## **Course content:**

Course 3 – Prepare Data for Exploration

1. **Understanding data types and structures:** We all generate lots of data in our daily lives. In this part of the course, you will check out how we generate data and how analysts decide which data to collect for analysis. You’ll also learn about structured and unstructured data, data types, and data formats as you start thinking about how to prepare your data for exploration.
2. **Understanding bias, credibility, privacy, ethics, and access:** When data analysts work with data, they always check that the data is unbiased and credible. In this part of the course, you will learn how to identify different types of bias in data and how to ensure credibility in your data. You will also explore open data and the relationship between and importance of data ethics and data privacy.
3. **Databases: Where data lives:** When you are analyzing data, you will access much of the data from a database. It’s where data lives. In this part of the course, you will learn all about databases, including how to access them and extract, filter, and sort the data they contain. You will also check out metadata to discover the different types and how analysts use them.
4. **Organizing and protecting your data:** Good organization skills are a big part of most types of work, and data analytics is no different. In this part of the course, you will learn the best practices for organizing data and keeping it secure. You will also learn how analysts use file naming conventions to help them keep their work organized.
5. **Engaging in the data community (optional):** Having a strong online presence can be a big help for job seekers of all kinds. In this part of the course, you will explore how to manage your online presence. You will also discover the benefits of networking with other data analytics professionals.
6. **Completing the Course Challenge**: At the end of this course, you will be able to apply what you have learned in the Course Challenge. The Course Challenge will ask you questions about the key concepts and then will give you an opportunity to put them into practice as you go through two scenarios.

### 

### **How the data will be collected**

Decide if you will collect the data using your own resources or receive (and possibly purchase it) from another party. Data that you collect yourself is called first-party data.

## **Data sources**

If you don’t collect the data using your own resources, you might get data from second-party or third-party data providers. **Second-party data** is collected directly by another group and then sold. **Third-party data** is sold by a provider that didn’t collect the data themselves. Third-party data might come from a number of different sources.

## **Solving your business problem**

Datasets can show a lot of interesting information. But be sure to choose data that can actually help solve your problem. For example, if you are analyzing trends over time, make sure you use time series data — in other words, data that includes dates.

## **How much data to collect**

If you are collecting your own data, make reasonable decisions about sample size. A random sample from existing data might be fine for some projects. Other projects might need more strategic data collection to focus on certain criteria. Each project has its own needs.

## **Time frame**

If you are collecting your own data, decide how long you will need to collect it, especially if you are tracking trends over a long period of time. If you need an immediate answer, you might not have time to collect new data. In this case, you would need to use historical data that already exists.

### 

**Four types of factors are responsible in data collection, which are as follows:**

* Accuracy,
* Validity and Reliability,
* Time and Cost.
* Utility.

**The top 5 data collection methods :**

* Surveys and Questionnaires. ...
* Interviews. ...
* Observations. ...
* Records and Documents. ...
* Focus Groups.