User Guide: Navigating Consumption Data as Alternative Data

Introduction

The review paper "Alternative data in finance and business: emerging applications and theory analysis (review)" by Sun et al. discusses ten subcategories of alternative data, including consumption data, which is examined in this guide. In addition to typical financial data sources like financial statements, consumption data provides special insights into market circumstances and firm performance. Businesses and investors may be able to obtain a more thorough and timely perspective for decision-making by comprehending the features and uses of consumption data.

1. Sources of Consumption Data

The main source of consumption statistics is the shopping habits of customers. These actions can take place in retail markets or through online payment platforms.
Among the specific sources cited in the paper are retailers.

e-commerce sites like YHD.com, JD.com, T-mall, Amazon.com, and Taobao.com.
sources that produce records of transactions.
sources that offer credit card spending data at the transaction level.
These sources give a clear picture of enterprise sales performance by documenting the specifics of purchasing or selling actions.

2. Types of Consumption Data

The information derived from consumers' shopping behaviors is diverse and granular. It includes a wealth of information that goes beyond simple sales figures.
Types of data obtained from consumption activities include:

- Product prices.
- Purchase volumes.
- Geographical locations where purchases occur.
- Payment methods used.
- Purchase time.
- Number of items purchased.

- Logistics information.
- Transaction records.
- Anomalies in purchasing patterns.
- Purchase frequency.
- Customer segmentation based on factors like demographics, geography, or shopping habits.
- Real-time quarterly retail sales data.
- E-commerce sales data.
- Transaction-level credit card spending data.
- Information allowing for customer analysis based on spending capability and loyalty levels.
- Data on customers with specific financial profiles, such as high FICO scores, substantial liquidity, and intense loyalty.

This rich dataset can provide a dynamic snapshot of market conditions and enterprise revenue.

3. Quality of Consumption Data

When it comes to data quality, consumption data has a number of advantages over traditional data sources, but it also has drawbacks.

- Benefits of Quality Consumption Data:
- Direct Reflection: Enterprise sales performance can be directly reflected in consumption data.
- Real-time Nature: It has the ability to give an instantaneous picture of the state of the market. Compared to quarterly or annual reports, more real-time performance insights are possible due to the availability of transaction records and e-commerce sales data.
- Timely and Direct Evaluation: It provides a prompt and straightforward evaluation of retail business performance.
- Predictive Power: For businesses that mostly depend on consumer spending, consumption data and related statistics, such as irregularities in buying trends or client segmentation, are excellent indicators of sales, unforeseen income, and excess returns.
- Granularity: In comparison to conventional reports, it may provide performance statistics with a finer level of granularity.

Challenges to Consumption Data Quality:

- Noise: Although specific noise categories for consumption data are not specified by the source, noise may be present in consumption data, just like in other alternative data. For alternative data, handling noise is a general difficulty.

- Bias: Although the source does not particularly address consumption data, alternative data in general may be prone to a number of biases, including selection bias, which occurs when the data being analyzed is merely a fraction of the whole population and may result in inaccurate interpretations. Inferences made from these subsets may not be representative of the total population.
- Potential Misinformation: Just as social media can contain inaccurate or unofficial information, consumption data from transaction records is probably factual about the transaction itself, but the context or conclusions drawn from them may be impacted by outside variables or analysis techniques.

Overall, users should be aware of the possibility of data quality problems in alternative data sources and the necessity of careful processing and analysis, even though consumption data offers substantial quality benefits in terms of timeliness, directness, and granularity.

4. Ethical Issues with Consumption Data

A significant ethical concern specifically highlighted in the source regarding consumption data is privacy.

- Collecting consumption data, which details individual or aggregate shopping behaviors, may raise privacy concerns. This falls under the broader category of personal data, which is valuable but individuals are often unwilling to share due to privacy protection needs.
- The source suggests that privacy computing is a viable solution to address the privacy problem associated with consumption data, allowing for data analysis without exposing the actual data. This technology is expected to be critical for extending alternative data application scenarios by enhancing data protection and reducing leakage risk through encryption mechanisms.

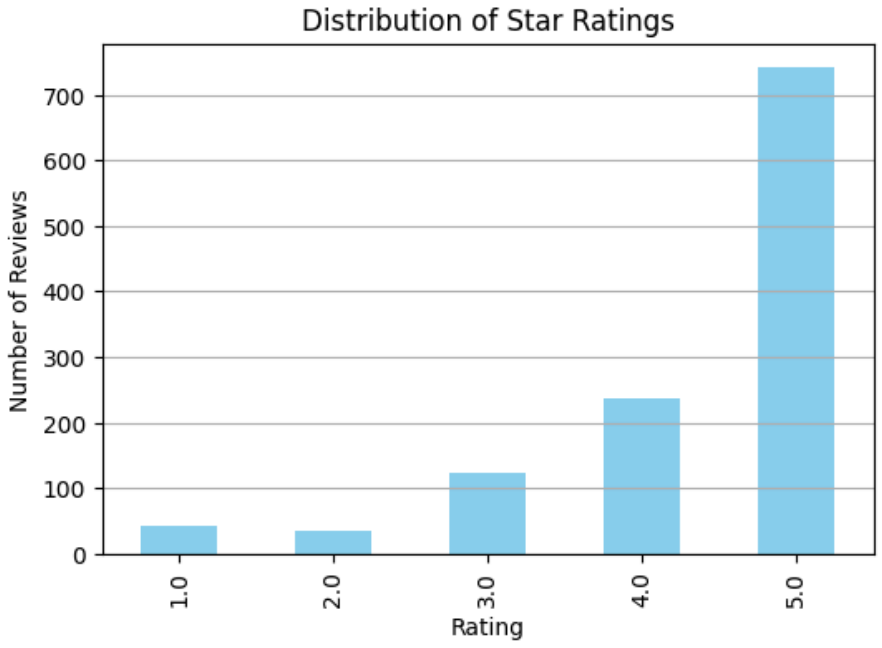
Addressing privacy concerns is crucial for the ethical use and broader adoption of consumption data.

5. Python Code to Import and Structure Data

From your dataset, these columns are most useful for financial or business insights:

Column	Use Case

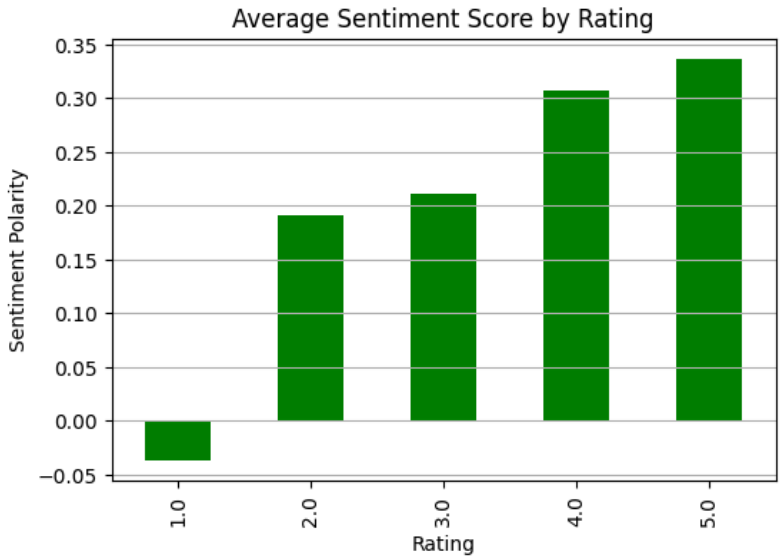
`reviews.rating`	Quantitative sentiment score
`reviews.text`	Unstructured text for NLP/sentiment analysis
`reviews.title`	Short summary sentiment
`dateAdded`/`dateUpdated`	Time-based trends
`brand`,`categories`	Categorization and filtering
`id`,`asins`	Grouping by product
`dimension`,`weight`	Product features influencing perception



fig(1)

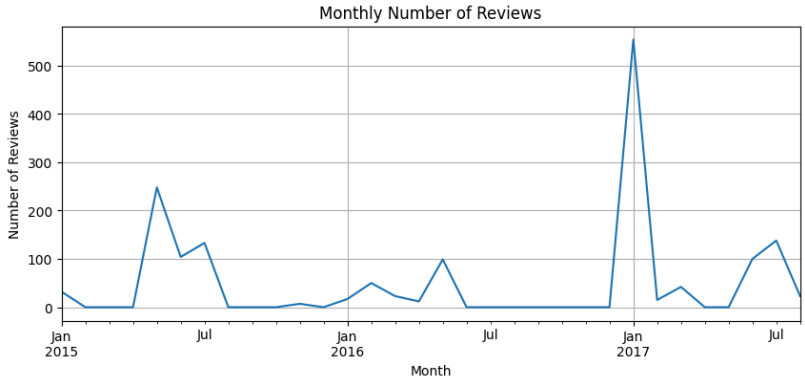
- Significant variations in the quantity of reviews over time are depicted in the graph.

- There are noticeable upticks, especially in January 2017, when there are more than 500 reviews.
- Review counts in other months are significantly lower, frequently ranging from zero to one digit.



fig(2)

- More than 700 reviews fall into the 5-star rating category, which makes up the bulk of reviews.
- Positive ratings (4 and 5 stars) make up a large portion of the distribution.



fig(3)

- There is a strong positive correlation between star ratings and sentiment scores.

- This confirms that higher ratings generally correspond to more positive textual feedback.

6. Exploratory Data Analysis (EDA) of Sample Data

The provided source document does not include any sample consumption data or examples of exploratory data analysis performed on such data. Therefore, I cannot demonstrate EDA using sample data from this source.
Exploratory Data Analysis on consumption data would typically involve summarizing key statistics (e.g., total sales, average purchase value, purchase frequency), visualizing distributions (e.g., sales over time, geographical sales patterns, price ranges), identifying trends and anomalies (e.g., spikes in sales, changes in purchasing patterns), and exploring relationships between different data points (e.g., purchase volume vs. location, customer segmentation vs. spending).

7. Short Literature Review Linking to Cited Papers

The provided source is a review article that extensively cites other research papers demonstrating the use of consumption data in finance and business. Here are some examples of research cited in the source that use or discuss consumption data:

- Froot, Kang, Ozik, & Sadka (2017): This paper is cited in relation to the use of real-time corporate sales data (a form of consumption data) and what it says about earnings surprises and post-announcement returns.
- Agarwal, Qian, & Zou (2021): This study uses disaggregated sales and transaction-level credit card spending data to predict excess returns, finding a correlation between increased customer spending and cumulative abnormal returns.
- Niu, Xie, Chen, Zhao, & Wu (2023): This research examines the information gain from e-commerce sales data (consumption data) for analyst earnings forecasts and enterprise value evaluation.
- Li & Liu (2023): This paper demonstrates how leveraging online sales data from major e-commerce sites can reduce information asymmetry and mitigate stock price crash risk by lowering the possibility of managers concealing bad news.
- Sun, Y., et al. (2024). *Alternative data in finance and business: emerging applications and theory analysis (review)* . Financial Innovation, 10(1), 43. https://doi.org/10.1186/s40854-024-00652-0
- Zhu, F., & Zhang, X. (2010). Impact of online consumer reviews on sales: The moderating role of product characteristics. *Journal of Marketing Research* , 47(1), 133–148.

- Archak, N., Ghose, A., & Ipeirotis, P. G. (2011). Deriving the pricing power of product features by mining consumer reviews. *Management Science* , 57(8), 1485–1509.
- Niu, B., et al. (2023). E-commerce big data analytics for economic forecasting: Evidence from Taobao traffic. *Technological Forecasting and Social Change* , 188, 122251.