

Centro Universitário Carlos Drummond de Andrade

Barbara de Souza Forte (201000960)
Danilo Cassiano Dias (201000329)
Eduardo Marcos dos Santos (201000006)
Rodrigo Henrique de Souza (201000435)
Wesley Brito Souza (192000961)

**Predição de Salário
(Base Adult)**

Cidade - SP
2021

Centro Universitário Carlos Drummond de Andrade

Predição de Salário (Base Adult)

Relatório Técnico-Científico apresentado na disciplina de Projeto Integrador para o curso de Análise e Desenvolvimento de Sistemas da Centro Universitário Carlos Drummond de Andrade (UNIDRUMMOND).

São Paulo – SP
2021

de SOUZA FORTE, Barbara; CASSIANO DIAS, Danilo; MARCOS dos SANTOS, Eduardo; HENRIQUE de SOUZA, Rodrigo; BRITO SOUZA, Wesley. **Predição de Salário**. 00f. Relatório Técnico-Científico. Análise e Desenvolvimento de Sistemas – Centro Universitário Carlos Drummond de Andrade. Tutor: Eduardo Palhares. Ponte Rasa, 2021.

RESUMO

Neste trabalho foi desenvolvido um código na linguagem de programação Python, onde foi utilizado a ferramenta Collab. Utilizamos a base de dados base Adult, do site da Kaggle que apresenta informações cadastrais como estado civil, gênero, região, raça, entre outros dados. O código desenvolvido tem como objetivo final demonstrar o resultado da análise, trazendo a quantidade de pessoas com renda superior ou inferior a \$50.000,00.

PALAVRAS-CHAVE: Dados, Informações. Análise.

SUMÁRIO

<u>1. INTRODUÇÃO</u>	1
<u>2. DESENVOLVIMENTO</u>	2
<u>2.1 Problema e objetivos</u>	2
<u>2.2. Justificativa</u>	2
<u>2. 3. Fundamentação teórica</u>	2
<u>2.4. Aplicação das disciplinas estudadas no Projeto Integrador</u>	3
<u>2.5. Metodologia</u>	3
<u>3. RESULTADOS</u>	4
<u>3.1. Solução inicial</u>	4
<u>3.2. Solução Final</u>	4
<u>4. CONSIDERAÇÕES FINAIS</u>	5
<u>REFERÊNCIAS</u>	6
<u>ANEXOS</u>	7
<u>APÊNDICES</u>	8
<u>Apêndice A –</u>	8

1. INTRODUÇÃO

Foram coletadas informações cadastrais sobre um grupo de pessoas dentro do Base Adult, utilizando algumas possíveis variáveis.

Iniciando pela variável de gênero, tornando possível analisar as diferenças salariais existente entre pessoas do sexo oposto.

A partir da análise do gênero foi possível considerar outras variáveis como cor, raça, região, entre outros dados.

Foi identificado que a variável raça pode influenciar nos resultados das diferenças salarias.

Com base nas variáveis citadas, foi identificado um grande problema, que é a diferença de renda entre determinados grupos, demonstrando a desigualdade social.

2. DESENVOLVIMENTO

2.1 Objetivos

Demonstrar o resultado da análise, trazendo a quantidade de pessoas com renda superior ou inferior a \$50.000,00.

- Levantamento dos dados a serem utilizados como base

2.2. Justificativa e delimitação do problema

O problema analisado é de suma importância para a sociedade em geral, pois a desigualdade social acarreta em diversas complicações educacionais, culturais que contribuem para o não desenvolvimento do indivíduo.

2.3. Fundamentação teórica

Aulas e vídeos apresentados pelo professor Eduardo Palhares, que administra o curso de Estrutura de Dados.

Site de base de dados Kaggle.

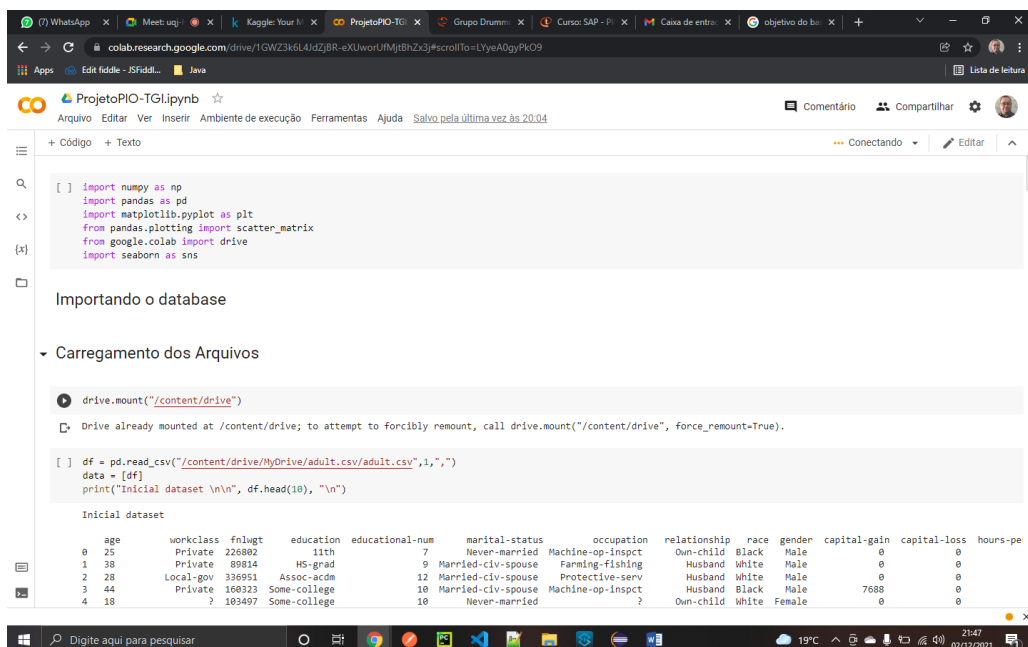
2.4. Aplicação das disciplinas estudadas no Projeto Integrador

Para realização do projeto integrador foi utilizado os conteúdos disponibilizados pelo professor orientador do projeto, no YouTube, na SAP, como também nas aulas ministradas pelo professor.

2.5. Metodologia

Para realização desse trabalho nos baseamos nos vídeos disponibilizados sobre a base Adult, utilizando a linguagem Python dentro do Google Collab para desenvolvimento do projeto integrador

3. RESULTADOS



```
[ ] import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from pandas.plotting import scatter_matrix
from google.colab import drive
import seaborn as sns
```

Importando o database

Carregamento dos Arquivos

```
drive.mount("/content/drive")
Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).
```

```
[ ] df = pd.read_csv("/content/drive/MyDrive/Adult.csv/adult.csv",1,")
data = [df]
print("Inicial dataset \n\n", df.head(10), "\n")
```

Inicial dataset

	age	workclass	fnlwt	education	educational-num	marital-status	occupation	relationship	race	gender	capital-gain	capital-loss	hours-pe
0	25	Private	226802	11th	7	Never-married	Machine-op-inspct	Own-child	Black	Male	0	0	
1	38	Private	89814	HS-grad	9	Married-civ-spouse	Farming-fishing	Husband	White	Male	0	0	
2	28	Local-gov	336951	Assoc-acdm	12	Married-civ-spouse	Protective-serv	Husband	White	Male	0	0	
3	44	Private	160323	Some-college	10	Married-civ-spouse	Machine-op-inspct	Husband	Black	Male	7688	0	
4	18	?	103497	Some-college	10	Never-married	?	Own-child	White	Female	0	0	

colab.research.google.com/drive/1GWZ3k6L4jDZjBR-eXUworUfMjtBhZx3j#scrollTo=LtYeA0gyPkO9

ProjetoPIO-TGI.ipynb

Arquivo Editar Ver Inserir Ambiente de execução Ferramentas Ajuda Salvo pela última vez às 20:04

+ Código + Texto

SEARCH STACK OVERFLOW

Análise das Variáveis

Função que plota a matriz de correlação

```
[ ] def plot_correlation(df, size=10):
    corr = df.corr()
    fig, ax = plt.subplots(figsize=(size,size))
    ax.matshow(corr)
    plt.xticks(range(len(corr.columns)),corr.columns)
    plt.yticks(range(len(corr.columns)),corr.columns)
    plt.show()
```

Correlação entre as variáveis numéricas

plot_correlation(df)

	age	fnlwgt	educational-num	capital-gain	capital-loss	hours-per-week
age	1.0	0.0	0.0	0.0	0.0	0.0
fnlwgt	0.0	1.0	0.0	0.0	0.0	0.0
educational-num	0.0	0.0	1.0	0.0	0.0	0.0
capital-gain	0.0	0.0	0.0	1.0	0.0	0.0
capital-loss	0.0	0.0	0.0	0.0	1.0	0.0
hours-per-week	0.0	0.0	0.0	0.0	0.0	1.0

19°C 21:47 02/12/2021

colab.research.google.com/drive/1GWZ3k6L4jDZjBR-eXUworUfMjtBhZx3j#scrollTo=LtYeA0gyPkO9

ProjetoPIO-TGI.ipynb

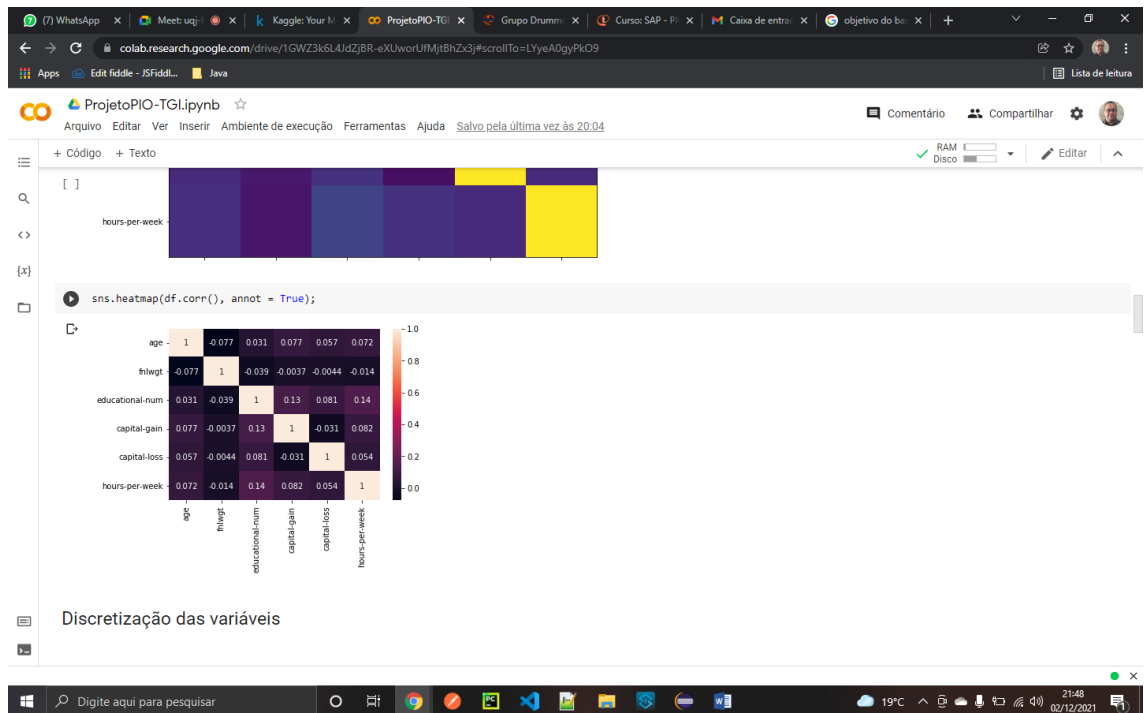
Arquivo Editar Ver Inserir Ambiente de execução Ferramentas Ajuda Salvo pela última vez às 20:04

+ Código + Texto

plot_correlation(df)

	age	fnlwgt	educational-num	capital-gain	capital-loss	hours-per-week
age	1.0	0.0	0.0	0.0	0.0	0.0
fnlwgt	0.0	1.0	0.0	0.0	0.0	0.0
educational-num	0.0	0.0	1.0	0.0	0.0	0.0
capital-gain	0.0	0.0	0.0	1.0	0.0	0.0
capital-loss	0.0	0.0	0.0	0.0	1.0	0.0
hours-per-week	0.0	0.0	0.0	0.0	0.0	1.0

19°C 21:47 02/12/2021



4. CONSIDERAÇÕES FINAIS

O código foi desenvolvido com base no vídeo passado, porém foram encontrados alguns bugs no código e a necessária avaliação.

REFERÊNCIAS

- ABNT – Associação Brasileira de Normas Técnicas. NBR 14724: Informação e documentação.
- Trabalhos Acadêmicos - Apresentação. Rio de Janeiro: ABNT, 2002.
- PALHARES, Eduardo. Predição de Salário - Base Adult. Youtube, 2021