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

**Trabalho 1 – Tipo 4 - 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 SimMultiple Linear Regression com a técnica de
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

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **bedrooms** | **bathrooms** | **floors** | **condition** | **grade** | **yr\_built** | **yr\_renovated** | **zipcode** | **Location** | House price |
| 3 | 1 | 1 | 3 | 7 | 1955 | 0 | 98178 | Paulista | 221900 |
| 3 | 225 | 2 | 3 | 7 | 1951 | 1991 | 98125 | Aclimação | 538000 |
| 2 | 1 | 1 | 3 | 6 | 1933 | 0 | 98028 | Campo Belo | 180000 |
| 4 | 3 | 1 | 5 | 7 | 1965 | 0 | 98136 | Mooca | 604000 |
| 3 | 2 | 1 | 3 | 8 | 1987 | 0 | 98074 | Paulista | 510000 |
| 4 | 45 | 1 | 3 | 11 | 2001 | 0 | 98053 | Campo Belo | 123000000 |
| 3 | 225 | 2 | 3 | 7 | 1995 | 0 | 98003 | Mooca | 257500 |
| 3 | 15 | 1 | 3 | 7 | 1963 | 0 | 98198 | Aclimação | 291850 |
| 3 | 1 | 1 | 3 | 7 | 1960 | 0 | 98146 | Centro | 229500 |
| 3 | 25 | 2 | 3 | 7 | 2003 | 0 | 98038 | Mooca | 323000 |
| 3 | 25 | 1 | 3 | 8 | 1965 | 0 | 98007 | Paulista | 662500 |
| 2 | 1 | 1 | 4 | 7 | 1942 | 0 | 98115 | Campo Belo | 468000 |
| 3 | 1 | 15 | 4 | 7 | 1927 | 0 | 98028 | Centro | 310000 |
| 3 | 175 | 1 | 4 | 7 | 1977 | 0 | 98074 | Paulista | 400000 |
| 5 | 2 | 15 | 3 | 7 | 1900 | 0 | 98107 | Centro | 530000 |
| 4 | 3 | 2 | 3 | 9 | 1979 | 0 | 98126 | Aclimação | 650000 |
| 3 | 2 | 2 | 3 | 7 | 1994 | 0 | 98019 | Campo Belo | 395000 |
| 4 | 1 | 15 | 4 | 7 | 1916 | 0 | 98103 | Mooca | 485000 |
| 2 | 1 | 1 | 4 | 7 | 1921 | 0 | 98002 | Paulista | 189000 |
| 3 | 1 | 1 | 4 | 7 | 1969 | 0 | 98003 | Paulista | 230000 |
| 4 | 175 | 1 | 4 | 7 | 1947 | 0 | 98133 | Centro | 385000 |
| 3 | 275 | 1 | 3 | 9 | 1968 | 0 | 98040 | Aclimação | 200000000 |
| 5 | 25 | 2 | 3 | 8 | 1995 | 0 | 98092 | Mooca | 285000 |
| 2 | 15 | 1 | 3 | 7 | 1985 | 0 | 98030 | Centro | 252700 |
| 3 | 225 | 2 | 4 | 8 | 1985 | 0 | 98030 | Mooca | 329000 |
| 3 | 2 | 15 | 5 | 6 | 1941 | 0 | 98002 | Centro | 233000 |
| 3 | 175 | 2 | 3 | 8 | 1915 | 0 | 98119 | Paulista | 937000 |
| 3 | 1 | 15 | 5 | 8 | 1909 | 0 | 98112 | Campo Belo | 667000 |

**Predição: House Price**

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 SimMultiple Linear Regression com a técnica de
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