Analyzing E-commerce Trends through Data Wrangling

Introduction:
The rapid growth of e-commerce has generated extensive datasets capturing various aspects of online shopping behavior over time. For our company, FP E-commerce, effectively leveraging this data is crucial to understanding market trends, customer preferences, and sales performance. This project aims to analyze and transform our diverse e-commerce dataset to provide actionable insights into current and emerging trends of our product offerings.

Goals/Objectives:

1. Data Analysis/Cleanup: Review data set to make sure all the data is uniform and ready to use.
2. Data Transformation: Standardize and enhance the data for trend analysis and reporting.
3. Data Analysis Preparation: Prepare the cleaned and integrated data for in-depth analysis and visualization.
    a. The main feature we want to isolate is the Sales column, which would assist in analyzing the other independent features to review correlations and predict trends over time.

Scope:

1. Data Source: Sales transactions from our inner reporting system
2. Data Volume: 1 thousand records collected over the past year, with ~10 independent features.
3. Tools: Python with the help of modules like pandas & Matplotlib

Methodology:

1. Data Assessment:
    a. Conduct a thorough review of e-commerce data source.
    b. Identify and document key data quality issues, such as missing values, duplicates, and inconsistencies.

2. Data Cleaning:
    a. Use Python and Pandas to remove duplicate records and correct errors.
    b. Handle missing values through imputation or deletion as appropriate.
    c. Standardize data formats (e.g., dates, currency).

3. Data Transformation:
    a. Derive new metrics relevant to e-commerce trends.
    b. Normalize and scale data to facilitate analysis.

4. Data Analysis Preparation:
    a. Organize the transformed data into a structured format suitable for trend analysis.
    b. Create preliminary visualizations to identify key e-commerce trends using the Sales feature along with the other independent features.

Deliverables:

1. Cleaned Data: A dataset devoid of duplicates, errors, and inconsistencies.
2. Integrated Data: A unified dataset encompassing all relevant e-commerce data sources.
3. Transformation Scripts: Documented scripts and methodologies for data cleaning, integration, and transformation.
4. Preliminary Insights: Initial findings and visualizations highlighting key e-commerce trends through regression modules on key Sales features.

Conclusion:

By executing this data wrangling project, FP E-commerce will gain a robust and reliable dataset that uncovers vital e-commerce trends through its knowledge of past Sales. This enhanced understanding of market dynamics, customer behavior, and sales performance will enable data-driven decision-making, optimize marketing strategies as well as inventory, and ultimately drive business growth through profitability.