# Data Ethics & Applied Analytics Project

Anti-money laundering projects at a global bank

*Case Study 5*

# Project Overview

## Context

Pig E. Bank is a hypothetical well-known global bank with an anti-money-laundering compliance department.

## Objective

Provide analytical support to a variety of data-related projects that help the bank assess client risk and transaction risk. Help building and optimising models that assist the bank in running their compliance program more efficiently.

## Key Topics

- The characteristics of big data, how data analysts use big data, and the challenges of
- Extracting knowledge from big data
- The impact of data bias and ethics on how data is used, shared, collected, and protected
- The fundamentals of data mining, including techniques for data mining and how it drives decision-making
- Predictive analysis and models such as linear regression
- Time-series analysis and time-series forecasting

## Data & Tools

Understanding of advanced data analytics concepts & processes based on fictional scenarios.

No actual data was processed.

# Discussion of Advanced Data Analytics Topics

- What does the **CRISP-DM methodology** for data mining entail?
- Explain what **time series analysis** involves
- Discuss the **main principles of data ethics**, as well as what companies or organisations could do to uphold these principles.
- If you had concerns about bias in the way a predictive model was measured. How would you raise your concerns with your manager? How could **measurement bias** be prevented in the future?
- Explain the **difference between regression and classification models** in predictive analysis.

# My Project Reflections

*Having worked in marketing analytics for several years, the topic of data privacy mostly came up as a somewhat annoying constraint to deriving more insights on customer behaviour. Looking at this topic again from the perspective of big data in finance, I found it very eye-opening and absolutely crucial to consider data ethics in regards to data bias, security and privacy.*

*I think there are certain steps & processes in data analytics that can be neglected because they are time-consuming and slow you down ( e.g. meticulous documentation or uncomfortable discussions), but without knowing where the data comes from, how it was processed and if it was checked for bias and otherwise evaluated, there can be no proper understanding and valid conclusions to any analysis.*