**Final Project Preliminary Proposal**

For my DATA512: Human-centered Data Science subject, I am interested in researching and analyzing into the topics of People Analytics. I am interested in looking into what are the key reasons that make an employee leave and if there are possible indicators (existing or engineered during the course of analysis) that can help us predict employee attrition beforehand.

My Motivation to pursue this project comes from my 2 years of professional work experience as a Business Analyst for OYO, a travel & hospitality related startup. I handled their Sales Operation Analaytics and was the owner for Incentives and Performance Management project for the entire 2 years. With this project, I want to further explore into the causes of employee attrition and if possible identify indicators leading to it. This project is interesting from any organisation's purview because companies not only money but time and resources in grooming and training their employees. A single employee's resignation is not only a loss of head count but also a loss in terms of times, money and bandwidth. I also found this topic to be in line to what I learnt about human centred data science from this class. This project can provide useful insights into what people value the most in their job: salary, evauluation or work quality. I hope to generate insights that can be readily utilized by an organisation to strengthen their decision in terms of improving employee satisfaction, and boost their enthusiasm if found low.

**Data**

For my analysis, I wanted a simpler data and one publicly available. I came across this Kaggel Data Challenge: **HR Data for Analytics**. Link to the dataset: <https://www.kaggle.com/jacksonchou/hr-data-for-analytics>.

The data is publicly available and under the CC0: Public Domain license.The dataset contains employee profiles of a large company, where each record is an employee. It has the follwoing columns available:

* satisfaction\_level
* last\_evaluation
* number\_project
* average\_montly\_hours
* time\_spend\_company
* Work\_accident
* left
* promotion\_last\_5years
* sales
* salary

**Unknown Dependency**

CThe above dataset contains only 10 features and has the information for a limited time duration only. So, our analysis will be limited to the insights from these 10 features only. But I belive, this should still give us important insights and answer our research questions.