

Aluno: Iago Batista Antunes Leobas

Disciplina: Mineração de Dados

Data: 29/08/2022

Aula 03 - Prática 1 e 2 de mineração de dados

Prática 1:

- Descrever os tipos de dados e informações dos atributos

Atributo	Valores	Tipos
1 - age	Quantitativa	Discreto
2 - job	Qualitativo	Nominal
3 - marital	Qualitativo	Nominal
4 - education	Qualitativo	Ordinal
5 - default	Qualitativo	Nominal
6 - housing	Qualitativo	Nominal
7 - loan	Qualitativo	Nominal
8 - contact	Qualitativo	Nominal
9 - month	Qualitativo	Ordinal
10 - day_of_week	Qualitativo	Ordinal
11 - duration	Qualitativo	Ordinal
12 - campaign	Qualitativo	Ordinal
13 - pdays	Qualitativo	Ordinal
14 - previous	Quantitativo	Discreto
15 - poutcome	Qualitativo	Nominal

Prática 2:

- Distinção entre os datasets: tamanho e estrutura de cada
- Verificar valores nulos
- Criar categorias a partir das existentes
- Filtros para verificar atributos relacionados e com significância
- Agrupar atributos relacionados

- Histograma e outliers

1. Inicialmente é importado a biblioteca pandas e o banco completo para o notebook.
2. São apresentados os primeiros 30 dados da base completa.
3. São apresentados respectivamente a quantidade de dados e colunas na base de dados.
4. É verificado então se há dados nulos e para garantir já é dado o comando de apagar caso haja.
5. É então importado o outro banco, com menos dados, e apresentado os 30 primeiros itens.
6. É feita também a verificação de campos nulos.
7. É feito então o primeiro filtro, no caso filtrando os clientes que não trabalham, pois a partir do momento que um cliente não possui renda fixa, é pouco provável que este aplicará dinheiro no banco.
8. É importada então a biblioteca pandas e feita uma validação condicional adicionando uma nova coluna na base de dados mostrando se o cliente “é capaz para aplicar” passando como parâmetro o campo “poutcome” como diferente de “failure” pois se o cliente que já demonstraram que não tem interesse ao serviço no passado possivelmente não aceitará no futuro.
9. É feito então um agrupamento com os dados “Educação”, “Idade” e “Trabalho”, para melhor visualizar as informações dos clientes primeiramente por suas informações primordiais relacionando-as.
10. É importada então a biblioteca plotly.express para apresentar um histograma de idade, para melhor visualizar qual o maior idade entre os clientes.
11. E por fim é utilizada a biblioteca matplotlib.pyplot para encontrar outliers dentro da lógica onde é vista a duração das ligações com os clientes, visando a idade. É apontado dessa forma que quanto mais novo o cliente, mais tempo ele passa em ligação, salvo certos outliers.