The Challenge

Background

UBS Evidence Lab tracks the popularity of brands on Instagram based on metrics such as followers, posts, likes, etc. It includes 730+ brands in categories including Apparel Retail, Beauty & Boutique, Beverages, Building Products, Fitness & Exercise, Food Products, Luxury & Premium & Mainstream, Mattresses, Outdoor Gear, Restaurants, Pet Care, Sporting Goods, and Sportswear & Athleisure.

Challenge

Monitoring brand popularity trends on social platforms can provide valuable insights into how a brand's perception is related to its future success.

Your challenge is to develop a model that not only identifies significant deviations from these observed trends but also highlights noteworthy brands based on the provided dataset. You're free to choose the brand(s) for your analysis and to define what you consider as an "interesting" deviation. Examples of deviations could include sudden metric spikes, substantial changes in the 1st or 2nd derivative, or a significant shift in rank relative to competitors, among others.

Beyond identifying these deviations, consider their broader implications. Are they indicators of future potential or simply anomalies? The goal is not just to detect them, but also to provide interpretations that could inform strategic decisions.

UBS Evidence Lab

UBS Evidence Lab is a global team of alternative data experts who work across numerous specialized areas creating insight-ready datasets. The experts turn data into evidence by applying a combination of tools and techniques to harvest, cleanse, and connect billions of data items each month. The library of assets, covering 1000s of companies of all sizes, across all sectors and regions, is designed to help investors answer the questions that matter to their investment analysis.

The data

NAME	TYPE	DESCRIPTION
period	string	Frequency of the measurement ('Weekly')
period_end_date	date	End date of the period under consideration
compset_group	string	Main competitive set
compset	string	Competitive set
business_entity_doing_business_as_name	string	Brand
legal_entity_name	string	Company
domicile_country_name	string	Country of domicile
ultimate_parent_legal_entity_name	string	Ultimate parent company
primary_exchange_name	string	Exchange where stock is traded
calculation_type	string	Type of calculation ('Metric Value')
followers	int	Number of followers at period_end_date
pictures	int	Number of pictures during the period
videos	int	Number of videos during the period
comments	int	Number of comments during the period
likes	int	Number of likes during the period

Data starts in 2015 and ends 2023

Deliverables and Scoring

Deliverables

Code and documentation submission guidelines:

Participants are granted the flexibility to select their preferred format for project submission. Acceptable formats include, but are not limited to, Jupyter Notebooks, R Markdown files, or standalone Python scripts. Regardless of the chosen format, it is imperative that all submitted code is commented and logically structured to facilitate ease of understanding. The culmination of your submission should be a technical report that outlines your analytical journey, highlighting the methodologies employed, any obstacles encountered along the way, and the strategies adopted to overcome these challenges. This comprehensive report should consist of the following key components:

1. **Data understanding summary**

    i. **Data cleaning and preprocessing:** Provide a detailed account of the initial steps taken to prepare the data for analysis. This should include a description of how data quality issues, such as missing values or outliers, were addressed.

    ii. **Assumptions:** Clearly articulate any assumptions that were made during the data preparation phase.

    iii. **Feature engineering and data augmentation:** Describe any techniques employed to enhance the dataset, whether through the creation of new features or augmentation of the existing data.

2. **Modelling approach summary**

    i. **Model selection:** Explain on the choice(s) of statistical or machine learning model(s) utilized in your analysis. Provide a compelling justification for each model selected, emphasizing how they align with the objectives of the challenge.

    ii. **Recommendations for model enhancement:** Conclude with a thoughtful reflection on potential avenues for further improving your model. Propose specific modifications or additional analyses that could refine your predictions and insights.

Scoring

Maximum of 100 points total

Data Understanding (40pts):

1. Exploratory Data Analysis

    • Examination and understanding of the dataset's structure and content

    • Performing exploratory data analysis to understand data patterns, outliers, and relationships between variables

2. Data Cleaning

    • Data preprocessing, include, but not limited to, handling missing values, data conversions, and normalization

Modelling Approach (30pts)

1. Feature engineering to create relevant features for identifying deviations

2. Development of a model to identify significant deviations from observed trends

3. Evaluation of the model's performance and its ability to identify deviations

Creativity and Presentation (30pts):

1. Analysis of deviations detected by your model, understanding potential causes

2. Interpretation of the results, providing strategic insights based on deviations

3. Use of alternative approaches (potential enhancement of dataset with external data)

Rules and recommendations

1. If there is a commit after the deadline (12:00), the team is automatically disqualified

2. The GitHub repo should include all the code

3. The GitHub repo must include a "docs" directory containing the technical report and the presentation of the solution

4. Please make sure pushing your codes well in advance