UDACITY

Data Analyst

Overview

This program prepares you for a career as a data analyst by helping you learn to organize data, uncover patterns and insights, draw meaningful conclusions, and clearly communicate critical findings. You'll develop proficiency in Python and its data analysis libraries (Numpy, pandas, Matplotlib) and SQL as you build a portfolio of projects to showcase in your job search.

Depending on how quickly you work through the material, the amount of time required is variable. We have included an hourly estimation for each section of the program. The program covers one term of three month (approx. 13 weeks). If you spend about 10 hours per week working through the program, you should finish the term within 13 weeks. Students will have an additional four weeks beyond the end of the term to complete all projects.
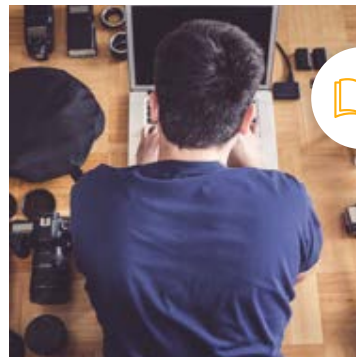
In order to succeed in this program, we recommend having experience working with data in Python (Numpy and Pandas) and SQL.
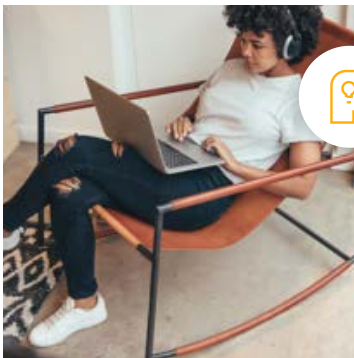
IN COLLABORATION WITH

kaggle   MODE

Estimated Time:
4 Months at
10hrs/week

Prerequisites:
Python & SQL

Flexible Learning:
Self-paced, so
you can learn on
the schedule that
works best for you

Need Help?
udacity.com/advisor
Discuss this program
with an enrollment
advisor.

Course 1: Introduction to Data Analysis

Learn the data analysis process of wrangling, exploring, analyzing, and communicating data. Work with data in Python, using libraries like NumPy and Pandas.

Course Project
Explore Weather Trends

This project will introduce you to the SQL and how to download data from a database. You'll analyze local and global temperature data and compare the temperature trends where you live to overall global temperature trends.

Course Project
Investigate a Dataset

In this project, you'll choose one of Udacity's curated datasets and investigate it using NumPy and pandas. You'll complete the entire data analysis process, starting by posing a question and finishing by sharing your findings.

	LEARNING OUTCOMES	
LESSON ONE	**Anaconda**	• Learn to use Anaconda to manage packages and environments for use with Python
LESSON TWO	**Jupyter Notebooks**	• Learn to use this open-source web application to combine explanatory text, math equations, code, and visualizations in one sharable document
LESSON THREE	**Data Analysis Process**	• Learn about the keys steps of the data analysis process. • Investigate multiple datasets using Python and Pandas.

|		
LESSON FOUR	**Pandas and AND NumPy: Case Study 1**	• Perform the entire data analysis process on a dataset • Learn to use NumPy and Pandas to wrangle, explore, analyze, and visualize data
LESSON FIVE	**Pandas and AND NumPy: Case Study 2**	• Perform the entire data analysis process on a dataset • Learn more about NumPy and Pandas to wrangle, explore, analyze, and visualize data
LESSON SIX	**Programming Workflow for Data Analysis**	• Learn about how to carry out analysis outside Jupyter notebook using IPython or the command line interface
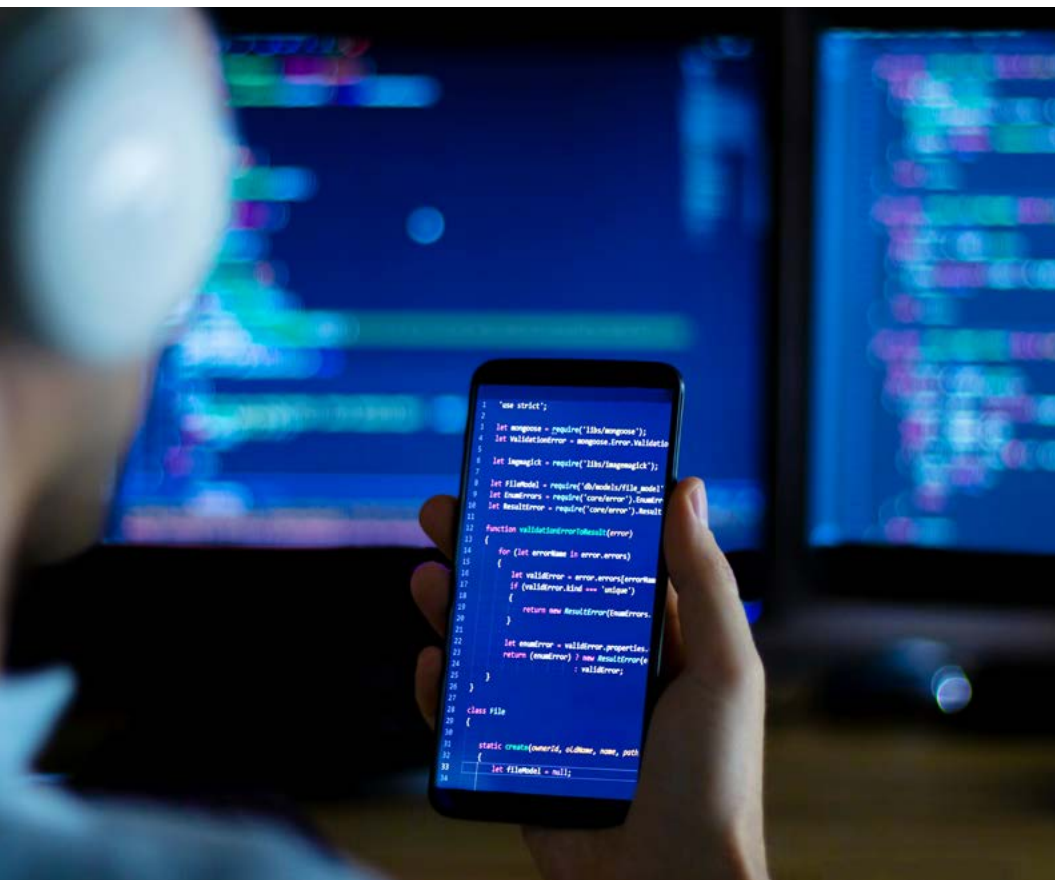
Course 2: Practical Statistics

Learn how to apply inferential statistics and probability to real-world scenarios, such as analyzing A/B tests and building supervised learning models.

Course Project
Analyze Experiment Results

In this project, you will be provided a dataset reflecting data collected from an experiment. You'll use statistical techniques to answer questions about the data and report your conclusions and recommendations in a report.

		LEARNING OUTCOMES
LESSON ONE	**Simpson's Paradox**	• Examine a case study to learn about Simpson's Paradox
LESSON TWO	**Probability**	• Learn the fundamental rules of probability.
LESSON THREE	**Binomial Distribution**	• Learn about binomial distribution where each observation represents one of two outcomes • Derive the probability of a binomial distribution
LESSON FOUR	**Conditional Probability**	• Learn about conditional probability, i.e., when events are not independent.
LESSON FIVE	**Bayes Rule**	• Build on conditional probability principles to understand the Bayes rule • Derive the Bayes theorem
LESSON SIX	**Standardizing**	• Convert distributions into the standard normal distribution using the Z-score. • Compute proportions using standardized distributions.

LESSON SEVEN	**Sampling Distributions and Central Limit Theorem**	• Use normal distributions to compute probabilities • Use the Z-table to look up the proportions of observations above, below, or in between values
LESSON EIGHT	**Confidence Intervals**	• Estimate population parameters from sample statistics using confidence intervals
LESSON NINE	**Hypothesis Testing**	• Use critical values to make decisions on whether or not a treatment has changed the value of a population parameter.
LESSON TEN	**T-Tests and A/B Tests**	• Test the effect of a treatment or compare the difference in means for two groups when we have small sample sizes
LESSON ELEVEN	**Regression**	• Build a linear regression model to understand the relationship between independent and dependent variables. • Use linear regression results to make a prediction.
LESSON TWELVE	**Multiple Linear Regression**	• Use multiple linear regression results to interpret coefficients for several predictors
LESSON THIRTEEN	**Logistic Regression**	• Use logistic regression results to make a prediction about the relationship between categorical dependent variables and predictors.

Course 3: Data Wrangling

Learn the data wrangling process of gathering, assessing, and cleaning data. Learn to use Python to wrangle data programmatically and prepare it for analysis.

Course Project
Wrangle and Analyze Data

Real-world data rarely comes clean. Using Python, you'll gather data from a variety of sources, assess its quality and tidiness, then clean it. You'll document your wrangling efforts in a Jupyter Notebook, plus showcase them through analyses and visualizations using Python and SQL.

	LEARNING OUTCOMES	
LESSON ONE	**Intro to Data Wrangling**	• Identify each step of the data wrangling process (gathering, assessing, and cleaning). • Wrangle a CSV file downloaded from Kaggle using fundamental gathering, assessing, and cleaning code.
LESSON TWO	**Gathering Data**	• Gather data from multiple sources, including gathering files, programmatically downloading files, web-scraping data, and accessing data from APIs. • Import data of various file formats into pandas, including flat files (e.g. TSV), HTML files, TXT files, and JSON files. • Store gathered data in a PostgreSQL database.
LESSON THREE	**Assessing Data**	• Assess data visually and programmatically using pandas • Distinguish between dirty data (content or "quality" issues) and messy data (structural or "tidiness" issues) • Identify data quality issues and categorize them using metrics: validity, accuracy, completeness, consistency, and uniformity
LESSON FOUR	**Cleaning Data**	• Identify each step of the data cleaning process (defining, coding, and testing) • Clean data using Python and pandas • Test cleaning code visually and programmatically using Python

Course 4: Data Visualization with Python

Learn to apply visualization principles to the data analysis process. Explore data visually at multiple levels to find insights and create a compelling story.

Course Project Communicate Data Findings	Real-world data rarely comes clean. Using Python, you'll gather data from a variety of sources, assess its quality and tidiness, then clean it. You'll document your wrangling efforts in a Jupyter Notebook, plus showcase them through analyses and visualizations using Python and SQL.

		LEARNING OUTCOMES
LESSON ONE	**Data Visualization in Data Analysis**	• Understand why visualization is important in the practice of data analysis. • Know what distinguishes exploratory analysis from Explanatory analysis, and the role of data visualization in each.
LESSON TWO	**Design of Visualizations**	• Interpret features in terms of level of measurement. • Know different encodings that can be used to depict data in visualizations. • Understand various pitfalls that can affect the effectiveness and truthfulness of visualizations.
LESSON THREE	**Univariate Exploration of Data**	• Use bar charts to depict distributions of categorical variables. • Use histograms to depict distributions of numeric variables • Use axis limits and different scales to change how your data is interpreted
LESSON FOUR	**Bivariate Exploration of Data**	• Use scatterplots to depict relationships between numeric variables. • Use clustered bar charts to depict relationships between categorical variables • Use violin and bar charts to depict relationships between categorical and numeric variables • Use faceting to create plots across different subsets of the data

|		
LESSON FIVE	**Multivariate Exploration of Data**	• Use encodings like size, shape, and color to encode values of a third variable in a visualization. • Use plot matrices to explore relationships between multiple variables at the same time. • Use feature engineering to capture relationships between variables.
LESSON SIX	**Explanatory Visulizations**	• Understand what it means to tell a compelling story with data. • Choose the best plot type, encodings, and annotations to polish your plots. • Create a slide deck using a Jupyter Notebook to convey your findings.
LESSON SEVEN	**Visulization Case Study**	• Apply your knowledge of data visualization to a dataset involving the characteristics of diamonds and their prices.
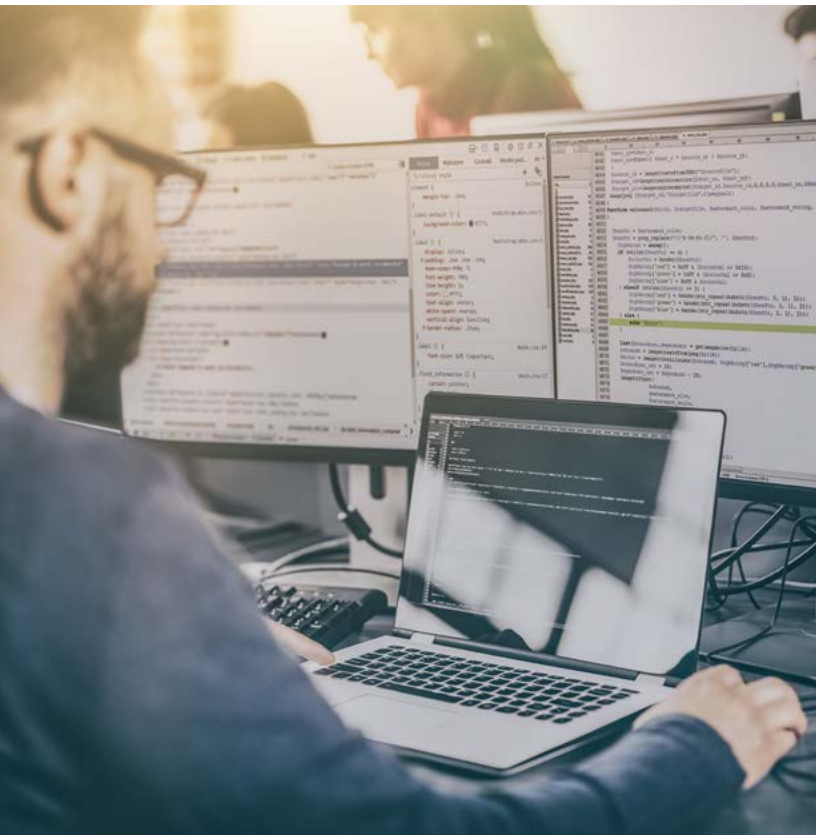
Our Classroom Experience
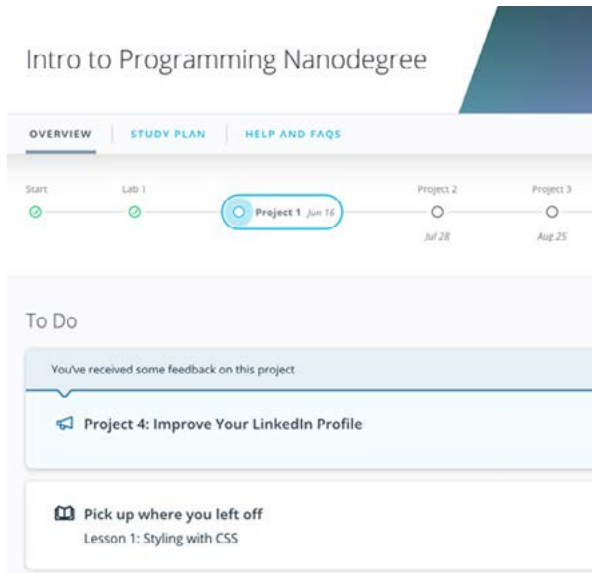
REAL-WORLD PROJECTS
Build your skills through industry-relevant projects. Get personalized feedback from our network of 900+ project reviewers. Our simple interface makes it easy to submit your projects as often as you need and receive unlimited feedback on your work.

KNOWLEDGE
Find answers to your questions with Knowledge, our proprietary wiki. Search questions asked by other students, connect with technical mentors, and discover in real-time how to solve the challenges that you encounter.

WORKSPACES
See your code in action. Check the output and quality of your code by running them on workspaces that are a part of our classroom.
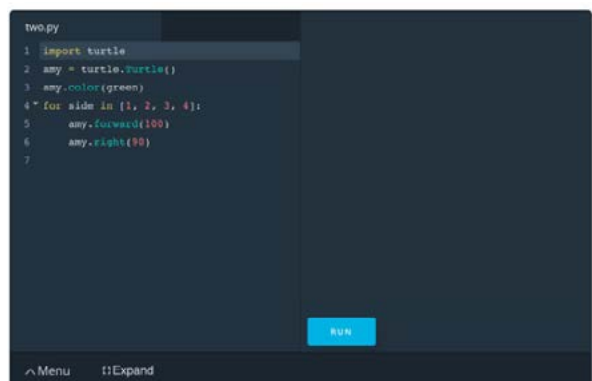
QUIZZES
Check your understanding of concepts learned in the program by answering simple and auto-graded quizzes. Easily go back to the lessons to brush up on concepts anytime you get an answer wrong.

CUSTOM STUDY PLANS
Create a custom study plan to suit your personal needs and use this plan to keep track of your progress toward your goal.

PROGRESS TRACKER
Stay on track to complete your Nanodegree program with useful milestone reminders.

UDACITY

Learn with the Best

Josh Bernhard

DATA SCIENTIST
AT NERD WALLET

Josh has been sharing his passion for data for nearly a decade at all levels of university, and as Lead Data Science Instructor at Galvanize. He's used data science for work ranging from cancer research to process automation.

Sebastian Thrun

PRESIDENT OF
UDACITY

As the founder and president of Udacity, Sebastian's mission is to democratize education. He is also the founder of Google X, where he led projects including the Self-Driving Car, Google Glass, and more.

Derek Steer

CEO
AT MODE

Derek is the CEO of Mode Analytics. He developed an analytical foundation at Facebook and Yammer and is passionate about sharing it with future analysts. He authored SQL School and is a mentor at Insight Data Science..

Juno Lee

CURRICULUM LEAD
AT UDACITY

Juno is the curriculum lead for the School of Data Science. She has been sharing her passion for data and teaching, building several courses at Udacity. As a data scientist, she built recommendation engines, computer vision and NLP models, and tools to analyze user behavior.

Learn with the Best

Mike Yi

DATA ANALYST
INSTRUCTOR

Mike is a Content Developer with a multidisciplinary academic background, including math, statistics, physics, and psychology. Previously, he worked on Udacity's Data Analyst Nanodegree program as a support lead.

David Venturi

DATA ANALYST
INSTRUCTOR

Formerly a chemical engineer and data analyst, David created a personalized data science master's program using online resources. He has studied hundreds of online courses and is excited to bring the best to Udacity students.

Sam Nelson

PRODUCT LEAD

Sam is the Product Lead for Udacity's Data Analyst, Business Analyst, and Data Foundations programs. He's worked as an analytics consultant on projects in several industries, and is passionate about helping others improve their data skills.

All Our Nanodegree Programs Include:

EXPERIENCED PROJECT REVIEWERS

REVIEWER SERVICES

- Personalized feedback & line by line code reviews
- 1600+ Reviewers with a 4.85/5 average rating
- 3 hour average project review turnaround time
- Unlimited submissions and feedback loops
- Practical tips and industry best practices
- Additional suggested resources to improve

TECHNICAL MENTOR SUPPORT

MENTORSHIP SERVICES

- Questions answered quickly by our team of technical mentors
- 1000+ Mentors with a 4.7/5 average rating
- Support for all your technical questions

PERSONAL CAREER SERVICES

CAREER SUPPORT

- Resume support
- Github portfolio review
- LinkedIn profile optimization

Frequently Asked Questions

PROGRAM OVERVIEW

WHY SHOULD I ENROLL?

The Data Analyst Nanodegree program offers you the opportunity to master data skills that are in demand by top employers, such as Python and Statistics. By the end of the program you will have created a portfolio of work demonstrating your ability to solve complex data problems. After graduating, you will have the skills needed to join a large corporation or a small firm, or even go independent as a freelance data analyst.

You'll have personalized support as you master in-demand skills that qualify you for high-value jobs in the data field. You'll also receive career support via profile and portfolios reviews to help make sure you're ready to establish a successful career in data, and land a job you love.

WHAT JOBS WILL THIS PROGRAM PREPARE ME FOR?

Graduates will be well prepared to fill a wide array of data related roles. These include: Data Analyst, Analytics Consultant, Product Manager, and Management Consultant.

HOW DO I KNOW IF THIS PROGRAM IS RIGHT FOR ME?

If you're someone who wants to make data driven decisions or work with various types of data to conduct analyses, or is interested in becoming an data analyst, this program is ideal for you, because you'll learn applied statistics, data wrangling with Python, and data visualization with Matplotlib, which will enable you to work with any data set and find and showcase meaningful insights. This will qualify you for roles such as a Data Analyst and Analytics Consultant. You'll need to have some experience with python and pandas to succeed in this program, and if that's you, and you're ready to apply those skills to real world projects, then we encourage you to enroll today.

WHAT IS THE SCHOOL OF DATA SCIENCE, AND HOW DO I KNOW WHICH PROGRAM TO CHOOSE?

Udacity's School of Data consists of several different Nanodegree programs, each of which offers the opportunity to build data skills, and advance your career. These programs are organized around three main career roles: Business Analyst, Data Analyst, and Data Scientist.
The School of Data currently offers two clearly-defined career paths. These paths are differentiated by whether they focus on developing programming skills or not. Whether you are just getting started in data, are looking to augment your existing skill set with in-demand data skills, or intend to pursue advanced studies and career roles, Udacity's School of Data has the right path for you! Visit How to Choose the **Data Science Program That's Right for You** to learn more.

FAQs Continued

ENROLLMENT AND ADMISSION

DO I NEED TO APPLY? WHAT ARE THE ADMISSION CRITERIA?
No. This Nanodegree program accepts all applicants regardless of experience and specific background.

WHAT ARE THE PREREQUISITES FOR ENROLLMENT?
In order to succeed in this program, we recommend having the following experience:
• Python programming, including common data analysis libraries (e.g., Numpy and Pandas)
• SQL Programming

You should also be able to read and write in English.

TUITION AND TERM OF PROGRAM

HOW IS THIS NANODEGREE PROGRAM STRUCTURED?
The Data Analyst Nanodegree program is comprised of content and curriculum to support five (5) projects. We estimate that students can complete the program in four (4) months working 10 hours per week.
Each project will be reviewed by the Udacity reviewer network and platform. Feedback will be provided and if you do not pass the project, you will be asked to resubmit the project until it passes.

HOW LONG IS THIS NANODEGREE PROGRAM?
Access to this Nanodegree program runs for the length of time specified in the payment card above. If you do not graduate within that time period, you will continue learning with month to month payments. See the **Terms of Use** and **FAQs** for other policies regarding the terms of access to our Nanodegree programs.

I HAVE GRADUATED FROM THE DATA ANALYST PROGRAM AND I WANT TO KEEP LEARNING. WHERE SHOULD I GO FROM HERE?
Check out our Data Scientist Nanodegree program to take the concepts you have learned in Data Analyst and build upon them using machine learning and neural networks. Learning these advanced concepts will not only enhance your knowledge it will make you a more attractive candidate to be hired as an analyst or data scientist.

WHAT IS THE SCHOOL OF DATA SCIENCE, AND HOW DO I KNOW WHICH PROGRAM TO CHOOSE?
Udacity's School of Data consists of several different Nanodegree programs, each of which offers the opportunity to build data skills, and advance your career. These programs are organized around career roles like Business Analyst, Data Analyst, Data Scientist, and Data Engineer.

FAQs Continued

The School of Data currently offers three clearly-defined career paths in Business Analytics, Data Science, and Data Engineering. Whether you are just getting started in data, are looking to augment your existing skill set with in-demand data skills, or intend to pursue advanced studies and career roles, Udacity's School of Data has the right path for you! Visit How to Choose the Data Science Program That's Right for You to learn more.

SOFTWARE AND HARDWARE

WHAT SOFTWARE AND VERSIONS WILL I NEED IN THIS PROGRAM?
For this Nanodegree program you will need access to the Internet, and a 64 bit computer.

Additional software such as Python and its common data analysis libraries (e.g., Numpy and Pandas) will be required, but the program will guide students on how to download once the course has begun.