**IB Computer Science HL - KNN Project**

**Salary Predictions Based on Personal Attributes - Margaret Jackson and Zara Tekmen**

Introduction

**Research Objective**

In this project, we are exploring salary data in order to better predict salary for future hires. The list of people given are described by the following criteria: Education Level, Experience, Location, Job Title, Age, Gender, and Current Salary.

Our research objective is to use a kNN model trained on the salary dataset to give a salary number based on a person’s attributes. The kNN model will measure similarity of attributes to predict a salary for a given employee. This will help employers more accurately compensate their employees based on their years of experience, degree, etc. To avoid bias in salary prediction, we will build a kNN model. When an employer needs to make a salary offer, they can use the model to find comparable employees to base their salary offer on.

**Research Questions**

1. Can we use a limited set of personal attributes: Education, Experience, Location Type, Age, and Gender to predict the salary a person will earn?
2. Can we optimize the predictions of a person's salary by comparing alternative distance algorithms - Manhattan, Euclidean, and Cosine?

**Data Source**

The dataset list of salaries contains string values for Education, Location, Job Title, and Gender as well as numerical values for Experience, Age, and Salary. This dataset was sourced from kaggle.com. We chose this data because we are soon entering the workforce and we wanted to explore the predictions this model makes, especially for people with little experience like us. The dataset was selected because it has attributes that are relevant to our questions. Original data source: <https://www.kaggle.com/datasets/mrsimple07/salary-prediction-data>.

The dataset needed to be altered to be able to run through our kNN model. We needed to remove the the

**Altered Data Source:**

Three key attributes