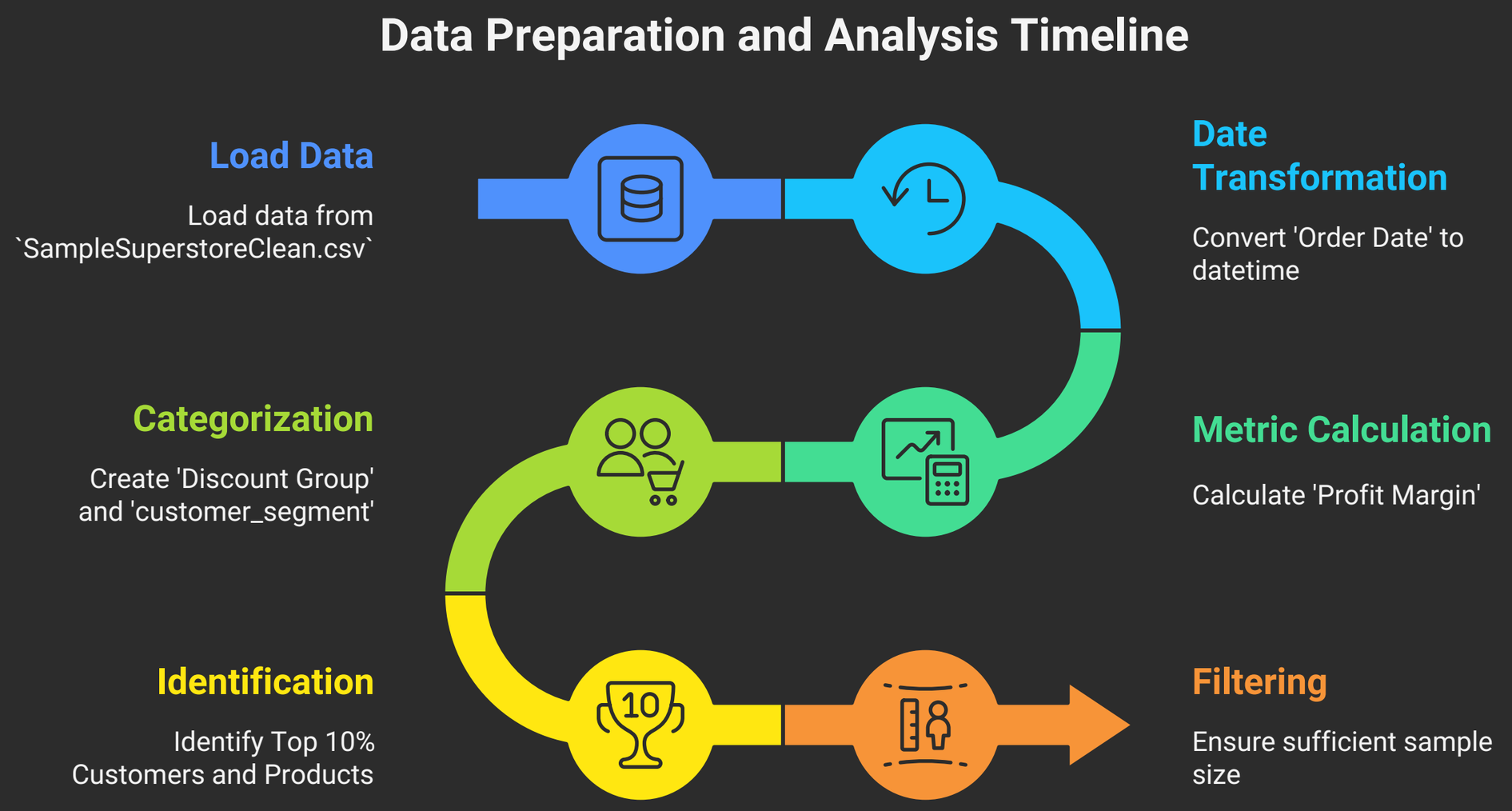
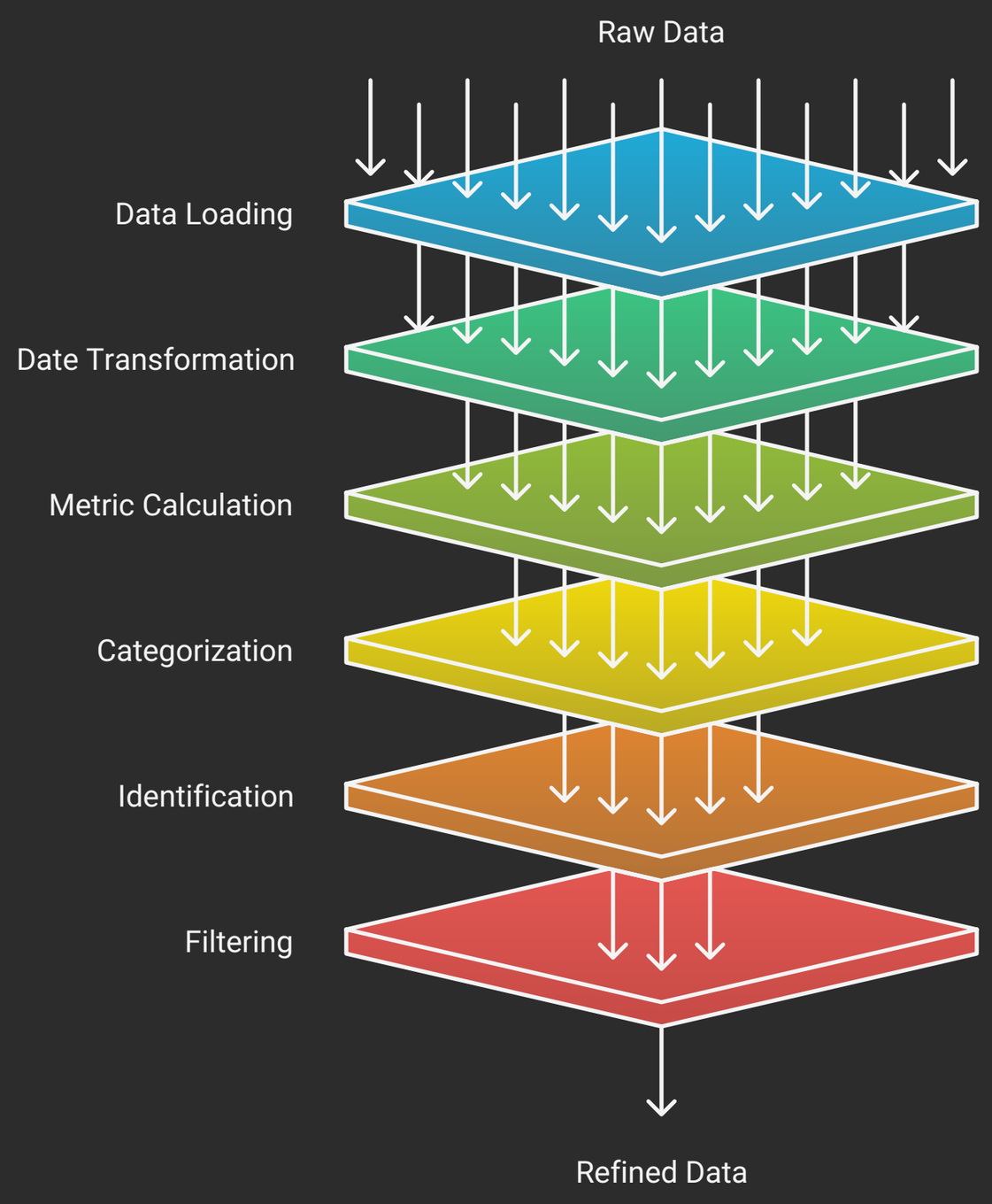
Superstore Business Insights: Analytical Workflow

Here is a mind map illustrating the "Superstore Business Insights: Analytical Workflow and Recommendations," drawing upon the provided sources and our conversation history:

Superstore Business Insights: Analytical Workflow Mind Map

- **Central Theme: Superstore Business Insights Project**
    - **Objective:** Identify **key profitability drivers** & provide **evidence-based recommendations**
- **Phase 1: Data Loading & Preprocessing**
- **Load Data:** SampleSuperstoreClean.csv
- **Date Transformation:**
    - Convert 'Order Date' to datetime
    - Extract 'Order Month' & 'Order Quarter'
- **Metric Calculation:**
    - Calculate **'Profit Margin'** (Profit / Sales)
- **Categorization:**
    - Create **'Discount Group'** (e.g., '0-20%', '20-30%')
    - Compute RFM metrics & assign **'customer_segment'** (e.g., 'High-Value', 'Loyal', 'At-Risk')
    - Create **'Group'** ('Furniture' vs. 'Others')
- **Identification:** Identify **Top 10% Customers** and **Top 10% Products** by Profit
- **Filtering:** Ensure **sufficient sample size** (min. 20 observations per group)





- **Phase 2: Key Finding - Statistical Tests** (Repeated for each Key Area, e.g., Regional, Category, Discount, Customer, Furniture, Seasonal, Monthly)
    - **Objective & Hypotheses:**
        - **Objective:** Analyze specific business aspects
        - **H_0:** Distributions are equal across groups
        - **H_1:** Distributions differ across groups
    - **Assumption Check (Parametric Tests):**
        - **Normality Tests:**
            - **Shapiro-Wilk** (for n < 50)
            - **D'Agostino K^2** (for n $\geq$ 50)
    - **Homogeneity of Variances Test: Levene's test**
    - **Decision:** Are **BOTH Normality AND Homogeneity assumptions met?**
        - **IF YES (Assumptions Met):**
            - **Action:** Perform **Parametric Test:**
                - **ANOVA** (for 3+ groups)
                - **T-test** (for 2 groups, e.g., Furniture vs. Others)
            - **Decision:** Is p-value < **alpha (0.05)**?
                - **IF YES (Significant):** Perform **Tukey HSD post-hoc test**
                - **IF NO (Not Significant):** Fail to reject H_0
        - **IF NO (Assumptions NOT Met):**
            - **Action:** Perform **Non-Parametric Test:**

        - **Kruskal-Wallis H-test** (for 3+ groups)
        - **Mann-Whitney U test** (for 2 groups)
    - **Decision:** Is p-value < **alpha (0.05)**?
        - **IF YES (Significant):** Perform **Dunn's post-hoc test**
        - **IF NO (Not Significant):** Fail to reject H_0
- **Statistical Conclusion & Insights:**
    - State test statistic & p-value
    - **Conclusion: Reject H_0** or **Fail to Reject H_0**
    - Summarize **significant pairs** from post-hoc tests
    - Rank groups by **median metric**
    - Generate **visualizations** (Box Plots, Median Bar Charts, Post-hoc Heatmaps)
    - **Derive Business Insights** (e.g., "West and East are high-performing regions")

Statistical Tests Comparison

Characteristic	Parametric Test	Non-Parametric Test
Number of Groups	ANOVA (3+)	Kruskal-Wallis H-test (3+)
Number of Groups	T-test (2)	Mann-Whitney U test (2)
Post-Hoc Test (Significant)	Tukey HSD	Dunn's Test
Conclusion (Significant)	Reject H0	Reject H0
Conclusion (Not Significant)	Fail to Reject H0	Fail to Reject H0

- **Phase 3: Quantile Regression Modeling & Assumption Checks** (Repeated for areas desiring quantified median impact)
    - **Purpose:** Quantify **median impact** of predictors on target variable
    - **Data Preparation:**
        - Encode categorical variables as **dummy variables**
        - Define a **baseline group** for each
    - **Quantile Regression Assumption Checks:**
        - **Data Availability & No Missing Values:** Check. If NOT MET, **Action:** Impute
        - **Sufficient Sample Size:** Check. If NOT MET, **Recommendation:** Collect more data/simplify model
        - **No Extreme Outliers:** Check. If NOT MET, **Recommendation:** Verify data/sensitivity analysis

    - **No Multicollinearity:** Check **VIF < 5**. If NOT MET, **Recommendation:** Remove/combine predictors
    - **Correct Model Specification:** Check **all p-values** ≤ 0.05 & **Pseudo** $R^2 \geq 0.1$. If NOT MET, **Recommendation:** Add predictors/check non-linearity
    - **Independence of Observations:** Check **Durbin-Watson (1.5-2.5)**. If NOT MET, **Recommendation:** Check clustering/time-series
    - **Linearity at Quantile:** Check **p-values** ≤ 0.05 at q=0.25 & q=0.75. If NOT MET, **Recommendation:** Add non-linear terms/additional predictors
    - **Monotonicity of Conditional Quantiles:** Check **q=0.25** $\leq$ q=0.5 $\leq$ q=0.75. If NOT MET, **Recommendation:** Simplify model/explore
    - **Action: Refit models with robust standard errors** for reliable inference
- **Model Building & Interpretation:**
    - Fit **Quantile Regression model** (median, **q=0.5**)
    - Print **model summary**
    - Provide **business-friendly interpretations** of coefficients (quantified impact)

Quantile Regression Modeling Checks

Assumption	Check	If Not Met	Action/Recommendation
Data Availability & Missing Values	Check	NOT MET	Impute missing values
Sufficient Sample Size	Check	NOT MET	Collect more data/simplify model
No Extreme Outliers	Check	NOT MET	Verify data/sensitivity analysis
No Multicollinearity	Check VIF < 5	NOT MET	Remove/combine predictors
Correct Model Specification	Check all p-values ≤ 0.05 & Pseudo R2 ≥ 0.1	NOT MET	Add predictors/check non-linearity
Independence of Observations	Check Durbin-Watson (1.5-2.5)	NOT MET	Check clustering/time-series
Linearity at Quantile	Check p-values ≤ 0.05 at q=0.25 & q=0.75	NOT MET	Add non-linear terms/additional predictors
Monotonicity of Conditional Quantiles	Check q=0.25 ≤ q=0.5 ≤ q=0.75	NOT MET	Simplify model/explore constrained QR

- **Phase 4: Conclusion & Recommendations**
    - **Conclusion:** Synthesize insights from **statistical tests** and **quantile regression**, emphasizing **statistically robust findings** ($p < 0.0001$)
    - **Evidence-Based Recommendations:**
        - **Regional Focus:** Increase marketing/inventory in West & East
        - **Category Prioritization:** Allocate resources to **Technology & Office Supplies**; reduce emphasis on Furniture
        - **Discount Strategy:** Implement **20% discount cap** (especially for Furniture); test thresholds
        - **Customer Engagement:** Launch **targeted campaigns** for At-Risk & Lost segments
        - **Furniture Optimization:** Reduce costs or adjust pricing strategies (e.g., bundling)
        - **Seasonal Strategies:** Maintain **uniform marketing** across quarters (minimal trends)
        - **Performance Monitoring:** Shift to **quarterly/annual analysis** for top customers/products; focus promotions on high-profit months (May/Sept)
    - **Model Limitations:**
        - **Low Explanatory Power** (Pseudo R^2 often < 0.01) (based on conversation history, not explicitly in current source)
        - **Linearity Issues** at specific quantiles for some models (based on conversation history)
        - Interpretations focus on **trends** rather than precise predictions
        - Non-parametric tests (Kruskal-Wallis, Dunn's) used to validate group differences
        - No multicollinearity and acceptable outlier levels confirmed
    - **Proposed Improvements (Future Work):**
        - Add **more predictors** (e.g., store size, marketing spend)
        - Explore **non-linear relationships** (e.g., polynomial terms)

- Apply **scaling or transformations** (e.g., log-transform, standardize) (based on conversation history, not explicitly in current source)

Recommendations based on statistical analysis

Recommendation Area	Recommendation
Regional Focus	Increase marketing/inventory in West & East
Category Prioritization	Allocate resources to Technology & Office Supplies; reduce emphasis on Furniture
Discount Strategy	Implement 20% discount cap (especially for Furniture); test thresholds
Customer Engagement	Launch targeted campaigns for At-Risk & Lost segments
Furniture Optimization	Reduce costs or adjust pricing strategies (e.g., bundling)
Seasonal Strategies	Maintain uniform marketing across quarters (minimal trends)
Performance Monitoring	Shift to quarterly/annual analysis for top customers/products; focus promotions on high-profit months (May/Sept)

Superstore Business Insights: Analytical Workflow and Recommendations

Superstore Business Insights Project

Data Loading & Preprocessing

- Load Data
- Date Transformation
- Metric Calculation
- Categorization
- Identification
- Filtering

Statistical Tests

- Objective & Hypotheses
- Assumption Check
- Statistical Conclusion & Insights

Quantile Regression Modeling

- Data Preparation
- Assumption Checks
- Model Building & Interpretation

Conclusion & Recommendations

- Regional Focus
- Category Prioritization
- Discount Strategy
- Customer Engagement
- Furniture Optimization
- Seasonal Strategies
- Performance Monitoring