**what data do you plan to use and what would be your target variable(s)? Please include link to data source in your response.**

Kenyan Youtube data API & Web Scrapping.  Our target variable is the viewership.

<https://socialblade.com/>

**Project Name: NLP-Based Sentiment Dissection of Apple Vision Pro Reception on Marques Brownlee’s YouTube Channel.**

**Business Understanding**

**Problem Statement:** The Product Development Team at Apple has subcontracted us to assess how the Apple Vision Pro is being received by Marques Brownlee's viewers on YouTube. The goal is to use Natural Language Processing (NLP) to analyze viewer comments and determine their sentiments, particularly identifying whether viewers are inclined towards purchasing the product.

**Why This Topic?**

* **Relevance**: Marques Brownlee, a significant influencer in the tech domain, has a considerable impact on consumer decisions. Understanding his audience's reaction can guide marketing strategies and product improvements.
* **Impact**: Insights derived could influence product marketing strategies, feature enhancements, and customer engagement practices.

**Domain**: This project applies to the technology and consumer electronics industry, specifically in the domain of wearable and augmented reality devices.

**Target Audience**: The primary stakeholders are the product development team, marketing team, and strategic decision-makers within Apple.

**Real World Impact**: Implementing the findings could enhance product positioning, improve customer satisfaction, and increase sales through targeted improvements and marketing strategies.

**Pre-existing Projects/Domain Knowledge**:

* Review of existing sentiment analysis models applied on social media platforms.
* Research on consumer behavior analysis using comments and feedback from social media.

**Data Understanding**

**Data Collection**:

* **Data Source**: YouTube comments from Marques Brownlee's review videos specifically for the Apple Vision Pro.
* **Collection Method**: Using the YouTube Data API to fetch comments.

**Features**:

* **Comment Text**: Main text of the comment.
* **Likes**: Number of likes on the comment.
* **Timestamp**: Time the comment was posted.

**Previous Work**: While there are general sentiment analysis projects on product reviews, this specific analysis on YouTube tech reviews particularly focusing on Augmented Reality devices like Apple Vision Pro is relatively niche.

**Data Preparation**

**Data Storage**: Raw data will be stored in CSV format.

**Data Types**:

* Comment text: String
* Likes: Integer
* Timestamp: DateTime

**Preprocessing Steps**:

* Text cleaning (removing URLs, special characters).
* Tokenization and lemmatization for NLP.
* Sentiment scoring using pre-trained models like VADER.

**Challenges**:

* Handling large volume of data and comments.
* Managing noisy or irrelevant comments.

**Minimum Rows**: Approximately 1,000 comments to ensure a robust analysis.

**Visualization**: Initial exploratory data analysis with histograms, pie charts for sentiment distribution, and time-series analysis of sentiment over time.

**Modeling**

**Techniques**: Sentiment analysis using NLP.

**Target Variable**: Sentiment score (Positive, Neutral, Negative).

**Baseline Model**: Use VADER for initial sentiment analysis as a baseline.

**Problem Type**: Classification problem (categorizing sentiments).

**Evaluation**

**Metrics**:

* Accuracy and F1-Score for sentiment classification.
* Analysis of sentiment distribution and correlation with product features.

**MVP**: Basic sentiment analysis using VADER or TextBlob and visualizing sentiment distribution.

**Stretch Goals**:

* Enhance the model by training a custom classifier using a labeled dataset.
* Implement more complex NLP techniques like BERT/ROBERTA or fine-tuning transformer models.

**Deployment**

**Final Results Reporting**: Parsing statements to determine whether the comment is Positive, Neutral or Negative.

Interactive dashboards using libraries like Plotly or Dash.

**Deployment Plan**:

* Development of a web application to present the analysis and real-time sentiment tracking.
* Option to input new video URLs for sentiment analysis.

**Tools/Methodologies**

**Libraries**:

* **pandas** for data manipulation.
* **nltk** or **spaCy** for NLP.
* **scikit-learn** for modeling.
* **matplotlib** and **plotly** for visualization.

**Algorithms**: Sentiment analysis models, possibly LSTM or BERT if custom training is undertaken.

**Analysis Environment**: Local machine for initial development and testing, considering cloud platforms for scalability if real-time analysis is required.

**Data Storage**: Local and cloud storage options will be evaluated based on the volume of data and processing requirements.

This proposal outlines a structured approach to using NLP for analyzing consumer sentiments on YouTube, specifically for a high-profile tech influencer’s reviews on a new Apple product. The plan covers the entire pipeline from data collection to deployment with considerations for scalability and real-world application.

Top of Form