



# Big Data

**Cesto de Compras**





## BUSINESS SCHOOL

Graduação, pós-graduação, MBA, Pós- MBA, Mestrado Profissional, Curso In Company e EAD



## CONSULTING

Consultoria personalizada que oferece soluções baseadas em seu problema de negócio



## RESEARCH

Atualização dos conhecimentos e do material didático oferecidos nas atividades de ensino



Líder em Educação Executiva, referência de ensino nos cursos de graduação, pós-graduação e MBA, tendo excelência nos programas de educação. Uma das principais **escolas de negócio do mundo**, possuindo convênios internacionais com Universidades nos EUA, Europa e Ásia. +8.000 **projetos de consultorias** em organizações públicas e privadas.



Único curso de graduação em administração a receber as notas máximas



A primeira escola brasileira a ser finalista da maior competição de MBA do mundo



Única *Business School* brasileira a figurar no *ranking* LATAM



Signatária do Pacto Global da ONU



Membro fundador da ANAMBA - Associação Nacional MBAs



Credenciada pela AMBA - Association of MBAs



Credenciada ao Executive MBA Council



Filiada a AACSB - Association to Advance Collegiate Schools of Business



Filiada a EFMD - European Foundation for Management Development



Referência em cursos de MBA nas principais mídias de circulação



## Breve Descrição Profissional

Frequente atuação em projetos analíticos para empresas dos seguimentos: financeiro, varejo e Telecomunicação. Implementação de soluções de Risco (Banco do Brasil, Bradesco e Itaú), Prevenção a Fraude (SEFAZ e Santander), Forecasting (Prismah e Natura), Retenção de clientes (Claro), Big Data Analytics(Petrobrás). Além de ter atuado, como pré-vendas e Gerente de Produto para Machine Learning realizando apresentações, Workshops e PoC's para clientes de diversos segmentos de mercado.



## Educação


- Graduação em Estatística (USP)
- Professor - Data Mining (FIA)
- Pós G. em Economia (FGV)
- Mestrando(USP)

**Superintendente de Data Analytics**

<https://www.linkedin.com/in/daniel-martins-9599727a/>



# Conteúdo da Aula

- 
1. Introdução
  2. Análise de Associação (Exemplos)
  3. Conceitos
  4. Aplicação no R
  5. Caso de Uso
  6. Material Adicional



# 1. Introdução



# Análise de Cesto de Compra

A análise de Cesto de Compras é um campo dentro da área de “Social Network Analysis” que possibilita detectar associação entre os dados.

Esta técnica é muito utilizada para detectar Padrões entre a mesma classe de entidades (produtos por exemplo) e também duas classes distintas de entidades ( clientes e produtos).

Esta técnica é muito utilizada para recomendar ações de compra conjunta ( veremos exemplos no decorrer da aula).





## Análise de Cesto de Compra

### Motivadores



- ? Onde os detergentes devem estar nas prateleiras para maximizar a venda deste produto?
- ? Limpa vidros são comprados quando detergentes e suco de laranja são comprados juntos?
- ? Refrigerante tipicamente é comprado junto com bananas? A marca do refrigerante faz diferença?
- ? Como a região de moradia afeta o que os clientes estão comprando?



## Analise de Associação

### Vantagens e Desvantagens

- ☐ Computacionalmente Eficiente
- ☐ Regras individuais são fáceis de interpretação
- ☐ Métodos comprovados
- ☐ Fácil Aplicação



- ☐ O numero de regras pode ser grande
- ☐ Algumas regras são triviais
- ☐ Há necessidade de Isolar padrões interessantes





## 2. Análise de Associação(Exemplos)



## Análise de Associação – Compras Mercado

### Exemplo1

Um cliente que vai ao mercado e coloca em seu carrinho :  
então com grande chance ele vai colocar também .....



### Tomada de decisão

Colocar o carvão próximo à picanha.

Colocar o carvão longe da picanha.

Vender o carvão e a picanha de forma casada.

Vender o carvão, a picanha e mais alguns produtos de forma casada.



## Exemplo2



Um cliente de Telecom que  
Reduz o uso de voz  
Reduz o uso de dados  
Então com grande chance ele vai .....



- Verificar se o cliente efetuou alguma reclamação
- Oferecer um novo plano ao cliente
- Oferecer um combo ao cliente
- Oferecer um pacote de produtos e serviços a família



## Análise de Associação – Investimentos

### Exemplo3

Após a análise dos dados pode-se concluir que, para um dado segmento, um cliente que possui : CDB, Previdência e Renda Fixa. Então com grande chance ele vai adquirir um seguro.

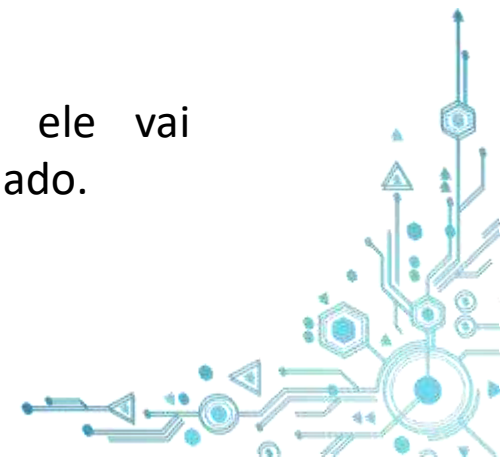


### Tomada de decisão

Após a análise dos dados pode-se concluir que, para um dado segmento, um cliente que possui

Empréstimo pessoal  
Usa o rotativo do cartão

Então com grande chance ele vai adquirir um automóvel financiado.



## Análise de Associação - Eletrodomésticos

### Exemplo4

Pode-se verificar que um cliente que adquiri no site: Geladeira, Micro-ondas com grande chance ele vai adquirir um fogão.



## Análise de Associação – E-Commerce

### Exemplo5



Diariamente milhões de compras online são realizadas em diferentes sites, possibilitando identificar o perfil do comprador de acordo com o “Rastro Digital” gerado nas diferentes transações.

Desta forma é possível identificar o perfil de cliente para realizar recomendações assertivas em momentos oportunos.





## 3. Conceitos

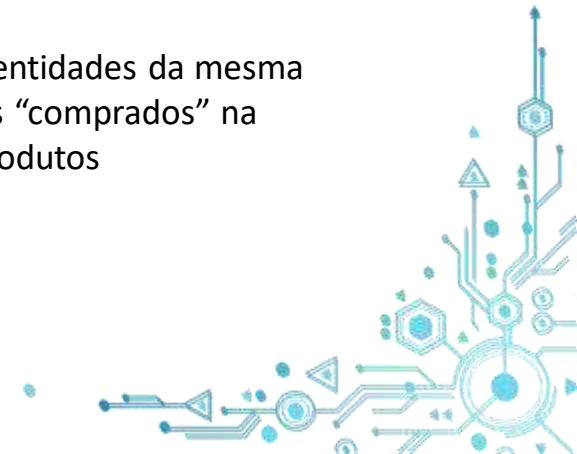


## Dados de Associação

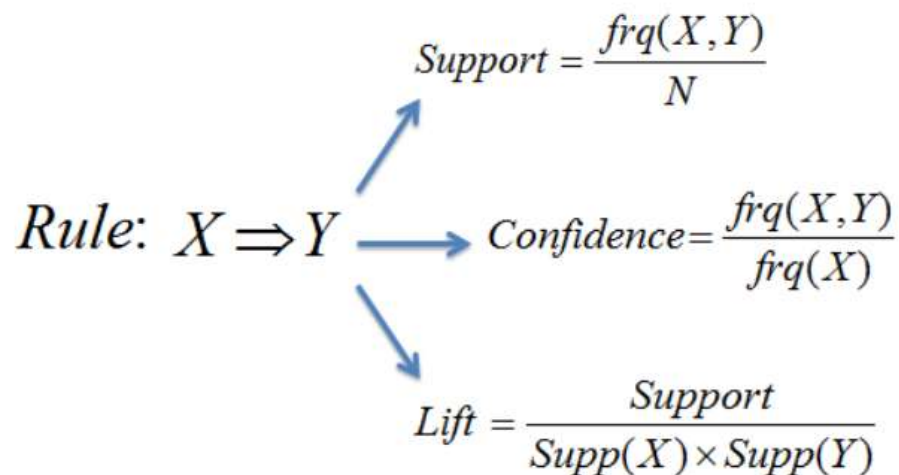
ID	Items
1	{Bread, Milk}
2	{Bread, <b>Diapers</b> , <b>Beer</b> , Eggs}
3	{Milk, <b>Diapers</b> , <b>Beer</b> , Cola}
4	{Bread, Milk, <b>Diapers</b> , <b>Beer</b> }
5	{Bread, Milk, Diapers, Cola}
...	...

Tabela Transacional

Análise de Associação é realizada com base em dados transacionais que descrevem a associação entre entidades da mesma classe(Produtos) ou diferentes entidades (clientes e Produtos). Uma compra é um conjunto de produtos “comprados” na mesma transação, ao passo que uma regra de associação representa a associação entre dois ou mais produtos

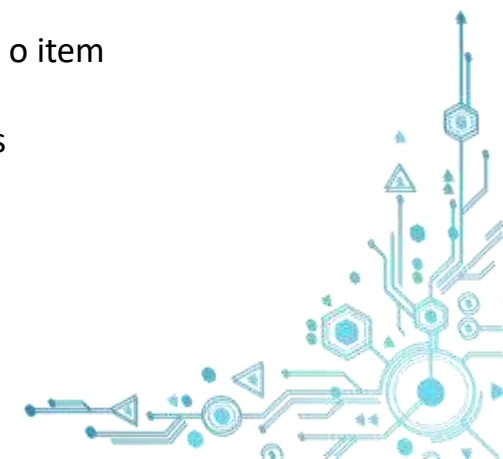


# Medidas de Associação



Rule	Support	Confidence	Lift
$A \Rightarrow D$	2/5	2/3	10/9
$C \Rightarrow A$	2/5	2/4	5/6
$A \Rightarrow C$	2/5	2/3	5/6
$B \& C \Rightarrow D$	1/5	1/3	5/9

- ❑ **Suporte:** Mede a frequência que uma combinação ocorre considerando-se todos os cupons fiscais (compras).
- ❑ **Confidence:** Mede a frequência de vezes em que os 2 itens foram comprados juntos, em relação as vezes que o item da esquerda foi comprado.
- ❑ **Lift:** Mede a associação entre os itens. Quando **Lift**  $\geq 1$  há **atração**, ao passo que **Lift**  $< 1$  há **retração** entre os produtos.



## Apriori Algoritmo

O algoritmo Apriori é o algoritmo de regras de associação mais conhecido e utilizado em diversas aplicações para diferentes indústrias e áreas do conhecimento. O algoritmo pode ser dividido em duas etapas:

- I. Encontrar os Itemsets(produtos) frequentes – aqueles que têm o Suporte maior que um valor mínimo;
- II. Gerar Regras de associação a partir dos itemsets frequentes – regras com Confiança maior do que um valor mínimo;

Para reduzir o número de regras evita-se a geração de regras que abrangem poucos exemplos . Para tanto, é definido um valor Mínimo para o Suporte assim como para as regras é denominado um valor mínimo de Confiança.



## Estudo de Caso

Neste **estudo de Caso** será apresentado o cálculo das medidas de associação para compras na Loja Renner. Deve-se verificar quais itens foram comprados em cada Cupom Fiscal.



LOJAS RENNER S.A.					
AV. ASSIS BRASIL, 1010 - PORTO ALEGRE - RS					
CNPJ:92.754.738/0013-04			IE:096/0372202		
03/05/2013 10:05:40		CCF:002260		COO:010169	
CUPOM FISCAL					
ITEM	CODIGO	DESCRIÇÃO			
QTD.	UN.	VL.UNIT( R\$)	ST	IAT	VL.ITEM( R\$)
001	530236310	Blusa n3 4			
	1,000Un x	89,90	T04	A	89,90%
002	520196220	PERFUME PA			
	1,000Un x	199,00	F1	A	199,00%
TOTAL		R\$		288,90	
Dinheiro				288,90	
T04=17,00%					



## Estudo de Caso - Renner

Para exemplificar o cálculo das medidas de associação são considerados 5 cupons(compras) em que os clientes compram meias de 5 tipos distintos.



<u>Regra</u>	<u>Support</u>	<u>Confidence</u>	<u>Lift</u>
A → C	2/5	2/3	0.83
A → D	2/5	2/3	1.11
A & D → E	1/5	1/2	1.25





## Exercícios Teóricos

**Cesto de Compras**

Exercícios Teóricos

Nº	Leite	Café	Cerveja	Pão	Manteiga	Arroz	Feijão
1	Não	Sim	Não	Sim	Sim	Não	Não
2	Sim	Não	Sim	Sim	Sim	Não	Não
3	Não	Sim	Não	Sim	Sim	Não	Não
4	Sim	Sim	Não	Sim	Sim	Não	Não
5	Não	Não	Sim	Não	Não	Não	Não
6	Não	Não	Não	Não	Sim	Não	Não
7	Não	Não	Não	Sim	Não	Não	Não
8	Não	Não	Não	Não	Não	Não	Sim
9	Não	Não	Não	Não	Não	Sim	Sim
10	Não	Não	Não	Não	Não	Sim	Não

**Exercício 1**

Execute os passos seguintes considerando que:

**Suporte** = Número de registros com X e Y / Número total de registros

Suporte  $\geq 0,3$  (Ponto de Corte da Medida de Associação Suporte)

Passo 1: Calcule o suporte de conjuntos com 1 item (itens/sets/compras)

Item	Suporte
Leite	

## Medidas de Associação

Com base em uma tabela transacional contendo 10 compras referente aos produtos: Leite, Café, Cerveja, Pão, Manteiga, Arroz e Feijão as Medidas de associação serão calculadas e as regras de maiores destaque serão consideradas para analisar as respectivas associações.

O Arquivo “Exercícios Teóricos.docx” tem as instruções detalhadas para cada etapa dos exercícios mencionados.



## 4. Aplicação no R



## Pacote “arules” e função apriori



### Função Apriori

A função “Apriori” do pacote arules permite implementar o algoritmo de mesmo nome em dados transacionais para análise de associação. Os principais parâmetros da função são:

**Data:** Tabela Transacional

**Parameter:** Lista dos parâmetros de Suporte, Confidence e target

**Apperance:** restrição de itens

**Control:** regula a performance do algoritmo

### Exemplo de Aplicação:

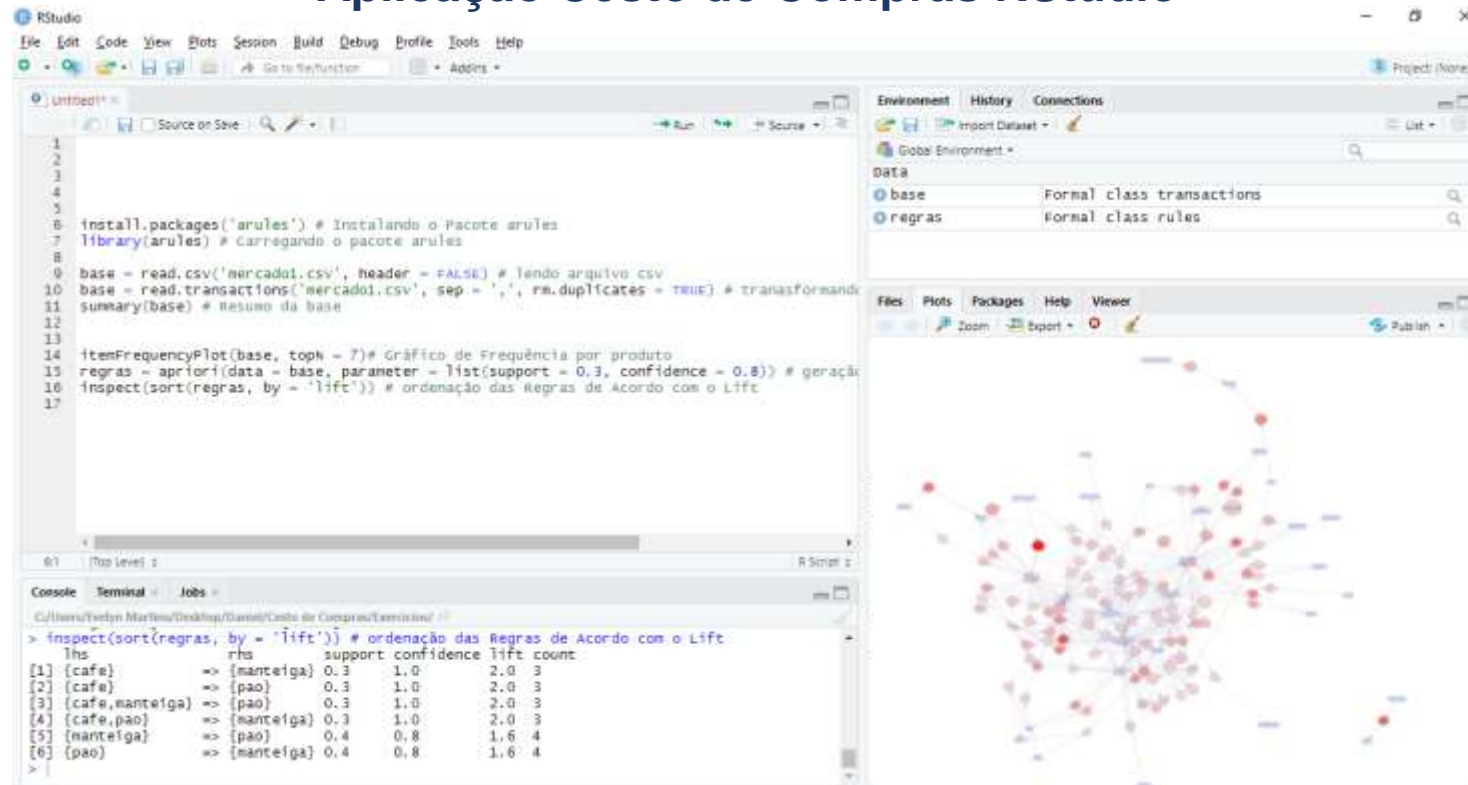
```
data("Adult")
rules <- apriori(Adult, parameter = list(supp = 0.5, conf = 0.9, target = "rules"))
summary(rules)
```

<https://cran.r-project.org/web/packages/arules/arules.pdf>

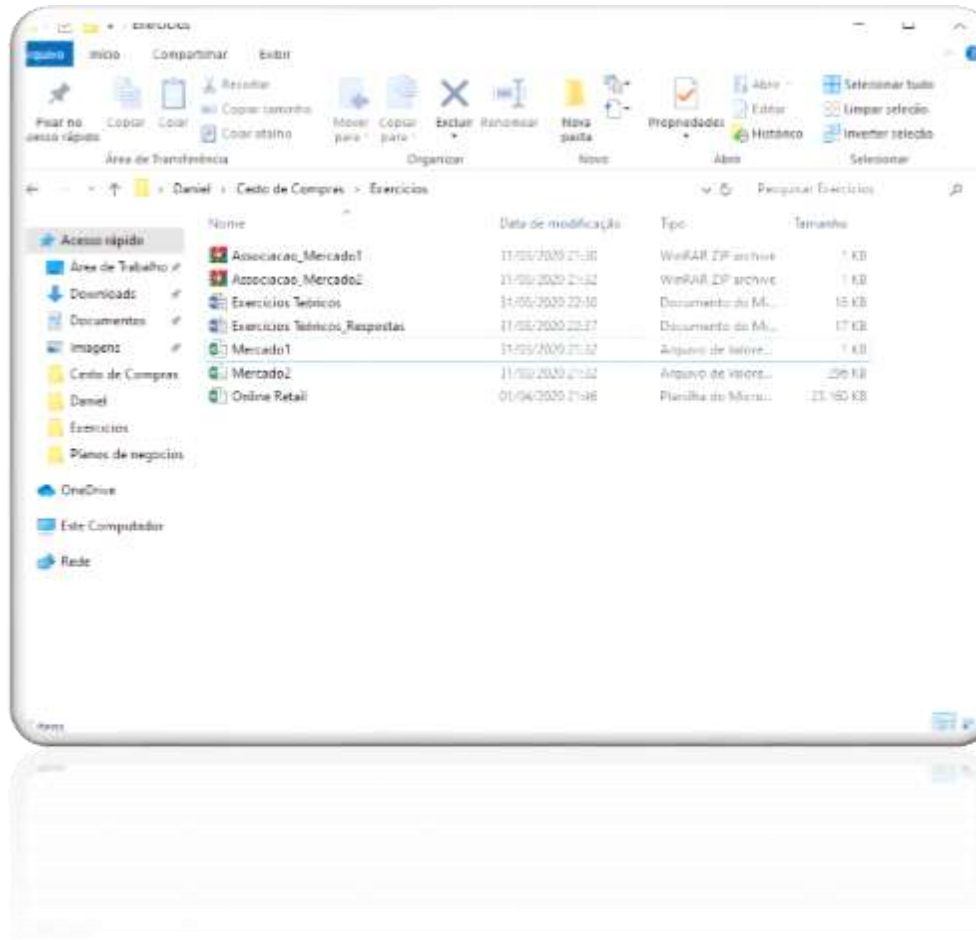


# Demonstração - Mercado1

## Aplicação Cesto de Compras RStudio



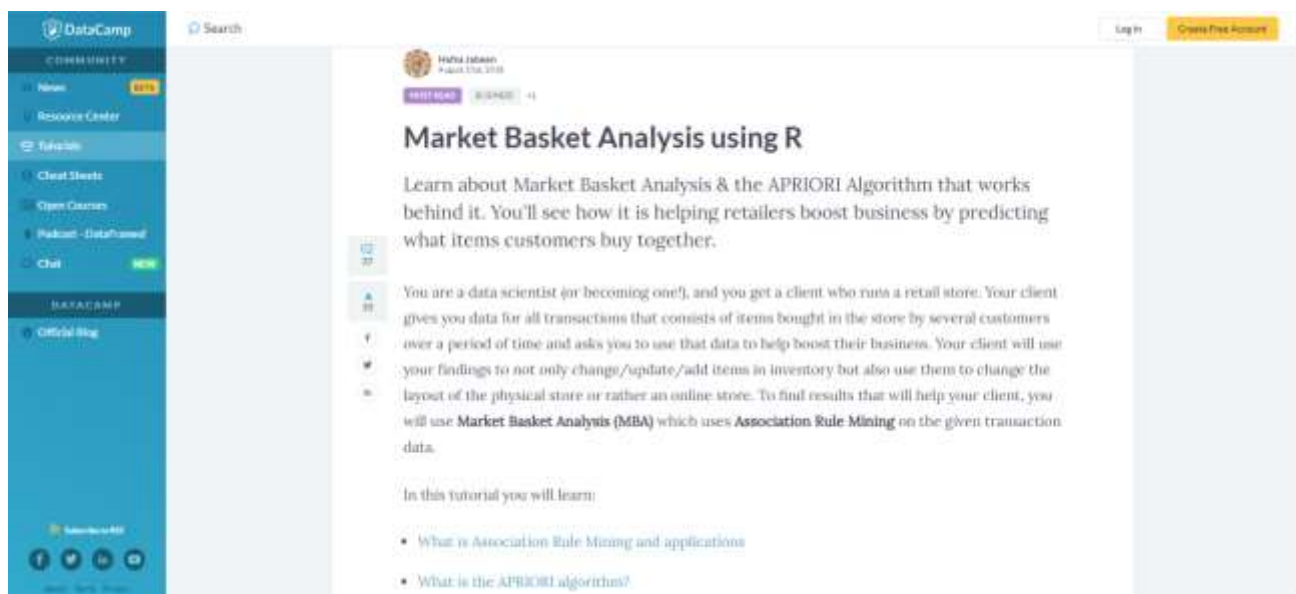
## Exercícios



Na pasta Exercícios abra o arquivo “Associação\_Mercado1.r” que contem o código comentado da aplicação de ‘Market Basket Analysis’ para a base transacional de compras denominada Mercado1. O arquivo “Associação\_Mercado2.r” é referente a base Mercado2.



## Demonstração - Caso de Uso no R



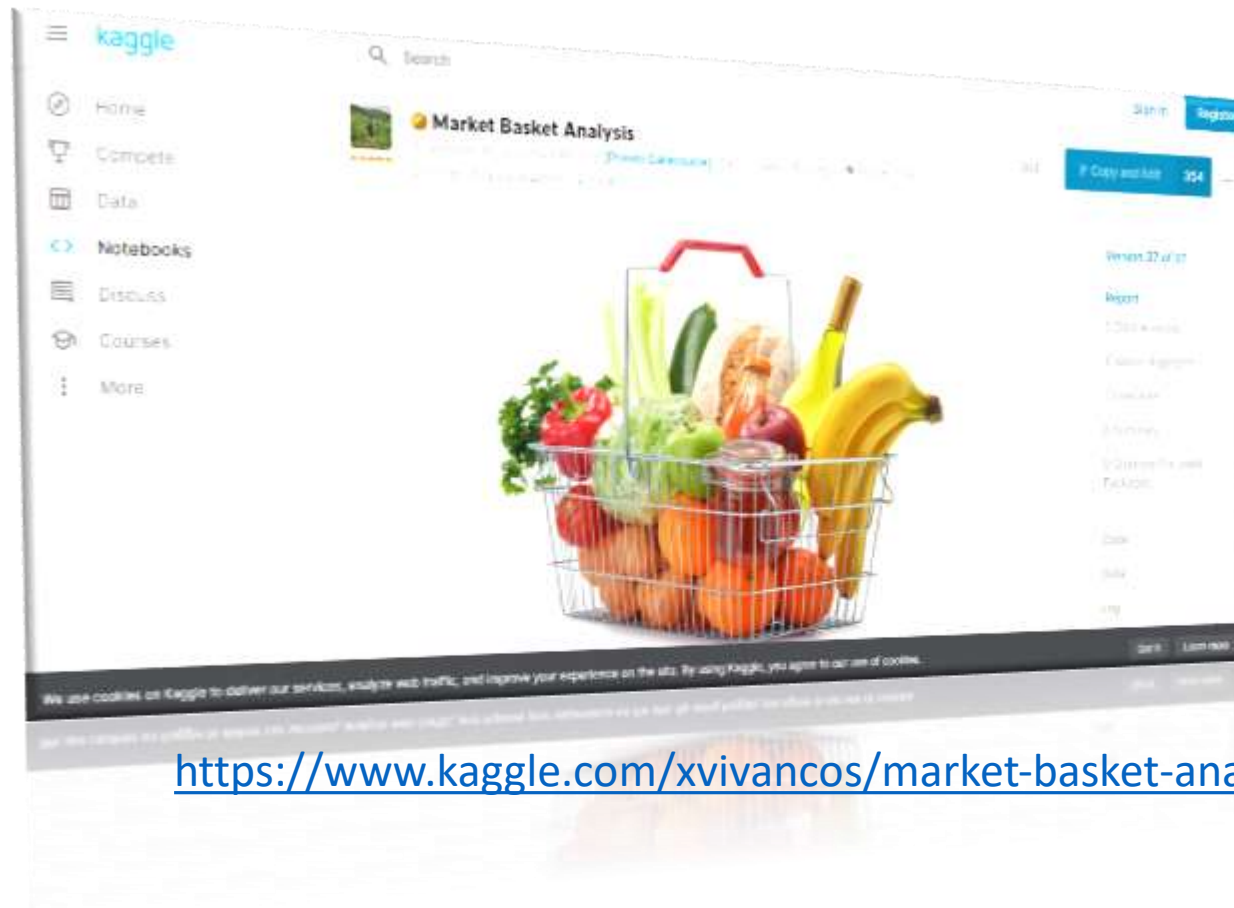
O código e a base de dados do caso de Uso de Cesto de Compras denominado “CasodeUso\_DataCamp” segue na pasta Exercícios em uma versão simplificada e comentada.

<https://www.datacamp.com/community/tutorials/market-basket-analysis-r>





## Exercícios Caso de Uso



<https://www.kaggle.com/xvivancos/market-basket-analysis>

O caso de Uso disponível no Kaggle sobre Market Basket Analysis será o enredo do nosso exercício. O conteúdo pode ser acessado através do link disponível neste slide.



# Cesto de Compras

## *Papers sobre Cesto de Compras*

[https://www.maxwell.vrac.puc-rio.br/14008/14008\\_4.PDF](https://www.maxwell.vrac.puc-rio.br/14008/14008_4.PDF)

[https://en.wikipedia.org/wiki/Apriori\\_algorithm](https://en.wikipedia.org/wiki/Apriori_algorithm)

<https://cran.r-project.org/web/packages/arules/vignettes/arules.pdf>

## *Casos de Uso ( Aplicações em R)*

[https://rpubs.com/rpandey\\_20/310664](https://rpubs.com/rpandey_20/310664)

<https://datascienceplus.com/visualize-market-basket-analysis-in-r/>

<https://www.datacamp.com/community/tutorials/market-basket-analysis-r>

<https://towardsdatascience.com/a-gentle-introduction-on-market-basket-analysis-association-rules-fa4b986a40ce>

<https://datascienceplus.com/a-gentle-introduction-on-market-basket-analysis%E2%80%8A-%E2%80%8Aassociation-rules/>

