Munich Personal RePEc Archive

Occupation-personality fit is associated with higher employee engagement and happiness

McCarthy, Paul X. and Kern, Margaret L. and Gong, Xian and Parker, Michael and Rizoiu, Marian-Andrei

1 June 2022

Online at https://mpra.ub.uni-muenchen.de/114458/
MPRA Paper No. 114458, posted 07 Sep 2022 00:41 UTC

Occupation-Personality Fit is Associated with Higher Employee Engagement and Happiness

Paul X McCarthy,

UNSW Sydney

Computer Science and Engineering

paul@onlinegravity.com

Margaret L. Kern

Melbourne Graduate

School of Education

The University of Melbourne

Peggy.Kern@unimelb.edu.au

Xian Gong

UTS Data Science Institute,

University of Technology

Sydney

Michael Parker

Praxis Communication

Sydney, Australia

Marian-Andrei Rizoiu

UTS Data Science Institute,

University of Technology

Sydney

Keywords: Personality, Occupation-fit, Engagement, Wellbeing.

OCCUPATION-PERSONALITY FIT IS ASSOCIATED WITH HIGHER EMPLOYEE ENGAGEMENT AND HAPPINESS

ABSTRACT

Using large scale data sets about Australians (n=99,897) active on social media in a variety of occupations (n=624) across all industries, we used a variety of linguistic analysis techniques to infer user's happiness, engagement and Big5 personality traits across 30 dimensions, as well as their occupational-personality fit when compared to others in the same role.

We found that: (a) when roles are clustered by the personality traits of those in them there appears to be eight groups or 'tribes' made up of roles with similar personality trait combinations; (b) happiness, as measured by inferred current happiness, is positively correlated with occupation-personality fit and; (c) engagement is significantly correlated with occupation-personality fit and can explain over 25% of the variance in engagement in a sample of 18k people across 624 roles.

These findings show that occupation and personality fit play a material and significant role in employee engagement, which in turn is known to have many firm-level and economy-wide outcomes.

INTRODUCTION

Ensuring employees are in roles that are aligned with the strengths of their personality, and preferred mode of engaging with the world, has a natural and intuitive appeal for both organisations and individuals alike.

While there is increasing research showing beneficial links between the matching of occupations suited to people's personality in specific occupations or industries such as accountants, school teachers and veterinary surgeons (Dole and Schroeder 2001; Kokkinos 2007 and Dawson and Thompson 2017) to date there has not been any at-scale evidence that aims to comprehensively explore these links across a broad range of occupations.

There is a large and growing body of research that explores occupational fit in terms of skills, experience or cognitive fit. Yet none of these directly focus on who we are as individuals. Many of these features, even mental ones like cognitive ability, have known gaps in their ability to predict fit and future performance.

For example, cognitive ability and previous academic performance are a good, but not perfect, predictor of achievement in many key roles. Doctors, for example, in many countries are largely selected on the basis of academic performance yet, in a highly cited British Medical Journal meta-study, it only accounts for 23% of the variance in performance in undergraduate medical training and 6% of that in postgraduate competency (Ferguson, James, and Madeley 2002).

Occupational Tribes

Research over the last four decades has shown personality traits — notably, those measured using the Big5 framework — have been found to be consistent, reliable and robust predictors of a variety of health, education and other significant life outcomes such as longevity and divorce (Strickhouser, Zell, and Krizan 2017; Noftle and Robins 2007; Roberts, Kuncel, Shiner, Caspi and Goldberg 2007).

The language people use has been shown to be a reliable and useful predictor of people's personality traits and, in particular, using machine learning and computational linguistics (Park et al. 2015; Kern et al. 2016). Research has demonstrated that the conversational language of social media users has been shown to be an accurate way of predicting a user's personality trait, at least as accurate as the judgement of a spouse, and more accurate than predictions of co-workers, friends and family (Youyou, Kosinski, and Stillwell 2015).

A recent study that uses data from over 100,000 active social media users shows that many people in the same occupations (Kern et al. 2019) share similar personality trait patterns. Additionally, many related occupations, such as technology roles, are adjacent or close to one another when measured in terms of the differences between the median personality traits of people in those in the occupations.

Here we show using a new set of more detailed occupation-personality traits data from a large number of Australians (n=99,897) in a wide variety of occupations (n=624) across all industries that, consistent with previous research, many people in the same role share similar combinations of personality traits, and that the resulting personality footprints of occupations cluster into eight different types of roles or tribes.

Understanding more about groups of occupations or tribes that share similar personality footprints could become an important future key to unlocking better job-fit — one that is not simply based upon an individual's skills and experience but on their preferred way of engaging with the world.

An important aspect of this study was the discovery of occupational ‘tribes.’

When personality is overlaid on a large set of occupations, patterns appear so that seemingly unrelated occupations are in fact related according to shared personality traits of people in those roles. Using machine learning to cluster occupations based on shared personality trait combinations reveals eight occupational tribes each containing different sets of roles:

1. *Leaders*: teaching, advisory, and agents
2. *Listeners*: justice, social work, journalism
3. *Umpires*: hospitality, beauty, personal services
4. *Rebels*: artists, writers, film and television
5. *Experts*: medicine, technology, science
6. *Observers*: design, music, entertainment
7. *Accomplishers*: services, trades, and operations
8. *Fighters*: construction, finance, professional sport

The Importance of Engagement

The word ‘engagement’ has enjoyed a steep rise in popularity since the early 1990s. At least in books published in English. And we get to hear it a lot, possibly too much, in corporate presentations. At a time when engagement is at risk of becoming cliché, our research illustrates that engagement still has more to tell us about work, about people, and most of all how it relates to personality and occupation. Our research reveals how individuals with high or low engagement in the same roles can be directly related to how well their personality and occupation are aligned.

Our measurement of engagement uses the five-dimensional framework designed to measure individuals’ happiness and wellbeing known as PERMA (Seligman 2012) where Engagement is one of five independent dimensions of overall well-being namely: Positive emotions, Engagement, Relationships, Meaning and Accomplishment.

In this framework, Engagement is defined as an experience in which someone fully uses their skills, strengths, and attention for a challenging task, which produces an experience called “flow” (Csikszentmihalyi 1990).

There are a number of steps in our research that we undertook to explore the relationship between *Engagement* and *Occupation-Personality* fit or how well suited to one's personality their current role is.

Firstly, we conducted linguistic analysis of tweets unobtrusively collected from Twitter users, matched to occupations in Australia, using the Australian and New Zealand Standard Classification of Occupations (ANZSCO). For each occupation title, we used web search and twitter search to find users who mentioned the job title in their biotext and were most thus likely to be currently working as ACCOUNTANTs, ACTORs, ACTUARYs etc. In total we collected data from the public domain postings of 128,279 twitter users across 624 of the most common occupations.

Then, we automatically inferred the personality traits of these users by the words they used on social media using established and proven techniques. Then by sampling the median personality trait scores of all the people in the same occupation, we created an Occupation-Personality map or *Occupation Matrix,* a table with 30 personality dimensions (Big5 Subdomains) for each of the 624 occupations.

Once we had personality characteristics for each of the occupations in our study, we clustered the occupations themselves using machine learning based on their personality signatures and found there to be eight types of occupations, each of which share similar personality traits.

Now with a map of the typical personality traits of each role, using a test set of users not involved in creating the map (n=18k) we measured how well each user's personality aligned with the typical personality of others in their current occupation. We ranked each of the 624 occupations in terms of their predicted ideal occupation-personality match and compared where their current job sat in this ranking. If they already had the job with most similar personality traits to theirs - this was considered a "bullseye" or they were likely already in their ideal or close to ideal job and conversely if their current job ranked 624 they were considered a misfit. Where their current job ranked in a list of ideal jobs provided a new consistent way of measuring occupation-personality fit across a wide range of users and a wide range of occupations.

Lastly, we explored if there were any correlations between occupation-personality fit and a number of other established measures of happiness and wellbeing.

One key wellbeing measure we found to be consistently linked with occupation-fit was Engagement. We found that when an individual's personality closely matches an occupation's personality, levels of engagement are measurably higher and conversely when someone's personality does not match the personality of their occupation there is a greater chance of disengagement.

A recent large-scale metastudy has shown that almost half of the overall variance in engagement can be explained by personality features (Young, Glerum, Wang and Joseph, 2018) yet most research to date has been focused on understanding and identifying which, if any, individual personality traits are most related to engagement.

Instead, here we explore how well the pattern of personality traits of an individual matches the pattern of personality traits of their occupation and how occupation-personality fit relates to engagement. In other words, are people in roles that suit their personality more engaged?

Another significant metastudy (Cole, Walter, Bedeian, O'Boyle 2012) that draws upon 37 other studies found there to be clear evidence that the specific concept of Employee Engagement as measured by the Utrecht Work Engagement Scale (UWES) is highly correlated to and in many cases effectively the polar opposite of

Job Burnout often measured by the Maslach Burnout Inventory (MBI). This evidence supports the view that the three dimensions of employee engagement in the UWES Scale (work engagement energy, involvement and efficacy) are functionally direct opposites of the dimensions of Burnout (exhaustion, cynicism and inefficiency).

In this research, we tested the associations between the engagement of individuals using the PERMA wellbeing model (Seligman 2012) and our own measure of occupation-personality fit. We also tested for correlation between occupation-personality fit for the other four dimensions in this wellbeing model (Positive emotions, Relationships, Meaning and Accomplishment) as well as happiness using the University of Vermont Hedonometer schema (Dodds, Peter Sheridan, et al 2011).

Using a large sample of people (n=18k) in 624 different occupations, we found that occupation-personality fit explained statistically significant variance in three of these dimensions: Happiness, Engagement and Accomplishment as well as the variations in the overall composite PERMA wellbeing score.

The most significant of these was *Engagement* where occupation-personality fit explained 25.7% of overall variance with a p value of less than 0.001.

We recognise that Engagement as measured here in the PERMA model, refers to a broader concept of engagement beyond simply engagement at work, but nonetheless, the fact that occupation-personality fit alone explains a large degree of the variance suggests people in well suited roles are more likely to experience ‘flow’ (the perfect combination of challenge and skill/strength) both during and outside their work lives.

Going further, we found that when a person is matched to a personality suited role (perhaps because they are engaged in their occupation) are typically happier (as measured by hedonic happiness) and more likely to pursue accomplishment for its own sake (higher PERMA Accomplishment scores). As a result of these higher engagement and accomplishment scores, people in personality-aligned roles also have higher overall wellbeing scores (higher overall PERMA scores).

At the enterprise level, and within the labour market more broadly, having people with personalities that match occupations has clear potential benefits in terms of overall workforce efficiency and effectiveness.

A clearer understanding of the relationship between personality and occupation is likely to lead to significant improvement at three levels: in the career choices of individuals; the recruitment practices of enterprises; and in the labour force policies and programs developed by governments.

The most obvious benefit is across-the-board productivity improvement, but there is even greater potential than economic productivity.

At an individual level, a higher level of engagement resulting from personality and occupational fit would also mean a person was happier, more satisfied and, generally, more successful in their occupation.

At the enterprise level, one would expect retention rates would be higher for more engaged employees, but it is also likely – with the flow concept in mind – that the quality of work performed by engaged people would also be greater. Such people would be more efficient and effective. Indeed, even passionate.

In terms of government policy and programs, the prospect of far fewer people stopping and starting different occupations, including changing courses during tertiary education, would see a considerable reduction in wasted resources as higher education completion rates would most likely rise.

If the relationship between personality and occupation were better understood in the education system, especially by career planners and other advisers, young people might enjoy a much more satisfying journey from school to work.

Finally, at a national level, with a more mature and widespread understanding of how important the relationship between personality and occupation is, we could see greater engagement create a form of national competitive advantage.

Building on previous insights

In this new work, we looked to explore the relationship between occupations and personality in more detail using four significant methodological advances beyond our previous work (Kern et al. 2019) in this domain namely:

1. While the initial study used a global sample of social media users to create our vocation map, this time we worked with a large corpus of users (100k+) located in a specific geographic market (Australia) and with any one of hundreds of diverse jobs (n=624) aligned to that geography using the Australian and New Zealand Standard Classification of Occupations (ANZSCO codes).
2. We examined Big5 facets or subdomains representing thirty trait features (30 dimensions) for each user as opposed to the Big5 domains (5 dimensions) used in the initial study giving a more subtle and accurate personality trait signature for each user.
3. We created a new occupation-personality map based on a subset of the total corpus of users using web-search to include only the top-ranked professionals in each role (n=15k), thus ensuring the sample of people used to create the *Occupation Matrix* was consistent and authoritative, relative to their peers.
4. Using an independent sample of users (n=18k) not used in the creation of the *Occupation Matrix* went on to create a new measure Occupation-Personality fit and we tested the correlation of Occupation-Personality fit with a range of established happiness and wellbeing measures.

METHOD

In this new research, we established at-scale relationships between occupations and combinations of Big5 personality traits. We explored the relationships between the detailed personality features across thirty dimensions of large cohorts of individuals in the same role in a universe of diverse industry occupations.

In addition to exploring the relationships between the personality patterns of individuals in the same role, we went on to explore relationships between occupation-personality fit and a variety of wellbeing measures.

Our detailed research method involved eight key steps:

1. Determine the reference universe of occupations
2. Curate a collection of social media profiles by occupation
3. Infer personality traits of people in the same occupation
4. Cluster occupations by personality features
5. Analyse the features of each cluster of occupations or tribe
6. Infer users' happiness and other wellbeing measures
7. Measure Occupation-Personality fit
8. Explore how occupation-personality fit relates to wellbeing

1. Determine a reference Universe of Occupations

To determine the reference universe of occupations, we started with the Australian Government's Job Outlook List with 1646 formal occupation titles, descriptions and corresponding to the Australian occupational standard – ANZSCO. Our approach, as with our previous study, looks to find people who are active on the public social media platform Twitter and who in their biographical profile state their occupation explicitly as one of the 1646 formal occupations mentioned above.

Through advanced domain-controlled web search queries, we collected the top 100 Twitter users suggested by Google and other web search engines and filtered by occupation titles.

The locations (Australia) and Twitter handles by different occupations could be roughly reliably ascertained from Google based on the query. After removing those jobs with an insufficient sample size of Australian users on Twitter, we settled on 624 occupations across a diverse range of industries and skill levels in the final list for further analysis. This list covers the most common jobs in Australia. The 624 occupations chosen represent a wide range of diverse occupations across 216 different 4-digit ANZSCO codes, which, when taken together, represent 73% of Australia's workforce.

2. Curate a collection of social media profiles by occupation

Since web search ranking, such as Google search, works via a combination of authority and relevance, we postulate that users with a higher search ranking within an occupation are likely to be more relevant and hence representative of people in that occupation. On the authority dimension we also postulate that higher search ranked users are also more likely to be associated with a greater degree of extrinsic success in that occupation. We tested this with tennis players and found higher search-ranked tennis players to be more likely to be among the top tennis professionals worldwide whereas low-search ranked tennis professionals are likely to be lower ranked professionals or unranked amateurs.

Based on this approach we took the top 25 web search-ranked users as the representative of employees for each job and the 16K social media users (n=15,600) and their accounts would make up the universe of employees for further analysis, which we call the Occupation Mapping data.

An additional independent sample of 18k users (n=18,000) who are also matched to the 624 reference occupations, but are outside the first 25 in terms of web search ranking for each occupation is kept for testing occupation-fit.

The Twitter application programming interface (API) is used to collect the basic twitter information and statuses of the underlying dataset of the 16K active Twitter users each with one of the 624 reference occupations. For each user, we collected the most recent 500 public posts and ignored any users who elect to keep their posts private.

3. *Infer personality traits of people in the same occupation*

Much previous research, including our own, has shown the efficacy and accuracy of personality inference using social media. In this study, we used a proven, accurate and established commercial online service (IBM Personality Insights) to infer the

Big5's domains (5) and facets (30 numeric features) of each user from social media, which, in our case, used Twitter.

The context of the recent 500 statuses of each employee is the input text of the service. Although IBM's personality tool supports several languages such as English, Japanese, Spanish and so on, we only use English language accounts. The minimum number of words of the input text is 100, which means we only have personality features if the user has sufficient activity on Twitter to accurately infer the personality traits of the user.

For each user, we used the IBM personality inference engine to return the Big5 personality domains scores (openness, conscientiousness, extraversion, agreeableness and emotional range) and their corresponding personality facets (30 dimensions in total – with six facets within each Big5 domain) to explore the relationship between personality and occupation.

The Big5 domains and the corresponding six facets of each domain are the primary and secondary dimensions (see Appendix A for full list). Domains are more likely to be the high-level overview of personality while the 30 facets describe how a person engages with the world in more detail. Therefore, 30 personality facets are more comprehensive and provide a fuller description of personality.

4. Cluster occupations by personality features

We used an unsupervised machine learning algorithm, PAM (Partition Around Medoids) also known as K-Medoids, to help to discover interesting patterns in data, grouping occupations based on the personality traits of those in those roles.

Firstly, for each of 624 different jobs we identified data on the top 25 search ranked employees in each of these roles from the sample. Each of these groups were regarded as the representatives of each occupation in the universe and we used the median personality scores for each facet for these cohorts of employees to measure the personality of that occupation.

From these steps we created a table of data called the *Occupation Matrix* — a map of the personality dimensions of each occupation comprising 624 rows and 30 columns, where rows represent the occupation list and columns represent the 30 personality facets.

Does the Data Lend Itself to Clustering?

To measure the Clustering Tendency of this data we used the Hopkins test to assess the non-random structure of the Occupation matrix. The Hopkins score for the occupations matrix data is 0.79 with scores that are closer to 1 representing more naturally predisposed to clustering. The result indicates that there are natural and meaningful clusters in the Occupation Matrix.

What are the optimum number of clusters?

To determine the optimum number of clusters to choose — where there are the best, most distinct yet cohesive naturally occurring groupings of occupation and personality combinations — we use intrinsic measurements of clustering evaluation that do not require ground truth labels, which includes Davies-Bouldin Index, Silhouette coefficients, Calinski-Harabas Index and Dunn Index.

These indices measure the difference between within-cluster distances and between-cluster distances. We created a composite index of all the indices in Figure 1 (below). Note the Davies-Bouldin Index has been inverted to match others.

The ideal number of clusters is a trade-off between maximizing the number of clusters while maintaining their distinctiveness as measured by the composite index. We can see from the figure below that eight clusters is the optimum number as after eight clusters, the composite index drops significantly.

\-\-\-\-\-\-\-\-\-\-\-\-\-\-\-\-

Insert Figure 1 about here

\-\-\-\-\-\-\-\-\-\-\-\-\-\-\-

5. Analyse the features of each cluster of occupations or tribe

We call these clusters the ‘Eight Tribes’ and the occupations within each cluster or tribe are defined as a set of occupations that share similar personality traits.

The clustering is based on how similar roles are to one another in personality terms.

We measured the personality similarity of roles using the Euclidean distances between the 30 facets of one role to another, so that the closer together occupations are the more similar the personality footprint is with those roles. This proximity of roles (a representation of personality similarity) can be seen clearly in a two-dimensional mapping of the dimensions using a T-SNE plot (**Figure 2**).

We looked at the clustering of occupations by tribes and the personality characteristics of occupations at the XY extremes (principal components) when mapped to two-dimensions and explored patterns. In other words, we looked at the occupations at the 'north pole' versus 'south pole' in our clustering. And then, we looked at the occupations at the 'western pole' versus 'eastern pole' in our clustering.

Patterns were consistent with variations in personality facets along each dimension. High openness is associated with occupations in the eastern pole, while high extraversion is associated with occupations at the western pole.

Looking at the north-south axis, higher emotional range is associated with occupations closer to the north pole, and high conscientiousness is associated with occupations on the south pole.

We then ranked all 624 occupations by their two principal components represented by their horizontal and vertical position in figure 2 and found there to be clear patterns in each quadrant as follows:

- North: Artistic and performance roles such as OPERA SINGER, COMEDIAN and BALLET DANCER

- South: Business roles such as BANK MANAGER, ACCOUNTANT, BUSINESS BROKER and small businesses roles such as LOCKSMITH, PLUMBERS and ELECTRICIANS
- East: Medical specialists and scientists such as INFECTIOUS DISEASES PHYSICIAN, VASCULAR SURGEON and BIOCHEMIST
- West: Athletes including TENNIS PLAYER, RACING DRIVER and FOOTBALLER.

This analysis further supports our view that no one single personality trait explains how engaged a person might be. Rather, it is the matching of a person's whole personality with the personality of the occupation defines occupation-personality fit and related benefits.

Insert Figure 2 about here

Sometimes, occupations from the same industry are in the same tribe. For example, athletes, cyclists, footballers, golfer and tennis players are in the same tribe (**Appendix A**) and in other cases, functional roles, such as executive leadership roles like CEO, CTO, COO, CFO and CIO, are also found in the same tribe.

For clarity, we labelled each tribe: Leaders, Listeners, Fighters and so on (**see Table 1**) by examining the signature personality characteristics of each tribe a preferred way of engaging with the world or 'modus operandi' for each tribe is noted based on the distinctive personality trait combinations associated with that group of occupations as well as the types of unique roles within each cluster.

Insert Table 1 about here

We went on to explore the hierarchical relationships between tribes in terms of the similarity of personality footprints as well as the relationship between personality facets mapping them in the dendrogram (see **Figure 3**).

By analysing the rows of this dendrogram we can see facets that have a natural correlation within aligning and clustering together. So, for example, the facet of Modesty within Agreeableness and facet of Immoderation within Emotional Range are related. Also, perhaps unsurprisingly the facets of Artistic interest and Imagination within Openness are also closely related.

By analysing the clustering of the columns of the dendrogram we note that *Experts* and *Listeners* were most similar to each other - this is consistent with the fact these two tribes are adjacent on the 2-Dimensional TSNE Plot (Figure 2) and while most medical specialist roles are in *Experts* tribe there are a number among *Listeners* too such as *Dermatologist, Diabetologist and Veterinarian.*

The tribe most different to all other tribes is *Accomplishers* with their notable low scores on a number of facets within the domain of Emotional Range. Perhaps services-oriented and operations roles occupied by people in this cluster such as *Accountant, Chauffeur* and *Electricians* benefit from the inferred patience, stability and conscientiousness of people in these roles.

- - - - - - - - - - - - - - - -

Insert Figure 3 about here

- - - - - - - - - - - - - - -

6. Infer users' happiness and other wellbeing measures

Happiness, along with personality traits, has been shown to be accurately inferred from social media. (Dodds, Peter Sheridan, et al 2011) developed a method to measure the happiness of social media users known as a 'hedonometer' based on their social media text posts. We have reproduced the methods described by Dodds et al and applied it to our corpus of users with known occupations and already inferred personality traits to better understand the relationships between happiness, occupation and personality. (Details on method are in **Appendix C**).

Overall wellbeing and happiness can also be understood as more than simply a series of positive emotions as illustrated by Seligman's more sophisticated model of wellbeing outlined in his book Flourish.

Seligman's multidimensional model involves positive emotions but the broader concepts of engagement, relationships, meaning and accomplishment. Human well-being can be measured along five separate dimensions: positive emotion (P); engagement (E); relationships (R); meaning (M); and accomplishment (A) and a method for inferring these scores from a linguistic analysis is explained in (Schwartz, et al 2016).

Those five dimensions can be subdivided into 10 dimensions (positive and negative for each dimension) inferred based on their social media text. The process of calculating PERMA scores is outlined below, using the dimension of positive engagement as an example, with the other nine dimensions computed using the same approach. Note that the 10 different dimensions of PERMA each use different term lists.

Firstly, we created a database from the reference corpus of terms in (Schwartz, et al 2016) with corresponding weights for each of the words in the positive engagement dimension. The range of weights for terms in 10 dimensions are from around -0.4 to 0.9.

Secondly, we parsed input text — social media postings from Twitter – by removing punctuation and breaking posts into words and phrases (tokenization), and then measuring the relative frequency distribution of those tokens for each user. The input text used for inference of PERMA scores is the same as that used for inference of personality traits and happiness scores.

Lastly, we used the sum of multiples between the normalised percentage and corresponding weights for each tokens as the final measurement of each of the PERMA Scores for each individual.

7. Measure Occupation-Personality fit

Here we created a measure of *Occupation-Personality* fit which is a relative ranking of how closely a user's inferred personality is with their current role relative to other possible roles in our universe.

To create an occupation-independent and consistent measure of the current Occupation-Personality fit of each user we:

A. *Determined which roles were most aligned to each user*

   For each user we determined which of the 624 roles was the closest in terms of personality fit with their inferred personality and then ranked all other roles in terms of their closeness of fit from second closest (ranked #2), third closest to farthest (rank #624).

B. *Noted where each user's current role appears within this ranking*

   With 1 being the best fit (bullseye) and 624 being the worst fit (misfit). This provides a simple way of comparing how well individuals' current roles are suited to their personality independent of occupation within our universe of 624 occupations.

The occupation-fit is measured using Euclidean distances between each user's personality (inferred from their social media text) and the personality of each of 624 different jobs (based on 30 Big5 personality facet scores).

8. Explore how occupation-personality fit relates to wellbeing

To explore the relationships where they exist between happiness and occupation-fit, we used a test user data set, which comprises 18,000 Australian Twitter users each of whom have one of the 624 reference occupations but none of whom are among those used in the creation of the Occupation Matrix — the map used to define the personalities of each occupation.

We applied linear regression to test the relationship between happiness of users and their current occupation-personality fit (see **Figure 4**) since the average happiness scores and corresponding occupation-fit ranks have each been measured.

To do this, we averaged the happiness scores of individuals in each occupation-fit rank for example the average happiness score of those users who are in their best ranked role for them, then those in who are in 2nd best ranked role for them and so on up to those who are currently in the worst possible matched role – ranked 624. Based on the data comprising 18k users, there is a significant correlation between happiness and occupation-fit (Full results see **Table 2**). The correlation coefficient is 0.463 with p-value of 0.013%.

\- - - - - - - - - - - - - - - -

Insert Figure 4 about
here

\- - - - - - - - - - - - - - -

Caution needs to be taken about over interpreting these results as the high variance in the data suggests the results may not be stable or representative.

However, we also note the variance in happiness in people who are "misfits" is highest indicating there is little relationship between happiness and occupation-fit when people are completely mismatched to roles by personality. Conversely, the correlation and effect size of occupation-fit and happiness among those who are in the Top 100 of people by occupation-personality fit is significantly higher.

Relationship Between Positive Engagement and Occupation-Fit

The same 18k user test set is used to measure the linear relationship between positive engagement and occupation-fit. We calculated the average and median value of positive engagement of each group of people whose current job shares the same occupation-fitness rank in the 624-job universe. It's clear to see that there are declines in mean and median of positive engagement with the larger occupation-fit. The correlation coefficient between median of engagement and occupation-fit rank is 0.5 with a 0.000% p-value (Full results see **Table 2**).

Insert Figure 5 about here

RESULTS

There are three interrelated aspects to the results of our work.

Firstly, we confirmed that occupations have personalities. Many people in the same role share similar combinations of personality traits and thus different occupations can be characterised as having their own ‘personality.’ In our work, we used The Big Five personality framework to explore 30 dimensions of personality comprising six facets across each of the five domains of openness, contentiousness, extraversion, agreeableness and emotional stability.

We inferred the ‘personality fingerprints’ of 624 occupations using public data created by a large number of Twitter users (n=99,897) in a broad range of different roles across a range of industries. To create a map of the personality of each occupation, we used a subset of this data (n=16k) comprising the top 25 users best matched to each role using search-ranking. For each role, we used the distribution of personality features (30 facets of Big5 Domains) to assign each occupation its own personality footprint.

Next, we ran a series of tests to reveal that the occupations naturally cluster into ‘tribes’ that share a similar personality profile. Our cluster analysis shows occupations fall into eight broad tribes, and our summaries identify the central or most typical occupation for each tribe (medoid) as well as the most defining personality attributes of each tribe (see Appendix B).

We found patterns among the tribes clustering of occupations to be consistent with expectations with some tribes more similar to each other (*Experts* and *Listeners*) while others being more distinctive (*Accomplishers*) — the most different to all others.

Lastly, using a large sub-sample of test users (n=18k) that were not used in the map making, we found there to be statistically significant correlations between occupation-personality fit and happiness as well as engagement, accomplishment and overall wellbeing.

The most significant correlation in terms of effect size and confidence is the relationship between Occupation-Personality fit and Engagement and a key finding is that 25.7% of the variance in people's Engagement can be explained by variations in occupation-personality fit.

Wellbeing scores are inferred based on established linguistic techniques in previously published research: PERMA (Schwartz, et al 2016); Hedonic Happiness (Dodds, Peter Sheridan, et al 2011). Occupation-personality "fitness-scores" represent the ranked distance between individual's current role and their predicted ideal role or "bullseye" based on their inferred personality traits and the known traits of the roles. Fitness scores are in the range from 1 to 624 binned by deciles with larger scores indicating a better fit. Spearman correlations are based on occupation-personality fit and inferred wellbeing scores of people (n=18k) in 624 occupations across all industries.

Insert Table 2 about here

DISCUSSION

The tribes, and the attributes that define them, offer great potential when considering the relationship between personality and engagement.

The questions ‘what personality traits are most important to engagement?’ may not be the best question to ask.

A more helpful question might be, ‘to what extent does the alignment of personality and occupation influence engagement?’

The combination of attributes that define each tribe is very different. So much so that it is safe to say no specific personality traits are more or less important to engagement, and no personality trait is universal.

For example, within the Big Five’s personality trait of openness, the tribe of Leaders scores highest of all tribes for the facet of intellectual curiosity, while Rebels score highest for the openness facet of imagination.

We do not see that there can be a general rule that intellectual curiosity is more important to engagement than imagination. But we can say that imagination is important to occupations such as photographer, novelist, and illustrator, just as intellectual curiosity is important to an economic historian, legal researcher, and commodities trader.

The combination of traits and facets is what defines the tribes and the occupations that fall within the tribe. So, an individual's entire personality is important to engagement in that, the better their personality matches the personality of the occupation, the greater the engagement is likely to be.

What do the tribes tell us about engagement?

Perhaps the first thing they tell us is that the alignment of our individual personality and the personality of an occupation is the key to engagement.

When there is alignment, we have people who love what they do. And we have the conditions for 'flow' to be more likely.

Furthermore, we are now able to predictively match individuals to occupations that they are most likely to have a high level of alignment with. The potential this offers individuals, enterprises and whole economies are significant.

For the individual, personality and occupational alignment mean they are more likely to find their work satisfying, be more successful at it, and be happier, provided other aspects of their lives and relationships are in good order.

For the enterprise or organisation, it means higher performance, greater productivity, and – provided all roles in the enterprise are aligned with personality – there will be a greater sense of esprit de corps across the organisation. It would be relatively simple to apply the alignment of personality and occupations to recruitment, retention and development strategies.

In terms of a nation's economy, it is hard to overestimate what could be possible. Imagine if the current 'wasted' resource created by people abandoning study or work because they determine that it is "not for them" could be dramatically reduced. Instead, imagine if individuals could apply a reliable personality lens to their career planning and related decisions. More people would be pursuing careers and be in occupations that suited them. If personality became a mainstream part of a nation's career planning and skill development effort it would be akin to all boats rising on a new tide.

CONCLUSION

The most important finding in our research is that matching the personality of occupations to the personalities of individuals offers enormous potential to improve engagement.

With 25% of the variance in engagement explained by occupation-personality fit – and a p-value of 0.000% this relationship is highly significant. And while Engagement varies with Occupation-personality fit, it also varies by occupation. Prior research (Ascenso et al 2018) and our own shows for example that concert musicians are highly engaged — more so than those in most other professions. What this research has demonstrated is that anyone in an occupation that is aligned with their personality can enjoy a level of career-related engagement boost equivalent to putting them in the Engagement league of concert musicians.

What is also important in our findings is that a more holistic view of the relationship between occupation and personality, especially in the context of the tribes we have identified, is likely to tell us more, and be more applied, than are studies of particular personality traits.

We have effectively built a platform that would allow individuals, organisations, and governments to develop much more effective career planning and management strategies.

Moreover, this work can readily be applied to study other aspects of careers. For example, it is now possible to identify the personality characteristics of the most successful technology entrepreneurs or start-up teams. Would-be entrepreneurs could be assessed in terms of the tribes they come from and how well their own personality matches this special cohort.

The interplay of data from occupations, personality testing, and social media has given us the opportunity to see patterns that have not been possible to see before and, though some may be faint or crude, they are nonetheless there and can teach us a lot about the relationship between work, engagement and happiness.

REFERENCES

Ascenso, S., Perkins, R. and Williamon, A., 2018. Resounding meaning: a PERMA wellbeing profile of classical musicians. *Frontiers in Psychology*, p.1895.

Csikszentmihalyi, M. and Csikzentmihaly, M., 1990. *Flow: The psychology of optimal experience* (Vol. 1990). New York: Harper & Row.

Cole, M.S., Walter, F., Bedeian, A.G. and O'Boyle, E.H., 2012. Job burnout and employee engagement: A meta-analytic examination of construct proliferation. *Journal of management*, *38*(5), pp.1550-1581.

Dawson, B.F. and Thompson, N.J., 2017. The effect of personality on occupational stress in veterinary surgeons. *Journal of Veterinary Medical Education*, *44*(1), pp.72-83.

Dodds, P.S., Harris, K.D., Kloumann, I.M., Bliss, C.A. and Danforth, C.M., 2011. Temporal patterns of happiness and information in a global social network: Hedonometrics and Twitter. *PloS one*, *6*(12), p.e26752.

Dole, C. and Schroeder, R.G., 2001. The impact of various factors on the personality, job satisfaction and turnover intentions of professional accountants. *Managerial auditing journal*.

Ferguson, E., James, D. and Madeley, L., 2002. Factors associated with success in medical school: systematic review of the literature. *Bmj*, *324*(7343), pp.952-957.

Kern, M.L., Park, G., Eichstaedt, J.C., Schwartz, H.A., Sap, M., Smith, L.K. and Ungar, L.H., 2016. Gaining insights from social media language: Methodologies and challenges. *Psychological methods*, *21*(4), p.507.

Kern, M.L., McCarthy, P.X., Chakrabarty, D. and Rizoiu, M.A., 2019. Social media-predicted personality traits and values can help match people to their ideal jobs. *Proceedings of the National Academy of Sciences*, *116*(52), pp.26459-26464.

Kokkinos, C.M., 2007. Job stressors, personality and burnout in primary school teachers. *British journal of educational psychology*, *77*(1), pp.229-243.

Noftle, E.E. and Robins, R.W., 2007. Personality predictors of academic outcomes: big five correlates of GPA and SAT scores. *Journal of personality and social psychology*, *93*(1), p.116.

Park, G., Schwartz, H.A., Eichstaedt, J.C., Kern, M.L., Kosinski, M., Stillwell, D.J., Ungar, L.H. and Seligman, M.E., 2015. Automatic personality assessment through social media language. *Journal of personality and social psychology*, *108*(6), p.934.

Roberts, B.W., Kuncel, N.R., Shiner, R., Caspi, A. and Goldberg, L.R., 2007. The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes. *Perspectives on Psychological science*, *2*(4), pp.313-345.

Seligman, M.E. and Flourish, P., 2012. A visionary new understanding of happiness and well-being. *M. Selligman, Flourish*, pp.1-368.

Strickhouser, J.E., Zell, E. and Krizan, Z., 2017. Does personality predict health and well-being? A metasynthesis. *Health Psychology*, *36*(8), p.797.

Young, H.R., Glerum, D.R., Wang, W. and Joseph, D.L., 2018. Who are the most engaged at work? A meta-analysis of personality and employee engagement. *Journal of Organizational Behavior*, *39*(10), pp.1330-1346.

Youyou, W., Kosinski, M. and Stillwell, D., 2015. Computer-based personality judgments are more accurate than those made by humans. *Proceedings of the National Academy of Sciences*, *112*(4), pp.1036-1040.

Table 1

Characteristics of each Occupation-personality cluster

The Eight "Tribes"of Occupations Clusters of occupations that share similar Big5 personality facets		
Tribe Label	Modus Operandi	Roles in
Leaders	We are born leaders.	Teaching, advice and as agents.
Listeners	We are good listeners.	Justice, social work and journalism.
Umpires	We care about the details.	Hospitality, beauty & personal service.
Rebels	We are game changers.	Making art, writing, film and TV.
Experts	We know what we are doing.	Medicine, technology and science.
Observers	We understand patterns, trends and motifs.	Design, music & entertainment.
Accomplishers	We are reliable. We get things done.	Services, trades and operations.
Fighters	We are tough, competitive and like working with our body.	Construction, finance and professional sport.

Table 2

Correlation between Occupation-Personality Fit and Wellbeing features

Occupation-Personality Fit and Well Being

Features	Correlation	p-value	R-squared
Hedonic Happiness (Happiness_index_norm)	**0.463**	**0.013% ** **	**21.431%**
PERMA Scores			
Positive Emotion (POS_P_SCORE_NORM)	0.125	32.751%	0.005%
Engagement (POS_E_SCORE_NORM)	**0.507**	**0.000% *****	**25.740%**
Relationships (POS_R_SCORE_NORM)	0.087	49.596%	0.763%
Meaning (POS_M_SCORE_NORM)	0.000	99.732%	0.000%
Accomplishment (POS_A_SCORE_NORM)	**0.337**	**0.694% ** **	**11.346%**
Overall Wellbeing (PERMA_NORM)	**0.301**	**1.666%***	**9.039%**

* P < 0.05

** P < 0.01

*** P < 0.001.

Figure 1

Clustering Quality Measurements Plot

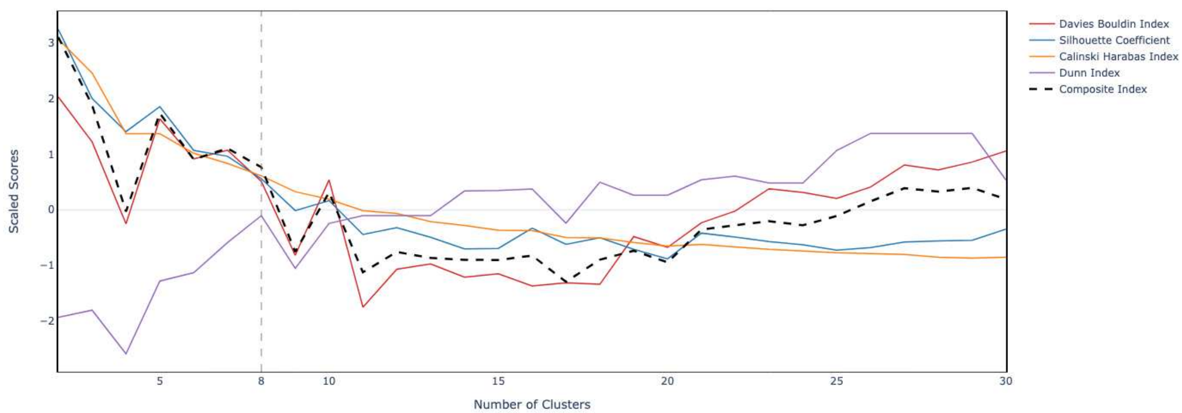


Figure 2

T-SNE 2-D plot of 8 Tribes (Interactive version here)

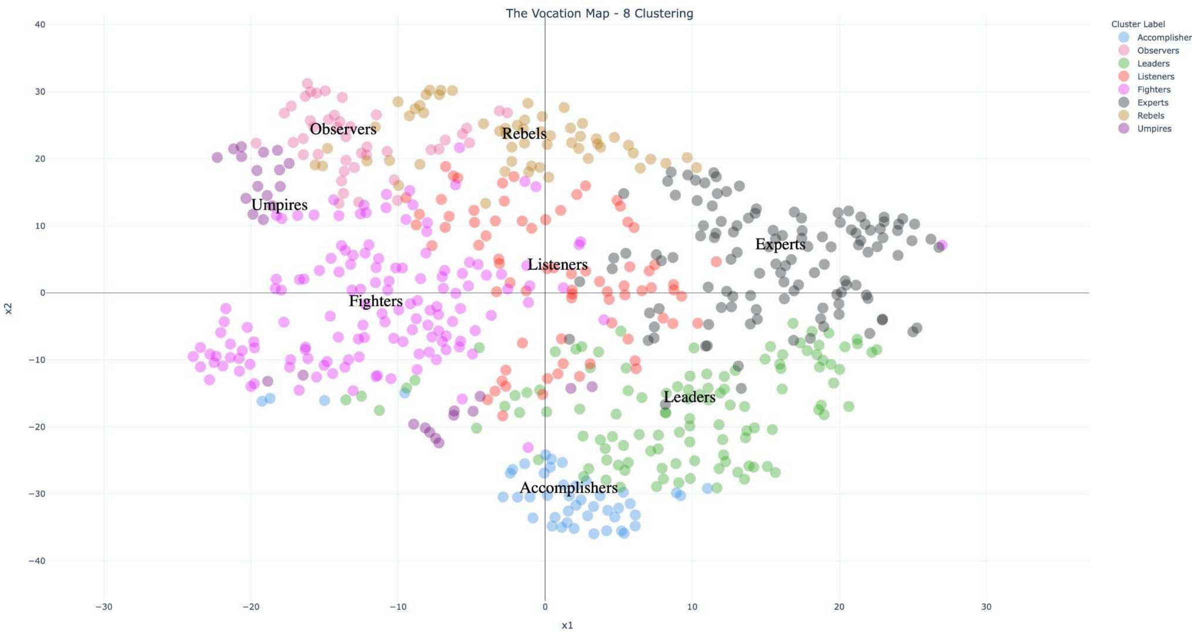


Figure 3

Eight occupational tribes each have their own distinctive patterns of personality traits as revealed by this tribes-facets heatmap and dendrogram that reveals hierarchical relationships between features and tribes.

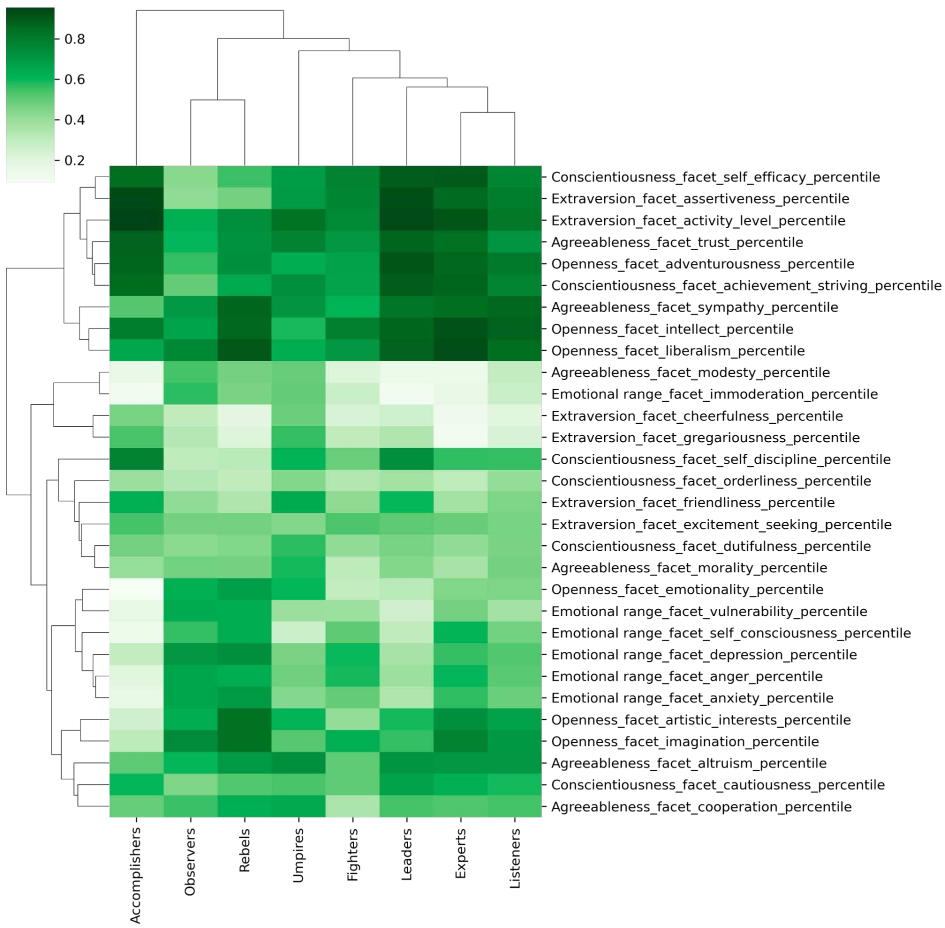


Figure 4

Happiness correlates to Occupation-Personality fit. The occupation-personality fit ranking is on the X-axis with best fit (1st ranked) on left and worst fit (624 ranked) on right. Each blue dot represents the average happiness of all users in each occupation-fit ranking from left to right and despite some noise, we can see clear statistically significant correlation between occupation-personality fit and happiness.

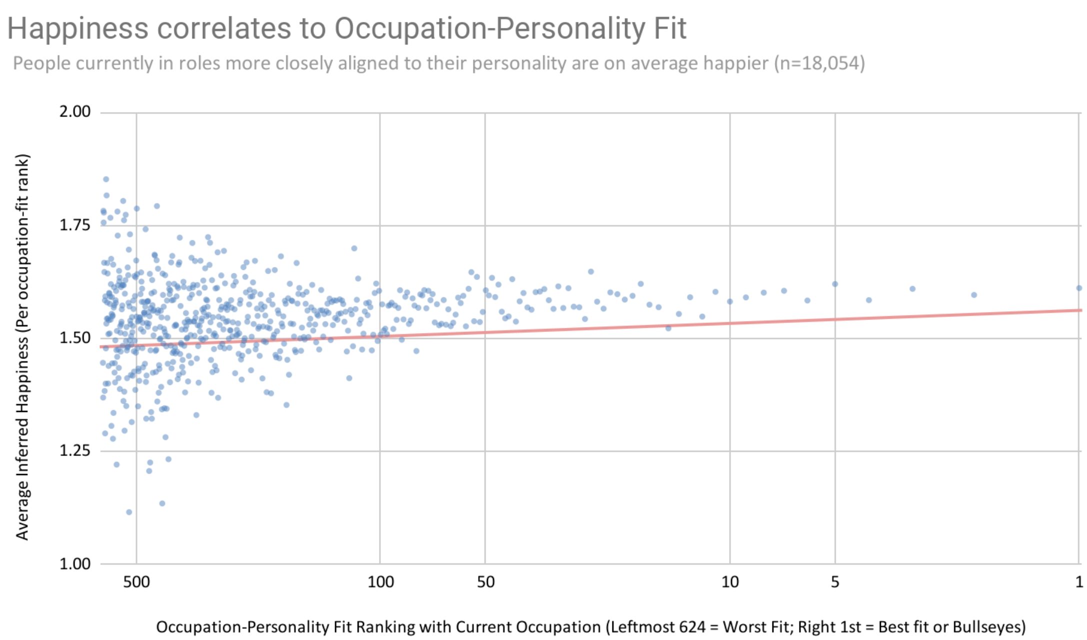
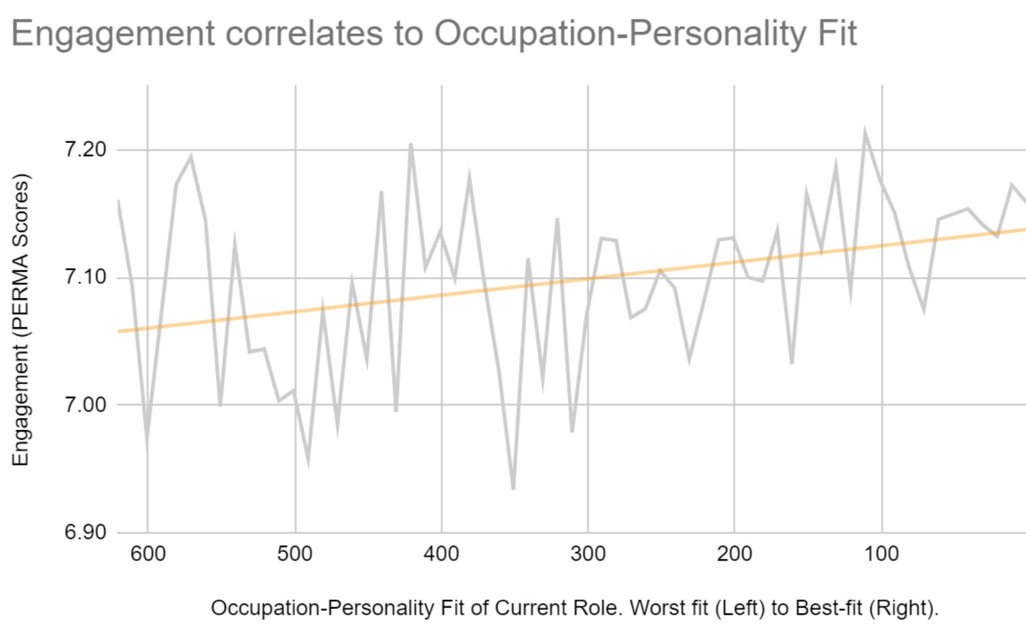


Figure 5

Engagement correlates to Occupation-Personality fit.

The occupation-personality fit ranking is on the X-axis with best fit (1st ranked) on right and worst fit (624 ranked) on left. With 25.7% of Engagement explained by Occupation-Personality fit, it's clear that it's an important association.



Appendix A - Occupation Tribes Clustering Full List

leaders	listeners	umpires	rebels	experts	observers	accomplishers	fighters
actuary	acupuncturist	aromatherapist	animator	aerospace engineer	actor	accountant	aerobics instructor
agricultural engineer	aged carer	beauty therapist	antique dealer	anaesthetist	band leader	advertising specialist	aeronautical engineer
agricultural scientist	animal trainer	blogger	art historian	analytical chemist	cake decorator	asset manager	agronomist
ambassador	aquarist	cabin crew	art therapist	anatomist	candle maker	bank manager	aircraft maintenance engineer
ambulance paramedic	arborist	car detailer	artistic director	anthropologist	child care worker	barista	ambulance officer
apiarist	art director	caterer	author	archaeologist	cinematographer	basketball coach	analyst programmer
assistant commissioner	barber	chef	ballet dancer	architect	comedian	bookkeeper	applications developer
astrologer	beekeeper	chef de partie	ballet teacher	archivist	costume designer	building inspector	athlete
building manager	caretaker	choreographer	book editor	art teacher	costume maker	business broker	auctioneer
business consultant	chemist	civil celebrant	cartoonist	associate editor	dental nurse	business development manager	bank worker
cabinetmaker	chinese medicine practitioner	commis chef	casting director	astronomer	director of photography	chauffeur	bicycle mechanic
cafe manager	chiropractor	dance teacher	ceramic artist	audiologist	dog breeder	chief information officer	bodyguard
careers adviser	chocolate maker	dancer	chorister	bacteriologist	dog trainer	chief operating officer	bookmaker
chief executive	climatologist	dental assistant	circus artist	baker	drainer	coffee roaster	bouncer
chief financial officer	columnist	fashion designer	community artist	barrister	drummer	contract administrator	bricklayer
chief technology officer	communications specialist	fitness instructor	composer	biochemist	entertainer	conveyancer	building surveyor
childbirth educator	community worker	florist	critic	bioengineer	fashion photographer	courier	bus driver
clinical nurse educator	cook	lifeguard	drama teacher	biologist	graphic artist	customer service manager	business analyst
commodities trader	copywriter	make up artist	dressmaker	biomedical engineer	graphic designer	diving coach	business systems analyst
communications manager	correctional officer	merchandiser	embalmer	botanist	guitarist	electrical contractor	camera operator
community development officer	creative director	model	embroiderer	cardiologist	hairdresser	electrician	carpenter
community pharmacist	criminologist	nanny	feature writer	cartographer	hotel receptionist	finance broker	cashier
company secretary	dental surgeon	office manager	features editor	chemical engineer	instrumentalist	fitter	civil engineer
demographer	dermatologist	orthodontist	film director	clinical psychologist	jeweller	forklift operator	communications engineer
dentist	detective	personal assistant	film producer	combat engineer	jewellery designer	furniture removalist	concreter
diabetes educator	diabetologist	personal shopper	illustrator	commando	milliner	hotelier	construction project manager
dietitian	dive master	receptionist	interior decorator	conservator	mortician	immigration consultant	construction worker
director of nursing	diver	sous chef	interviewer	councillor	multimedia designer	insurance broker	corrections officer
early childhood teacher	diving instructor	waiter	landscape photographer	counselling psychologist	music director	locksmith	cost accountant
economic historian	editorial assistant		literary agent	cultural historian	musician	marketing consultant	crane operator
economist	environmental engineer		manicurist	ecologist	nail technician	marketing specialist	cricket coach
education officer	executive producer		museum curator	educational psychologist	nursing assistant	migration agent	cricketer
employment consultant	first aid instructor		music teacher	emergency medicine specialist	pet groomer	mortgage broker	customs officer
environmental economist	first aid trainer		music therapist	emergency physician	phlebotomist	motor mechanic	cyclist
environmental educator	forester		musicologist	endocrinologist	portrait photographer	network support	data entry operator
event manager	funeral director		novelist	entomologist	production assistant	plant manager	delivery driver
executive director	furniture designer		opera singer	environmental scientist	stunt performer	plumber	dental technician
facilities manager	gardener		painter	epidemiologist	undertaker	portfolio manager	derivatives trader
finance director	goldsmith		palaeontologist	exercise physiologist	veterinary assistant	professional builder	disc jockey
food scientist	grape grower		photographer	eye surgeon	veterinary nurse	property manager	driving instructor
head teacher	herbalist		photographic artist	flight nurse	visual merchandiser	purchasing officer	duty manager
headmaster	hospitality worker		pianist	gastroenterologist	vlogger	quantity surveyor	electrical engineer
health educator	hypnotherapist		picture framer	general practitioner	vocalist	recruitment consultant	engineering manager
health promotion officer	industrial designer		playwright	genetic counsellor		retail store manager	enrolled nurse
hydrogeologist	interior designer		poet	geographer		travel agent	fencer
importer	journalist		potter	geologist		valuer	finance manager
industrial engineer	judge		production designer	geophysicist			financial analyst
industrial psychologist	kinesiologist		screenwriter	geriatrician			fire fighter
industry analyst	law clerk		script writer	health psychologist			food technologist
information officer	library technician		sculptor	heritage consultant			football coach
insolvency practitioner	livestock farmer		sound recordist	historian			footballer
instructional designer	logistics officer		taxidermist	hospital pharmacist			forensic pathologist
insurance agent	magazine editor		television producer	immunologist			futures trader
insurance consultant	managing editor		textile artist	infectious diseases physician			gatekeeper
internal auditor	media producer		violinist	intensivist			geotechnical engineer
interpreter	newspaper reporter		zookeeper	landscape architect			golfer
joiner	orchardist		zoologist	librarian			greenkeeper
kindergarten teacher	osteopath			life scientist			gym instructor
knowledge manager	park ranger			linguist			gymnastics coach
language tutor	pig farmer			marine biologist			handicapper
legal officer	police officer			materials engineer			harbour master
legal researcher	prison officer			materials scientist			health care assistant
liaison officer	product designer			mathematician			horse breeder
lobbyist	production manager			medical oncologist			horse trainer
management consultant	proof reader			medical physicist			hotel manager
managing director	public speaker			medical practitioner			ict manager
manufacturer	reference librarian			metallurgist			immigration officer
media relations manager	reflexologist			microbiologist			industrial chemist
medical officer	scuba instructor			molecular biologist			jockey
member of parliament	sound technician			molecular geneticist			labourer
midwife	special needs teacher			multimedia artist			landscape gardener
miner	television journalist			neonatologist			logistics manager
motivational speaker	theatre manager			nephrologist			management accountant
multimedia specialist	veterinarian			neurologist			marketing coordinator
naturopath	welfare worker			neuropsychologist			martial arts instructor
network manager				neuroscientist			maths tutor
newsagent				neurosurgeon			mechanic
nurse manager				nurse educator			mechanical engineer
nurse practitioner				nurse researcher			mechatronics engineer
nurse unit manager				obstetrician and gynaecologist			media buyer
nutritionist				oceanographer			medical assistant
occupational therapist				ophthalmologist			medical intern
operations manager				optometrist			medical laboratory scientist
organisational psychologist				ornithologist			medical receptionist
patent attorney				orthopaedic surgeon			medical technician
personal carer				orthotist			metal fabricator
policy adviser				paediatric surgeon			meteorologist
policy analyst				paediatrician			mining engineer
postie				parasitologist			musculoskeletal physiotherapist
practice manager				pathologist			naval architect
practice nurse				pharmacologist			news editor
procurement manager				physical therapist			news photographer
procurement officer				physicist			parole officer
program administrator				physiologist			petroleum engineer
program director				physiotherapist			pharmacy technician
public affairs manager				plant pathologist			poultry farmer
quality assurance manager				plant physiologist			pr manager
registered nurse				podiatrist			private investigator
restaurateur				political scientist			probation officer
sales and marketing manager				primary school teacher			product manager
school administrator				project coordinator			production engineer
school principal				psychotherapist			programmer analyst
seafarer				public health physician			project administrator
secretary				rheumatologist			psychiatric nurse
senior counsel				secondary school teacher			public relations officer
service manager				seismologist			racing driver
solicitor				social worker			radio presenter
speech therapist				sociologist			referee
sport psychologist				soil scientist			rigger
sports physician				sonographer			sales assistant
tafe teacher				specialist physician			salesperson
town planner				speech pathologist			school bus driver
translator				sports physiotherapist			security consultant
union organiser				statistician			security guard
				surgeon			signwriter
				toxicologist			singer
				university lecturer			soccer player
				university tutor			software developer
				vascular surgeon			sports photographer
							stockbroker
							storeperson
							surfer
							surveyor
							systems administrator
							systems analyst
							tanker driver
							technical director
							technical writer
							telecommunications engineer
							tennis coach
							tennis player
							toolmaker
							transport engineer
							truck driver
							upholsterer
							veterinary surgeon
							weather forecaster
							web designer
							web developer
							webmaster
							welder

Link to full occupations tribes table.

Appendix B - Occupation Tribes, Medoid Roles, Personality Features

Features of each of the eight occupation tribes

Tribe	Distinctive personality attributes.	Medoid Role
Leaders	Adventurous, persistent, dispassionate, assertive, self-controlled, calm under pressure, philosophical, excitement-seeking & confident.	Mining Manager (Miner)
Listeners	Empathetic and compassionate - can understand others feelings, needs and suffering.	Journalist
Umpires	Bold, altruistic, respectful of authority, outgoing & uncompromising. Pays attention to the fine details.	Personal Assistant
Rebels	Authority-challenging, imaginative and appreciative of art.	Film Producer
Experts	Curious, reserved, independent, trusting of others, self-assured, organised, carefree, deliberate, driven & energetic.	Scientist (Plant physiologist)
Observers	Compassionate, modest, consistent, cheerful, sociable, emotionally aware, laid-back, content.	Comedian (Vlogger)
Accomplishers	Organized & outgoing. Confident, Down-to-earth, Content, Accommodating, Mild-tempered & Self-assured.	Advertising Executive (Avertising Specialist)
Fighters	Spontaneous and impulsive. Tough, skeptical, and uncompromising.	Labourer

Appendix C - Detailed Calculations

Wellbeing Inference Formula

In formulaic terms, each of the PERMA wellbeing dimensions such as engagement are calculated as follows:

$$PERMA_{Dimensions} = \sum_{word \in lex} w_{lex}(word) \frac{freq(word, doc)}{freq(*, doc)}$$

- $w_{lex}(word)$: lexicon (lex) weight for the word
- $freq(word, doc)$: frequency of the word in the document of a given user (twitter statuses of each user in our case)

$freq(*, doc)$: the total word count for that document

Hedonic Happiness Inference

The Hedonic Happiness scores of each user were calculated after Dodds et al as follows.

The method starts with an initial word list that is created from four separate text sources (Twitter, Google Books, music lyrics and the New York Times). The word list comprises the top 5,000 most frequently used words in each of these sources. To evaluate the happiness score of individual words, the authors asked Mechanical Turk users to rate a given word from 1 to 9 as sad to happy. They created average ratings for each word based on multiple users' perspectives. Each word in the corpus had over 50 independent evaluations.

The hedonometer corpus is improved by removing stop words — or words that are ‘neutrally’ rated in terms of happiness by the users, those ranging from 4 to 6 in average happiness scores. The final word list can then be used to provide a relative measure of happiness of twitter users based on the language they use in their tweets.

It's reasonable to compare the happiness and personality traits in our further analysis as both are based on inference from text from social media.

The process of measuring the inferred happiness of each user using the hedonometer is as follows:

Step 1: The paper provides a word list and the corresponding average happiness scores for those words. The range of average happiness scores in the word list is from around 1 to 9, where 5 means neutral for happiness. Therefore, we remove words with the neutral scores (5) from the list and use the revised list as the final corpus when we measure the happiness of an individual employee.

Step 2: To prepare the input of the happiness tool, we remove punctuations and split the sentence into separate words known as tokenize. Then we measure the relative frequency distribution of the tokens in the input and use the sum of multiples between the percentage of frequency and the corresponding average happiness scores as the final happiness score for the individual.

In formulaic terms:

$$h_{avg}(T) = \frac{\sum_{i=1}^{N} h_{avg}(w_i) f_i}{\sum_{i=1}^{N} f_i} = \sum_{i=1}^{N} h_{avg}(w_i) p_i$$

T : given text, twitter status in our case (the same input as IBM Personality Tool)

w_i : the ith word w_i in the given text T

$h_{avg}(w_i)$: estimate of average happiness of w_i

f_i : the frequency of the ith word w_i for which we have an estimate of average happiness

$p_i = \frac{f_i}{\sum_{i=1}^{N} f_i}$: corresponding normalised frequency

We treat all social media posts from Twitter users as the total corpus. To maintain a fixed, ordered list of words, we took the most frequent 50,000 words from the corpus. Using this list, we transformed the given text T into vectors of word frequencies.

Clustering Analysis Formula

Below are the detailed formulae for the clustering analysis:

Davies-Bouldin Index:

$$DB = \frac{1}{k}\sum_{i=1}^{k} maxR_{ij} \quad (i \neq j)$$

k : number of clusters

R_{ij} : similarity between cluster i (C_i) and j (C_j) $\Rightarrow$ $R_{ij} = \frac{s_i + s_j}{d_{ij}}$

s_i : the average distance between each point of cluster i (C_i) and the centroid of that cluster – also known as cluster diameter.

d_{ij} : the distance between centroids of cluster i (C_i) and j (C_j)

Lower scores indicate that the clusters are not similar with each other.

Silhouette coefficients:

$$s = \frac{b - a}{max(a, b)}$$

a : the mean distance between a sample and all other points in the same cluster

b : the mean distance between a sample and all other points in the next nearest cluster

Silhouette coefficient of the model is the average silhouette coefficients of all data points

Higher scores indicate that maximise distance among clusters and minimise distance within the cluster.

Calinski-Harabas Index:

$$ch = \frac{tr(B_k)}{tr(W_k)} \times \frac{n_E - k}{k - 1}$$

n_E : the size of Dataset E

k : number of clusters

B_k : groups dispersion matrix $\Rightarrow$ $B_k = \sum_{q=1}^{k} n_q(c_q - c_E)(c_q - c_E)^T$

c_q : the centre of cluster q

n_q : number of points in cluster q

c_E : the centre of Dataset E

W_k : the within-cluster dispersion matrix $\Rightarrow$ $W_k = \sum_{q=1}^{k} \sum_{x \in C_q} (c_q - c_E)(c_q - c_E)^T$

C_q : the set of points in cluster q

$tr(B_k)/tr(W_k)$: the trace of the Matrix

Higher scores indicate that clusters are dense and well separate

Dunn Index:

$$DI_m = \frac{min_{1 \le i < j \le m} \delta(C_i, C_j)}{max_{1 \le k \le m} \Delta k}$$

m : number of clusters

$\delta(C_i, C_j)$: the intercluster distance between clusters C_i and C_j

Δk : diameter of the kth cluster

Higher scores indicate the maximise distance between clusters and minimise distance within cluster