Project Report

TASK 4: ANALYZING CONSUMER SMARTPHONE APP REVIEWS

Introduction:

In this project, we analyzed user reviews for popular smartphone apps using data cleaning, advanced analysis, and sentiment analysis techniques. We visualized our findings using Kaggle notebook to gain insights into sentiment trends, frequently mentioned keywords, and app rating trends over time.

Data Cleaning and Preparation:

We began by loading the dataset into a pandas dataframe and performing initial data cleaning and preparation. We handled missing values by filling them with an empty string and removing duplicates. We then converted the "at" column to datetime format.

Sentiment Analysis:

We used the Natural Language Toolkit (NLTK) to perform sentiment analysis on the review content. We created a SentimentIntensityAnalyzer object and applied it to the review content to calculate sentiment scores and labels. We then added these as new columns to the dataframe.

Sentiment Distribution:

We calculated the distribution of sentiment scores and labels using the describe() and value_counts() methods. We found that the majority of reviews had a positive sentiment score, but there were also a significant number of negative and neutral reviews.

Frequently Mentioned Keywords:

We tokenized the review content using NLTK's word_tokenize() function and removed stopwords. We then calculated the frequency of keywords using Counter() and plotted the most common keywords using a bar chart. We also created a word cloud to visualize the frequently mentioned keywords.

Rating Trends:

We plotted the distribution of app ratings using a histogram and the trend of app ratings over time using a bar chart. We found that the majority of ratings were positive, but there were also some negative ratings. The trend of app ratings over time showed that the app received more positive ratings over time.

Sentiment Trends:

We grouped the data by review created version and calculated the sentiment scores for each version. We plotted the sentiment scores by review created version using a bar chart. We found that the sentiment scores were generally positive, but there were some negative scores for earlier versions.

Conclusion:

In this project, we analyzed consumer smartphone app reviews using data cleaning, advanced analysis, and sentiment analysis techniques. We visualized our findings using kaggle notebook and gained insights into sentiment trends, frequently mentioned keywords, and app rating trends over time. Our analysis can help app developers understand user feedback and improve their apps accordingly.

Limitations:

There are several limitations to this study. First, we only analyzed a subset of the available data, focusing on the review content, sentiment scores, and app ratings. We did not analyze other factors such as user demographics or app features. Second, our sentiment analysis was based on a pre-trained model and may not accurately reflect the true sentiment of the reviews. Finally, our visualizations were created using Tableau Public, which may not be accessible to all users.

Future Work:

In the future, we could explore user demographics and app features to gain a more complete understanding of consumer feedback. We could also use more advanced sentiment analysis techniques, such as deep learning models, to improve the accuracy of our analysis. Additionally, we could explore other visualization tools to make our findings more accessible to a wider audience.