**Tableau Data Analytics assignment**

**1. Basics:**

1. **What is the difference between Discrete and Continuous Data?**

Discrete data

It is the information that can only take certain values. These values don’t have to be whole numbers but they are fixed values

The number of each type of treatment a salon needs to schedule for the week, the number of children attending a nursery each day or the profit a business makes each month are all examples of discrete data. This type of data is often represented using tally charts, bar charts or pie charts.

Continuous data

It is the data that can take any value. Height, weight, temperature and length are all examples of continuous data. Some continuous data will change over time; the weight of a baby in its first year or the temperature in a room throughout the day. This data is best shown on a line graph as this type of graph can show how the data changes over a given period of time. Other continuous data, such as the heights of a group of children on one particular day, is often grouped into categories to make it easier to interpret.

1. **What is the criteria for data to land into dimensions and measures?**

A dimension is a field that can be considered an independent variable. Dimensions in Tableau produce headers when added to the **Rows** or **Columns shelves** in the view. By default, Tableau treats any field containing qualitative, categorical information as a dimension. However, in relational data sources, the actual definition of a Dimension is slightly more complex.

A measure is a field that is a dependent variable; that is, its value is a function of one or more dimensions. Measures typically produce axes when added to the rows or columns shelves. By default, Tableau treats any field containing numeric (quantitative) information as a measure.

1. **What is Metadata, where is it present in the workbook?**

After connecting to the data source, Tableau captures the metadata details of the source like the columns and their data types. This is used to create the dimensions, measures, and calculated fields used in views. You can browse the metadata and change some of its properties for some specific requirements.

1. **What happens when you aggregate or disaggregate the Data?**

The process of viewing numeric values or measures at higher and more summarized levels of the data is called **aggregation.**

When you place a measure on a shelf, Tableau automatically aggregates the data, usually by summing it.

You can easily determine the aggregation applied to a field because the function always appears in front of the field‘s name when it is placed on a shelf.

For example, Sales becomes SUM(Sales).

You can aggregate measures using Tableau only for relational data sources. Multidimensional data sources contain aggregated data only. In Tableau, multidimensional data sources are supported only in Windows. ([Source)](http://onlinehelp.tableau.com/current/pro/online/en-us/help.htm#calculations_aggregation_aggregatingdata.html%3FTocPath%3DDo%2520More%2520with%2520Views%7CAggregations%7C_____3)

According to Tableau, **disaggregating** your data allows you to view every row of the data source which can be useful when you are analyzing measures that you may want to use both independently and dependently in the view.

1. **You are working on a dataset; the client adds in more data to the dataset. What happens to the Visualization that you had created? Give the explanation for both Live and Extracted data.**

Data blending is a method for combining data from multiple sources. Data blending brings in additional information from a secondary data source and displays it with data from the primary data source directly in the view.

Blends, unlike relationships or joins, never truly combine the data. Instead, blends query each data source independently, the results are aggregated to the appropriate level, then the results are presented visually together in the view. Because of this, blends can handle different levels of detail and working with published data sources. Blends are also established individually on every sheet and can never be published, because there is no true “blended data source”, simply blended results from multiple data sources in a visualisation.

Data blending is particularly useful when the blend relationship – linking fields – need to vary on a sheet-by-sheet basis, or when combining published data sources.

**Extracts**

**Tableau Data Extracts:**

 (tde or hyper) These are snapshots of data optimised for aggregation. These snapshots are loaded into Tableau’s memory to be quickly recalled/queried for visualisation, accordingly, the database isn’t required to build your viz.

**What is a TDE?** A TDE is a columnar store. A rudimentary definition of a columnar store is a database that stores values together within a column rather than a row. By only reading in the relevant data (columns) necessary to answer the question, the input and output necessary to query and aggregate data is largely decreased. Therefore, extracts are **often faster**(but not always!).

Extracts often have several

**benefits**:

**Speeding up the workbook through optimisation –**The extract is embedded within the workbook (see above). Thus, extracts are often much faster than live connections, especially when concerning complex visualizations e.g. large data sets or many filters. Therefore, extracts are **great for the flow of visual analytics.**

**Offline –** As the extract is embedded, your data source can be utilised when you’re offline or your network connection is poor.

**Privacy** – By hiding certain fields within your data source and then creating an extract, you can hide the fields that aren’t used, thereby ensuring data can only be viewed by the people intended.

And several

**Downsides**:

**Snapshot** – Given the nature of the extract, the data will remain the same, unless the extract is refreshed. This can be scheduled to take place (either fully or incrementally) via cloud applications such as Salesforce, Google Analytics and Google Sheets.

**Size/Structure:** Extracts can become **slow** to refresh and query depending on the data structure, e.g. many columns and many rows.

**Live connections:**

Data source that contains direct connection to underlying data. relies on database for all queries and undergoes real-time updates.

**Real-time updates –**As your viz is directly connected to the underlying data, this ensures data freshness.

**Databases are not always optimized for fast performance** (unlike extracts) – As data queries go through the database, they can only as fast as the database itself. Accordingly, working with a live connection may be slow.

**Other factors can affect speed –** e.g. Poor network speed and network traffic can slow down your workbook.

**Stress –**Live connections, especially within complex workbooks, can stress some traditional databases.

**Example of use:**A business that needs incoming sales data to make real-time decisions would require a live database connection.

1. **What are the file extensions in Tableau and how each one is different?**

* **Workbooks (.twb)**

Tableau workbook files have the .twb file extension. Workbooks hold one or more worksheets, plus zero or more dashboards and stories.  
(Saves the all the sheets and their connection information in a workbook file. The data is not included.)

* **Packaged Workbooks (.twbx)**

Tableau packaged workbooks have the .twbx file extension. A packaged workbook is a single zip file that contains a workbook along with any supporting local file data sources and background images. This format is the best way to package your work for sharing with others who don’t have access to the data.  
(Saves all the sheets, their connection information and any local resources (e.g., local file data sources, background images, custom geocoding, etc.).

* **Data Extract (.tde)**

Tableau data extract files have the .tde file extension. Extract files are a local copy of a subset or entire data source that you can use to share data, work offline, and improve database performance.

* **Data Source (.tds)**

Tableau data source files have the .tds file extension. Data source files are shortcuts for quickly connecting to data sources that you use often. Data source files do not contain the actual data but rather the information necessary to connect to the data source as well as modifications you've made in the Data pane such as default properties, calculated fields, groups, and so on.

* **Packaged Data Source (.tdsx)**

Tableau packaged data source files have the .tdsx file extension. A packaged data source is a zip file that contains the data source file (.tds) described above as well as any local file data sources such as Extract files (.tde), text files, Excel files, Access files, and local cube files. Use this format to create a single file that you can then share with others who may not have access to the original data stored locally on your computer.

* **Bookmarks (.tbm)**

Tableau bookmark files have the .tbm file extension. Bookmarks contain a single worksheet and are an easy way to quickly share your work.

**8. Filters:**

1. What are the different types of filters and give their working order?

Types of Filters and Order of Operation

There are basically, 6 types of filters and by order of operation they are:

1. Extract Filter
2. Data Source Filter
3. Context Filter
4. Dimension Filter
5. Measure Filter
6. Table Calculation Filter

Extract Filter

When you’re loading in your data you can choose to extract it, saving a snapshot of how it looks in your workbook and ultimately reducing the number of times Tableau queries the data source. To further reduce the size of the data going into Tableau, you can apply filters to the extract, which can be either by a certain dimension or measurement.

Data Source Filter

Data Source Filters reduce the amount of data being fed into Tableau and restrict what data the viewer sees. With certain access rights, the viewer can view all of the underlying data, so if not done in the data source, Data Source Filters can be used to control sensitive data. One thing that is important to note is that, Extract and Data Source Filters are not linked. So, if you change back to a live connection your Data Source Filters will still be intact.

**Context Filter**

All filters in Tableau are applied to all rows of your data without regards to other filters.

If for example, you need one filter to be applied before other filters, either for performance reasons (filter out a certain category to show Top X) or if you have a **FIXED** function in your view that needs to be filtered, making this a Context filter will make sure it is processed first. Context Filters are limited to the view but can be applied to Selected Sheets, All Using Same Data Source or All Using Related Data Source. While Context filters can improve performance, if they do not reduce the data enough (the rule is by 1/10 or more), the cost of computing them is too high to be beneficial.

**Dimension Filter**

Another name for non-aggregated filters (blue pills), such as Dimensions, Groups, Bins, Sets, etc. These are applied by both dragging them on the Filters pane or right-clicking on the specific dimension and selecting **Show Filter**.

You can choose to either only show the things highlighted or by ticking **Exclude** it will instead filter out the Dimensions selected. This will be shown by a strikethrough. If you have many dimensions you can search for it, but make sure to click **All** or **None** to select/deselect all depending on what you want to do.

In the filter dialog that pops up, there are three tabs for **Wildcard**, **Condition**, and **Top**. Here you can choose if you want to show the Top 10 (or by Parameter) by certain measurement or by a certain condition. The filters can be edited at any time by right-clicking on the pill and **Edit Filter**.

**Measure Filters**

Aggregated filters are applied after non-Aggregated filters, no matter what order they are shown on the Filters pane. When dragging it on, Tableau will ask you *how* you want to filter – in other words, what aggregation to use (**Sum**, **Avg**, **Median**, **Standard Deviation**, etc.) The second step will give you four options: **Range of values**, **At least**, **At most** and **Special**. You can choose to drag or type in the number you want to filter on. **Special** is if you want to include **Null** values or not.

**Table Calculation Filter**

A Table Calculation is the last filter applied and it applies the filters after the view has been produced. So if you want to filter the view without filtering the underlying data, Table Calculations Filters are the way to go. For example, if you are showing a reference line and don’t want this to change when using a quick filter.

**9. Dashboards & story:**

1. What are the different device type preview that Dashboards can use?

* There are four general Device Types
* Desktop
* Tablet
* Phone
* Default

**11. Sets, Parameters, Groups:**

1. Parameters can be used in?

Parameters in [Tableau](https://www.simplilearn.com/learn-tableau-tips-to-start-article) enable users to add some advanced calculations and calculated fields. Parameters provide adding a non-existing variable to the entire work and simplify the needs and requirements to analyze and [visualize](https://www.simplilearn.com/value-of-mastering-data-visualization-storytelling-for-data-scientists-article) the data.

1. What are the different ways to create a Parameter?

There are multiple possibilities of creating the Parameters in Tableau based on the user's requirements

* Top N Parameters in Tableau
* Date Field Parameters in Tableau
* Dynamic Measures
* Dynamic Dimensions

**12. Forecast:**

1. You are provided with the dataset for the past 10yrs. How can you forecast the data for next 4 years, Quarter wise?

* To generate the forecast, go to the “Analysis” menu. In the “Forecast” option, click on “Show Forecast”.
* Following the above steps will create a forecast as below. Forecast measure gets created applied over Colour in the Marks card. On the right side, the Forecast Indicator can be seen. There are two indicators Actual and Estimate. The actual indicator is the point prediction while in the case of Estimate lower and upper bounds are calculated by the algorithm based on the confidence interval e.g. 95%, 90%, etc.
* forecasts appear for the next months. The Actual is represented as a dark line, while the Estimate appears in the form of blue shade around the Actual forecast. To explore various aspects pertaining to forecast, we will go through the below steps:
* In the Forecast option, click on “Forecast Options…”. It will pop-up the Forecast Option dialogue box as below. There are various sections and options under those sections in the dialogue box. We shall see them one-by-one.

In the “Forecast Length” section, “Automatic” is by default selected. It generates a forecast for the next 12 months. “Exactly” allows us to extend the forecast for the specified number of time units, and “Until” extends the forecast to the specified point in the future. The two options Exactly and Until should be experimented by passing different values. Whenever the forecasts are generated, the description under “Show prediction” interval changes.

In the Source Data section, there is an option, aggregate by followed by a drop-down menu. We can aggregate by Years, Quarters, Months, Weeks Days, Hours, Minutes, and Seconds.

**By selecting 4 in Forecast Length (Exactly)** has generated forecasts for the next two years. Also, the description has changed. Remember in Exactly and Until we can select the Forecast length for any unit of time from Years to Seconds.