Exploring the Impact of Personality on Food Preferences

Alice Thwaites, Ava Sawers, Alisha Sajjad, Jenessa Entz, and Jaiveer Toor

Department of Computer Science, University of Calgary

DATA 201: Thinking with Data

Dr. Nelson Wong

June 16th, 2025

Section A: Introduction

Topic and Main Goal of Study

Our project focuses on personality types and individuals' preferences for food flavours. Through the exploration of this topic, we aim to identify potential correlations between a preference for specific food flavours, such as spicy, sweet, or salty, and whether a person identifies as an introvert, extrovert, or ambivert. We have also examined the extent to which food flavour preferences and other personality traits may be related, such as impulsivity and sensitivity. We selected this topic out of curiosity about whether our own flavour profiles might reflect something deeper about our personalities, specifically regarding extroversion and introversion. Additionally, much of the existing literature lacks information on how food flavour preferences specifically relate to introversion, extroversion, or ambiversion. This gap in the research inspired us to investigate what insights our collected data might provide.

To collect data for this project, we used Qualtrics to create and distribute a survey. The survey consists of 11 questions designed to gather meaningful information that will help us determine potential correlations between food flavour preferences and personality traits. Some of the questions focused on evaluating personality traits, while others assessed participants' levels of preference for various food flavours.

Background Research

Individual preferences for food flavours, including sweet, sour, salty, savory (umami), bitter, and spicy, vary significantly from person to person. Some individuals strongly prefer certain flavours or are more sensitive to them than others. This variation prompts the question of whether personality traits contribute to these differences in flavour preferences. Specifically, our study aims to explore whether distinct flavour preference patterns are associated with extroversion and introversion.

Existing research indicates that specific personality traits are linked to preferences for certain flavours. For instance, individuals who score high in novelty seeking tend to prefer salty foods, while those high in reward dependence often prefer sweet foods (Day et al., 2008). Both novelty seeking and reward dependence have been found to correlate positively with extroversion (De Fruyt et al., 2000; Gocłowska et al., 2019), suggesting that extroverted individuals may have a greater liking for salty and sweet foods. Similarly, a preference for spicy foods has been associated with higher levels of sensation seeking and sensitivity to reward (Byrnes & Hayes, 2013). These traits are also positively associated with extroversion (Aluja et al., 2003; Blain et al., 2021). Furthermore, individuals with high levels of anxiety tend to report more food aversions and are often considered picky eaters (Smith et al., 1955).

Additionally, individuals who prefer bitter foods are more likely to exhibit antisocial tendencies, including increased aggression (Wu, 2023). In contrast, people who prefer sweet foods often score higher on the agreeableness dimension of personality (Wu, 2023). A contributing factor in the relationship between personality and flavour preferences is

neurotransmitters (Spence, 2022). For example, higher levels of serotonin, a neurotransmitter that affects happiness and anxiety, have been shown to increase a person's sensitivity to both bitterness and sweetness, meaning they would perceive these flavours as more intense (Scaccia, 2024; Spence, 2022).

These findings suggest that personality traits may play a meaningful role in shaping individual flavour preferences. Further research into this connection can enhance our understanding of the complex links between personality and dietary choices.

Section B: Obtaining Data

Recruitment Message

Subject: Invitation to Participate in Food Preferences and Personality Survey

Hello [Recipient Name],

You are receiving this email to invite you to participate in a research survey conducted by students at the University of Calgary. The purpose of this survey is to gather data on individuals' perspectives regarding whether their preferred food flavours (sweet, salty, savoury, spicy, sour, bitter) may correlate with personality types (introvert, extrovert, or ambivert), and to explore the nature of those correlations. We are seeking responses from individuals of all ages to ensure a diverse data set.

Your participation in this survey is entirely voluntary and completely anonymous. This survey consists of 11 short questions that will help us determine your perspective on the impact of technology becoming a main learning resource in schools, as well as gather intel securely and safely. It should only take a few minutes to complete.

If you have any questions, please feel free to contact our team using any of the following email addresses: ava.sawers@ucalgary.ca, jenessa.entz@ucalgary.ca, alice.thwaites@ucalgary.ca, alisha.sajjad@ucalgary.ca, and jaiveer.toor@ucalgary.ca.

We would be happy to assist you and answer any questions.

You can review the consent form here: DATA 201 Consent Form.docx

Thank you for considering our invitation to participate!

Sincerely,

Ava Sawers, Alisha Sajjad, Jenessa Entz, Jaiveer Toor, and Alice Thwaites

UNIVERSITY OF CALGARY

Name of Researcher, Faculty, Department, Telephone & Email:

Jenessa Entz, Faculty of Arts, Department of Psychology, 587-834-7226, jenessa.entz@ucalgary.ca

Alisha Sajjad, Faculty of Arts, Department of Sociology, 587-839-9144, alisha.sajjad@ucalgary.ca

Ava Sawers, Faculty of Arts, Department of Anthropology, 403-478-5797, ava.sawers@ucalgary.ca

Alice Thwaites, Faculty of Arts, Department of Psychology, 587-224-2707, alice.thwaites@ucalgary.ca

Jaiveer Toor, Faculty of Science, Department of Computer Science, 825-712-8384, jaiveer.toor@ucalgary.ca

Supervisor:
Dr. Nelson Wong

Title of Project: Exploring the Impact of Personality on Food Preferences

Sponsor:
N/A

This consent form, a copy of which has been given to you, is only part of the process of informed consent. If you want more details about something mentioned here, or information not included here, you should feel free to ask. Please take the time to read this carefully and to understand any accompanying information.

The University of Calgary Conjoint Faculties Research Ethics Board has approved this research study

Participation is completely voluntary, anonymous, and confidential.

Purpose of the Study

The purpose of this survey is to determine if there is any correlation between personality types (extrovert, introvert, ambivert) and different flavours of food (sweet, salty, sour, spicy, umami, bitter).

What Will I Be Asked To Do?

You are asked to complete an online survey for an assignment in an undergraduate university course. The survey is designed to be completed in approximately 5 minutes. The survey is anonymous and does not contain questions asking for specific identifying information.

Sample question from the survey: Rate the extent to which you like salty foods.

Participation is completely voluntary. You may refuse to participate altogether, may refuse to participate in parts of the survey, may decline to answer any and all questions, and may withdraw from the survey at any time.

What Type of Personal Information Will Be Collected

No personal identifying information will be collected in this study, and all participants shall remain anonymous. Should you agree to participate, you will be asked to provide your personality type.

Are there Risks or Benefits if I Participate?

There are no foreseeable risks, harms, or inconveniences to the participant. By participating in the survey, you are helping university undergraduate students to complete a course assignment.

What Happens to the Information I Provide?

Only the students who distribute the survey, as well as the instructor and teaching assistants of the course will have access to the information collected.

Withdrawal of data is not possible after data collection, as all collected data is anonymous and identifiable data cannot be retrieved. All data collected will be destroyed by the end of the course.

No one except the students who have distributed the survey, the instructor, and the teaching assistants will be allowed to see any of the answers to the survey. There are no names on the survey. Only group information will be summarized for any presentation or publication of results. The survey data will be deleted by the end of the course.

Signatures

Your response below indicates that 1) you understand to your satisfaction the information provided to you about your participation in this survey, 2) you agree to participate in the survey, and 3) you are 18 years of age or older.

In no way does this waive your legal rights nor release the investigators, sponsors, or involved institutions from their legal and professional responsibilities. You are free to withdraw from this survey at any time. You should feel free to ask for clarification or new information throughout your participation.

I consent to participate in this survey -> https://survey.ucalgary.ca/jfe/form/SV_1z5w53GpnCzhVHg

I do not wish to participate in the survey -> https://ucalgary.ca

Questions/Concerns

If you have any further questions or want clarification regarding this survey and/or your participation, please contact:

Jenessa Entz, jenessa.entz@ucalgary.ca

Alisha Sajjad, alisha.sajjad@ucalgary.ca

Ava Sawers, ava.sawers@ucalgary.ca

Alice Thwaites, alice.thwaites@ucalgary.ca

Jaiveer Toor, jaiveer.toor@ucalgary.ca

If you have any concerns about the way you've been treated as a participant, please contact the Research Ethics Analyst, Research Services Office, University of Calgary at 403.220.6289 or 403.220.8640; email cfreb@ucalgary.ca.

Survey Explanation

Our goal was to explore whether personality traits, particularly extroversion, introversion, and ambiversion, correlate with preferences for specific food flavours including spicy, sweet, salty, sour, savoury (umami), and bitter. The survey data we collected plays a critical role in helping us analyze whether individuals with different personality characteristics (e.g., extroversion, introversion, sensitivity, impulsivity) are statistically more likely to prefer particular flavours. This directly supports our main research goal which is to identify meaningful patterns between food preferences and personality types.

Target Participants

We targeted a broad range of participants, primarily university students and adults within our social and academic circles. These individuals were chosen because they represent a diverse and accessible population familiar with self-report surveys. This approach enabled us to gather responses from people with varying personality traits and flavour preferences, helping to increase the representativeness and validity of our findings. The survey, created and distributed through Qualtrics, was completely anonymous, ensuring honest responses and ethical data collection. All participation was voluntary and privacy was maintained throughout. Our survey had a total of 53 responses ($n = 53$).

Survey Questions

Name of Researcher, Faculty, Department, Telephone & Email:

Jenessa Entz, Faculty of Arts, Department of Psychology, 587-834-7226, jenessa.entz@ucalgary.ca

Alisha Sajjad, Faculty of Arts, Department of Sociology, 587-839-9144, alisha.sajjad@ucalgary.ca

Ava Sawers, Faculty of Arts, Department of Anthropology, 403-478-5797, ava.sawers@ucalgary.ca

Alice Thwaites, Faculty of Arts, Department of Psychology, 587-224-2707, alice.thwaites@ucalgary.ca

Jaiveer Toor, Faculty of Science, Department of Computer Science, 825-712-8384, jaiveer.toor@ucalgary.ca

Supervisor:
Dr. Nelson Wong

Title of Project: Exploring the Impact of Personality on Food Preferences

Sponsor:
N/A

This consent form, a copy of which has been given to you, is only part of the process of informed consent. If you want more details about something mentioned here, or information not included here, you should feel free to ask. Please take the time to read this carefully and to understand any accompanying information.

The University of Calgary Conjoint Faculties Research Ethics Board has approved this research study.

Participation is completely voluntary, anonymous, and confidential.

Purpose of the Study

The purpose of this survey is to determine if there is any correlation between personality types (extrovert, introvert, ambivert) and different flavours of food (sweet, salty, sour, spicy, umami, bitter).

What Will I Be Asked To Do?

You are asked to complete an online survey for an assignment in an undergraduate university course. The survey is designed to be completed in approximately 5 minutes. The survey is anonymous and does not contain questions asking for specific identifying information.

Sample question from the survey: Rate the extent to which you like salty foods.

Participation is completely voluntary. You may refuse to participate altogether, may refuse to participate in parts of the survey, may decline to answer any and all questions, and may withdraw from the survey at any time.

What Type of Personal Information Will Be Collected

No personal identifying information will be collected in this study, and all participants shall remain anonymous. Should you agree to participate, you will be asked to provide your personality type.

Are there Risks or Benefits if I Participate?

There are no foreseeable risks, harms, or inconveniences to the participant. By participating in the survey, you are helping university undergraduate students to complete a course assignment.

What Happens to the Information I Provide?

Only the students who distribute the survey, as well as the instructor and teaching assistants of the course will have access to the information collected.

Withdrawal of data is not possible after data collection, as all collected data is anonymous and identifiable data cannot be retrieved. All data collected will be destroyed by the end of the course.

No one except the students who have distributed the survey, the instructor, and the teaching assistants will be allowed to see any of the answers to the survey. There are no names on the survey. Only group information will be summarized for any presentation or publication of results. The survey data will be deleted by the end of the course.

Signatures

Your response below indicates that 1) you understand to your satisfaction the information provided to you about your participation in this survey, 2) you agree to participate in the survey, and 3) you are 18 years of age or older.

In no way does this waive your legal rights nor release the investigators, sponsors, or involved institutions from their legal and professional responsibilities. You are free to withdraw from this survey at any time. You should feel free to ask for clarification or new information throughout your participation.

I consent to participate in this survey -> https://survey.ucalgary.ca/jfe/form/SV_1z5w53GpnCzhVHg

I do not wish to participate in the survey -> https://ucalgary.ca

Questions/Concerns

If you have any further questions or want clarification regarding this survey and/or your participation, please contact:

Jenessa Entz, jenessa.entz@ucalgary.ca

Alisha Sajjad, alisha.sajjad@ucalgary.ca

Ava Sawers, ava.sawers@ucalgary.ca

Alice Thwaites, alice.thwaites@ucalgary.ca

Jaiveer Toor, jaiveer.toor@ucalgary.ca

If you have any concerns about the way you've been treated as a participant, please contact the Research Ethics Analyst, Research Services Office, University of Calgary at 403.220.6289 or 403.220.8640; email cfreb@ucalgary.ca.

Do you wish to participate in this survey?

- ◯ Yes, I consent to participate in this survey
- ◯ No, I do not consent to participate in this survey

People often describe extroverts as recharged by social time, and introverts as recharged by alone time. I would consider myself:

- ◯ Mostly extroverted
- ◯ Mostly introverted
- ◯ An equal mixture of both
- ◯ I don't know

To what extent do you agree with the following statements.

	Strongly agree	Somewhat agree	Neither agree nor disagree	Somewhat disagree	Strongly disagree
I feel energized when surrounded by people.	◯	◯	◯	◯	◯
I thrive in large teams and crowds.	◯	◯	◯	◯	◯
I prefer solitary activities over social ones.	◯	◯	◯	◯	◯
I am shy and reserved in social settings.	◯	◯	◯	◯	◯

Rate how well each of the following personality traits describes you.

	Does not describe me	Describes me slightly well	Describes me moderately well	Describes me very well	Describes me extremely well
Introspective	○	○	○	○	○
Talkative	○	○	○	○	○
Carefree	○	○	○	○	○
Reserved	○	○	○	○	○
Sensitive	○	○	○	○	○
Impulsive	○	○	○	○	○

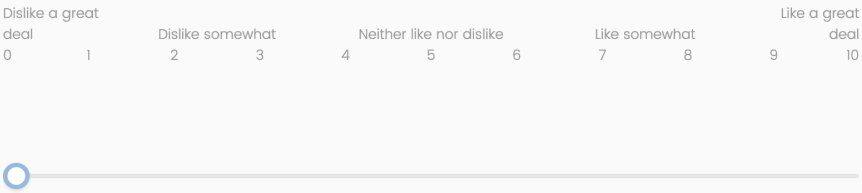
How much do you agree with the following statement: "I am a picky eater."

- ○ Strongly disagree
- ○ Somewhat disagree
- ○ Neither agree nor disagree
- ○ Somewhat agree
- ○ Strongly agree

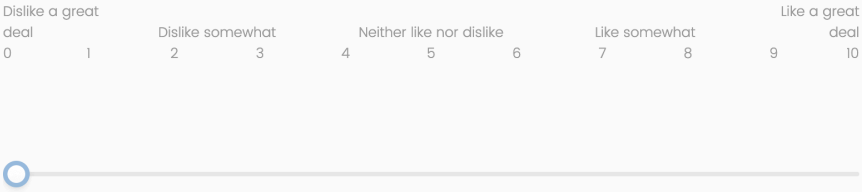
Which food flavor do you tend to eat most often?

- ○ Sweet
- ○ Spicy
- ○ Bitter
- ○ Sour
- ○ Savoury (umami)
- ○ Salty

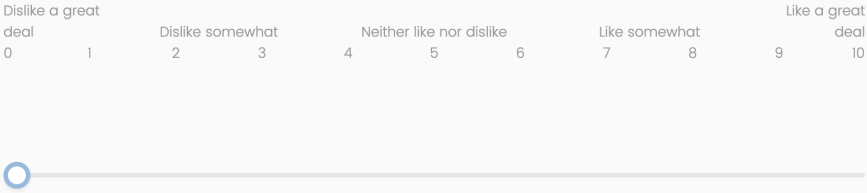
Rate the extent to which you like salty foods.



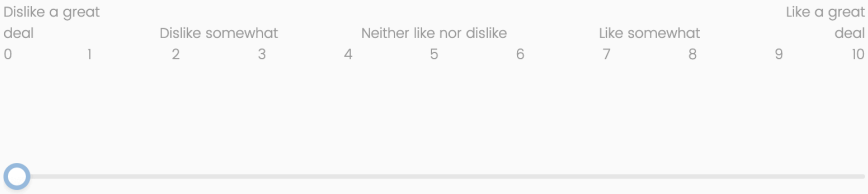
Rate the extent to which you like sweet foods.

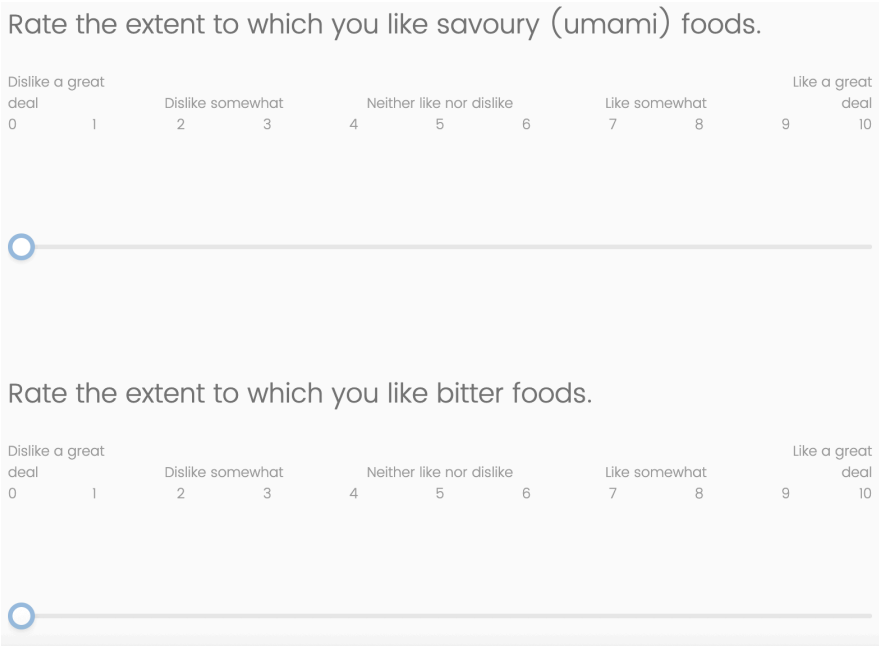


Rate the extent to which you like sour foods.



Rate the extent to which you like spicy foods.





We thank you for your time spent taking this survey.
Your response has been recorded.

Section C: Analysis Questions

Analysis Question 1
What food flavours are preferred by individuals who identify as highly sensitive (i.e., those who selected “describes me very well” or “describes me extremely well”), and how do these flavour preferences compare in terms of likeability?

Analysis Question 2
Do more intense flavours (like spicy, bitter, etc.) tend to be associated with certain personality types versus less intense flavours (like savoury or sweet), which may be associated with other personality types?

Analysis Question 3
To what extent do individuals who describe themselves as carefree or impulsive prefer strong and unconventional flavours (e.g. spicy, bitter, sour, salty) over less bold ones (e.g. sweet, savoury)?

Analysis Question 4
Do individuals’ statements about being extroverted, introverted, or in between align with their answers about their personality characteristics?

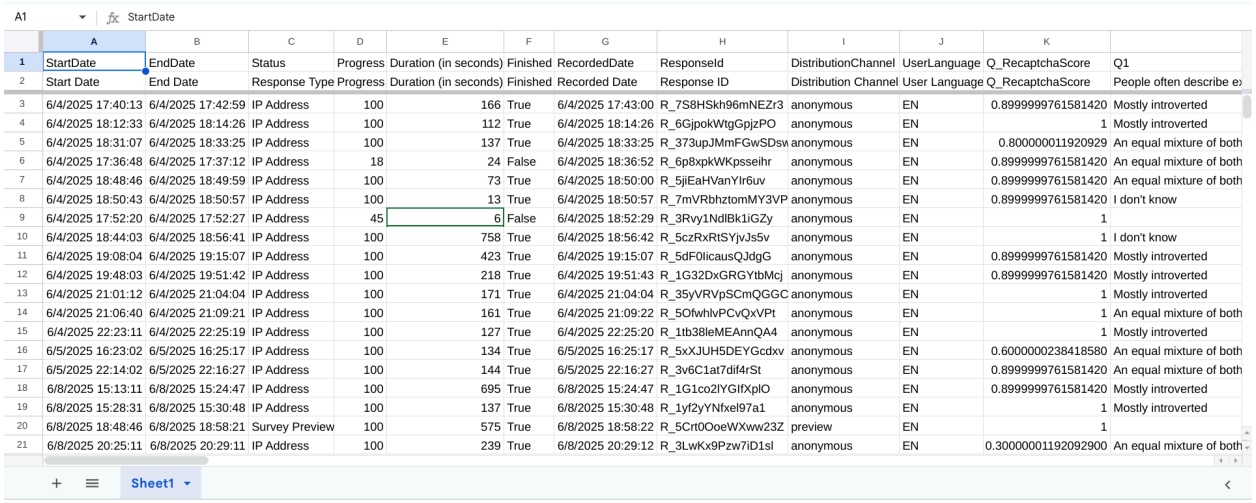
Analysis Question 5
Is there a noticeable difference in flavour preferences between individuals who identify as introverts versus those who identify as ambiverts?

Section D: Cleaning Data

Dirty Attribute 1: "Finished" Column

The "finished" column contained dirty data because some of the responses showed that the survey was not completed, which was visible through the responses labelled "false." This data is dirty because with incomplete surveys, we lack responses for certain questions and therefore lack data to analyze. To clean this data, we removed the two responses that were labelled "false" in the "finished" column. This made the data easier to read and ensured that the remaining data in this column contained only valid responses where all of the questions were answered and the survey was completed. With this data now being clean, it is easier for us to read and analyze, as well as ensure that we don't have incomplete responses in the columns for designated questions.

Dirty Attribute 1 (Before Cleaning)

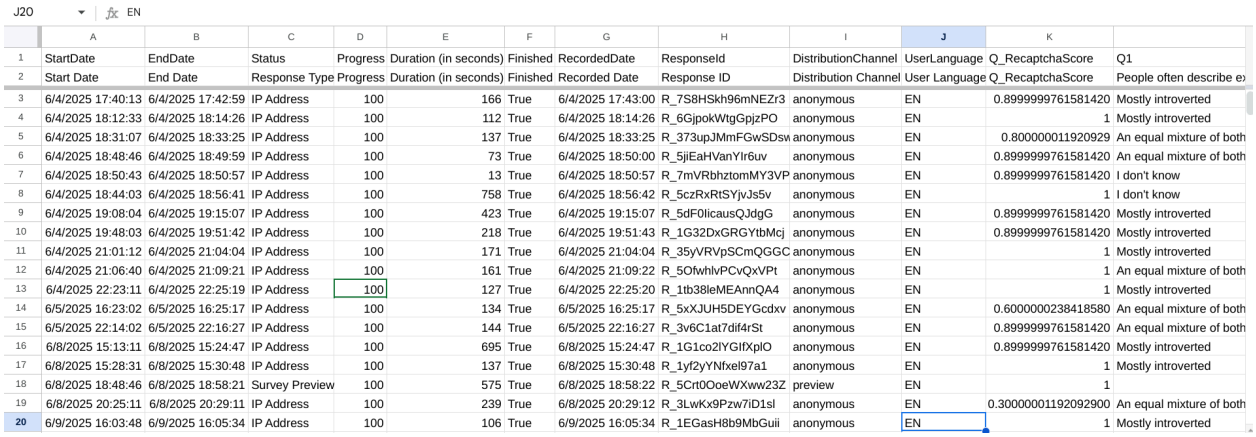


A1 | fx StartDate

	A	B	C	D	E	F	G	H	I	J	K	
1	StartDate	EndDate	Status	Progress	Duration (in seconds)	Finished	RecordedDate	ResponseId	DistributionChannel	UserLanguage	Q_RecaptchaScore	Q1
2	Start Date	End Date	Response Type	Progress	Duration (in seconds)	Finished	Recorded Date	Response ID	Distribution Channel	User Language	Q_RecaptchaScore	People often describe ex
3	6/4/2025 17:40:13	6/4/2025 17:42:59	IP Address	100	166	True	6/4/2025 17:43:00	R_7S8HSkh96mNEZr3	anonymous	EN	0.8999999761581420	Mostly introverted
4	6/4/2025 18:12:33	6/4/2025 18:14:26	IP Address	100	112	True	6/4/2025 18:14:26	R_6GjpokWtgGpjzPO	anonymous	EN	1	Mostly introverted
5	6/4/2025 18:31:07	6/4/2025 18:33:25	IP Address	100	137	True	6/4/2025 18:33:25	R_373upJMmFGwSDsw	anonymous	EN	0.800000011920929	An equal mixture of both
6	6/4/2025 17:36:48	6/4/2025 17:37:12	IP Address	18	24	False	6/4/2025 18:36:52	R_6p8xpkWKpsseihr	anonymous	EN	0.8999999761581420	An equal mixture of both
7	6/4/2025 18:48:46	6/4/2025 18:49:59	IP Address	100	73	True	6/4/2025 18:50:00	R_5jiEaHVanYlr6uv	anonymous	EN	0.8999999761581420	An equal mixture of both
8	6/4/2025 18:50:43	6/4/2025 18:50:57	IP Address	100	13	True	6/4/2025 18:50:57	R_7mVRbhztomMY3VP	anonymous	EN	0.8999999761581420	I don't know
9	6/4/2025 17:52:20	6/4/2025 17:52:27	IP Address	45	6	False	6/4/2025 18:52:29	R_3Rvy1NdlBk1iGZy	anonymous	EN	1	
10	6/4/2025 18:44:03	6/4/2025 18:56:41	IP Address	100	758	True	6/4/2025 18:56:42	R_5czRxRtSYjvJs5v	anonymous	EN	1	I don't know
11	6/4/2025 19:08:04	6/4/2025 19:15:07	IP Address	100	423	True	6/4/2025 19:15:07	R_5dF0licausQJdgG	anonymous	EN	0.8999999761581420	Mostly introverted
12	6/4/2025 19:48:03	6/4/2025 19:51:42	IP Address	100	218	True	6/4/2025 19:51:43	R_1G32DxGRGYtbMcj	anonymous	EN	0.8999999761581420	Mostly introverted
13	6/4/2025 21:01:12	6/4/2025 21:04:04	IP Address	100	171	True	6/4/2025 21:04:04	R_35yVRVpSCmQGGC	anonymous	EN	1	Mostly introverted
14	6/4/2025 21:06:40	6/4/2025 21:09:21	IP Address	100	161	True	6/4/2025 21:09:22	R_5OfwhlvPCvQxVPt	anonymous	EN	1	An equal mixture of both
15	6/4/2025 22:23:11	6/4/2025 22:25:19	IP Address	100	127	True	6/4/2025 22:25:20	R_1tb38leMEAnnQA4	anonymous	EN	1	Mostly introverted
16	6/5/2025 16:23:02	6/5/2025 16:25:17	IP Address	100	134	True	6/5/2025 16:25:17	R_5xXJUH5DEYGcdxv	anonymous	EN	0.6000000238418580	An equal mixture of both
17	6/5/2025 22:14:02	6/5/2025 22:16:27	IP Address	100	144	True	6/5/2025 22:16:27	R_3v6C1at7dif4rSt	anonymous	EN	0.8999999761581420	An equal mixture of both
18	6/8/2025 15:13:11	6/8/2025 15:24:47	IP Address	100	695	True	6/8/2025 15:24:47	R_1G1co2lYGIfXplO	anonymous	EN	0.8999999761581420	Mostly introverted
19	6/8/2025 15:28:31	6/8/2025 15:30:48	IP Address	100	137	True	6/8/2025 15:30:48	R_1yf2yYNfxel97a1	anonymous	EN	1	Mostly introverted
20	6/8/2025 18:48:46	6/8/2025 18:58:21	Survey Preview	100	575	True	6/8/2025 18:58:22	R_5Crt0OoeWXww23Z	preview	EN	1	
21	6/8/2025 20:25:11	6/8/2025 20:29:11	IP Address	100	239	True	6/8/2025 20:29:12	R_3LwKx9Pzw7iD1sl	anonymous	EN	0.30000001192092900	An equal mixture of both

Sheet1

Dirty Attribute 1 (After Cleaning)



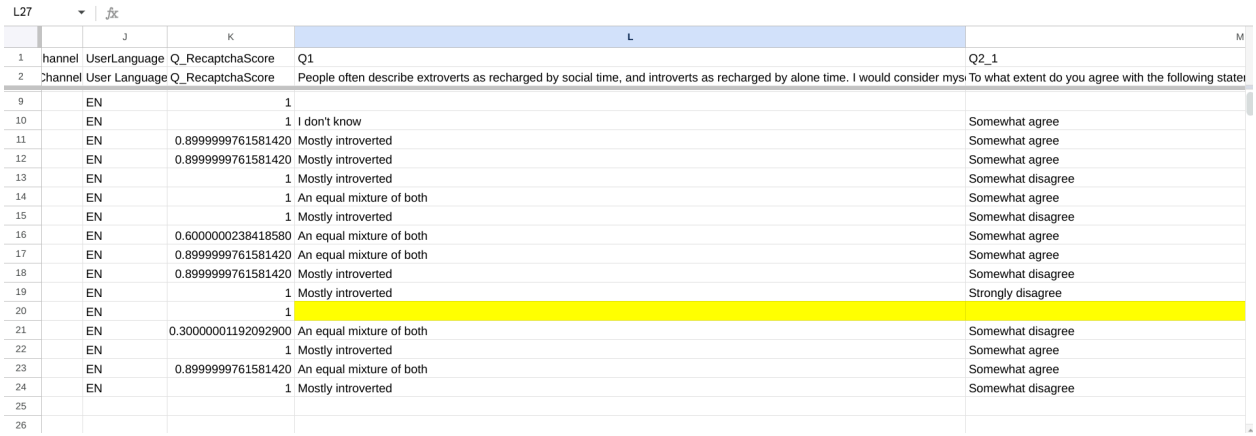
J20 | fx EN

	A	B	C	D	E	F	G	H	I	J	K	
1	StartDate	EndDate	Status	Progress	Duration (in seconds)	Finished	RecordedDate	ResponseId	DistributionChannel	UserLanguage	Q_RecaptchaScore	Q1
2	Start Date	End Date	Response Type	Progress	Duration (in seconds)	Finished	Recorded Date	Response ID	Distribution Channel	User Language	Q_RecaptchaScore	People often describe ex
3	6/4/2025 17:40:13	6/4/2025 17:42:59	IP Address	100	166	True	6/4/2025 17:43:00	R_7S8HSkh96mNEZr3	anonymous	EN	0.8999999761581420	Mostly introverted
4	6/4/2025 18:12:33	6/4/2025 18:14:26	IP Address	100	112	True	6/4/2025 18:14:26	R_6GjpokWtgGpjzPO	anonymous	EN	1	Mostly introverted
5	6/4/2025 18:31:07	6/4/2025 18:33:25	IP Address	100	137	True	6/4/2025 18:33:25	R_373upJMmFGwSDsw	anonymous	EN	0.800000011920929	An equal mixture of both
6	6/4/2025 18:48:46	6/4/2025 18:49:59	IP Address	100	73	True	6/4/2025 18:50:00	R_5jiEaHVanYlr6uv	anonymous	EN	0.8999999761581420	An equal mixture of both
7	6/4/2025 18:50:43	6/4/2025 18:50:57	IP Address	100	13	True	6/4/2025 18:50:57	R_7mVRbhztomMY3VP	anonymous	EN	0.8999999761581420	I don't know
8	6/4/2025 18:44:03	6/4/2025 18:56:41	IP Address	100	758	True	6/4/2025 18:56:42	R_5czRxRtSYjvJs5v	anonymous	EN	1	I don't know
9	6/4/2025 19:08:04	6/4/2025 19:15:07	IP Address	100	423	True	6/4/2025 19:15:07	R_5dF0licausQJdgG	anonymous	EN	0.8999999761581420	Mostly introverted
10	6/4/2025 19:48:03	6/4/2025 19:51:42	IP Address	100	218	True	6/4/2025 19:51:43	R_1G32DxGRGYtbMcj	anonymous	EN	0.8999999761581420	Mostly introverted
11	6/4/2025 21:01:12	6/4/2025 21:04:04	IP Address	100	171	True	6/4/2025 21:04:04	R_35yVRVpSCmQGGC	anonymous	EN	1	Mostly introverted
12	6/4/2025 21:06:40	6/4/2025 21:09:21	IP Address	100	161	True	6/4/2025 21:09:22	R_5OfwhlvPCvQxVPt	anonymous	EN	1	An equal mixture of both
13	6/4/2025 22:23:11	6/4/2025 22:25:19	IP Address	100	127	True	6/4/2025 22:25:20	R_1tb38leMEAnnQA4	anonymous	EN	1	Mostly introverted
14	6/5/2025 16:23:02	6/5/2025 16:25:17	IP Address	100	134	True	6/5/2025 16:25:17	R_5xXJUH5DEYGcdxv	anonymous	EN	0.6000000238418580	An equal mixture of both
15	6/5/2025 22:14:02	6/5/2025 22:16:27	IP Address	100	144	True	6/5/2025 22:16:27	R_3v6C1at7dif4rSt	anonymous	EN	0.8999999761581420	An equal mixture of both
16	6/8/2025 15:13:11	6/8/2025 15:24:47	IP Address	100	695	True	6/8/2025 15:24:47	R_1G1co2lYGIfXplO	anonymous	EN	0.8999999761581420	Mostly introverted
17	6/8/2025 15:28:31	6/8/2025 15:30:48	IP Address	100	137	True	6/8/2025 15:30:48	R_1yf2yYNfxel97a1	anonymous	EN	1	Mostly introverted
18	6/8/2025 18:48:46	6/8/2025 18:58:21	Survey Preview	100	575	True	6/8/2025 18:58:22	R_5Crt0OoeWXww23Z	preview	EN	1	
19	6/8/2025 20:25:11	6/8/2025 20:29:11	IP Address	100	239	True	6/8/2025 20:29:12	R_3LwKx9Pzw7iD1sl	anonymous	EN	0.30000001192092900	An equal mixture of both
20	6/9/2025 16:03:48	6/9/2025 16:05:34	IP Address	100	106	True	6/9/2025 16:05:34	R_1EGasH8b9MbGuii	anonymous	EN	1	Mostly introverted

Dirty Attribute 2: Incomplete Responses

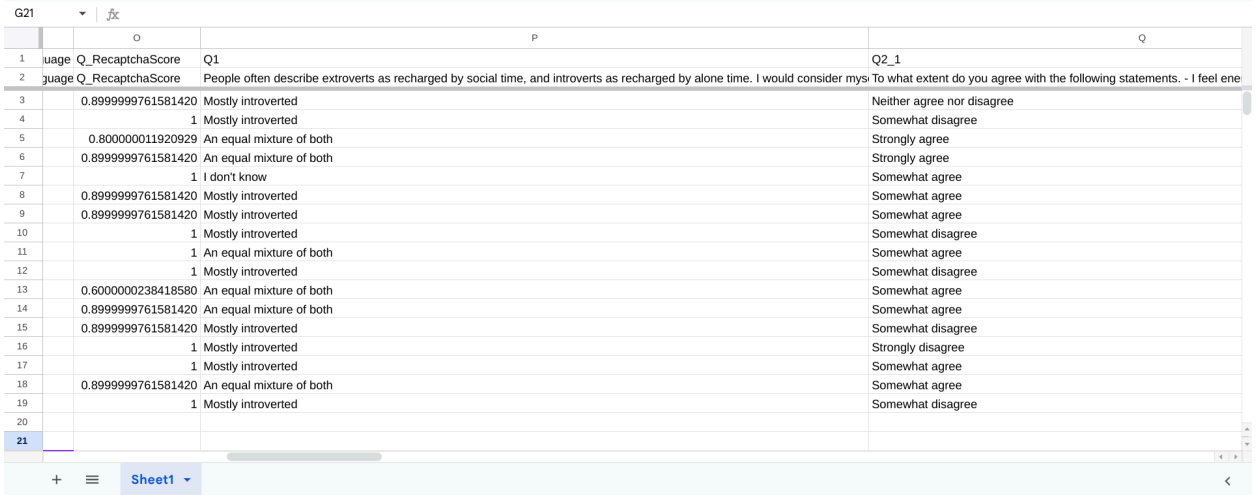
The dirty attribute being cleaned were any responses with two or more blanks in their responses. To clean this data we went through all the rows to find any responses with more than one unanswered question and deleted that entire row (response). Any responses which had two or more unanswered questions did not provide accurate and complete data, which made it unable to be analyzed. Responses with only one blank in their responses were considered clean, since they still provided enough data to be analyzed. Once all entries were removed, each group member went through the rows to ensure that there were no responses with more than one blank left. By removing these responses, we ensured that the data was clean, and all responses being analyzed had enough information to reach valuable conclusions.

Dirty Attribute 2 (Before Cleaning)



	J	K	L	M
1	UserLanguage	Q_RecaptchaScore	Q1	Q2_1
2	User Language	Q_RecaptchaScore	People often describe extroverts as recharged by social time, and introverts as recharged by alone time. I would consider mys	To what extent do you agree with the following state
9	EN	1		
10	EN	1	I don't know	Somewhat agree
11	EN	0.8999999761581420	Mostly introverted	Somewhat agree
12	EN	0.8999999761581420	Mostly introverted	Somewhat agree
13	EN	1	Mostly introverted	Somewhat disagree
14	EN	1	An equal mixture of both	Somewhat agree
15	EN	1	Mostly introverted	Somewhat disagree
16	EN	0.6000000238418580	An equal mixture of both	Somewhat agree
17	EN	0.8999999761581420	An equal mixture of both	Somewhat agree
18	EN	0.8999999761581420	Mostly introverted	Somewhat disagree
19	EN	1	Mostly introverted	Strongly disagree
20	EN	1		
21	EN	0.30000001192092900	An equal mixture of both	Somewhat disagree
22	EN	1	Mostly introverted	Somewhat agree
23	EN	0.8999999761581420	An equal mixture of both	Somewhat agree
24	EN	1	Mostly introverted	Somewhat disagree
25				
26				

Dirty Attribute 2 (After Cleaning)



	O	P	Q
1	Q_RecaptchaScore	Q1	Q2_1
2	Q_RecaptchaScore	People often describe extroverts as recharged by social time, and introverts as recharged by alone time. I would consider mys	To what extent do you agree with the following statements. - I feel ene
3	0.8999999761581420	Mostly introverted	Neither agree nor disagree
4	1	Mostly introverted	Somewhat disagree
5	0.800000011920929	An equal mixture of both	Strongly agree
6	0.8999999761581420	An equal mixture of both	Strongly agree
7	1	I don't know	Somewhat agree
8	0.8999999761581420	Mostly introverted	Somewhat agree
9	0.8999999761581420	Mostly introverted	Somewhat agree
10	1	Mostly introverted	Somewhat disagree
11	1	An equal mixture of both	Somewhat agree
12	1	Mostly introverted	Somewhat disagree
13	0.6000000238418580	An equal mixture of both	Somewhat agree
14	0.8999999761581420	An equal mixture of both	Somewhat agree
15	0.8999999761581420	Mostly introverted	Somewhat disagree
16	1	Mostly introverted	Strongly disagree
17	1	Mostly introverted	Somewhat agree
18	0.8999999761581420	An equal mixture of both	Somewhat agree
19	1	Mostly introverted	Somewhat disagree
20			
21			

Dirty Attribute 3: "Distribution Channel" Column

In the column "Distribution Channel" there were responses which indicated they were "previews" rather than "anonymous." We removed all these rows by selecting them and backspacing to delete them. Any responses in the "preview" column were blank or mostly blank because it was us previewing the survey to ensure it was working which caused the entire row to be blank or only a few questions to be answered. Removing this dirty data caused only responses labeled "anonymous" to be left behind and reviewed, ensuring that only complete responses were analyzed. The before photo shows the preview response highlighted in yellow, and the after photo shows our data without the "preview" response under "Distribution Channel." Removing the dirty data ensures that only responses which were completed by actual respondents are being included in the analyses.

Dirty Attribute 3 (Before Cleaning)

	StartDate	EndDate	Status	Progress	Duration (in seconds)	Finished	RecordedDate	ResponseId	DistributionChannel	UserLanguage	Q_RecaptchaScore	Q1
2	Start Date	End Date	Response Type	Progress	Duration (in seconds)	Finished	Recorded Date	Response ID	Distribution Channel	User Language	Q_RecaptchaScore	People often describe ex
7	6/4/2025 18:50:43	6/4/2025 18:50:57	IP Address	100	13	True	6/4/2025 18:50:57	R_7mVRbhztomMY3VP	anonymous	EN	0.8999999761581420	I don't know
8	6/4/2025 18:44:03	6/4/2025 18:56:41	IP Address	100	758	True	6/4/2025 18:56:42	R_5czRxRtSYjvJs5v	anonymous	EN	1	I don't know
9	6/4/2025 19:08:04	6/4/2025 19:15:07	IP Address	100	423	True	6/4/2025 19:15:07	R_5dF0licausQJdgG	anonymous	EN	0.8999999761581420	Mostly introverted
10	6/4/2025 19:48:03	6/4/2025 19:51:42	IP Address	100	218	True	6/4/2025 19:51:43	R_1G32DxGRGYtbMcj	anonymous	EN	0.8999999761581420	Mostly introverted
11	6/4/2025 21:01:12	6/4/2025 21:04:04	IP Address	100	171	True	6/4/2025 21:04:04	R_35yVRVpSCmQGGC	anonymous	EN	1	Mostly introverted
12	6/4/2025 21:06:40	6/4/2025 21:09:21	IP Address	100	161	True	6/4/2025 21:09:22	R_5OfwhlvPCvQxVPt	anonymous	EN	1	An equal mixture of both
13	6/4/2025 22:23:11	6/4/2025 22:25:19	IP Address	100	127	True	6/4/2025 22:25:20	R_1tb38leMEAnnQA4	anonymous	EN	1	Mostly introverted
14	6/5/2025 16:23:02	6/5/2025 16:25:17	IP Address	100	134	True	6/5/2025 16:25:17	R_5xXJUH5DEYGcdxv	anonymous	EN	0.6000000238418580	An equal mixture of both
15	6/5/2025 22:14:02	6/5/2025 22:16:27	IP Address	100	144	True	6/5/2025 22:16:27	R_3v6C1at7dif4rSt	anonymous	EN	0.8999999761581420	An equal mixture of both
16	6/8/2025 15:13:11	6/8/2025 15:24:47	IP Address	100	695	True	6/8/2025 15:24:47	R_1G1co2lYGIfXplO	anonymous	EN	0.8999999761581420	Mostly introverted
17	6/8/2025 15:28:31	6/8/2025 15:30:48	IP Address	100	137	True	6/8/2025 15:30:48	R_1yf2yYNfxel97a1	anonymous	EN	1	Mostly introverted
18	6/8/2025 18:48:46	6/8/2025 18:58:21	Survey Preview	100	575	True	6/8/2025 18:58:22	R_5Crt0OoeWXww23Z	preview	EN	1	
19	6/8/2025 20:25:11	6/8/2025 20:29:11	IP Address	100	239	True	6/8/2025 20:29:12	R_3LwKx9Pzw7iD1sl	anonymous	EN	0.30000001192092900	An equal mixture of both
20	6/9/2025 16:03:48	6/9/2025 16:05:34	IP Address	100	106	True	6/9/2025 16:05:34	R_1EGasH8b9MbGuii	anonymous	EN	1	Mostly introverted
21	6/9/2025 16:03:50	6/9/2025 16:05:53	IP Address	100	122	True	6/9/2025 16:05:53	R_50jmSc2brtMlqDQ	anonymous	EN	0.8999999761581420	An equal mixture of both
22	6/9/2025 16:07:26	6/9/2025 16:09:36	IP Address	100	129	True	6/9/2025 16:09:36	R_5OEaF11YZrtIiil	anonymous	EN	1	Mostly introverted

Dirty Attribute 3 (After Cleaning)

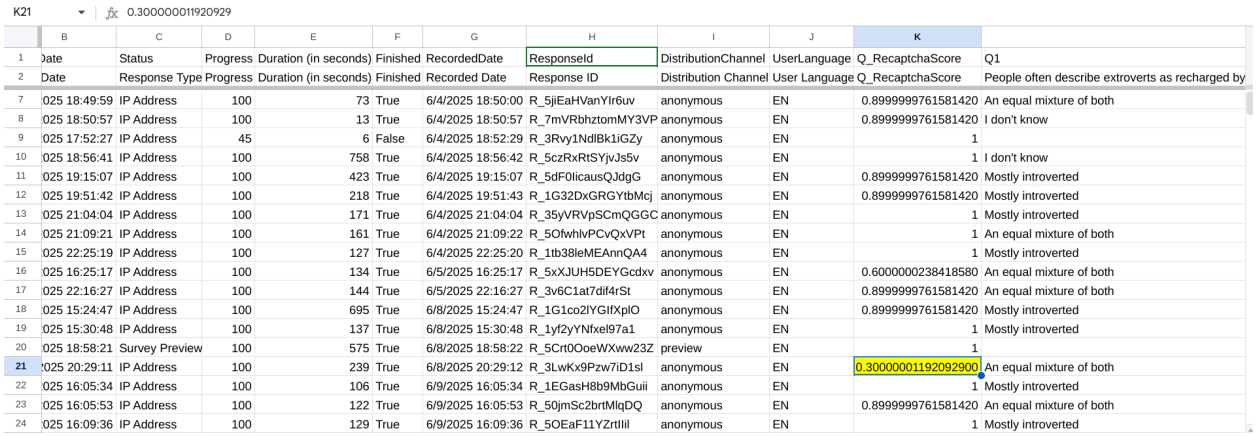
	A	B	C	D	E	F	G	H	I	J	K	
1	StartDate	EndDate	Status	Progress	Duration (in seconds)	Finished	RecordedDate	ResponseId	DistributionChannel	UserLanguage	Q_RecaptchaScore	Q1
2	Start Date	End Date	Response Type	Progress	Duration (in seconds)	Finished	Recorded Date	Response ID	Distribution Channel	User Language	Q_RecaptchaScore	People often describe ex
7	6/4/2025 18:50:43	6/4/2025 18:50:57	IP Address	100	13	True	6/4/2025 18:50:57	R_7mVRbhztomMY3VP	anonymous	EN	0.8999999761581420	I don't know
8	6/4/2025 18:44:03	6/4/2025 18:56:41	IP Address	100	758	True	6/4/2025 18:56:42	R_5czRxRtSYjvJs5v	anonymous	EN	1	I don't know
9	6/4/2025 19:08:04	6/4/2025 19:15:07	IP Address	100	423	True	6/4/2025 19:15:07	R_5dF0licausQJdgG	anonymous	EN	0.8999999761581420	Mostly introverted
10	6/4/2025 19:48:03	6/4/2025 19:51:42	IP Address	100	218	True	6/4/2025 19:51:43	R_1G32DxGRGYtbMcj	anonymous	EN	0.8999999761581420	Mostly introverted
11	6/4/2025 21:01:12	6/4/2025 21:04:04	IP Address	100	171	True	6/4/2025 21:04:04	R_35yVRVpSCmQGGC	anonymous	EN	1	Mostly introverted
12	6/4/2025 21:06:40	6/4/2025 21:09:21	IP Address	100	161	True	6/4/2025 21:09:22	R_5OfwhlvPCvQxVPt	anonymous	EN	1	An equal mixture of both
13	6/4/2025 22:23:11	6/4/2025 22:25:19	IP Address	100	127	True	6/4/2025 22:25:20	R_1tb38leMEAnnQA4	anonymous	EN	1	Mostly introverted
14	6/5/2025 16:23:02	6/5/2025 16:25:17	IP Address	100	134	True	6/5/2025 16:25:17	R_5xXJUH5DEYGcdxv	anonymous	EN	0.6000000238418580	An equal mixture of both
15	6/5/2025 22:14:02	6/5/2025 22:16:27	IP Address	100	144	True	6/5/2025 22:16:27	R_3v6C1at7dif4rSt	anonymous	EN	0.8999999761581420	An equal mixture of both
16	6/8/2025 15:13:11	6/8/2025 15:24:47	IP Address	100	695	True	6/8/2025 15:24:47	R_1G1co2lYGIfXplO	anonymous	EN	0.8999999761581420	Mostly introverted
17	6/8/2025 15:28:31	6/8/2025 15:30:48	IP Address	100	137	True	6/8/2025 15:30:48	R_1yf2yYNfxel97a1	anonymous	EN	1	Mostly introverted
18	6/8/2025 20:25:11	6/8/2025 20:29:11	IP Address	100	239	True	6/8/2025 20:29:12	R_3LwKx9Pzw7iD1sl	anonymous	EN	0.30000001192092900	An equal mixture of both
19	6/9/2025 16:03:48	6/9/2025 16:05:34	IP Address	100	106	True	6/9/2025 16:05:34	R_1EGasH8b9MbGuii	anonymous	EN	1	Mostly introverted
20	6/9/2025 16:03:50	6/9/2025 16:05:53	IP Address	100	122	True	6/9/2025 16:05:53	R_50jmSc2brtMlqDQ	anonymous	EN	0.8999999761581420	An equal mixture of both
21	6/9/2025 16:07:26	6/9/2025 16:09:36	IP Address	100	129	True	6/9/2025 16:09:36	R_5OEaF11YZrtIiil	anonymous	EN	1	Mostly introverted

Sheet1

Dirty Attribute 4: Bot Detection

All responses with a "Q_RecaptchaScore" of less than 0.5 were removed, because a score less than this value indicates the respondent is likely a bot and not human. Having data points that are produced by a bot would skew our data, as it would not accurately measure the constructs we are seeking to study. We cleaned the data by looking for entries with a Q Recaptcha score of less than 0.5, and removed any entries that fell within this range. One entry had a Q Recaptcha score of 0.3 (which is highlighted in yellow in the image), and this entry was deleted.

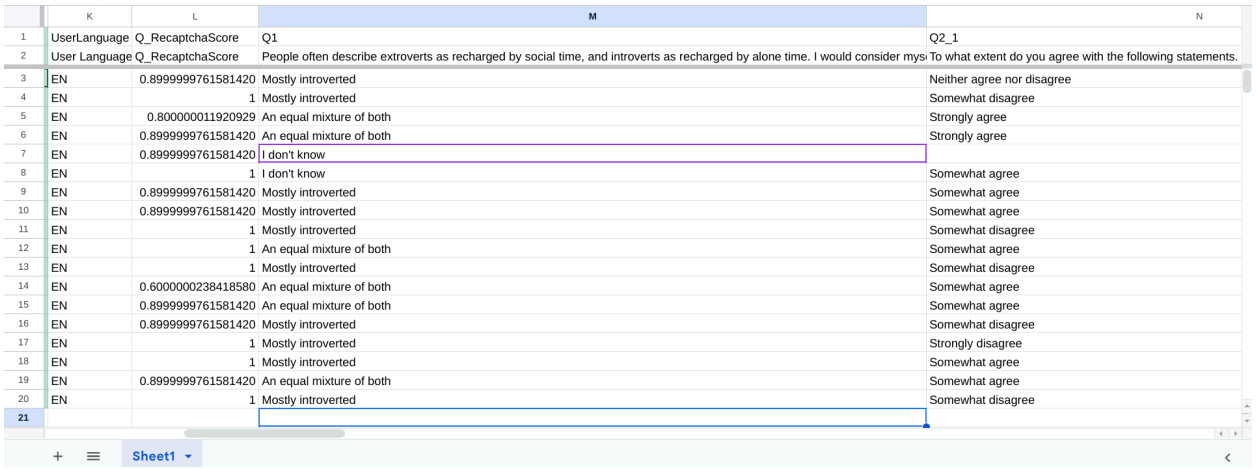
Dirty Attribute 4 (Before Cleaning)



K21 | 0.300000011920929

	B	C	D	E	F	G	H	I	J	K	
1	Date	Status	Progress	Duration (in seconds)	Finished	RecordedDate	ResponseId	DistributionChannel	UserLanguage	Q_RecaptchaScore	Q1
2	Date	Response Type	Progress	Duration (in seconds)	Finished	Recorded Date	Response ID	Distribution Channel	User Language	Q_RecaptchaScore	People often describe extroverts as recharged by
7	025 18:49:59	IP Address	100	73	True	6/4/2025 18:50:00	R_5jiEaHVanYIr6uv	anonymous	EN	0.8999999761581420	An equal mixture of both
8	025 18:50:57	IP Address	100	13	True	6/4/2025 18:50:57	R_7mVRbhztomMY3VP	anonymous	EN	0.8999999761581420	I don't know
9	025 17:52:27	IP Address	45	6	False	6/4/2025 18:52:29	R_3Rvy1NdlBk1iGZy	anonymous	EN	1	
10	025 18:56:41	IP Address	100	758	True	6/4/2025 18:56:42	R_5czRxRtSYjvJs5v	anonymous	EN	1	I don't know
11	025 19:15:07	IP Address	100	423	True	6/4/2025 19:15:07	R_5dF0licausQJdgG	anonymous	EN	0.8999999761581420	Mostly introverted
12	025 19:51:42	IP Address	100	218	True	6/4/2025 19:51:43	R_1G32DxGRGYtbMcj	anonymous	EN	0.8999999761581420	Mostly introverted
13	025 21:04:04	IP Address	100	171	True	6/4/2025 21:04:04	R_35yVRVpSCmQGGC	anonymous	EN	1	Mostly introverted
14	025 21:09:21	IP Address	100	161	True	6/4/2025 21:09:22	R_5OfwhlvPCvQxVPt	anonymous	EN	1	An equal mixture of both
15	025 22:25:19	IP Address	100	127	True	6/4/2025 22:25:20	R_1tb38leMEAnnQA4	anonymous	EN	1	Mostly introverted
16	025 16:25:17	IP Address	100	134	True	6/5/2025 16:25:17	R_5xXJUH5DEYGcdxv	anonymous	EN	0.6000000238418580	An equal mixture of both
17	025 22:16:27	IP Address	100	144	True	6/5/2025 22:16:27	R_3v6C1at7dif4rSt	anonymous	EN	0.8999999761581420	An equal mixture of both
18	025 15:24:47	IP Address	100	695	True	6/8/2025 15:24:47	R_1G1co2lYGIfXplO	anonymous	EN	0.8999999761581420	Mostly introverted
19	025 15:30:48	IP Address	100	137	True	6/8/2025 15:30:48	R_1yf2yYNfxel97a1	anonymous	EN	1	Mostly introverted
20	025 18:58:21	Survey Preview	100	575	True	6/8/2025 18:58:22	R_5Crt0OoeWXww23Z	preview	EN	1	
21	025 20:29:11	IP Address	100	239	True	6/8/2025 20:29:12	R_3LwKx9Pzw7iD1sl	anonymous	EN	0.30000001192092900	An equal mixture of both
22	025 16:05:34	IP Address	100	106	True	6/9/2025 16:05:34	R_1EGasH8b9MbGuii	anonymous	EN	1	Mostly introverted
23	025 16:05:53	IP Address	100	122	True	6/9/2025 16:05:53	R_50jmSc2brtMlqDQ	anonymous	EN	0.8999999761581420	An equal mixture of both
24	025 16:09:36	IP Address	100	129	True	6/9/2025 16:09:36	R_5OEaF11YZrtIlil	anonymous	EN	1	Mostly introverted

Dirty Attribute 4 (After Cleaning)



	K	L	M	N
1	UserLanguage	Q_RecaptchaScore	Q1	Q2_1
2	User Language	Q_RecaptchaScore	People often describe extroverts as recharged by social time, and introverts as recharged by alone time. I would consider mys	To what extent do you agree with the following statements.
3	EN	0.8999999761581420	Mostly introverted	Neither agree nor disagree
4	EN	1	Mostly introverted	Somewhat disagree
5	EN	0.800000011920929	An equal mixture of both	Strongly agree
6	EN	0.8999999761581420	An equal mixture of both	Strongly agree
7	EN	0.8999999761581420	I don't know	
8	EN	1	I don't know	Somewhat agree
9	EN	0.8999999761581420	Mostly introverted	Somewhat agree
10	EN	0.8999999761581420	Mostly introverted	Somewhat agree
11	EN	1	Mostly introverted	Somewhat disagree
12	EN	1	An equal mixture of both	Somewhat agree
13	EN	1	Mostly introverted	Somewhat disagree
14	EN	0.6000000238418580	An equal mixture of both	Somewhat agree
15	EN	0.8999999761581420	An equal mixture of both	Somewhat agree
16	EN	0.8999999761581420	Mostly introverted	Somewhat disagree
17	EN	1	Mostly introverted	Strongly disagree
18	EN	1	Mostly introverted	Somewhat agree
19	EN	0.8999999761581420	An equal mixture of both	Somewhat agree
20	EN	1	Mostly introverted	Somewhat disagree
21				

Sheet1

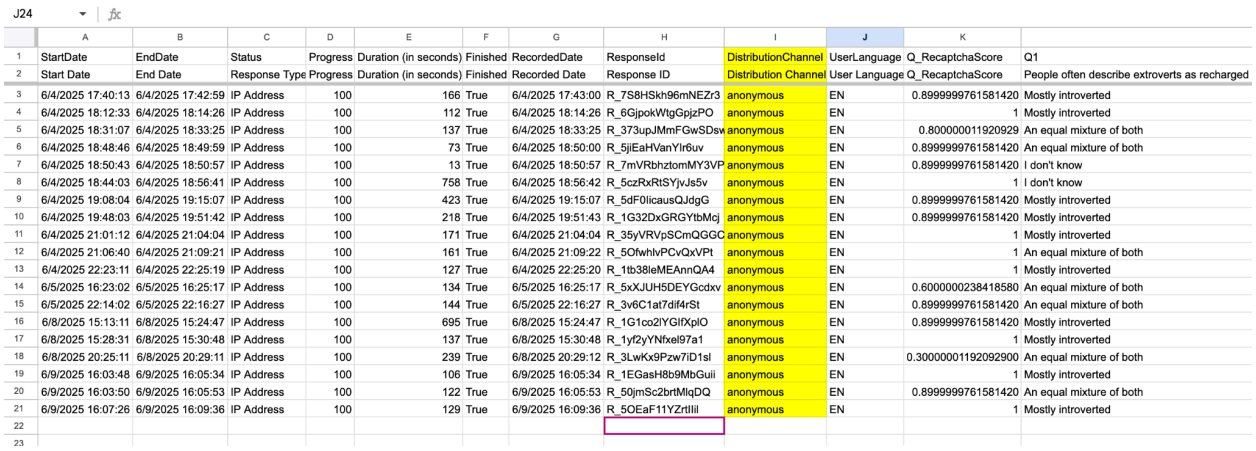
Dirty Attribute 5: Proper Capitalization

Standardizing capitalization within the dataset is important for maintaining consistency and reducing errors during analysis. Inconsistent formatting can cause software to misinterpret similar terms or group values incorrectly, which may lead to skewed results and can negatively affect the validity of the study. Ensuring consistent formatting such as capitalization before

analysis also improves efficiency, as it eliminates the need for corrections after visualizations are created and simplifies the overall process.

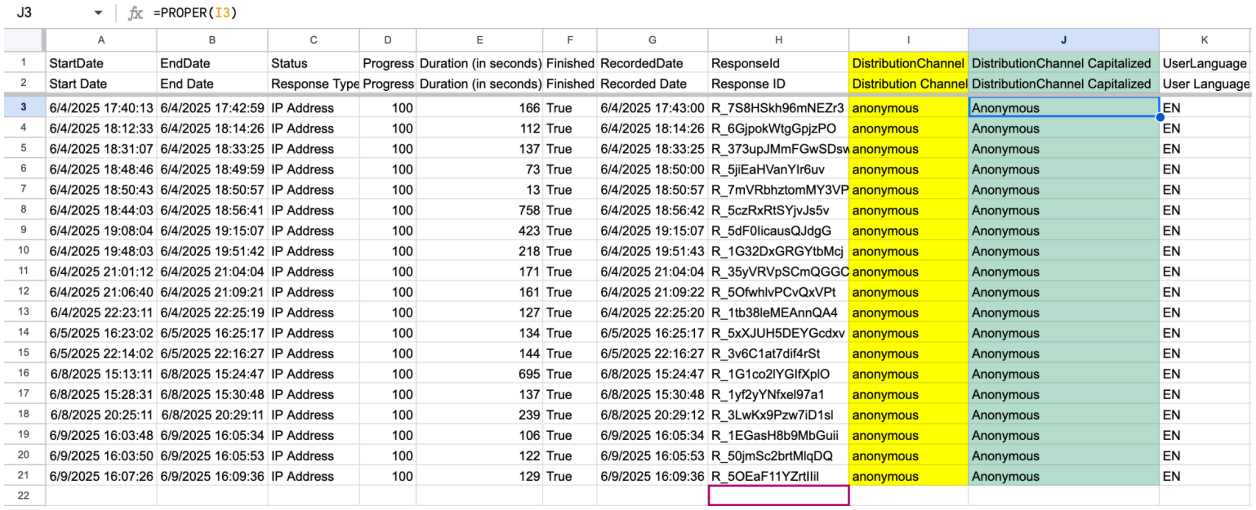
To clean this data, capitalization needed to be corrected within the “Distribution Channel” column by changing the term “anonymous” to “Anonymous.” To do this, a new column was inserted to the right of the “Distribution Channel” column and was labeled “Distribution Channel Capitalized.” The formula =PROPER(I#) was then entered into each cell in the new column, with # replaced by the appropriate row number. For example, as shown in the image, the formula in cell J3 is =PROPER(I3).

Dirty Attribute 5 (Before Cleaning)



	A	B	C	D	E	F	G	H	I	J	K	
1	StartDate	EndDate	Status	Progress	Duration (in seconds)	Finished	RecordedDate	ResponseId	DistributionChannel	UserLanguage	Q_RecaptchaScore	Q1
2	Start Date	End Date	Response Type	Progress	Duration (in seconds)	Finished	Recorded Date	Response ID	Distribution Channel	User Language	Q_RecaptchaScore	People often describe extroverts as recharged
3	6/4/2025 17:40:13	6/4/2025 17:42:59	IP Address	100	166	True	6/4/2025 17:43:00	R_7S8HSkh96mNEZr3	anonymous	EN	0.8999999761581420	Mostly introverted
4	6/4/2025 18:12:33	6/4/2025 18:14:26	IP Address	100	112	True	6/4/2025 18:14:26	R_6GjpokWtgGpjzPO	anonymous	EN	1	Mostly introverted
5	6/4/2025 18:31:07	6/4/2025 18:33:25	IP Address	100	137	True	6/4/2025 18:33:25	R_373upJMmFGwSDsw	anonymous	EN	0.800000011920929	An equal mixture of both
6	6/4/2025 18:48:46	6/4/2025 18:49:59	IP Address	100	73	True	6/4/2025 18:50:00	R_5jiEaHVanYlr6uv	anonymous	EN	0.8999999761581420	An equal mixture of both
7	6/4/2025 18:50:43	6/4/2025 18:50:57	IP Address	100	13	True	6/4/2025 18:50:57	R_7mVRbhztomMY3VP	anonymous	EN	0.8999999761581420	I don't know
8	6/4/2025 18:44:03	6/4/2025 18:56:41	IP Address	100	758	True	6/4/2025 18:56:42	R_5czRxRtSYjvJs5v	anonymous	EN	1	I don't know
9	6/4/2025 19:08:04	6/4/2025 19:15:07	IP Address	100	423	True	6/4/2025 19:15:07	R_5dF0licausQJdgG	anonymous	EN	0.8999999761581420	Mostly introverted
10	6/4/2025 19:48:03	6/4/2025 19:51:42	IP Address	100	218	True	6/4/2025 19:51:43	R_1G32DxGRGYtbMcj	anonymous	EN	0.8999999761581420	Mostly introverted
11	6/4/2025 21:01:12	6/4/2025 21:04:04	IP Address	100	171	True	6/4/2025 21:04:04	R_35yVRVpSCmQGGC	anonymous	EN	1	Mostly introverted
12	6/4/2025 21:06:40	6/4/2025 21:09:21	IP Address	100	161	True	6/4/2025 21:09:22	R_5OfwhlvPCvQxVPt	anonymous	EN	1	An equal mixture of both
13	6/4/2025 22:23:11	6/4/2025 22:25:19	IP Address	100	127	True	6/4/2025 22:25:20	R_1tb38leMEAnnQA4	anonymous	EN	1	Mostly introverted
14	6/5/2025 16:23:02	6/5/2025 16:25:17	IP Address	100	134	True	6/5/2025 16:25:17	R_5xXJUH5DEYGcdxv	anonymous	EN	0.6000000238418580	An equal mixture of both
15	6/5/2025 22:14:02	6/5/2025 22:16:27	IP Address	100	144	True	6/5/2025 22:16:27	R_3v6C1at7dif4rSt	anonymous	EN	0.8999999761581420	An equal mixture of both
16	6/8/2025 15:13:11	6/8/2025 15:24:47	IP Address	100	695	True	6/8/2025 15:24:47	R_1G1co2lYGlfXplO	anonymous	EN	0.8999999761581420	Mostly introverted
17	6/8/2025 15:28:31	6/8/2025 15:30:48	IP Address	100	137	True	6/8/2025 15:30:48	R_1yf2yYNfxel97a1	anonymous	EN	1	Mostly introverted
18	6/8/2025 20:25:11	6/8/2025 20:29:11	IP Address	100	239	True	6/8/2025 20:29:12	R_3LwKx9Pzw7iD1sl	anonymous	EN	0.30000001192092900	An equal mixture of both
19	6/9/2025 16:03:48	6/9/2025 16:05:34	IP Address	100	106	True	6/9/2025 16:05:34	R_1EGasH8b9MbGuii	anonymous	EN	1	Mostly introverted
20	6/9/2025 16:03:50	6/9/2025 16:05:53	IP Address	100	122	True	6/9/2025 16:05:53	R_50jmSc2brtMlqDQ	anonymous	EN	0.8999999761581420	An equal mixture of both
21	6/9/2025 16:07:26	6/9/2025 16:09:36	IP Address	100	129	True	6/9/2025 16:09:36	R_5OEaF11YZrtIlil	anonymous	EN	1	Mostly introverted
22												
23												

Dirty Attribute 5 (After Cleaning)



	A	B	C	D	E	F	G	H	I	J	K
1	StartDate	EndDate	Status	Progress	Duration (in seconds)	Finished	RecordedDate	ResponseId	DistributionChannel	DistributionChannel Capitalized	UserLanguage
2	Start Date	End Date	Response Type	Progress	Duration (in seconds)	Finished	Recorded Date	Response ID	Distribution Channel	DistributionChannel Capitalized	User Language
3	6/4/2025 17:40:13	6/4/2025 17:42:59	IP Address	100	166	True	6/4/2025 17:43:00	R_7S8HSkh96mNEZr3	anonymous	Anonymous	EN
4	6/4/2025 18:12:33	6/4/2025 18:14:26	IP Address	100	112	True	6/4/2025 18:14:26	R_6GjpokWtgGpjzPO	anonymous	Anonymous	EN
5	6/4/2025 18:31:07	6/4/2025 18:33:25	IP Address	100	137	True	6/4/2025 18:33:25	R_373upJMmFGwSDsw	anonymous	Anonymous	EN
6	6/4/2025 18:48:46	6/4/2025 18:49:59	IP Address	100	73	True	6/4/2025 18:50:00	R_5jiEaHVanYlr6uv	anonymous	Anonymous	EN
7	6/4/2025 18:50:43	6/4/2025 18:50:57	IP Address	100	13	True	6/4/2025 18:50:57	R_7mVRbhztomMY3VP	anonymous	Anonymous	EN
8	6/4/2025 18:44:03	6/4/2025 18:56:41	IP Address	100	758	True	6/4/2025 18:56:42	R_5czRxRtSYjvJs5v	anonymous	Anonymous	EN
9	6/4/2025 19:08:04	6/4/2025 19:15:07	IP Address	100	423	True	6/4/2025 19:15:07	R_5dF0licausQJdgG	anonymous	Anonymous	EN
10	6/4/2025 19:48:03	6/4/2025 19:51:42	IP Address	100	218	True	6/4/2025 19:51:43	R_1G32DxGRGYtbMcj	anonymous	Anonymous	EN
11	6/4/2025 21:01:12	6/4/2025 21:04:04	IP Address	100	171	True	6/4/2025 21:04:04	R_35yVRVpSCmQGGC	anonymous	Anonymous	EN
12	6/4/2025 21:06:40	6/4/2025 21:09:21	IP Address	100	161	True	6/4/2025 21:09:22	R_5OfwhlvPCvQxVPt	anonymous	Anonymous	EN
13	6/4/2025 22:23:11	6/4/2025 22:25:19	IP Address	100	127	True	6/4/2025 22:25:20	R_1tb38leMEAnnQA4	anonymous	Anonymous	EN
14	6/5/2025 16:23:02	6/5/2025 16:25:17	IP Address	100	134	True	6/5/2025 16:25:17	R_5xXJUH5DEYGcdxv	anonymous	Anonymous	EN
15	6/5/2025 22:14:02	6/5/2025 22:16:27	IP Address	100	144	True	6/5/2025 22:16:27	R_3v6C1at7dif4rSt	anonymous	Anonymous	EN
16	6/8/2025 15:13:11	6/8/2025 15:24:47	IP Address	100	695	True	6/8/2025 15:24:47	R_1G1co2lYGlfXplO	anonymous	Anonymous	EN
17	6/8/2025 15:28:31	6/8/2025 15:30:48	IP Address	100	137	True	6/8/2025 15:30:48	R_1yf2yYNfxel97a1	anonymous	Anonymous	EN
18	6/8/2025 20:25:11	6/8/2025 20:29:11	IP Address	100	239	True	6/8/2025 20:29:12	R_3LwKx9Pzw7iD1sl	anonymous	Anonymous	EN
19	6/9/2025 16:03:48	6/9/2025 16:05:34	IP Address	100	106	True	6/9/2025 16:05:34	R_1EGasH8b9MbGuii	anonymous	Anonymous	EN
20	6/9/2025 16:03:50	6/9/2025 16:05:53	IP Address	100	122	True	6/9/2025 16:05:53	R_50jmSc2brtMlqDQ	anonymous	Anonymous	EN
21	6/9/2025 16:07:26	6/9/2025 16:09:36	IP Address	100	129	True	6/9/2025 16:09:36	R_5OEaF11YZrtIlil	anonymous	Anonymous	EN
22											

Dirty Attribute 6: Split Columns

When cleaning data, it is necessary for each column to contain only one data type to ensure that analysis can be conducted accurately. In our data file, the columns “StartDate,” “EndDate,” and “RecordedDate” each contained the date and time, which are two distinct types

of data. If the columns were not split and analysis was run on the variable of time, for example, the analysis could not be conducted, as the column contains both time and date. As such, it was vital for the three columns indicated above to be split so that each data type was held in a distinct column.

The data was split by first adding an additional column to the right of the "StartDate," "End Date," and "RecordedDate" columns. We then selected "Data," "Split text to columns," and specified a space to be the separator. The column headings were then reformatted to accurately represent the data within them, now labelled "StartDate," "StartTime," "EndDate," "EndTime," "RecordedDate," and "RecordedTime."

Dirty Attribute 6 (Before Cleaning)

H23 | fx

	A	B	C	D	E	F	G	H
1	StartDate	EndDate	Status	Progress	Duration (in seconds)	Finished	RecordedDate	ResponseId
2	Start Date	End Date	Response Type	Progress	Duration (in seconds)	Finished	Recorded Date	Response ID
3	6/4/2025 17:40:13	6/4/2025 17:42:59	IP Address	100	166	True	6/4/2025 17:43:00	R_7S8HSkh96mNEZr3
4	6/4/2025 18:12:33	6/4/2025 18:14:26	IP Address	100	112	True	6/4/2025 18:14:26	R_6GjpokWtgGpjzPO
5	6/4/2025 18:31:07	6/4/2025 18:33:25	IP Address	100	137	True	6/4/2025 18:33:25	R_373upJMmFGwSDsv
6	6/4/2025 18:48:46	6/4/2025 18:49:59	IP Address	100	73	True	6/4/2025 18:50:00	R_5jiEaHVanYlr6uv
7	6/4/2025 18:50:43	6/4/2025 18:50:57	IP Address	100	13	True	6/4/2025 18:50:57	R_7mVRbhztomMY3VP
8	6/4/2025 18:44:03	6/4/2025 18:56:41	IP Address	100	758	True	6/4/2025 18:56:42	R_5czRxRtSYjvJs5v
9	6/4/2025 19:08:04	6/4/2025 19:15:07	IP Address	100	423	True	6/4/2025 19:15:07	R_5dF0licausQJdgG
10	6/4/2025 19:48:03	6/4/2025 19:51:42	IP Address	100	218	True	6/4/2025 19:51:43	R_1G32DxGRGYtbMcj
11	6/4/2025 21:01:12	6/4/2025 21:04:04	IP Address	100	171	True	6/4/2025 21:04:04	R_35yVRVpSCmQGGC
12	6/4/2025 21:06:40	6/4/2025 21:09:21	IP Address	100	161	True	6/4/2025 21:09:22	R_5OfwhlvPCvQxVPt
13	6/4/2025 22:23:11	6/4/2025 22:25:19	IP Address	100	127	True	6/4/2025 22:25:20	R_1tb38leMEAnnQA4
14	6/5/2025 16:23:02	6/5/2025 16:25:17	IP Address	100	134	True	6/5/2025 16:25:17	R_5xXJUH5DEYGcdxv
15	6/5/2025 22:14:02	6/5/2025 22:16:27	IP Address	100	144	True	6/5/2025 22:16:27	R_3v6C1at7dif4rSt
16	6/8/2025 15:13:11	6/8/2025 15:24:47	IP Address	100	695	True	6/8/2025 15:24:47	R_1G1co2lYGlfXplO
17	6/8/2025 15:28:31	6/8/2025 15:30:48	IP Address	100	137	True	6/8/2025 15:30:48	R_1yf2yYNfxel97a1
18	6/9/2025 16:03:48	6/9/2025 16:05:34	IP Address	100	106	True	6/9/2025 16:05:34	R_1EGasH8b9MbGuii
19	6/9/2025 16:03:50	6/9/2025 16:05:53	IP Address	100	122	True	6/9/2025 16:05:53	R_50jmSc2brtMlqDQ
20	6/9/2025 16:07:26	6/9/2025 16:09:36	IP Address	100	129	True	6/9/2025 16:09:36	R_5OEaF11YZrtIlil
21								

Dirty Attribute 6 (After Cleaning)

I21

	A	B	C	D	E	F	G	H	I	J	K
1	StartDate	StartTime	EndDate	EndTime	Status	Progress	Duration (in seconds)	Finished	RecordedDate	Recorded Time	ResponseId
2	Start Date	Start Time	End Date	End Time	Response Type	Progress	Duration (in seconds)	Finished	Recorded Date	Recorded Time	Response ID
3	6/4/2025	17:40:13	6/4/2025	17:42:59	IP Address	100	166	True	6/4/2025	17:43:00	R_7S8HSkh96mNEZr3
4	6/4/2025	18:12:33	6/4/2025	18:14:26	IP Address	100	112	True	6/4/2025	18:14:26	R_6GjpokWtgGpjzPO
5	6/4/2025	18:31:07	6/4/2025	18:33:25	IP Address	100	137	True	6/4/2025	18:33:25	R_373upJMmFGwSDsw
6	6/4/2025	18:48:46	6/4/2025	18:49:59	IP Address	100	73	True	6/4/2025	18:50:00	R_5jiEaHVanYlr6uv
7	6/4/2025	18:44:03	6/4/2025	18:56:41	IP Address	100	758	True	6/4/2025	18:56:42	R_5czRxRtSYjvJs5v
8	6/4/2025	19:08:04	6/4/2025	19:15:07	IP Address	100	423	True	6/4/2025	19:15:07	R_5dF0licausQJdgG
9	6/4/2025	19:48:03	6/4/2025	19:51:42	IP Address	100	218	True	6/4/2025	19:51:43	R_1G32DxGRGYtbMcj
10	6/4/2025	21:01:12	6/4/2025	21:04:04	IP Address	100	171	True	6/4/2025	21:04:04	R_35yVRVpSCmQGGC
11	6/4/2025	21:06:40	6/4/2025	21:09:21	IP Address	100	161	True	6/4/2025	21:09:22	R_5OfwhlvPCvQxVPt
12	6/4/2025	22:23:11	6/4/2025	22:25:19	IP Address	100	127	True	6/4/2025	22:25:20	R_1tb38leMEAnnQA4
13	6/5/2025	16:23:02	6/5/2025	16:25:17	IP Address	100	134	True	6/5/2025	16:25:17	R_5xXJUH5DEYGcdxv
14	6/5/2025	22:14:02	6/5/2025	22:16:27	IP Address	100	144	True	6/5/2025	22:16:27	R_3v6C1at7dif4rSt
15	6/8/2025	15:13:11	6/8/2025	15:24:47	IP Address	100	695	True	6/8/2025	15:24:47	R_1G1co2lYGlfXplO
16	6/8/2025	15:28:31	6/8/2025	15:30:48	IP Address	100	137	True	6/8/2025	15:30:48	R_1yf2yYNfxel97a1
17	6/9/2025	16:03:48	6/9/2025	16:05:34	IP Address	100	106	True	6/9/2025	16:05:34	R_1EGasH8b9MbGuii
18	6/9/2025	16:03:50	6/9/2025	16:05:53	IP Address	100	122	True	6/9/2025	16:05:53	R_50jmSc2brtMlqDQ
19	6/9/2025	16:07:26	6/9/2025	16:09:36	IP Address	100	129	True	6/9/2025	16:09:36	R_5OEaF11YZrtIiil
20											

Clean Attribute 1: “Progress” Column

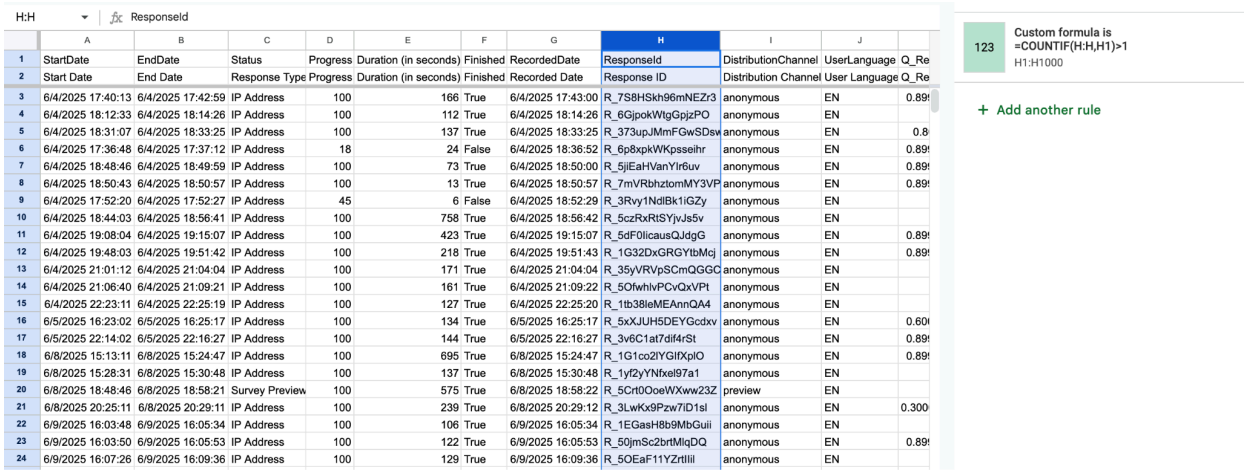
The “Progress” column contains percentage values that indicate the completion level of the survey for each respondent. All values are 100%, with no nulls or formatting issues. This confirms that every recorded entry represents a fully completed survey, making the attribute clean and reliable for analysis.

Clean Attribute 2: “Duration (In Seconds)” Column

The “Duration (In Seconds)” column is clean because all values are non-negative integers representing how long each respondent took to complete the survey. There are no missing values, outliers, or inconsistencies, ensuring the data can be used to evaluate engagement time if needed.

Clean Attribute 3: Duplicate Responses

To ensure there were no duplicate responses, the “ResponseID” column was examined for multiple submissions. This was accomplished using the conditional formatting tool and applying a custom formula, which can be seen in the upper right corner of the image. The formula used was =COUNTIF(H:H,H1)>1. Green cells indicated duplicate responses, while white cells indicated unique entries. After applying the formula, all cells remained white, signifying that there were no duplicate submissions and that the data in the “ResponseID” column was already clean.



H:H | ResponseId

	A	B	C	D	E	F	G	H	I	J	
1	StartDate	EndDate	Status	Progress	Duration (in seconds)	Finished	RecordedDate	ResponseId	DistributionChannel	UserLanguage	Q_Re
2	Start Date	End Date	Response Type	Progress	Duration (in seconds)	Finished	Recorded Date	Response ID	Distribution Channel	User Language	Q_Re
3	6/4/2025 17:40:13	6/4/2025 17:42:59	IP Address	100	166	True	6/4/2025 17:43:00	R_7S8HSkh96mNEZr3	anonymous	EN	0.89
4	6/4/2025 18:12:33	6/4/2025 18:14:26	IP Address	100	112	True	6/4/2025 18:14:26	R_6GjpokWtgGpjzPO	anonymous	EN	
5	6/4/2025 18:31:07	6/4/2025 18:33:25	IP Address	100	137	True	6/4/2025 18:33:25	R_373upJMmFGwSDsv	anonymous	EN	0.8
6	6/4/2025 17:36:48	6/4/2025 17:37:12	IP Address	18	24	False	6/4/2025 18:36:52	R_6p8xpkWKpsseihr	anonymous	EN	0.89
7	6/4/2025 18:48:46	6/4/2025 18:49:59	IP Address	100	73	True	6/4/2025 18:50:00	R_5jiEaHVanYlr6uv	anonymous	EN	0.89
8	6/4/2025 18:50:43	6/4/2025 18:50:57	IP Address	100	13	True	6/4/2025 18:50:57	R_7mVRbhztomMY3VP	anonymous	EN	0.89
9	6/4/2025 17:52:20	6/4/2025 17:52:27	IP Address	45	6	False	6/4/2025 18:52:29	R_3Rvy1NdlBk1iGZy	anonymous	EN	
10	6/4/2025 18:44:03	6/4/2025 18:56:41	IP Address	100	758	True	6/4/2025 18:56:42	R_5czRxRtSYjvJs5v	anonymous	EN	
11	6/4/2025 19:08:04	6/4/2025 19:15:07	IP Address	100	423	True	6/4/2025 19:15:07	R_5dF0licausQJdgG	anonymous	EN	0.89
12	6/4/2025 19:48:03	6/4/2025 19:51:42	IP Address	100	218	True	6/4/2025 19:51:43	R_1G32DxGRGYtbMcj	anonymous	EN	0.89
13	6/4/2025 21:01:12	6/4/2025 21:04:04	IP Address	100	171	True	6/4/2025 21:04:04	R_35yVRVpSCmQGGC	anonymous	EN	
14	6/4/2025 21:06:40	6/4/2025 21:09:21	IP Address	100	161	True	6/4/2025 21:09:22	R_5OfwhlvPCvQxVPt	anonymous	EN	
15	6/4/2025 22:23:11	6/4/2025 22:25:19	IP Address	100	127	True	6/4/2025 22:25:20	R_1tb38leMEAnnQA4	anonymous	EN	
16	6/5/2025 16:23:02	6/5/2025 16:25:17	IP Address	100	134	True	6/5/2025 16:25:17	R_5xXJUH5DEYGcdxv	anonymous	EN	0.60
17	6/5/2025 22:14:02	6/5/2025 22:16:27	IP Address	100	144	True	6/5/2025 22:16:27	R_3v6C1at7dif4rSt	anonymous	EN	0.89
18	6/8/2025 15:13:11	6/8/2025 15:24:47	IP Address	100	695	True	6/8/2025 15:24:47	R_1G1co2lYGlfXplO	anonymous	EN	0.89
19	6/8/2025 15:28:31	6/8/2025 15:30:48	IP Address	100	137	True	6/8/2025 15:30:48	R_1yf2yYNfxel97a1	anonymous	EN	
20	6/8/2025 18:48:46	6/8/2025 18:58:21	Survey Preview	100	575	True	6/8/2025 18:58:22	R_5Crt0OoeWXww23Z	preview	EN	
21	6/8/2025 20:25:11	6/8/2025 20:29:11	IP Address	100	239	True	6/8/2025 20:29:12	R_3LwKx9Pzw7iD1sI	anonymous	EN	0.300
22	6/9/2025 16:03:48	6/9/2025 16:05:34	IP Address	100	106	True	6/9/2025 16:05:34	R_1EGasH8b9MbGuii	anonymous	EN	
23	6/9/2025 16:03:50	6/9/2025 16:05:53	IP Address	100	122	True	6/9/2025 16:05:53	R_50jmSc2brtMlqDQ	anonymous	EN	0.89
24	6/9/2025 16:07:26	6/9/2025 16:09:36	IP Address	100	129	True	6/9/2025 16:09:36	R_5OEaF11YZrtIlil	anonymous	EN	

123 Custom formula is =COUNTIF(H:H,H1)>1
H1:H1000

+ Add another rule

Clean Attribute 4: "User Language" Column

The "User Language" column includes standardized two-letter language codes (e.g., "EN" for English). All entries are formatted consistently with no missing or invalid codes, confirming the data is already clean and uniform across responses.

Clean Attribute 5: Question 1

This attribute is clean because there are no empty responses, and the completed data we gathered from this question will be necessary during analysis and when determining the answers to our research questions. The formatting and capitalization in this column was all correct as well, making this column clean.

Clean Attribute 6: Question 2 (Parts 1-4)

The data represented in question 2 (Q2_1, Q2_2, Q2_3, Q2_4) is clean as there are no incomplete responses. The data gathered from this question is necessary to further understand the constructs we are seeking to explore and to answer our research questions. All of the formatting and capitalization was also correct in this column making the data clean.

Clean Attribute 7: Question 3 (Parts 1-6)

The data represented in question 3 (Q3_1, Q3_2, Q3_3, Q3_4, Q3_5, Q3_6) is clean because every response is complete. The data from this question is necessary to further understand the constructs we are studying and to answer our research questions. All of the formatting and capitalization was also correct in this column making the data clean.

Clean Attribute 8: Question 4

The data represented in Q4 is clean as all entries have proper formatting, correct spelling and capitalization. All entries make logical sense and are complete, making the data clean.

Clean Attribute 9: Question 5

The data represented in Q5 is clean as all entries have proper formatting, correct spelling and capitalization. All entries make logical sense and are complete, making the data clean.

Clean Attribute 10: Questions 6-11

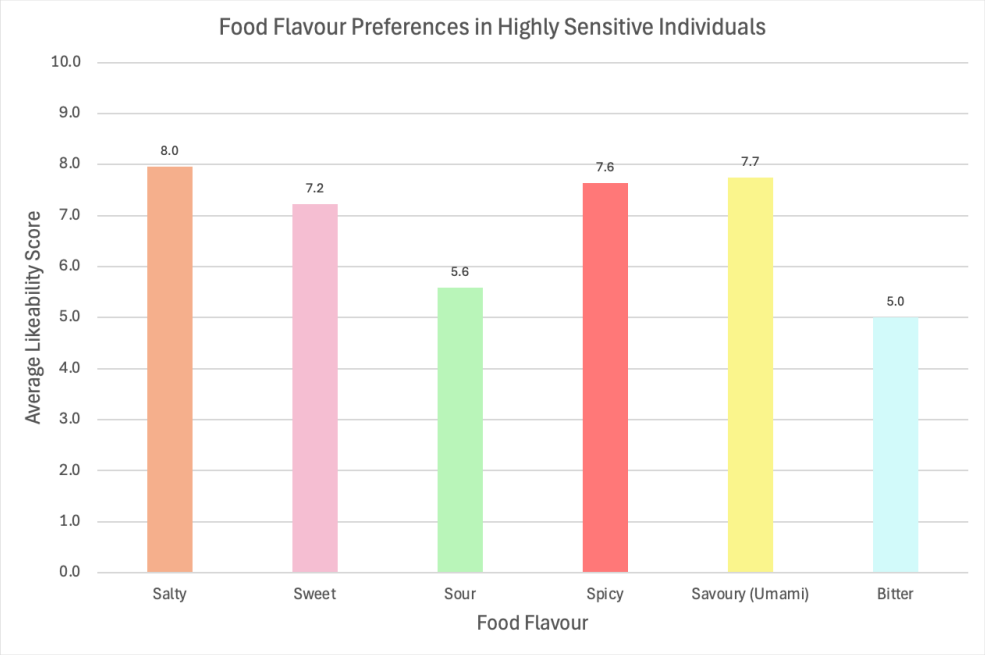
The data represented in Q6, Q7, Q8, Q9, Q10, and Q11 are clean. They all have proper formatting, and all values are numerical and between the appropriate range, which is between 0 and 10. Q11 contains one blank entry. However, we defined clean responses as those with fewer than two blank answers. Since this was the only question left unanswered by the respondent, Q11 was still considered clean.

Section E: Visualizations and Findings

Visualization and Findings for Analysis Question 1

What food flavours are preferred by individuals who identify as highly sensitive (i.e., those who selected “describes me very well” or “describes me extremely well”), and how do these flavour preferences compare in terms of likeability?

Visualization



Note. Likeability scores range from 0 to 10, with 0 representing a strong dislike for the food flavour and 10 representing a strong preference for it.

Rationale

A vertical bar graph was chosen to present the data for the first analysis question because it best showcases the data and consequent findings. The average likeability score is presented on the y-axis, and the six distinct food flavours are shown on the x-axis, with data labels indicating the mean likeability score above each bar. This format enables easy side-by-side comparisons of different flavors and their likeability scores, and presents the data clearly and visually, making it easier to understand at a glance. By organizing the information in this format, viewers can quickly and effectively compare the likeability scores of various flavours, leading to a deeper understanding of the data. Overall, the vertical bar graph simplifies complex data, making the analysis more accessible and straightforward.

Discussion of Findings

The visualization effectively addresses the first analysis question by revealing whether highly sensitive individuals exhibit preferences for specific food flavours, and if so, the order in which these flavours are ranked. Based on the visualization created from the data, it is evident that individuals categorized as highly sensitive exhibit clear preferences for certain food flavours, with a distinct ranking among them. The graph indicates that salty flavours are the most preferred ($M = 8.0$), followed closely by savoury ($M = 7.7$) and spicy options ($M = 7.6$). Additionally, highly sensitive individuals appear to have a strong preference for sweet flavours ($M = 7.2$). In contrast, bitter ($M = 5.0$) and sour flavours ($M = 5.6$) are less favoured by this group. Overall, the data suggests that highly sensitive individuals tend to prefer salty, sweet, spicy, and savoury flavours over bitter and sour ones.

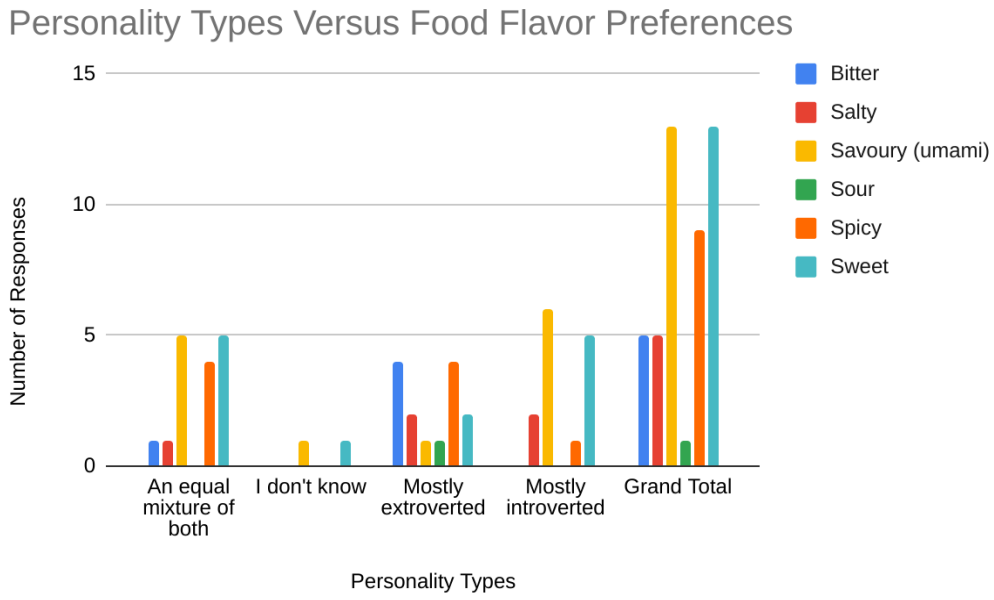
Although existing research does not demonstrate a positive correlation between high sensitivity and traits such as novelty seeking, sensation seeking, or sensitivity to rewards, it does indicate that individuals high in reward dependence tend to prefer sweet flavours (Day et al., 2008). Additionally, increased reward dependence has been observed in individuals classified as highly sensitive (Bjørnebekk et al., 2012). These findings align with our results, which revealed that highly sensitive individuals exhibited a strong preference for sweet foods ($M = 7.2$). Furthermore, while a direct correlation has not been established between antisocial tendencies and high sensitivity, research has shown that highly sensitive individuals typically score high on the agreeableness dimension of personality (Trå et al., 2022). Similarly, individuals who prefer sweet flavours have also been found to be more agreeable (Wu, 2023). These findings support our results, as highly sensitive respondents in our study showed a notable preference for sweet flavours ($M = 7.2$).

Future research should examine whether the preference for sweet flavours among highly sensitive individuals is directly influenced by reward dependence and agreeableness, or whether these associations are mediated by other personality traits or environmental factors.

Visualization and Findings for Analysis Question 2

Do more intense flavours (like spicy, bitter, etc.) tend to be associated with certain personality types versus less intense flavours (like savoury or sweet), which may be associated with other personality types?

Visualization



Rationale

I chose to use a bar graph to visually represent the data shown above (i.e., the number of responses for each flavour based on whether the participant identified as introverted, extroverted, both, or were unsure). This format was chosen as it easily displayed the number of responses each flavour received as well as the number of responses for each personality type. By using this type of graph, I was able to see how many introverts prefer salty foods or spicy foods, and do the same for all the other categories as well, which allowed me to then easily compare the numbers.

Discussion of Findings

Through this visual representation of our data, I was able to determine which food flavors were most popular per personality type and make comparisons to answer the analysis question which asked whether more intense flavours (like spicy, bitter, etc.) tend to be associated with certain personality types versus less intense flavours (like savoury or sweet).

The data displayed shows that for individuals who identified as “mostly introverted,” the most popular flavour was savoury with 6 responses, and sweet was a close second with 5 responses. For those who identified as introverted, the least popular option was spicy flavors, with only 1 response. For “mostly extroverted” individuals, the most popular flavours were bitter and spicy, with 4 responses each, and sour and savoury were the least popular. For participants who answered “an equal mixture of both,” meaning that they are ambiverts, the most popular flavours were sweet and savoury, with 5 responses.

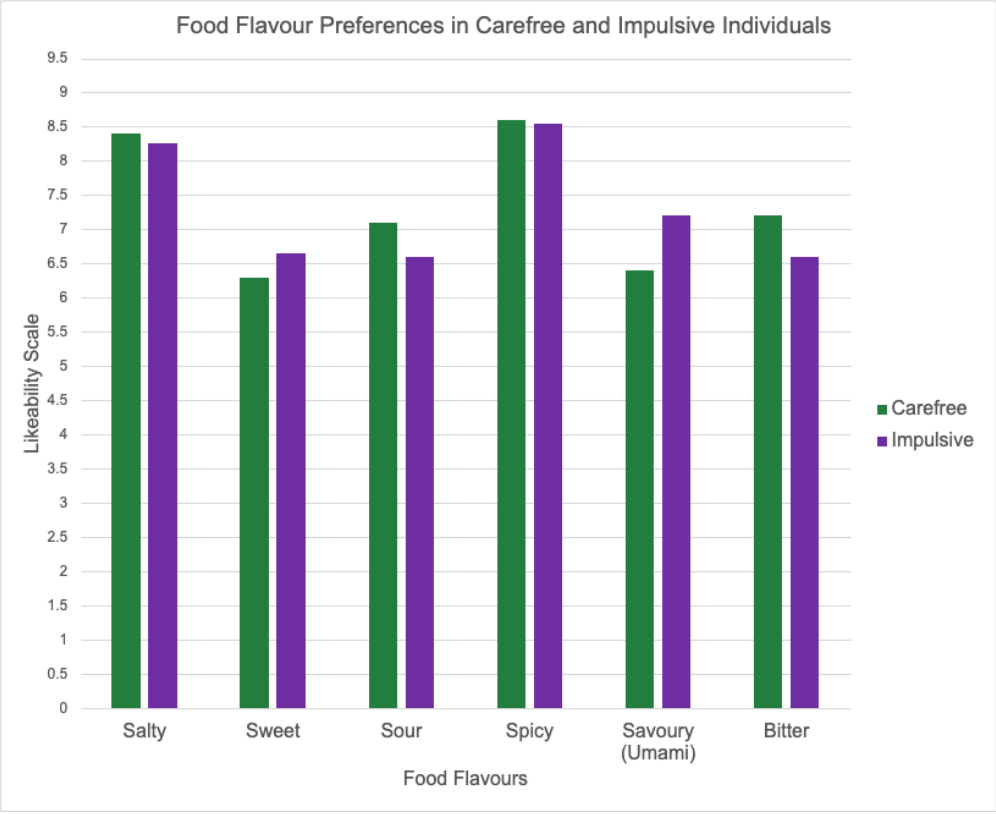
In analyzing this data, we see that introverts tend to like savoury or sweet flavours, which are typically milder, and their least popular flavor was spicy, which is bolder. This correlates with some of the research we did for this project (Day et al., 2008; De Fruyt et al., 2000; Gocłowska et al., 2019), and also demonstrates a clear correlation of individuals who identify as introverted

preferring certain flavors over others (namely, more mild flavours). In this visualization, we also see that extroverted individuals, or participants who identified as 'mostly extroverted,' preferred bolder flavours like spicy or bitter. In comparison, milder flavours like sweet were less popular. This also demonstrates that extroverts have clear flavor preferences for bolder flavors compared to milder ones, as proposed by our analysis question 2. The data displaying that extroverts prefer bolder flavors also aligns with our research on the topic, which suggested that those with more impulsive and outgoing personality traits like bolder flavors (Day et al., 2008). In reviewing this visualization, it is clear that there are correlations between extroverts liking spicy or bold flavors and introverts liking sweet or other mild flavors, answering our earlier proposed analysis question.

Visualization and Findings for Analysis Question 3

To what extent do individuals who describe themselves as carefree or impulsive prefer strong and unconventional flavours (e.g. spicy, bitter, sour, salty) over less bold ones (e.g. sweet, savoury)?

Visualization



Rationale

The reason why I chose a side by side bar graph to visualize this analysis question was to simply demonstrate the direct comparison between two groups across multiple different

categories. Since we were comparing two different types of people (carefree and impulsive) to preferences of flavours (sweet, spicy, sour, bitter, umami), having a side by side bar chart made it easy to identify any similarities or differences between the two personality traits. Both traits are similar, so seeing the range of responses is important. This visualization also highlights the likeability of different flavours, showcasing which are more preferred by what personality type. This is crucial to answering the main question and allows for an easy glance over the data to understand. This visualization is created by taking an average of each likeability scale, and using that to compare between the different flavours. This creates a strong understanding of the answer to the question, while maintaining an organized and clear chart.

Discussion of Findings

The third analysis question is effectively answered through this visualization of a side-by-side bar chart which visualizes the comparison of individuals who identify themselves as either carefree or impulsive (described them very or extremely well) and how they rate the likeability of different food flavours on a scale of 1 to 10. There were six food flavour categories; salty, sweet, sour, spicy, savoury, and bitter, which work together to show how impulsive or carefree personalities may favour bold and unconventional flavours over less intense ones.

As shown in the visualization created from the survey data, both carefree and impulsive individuals prefer salty and spicy food over any other. The likeability scale is based on an average of the scale administered in the survey, and shows that people who describe themselves as carefree and impulsive both prefer spicy food the most, with a mean of 8.6 and 8.5, respectively. This is the maximum likeability of both personality types, indicating that there may be a connection between preferring spicy food and having a risky-carefree personality type. Both personality types were also more inclined to prefer salty foods, at an average just under the maximum. Those who describe themselves as carefree have a mean of 8.4, while impulsive individuals have a mean of 8.3. This correlation for salty and spicy suggests a shared tendency to prefer bold and stimulating tastes, which aligns with the background research, stating that spicy food is associated with higher levels of sensation (risk) seeking, which is a common factor in impulsive and carefree individuals. These data points also highlight how similar both carefree and impulsive individuals are with their likeability scales, as the majority of the data is within 0.5 units from one another. This indicates that individuals who are more likely to take risks, have similar food preferences, further solidifying the correlation between food flavour preferences and personality types.

Individuals who describe themselves as having a carefree personality have sweet as their lowest rated food flavour, with a mean of 6.3. This is similar to individuals who are impulsive, as it is also the minimum, but with a mean of 6.6. This indicates that personality types based on being carefree and impulsive may dislike more conventional and less bold flavours. Food flavours that are conventional may be too boring or bland for personalities drawn to boldness.

Preferences for sour and bitter flavours follow similar patterns, where carefree individuals rated higher on the likeability scale than impulsive ones. These likeability

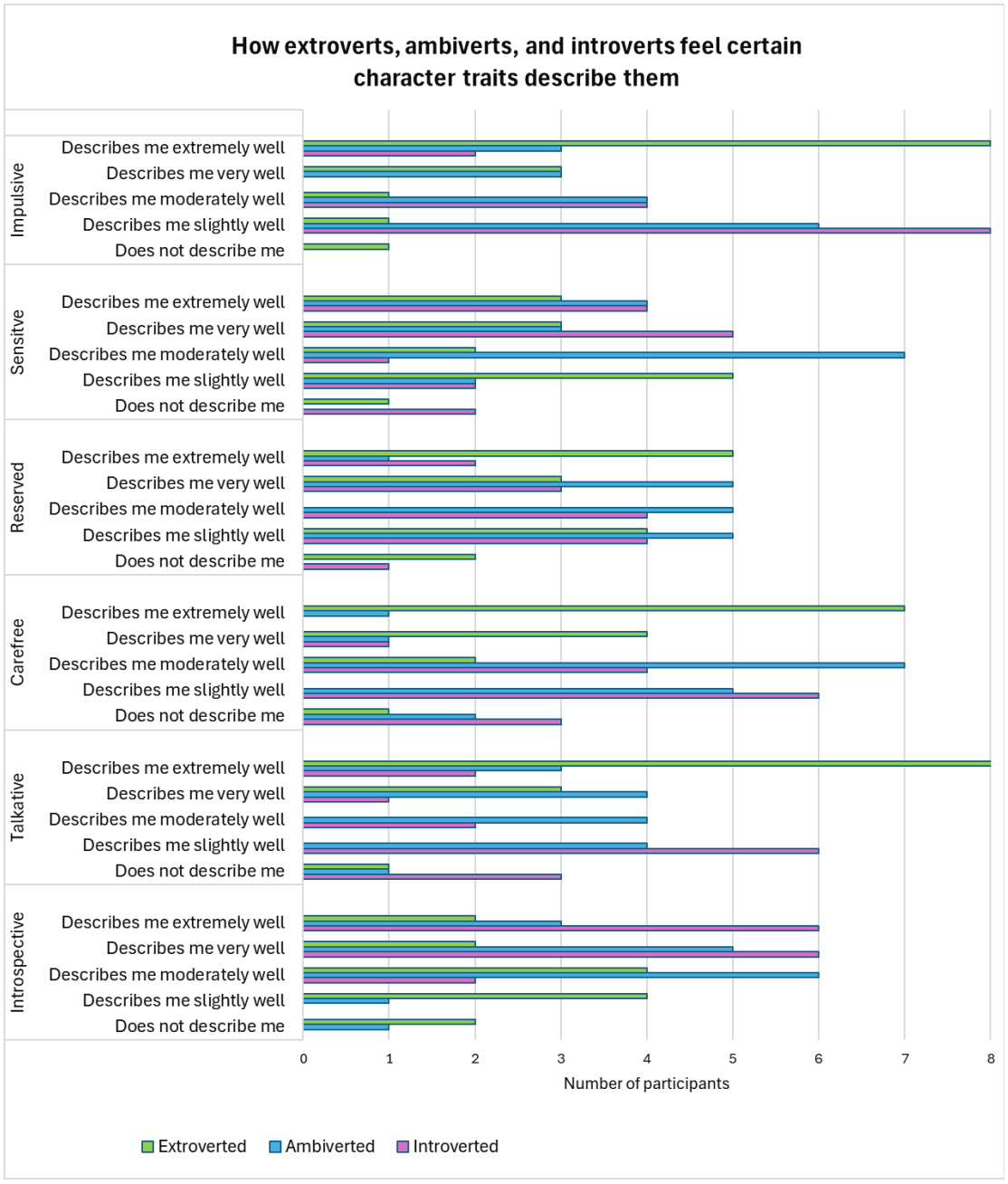
preferences are also very similar in data points, as carefree individuals have a mean of 7.1 for both sour and bitter, while impulsive individuals have a mean of 6.6 for both. This highlights that carefree individuals may be more open to bold flavours. For savoury foods, impulsive individuals have a higher average on the likeability scale ($M = 7.2$) whereas carefree individuals were much lower ($M = 6.4$).

Overall, both personality types favour some of the strong and bold flavours (spicy and salty), but have different preferences when it comes to more conventional flavours. There is also a range of responses for some aspects of bold flavours, such as bitter. Impulsive individuals tend to rank flavours where they are immediately gratified (such as sweet), higher than carefree individuals. Both personality types are quite similar in responses, indicating the correlation between flavour preferences and personality types which are more inclined to risks. The findings mainly relate to background research and knowledge, highlighting that the correlation between food flavours and personality types is one of significance.

Visualization and Findings for Analysis Question 4

Do individuals' statements about being extroverted, introverted, or in between align with their answers about their personality characteristics?

Visualization



Rationale

I chose a clustered bar chart because it effectively communicates the nominal and ratio information I want to convey. Because the bars are clustered, it is easy to tell which category each bar is a part of. Furthermore, the use of distinctive colours helps identify which bars correspond to extroverts, ambiverts, and introverts. Bar graphs are also a simple way to compare different values.

Discussion of Findings

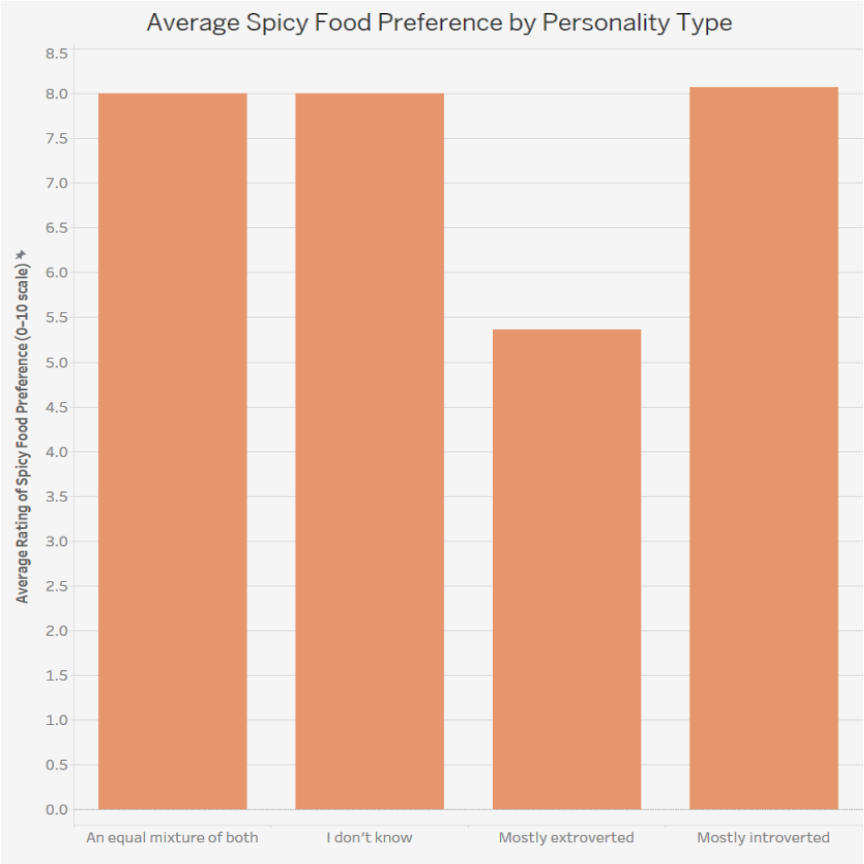
From the graph, it appears that being talkative and impulsive are the strongest character traits in extroverts. In contrast, introverts generally have more variation for each characteristic. However, they are always on the opposite end of each characteristic compared to extroverts. For instance, for carefree, most extroverts were either extremely or very carefree, whereas most introverts were slightly or not at all carefree. Ambiverts were typically somewhere between extroverts and introverts but were the same as introverts or extroverts a few times. For example, most ambiverts were in the middle with respect to carefree and sensitive with moderately describing them.

With the exception of reserved, most traits align with what would be expected. Extroverts were more carefree, talkative, and impulsive, whereas introverts were generally more introspective and sensitive. According to our study, extroverts identified as more reserved than introverts. However, introverts are typically considered more reserved than extroverts. Our small sample size may explain this anomaly.

Visualization and Findings for Analysis Question 5

Is there a noticeable difference in flavour preferences between individuals who identify as introverts versus those who identify as ambiverts?

Visualization



Rationale

I chose a vertical bar chart because it allows for an easy visual comparison of average spicy food preferences across distinct personality groups. Since the variable for personality type is categorical (e.g., “Mostly introverted” or “An equal mixture of both”) and the spicy preference rating is numerical (0–10), a bar chart clearly highlights which group rates spicy food more highly on average. This format is also simple to interpret and visually intuitive for spotting differences between categories.

Discussion of Findings

From the visualization, we can observe that introverts reported the highest average preference for spicy food. Ambiverts and individuals who selected “I don’t know” followed closely behind. Extroverts, surprisingly, had the lowest average preference for spicy food. These results are interesting because they slightly contradict existing research which states that extroverts, due to their sensation-seeking traits, would be more inclined toward bold or spicy flavours (Byrnes & Hayes, 2013). This finding suggests that flavour preference may depend on more than just personality alone, and may be partially due to other factors like sensitivity or cultural exposure.

Section F: Contributions

Table 1

Individual Contributions of Group Members

Group Member	Contributions
Alice Thwaites	Project proposal, survey creation, introduction, analysis question 1, corresponding visualization and write-up, data cleaning, 2 dirty data write-ups & 1 clean data write-up, individual reflection, powerpoint
Ava Sawers	Project proposal, recruitment message, introduction, analysis question 2, corresponding visualization and write-up, data cleaning, 1 dirty data write-up & 3 clean data write-ups, individual reflection, powerpoint
Alisha Sajjad	Project proposal, consent form, analysis question 3,

	corresponding visualization and write-up, data cleaning, 2 dirty data write-ups, individual reflection, powerpoint
Jenessa Entz	Project proposal, introduction, analysis question 4, corresponding visualization and write-up, data cleaning, 1 dirty data write-up & 3 clean data write-ups, individual reflection, powerpoint
Jaiveer Toor	Project proposal, obtaining data write-up, analysis question 5, corresponding visualization and write-up, data cleaning, 3 clean data write-ups, individual reflection, powerpoint

Section G: Individual Reflections

Individual Reflection: Alice Thwaites

Obtaining Data

Upon reflecting on my Assignment 2 submission, I concluded that creating a more effective survey for the group project required incorporating a greater variety of question types, ensuring that each question included the necessary background information for participants to answer accurately, and improving overall question quality to enhance the reliability of the data collected.

For example, in my Assignment 2 submission, I included a self-report question asking participants whether they considered themselves mostly extroverted, mostly introverted, or an equal mixture of both. However, I failed to provide definitions for the terms "extrovert" and "introvert," which could have led to misinterpretation and skewed data which did not accurately represent the participants. Additionally, I did not include an "I don’t know" response option, which may have forced participants to choose an answer that did not properly reflect themselves simply because they were unsure.

In the survey for the group project, we addressed these issues by including definitions for extroversion and introversion to help participants provide informed responses. We also added an "I don’t know" option to ensure participants were not pressured into selecting an inaccurate

response. These modifications significantly improved the quality of the data by increasing its accuracy and reducing the likelihood of skewed results. Furthermore, we expanded the range of question types used in the group project survey. In addition to multiple choice questions, we incorporated matrix-style questions and slider scale questions. This variety allowed us to capture more complete and nuanced data, contributing to a stronger analysis.

Cleaning Data

In my Assignment 3 submission, one major issue that needed improvement was my failure to properly address incomplete responses in the dataset. Specifically, I did not remove all incomplete entries, nor did I clearly define what would qualify as an incomplete response. Although there were several rows in the dataset with multiple blank fields, I did not remove them during data cleaning. This error could significantly impact the accuracy of consequent analysis by allowing incomplete or potentially invalid responses to remain in the dataset. Including these entries can skew the results, as the missing values may lead to unreliable conclusions.

In our group assignment, we took a more structured approach to this issue and collectively agreed that any response containing two or more blank fields would be considered incomplete. Based on this criterion, those rows were removed from the dataset. This decision ensured that only complete responses were included in our analysis. As a result, we were able to draw more accurate conclusions from the cleaned data. By establishing clear rules for data inclusion, we increased the reliability of our findings and demonstrated a more thorough approach to data handling.

Digital Visualization

After reviewing my Assignment 4 submission, I concluded that several modifications could have been made to my visualizations to more accurately reflect the data in a concise and visually digestible format. I addressed and improved upon all of these aspects in the visualization created for the group project. In my Assignment 4 submission, I failed to include data labels at the tops of the bars in the bar charts. This exclusion made it more difficult to interpret the data accurately, as the exact values for each bar were not visible, potentially leading to inaccurate comparisons. Additionally, some of the visualizations were difficult to read and interpret because they attempted to compare variables with different scales, which made the charts visually confusing. Finally, all the bars in the bar graphs were the same color, making it challenging to distinguish between the different categories being compared.

In the visualization I created for the group project, I addressed these issues by including data labels above each bar, assigning distinct colors to each bar to clearly differentiate categories, and ensuring the visualization was visually easy to understand and interpret. These modifications contributed to more accurate analyses, allowed for quicker and more precise comparisons among groups, and made the data presentation more visually appealing and accessible.

Individual Reflection: Ava Sawers

Obtaining Data

In reviewing my Assignment 2 submission, I concluded that some revisions could have been made to add to the integrity of the assignment. For my survey, I could have had a few more tailored questions into the specific impacts noticed due to the addition of technology as a main learning resource in schools instead of having multiple broader-scope questions. I could have replaced one or two of my broad-scale questions with more specific questions about the effects of technology with the multiple choice answers as specific examples. This would have allowed for more concise data regarding the participants' experience with technology in education as well as more specific answers in regards to some of the effects that have transpired due to technology in education in participants' lives. For example, instead of just having broad questions like "have you noticed a positive or negative shift in the level of education since the implementation of technology as a main learning resource in schools?" I could have also had more tailored questions asking "have you noticed any of these changes in education levels since the implementation of technology as a main learning resource in schools?" and provided multiple choice answers with specific changes where participants could select any that applied.

In the survey for our group project, we had a variety of questions, some of them more broad, like the sliding scale rankings of how much the participants like specific flavours, as well as more detailed and concise questions about their personality types and what sort of traits described the participants best. Having this variety of questions allowed for a more encompassing analysis of our participants personality types and preferences and added more context to their selections throughout the survey.

Cleaning Data

In my cleaning data assignment, one thing that I could have improved on as I have reviewed my assignment after completing this group project is better organization of my descriptions and rationales regarding why the data was deemed dirty. While the information I provided was good, the organization of my rationale distracted from my points and should have been separated and defined more appropriately for the best readability and organization. Instead of having one paragraph for each piece of dirty data I cleaned, I could have separated it into two paragraphs, the first containing information on why the data was dirty, and the second containing information on how I cleaned the data, and finally, why it was now clean.

Our organization of data for our group project was much easier to read and follow. The data was separated, and then descriptions of why the data was dirty were applied, with separate paragraphs for how it was cleaned and why it was now clean. The split columns section of our data cleaning was also an improvement that I could have applied to my previous assignment in hindsight, as it added to clarify the data and make it easier to analyze.

Digital Visualization

Something from my Assignment 4 I noticed I could have improved on after reviewing my submission, is the quality of my visualizations. I could have made a few different improvements, one of which would be changing the colours of my bar graph so that different options were coloured differently, making the graph easier to read and analyze. In not having all of the bars the same colour in my visualization for this group assignment, it made a big difference as the readability was much better and the data was easier to interpret. I also forgot to include titles on a few of my graphs, which made it more difficult to determine what the graph was demonstrating, so I made sure to amend that issue for this project.

In my visualization for this project, I made sure to improve on these issues. I included multiple colours for my bar graph as well as a legend, which made reading and analyzing the data represented in my bar graph easy and allowed for the most efficient and concise representation. I also made sure to include a title so that just by looking at the graph, you are already aware of what is being compared and analyzed, which I then followed up on with my description, rationale, and findings underneath the graph.

Individual Reflection: Alisha Sajjad

Obtaining Data

In my obtaining data assignment there were some areas which could have been improved, mainly relating to the organization of the document. When presenting my rationale for each survey question I included all aspects of the assignment in one paragraph. This made it harder to follow and read. By splitting up each paragraph into two, one on the attribute and one on the level of measurement, the overall organization would have improved. This would have made the information easier to read and understand. Additionally, for some of the multiple choice questions, I did not provide a response choice of “I don't know.” Including this would have been beneficial to participants, as it acknowledges that they may be unsure of an answer. This allows for more reliable data, as it makes them aware they don't have to answer anything, improving honesty.

To improve from the assignment, we prioritized presentation and readability of our document for the group project. Each paragraph was divided into smaller sections based on the content, which greatly improved clarity. The entire document for the group project was structured according to the outline provided, ensuring that it is easy to read and follow. When creating the survey, we were aware of how participants may respond to questions and the choices provided. We included response options such as "I don't know” or equivalent. This allows us to follow ethical practices (ensuring that no participants feel as if they are forced to answer) as well as reliability and honesty in our data.

Cleaning Data

In the cleaning data assignment, I lacked an understanding of proper data cleaning techniques. I was unaware that deleting a column is not considered an appropriate data cleaning

technique. Looking back, I should have utilized a different method which followed lab guidelines. Additionally, similar to the previous assignment (obtaining data) I did not have the best organization and presentation of my ideas. Rather than breaking up paragraphs based on the separate aspects of the questions, I presented the rationale in one paragraph. This causes issues in readability, as it affects how easy the document is to follow through.

For the group project, we were more aware of what counts as a proper data cleaning method. We ensured we were using all of the data cleaning aspects discussed in class, to allow for clean data. We also carefully reviewed all data cleaning methods to confirm they were actually cleaning the data so it could be analyzed properly. When creating paragraphs for rationales we split up large amounts of texts which allowed for clear readability, as well as a more organized document.

Digital Visualization

In the digital visualization assignment I could have improved on the quality of some of my visualizations. Upon reflection, several of the visualizations created were difficult to read, ultimately making them ineffective, as the data can not be interpreted or analyzed. The font of the graphs were hard to read due to the small font size. Some of the colours used for the visualizations could have been improved also, to ensure that there is clear separation between different categories.

For the group project, I prioritized clearly conveying the most important information for the visualizations. I ensured that the graphs included appropriate titles, axes, and legends. I also ensured that the font size and colour was easily readable. This allowed for the visualization to be easy to interpret, assisting us in reaching our conclusions. When selecting the colours of the different elements, I intentionally chose ones which were opposite and provided contrast, to allow viewers to be able to take a quick glance and understand.

Individual Reflection: Jenessa Entz

Obtaining Data

In Assignment 2, I had a question about how much money people spend on their pets. I grouped the amount of money spent into groups, $0-$100, $101-$500, etc. However, I could get a more precise estimate on how much people spend on their pets if I changed the question to a slider. This way, people could input a more exact value.

In our project, we included slider questions for how much people liked each flavour type. This allows for a precise understanding of the extent people like or dislike certain flavours.

Cleaning Data

For my cleaning, in Assignment 3, I could have improved on the descriptions for what I did. So, for this project, I included clear descriptions as to what we did for cleaning. I also

explained why certain columns were already clean. Also, in Assignment 3, not everything was cleaned. However, for this project, we made sure to clean the entire dataset.

Digital Visualization

For Assignment 4, I could have improved my reasoning for why I chose the certain type of graph. For this project, I ensured that I had a detailed description explaining why I chose the graph and how it can be used to effectively portray the information I would like to convey.

Individual Reflection: Jaiveer Toor

Obtaining Data

In Assignment 2, my data collection approach could have been improved by being more intentional with the alignment between each question and the research objective. For example, I initially lacked a clear link between personality traits and flavour preferences, which made it harder to analyze correlations. Additionally, my earlier questions did not define key personality terms or include flexible response options like “I don’t know,” which could lead to biased or forced answers.

In the group project, I improved on these areas by ensuring the survey included well-defined concepts, such as “introvert,” “extrovert,” and “ambivert,” along with an “I don’t know” option. We also used a greater variety of question formats (Likert scale, sliders, and multiple choice), which allowed us to gather more diverse and nuanced data. This improved the clarity and depth of our analysis.

Cleaning Data

In Assignment 3, one area for improvement was not documenting each step of the cleaning process in enough detail. I focused too much on technical actions (like deleting rows) without clearly explaining why the data was considered dirty or how it would affect the analysis if left unchanged. I also didn’t use column splitting techniques or bot filtering, which would have made my dataset cleaner and more structured.

In this group project, I clearly explained the rationale behind each data cleaning decision. For example, I cleaned both data using the Recaptcha score, split columns containing date and time to make them more analyzable, and wrote detailed justifications for removing rows with incomplete responses. I also contributed clean data attribute write-ups, which helped ensure the final dataset was valid, consistent, and ready for visualization.

Digital Visualization

My Assignment 4 submission had several issues, most notably, the visualizations lacked proper labeling, used a single color scheme, and didn’t always use the best chart type for the data. For instance, I used default settings instead of tailoring the visualization to highlight

comparisons effectively, and the absence of axis labels or descriptive titles made the charts harder to interpret.

In the group project, I made significant improvements. I used Tableau to create a clear and well-labeled bar chart comparing average spicy food preference across personality types. I gave the chart a meaningful title, labeled the y-axis clearly, and used color formatting to enhance readability. This made the visualization both visually appealing and analytically strong. The result helped directly answer our research question by showing a clear trend: ambiverts and extroverts tend to prefer spicy foods more than introverts.

References

Aluja, A., García, Ó., & García, L. F. (2003). Relationships among extraversion, openness to experience, and sensation seeking. *Personality and Individual Differences*, *35*(3), 671–680. https://doi.org/10.1016/S0191-8869(02)00244-1

Bjørnebekk, A., Westlye, L. T., Fjell, A. M., Grydeland, H., & Walhovd, K. B. (2012). Social reward dependence and brain white matter microstructure. *Cerebral Cortex (New York, N.Y. 1991)*, *22*(11), 2672–2679. https://doi.org/10.1093/cercor/bhr345

Blain, S. D., Sassenberg, T. A., Xi, M., Zhao, D., & DeYoung, C. G. (2021). Extraversion but Not Depression Predicts Reward Sensitivity: Revisiting the Measurement of Anhedonic Phenotypes. *Journal of Personality and Social Psychology*, *121*(2), e1–e18. https://doi.org/10.1037/pspp0000371

Byrnes, N. K., & Hayes, J. E. (2013). Personality factors predict spicy food liking and intake. *Food Quality and Preference*, *28*(1), 213–221. https://doi.org/10.1016/j.foodqual.2012.09.008

Day, C. J., Francis, J., & Mchale, S. (2008). Associations between temperamental personality and taste preferences in a non-clinical population. *Appetite*, *50*(2), 557–557. https://doi.org/10.1016/j.appet.2007.09.023

De Fruyt, F., Van De Wiele, L., & Van Heeringen, C. (2000). Cloninger's Psychobiological Model of Temperament and Character and the Five-Factor Model of Personality. *Personality and Individual Differences*, *29*(3), 441–452. https://doi.org/10.1016/S0191-8869(99)00204-4

Gocłowska, M. A., Ritter, S. M., Elliot, A. J., & Baas, M. (2019). Novelty seeking is linked to openness and extraversion, and can lead to greater creative performance. *Journal of Personality*, *87*(2), 252–266. https://doi.org/10.1111/jopy.12387

Scaccia, A. (2024, July 3). Everything you need to know about serotonin. *Healthline*. https://www.healthline.com/health/mental-health/serotonin

Smith, W., Powell, E. K., & Ross, S. (1955). MANIFEST ANXIETY AND FOOD AVERSIONS. *Journal of Abnormal and Social Psychology*, *50*(1), 101–104. https://doi.org/10.1037/h0049253

Spence, C. (2022). What is the link between personality and food behavior? *Current Research in Food Science*, *5*, 19–27. https://doi.org/10.1016/j.crfs.2021.12.001

Trå, H. V., Volden, F., & Watten, R. G. (2022). High Sensitivity: Factor structure of the highly sensitive person scale and personality traits in a high and low sensitivity group. Two gender-matched studies. *Nordic Psychology*, *ahead-of-print*(ahead-of-print), 1–23. https://doi.org/10.1080/19012276.2022.2093778

Wu, K. (2023, January 8). 9 things that your food might say about your personality. *Psychology Today*. https://www.psychologytoday.com/ca/blog/the-modern-heart/202301/9-things-that-your-food-might-say-about-your-personality