



DATA SCIENCE IMMERSIVE SYLLABUS

COURSE OVERVIEW

By the end of this course, students will be able to:

- › Collect, extract, query, clean, and aggregate data for analysis
- › Perform visual and statistical analysis on data using Python and its associated libraries and tools
- › Build, implement, and evaluate data science problems using appropriate machine learning models and algorithms
- › Use appropriate data visualization to communicate findings
- › Create clear and reproducible reports to stakeholders
- › Identify big data problems and articulate how distributed systems and parallel computing technologies are solving these challenges
- › Apply question, modeling, and validation problem-solving processes to datasets from various industries in order to provide insight into real-world problems and solutions

UNIT 1: FUNDAMENTALS

Project 1	<i>Students will apply their skills in NumPy and Python in order to answer several questions provided from a clean dataset.</i>
------------------	---

Python & NumPy	<i>Demonstrate introductory programming concepts using Python and NumPy as a tool to navigate data sources and collections</i>
----------------	--

UNIX	<i>Utilize UNIX commands to navigate file systems and modify files</i>
------	--

git	<i>Learn to keep track of changes and iterations using git version control from your terminal</i>
-----	---

Descriptive Statistics	<i>Define and apply descriptive statistical fundamentals to sample datasets</i>
------------------------	---

Intro to Plotting and Visualization	<i>Practice plotting and visualizing data using Python libraries like matplotlib and Seaborn</i>
-------------------------------------	--

UNIT 2: EXPLORATORY DATA ANALYSIS

Project 2	<i>Students will use Pandas to apply advanced NumPy and Python skills in order to clean, analyze, and test data from multiple messy datasets</i>
------------------	--

Experiment Design	<i>Plan experimental study design with a well thought out problem statement and data framework</i>
-------------------	--

Pandas & Pivot Tables	<i>Use Pandas to read, clean, parse, and plot data using functions such as boolean, indexing, math series, joins, and others</i>
-----------------------	--

SciPy & Statsmodels	<i>Review statistical testing concepts (p-values, confidence intervals, lambda functions, correlation/causation) with SciPy and Statsmodels</i>
---------------------	---

Web Scraping	<i>Learn to scrape website data using popular scraping tools</i>
--------------	--

Bootstrapping	<i>Practice resampling and building inferences about your data</i>
---------------	--



DATA SCIENCE IMMERSIVE SYLLABUS

UNIT 3: CLASSICAL STATISTICAL MODELING

Project 3	<i>Using a provided dataset, students will explore, clean, and model data, outlining their strategy and explaining their results.</i>
Linear &, logistic regression	<i>Use scikit learn and statsmodels to run linear and logistic regression models and learn to evaluate model fit</i>
Bias-Variance Tradeoff	<i>Articulate the bias-variance trade-off as you practice evaluating classical statistical models</i>
Gradient Descent	<i>Dive into the math and theory of how gradient descent helps to optimize loss function for regression models</i>
Feature Selection	<i>Use feature selection to deepen your knowledge of study design and model evaluation</i>
Regularization & Optimization	<i>Learn to apply regularization and optimization when evaluating model fit</i>
K-Nearest Neighbors	<i>Begin to look at classification models through an application of the kNN algorithm</i>

UNIT 4: MACHINE LEARNING MODELS

Project 4	<i>Students will scrape and model their own data using multiple methods, outlining their approach and evaluating any risks or limitations.</i>
Clustering	<i>Define clustering and its advantages and disadvantages from classification models</i>
Ensemble Models	<i>Build and evaluate ensemble models, using decision trees, random forests, bagging, and boosting</i>
NLP	<i>Get introduced to natural language processing through sentiment analysis of scraped website data.</i>
Naive Bayes	<i>Learn how Naive Bayes can simplify the process of analyzing data for supervised learning algorithms</i>
Hadoop & MapReduce	<i>Get introduced to the history and use of Hadoop as well as the advantages and disadvantages of using parallel or distributed systems to store, access, and analyze big data</i>
Hive & Spark	<i>Gain an introductory understanding of how Hive interacts with Hadoop and learn about Spark's advantages through big data case studies</i>
Time Series Analysis	<i>Analyze and model time series data using the ARIMA model in Pandas</i>

UNIT 5: ADVANCED TOPICS & TRENDS

Capstone Project	<i>Students will choose a dataset to explore and model, providing detailed notebook of their technical approach and a public presentation on their findings.</i>
Neural Network Basics	<i>Learn the differences between different types of neural networks and demonstrate how they are fit with backpropagation.</i>
Recommender Systems	<i>Build and apply basic recommender systems in order to predict on sample user data</i>
Multi-Armed Bandit	<i>Practice split testing with Bayesian multi arm bandits</i>
Portfolio Development	<i>Work with career coaches to create and polish your portfolio for employers</i>
Interview Prep	<i>Practice data science case studies to prep for job interviews</i>