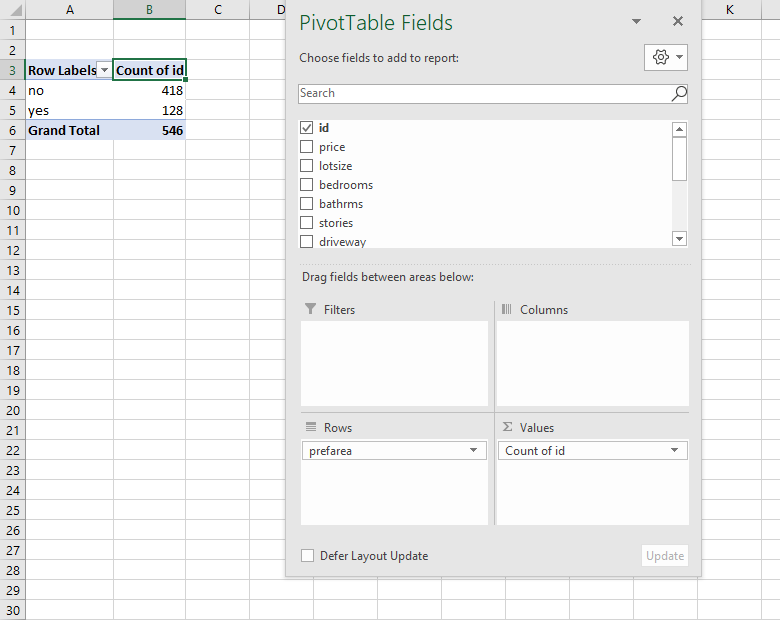
**EXPLORATORY DATA ANALYSIS IN EXCEL– DEMO NOTES**

**Frequencies**

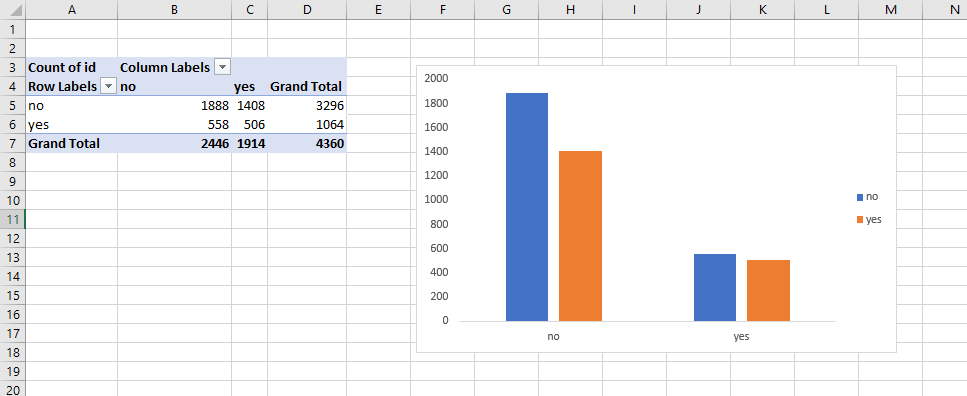
Create a PivotTable from the source data.

1. Make a frequency table by selecting categories of interest in the Rows/Columns field, then place a Count of the ID field in the Values section.
   1. To convert a field from a Sum to a Count, double-click on that variable header, and select Count in the “Summarize value field by” menu.



* 1. This is a *one-way* frequency table. We can make it a two-way table by placing another categorical variable along the Columns section.

1. Finally, we can visualize our categorical frequencies with a bar chart.
   1. We’ll remove some of the “chartjunk” by deleting gridlines and removing the PivotChart field buttons.

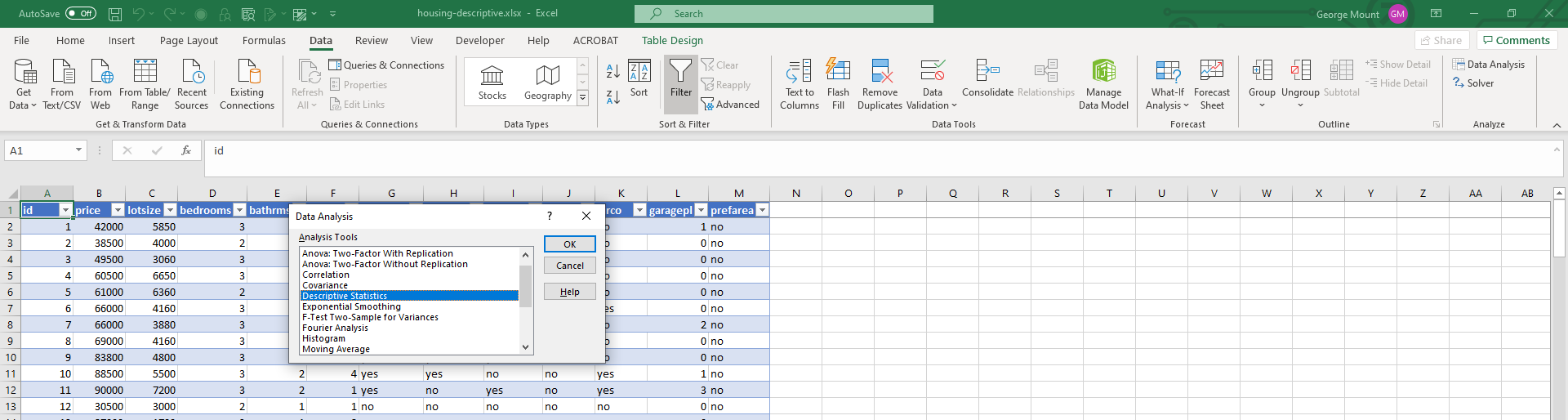


**Downloading the Analysis ToolPak**

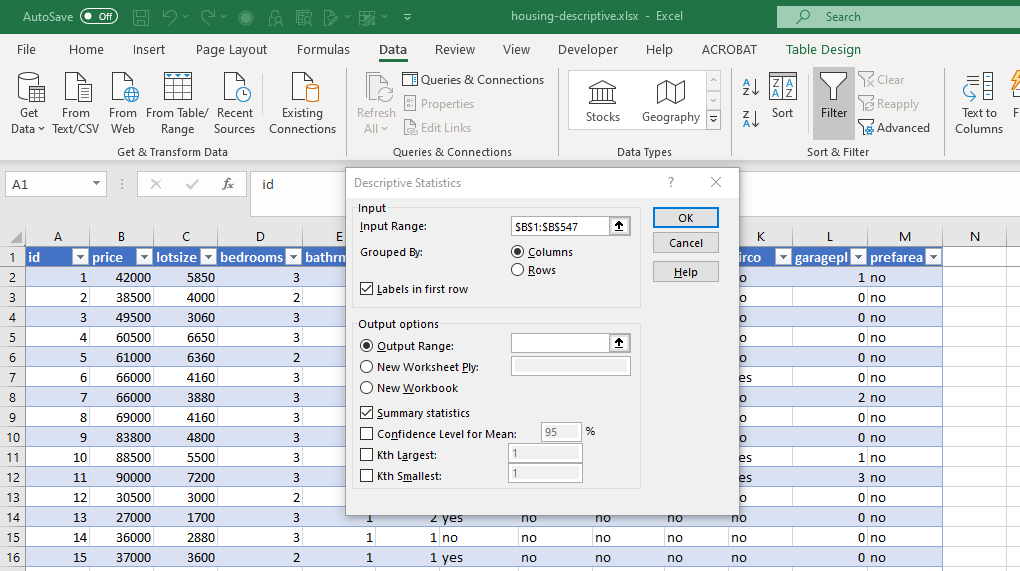
[See instructions from Microsoft here](https://support.office.com/en-us/article/Load-the-Analysis-ToolPak-in-Excel-6a63e598-cd6d-42e3-9317-6b40ba1a66b4). Note the process is different for Windows and Mac.

**Descriptive Statistics**

1. Go to the Data tab on the home ribbon.
2. Select Data Analysis from the Analyze group (far right of the menu).
3. Select Descriptive Statistics from the menu.

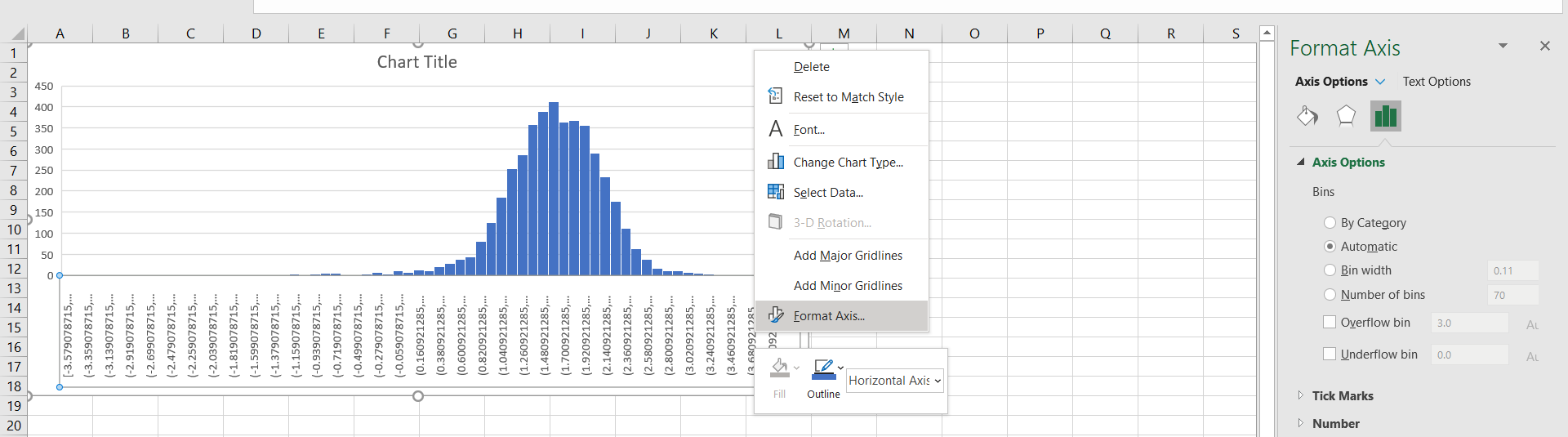


1. Select your Input Range. If your selection includes a header row, make sure to check on the “Labels in First Row” option.
2. By default, the output will be placed in a new worksheet. If you want it elsewhere, click inside “Output Range.” Make sure to double-click inside the dialog box before selecting a new range, otherwise the input range will be re-written.
3. Check on “Summary Statistics.”



**Histograms**

1. Select your input range and go to Insert -> Charts. Histogram should be your third option. Select that. You can cut and paste the resulting histogram elsewhere in the workbook.
2. To change the number of bins in the histogram, right-click on the X-axis and select Format Axis. You can then customize the X-axis on the side menu. *Note: these features are not available on Excel for Mac.*



Demo: stat-roulete.xlsx

This is to demonstrate the central limit theorem.

1. First, simulate 500 rounds of a roulette spin using RANDBETWEEN(0,36).
   1. Plot the resulting frequency distribution as a histogram.
   2. Hit F9 to recalculate the results. They will randomly recalculate each time.
   3. This is a *uniform distribution*. It’s equally likely to get any of the numbers between 0 and 36.

A screenshot of a cell phone

Description automatically generated

1. Now, simulate a roulette spin 100 times.
   1. Take the average spin of the 100.
   2. Do this 500 times.
   3. Plot the distribution of means and recalculate.
   4. We now get a “bell-shaped” normal distribution. This is due to the central limit theorem which states that the sampling distribution of the mean of any independent, random variable will be normal or nearly normal if the sample size is large enough.

A screenshot of a cell phone

Description automatically generated

See workbook

Demo: large-numbers.xlsx

This time we will simulate what the average roulette spin is with increasingly more spins. Go to the start worksheet to start from the beginning.

1. Column B will be randomly simulated with RANDBETWEEN(0,36). We will do that 500 times.
2. Column C will find the cumulative total. We will do this with a mixed reference sum: SUM($B$2:B2)
3. We will then take the cumulative average in Column D: C2/A2.
4. Plot a line chart of Column D to see the law of large numbers in effect. This states that the average trial will become closer to the expected value as more trials are performed.