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

Importing

Preparation

* Create Pandas DataFrame
  + Called DataFrame – employee\_data
* Sorting, Indexing, Grouping
  + Added an index of years\_comp to represent Number of Years at the company
  + Grouped by Department and ungrouped
  + Sort by Column AF, most tenured to least tenured employee
* Drop Duplicates, replace missing values
  + Employee # 2068 is a duplicate, removed this to prepare the data, line 1472
  + 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

**Analysis**

* Conditional statements, looping, groupby
* Define a custom function to create reusable code
* Use NumPy functions
* Dictionary or Lists

# Results

(Include the charts and describe them)

**Visualisations**

# Insights

(Point out at least 5 insights in bullet points)

Drivers of high performance (star rating, uptake of training & EA, employee engagement survey, stock options, relationship satisfaction, years with current manager). percent salary hike

Factors causing attrition (timesheet, hours, manager satisfaction, department, employment satisfaction, gender, job involvement, education field, business travel, work/life balance, years at company, current role, overtime, percent salary hike, number of companies worked, marital status)

Predictive factors, employee tenure (does it max out, engagement survey scores),

Identify high potential employees for career advancement opportunities, establish clear career pathways, foster a culture of open communication,

What are mid-level managers doing – is enough being put into them? Are they burned out? If manager has an impact on employee attrition, is enough support being put in place into management to prevent burnout, so they can support and develop others.

Call out potential biases

# Machine Learning Questions

# References

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/)

