

Datová Akademie

Lekce #1

- uvedení výuky, nastavení očekávání a organizační detaily
- demonstrace workflow datového analytika
- představení projektů a oblastí, na kterých se bude v průběhu kurzu pracovat
- komunikace s API pomocí prohlížeče a Pythonu, stažení dat
- práce se soubory, datové typy - CSV
- základy vizualizace v Tableau

Lekce #2

- praktická ukázka práce s CSV, výhody a nevýhody práce s CSV
- základy práce s databázemi
- úvod do Pandas, praktická demonstrace workflow
- práce s PostgreSQL, pgAdmin, SQLAlchemy
- selekce a filtrace dat z databáze
- praktická vizualizace dat v Tableau
- zadání úkolu na doma

Lekce #3

- navrhování databází, práce s ERD
- typy vztahů v databázi
- pokročilá práce s databázemi
 - datové typy, agregační funkce
 - JOIN (left/right/inner)
 - práce s datem a časem
 - GROUP BY, HAVING
- praktické úkoly v průběhu lekce

Lekce #4

- revize domácího úkolu
- další konstrukce v jazyku SQL
 - vytvoření tabulky, indexy, klíče
- efektivní dotazování v databázi
- práce s daty v Pandas - SELECT, WHERE, GROUP BY, ...
- porovnání zpracování dat z databáze vs. Pandas
- úvod do technologie Git pro práci na projektu

Lekce #5

- absolutní základ vizualizací - best practices
- vysvětlení, proč používáme vizualizace a typy vizualizace
- základy vizuální percepce ve vizualizaci
- arbitrární vs. senzorické vizualizace
- manipulace ve vizualizaci a jak je spozorovat
- praktické příklady

Lekce #6

- pokročilé SQL konstrukce
 - SELECT - N-tiles a window functions
- základy scrapingu, příklad na transparentních účtech
- úvod do Machine Learning
- definice a hranice Machine Learningu
- cheat-sheet k ML

Lekce #7

- tvorba prvního modelu v Machine Learningu
- lineární regrese - intuice, příprava dat, normalizace, model
- logistická regrese - intuice, model
- distribuované funkce, confidence intervaly
- Gradient descent
- začátek práce na projektech

Lekce #8

- práce s časem - time series
- práce s modulem matplotlib na vizualizaci v Pythonu
- trend a seasonality
- funkce moving average
- predikce (SARIMA)
- práce na projektu

Lekce #9

- práce v nástroji BigML
 - decision trees
- evaluace modelů
- vysvětlení confusion matrix
- úvod do neuronových sítí a deep learning
- práce na projektu

Lekce #10

- úvod do nástroje dask na práci s datami
- paralelní zpracování dat
- práce s Numpy a Pandas v Dasku
- vizualizace a grafy
- úvod do Hadoop pro práci s Big Data
- práce na projektu

Lekce #11

- NoSQL databáze - úvod
- praktický use case v MongoDB
- technologie Docker, nasazení projektu v dockeru
- další oblasti relevantní pro datového analytika v systémové administraci
- práce na projektu

Lekce #12

- prezentace projektů
- poskytnutí zpětné vazby na projekty
- best practices pro datovou analýzu
- příprava na interview na pozici datového analytika
- business case na základě reálných dat
- další materiály na studium