Invictus Engineers

DATA WRANGLING

ACTIONABLE INSIGHTS FROM RAW DATA SOURCES

Invictus Engineers

3rd Floor, Surviba Towers, Plot No.247/3RT, Near Umesh Chandra Statue, Sanjeeva Reddy Nagar, Hyderabad, Telangana. 500038 **Email Us :** info@invictusengineers.com **Phone:** +91 99666 33097
www.invictusengineers.com

Topic 1: Introduction to Data Wrangling

Importance of Data Wrangling

Python for Data Wrangling

Lists, Sets, Strings, Tuples, and Dictionaries

List Functions

- Sets
- Introduction to Sets
- Union and Intersection of Sets
- Creating Null Sets
- Dictionary
- Tuples
- Creating a Tuple with Different Cardinalities
- Unpacking a Tuple
- Strings
- String Functions

Topic 2: Advanced Operations on Built-In Data Structures

Introduction

Advanced Data Structures

- Iterator
- Stacks
- User-Defined Methods
- Lambda Expressions
- Queue

Basic File Operations in Python

- File Handling
- The with Statement
- Opening a File Using the with Statement

Topic 3: Introduction to NumPy, Pandas, and Matplotlib

Introduction

NumPy Arrays

- NumPy Arrays and Features

Advanced Mathematical Operations

- Conditional SubSetting
- Stacking Arrays
- Pandas DataFrames
- Indexing and Slicing Columns
- Indexing and Slicing Rows

Statistics and Visualization with NumPy and Pandas

- Refresher on Basic Descriptive Statistics

The Definition of Statistical Measures – Central Tendency and Spread

- Random Variables and Probability Distribution
- What is a Probability Distribution?
- Discrete Distributions
- Continuous Distributions

Data Wrangling in Statistics and Visualization

- Using NumPy and Pandas to Calculate Basic Descriptive Statistics
- Random Number Generation Using NumPy

Topic 4: A Deep Dive into Data Wrangling with Python

Introduction

Subsetting, Filtering, and Grouping

- Sales Data in an Excel File
- Subsetting the DataFrame
- An Example Use Case – Determining Statistics on Sales and Profit
- Conditional Selection and Boolean Filtering
- The GroupBy Method

Detecting Outliers and Handling Missing Values

- Missing Values in Pandas
- Using the fillna Method
- The dropna Method
- Outlier Detection Using a Simple Statistical Test

Concatenating, Merging, and Joining

- Merging by a Common Key
- The join Method

Useful Methods of Pandas

- Randomized Sampling
- The value_counts Method
- Pivot Table Functionality
- Functions with the apply Method

Topic 5: Getting Comfortable with Different Kinds of Data Sources

Introduction

Reading Data from Different Sources

- Data Files Provided with This Chapter
- Libraries to Install for This Chapter
- Reading Data Using Pandas
- Data from a CSV File
- Delimiters Are Not Commas
- Headers of a CSV File
- Reading Only the First N Rows
- Setting the skip_blank_lines Option
- Reading CSV Data from a Zip File
- Reading from an Excel File Using sheet_name and Handling a Distinct sheet_name
- Reading HTML Tables Directly from a URL
- Reading from a JSON file
- Reading a PDF File

Introduction to Beautiful Soup and Web Page Parsing

- Structure of HTML

Topic 6: Learning the Hidden Secrets of Data Wrangling

Introduction

Advanced List Comprehension and the zip Function

- Introduction to Generator Expressions

Data Formatting

- The % operator
- Using the format Function

Identifying and Cleaning Outliers

- Z-score

Levenshtein Distance

- Additional Software Required

Topic 7: Advanced Web Scraping and Data Gathering

Introduction

The Requests and BeautifulSoup Libraries

- Checking the Encoding of a Web Page
- Extracting Text from a Section

- Extracting Important Historical Events that Happened on Today's Date

Reading Data from XML

- Reading from a Local XML File into an ElementTree Object
- Extracting and Printing the GDP/Per Capita Information Using a Loop
- Finding All the Neighboring Countries for Each Country and Printing Them

Reading Data from an API

- Defining the Base URL (or API Endpoint)
- Using the Built-In JSON Library to Read and Examine Data
- Printing All the Data Elements
- Using a Function that Extracts a DataFrame Containing Key Information

Fundamentals of Regular Expressions (RegEx)

- RegEx in the Context of Web Scraping
- Using the compile Method to Create a RegEx Program
- Finding the Number of Words in a List That End with "ing"
- The search Method in RegEx
- Sets of Matching Characters
- The findall Method

TOPIC 8: RDBMS AND SQL

Introduction

Revisiting of RDBMS and SQL

- How Is an RDBMS Structured?
- SQL
- Using an RDBMS (MySQL/PostgreSQL/SQLite)
- DDL and DML Commands in SQLite
- Reading Data from a Database in SQLite
- The ALTER Command
- The GROUP BY clause

Relation Mapping in Databases

- Adding Rows in the comments Table

Joins

- Retrieving Specific Columns from a JOIN Query
- Deleting Rows from Tables
- Updating Specific Values in a Table

Invictus Engineers

TOPIC 9: APPLICATIONS IN BUSINESS USE CASES

Introduction

Applying Your Knowledge to a Data Wrangling Task

An Extension to Data Wrangling

Additional Skills Required to Become a Data Scientist

Basic Familiarity with Big Data and Cloud Technologies

What Goes with Data Wrangling?

Tips and Tricks for Mastering Machine Learning

Invictus Engineers

3rd Floor, Surviba Towers, Plot No.247/3RT,
Near Umesh Chandra Statue, Sanjeeva Reddy Nagar,
Hyderabad, Telangana. 500038

+91 99666 33097

info@invictusengineers.com