

**ANL252**

**Python for Data Analysis**

**July 2022 Presentation**

**Tutor-Marked Assignment**

**Submitted by:**

|  |  |
| --- | --- |
| **Name** | **PI No.** |
| **Tan Wee Min Andre** | **E2070932** |

**Tutorial Group: T09**

**Instructor’s Name: Munish Kumar**

**Submission Due Date: 14 August 2022**

“We declare that this assignment is our own work, unless otherwise acknowledge or credited by appropriate referencing. We have read and abide by the SUSS Honour Code and we are aware of the penalties associated with plagiarism and collusion listed in the SUSS Student Handbook.”

Q1)

(a)

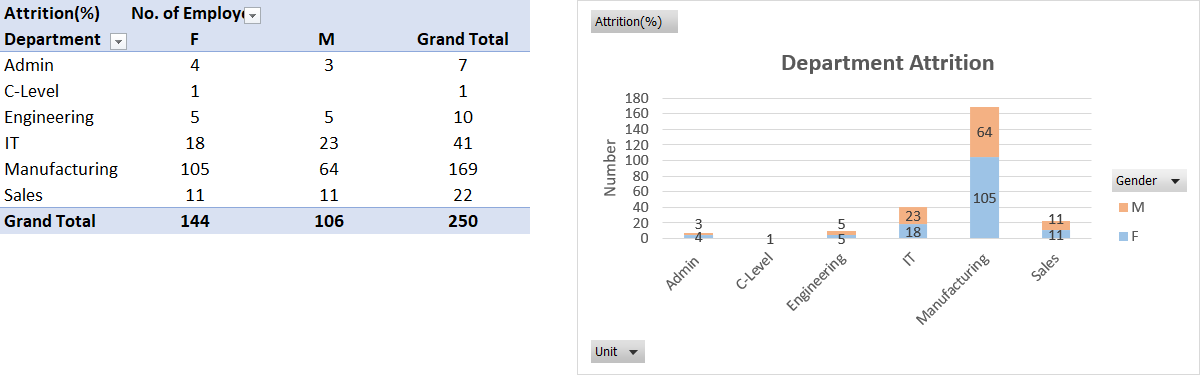


Fig 1. Attrition table and corresponding chart

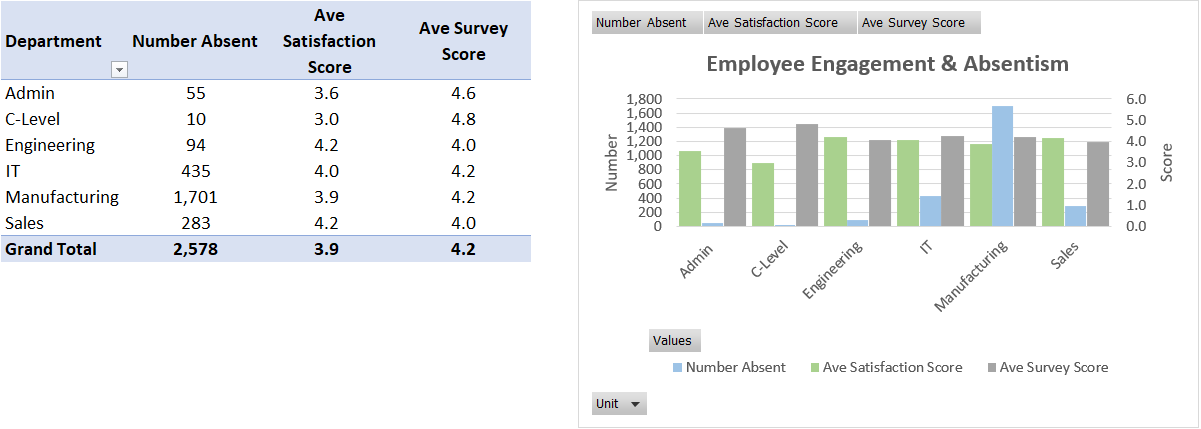


Fig 2. Employee engagement & absenteeism table and corresponding chart

Figure 1 shows the table and chart of employee attrition of each department in the company. Figure 2 shows the table and chart of employee engagement and absenteeism of each department in the company. After comparison, a trend seemed to have arisen between the two charts. The trend of absenteeism level of each department follows closely with the attrition level. It may be possible that employees losing their sense of belonging within the company over time tries to absent themselves from work whether incidental or otherwise. We cannot automatically conclude that absenteeism drives attrition levels but it can be said that there is plausible relationship and further study into the two outcome is necessary.

(b)

import pandas as pd

import numpy as np

import os as path

import matplotlib.pyplot as plt

from datetime import datetime as dt

path.chdir("D:\My Docs\Studies\Singapore Uni of Social Science (SUSS)\Jul'22 registration\Study Materials\ANL252\TMA")

path.getcwd()

df\_staff = pd.read\_csv("TMA\_Data.csv")

attrit\_pvtbl = pd.pivot\_table(df\_staff, index=["Unit"], columns=["Gender"], values=["ID"], aggfunc="count")

print(attrit\_pvtbl)

ax = attrit\_pvtbl.plot.bar(stacked=True, title="Department Attrition", figsize=(8,4), xlabel="Department")

ax.legend(loc="right",bbox\_to\_anchor=(1.2,0.5))

ax.plot()

resp\_pvtbl = pd.pivot\_table(df\_staff, index=["Unit"], values=["Absence", "Satisfaction", "Survey"], aggfunc={"Absence":np.sum, "Satisfaction":np.mean, "Survey":np.mean})

resp\_pvtbl = resp\_pvtbl.round(decimals = 1)

print(resp\_pvtbl)

ax = resp\_pvtbl.plot.bar(stacked=False, title="Employee Engagement & Absentism", figsize=(8,4), secondary\_y=["Satisfaction", "Survey"], xlabel="Department")

ax.plot()

(c)

nowdate = dt(2022,5,1).date()

length\_col = []

#loop through entire dataframe. using timedelta to compute day differences and convert to years

for i in range(len(df\_staff)):

if(pd.isnull(df\_staff.iloc[i]["LeftDate"]) == True):

timedelta = nowdate - dt.strptime(df\_staff.iloc[i]["JoinDate"],"%m/%d/%Y").date()

length\_col.append(round(timedelta.days/365,1))

else:

timedelta = dt.strptime(df\_staff.iloc[i]["LeftDate"],"%m/%d/%Y").date() - dt.strptime(df\_staff.iloc[i]["JoinDate"],"%m/%d/%Y").date()

length\_col.append(round(timedelta.days/365,1))

#adding service length as new column to right most column of dataframe

df\_staff["ServiceLength"] = length\_col

print("Minimum length of service is {} year(s)".format(df\_staff["ServiceLength"].min(axis=0)))

print("Maximum length of service is {} year(s)".format(df\_staff["ServiceLength"].max(axis=0)))

print("Average length of service is {:.1f} year(s)".format(df\_staff["ServiceLength"].mean(axis=0)))

(d)

proceed = "Y"

while(proceed == "Y" or proceed == "y"):

name\_query = input("Full name of person to search: ")

for j in range(len(df\_staff)):

if(name\_query.lower() == df\_staff.loc[j]["Staff"].lower()):

print("Details retrieved...")

print(df\_staff.loc[j])

break

elif(j == len(df\_staff)-1):

print("Staff not found.")

#else not require because "Staff not found" response to show only when all record search through with no match.

#all conditions covered so else statement not necessary.

proceed = input("Press Y or y to continue querying. Press any other keys to quit.")

print("Stop querying.")

# References

Wu, K. Y. (2022). *ANL252 Python for data analytics (study guide).* Singapore University of Social Sciences.