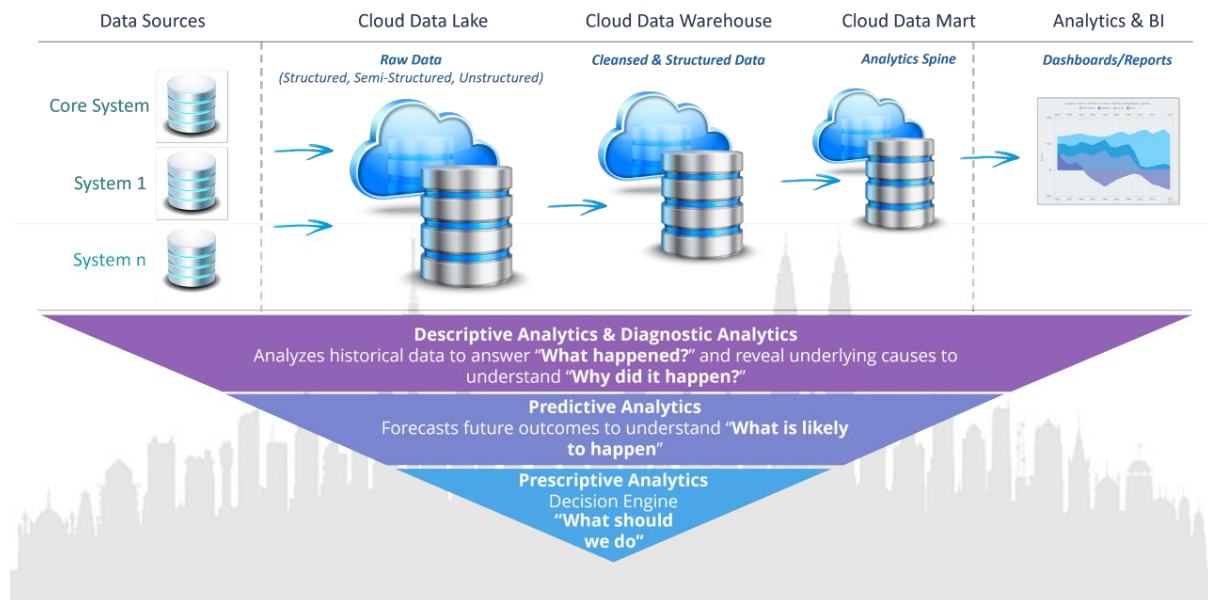


DATA ANALYTICS JOURNEY

Article by Dylan Tan

14 minute read



Data Analytics Journey

Many different types of data analytics can be used to examine raw data, big data, or statistics to uncover valuable insights that businesses can use for decision-making. There are primarily four main types of data analytics which offer distinctive perspective, insights, patterns, correlations, and trends. They are descriptive analytics, diagnostic analytics, predictive analytics and prescriptive analytics.

Descriptive analytics is a form of analytics that examines data or content to answer the question, "**What happened?**"

Descriptive Analytics

Descriptive analytics focuses on summarizing historical data to provide insights into what has happened in the past. It answers the question "**What happened?**" and lays the foundation for further evaluation by establishing a baseline understanding of the data's historical context. Some of the typical use cases involve:

- Reporting on sales performance.
- Visualizing website traffic over time.

- Demand trends.
- Summarizing customer feedback.
- Customer segmentation.
- Financial reporting.

Scenario Example:

A retail company uses descriptive analytics to create weekly reports on sales performance across different product categories, helping managers track trends and make informed decisions on inventory management.

Advantages

- Provides a snapshot of historical data.
- Simplifies data interpretation making it easier to understand
- Helps identify anomalies in data
- Forms the foundation for more advanced data analytics

Disadvantages

- Limited in providing actionable insights for the future.
- May oversimplify complex scenarios as it is highly dependent on the quality of the data sample; unsuitable or biased samples may distort conclusions

The below are some of the most common ***Descriptive Analytics Techniques***:

- **Data Aggregation** - Summarizes data by grouping and combining information based on certain criteria. An example of use is aggregating sales data by month to analyze monthly performance.
- **Data Summarization** - Provides a concise overview of key metrics such as mean, median, mode, and standard deviation. An example of use is summarizing employee performance ratings for a given period.
- **Data Visualization** - Represents data visually using charts, graphs, and dashboards for better understanding. An example of use is creating a bar chart to visualize sales performance across different product categories.
- **Histograms** - Displays the distribution of a dataset by dividing it into bins and representing the frequency of values in each bin. An example of use is analyzing the distribution of customer ages in a given dataset.
- **Pie Charts** - Represents the proportion of each category in a dataset as a slice of a pie. An example of use is showing the market share of different product lines.
- **Frequency Distribution** - Lists the number of occurrences of each unique value in a dataset. An example of use is creating a frequency distribution table for customer purchase amounts.

- **Heatmaps** - Displays the intensity of data values with color gradients. An example of use is analyzing website traffic patterns over different hours of the day.
- **Time-Series Analysis** - Examines data points collected over time to identify trends, seasonality, and recurring patterns. An example of use is plotting monthly revenue over a year to identify seasonal trends.
- **Percentiles and Quartiles** - Divides a dataset into percentile ranges to understand the distribution of values. An example of use is analyzing salary distributions by calculating quartiles.
- **Central Tendency Measures** - Describes the center of a dataset using metrics like mean, median, and mode. An example of use is calculating the average sales price for a specific product.
- **Cross-Tabulation (Contingency Tables)** - Summarizes data by creating a table that shows the relationship between two categorical variables. An example of use is analyzing the correlation between product satisfaction and customer demographics.
- **Summary Reports** - Provides a comprehensive summary of key metrics and performance indicators. An example of use is **g**enerating a monthly report summarizing sales, expenses, and profit.

When applying descriptive analytics, it is important to choose the most appropriate techniques based on the nature of the data and the specific objectives of the analysis. Data cleaning and preparation are also crucial steps to ensure accurate and meaningful results.

Some of the ***most commonly used tools for Descriptive Analytics*** are *Excel Spreadsheets and Business Intelligence/Visualization Tools like Qlik Sense, Tableau and Power BI.*

How to Get Started with Descriptive Analytics

To get started with descriptive analytics, one shall first collect relevant data, learn data visualization tools and techniques as well as understand basic statistics for summarizing data.

Diagnostic analytics is a form of advanced analytics that examines data or content to answer the question, **“Why did it happen?”**

Diagnostic Analytics

Diagnostic analytics uncovers relationships between variables to understand factors contributing to specific outcomes by examining and scrutinizing historical data. It

involves root cause analysis and often follows descriptive analytics. It aims to answer the question, “**Why did it happen?**”. Some of the typical use cases involve:

- Investigating reasons for a sudden drop in sales.
- Analyzing Employee turnover.
- Identifying factors leading to customer churn.
- Analyzing the root cause of production issues.
- Fraud detection.

Scenario Example:

An e-commerce platform uses diagnostic analytics to investigate a sudden increase in customer complaints, discovering that a recent website update caused usability issues, leading to a swift resolution.

Advantages

- Helps in problem-solving and identifying issues.
- Provides insights into the "why" behind specific outcomes.
- Early detection of anomalies/trends/patterns reduces the likelihood of issues escalating into more significant challenges.
- Supports informed decision-making and development of targeted solutions based on a deeper understanding of events.

Disadvantages

- Requires deep domain knowledge to interpret results accurately.
- Requires in-depth analysis and expertise for accurate interpretation.
- Can be resource and time-intensive, particularly when dealing with large datasets or complex algorithms thus making it less practical for frequent use.
- Mainly focused on understanding historical patterns rather than predicting future ones.

The below are some of the most common ***Diagnostic Analytics Techniques***:

- **Root Cause Analysis** - Investigates the primary factor or factors that contribute to a particular outcome or issue. An example of use is identifying why a manufacturing process is yielding defective products.
- **Drill-Down Analysis** - Examines data at a more granular level to get a detailed view of specific components contributing to a broader trend. An example of use is investigating which specific products are causing a decline in overall sales.
- **Correlation Analysis** - Identifies statistical relationships between variables, helping to understand how changes in one variable may impact another. An

example of use is determining if there is a correlation between advertising spending and sales revenue.

- **Pareto Analysis** - Applies the 80/20 rule to identify the most significant factors contributing to a problem or trend. An example of use is identifying the key products or customers that contribute most to revenue, expenses or issues.
- **Comparative Analysis** - Compares different sets of data to identify patterns, differences, or anomalies. An example of use is comparing the performance of different regions or teams to understand variations in sales or productivity.
- **Regression Analysis** - Examines the relationship between a dependent variable and one or more independent variables, helping to understand how changes in one variable affect another. An example of use is analyzing how changes in marketing spending impact overall revenue.
- **Time-Series Analysis** - Examines data points collected over time to identify trends, seasonality, and recurring patterns. An example of use is understanding the monthly variations in website traffic or sales.
- **Fishbone Diagrams (Ishikawa or Cause-and-Effect Diagrams)** - Visual representation that helps identify and categorize potential causes of a specific problem. An example of use is investigating the root causes of delays in product delivery.
- **Scatter Plots** - Graphically represents the relationship between two variables to identify potential correlations. An example of use is analyzing the relationship between employee training hours and performance metrics.
- **Hypothesis Testing** - Formulates and tests hypotheses to determine if there is a statistically significant relationship between variables. An example of use is testing whether changes in pricing strategy have a significant impact on customer satisfaction.

When performing diagnostic analytics, it is essential to combine these techniques judiciously, considering the nature of the data and the specific objectives of the analysis. Domain knowledge and a thorough understanding of the business context are crucial for effective diagnostic analytics.

Some of the ***most commonly used tools for Diagnostic Analytics*** are statistical analysis software, data drilling and data mining software, and software for anomaly detection.

How to Get Started with Diagnostic Analytics

To get started with diagnostic analytics, one shall need to acquire skills in SQL for data querying as well as learn Python/R etc. for statistical analysis for identifying patterns as the diagnostic analytics process entails defining the problem, gathering relevant data, formulating hypotheses, evaluating hypotheses, interpreting results, recommending solutions, implementing changes, and tracking outcomes.

Predictive analytics is a form of advanced analytics that examines data or content to answer the question, **“What is likely to happen?”**

Predictive Analytics

Predictive analytics uses statistical algorithms, machine learning (ML) and data mining to analyze historical data to make predictions about future outcomes. It aims to answer the question **“What is likely to happen?”** and it empowers organizations to find potential opportunities, assess risks, and make proactive and informed decisions. Some of the typical use cases involve:

- Forecasting future sales and demand based on historical data.
- Risk assessment.
- Predicting equipment failures in manufacturing.
- Anticipating customer preferences for personalized recommendations.
- Predicting consumer trends.
- Predicting employee attrition.
- Healthcare risk classification.

Scenario Example:

An energy company utilizes predictive analytics to forecast future energy consumption patterns, enabling them to optimize energy production and distribution, reducing costs and improving efficiency.

Advantages

- Enables proactive and informed decision-making based on data-driven forecasts and predictions.
- Streamlines planning through anticipating future demands and trends.
- Improves strategies by targeting specific demographics and personalizing campaigns for better customer engagement.
- Mitigates risks by predicting potential challenges.

Disadvantages

- Requires high-quality data and may be sensitive to outliers.
- Heavily relies on the quality of input data; inaccurate, incomplete, or biased data can lead to unreliable predictions and skewed outcomes.
- Handling sensitive or personal data in predictive analytics raises privacy concerns and requires compliance with data protection regulations to avoid legal and ethical issues.

- Predictive models might lose their precision over time due to shifts in market conditions and changes in consumer behaviors; regular updates are necessary to ensure accuracy.
- Developing and maintaining predictive models can be resource-intensive and time consuming, requiring skilled professionals, computational power, and ongoing updates to remain effective.

The below are some of the most common ***Predictive Analytics Techniques***:

- **Linear Regression** - Establishes a linear relationship between a dependent variable and one or more independent variables. An example of use is predicting sales based on factors like advertising expenditure and seasonality.
- **Logistic Regression** - Used when the dependent variable is binary, predicting the probability of an event occurring. An example of use is predicting customer churn (yes/no) based on various customer attributes.
- **Decision Trees** - Creates a tree-like model to make decisions based on input features. An example of use is predicting whether a customer will purchase a product based on demographic and purchase history.
- **Random Forest** - Ensemble learning method that builds multiple decision trees and combines their predictions. An example of use is predicting stock prices based on various economic indicators.
- **Gradient Boosting** - Builds a series of weak models and combines them to create a strong predictive model. An example of use is predicting customer satisfaction based on feedback and historical data.
- **Time Series Analysis** - Analyzes data points collected over time to identify patterns, trends, and seasonality. An example of use is forecasting future sales based on historical sales data.
- **ARIMA (AutoRegressive Integrated Moving Average)** - A time series forecasting method that combines autoregression and moving averages. An example of use is predicting future demand for a product based on historical sales data.
- **K-Nearest Neighbors (KNN)** - Classifies data points based on the majority class among their nearest neighbors. An example of use is predicting customer preferences based on similarities with other customers.
- **Support Vector Machines (SVM)** - Finds a hyperplane that best separates data into different classes. An example of use is predicting whether an email is spam or not based on various features.
- **Neural Networks** - Deep learning models composed of interconnected nodes that mimic the structure of the human brain. An example of use is predicting customer behavior using a neural network with multiple hidden layers.
- **Clustering Analysis** - Groups similar data points into clusters based on certain features. An example of use is predicting market segments and customer behavior through clustering analysis.

- **Ensemble Methods** - Combines multiple models to improve predictive performance. An example of use is using an ensemble of models, such as bagging or stacking, to predict customer churn.
- **Naive Bayes** - A probabilistic classifier based on Bayes' theorem, often used for classification problems. An example of use is predicting whether an email is spam or not based on the probability of certain words.
- **Association Rule Mining** - Identifies relationships and patterns within large datasets. An example of use is predicting product recommendations based on the association between different items in a shopping cart.

When applying predictive analytics, it is essential to select the most suitable technique based on the specific characteristics of the data and the goals of the prediction task. Additionally, careful validation and evaluation of the predictive models are crucial to ensure their accuracy and reliability.

Some of the ***most commonly used tools for Predictive Analytics*** include ML libraries, data mining tools, data visualization tools, BI software, statistical modeling tools, and predictive analytics software.

How to Get Started with Predictive Analytics

To get started with predictive analytics, one shall need to learn programming languages like Python or R as well as explore machine learning algorithms and techniques as predictive analytics is an ongoing process that requires defining goals, data collection and cleaning, selecting prediction techniques, training and validating models, model deployment, making predictions, performance assessment, and continuous improvement.

Prescriptive analytics is a form of advanced analytics that examines data or content to answer the question, **“What should we do?”**

Prescriptive Analytics

Prescriptive analytics provides recommendations on what actions to take to optimize outcomes. It combines insights from descriptive, diagnostic, and predictive analytics. It is an advanced form of analytics that offers recommendations to achieve a specific result. It uses data, mathematical algorithms, and business rules to suggest the best course of action that your organization should take to refine decision-making processes. It aims to answer the question **“What should we do?”** by giving actionable recommendations. Some of the typical use cases involve:

- Optimizing supply chain routes for efficiency.

- Recommending personalized marketing strategies.
- Prescribing actions for fraud detection and prevention.
- Provide recommendations for dynamic pricing strategies.
- Provide recommendations for personalized healthcare treatment plan.

Scenario Example:

A logistics company employs prescriptive analytics to optimize delivery routes based on real-time traffic data, reducing delivery times and fuel costs while improving overall service quality.

Advantages

- Guides decision-making by recommending actionable optimal actions.
- Incorporates real-time data for dynamic decision support.
- Improves resource efficiency and reduces costs by recommending efficient resource allocation.
- Increases revenue by suggesting best pricing strategies and targeted marketing approaches.
- Enhances strategic planning with the ability to adapt strategies to changing conditions.

Disadvantages

- Requires advanced analytics capabilities.
- Significantly dependent on accurate and comprehensive data; if the input data is incomplete, inaccurate, or outdated, it can lead to flawed recommendations.
- Implementing recommended actions can be very complex.
- Raises ethical concerns related to privacy, particularly in areas like personalized marketing.
- Deployment includes investment in technology, data infrastructure, and skilled personnel, which can have substantial costs.
- Requires continuous monitoring and maintenance to guarantee relevance and accuracy.

The below are some of the most common ***Prescriptive Analytics Techniques***:

1. **Optimization Algorithms** - Mathematical algorithms designed to find the best solution among a set of feasible options. An example of use is optimizing supply chain routes, production schedules, or resource allocation.
2. **Simulation Modeling** - Creates a model to simulate real-world scenarios and understand the potential outcomes of different decisions. An example of use is simulating the impact of changes in manufacturing processes on production efficiency.

3. **Game Theory** - Analyzes interactions between different entities to identify optimal strategies and outcomes. An example of use is determining pricing strategies in a competitive market.
4. **Prescriptive Machine Learning** - Combines predictive analytics and optimization techniques to recommend actions that maximize or minimize a specific objective. An example of use is recommending pricing adjustments to maximize revenue based on predicted customer behavior.
5. **Decision Support Systems** - Computer-based tools that aid decision-making by providing insights, analysis, and recommended actions. An example of use is supporting executives in making strategic decisions, such as market entry strategies.
6. **Expert Systems** - Rule-based systems that mimic the decision-making process of a human expert in a particular domain. An example of use is providing expert advice in fields like healthcare for treatment recommendations.
7. **Prescriptive Analytics Software Platforms** - Integrated platforms that leverage various techniques for optimization and decision support. An example of use is using dedicated software to optimize inventory levels in a retail supply chain.
8. **Goal-Seeking Models** - Seeks the optimal input values to achieve a desired outcome. An example of use is adjusting marketing budgets to maximize customer acquisition within a given cost constraint.
9. **Constraint Programming** - Solves problems subject to constraints, ensuring that decisions comply with predefined rules and limitations. An example of use is developing schedules that adhere to workforce constraints and operational requirements.
10. **Prescriptive Analytics Dashboards** - Interactive dashboards that provide real-time insights and recommended actions based on ongoing data analysis. An example of use is monitoring key performance indicators and receiving recommendations for process improvements.
11. **Heuristic Approaches** - Rule-of-thumb methods that provide practical solutions, especially when optimization is challenging. An example of use is using heuristic algorithms for vehicle routing in delivery logistics.
12. **A/B Testing and Multivariate Testing** - Experimentation techniques that compare different versions of a product or process to determine the most effective one. An example of use is optimizing website design or marketing campaigns through controlled experiments.
13. **Monte Carlo simulation** - A sub-type of simulation that involves generating repeated random samples to model the probability distribution of various outcomes and obtain numerical results. An example of use is analyzing the potential outcomes and uncertainties associated with a construction project schedules and costs.

Prescriptive analytics often involves a combination of these techniques to provide comprehensive and actionable recommendations. It requires a deep understanding of business objectives, constraints, and the ability to interpret and act on the insights provided by the models and tools.

Some of the ***most commonly used tools for Prescriptive Analytics*** include ML libraries, data mining tools, data visualization tools, BI software, statistical modeling tools, and predictive analytics software.

Tools used for prescriptive analytics include optimization software, simulation tools, neural networks, AI/ML, prescriptive analytics platforms, advanced analytics platforms and BI tools.

How to Get Started with Prescriptive Analytics

To get started with predictive analytics, one must first build a solid foundation in descriptive, diagnostic, and predictive analytics and then learn optimization techniques and algorithms since prescriptive analytics revolves around defining objectives, collecting and integrating relevant data, preparing data, selecting prescriptive analytics tools, building models, implementing recommendations, monitoring and assessing performance, and constantly enhancing predictive analytics models.