Customer Satisfaction Analysis Based on Trustpilot Reviews

(Focus: Customer Rating Prediction)

1. Context and Objectives

In today's e-commerce landscape, customer reviews provide crucial insights into service quality. Trustpilot, a Danish consumer platform, hosts millions of reviews about businesses worldwide. These reviews often reflect key factors influencing customer satisfaction, such as delivery delays, product quality, or customer service issues.

Motivation: By analyzing these reviews through text mining, companies can identify operational weaknesses and improve the customer experience. In this two-month project, our data science team (Frank, Sebastian, Mohamed) analyzes English-language reviews for AliExpress (~57,000), Wish (~99,000), and Temu (~14,000) to investigate drivers of customer satisfaction and dissatisfaction.

Primary Objective: The primary objective of this project is predicting customer star ratings based on review texts. Given that star ratings range from 1 to 5, this is primarily addressed as a regression problem to predict continuous ratings. However, classification methods may also be explored to categorize reviews clearly into rating classes.

Project Goals:

- **Rating Prediction:** Predict star ratings using regression models.
- **Text Analysis:** Collect and analyze review texts.
- **Key Theme Extraction:** Identify common customer complaints (e.g., delivery, refunds).
- **Customer Classification:** Classify loyal customers and those frequently submitting negative feedback.
- **AI-Generated Reply Suggestions:** Detect emotionally negative reviews and generate calming, empathetic replies.
- **Dashboard Visualization:** Create an interactive dashboard using Plotly and Streamlit to summarize findings.

2. Framework: Data and Tools

Data Source:

- Trustpilot English-language reviews for AliExpress, Wish, and Temu.
- CSV fields: UserId, UserName, UserCountry, ReviewCount, ReviewRating, ReviewTitle, ReviewText, ReviewDate, ReviewExperienceDate, ReplyText, ReplyDate.

Challenges:

- Highly imbalanced star rating distribution (many 1-star or 5-star ratings).
- Informal and grammatically inconsistent language.
- Emotional and subjective text complicating automated analysis.

Tools & Technologies:

- **Python:** Pandas, NumPy for data manipulation.
- **Web Scraping:** BeautifulSoup to extract Trustpilot data.
- **Natural Language Processing**: NLTK for preprocessing (tokenization, lemmatization, sentiment analysis).
- **Machine Learning:** scikit-learn for regression/classification; potential use of Random Forest, XGBoost, and deep learning methods (e.g., Transformers like BERT for contextual understanding).
- **Visualization:** Matplotlib, Seaborn, Plotly.
- **Deployment:** Interactive dashboard via Streamlit.
- **Version Control:** GitHub (Master branch provided by mentor Kylian Santos).

3. Preprocessing and Feature Engineering

Text Cleaning:

- Lowercasing, removal of HTML tags, URLs, and special characters.
- Stopword removal ("the", "and", "is", etc.).
- Tokenization and lemmatization for word normalization.
- Optional spelling and abbreviation corrections (e.g., "u" → "you").

Feature Creation:

- TF-IDF vectors for converting text into numerical data.
- Sentiment analysis to score emotional tone.
- Features capturing review length, sentence structure, and complexity.
- Metadata features: date, language, experience context.

Rating Prediction:

- Begin with linear regression models on TF-IDF vectors.
- Extend to more complex models like Random Forest or XGBoost to capture nonlinear relationships and improve prediction accuracy. These models are recommended due to their capability to handle complex feature interactions and their robustness in predicting customer ratings.
- Deep learning methods such as Transformer-based models (e.g., BERT) are particularly suitable for this task because they effectively capture linguistic nuances, contextual information, and emotional undertones present in customer reviews.
- Evaluation: Mean Absolute Error (MAE) or Root Mean Squared Error (RMSE).
- Address rating imbalance using downsampling, class weighting, or classification approaches.

Topic Modeling:

- Cluster similar reviews to uncover frequent complaint categories.
- Visualize frequency of identified issues.

Customer Classification:

- Analyze user review frequency (e.g., frequent reviewers).
- Tone analysis for aggressive vs. constructive reviews.
- Identify loyal customers through linguistic cues ("regular customer", "order monthly").

AI-Powered Reply Suggestions:

- Combine predefined response templates with NLP-based analysis.
- Optionally integrate LLM-generated empathetic responses.

Visualization & Dashboard:

- Rating distribution histograms.
- Geographical and temporal sentiment analysis.
- Bar charts of complaint frequencies.
- Positive vs. negative review word clouds.
- Interactive Streamlit dashboard with filtering options.

4. Value of Platform Comparison

Including Temu and Wish alongside AliExpress enhances analytical depth:

- **Benchmarking:** Determine predominant issues per platform (e.g., delivery delays vs. quality concerns).
- **Linguistic Variation:** Helps fine-tune sentiment analysis models.
- **Model Robustness:** Test AliExpress-trained models for generalizability on Temu and Wish.
- **Comparative Dashboard:** Visualize differences

We chose AliExpress, Temu, and Wish because all three operate the same low-cost, cross-border marketplace model, yet differ in market maturity, AliExpress is long-established, Wish is a repositioned veteran, and Temu is a hyper-growth newcomer. Each platform offers a large pool of English Trustpilot reviews (≥ 10 k), ensuring consistent language for NLP while providing a broad sentiment spectrum: Wish skews positive, whereas AliExpress and Temu skew strongly negative. This combination gives us comparable data, shared supply-chain pain points, and contrasting customer sentiment,ideal conditions for meaningful, generalisable customer rating predictions.

AliExpress is the project's central focus, but comparisons with Temu and Wish provide broader insights and validation of analytical approaches.

Conclusion:

This project analyzes customer satisfaction drivers through genuine reviews, aiming to reveal critical service and communication weaknesses. Incorporating multiple platforms highlights cross-platform insights, enhancing data-driven decision-making. Additionally, optional AI-generated reply suggestions illustrate practical customer service improvements. The interactive dashboard and final report offer stakeholders clear, actionable insights from direct customer voices.

5. Exploratory Insights (first-pass statistics)

The following descriptive figures are based on the raw datasets and will be refined after further cleaning.

Wish shows the most positive sentiment: nearly two-thirds of its reviews are 5-star, and the bulk of feedback comes from North America (≈ 62 %), with the United States alone accounting for more than half of all Wish reviews.

Temu is the opposite extreme: a majority (≈ 52 %) of reviews are 1-star; reviewers are predominantly European (≈ 58 %), led by the United Kingdom and the US.

AliExpress also skews negative (≈ 52 % 1-star) but its audience is more geographically mixed—Europe (~48 %), Asia (~21 %) and North America (~20 %).

Across all three platforms, the US and UK consistently dominate the top-country slice, a pattern that reflects the fact that only English-language reviews were analysed. Within that English-speaking subset.