PY: Project Instructions

Created	@October 1, 2024 3:51 PM
Class	Python

Deliverables

The following deliverables are expected from each group:

- A 5–10 page report detailing your main findings.
- The source code of your exploratory data analysis:
  - Hosted in a GitHub or GitLab repository (preferably), or as a zip archive.
  - Both notebooks and Python scripts are accepted.
- An oral presentation of your work and main findings (date TBA):
  - 10 min presentation.
  - 5 min questions.

Task 1	# Rows	# Features
Customer Churn Analysis (Banking)	10000	13

Context

A european bank wants to identify the main factors contributing to customer churn (i.e., customers leaving the bank and closing their accounts). By doing so, the bank can target such customers with incentives, or use this knowledge to propose new products that are better suited to their needs.

Features

The dataset contains the following features.

Feature	Description
`CustomerId`	The customer's unique identifier.
`Surname`	The customer's last name.
`CreditScore`	The customer's credit score.
`Geography`	The customer geographic location (country).
`Gender`	The customer's gender.
`Age`	The customer's age.
`Tenure`	The number of years the customer has been a client of the bank.
`Balance`	The customer's bank account's balance.
`NumOfProducts`	The number of products the customer contracted with the bank.
`HasCrCard`	Whether the customer has a credit card or not.
`IsActiveMember`	Whether the customer has been recently active (i.e., made transactions) or not.
`EstimatedSalary`	An estimate of the customer's annual salary.
`Exited` (target)	Whether the customer closed their account (1) or not (0).

Questions

- What is the churn rate among the bank's customers?
- How are the different variables (gender, age, geography, etc.) distributed in the dataset?
- How do the different variables interact with each other?
    - What are the age, salary, balance, number of products, etc. distributions for each gender group?

    - How are different indicators distributed by country?
    - etc.
- How do the different variables affect churn? What are the causes that can lead to increased (or reduced) customer churn?
- (Optional) Build a simple machine learning classification model that predicts churn based on a customer's features.