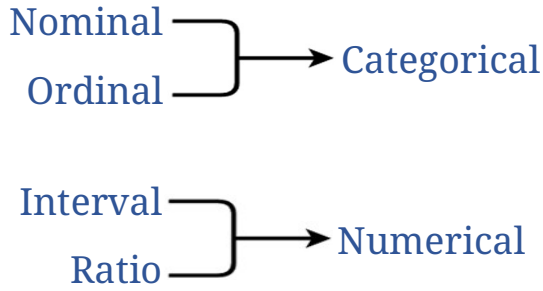




০০১-০০৮ Categorical Data

আগে আমাদের জানতে হবে যে,



ফ্রিকুয়েন্সি ডিস্ট্রিবিউশনঃ

যেই পদ্ধতি ব্যবহার করে উপাত্তগুলো কে বিভিন্ন শ্রেণিতে বিভক্ত করা হয় এবং তাদের পুনরাবৃত্তির সংখ্যা হিসাব করা হয়, তাকে ফ্রিকুয়েন্সি ডিস্ট্রিবিউশন বলে।

যেমনঃ A, A, C, A, C, B, A, B, B, A, D, D, D, C, B

এখানে A, B, C, আর D হলো উপরের উপাত্তের শ্রেণী, ইংরেজিতে যাকে Category বলা হয়।

এক্ষেত্রে নিম্নে এর ফ্রিকুয়েন্সি টেবিল বানানো হলো।

শ্রেণী	ফ্রিকুয়েন্সি
A	5
B	4
C	3
D	3
Total	15

এখানে লক্ষ্য করুন, প্রত্যেকটা শ্রেণীর repetitions তাদের ফ্রিকুয়েন্সি।

যেমনঃ A শ্রেণী উপরের উপাত্তে 5 বার repetition হয়েছে, তাই A এর ফ্রিকুয়েন্সি হলো 5। একইভাবে B এর ফ্রিকুয়েন্সি 4; C এর ফ্রিকুয়েন্সি 3; D এর ফ্রিকুয়েন্সি 3।

মোট ফ্রিকুয়েন্সি 15

ট্যালি মার্কঃ

আমরা A, A, C, A, C, B, A, B, B, A, D, D, D, C, B এর ফ্রিকুয়েন্সি টেবিল বানাতে পেরেছি। এখন আমরা এর ট্যালি মার্ক টেবিল বানাবোঃ

শ্রেণী	ট্যালি মার্ক	ফ্রিকুয়েন্সি
A		5
B		4
C		3
D		3
Total		15

ফ্রিকুয়েন্সি 5 এর ক্ষেত্রে ট্যালি মার্ক হয় ||||

ফ্রিকুয়েন্সি 6 এর ক্ষেত্রে ট্যালি মার্ক হয় |||| |

ফ্রিকুয়েন্সি 7 এর ক্ষেত্রে ট্যালি মার্ক হয় |||| ||

ফ্রিকুয়েন্সি 10 এর ক্ষেত্রে ট্যালি মার্ক হয় |||| ||||

রিলেটিভ ফ্রিকুয়েন্সিঃ

যে পদ্ধতি অনুযায়ী একটি নির্দিষ্ট ঘটনা বা ফলাফল আনুপাতিক হারে উপস্থাপনা করা হয়, তাকে রিলেটিভ ফ্রিকুয়েন্সি বলে। একে Rf দ্বারা প্রকাশ করা হয়।

$$\text{সূত্রঃ } Rf = \frac{\text{Frequency}}{\text{Total Frequency}}$$

যেমনঃ A, A, C, A, C, B, A, B, B, A, D, D, D, C, B এর রিলেটিভ ফ্রিকুয়েন্সি টেবিল দেখানো হলো

শ্রেণী	ট্যালি মার্ক	ফ্রিকুয়েন্সি	রিলেটিভ ফ্রিকুয়েন্সি
A		5	$5/15 = 1/3 = 0.333$
B		4	$4/15 = 0.267$
C		3	$3/15 = 1/5 = 0.2$
D		3	$3/15 = 1/5 = 0.2$
Total		15	1

রিলেটিভ ফ্রিকুয়েন্সি এর Total মান অবশ্যই 1 হতে হবে।

পাই চার্টের জন্য রিলেটিভ ফ্রিকুয়েন্সি অবশ্যই গুরুত্বপূর্ণ।

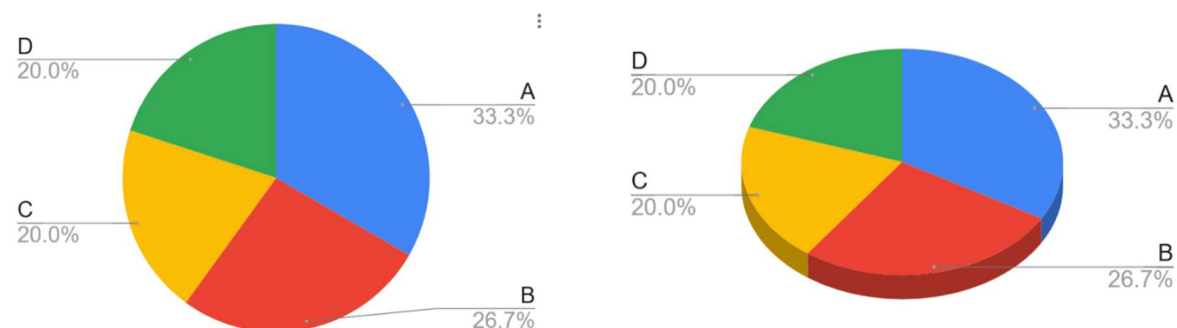
ডেটা ভিজুয়ালাইজেশনঃ

পাই চার্টঃ পাই চার্ট (Pie Chart) হলো এমন এক ডেটা ভিজুয়ালাইজেশন, যেখানে একটা বৃত্তকে কয়েক অংশে বিভক্ত করে ডেটাকে ভিজুয়াল রিপ্রেজেন্ট করা হয়। এক্ষেত্রে আমাদের **অবশ্যই অবশ্যই** রিলেটিভ ফ্রিকুয়েন্সি নির্ণয় করতে হবে, এরপর সেই রিলেটিভ ফ্রিকুয়েন্সি কে 360° দ্বারা গুন করে ডিগ্রি নির্ণয় করতে হবে। সেই ডিগ্রি অনুযায়ী পাইচার্ট বানাতে হবে।

যেমনঃ A, A, C, A, C, B, A, B, B, A, D, D, D, C, B এর ক্ষেত্রে

শ্রেণী	ফ্রিকুয়েন্সি	রিলেটিভ ফ্রিকুয়েন্সি	ডিগ্রি
A	5	$5/15 = 1/3 = 0.333$	$0.333 * 360^\circ = 119.88^\circ$
B	4	$4/15 = 0.267$	$0.267 * 360^\circ = 96.12^\circ$
C	3	$3/15 = 1/5 = 0.2$	$0.2 * 360^\circ = 72^\circ$
D	3	$3/15 = 1/5 = 0.2$	$0.2 * 360^\circ = 72^\circ$
Total	15	1	360°

নিচে পাইচার্ট দুইভাবে বানিয়ে দেখানো হলো (Courtesy: This visualization was generated using Google Sheets.)



এখানে আরও একটা জিনিস লক্ষ্য করুন, পাইচার্টে পার্সেন্টেজ করে A, B, C, D দেখানো হয়েছে।

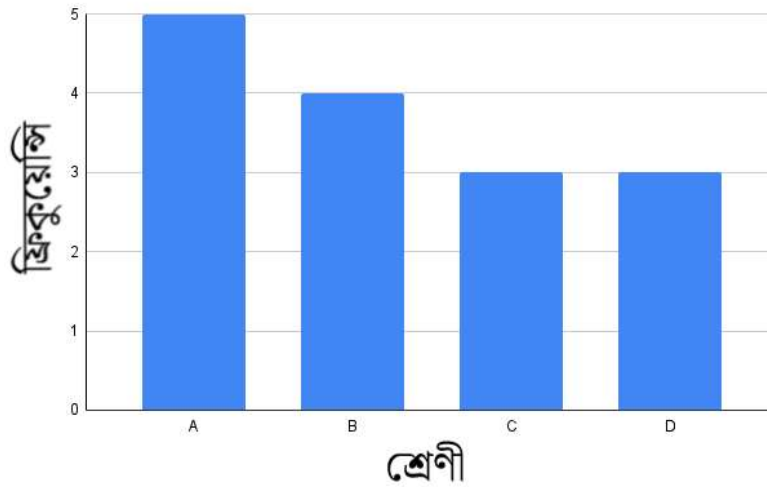
এই পার্সেন্টেজ নির্ণয় করা হয়েছে রিলেটিভ ফ্রিকুয়েন্সি কে 100% দ্বারা গুন করে।

শ্রেণী	ফ্রিকুয়েন্সি	রিলেটিভ ফ্রিকুয়েন্সি	ডিগ্রি	পার্সেন্টেজ
A	5	0.333	119.88°	$0.333 * 100\% = 33.3\%$
B	4	0.267	96.12°	$0.267 * 100\% = 26.7\%$
C	3	0.2	72°	$0.2 * 100\% = 20.0\%$
D	3	0.2	72°	$0.2 * 100\% = 20.0\%$
Total	15	1	360°	100.0%

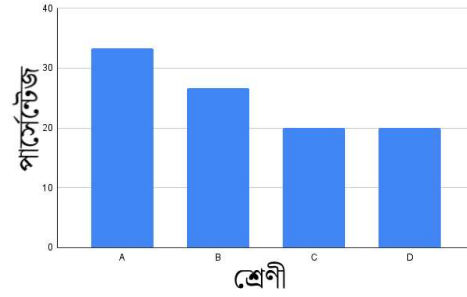
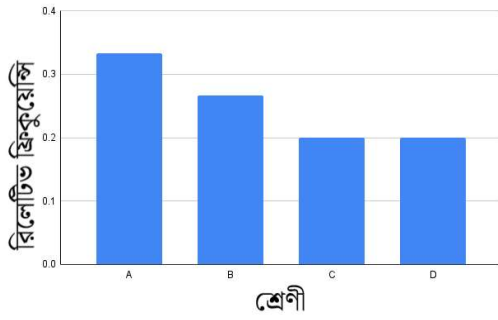
বার চার্টঃ বার চার্ট (Bar Chart) হলো এমন এক ডেটা ভিজুয়ালাইজেশন, যেখানে সকল শ্রেণি আর তাদের ফ্রিকুয়েন্সি কে লম্বালম্বিভাবে রিপ্রেজেন্ট করা হয়।

যেমনঃ ঐ একই A, A, C, A, C, B, A, B, B, A, D, D, D, C, B এর ক্ষেত্রে বার চার্ট দেখানো হলো।

(Courtesy: This visualization was generated using Google Sheets.)



এছাড়া ফ্রিকুয়েন্সি বাদেও পার্সেন্ট, রিলেটিভ ফ্রিকুয়েন্সি দ্বারা বার চার্ট গঠন করা যায়।



(Courtesy: This visualization was generated using Google Sheets.)

বার চার্ট কে অনেক সময় কলাম চার্ট (column chart) বলাও হয়।

সেন্ট্রাল টেন্ডেন্সি (Central Tendency)

সেন্ট্রাল টেন্ডেন্সি তিন ধরনের। যথাঃ গড় (Mean), মধ্যক (Median), প্রচুরক (Mode)

Categorical data তে গড় (Mean) পসিবল নয়।

প্রচুরক (Mode): Categorical data এর ক্ষেত্রে যার ফ্রিকুয়েন্সি বেশি, সেই শ্রেণী হলো mode।

যেমনঃ A, A, C, A, C, B, A, B, B, A, D, D, D, C, B এর ক্ষেত্রে,

শ্রেণী	ফ্রিকুয়েন্সি
A	5
B	4
C	3
D	3
Total	15

এখানে A এর ফ্রিকুয়েন্সি বাকিদের তুলনায় বেশি।

সেই অনুযায়ী A হলো mode।

বার চার্ট আর পাই চার্ট এর বৃহত্তর অংশ কে mode হিসেবে ধরা হয়।

কোনো ডেটা তে দুইটি mode থাকলে তাকে bi-modal data বলা হয়।

যেমনঃ B, A, A, C, A, C, B, A, B, B, A, D, D, D, C, B এর ক্ষেত্রে,

শ্রেণী	ফ্রিকুয়েন্সি
A	5
B	5
C	3
D	3
Total	16

এখানে A আর B উভয়ের ফ্রিকুয়েন্সি বাকিদের তুলনায় বেশি। দুইজনের ফ্রিকুয়েন্সি হলো 5।

সেই অনুযায়ী A আর B হলো mode। আর এই ডেটা হলো bi-modal।

দুইয়ের বেশি mode থাকলে সেই ডেটা কে multi-modal data বলা হয়।

মধ্যক (Median): Categorical data দুই ধরনের। Ordinal data এবং nominal data।

এক্ষেত্রে ordinal data তে মধ্যক ব্যবহৃত হয়। এক্ষেত্রে ডেটা কে sort করে median বের করতে হবে।

যেমনঃ ধরে নিন A, A, C, A, C, B, A, B, B, A, D, D, D, C, B হলো ordinal data। এক্ষেত্রে ordinal হিসেবে দেওয়া আছে $A > B > C > D$ । তাহলে আমরা এই ডেটা কে সর্ট করে পাই,

A, A, A, A, A, B, B, B, B, C, C, C, D, D, D

এখানে total ফ্রিকুয়েন্সি হলো 15। তাহলে,

A, A, A, A, A, B, B, B, B, C, C, C, D, D, D
↑
Median = B = 8th

এক্ষেত্রে median বের করা হয়েছে $\frac{n+1}{2} = \frac{15+1}{2} = 8^{\text{th}}$ এভাবে (এখানে n এর মান 15 কারণ

উপাত্ত 15 টা দেওয়া আছে। আর 15 বিজোড় সংখ্যা। n বিজোড় হলে median হবে $\frac{n+1}{2}$ তম

উপাত্ত।)

যদি জোড় হতো, যেমনঃ B, A, A, C, A, C, B, A, B, B, A, D, D, D, C, B তখন উপাত্ত সংখ্যা বা total ফ্রিকুয়েন্সি হতো 16

এক্ষেত্রে median হবে $\frac{n}{2}$ তম উপাত্ত এবং $\frac{n}{2} + 1$ তম উপাত্ত। অর্থাৎ B, A, A, C, A, C, B, A, B,

B, A, D, D, D, C, B এর median হবে $\frac{16}{2} = 8^{\text{th}}$ তম উপাত্ত আর $\frac{16}{2} + 1 = 9^{\text{th}}$ তম

উপাত্ত।

প্রথমে এই উপাত্তগুলোকে sort করি।

A, A, A, A, A, B, B, B, B, B, C, C, C, D, D, D

Median বের করে দেওয়া হলোঃ

A, A, A, A, A, B, B, B, B, B, C, C, C, D, D, D

Median = B = 8th Median = B = 9th

$\frac{16}{2} = 8^{\text{th}}$ $\frac{16}{2} + 1 = 9^{\text{th}}$

বাস্তবতাঃ

Categorical data এর ক্ষেত্রে আমরা আগেই বলেছিলাম যে “যেই পদ্ধতি ব্যবহার করে উপাত্তগুলোকে বিভিন্ন শ্রেণিতে বিভক্ত করা হয় এবং তাদের পুনরাবৃত্তির সংখ্যা হিসাব করা হয়, তাকে ফ্রিকুয়েন্সি ডিস্ট্রিবিউশন বলে।” আবার এই ফ্রিকুয়েন্সি ডিস্ট্রিবিউশনের মাধ্যমে দেখিয়েছি পাই চার্ট, বার চার্ট ইত্যাদি।

প্রশ্ন হতে পারে, তাহলে কি সবসময় আমাদের ডেটার ফ্রিকুয়েন্সি দিয়েই কি পাই চার্ট, বার চার্ট বানাতে হয়?

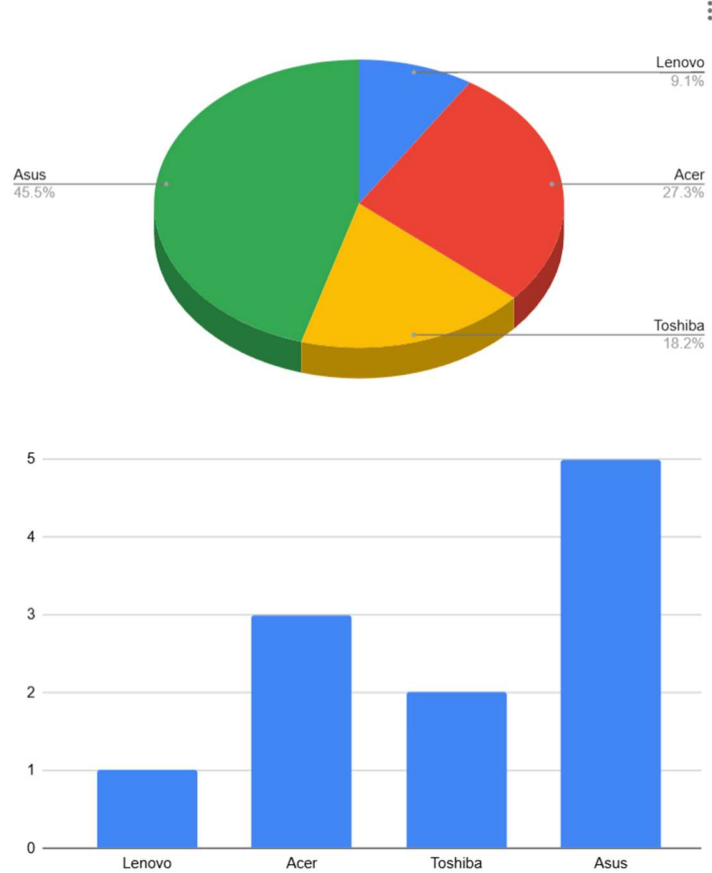
আসলে বিষয়টা সেরকম নয়। যেমনঃ

কোম্পানি	ল্যাপটপ উৎপাদন
Lenovo	1 বিলিয়ন
Acer	3 বিলিয়ন
Toshiba	2 বিলিয়ন
Asus	5 বিলিয়ন

এখানে আমরা পুনরাবৃত্তি করে উপাত্ত লিখছি না, এবং “ল্যাপটপ উৎপাদন” কে ফ্রিকুয়েন্সি বলছি না।

যদিও কথার কথা ধরলাম যে, বিলিয়ন এখানে আনুপাতিক হারে রেখে দিলাম কিন্তু আমরা এভাবে Lenovo, Acer, Acer, Acer, Toshiba, Asus লেখছি না।

কিন্তু এরপরেও এর পাই চার্ট, বার চার্ট বানানো সম্ভব। নিচেই দেখুনঃ (Courtesy: This visualization was generated using Google Sheets.)



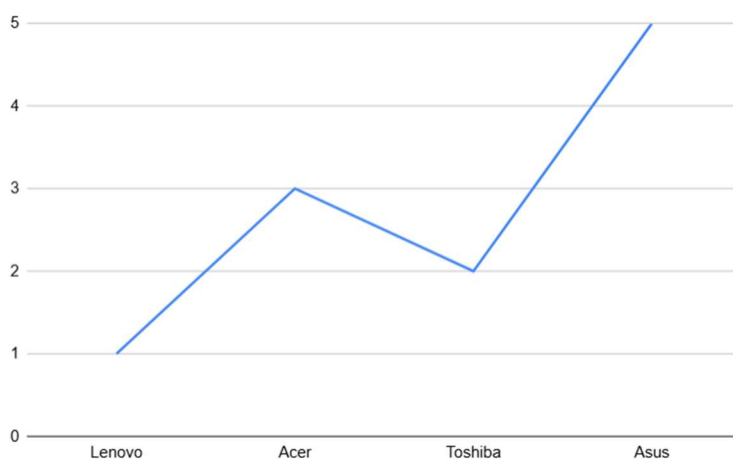
এর মানে এটাই বুঝায় যে, ফ্রিকুয়েন্সির কথা উল্লেখ না থাকলেও ডেটা ভিজুয়ালাইজেশন সম্ভব।

এবার আপনার প্রশ্ন জাগতে পারে, রিলেটিভ ফ্রিকুয়েন্সি ছাড়া কিভাবে পাই চার্ট বানানো হলো?

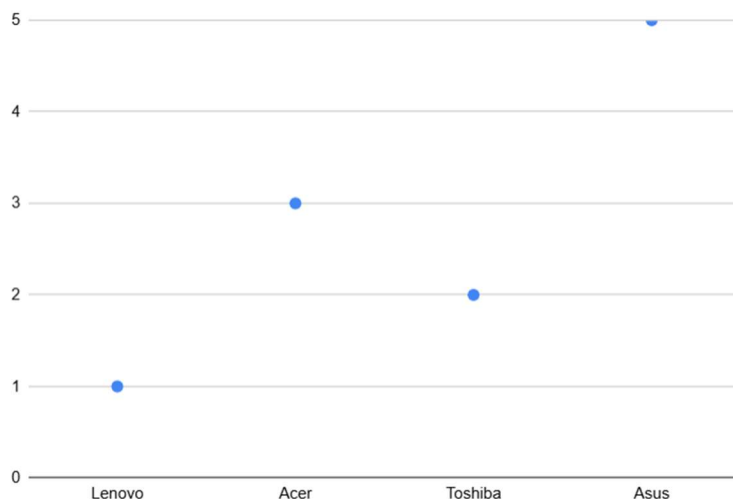
এক্ষেত্রে আমরা “ল্যাপটপ উৎপাদন” কে ফ্রিকুয়েন্সির মত কাজে লাগিয়েছি। Total ল্যাপটপ উৎপাদন হলো 11 বিলিয়ন। এরপর আমরা “রিলেটিভ ল্যাপটপ উৎপাদন” বের করেছি “ল্যাপটপ উৎপাদন” এর সাথে “Total ল্যাপটপ উৎপাদন” দিয়ে ভাগ করে। সেই “রিলেটিভ ল্যাপটপ উৎপাদন” কে আমরা ডিগ্রি বা পার্সেন্টেজ এ কনভার্ট করতে পারি।

পাই চার্ট, বার চার্ট বাদেও আরও কয়েকটি চার্ট (ডেটা ভিজুয়ালাইজেশন) দেখানো হলোঃ

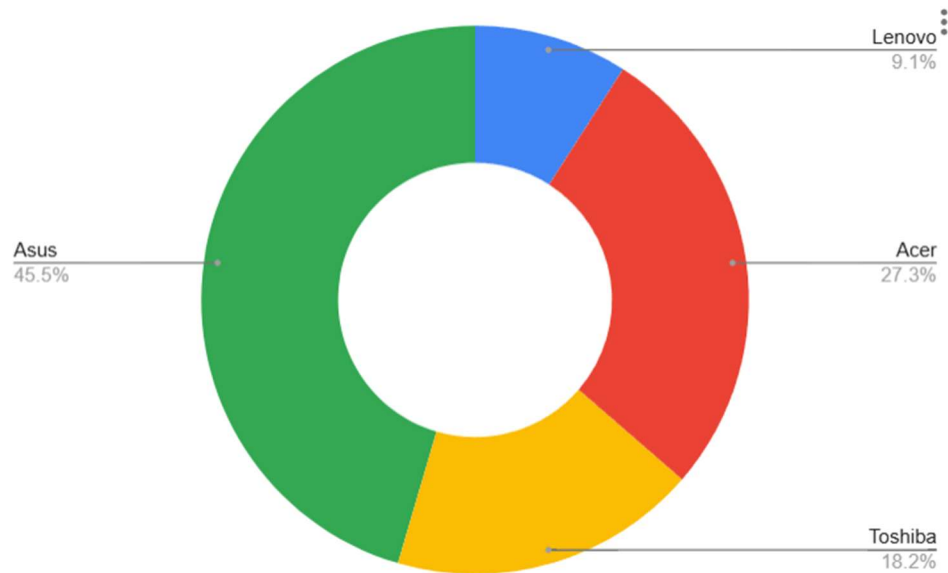
লাইন চার্ট (Line Chart): (Courtesy: This visualization was generated using Google Sheets.)



স্ক্যাটার্ড প্লট (Scatter Plot): (Courtesy: This visualization was generated using Google Sheets.)



ডোনাট চার্ট (Doughnut Chart): (Courtesy: This visualization was generated using Google Sheets.)



তো, এই পর্যন্ত আমরা আজ শিখেছি কিভাবে Categorical data এর ফ্রিকুয়েন্সি বানাতে হয়, কিভাবে পাই চার্ট, বার চার্ট বানাতে হয়। কিভাবে Categorical data তে মধ্যক (Median), প্রচুরক (Mode) বের করতে হয়। ধন্যবাদ।