Datasheet for Pizza Hut Ratings and Reviews Dataset

Chenyiteng Han

April 2, 2024

Dataset Description

Dataset Name

Pizza Hut Reviews and Ratings: Insights from One of Sri Lanka's Pioneer Branches

Dataset Author

Kanchana Gajamuthu

Dataset Source

This dataset is a detailed compilation of customer feedback for the Pizza Hut outlet at Union Place, Colombo, one of the inaugural Pizza Hut locations in Sri Lanka. The dataset is available on Kaggle, provided by Kanchana Gajamuthu. Access the dataset directly at Pizza Hut Reviews and Ratings on Kaggle.

Motivation

The dataset was created to analyze customer experiences at the Union Place, Colombo branch of Pizza Hut, aiming to utilize data science applications such as sentiment analysis, trend identification, and customer satisfaction metrics through ratings and textual reviews.

Composition

The dataset comprises customer ratings and textual reviews, capturing a wide array of experiences at Pizza Hut's Union Place, Colombo branch. It includes columns for:

- `title`: Specifies the Pizza Hut location reviewed.

- `stars`: Customer-provided rating on a scale from 1 to 5.

- `text`: The customer's written review.

Data Collection Process

Data Collection Methods

The data was scraped using Apify, utilizing their Tool for Google Reviews. This method allowed for the collection of customer reviews from the Union Place branch in Colombo, which is one of the oldest Pizza Hut branches in Sri Lanka. The dataset creator has cited the use of Apify under secondary sources.

Data Collection Date

The dataset was mined on 11th March 2024, capturing customer reviews and ratings up to this date. This timestamp is critical for understanding the dataset's context and relevancy, especially for analyses focusing on trends over time.

Data Preprocessing

The dataset has been processed to remove personal identifiers, ensuring compliance with GDPR and maintaining reviewer privacy.

Annotation Process

The dataset presents customer ratings and comments in their raw, unaltered form, with no additional post-processing or annotation applied to the content of the reviews beyond the removal of commenter names to protect privacy. This approach ensures the integrity and authenticity of the feedback are maintained, providing direct insights into customer experiences and perceptions. The only alteration made is the removal of commenter names, a step taken to guard their privacy and adhere to ethical data handling practices. No sentiment labels or other modifications have been added, preserving the original context and nuance of each review.

Ethical Considerations

In compiling this dataset, the highest levels of data privacy have been adhered to. Names of the commenters have been removed, leaving only the content of the reviews to protect user privacy. This approach ensures that the dataset respects the anonymity and privacy of the individuals whose feedback is included.

Data Structure

Schema

The dataset includes the following columns: `title`, `stars`, and `text`, designed to offer insights into customer experiences at the Pizza Hut Union Place, Colombo branch. Below is a sample table showing a snippet of the dataset to provide a clear example of its structure and the type of information each entry contains.

Table 1: Sample Data from the Dataset

Title	Stars	Text
Pizza Hut - Union Place	5	
Pizza Hut - Union Place	5	
Pizza Hut - Union Place	4	
Pizza Hut - Union Place	5	
Pizza Hut - Union Place	4	
Pizza Hut - Union Place	5	
Pizza Hut - Union Place	5	Good place to visit to have some food

3

Table 1: Sample Data from the Dataset

Title	Stars	Text
Pizza Hut - Union Place	5	Had a personal pan chicken tandoori, it was quite on point just like the pizza hut one looks forward to. The staff was courteous and had parking space too.
Pizza Hut - Union Place	5	
Pizza Hut - Union Place	4	

Table 1 is intended as a representative example; the actual dataset contains a more comprehensive collection of reviews and ratings.

Missing Data

The dataset has been maintained in its raw state, and no specific measures have been taken to address or impute missing data. Users of the dataset should be aware of this and consider appropriate methods for handling any missing values according to their analysis needs.

Data Usage

Intended Use

The dataset is intended for applications such as sentiment analysis, trend identification, and measuring customer satisfaction through review analysis.

Expected Update Frequency

This dataset is not expected to be updated. It represents a snapshot of customer reviews and ratings up to the point of its last update.

Limitations

While this dataset provides valuable insights into customer experiences and perceptions at the Union Place branch in Colombo, several limitations should be considered when using and interpreting the data:

1. **Branch-Specific Insights**: The dataset focuses exclusively on one of the oldest Pizza Hut branches in Sri Lanka. Consequently, the feedback and ratings captured may reflect unique aspects of this specific location, such as management practices, customer service quality, and local consumer preferences, which are not necessarily indicative of the Pizza Hut brand as a whole. Users should exercise caution when generalizing findings across other Pizza Hut locations or the broader fast-food industry.

2. **Temporal Scope and Trends**: Given that the dataset represents a snapshot in time without updates, it may not effectively capture evolving trends or changes in customer sentiment over longer periods. Seasonal variations, changes in management or menu offerings, and external factors like economic shifts or public health concerns can significantly impact customer experiences and perceptions. Users analyzing this dataset for trend identification should be aware of its limited temporal range and the potential for these dynamics to influence the relevance and accuracy of their findings.

3. **Lack of Demographic Data**: The dataset's adherence to privacy standards, while crucial for ethical considerations, means that it does not include demographic information about the reviewers. Without this context, it is challenging to analyze how different segments of the population may perceive or experience the service at this Pizza Hut branch. Insights into age, gender, or locality could enhance understanding of customer feedback but are absent from the data collected.

These limitations underscore the importance of a nuanced approach to data analysis and interpretation. Researchers and analysts should consider these factors when drawing conclusions, especially when attempting to extrapolate insights beyond the specific context of the Union Place branch in Colombo. Acknowledging these limitations helps to ensure that interpretations of the data are grounded in the dataset's scope and nature.

Licensing

This dataset is made available under the ODC Attribution License (ODC-By). Users are free to share and adapt the data, provided that appropriate credit is given, and any changes to the dataset are indicated.

Additional Notes

Ethically Mined Data

The dataset adheres to strict privacy standards, including GDPR, to ensure the removal of personal identifiers.

Acknowledgments

This datasheet was created by Chenyiteng Han based on the "Pizza Hut Reviews and Ratings: Insights from One of Sri Lanka's Pioneer Branches" dataset originally posted by Kanchana Gajamuthu. We extend our gratitude to Kanchana Gajamuthu for compiling the dataset and to the platforms that enabled its collection, as well as to Pizza Hut for the use of their logo for illustrative purposes. This does not imply any affiliation or endorsement. Further works by Kanchana Gajamuthu can be explored at Kanchana Gajamuthu's Kaggle Profile.