The project detailed in the "DVA Final Report" is titled "Dynamic Amazon Product Review Visualization" and is aimed at developing an innovative online tool designed to analyze product reviews on Amazon, providing users with comprehensive insights into product quality and satisfaction trends. Here's a brief overview of the project:

Abstract/Introduction

* **Objective**: To bridge the analytical gap in existing online shopping platforms that predominantly track price histories but lack in-depth analysis of product quality and satisfaction trends.
* **Solution**: A user-friendly online tool dissecting Amazon product reviews, offering insights on various dimensions like sentiment analysis, rating breakdowns, price variations, and comparative product rankings.
* **Goal**: Enable users to make more informed purchase decisions by discerning trends in product quality or popularity over selected time frames.

Methods and Innovation

* **Data Source**: Utilizing an amazon review dataset provided by UCSD.
* **Key Innovations:**
  + Moving Average: Displays average rating and sentiment over time for a product, helping visualize the product’s reception evolution.
  + Sentiment Analysis: Shows a weighted ratio of "Positive" and "Negative" reviews in a defined time window, allowing assessment of sentiments in product feedback.
  + Product Relation Graph: Constructs a graph of similar products with prices and ratings, aiding in exploring alternative items.
* **Interface and Tools**
  + Interface: Developed using Streamlit, an open-source framework for Machine Learning applications in Python.
  + Database: MySQL for data storage and retrieval, hosted on PlanetScale, a cloud-provider offering scalable database solutions.
  + Sentiment Analysis Technique: Leveraging the Hugging Face library and models like BERT and GPT for sentiment analysis, specifically using the Happy Transformer package for sentiment inference.
  + Efficiency Measures: Implemented multithreading for processing large volumes of data and incorporated error handling within this environment.

Sentiment Analysis

* **Objective**: To gauge the emotional tone behind reviews, providing an additional layer of insight beyond basic star ratings.
* **Approach:**
  + Data Processing: Reviews are pre-processed to remove noise and prepare text for analysis.
  + Model Deployment: Utilizes advanced NLP models (BERT, GPT) from Hugging Face's Happy Transformer package for accurate sentiment detection.
* **Output:**
  + Sentiment Scores: Each review is assigned a sentiment score, categorizing it as positive, neutral, or negative.
  + Visual Representation: Bar graphs or pie charts display the proportion of sentiments across a product’s reviews over selected time frames.
* **Impact:**
  + Qualitative Insights: Offers nuanced understanding of customer opinions, supplementing quantitative rating data.
  + Trend Analysis: Tracks shifts in sentiment over time, highlighting changes in consumer satisfaction or product quality.

Moving Average

* **Objective**: To provide a clear, quantitative analysis of product review trends over time, aiding in understanding how public perception of a product has evolved.
* **Approach**:
  + Data Processing: Aggregates review ratings over specific time intervals, smoothing out short-term fluctuations to highlight longer-term trends.
  + Average Calculation: Calculates the moving average of ratings, offering a more stabilized view of the product's reception.
* **Visualization:**
  + Line Graphs: Utilizes line graphs to depict the moving average of product ratings across time, making it easy to spot trends or significant changes.
  + Time-Frame Selection: Allows users to adjust the time frame for analysis, catering to different needs for short-term or long-term trend observation.
* **Features**:
  + Rating Distribution: Accompanies the moving average with a breakdown of ratings distribution, providing a comprehensive view of the overall review landscape.
  + Interactive Elements: Interactive sliders or selectors for users to customize the period of analysis and the granularity of data.
* **Impact:**
  + Market Analysis: Offers valuable insights for market analysis, showing how products fare over time in the eyes of consumers.
  + Comparison Tool: Serves as a comparison tool, enabling users to juxtapose different products based on their moving average trends.

Product Relation Graph

* **Objective:** To provide users with a visual network of related products based on similarities in customer reviews, ratings, and features.
* **Functionality:**
  + Relation Mapping: Analyzes similarities between products to construct a graph, connecting items that share similar characteristics or are frequently compared in reviews.
  + Interactive Graph: Allows users to explore related products visually, with nodes representing individual products and edges indicating relationships.
* **Features:**
  + Dynamic Filtering: Users can filter the graph based on criteria like price range, ratings, or specific product features.
  + Product Comparison: Facilitates easy comparison of related products by displaying key metrics like average rating and price.
* **Benefits:**
  + Discoverability: Enhances product discoverability, guiding users to explore alternatives they might not have found independently.
  + Informed Decision Making: Helps users make better-informed choices by providing a broader context of product options and their comparative standings.

Scalability and Evaluation

* **Scalability Testing**: Essential due to the substantial volume of data (approx. 1.6 million records), ensuring system performance and responsiveness.
* **User Study**: To evaluate the effectiveness and user experience of the dashboard, observing natural interactions and conducting interviews for qualitative insights.