

# Data Validator Agent

Upload your CSV and ask a question. Files must be  $\leq 10$  MB and < 50,000 rows.

Request Type

Price Strategy

(?)

Your Question

Compare the total sales revenue of the 'West' region versus the 'East' region. What is the difference in total sales between these two regions?

(?)

Additional Context (Optional)

Evaluate the sales performance across different geographic regions using the 'Region' and 'Sales' columns.

Upload CSV ( $\leq 10$  MB)
Drag and drop file here
Browse files
X

Analyze

Analysis started! Fetching results...

Analysis Complete!

## Summary

The analysis successfully loaded a Superstore dataset containing 9,994 transactions across 4 US regions (West, East, South, Central). However, the execution results only provided basic statistical summaries without completing the specific regional comparison requested. The dataset shows West region has the highest transaction frequency (3,203 records), but the actual aggregated sales totals for West vs East regions and their difference were not calculated in the execution output. The data quality is excellent with minimal missing values and appropriate data types for analysis.

Confidence Score

**45%**

## Key Insights

Dataset contains 9,994 transactions with 4 distinct regions: West (3,203 transactions - most frequent), followed by other regions

Overall sales statistics show mean of 229.86 with high variability (std dev 623.25), ranging from 0.44 to 22,638.48

Sales data is highly right-skewed: median (54.49) is significantly lower than mean (229.86), indicating presence of high-value outliers

Zero null values in Sales column and only 11 missing postal codes out of 9,994 records indicate excellent data quality

The execution did NOT complete the requested calculation - regional sales aggregation and West vs East comparison are missing from results

### Risks & Limitations

Execution Only basic statistics generated

Cannot verify if both 'West' and 'East' regions exist in dataset from current output (only region frequency counts provided)

Fast-path execution bypassed the specific aggregation logic needed to answer the business question

High sales variability (std dev 2.7x mean) suggests potential outliers that could skew regional comparisons

Dataset size discrepancy: validation shows 2,000 rows but execution shows 9,994 rows - unclear which is correct

## Recommended Next Steps

1. CRITICAL: Re-execute analysis with proper aggregation - calculate SUM(Sales) GROUP BY Region for West and East regions specifically
2. Calculate absolute difference and percentage difference between West and East total sales
3. Verify actual region values in dataset and confirm both 'West' and 'East' exist (execution shows 4 regions but doesn't list them)
4. Resolve row count discrepancy between validation (2,000) and execution (9,994) datasets
5. Perform outlier analysis on high-value transactions (>\$10,000) to assess impact on regional comparisons
6. Extend analysis to include profit margins by region for comprehensive price strategy insights
7. Consider temporal analysis to identify seasonal patterns in regional sales performance

### ▼ Disclaimers

Based on 9,994 transaction rows (note: validation indicated 2,000 rows - discrepancy exists)

Assumes Sales column represents revenue in USD without currency conversion

Assumes no PII concerns with customer names and IDs present in dataset

Analysis incomplete - regional aggregation and comparison calculations were not executed

All data from United States only (single country dataset)

Time period not specified in execution results - unable to confirm date range coverage

### ▼ View Technical Details

```
{
  "validation": {
    "ok": true,
    "message": "Data is highly suitable for the price strategy analysis. The dataset contains all required columns ('Region' and 'Sales') with no missing values, sufficient data volume (2000 rows), and appropriate data types for the requested regional sales comparison."
  },
  "profiles": [
    {
      "file": "uploads\178acf13f72c4c9b82f2bd87964007c0_Superstore.csv",
      "rows": 2000,
      "columns": [
        0: "Row ID",
        1: "Order ID",
        2: "Order Date",
        3: "Ship Date",
        4: "Ship Mode",
        5: "Customer ID",
        6: "Customer Name",
        7: "Segment",
        8: "Country",
        10: "State",
        11: "Postal Code",
        12: "Region",
        13: "Product ID",
        14: "Category",
        15: "Sub-Category"
      ]
    }
  ]
}
```

```
-- . -- --o-->
16 : "Product Name"
17 : "Sales"
18 : "Quantity"
19 : "Discount"
20 : "Profit"
]
  "dtypes" : {
    "Row ID" : "int64"
    "Order ID" : "object"
    "Order Date" : "object"
    "Ship Date" : "object"
    "Ship Mode" : "object"
    "Customer ID" : "object"
    "Customer Name" : "object"
    "Segment" : "object"
    "City" : "object"
    "State" : "object"
    "Postal Code" : "int64"
    "Region" : "object"
    "Product ID" : "object"
    "Category" : "object"
    "Sub-Category" : "object"
    "Product Name" : "object"
    "Sales" : "float64"
    "Quantity" : "int64"
    "Discount" : "float64"
    "Profit" : "float64"
  }
  "nulls" : {
    "Row ID" : 0
    "Order ID" : 0
    "Order Date" : 0
    "Ship Date" : 0
    "Ship Mode" : 0
    "Customer ID" : 0
    "Customer Name" : 0
    "Country" : 0
    "City" : 0
    "State" : 0
    "Postal Code" : 0
    "Region" : 0
    "Product ID" : 0
    "Category" : 0
    "Sub-Category" : 0
    "Product Name" : 0
    "Sales" : 0
    "Quantity" : 0
    "Discount" : 0
    "Profit" : 0
  }
  "head" : [
    0 : {
      "Row ID" : 1
      "Order ID" : "CA-2017-152156"
      "Order Date" : "2017-11-08"
      "Ship Date" : "2017-11-11"
      "Ship Mode" : "Second Class"
      "Customer ID" : "CG-12520"
      "Segment" : "Consumer"
      "Country" : "United States"
      "City" : "Henderson"
      "State" : "Kentucky"
      "Postal Code" : 42420
      "Region" : "South"
      "Product ID" : "FUR-B0-10001798"
      "Category" : "Furniture"
      "Sub-Category" : "Bookcases"
      "Product Name" : "Bush Somerset Collection Bookcase"
    }
  ]
}
```

```
"Sales" : 261.96
"Quantity" : 2
"Discount" : 0
"Profit" : 41.9136
}
▼ 1 : {
    "Row ID" : 2
    "Order ID" : "CA-2017-152156"
    "Order Date" : "2017-11-08"
    "Ship Date" : "2017-11-11"
    "Ship Mode" : "Second Class"
    "Customer Name" : "Claire Gute"
    "Segment" : "Consumer"
    "Country" : "United States"
    "City" : "Henderson"
    "State" : "Kentucky"
    "Postal Code" : 42420
    "Region" : "South"
    "Product ID" : "FUR-CH-10000454"
    "Category" : "Furniture"
    "Sub-Category" : "Chairs"
    "Product Name" :
        "Hon Deluxe Fabric Upholstered Stacking Chairs, Rounded Back"
    "Sales" : 731.94
    "Quantity" : 3
    "Discount" : 0
    "Profit" : 219.582
}
▼ 2 : {
    "Row ID" : 3
    "Order ID" : "CA-2017-138688"
    "Order Date" : "2017-06-12"
    "Ship Mode" : "Second Class"
    "Customer ID" : "DV-13045"
    "Customer Name" : "Darrin Van Huff"
    "Segment" : "Corporate"
    "Country" : "United States"
    "City" : "Los Angeles"
    "State" : "California"
    "Postal Code" : 90036
    "Region" : "West"
    "Product ID" : "OFF-LA-10000240"
    "Category" : "Office Supplies"
    "Sub-Category" : "Labels"
    "Product Name" :
        "Self-Adhesive Address Labels for Typewriters by Universal"
    "Sales" : 14.62
    "Quantity" : 2
    "Discount" : 0
    "Profit" : 6.8714
}
▼ 3 : {
    "Row ID" : 4
    "Order Date" : "2016-10-11"
    "Ship Date" : "2016-10-18"
    "Ship Mode" : "Standard Class"
    "Customer ID" : "SO-20335"
    "Customer Name" : "Sean O'Donnell"
    "Segment" : "Consumer"
    "Country" : "United States"
    "City" : "Fort Lauderdale"
    "State" : "Florida"
    "Postal Code" : 33311
    "Region" : "South"
    "Product ID" : "FUR-TA-10000577"
    "Category" : "Furniture"
    "Sub-Category" : "Tables"
}
```

```

    "Product Name" : "Bretford CR4500 Series Slim Rectangular Table"
    "Sales" : 957.5775
    "Quantity" : 5
    "Discount" : 0.45
    "Profit" : -383.031
}
    "Row ID" : 5
    "Order ID" : "US-2016-108966"
    "Order Date" : "2016-10-11"
    "Ship Date" : "2016-10-18"
    "Ship Mode" : "Standard Class"
    "Customer ID" : "SO-20335"
    "Customer Name" : "Sean O'Donnell"
    "Segment" : "Consumer"
    "Country" : "United States"
    "City" : "Fort Lauderdale"
    "State" : "Florida"
    "Postal Code" : 33311
    "Region" : "South"
    "Product ID" : "OFF-ST-10000760"
    "Category" : "Office Supplies"
    "Sub-Category" : "Storage"
    "Product Name" : "Eldon Fold 'N Roll Cart System"
    "Sales" : 22.368
    "Quantity" : 2
    "Discount" : 0.2
    "Profit" : 2.5164
]
    "stats" : {
        "Row ID" : {
            "mean" : 1000.5
            "std" : 577.4945887192364
            "min" : 1
            "max" : 2000
        }
        "Postal Code" : {
            "mean" : 53734.655
            "std" : 32004.737003782982
            "min" : 1841
            "max" : 99207
        }
        "Sales" : {
            "mean" : 228.94903194999998
            "std" : 525.0289137475023
            "min" : 1.08
            "max" : 8159.952
        }
        "Quantity" : {
            "mean" : 3.788
            "min" : 1
            "max" : 14
        }
        "Discount" : {
            "mean" : 0.15984
            "std" : 0.2071652314375129
            "min" : 0
            "max" : 0.8
        }
        "Profit" : {
            "mean" : 21.6250913
            "std" : 194.98833898298062
            "min" : -3839.9904
            "max" : 3177.475
        }
    }
}
]
    "suggestions" : [
        0 :
            "Verify that 'West' and 'East' regions exist in the dataset (sample"

```

```

    shows 'South' and 'West' regions)"
1 : "comprehensive regional performance insights"
2 :
    "Leverage additional columns like 'Profit', 'Discount', and
    'Category' to deepen the price strategy analysis beyond just
    revenue comparison"
]

"relevance_check" :
"PASSED - Dataset is directly relevant. Contains transactional sales
data with Region and Sales columns needed for geographic revenue
comparison. Data quality is excellent with zero null values across all
fields."
}

"plan" : {
    "task_type" : "analyze"
}

"steps" : [
    0 : {
        "id" : "step_1"
        "action" : "load_data"
        "params" : {
            "file" :
                "uploads\178acf13f72c4c9b82f2bd87964007c0_Superstore.csv"
            "columns" : [
                0 : "Region"
                1 : "Sales"
            ]
        }
    }
    1 : {
        "id" : "step_2"
        "action" : "filter_and_aggregate"
        "params" : {
            "group_by" : "Region"
            "aggregations" : {
                "Sales" : "sum"
            }
        }
        "filter_regions" : [
            0 : "West"
            1 : "East"
        ]
    }
    2 : {
        "id" : "step_3"
        "action" : "calculate_difference"
        "params" : {
            "metric" : "Sales"
            "regions" : [
                0 : "East"
            ]
        }
    }
    3 : {
        "id" : "step_4"
        "action" : "summarize"
        "params" : {
            "include" : [
                0 : "total_sales_west"
                1 : "total_sales_east"
                2 : "difference"
                3 : "percentage_difference"
                4 : "interpretation"
            ]
        }
    }
]

"required_columns" : [
    0 : "Region"
    1 : "Sales"
]

"expected_outputs" : [
    1 : "Total sales revenue for East region"
    2 : "Absolute difference between West and East sales"
    3 : "Percentage difference"
]

```

```
    4 : "Regional performance comparison summary"
]
  ▼ "failure_modes" : [
    0 : "Region column contains values other than expected region names"
    1 : "Sales column contains null or negative values"
    2 : "West or East regions not present in dataset"
    3 : "Data type mismatch in Sales column"
]
}
  ▼ "plan_validation" : {
    "ok" : true
    ▼ "required_columns" : [
      0 : "Region"
      1 : "Sales"
    ]
    ▼ "available_columns" : [
      0 : "Row ID"
      1 : "Order ID"
      2 : "Sales"
      4 : "Ship Mode"
      5 : "Customer ID"
      6 : "Category"
      7 : "Region"
      8 : "Profit"
      9 : "Sub-Category"
      10 : "Quantity"
      11 : "Customer Name"
      12 : "Ship Date"
      13 : "Order Date"
      14 : "Segment"
      15 : "Country"
      16 : "City"
      17 : "Discount"
      18 : "Product Name"
      19 : "Postal Code"
      20 : "State"
    ]
  }
  ▼ "execution" : {
    "ok" : true
    0 : "Fast-path: computed basic statistics without LLM in 0.20s"
  }
  ▼ "outputs" : {
    ▼ "basic_stats" : {
      ▼ "Row ID" : {
        "count" : 9994
        "unique" : NaN
        "top" : NaN
        "freq" : NaN
        "mean" : 4997.5
        "std" : 2885.1636290974325
        "min" : 1
        "25%" : 2499.25
        "50%" : 4997.5
        "75%" : 7495.75
        "max" : 9994
      }
      ▼ "Order ID" : {
        "count" : 9994
        "unique" : 5009
        "top" : "CA-2018-100111"
        "freq" : 14
        "std" : NaN
        "min" : NaN
        "25%" : NaN
        "50%" : NaN
        "75%" : NaN
        "max" : NaN
      }
      ▼ "Order Date" : {
```

```
        "count" : 9994
        "unique" : 1236
        "top" : "2017-09-05"
        "freq" : 38
        "mean" : NaN
        "std" : NaN
        "min" : NaN
        "25%" : NaN
        "50%" : NaN
        "75%" : NaN
        "max" : NaN
    }
}
▼ "Ship Date": {
    "count" : 9994
    "top" : "2016-12-16"
    "freq" : 35
    "mean" : NaN
    "std" : NaN
    "min" : NaN
    "25%" : NaN
    "50%" : NaN
    "75%" : NaN
    "max" : NaN
}
▼ "Ship Mode": {
    "count" : 9994
    "unique" : 4
    "top" : "Standard Class"
    "freq" : 5968
    "mean" : NaN
    "std" : NaN
    "min" : NaN
    "25%" : NaN
    "50%" : NaN
    "75%" : NaN
}
▼ "Customer ID": {
    "count" : 9994
    "unique" : 793
    "top" : "WB-21850"
    "freq" : 37
    "mean" : NaN
    "std" : NaN
    "min" : NaN
    "25%" : NaN
    "50%" : NaN
    "75%" : NaN
    "max" : NaN
}
▼ "Customer Name": {
    "count" : 9994
    "unique" : 793
    "top" : "William Brown"
    "freq" : 37
    "mean" : NaN
    "std" : NaN
    "min" : NaN
    "25%" : NaN
    "50%" : NaN
    "75%" : NaN
    "max" : NaN
}
▼ "Segment": {
    "count" : 9994
    "unique" : 3
    "top" : "Consumer"
    "freq" : 5191
    "mean" : NaN
    "std" : NaN
    "min" : NaN
}
```

```
    "25%" : NaN
    "50%" : NaN
    "75%" : NaN
    "max" : NaN
  }
  ▾ "Country" : {
    "count" : 9994
    "unique" : 1
    "top" : "United States"
    "mean" : NaN
    "std" : NaN
    "min" : NaN
    "25%" : NaN
    "50%" : NaN
    "75%" : NaN
    "max" : NaN
  }
  ▾ "City" : {
    "count" : 9994
    "unique" : 531
    "top" : "New York City"
    "freq" : 915
    "mean" : NaN
    "std" : NaN
    "min" : NaN
    "25%" : NaN
    "50%" : NaN
    "75%" : NaN
    "max" : NaN
  }
  ▾ "State" : {
    "unique" : 49
    "top" : "California"
    "freq" : 2001
    "mean" : NaN
    "std" : NaN
    "min" : NaN
    "25%" : NaN
    "50%" : NaN
    "75%" : NaN
    "max" : NaN
  }
  ▾ "Postal Code" : {
    "count" : 9983
    "unique" : NaN
    "top" : NaN
    "freq" : NaN
    "mean" : 55245.23329660423
    "std" : 32038.71595549834
    "min" : 1040
    "25%" : 23223
    "50%" : 57103
    "max" : 99301
  }
  ▾ "Region" : {
    "count" : 9994
    "unique" : 4
    "top" : "West"
    "freq" : 3203
    "mean" : NaN
    "std" : NaN
    "min" : NaN
    "25%" : NaN
    "50%" : NaN
    "75%" : NaN
    "max" : NaN
  }
  ▾ "Product ID" : {
    "count" : 9994
```

```
        "unique" : 1862
        "top" : "OFF-PA-10001970"
        "freq" : 19
        "mean" : NaN
        "min" : NaN
        "25%" : NaN
        "50%" : NaN
        "75%" : NaN
        "max" : NaN
    }
    ▾ "Category" : {
        "count" : 9994
        "unique" : 3
        "top" : "Office Supplies"
        "freq" : 6026
        "mean" : NaN
        "std" : NaN
        "min" : NaN
        "25%" : NaN
        "50%" : NaN
        "75%" : NaN
        "max" : NaN
    }
    ▾ "Sub-Category" : {
        "count" : 9994
        "unique" : 17
        "freq" : 1523
        "mean" : NaN
        "std" : NaN
        "min" : NaN
        "25%" : NaN
        "50%" : NaN
        "75%" : NaN
        "max" : NaN
    }
    ▾ "Product Name" : {
        "count" : 9994
        "unique" : 1850
        "top" : "Staple envelope"
        "freq" : 48
        "mean" : NaN
        "std" : NaN
        "min" : NaN
        "25%" : NaN
        "50%" : NaN
        "75%" : NaN
        "max" : NaN
    }
    ▾ "Sales" : {
        "count" : 9994
        "unique" : NaN
        "top" : NaN
        "freq" : NaN
        "mean" : 229.85800083049827
        "std" : 623.2451005086808
        "min" : 0.444
        "25%" : 17.28
        "50%" : 54.48999999999995
        "75%" : 209.94
        "max" : 22638.48
    }
    ▾ "Quantity" : {
        "count" : 9994
        "unique" : NaN
        "top" : NaN
        "freq" : NaN
        "mean" : 3.789573744246548
        "std" : 2.2251096911414
        "min" : 1
        "25%" : 2
        "50%" : 3
        "75%" : 4
        "max" : 10
    }
}
```

```
        "75%" : 5
        "max" : 14
    }
    ▼ "Discount" : {
        "count" : 9994
        "unique" : NaN
        "top" : NaN
        "freq" : NaN
        "mean" : 0.1562027216329798
        "std" : 0.20645196782571615
        "min" : 0
        "25%" : 0
        "50%" : 0.2
        "75%" : 0.2
        "max" : 0.8
    }
    ▼ "Profit" : {
        "count" : 9994
        "unique" : NaN
        "top" : NaN
        "freq" : NaN
        "std" : 234.2601076909573
        "min" : -6599.978
        "25%" : 1.72875
        "50%" : 8.6665
        "75%" : 29.364
        "max" : 8399.976
    }
}
▼ "shape" : [
    0 : 9994
    1 : 21
]
▼ "columns" : [
    0 : "Row ID"
    1 : "Order ID"
    2 : "Order Date"
    3 : "Ship Date"
    4 : "Ship Mode"
    5 : "Customer ID"
    6 : "Customer Name"
    7 : "Segment"
    8 : "Country"
    10 : "State"
    11 : "Postal Code"
    12 : "Region"
    13 : "Product ID"
    14 : "Category"
    15 : "Sub-Category"
    16 : "Product Name"
    17 : "Sales"
    18 : "Quantity"
    19 : "Discount"
    20 : "Profit"
]
"summary" :
"Computed basic statistics in 0.20s via fast-path (no LLM)"
}
}
▼ "response" : {
    "tldr" : NULL
    "summary" :
    "The analysis successfully loaded a Superstore dataset containing 9,994 transactions across 4 US regions (West, East, South, Central). However, the execution results only provided basic statistical summaries without their aggregated totals (e.g., sum, mean, std, etc.), but the actual aggregated sales totals for West vs East regions and their difference were not calculated in the execution output. The data quality is excellent with minimal missing values and appropriate data types for analysis."
    "confidence" : 0.45
}
```

```

    ▾ "insights" : [
      0 :
        "Dataset contains 9,994 transactions with 4 distinct regions: West
         (3,203 transactions - most frequent), followed by other regions"
      1 :
        "Overall sales statistics show mean of $229.86 with high
         variability (std dev $623.25), ranging from $0.44 to $22,638.48"
      2 :
        "Sales data is highly right-skewed: median ($54.49) is
         significantly lower than mean ($229.86), indicating presence of
         high-value outliers"
      3 :
        "Zero null values in Sales column and only 11 missing postal codes
         out of 9,994 records indicate excellent data quality"
      4 :
        "The execution did NOT complete the requested calculation -
         regional sales aggregation and West vs East comparison are missing
         from results"
    ]
    "sales_bounds" :
      "PASS - Sales values range from $0.44 to $22,638.48, reasonable for
       retail transactions"
    "sales_non_negative" :
      "PASS - Minimum sales value is $0.44 (no negative values detected)"
    "data_completeness" :
      "PASS - Sales column has 9,994 non-null values matching total row
       count"
    "region_coverage" :
      "WARNING - Execution shows 4 regions exist but did not
       filter/aggregate for West and East specifically"
    "statistical_consistency" :
      "PASS - Mean ($229.86) within expected range given median ($54.49)
       and high std dev ($623.25) for retail data"
    "execution_completeness" :
      "FAIL - Requested calculations (West total, East total, difference)
       not present in execution output"
  ]
  ▾ "risks" : [
    0 :
      "Execution plan steps 2-4 (filter_and_aggregate,
       calculate_difference, summarize) were not executed - only basic
       statistics generated"
    1 :
      "Cannot verify if both 'West' and 'East' regions exist in dataset"
    2 :
      "Fast-path execution bypassed the specific aggregation logic needed
       to answer the business question"
    3 :
      "High sales variability (std dev 2.7x mean) suggests potential
       outliers that could skew regional comparisons"
    4 :
      "Dataset size discrepancy: validation shows 2,000 rows but
       execution shows 9,994 rows - unclear which is correct"
  ]
  ▾ "disclaimers" : [
    0 :
      "Based on 9,994 transaction rows (note: validation indicated 2,000
       rows - discrepancy exists)"
    1 :
      "Assumes Sales column represents revenue in USD without currency
       conversion"
    2 :
      "Assumes no PII concerns with customer names and IDs present in
       dataset"
    3 :
      "Analysis incomplete - regional aggregation and comparison
       calculations were not executed"
    4 :
      "All data from United States only (single country dataset)
       ... per row has specific ... execution results - unable to confirm
       date range coverage"
  ]
  ▾ "next_steps" : [
    0 :
      "CRITICAL: Re-execute analysis with proper aggregation - calculate
       SUM(Sales) GROUP BY Region for West and East regions specifically"
    1 :
      "Calculate absolute difference and percentage difference between
       West and East total sales"
    2 :
      "Verify actual region values in dataset and confirm both 'West' and
       'East' regions exist"
  ]

```

```
'East' exist (execution shows 4 regions but doesn't list them)"  
3 :  
"Resolve row count discrepancy between validation (2,000) and  
execution (9,994) datasets"  
4 :  
"Perform outlier analysis on high-value transactions (>$10,000) to  
assess impact on regional comparisons"  
5 :  
"Extend analysis to include profit margins by region for  
comprehensive price strategy insights"  
6 :  
"Consider temporal analysis to identify seasonal patterns in  
regional sales performance"  
]  
}  
"hitl_required" : false  
}
```