COLLECTING, ANALYZING, AND VISUALIZING DATA WITH PYTHON PART II

DR. MICHAEL FIRE

Turi Create - SFrame

Pandas
- In-memory data structure - cannot work (usually) with DataFrames bigger than the RAM
- Usually use a single thread
- Extensive functionality
- Well maintained
- Many tutorials

TuriCreate
- Out-of-core data structure (uses both disk and memory)
- Multithreaded
- **Can scale to big data**
- Uses similar syntax to DataFrame
- Easy to understand and uses *Group By* function
- Well maintained (by Apple)
- Part of the Turi Create Platform

Turi Create in Practice

I used SFrame to:

- [Analyze Reddit dataset of with data of over 10 years](#)
- [Analyze Microsoft Academic Dataset with 120 million papers](#)
- Analyze Amazon Reviews Dataset
- and more

It is a great easy-to-use tool to work with Big Data.

Let's move to running code using Jupyter Notebook

Recommended Reading

- [TuriCreate Userguide](#)
- [TuriCreate API - turicreate.SFrame](#)
- [A Guide to Turi Create - WWDC 2018 - Videos - Apple Developer](#)