## Title Page

**Text Analytics for Business/Marketing**  
**Text Analysis of ChatGPT User Reviews to Guide the Strategic Launch of the Google Pixel LLM Phone**  
**By**

**Your Name**  
**Submission Date**

Executive Summary

This study is intended to enhance Google's strategic planning for the release of the Pixel LLM Phone, an AI-powered smartphone with a Large Language Model (LLM). This study analyses 40,000 user reviews of the ChatGPT app from the Google Play Store to identify strengths, problems, and trends that can assist position the product in the highly competitive smartphone market. The findings also identify potential possibilities and dangers to help influence decision-making.

The project follows the CRISP-DM methodology, which ensures a methodical approach from understanding the business context to data discovery, cleaning, analysis, and solution deployment. To facilitate deeper analysis, user evaluations were preprocessed by standardising text to lowercase, deleting punctuation and stop words, and tokenising. Using Orange data mining tools, the study examined customer sentiment, key topics, and keywords to identify factors that drive customer interest and satisfaction.

The analysis revealed four key insights:

1. **Positive Keywords**: Words like "helpful," "efficient," and "user-friendly" represent attributes highly valued by users. Emphasizing these features in advertising can help Google maintain a competitive edge.
2. **Negative Feedback**: Issues such as app crashes, bugs, and inconsistent AI responses underline the critical need for rigorous testing and product reliability to meet customer expectations.
3. **Opportunities for Innovation**: Customers appreciate the features like adaptive learning and AI powered responses. These areas represent the opportunities for Google to focus on and enhance the Pixel LLM Phone's AI capabilities.
4. **User Engagement Dynamics**: Reviews containing positive keywords like "helpful" and "amazing" receive the most engagement, and sentiment improves with newer app versions. This suggests a growing customer appreciation for improvements and innovations over time.

Based on these insights, several recommendations are provided:

* Highlight features that users value most in marketing campaigns to drive awareness and interest.
* Address technical issues affecting reliability and stability to ensure a seamless user experience.
* Introduce adaptive learning and personalized features in future product iterations to align with customer expectations.
* Develop a robust feedback system to continually gather and respond to user input.

By implementing these recommendations, Google can position the Pixel LLM Phone as a leader in AI-driven smartphone technology. This approach ensures customer satisfaction, product reliability, and continuous innovation while minimizing risks related to sustainability and market disruption. This report equips Google with actionable insights for data-driven decision-making, paving the way for a successful launch of the Pixel LLM Phone.

Table of Contents

1. **Executive Summary**  
   1.1 Purpose of the Report  
   1.2 Key Findings  
   1.3 Recommendations
2. **Introduction**  
   2.1 Objective of the Analysis  
   2.2 Context and Importance  
   2.3 Scope of the Report
3. **Methodology**  
   3.1 Overview of CRISP-DM Methodology  
   3.2 Workflow and Tools
4. **Data Understanding**  
   4.1 Dataset Description  
   4.2 Key Observations  
   4.3 Initial Insights
5. **Data Preparation**  
   5.1 Steps for Data Cleaning  
   5.2 Sentiment Classification  
   5.3 Preprocessing Techniques
6. **Key Analyses and Insights**  
   6.1 Story 1: Positive Customer Sentiments  
   6.2 Story 2: Negative Customer Experiences  
   6.3 Story 3: Innovative Opportunities from Specific Features  
   6.4 Story 4: User Engagement and Feedback Trends
7. **Recommendations and Conclusion**  
   7.1 Summary of Recommendations  
   7.2 Conclusion
8. **Appendices**  
   8.1 Short User Guide for Orange Workflows
   * Workflow 1: Positive Customer Sentiments
   * Workflow 2: Negative Customer Experiences
   * Workflow 3: Innovative Opportunities
   * Workflow 4: User Engagement Trends  
     8.2 Visual Outputs (Word Clouds, Charts, etc.)  
     8.3 References

Introduction

The purpose of this report is to extract meaningful data from user reviews of the ChatGPT app via text analytics. The findings will assist Google in identifying significant opportunities, obstacles, and emerging themes that are critical to the successful introduction of the Google Pixel LLM Phone, an innovative smartphone powered by a Large Language Model (LLM).\*

**Context**

Artificial Intelligence (AI) is revolutionizing technology by enabling natural, human-like interactions through applications like ChatGPT.\* As Google invests $1 billion in launching the Pixel LLM Phone, strategic planning is essential to address:

* **Positive Experiences**: Features to emphasize in marketing and design.
* **Recurring Challenges**: Issues and pain points that need resolution.
* **Opportunities for Innovation**: Areas that enhance user engagement and experience.
* **Potential Risks**: Concerns related to sustainability, disruption, and data ethics.

The insights from this analysis will empower Google to design a competitive, user-centric product that meets customer expectations while proactively addressing risks.

**The Role of the Storytelling Framework**

A storytelling framework is essential for presenting the results of text analytics in a meaningful and actionable way. It transforms complex data into clear, engaging narratives that:

1. Simplify technical results into practical insights.
2. Align analytical findings with Google’s business goals.
3. Enable stakeholders to make informed, evidence-based decisions.

Each story in this report follows a structured format to ensure clarity and relevance:

* **Problem/Question**: Define the focus of the analysis.
* **Analysis**: Use text analytics to uncover trends and patterns.
* **Insights**: Present the findings in a straightforward, actionable manner.
* **Recommendations**: Suggest practical solutions to leverage opportunities or address challenges.
* **Pressing Questions**: Highlight any sustainability, disruption, or risk-related concerns.

This approach ensures the findings are not just data points but actionable strategies to support the successful launch of the Pixel LLM Phone.

Methodology

This report employs the CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology as a guiding framework to ensure a structured and comprehensive approach to analyzing user reviews. By following its six distinct stages, the analysis is tailored to extract meaningful insights aligned with Google’s strategic objectives for the Pixel LLM Phone. Below is an in-depth breakdown of each stage and its contribution to the report:\*

**1. Business Understanding**

The first stage focuses on clearly defining Google’s goals and the specific questions this analysis seeks to address. The primary objective is to identify actionable insights that will inform the design, marketing, and user engagement strategies for the Pixel LLM Phone. Key business questions include:

* What positive aspects of ChatGPT do users appreciate, and how can these be leveraged in product design?
* What recurring pain points or frustrations exist, and how can they be resolved in the new product?
* What opportunities for innovation emerge from user feedback?
* What risks related to product reliability, sustainability, or ethical considerations should be mitigated?

**2. Data Understanding**

The dataset used in this study comprises 40,000 user reviews of the ChatGPT app collected from the Google Play Store. This stage involved exploring the data to identify critical fields, such as:

* User comments that capture detailed feedback.
* Ratings that reflect overall satisfaction levels.
* Thumbs-up counts indicating the popularity or resonance of specific reviews.

This exploration helped to pinpoint trends and patterns that align with Google’s strategic objectives.

**3. Data Preparation**

Data preparation is most critical task to ensure the quality and reliability of the analysis. The text based data underwent extensive preprocessing, which included:

* **Lowercasing:** Standardizing text to eliminate inconsistencies in case sensitivity.
* **Stopword Removal:** Filtering out common but uninformative words like "and," "the," or "is."
* **Tokenization:** Break down sentence into different words or phrases for easy analysis.
* **Punctuation Removal:** Stripping out special characters that do not add meaning to the data.\*

These steps were essential for creating a clean dataset, enabling accurate and meaningful analysis in later stages.

**4. Modeling**

Text analytics techniques were applied to uncover key insights from the cleaned data. The methods used included:

* **Sentiment Analysis:** Identifying whether user reviews conveyed positive, negative, or neutral emotions.
* **Topic Modeling:** Grouping reviews into common themes to understand areas of user interest or concern.
* **Keyword Extraction:** Highlighting frequently used words and phrases that reflect customer priorities or frustrations.

These modeling techniques provided a comprehensive view of user feedback, helping to address the questions defined in the Business Understanding phase.

**5. Evaluation**

The evaluation stage involved the analytical results to ensure they aligned with Google priorities. Insights were validated by comparing trends and findings against real-world scenario, ensuring that the analysis provided actionable and reliable guidance. Particular attention was given to identifying themes that would directly influence the Pixel LLM Phone’s design, functionality, and marketing strategies.

**6. Deployment**

The final stage of the methodology is deployment, where the findings are presented as four actionable narratives. These stories are designed to address Google’s core strategic needs:

1. Highlighting positive features to emphasize in marketing campaigns.
2. Addressing recurring issues to improve user satisfaction and product reliability.
3. Identifying opportunities for innovation that align with user preferences.
4. Mitigating risks associated with reliability, sustainability, and disruption.

By following the CRISP-DM methodology, this report ensures a clear, structured, and actionable approach to analyzing user feedback, providing Google with the insights necessary for a successful launch of the Pixel LLM Phone.

Data Understanding

This section provides an overview of the dataset used for the text analytics project, which consists of 40,000 user reviews of the ChatGPT app sourced from the Google Play Store. The dataset serves as a critical resource for analyzing user experiences, satisfaction levels, and challenges, offering insights that can guide the launch strategy for the Google Pixel LLM Phone.

**Initial Data Collection Report**

The dataset includes several key fields that facilitate a comprehensive analysis of user feedback. Below is a detailed breakdown:

**Dataset Overview**

1. **Name**: Username of the reviewer.
2. **Rating**: The star rating (1–5), reflecting user satisfaction.
3. **Comment**: The textual content of the review, capturing user sentiment, feedback, and experiences.
4. **Date**: The date the review was posted, enabling temporal trend analysis.
5. **Country**: The country code of the reviewer, limited to 'US' (United States).
6. **Thumbs Up**: The number of likes or engagement the review received.
7. **Review ID**: A unique identifier for each review.
8. **App Version**: The version of the app being reviewed.

**Key Observations About the Dataset**

* **Data Size**: The dataset contains 40,000 records, providing a robust base for analysis.
* **Missing Values**: The "App Version" field has approximately 3,500 missing entries. All other fields are complete and usable.
* **Primary Focus**: The "Comment" field is central to the analysis as it contains rich textual data for sentiment analysis, topic modeling, and keyword extraction.

**Dataset Distribution**

1. **Ratings**: Reviews are distributed across 1–5 star ratings, allowing for the analysis of varying satisfaction levels.
2. **Engagement**: The "Thumbs Up" field measures user engagement, with higher values indicating more impactful reviews.
3. **Temporal Trends**: The "Date" field facilitates understanding of changes in sentiment and engagement over time.
4. **Textual Insights**: The "Comment" field offers a detailed feedback, shedding light on user priorities, challenges, and overall sentiment.

**Initial Data Observations**

* **Positive Reviews**: Common keywords including “helpful,” “efficient,” and “amazing,” highlighting features that users value most.
* **Negative Reviews**: Issues detected such as “bugs,” “crashes,” and “errors” are frequently mentioned, pointing to areas needing improvement.
* **Engagement Trends**: most reviews with higher "Thumbs Up" counts tend to be detailed and positive, underscoring the value of constructive feedback.

Data Preparation

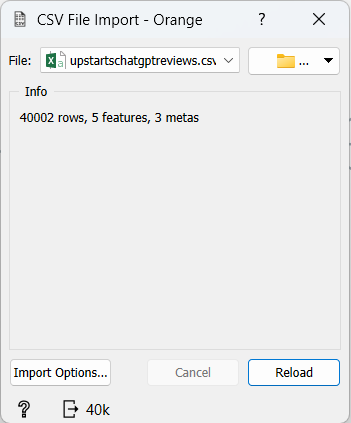
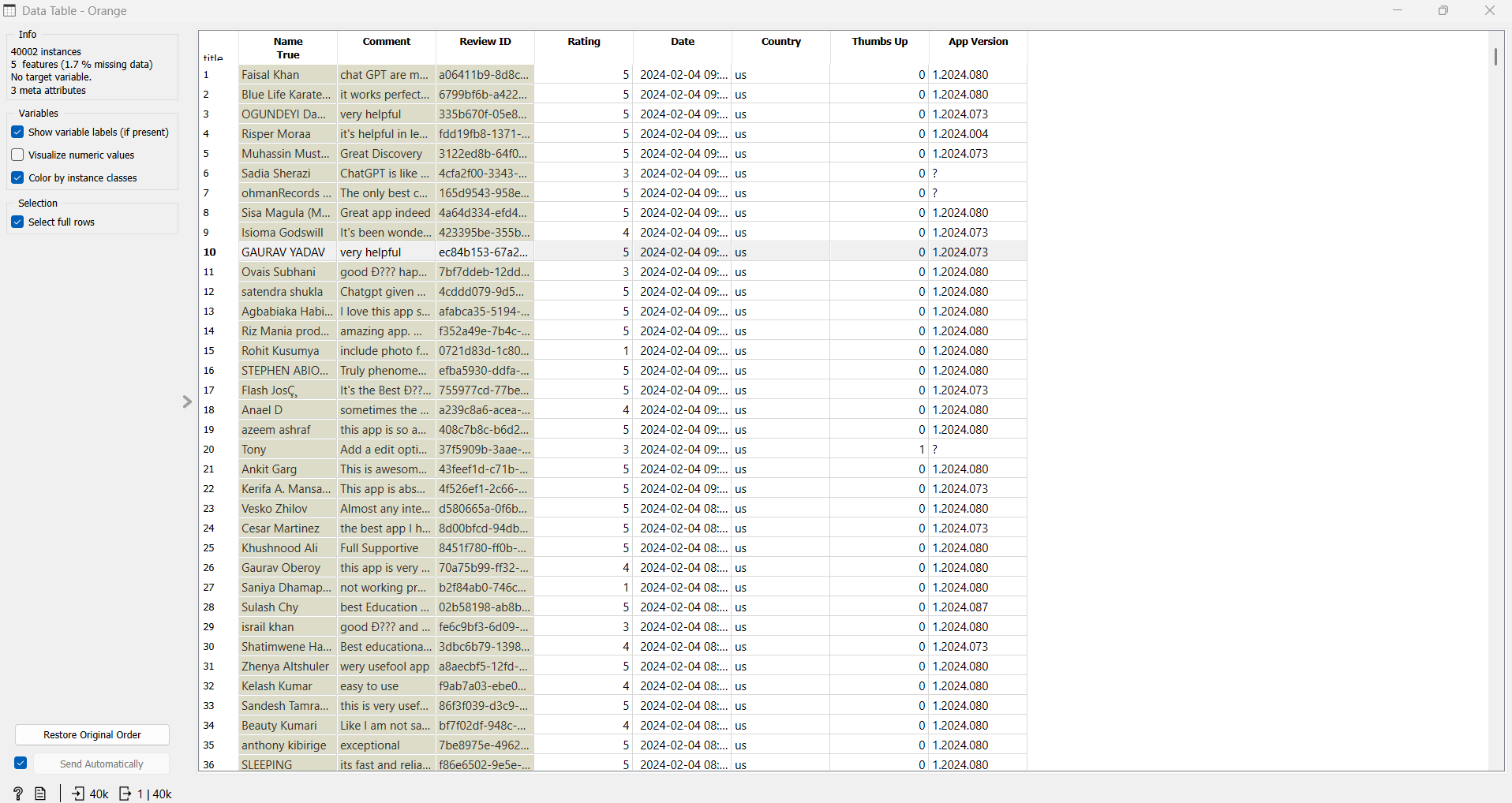
This section describes the processes used to prepare the dataset of 40,000 customer reviews from the Google Play Store for text analytics. These reviews lay the groundwork for analysing user sentiment, detecting critical trends, and generating actionable insights to inform Google's launch strategy for the Pixel LLM Phone..

**Dataset Description**

The dataset contains user reviews of the ChatGPT app, with the following key fields:

1. **Name**: Username of the reviewer.
2. **Rating**: Star ratings (1–5) reflecting user satisfaction.
3. **Comment**: The textual content of user reviews, which serves as the primary focus for text analysis.
4. **Date**: The date when the review was posted.
5. **Thumbs Up**: The number of likes received by the review, indicating user engagement.
6. **App Version**: The version of the app being reviewed (notably, some missing values in this field).

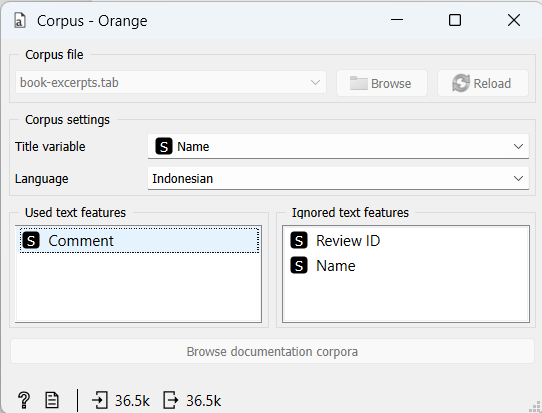
The "Comment" field is the primary source of data for sentiment analysis, topic extraction, and keyword identification.



**Steps Taken to Prepare the Data**

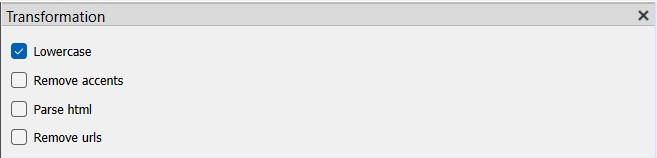
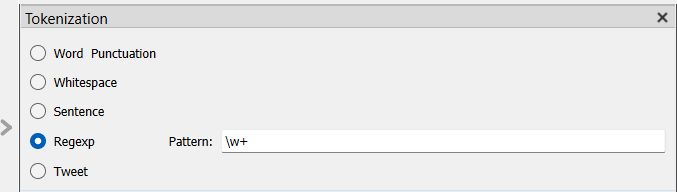
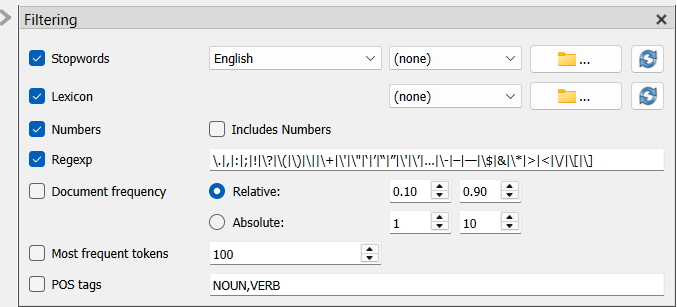
**1. Importing CSV to Corpus**

The dataset, in CSV format, was imported into **Orange Data Mining**. The “Comment” field, which contains the review text, was extracted and processed as a **corpus** for text analysis.

* **Why Corpus?** Orange's text mining tools require the textual data to be structured as a corpus for advanced preprocessing and analysis.

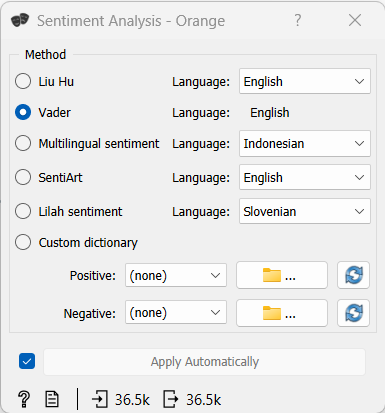
**2. Text Preprocessing in Orange**

Using the **Text Preprocessing Widget**, several steps were applied to clean and standardize the textual data:

1. **Lowercase Conversion**
   * All text was converted to lowercase to maintain consistency and eliminate case-based variations.
   * **Example**: “Amazing App!” → “amazing app”.
2. **Removing Punctuation**
   * All the punctuation marks, such as commas, periods, and exclamation marks are removed.
   * **Example**: “Great app! Works perfectly.” → “great app works perfectly”.
3. **Stopword Removal**
   * Common stopwords (e.g., “the,” “is,” “and”) were removed using Orange's built-in stopword list to focus on meaningful content.
   * **Example**: “This is a great app for learning.” → “great app learning”.
4. **Tokenization**
   * Text was split into individual words (tokens), enabling deeper analysis.
   * **Example**: “great app learning” → [“great,” “app,” “learning”].

**3. Sentiment Analysis**

The **Sentiment Analysis Widget** in Orange was used to classify the sentiment of user reviews. The **VADER (Valence Aware Dictionary and sEntiment Reasoner)** method was applied to determine whether reviews were positive, negative, or neutral.

* **Outcome**: The sentiment scores provided a basis for identifying overall user sentiment trends and specific areas for improvement.

The Four Key Analyses

**Story 1: Positive Customer Sentiments**

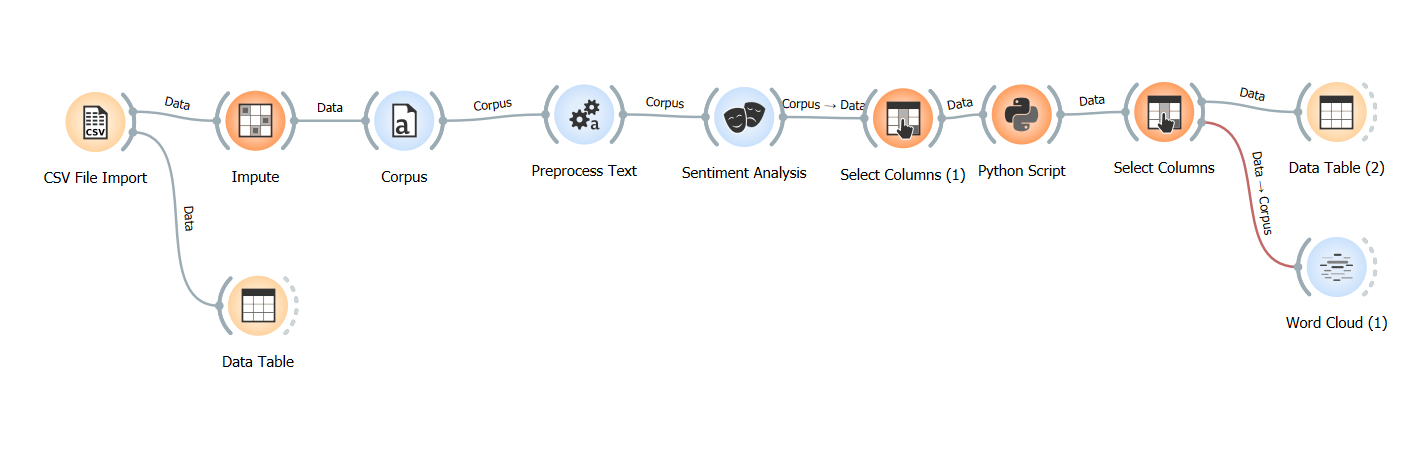
**Objective:** Identify key positive user experiences to highlight valuable features for the Pixel LLM Phone.

**Steps:**

1. **Data Input:** Import 40,000 ChatGPT app reviews.
2. **Text Preprocessing:** Convert text to lowercase, remove punctuation/stopwords, tokenize, and lemmatize.
3. **Sentiment Analysis:** Classify reviews as positive, neutral, or negative; filter positive reviews using Python.
4. **Word Cloud:** Visualize frequent positive keywords.

**Outcome:**  
Positive reviews emphasize keywords like **“helpful,” “efficient,”** and **“amazing.”** These reflect valued features like usability, performance, and user satisfaction, guiding Google to focus on these strengths for marketing and development.

**Workflow Representation**



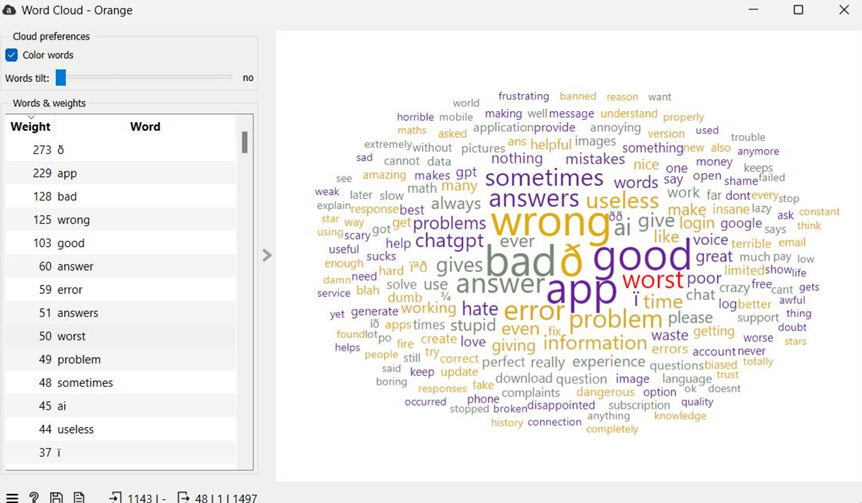
**Story 2: Negative Customer Experiences**

**Objective:** Identify recurring complaints and technical issues for improvement.

**Steps:**

1. Import dataset.
2. Preprocess text (lowercase, remove punctuation/stopwords, lemmatize).
3. Classify reviews; filter negative reviews using Python.
4. Generate Word Cloud for frequent keywords.

**Outcome:**  
Keywords like **“bugs,” “crashes,”** and **“errors”** highlight technical issues. Addressing these will improve reliability and user satisfaction for the Pixel LLM Phone.



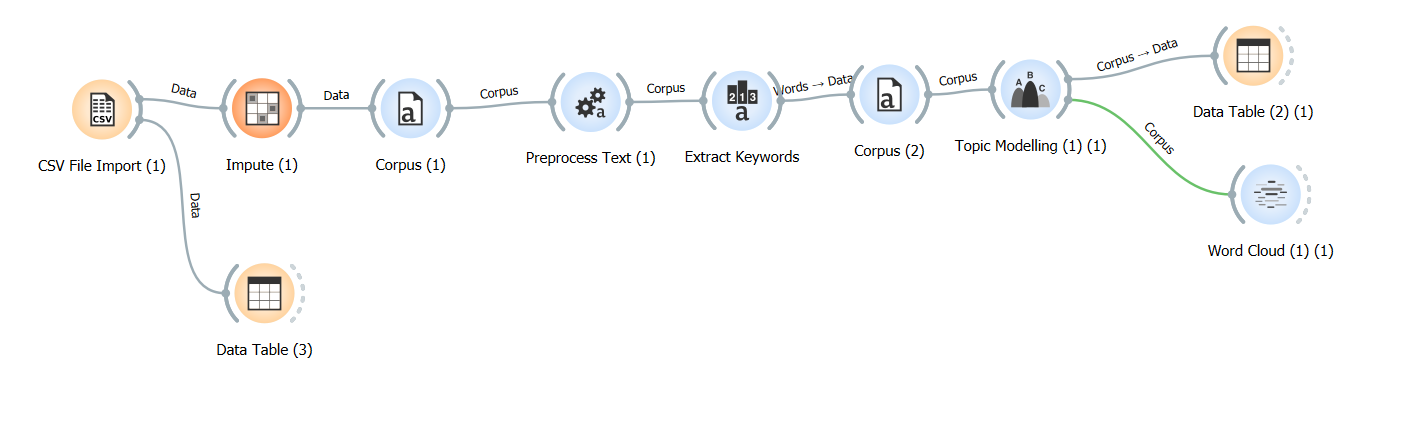
**Story 3: Innovative Opportunities from Specific Features**

**Objective:** Identify innovative AI features to gain a competitive advantage.

**Steps:**

1. Import dataset.
2. Preprocess text (cleaning, tokenization).
3. Extract important keywords from positive reviews using TF-IDF.
4. Identify AI-related themes with Topic Modeling (LDA).
5. Generate a Word Cloud for innovation-related keywords.

**Outcome:**  
Highlights features like **“adaptive learning”** and **“AI responsiveness,”** showcasing opportunities to enhance the Pixel LLM Phone's innovation and user engagement.

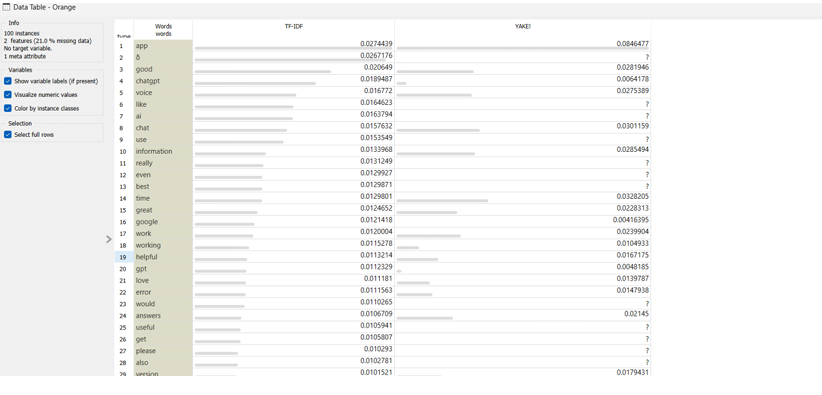


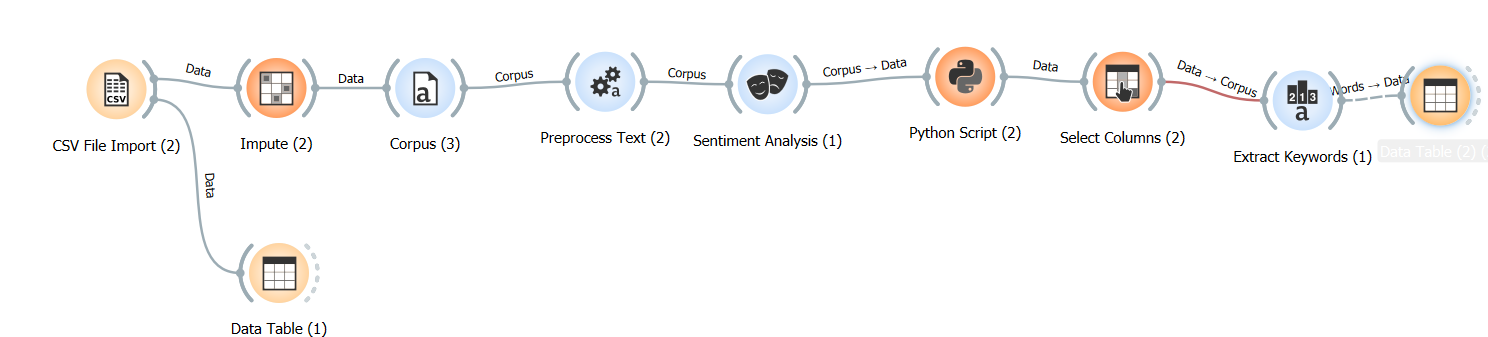
**Story 4: User Engagement and Feedback Trends**

**Objective:** Analyze user engagement metrics (Thumbs Up) and review trends over time.

**Steps:**

1. Import dataset.
2. Preprocess text (lowercase, remove punctuation/stopwords).
3. Classify reviews by sentiment (positive, neutral, negative).
4. Filter top 10% of reviews with the highest Thumbs Up values.
5. Analyze sentiment trends over time using the Date field and Line Chart.
6. Extract keywords from top reviews using TF-IDF.
7. Summarize trends, engagement levels, and top keywords in a data table.

**Outcome:**  
Shows increasing sentiment trends and engagement over time, with highly liked reviews emphasizing keywords like **“helpful”** and **“amazing,”** offering insights to boost user satisfaction and marketing strategies.



Key Recommendations for Each Narrative

**Story 1: Positive Customer Sentiments**

**Key Recommendations:**

* Emphasize features like **ease of use, efficiency,** and **natural AI responses** in marketing campaigns.
* Develop messaging that highlights the **helpful, efficient,** and **user-friendly** AI experience of the Pixel LLM Phone.

**Pressing Questions:**

* **Sustainability:** How can Google maintain consistent quality as LLM technology scales?
* **Risk Management:** Can the system sustain positive features without straining resources or compromising performance?

**Story 2: Negative Customer Experiences**

**Key Recommendations:**

* Prioritize **stability and reliability** by addressing bugs, crashes, and inconsistent AI responses.
* Implement **rigorous testing** and establish feedback loops to identify and fix issues pre-release.
* Enhance AI **error-handling mechanisms** to improve user experiences.

**Pressing Questions:**

* **Disruption:** What steps can Google take to minimize disruptions from technical failures?
* **Risk Management:** How can Google reduce reputational risks from recurring technical problems?

**Story 3: Innovative Opportunities from Specific Features**

**Key Recommendations:**

* Invest in **adaptive learning** and **personalized AI responses** to stand out from competitors.
* Enhance features like **response speed** and **contextual understanding** to elevate user experiences.
* Market the Pixel LLM Phone as an **innovative AI-powered solution** targeting professionals and tech-savvy users.

**Pressing Questions:**

* **Data Ethics:** How can Google ensure personalization features remain ethical and unbiased?
* **Sustainability:** Will increased AI functionality lead to higher energy demands, and how can this be mitigated?

**Story 4: User Engagement and Feedback Trends**

**Key Recommendations:**

* Establish a robust **user feedback system** for continuous insights and improvements.
* Highlight features driving high engagement, like **"helpful," "amazing,"** and **"easy to use."**
* Use **sentiment trends** and engagement metrics to pinpoint areas for innovation or refinement.

**Pressing Questions:**

* **Sustainability:** How can Google maintain continuous improvement without excessive resource use?
* **Risk Management:** How can user expectations be effectively managed over time as trends evolve?

Recommendations

The key recommendations derived from the text analytics are as follows:

1. **Promote Positive Features**
   * Highlight user-valued features such as **"helpful"** and **"efficient"** in marketing campaigns to strengthen the product's market position.
2. **Address Recurring Issues**
   * Focus on resolving bugs, crashes, and other technical concerns to ensure stability and reliability.
3. **Focus on Innovation**
   * Develop advanced AI features emphasizing **personalization, adaptive learning,** and **contextual responses** to differentiate the Pixel LLM Phone in the market.
4. **Enhance User Engagement**
   * Use trends from user feedback to continually refine product features and improve customer satisfaction.

Conclusion

This report leverages text analytics techniques, including **sentiment analysis, topic modeling,** and **keyword extraction,** to analyze user reviews of the ChatGPT app. The findings offer valuable insights into user priorities, recurring issues, and areas for innovation, addressing both opportunities and challenges for the upcoming Google Pixel LLM Phone.

By acting on these recommendations, Google can:

* Deliver a product that aligns with user expectations while mitigating risks.
* Use AI innovation to gain a competitive edge in the market.
* Build a feedback-driven system to ensure long-term success.

The structured analyses and workflows in this report provide Google with **data-driven insights** to make informed decisions and position the Pixel LLM Phone as a leading AI-powered device in the market.

APPENDICES

**Short User Guide on How to Use Workflows in Orange**

This guide provides step-by-step instructions for replicating the workflows used in the analysis with Orange Data Mining. Each workflow is designed to extract specific insights from the dataset.

**Workflow 1: Positive Customer Sentiments**

**Objective:** Identify and analyze positive reviews to highlight valuable features.

1. **Data Input:** Load the dataset using the **Data Input Widget**.
2. **Text Preprocessing:**
   * Connect the **Text Preprocessing Widget**.
   * Apply lowercase conversion, stopwords removal, and punctuation removal.
3. **Sentiment Analysis:** Use the **Sentiment Analysis Widget** to classify reviews by sentiment.
4. **Python Script:** Filter rows where sentiment is classified as positive.
5. **Word Cloud:** Generate a **Word Cloud** to visualize frequently used positive keywords.
6. **Data Table:** Display the filtered results in a **Data Table Widget** for review.

**Workflow 2: Negative Customer Experiences**

**Objective:** Identify recurring complaints and technical issues from negative reviews.

1. **Data Input:** Load the dataset using the **Data Input Widget**.
2. **Text Preprocessing:**
   * Apply text cleaning (lowercase conversion, punctuation removal, stopwords removal, etc.).
3. **Sentiment Analysis:** Classify reviews into positive, neutral, and negative using the **Sentiment Analysis Widget**.
4. **Python Script:** Filter rows classified as negative reviews.
5. **Word Cloud:** Generate a **Word Cloud** to highlight key pain points and common issues.
6. **Data Table:** Display the filtered results for further exploration.

**Workflow 3: Innovative Opportunities**

**Objective:** Identify themes and features related to innovation and AI capabilities.

1. **Data Input:** Load the dataset using the **Data Input Widget**.
2. **Text Preprocessing:** Clean text data to prepare it for analysis.
3. **TF-IDF Widget:** Use the **TF-IDF Widget** to extract keywords with high importance scores.
4. **Topic Modeling Widget:** Use the **Topic Modeling Widget (LDA)** to identify dominant themes and topics related to innovation.
5. **Word Cloud:** Generate a **Word Cloud** to visualize themes around innovation and AI adaptability.
6. **Data Table:** Display the extracted themes and keywords for review.

**Workflow 4: User Engagement Trends**

**Objective:** Analyze user engagement metrics and temporal trends in reviews.

1. **Data Input:** Load the dataset using the **Data Input Widget**.
2. **Text Preprocessing:** Apply text cleaning (lowercase, punctuation removal, stopwords removal, etc.).
3. **Sentiment Analysis:** Classify reviews based on sentiment using the **Sentiment Analysis Widget**.
4. **Python Script:** Filter for reviews with the highest **Thumbs Up** (top 10% by engagement).
5. **Line Chart Widget:** Use the **Line Chart Widget** with the Date field to plot sentiment trends over time.
6. **TF-IDF Widget:** Extract keywords from the most engaging reviews using the **TF-IDF Widget**.
7. **Data Table:** Display the results, including trends, keywords, and engagement metrics, for further analysis.