**SEHS4697 Big Data and Cloud Analytics**

**Group Project (2024/25 Semester 2)**

**Group ID: 02 (4 - 5 students in each group)**

|  |  |  |
| --- | --- | --- |
| **Student Name** | **Student ID** | **Signature** |
| Leung Man Yuk | 23068421s | Leung Man Yuk |
| Wong Wang Bun | 23063960s | Wong Wang Bun |
| Cheung Ho Kin | 23054127s | Cheung Ho Kin Jack |
| Fok Lik Hang | 23052447s | Fok Lik Hang |
| Zhang Chengtao | 23059264s | Zhang Chengtao |

**Project Title:**

House price analysis: Compare with the ensemble models for the Ames Housing Market during data analysis tactic

**Abstract: A brief abstract of the proposed research problem and choice of dataset, no more than 250 words.**

This study offers the prediction of residential property prices critical for diverse stakeholders, including homebuyers, real estate agents, lenders, and policymakers in an era where housing markets drive global economies. Research illustrates the gap between traditional appraisal methods (e.g., comparative market analysis) and data-driven approaches, with the purpose of enhancing transparency and scalability in price estimation.

Choice of Dataset:   
This study uses Ames Housing dataset. This a comprehensive collection of data on the residential property sales in Ames, lowa, from 2006 to 2010. This dataset includes 79 different variables and sale prices for 2,930 residential properties, ideal for multiple layer data analysis. This dataset also covers fine-grained features of categorical (e.g., neighbourhood zoning, house style) and numerical variables (e.g., lot size, year built), which provides a solid foundation for the exploration of multidimensional data integration strategies.

Secondly, as a recognised benchmark dataset in the field of real estate analytics, the Ames data supports direct comparisons with established research. Its public availability guarantees a transparent and reproducible analysis process.

Further, compared to the Boston house price dataset (13 features) or the California house price dataset (8 features), the Ames dataset has higher data richness (e.g., basement/garage details), which is crucial for our study.