

# **Data Analytics Process**

## **Application in Real Life Scenario**

Data analytics is the process of examining, cleaning, transforming, and modeling data to discover useful information, draw conclusions, and support decision-making.

For instance, in healthcare, data analytics can predict patient outcomes, optimize treatment plans, and streamline hospital operations. In retail, businesses use analytics to understand customer behavior, manage inventory, and personalize marketing strategies. Financial institutions rely on data analytics for risk assessment, fraud detection, and investment planning. Additionally, governments employ analytics to improve public services, monitor environmental changes, and allocate resources effectively.

By turning raw data into actionable insights, data analytics plays a critical role in enhancing efficiency, innovation, and competitiveness in real-world scenarios. Data analytics finds applications across various industries and sectors, transforming the way organizations operate and make decisions. Here are some examples of how data analytics is applied in everyday life:

### **1. Plan a successful family picnic.**

#### **Plan :**

What is the best location? What food should be brought? What activities will be enjoyed by everyone? What's the ideal date and time?

#### **Prepare :**

Gather data on each family member's preferences for picnic location, food, and activities. Use surveys, polls, or discussions to gather input. Collect weather forecasts for Preferred locations, favorite foods, preferred activities, dietary restrictions, preferred dates and times.

#### **Process :**

Ensure all data is relevant, accurate, and properly formatted. For example, if someone lists a location that is too far or unavailable on the preferred date, it should be filtered out. Follow up with family members who haven't provided preferences.

### **Analyze :**

Identify trends in family preferences. For eg. Determine the most popular picnic spots, favored foods, and common activity interests. Analyze weather data to find a date with the most favorable conditions. Assess the locations for distance from home and accessibility for all family members. Estimate the cost of the picnic, including travel, food, and supplies.

### **Share :**

Share the final plan with the family, including the date, time, location, and what each person should bring. Shop for food, prepare activities, and ensure all logistics are handled.

### **Act :**

Based on the analysis, choose the best date and location for the picnic. Select the food items and activities that cater to the majority's Preferences. Create a checklist of items to bring and tasks to complete before the picnic.

## **2. Selecting a good project team.**

### **Define the Problem**

Clearly define what the project needs in terms of skills, experience, and knowledge. Determine what a "good" project team looks like, considering factors like technical skills, communication abilities, and past performance.

## Data Collection

Collect relevant data on potential team members. This could include:

- **Skills and Certifications:** Technical and soft skills, relevant certifications, and educational background.
- **Experience:** Past project experience, duration of work in specific roles, and industry experience.
- **Personality and Behavioral Data:** Assessment results, teamwork capabilities, and leadership qualities.

## Data Cleaning

Remove duplicates, correct errors, and handle missing values. Group similar skills, experiences into categories for easier analysis.

## Data Analysis

Use charts, graphs, and other visualization tools to understand the distribution of skills, experience, and other relevant factors among candidates.

## Model Building

Use statistical or machine learning models to predict the potential success of different team combinations. Examples of models include: Regression model to determine the impact of different factors on project success and classification model to categorize candidates as high, medium, or low fit for the project.

## Decision Making

Based on the model results, choose the team composition that best aligns with your project goals. Present your findings and recommendations to key stakeholders for feedback and final decision-making.

## Implementation

Assemble the team based on the data-driven insights.

### 3. Recommendation System

#### Define the Problem

Suppose we want to build a recommendation system that suggests songs or artists based on user preferences and listening history.

#### Data Collection

- **User Data:** Collect user-specific data, such as listening history, playlists, likes, and dislikes.
- **Song Data:** Gather metadata on songs, including genres, artists, albums, and acoustic features like tempo, energy, and danceability.
- **External Data:** Optionally, include external data like social media trends.

#### Data Cleaning

Handle missing values, remove duplicates, and filter out irrelevant data.

#### Model Selection

- **Collaborative Filtering:**
  - **User-Based Filtering:** Recommend songs based on similar users.
  - **Item-Based Filtering:** Recommend songs similar to the ones a user has liked.
- **Content-Based Filtering:** Recommend songs with similar acoustic features to the ones a user has liked.
- **Hybrid Models:** Combine collaborative filtering with content-based methods to improve recommendations.

#### Model Deployment

Decide whether to deploy the model in batch or real-time. Integrate the recommendation system with the Spotify-like application or platform. Ensure that the system can handle large volumes of users and songs efficiently.

## 4. Gain a deeper understanding of how customers interact with your business.

### Define Objectives

Determine what specific aspects of customer behavior you want to understand (e.g., purchasing patterns, product preferences, churn rate).

### Collect Data

Identify data sources (e.g., transaction records, customer surveys, website analytics) and gather data from these sources. This might involve using tools like CRM systems, web analytics platforms, or social media insights.

### Data Preparation

Remove duplicates, handle missing values, and correct errors in the data. Convert data into a suitable format for analysis (e.g., aggregating data, normalizing values).

### Analyze Data

- **Segmentation:** Divide customers into groups based on similarities (e.g., demographic segments, purchasing behavior).
- **Behavioral Analysis:** Look at specific behaviors such as buying frequency, average order value, or response to marketing campaigns.
- **Predictive Analysis:** Use models to predict future behavior based on historical data (e.g., churn prediction, lifetime value estimation).

### Results

Extract meaningful insights from the analysis (e.g., which products are most popular among different segments). Use graphs, charts, and dashboards to present the data effectively.

### Take Action

Track the results of any changes made to see if they have the desired effect on customer behavior. Use graphs, charts, and dashboards to present the data effectively.

## **5. Analyze the pattern of money collection by kids in a piggy bank.**

### **Define the Objective**

Understand and analyze the pattern of money collection by kids in a piggy bank.

### **Data Collection**

- Date and time of each deposit.
- Amount deposited (coins, bills).
- Source of the money (gifts, allowance, etc.).
- Motivation or reason for the deposit (saving for a toy, holiday money, etc.).

### **Data Preprocessing**

- Remove duplicates. Correct any inconsistencies (e.g.wrong dates)

### **Data Analysis**

Create line graphs showing the amount of money accumulated over time. Use bar charts to visualize the frequency of deposits by day, week, or month. Pie charts to represent the proportion of different types of deposits (coins vs. bills). Identify any seasonal trends or patterns (e.g., more deposits during holidays or birthdays).

### **Results**

Determine key factors that influence savings behavior. Identify which kids are more consistent savers and why.

### **Monitoring**

Based on the insights, design an action plan to encourage better saving habits. Suggest ways to encourage saving behavior (e.g., reward systems, setting goals).

## 6. Understand and optimize the daily routine of an Uber driver

### Define the Objective

Understand and optimize the daily routine of an Uber driver for better efficiency and earnings.

- How many trips does a driver complete daily?
- What are the peak hours for earnings?
- What routes are most profitable?
- How does time spent online correlate with earnings?

### Data Collection

- **Trip Data:** Start and end times, pickup and drop-off locations, trip duration, fare amount.
- **Driver Activity:** Online/offline times, idle times, breaks, time spent at each location.
- **Environmental Factors:** Traffic conditions, weather, local events.
- **Customer Data:** Ratings, feedback, trip cancellations.

### Data Cleaning

Ensure that there are no duplicate entries in the dataset. Identify and remove data points that don't make sense (e.g., unusually high trip durations).

### Visualization

- **Heatmaps:** Display areas with high pickup/drop-off frequency.
- **Time Series Analysis:** Plot earnings over time to identify peak hours.
- **Bar Charts:** Compare earnings across different days of the week.

## Data Modeling

- **Regression Models:** Predict earnings based on hours worked, locations, and other factors.
- **Clustering:** Group similar trips or times of day to identify patterns.

## Reporting

Create a dashboard to display key metrics like daily earnings, number of trips, and active hours. Generate detailed reports highlighting the findings and actionable insights.

## Monitoring and Refinement

Regularly update the data and refine the models as new data comes in. Continuously collect feedback from drivers to improve recommendations.

## 7. Diagnosing and treating a patient's disease.

### Data Collection

- **Patient History:** Collecting information on the patient's medical history, including previous illnesses, family medical history, lifestyle habits, and any current symptoms.
- **Clinical Data:** Gathering clinical data through physical exams, lab tests (blood tests, urine tests, etc.), imaging (X-rays, MRIs, CT scans), and other diagnostic tools.
- **Health Records:** Accessing existing medical records, including previous diagnoses, treatments, and medications.

### Data Cleaning

Removing or correcting errors in the data, such as duplicates, incomplete records, or outliers that might skew the analysis. Combining data from various sources, like lab results and patient history, into a unified dataset.



## **Visualization**

Analyzing the data to identify patterns, correlations, or anomalies that may indicate a specific disease. Using charts, graphs, and heatmaps to visualize data trends, such as tracking the progression of symptoms or comparing patient data against population averages.

## **Predictive Modeling**

Applying machine learning models to predict possible diseases based on the patient's data. These models can be trained on large datasets to recognize disease markers. Evaluating the probability of various diseases and identifying high-risk conditions that require immediate attention.

## **Diagnosis**

Ordering additional tests to confirm the presence of a suspected disease, such as a biopsy for cancer or a genetic test for inherited conditions.

## **Treatment Planning**

Developing a treatment plan tailored to the patient's specific condition, health status, and preferences.

Using data analytics to simulate different treatment scenarios and predict outcomes, helping doctors choose the most effective therapy.

## **Monitoring and Evaluation**

Continuously monitoring the patient's response to treatment through follow-up tests and tracking symptoms. Analyzing treatment outcomes to assess effectiveness, adjust the treatment plan, and identify any complications early.

## **Reporting**

Documenting the entire process in the patient's health record for future reference. Generating reports for the patient, healthcare providers including treatment summaries, progress reports, and future recommendations.

## **8. Managing student information efficiently in schools**

### **Data Collection**

Collect data on student demographics, attendance, grades, behavior, health records, and extracurricular activities. Ensure accurate and consistent data entry by staff, using standardized formats.

### **Data Cleaning**

Identify and correct errors or inconsistencies in the data (e.g., missing values, duplicate records).

### **Data Storage and Management**

Store the cleaned and integrated data in a centralized database, ensuring security and access control. For larger datasets, consider a data warehouse that supports querying and reporting.

### **Data Analysis**

Analyze trends in attendance, performance, and behavior using statistical tools. Use machine learning models to predict student outcomes like graduation rates or at-risk students.

### **Data Reporting**

Generate reports for different stakeholders, such as teachers, administrators, and parents, focusing on relevant data. Set up automated reports to be generated at regular intervals (e.g., weekly attendance reports).

### **Data Utilization**

Use insights from data analysis to inform decisions on curriculum changes, resource allocation, and interventions for at-risk students. Provide educational support to individual student needs based on their data profiles.

## **Conclusion**

Data analytics plays a crucial role in real life by transforming raw data into actionable insights that drive decision-making and strategy across various domains.

In business, it helps companies understand customer behavior, optimize operations, and predict market trends, leading to more informed decisions and competitive advantages. In healthcare, data analytics enables better patient outcomes through personalized medicine and efficient resource management. In finance, it aids in risk assessment and fraud detection, while in education, it supports personalized learning and institutional improvements.

Overall, data analytics empowers individuals and organizations to make evidence-based decisions, enhance efficiency, and achieve better outcomes in their respective fields.