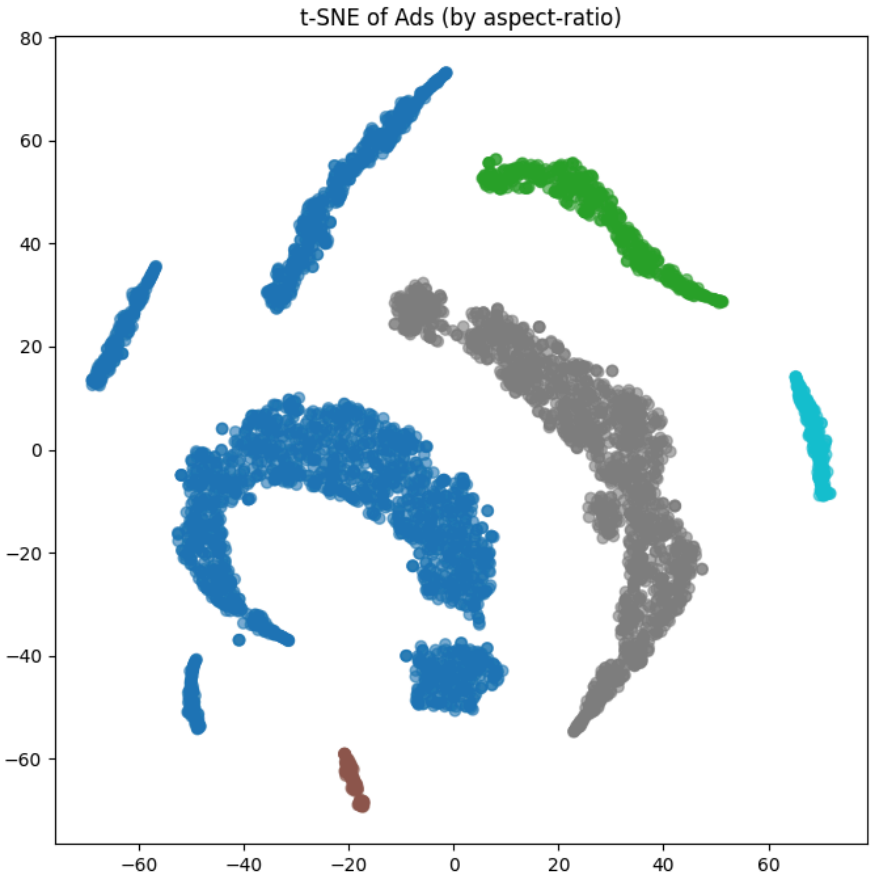
Data Appendix

I. train-00001-of-00002-823ac5dae71e0e87.parquet

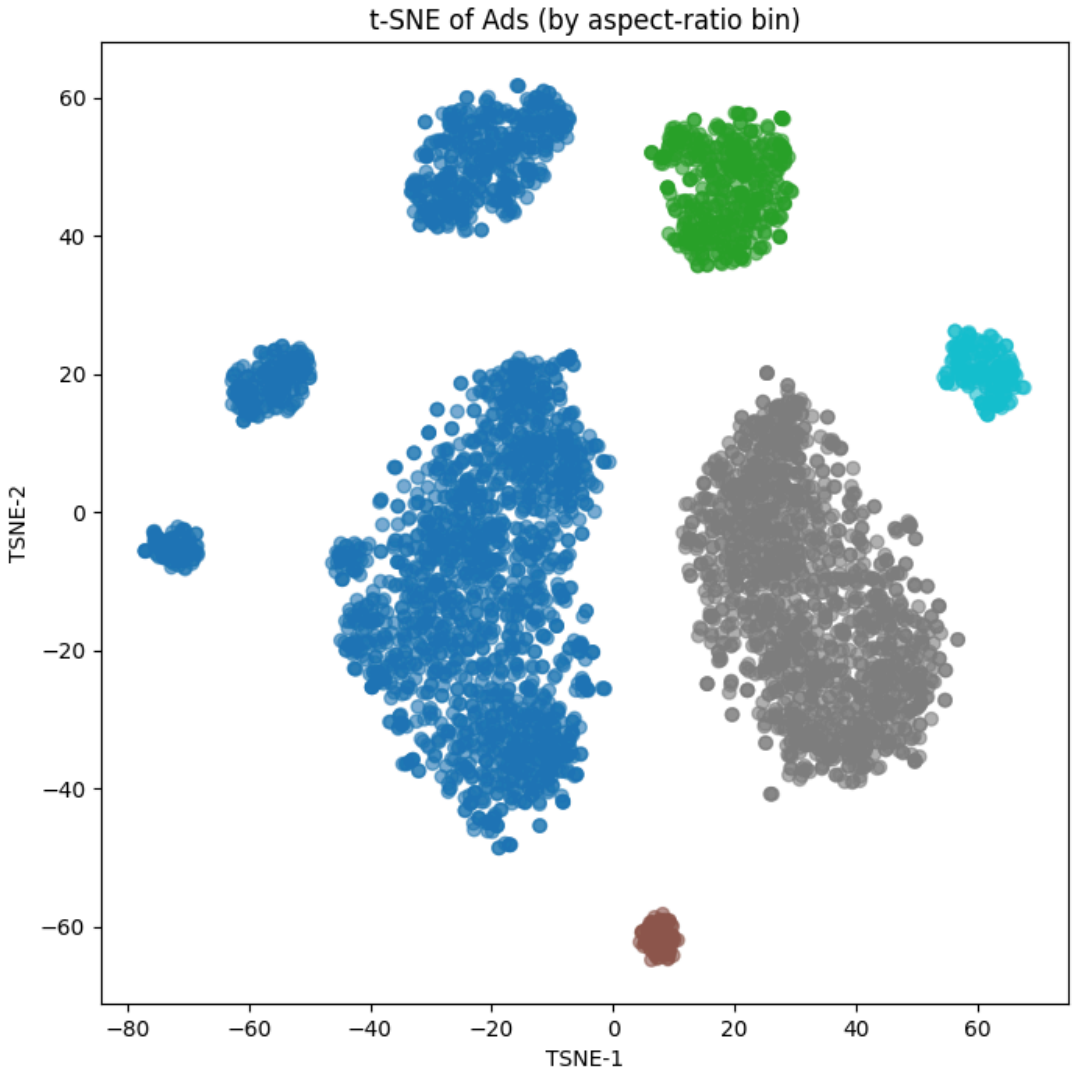
Each row represents a single online advertisement. The core content of each row is a byte-encoded image of the ad, along with metadata fields describing its dimensions and structure.

A. image
   a. Byte-based images which are stored in a dictionary
B. text
   a. String-based text information that each ad contains
C. dimensions
   a. Tuple-based image dimensions

Outputs/Figures and Analysis:

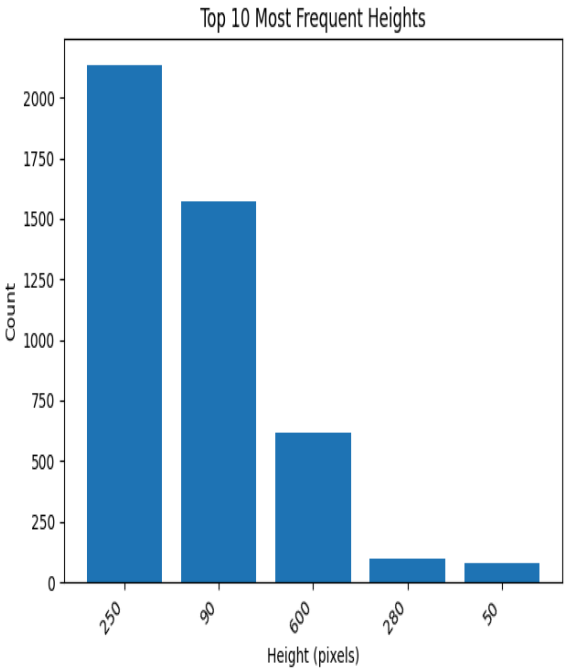


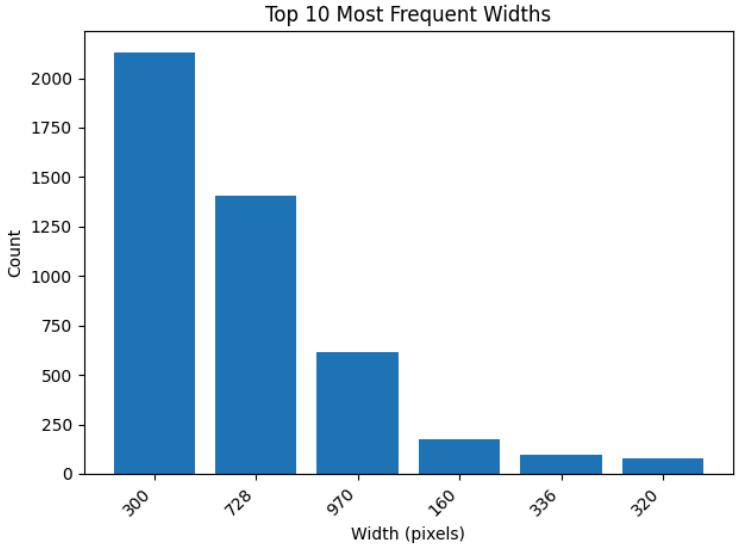
The t-SNE visualization above is of ad embeddings by aspect ratio reveals distinct clusters of ad images based on their structural features, particularly shape. Each color represents a different aspect ratio bin. The separation of clusters suggests that ads with similar proportions tend to group closely in the high-dimensional feature space, supporting the idea that layout significantly impacts visual embedding similarity.

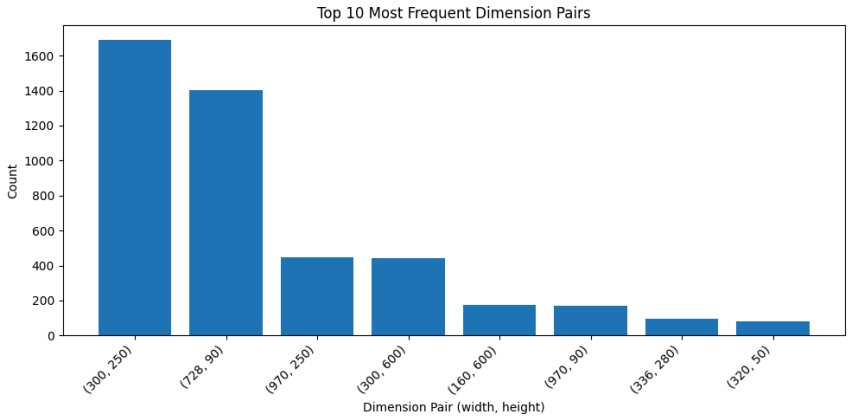


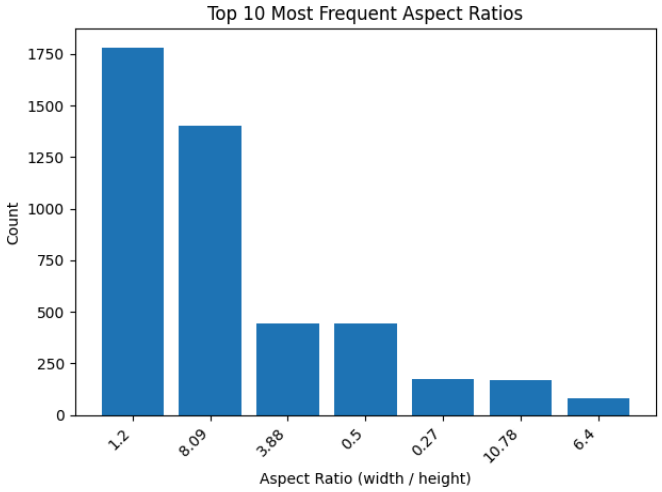
This t-SNE above offers the same dimensionality-reduced view but with a normalized layout, resulting in clearer separation between aspect ratio groups and more compact clusters. This graph further confirms that visual layout, particularly shape and scale, plays a dominant role in how ads

are distributed in feature space. The tightness of certain clusters also suggests heavy reuse of specific templates or dimensions, reinforcing insights drawn from the EDA charts.









The bar chart above of the top 10 most frequent widths reveals that the most common ad widths are 300, 728, and 970 pixels, with 300px being the dominant size. These widths align with standard web display formats, such as medium rectangles, leaderboards, and wide skyscrapers. The high frequency of these widths suggests strong adherence to conventional ad sizing for compatibility across advertising platforms.

The top 10 most frequent heights chart above mirrors this pattern, with 250px and 90px as the most used values. These dimensions match popular vertical ad formats and reinforce the prevalence of fixed, standardized templates. Heights such as 600px and 280px also appear, indicating some diversity in layout strategies, especially for taller sidebar ads.

The dimension pair distribution above shows that combinations like (300, 250), (728, 90), and (970, 250) are the most frequently used ad sizes. These pairs correspond to commonly accepted display ad standards by platforms such as Google Ads and IAB. Their repetition across the dataset highlights a strong trend toward uniformity in ad design and dimensions, likely driven by platform compatibility requirements and established design best practices.

The aspect ratio analysis above shows that ads with a 1.2 or 8.09 width-to-height ratio dominate the dataset. These ratios reflect rectangular banners (wide formats) and nearly square designs, indicating two prevailing formats in digital advertising. Less common but present ratios include 3.88, 0.5, and 10.78, demonstrating that while some creative variation exists, advertisers largely stick to consistent proportional templates.