

IFT3913 QUALITÉ DE LOGICIEL ET MÉTRIQUES – HIVER 2022 – TRAVAIL PRATIQUE 2

- Travail à remettre le **vendredi 18 mars 23h59** via StudiUM. Aucun retard ne sera accepté.
- Travaillez en groupes de 2. Soumettez un zip nommé comme suit : *prenom1_nom1_prenom2_nom2.zip*
- Le membre de l'équipe qui ne soumet pas le fichier Zip principal de la soumission doit soumettre un zip (même nom) avec le fichier *readme.txt*, qui devrait contenir les noms de l'équipe.
- Vous pouvez utiliser n'importe quel logiciel qui appuie les opérations statistiques requises (Excel, R, Octave, SPSS, Python, Java, ...). Cependant, nous n'offrons pas de conseils sur l'utilisation des logiciels statistiques. **Ajoutez votre code/script/fichiers dans le Zip de votre soumission.**
- Le livrable le plus important est le **rapport**, en format PDF. Assurez-vous de communiquer clairement.
- Rapports de plus de 10 pages (12p, simple espace) vont être éliminés d'office.

TRAVAIL À RÉALISER

Ce travail a pour objectif **d'étudier** un échantillon de données fourni dans le fichier *jfreechart-stats.csv*. Cet échantillon donne les valeurs de métriques collectées par les classes du dossier *main* du JFreechart:

- NCLOC : nombre de lignes de code qui ne sont pas ni vides ni commentaires
- DCP : densité de commentaires (CLOC/LOC) donnée en pourcentage
- NOCom : nombre de *commits* (combien de fois la classe a été changée) dans l'historique git de la classe
- WMC : la métrique de complexité *Weighed Methods per Class*

Les tâches à effectuer :

- T1. (15%) **Visualisez** chacune des métriques de l'échantillon en créant les boîtes à moustaches. Calculez les informations pertinentes et décrivez les distributions.
- T2. (25%) **Évaluer l'hypothèse** selon laquelle *les classes qui ont été modifiées plus de 10 fois sont mieux commentées que celles qui ont été modifiées moins de 10 fois*. Décrire d'abord la conception de l'étude et discuter par la suite les résultats. Suivez les étapes d'une étude empirique (choix d'étude, énoncé des hypothèses, définition des variables, interprétation et généralisation des résultats, discussion des menaces à la validité).
- T3. (15%) **Étudier les corrélations** entre NCLOC et WMC, DCP et WMC, NOCom et WMC. Visualisez les données, les droits de régression, etc., et expliquez pourquoi (ou pourquoi pas) ces visualisations sont significatives (ou pas). Dans cette étape, vous ne prenez pas de décisions: vous explorez et vous étudiez l'ensemble de données.
- T4. (30%) **Évaluer les hypothèses suivantes :**
 - a. WMC est une fonction linéaire du NCLOC
 - b. WMC est une fonction linéaire du DCP
 - c. WMC est une fonction linéaire du NOComDécrire d'abord la conception de l'étude (comme en T2) et discuter les résultats par la suite.
- T5. (5%) **Décrivez vos conclusions** dans un court paragraphe.

La qualité du rapport vaut 10%.