

# Data Analyst Nanodegree Syllabus

*Discover Insights from Data with Python, R, SQL, and Tableau*



## Before You Start

**Prerequisites:** In order to succeed in this program, we recommend having experience working with data in SQL and/or a spreadsheet tool like Microsoft Excel. You should also have a good understanding of descriptive statistics, including how to calculate and interpret measures of center (mean, median, mode); measures of spread (standard deviation, 5-number summary); and build bar charts, histograms, boxplots, and scatterplots.

**Educational Objectives:** Learn to organize data, uncover patterns and insights, draw meaningful conclusions, and clearly communicate critical findings. Learn to use Python, R, SQL and Tableau. Gain all the skills necessary to get a job as a data analyst.

### Program Design

*Length of Program\**: The program is divided into two terms of three months each (approx. 13 weeks). We expect students to work 10 hours/week on average. Estimated time commitment is 130 hours per term.

*Textbooks required:* None

*Instructional Tools Available:* Video lectures, personalized project reviews, live chat help, dedicated mentor

\*The length is an estimation of total hours the average student may take to complete all required coursework, including lecture and project time. Actual hours may vary.

## TERM 1: DATA ANALYSIS WITH PYTHON AND SQL

### Intro Project: Explore Weather Trends (5 hrs)

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

## Project: Explore US Bikeshare Data (40 hrs)

You will use Python to perform steps of the data analysis process on bikeshare trip data collected from three US cities. You will write code to clean the data, compute descriptive statistics, and create basic visualizations of the distribution of data.

### Supporting Lesson Content: Introduction to Data Analysis

Lesson Title	Learning Outcomes
NUMBERS AND STRINGS	<ul style="list-style-type: none"><li>→ Learn about Python's numeric and string data types</li><li>→ Use variables to store data</li><li>→ Use built-in functions and methods</li></ul>
FUNCTIONS, INSTALLATION, AND CONDITIONALS	<ul style="list-style-type: none"><li>→ Install Python on your computer</li><li>→ Organize your code into functions</li><li>→ Use conditionals to make decisions</li></ul>
DATA STRUCTURES AND LOOPS	<ul style="list-style-type: none"><li>→ Use collection data types: lists, sets, and dictionaries</li><li>→ Write `for` and `while` loops to express repetition</li><li>→ Practice refactoring and problem solving</li></ul>
FILES AND MODULES	<ul style="list-style-type: none"><li>→ Use modules from the Python standard library and from third-party libraries</li><li>→ Read data from files on disk</li><li>→ Use online resources to help solve problems</li></ul>

## Project: Investigate a Dataset (40 hrs)

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.

### Supporting Lesson Content: Introduction to Data Analysis

Lesson Title	Learning Outcomes
Data Analysis in Python	
DATA ANALYSIS PROCESS	<ul style="list-style-type: none"><li>→ Learn about the keys steps of the data analysis process</li><li>→ Investigate multiple datasets using Python and Pandas</li></ul>
PANDAS AND NUMPY: CASE STUDY 1	<ul style="list-style-type: none"><li>→ Perform the entire data analysis process on a dataset</li><li>→ Learn to use NumPy and Pandas to wrangle, explore, analyze,</li></ul>

---

and visualize data

---

**PANDAS 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.
- 

**Introduction to SQL**

---

**Basic SQL**

- Write common SQL commands including SELECT, FROM, and WHERE, as well as corresponding logical operators
- 

**SQL Joins**

- Write JOINS in SQL, as you are now able to combine data from multiple sources to answer more complex business questions
- 

**SQL Aggregations**

- Write common aggregations in SQL including COUNT, SUM, MIN, and MAX
  - Write CASE and DATE functions, as well as work with NULLs
- 

**Advanced SQL Queries**

- Edit a database using CREATE TABLE, INSERT INTO, UPDATE, and other statements
  - Use window functions and subqueries to add steps to a query
  - Use documentation to learn new functions and complete complex tasks
- 

## Project: Analyze Experiment Results (45 hrs)

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.

### Supporting Lesson Content: Inferential Statistics

Lesson Title	Learning Outcomes
STANDARDIZING	<ul style="list-style-type: none"><li>→ Convert distributions into the standard normal distribution using the Z-score</li><li>→ Compute proportions using standardized distributions</li></ul>
NORMAL DISTRIBUTION	<ul style="list-style-type: none"><li>→ Use normal distributions to compute probabilities</li><li>→ Use the Z-table to look up the proportions of observations above, below, or in between values</li></ul>
SAMPLING DISTRIBUTIONS	<ul style="list-style-type: none"><li>→ Apply the concepts of probability and normalization to sample data sets</li></ul>
ESTIMATION	<ul style="list-style-type: none"><li>→ Estimate population parameters from sample statistics using confidence intervals</li></ul>

---

HYPOTHESIS TESTING	→ Use critical values to make decisions on whether or not a treatment has changed the value of a population parameter
T-TESTS	→ Test the effect of a treatment or compare the difference in means for two groups when we have small sample sizes
REGRESSION	→ Build a linear regression model to understand the relationship between independent and dependent variables → Use linear regression results to make a prediction

## TERM 2: ADVANCED DATA ANALYSIS

### Intro Project: Test a Perceptual Phenomenon (10 hrs)

In this project, you'll use descriptive statistics and a statistical test to analyze the Stroop effect, a classic result of experimental psychology. Communicate your understanding of the data and use statistical inference to draw a conclusion based on the results.

#### Supporting Lesson Content: Inferential Statistics

### Project: Wrangle and Analyze Data (50 hrs)

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.

#### Supporting Lesson Content: Data Wrangling

Lesson Title	Learning Outcomes
INTRO TO DATA WRANGLING	<ul style="list-style-type: none"> <li>→ Identify each step of the data wrangling process (gathering, assessing, and cleaning)</li> <li>→ Wrangle a CSV file downloaded from Kaggle using fundamental gathering, assessing, and cleaning code</li> </ul>
GATHERING DATA	<ul style="list-style-type: none"> <li>→ Gather data from multiple sources, including gathering files, programmatically downloading files, web-scraping data, and accessing data from APIs</li> <li>→ Import data of various file formats into pandas, including flat files (e.g. TSV), HTML files, TXT files, and JSON files</li> <li>→ Store gathered data in a PostgreSQL database</li> </ul>

<b>ASSESSING DATA</b>	<ul style="list-style-type: none"> <li>→ Assess data visually and programmatically using pandas</li> <li>→ Distinguish between dirty data (content or “quality” issues) and messy data (structural or “tidiness” issues)</li> <li>→ Identify data quality issues and categorize them using metrics: validity, accuracy, completeness, consistency, and uniformity</li> </ul>
<b>CLEANING DATA</b>	<ul style="list-style-type: none"> <li>→ Identify each step of the data cleaning process (defining, coding, and testing)</li> <li>→ Clean data using Python and pandas</li> <li>→ Test cleaning code visually and programmatically using Python</li> </ul>

## Project: Explore and Summarize Data (50 hrs)

In this project, you’ll use R and apply exploratory data analysis techniques to explore a selected data set for distributions, outliers, and anomalies.

### Supporting Lesson Content: Data Analysis with R

Lesson Title	Learning Outcomes
<b>WHAT IS EDA?</b>	<ul style="list-style-type: none"> <li>→ Define and identify the importance of exploratory data analysis (EDA)</li> </ul>
<b>R BASICS</b>	<ul style="list-style-type: none"> <li>→ Install RStudio and packages</li> <li>→ Write basic R scripts to inspect datasets</li> </ul>
<b>EXPLORE ONE VARIABLE</b>	<ul style="list-style-type: none"> <li>→ Quantify and visualize individual variables within a dataset</li> <li>→ Create histograms and boxplots</li> <li>→ Transform variables</li> <li>→ Examine and identify tradeoffs in visualizations</li> </ul>
<b>EXPLORE TWO VARIABLES</b>	<ul style="list-style-type: none"> <li>→ Properly apply relevant techniques for exploring the relationship between any two variables in a data set</li> <li>→ Create scatter plots</li> <li>→ Calculate correlations</li> <li>→ Investigate conditional means</li> </ul>
<b>EXPLORE MANY VARIABLES</b>	<ul style="list-style-type: none"> <li>→ Reshape data frames and use aesthetics like color and shape to uncover information</li> </ul>
<b>DIAMONDS AND PRICE PREDICTIONS</b>	<ul style="list-style-type: none"> <li>→ Use predictive modeling to determine a good price for a diamond</li> </ul>

## Project: Create a Tableau Story (20 hrs)

In this project, you'll create a data visualization, using Tableau, from a data set that tells a story or highlights trends or patterns in the data. Your work should be a reflection of the theory and practice of data visualization, harnessing visual encodings and design principles for effective communication.

### Supporting Lesson Content: Data Visualization with Tableau

Lesson Title	Learning Outcomes
DATA VISUALIZATION FUNDAMENTALS	<ul style="list-style-type: none"><li>→ Understand the importance of data visualization</li><li>→ Know how different data types are encoded in visualizations</li></ul>
DESIGN PRINCIPLES	<ul style="list-style-type: none"><li>→ Select the most effective chart or graph based on the data being displayed</li><li>→ Use color, shape, size, and other elements effectively</li></ul>
CREATING VISUALIZATIONS WITH TABLEAU	<ul style="list-style-type: none"><li>→ Become proficient in basic Tableau functionality, including charts, filters, hierarchies, etc.</li><li>→ Create calculated fields in Tableau</li></ul>
TELLING STORIES WITH TABLEAU	<ul style="list-style-type: none"><li>→ Create Tableau dashboards and stories to effectively communicate data</li></ul>