Advanced E-commerce Analytic and Sales Prediction

Problem Statement:

Advanced Exploratory Data Analysis (EDA)

a. **Multivariate Analysis**: Use advanced visualization techniques like parallel coordinates plot to analyze the relationships between multiple numerical variables.
b.Time Series Analysis: Explore the temporal aspects of the data, considering the 'Order Date' column. Utilize techniques like seasonal decomposition and autocorrelation to identify patterns over time.
b. **Geospatial Analysis**: If the data includes geographic information, perform a geospatial analysis to understand sales patterns across different regions. Use maps and heatmaps for visualization.
c. **Advanced Correlation Analysis**: Besides a traditional heatmap, use techniques like partial correlation analysis to identify relationships between variables while controlling for other factors.
d. **Hierarchical Clustering**: Apply hierarchical clustering to group similar products or categories based on sales patterns and explore relationships within these clusters.
e. **Distribution Analysis**: Utilize advanced statistical distributions like Poisson or Gamma distributions to model the distribution of sales and identify potential patterns.
f. **Interactive Visualizations**: Create interactive dashboards using tools like Plotly or Tableau for a dynamic exploration of data patterns.
g. **Advanced Pair Plot**: Enhance the pair plot by including additional information such as product categories or customer segments to observe multidimensional relationships.

Feature Engineering and Selection

a. **Dimensionality Reduction**: Implement dimensionality reduction techniques like Principal Component Analysis (PCA) to reduce the number of features while preserving essential information.
b. **Feature Importance Analysis**: Use advanced feature importance techniques such as recursive feature elimination with cross-validation to identify and select the most relevant features for sales prediction.

Outlier Detection and Treatment

a. **Advanced Outlier Detection**: Apply sophisticated outlier detection algorithms such as Isolation Forest or Local Outlier Factor to identify and handle outliers.
b. **Anomaly Detection**: Explore anomaly detection techniques to identify unusual patterns or behaviors in sales data that may impact predictions.

Imputation Techniques

a. **Predictive Imputation**: Use machine learning models to predict missing values based on other relevant features for a more accurate imputation.
b. **Time-Series Imputation**: If applicable, leverage time-series imputation methods to fill missing values considering the temporal aspect of the data.

Advanced Data Cleaning

a. **Regular Expressions**: Implement regular expressions to identify and clean textual data, ensuring consistency and uniformity.
b. **Advanced Data Transformation**: Apply advanced data transformation techniques such as logarithmic transformations or Box-Cox transformations to achieve better normalization.

Statistical Testing and Hypothesis Analysis

a. **A/B Testing**: If there are different strategies or interventions in the data, conduct A/B testing to assess their impact on sales.
b. **Hypothesis Testing**: Formulate hypotheses about factors influencing sales and perform statistical tests to validate or reject these hypotheses.