Machine Learning Foundation Syllabus

Before You Start

Educational Objectives: You will learn python-numpy and pandas, descriptive and inferential statistics along with evaluation & verification of the data. We will teach Data analysis process: Learn how to use data to answer questions.

Program Design

*Length of Program**: The program is of 5 months We expect students to work 10 hours/week on average.

Project 1: Explore US Bikeshare Data

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	➔ Learn about Python's numeric and string data types ➔ Use variables to store data ➔ Use built-in functions and methods
FUNCTIONS, INSTALLATION, AND CONDITIONALS	➔ Install Python on your computer ➔ Organize your code into functions ➔ Use conditionals to make decisions
DATA STRUCTURES AND LOOPS	➔ Use collection data types: lists, sets, and dictionaries ➔ Write `for` and `while` loops to express repetition ➔ Practice refactoring and problem solving
FILES AND MODULES	➔ Use modules from the Python standard library and from third-party libraries ➔ Read data from files on disk ➔ Use online resources to help solve problems

Project 2: Test a Perceptual Phenomenon

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

Lesson Title	Learning Outcomes
STANDARDIZING	➔ Convert distributions into the standard normal distribution using the Z-score ➔ Compute proportions using standardized distributions
NORMAL DISTRIBUTION	➔ Use normal distributions to compute probabilities ➔ Use the Z-table to look up the proportions of observations above, below, or in between values
SAMPLING DISTRIBUTIONS	➔ Apply the concepts of probability and normalization to sample data sets
ESTIMATION	➔ Estimate population parameters from sample statistics using confidence intervals
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

Project 3: 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.

Supporting Lesson Content: Introduction to Data Analysis

Lesson Title	Learning Outcomes
Data Analysis in Python	

DATA ANALYSIS PROCESS	➔ Learn about the keys steps of the data analysis process ➔ Investigate multiple datasets using Python and Pandas
PANDAS 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
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.

Project 4: Predicting Boston Housing Prices

The Boston housing market is highly competitive, and you want to be the best real estate agent in the area. To compete with your peers, you decide to leverage a few basic machine learning concepts to assist you and a client with finding the best-selling price for their home. Luckily, you've come across the Boston Housing dataset which contains aggregated data on various features for houses in Greater Boston communities, including the median value of homes for each of those areas. Your task is to build an optimal model based on a statistical analysis with the tools available. This model will then be used to estimate the best-selling price for your clients' home

Supporting Lesson Content: Model Evaluation and Validation

Lesson Title	Learning Outcomes
STATISTICAL ANALYSIS	➔ Identify key features of datasets, such as average, mean, median, mode, standard deviation, and quantiles.
DATA MODELING	➔ Learn the basic types of data. ➔ Learn how to handle datasets in sklearn.
EVALUATION AND VALIDATION	➔ Test a model and use metrics such as accuracy and recall to compare and improve models.
MANAGING ERROR AND COMPLEXITY	➔ Learn the types of error such as overfitting and ➔ underfitting. Learn to identify them using learning curves and model complexity. ➔ Apply techniques such as cross validation to improve your models.