

Project Overview

Goal

The aim of this project was to extract data from the Reddit API, specifically from the r/analog subreddit, to uncover trends and patterns in analog photography. The focus was to identify the most commonly used film and camera brands and assess how their use correlates with post scores (upvotes). This information could offer insights into popular gear choices and help inform future purchasing decisions.

Data Source

The dataset was derived from the top posts in the r/analog subreddit, with a limit of 1000 posts from the previous year. Each post includes fields such as:

- Post title
- Score (upvotes)
- Author
- Time of post (created_utc)
- Additional details (num_comments, URL...)

Objective

1. Identify the most popular film and camera brands based on post count.
2. Analyze how these brands influence post scores
3. Present a statistical understanding of popular equipment to guide future purchasing decisions.

Process

1. **Data Extraction:**
 - The Reddit API was accessed to pull top posts from the r/analog subreddit.
 - Data was stored in raw table in a PostgreSQL database
2. **Data Processing:**
 - Camera and film names were extracted from the post titles and descriptions using a custom-trained SpaCy model. The result of the extraction is stored in reddit_posts_transformed table.
 - The extracted names were mapped to a standardized values using fuzzywuzzy library for calculating similarity. The results were stored in reddit_posts_final table
 - Aggregations were performed to calculate the post count and average score for each camera and film brand.
3. **Analysis:**
 - The number of posts per camera and film brand was visualized to identify the most popular brands.

- The average score for each brand was calculated to understand how post scores vary by equipment.

Results

- **Top Camera Brands:** The most frequently used camera brands were Nikon, Canon, Mamiya and Pentax.
- **Top Film Brands:** The most frequently used film brand is by far Kodak with almost 75%. Which is expected because Kodak Portra is the most popular and versatile film stock in professional and amateur use.
- **Post Scores:** The analysis showed that pictures taken with Leica, Olympus and Fujica cameras tend to have the most upvotes, but are used less often in comparison to Nikon and Canon. Also, if Fujifilm or Fomapan film stocks are used, post tend to have more upvotes, but those films stock are generally rarely used, because they are out of production and harder to find.