OJO, C

ANALYSIS CRITIQUE

ANALYSIS CRITIQUE

Introduction

This is a brief evaluation of the data associate's *chatflow* analysis document. The evaluation assumes that

- Members of the Digital Marketing Team & Contact Centre Team, herein the teams, are the document's audience.

- Within Proximity Designs it is these teams that use *Many-Chat*, therefore their members are the organisation's internal *ManyChat* experts.

- Thus far, the data associate has not shared any analysis results with either team.

- The data associate intends to share the analysis findings via a presentation session; considering the document's format.

- Each team's members have varying levels of experience at Proximity Designs, and have minimal knowledge of statistics & mathematical modelling.

Analysis Feedback

Background & Audience

The digital marketing & contact centre teams are familiar with *ManyChat*. Hence, instead of **slides** 3, 4, 5, & 6, consider two or three background slides that illustrate the associate's understanding of the teams' concerns, alongside the mechanics of a *chatflow*.

Problem Statement, Goals, & Objectives

The problem statement of the analysis, and the goals & objectives thereof, are not stated. These items are critical because they are the fundamental guides of an analysis. A problem statement summarises an overarching issue in focus, and should be informed by the concerns posed by the teams, as illustrated in the background slides. The analysis' goal/s should outline the purpose of the analysis in relation to the problem statement. For example, if the problem statement focuses on incomplete *chatflow* sessions, an analysis goal might be to

> Identify the probable root cause/s of incomplete *chatflow* sessions.

Usually, an analysis focuses on one or two goals only. Finally, each goal is investigated via a set of objectives. For example, an accompanying objective of the goal above might be to

> Investigate whether there is a statistically significant overlap between the *LiveChat* enquiries of users that do not complete a *chatflow* session, and the content of the skipped content.

This suggestion assumes that the customers are not shown a flow diagram of steps at the beginning of a *chatflow* session.

Once the problem statement, goals, and objectives are clear, consider replacing the "challenges we anticipated'' slide with a slide that outlines these items; and re-write the methodology page in line with the objectives being investigated. Additionally, consider re-writing the title; especially because *AgronomyBot* seems unrelated.

The Analysis

The concerns posed by the teams suggests at least two overarching, but interconnected, lines of investigation.

Foremost, the digital marketing team is concerned with the volume of incomplete *chatflow* sessions, i.e., sessions wherein the users opt for obtaining more farm practice details via *LiveChat*, rather than finishing a *chatflow* session.

Second, the contact centre team is concerned about the volume of *LiveChat* enquiries it receives during digital campaigns; this hints at the interconnection. The team's hypothesis is that the number of enquiries increases during digital campaigns. The interpretation of this hypothesis, which should be discussed/confirmed with the team, is

> If the *LiveChat* enquiries of *digital campaign periods* & *campaign free periods* - of the same length - are compared, (number of enquiries)/(number of *chatflow* sessions) **(1)** is higher during a digital campaign period compared to a campaign free period, and **(2)** increases with number of campaigns per period.

These concerns/hypotheses suggests the need to investigate, at least,

- Why users do not complete *chatflow* sessions.

- Whether the ratio of *LiveChat enquires* to *chatflow sessions* varies with the absence/presence of digital campaigns.

However, the first point of the methodology page, **slide 8**, suggests that the predominant focus of the associate's analysis is → users that complete a *chatflow* session during campaign periods only. The slides that focus on the analysis results, **slide 9** onward, reflect this suggestion. However, in contrast to **slide 8**, their footnotes do not stipulate *chatflow* session completion.[1]

[1] Slide 17 does not have a footnote.

Altogether, consider re-visiting the foci of the analysis in relation to the concerns/hypotheses of the teams; this is akin to addressing the missing problem statements, goals, and objectives. At present, the analysis does not yet investigate the concerns/hypotheses raised. **In general:**

Slides 9 & 10

These slides focus on users that complete a *chatflow* session per digital campaign, they compare the number/proportion of enquires made by users across campaigns. It is stated that some campaigns have a *relatively low number of participants*, but this statement cannot be evaluated in the absence of participants counts.[2] Altogether, the slides do not focus on the teams' concerns/hypotheses.

[2] For comparison purposes **slide 10** is the more appropriate graph, especially if separated into graphs by *chatflow* steps counts.

Slides 11, 12, 13, 14, & 15

The purpose of each slide is unclear, hence consider excluding these slides.

Setting purpose aside, the graph of **slide 11** is probably inappropriate vis-à-vis the underlying data, e.g., it disregards the varying number of *chatflow* steps per *chatflow*. And, including the definition of **link click** in **slide 13** would clarify the population in focus.

These slides focus on a very useful investigation, but there are 2 caveats.

Foremost, the analysis only focuses on the users that complete a *chatflow* session during a digital campaign. Second, the categorisation was performed manually.

Expanding this investigation to include - **(a)** all users that participated in a digital campaign during a specified study period, and **(b)** a second study period, of the same length, but free of digital campaigns - will allow for much richer, and more relevant, root cause investigations. Especially if *LiveChat* enquiries categorisations are conducted via word clustering algorithms.[3]

Two investigation examples, in relation to users who do not complete a *chatflow* session but proceed to a *LiveChat* enquiry, are → is there a significant overlap between the content of *chatflow* steps

- skipped, and the *LiveChat* enquiries of these users? [Impatience? This might be addressed via *chatflow* design adjustments.]

- prior to skipping, and the *LiveChat* enquiries made by these users? [This might indicate unclear content.]

The data required for this is available, as noted by point 3 of *a note on the data available.*

To investigate whether there is a relationship between

1. the number of users that made an enquiry
   after completing a *chatflow* session, **and**

2. the total number of users that completed a
   chatflow session

during a digital campaign, it is much more appropriate to consider rates instead of raw numbers because each campaign has a different number of participants.[4] Additionally, the outlined linear regression model is probably inappropriate because

[3] A helpful text is *Leskovec, J., Rajaraman, A., & Ullman, J. (2020). Mining of Massive Datasets (3rd ed.). Cambridge: Cambridge University Press.*

[4] Per case, point, the rate is calculated by dividing by the total number of digital campaign participants.

- There are varying number of *chatflow* steps per *chatflow*.

- The digital campaigns probably occurred over different lengths of time.

- The general knowledge level per farming practice probably varies.

To directly, or indirectly, account for these variables, and other unknowns, consider multilevel models.[5]

[5] A very helpful text is *Gelman, A., & Hill, J. (2006). Data Analysis Using Regression and Multilevel/Hierarchical Models (Analytical Methods for Social Research). Cambridge: Cambridge University Press*

Language, Grammar

Consider reviewing the grammar of each page via a text editor that has a grammar inspection tool; I always use such tools when writing. Please, do contact me for tools recommendations. Additionally, once ready, I suggest organising at least two grammar review session - one with me, and another with other colleagues.

Further Suggestions

Prior to the presentation consider sharing a brief summary that outlines the goal & objectives of the analysis, and anything else that would make it easy to understand the presentation. A grammar review session will suffice before sharing.