KNN Project Proposal

Zara Tekmen & Margaret Jackson

Potter IB CS HL

**Project Description**

The "Salary Prediction Dataset" is meant to predict the integer value of a person's yearly salary based on various factors.

[Kaggle Link](https://www.kaggle.com/datasets/mrsimple07/salary-prediction-data) [Google Sheet Data Set](https://docs.google.com/spreadsheets/d/19MFEB6VyU_PNyE5tB3wATJ5RWmgr65xfKa_ucBQiguo/edit#gid=1599316116)

**Data Values:**

* Education: String value indicating the highest level of education by the individual
  + Possible Values: High School, Bachelor, Master, PhD
  + String values will be converted to integer values ranked by level of degree (HS = 0, Bachelor = 1, etc.)
* Experience: Integer value representing years of professional experience by the individual
  + Integer values range from 1 to 29
* Location: String value indicating the geographic location of the individual
  + Possible Values: Urban, Suburban, Rural
  + String values will be converted to integer values ranked proximity to city (Urban = 0, Suburban = 1, Rural = 2)
* Age: Integer value representing the age of the individual
  + Integer values range from 20 to 64
* Gender: String value indicating identified gender of the individual
  + Possible Values: Male = 1, Female = 0
* (**Target**) Salary: Floating point value representing annual salary of the individual
  + Floating point values range from $33,510.51 to $193,016.60