

Hillsboro Python Machine Learning Meetup

May/2017

Ernest Bonat, Ph.D.

Senior Software Engineer

Senior Data Scientist

DAT Wi-Fi

Username: DAT Guest

Password: beaverton dat

- 6:00 – 6:40 pm: Pizza, **water only** and networking.
- 6:40 – 6:45 pm: Welcome message by Ernest Bonat, Ph.D.
- 6:45 – 8:00 pm: Presentation and open discussions.
- 8.00 pm – 9.00 pm: Coding and learning session. Bring your Python development laptop!

Why did I create this meetup?

1. Bad traffic to Portland downtown.
2. Vert hard to find a parking lot.
3. Bad Python presentation code.
4. No time at all to review the presentation and learn something after the meeting.

We need your support:

1. Need 1 Senior Python Developers for presentation and code review every month (Co-organizers, 4-6 hours a month).
2. Email Ernest at ebonat@15itresources.com

Our Meetup Mission:

1. *“Come, Listen, Code and Learn”.*
2. Finding and presenting best practices of Machine Learning using Python Data Stack.
3. Create great networking place for Hillsboro-Beaverton Data Scientists.

Today Presentation

“Advanced
Extract-Transform-Load (ETL)
Systems Design using Python Data Stack”

Fun Question Today:

What kind of dog breed is this?



The answer is?

How Many Breeds of Dogs Are There in the World today?

Around 340 breeds!

Do you need to know all of them? **NO! WHY?**

Email me: **ernest.bonat@gmail.com**

Most popular Open Source Relational Database: **MySQL or PostgreSQL or SQLite?**

Most popular Open Source NoSQL Database: **MongoDB or Casandra?**

We'll be covering later **MongoDB/PyMongo** – a lot of information to cover to!

DB-Engines Ranking - Trend Popularity (https://db-engines.com/en/ranking_trend)

MySQL Database APIs Packages:

1. MySQLdb (<https://pypi.python.org/pypi/MySQL-python/>)

2. **SQLAlchemy for ORM** (<http://www.sqlalchemy.org/>)

3. PyMySQL (<https://github.com/PyMySQL/PyMySQL>)

4. **MySQL Connector/Python**
(<https://dev.mysql.com/downloads/connector/python/>)

MySQL CRUD Operations using Connector/Python

1. Create
2. Read (Select)
3. Update
4. Delete

Extract-Transform-Load (ETL) – process of extracting data from source systems and bringing it into another source systems.

1. Extract – the desired data is identified and extracted from many different sources, including **database** systems and applications (MySQL, SQLite, Oracle, SQL Server, **JSON**, CSV, XML, TXT, etc.)
2. Transform – based on business requirements the data is transform (**pre-processing**) in another format or structure (in Data Science the transform is done using **pandas** library – very important!)
3. Load – the data is loaded into the final target source (database, files (**JSON**, CSV, XML, TXT), etc.)

There is in general four types of writing Python code for Data Science today:

- 1.Top-bottom Code (Jupyter Notebook!)
- 2.Procedures Language Code
- 3.Object-Oriented Programming (OOP)
- 4.MVC Design Pattern

Training

“Business Statistics Course for Python Programmers”

(<http://15itresources.com/training/>)

What do we do different?

1. Corporate/in-person/hands-on training.
2. Direct business data analysis for your company needs.
3. Own and use the Ernest's Python Data Science libraries which offer full proof programming recipes.

Presentation Source Code

(https://github.com/ebonat/hillsboro_machine_learning_05_2017)