

Bernardo Bouzas

David Méndez

Claudio Sotillos

Laura Torregrosa



# uc3m x IronIA Fintech

robo-advisor algorithm solution for drawdown-based  
**optimization of investment portfolios.**

B. Sc. in Data Science and Engineering

Coordinator: David Delgado



# the team



Universidad  
Carlos III de Madrid



**Bernardo**  
Bouzas



**David**  
Méndez



**Claudio**  
Sotillos



**Laura**  
Torregrosa



# the team



**Bernardo**  
Bouzas



**David**  
Méndez



**Claudio**  
Sotillos



**Laura**  
Torregrosa



Universidad  
Carlos III de Madrid

# the partner

create your **investment portfolio** from over  
**+20 000 funds**



# the problem

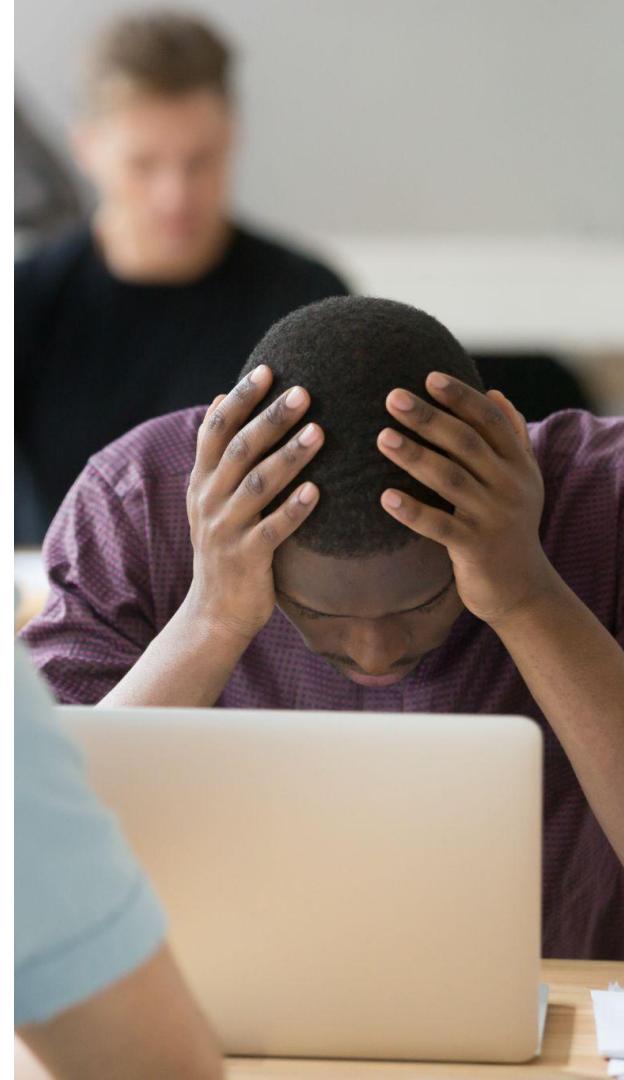
**+20 000 funds** is an overwhelming number of choices



# the problem

**+20 000 funds** is an overwhelming number of choices

investors have different **risk aversion profiles**

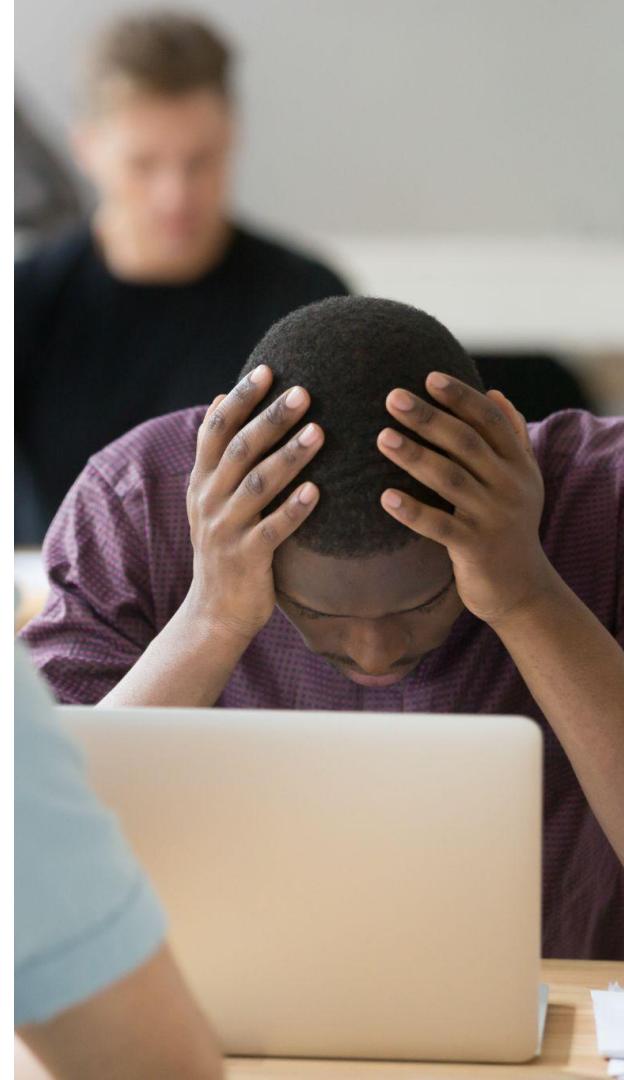


# the problem

**+20 000 funds** is an overwhelming number of choices

investors have different **risk aversion profiles**

53% of spanish people\* have a **minimum financial literacy** level



\* AXA INVESTMENT MANAGERS research



# **the solution**



# the solution

classic in the basics ...



Portfolio optimization theory: **Markovitz**

"COMPARATIVE ANALYSIS OF LINEAR PORTFOLIO REBALANCING STRATEGIES: AN APPLICATION TO HEDGE FUNDS" (Nov 2001)

*Pavlo Krokhmal, Stanislav Uryasev, Grigory Zrazhevsky*



# the solution

**classic in the basics ...**



Portfolio optimization theory: **Markovitz**

"COMPARATIVE ANALYSIS OF LINEAR PORTFOLIO REBALANCING STRATEGIES: AN APPLICATION TO HEDGE FUNDS" (Nov 2001)

*Pavlo Krokhamal, Stanislav Uryasev, Grigory Zrazhevsky*

**... innovative in the means**



python solvers

machine learning

web app SDK

# steps

01

## Work with the available data

- Big Data processing
- Cloud Environment
- Training and test sets

03

## Upgrade the model: clustering

- Global market characterization
- Principal Component Analysis
- K-Means clustering

02

## Design an optimization algorithm

- Formulate linear risk
- Allocate funds in the portfolio
- Enhance the tool

04

## Deploy a Minimum Viable Product

- Assess user risk profile
- Implement interface with SDK
- Get feedback

# massive amounts of financial data

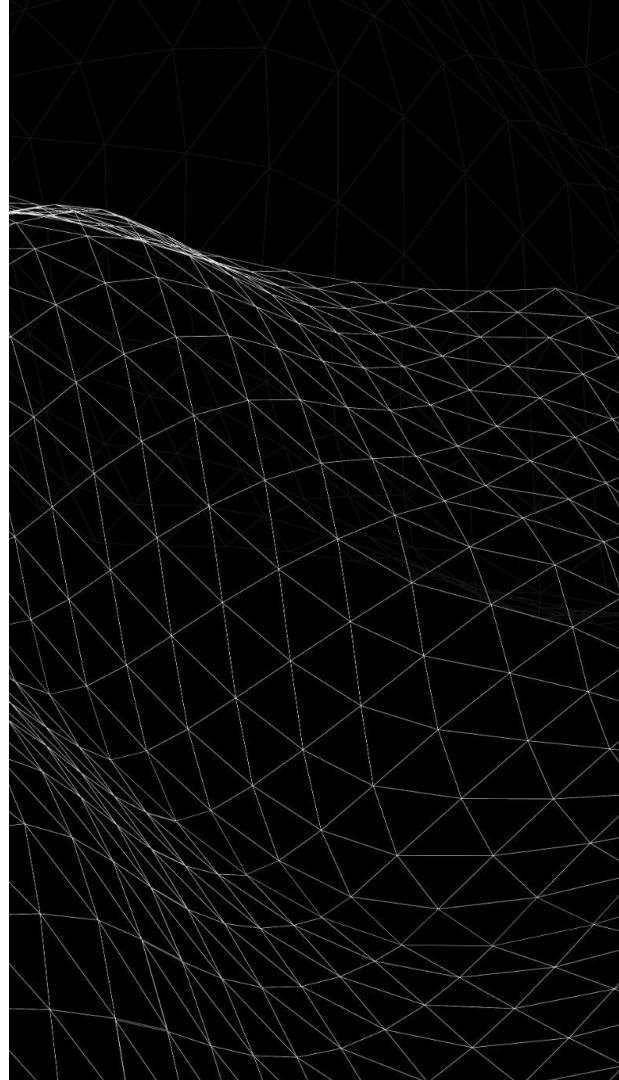
we need to **work in cloud**



Google  
BigQuery

colab git

we need to **assume, filter, and clean**



# dataset considerations

covid has impacted



train set: **2016-2018**

test set: **2019**

# dataset considerations

covid has impacted



train set: **2016-2018**  
test set: **2019**

we need time series

	Fund A	Fund B
2016	✓	✓
2017	✓	✓
2018	✗	✓
2019	✗	✓

# dataset considerations

covid has impacted

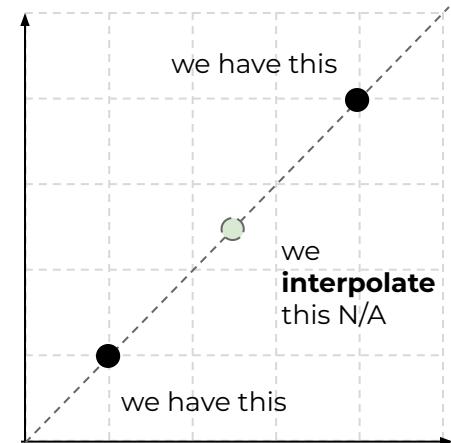


train set: **2016-2018**  
test set: **2019**

we need time series

	Fund A	Fund B
2016	✓	
2017	✓	✓
2018	✗	✓
2019	✗	✓

null values are annoying



A close-up photograph of a person's hand placing a small wooden block with a black arrow pointing upwards onto a stack of two other wooden blocks. The blocks are light-colored wood with dark vertical stripes. The background is a warm, out-of-focus brown.

# challenge

generate an **optimal portfolio allocation**



maximize **investment return**



minimize **risk**

A close-up photograph of a person's hand placing a small wooden block with a black arrow pointing upwards onto a stack of similar blocks. The blocks are light-colored wood with dark vertical stripes. The background is a warm, out-of-focus brown.

# challenge

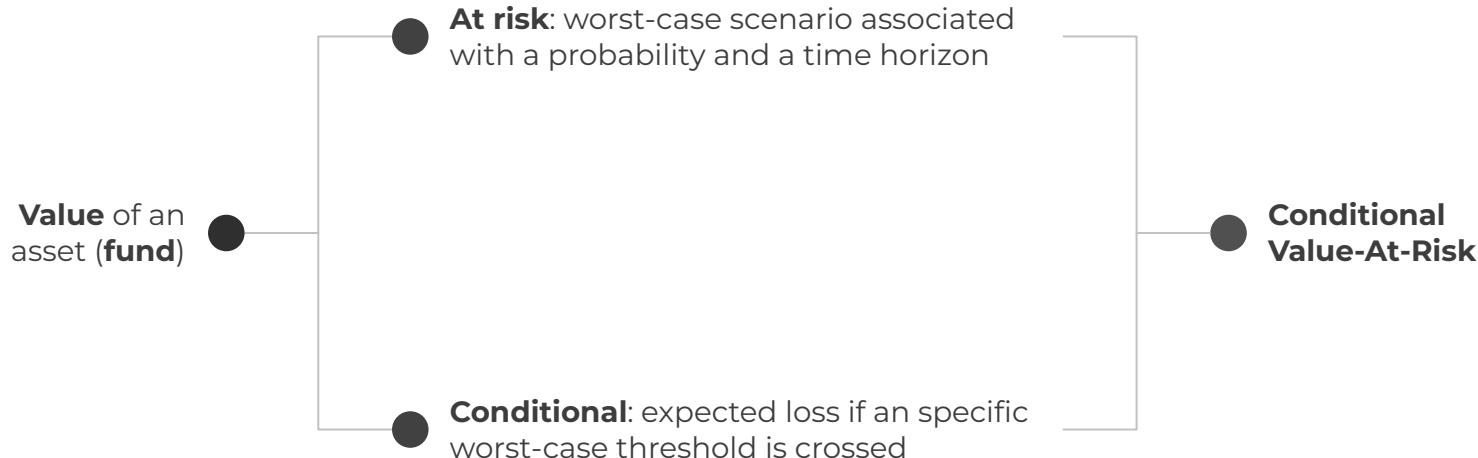
generate an **optimal portfolio allocation**

=

maximize **investment return** + minimize **risk**

**but what is  
actually ‘risk’?**

# how to approach risk

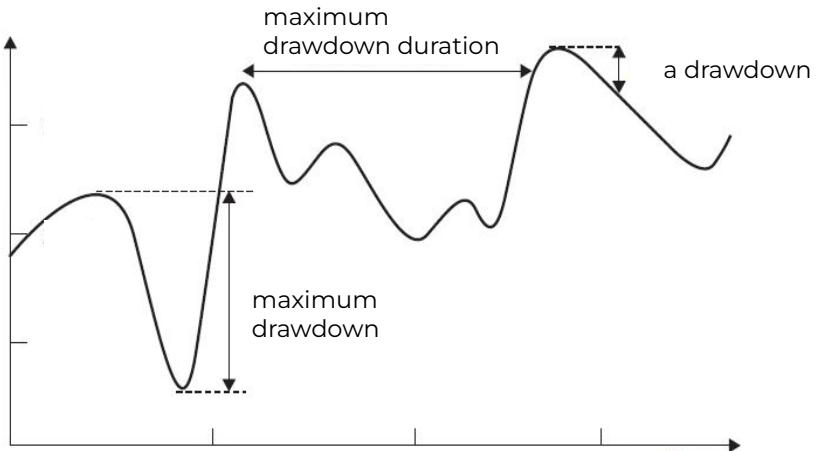


# how to approach risk

same applies for drawdown

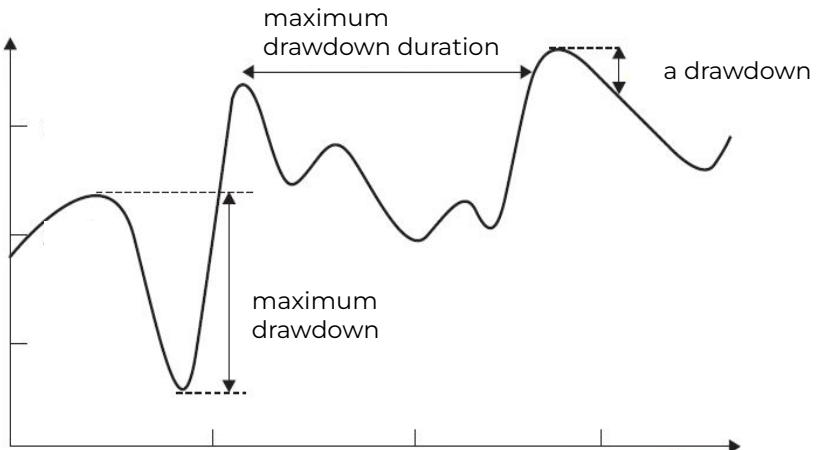
# how to approach risk

## same applies for drawdown



# how to approach risk

same applies for drawdown



**Conditional  
Drawdown-at-Risk**  
vs (just) Drawdown-at-Risk

- superior **mathematical** properties
- more **efficient**
- accounts for **losses** exceeding DaR
- strong competitor to **Standard Deviation**

# how to approach risk

now we need some constraints

# how to approach risk

now we need some constraints

OBJECTIVE FUNCTION

$$\max_x E \left[ \sum_{i=1}^n r_i x_i \right]$$

# how to approach risk

now we need some constraints

## OBJECTIVE FUNCTION

$$\max_x E \left[ \sum_{i=1}^n r_i x_i \right]$$

- 01 fund limitation
- 02 budget constraint
- 03 **risk function**
- 04 market neutrality

# how to approach risk

## now we need some constraints

### OBJECTIVE FUNCTION

$$\max_x E \left[ \sum_{i=1}^n r_i x_i \right]$$

01

fund limitation

02

budget  
constraint

03

risk function

04

market  
neutrality

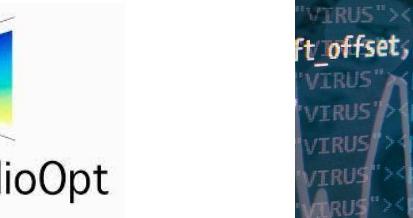
$$\begin{aligned}\zeta + \frac{1}{1-\alpha} J \sum_{j=1}^J w_j &\leq \omega, \\ -\sum_{i=1}^n r_{ij} x_i - \zeta &\leq w_j, \quad j = 1, \dots, J, \\ \zeta \in \mathbb{R}, \quad w_j &\geq 0, \quad j = 1, \dots, J.\end{aligned}$$

$$\begin{aligned}w &\leq \omega, \\ -\sum_{i=1}^n r_{ij} x_i &\leq w, \quad j = 1, \dots, J.\end{aligned}$$

$$\begin{aligned}\eta + \frac{1}{1-\alpha} J \sum_{j=1}^J w_j &\leq \omega; \\ Z - \sum_{i=1}^n \left( \sum_{s=1}^j r_{is} \right) x_i - \eta &\leq w_j \quad j = 1, \dots, J; \\ \sum_{i=1}^n \left( \sum_{s=1}^K r_{is} \right) x_i &\leq Z \quad K = 1, \dots, j; \\ \eta \in \mathbb{R}, \quad w_j &\geq 0 \quad j = 1, \dots, J.\end{aligned}$$

### LINEARIZATION

# let's get down to coding

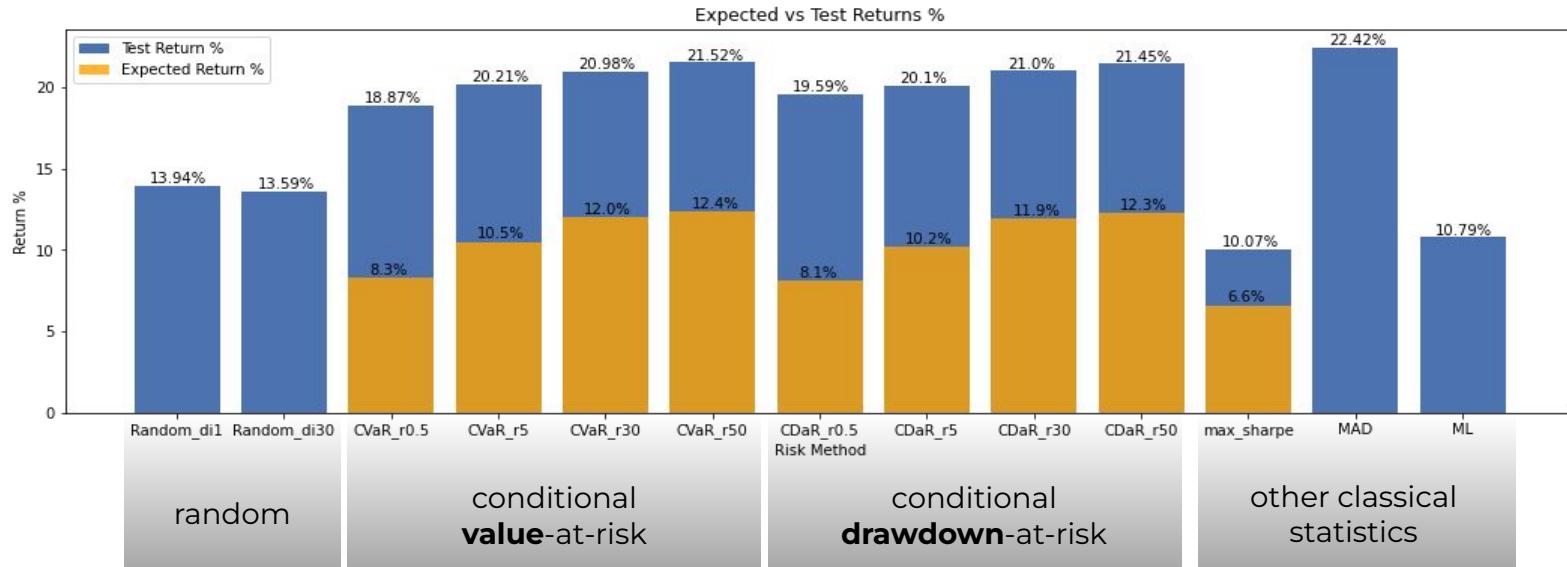


# **preliminary results**

## **comparing risk metrics**

# preliminary results

## comparing risk metrics

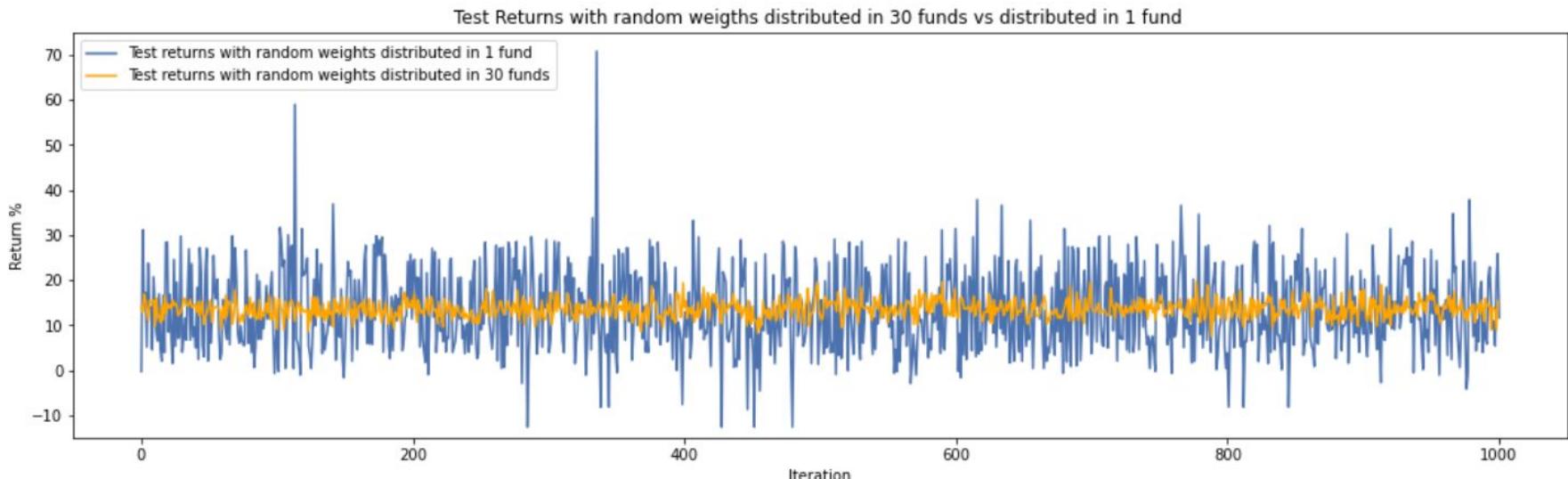


# **preliminary results**

the importance of diversification

# preliminary results

## the importance of diversification



A close-up photograph of a person's lower torso and hand. The person is wearing brown jeans with a visible pocket. A silver wrench is stuck vertically through the back pocket. On their left wrist, they wear a yellow elastic band. Their right hand is partially visible, showing the fingers.

# time to make improvements

- only works with **500 funds**
- **fund classes** are pseudoreplicates
- we need to further **risk-balancing**

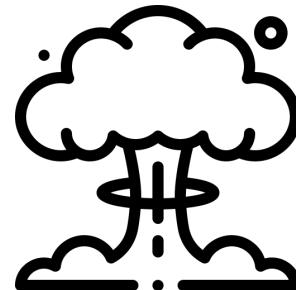
# **hierarchical computation**

**why do we need this?**

# hierarchical computation

why do we need this?

>500 funds      =

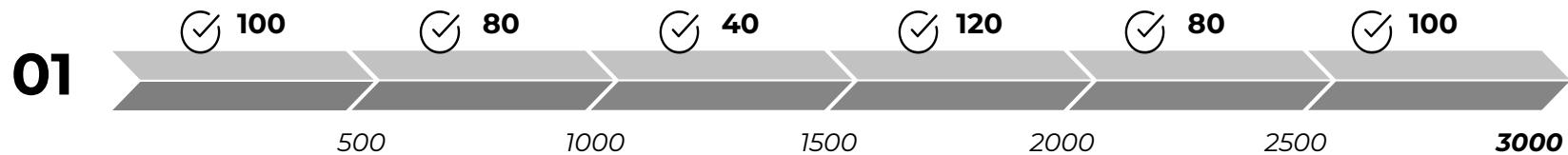


# **hierarchical computation**

**segment, optimize, repeat.**

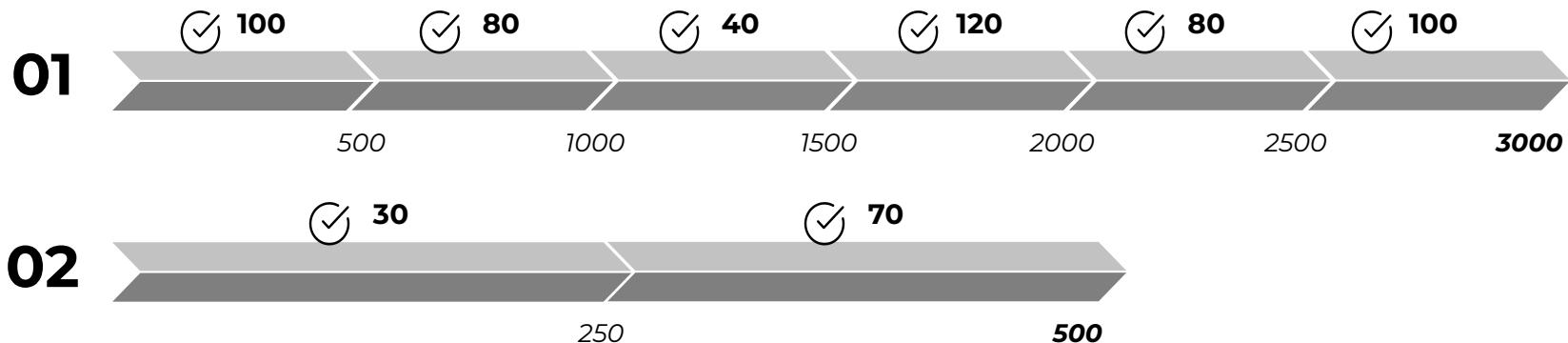
# hierarchical computation

segment, optimize, repeat.



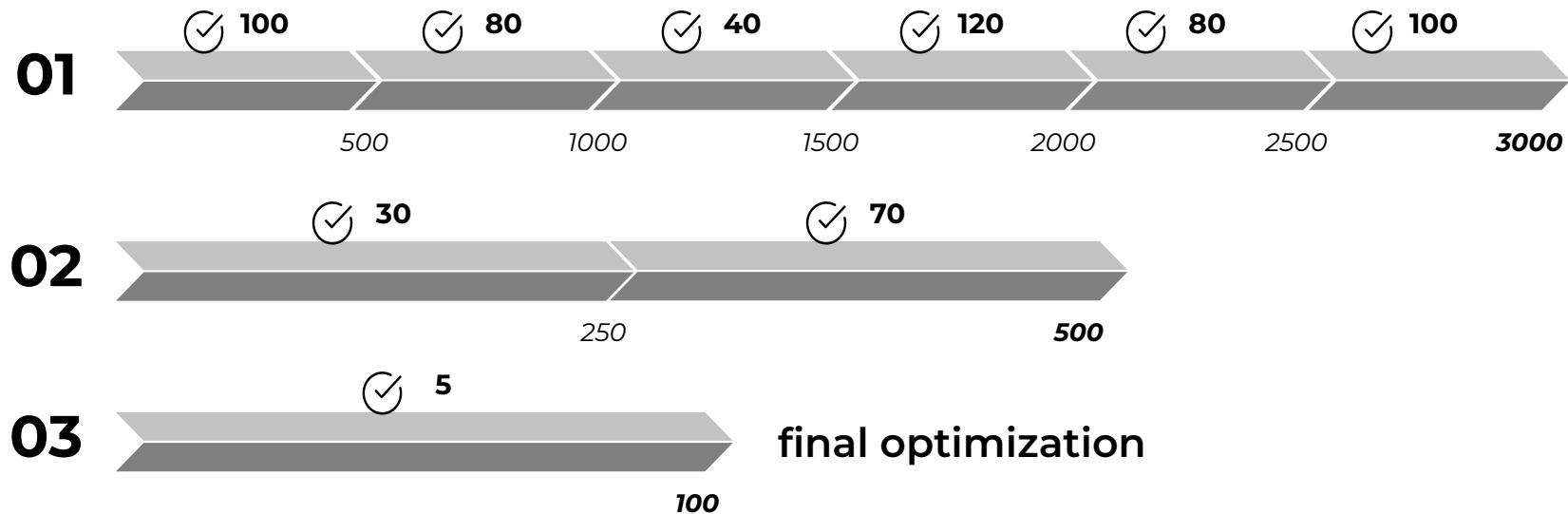
# hierarchical computation

segment, optimize, repeat.



# hierarchical computation

segment, optimize, repeat.



# what is going on with fund names...

different **currency**,  
different **min. capital**

JPM US TECHNOLOGY "R" (USD)

JPM US TECHNOLOGY "I" (USD)

JPM US TECHNOLOGY "I" (GBP)

same **currency**,  
different **min. capital**

same **min. capital**,  
different **currency**

---

same **asset composition**,  
different **for the algorithm**

**what is going on with  
fund names...**

**... solution?**

# what is going on with fund names...

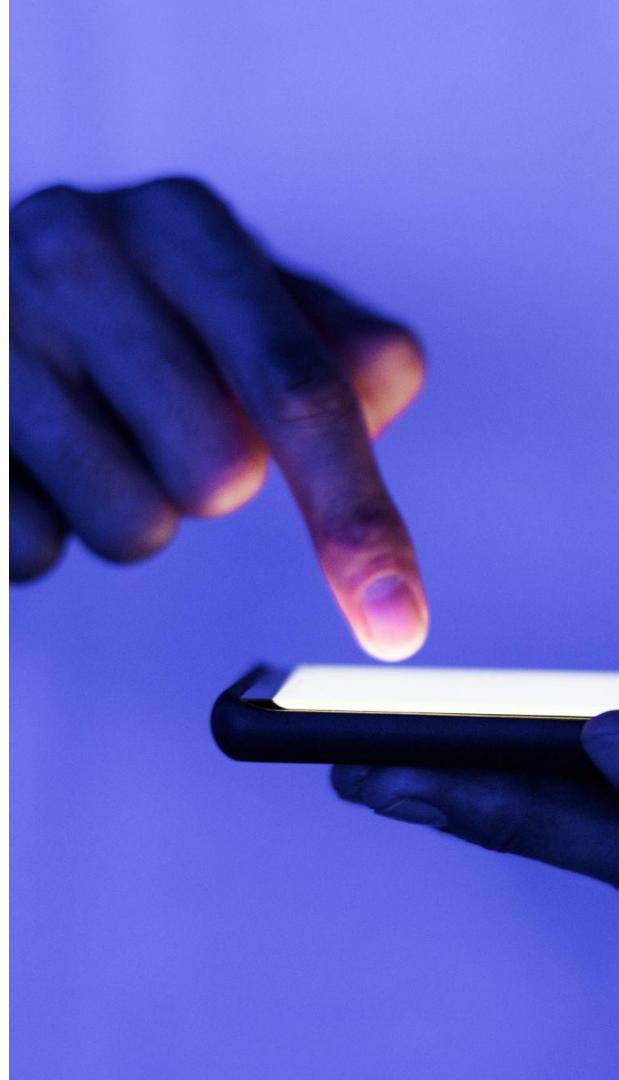
... solution?

## Levenshtein distance

H E L L O   W O R L D

H E L L O , G I R L S

**key point:**  
**broadening risk**  
**customization**

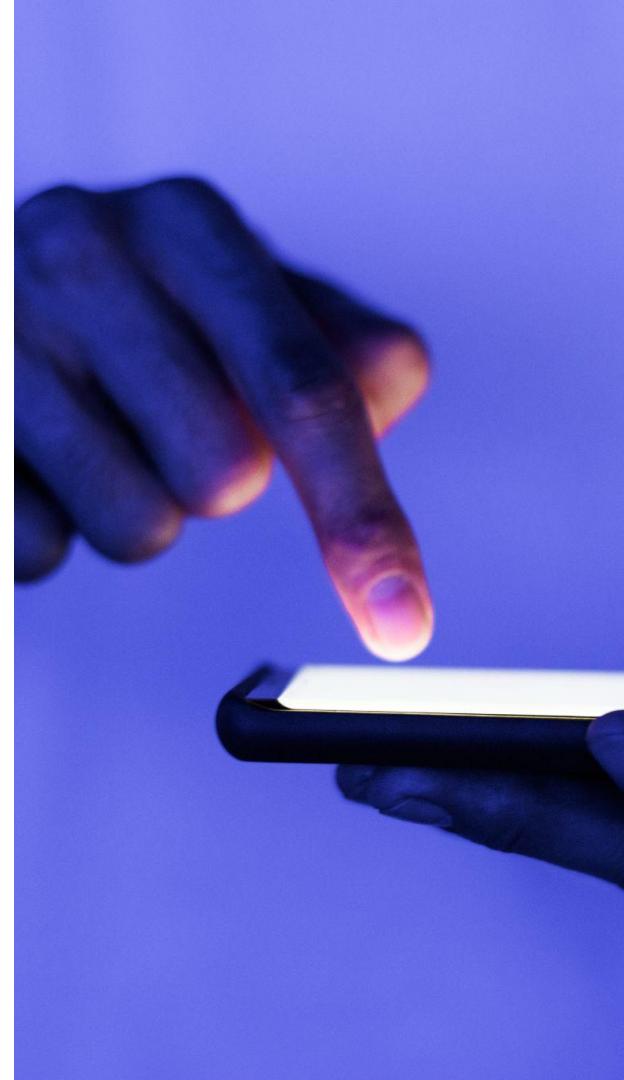


# key point: broadening risk customization

the **algorithm**'s risk tolerance

should depend on

the **investor**'s risk tolerance



# **problem:** *optimal portfolio* **is not optimal for everyone**

different **people**

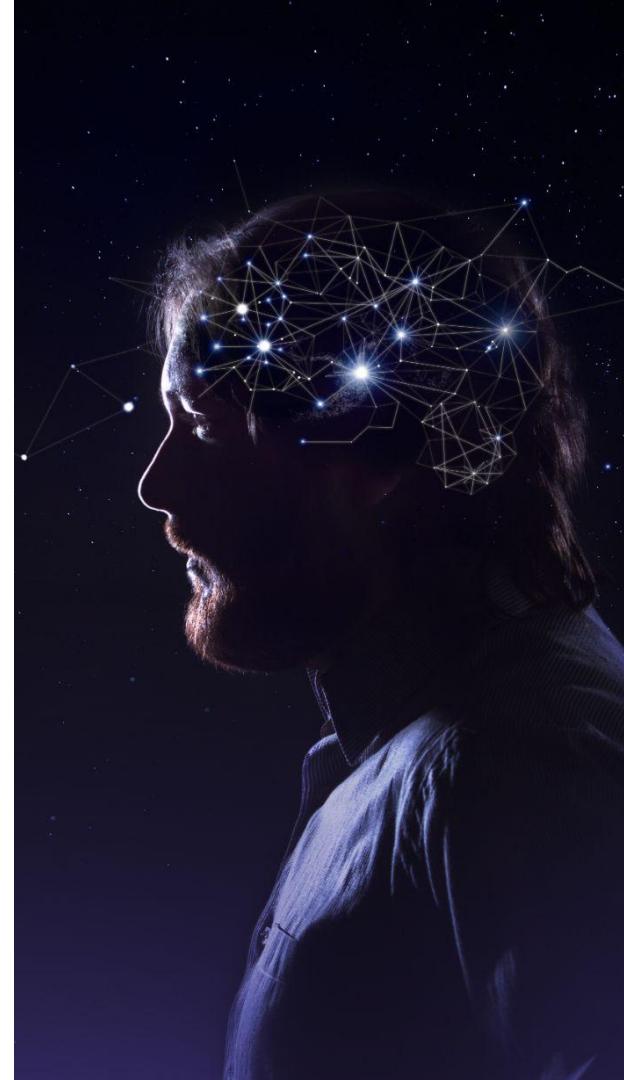
different **psychological traits**

different **risk profiles**

different **funds**

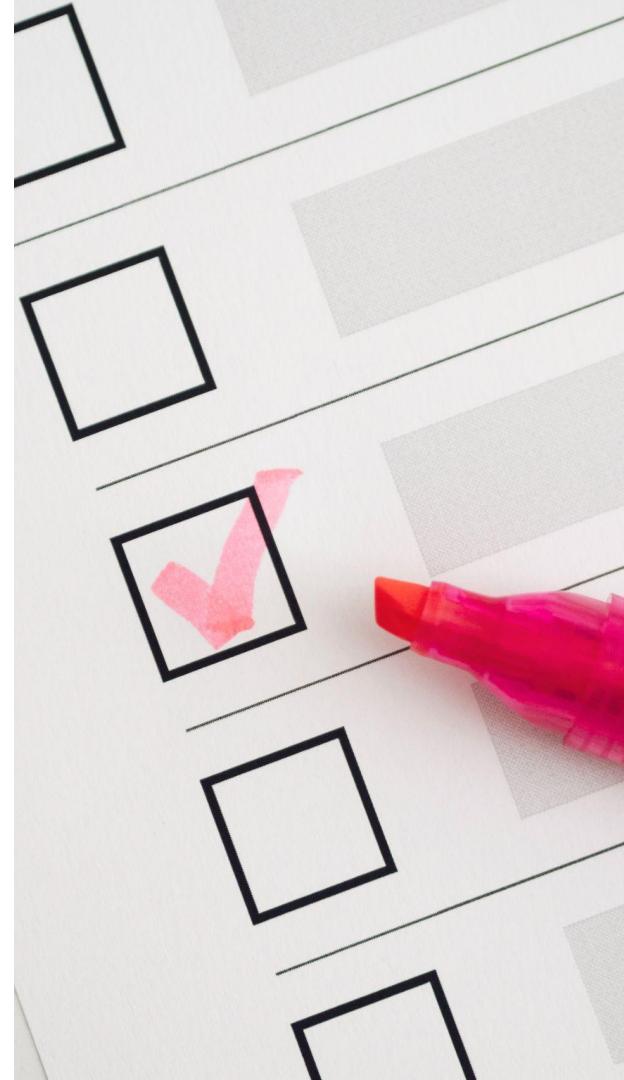
different **asset compositions**

different **risk levels**



# solution: assess risk qualitatively

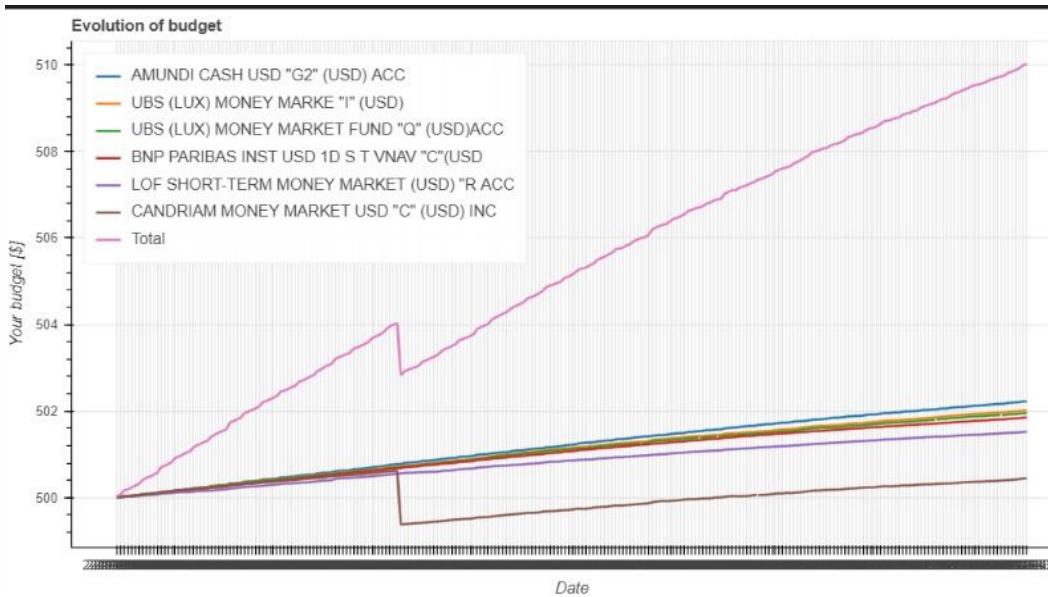
- a. fund **risk level**  $\propto$  fund **composition**
  
- b. investor **risk tolerance** - questionnaire



a. **determine** a risk level **for each** fund type

a. **determine** a risk level **for each** fund type

## Risk 1: **Monetary** Funds



a. **determine** a risk level **for each** fund type

Risk 1 : **Monetary** Funds

Risk 2 : **Fixed-Income** Funds

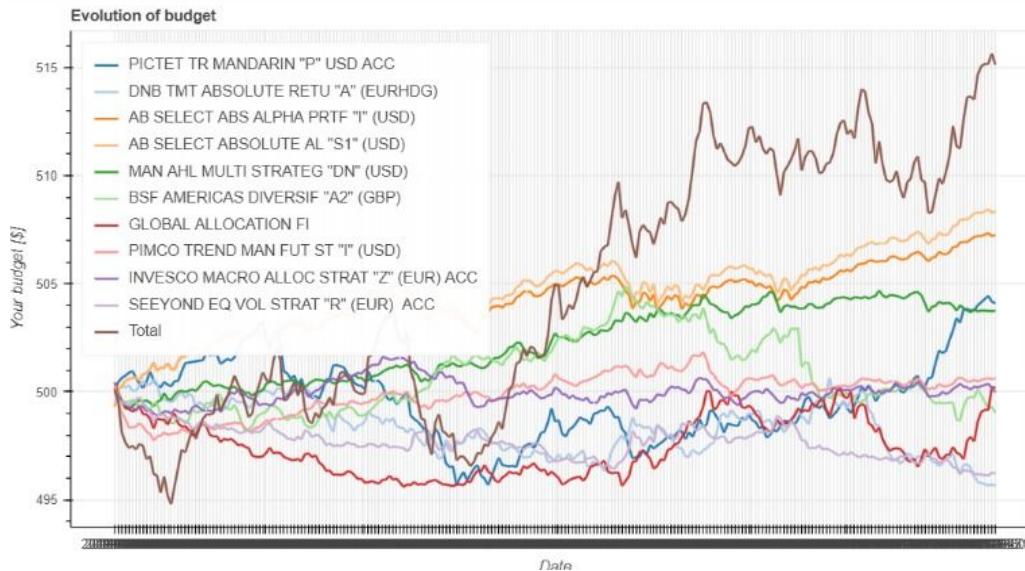


a. **determine** a risk level **for each** fund type

Risk 1 : **Monetary** Funds

Risk 2 : **Fixed-Income** Funds

Risk 3 : **Alternative** Funds



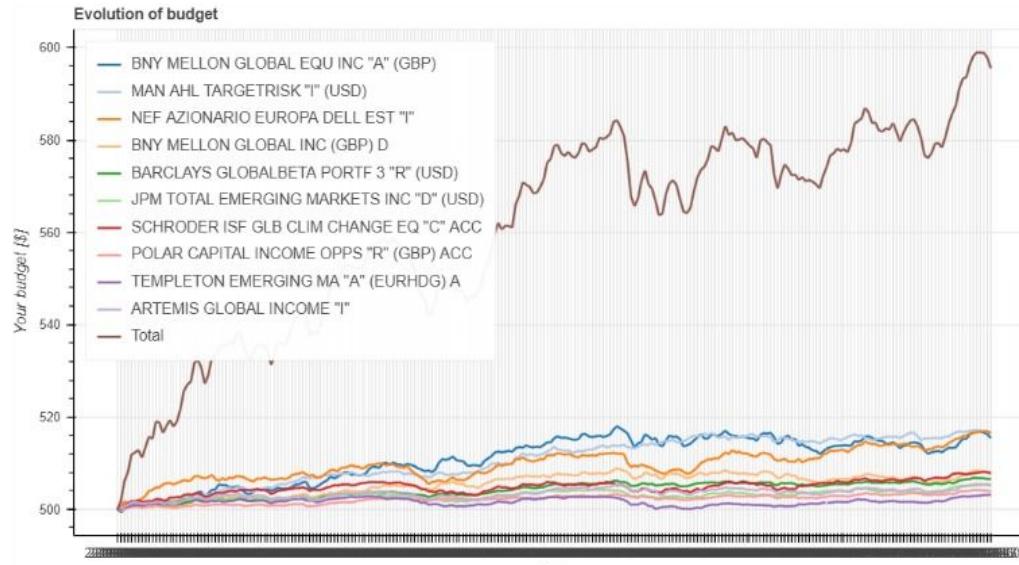
a. **determine** a risk level **for each** fund type

Risk 1 : **Monetary** Funds

Risk 2 : **Fixed-Income** Funds

Risk 3 : **Alternative** Funds

Risk 4 : **Mixed** Funds



a. **determine** a risk level **for each** fund type

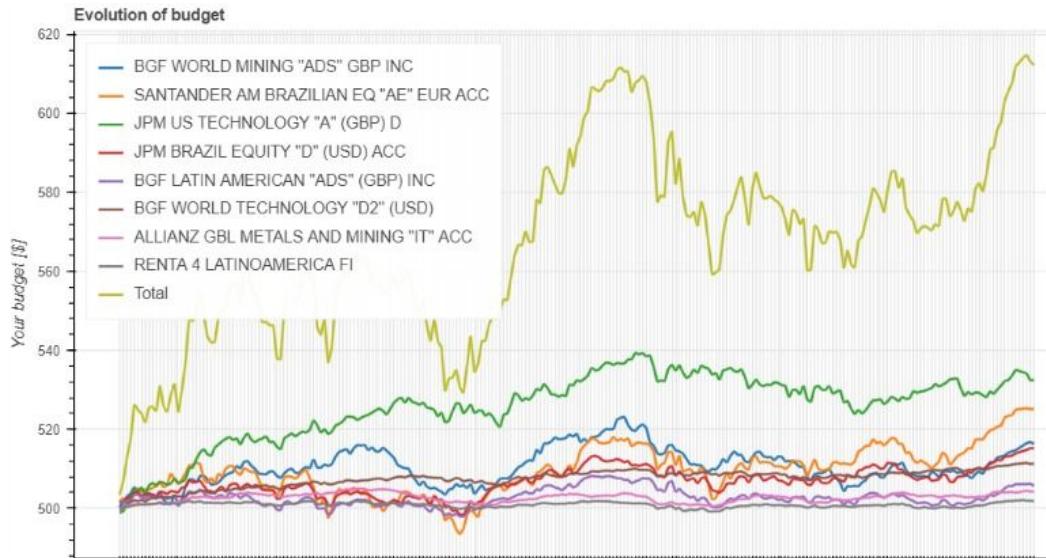
Risk 1 : **Monetary** Funds

Risk 2 : **Fixed-Income** Funds

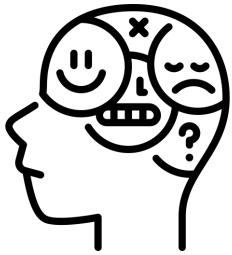
Risk 3 : **Alternative** Funds

Risk 4 : **Mixed** Funds

Risk 5 : **Equity** Funds

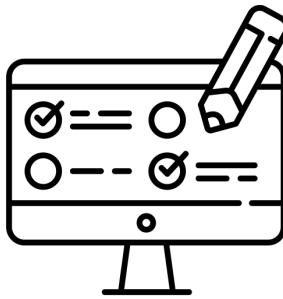
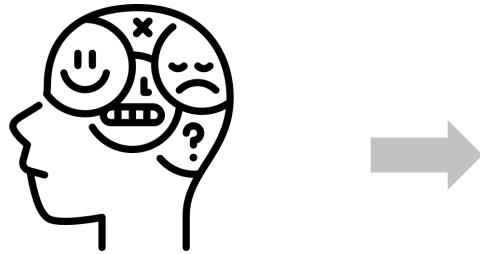


b. **determine** the risk aversion **for each** investor



**b. determine the risk aversion **for each** investor**

Risk Assessment Questionnaire

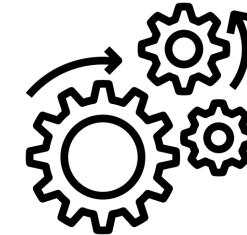
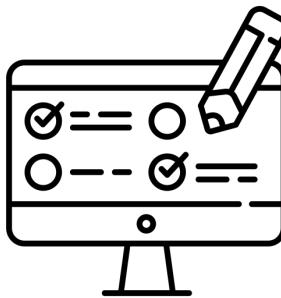


7 questions

- Answer **A** (high aversion)
- Answer **B** (mid-high aversion)
- Answer **C** (mid-low aversion)
- Answer **D** (low aversion)

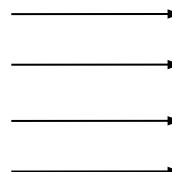
**b. determine the risk aversion **for each** investor**

Risk Assessment Questionnaire



7 questions

- Answer **A** (high aversion)
- Answer **B** (mid-high aversion)
- Answer **C** (mid-low aversion)
- Answer **D** (low aversion)



**4 points** each  
**3 points** each  
**2 points** each  
**1 points** each

sum of the user answers



risk aversion score	result of the risk assessment	interval constraining the risk mean
24-28	minimum risk profile	$R \in [1, 2]$
18-23	mid-low risk profile	$R \in [2, 3]$
12-17	mid-high risk profile	$R \in [3, 4]$
7-11	maximum risk profile	$R \in [4, 5]$

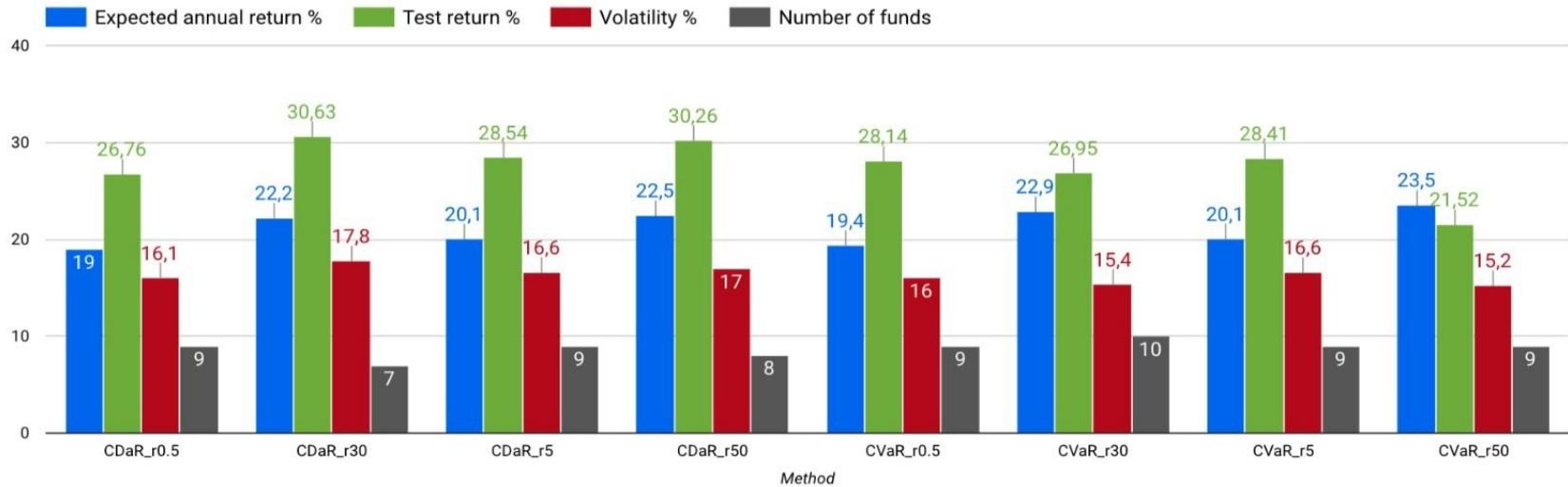


risk mean of the funds of the portfolio

# risk tailoring results

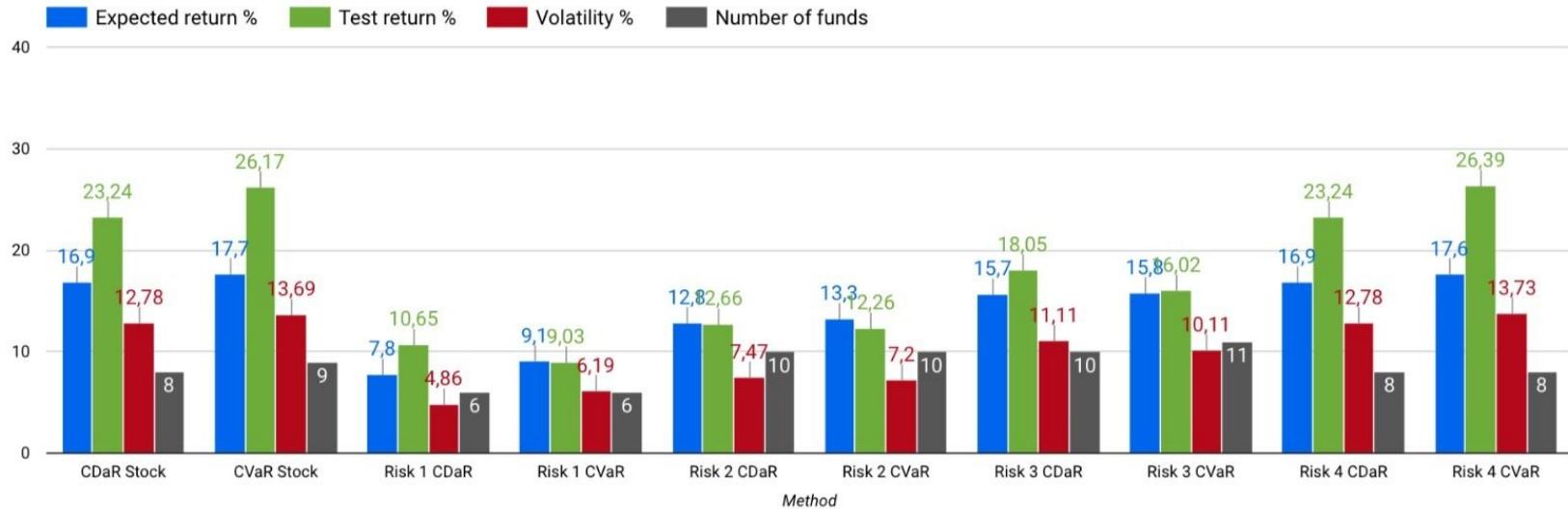
# risk tailoring results

## previous approach



# risk tailoring results

## improved approach



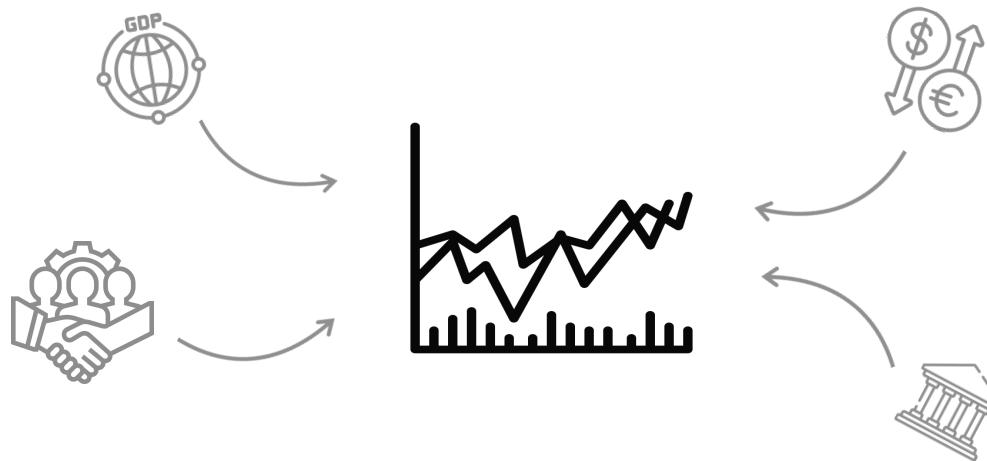


**no more funds**  
**let's talk about**  
**the market**



# **no more funds**

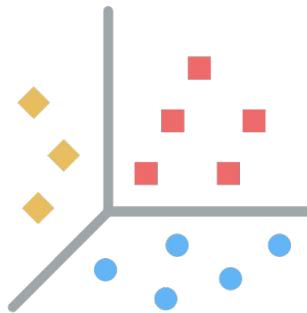
## let's talk about the market



## web scraping



## clustering



## optimization

**OPTIMAL  
PORTFOLIO**  
for each cluster

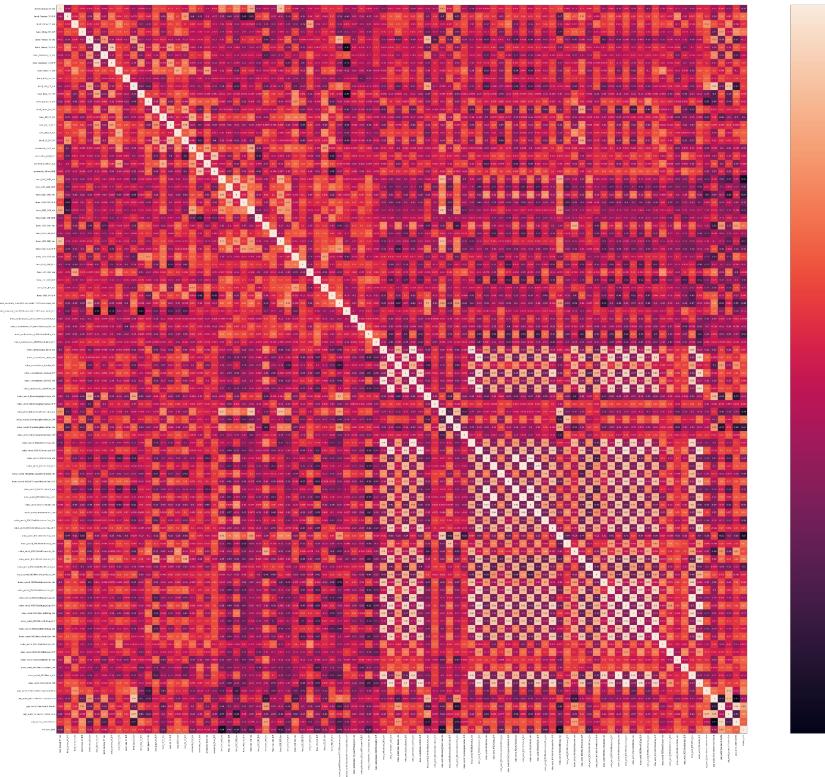
# but that is...

- **6** world regions **GDP**
- **8** currencies **exchange** rate
- 1 country (**US**) **inflation** rate
- **25** global industries **indexes**
- **9** countries 5-year **bonds**
- **2 commodities**

x 365 days  
x 20 years

x 5 timeframes for which weighted variance is computed (1 month ago, 3 months ago...)

# but that is...

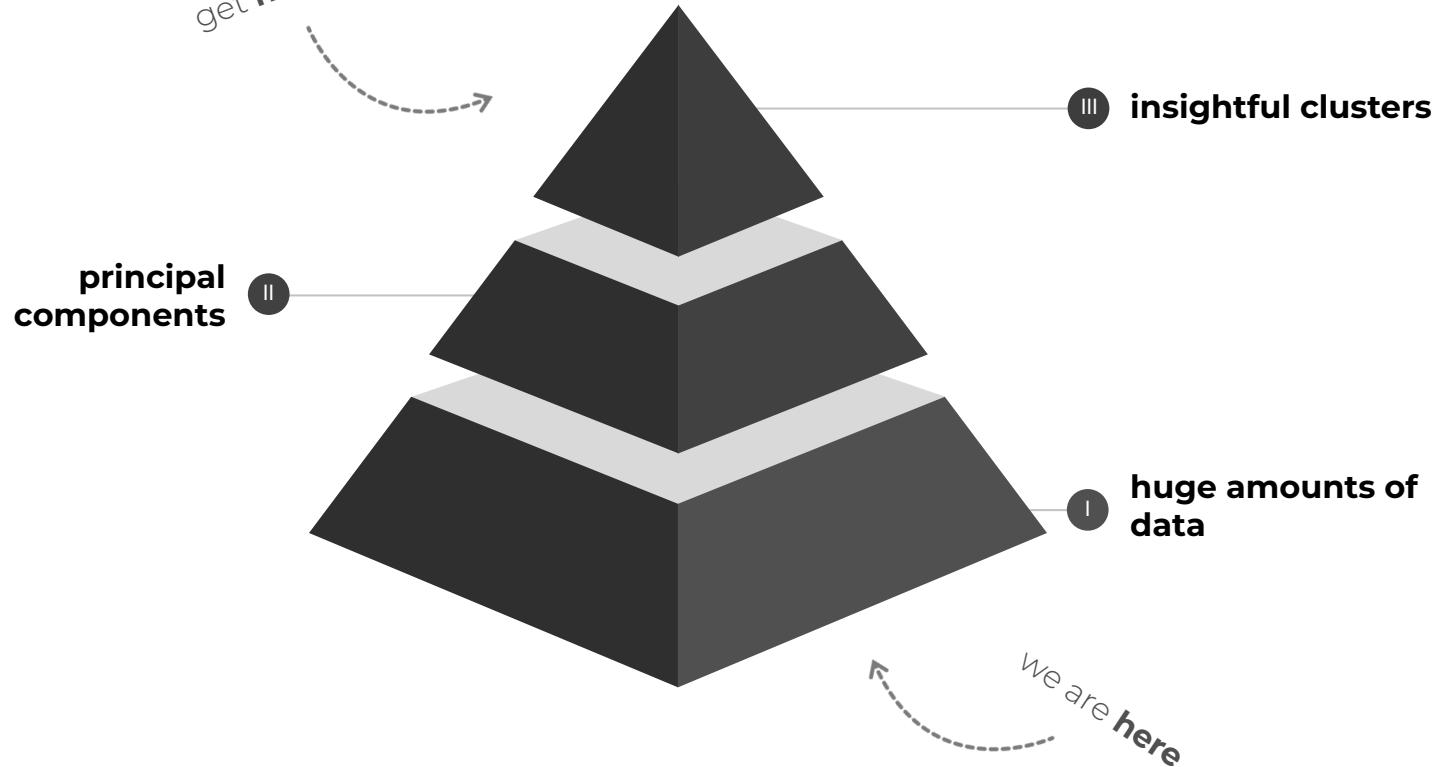


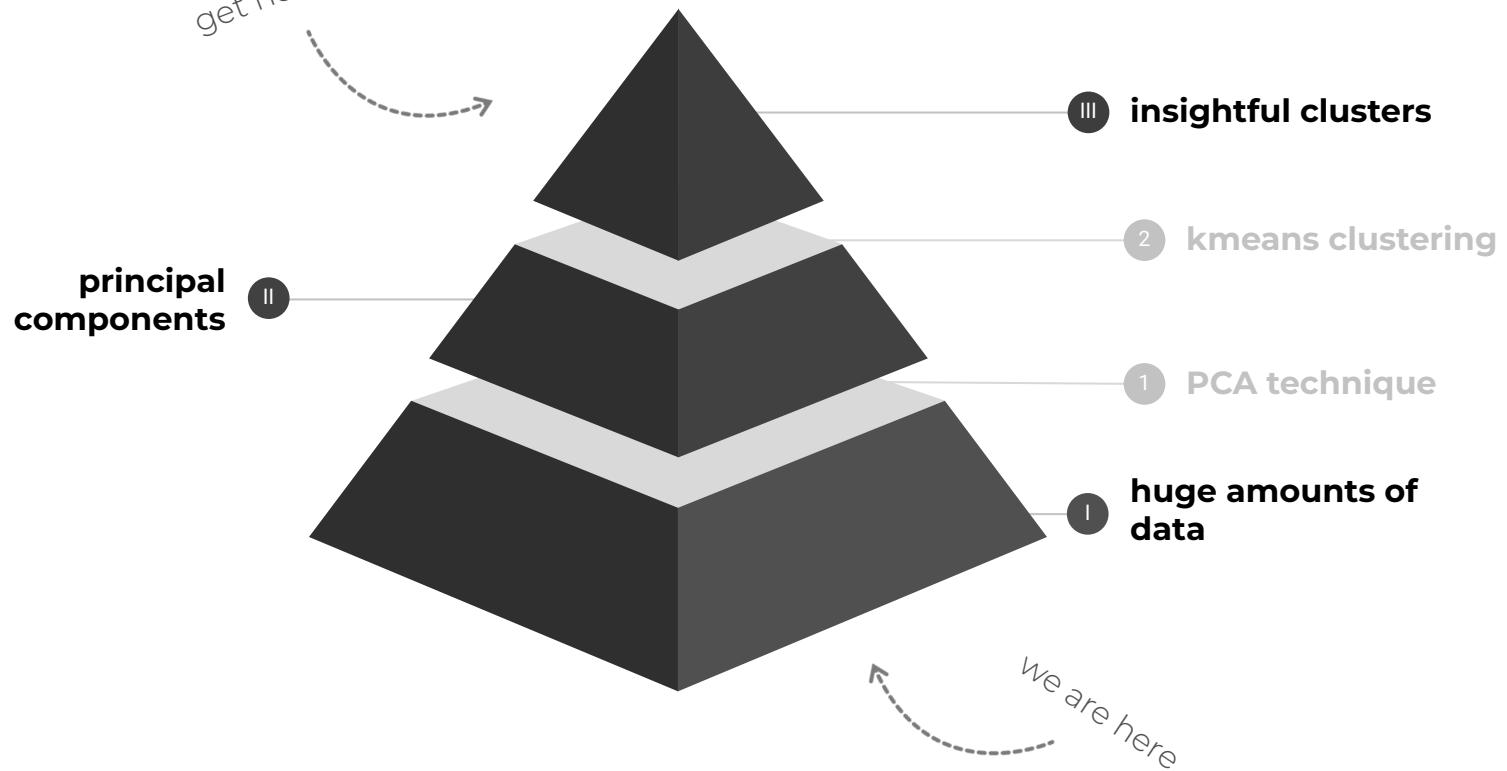
- **6** world regions **GDP**
- **8** currencies **exchange** rate
- 1 country (**US**) **inflation** rate
- **25** global industries **indexes**
- **9** countries 5-year **bonds**
- **2 commodities**

x 365 days  
x 20 years

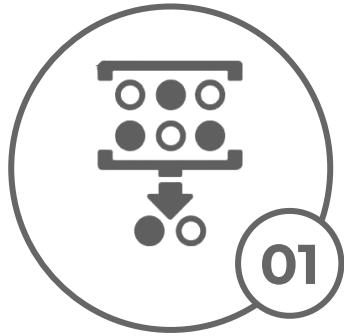
x 5 timeframes for which weighted variance is computed (1 month ago, 3 months ago...)

**... a lot of data.**





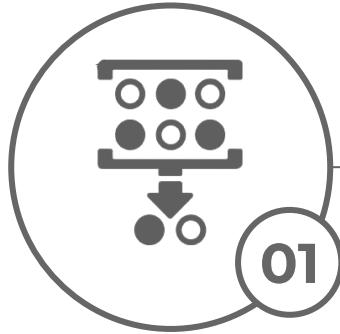
# untangling the data



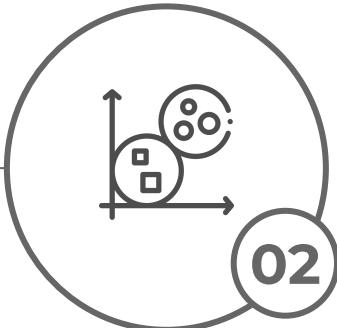
01

**principal  
component** analysis

# untangling the data

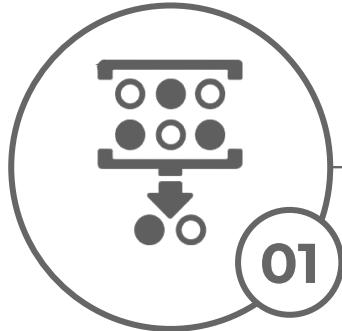


**principal  
component** analysis

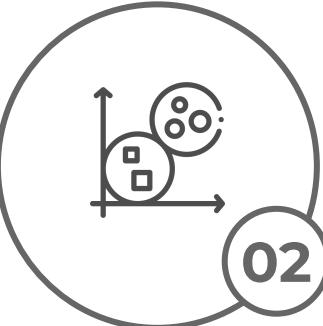


**k-means** clustering

# untangling the data



**principal**  
**component** analysis



**k-means** clustering

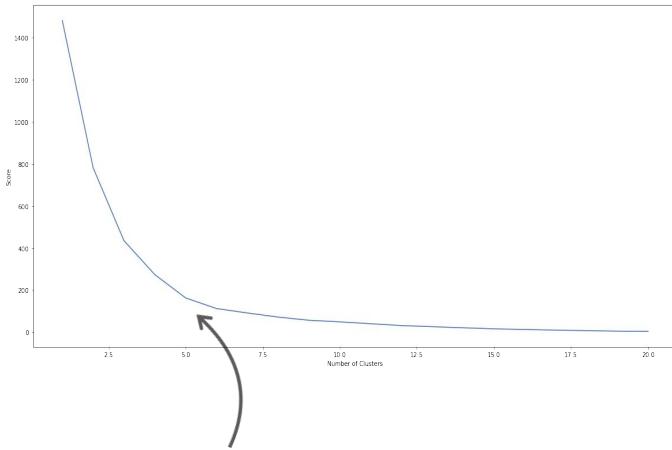


**distance** assimilation

# clustering

## k-means

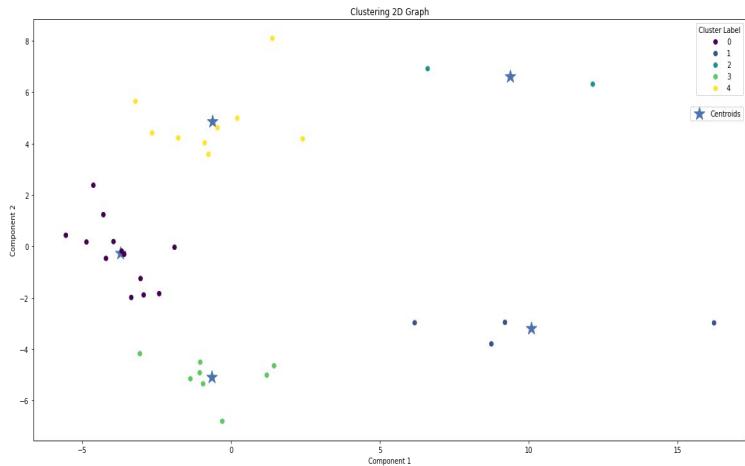
Elbow Curve



number of **centroids** to parametrize

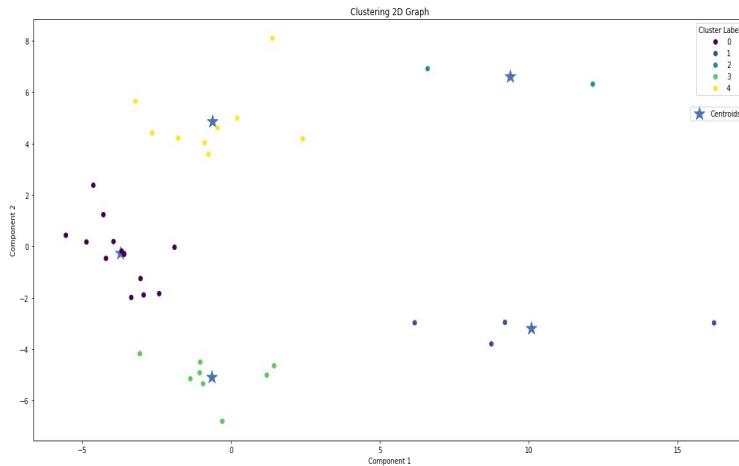
# clustering

## k-means

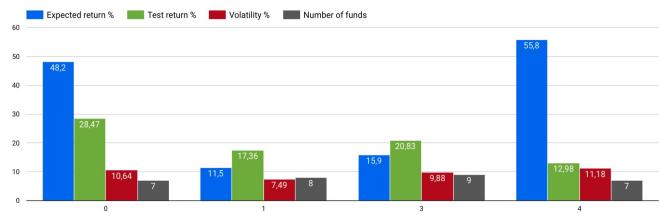
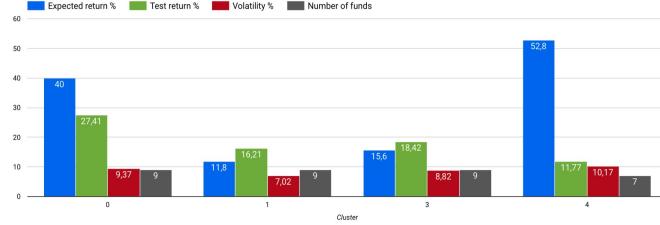
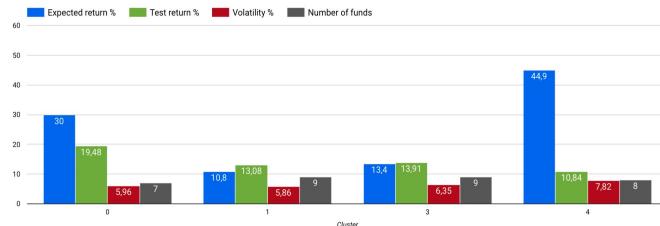
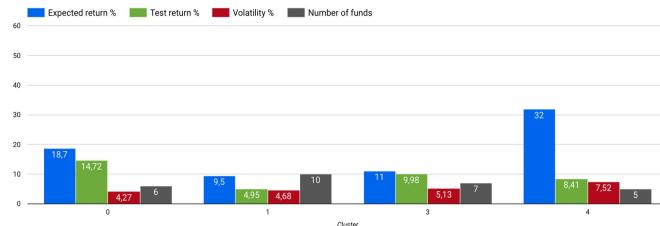


**5 centroids = 5 clusters**

# clustering k-means



5 centroids = 5 clusters



**now what ?**

**¿ahora qué ?**

# IroAdvisor\_v1.01

## **producto minimo viable**



# IroAdvisor\_v1.01

## producto minimo viable



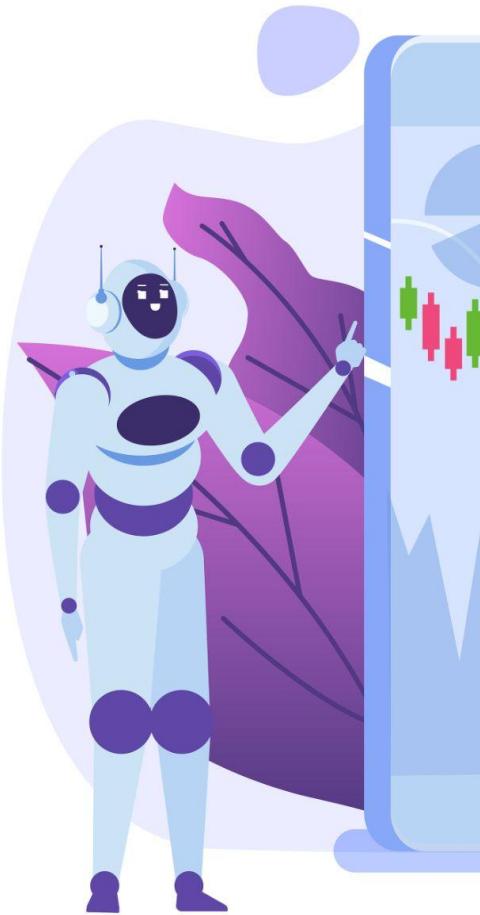
front-end

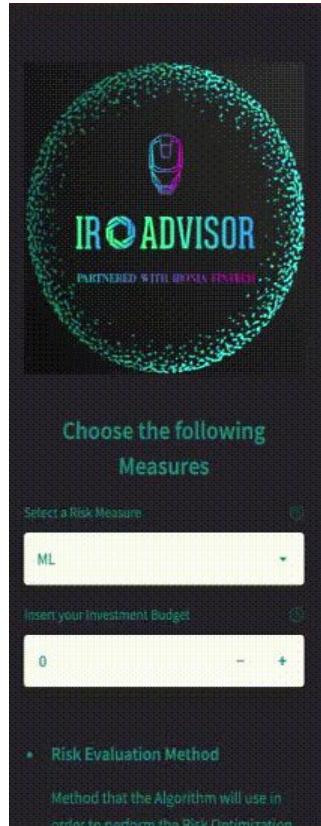


- 01. recibir inputs (**capital** y **riesgo**)
- 02. asignar perfil de **aversión**
- 03. calcular **intervalo de riesgo**
- 04. aplicar **restricciones**
- 05. **optimización** jerárquica
- 06. **portfolio** óptimo



back-end





# IROADVISOR (Beta)

## Your personal fund portfolio optimizer

### Risk Assessment Questionnaire

If you have low financial knowledge, we recommend you to fill this Questionnaire

Running Questionnaire

### User Portfolio

Manage app



**IRO ADVISOR**  
Your personal fund portfolio optimizer

Choose the following Measures

Select the Measure:  
CVar

Enter your investment budget:  
2000

Risk Evaluation Method  
Method that the algorithm will use to identify suitable funds for optimization.  
(We recommend to use CVar or CVaR)

Budget  
Amount of money the Client is willing to invest.

# IROADVISOR (Beta)

## Your personal fund portfolio optimizer

---

### Risk Assessment Questionnaire

If you have low financial knowledge, we recommend you to fill this Questionnaire

We will ask you 7 questions with the aim of getting to know you better and in this way discard certain funds.

1. If you had to choose between more job security with a small pay increase and less job security with a big pay increase, which would you pick?

A. Definitely more job security with a small pay increase  
 B. Probably more job security with a small pay increase  
 C. Probably less job security with a big pay increase  
 D. Definitely less job security with a big pay increase

2. Imagine you were in a job where you could choose to be paid salary, commission, or a mix of both. Which would you pick?

A. All salary  
 B. Mainly salary

3. Which of the statements better reflect the way you feel in situations in which you have little to no control over the outcome?

A. I tend to panic and start making bad decisions.  
 B. I feel powerless and start overthinking.  
 C. I get a bit nervous but I let the situation develop.  
 D. I remain completely calm.

4. Of the following investments, which of the following scenarios would you be most comfortable with?

A. You can lose down to -2%, and gain up to +9%  
 B. You can lose down to -17%, and gain up to +13%

[Manage app](#)

**IRO ADVISOR**  
PARTNERED WITH IBERIA FINTECH

Choose the following Measures

Select a Risk Measure: CVaR

insert your Investment Budget: 2000

Risk Evaluation Method: Method that the Algorithm will use in order to perform the Risk Optimization (we recommend to use CVaR or CDaR).

Budget: Amount of money the Client is willing to invest.

PARTNERED WITH



### Volatility

0.133%

### Total Returns

23.79%

### Capital Earned

475.86\$

### Capital after 1 Year

2475.86\$

### Additional Fund's Information

Names	Benchmark Id	Budget Inversion	Risk Level	Category	Benchmark	Morningstar Category Id
THREADNEEDLE LATIN AMERICA "Z"	5e809c91c6805b4d89f1686b	187.256\$	5	Renta Variable	MSCI EM Latin America Net Total Return USD Index	EUCA000524
JPM BRAZIL EQUITY "D" (USD) ACC	5f5f81bba2871d8612946884	342.576\$	5	Renta Variable	MSCI Brazil Net Total Return USD Index	EUCA000699
BGF WORLD TECHNOLOGY "D2" (USD)	5e32bbf59fd58c15144cdcf8	227.676\$	5	Renta Variable	NASDAQ Composite Index	EUCA000542
T ROWE P US BLUE CHIP EQ "Q" (GBP)	5e7ff052bb5924d830f3ccb	98.436\$	5	Renta Variable	S&P 500 Growth Index	EUCA000527
JPM US TECHNOLOGY "D"	5e32bbf59fd58c15144cdcf8	101.176\$	5	Renta Variable	NASDAQ Composite Index	EUCA000542
TEMPLETON LATIN AMERICA "W" (GBP)	5e809c91c6805b4d89f1686b	193.596\$	5	Renta Variable	MSCI EM Latin America Net Total Return USD Index	EUCA000524
BGF WORLD MINING "D4" (GBP) D	5e808531c6805b4d89f16867	256.356\$	5	Renta Variable	S&P Global Natural Resources Net Total Return Index	EUCA000704
BGF WORLD MINING "ADS" GBP INC	5e808531c6805b4d89f16867	247.816\$	5	Renta Variable	S&P Global Natural Resources Net Total Return Index	EUCA000704
BGF WORLD TECHNOLOGY "E2"	5e32bbf59fd58c15144cdcf8	134.196\$	5	Renta Variable	NASDAQ Composite Index	EUCA000542
T ROWE P US LAR CAP GRW EQ "Q" (GBP)	5e7ff052bb5924d830f3ccb	210.936\$	5	Renta Variable	S&P 500 Growth Index	EUCA000527

**thank you.**