**Course: Practical Data Science**

Goal: Be able to pose technological, business questions that can be solved with big data; collect necessary data, process and analyze it and provide actionable insights using this data.

Prerequisites: working knowledge of any coding language, basic math, statistics.

**Module 1: Data Science Introduction**

* Introduction to Data Science: What Data Science can do for you?
* Goals, methods, structure of Data Science.
* Characteristics and technologies for Big Data.
* Install environment (Anaconda distribution for different operation systems (Windows, Mac OS, Linux)).
* Employ Github to organize your projects.
* Python programming language.
* Comments, indentation, modules.
* Basic data types, type conversion.
* Collections (lists, tuples, sets, dictionaries).
* Conditional operations.
* Loops and looping through the elements of different collections.
* Errors and exceptions.
* Built-in, User-defined and inline functions.
* Working with files.

Goal: Be able to organize DS projects, set up Python environment with necessary modules, run Python, iPython, write basic Python programs.

**Week 2: Collecting formats**

* Types of data (structured and unstructured).
* Structured data formats (CSV, JSON, XML).
* Graphics formats.
* Audio formats.
* Video formats.
* Parsing unstructured data with regular expressions in Python.
* Parsing HTML with Beautiful Soup.
* Web crawling (HTTP requests, URLLIB).
* API (REST, SOAP), API Security and Rate Limits.
* Relational databases, terminology, structure, systems.
* Designing RDB (ER Diagram, Relational Schema)
* Key-value pairs, Relationships.
* SQL syntax (create, insert, delete, update tables)
* Retrieve records with SQL.
* Join tables with SQL.
* SQL with Python (on example of SQLite).
* Places to find data.

Goal: Be able to find and collect data in readable format for further processing

**Week 3: Processing and analyzing data**

* Numpy module.
* Operations with Numpy arrays.
* Matrix and vector operations in Numpy.
* Pandas module.
* Series (creation, query, functions).
* DataFrames (creation, query, functions)
* Continuous, discrete, categorical data.
* Time stamps.
* Dealing with missing values.
* Join DataFrames.
* GroupBy, Aggregate and Pivot Tables in Pandas.
* Statistics in Pandas.
* Visualization with DataFrames.
* Descriptive statistics with Scipy module.
* Distributions.
* Hypothesis testing.
* A/B testing.
* Outliers and inliers. Anomaly detection in static data and time series.
* Data science pipeline (find, collect, store, verify, clean, transform, combine, analyze, visualize, present)

Goal: Be able to clean, manipulate data and provide actionable insights from the data.

**Week 4: Data visualization and presentation**

* Visualization wheel.
* Effect of the audience.
* Qualities of great visualization (truthful, functionality, beauty, insightful, enlightening)
* Matplotlib.
* Scatter plots, line plots, box plots, barcharts, histograms, heatmaps.
* Animation with matplotlib.
* Interactivity with matplotlib.
* Seaborn.
* Interactive infographics with d3.js
* Business dashboards with Tableau.
* Where to learn more and how to stay current in Data Science.

Goal: Be able to present data analysis and actionable insights for managers, stakeholders, etc.