



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 tools to communicate findings
- › Create clear and reproducible reports to stakeholders
- › Identify big data problems and understand 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

WEEK 1: MATH, & PROGRAMMING 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>Maintain a git repository in order to keep track of changes and iterations as your project evolves.</i>
-----	--

Descriptive Statistics	<i>Define and apply descriptive statistic fundamentals</i>
------------------------	--

Intro to Plotting and Visualization	<i>Practice using plot.ly, iPython notebook and Tableau to plot and visualize data</i>
-------------------------------------	--

WEEK 2: EDA, PANDAS & SCIPY

Project 2	<i>Students will use Pandas to apply advanced NumPy and Python skills in order to acquire, parse, mine, and refine data from a larger, messier dataset</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	<i>Review statistical testing concepts (p-values, confidence intervals, lambda functions, correlation/causation) with SciPy</i>
-------	---



DATA SCIENCE IMMERSIVE SYLLABUS

WEEK 3: LINEAR REGRESSIONS, SCI-KIT-LEARN, GRADIENT DESCENT, & MODEL FIT

Project 3	<i>Using a provided dataset, students will create a Linear Regression model and an executive summary writeup of their findings.</i>
Linear regression, stats models, and scikit learn	<i>Use scikit learn and statsmodels to run linear regression models and evaluate model fit</i>
Bias-Variance Tradeoff	<i>Understand bias-variance trade-off to evaluate machine learning models</i>
Gradient Descent	<i>Look behind the hood at the math and theory of how gradient descent helps to optimize the loss function for models</i>
Regularization & Optimization	<i>Learn to apply regularization and optimization when evaluating model fit</i>

WEEK 4: CLASSIFICATION, KNN, AND WEB SCRAPING

Project 4	<i>Students will build a logistic regression model while applying all of the skills learned during Weeks 1-3.</i>
Web Scraping	<i>Learn to scrape data using popular scraping tools</i>
kNN	<i>Begin to look at classification models through an application of the kNN algorithm</i>
Data Visualization	<i>Dive deeper into Tableau and plot.ly to create more robust plots and visualizations</i>
Intro to Capstone Project	<i>Get introduced to the final project. Begin to plan your experiment design and look for data sources</i>

WEEK 5: SQL, DATABASES, & LOGISTIC REGRESSIONS

Project 5	<i>Students will build a logistic regression classification model using data pulled from a remote database.</i>
SQL & Remote Databases	<i>Get introduced to different types of databases, review SQL commands, and practice connecting to and pulling data from a remote database</i>
Feature Selection	<i>Use feature selection to deepen knowledge of model evaluation</i>
Logistic Regressions	<i>Build, evaluate, and refine a logistic regression model for a given business case study</i>



DATA SCIENCE IMMERSIVE SYLLABUS

WEEK 6: APIS, TREES & ENSEMBLE METHODS

Project 6	<i>Students will build a decision tree and a random forest model using data pulled from a remote database.</i>
JSON & APIs	<i>Learn to pull JSON data from APIs as another potential data source</i>
Ensemble Models	<i>Build and evaluate ensemble models, using decision trees, random forests, bagging, and boosting</i>

WEEK 7: PCA, CLUSTERING, K-MEANS & AWS

Project 7	<i>Students will apply skills from Weeks 1-6 while building a PostgreSQL database and performing a PCA (“principal components analysis”).</i>
Clustering	<i>Define clustering and it’s advantages and disadvantages from classification models</i>
K-Means	<i>Practice building and evaluating a k-means algorithm</i>
PCA	<i>Convert a set of observations or variables into principal components in order to improve predictive analysis.</i>
Intro to AWS	<i>Learn to access and deploy databases to AWS</i>

WEEK 8: BAYESIAN INFERENCE & LDA

Capstone Project, Part 1	<i>During the first part of the capstone project, students will select and verify topics and datasets, in order to develop problem statements, goals, and success criteria.</i>
Bayesian Methods	<i>Build a linear regression model with bayesian methods</i>
LDA	<i>Refine data using latent dirichlet allocation (LDA)</i>



DATA SCIENCE IMMERSIVE SYLLABUS

WEEK 9: WORKING WITH TIME SERIES

Capstone Project, Part 2	<i>During the second part of the capstone project, students will create their own database, acquire and clean their data, and perform exploratory data analysis using real world tools and processes.</i>
Time Series & Autocorrelation	<i>Analyze and visualize time series data using Pandas, plot.ly and Tableau</i>
ARIMA Model	<i>Use the ARIMA model to make predictions with time series data</i>

WEEK 10: SVMS, NAIVE BAYES & INTRO TO BIG DATA

Capstone Project, Part 3	<i>During the third part of the capstone project, students will build and tune a predictive model from their data using machine learning algorithms to analyze, classify, and plot their data.</i>
SVMS & Naive Bayes	<i>Learn how concepts like SVMS and Naive Bayes can simplify the process of analyzing data for supervised learning algorithms</i>
Hadoop, Hive, Spark	<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>

WEEK 11 - 12: JOB PREP & CAPSTONE

Capstone Project, Part 4 & 5	<i>Students will create a detailed technical model notebook to share with peer audiences, while adapting their findings into a visual narrative for non-technical stakeholders.</i>
Interview Prep	<i>Practice data science case studies to prep for job interviews</i>