i.e. the rule is:

$\text{ArgMin}\{ \text{dist}[z(i), C(j)] \}$

j

Client(i) Personas(j)



Dist[Client, Personas]

Search for the nearest Personas

Real example – inside clusters (it's not your dataset)

Features	Cluster (2) = 21% “Wealthy widow”
Age	55-70
Gender	F
Job	Housewife, retired
Marital status	Widow, separated, divorced
Family	1
Financial education	Below average
Geographical area	Italy
Size of the municipality	20k÷50k
Income	Above average
Mortgage	N
Short term loans	N
Real estate wealth	Huge
Financial assets	Above average
Socio-demographic risk	Average
Geo-seismic risk	Above average
Digital propensity	Low



Service model:

- Physical
- Physical+Call center



Main needs:

- Long Term Care
- Inheritance
- Investments (low risk, capital protection)
- Premium credit cards

Real example – inside clusters

Features	Cluster (3) = 2% “Top jobs”
Age	50-65
Gender	M
Job	Manager, professional, business owner
Marital status	Married
Family	2-3
Financial education	High
Geographical area	North Italy, large cities
Size of the municipality	Above 200k
Income	High
Mortgage	Y
Short term loans	N
Real estate wealth	Above average
Financial assets	High
Socio-demographic risk	Above average
Geo-seismic risk	Low
Digital propensity	High



Service model:

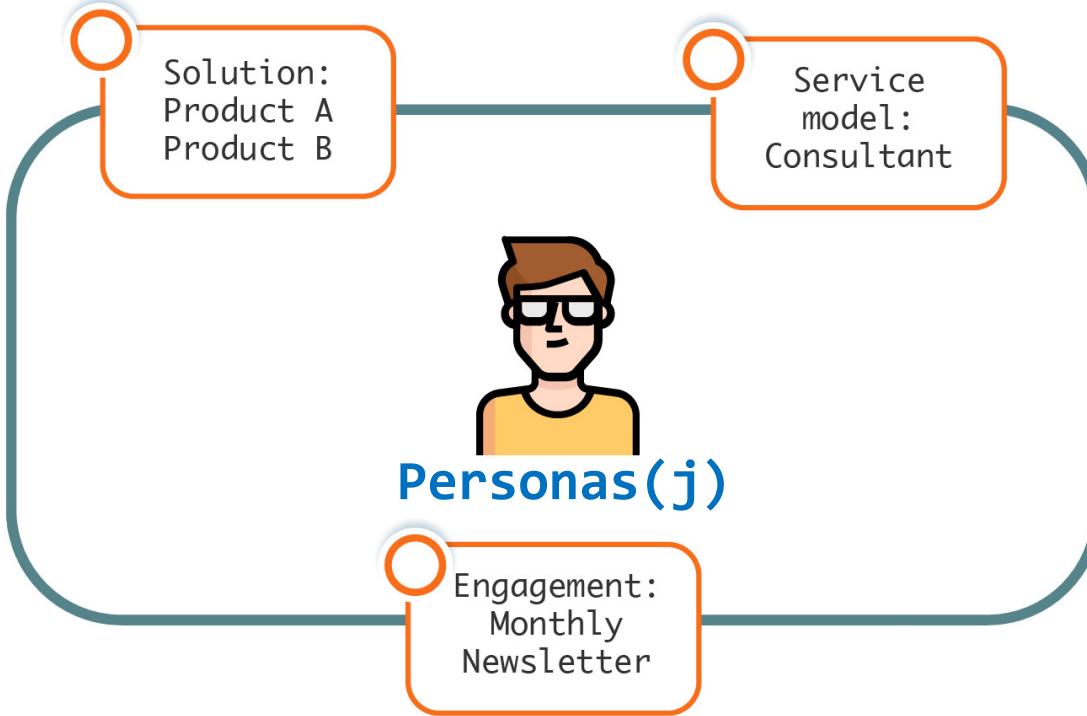
- Digital
- Physical+call center



Main needs:

- Long Term Care
- Family/Home protection
- Death insurance
- Investments
- Premium credit cards

Using Personas to customize financial services in an industrial way



- Each Persona has her service/communication model, products, etc
- But Personas are typically 5÷15 → you reduce business complexity

Take home on data-driven personas

- Application of Unsupervised ML + qualitative overlay (strongly suggested)
- Much better segmentation, based on empirical evidence
- Several business applications:
 - customized products, services, channels
 - dedicated communication tools for each personas
 - precision marketing online/offline
 - data enrichment if we have a limited amount of information (a handful of features) → then you can start engaging customers
 - generating synthetic data, for ML training and simulations
 - ...

Now **YOU**

- **It's your turn: use your favorite techniques to segment customers, write the code (use my code, or start from scratch, or whatever...as you like), and we'll talk about it next time**
- **Work in small groups: use collective intelligence (Data Science = teamwork)**
- **Get your hands on that data...**

Next steps

- **Office Hours = mentorship**
 - Monday 3 March from 4:00pm to 5:30pm
 - Wednesday 5 March from 4:45pm to 5:45pm
 - It's ONLINE:
<https://politecnicomilano.webex.com/meet/raffaele.zenti>
- Book if you want to talk to me; write to me at this email address:
raffaele.zenti@wealthype.it
(NOT the Politecnico email)
- **Poster session**
 - Thursday 13 March from 8:30am to 11:00am
 - Refer to what was said in the previous slides