

Business Analytics

Chapter 3 Data Visualization



Introduction

 Data visualization involves:

 Creating a summary table for the data.

 Generating charts to help interpret, analyze, and learn from the data.

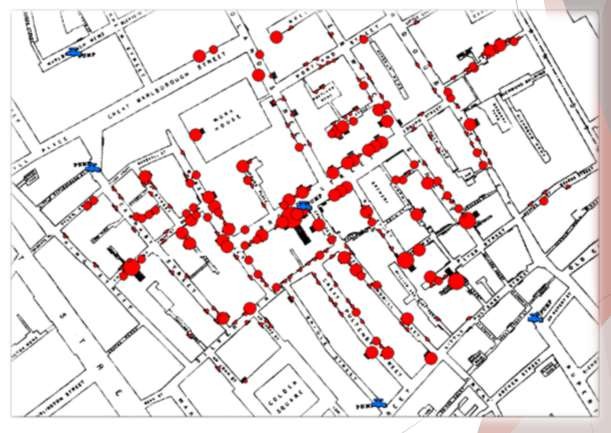
 Uses of data visualization:

 Helpful for identifying data errors.

 Reduces the size of your data set by highlighting important relationships and trends in the data.



Why is Data Visualization Important?



It Can Save Lives!

**London 1854**

 John Snow (Not that one!)

 Mapped Cholera cases

**What do you notice about the data?**

 Clustering around water pump!

 Cholera was water-borne!

**Results?**

 Transformed the medical profession

 Saved Lives!



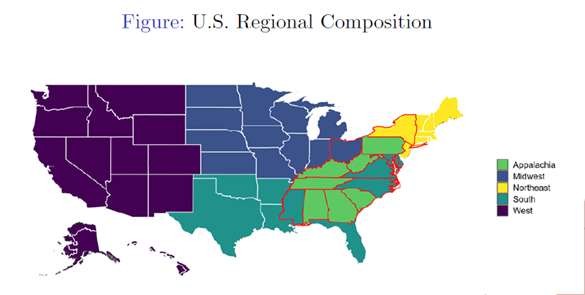
1. Know Your Audience

**What is the purpose?**

**Who will use the What are their data? objectives?**

**How will they interact with the data?**

**What business questions do users need answered?**



2. Use Space and Colors Well

**Use of Color is Powerful!**

 Aesthetics

 Higher Memorability

 Draws Attention

**Avoid Clutter**

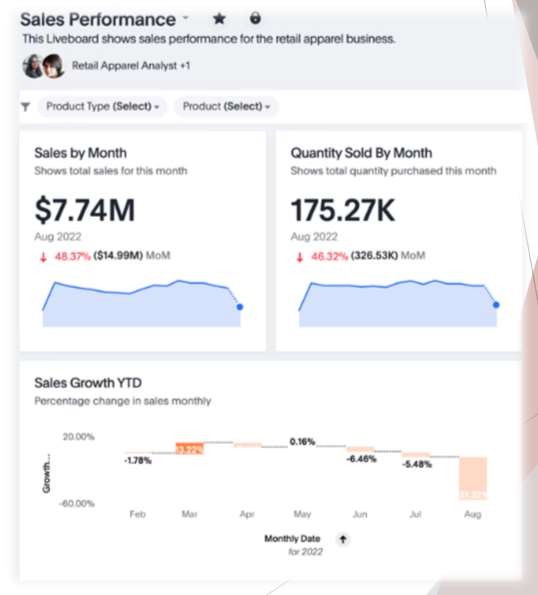
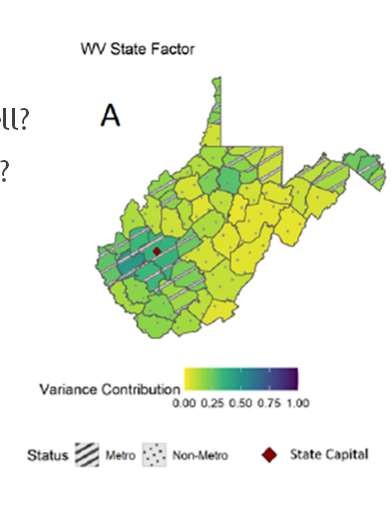
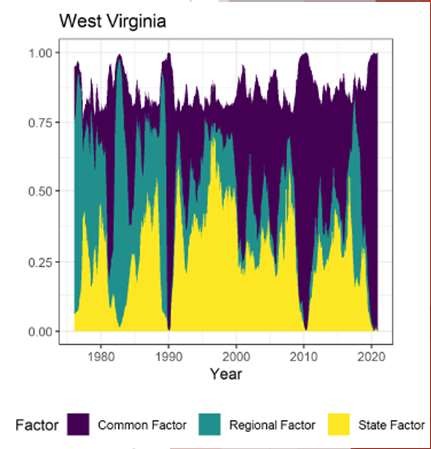
 Lose Actionable Information

**Used For:**

 Sequential Identification

 Show Divergence

 Categories



3. Highlight Important Information

**Tell a Story**

 **Point out what is important**

 **Make it easy to follow**

**Placement**

 **For Western Audiences:**

 **Top Left = Most Important**



4. Keep it Simple and Straightforward (K.I.S.S)

**Use Correct Figure Types**

 What story are you trying to tell?

 What type of data do you have?

**Make It Easy For Viewers**

 Less Distractions

 Less Conscious Thought



Overview of Data Visualization

Effective Design Techniques:

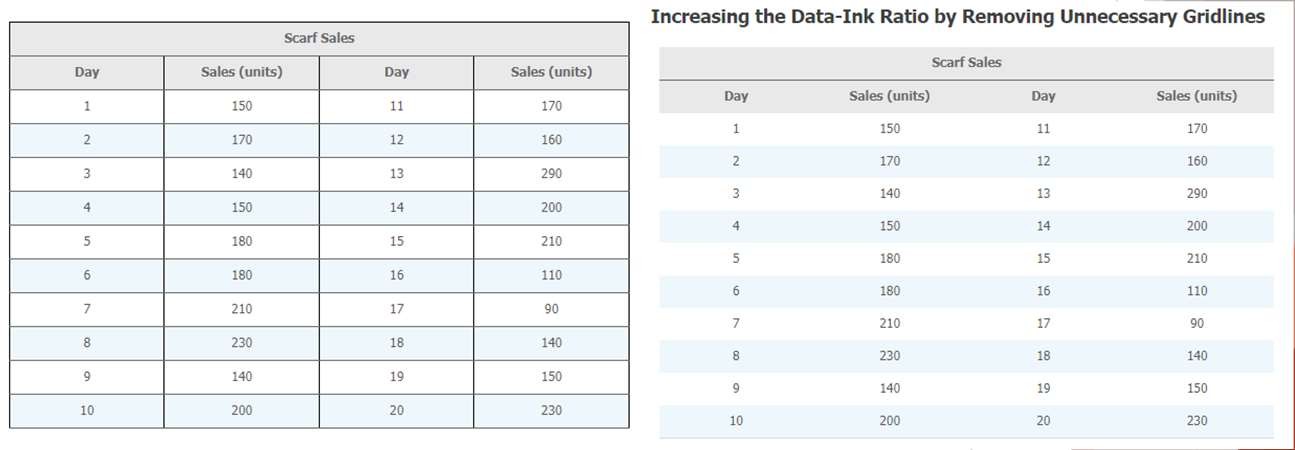
 **Data-ink ratio**:

 Measures the proportion of what Tufte terms “data-ink” to the total amount of ink used in a table or chart.

 Helpful for creating effective tables and charts for data visualization:

 Data-ink: Ink used in a table or chart that is necessary to convey the meaning of the data to the audience.

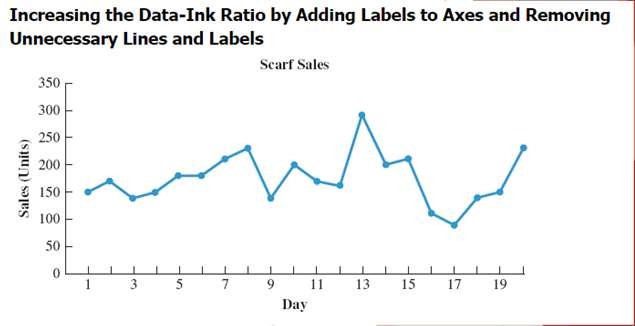
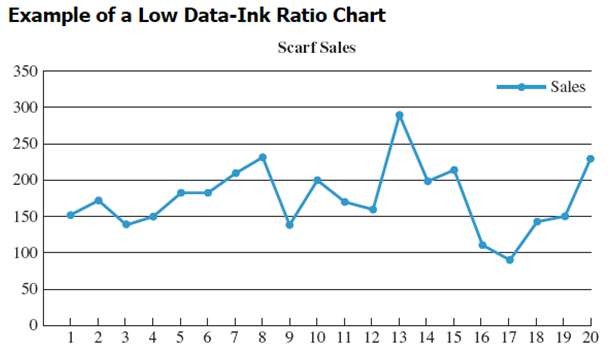
 Non-data-ink: Ink used in a table or chart that serves no useful purpose in conveying the data to the audience.



Which One is Better?

A

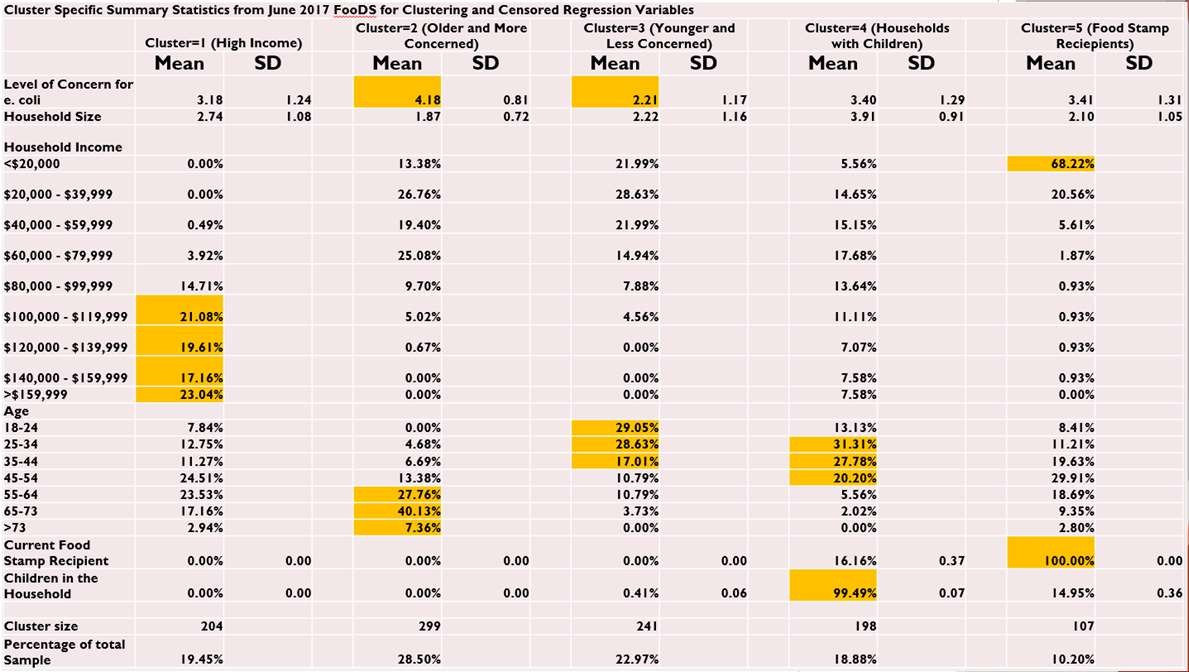
B



Which One is Better?

A

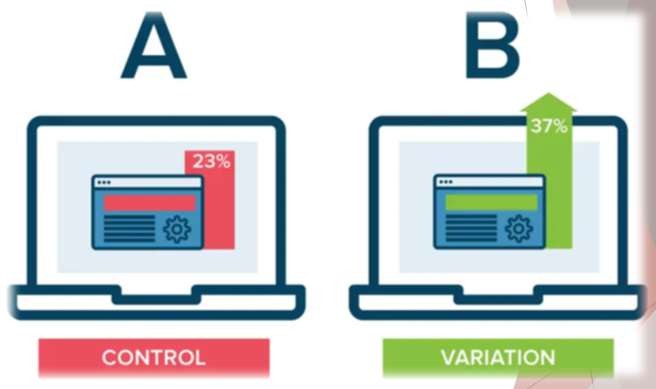
B



What’s Wrong With This?



|  |  |  |
| --- | --- | --- |
| **Choosing FSMA Exempt Tomatoes when ONLY Food Safety Label is Present** | | |
| **Variable** | Estimate | SE |
| **Intercept** | -2.0587\*\*\* | 0.4773 |
| **Importance of Food Safety** | -0.1234 | 0.1405 |
| **Female** | **0.2467\*** | **0.1486** |
| **Vegan and/or Vegetarian** | **-0.9259\*\*** | **0.4724** |
| **Spanish Origin** | 0.3266 | 0.236 |
| **Liberal** | -0.0143 | 0.1521 |
| **Food at home expenditure** | **0.3326\*** | **0.1738** |
| **Food at home expenditure squared** | **-0.0261\*** | **0.0152** |
| **\*, \*\*, \*\*\* indicated significance at the 10%, 5%, and 1% levels** | | |



5. Get Other Opinions

**Ask someone!**

 Ask a lot of people

 What looks better

 What performs better

**Example:** A/B Testing

 Compare 2 versions of something to figure out which performs better

 Typically used for websites and apps

 Started 1920



Tables

Table Design Principles Crosstabulation PivotTables in Excel

Recommended PivotTables in Excel



Should I Use Tables or Charts?

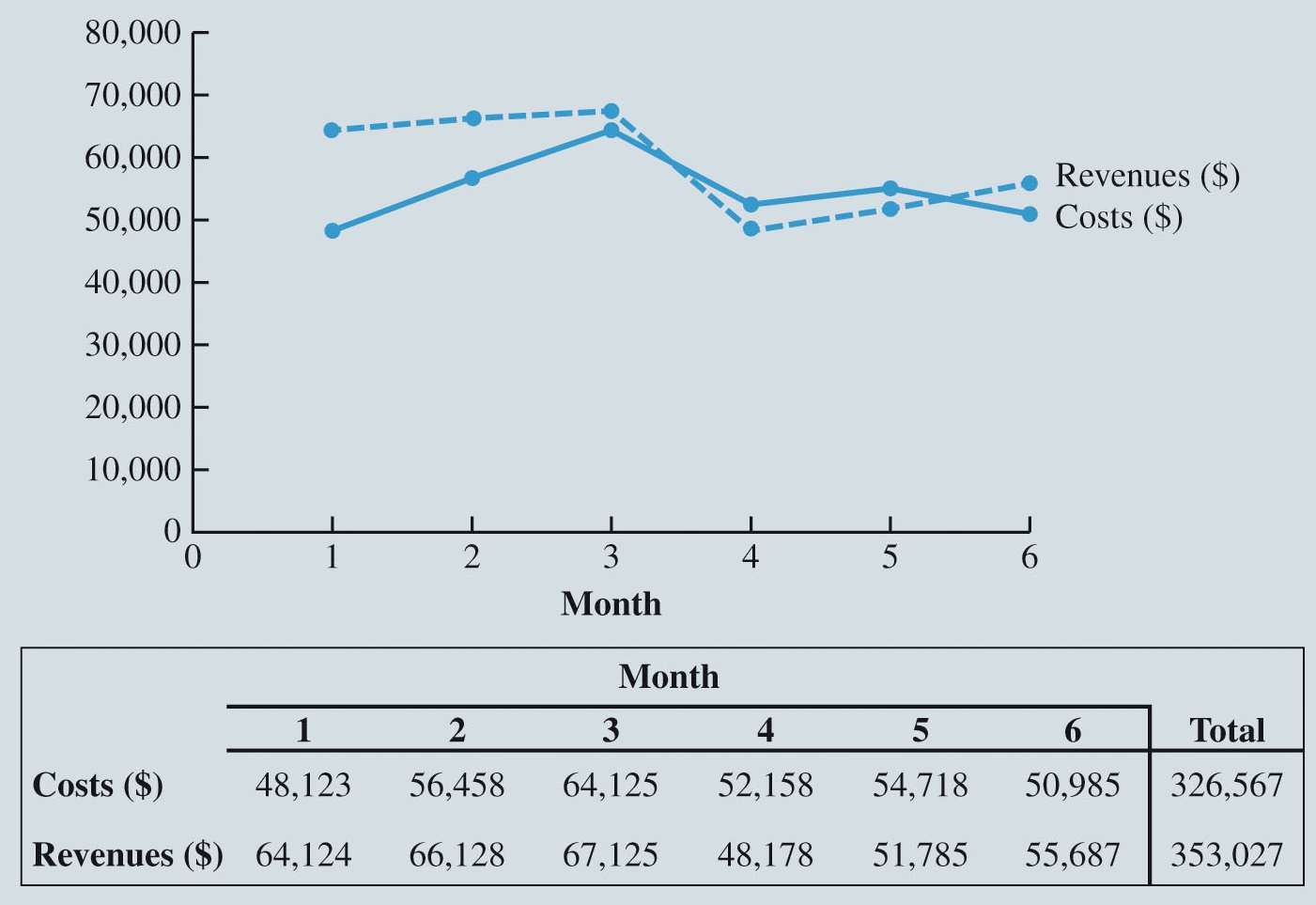
Tables should be used when:

1. The reader needs to refer to specific numerical values.
2. The reader needs to make precise comparisons between different values and not just relative comparisons.
3. The values being displayed have different units or very different magnitudes.



Tables

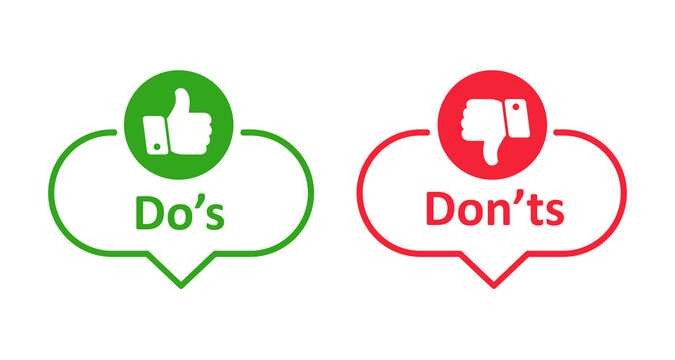
Table 3.4: Table Displaying Head Count, Costs, and Revenues at Gossamer Industries



Tables (4 of 18)

Figure 3.6: Combined Line Chart and Table for Monthly Costs and Revenues at Gossamer Industries

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Month 1** | **Month 2** | **Month 3** | **Month 4** | **Month 5** | **Month 6** | **Total** |
| **Head Count** | 8 | 9 | 10 | 9 | 9 | 9 |  |
| **Costs ($)** | 48,123 | 56,458 | 64,125 | 52,158 | 54,718 | 50,985 | 326,567 |
| **Revenues ($)** | 64,124 | 66,128 | 67,125 | 48,178 | 51,785 | 55,687 | 353,027 |

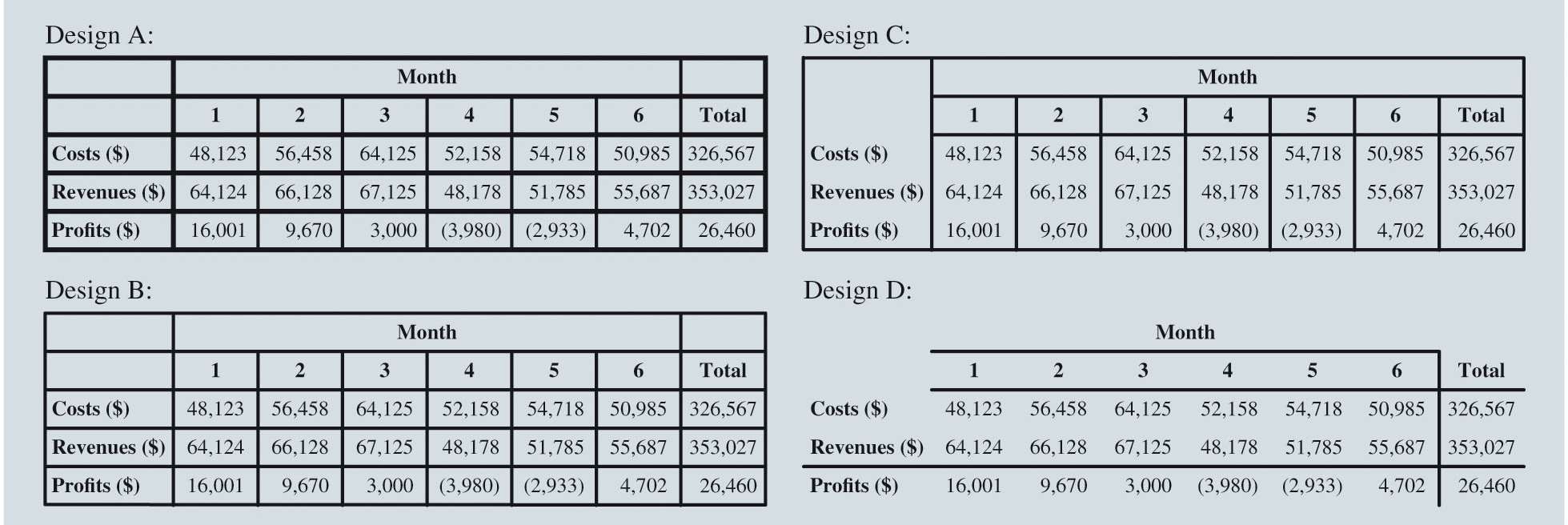


Tables

Table Design Principles:

Avoid using vertical lines in a table unless they are necessary for clarity.

Horizontal lines are generally necessary only for separating column titles from data values or when indicating that a calculation has taken place.



Tables

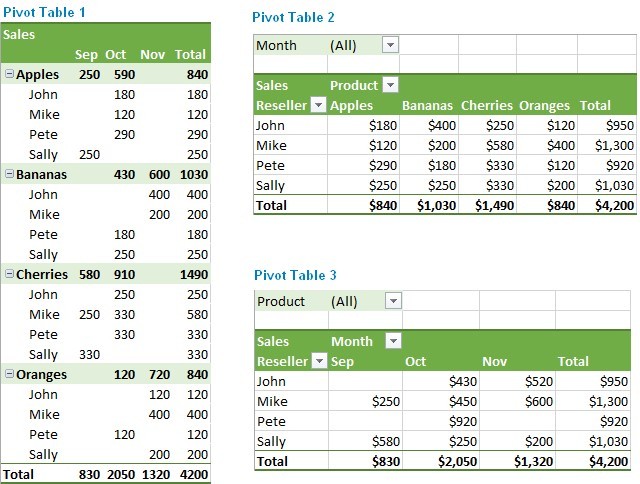
Figure 3.7: Comparing Different Table Designs



Tables

Table 3.5: Larger Table Showing Revenues by Location for 12 Months of Data

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Revenues by Location ($)** | **Month 1** | **Month 2** | **Month 3** | **Month 4** | **Month 5** | **Month 6** |
| **Temple** | 8,987 | 8,595 | 8,958 | 6,718 | 8,066 | 8,574 |
| **Killeen** | 8,212 | 9,143 | 8,714 | 6,869 | 8,150 | 8,891 |
| **Waco** | 11,603 | 12,063 | 11,173 | 9,622 | 8,912 | 9,553 |
| **Belton** | 7,671 | 7,617 | 7,896 | 6,899 | 7,877 | 6,621 |
| **Granger** | 7,642 | 7,744 | 7,836 | 5,833 | 6,002 | 6,728 |
| **Harker Heights** | 5,257 | 5,326 | 4,998 | 4,304 | 4,106 | 4,980 |
| **Gatesville** | 5,316 | 5,245 | 5,056 | 3,317 | 3,852 | 4,026 |
| **Lampasas** | 5,266 | 5,129 | 5,022 | 3,022 | 3,088 | 4,289 |
| **Academy** | 4,170 | 5,266 | 7,472 | 1,594 | 1,732 | 2,025 |
| **Total** | 64,124 | 66,128 | 67,125 | 48,178 | 51,785 | 55,687 |
| **Costs ($)** | 48,123 | 56,458 | 64,125 | 52,158 | 54,718 | 50,985 |



Tables

 **Crosstabulation**:

 A useful type of table for describing data of two variables.

 **PivotTable**:

 A crosstabulation in Microsoft Excel.



Tables

Table 3.6: Quality Rating and Meal Price for 300 Los Angeles Restaurants

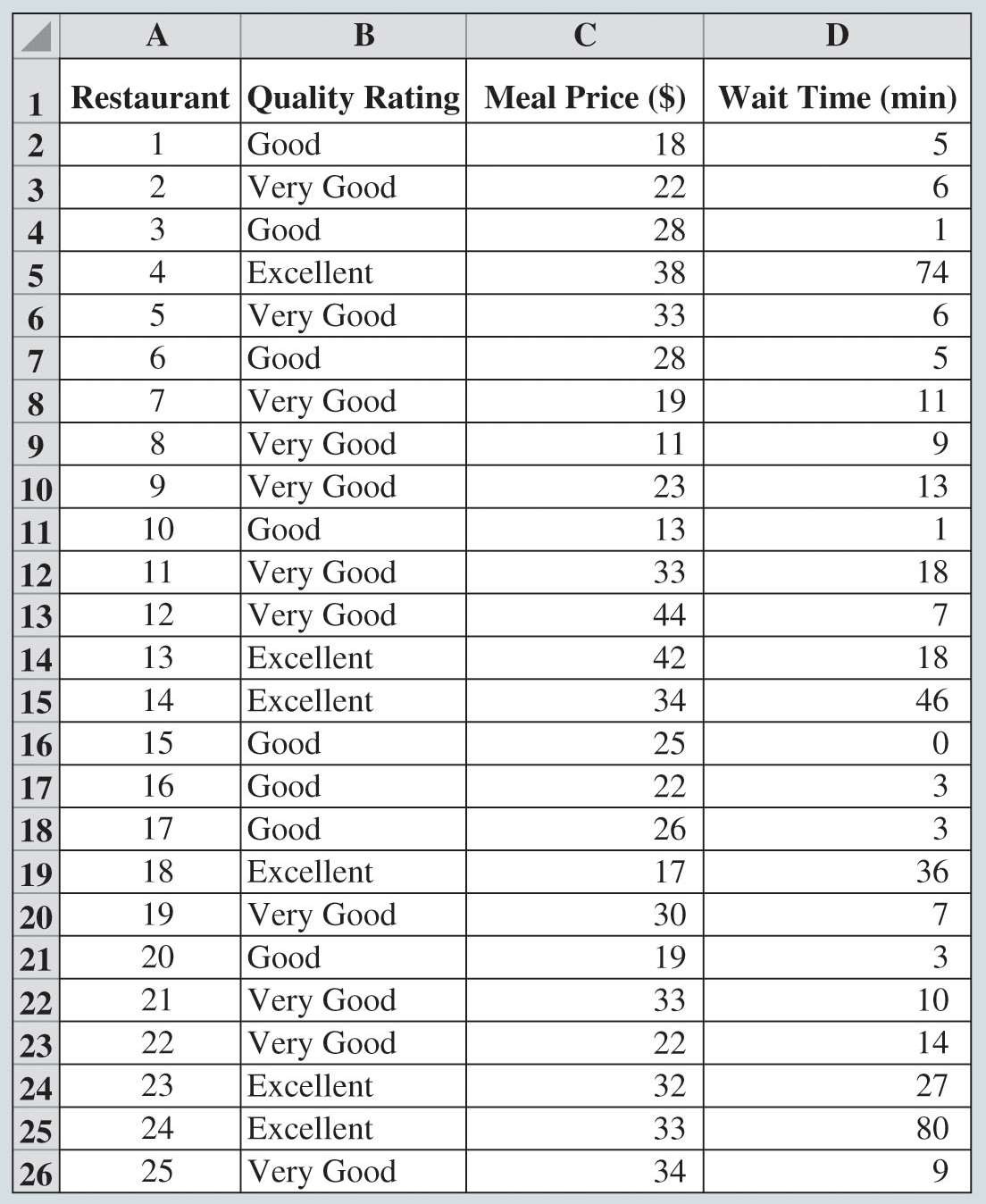


Tables

Table 3.7: Crosstabulation of Quality Rating and Meal Price for 300 Los Angeles Restaurants

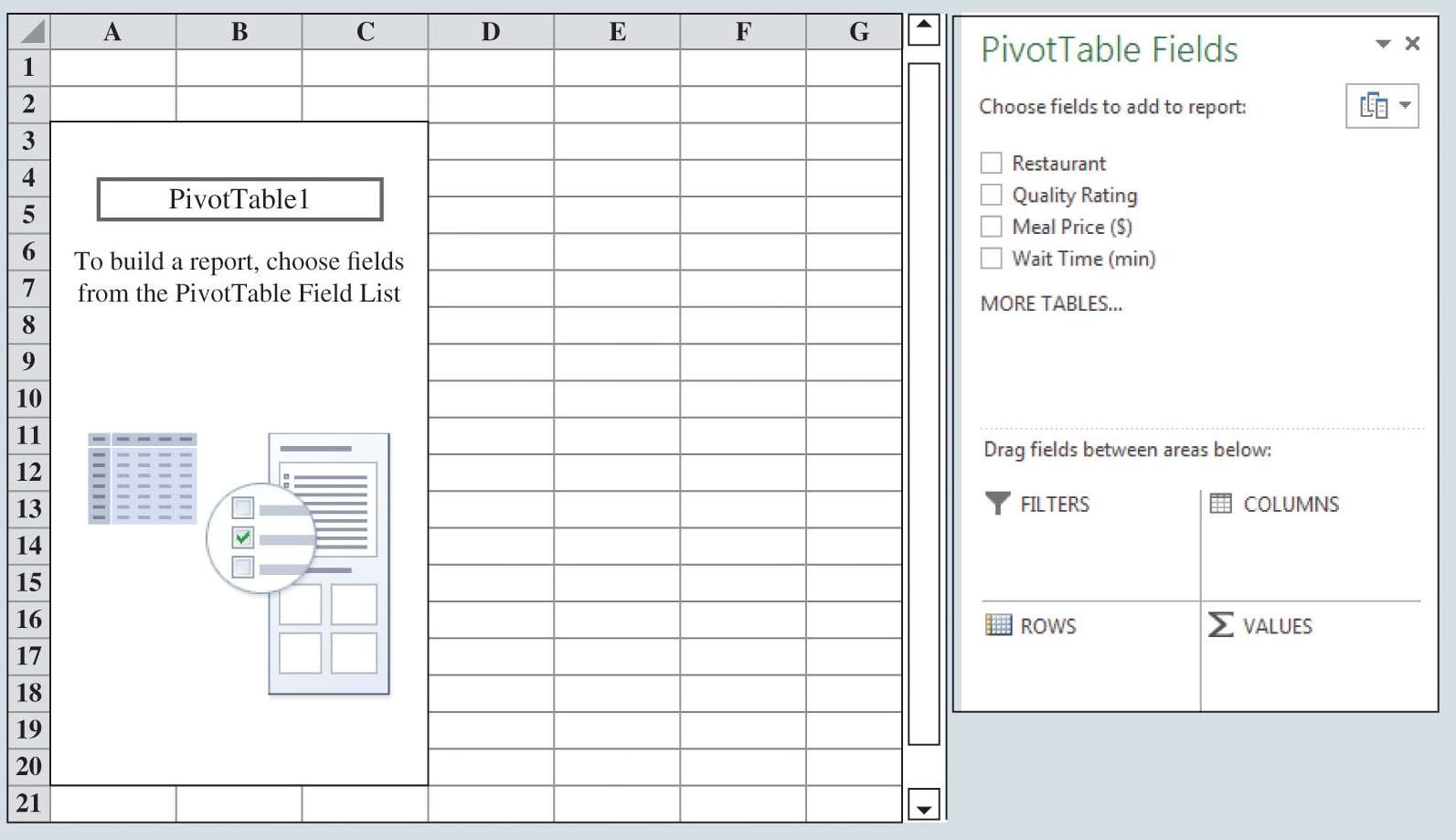
|  |  |  |  |
| --- | --- | --- | --- |
| **Restaurant** | **Quality Rating** | **Meal Price ($)** | **Wait Time (min)** |
| 1 | Good | 18 | 5 |
| 2 | Very Good | 22 | 6 |
| 3 | Good | 28 | 1 |
| 4 | Excellent | 38 | 74 |
| 5 | Very Good | 33 | 6 |
| 6 | Good | 28 | 5 |
| 7 | Very Good | 19 | 11 |
| 8 | Very Good | 11 | 9 |
| 9 | Very Good | 23 | 13 |
| 10 | Good | 13 | 1 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Meal Price** |  | | | |
| **Quality Rating** | **$10–19** | **$20–29** | **$30–39** | **$40–49** | **Total** |
| Good | 42 | 40 | 2 | 0 | 84 |
| Very Good | 34 | 64 | 46 | 6 | 150 |
| Excellent | 2 | 14 | 28 | 22 | 66 |
| Total | 78 | 118 | 76 | 28 | 300 |



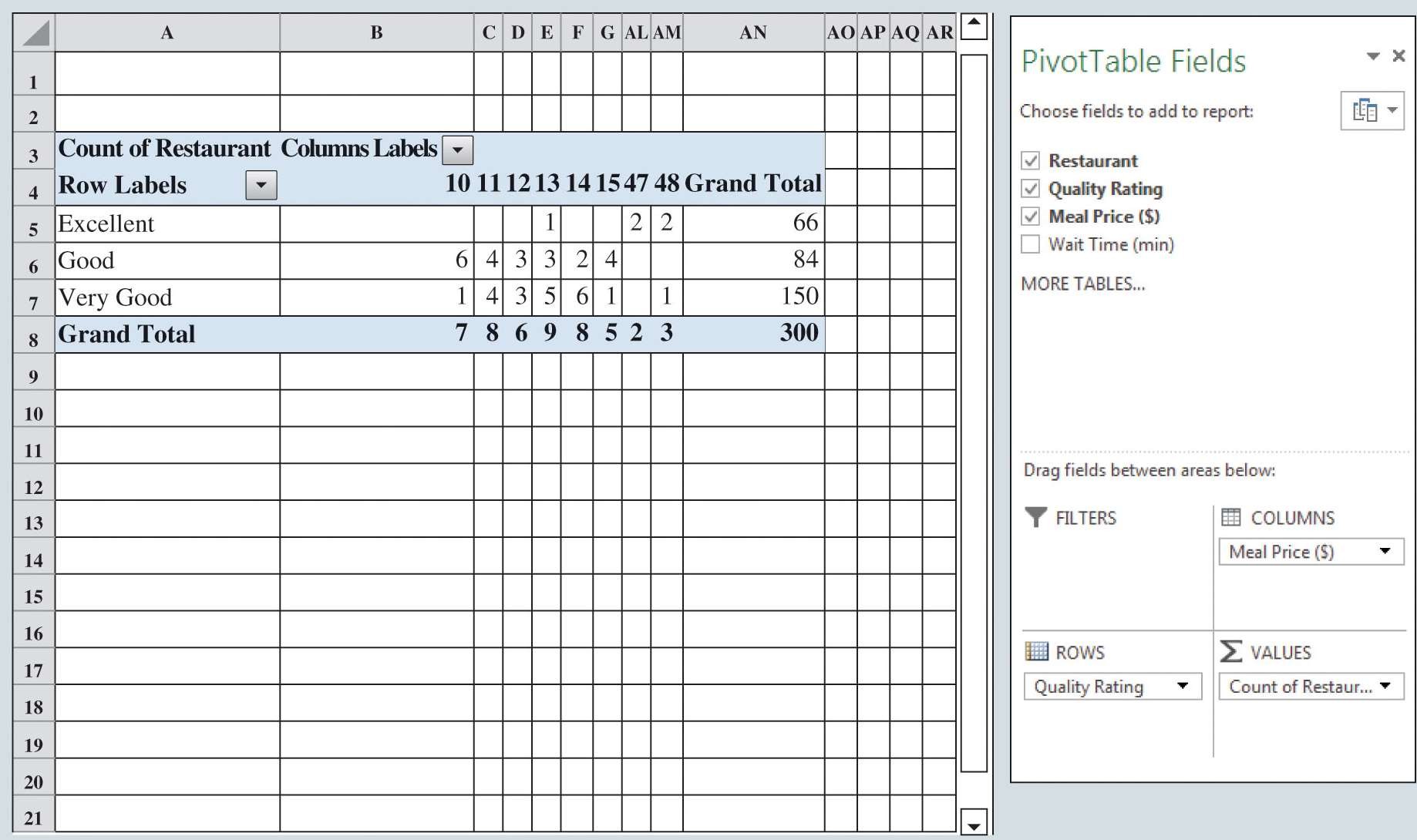
Tables

Figure 3.8: Excel Worksheet Containing Restaurant Data



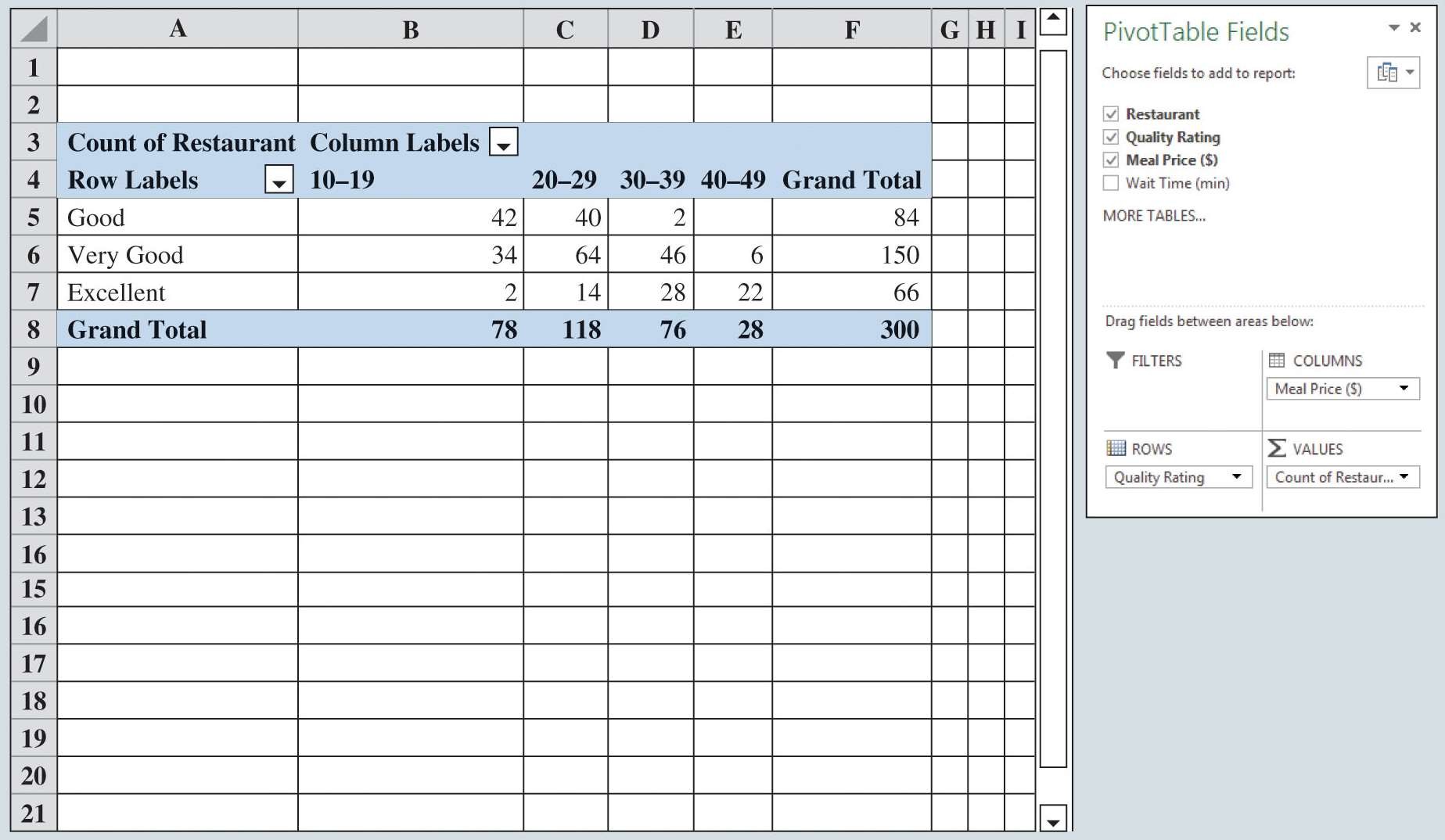
Tables

Figure 3.9: Initial PivotTable Field List and PivotTable Field Report for the Restaurant Data



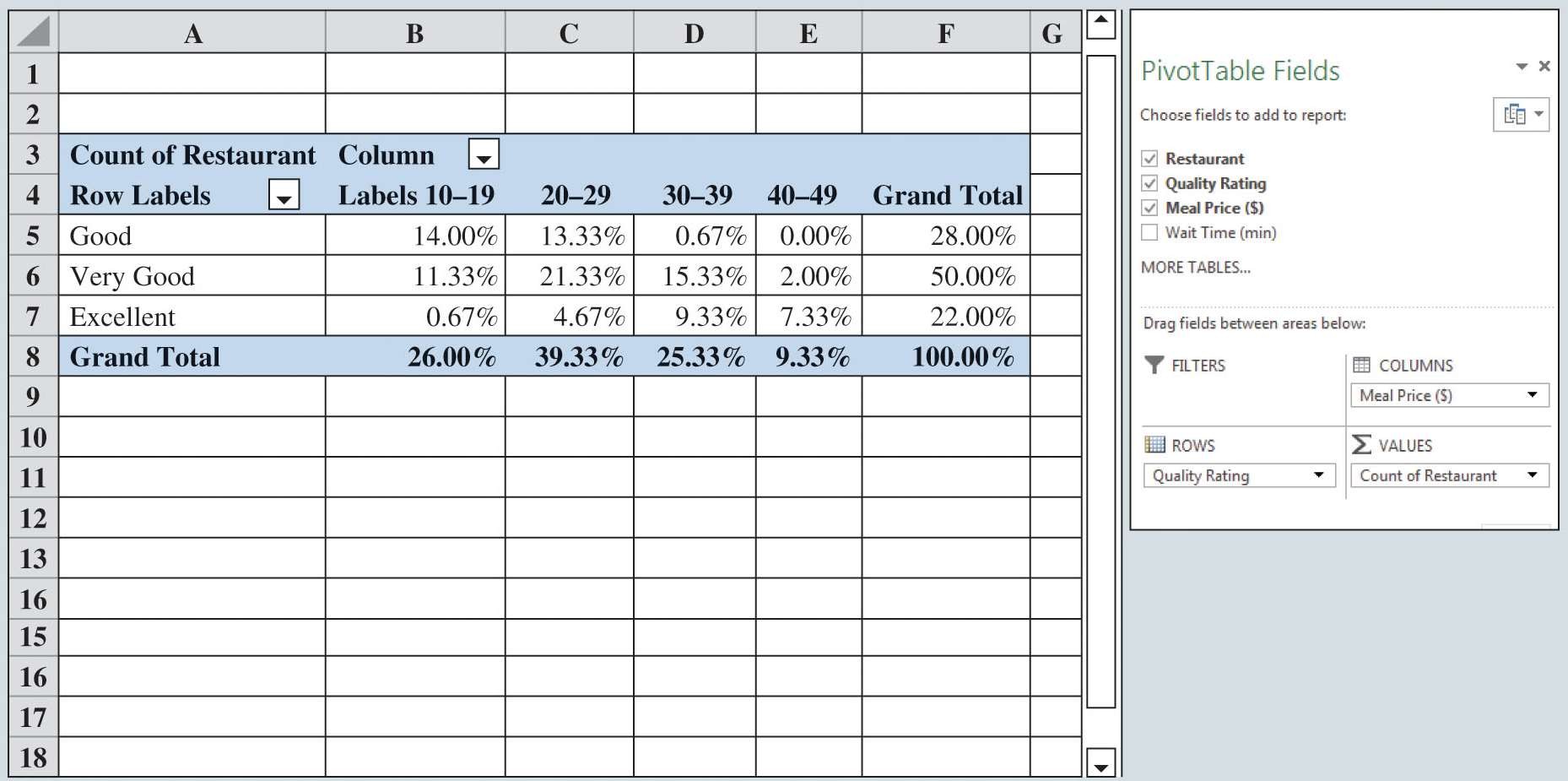
Tables

Figure 3.10: Completed PivotTable Field List and a Portion of the PivotTable Report for the Restaurant Data (Columns H:AK Are Hidden)



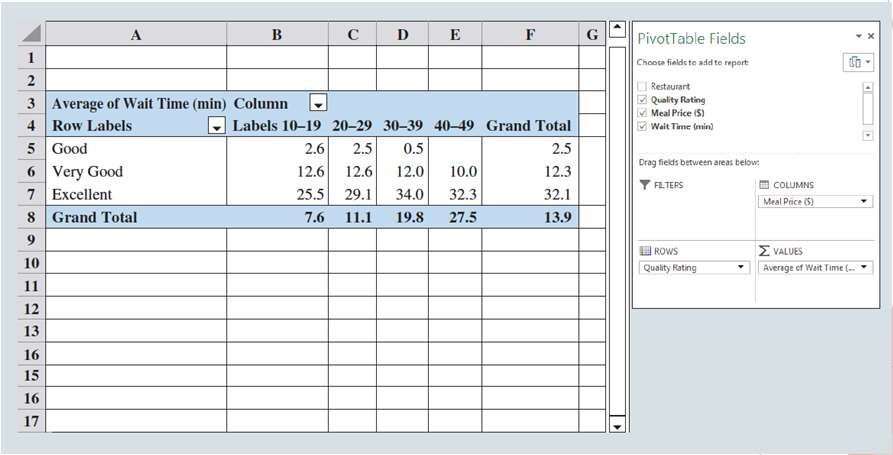
Tables

Figure 3.11: Final PivotTable Report for the Restaurant Data



Tables

Figure 3.12: Percent Frequency Distribution as a PivotTable for the Restaurant Data



Tables

Figure 3.13: PivotTable Report for the Restaurant Data with Average Wait Times Added