What is Data Analysis

Data Analysis

लागत लाभ विश्लेषण (cost benefit analysis) के लिए डाटा विश्लेषण आवश्यक है | सिस्टम की जांच-पड़ताल तथा डाटा संग्रहण वर्तमान उपलब्धियों का आंकलन होता है | हमारी रुचि यह पता करने में है की कैसे कुछ कदम पूरी कार्यकुशलता के साथ संपन्न किये जाते हैं, कैसे वे वांछित लक्ष्यों की प्राप्ति में मदद कर सकते हैं तथा कैसे निर्माण की लागत सुधारी जा सकती है | डेटा विश्लेषण निर्णय को अधिक वैज्ञानिक बनाने और व्यवसाय को प्रभावी संचालन करने में मदद करता है। इसका उपयोग विभिन्न व्यवसाय, विज्ञान और सामाजिक विज्ञान डोमेन में किया जा रहा है।

विश्लेषण से सिस्टम डिजाईन संबंधी आवश्यकताओं को पहचाना जाता है। आवश्यक सुधार करने के लिए कैंडिडेट सिस्टम में इन फीचर को शामिल किया जाना चाहिए |

सिस्टम की आवश्यकताएँ निम्नांकित हैं -

- बेहतर ग्राहक सेवा |
- सूचना को फिर से प्राप्त करने की तीव्र गति |
- नोटिस की त्वरित तैयारी |
- बिलिंग की बेहतर विशुद्धता |
- प्रोसेसिंग और ऑपरेटिंग में सुधार |
- स्टाफ की कार्यकुशलता में सुधार |
- त्रुटियाँ हटाने के लिए ससंगत बिलिंग प्रक्रिया |

डिजाईन सम्बन्धी इन उद्देश्यों को प्राप्त करने के लिए विभिन्न विकल्पों का पता करना होगा | यदि साधारणतः एक से अधिक विकल्प हो | एनालिस्ट उनमें से केवल उनका चयन करता है जो आर्थिक, तकनीकी और संचालन की दृष्टि से उपयुक्त होते हैं | प्रत्येक विधि के अपने लाभ और नुकसान हैं |

Process of Data Analysis

- Data requirements
 - डेटा विश्लेषण के लिए इनपुट के रूप में आवश्यक हैं, जो विश्लेषक या ग्राहकों (जो विश्लेषण के तैयार उत्पाद का उपयोग करेंगे) की आवश्यकताओं के आधार पर बनाया जाता है।
- Data collection
 - विभिन्न स्रोतों से डेटा एकत्र किया जाता है।
 - डेटा को विभिन्न सेंसर के माध्यम से भी एकत्र किया जा सकता है, जैसे ट्रैफिक कैमरा, उपग्रह, रिकॉर्डिंग डिवाइस आदि।
 - यह इंटरव्यू, ऑनलाइन डाउनलोड, प्रश्नावली के माध्यम से भी प्राप्त किया जा सकता है।
- Data processing

- प्रारंभ में प्राप्त डेटा को विश्लेषण के लिए संसाधित या व्यवस्थित किया जाना चाहिए। उदाहरण के लिए, प्राप्त डेटा को टेबल के रूप में रखा जा सकता है जिसके लिए विभिन्न स्प्रेडशीट या डेटाबेस मैनेजमेंट सॉफ्टवेयर का प्रयोग किया जा सकता है।
- Data cleaning
 - एक बार संसाधित और व्यवस्थित होने के बाद, डेटा अधूरा, डुप्लिकेट या उसमें त्रुटियां हो सकती हैं।
 - **Data cleaning** इन त्रुटियों को रोकने और ठीक करने की प्रक्रिया है। सामान्य कार्यों में शामिल हैं
 - रिकॉर्ड मिलान (record matching)
 - डेटा की सटीकता की पहचान करना (identifying accuracy of data)
 - मौजूदा डेटा की समग्र गुणवत्ता (overall quality of existing data)
 - डिडुप्लीकेशन (de-duplication) - इसके अंतर्गत डेटा में से डुप्लिकेट डेटा को अलग किया जाता है।
 - डेटा को कॉलम में विभाजन (column segmentation)
- Exploratory data analysis
 - एक बार डेटा स्पष्ट या डेटा के clean हो जाने के बाद, इसका विश्लेषण किया जा सकता है।
 - विश्लेषक डेटा में निहित संदेशों को समझने के लिए डेटा विश्लेषण के रूप में विभिन्न तकनीकों का प्रयोग कर सकता है।
- Modeling and algorithms
 - गणितीय सूत्र या मॉडल जिसे एल्गोरिदम कहा जाता है, एल्गोरिदम को चर (variable) के बीच संबंधों की पहचान करने के लिए डेटा पर लागू किया जा सकता है।
- Data product
 - डेटा उत्पाद एक कंप्यूटर अनुप्रयोग है जो डेटा को इनपुट के रूप में लेता है और आउटपुट उत्पन्न करता है।
 - यह एक मॉडल या एल्गोरिथम पर आधारित हो सकता है।
 - उदाहरण के लिए कोई एक एप्लीकेशन कस्टमर की पुरानी खरीदों का विश्लेषण करता है और उन्हें उनके जरूरतों के आधार पर नए प्रोडक्ट दिखाता है।
- Communication
 - एक बार डेटा का विश्लेषण हो जाने के, उपयोगकर्ताओं को उनकी आवश्यकताओं के आधार पर कई स्वरूपों में प्रस्तुत किया जा सकता है। उपयोगकर्ताओं की प्रतिक्रिया के अनुसार इसे दुबारा अतिरिक्त विश्लेषण के लिए भेजा जा सकता है।
 - विश्लेषक उपयोगकर्ताओं को डेटा को स्पष्ट रूप से और कुशलता से संवाद करने के लिए विभिन्न डेटा विज़ुअलाइज़ेशन तकनीकों का प्रयोग कर सकता है। जैसे चार्ट, टेबल इत्यादी