



Lisbon School
of Economics
& Management
Universidade de Lisboa

Mestrado

Gestão de Sistemas de Informação

Trabalho Final de Mestrado

Projeto

Relatórios Semi-Automatizados

Gonçalo Garcia Gonçalves

outubro - 2024



Mestrado

Gestão de Sistemas de Informação

Trabalho Final de Mestrado

Projeto

Relatórios Semi-Automatizados

Por Gonçalo Garcia Gonçalves

Orientador: Carlos J. Costa

outubro - 2024

Agradecimentos

Começo por agradecer a todos que, de alguma forma, se cruzaram comigo neste percurso e contribuíram, direta ou indiretamente, para que eu pudesse adquirir o conhecimento e as condições necessárias para finalizar este projeto.

Em particular manifesto gratidão aos meus pais, que me deram apoio incondicional especialmente nos momentos difíceis desta jornada, encorajando-me e garantindo as melhores condições para ter sucesso.

Ao Professor Carlos J. Costa, que me orientou e acompanhou neste Trabalho Final de Mestrado, agradeço sinceramente pela paciência, pelo valioso apoio e conhecimento prestado.

Por último, mas igualmente significativo, expresso o meu reconhecimento à empresa Qmetrics, pelo acolhimento e pela disponibilidade demonstrada, que foram fundamentais para o desenvolvimento deste estudo.

Resumo

A tomada de decisão depende de informação de qualidade e atualizada. Neste sentido torna-se especialmente importante recolher e tratar dados e apresentar informação de forma mais adequada e próxima do tempo-real.

É neste contexto que se desenvolve este projeto que tem por objetivo principal a proposta de um modelo preliminar de relatório gerado de forma semi-automática, respeitando boas práticas de visualização de dados. Para atingir o objetivo foi utilizada uma metodologia que segue o paradigma de *Design Science*. Este recurso é relevante uma vez que, sem descurar o rigor científico, é orientado para a prática, contribuindo para a resolução de problemas concretos.

No contexto de visualização de dados, a automação é uma ferramenta valiosa para empresas de consultoria, pois permite uma resposta rápida e eficiente às necessidades dos clientes.

Este trabalho pretende também contribuir para o desenvolvimento tecnológico da empresa, proporcionando a perspetiva de integrar tecnologias de automação visando aprimorar ainda mais a qualidade e a rapidez na elaboração de relatórios.

Palavras-Chave

Visualização de dados; *Design Science*; relatórios; *Business Intelligence*; *Dashboards*.

Índice

1.Introdução	1
1.1 Enquadramento e motivação	1
1.2 Objetivos do projeto	3
1.3 Relevância dos objetivos.....	3
2. Revisão de Literatura	5
2.1 Visualização de dados	5
2.2 Business Intelligence.....	8
2.3 Dashboards	9
3. Metodologia	12
4. Implementação.....	15
5. Conclusões.....	26
Referências Bibliográficas.....	28
ANEXOS.....	35
Anexo I - Manual do projeto no RStudio	35

Lista de Figuras

Figura 1- Processo de Design Science.....	14
Figura 2- Diagrama UML das atividades da ferramenta.....	16
Figura 3- Modelo Global de PowerPoint	22
Figura 4- Integração dos dados no software	23
Figura 5- Exemplo de um slide e gráfico.....	24

Lista de Tabelas

Tabela 1- Definições dos conceitos de dados, informação e conhecimento, segundo Russel Ackoff, a nível cognitivo	6
Tabela 2- Definições dos conceitos de dados, informação e conhecimento, segundo M.Chen et al., a nível computacional	6

Lista de abreviaturas e expressões

BI: Business Intelligence

UML: Linguagem de Modelagem Unificada

PPT: PowerPoint

Chunk de código- é um bloco ou segmento de código que realiza uma tarefa específica dentro de um programa

1.Introdução

1.1 Enquadramento e motivação

Na dinâmica das organizações modernas é desejo generalizado a garantia do sucesso da produtividade empresarial. Este desejo é o pilar que sustenta a motivação deste trabalho. Embora não haja uma solução única para se alcançar esse desejo, existem vários princípios que podem ajudar a sua concretização.

Na sociedade atual o recurso à tecnologia e à informação é uma realidade recorrente e eficaz. Nos últimos vinte anos verificou-se que a revolução dos dados é uma das qualidades fundamentais e adequadas a um funcionamento organizacional competente.

Inúmeros especialistas enquadram as preferências de decisão profissional e de gestão num ambiente de crescente mudança, de uma prática fundamentada na intuição, para uma prática progressivamente centrada em dados. (May 2009; Davenport 2010; Brynjolfson et al. 2011).

O progresso tecnológico e a transformação digital provocaram um aumento extraordinário na criação de dados; assim, tornou-se essencial usar esses dados para adquirir informação e conhecimento. (Vázquez-Ingelmo et al., 2019).

A competência em obter, guardar e pesquisar dados promoveu categoricamente o método científico na prática de gestão. Hoje as instituições têm a capacidade de recolher, de forma contínua e ao longo do tempo, dados empíricos para examinar ou aferir hipóteses e, a partir daí, produzir novas observações e novos *insights* sobre processos organizacionais (Tavares & Costa, 2007, Parente & Costa, 2022).

Uma vez que a Inteligência Empresarial (BI) tem as suas origens na abordagem científica, o seu uso foi validado e difundido com celeridade pelas instituições. Por isso, o conceito e o uso das tecnologias de BI adquiriram relevância (Marques & Costa,2019).).

Ultimamente, os investimentos em BI cresceram em comparação com os orçamentos gerais de Tecnologias de Informação (TI): apenas no ano anterior, as despesas com BI

tiveram um acréscimo a escala global de 10,5 bilhões de dólares, ao passo que os orçamentos de TI se mantiveram inalterados (Gartner 2011a).

Sobre este assunto, Luftman e Zadeh (2011) reconheceram o BI como a tecnologia mais impactante nas organizações. Por sua vez, a Gartner (2011) posicionou o BI entre as dez principais tecnologias estratégicas para 2012. Enquanto isso, Brynjolfson et al. (2011) provaram que a aceitação de tecnologias de BI e que as decisões apoiadas em dados provocam um crescimento na produtividade das empresas na ordem de 5% a 6%.

A necessidade empresarial de desenvolver um produto que agilize o processo de tratamento de dados para responder o mais rápido e eficazmente possível às exigências dos clientes foi sem dúvida o fator impulsionador deste trabalho. É importante criar um relatório preliminar simples e rápido que qualquer colaborador da empresa possa replicar, antes dos resultados estarem fechados.

A missão central da empresa Qmetrics é fornecer produtos e/ou serviços de qualidade que assegurem uma experiência positiva aos seus clientes. A pesquisa deste estudo recaiu nesta empresa pelo facto de eu ter sido colaborador e ter integrado na prática a sua equipa de trabalho.

Muitos dos estudos concebidos na empresa têm um relatório preliminar simples, antes dos resultados estarem fechados. Este produto dá a possibilidade aos clientes de já terem alguma informação e compreender o projeto. No contexto profissional, foi preciso desenvolver uma possível solução mais rápida para a criação de apresentações PowerPoint com dados para o cliente. Sendo que esta apresentação não é a final, mas sim uma apresentação preliminar para deixar o cliente informado com os dados até ao momento.

1.2 *Objetivos do projeto*

O objetivo principal é propor uma solução que integre dados com a reprodução de relatórios a partir de dados estatísticos, sempre com foco no sucesso empresarial, permitindo através do artefacto a reprodução de modelos standard de relatórios simples.

Com vista a atingir este objetivo, é utilizada uma abordagem metodológica baseada no paradigma de *Design Science* (Aparicio, et al,2023) . Para alcançar na integra a finalidade deste projeto é preciso concluir os seguintes passos:

1. Escolher a ferramenta/softawre para o desenvolvimento do projeto;
2. Definir gráficos relevantes para utilização;
3. Codificar a solução em R;
4. Apresentar conclusões e ilações.

1.3 *Relevância dos objetivos*

Em contexto empresarial a relevância dos objetivos é importante pois orienta esforços e recursos, dirigindo o foco de todos para um propósito comum. Na verdade, quando os objetivos são claros e significativos todos os envolvidos sabem para onde vão e quais são as prioridades. Isso proporciona um ambiente de trabalho colaborativo entre os integrantes, otimizando recursos e tempo. Além disso, a clareza nos objetivos minimiza mal-entendidos e conflitos, já que promove um entendimento mútuo entre os membros das equipas. Por consequência a produtividade aumenta, a motivação fortalece e cresce uma consciência de meta coletiva.

A proposta do projeto concebida para a empresa Qmetrics pretende reforçar a sua competitividade no mercado através de um processo mais rápido, simples e sem custos adicionais na sua produção. Além disso, propõe semi-automatizar o processo de produção de relatórios dos estudos encomendados.

A nível da revisão da literatura, este projeto procura contribuir para o enriquecimento de temáticas importantes, nomeadamente visualização de dados, *Design Science*, BI e

dashboards. A intenção é analisar e desenvolver o conhecimento e a interligação entre os temas que têm sido bastante debatidos no contexto atual de inovação tecnológica e transformação digital.

2. Revisão de Literatura

Neste capítulo, é realizada uma revisão de literatura abrangendo temas e tópicos relevantes e relacionados com a temática desenvolvida no projeto.

São apresentadas de forma concisa as principais análises e considerações feitas por diversos autores, além de trabalhos relacionados com o tema em estudo.

Sendo, o objetivo deste projeto a proposta de uma solução capaz de conceber relatórios alimentados por dados estatísticos.

Assim, faz-se uma abordagem ao sistema de visualização de dados. De seguida, descreve-se de forma breve o conceito de BI e o trabalho prossegue com a apresentação do conceito de *dashboards*.

2.1 Visualização de dados

Habitualmente, define-se a visualização de dados como um sistema de métodos para representar informação gráfica de modo acessível e compreensível, cuja estética deve ser trabalhada de forma que a sua apresentação prenda o interesse do destinatário (Aparicio & Costa, 2015).

Independentemente das etapas que compõem o processo de visualização, há dois aspetos fundamentais a ter em conta ao longo do mesmo: “O Quê?” e “Como?”. Em primeiro lugar, é essencial definir o público-alvo, o domínio envolvido e as ações necessárias para a comunicação visual dos dados. Em seguida, é preciso estabelecer a codificação visual dos dados, as diretrizes ou recomendações de *design*, os elementos visuais que auxiliam as tarefas e requisitos, bem como o nível de aceitabilidade por parte do público-alvo (Costa & Aparicio, 2005. Reynoso & Diván, 2020).

Os conceitos de dados, informação e conhecimento podem ter, ainda, significados diferentes conforme a área retratada: área de domínio, área computacional (Tabela 2) ou área cognitiva (Tabela 1) (Chen et al., 2009). Não obstante, estes indicadores são

colaboradores na produção de visuais e proporcionam a tomada de decisão mais correta consoante as necessidades das empresas.

Categoria	Definição
Dados	Símbolos que traduzem acontecimentos ou objetos
Informação	Dados processados com o intuito de serem úteis para responder às questões “quem”, “o quê”, “quando” e “quanto”
Conhecimento	Responde à questão “como” aplicando os conceitos de dados e informação

Tabela 1- Definições dos conceitos de dados, informação e conhecimento, segundo Russel Ackoff, a nível cognitivo (1989)

Categoria	Definição
Dados	Representações computadorizadas de modelos e atributos de entidades reais ou simuladas
Informação	Dados que representam os resultados de um processo computacional, como uma análise estatística, para atribuir significados aos dados ou as transcrições de alguns significados atribuídos por seres humanos
Conhecimento	Dados que representam os resultados de um processo cognitivo simulado por computador, como percepção, aprendizagem, associação e raciocínio ou as transcrições de algum conhecimento adquirido por seres humanos.

Tabela 2- Definições dos conceitos de dados, informação e conhecimento, segundo Chen et al (2009)

No âmbito computacional estes conceitos ostentam definições diferenciadas. Todavia, no decorrer do processo de visualização os domínios cognitivo e computacional relacionam-se entre si (Aparicio et al, 2024).

Nunca será demais referir que a visualização de dados é uma área fundamental da ciência de dados, que engloba a representação gráfica de informações complicadas que permitem a compreensão e a tomada de decisões.

Com o constante aumento de dados disponíveis, a visualização de dados tem apresentado uma crescente importância, possibilitando aos utilizadores o entendimento de padrões, tendências e *outliers* de modo notável.

De acordo com a visualização de dados deve-se converter dados abstratos numa configuração visual que sustente a compreensão humana (Card, Mackinlay e Shneiderman, 1999). Para representar dados quantitativos utilizam-se técnicas como gráficos de linhas, dispersão e barras, ao passo que mapas de calor e gráficos em rede são utilizados para representações mais complexas, nomeadamente visualizações multidimensionais e relacionais (Costa e Aparicio, 2019).

Cleveland e McGill (1984) descobriram que a exatidão na interpretação de gráficos muda conforme a categoria de gráfico usada, sendo os gráficos de barras e linhas mais meticulosos do que os gráficos circulares para comparações de magnitude. Essa investigação realça o interesse de preferir o método de representação visual apropriado ao tipo de dado e ao objetivo da investigação.

Os progressos tecnológicos promoveram o aparecimento de várias ferramentas e softwares que ajudam a visualização de dados. Wickham (2010) inovou a forma como os gráficos são produzidos lançando o pacote ggplot2 no R. A produção é mais instintiva e individualizada mediante uma gramática de gráficos. Esse pacote é bastante usado na academia e na indústria por ser adaptável e produzir visualizações complexas com código moderadamente simples.

Em ambiente empresarial ferramentas como o Tableau e o Power BI têm-se evidenciado pela sua faculdade de incluir amplos volumes de dados e gerar *dashboards* interativos. Few (2013) defende que a interatividade cedida por estas ferramentas permite que os utilizadores analisem dados em tempo real, adaptando visualizações para adquirir *insights* mais profundos.

É certo que a visualização de dados tem um papel fundamental na escolha de decisões informadas. Na ótica de Tufte (2001) boas visualizações de dados reduzem o "*ruído gráfico*" e aumentam a "*taxa de dados*", prendendo o foco dos utilizadores nas informações importantes. Em contextos organizacionais, visualizações eficientes podem aprimorar positivamente a comunicação de resultados analíticos, reconhecendo oportunidades e perigos.

Não obstante os progressos, a representação de dados lida com obstáculos, como a exigência de harmonizar simplicidade e complexidade. Na interpretação de Cairo (2016) as visualizações demasiado simplificadas podem deixar de lado informações essenciais, enquanto visualizações demasiadamente complexas podem ocupar excessivamente o utilizador, afetando a interpretação. Heer e Bostock (2010) enfocam outro desafio que é a interpretação incorreta de representações visuais, que pode acontecer por causa de escolhas desajustadas de escala e cores ou tipos de gráfico.

2.2 *Business Intelligence*

O conceito de *Business Intelligence* (BI) foi divulgado pela primeira vez por Luhn (1958). No início dos anos 70, os sistemas de apoio à decisão foram as primeiras aplicações para o auxílio à tomada de decisão, advindo da criação do BI (Watson & Wixom, 2007).

De acordo com Howard Dresner, da Gartner Research, designado usualmente como o pai da BI, esse termo define-se como sendo “*uma coleção de softwares e soluções para recolha, consolidação, análise e facilitar o acesso a dados de forma a permitir que os utilizadores tomem melhores decisões de negócio*” (Chan et al., 2009, p.96).

O BI consiste num sistema de processos, tecnologias e ferramentas que transformam dados não processados em informações importantes e necessárias para a análise de negócios. O uso de BI tem-se desenvolvido celeremente devido à exigência das empresas de fazer escolhas mais informadas e apoiadas em dados (Davenport, 2010).

Apesar das diversas definições atribuídas ao termo de BI, este conceito é geralmente compreendido como um conjunto de métodos, tecnologias e ferramentas utilizadas para otimizar a tomada de decisões empresariais. Nos últimos anos, este termo tem sido gradualmente substituído por “*Business Analytics*”. No entanto, a análise de dados não é um conceito recente – existem referências à análise empresarial desde a década de 1940. O interesse por essa abordagem intensificou-se no final dos anos 1960, quando os computadores começaram a ser aplicados em sistemas de apoio à decisão. No entanto, nesse período, as fontes de dados eram relativamente pequenas e estruturadas, e a maioria das atividades analíticas concentrava-se em análises

descritivas. A análise descritiva, também conhecida como BI ou relatórios empresariais, representa o nível mais básico da taxonomia da análise de dados. As atividades analíticas desta categoria concentram-se, essencialmente, na criação de relatórios que sintetizam as operações empresariais, permitindo obter informações sobre o estado atual ou passado da organização (Delen et al., 2018).

A literatura sobre BI enfatiza a sua relevância para a escolha de decisões estratégicas. Na perspetiva de Sharma et al. (2014), as tecnologias de BI facilitam às empresas reconhecer tendências, antecipar a procura e aperfeiçoar operações. Todavia, tais competências são censuráveis em ambientes de negócios fortemente competitivos, aí a capacidade de solução rápida e a previsão de mudanças são fundamentais para o êxito.

Outro fator importante na literatura é a relação entre BI e o desempenho organizacional. Segundo Popovič et al. (2012) a eficácia dos sistemas de BI deriva da qualidade dos dados, da infraestrutura tecnológica e da capacidade analítica dos utilizadores. Além disso, Elbashir et al. (2008) acentuam que a implementação de BI deve ser estruturada de acordo com as finalidades estratégicas da organização para potenciar repercussões marcantes no seu desempenho.

A aplicação de BI nas pequenas e médias empresas (PMEs) é ainda uma realidade em desenvolvimento. Scholz et al. (2010) estudaram os desafios e benefícios que as PMEs enfrentam ao adotar soluções de BI. Na sua perspetiva, apesar das barreiras que as PMEs encaram, como custo e complexidade, o BI é capaz de dar uma vantagem competitiva considerável ao proporcionar decisões assentes em dados mais explícitos.

2.3 Dashboards

Nos últimos tempos diversas investigações académicas relataram que o uso de *dashboards* como instrumentos de base à tomada de decisão têm evoluído de forma crescente. Os dashboards tornaram-se o objeto principal da emergência do campo da Gestão de Informação (Malik, 2005).

Dashboards são um tipo particular de Sistemas de Suporte à Decisão (Arnott e Pervan, 2005). Um *dashboard* é uma interface visual interativa que apresenta informações

agrupadas de forma precisa e atingível, possibilitando aos utilizadores examinar, ponderar e tomar decisões apoiadas em dados em tempo real (Few, 2006). Outra designação define *dashboards* como sendo painéis interativos de controle de gestão que apresentam a informação mais importante numa janela, de modo a atingir os objetivos determinados, possibilitando ao utilizador explorar, identificar e comunicar ações corretivas. A eficácia de um *dashboard* está claramente conectada à sua propriedade de sintetizar informações enigmáticas em representações visuais de fácil compreensão (Yigitbasioglu e Velcu, 2012). Sendo que, os *dashboards* servem principalmente para monitorizar e rastrear métricas, analisar tendências e identificar ações corretivas (Vora, 2009). Os *dashboards* de informação são atualmente ferramentas fundamentais para a compreensão e extração de conhecimento a partir de grandes volumes de dados, podendo apresentar-se de diversas formas. São utilizados para múltiplas finalidades, como a análise de dados em diferentes domínios, a explicação de conceitos, a geração de conhecimento e a validação de hipóteses, entre outros. A sua ampla disseminação e aplicação em diversos contextos tornam a definição destes sistemas uma tarefa complexa. Embora, por vezes, possa ser desafiador distinguir o que constitui ou não um dashboard de informação, este pode ser descrito como um conjunto de recursos visuais que permitem ao utilizador interpretar e obter *insights* a partir dos dados apresentados (Vázquez-Ingelmo et al., 2019).

Um dos tópicos fundamentais na análise de *dashboards* é o *design* e a facilidade de utilização. Bach et al. (2023) pesquisaram as técnicas preferíveis para o *design* de *dashboards*, destacando que a transparência e a clareza são fundamentais para a produtividade desses instrumentos. Propuseram que a decisão de gráficos ajustados e a organização coerente dos elementos na interface são cruciais, permitindo que os utilizadores interpretem os dados com precisão e rapidez.

Por outro lado, Wu et al (2019) pesquisaram a ligação entre a interatividade dos *dashboards* e o contentamento do utilizador. Concluíram que a personalização e a interação em tempo real aumentam substancialmente a praticidade e a credibilidade dos utilizadores nas decisões tomadas com base nos dados apresentados.

Várias pesquisas investigaram o efeito dos *dashboards* na tomada de decisão. Zingde e Shroff (2020) dirigiram uma pesquisa empírica em meios empresariais, verificando que o uso de *dashboards* leva a decisões mais informadas e ágeis, especialmente em situações onde a rapidez na resposta é fundamental. Também identificaram que *dashboards* com funcionalidades preditivas, podem oferecer *insights*, facilitando a previsão de tendências e a deteção de riscos.

3. Metodologia

A metodologia utilizada neste projeto de Trabalho Final de Mestrado baseou-se na *Design Science*. Este termo apareceu pela primeira vez num capítulo de Herbert Simon em *The Sciences of the Artificial* de 1996. Diversos autores, nomeiam *Design Science* para referir o campo de estudo que se foca na investigação de artefactos, construções e outros conceitos artificiais. Sendo que na fabricação de novos artefactos, existem sempre duas questões que estão agregadas “Porquê?” e “Como?”. Além dessas questões, também é necessário compreender a eficácia e eficiência do artefacto concebido. (Jarvien, 2000).

Cada projeto de Sistemas de Informação tem a sua natureza específica, o que pode tornar as pesquisas na área correspondente bastante diferente de projeto para projeto. Para que o projeto seja bem implementado é necessário ampliar o foco do estudo para abordar considerações organizacionais e comportamentais. De modo que, a implementação do sistema, seja o mais capaz possível e futuramente possa analisar o impacto nos indivíduos e na organização (Galliers e Land, 1987, Aparicio et al., 2023).

De modo a concretizar o paradigma de *Design Science*, Peffers et al., (2009), propõe o *Design Science Research* (DSR), apresentando um processo que contém seis fases que são iterativas. Essas fases são: 1ª Identificação do problema e Motivação; 2ª Definir objetivos para uma solução; 3ª Design e Desenvolvimento; 4ª Demonstração; 5ª Avaliação e 6ª Comunicação (Figura 1).

A 1ª fase, Identificação do problema e Motivação, consiste no seguinte: quanto melhor for definido e descrito o problema, melhor será o artefacto criado, a partir desse processo. A justificação do valor da solução motiva ao investigador a sua procura e a aceitação dos resultados. Nesta fase, apenas são necessários dois recursos, conhecimento relativo ao estado do problema e do valor da sua solução.

A 2ª fase consiste na definição dos objetivos da solução. Nesta fase, são inferidos os objetivos da solução a propor, a partir da definição do problema e conhecimento de

possíveis métricas. Os recursos necessários são o conhecimento do estado do problema e das soluções atuais.

A 3ª fase, é a parte do *design* e desenvolvimento. Nesta etapa, finalmente, é criado o artefacto como construções, modelos, métodos ou instanciações. Neste ponto os recursos necessários são os conhecimentos teóricos capazes de realizar um artefacto que pode solucionar o problema.

Na 4ª fase é demonstrado o artefacto concebido para resolver o problema. O único recurso necessário nesta etapa é o conhecimento de utilização do artefacto para solução da problemática.

A 5ª fase consiste na observação e avaliação da eficácia do artefacto na resolução do problema. Aqui, são comparados os objetivos propostos na 2ª fase com os resultados obtidos na demonstração do artefacto. Os recursos necessários são as técnicas de análise e as comparações métricas. Como referido, este processo de *Design Science* é iterativo, sendo que, a partir deste ponto, o processo pode voltar à 3ª fase de modo a melhorar o artefacto concebido até ao momento.

Na 6ª fase, Comunicação, é divulgado a sua importância, a problemática, o artefacto e a sua utilidade, o rigor da sua configuração e eficácia. A comunicação pode ser feita para outros investigadores, docentes, administradores e outras audiências relevantes. Para que tal aconteça, o recurso necessário é a cultura disciplinar. Contudo, o processo de *Design Science* não termina na 6ª fase. Devido à rápida evolução computacional e a novas necessidades que surgirão, poderá ser necessário regressar à 1ª fase, caso exista uma nova problemática que solicite um novo artefacto para o solucionar.

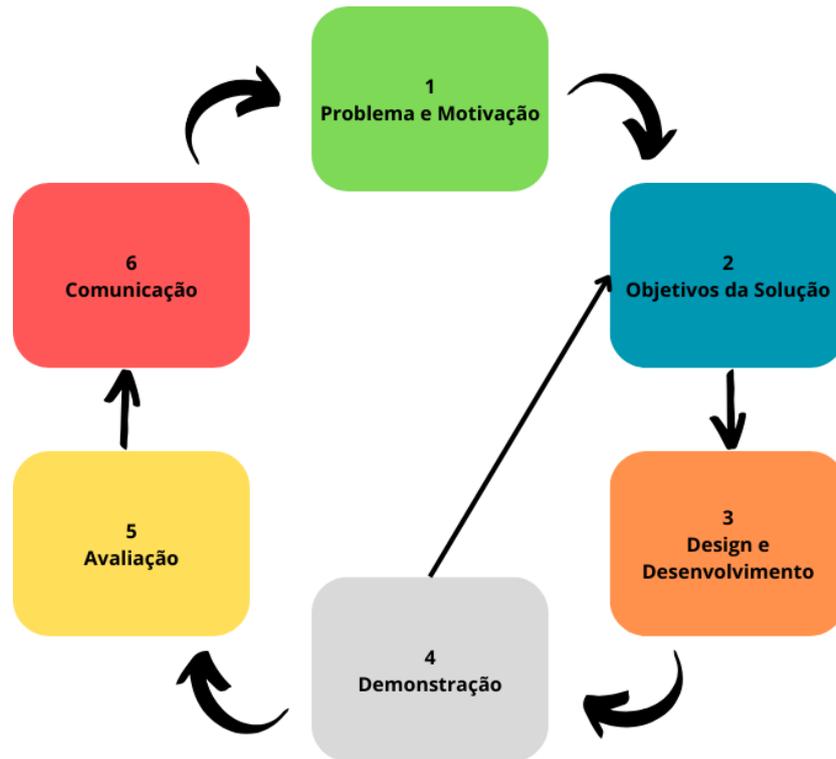


Figura 1- Processo de Design Science

Fonte: Elaboração própria

4. Implementação

O modelo proposto no processo de *Design Science* pressupõe a aplicação de seis passos os quais foram aplicados no âmbito deste Trabalho Final de Mestrado e no âmbito de propor um modelo capaz de dar resposta a uma necessidade empresarial.

- 1ª Fase - Identificação do problema e Motivação

Para identificação do problema o estudo dirigiu a sua atenção para a Qmetrics verificando que nesta empresa o processo de criação de relatórios preliminares era realizado manualmente. Ora os processos manuais são mais propensos a falhas, requerem maior coordenação e, além disso, consomem mais tempo. Neste enquadramento a empresa enfrentava alguns desafios para atender às exigências dos clientes e oferecer-lhes o acesso rápido e eficiente a informações essenciais para a sua tomada de decisão.

É a partir da tomada de consciência desta necessidade, ou seja, da importância de agilizar o processo de criação de relatórios preliminares que surgiu o problema e foi proposto no projeto.

Em relação à motivação principal do projeto, esta prendeu-se na melhoria do processo através da automatização, utilizando ferramentas de BI e visualização de dados, para gerar relatórios mais rapidamente sem comprometer a qualidade da informação fornecida.

- 2ª Fase - Definir objetivos para uma solução

Após o reconhecimento do problema, definiu-se como objetivo a criação de uma solução que permitisse a produção automatizada de relatórios preliminares de forma eficiente e de fácil utilização por qualquer colaborador da empresa.

Era essencial que a solução incluísse a implementação de uma ferramenta ou software que pudesse processar os dados recolhidos e gerar visualizações apropriadas em apresentações, a serem fornecidas aos clientes antes do relatório final. Deste modo, a

ferramenta deveria agregar os dados para, posteriormente, alimentar os gráficos criados no relatório ou *dashboard*. Para explicar de melhor forma o seguinte modelo Unified Modeling Language (UML) de atividades foi desenvolvido (Figura 2).

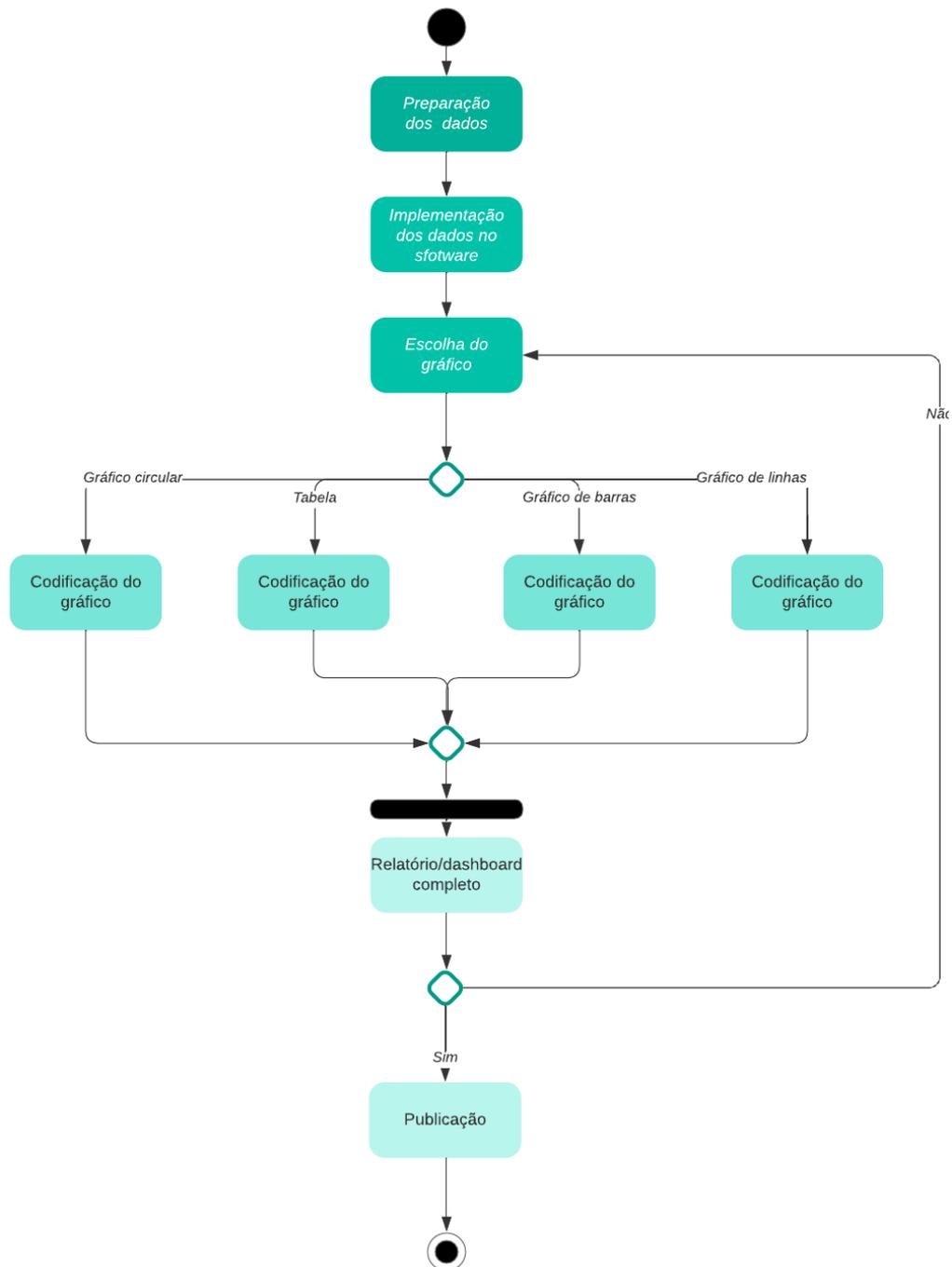


Figura 2- Diagrama UML das atividades da ferramenta

Fonte: Elaboração própria

A escolha de gráficos e a estrutura dos relatórios foram definidos com base em boas práticas de visualização de dados e de BI. Assim as projeções visuais continham informações visualmente atraentes e fáceis de entender.

A aprendizagem visual é uma das principais formas de interpretar informação, tendo historicamente combinado imagens, como gráficos e diagramas, com texto (Midway, 2020).

As organizações dependem cada vez mais de análises avançadas para extrair *insights* a partir de grandes volumes de dados. Esses *insights* ajudam a orientar a tomada de decisões e a criar vantagem competitiva. Como consequência, muitas organizações procuram formas eficazes para que os seus colaboradores analisem fluxos de dados e reajam com rapidez. A visualização de dados incorpora estratégias capazes de representação da informação para integrar, unificar e padronizar dados provenientes de diferentes fontes. Deste modo, facilita a interpretação dos dados pelos colaboradores, especialmente quando as visualizações estão incorporadas em painéis de controlo intuitivos. Representações visuais bem concebidas permitem o acesso a grandes quantidades de dados de forma segmentada e facilmente assimilável. Tal fato incentiva uma maior utilização dos dados, melhora a aprendizagem e facilita um melhor desempenho nas tarefas (Naidoo & Campbell, 2016).

As boas práticas tiveram em conta alguns dos princípios fundamentais apresentados no artigo “Princípio de Visualização de dados efetiva” (Midway, 2020):

- Esquematizar primeiro – Antes de criar um gráfico, determinar a mensagem central a transmitir, planear a estrutura da visualização antes de usar qualquer software;
- Usar o *software* certo – Escolher ferramentas adequadas para criar gráficos eficazes, *softwares* simples podem limitar a qualidade da visualização;
- Utilizar geometria e mostrar a informação – Escolher a melhor forma gráfica (barras, pontos, linhas etc.) para representar os dados e evitar gráficos que ocultem informações importantes;

- As cores representam significados – As cores devem ser usadas intencionalmente para transmitir informações e devem ser acessíveis para leitores daltónicos;
- Dados e modelos são conceitos diferentes – Diferenciar dados brutos de modelos estatísticos nos gráficos para evitar confusão na interpretação;
- Visuais simples, legendas detalhadas – Manter os gráficos simples e as legendas detalhadas para que a visualização seja autoexplicativa;
- Obter opinião – Solicitar *feedback* de colegas para garantir que os gráficos sejam compreensíveis e transmitam a mensagem corretamente.

Ainda que estas representações se concentrem no design para atrair o público, este não deve levar em conta a tecnologia de reprodução, a metodologia e o design, para que a sua perceção seja a mais verdadeira possível. Essas distrações podem afetar a experiência de visualização do espetador. É importante que vastos conjuntos de dados sejam estruturados e coerentes para serem interpretados da forma que se deseja.

- 3ª Fase - Design e Desenvolvimento

No contexto do projeto foi desenvolvida uma ferramenta que visa a criação de relatórios semi-automatizados e a organização de dados em informações visualmente organizadas, facilitando a interpretação dos resultados.

A ferramenta mantém o foco na simplicidade, oferecendo ao utilizador modelos de gráficos e tabelas já definidos, minimizando o processo de programação e de edição dos mesmos.

Neste sentido, a ferramenta rege-se por diversas atividades, as quais são estruturadas como está representado na Figura 2. Assim, a primeira atividade é a preparação e integração dos dados no software; de seguida as escolhas das visualizações mais indicadas para cada tipo de dados e variáveis e, por fim, a renderização do relatório.

Em primeiro lugar, é feita a recolha dos dados, que podem ser provenientes de diversas fontes, como base de dados, questionários, entrevistas, grupos focais, entre outros.

De seguida, é efetuado uma preparação dos dados, que passa pelos processos de transformação, organização e limpeza dos mesmos. Deste modo, tendo como *output* uma tabela de dados simples e de qualidade.

Subsequentemente, é necessário ligar os dados limpos ao *software* a ser utilizado. Este conjunto de dados permite que os utilizadores façam análises, treinem modelos ou que alimentem outras funcionalidades específicas.

Posteriormente, o utilizador começa a pensar na estrutura do relatório, através de uma análise exploratória dos dados ou de um questionário utilizado na extração dos dados. Nesta fase, inicia a criação de gráficos ou *dashboards* com recurso a modelos gráficos pré-concebidos. Estes modelos são o ponto fulcral do projeto, sendo esta a secção mais específica em termos de conhecimento, mas que permite ao utilizador selecionar um tipo de gráfico e as variáveis desejadas.

Quando, todos os gráficos ou *dashboards* necessários estão construídos é feita a renderização do relatório para partilha com o utilizador final.

- 4ª Fase - Demonstração

O seu desenvolvimento foi realizado utilizando a linguagem R, em conjunto com o *RStudio* e *R Markdown*, ferramentas que permitiram a criação automatizada dos relatórios em *PowerPoint*. Durante o desenvolvimento, foram selecionados gráficos relevantes e implementadas funcionalidades específicas para a organização e visualização dos dados de forma clara e atraente (Anexo I).

O *software* utilizado no projeto foi o *RStudio*, utilizando a ferramenta de *R Markdown* para produzir os visuais no relatório. Este *software* foi escolhido por diversos fatores, como a sua integração na unidade curricular de estatística, o que já favorece uma certa aptidão para trabalhar com o *software* e com a linguagem de R. Por outro lado, o facto de ser um *software* gratuito com a capacidade de trabalhar em ambiente de *desktop* e servidor remoto, dá a possibilidade de outros colaboradores utilizarem-no em qualquer computador e sem a necessidade de adquirir licenças adicionais para novos *softwares*.

Para mais, com a criação da ferramenta do *R Markdown*, apareceu uma forma simples de traduzir os gráficos criados em apresentações *PowerPoint*, e ainda é possível através do *RStudio* a programação em diferentes linguagens como *R*, *Python*, *Julia*, *SQL*, *Stan*, *JavaScript*, *C*, *C++*, *Fortran*, entre outras, exibindo um grande leque de escolhas para os diferentes graus de conhecimento dos utilizadores.

O *RStudio* é um Ambiente de Desenvolvimento Integrado para linguagem de programação *R* concebido para ser de fácil utilização para novos utilizadores e, simultaneamente, ser uma ferramenta de grande produtividade para os utilizadores mais experientes (Allaire, 2019).

A literatura sugere que o *RStudio* tem sido fundamental para a popularização do *R* na comunidade científica e empresarial. Wickham e Golemund (2017) destacaram que, ao fornecer uma interface gráfica intuitiva e ferramentas integradas como o *RMarkdown* e o *Shiny*, o *RStudio* tornou a programação em *R* mais acessível para utilizadores com diferentes níveis de conhecimento. Segundo Allaire et al. (2019), a interface do *RStudio* facilita a reprodução e a documentação dos resultados, promovendo práticas de ciência aberta.

O *RStudio* é amplamente utilizado para análises estatísticas em diversas disciplinas, desde as ciências sociais até à biologia. Osborne (2015) relata que o *RStudio*, ao integrar pacotes estatísticos avançados, permite que os investigadores conduzam análises complexas de forma eficiente. A interface também suporta a integração de pacotes de visualização de dados, como o *ggplot2*, que são fundamentais para a exploração e apresentação de dados (Wickham, 2016).

Na era do *Big Data*, o *RStudio* tem sido crucial para a realização de análises de dados em larga escala. De acordo com Xie, Allaire e Golemund (2018), o *RStudio* facilita a implementação de fluxos de trabalho completos de ciência de dados, desde a aquisição de dados até à modelagem preditiva e à comunicação de resultados. O uso de ferramentas como o *R Markdown* permite a criação de documentos dinâmicos que combinam código, texto e visualizações, o que é essencial para a comunicação de resultados em ciência de dados (Baumer, 2017).

Apesar destas vantagens, o *RStudio* não está isento de adversidades. Algumas pesquisas indicam que o ambiente pode ser intimidante para iniciantes, devido à necessidade de familiaridade com a programação em R (Fox & Weisberg, 2018). Aliás, a *performance* do *RStudio* pode ser um problema em análise de dados, em conjuntos de dados grandes ou complexos, exigindo *hardware* mais potente ou soluções de paralelização (Weston, 2015).

Numa descrição detalhada do processo para automatizar a criação de relatórios no *RStudio*, usando dados provenientes de ficheiros *Excel*, e gerar gráficos e tabelas, que serão posteriormente exportados para uma apresentação em *PowerPoint*. O processo é composto por seis passos.

1. Criar um projeto no *RStudio*: O primeiro passo é abrir o *RStudio* e iniciar um novo projeto. Para isso, navega-se até à aba "*File*" e seleciona-se a opção "*New Project*".

O utilizador pode optar por criar o projeto num diretório existente ou num novo diretório. Após definir o local, o projeto será criado e o ambiente de trabalho estará pronto para o desenvolvimento.

2. Criar um *template* de *PowerPoint* (PPT): Antes de gerar o relatório, é necessário criar um modelo de apresentação (*template*) no *PowerPoint*. Esse *template* incluirá slides padrão como título, conteúdo, cabeçalhos, etc (Figura 3).

Este modelo será utilizado pelo *RStudio* para formatar automaticamente os elementos gráficos e textuais conforme o conteúdo do relatório.

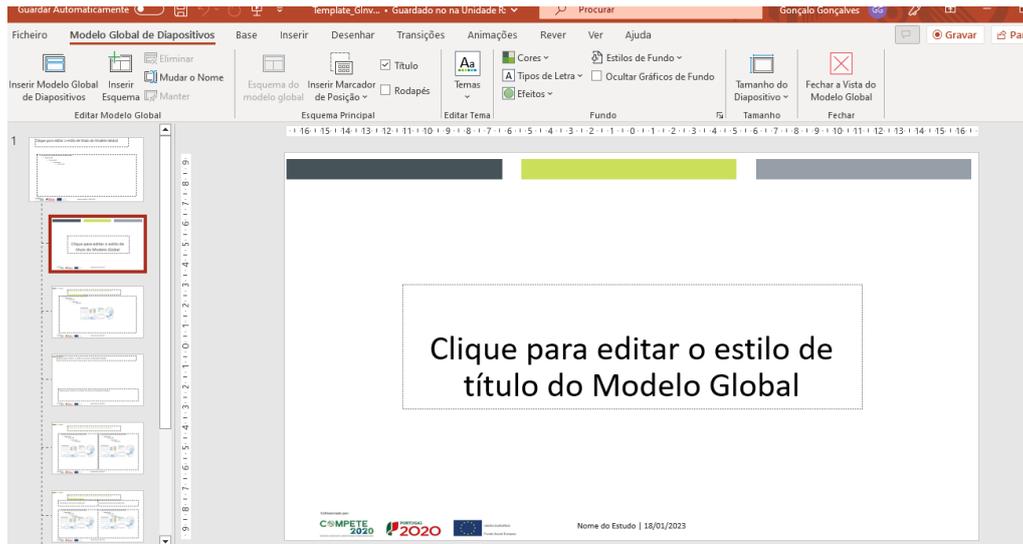


Figura 3- Modelo Global de PowerPoint

Fonte: Elaboração própria

3. Conectar e importar dados ao *RStudio*: Para começar a trabalhar com os dados, é necessário criar cópias dos ficheiros *Excel* de apuramento de frequências e médias e colocá-los na pasta onde está o projeto.

Estes ficheiros de cópias estão ligados aos ficheiros originais, permitindo a atualização automática sempre que o projeto for aberto.

Adicionalmente, é preciso criar um ficheiro *Excel* com *labels* e um ficheiro com as questões e respetivas variáveis. Esses dados são então importados para o *RStudio* utilizando a função "*From Excel*" na aba "*Environment*" (Figura 4).

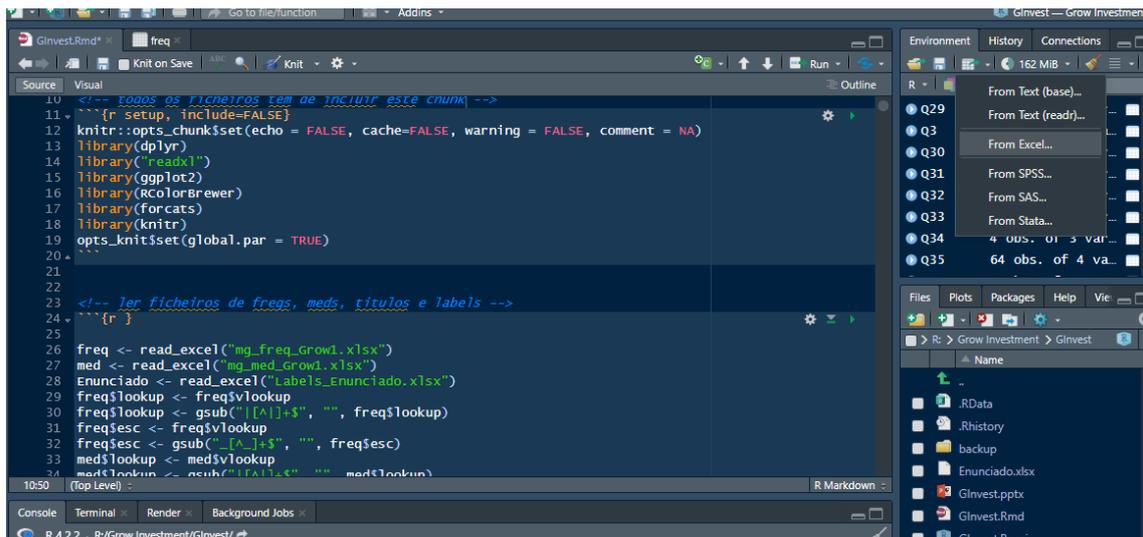


Figura 4- Integração dos dados no software

Fonte: Elaboração própria

4. Gerar gráficos em R: A forma semi-automatizada do relatório são os *chunks* de código já concebidos para posterior utilização de qualquer colaborador.

O documento descreve como construir diferentes tipos de gráficos em R utilizando o pacote ggplot2. Os *chunks* incluem gráficos de barras verticais e horizontais, gráficos de barras empilhadas e gráficos circulares.

O código em R organiza e processa os dados, formatando e visualizando as informações de acordo com o tipo de gráfico adequado. Dependendo do tipo de pergunta (escolha múltipla, resposta única, ou escalas), são utilizados métodos específicos para construir os gráficos corretamente.

Para garantir a qualidade de exportação, o DPI (resolução) dos gráficos é ajustado e são aplicadas várias opções de formatação, como *labels* de percentagens e cores diferenciadas (Figura 5).

Grafico horizontal

C6. Qual a dimensão da empresa?

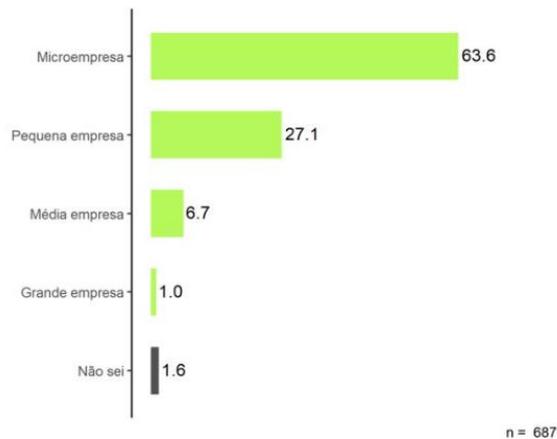


Figura 5- Exemplo de um slide e gráfico

Fonte: Elaboração própria

5. Construir tabelas em R: As tabelas são realizadas da mesma maneira que os gráficos, utilizando já um *chunk* de código que irá criar uma tabela.

Além dos gráficos, o processo envolve a criação de tabelas com dados formatados, utilizando pacotes como *knitr*. Estas tabelas são geradas a partir dos dados importados e ordenadas ou formatadas conforme necessário para uma apresentação clara.

As tabelas podem incluir percentagens, labels e ser formatadas de forma a ficarem visualmente atrativas nas apresentações.

6. Exportar para *PowerPoint*: Por fim, o relatório, incluindo gráficos e tabelas, é exportado para um ficheiro *PowerPoint* utilizando o pacote *Office*. Este pacote permite criar e personalizar slides de forma programática, gerando automaticamente uma apresentação com todos os elementos criados no *RStudio*.

O utilizador pode visualizar a apresentação e fazer ajustes finais, garantindo que o conteúdo está conforme esperado.

Este processo permite a automação de relatórios, tornando mais eficiente a visualização e apresentação de dados, além de reduzir o esforço manual na criação de gráficos, tabelas e slides.

- 5ª Fase – Avaliação

Após a demonstração, validou-se a capacidade do artefacto solucionar o problema através de relatórios preliminares automatizados mais simples. No caso prático, esses relatórios foram propostos com base em dados reais da empresa, e foram criadas e entregues apresentações *PowerPoint* aos clientes de maneira mais rápida. O *feedback* dos colaboradores permitiu concluir que a solução era viável e satisfazia as suas necessidades. Em termos de resultados da aplicação destes relatórios, notou-se uma otimização do tempo dos recursos, reduzindo em cerca de 30% do tempo gasto em comparação com o método tradicional. Este novo método permite que os gráficos não tenham de ser ajustados manualmente e o processo de renderização acontece de forma mais suave. Em contrapartida, no processo tradicional é renderizado a partir de uma macro de *Excel* que cria o *PowerPoint*.

- 6ª Fase - Comunicação

A última fase foi dedicada à comunicação dos resultados do projeto, tanto para os *stakeholders* internos da empresa quanto para a comunidade académica. A solução e os seus benefícios foram apresentados aos gestores da empresa, enfatizando a importância da inovação para melhorar a sua competitividade no mercado. O trabalho foi metodicamente documentado no relatório final de mestrado, e a proposta foi compartilhada para futuras implementações e possíveis melhorias, abrindo caminho para novas pesquisas e inovações na área de BI e visualização de dados.

5. Conclusões

O Trabalho Final de Mestrado desenvolvido teve como objetivo propor uma solução para a produção de relatórios semi-automatizados, com base na visualização de dados e nas práticas de BI. A partir da pesquisa realizada, foi possível propor uma solução eficiente e prática para a criação de relatórios automáticos utilizando a linguagem R, o que resultou numa apresentação preliminar de dados mais ágil e acessível para os clientes da empresa.

A metodologia de *Design Science* facilitou a construção e avaliação de artefactos inovadores, assegurando que a solução proposta atendesse não só às necessidades atuais, como também abrisse caminho para novas abordagens dentro da empresa.

Em virtude dos resultados apresentados é possível concluir que a ferramenta de automação no contexto de visualização de dados é valiosa. No caso concreto, esta ferramenta mostrou-se valiosa pois permitiu uma resposta rápida e eficiente às necessidades dos clientes.

Por fim, a relevância deste projeto centra-se na sua contribuição tanto para o desenvolvimento tecnológico da empresa quanto para a literatura académica, ao interligar conceitos de BI, *Design Science* e Visualização de dados.

6. Limitações e Trabalhos Futuros

Em fase de término deste projeto, pretende-se delinear certas limitações verificadas ao longo deste estudo, assim como, sugerir algumas recomendações para a organização e promoção de futuras investigações.

Este projeto apresenta aspetos que exigem maior desenvolvimento e investigação adicional, além de necessitar de melhorias e futuras implementações para superar certas limitações pré-existentes. Sendo um projeto de cariz profissional, existem também algumas limitações e restrições organizacionais. Entre as principais dificuldades encontradas nesta temática estão: orçamento da empresa para aquisição de licenças de determinadas tecnologias, colaboradores sem experiência em certas tecnologias, práticas internas da empresa e acordos estabelecidos com clientes.

Como recomendações relevantes para a promoção de futuras pesquisas, são propostas algumas ações pertinentes que podem dar continuidade ao trabalho aqui iniciado:

1. Aplicar e testar a ferramenta noutro contexto empresarial, de modo a compreender a sua eficácia e eficiência.
2. Testar diferentes *softwares* com funcionalidades semelhantes no contexto deste projeto, a fim de identificar as melhores opções para cada caso prático específico.
3. Criar soluções com a ajuda de Inteligência Artificial e bases de dados cada vez maiores e compará-las com as soluções criadas manualmente.

Referências Bibliográficas

- Ackoff, R. L. (1989). From data to wisdom. *Journal of Applied Systems Analysis*, 16, 3–9.
- Allaire, J. (2012). RStudio: integrated development environment for R. *Use R Boston, MA*, 770(394), 165-171.
- Aparicio, J.T., Aparicio, M., Costa, C.J. (2023). Design Science in Information Systems and Computing. In: Anwar, S., Ullah, A., Rocha, Á., Sousa, M.J. (eds) Proceedings of International Conference on Information Technology and Applications. Lecture Notes in Networks and Systems, vol 614. Springer, Singapore. https://doi.org/10.1007/978-981-19-9331-2_35
- Aparicio, J. T., Karatsoli, A., & Costa, C. J. (2024). Network visualization techniques for story charting. *arXiv preprint arXiv:2406.14734*. <https://doi.org/10.48550/arXiv.2406.14734>
- Aparicio, M., & Costa, C. J. (2015). Data visualization. *Communication design quarterly review*, 3(1), 7-11. <https://doi.org/10.1145/2721882.2721883>
- Araujo, H, Costa, C. & Aparicio, M (2017) "Modelo de competitive intelligence (CI) competitive intelligence (CI) model." 2017 12th Iberian Conference on Information Systems and Technologies (CISTI). IEEE, 2017. <https://doi.org/10.23919/CISTI.2017.7975787>.
- Arnott, D., & Pervan, G. (2005). A critical analysis of decision support systems research. *Journal of information technology*, 20(2), 67-87.
- Bach B. et al., "Dashboard Design Patterns," in IEEE Transactions on Visualization and Computer Graphics, vol. 29, no. 1, pp. 342-352, Jan. 2023, <https://doi.org/10.1109/TVCG.2022.3209448>.
- Baskerville, R. (2008). What *Design Science* is not. *European Journal of Information Systems*, 17(5), 441–443. <https://doi.org/10.1057/ejis.2008.45>

- Baskerville, R. L. (1999). Investigating information systems with action research. *Communications of the association for information systems*, 2(1), 19.
- Baumer, B. S., Kaplan, D. T., & Horton, N. J. (2017). *Modern data science with R*. Chapman and Hall/CRC.
- Brynjolfsson, E., Hitt, L. M., & Kim, H. H. (2011). Strength in numbers: How does data-driven decisionmaking affect firm performance?. *Available at SSRN 1819486*.
- Cairo, A. (2016). *The truthful art: Data, charts, and maps for communication*. New Riders.
- Card, S. K., Mackinlay, J., & Shneiderman, B. (Eds.). (1999). *Readings in information visualization: using vision to think*. Morgan Kaufmann.
- Chee, T., Chan, L. K., Chuah, M. H., Tan, C. S., Wong, S. F., & Yeoh, W. (2009, December). Business intelligence systems: state-of-the-art review and contemporary applications. In *Symposium on progress in information & communication technology* (Vol. 2, No. 4, pp. 16-30).
- Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS quarterly*, 1165-1188.
- Chen, M., Floridi, L., Greenfield, P., & Keim, D. A. (2009). Data, information, and knowledge in visualization. *IEEE Computer Graphics and Applications*, 29(1), 12-19. <https://doi.org/10.1109/MCG.2009.6>
- Cleveland, W. S., & McGill, R. (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American statistical association*, 79(387), 531-554.
- Costa, C. (2007). *Desenvolvimento para web*. ITML press/Lusocredito.
- Costa, C. J., & Aparcio, M. (2005). Visualization of balanced scorecard on PDAs. In *Proceedings of the 23rd annual international conference on Design of communication: documenting & designing for pervasive information* (pp. 103-107). <https://doi.org/10.1145/1085313.1085339>

- Costa, C. J., & Aparicio, M. (2019). Supporting the decision on dashboard design charts. In *Proceedings of 254th The IIER International Conference*.
- Davenport, Thomas. (2010). Business Intelligence and Organizational Decisions. *IJBIR*. 1. 1-12. 10.4018/jbir.2010071701.
- Davison, R., Martinsons, M. G., & Kock, N. (2004). Principles of canonical action research. *Information systems journal*, 14(1), 65-86.
- Delen, D., Moscato, G., & Toma, I. L. (2018, January). The impact of real-time business intelligence and advanced analytics on the behaviour of business decision makers. In 2018 International Conference on Information Management and Processing (ICIMP) (pp. 49-53). IEEE. <https://doi.org/10.1109/ICIMP1.2018.8325840>.
- Elbashir, M. Z., Collier, P. A., & Davern, M. J. (2008). Measuring the effects of business intelligence systems: The relationship between business process and organizational performance. *International journal of accounting information systems*, 9(3), 135-153.
- Few, S. (2006). *Information dashboard design: The effective visual communication of data*. O'Reilly Media, Inc..
- Few, S. (2013). *Information dashboard design : displaying data for at-a-glance monitoring* (Second edition). Analytics Press.
- Fox, J., & Weisberg, S. (2018). *An R companion to applied regression*. Sage publications.
- Galliers, R. D., & Land, F. F. (1987). Choosing appropriate information systems research methodologies. *Communications of the ACM*, 30(11), 901-902.
- Heer, J., & Bostock, M. (2010, April). Crowdsourcing graphical perception: using mechanical turk to assess visualization design. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 203-212).
- Järvinen, P. (2007). Action research is similar to design science. *Quality & quantity*, 41, 37-54.

- Jarvinen, P. H. (2000). Research questions guiding selection of an appropriate research method. In Mahrer (Ed.), *Proceedings of ECIS 2000, 3-5 July (1st ed., pp. 124-131)*.
- LeCun, Y. (2022). A path towards autonomous machine intelligence version 0.9. 2, 2022-06-27. *Open Review*, 62(1), 1-62.
- Luftman, J., & Zadeh, H. S. (2011). Key information technology and management issues 2010–11: an international study. *Journal of Information Technology*, 26(3), 193-204.
- Luhn, H. P. (1958). A business intelligence system. *IBM Journal of research and development*, 2(4), 314-319.
- Malik, S. (2005). *Enterprise dashboards: design and best practices for IT*. John Wiley & Sons.
- Marques, M. A. & Costa, C. J. (2019). Social CRM Analytics Challenges. *Journal of Information Systems Engineering & Management*, 4(4), em0105. <https://doi.org/10.29333/jisem/6349>
- May, T. (2009). *The new know: innovation powered by analytics*. John Wiley & Sons.
- Midway, S. R. (2020). Principles of effective data visualization. *Patterns*, 1(9). 100141December 11, <https://doi.org/10.1016/j.patter.2020.100141>
- Naidoo, J., & Campbell, K. (2016). Extended abstract: Best practices for data visualization. IEEE International Professional Communication Conference, 2016-November. Austin, TX, USA, 2016, pp. 1-3, <https://doi.org/10.1109/IPCC.2016.7740509>.
- Ogan M. Yigitbasioglu, Oana Velcu, A review of dashboards in performance management: Implications for design and research, *International Journal of Accounting Information Systems*, Volume 13, Issue 1, 2012, Pages 41-59, ISSN 1467-0895, <https://doi.org/10.1016/j.accinf.2011.08.002>.
- Osborne, J. W. (2014). *Best practices in logistic regression*. Sage Publications.

- Parente, C & Costa, C (2022) "Comparing Process Mining Tools and Algorithms," 2022 17th Iberian Conference on Information Systems and Technologies (CISTI), Madrid, Spain, pp. 1-7, <https://doi.org/10.23919/CISTI54924.2022.9820570>.
- Pawan Vora, Chapter 4 - Application Main Page, Editor(s): Pawan Vora, In Interactive Technologies, Web Application Design Patterns, Morgan Kaufmann, 2009, Pages 79-109, ISBN 9780123742650, <https://doi.org/10.1016/B978-0-12-374265-0.00004-9>.
- Peffer, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of management information systems*, 24(3), 45-77.
- Peng, R. D. (2016). *R programming for data science* (pp. 86-181). Victoria, BC, Canada: Leanpub.
- Popovi, Aleš & Coelho, P. & Jakli, Jurij. (2009). The impact of business intelligence system maturity on information quality. *Information Research*. 14.
- Reynoso, M & M. J. Div3n, M. (2020) Applying Data Visualization Guideline on Forest Fires in Argentina," 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, pp. 617-622, <https://doi.org/10.1109/Confluence47617.2020.9058174>.
- Scholz, Patrick & Schieder, Christian & Kurze, Christian & Gluchowski, Peter & B3hringer, Martin. (2010). Benefits and Challenges of Business Intelligence Adoption in Small and Medium-Sized Enterprises. 18th European Conference on Information Systems, ECIS 2010.
- Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., ... & Hassabis, D. (2016). Mastering the game of Go with deep neural networks and tree search. *nature*, 529(7587), 484-489.
- Simon, H. A. (1996). *The sciences of the artificial* MIT Press.
- Tavares, M & Costa, C. (2007). Knowledge management process in the local government. In Proceedings of the 25th annual ACM international conference on Design of

- communication (SIGDOC '07). Association for Computing Machinery, New York, NY, USA, 182–188. <https://doi.org/10.1145/1297144.1297183>
- Tufte, E. R. (1985). The visual display of quantitative information. *The Journal for Healthcare Quality (JHQ)*, 7(3), 15.
- Vázquez-Ingelmo, Andrea & García-Peñalvo, Francisco & Therón, Roberto. (2019). Information Dashboards and Tailoring Capabilities - A Systematic Literature Review. *IEEE Access*. 7. 109673-109688. <https://doi.org/10.1109/ACCESS.2019.2933472>.
- Watson, H. J., & Wixom, B. H. (2007). The current state of business intelligence. *Computer*, 40(9), 96-99.
- Wickham, H. (2010). A Layered Grammar of Graphics. *Journal of Computational and Graphical Statistics*, 19(1), 3–28. <https://doi.org/10.1198/jcgs.2009.07098>
- Wickham, H. (2016). Getting Started with ggplot2. In: ggplot2. Use R!. Springer, Cham. https://doi.org/10.1007/978-3-319-24277-4_2
- Wickham, H., & Wickham, H. (2016). *Data analysis* (pp. 189-201). Springer International Publishing.
- Wickham, H., Çetinkaya-Rundel, M., & Grolemund, G. (2023). *R for data science*. "O'Reilly Media, Inc."
- Wu, D. T., Vennemeyer, S., Brown, K., Revalee, J., Murdock, P., Salomone, S., ... & Hanke, S. P. (2019). Usability testing of an interactive dashboard for surgical quality improvement in a large congenital heart center. *Applied Clinical Informatics*, 10(05), 859-869. <https://doi.org/10.1055/s-0039-1698466>
- Xie, Y., Allaire, J. J., & Grolemund, G. (2018). *R markdown: The definitive guide*. Chapman and Hall/CRC.
- Xie, Y., Allaire, J.J., & Grolemund, G. (2018). *R Markdown: The Definitive Guide* (1st ed.). Chapman and Hall/CRC. <https://doi.org/10.1201/9781138359444>

Zingde, S., & Shroff, N. (2020). The Role of Dashboards in Business Decision Making and Performance Management. *A Road Map to Future Business; Institute of Management, Nirma University: Ahmedabad, India, 227.*

ANEXOS

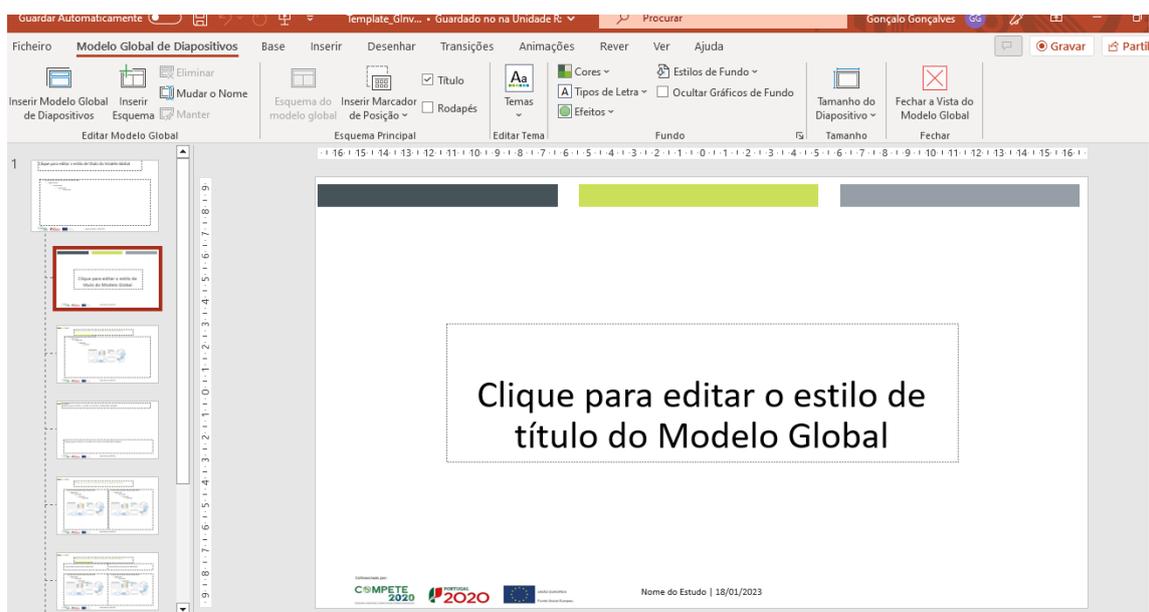
Anexo I - Manual do projeto no RStudio

Criar template ppt para o design do relatório

Algumas orientações para criar um template de PowerPoint (PPT) para o design do relatório:

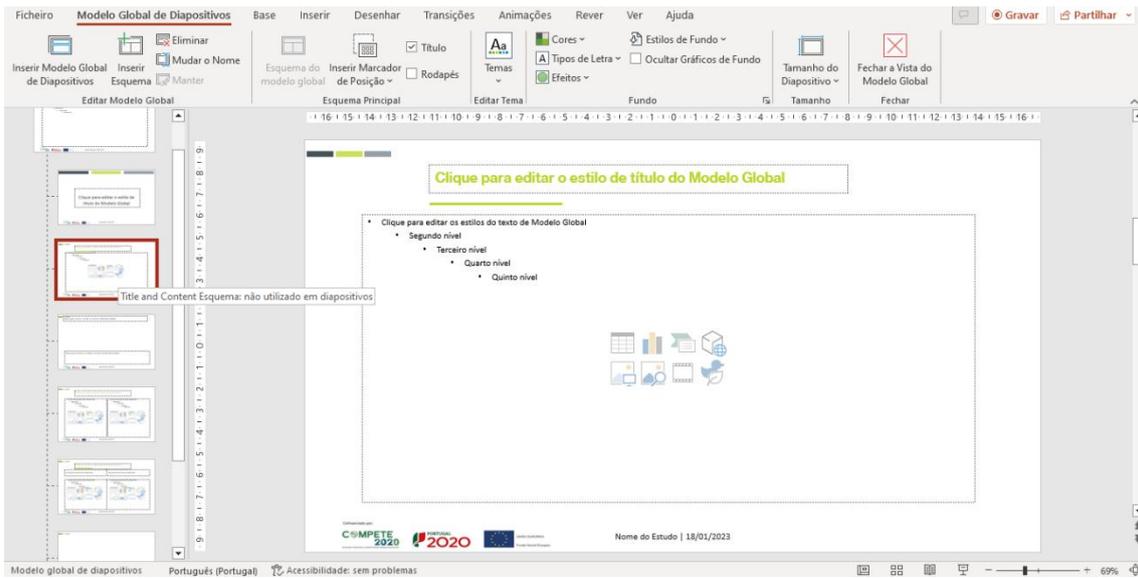
Estrutura do Template PPT

O primeiro passo é criar um ppt e alterar no modelo global o seu design, este será aplicado no ppt do relatório.

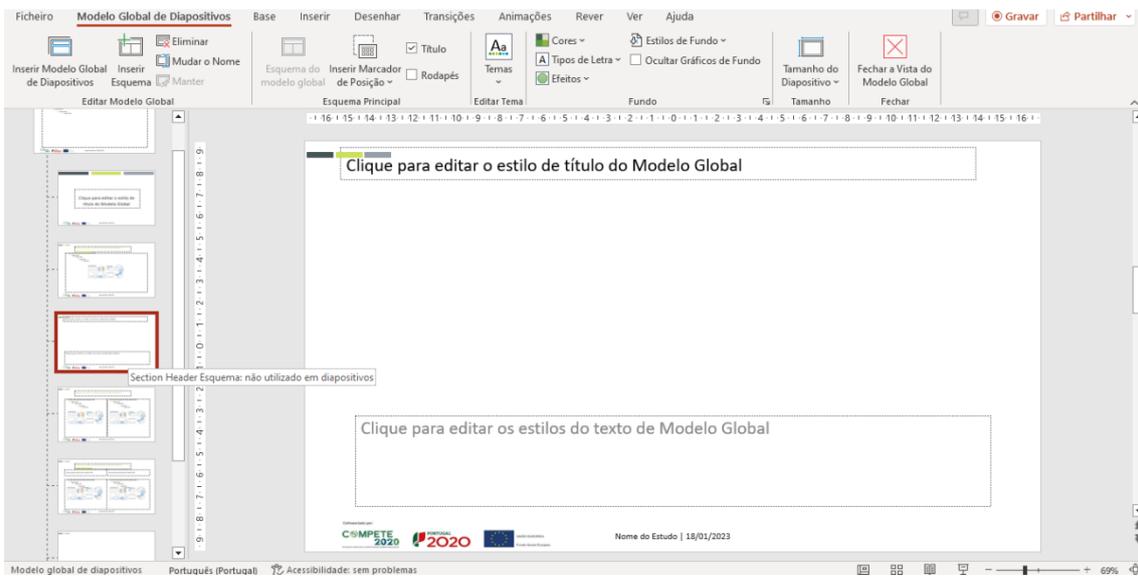


Fonte: Elaboração própria

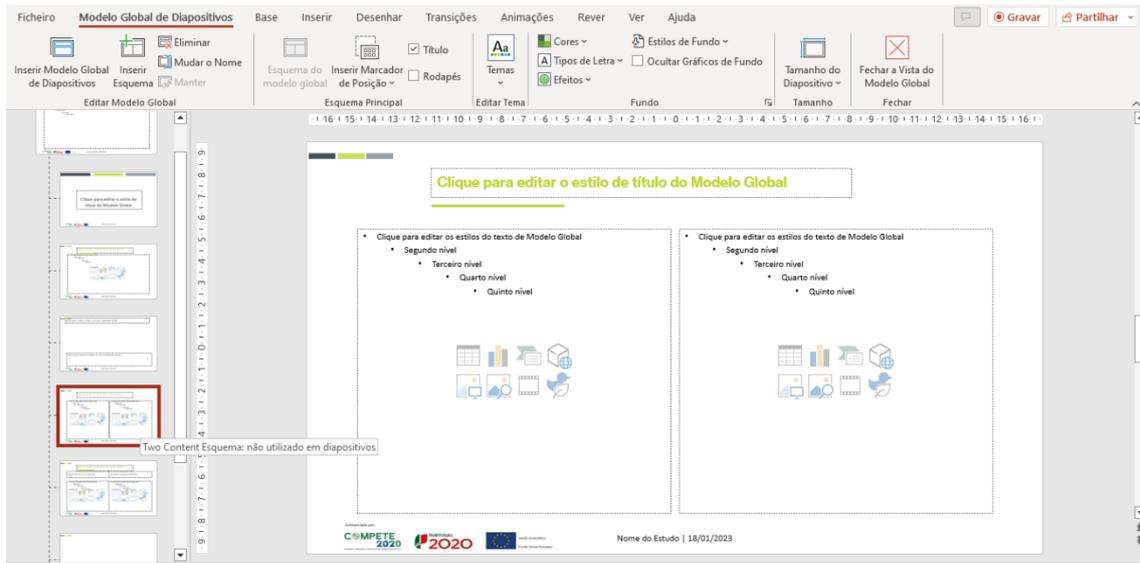
Deve conter slide global para: Slide de título, Slide de título e conteúdo, Slide de cabeçalho, Slide de dois conteúdos, Slide de comparação e Slide de conteúdo e texto. Estes slides depois são aplicados de forma automática pelo *RStudio*, consoante o que for mais adequado para os elementos inseridos.



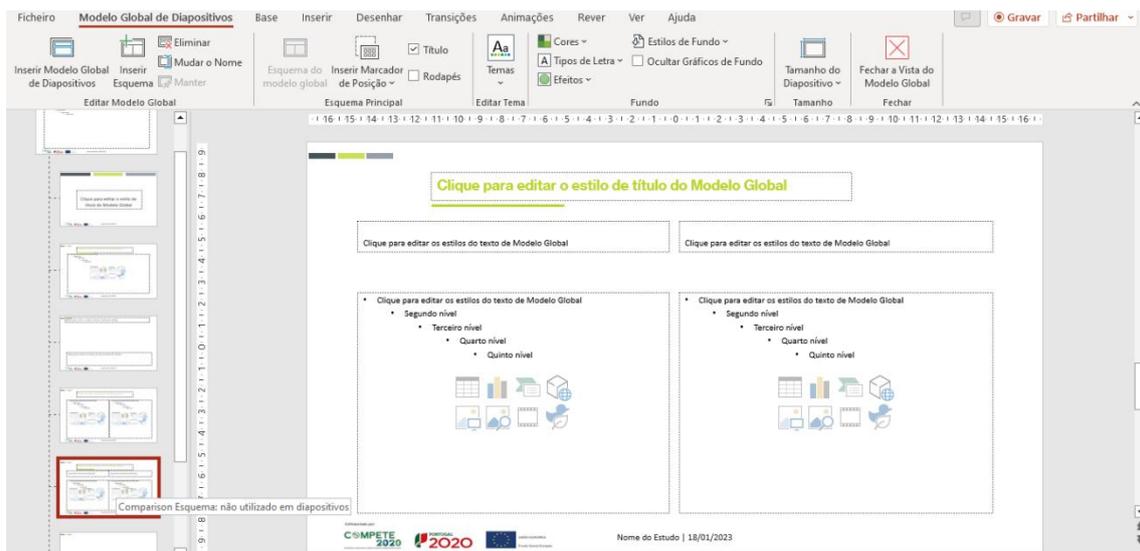
Fonte: Elaboração própria



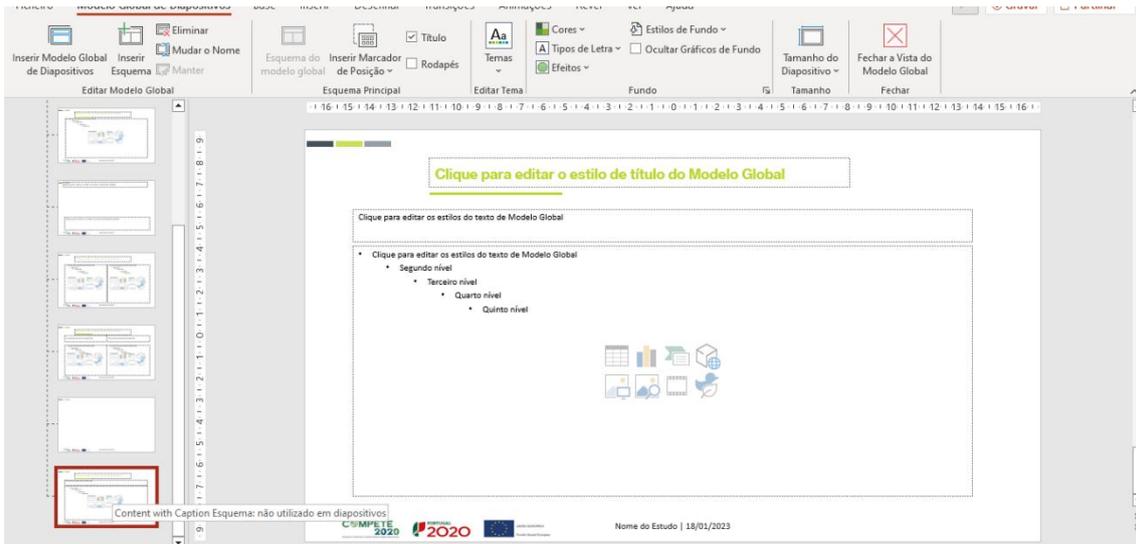
Fonte: Elaboração própria



Fonte: *Elabora3o pr3pria*



Fonte: *Elabora3o pr3pria*



Fonte: Elaboração própria

Daqui em diante tudo o que estiver realçado a amarelo é texto de programação utilizado no *RStudio*.

Para passar a informação deste ppt para o *RStudio* é necessário o bloco YAML (espécie de cabeçalho do ficheiro) deve ser escrito da seguinte forma:

```
---
```

```
title: "Grow Investment"
```

```
author: "Qmetrics"
```

```
date: "2023-02-13"
```

```
output:
```

```
powerpoint_presentation:
```

```
reference_doc: Template_GInvest.pptx
```

```
---
```

RStudio

Packages

Apenas na primeira utilização.

Colocar os packages na consola. Exemplo:

```

1  ---
2  title: "Grow Investment"
3  author: "Qmetrics"
4  date: "2023-02-13"
5  output:
6    powerpoint_presentation:
7      reference_doc: Template_Ginvest.pptx
8  ---
9
10 <!-- todos os ficheiros têm de incluir este chunk -->
11 [r setup, include=FALSE]
12 knitr::opts_chunk$set(echo = FALSE, cache=FALSE, warning = FALSE, comment = NA)
13
14 library(dplyr)
15 library(readr)
16 library(ggplot2)
17 library(RColorBrewer)
18 library(forcats)
19 library(knitr)
20 opts_knit$set(global.par = TRUE)
21

```

```

R 4.2.2 R:/Grow Investment/Ginvest/
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

Error in file.exists(pythonPath) :
  file name conversion problem -- name too long?
[workspace loaded from R:/Grow Investment/Ginvest/.RData]
> install.packages(ggplot2)

```

Fonte: Elaboração própria

Conectar os ficheiros de apuramentos ao RStudio

```

10 <!-- todos os ficheiros têm de incluir este chunk -->
11 [r setup, include=FALSE]
12 knitr::opts_chunk$set(echo = FALSE, cache=FALSE, warning = FALSE, comment = NA)
13
14 library(dplyr)
15 library(readr)
16 library(ggplot2)
17 library(RColorBrewer)
18 library(forcats)
19 library(knitr)
20 opts_knit$set(global.par = TRUE)
21
22
23 <!-- ler ficheiros de freqs, meds, titulos e labels -->
24 [r ]
25
26 freq <- read_excel("mg_freq_growd.xlsx")
27 med <- read_excel("mg_med_growd.xlsx")
28 Enunciado <- read_excel("Labels_Enunciado.xlsx")
29 freq$lookup <- freq$vlookup
30 freq$lookup <- gsub("[A]+$", "", freq$lookup)
31 freq$esc <- freq$vlookup
32 freq$esc <- gsub("[A]+$", "", freq$esc)
33 med$lookup <- med$vlookup
34 med$lookup <- gsub("[A]+$", "", med$lookup)

```

Fonte: Elaboração própria

Selecionar o ícone de tabela na aba de “Environment” e selecionar a opção “From Excel”. De seguida, selecionar o excel de apuramentos pretendido.

Além disto, é necessário criar um ficheiro excel de labels. Para que seja possível atribuir as labels às vars do gráfico. Este ficheiro de labels é similar ao ficheiro de filtros, exemplo “R:\Grow Investment\GInvest\Labels”, que posteriormente é utilizado no ficheiro cópia de freqs e meds.

Quando todos os ficheiros necessários estiverem criados e adicionados ao *RStudio* é necessário colocar o seguinte chunk de código:

```
```{r }
```

```
freq <- read_excel("mg_freq_Grow1.xlsx")
```

← cria a variavel freq que guarda todos os valores do excel de frequências

```
med <- read_excel("mg_med_Grow1.xlsx")
```

← cria a variavel med que guarda todos os valores do excel de médias

```
Enunciado <- read_excel("Labels_Enunciado.xlsx")
```

← cria a variável Enunciado que guarda

```
freq$lookup <- freq$vlookup
```

← cria uma nova coluna, no ficheiro de frequências, com o nome lookup, que é uma cópia da coluna vlookup

```
freq$lookup <- gsub("[^|]+$", "", freq$lookup)
```

← na coluna lookup substitui o simbolo “|” por nada ou “”

```
freq$esc <- freq$vlookup
```

← cria uma nova coluna, no ficheiro de frequências, com o nome esc com os mesmos valores que a vlookup

```
freq$esc <- gsub("[^_]+$", "", freq$esc)
```

← na coluna esc substituí o símbolo “\_” por nada ou “”

```
med$lookup <- med$vlookup
```

← cria uma nova coluna, no ficheiro de médias, com o nome lookup, que é uma cópia da coluna vlookup

```
med$lookup <- gsub("[^|]+$", "", med$lookup)
```

← na coluna lookup substitui o simbolo “|” por nada ou “”

`med$esc <- med$vlookup` ← cria uma nova coluna, no ficheiro de médias, com o nome esc com os mesmos valores que a vlookup

`med$esc <- gsub("_[^_]+$", "", med$esc)` ← na coluna esc substituí o símbolo "\_" por nada ou ""

'''

Este código é que irá ler e obter os valores dos ficheiros de apuramentos. Ainda, este chunk também cria colunas, dataset de freqs e meds, duas colunas a lookup e a esc. A coluna lookup é uma simplificação da coluna vlookup, onde não aparece a última componente, a mod. De outra forma, a coluna esc retira nas vars de escala o indicativo do agrupamento de valores da escla (esc\_Q35\_1 passa para esc\_Q35, isto é feito para todos os agrupamentos de valores 2, 3 e 4) desta forma é possível utilizar a var esc\_Q35 para passar os valores para o gráfico.

## Construção de gráficos

Algumas orientações sobre como construir gráficos no *RStudio* usando a linguagem R, com foco em alguns pacotes populares, como **ggplot2** e funções base.

Algumas noções básicas para a construção de gráficos em R, especialmente usando o pacote **ggplot2**, que é amplamente utilizado para visualização de dados:

- Noções básicas

Questões de escolha múltipla-Utilizar o código `Q41<-`

`freq[which(freq$esc=="i|1|ii|1|Q41" & freq$mod=="1")]` no caso de questões de resposta múltipla, é necessário utilizar a coluna esc para obter todos os indicadores da Q41, é também, sempre necessário indicar que a modalidade é =1. Este é apenas um exemplo da Q41.

Questões de resposta única-Utilizar o código `C3<-`

`freq[(freq$lookup=="i|1|ii|1|C3_aux|"),c("label","PERCENT")]`, deste modo todas as modalidades são abrangidas e guardas na var C3.

Questões de escalas-Utilizar o código `Q42<-`

`freq[freq$esc=="i|1|ii|1|esc_Q42",c("label","mod","PERCENT","var")]` é utilizado do mesmo modo que as questões de escolha múltipla com a exceção da seleção de modalidade.

- Construção de tabelas

Aqui estão algumas noções básicas sobre a construção de tabelas em R, incluindo como usar pacotes populares como **knitr** e **gt** para formatar e apresentar dados de forma clara e atraente.

`## Tabela` ← o duplo # indica que um novo slide é iniciado, o que estiver escrito é o título do slide

```{r}`

`cat(as.character(Enunciado[Enunciado$Pergunta=="CAE",2]))` ← coloca o título da questão

`C3<-freq[(freq$lookup=="i|1|ii|1|C3_aux|"),c("label","PERCENT")]` ← cria a var C3 que guarda os valores das colunas "label" e "PERCENT" que correspondem ao lookup da var desejada

`C3$PERCENT <- format(C3$PERCENT, decimal.mark = ".", digits =1)` ← altera o número de casa decimais para 1, esta parte do código nem sempre funciona como deve ser, por isso, é preciso testar e ir alterando o número de dígitos para 2 ou 3.

`C3$PERCENT=as.numeric(C3$PERCENT)` ← passar os valores a numérico

`C3<-C3[order(C3$PERCENT,decreasing = TRUE),]` ← ordenar os valores por ordem decrescente

```
colnames(C3)<-c("CAE","%") ← alterar o nome das colunas
```

```
knitr::kable(C3, align = c("c","c")) ← reproduzir a tabela e centrar os valores de ambas a colunas
```

```
'''
```

Escolha do Tipo de Gráfico: A escolha entre os diferentes tipos de gráficos de barras deve depender dos dados que você tem e da mensagem que deseja comunicar.

- Gráfico de barras Horizontal e Vertical

Aqui estão exemplos de como criar gráficos de barras horizontais e verticais em R usando o pacote **ggplot2**.

Gráfico de Barras Vertical

Um gráfico de barras vertical é o padrão em ggplot2 e pode ser criado facilmente com a função `geom_bar()`. Aqui está um exemplo:

Os gráficos de barras são uma forma popular de visualizar dados categóricos e podem ser usados de várias maneiras. Aqui estão alguns tipos comuns de gráficos de barras:

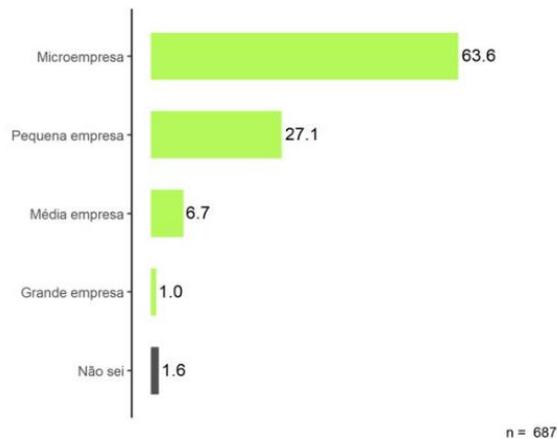
Um dos tipos mais comuns, onde as barras são desenhadas verticalmente. É útil para comparar diferentes categorias.

Gráfico de Barras Horizontal

As barras são desenhadas horizontalmente. Útil quando os rótulos das categorias são longos ou quando se tem muitas categorias.

Grafico horizontal

C6. Qual a dimensão da empresa?



Fonte: Elaboração própria

O mesmo processo para gráficos de barras verticais, no entanto não utiliza o último passo que inverte a ordem dos eixos.

```
## Grafico horizontal ← título do slide
```

```
``{r, dpi=300, echo=FALSE, message=FALSE} ← colocar o dpi=300 é muito importante para o gráfico ser exportado com boa qualidade
```

```
cat(as.character(Enunciado[Enunciado$Pergunta=="Concelho",2])) ← título da questão
```

```
Concelho<-
```

```
freq[freq$lookup=="i | 1 | ii | 1 | Concelho | ",c("mod","PERCENT","label")] ← cria a var Concelho para guardar os valores da var Concelho do lookup desejado, guarda os valores das colunas mod, PERCENT e label
```

```
ggplot(Concelho, aes(x=reorder(mod, +PERCENT), y=PERCENT, fill=ifelse(mod=="99", "Diferente", "Normal")))+ ← seleciona as variáveis do gráfico, ordena de forma
```

decrecente em relação ao valor de frequência e coloca neste caso o mod “99” de cor diferente das restantes

`geom_bar(stat = "identity", width =.6, show.legend = FALSE)+` ← cria as barras do gráfico

`scale_fill_manual(values = c("Normal"="#b4f959", "Diferente"="#515151"))+` ← pinta a barras com a cor desejada e aplica a condição utilizada anteriormente

`geom_text(aes(label=sprintf(PERCENT,fmt="%1.1f")), vjust=0.4, hjust=-0.2)+` ← adicionada as labels de valores ao gráfico e formata, através do “fmt=%1.1f” o valor para ser apresentado com uma casa decimal

`theme_classic()+` ← tema do gráfico, pode ser sempre este, é o básico

`labs(caption = (paste("n = ",
freq[which(freq$lookup=="i | 1 | ii | 1 | Antiguidade_da_empresa | Nãodisponível"),c(11)])` ← coloca uma pequena legenda a indicar a base, é necessário fazer referência do valor pretendido nas questões com filtros, as restantes pode fazer referência à mesma questão que abranja todos os questionados

`x=NULL,` ← remove a labels do eixo x

`y=NULL)+` ← remove a labels do eixo y

`guides(x="none")+` ← remove o eixo x neste caso

`ylim(0,80)+` ← alterar os limites de valores do gráfico, podem ser outros valores, este é o mais adequado para este estudo

`coord_flip()` ← inverter a orientação do gráfico, utilizado para gráficos horizontais

'''

- Gráfico circular

Os gráficos circulares, ou gráficos de pizza, são uma maneira comum de visualizar a proporção de diferentes categorias em um conjunto de dados. Aqui está um guia sobre como criar gráficos circulares em R, especialmente usando o pacote **ggplot2**.

```{r, dpi=300}` ← resolução do gráfico

`cat(as.character(Enunciado[Enunciado$Pergunta=="C5",2]))` ← título da questão

`hsize <- 1.5` ← tamanho do buraco do gráfico, vai também determinar a espessura da secção de valores

`C5<-freq[freq$lookup=="i|1|ii|1|C5|",c("label","PERCENT")]` ← variável que guarda os valores desejados

`C5 <- C5 %>%`

`mutate(x = hsize)` ← adiciona novas variáveis que são funções de variáveis existentes, neste caso o tamanho da parte central

`ggplot(C5, aes(x = hsize, y = PERCENT, fill = label)) +` ← variáveis a serem utilizadas no gráfico

`geom_col()` + ← é necessário adicionar para formar o gráfico

`geom_text(aes(label = sprintf(PERCENT, fmt = "%1.1f")))`, ← adiciona as labels dos valores e coloca a sua formatação com uma casa decimal

`position = position_stack(vjust = 0.5)) +` ← ainda em relação à label coloca a posição dos valores centrada no eixo vertical, neste caso (0,5)

`coord_polar(theta = "y") +` ← é utilizado para mudar de retângulos empilhados para um anel

`xlim(c(0.2, hsize + 0.5)) +` ← limites, importante para colocar o buraco, consoante maior seja o valor de hsize menor é a parte colorida do gráfico

```
scale_fill_manual(values = c("Nacional"="#b4f959", "Estrangeira"="green4", "Não sei"="#515151"), name=NULL)+
```

```
labs(caption = (paste("n = ",
freq[which(freq$lookup=="i|1|ii|1|Antiguidade_da_empresa|Nãodisponível"),c(11)])
),
```

```
x=NULL,
```

```
y=NULL)+
```

```
theme_void()
```

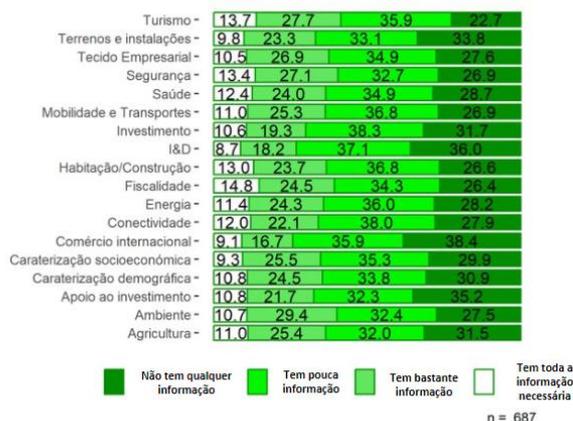
```
''''
```

- Gráfico de barras empilhadas horizontal e vertical

As barras empilhadas mostram diferentes categorias dentro de uma única barra, permitindo ver a composição de cada grupo.

### Barras empilhadas

Q46. Em que medida considera ter acesso à informação suficiente, relativa a cada um dos temas elencados, para poder investir na região Oeste? Utilize a escala de 1 a 10, em que 1 significa "não tenho qualquer informação" e 10 significa "tenho toda a informação necessária".



Fonte: Elaboração própria

A única diferença entre a orientação deste gráfico é a função “coord\_flip()”

```
Gráfico de barras empilhadas ← novo slide
```

```
` `{r echo=FALSE, dpi=300} ← resolução do gráfico
```

```
cat(as.character(Enunciado[Enunciado$Pergunta=="Q42",2])) ← título da questão
```

```
Q42<-freq[freq$esc=="i|1|ii|1|esc_Q42",c("label","mod","PERCENT","var")] ← cria variável para guardar os valores desejados, nestes casos de vars de escala é necessário utilizar a coluna esc “freq$esc” para selecionar a var corretamente
```

```
ggplot(Q42, aes(fill = factor(mod, levels = c("1","2","3","4")), x=label, y=PERCENT))+ ← define as variáveis a utilizar no gráfico
```

```
geom_bar(stat = "identity", position = "stack", width =.5, color="#058d02")+ ← aqui fica definido que as barras devem ficar empilhadas, a sua espessura e a cor da linha de contorno
```

```
scale_fill_manual(values = c("4"="white","3"="#63e660","2"="#05f500","1"="#058d02"), labels=c("Não tem qualquer\n informação","Tem pouca\n informação","Tem bastante\n informação","Tem toda a informação\n necessária"))+ ← define a escala de cores
```

```
scale_y_continuous(labels = scales::percent)+ ← altera a escala do eixo y
```

```
geom_text(aes(label=sprintf(PERCENT,fmt="%1.1f")), ← coloca um label de valores no gráfico com apenas uma casa decimal
```

```
position = position_stack(vjust=0.5))+ ← formatação da label de valores e a sua posição
```

```
theme_classic()+ ← tema de gráfico clássico/normal
```

```
labs(x=NULL, ← remove a labels do eixo x
```

```
y=NULL, ← remove a labels do eixo y
```

```
caption = (paste("n = ", freq[which(freq$lookup=="i|1|ii|1|Q1|1")],c(11))))+←
```

adiciona uma pequena legenda com a base

```
guides(x="none")+← remove o eixo x
```

```
theme(legend.title = element_blank(), legend.position = "bottom", legend.key.size =
unit(0.5,"cm"),axis.line = element_blank()+← altera o tema do gráfico, coloca a
```

legenda abaixo do gráfico e altera o seu tamanho

```
coord_flip()←troca a ordem dos eixos
```

```
....
```

- Colocar conteúdo lado a lado

### Tabela

Q19. Para os bens exportados pela empresa, indique por favor qual a respetiva % do volume de negócios.



Q19.1. Quais os bens exportados pela empresa?

Var	valor médio (%)	n
Eletrodomésticos	100.0	1
Alojamento/ Hotelaria	100.0	4
Cerâmica	100.0	3
Serviços (consultoria, publicidade, turismo, etc)	78.7	14
Plásticos e Caixilharia	70.2	6
Moldes na área tecnológica	65.0	2
Outro	62.2	33
Produtos hortícolas	59.0	5
Componentes de automóveis	55.2	4
Transformação de fruta (desidratados, sumos e polpas)	55.0	2
Pera Rocha	50.0	1
Vinho	35.0	1
Metalomecânica (alfaias agrícolas de grande porte)	15.0	3
Pedra	10.0	2

Fonte: Elaboração própria

Para colocar conteúdo lado a lado em um relatório ou apresentação em R, você pode usar algumas abordagens dependendo do contexto (por exemplo, R Markdown, Shiny ou gráficos). Aqui estão algumas maneiras comuns de fazer isso:

Para colocar conteúdos lado a lado basta utilizar as “fences”, é mesmo necessário respeitar a quantidade de dois-pontos.

```
::::{.columns}
```

```
::{.column}
```

Colocar aqui o código do gráfico desejado

```
...
```

```
::{.column}
```

Colocar aqui o código do gráfico desejado

```
...
```

```
.....
```

## Criar o PPT

Criar uma apresentação em PowerPoint (PPT) usando *RStudio* é muito simples através da ferramenta RMarkdown.

Para criar o ppt basta clicar no botão “*knit*”.