

index

November 4, 2022

CSE 817: Data Engineering

0.1 Course Contents

- **Introduction to Data Science**
 - Defining Data Science
 - * What is Data Science?
 - * Fundamentals of Data Science
 - * The Many Path of Data Science
 - What Do Data Scientists Do?
 - * A day in the Life of a Data Scientist
 - * Old Problems, new problems, Data Science solutions
 - * Data Science Topics and Algorithms
 - * Cloud for Data Science
 - Big Data and Data Mining
 - * Foundations of Big Data
 - * What is Hadoop?
 - * How Big Data is Driving Digital Transformation
 - * Data Mining
 - * Quiz: Data Mining
 - Deep Learning and Machine Learning
 - * What is the difference?
 - * Neural Networks and Deep Learning
 - * Applications of Machine Learning
 - * Regression
 - * Quiz: Regression
 - Data Science in Business
 - * How Data Science is saving lives
 - * Applications of Data Science
- **Python for Data Science**
 - Fundamentals of Data Manipulation with Python
 - * Python Functions
 - * Python Types and Sequences
 - * Python Demonstration: Reading and Writing CSV Files
 - * Python Dates and Times
 - * Python: Map, Lambda, and List Comprehensions
 - * Numerical Python Library (numpy)
 - * Manipulating Text with Regular Expression

- * Programming Assignment: 1
- Basic Data Processing with Pandas
 - * Introduction to Pandas
 - * The Series Data Structure
 - * Querying a Series
 - * DataFrame Data Structure
 - * DataFrame Indexing and Loading
 - * Querying a DataFrame
 - * Indexing Dataframes
 - * Missing Values
 - * Example: Manipulating DataFrame
- More Data Processing with Pandas
 - * Merging Dataframes
 - * Pandas Idioms
 - * Group By
 - * Scales
 - * Pivot Table
 - * Date/Time Functionality
 - * Programming Assignment: 2
- Answering Questions with Messy Data
 - * Basic Statistical Testing
 - * Other Forms of Structured Data
 - * **Reading** – The 5 Graph Algorithms that you should know
- **Plotting, Charting & Data Representation in Python-**
 - Principles of Information Visualization
 - Basic Charting
 - * Matplotlib
 - * Scatterplots
 - * Line Plots
 - * Bar Charts
 - Charting Fundamentals
 - * Subplots
 - * Histograms
 - * Box Plots
 - * Heatmaps
 - * Animation
 - * Interactivity
 - * Plotting with Pandas
- **Social Network Analysis in Python-**
 - Introduction to NetworkX
 - Network Connectivity
 - Influence Measure and Network Centralization
 - Network Evolution

Optional

- ****Text Mining in Python****-
 - Working with Text in Python
 - Basic Natural Language Processing

- Classification of Text
- Topic Modeling
- ****Machine Learning in Python****-
 - Fundamentals of Machine Learning
 - Introduction to SciKit Learn
 - Supervised Machine Learning
 - Evaluation
 - Unsupervised Machine Learning

0.2 Textbooks

- [PythnDS] Wes McKinney **Python for Data Analysis** *Data Wrangling with pandas, NumPy & Jupyter*, O'Reilly.

0.3 Environment Setup:

- Python 3 and Jupyter Installation - [Python 3 Installation & Setup Guide](#) - [Anaconda Installation](#) - [Jupyter Installation Guide](#)

0.4 Python Tutorials

- Part 1: [Slides](#), [Notebook](#), [HTML](#)
- Part 2: [Slides](#), [Notebook](#), [HTML](#)
- Part 3: [Slides](#), [Notebook](#), [HTML](#)
- List in Python: [Notebook](#), [HTML](#)