BI COURSE FIRST HALF

Terms, Themes, Techniques, and Tools

Python

Introduction to Python and iPython

- Anaconda and Jupyter
- language syntax
- program structures
- functions and packages for working with data
- data visualisation packages
  - `matplotlib`
  - `seaborn`
  - `plotly` (includes 3D plots)
  - `folium` (for geo data and maps)
  - `PyGWalker` (for operations and visualisation of data cubes)

Data

Data in BI

- data types
- data structures in Python – data frames, numeric arrays, lists, maps, dictionaries
- converting structures from one into another (e.g. data frame to array and vice versa)
- data cubes, dimensions and measures
- dependent and independent variables

Data Ingestion and Cleaning

- Reading data from external sources into Python program
- Getting to know your data – shape, attributes, types
- Cleaning the data
  - dealing with missing values
  - dealing with wrong or damaged values
  - binning (rows)
  - grouping (columns)
  - dropping columns and creating new
  - joining data frames
- Data Visualisation
  - scatter plots
  - bar chart, pie charts

Data Exploration and Wrangling

Statistic

- Descriptive statistics
  - data distribution (normal distribution)
  - statistical measures
    - central tendency measures (mean, mode, median)
    - measures of dispersion (min, max, range, standard deviation, variance, quantiles, quartiles, outliers, Z-value)
- Correlation
  - covariance
  - correlation
  - correlation matrix
- Visualisation
  - histograms
  - box-whisker plot
  - heat map

Feature Engineering

- encoding categorical data with numeric codes
  - label encoding – with category numbers
  - one-hot encoding – with zeros and one one
  - count (occurrences) and target encoding (with specific values)
- detecting and handling of outliers
- scaling the data
  - standard scaler
  - data normalisation
- *PCA – principal components analysis -postponed*

Reporting Results from Data Analysis

- Design of reports
  - data visuals – 2D and 3D charts, animation, geo-maps
  - dashboards – as a collection of diagrams and narratives
  - data stories – as a collection of dashboards
- *Publishing platforms - postponed*
  - Streamlit
  - Tableau