"Advanced Python Programming for Everybody"

Instructor: Ernest Bonat, Ph.D.
Senior Software Engineer
Senior Data Scientist
ebonat@15itresources.com
Cell: 503.730.4556

Module 4 Source Code

https://github.com/ebonat/intel_module_4

Hillsboro Python Machine Learning Meetup

https://www.meetup.com/Hillsboro-Python-Machine-Learning-Meetup/

Title: "Advanced XGBoost Algorithm for Data Classification and Regression"

• Tuesday, May 29, 2018

6:00 PM to 9:00 PM

• DAT Solutions

8405 SW Nimbus Avenue, Beaverton, OR (edit map)

• Meeting Agenda:

6:00 – 6:40 pm: Pizza, water and networking.

6:40 – 6:45 pm: Welcome message by Ernest Bonat, Ph.D.

6:45 – 8:15 pm: Presentation and open discussions. ““Advanced XGBoost Algorithm for Data Classification and Regression”” by Ernest Bonat, Ph.D., Senior Data Scientist.

Module 4. Extract-Transform-Load (ETL) System Design and Development

<u>Extract-Transform-Load</u> (ETL) – process of extracting data from source systems and bringing it into other source systems.

1. <u>Extract</u> – the desired data is identified and extracted from many different sources, including **database** systems and applications (MySQL, SQLite, MongoDB, Oracle, SQL Server, JSON, CSV, XML, etc.)

2. <u>Transform</u> – based on business requirements the data is transform (**pre-processing or clean up**) in another format or structure (in Data Science this transform is done using Python **pandas** library – very important!)

3. <u>Load</u> – the data is loaded into the final target source (database, files (JSON, CSV and XML), etc.)

Exercise 1

Create a simple Python program to generate a JSON "Publisher Report" file from `magazine`.`publisher` MySQL table.