

**Project Non-Technical Summary Report**  
**CSC423**  
**Group Name: TBD**

Manish Singh section:

**Goal:**

The dataset we are using is collection of responses of 1010 participants of ages between 15-30. The survey had a total of 151 questions covering a broad range of topics amongst which I have focused on questions and their responses which will help me analyze if a person is a money saver or not.

**Process:**

Few examples of questions I have chosen are, whether one likes to spend on entertainment, shopping centers, branded clothes, healthy eating, gadgets, and, also, questions like their education, whether they spend most of his time in a city or town, number of siblings they have, etc. These questions can also be called as variables.

Based on the statistical analysis I performed using SAS, I was able to filter even more amongst these variable and the three major variable I was left with throughout the analysis were Spending on entertainment, spending on looks, and whether the person likes spending most time in the village or a city.

Then, I calculated significance and other parameters as to how these variables were impacting our end goal.

- If the person was spending more on entertainment, the probability if him being a money saver was less.
- If the person was spending more on looks, the probability if him being a money saver was less.
- If the person preferred spending most of his/her time in a village, the probability if him being a money saver was more.

**Applications:**

Huge organizations like, for example, Google, Apple, Facebook, Amazon always like to go an extra mile for their employees. They can use a model like this to analyze if their employees are money savers or not and have guidance sessions to help one's who are not and make them understand the importance of saving.

Michal Chowaniak section:

Our group's dataset was created by Miroslav Sabo, based on a survey of 1010 participants age of 15 to 30, done in 2013 in Slovakia. The survey asked 151 questions from several areas like music, movie, hobbies preferences, phobias, health habits, personality and the most important for our group spending habits and demographics. I decided to focus on a question about saving or not saving money. I used different statistical techniques to find out what are characteristics of a person who save or does not save money.

My analysis resulted in a confirmation for some of general common sense ideas regarding saving money, and it gave the biggest weight to following characteristics: entertainment, appearance, healthy food and childhood hometown.

This is how saving money is influenced by above characteristics:

1. More money people spend on entertainment less money they save.
2. More money they spend on appearance less money they save.
3. If they spend money on healthy food more likely they also save money.
4. If they grew up in a village (small town) more likely they save money.

My statistical model can do a good job in predicting if a person is NOT saving money, and the model is worse in predicting if a person is saving money. This is actually a good thing, I would rather know who needs help in changing spending behavior then otherwise.

This model would be helpful for a company or non-profit organization interested in educating people about benefits of saving money or targeting people with product advertisements, because they spend money instead of saving them.

Priyank Beno Cerejo section:

### **Non-technical Report**

This project aims to determine whether a person is either someone who saves money or spends it based on various factors, using SAS. The possible and potential factors are: 'I am fond of visiting large shopping areas', 'I choose branded clothes over the non-branded ones', 'I spend most of my money on partying with people and socializing', 'I spend a considerable measure of cash on the way I look', 'I spend a great deal of money on gadgets and electronics', 'I will readily spend more money for good, quality or healthy nourishment without any hesitation', age, height, weight, number of siblings, gender, left handed or right handed, highest degree completed, single kid or not, childhood spent in a city or a village, childhood in house or block of flats. All these factors are taken from a study at <https://www.kaggle.com/miroslavsabo/young-people-survey>, and we will use the above factors to predict whether a specific person is either cash saver or high roller. 'Finances' is the expression used for this prediction, and if that person is categorized as a money saver, then 'Finances' will be 4 or 5, else, it will be 1, 2 or 3.

The above components just have possible association with 'Finances', so we will need to evaluate the significance of every factor's effect for it. For instance, judging by basic common sense, a person's left handedness or right handedness shouldn't be counted as an important factor to check if he/she is a money spender or saver. Hence, this factor can be expelled from the list and we can go on to evaluate the next one. After every one of the elements are checked; in the long run we will have a rundown with factors which have most impact on 'Finances', and we can utilize them to make assumption about one's financial behavior.

However, to know if one factor can truly have significant relationship, the fundamental knowledge like common sense might not be enough, and we need to have a greater deal of understanding about the factors and statistical results before we use those factors to jump to a conclusion. Therefore, statistic methods are required and the variables which are incorporated into the final outcome from the model will be the ones we use for our assumption.

To conclude, the complete process of this project can be summarized as the following depiction:

1. We generate the dataset with factors that has probable relationship to a person's financial behavior – either money high roller or saver.
2. Using the data as the input to the system, statistical relationship between the factors and 'Finances' is expected as the output. This is the training data which is used to prepare the system for more accurate results.
3. Using factors with robust statistical importance, an equation was formulated which helps ascertain people's financial behavior and determine the efficiency of the equation.
4. Based on the equation's performance and using of make-up values wherever required, the result from this equation will give the probability of 'Finances'. People with higher value of probability will be a money saver else a money spender.

Rushabh Shah section:

Our aim was to take the dataset of survey results, and develop a model to predict a person's financial behavior, aka are they money spenders, or money savers? This model would be created using the wide variety of variables that survey participants were asked about. These variables included financial spending or saving habits, preferences regarding branded/non-branded clothing, spending habits on looks, entertaining, healthy eating. Other variables included age, gender, height, weight, number of siblings, highest education achieved, whether the participant grew up in a village or city, and whether they grew up in a house or block of flats. These variables, as well as many others were collected by surveyors, and organized into a table. The other variables were deemed as unimportant to our predictive model, so we selected only the variables listed above to help create our predictive model. For many of the variables, participants answered with a 1-5, with 1 being strongly disagree, and 5 being strongly agree. Other variables were answered with categorical answers, such as male/female, house/block of flats, etc.

Amongst all the variables, only the finance variable was dependent, all the other variables are independent of each other. With this knowledge, we can try to predict what this dependent variable would be, given values for all the other variables. Using the SAS software, I generated boxplots of each variable against finance, which helped me narrow down which variables I would use. After selecting the variables that would be included in the initial model, I generated a regression model, and went through forward selection, narrowing down my predictive models variable list. Eventually my model was narrowed down to include just urban (village or city) and entertainment spending habit. The final regression model was  $\text{finance} = 0.8075 - (0.4890 * \text{Urban}) - (0.3598 * \text{Entertainment\_spending})$ . This equation indicates that if a person grew up in the city and spent a lot on entertainment, they were more than likely going to be spenders, whereas if they grew up in a village and spent very little on entertainment, they were more than likely savers.

Valentine Silvester Correia section:

This project is being performed in SAS which is a software for advanced data analytics, data management and predictive analytics. The chosen data is Young People Survey and is a survey that was done in 2013 by students from Slovakia aged between 15-30 years. The data file basically consists of 1010 rows and 150 columns consisting of variables and had missing values.

I have focused on 15 variables to determine which person is a saver or not saver when it comes to saving money and for predicting this we are considering the variable Finances for which the values range from 1 to 5 and person who has Finance value as 4 or 5 will be a saver and the person who has Finance value as 1, 2 or 3 will not be a saver. Some of the potential factors are "I spend a lot of money on my appearance", "I spend a lot of money on Entertainment" and "childhood in city or village".

I have cleaned the data, used Logistic regression to perform analysis by creating the needed dummy variable i.e. modifying the data to use it as per the regression technique. Then I have split my data for training and testing it for validation. I have used two models with different models in it to predict the person is a saver or a spender and chose the best of the two.

The analysis I performed was influenced by variables like spending on entertainment, spend on looks and lived in village or town and predicted that if a person spent more money on entertainment and look they were spender and they people who grew up in Villages were probably the savers. This prediction is very obvious.

## Yesheng Qin section

The purpose of this project is to use SAS to determine whether one person is either a money saver or spender based on multiple factors. The potential factors are: 'I enjoy going to large shopping centres', 'I prefer branded clothing to non-branded', 'I spend a lot of money on partying and socializing', 'I spend a lot of money on my appearance', 'I spend a lot of money on gadgets', 'I will happily pay more money for good, quality or healthy food', age, height, weight, number of siblings, gender, left handed or right handed, highest education achieved, only child or not, childhood in city or village, childhood in house or block of flats. These factors are from a survey at <https://www.kaggle.com/miroslavsabo/young-people-survey>, and the above are the ones we will use to predict whether a person is either money saver or spender. We use the term 'Finances' for this prediction, and if a person is considered as money saver, then 'Finances' will be 4 or 5, if not then it will be 1, 2 or 3.

The above factors only have possible relationship with 'Finances', so we will need to calculate if the effect of each factor is considered significant for it. For example, based on common sense, a person is either left handed or right handed shouldn't count as a significant factor to determine he or she is a money spender or saver. Therefore, we can remove this factor from the list and check the next one. After all the factors are checked and eventually we will have a list with factors have most significant effect to 'Finances', and we can use them to make assumption about one's financial behavior.

However, the basic knowledge such as common sense is not enough to indicate whether one factor is truly having meaningful relationship, and we need to have statistic result for the factors before we use them to make prediction. Therefore, statistic methods are needed and the factors which are included in the final result from the model will be the one we use for our assumption.

The final result I come out from my model represents that only two factors have significant relationship with the identifier to determine one's financial behavior: willingness of spending on entertainment and appearance. The equation of my final model is:  $\text{Log}(p/(1-p)) = 0.8129 - 0.2302 * \text{willingness of spending on entertainment} - 0.2362 * \text{willingness of spending on appearance}$ , this equation indicates that if a person is more willing to spend either or both on entertainment and appearance, the chance he or she is a money saver will decrease.

Therefore, people who want to save money would need to consider limit the spending on entertainment and appearance.