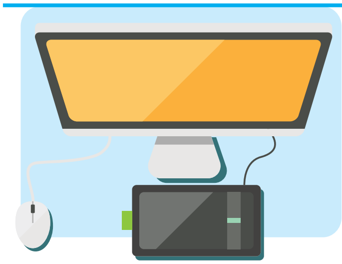


یونٹ

# 5 ڈیٹا سائنس کا تعارف

**حاصلاتِ تعلم:** اس باب کے اختتام تک، آپ اس قابل ہو جائیں گے کہ:

- ڈیٹا سائنس کے تصور اور روز مرہ زندگی میں اس کی اہمیت کی وضاحت کر سکیں۔
- ڈیٹا سائنس لائف سائیکل (DSLC) کے ہر مرحلہ کو بیان کر سکیں۔
- حقیقی دنیا کے ایسے مسائل کی شناخت اور تعریف کر سکیں جنہیں ڈیٹا کا استعمال کرتے ہوئے حل کیا جا سکتا ہے۔
- سروے، آن لائن فارمز، اور مشاہداتی تکنیک کا استعمال کرتے ہوئے ڈیٹا اکٹھا کر سکیں۔
- درستگی (accuracy) کو بہتر بنانے کے لیے ڈیٹا کلیننگ اور تصدیق (validation) کا عمل انجام دے سکیں۔
- ڈیٹا بیس سے ڈیٹا محفوظ کرنے اور بازیافت کرنے کے لیے سادہ SQL سوالات (queries) (جیسے SELECT، INSERT) کا استعمال کر سکیں۔
- بنیادی الگورتھمز جیسے ترتیب (sorting)، درجہ بندی (classification)، کلسٹرنگ کا استعمال کرتے ہوئے ڈیٹا کا تجزیہ کر سکیں۔
- ڈیٹا سے حاصل ہونے والے بصیرت (insights) کو بصری (visual) طور پر ظاہر کرنے کے لیے بار چارٹس اور ڈیش بورڈز بنا سکیں۔
- چارٹس اور سمریز کا استعمال کرتے ہوئے نتائج کو واضح طور پر بیان کر سکیں۔

## تعارف

ڈیٹا سائنس ڈیٹا کا استعمال کرتے ہوئے مسائل کو سمجھنے، نمونوں (patterns) کو دریافت کرنے اور دانشمندانہ فیصلے (smart decisions) کرنے کا عمل ہے۔ موجودہ ڈیجیٹل دور میں، ڈیٹا ہر جگہ موجود ہے، جیسے سروے، موبائل ایپس، سوشل میڈیا، آن لائن شاپنگ، اور صحت کی خدمات۔ ڈیٹا سائنس ہمیں اس خام ڈیٹا کو مفید معلومات میں بدلنے میں مدد کرتی ہے۔

چاہے موسم کی پیش گوئی کرنا ہو، طلباء کی دلچسپیوں کو سمجھنا ہو، یا صحت کی دیکھ بھال کو بہتر بنانا ہو، ڈیٹا سائنس ہمارے رہنے، سیکھنے اور کام کرنے کے طریقوں کو تبدیل کر رہی ہے۔ دنیا بھر کی تنظیمیں فیصلے سازی کو بہتر بنانے، اخراجات بچانے اور بہتر خدمات تخلیق کرنے کے لیے ڈیٹا کا استعمال کرتی ہیں۔

یہ باب ڈیٹا سائنس لائف سائیکل (DSLC) کا تعارف کراتا ہے، جس میں مسئلہ کو سمجھنا، ڈیٹا اکٹھا کرنا اور کلین (cleaning) کرنا، ڈیٹا کا تجزیہ اور تشریح کرنا، اور آخر میں چارٹس اور بصری (visual) معاونتوں کا استعمال کرتے ہوئے نتائج فراہم کرنا شامل ہے۔ ہر قدم تنقیدی سوچ اور مسئلہ حل کرنے کی مہارتوں کو فروغ دیتا ہے۔

ہم سیکھیں گے کہ سادہ ٹولز جیسے سروے، سپریڈ شیٹس، اور چارٹس کا استعمال کرتے ہوئے اپنا ڈیٹا کیسے اکٹھا کیا جاتا ہے۔ اس باب میں حقیقی دنیا

Web version of PECTA Textbook Not for Sale


NOT FOR SALE-PESRP

کی مثالیں اور عملی سرگرمیاں شامل ہیں جو انہیں کلاس روم اور روزمرہ زندگی میں ڈیٹا سائنس کا اطلاق کرنے میں مدد دیں گی۔

## 5.1 ڈیٹا سائنس کا تعارف

ہم ایک ڈیٹا سے بھری دنیا میں رہ رہے ہیں۔ چاہے ہم موبائل فون استعمال کر رہے ہوں، گوگل پر تلاش کر رہے ہوں، آن لائن شاپنگ کر رہے ہوں، یا یہاں تک کہ فٹنس ٹریکر کے ساتھ چل رہے ہوں، ڈیٹا اکٹھا کیا جا رہا ہے۔ سوال یہ پیدا ہوتا ہے کہ یہ ڈیٹا کس طرح مفید ہے؟

ڈیٹا سائنس وہ شعبہ ہے جہاں ہم نمونے (Pattern) تلاش کرنے، سوالات کے جواب دینے، اور لوگوں یا کاروباری اداروں کو بہتر فیصلے کرنے میں مدد کے لیے ڈیٹا کا استعمال کرتے ہیں۔ اس میں ڈیٹا کو اکٹھا کرنا، منظم کرنا، تجزیہ کرنا، اور ویژولائیزیشن (visualizing) شامل ہے۔

### فیصلہ سازی میں ڈیٹا کی اہمیت

ڈیٹا لوگوں اور تنظیموں کو بہتر فیصلے کرنے میں بہت بڑا کردار ادا کرتا ہے۔ ڈیٹا کے بغیر، فیصلے اندازوں یا ذاتی آراء پر مبنی ہو سکتے ہیں جو غلطیوں کا باعث بن سکتے ہیں۔ مثال کے طور پر، اگر سکول انتظامیہ کوئی نیا سپورٹس کلب شروع کرنا چاہتی ہے، تو وہ پہلے اس بارے میں ڈیٹا اکٹھا کر سکتے ہیں کہ طلباء کون سے کھیل سب سے زیادہ پسند کرتے ہیں۔ یہ انہیں ایک سمارٹ اور منصفانہ انتخاب کرنے میں مدد کرتا ہے۔

### ڈیٹا سائنس کے حقیقی زندگی میں اطلاقات

ڈیٹا سائنس ہماری روزمرہ کی زندگی کے کئی پہلوؤں میں اکثر ہمارے علم میں آئے بغیر استعمال ہوتی ہے،:

- **سوشل میڈیا:** فیس بک، یوٹیوب اور انسٹاگرام جیسی ایپلی کیشنز ڈیٹا کا استعمال کرتے ہوئے ہمیں ایسی پوسٹس دکھاتی ہیں جو ہمیں پسند آ سکتی ہیں۔
- **آن لائن شاپنگ:** ای-کامرس ویب سائٹس ہماری پچھلی خریداریوں کے ڈیٹا کا استعمال کرتے ہوئے مصنوعات تجویز کرتی ہیں۔
- **صحت کی دیکھ بھال:** ہسپتال بیماریوں کی پیش گوئی کرنے اور مریضوں کی دیکھ بھال کو زیادہ مؤثر طریقے سے منظم کرنے کے لیے ڈیٹا سائنس کا استعمال کرتے ہیں۔
- **نقل و حمل (Transport):** گوگل میپس جیسی ایپلی کیشنز ٹریفک ڈیٹا کا استعمال کرتے ہوئے تیز ترین راستوں کی تجویز دیتی ہیں۔
- **تعلیم:** سکولوں میں، طلباء کی حاضری اور ٹیسٹ کے نتائج سے حاصل کردہ ڈیٹا کو تدریسی حکمت عملیوں کو بہتر بنانے اور پیش رفت کو ٹریک کرنے کے لیے استعمال کیا جاتا ہے۔
- **ڈیٹا سائنس:** یوٹیوب ویڈیوز دیکھنے سے لے کر صحت کے مشورے حاصل کرنے یا بہترین راستہ منتخب کرنے تک، ہماری زندگی کو زیادہ سمارٹ (smart)، آسان اور زیادہ شخصی نوعیت (personalized) کا بنانے میں مدد کرتی ہے۔

Web version of PECTAA Textbook Not for Sale

## 5.2 ڈیٹاسائنس لائف سائیکل (DSLC) کا ایک جائزہ

ڈیٹاسائنس صرف ایک مرحلہ میں مکمل نہیں ہوتی۔ یہ مسائل کو حل کرنے میں مددگار اقدامات کا ایک سلسلہ ہے، جسے ہم ڈیٹاسائنس لائف سائیکل (DSLC) کہتے ہیں۔ ڈیٹاسائنس لائف سائیکل (DSLC) عام طور پر درج ذیل مراحل پر مشتمل ہوتا ہے:

ڈیٹاسائنس لائف سائیکل (DSLC) عام طور پر درج ذیل چھ (6) مراحل پر مشتمل ہوتا ہے:

- **مرحلہ 1:** مسئلہ کو سمجھنا (Understanding the Problem)
- **مرحلہ 2:** ڈیٹا اکٹھا کرنا (Collecting Data)
- **مرحلہ 3:** ڈیٹا کو کلین کرنا اور تصدیق کرنا (Cleaning and Validating Data)
- **مرحلہ 4:** ڈیٹا کا تجزیہ کرنا (Analyzing Data)
- **مرحلہ 5:** ڈیٹا کی تشریح کرنا (Interpreting Data)
- **مرحلہ 6:** نتائج کو بصری شکل دینا اور ابلاغ کرنا / ویژولائزنگ اور کمیونیکیٹنگ رزلٹ

Web version of PECTAA Textbook Not for Sale

### مسئلہ کو سمجھنا (Understanding the Problem)

اس سے پہلے کہ ہم ڈیٹا اکٹھا کرنے کے بارے میں بات کرنا شروع کریں، زیر غور مسئلہ کو واضح طور پر سمجھنا ضروری ہے۔ مسئلہ کو سمجھنا DSLC میں سب سے اہم مرحلہ ہے۔ اگر ہم مسئلہ کو صحیح طریقے سے نہیں سمجھتے، تو ہم غلط ڈیٹا اکٹھا کر سکتے ہیں جس کے نتیجے میں غیر مفید یا غلط نتائج برآمد ہوں گے۔

اس ابتدائی مرحلے میں، ہمیں کچھ سوالات (کیا، کیوں، کون) پوچھنے کی ضرورت ہوتی ہے، جیسے کہ:

- ہم کیا معلوم کرنے کی کوشش کر رہے ہیں؟
- یہ مسئلہ کیوں اہم ہے؟
- ان نتائج کو کون استعمال کرے گا؟

مندرجہ ذیل ذیلی حصے DSLC کے پہلے مرحلے کا مزید تفصیل سے احاطہ کرتے ہیں۔

### مسئلہ کی شناخت اور تعریف کرنا

سب سے پہلے، ہمیں مسئلے کی شناخت کرنے کی ضرورت ہے جسے ہم ڈیٹا کا استعمال کرتے ہوئے حل کرنا چاہتے ہیں۔ یہ مسئلہ ہمارے اسکول، گھر، شہر، یا زندگی کے کسی بھی شعبے سے متعلق ہو سکتا ہے جہاں معلومات کی بنیاد پر فیصلے کیے جاتے ہیں۔ ایک بار جب کسی مسئلے کی شناخت ہو جائے، تو ہمیں اسے واضح اور سادہ طریقے سے بیان کرنا چاہیے۔


NOT FOR SALE-PESRP

مثال: صرف یہ کہنے کے بجائے کہ، "طالب علم خوش نہیں ہیں،" ہم مسئلہ کو بہتر طریقے سے یوں بیان کر سکتے ہیں:

"طالب علموں کو اسکول کی کون سی سرگرمیوں میں سب سے زیادہ لطف آتا ہے؟"

کسی مسئلے کی تعریف کرتے وقت، ہمیں مندرجہ ذیل باتوں کو مدنظر رکھنے کی ضرورت ہے:

(i واضح اور سادہ الفاظ کا استعمال کریں،

(ii ایک مخصوص سوال پر توجہ مرکوز کریں، اور

(iii غیر واضح یا عام شرائط سے گریز کریں۔

**کیا آپ جانتے ہیں؟**

ایک اچھا مسئلہ وہ ہے جس کا جواب ڈیٹا اکٹھا کرنے اور تجزیہ کرنے سے دیا جا سکے۔

## واضح مقاصد کا تعین کرنا

ایک بار جب ہم مسئلے کی شناخت اور تعین کر لیتے ہیں، تو ہمیں یہ فیصلہ کرنا چاہیے کہ ہم کیا حاصل کرنا چاہتے ہیں۔ انہیں ہمارے مقاصد (Objectives) کہا جاتا ہے جو ڈیٹا سائنس کے باقی عمل میں رہنمائی کرتے ہیں۔ وہ بتاتے ہیں کہ:

I) ہمیں کس قسم کے ڈیٹا کی ضرورت ہے؟

(ii) ہم کیا سوالات پوچھیں گے؟

(iii) ہمیں کن نتائج کی تلاش ہے؟

مثال: اوپر بیان کردہ مسئلہ "طالب علموں کو اسکول کی کون سی سرگرمیوں میں سب سے زیادہ لطف آتا ہے؟" کے لیے، ہم اپنے مقاصد کو درج ذیل طریقے سے تشکیل دے سکتے ہیں:

- معلوم کریں کہ ہر سرگرمی میں کتنے طالب علم حصہ لیتے ہیں؟
- معلوم کریں کہ کون سی سرگرمی کو طالب علموں کی طرف سے سب سے زیادہ دلچسپ قرار دیا جاتا ہے؟
- سفارش کریں کہ کن سرگرمیوں کو بہتر یا تبدیل کیا جانا چاہیے؟

**کیا آپ جانتے ہیں؟**

اپنے مقاصد کو ایک چیک لسٹ کی طرح لکھیں۔ یہ آپ کے پراجیکٹ کو مرکوز رکھنے میں مدد کرتا ہے۔

### مثالیں: اسکول، صحت، کاروبار سے متعلق مسائل اور مقاصد

یہ ذیلی سیکشن تصور کو بہتر طریقے سے سمجھنے کے لیے کچھ حقیقت پسندانہ مثالوں کا احاطہ کرتا ہے۔ جدول 5.1 مختلف ڈومینز جیسے اسکول، صحت اور کاروبار سے متعلق مثال کے مسائل اور ان کے مقاصد فراہم کرتا ہے۔

### مثال: مسائل اور مقاصد

یہ جدول مختلف شعبوں میں ڈیٹا سائنس کے اطلاق کے لیے حقیقی مسائل کی مثالیں اور ان سے متعلقہ واضح مقاصد کی وضاحت کرتا ہے۔

Web version of PECTAA Textbook Not for Sale

جدول:5.1

| ڈومین | مثالی مسئلہ | مقاصد |
|---|---|---|
| سکول | بہت سے طالب علم اسکول دیر سے پہنچتے ہیں۔ | ایک مہینے میں کتنے طالب علم دیر سے پہنچتے ہیں، اس کا پتہ لگانا۔ |
| | | دیر سے آنے کی وجوہات کی شناخت کرنا (مثلاً، ٹرانسپورٹ، فاصلہ، ہنگامی صورتحال، یا موسمی مسائل)۔ |
| | | دیر سے آنے والے کیسز کو کم کرنے کے لیے حل تجویز کرنا۔ |
| صحت | سکول کے اوقات کے دوران بہت سے طالب علم کافی زیادہ پانی نہیں پیتے ہیں۔ | طالب علم روزانہ کتنا پانی پیتے ہیں، اس پر ڈیٹا اکٹھا کرنا۔ |
| | | پانی کی کم مقدار پینے کی وجوہات معلوم کرنا۔ |
| | | صحت مند عادات کو فروغ دینے کے لیے اقدامات کی سفارش کرنا۔ |
| کاروبار | سکول کی کینٹین یہ جاننا چاہتی ہے کہ طالب علم کون سے سنیکس سب سے زیادہ پسند کرتے ہیں۔ | طالب علموں کے پسندیدہ اسنیکس کے بارے میں سروے کرنا۔ |
| | | مقبولیت اور فروخت کے ڈیٹا کا موازنہ کرنا۔ |
| | | کینٹین مینیجر کو سمارٹ آرڈرنگ کے فیصلے کرنے میں مدد فراہم کرنا۔ |

Web version of PECTAA Textbook Not for Sale


NOT FOR SALE-PESRP

**یادرکھیں:**

1. کسی بھی ڈیٹا سائنس پراجیکٹ کا پہلا مرحلہ مسئلہ کو سمجھنا ہے۔
2. ہمیں مسئلہ کو واضح طور پر بیان کرنا چاہیے اور مخصوص اہداف کا تعین کرنا چاہیے۔
3. کلیئر (Clear) مسائل وہ ہوتے ہیں جنہیں ڈیٹا کا استعمال کر کے حل کیا جا سکتا ہے۔
4. واضح مقاصد صحیح ڈیٹا اکٹھا کرنے اور مفید نتائج تک پہنچنے میں مدد کرتے ہیں۔

## ڈیٹا اکٹھا کرنا

ایک بار جب ہم مسئلہ کو سمجھ لیتے ہیں اور واضح اہداف مقرر کر لیتے ہیں، تو DSLC میں اگلا مرحلہ مطلوبہ ڈیٹا اکٹھا کرنا ہے۔ ڈیٹا خام مال (raw material) کی طرح ہے، اور ہمیں اسے کلین کرنے، ذخیرہ کرنے یا تجزیہ کرنے سے پہلے صحیح قسم کا ڈیٹا جمع کرنا چاہیے۔

مرحلہ 1 : ہم نے سوالات پوچھے جیسے "طالب علموں کو اسکول کی کون سی سرگرمیوں میں سب سے زیادہ لطف آتا ہے؟" یا "طالب علم سکول دیر سے کیوں آرہے ہیں؟"

مرحلہ 2: ان سوالات کے جواب دینے میں مدد فراہم کرنے والی معلومات اکٹھا کرنے کے بارے میں ہے۔

ہم مندرجہ ذیل چیزوں کا جائزہ لیں گے:

i. ڈیٹا اکٹھا کرنے کے طریقے

ii. ڈیٹا کہاں سے اکٹھا کیا جائے (ذرائع)؟

iii. یہ کیسے یقینی بنایا جائے کہ یہ ڈیٹا مفید اور منصفانہ ہے (ڈیٹا کا معیار)؟

### (i) ڈیٹا اکٹھا کرنے کے طریقے

مسئلے اور دستیاب ٹولز پر منحصر، ڈیٹا اکٹھا کرنے کے کئی طریقے ہیں۔ آیئے اسکولوں، گھروں اور کاروباروں میں استعمال ہونے والے سب سے عام طریقوں پر بات کرتے ہیں۔

**سروے (آن لائن / پیپر بیسڈ):** سروے سب سے آسان اور عام طریقہ ہے۔ ہم لوگوں سے سوالات پوچھتے ہیں اور ان کے جوابات ریکارڈ کرتے ہیں۔

**مثال:** یہ جاننے کے لیے کہ طالب علم کن کھیلوں سے لطف اندوز ہوتے ہیں، ہم سوالات کے ساتھ ایک سروے بنا سکتے ہیں۔ جیسے

I) آپ کو کون سا کھیل سب سے زیادہ پسند ہے؟

ii) آپ اسے کتنی بار کھیلتے ہیں؟

iii) آپ دن کے کس وقت کھیلنا پسند کرتے ہیں؟

Web version of PECTAA Textbook Not for Sale



NOT FOR SALE-PESRP

## کیا آپ جانتے ہیں؟

سروے گوگل فارمز (آن لائن) اور پرنٹ شدہ کاغذی فارمز (مینؤل) کا استعمال کرتے ہوئے کیے جاسکتے ہیں۔

- APIs اور آن لائن ذرائع: ایپلیکیشن پروگرامنگ انٹرفیس (APIs) سافٹ ویئر کو ویب سائٹس یا ایپلی کیشنز (مختصراً ایپس) سے خودکار طور پر ڈیٹا اکٹھا کرنے کی اجازت دیتے ہیں۔
مثال: ایک اسکول روزانہ کی حاضری کا ڈیٹا کی رپورٹ بنانے کے لیے حاضری کا API استعمال کر سکتا ہے (یہ ڈویلپرز زیادہ استعمال کرتے ہیں)۔
- IoT اور سینسر ڈیٹا: انٹرنیٹ آف تھنگز (IoT) ڈیوائسز، جیسے سمارٹ واچز، واٹر ڈسپنسر، یا درجہ حرارت کے سینسر، خود بخود ڈیٹا اکٹھا کر سکتے ہیں۔
مثال: پانی پینے کی عادات کا مطالعہ کرنے کے لیے، سکول پانی کے فواروں (پوائنٹس) کے قریب سمارٹ واٹر میٹر لگا سکتے ہیں۔

### ویب سکریپنگ (Web Scraping):

**SMART TIPS** سمارٹ ٹپس

IoT ڈیوائسز صحت، حفاظت، اور ماحولیاتی نگرانی میں مفید ہیں۔

یہ ویب سائٹس سے عوامی ڈیٹا کو خودکار طور پر اکٹھا کرنے کا ایک طریقہ ہے۔

مثال: ایک کاروباری طالب علم سکول کی کینٹین کی قیمتوں سے موازنہ کرنے کے لیے آن لائن اسٹورز سے سنیکس کی قیمتیں اکٹھی کر سکتا ہے۔

**SMART TIPS** سمارٹ ٹپس

ویب سکریپنگ صرف اساتذہ کی رہنمائی میں اور ان ویب سائٹس کے لیے کی جانی چاہیے جو اس کی اجازت دیتی ہیں۔

### (ii) درست ڈیٹا سورس کا انتخاب

ہر قسم کا ڈیٹا مفید نہیں ہوتا۔ ہمیں قابل اعتماد، درست اور متعلقہ ذرائع کا انتخاب کرنا چاہیے۔

ڈیٹا سورس کا انتخاب کرتے وقت، چند سوالات پوچھیں جیسے:

i) کیا یہ ڈیٹا ایک بھروسہ مند جگہ سے ہے؟ ii) کیا یہ حالیہ ہے یا پرانا؟

iii) کیا یہ اس مسئلے سے مطابقت رکھتا ہے جسے میں حل کرنا چاہتا ہوں؟

**مثال:** اگر ہم 2025 میں طالب علموں کے پسندیدہ سنیکس جاننا چاہتے ہیں، تو ہمیں 2020 میں کیے گئے سروے کا استعمال کرنے سے گریز کرنا چاہیے۔

Web version of PECTAA Textbook Not for Sale

## iii) ڈیٹا کے معیار اور جانبداری کا جائزہ لینا Assessing Data Quality and Bias

ڈیٹا کا معیار کا مطلب ہے کہ ڈیٹا

(i مکمل ہے (کوئی غیر موجود ڈیٹا نہیں)،

(ii درست ہے (کوئی غلطیاں نہیں)،

(iii متعلقہ ہے (ہمارے مسئلے سے مطابقت رکھتا ہے)۔

**SMART TIPS سمارٹ ٹپس**

ہمیشہ جدید اور مسئلے سے متعلق ڈیٹا اکٹھا کریں۔

جانبداری (Bias) کا مطلب ہے کہ ڈیٹا صرف حقیقت کا ایک حصہ دکھاتا ہے اور پوری تصویر کو نظر انداز کر دیتا ہے۔

مثال: اگر ہم صرف لڑکوں کا ان کے پسندیدہ کھیلوں کے بارے میں سروے کرتے ہیں، تو ہم لڑکیوں کی رائے سے محروم ہو جائیں گے۔ نتائج غیر منصفانہ یا لڑکوں کی طرف جانبدار ہوں گے۔

جانبداری سے بچنے کے لیے درج ذیل باتوں کو یاد رکھیں:

- لوگوں کے مختلف گروپس سے پوچھیں۔
- سوالات کو غیر جانبدار رکھیں اور رہنمائی کرنے والے (لیڈنگ) نہ ہوں۔
- جوابات میں عجیب و غریب پیٹرنز کی جانچ کریں۔

Web version of PECTAA Textbook Not for Sale

### سرگرمی:

**ایک سادہ سروے ڈیزائن کرنا**

**مقصد:** اپنے ہم جماعتوں کی فارغ وقت کی عادات کے بارے میں ایک مختصر سروے بنا کر ڈیٹا اکٹھا کرنے اور ذخیرہ کرنے کا طریقہ سیکھیں۔

**ہدایات:**

1. **اپنا سروے ڈیزائن کریں:** اپنے ہم جماعتوں سے پوچھنے کے لیے 3 سے 5 سوالات سوچیں۔
2. **جوابات اکٹھا کرنے کی تیاری کریں:** ایک طریقہ کا انتخاب کریں (مثلاً، کاغذی فارم یا گوگل فارم)۔
3. **ڈیٹا اکٹھا کریں:** کم از کم 5 ہم جماعتوں سے اپنا سروے مکمل کرنے کو کہیں اور ان کے جوابات کو واضح طور پر لکھیں۔
4. **ڈیٹا کو ذخیرہ کریں:** (i) اگر آپ کاغذ استعمال کر رہے ہیں، تو اسے صاف ستھرا اور کالموں میں منظم رکھیں۔
   (ii) اگر ڈیجیٹل ٹولز استعمال کر رہے ہیں، تو فائل کو ایک واضح نام جیسے FreeTime_your name.xlsx سے محفوظ کریں۔
5. (اختیاری طور پر) اپنے استاد یا ہم جماعتوں کو دکھائیں کہ آپ نے ڈیٹا کیسے اکٹھا اور ذخیرہ کیا۔

### MEMORY BOX یاد رکھیں:

1. کسی مسئلے کی تعریف کرنے کے بعد، ہمیں وہ ڈیٹا اکٹھا کرنے کی ضرورت ہے جو مسئلہ حل کرنے میں مدد کر سکے۔
2. ڈیٹا اکٹھا کرنے کے مختلف طریقے ہیں جیسے سروے، APIs، سینسرز، اور ویب سکریپنگ۔
3. اگلے مرحلے پر جانے سے پہلے ہمیں ہمیشہ ڈیٹا کے معیار اور غیر جانبداری جانچنا چاہیے۔
4. وہ ڈیٹا جو نامکمل یا متعصب ہو، غلط فیصلوں کا باعث بن سکتا ہے۔



NOT FOR SALE-PESRP

## ڈیٹا کلیننگ اور تصدیق (ویلیڈیٹنگ)

ڈیٹا اکٹھا کرنے کے بعد، اگلا اہم مرحلہ اسے کلین کرنا اور تصدیق کرنا ہے۔ خام ڈیٹا بے ترتیب ہوتا ہے۔ اس میں غلطیاں، غائب ویلیوز، یا یہاں تک کہ غلط اندراجات بھی شامل ہو سکتے ہیں۔ اگر ہم اسے جانچے بغیر استعمال کرتے ہیں، تو ہمارا تجزیہ غلط نتائج دے گا۔

یاد رہے کہ اس سے پہلے ہم نے سروے اور دیگر طریقوں سے ڈیٹا اکٹھا کیا تھا۔ اب، ہمیں یہ جانچنے کی ضرورت ہے کہ آیا وہ ڈیٹا درست، مکمل اور استعمال کے لیے تیار ہے۔ یہ مرحلہ یقینی بناتا ہے کہ جس ڈیٹا کا ہم بعد میں تجزیہ کریں گے، وہ قابلِ اعتماد ہے۔

### خام ڈیٹا (Raw Data)

خام ڈیٹا اصل معلومات ہے جو ہم لوگوں، سینسرز، ویب سائٹس، یا ایپس سے اکٹھا کرتے ہیں۔ اسے ابھی تک جانچا یا کلین نہیں کیا گیا ہے۔

مثال کے طور پر، فرض کریں کہ 20 طالب علموں نے آپ کے پسندیدہ کھیلوں کے سروے کو پُر کیا۔ کچھ اندراجات جدول میں دکھائے گئے ہیں:

جدول 5.2: طالب علموں کے پسندیدہ کھیل سے متعلق خام ڈیٹا

| ایک دن میں پڑھنے کا وقت | فریکوئنسی (Frequency) | پسندیدہ کھیل | نام | سیریل نمبر |
|---|---|---|---|---|
| Evening | Daily | cricket | Ahmad | 1 |
| Afternoon | Weekly | | Ali | 2 |
| | Daily | Footbal | Sara | 3 |
| Evening | Daliy | Football | Sana | 4 |

جدول 5.2 میں مندرجہ ذیل چیزیں شامل ہیں:

(i) سپیلنگ کی غلطیاں (Spelling Errors): مثلاً، Footbal سیریل نمبر 3 اور Daliy سیریل نمبر 4۔

(ii) (Missing Entries): میسنگ اینٹریز سیریل نمبر 2 اور 3 میں خالی خانے موجود ہیں۔

مکسڈ کیپیٹلائزیشن (Mixed Capitalizations): مثلاً، cricket سیریل نمبر 1 اور Football سیریل نمبر 4۔ اس طرح کے ڈیٹا کو کلیننگ (Cleaning) کی ضرورت ہوتی ہے۔

### عمومی ڈیٹا ایررز (غائب، نقل ڈپلیکیٹ، غیر معمولی ویلیوز)

جدول 5.3 خام ڈیٹا میں پائے جانے والے کچھ عام مسائل کو ظاہر کرتا ہے۔

Web version of PECTAA Textbook Not for Sale

جدول:5.3 خام ڈیٹا میں عام مسائل

| سیریل نمبر | غلطی کی قسم | مثال |
|---|---|---|
| 1 | مِسنگ ڈیٹا | خالی خانے (مثلاً، کھیل کا نام یا وقت درج نہیں)۔ |
| 2 | ڈوپلیکیٹ اندراجات | ایک ہی طالب علم کو دو بار درج کیا جانا۔ |
| 3 | غیر مستحکم فارمیٹس (Inconsistent formats) | "CRICKET"،""cricket،Cricket"" (حروف کی کیپیٹلائزیشن میں فرق)۔ |
| 4 | ہجوں کی غلطیاں (Spelling mistakes) | "Football" کی بجائے "Footbal" یا "Daily" کی بجائے "Daliy"۔ |
| 5 | غیر معمولی ویلیوز (Outliers) (عجیب و غریب اقدار) | فریکئیونسی =1000 بار فی ہفتہ۔ |

Web version of PECTAA Textbook Not for Sale

## ڈیٹا کی تصدیق کی تکنیکیں (Data Validation Techniques)

ڈیٹا کی تصدیق کا مطلب یہ جانچنا ہے کہ آیا ڈیٹا:

i. درست قواعد کی پیروی کرتا ہے (مثلاً، صرف درست کھیلوں کے نام)،

ii. ویلیوز کی صحیح قسم رکھتا ہے (مثلاً، جہاں نمبرز کی ضرورت ہو وہاں نمبرز ہوں)، اور

iii. فیک یا غیر سنجیدہ (fake or silly) نہیں ہے۔

**مثال کے طور پر، ویلیڈیشن رولز:**

"پسندیدہ کھیل" کے لیے صرف ان ویلوز میں سے کسی ایک کی اجازت دیں: فٹ بال، کرکٹ، بیڈمنٹن، ہاکی۔

**سروے میں ویلیڈیشن:** اس مقصد کے لیے، مندرجہ ذیل باتوں پر غور کریں:

- خالی ٹیکسٹ باکسز کے بجائے کثیر الانتخابی سوالات کا استعمال کریں۔
- اہم سوالات کو لازمی قرار دیں، تاکہ کوئی بھی انہیں نظر انداز نہ کرے۔
- ہجوں کی غلطیوں کو کم کرنے کے لیے ڈراپ ڈاؤنز یا چیک باکسز استعمال کریں۔

**SMART TIPS**

سمارٹ ٹپس

گوگل شیٹس، ایکسل، یا آن لائن فارمز جیسے ٹولز میں مہیا کئے گئے سادہ قواعد بھی ڈیٹا کی تصدیق میں مدد کر سکتے ہیں۔

## ڈیٹا کلیننگ کے طریقے

ایک بار جب ہمیں غلطیاں مل جاتی ہیں، تو ہمیں انہیں درست کرنے یا ہٹانے کی ضرورت ہوتی ہے۔ اسے ڈیٹا کلیننگ کہتے ہیں۔
ڈیٹا کلین کرنے کے لیے چند بنیادی اقدامات درج ذیل ہیں:

- سپیلنگ کی غلطیاں درست کریں: "Footbal" → "Football"
- فارمیٹس کو معیاری بنائیں: "CRICKET" → "Cricket"
- نقل کو ہٹائیں (تکرار-Redundancy): ایک ہی طالب علم کے لیے بار بار آنے والے اندراجات کو حذف کر دیں۔
- گمشدہ اقدار کو ہینڈل کرنا: اس شخص سے دوبارہ پوچھیں یا اس قطار (row) کو ہٹا دیں۔
- غلط ڈیٹا کو درست کریں: اگر فریکیوئنسی (frequency) "دس سو" (ten hundred) کے طور پر لکھا گیا ہے، تو اسے درست کریں یا ہٹا دیں۔

**ڈیٹا کلیننگ کیوں ضروری ہے؟**

فرض کریں کہ ہم اس مرحلے کو چھوڑ دیتے ہیں اور ایسے ڈیٹا کا تجزیہ کرتے ہیں جس میں شامل ہیں:

- سپیلنگ کی غلطیوں والے کھیل ("Cricket")
- گمشدہ اوقات / سٹڈی ٹائم
- ڈوپلیکیٹ اندراجات

**ہمارا نتیجہ شاید کہے:**

"فٹ بال: 3 ووٹ، cricket: 1 ووٹ، Cricket: 2 ووٹ"

یہ غلط تصویر دیتا ہے، کیونکہ "cricket" اور "Cricket" ایک ہی کھیل ہیں!

خلاصہ: ہم کہہ سکتے ہیں کہ کلین ڈیٹا سمارٹ فیصلوں کی طرف لے جاتا ہے۔ ڈرٹی (Dirty) ڈیٹا الجھن کا باعث بنتا ہے۔

## ڈیٹا کا تجزیہ کرنا

اب جب کہ ہمارا ڈیٹا صحیح طریقے سے اکٹھا اور کلین ہو چکا ہے، ہم تجزیہ کے مرحلے کی طرف بڑھ سکتے ہیں۔ یہ وہ جگہ ہے جہاں ہم پیٹرن (patterns)، رجحانات (trends)، اور سوالات کے جوابات تلاش کرنے کے لیے ڈیٹا کا جائزہ لیتے ہیں جو ہم نے مرحلہ 1 میں پوچھے تھے۔
فرض کریں کہ آپ کے پاس ڈیٹا کا ایک کلین اور منظم جدول ہے، مثال کے طور پر، طالب علموں کے پسندیدہ کھیل یا پڑھائی کے اوقات کے بارے میں جوابات۔ اب آپ پوچھ سکتے ہیں، "یہ ڈیٹا ہمیں کیا بتا رہا ہے؟" یہ مرحلہ ہمیں ڈیٹا میں جو کچھ بھی ہم دریافت کرتے ہیں، اس کی بنیاد پر فیصلے کرنے میں مدد کرتا ہے۔

Web version of PECTAA Textbook Not for Sale

## ڈیٹا کا تجزیہ کرنے کا کیا مطلب ہے؟

ڈیٹا کا تجزیہ کرنے کا مطلب ہے احتیاط سے ڈیٹا کی جانچ کرنا تا کہ:

(i) اس کا خلاصہ کیا جائے (مثلاً، کتنے طالب علموں نے ہر آپشن کا انتخاب کیا)،

(ii) مختلف گروپوں کا موازنہ کیا جائے (مثلاً، لڑکے بمقابلہ لڑکیاں صبح بمقابلہ شام)،

(iii) نمونے تلاش کیے جائیں (مثلاً، زیادہ تر طالب علم فٹ بال کو ترجیح دیتے ہیں اور شام کو کھیلتے ہیں)۔

آپ شمار (counts-کتنے؟)، اوسط (averages)، فریکوینسی (frequency-کتنی بار؟) وغیرہ جیسے پیمانوں کا استعمال کر سکتے ہیں۔

### تجزیہ کے لیے سادہ ٹولز کا استعمال

آپ سادہ ٹولز جیسے گوگل شیٹس، مائیکروسافٹ ایکسل، یا نوٹ بک میں لکھتے ہوئے ڈیٹا کا تجزیہ شروع کر سکتے ہیں۔

مثال: آپ نے 20 طالب علموں کا سروے کیا۔ جدول 5.4 دکھاتا ہے کہ آپ کو کیا ملا ہو گا۔

اس ڈیٹا سے، ہم یہ مشاہدہ کر سکتے ہیں کہ زیادہ تر طالب علم کرکٹ کو ترجیح دیتے ہیں۔

جدول 5.4: طالب علم کے سروے سے ماخوذ ڈیٹا

| پسندیدہ کھیل | فریکیونسی |
|---|---|
| کرکٹ | 10 |
| فٹ بال | 5 |
| ہاکی | 4 |
| بیڈمنٹن | 1 |

Web version of PECTAA Textbook Not for Sale

### سرگرمی:

ڈیٹا کلیننگ

**مقصد:** ڈیٹا انٹری کی عام غلطیوں (جیسے ہجوں کی غلطیاں، Null اقدار، اور غیر مستحکم فارمیٹنگ) کی شناخت اور درستگی کا طریقہ سیکھنا۔

ہدایات:

**1. بے ترتیب جدول کا جائزہ:**

ذیل میں طالب علموں کے پسندیدہ کھیل سے متعلق خام ڈیٹا کا جدول دیا گیا ہے جس میں غلطیاں موجود ہیں۔

جدول 5.5: طالب علموں کے پسندیدہ کھیل سے متعلق خام ڈیٹا (بے ترتیب ڈیٹا)

| سیریل نمبر | نام | پسندیدہ کھیل | فریکیوئنسی (Frequency) | دن کا وقت |
|---|---|---|---|---|
| 1 | Sadiq | cricket | Daliy | Evening |
| 2 | Noor | Cricket | Weekly | |
| 3 | Maryam | Footbal | | morning |
| 4 | Zara | Football | Daliy | Afternoon |



NOT FOR SALE-PESRP

2. گروپس (pair) میں کام کریں
اپنے ساتھی کے ساتھ بیٹھیں اور جدول کا ایک ساتھ مطالعہ کریں۔

3. غلطیوں کی نشاندہی اور فہرست سازی:
کم از کم 5 غلطیاں تلاش کریں۔ ان میں شامل ہو سکتی ہیں:
- سپیلنگ کی غلطیاں: مثلاً، Daliy، یا ممکنہ طور پر Cricket۔
- NULL ڈیٹا: مثلاً، خالی خانے۔
- غیر مستحکم فارمیٹنگ: مثلاً، Football بمقابلہ football، یا morning کو کیپیٹلائزنہ کرنا۔

4۔ غلطیوں کی درستگی:
i) ہر غلطی کو منطقی اور منصفانہ طریقے سے ٹھیک کرنے پر تبادلہ خیال کریں۔
ii) کلین شدہ جدول کو اپنی نوٹ بک میں دوبارہ لکھیں۔

5. کلاس میں اشتراک (اختیاری)
I) اپنے کلین شدہ جدول کا دوسرے گروپ کے ساتھ موازنہ کریں۔
ii) دیکھیں کہ آیا کہ انہوں نے بھی وہی غلطیاں تلاش کیں یا کچھ مختلف غلطیاں۔

MEMORY BOX
یاد رکھیں:
1. ڈیٹا کا تجزیہ ہمیں نمونے تلاش کرنے اور فیصلے کرنے میں مدد کرتا ہے۔
2. ڈیٹا کو سمجھنے کے لیے ہم گنتی (counts) اور موازنہ (comparisons) کا استعمال کرتے ہیں۔
3. یہاں تک کہ سادہ ٹولز جیسے جدول (Table) یا اسپریڈ شیٹس بھی ڈیٹا کا تجزیہ کرنے میں ہماری مدد کر سکتے ہیں۔

سرگرمی:
ڈیٹا کا تجزیہ (Data Analysis)
مقصد: ٹرینڈ کی شناخت اور فریکوئنسی جدول سے بنیادی نتائج اخذ کرنے کے ذریعے ڈیٹا کا تجزیہ کرنے کا طریقہ سیکھنا۔
ہدایات:
اپنا ذاتی سروے ڈیٹا (اگر دستیاب ہو) یا ذیل میں جدول 5.6 میں دیا گیا نمونہ ڈیٹا استعمال کریں۔
احتیاط سے اس جدول کو پڑھیں جو ظاہر کرتا ہے کہ کتنے طالب علم ہر کھیل کو پسند کرتے ہیں۔
شناخت کریں:
i) سب سے زیادہ مقبول کھیل
ii) سب سے کم مقبول کھیل
ایک وجہ لکھیں کہ ہر کھیل آپ کی کلاس میں سب سے زیادہ یا سب سے کم مقبول کیوں ہو سکتا ہے۔
اپنے نتائج کو سادہ جملوں کا استعمال کرتے ہوئے اپنی نوٹ بک میں پیش کریں۔

جدول 5.6: سروے ڈیٹا سے ماخوذ نمونہ ڈیٹا

| کھیل | فریکوئنسی یا طلباء کی تعداد |
|---|---|
| کرکٹ | 7 |
| فٹ بال | 6 |
| ہاکی | 5 |
| بیڈمنٹن | 2 |

Web version of PECTAA Textbook Not for Sale


NOT FOR SALE-PESRP

## انٹرپرٹینگ ڈیٹا

اس مرحلے کا مقصد ڈیٹا سے معنی تلاش کرنا ہے (یعنی رجحانات اور پیٹرنز کی شناخت کرنا)۔ ڈیٹا کا خلاصہ کرنے کے بعد، ہم خود سے یہ سوال پوچھ سکتے ہیں:

- یہ نمبرز ہمیں کیا بتا رہے ہیں؟
- کچھ آپشنز زیادہ مقبول کیوں ہیں؟
- کیا کوئی حیرت انگیز بات یا کوئی واضح پیٹرن ہے؟

مثال: اگر ہم یہ معلوم کرنا چاہتے ہیں کہ دیر سے آنے والے زیادہ تر طالب علم اسکول سے دور رہتے ہیں، تو ہم ایک حل کے طور پر نقل و حمل (transport) کی مدد فراہم کرنے کی سفارش کر سکتے ہیں۔

## ویژولائزیشن اور ابلاغ کے نتائج

ڈیٹا کا تجزیہ کرنے کے بعد، آخری مرحلہ نتائج کو اس طرح سے شیئر کرنا ہے کہ دوسرے لوگ اسے آسانی سے سمجھ سکیں۔ لوگ عام طور پر نمبروں کے مقابلے میں تصاویر کو زیادہ بہتر سمجھتے ہیں، اسی لیے ہم ڈیٹا ویژولائزیشن کا استعمال کرتے ہیں۔

یاد رہے کہ پچھلے مرحلے میں، ہم نے پایا کہ کرکٹ سب سے زیادہ مقبول کھیل تھا۔ اس مرحلے میں، ہم ایک ایسا چارٹ بنائیں گے جو اسے واضح طور پر ظاہر کرے گا۔ یہ ہمارے نتائج کو ہم جماعتوں، اساتذہ یا دوسروں کے سامنے پیش کرنا آسان بنا دیتا ہے۔

## ڈیٹا ویژولائزیشن کی اہمیت

ڈیٹا ویژولائزیشن اہم ہے کیونکہ یہ:

i) ڈیٹا کو آسانی سے سمجھنے کے قابل بناتا ہے،

ii) پیٹرنز کو ایک نظر میں دکھاتا ہے،

iii) پریزنٹیشنز یا رپورٹس میں خیالات کو واضح طور پر شیئر کرنے میں مدد کرتا ہے۔

مثال: یہ کہنے کے بجائے کہ "10 طالب علم کرکٹ پسند کرتے ہیں،" ہم اسے بار چارٹ میں دکھا سکتے ہیں۔

## چارٹس اور گرافس کی عام اقسام (بار چارٹ، پائی چارٹ، لائن چارٹ وغیرہ)

ہمیشہ ایک ایسا چارٹ منتخب کریں جو آپ کے ڈیٹا سے مطابقت رکھتا ہو۔ جدول 5.7 عام طور پر استعمال ہونے والے چارٹس کی اقسام، ان کے استعمال کی صورتحال اور مثالیں دکھاتا ہے۔

Web version of PECTAA Textbook Not for Sale



NOT FOR SALE-PESRP

جدول 5.7: چارٹس کی عام اقسام

| چارٹ / گراف کی قسم | استعمال کا مقصد | مثال |
|---|---|---|
| بار چارٹ | گروپس کا موازنہ کرنا | ہر کھیل کو پسند کرنے والے طالب علموں کی تعداد۔ |
| پائی چارٹ | پوری چیز کے حصے دکھانا | پسندیدہ اسنیکس کا فیصد۔ |
| لائن چارٹ | وقت کے ساتھ تبدیلی دکھانا | ہر ہفتے دیر سے آنے والے طالب علموں کی تعداد۔ |

Web version of PECTAA Textbook Not for Sale

**SMART TIPS**

سمارٹ ٹپس

اپنے چارٹس کو سادہ، کلین اور درست طریقے سے لیبل کریں۔

## کمیونیکیٹنگ رزلٹ:

ایک بار جب آپ کا چارٹ تیار ہو جائے تو، اپنے نتائج کی وضاحت کریں۔ آپ 2 سے 3 سطریں لکھ سکتے ہیں:
مثال: "زیادہ تر طالب علم 10 کرکٹ پسند کرتے ہیں، جس کے بعد فٹ بال 6 آتا ہے۔ بہت کم طالب علم 1 بیڈ منٹن کو پسند کرتے ہیں۔"

**SMART TIPS**

سمارٹ ٹپس

ڈیٹا کی کہانی سنانے کے لیے بصری معاونتوں اور الفاظ کو ایک ساتھ استعمال کریں (واضح، مختصر اور مفید)۔

**MEMORY BOX**

یاد رکھیں:

1. ڈیٹا ویژولائیزیشن کا مطلب نتائج دکھانے کے لیے چارٹس اور گرافس کا استعمال کرنا ہے۔
2. یہ ڈیٹا کو سمجھنے اور شیئر کرنے میں آسان بناتا ہے۔
3. اپنے ڈیٹا کی قسم کے لیے صحیح چارٹ استعمال کریں۔
4. اپنی بصری معاونتوں کو ہمیشہ لیبل لگائیں اور وضاحت کریں۔



NOT FOR SALE-PESRP

## 5.3 ویژولائزیشن کے لیے ٹولز

جیسے ہی آپ جان لیں کہ کس قسم کا چارٹ استعمال کرنا ہے۔ اگلا مرحلہ اسے مختلف ٹولز کی مدد سے تیار کرنا ہوتا ہے۔
ہم نے سیکھا کہ کس طرح چارٹس ڈیٹا کو سمجھنے اور پیش کرنے میں آسانی پیدا کرتے ہیں۔ اس سیکشن میں، آپ سیکھیں گے کہ سافٹ ویئر کا استعمال کرتے ہوئے درحقیقت وہ چارٹس کیسے بنائے جاتے ہیں۔
آیئے سب سے زیادہ مفید اور ابتدائی افراد کے لیے دوستانہ ٹولز کا جائزہ لیتے ہیں۔

### اسپریڈ شیٹس (Excel/Google Sheets) میں چارٹس بنانا

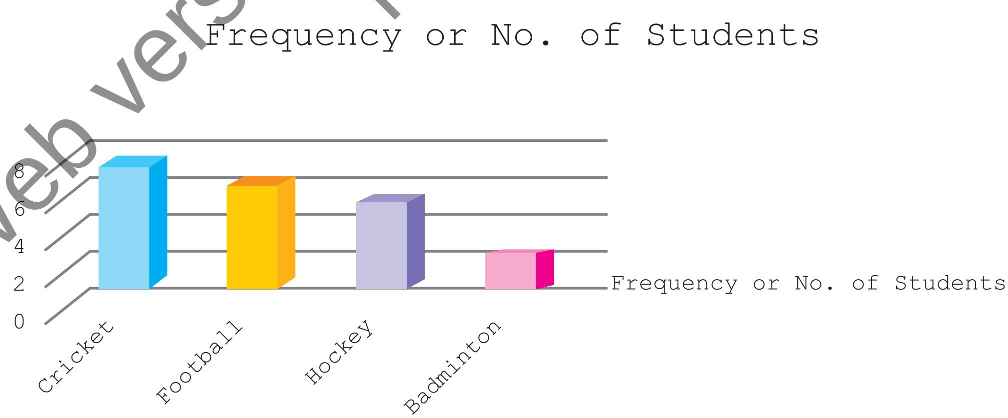
گوگل شیٹس (Google Sheets) اور مائیکروسافٹ ایکسل (Microsoft Excel) ٹیبلز سے چارٹس اور گرافس بنانے کے لیے سادہ مگر طاقتور ٹولز ہیں۔

**گوگل شیٹس میں چارٹ بنانے کے اقدامات:**

i. اپنا ڈیٹا قطاروں (rows) اور کالموں (columns) میں درج کریں۔
ii. ڈیٹا کی حد (data range) کو منتخب کریں۔
iii. "Insert" مینیو پر کلک کریں پھر "Chart" کو منتخب کریں۔
iv. چارٹ کی قسم (بار، پائی، لائن) کا انتخاب کریں۔
v. چارٹ کو ایک عنوان (title)، لیبل (label)، اور رنگ کے ساتھ حسب ضرورت (Customize) بنائیں۔

مثال: فرض کریں کہ آپ کے پاس گوگل شیٹس میں ڈیٹا ہے (براہ کرم جدول 5.6 سے رجوع کریں)۔ آپ کی حوصلہ افزائی کی جاتی ہے کہ آپ جدول 5.6 کے ڈیٹا کو بار چارٹ میں تبدیل کرنے کے لیے ان اقدامات کا استعمال کریں۔

Web version of PECTAA Textbook
Not for Sale




شکل 5.1: طالب علموں کے پسندیدہ کھیل (جدول 5.6 کے ڈیٹا پر مبنی)


NOT FOR SALE-PESRP

ہم اسپریڈ شیٹس کا استعمال اس لیے کرتے ہیں کیونکہ وہ:

i. استعمال میں آسان ہیں،
ii. انسٹالیشن کی ضرورت نہیں (گوگل شیٹس کے لیے)،
iii. چھوٹے سے درمیانے سائز کے پراجیکٹس کے لیے بہترین ہیں۔
iv. شکل 5.1: ہمارے جدول 5.6 کے مطابق ایک سادہ بار گراف دکھاتی ہے۔

**سرگرمی:**

سروے ڈیٹا جدول 5.4 سے بار چارٹ، پائی چارٹ اور لائن چارٹ بنانا

مقصد: یہ سیکھنا کہ سروے ڈیٹا ویژولائیزیشن پیش کرنے کے لیے فریکوئنسی ٹیبل کو بار چارٹ، پائی چارٹ، اور لائن چارٹ میں کیسے تبدیل کیا جاتا ہے۔

ہدایات:

- جدول 5.4 میں دیے گئے نمونہ ڈیٹا کا استعمال کریں۔
- اپنی نوٹ بک میں درج ذیل تینوں چارٹس کو بنائیں (بار چارٹ، پائی چارٹ، لائن چارٹ)۔
- اپنے نتائج کو سادہ جملوں کا استعمال کرتے ہوئے اپنی نوٹ بک میں پیش کریں۔

MEMORY BOX

**یاد رکھیں:**

1. گوگل شیٹس اور ایکسل چارٹس بنانے کے لیے آسان ٹولز ہیں۔
2. ایک ڈیش بورڈ ایک ہی جگہ پر متعدد ویژولز (multiple visuals) دکھاتا ہے تاکہ ایک واضح ڈیٹا کہانی بیان کی جاسکے۔

نوٹ: معیاری DSLC چھ اہم مراحل پر مشتمل ہوتا ہے، تاہم، حقیقی دنیا کے کچھ اطلاقات / منظرناموں میں، کچھ اضافی مراحل بھی شامل کیے جاسکتے ہیں۔ ایسا ہی ایک مرحلہ ڈیٹا کو محفوظ کرنا اور منظم کرنا ہوسکتا ہے، خاص طور پر بڑے منصوبوں یا ٹیم کے ماحول میں جہاں ڈیٹا بیسز اور سٹرکچرڈ سٹوریج اہم ہوتی ہے۔ انہیں عام طور پر DBMS اور SQL جیسے نظاموں کے ذریعے منظم کیا جاتا ہے۔

## 5.4 ڈیٹا بیس اور ڈیٹا کو محفوظ کرنا

ایک بے ترتیب سروے کے جدول کو کلین کرنے کے بعد، کلین شدہ ڈیٹا کو ایک مناسب فارمیٹ میں (یا تو اسپریڈ شیٹ میں یا ایک ڈیٹا بیس میں) ذخیرہ کرنے کی ضرورت ہوتی ہے، تاکہ یہ منظم، قابل تلاش، اور تجزیہ کے لیے تیار ہو۔ آیئے دیکھتے ہیں کہ ڈیٹا کو کیسے اور کہاں ذخیرہ کیا جاتا ہے۔



NOT FOR SALE-PESRP

Web version of PECTAA Textbook Not for Sale

سٹرکچرڈ ڈیٹا کو ذخیرہ کرنے کے دو اہم طریقے ہیں:

.i اسپریڈ شیٹس (Spreadsheets)

.ii ڈیٹا بیسز (Databases)

دونوں کے درمیان فرق جدول 5.8 میں فراہم کیا گیا ہے۔

جدول 5.8 اسپریڈ شیٹ بمقابلہ ڈیٹا بیس

| طریقہ | وضاحت | مثالی ٹول |
|---|---|---|
| اسپریڈ شیٹس | چھوٹے ڈیٹا سیٹس کے لیے قطاروں اور کالموں پر مشتمل ایک گرڈ۔ | ایکسل (Excel)، گوگل شیٹس (Google Sheets) |
| ڈیٹا بیسز | بڑے ڈیٹا کو ٹیبلز میں محفوظ کرنے اور منظم کرنے کے لیے ایک سٹرکچرڈ نظام۔ | مائی ایس کیو ایل (MySQL)، ایس کیو لائٹ (SQLite) |

**SMART TIPS**

سمارٹ ٹپس

اسپریڈ شیٹس چھوٹے پراجیکٹس کے لیے اچھی ہیں۔ ڈیٹا بیسز بڑے یا ملٹی یوزر ڈیٹا سسٹمز (multi-user data systems) کے لیے بہتر ہوتے ہیں۔

## ڈیٹا بیس کیا ہے؟

ایک ڈیٹا بیس (Database) ڈیٹا کا ایک منظم مجموعہ ہے۔ یہ ڈیٹا کو ٹیبلز کی شکل میں محفوظ کرتا ہے۔ ہر ٹیبل قطاروں (rows) اور کالموں (columns) کے ساتھ ایک چھوٹی اسپریڈ شیٹ کی طرح ہوتا ہے۔ ڈیٹا بیسز ڈیٹا کو مؤثر طریقے سے محفوظ کرنے، بازیافت کرنے، اور اپ ڈیٹ کرنے میں مدد دیتے ہیں۔ جدول 5.9 ڈیٹا بیس سے متعلق اصطلاحات کی وضاحت کرتا ہے۔

جدول 5.9 ڈیٹا بیس اصلاحات: معنی اور مثالیں

| اصطلاح | معنی | مثال |
|---|---|---|
| ٹیبل | متعلقہ ڈیٹا کا ایک گروپ۔ | اسپورٹس سروے |
| ریکارڈ | ٹیبل میں ایک واحد اندراج (قطار-row)۔ | علی کے ڈیٹا کی مکمل قطار۔ |
| فیلڈ | معلومات کا ایک مخصوص حصہ (کالم-column)۔ | نام، کھیل، فریکوئنسی |

Web version of PECTAA Textbook Not for Sale



NOT FOR SALE-PESRP

**سرگرمی:**

ڈیٹا کو منظم کرنا

مقصد: مستقل فارمیٹنگ استعمال کرتے ہوئے ڈیٹا کو منظم اور بہتر ساخت والے جدول میں ترتیب دینا سیکھیں۔

ہدایات: پہلے سے حاصل کیے گئے کلین ڈیٹا کو استعمال کریں۔

- اپنی نوٹ بک پر ایک نیا جدول بنائیں جس کے کالم ہیڈنگز درج ذیل ہوں۔

Frequency, Time of day, Favorite sport Name, Roll No

- کم سے کم 5 قطاروں میں ڈیٹا لکھیں۔
- اپنا ڈیٹا صحیح طریقے سے کالمز میں لکھیں۔
- اس بات کو یقینی بنائیں کہ تمام انٹریز کے صحیح لکھے گئے ہوں اور غلطیوں سے پاک ہوں۔

## ڈی بی ایم ایس اور ایس کیو ایل (DBMS and SQL)

ڈیٹابیس مینجمنٹ سسٹم (DBMS) وہ سافٹ ویئر ہے جو ڈیٹابیسز کو بنانے، منظم کرنے اور ان کے ساتھ تعامل کرنے کے لیے استعمال ہوتا ہے۔ اس کی مثالوں میں MySQL، SQLite، اور مائیکروسافٹ ایکسیس (Microsoft Access) شامل ہیں۔

ڈیٹابیس میں ڈیٹا کے ساتھ کام کرنے کے لیے، ہم ایک لینگوئج استعمال کرتے ہیں جسے SQL یا سٹرکچرڈ قیوری لینگوئج (Structured Query Language - SQL) کہا جاتا ہے۔

SQL ہمیں سادہ کمانڈز کا استعمال کرتے ہوئے ڈیٹا کو شامل کرنے (add)، اپ ڈیٹ کرنے، حذف کرنے، یا بازیافت کرنے (retrieve) میں مدد کرتی ہے۔

**چند بنیادی SQL قیوریز (Queries) درج ذیل ہیں:**

1. SELECT * FROM Sports

یہ سادہ قیوری (Query) Sports نامی ٹیبل سے تمام ڈیٹا دکھاتی ہے۔

2.SELECT Name FROM Sports WHERE Favorite_Sport = 'Cricket'

یہ قیوری Sports ٹیبل میں موجود ان طلباء کے نام دکھاتی ہے جو کرکٹ پسند کرتے ہیں۔

3.INSERT INTO Sports (Name, Favorite_Sport) VALUES ('Ali', 'Football')

یہ قیوری ایک نیا اندراج (entry) شامل کرتی ہے۔

یہ بنیادی SQL اسٹیٹمنٹس بڑے ڈیٹا بیسز سے مخصوص ڈیٹا کو مؤثر طریقے سے منظم کرنے اور بازیافت کرنے میں مدد کرتی ہیں۔

**MEMORY BOX**

**یاد رکھیں:**

1. آسان رسائی اور تجزیہ کے لیے ڈیٹا کو منظم طریقے سے محفوظ کیا جانا چاہیے۔
2. اسپریڈ شیٹس چھوٹے ڈیٹا سیٹس کے لیے مفید ہیں۔
3. ڈیٹا بیسز بڑے اور زیادہ پیچیدہ ڈیٹا کے لیے استعمال ہوتے ہیں۔
4. ایک ٹیبل ریکارڈز (قطاروں) اور فیلڈز (کالموں) کا استعمال کرتے ہوئے ڈیٹا کو ذخیرہ کرتا ہے۔
5. منظم ڈیٹا غلطیوں کو کم کرتا ہے اور تجزیہ کو آسان بناتا ہے۔

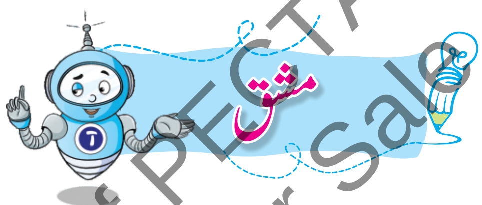
Web version of PECTAA Textbook Not for Sale



NOT FOR SALE-PESRP

## خلاصہ (Summary)

- اسپریڈ شیٹس (Spreadsheets): چھوٹے ڈیٹا سیٹس کے لیے قطاروں اور کالموں کا گرڈ (جیسے ایکسل، گوگل شیٹس)۔
- ڈیٹا بیس (Database): ٹیبلز میں ذخیرہ شدہ ڈیٹا کا ایک منظم مجموعہ۔
- ٹیبل (Table): متعلقہ ڈیٹا کا ایک گروپ (ایک چھوٹی اسپریڈ شیٹ کی طرح)۔
- ریکارڈ (Record): ایک ٹیبل میں ایک واحد اندراج (قطار)۔
- فیلڈ (Field): ایک ٹیبل میں معلومات کا ایک مخصوص حصہ (کالم)۔
- SQL (سٹرکچرڈ قیوری لینگویج): وہ لینگویج جو ڈیٹا بیسز میں ڈیٹا کے ساتھ کام کرنے کے لیے استعمال ہوتی ہے۔




### کثیر انتخابی سوالات

**i** ڈیٹا سائنس لائف سائیکل (DSLC) میں پہلا مرحلہ کیا ہے؟

- (الف) ڈیٹا اکٹھا کرنا
- (ب) ڈیٹا کی بصری پیشکش کرنا
- (ج) مسئلہ کو سمجھنا
- (د) ڈیٹا کا تجزیہ کرنا

**ii** بہت سے طالب علموں کی آراء اکٹھا کرنے کا بہترین طریقہ کون سا ہے؟

- (الف) IoT آلات
- (ب) سروے
- (ج) ویب سکریپنگ
- (د) ڈیٹا بیس

**iii** ڈیٹا میں جانبداری (Bias) کیا ہے؟

- (الف) ڈیٹا کا بہت بڑا سائز
- (ب) غیر منصفانہ یا یک طرفہ ڈیٹا
- (ج) اچھی طرح کلین شدہ ڈیٹا
- (د) تیز رفتار ڈیٹا

Web version of PECTAA Textbook Not for Sale



NOT FOR SALE-PESRP

iv چارٹ بنانے کے لیے ابتدائی افراد کے لیے مثالی ٹول کون سا ہے؟

الف Tableau
ب Python
ج Google Sheets
د Hadoop

v ایک ڈیٹابیس معلومات کو ------------------ کی شکل میں محفوظ کرتا ہے۔

الف سلائیڈز
ب ٹیبلز
ج بلاکس
د لئیرز

vi درج ذیل چارٹس میں سے کون سا گروہوں کا موازنہ کرنے کے لیے بہترین ہے؟

الف پائی چارٹ
ب بار چارٹ
ج لائن چارٹ
د ایریا چارٹ

vii ڈیٹا کا تجزیہ کرنے کا بنیادی مقصد کیا ہے؟

الف پرانا ڈیٹا حذف کرنا
ب ڈیٹا کو سجانا
ج پیٹرنز تلاش کرنا اور فیصلے کرنا
د انٹرنیٹ پر اپ لوڈ کرنا

viii درج ذیل میں سے کون سا ٹول ایک جدید ڈیٹا ویژولائزیشن پلیٹ فارم ہے؟

الف MS Paint
ب Google data Studio
ج Note Pad
د WhatsApp

ix ڈیٹا کو ذخیرہ کرنے سے پہلے کیا کرنا چاہیے؟

الف اینیمیشنز شامل کرنا
ب اسے کلین اور تصدیق کرنا
ج اسے انکرپٹ کرنا
د اس کا ترجمہ کرنا

x درج ذیل میں سے کون سی ایک ڈیٹا سورس کی مثال ہے؟

الف یوٹیوب
ب سروے فارم
ج وائٹ بورڈ
د سلائیڈ شو

Web version of PECTAA Textbook Not for Sale



NOT FOR SALE-PESRP

## مختصر سوالات

1. ڈیٹا اکٹھا کرنے سے پہلے مسئلہ کو سمجھنے کی کیا اہمیت ہے؟
2. ڈیٹا اکٹھا کرنے کے کوئی دو طریقوں کے نام بتائیں اور ہر ایک کی ایک مثال دیں۔
3. "ڈیٹا کلیننگ" سے ہمارا کیا مطلب ہے؟
4. ڈیٹا استعمال کرنے سے پہلے اسے تصدیق (validate) کرنا کیوں ضروری ہے؟
5. چارٹس کا استعمال کرتے ہوئے ڈیٹا کو بصری شکل دینے کا کیا مقصد ہے؟
6. ڈیٹا میں جانبداری نتائج کو کیسے متاثر کرتی ہے؟
7. ایک بے ترتیب سروے کو کلین کرنے کے لیے کوئی سے دو مراحل / اقدامات لکھیں۔
8. چارٹ بنانے کے لیے استعمال ہونے والے کوئی سے دو ٹولز کے نام بتائیں۔

## تفصیلی سوالات

1. ڈیٹا سائنس لائف سائیکل (DSLC) کے چھ مراحل کو ہر ایک کے لیے ایک سطر میں بیان کریں۔
2. بار چارٹ، پائی چارٹ، اور لائن چارٹ کا مثالوں کے ساتھ تفصیلی موازنہ کریں۔
3. DBMS اور SQL کے درمیان فرق واضح کریں۔
4. SQL میں ایک کیوری (Query) لکھیں جس کے ذریعے ایک طالب علم کا تمام ڈیٹا حاصل کیا جائے جس کا رول نمبر 15 ہے۔

## جوابات

1. ج – مسئلہ کو سمجھنا
2. ب – سروے
3. ب – غیر منصفانہ یا یک طرفہ ڈیٹا
4. ج – Google Sheets
5. ب – ٹیبلز
6. ب – بار چارٹ
7. ج – نمونے تلاش کرنا اور فیصلے کرنا
8. ب – Google data Studio
9. ب – اسے کلین اور تصدیق کرنا
10. ب – سروے فارم

Web version of PECTAA Textbook
Not for Sale



NOT FOR SALE-PESRP