

Yixuan Wang				
No.	Activity	Time	Challenges and Solutions	
			Challenges	Solutions
1	Datasets Selection	4h	1. Research topic selection challenges: dataset and diverse expectations	1. Group discussion, internet dataset exploration.
2	Download and Merge Yearly CSV Files Spanning 15 Years	1h	1. Efficient file merging 2. Decoding errors for applying Pandas.read_csv()	1. Concatenate DataFrames using pandas.concat() with map() and Lambda functions 2. Specify "encoding='unicode_escape'" for file reading
2	Store and Retrieve CSV File in/from MySQL Database	3h	1. Forgot 'root' password in MySQL 2. Connecting MySQL database and write file	1. Reinstall MySQL server 2. Use dataframe.to_sql
3	Adjust Data Types, Rename Columns and Standardise Formats	1h	1. Day-month swap issues in pd.to_datetime() 2. Incorrect Price Formatting representation in pandas 3. Duplicate descriptions due to mixed English and Irish labels in columns.	1. Set "dayfirst=True" 2. Use .str.lstrip(), .str.replace() and .astype() for DataFrame column. 3. Apply DataFrame.replace() function to standardise columns.
4	Exploratory Data Analysis and Visualisation	12h	1. Undesirable format and colour scheme in matplotlib pie chart. 2. Year-only labels, unappealing tick marks for line charts.	1. Read documentation, explore matplotlib colour options and customise pie chart parameters. 2. Apply MonthLocator, DateFormatter from matplotlib.dates module for x axis; set tick_params() to customise ticks.
5	Forecasting Model Development	12h	1. Forecast methodology selection 2. Training, validation and forecast data selection. 3. Slow gradient descent convergence 4. Poor performance	1. Experiment with the moving average and first differencing techniques for daily and monthly data. Identify the most effective approach by comparing the performance of different models. 2. Implement rolling window techniques by slicing appropriate length periods and use Tensorflow Dataset objects to prepare the forecast and training datasets. 3. Normalise data with MinMaxScaler 4. Switching the activation function from 'relu' to 'tanh', and incorporating Dropout layers and Layer Normalization significantly enhanced the performance.
6	Code Selection for Submission	2h	NA	NA
7	Preparing Report and Journal and Presentation	5h	NA	NA