![LaGuardia Community College logo]

DATA ANALYTICS

Micro Credential Program

at LaGuardia Community College

TABLE OF CONTENTS

Course Description:

Data analytics is a process of extracting information from raw data and generating actionable insights. It is one of the most sought-after skills in the information technology marketplace, with demand spanning across several sectors including but not limited to healthcare, finance, retail, real estate, education, and gaming.

This is an applied, project-based course which prepares students to gather, wrangle, analyze and visualize data using various tools and technologies helping business in decision-making and strategy. Students will also have opportunity to work on multiple real-world projects throughout the to advance their career in the analytics field.

Course Learning Objectives:

At the completion of this course, students will be able to use:

- SQL to interact with relational databases, perform database operations and extract information
- Python as a primary programming language to gather, clean, analyze and visualize data using scientific libraries
- Advanced techniques and statistical tools to conduct thorough and insightful analyses, and to interpret the results in an accurate and meaningful way
- Tableau to build interactive dashboards using data visualization techniques

How is this course relevant for analytics professionals?

The ability to gather, prepare, and analyze data is one of the most important, and most sought-after skills in the information technology and analytics marketplace. Skills and projects completed during the course will relate directly to actual every-day job functions in the data analytics field and are designed around employer needs. The skills learned during the course will allow students to be better prepared for roles such as:

- Data Analyst
- Business Analyst
- Data Scientist

Data analytics positions can be found in every sector such as: healthcare, finance, insurance, marketing, real estate, retail, and even gaming! Data analytics professionals can be found working at companies such as: Facebook, Twitter, Google, ConEdison, Northwell Health, NYC Agencies, AIG, Tishman Speyer, TIAA and more

OUR INDUSTRY PARTNERS

Our partners in this journey speak volume

We have partnered with industry leaders who are not only building data analytics solutions to make better business decisions but also transforming their business models, creating new line of businesses, products, revenue streams and services. From sourcing materials and forecasting demand to the accounting and human resource activities, every aspect of the business using data analytics. Our course topics and industry use cases are carefully selected and meeting the current market needs and expectations.

With continuous support from NYC Jobs CEO Council and our strategic partners this program will prepare students for future job market.

CURRICULUM OVERVIEW

The program will give students the breadth and depth of knowledge needed to enter or advance their careers in data analytics. From SQL, python and tableau, to statistical concepts – students will feel more confident and prepared with their analytics abilities. Students will explore and answer real-life questions from publicly available datasets and practice explaining and presenting their findings to their peers.

UNITS	HOURS
Unit 1: Data Analytics Foundation	14 hours
Unit 2: Python Programming Fundamentals	6 hours
Unit 3: Data Analytics and Descriptive Statistics in Python	12 hours
Unit 4: Machine Learning in Python	10 hours
Unit 5: Data, Databases, and SQL	20 hours
Unit 6: Data Visualization & Commination with Tableau	18 hours
Mini Projects & Presentations	10 hours
Test, Quiz and Support	10 hours
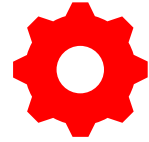
Unit 1: Data Analytics Foundation

Students will be able define Data Science, understand the importance of data science as a discipline and discuss the roles and responsibilities of a data scientist, Identify the tools used for data analysis and their limitations.

Students will explore Data Analytics process, understand the fundamentals of Exploratory Data Analysis (EDA) and learn about hypothesis Building and explain the process of Hypothesis Testing.

Students will build solid data analytics foundation by understanding the basics of descriptive statistics, inferential statistics, data distributions and dive into statistical algorithms and models,

Some topics this unit will cover:

- What is data science? How different industry and organization uses data science?
- Data Science and Big Data
- Tools for Data Analysis
- Data Analytics Process and Exploratory Data Analysis (EDA)
- Hypothesis building and hypothesis testing
- Descriptive statistics, data distribution
- Statistical theorems, models and correlation matrix

Outcomes of this unit:

- A solid foundation of data science and data analytics process
- How to use statistical functions and models for statistical data analysis

Unit 2: Python Programming Fundamentals

Students will begin to explore Python 3. Python is high level, general purpose programming language – used in application and web development, data analysis, artificial intelligence, and more. This until will expose students to the fundamentals of using Python – setting the stage for more advanced data analytics topics.

Some topics this unit will cover:

- Variable assignments and data types
- Data Structure: Lists, Dictionaries, Sets, Tuple
- Conditional Operations: If Statements, else, elif
- Logical Operations: AND, OR, NOT Functions
- Built-in Functions and Custom Functions

Outcomes of this unit:

- A solid foundation of Python Programming fundamentals
- Start using Python as a primary data analysis programming language
- How to loop through a dataset and perform different tasks
- Create, use, and understand why functions are useful and necessary for programming.

Projects:

In addition to multiple assignments that provide practice on core concepts, students will be engaged in projects that develop their skills using functions and looping through datasets. Students will select from a list of projects such as: creating a random number generator, build custom functions and re-usable software components.

Unit 3: Data Analytics & Descriptive Statistics in Python

Building on the topics learned in the previous units, students will dive deeper into data analysis and statistics using Python. Students will be importing data from multiples sources, including excel and csv files. Students will explore real-world business cases using popular and relevant scientific data analysis libraries.

Some topics this unit will cover:

- NumPy
- Pandas
- Matplotlib, Seaborn
- Data wrangling
- Regression

Outcomes of this unit:

- Use NumPy and Pandas to gather data from various sources (excel, csv, etc.), and clean and transform that data.
- Create visualizations to communicate findings in Matplotlib, Seaborn, and other visualization tools.
- Understand hypothesis testing and how it's used in solving business use cases.

Projects:

- Analyze Covid-19 dataset and gain insights from underlying data
- Perform stock data analysis using various data analysis libraries
- Perform analysis on CDC data to answer questions such as: does the data show significant evidence that certain demographic groups are more prone to diabetes? How and does diabetes impact the length of stay in a hospital?
- Perform analysis on various tasks and activities performed by a data analyst professional and summarize the analysis

CURRICULUM

Unit 4: Machine Learning and Modeling Techniques

Students will learn how to build and implement machine learning models that can help their organization answer big data questions. Students will learn some of the most popular and used algorithms for supervised and unsupervised machine learning.

Some topics this unit will cover following machine learning algorithms:

- Linear regression
- Logistic regression
- KNN
- Decision Trees
- Random Forests
- Scikit-Learn

Outcomes of this unit:

- A solid foundation of machine learning techniques, and how to use them in Python.
- Understand and use the modeling process and why it's important for all machine learning.
- How to leverage Python's advanced libraries to perform Machine Learning.

Projects:

- Use Machine Learning techniques to predict if Boston housing market prices
- Use Machine Learning techniques to predict if a passenger was more likely to survive the Titanic.
- Use Machine Learning to explore Amazon Alexa Review data and perform sentiment analysis
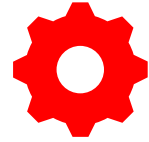
Unit 5: Data, Databases, and SQL

Students will gain an overview of data, databases and SQL using PostgreSQL. The skills learned during this unit are applicable to any other major SQL database, such as Microsoft SQL Server, MySQL, Oracle and others.

Students will explore what exactly data is, why it's useful, and what types of hidden questions data can help bring to light.

SQL is one of the most in demand tech skills – being able to extract data from your organization allows you to manipulate, transform, visualize and answer business related questions.

Some topics this unit will cover:

- What is data? How does it relate to business?
- SQL Statement fundamentals
- GROUP BY Statements
- JOINS
- Creating Databases and Tables

Outcomes of this unit:

- A solid foundation of gathering data with SQL – from simple to complex queries.
- How to create databases, tables, and import data into that database and perform queries on it.

Projects

In addition to multiple assignments that provide practice on core concepts, students will be engaged in projects that develop their skills joining tables, updating tables, and querying data.

Unit 6: Data Visualization & Commination with Tableau

Organizations use dashboards to keep track of their day-to-day, and long-term business objectives – from keeping track of patient admissions in an emergency department, to tracking spend vs. budgets. Students will learn how to create visualizations that tell meaningful stories that viewers and users can clearly identify and understand to help keep track of business objectives.

Some topics this unit will cover:

- Data-Ink Ratio
- ChartJunk
- Small Multiples
- Multifunctioning Graph Elements
- Aesthetics and technique
- Tools tips
- Exploratory Data Analysis in Tableau
- Tableau parameters and calculations

Outcomes of this unit:

- A solid foundation of data visualization techniques – presenting data in a meaningful, impactful way to present an easy-to-understand story.
- Creating visualizations and dashboards using tools such as tableau and communicating findings to peers.

Projects:

- Students will create dashboards and stories in tableau. The projects should incorporate the various visualization skills, tell a story, and be business driven. Students can select projects from the provided project list, or work with the instructor on a passion-related dataset.

Mini Projects

Mini Project

The mini project involves an analysis of the students choosing relating to their field of interest or current industry, creating a visualization (static or interactive) using relevant data and presentation to their classmates. All mini projects are guided projects, so you will be working very closely with your instructors and teammates to build and deliver projects. A short 2-3 paragraph proposal on why this dataset is relevant to your industry is required.

You may work with a team of up to 3-4 people, or individually. Each team member is responsible for understanding and being able to explain all portions of the project.

You should expect to share your work with classmates, your instructor, as well as your employer.

Your project should include the following guidelines:

- Describe your motivation for performing this analysis
- Describe your data sources
- Project includes at least one data transformation (example: converting columns to different data type)
- Project includes at least one statistical analysis and one graphic that supports your data.
- One graph that supports your conclusions.
- Each section of your project should be accompanied by an explanation. For example, a 1-2 paragraph explaining what is happening in that section or graph – so the reader understandings the story being told.
- Presentation 3-5 minutes.
- Presentation: did the audience come away with a clear understanding of why you undertook this project?
- Presentation: did the audience come away with a clear understanding of one insight or conclusion you've found?
- Delivered code and data should run without errors.
- Code should be hosted on github.

PROJECT LIST - INDUSTRY USE CASES

Use Case #	Industry	Description	Type
1	Healthcare	Analyze Covid -19 data	Statistical Data Analysis
2	Finance	Stock market data analysis	Statistical Data Analysis
3	Real Estate	Boston housing market pricing analysis	Predictive Analytics
4	Insurance	Cross sell insurance products data analysis	Statistical Data Analysis
5	Social Media	Sentiment Analysis	Predictive Analytics
6	Marketing and Advertising	Analyze and predict ad click advertising data	Predictive Analytics
7	US State - Energy	Exploratory Data Analysis of State Wind Energy Facts using Tableau	Analytics Dashboard

Students can also select their own passion projects. If there is a project you feel passionate about, and it directly relates to your current career path. Please discuss with the instructor for approval. Projects will be approved on a case-by-case basis; however, we welcome students to take ownership of the learning and explore and find value from meaningful datasets.

Sites such as:

- https://opendata.cityofnewyork.us/
- https://data.world/
- https://archive.ics.uci.edu/ml/index.php
- https://www.kaggle.com/datasets
- https://www.data.gov/

Offer almost endless amounts of meaningful datasets waiting to be explored.   These websites act as essential tools in helping you find projects to explore during class. All projects should relate to your current line of work.

REQUIREMENTS

Grading: You will be graded on your understanding of the theory of the various data analytics topics, your presentation of information, and how you apply the tools we will be using in class.

- Assignments: 20%
- Classroom Exercise and Quiz: 30%
- Mini Project: 30%
- Project presentation: 10%
- Class Participation: 10%

Course Completion Requirements: To pass this course, you must complete all projects, including the mini projects, and make the project presentation.

Code Quality: Students are responsible for providing all code when necessary for assignments and projects. Code should run – unless you provide an explanation on why your code isn't working, you will not receive credit. Code does not need to be perfect – credit will still be earned with code that is submitted and an effort is clearly shown.

Textbooks

- Think Python: How to Think Like a Computer Scientist by Allen B. Downey
- Data Science from Scratch: First Principles with Python by Joel Grus
- Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython by Wes McKinney
- The Language of SQL: How to Access Data in Relational Databases by Larry Rockoff
- Practical Tableau: 100 Tips, Tutorials, and Strategies by Ryan Sleeper
- https://www.openintro.org/book/os/