

# **“Advanced Python Programming for Machine Learning Projects”**

**Instructor: Ernest Bonat, Ph.D.**

**Senior Software Engineer**

**Senior Data Scientist**

**[ebonat@15itresources.com](mailto:ebonat@15itresources.com)**

**Cell: 503.730.4556**

## 2. Data Visualization and Pre-processing 2

GitHub: [https://github.com/ebonat/intel\\_session\\_3](https://github.com/ebonat/intel_session_3)

### Data Science Two Main Tasks:

1	<b>Data Pre-processing (Cleansing)</b>	<b>60% - 70% work</b>
2	<b>Data Analytics</b>	<b>40% - 30% work</b>

**Data Pre-processing very important task. Be careful with “Garbage IN – Garbage OUT”**

1. Create a new Python environment folder name `intel_data_preprocessing_2`
2. Install latest Python, numpy, pandas and scikit-learn libraries
3. Create a Python project `intel_data_preprocessing_2`
4. Create a Python file `version.py` to print the versions
5. Create a Python file `data_preprocessing_2.py` file. Copy the following code:

```
import sys  
  
import pandas as pd  
  
import numpy as np  
  
import sklearn  
  
import scipy
```

```
if __name__ == '__main__':  
    print("Python: {}".format(sys.version))  
    print("NumPy: {}".format(np.__version__))  
    print("pandas: {}".format(pd.__version__))  
    print("SciPy: {}".format(scipy.__version__))  
    print("scikit-learn: {}".format(sklearn.__version__))  
  
    email = np.array(["ernest.bonat@gmail.com"])  
    df_data = pd.DataFrame(data=email, columns=["email"])  
    print(df_data)
```

Follow CHRIS ALBON site (<https://chrisalbon.com/>) sessions:

1.Preprocessing Structured Data

2.Data Wrangling

**Follow the Instructor and let's write good Python code!**