



Escola Politécnica de Pernambuco
Especialização em Ciência de Dados e Analytics

Estatística Computacional

Aula 4.2 – Avaliação Prática de Análise de Conglomerados

Prof. Dr. Rodrigo Lins Rodrigues

rodrigo.linsrodrigues@ufrpe.br

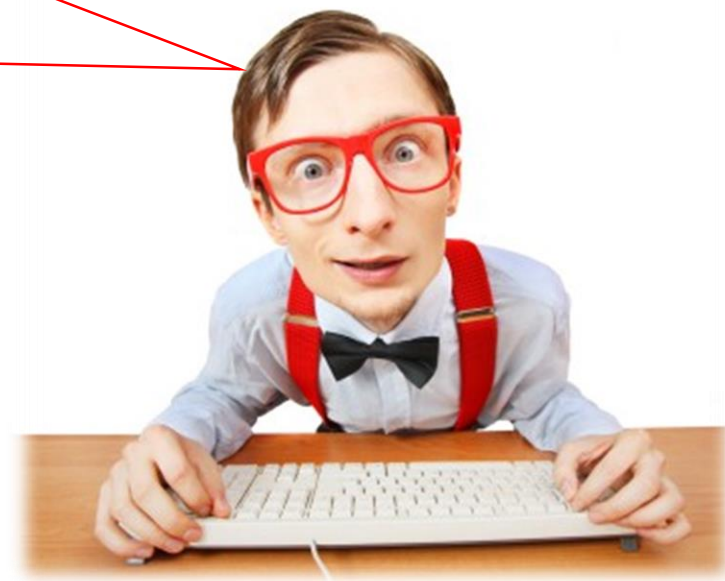
Como será esta aula?

- ✓ Avaliação prática em laboratório;
- ✓ Deve ser realizada em **grupos de 4** alunos;
- ✓ **Serei o “mentor”** na execução da avaliação prática.



O que você aprenderá ?

Colocará em prática todo o conhecimento adquirido sobre **análise de conglomerados**



[illegible]

Bases de Dados

- Plataformas de aprendizagem online e EAD.



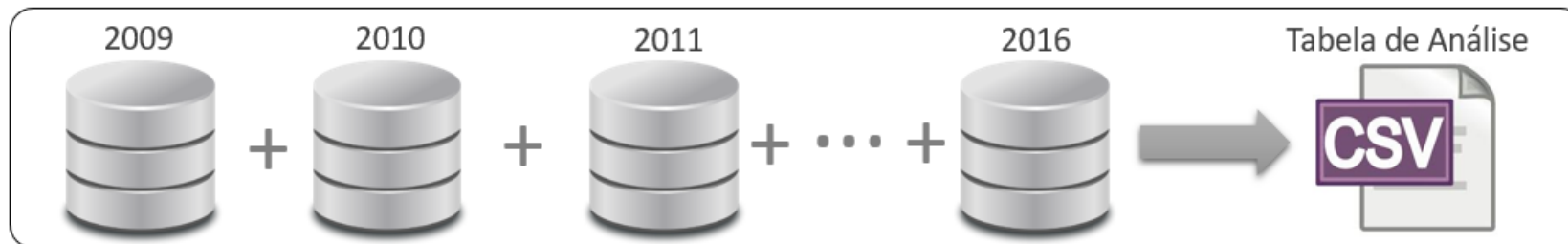
Bases de Dados

- Comportamento de alunos em sistema de EAD.
 - ✓ Logs de acessos diários de alunos;
 - ✓ Interações com colegas;
 - ✓ Acesso a recursos e materiais;
 - ✓ Tempos de permanência na plataformas;
 - ✓ Desempenho dos alunos;
 - ✓ ...



Bases de Dados

- Sobre a base de dados:
 - ✓ Foi realizada a junção das bases de dados referente a um período de sete anos;
 - ✓ Dados armazenados em bases de dados relacionais;
 - ✓ Construção de 37 scripts SQL;



Bases de Dados

- Sobre as variáveis comportamentais:

Variável	Descrição sobre as variáveis
VAR01	Quantidade de diferentes locais (IP's) a partir dos quais a(o) aluna(o) acessou o ambiente.
VAR02	Quantidade de mensagens enviadas por aluna(o) às(os) Professoras(es) pelo ambiente.
VAR03	Quantidade de mensagens enviadas por aluna(o) às(os) Tutor(es) pelo ambiente.
VAR04	Quantidade geral de mensagens enviadas pela(o) aluna(o) dentro do ambiente.
VAR05	Quantidade geral de mensagens recebidas pela(o) aluna(o) dentro do ambiente.
VAR06	Quantidade de tópicos criados pelo(a) aluno(a) em fórum do tipo "tira-dúvidas".
VAR07	Quantidade de postagens no "Fórum tira dúvidas";
VAR08	Quant. de postagens de um(a) aluno(a) em fóruns que foram respondidas por outros(as) alunos(as).
VAR09	Quantidade de postagens de um(a) aluno(a) em fóruns que foram respondidas pelo(a) professor(a) ou tutor(a).
VAR10	Quantidade de colegas diferentes para quem o(a) aluno(a) enviou mensagens dentro do ambiente.
VAR12	Quantidade de visualizações da aba "Conteúdo" do curso, onde constam os arquivos com o conteúdo programático do curso
VAR13	Horário que mais realizou atividades;
VAR14	Turno do dia em que realizou mais atividades.
VAR16	Quantidade de atividades entregues por um(a) aluno(a) fora do prazo, por disciplina;
VAR17	Tempo médio entre a abertura da atividade e sua submissão;
VAR18	Quantidade de leituras feitas ao fórum (pageviews);

Bases de Dados

- Sobre as variáveis comportamentais:

VAR20	Quantidade de respostas ao tópico principal (refazer opinião em fórum);
VAR21	Quantidade de pageviews ao quadro de notas;
VAR22	Quantidade de vezes que o aluno visualiza o (Cheklist Atividades)
VAR23	Quantidade de visualizações de notas por atividade;
VAR24	Média semanal da quantidade de acessos de um(a) aluno(a) ao ambiente.
VAR25	Tempo médio entre a criação de um tópico no fórum temático e a primeira postagem do aluno;
VAR28	Quantidades de Time Out;
VAR31	Quantidade de acessos do(a) aluno(a) ao ambiente.
VAR31b	Quantidade de dias distintos que o aluno entrou na disciplina
VAR31c	Quantidade de dias distintos que o aluno entrou na plataforma
VAR32a	Quantidade de acessos do(a) aluno(a) ao ambiente por turno (Manhã).
VAR32b	Quantidade de acessos do(a) aluno(a) ao ambiente por turno (Tarde).
VAR32c	Quantidade de acessos do(a) aluno(a) ao ambiente por turno (Noite).
VAR32d	Quantidade de acessos do(a) aluno(a) ao ambiente por turno (Madrugada).
VAR33	Quantidade de atividades entregues por um(a) aluno(a) no prazo, por disciplina;
VAR34	Quantidade geral de postagens de um(a) aluno(a) em fóruns.
VAR35	Quantidade de respostas de um(a) professor(a) para as dúvidas de alunos(as) em fóruns.



Projeto prático

Projeto prático

- Proposta do projeto prático.
 - ✓ Realizar uma **análise de conglomerados**, baseado em dados comportamentais, com o objetivo de descobrir padrões comportamentais de alunos;
 - ✓ Cada grupo deverá utilizar o máximo de conhecimentos de pré-processamento para otimizar a análise.



Projeto prático

- Fase de pré-processamento:
 - ✓ Quando necessário padronizar variáveis;
 - ✓ Retirar variáveis quando necessário;
 - ✓ Verificar variáveis que tem uma grande quantidade de zeros e decidir o que fazer;
 - ✓ Verificar a presença de *outliers*;
 - ✓ ...

Projeto prático

- **Quais análises deverão ser feitas?**
 - ✓ Selecionar um grupo de 10 variáveis mais importantes;
 - ✓ Padronizar as variáveis pelo método do *Z score*;
 - ✓ Aplicar a melhor medida de similaridade;
 - ✓ Aplicar método hierárquico para identificar a quantidade de grupos;
 - ✓ Aplicar o método não-hierárquico para separar os grupos;
 - ✓ Validar o agrupamento;
 - ✓ Aplicar medidas descritivas em cada grupo encontrado.

Dúvidas



- **Contatos:**

- ✓ Email: rodrigo.linsrodrigues@ufrpe.br
- ✓ Facebook: [/rodrigomuribec](https://www.facebook.com/rodrigomuribec)