

Estatística Aplicada

EAD Ao Vivo

Tema da aula
Cases em Python



13/01/2021



APLICAÇÕES DE ESTATÍSTICA PARA TOMADA DE DECISÃO



Professora:
Dr^a Karin Ayumi Tamura

Coordenadores:
Prof^a Dr^a Alessandra de Ávila Montini
Prof^a Dr. Adolpho Walter Pimazoni Canton





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

O **Laboratório de Análise de Dados** – LABDATA é um Centro de Excelência que atua nas áreas de ensino, pesquisa e consultoria em análise de informação utilizando técnicas de **Big Data, Analytics** e **Inteligência Artificial**.



O LABDATA é um dos pioneiros no lançamento dos cursos de *Big Data* e *Analytics* no Brasil

Os diretores foram professores de grandes especialistas do mercado

+10 anos de atuação

+1000 alunos formados

Docentes

- Sólida formação acadêmica: doutores e mestres em sua maioria
- Larga experiência de mercado na resolução de *cases*
- Participação em Congressos Nacionais e Internacionais
- Professor assistente que acompanha o aluno durante todo o curso

Estrutura

- 100% das aulas realizadas em laboratórios
- Computadores para uso individual durante as aulas
- 5 laboratórios de alta qualidade (investimento +R\$2MM)
- 2 Unidades próximas a estação de metrô (com estacionamento)



Profª Dra.
Alessandra Montini

Diretora do LABDATA-FIA, apaixonada por dados e pela arte de lecionar. Têm muito orgulho de ter criado na FIA cinco laboratórios para as aulas de Big Data e inteligência Artificial. Possui mais de 20 anos de trajetória nas áreas de Data Mining, Big Data, Inteligência Artificial e Analytics. Cientista de dados com carreira realizada na Universidade de São Paulo. Graduada e mestra em estatística aplicada pelo IME-USP e doutora pela FEA-USP. Com muita dedicação chegou ao cargo de professora e pesquisadora na FEA-USP, ganhou mais de 30 prêmios de excelência acadêmica pela FEA-USP e mais de 30 prêmios de excelência acadêmica como professora dos cursos de MBA da FIA. Orienta alunos de mestrado e de doutorado na FEA-USP. Membro do Conselho Curador da FIA, Coordenadora de Grupos de Pesquisa no CNPQ, Parecerista da FAPESP e Colunista de grandes Portais de Tecnologia.

 [linkedin.com/in/alessandramontini/](https://www.linkedin.com/in/alessandramontini/)



Prof. Dr.
Adolpho Walter Canton

Diretor do LABDATA-FIA. Consultor em Projetos de *Analytics*, *Big Data* e Inteligência Artificial. Professor FEA – USP. PhD em Estatística Aplicada pela *University of North Carolina at Chapel Hill*, Estados Unidos.



Currículo - Profª Drª Karin Ayumi Tamura

FORMAÇÃO ACADÊMICA | EXPERIÊNCIA PROFISSIONAL

6



Profª Dra.
Karin Ayumi Tamura

Contato: karin.tamura@fia.com.br

- **FORMAÇÃO ACADÊMICA:** Pós-doutora (2015), Doutora (2012), mestre (2007) e bacharel (2003) em Estatística pelo Instituto de Matemática e Estatística da USP, tendo como área de pesquisa modelos de regressão, análise multivariada de dados e algoritmos de *machine learning*.
- **ATUAÇÃO PROFISSIONAL:** Foi *Head* de *Analytics* por 14 anos, e atualmente é Conselheira Executiva e *Head* de Inovação na *Marketdata Solutions*, uma empresa do grupo WPP, e Professora Doutora no LABDATA FIA.
- **HISTÓRICO:** Atuação no mercado por 17 anos, com experiência profissional no segmento bancário (Bradesco) e consultoria (*Marketdata Solutions*). Atuou como docente em cursos de pós-graduação (2010-15) no LABDATA FIA e ABEMD. Especialista em Estatística e *Advanced Analytics* trabalhando em projetos de diversos segmentos do mercado. Participante de congressos nacionais e internacionais voltados a área de Estatística, Dados e Algoritmos de *Machine Learning*.

"Tenho duas paixões no meu trabalho: dados e pessoas. Voltar a lecionar no LABDATA FIA está sendo a realização de um sonho planejado desde a minha época de aluna de pós-graduação. Meu objetivo como professora é integrar a visão do mercado com as técnicas e tecnologias de análise de dados, por meio de uma atuação humanista no ensino aos alunos"

Projetos atendidos



Conteúdo Programático do Curso

21 AULAS AO VIVO COM PROFA. KARIN | 27 PLANTÕES AO VIVO COM PROF. STEPHAN, 7 LISTAS DE EXERCÍCIOS E EAD VIDEO AULA EM PYTHON

7

Dia	Mês	Aula	EAD Ao Vivo	Plantão Prof. Stephan
5	Agosto	Introdução ao Curso e Análise Exploratória de Dados	Aula Prof. Karin	06/ago
12	Agosto	Análise Exploratória de Dados	Aula Prof. Karin	13/ago
19	Agosto	Análise Exploratória de Dados - Introdução ao R	Aula Prof. Karin	20/ago
26	Agosto	Lista de Exercícios em Sala de Aula (19hs-23hs - com presença obrigatória)	-	27/ago
2	Setembro	Regressão Linear Simples	Aula Prof. Karin	03/set
9	Setembro	Regressão Linear Simples e Múltipla	Aula Prof. Karin	10/set
16	Setembro	Regressão Linear Simples e Múltipla	Aula Prof. Karin	17/set
23	Setembro	Lista de Exercícios em Sala de Aula (19hs-23hs - com presença obrigatória)	-	24/set
30	Setembro	Análise de Cluster	Aula Prof. Karin	01/out
7	Outubro	Análise de Cluster	Aula Prof. Karin	08/out
14	Outubro	Lista de Exercícios em Sala de Aula (19hs-23hs - com presença obrigatória)	-	15/out
21	Outubro	Arvore de Decisão	Aula Prof. Karin	22/out
28	Outubro	Lista de Exercícios em Sala de Aula (19hs-23hs - com presença obrigatória)	-	29/out
4	Novembro	Regressão Logística	Aula Prof. Karin	05/nov
11	Novembro	Regressão Logística	Aula Prof. Karin	11/nov
18	Novembro	Lista de Exercícios em Sala de Aula (19hs-23hs - com presença obrigatória)	-	19/nov
25	Novembro	estudo de caso	Aula Prof. Karin	26/nov
2	Novembro	estudo de caso	Aula Prof. Karin	30/dez
9	Dezembro	estudo de caso	Aula Prof. Karin	10/dez
16	Dezembro	Análise de Série Temporal - modelo auto regressivo	Aula Prof. Karin	17/dez
23	Dezembro	Lista de Exercícios em Sala de Aula (Frequência Liberada - véspera Natal)	-	-
Recesso Escolar		EAD - INTRODUÇÃO AO PYTHON	EAD Video Aula	-
		EAD - INTRODUÇÃO AO PYTHON	(8 horas)	-
6	Janeiro	Modelos estatísticos em Python	Aula Prof. Karin	07/jan
13	Janeiro	Modelos estatísticos em Python	Aula Prof. Karin	14/jan
20	Janeiro	Modelos estatísticos em Python	Aula Prof. Karin	20/jan
27	Janeiro	Introdução a Big Data - Aplicações de Machine Learning e Deep Learning	Aula Prof. Karin	28/jan
3	Fevereiro	Aplicações de Machine Learning	Aula Prof. Karin	04/fev
10	Fevereiro	Aplicações de Machine Learning	Aula Prof. Karin	11/fev
17	Fevereiro	Lista de Exercícios (Frequência Liberada - quarta de cinzas)	-	18/fev
24	Fevereiro	EXERCICIOS DE REVISÃO - EAD (19hs e 23hs - com presença obrigatória)	-	24/fev
3	Março	Prova (Plataforma On Line: 19hs e 23hs)	-	

Cases

- 1. Análise Exploratória
 - i. Companhia MB
- 2. Regressão Linear
 - i. Limite de Crédito
 - ii. Start up
- 3. Análise de Cluster
 - i. Consumo de Alimentos
 - ii. Varejo
 - iii. Serviço de Entregas
- 4. Árvore de Decisão
 - i. Pedidos On Line
 - ii. Churn Telecom

1. Cases



Descreva de forma sucinta (com apoio de gráficos e tabelas) o perfil sócio demográfico dos funcionários da Companhia MB:

- Qual o tamanho da base de dados?
- Quais informações temos disponível para análise?
- Qual a idade da pessoas?
- Qual o grau de instrução dos funcionários desta empresa?
- De onde os funcionários vieram?
- Qual a renda (em salários mínimos) deles?
- Quantas pessoas têm filhos? Qual a quantidade de filhos?



N	estado_civil	grau_instrucao	n_filhos	salario	idade_anos	reg_procedencia
1	solteiro	ensino fundamental		4	26	interior
2	casado	ensino fundamental	1	4,56	32	capital
3	casado	ensino fundamental	2	5,25	36	capital
4	solteiro	ensino médio		5,73	20	outra
5	solteiro	ensino fundamental		6,26	40	outra
6	casado	ensino fundamental	0	6,66	28	interior
7	solteiro	ensino fundamental		6,86	41	interior
8	solteiro	ensino fundamental		7,39	43	capital
9	casado	ensino médio	1	7,59	34	capital
10	solteiro	ensino médio		7,44	23	outra
11	casado	ensino médio	2	8,12	33	interior
12	solteiro	ensino fundamental		8,46	27	capital
13	solteiro	ensino médio		8,74	37	outra
14	casado	ensino fundamental	3	8,95	44	outra
15	casado	ensino médio	0	9,13	30	interior
16	solteiro	ensino médio		9,35	38	outra
17	casado	ensino médio	1	9,77	31	capital
18	casado	ensino fundamental	2	9,8	39	outra
19	solteiro	superior		10,53	25	interior
20	solteiro	ensino médio		10,76	37	interior

Utilize todas as ferramentas aprendidas até o momento para tirar suas conclusões de negócio.

Arquivo: Companhia_MB.txt

Código Python: 01-Analise_Exploratoria-CompanhiaMB

@2020 LABDATA FIA. Copyright all rights reserved.



Limite de Cheque Especial

BANCO DE DADOS EM .TXT | FAZER ANÁLISE NO PYTHON

11

Uma instituição financeira tem objetivo de estimar o valor de **Limite de Cheque Especial** para seus novos clientes, com base em informações disponíveis em seu banco de dados. Para o estudo, foi disponibilizado uma amostra histórica de clientes com as informações de **Idade**, **Rendimento Total**, **Salário**, **Limite de Crédito Imediato**, **Escolaridade** para investigar se é possível estimar o Limite do Cheque Especial com base nas características disponibilizadas. Avalie a possibilidade de fornecer uma "regra" por meio de um modelo estatístico, interprete como as informações predizem o evento de interesse e qual a performance desta "regra".

Fonte: Base de dados inspirada em cases reais.



Utilize todas as ferramentas aprendidas até o momento para tirar suas conclusões de negócio.

Vamos fazer juntos?

R Studio®

Arquivo: Limite_Credito_Escolaridade.txt
Código Python: 02-Regressao_Linear-Limite_Credito

@2020 LABDATA FIA. Copyright all rights reserved.



Start up

BANCO DE DADOS EM .TXT | FAZER ANÁLISE NO PYTHON

12

Um investidor deseja estimar o lucro de Startups de acordo com suas características, com objetivo de tomada de decisão de investimento baseada no lucro projetado da empresa. A base de dados possui características de investimento e região das empresas já investidas do histórico do investidor.



Utilize todas as ferramentas aprendidas até o momento para tirar suas conclusões de negócio.

Arquivo: Startup.txt

Código Python: 02-Regressao_Linear-Startup

@2020 LABDATA FIA. Copyright all rights reserved.



Consumo Alimentar

BANCO DE DADOS EM .TXT | FAZER ANÁLISE NO PYTHON

13

Os dados são de uma pesquisa de consumo de alimentos em 25 países da Europa. Nove grupos de alimentos foram analisados: carne vermelha, carne branca, ovos, leite, peixes, cereais, carboidratos, grãos, frutas e vegetais. Os dados foram obtidos da DASL (The Data and Story Library).

O objetivo do estudo é agrupar os países segundo comportamentos de hábitos alimentares semelhantes, associando os hábitos alimentares com indicadores de longevidade e doenças crônicas de cada grupo de países.



Utilize todas as ferramentas aprendidas até o momento para tirar suas conclusões de negócio.

Arquivo: Consumo_Alimentos..txt

Código Python: 03-Analise_Cluster-ConsumoAlimentar

@2020 LABDATA FIA. Copyright all rights reserved.



Uma empresa de e-commerce deseja agrupar seus clientes para criar diferentes ações de marketing, com base em três variáveis: quantidade de compras, valor dos produtos e nota de revisão.



Utilize todas as ferramentas aprendidas até o momento para tirar suas conclusões de negócio.

Arquivo: Varejo.txt

Código Python: 03-Analise_Cluster-Varejo



Serviço de Entregas

BANCO DE DADOS EM .TXT | FAZER ANÁLISE NO PYTHON

15

Uma empresa de serviço de delivery de comida pronta tem o objetivo de fazer ações de relacionamento e reconhecimento com seus clientes. Para isso gostaria de identificar os perfis de clientes referentes à frequência de pedidos, o valor desses pedidos, a distância entre o estabelecimento-residência do cliente, e o tempo de entrega.

Fonte da Base: <https://www.kaggle.com/asaumya/k-means-clustering-food-delivery-case-study/data>



Utilize todas as ferramentas aprendidas até o momento para tirar suas conclusões de negócio.

Arquivo: Servico_Entrega.txt

Código Python: 03-Analise_Cluster-Servico_Entrega

@2020 LABDATA FIA. Copyright all rights reserved.



Pedidos de refeições

BANCO DE DADOS EM .TXT | FAZER ANÁLISE NO PYTHON

16

Uma empresa de tecnologia, atuante no ramo de entrega de refeições solicitadas online, gostaria de entender quais os perfis de consumidores que mais se deixam influenciar pela nota de avaliação dos restaurantes parceiros, no momento de realizarem seu pedido. Para entender esse aspecto, realizaram uma pesquisa com 2.744 clientes que fizeram algum pedido nos últimos 30 dias.

Utilize todas as ferramentas aprendidas até o momento para tirar suas conclusões de negócio.



Arquivo: Pedidos_Refeicoes.txt
Código Python: 04-Arvore_Decisao-PedidosOnline

@2020 LABDATA FIA. Copyright all rights reserved.



Estudo de Caso - *Churn* em Telefonia

BANCO DE DADOS EM .TXT | FAZER ANÁLISE NO PYTHON

17

Uma empresa de Telecom com 318.463 mil clientes está preocupada com o aumento do cancelamento voluntário das assinaturas de telefonia fixa. No período de 1 mês, existiram 2.766 cancelamentos voluntários (0,87% dos clientes cancelaram voluntariamente). Apesar deste número parecer pequeno, em um ano se nada for feito, 10% da base de clientes deixarão a companhia. Por meio do estudo da base de dados, o objetivo da empresa é identificar os clientes com maior probabilidade de cancelar voluntariamente suas linhas telefônicas e fazer ações segmentadas de marketing de acordo com seu perfil. Serão estudadas as características transacionais dos clientes para explicar o evento de cancelamento voluntário. O gestor da área de CRM gostaria de saber qual o perfil dos clientes que mais cancelam, quais suas características e se essa regra é eficaz para predizer o comportamento daqueles clientes que cancelam no próximo mês.

Fonte: base simulada, inspirada em problemas reais de consultoria da Prof^a Karin Ayumi Tamura.



Utilize todas as ferramentas aprendidas até o momento para tirar suas conclusões de negócio.

Arquivo: Churn_Telecom.txt

Código Python: 04-Arvore_Decisao-ChurnTelecom

@2020 LABDATA FIA. Copyright all rights reserved.

