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	*Students will apply their skills in NumPy and Python in order to answer several questions provided from a clean dataset.*
Python & NumPy	*Demonstrate introductory programming concepts using Python and NumPy as a tool to navigate data sources and collections*
UNIX	*Utilize UNIX commands to navigate file systems and modify files*
git	*Maintain a git repository in order to keep track of changes and iterations as your project evolves.*
Descriptive Statistics	*Define and apply descriptive statistic fundamentals*
Intro to Plotting and Visualization	*Practice using plot.ly, iPython notebook and Tableau to plot and visualize data*

WEEK 2: EDA, PANDAS & SCIPY

Project 2	*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*
Experiment Design	*Plan experimental study design with a well thought out problem statement and data framework*
Pandas & Pivot Tables	*Use Pandas to read, clean, parse, and plot data using functions such as boolean, indexing, math series, joins, and others*
SciPy	*Review statistical testing concepts (p-values, confidence intervals, lambda functions, correlation/causation) with SciPy*

DATA SCIENCE IMMERSIVE SYLLABUS

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

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

WEEK 4: CLASSIFICATION, KNN, AND WEB SCRAPING

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

WEEK 5: SQL, DATABASES, & LOGISTIC REGRESSIONS

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

DATA SCIENCE IMMERSIVE SYLLABUS

WEEK 6: APIS, TREES & ENSEMBLE METHODS

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

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

Project 7	*Students will apply skills from Weeks 1-6 while building a PostgresSQL database and performing a PCA ("principal components analysis").*
Clustering	*Define clustering and it's advantages and disadvantages from classification models*
K-Means	*Practice building and evaluating a K-means model*
PCA	*Convert a set of observations or variables into principal components in order to improve predictive analysis.*
Intro to AWS	*Learn to access and deploy databases to AWS*

WEEK 8: BAYESIAN INFERENCE & LDA

Capstone Project, Part 1	*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.*
Bayesian Methods	*Build a linear regression model with bayesian methods*
LDA	*Refine data using latent dirichlet allocation (LDA)*

DATA SCIENCE IMMERSIVE SYLLABUS

WEEK 9: TIMES SERIES & FINAL PROJECTS

Capstone Project, Part 2	*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.*
Time Series & Autocorrelation	*Analyze and visualize time series data using Pandas, plot.ly and Tableau*
ARIMA Model	*Use the ARIMA model to make predictions with time series data*

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

Capstone Project, Part 3	*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.*
SVMs & Naive Bayes	*Learn how concepts like SVMs and Naive Bayes can simplify the process of analyzing data for supervised learning algorithms*
Hadoop, Hive, Spark	*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*

WEEK 11 - 12: JOB PREP & FINAL PROJECTS

Capstone Project, Part 4 & 5	*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.*
Interview Prep	*Practice data science case studies to prep for job interviews*