Statement of Independent Work

1. Declaration of Original Work. By entering our Student ID below, we certify that we completed our assignment independently of all except as stated below.

Signed, Olivia (Alex) Xiao (E1497248), Shen Tianwei (A0313045U), Joshua Harsha Dass (A0217678N), Rayan Maknojia (A0313081U), and Roderick Kong Zhang (A0286550Y).

2. References. We give credit where credit is due. We acknowledge that we used the following websites, AI model, or contacts to complete this assignment:

Data Cleaning

- Kaggle website. https://www.kaggle.com/code/rikdifos/eda-vintage-analysis/notebook, for exploratory data analysis and vintage analysis.
- Machine Learning Mastery website. https://machinelearningmastery.com/one-hot-encoding-for-categorical-data/, for the code implementation for one-hot encoding.

Feature Extraction

- 365 Data Science website. https://365datascience.com/tutorials/python-tutorials/pca-k-means/, for the code implementation for PCA.
- Statology website. https://www.statology.org/scree-plot-python/, for the code implementation of PCA scree plots.
- Built In website. https://builtin.com/machine-learning/pca-in-python, for the code implementation for feature extraction using PCA.
- GeeksforGeeks website. https://www.geeksforgeeks.org/ml-linear-discriminant-analysis/?ref=rp, for the code implementation of LDA.
- DataCamp website. https://www.datacamp.com/tutorial/tutorial-lasso-ridge-regression, for the code implementation for Lasso regression.

Machine Learning Models

- Towards Data Science website. https://towardsdatascience.com/how-to-use-unsupervised-learning-to-cluster-well-log-data-using-python-a552713748b5/, for the code implementation for clustering.
- GeeksforGeeks website. https://www.geeksforgeeks.org/linear-regression-python-implementation/, for the code implementation for linear regression.
- GeeksforGeeks website. https://www.geeksforgeeks.org/ml-logistic-regression-using-python/, for the code implementation for logistic regression.
- EE2211 Introduction to Machine Learning lecture materials and code. For the explanation and code implementation for polynomial ridge regression.
- GeeksforGeeks website. https://www.geeksforgeeks.org/decision-tree-implementation-python/, for the code implementation for decision trees.
- GeeksforGeeks website. https://www.geeksforgeeks.org/random-forest-algorithm-in-machine-learning/, for the code implementation for random forests.

Evaluation Metrics

- GeeksforGeeks website. https://www.geeksforgeeks.org/gap-statistics-for-optimal-number-of-cluster/, for the code implementation of gap statistics.
- GeeksforGeeks website. https://www.geeksforgeeks.org/calculate-roc-auc-for-classification-algorithm-such-as-random-forest/, for the code implementation for ROC curves.
- Sin-Yi Chou website. https://sinyi-chou.github.io/python-sklearn-precision-recall/, for the code implementation for PR curves.

Debugging

- ChatGPT, for debugging Python code.