



Departamento de  
Informática

# Trabalho Prático - Relatório

**Adriana Abreu**

**Leonor Rebola**

Faculdade de Engenharias  
Departamento de Informática  
Licenciatura em Inteligência Artificial e Ciência de Dados

Covilhã, Junho 2025





Departamento de  
Informática

## Trabalho Prático - Relatório

**Adriana Abreu**

*Nº Estudante 53672*

**Leonor Rebola**

*Nº Estudante 53663*

**Professor:** João Carlos Raposo Neves  
*Professor de Elementos de Inteligência Artificial e  
Ciência de Dados, Universidade da Beira Interior*

Faculdade de Engenharias  
Departamento de Informática  
Licenciatura em Inteligência Artificial e Ciência de Dados

Covilhã, Junho 2025



# CONTEÚDO

|  |           |
|--|-----------|
| <b>Conteúdo</b>                                  | <b>4</b>  |
| <b>Lista de Figuras</b>                          | <b>5</b>  |
| <b>1 Introdução</b>                              | <b>6</b>  |
| <b>2 Recolha de Dados</b>                        | <b>7</b>  |
| <b>3 Integração de Dados</b>                     | <b>8</b>  |
| <b>4 Análise Exploratória de Dados</b>           | <b>10</b> |
| 4.1 Estatísticas Descritivas . . . . .           | 10        |
| 4.2 Identificação de Outliers . . . . .          | 10        |
| <b>5 Limpeza e Preprocessamento de Dados</b>     | <b>14</b> |
| 5.1 Identificação dos valores em falta . . . . . | 14        |
| 5.2 Tratamento de Outliers . . . . .             | 15        |
| 5.3 Normalização dos Dados . . . . .             | 15        |
| <b>6 Análise Descritiva</b>                      | <b>16</b> |
| 6.1 Correlações . . . . .                        | 16        |
| 6.2 Análise Preditiva . . . . .                  | 21        |
| 6.3 Conclusões a tirar . . . . .                 | 25        |
| <b>7 Conclusão</b>                               | <b>26</b> |
| <b>8 Webgrafia</b>                               | <b>27</b> |

## LISTA DE FIGURAS

|      |   |    |
|------|---|----|
| 3.1  | Organização dos dados em 6 linhas distintas: combinações entre os anos escolares e tipos de curso . . . . . | 9  |
| 4.1  | Identificação de Outliers na variável Valor Alunos utilizando o método IQR . . . . .                        | 12 |
| 4.2  | Identificação de Outliers na variável Valor RSI utilizando o método IQR                                     | 12 |
| 4.3  | Identificação de Outliers na variável Valor GMM utilizando o método IQR . . . . .                           | 13 |
| 6.1  | csv dadostratados.csv com a Taxa de Retenção e Desistência separada por cursos e anos . . . . .             | 16 |
| 6.2  | Matriz de Correlação . . . . .  | 17 |
| 6.3  | Matriz de Correlação com Rendimento de Inserção Social entre 10 a 404                                       | 18 |
| 6.4  | Matriz de Correlação com Rendimento Social de Inserção Social entre 405 a 799 . . . . .                     | 19 |
| 6.5  | Matriz de Correlação com Ganho Médio Mensal entre 683.5 a 1014.6 . .  | 20 |
| 6.6  | Matriz de Correlação com Ganho Médio Mensal entre 1015.1 a 1346.7 .   | 21 |
| 6.7  | Gráfico de Barras das Médias da Taxa de Desistência Escolar . . . . .                                       | 22 |
| 6.8  | Gráfico de Barras das Médias do Rendimento Social de Inserção . . . .                                       | 23 |
| 6.10 | Gráfico de Barras das Médias dos Ganhos Médios Mensais . . . . .  | 24 |

## INTRODUÇÃO

O presente trabalho tem como tema a **Relação entre Desistência Escolar, Ganho Médio Mensal e Rendimento Social de Inserção**, mais concretamente em como os alunos do secundário são influenciados a desistir do ensino pela zona onde vivem ou pelos benefícios que ganham. Este grande tópico é de extrema relevância nos dias hodiernos, pois cada vez mais a educação em Portugal e no Mundo é valorizada.

O objetivo deste trabalho é que os educandos demonstrem os seus conhecimentos aprendidos ao longo do semestre, através da aplicação prática dos mesmos.

Para esse efeito este relatório será dividido em cinco tópicos com a sua devida importância, sendo o primeiro relativamente à recolha de dados, onde indicamos quais dados foram recolhidos. No segundo capítulo falaremos superficialmente acerca da integração de dados, onde explicaremos como reunimos todos os dados utilizados. Em seguida, iremos explicitar que métodos utilizamos para a análise exploratória de dados. A posteriori, passaremos à limpeza e processamento de dados, onde abordaremos as técnicas utilizadas para o efeito. Por fim, tiramos as conclusões finais deste trabalho e analisamos as futuras perspetivas.

## RECOLHA DE DADOS

Na fase preliminar deste trabalho, tivemos diversas discussões acerca do tema que gostaríamos de abordar, onde decidimos focarmo-nos na relação entre desistência escolar, ganho médio mensal e Rendimento Social de Inserção.

Após meticulosa reflexão, formulámos a ideia que queríamos analisar, sendo ela a hipótese de que, nas grandes cidades portuguesas, onde se registam ganhos médios mensais mais elevados, ocorre simultaneamente uma maior taxa de desistência escolar e uma maior atribuição de Rendimento Social de Inserção (RSI).

Para tal, foram recolhidos do PORDATA os seguintes ficheiros:

- Taxa de retenção e desistência no ensino secundário por modalidade de ensino e ano de escolaridade
- Ganho médio mensal
- Beneficiários do Rendimento Social de Inserção (RSI) por grupo etário



## INTEGRAÇÃO DE DADOS

Após a recolha de dados, passámos para a integração dos mesmos, onde agrupamos tudo num só ficheiro Excel chamado *dados\_inputados.csv*, como sugerido.

As variáveis das colunas são as seguintes:

- **Ano:** Esta coluna é referente aos anos em análise, mais concretamente ao período de 2012 a 2022.
- **Regiao:** Esta coluna é referente a todos os municípios de Portugal.
- **Valor Alunos:** Esta coluna é referente aos dados extraídos do Excel sobre a Taxa de retenção e desistência no ensino secundário por modalidade de ensino e ano de escolaridade. Estes valores encontram-se em percentagem.
- **Valor RSI:** Esta coluna é referente aos dados extraídos do Excel sobre os Beneficiários do Rendimento Social de Inserção por grupo etário. Para a nossa análise, escolhemos a faixa etária dos Menos de 25 anos.
- **Valor GMM:** Esta coluna é referente aos dados extraídos do Excel sobre o Ganho médio mensal. Estes valores encontram-se em Euros.

Como a nossa análise passava por comparar os Cursos Gerais e os Cursos Tecnológicos Profissionais relativamente a cada ano escolar separadamente, dividimo-los através da seguinte fórmula proveniente do Excel:

$$= SE(F_{602} = , \text{PROCV}(B_{602}, B : F, 5, \text{FALSO}), F_{602})$$

Ao utilizarmos a fórmula, ficamos com seis linhas correspondentes a cada ano e tipo de curso facilitando, assim, a nossa análise. Na seguinte imagem podemos constatar esse facto.

|   | A    | B        | C                                   | D           | E            | F         | G         |
|---|------|----------|-------------------------------------|-------------|--------------|-----------|-----------|
| 1 | Ano  | Regiao   | Curso                               | Ano Escolar | Valor Alunos | Valor RSI | Valor GMM |
| 2 | 2012 | Abrantes | Cursos gerais                       | 10º ano     | 19.3         | 556.0     | 975.8     |
| 3 | 2012 | Abrantes | Cursos gerais                       | 11º ano     | 16.5         | 556.0     | 975.8     |
| 4 | 2012 | Abrantes | Cursos gerais                       | 12º ano     | 38.5         | 556.0     | 975.8     |
| 5 | 2012 | Abrantes | Cursos Tecnológicos e Profissionais | 10º ano     | 12.9         | 556.0     | 975.8     |
| 6 | 2012 | Abrantes | Cursos Tecnológicos e Profissionais | 11º ano     | 7.8          | 556.0     | 975.8     |
| 7 | 2012 | Abrantes | Cursos Tecnológicos e Profissionais | 12º ano     | 32.9         | 556.0     | 975.8     |

Figura 3.1: Organização dos dados em 6 linhas distintas: combinações entre os anos escolares e tipos de curso

## ANÁLISE EXPLORATÓRIA DE DADOS

Nesta fase, tentámos compreender a estrutura, distribuição e relações preliminares entre as variáveis do estudo, onde cumprimos os seguintes passos:

### 4.1 Estatísticas Descritivas

Ao longo das nossas pesquisas, deparámo-nos com a existência do método `describe()` do Pandas, que nos foi bastante útil para saber quais os valores para as estatísticas descritivas. Apresentamos a seguir o código:

```
with pd.option_context('display.max_columns', None):
    print(df.describe())
```

Com este código, obtivemos os seguintes dados para cada coluna:

Tabela 4.1: Estatísticas descritivas de todas as variáveis em estudo

| Variável            | count | mean       | std        | min   | 25% | 50% | 75%    | max    |
|---------------------|-------|------------|------------|-------|-----|-----|--------|--------|
| Taxa de Desistência | 17379 | 13.629789  | 12.862973  | 0     | 4.1 | 9.9 | 19.7   | 100    |
| RSI                 | 20160 | 422.890873 | 964.508969 | 0     | 56  | 132 | 359    | 12066  |
| GMM                 | 19986 | 965.986040 | 185.919367 | 683.5 | 843 | 933 | 1047.1 | 2331.2 |

Com a obtenção destes valores, conseguimos perceber melhor como as variáveis se comportavam e também conseguimos, a partir da mediana, tratar dos valores em falta.

### 4.2 Identificação de Outliers

Na fase inicial de detecção de Outliers, aplicámos o método do Z-Score porém, esta forma de detecção mostrou-se incompatível para as nossas variáveis. Contudo, e apesar desta limitação, a classificação booleana (valores True/False para Outliers) provou ser extremamente útil.

Devido a isso, mudamos de abordagem, passando para o método IQR, que nos deu resultados mais robustos para a nossa análise.

```
def iqr_alunos():
    sns.boxplot(y=df["Valor Alunos"])
    plt.show()

    coluna = "Valor Alunos"

    # 3. Calcula Q1, Q3 e IQR
    Q1 = df[coluna].quantile(0.25)
    Q3 = df[coluna].quantile(0.75)
    IQR = Q3 - Q1

    # 4. Define limites
    LI = Q1 - 1.5 * IQR
    LS = Q3 + 1.5 * IQR

    # 5. Identifica outliers
    outliers_sup = df[df[coluna] > LS][coluna]
    outliers_inf = df[df[coluna] < LI][coluna]

    # 6. Resultados
    min_sup = outliers_sup.min() if not outliers_sup.empty else None
    max_inf = outliers_inf.max() if not outliers_inf.empty else None

    print(f"Limite inferior (LI): {LI:.2f}")
    print(f"Limite superior (LS): {LS:.2f}")
    print(f"Maior outlier inferior: {max_inf}")
    print(f"Menor outlier superior: {min_sup}")
    '''Name: Valor Alunos, dtype: float64
    Maior outlier inferior: None
    Menor outlier superior: 38.7
    '''

    max_inf = max_inf if max_inf is not None else 0
    return min_sup, max_inf
```

Neste código, podemos ver a implementação do IQR para a variável Valor Alunos, onde depois obtivemos o seguinte diagrama de caixa para a visualização dos Outliers:

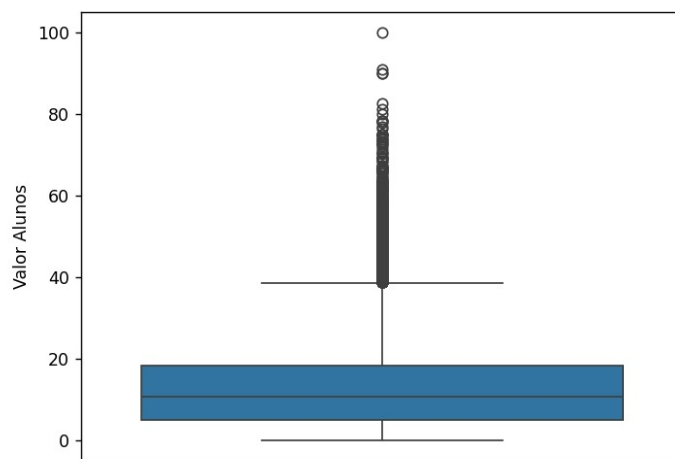


Figura 4.1: Identificação de Outliers na variável Valor Alunos utilizando o método IQR

Para as outras variáveis, foram também gerados diagramas de caixas, sendo eles os seguintes:

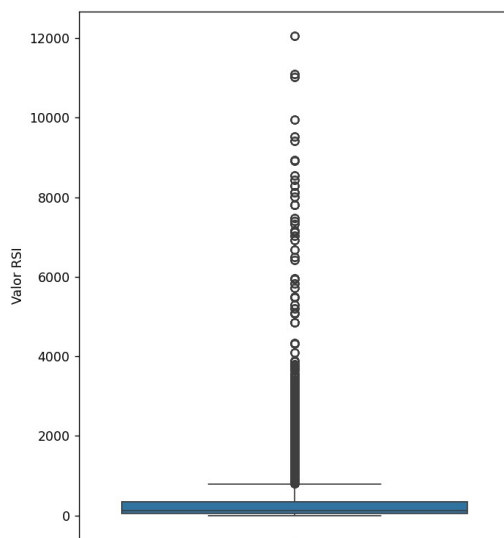


Figura 4.2: Identificação de Outliers na variável Valor RSI utilizando o método IQR

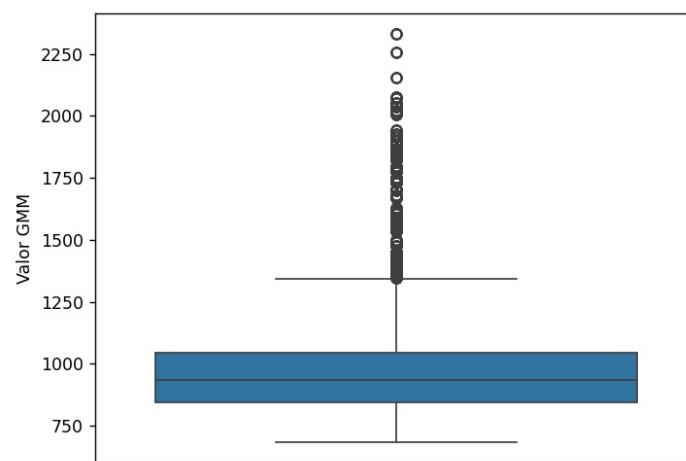


Figura 4.3: Identificação de Outliers na variável Valor GMM utilizando o método IQR

## LIMPEZA E PREPROCESSAMENTO DE DADOS

Nesta fase, preparámos os dados para análise através de um rigoroso processo de tratamento, garantindo qualidade e consistência. Os principais passos incluíram:

### 5.1 Identificação dos valores em falta

Para a deteção dos valores em falta utilizamos o código apresentado a seguir:

```
missing_values = df.isnull().sum()
print(missing_values)
```

e obtivemos os seguintes valores:

Tabela 5.1: Valores em Falta das Variáveis em Estudo

| Variável            | count |
|---------------------|-------|
| Taxa de desistência | 2949  |
| RSI                 | 168   |
| GMM                 | 342   |

De seguida, através da média, tratámos dos valores em falta para cada variável, guardando os valores num *csv* novo. No código a seguir, podemos verificar isso mesmo para a variável “Valor Alunos”, sendo o mesmo replicado para as restantes:

```
def faltar_valores_alunos(valor_alunos):
    '''valores em falta da desistência escolar em percentagem'''
    media_valores = dff['Valor Alunos'].mean()
    dff['Valor Alunos'] = dff['Valor Alunos'].fillna(media_valores)

    # 4. Salvar o DataFrame com os vlores imputados num novo CSV
    dff.to_csv('dados_imputados_media.csv', index=False)

    return valores_em_falta(valor_alunos)
```

## 5.2 Tratamento de Outliers

Após aplicar o método IQR para a detecção de Outliers, passamos para o tratamento dos mesmos, através da Winsorização, o que nos permitiu limitar o impacto de valores extremos sem perder informação relevante e garantir robustez na análise dos dados. O exemplo de código para este passo encontra-se apresentado a seguir:

```
def winsorizersi(max_infr, min_supr):  
    def winsorize_by_value(data, lower_bound, upper_bound):  
        data = np.array(data)  
        data[data < lower_bound] = lower_bound  
        data[data > upper_bound] = upper_bound  
        return data  
  
    # Aplica a winsorização à coluna desejada (substitui a original)  
    dfp['Valor RSI'] = winsorize_by_value(dfp['Valor RSI'],  
        lower_bound=10, upper_bound=1000)  
    winsorize_by_value(dfp["Valor RSI"], max_infr, min_supr)  
    # Salva no mesmo arquivo (sobrescreve)  
    dfp.to_csv('dados_com_zscore21.csv', index=False)
```

## 5.3 Normalização dos Dados

Por fim, normalizamos os dados através da técnica Min-Max, que nos permitiu transformar todas as variáveis numéricas para uma escala comum entre [0, 1], permitindo uma comparação direta entre variáveis com unidades distintas.

```
def normalisealunos():  
    scaler = MinMaxScaler()  
  
    # Aplicar a normalização  
    df['valor_alunos_minmax'] = scaler.fit_transform(df[['Valor  
        Alunos']])  
    df["valor_alunos_minmax"] = df["valor_alunos_minmax"].round(3)  
  
    # 3. Salvar o novo CSV com a nova coluna  
    df.to_csv("dados_com_zscore1.csv", sep=";", index=False)  
  
normalisealunos()
```



## ANÁLISE DESCRITIVA

Após termos os dados todos tratados elaboramos novos *csv*, desta vez separando tanto a taxa de desistência e retenção por cursos e por anos escolares como os valores do RSI e do GMM por superiores e inferiores. Ou seja, no RSI valores de 10 a 404 para inferiores entre 405 a 799 e para o GMM para inferiores entre 683.5 e 1014.6 e superiores de 1015.1 a 1346.7.

| Ano  | Região     | Alunos10C | Alunos10C | Alunos11C | Alunos11C | Alunos12C | Alunos12C | valor_RSI | valor_GMM |
|------|------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| 2012 | Abrantes   | 0.324     | 0.101     | 0.226     | 0.0       | 0.993     | 0.798     | 0.692     | 0.441     |
| 2012 | Águeda     | 0.359     | 0.014     | 0.066     | 0.0       | 1.0       | 0.15      | 0.622     | 0.365     |
| 2012 | Aguar da F | 0.038     | 0.0       | 0.265     | 0.0       | 0.568     | 0.882     | 0.048     | 0.023     |
| 2012 | Alandroal  | 0.334     | 0.334     | 0.334     | 0.0       | 0.334     | 0.334     | 0.106     | 0.162     |
| 2012 | Albergaria | 0.334     | 0.028     | 0.237     | 0.0       | 1.0       | 0.777     | 0.421     | 0.399     |
| 2012 | Albufeira  | 0.564     | 0.0       | 0.425     | 0.0       | 0.909     | 0.767     | 0.431     | 0.343     |
| 2012 | Alcácer do | 0.383     | 0.345     | 0.045     | 0.0       | 0.631     | 0.0       | 0.332     | 0.295     |
| 2012 | Alcanena   | 0.066     | 0.373     | 0.139     | 0.0       | 0.585     | 0.812     | 0.157     | 0.451     |
| 2012 | Alcobaça   | 0.0       | 0.0       | 0.077     | 0.0       | 0.972     | 0.659     | 0.521     | 0.293     |

Figura 6.1: csv dadostratados.csv com a Taxa de Retenção e Desistência separada por cursos e anos

### 6.1 Correlações

Para começar a nossa análise, procedemos a realização de várias matrizes de correlação, para podermos avaliar todas as relações possíveis. Para a criação dos gráficos, usámos o seguinte código:

```
def principal():
    dados = pd.DataFrame({
        "10º ano CG": alunos1,
        "10º ano CP": alunos2,
        "11º ano CG": alunos3,
        "11º ano CP": alunos4,
        "12º ano CG": alunos5,
        "12º ano CP": alunos6,
        "Rendimento Social de Inserção()": valor_rsi,
        "Ganho médio mensal": valor_gmm
    })
```

```
corr_df = dados.corr(numeric_only=True)

plt.figure(figsize=(10, 8))
sns.heatmap(corr_df, annot=True, fmt=".2f", cmap="mako", square=True)
plt.title("Matriz de Correlação")
plt.tight_layout()
plt.show()
```

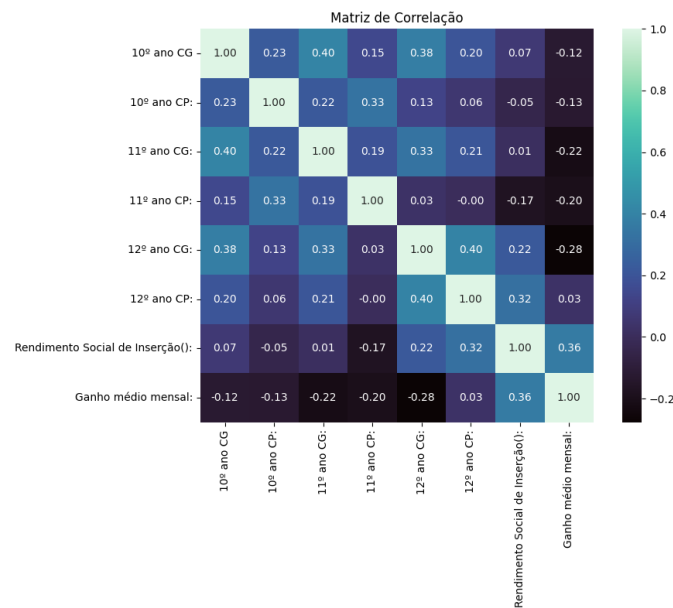


Figura 6.2: Matriz de Correlação

A matriz de correlação revela algumas relações relevantes entre as variáveis analisadas. Observa-se uma correlação moderada entre os cursos gerais (CG) dos diferentes anos letivos, especialmente entre o 10º e o 11º ano (0.40) e entre o 11º e o 12º ano (0.33), sugerindo consistência no desempenho dos alunos ao longo dos anos.

O Rendimento Social de Inserção (RSI) apresenta uma correlação positiva com o Ganho médio mensal (0.36), o que pode indicar que regiões com maior acesso ao RSI também tendem a ter ganhos médios mais elevados, possivelmente por efeitos de compensação social.

Por outro lado, o Ganho médio mensal tem correlação negativa com os desempenhos escolares, em especial com o 12.º ano CG (-0.28) e o 11.º ano CG (-0.22), o que pode apontar para um possível desequilíbrio entre condições socioeconómicas e resultados escolares em determinadas zonas.

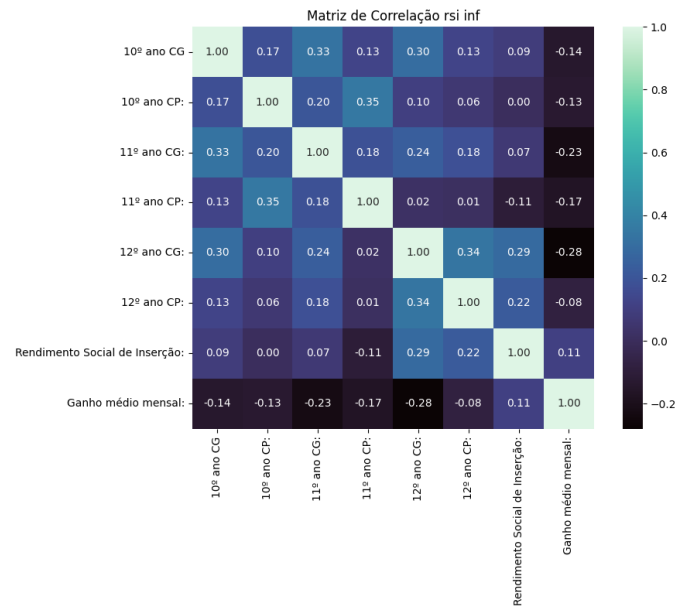


Figura 6.3: Matriz de Correlação com Rendimento de Inserção Social entre 10 a 404

Neste mapa, observa-se uma correlação negativa moderada entre o ganho médio mensal e o percentual de alunos que concluem o 12º ano no ensino geral (CG), com um valor de -0.28. Este padrão sugere que, em concelhos com RSI inferior (presumivelmente com menor apoio social), a presença de jovens a concluir o ensino secundário geral está associada a rendimentos mais baixos. Isso pode refletir dinâmicas locais onde a continuação dos estudos não se traduz automaticamente em melhores salários, talvez por falta de oportunidades profissionais qualificadas ou pela migração dos mais escolarizados para outras regiões.

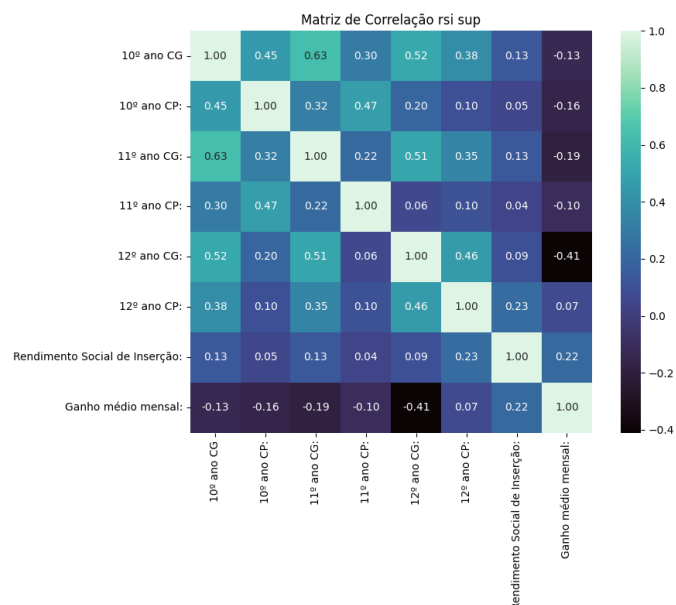


Figura 6.4: Matriz de Correlação com Rendimento Social de Inserção Social entre 405 a 799

No grupo com RSI superior, a correlação entre o ganho médio mensal e o 12º ano CG intensifica-se para -0.41, reforçando a ideia de que há uma desconexão entre escolaridade formal e rendimentos locais. Este padrão pode ser interpretado como um efeito de “fuga de cérebros”, em que os jovens mais qualificados abandonam estas zonas para procurar melhores oportunidades noutras territórios, deixando para trás um perfil populacional com menor rendimento. Curiosamente, o 12º ano profissional (CP) mantém correlações menos negativas ou neutras, o que indica que essa via pode estar mais bem adaptada às necessidades e estruturas de emprego locais.

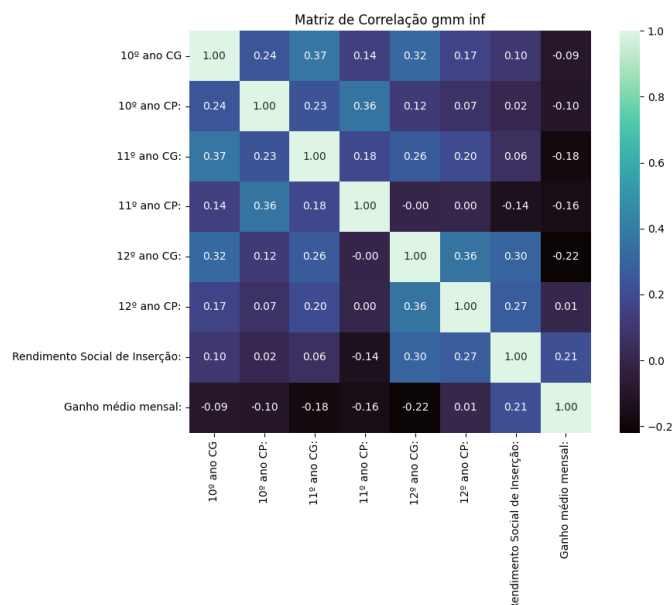


Figura 6.5: Matriz de Correlação com Ganho Médio Mensal entre 683.5 a 1014.6

No grupo com ganho médio mensal inferior (gmm inf), as correlações entre as taxas de desistência e o ganho médio mensal são negativas mas bastante fracas, variando entre -0.09 e -0.27. Isto revela que, embora o rendimento ainda influencie negativamente a desistência, o seu efeito é menos pronunciado neste grupo, possivelmente devido à prevalência generalizada de dificuldades económicas que tornam o rendimento um fator menos diferenciador. Por outro lado, as correlações com o RSI são positivas, ainda que mais suaves do que no grupo superior, o que indica que a vulnerabilidade social afeta negativamente a continuidade escolar, mas de forma menos clara. Este cenário pode refletir uma realidade em que a exclusão social é mais difusa e estrutural, reduzindo o contraste entre diferentes subgrupos socioeconómicos.

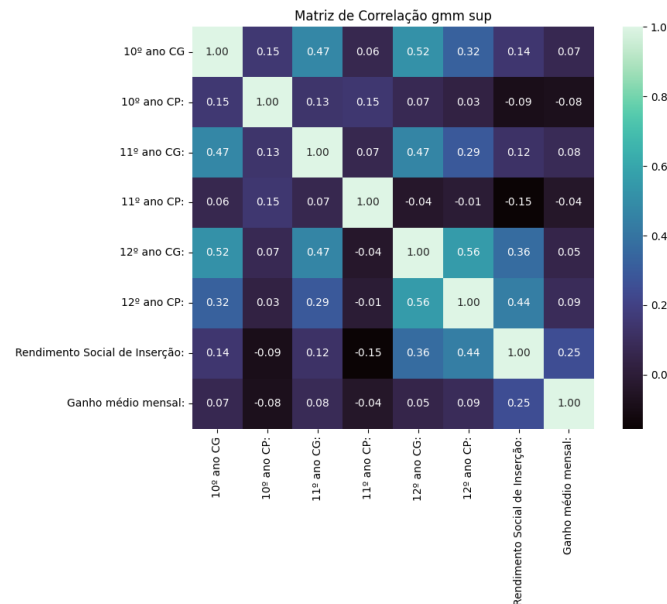


Figura 6.6: Matriz de Correlação com Ganho Médio Mensal entre 1015.1 a 1346.7

No grupo com ganho médio mensal superior (gmm sup), observa-se uma correlação negativa moderada entre as taxas de desistência escolar e o ganho médio mensal, particularmente no 12º ano do ensino profissional, onde a correlação atinge -0.60. Isto sugere que, em contextos socioeconómicos mais favorecidos, os alunos têm maior probabilidade de concluir os estudos, provavelmente devido a maior estabilidade financeira, acesso a recursos educativos e menor necessidade de ingressar precocemente no mercado de trabalho. Em contraste, a correlação entre desistência e o Rendimento Social de Inserção (RSI) é positiva e moderadamente forte, atingindo valores de 0.56 no 12º ano CP. Este padrão indica que, mesmo em regiões com rendimento médio mais alto, a presença de bolsas de pobreza social continua a ter impacto negativo na permanência escolar, funcionando como um indicador de risco.

## 6.2 Análise Preditiva

Finalmente, passaremos a analisar as várias variáveis ao longo dos anos para prever o que irá acontecer futuramente. Para isso, criámos vários gráficos de barras e utilizámos o seguinte código:

```
def grafico_media_alunos():
    df = pd.read_csv("dados_com_zscore1.csv", sep=";", encoding="latin1")
    df_grouped = df.groupby('Ano')['Valor Alunos'].mean().reset_index()

    # Plotar gráfico de barras
    plt.figure(figsize=(10, 6))
```

```
plt.bar(df_grouped['Ano'], df_grouped['Valor Alunos'], color='#
CD7BEB', edgecolor='black')
plt.title('Média da Taxa Retenção e Desistência no Ensino
Secundário Por Ano')
plt.xlabel('Ano')
plt.ylabel('Taxa Retenção e Desistência no Ensino Secundário')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```

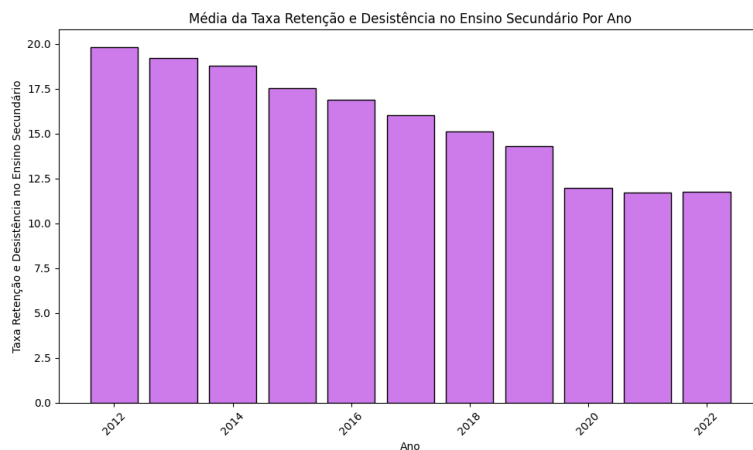


Figura 6.7: Gráfico de Barras das Médias da Taxa de Desistência Escolar

O seguinte gráfico apresenta a evolução da média da taxa de retenção e desistência no ensino secundário. Observa-se uma tendência decrescente, com a taxa a situar-se perto dos 20% em 2012 e a descer para cerca de 12,5% em 2022. Esta redução progressiva sugere a implementação de políticas educativas mais eficazes, bem como o possível reforço de medidas de apoio ao sucesso escolar. A diminuição das taxas de abandono escolar poderá também refletir uma maior consciencialização da importância da educação, além de potenciais melhorias na qualidade do ensino e nas condições socioeconómicas dos alunos.

Este gráfico ilustra a média anual de beneficiários do Rendimento Social de Inserção (RSI). Verifica-se uma tendência de decréscimo ao longo do período analisado, com os valores a descerem de mais de 300 mil beneficiários em 2012 para valores inferiores a 250 mil em 2022. Esta redução pode estar associada a uma melhoria gradual das condições económicas e sociais no país, resultando num menor número de pessoas em situação de vulnerabilidade extrema.

Os gráficos (a) e (b) representam a média da taxa de inserção social (RSI) em dois intervalos distintos: entre 10 a 404 beneficiários e entre 405 a 799 beneficiários.

No gráfico (a), observa-se uma maior variabilidade ao longo dos anos, com oscilações mais acentuadas, especialmente entre 2016 e 2021.

Já no gráfico (b), os valores são mais estáveis ao longo do tempo, apresentando variações ligeiras entre os anos. Isto deve-se ao facto de se tratar de regiões com maior

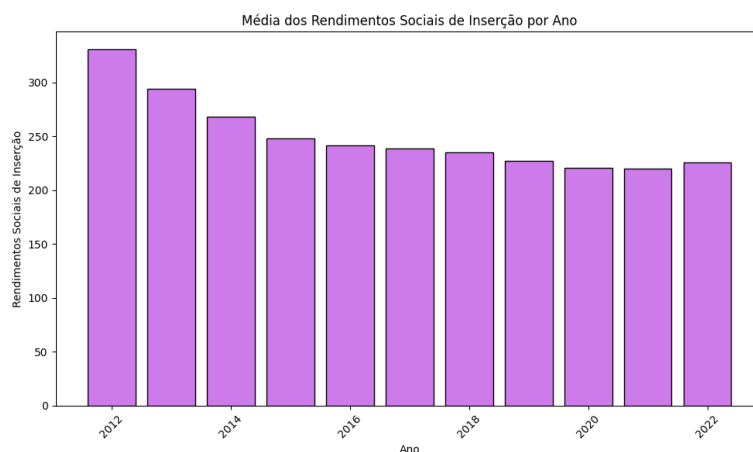
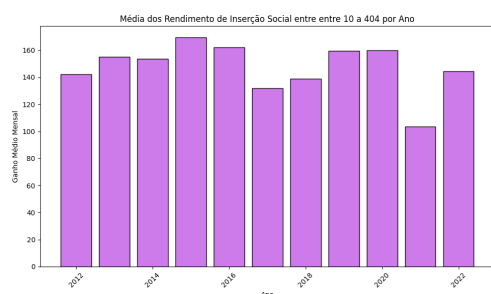
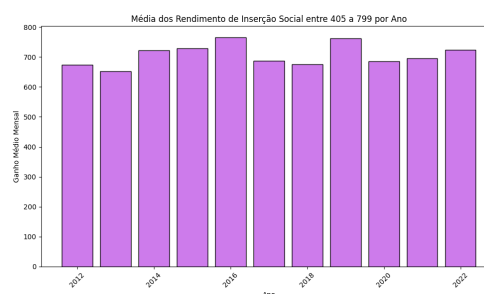


Figura 6.8: Gráfico de Barras das Médias do Rendimento Social de Inserção



(a) RSI entre 10 a 404



(b) RSI entre 405 a 799

número de beneficiários, o que tende a suavizar as flutuações.

Em conjunto, ambos os gráficos mostram que, apesar das diferenças na escala, existe uma certa estabilidade nos níveis de RSI, embora com algumas oscilações nas faixas mais baixas.



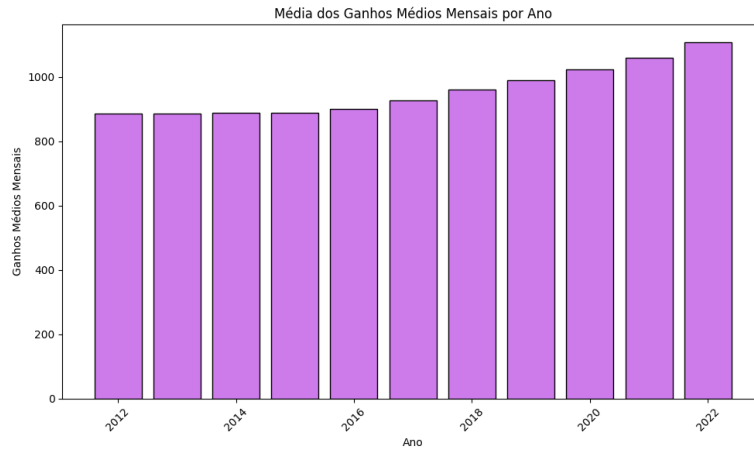
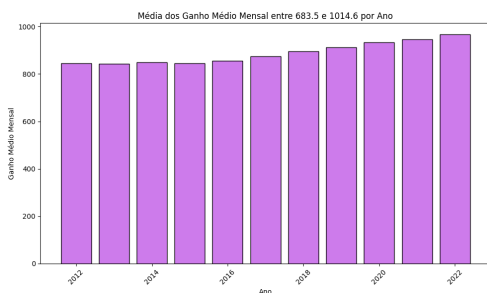
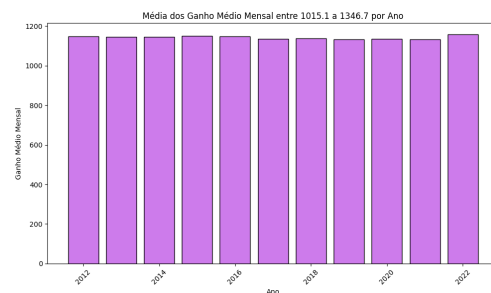


Figura 6.10: Gráfico de Barras das Médias dos Ganhos Médios Mensais

Este gráfico mostra a média anual do ganho médio mensal (GMM) em Portugal. O gráfico mostra uma tendência de crescimento constante ao longo do período em análise, com os valores a subirem de cerca de pouco mais de 800 euros em 2012 para mais de 1000 euros em 2022. Este aumento pode refletir diversos fatores, como a valorização gradual dos salários mínimos e o desenvolvimento económico geral do país. Além disso, este crescimento pode também estar relacionado com a melhoria da educação.



(a) GMM entre 683.5 a 1014.6



(b) GMM entre 1015.1 a 1346.7

Os gráficos (a) e (b) representam os ganhos médios mensais divididos por dois intervalos salariais: entre 683.5 a 1014.6 euros e entre 1015.1 a 1346.7 euros. No gráfico (a), referente aos salários mais baixos, observa-se uma subida moderada e constante ao longo dos anos, com os ganhos a aproximarem-se progressivamente do limite superior da faixa. Isto pode indicar uma melhoria nas condições dos trabalhadores com rendimentos mais baixos. No gráfico (b), correspondente a salários mais elevados, o crescimento é também visível, embora mais acentuado nos anos mais recentes (a partir de 2019). Tal evolução pode estar relacionada com valorização de funções mais qualificadas ou com o impacto de reajustes salariais em setores com maior capacidade económica.

Em conjunto, os dois gráficos mostram uma tendência geral de valorização salarial, abrangendo tanto os rendimentos mais baixos como os mais altos, embora com ritmos de crescimento ligeiramente distintos.

### 6.3 Conclusões a tirar

Após o tratamento e limpeza dos dados, conseguimos concluir que:

- A taxa de desistência escolar no ensino secundário diminuiu ao longo do período de 2012 a 2022.
- O Rendimento Social de Inserção (RSI) apresenta correlações fracas, mas positivas com os resultados escolares, sobretudo no 12º ano profissional, sugerindo um impacto mais relevante nas fases finais do percurso académico.
- O Ganho Médio Mensal (GMM) tem correlações negativas com os resultados escolares, especialmente no ensino geral.
- Existe consistência no desempenho escolar, com correlações fortes entre os resultados desses anos, o que aponta para estabilidade no percurso académico.
- No futuro, é possível que a taxa de desistência diminua, que o ganho médio mensal aumente e que aqueles que beneficiam do RSI tendem a ser cada vez menos.

## CONCLUSÃO

Neste trabalho, abordámos e pomos em prática diversas técnicas para o tratamento e análise de dados. Foi feito o estudo de diversas variáveis onde, após a verificação de Outliers e da análise exploratória inicial, aplicámos técnicas de limpeza e transformação de dados. Utilizámos estatísticas descritivas para compreender a distribuição das variáveis e normalização para garantir a comparabilidade entre métricas distintas. Adicionalmente, usamos métodos de visualização, como diagramas de caixa, para identificar padrões e anomalias. Por fim, realizámos testes de correlação e modelação estatística para explorar relações entre as variáveis, o que nos permitiu extrair insights valiosos.

Durante a realização deste relatório, tivémos numa posição onde foi-nos exposta diversa informação que proporcionou-nos uma grande curiosidade por aprender mais sobre o que foi lecionado em sala de aula, alguma sabedoria e conhecimento acerca do assunto. Após a finalização deste trabalho concluímos com uma certa realização a aquisição de temas que no futuro, não só no percurso académico mas também como no profissional, nos proporcionam competências importantes.

Para a realização deste trabalho, foram utilizados recursos digitais.

Repositório GitHub: <https://github.com/adriana1005/elementos>

## WEBGRAFIA

Taxa de retenção e desistência no ensino secundário por modalidade de ensino e ano de escolaridade. (2025). PORDATA. , 21 de Abril 2025

Ganho médio mensal. (2025). PORDATA., 21 de Abril 2025

Beneficiários do Rendimento Social de Inserção (RSI) por grupo etário. (2025). PORDATA., 21 de Abril 2025

Pandas DataFrame describe() Method. (n.d.). Wwww.w3schools.com. <https://www.w3schools.com/p>  
7 de maio 2025

How to create a seaborn correlation heatmap in Python? (2020, November 9). GeeksforGeeks., 28 de Maio 2025

Arunn Thevapalan. (2024, July 30). Matriz de correlação no Excel: Um guia completo para criar e interpretar. Datacamp.com; DataCamp. <https://www.datacamp.com/pt/tutorial/correlation-matrix-excel>, 29 de maio 2025