01. Introduction to 3ikakke's course on data science with python

3ikakke

2022-02-21

Outline

- Learning Objectives
- Course Overview
- Understanding the data science domain

- Tools/skills in data science
- What you will learn
- What we won't focus on
- Who will benefit from this course
- Personal portfolio/project
- House keeping
- Q&A

Learning Objectives

- Getting to know the students
  - What do you do?
  - What do you hope to learn?
  - One fun fact about yourself
- Setting expectations

Overview

- What is data?
- What is data science?
- Real life examples
- Careers in data science

Understanding the domain

- **Data collection**
- **Data analysis** - exploratory, descriptive/diagnostic analysis
- **Data engineering** - predictive/prescriptive analysis
- **Data science** - Advanced including deep learning, complex algorithms, leading a data science team

Tools/Skills

- **Excel/Spreadsheet application**
- **SQL** (Structured query language)
- **Data visualization software**
  - PowerBI
  - Tableau
  - **Metabase**
- **Statistical software**
  - SAS
  - Stata
  - SPSS
- **Programing languages**
  - Python
  - R
  - Julia

What you will learn

- Python - core language construct
  - `print('Hello world')`
- Basic statistics
  - $$\sigma = \sqrt{\frac{1}{n-1}\sum(x-\bar{x})^2}$$
- Machine Learning
  - Regression, Classification, Random Forests, Principal Component Analysis, Ensemble, Natural Language Processing
- Deep Learning
  - Artificial Neural Networks, Convolutional Neural Networks

What we won't focus on

- Excel (*functions, data filtering, pivot tables, x-lookup, macros*)
- SQL (*SQLAlchemy*)
- Visualization software (*Metabase*)
- Statistical packages

Who will benefit from this course?

- First timers
- Career people

- Academics/researchers
- People giving data science a second chance
- What will you benefit?

Personal Portfolio/Project

- Everyone will be expected to create a GitHub account
- Come up with a project of interest
- Keep your reusable code stored in private git repositories
- Keep your portfolio in public repositories

Housekeeping

- Class days
- Start times
- Duration
- Recording of classes
- **Our tools**
  - Python
  - Jupyter Notebook - great IDE! for displaying code while teaching
  - Spyder - optional IDE
  - GitHub - for your codes and projects along with GitHub gists
  - Slack - for interaction and video recordings
  - Kaggle - primary data source
  - Visual Studio Code

Housekeeping (contd)

- There will be no graded homework
- There will be simple tasks at the end of every other class

Gist for the day

1. 01.Setting.Up

Q&A

Thank for signing up!