

TI do Zero ao Pro

- Sua escolha vai ser entre Analista de Dados ou Cientista de Dados. Incluso também uma especialização a escolha entre:
 - Power BI: **Licença inclusa**
 - Tableu: **Necessário comprar licença a parte**

Especialização em Analista de Dados, professor André Perez



- Senior Data Engineer na SumUp
- Especialista em Engenharia de Dados na SumUP, já passou por empresas como **Serasa Experian** e **Amdocs**, onde coordenou projetos de ciência de dados em Tel-Aviv, Israel. É formado pela **Universidade de São Paulo** e pela **Universidade de Glasgow**, no Reino Unido. Com experiências internacionais, é professor e já foi tutor do MBA em ciência de dados do Instituto de Ciências Matemáticas e Computação (ICMC/USP), onde ajudou profissionais de diversas áreas a ingressarem no mercado de dados.

Python para análise de Dados.

- Aprenda a programar em Python para coletar e manejar dados. Conheça a sintaxe, crie funções e módulos, baixe pacotes, aplique uma programação orientada a objetos e execute scripts em Python.
 - Variáveis e tipos de dados (Data types)
 - Estrutura de dados
 - Fluxo de controle e fluxo condicional
 - Arquivos e funções
 - Programação funcional
 - Programação orientada a objetos
 - Módulos e pacotes
 - Lidando com erros
 - Transformação, limpeza e visualização de dados em Python

Atividade: Exercícios de programação e armazenamento em conta Github.

O processo de análise de dados

- Aprenda como coletar, processar e preparar os dados para análise e visualizações posteriores
 - Mineração de dados
 - Preparação de dados
 - Visualização de dados
 - Análise exploratória com Seaborn e Pandas

Atividade: Exercícios práticos de Data Mining, preparação e visualização de dados com Python, Pandas, Seaborn e armazenamento em conta Github.

Controle de versão

- prenda a trabalhar com Git e Github para controlar as versões do código da análise de dados.
Desenvolva o trabalho colaborativamente em equipe.
 - Introdução ao Git e ao Github
 - Criando um repositório Github
 - Executando mudanças no seu repositório
 - Salvando e baixando as mudanças
 - Trabalhando com Git branches

Atividade: Criar e publicar trabalhos no GitHub.

Fundamentos de matemática para análise de dados

- Aprenda a operar com Numpy e as habilidades matemáticas necessárias para trabalhar com Análise de dados.
 - Trabalhando com Numpy
 - Objetos array do Numpy
 - Vetorização de problema
 - Fundamentos da estatística

Atividade: Utiliza a biblioteca do Numpy para executar funções matemáticas em dados e resolver problemas.

Machine Learning

- Familiarize-se com o machine learning e construa modelos de análise usando inteligência artificial.
 - Fundamentos do machine learning
 - Regressão
 - Classificação
 - Agrupamento
 - Dados em série temporal
- **Atividade:** Criar modelos de Machine Learning de classificação não supervisionado com k-means e outros métodos, configurar e classificar dados.

SQL para análise de dados

- Aprenda a operar em ambiente SQL, criar scripts a partir de seleção, filtragem, agregação e mesclagem de dados, executar comandos de descrição, articular SQL com métodos quantitativos e utilizar bancos de dados.
 - Base de dados e linguagem SQL
 - Trabalhando com tabelas
 - Seleção e ordenação
 - Filtragem e seleção condicional
 - Agregação
 - Trabalhando com diversas tabelas
 - Técnicas avançadas em SQL

Atividade: Análise exploratória em bancos de dados SQL com exercícios com Pandas e Seaborn.

Visualização interativa de Dados

- Saiba como criar e operar dashboards de visualização de dados usando Google Data Studio.
 - Visualização interativa de dados
 - Introdução ao Google Data Studio
 - Medidas e dimensões
 - Criação e publicação de dashboards para visualização
- **Projeto:** Exercícios em Google Data Studio e visualização interativa de dados

Trabalhando em ambiente de Nuvem

- Aprenda a operar a plataforma de nuvem da Amazon Web Services e seus elementos relacionados à análise de dados.
 - Iniciando-se no Amazon Web Services
 - Rodando Python com funções de AWS Lambda
 - Gerenciamento de tarefas com AWS
 - Serviços de gerenciamento com AWS CloudWatch
 - Trabalhando com Sagemaker notebooks
 - Rodando aplicações em máquinas virtuais de AWS EC2

Atividade: Criar um AWS Datalake e realizar a automação de pipeline de ingestão de dados.

Big Data e Data Lake no AWS

Saiba como trabalhar com Spark, combinando-o com diferentes bibliotecas e estratégias de manipulação e processamento de dados em grandes quantidades. Crie Data lakes.

- Introdução ao Spark
- Trabalhando com Apache Spark
- Processamento de grande quantidade de dados
- Processamento de dados usando Pandas, PySpark e Pyarrow
- Trabalhando com Data lakes
- Extract load transform (ETL)

Atividade: Crie um Data lake no AWS e utiliza Python para extrair, transformar e carregar dados nele.

Especialização em Cientista de Dados, professor André Perez



- Senior Data Engineer na SumUp
- Especialista em Engenharia de Dados na SumUP, já passou por empresas como **Serasa Experian** e **Amdocs**, onde coordenou projetos de ciência de dados em Tel-Aviv, Israel. É formado pela **Universidade de São Paulo** e pela **Universidade de Glasgow**, no Reino Unido. Com experiências internacionais, é professor e já foi tutor do MBA em ciência de dados do Instituto de Ciências Matemáticas e Computação (ICMC/USP), onde ajudou profissionais de diversas áreas a ingressarem no mercado de dados.

Especialização em Cientista de Dados, professor João Serrajordia



- Cientista de Dados e Estatístico Msc
- Trabalhou os últimos oito anos no Banco Votorantim, desenvolvendo novos métodos de análise de dados, orientando metodologias internas e responsável pelo controle de qualidade dos projetos de ciência de dados com foco em crédito.
É graduado e mestre em estatística pela Universidade de São Paulo.

Especialização em Cientista de Dados, professor Lucas Serra de Assis



- Cientista de Dados na Saint-Gobain
- Atuou como cientista de dados de crédito e CRM no Banco Votorantim e hoje é cientista de dados na multinacional Saint-Gobain. É formado em Engenharia da Computação pela Universidade Federal de São Carlos

Python para análise de Dados.

- Aprenda a programar em Python para coletar e manejar dados. Conheça a sintaxe, crie funções e módulos, baixe pacotes, aplique uma programação orientada a objetos e execute scripts em Python.
 - Variáveis e tipos de dados (Data types)
 - Estrutura de dados
 - Fluxo de controle e fluxo condicional
 - Arquivos e funções
 - Programação funcional
 - Programação orientada a objetos
 - Módulos e pacotes
 - Lidando com erros
 - Transformação, limpeza e visualização de dados em Python

Atividade: Exercícios de programação e armazenamento em conta Github.

Introdução à ciência de dados

Conheça o mercado de ciência de dados, a experiência dos professores em diferentes setores e os primeiros passos para configurar o seu ambiente e iniciar o aprendizado na área.

- Conheça os professores e entenda os objetivos do curso
- Mercado de ciência de dados e possibilidades de carreira
- Como configurar seu ambiente de ciência de dados
- Criação de scripts para ciência de dados e publicação de scripts em Streamlit

Atividade

Configurar seu ambiente de ciência de dados integrando a ferramenta Streamlit.

Introdução à ciência de dados

Conheça o mercado de ciência de dados, a experiência dos professores em diferentes setores e os primeiros passos para configurar o seu ambiente e iniciar o aprendizado na área.

- Conheça os professores e entenda os objetivos do curso
- Mercado de ciência de dados e possibilidades de carreira
- Como configurar seu ambiente de ciência de dados
- Criação de scripts para ciência de dados e publicação de scripts em Streamlit

Atividade

Configurar seu ambiente de ciência de dados integrando a ferramenta Streamlit.

Etapas de mineração de dados, Data Mining

Aprenda as principais etapas do CRISP em um projeto prático de compreensão de negócios e dados com preparação e modelagem de dados para avaliação e implantação. Escolha dados de uma das áreas em que deseja trabalhar.

- Visão geral da ciência de dados, classificação de modelos e metodologias existentes
- Estágios de mineração de dados para resolução de problemas
- Processos e metodologias intersetoriais para mineração de dados
- Combinação de modelos: Bagging e floresta aleatória
- Planejamento e modelagem de soluções analíticas

Atividade

Execute e teste seu primeiro modelo de classificação.

Modelos de árvores de decisão em ciência de dados

Conheça o que são árvores de decisão e como elas são usadas para solucionar problemas. Aprenda como realizar tratamento de dados, análise descritiva e técnicas de visualização de dados em ordem para tratar e classificar transações bancárias para detecção de fraude.

- Introdução às Árvores
- Processo padrão de indústria cruzada para mineração de dados (CRISP-DM)
- Python básico
- Tratamento de dados: dados estruturados e não estruturados. Leitura de arquivos estruturados e dataframes
- Análise de dados e resumo com gráficos
- Github para ciência de dados I

Atividades

Desenvolva um modelo de classificação de detecção de doenças utilizando a árvores de decisão.

Árvores de regressão e regressão linear

Saiba como explorar e exibir dados com tabelas e gráficos por meio do Pandas e realizar seleção de variáveis por meio de indicadores. Aprenda regressão básica, melhores práticas de script e como trabalhar com o algoritmo chaid, usar Github e Streamlit para gerenciamento e visualização de projetos de Ciência de Dados.

- Exploração de dados básicos no Pandas: tabelas e Gráficos
- Seleção de variáveis pela primeira vez – indicadores
- Algoritmo de base de árvores de decisão (Chaid) I
- Regressão básica
- Testes, intervalos e previsões, parametrização de dummies
- Github para ciência de dados II
- Melhores práticas de script
- Publicação de relatórios no Streamlit

Atividade

Desenvolva um modelo de regressão e publique-o com o Streamlit usando suas habilidades em Python.

Classificação logística e árvores de regressão

Aprenda a criar novas variáveis, combinar tabelas, aplicar o algoritmo Chaid em situações mais complexas, o básico de regressão logística e mais componentes de gerenciamento de Github e publicação Streamlit. Descubra como usar os recursos de Terminal e Markdown dos notebooks Jupyter.

- Criação de novas variáveis de regressão, mudança de forma e combinação de tabelas
- Algoritmo de base de árvores de decisão (Chaid) II
- Introdução à Regressão Logística
- Componentes Streamlit para publicação de relatórios
- Jupyter Markdown e Terminal
- Github para ciência de dados III

Atividade

Parte I: Desenvolver um projeto de classificação de dados de cartão de crédito.

Combinando diferentes modelos de ciência de dados

Saiba como usar SQL para análise exploratória de ciência de dados, descrever dados com boxplots, gráficos e histogramas e trabalhar com diferentes variações das combinações de algoritmo de Bagging e Boosting. Aprenda a fazer Análise de componentes principais (PCA), conceitos de cálculo e álgebra linear, trabalho com terminal Linux e publicação de relatórios de regressão e clusters em Streamlit.

- Análise Exploratória com SQL
- Descrever dados com boxplots, gráficos e histogramas
- Combinações populares de Bagging (floresta aleatória)
- Combinações populares de Boosting (XGBoosting)
- Conceitos intuitivos de cálculo e álgebra linear
- Terminal Linux I
- Análise de componentes principais (PCA)
- Publique relatórios de regressão logística e clusters no Streamlit

Atividade

Parte II: Desenvolver um projeto de introdução ao reconhecimento de imagens.

Algoritmos não supervisionados

Saiba como aplicar diferentes soluções de regressão para compreender e resolver um problema orientado a dados do início ao fim. Aprenda também a diagnosticar, interpretar um problema e realizar previsões de resultados usando regressão logística e suas equações junto com a tabela de contingência.

- K-means
- Decomposição de valor singular (SVD)
- Incorporação vizinha estocástica com distribuição t (t-SNE)
- Estimativa de densidade de kernel (KDE) e DBScan
- Hierárquicos e Aglomerativos
- Análise fatorial exploratória (EFA)
- Análise de correspondência múltipla (MCA)
- Publicação Streamlit de relatórios de redução de dimensionalidade e relatório final

Atividade

Parte I: Projeto de agrupamento de clientes para a área de marketing.

Regressão linear

Aprenda como escolher variáveis de regressão, detectar e lidar com a multicolinearidade, aplicar inferência e probabilidade a problemas de ciência de dados e usar comandos de terminal Linux mais avançados.

- Seleção de variáveis de regressão, interações
- Probabilidade e inferência
- Conceito de cálculo diferencial por trás de algoritmos DS
- Diagnóstico, multicolinearidade e resolução de problemas
- Interações e diagnósticos em regressão
- Variáveis fictícias (Dummy)

Atividade

Parte II: Projeto de classificação para pontuação de crédito bancário.

SQL para análise de dados

- Introdução à programação em SQL
- Descreva dados com SQL
- Manipule dados com SQL
- SQL Avançado
- Análise Exploratória de Dados com Pandas e Seaborn

Obrigado!



<https://www.linkedin.com/in/pedrobrocaldi/>