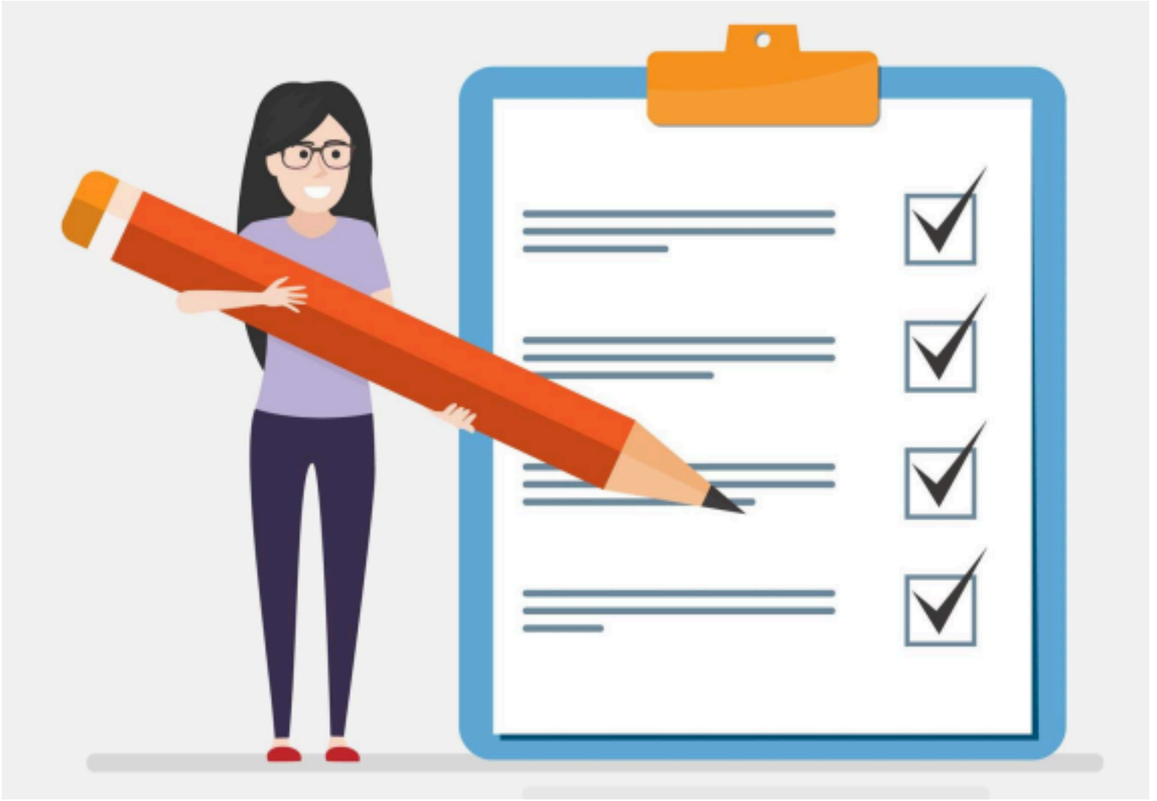
Incremental Capstone - Session 1
Capstone Project

The capstone project is a platform for the learners to implement the skills learned through the program.

Why Capstone?

The project will enable you to:

- Apply the skills learned through the program
- Build a project end-to-end
- Make better data-driven decisions
- Gain exposure to real-world data science challenges



Milestones

The project milestones are as follows:

Agenda

Import, export, and clean data	Supervised learning with regression	Solving a binary classification problem
Perform linear algebraic operations	Supervised learning with classification	Solving a multiclass image classification problem using CNN
Working with Pandas	Modeling with unsupervised learning	Performing image classification and object detection
Analyze and cleanse data	Performing data management and ensemble learning	Building a CNN-LSTM hybrid model
Visualize data	Developing recommendation engines	Denoising images with autoencoders

Aura - The Intelligent Marketing Data Engine
Introduction

ClickO is a Boston, MA-based email marketing company. ClickO delivers integrated data analytics services to its customers, primarily large enterprises in the retail space.

ClickO plans to expand from being purely an email marketing platform to a comprehensive AI/ML-based digital marketing and recommendation platform. This new platform will offer tools and analytics to help digital
marketers predict demand, optimize content, and customize digital campaigns that reach audiences on multiple channels.

To this end, ClickO is developing a new product called Aura. It aims to collate audience information from multiple sources and provides a toolkit for digital marketers to reach similar audiences across multiple platforms.

Introduction

Aura toolkit consists of the following:

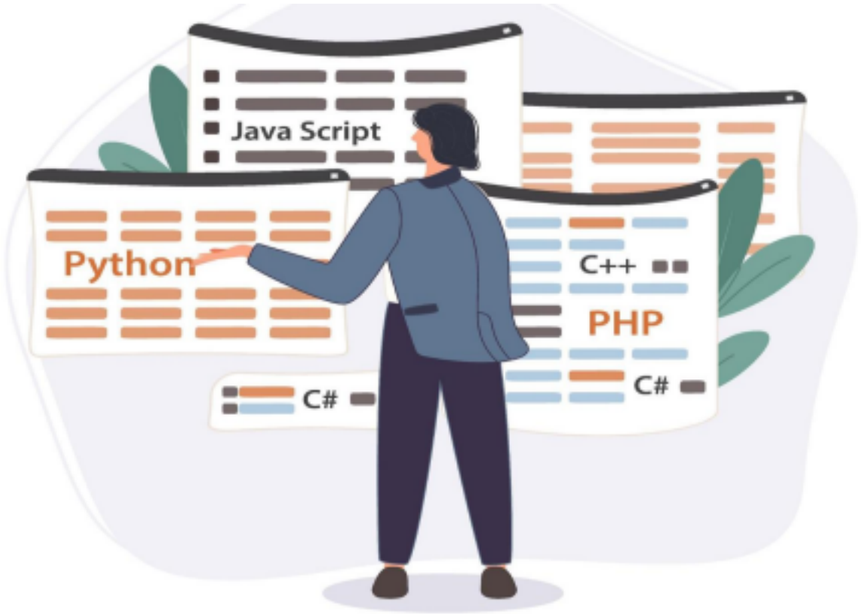
1. Real-time recommendation system: It ingests disparate data on products, people, geography, and other factors to make relevant recommendations.
2. Content optimization engine: Aura's marketing ML algorithm can estimate which types of content, questions, and headlines are most likely to resonate with your target audience.
3. Custom promotion engine: Aura leverages deep learning to help customers tailor marketing promotions for new products or services.
4. Demand forecasting engine: Using historical data, this engine can predict demand for certain products and services.

Project Goal

Build **Aura - The Intelligent Prediction and Recommendation Engine** for digital marketers.

Python for Data Science
Data Preprocessing with Python

Aura must be built to receive and process marketing campaign and user behavior data from various sources such as healthcare, technology and manufacturing domains.



Aggregate data

Shape data

Understand the statistical analysis of data

Normalize data

Project Statement

Develop a comprehensive solution for data aggregation, wrangling, and visualization using a healthcare dataset for the Aura platform.

The goal is to effectively manage and process complex healthcare data to enable insightful analysis and enhance data-driven decision-making capabilities within Aura.

Dataset Description

NSMES1988.csv

visits	Number of physician office visits	health	Factor indicating self-perceived health
nvisits	Number of non-physician office visits	chronic	Number of chronic conditions
ovisits	Number of physician hospital outpatient visits	adl	Factor indicating whether the individual has a condition that limits activities of daily living
novisits	Number of non-physician hospital outpatient visits	region	Factor indicating region
emergency	Emergency room visits	age	Age in years (divided by 10)

hospital	Number of hospital stays	married	Factor. Is the individual married?
gender	Factor indicating gender	income	Family income in USD 10000

school	Number of years of education	insurance	Factor. Is the individual covered by private insurance?
employed	Factor. Is the individual employed?		
medicaid	Factor. Is the individual covered by Medicaid?		

Session 1: Data Import and Cleaning Tasks

Task: Import and Export data, clean data

- Import relevant Python libraries necessary for Python programming and Numpy for doing numerical operations.
- Import the CSV file – NSMES1988.csv into a dataframe.
- Inspect the data and report the details from physical inspection – rows, columns, data types etc.
- Find out if the data is clean or if the data has missing values.
- Comment on the data types, their values and range, specifically on age and income columns.
- Export the data to JSON as NSMES1988.json format file and view and enter your comments.

Session 1: Data Import and Cleaning Tasks

- Perform memory information on the data and recommend what non-default data types would you recommend to optimize memory settings for the dataframe.

- What changes would you recommend on the dataframe before attempting a detailed data analysis?
- Export the data frame as a new CSV file NSMES1988new.csv and store it in the local space for likely use in other assignments.
- Write a short report on the visual observations of the data.

Thank You