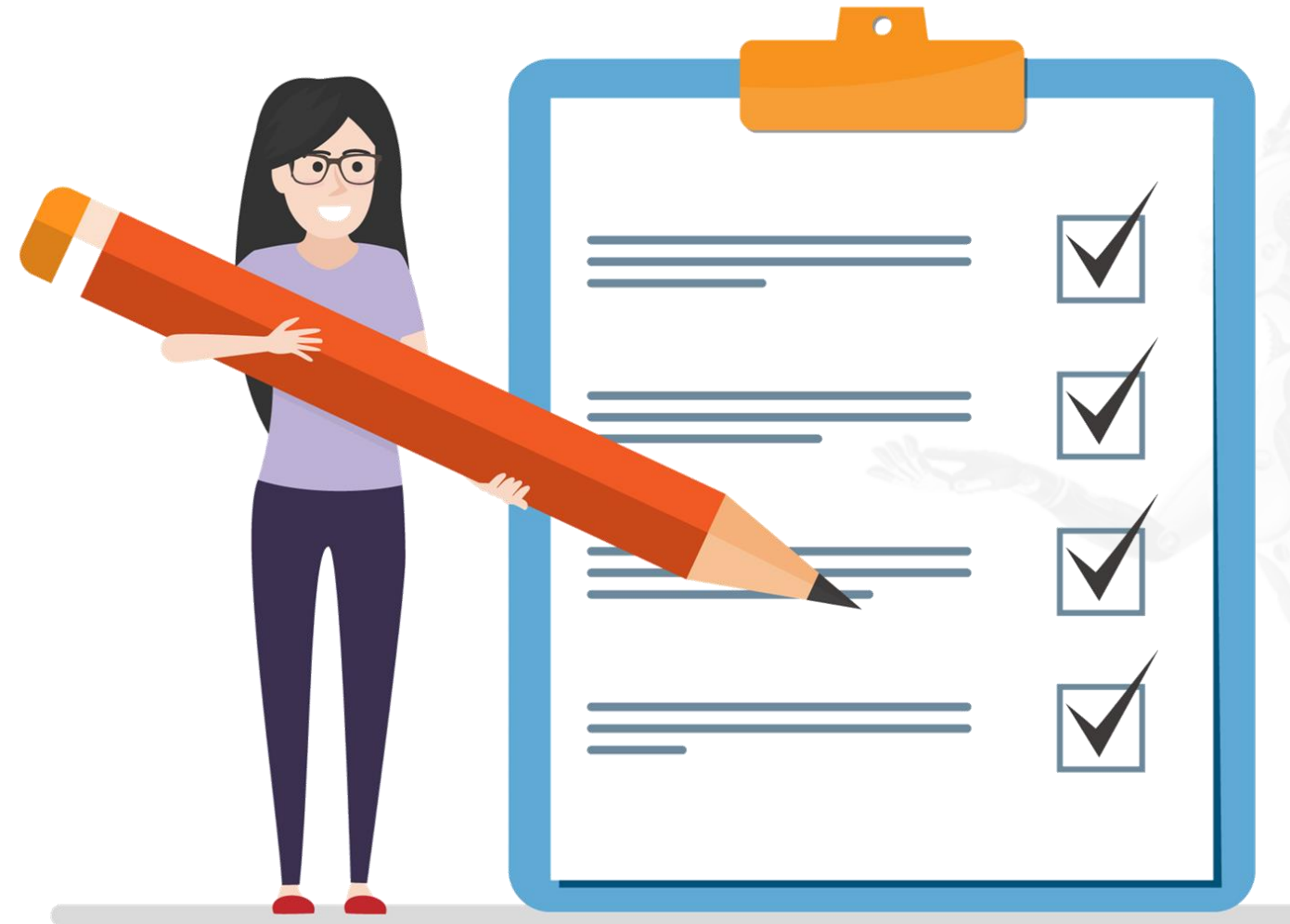Capstone Session 1

Capstone Project

This capstone project is a platform for the learners to implement the skills learned through the program.
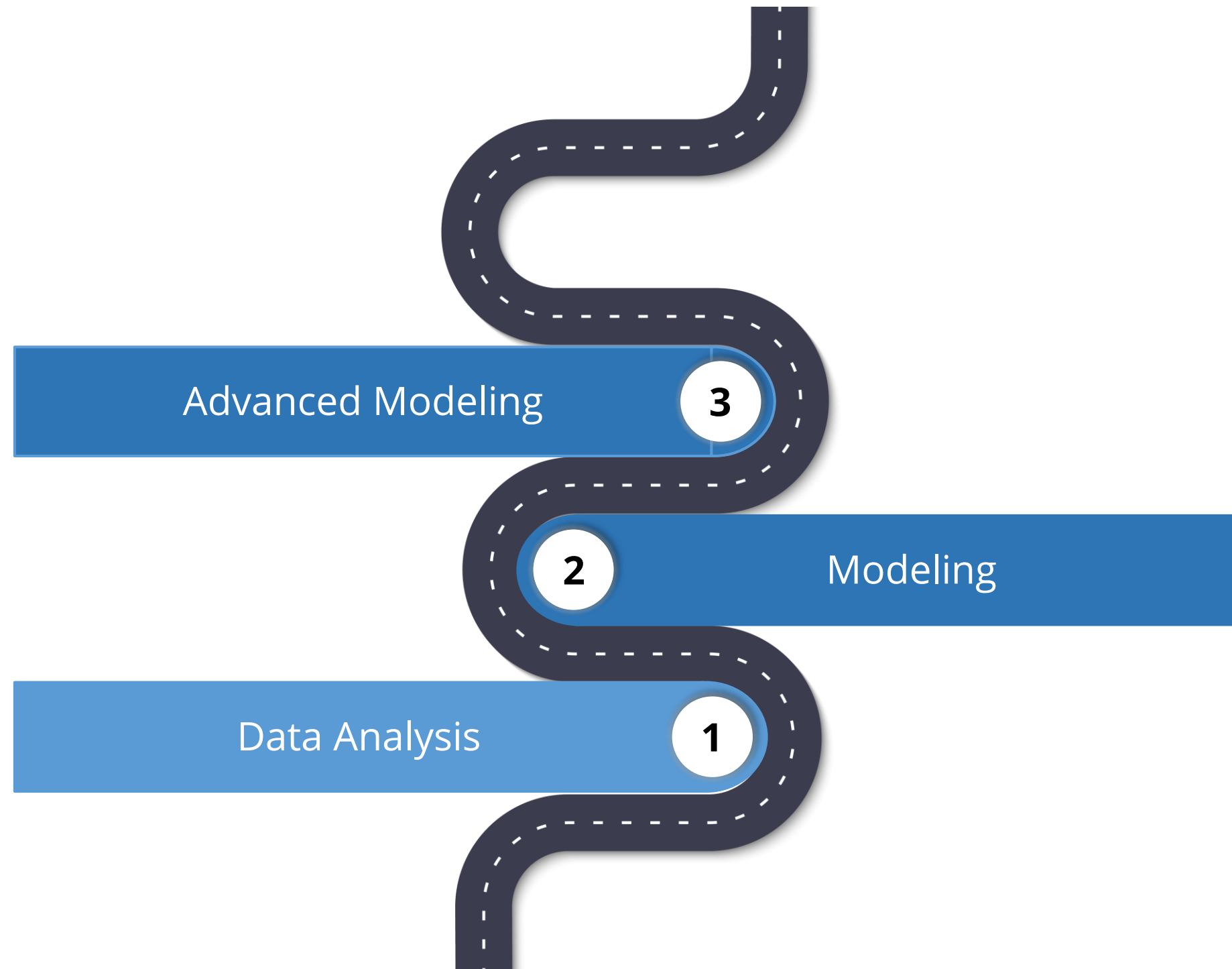
Why Capstone?

The project will enable you to:

- Apply the skills learned through the program

- Build a project end-to-end

- Make better data-driven decisions

- Get an exposure to real-world data science challenges

Milestones

The project milestones are as follows:

3 Advanced Modeling

2 Modeling

1 Data Analysis

Agenda

Data Analysis Weeks 1 - 4	Modeling Weeks 5 - 8	Advanced Modeling Weeks 9 - 12
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
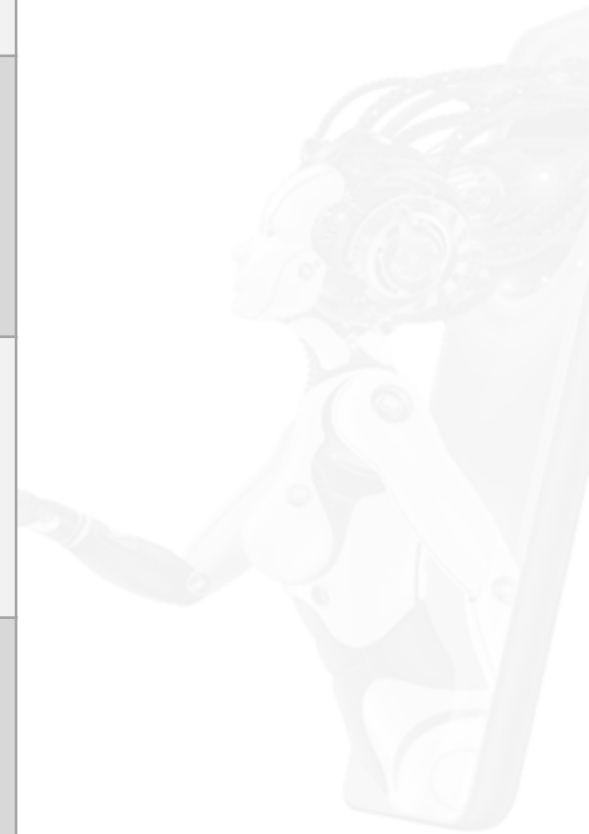
Aura - The Intelligent Marketing Data Engine

Introduction

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

ClickO plans to expand from being a pure email marketing platform to a full-fledged AI/ML-based digital marketing and recommendation platform with tools and analytics to help digital marketers predict demand, optimize content and custom-stitch digital campaigns that reach audiences on multiple channels.

Towards this end, ClickO is developing a new product called Aura. Aura collates audience information from multiple sources and provides a toolkit for digital marketers to market to similar audiences on multiple platforms.

Introduction

Among other tools, Aura toolkit consists of -

1. Real-time recommendation system to take in a disparate product, people, geographic and other data and make relevant recommendations.

2. Content optimization engine - Aura's marketing ML algorithm can help you to estimate which types of content, questions, and headlines are most probable to become popular among 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 - Based on input datasets of historical data, the engine can predict demand for certain products and services.

Project Goal

Build **Aura - The Intelligent Prediction and Recommendation Engine** for Product Marketers.

Digital marketers can leverage the insights derived from Aura to:

- Predict demand
- Optimize content
- Custom-stitch digital campaigns

simplilearn

Python for Data Analysis

simplilearn

Python End Goal

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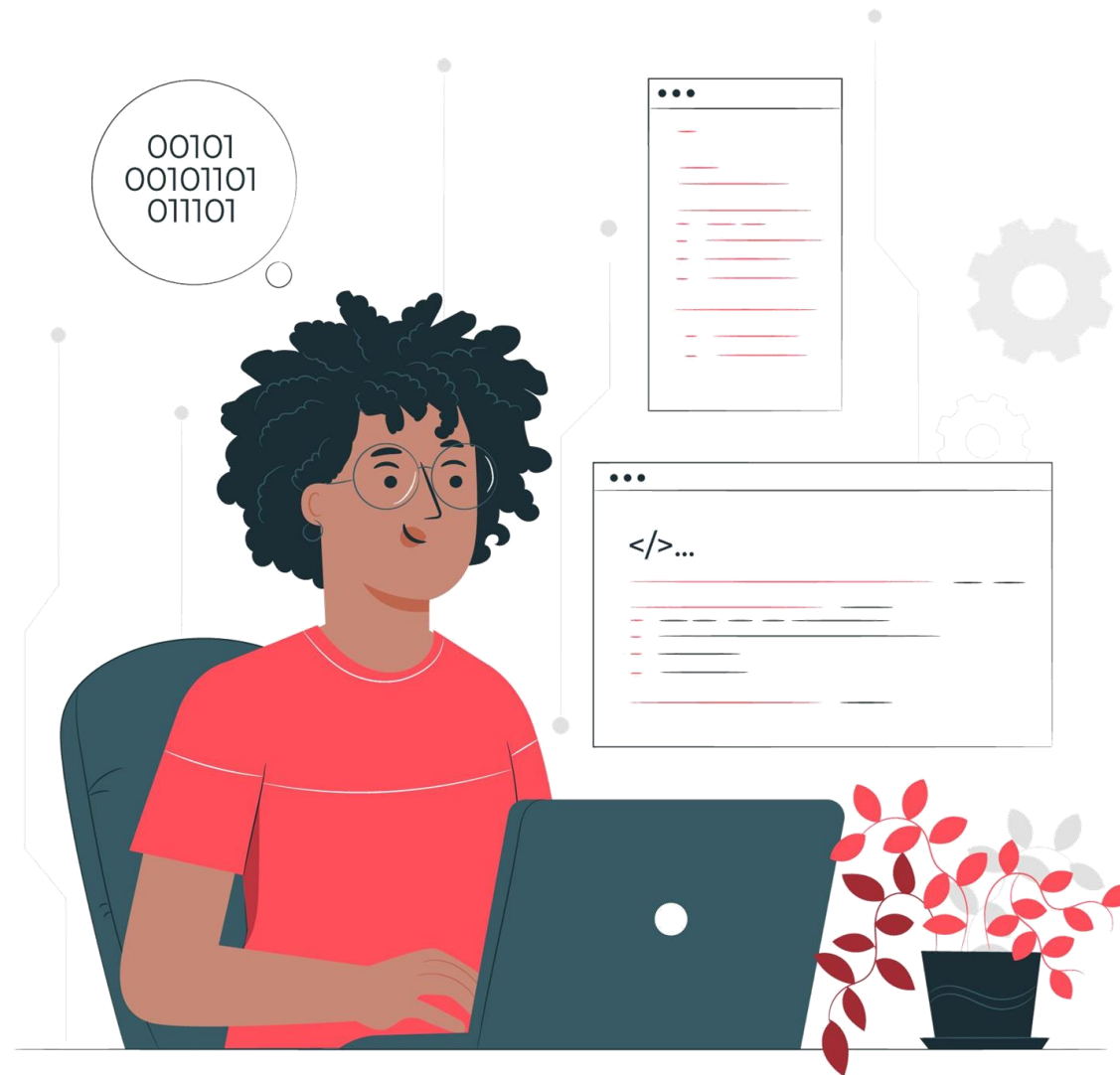

- Aggregate data
- Shape data
- Understand the statistical analysis of data
- Normalize data

Project Statement

Build necessary data aggregation, wrangling and visualization modules for Aura using the Healthcare dataset.

Dataset Description

NSMES1988.csv

| Variable	Description	Variable	Description
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)

Dataset Description

NSMES1988.csv

Variable	Description	Variable	Description
hospital	Number of hospital stays	afam	Factor. Is the individual African-American?
gender	Factor indicating gender	married	Factor. Is the individual married?
school	Number of years of education	income	Family income in USD 10000
employed	Factor. Is the individual employed?	insurance	Factor. Is the individual covered by private insurance?
medicaid	Factor. Is the individual covered by Medicaid?		

Week 1

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 their range, specifically on age and income columns.

- Export the data to JSON as NSMES1988.json format file and view and enter your comments.
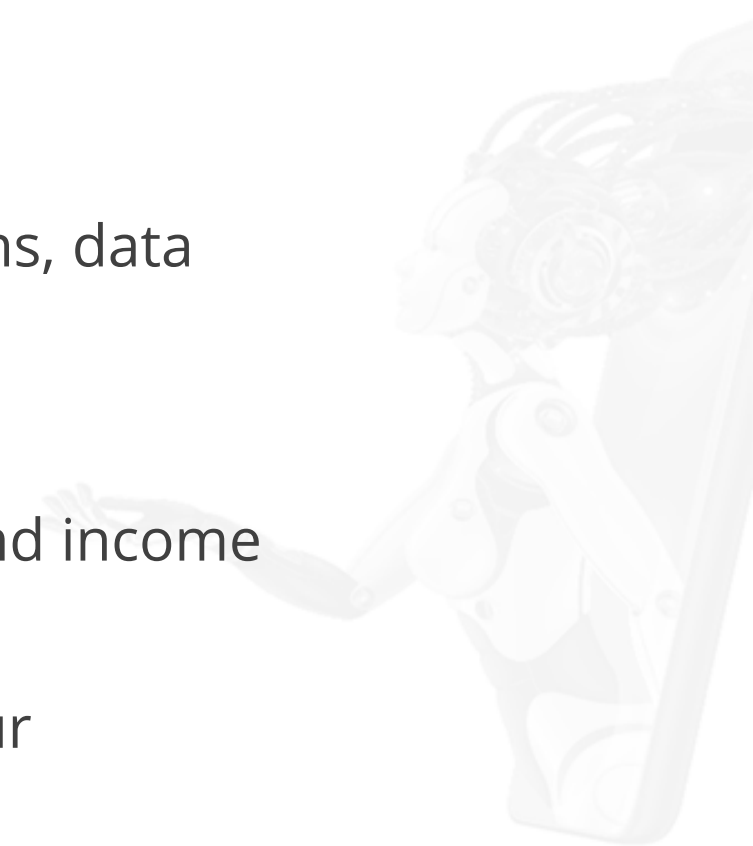
Week 1

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

- What changes you would 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 possible use in other assignments.

- Write a short report on the visual observations of the data.

Thank You