LabFCC Team at Rest-Mex 2026: A Synthetic Corpus Based on Tourist Typologies

Erick Barrios González[1,*,†], Danna Patricia Riveroll Martínez[1,†], Alan Kenath Sodi Hernández[1,†] and Mireya Tovar Vidal[1,†]

[1]*Benemerita Universidad Autonoma de Puebla, Av. San Claudio y 14 Sur, Ciudad Universitaria, 72592 Puebla, Pue., México*

Abstract

In the field of tourist opinion analysis, the limited availability of specialized corpora presents a challenge for the development and evaluation of robust models. In this context, the present work explores the creation of three synthetic tourist review corpora based on Plog's theory of tourist typologies. To this end, three text generation strategies were designed and evaluated using large language models. The first, corresponding to the base proposal, is based solely on Plog's theory of tourist typologies. The second incorporates categories of tourist attractions and Likert-scale ratings, while the third is based on iterative prompt refinement. These two latter strategies also rely on Plog's theory of tourist typologies. From these strategies, multiple synthetic corpora were generated, which were analyzed and compared to identify the factors contributing to the production of more representative and useful data for subsequent tasks. The results show that the strategy based on systematic prompt refinement produces higher-quality corpora, achieving a Track Score of 0.5032 on the corpus named submission v3. Its performance indicates that improving the prompts provided to the language model has a positive impact on the quality of the generated synthetic corpora.

Keywords

Synthetic Data Generation, Large Language Models, Prompt Engineering, Tourist Typology

1. Introduction

Artificial intelligence (AI) in recent years has redefined trends in content creation, as it has become increasingly common for internet content to be generated by AI. In the field of Natural Language Processing (NLP), the synthetic generation of Spanish text has emerged as an alternative solution for the tourism industry.

In tourism, online reputation is the most critical asset influencing consumer decision-making, offering deeper insights into all aspects of tourist travel while driving innovation and tourism development [1]. The automatic detection of polarity in user reviews the process of classifying opinions as positive, negative, or neutral is a fundamental task in Natural Language Processing (NLP). However, training high-performing supervised models faces challenges related to labeled data scarcity and class imbalance, particularly in specific market niches or regional variants of Spanish [2, 3, 4, 5].

For this reason, REST-MEX 2026 proposes an evaluation strategy centered on synthetic data generation. The challenge involves developing generative models capable of producing original tourist reviews about Mexico's Magical Towns, accompanied by explicit polarity labels. The primary objective is to assess whether a completely synthetic dataset can serve as a substitute or complement to real data for training sentiment analysis systems [6]. This task forms part of IberLEF 2026, a shared evaluation campaign for Natural Language Processing systems targeting Spanish and other Iberian languages, held in the context of the International Congress of the Spanish Society for Natural Language Processing (SEPLN), whose purpose is to promote new research challenges and advance the state of the art in processing, understanding, and automatic generation of text [7].

IberLEF 2026, September 2026, León, Spain

*Corresponding author.

†These authors contributed equally.

erick.barriosgo@correo.buap.mx (E. Barrios González); rm202455570@alm.buap.mx (D. P. Riveroll Martínez); sh202473505@alm.buap.mx (A. K. Sodi Hernández); mireya.tovar@correo.buap.mx (M. Tovar Vidal)

0000-0002-2077-7541 (E. Barrios González); 0000-0002-9086-7446 (M. Tovar Vidal)

This article examines the technical challenges and strategic opportunities associated with generating Spanish-language tourist texts, based on the REST-MEX 2026 event [6]. For this work, critical aspects such as the following have been explored:Preservation of lexical richness: Managing regionalisms and dialectal variants of Spanish within global contexts; Factual accuracy: Control mechanisms to mitigate "hallucinations" in the description of infrastructure and service.

2. Related works

The generation of synthetic data has been addressed in previous studies [8, 9], yet it is widely acknowledged that the use of synthetic data may lead to biases that can ultimately affect our results when training models with such data. For this reason, the integration of mixed methods and triangulation with interviews, surveys, or observations is recommended to ensure that generated texts remain grounded in real-world contexts. To mitigate biases in synthetic data, it is advised to integrate mixed methods, such as transparency and accountability, methodological rigor, ethical reflexivity, and epistemological integrity [8].

One of the most well-known approaches to contextualize this work is through tourist typology, whose literature has evolved from the 1970s to the present day [10, 11, 12]. Within the field of tourism, Plog's [11] tourist typology provides a framework for applying synthetic data, as it enables the segmentation of tourist behaviors based on personality, a key factor for generating validated synthetic data.

The theory underpinning this work originates from Stanley Plog's 1970 study [11], which posits that tourist preferences and behaviors are closely linked to individual personality. Plog categorized tourists along a psychographic continuum, with psychocentric and allocentric as its extremes. Psychocentric tourists prefer familiar, safe destinations with well-developed tourism infrastructure, seeking predictable and familiar experiences. In contrast, allocentric tourists are characterized by their interest in exploration, adventure, and discovering less-visited locations, demonstrating greater tolerance for uncertainty and cultural differences.

In his 2002 article [13], Plog revised and strengthened his original proposal using empirical evidence gathered from thousands of American travelers. One of the main modifications involved replacing the terms psychocentric and allocentric with more accessible concepts: dependable (conservative or security-dependent) and venturer (adventurous or explorer). The study demonstrated that personality is a better predictor than income for understanding travel frequency and activities undertaken during vacations.

One of the most recent works on Plog's theory [10] indicates that this theory has shortcomings in accurately predicting the destinations travelers actually visit. Nevertheless, it is important to emphasize that although this theory has gradually lost strength as a predictor of travel decisions, it can still serve as a useful tool for understanding general trends and segmenting markets.

Parallel to the revisions made to Plog's theory, research in psychometrics and the development of measurement instruments has experienced significant advancements. In this context, Jebb, Ng, and Tay [14] conducted a review of the primary methodological developments in the creation of Likert-type scales between 1995 and 2019. The authors note that scale construction should no longer focus solely on question formulation and the calculation of traditional indicators, but instead requires a more rigorous process beginning with a clear and precise definition of the construct to be measured. They also emphasize the importance of ensuring that items are understandable to participants through readability tests and more comprehensive content validation procedures.

The generation of text via prompts has evolved from manual engineering techniques to automated optimization approaches aimed at reducing human intervention and improving the quality of generated outputs. Recent studies demonstrate that language models can produce realistic synthetic data even in specialized domains where access to real data is limited [15, 16], such as healthcare or tourism. In this context, automated prompt optimization methods employ feedback mechanisms, error analysis, and evolutionary strategies to iteratively refine the instructions provided to the model. Moreover, approaches such as Auto-CoT (Automatic Chain-of-Thought Prompting) demonstrate that it is possible

to automatically generate reasoning examples that match or even surpass the performance of manually designed demonstrations [15], adapting to the specific characteristics of each task.

3. Methodology

In this work, a three-stage methodology was employed: (i) analysis of the corpus, (ii) definition of profiles for synthetic data generation, and (iii) generation of the dataset. As a result of this methodology, three datasets were obtained to address the proposed task. The three stages are described below.

3.1. Analysis of the Corpora

An exploratory analysis was conducted on the provided corpus containing real tourist reviews and the synthetically generated corpora.

Evaluating the structural and semantic quality of the data is a critical step to validate the utility of the generation models. To this end, an exploratory analysis was performed based on vector representations (embeddings), using the model *dccuchile/bert-base-spanish-wwm-cased* (BERT base cased in Spanish). The objective is to transform textual information into 768-dimensional vectors that capture contextual relationships.

To facilitate the interpretation of these high-dimensional spaces, the UMAP (Uniform Manifold Approximation and Projection) dimensionality reduction technique was employed, projecting the data into a 3- or 2-dimensional space. This approach allows for the visualization of the topological structure and the degree of cohesion among classes, providing a visual metric for the interpretability of the language and the consistency of the polarity labels.

Additionally, information about the texts was obtained, including the number of texts per class, text length, lexical richness, and diversity.

In this experiment, various metrics of lexical richness and diversity were employed to characterize language use in reviews corresponding to different rating categories. The Moving-Average Type-Token Ratio (MATTR) was incorporated, which calculates lexical diversity over sliding windows and provides more stable and comparable estimates across corpora of different sizes. Complementarily, the Measure of Textual Lexical Diversity (MTLD) allows for the evaluation of lexical diversity while minimizing the effect of document length, which is particularly useful in collections of heterogeneous texts. Finally, Honoré indices were employed, a classic metric of lexical richness that combines information about vocabulary size and word frequency, offering a complementary perspective on the complexity and variety of the language used. Together, these measures allow for a more robust characterization of lexical richness, avoiding reliance on a single metric and reducing biases associated with the size of the analyzed texts.

The general pipeline was as follows: the corpus was preprocessed using spaCy for Spanish, performing tokenization, removal of stop words, punctuation, and numbers, as well as lemmatization of the terms. Subsequently, the documents were grouped by their polarity class, and descriptive statistics were obtained, including the number of documents, average length, median, standard deviation, total number of tokens, and vocabulary size. Finally, a lexical richness analysis was conducted using the MATTR, MTLD, and Honoré metrics, complemented by term frequency analysis and visualizations of class distribution, text lengths, and lexical diversity.

3.2. Synthetic Generation Based on Profiles

From the analysis, a primary strategy (Strategy 1) was defined, from which two complementary synthetic generation strategies were developed after its analysis. Each strategy aims to capture other aspects of language and semantic variability present in real reviews. These strategies seek to explore textual structure while establishing a comparative framework among all corpora.

3.2.1. Strategy 1

To ensure representative variability in language and polarity, a traveler psychography-based approach was employed as the base strategy. Following Plog's tourist personality continuum theory, generation was parameterized based on three distinct profiles:

Allocentric: Characterized by exploratory language, a focus on unique experiences, and greater tolerance for unexpected situations, influencing more nuanced polarity even in low ratings.

Midcentric: The majority of tourists can be classified as midcentric, seeking a balance between novelty and comfort, using standard and comparative vocabulary.

Psychocentric (or Dependent): Focused on safety, familiarity, and conventional services; their reviews tend to be more critical of deviations from expected standards.

3.2.2. Strategy 2

Derived from Strategy 1, fourteen factual, non-geographically-specific tourist attraction categories were selected, including historic centers, archaeological sites, craft markets, beaches, and local cafés. Real toponyms were deliberately avoided to prevent the classifier from developing location-based biases, favoring instead the analysis of the textual content of experiences.

The combination of these categories with Plog's tourist personality continuum theory led to Strategy 2, aimed at representing different tourist behavior profiles. Finally, a five-level semantic scale, inspired by the Likert scale, was defined to measure the sentiment polarity expressed in each experience. This scale spans a full spectrum from highly negative experiences to fully positive recommendations.

3.2.3. Strategy 3

The main focus of Strategy 3 is improving prompts: Many models such as ChatGPT, Gemini, and others have the capability to provide highly informative content, but in most cases, this is limited by poorly implemented prompts. The problem lies in the limited understanding of what yields the best results, a difference that becomes more evident when more complex information is required from AI models.

Starting from Strategy 1, only the midcentric tourist profile was considered, as midcentric tourists represent the largest proportion of tourists within Plog's typology.

3.3. Structuring the Synthetic Dataset

The following outlines the process of creating prompts for each strategy and the three resulting corpora.

3.3.1. Submission v1 - Strategy 1

Submission v1 was generated using Strategy 1, where controlled variations of prompts explicitly incorporated the tourist type (allocentric, midcentric, and psychocentric), the rating, and narrative guidelines regarding the experience—whether service, ambiance, or overall opinion. This approach enabled inducing systematic differences in discursive style, aligned with the theoretical characteristics of each profile. Additionally, the generated texts underwent a manual review to validate coherence, followed by an automated cleaning and structuring phase, where the textual content was normalized and relevant fields were extracted using regular expressions.

This approach aims to ensure that the polarity detection model learns not only keywords (such as positive or negative adjectives), but also discursive patterns associated with different types of consumer expectations and behaviors, thereby enhancing its robustness against unseen real-world data.

A representative example of this approach is as follows: the model was instructed to generate a set of reviews corresponding to a psychocentric tourist with a 1-star rating, specifying that the content must reflect a realistic experience including aspects such as service, ambiance, and an overall opinion.

The prompts were designed to generate 100 reviews per query, 6 queries per class, totaling 3,000 text instances, evenly distributed across the 5 rating classes (1 to 5 stars). The generation process ensured

that each synthetic text preserved the structural triad initially proposed by the task (Rating, Title, and Review).

3.3.2. Submission v2 - Strategy 2

To rigorously mitigate the risk of spurious correlation a critical phenomenon in text mining where a machine learning model erroneously associates a noun or attraction with a fixed rating (e.g., assuming museums always receive low ratings and beaches always high ratings rather than interpreting actual adjectives a random sampling algorithm without replacement was implemented.

The Cartesian product of all possible combinations between the fourteen attractions and the five available sentiment classes was computed. This base matrix was subsequently randomized through a random shuffle to ensure randomness in execution flow and to maintain class balance.

Special attention was also given to the simplicity bias often exhibited by Large Language Models (LLMs) when generating extreme ratings, a well-documented issue in academic contexts where models tend to produce flat or repetitive text (e.g., simply stating "the place was nice" for the highest rating). To counteract this lack of lexical depth, the system prompt was parameterized by increasing the model's sampling temperature to 0.85, enhancing creativity and vocabulary richness. Additionally, constraints were added to compel the AI to provide highly detailed descriptions, featuring substantial arguments, expressions of high enthusiasm, and contextual justifications explaining why the establishment or location met or exceeded the expectations of the assigned tourist profile.

The inference and extraction of texts were fully executed locally using the Ollama API integrated with the open-source Llama 3 model. The process concluded with a data post-processing and curation phase assisted by Python-based regular expressions. This module automatically corrected any responses originally expressed in text format or fractional values into purely integer numerical values ranging from one to five.

3.3.3. Submission v3 - Strategy 3

In Strategy 3, a prompt was created and refined further to fine-tune each query by class. Each class was tailored based on the sentiment assigned to the query, with modifications reflected in the selected attribute types and the intensity sought in the synthetic text. For example, reviews with sentiment 5 incorporated one or more exclamation marks to emphasize the positivity of the review (based on the original dataset provided by "REST-MÉX 2026").

The "Thinking" version of ChatGPT was instructed to generate 33 synthetic texts per sentiment class, ensuring an equal distribution across the five required sentiment classes, for a total of 495 synthetic reviews. A general prompt was developed, incorporating improvements observed in each iteration and text generation.

Over 25 responses were generated per query, with refinements made until a prompt was achieved that encompassed the following information: Domain, Sentiment, Intensity, Main Aspect (cleanliness, service, comfort, location, price, facilities), User Type (solo, couple, family, friends), Narrative Style (narrative, descriptive, critical, emotional), Length, Accommodation Type (hotel, hostel, cabin, boutique), Stay Duration (short, medium, long), Linguistic Attributes (register: formal, casual; regionalism: light, natural; e.g., “the truth”, “worth it”), and Generation Rules.

4. Results

In this section, the results of the analysis of the Rest-Mex 2026 corpus, the generated corpora, and the findings derived from the respective analysis are presented. Finally, the results from this year's event are shown.

Table 1
Document length and lexical diversity metrics across review rating classes.

Class	Docs	Tokens	Vocabulary	MATTR	MTLD	Honoré
1	5,441	184,936	16,236	0.886	364.08	2524.44
2	5,496	186,751	15,729	0.891	401.68	2455.13
3	15,519	463,039	23,877	0.893	394.31	2638.84
4	45,034	1,290,576	39,623	0.897	387.69	2892.60
5	136,561	3,687,955	72,646	0.896	366.64	3217.49

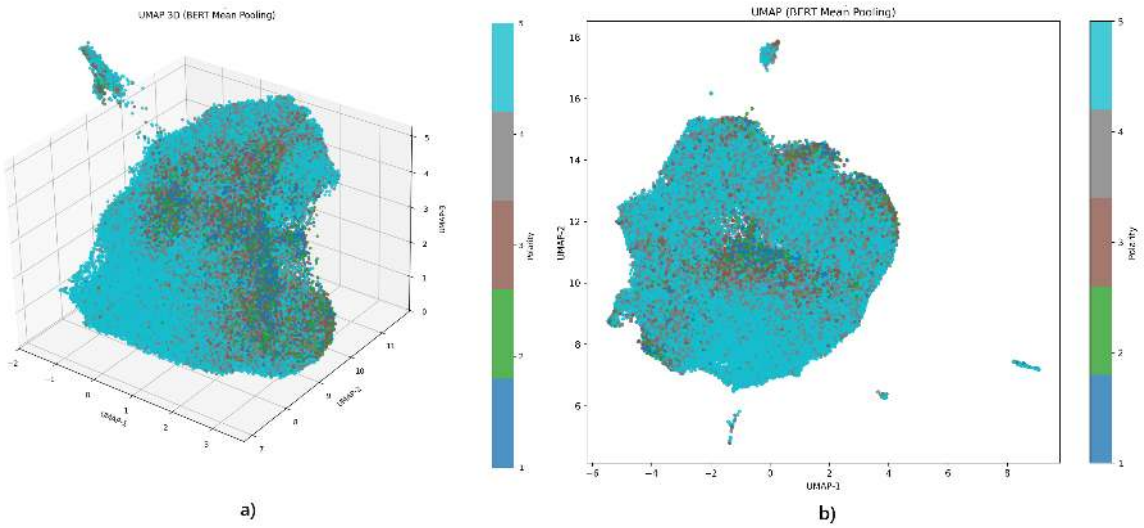

Figure 1: Visualization of polarity-based clustering in the RestMex 2026 corpus. a) Three-dimensional representation, b) Two-dimensional representation.

4.1. Corpus Analysis

For the analysis of the Rest-Mex 2026 corpus, a table summarizing the main characteristics of the corpus has been obtained.

Table 1 summarizes the document length and lexical diversity characteristics of the review corpus across the five rating classes. In addition to corpus size statistics, the analysis includes MATTR, MTLD, and Honoré's statistic, which are widely used measures of lexical richness and are less sensitive to text length than traditional type-token ratio approaches. The results indicate that lower-rated reviews tend to be longer, while lexical diversity remains relatively stable across rating categories.

Regarding the analysis of the two-dimensional and three-dimensional projection generated from dimensionality reduction using UMAP, the results are observed in Figure 1, which reveals a structure where all classes intermingle, suggesting that the corpus structure is complex and relies on multiple components to define the classes. It is evident that the volume of text from the 'Polarity 5' class dominates the space, which also prevents clearly appreciating whether a more distinct separation exists with the other classes.

4.2. Analysis of Synthetic Corpora

Following the corpus provided in Rest-Mex 2026, tables have been created to summarize the characteristics of the other generated corpora.

In Tables 2, 3, 4, the MATTR index shows a clear progression across the three submissions, as this index exhibits a clear increase from Submission 1 to Submission 2, followed by another increase from Submission 2 to Submission 3. The values for Submission 3 are exceptionally high, indicating a high level of lexical variation within each document. Conversely, Submission 1 exhibits more limited diversity,

Table 2
Document length and lexical diversity metrics across review rating classes, submission v1.

Class	Docs	Tokens	Vocabulary	MATTR	MTLD	Honoré
1	460	2,505	400	0.727	48.41	1293.56
2	396	2,070	317	0.680	36.32	1186.47
3	390	1,855	276	0.655	30.08	1236.36
4	366	1,741	228	0.609	25.99	1189.78
5	335	1,665	276	0.635	27.19	1355.80

Table 3
Document length and lexical diversity metrics across review rating classes, submission v2.

Class	Docs	Tokens	Vocabulary	MATTR	MTLD	Honoré
1	360	3,762	991	0.834	110.75	1747.03
2	380	3,648	859	0.799	85.12	1705.92
3	437	4,159	1,011	0.839	139.59	1568.84
4	304	3,003	620	0.771	65.90	1566.11
5	295	4,102	848	0.810	87.60	1543.70

Table 4
Document length and lexical diversity metrics across review rating classes, submission v3.

Class	Docs	Tokens	Vocabulary	MATTR	MTLD	Honoré
1	52	1,544	521	0.948	220.57	1488.42
2	68	1,433	455	0.933	179.13	1425.31
3	66	1,267	447	0.933	186.70	1458.24
4	66	2,027	422	0.911	126.69	1391.01
5	64	1,277	293	0.872	85.13	1218.38

likely due to the brevity of the texts, which restricts the occurrence of varied vocabulary.

Similarly, the MTLD index confirms the pattern observed with MATTR; the values for Submission 3 (Table 4) are double or even quadruple those observed in Submission 1 (Table 2). This suggests that longer texts allow for a more diverse vocabulary to develop before significant lexical repetitions appear. It is particularly interesting that lower-rated classes (1–3) typically exhibit higher MTLD levels than higher-rated classes (4–5). This could indicate that negative or critical reviews tend to be more descriptive and detailed, while positive reviews use a more repetitive vocabulary.

Honoré's index displays a different pattern; Submission 2 (Table 3) systematically presents the highest Honoré values, indicating a higher proportion of rare or low-frequency words. Although Submission 3 (Table 4) possesses greater global diversity (MATTR and MTLD), Submission 2 appears to contain a relatively more specialized or less repetitive vocabulary. This suggests that lexical richness and the presence of infrequent vocabulary do not always evolve in parallel.

Comparing the vocabulary, Submission 2 (Table 3) achieves the broadest vocabulary (1,011 in class 3), surpassing even Submission 3 (Table 4). This aligns with the previously observed high Honoré values and suggests that this corpus contains a greater variety of distinct terms at the global level.

For the UMAP visualization of the synthetic corpora, two-dimensional representations have been chosen. In Figure 2, the structure of Submission 1 is displayed in two dimensions. This submission stands out for having simple and short texts, which likely explains why the class separation is well-defined, unlike in the Rest-Mex 2026 corpus.

In Figure 3, the class separation remains evident despite modifications made to the original proposal.

Finally, in Figure 4, some class separation is visible, but the limited number of texts prevents clear visualization. It is evident that synthetic texts typically exhibit better class separation, whereas in real texts, this separation becomes increasingly ambiguous.

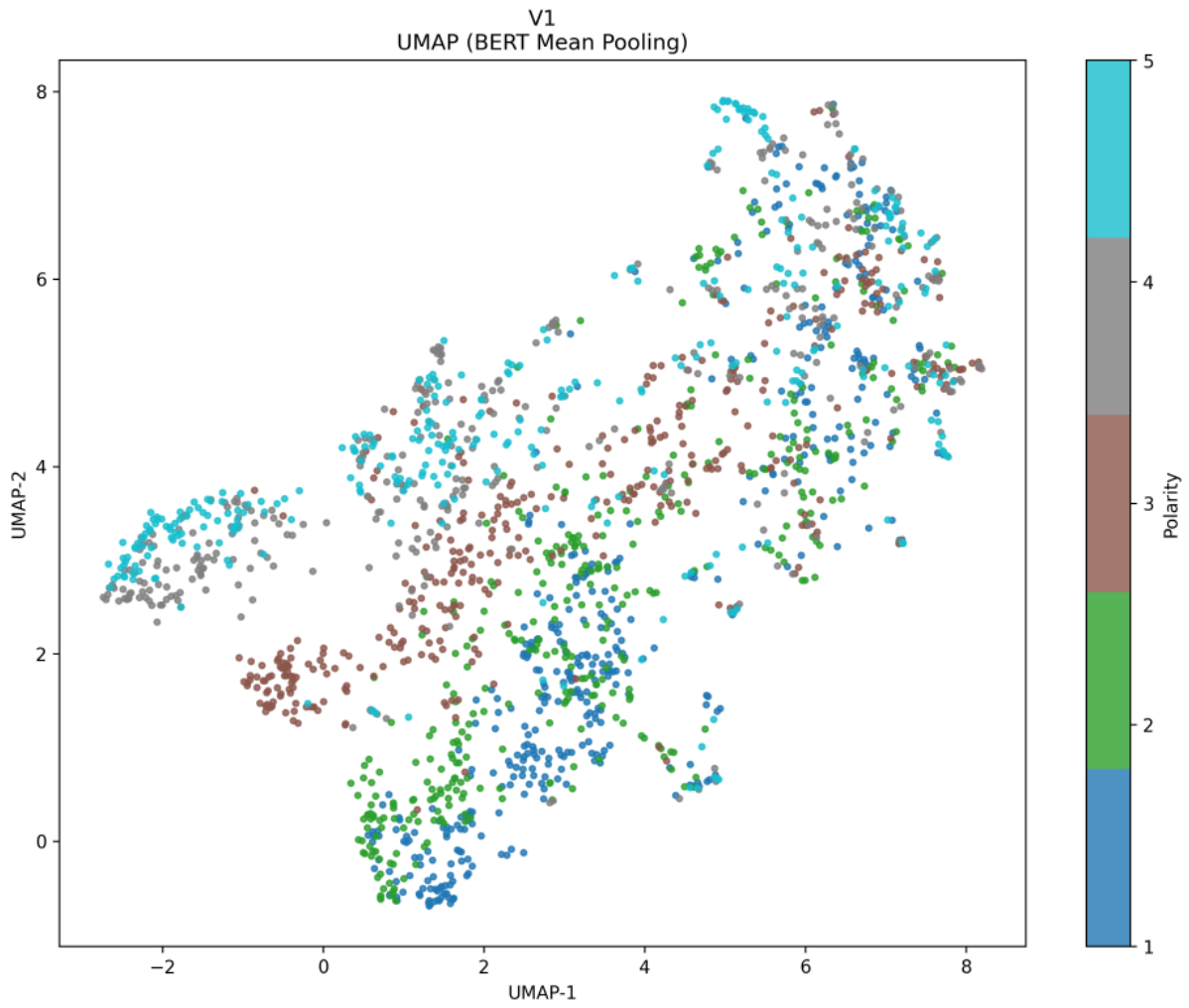

Figure 2: Visualization of polarity-based clustering in synthetic corpus v1.

4.3. Results in RestMex-2026

4.3.1. Submission v1

The first submission achieved a Track Score of 0.4383, with a Macro F_1 of 0.4922 and an accuracy of 70.45%. The model exhibited a clearly biased behavior towards class 5, reaching an F_1 of 0.8510, while facing significant difficulties in the intermediate classes, particularly class 2 (F_1 = 0.2349) and class 3 (F_1 = 0.3338). This suggests that the model tended to correctly recognize the extremes of polarity but struggled to distinguish moderate opinions.

4.3.2. Submission v2

The second version achieved a slight increase in Track Score (0.4415), although the Macro F_1 decreased to 0.4849 and accuracy dropped to 63.72%. The main issue was a significant degradation in class 3, whose F_1 fell to 0.2489. However, it improved precision for class 1 (0.7243) and maintained good performance in class 5 (F_1 = 0.7772). This behavior indicates that the changes introduced favored specific classes but compromised the overall balance of the classifier.

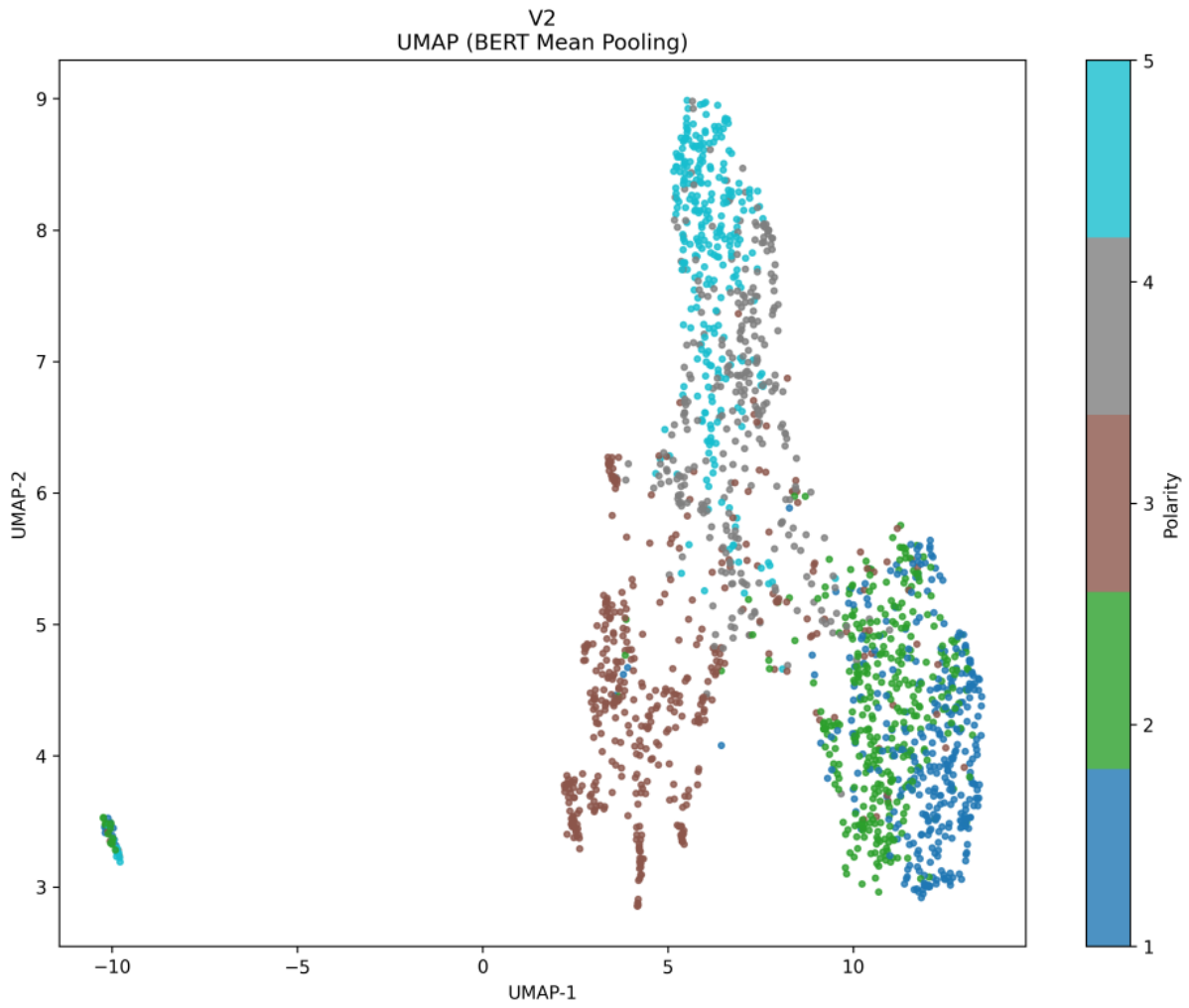

Figure 3: Visualization of polarity-based clustering in synthetic corpus v2.

4.3.3. Submission v3

The third version represented a significant improvement over the previous two. It achieved a Track Score of 0.5032, a Macro F_1 of 0.5456, and an accuracy of 71.26%. All classes showed improved performance, especially classes 1, 2, and 5. The improvement across these classes allowed the Macro F_1 to increase by more than five percentage points compared to earlier versions, making it the best submission.

Additionally, it is necessary to note that the corpus "submission v3" was generated under time constraints, using solely the midcentric tourism typology. This typology corresponds to the profile that allows classifying the largest number of tourists among the three base types defined by Plog. This likely enabled the language model (ChatGPT) to have more information available for generating the various reviews.

4.4. Conclusion

This work explored the creation of a synthetic corpus of tourist reviews based on Plog's theory of tourist typologies. To this end, various synthetic text generation techniques were analyzed, supported by categories of tourist attractions, Likert-scale evaluations, and iterative prompt refinement processes. Additionally, an analysis of the generated synthetic corpora was conducted to better understand and explain their performance.

The best corpus presented in this work, named "submission v3," achieved a Track Score of 0.5032 in

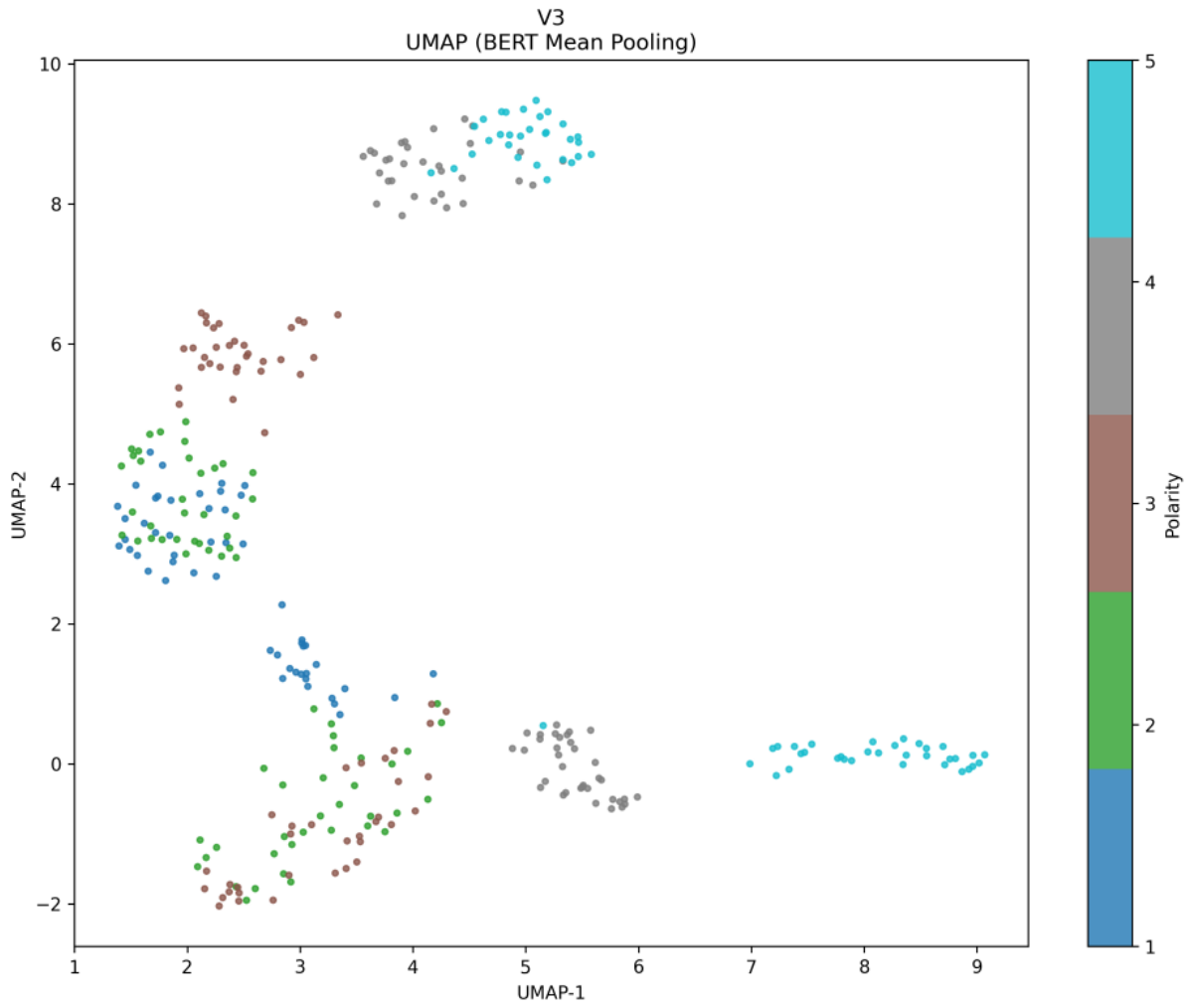

Figure 4: Visualization of polarity-based clustering in synthetic corpus v3.

the evaluation table. This result suggests that a synthetic text generation strategy based on iterative prompt refinement constitutes the most viable alternative. However, this strategy requires additional time for review and adjustment, which complicates its initial implementation.

As future work, it is proposed to extend this strategy to the tourist typologies not explored in this study. Furthermore, it is proposed to increase the number of reviews generated using the strategy that showed the best results and conduct a comparative analysis with other proposals developed by the participants of Rest-Mex 2026.

Declaration on Generative AI

Generative AI tools were used solely to improve the language, grammar, and readability of the manuscript. The authors reviewed and edited all AI-assisted content and take full responsibility for the final version of the manuscript.

References

[1] Q. Li, S. Li, S. Zhang, J. Hu, J. Hu, A review of text corpus-based tourism big data mining, Applied Sciences 9 (2019). URL: https://www.mdpi.com/2076-3417/9/16/3300. doi:10.3390/app9163300.

[2] M. Á. Álvarez-Carmona, R. Aranda, S. Arce-Cárdenas, D. Fajardo-Delgado, R. Guerrero-Rodríguez,

Table 5
Performance comparison between the LabFCC submissions and the top-ranked systems in the REST-MEX 2026 shared task. Systems are sorted according to the official Track Score ranking.

Rank	Run	Track Score	Macro F_1	$F_1(1)$	$F_1(2)$	$F_1(3)$	$F_1(4)$	$F_1(5)$
1	Combination (MLP)	0.6717	0.6471	0.6694	0.3545	0.6360	0.6548	0.9206
2	Combination (Best per class)	0.5515	0.5676	0.6366	0.4498	0.4907	0.4904	0.8127
3	gradiant_submission6	0.5223	0.5564	0.6103	0.3949	0.4734	0.5101	0.7933
4	Combination (Decision Tree)	0.5212	0.5468	0.6584	0.2439	0.5294	0.4343	0.8679
5	gradiant_submission4	0.5178	0.5458	0.6117	0.3939	0.4916	0.4852	0.7467
6	gradiant_submission3	0.5177	0.5489	0.6094	0.3969	0.4734	0.4955	0.7692
7	gradiant_submission1	0.5121	0.5465	0.5839	0.3895	0.4706	0.5052	0.7835
8	Axolotux_qwen_27b	0.5114	0.5451	0.5989	0.3916	0.4738	0.4750	0.7862
18	LabFCC_submission_v3	0.5032	0.5456	0.6015	0.3863	0.4415	0.4447	0.8537
46	LabFCC_submission_v2	0.4415	0.4849	0.5839	0.3229	0.2489	0.4918	0.7772
48	LabFCC_submission_v1	0.4383	0.4922	0.5615	0.2349	0.3338	0.4798	0.8510

A. P. López-Monroy, J. Martínez-Miranda, H. Pérez-Espinosa, A. Rodríguez-González, Overview of rest-mex at iberlef 2021: Recommendation system for text mexican tourism, Procesamiento del Lenguaje Natural 67 (2021). doi:https://doi.org/10.26342/2021-67-14.

[3] M. Á. Álvarez-Carmona, Á. Díaz-Pacheco, R. Aranda, A. Y. Rodríguez-González, D. Fajardo-Delgado, R. Guerrero-Rodríguez, L. Bustio-Martínez, Overview of rest-mex at iberlef 2022: Recommendation system, sentiment analysis and covid semaphore prediction for mexican tourist texts, Procesamiento del Lenguaje Natural 69 (2022) 289–299.

[4] M. Á. Álvarez-Carmona, Á. Díaz-Pacheco, R. Aranda, A. Y. Rodríguez-González, V. Muñiz-Sánchez, A. P. López-Monroy, F. Sánchez-Vega, L. Bustio-Martínez, Overview of rest-mex at iberlef 2023: Research on sentiment analysis task for mexican tourist texts, Procesamiento del Lenguaje Natural 71 (2023) 425–436.

[5] M. Á. Álvarez-Carmona, Á. Díaz-Pacheco, R. Aranda, A. Y. Rodríguez-González, L. Bustio-Martínez, V. Herrera-Semenets, Overview of rest-mex at iberlef 2025: Researching sentiment evaluation in text for mexican magical towns, Procesamiento del Lenguaje Natural 75 (2025) 499–511.

[6] M. Á. Álvarez-Carmona, Á. Díaz-Pacheco, R. Aranda, M.-M. Reyes-Sierra, A. Y. Rodríguez-González, L. Bustio-Martínez, V. Herrera-Semenets, Overview of rest-mex at iberlef 2026: Research on synthetic text modeling for mexican magical towns, Procesamiento del Lenguaje Natural 77 (2026).

[7] A. Bonet-Jover, J. Á. González-Barba, L. Chiruzzo, Overview of IberLEF 2026: Natural Language Processing Challenges for Spanish and other Iberian Languages, in: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2026), co-located with the 42nd Conference of the Spanish Society for Natural Language Processing (SEPLN 2026), CEUR-WS.org, 2026.

[8] F. Ali, Rethinking synthetic data in tourism research: Ethical risks, epistemic shifts, and the rsdu-t framework, Annals of Tourism Research 114 (2025) 104009. URL: https://www.sciencedirect.com/science/article/pii/S016073832500115X. doi:https://doi.org/10.1016/j.annals.2025.104009.

[9] N. M. S. Iswari, N. Afriliana, Enhancing aspect-based sentiment analysis in tourism reviews through hybrid data augmentation, Journal of Applied Data Sciences 6 (2025) 2192–2206. URL: https://www.bright-journal.org/Journal/index.php/JADS/article/view/842. doi:10.47738/jads.v6i3.842.

[10] S. W. Litvin, W. W. Smith, A new perspective on the plog psychographic system, Journal of Vacation Marketing 22 (2016) 89–97. URL: https://doi.org/10.1177/1356766715580187. doi:10.1177/1356766715580187. arXiv:https://doi.org/10.1177/1356766715580187.

[11] S. C. Plog, Why destination areas rise and fall in popularity, Cornell Hotel and Restaurant Administration Quarterly 14 (1974) 55–58. URL: https://doi.org/10.1177/001088047401400409. doi:10.1177/001088047401400409. arXiv:https://doi.org/10.1177/001088047401400409.

[12] S. Prince, Cohen's model of typologies of tourists, in: The SAGE International Encyclopedia of Travel and Tourism, volume 4, SAGE Publications, Inc., 2017, pp. 280–282. URL: https://sk.sagepub.com/ency/edvol/the-sage-international-encyclopedia-of-travel-and-tourism/chpt/cohen-s-model-typologies-tourists. doi:10.4135/9781483368924.n109.

[13] S. C. Plog, The power of psychographics and the concept of venturesomeness, Journal of Travel Research 40 (2002) 244–251. URL: https://doi.org/10.1177/004728750204000302. doi:10.1177/004728750204000302. arXiv:https://doi.org/10.1177/004728750204000302.

[14] A. T. Jebb, V. Ng, L. Tay, A review of key likert scale development advances: 1995–2019, Frontiers in Psychology Volume 12 - 2021 (2021). URL: https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2021.637547. doi:10.3389/fpsyg.2021.637547.

[15] N. Freise, M. Heitlinger, R. Nuredini, G. Meixner, Automatic prompt optimization techniques: Exploring the potential for synthetic data generation, in: M. Kurosu, A. Hashizume (Eds.), Human-Computer Interaction, Springer Nature Switzerland, Cham, 2025, pp. 190–201.

[16] Z. Zhang, A. Zhang, M. Li, A. Smola, Automatic chain of thought prompting in large language models, in: The Eleventh International Conference on Learning Representations, 2023. URL: https://openreview.net/forum?id=5NTt8GFjUHkr.