**BÁO CÁO ĐỒ ÁN MÔN HỌC IE221.M11.VB2**

**SINH VIÊN THỰC HIỆN**

**Mã sinh viên: 20520010 Họ và tên: Phan Thanh Hải**

**TÊN ĐỀ TÀI PHÂN TÍCH DỮ LIỆU HƯỚNG NGHIÊN CỨU (Các nhân tố ảnh hưởng đến tính hữu hiệu của kiểm soát nội bộ đối với Doanh nghiệp sản xuất Tôn tại Việt Nam)**

CÁC NỘI DUNG CẦN BÁO CÁO:

1. **Giới thiệu đồ án**

Đồ án sử dụng Python để phân tích dữ liệu nghiên cứu dựa trên dữ liệu thu thập từ khảo sát người dùng.

Em sử dụng thư viện pandas, numpy, pingouin để hỗ trợ cho phân tích dữ liệu

1. **Tóm tắt quá trình thực hiện**

* Khảo sát thu thập dữ liệu
* Task 1: Nhập khẩu dữ liệu
* Task 2: Làm sạch dữ liệu
* Task 3: Phân tích dữ liệu
* Task 4: Vẽ biểu đồ

1. **Kết quả đạt được**

* Thu thập dữ liệu
* Dữ liệu được làm sạch
* Chứng minh có sự tương quan của các biến trong mô hình

1. **Tài liệu tham khảo**
2. **Phụ lục 1: Giới thiệu (demo) kết quả**

import pandas as pd

import numpy as np

import pingouin as pg

**TASK 1: NHẬP KHẨU DỮ LIỆU**

path = 'data\_python.csv'

dataframe = pd.read\_csv(path)

df = dataframe

df.head()

**TASK 2: LÀM SẠCH DỮ LIỆU**

#2.1 Check Null Values

df.isnull().sum()

df.info()

--> Do đó, dữ liệu không có khoản trắng. Ngoại từ cột "So\_dien\_thoai\_hoac\_email", trường dữ liệu cột này không quan trọng cho nghiên cứu, do đó không nhất thiết phải đầy đủ.

##2.2 Check values of all columns

column\_headers = list(df.columns.values)

n = len(df.columns)

for i in range(n):

print(i, end = " ")

print('- ', end = " ")

print(column\_headers[i], end = " ")

print(set(df[column\_headers[i]]))

Lọc dữ liệu nghiên cứu:

+ Môi trường (đang công tác hoặc đã công tác)

Một trong các môi trường sau:

1.1. Công ty sản xuất Tôn

1.2. Công ty dịch vụ kiểm toán/ kế toán/ tư vấn (Có thực hiện dịch vụ kiểm toán/ kế toán/ tư vấn công ty sản xuất Tôn)

1.3. Trường đại học/ học viện đào tạo khối ngành kinh tế

+ Kinh nghiệm (đang công tác hoặc đã công tác)

Đáp ứng đồng thời các điều kiện kinh nghiệm sau:

2. Ít nhất đã hoặc đang làm/ giảng dạy một trong các lĩnh vực về: kế toán, tài chính, kiểm toán, kiểm soát nội bộ, bán hàng, sản xuất, nhân sự, tổng hợp, quản lý

3. Ít nhất 1 năm kinh nghiệm trong lĩnh vực thuộc mục (2) và trong môi trường thuộc mục (1)

4. Chức danh công việc từ nhân viên trở lên

2.2a Linh\_vuc\_hoat\_dong Column

Chỉ chấp nhận công ty sản xuất, công ty dịch vụ kiểm toán/ kế toán/ tư vấn và các trường đại học/ học viện đào tạo khối ngành kinh tế

+ San xuat

+ Giao duc

print(set(df['Linh\_vuc\_hoat\_dong']))

## Thống kê số lượng các phần tử có trong cột

count = df['Linh\_vuc\_hoat\_dong'].value\_counts()

print(count)

## Xóa các dòng có 'Gia cong', 'Thuong mai'

df = df[df['Linh\_vuc\_hoat\_dong'] != 'Gia cong']

df = df[df['Linh\_vuc\_hoat\_dong'] != 'Thuong mai']

df

## Check đã drop được các phần tử không phải là 'Gia cong', 'Thuong mai' trong cột Linh\_vuc\_hoat\_dong hay chưa

count = df['Linh\_vuc\_hoat\_dong'].value\_counts()

print(count)

2.2b Tham\_nien\_cong\_tac column

Dữ liệu có thâm niên công tác dưới 1 năm sẽ bị loại bỏ

print(set(df['Tham\_nien\_cong\_tac']))

## Thống kê số lượng các phần tử có trong cột

count = df['Tham\_nien\_cong\_tac'].value\_counts()

print(count)

--> Ta thấy không có thâm niên công tác dưới 1 năm

==> DỮ LIỆU ĐÃ SẠCH VÀ CÓ THỂ TIẾP TỤC SỬ DỤNG CHO PHÂN TÍCH TIẾP THEO

**TASK 3: PHÂN TÍCH DỮ LIỆU**

Theo kết quả nghiên cứu định tính, ta có mô hình hồi quy đa biến sau:

HH = β0 + β1 x MT + β2 x RR + β3 x KS + β4 x TT + β5 x GS + e

Trong đó:

HH : Tính hữu hiệu (biến phụ thuộc)

MT : Môi trường kiểm soát (biến độc lập)

RR : Đánh giá rủi ro (biến độc lập)

KS : Hoạt động kiểm soát (biến độc lập)

TT : Hệ thống thông tin và truyền thông (biến độc lập)

GS : Hoạt động giám sát (biến độc lập)

β0 : Hằng số

β1, β2, β3, β4, β5 : Hệ số các biến độc lập (hệ số hồi quy)

e : Phần dư

## 3.1 Thống kê mô tả các biến

df.describe()

# 3.2 Kiểm định độ tin cậy bằng hệ số Cronbach’s Alpha

Kiểm định độ tin cậy bằng hệ số Cronbach’s Alpha là phép kiểm định mức độ tương quan chặt chẽ giữa các biến quan sát với nhau trong cùng một nhân tố, nghĩa là, Cronbach’s Alpha cho biết các biến quan sát của một nhân tố, biến nào đóng góp vào việc đo lường khái niệm nhân tố, biến nào không.

Điều kiện để xét hệ số Cronbach's Alpha phù hợp là:

+ Theo Nguyễn Đình Thọ (2013) có chỉ ra một số điều kiện lý thuyết:

- Hệ số Cronbach’s Alpha chỉ đo lường độ tin cậy của thang đo (từ 3 biến quan sát trở lên), không tính được độ tin cậy cho từng biến quan sát;

- Hệ số Cronbach’s Alpha có giá trị biến thiên trong đoạn từ 0 đến 1. Về lý thuyết, hệ số này càng cao thì thang đo có độ tin cậy càng cao. Tuy nhiên nếu hệ số quá lớn ≥ 0,95 có khả năng có nhiều biến trong thang đo bị trùng lắp.

+ Theo Nunnally (1978), hệ số Cronbach’s Alpha > 0,6 thì thang đo lường đủ điều kiện.

=> Nghĩa là Cronbach’s Alpha cần > 0,6 và < 0,95 thì đạt độ tin cậy.

##Drop các dữ liệu không phải dạng số để tính Cronbach's Alpha

data = df.drop(columns=['Ho\_va\_ten', 'So\_dien\_thoai\_hoac\_email', 'Gioi\_tinh', 'Do\_tuoi', 'Tham\_nien\_cong\_tac', 'Noi\_lam\_viec', 'Nganh', 'Chuc\_vu', 'Bo\_phan', 'Loai\_hinh\_doanh\_nghiep', 'Linh\_vuc\_hoat\_dong', 'Quy\_mo'])

data

data.columns[1]

##Tính Cronbach Alpha

pg.cronbach\_alpha(data)

==> Cronbach's Alpha = 0.916 > 0.65 và < 0.95 cho nên dữ liệu đạt độ tin cậy và có thể sử dụng cho phân tích tiếp theo.

1. Phụ lục 2: docstring