

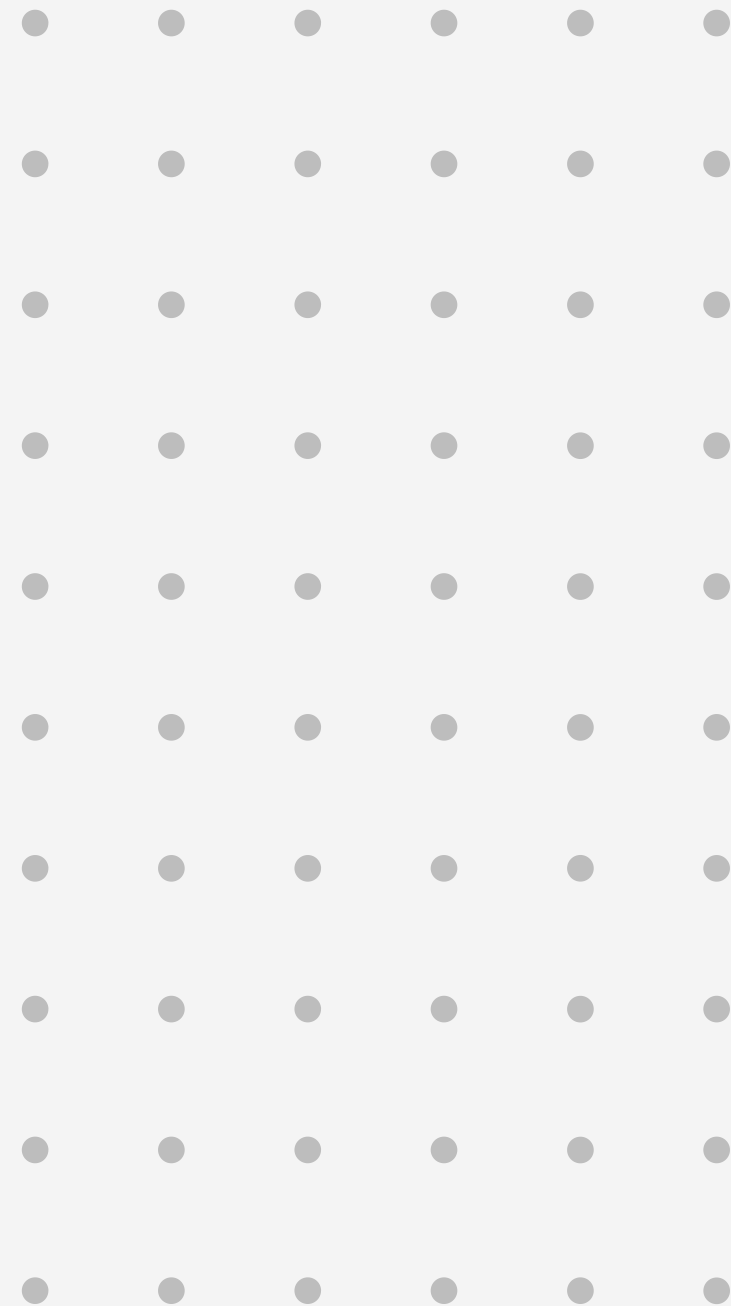


# IAVE PROJECT

ISCTE-IUL

12/03/2025

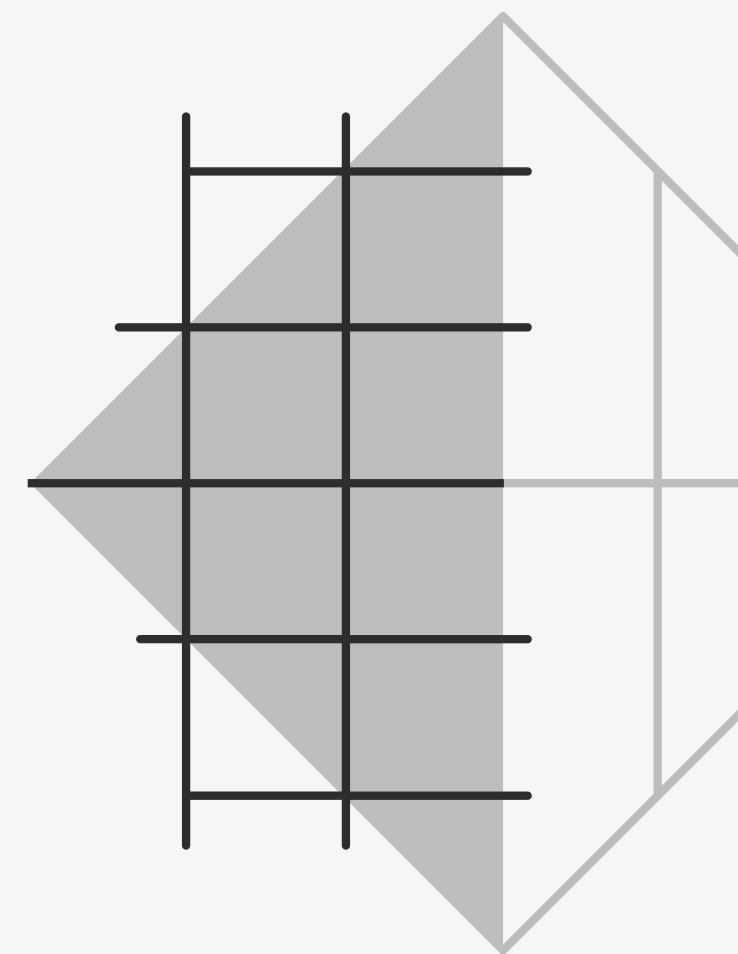
André Costa, Celso Ribeiro  
e David Mendes






# Objetivos

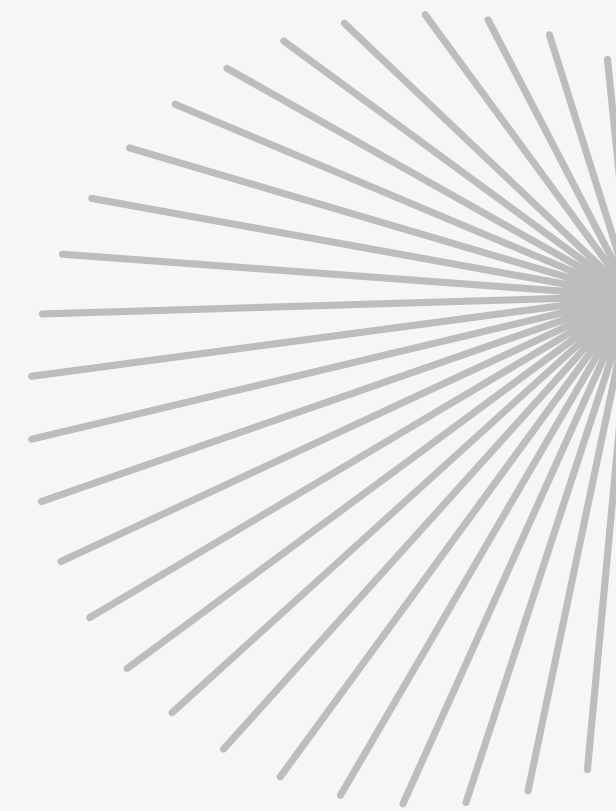
- Escolher uma base de dados otimizada para Big Data
- Descodificar e perceber os dados
- Efetuar uma limpeza nos dados que não sejam úteis ou estejam vazios.
- Visualizar graficamente os dados (matriz de correlações, feature importances, pca).





# Avanços

- Criação de um MongoDB
  - Decodificação dos dados com base na documentação oficial
  - Limpeza inicial dos dados
- 



# Filtros

REPEAT



Filtrar o número total de alunos para apenas alunos repetentes

- De ~600k para ~60k

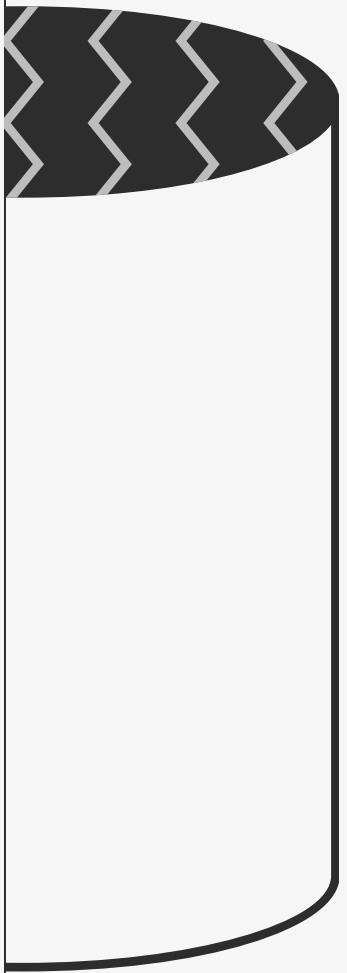
PV1MATH – PV10MATH

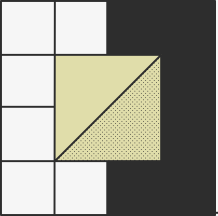


Nova coluna **MATH\_RESULT** com a média dos valores a matemática

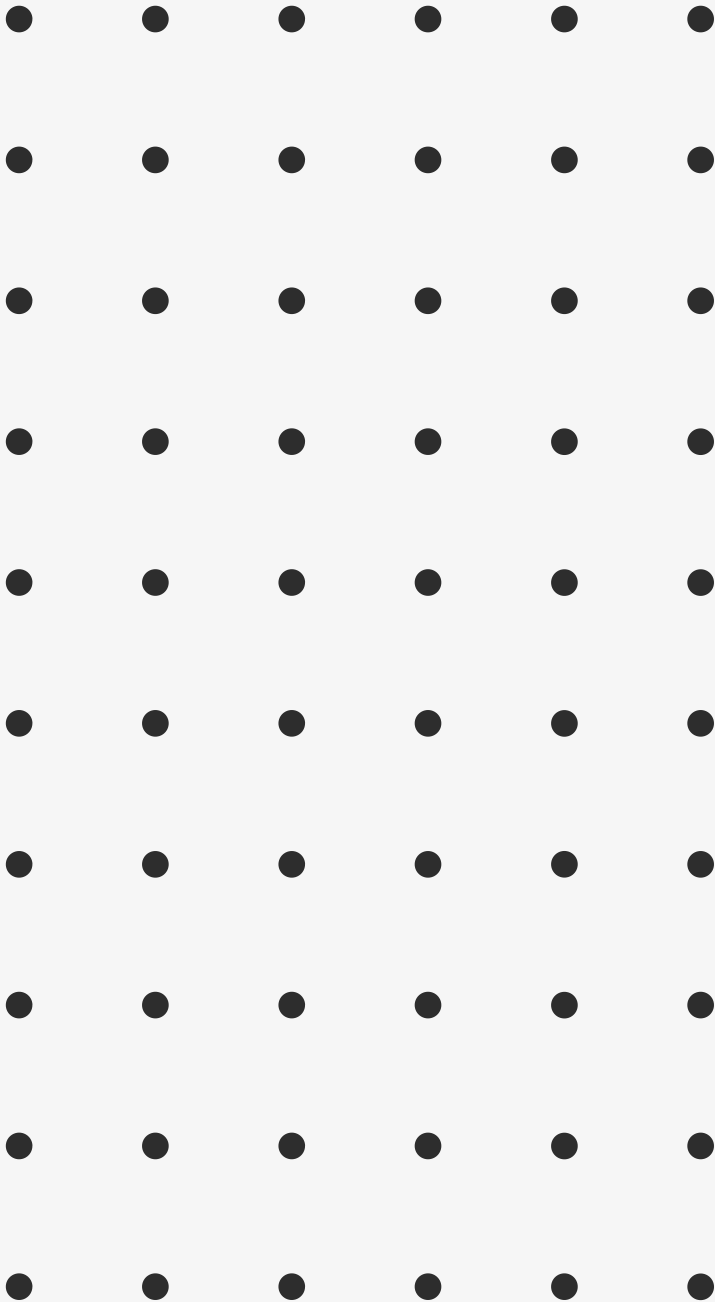


# Próximas implementações

- 
- Criar um notebook com os objetivos de negócio detalhados
  - Conclusão da parte “Data Understanding”
  - Visualizar graficamente os dados (matriz de correlações, feature importances)



# Obrigado!



ISCTE-IUL

12/03/2025  
André Costa, Celso Ribeiro  
e David Mendes