

Fluxo de Grandes Dados em Ambientes Corporativos

Felipe Manikowski, Gustavo C. de P. Santos e Luiz A. C.
Junior

Departamento Acadêmico de Informática – Universidade Tecnológica Federal
do Paraná (UTFPR)

{felipem,gustavosantos,ljunior}@alunos.utfpr.edu.br

22 jun. 2016

Descrição

Apresentação
Final

Felipe
Manikowski,
Gustavo C. de
P. Santos e
Luiz A. C.
Junior

- *Business Intelligence*: Transformar dados em informações relevantes.
- *Big Data* (Grandes Dados): Dados em elevado volume e quantidade, estruturados ou não, presentes nas mais diversas fontes.
- Desafio: Como obter vantagens competitivas e de mercado a partir destes dados?

Objetivos

Apresentação
Final

Felipe
Manikowski,
Gustavo C. de
P. Santos e
Luiz A. C.
Junior

Objetivo Geral

Abranger, através de um ambiente corporativo simulado, os processos de um fluxo de grandes dados neste, desde a coleta até a exibição utilizando ferramentas tecnológicas atuais.

Objetivos Específicos

- Explicar de forma geral como é o algoritmo de extração de dados do *Google Analytics*;
- Tratar dados alegóricos e armazená-los de forma simulada em um *Data Warehouse*, fazendo uso do *software MySQL*;
- Manipular os dados de forma teórica e exibi-los, para que facilitem a geração de estratégias e tomadas de decisão por parte dos responsáveis, a partir do *software Tableau*.

Referencial Teórico

Apresentação
Final

Felipe
Manikowski,
Gustavo C. de
P. Santos e
Luiz A. C.
Junior

- *Data Warehouse*: Após os dados identificados e coletados, como e onde devem ser armazenados de forma que facilite a sua manipulação?
- *Business Intelligence*: Quem são os responsáveis pelas análises, *insights* e exibição destes dados?
- *Big Data*: Onde estão, quais são os tipos, e qual o nível de relevância destes dados para análises e estudos?

Metodologia

Apresentação
Final

Felipe
Manikowski,
Gustavo C. de
P. Santos e
Luiz A. C.
Junior

A metodologia será baseada no estudo de caso do fluxo de grandes dados em um ambiente corporativo simplificado, onde será exemplificado como é possível alcançar dados diários de visualizações, rejeições e sessões nas páginas de produtos de um *e-commerce* e com isto ditar os rumos da empresa.

Ferramentas

- *Google Analytics* - Extração dos dados.
- *Data Warehouse (MySQL)* - Gerenciamento dos dados.
- *Tableau* - Exibição dos dados.

O Cenário

- *E-commerce* - 100.000 produtos divididos em 100 categorias.
- Formato das URLs -
“<http://www.lojasimulada.com.br/nome-produto-id.html>”.
- Demanda - Melhor controle da situação de cada produto a venda na loja.

Desenvolvimento - Extração e Inserção no *Data Warehouse* (DW)

Apresentação
Final

Felipe
Manikowski,
Gustavo C. de
P. Santos e
Luiz A. C.
Junior

- API do *Google Analytics*.
- Algoritmo de extração e inserção executado diariamente.
- Diariamente inseridas entre 0 e 100.000 novas linhas nas tabelas de dados do DW.

Desenvolvimento - DW

Apresentação
Final

Felipe
Manikowski,
Gustavo C. de
P. Santos e
Luiz A. C.
Junior

- Informações provenientes de vários ambientes da empresa.
- *MySQL* - flexibilidade e escalabilidade (ferramenta que suporta os mais variados tipos de necessidades).

Desenvolvimento - *Data Warehouse* - Esquema Global

Apresentação Final

Felipe
Manikowski,
Gustavo C. de
P. Santos e
Luiz A. C.
Junior

```
dw_venda (_id_pedido_, _id_produto_, datahora,  
status, quantidade, frete, valor_absoluto,  
_id_cliente_);
```

```
dw_cliente (_id_cliente_, nome, endereco,  
endereco_entrega, cep, cpf, celular);
```

```
dw_produto (_id_produto_, nome, marca,  
categoria, preco_custo, preco_venda);
```

```
dw_analytics (_id_produto_, data, visualizacoes,  
rejeicoes, sessoes);
```

Desenvolvimento - Análise e Exibição

Apresentação
Final

Felipe
Manikowski,
Gustavo C. de
P. Santos e
Luiz A. C.
Junior

- Taxa de Conversão ($C = (P/V) * 100$).
- Nível de loja, categoria e produto.
- Ferramentas tradicionais deixam a desejar.

Desenvolvimento - Análise e Exibição - *Tableau*

Apresentação
Final

Felipe
Manikowski,
Gustavo C. de
P. Santos e
Luiz A. C.
Junior

- Painéis interativos, manipuláveis no próprio navegador.
- Total integração com o banco de dados.

Desenvolvimento - Análise e Exibição - *Query* (Consulta)

Apresentação
Final

Felipe
Manikowski,
Gustavo C. de
P. Santos e
Luiz A. C.
Junior

```
SELECT
v.id_produto,
a.data,
COUNT(DISTINCT v.id_pedido) as pedidos,
SUM(a.visualizacoes) as visualizacoes,
COUNT(DISTINCT v.id_pedido)/
SUM(a.visualizacoes) as taxa_de_conversao
FROM
dw_analytics a
LEFT JOIN dw_venda v ON v.id_produto = a.id_produto
AND DATE_FORMAT(v.datahora, '%d-%m-%Y') = a.data
WHERE a.data > SUBDATE(NOW(),30)
GROUP BY v.id_produto, a.data;
```

Desenvolvimento - Análise e Exibição

Apresentação
Final

Felipe
Manikowski,
Gustavo C. de
P. Santos e
Luiz A. C.
Junior

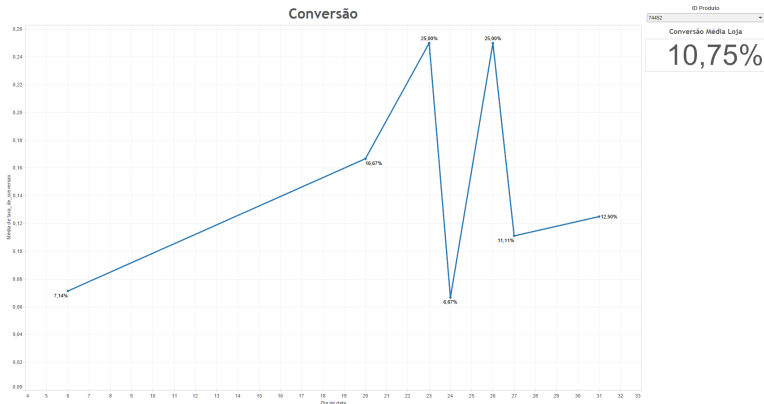


Figura: Gráfico de evolução da conversão de um produto no *Tableau* facilmente comparado com a média geral da loja

Resultados - Matriz de Quadrantes

Apresentação
Final

Felipe
Manikowski,
Gustavo C. de
P. Santos e
Luiz A. C.
Junior

- Matriz - quantidade de pedidos X quantidade de visualizações.
- Categoria A - muitos pedidos X muitas visualizações. (Exemplar)
- Categoria B - grande quantidade de visualizações, porém poucos pedidos. (Problema)
- Categoria C - apesar do baixo número de visualizações, possuem um alto número de pedidos. (Potencial)
- Categoria D - vendem pouco e recebem poucas visualizações. (Ruim)

Resultados - Matriz de Quadrantes

Apresentação
Final

Felipe
Manikowski,
Gustavo C. de
P. Santos e
Luiz A. C.
Junior

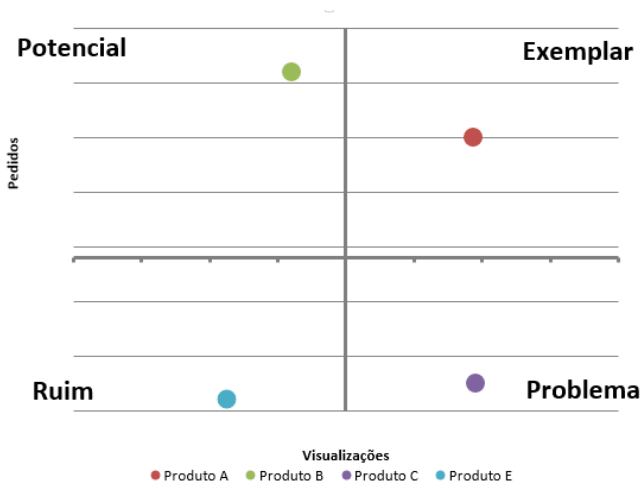


Figura: Matriz dos produtos de acordo com suas métricas relacionadas à média da loja

Resultados - Exibição no *Tableau*

Apresentação
Final

Felipe
Manikowski,
Gustavo C. de
P. Santos e
Luiz A. C.
Junior

The screenshot shows a Tableau worksheet with a table view. The table has three columns: ID Produto, Quadrante, and a third column (likely a measure). The table is filtered by 'Quadrante' (A, B, C, D) and 'ID_categoria' (1, 7, 10, 31, 47, 48, 49, 51, 52, 54, 55, 57, 59, 60, 61, 64, 62, 64, 96, 115, 118, 119, 126, 121, 125, 185, 186, 188, 189, 224, 226, 254, 255, 263, 264, 265, 266). The table is sorted by 'ID Produto' in ascending order.

ID Produto	Quadrante	
181	A	Abc
184	A	Abc
187	A	Abc
189	A	Abc
202	A	Abc
208	A	Abc
213	D	Abc
214	A	Abc
216	A	Abc
217	A	Abc
219	A	Abc
779	A	Abc
788	A	Abc
2027	A	Abc
11871	A	Abc
11872	A	Abc
11873	A	Abc
22471	C	Abc
22472	D	Abc
22473	C	Abc
22474	C	Abc
26282	A	Abc
31568	A	Abc
31571	A	Abc
31572	A	Abc
31574	A	Abc
72156	A	Abc
72158	A	Abc
72159	A	Abc
72160	A	Abc
72161	A	Abc
72162	A	Abc
72163	A	Abc
72164	A	Abc
72165	A	Abc
72169	A	Abc
72175	A	Abc
72463	C	Abc
72465	A	Abc
72467	A	Abc

Figura: Tabela que exibe os produtos de acordo com seus quadrantes e/ou de acordo com sua categoria, dependendo da aplicação dos filtros

Conclusões

Apresentação
Final

Felipe
Manikowski,
Gustavo C. de
P. Santos e
Luiz A. C.
Junior

- Estruturação dos processos de obtenção de dados.
- Junção com outros dados da empresa.
- Situação dos produtos a venda perante ao mercado.
- Outros dados para outras análises já presentes no *Data Warehouse*.

Apresentação
Final

Felipe
Manikowski,
Gustavo C. de
P. Santos e
Luiz A. C.
Junior

Obrigado!