Aim: To perform Data Exploration using Microsoft Excel.

Lab Outcome:

• Understand the concept of data exploration and its importance in business decision-making.

• Identify and classify different types of data such as nominal, ordinal, interval, and ratio.

• Create and interpret visualizations such as Pie Charts, Bar Charts, and Box Plots.

Theory:

Q1. What is Data Exploration?

Data exploration is the process of analysing and summarizing data to understand its main characteristics before applying any advanced analysis or modelling. It involves examining patterns, trends, outliers, and relationships in data using statistical and visual tools.

Q2. What are the Types of Data?

Data can be classified into four main types:

1. Nominal Data: Nominal data represents categories or labels without any specific order. Example: City names (Boston, New York, Los Angeles), Region (East, West).

2. Ordinal Data: Ordinal data represents categories with a meaningful order but without equal intervals between values. Example: Product ratings such as Poor, Average, Good, Excellent.

3. Interval Data: Interval data has ordered values with equal differences between them, but no true zero point. Example: Temperature in Celsius or Fahrenheit.

Q3. Data set used:

Customer	Region	City	Category	Product	Quantity	Price ($)	Product Rating
1	East	Boston	Bars	Carrot	27	3.5	4.2
2	East	Boston	Crackers	Whole Wheat	40	2.8	3.8
3	West	Los Angeles	Cookies	Chips	20	4.2	4
4	East	New York	Bars	Carrot	43	5.1	4.5
5	West	San Diego	Snacks	Whole Wheat	60	6.4	3.9
6	East	Los Angeles	Cookies	Oat meal	52	3.9	4.1
7	East	Boston	Crackers	Oat meal	100	7.5	4.6
8	West	Los Angeles	Bars	Carrot	72	4.8	4.3
9	East	New York	Snacks	Potato Chips	91	6.9	4.7
10	West	San Diego	Crackers	Whole Wheat	102	8.2	4.4
11	East	Boston	Snacks	Granola	55	5.6	4
12	West	Los Angeles	Cookies	Chocolate Chip	68	7.9	4.8
13	East	New York	Bars	Oats	49	4.1	3.7
14	West	San Diego	Snacks	Nachos	77	6.3	4.2
15	East	Los Angeles	Crackers	Multigrain	88	8.7	4.6
16	East	Boston	Cookies	Butter	34	2.4	3.5
17	West	Los Angeles	Snacks	Corn Chips	95	9.1	4.9
18	East	New York	Crackers	Oat meal	66	5.2	4.1
19	West	San Diego	Bars	Peanut	81	6.8	4.4
20	East	Boston	Snacks	Trail Mix	59	7.3	4.7

Steps to create a pie chart:

1. Enter the list of cities in column J:

    - Boston
    - Los Angeles
    - San Diego
    - New York

2. In cell K1, type Customer Count.

3. In cell K2, enter the formula: =COUNTIF(C2:C21, J2)

4. Drag the formula down to get the count for all cities.

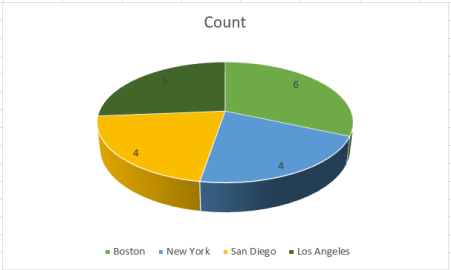
5. Select the range J1:K5.

6. Go to Insert → Pie Chart → 2-D Pie.

7. Give the chart title as: Customer Distribution by City The pie chart shows how customers are distributed among different cities. Boston and Los Angeles have the highest number of customers, while San Diego and New York have fewer customers.

Customer	Region	City	Category	Product	Quantity	Price ($)	Product Rating
1	East	Boston	Bars	Carrot	27	3.5	4.2
2	East	Boston	Crackers	Whole Wheat	40	2.8	3.8
3	West	Los Angeles	Cookies	Chips	20	4.2	4
4	East	New York	Bars	Carrot	43	5.1	4.5
5	West	San Diego	Snacks	Whole Wheat	60	6.4	3.9
6	East	Los Angeles	Cookies	Oat meal	52	3.9	4.1
7	East	Boston	Crackers	Oat meal	100	7.5	4.6
8	West	Los Angeles	Bars	Carrot	72	4.8	4.3
9	East	New York	Snacks	Potato Chips	91	6.9	4.7
10	West	San Diego	Crackers	Whole Wheat	102	8.2	4.4
11	East	Boston	Snacks	Granola	55	5.6	4
12	West	Los Angeles	Cookies	Chocolate Chip	68	7.9	4.8
13	East	New York	Bars	Oats	49	4.1	3.7
14	West	San Diego	Snacks	Nachos	77	6.3	4.2
15	East	Los Angeles	Crackers	Multigrain	88	8.7	4.6
16	East	Boston	Cookies	Butter	34	2.4	3.5
17	West	Los Angeles	Snacks	Corn Chips	95	9.1	4.9
18	East	New York	Crackers	Oat meal	66	5.2	4.1
19	West	San Diego	Bars	Peanut	81	6.8	4.4
20	East	Boston	Snacks	Trail Mix	59	7.3	4.7

City	Count
Boston	6
New York	4
San Diego	4
Los Angeles	5



Steps to create a bar chart:

1. Use the same city list in column J.

2. In cell K1, type Average Rating.

3. In cell K2, enter this formula: =AVERAGEIF(C2:C21, J2, H2:H21)

4. Drag the formula down to calculate average ratings for all cities.

5. Select the range J1:K5.

6. Go to Insert → Column Chart.

7. Add titles:

o Chart Title: Average Product Rating by City

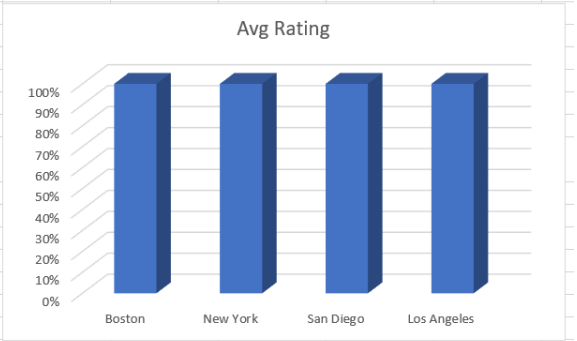
o X-Axis: City

o Y-Axis: Average Rating

The graph compares the average customer ratings of products across different cities. Los Angeles has the highest average rating, indicating higher customer satisfaction.

Customer ID	Region	City	Category	Product	Quantity	Price ($)	Product Rating
1	East	Boston	Bars	Carrot	27	3.5	4.2
2	East	Boston	Crackers	Whole Wheat	40	2.8	3.8
3	West	Los Angeles	Cookies	Chips	20	4.2	4
4	East	New York	Bars	Carrot	43	5.1	4.5
5	West	San Diego	Snacks	Whole Wheat	60	6.4	3.9
6	East	Los Angeles	Cookies	Oat meal	52	3.9	4.1
7	East	Boston	Crackers	Oat meal	100	7.5	4.6
8	West	Los Angeles	Bars	Carrot	72	4.8	4.3
9	East	New York	Snacks	Potato Chips	91	6.9	4.7
10	West	San Diego	Crackers	Whole Wheat	102	8.2	4.4
11	East	Boston	Snacks	Granola	55	5.6	4
12	West	Los Angeles	Cookies	Chocolate Chip	68	7.9	4.8
13	East	New York	Bars	Oats	49	4.1	3.7
14	West	San Diego	Snacks	Nachos	77	6.3	4.2
15	East	Los Angeles	Crackers	Multigrain	88	8.7	4.6
16	East	Boston	Cookies	Butter	34	2.4	3.5
17	West	Los Angeles	Snacks	Corn Chips	95	9.1	4.9
18	East	New York	Crackers	Oat meal	66	5.2	4.1
19	West	San Diego	Bars	Peanut	81	6.8	4.4
20	East	Boston	Snacks	Trail Mix	59	7.3	4.7

City	Avg Rating
Boston	4.133333333
New York	4.25
San Diego	4.225
Los Angeles	4.54



Steps for box and whisker:

East Column (J2) formula to enter to get the quantity of Bars sold in the east region:

=FILTER(F:F,(B:B="East")*(D:D="Bars"))

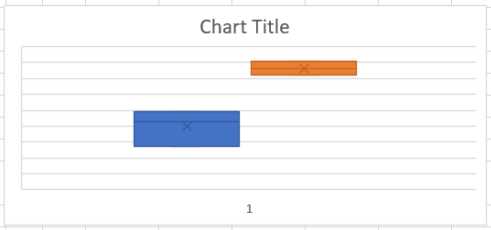
West Column (K2) formula to enter to get the quantity of Bars sold in the west region:

=FILTER(F:F,(B:B="West")*(D:D="Bars"))

1. Create headers East and West.
2. Enter the above formulas.
3. Select the generated data.
4. Go to Insert → Statistical → Box & Whisker.

The box plot shows that the West region has higher median and spread of quantity ordered for Bars compared to East.

Customer ID	Region	City	Category	Product	Quantity	Price ($)	Product Rating		East	West
1	East	Boston	Bars	Carrot	27	3.5	4.2		27	72
2	East	Boston	Crackers	Whole Wheat	40	2.8	3.8		43	81
3	West	Los Angeles	Cookies	Chips	20	4.2	4		49	
4	East	New York	Bars	Carrot	43	5.1	4.5			
5	West	San Diego	Snacks	Whole Wheat	60	6.4	3.9			
6	East	Los Angeles	Cookies	Oat meal	52	3.9	4.1			
7	East	Boston	Crackers	Oat meal	100	7.5	4.6			
8	West	Los Angeles	Bars	Carrot	72	4.8	4.3			
9	East	New York	Snacks	Potato Chips	91	6.9	4.7			
10	West	San Diego	Crackers	Whole Wheat	102	8.2	4.4			
11	East	Boston	Snacks	Granola	55	5.6	4			
12	West	Los Angeles	Cookies	Chocolate Chip	68	7.9	4.8			
13	East	New York	Bars	Oats	49	4.1	3.7			
14	West	San Diego	Snacks	Nachos	77	6.3	4.2			
15	East	Los Angeles	Crackers	Multigrain	88	8.7	4.6			
16	East	Boston	Cookies	Butter	34	2.4	3.5			
17	West	Los Angeles	Snacks	Corn Chips	95	9.1	4.9			
18	East	New York	Crackers	Oat meal	66	5.2	4.1			
19	West	San Diego	Bars	Peanut	81	6.8	4.4			
20	East	Boston	Snacks	Trail Mix	59	7.3	4.7			



Conclusion

This practical demonstrated how Excel can be used for effective data exploration and visualization. By using charts such as Pie Charts, Bar Graphs, and Box Plots, meaningful insights about customer distribution, product ratings, and regional demand were obtained. These visual tools help in better understanding of business data and support informed decision-making.