BSBINS401 - Analyse and Present Research Information

Session 1: Introduction to the Course and Data Science

Lecturer: Jordan Hill

Learning Objectives

- Understand the course structure and expectations.
- Explore the role of data science in various industries.
- Familiarize with unit requirements and assessments.
- Initiate application for Google Cloud Credits.

Welcome to the Course!

- **Course Code**: BSBINS401
- **Course Title**: Analyse and Present Research Information
- **Focus**: Introduction to Data Visualization Tools using Python

Course Overview

- **Duration**: 18 weeks
- **Delivery Mode**: Face-to-face learning
- **Primary Tools**:
  - Python
  - Jupyter Notebooks
  - Kaggle
  - Matplotlib and Seaborn libraries

Course Structure

- **First Cycle**:
  - Work with a standard dataset provided.
  - Conduct research, analyze data, and present findings.
- **Second Cycle**:
  - Scope your own research project.
  - Find or collect your own dataset.
  - Repeat the cycle independently.

Assessments Overview

Assessment	Description	Due Date
Assessment 1: Research and Presentation on Standard Dataset	Analyze provided dataset and present findings through a report and presentation	Week 7
Assessment 2: Independent Research Project	Conduct independent research and present findings in a comprehensive report and presentation	Week 16
Assessment 3: In-Class Closed Book Exam	Complete closed book exam to assess understanding of research principles and methodologies	Week 17

Learning Outcomes

By the end of this course, you will be able to:

- **Identify research requirements and objectives.**
- **Collect, organize, and present research information.**
- **Maintain information securely according to policies.**
- **Prepare and produce research reports.**

Role of Data Science

- **What is Data Science?**
  - Interdisciplinary field combining domain knowledge, programming skills, and math/statistics.
- **Data Science in Industries**:
  - **Healthcare**: Predictive analytics for patient care.
  - **Finance**: Fraud detection and risk management.
  - **Retail**: Customer segmentation and recommendation systems.
  - **Transportation**: Route optimization and autonomous vehicles.

Why Data Visualization?

- **Understand Complex Data**: Visual representations make patterns and trends easier to comprehend.
- **Communicate Insights**: Effectively share findings with stakeholders.
- **Facilitate Decision Making**: Support data-driven decisions in organizations.

Tools We'll Use

- **Python**: Versatile programming language.
- **Jupyter Notebooks**: Interactive computing environment.
- **Kaggle**: Platform for data science competitions and datasets.
- **Matplotlib and Seaborn**: Libraries for creating static, animated, and interactive visualizations.

What is Kaggle?

- **Kaggle.com**: A platform offering datasets, competitions, and a community of data scientists.
- **Features**:
  - Access to public datasets.
  - Online coding environment.
  - Learning resources and tutorials.

Cloud Credits for Education

Why Google?: Google Cloud is the backbone of Kaggle— so if we can get some credits it will give us some extra flexibility with running our workloads.

Why AWS?: AWS has a number of great integrations with the open source AI & Machine learning community. For example, Huggingface, via AWS SageMaker.

Why Azure?: Azure has a number of great integrations with the open source AI & Machine learning community. For example, Huggingface, via Azure ML.

These three are the major players when it comes to AI & cloud computing. If you continue in this field you will likely need some form of support from these providers.

Applying for Google Cloud Credits

- **Benefits**:
  - Access to cloud computing resources.
  - Ability to run computing-intensive tasks.
- **How to Apply**:
  - Visit Google Cloud for Education.
  - Follow instructions to redeem credits.

AWS Educate

- **AWS Educate**:
  - Access to cloud computing resources.
  - Ability to run computing-intensive tasks.
- **How to Apply**:
  - Visit AWS Educate.
  - Follow instructions to redeem credits.

Activities for Today

- **Introductions**:
  - Share your background and interests.
  - Discuss what you hope to learn from this course.
- **Case Studies**:
  - Explore examples of data science impacting industries.
  - Discuss the societal implications of data science.
- **Set Up Accounts**:
  - Create accounts on Kaggle and Google Cloud.
  - Apply for Google Cloud Credits.

Getting Started with Python

- **Check Installation**:
  - Ensure Python and Jupyter Notebook are installed on your computer.
  - Alternatively, set up a cloud environment via Kaggle or Google Colab.
- **First Notebook**:
  - Create a simple "Hello, World!" notebook.
  - Explore basic Python commands.

Unit Requirements

- **Performance Evidence**:
  - Research, analyze, and present findings on at least two occasions.
  - Align with organizational requirements.
- **Knowledge Evidence**:
  - Understand organizational policies.
  - Recognize reliable and valid research practices.
  - Familiarity with research strategies and information sources.

Policies and Procedures

- **Confidentiality and Privacy**:
  - Adhere to privacy laws and organizational policies.
  - Secure handling of data.
- **Data Security**:
  - Best practices for storing and accessing information.
- **Academic Integrity**:
  - Avoid plagiarism.
  - Proper citation of sources.

Support and Resources

Lecturers		Office Hours	Resources
Jordan Hill	jordan.hill@nmtafe.wa.edu.au	Mon-Fri, 9 AM - 5 PM	• Course materials on LMS • Access to library and online databases
Hong Fu	hong.fu@nmtafe.wa.edu.au		

Out-of-Class Activities

- **Required Reading**:
  - "What is Data Science?"
  - *Python for Data Analysis* by Wes McKinney, Chapter 1: Preliminaries
- **Tasks**:
  - Ensure Python and Jupyter Notebook are set up.
  - Familiarize yourself with basic Jupyter Notebook functionality.

Any Questions?

- **Open Floor**:
  - Any questions about the course structure?
  - Concerns or topics you'd like to explore?
- **Contact Us**:
  - Feel free to reach out via email or during office hours.

Next Session Preview

- **Topic**: Data Visualization Basics and Tools Overview
- **What to Expect**:
  - Learn about different types of graphs.
  - Introduction to Matplotlib and Seaborn.
  - Hands-on activities with Jupyter Notebooks.
- **Preparation**:
  - Complete the required readings.
  - Bring your laptops for in-class exercises.