Assignment 1 Guidelines

Machine Learning and Artificial Intelligence in Finance
FINC 5322
Fall 2025
The American University in Cairo
Onsi Sawiris School of Business
Heikal Department of Management

Customer Churn Prediction – Assignment 1

In business, customer churn (or attrition) refers to when customers stop doing business with a company.

For example, in a subscription-based telecom company, a customer is considered churned if they cancel their subscription.

Predicting churn is critical because:
- Retaining an existing customer is often cheaper than acquiring a new one.
- Companies can take action (targeted campaigns, discounts, improved services) to prevent churn.
- Understanding who is likely to churn helps businesses allocate resources more effectively.

In this assignment, you will work with a telecom churn dataset. Your goal is not to build the most accurate model but to practice the full data preparation pipeline: exploring, cleaning, engineering features, and running a simple model.

Accessing the Assignment
- All Assignments are hosted on our course Code Repository (GitHub): https://github.com/kelkess43/AUC-Material
- The Assignment will also be shared through Canvas modules for easy access

---

Structure of the Assignment
Your notebook is divided into five sections.
You must complete the TODOs in the notebook and keep all work within the. ipynb file.

Section 1: Exploratory Data Analysis (EDA)
- Count the number of unique customers.
- Check for missing values.
- Visualize numeric distributions (tenure, charges)
- Visualize churn distribution (how many customers stayed vs churned).

Section 2: Data Cleaning
- Convert TotalCharges to numeric and handle missing values.
- Remove duplicate rows.
- Reflect briefly: Why did you choose your method for handling missing values?

Section 3: Feature Engineering
- Create a tenure_group feature (e.g., 0–12 months, 13–24 months, etc.).
- Encode categorical variables into numeric form.

Section 4: Simple Modelling
- Perform train/test split (70/30).
- Train a Logistic Regression model.
- Evaluate with accuracy, precision, and recall.

Section 5: Research & Reflection
- Explain why accuracy is not enough in churn prediction.
- Research and explain in your own words what precision, recall, and the classification report mean.

Deliverables and Submission
- Submit one Jupyter Notebook (.ipynb) file with all your work.
- Make sure all cells are <u>executed</u>, and outputs are <u>visible</u>.
- Add short comments/markdown where required
- Deadline: Tuesday, 7th October at 11:59 PM
- Submission: Upload your completed. ipynb notebook on Canvas

---

Grading
- EDA (20 points)
- Cleaning (20 points)
- Feature Engineering (20 points)
- Modelling (25 points)
- Research & Reflection (15 points)