Factors influencing the choice of a young Lviv resident to become a vegetarian or vegan

Khrystyna Sliusarchuk

April 2021

1   Introduction

Vegetarianism and veganism become more and more popular in the modern world. By now, there are almost 1 billion people who follow a plant-based diet on our planet. Those who choose this lifestyle do so for a vast variety of reasons. The most popular of them include the desire to reduce the suffering of animals which they experience at the factories and plants where our 'younger brothers' are used for different purposes: production of animal based food, production of natural fur and leather, testing of cosmetics, etc. Another significant reasons are those connected with the environmental issues and the desire to reduce the greenhouse gas emissions because a third of all emissions on our planet are caused by food production. This list also contains many other factors, including religious, medical and economic ones.

When talking about Ukraine, our country ranks 11th in Europe in terms of interest of population in this lifestyle. According to the survey, conducted in 2017 by NGO "Open cages" and Kiev International Institute of Sociology, almost 2 million Ukrainians have consciously given up eating meat and are vegetarians or vegans. When the survey was conducted again in 2019, the results changed. Actually, they showed that over the past two years the number has increased by 1.9%, which means that the number of people on a plant-based diet has increased by more than half a million. What is more, the survey revealed that most Ukrainians who choose to become a vegetarian or vegan are young people between the ages of 18 and 29 and, depending on the region, vegetarianism and veganism are more popular in the western and central regions of Ukraine.

Thus, we can assume that in Ukraine these two lifestyles are only gaining popularity and will become more and more widespread every year. Since, according to the results of the survey described above, the majority of vegetarians and vegans in Ukraine are young people from the Western regions of Ukraine, trying to identify the factors that influence a young Lviv resident's choice to follow the plant-based diet is a very interesting and urgent task.

2 Data Description

Since the data or some examples of researches on the same or similar topic were not available on the Internet, it was decided to collect the data independently by conducting a survey. A form was created, where the respondent had to answer the questions concerning his/her gender, age, place of residence, levels of education, educational specialization, number of foreign languages the respondent knows, religion, marital status, income. The next part of the survey included questions about the presence of vegetarians and vegans, pets in respondent's life, empathy level and existence of problems with overweight, attitude towards the LGBT community and vaccine and the final one about trash sorting. One can view the format of the survey and questions in more details by following this link: `https://forms.gle/mu39iEVaP9ZFkPxX6`.

As a result of the survey, 194 responses were collected. All answers were saved in a .xlsx file, which was used for the research after some data processing.

veg_	sex_	age	educ_	educ_spec_	foreign_lang	religious_	mar_status_
0	male	25	17	human	2	no	not married
1	female	19	11	medic	2	no	not married
0	female	19	11	human	4	yes	not married
0	male	18	11	tech	2	yes	not married
0	female	18	11	tech	4	yes	not married

Table 1: Example of data

To confirm that the research is about **young Lviv** resident's choice to become a vegetarian or vegan, one can look at the following plots of the age-gender and place of residence distributions in the sample.
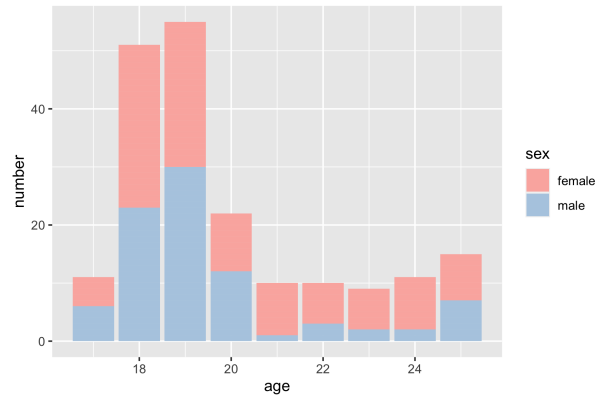

Figure 1: Age-sex distribution

On this plot one can see that the age of individuals in the sample ranges from 17, as the youngest one, to 25, as the oldest one. Based on this, it can be assumed that research is about **young** people.
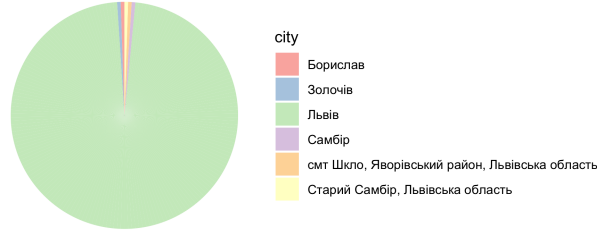

Figure 2: Place of residence distribution

On this plot one can see that the place of residence of almost all individuals in the sample is Lviv. The rest of the individuals live in the cities located in Lviv region. Based on this, it can be assumed that research is about **Lviv** residents.

The percentage of vegetarians or vegans in the sample is equal to 12.4%. This result can be seen on the following plot.
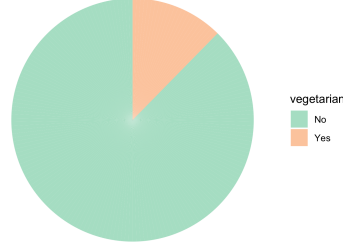

Figure 3: Vegetarians or vegans distribution

3 Methodology Explanation and Results

To identify what factors have an effect on the choice of a young Lviv resident to become a vegetarian or vegan, it was decided to build a model. Since in this case one deals with the binary dependent variable (whether a person is a vegetarian/vegan or not), the linear probability model is not applicable. The alternative is to use one of the binary response models - the logistic one. This model consists of 2 parts: in the first one it estimates the probability p of a binary outcome and in the second one this probability is linked to a linear equation using

logit function, also called log odds ratio, which actually transforms probabilities to real numbers from $-\infty$ to $+\infty$: $logit(p) = log(\frac{p}{1-p}) = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k$

Firstly, it was decided to include into the model almost all of the factors available, so that the model looked like:

$$veg\_ = \beta_0 + \beta_1 sex\_ + \beta_2 age + \beta_3 educ + \beta_4 educ\_spec\_ + \beta_5 foreign\_lang +$$

$$+ \beta_6 religious\_ + \beta_7 veg\_surround\_ + \beta_8 veg\_past\_ + \beta_9 pet\_ +$$

$$+ \beta_{10} empathy\_ + \beta_{11} overweight\_ + \beta_{12} vaccine\_ + \beta_{13} trash\_$$

What is important to mention here is that $sex\_$, $religious\_$, $veg\_surround\_$, $veg\_past\_$, $pet\_$, $empathy\_$, $overweight\_$ and $trash\_$ are binary categorical variables, $educ\_spec\_$ and $vaccine\_$ - categorical variables and age, educ and foreign_lang - numerical variables.

Marital status was not included, since the respondents are young and almost all of them are unmarried. Also, due to the young age of the respondents, it makes no sense to include wage and rent factors, many of them still do not earn anything. Individuals' attitude to LGBT community and vaccine are both proxies to the same hypothesis, so including them simultaneously into the model makes little sense.

To determine whether this model is applicable in this case, the following assumptions need to be met:

1. *The outcome is a binary dependent variable.* It is true in this model, since the dependent variable denotes whether a person is a vegetarian/vegan or not.

2. *The observations are independent.* It is true in this model, since the data is a random sample of individuals.

3. *There is no multicollinearity.* To check this, one can try to calculate a so called variance inflation factor (VIF). This indicator measures the strength of the correlation between the explanatory variables. The smallest value of VIF, which one can get, is 1. This would mean that there is no multicollinearity in the model. However, even if the value of VIF is in range from 1 to 5, the model is still applicable. When calculating the values of this indicator in this model, the variables $age\_$ and $educ\_$ revealed to be highly correlated, since the values of VIF for them were equal to 7.997246 and 8.255510, respectively. Based on this result, it was decided to exclude one of them, $educ\_$, from the model. After doing so, all of the values of VIF for all of the variables became in range from 1 to 5.

After running the logistic regression, this model turned out to be a little bit poor, since AIC was quite big - 115.16. That's why, it was decided to use R function *step*, which selects a formula-based model by Akaike Information Criterion. The result suggested to use the following model:

4

$$\text{veg}_- = \beta_0 + \beta_1 sex_- + \beta_2 educ\_spec_- + \beta_3 religious_- +$$

$$+ \beta_4 veg\_surround_- + \beta_5 empathy_- + \beta_6 trash_-$$

AIC decreased from 115.16 to 104.1.

After running the logistic regression on this model, the following estimates of the coefficients β_k, $k = 0, ..., 6$ were obtained:

	Estimate	Std. Error	z value	Pr($> |z|$)	
(Intercept)	-2.2486	0.8766	-2.565	0.01032	*
sex_male	-2.0676	0.6815	-3.034	0.00241	**
educ_spec_medic	-0.7195	0.6762	-1.064	0.28731	
educ_spec_tech	-1.5312	0.7144	-2.143	0.03209	*
religious_yes	-2.8598	0.6957	-4.111	3.95e-05	***
veg_surround_yes	1.6238	0.6862	2.366	0.01797	*
empathy_yes	1.1142	0.6266	1.778	0.07539	.
trash_yes	1.9209	0.6697	2.868	0.00413	**

Table 2: Estimated coefficients

When analyzing the table above, one can start from the last two columns, which are the p-values and significance codes, respectively. Based on these, one can state that almost all of the variables, except for the educ_spec_medicine are statistically significant.

The estimates of the coefficients, actually, do not provide one with some useful information, since they represent the change in the log odds ratio, which is something hard to understand. An alternative is to exponentiate all the estimated coefficients and thus obtain a new interpretation: a one unit change in x_j makes the odds ratio change by a factor of $e^{\hat{\beta}_j}$. Whether the odds ratio decreases or increases depends on whether the coefficient is negative or positive: if negative - increase in x_j will cause decrease in the odds ratio, and vice versa. The exponentiated estimated coefficients are as follows:

	Exponentiated estimate
(Intercept)	0.10555019
sex_male	0.12648871
educ_spec_medic	0.48698284
educ_spec_tech	0.21628297
religious_yes	0.05727811
veg_surround_yes	1.6238
empathy_yes	3.04712370
trash_yes	6.82701568

Table 3: Exponentiated estimated coefficients

Now, the coefficients can be interpreted in the following way: holding all other factors fixed, for those people who have vegetarians or vegans in their surrounding, the odds of becoming a vegetarian or vegan compared to the odds of not becoming a vegetarian or vegan are by a factor of 1.6238 larger, compared to those who do not have and so on.

What is more, one can calculate the confidence intervals for odds ratio. Since the odds ratio are not normally distributed, one can try to construct the confidence interval for the log odds ratio, which are normally distributed, and then exponentiate the obtained values. The formula for the 95% confidence interval for the log odds ratio is:

$$\hat{\beta}_j \pm 1.96 \times SE(\hat{\beta}_j)$$

After exponentiating the values obtained, the confidence intervals for the odds ratio are as follows:

	2.5%	97.5%
(Intercept)	0.01589157	0.5167129
sex_male	0.02962662	0.4412434
educ_spec_medic	0.12308601	1.8072584
educ_spec_tech	0.04907145	0.8389753
religious_yes	0.01273227	0.2016938
veg_surround_yes	1.44558128	22.4106799
empathy_yes	0.94190574	11.3520218
trash_yes	2.00620912	28.8718693

Table 4: Confidence intervals for the odds ratio

When it comes to determining the goodness-of-fit of this model, one can not compute usual R^2 like in the linear probability model. An alternative is to compute a so called pseudo R^2 squared, which is defined as $1 - \frac{L_{ur}}{L_r}$. In this formula, L_{ur} denotes the log-likelihood function for the whole estimated model, whereas L_r - the log-likelihood function for the model only with the intercept included. For the model, which was described above, the value of the pseudo R^2 squared is equal to 0.3933032, which is enough to state that this model represents a good fit.

Another way to determine the goodness-of-fit oh the model is to calculate the percent correctly predicted. The idea is quite straightforward: if the predicted probability is greater than some value t, then $\tilde{y}_i = 1$ and if less - $\tilde{y}_i = 0$. Then the accuracy of the model is calculated as the percentage of times when $\tilde{y}_i = y_i$.
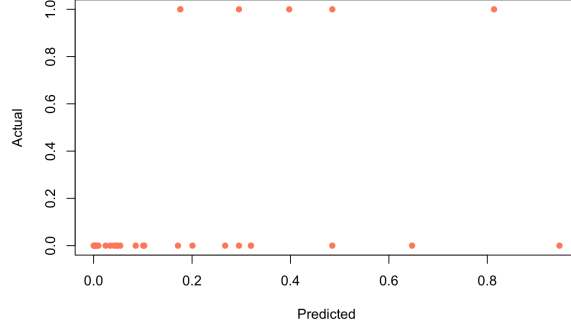
Figure 4: Predicted probabilities

One of the common values for t is 0.5. In such case, the confusion matrix will be of the form:

Predicted / Actual	0	1
0	31	5
1	2	1

The accuracy of the model when $t = 0.5$ is 82%.

An alternative to setting t equal to 0.5 is to set it to be equal to the fraction of the vegetarians or vegans in the sample - 0.1237113. The accuracy of the model will decrease, because there will be more mistakes in predicting non vegetarians or vegans, but the number of predicted vegetarians or vegans will increase. In such case, the confusion matrix will be of the form:

Predicted / Actual	0	1
0	24	0
1	9	6

The accuracy of the model when $t = 0.1237113$ is 77%. It decreased, but at the same it can be seen that the number of predicted vegetarians or vegans increased from 1 to 6, as it was stated before.

Also, one can try to introduce a logistic regression model with the multiplication term of some independent variables. One of the interesting choices to test is to multiply the empathy_ factor by the sex_ one. By doing so, one can receive the model of the form:

$$veg_- = \beta_0 + \beta_1 sex_- + \beta_2 educ\_spec_- + \beta_3 religious_- + \beta_4 veg\_surround_- +$$

$$+ \beta_5 empathy_- + \beta_6 empathy_- \times sex_- + \beta_7 trash_-$$

7

Then, when the individual is male and empathic, the odds of becoming a vegetarian or vegan comparing to the odds of not becoming are by a factor of $e^{\beta_5+\beta_6}$ smaller/larger, compared to those males who are not empathic, since the equation will be of the form:

$$\text{veg}_{-} = (\beta_0 + \beta_1) + \beta_2 educ\_spec_{-} + \beta_3 religious_{-} + \beta_4 veg\_surround_{-} +$$

$$+(\beta_5 + \beta_6) empathy_{-} + \beta_7 trash_{-}$$

On the other hand, when the individual is female and empathic , the odds of becoming a vegetarian or vegan compared to the odds of not becoming are by a factor of e^{β_5} smaller/larger, compared to those females who are not empathic, since the equation will be of the form:

$$\text{veg}_{-} = \beta_0 + \beta_2 educ\_spec_{-} + \beta_3 religious_{-} + \beta_4 veg\_surround_{-} +$$

$$+ \beta_5 empathy_{-} + \beta_7 trash_{-}$$

Unfortunately, after running the logistic regression, the variables $empathy_{-}$ and $empathy_{-} \times sex_{-}$ revealed to be insignificant, so there is no point in doing so.

4   Conclusions and limitations

All in all, one can state that the factors influencing the choice of a young Lviv resident to become a vegetarian or vegan include his gender, education specialization, whether he is a religious person, whether he has vegetarians or vegans in his surrounding, whether he is an empathic person and whether he sorts trash or not. For religious individuals, the odds of becoming a vegetarian or vegan compared to the odds of not becoming, will by some factor smaller. It is an interesting result, which is not quite easy to explain, but one can try to rely on the following facts: Ukraine is still quite a religious ountry, even in the modern rather secular world, the vast majority of all religious organizations in Ukraine are Christian and a big part of the individuals in the sample are the students of the Ukrainian Catholical University. Taking them and the fact that in the Christianity the attitude towards vegetarianism is not explicitly formulated into account, one can try to understand the result intuitively . For individuals, who have vegetarians or vegans in their surrounding, the odds will be by some factor larger. When trying to align this result with some intuition, one can try to refer to a famous Ukrainian proverb: "Tell me who your friend is and I will tell you who you are". This means that those people, who surround us, strongly affect our perceptions and beliefs. As it turned out, the topic of vegetarianism is no exception. Also, when talking about empathic individuals, the odds will be by some factor larger. It is quite intuitive, since those people, who are empathic, will be shocked more after, for example, watching a video from some slaughter

plant and thus will have bigger chances to become a vegetarian or even vegan. Finally, when it comes to those people, who are sorting trash , the odds will also be by some factor larger. One can try to explain it in the following way: those people by sorting trash are trying to take care of our planet. Since, as it was stated in the beginning, a third of all emissions on our planet are caused by food production, those people have bigger chances to stop eating meat, because by doing so they will try to reduce those emissions and, as a result, save our planet.

One of the main limitations of this research is, of course, data. The lack of such on the Internet led to the decision to collect it independently, which is quite a difficult task. As a result, 194 responses were collected. This quantity is enough to start making some conclusions, but to develop this project in the future more data maybe even of individuals of all ages and from all regions of Ukraine need to be collected.