# AI Survey Analysis: Effectiveness Report

Below the following report is a prototype of the survey analysis tool used on a fictitious, AI-generated response set of 100 respondents. Please see the “Notes” section for more information about the tool and sample data.

## Executive Summary

This AI survey analysis tool was developed to make response analysis more accessible by making it easier for teams without advanced data analysis expertise to assess non-quantitative survey responses. This report analyzes its effectiveness using five major criteria: (1) Summarization and Synthesis, (2) Sentiment Analysis, (3) Outlier Analysis, (4) Demographic Analysis, and (5) Ease of Use.

## 1. Summarization and Synthesis:

**Proposal Idea:**

The proposal emphasized the tool’s ability to accurately distill responses into a short paragraph.

**Effectiveness:**

The AI tool correctly identified that there was a mix of positive and negative feedback about the project and cited the sources of negative and positive feedback. However, it failed to identify whether the overall feedback was positive, negative, or mixed. Though sentiment is further explored in the second criterion, a summary of respondent perception is an important facet of the overall summary. Nevertheless, the tool correctly identified major points.

## 2. Sentiment Analysis

**Proposal Idea:**

The proposal posited that this tool would be able to analyze respondent sentiment outside of overall trends. This is intended to highlight the emotion behind responses, not just their content, and is a task uniquely suited to the unbiased nature of AI (in lieu of a data scientist).

**Effectiveness:**

The AI tool correctly identified mean sentiment as being generally neutral to slightly positive and proceeded to detail sources of positive and negative sentiment. This section looks similar to the synthesization summary. The tester suspects that this is because the AI-generated responses are not as descriptive as real ones would be and don’t provide the emotional cues that humans would. For example, a human might say they “didn’t like” the project, whereas the tool must infer sentiment from emotionless descriptions about the project like “it wasn’t on time” or “it exceeded expectations,” leading to similar responses in the summary (which is meant to assess project feedback) and sentiment analysis (emotional summary). Further testing is needed with real data, but the tool’s ability to infer sentiment from emotionless response content is a promising sign.

## 3. Outlier Analysis:

**Proposal Idea:**

The project will note outliers so that especially bad experiences can be addressed and their root causes found, while especially good experiences can be studied and replicated. Because means are not statistically representative of the entire dataset, this section meant to catch extremes is an important part of proper analysis.

**Effectiveness:**

The tool’s effectiveness was acceptable. It noted that there were some extremely positive and negative responses, with examples of each. It explained the results’ significance, which was especially helpful. Through the program’s iterations, the tester saw that it chose to quantify extremes by finding sentiment outliers. In real-life application, this portion should detail responses that brought up concerning behavior or patterns within a team as well as responses that provide examples of very good processes. More testing is needed with real data to see if it would catch these specific types of responses. However, the prototype’s ability to correctly identify emotional outliers is a promising sign.

## 4. Demographic Analysis:

**Proposal Idea:**

This analysis segments respondents based on demographic features to best analyze responses within certain blocks, as it is likely that respondent features influence the experiences they had with a given survey subject.

**Effectiveness:**

In this dataset, respondents were separated by seniority with three levels: frontline workers, middle management, and upper management. While a basic summary was given, the tool failed to recognize some important information from within the dataset, such as what issues primarily concerned different demographics (ex: upper management was concerned with budget adherence). This portion needs to be refined to be more descriptive.

## 5. Ease of Use

**Proposal Idea:** this AI tool is meant to be intuitive and make data analysis of non-quantitative data accessible to laypeople.

**Effectiveness:**

The tool is easy to use and requires no prior coding experience. Many survey tools, including the free *Google Forms* application, can automatically update a spreadsheet as responses come in. After all responses are collected, the spreadsheet can be easily uploaded to the survey analysis tool and paired with the descriptive query. The ensuing report is well-organized and short, but descriptive and easy to read. However, using this summary effectively still requires some statistical knowledge (see “Conclusion”).

## Notes

The prototype has been refined over several iterations. Earlier iterations preferred using quantitative measures when rating the data, like tracking the number of times a phrase was said throughout the survey and rating sentiment on a scale of -1 to 1, then noting outliers based on the average rating and standard deviation. The format the AI tool used for this was slightly neater than the final prototype, but the information therein was overall less descriptive. In future testing, it would be prudent to find a way to maintain sentiment rating alongside the paragraph structure of the sentiment analysis deliverable, but otherwise keep the paragraph format.

Additionally, it is important to note that the data used to test this tool was itself AI generated, so it lacked the variability, specificity, and errors that would have been made by human respondents. The choice to use AI data despite these drawbacks was made for several reasons. First, the tester did not have access to a real dataset or a pre-existing survey analysis tool to benchmark performance. The AI that created the set was able to synthesize many responses quickly and specify what answers the new tool was expected to give based on the data it provided. Additionally, using AI removed any subconscious bias the tester may have had towards making data perfectly suited for the new tool.

## Conclusion

Overall, the prototype performs acceptably in the tested categories. It is a useful tool that makes soliciting descriptive feedback more feasible for teams without data analysis professionals. However, there are some things to consider.

First, the prototype needs further testing with real datasets to gauge performance, especially those larger than 100 responses. In the sample data, all three chosen demographics were represented fairly evenly, which is somewhat unrealistic. It is far more likely that frontline workers would vastly outnumber more senior workers. In large, more realistically balanced sets, it is possible that the sentiments expressed by minority demographics would be understated in sections other than “Demographic Analysis.”

Additionally, some statistical knowledge is still necessary to effectively use this tool. Because data is gathered in-house, survey creators must know how to collect good data for the analysis to be useful. If the underlying survey is optional, then the tool’s users must know how to account for the fact that one’s decision to respond may be influenced by their experience with the survey subject, and the data is non-representative of overall sentiment. Additionally, answer length may also reflect bias, and further testing is necessary to determine whether the tool places more weight on longer responses.

In conclusion, the AI survey analyzer is a useful tool to lower the minimum statistical skill required and cost of qualitative survey response analysis, but some basic statistical knowledge is still necessary for effective use.

# Prototype

## Input Prompt:

A close-up of a blue background

Description automatically generated

## Output:

A white paper with black text

Description automatically generated

(Copilot, Personal Communication, November 22, 2024)