

ISPGAYA

instituto superior politécnico

Escola Superior de Ciência e Tecnologia

Licenciatura EI- 2023/2024

Processamento de Informação – 2º momento de avaliação



Aluno:

Miguel Magalhães (2021103166)

Gonçalo Silva (2020101055)

Gonçalo Pereira (2020101880)

Docente: Prof. José Monteiro

12 de Janeiro de 2024

ÍNDICE DE CONTEÚDOS

Índice de Conteúdos.....	2
Índice de Ilustrações	3
Introdução	4
Estudo das questões de investigação	5
Questões de Investigação (resultados alcançados, limitações e trabalho futuro)	5
1ª Questão:	5
2ª Questão:	7
3ª Questão:	9
4ª Questão:	11
Materiais e Métodos.....	15
Linguagem Utilizada:	15
HTML (HyperText Markup Language):	15
PHP (Hypertext Preprocessor):.....	15
phpMyAdmin:	15
JavaScript:.....	15
Manipulação de dados:	16
Formato de Saída:	16
Conclusão.....	18
Bibliografia	19
Anexos	20

ÍNDICE DE ILUSTRAÇÕES

Figura 1.: Resultado da pergunta 1 executada na query.	5
Figura 2.: Gráfico referente à pergunta 1.....	6
Figura 3.: Resultado da pergunta 2 executada na query.	7
Figura 4.: Gráfico referente à pergunta 2.....	8
Figura 5.: Resultado da pergunta 3 executada na query.	9
Figura 6.: Gráfico referente à pergunta 3.....	10
Figura 7.: Resultado da pergunta 4 executada na query.	12
Figura 8.: Gráfico referente à pergunta 4.....	13

INTRODUÇÃO

No contexto da Unidade Curricular de Processamento de Informação, sob a orientação do professor José Monteiro, este trabalho de aula final propõe-se abordar dois momentos distintos de avaliação, cada um centrado em objetivos específicos relacionados com a extração, processamento e análise de dados provenientes de um conjunto de ficheiros no formato Ms Excel.

No primeiro momento de avaliação, realizado anteriormente, a ênfase recaiu sobre o desenvolvimento de um modelo de extração de conteúdo de informação, compreendendo desde a análise da fonte de dados até a automação da obtenção do conjunto de dados completo num sistema de gestão de base de dados. O resultado desse esforço culminou na entrega de um código fonte e relatório técnico que documenta a abordagem, princípios teóricos e técnicas adotadas durante essa fase inicial do projeto.

O segundo momento de avaliação, o foco principal desta etapa do trabalho, estende a complexidade do projeto ao envolver a normalização da base de dados e a garantia da qualidade dos dados extraídos. Durante a aula do dia 9 de janeiro, o grupo realizou tarefas fundamentais, tais como a associação em grupos, a proposição de questões de investigação relevantes para o conjunto de dados e a realização de consultas estruturadas à base de dados para responder a essas questões.

Neste relatório, detalharemos minuciosamente as etapas específicas realizadas durante a aula do dia 9 de janeiro, destacando a preparação do relatório, a formulação e refinamento das questões de investigação, a descrição dos materiais e métodos utilizados nas consultas à base de dados, a apresentação visual dos resultados obtidos e, por fim, as conclusões alcançadas.

Cada secção subsequente fornecerá uma visão aprofundada das tarefas executadas durante esta fase do projeto, destacando as contribuições individuais dos membros do grupo e culminando na entrega dos elementos essenciais para avaliação, incluindo o conjunto de dados, código-fonte e o relatório técnico final.

ESTUDO DAS QUESTÕES DE INVESTIGAÇÃO

QUESTÕES DE INVESTIGAÇÃO (RESULTADOS ALCANÇADOS, LIMITAÇÕES E TRABALHO FUTURO)

1ª QUESTÃO:

O código fornecido parece ser uma página da web escrita em PHP e HTML, que utiliza a biblioteca Chart.js para exibir um gráfico de rosquinha (doughnut chart) com dados provenientes de uma consulta SQL. Vou abordar os diferentes tópicos mencionados.

Pergunta SQL:

```
"SELECT
(SUM(CASE WHEN u.codigo = '1444200' OR u.codigo = '1444400' OR u.codigo = '4441'
OR u.codigo = '4442' OR u.codigo = ' 4440' OR u.codigo = '1444100'
OR u.codigo = '4443' THEN 1 ELSE 0 END) / COUNT(d.id)) * 100 AS
percentagem_pertence,
(SUM(CASE WHEN u.codigo != '1444200' OR u.codigo != '1444400' OR u.codigo != '4441'
OR u.codigo != '4442' OR u.codigo != ' 4440' OR u.codigo != '1444100'
OR u.codigo != '4443' OR u.codigo IS NULL THEN 1 ELSE 0 END) / COUNT(d.id)) *
100 AS percentagem_nao_pertence
FROM
docente d
LEFT JOIN habilitacoes h ON d.id = h.id_docente
LEFT JOIN unidadeorganica u ON h.cod_unidadeorganica = u.codigo;"
```

Resultados Alcançados:

percentagem_pertence	percentagem_nao_pertence
88.0315	11.3123

Figura 1.: Resultado da pergunta 1 executada na query.

O código realiza uma consulta ao banco de dados para obter percentagens relacionadas a diferentes códigos de unidade orgânica de docentes. Essas percentagens são usadas para gerar dados para o gráfico de rosquinha. Os resultados são visualizados na página da web por meio da biblioteca Chart.js.

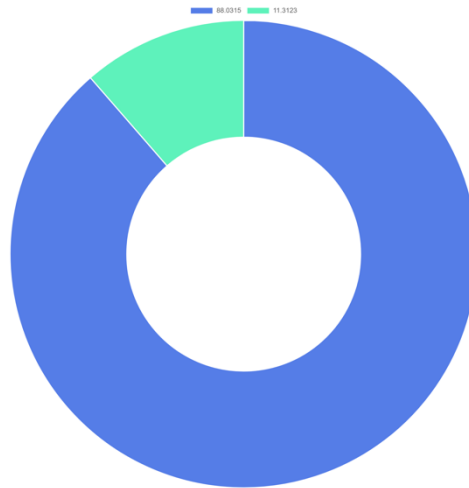


Figura 2.: Gráfico referente à pergunta 1.

Limitações:

- **Lógica da Consulta SQL:** Há um problema na lógica da consulta SQL. As comparações "=" e "!=" são incorretas. A forma correta seria "<=" e "<>" ou "!=". Corrigindo essas comparações, a lógica da consulta deve ser revista.
- **Estrutura de Dados do Gráfico:** A estrutura dos dados para o gráfico pode não ser ideal. Parece haver um problema na construção dos arrays \$labels, \$data, e \$backgroundColor dentro do loop foreach. Esses arrays podem estar a ser sobrescritos a cada iteração, e o gráfico pode não ser exibido conforme o esperado.

Trabalho Futuro:

- **Melhoria na Lógica da Consulta:** Refatorar a lógica da consulta SQL para garantir que as comparações estejam corretas e que os resultados reflitam as expectativas.
- **Revisão da Estrutura de Dados:** Revir a construção dos arrays \$labels, \$data, e \$backgroundColor para garantir que os dados sejam corretamente estruturados para o gráfico de rosquinha.
- **Aprimoramento do Layout:** Considerar melhorias no layout da página da web, como a adição de informações explicativas, legendas e outros elementos visuais para tornar a apresentação mais compreensível.
- **Tratamento de Erros:** Implementar tratamento de erros para lidar com possíveis falhas na execução da consulta SQL ou na geração do gráfico.
- **Segurança:** Avaliar a segurança do código, especialmente em relação à manipulação de consultas SQL para evitar vulnerabilidades de injeção de SQL.
- Essas sugestões visam melhorar a funcionalidade, a legibilidade e a segurança do código, proporcionando uma experiência mais robusta para os utilizadores.

2ª QUESTÃO:**Pergunta SQL:**

```
SELECT
  (SELECT COUNT(DISTINCT id_docente)
   FROM habilitacoes
   WHERE id_especialidade IS NOT NULL) / total_docentes.count * 100 AS
media_com_especialidade,
  (SELECT COUNT(DISTINCT id_docente)
   FROM habilitacoes
   WHERE id_especialidade IS NULL) / total_docentes.count * 100 AS
media_sem_especialidade
FROM
  (SELECT COUNT(DISTINCT id_docente) AS count FROM habilitacoes) AS
total_docentes;
```

Resultados Alcançados:

media_com_especialidade	media_sem_especialidade
17.4877	89.9015

Figura 3.: Resultado da pergunta 2 executada na query.

O código calcula a percentagem de docentes com e sem especialidade relativamente ao total de docentes. Esses resultados são então utilizados para preencher os dados do gráfico de rosquinha, exibido na página HTML. O gráfico fornece uma representação visual das proporções de docentes com e sem especialidade.



Figura 4.: Gráfico referente à pergunta 2.

Limitações:

- **Divisão de Rótulos e Dados:** No loop foreach, há uma duplicação de rótulos e dados, o que pode originar uma representação incorreta no gráfico. Isso ocorre porque os rótulos e dados para "com especialidade" e "sem especialidade" estão a ser adicionados duas vezes.
- **Estilo Padrão:** Da mesma forma que no código anterior, a geração aleatória de cores pode resultar numa falta de consistência visual no gráfico.

Trabalho Futuro:

- **Correção dos Dados:** Corrigir o loop foreach para garantir que os rótulos e dados sejam adicionados corretamente ao gráfico, evitando duplicações.
- **Melhoria na Legibilidade:** Adicionar rótulos ao gráfico para indicar claramente as percentagens associadas a "com especialidade" e "sem especialidade".
- **Aprimoramento Estético:** Considerar o uso de cores mais significativas ou uma paleta de cores específica para melhorar a estética do gráfico.
- **Personalização Adicional:** Adicionar opções de personalização, como a capacidade de escolher cores ou estilos para o gráfico.
- **Validação de Dados:** Realizar uma validação de dados para garantir que os resultados da consulta SQL sejam coerentes e representativos.

3ª QUESTÃO:**Pergunta SQL:**

“SELECT TRIM(f.formacao) as formacao, ROUND((COUNT(h.id_formacao) / (SELECT COUNT(id_docente) FROM habilitacoes))*100,2) AS media FROM 'habilitacoes' h INNER JOIN formacao f ON h.id_formacao = f.id GROUP BY h.id_formacao;”

Resultados Alcançados:

formacao	media
Mestre	25.20
Licenciado (grau obtido antes de 2008)	1.60
Grau de Mestre ou equivalente obtido no estrangeir...	0.29
Doutor	12.07
Grau de Doutor ou equivalente obtido no estrangeir...	1.10
Licenciado (grau obtido em 2008 ou após)	0.34
DESE	2.41
Licenciado	35.41
Bacharel	5.30
Master	0.47
Especialização de pós-licenciatura	1.65
DEA	0.58
Outros-Curso de qualificação profissional	0.21
Outros-Nível de ensino não superior	0.13
Outros-Frequência de ensino superior	0.05
Doctor of Philosophy	0.08
Outros-Curso médio	0.08
Licenciatura	6.30
Doutoramento	0.68
Bacharelato	1.18
Mestrado	3.46
Pós-graduação	0.47
Prov. Apt. Ped. Cien.	0.05
Maîtrise	0.05
Docteur	0.03

Figura 5.: Resultado da pergunta 3 executada na query.

O código realiza uma consulta ao banco de dados para obter a contagem de docentes agrupados por formação, calculando a média percentual para cada formação relativamente ao total de docentes. Os resultados são utilizados para gerar dados para um gráfico de rosquinha (doughnut chart) utilizando a biblioteca Chart.js.

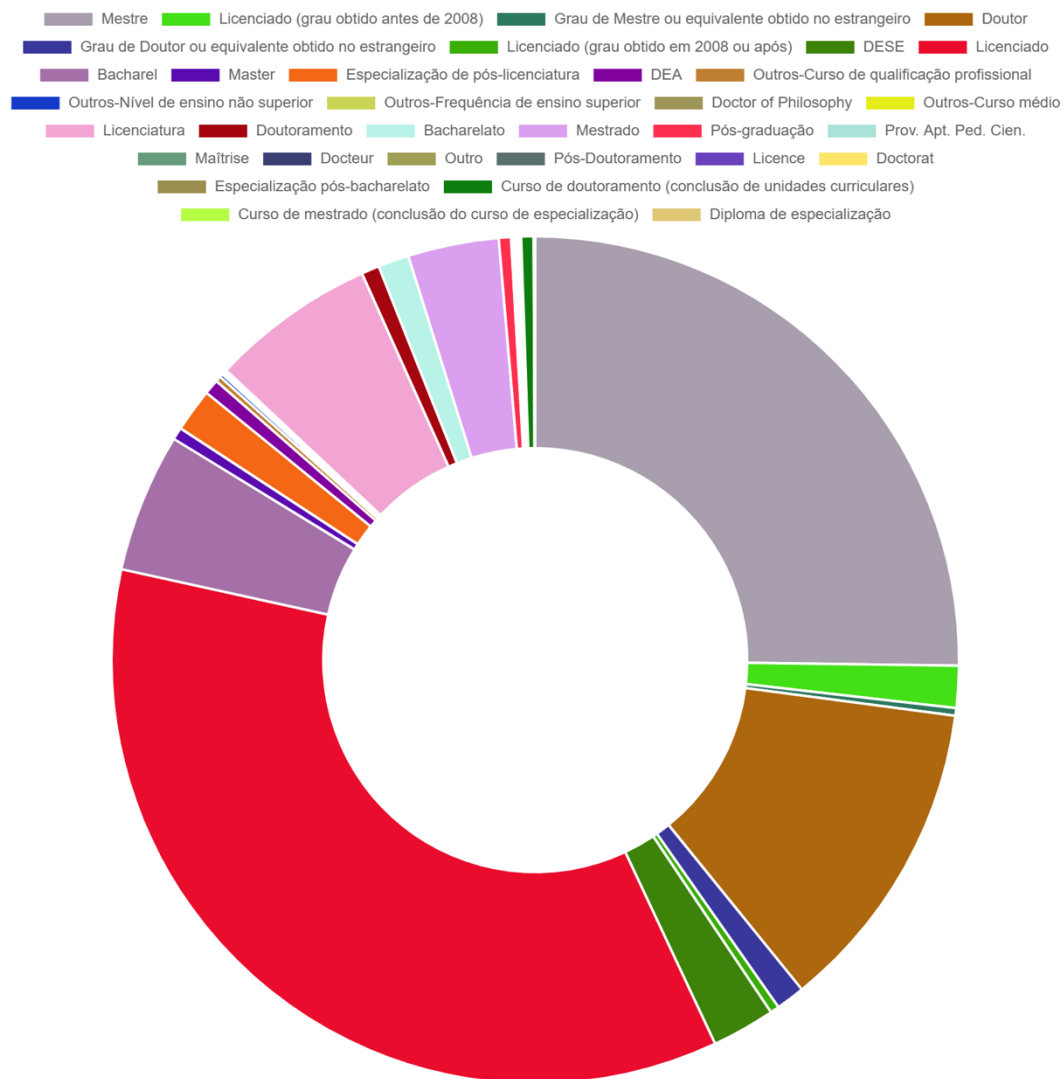


Figura 6.: Gráfico referente à pergunta 3.

Limitações:

- **Nomenclatura e Semântica:** O uso de nomes de variáveis como \$labels, \$data, e \$backgroundColor sugere uma estrutura semelhante à anterior, mas a variável \$labels aqui representa as formações, enquanto na anterior representava as percentagens pertencentes e não pertencentes. A consistência na nomenclatura pode melhorar a legibilidade.
- **Tratamento de Dados:** Se a consulta SQL não retornar resultados para uma formação específica, o código pode não lidar adequadamente com essa situação. Pode ser útil incluir verificações para garantir que os arrays utilizados na construção do gráfico estejam completos.

Trabalho Futuro:

- **Comentários e Documentação:** Incluir comentários explicativos no código, detalhando a lógica por trás da consulta SQL, do processamento dos resultados e da geração do gráfico. Isso facilitará a manutenção futura e a compreensão do código por outros desenvolvedores.
- **Aprimoramento Visual:** Considere adicionar rótulos, legendas ou outras informações visuais ao gráfico para tornar a apresentação mais informativa e compreensível para os utilizadores.
- **Reestruturação de Código:** Avalie a possibilidade de reestruturar o código para torná-lo mais modular e reutilizável. Isso pode facilitar a manutenção e expansão futuras.
- **Testes e Validação:** Realize testes rigorosos para garantir que o código funcione corretamente com diferentes conjuntos de dados e em diferentes cenários.
- **Melhorias de Layout:** Considere ajustes no layout da página da web para garantir uma experiência de utilizador mais agradável e responsiva.

Ao abordar esses pontos, o código pode ser aprimorado em termos de funcionalidade, legibilidade e usabilidade.

4ª QUESTÃO:

Pergunta SQL:

```
SELECT
    COALESCE(e.especialidade, 'Sem Especialidade') AS especialidade,
    ROUND((COUNT(DISTINCT h.id_docente) / (SELECT COUNT(DISTINCT id_docente)
FROM habilitacoes) * 100), 2) AS Percentagem_especialidade
FROM
    habilitacoes h
LEFT JOIN
    especialidade e ON h.id_especialidade = e.id
GROUP BY
    e.especialidade;
```

Resultados Alcançados:

especialidade	Percentagem_especialidade	especialidade	Percentagem_especialidade
Sem Especialidade	89.90	especialidade de Mecânica	0.41
Área de especialização: Automação Industrial	0.08	especialidade de Política Internacional	0.25
Área de especialização: Ciências Jurídico-Política...	0.08	Especialidade de Pré-História e Arqueologia	0.57
Área de especialização: Engenharia Térmica	0.08	especialidade de Psicologia da Educação	0.08
Área de especialização: Informática Industrial	0.16	Especialidade: Engenharia e Gestão de Sistemas de ...	0.16
Área de especialização: Marketing e Negócios Inter...	0.08	Especialidade: Geografia Humana	0.08
Área de especialização: Sistemas de Energia	0.16	especialização em Ética e Filosofia Social	0.16
Área de especialização: Telecomunicações	0.16	especialização em Física Nuclear	0.41
Art, Craft & Design Education	0.08	especialização em Formação Psicológica de Professo...	0.08
Auditoria	0.08	especialização em Geografia Regional	0.25
Automação, Controlo e Instrumentação	0.08	especialização em Intervenção Precoce	0.08
Ciências do Desporto	0.08	especialização em Pedagogia	0.08
Científico	0.16	especialização em Telecomunicações	0.33
Construções Mecânicas	0.08	especialização em Teoria e Desenvolvimento do Curr...	0.08
Consulta Psicológica de Jovens e Adultos	0.08	Estudos Ingleses	0.08
Desigualdades, Cultura e Território	0.08	Finanças	0.08
Desporto de Recreação e Lazer	0.08	Formação Educacional	0.08
Desporto de Rendimento	0.08	Geografia	0.08
Educação Artística na Infância	0.08	Geografia Regional	0.16
Energia	0.08	Gestão Desportiva	0.08
Engenharia e Gestão de Sistemas de Informação	0.08	Informática e Sistemas	0.08
Engenharia Térmica	0.08	Informática Industrial	0.16
Ensino da Música - Instrumento	0.16	Literaturas Africanas de Língua Portuguesa	0.08
especialidade de Ciências do Desporto	0.08	Marketing	0.08
especialidade de Geografia	0.25	Métodos Quantitativos de Gestão	0.16

Figura 7.: Resultado da pergunta 4 executada na query.

O código realiza uma consulta SQL para obter a contagem de docentes por especialidade e calcula a percentagem correspondente. Os resultados são exibidos num gráfico de rosquinha usando a biblioteca Chart.js. O gráfico fornece uma visualização clara da distribuição de docentes por especialidade, destacando a percentagem em cada categoria.

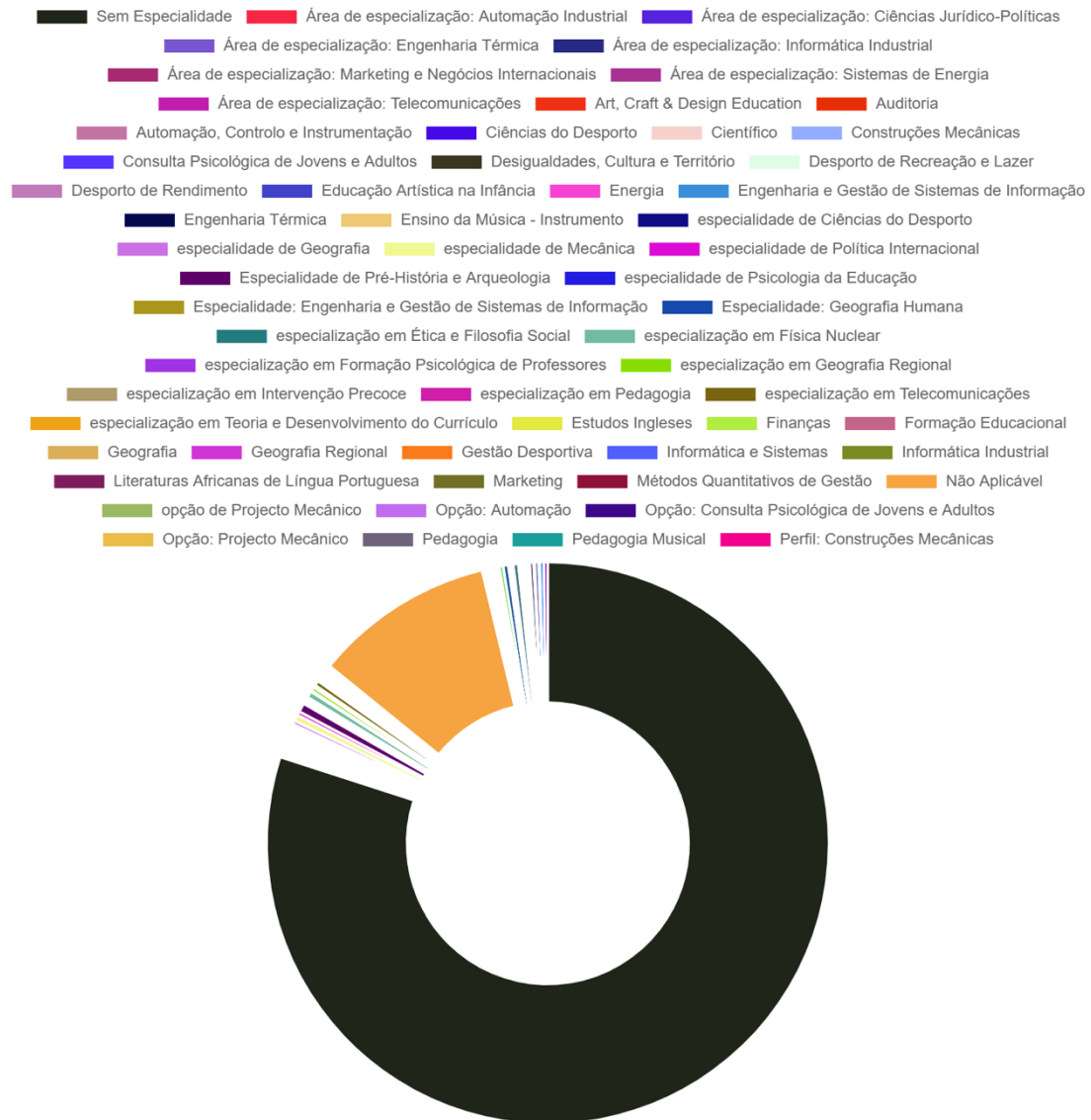


Figura 8.: Gráfico referente à pergunta 4.

Limitações:

- **Dados de Exemplo:** O código depende de dados específicos do banco de dados, e as limitações podem surgir se os dados de exemplo não forem representativos ou houver-se falta de dados em algumas áreas.
- **Estilo Padrão:** A geração aleatória de cores para o gráfico pode resultar em cores não relacionadas entre si. Isso pode afetar a legibilidade e a estética do gráfico.

Trabalho Futuro:

- **Aprimoramento da Interface:** Considerar aprimoramentos na interface do utilizador, como adicionar rótulos ao gráfico para melhorar a compreensão.

- **Interatividade:** Adicionar recursos interativos ao gráfico, como a capacidade de clicar numa fatia para obter mais detalhes.
- **Melhorias de Desempenho:** Otimizar consultas SQL e o código PHP para garantir um desempenho eficiente, especialmente com grandes conjuntos de dados.
- **Personalização de Cores:** Permitir que o utilizador escolha ou defina cores para as fatias do gráfico.
- **Responsividade:** Garantir que a página e o gráfico sejam responsivos para diferentes tamanhos de tela.

MATERIAIS E MÉTODOS

LINGUAGEM UTILIZADA:

HTML (HYPERTEXT MARKUP LANGUAGE):

- O HTML é a espinha dorsal de qualquer página web, sendo responsável pela estruturação e marcação do conteúdo. Além das tags básicas, você pode utilizar recursos avançados, como HTML5, para incorporar elementos multimídia, formulários mais avançados e até mesmo geolocalização.

PHP (HYPERTEXT PREPROCESSOR):

- O PHP é uma linguagem server-side, amplamente utilizada para processar dados do lado do servidor. Além da conexão com o banco de dados, ele pode ser usado para manipular arquivos, enviar e-mails e realizar diversas tarefas.

PHPMYADMIN:

- O phpMyAdmin simplifica a administração de bancos de dados MySQL, proporcionando uma interface intuitiva. Além de gerir tabelas, você pode realizar operações avançadas, como otimização de tabelas e execução de consultas SQL complexas.

JAVASCRIPT:

- O JavaScript é uma linguagem client-side que adiciona interatividade às páginas web. Além das funcionalidades básicas, você pode utilizar bibliotecas como jQuery ou frameworks como React e Vue.js para desenvolvimento mais avançado.

BASE DE DADOS:

- **MySQL:** O código está configurado para interagir com um banco de dados MySQL. Ele se conecta a um banco de dados local (localhost) usando as credenciais fornecidas ('root' como usuário). O nome do banco de dados é especificado como 'dcw'. Além disso, há um código SQL para inserir dados do DataFrame do Pandas no banco de dados.

MANIPULAÇÃO DE DADOS:

- **Chart.js:** O Chart.js é uma biblioteca JavaScript para criação de gráficos interativos e dinâmicos no navegador. Ele oferece suporte a vários tipos de gráficos, como barras, linhas, circular, entre outros.

FORMATO DE SAÍDA:

- **Chart.js:** Este código PHP e HTML visa criar um gráfico de rosquinha dinâmico, utilizando a biblioteca Chart.js, para visualizar as médias de formações a partir de dados armazenados num banco de dados MySQL.
- **Conexão com o Banco de Dados:** O código inicia a estabelecer uma conexão com o banco de dados, utilizando um arquivo de configurações (settings.php) e um script de banco de dados (db.php).
- **Consulta ao Banco de Dados:** Uma consulta SQL é executada para obter informações relacionadas a formações e as suas médias a partir de tabelas específicas no banco de dados. Os resultados da consulta são armazenados em um array associativo (\$arr). Um loop é utilizado para percorrer os resultados da consulta e preparar três arrays: \$labels para rótulos das fatias, \$data para os valores das fatias e \$backgroundColor para cores aleatórias das fatias.
- **HTML e Inclusão do Chart.js:** A estrutura básica do HTML é definida, incluindo referências a bibliotecas externas, como o Bootstrap e o Chart.js.
- **Criação do Gráfico circular:** Um script JavaScript é incorporado ao final do documento HTML para criar um gráfico de rosquinha usando o Chart.js. O contexto do canvas é obtido, e um objeto Chart é instanciado, configurando o tipo do gráfico como 'doughnut'. Os rótulos, dados e cores são passados para o gráfico, resultando num gráfico de rosquinha interativo.

Esse código permite visualizar de forma dinâmica e atrativa as médias de formações a partir de dados armazenados num banco de dados MySQL. A utilização do Chart.js facilita a criação de gráficos interativos diretamente na página web, proporcionando uma representação visual das informações extraídas do banco de dados. A inclusão do Bootstrap ajuda a criar um layout responsivo e esteticamente agradável.

CONCLUSÃO

Ao longo deste trabalho, abordamos dois momentos distintos de avaliação na Unidade Curricular de Processamento de Informação, guiados pelo professor José Monteiro. No primeiro momento, concentramo-nos no desenvolvimento de um modelo de extração de conteúdo de informação, resultando na automação da obtenção do conjunto de dados completo num sistema de gestão de base de dados. Agora, na segunda fase, ampliamos a complexidade do projeto ao envolver a normalização da base de dados e garantir a qualidade dos dados extraídos.

Durante a aula do dia 9 de janeiro, o nosso grupo empreendeu tarefas cruciais, desde a associação em grupos até a formulação de questões de investigação relevantes. Realizamos consultas estruturadas à base de dados para responder a essas questões, culminando numa análise detalhada dos resultados obtidos.

A preparação metódica deste relatório reflete o nosso compromisso com a clareza e organização na comunicação dos processos e resultados. A formulação e refinamento das questões de investigação demonstram a nossa habilidade em direcionar a análise para a normalização da base de dados e a garantia da qualidade dos dados.

Os materiais e métodos utilizados para as consultas à base de dados revelam uma abordagem técnica sólida e a capacidade de aplicar conceitos teóricos à prática. A apresentação visual dos resultados, por meio de gráficos e tabelas, visa facilitar a compreensão e interpretação dos dados.

Em conclusão, este relatório representa não apenas o produto final dos nossos esforços durante esta fase, mas também reflete a colaboração eficaz e as contribuições individuais de cada membro do grupo. As conclusões obtidas oferecem insights valiosos, não apenas em resposta às questões de pesquisa formuladas, mas também na compreensão mais profunda da normalização da base de dados e da qualidade dos dados. Este trabalho reforça a importância do processamento de informação de maneira eficiente e estruturada para extrair informações significativas e confiáveis a partir de conjuntos de dados complexos no contexto do Ms Excel.

BIBLIOGRAFIA

ChartsJS. (2017, October 3). Chart.js. <https://www.chartjs.org>

Jonatas Silva do Espirito Santo. (2021, March 21). *Python: Manipulando dados armazenados Em bancos MySQL.* Medium. <https://jonates.medium.com/python-manipulando-dados-armazenados-em-bancos-mysql-45ba477b19d1>

XAMPP. (2012, July 4). XAMPP Installers and Downloads for Apache Friends. <https://www.apachefriends.org>

W3school. (2018, February 4). W3Schools Online Web Tutorials. <https://www.w3schools.com>

ANEXOS

(Todos os anexos deste trabalho encontram-se na devida pasta inserida no inforestudante.)