Summary – Day 6

- In real-time scenario the data is stored in different sources. e.g., Excel files, CSV files, databases etc.

- If one requires the data from all these sources, it becomes essential to merge or append the same in tableau.

- Hence there is a need to combine the data in order to solve the business problems.

- In tableau there are various ways to achieve this requirement.

- All the identified data sources are imported in tableau and then merged together based on common columns as specified by the business.

- The charts are then created on top of this merged data so formed.

- There are 4 ways to combine the data.
  - **Joins**
  - **Relationships**
  - **Blending**
  - **Union**

- **Joins**:

  - In Tableau, when one works with vast data sets, it is natural that data sets will have multiple tables with different data fields. This means that data does not reside on a single table instead, there are different tables. However, some tables might have common fields.
  - One can join such tables by means of columns that are common or related. Such related fields or identifiers used to form joins are known as Key fields or records.
  - Thus, related tables are joined or combined from the same or different sources to form a single virtual table which is an extended table having added columns from different tables. It is convenient for the users to use this joined table in data analysis works.
  - It is necessary to have the same datatype for the columns used under joins.

- o   At least one column should be common between tables that one needs to join.
- o   Joins are always performed in the physical layer of data source pane. In order to go into physical layer, one needs to right click the table in logical layer and select Open option.
- o   Types of Joins:
    - ▪   Inner Join: Only common data would come.
    - ▪   Left Join: All the data from left table and matching data from the right table
    - ▪   Right Join: All the data from right table and matching data from left table
    - ▪   Full Outer Join: All the data from left and right tables

- **Relationships**:

  - o   Relationships is more flexible way of combining the data in Tableau.
  - o   One can create a relationship between two or more data tables from multiple sources, and tableau brings in data from these tables using relationships to build a data query with the appropriate join between the tables.
  - o   When one creates relationships, tableau automatically determines the common field between these tables, the right aggregation, and how to handle nulls.
  - o   There must be at least one common field between the tables having same datatype.
  - o   Relationships are always performed in logical layer.
  - o   Relationships are faster than joins in terms of performance.
  - o   Cardinality means whether a row from one table could be linked with more than one row in another table. Below are its types.
      - ▪   One to One
      - ▪   One to Many
      - ▪   Many to One
      - ▪   Many to Many
  - o   Referential Integrity a row in one table will always have a matching row in the other table, as determined by the value of shared fields.
  - o   In Tableau, referential integrity is configured on each side of the relationship. In the Performance Options settings, Some records match means there isn't (or you don't know if there is) referential integrity. All records match means there is referential integrity. The default setting is to not assume referential integrity (Some records match).
  - o   Correctly configuring the cardinality or referential integrity settings can boost performance through query optimization. Incorrect configurations, however, can lead to issues of aggregation due to loss or duplication of data. The default Performance Option settings are Many for cardinality and Some records match for referential integrity. These should only be adjusted if you are sure of the correct characteristics of your data.

- **Blending**:

  - o   Data blending is a way to combine data in Tableau. Blending provides a quick and simple way to bring information from multiple data sources into a view. If one wants to achieve different joins on different worksheets then data blending can be used.

- o By default data blending shows left join which can later be converted to inner join by removing null values. Data blending is performed in the worksheet as against data source pane.
- o When one uses data blending to combine data, a query is sent to the database for each data source that is used on the sheet. The results of the queries, including the aggregated data, are sent back and combined by Tableau.
- o The view uses all rows from the primary data source, the left table, and the aggregated rows from the secondary data source, the right table, based on the dimension of the linking fields.
- o Whichever column is utilized first in the worksheet, the table belonging to that column becomes primary data source. It is always identified using blue tick mark. The other data source automatically becomes secondary data source, identified by orange tick mark.
- o A key difference between a data blend and a join is the order it occurs. With a join, the data is joined and then aggregated. With data blending, the data is aggregated and then joined.
- o Blending can be performed on multiple columns as well in Edit Blend Relationship.
- o If we remove the null values from left join the result we get is of inner join.
- o Linking field is used as a symbol to blend primary with secondary and it should always be on. Linking field will always be shown on secondary source matching column.

- **Union**:

  - o *Union* is used in cases where the identical set of data has been spread over multiple tables and one wants to append all such data in one big table without mentioning any common column condition.
  - o Typically, union is also preferred in cases where the tables under the SAME data source do not bear any common columns and still user wants to see the data from both tables under one chart.
  - o Union in tableau mandatorily requires the data to come from the same data source.
  - o Data coming from multiple connections can't be made into union.