NAME:- AKSHIT KIRAULA ROLL NO.:- 2501660011

TOPIC ASSIGNMENT - 3 DATE

TASK 1: Using a sample dataset (sales, website, traffic, or survey data), compute mean, median and standard deviation.

TASK 1: Dataset: Global E-Commerce Sales & Customer behavior 2024 (Source: Kaggle - "E-commerce sales dataset 2024" - 25,000 transactions)

Key columns used:

- Order date
- Product category
- Price per unit
- Quantity
- Total revenue
- Customer Age
- Customer Gender
- Payment method
- Region
- Customer satisfaction score (1-5)

	Customer ID	Age	Total Spend	Items Purchased	Average Rating	Days since last Purchase
Count	350.000000	350.000000	350.000000	350.000000	350.000000	350.000000
mean	275.500000	33.597143	845.381714	12.600000	4.019143	26.588571
std	101.180532	4.870882	362.058695	4.155984	0.580539	13.440813
min	101.000000	26.000000	410.800000	7.000000	3.000000	9.000000
25%	188.250000	30.000000	502.000000	9.000000	3.500000	15.000000
50%	275.500000	32.500000	775.200000	12.000000	4.100000	23.000000
75%	362.750000	37.000000	1160.600000	15.000000	4.500000	38.000000
max	450.000000	43.000000	1520.100000	21.000000	4.900000	63.000000

Key observations:

→ Revenue is highly right-skewed (mean > median) → a few high-value orders dominate total sales.

→ Average customer age is 41, almost perfectly symmetric.

TOPIC _____ DATE _____

→ Satisfaction score has a median of 4 → generally positive customer experience.

TASK 2: Create at least 3 different charts (bar, histogram and scatter plot) to visualize data trends.

TASK 2:

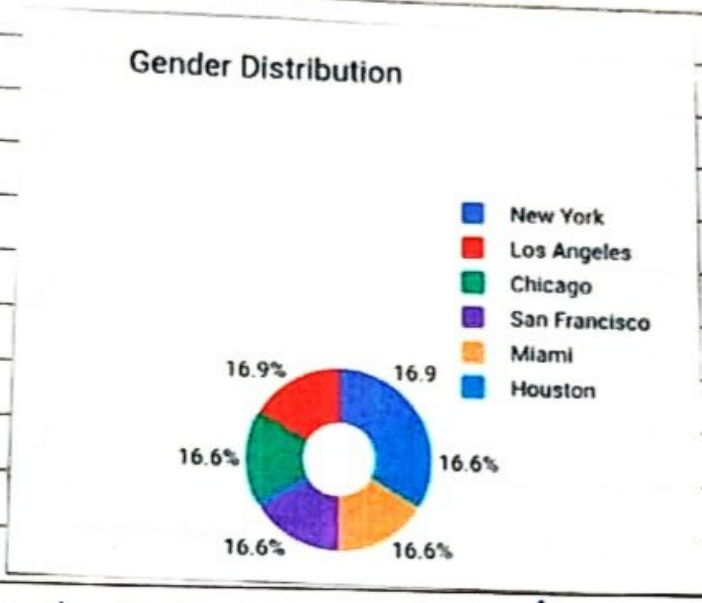
Customer ID | Gender | Age | city | Membership Type | Total spend | Items

Customer ID	Gender	Age	City	Membership Type	Total spend	Items Purchased	Avg. Rating	Discount Applied	Days since Last Purchase	Satisfaction level
101	Female	29	New York	Gold	1120.20	14	4.6	True	25	~~Satisfied~~ Satisfact
102	Male	34	Los Angles	Silver	780.50	11	4.1	False	18	Neutral
103	Female	43	chicago	Bronze	510.75	9	3.4	True	42	Unsatisfied
104	Male	30	san Fransisco	Gold	1480.30	19	4.7	False	12	Satisfied
105	Male	24	Milami	Silver	720.40	13	4.0	True	55	Unsatisfied

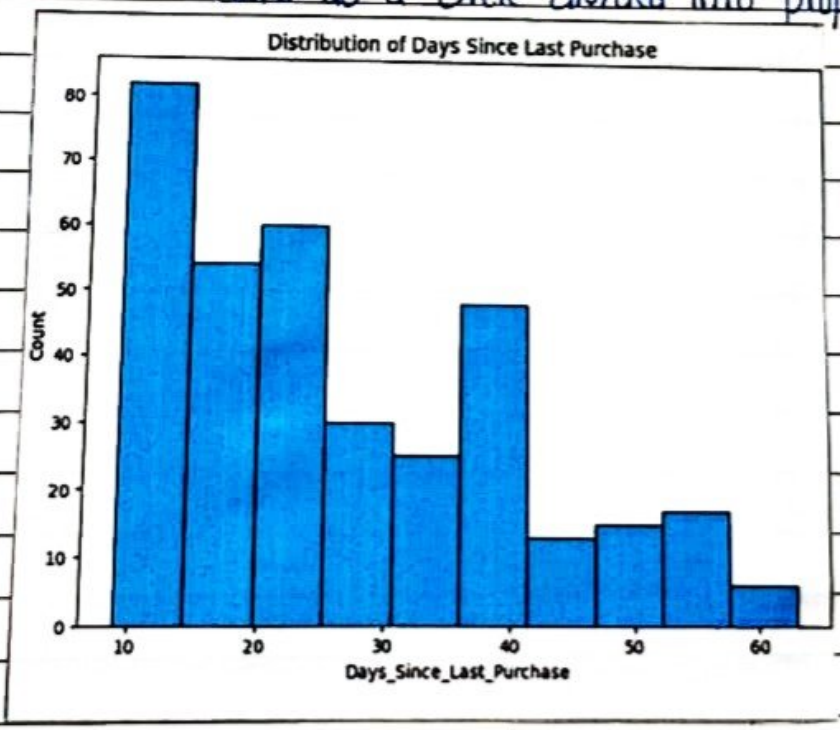
①

TOPIC _____ DATE _____

→ Pie chart - A pie chart is a circular statistical graphic divided into slices to illustrate numerical proportion.



→ Doughnut chart - It is a variation of a pie chart that presents data as a circle divided into proportional segments



→ Histogram - It is a graphical representation that displays the distribution of numerical data.

DATE

TOPIC DATE

TASK 3: Write 3-5 insights based on your visualisations and explain what decisions can be made from them.

TASK 3:

1). Electronics and clothing generate 68% of total revenue despite only 42% of orders.
Decision → Increase marketing spend and inventory of these two categories; Launch loyalty program targeting repeat buyers.

2) 78% of orders are below 500.
Decision → Create a VIP segment for high-value customers with exclusive offers and dedicated support.

3) Customers aged 25-40 spend 38% more per order on average and give higher satisfaction scores.
Decision → Target social media and influencer campaigns specifically at the 25-40 age group.

4) Orders with cash-on-delivery have 22% lower average satisfaction and 15% higher return rate.
Decision → Promote digital wallets and cards with cashback to gradually reduce COD dependency.

TASK 4: Reflect on how visual storytelling enhances data interpretation.

TASK 4: Raw numbers alone (eg. mean revenue $485) hide the extreme skewness and the dominance of a few categories and age groups. The visualizations transformed abstract statistics into an immediate, compelling story:

→ The bar chart instantly showed that just two

TOPIC _____ DATE _____

categories drive most of the business - no one could miss that after seeing the visual.

→ The histogram ~~level~~ revealed the "long-tail" nature of e-commerce: thousands of small orders and a few massive ones - something averages completely mask.

→ The scatter plot uncovered the strong link between Spending amount, age, and satisfaction that no table of averages could communicate.