**Exercício 1**

A companhia “We-Commerce” captura e guarda todos os dados gerados pelos visitantes que navegam nos seus sites comerciais. Estes dados são maioritariamente dados característicos da web.

De entre os dados existem os seguintes tipos de dados:

tracking\_record\_id -. Identificador da transação, serve para distinguir todas as transações de forma única.

date\_time – Data em que o utilizador visitou a página Web, ou seja, sempre que um utilizador entrou numa página da companhia foi registada a data.

user\_gui - Identificador exclusivo dos utilizadores que já estejam inscritos. Este identificador apenas varia entre utilizadores já registados nas páginas da companhia.

campaign\_id- Identificador da campanha promocional, sempre que um ou mais produtos estejam em campanha, a mesma é identificada por um id.

producto\_gui – Identificador exclusivo do produto que é visitado pelo visitante numa sessão.

Company - Nome da empresa que fornece o produto.

Link – URL da página da Web que foi visitada

tracking\_id- Identificador que é atribuído por cada registo como identificador único

Meio – Meio pelo qual o visitante navegou

IP – Endereco de protocol (Internet Protocol) de cada visitante

Brower- Navegador que o visitante utilizou.

session\_id- Identificador criado quando é iniciada uma nova sessão, podendo ser diferente para o mesmo utilizador, indicando que o utilizador visitou em diferentes alturas.

cookie\_id – Identificador exclusivo global que serve para identificar o visitante.

**Exercício 2**

As regras de associação têm como objetivo usar os dados das transações para encontrar afinidades entre os produtos que são vendidos em simultâneo, assim como um problema “market-basket analysis”.

Market Basket Analysis descreve o Business intelligence, ou seja, são as informações do desempenho passado da empresa que são usadas para ajudar a prever o desempenho futuro da empresa. Isto pode revelar tendências emergentes das quais a empresa pode lucrar no futuro.

O suporte e a confiança das regras são duas medidas de associação que refletem a utilidade e a certeza das regras de associação descobertas. Normalmente, as regras de associação são consideradas interessantes se satisfazem um limite mínimo de suporte e um limite mínimo de confiança, o que é definido no software Orange.

O suporte mede a frequência com que os produtos da associação, ou seja mede a frequência com que os produtos ocorrem juntos numa transação.

A confiança mede a probabilidade de que o antecedente ocorra quando o consequente ocorre.

Existem algumas vezes onde as medidas de suporte e de confiança são muito altas pelo que podem produzir uma regra não muito útil, por isso utiliza-se uma outra medida o “lift” que indica a força de uma regra. Caso a medida “lift” seja maior do que 1, significa que a regra de associação prevê um bom resultado, caso seja inferior a 1 a regra prevê um resultado não tao significativo.

Uma imagem com mesa

Descrição gerada automaticamente

Na imagem anterior, podemos observar os antecedentes e os consequentes que demostram os produtos que quando ocorrem numa transação (antecedentes), tem maior probabilidade de ocorrer juntos com outros produtos (consequente).

**Exercício 3**

Neste exercício, utilizou-se os passos utilizados no Modelo pratico 7 (MOP7) onde se criaram views que agregaram so dados para se poder analisar o dataset. Ao analisar o dataset criou-se o ficheiro remove\_useless\_produts para remover o lixo.

Os scripts utilizados neste exercício encontram-se na pasta com o nome scripts.

**Exercício 4**

Para gerar um subconjunto de dados mais relevante e focar na nossa análise, neste exercício foi criado um script que mostra todos os eventos gerados pelos visitantes com o número de sessões. Este script é enviado em anexo com o nome de

**Exercício 5**

Neste exercício, foi gerado um dataset como as transações registadas que foram filtradas de acorda com o critério definido. Este script é enviado em anexo com o nome de

**Exercício 6**

Neste exercício, foi necessário fazer primeiramente, um processo de normalização, capaz de descrever os produtos, como por exemplo foi necessário eliminar os espaços, os acentos, e colocar todas as letras em minúsculas. De segundo lugar, foi necessário utilizar o software Orange para gerar um ficheiro Orange “. basket”. O script utilizado neste exercício é enviado em anexo com o nome de

**Exercício 7**

**Which few products should be kept together in a Web page?**

Ao analisar a tabela das regras de associação do Orange podemos verificar que no topo da tabela os produtos nas colunas consequente são os que tem maior probabilidade de serem visitados depois de o utilizador ter visitado os produtos na coluna antecedentes. Logo, os produtos no topo destas colunas devem continuar juntos pois contem mais visualizações, que se podem traduzir em mais lucro para a empresa.

Uma imagem com mesa

Descrição gerada automaticamente

**Which products are rarely visited but often visited together?**

Ao analisar a tabela das regras de associação do Orange do exercício anterior, podemos verificar que os produtos nas colunas consequente não estejam na coluna antecedente são produtos que raramente visitados logo quando estão aos pares são visitados com mais frequência.

**Which products are the most visited?**

Para se saber quais foram os produtos que mais foram visitados, criou-se a seguinte view. Esta view demonstra todas as visitas aos produtos de todas as transações registadas.

Desta forma, podemos observar que os produtos mais procurados foram o calcado, tecnologia e divertimento.

Uma imagem com mesa

Descrição gerada automaticamente

**Exercício 8**