



UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE
INSTITUTO METRÓPOLE DIGITAL
PROGRAMA DE RESIDÊNCIA EM TECNOLOGIA DA INFORMAÇÃO

Título

Autor

Natal-RN, Brasil
2020

Autor

Título

Trabalho de Conclusão de Curso apresentado ao Programa de Residência em Tecnologia da Informação do Instituto MetrÓpole Digital da Universidade Federal do Rio Grande do Norte como requisito parcial para a obtenção do título de Especialista em Tecnologia da Informação. Área de Concentração:

Orientador: Nome do orientador
Coorientador: Nome do coorientador

Natal-RN, Brasil
2020

Autor

Título

Trabalho de Conclusão de Curso apresentado ao Programa de Residência em Tecnologia da Informação do Instituto Metr pole Digital da Universidade Federal do Rio Grande do Norte como requisito parcial para a obten  o do t tulo de Especialista em Tecnologia da Informa  o.  rea de Concentra  o:

Trabalho aprovado. Natal-RN, Brasil, 22 de setembro de 2020:

Nome do orientador

Orientador

Nome do coorientador

Coorientador

Professor

Examinador

Professor

Examinador

Natal-RN, Brasil
2020

Dedicatória.

Agradecimientos

Agradecimientos

Resumo

A Tecnologia da Informação (TI) vem se tornando cada vez mais importante em empresas e órgãos, as aplicações vão desde a infraestrutura que busca conectar os diferentes setores, mantendo a segurança da rede, até a automação de processos simples. Além disso, nesse espectro de aplicações da TI podemos incluir o melhoramento da gestão usando a computação, atualmente a quantidade de dados e variáveis disponíveis para o gestores é muito grande, e é extremamente difícil de se gerenciar essa massa de dados, para isso existem várias ferramentas que têm por objetivo auxiliar na visualização e futura tomada de decisão dos administradores.

Uma área da TI que tem crescido bastante é a *Business Intelligence* (BI), que reúne uma série de conceitos que podem ser aplicados em empresas, de qualquer tamanho e área, com o objetivo de dar suporte à tomada de decisão, trazendo dados e gerando informação, que serve de base para as escolhas das estratégias de um determinado negócio.

Esses conceitos devem ser usados para desenvolver um painel para o Centro de Inteligência, que monitora os processos que entram na JFRN, a fim de evitar a multiplicação de demandas repetitivas.

Palavras-chave: BI. Dados. Informação.

Abstract

O resumo em língua estrangeira (em inglês Abstract, em espanhol Resumen, em francês Résumé) é uma versão do resumo escrito na língua vernácula para idioma de divulgação internacional. Ele deve apresentar as mesmas características do anterior (incluindo as mesmas palavras, isto é, seu conteúdo não deve diferir do resumo anterior), bem como ser seguido das palavras representativas do conteúdo do trabalho, isto é, palavras-chave e/ou descritores, na língua estrangeira. Embora a presente especificação considere o inglês como língua estrangeira (o mais comum), não fica impedida a adoção de outras línguas (a exemplo de espanhol ou francês) para redação do resumo em língua estrangeira.

Keywords: Keywords separated by dots.

Résumé

O resumo em língua estrangeira (em inglês Abstract, em espanhol Resumen, em francês Résumé) é uma versão do resumo escrito na língua vernácula para idioma de divulgação internacional. Ele deve apresentar as mesmas características do anterior (incluindo as mesmas palavras, isto é, seu conteúdo não deve diferir do resumo anterior), bem como ser seguido das palavras representativas do conteúdo do trabalho, isto é, palavras-chave e/ou descritores, na língua estrangeira. Embora a presente especificação considere o inglês como língua estrangeira (o mais comum), não fica impedida a adoção de outras línguas (a exemplo de espanhol ou francês) para redação do resumo em língua estrangeira.

Mots-clés : Mots-clés séparés par points.

Lista de ilustrações

Figura 1 – Resumo de um sistema BI	17
Figura 2 – Preço do Power BI Premium	18
Figura 3 – Linguagens de programação mais usadas em 2020	21
Figura 4 – Estrutura básica do painel	21

Lista de quadros

Lista de tabelas

Lista de abreviaturas e siglas

IMD Instituto Metrópole Digital

UFRN Universidade Federal do Rio Grande do Norte

Lista de símbolos

Γ	Letra grega Gama
Λ	Lambda
ζ	Letra grega minúscula zeta

Sumário

	Introdução	15
1	INTRODUÇÃO AO BI	16
1.1	Estrutura básica do BI	16
1.2	Visualização de dados no BI	17
1.3	Ferramentas de BI	18
1.3.1	Power BI	18
1.3.2	Qlikview	19
1.3.3	Tableau	19
2	A CONSTRUÇÃO DO PAINEL	20
2.1	Python	20
2.2	Estrutura básica do painel	21
2.2.1	Estrutura dos dados	22
2.2.2	Análise de anomalias	22
3	DETALHES DO PAINEL	24
3.1	Dash	24
4	CONTEÚDOS ESPECÍFICOS DO MODELO DE TRABALHO ACADÊMICO	25
5	CONCLUSÃO	26
	REFERÊNCIAS	27

Introdução

A Tecnologia da Informação (TI) vem se tornando cada vez mais importante em empresas e órgãos, as aplicações vão desde a infraestrutura que busca conectar os diferentes setores, mantendo a segurança da rede, até a automação de processos simples. Além disso, nesse espectro de aplicações da TI podemos incluir o melhoramento da gestão usando a computação, atualmente a quantidade de dados e variáveis disponíveis para o gestores é muito grande, e é extremamente difícil de se gerenciar essa massa de dados, para isso existem várias ferramentas que têm por objetivo auxiliar na visualização e futura tomada de decisão dos administradores.

Uma área da TI que tem crescido bastante é a *Business Intelligence* (BI), que reúne uma série de conceitos que podem ser aplicados em empresas, de qualquer tamanho e área, com o objetivo de dar suporte à tomada de decisão, trazendo dados e gerando informação, que serve de base para as escolhas das estratégias de um determinado negócio.

Esses conceitos devem ser usados para desenvolver um painel para o Centro de Inteligência, que monitora os processos que entram na JFRN, a fim de evitar a multiplicação de demandas repetitivas.

1 Introdução ao BI

De forma geral, o termo *Business Intelligence* (BI) ainda não é bem definido na literatura, mas alguns dos principais estudiosos da área, como Solomon Negash, apresentam o BI como um sistema que combina dados operacionais com ferramentas analíticas para apresentar informações para os gerentes do negócio. O objetivo disso é melhorar a qualidade das decisões do processo, tornando-as melhor embasadas, então os métodos do BI podem ser usados para melhor entender o estado da empresa, organização ou órgão onde estiver sendo aplicado, resultando em melhores decisões. Apesar disso, ainda existe bastante empresa que emprega o BI num viés muito mais ligado à gestão, sem exigir muito conhecimento da parte de tecnologia da informação, e usando conceitos sobre mercado, marketing, administração da produção etc.

Nesse trabalho serão usados os conceitos que se aproximam mais ao que Negash apresenta como BI, e dentro do que ele define como sendo importante na Inteligência de Negócios podemos citar algumas áreas do conhecimento como:

- *Data Warehouse*
- Visualização
- Mineração de dados
- *Online Analytic Processing* (OLAP)
- Gerenciamento do conhecimento
- Probabilidade
- Estatística
- Análises preditivas
- Detecção de anomalias

1.1 Estrutura básica do BI

Como foi apresentado anteriormente, o BI serve para auxiliar nas tomadas de decisão, e isso é alcançado usando os dados da empresa ou órgão em que estiver sendo empregado. Os dados são armazenados em *Data Warehouses*, que são armazéns de dados, em tradução livre, esses armazéns guardam dados históricos, então a partir deles é possível analisar o desenvolvimento de variáveis importantes e o comportamento delas de acordo

ao passar dos anos e tentar estabelecer padrões, isso por si só já poderia ser usado para prever possíveis mudanças em estratégia de negócios.

Após o armazém, os dados devem ser coletados e limpos, a limpeza corresponde a remoção de linhas erradas, que contenham dados ou errados ou faltosos, que podem atrapalhar na análise e apresentação ao gestor. Em seguida, os dados são apresentados à pessoa do negócio, que a partir das suas análises irá tomar alguma decisão que afeta a estratégia da empresa. Com alguma sorte essa decisão será pautada em matemática, estatística e informações úteis, sendo bem embasada em dados e números.

De forma resumida, o BI usa o Data Warehouse para guardar os dados, usa um conjunto de ferramentas e técnicas para limpar e extrair os dados, essa técnica também é conhecida como *Extraction, Transform, Load* (ETL), e ,finalmente, apresenta gráficos que mostram o comportamento de variáveis de interesse da empresa para o gestor, que a partir disso escolhe alguma estratégia para os rumos da empresa/órgão/setor que gerencia.

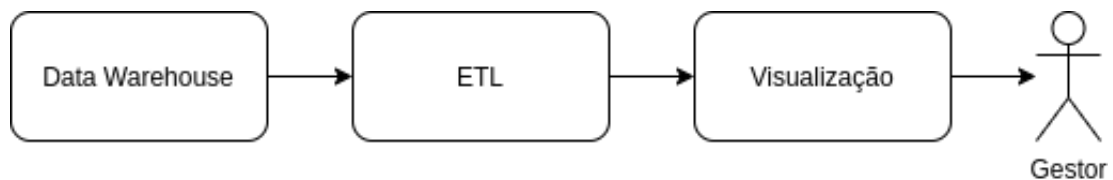


Figura 1 – Resumo de um sistema BI

Isso tudo que foi tratado acima corresponde às etapas do processamento de dados estruturados, ou seja, dados que podem ser organizados e categorizados em linhas e colunas, e que, muitas vezes, possuem relações entre si. O processo para "manusear" dados não estruturados é um pouco diferente porque eles não são tão bem organizados, e alguns passos precisam ser inseridos nesse caminho para que eles sejam apresentados e tratados da melhor forma, evitando distorções.

1.2 Visualização de dados no BI

Todas as etapas do processo são importantes, mas o gestor só enxerga o último estágio: a visualização, e as decisões serão tomadas a partir das conclusões tiradas da visualização, portanto, a visualização precisa de um cuidado especial, o visual é importante porque ele pode levar a conclusões erradas, então é necessário ter atenção com as cores, os eixos entre outros detalhes. Mais adiante serão abordados mais características da visualização de dados.

1.3 Ferramentas de BI

Existem vários programas feitos para se desenvolver painéis BI, entre os mais usados podemos citar o PowerBI, Pentaho, Tableau entre outros. Na JFRN o software usado é o Qlikview.

Ultimamente as empresas que fornecem os softwares de BI têm notado que uma abordagem de serviços é mais rentável, então o cliente paga por um serviço e não por um produto. As mensalidades variam de acordo com o serviço que o cliente quiser, ficando mais caro à medida que mais funções são adicionadas ao pacote que será usado. Abaixo serão listados alguns dos principais serviços de BI do mercado.

1.3.1 Power BI

O Power BI é um serviço da Microsoft, ele tem por objetivo fornecer visualização de dados, customizável pelo próprio usuário, e também funções de BI que sejam simples de serem feitas.

Em 2020 o preço dos serviços partem de \$9,99 para o Power BI Pro e \$4,995 para o Power BI Premium, é importante ressaltar que em ambos os casos esses valores são mensais, e o pacote que atende às necessidades da JFRN é o mais caro, que pode aumentar o preço de acordo com as funcionalidades extra que serão requeridas para atender às demandas.

Você precisa de Advanced Analytics, suporte de Big Data, relatórios na infraestrutura local e na nuvem?

Adicionar o Power BI Premium

- Enterprise BI, análise de Big Data, relatórios na nuvem e na infraestrutura local
- Controles avançados de administração e implantação
- Computação de nuvem dedicada e recursos de armazenamento
- Permite que o usuário consuma o conteúdo do Power BI

Power BI Premium

\$4.995

Preço mensal por computação de nuvem dedicada e recursos de armazenamento com assinatura anual

[Solicitar consulta >](#)

Figura 2 – Preço do Power BI Premium

1.3.2 Qlikview

O Qlikview é um software desenvolvido pela Qlik, que tem focado os esforços no desenvolvimento do Qlik Sense, que começou como um produto paralelo e agora é o principal. Então, o suporte ao Qlikview está entrando no fim e nem é mais possível consultar os preços dele no site da Qlik.

Apesar disso, o Qlikview é uma boa ferramenta, oferece um bom conjunto de funções para analisar e apresentar os dados. O Qlik Sense é a evolução dele, com muito mais recursos.

1.3.3 Tableau

Tableau é uma solução que fornece funcionalidades clique-e-arraste para ajustar dados, criar painéis e visualizações. Existem três produtos dentro dele, o Tableau Desktop, Server e o Online. O preço entre eles varia de \$12,00 até \$70,00 por mês, por licença, podendo ajustar anualmente. É uma das ferramentas mais simples e completas do mercado, exigindo pouco conhecimento do analista por ser clique-e-arraste, por outro lado isso torna um pouco mais difícil o desenvolvimento de análises mais detalhadas.

De modo geral, as principais ferramentas disponíveis são muito similares entre si, todas são competentes, fornecem ótimas funcionalidades, possuem um bom suporte e custam um valor razoavelmente alto, principalmente com o câmbio desfavorecendo o Real. Porém, as ferramentas proprietárias não são as únicas, e para o desenvolvimento do Painel do Centro de Inteligência o Python foi escolhido.

2 A construção do painel

Antes de avançar para a parte técnica, é importante explicar o que é o Centro de Inteligência da JFRN e como um painel poderia ajudar na tarefa deles.

De acordo com o site do Centro de Inteligência, ele existe para otimizar o trabalho jurisdicional a partir de demandas repetitivas. Esse tipo de demanda deve ser comunicado às autoridades para que a multiplicação desses casos diminuam, e não haja uma sobrecarga. Portanto, o painel será usado para auxiliar na análise dessas demandas, tentando acompanhar a evolução e desenvolvimento delas, agindo para que elas não sobrecarreguem a JFRN.

2.1 Python

Python é uma linguagem de programação de alto nível e de aplicações gerais, foi criada por Guido van Rossum e lançada em 1991. Suporta vários paradigmas, incluindo orientação a objeto e programação funcional. Essas características a tornaram muito popular entre programadores, porque é possível construir muita coisa usando poucas linhas de código, e esses projetos vão desde um timer simples, até grandes análises de dados e sistemas com inteligência artificial.

No final do capítulo 1 foram detalhadas algumas ferramentas BI, entre elas Tableau e Power BI, essas ferramentas já vem prontas com todas as funcionalidades que o usuário vai precisar, já lê os dados automaticamente, identifica campos e cria gráficos de forma muito rápida, porém os custos de implementação são altos, as licenças também são caras e é difícil encontrar profissionais que mexam nessas ferramentas tão específicas. Já no Python alguns desses problemas são resolvidos, o Python é uma linguagem de programação, portanto não tem nada pronto, tudo precisa ser construído, desde o leitor de dados, até o construtor de gráficos, o trabalho para se desenvolver um painel usando uma linguagem de programação é difícil, mas uma vez desenvolvido, é muito fácil de se manter, e o custo é zero porque não há licenças que limitem a quantidade de usuários que podem acessar o que foi feito. Além disso, é relativamente fácil encontrar desenvolvedores de Python no Rio Grande do Norte (e no Brasil), porque é uma linguagem muito usada no mundo todo, isso pode ser visto no ranking elaborado pelo *Institute of Electrical and Electronics Engineers* (IEEE).

Portanto, o uso do Python para elaborar um painel pode ser difícil no início, porque muita coisa precisa ser construída do zero, mas uma vez que isso esteja desenvolvido é, relativamente, fácil de se manter porque não há custos envolvidos com o uso da ferramenta em si. Apesar disso, um dos problemas seria o suporte, nas opções pagas é possível requisitar suporte caso alguma coisa não funcione, já no Python se acontecer algum imprevisto, os

próprios desenvolvedores é que devem solucionar.

Rank	Language	Type	Score
1	Python ▼	🌐 🖥️ ⚙️	100.0
2	Java ▼	🌐 📱 🖥️	95.3
3	C ▼	📱 🖥️ ⚙️	94.6
4	C++ ▼	📱 🖥️ ⚙️	87.0
5	JavaScript ▼	🌐	79.5
6	R ▼	🖥️	78.6
7	Arduino ▼	⚙️	73.2
8	Go ▼	🌐 🖥️	73.1
9	Swift ▼	📱 🖥️	70.5
10	Matlab ▼	🖥️	68.4

Figura 3 – Linguagens de programação mais usadas em 2020

Outro ponto positivo de se usar Python é a replicabilidade, é possível criar painéis que atendam as diferentes Varas da JFRN, mostrando os dados que os gestores preferirem e acharem que são mais relevantes para suas realidades.

2.2 Estrutura básica do painel

Com os dados da JFRN em mãos e a ferramenta escolhida, passou-se a pesquisar quais seriam as bibliotecas usadas, portanto, a estrutura ficou da seguinte forma:



Figura 4 – Estrutura básica do painel

O painel que se propôs ao Centro de Inteligência não tem a estrutura clássica com diferentes tabelas (fato e dimensão), com as quais se geram as visualizações, no lugar disso, existe um arquivo .csv que contém os dados que serão usados, esses dados .csv fazem parte uma extração que veio do PJe, e a partir desse arquivo o painel vai criar subconjuntos de acordo com o ano e órgão julgador escolhidos pelo usuário. Portanto, esse painel se aproxima mais de um visualizador de dados do que de um painel BI.

Nele é possível selecionar duas variáveis: o Órgão Julgador e o Ano. A partir dessas escolhas o sistema vai fatiar os dados recebidos e mostrará algumas análises. Essas análises são mostradas em forma de tabelas condicionais, que mudam as cores das células de acordo com a frequência de aparição dos Assuntos.

2.2.1 Estrutura dos dados

Nessa primeira versão do painel os dados virão de um arquivo .csv gerado a partir do Qlikview, que é o software de BI padrão da JF. Esse arquivo carrega várias colunas, entre elas podemos citar número do processo, status, classe judicial, documento da parte, data do trânsito em julgado. Porém, para fazer a análise dos dados serão usadas as seguintes colunas:

- Órgão Julgador - os órgão julgadores são as Varas da JFRN que ficam espalhadas pelo Estado, o usuário precisa selecionar um desses órgãos para visualizar os dados.
- Data Primeira Distribuição - essa é a data em que o processo chega na JFRN, mesmo que caia numa Vara que não seja da competência dele essa data é importante para analisar que Vara o recebeu e quando ele chegou na JFRN.
- Assunto - é o tema do processo, existem diferentes categorias em que um processo pode ser categorizado, e a partir desse campo é possível contar quantos processos de cada tipo deram entrada na JFRN.
- Assunto Código - diferentes assuntos possuem diferentes códigos, e a contagem dos processos se dá usando esse campo, que agrupa os códigos que são iguais e conta o total para saber quantos deram entrada na JFRN.

A partir da escolha do Ano e do Órgão Julgador, o painel irá fazer as análises e seleções relevantes, populando a tabela e mostrando ao usuário quais são os processos mais frequentes de cada mês, no Ano e Vara escolhidos.

2.2.2 Análise de anomalias

A detecção de anomalias é, basicamente, uma técnica (ou um conjunto de técnicas) que servem para identificar comportamentos que fogem do que é esperado. Um dos desafios

do trabalho foi encontrar uma forma de se detectar os Assuntos que possuísem alta frequência de entrada na JFRN, porque, teoricamente, cada Ano e cada Vara possuem diferentes distribuições de probabilidade, e um modelo de detecção de anomalia que se encaixa bem em um determinado período, pode não se encaixar em outros. São 15 órgãos julgadores diferentes, e os anos que podem ser consultados são de 2014 até 2020, então são 90 distribuições diferentes. Portanto, usamos uma abordagem simples mas eficaz.

Primeiro, há uma análise da média (\bar{x}) de Assuntos que entraram na Vara, essa análise leva em conta o ano selecionado e o ano anterior, após isso, o desvio padrão (σ) é calculado e novas variáveis são geradas.

As variáveis são:

- $anom_2$ que é a média mais duas vezes o desvio padrão, definida como:

$$anom_2 = \bar{x} + 2 * \sigma$$

- $anom_1$ que é a média mais uma vez o desvio padrão, definida como:

$$anom_1 = \bar{x} + \sigma$$

- $media_{assuntos}$ que é a média simples dos assuntos, a cada dois anos:

$$media_{assuntos} = \sum_{ano}^{ano-1} \frac{assuntos}{total_{meses}}$$

Com essas variáveis encontradas, a distribuição das cores segue as regras a seguir, em que *total* significa a quantidade total de Assuntos de determinada categoria:

$$F_{cores} = \begin{cases} Vermelho & \text{se } total \geq anom_2 \\ Amarelo & \text{se } total \geq anom_1 \text{ e } total < anom_2 \\ Verde & \text{se } total \geq media_{assuntos} \text{ e } total < anom_1 \end{cases} \quad (2.1)$$

3 Detalhes do painel

Como foi dito anteriormente, o painel usa diferentes bibliotecas na sua construção, são elas o Dash, Plotly Express e Pandas.

3.1 Dash

A biblioteca Dash

4 Conteúdos específicos do modelo de trabalho acadêmico

5 Conclusão

Sed consequat tellus et tortor. Ut tempor laoreet quam. Nullam id wisi a libero tristique semper. Nullam nisl massa, rutrum ut, egestas semper, mollis id, leo. Nulla ac massa eu risus blandit mattis. Mauris ut nunc. In hac habitasse platea dictumst. Aliquam eget tortor. Quisque dapibus pede in erat. Nunc enim. In dui nulla, commodo at, consectetur nec, malesuada nec, elit. Aliquam ornare tellus eu urna. Sed nec metus. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas.

Phasellus id magna. Duis malesuada interdum arcu. Integer metus. Morbi pulvinar pellentesque mi. Suspendisse sed est eu magna molestie egestas. Quisque mi lorem, pulvinar eget, egestas quis, luctus at, ante. Proin auctor vehicula purus. Fusce ac nisl aliquam ante hendrerit pellentesque. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Morbi wisi. Etiam arcu mauris, facilisis sed, eleifend non, nonummy ut, pede. Cras ut lacus tempor metus mollis placerat. Vivamus eu tortor vel metus interdum malesuada.

Sed eleifend, eros sit amet faucibus elementum, urna sapien consectetur mauris, quis egestas leo justo non risus. Morbi non felis ac libero vulputate fringilla. Mauris libero eros, lacinia non, sodales quis, dapibus porttitor, pede. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Morbi dapibus mauris condimentum nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Etiam sit amet erat. Nulla varius. Etiam tincidunt dui vitae turpis. Donec leo. Morbi vulputate convallis est. Integer aliquet. Pellentesque aliquet sodales urna.

Referências