

Исследование пищевых привычек студентов университета

Цель: на основе результатов тестирования о пищевых привычках студентов очистить и подготовить данные для последующего анализа.

Исходные данные: dataset Food preferences (Food choices) students of Mercyhurst University, источник:

https://www.kaggle.com/datasets/borapajo/food-choices/data?select=food_coded.csv

Описание набора данных: содержит информацию о предпочтениях продуктов, пищевых привычках в детстве и в настоящее время 126 студентов университета.

Исходные данные в табличном формате:

https://docs.google.com/spreadsheets/d/1kx7klKPtmaLdlLoFCnxeXT9v0wgHgRPTa16DkyFe_1E/edit?usp=sharing

файл с неочищенными данными Food preferences

файл с очищенными данными Food preferences cleaned

Порядок подготовки и очистки данных

1. **Этап предварительной очистки:** удаление лишних пробелов в ячейках с текстовым форматом, выбор правильного формата для столбцов, поиск пустых значений и замена их на средние значения (при возможности), унификация регистра текста в ячейках для упрощения анализа информации.

#Change log

Version 1.0

Changes

Данные-Очистка-Удалить пробелы

Удалены пробелы из текста столбцы H, I, N, Q, AC, AI, AJ, AQ, AA, AB

##Fix

Fix data type in columns A-G, J-M, O, P, R-X and all other columns with numbers from text to number. (*function VALUE*)

##Changes

Found empty cell in column A with data about GPA, decided to replace them with average meaning in that column (*instrument conditional formatting and filtering*)

##Changes

Column E “Calories per day”, found empty cells replace them by meaning 2 - it is not at all important. It is understandable that a person doesn't answer the question because he doesn't consider it important. (*instrument conditional formatting and filtering*)

##Changes

Column F, replace the meaning 'nan' with meaning 0.

##Changes

Made spellcheck of columns H,I, N, Q, Y, Z, AC, AI, AJ, AQ, AS, BE (*Инструменты-проверка правописания*)

#Changes

Found empty cells J108-J126 replace 9 - none, the question was - comfort_food_reasons_coded, so if student didn't answer it means that there are no reasons for him (*instrument conditional formatting and filtering*)

#Changes

Column U replaced empty cells with answer 4 (other), the question was 'do you work.

#Check duplicate

Duplicate didn't found

Changes that need manipulation with columns

##Changes

the capitalization of letters in columns to lowercase in columns H,I, Y,AC,AI, AJ,AQ, AS, BE

#Changes

Changed the capitalization of the first letter in the cell (the answers are written like sentences) in the column Q, AA with next formula =ЗАМЕНИТЬ(R3,1,1, ПРОПИСН(ЛЕВСИМВ(R3,1)))
(*R because the copy of the column*)

2. **Этап структурирования данных** для упрощения анализа: добавлена строка с номером вопроса для быстрой идентификации. Далее классифицированы вопросы по тематике и типам ответов (численное значение, варианты ответов, короткий текстовый ответ, развернутый текстовый ответ). Разделение вопросов на группы по тематика позволит проанализировать различные аспекты пищевых привычек студентов, а разделение в зависимости от типа ответа - применить одинаковые методы для их анализа.

Предложены группы ответов по тематике:

1. общие параметры студентов (балл, пол, характеристики родителей, доход, работа)
2. суточные калории
3. характеристики диеты и рациона (калории, кухня, тип диеты, количество фруктов и овощей в рационе, любимая еда)

4. пищевые привычки (еда вне дома, причины поесть, самостоятельное приготовление пищи, отслеживание количества поступающих микронутриентов)
5. выбор еды из предложенной в вариантах ответов
6. здоровье и стиль жизни (активность, уровень счастья, витамины)
7. развернутый текстовый ответ)

#Change log

##New

Added the row 'question number' on the sheet 'food_coded' in order to find the text of the exact question faster

##New

Created sheet "Question_type"

Divided all questions to 7 groups in order to simplify analysis

1. general parameter (GPA, gender, mother and father characteristics, income, employment, grade level)
2. calories guessing
3. diet characteristics (calories, cuisine, type of diet, number of fruits and veggies in the diet, favorite food)
4. eating habits (meals out, reasons for food, who and when cooks, nutritional check)
5. meal choice
6. health and lifestyle (sporting activities, happiness level, weight, vitamins)
7. open ended

##New

Added the addicted row on the sheet food_coded - type of question. Returned the value of question type for every question. Used the function the VLOOKUP

##New

created new sheets to locate there separately answers to question of every type, transposed the data from the sheet food_coded and filtered needed questions

3. Этап формулирования вопросов и выявления закономерностей. Ниже приведены возможные направления исследования данных с указанием в скобках номеров вопросов, которые используются в анализе. Как пример дальнейшего анализа данных, в файле с очищенными данными лист 'Eating habits' построена круговая диаграмма причин приемов пищи студентами.

Анализ открытых вопросов реализован wordcloud in Tableau. Это позволило увидеть наиболее популярные ответы и все ответы в целом. В Google применила функцию SPLIT для того, чтобы поместить отдельные слова ответов в отдельные ячейки. Далее в EXCEL (не нашла такую функцию в Google) с помощью функции "Найти и Заменить" заменила запятые на пробелы (запятые мешали анализировать отдельные слова). С помощью функции СЧЕТЕСЛИ определила частоту употребления слов.

boredom usually sadness and cravings bored, stress stress, bored, anger
 we don't have comfort boredom, laziness, anger, sadness, stress, cold weather happiness, satisfaction
 sadness, depression boredom, stressed, sad depression, comfort, accessibility stress, anger and sadness happiness, hunger, sadness
 boredom, sadness, anxiety stress, boredom, cold weather hormones, premenstrual syndrome. boredom, anger, drunkenness sadness, stressed, boredom
 boredom, sadness, hungry boredom, stress tired boredom and stress boredom, sadness, and if it has a good taste. loneliness, homework, boredom boredom, sadness and anger
 friends, environment and boredom just cause boredom, anger and just being hungry in general. all of the above; sadness, boredom and confusion boredom, happiness, sadness
 survival, bored stress, boredom and physical activity none, i don't eat comfort food. i just eat when i'm hungry. stressed, upset, or just craving a cheat meal boredom. celebration.
 no reasons i do not really eat "comfort food" but i guess sadness, special occasions, and anxiety stress and sadness they taste better than other food. they are a pick me up. they are easy to make
 anger, sadness they are yummy, my boyfriend sometimes makes me sad, boredom i usually only eat comfort food when i'm bored, if i am doing something, i can go for hours without eating
 lazy boredom, being in your period, and long bus rides for softball i eat comfort food when i'm stressed out from school (finals week), when im sad, or when i am dealing with personal family issues.
 stress, boredom, cravings i usually only eat comfort foods when i am bored. i will also eat them when i am happy to celebrate and then when i am sad to comfort me. stress, boredom none
 boredom, sadness i am always stressed out, and bored when i am in my apartment. sadness boredom, stress, and it tastes good
 stress, boredom, and nighttime when i'm eating with my close friends/ food smell or look good/ when i feel tired boredom, sadness, or with friends stress
 boredom, hunger, snacking. stress, sadness, boredom anxiousness, watching tv i desire "comfort food" stress and boredom sadness, stress usually if i'm sad or depressed.
 sadness, boredom, & anger boredom, happiness, distraught happiness, they are some of my favorite foods stress, boredom, special occasions sadness, happiness and hunger stress, boredom, craving
 stress, sadness, bored sadness, boredom, late night snack boredom and sadness a long day, not feeling well, winter boredom, stress, mood swings sad, bored, excited
 hunger and boredom nan stress, anger and boredom stress, boredom, college as whole happiness, boredom, social event sadness, loneliness, boredom
 stress, sadness boredom, anger, happy sadness, happiness and boredom stress, boredom, sadness boredom comfort hunger boredom, laziness
 hunger, boredom when i am sad or craving bad day, bored, sadness boredom!, sadness laziness and hungover
 mostly boredom, mostly stress

Exploratory Data Analysis questions

- about the current diet and popular food
1. What does the ideal diet look like on behalf of the students?
 the pie chart diagram (36)
 word cloud of healthy meal and ideal diet (34, 35)
 2. What are the cuisine preferences of students? (25, 26, 32, 38, 39, 49) 21
 3. What is the most popular (comfort) food (top 10) or meals among the students? (13, 8)
 4. Are the favorite foods of the students cooked by them or bought? (27,)
 5. How do the students estimate their own diet, as healthy or not? (14)
 Did such students have any general parameters? Who has healthy diet (who more go out, cook, parents cook)? Connection with sport, income, cooking or going out, life rewarding?
 6. How important is it for students to have fruits and vegetables in everyday diet? (30, 57)
- about the diet changes during the life

in childhood, after entering the college

What cuisine did the most part of students eat growing up? 12

Top favorite food of students in childhood

How do the students describe their eating changes after entering college?

Top 5 changes 17,16

- about eating habits

9,10

What percentage of students eat comfort food only when they are hungry or don't have an exact reason to eat?

What is the most popular reason that makes students eat comfort food? Is it negative and connected with a low level of happiness?

Is there any correlation between negative food reasons and sporting activities?

11, 19,27

What is more popular among the students to cook or to go out? With 20

Who are more likely to go out students who work or who not?47 What part of students who prefer home food eat food cooked by parents?47

What part of their income students spend on going out? 48

31,37 Is there any correlation of grade level of the student, his income and the frequency of going out?

41 Who is more likely to go out/cook single students or married ones?

46 Is there any correlation of living situation and going out?

- about the health and the happiness

50

What percentage of students have problems/don't have problems with weight perception?

40

What is the average estimation of life rewarding among the students?

33

What is the average estimation of health among the students?

Is there any correlation of income and health estimation of students?

22,52

What percentage of student have active life (exercise more than twice a week) and don't have exercises?

56 Top 5 sports

Are students who have sporting activities more likely to feel healthy/happy?

Are students who have sporting activities have problems with weight perception?

What part of students take supplements or vitamins?

- about food consciousness

What percentage of students choose the right answers in more than 80% of questions with guessing the calories? (4,6 ,54,55, 59)

What percentage of students consider that knowing the day calories is important? (5)

Is there any correlation of successfull calories choice and knowing the day calories?

What percentage of students check nutritional values? What is the most popular frequency? (45)

According to general principles of healthy diet meal choice questions to my mind are not an informative part of the survey. The main reason for this is that junk food in some volume is not dangerous and such questions don't give us information on how often students prefer junk food. 3,7, 15, 29,51

In this question we can only count the sum of the calories each student has chosen.

Don't know how to use information about parents, GPA (happyness?)

Weight in general doesn't give us any information about the health of students without parameters of high, percentage of fat in the body.