

Microsoft Professional Program

Data Science 과정 소개서



Transforming Education

“How can you prepare for a job and Career in an Era of the fourth industrial revolution?”

Microsoft Professional Program – Data Science

MPP 란 무엇인가요?

4 차산업혁명시대의 기업과 기관에서 필요로 하는 가장 최신 직무의 수행에 도움이 되는 핵심 기술을 가르치는 코스의 모음입니다. 각 코스는 동영상 강의와 퀴즈, 클라우드 기반의 실습 랩을 포함하고 있어 매우 재미있게 학습할 수 있는 온라인 강좌입니다. 또한 각 코스를 우수한 성적으로 마쳤을 경우 이를 입증할 수 있는 마이크로소프트 공식 인증서를 획득할 수 있어 취업/재취업/창업에 큰 도움이 될 수 있습니다.



Microsoft's mission

Empower every person and every organization on the planet to achieve more.

Microsoft wants to **empower every student today to succeed in tomorrow's job market**, regardless of socio-economic status or ability.

– Microsoft Satya Nadella

Microsoft 는 오늘날의 모든 학생들이 사회 경제적 지위 또는 능력에 관계없이 내일의 직업 시장에서 성공할 수 있도록 힘을 북돋고 싶습니다.

왜 마이크로소프트 데이터 사이언스 과정을 학습해야 하나요?

최근 전세계적으로 가장 인기있는 직업 분야인 데이터 사이언티스트에 도전해 보십시오.

Microsoft 에서 만든 독창적인 MPP 온라인 학습 프로그램은 산업계의 데이터 사이언티스트 및 학계 전문가가 제공하는 데이터 사이언스의 기초, 주요 데이터 사이언스 기술 및 광범위하게 사용되는 프로그래밍 언어를 교육 커리큘럼으로 제공하고 있습니다.

데이터 사이언티스트를 필요로 하는 산업계의 요구를 반영하여 산학공동으로 구축된 Microsoft 전문 프로그램인 데이터 사이언스 과정은 현재 전세계적으로 데이터 사이언스 분야에서 제공되는 150 만 건의 취업 기회를 활용하는 데 필요한 직무 기술 및 프로그래밍 기술을 제공합니다.

과정을 통해 학습 할 수 있는 기술

- Use Microsoft Excel to explore data
- Use Transact-SQL to query a relational database
- Create data models and visualize data using Excel or Power BI
- Apply statistical methods to data
- Use R or Python to explore and transform data
- Follow a data science methodology
- Create and validate machine learning models with Azure Machine Learning
- Write R or Python code to build machine learning models
- Apply data science techniques to common scenarios
- Implement a machine learning solution for a given data problem

학습방법안내

마이크로소프트에서 개발한 데이터 사이언스 교육 과정은 총 9 개의 Track 으로 운영되고 있으며 각각의 순서대로 학습을 진행하는 것을 제안하지만, 본인의 관심도와 학습 경험에 따라 자유롭게 과정을 선택하여 학습 할 수도 있습니다. 또한 온라인 학습은 자기주도적으로 이루어지지만 교육기관에서 제공하는 Guide 에 따라 학습 계획을 세울 수도 있습니다.

MS Certification 인증서 취득

전체 9 개 Track 13 과목으로 진행되는 Data Science 과정은 학습자가 목표로 하는 진로와 경력에 맞추어 학습을 완료한 이후 인증서를 취득할 수 있습니다. 우수한 성적으로 학습을 완료하면 academy.microsoft.com 과 연동되어 학습자의 학습 정보를 관리하고, 추후 전세계적으로 채용에 활용되고 있는 LinkedIn 과 연동을 통해 개인의 커리어를 취업과 연동할 수도 있습니다.



[선문대학교 글로벌소프트웨어학과 데이터사이언스 학습 동아리]

Data Science Professional Program – 과정 개요

Unit 1 – 기본과정

데이터 사이언스에 대한 기초 과정입니다. 데이터 쿼리, 데이터 분석, 데이터 시각화 및 통계 데이터를 활용하는 데이터 과학의 기본 과정을 학습할 수 있습니다. **Course 3a 와 Course 3b 는 선택하여 과정을 학습하면 됩니다.**

Course 1: Data Science Orientation

Course 2: Querying Data with Transact-SQL

Course 3a: Analyzing and Visualizing Data with Excel

Course 3b: Analyzing and Visualizing Data with Power BI

Course 4: Essential Statistics for Data Analysis using Excel

Unit 2 – 데이터 사이언스의 핵심

머신러닝을 활용하여 데이터를 분석하기 위한 기초적인 프로그래밍 언어를 학습할 수 있습니다.. **Course 5a 와 Course 5b 는 선택하여 과정을 학습하면 됩니다.**

Course 5a: Introduction to R for Data Science

Course 5b: Introduction to Python for Data Science

Course 6: Data Science Essentials

Course 7: Principles of Machine Learning

Unit 3 – 데이터 사이언스의 응용

데이터 분석을 위한 좀더 깊이 있는 프로그래밍 언어를 학습하며, 데이터를 활용하여 지능형 솔루션을 개발하는 방법을 배울 수 있습니다. **Course 8a 와 Course 8b 는 선택하여 학습할 수 있고 Unit 2 에서 선택한 프로그래밍 언어를 지속하는 것이 좋습니다. Course 9a 와 9b 역시 선택하여 학습을 진행합니다.**

Course 8a: Programming with R for Data Science

Course 8b: Programming with Python for Data Science

Course 9a: Implementing Predictive Analytics with Spark in Azure HDInsight

Course 9b: Analyzing Big Data with Microsoft R

Unit 1 – 기본과정

Course 1 : Data Science Orientation



- 예상 학습 기간: 6 주
- 1 주일에 2~4 시간 학습 권고
- 주관기관: Microsoft
- 주제: Computer Science
- 레벨: 입문
- 동영상 강의: 한글/영어

과정 소개

이 과정은 [Microsoft Professional Program Certificate in Data Science](#) 과정에 포함되어 있습니다.

데이터 사이언티스트가 되기 위해 무엇을 준비해야 할까요?

이 과정은 마이크로소프트 데이터 사이언스 과정의 첫번째 시작점입니다. 데이터 사이언스 오리엔테이션을 통해 프로그램을 시작하고, 전체 학습 일정을 계획할 수 있습니다. 정해진 학습 목표에 따라 다양한 시각화 방법, 분석 및 통계 기법을 사용하여 데이터를 다루는 방법에 대해 소개합니다.

무엇을 배우나요?

- How the Microsoft Data Science curriculum works
- How to navigate the curriculum and plan your course schedule
- Basic data exploration and visualization techniques in Microsoft Excel
- Foundational statistics that can be used to analyze data

강사 소개



Graeme Malcolm

Senior Content Developer
Microsoft Learning Experiences



Liberty J. Munson

Principal Psychometrician and
Quality Lead Microsoft

Course 2 : Querying Data with Transact-SQL



- 예상 학습 기간 : 6 주
- 1 주일에 4~5 시간 학습 권고
- 주관기관: Microsoft
- 주제: Computer Science
- 레벨: 중급
- 동영상 강의: 한글/영어

과정 소개

이 과정은 [Microsoft Professional Program Certificate in Data Science 및 Big Data](#) 과정에 포함되어 있습니다.

트랜잭트 SQL 은 데이터베이스로 작업을 하는 데이터 전문가 및 개발자에게 필수적인 기술입니다. 전문가의 동영상 강의, DEMO, 실습을 통해 가장 처음 SELECT 구문에서부터 트랜잭션 방식의 다양한 논리 구현을 학습합니다. 여러 개의 코스 모듈을 학습하며 Microsoft SQL Server 또는 Azure SQL Database 의 데이터 쿼리 및 업데이트 등 트랜잭트 SQL 의 주요 영역을 경험할 수 있습니다. 특히 이 과정은 실습환경을 제공하고 있으며 Azure Cloud 기반의 데이터베이스를 쉽게 배포할 수 있는 샘플 데이터베이스를 사용하므로 SQL 을 설치 및 구성하지 않아도 트랜잭트 SQL 을 학습할 수 있습니다.

선행 학습

데이터베이스 및 IT 시스템에 대한 기본적인 이해

무엇을 배우나요?

- Create Transact-SQL SELECT queries
- Work with data types and NULL
- Query multiple tables with JOIN
- Explore set operators
- Use functions and aggregate data
- Work with subqueries and APPLY
- Use table expressions
- Group sets and pivot data
- Modify data
- Program with Transact-SQL
- Implement error handling and transactions

강사 소개



Geoff Allix
Microsoft Certified IT
Professional for SQL Server
Content Master



Graeme Malcolm
Senior Content Developer
Microsoft Learning
Experiences

Course 3a : Analyzing and Visualizing Data with Excel



- 예상 학습 기간: 6 주
- 1 주일에 2~4 시간 학습 권고
- 주관기관: Microsoft
- 주제: Business & Management
- 레벨: 중급
- 동영상 강의: 한글/영어

과정 소개

이 과정은 [Microsoft Professional Program Certificate in Data Science 및 Big Data](#) 과정에 포함되어 있습니다.

엑셀은 데이터를 분석하고 시각화 하는데 가장 널리 사용되는 솔루션 중 하나입니다. 향상된 시각화 기능과 보다 정교한 비즈니스 로직을 반영하여 더 많은 데이터를 분석 할 수 있는 방법을 학습하게 됩니다. 이 데이터 사이언스 과정에서는 Microsoft의 Excel Product Team 전문가가 Excel 2016에서의 최신 기능을 소개합니다.

서로 다른 소스에서 데이터를 가져 오는 방법, 데이터 원본 간에 매시업을 만드는 방법, 분석을 위해 데이터를 준비하는 방법에 대해 학습합니다. 데이터를 준비한 후 DAX 계산 엔진을 사용하여 비즈니스 결과를 표현하는 방법을 알아보십시오. Power BI 클라우드 서비스를 활용하여 데이터를 시각화 하고 공유 할 수 있는 방법을 확인한 후, 대시보드에서 사용하거나 또는 일반 영어로 된 명령문을 사용하여 쿼리할 수 있으며, 모바일 장치에서도 사용할 수 있습니다.

선행 학습

테이블, 피벗 테이블 및 피벗 차트와 같은 엑셀 분석 도구에 대한 이해 또한 DB의 데이터 및 텍스트 파일 작업에 대한 선행 지식이 도움이 됩니다. (DAT205x: Introduction to Data Analysis using Excel)

강의 계획서

1 주

Office 응용 프로그램을 설치하여 실습 환경을 설정하십시오. 이미 워크 시트 / 그리드 데이터에 준비되어 있는 데이터에 대해 피벗 테이블, 피벗 차트 및 슬라이서와 같은 도구를 사용하여 엑셀에서 데이터 분석을 수행하는 방법을 학습하십시오. Power Pivot 추가 기능을 사용하여 엑셀 데이터 모델, 콘텐츠 및 구조를 탐색하십시오. 또한 컬럼의 데이터를 측정하기 위한 DAX(Data Analysis Expressions)에 대해서 배웁니다.

2 주

쿼리(Excel 2013 및 Excel 2010의 Power Query 추가 기능)에 대해 알아보고 단일 플랫폼 테이블에서

엑셀 데이터 모델을 작성하십시오. SQL 데이터베이스에서 여러 테이블을 가져 오는 방법과 가져온 데이터에서 엑셀 데이터 모델을 만드는 방법에 대해 알아보십시오. 텍스트 파일의 데이터와 SQL 데이터베이스의 데이터 사이에 매시업을 만드는 방법을 학습합니다.

3 주

각 셀에 대해 계산할 측정 값을 만드는 방법, 계산을 위해 컨텍스트를 필터링하는 방법 및 여러 가지 고급 DAX 함수를 탐색하는 방법에 대한 세부 내용을 학습할 수 있습니다. 고급 텍스트 쿼리를 사용하여 서식이 지정된 엑셀 보고서에서 데이터를 가져 오는 방법을 배우십시오. 또한 표준 사용자 인터페이스 이상으로 난이도가 높은 쿼리를 배웁니다.

4 주

Excel 에서 멋진 시각화를 만드는 방법을 살펴보십시오. 큐브 함수를 사용하여 전년 대비 비교를 수행하게 됩니다. 타임라인, 계층구조 및 슬라이서를 만들어 시각화를 보다 향상시킵니다. 엑셀이 Power BI 와 함께 어떻게 작동하는지 알아볼 수 있습니다. 엑셀 통합 문서를 Power BI 서비스에 업로드하는 방법을 배웁니다. 마지막으로 모바일 플랫폼에서 엑셀 사용법을 살펴보게 됩니다.

무엇을 배우나요?

- Gather and transform data from multiple sources
- Discover and combine data in mashups
- Learn about data model creation
- Explore, analyze, and visualize data

강사 소개



Dany Hoter

Solutions Architect DataRails
Microsoft



Jonathan Sanito

Senior Content Developer
Microsoft

Course 3b : Analyzing and Visualizing Data with Power BI



- 예상 학습 기간: 6 주
- 1 주일에 2~4 시간 학습 권고
- 주관기관: Microsoft
- 주제: Computer Science
- 레벨: 입문
- 동영상 강의: 한글/영어

과정 소개

이 과정은 [Microsoft Professional Program Certificate in Data Science 및 Big Data](#) 과정에 포함되어 있습니다.

클라우드 기반으로 제공되는 Power BI 는 기업 내에서 필요로 하는 데이터를 직관적으로 분석하고 분석 결과를 시각화 하여, 이를 목적에 맞게 공유하는 데 큰 인기를 얻고 있습니다.

이 과정에서는 Microsoft 의 Power BI 제품 팀에서 직접 개발한 동영상 강의와 데모, 퀴즈 및 실습을 통해 학습을 진행하게 됩니다. Power BI 를 통해 데이터를 연결하고 가져 오는 방법부터 Power BI Desktop 을 사용하여 보고서를 작성한 다음 Power BI 클라우드 서비스에 게시 할 수 있습니다. 또한 웹 및 모바일 장치에서 사용자를 위한 대시보드를 만들고 공유하는 방법을 배울 수 있습니다.

선행 학습

엑셀, 데이터베이스, 혹은 텍스트 기반의 데이터 작업 경험

강의 계획서

1 주

- 비즈니스 인텔리전스, 데이터 분석 및 데이터 시각화의 주요 개념 이해
- Marketo, Salesforce 및 Google Analytics 와 같은 서비스에서 데이터 가져 오기 및 대시보드 자동 생성
- 데이터 연결 및 가져오기를 수행 한 다음 해당 데이터를 구성하고 변환
- 비즈니스 계산을 통한 데이터의 가치 증대

2 주

- 데이터 시각화 및 보고서 작성
- 업데이트 일정에 따라 자동 보고서 새로 고침
- 보고서 및 자연언어 쿼리를 기반으로 대시보드 만들기
- 조직 전체에 대시보드 공유
- 모바일 앱에서 대시보드 활용

3 주

- Power BI 내에서 엑셀 보고서 활용
- 대시보드 및 보고서에서 사용할 수 있는 사용자 지정 시각화 만들기
- 보고서 및 대시보드 작성을 위해 그룹 내에서 협업
- 조직의 필요에 따라 대시보드를 효과적으로 공유

4 주

- Power BI 를 사용하여 데이터에 대한 실시간 연결 탐색
- SQL Azure, HD Spark 및 SQL Server Analysis Services 에 직접 연결
- Power BI Development API 소개
- Power BI 의 사용자 지정 비주얼 활용 방법

무엇을 배우나요?

- Connect, import, shape, and transform data for business intelligence (BI)
- Visualize data, author reports, and schedule automated refresh of your reports
- Create and share dashboards based on reports in Power BI desktop and Excel
- Use natural language queries
- Create real-time dashboards

강사 소개



Will Thompson

Program Manager, Power BI
Microsoft



Kim Manis

Program Manager Microsoft



Jonathan Sanito

Senior Content Developer
Microsoft



Miguel Llopis

Senior Program Manager
Microsoft

Course 4 : Essential Statistics for Data Analysis using Excel



- 예상 학습 기간: 6 주
- 1 주일에 2~4 시간 학습 권고
- 주관기관: Microsoft
- 주제: Computer Science
- 레벨: 중급
- 동영상 강의: 한글/영어

과정 소개

이 과정은 [Microsoft Professional Program Certificate in Data Science](#) 과정에 포함되어 있습니다.

데이터 분석가로서의 취업과 경력을 쌓기를 고려하고 있다면 히스토그램, 파레토 차트, 박스 플롯, 베이즈 이론 등을 알아야합니다. Microsoft Excel 데이터 사이언스 두 번째 응용 통계 과정에서는 엑셀에 내장된 강력한 도구를 활용하고 개념적 관점과 응용 관점에서 통계 및 기본 확률의 핵심 원칙을 탐구합니다. 기술통계, 기본확률, 무작위 변수, 샘플링 및 신뢰구간 및 가설 테스트에 대해 배웁니다. 엑셀의 환경, 기능 및 시각화를 사용하여 이러한 개념과 원칙을 적용하는 방법을 살펴보십시오.

데이터 사이언티스트가 데이터를 분석하는 기능은 의사결정을 내리는데 큰 지표가 되며 통계 및 기본 확률의 견고한 기초는 데이터를 더 잘 이해하는 데 도움이 됩니다. 의료, 비즈니스, 스포츠, 보험 등 많은 업계에 적용 할 수 있는 실제 개념을 사용하여 엑셀이 데이터 분석을 위한 최고의 도구 중 하나 인 이유와 엑셀의 기본 기능이 엑셀을 훌륭한 방법으로 만드는 이유에 대해 최고의 전문가로부터 배우십시오.

선행 학습

- 중등학교(고등학교) 대수학
- 엑셀에서 테이블, 수식 및 차트로 작업 할 수 있는 기능
- 테이블, 피벗 테이블 및 피벗 차트와 같은 엑셀 분석도구를 사용하여 데이터를 구성하고 요약하는 능력

강의 계획서

모듈 1: 기술 통계

차트 및 기본 통계 측정을 사용하여 데이터를 분석하는 방법을 배우게 됩니다. Excel 2016 의 새로운 히스토그램, 파레토 차트, 박스플롯 및 트리맵 및 선버스트 차트를 모두 사용합니다.

모듈 2: 기본 확률

보완법칙, 독립 사건, 조건부 확률 및 베이즈 이론을 비롯한 기본적인 확률을 배웁니다.

모듈 3: 무작위 변수

임의 변수의 평균과 분산을 찾는 방법을 배우고 이항 분포, 포아송 분포, 정규 확률 변수에 대해 배웁니다. 매우 중요한 중심 극한 정리에 대한 논의로 끝맺습니다.

모듈 4: 샘플링 및 신뢰 구간

샘플링, 점추정 및 모집단 매개 변수의 간격 추정에 대한 메커니즘을 학습합니다.

모듈 5: 가설 테스트

귀무가설과 대립가설, 타입 I 오류와 타입 II 오류, 평균 및 비율에 대한 표본검정, 두 모집단 평균의 차이에 대한 검정 및 독립변수에 대한 카이 제곱 검정을 배우게 됩니다.

무엇을 배우나요?

- Descriptive statistics
- Basic probability
- Random variables
- Sampling and confidence intervals
- Hypothesis testing

강사 소개



Liberty J. Munson

Principal Psychometrician and
Quality Lead Microsoft



Matthew Minton

Senior Content Publishing
Manager Microsoft



Wayne Winston

Professor Emeritus of Decision Sciences at the Kelly School of Business Indiana
University

UNIT2 – 데이터 사이언스의 핵심

Course 5a : Introduction to R for Data Science



- 예상 학습 기간: 4 주
- 1 주일에 2~3 시간 학습 권고
- 주관기관: Microsoft
- 주제: Computer Science
- 레벨: 입문
- 동영상 강의: 한글/영어

과정 소개

이 과정은 [Microsoft Professional Program Certificate in Data Science](#) 과정에 포함되어 있습니다.

R은 데이터 사이언스 및 통계 분야의 선도 언어로 빠르게 자리 매김하고 있습니다. 오늘날 R은 모든 업계 및 분야의 데이터 사이언스 전문가에게 가장 적합한 도구입니다. 대형 컴퓨터 또는 가끔 데이터 분석가의 필요에 의해 사용하던 상관없이 R은 다양한 사용자의 요구에 맞을 것입니다.

R 프로그래밍 과정에 대한 소개는 R의 기본 사항을 익히는 데 도움이 됩니다. 총 7개 섹션에서 R을 학습하며 자신의 첫 번째 데이터 분석을 수행할 수 있도록 기본적인 구문을 다룰 것입니다. 변수 및 기본 연산을 시작으로 벡터, 행렬, 데이터 프레임과 리스트와 같은 데이터 구조를 처리하는 방법을 배우게 됩니다. 마지막 섹션에서는 R의 그래픽 기능에 대해 자세히 알아보고 멋진 시각적 데이터를 표현하기 위해 시각화를 배웁니다. 이 과정을 학습하기 위해서는 프로그래밍 언어나 데이터 사이언스에 대한 사전 지식이 필요하지 않습니다.

이 과정의 학습자들이 더욱 유익하게 학습할 수 있는 이유는 **DataCamp 플랫폼**을 사용하여 양방향 대화식 브라우저 코딩 문제를 통해 새로 습득한 기술을 지속적으로 연습할 수 있도록 코딩 실습 환경을 제공한다는 것입니다. 수동적으로 동영상 강의를 보는 대신 실제 데이터 문제를 풀어보고 동시에 올바른 솔루션을 안내하는 즉각적이고 개인화 된 피드백을 받게 됩니다

선행 학습

특별히 없지만 기본 수학 능력은 도움이 됩니다.

강의 계획서

섹션 1: 기본 소개

R을 사용하여 첫 번째 단계를 학습하며, R의 기본 데이터 유형을 발견하고 첫 번째 변수를 지정하십시오.

섹션 2: 벡터

벡터를 사용하여 도박 행위를 분석합니다. 벡터에서 요소를 만들고 이름을 지정하고 선택합니다.

섹션 3: 행렬

R의 행렬 작업 방법을 배우십시오. 스타워즈 박스 오피스 통계를 분석하여 기본 계산을 수행하고 학습 내용을 검증하십시오..

섹션 4: 요소

R은 범주 데이터를 요소로 저장합니다. 카테고리 데이터를 생성, 하위 집합 및 비교하는 방법에 대해 알아보십시오.

섹션 5: 데이터 프레임

R 작업을 할 때는 항상 데이터 프레임을 처리해야 합니다. 그러므로 데이터 프레임을 생성하고, 그중 가장 흥미로운 부분을 선택하고, 그것들을 처리하는 방법을 알아야 합니다.

섹션 6: 목록

목록을 사용하면 여러 유형의 구성 요소를 저장할 수 있습니다. 6 절에서는 목록을 다루는 방법을 보여줍니다.

섹션 7: 기본 그래픽

R 패키지를 통해 그래픽 작업을 하고 자신만의 데이터 시각화를 만들 수 있습니다

무엇을 배우나요?

- Introductory R language fundamentals and basic syntax
- What R is and how it's used to perform data analysis
- Become familiar with the major R data structures
- Create your own visualizations using R

강사 소개



Filip Schouwenaars

Filip은 DataCamp의 대화형 교육 과정의 많은 주요 배움터 개발자입니다. 그의 과정과 지도서는 이미 전세계 수천 명의 학생들이 수강했습니다.

DataCamp는 370만 회 이상의 양방향 연습을 완료한 250,000명 이상의 데이터 과학자를 교육했습니다



Jonathan Sanito

Jonathan은 마이크로소프트의 콘텐츠 개발자 및 프로젝트 관리자로 Data and Analytics 온라인 교육에 중점을 둡니다.

마이크로소프트 Dynamics NAV에서 Windows Active Directory에 이르기까지 개발자 및 IT 전문가를 대상으로 한 교육을 담당했습니다.

Course 5b : Introduction to Python for Data Science



- 예상 학습 기간: 6 주
- 1 주일에 2~4 시간 학습 권고
- 주관기관: Microsoft
- 주제: Computer Science
- 레벨: 입문
- 동영상 강의: 한글/영어

과정 소개

이 과정은 [Microsoft Professional Program Certificate in Data Science 및 Artificial Intelligence](#) 과정에 포함되어 있습니다.

파이썬은 많은 다른 응용 프로그램에 사용되는 매우 강력한 프로그래밍 언어입니다. 시간이 지남에 따라 이 오픈 소스 언어를 둘러싼 거대한 커뮤니티는 파이썬으로 보다 효율적으로 작업 할 수 있는 도구를 많이 만들고 있습니다. 최근 몇 년 동안 데이터 사이언스를 위해 특별히 제작 된 여러 도구가 있습니다. 결과적으로 파이썬으로 데이터를 분석하는 것이 결코 쉬운 일은 아닙니다.

이 실습에서는 처음부터 기본 산술 및 변수를 사용하여 시작하고 파이썬 목록, Numpy 배열 및 Pandas DataFrames 와 같은 데이터 구조를 처리하는 방법을 배웁니다. 순서에 따라 파이썬 함수와 제어 흐름을 배우게 됩니다. 또한 파이썬을 사용하여 데이터 시각화 방법을 살펴보고 실제 데이터를 기반으로 멋진 시각화를 배우게 됩니다.

선행 학습

엑셀, 데이터베이스 또는 텍스트 파일을 활용한 데이터 작업 경험

강의 계획서

섹션 1: Python Basics

파이썬 세계에 첫발을 내딛으십시오. 다른 데이터 유형을 발견하고 첫 번째 변수를 만듭니다.

섹션 2: Python Lists

하나의 이름으로 다양한 데이터 요소를 저장하는 첫 번째 방법을 알 수 있습니다. 모든 종류의 방법으로 목록을 작성, 부분 집합 및 생성하는 방법을 배웁니다.

섹션 3: Functions and Packages

파이썬 패키지 가져오기와 함수 호출을 통하여 다른 사람들의 개발 모듈을 최대한 활용하는 방법을 배웁니다.

섹션 4: Numpy

방대한 양의 데이터를 효율적으로 저장하고 계산할 수 있는 방법인 Numerical Python 으로 초고속 코드를 작성하십시오.

섹션 5: Matplotlib

전달하려는 메시지에 따라 다양한 시각화 유형을 생성하십시오. 실제 데이터를 기반으로 복잡하고 사용자 정의 된 플롯을 작성하는 방법을 익히십시오.

섹션 6: Control flow and Pandas

조건부 구문을 작성하여 스크립트의 실행을 조정하고 Pandas DataFrame(파이썬의 데이터 사이언스 핵심 데이터 구조)을 익히십시오.

무엇을 배우나요?

- Explore Python language fundamentals, including basic syntax, variables, and types
- Create and manipulate regular Python lists
- Use functions and import packages
- Build Numpy arrays, and perform interesting calculations
- Create and customize plots on real data
- Supercharge your scripts with control flow, and get to know the Pandas DataFrame

강사 소개



Filip Schouwenaars

Filip 은 DataCamp 의 대화 형 교육 과정의 많은 주요 배움터 개발자입니다. 그의 과정과 지도서는 이미 전세계 수천 명의 학생들이 수강했습니다.

DataCamp 는 370 만 회 이상의 양방향 연습을 완료 한 250,000 명 이상의 데이터 과학자를 교육했습니다



Jonathan Sanito

Jonathan 은 마이크로소프트의 콘텐츠 개발자 및 프로젝트 관리자로 Data and Analytics 온라인 교육에 중점을 둡니다.

마이크로소프트 Dynamics NAV 에서 Windows Active Directory 에 이르기까지 개발자 및 IT 전문가를 대상으로 한 교육을 담당했습니다

Course 6 : Data Science Essentials



- 예상 학습 기간 : 6 주
- 1 주일에 3~4 시간 학습 권고
- 주관기관: Microsoft
- 주제: Computer Science
- 레벨: 중급
- 동영상 강의: 한글/영어

과정 소개

이 과정은 [Microsoft Professional Program Certificate in Data Science 및 Artificial Intelligence](#) 과정에 포함되어 있습니다.

데이터 사이언스 인재에 대한 수요가 전세계적으로 급증하고 있습니다. Duke University 및 Microsoft의 전문가들과 함께 필수기술과 원칙을 학습하면서 데이터 사이언티스트가 되기 위한 경력을 개발하십시오.

이 데이터 사이언스 과정에서는 Microsoft Azure Machine Learning 플랫폼을 사용하여 클라우드 기반 데이터 사이언스 솔루션을 구축하는 방법, Azure stack 기반에 R 및 파이썬과 같은 실용적인 응용 프로그램 지향적인 내용을 학습하며, 데이터 수집, 준비, 탐색 및 시각화에 대한 주요 개념들을 배우게 됩니다.

선행 학습

- 기초 수학 능력
- R 또는 파이썬에 대한 입문 레벨 지식

강의 계획서

데이터 사이언스 프로세스 탐구 - 소개

- 데이터 사이언스 사고 이해
- 데이터 사이언스 과정 이해
- AML(Azure Machine Learning)을 사용하여 첫 번째 머신 러닝 실험을 작성하는 방법
- Lab: Azure Machine Learning 에서 첫 번째 모델 만들기, 데이터 사이언스에서 확률 및 통계
- 신뢰구간과 가설검정을 이해하고 적용
- 상관관계의 의미와 적용 이해하기, 시뮬레이션을 적용하는 방법
- Lab : 확률 및 통계 작업
- Lab : 시뮬레이션 및 가설 테스트 데이터 작업
- 데이터 처리 및 선택 방법의 기초 파악
- 데이터 정리, 통합 및 변환의 중요성과 프로세스 이해

- Lab : 데이터 처리 및 선택
- Lab : Azure Machine Learning 을 활용한 데이터 통합, Azure stack 의 R 및 파이썬을 활용한 데이터 탐색 및 시각화
- 기본 플롯 유형을 생성하고 해석하는 방법을 익히십시오.
- Datasets 탐색 프로세스 이해
- Lab: Azure Machine Learning 을 활용한 데이터 탐색 및 시각화, 지도학습을 위한 R 과 파이썬 소개
- 지도학습의 기본 개념 이해
- 자율학습(비지도학습)의 기본 개념 이해
- AML 을 이용하여 간단한 머신 러닝 모델 만들기
- Lab: 소득별 국민 분류
- Lab: 회귀분석을 통한 자동차 가격 예측
- Lab: Azure Machine Learning 을 통한 K-평균 클러스터링

무엇을 배우나요?

- Explore the data science process
- Probability and statistics in data science
- Data exploration and visualization
- Data ingestion, cleansing, and transformation
- Introduction to machine learning
- The hands-on elements of this course leverage a combination of R, Python, and Microsoft Azure Machine Learning

강사 소개



Graeme Malcolm

Senior Content Developer
Microsoft Learning
Experiences



Steve Elston

Managing Director Quantia
Analytics, LLC



Cynthia Rudin

Associate Professor MIT and Duke

Course 7 : Principles of Machine Learning



- 예상 학습 기간: 6 주
- 1 주일에 3~4 시간 학습 권고
- 주관기관: Microsoft
- 주제: Computer Science
- 레벨: 중급
- 동영상 강의: 한글(7 월)/영어

과정 소개

이 과정은 [Microsoft Professional Program Certificate in Data Science 및 Artificial Intelligence](#) 과정에 포함되어 있습니다.

머신 러닝은 컴퓨터를 사용하여 미래의 행동, 결과 및 추세를 예측하기 위해 기존 데이터로부터 학습하는 예측 모델을 실행합니다.

이 데이터 사이언스 과정에서는 실용적인 시나리오 및 실습 경험을 바탕으로 머신 러닝 이론을 명확하게 설명하고 머신 러닝 모델을 검증 및 배포합니다. R, Python 및 Azure Machine Learning 을 사용하여 이러한 모델에서 통찰력을 구축하고 파생시키는 방법을 배우게 됩니다.

강의 계획서

분류 탐색 (Explore classification)

- 분류기의 작동 이해
- 분류기준으로 로지스틱 회귀 분석 사용
- 분류기준 평가에 사용되는 측정 항목 이해
- Lab : 로지스틱 회귀 분석

머신 러닝의 회귀 (Regression in machine learning)

- 회귀 모델의 이해
- 예측을 위한 선형 회귀분석 사용
- 회귀모델을 평가하는데 사용되는 지표 이해
- Lab : Azure Machine Learning 을 사용하여 선형 회귀로 자전거 수요 예측

지도학습 모델을 개선하는 방법

- 피처 선택을 위한 프로세스
- 매개변수 초과 문제와 차원의 저주에 대한 이해
- 매개변수 초과모델에 대해 정규화 사용
- 차원감소 방법 - 추정모델 유효성 평가를 위한 교차 검증 적용
- Lab : Azure Machine Learning 을 사용하여 당뇨병 환자 분류 개선

- Lab : Azure Machine Learning 을 사용하여 자전거 수요 예측 개선

비선형 모델링에 대한 세부 정보

- 일반적으로 지도된 머신 러닝 모델을 언제 어떻게 사용해야 하는지 이해하기 - ML 모델을 당뇨병 환자 분류에 적용
- ML 모델을 자전거 수요 예측 클러스터링에 적용
- 비지도 학습 모델의 원리 이해
- k-평균 클러스터링 모델을 올바르게 적용하고 평가
- 첨단 클러스터링 모델을 올바르게 적용하고 평가
- Lab : AML, R, 파이썬을 활용한 클러스터 모델

추천 시스템

- 추천 시스템의 동작원리 이해
- 추천 시스템을 평가하는 방법 이해
- 추천을 위한 공동 작업 필터링 대안 사용 방법
- Lab : 추천 작업 및 평가

무엇을 배우나요?

- Explore classification
- Regression in machine learning
- How to improve supervised models
- Details on non-linear modeling
- Clustering
- Recommender systems
- The hands-on elements of this course leverage a combination of R, Python, and Microsoft Azure Machine Learning

강사 소개



Graeme Malcolm

Senior Content Developer
Microsoft Learning
Experiences



Steve Elston

Managing Director Quantia
Analytics, LLC



Cynthia Rudin

Associate Professor MIT and Duke

UNIT3 – 데이터 사이언스의 응용

Course 8a : Programming with R for Data Science



- 예상 학습 기간: 6 주
- 1 주일에 4~8 시간 학습 권고
- 주관기관: Microsoft
- 주제: Computer Science
- 레벨: 중급
- 동영상 강의: 한글(7 월)/영어

과정 소개

이 과정은 [Microsoft Professional Program Certificate in Data Science](#) 과정에 포함되어 있습니다.

덴마크의 Technical University (DTU)와 공동으로 개발한 이 과정은 컴퓨터 사이언스 분야에서 필요한 지식과 기술을 습득합니다.

이 과정에서는 R로 프로그래밍하는 방법을 배우게 됩니다. R 데이터 구조와 구문을 탐색하고, 로컬 파일에서 cloud-hosted 데이터베이스로 데이터를 읽고 쓰는 방법, 데이터로 작업하는 방법, 요약하는 방법, 또한 필요에 맞게 데이터를 변형하는 방법을 배우게 됩니다. 또한 R을 사용하여 예측 분석을 수행하는 방법과 널리 사용되는 ggplot2 패키지를 사용하여 시각화를 만드는 방법을 학습해 보십시오.

선행 학습

UNIT2: 데이터 사이언스를 위한 R 입문 과정

강의 계획서

Section 1: Introduction

Section 2: Functions

Section 3: Control flow and Loops

Section 4: Working with Vectors and Matrices

Section 5: Reading in Data

Section 6: Writing Data

Section 7: Reading from SQL Server

Section 8: Working with Data

Section 9: Manipulating Data

Section 10: Simulation

Section 11: Linear model

Section 12: Graphics in R

무엇을 배우나요?

- Explore R language fundamentals, including basic syntax, variables, and types
- How to create functions and use control flow.
- Details on reading and writing data in R
- Work with data in R
- Create and customize visualizations using ggplot2
- Perform predictive analytics using R

강사 소개



Anders Stockmarr
Statistician



Jonathan Sanito
Senior Content Developer
Microsoft

Course 8b : Programming with Python for Data Science



- 예상 학습 기간: 6 주
- 1 주일에 8~9 시간 학습 권고
- 주관기관: Microsoft
- 주제: Computer Science
- 레벨: 중급
- 동영상 강의: 한글(8 월)/영어

과정 소개

이 과정은 [Microsoft Professional Program Certificate in Data Science](#) 과정에 포함되어 있습니다.

Coding Dojo 와 공동으로 개발한 이 실용적인 과정은 Python 프로그래밍 입문 학습을 선행한 학습자를 대상으로 합니다. 데이터 사이언티스트에게 가장 인기있는 언어 중 하나인 파이썬은 유용한 정보를 발굴하기 위한 잘 알려진 마이닝 모델을 적용하여 데이터 사이언티스트로서 데이터를 다루는 방법을 배울 수 있습니다. 데이터 시각화, 피처의 중요성 및 선택, 차원 감소, 클러스터링, 분류 등의 주제가 포함됩니다. 이 과정에서 사용되는 모든 data sets 는 실시간 데이터로 수집되거나 머신 러닝 으로부터 실제 유용한 정보를 얻을 수 있는 방법을 학습합니다.

선행 학습

UNIT2: 데이터 사이언스를 위한 파이썬 입문 과정

무엇을 배우나요?

- What machine learning is and the types of problems it is adept to solving
- How to represent raw data in a manner conducive to deriving valuable information
- How to use various data visualization techniques
- How to use principal component analysis and isomap intelligently to simplify your data
- How to apply supervised learning algorithms to your data, such as random forest and support vector classifier
- Concepts such as model selection, pipelining, and cross validation

강사 소개



Autjman Apatira

Lead Instructor Coding Dojo



Jonathan Sanito

Senior Content Developer
Microsoft

Course 9a : Implementing Predictive Analytics with Spark in Azure HDInsight



- 예상 학습 기간: 6 주
- 1 주일에 3~4 시간 학습 권고
- 주관기관: Microsoft
- 주제: Computer Science
- 레벨: 고급
- 동영상 강의: 한글(7 월)/영어

과정 소개

이 과정은 [Microsoft Professional Program Certificate in Data Science 및 Big Data](#) 과정에 포함되어 있습니다.

빅 데이터 과학에 대한 학습 준비가 되셨습니까? 이 과정에서는 Microsoft Azure HDInsight 에서 아파치 스파크를 사용하여 대용량 데이터에 대한 예측 분석 솔루션을 구현하는 방법을 학습합니다. 스칼라 또는 파이썬을 사용하여 데이터를 정리 및 변환하고 Spark ML 로 머신 러닝 모델을 작성하는 방법을 배우보십시오.

참고: 이 과정의 실습을 진행하려면 Azure 구독과 Windows 클라이언트 컴퓨터가 필요합니다. 무료 Azure 평가 판 가입을 신청할 수 있습니다 (확인을 위해서는 유효한 신용 카드가 필요하지만 Azure 서비스에 대해서는 비용이 청구되지 않습니다).

선행 학습

- Azure HDInsight 경험
- 데이터베이스 및 SQL 에 대한 지식.
- 프로그래밍 경험.

강의 계획서

Spark를 활용한 데이터 사이언스 개론

Azure HDInsight에서 스파크 클러스터를 시작하고 스파크를 사용하여 파이썬 또는 스칼라를 실행하여 데이터 작업을 수행하십시오.

머신 러닝 시작하기

Spark ML 라이브러리를 사용하여 분류 및 회귀 모델을 작성하는 방법을 배웁니다.

머신 러닝 모델 평가

지도학습 모델을 평가하는 방법과 모델 매개 변수를 최적화하는 방법에 대해 알아보십시오.

추천자 및 비지도 학습 모델

Spark ML을 사용하여 추천 구축 및 클러스터링 모델을 작성하는 방법을 익히십시오

무엇을 배우나요?

- Using Spark to explore data and prepare for modeling
- Build supervised machine learning models
- Evaluate and optimize models
- Build recommenders and unsupervised machine learning models

강사 소개



Graeme Malcolm

Senior Content Developer
Microsoft Learning
Experiences

Course 9b : Analyzing Big Data with Microsoft R



- 예상 학습 기간: 4 주
- 1 주일에 2~4 시간 학습 권고
- 주관기관: Microsoft
- 주제: Data Analysis & Statistics
- 레벨: 고급
- 동영상 강의: 한글(8 월)/영어

과정 소개

이 과정은 [Microsoft Professional Program Certificate in Data Science 및 Big Data](#) 과정에 포함되어 있습니다.

오픈 소스 프로그래밍 언어 R은 데이터 처리 및 통계 분석을 위해 오랜 기간(특히 학계에서) 인기 있는 언어입니다. R의 장점 중 하나는 간결한 프로그래밍 언어이며 또한 모든 종류의 분석을 수행 할 수 있는 제 3자 라이브러리의 광범위한 저장소가 있다는 것입니다. 이 두 기능을 함께 사용하면 데이터 데이터 사이언티스트가 원시 데이터에서 요약, 차트 및 전체 보고서로 신속하게 만들어 낼 수 있습니다. 그러나 R의 한 가지 부족한 점은 전통적으로 많은 양의 메모리를 사용한다는 것입니다. 왜냐하면 데이터 복사본을 data.frame 개체로 전체적으로 로드해야하기 때문에 데이터를 처리할 때 종종 추가 복사본을 생성하기 때문입니다(copy-on-modify 라고 함). 이것은 R이 학계에 비해 산업계에 더 마지 못해 받아 들여진 이유 중 하나입니다.

Microsoft R Server (MRS)의 주요 구성 요소는 R 라이브러리에있는 RevoScaleR 패키지입니다. 이 라이브러리는 대규모 데이터 세트를 메모리에 한꺼번에 로드하지 않고도 처리 할 수 있는 일련의 기능을 제공합니다. RevoScaleR은 시간이 지남에 따라 추가되는 다양한 통계 및 머신 러닝 알고리즘을 공유합니다. 마지막으로, RevoScaleR은 우리가 랩톱에서 개발한 코드를 가져 와서 최소한의 노력으로 SQL Server 또는 스파크(인프라가 전혀 다른 곳)에 원격 배포 할 수 있는 메커니즘을 제공합니다. 이 과정에서는 MRS를 사용하여 대규모 데이터 집합에 대한 분석을 실행하고 SQL Server 데이터베이스에 이를 배포하는 방법에 대한 몇 가지 예를 제공합니다. 이 과정을 완료하면 빅데이터 처리를 위해 R을 사용하는 방법을 알게 됩니다.

RevoScaleR은 R 패키지이므로 과정 학습자는 R에 익숙하다고 가정합니다. R 데이터 구조 (벡터, 행렬, 목록, 데이터 프레임, 환경)에 대한 확실한 이해가 필요합니다. dplyr과 같은 타사 패키지에 익숙한 것도 학습에 도움이 됩니다

선행 학습

- 데이터 사이언스를 위한 R 입문 및 데이터 사이언스를 위한 R 프로그래밍

무엇을 배우나요?

MRS(Microsoft R Server)를 사용하여 대형 datasets 을 읽고, 처리하고, 분석하는 방법을 배우게 됩니다.

- Read data from flat files into R's data frame object, investigate the structure of the dataset and make corrections, and store prepared datasets for later use
- Prepare and transform the data
- Calculate essential summary statistics, do crosstabulation, write your own summary functions, and visualize data with the ggplot2 package
- Build predictive models, evaluate and compare models, and generate predictions on new data

강사 소개



Jonathan Sanito

Senior Content Developer
Microsoft



Seth Mottaghinejad

Data Scientist Microsoft