**Machine Learning Project # 2: Public Comment**

Professor Terence Johnson

Group 6: Nina Ysabel Alinsonorin & Daphne Pfoser

The U.S. Census Bureau is currently seeking public input and comments on its proposed information collection activities. This request for public comments is an opportunity for individuals and organizations to provide valuable insights and feedback on the methodology, assumptions, and estimates related to the proposed collection, particularly regarding the Sexual Orientation and Gender Identity (SOGI) test. The Census Bureau encourages public engagement to ensure that the information collected is accurate, relevant, and aligned with the diverse needs of the population.

The proposed information collection through the SOGI survey is crucial for the proper functions of the Census Bureau, aligning with its mandate to provide comprehensive and inclusive demographic data. The census plays a fundamental role in shaping the nation's trajectory by informing decisions on resource allocation and community development. It aids in determining where to build essential infrastructure, such as schools, hospitals, and public works, and influences the distribution of federal funds to states and localities. The SOGI data has practical utility, as demonstrated by its potential implications in various areas. For instance, understanding the healthcare needs of transgender individuals can guide healthcare planning, ensuring access to gender-affirming care services. Additionally, the data can inform the allocation of resources for social services, educational facilities, fair resource distribution, and targeted economic development initiatives, promoting inclusivity and addressing disparities.

Regarding the accuracy of the time and cost burden estimate, the validity of the methodology and assumptions used require scrutiny. The estimated time per response, particularly 40 minutes for the average household questionnaire, may be questionable, as the general U.S. survey, which covers more topics, takes about 52 minutes (as given by the Bureau’s website). The methodology, while consistent with the format of the general survey, may need reassessment to align better with the actual time required for the SOGI survey. The assumption of a $0 cost to the public is accurate, considering the provided clarification that it excludes the cost of respondents' time and that there is a $0 cost to the public for the general census survey. Furthermore, the validity of assumptions related to gender identity, specifically the omission of a 'don't know' option for sexual orientation, should be revisited. Including an option like 'questioning' could enhance the accuracy and inclusivity of the survey while avoiding non-response answers, which could have an answer if provided.

To improve data collection from the American Community Survey’s four models (the Internet, paper, in-person interviews, and over the phone), ordinary people and the Census Bureau have proposed suggestions. For the Internet model, a suggestion was to simplify the user interface so that the respondents would have an easier time completing the survey. More specifically, there should be clear navigation signs that help the user toggle back and forth between questions. The last ordinary people's suggestions are to have bigger text, a minimum number of colors, and a color-blind-friendly palette. The next suggestions are about word choice and language. Word choice should be very simple so that every person can understand. Secondly, the survey should be done in English and Spanish since the respondents are from the US and Puerto Rico. Next are the specific suggestions that the Census Bureau has made for gender identity and sexual orientation questions. The first one is splitting the gender identification question into two parts: the gender assigned at birth and the present gender. Next is not referencing the birth certificate in the gender assigned to the birth question to be culturally sensitive to non-English respondents who might not have their certificate. After, include “X” for nonbinary in the gender assigned to the birth question. Next is adding “non-binary” as an answer option to the current gender question. Finally, when asking about the sexual orientation of the person, the response "I don't know” is excluded to reduce the nonresponse answers. Enhanced quality comes from the more accurate expression of identity among diverse groups in the survey. Improved utility arises from the numerous inclusive answer options and results in a more meaningful data group and results. Finally, clarity comes from the fact that questions are easy to understand, and there are a lot of answer options that will make it clear which one the respondents should choose. Flashcards and numbered responses were suggested to ensure accurate responses. This will then produce usable data for analysis, and respondents will be trained on how to use it.

The Census Bureau is aware of the burden associated with completing surveys and is thus trying to come up with technological ways to reduce it. Several technological methods include the integration of AI chat boxes to help out with respondents’ questions, interacting widgets such as process bars or rewards to make the process entertaining, the option of a video response instead of typing, and finally, the option to pause and come back to the survey. Another helpful technology tool would be automated collection techniques. These include online survey forms such as Google Forms or Survey Monkey. Others are sending surveys via email along with automatic reminders. The last and probably best option would be to just scrape information from public websites without the need to do a survey.

In conclusion, the SOGI survey is an important tool that the Census Bureau uses to collect demographic data. Its influence is evident in the decision-making regarding the locations and funding of essential structures. The survey’s practical utility is emphasized by its future implications, particularly in healthcare for transgender individuals. However, timing accuracy and the cost burden estimate, along with the methodology, require careful examination. The ultimate goal is to improve the survey’s quality, utility, and clarity in order to make it a credible and valuable source of information.