# Perceber o negócio

## Determinar os Objetivos do negócio

### Background

## Avaliação da Situação

## Determinar os objetivos de Data Mining

## Produzir um plano de projeto

# Perceber os dados

## Recolher dados iniciais

### Relatório da recolha inicial de dados

Nesta recolha inicial a equipa de trabalho recolheu 5 ficheiros diferentes sendo eles **Train.csv**, **Features.csv**, **Test.csv**, **Stores.csv**, **sampleSubmission.csv.**

De uma maneira geral podemos dizer que os ficheiros recolhidos contêm dados acerca das 45 lojas Walmart e que será possível fazer o cruzamento entre todos uma vez que todos os ficheiros contêm a coluna “Store”. Destes ficheiros podemos obter informações acerca das vendas semanais de cada departamento das diferentes lojas e cruzar esta informação com dados adicionais.

Da análise dos ficheiros a equipa de desenvolvimento, selecionou quais os ficheiros que seriam necessários de acordo com os objetivos de Data Mining. O resultado da análise está presente na tabela que se segue.

|  |  |  |
| --- | --- | --- |
| Ficheiro | Usado | Descartado |
| Train.csv |  |  |
| Features.csv |  |  |
| Test.csv |  |  |
| Stores.csv |  |  |
| sampleSubmission.csv |  |  |

## Descrever os dados

De seguida é apresentada uma tabela que contém o número de linhas e colunas dos ficheiros que a equipa de desenvolvimento achou relevante usar neste projeto.

|  |  |  |
| --- | --- | --- |
| Ficheiro | Nº de Linhas | Nº de Colunas |
| Train.csv | 421571 | 5 |
| Features.csv | 8191 | 12 |
| Test.csv | 115065 | 4 |
| Stores.csv | 46 | 3 |

Para cada ficheiro foram analisadas as respetivas colunas.

**Train.csv**

|  |  |  |
| --- | --- | --- |
| Coluna | Tipo | Descrição |
| Store | Int | Identificador da loja |
| Dept | Int | Identificador do departamento de determinada loja |
| Date | Date | Data que define o início da contagem do volume vendas. Importante referir que a contagem é feita semana a semana e todas as sextas feiras |
| Weekly\_Sales | Decimal | Volume de vendas no período |
| isHoliday | Boolean | Indica se ocorreu algum feriado no período considerado |

**Features.csv**

|  |  |  |
| --- | --- | --- |
| Coluna | Tipo | Descrição |
| Store | Int | Identificador da loja |
| Date | Date | Data que define o início da contagem do volume vendas. Importante referir que a contagem é feita semana a semana e todas as sextas feiras |
| Temparature | Decimal | Temperatura média da região |
| Fuel\_Price | Decimal | Custo do combustível na região |
| MarkDown1 | Decimal | Dado acerca de descontos promocionais |
| MarkDown2 | Decimal | Dado acerca de descontos promocionais |
| MarkDown3 | Decimal | Dado acerca de descontos promocionais |
| MarkDown4 | Decimal | Dado acerca de descontos promocionais |
| MarkDown5 | Decimal | Dado acerca de descontos promocionais |
| CPI | Decimal | Indíce de preços ao consumidor |
| Unemployement | Decimal | Taxa de desemprego |
| isHoliday | Boolean | Indica se ocorreu algum feriado no período considerado |

**Test.csv**

|  |  |  |
| --- | --- | --- |
| Coluna | Tipo | Descrição |
| Store | Int | Identificador da loja |
| Dept | Int | Identificador do departamento de determinada loja |
| Date | Date | Data que define o início da contagem do volume vendas. Importante referir que a contagem é feita semana a semana e todas as sextas feiras |
| isHoliday | Boolean | Indica se ocorreu algum feriado no período considerado |

**Stores.csv**

|  |  |  |
| --- | --- | --- |
| Coluna | Tipo | Descrição |
| Store | Int | Identificador da loja |
| Type | String | Tipificação da loja |
| Size | Int | Dimensão da loja |

## Explorar os dados

## Verificar a qualidade dos dados

# Preparação dos dados

## Selecionar os dados

## Limpar os dados

## Construir dados

## Integrar os dados

## Formatar dados

# Modelação

## Selecionar técnicas de modelação

## Gerar design de teste

## Construir o modelo

## Avaliar o modelo

# Avaliação

## Avaliar Resultados

## Rever o processo

## Determinar os próximos passos

# Deployment

## Plano de Deploy

## Plano de monitorização e manutenção

## Produzir relatório final

## Rever Projeto