

BIG DATA AND STRATEGIC MARKETING

EQUALS 3 COURSE PROJECT



Team - 1

Debaleena Saha debaleena.saha@uconn.edu

Kunja Dutta kunja.dutta@uconn.edu

Mayank Gupta mayank.gupta@uconn.edu

Samarpita Chakrabarty samarpita.chakrabarty@uconn.edu

Table of Contents

1. Business Challenge.....	2
2. Modeling.....	3
2.1 Data Set	3
2.2 Data Pre-Modeling.....	3
2.3 Data Extraction	4
2.4 Assumption	5
2.5 Data Dictionary Generation	7
2.6 Identifying Archetypes	7
2.7 Testing	8
3. Marketing Component.....	9
4. Executive Summary and Conclusion.....	13
5. Recommendations and Future Scope.....	14
6. Reference.....	15
7. Appendix.....	15
7.1 R Code:.....	15
7.2 SPSS Models.....	19

1. Business Challenge

Words have power. A clear, consistent message can be the make or break for a successful marketing campaign or an utter waste of time and money. In order to establish an effective marketing communication with the target audience, it is important to convey the intended message in a language that would resonate with their own. This communication can be in the form of advertisement, text message, social media, or other methods. The target audience has different characteristics, different diction, and different ways to perceive. Hence, accomplishing this is a real challenge since it is not a one size fits all situation here.

It is of utmost importance to lay down a methodology with the help of which the content of the marketing communication can be made more effective. This effectiveness can only be increased by increasing the relevance of the communication message with the interests of the target segment. The message needs to be as appropriate as possible for the market segment's personality type. Establishment of such communication is required for the company to be able to persuade the market.

Identifying personality traits of each market segment is needed for achieving message congruence. Communication goal can only be achieved by the successful identification of the words, that are used by each archetype, and laying down a methodology to communicate with a person belonging to that archetype, in a marketing language that is of most relevance to them. This relevance will make it quicker, better and easier for the target to acknowledge and appreciate the message.

2. Modeling

2.1 Data Set

Text files for 52 personality traits were provided. These 52 traits included 35 Big Fives, 12 Needs and 5 Values. Big Five reflects how a person generally interacts with the world such as cautiousness, anxiety, modesty, etc. Needs reflect an important aspect of human behavior such as love, liberty, challenge, etc. Values reflect what is most important to an individual such as hedonism, conservation, etc. These text files had comments picked up from various sites such as Twitter, Instagram. This data was unstructured and contained some alphanumerics, URLs, bad words and special characters.

Equals 3 also provided information about the 12 Jungian Archetypes. Examples of typical traits for each archetype has also been provided. The 12 identified archetypes are the Innocent, the Hero, the Regular Guy, the Citizen, the Nurturer, the Creator, the Explorer, the Rebel, the Lover, the Magician, the Ruler, the Jester and, the Sage.

2.2 Data Pre-Modeling

Since the data provided was unstructured, it was necessary to structure the data to work with it.

The following steps were performed to prepare a structured dataset:

Removal of URLs: URLs of different websites like Twitter and Instagram, containing “http”, “https”, “.com” etc. were identified as unimportant and were removed first.

Removal of special characters: Characters such as !, @, #, \$ were not of any use to the dataset.

Hence, they were removed.

Removal of single and double character words: Single character words such as w, t, etc. and double character words such as so, no, etc. were removed from being a part of the dataset.

Removal of numbers: Some alphanumeric words which were of no importance to the dataset were also removed.

Removal of English stop words: All stop words in the English language, such as a, an, the, etc. were removed with the help of a common stop-word list identified by R programming language.

Removal of bad words: All the bad words such as abuse, killing, etc. are not relevant to the dataset, were removed. The complete list of these bad words is provided in the appendix.

Once the data was free from the above-mentioned words/characters, it was ready to be utilized for the generation of the data dictionary.

2.3 Data Extraction

Since the dataset was huge, a random sample of 10,000 sentences was extracted from each personality trait text file to build the list of words. It was decided to extract the adjectives from each sentence of the trait texts, as adjectives were found the best way to identify the target's personality type.

As the first step of this data extraction process, a function called "tagPOS" was created to tag the parts of speech(POS) of each sentence. The objective of this function was to separate the meaningful words from each sentence. The final aim of this extraction was to have adjectives for

each trait ready in their respective trait text files.

The “tagPOS” function tokenizes the sentences which means that it converts the sentences into a bag of words. Each term in this bag of words was associated with a part of speech tag (POS Tag). These tags are such as JJ for the adjective, RP for the particle, RB for an adverb. The complete list is available in the appendix.

Each term was taken out and from their tag, adjectives were identified by using a regular expression. These adjectives were converted to lowercase to maintain consistency throughout the process. Next, they were sorted in decreasing order of their frequency of occurrence and unique adjectives were selected. Based on the decreasing frequency of the words, top 10% was extracted. These words were written in a text file. This text file was a bag of words for a particular trait. Finally, 52 such text files with relevant bag of words were generated and ready for data dictionary generation by mapping each archetype to the traits.

2.4 Assumption

Before proceeding for data dictionary generation, it was required to identify the mapping of each Jungian archetype to the personality traits. It has been assumed that there exists a one-to-one correspondence between the 12 Jungian archetypes and the 12 Users created by Equals 3 (Zoe, Michael, Matt, Steven, Ann, Bonnie, Richard, Heather, Ryan, Robert, Gwen, Paul). Based on that assumption, the seven traits (including Big 5, Needs and Values) identified by Equals 3 for each User Type have been mapped to each of the Jungian Archetype for generating the data dictionary for the corresponding Archetype.

The Innocent: Dutifulness, friendliness, morality, sympathy, trust, ideal, self-transcendence

The Hero: Self-enhancement, challenge, self-discipline, excitement-seeking, assertiveness, activity level, achievement striving

Regular Guy: Cheerfulness, cooperation, friendliness, modesty, trust, self-expression, hedonism

The Nurturer: Self-transcendence, love, vulnerability, trust, dutifulness, cheerfulness, altruism

The Creator: Artistic interests, emotionality, imagination, immoderation, vulnerability, self-expression, self-enhancement

The Explorer: Hedonism, excitement, immoderation, self-excitement, assertiveness, adventurousness, activity level

The Rebel: Adventurousness, anger, emotionality, immoderation, liberalism, liberty, self-enhancement

The Lover: Self-enhancement, love, sympathy, self-consciousness, friendliness, emotionality, cooperation

The Magician: Achievement striving, activity level, adventurousness, imagination, intellect, curiosity, openness to change

The Ruler: Hedonism, conservation, self-discipline, orderliness, morality, assertiveness, anger

The Jester: Self-transcendence, harmony, immoderation, gregariousness, friendliness, altruism, artistic interest

The Sage: Hedonism, conservation, self-discipline, morality, dutifulness, assertiveness,

achievement striving

These traits of each Jungian Archetype were utilized to generate data dictionaries of each archetype.

2.5 Data Dictionary Generation

Once the text files with adjectives for each trait was created, it was utilized by assigning above assumed traits provided by Equals 3 to a particular Jungian Archetype. The assumption matrix was provided in the Appendix section of this document.

For data dictionary generation, IBM SPSS Modeler is used. The modeler merged the text files, containing the bag of words of each trait relevant to a particular Jungian Archetype, to form the data dictionary for that Archetype. This way 12 data dictionaries, one for each Jungian archetype, was generated following the assumptions made.

As per the assumption, each Jungian Archetype was mapped to 7 different traits. After merging all the trait text files to each of the archetypes, it was filtered to remove unwanted fields keeping only distinct words in the bag of words.

The data dictionaries for each of the 12 archetypes were ready for predicting individual personality types.

2.6 Identifying Archetypes

The next part of the model was used to predict personality types of individuals from an input text file. This was generated using R programming. A text file containing minimum 6500 words was

needed to give as input to predict the individual personalities. The input file was again pre-processed following the steps of data cleaning mentioned above, to obtain a structured dataset.

This structured data was then tokenized in the same fashion by a function “tagPOS” as described in the previous sections. Finally, the adjectives were extracted from the input text file.

Next, this bag of words is checked against the data dictionary for each archetype and decided which archetype did the sentence correspond to the most. A term document matrix, containing the similarity value for each of the traits with the input text, was generated using R. The highest similarity value decided the personality type of the input text. The similarity values between the archetypes and the text input was calculated using Cosine similarity distance function under Latent semantic analysis(LSA) package in R. The latent semantic analysis (LSA) is a package which considers the fact that text have a higher order (=latent semantic) structure which is obscured by word usage (e.g. through the use of synonyms or polysemy). By using conceptual indices that are derived statistically via a truncated singular value decomposition (a two-mode factor analysis) over a given document-term matrix.

2.7 Testing

A text file whose archetype was pre-identified has been given as an input to the above-described model. The model has applied all the techniques to process the unstructured data into a structured form of a dataset. Next, it has extracted the bag-of-words (adjectives in this case) from that structured data. Finally, the semantic analysis has been done on the bag-of-words from the input text against the data dictionaries each of the archetypes. This semantic analysis generates

similarity scores for each archetype with the input text file. The highest similarity score decided the archetype of the text file which is very close to the pre-defined archetype.

3. Marketing Component

According to a recent report, 52% of marketers find producing the kind of content that engages the target audience as their biggest challenge. The goal of effective marketing is to ensure that each customer receives a consistent story in accordance to their personality. The communication should effectively be aligned with the interests of the target audience. It should reflect the same values, interests or products that the customers seek. In one word, the message should resonate with the same core needs or desires of its target audience. If the message isn't targeted to a specific audience, it means that the company is "spraying and praying" where the message is thrown out to everyone, like a fishing net hoping that some fish would get caught in the net. This technique would only work if you possess an unlimited budget. Thus, the importance of a targeted ad campaign which helps us maximize the budget and create a loyal set of customers who can relate to our brand is huge.

The data dictionaries or the bag-of-words generated by this project for each archetype will help to identify the top words that resonate with the audience's personality types. These identified words can then be used to design the marketing campaign which can engage the target segment accordingly.

The following steps have been identified, which is believed to help Equals 3 launch a successful marketing communications campaign.

Define Marketing Goals

The first step is to define what the goal of their marketing component is. Is this about promoting a high technology product? or is to create brand consciousness among customers? The strategies for both would be different in terms of budget and the scale of implementation. Once this goal has been defined, targeted marketing becomes more effective. It's easier to come up with messaging if you know

Create Personas

The first step to creating content that resonates with readers is getting a better understanding of who the readers are, especially the ones you want to convert into leads and customers. A major portion of this project is to create dictionaries of words which echo closely with the 12 archetypes that have been defined in the beginning of this report.

It is recommended that the personas be differentiated into 2 categories: Buyer and Reader. While companies may probably care most about appealing to the readers who will convert into leads and customers, there's also a segment of readers who will probably never buy from their business but will continue to read and share the content. In an ideal world, companies should understand and appeal to both segments.

Also during the marketing campaign, "lookalike audiences" will get attracted. This audience comprises of new people who are interested in the product/service as they are similar to the existing customers. The cost of acquiring a new customer is high and hence it would be beneficial to understand these lookalikes to see if they fit into the existing persona of our buyer or would the personas need to be altered to incorporate this group.

Delivering Content in The Right Format

Once the definitions for the personas has been finalized, the next step in this regard would be to choose what content format type would be the best way to get the message across. In what content type, will the target audience be most likely to consume the intended message?

There can be several formats that can be used i.e., messages on social media, e-books, blogs, videos, graphs etc. Whether it's deciding to create an e-book instead of a blog post, a video instead of a manual, the focus should be on determining what content format will get the message across in 1) the clearest way, and 2) the way audience will most enjoy reading it.

Using the Right Tone

This is probably the most important element of creating impactful marketing content. It takes not only experimentation to get it just right, but also an in-depth knowledge of your target audience's personas. This includes things like:

- Sentence structure
- Sense of humor
- Word choice
- Punctuation usage
- Content formatting

If an improper message is delivered to the target audience, it will likely be dismissed. For example, a customer who is risk averse will unlikely respond to a message that indicates anything to do with risk. Hence, to strike a chord with this customer, it is essential to propagate the low risks involved in the product (for example a highly trusted mutual fund) which will resonate with

the customer and produce a response.

As another example, REI has done a stellar job appealing to their target audience with their “Find Out” campaign, aligning to the Explorer archetype. The messaging and visuals appeal to the outdoor enthusiast through independence and fulfillment.

Similarly, the top keywords need to be identified from the bag of words that are generated as a result of this project and the campaign should be designed in a way which resonates closely with their targeted segments.

Start Small

The first full strategic messaging map is likely to be far from perfect. It is always good to start small, rather than starting a full-fledged campaign from the beginning itself. It gives time to gather sufficient, high-quality feedback from the stakeholders.

The wider the range of ideas that can be collected, the better and more thoughtful the strategic messaging will become. Not every idea can be satisfied by a strategic messaging map, but analyzing and evaluating as many different perspectives as possible, will definitely be beneficial. Different metrics need to be set to test the effectiveness of the campaign and then the components need to be scaled up.

To understand the success following three steps need to be followed: Review > Get Feedback > Iterate.

Track everything, do more of what works and fix what doesn't. Maybe a specific persona isn't engaging as much as it should. Try a different tactic or channel to reach them. If that still doesn't work, there's a chance that's not the right persona.

Retargeting is critical here to ensure that the brand stays in the minds of the customers. One way to do that is to create different ads that are similar but evoke different character traits within the same archetype. This way we can ensure that all customers that belong to an archetype have been targeted.

In conclusion, the success of a marketing campaign lies in understanding the minds of the customers and to position ourselves such that they can strike a chord with us. By understanding the target customer and knowing the language that he speaks can help create profitable and successful marketing campaign.

4. Executive Summary and Conclusion

An effective marketing communication depends on the relevance and relatability of the content of the communication to its target audience. The ability of a company to persuade a market segment requires the proper identification of words which would resonate with the words of the target audience. In this project, the message resonance model is built with the objective of identifying such words which can be used to establish an effective marketing communication with the target audience.

The model generates a data dictionary for each of the 12 archetypes, who are representatives of different market segments, from the 52 characteristic traits provided. The key traits for each archetype have been identified and semantic content from the traits have been extracted to build the vocabulary for each archetype. At the end, the model takes a text file as input to find out the percentage of similarity for the language used in the text with each of the archetypes. The one

with the highest percentage will be identified as the archetype that the author of the text belongs to. This will aid a company in establishing effective marketing campaigns by communicating with its prospective customer in a language that resonates with their own language.

The model developed is simple and straightforward but at the same time highly accurate. It has been achieved through a thorough data cleaning process which left only meaningful words, easy to work with and then selecting the highest percentage of the most frequently used words (adjectives) among people displaying a particular personality trait. Using this model, the effectiveness of a marketing communication can be understood easily. Subsequently, words from the list intended for that particular archetype can be used to communicate with the person.

5. Recommendations and Future Scope

Certain enhancements have been identified which can intensify the impact of the message resonance model. These can be implemented in future, to increase the accuracy of the model. The top three key recommendations which if implemented in the future will give more precise results would be

Increasing the sample size: In this project, a random sample size of 10,000 sentences have been used for each personality trait. If a larger sample can be used while building the model, the resultant model will definitely become more robust.

Taking phrases into account: In this project, only one-word adjectives relevant to a particular Jungian Archetype has been laid down in a data dictionary. Including associated phrases, bigrams, n-grams in the data dictionary for the archetypes will strengthen the model further.

Better segregation of Traits: The initial identification of traits from the users' comments which have been used for the purpose of building the data dictionary, had been done crudely, resulting in the traits being highly concordant with each other. If the segregation of the traits is done more efficiently, it would result in a much more effective model.

Better technique for validation: The method of validation used in this project can be improved to a much greater extent by using many more sophisticated tools and techniques for validation.

6. Reference

<https://courses.lumenlearning.com/marketing-spring2016/chapter/reading-defining-the-message/>

<https://blog.hubspot.com/blog/tabid/6307/bid/33747/How-to-Create-Content-That-Actually-Resonates-With-Your-Readers.aspx#sm.00001xav6xuzj3dz8w535hzn89bpy>

<http://buzzbee.biz/blog/2015/03/relevant-messaging-through-archetypes>

<https://medium.com/startup-grind/strategic-communication-how-to-develop-strategic-messaging-and-positioning-3cc59689ca28>

<https://cran.r-project.org/web/packages/lsa/lsa.pdf>

<http://www.thefactmachine.com/cosine-similarity/>

<https://artax.karlin.mff.cuni.cz/r-help/library/lsa/html/cosine.html>

7. Appendix

7.1 R Code:

Generating the Data Dictionary

Loading the required R libraries:


```
install_github("trinker/qdapRegex")
install.packages("lsa")
install.packages("SnowballC")
install.packages("tm")
install.packages("openNLP")
library(lsa)
library(openNLP)
library(NLP)
library(devtools)
library(qdapRegex)
library(stringr)
library(tm)
```

Importing the Text File (containing one of the 52 Traits) and taking a sample of 10000 records.

```
char_vector <- readLines("F:/UConn/Spring/Big Data Strategic Mktg/
Project/Data Set/facet_morality.txt")
data.subset <- sample(char_vector,10000)
```

The Data Pre-processing component containing functions for data cleaning as mentioned in the report.

```
##### Data Preprocessing #####
# removing urls
data.subset <- rm_url(data.subset, pattern=pastex("@rm_twitter_url", "@rm_url"))
# removing special characters and numbers
data.subset <- gsub("[^A-Za-z//'"", "", data.subset)
# removing stopwords
data.subset <- removeWords(data.subset, stopwords('en'))
# removing badwords
badwords<-readLines("http://www.cs.cmu.edu/~biglou/resources/bad-words.txt")
badwords<-badwords[-(which(badwords%in%c("refugee","reject","remains","screw",
"welfare","sweetness","shoot","sick","shooting","servant",
"radical","racial","racist","republican","public","molestation","mexican",
"looser","lesbian","liberal","kill","killing","killer","heroin","fraud",
"fire","fight","fairy","^die","death","desire","deposit","crash","^crim",
"crack","^color","cigarette","church","^christ","cancer",
"^catholic","cemetery","buried","burn","^bomb","^beast","attack",
"australian","balls","baptist","^addict","abuse","abortion","amateur","asian",
"aroused","angry","arab","bible")==TRUE))]
data.subset <- removeWords(data.subset,badwords)
# removing two letters and one letter words
data.subset <- gsub(".*\\b[[:alpha:]]{1,2}\\b.*", "", data.subset)
```

The function used for tokenizing sentences and POS Tagging. The subset of the dataset taken is sent as an input to the function.

```
##### POS Tagging #####
tagPOS <- function(x, ...) {
  str <- as.String(x)
  word_token_annotator <- Maxent_Word-Token_Annotator()
  antn <- Annotation(1L, "sentence", 1L, nchar(str))
  antn <- annotate(str, word_token_annotator, antn)
  antnPOSTag <- annotate(str, Maxent_POS_Tag_Annotator(), antn)
  antnPOSTagWord <- antnPOSTag[antnPOSTag$type == "word"]
  tgPOS <- unlist(lapply(antnPOSTagWord$features, `[`, "pos"))
  tgdPOS <- paste(sprintf("%s/%s", str[antnPOSTagWord], tgPOS), collapse = " ")
  list(tgdPOS = tgdPOS, tgPOS = tgPOS)
}

tagged_str <- tagPOS(data.subset)
```

Function for extracting the adjectives from the POS-tagged data.

```
# Extracting adjectives
adjopnss <- unlist(str_extract_all(unlist(tagged_str[[1]]), "\\w+(?=[\\s\\p{P}])"))
adjopnss <- tolower(adjopnss)
adjopnss <- unique(names(sort(table(adjopnss), decreasing=T)[1:(0.1*length(adjopnss))]))
```

Identifying archetypes

Reading the generated bag-of-words for each trait.

```
#### Reading Traits####
char_achievement_striving <- readLines("F:/UConn/Spring/Big Data Strategic Mktg/Project/TextFile/adj_achievement_striving.txt")
# char_achievement_striving <- read.table("F:/UConn/Spring/Big Data Strategic Mktg/Project/TextFile/adj_achievement_striving.txt",header=TRUE, strip.white=TRUE)
char_activity_level <- readLines("F:/UConn/Spring/Big Data Strategic Mktg/Project/TextFile/adj_activity_level.txt")
# char_activity_level <- read.table("F:/UConn/Spring/Big Data Strategic Mktg/Project/TextFile/adj_activity_level.txt",header=TRUE, strip.white=TRUE)
char_adventurousness <- readLines("F:/UConn/Spring/Big Data Strategic Mktg/Project/TextFile/adj_adventurousness.txt")
char_altruism <- readLines("F:/UConn/Spring/Big Data Strategic Mktg/Project/TextFile/adj_altruism.txt")
char_anger <- readLines("F:/UConn/Spring/Big Data Strategic Mktg/Project/TextFile/adj_anger.txt")
char_anxiety <- readLines("F:/UConn/Spring/Big Data Strategic Mktg/Project/TextFile/adj_anxiety.txt")
char_artistic_interests <- readLines("F:/UConn/Spring/Big Data Strategic Mktg/Project/TextFile/adj_artistic_interests.txt")
char_assertiveness <- readLines("F:/UConn/Spring/Big Data Strategic Mktg/Project/TextFile/adj_assertiveness.txt")
char_agreeableness <- readLines("F:/UConn/Spring/Big Data Strategic Mktg/Project/TextFile/adj_agreeableness.txt")
char_conscientiousness <- readLines("F:/UConn/Spring/Big Data Strategic Mktg/Project/TextFile/adj_big5_conscientiousness.txt")
char_extraversion <- readLines("F:/UConn/Spring/Big Data Strategic Mktg/Project/TextFile/adj_big5_extraversion.txt")
char_neuroticism <- readLines("F:/UConn/Spring/Big Data Strategic Mktg/Project/TextFile/adj_big5_neuroticism.txt")
char_cautiousness <- readLines("F:/UConn/Spring/Big Data Strategic Mktg/Project/TextFile/adj_cautiousness.txt")
char_challenge <- readLines("F:/UConn/Spring/Big Data Strategic Mktg/Project/TextFile/adj_challenge.txt")
char_cheerfulness <- readLines("F:/UConn/Spring/Big Data Strategic Mktg/Project/TextFile/adj_cheerfulness.txt")
char_closeness <- readLines("F:/UConn/Spring/Big Data Strategic Mktg/Project/TextFile/adj_closeness.txt")
char_conservation <- readLines("F:/UConn/Spring/Big Data Strategic Mktg/Project/TextFile/adj_conservation.txt")
char_cooperation <- readLines("F:/UConn/Spring/Big Data Strategic Mktg/Project/TextFile/adj_cooperation.txt")
char_curiosity <- readLines("F:/UConn/Spring/Big Data Strategic Mktg/Project/TextFile/adj_curiosity.txt")
char_depression <- readLines("F:/UConn/Spring/Big Data Strategic Mktg/Project/TextFile/adj_depression.txt")
char_dutifulness <- readLines("F:/UConn/Spring/Big Data Strategic Mktg/Project/TextFile/adj_dutifulness.txt")
char_emotionality <- readLines("F:/UConn/Spring/Big Data Strategic Mktg/Project/TextFile/adj_emotionality.txt")
char_excitement <- readLines("F:/UConn/Spring/Big Data Strategic Mktg/Project/TextFile/adj_excitement.txt")
char_excitement_seeking <- readLines("F:/UConn/Spring/Big Data Strategic Mktg/Project/TextFile/adj_excitement_seeking.txt")
char_friendliness <- readLines("F:/UConn/Spring/Big Data Strategic Mktg/Project/TextFile/adj_friendliness.txt")
char_gregariousness <- readLines("F:/UConn/Spring/Big Data Strategic Mktg/Project/TextFile/adj_gregariousness.txt")
char_harmony <- readLines("F:/UConn/Spring/Big Data Strategic Mktg/Project/TextFile/adj_harmony.txt")
char_hedonism <- readLines("F:/UConn/Spring/Big Data Strategic Mktg/Project/TextFile/adj_hedonism.txt")
char_ideal <- readLines("F:/UConn/Spring/Big Data Strategic Mktg/Project/TextFile/adj_ideal.txt")
char_imagination <- readLines("F:/UConn/Spring/Big Data Strategic Mktg/Project/TextFile/adj_imagination.txt")
char_immoderation <- readLines("F:/UConn/Spring/Big Data Strategic Mktg/Project/TextFile/adj_immoderation.txt")
char_intellect <- readLines("F:/UConn/Spring/Big Data Strategic Mktg/Project/TextFile/adj_intellect.txt")
char_liberalism <- readLines("F:/UConn/Spring/Big Data Strategic Mktg/Project/TextFile/adj_liberalism.txt")
char_liberty <- readLines("F:/UConn/Spring/Big Data Strategic Mktg/Project/TextFile/adj_liberty.txt")
char_love <- readLines("F:/UConn/Spring/Big Data Strategic Mktg/Project/TextFile/adj_love.txt")
char_modesty <- readLines("F:/UConn/Spring/Big Data Strategic Mktg/Project/TextFile/adj_modesty.txt")
char_morality <- readLines("F:/UConn/Spring/Big Data Strategic Mktg/Project/TextFile/adj_morality.txt")
char_openness <- readLines("F:/UConn/Spring/Big Data Strategic Mktg/Project/TextFile/adj_big5_openness.txt")
char_openness_to_change <- readLines("F:/UConn/Spring/Big Data Strategic Mktg/Project/TextFile/adj_openness_to_change.txt")
char_orderliness <- readLines("F:/UConn/Spring/Big Data Strategic Mktg/Project/TextFile/adj_orderliness.txt")
char_practicality <- readLines("F:/UConn/Spring/Big Data Strategic Mktg/Project/TextFile/adj_practicality.txt")
char_self_consciousness <- readLines("F:/UConn/Spring/Big Data Strategic Mktg/Project/TextFile/adj_self_consciousness.txt")
char_self_discipline <- readLines("F:/UConn/Spring/Big Data Strategic Mktg/Project/TextFile/adj_self_discipline.txt")
char_self_efficacy <- readLines("F:/UConn/Spring/Big Data Strategic Mktg/Project/TextFile/adj_self_efficacy.txt")
char_self_enhancement <- readLines("F:/UConn/Spring/Big Data Strategic Mktg/Project/TextFile/adj_self_enhancement.txt")
char_self_expression <- readLines("F:/UConn/Spring/Big Data Strategic Mktg/Project/TextFile/adj_self_expression.txt")
char_stability <- readLines("F:/UConn/Spring/Big Data Strategic Mktg/Project/TextFile/adj_stability.txt")
char_structure <- readLines("F:/UConn/Spring/Big Data Strategic Mktg/Project/TextFile/adj_structure.txt")
char_sympathy <- readLines("F:/UConn/Spring/Big Data Strategic Mktg/Project/TextFile/adj_sympathy.txt")
char_transcendence <- readLines("F:/UConn/Spring/Big Data Strategic Mktg/Project/TextFile/adj_transcendence.txt")
char_trust <- readLines("F:/UConn/Spring/Big Data Strategic Mktg/Project/TextFile/adj_trust.txt")
char_vulnerability <- readLines("F:/UConn/Spring/Big Data Strategic Mktg/Project/TextFile/adj_vulnerability.txt")
```

A temporary directory is created.

```
##### Archetypes #####
# rm(td)
td = tempfile()
dir.create(td)
```

The 52 bag of words is mapped with the 12 Archetypes and inserted into the directory created.

```

#### Innocent #####
write( unique(c(char_dutifulness,char_friendlyness,char_morality,char_openness,char_sympathy,char_trust,
char_ideal,char_transcendence)), file=paste(td, "D1", sep="/t"))

#### Hero #####
write( unique(c(char_achievement_striving,char_activity_level,char_assertiveness,char_excitement_seeking,
char_self_discipline,char_challenge,char_self_enhancement)), file=paste(td, "D2", sep="/t"))

#### Regular Guy #####
write( unique(c(char_cheerfulness,char_cooperation,char_friendlyness,char_modesty,char_trust,
char_self_expression,char_hedonism)), file=paste(td, "D3", sep="/t"))

#### Nurturer #####
write( unique(c(char_altruism,char_cheerfulness,char_dutifulness,char_sympathy,char_trust,char_love,
char_transcendence)), file=paste(td, "D4", sep="/t"))

#### Creator #####
write( unique(c(char_artistic_interests,char_emotionality,char_imagination,char_immoderation,char_vulnerability,
char_self_expression,char_self_enhancement)), file=paste(td, "D5", sep="/t"))

#### Explorer #####
write( unique(c(char_activity_level,char_adventurousness,char_assertiveness,char_excitement_seeking,
char_immoderation,char_excitement,char_hedonism)), file=paste(td, "D6", sep="/t"))

#### Rebel #####
write( unique(c(char_anger,char_adventurousness,char_liberalism,char_excitement_seeking,
char_immoderation,char_liberty,char_hedonism)), file=paste(td, "D7", sep="/t"))

#### Lover #####
write( unique(c(char_cooperation,char_emotionality,char_friendlyness,char_self_consciousness,char_sympathy,
char_love,char_self_enhancement)), file=paste(td, "D8", sep="/t"))

#### Magician #####
write( unique(c(char_achievement_striving,char_activity_level,char_adventurousness,char_imagination,
char_intellect,char_curiosity,char_openness_to_change)), file=paste(td, "D9", sep="/t"))

#### Ruler #####
write( unique(c(char_anger,char_assertiveness,char_morality,char_orderliness,
char_self_discipline,char_structure,char_conservation)), file=paste(td, "D10", sep="/t"))

#### Jester #####
write( unique(c(char_altruism,char_artistic_interests,char_friendlyness,char_gregariousness,char_immoderation,
char_harmony,char_transcendence)), file=paste(td, "D11", sep="/t"))

#### Sage #####
write( unique(c(char_achievement_striving,char_assertiveness,char_morality,char_dutifulness,
char_self_discipline,char_structure,char_conservation)), file=paste(td, "D12", sep="/t"))

```

The directory created is read into a Document-Term matrix.

```

##### read files into a document-term matrix #####
myMatrix = textmatrix(td, minwordlength=1)

```

Calculating similarity scores of the input text file against each of the 12 archetypes.

```

##### Calculate cosine similarity #####
resMat <- array()
for (i in 1:12)
{
  resMat[i] <- lsa::cosine(myMatrix[,i], myMatrix[,13])
}

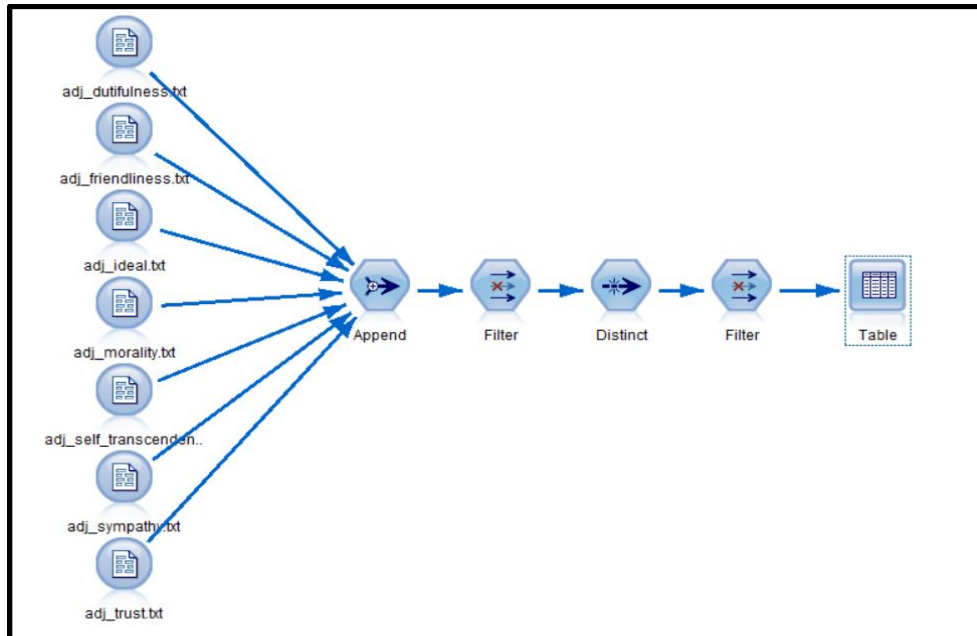
names(resMat) <- c("Innocent","Hero","Regular Guy","Nurturer","Creator","Explorer","Rebel",
" Lover","Magician","Ruler","Jester","Sage")
resMat

```

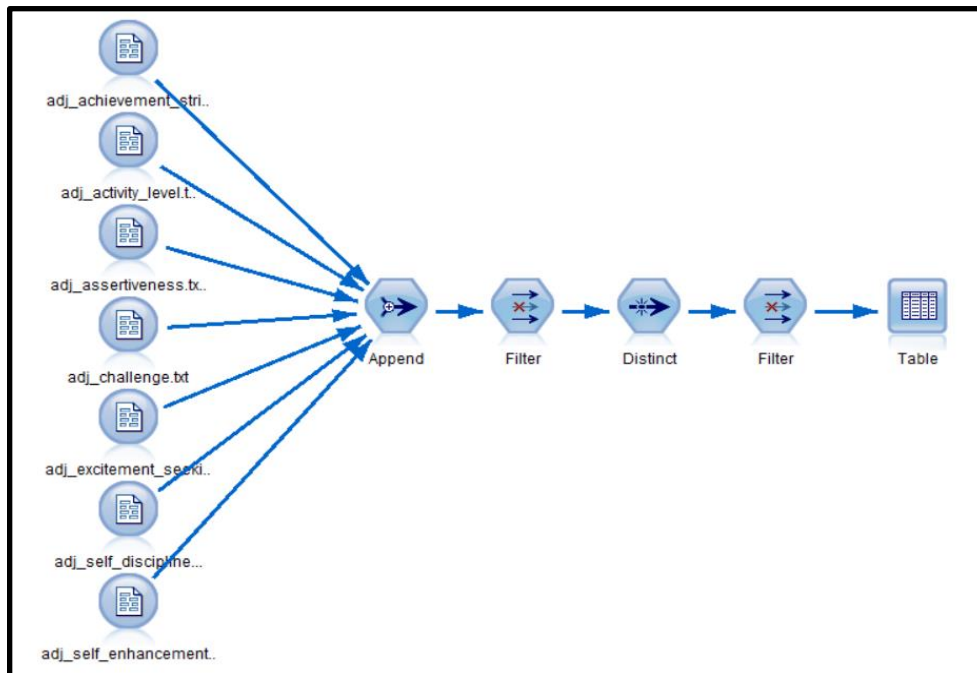
7.2 SPSS Models

Mapping 52 personality traits to 12 archetypes using IBM SPSS Modeler

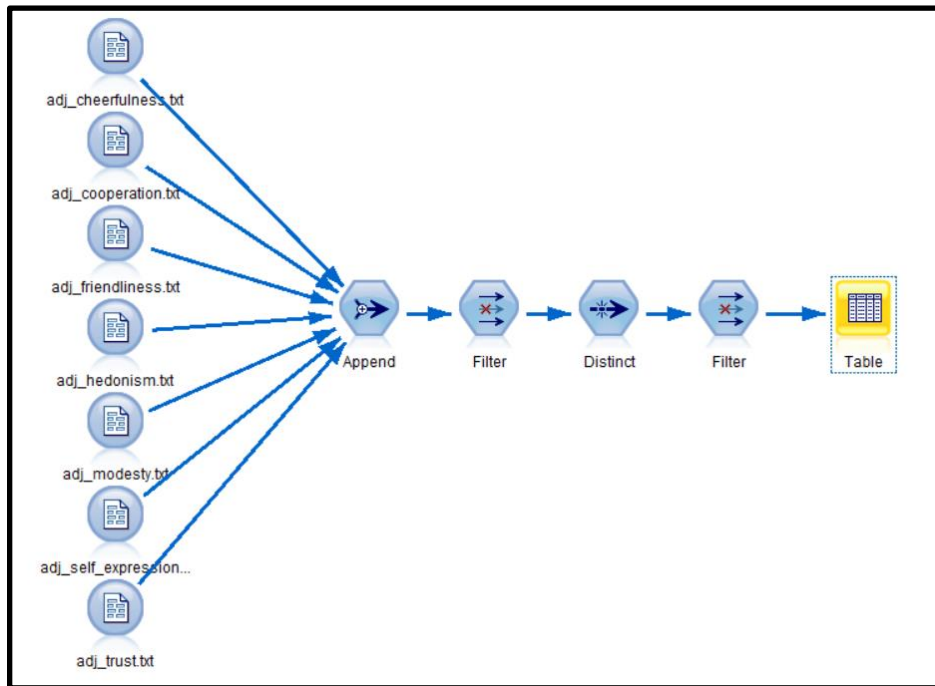
The Innocent



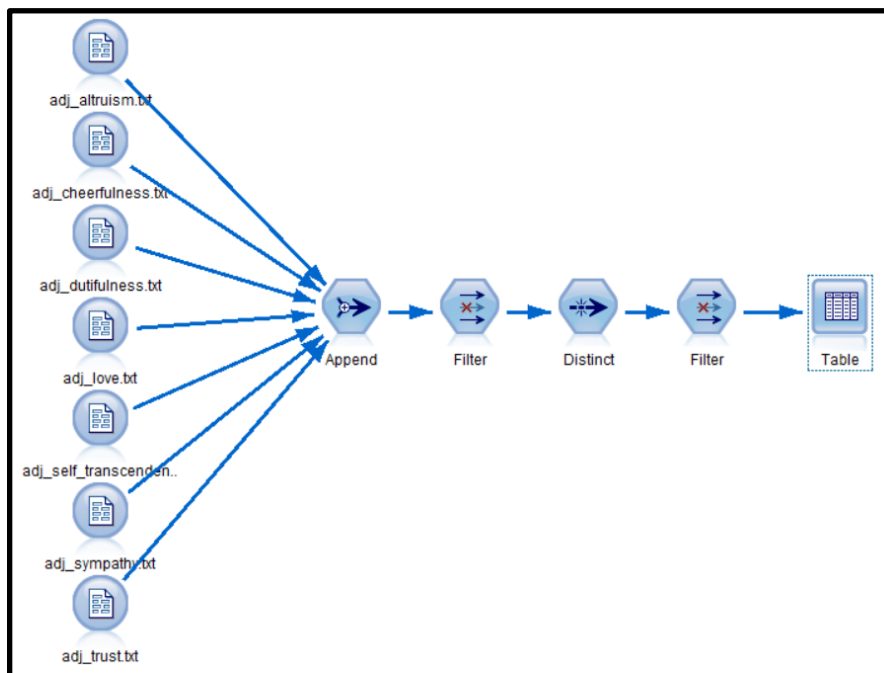
The Hero



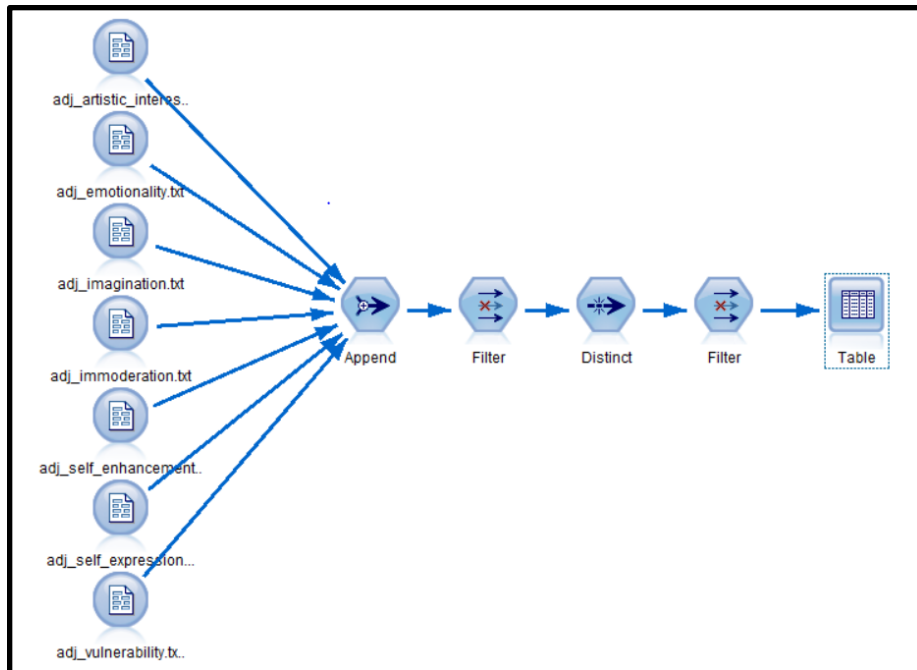
The Regular Guy



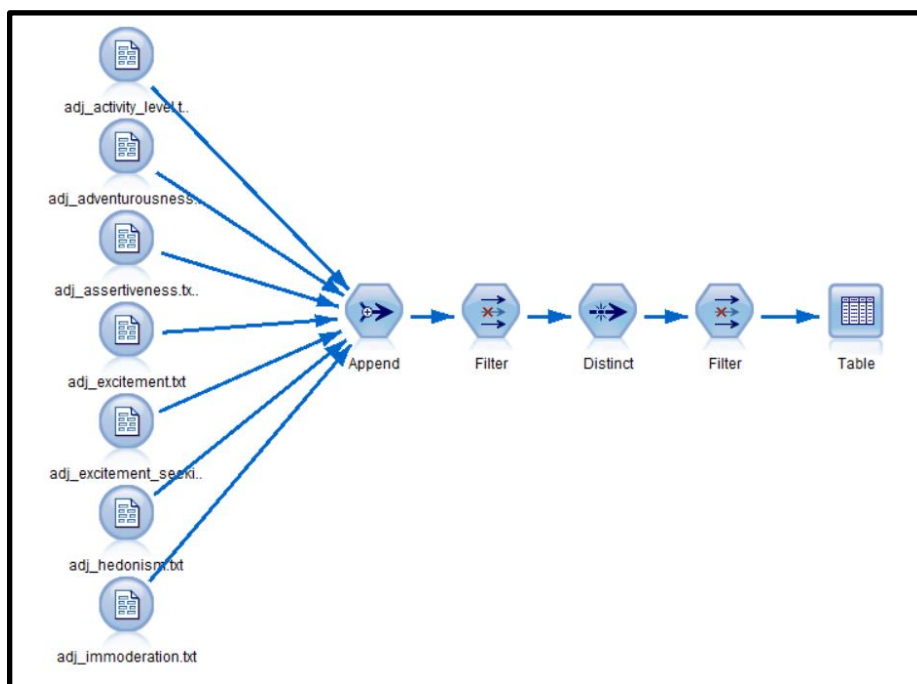
The Nurturer



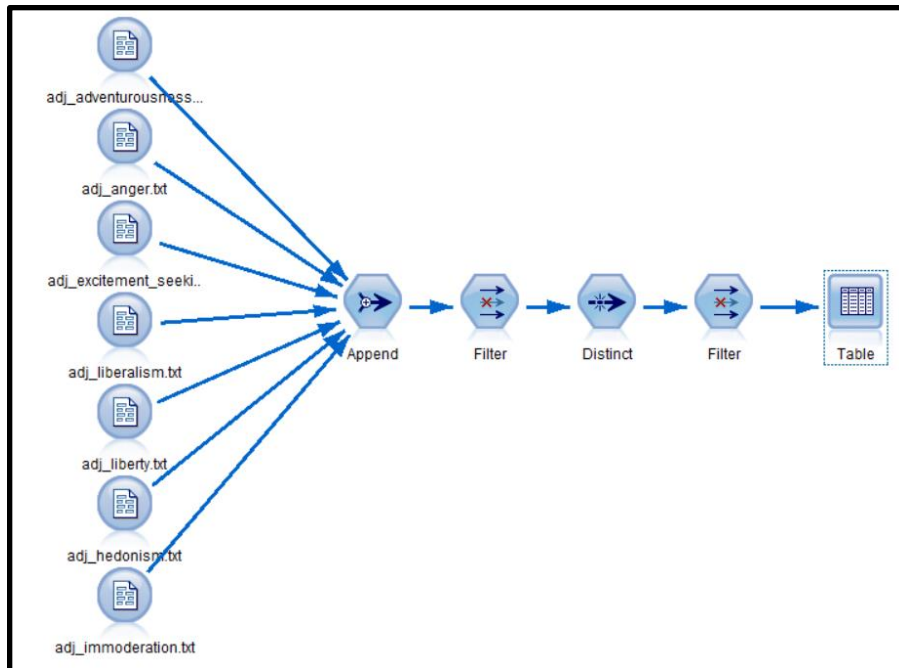
The Creator



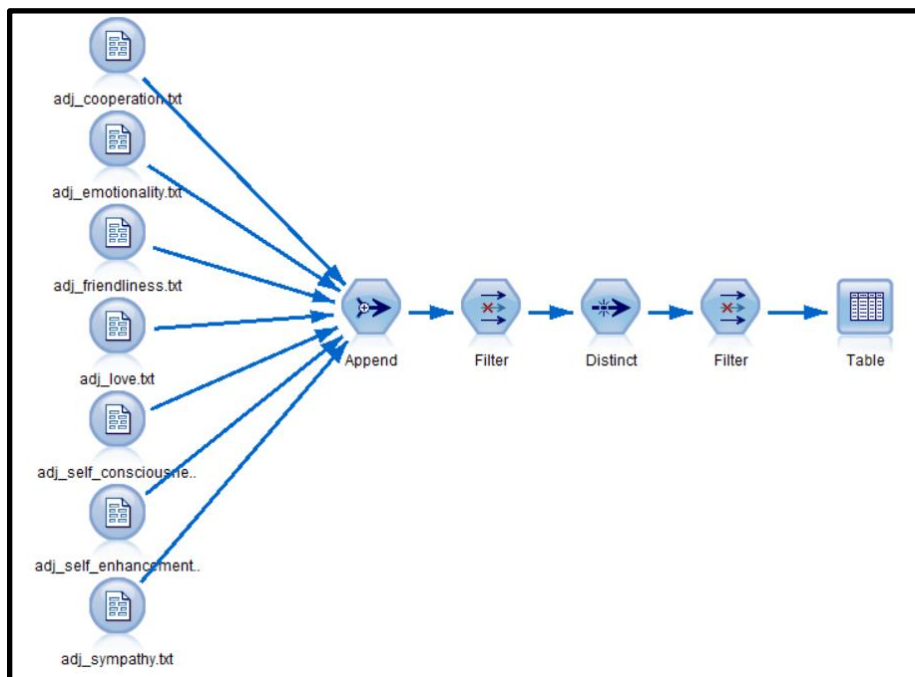
The Explorer



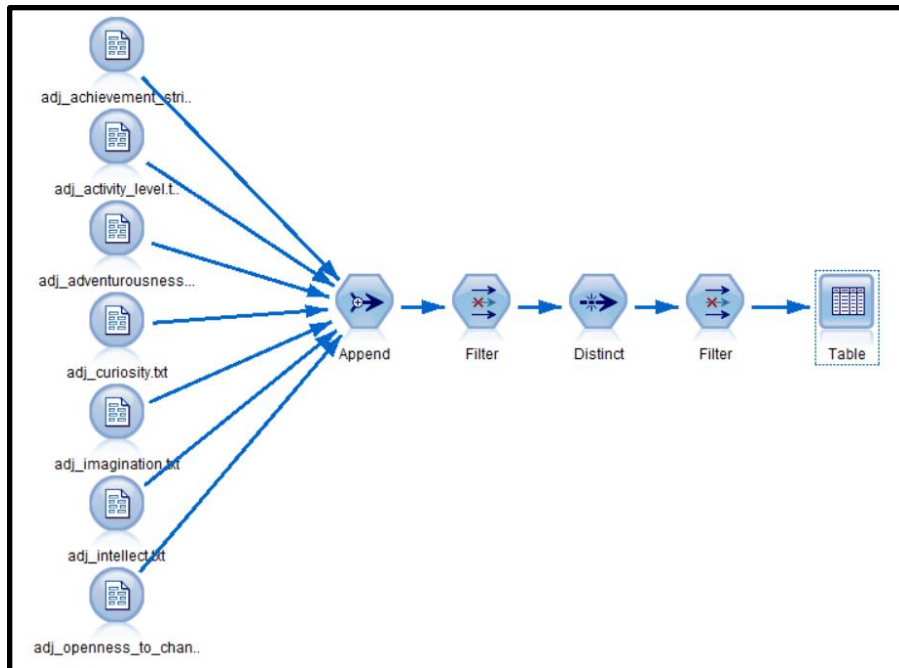
The Rebel



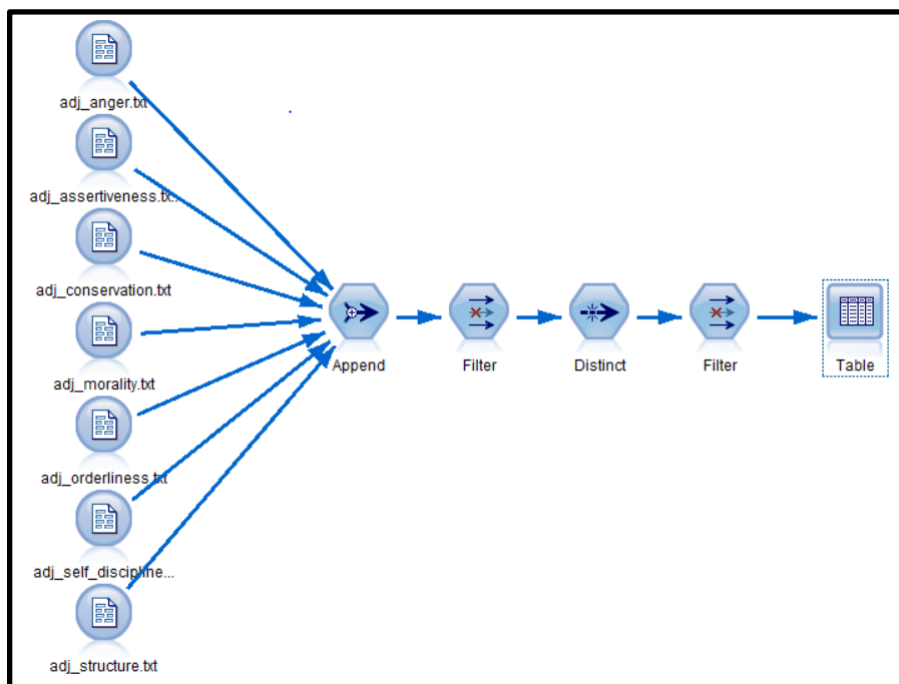
The Lover



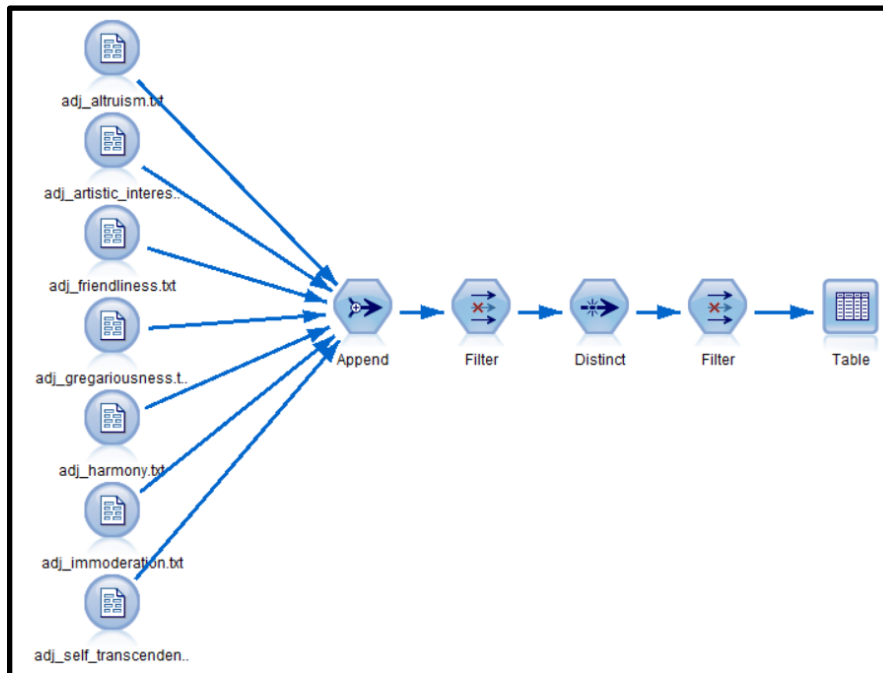
The Magician



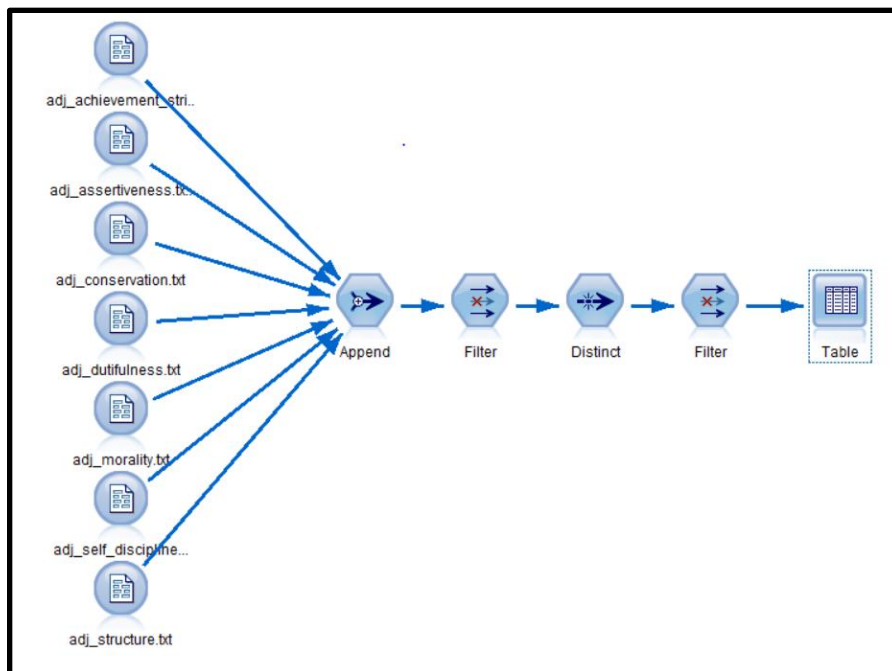
The Ruler



The Jester



The Sage



Authors:



Kunja Dutta

Bachelors of Engineering in Electrical and Electronics from M.S.Ramaiah Institute of Technology, India (2008-2012). Master of Science in Business Analytics and Project Management from the University of Connecticut (2016-2017).

4 years of work experience with Schneider Electric as a Software Engineer (2012-2016).

Business analytics enthusiast with excellent problem solving skills. Industry experience in requirement analysis and planning, programming, testing and energy management across different domains such as Electrical and Information Technology.

Industry experience includes interaction with cross functional teams located in the US.

Recognized as a team player, effective communicator, and an excellent organizer.

Hobbies: Painting, Playing the Piano, Trekking



Debaleena Saha

Bachelor of Technology in Information Technology from West Bengal University of Technology, India (2003-2007). Master of Science in Business Analytics and Project Management from the University of Connecticut (2016-2017)

Graduate student with a focus on Big Data Analytics and Predictive Modeling and over 8 years of IT experience in Microsoft Dot Net Technologies and SQL Server; Have varied experience in Techno-functional role - Development, Production support, Maintenance, Project Management experience as a Team Lead; Excellent analytical and management skills for resolving standing problems and creating solutions; Wide range of experience in working with top Insurance and Travel & Tourism customers across USA and UK.

Companies worked for:

IBM(2011-2016), CGI Information Systems(2009-2011), Satyam Computer Services Ltd. (2008-2009)

Hobbies: Photography, Singing



Samarpita Chakrabarty

Bachelor of Technology in Information Technology from West Bengal University of Technology

Master of Science in Business Analytics and Project Management from the University of Connecticut (2016-2017)

4 years of work experience with Cognizant Technology Solutions as a Data Warehousing and Business Intelligence Developer. Worked with multiple databases and BI Tools. Have worked with global clients in the domains of Banking & Finance and Insurance.

Responsibilities: Requirement gathering and Analysis, Coding, Implementation, Integration with UI, Unit Testing, Defect fixing and Maintenance, Cost estimation, Risk analysis, Quality Control, and Project planning for multiple projects.

Hobbies: Travelling, Cooking, Dancing, Reading



Mayank Gupta

Bachelor of Engineering in Electronics and Communications from Jaypee University Of Information Technology, India (2006-2010).

Master of Business Administration (MBA) from University of Connecticut - School of Business with specializations in Business Analytics & Digital Marketing Strategy (2015-2017).

Experienced professional with expertise in Analytics, Consulting and Product/Project Management having worked in the IT Industry for more than 5 years. I have helped customers solve their business challenges by providing impactful solutions. Have experience working with global clients like Morgan Stanley and Royal Bank of Scotland (RBS).

Prior to my MBA, I was working with Infosys Limited, where my responsibilities included people management, leading teams, cross-functional collaboration, problem solving through data analysis and coordinating test planning strategies.

Hobbies: Travelling, Watching F1 and Soccer, playing all kinds of sports, music.