|  |  |
| --- | --- |
|  | Curso de Data Science  Prof. MSc. Eng. Marcelo Bianchi |

**Trabalho 1 – Tipo 1 - Curso de Data Science**

Integrantes do Grupo:

Nome:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Nome:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Nome:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Nome:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Nome:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. **(5,0 pontos) Regressão Linear Simples**

Dado o dataset abaixo, calcule a regressão linear simples programando em Python

|  |  |
| --- | --- |
| **Horas\_de\_Academia Semana** | **Nota Teste Aptidão Fisica** |
| 1 | 53 |
| 5 | 74 |
| 7 | 59 |
| 8 | 43 |
| 10 | 56 |
| 11 | 84 |
| 14 | 96 |
| 15 | 69 |
| 15 | 84 |
| 19 | 83 |

**Predição: Nota Teste Aptidão Física**

1. ( 0,5 ponto ) Importar o Data Set e aplicar a técnica Missing Data
2. ( 0,5 ponto ) Dividir o dataset entre o Training Set e o Test Set
3. ( 0,5 ponto ) Aplicar Feature Scaling (Se for aplicável, senão for então justificar)
4. ( 0,5 ponto ) Aplicar Dummy Variable (Se for aplicável, senão for então justificar)
5. ( 0,5 ponto) Aplicar a Simple Linear Regression com a técnica
6. ( 0,5 ponto ) Construir o Gráfico (Scatter Plot)
7. (0,5 ponto) Criar a tabela no banco de dados SQLite
8. (0,5 ponto) Aplicar uma consulta em linguagem SQL que irá trazer uma listagem da tabela
9. (0,5 ponto) Apresentação e explicação do exercício ao professor
10. (0,5 ponto) Responder uma dúvida ou questão do professor

**2) (5,0 pontos) Regressão Linear Múltipla**

Dado o dataset abaixo, calcule a regressão linear múltipla programando em python

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Idade** | **Sexo** | **Filhos** | **Região** | **Custo Seguro Saude** |
| 19 | female | 0 | southwest | 1688 |
| 18 | male | 1 | southeast | 1726 |
| 28 | male | 3 | southeast | 445 |
| 33 | male | 0 | northwest | 219845 |
| 32 | male | 0 | northwest | 3867 |
| 31 | female | 0 | southeast | 3757 |
| 46 | female | 1 | southeast | 8241 |
| 37 | female | 3 | northwest | 7282 |
| 37 | male | 2 | northeast | 6406 |
| 60 | female | 0 | northwest | 289231 |
| 25 | male | 0 | northeast | 2721 |
| 62 | female | 0 | southeast | 27809 |
| 23 | male | 0 | southwest | 183 |
| 56 | female | 0 | southeast | 11091 |
| 27 | male | 0 | southeast | 39612 |

**Predição: Custo Seguro Saúde**

1. ( 0,5 ponto ) Importar o Dataset e aplicar a técnica Missing Data
2. ( 0,5 ponto ) Dividir o dataset entre o Training Set e o Test Set
3. ( 0,5 ponto ) Aplicar Feature Scaling (Se for aplicável, senão for então justificar)
4. ( 0,5 ponto ) Aplicar Dummy Variable (Se for aplicável, senão for então justificar)
5. ( 0,5 ponto) Aplicar a Multiple Linear Regression com a técnica
6. ( 0,5 ponto ) Construir o Gráfico (Scatter Plot)
7. (0,5 ponto) Criar a tabela no banco de dados SQLite
8. (0,5 ponto) Aplicar uma consulta em linguagem SQL que irá trazer uma listagem da tabela
9. (0,5 ponto) Apresentação e explicação do exercício ao professor
10. (0,5 ponto) Responder uma dúvida ou questão do professor