Data Science Workshop:

Instructor: Robby Grodin [robby@toypig.co](mailto:robby@toypig.co)

Python: high-level generic scripting language (interpretive programming)

Pandas: open-source module to use data frames

Scikitlearn: statistical and machine learning algorithms

Matplotlib: graphic package

Anaconda: Environment for Python

Data frame = set of series that share an index

Neural Networks implement Collaborative Filtering (this is what Netflix does)

Isolating features -> Building Models

Cross Validation Analysis: How to validate model works?

Splits data into training and test sets

Classification Algorithms:

K Neighbors Classification: clusters based on a system of distance-based weighting

Requires tuning while searching for optimal value of K = number of clusters

Can be visualized as a Voronoi Diagrams