Excel Data Analysis

https://www.depts.ttu.edu/itts/training/shortcourses/handouts.php

OR

Google "TTU Short Course", click the link and then "Handouts"

Benjamin Chamness

Accessibility & Data Analyst

ITTS Technology Support, IT Team Web

Material Notes

- Information here may apply to other spreadsheet software tools (LibreOffice, OpenOffice, Google Sheets, etc.).
  - These other tools are not approved for use with TTU data.
- All sample data here is either fully anonymized and random OR publicly available.
  - Any connection to real TTU data / events is not intentional.
  - I do not intend to make any statements with the sample data presented.
- I don’t know your data – so my examples may not be 100% accurate with your processing.
- I may touch on concepts in statistics, but this course isn’t meant to teach you statistics.

Outline

- Our Test Data
- Pivot Tables and Charts
- Analysis ToolPak

Our Test Data

Outline

- **Our Test Data**
- Pivot Tables and Charts
- Analysis ToolPak

Our Test Data – Scoop! There It Is!

- Scoop! There It Is! is an ice cream truck serving cool treats to people all over the city.
- We have 1 year of sales data, which includes:
  - How much we posted on social media
  - Where the truck was located
  - What the weather was like (high temperature and if it was raining)
  - How many people passed the truck
  - How many people made a purchase
  - The total sales from the day
- Not every day is tracked, because even ice cream trucks need a break!

Our Test Data – Scoop! There It Is!

	A	B	C	D	E	F	G	H	I	J
1	Date	Weekday	Location	Social Media Posts	High Temp	Rain	Foot Traffic	Customer Count	Total Sales	Customer Conversion
2	1/1/2022	Saturday	Park	4	52	1	71	31	$ 640.50	43.66%
3	1/2/2022	Sunday	Park	5	38	1	78	21	$ 433.59	26.92%
4	1/3/2022	Monday	Downtown	2	52	0	96	24	$ 211.27	25.00%
5	1/4/2022	Tuesday	Campus	2	68	0	231	41	$ 246.55	17.75%
6	1/7/2022	Friday	Campus	0	52	0	232	14	$ 92.42	6.03%
7	1/8/2022	Saturday	Mall	4	73	0	483	386	$ 5,667.06	79.92%
8	1/9/2022	Sunday	Park	0	54	0	305	59	$ 1,252.10	19.34%
9	1/10/2022	Monday	Campus	4	54	0	239	17	$ 90.86	7.11%
10	1/11/2022	Tuesday	Downtown	2	56	0	68	19	$ 182.68	27.94%
11	1/12/2022	Wednesday	Downtown	5	67	0	89	53	$ 504.13	59.55%
12	1/13/2022	Thursday	Campus	2	69	0	238	8	$ 59.10	3.36%
13	1/19/2022	Wednesday	Campus	3	46	0	246	26	$ 151.57	10.57%
14	1/20/2022	Thursday	Downtown	0	28	0	107	3	$ 31.59	2.80%
15	1/21/2022	Friday	Campus	1	47	0	241	17	$ 112.42	7.05%

About Our Data

Categories:

Things that describe our data. These can help “Group” our data for insights.

- Date
- Weekday
- Location
- Social Media Posts
- High Temp
- Rain
- Foot Traffic

Facts:

The values we care about for our business needs. These values are often summarized.

- Foot Traffic
- Customer Count
- Total Sales
- Customer Conversion (*careful when summarizing percentages)

Pivot Tables and Charts

Outline

- Our Test Data
- **Pivot Tables and Charts**
- Analysis ToolPak

Pivot Tables and Charts

- Pivot Tables are an easy way to build reports on your data in Excel
- Allow you to group your data by category (date, location, temperature, college, major, residence hall, etc)
- Allows you to summarize the data by some calculation (total sales, average traffic, count of students, etc.)

3	Sum of Total Sales	Column Labels						
4	Row Labels	Brewery	Campus	Downtown	Mall	Park	(blank)	Grand Total
5	<1/1/2022							
6	Jan	$ 862.48	$ 1,169.73	$ 1,220.11	$ 8,570.52	$ 6,708.93		$ 18,531.77
7	Feb	$ 344.85	$ 941.85	$ 1,289.65	$ 7,673.75	$ 3,656.32		$ 13,906.43
8	Mar	$ 1,449.68	$ 1,727.61	$ 3,489.60	$ 17,263.23	$ 3,294.58		$ 27,224.71
9	Apr	$ 1,693.89	$ 1,115.22	$ 3,893.03	$ 9,431.57	$ 4,271.19		$ 20,404.90
10	May	$ 2,065.53	$ 2,251.86	$ 8,971.35	$ 2,589.01	$ 37,382.39		$ 53,260.14
11	Jun	$ 769.21	$ 3,577.08	$ 8,742.32	$ 25,234.73	$ 16,289.49		$ 54,612.83

What reports may be relevant with our data?

- Total sales and customers by date and location
- Average sales and customers by day of week and location
- Average sales by temperature

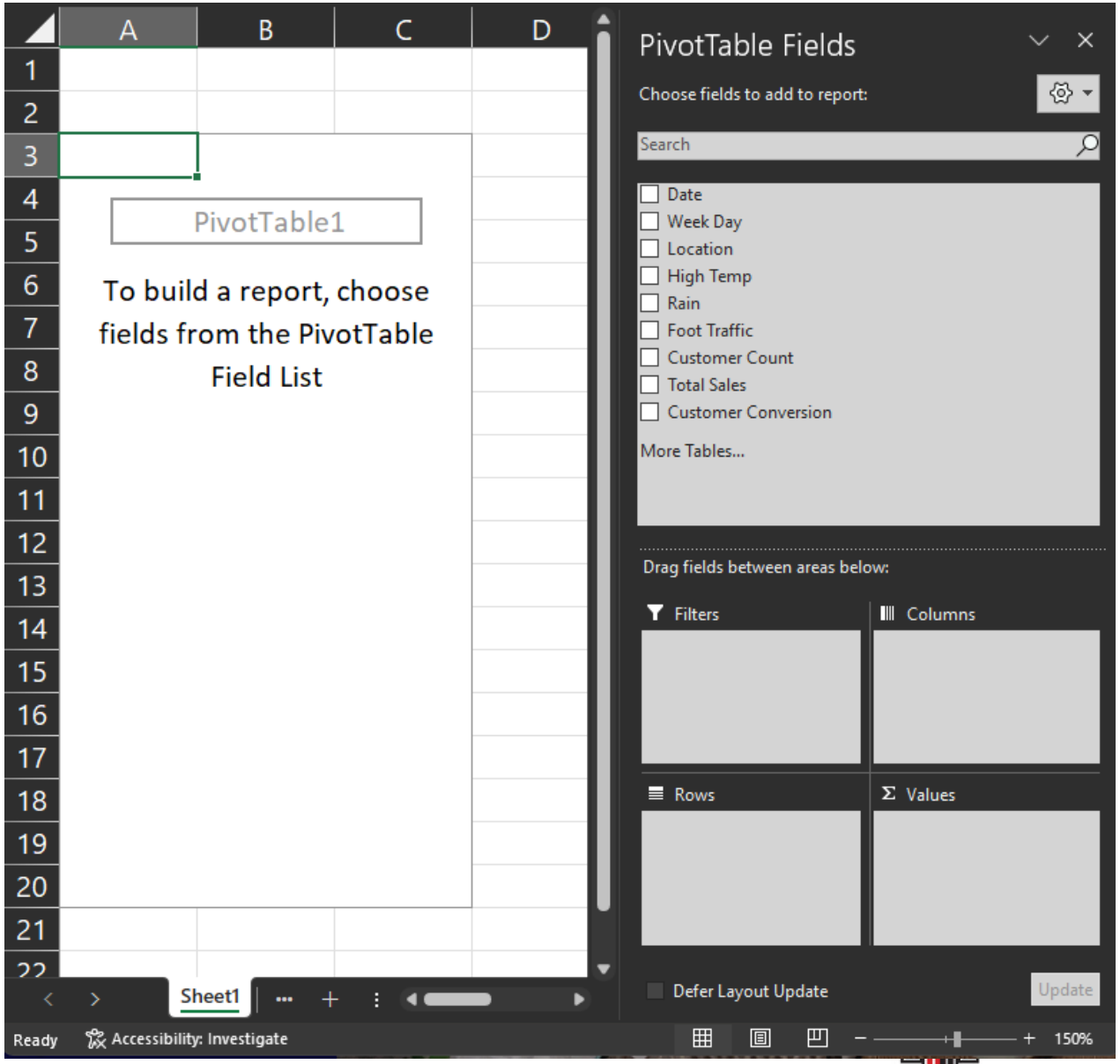
	A	B	C	D	E	F	G	H	I	J
1	Date	Weekday	Location	Social Media Posts	High Temp	Rain	Foot Traffic	Customer Count	Total Sales	Customer Conversion
2	1/1/2022	Saturday	Park	4	52	1	71	31	$ 640.50	43.66%
3	1/2/2022	Sunday	Park	5	38	1	78	21	$ 433.59	26.92%
4	1/3/2022	Monday	Downtown	2	52	0	96	24	$ 211.27	25.00%
5	1/4/2022	Tuesday	Campus	2	68	0	231	41	$ 246.55	17.75%
6	1/7/2022	Friday	Campus	0	52	0	232	14	$ 92.42	6.03%
7	1/8/2022	Saturday	Mall	4	73	0	483	386	$ 5,667.06	79.92%
8	1/9/2022	Sunday	Park	0	54	0	305	59	$ 1,252.10	19.34%

Pivot Table Creation

1. Select all your data (ideally by selecting the column letters)
2. Go to the “Insert” menu and select “PivotTable”
3. Confirm that the “Table/Range” field is the correct value (for our test data, it should be “SampleData!$A:$J”)
4. State where you want the PivotTable to be created. The default of “New Worksheet” is fine for our purposes, but you could also tell Excel to crate the PivotTable in a specific cell.

PivotTable Parts

- Field List shows all of your available data fields.
- Filters will filter your data by a category.
- Rows and Columns break data out by categories.
- Values will hold your data summaries.



About Our Data

- Categories:
  - Date
  - Weekday
  - Location
  - Social Media Posts
  - High Temp
  - Rain
  - Foot Traffic

Use these in “Rows”, “Columns”, and “Filters”

- Facts:
  - Foot Traffic
  - Customer Count
  - Total Sales
  - Customer Conversion (*careful when summarizing percentages)

Use these in “Values”

Rows and Columns

- Will identify distinct category values in your dataset
- Will separate your data points by the categories
- Will NOT fill in any gaps (days of the week not represented)

3	Sum of Total Sales	Column Labels					
4	Row Labels	Brewery	Campus	Downtown	Mall	Park	Grand Total
5	Sunday	$ 6,146.99			$ 90,325.86	$ 68,105.85	$164,578.71
6	Monday		$ 5,891.53	$11,198.12			$ 17,089.65
7	Tuesday		$ 6,716.47	$11,539.05			$ 18,255.53
8	Wednesday		$ 3,444.39	$13,527.73			$ 16,972.12
9	Thursday		$ 4,668.35	$10,670.88			$ 15,339.22
10	Friday		$ 5,034.17	$12,900.49			$ 17,934.66
11	Saturday	$ 6,411.13			$ 68,586.30	$103,936.05	$178,933.47
12	**Grand Total**	**$ 12,558.12**	**$25,754.91**	**$59,836.27**	**$158,912.16**	**$172,041.90**	**$429,103.36**

Values

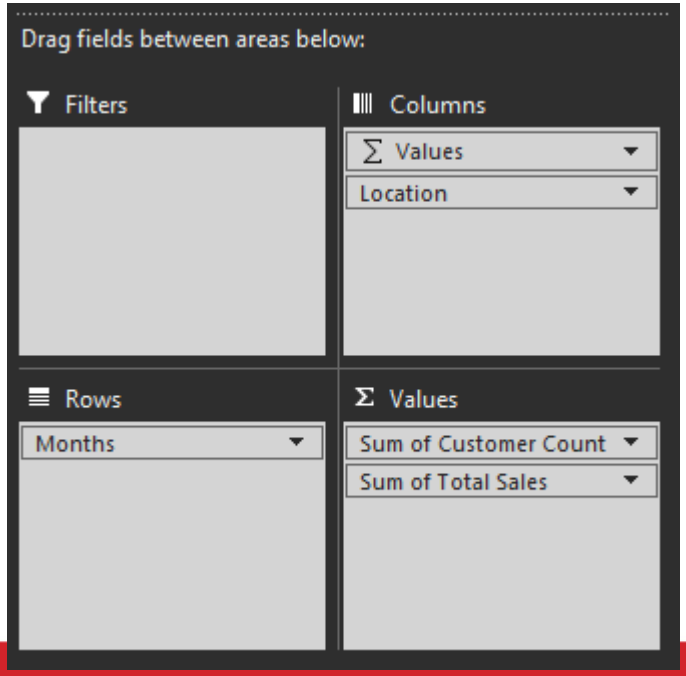
- Perform some summary (count, average, sum, etc) on a "fact" in your dataset
- Will NOT fill in any gaps (data intersections not represented in your data)
- Do not summarize on percentage data

3	Sum of Total Sales	Column Labels					
4	Row Labels	Brewery	Campus	Downtown	Mall	Park	Grand Total
5	Sunday	$ 6,146.99			$ 90,325.86	$ 68,105.85	$164,578.71
6	Monday		$ 5,891.53	$11,198.12			$ 17,089.65
7	Tuesday		$ 6,716.47	$11,539.05			$ 18,255.53
8	Wednesday		$ 3,444.39	$13,527.73			$ 16,972.12
9	Thursday		$ 4,668.35	$10,670.88			$ 15,339.22
10	Friday		$ 5,034.17	$12,900.49			$ 17,934.66
11	Saturday	$ 6,411.13			$ 68,586.30	$103,936.05	$178,933.47
12	**Grand Total**	**$ 12,558.12**	**$25,754.91**	**$59,836.27**	**$158,912.16**	**$172,041.90**	**$429,103.36**

Report 1: Total sales and customers by time and location

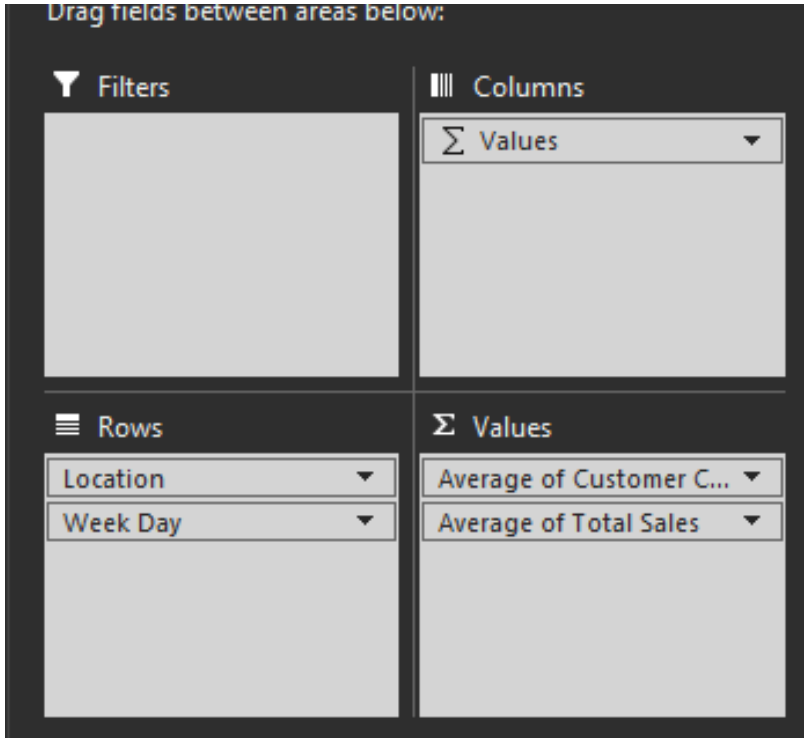
3		Column Labels											
4		Sum of Customer Count					Sum of Total Sales					Total Sum of Customer Count	Total Sum of Total Sales
5	Row Labels	Brewery	Campus	Downtown	Mall	Park	Brewery	Campus	Downtown	Mall	Park		
6	Jan	78	209	128	554	302	$ 862.48	$ 1,169.73	$ 1,220.11	$ 8,570.52	$ 6,708.93	1271	$ 18,531.77
7	Feb	34	179	129	494	188	$ 344.85	$ 941.85	$ 1,289.65	$ 7,673.75	$ 3,656.32	1024	$ 13,906.43
8	Mar	130	361	367	1023	171	$ 1,449.68	$ 1,727.61	$ 3,489.60	$ 17,263.23	$ 3,294.58	2052	$ 27,224.71
9	Apr	149	247	376	652	208	$ 1,693.89	$ 1,115.22	$ 3,893.03	$ 9,431.57	$ 4,271.19	1632	$ 20,404.90
10	May	135	460	780	193	1308	$ 2,065.53	$ 2,251.86	$ 8,971.35	$ 2,589.01	$ 37,382.39	2876	$ 53,260.14
11	Jun	49	627	717	1296	791	$ 769.21	$ 3,577.08	$ 8,742.32	$ 25,234.73	$ 16,289.49	3480	$ 54,612.83

- Drag “Date” to rows, which automatically summarizes by month
- Drag Location to columns
- Drag Sales and Customers to values
- Change Sales to report in dollar values



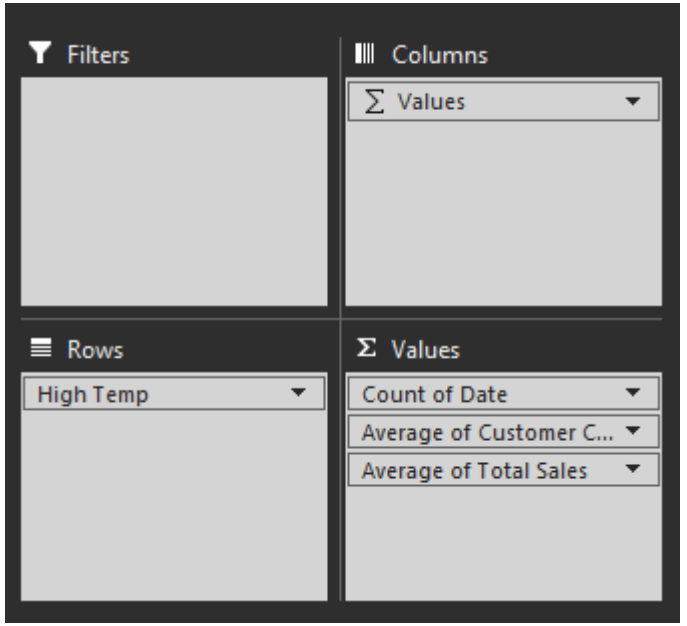
Report 2: Average sales and customers by day of week and location

3	Row Labels	Average of Customer Count	Average of Total Sales
4	**Brewery**	**37.15384615**	**$ 483.00**
5	Sunday	36.85714286	$ 439.07
6	Saturday	37.5	$ 534.26
7	**Campus**	**46.14678899**	**$ 236.28**
8	Monday	46.30769231	$ 226.60
9	Tuesday	59.76190476	$ 319.83
10	Wednesday	35.71428571	$ 164.02
11	Thursday	43.4	$ 233.42
12	Friday	45.38095238	$ 239.72
13	**Downtown**	**49.05504587**	**$ 548.96**
14	Monday	48.71428571	$ 533.24
15	Tuesday	42.52	$ 461.56



Report 3: Average sales by temperature

Row Labels	Count of Date	Average of Customer Count	Average of Total Sales
(blank)			
16-25	3	4.333333333	$ 28.19
26-35	4	8	$ 75.27
36-45	12	33.33333333	$ 435.35
46-55	38	46.76315789	$ 644.31
56-65	30	40.36666667	$ 586.91
66-75	46	61.63043478	$ 707.92
76-85	46	110.2826087	$ 1,558.98
86-95	75	99.74666667	$ 1,501.59
96-106	59	147.5254237	$ 2,788.23
Grand Total	**313**	**87.94249201**	**$ 1,370.94**



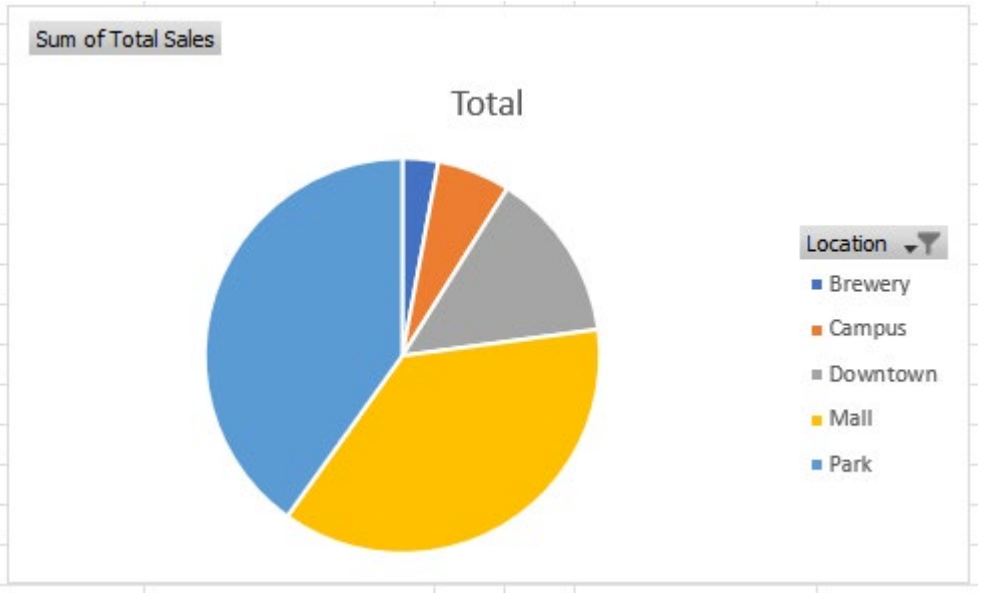
Charts

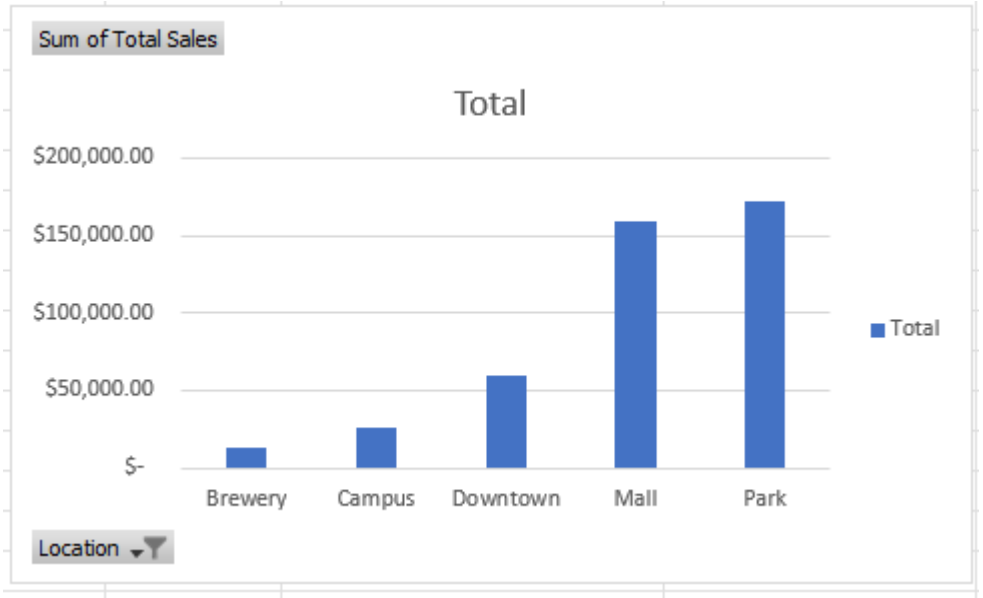
- Bar charts show counts or totals grouped by a category
- Pie charts show how distinct categories make up parts of a whole
- Line charts show changes in value over time

- As with other charts in Excel, you have all of the same tools to adjust labels, titles, colors, and other chart elements

Bar and Pie Chart Examples

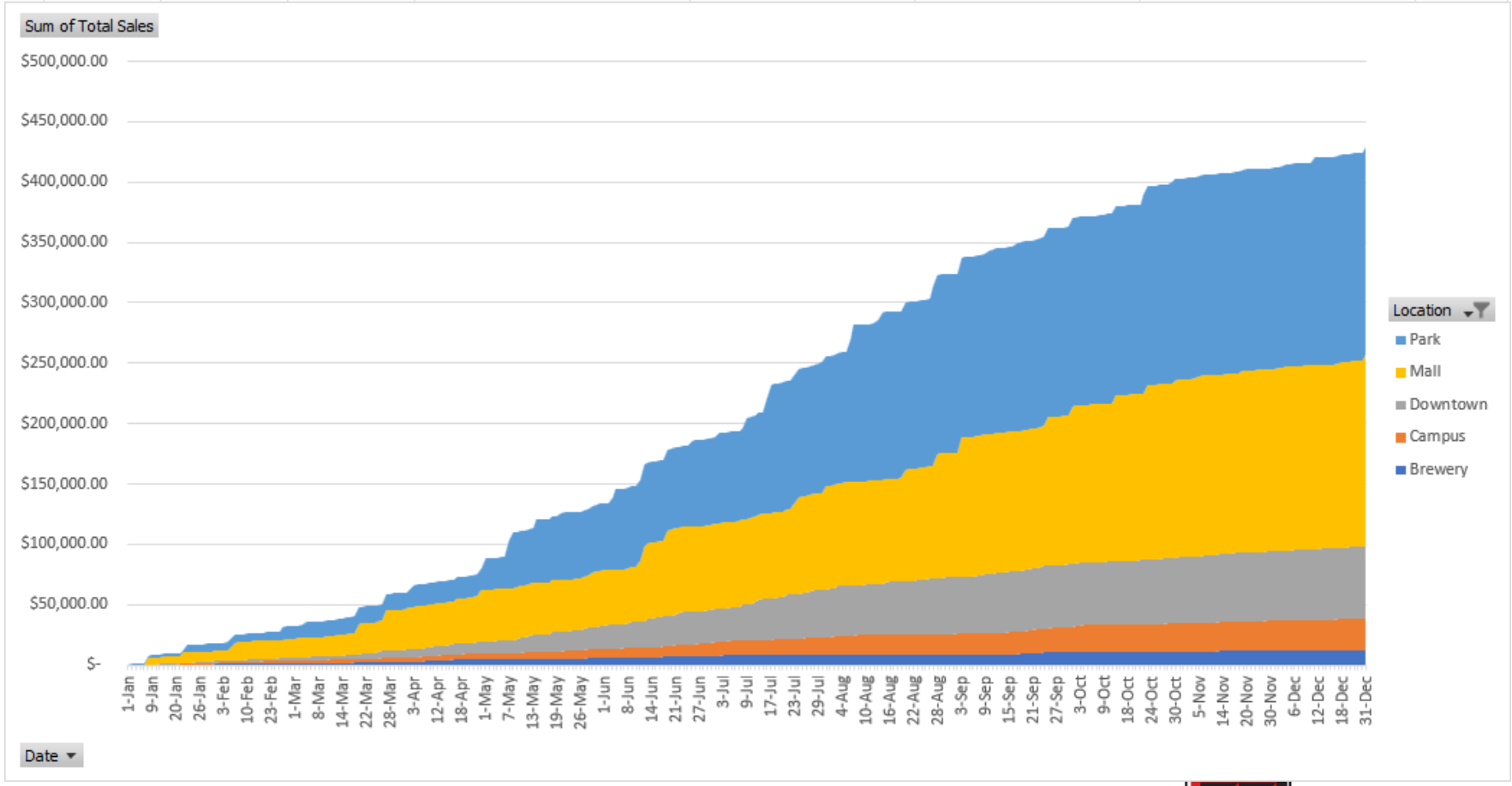
3	Row Labels	Sum of Total Sales	
4	Brewery	$	12,558.12
5	Campus	$	25,754.91
6	Downtown	$	59,836.27
7	Mall	$	158,912.16
8	Park	$	172,041.90
9	**Grand Total**	**$**	**429,103.36**





Stacked Area Chart Example

- Set "Sum of Total Sales" to be a "Running Total In > Date"
- Notice date gap: Jan 9 - 20
  - Solution: Build a list of dates and use GetPivotData to fetch values



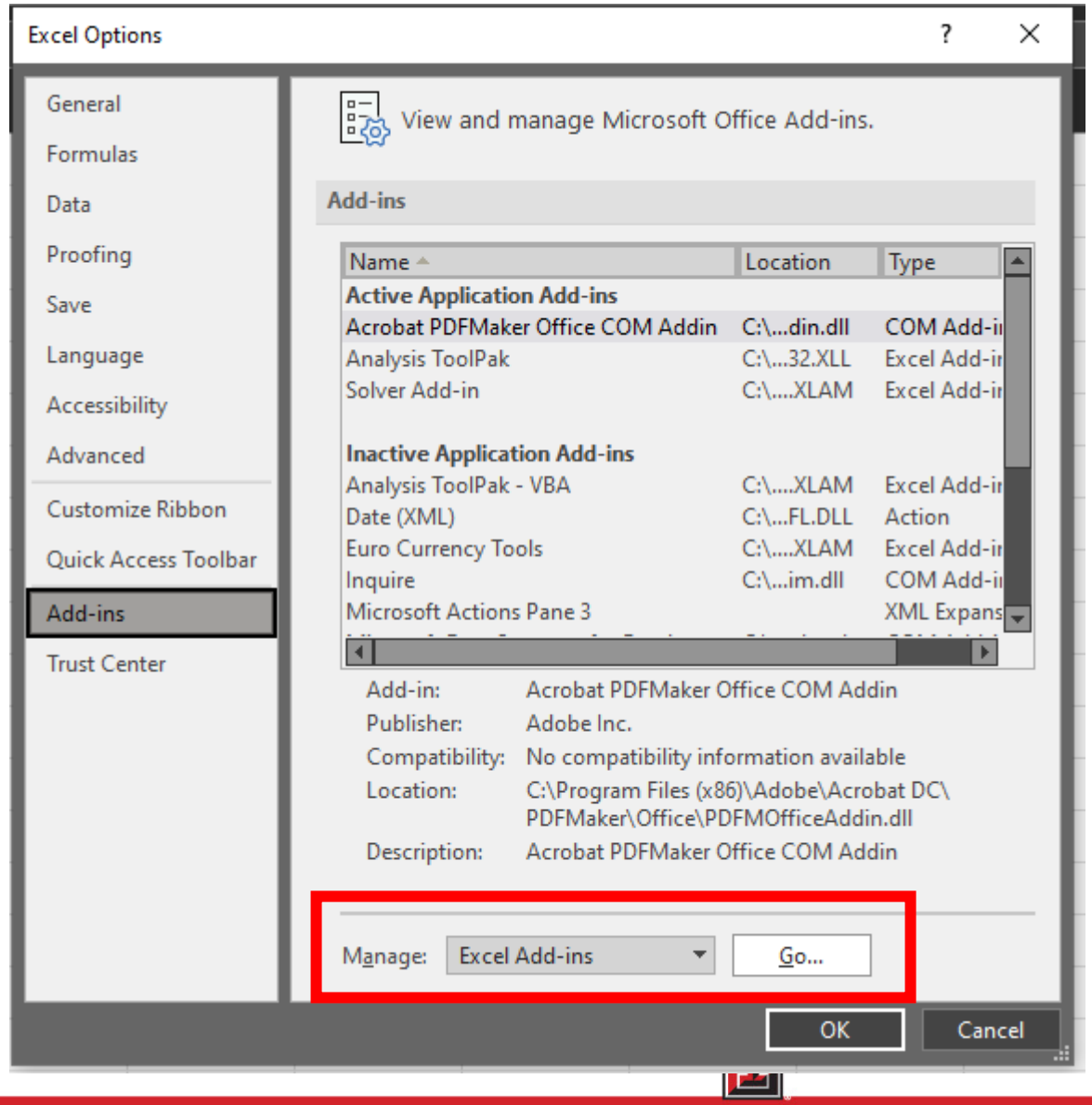
Analysis ToolPak

Outline

- Our Test Data
- Pivot Tables and Charts
- **Analysis ToolPak**

To Enable:

1. Click on "File" in the top left and then "Options" towards the bottom.
2. In the "Excel Options" window that opens, click on "Add-ins" on the left, and then click the button that says "Go…"
3. Check the box next to "Analysis ToolPak" and then click "OK".
4. When enabled, it should be in the "Data" ribbon menu.



Analysis Limitations: Text

- String data is hard to analyze
- Binary strings (yes/no, true/false) can be mapped to 0 and 1
- Categories with some defined order can sometimes be mapped to integers
  - Freshman, Sophomore, Junior, Senior
  - Strongly Agree, Agree, Disagree, Strongly Disagree
- Our “Rain” column is a binary yes/no, so it is coded as 1s and 0s
- Weekday and Location can’t be mapped to numbers as easily

Correlation

- Helps describe a relationship between variables in your dataset.
- A positive correlation means two variables will change in the same direction (increase/decrease together).
- A negative correlation means two variables will change in opposite directions (increase in one, decrease in the other).
- Values close to +/-1 show a strong correlation, values close to 0 show weak / no correlation

- Correlation is not causation: the relationship does not mean that one change causes the other.

Correlation

1. Click the "Data Analysis" tool in the "Data" menu.
2. Select "Correlation"
3. Select your data for the "Input Range" (entering in column letters), check the "Labels in First Row" box, and click "OK"

	Social Media Posts	High Temp	Rain	Foot Traffic	Customer Count	Total Sales
Social Media Posts	1.00					
High Temp	-0.07	1.00				
Rain	0.01	-0.10	1.00			
Foot Traffic	0.02	0.07	**-0.40**	1.00		
Customer Count	**0.25**	**0.35**	-0.25	**0.73**	1.00	
Total Sales	0.24	0.29	-0.17	**0.60**	0.91	1.00

- Green cells are our high(er) correlated values.
- Blue cells may be moderately correlated.

Regression

- Attempts to create a function that can predict a given value.
- In our case, given known values for temperature, rain, traffic, etc – how can we predict our daily sales?
- Values:
  - R Square: How well the model fits the given data, ideally above 0.95.
  - Coefficient: This value will be multiplied by your variable to produce the predicted output.
  - P-Value: A p-value less than 0.05 means the variable is a likely predictor of your output.

Regression

1. Click the "Data Analysis" tool in the "Data" menu.
2. Select "Regression"
3. Select the value you want to predict as the "Input Y Range"
4. Select the values that will influence this output as the "Input X Range"
5. Check the "Labels" box and the "Confidence Interval" box, then "OK"

Note: In items 3 and 4, you will need to specifically select the cells with data (H1:H314), not just the full column (H:H)

Regression (R Square: 0.83)

	Coefficients	P-value	*Note*
Intercept	335.88	0.32	Ignore: High P-Value
Social Media Posts	-23.07	0.55	Ignore: High P-Value
High Temp	-7.08	0.04	
Rain	152.38	0.40	Ignore: High P-Value
Foot Traffic	-2.55	0.00	
Customer Count	23.65	6.09E-75	

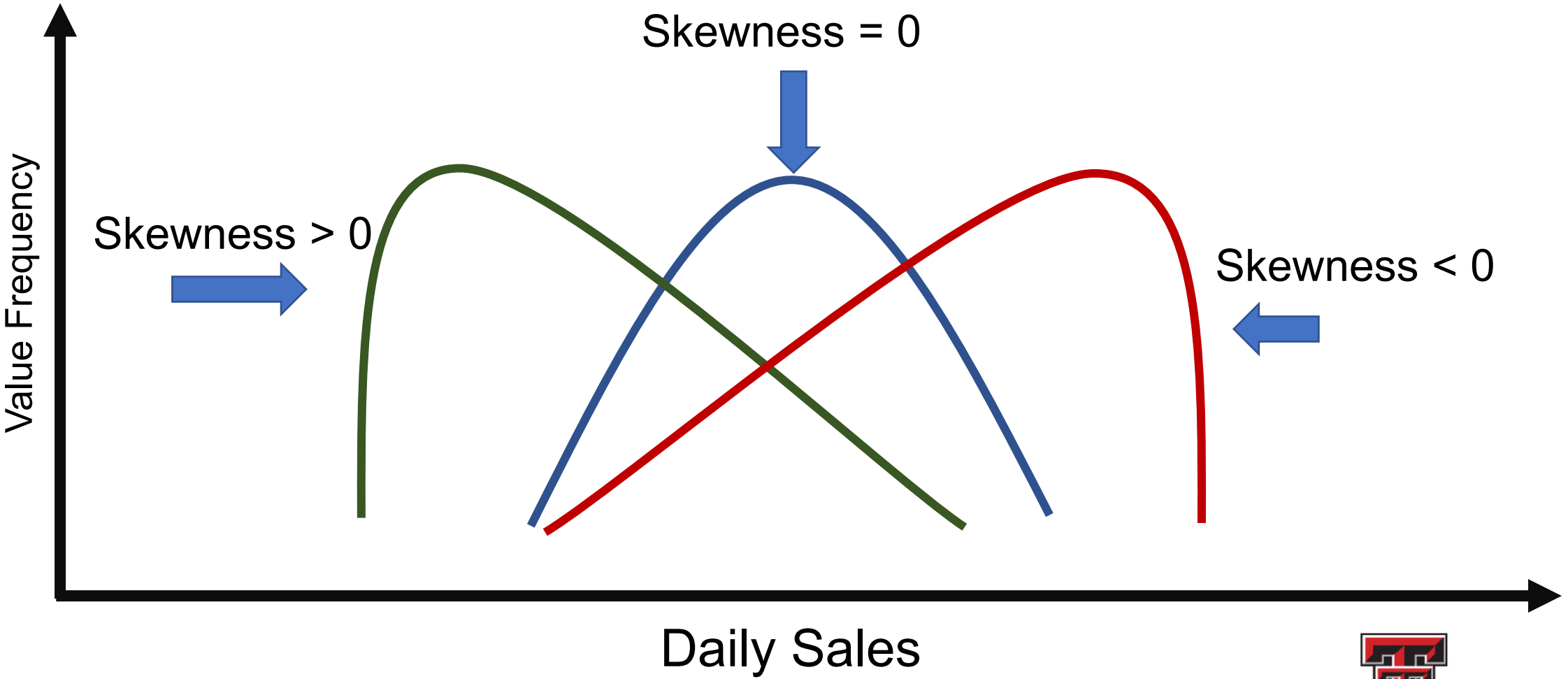
- Theoretically, we can predict sales with this formula:

  dailySales = $-7.08 * highTemp + $152.38 * rain + $-7.08 * footTraffic + $23.65 * customerCount
- This fails to consider the text category of "Location", which was a large contributor of sales.

Descriptive Statistics

- Allows you to see some various values that describe a given variable.
- Statistics:
  - Mean – average
  - Median – "middle" number when sorted low to high
  - Standard Deviation – how much the value can vary
  - Skewness – is the distribution of your data shifted to one side

Skewness


Skewness = 0
Skewness > 0
Skewness < 0
Value Frequency
Daily Sales

Descriptive Statistics

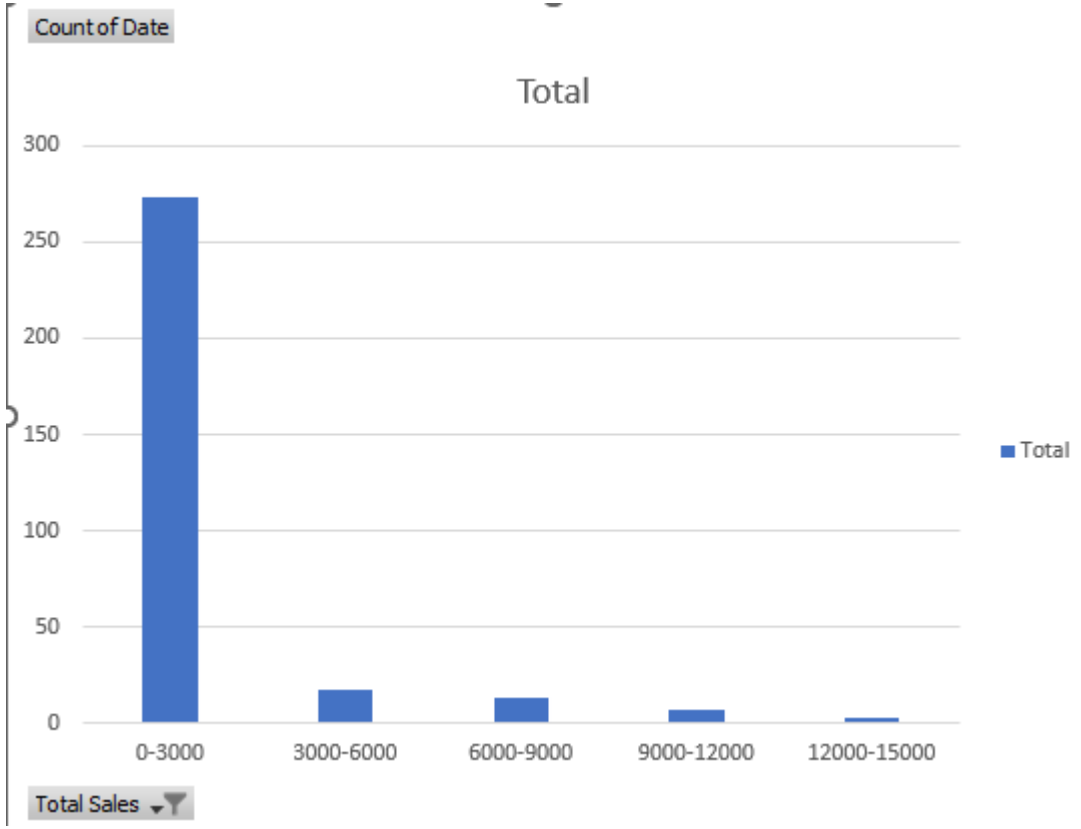
1. Click the "Data Analysis" tool in the "Data" menu.
2. Select "Descriptive Statistics"
3. Select your data for the "Input Range" (entering in the column letters)
4. Check the "Labels in First Row" box and the "Summary Statistics" box, then "OK"

Descriptive Statistics

Total Sales	
Mean	1370.937
Standard Error	139.6033
Median	399.0955
Mode	
Standard Deviation	2469.834
Sample Variance	6100081
Kurtosis	7.653245
Skewness	2.805633
Range	12895.57
Minimum	0
Maximum	12895.57
Sum	429103.4
Count	313

- Average sales is $1371, but median is $399.
  - Lots of small sale days, a few REALLY large sale days
- Standard Deviation is $2469, so the daily sale value varies quite a bit.
- Skewness is fairly large

Actual Data Distribution



- Heavily skewed towards smaller daily sales
- Positive Skewness

Questions? Answers? Favorite PivotTable Uses?