Artificial Intelligence Opinion Survey

DATA 490 Independent Study

Mentor: Dr. Nicholas Dietrich
Aadarsha Gopala Reddy, Darren Lo, Ethan Love, Jose Mancilla, Santo Sumo

April 20, 2023

Contents

1. Load Data

```r
library(tidyverse)
library(ggplot2)

# Load data
edu <- read.csv("prolific_edu.csv")
health <- read.csv("prolific_health.csv")
retail <- read.csv("prolific_retail.csv")
tech <- read.csv("prolific_tech.csv")
```

2. Data Cleaning

```r
# Combine data into one data frame after mutating Age to be one data type
edu <- edu %>% mutate(Age = as.character(Age))
health <- health %>% mutate(Age = as.character(Age))
retail <- retail %>% mutate(Age = as.character(Age))
tech <- tech %>% mutate(Age = as.character(Age))

combined <- bind_rows(edu, health, retail, tech)

# Keep only rows with United States as the Country of residence
combined <- combined %>% filter("Country of residence" == "United States")
```

3. Data Exploration

The columns in the dataset are:

- Submission id: Unique identifier for each submission

- Participant id: Unique identifier for each participant
- Status: Status of the submission
- Started at: Time the survey was started
- Completed at: Time the survey was completed
- Reviewed at: Time the survey was reviewed
- Archived at: Time the survey was archived
- Time taken: Time taken to complete the survey
- Completion code: Completion code for the survey
- Total approvals: Total approvals for the survey
- Employment-sector: Employment sector of the participant
- Age: Age of the participant
- Sex: Sex of the participant
- Ethnicity simplified: Ethnicity of the participant
- Country of birth: Birth country of the participant
- Country of residence: Residence country of the participant
- Nationality: Nationality of the participant
- Language: Language of the participant
- Student status: Is participant a student?
- Employment status: Is participant a student?