# Configuration Manual

MSc Research Project
Msc. in Data Analytics

Vishal Chikne
Student ID: x21204829

School of Computing
National College of Ireland

Supervisor: Dr. Vikas Tomer

**National College of Ireland**
**Project Submission Sheet**
**School of Computing**

| | |
|---|---|
| **Student Name:** | Vishal Chikne |
| **Student ID:** | x21204829 |
| **Programme:** | Msc. in Data Analytics |
| **Year:** | 2023 |
| **Module:** | MSc Research Project |
| **Supervisor:** | Dr. Vikas Tomer |
| **Submission Due Date:** | 14/08/2023 |
| **Project Title:** | Configuration Manual |
| **Word Count:** | XXX |
| **Page Count:** | 6 |

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

**ALL** internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

| | |
|---|---|
| **Signature:** | Vishal Chikne |
| **Date:** | 18th September 2023 |

## PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

| | |
|---|---|
| Attach a completed copy of this sheet to each project (including multiple copies). | □ |
| **Attach a Moodle submission receipt of the online project submission**, to each project (including multiple copies). | □ |
| **You must ensure that you retain a HARD COPY of the project**, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer. | □ |

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

| **Office Use Only** | |
|---|---|
| Signature: | |
| Date: | |
| Penalty Applied (if applicable): | |

# Configuration Manual

Vishal Chikne
x21204829

## 1 Introduction

This is a comprehensive configuration manual for Employees Engagement in Hybrid World Within Organisation Using Recommendation System will guide you with the requirements like software and hardware to run this system with complete process from installing libraries to execution of last step through process of setting up and configuration for the Recommendation System including the integration of Tableau for visualization and dashboard creation.

## 2 Project file details

This section provides a information of file to better understating structure of files, usage and to run modeling evaluations.

Dataset(HyWorx_IDEX_Survey_Results_Staff_Dashboard) : This is the dataset with multiple sheets the preprocessing steps applied to categories data into multiple sheets. The sheet using for recommdation system is name with Users and rest of visualise in the tableau for organisation to identify the metrics.

Recommeded_users.xlsx : This file has results of suggestions given by our research. Tableau file : This contains the dashboard and worksheets for organisation.

Google colab file(RP final): This contents the code to run in colab by importing dataset at the end will get output as recommeded_users into xlsx excel file.

## 3 System Specifications

This section will provides a idea of hardware and software minimum requirements to run this system.

### 3.1 Software Requirements:

- Operating System: Windows, Linux, mac
- Python(version is 3.11 or higher)
- Required Python Libraries: Pandas, Numpy, seaborn
- Jupyter Notebook (Version 7.0.0) or Google Colab
- Tableau Desktop or Tableau Public (for visualization)

### 3.1.1 Installation steps

**Install Python:**

Download Python from pythons website(https://www.python.org/downloads/). During installation make sure option to add Python to system PATH is selected. This allows easy command line access and follow the steps. [1]

**Install Jupyter notebook**

open the command prompt or terminal enter following command to install the jupyter notebook with the help of Python package by typing:

**pip install jupyter**

open Jupyter Notebook: whenever the Jupyter Notebook install you can open by the typing following command in command prompt or terminal and pressing Enter:

**jupyter notebook**

These will open the your website browser with notebook interface showing the contents of the current directory.

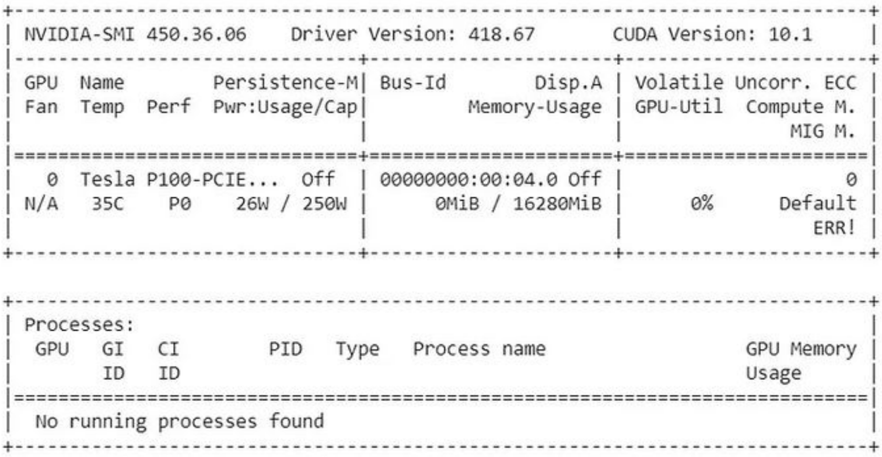
**Install Tableau:**

Navigate to Tableau website (https://www.tableau.com/products/desktop) to download Tableau Desktop into the system. Follow provided installation instructions as per operating system.

[2]

## 3.2 Hardware Requirements:

The Figure 1 showing the specification of google colab that needed to run the code.



```
+-----------------------------------------------------------------------------+
| NVIDIA-SMI 450.36.06    Driver Version: 418.67       CUDA Version: 10.1     |
|-------------------------------+----------------------+----------------------+
| GPU  Name        Persistence-M| Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp  Perf  Pwr:Usage/Cap|         Memory-Usage | GPU-Util  Compute M. |
|                               |                      |               MIG M. |
|===============================+======================+======================|
|   0  Tesla P100-PCIE...  Off  | 00000000:00:04.0 Off |                    0 |
| N/A   35C    P0    26W / 250W |      0MiB / 16280MiB |      0%      Default |
|                               |                      |                 ERR! |
+-------------------------------+----------------------+----------------------+

+-----------------------------------------------------------------------------+
| Processes:                                                                  |
|  GPU   GI   CI        PID   Type   Process name                  GPU Memory |
|        ID   ID                                                   Usage      |
|=============================================================================|
|  No running processes found                                                 |
+-----------------------------------------------------------------------------+
```

Figure 1: specifications of google colab

# 4 Data Collection

The dataset is taken by open source dataset library from survey monkey. the dataset has data which collected by employees surveys.

[1] helpful urls : `https://www.python.org/downloads/`

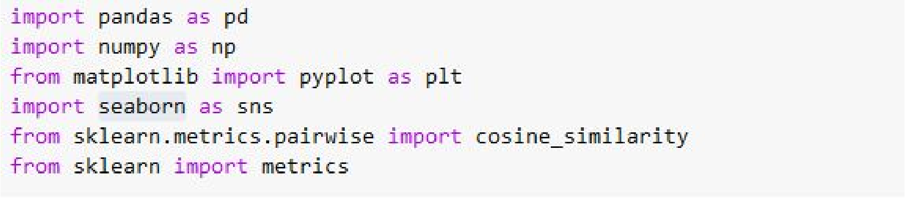
[2] https://www.tableau.com/products/desktop

# 5 Project Development

After successfully installing the software open the dataset and understand the information then open the google colab or jupyter notebook and follow the steps.

## 5.1 Importing libraries

imported all require library that use in this project.



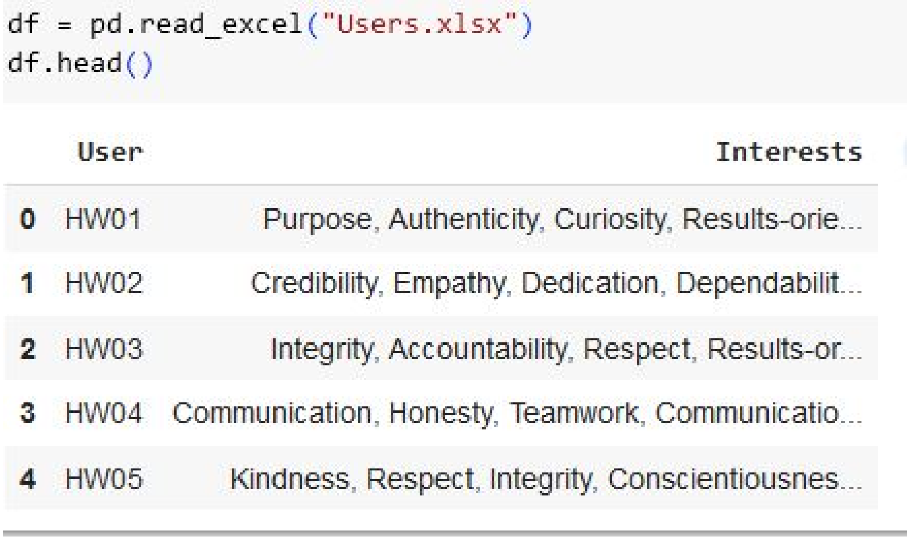
```
import pandas as pd
import numpy as np
from matplotlib import pyplot as plt
import seaborn as sns
from sklearn.metrics.pairwise import cosine_similarity
from sklearn import metrics
```

Figure 2: Required Python Libraries

## 5.2 Data cleaning and prepossessing

Data prepossessing and cleaned perform on the raw dataset to make in proper format that needed for the algorithm and unnecessary data, null values are removed. lastly organise data to structure format ensuring that each employee has unique identifier and corresponding interest data.

**Dataset importing** here user are name with HW01, HW02 and so on as shown



```
df = pd.read_excel("Users.xlsx")
df.head()
```

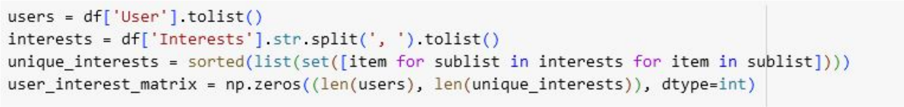
| | User | Interests |
|---|---|---|
| 0 | HW01 | Purpose, Authenticity, Curiosity, Results-orie... |
| 1 | HW02 | Credibility, Empathy, Dedication, Dependabilit... |
| 2 | HW03 | Integrity, Accountability, Respect, Results-or... |
| 3 | HW04 | Communication, Honesty, Teamwork, Communicatio... |
| 4 | HW05 | Kindness, Respect, Integrity, Conscientiousnes... |

Figure 3: Overview of dataset

Figure 3 because raw dataset doesn't have the name and interests are mentioned in front them.

## 5.3 Similarity Calculation and Correlation Matrix

Use Pandas to create correlation matrix in reflecting relationships between the different interest categories. user interest matrix build to find co relations.

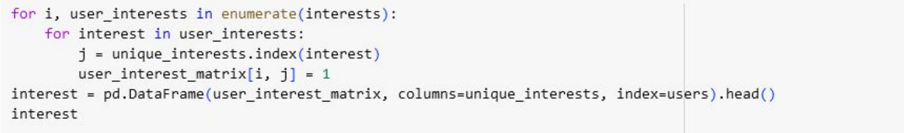


```
users = df['User'].tolist()
interests = df['Interests'].str.split(', ').tolist()
unique_interests = sorted(list(set([item for sublist in interests for item in sublist])))
user_interest_matrix = np.zeros((len(users), len(unique_interests)), dtype=int)
```

Figure 4: interest matrix

With the help this easy to understand the results of co relation between the employees and the organisation can better metrics for user engagements this calculated by this logic as shown in Figure 5.

The similarity score of the employee interest matrix is help in identify the trends pattern within the data.

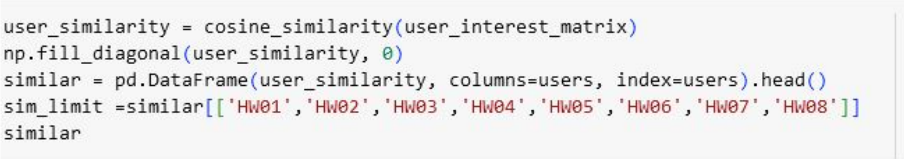


```
for i, user_interests in enumerate(interests):
    for interest in user_interests:
        j = unique_interests.index(interest)
        user_interest_matrix[i, j] = 1
interest = pd.DataFrame(user_interest_matrix, columns=unique_interests, index=users).head()
interest
```

Figure 5: Building employee interest matrix

The cosine similarity as the Figure 6 applied to calculate the similarity score with few users check how score is effective with interests.

Cosine similarity (Lahitani et al.; 2018) is a very versatile and valuable for metric to make calculations of interests between the vectors in multiple dimensional spaces. Its a applications span across the diverse domains from the NLP to recommendation systems to offering insights into the relationships that can not apparent through the other measure. Understanding and to utilizing of a cosine similarity to empowers the data analysts and machine learning users to unravel the patterns and optimise the solutions in the complex data of scenarios. [3]



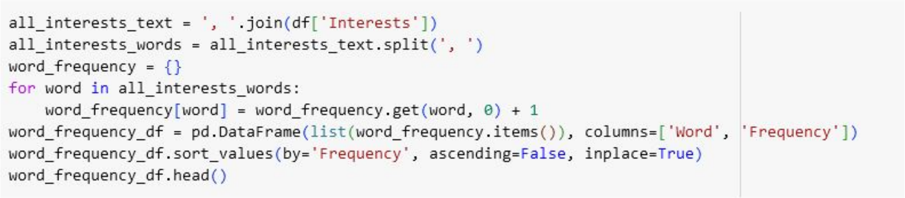
```
user_similarity = cosine_similarity(user_interest_matrix)
np.fill_diagonal(user_similarity, 0)
similar = pd.DataFrame(user_similarity, columns=users, index=users).head()
sim_limit =similar[['HW01','HW02','HW03','HW04','HW05','HW06','HW07','HW08']]
similar
```

Figure 6: cosine similarity

Further the frequency here Figure 7 by overall employees is calculated to identify how many users like same interest so organisation can take decisions and make good engagement within the organisation.

Interest frequency calculation is the very pivotal step in the understanding of preferences and in trends of individuals to a dataset. Whether in the of contex of market research, social networks or personalised recommendations to analszing interest frequencies provides a valuable insights that into user behavior and preferences. This is process involves quantifying the how often can specific interests can occur within a given dataset allowing to uncover patterns, popular trends and areas of interest concentration of egagements.

[3] https://en.wikipedia.org/wiki/Cosine$_{s}imilarity$

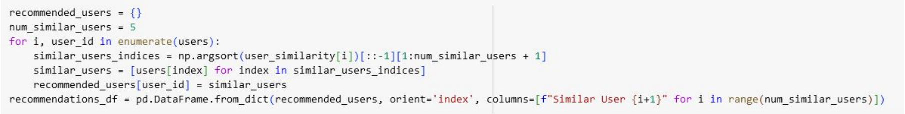


```
all_interests_text = ', '.join(df['Interests'])
all_interests_words = all_interests_text.split(', ')
word_frequency = {}
for word in all_interests_words:
    word_frequency[word] = word_frequency.get(word, 0) + 1
word_frequency_df = pd.DataFrame(list(word_frequency.items()), columns=['Word', 'Frequency'])
word_frequency_df.sort_values(by='Frequency', ascending=False, inplace=True)
word_frequency_df.head()
```

Figure 7: frequency calculation

## 5.4 Recommendation Engine

Create a recommendation algorithm using Python that utilizes similarity scores and the correlation matrix to suggest connections.



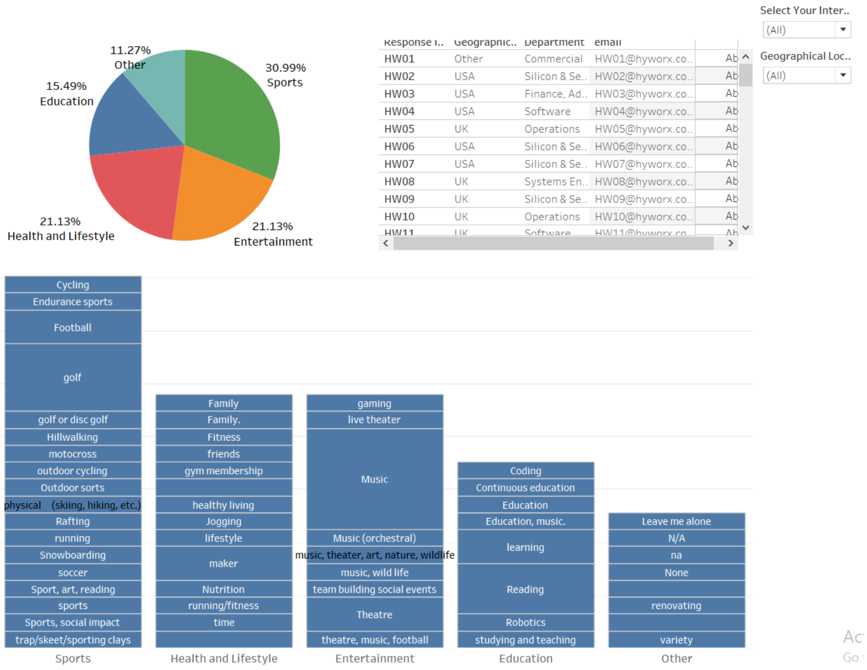
```
recommended_users = {}
num_similar_users = 5
for i, user_id in enumerate(users):
    similar_users_indices = np.argsort(user_similarity[i])[::-1][1:num_similar_users + 1]
    similar_users = [users[index] for index in similar_users_indices]
    recommended_users[user_id] = similar_users
recommendations_df = pd.DataFrame.from_dict(recommended_users, orient='index', columns=[f"Similar User {i+1}" for i in range(num_similar_users)])
```

Figure 8: Recommendations

The scores is calculated and basis of that users recommeded as Figure 8 to other to make a connection berween organisation.

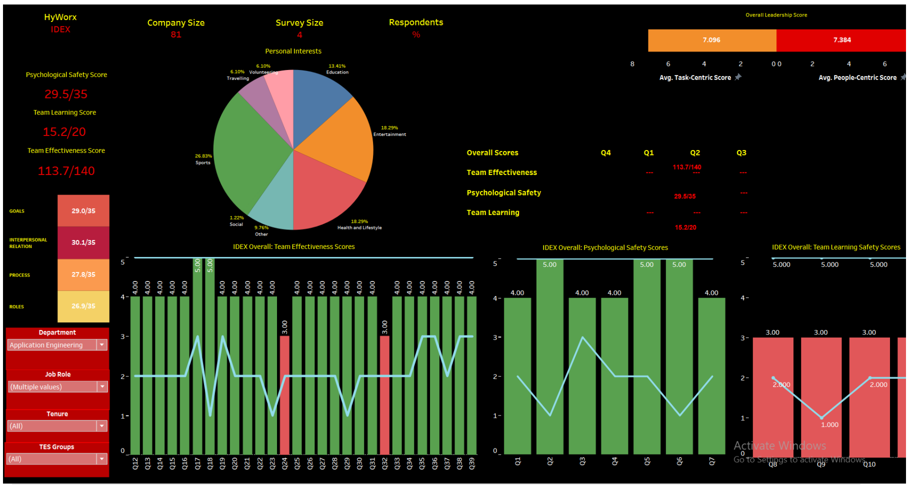
## 5.5 Tableau Integration

Next the results of the system imported to tableau and then with dataset users categories as in shown Figure 9 to manage and filter easily for organisation. here the metrics and



| Response I.. | Geographic.. | Department | email | |
|---|---|---|---|---|
| HW01 | Other | Commercial | HW01@hyworx.co.. | Ab |
| HW02 | USA | Silicon & Se.. | HW02@hyworx.co.. | Ab |
| HW03 | USA | Finance, Ad.. | HW03@hyworx.co.. | Ab |
| HW04 | USA | Software | HW04@hyworx.co.. | Ab |
| HW05 | UK | Operations | HW05@hyworx.co.. | Ab |
| HW06 | USA | Silicon & Se.. | HW06@hyworx.co.. | Ab |
| HW07 | USA | Silicon & Se.. | HW07@hyworx.co.. | Ab |
| HW08 | UK | Systems En.. | HW08@hyworx.co.. | Ab |
| HW09 | UK | Silicon & Se.. | HW09@hyworx.co.. | Ab |
| HW10 | UK | Operations | HW10@hyworx.co.. | Ab |
| HW11 | UK | Software | HW11@hyworx.co.. | Ab |

Figure 9: dashboard for organisation filter employees within organsiation



| Overall Scores | Q4 | Q1 | Q2 | Q3 |
|---|---|---|---|---|
| Team Effectiveness |  | --- | 113.7/140<br>--- | --- |
| Psychological Safety |  |  | 29.5/35 | --- |
| Team Learning |  | --- | ---<br>15.2/20 | --- |

Department
Application Engineering

Job Role
(Multiple values)

Tenure
(All)

TES Groups
(All)

Figure 10: metrics for organisation to asses performance of engagement

visualisation shown in form of dashboard to examine the performance and engagement of employees within organisation as Figure 10

# References

Lahitani, A. R., Permanasari, A. E. and Setiawan, N. A. (2018). Cosine similarity to determine similarity measure: Study case in online essay assessment, *2016 4th International Conference on Cyber and IT Service Management* pp. 1–6.