



Data Science Community (DSC)

Študentský klub pôsobiaci pri Ekonomickej Univerzite v Bratislave

Harmonogram výučby:	Zimný Semester 2019/2020
Stretnutia:	Každý utorok 17:00-18:30
Miesto:	EUBA (učebňa D117)
Čo potrebujete:	Vlastný notebook s JUPYTER-om

Prihlášky a otázky, prosím, smerujte na:

<https://docs.google.com/forms/d/1cB3eOnQRUTU10bG7c5M45lQ5y0yC0ynZKn4a4kuNfjE/>

0. Kick-Off Semestra (24. Sep. 2019)

- Stručne o tom, ako bude vyzerat' semester a rôzne organizačné záležitosti.
- Prečo práve Python, v čom je jeho sila, aké nástroje ponúka a jeho aplikácia v praxi.
- Úvod do programovacieho prostredia JUPYTER notebook.

Prednáška: **What is cognitive?**

- Speaker: **Ján Sehnal**, Head of Innovations at ERSTE GROUP IT, alumni NHF.
- Súčasťou prvej úvodnej hodiny bude aj prednáška na tému umelej inteligencie a jej špecifik, či budúcnosti.
- Ako banky reagujú na technológie, ktoré menia podnikateľské prostredie.
- Q&A

1. Introduction to Python (1. Okt. 2019)

- Lektor: **Michael Strauß**, Business Architect at ERSTE GROUP BANK, štúdium: Business Informatics at JKU Linz, Rakúsko (prednáška v **ENG**).
- Práca s pythonom v jupyter notebook-u a základný syntax.
- Tvorba premenných a následné operácie s nimi.
- Deklarácia a využitie numerických dát: int, float.
- Formátovanie string dát a operácie s nimi.
- Zoznamy: list, tuple, dictionary, set. Funkcie s nimi spojené.

2. Python Workflow Control (8. Okt. 2019)

- Lektor: **Michael Strauß**, Business Architect at ERSTE GROUP BANK, štúdium: Business Informatics at JKU Linz, Rakúsko (prednáška v **ENG**).
- Opakované výpočty: využitie cyklov FOR a WHILE.
- Využitie podmienok v bloku: IF, ELIF, ELSE a vnorený IF.

Prednáška: **Importance of Problem Solving in Business.**

- Speaker: **Michaela Albrecht**, Head of Enterprise Architecture at ERSTE GROUP BANK, alumni University of Vienna
- Ako uchopiť problém a ako ho rozbiť ho na malé časti, tak aby sa dal postupne riešiť.
- Aké zručnosti sú nevyhnutné na prepojenie sveta financií a technológií a či je na to nevyhnutné ovládať všetky programovacie jazyky sveta.
- Aká je cesta od písania básni až po vedenie IT oddelenia v rámci najväčšej bankovej skupiny v strednej a východnej Európe.
- Q&A

3. NumPy – Arrays (15. Okt. 2019)

- Lektor: **Filip Dzuroška**, Big Data Science Lab Team, ERSTE GROUP IT.



NumPy:

- Arrays (ako volať polia a ako iterovať polia). Funkcie spojené s týmto typom dátového formátu.
- Shaping, reshaping.
- Výpočtové operácie s využitím Arrays. Univerzálne funkcie. Broadcasting.
- Indexing, Fancy indexing, Combined indexing, Masking,
- Sorting

5. Pandas – Data Frame (22. Okt. 2019)

- Lektor: **Filip Dzuroška**, Big Data Science Lab Team, ERSTE GROUP IT.



Pandas:

- Úvod do dátových typov dostupných v tejto knižnici: Series, DataFrame.
- Index, reindex operácie.
- Ako dostať dáta z tabuľky – loc[], iloc[].
- Základne funkcie: split-apply-combine-merge aplikované na DataFrames.
- NaN hodnoty a práca s nimi.
- Aggregation, pivoting, grouping a iné operácie na sumárnu prácu s dátami.
- Summary statistics, simple a rolling statistic a ako pomôžu lepšie porozumieť dátam.

5. Exploratívna Dátová Analýza a Machine Learning. (29. Okt. 2019)

- Lektor: **Matúš Cimermann**, AI & ML engineer at EXPONEA, alumni FIIT STU.
- Data Wrangling: Príprava dát a ich modifikácia, ktorá jednoduchší prácu s nimi.
- Data Visualization: Rôzne typy grafov a plotov. Python knižnice: Matplotlib (základná vizualizácia) a Seaborn (štatistické problémy).
- Exploratory Data Analysis: Hľadanie súvislosti v dátach (Titanic dataset).
- Machine Learning: Vytvorenie modelu, ktorý sa učí z dát bez explicitného definovania pravidiel. Model bude predikovať jedincov, ktorí majú najväčšiu pravdepodobnosť prežitia. Aplikujeme pár pomerne základných techník machine learningu.



Zadanie pre študentov: Semestrálny projekt

- Prezentácia výsledkov v 10. týždni.
- Je možné pracovať sám alebo sa spojiť do tímov.
- Možnosť vybrať si tému podľa vlastnej chuti. Odporúčaný je štruktúrovaný dataset z FED.com/OECD.org/KAGGLE.com/DATA.gov/sdw.ECB.europa.eu
- Úprava dát, data munging, vytvoriť pár grafov, štatistická analýza, socio-ekonomický rozmer.

6. Modelovanie kreditného rizika : IFRS9 prípadová štúdia (5. Nov. 2019)

- Lektor: **Marko Vateha** (EUBA, 5. ročník).
- TtC and PiT model na predikciu zlyhávania korporátnych klientov.
- Input: migračná matica klientov v rámci ratingových tried a makroekonomické premenné.



7. Modelovanie kreditného rizika : IFRS9 prípadová štúdia (12. Nov. 2019)

- Lektor: **Marko Vateha** (EUBA, 5th grade).
- Dokončenie modelu kreditného rizika.



8. Časové rady – Predikcia cien akcií (19. Nov. 2019)

- Lektor: **Juraj Krivda**, ERSTE GROUP IT, alumni FHI.
- Načítavať údaje o vývoji cien akcií napr. Apple, Microsoft, Google, Amazon.
- Plot grafov kde uvidíme ako sa ceny vyvíjali a analýza korelácie medzi firmami.
- Analyzovať denné zmeny cien a zobrazíť ich graficky, plus histogram ich frekvencie.
- Vyvodíť riziko z investovania na základe ich denných zmien.
- Monte Carlo metóda simulácie na predikciu budúcej ceny akcií.
- ARIMA model pre časové rady.



9. Príbehy ukryté za dátami a číslami (26. Nov. 2019)

- Lektorka: **Michaela Lacko-Vrabcová**, Enablement Manager v EMARK, štúdium UK v BA.
- Na workshope si vyskúšaš prácu s analytickým nástrojom Qlik Sense, ktorý patrí medzi najlepšie BI nástroje súčasnosti.
- Naučíš sa vizualizovať dáta tak, aby boli rýchlo pochopiteľné a osvojíš si prácu s programom QS na úrovni tvorby a základnej optimalizácie dátového modelu.



10. Prezentácia projektov (3. Dec. 2019)

- V závislosti ako budú skúšky a predtermíny na škole, bude buď iba odovzdanie projektu alebo aj prezentácia.
- Plánovaná je však krátka prezentácia projektov a vlastných výstupov z analýz pred celou skupinou študentov.



Študentská spätná väzba: [Feedback na DSC aktivity](#)

- Zhodnotenie semestra, návrhy na zlepšenie a ďalšie kroky klubu.

[Odporúčaná literatúra](#)

1. Python Data Science Handbook: Essential Tools for Working with Data, Author: Jake VanderPlas, 2016 (available On-line)
2. Hands-On Machine Learning with Scikit-Learn and TensorFlow Concepts, Tools, and Techniques to Build Intelligent Systems, Author: Aurélien Géron, 2017

