

Limpeza de Dados -Valores Faltantes





- Aula 14 -Coleta, Preparação e Análise de Dados

Prof. Me. Lucas R. C. Pessutto

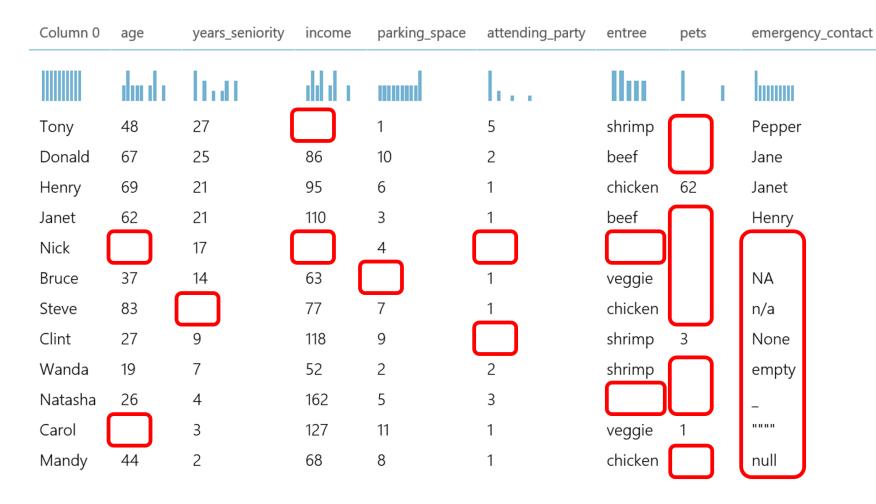
Slides adaptados do material do Prof. Lucas Silveira Kupssinskü e do Prof. Luan Fonseca Garcia

Data Cleaning

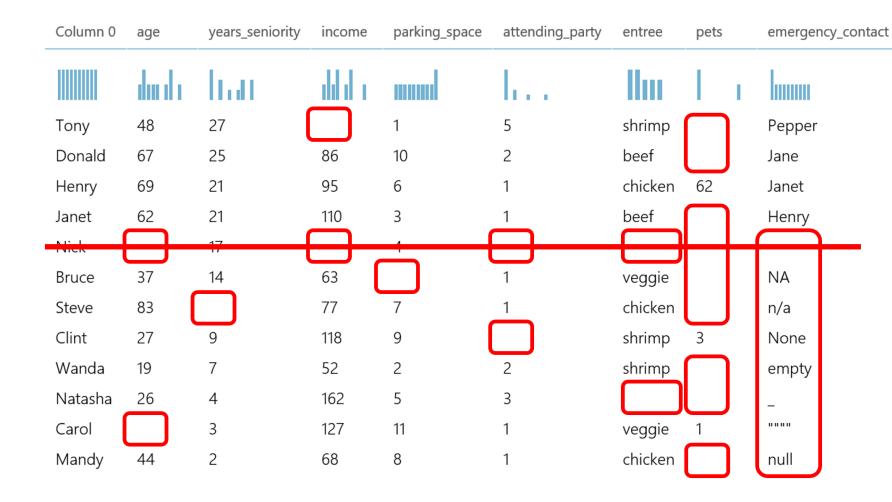
- O processo de limpeza dos dados pode ser separado em dois processos:
 - Pré-processamento (ou imputação) de valores faltantes.
 - Remoção de Ruído

Dados faltantes

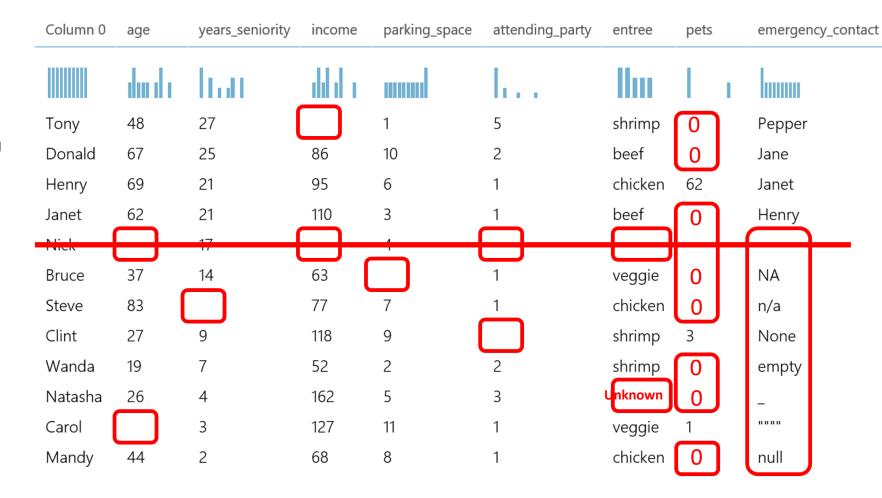
- Dados faltantes podem ser ocasionados por diversos motivos
 - Problemas na coleta;
 - Observações que faltaram coletar;
 - Podem apresentar informação.
- Mesmo modelos robustos contra dados faltantes podem se beneficiar desse tipo de processamento.



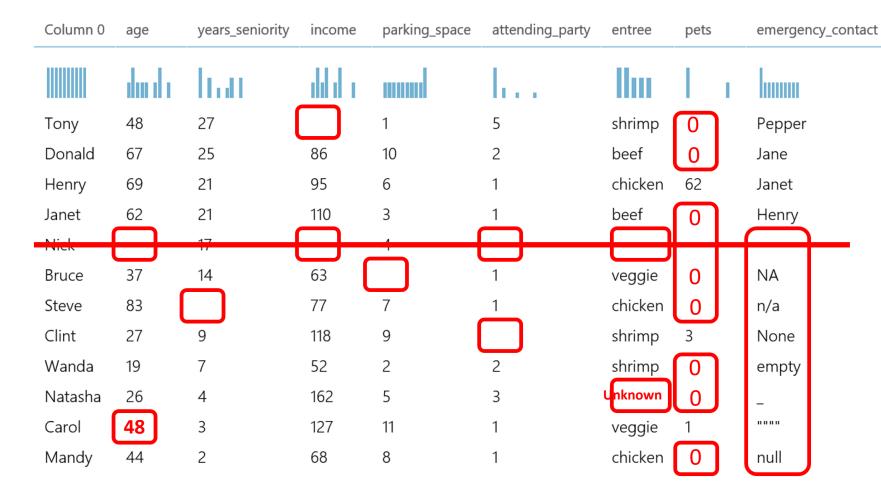
- Deleção:
 - Remover toda a observação.



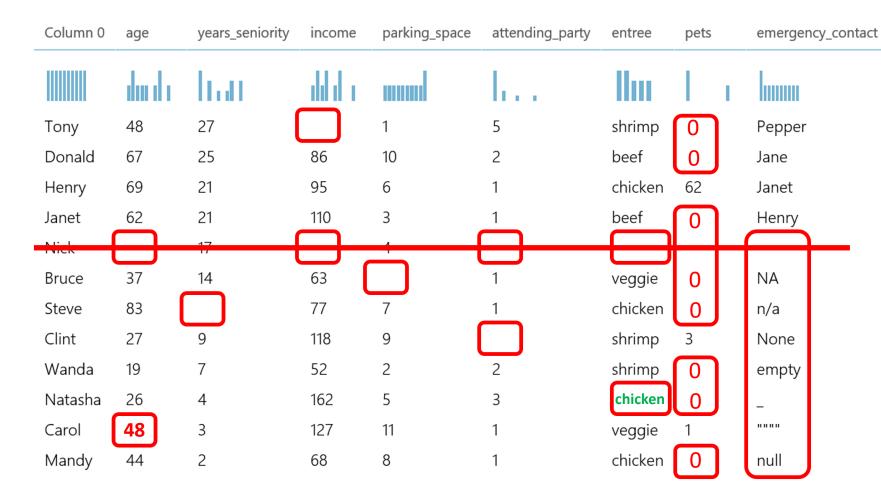
- Deleção:
 - Remover toda a observação.
- Substituição Dummy:
 - Substituir por um valor dummy (0 ou Unknow).



- Deleção:
 - Remover toda a observação.
- Substituição Dummy:
 - Substituir por um valor dummy (0 ou Unknow).
- Substituição pela Média:
 - Substituir pelo valor médio.



- Deleção:
 - Remover toda a observação.
- Substituição Dummy:
 - Substituir por um valor dummy (0 ou Unknow).
- Substituição pela Média:
 - Substituir pelo valor médio.
- Substituição por item frequente:
 - Substituir pela moda.



- Deleção:
 - Remover toda a observação.
- Substituição Dummy:
 - Substituir por um valor dummy (0 ou Unknow).
- Substituição pela Média:
 - Substituir pelo valor médio.
- Substituição por item frequente:
 - Substituir pela moda.
- Substituição por regressão/class.
 - Usar um método de regressão ou classificação para estimar o valor.

