
Professional Certificate in UK-Based HR Analytics

Data Collection and Preparation

Data Collection and Preparation are essential steps in the HR analytics process. In this explanation, we will discuss key terms and vocabulary related to data collection and preparation in the context of the Professional Certificate in UK-Based HR Analytics.

1. Data Collection:

Data collection is the process of gathering and measuring information on variables of interest, in an established systematic fashion that enables one to answer stated research questions, test hypotheses, and evaluate outcomes.

1.1 Primary Data:

Primary data is the data that is collected specifically for the purpose of the research or analysis at hand. It can be collected through various methods such as surveys, interviews, or observations. In the context of HR analytics, primary data can be collected through employee surveys, focus groups, or interviews with HR managers.

1.2 Secondary Data:

Secondary data is the data that has already been collected by someone else, for a different purpose. It can be obtained from various sources such as internal company databases, government databases, or industry reports. In the context of HR analytics, secondary data can be obtained from internal sources such as HRIS (Human Resource Information Systems), payroll systems, or performance management systems. External sources of secondary data can include government labor statistics, industry benchmarking reports, or academic research studies.

1. Data Preparation:

Data preparation is the process of cleaning, transforming, and organizing data for analysis. It involves removing errors, handling missing values, and transforming data into a format that is suitable for analysis.

2.1 Data Cleaning:

Data cleaning is the process of identifying and correcting or removing errors, inconsistencies, and inaccuracies in the data. It involves checking for missing values, outliers, and duplicate records. In the context of HR analytics, data cleaning might involve checking for missing employee records, correcting errors in job titles, or removing duplicate records from a database.

2.2 Data Transformation:

Data transformation is the process of converting data from one format to another, or from one type of variable to another. It involves aggregating data, calculating new variables, and normalizing data. In the context of HR analytics, data transformation might involve aggregating data by department, calculating turnover rates, or normalizing data to a common scale.

2.3 Data Organisation:

Data organization is the process of structuring data in a way that is easy to analyze. It involves creating tables, charts, and graphs that clearly display the data. In the context of HR analytics, data organization might involve creating a table that shows turnover rates by department, or creating a graph that displays the distribution of employee salaries.

3. Data Quality:

Data quality refers to the overall quality of the data, including its accuracy, completeness, and relevance. Poor data quality can lead to incorrect analysis and decision-making.

3.1 Data Accuracy:

Data accuracy refers to the degree to which data accurately reflects the real-world phenomenon that it is intended to measure. In the context of HR analytics, data accuracy might refer to the accuracy of employee records, such as job titles, salaries, or performance ratings.

3.2 Data Completeness:

Data completeness refers to the degree to which all relevant data is included in the dataset. In the context of HR analytics, data completeness might refer to the inclusion of all employees in a turnover analysis, or the inclusion of all departments in a salary comparison.

3.3 Data Relevance:

Data relevance refers to the degree to which the data is relevant to the research question or analysis at hand. In the context of HR analytics, data relevance might refer to the inclusion of data on employee engagement, turnover, or performance.

4. Data Integration:

Data integration is the process of combining data from different sources into a single dataset. It involves reconciling differences in data formats, resolving data inconsistencies, and ensuring data compatibility. In the context of HR analytics, data integration might involve combining data from HRIS, payroll, and performance management systems.

5. Data Security:

Data security refers to the measures taken to protect data from unauthorized access, theft, or damage. It includes physical security measures, such as locked doors and servers, as well as electronic security measures, such as encryption and firewalls.