# PERSONALITY DETECTION - DATA PREPROCESSING (APPROACH)

Complete_colabnotebook_code-*Colab_notebook_link*
Dataset_link-*Essays*

**The Essay** dataset is chosen over the Facebook **My Personality** dataset for the research.

Because the **Essays** dataset contains highly meaningful content, in contrast to the **MyPersonality** dataset, which only includes status updates with little meaningful content.

Below is the list of personalities and their corresponding counts on the **ESSAY** dataset is presented.
Totally it contains 2467 data

Regarding **single personalities** following data points are presented
**neuroticism**          - 141
**openness**          - 59
**agreeableness**     - 53
**extraversion**       - 47
**conscientiousness** - 43
Total               = 343 datas

## All combinations of personalities and no of datas in the ESSAYS dataset

(extraversion, agreeableness, conscientiousness, openness)  - 189

(neuroticism,) - 141

(neuroticism, openness) - 134

(extraversion, agreeableness, conscientiousness) - 134

(extraversion, agreeableness, openness)-  90

(neuroticism, conscientiousness) - 89

(agreeableness, conscientiousness) - 89

(extraversion, neuroticism, openness) - 80

(neuroticism, agreeableness) - 78

(agreeableness, conscientiousness, openness) - 73

(agreeableness, openness) - 73

(extraversion, neuroticism, agreeableness, conscientiousness, openness) - 73

(neuroticism, agreeableness, conscientiousness) - 73

(extraversion, conscientiousness) - 72

(neuroticism, agreeableness, openness) - 72

(extraversion, neuroticism, agreeableness, conscientiousness) - 72

(extraversion, neuroticism, agreeableness, openness) - 71

(extraversion, agreeableness) - 71

(extraversion, conscientiousness, openness) - 71

(extraversion, neuroticism, conscientiousness, openness) - 69

(extraversion, neuroticism) - 68

(extraversion, openness) - 63

(neuroticism, conscientiousness, openness) - 62

() or no personality - 60

(openness,) - 59

(extraversion, neuroticism, agreeableness) - 54

(agreeableness,) - 53

(extraversion, neuroticism, conscientiousness) - 52

(extraversion,) - 47

(conscientiousness, openness) - 47

(neuroticism, agreeableness, conscientiousness, openness) - 45

(conscientiousness,) - 43

Above I grouped the top personalities and multi-personalities. And the Approach method is briefly stated below.

**Problem statement**
<mark>Is it possible to have a high score in more than one Big Five trait?</mark>

- The **Big Five traits are independent** of each other, which means they don't necessarily influence one another. This means that a person's score in one trait doesn't predict or determine their score in another. For instance, just because someone is **highly open to experiences doesn't mean they'll be extroverted or agreeable**. Each trait is its own distinct aspect of personality.

- Every person's personality is unique, and each characteristic is evaluated along a range or spectrum. This means that a **person's score for a certain attribute can range from very low to very high**. The set of **scores one will receive for each of the five qualities is not fixed.**
- Due to the independent nature of these traits, it's entirely possible for someone to score highly in multiple areas.

**Examples:-**
1. One could be both **highly extroverted** (meaning they're outgoing and gain energy from social interactions) and **highly conscientious** (indicating they're organized, responsible, and hardworking).

2. a person might be very open to new experiences (**high openness**) but might prefer solitary activities and introspection over social interactions (**low extraversion**)

# Research papers

These papers present research on how each of the **Big Five personality traits correlates with one another**

1. **'Oz, H. (2015). PERSONALITY TRAITS AND IDEAL L2 SELF AS PREDICTORS OF ACADEMIC ACHIEVEMENT AMONG PROSPECTIVE ENGLISH TEACHERS. *ResearchGate*. '(*Paper_link*)**

| | Variables | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| 1 | Extraversion | 1 | .332** | .129 | .049 | .270** |
| 2 | Agreeableness | .332** | 1 | .240* | .031 | .507** |
| 3 | Conscientiousness | .129 | .240* | 1 | .253* | .403** |
| 4 | Neuroticism | .049 | .031 | .253* | 1 | .074 |
| 5 | Openness | .270** | .507** | .403** | .074 | 1 |

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).

2. **Fralenko, V., Khachumov, V., & Khachumov, M. V. (2019). Correlation analysis and prediction of personality traits using graphic data collections. SHS Web of Conferences, 72, 01012. (*Paper_link*)**

| | O | C | E | A | N |
|---|---|---|---|---|---|
| O | 1 | 0.2198 | 0.1810 | 0.1540 | 0.1201 |
| C | 0.2198 | 1 | 0.1036 | **0.3212** | -0.0737 |
| E | 0.1810 | 0.1036 | 1 | 0.2125 | -0.0189 |
| A | 0.1540 | **0.3212** | 0.2125 | 1 | **-0.1730** |
| N | 0.1201 | -0.0737 | -0.0189 | **-0.1730** | 1 |

3. Emelyanova, Y., Talalaev, A., Fralenko, V., Khachumov, V., & Хачумов, М. В. (2019). ABOUT OF THE CORRELATION RELATIONS OF «BIG FIVE» PERSONALITY FACTORS. Научное Обозрение. Педагогические Науки.(*Paper link*) - The paper is in the Russian Language

|   | O | C | E | A | N |
|---|---|---|---|---|---|
| O | 1 | 0,25091 | −0,15924 | −0,09228 | 0,19647 |
| C | 0,25091 | 1 | −0,02079 | **0,45647** | −0,50726 |
| E | −0,15924 | −0,02079 | 1 | 0,31929 | −0,25016 |
| A | −0,09228 | **0,45647** | 0,31929 | 1 | **−0,76511** |
| N | 0,19647 | −0,50726 | −0,25016 | **−0,76511** | 1 |

The comparison in **paper 2** and **paper 3** looks relatively similar I choose the table from paper 2 over paper 3 for our research purpose. Because paper have more citation than paper.

I Generated a Top 10 correlation score among the Personalities from the Tables obtained from the above research paper.

**From paper 1.**
1. ('Agreeableness', 'Openness', 0.507)
2. ('Conscientiousness', 'Openness', 0.403)
3. ('Agreeableness', 'Conscientiousness', 'Openness', 0.3833333333333333)
4. ('Extraversion', 'Agreeableness', 'Openness', 0.36966666666666664)
5. ('Extraversion', 'Agreeableness', 0.332)
6. ('Extraversion', 'Agreeableness', 'Conscientiousness', 'Openness', 0.31350000000000006)
7. ('Extraversion', 'Openness', 0.27)
8. ('Extraversion', 'Conscientiousness', 'Openness', 0.26733333333333337)
9. ('Conscientiousness', 'Neuroticism', 0.253)
10. ('Agreeableness', 'Conscientiousness', 'Neuroticism', 'Openness', 0.25133333333333335)

**From paper2.**
('C', 'A', 0.3212)
('O', 'C', 'A', 0.23166666666666667)
('O', 'C', 0.2198)
('E', 'A', 0.2125)
('C', 'E', 'A', 0.21243333333333334)
('O', 'C', 'E', 'A', 0.19868333333333332)
('C', 'A', 'N', 0.1893)
('O', 'E', 'A', 0.1825)
('O', 'E', 0.181)
('O', 'C', 'A', 'N', 0.17696666666666663)