# Data Science
## ESSENTIALS

***Data Science Essentials*** from The Data Incubator is an introductory, 8-week, part-time, online program geared towards giving working professionals an immersive, hands-on experience with fundamental data science techniques. Developed with feedback from industry partners and based on the same rigorous methodology as our fellowship program, *Data Science Essentials* prepares data novices for the future of data in business. Students will learn how to analyze data using Python, predict trends and build models to make better business decisions.

### Our Curriculum

The Data Incubator uses open source tools, and provides each student with a Jupyter server—hosted by DigitalOcean—for the duration of the course for coursework completion. Jupyter notebooks are used so students can follow along in lectures and see and edit the raw code that's being run.

### Weekly Schedule

**Week 1**
- Files and data formats
- Object-oriented programming

**Week 2**
- NumPy
- Introduction to pandas

**Week 3**
- Data transformations and aggregation with pandas
- Time series with pandas

**Week 4**
- Data visualization
- Text and regular expressions

**Week 5**
- Introduction to machine learning
- Scikit-learn API

**Week 6**
- Regression
- Bias, variance and overfitting

**Week 7**
- Introduction to classification
- Classification example: predicting customer churn

**Week 8**
- Dimensionality reduction
- Clustering
- Scikit-learn workflow

Register for *Data Science Essentials* or learn more by visiting **thedataincubator.com**.

The Data Incubator
A Pragmatic Institute Company

# Data Science Essentials

MODULE/WEEK 1–4

## Data Essentials

Answer complicated questions, uncover greater insights and learn in-demand skills with Python. This part of the course provides the foundational knowledge needed to move beyond the limitations of traditional spreadsheets to streamline and automate time-consuming tasks and encourage data-driven decision making.

**Hands-on Lab**
Students work with pandas and the College Scorecard data set to determine the top universities by performing common pandas operations like filtering and merging data sets.

Students will also answer questions about faculty pay and tuition, using grouping and aggregating pandas operations, and applying statistical analysis.

MODULE/WEEK 5–8

## Machine Learning

Build useful machine learning models that deliver data-driven insights to help your company make better decisions that can improve revenue, create new opportunities, identify new ideas, improve the customer experience and more.

**Hands-on Lab**
Students work with housing data sets to develop a model to predict house prices based on features, using linear regression models; scikit-learn's transformers, predictors and pipelines; and more.

Students develop a model to predict customer churn based on customer features, starting with a logistic regression model and improving on it with a random forest classifier.

Students implement unsupervised learning algorithms on a real-world data set of credit card company customers. They will perform principal components analysis, train a K-Means clustering model, find cluster centroids and extract characteristic customer information