Project Report

# GitHub URL

<https://github.com/mariemcsweeney/UCDPA_mariemcsweeney>

# Abstract

(Short overview of the entire project and features)

# Introduction

(Explain why you chose this project use case)

Personal interest in the most valuable resource to a company – people. People analytics is key to bolstering strategic business partnership. I also have a personal interest in the organizational power of people. It has been proven (cite reference) that organisations that undertake People Analytics financially outperform those who do not.

# Dataset

(Provide a description of your dataset and source. Also justify why you chose this source)

Kaggle

# Implementation Process

(Describe your entire process in detail)

Data – Brought in Kaggle CSV

Importing – Brought in Gallup Survey on Employee Wellness data

Preparation and Analysis

* Create Pandas DataFrame
  + Called DataFrame – employee\_data
  + Called Wellness\_df
* Sorting, Indexing, Grouping
  + Be able to compare the IBM Work Life Balance score based on age and gender to Various Aspects of Life in Balance on the Global Wellness Score and Job Satisfaction Score in IBM data (Column Q) to Column AS to AV (Enjoy the work you do every day) by putting a pivot table through python and then merging this pivot table with the dataframe on the global wellness index – then do a line graph to visually indicate if a linear relationship
  + Employee data -> Added an index of years\_comp to represent Number of Years at the company
  + Employee data -> Grouped by Department and ungrouped
  + Employee data -> Sort by Column AF, most tenured to least tenured employee
* Drop Duplicates, replace missing values
  + Employee data -> Employee # 2068 is a duplicate, removed this to prepare the data, line 1472. This code is not specific to employee #2068 as it simply sorts by the employee number and then drops the duplicate and in this way, is reusable code.
  + Employee data -> Column V missing some entries of ‘Over 18’, as column A has the employee ages, we can interpolate data of ‘Y’ for these, could use a fill statement but as it is conditional, I use a ‘loc’ function.
* Merge DataFrames
  + Brought in inner join using a pivot table and merged on ‘EnjoyJob’
* Covered Conditional statements by the way we added a new age category, for ‘if’, ‘then’ rules, we assigned a category based on the value of age,
* Groupby – we pivoted up the age categories and grouped together the higher elements of the job satisfaction scores to broadly suggest if the employee enjoyed their job.

# Results

(Include the charts and describe them)

**Visualisations**

# Insights

Engagement and Attrition analysis

Comparing to a benchmarked Global Wellness Index, the employees are well below the job satisfaction levels of the global population, in all age groups, but in particular

45-64 Age Group, with 59% reporting a 3 or 4 in job satisfaction and 86.1% of the general population reporting job satisfaction. In IBM, there is actually an inverse impact of job satisfaction vs general population – the older an employee gets, the less satisfied they are with their role. As a result, I would make a recommendation to Senior Management to target initiatives towards all groups around engagement

start with those particularly in mid-career in particular that might need their needs at their particular age and stage of life. It is important not to let any biases around their attitudes to work cloud any grasps from the data.

A screenshot of a computer

Description automatically generated

It is intuitive to believe that an employee’s job satisfaction has a linear relationship with likelihood of attrition. In particular, IBM’s management group should take note of an employee’s satisfaction rating of 2.47 or below as this is a high predictor of the employee leaving the business and as a general rule, People Managers and/or HR should focus in on an individual employee. One suggestion would be to review the telecommuting policies and work-from-home flexibility – the distance the employees have to travel is an even more meaningful predictor of an employee leaving than their own perception of work/life balance. It is hard to state from the data the commuting time impact on work/life balance due to the variability (so suggesting a number of other factors have an influence on the employees perception of his/her work/life balance). That being said, the commute does appear to have a significant impact for some.

Broadly looking at a three-way visualization between attrition, job satisfaction and job level, the more senior an employee is, the less likely they are to turnover. An interesting exception to this is in the mid-level job level. This insight is predictive in nature and may identify some key personnel at risk due to low job satisfaction – it may potentially signify mid-level manager burnout – is there enough support being put in place so they in turn, can develop and support others? This is a recommendation for HR for further investigation.

Performance

The relationship between the amount of times training was implemented last year and the performance rating is a negative correlation in the aggregate. This would potentially indicate that mangers at IBM spend a lot of their training resources/budget etc augmenting the knowledge of underperformers to get them to a basic level of understanding or fill in particular gaps but then no more. This finding appears to be unique to those who work in Research & Development department.

Conversely, however, those who got the highest performance ratings, also were in receipt of the most training and this effect was magnified in Sales roles.

Suggesting to HR that Learning & Development department should market themselves as training for continuous growth mindset, instead of just filling in the gaps. Some cross-functional learning may also be relevant as any sales training appears to be an investment.

Overtime in the aggregate does not have any effect on an employee’s performance. The visualisation indicates that it may do as much harm as good as there are those doing overtime that have low performance and high performance and the opposite. This may signify to IBM that in knowledge-type roles, working longer hours doesn’t actually produce higher results and HR could do analysis on employee’s overtime to understand the factors driving that and put supports in place to reduce it as the cost of overtime does not appear to be recouping many or any benefits.

Such as the analysis I discovered while looking through the data to identify concerns around attrition, I see a similar pattern when appreciating data around performance. Broadly speaking, there is a slight upswing in performance ratings the more senior an employee but again, we see a dip in performance in the mid-level job role. On deeper analysis, this appears to be very prevalent within the HR department. Suggesting therefore Senior leadership within HR, open communication lines with mid-level employe

# Machine Learning Questions

# References

**There** Datacamp – Podcast ‘DataFramed’ episode -Kaggle and the future of datascience

UCD Class Notes

[How people analytics is transforming the HR landscape | McKinsey](https://www.mckinsey.com/capabilities/people-and-organizational-performance/our-insights/how-to-be-great-at-people-analytics)

[Using people analytics to drive business performance: A case study | McKinsey](https://www.mckinsey.com/capabilities/quantumblack/our-insights/using-people-analytics-to-drive-business-performance-a-case-study)

[Organizational Health Index | McKinsey & Company](https://www.mckinsey.com/solutions/orgsolutions/overview/organizational-health-index)

Kaggle.com – level up with the largest AI and Machine Learning Community

[HR Analytics Case Study | Kaggle](https://www.kaggle.com/datasets/vjchoudhary7/hr-analytics-case-study)

[5 Ways Technology Can Help Build a Strong Company Culture - Spiceworks](https://www.spiceworks.com/hr/engagement-retention/articles/5-ways-technology-can-help-build-a-strong-company-culture/)

Podcast – the path to becoming a Kaggle Grandmaster – ‘DataFramed’ the podcast by DataCamp

[IBM HR Analytics Employee Attrition & Performance | Kaggle](https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset)

[Collecting HR Data With Web Scraping (Why It Matters) | Scraping Robot](https://scrapingrobot.com/blog/hr-data/)

<https://dummyapis.com/dummy/employee>

note:data is always random]

<https://www.gallup.com/file/education/468923/Global%20Wellbeing%20Initiative%20Dataset.xlsx>

<https://www.gallup.com/467702/indicator-employee-retention-attraction.aspx>

**are no sources in the current document.**