

---

# PROJECT REPORT

---

## **Personality Trait Detection**

Submitted to : Dr. Vikram Pudi

Submitted by: Aditi Shrivastava - 2018201056 Padma Dhar - 2018201011

# Contents

0.1	Problem Description . . . . .	2
0.2	Approach to Problem . . . . .	2
0.3	Project Goals . . . . .	3
0.4	Data Collection . . . . .	4
0.5	Data Processing . . . . .	4
0.6	Data Analysis and Visualization . . . . .	7
0.7	Observations and Conclusion . . . . .	20
0.8	Future Scope . . . . .	22
0.9	References . . . . .	23
0.10	Project Github Repository . . . . .	24

## 0.1 PROBLEM DESCRIPTION

*It is a belief that the alignment of the stars and planets, the date and time a person is born affects his/her personality. In this project, we attempted to verify if there is any connection between the date of birth and personality or not?*

## 0.2 APPROACH TO PROBLEM

*We approached the problem in the following manner :*

1. **Deciding the Goals** : We firstly decided about the goals of the project and tried to achieve those goals. Our goal was to collect the data from the college students and then study the different patterns in the data by various means of visualizing it.
2. **Collection Of Data**: We collected the data for our project by creating two Google Forms (one for IIIT-Hyderabad and second one for other college students all over India) and asked them to fill it honestly and then saved that data in the csv format.
3. **Data Processing and Calculation of BIG-5 Score using BFI-2S Questionnaire and Scoring Key**: The data collected has the following information: the answers to the **30 Questions** of the **BFI-2S Question set** and the various birth details such as : Date Of Birth, Zodiac Signs, Place Of Birth, Time of Birth and Gender information.  
**Based on the responses** of the questionnaire that a student gives, he is given a **score** for each of the BIG-Five domains, i.e. **Extraversion** (with facets of Sociability, Assertiveness, and Energy Level), **Agreeableness** (Compassion, Respectfulness, and Trust), **Conscientiousness** (Organization, Productiveness, and Responsibility), **Negative Emotionality** (Anxiety, Depression, and Emotional Volatility), and **Open-Mindedness** (Intellectual Curiosity, Aesthetic Sensitivity, and Creative Imagination).
4. **Analysis of Data and its Visualization**: We analyzed the data obtained by means of creating a Python Notebook and made different pie plots and bar graphs.

## 5. **Observations and Conclusion:**The main observations are

**A)** There have been cases when our generated graph agrees with the things being said by the different horoscope online predictors, as well as disobeys under certain parameters.

**B)** When we ignore values which are closer to the mean(i. e. the attribute doesn't define the major characteristics of a particular question. This is seen between second type of graph and the third type of graph.

Here the second type of graph refers to labelling all people having greater than or equal to the mean of dataset as high and all others as low(for each domain) and then plotting according to the zodiac signs.

The third type of graph considers high as all individuals having a score of greater than or equal to the mean+ std. deviation and all individuals having a score of less than mean-std.deviation as low. All individuals between these 2 ranges are ignored.

## 6. **Future Scope:**We will try to get more responses and try to analyse the BFI-2-S on the facet level(which are 15.i.e. Sociability, Assertiveness, and Energy Level,Compassion, Respectfulness, and Trust,Organization, Productiveness, and Responsibility,Anxiety, Depression, and Emotional Volatility,(Intellectual Curiosity, Aesthetic Sensitivity,and Creative Imagination) rather than on the Domain Level(Extraversion,Agreeableness, Conscientiousness,Negative Emotionality and Open-Mindedness)

## 0.3 **PROJECT GOALS**

- Our goal was to **Collect the Data** from the college students and then study the different patterns in the data by various means of visualizing it.
- The decision for **What Data to be Collected** played a major role in solving this problem.
- After collection of the data we did the **Data Processing** by the scoring keys of BFI-2-S and mapped with the Zodiac signs to see the interesting patterns.

- The processed data was **Analyzed** by finding the interesting patterns that showed the relationship among the zodiac signs and the Big-5 domains.
- The **Visual Analysis** of the patterns were done with the help of pie plots and bar graphs.
- Then we made the **Observations** and discussed about the **Future Scope** of the project.

## 0.4 DATA COLLECTION

The dataset consisted of 270 responses. These were collected from IIT H and other colleges. The link to these forms are:

1. **For IIT Students**

2. **For Other College Students**

## 0.5 DATA PROCESSING

Individual differences in people's characteristic patterns of thinking, feeling, and behaving can be organized in terms of the Big Five personality trait domains. Moreover, these five broad domains can be conceptualized hierarchically, with each domain subsuming several more-specific facet traits. We have used the BFI-2-S which is the short form of the Big Five Inventory - 2 which consisted of 30 questions.

The BFI-2 is a 60-item questionnaire that hierarchically assesses the Big Five personality domains and 15 more-specific facet traits. The present research developed and validated a 30-item short form (the BFI-2-S) and a 15-item extra-short form (the BFI-2-XS) of the BFI-2. These abbreviated forms offer savings in assessment time over the full BFI-2. They also retain most of the full measure's reliability and validity, especially at the domain level. The BFI-2-S and BFI-2-XS should therefore prove useful for assessing personality traits in research contexts where, due to pressing concerns about assessment time or respondent fatigue, administering the full BFI-2 would not be feasible.

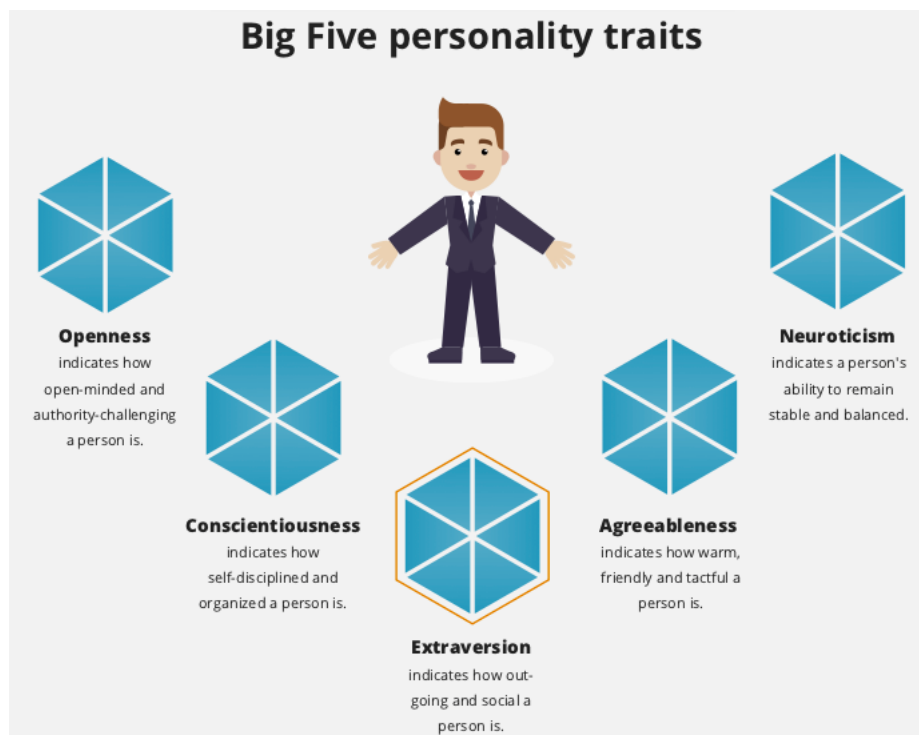
### The Big Five Inventory–2 Short Form (BFI-2-S)

Here are a number of characteristics that may or may not apply to you. For example, do you agree that you are someone who *likes to spend time with others*? Please write a number next to each statement to indicate the extent to which you agree or disagree with that statement.

1 Disagree strongly	2 Disagree a little	3 Neutral; no opinion	4 Agree a little	5 Agree strongly
<i>I am someone who...</i>				
1. ___ Tends to be quiet.				16. ___ Is outgoing, sociable.
2. ___ Is compassionate, has a soft heart.				17. ___ Can be cold and uncaring.
3. ___ Tends to be disorganized.				18. ___ Keeps things neat and tidy.
4. ___ Worries a lot.				19. ___ Is relaxed, handles stress well.
5. ___ Is fascinated by art, music, or literature.				20. ___ Has few artistic interests.
6. ___ Is dominant, acts as a leader.				21. ___ Prefers to have others take charge.
7. ___ Is sometimes rude to others.				22. ___ Is respectful, treats others with respect.
8. ___ Has difficulty getting started on tasks.				23. ___ Is persistent, works until the task is finished.
9. ___ Tends to feel depressed, blue.				24. ___ Feels secure, comfortable with self.
10. ___ Has little interest in abstract ideas.				25. ___ Is complex, a deep thinker.
11. ___ Is full of energy.				26. ___ Is less active than other people.
12. ___ Assumes the best about people.				27. ___ Tends to find fault with others.
13. ___ Is reliable, can always be counted on.				28. ___ Can be somewhat careless.
14. ___ Is emotionally stable, not easily upset.				29. ___ Is temperamental, gets emotional easily.
15. ___ Is original, comes up with new ideas.				30. ___ Has little creativity.

Please check: Did you write a number in front of each statement?  
BFI-2 items copyright 2015 by Oliver P. John and Christopher J. Soto.

The data collected has the following information: the answers to the **30 Questions** of the **BFI-2S Question set** and the various birth details such as : Date Of Birth, Zodiac Signs, Place Of Birth, Time of Birth and Gender information. **Based on the responses** of the questionnaire that a student gives, he is given a **score** for each of the BIG-Five domains, i.e. **Extraversion** (with facets of Sociability, Assertiveness, and Energy Level), **Agreeableness** (Compassion, Respectfulness, and Trust), **Conscientiousness** (Organization, Productiveness, and Responsibility), **Negative Emotionality** (Anxiety, Depression, and Emotional Volatility), and **Open-Mindedness** (Intellectual Curiosity, Aesthetic Sensitivity, and Creative Imagination).



Referenced from: <https://www.123test.com/personality-extraversion/>  
The Scoring is done based on the following scoring key:

#### Scoring Key

Item numbers for scoring the BFI-2-S domain and facet scales are listed below. Reverse-keyed items are denoted by "R." Due to the limited reliability of the two-item facet scales, we only recommend using them in samples with approximately 400 or more observations. For more information about the BFI-2, visit the Colby Personality Lab website (<http://www.colby.edu/psych/personality-lab/>).

#### Domain Scales

Extraversion: 1R, 6, 11, 16, 21R, 26R  
Agreeableness: 2, 7R, 12, 17R, 22, 27R  
Conscientiousness: 3R, 8R, 13, 18, 23, 28R  
Negative Emotionality: 4, 9, 14R, 19R, 24R, 29  
Open-Mindedness: 5, 10R, 15, 20R, 25, 30R

#### Facet Scales

Sociability: 1R, 16  
Assertiveness: 6, 21R  
Energy Level: 11, 26R  
Compassion: 2, 17R  
Respectfulness: 7R, 22  
Trust: 12, 27R  
Organization: 3R, 18  
Productiveness: 8R, 23  
Responsibility: 13, 28R  
Anxiety: 4, 19R  
Depression: 9, 24R  
Emotional Volatility: 14R, 29  
Aesthetic Sensitivity: 5, 20R  
Intellectual Curiosity: 10R, 25  
Creative Imagination: 15, 30R

#### Citations for the BFI-2 and BFI-2-S

Soto, C. J., & John, O. P. (2017). The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of Personality and Social Psychology*, 113, 117-143.

Soto, C. J., & John, O. P. (2017). Short and extra-short forms of the Big Five Inventory-2: The BFI-2-S and BFI-2-XS. *Journal of Research in Personality*, 68, 69-81.

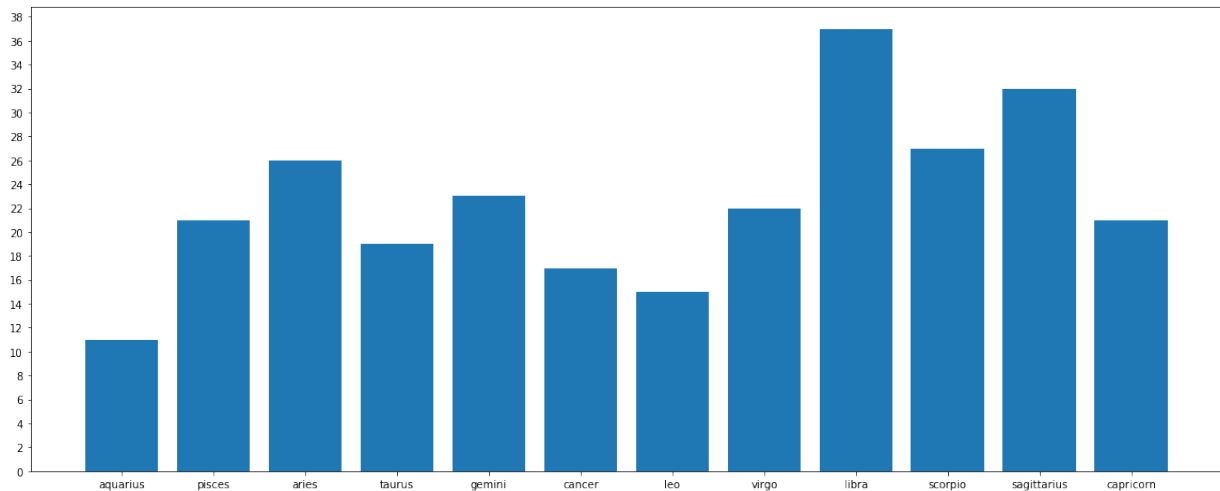
## o.6 DATA ANALYSIS AND VISUALIZATION

### Observations on Dataset:

A) The dataset was biased with a larger number of Libras(37 responses) in comparison to Aquarius(11 responses).

This can be attributed to the reasoning, that conception during cooler months is higher in humans while birth of children during cold months(like January) is less due to extreme cold conditions stillbirth rates are higher.

Further cooler months lie in the period around October - December. Thus children conceived during that period are likely to be born in September/October/November Period(Libra, Scorpio, Sagittarius). Please find figure below that shows the distribution of the dataset as per zodiac signs

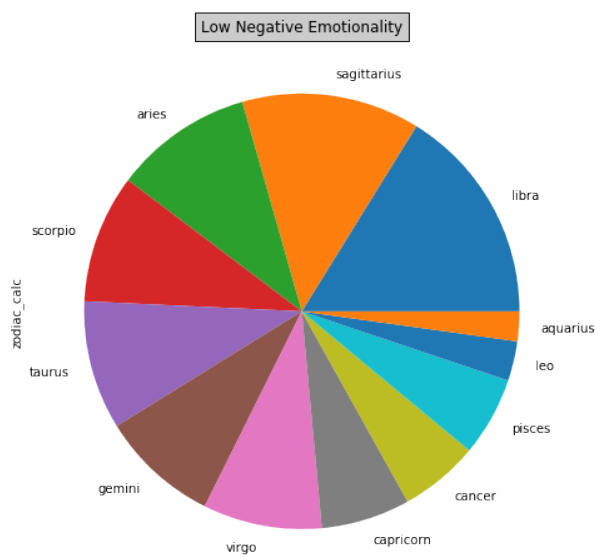


Date Range	Zodiac sign	Constellation
March 21 <sup>st</sup> – April 19 <sup>th</sup>	<b>Aries</b>	Aries
April 20 <sup>th</sup> – May 20 <sup>th</sup>	<b>Taurus</b>	Taurus
May 21 <sup>st</sup> – June 20 <sup>th</sup>	<b>Gemini</b>	Gemini
June 21 <sup>st</sup> – July 22 <sup>nd</sup>	<b>Cancer</b>	Cancer
July 23 <sup>rd</sup> – August 22 <sup>nd</sup>	<b>Leo</b>	Leo
August 23 <sup>rd</sup> – September 22 <sup>nd</sup>	<b>Virgo</b>	Virgo
September 23 <sup>rd</sup> – October 22 <sup>nd</sup>	<b>Libra</b>	Libra
October 23 <sup>rd</sup> – November 21 <sup>st</sup>	<b>Scorpio</b>	<u>Scorpius</u>
November 22 <sup>nd</sup> – December 21 <sup>st</sup>	<b>Sagittarius</b>	Sagittarius
December 22 <sup>nd</sup> – January 19 <sup>th</sup>	<b>Capricorn</b>	<u>Capricornus</u>
January 20 <sup>th</sup> – February 18 <sup>th</sup>	<b>Aquarius</b>	Aquarius
February 19 <sup>th</sup> – March 20 <sup>th</sup>	<b>Pisces</b>	Pisces

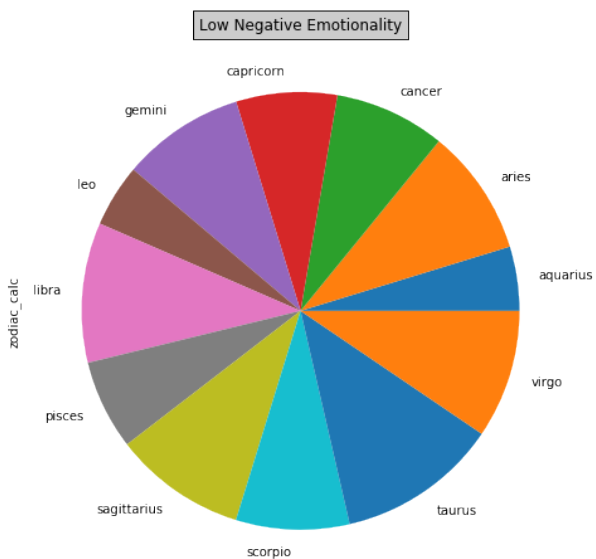


B) This biasing gave rise to a problem. A class as large as Libra was dominating in every scope while classes with lesser data was being left with a smaller share just because the number of people who had filled the form with that particular zodiac is lesser.

C) In order to handle this problem, we have used normalisation. For every domain, for every zodiac sign, we divide the number of people of that zodiac sign classified into the particular domain by the total number of responses of that zodiac sign.

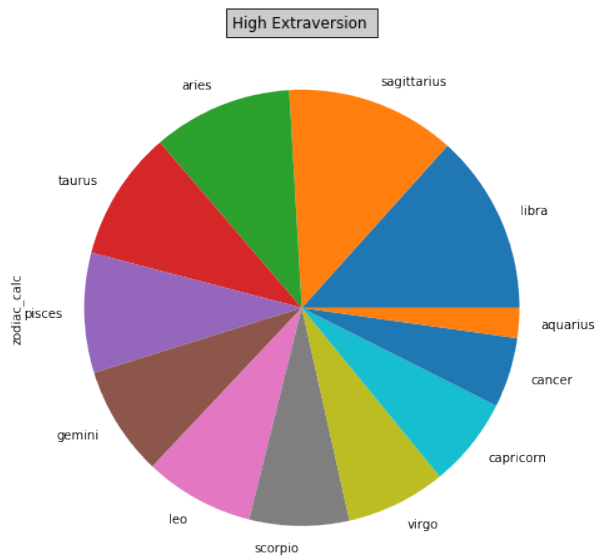


**Figure 1:** Before Normalisation

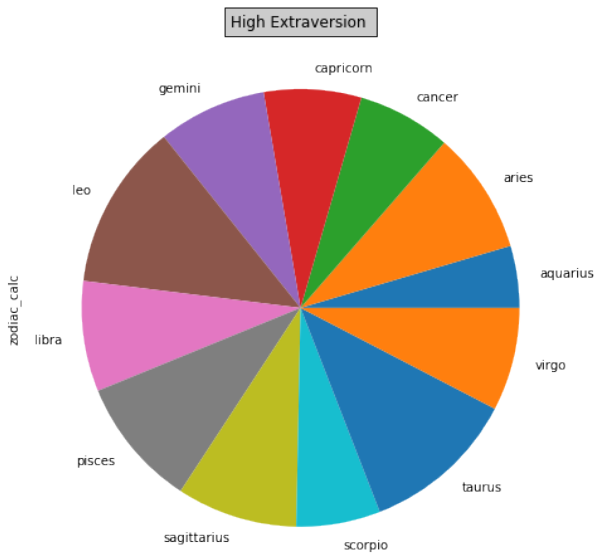


**Figure 2:** After Normalisation

In the above figure(low negative emotionality), we can see for an example, the older data distribution(the one on the left) was not normalised, which caused the graphs to tend towards the zodiac signs which had a larger number of respondents(in this case Libra and Sagittarius). On normalisation, we observe that Taurus which occupied a smaller chunk in the un-normalised data made a huge jump to obtain the largest chunk of the pie chart.



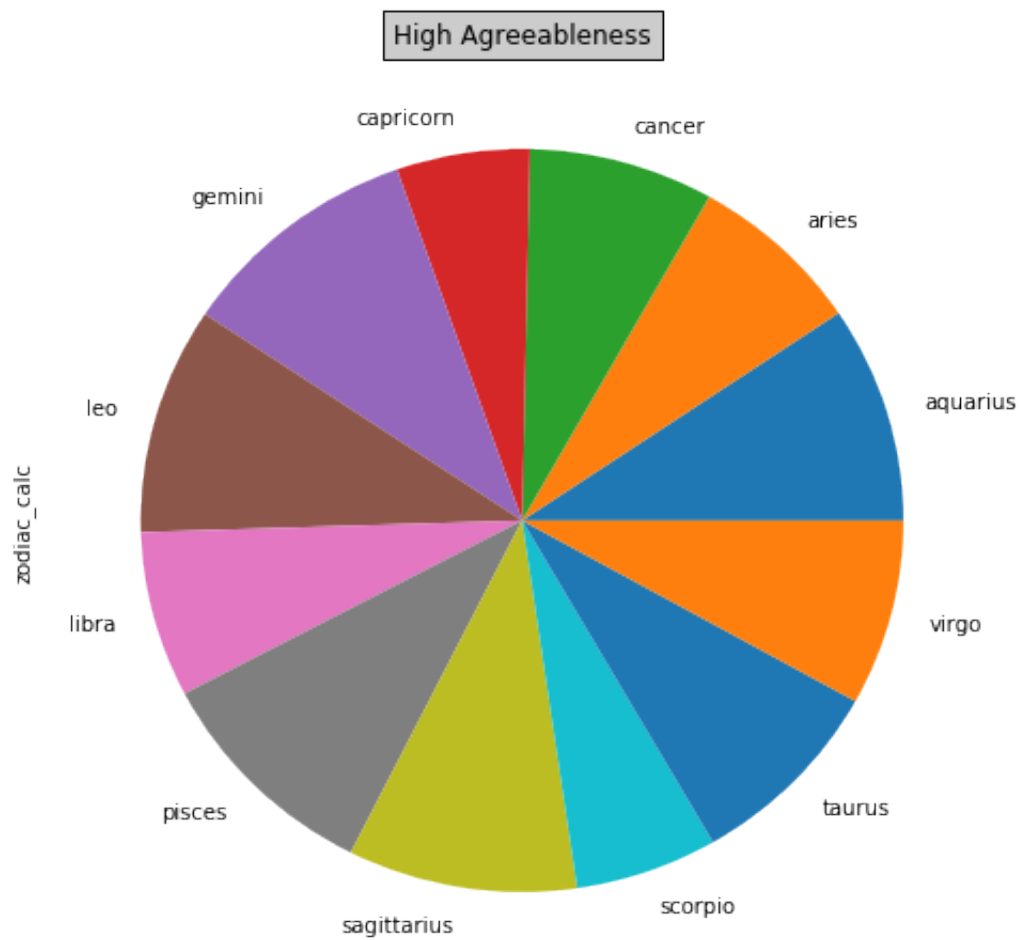
**Figure 3:** Before Normalisation



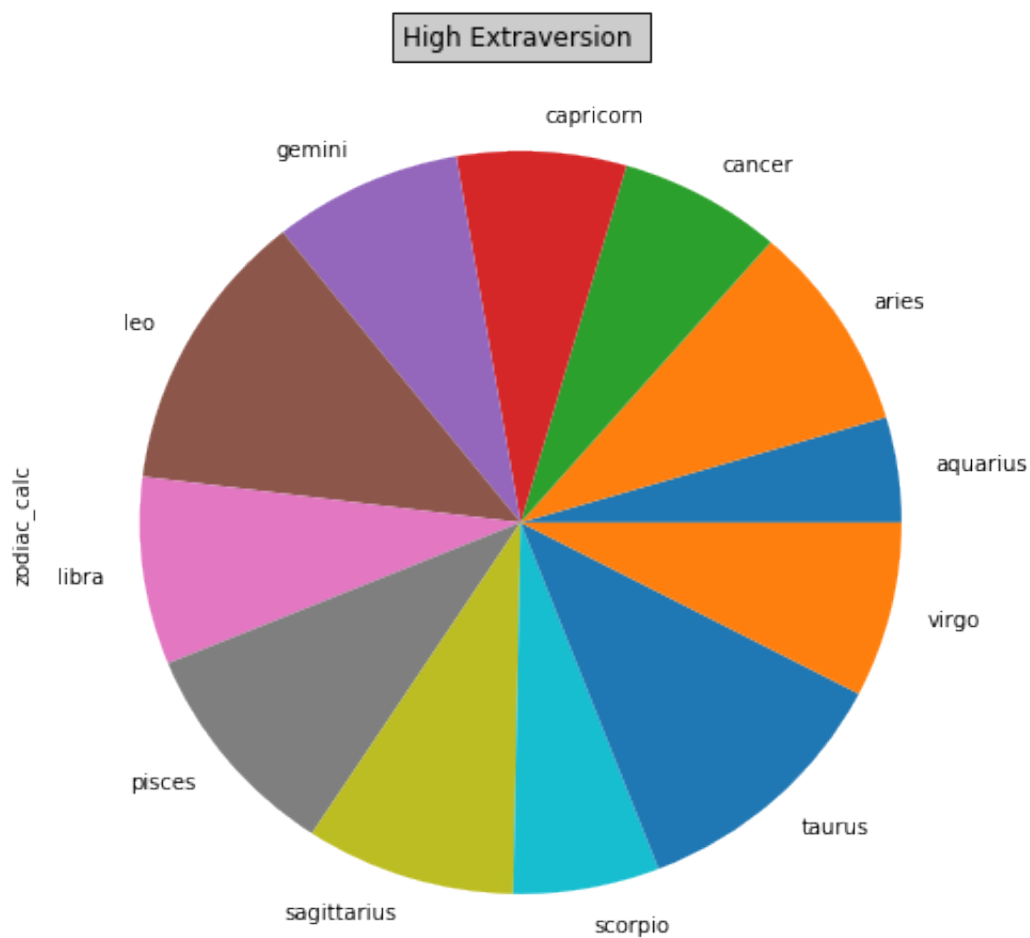
**Figure 4:** After Normalisation

In figure 3 and 4, that corresponds to the unnormalised high extroversion and normalised high extroversion, we observe a similar trend of zodiac signs with a larger chunk of responses(Sagittarius and Libra) forming a major share and losing out on it when the data is normalised. Smaller contributors like Leo and Taurus(Fig. 3) are now the major share of the pie chart.(Fig. 4)

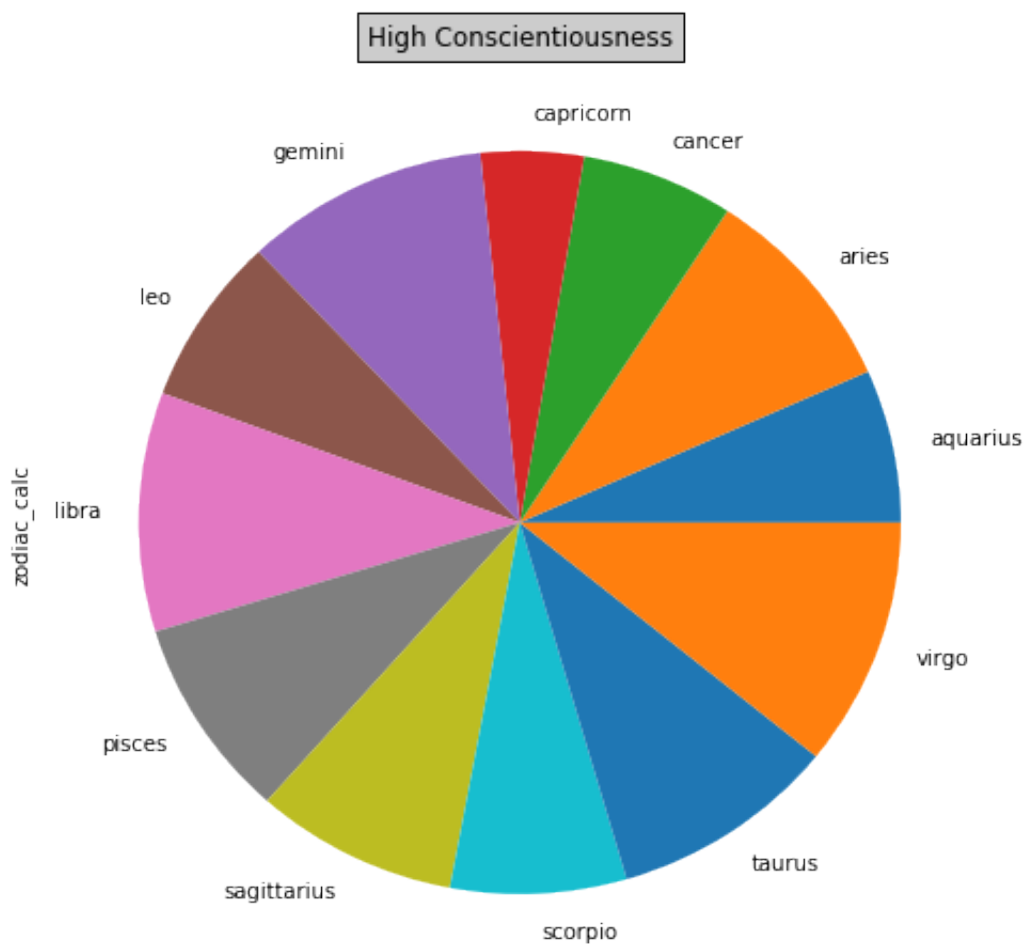
Coming to the visualisation of the data set and retrieval of useful information from it, firstly, we shall plot our domain as high and low for each domain with the threshold decided by the mean of the domain from the entire data set. Every person is classified into either the high of the domain or the low of the domain.



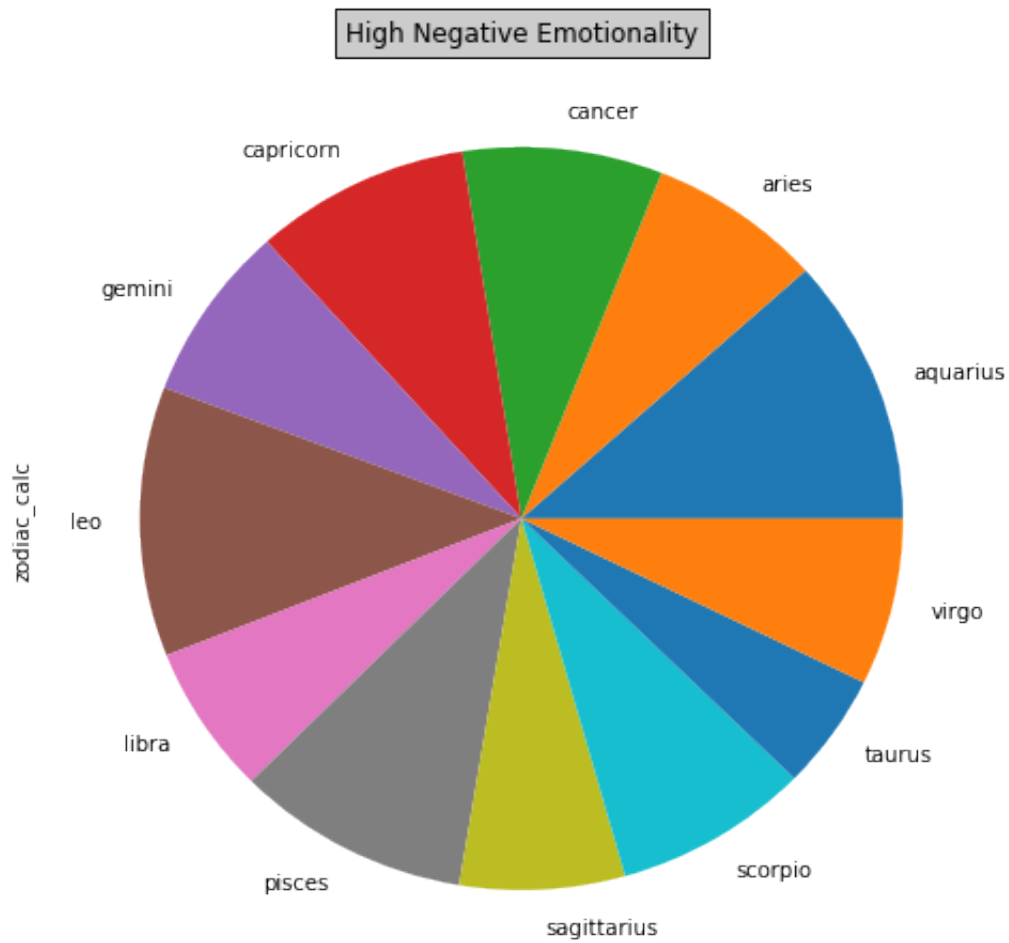
In the Above figure of High Agreeableness, We find that Pisces and Leo tie for the first place with 66.66 % of the respondents being classified as highly agreeable. Capricorns form the lowest chunk with a mere 38% being classified as highly agreeable.



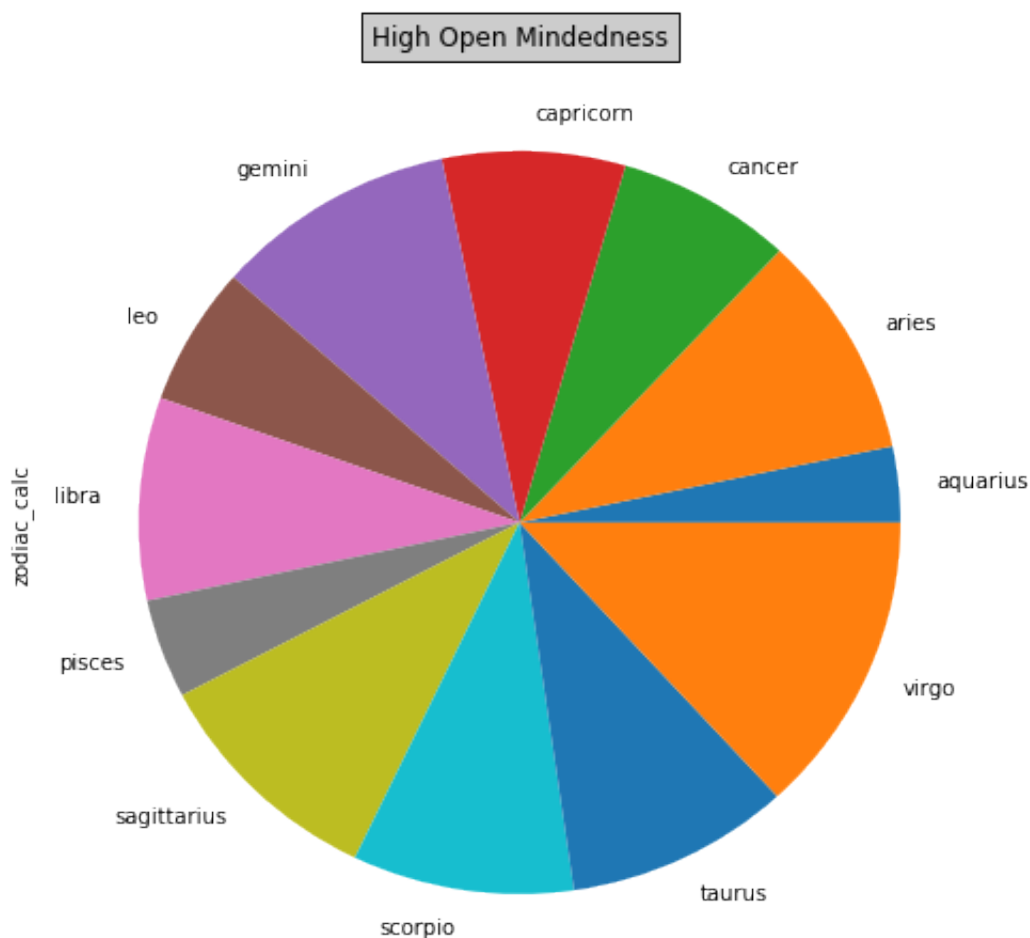
In the above figure of High Extroversion, Leos are highly extroverted with a 73.33 % of all Leos that gave the questionnaire being classified as Extroverted followed by Taurus at 68.42%. Aquarius with a 27.27 % formed the lowest category followed by Scorpio at 37%



In the domain of High Conscientiousness, it is observed that Virgos are highly conscientious with the majority chunk of 59.09% followed closely by Libra at 56.7% and Gemini at 56.5% Capricorns seem to have the least conscientiousness at a mere 23.8 %



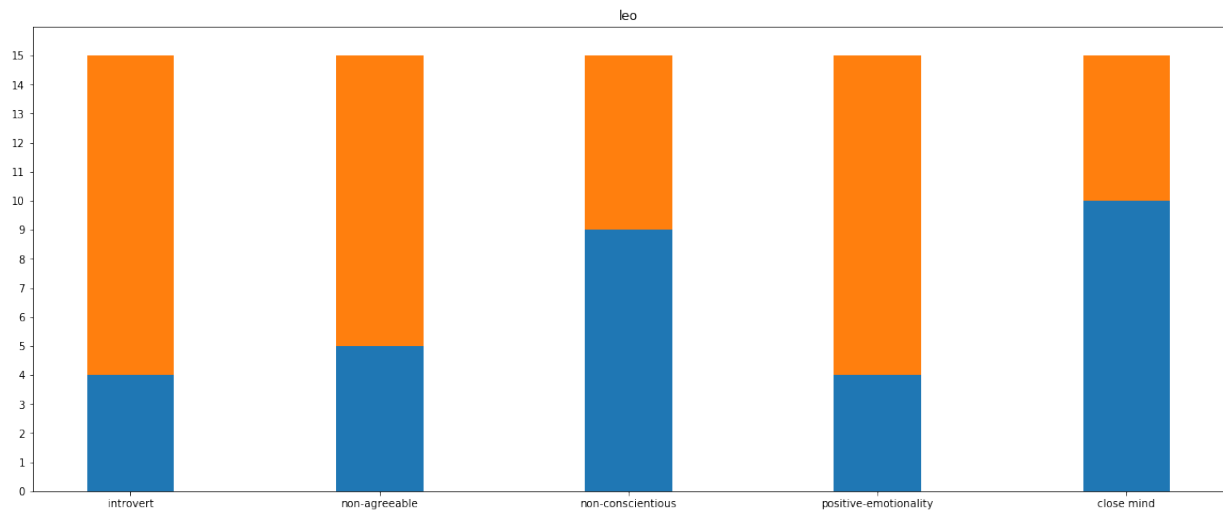
In the given domain of High Negative Emotionality, Leos and Aquarius with a percentage of 73.33 % and 72.7 % have the highest points in negative emotionality while Taurus and Libra have the least negative emotionality of 31.5% and 40.5%



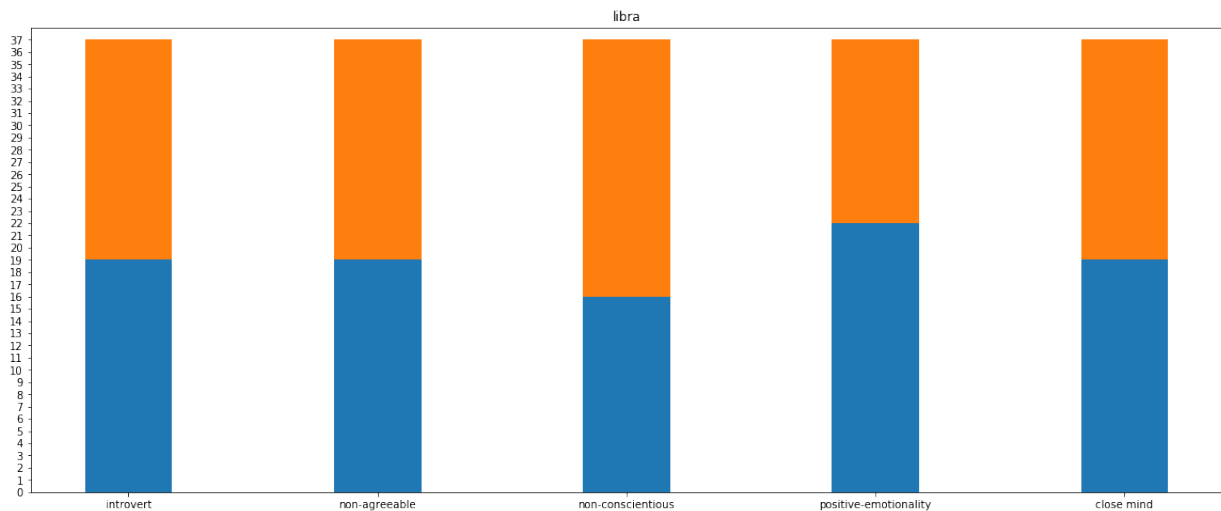
In the above pie chart, Virgos are extremely open minded with a 72.27% being classified as open minded. Aquarius and Pisces were a bit more conservative with 18.18% and 23.8% of all respondents of the zodiac sign classified as Open Minded.

In the second type of graph, we needed to for each zodiac sign find distribution of the people in the domains. Here as well, we have taken the mean of the data set as the threshold criteria. For each zodiac sign, we have taken all the entries of that zodiac sign and classified each person in either high or low of each domain. This count is then accumulated across all people of a particular zodiac sign and is plotted as a bar graph.

Thus, we can see the distinguishing characteristics of each zodiac sign. Here in the report, we are taking only 4 signs- Leo, Libra, Aquarius and Taurus. Rest can be found in the ipynb of the same. In order to understand the graph,

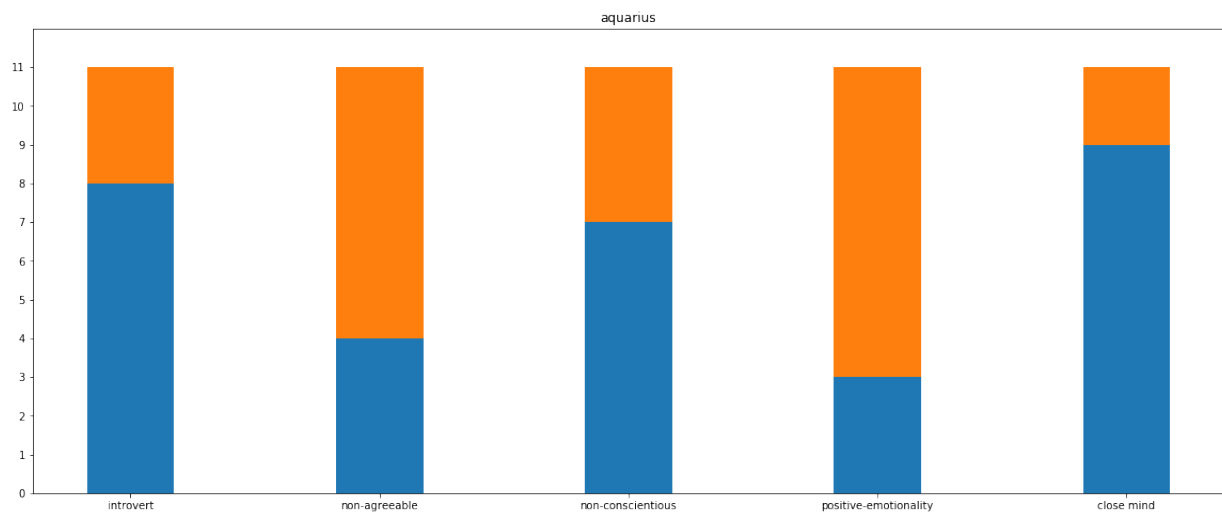


In the above figure of Leo, we can see that a large chunk of Leos are extroverted(11 of 15) , traditional(10 of 15 responses) and agreeable(10 of 15 responses) set of people.

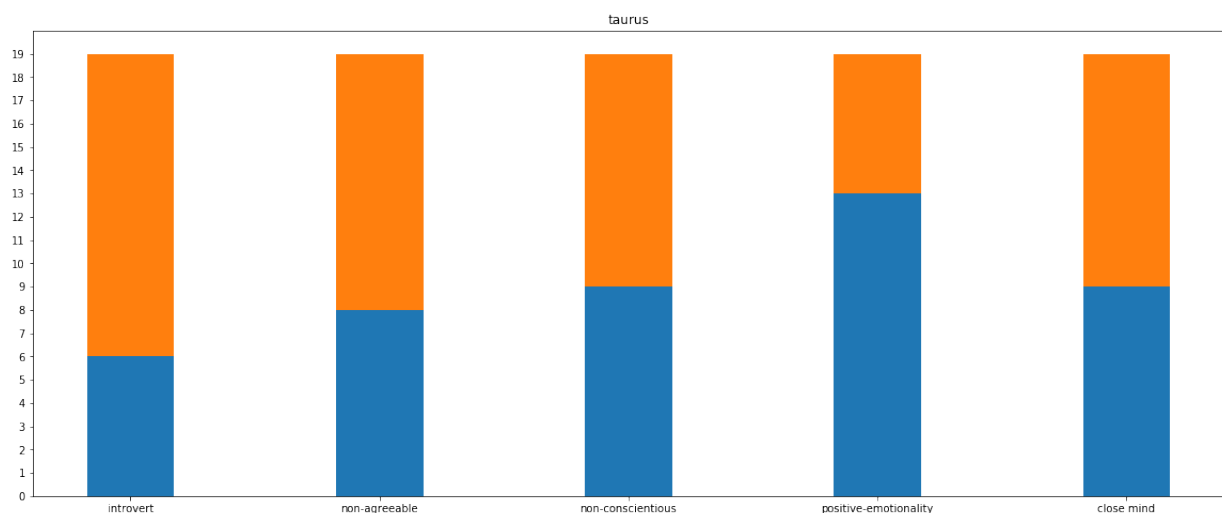


Libras seem to be higher in the positive emotionality domain. They seem to have a tendency towards conscientiousness. otherwise they are a relatively balanced set of people.





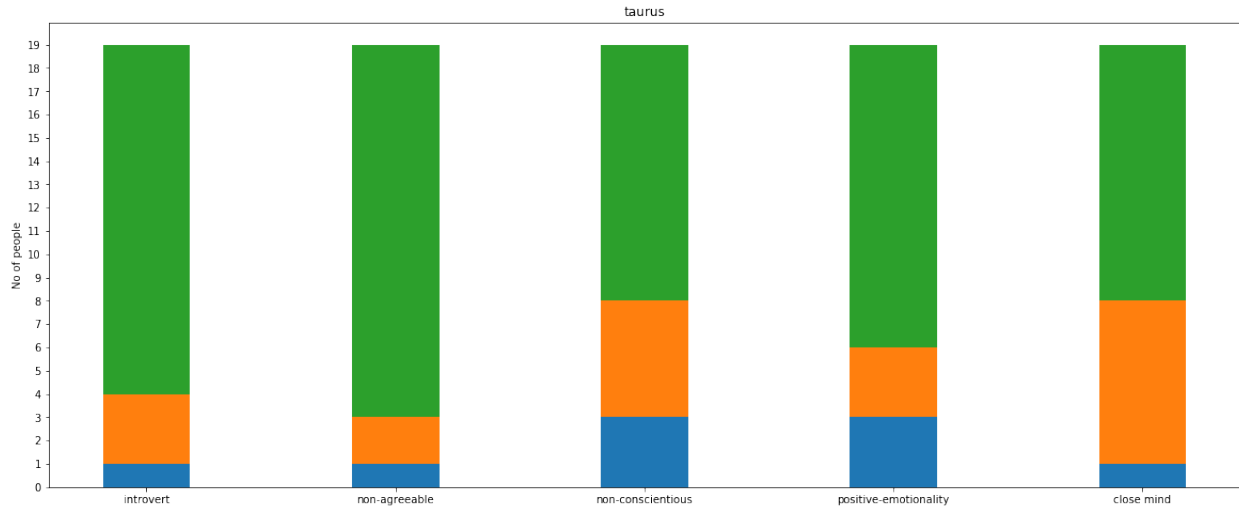
Aquarius have a dominant negative emotionality(8 of 11) and are traditionalists(9 of 11 responses) and introverted(8 of 11).



Taurus are a higher positive emotionality(13 of 19) and more of extroverted people(13 of 19 responses)

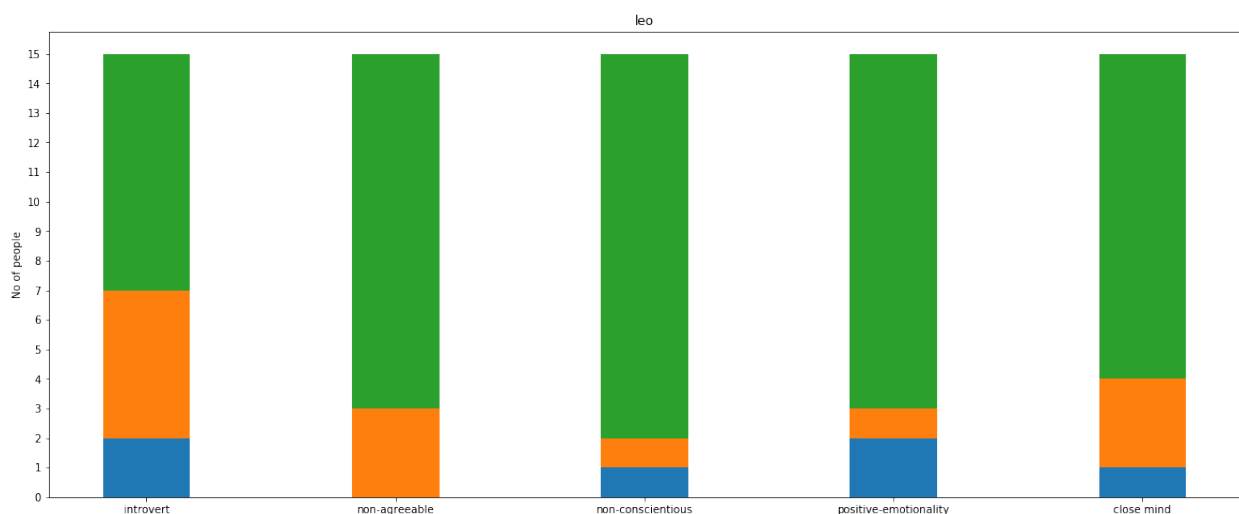
Now, in the next type of graphs, we shall ignore all values that are close to mean(that is people who are pretty much average in that domain and that domain does not form a defining facet of their personality. Thus, all values lying between mean + standard deviation and mean- standard deviation are ignored. For each domain, only values equal and above mean + std deviation are considered as a part of high for that domain and the values that are less than mean - standard deviation are considered as part of low for that domain.

This distribution skews the gap between the various domains for the different zodiac signs and different domains.



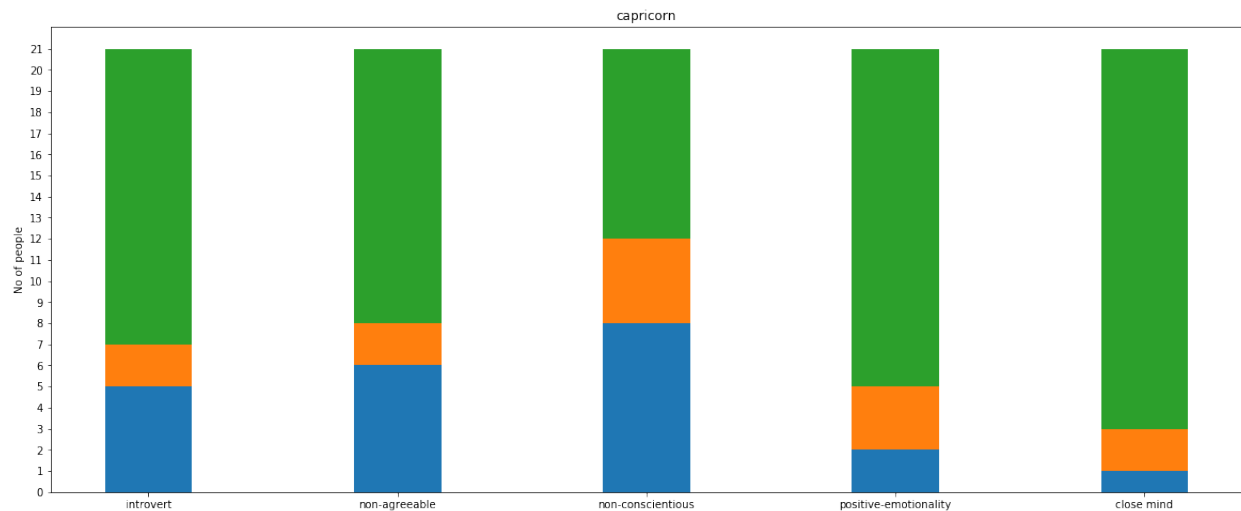
Here, blue colour indicates the fraction of the domain mentioned in the x-axis, orange indicates the fraction of the opposite domain to the domain listed in the x-axis and the green colour represents the fraction of people being discarded because they are out of range of the thresholds.

Taurus seem to be an extremely open minded set of people, who tend towards extroversion and are conscientious. Please note this is similar to what was predicted by the earlier graphs, however, here we can see a higher skew in favour of the prediction.

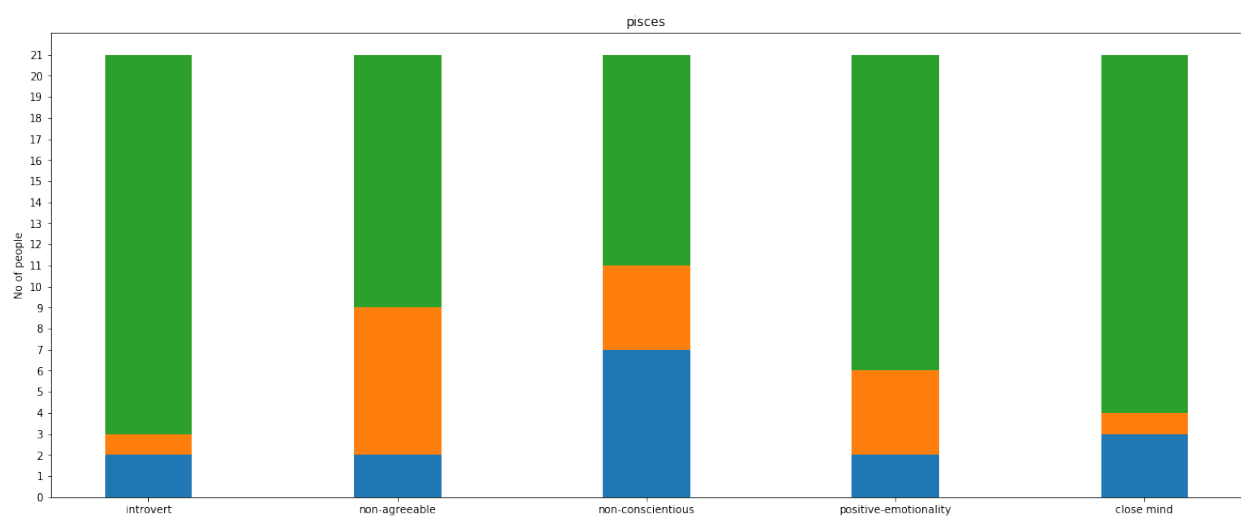


Leos are an extroverted set of people who are agreeable and open minded.

Here we are also seeing that positive emotionality is comparatively higher than negative emotionality. This was not true in the earlier graphs. That could have been because a lot of people were a slightly bit below the mean(which doesn't really define a person) and that had caused them to be classified as negative emotionality.

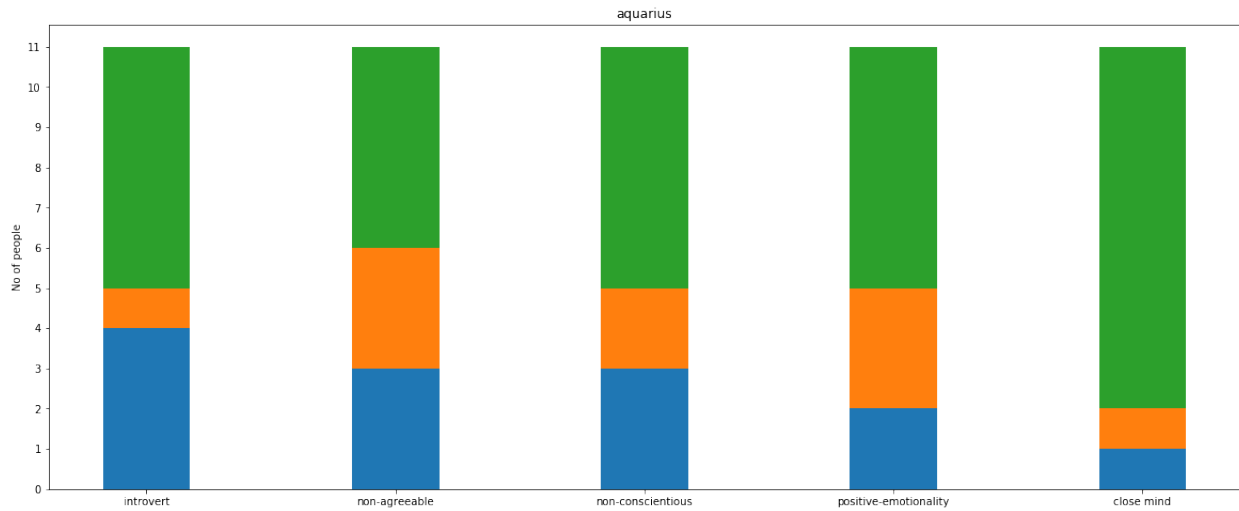


Capricorns have a huge tendency towards non conscientiousness, introversion and non agreeableness. These characteristics were existent in the earlier graphs as well but were accentuated in this.



As observed before, Pisces are agreeable though non conscientious people.

They have a tendency towards being traditionalists.



Aquarius are introverted bunch of people. Now they are being shown as equally agreeable and non agreeable. This was not in the earlier graphs. Though they still do consist of more people who are negative emotionality.

## 0.7 OBSERVATIONS AND CONCLUSION

We had generated 3 types of outputs

### 1. **Based on the normalised distribution of the zodiac signs for each domain(both high and low)**

Here, we had for each domain, found the mean over the entire dataset and using the mean as a threshold, classified each person as high or low in the domain(used the BFI-2S Score for both mean and individuals). Then for each domain, we accumulated the count of the high and the lows as a cluster of the zodiac signs i. e we plotted fraction of people classified in high over the entire number of people of that zodiac sign who were filling the questionnaire. Thus, we can say every person is classified into high or low of every domain.

### 2. **Based on the zodiac sign, the classification of each person into high and low of the domain using the mean value of the entire dataset as the cut off criteria.**

Here, we see the quantity of people classified into the high or low inn each zodiac sign. Thus we create 12 bar graphs( one for each zodiac sign) which consists of the division of people of that zodiac sign into the high and low. Thus, we are able to observe the general trend of the particular zodiac sign.

### 3. **Based on the zodiac sign, classification of each person into high and low of the domain using mean plus standard deviation of the domain and the mean minus the standard deviation of the domain.**

Here, some of the data(which was around the mean) was not considered since it was not an extremity. We performed the same as in 2, except the extrema values were considered( greater than mean+std.deviation as high and lesser than mean-std. deviation as low).

Based on the above three, we observed that:

For the First graph for the given domains :

#### 1. **Agreeableness:**

We find that Pisces and Leo tie for the first place with 66.66% of the respondents being classified as highly agreeable. Capricorns form the lowest chunk with a mere 38% being classified as highly agreeable.

#### 2. **Extroversion**

Leos are highly extroverted with a 73.33 % of all Leos that gave the question-

naire being classified as Extroverted followed by Taurus at 68.42%. Aquarius with a 27.27 % formed the lowest category followed by Scorpio at 37%.

### 3. **Conscientiousness**

Virgos are highly conscientious with the majority chunk of 59.09 followed closely by Libra at 56.7% and Gemini at 56.5.% Capricorns seem to have the least conscientiousness at a mere 23.8 %

### 4. **Negative Emotionality**

Leos and Aquarius with a percentage of 73.33 % and 72.7 % have the highest points in negative emotionality while Taurus and Libra have the least negative emotionality of 31.5% and 40.5%.

### 5. **Open Mindedness**

Virgos are extremely open minded with a 72.27% being classified as open minded. Aquarius and Pisces were a bit more conservative with 18.18% and 23.8% of all respondents of the zodiac sign classified as Open Minded.

**Second Type of Graph** We found a higher tendency of the given signs towards one or more of the domains which are as follows:

**Table 1:** Second Type of Graph

<b>Sun Sign</b>	<b>Characteristics</b>
Pisces	traditional and highly agreeable
Aries	balanced personality
Taurus	higher positive emotionality, introverted
Gemini	Agreeable, open minded
Libra	High Positive Emotionality
Sagittarius	Agreeable, open minded, extroverted
Virgo	Open Minded, Conscientious, Agreeable
Aquarius	Negative Emotionality
Scorpio	Introverts
Leo	Extroverted, Traditional, Agreeable, Negative Emotionality
Cancer	Non Conscientiousness
Capricorn	Non-conscientiousness, introverts, non-agreeable, negative emotionality

From the third type of Graph, We found a higher tendency of the given signs towards one or more of the domains which are as follows:

**Table 2:** Third Type of Graph

<b>Sun Sign</b>	<b>Characteristics</b>
Pisces	non-conscientious and highly agreeable
Aries	balanced personality with an incline towards positive emotionality
Taurus	open minded, conscientious
Gemini	Conscientious , Introverted
Libra	Conscientious, High Positive Emotionality
Sagittarius	Positive Emotionality
Virgo	Open Minded, Conscientious, Non-Agreeable
Aquarius	Introverted
Scorpio	Non conscientious, non agreeable
Leo	Extroverted, Traditional, Agreeable, Positive Emotionality
Cancer	Introverted, open minded, Conscientiousness
Capricorn	Non-conscientiousness, introverts, non-agreeable

So, summing up, we can conclude from our observations that :

1. If we take the intersection (**Strong Qualities**) of the qualities of the respective Zodiac signs from the Tables obtained from the above two graphs:

Pisces are highly agreeable, Aries have a balanced personality, Libra has High Positive Emotionality, Virgo are Open Minded and Conscientious, Leo are Extroverted, Traditional, Agreeable and Capricorn are introverts, non-agreeable and have Non-conscientiousness as a dominant feature.

2. But there were also some unexpected results which showed the opposite qualities by the above two different methods such as:

Virgo are Agreeable as observed in first type of graph while it showed Non-Agreeable in second type of graph. For Leo, the first graph showed Negative Emotionality while the second graph showed the Positive Emotionality. For Cancer, first one showed Non-Conscientiousness while the second one showed Conscientiousness. So there is a kind of disagreement between the two type of graphs. But we consider the second type of graph as the better predictor as it was calculated based under consideration of both mean and variance.

## **0.8 FUTURE SCOPE**

1. A big limitation that we face here was limited amount of responses. We had 270 responses only which was highly biased with as less as 11 Aquarius

answering the questionnaire. This may have affected the accuracy of the particular zodiac sign and also did not allow us to break down the domain into facets. As per the research paper read, we need 400+ responses to be able to break down into the various facets. Thus, the next step would be to increase the number of data collected.

**2.** The data collected was majorly from Mumbai, Bhopal and IIIT H. Since the region was not scattered, some bias may have been introduced into the data. We should try to get a more distributed version of the dataset.

**3.** A major chunk of the responses from other colleges were filled by females, while in IIIT, a major chunk was filled by males. This introduces some bias. We should try to get a more balanced dataset.

We will try to get more responses and try to analyse the BFI-2-S on the facet level (which are 15, i.e. Sociability, Assertiveness, and Energy Level, Compassion, Respectfulness, and Trust, Organization, Productiveness, and Responsibility, Anxiety, Depression, and Emotional Volatility, Intellectual Curiosity, Aesthetic Sensitivity, and Creative Imagination) rather than on the Domain Level (Extraversion, Agreeableness, Conscientiousness, Negative Emotionality and Open-Mindedness).

## 0.9 REFERENCES

**Short and extra-short forms of the Big Five Inventory-2: The BFI-2-S and BFI-2-XS :**

<http://www.colby.edu/psych/>

[wp-content/uploads/sites/50/2013/08/Soto\\_John\\_2017b.pdf](http://www.colby.edu/psych/wp-content/uploads/sites/50/2013/08/Soto_John_2017b.pdf)

<https://www.123test.com/personality-extraversion/>

**For the variation in birth months and common birth months :**

<https://www.ncbi.nlm.nih.gov/pubmed/721846>

<https://indianpediatrics.net/march2000/march-306-312.htm>



## **0.10 PROJECT GITHUB REPOSITORY**

<https://github.com/git-aditi/PERSONALITY-TRAIT-DETECTION>