

Projeto: Sistema de Suporte à Produção e Distribuição de Bebidas

OBJECTIVOS:

- Utilizar técnicas de previsão e otimização num problema do mundo real, gerindo e documentando todo o processo.
- Corresponde a 80% da nota final;

> Grupos



- Os grupos serão constituídos de modo ideal por **4 alunos**.
- **Todos membros do grupo** têm de participar no desenvolvimento do projeto.
- Os grupos devem eleger um representante/líder do grupo;
- Durante as aulas, irá haver uma **avaliação do funcionamento do grupo...**

> Autoavaliação do Projeto (A)

- No final da execução do projeto, cada grupo propõe uma Autoavaliação, nota sugerida para o projeto (de 0 a 20 valores): **A**.
- Cada projeto é avaliado pelo docente numa nota final de **P**, que pode ser igual ou diferente de **A**.
- **Exemplo de autoavaliação a inserir no relatório do projeto:** “O grupo considera que o projeto merece 16 valores, porque cumpriu este objetivos ..., mas não consegui realizar os objetivos ... e em termos de execução do projeto ao longo do tempo teve um desempenho”

Notas:

- Realizar todos os objetivos significa ter uma boa avaliação, mas não necessariamente um 18/19/20, podem existir diferentes formas, mais ou menos interessantes, de realizar os objetivos pedidos. E para além do código desenvolvido existe também a análise, o relatório, a apresentação, a execução ao longo das aulas, etc.
- Não é necessário realizar todos objetivos para ter aprovação à Unidade Curricular.

> Autodiferenciação Individual



- Cada grupo propõe uma autodiferenciação individual, com a contribuição individual para cada membro do seu grupo.
- A autodiferenciação individual segue estas regras:
 - o **somatório individual das notas** não pode ser superior a **MxP**, onde M é o número de elementos do grupo e P é a nota final do projeto (avaliado pelo docente).
 - **não se pode reprovar um aluno** por subida de notas (para isso o grupo teria de “despedir” o aluno com antecedência);
- A autodiferenciação individual tem de ser **justificada no relatório**, com descrição do desempenho individual (ver slide sobre relatório).
- Esta autoavaliação pode ser utilizada para **diferenciar as notas, sendo a decisão final dependente do docente** (e.g., via presenças e participação nas aulas práticas, desempenho na entrega do projeto).
- Por exemplo, para uma nota **P** de 14 e 4 alunos A, B, C, D:
 - **A** trabalhou mais do que a média, {**B, C**} trabalharam um valor médio e **D** trabalhou menos: A=15, B=14, C=14, D=13.

Procedimentos éticos



- Não falsificar resultados, nem plagiar projetos de outros grupos ou conteúdos da Internet sem que estes sejam devidamente identificados e referenciados (quem é o autor, onde foi publicado). Esta utilização não pode ser exagerada face ao total trabalho desenvolvido. **Os alunos são responsáveis por garantir que todos os trabalhos enviados sejam originais e as referências consultadas sejam devidamente citadas.**
- Podem utilizar ferramentas de Inteligência Artificial (IA), como ChatGPT, Bard, DeepL e outras. No entanto, devem usar estas ferramentas com cautelas éticas. Copiar e colar informações de ferramentas de IA sem a sua compreensão e atribuição é considerado plágio, quebrando a honestidade acadêmica. Em particular, **o uso das ferramentas de IA deve ser explicitado no relatório e código** (e.g., quais as consultas utilizadas em que ferramentas, quais os resultados obtidos, como textos ou códigos gerados). Estas informações fornecem transparência e responsabilidade pelo uso de IA em trabalhos acadêmicos e ajudam a evitar possíveis problemas de plágio.

> Enunciado



- Uma empresa distribui diariamente diversas bebidas por lojas na cidade de Goiânia (Brasil).
- Tentando adaptar-se ao contexto da Industria 4.0, a empresa começou a registrar as suas vendas diárias (quantidades totais) para duas das suas bebidas mais vendidas (cerveja):
 - STELLA - “STELLA ARTOIS LONG NECK 275ML”;
 - BUD - “BUDWEISER OW 330ML”;
- Registrou-se ainda a **DATA**, dados entre: 2/jan/2019 e 31/dez/2020; e o dia da semana (**DIA_SEMANA**), onde 1 = domingo, 2 = segunda, ...
- Via um *Webservice*, foi possível aceder ainda a dados meteorológicos relativos ao estado do tempo da cidade:
 - **PRECIPITACAO** – total de precipitação, em milímetros (mm) de água líquida.
 - **MAX_TEMP** – temperatura máxima diária, em graus Celcius.

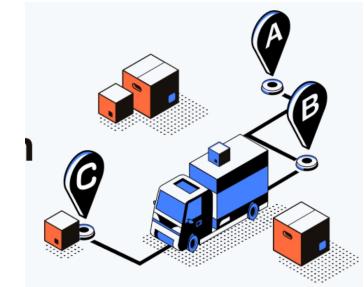
> Dados

- Os dados estão disponíveis via um ficheiro xlsx (MS Excell): **bebidas.xlsx**
- Na ferramenta R, os dados podem ser lidos com estes comandos:

```
# install.packages("openxlsx")
library(openxlsx)
db=read.xlsx(xlsxFile="bebidas.xlsx",sheet=1,skipEmptyRows=FALSE,colNames=TRUE,detectDates=TRUE)
# notas:
class(db[,1]) # "Date" - R data type that handles dates
summary(db)
```

DATA	DIA_SEMANA	PRECIPITACAO	TEMP_MAX	STELLA	BUD
Min. :2019-01-02	Min. :1.000	Min. : 0.000	Min. :21.40	Min. : 0.0	Min. : 0.0
1st Qu.:2019-07-03	1st Qu.:2.000	1st Qu.: 0.000	1st Qu.:29.20	1st Qu.: 13.0	1st Qu.: 22.0
Median :2020-01-01	Median :4.000	Median : 0.000	Median :30.90	Median : 47.0	Median : 58.0
Mean :2020-01-01	Mean :4.001	Mean : 3.669	Mean :31.13	Mean : 105.4	Mean : 101.4
3rd Qu.:2020-07-01	3rd Qu.:6.000	3rd Qu.: 1.550	3rd Qu.:32.80	3rd Qu.: 128.8	3rd Qu.: 125.8
Max. :2020-12-31	Max. :7.000	Max. :66.000	Max. :40.50	Max. :1335.0	Max. :1280.0

> Distribuição



Para despachar o produto a ser vendido, é necessário dois tipos de recursos:

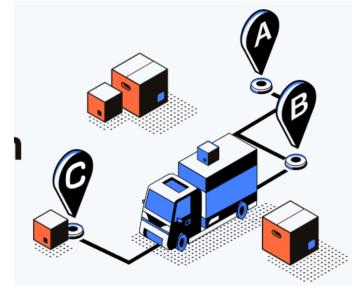
- Recurso de **Armazém** para empacotamento e carga: funcionários e necessidades de empacotamento (e.g., caixas), carregamento ou descarregamento para o veículo, a contratar para o armazém principal;
- Recursos de **Distribuição** (para transporte e entrega). Existem diversos veículos (e.g., mota, carrinha), sendo que cada veículo leva um motorista.

> Recurso de Armazém



- Cada recurso de Armazém custa (funcionário, fitas, caixas para cervejas):
 - 10 R\$ por dia normal de trabalho.
 - 15 R\$ ao fim de semana (Sábado ou Domingo).
- Pode ser alocado numa base diária, N_A recursos por dia (recorrendo a uma bolsa de trabalhadores temporários).
- Cada recurso permite empacotar e carregar veículos até 72 bebidas por dia.
- Só as bebidas empacotadas é que podem ser distribuídas.

> Distribuição



- Pode ser alocado numa base diária, N_D recursos por dia.
- Cada recurso de Distribuição custa (aluguer veículo, motorista):
 - **V1**: custo fixo 40 R\$ por dia normal, distribui até 60 produtos (e.g., cervejas) por dia.
 - **V2**: custo fixo 50 R\$ por dia normal, distribui até 90 produtos por dia.
 - **V3**: custo fixo 53 R\$ por dia normal, distribui até 120 produtos por dia.
 - Se fim de semana, o custo acresce em 5 R\$ ao dia.



> Proveitos e custos nas lojas

Cada bebida vendida no próprio dia (uma unidade) tem este ganho para a empresa distribuidora:

- **STELLA:** 5.7 R\$.
- **BUD:** 4.4 R\$.

Cada bebida não vendida no próprio dia (em qualquer loja), tem o custo extra de refrigeração e armazenagem em loja (por dia de não venda) de 1 R\$, podendo ser vendida num dos dias seguintes.

Assumir também que no final da semana planeada, se existir stock, o custo de refrigeração e armazenamento final em loja é de 1 R\$ por unidade ainda existente.

No início do planeamento, as lojas estão vazias (sem cervejas nos frigoríficos, logo stock = 0).

Se o total de vendas previstas para um dia for X e o que existir nas lojas for Y, tal que $X > Y$, então só é vendido Y nesse dia.

> Objetivo Previsão:



Prever o valor diário de vendas de cada tipo de bebida

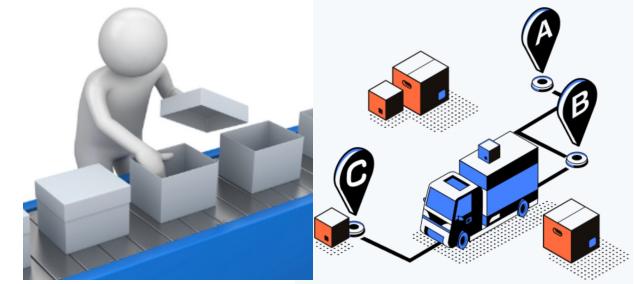
- Podem usar modelos univariados (e.g., somente atributo STELLA) ou multivariados de séries temporais, bem como modelos de *machine learning*.
- Os modelos devem ser treinados com os dados mais antigos e testados com dados mais recentes.
- Ter em atenção o horizonte temporal (prever até $H=7$ períodos em avanço – **uma semana**).
- Para modelos que utilizem dados meteorológicos, podem assumir que é possível obter com uma antecedência de $H=7$ dias uma previsão fiável dos mesmos (1 semana).

Requisitos Previsão



- Deve ser utilizado pelo menos **um modelo** de previsão por cada **membro do grupo**. Cada membro deve ser responsável pela implementação (código), configuração (ajuste de parâmetros) e experimentação desse modelo de previsão.
- Devem usar as métricas e métodos de avaliação que considerarem mais adequados para **avaliar a qualidade** dos modelos.

> Objetivo Otimização:



Para um dado dia d pretende-se definir um **plano de uso de recursos (armazém e distribuição)** para os próximos 7 dias (uma semana): entre $d+1$ e $d+7$.

Nota: **lucro final empresa = vendas – custo.**

> Exemplo de um plano para penúltima semana:

```
Fim de semana  
(19 e 20)
```

```
[1] "datas:"  
[1] "2020-12-18" "2020-12-19" "2020-12-20" "2020-12-21" "2020-12-22" "2020-12-23" "2020-12-24" "2020-12-25"  
valores estimados para venda:  
stella: 141 154 18 102 211 69 37 0  
bud: 211 172 220 330 39 45 125 0  
arm.: 6 1 0 1 1 0 1 > custo: 105  
v1: 2 0 0 1 0 0 0  
v2: 2 1 0 0 1 0 0  
v3: 2 1 0 0 0 0 0  
bebidas empacotadas e distribuidas:  
stella: 160 8 0 52 20 0 0  
bud: 200 200 0 0 30 0 0  
custo vei.: 120 + 205 + 170 = custo vei.: 495  
> vendas stella: 141 27 0 52 20 0 0 > lucro: 1368  
> vendas bud: 200 172 28 0 30 0 0 > lucro: 1892  
> stock stella: 19 0 0 0 0 0 0 > custo: 19  
> stock bud: 0 28 0 0 0 0 0 > custo: 28  
>> LUCRO FINAL: 2198 R$  
>> recursos= 10 + 3 + 4 + 3 = 20
```

The diagram consists of three blue arrows pointing from specific sections of the R code to their respective labels. The first arrow points from the highlighted section of the 'valores estimados para venda' block to the label 'Elementos do plano'. The second arrow points from the highlighted section of the 'bebidas empacotadas e distribuidas' block to the same label 'Elementos do plano'. The third arrow points from the highlighted section of the 'valores estimados para venda' block to the label 'Fim de semana (19 e 20)'.

Requisitos otimização:

- Para a otimização, assumir: valores previstos para os 7 próximos dias (**melhor**) OU usar valores reais;
- Assumir 2 possibilidades de otimização:
 - Maximizar o lucro da empresa para uma dada semana (**O1**).
 - Multiobjetivo: maximizar o lucro da empresa (**O1**) e minimizar o número total de recursos de armazém e distribuição (**O2**).
- Deve ser utilizado pelo menos um modelo de otimização por cada membro do grupo. Cada membro deve ser responsável pela implementação (código), configuração (ajuste de parâmetros) e experimentação desse modelo de otimização.

Valorização



- Estudo (relatório) sobre a previsão e **otimização**.
- Maior valorização usarem previsões e não valores reais na otimização, se testarem e compararem diversos modelos de otimização (maximizar o lucro, multiobjectivo, métodos de otimização moderna explicados nas aulas, outros métodos de outros pacotes do R ou de outras ferramentas), estudar o efeito de alterar algumas das variáveis (e.g., data em que se faz a simulação, valores de custo ou lucro de vendas).
- Acrescento de realismo ao problema, criando outras versões a otimizar.
- Sistema desenvolvido, incluindo **interface** em modo de consola ou **gráfica** (e.g., via shiny.rstudio.com) para demonstrar o sistema completo de Análise Inteligente de Dados (previsão e otimização) a funcionar.

> Notas:

- Não existe uma única forma ideal de resolver o projeto, é esperado que cada grupo execute um projeto diferente;
- Aceito alguma flexibilidade/originalidade em cada grupo, desde que **justifiquem as decisões tomadas**;
- Não necessitam de executar tudo o que é pedido para ter nota positiva.

> Execução

- A partir de hoje até data de entrega, apoio durante aulas.
- Durante as aulas será efetuada uma avaliação preliminar (não formal, com um peso reduzido) a cada grupo e projeto (convém estarem presentes todos elementos do grupo).
- A partir do dia **29/3/2022** (quarta), todos grupos devem mostrar ao docente um “semanário” de projeto (via Googledocs, overleaf, etc.), que deve incluir para cada semana uma entrada com:
 - Resumo do que foi **realizado nessa semana para o projeto**. Indicação de que **reuniões de trabalho** tiveram lugar (com o grupo todo) e sua duração.
 - Para **cada elemento do projeto**: o que fez em termos individuais e quantas horas dedicou ao projeto.

> Exemplo de um semanário

Semana 2:

1 de novembro de 2022 a 7 de novembro de 2022

Trabalho do grupo		Tarefas individuais		
Resumo	Reuniões	Responsável	Contribuição para o projeto	Horas dedicadas
- Ajustes à Análise Exploratória de Dados; - Aprofundamento da elaboração dos modelos preditivos; - Definição e cálculo das métricas.	Reunião no dia 01/11/2022 com uma duração total de 2h30 . Reunião no dia 07/11/2022 com uma duração de 2h15 .	João  	- Correção de alguns detalhes na análise exploratória; - Aperfeiçoamento do Modelo Elman; - Estudo detalhado do tipo de erros e elaboração de documento de suporte; - Análise dos modelos realizados pelos outros elementos do grupo.	9h
			- Continuação da implementação do modelo ARIMA; - Cálculo de métricas para o modelo implementado; - Estudo do modelo <u>ARIMAX</u> para possível implementação; - Análise dos modelos realizados pelos outros elementos do grupo.	6h
			- Ajustes em diversos gráficos da Análise Exploratória de Dados; - Implementação de diferentes variações do modelo Random Forest com a base de dados completa; - Otimizações no código global do projeto.	12h

> Exemplo de um semanário:

Semana 4

Reunião 8 - 17 de novembro - duração de 180 min (online)

Tópicos discutidos

- Correção dos modelos ARIMA e Holt-Winters
- Aplicação dos modelos ARIMAX e VAR
- Debugging do rolling window para modelos multivariados ML

Inputs individuais

- [REDACTED]: 250 min
 - Extensão do rolling window para modelos univariados de ML
 - Ajuste das funções de rolling window para retornarem listas com os valores previstos e reais em todos os ciclos e criar os gráficos REC e REG
- Alexandre [REDACTED] 180 minutos
 - Extensão do rolling window para HW
 - Escrita do relatório
- [REDACTED]: 180 min
 - Extensão do rolling window para ARIMA
 - Criação de mais time lags (7) para ambos os tipos de cerveja e integração em todos os modelos (holdout e rolling window)
- [REDACTED]: 190 min
 - Extensão do rolling window para ARIMAX
 - Alteração da estrutura das funções holdout para retornarem as métricas de desempenho e os valores previstos de Stella e Bud

> Entrega

- Dia **24/5/2023** (quarta – TP2 ou TP1).
- Cada grupo (todos elementos presentes) deve efetuar uma **apresentação** do trabalho realizado, respondendo a questões.
- Entrega de relatório (em pdf), slides da apresentação (em formato pdf), código R em anexo do ficheiro zip, anexos relevantes => ficheiro **.zip** a depositar no sistema **elearning**.

Video de Demonstração

- Deve ser **gravado um vídeo** (até máximo de 5 minutos, idealmente com narração) onde se demonstra o sistema desenvolvido (em R e/ou outro), sendo efetuado o upload no **youtube**.
- Incluir no relatório o link (ou links) do vídeo **youtube**.
- Existe diverso software gratuito de captura de ecrã para vídeo, por exemplo, fazer a pesquisa google:
screen capture video

> Relatório

- O relatório deve explicar aquilo que foi realizado:
 1. **Introdução** (breve descrição do que foi feito)
 2. **Execução do Projeto**: Os alunos que constituem o grupo, planeamento (que tipo de tarefas cada um executou ao longo do tempo), como funcionou o grupo, qual a autoavaliação (nota proposta para o projeto e diferenciação de notas). Ter 1 página por elemento do grupo, onde se descreve o que esse elemento fez em termos do projeto (e.g., esforço, tarefas, código, testes).
 3. **Objetivo de Previsão** (o que foi feito e resultados)
 4. **Objetivo de Otimização** (o que foi feito e resultados)
 5. **Demonstração do Sistema Desenvolvido** (1 frase com o link [youtube](#))
 6. **Conclusões** (breve apreciação do que foi realizado).
 - **Bibliografia** (opcional)
 - **Anexos** (opcional)

Tentar ter um máximo a rondar as 20 a 40 páginas de corpo de documento (sem capa, índices, bibliografia e anexos). Escrita direta, **24** sem “palha”!