Nanodegree Fundamentos de Data Science II



Ementa detalhada do programa

Oferecido por: Udacity do Brasil Ltda.

CNPJ: 24.334.521/0001-43

<https://br.udacity.com/course/fundamentos-data-science-ii--nd111>

# Módulo1: Introdução

# 

Databases, ou bancos de dados, são essenciais em data science. Neste módulo, você aprenderá a utilizar e extrair informações utilizando o banco de dados SQL e firmará seus conhecimentos ao avaliar um banco do tipo de uma loja virtual para tirar conclusões de negócios.

## Aprenda o que é e como utilizar o banco de dados SQL

## Aprenda a limpar, combinar e agregar dados

## Saiba como fazer buscas mais complexas através de subqueries

## Aprenda a fazer comparações entre linhas sem junção adicional e utilizando janelamento (opcional)

## Melhore a performance de seu banco de dados e faça Junções avançadas

Módulo 2: Estatística Prática

Data science precisa de estatística para funcionar. Neste módulo, aprenda conceitos estatísticos diversos, como intervalos de confiança, regra de Bayes e regressão, de forma prática e focada com exercícios de Python e estudos de caso. Fortaleça seus conhecimentos ao analisar resultados de um teste A/B para embasar a tomada de decisão de um negócio.

* Aprenda conceitos de estatística descritiva como medidas de centro e dispersão, formato e outros parâmetros quantitativos
* Aprenda conceitos básicos de probabilidade, distribuição binomial e probabilidade condicional
* Entenda o que é a regra de Bayes, uma das regras mais poderosas da estatística
* Pratique probabilidade com exercícios de Python, Entenda os conceitos básicos de testes de hipóteses e como criar intervalos de confiança
* Aprenda a construir modelos de regressão, regressão linear múltipla e regressão logística usando Python

Módulo 3: Data Wrangling

Dados do mundo real raramente estão limpos, formatados e prontos para uso. É aí que entra o data wrangling. Aprenda a tratar os dados (o que inclui reuni-los, avaliálos e limpá-los) e fortaleça seus conhecimentos avaliando um conjunto de dados de tweets.

* Aprenda quais processos compõe data wrangling
* Entenda como reunir dados de fontes diversas, avaliar sua qualidade e utilizar a biblioteca Pandas para limpá-los
* Pratique seus conhecimentos com exercícios de linguagem Python

Módulo 4: Introdução a Machine Learning

Cientistas de dados utilizam algoritmos de machine learning no dia a dia para criar modelos capazes de prever comportamentos futuros. Adquira conhecimentos iniciais sobre a área, aprenda as principais técnicas de aprendizado de máquina e construa seu primeiro algoritmo!

Aprenda a classificar dados utilizando Naive Bayes e a biblioteca Scikit Learn

* Aprenda o que são máquinas de vetores de suporte (SVM) e árvores de decisão
* Entenda como utilizar regressão linear para mapear dados contínuos
* Descubra o que são outliers, k-means e como ajustar atributos
* Saiba quais são as técnicas de validação de algoritmos e que métricas de avaliação usar
* Aprenda sobre análise de componentes principais para reduzir a dimensionalidade das bases de dados

Módulo 5: Visualização de Dados

Neste módulo opcional, aprenda por que a visualização de dados, os princípios de design e apresentações eficazes são importantes em data science e crie visualizações efetivas com a ferramenta Tableau.

* Aprenda fundamentos de visualização de dados e as diferentes ferramentas que existem
* Saiba como escolher a visualização correta e qual é a relação entre dados
* Crie visualizações de dados utilizando o Tableau