PERSONALITY PREDICTION

A MINI PROJECT REPORT

18CSC305J - ARTIFICIAL INTELLIGENCE

Submitted by

PRASTUTI SANDILYA SARMA
[RA2011003010164]
MANU SRIVASTAVA[RA2011003010150]

Under the guidance of

RAJALAKSHMI M

Assistant Professor, Department of Computer Science and Engineering

in partial fulfillment for the award of the degree

of

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE & ENGINEERING

of

FACULTY OF ENGINEERING AND TECHNOLOGY

S.R.M. Nagar, Kattankulathur, Chengalpattu District

MAY 2023

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY

(Under Section 3 of UGC Act, 1956)

BONAFIDE CERTIFICATE

Certified that Mini project report titled **"PERSONALITY PREDICTION"** is the bonafide work of **PRASTUTI S SARMA (RA2011003010164)** and **MANU SRIVASTAVA (RA2011003010150)** who carried out the minor project under my supervision. Certified further, that to the best of my knowledge, the work reported herein does not form any other project report or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

M. Rajalakshmi 12/5/23

SIGNATURE

Rajalakshmi M
GUIDE
Assistant Professor
Department of Computing Technologies

Department of Computing Technologies Kattankulathur 603 203 SRM IST

M. Pushpalatha

SIGNATURE

Dr. M. Pushpalatha
HOD
Professor and Head
Department of Computing Technologies

ABSTRACT

This project is based on Personality Prediction using AI. Curiosity to predict personality, behavior and need for this is not as new as invent of social media. Personality prediction to better accuracy could be very useful for society.The personality of a human plays a major role in his personal and professional life. Many organizations have also started shortlisting the candidates based on their personality as this increase efficiency of work because the person is working in what he is good at than what he is forced to do.

Personality prediction is a growing field that utilizes machine learning algorithms and natural language processing techniques to analyze a variety of factors, such as behavior, communication style, and self-reported personality traits, in order to predict an individual's personality. The goal of personality prediction is to better understand human behavior and to inform personalized recommendations and interventions, such as in the areas of mental health and career counseling. However, personality prediction must be developed ethically and with privacy considerations in mind, as personal information should be kept confidential and individuals should have the option to opt out of having their data analyzed or used for research purposes. While there are limitations to predicting personality based on self-reported responses, continued research in this area holds promise for improving our understanding of human behavior and facilitating personalized interventions.

TABLE OF CONTENTS

LIST OF FIGURES

ABBREVIATIONS

PP	Personality Prediction
AI	Artificial Intelligence
CCS	Cross-cultural studies
KNN	K-nearest neighbour

CHAPTER 1

INTRODUCTION

Personality prediction is a field of study that seeks to predict an individual's personality traits using machine learning algorithms and natural language processing techniques. Personality traits are the enduring patterns of thought, feeling, and behavior that characterize an individual and influence their responses to various situations. The most widely used model of personality is the "Big Five" personality traits, which include openness, conscientiousness, extra-version, agreeableness, and neuroticism.

Personality prediction can be done using a variety of data sources, such as self-reported personality questionnaires, online behavior, social media activity, and speech patterns. Machine learning algorithms are trained on these datasets to identify patterns and make predictions about an individual's personality.

The applications of personality prediction are vast and include fields such as mental health, career counseling, marketing, and education. For example, personality prediction can be used to identify individuals who are at risk for mental health issues, to provide personalized career guidance based on personality traits, to tailor marketing messages to different personality types, and to develop personalized educational strategies for students based on their personality profiles.

However, personality prediction also raises ethical concerns regarding the privacy and confidentiality of personal information. It is important to ensure that individuals have the option to opt out of having their data analyzed or used for research purposes, and that the data is stored securely. Additionally, it is important to recognize that predicting personality based on self-reported responses or other data sources is not always accurate and should be taken with a grain of salt.

Despite these limitations and ethical considerations, continued research in the field of personality prediction holds promise for improving our understanding of human behavior and for facilitating personalized interventions that can improve people's lives.

CHAPTER 2

LITERATURE SURVEY

Personality prediction is a field of study that seeks to predict an individual's personality traits using machine learning algorithms and natural language processing techniques. Personality traits are the enduring patterns of thought, feeling, and behavior that characterize an individual and influence their responses to various situations. The most widely used model of personality is the "Big Five" personality traits, which include openness, conscientiousness, extraversion, agreeableness, and neuroticism.

Personality prediction can be done using a variety of data sources, such as self-reported personality questionnaires, online behavior, social media activity, and speech patterns. Machine learning algorithms are trained on these datasets to identify patterns and make predictions about an individual's personality.

The applications of personality prediction are vast and include fields such as mental health, career counseling, marketing, and education. For example, personality prediction can be used to identify individuals who are at risk for mental health issues, to provide personalized career guidance based on personality traits, to tailor marketing messages to different personality types, and to develop personalized educational strategies for students based on their personality profiles.

However, personality prediction also raises ethical concerns regarding the privacy and confidentiality of personal information. It is important to ensure that individuals have the option to opt out of having their data analyzed or used for research purposes, and that the data is stored securely. Additionally, it is important to recognize that predicting personality based on self-reported responses or other data sources is not always accurate and should be taken with a grain of salt.

Despite these limitations and ethical considerations, continued research in the field of personality prediction holds promise for improving our understanding of human behavior and for facilitating personalized interventions that can improve people's lives.

CHAPTER 3

SYSTEM ARCHITECTURE AND DESIGN

3.1 ARCHITECHTURE DIAGRAM OF A PROPOSED PERSONALITY PREDICTION PROJECT

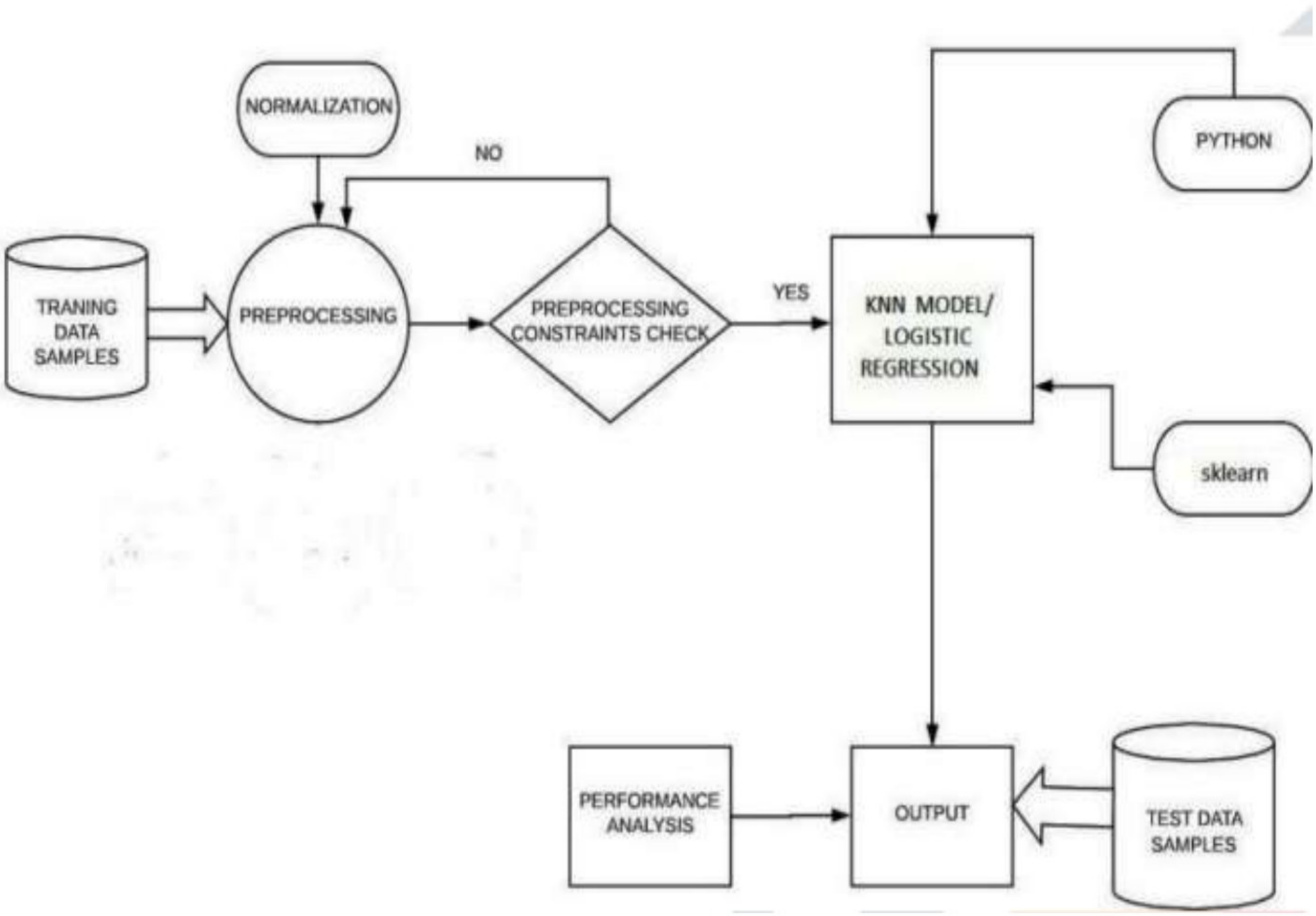


FIG: 3.1.1

3.2 DESCRIPTION OF MODULES AND COMPONENTS

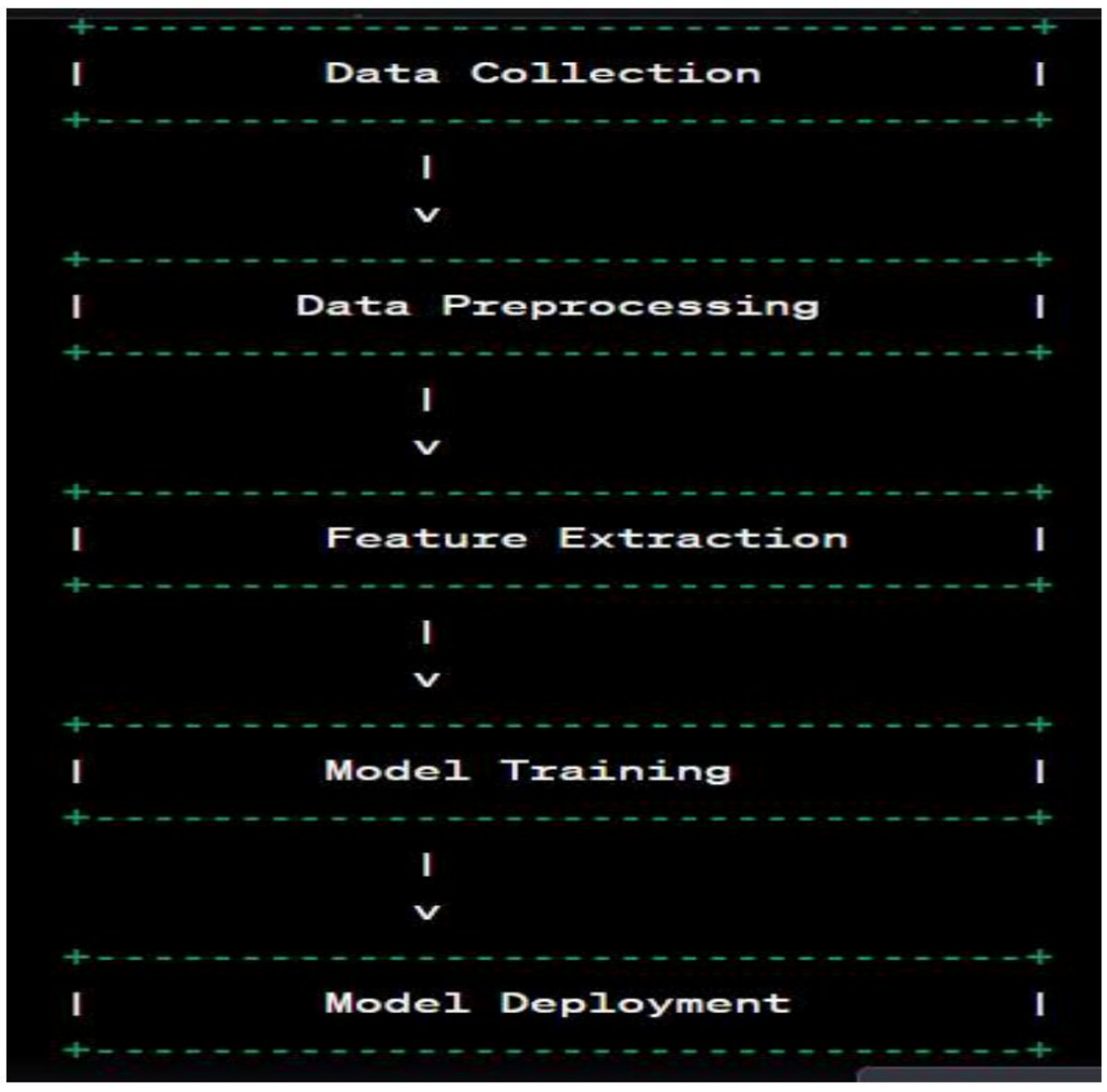


FIG: 3.2.1

Data Collection: The first component of the system architecture involves collecting data from various sources, such as self-reported questionnaires, social media activity, and speech patterns. The data is then processed and cleaned to remove any irrelevant information.

Data Preprocessing: The next component involves processing the data to make it usable for feature extraction. This includes tasks such as cleaning and normalizing the data, dealing with missing values, and transforming it into a suitable format for feature extraction.

Feature Extraction: In this component, relevant features are extracted from the preprocessed data. This includes identifying patterns and correlations in the data that are predictive of personality traits.

Model Training: The extracted features are then used to train machine learning models that predict personality traits. Various algorithms such as Random Forest, Decision Trees, and Support Vector Machines can be used for this purpose.

Model Deployment: The final component involves deploying the trained models into production systems, such as websites or mobile applications, where they can be used to make predictions about the personality traits of users.

CHAPTER 4

METHODOLOGY

4.1 METHODOLGICAL STEPS

Self-report questionnaires: This is a common method for assessing personality traits, where individuals answer a series of questions about themselves and their behavior. The Big Five Inventory is a widely used self-report questionnaire that assesses the five major personality traits.

Behavioral observation: This involves observing individuals in different situations and contexts to assess their behavior and personality traits. This method can provide more objective data than self-report measures, but can be more time-consuming and resource-intensive.

Machine learning algorithms: These can be used to analyze large datasets of personality-related information, such as social media posts or online behavior, to predict personality traits. This method has the advantage of being able to analyze large amounts of data quickly and efficiently, but may require more technical expertise.

Projective techniques: These are methods that involve presenting individuals with ambiguous stimuli, such as inkblots or incomplete sentences, and asking them to interpret or complete them. These techniques can provide insights into an individual's unconscious thoughts and feelings, but may be more subjective and difficult to interpret

CHAPTER 5

CODING AND TESTING


```
{
 "cells": [
  {
   "cell_type": "code",
   "execution_count": 3,
   "metadata": {},
   "outputs": [],
   "source": [
    "#Librerias\n",
    "import pandas as pd \n",
    "import numpy as np\n",
    "import matplotlib.pyplot as plt "
   ]
  },
  {
   "cell_type": "code",
   "execution_count": 4,
   "metadata": {},
   "outputs": [
    {
     "data": {
      "text/plain": [
       "(1010, 150)"
      ]
     },
     "execution_count": 4,
     "metadata": {},
     "output_type": "execute_result"
    }
   ],
   "source": [
    "#Data\n",
    "df = pd.read_csv(\"responses.csv\")\n",
    "df.shape"
   ]
  },
  {
   "cell_type": "markdown",
   "metadata": {},
   "source": [
    "MUSIC PREFERENCES (19) 0:19\n",
    "\n",
    "MOVIE PREFERENCES (12) 19:31\n",
    "\n",
    "HOBBIES & INTERESTS (32) 31:63\n",
    "\n",
    "PHOBIAS (10) 63:73\n",
    "\n",
    "HEALTH HABITS (3) 73:76\n",
    "\n",
```

```
 "PERSONALITY TRAITS, VIEWS ON LIFE & OPINIONS (57) 76:133\n",
 "\n",
 "SPENDING HABITS (7) 133:140\n",
 "\n",
 "DEMOGRAPHICS (10 ) 140:150\n",
 "\n",
 "We will take only: PERSONALITY TRAITS, VIEWS ON LIFE & OPINIONS (57) 76:133"
]
},
{
"cell_type": "code",
"execution_count": 5,
"metadata": {},
"outputs": [
 {
  "data": {
   "text/html": [
    "<div>\n",
    "<style scoped>\n",
    "    .dataframe tbody tr th:only-of-type {\n",
    "        vertical-align: middle;\n",
    "    }\n",
    "\n",
    "    .dataframe tbody tr th {\n",
    "        vertical-align: top;\n",
    "    }\n",
    "\n",
    "    .dataframe thead th {\n",
    "        text-align: right;\n",
    "    }\n",
    "</style>\n",
    "<table border=\"1\" class=\"dataframe\">\n",
    "  <thead>\n",
    "    <tr style=\"text-align: right;\">\n",
    "      <th></th>\n",
    "      <th>Daily events</th>\n",
    "      <th>Prioritising workload</th>\n",
    "      <th>Writing notes</th>\n",
    "      <th>Workaholism</th>\n",
    "      <th>Thinking ahead</th>\n",
    "      <th>Final judgement</th>\n",
    "      <th>Reliability</th>\n",
    "      <th>Keeping promises</th>\n",
    "      <th>Loss of interest</th>\n",
    "      <th>Friends versus money</th>\n",
    "      <th>...</th>\n",
    "      <th>Happiness in life</th>\n",
    "      <th>Energy levels</th>\n",
    "      <th>Small - big dogs</th>\n",
    "      <th>Personality</th>\n",
    "      <th>Finding lost valuables</th>\n",
    "      <th>Getting up</th>\n",
    "      <th>Interests or hobbies</th>\n",
    "      <th>Parents' advice</th>\n",
    "      <th>Questionnaires or polls</th>\n",
```

```
"      <th>Internet usage</th>\n",
"    </tr>\n",
"  </thead>\n",
"  <tbody>\n",
"    <tr>\n",
"      <td>0</td>\n",
"      <td>2.0</td>\n",
"      <td>2.0</td>\n",
"      <td>5.0</td>\n",
"      <td>4.0</td>\n",
"      <td>2.0</td>\n",
"      <td>5.0</td>\n",
"      <td>4.0</td>\n",
"      <td>4.0</td>\n",
"      <td>1.0</td>\n",
"      <td>3.0</td>\n",
"      <td>...</td>\n",
"      <td>4.0</td>\n",
"      <td>5.0</td>\n",
"      <td>1.0</td>\n",
"      <td>4.0</td>\n",
"      <td>3.0</td>\n",
"      <td>2.0</td>\n",
"      <td>3.0</td>\n",
"      <td>4.0</td>\n",
"      <td>3.0</td>\n",
"      <td>few hours a day</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>1</td>\n",
"      <td>3.0</td>\n",
"      <td>2.0</td>\n",
"      <td>4.0</td>\n",
"      <td>5.0</td>\n",
"      <td>4.0</td>\n",
"      <td>1.0</td>\n",
"      <td>4.0</td>\n",
"      <td>4.0</td>\n",
"      <td>3.0</td>\n",
"      <td>4.0</td>\n",
"      <td>...</td>\n",
"      <td>4.0</td>\n",
"      <td>3.0</td>\n",
"      <td>5.0</td>\n",
"      <td>3.0</td>\n",
"      <td>4.0</td>\n",
"      <td>5.0</td>\n",
"      <td>3.0</td>\n",
"      <td>2.0</td>\n",
"      <td>3.0</td>\n",
"      <td>few hours a day</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>2</td>\n",
"      <td>1.0</td>\n",
```

```
"     <td>2.0</td>\n",
"     <td>5.0</td>\n",
"     <td>3.0</td>\n",
"     <td>5.0</td>\n",
"     <td>3.0</td>\n",
"     <td>4.0</td>\n",
"     <td>5.0</td>\n",
"     <td>1.0</td>\n",
"     <td>5.0</td>\n",
"     <td>...</td>\n",
"     <td>4.0</td>\n",
"     <td>3.0</td>\n",
"     <td>3.0</td>\n",
"     <td>3.0</td>\n",
"     <td>4.0</td>\n",
"     <td>5.0</td>\n",
"     <td>3.0</td>\n",
"     <td>1.0</td>\n",
"     <td>few hours a day</td>\n",
"   </tr>\n",
"   <tr>\n",
"     <td>3</td>\n",
"     <td>4.0</td>\n",
"     <td>4.0</td>\n",
"     <td>4.0</td>\n",
"     <td>5.0</td>\n",
"     <td>3.0</td>\n",
"     <td>1.0</td>\n",
"     <td>3.0</td>\n",
"     <td>4.0</td>\n",
"     <td>5.0</td>\n",
"     <td>2.0</td>\n",
"     <td>...</td>\n",
"     <td>2.0</td>\n",
"     <td>2.0</td>\n",
"     <td>1.0</td>\n",
"     <td>2.0</td>\n",
"     <td>1.0</td>\n",
"     <td>1.0</td>\n",
"     <td>NaN</td>\n",
"     <td>2.0</td>\n",
"     <td>4.0</td>\n",
"     <td>most of the day</td>\n",
"   </tr>\n",
"   <tr>\n",
"     <td>4</td>\n",
"     <td>3.0</td>\n",
"     <td>1.0</td>\n",
"     <td>2.0</td>\n",
"     <td>3.0</td>\n",
"     <td>5.0</td>\n",
"     <td>5.0</td>\n",
"     <td>5.0</td>\n",
"     <td>4.0</td>\n",
```

```
"      <td>2.0</td>\n",
"      <td>3.0</td>\n",
"      <td>...</td>\n",
"      <td>3.0</td>\n",
"      <td>5.0</td>\n",
"      <td>3.0</td>\n",
"      <td>3.0</td>\n",
"      <td>2.0</td>\n",
"      <td>4.0</td>\n",
"      <td>3.0</td>\n",
"      <td>3.0</td>\n",
"      <td>few hours a day</td>\n",
"    </tr>\n",
"  </tbody>\n",
"</table>\n",
"<p>5 rows × 57 columns</p>\n",
"</div>"
],
"text/plain": [
"   Daily events  Prioritising workload  Writing notes  Workaholism  \\\n",
"0           2.0                    2.0            5.0          4.0   \n",
"1           3.0                    2.0            4.0          5.0   \n",
"2           1.0                    2.0            5.0          3.0   \n",
"3           4.0                    4.0            4.0          5.0   \n",
"4           3.0                    1.0            2.0          3.0   \n",
"\n",
"   Thinking ahead  Final judgement  Reliability  Keeping promises  \\\n",
"0             2.0              5.0          4.0               4.0   \n",
"1             4.0              1.0          4.0               4.0   \n",
"2             5.0              3.0          4.0               5.0   \n",
"3             3.0              1.0          3.0               4.0   \n",
"4             5.0              5.0          5.0               4.0   \n",
"\n",
"   Loss of interest  Friends versus money  ...  Happiness in life  \\\n",
"0               1.0                   3.0  ...                4.0   \n",
"1               3.0                   4.0  ...                4.0   \n",
"2               1.0                   5.0  ...                4.0   \n",
"3               5.0                   2.0  ...                2.0   \n",
"4               2.0                   3.0  ...                3.0   \n",
"\n",
"   Energy levels  Small - big dogs  Personality  Finding lost valuables  \\\n",
"0            5.0               1.0          4.0                     3.0   \n",
"1            3.0               5.0          3.0                     4.0   \n",
"2            4.0               3.0          3.0                     3.0   \n",
"3            2.0               1.0          2.0                     1.0   \n",
"4            5.0               3.0          3.0                     2.0   \n",
"\n",
"   Getting up  Interests or hobbies  Parents' advice  Questionnaires or polls  \\\n",
"0         2.0                   3.0              4.0                      3.0   \n",
"1         5.0                   3.0              2.0                      3.0   \n",
"2         4.0                   5.0              3.0                      1.0   \n",
"3         1.0                   NaN              2.0                      4.0   \n",
"4         4.0                   3.0              3.0                      3.0   \n",
"\n",
```

```
 "    Internet usage  \n",
 "0  few hours a day  \n",
 "1  few hours a day  \n",
 "2  few hours a day  \n",
 "3  most of the day  \n",
 "4  few hours a day  \n",
 "\n",
 "[5 rows x 57 columns]"
]
},
"execution_count": 5,
 "metadata": {},
 "output_type": "execute_result"
}
],
"source": [
 "df = df.iloc[:, 76:133]\n",
 "df.head(5)"
]
},
{
"cell_type": "markdown",
"metadata": {},
"source": [
 "# 1. Prepare the Data"
]
},
{
"cell_type": "code",
"execution_count": 6,
"metadata": {},
"outputs": [
 {
  "data": {
   "text/html": [
    "<div>\n",
    "<style scoped>\n",
    "    .dataframe tbody tr th:only-of-type {\n",
    "        vertical-align: middle;\n",
    "    }\n",
    "\n",
    "    .dataframe tbody tr th {\n",
"        vertical-align: top;\n",
"    }\n",
"\n",
"    .dataframe thead th {\n",
"        text-align: right;\n",
"    }\n",
"</style>\n",
"<table border=\"1\" class=\"dataframe\">\n",
"  <thead>\n",
"    <tr style=\"text-align: right;\">\n",
"      <th></th>\n",
"      <th>Daily events</th>\n",
"      <th>Prioritising workload</th>\n",
```

```
"      <th>Writing notes</th>\n",
"      <th>Workaholism</th>\n",
"      <th>Thinking ahead</th>\n",
"      <th>Final judgement</th>\n",
"      <th>Reliability</th>\n",
"      <th>Keeping promises</th>\n",
"      <th>Loss of interest</th>\n",
"      <th>Friends versus money</th>\n",
"      <th>...</th>\n",
"      <th>Happiness in life</th>\n",
"      <th>Energy levels</th>\n",
"      <th>Small - big dogs</th>\n",
"      <th>Personality</th>\n",
"      <th>Finding lost valuables</th>\n",
"      <th>Getting up</th>\n",
"      <th>Interests or hobbies</th>\n",
"      <th>Parents' advice</th>\n",
"      <th>Questionnaires or polls</th>\n",
"      <th>Internet usage</th>\n",
"    </tr>\n",
"  </thead>\n",
"  <tbody>\n",
"    <tr>\n",
"      <td>0</td>\n",
"      <td>1</td>\n",
"      <td>1</td>\n",
"      <td>4</td>\n",
"      <td>3</td>\n",
"      <td>1</td>\n",
"      <td>4</td>\n",
"      <td>3</td>\n",
"      <td>3</td>\n",
"      <td>0</td>\n",
"      <td>2</td>\n",
"      <td>...</td>\n",
"      <td>3</td>\n",
"      <td>4</td>\n",
"      <td>0</td>\n",
"      <td>3</td>\n",
"      <td>2</td>\n",
"      <td>1</td>\n",
"      <td>2</td>\n",
"      <td>3</td>\n",
"      <td>2</td>\n",
"      <td>0</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>1</td>\n",
"      <td>2</td>\n",
"      <td>1</td>\n",
"      <td>3</td>\n",
"      <td>4</td>\n",
"      <td>3</td>\n",
"      <td>0</td>\n",
"      <td>3</td>\n",
"      <td>2</td>\n",
"      <td>1</td>\n",
"      <td>3</td>\n",
"      <td>4</td>\n",
"      <td>3</td>\n",
"      <td>0</td>\n",
"      <td>3</td>\n",
```

```
"      <td>3</td>\n",
"      <td>2</td>\n",
"      <td>3</td>\n",
"      <td>...</td>\n",
"      <td>3</td>\n",
"      <td>2</td>\n",
"      <td>4</td>\n",
"      <td>2</td>\n",
"      <td>3</td>\n",
"      <td>4</td>\n",
"      <td>2</td>\n",
"      <td>1</td>\n",
"      <td>2</td>\n",
"      <td>0</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>2</td>\n",
"      <td>0</td>\n",
"      <td>1</td>\n",
"      <td>4</td>\n",
"      <td>2</td>\n",
"      <td>4</td>\n",
"      <td>2</td>\n",
"      <td>3</td>\n",
"      <td>4</td>\n",
"      <td>0</td>\n",
"      <td>4</td>\n",
"      <td>...</td>\n",
"      <td>3</td>\n",
"      <td>3</td>\n",
"      <td>2</td>\n",
"      <td>2</td>\n",
"      <td>2</td>\n",
"      <td>3</td>\n",
"      <td>4</td>\n",
"      <td>2</td>\n",
"      <td>0</td>\n",
"      <td>0</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>4</td>\n",
"      <td>2</td>\n",
"      <td>0</td>\n",
"      <td>1</td>\n",
"      <td>2</td>\n",
"      <td>4</td>\n",
"      <td>4</td>\n",
"      <td>4</td>\n",
"      <td>3</td>\n",
"      <td>1</td>\n",
"      <td>2</td>\n",
"      <td>...</td>\n",
"      <td>2</td>\n",
"      <td>4</td>\n",
"      <td>2</td>\n",
"      <td>2</td>\n",
```

```
"       <td>2</td>\n",
"       <td>1</td>\n",
"       <td>3</td>\n",
"       <td>2</td>\n",
"       <td>2</td>\n",
"       <td>2</td>\n",
"       <td>0</td>\n",
"     </tr>\n",
"     <tr>\n",
"       <td>5</td>\n",
"       <td>1</td>\n",
"       <td>1</td>\n",
"       <td>2</td>\n",
"       <td>2</td>\n",
"       <td>2</td>\n",
"       <td>0</td>\n",
"       <td>2</td>\n",
"       <td>3</td>\n",
"       <td>2</td>\n",
"       <td>1</td>\n",
"       <td>...</td>\n",
"       <td>2</td>\n",
"       <td>3</td>\n",
"       <td>3</td>\n",
"       <td>2</td>\n",
"       <td>2</td>\n",
"       <td>2</td>\n",
"       <td>4</td>\n",
"       <td>2</td>\n",
"       <td>3</td>\n",
"       <td>0</td>\n",
"     </tr>\n",
"     <tr>\n",
"       <td>...</td>\n",
"       <td>...</td>\n",
"       <td>...</td>\n",
"       <td>...</td>\n",
"       <td>...</td>\n",
"       <td>...</td>\n",
"       <td>...</td>\n",
"       <td>...</td>\n",
"       <td>...</td>\n",
"       <td>...</td>\n",
"       <td>...</td>\n",
"       <td>...</td>\n",
"       <td>...</td>\n",
"       <td>...</td>\n",
"       <td>...</td>\n",
"       <td>...</td>\n",
"       <td>...</td>\n",
"       <td>...</td>\n",
"       <td>...</td>\n",
"       <td>...</td>\n",
"       <td>...</td>\n",
"       <td>...</td>\n",
"       <td>...</td>\n",
```

```
"    </tr>\n",
"    <tr>\n",
"      <td>1005</td>\n",
"      <td>2</td>\n",
"      <td>1</td>\n",
"      <td>0</td>\n",
"      <td>3</td>\n",
"      <td>1</td>\n",
"      <td>2</td>\n",
"      <td>2</td>\n",
"      <td>2</td>\n",
"      <td>3</td>\n",
"      <td>3</td>\n",
"      <td>...</td>\n",
"      <td>3</td>\n",
"      <td>2</td>\n",
"      <td>2</td>\n",
"      <td>2</td>\n",
"      <td>3</td>\n",
"      <td>4</td>\n",
"      <td>3</td>\n",
"      <td>3</td>\n",
"      <td>2</td>\n",
"      <td>0</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>1006</td>\n",
"      <td>0</td>\n",
"      <td>2</td>\n",
"      <td>0</td>\n",
"      <td>4</td>\n",
"      <td>4</td>\n",
"      <td>4</td>\n",
"      <td>4</td>\n",
"      <td>3</td>\n",
"      <td>0</td>\n",
"      <td>1</td>\n",
"      <td>...</td>\n",
"      <td>3</td>\n",
"      <td>3</td>\n",
"      <td>2</td>\n",
"      <td>4</td>\n",
"      <td>2</td>\n",
"      <td>0</td>\n",
"      <td>2</td>\n",
"      <td>3</td>\n",
"      <td>2</td>\n",
"      <td>1</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>1007</td>\n",
"      <td>2</td>\n",
"      <td>0</td>\n",
"      <td>0</td>\n",
"      <td>0</td>\n",
"      <td>0</td>\n",
```

```
"      <td>3</td>\n",
"      <td>0</td>\n",
"      <td>2</td>\n",
"      <td>4</td>\n",
"      <td>0</td>\n",
"      <td>3</td>\n",
"      <td>...</td>\n",
"      <td>2</td>\n",
"      <td>0</td>\n",
"      <td>2</td>\n",
"      <td>1</td>\n",
"      <td>2</td>\n",
"      <td>4</td>\n",
"      <td>0</td>\n",
"      <td>3</td>\n",
"      <td>4</td>\n",
"      <td>2</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>1008</td>\n",
"      <td>2</td>\n",
"      <td>0</td>\n",
"      <td>4</td>\n",
"      <td>0</td>\n",
"      <td>2</td>\n",
"      <td>3</td>\n",
"      <td>3</td>\n",
"      <td>4</td>\n",
"      <td>2</td>\n",
"      <td>...</td>\n",
"      <td>2</td>\n",
"      <td>1</td>\n",
"      <td>1</td>\n",
"      <td>3</td>\n",
"      <td>0</td>\n",
"      <td>4</td>\n",
"      <td>2</td>\n",
"      <td>2</td>\n",
"      <td>2</td>\n",
"      <td>2</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>1009</td>\n",
"      <td>2</td>\n",
"      <td>4</td>\n",
"      <td>3</td>\n",
"      <td>4</td>\n",
"      <td>3</td>\n",
"      <td>2</td>\n",
"      <td>4</td>\n",
"      <td>4</td>\n",
"      <td>2</td>\n",
"      <td>3</td>\n",
"      <td>...</td>\n",
```

```
"      <td>3</td>\n",
"      <td>1</td>\n",
"      <td>2</td>\n",
"      <td>3</td>\n",
"      <td>0</td>\n",
"      <td>1</td>\n",
"      <td>1</td>\n",
"      <td>2</td>\n",
"      <td>4</td>\n",
"      <td>0</td>\n",
"    </tr>\n",
"  </tbody>\n",
"</table>\n",
"<p>864 rows × 57 columns</p>\n",
"</div>"
],
"text/plain": [
"      Daily events  Prioritising workload  Writing notes  Workaholism  \\\n",
"0                1                      1              4            3   \n",
"1                2                      1              3            4   \n",
"2                0                      1              4            2   \n",
"4                2                      0              1            2   \n",
"5                1                      1              2            2   \n",
"...            ...                    ...            ...          ...  \n",
"1005             2                      1              0            3   \n",
"1006             0                      2              0            4   \n",
"1007             2                      0              0            0   \n",
"1008             2                      0              4            0   \n",
"1009             2                      4              3            4   \n",
"\n",
"      Thinking ahead  Final judgement  Reliability  Keeping promises  \\\n",
"0                  1                4            3                 3   \n",
"1                  3                0            3                 3   \n",
"2                  4                2            3                 4   \n",
"4                  4                4            4                 3   \n",
"5                  2                0            2                 3   \n",
"...              ...              ...          ...               ...   \n",
"1005               1                2            2                 2   \n",
"1006               4                4            4                 3   \n",
"1007               3                0            2                 4   \n",
"1008               2                3            3                 3   \n",
"1009               3                2            4                 4   \n",
"\n",
"      Loss of interest  Friends versus money  ...  Happiness in life  \\\n",
"0                    0                     2  ...                  3   \n",
"1                    2                     3  ...                  3   \n",
"2                    0                     4  ...                  3   \n",
"4                    1                     2  ...                  2   \n",
"5                    2                     1  ...                  2   \n",
"...                ...                   ...  ...                ...   \n",
"1005                 3                     3  ...                  3   \n",
"1006                 0                     1  ...                  3   \n",
"1007                 0                     3  ...                  2   \n",
"1008                 4                     2  ...                  2   \n",
"1009                 2                     3  ...                  3   \n",
```

```
"\n",
"      Energy levels  Small - big dogs  Personality  Finding lost valuables  \\\n",
"0                 4                 0            3                       2  \n",
"1                 2                 4            2                       3  \n",
"2                 3                 2            2                       2  \n",
"4                 4                 2            2                       1  \n",
"5                 3                 3            2                       2  \n",
"...             ...               ...          ...                     ...  \n",
"1005              2                 2            2                       3  \n",
"1006              3                 2            4                       2  \n",
"1007              0                 2            1                       2  \n",
"1008              1                 1            3                       0  \n",
"1009              1                 2            3                       0  \n",
"\n",
"      Getting up  Interests or hobbies  Parents' advice  \\\n",
"0              1                     2                3  \n",
"1              4                     2                1  \n",
"2              3                     4                2  \n",
"4              3                     2                2  \n",
"5              2                     4                2  \n",
"...          ...                   ...              ...  \n",
"1005           4                     3                3  \n",
"1006           0                     2                3  \n",
"1007           4                     0                3  \n",
"1008           4                     2                2  \n",
"1009           1                     1                2  \n",
"\n",
"      Questionnaires or polls  Internet usage  \n",
"0                           2               0  \n",
"1                           2               0  \n",
"2                           0               0  \n",
"4                           2               0  \n",
"5                           3               0  \n",
"...                       ...             ...  \n",
"1005                        2               0  \n",
"1006                        2               1  \n",
"1007                        4               2  \n",
"1008                        2               2  \n",
"1009                        4               0  \n",
"\n",
"[864 rows x 57 columns]"
]
},
"execution_count": 6,
"metadata": {},
"output_type": "execute_result"
}
],
"source": [
"#Drop NAs\n",
"df = df.dropna()\n",
"#.....\n",
"#Encode categorical data\n",
"from sklearn.preprocessing import LabelEncoder\n",
"\n",
```

```
 "df = df.apply(LabelEncoder().fit_transform)\n",
 "df"
]
},
{
 "cell_type": "markdown",
 "metadata": {},
 "source": [
 "# 2. Choose the factors "
]
},
 {
  "cell_type": "code",
  "execution_count": 16,
  "metadata": {},
  "outputs": [
   {
    "name": "stdout",
    "output_type": "stream",
    "text": [
     "Requirement already satisfied: factor_analyzer in c:\\users\\dell\\anaconda3\\lib\\site-packages
(0.3.2)\n",
     "Requirement already satisfied: pandas in c:\\users\\dell\\anaconda3\\lib\\site-packages (from
factor_analyzer) (0.25.1)\n",
     "Requirement already satisfied: scipy in c:\\users\\dell\\anaconda3\\lib\\site-packages (from
factor_analyzer) (1.3.1)\n",
     "Requirement already satisfied: numpy in c:\\users\\dell\\anaconda3\\lib\\site-packages (from
factor_analyzer) (1.16.5)\n",
     "Requirement already satisfied: scikit-learn in c:\\users\\dell\\anaconda3\\lib\\site-packages (from
factor_analyzer) (0.21.3)\n",
     "Requirement already satisfied: pytz>=2017.2 in c:\\users\\dell\\anaconda3\\lib\\site-packages (from
pandas->factor_analyzer) (2019.3)\n",
     "Requirement already satisfied: python-dateutil>=2.6.1 in c:\\users\\dell\\anaconda3\\lib\\site-
packages (from pandas->factor_analyzer) (2.8.0)\n",
     "Requirement already satisfied: joblib>=0.11 in c:\\users\\dell\\anaconda3\\lib\\site-packages (from
scikit-learn->factor_analyzer) (0.13.2)\n",
     "Requirement already satisfied: six>=1.5 in c:\\users\\dell\\anaconda3\\lib\\site-packages (from
python-dateutil>=2.6.1->pandas->factor_analyzer) (1.12.0)\n",
     "Note: you may need to restart the kernel to use updated packages.\n"
    ]
   }
  ],
 "source": [
  "pip install factor_analyzer "
 ]
},
{
 "cell_type": "code",
 "execution_count": 9,
 "metadata": {},
 "outputs": [
  {
   "data": {
    "image/png": " [
     "<Figure size 432x288 with 1 Axes>"
```

```
   ]
  },
  "metadata": {
   "needs_background": "light"
  },
  "output_type": "display_data"
 }
],
"source": [
 "#Try the model with all the variables \n",
 "from factor_analyzer import FactorAnalyzer        # pip install factor_analyzer \n",
 "fa = FactorAnalyzer(rotation=\"varimax\")\n",
 "fa.fit(df) \n",
 "\n",
 "# Check Eigenvalues\n",
 "ev, v = fa.get_eigenvalues()\n",
 "ev\n",
 "\n",
 "# Create scree plot using matplotlib\n",
 "plt.scatter(range(1,df.shape[1]+1),ev)\n",
 "plt.plot(range(1,df.shape[1]+1),ev)\n",
 "plt.title('Scree Plot')\n",
 "plt.xlabel('Factors')\n",
 "plt.ylabel('Eigenvalue')\n",
 "plt.grid()\n",
 "plt.show()"
]
},
{
"cell_type": "code",
"execution_count": 15,
"metadata": {},
"outputs": [
 {
  "data": {
   "image/png":       "text/plain": [
    "<IPython.core.display.Image object>"
   ]
  },
  "metadata": {},
  "output_type": "display_data"
 }
],
"source": [
 "from IPython.display import display, Image\n",
 "display(Image(filename='Varimax.png'))"
]
},
{
"cell_type": "markdown",
"metadata": {},
"source": [
 "As you can see the most usefull factors for explain the data are between 5-6 until falling significantly.\n",
 "\n",
```

```
 "We will fit the model with 5 Factors:"
]
},
{
 "cell_type": "code",
 "execution_count": 18,
 "metadata": {},
 "outputs": [
  {
   "data": {
    "text/html": [
"<div>\n",
"<style scoped>\n",
"    .dataframe tbody tr th:only-of-type {\n",
"        vertical-align: middle;\n",
"    }\n",
"\n",
"    .dataframe tbody tr th {\n",
"        vertical-align: top;\n",
"    }\n",
"\n",
"    .dataframe thead th {\n",
"        text-align: right;\n",
"    }\n",
"</style>\n",
"<table border=\"1\" class=\"dataframe\">\n",
"  <thead>\n",
"    <tr style=\"text-align: right;\">\n",
"      <th></th>\n",
"      <th>0</th>\n",
"      <th>1</th>\n",
"      <th>2</th>\n",
"      <th>3</th>\n",
"      <th>4</th>\n",
"    </tr>\n",
"  </thead>\n",
"  <tbody>\n",
"    <tr>\n",
"      <td>Daily events</td>\n",
"      <td>0.250416</td>\n",
"      <td>0.058953</td>\n",
"      <td>0.206877</td>\n",
"      <td>0.026094</td>\n",
"      <td>0.028915</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>Prioritising workload</td>\n",
"      <td>-0.012803</td>\n",
"      <td>-0.150045</td>\n",
"      <td>0.555946</td>\n",
"      <td>0.078913</td>\n",
"      <td>0.128156</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>Writing notes</td>\n",
```

```
"      <td>-0.006039</td>\n",
"      <td>-0.015927</td>\n",
"      <td>0.420849</td>\n",
"      <td>0.225307</td>\n",
"      <td>0.261380</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>Workaholism</td>\n",
"      <td>0.069524</td>\n",
"      <td>0.029275</td>\n",
"      <td>0.527082</td>\n",
"      <td>0.088573</td>\n",
"      <td>0.032979</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>Thinking ahead</td>\n",
"      <td>0.023475</td>\n",
"      <td>0.127909</td>\n",
"      <td>0.530457</td>\n",
"      <td>0.035213</td>\n",
"      <td>0.055426</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>Final judgement</td>\n",
"      <td>0.046188</td>\n",
"      <td>0.112493</td>\n",
"      <td>0.119861</td>\n",
"      <td>0.381338</td>\n",
"      <td>-0.039756</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>Reliability</td>\n",
"      <td>0.061028</td>\n",
"      <td>-0.102481</td>\n",
"      <td>0.539373</td>\n",
"      <td>0.073534</td>\n",
"      <td>-0.003491</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>Keeping promises</td>\n",
"      <td>0.053358</td>\n",
"      <td>-0.034661</td>\n",
"      <td>0.420538</td>\n",
"      <td>0.121450</td>\n",
"      <td>-0.033511</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>Loss of interest</td>\n",
"      <td>0.273777</td>\n",
"      <td>0.226286</td>\n",
"      <td>0.003524</td>\n",
"      <td>-0.149262</td>\n",
"      <td>0.101882</td>\n",
"    </tr>\n",
"    <tr>\n",
```

```
"      <td>Friends versus money</td>\n",
"      <td>0.021279</td>\n",
"      <td>-0.111839</td>\n",
"      <td>0.022026</td>\n",
"      <td>0.381357</td>\n",
"      <td>-0.045824</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>Funniness</td>\n",
"      <td>0.312861</td>\n",
"      <td>0.131400</td>\n",
"      <td>-0.043014</td>\n",
"      <td>-0.018258</td>\n",
"      <td>-0.026083</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>Fake</td>\n",
"      <td>0.091188</td>\n",
"      <td>0.469616</td>\n",
"      <td>-0.024535</td>\n",
"      <td>-0.191798</td>\n",
"      <td>0.019356</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>Criminal damage</td>\n",
"      <td>0.154868</td>\n",
"      <td>0.177732</td>\n",
"      <td>-0.112659</td>\n",
"      <td>-0.240721</td>\n",
"      <td>0.266761</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>Decision making</td>\n",
"      <td>-0.287128</td>\n",
"      <td>0.102033</td>\n",
"      <td>0.267415</td>\n",
"      <td>0.129336</td>\n",
"      <td>0.158694</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>Elections</td>\n",
"      <td>0.074306</td>\n",
"      <td>-0.015585</td>\n",
"      <td>0.222003</td>\n",
"      <td>0.131404</td>\n",
"      <td>-0.083563</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>Self-criticism</td>\n",
"      <td>-0.016858</td>\n",
"      <td>0.398420</td>\n",
"      <td>0.229116</td>\n",
"      <td>0.114144</td>\n",
"      <td>0.069707</td>\n",
"    </tr>\n",
```

```
"    <tr>\n",
"      <td>Judgment calls</td>\n",
"      <td>0.182082</td>\n",
"      <td>-0.010461</td>\n",
"      <td>0.102263</td>\n",
"      <td>0.035675</td>\n",
"      <td>0.086474</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>Hypochondria</td>\n",
"      <td>-0.040254</td>\n",
"      <td>0.258913</td>\n",
"      <td>-0.034874</td>\n",
"      <td>0.042981</td>\n",
"      <td>0.213548</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>Empathy</td>\n",
"      <td>-0.050152</td>\n",
"      <td>-0.073697</td>\n",
"      <td>0.059441</td>\n",
"      <td>0.324982</td>\n",
"      <td>0.133754</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>Eating to survive</td>\n",
"      <td>-0.010608</td>\n",
"      <td>0.183045</td>\n",
"      <td>0.003261</td>\n",
"      <td>-0.015131</td>\n",
"      <td>-0.018874</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>Giving</td>\n",
"      <td>0.082276</td>\n",
"      <td>-0.154549</td>\n",
"      <td>0.112481</td>\n",
"      <td>0.376723</td>\n",
"      <td>0.234000</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>Compassion to animals</td>\n",
"      <td>-0.083505</td>\n",
"      <td>-0.002767</td>\n",
"      <td>-0.010424</td>\n",
"      <td>0.262183</td>\n",
"      <td>0.192734</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>Borrowed stuff</td>\n",
"      <td>-0.097017</td>\n",
"      <td>-0.023047</td>\n",
"      <td>0.323253</td>\n",
"      <td>0.171017</td>\n",
"      <td>0.071189</td>\n",
```

```
"    </tr>\n",
"    <tr>\n",
"      <td>Loneliness</td>\n",
"      <td>-0.199197</td>\n",
"      <td>0.542350</td>\n",
"      <td>-0.019272</td>\n",
"      <td>0.045942</td>\n",
"      <td>0.190369</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>Cheating in school</td>\n",
"      <td>0.216223</td>\n",
"      <td>-0.063183</td>\n",
"      <td>-0.384634</td>\n",
"      <td>-0.083940</td>\n",
"      <td>0.208210</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>Health</td>\n",
"      <td>-0.012267</td>\n",
"      <td>0.027867</td>\n",
"      <td>0.131645</td>\n",
"      <td>0.184296</td>\n",
"      <td>0.437826</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>Changing the past</td>\n",
"      <td>-0.016622</td>\n",
"      <td>0.482307</td>\n",
"      <td>-0.161320</td>\n",
"      <td>0.073843</td>\n",
"      <td>0.159231</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>God</td>\n",
"      <td>0.047894</td>\n",
"      <td>0.032281</td>\n",
"      <td>0.027136</td>\n",
"      <td>0.453873</td>\n",
"      <td>-0.025963</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>Dreams</td>\n",
"      <td>0.207076</td>\n",
"      <td>-0.187723</td>\n",
"      <td>0.078634</td>\n",
"      <td>0.037709</td>\n",
"      <td>-0.124853</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>Charity</td>\n",
"      <td>0.163161</td>\n",
"      <td>0.116834</td>\n",
"      <td>0.156898</td>\n",
"      <td>0.354953</td>\n",
```

```
"      <td>-0.067795</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>Number of friends</td>\n",
"      <td>0.514994</td>\n",
"      <td>-0.321738</td>\n",
"      <td>-0.086711</td>\n",
"      <td>0.241070</td>\n",
"      <td>-0.006859</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>Punctuality</td>\n",
"      <td>0.004662</td>\n",
"      <td>0.090531</td>\n",
"      <td>-0.143569</td>\n",
"      <td>0.069648</td>\n",
"      <td>0.078111</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>Lying</td>\n",
"      <td>-0.095933</td>\n",
"      <td>-0.193370</td>\n",
"      <td>0.001775</td>\n",
"      <td>0.138092</td>\n",
"      <td>0.006950</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>Waiting</td>\n",
"      <td>0.032019</td>\n",
"      <td>-0.067715</td>\n",
"      <td>-0.000820</td>\n",
"      <td>0.075966</td>\n",
"      <td>-0.329606</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>New environment</td>\n",
"      <td>0.470076</td>\n",
"      <td>-0.129745</td>\n",
"      <td>-0.058912</td>\n",
"      <td>0.005400</td>\n",
"      <td>-0.230743</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>Mood swings</td>\n",
"      <td>-0.086477</td>\n",
"      <td>0.353226</td>\n",
"      <td>-0.041005</td>\n",
"      <td>0.031490</td>\n",
"      <td>0.404388</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>Appearence and gestures</td>\n",
"      <td>0.227246</td>\n",
"      <td>-0.004762</td>\n",
"      <td>0.105894</td>\n",
```

```
"     <td>0.068825</td>\n",
"     <td>0.303119</td>\n",
"   </tr>\n",
"   <tr>\n",
"     <td>Socializing</td>\n",
"     <td>0.537811</td>\n",
"     <td>-0.096245</td>\n",
"     <td>-0.048127</td>\n",
"     <td>0.135323</td>\n",
"     <td>-0.039204</td>\n",
"   </tr>\n",
"   <tr>\n",
"     <td>Achievements</td>\n",
"     <td>0.252835</td>\n",
"     <td>0.048658</td>\n",
"     <td>-0.042799</td>\n",
"     <td>-0.082401</td>\n",
"     <td>0.111902</td>\n",
"   </tr>\n",
"   <tr>\n",
"     <td>Responding to a serious letter</td>\n",
"     <td>-0.126985</td>\n",
"     <td>0.087976</td>\n",
"     <td>-0.026876</td>\n",
"     <td>0.022940</td>\n",
"     <td>0.013346</td>\n",
"   </tr>\n",
"   <tr>\n",
"     <td>Children</td>\n",
"     <td>0.079877</td>\n",
"     <td>-0.134254</td>\n",
"     <td>0.033040</td>\n",
"     <td>0.440103</td>\n",
"     <td>0.075663</td>\n",
"   </tr>\n",
"   <tr>\n",
"     <td>Assertiveness</td>\n",
"     <td>0.353462</td>\n",
"     <td>-0.094372</td>\n",
"     <td>0.002509</td>\n",
"     <td>-0.067185</td>\n",
"     <td>0.044117</td>\n",
"   </tr>\n",
"   <tr>\n",
"     <td>Getting angry</td>\n",
"     <td>0.051167</td>\n",
"     <td>0.176922</td>\n",
"     <td>-0.086069</td>\n",
"     <td>-0.070837</td>\n",
"     <td>0.532025</td>\n",
"   </tr>\n",
"   <tr>\n",
"     <td>Knowing the right people</td>\n",
"     <td>0.478657</td>\n",
"     <td>0.022868</td>\n",
```

```
"     <td>0.113503</td>\n",
"     <td>-0.045359</td>\n",
"     <td>0.227230</td>\n",
"   </tr>\n",
"   <tr>\n",
"     <td>Public speaking</td>\n",
"     <td>-0.385674</td>\n",
"     <td>0.104662</td>\n",
"     <td>0.069712</td>\n",
"     <td>0.030447</td>\n",
"     <td>0.190834</td>\n",
"   </tr>\n",
"   <tr>\n",
"     <td>Unpopularity</td>\n",
"     <td>-0.082146</td>\n",
"     <td>0.229228</td>\n",
"     <td>0.079173</td>\n",
"     <td>0.241031</td>\n",
"     <td>-0.031212</td>\n",
"   </tr>\n",
"   <tr>\n",
"     <td>Life struggles</td>\n",
"     <td>-0.226293</td>\n",
"     <td>0.057892</td>\n",
"     <td>-0.059615</td>\n",
"     <td>0.384875</td>\n",
"     <td>0.392060</td>\n",
"   </tr>\n",
"   <tr>\n",
"     <td>Happiness in life</td>\n",
"     <td>0.288585</td>\n",
"     <td>-0.541050</td>\n",
"     <td>0.158473</td>\n",
"     <td>0.051235</td>\n",
"     <td>-0.064525</td>\n",
"   </tr>\n",
"   <tr>\n",
"     <td>Energy levels</td>\n",
"     <td>0.499978</td>\n",
"     <td>-0.478860</td>\n",
"     <td>0.037918</td>\n",
"     <td>0.122773</td>\n",
"     <td>-0.025001</td>\n",
"   </tr>\n",
"   <tr>\n",
"     <td>Small - big dogs</td>\n",
"     <td>0.206696</td>\n",
"     <td>0.040211</td>\n",
"     <td>-0.143225</td>\n",
"     <td>-0.203991</td>\n",
"     <td>-0.131298</td>\n",
"   </tr>\n",
"   <tr>\n",
"     <td>Personality</td>\n",
"     <td>0.259646</td>\n",
```

```
"      <td>-0.393197</td>\n",
"      <td>0.064236</td>\n",
"      <td>0.049013</td>\n",
"      <td>-0.056988</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>Finding lost valuables</td>\n",
"      <td>-0.127907</td>\n",
"      <td>-0.011367</td>\n",
"      <td>0.163354</td>\n",
"      <td>0.391951</td>\n",
"      <td>-0.101749</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>Getting up</td>\n",
"      <td>0.012217</td>\n",
"      <td>0.150551</td>\n",
"      <td>-0.312297</td>\n",
"      <td>0.082580</td>\n",
"      <td>0.121198</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>Interests or hobbies</td>\n",
"      <td>0.465627</td>\n",
"      <td>-0.253289</td>\n",
"      <td>0.065015</td>\n",
"      <td>0.144827</td>\n",
"      <td>-0.078694</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>Parents' advice</td>\n",
"      <td>0.022594</td>\n",
"      <td>-0.032871</td>\n",
"      <td>0.243628</td>\n",
"      <td>0.282252</td>\n",
"      <td>0.113225</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>Questionnaires or polls</td>\n",
"      <td>-0.045177</td>\n",
"      <td>0.114865</td>\n",
"      <td>0.154309</td>\n",
"      <td>0.188501</td>\n",
"      <td>-0.032532</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>Internet usage</td>\n",
"      <td>-0.046077</td>\n",
"      <td>0.075435</td>\n",
"      <td>-0.007799</td>\n",
"      <td>-0.081575</td>\n",
"      <td>0.048144</td>\n",
"    </tr>\n",
"  </tbody>\n",
"</table>\n",
```

```
 "</div>"
],
"text/plain": [
 "                          0         1         2         3  \\\n",
 "Daily events               0.250416  0.058953  0.206877  0.026094   \n",
 "Prioritising workload     -0.012803 -0.150045  0.555946  0.078913   \n",
 "Writing notes             -0.006039 -0.015927  0.420849  0.225307   \n",
 "Workaholism                0.069524  0.029275  0.527082  0.088573   \n",
 "Thinking ahead             0.023475  0.127909  0.530457  0.035213   \n",
 "Final judgement            0.046188  0.112493  0.119861  0.381338   \n",
 "Reliability                0.061028 -0.102481  0.539373  0.073534   \n",
 "Keeping promises           0.053358 -0.034661  0.420538  0.121450   \n",
 "Loss of interest           0.273777  0.226286  0.003524 -0.149262   \n",
 "Friends versus money       0.021279 -0.111839  0.022026  0.381357   \n",
 "Funniness                  0.312861  0.131400 -0.043014 -0.018258   \n",
 "Fake                       0.091188  0.469616 -0.024535 -0.191798   \n",
 "Criminal damage            0.154868  0.177732 -0.112659 -0.240721   \n",
 "Decision making           -0.287128  0.102033  0.267415  0.129336   \n",
 "Elections                  0.074306 -0.015585  0.222003  0.131404   \n",
 "Self-criticism            -0.016858  0.398420  0.229116  0.114144   \n",
 "Judgment calls             0.182082 -0.010461  0.102263  0.035675   \n",
 "Hypochondria              -0.040254  0.258913 -0.034874  0.042981   \n",
 "Empathy                   -0.050152 -0.073697  0.059441  0.324982   \n",
 "Eating to survive         -0.010608  0.183045  0.003261 -0.015131   \n",
 "Giving                     0.082276 -0.154549  0.112481  0.376723   \n",
 "Compassion to animals     -0.083505 -0.002767 -0.010424  0.262183   \n",
 "Borrowed stuff            -0.097017 -0.023047  0.323253  0.171017   \n",
 "Loneliness                -0.199197  0.542350 -0.019272  0.045942   \n",
 "Cheating in school         0.216223 -0.063183 -0.384634 -0.083940   \n",
 "Health                    -0.012267  0.027867  0.131645  0.184296   \n",
 "Changing the past         -0.016622  0.482307 -0.161320  0.073843   \n",
 "God                        0.047894  0.032281  0.027136  0.453873   \n",
 "Dreams                     0.207076 -0.187723  0.078634  0.037709   \n",
 "Charity                    0.163161  0.116834  0.156898  0.354953   \n",
 "Number of friends          0.514994 -0.321738 -0.086711  0.241070   \n",
 "Punctuality                0.004662  0.090531 -0.143569  0.069648   \n",
 "Lying                     -0.095933 -0.193370  0.001775  0.138092   \n",
 "Waiting                    0.032019 -0.067715 -0.000820  0.075966   \n",
 "New environment            0.470076 -0.129745 -0.058912  0.005400   \n",
 "Mood swings               -0.086477  0.353226 -0.041005  0.031490   \n",
 "Appearence and gestures    0.227246 -0.004762  0.105894  0.068825   \n",
 "Socializing                0.537811 -0.096245 -0.048127  0.135323   \n",
 "Achievements               0.252835  0.048658 -0.042799 -0.082401   \n",
 "Responding to a serious letter -0.126985  0.087976 -0.026876  0.022940   \n",
 "Children                   0.079877 -0.134254  0.033040  0.440103   \n",
 "Assertiveness              0.353462 -0.094372  0.002509 -0.067185   \n",
 "Getting angry              0.051167  0.176922 -0.086069 -0.070837   \n",
 "Knowing the right people   0.478657  0.022868  0.113503 -0.045359   \n",
 "Public speaking           -0.385674  0.104662  0.069712  0.030447   \n",
 "Unpopularity              -0.082146  0.229228  0.079173  0.241031   \n",
 "Life struggles            -0.226293  0.057892 -0.059615  0.384875   \n",
 "Happiness in life          0.288585 -0.541050  0.158473  0.051235   \n",
 "Energy levels              0.499978 -0.478860  0.037918  0.122773   \n",
 "Small - big dogs           0.206696  0.040211 -0.143225 -0.203991   \n",
 "Personality                0.259646 -0.393197  0.064236  0.049013   \n",
```

```
"Finding lost valuables       -0.127907 -0.011367  0.163354  0.391951   \n",
"Getting up                  0.012217  0.150551 -0.312297  0.082580   \n",
"Interests or hobbies         0.465627 -0.253289  0.065015  0.144827   \n",
"Parents' advice              0.022594 -0.032871  0.243628  0.282252   \n",
"Questionnaires or polls      -0.045177  0.114865  0.154309  0.188501   \n",
"Internet usage              -0.046077  0.075435 -0.007799 -0.081575   \n",
"\n",
"                           4  \n",
"Daily events                0.028915  \n",
"Prioritising workload         0.128156  \n",
"Writing notes                0.261380  \n",
"Workaholism                   0.032979  \n",
"Thinking ahead                0.055426  \n",
"Final judgement              -0.039756  \n",
"Reliability                -0.003491  \n",
"Keeping promises             -0.033511  \n",
"Loss of interest            0.101882  \n",
"Friends versus money          -0.045824  \n",
"Funniness                  -0.026083  \n",
"Fake                       0.019356  \n",
"Criminal damage               0.266761  \n",
"Decision making               0.158694  \n",
"Elections                  -0.083563  \n",
"Self-criticism              0.069707  \n",
"Judgment calls               0.086474  \n",
"Hypochondria                  0.213548  \n",
"Empathy                      0.133754  \n",
"Eating to survive             -0.018874  \n",
"Giving                     0.234000  \n",
"Compassion to animals           0.192734  \n",
"Borrowed stuff               0.071189  \n",
"Loneliness                   0.190369  \n",
"Cheating in school            0.208210  \n",
"Health                     0.437826  \n",
"Changing the past             0.159231  \n",
"God                        -0.025963  \n",
"Dreams                      -0.124853  \n",
"Charity                    -0.067795  \n",
"Number of friends             -0.006859  \n",
"Punctuality                  0.078111  \n",
"Lying                      0.006950  \n",
"Waiting                    -0.329606  \n",
"New environment              -0.230743  \n",
"Mood swings                  0.404388  \n",
"Appearence and gestures         0.303119  \n",
"Socializing                -0.039204  \n",
"Achievements                  0.111902  \n",
"Responding to a serious letter  0.013346  \n",
"Children                   0.075663  \n",
"Assertiveness                0.044117  \n",
"Getting angry                0.532025  \n",
"Knowing the right people        0.227230  \n",
"Public speaking              0.190834  \n",
"Unpopularity                -0.031212  \n",
"Life struggles               0.392060  \n",
```

```
 "Happiness in life            -0.064525  \n",
 "Energy levels                -0.025001  \n",
 "Small - big dogs              -0.131298  \n",
 "Personality                  -0.056988  \n",
 "Finding lost valuables        -0.101749  \n",
 "Getting up                   0.121198  \n",
 "Interests or hobbies          -0.078694  \n",
 "Parents' advice               0.113225  \n",
 "Questionnaires or polls       -0.032532  \n",
 "Internet usage                0.048144  "
]
  },
  "execution_count": 18,
  "metadata": {},
  "output_type": "execute_result"
 }
],
"source": [
 "#Factor analysis with 5 fators\n",
 "fa = FactorAnalyzer(5, rotation=\"varimax\")\n",
 "fa.fit(df)\n",
 "AF = fa.loadings_\n",
 "AF = pd.DataFrame(AF)\n",
 "AF.index = df.columns\n",
 "AF"
]
},
{
"cell_type": "code",
"execution_count": 19,
"metadata": {},
"outputs": [
 {
  "data": {
   "text/html": [
    "<div>\n",
    "<style scoped>\n",
    "    .dataframe tbody tr th:only-of-type {\n",
    "        vertical-align: middle;\n",
    "    }\n",
    "\n",
    "    .dataframe tbody tr th {\n",
    "        vertical-align: top;\n",
    "    }\n",
"\n",
"    .dataframe thead th {\n",
"        text-align: right;\n",
"    }\n",
"</style>\n",
"<table border=\"1\" class=\"dataframe\">\n",
"  <thead>\n",
"    <tr style=\"text-align: right;\">\n",
"      <th></th>\n",
"      <th>FACTOR</th>\n",
"      <th>Variable</th>\n",
```

```
"      <th>Varianza_Explica</th>\n",
"    </tr>\n",
"  </thead>\n",
"  <tbody>\n",
"    <tr>\n",
"      <td>0</td>\n",
"      <td>0</td>\n",
"      <td>New environment</td>\n",
"      <td>0.470076</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>1</td>\n",
"      <td>0</td>\n",
"      <td>Energy levels</td>\n",
"      <td>0.499978</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>2</td>\n",
"      <td>0</td>\n",
"      <td>Number of friends</td>\n",
"      <td>0.514994</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>3</td>\n",
"      <td>0</td>\n",
"      <td>Socializing</td>\n",
"      <td>0.537811</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>4</td>\n",
"      <td>0</td>\n",
"      <td>Knowing the right people</td>\n",
"      <td>0.478657</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>5</td>\n",
"      <td>1</td>\n",
"      <td>Mood swings</td>\n",
"      <td>0.353226</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>6</td>\n",
"      <td>1</td>\n",
"      <td>Self-criticism</td>\n",
"      <td>0.398420</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>7</td>\n",
"      <td>1</td>\n",
"      <td>Fake</td>\n",
"      <td>0.469616</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>8</td>\n",
"      <td>1</td>\n",
```

```
"    <td>Changing the past</td>\n",
"    <td>0.482307</td>\n",
"  </tr>\n",
"  <tr>\n",
"    <td>9</td>\n",
"    <td>1</td>\n",
"    <td>Loneliness</td>\n",
"    <td>0.542350</td>\n",
"  </tr>\n",
"  <tr>\n",
"    <td>10</td>\n",
"    <td>2</td>\n",
"    <td>Writing notes</td>\n",
"    <td>0.420849</td>\n",
"  </tr>\n",
"  <tr>\n",
"    <td>11</td>\n",
"    <td>2</td>\n",
"    <td>Workaholism</td>\n",
"    <td>0.527082</td>\n",
"  </tr>\n",
"  <tr>\n",
"    <td>12</td>\n",
"    <td>2</td>\n",
"    <td>Thinking ahead</td>\n",
"    <td>0.530457</td>\n",
"  </tr>\n",
"  <tr>\n",
"    <td>13</td>\n",
"    <td>2</td>\n",
"    <td>Prioritising workload</td>\n",
"    <td>0.555946</td>\n",
"  </tr>\n",
"  <tr>\n",
"    <td>14</td>\n",
"    <td>2</td>\n",
"    <td>Reliability</td>\n",
"    <td>0.539373</td>\n",
"  </tr>\n",
"  <tr>\n",
"    <td>15</td>\n",
"    <td>3</td>\n",
"    <td>Friends versus money</td>\n",
"    <td>0.381357</td>\n",
"  </tr>\n",
"  <tr>\n",
"    <td>16</td>\n",
"    <td>3</td>\n",
"    <td>Life struggles</td>\n",
"    <td>0.384875</td>\n",
"  </tr>\n",
"  <tr>\n",
"    <td>17</td>\n",
"    <td>3</td>\n",
"    <td>Finding lost valuables</td>\n",
```

```
"      <td>0.391951</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>18</td>\n",
"      <td>3</td>\n",
"      <td>Children</td>\n",
"      <td>0.440103</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>19</td>\n",
"      <td>3</td>\n",
"      <td>God</td>\n",
"      <td>0.453873</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>20</td>\n",
"      <td>4</td>\n",
"      <td>Appearence and gestures</td>\n",
"      <td>0.303119</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>21</td>\n",
"      <td>4</td>\n",
"      <td>Life struggles</td>\n",
"      <td>0.392060</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>22</td>\n",
"      <td>4</td>\n",
"      <td>Mood swings</td>\n",
"      <td>0.404388</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>23</td>\n",
"      <td>4</td>\n",
"      <td>Health</td>\n",
"      <td>0.437826</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>24</td>\n",
"      <td>4</td>\n",
"      <td>Getting angry</td>\n",
"      <td>0.532025</td>\n",
"    </tr>\n",
"  </tbody>\n",
"</table>\n",
"</div>"
],
"text/plain": [
"    FACTOR                Variable  Varianza_Explica\n",
"0        0         New environment         0.470076\n",
"1        0           Energy levels         0.499978\n",
"2        0         Number of friends         0.514994\n",
"3        0              Socializing         0.537811\n",
"4        0  Knowing the right people         0.478657\n",
```

```
"5       1          Mood swings       0.353226\n",
"6       1       Self-criticism       0.398420\n",
"7       1                 Fake       0.469616\n",
"8       1    Changing the past       0.482307\n",
"9       1          Loneliness       0.542350\n",
"10      2       Writing notes       0.420849\n",
"11      2          Workaholism       0.527082\n",
"12      2       Thinking ahead       0.530457\n",
"13      2  Prioritising workload       0.555946\n",
"14      2          Reliability       0.539373\n",
"15      3  Friends versus money       0.381357\n",
"16      3       Life struggles       0.384875\n",
"17      3  Finding lost valuables       0.391951\n",
"18      3             Children       0.440103\n",
"19      3                  God       0.453873\n",
"20      4  Appearence and gestures       0.303119\n",
"21      4       Life struggles       0.392060\n",
"22      4         Mood swings       0.404388\n",
"23      4               Health       0.437826\n",
"24      4        Getting angry       0.532025"
   ]
  },
  "execution_count": 19,
  "metadata": {},
  "output_type": "execute_result"
 }
],
"source": [
 "#Get Top variables for each Factor \n",
 "F = AF.unstack()\n",
 "F = pd.DataFrame(F).reset_index()\n",
 "F = F.sort_values(['level_0',0], ascending=False).groupby('level_0').head(5)    # Top 5 \n",
 "F = F.sort_values(by=\"level_0\")\n",
 "F.columns=[\"FACTOR\",\"Variable\",\"Varianza_Explica\"]\n",
 "F = F.reset_index().drop([\"index\"],axis=1)\n",
 "F"
]
},
{
 "cell_type": "code",
 "execution_count": 20,
 "metadata": {},
"outputs": [
 {
  "data": {
   "text/html": [
    "<div>\n",
    "<style scoped>\n",
    "    .dataframe tbody tr th:only-of-type {\n",
    "        vertical-align: middle;\n",
    "    }\n",
    "\n",
    "    .dataframe tbody tr th {\n",
    "        vertical-align: top;\n",
    "    }\n",
```

```
"\n",
"    .dataframe thead th {\n",
"        text-align: right;\n",
"    }\n",
"</style>\n",
"<table border=\"1\" class=\"dataframe\">\n",
"  <thead>\n",
"    <tr style=\"text-align: right;\">\n",
"      <th>FACTOR</th>\n",
"      <th>0</th>\n",
"      <th>1</th>\n",
"      <th>2</th>\n",
"      <th>3</th>\n",
"      <th>4</th>\n",
"    </tr>\n",
"  </thead>\n",
"  <tbody>\n",
"    <tr>\n",
"      <td>0</td>\n",
"      <td>New environment</td>\n",
"      <td>Mood swings</td>\n",
"      <td>Writing notes</td>\n",
"      <td>Friends versus money</td>\n",
"      <td>Appearence and gestures</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>1</td>\n",
"      <td>Energy levels</td>\n",
"      <td>Self-criticism</td>\n",
"      <td>Workaholism</td>\n",
"      <td>Life struggles</td>\n",
"      <td>Life struggles</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>2</td>\n",
"      <td>Number of friends</td>\n",
"      <td>Fake</td>\n",
"      <td>Thinking ahead</td>\n",
"      <td>Finding lost valuables</td>\n",
"      <td>Mood swings</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>3</td>\n",
"      <td>Socializing</td>\n",
"      <td>Changing the past</td>\n",
"      <td>Prioritising workload</td>\n",
"      <td>Children</td>\n",
"      <td>Health</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <td>4</td>\n",
"      <td>Knowing the right people</td>\n",
"      <td>Loneliness</td>\n",
"      <td>Reliability</td>\n",
"      <td>God</td>\n",
"
```

```
 "      <td>Getting angry</td>\n",
 "    </tr>\n",
 "  </tbody>\n",
 "</table>\n",
 "</div>"
],
"text/plain": [
 "FACTOR                    0             1                 2  \\\n",
 "0              New environment       Mood swings         Writing notes   \n",
 "1                Energy levels    Self-criticism           Workaholism   \n",
 "2            Number of friends              Fake       Thinking ahead   \n",
 "3                   Socializing  Changing the past  Prioritising workload   \n",
 "4       Knowing the right people        Loneliness           Reliability   \n",
 "\n",
 "FACTOR                    3                    4  \n",
 "0        Friends versus money  Appearence and gestures  \n",
 "1             Life struggles         Life struggles  \n",
 "2       Finding lost valuables           Mood swings  \n",
 "3                    Children               Health  \n",
 "4                         God        Getting angry  "
]
},
"execution_count": 20,
"metadata": {},
"output_type": "execute_result"
}
],
"source": [
 "#Show the Top for each Factor \n",
 "F = F.pivot(columns='FACTOR')[\"Variable\"]\n",
 "F.apply(lambda x: pd.Series(x.dropna().to_numpy()))"
]
},
{
"cell_type": "markdown",
"metadata": {},
"source": [
 "FACTOR 1: Energy levels, Number of friends, Socializing...\n",
 "\n",
 "Could be: Extraversion\n",
 "\n",
 "FACTOR 2: Self-ciricism, Fake, Loneliness...\n",
 "\n",
 "Looks very similar to \"Neuroticism\"\n",
 "\n",
 "Factor 3: Thinking ahead, Prioritising workload...\n",
 "\n",
 "very similar to \"Conscientiousness\"\n",
 "\n",
 "Factor 4: Children, God, Finding lost valuables\n",
 "\n",
 "This factor could be something like \"religious\" or \"conservative\", maybe have lowest scores of a \"Openness\" in Big Five model.\n",
 "\n",
 "Factor 5: Appearence and gestures, Mood swings\n",
```

```
 "\n",
 "Mmmm it could be \"Agreeableness\". What do you think it could be represent?"
]
},
{
 "cell_type": "markdown",
 "metadata": {},
 "source": [
  "# Conclusion:"
 ]
},
 {
  "cell_type": "markdown",
  "metadata": {},
  "source": [
   "The first three Factors are very clear: Extraversion, Neuroticism and Conscientiousness. The other two not to much. Anyway is a very interesting approximation\n",
   "\n",
   "Maybe doing first a PCA for remove hight correlate variables like \"God\" and \"Final judgement\"could help.\n",
   "\n",
   "What do you think?\n",
   "\n",
   "Thanks you!"
  ]
 }
],
"metadata": {
 "kernelspec": {
  "display_name": "Python 3",
  "language": "python",
  "name": "python3"
 },
 "language_info": {
  "codemirror_mode": {
   "name": "ipython",
   "version": 3
  },
  "file_extension": ".py",
  "mimetype": "text/x-python",
  "name": "python",
  "nbconvert_exporter": "python",
  "pygments_lexer": "ipython3",
  "version": "3.7.4"
 }
},
"nbformat": 4,
"nbformat_minor": 2
}
```

CHAPTER 6

SCREENSHOTS AND RESULTS

In [3]:
```
#Librerias
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

In [4]:
```
#Data
df = pd.read_csv("responses.csv")
df.shape
```

Out[4]: (1010, 150)

MUSIC PREFERENCES (19) 0:19

MOVIE PREFERENCES (12) 19:31

HOBBIES & INTERESTS (32) 31:63

PHOBIAS (10) 63:73

HEALTH HABITS (3) 73:76

PERSONALITY TRAITS, VIEWS ON LIFE & OPINIONS (57) 76:133

SPENDING HABITS (7) 133:140

DEMOGRAPHICS (10) 140:150

We will take only: PERSONALITY TRAITS, VIEWS ON LIFE & OPINIONS (57) 76:133

In [5]:
```
df = df.iloc[:, 76:133]
df.head(5)
```

Out[5]:

	Daily events	Prioritising workload	Writing notes	Workaholism	Thinking ahead	Final judgement	Reliability	Keeping promises	Loss of interest	Friends versus money	...	Happiness in life	Energy levels	Small - big dogs	Personality	Finding lost valuables	Ge
0	2.0	2.0	5.0	4.0	2.0	5.0	4.0	4.0	1.0	3.0	...	4.0	5.0	1.0	4.0	3.0	
1	3.0	2.0	4.0	5.0	4.0	1.0	4.0	4.0	3.0	4.0	...	4.0	3.0	5.0	3.0	4.0	
2	1.0	2.0	5.0	3.0	5.0	3.0	4.0	5.0	1.0	5.0	...	4.0	4.0	3.0	3.0	3.0	
3	4.0	4.0	4.0	5.0	3.0	1.0	3.0	4.0	5.0	2.0	...	2.0	2.0	1.0	2.0	1.0	
4	3.0	1.0	2.0	3.0	5.0	5.0	5.0	4.0	2.0	3.0	...	3.0	5.0	3.0	3.0	2.0	

5 rows × 57 columns

1. Prepare the Data

In [6]:
```
#Drop NAs
df = df.dropna()
#.....................................................................................
#Encode categorical data
from sklearn.preprocessing import LabelEncoder

df = df.apply(LabelEncoder().fit_transform)
df
```

Out[6]:

	Daily events	Prioritising workload	Writing notes	Workaholism	Thinking ahead	Final judgement	Reliability	Keeping promises	Loss of interest	Friends versus money	...	Happiness in life	Energy levels	Small - big dogs	Personality	Finding lost valuables
0	1	1	4	3	1	4	3	3	0	2	...	3	4	0	3	2
1	2	1	3	4	3	0	3	3	2	3	...	3	2	4	2	3
2	0	1	4	2	4	2	3	4	0	4	...	3	3	2	2	2
4	2	0	1	2	4	4	4	3	1	2	...	2	4	2	2	1
5	1	1	2	2	2	0	2	3	2	1	...	2	3	3	2	2
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1005	2	1	0	3	1	2	2	2	3	3	...	3	2	2	2	3
1006	0	2	0	4	4	4	4	3	0	1	...	3	3	2	4	2
1007	2	0	0	0	3	0	2	4	0	3	...	2	0	2	1	2
1008	2	0	4	0	2	3	3	3	4	2	...	2	1	1	3	0
1009	2	4	3	4	3	2	4	4	2	3	...	3	1	2	3	0

864 rows × 57 columns

2. Choose the factors

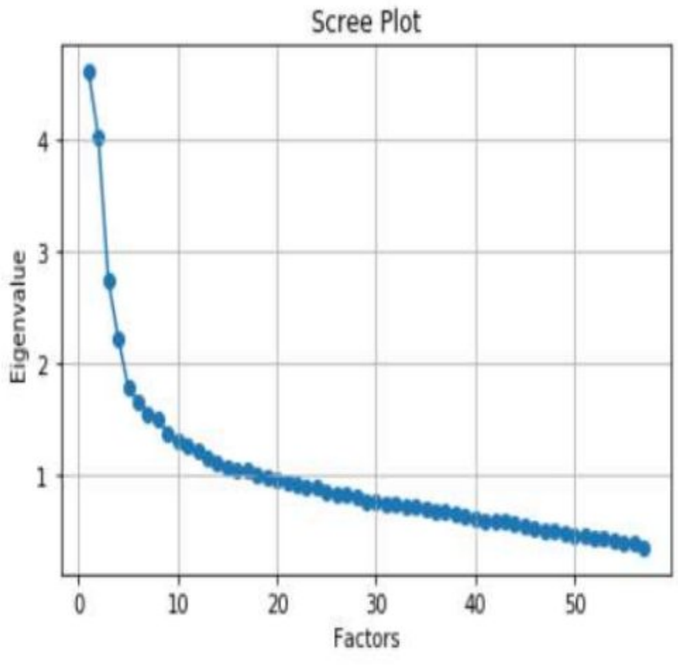
In [16]:
```
pip install factor_analyzer
```

```
Requirement already satisfied: factor_analyzer in c:\users\dell\anaconda3\lib\site-packages (0.3.2)
Requirement already satisfied: pandas in c:\users\dell\anaconda3\lib\site-packages (from factor_analyzer) (0.25.1)
Requirement already satisfied: scipy in c:\users\dell\anaconda3\lib\site-packages (from factor_analyzer) (1.3.1)
Requirement already satisfied: numpy in c:\users\dell\anaconda3\lib\site-packages (from factor_analyzer) (1.16.5)
Requirement already satisfied: scikit-learn in c:\users\dell\anaconda3\lib\site-packages (from factor_analyzer) (0.21.3)
Requirement already satisfied: pytz>=2017.2 in c:\users\dell\anaconda3\lib\site-packages (from pandas->factor_analyzer) (2019.3)
Requirement already satisfied: python-dateutil>=2.6.1 in c:\users\dell\anaconda3\lib\site-packages (from pandas->factor_analyzer) (2.8.0)
Requirement already satisfied: joblib>=0.11 in c:\users\dell\anaconda3\lib\site-packages (from scikit-learn->factor_analyzer) (0.13.2)
Requirement already satisfied: six>=1.5 in c:\users\dell\anaconda3\lib\site-packages (from python-dateutil>=2.6.1->pandas->factor_analyzer) (1.12.0)
Note: you may need to restart the kernel to use updated packages.
```

In [9]:
```
#Try the model with all the variables
from factor_analyzer import FactorAnalyzer        # pip install factor_analyzer
fa = FactorAnalyzer(rotation="varimax")
fa.fit(df)

# Check Eigenvalues
ev, v = fa.get_eigenvalues()
ev

# Create scree plot using matplotlib
plt.scatter(range(1,df.shape[1]+1),ev)
plt.plot(range(1,df.shape[1]+1),ev)
plt.title('Scree Plot')
plt.xlabel('Factors')
plt.ylabel('Eigenvalue')
plt.grid()
plt.show()
```



In [15]:
```
from IPython.display import display, Image
display(Image(filename='Varimax.png'))
```

$$R_{\text{VARIMAX}} = \arg\max_{R} \left(\frac{1}{p} \sum_{j=1}^{k} \sum_{i=1}^{p} (\Lambda R)_{ij}^{4} - \sum_{j=1}^{k} \left(\frac{1}{p} \sum_{i=1}^{p} (\Lambda R)_{ij}^{2} \right)^{2} \right).$$

As you can see the most usefull factors for explain the data are between 5-6 until falling significantly.

We will fit the model with 5 Factors:

In [18]:
```
#Factor analysis with 5 fators
fa = FactorAnalyzer(5, rotation="varimax")
fa.fit(df)
AF = fa.loadings_
AF = pd.DataFrame(AF)
AF.index = df.columns
AF
```

Out[18]:

	0	1	2	3	4
Daily events	0.250416	0.058953	0.206877	0.026094	0.028915
Prioritising workload	-0.012803	-0.150045	0.555946	0.078913	0.128156
Writing notes	-0.006039	-0.015927	0.420849	0.225307	0.261380
Workaholism	0.069524	0.029275	0.527082	0.088573	0.032979
Thinking ahead	0.023475	0.127909	0.530457	0.035213	0.055426
Final judgement	0.046188	0.112493	0.119861	0.381338	-0.039756
Reliability	0.061028	-0.102481	0.539373	0.073534	-0.003491
Keeping promises	0.053358	-0.034661	0.420538	0.121450	-0.033511
Loss of interest	0.273777	0.226286	0.003524	-0.149262	0.101882
Friends versus money	0.021279	-0.111839	0.022026	0.381357	-0.045824
Funniness	0.312861	0.131400	-0.043014	-0.018258	-0.026083
Fake	0.091188	0.469616	-0.024535	-0.191798	0.019356
Criminal damage	0.154868	0.177732	-0.112659	-0.240721	0.266761
Decision making	-0.287128	0.102033	0.267415	0.129336	0.158694
Elections	0.074306	-0.015585	0.222003	0.131404	-0.083563
Self-criticism	-0.016858	0.398420	0.229116	0.114144	0.069707
Judgment calls	0.182082	-0.010461	0.102263	0.035675	0.086474
Hypochondria	-0.040254	0.258913	-0.034874	0.042981	0.213548
Empathy	-0.050152	-0.073697	0.059441	0.324982	0.133754
Eating to survive	-0.010608	0.183045	0.003261	-0.015131	-0.018874
Giving	0.082276	-0.154549	0.112481	0.376723	0.234000
Compassion to animals	-0.083505	-0.002767	-0.010424	0.262183	0.192734
Borrowed stuff	-0.097017	-0.023047	0.323253	0.171017	0.071189
Loneliness	-0.199197	0.542350	-0.019272	0.045942	0.190369
Cheating in school	0.216223	-0.063183	-0.384634	-0.083940	0.208210

Responding to a serious letter	-0.126985	0.087976	-0.026876	0.022940	0.013346
Children	0.079877	-0.134254	0.033040	0.440103	0.075663
Assertiveness	0.353462	-0.094372	0.002509	-0.067185	0.044117
Getting angry	0.051167	0.176922	-0.086069	-0.070837	0.532025
Knowing the right people	0.478657	0.022868	0.113503	-0.045359	0.227230
Public speaking	-0.385674	0.104662	0.069712	0.030447	0.190834
Unpopularity	-0.082146	0.229228	0.079173	0.241031	-0.031212
Life struggles	-0.226293	0.057892	-0.059615	0.384875	0.392060
Happiness in life	0.288585	-0.541050	0.158473	0.051235	-0.064525
Energy levels	0.499978	-0.478860	0.037918	0.122773	-0.025001
Small - big dogs	0.206696	0.040211	-0.143225	-0.203991	-0.131298
Personality	0.259646	-0.393197	0.064236	0.049013	-0.056988
Finding lost valuables	-0.127907	-0.011367	0.163354	0.391951	-0.101749
Getting up	0.012217	0.150551	-0.312297	0.082580	0.121198
Interests or hobbies	0.465627	-0.253289	0.065015	0.144827	-0.078694
Parents' advice	0.022594	-0.032871	0.243628	0.282252	0.113225
Questionnaires or polls	-0.045177	0.114865	0.154309	0.188501	-0.032532
Internet usage	-0.046077	0.075435	-0.007799	-0.081575	0.048144

In [19]:

```
#Get Top variables for each Factor
F = AF.unstack()
F = pd.DataFrame(F).reset_index()
F = F.sort_values(['level_0',0], ascending=False).groupby('level_0').head(5)    # Top 5
F = F.sort_values(by="level_0")
F.columns=["FACTOR","Variable","Varianza_Explica"]
F = F.reset_index().drop(["index"],axis=1)
F
```

In [19]:

```
#Get Top variables for each Factor
F = AF.unstack()
F = pd.DataFrame(F).reset_index()
F = F.sort_values(['level_0',0], ascending=False).groupby('level_0').head(5)    # Top 5
F = F.sort_values(by="level_0")
F.columns=["FACTOR","Variable","Varianza_Explica"]
F = F.reset_index().drop(["index"],axis=1)
F
```

Out[19]:

	FACTOR	Variable	Varianza_Explica
0	0	New environment	0.470076
1	0	Energy levels	0.499978
2	0	Number of friends	0.514994
3	0	Socializing	0.537811
4	0	Knowing the right people	0.478657
5	1	Mood swings	0.353226
6	1	Self-criticism	0.398420
7	1	Fake	0.469616
8	1	Changing the past	0.482307
9	1	Loneliness	0.542350
10	2	Writing notes	0.420849
11	2	Workaholism	0.527082
12	2	Thinking ahead	0.530457
13	2	Prioritising workload	0.555946
14	2	Reliability	0.539373
15	3	Friends versus money	0.381357
16	3	Life struggles	0.384875
17	3	Finding lost valuables	0.391951
18	3	Children	0.440103

11	2	Workaholism	0.527082
12	2	Thinking ahead	0.530457
13	2	Prioritising workload	0.555946
14	2	Reliability	0.539373
15	3	Friends versus money	0.381357
16	3	Life struggles	0.384875
17	3	Finding lost valuables	0.391951
18	3	Children	0.440103
19	3	God	0.453873
20	4	Appearence and gestures	0.303119
21	4	Life struggles	0.392060
22	4	Mood swings	0.404388
23	4	Health	0.437826
24	4	Getting angry	0.532025

In [20]:
```
#Show the Top for each Factor
F = F.pivot(columns='FACTOR')["Variable"]
F.apply(lambda x: pd.Series(x.dropna().to_numpy()))
```

In [20]:
```
#Show the Top for each Factor
F = F.pivot(columns='FACTOR')["Variable"]
F.apply(lambda x: pd.Series(x.dropna().to_numpy()))
```

Out[20]:

FACTOR	0	1	2	3	4
0	New environment	Mood swings	Writing notes	Friends versus money	Appearence and gestures
1	Energy levels	Self-criticism	Workaholism	Life struggles	Life struggles
2	Number of friends	Fake	Thinking ahead	Finding lost valuables	Mood swings
3	Socializing	Changing the past	Prioritising workload	Children	Health
4	Knowing the right people	Loneliness	Reliability	God	Getting angry

FACTOR 1: Energy levels, Number of friends, Socializing...

Could be: Extraversion

FACTOR 2: Self-ciricism, Fake, Loneliness...

Looks very similar to "Neuroticism"

Factor 3: Thinking ahead, Prioritising workload...

very similar to "Conscientiousness"

Factor 4: Children, God, Finding lost valuables

This factor could be something like "religious" or "conservative", maybe have lowest scores of a "Openness" in Big Five model.

Factor 5: Appearence and gestures, Mood swings

Mmmm it could be "Agreeableness". What do you think it could be represent?

CHAPTER 7

CONCLUSION AND FUTURE ENHANCEMENT

7.1 CONCLUSION

In conclusion, personality prediction is a rapidly growing field that uses machine learning and natural language processing techniques to predict an individual's personality traits based on various sources of data. The potential applications of personality prediction are vast, ranging from mental health to career counseling, marketing, and education.

To build an effective personality prediction system, it is crucial to collect and pre process high-quality data, extract relevant features, and train accurate machine learning models. Additionally, it is important to consider the ethical implications of personality prediction and ensure that any predictions made by the system are fair, transparent, and unbiased.

Overall, personality prediction has the potential to revolutionize the way we understand human behavior and improve our ability to provide personalized services and support to individuals. As the field continues to evolve, we can expect to see many new and exciting applications of personality prediction emerge in the years to come.
By utilizing different methodologies such as self-report questionnaires, behavioral observation, machine learning algorithms, and projection techniques, we have gained insights into how personality develops and changes over time, how media preferences can inform personality prediction, and how different assessment tools can be used to measure personality traits. Our project has highlighted the importance of considering the strengths and limitations of each methodology, as well as ethical considerations such as informed consent and data privacy. Overall, our findings can contribute to a deeper understanding of human behavior and individual differences, and may have implications for fields such as psychology, education, and marketing.

7.2 FUTURE ENHANCEMENT

Here are some potential future enhancements for a personality prediction project:

Longitudinal studies: Conducting longitudinal studies that follow individuals over time can provide insights into how personality traits develop and change throughout the lifespan.

Cross-cultural studies: Comparing personality traits across different cultures can help to identify cultural differences in personality and inform more culturally sensitive personality assessment tools.

Integration of multiple data sources: Combining data from multiple sources, such as social media posts, behavioral observations, and self-report questionnaires, can provide a more comprehensive understanding of an individual's personality.

Personalized feedback: Providing individuals with personalized feedback based on their personality assessment results can help them to better understand their strengths and weaknesses, and make informed decisions about their personal and professional lives.

Integration with other fields: Integrating personality prediction with other fields, such as education or healthcare, can lead to more targeted interventions and better outcomes for individuals.

These future enhancements can help to improve the accuracy and usefulness of personality prediction tools, and contribute to a better understanding of human behavior and individual differences.

REFERENCES

Here are some references for a personality prediction project:

1. John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), Handbook of personality: Theory and research (2nd ed., pp. 102-138). Guilford Press.

2. Rentfrow, P. J., & Gosling, S. D. (2012). The do re mi's of everyday life: The structure and personality correlates of music preferences. Journal of Personality and Social Psychology, 100(6), 1139-1157.

3. McCrae, R. R., & Costa, P. T. (2003). Personality in adulthood: A five-factor theory perspective (2nd ed.). Guilford Press.

4. Butcher, J. N. (2003). The handbook of personality assessment. Wiley.

5. Corr, P. J., & Matthews, G. (Eds.). (2009). The Cambridge handbook of personality psychology. Cambridge University Press.

6. Mairesse, F., Walker, M. A., Mehl, M. R., & Moore, R. K. (2007). Using linguistic cues for the automatic recognition of personality in conversation and text. Journal of Artificial Intelligence Research, 30, 457-500.

7. Gosling, S. D., Ko, S. J., Mannarelli, T., & Morris, M. E. (2002). A room with a cue: Personality judgments based on offices and bedrooms. Journal of Personality and Social Psychology, 82(3), 379-398.

8. Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). The development and psychometric properties of LIWC2015. University of Texas at Austin, Austin, TX.