



**NANYANG
TECHNOLOGICAL
UNIVERSITY**
SINGAPORE

Centre for Professional and
Continuing Education

(SCTP) Advanced Professional Certificate **Data Science and AI**



NTU Wireless Network Login

93722830

- 1) SMS the word **register** to **93722830** to get free Wifi access for the day.
- 2) A user name and password will be sent to you via **SMS**.
- 3) Select wireless network: **NTUGUEST**
- 4) Open Web browser (IE, Fire fox, Safari). You will be directed to the page shown below.



Welcome to NTU Wireless Network

Username:

Password:

Domain:

Submit

- 5) Key in **User Name** (your phone number) and **Password** (which you have received via SMS)
(Note: Password is case sensitive. Please do not share your Username & Password with other users.)
- 6) Domain: **GUEST**
- 7) Click submit



NANYANG
TECHNOLOGICAL
UNIVERSITY
SINGAPORE

Module 2

Coaching Week 2

Jan 2025

Coaching Objectives

- Recap of lessons from past week(s)
- Assignment Q&A
- Focus on **difficult** or **interesting** topics
- Class activities

Module Overview

2.1 Introduction to Big Data and Data Engineering

2.2 Data Architecture

2.3 Data Encoding and Data Flow

2.4 Data Extraction and Web Scraping

2.5 Data Warehouse

2.6 Data Pipelines and Orchestration

2.7 Data Orchestration and Testing

2.8 Out of Core/Memory Processing

2.9 Big Data Ecosystem and Batch Processing

2.10 Event Streaming and Stream Processing



NANYANG
TECHNOLOGICAL
UNIVERSITY
SINGAPORE

Recap



NANYANG
TECHNOLOGICAL
UNIVERSITY
SINGAPORE

Assignment Q&A



NANYANG
TECHNOLOGICAL
UNIVERSITY
SINGAPORE

Focus Topics



NANYANG
TECHNOLOGICAL
UNIVERSITY
SINGAPORE

Class Activities

Mini Project 2

Web-Scraping and Data Analysis

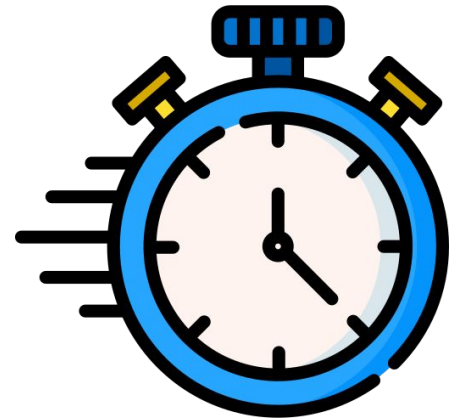
Project Objectives

- Collect data by performing web-scraping + data wrangling. Once you determined the source, make sure you are able to scrape the website successfully
- Formulate a data-science / business problem based on the data you have obtained, present your learnings and interesting take-away using appropriate visualizations.



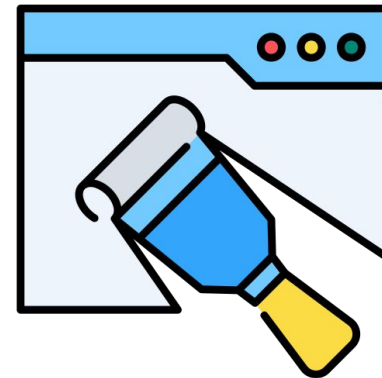
Business Case Development

- Set a strict 30-minute timer for final business case selection



Web Scraping

- Consider easier targets like product listings rather than complex data
- Use sites with cleaner HTML structure
- Assign specific scraping tasks to different team members



Presentation Time: **10 mins**

- You can use inspiration from the [ignite talks](#) framework to design your presentation (it's fun!).
- Use Google Slides or some other visual aid (Keynote, Powerpoint, etc).
- Consider the audience. Assume you are presenting to a non-technical audience.
- Start with the data science problem.
- Use visuals that are appropriately scaled and interpretable.
- Talk about your procedure/methodology (high level, CODE IS ALWAYS INAPPROPRIATE FOR A NON-TECHNICAL AUDIENCE).
- Talk about your primary findings.
- Make sure you provide clear recommendations that follow logically from your analyses and narrative and answer your data science problem.

Some Directions

1. Business Context / Data Science Problem
 - Problem statement
 - Why it matters
2. Data Approach
 - Data sources and scraping done
3. Key Findings
 - Visualization and clear learnings
4. Recommendations
 - 2-3 actionable recommendations or conclusion
 - Expected business impact

