

Data Analyst Program - Connect

BEGINNER **INTERMEDIATE** ADVANCED

The Data Analyst program is ideal for learners who want to develop the core skills required to be a Data Analyst.

PREREQUISITE REQUIREMENTS

Must meet *one* of the following requirements:

1. High school degree or GED or equivalent certificate.
2. Knowledge of college-level basic algebra or arithmetic proven through a course completion certificate in Algebra, Geometry, Statistics, and Algebra 2; or AP Chemistry, Biology or Physics; or one year related work experience AND Understanding of programming concepts, such as variables, functions, loops, and basic Python data structures like lists and dictionaries as demonstrated by completion of first 3 lessons of Introduction to Computer Science or Programming Foundations in Python offered by Udacity or equivalent work projects

The coursework and projects will allow students to:

- Wrangle, extract, transform, and load data from various databases, formats, and data sources
- Use exploratory data analysis to identify meaningful relationships, patterns, or trends from complex data sets
- Classify unlabeled data or predict into the future with applied statistics and machine learning algorithms
- Communicate data analysis and findings well through effective data visualizations

Data Analyst Course and Project Descriptions

By the end of this program, students will have a portfolio that includes the following projects. Each project is accompanied by a supporting course that is available online. Completion of the online course in its entirety is highly recommended but not required to create a satisfactory project portfolio for graduation. Students may use external resources as well as support during class from Session Leads to develop their projects.

PROJECT 0: Analyze Bay Area Bike Share Data

2 hrs in-person Connect class + 3 hrs online course

This project will introduce you to the key steps of the data analysis process. You'll do so by analyzing data from a bike share company found in the San Francisco Bay Area. You'll submit this project in your first 7 days, and by the end you'll be able to:

- Use basic Python code to clean a dataset for analysis
- Run code to create visualizations from the wrangled data
- Analyze trends shown in the visualizations and report your conclusions
- Determine if this program is a good fit for your time and talents

Intro to Descriptive Statistics - Online Course, 3 hrs

Descriptive statistics will teach you the basic concepts used to describe data.

Lessons will teach how to use statistical research methods, compute and interpret values like: Mean, Median, Mode, Sample, Population, and Standard Deviation. It will also teach students to compute simple probabilities, explore data through the use of bar graphs, histograms, box plots, and other common visualizations.

Students will investigate distributions and understand a distributions properties and learn to manipulate distributions to make probabilistic predictions on data.

This is a great beginner course for those interested in Data Science, Economics, Psychology, Machine Learning, Sports analytics and just about any other field.

Connect in-person class - 2 hrs

Student meet with peers in person for a 2 hour class each time to develop the project described above.

PROJECT 1: Test a Perceptual Phenomenon,

18 hrs of Classroom sessions

12 hrs of online UagdeW

Students will learn how to use descriptive statistics and a statistical test to analyze the Stroop effect, a classic result of experimental psychology. They will give readers a good intuition for the data and use statistical inference to draw a conclusion based on the results.

Intro to Inferential Statistics - Online Course, 12 hrs

Students will learn the basic concepts and tools of inferential and descriptive statistics, including how to use statistical research methods, compute and interpret values, investigate distributions and understand a distribution's properties, estimate parameters of a population using sample statistics, run hypothesis tests and calculate confidence intervals and run t-tests.

Connect in-person class - 18 hrs

Students meet with peers in person seven times (once a week) for a 2 hour class each time to develop the project described above.

PROJECT 2: Analyzing the NYC Subway Dataset

18 hrs of Classroom sessions

30 hrs of online UagdeW

Students will look at the NYC Subway data and figure out if more people ride the subway when it is raining versus when it is not raining. They will wrangle the NYC subway data, use statistical methods and data visualization to draw an interesting conclusion about the subway dataset that they've analyzed. Specifically, students will demonstrate skills in parametric and nonparametric statistical tests, linear regression, data visualization and MapReduce.

Intro to Data Analysis - Online Course, 20 hrs

Students will survey the foundational topics in data science and learn how to apply the basic skills of data wrangling, data analysis with statistics and machine learning, communication with effective data visualization, and processing big data to real-world data sets.

Connect in-person class - 14 hrs

Students meet with peers in person seven times (once a week) for a 2 hour class each time to develop the project described above.

PROJECT 3: Data Wrangle OpenStreetMaps Data,**14 hrs of classroom session****61.5 hrs of online courses**

Students will choose any area of the world in <https://www.openstreetmap.org> and use data-munging techniques to clean the OpenStreetMap data. After completing the project, students will be able to assess the quality of data, parse and gather data from popular file formats, process data from many files that can be cleaned with spreadsheet programs and learn how to store, query and aggregate data using MongoDB.

Data Wrangling with MongoDB - Online Course, 61.5 hrs

Students will learn how to gather and extract data from widely-used data formats. They will learn how to assess the quality of data and explore best practices for data cleaning. We will also introduce students to MongoDB, covering the essentials of storing data and the MongoDB query language together with exploratory analysis using the MongoDB aggregation framework.

Connect in-person class - 14 hrs

Student meet with peers in person seven times (once a week) for a 2 hour class each time to develop the project described above.

PROJECT 4: Explore and Summarize Data**24.5 hrs online courses****12 hrs in person courses**

Students will use R and apply exploratory data analysis techniques to explore relationships in one variable to multiple variables and to explore a selected data set for distributions, outliers, and anomalies. After completing the project, students will learn how to quantify and visualize individual variables within a dataset by using the appropriate plots. Students will also be able to explore variables to identify the most important variables and relationships with a data set before building predictive models; calculate correlations and investigate conditional means. Finally, students will learn powerful methods and visualizations for examining relationships among multiple variables, such as reshaping data frames and using aesthetics like color and shape to uncover more information.

Data Analysis with R - Online Course, 24.5 hrs

Students will learn about exploratory data analysis as an approach for summarizing and visualizing the important characteristics of a data set. They will use exploratory data analysis, as promoted by John Tukey, to explore data at both micro and macro levels using appropriate visualizations and acquire statistical knowledge for summarizing data. Students will also learn to demonstrate curiosity and skepticism when performing data analysis and to develop intuition around a data set and understand how the data was generated.

Connect in-person class - 12 hrs

Student meet with peers in person six times (once a week) for a 2 hour class each time to develop the project described above.

PROJECT 5: Identifying Fraud from the Enron Email Dataset**65.5 hrs online courses + 14hrs of in-person Connect classes**

In 2000, Enron was one of the largest companies in the United States. By 2002, it had collapsed into bankruptcy due to widespread corporate fraud. In the resulting Federal investigation, there was a significant amount of typically confidential information entered into public record, including tens of thousands of emails and detailed financial data for top executives. In this final project, students will play detective, and put their new machine-learning skills to use by building a person of interest, identifier-based on financial and email data made public as a result of the Enron scandal.

Intro to Machine Learning - Online Course, 65.5

Students will learn the end-to-end process of investigating data through a machine learning lens. It will teach them how to extract and identify useful features that best represent their data, a few of the most important machine-learning algorithms, and how to evaluate the performance of their machine-learning algorithms.

Connect in-person class - 14 hrs

Student meet with peers in person seven times (once a week) for a 2 hour class each time to develop the project described above.

PROJECT 6: Tell Stories with Data Visualization

46.5 hrs - Classroom session + 14 hrs of in-person Connect classes

Students will create a data visualization from a data set that tells a story or allows a reader to explore trends or patterns. Their work should be a reflection of the theory and practice of data visualization that they have learned throughout the course.

Data Visualization and D3.js - Online Course, 46.5 hrs

Students will learn the fundamentals of data visualization and practice communicating with data. This course covers how to apply design principles, human perception, color theory, and effective storytelling to data visualization. If you present data to others, aspire to be an analyst or data scientist, or if you'd like to become more technical with visualization tools, then you can grow your skills with this course.

Connect in-person class - 14 hrs

Student meet with peers in person seven times (once a week) for a 2 hour class each time to develop the project described above.

PROJECT 7: Design an A/B Test 24.5 hrs online courses + 12 hrs of in-person Connect classes

Make design decisions for an A/B test, including which metrics to measure and how long the test should be run. Analyze the results of an A/B test that was run by Udacity and recommend whether or not to launch the change.

Design and A/B Testing - Online Course, 24.5 hrs

This course will cover the design and analysis of A/B tests, also known as split tests, which are online experiments used to test potential improvements to a website or mobile application. Two versions of the website are shown to different users —usually the existing website and a potential change. Then, the results are analyzed to determine whether the change is an improvement worth launching. This course will cover how to choose and characterize metrics to evaluate your experiments, how to design an experiment with enough statistical power, how to analyze the results and draw valid conclusions, and how to ensure that the participants of your experiments are adequately protected.

Connect in-person class - 12 hrs

Student meet with peers in person six times (once a week) for a 2 hour class each time to develop the project described above.

Project Rubric

*Each project will be evaluated in key areas by a Udacity project reviewer. All criteria must "meet specifications" in order to pass. For example, to complete the project, **Data Wrangle OpenStreetMaps Data**, students must meet specifications for the following criteria:*

- 1. Code Functionality:** All required questions are correctly solved with the submitted code. Final project code functionality reflects the description in the project document.
- 2. Code Readability:** Final project code follows an intuitive, easy-to-follow logical structure.
- 3. Overview of the Data:** Student provides a statistical overview about their chosen dataset, such as: size of the file, number of unique users, number of nodes and ways, number of chosen type of nodes, like cafes, shops, etc.