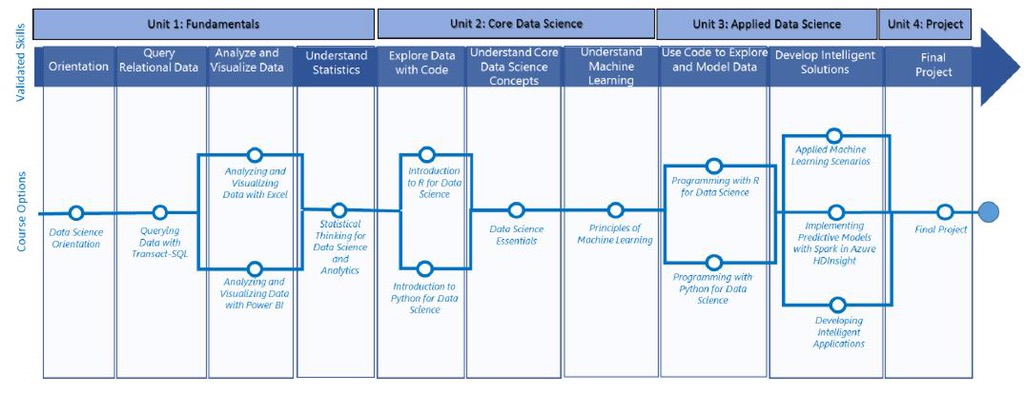


DATA SCIENCE

WINTER 2017 COHORT



COURSE SYLLABUS & CURRICULLUM

|  |  |
| --- | --- |
| WEEK | CONTENT |
| 1 | Data Science Fundamentals |
| 2 | Introductory Transact SQL |
| 3 | SQL – Joins, Functions and Set Operators |
| 4 | Statistical Thinking for Analytics |
| 5 | Conditional, Association, Regression, Clustering |
| 6 | Data Visualization |
| 7 | Bayesian Modeling |
| 8 | Python Basics |
| 9 | Python - Numpy, Plotting, Pandas |
| 10 | Statistics using Python |
| 11 | Hypothesis Testing |
| 12 | Exploring Data with Python |
| 13 | Data Cleansing & Manipulation with Python |
| 14 | Machine Learning |
| 15 | Review Week |
| 16 | Python Data & Features |
| 17 | Transforming Data (PCA, Isomap) |
| 18 | Data Modeling (Clustering, K-Nearest Neighbors, Regression, SVC, Random Forest |
| 19 | Review Week |
| 20 | Evaluating – Cross Validation |
| 21 | Final Quiz & Project |
| 22 | Model Improvement |
| 23 | Final Quiz & Project |
| 24 | Final Quiz & Project |

Week 1

**Lecture:** Intro to DS, MS Course, Expectations

**Assignment:**

Data Science Orientation DAT101x (Microsoft)

<https://courses.edx.org/courses/course-v1:Microsoft+DAT101x+5T2016/info>

1. Data Science Fundamentals
   * Getting Started with Data
     + Introduction to Data
     + Working with Data in Excel
     + Exploring Data
     + Exploring Data in Excel
     + Introduction to Data Visualization
     + Visualizing Data in Excel
     + Slicing and Dicing Data
     + Pivot Tables and Pivot Charts
2. Basic Introductions to Statistics
   * Basic Statistics
     + Getting Started with Statistics
     + Measures of Central Tendency
     + Measures of Variance
     + Descriptive Stats in Excel
     + Correlation
     + Hypothesis Testing
     + T-Tests
     + Two Sample T-Test
     + Paired T-Tests
     + Analysis of Variance
     + Regression
3. Lab: Exploring Data with Excel

Week 2

**Lecture:** Review concepts, answer questions, work problem(s) together

**Assignment:**

Querying with Transact SQL DAT201x (Microsoft)

<https://courses.edx.org/courses/course-v1:Microsoft+DAT201x+6T2016/info>

1. Section 1
   1. Module 1: Introduction to Transact SQL (View Demos and Summaries too)
      1. Getting Started with Transact SQL
      2. The Select Statement
      3. Using Select
      4. Working with Data Types
      5. Working with Nulls
   2. Complete Review: Introduction to SQL
   3. Module 2: Querying Tables with Select (View Demos and Summaries too)
      1. Removing Duplicates
      2. Sorting Results
      3. Filtering and Using Predicates
   4. Complete Lab & Review for Querying Tables with Select

Week 3

**Lecture:** Review concepts, answer questions, work problem(s) together

**Assignment:**

Querying with Transact SQL DAT201x (Microsoft)

<https://courses.edx.org/courses/course-v1:Microsoft+DAT201x+6T2016/info>

1. Section 2
   1. Module 3: Querying Multiple Tables with Joins (View Demos and Summaries too)
      1. Introduction to Joins
      2. Inner Joins
      3. Outer Joins
      4. Cross Joins
   2. Complete Review: Querying Multiple Tables with Joins
   3. Module 4: Using Set Operators (View Demos and Summaries too)
      1. Union Queries
      2. Intersect & Except Queries
   4. Complete Review for Using Set Operators
   5. Module 5: Using Functions and Aggregated Data
      1. Introduction to Functions
      2. Scalar
      3. Logical
      4. Window
      5. Aggregate
      6. Grouping Aggregated Data
      7. Filtering Groups

Week 4:

**Lecture:** Review concepts, answer questions, work problem(s) together

**Assignment:**

Statistical Thinking for Data Science and Analytics DS101x (Columbia)

<https://courses.edx.org/courses/course-v1:ColumbiaX+DS101X+1T2016/info>

1. Week 2 Statistics & Probability I
   1. 2a Statistical Thinking for Data Science
      1. Introduction
      2. Statistical Thinking for Data Science
   2. 2b Numerical Data 1 - Simple Visualization and Summaries
      1. Numerical Data 1 - Simple Visualization and Summaries
      2. Complete 2 assignments on Mathematical Notation
   3. 2c Numerical Data 2 - Simple Visualization and Summaries
      1. Numerical Data 2 - Simple Visualization and Summaries
      2. Learning Activity: Create side-by-side boxplots for comparison.
   4. 2d Numerical Data 3 - Association
      1. Numerical Data 3 Association Part 1
      2. Numerical Data 3 Association Part 2
   5. 2e Data Collection – Sampling
      1. Data Collection – Sampling
   6. 2f Introduction to Probability
      1. Introduction to Probability
      2. 2f Assignment
   7. 2g Statistical Inference - Confidence Intervals
      1. Statistical Inference - Confidence Intervals
   8. 2h Statistical Inference - Significance tests
   9. 2i Status of Current Observational Health Studies
      1. 2k Introduction
      2. 2k Status of Current Observational Health Studies
   10. 2j Statistical Terms Explained
   11. 2k Unknown Characteristics of Observational Health Studies
       1. 2m Unknown Characteristics of Observational Health Studies
       2. 2m Assignment
   12. 2l Lessons Learnt from OMOP Experiments
       1. 2n Lessons Learnt from OMOP Experiments
   13. 2m P-value Calibration

Week 5:

**Lecture:** Review concepts, answer questions, work problem(s) together

**Assignment:**

Statistical Thinking for Data Science and Analytics DS101x (Columbia)

<https://courses.edx.org/courses/course-v1:ColumbiaX+DS101X+1T2016/info>

1. Week 3 Statistics & Probability II
   1. 3a Conditional Probability
   2. 3b Bayes' Formula & Assignment
   3. 3c Studying Association: Two-way Table
   4. 3d Studying Association: Chi-square Test of Independence
   5. 3e Studying Association: One-way Analysis of Variance
   6. 3f Regression Analysis 1 and 2
   7. 3g Regression Analysis 3 and 4
   8. 3h Regression Analysis 4 and Concluding Remarks
   9. 3i Types of Data Analytics
   10. 3j Clustering Text
   11. 3k Topic Modeling
   12. 3l Metrics for Label Description
       1. 3l Assignment II
   13. 3m Concluding Remarks

Week 6:

**Lecture:** Review concepts, answer questions, work problem(s) together

**Assignment:**

Statistical Thinking for Data Science and Analytics DS101x (Columbia)

<https://courses.edx.org/courses/course-v1:ColumbiaX+DS101X+1T2016/info>

1. Week 4 Exploratory Data Analysis and Visualization
   1. 4a Graphs Are Comparisons
   2. 4b Use Data To Answer Questions
   3. 4c A Case Example
   4. 4d Decision Making Process of Data Visualization 1
   5. 4e Decision Making Process of Data Visualization 2 & Assignment
   6. 4f Decision Making Process Main Worked Example & Assignment
   7. 4g Why Visualize Data Worked Example 1
   8. 4h Why Visualize Data Worked Example 2 & Assignment 1 & 2
   9. 4i Dashboards
   10. 4j Dashboards Worked Example 1
   11. 4k Dashboards Worked Example 2

Week 7:

**Lecture:** Review concepts, answer questions, work problem(s) together

**Assignment:**

Statistical Thinking for Data Science and Analytics DS101x (Columbia)

<https://courses.edx.org/courses/course-v1:ColumbiaX+DS101X+1T2016/info>

1. Week 5 Introduction to Bayesian Modeling
   1. 5a Introduction to Bayesian Modeling
   2. 5b Probability Calibration
   3. 5c Probability As Measurement of Uncertainty
   4. 5d Bayesian Inference & Assignment
   5. 5e How To Use Prior Information & Assignment
   6. 5f Bayesian Modeling in Practice
   7. 5g Business Applications in Bayesian Statistics Introduction
   8. 5h Data Collection and Model Building 1
   9. 5i Data Collection and Model Building 2
   10. 5j Model Building Review & Assignment
   11. 5k Model Insights 1
   12. 5k Model Insights 1 & Assignment
   13. 5m Example Modeling Museum Membership Renewal
   14. 5n Example Modeling User Behavior on a Deals Website

Week 8:

**Lecture:** Review concepts, answer questions, work problem(s) together

**Assignment:**

Intro to Python for Data Science DAT208x (Microsoft)

<https://courses.edx.org/courses/course-v1:Microsoft+DAT208x+6T2016/info>

1. Python Basics
   1. Hello Python!
   2. Variables and Types
   3. Lab: Variables and Types
2. List - A Data Structure
   1. Python Lists
   2. Lab: Python Lists
   3. Subsetting Lists
   4. Lab: Subsetting Lists
   5. Manipulating Lists
   6. Lab: Manipulating Lists
3. Functions and Packages
   1. Functions
   2. Lab: Functions
   3. Methods
   4. Lab: Methods
   5. Packages
   6. Lab: Packages

Week 9:

**Lecture:** Review concepts, answer questions, work problem(s) together

**Assignment:**

Intro to Python for Data Science DAT208x (Microsoft)

<https://courses.edx.org/courses/course-v1:Microsoft+DAT208x+6T2016/info>

1. Numpy
   1. Numpy
   2. Lab: Numpy
   3. 2D Numpy Arrays
   4. Lab
   5. Basic Statistics with Numpy
   6. Lab
2. Plotting with Matplotlib
   1. Basic Plot with matplotlib
   2. Lab
   3. Histograms
   4. Lab
   5. Customization
   6. Lab
3. Control Flow and Pandas
   1. Boolean Logic and Control Flow
   2. Lab
   3. Pandas
   4. Lab

Week 10:

**Lecture:** Review concepts, answer questions, work problem(s) together

**Assignment:**

Data Science Essentials DAT203.1x (Microsoft)

<https://courses.edx.org/courses/course-v1:Microsoft+DAT203.1x+6T2016/info>

1. Introduction to Data Science
   1. Introduction to Data Science
      1. Data Analytic Thinking
      2. The Data Science Process
   2. Data Science Technologies
      1. An Overview of Data Science Technology
      2. Azure Machine Learning
      3. Using Code in Azure ML
      4. Jupyter Notebooks
      5. Creating a Machine Learning Model
   3. Lab
2. Probability and Statistics for Data Science
   1. Probability and Statistics for Data Science
      1. Introduction to Probability
      2. Discrete Random Variables
      3. Discrete Probability Distributions
      4. Binomial Distribution Examples
      5. Poisson Distributions
      6. Continuous Probability Distributions
      7. Cumulative Distribution Functions
      8. Central Limit Theorem
   2. Introduction to Statistics
      1. Descriptive Statistics
      2. Summary Stats
      3. Viewing Summary Stats
      4. Z-Scores
      5. Correlation
      6. Viewing Correlation
      7. Simpson’s Paradox
   3. Lab

Week 11:

**Lecture:** Review concepts, answer questions, work problem(s) together

**Assignment:**

Data Science Essentials DAT203.1x (Microsoft)

<https://courses.edx.org/courses/course-v1:Microsoft+DAT203.1x+6T2016/info>

1. Simulation and Hypothesis Testing
   1. Simulation and Confidence Intervals
      1. Simulation
      2. Performing a Simulation
      3. Confidence Intervals
      4. Demo Confidence Intervals
   2. Hypothesis Testing
      1. Introduction to Hypothesis Testing
      2. Z-Tests, T-Tests, and Other Tests
      3. Hypothesis Test Examples
      4. Type 1 and Type 2 Errors
      5. Demo: Hypothesis Testing
      6. Misconceptions About Hypothesis Testing
   3. Lab

Week 12:

**Lecture:** Review concepts, answer questions, work problem(s) together

**Assignment:**

Data Science Essentials DAT203.1x (Microsoft)

<https://courses.edx.org/courses/course-v1:Microsoft+DAT203.1x+6T2016/info>

1. Exploring and Visualizing Data
   1. Exploring Data
      1. Data and Data Frames
      2. Working with Data in Code
      3. Demo: Getting Started with Data Frames
      4. Working with Data Frames in Azure ML
      5. Demo: Working with Data Frames in Azure ML
      6. Working with Metadata
      7. Demo: Working with Metadata
   2. Visualizing Data
      1. Introduction to Data Visualization
      2. Conditioned Plots
      3. Plotting in Python
      4. Demo Plotting in Python
      5. Demo: Plotting in Azure ML
   3. Lab

Week 13:

**Lecture:** Review concepts, answer questions, work problem(s) together

**Assignment:**

Data Science Essentials DAT203.1x (Microsoft)

<https://courses.edx.org/courses/course-v1:Microsoft+DAT203.1x+6T2016/info>

1. Data Cleansing and Manipulation
   1. Data Ingestion and Flow
      1. Data Flow in Azure ML
      2. Joining Data Sets
      3. Demo: Ingesting and Joining Data
      4. Demo: Joins in Python
   2. Data Cleansing
      1. Overview of Data Cleansing
      2. Missing and Repeated Values
      3. Demo: Handling Missing and Repeated Values
      4. Feature Engineering
      5. Outliers and Errors
      6. Demo: Finding Outliers
      7. Demo: Handling Outliers in Azure ML
      8. Demo: Cleaning Data with Python
      9. Introduction to Data Scaling
      10. Demo: Scaling Data in Azure ML
      11. Demo: Scaling Data in Python
   3. Lab

Week 14:

**Lecture:** Review concepts, answer questions, work problem(s) together

**Assignment:**

Data Science Essentials DAT203.1x (Microsoft)

<https://courses.edx.org/courses/course-v1:Microsoft+DAT203.1x+6T2016/info>

1. Introduction to Machine Learning
2. Getting Started with Machine Learning
   * 1. Introduction to Machine Learning – Classification
     2. Evaluating Classifiers
     3. Demo: Creating a Classification Model in Azure ML
     4. Regression
     5. Evaluating Regression Models
     6. Demo: Creating a Regression Model
     7. Clustering
     8. Demo: K-Means Clustering
   1. Publishing a Machine Learning Web Service
      1. Overview of Publishing a Web Service
      2. Demo: Publishing a Web Service
      3. Demo: Consuming a Web Service
      4. Custom Code Considerations
   2. Lab
   3. Final Exam

Week 15:

**Lecture:** Review concepts, answer questions, work problem(s) together

**Assignment:**

REVIEW WEEK

Week 16:

**Lecture:** Review concepts, answer questions, work problem(s) together

**Assignment:**

Programming with Python for Data Science DAT210x (Microsoft)

<https://courses.edx.org/courses/course-v1:Microsoft+DAT210x+6T2016/info>

1. The Big Picture
   1. Data Science and Analysis
   2. Machine Learning
   3. The Possibilities (Lecture, Reading & Knowledge Check)
2. Data And Features
   1. Features Premiere
   2. Determining Features
   3. Manipulating Data
   4. Feature Regression
   5. Wrangling Data
   6. Data Features

Week 17:

**Lecture:** Review concepts, answer questions, work problem(s) together

**Assignment:**

Programming with Python for Data Science DAT210x (Microsoft)

<https://courses.edx.org/courses/course-v1:Microsoft+DAT210x+6T2016/info>

1. Exploring Data
   1. Visualizations
   2. Basic Plots
   3. Higher Dimensionality
   4. Lab: Visualizations
2. Transforming Data
   1. Transformations
   2. PCA
   3. Lab: PCA
   4. Isomap
   5. Lab: Isomap
   6. Data Cleansing

Week 18:

**Lecture:** Review concepts, answer questions, work problem(s) together

**Assignment:**

Programming with Python for Data Science DAT210x (Microsoft)

<https://courses.edx.org/courses/course-v1:Microsoft+DAT210x+6T2016/info>

1. Data Modeling
2. Clustering
3. Lab
4. Splitting Data
5. K-Nearest Neighbor
6. Lab
7. Regression
8. Lab
9. Data Modeling II
   1. SVC
   2. Lab
   3. Decision Trees
   4. Random Forest
   5. Lab

Week 19:

**Lecture:** Review concepts, answer questions, work problem(s) together

**Assignment:**

REVIEW WEEK

Week 20:

**Lecture:** Review concepts, answer questions, work problem(s) together

**Assignment:**

Programming with Python for Data Science DAT210x (Microsoft)

<https://courses.edx.org/courses/course-v1:Microsoft+DAT210x+6T2016/info>

1. Evaluating Data
2. Confusion
3. Cross Validation
4. Power Tuning

Week 21:

**Lecture:** Review concepts, answer questions, work problem(s) together

**Assignment:**

Programming with Python for Data Science DAT210x (Microsoft)

<https://courses.edx.org/courses/course-v1:Microsoft+DAT210x+6T2016/info>

FINAL QUIZ & PROJECT

Week 22:

**Lecture:** Review concepts, answer questions, work problem(s) together

**Assignment:**

Principles of Machine Learning DAT203.2x (Microsoft)

<https://courses.edx.org/courses/course-v1:Microsoft+DAT203.2x+6T2016/info>

1. Improving Machine Learning Models
   1. Principles of Model Improvement
      1. Improving Models
      2. Regularization
      3. Interpreting Features
   2. Techniques for Improving Models
      1. Sweeping Parameters
      2. Cross Validation
      3. Nested Cross Validation

Week 23:

**Lecture:** Review concepts, answer questions, work problem(s) together

**Assignment:**

Programming with Python for Data Science DAT210x (Microsoft)

<https://courses.edx.org/courses/course-v1:Microsoft+DAT210x+6T2016/info>

FINAL QUIZ & PROJECT

PLANNING FOR PROJECT COHORT

**Extra Credit:**

Principles of Machine Learning DAT203.2x (Microsoft)

<https://courses.edx.org/courses/course-v1:Microsoft+DAT203.2x+6T2016/info>

Modules 4,5,6

Week 24:

**Lecture:** Review concepts, answer questions, work problem(s) together

**Assignment:**

Programming with Python for Data Science DAT210x (Microsoft)

<https://courses.edx.org/courses/course-v1:Microsoft+DAT210x+6T2016/info>

FINAL QUIZ & PROJECT

PLANNING FOR PROJECT COHORT

**Extra Credit:**

Principles of Machine Learning DAT203.2x (Microsoft)

<https://courses.edx.org/courses/course-v1:Microsoft+DAT203.2x+6T2016/info>

Modules 4,5,6