

Prepared by Dr. Wesam Ahmed

IT393
Business Intelligence

Lecture 5



Practical guide to using Excel for Data Analytics and visualization

Lecture 5

IT393

Business Intelligence

Prepared by
Dr. Wesam Ahmed

01

Learning Objectives:

- 1.1 Using Excel for data analytics
- 1.2 Understand the importance of data/ information visualization
- 1.3 Learn different types of visualization techniques
- 1.4 Using Excel for data visualization

02

Using Excel for data analytics:

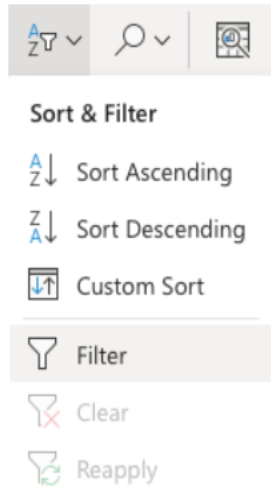
ID	Marital Status	Gender	Income	Children	Education	Occupation	Home Owner	Cars	Commute Dist	Region	Age	Purchased Bike
12496	M	F	\$40,000.00	1	Bachelors	Skilled Manual	Yes		0 0-1 Miles	Europe	42	No
24107	M	M	\$30,000.00	3	Partial College	Clerical	Yes		1 0-1 Miles	Europe	43	No
14177	M	M	\$80,000.00	5	Partial College	Professional	No		2 2-5 Miles	Europe	60	No
24381	S	M	\$70,000.00	0	Bachelors	Professional	Yes		1 5-10 Miles	Pacific	41	Yes
25597	S	M	\$30,000.00	0	Bachelors	Clerical	No		0 0-1 Miles	Europe	36	Yes
13507	M	F	\$10,000.00	2	Partial College	Manual	Yes		0 1-2 Miles	Europe	50	No
27974	S	M	\$160,000.00	2	High School	Management	Yes		4 0-1 Miles	Pacific	33	Yes
19364	M	M	\$40,000.00	1	Bachelors	Skilled Manual	Yes		0 0-1 Miles	Europe	43	Yes
22155	M	M	\$20,000.00	2	Partial High Sc	Clerical	Yes		2 5-10 Miles	Pacific	58	No
19280	M	M	\$120,000.00	2	Partial College	Manual	Yes		1 0-1 Miles	Europe	40	Yes
22173	M	F	\$30,000.00	3	High School	Skilled Manual	No		2 1-2 Miles	Pacific	54	Yes
12697	S	F	\$90,000.00	0	Bachelors	Professional	No		4 10+ Miles	Pacific	36	No
11434	M	M	\$170,000.00	5	Partial College	Professional	Yes		0 0-1 Miles	Europe	55	No
25323	M	M	\$40,000.00	2	Partial College	Clerical	Yes		1 1-2 Miles	Europe	35	Yes
23542	S	M	\$60,000.00	1	Partial College	Skilled Manual	No		1 0-1 Miles	Pacific	45	Yes
20870	S	F	\$10,000.00	2	High School	Manual	Yes		1 0-1 Miles	Europe	38	Yes
23316	S	M	\$30,000.00	3	Partial College	Clerical	No		2 1-2 Miles	Pacific	59	Yes
12610	M	F	\$30,000.00	1	Bachelors	Clerical	Yes		0 0-1 Miles	Europe	47	No
27183	S	M	\$40,000.00	2	Partial College	Clerical	Yes		1 1-2 Miles	Europe	35	Yes
25940	S	M	\$20,000.00	2	Partial High Sc	Clerical	Yes		2 5-10 Miles	Pacific	55	Yes
25598	M	F	\$40,000.00	0	Graduate Degr	Clerical	Yes		0 0-1 Miles	Europe	36	Yes
21564	S	F	\$80,000.00	0	Bachelors	Professional	Yes		4 10+ Miles	Pacific	35	No
19193	S	M	\$40,000.00	2	Partial College	Clerical	Yes		0 1-2 Miles	Europe	35	Yes
26412	M	F	\$80,000.00	5	High School	Management	No		3 5-10 Miles	Europe	56	No
27184	S	M	\$40,000.00	2	Partial College	Clerical	No		1 0-1 Miles	Europe	34	No

Figure 5.1 Bike data

03

Using Excel for data analytics:

The menu is accessed in the default Ribbon view or in the **Data** section in the navigation bar.



Filters can be applied to sort and hide data. It makes data analysis easier.

04

Using Excel for data analytics:

Remove duplicates

Duplicate entries can distort analysis and insights, so removing them ensures data accuracy.

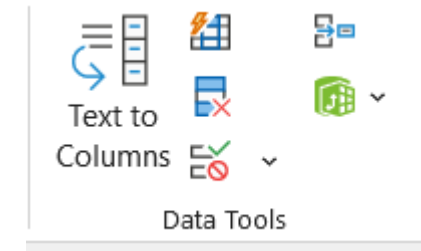
Select the worksheet or range of cells from which you want to remove duplicates.

Go to the **Data** tab on the **Ribbon**.

Click on **Remove Duplicates** in the **Data Tools** group.

In the **Remove Duplicates** dialog box, specify which column to check for duplicates.

Click **OK**. Excel will display a message showing the number of duplicate values removed.



05

Using Excel for data analytics:

Press Ctrl+H or go to Home > Find & Select > Replace

Find and Replace

Find Replace

Find what: M

Replace with: Male

Within: Sheet

Search: By Columns

Look in: Formulas

Match case

Match entire cell contents

Options <<

Replace All Replace Find All Find Next Close

06

Using Excel for data analytics:

We use a simple IF() statement to check one condition. But when you have to check more than one condition, you can combine several IF() statements — this creates a nested IF(). This will allow you to run several logical tests one after another and get different results for each test.

```
=IF(condition1, value_if_true, IF(condition2, value_if_true, value_if_false))
```

07

Using Excel for data analytics:

L	M
Age	Age Brackets
42	Middleage
43	Middleage
60	Old
41	Middleage
36	Middleage
50	Middleage
33	Middleage
43	Middleage
58	Old
40	Middleage
54	Middleage
36	Middleage
55	Old
35	Middleage
45	Middleage
38	Middleage
59	Old
47	Middleage
35	Middleage
55	Old
36	Middleage
35	Middleage
35	Middleage
56	Old
34	Middleage
63	Old
29	adolescent

```
=IF(L2<31, "adolescent", IF(L2<=54, "Middleage", "Old"))
```

08

Using Excel for data analytics:

Example 1: Data categorization with nested IF()

Say I want to assign a batch to each employee according to their years of service in the company.

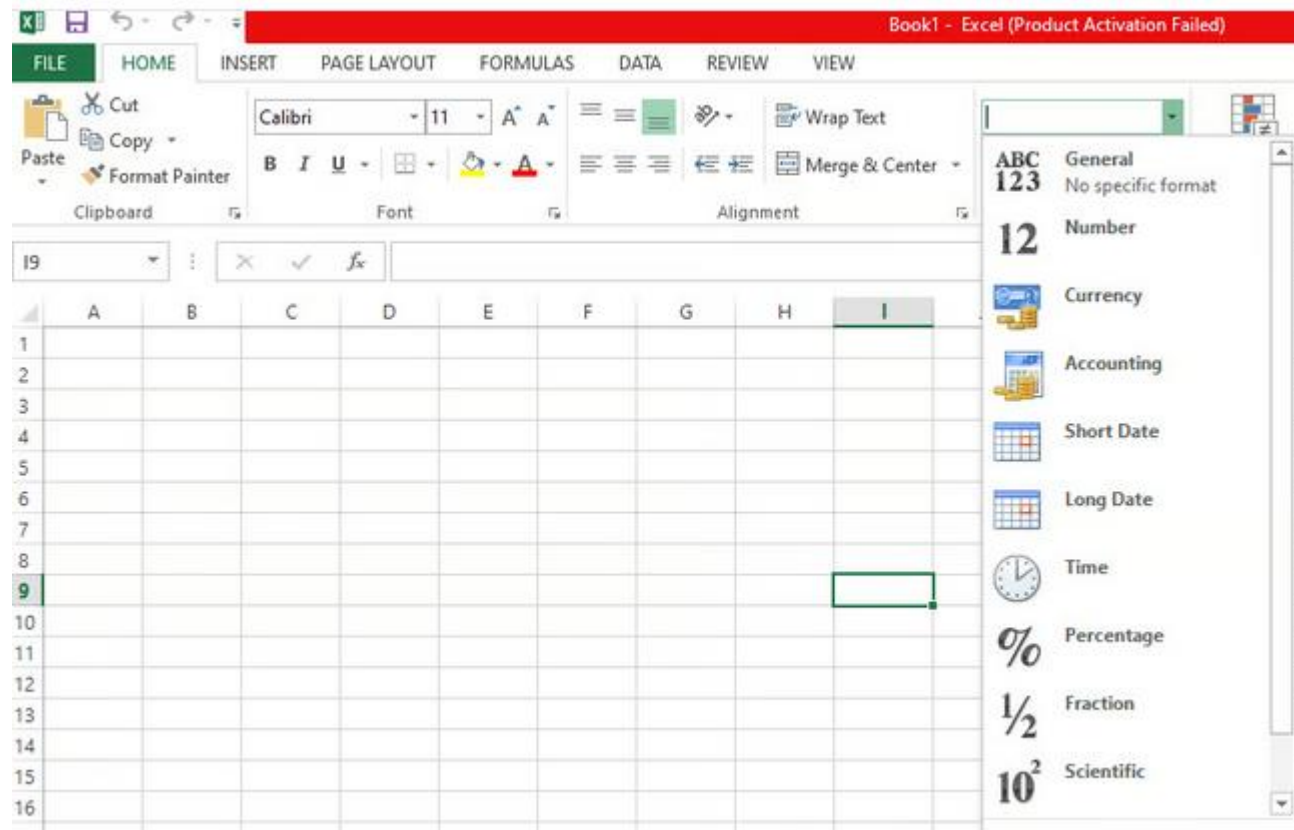
Years of service	Batch
≥ 10 years	Platinum
5–9 years	Gold
< 5 years	Silver

Here's the formula I'd use:

```
=IF(C2>=10, "Platinum", IF(C2>=5, "Gold", "Silver"))
```



Using Excel for data analytics:



You can go to the **Number** group from the **Home** tab and see the data types.

10

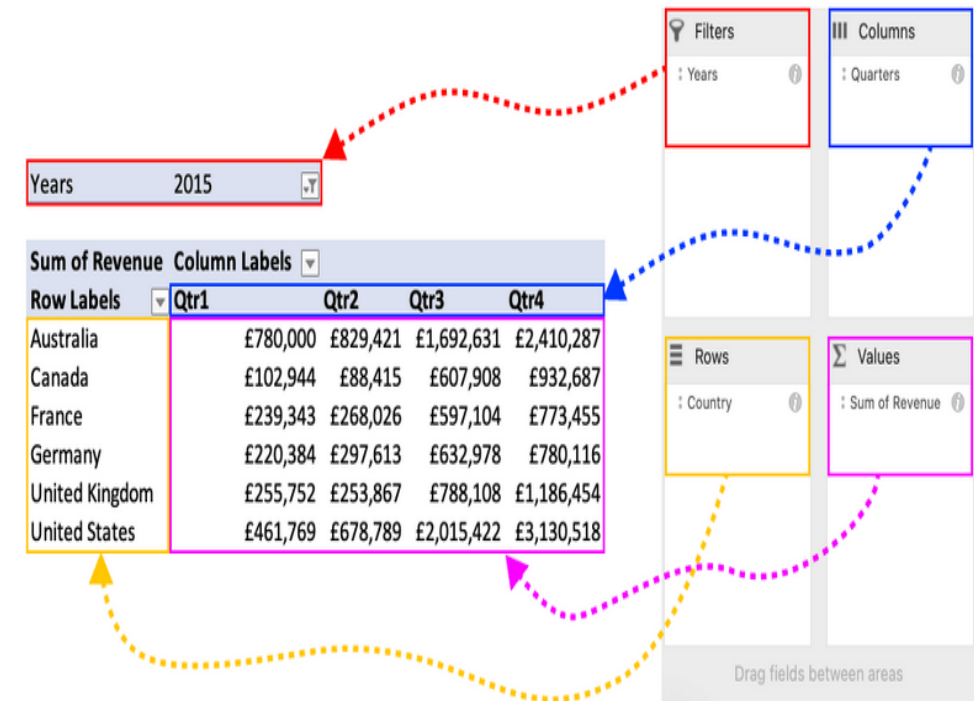
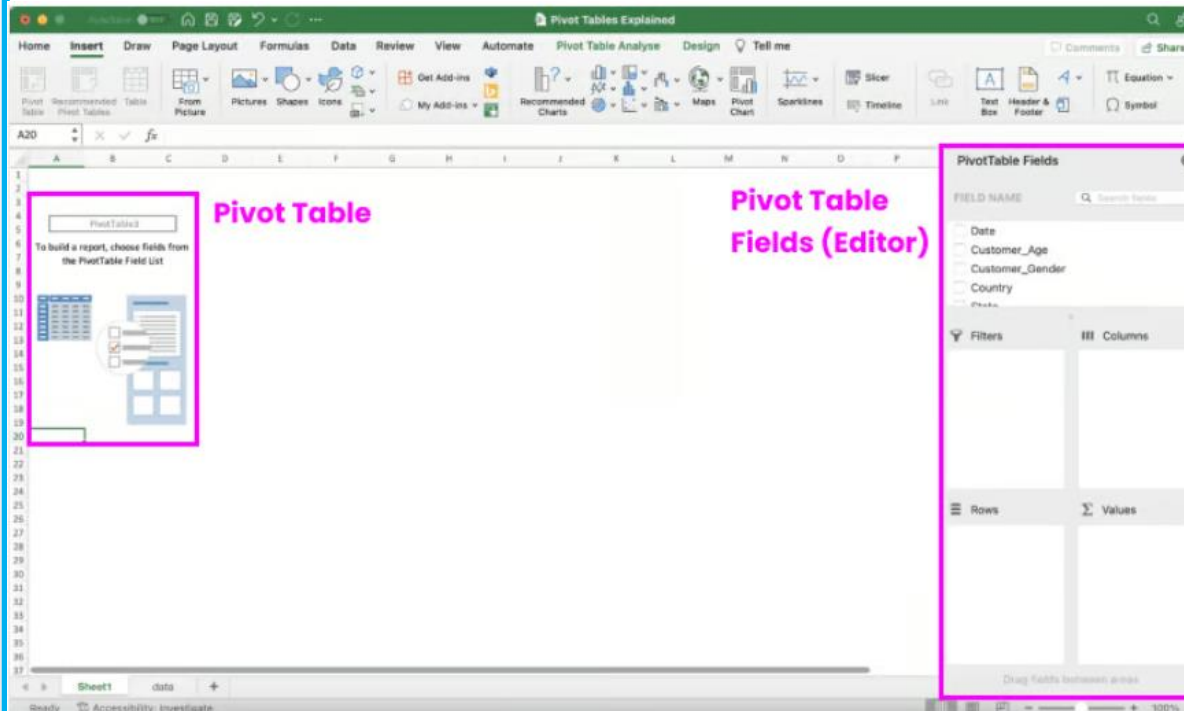
Using Excel for data analytics:

PivotTables in Excel offer several benefits that enable you to **summarize and analyze large amounts of data to help quickly identify trends and patterns**. They are a popular user-friendly tool enabling technical/non-technical people alike to dive deeper into their data.

At the most basic level, PivotTables enable you to **create a data matrix in a row and column formats**. You can apply filtering, sorting, aggregation, and summarization of your data in various ways.

11

Using Excel for data analytics:



12

Using Excel for data analytics:

Using LLM for Asking Business Questions



Excel Project Dataset.xlsx
Spreadsheet

Can you ask four questions about this data

Here are four questions based on this dataset:

1. How many people in the dataset are homeowners (Home Owner = "Yes")?
2. What is the average income of bike buyers (Purchased Bike = "Yes") compared to non-buyers?
3. Which marital status group has the highest number of bike buyers?
4. What is the most common commute distance among those who purchased a bike?

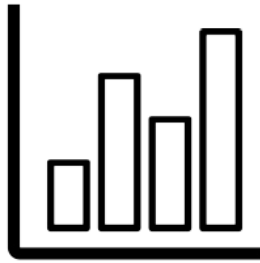
Here are four additional questions you can ask based on this dataset:

1. What is the average age of individuals who purchased a bike?
2. Which occupation category has the highest percentage of bike buyers?
3. How does the number of cars owned relate to bike purchasing behavior?
4. Which region has the highest number of bike buyers?

13

Data Visualization:

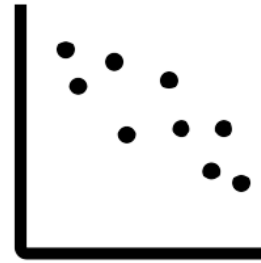
Data visualization (or more appropriately, information visualization) has been defined as “the use of visual representations to explore, make sense of, and communicate data” (Few, 2007).



Bar/column
charts



Line charts



Scatter plots



Stacked or grouped
bar/column charts

14

Data Visualization:

Although the name that is commonly used is *data visualization*, usually what is meant by this is information visualization. Because information is the aggregation, summarization, and contextualization of data (raw facts), what is portrayed in visualizations is the information and not the data.

Data visualization available in both BI applications have included charts and graphs, as well as the other types of visual elements used to create scorecards and dashboards.

Different Types of Charts and Graphs: Often end users of business analytics systems are not sure what type of chart or graph to use for a specific purpose. Some charts or graphs are better at answering certain types of questions. Some look better than others. Some are simple; some are rather complex and crowded.

15

Data Visualization:

Basic Charts and Graphs:

Line charts are perhaps the most frequently used graphical visuals for time series data.

Line charts (or a line graphs) show **the relationship between two variables**. They are most often used to **track changes or trends over time (having one of the variables set to time on the x-axis)**.

Line charts sequentially connect individual data points to help infer changing trends over a period of time.

Line charts are often used to **show time-dependent changes in the values of some measure**, such as **changes on a specific stock price over a 5-year period** or **changes on the number of daily customer service calls over a month**.

16

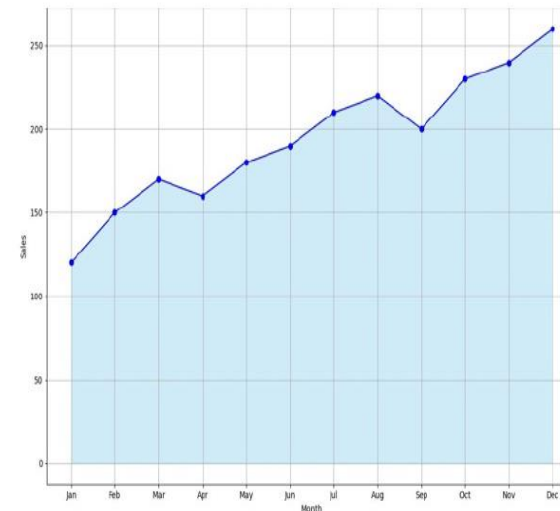
Data Visualization:

Line chart

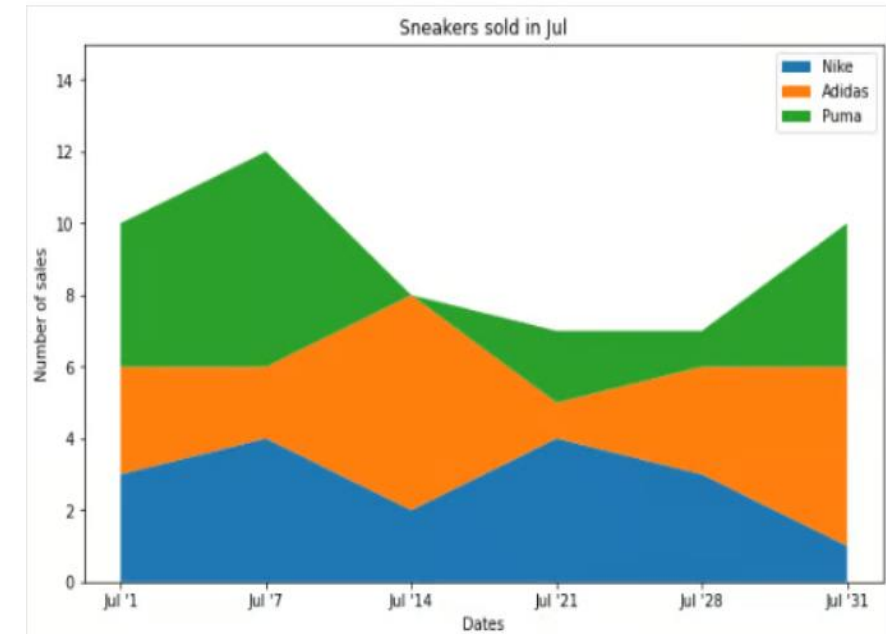


- Easier to see **how sharply** sales increase or decrease

Area chart



- Common variation of line chart
- Emphasizes volume of data over time



Purpose: To show **trends in a numerical feature over time**

17

Data Visualization:

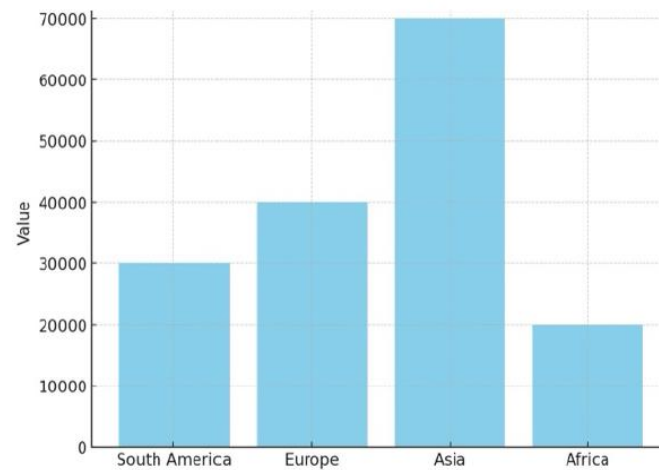
Bar charts are among the most basic visuals used for data representation. Bar charts are effective when you have nominal data or numerical data that splits nicely into different categories so you can quickly see comparative results and trends within your data.

Bar charts are often used to compare data across multiple categories such as percent of advertising spending by departments or by product categories. Bar charts can be vertically or horizontally oriented. They can also be stacked on top of each other to show multiple dimensions in a single chart.

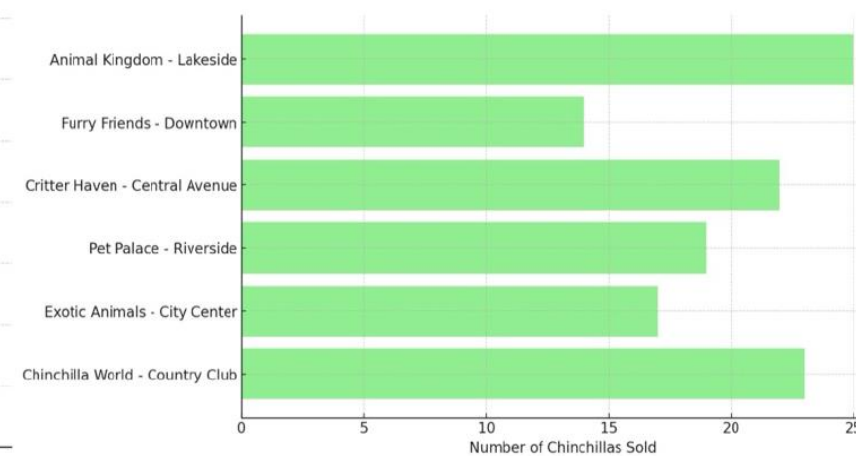
18

Data Visualization:

Album sales per region



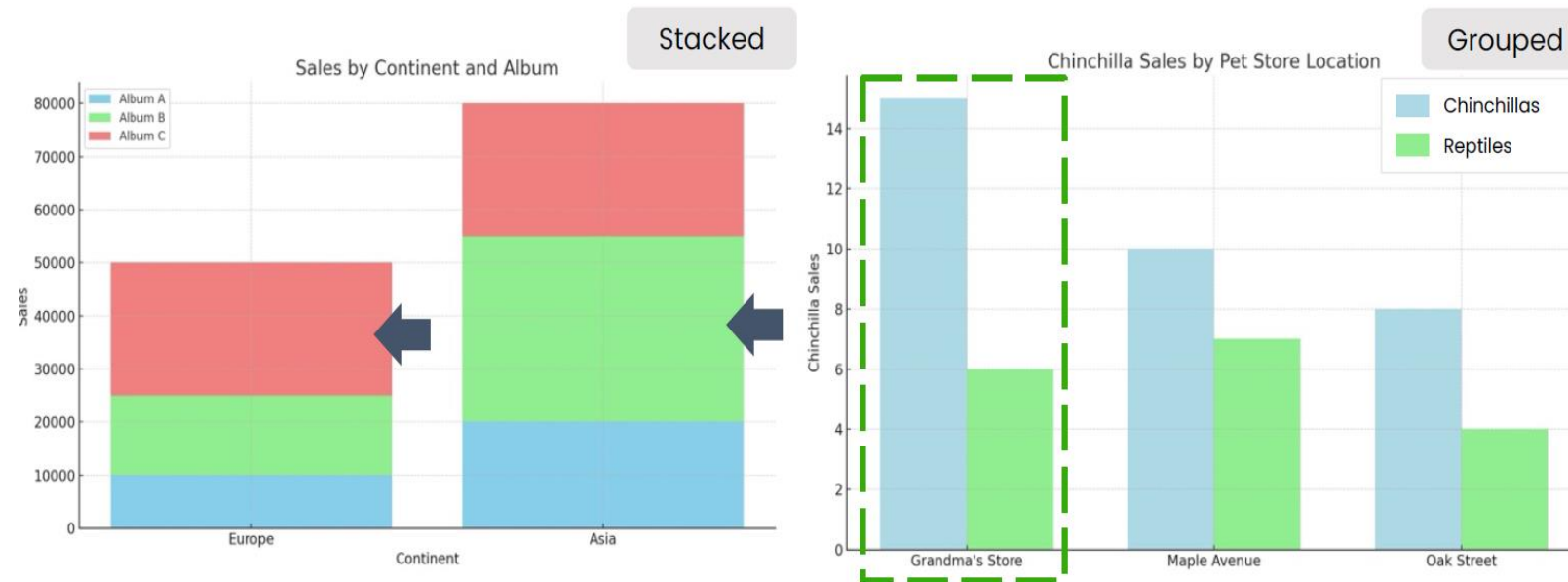
Number of chinchillas per store



Purpose: To compare a
numerical feature across
a categorical feature

19

Data Visualization:



- Shows a part-to-whole relationship

- Better for direct comparison between categories

Purpose: Compare a numerical feature across multiple categorical features

20

Data Visualization:

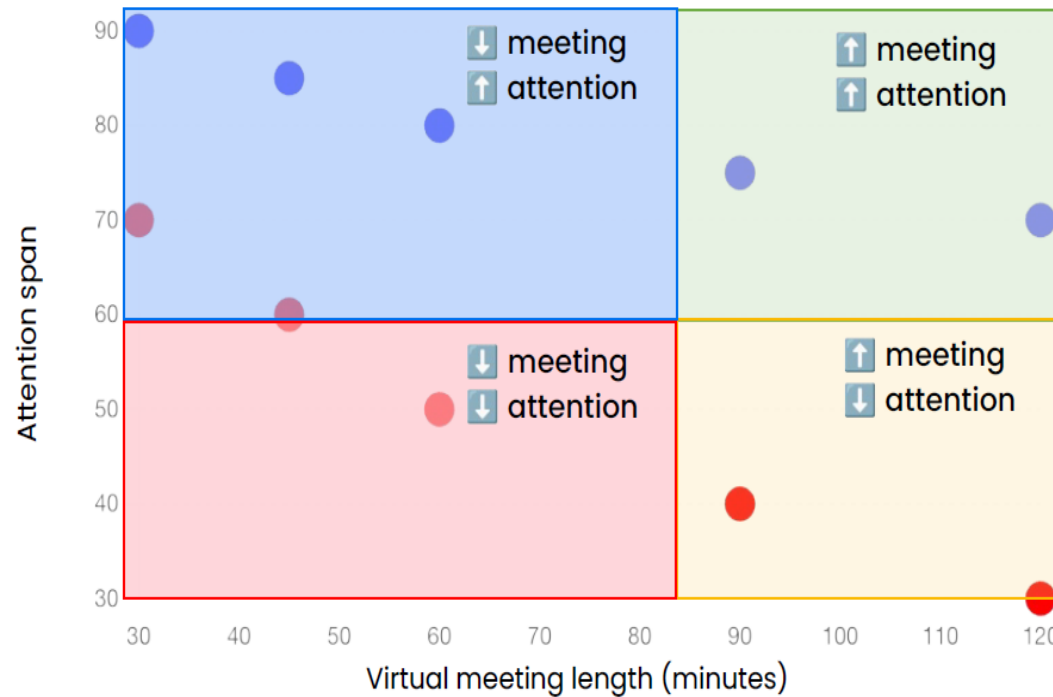
Scatter plots are often used to explore the relationship between two or three variables (in 2-D or 2-D visuals). Because they are visual exploration tools, having more than three variables, translating them into more than three dimensions is not easily achievable.

Scatter plots are an effective way to explore the existence of trends, concentrations, and outliers. For instance, in a two-variable (two-axis) graph, a scatter plot can be used to illustrate the corelationship between age and weight of heart disease patients or it can illustrate the relationship between the number of customer care representatives and the number of open customer service claims. Often, a trend line is superimposed on a two-dimensional scatter plot to illustrate the nature of the relationship.

21

Data Visualization:

Virtual meeting length and attention span



- Personal meetings
- Work meetings

Purpose: To compare two numerical features

22

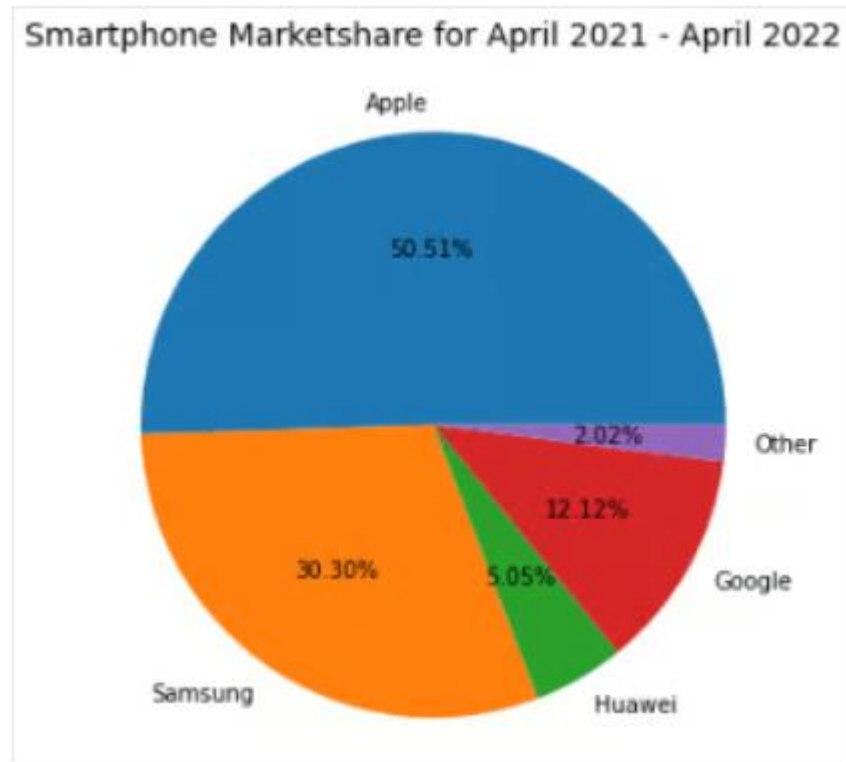
Data Visualization:

Pie charts are visually appealing, as the name implies, pie-looking charts. Because they are so visually attractive, they are often incorrectly used.

Pie charts should only **be used to illustrate relative proportions of a specific measure**. For instance, they can be used to show the **relative percentage of an advertising budget spent on different product lines**, or they can show relative proportions of majors declared by college students in their sophomore year. If the **number of categories to show is more than just a few (say more than four)**, one should seriously consider using a bar chart instead of a pie chart.

23

Data Visualization:



For example, a good use case for a pie chart would be to represent the market share for smartphones.

24

Data Visualization:

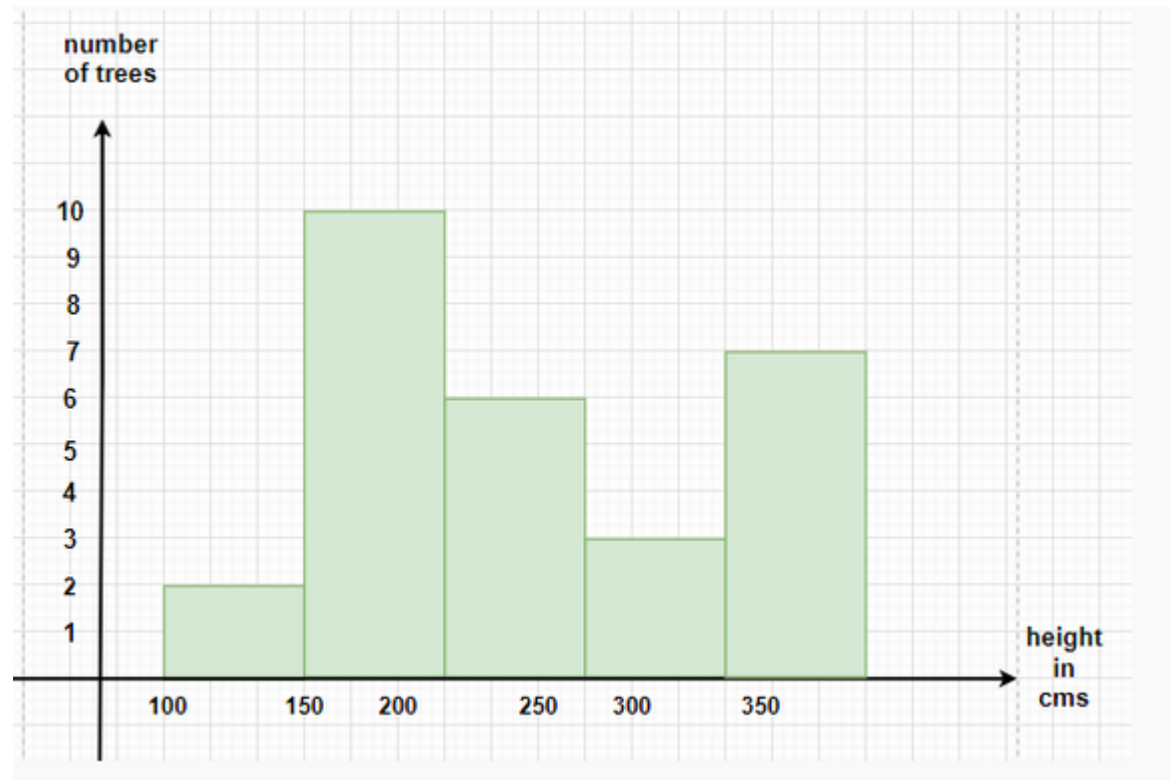
Histogram looks just like a bar chart. Histograms are used to show the frequency distribution of a variable or several variables.

In a histogram, the x -axis is often used to show the categories or ranges, and the y -axis is used to show the measures/values/frequencies. Histograms show the distributional shape of the data.

That way, one can visually examine if the data is normally or exponentially distributed. For instance, one can use a histogram to illustrate the exam performance of a class, where distribution of the grades as well as comparative analysis of individual results can be shown, or one can use a histogram to show age distribution of the customer base.

25

Data Visualization:



The histogram is similar to a bar graph, but histogram groups numbers into different ranges.

26

Data Visualization:

Geographic Map : when the data set includes any kind of location data (e.g., physical addresses, postal codes, state names or abbreviations, country names, latitude/longitude, or some type of custom geographic encoding), it is better and more informative to see the data on a map.

For instance, one can use maps to show distribution of customer service requests by product type by geographic locations.

Often a large variety of information (e.g., age distribution, income distribution, education, economic growth, or population changes) can be portrayed in a geographic map to help decide where to open a new restaurant or a new service station.

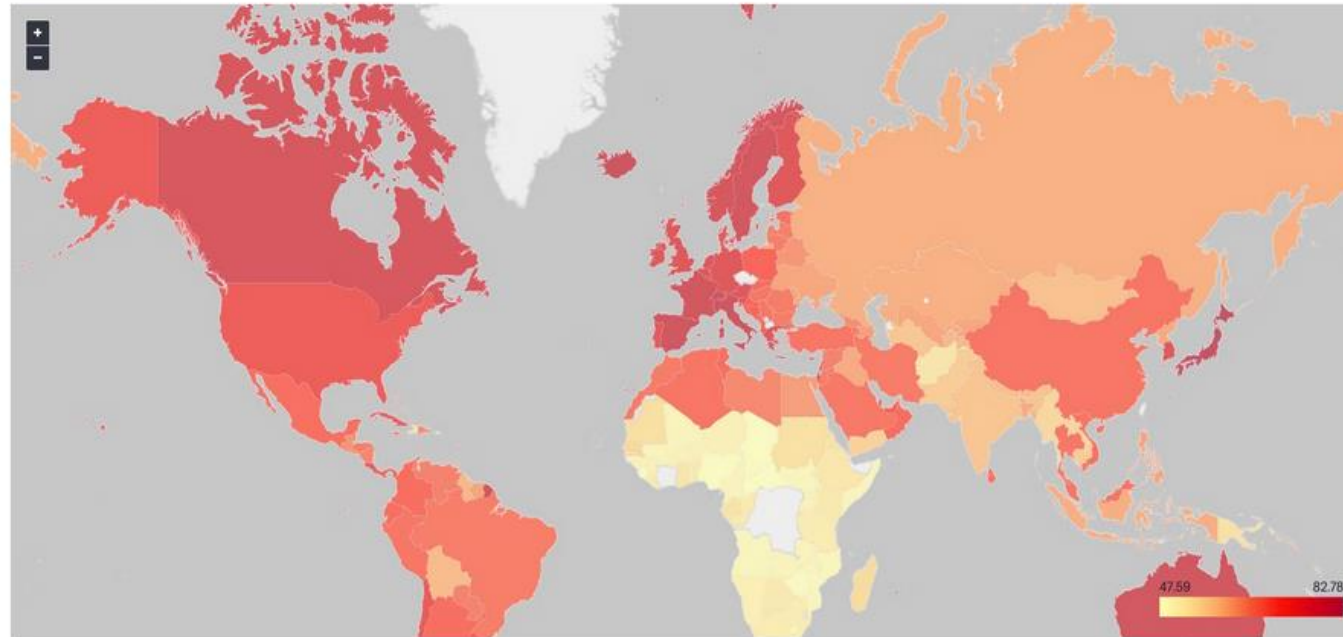
27

Data Visualization:

Average Life Expectancy At Birth, Total (years) by Country

Add description

Year >= 01/01/2001



They show spatial information, such as the distribution of values across different regions, countries, or states.

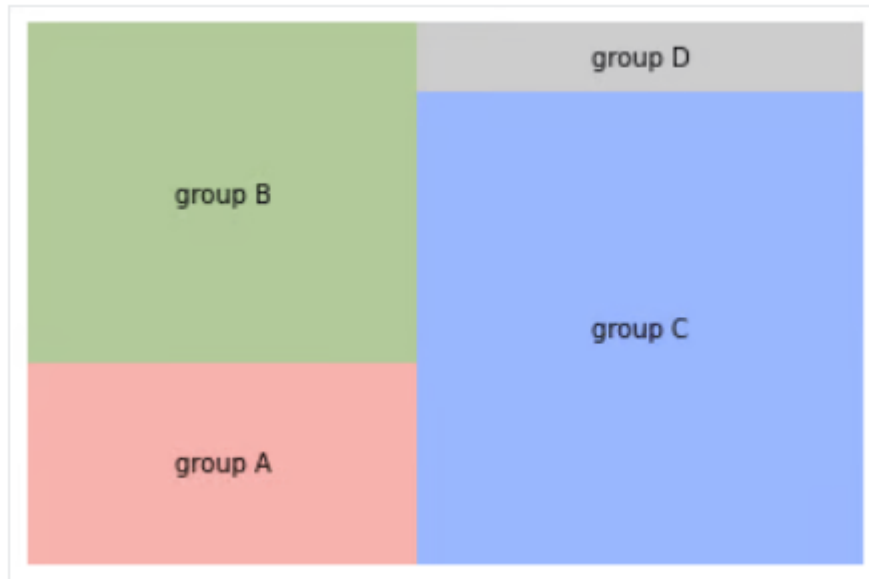
28

Data Visualization:

Tree maps display hierarchical (tree-structured) data as a set of nested rectangles. Each branch of the tree is given a rectangle, which is then tiled with smaller rectangles representing subbranches. A leaf node's rectangle has an area proportional to a specified dimension on the data. Often the leaf nodes are colored to show a separate dimension of the data. When the color and size dimensions are correlated in some way with the tree structure, one can often easily see patterns that would be difficult to spot in other ways, such as if a certain color is particularly relevant. A second advantage of tree maps is that, by construction, they make efficient use of space. As a result, they can legibly display thousands of items on the screen simultaneously.

29

Data Visualization:



Each rectangle is a category within a given variable, whereas the area of the rectangle is proportional to the size of that category.

30

Data Visualization and Best Practices:**Tips for Creating Effective Data Visualizations:**

1. Data visualizations should have a clear purpose and audience.
2. Choose the right type of viz or chart for your data.
3. Use text and labels to clarify, not clutter.
4. Use color to highlight important information or to differentiate or compare.
5. Avoid misleading visualizations.
6. Keep your visualizations simple. Less is more.

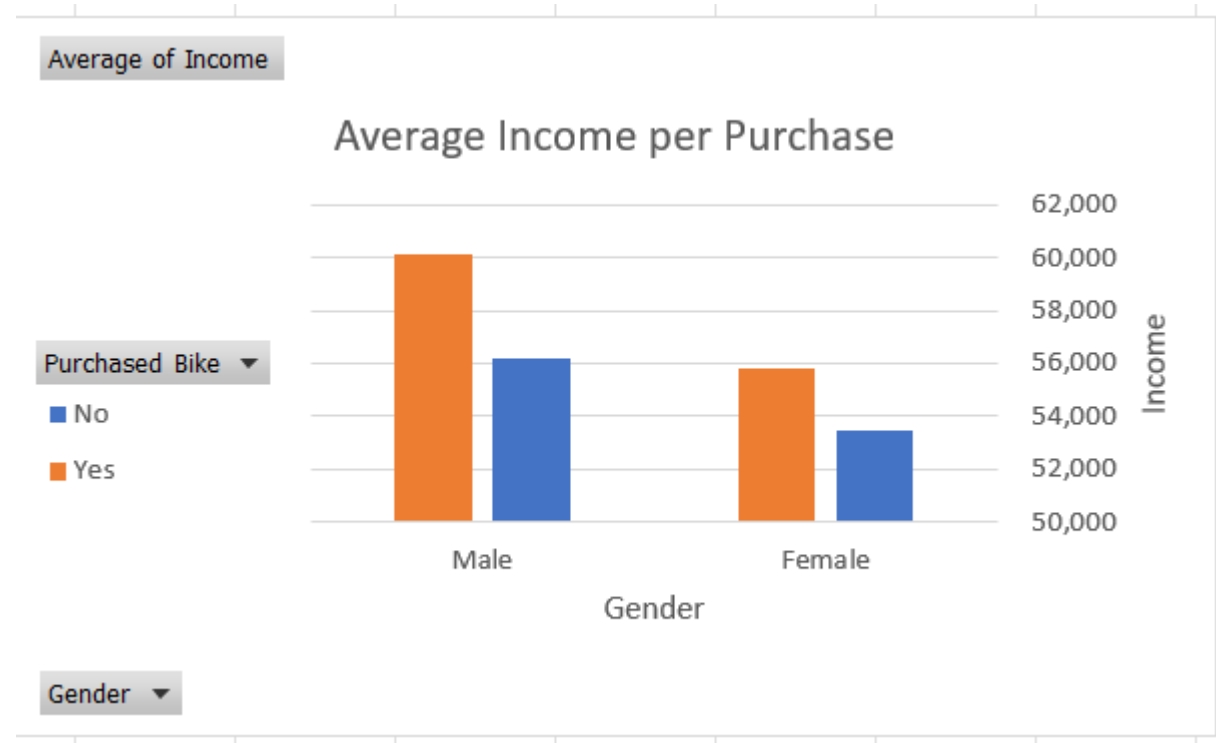
31

Data Visualization :

Business Question:

What is the average income for each gender among people who purchased a bike?

Average of Income			
Column Labels			
Row Labels	No	Yes	Grand Total
Female	53,440	55,774	54,581
Male	56,208	60,124	58,063
Grand Total	\$54,875	\$57,963	\$56,360



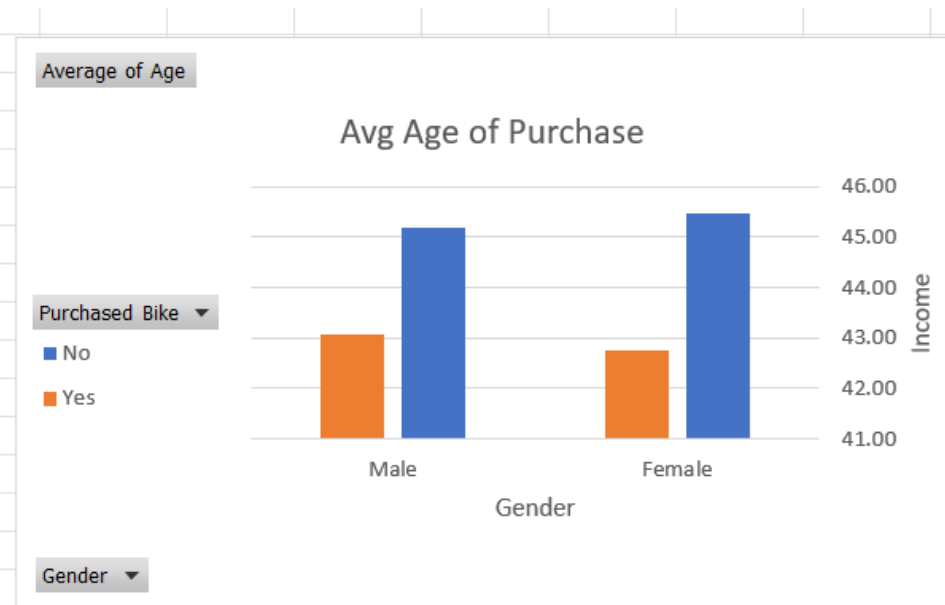
32

Data Visualization :

Business Question:

What is the average age of individuals who purchased a bike?

Average of Age	Column Labels		
Row Labels	No	Yes	Grand Total
Female	45.48	42.75	44.15
Male	45.19	43.06	44.18
Grand Total	45.33	42.91	44.16



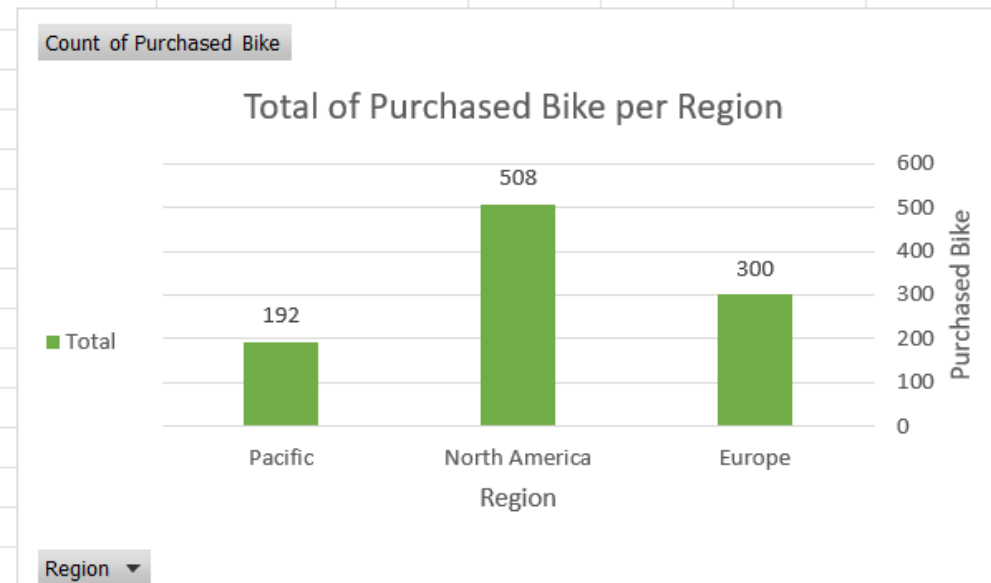
33

Data Visualization :

Business Question:

Which region has the highest number of bike buyers?

Row Labels	Count of Purchased Bike
Europe	300
North America	508
Pacific	192
Grand Total	1000



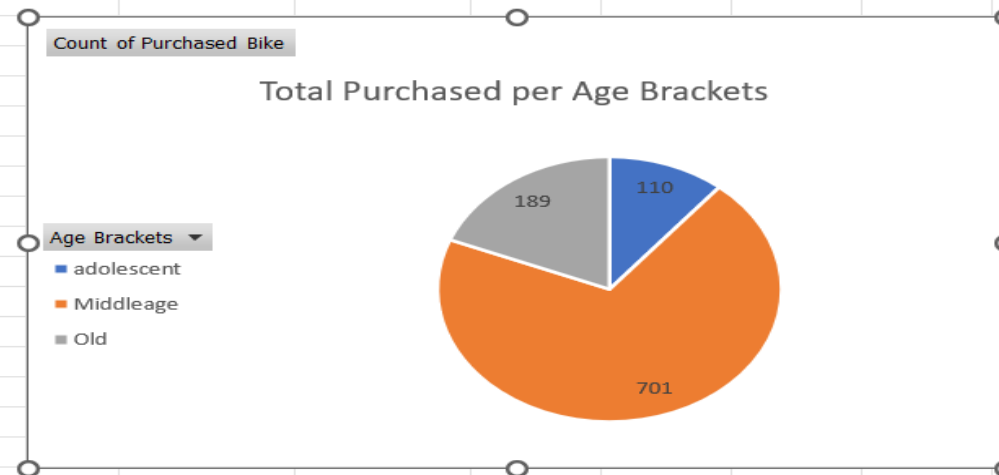
34

Data Visualization :

Business Question:

What is the total purchased Bike for each age brackets?

Row Labels	Count of Purchased Bike
adolescent	110
Middleage	701
Old	189
Grand Total	1000



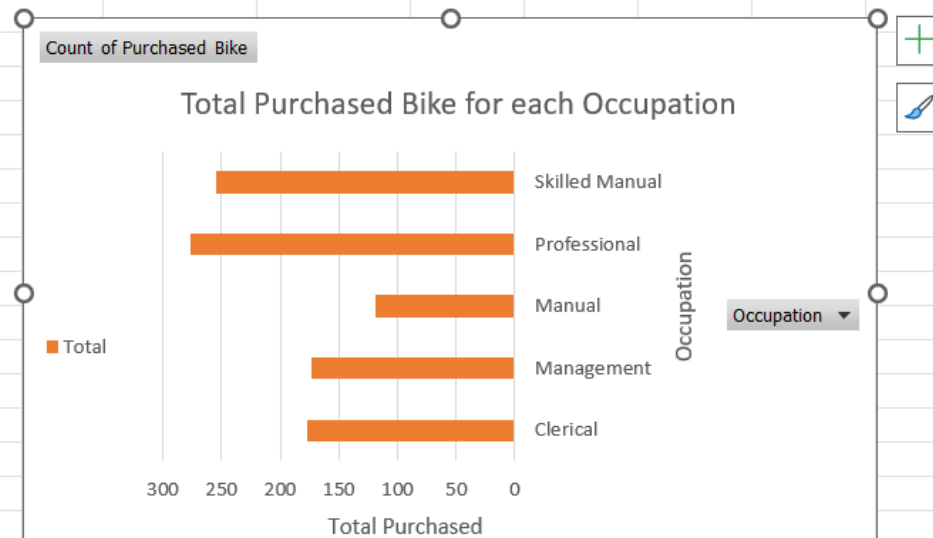
35

Data Visualization :

Business Question:

Which occupation category has the highest percentage of bike buyers?

Row Labels	Count of Purchased Bike
Clerical	177
Management	173
Manual	119
Professional	276
Skilled Manual	255
Grand Total	1000



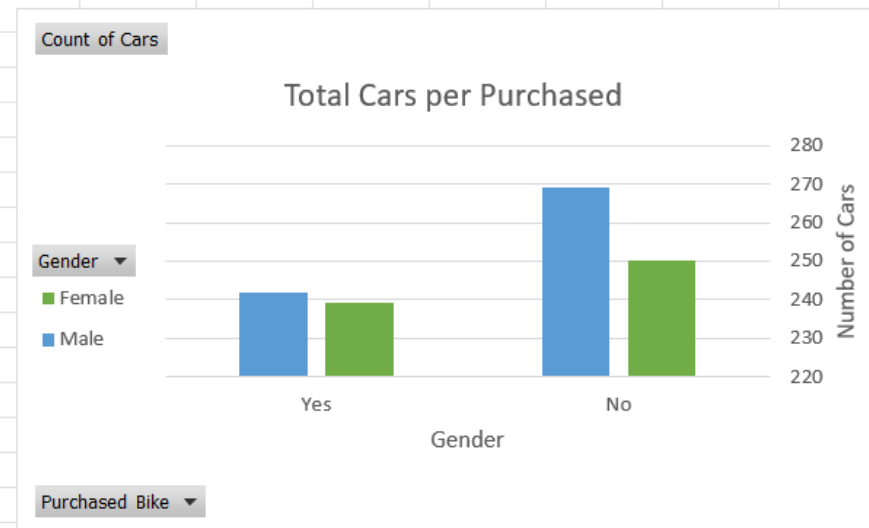
36

Data Visualization :

Business Question:

How does the number of cars owned relate to bike purchasing behavior?

Count of Cars			
Column Labels			
Row Labels	Female	Male	Grand Total
No	250	269	519
Yes	239	242	481
Grand Total	489	511	1000



37

Summary

1. Using Excel for data analytics.
2. data visualization and best practices.
3. How to use Excel for data visualization.



Thank You