LESSON HANDOUT

The Data Analytics Process

There is a four-step process to your analysis:

- **Define** - where we understand the problem and the available data.
- **Transform** - where we extract and prepare the data.
- **Analyse** - where we determine our findings.
- **Communicate** - where we present the data to the customer.

This process isn't strictly linear, meaning that you might frequently revisit previous stages of the process.

Define

During this stage, we plot the purpose of the analysis by structuring the problem in data-related terms. This stage might involve:

- Understanding what data is available and where it is stored.
- Validating the data for accuracy and completeness.
- Conduct a preliminary analysis of the dataset to determining limitations/opportunities.
- Develop a timeframe for the solution.

Transform

- **Source/extract** - source the data.
- **Join** - combine different datasets together.
- **Clean** - fix errors and inconsistencies in the data
- **Manipulate** - transform the data into a format ready for analysis

Analyse

In this stage, we take our prepared data, aggregate it (or summarise it) and apply statistical analysis. This stage might involve:

- **Classification** - add labels to the data to enable useful analysis
- **Aggregation** - summarise the data into meaningful values and metrics
- **Segment** - divide data into relevant groups (segments) and subsequently analyse them

Communicate

Finally, at the communication stage, we turn our findings into reports which answer the questions identified in the define stage (as best as possible) and are pitched correctly to the customer. Remember, the analysis must always have a purpose, and your customer must be able to understand, and ideally action, your findings.