**Chapter 1**

**Data Visualization**

Process of converting boring numbers and raw data into interesting graphical elements like bars, pie etc. Makes you master of story telling over insights hidden inside numbers.

**Benefits**

* Story telling which helps us to make right decision during right time.
* Tracking trends
* Making faster &quicker decisions
* Discovering unknown facts, patterns, trends
* Increasing engagement from end users

**Why tableau or BI over Excel?**

1. **Automation**

* If we made a report using excel and want to update the data, in excel we update data manually. Employee needs to sit and do it on a daily process of extracting data from source, importing on excel, do calculations, and then prepare report which is very time consuming.
* By using tableau, we can automate this process by creating schedule. E.g. we can create schedule on tableau everyday morning 6am tableau should automatically connect to the data sources, pulls the data, and prepares the reports.

Benefits:

By doing this we eliminate human errors which leads to wrong decisions and finance loss. We won’t need employees dedicated for this task for exporting and importing data on excel.

1. **Capacity**

* If sources produce massive amount of data, it is an issue, on excel we can handle around 1M records. Large data doesn’t process, and we will get errors stating data too large hence we split main file into multiple files which is hectic.
* No issues in Tableau as it is made for big data use cases and can easily handle massive amount of data. We can change connection type from extract to live to handle it.

1. **Security**

* Excel can be easily hacked as the users are used to share files on emails or store files in local computers which is not secure at all.
* Tableau provides security features like advance access control, data and network security. We don’t need to export data; we can share dashboards and report between company employees and only if access is granted, they can see data.

1. Row level Security- restricting the rows of data a certain users can see based on defined policies.

* Excel, we need to create separate files. E.g. we can’t show data of finance department with HR department.
* In Tableau we have RLS feature so we can implement using a single dashboard and giving access to different types of users based on their role.

1. **Advance Visuals**

* Not many options and too many steps in Excel
* In tableau easy and everything is automated and many options for visuals

**Types of files in Tableau**

1. **(.hpyer)**

Share only data without data source or visualizations.

1. **(.tds)- tableau data source**

Share just data source which is live, not data or visualizations.

1. **(.tdsx)-tableau packaged data source**

Share data and data source.

1. **(.twb)-tableau workbook**

Share data source which is live and visualizations.

1. **(.twbx)-tableau packaged workbook**

Share data, data source and visualizations.

**\*\*Note 1**- Using tableau desktop all these workbooks can be opened but using tableau public or reader only “**.twbx”** can be opened.

**Types of data tableau uses to store the workbook**

* **Metadata-** metadata information is stored in xml files.
* **Data**- Tableau stores in form of hyper file.

**Tableau Desktop Architecture**

1. **SOURCE**

Contains data in databases like oracle, My SQL or in files like excel, json or cloud like azure, AWS or API.

1. **DESKTOP**

**Data sources**

“Data connectors”, first we need to connect tableau to our data, offers around 90 different connectors. “Access information” is stored in the data source, e.g. path of file, username & password.

Type of connection- Extract and live.

Data model-we need to combine tables together by using relationships, joins and union. We can also set data types, renaming tables & columns.

**Visualizations**

3 levels

i)Worksheet- single view only one visual, each worksheet is directly connected to data source. We can create worksheet using two different data source using **data blending.**

ii)Dashboard- combine worksheet to form dashboard which displays multiple visual in one view.

iii)Story-sequence of worksheet and dashboards to tell users story based on data.

1. **CONSUMER**

Tableau readers- they use tableau to view and interact with visualizations. We need to send .twbx file.

Tableau desktop- team where they want to build analysis further. We need to send tableau file depending on their task.

Server consumers- they can use tableau server or cloud to view and interact with visuals. They can use web browser to access content of tableau server and can view, interact and edit visualizations depending on permissions.

Static users- extract from tableau as excel or pdf and send directly but they cannot interact with it.

1. **SERVER**

When we publish tableau workbook with extract then tableau desktop requests server to publish twbx then gateway server receives request and send to application server, the xml files are sent to repository and hyper files in file store.

**Chapter 2**

**First visualization**

* Download tableau desktop or public.
* Open tableau
* Download the datasets given in link.

A screenshot of a computer

Description automatically generated

* From left side, shown in image choose, text file.
* Use customers.csv from tableau sales dataset-small from the dataset you downloaded.
* Go to sheet1 in bottom left corner orange button.

A screenshot of a computer

Description automatically generated

**E.g.1**

* From left side drag and drop country on columns and count to rows.
* We can see how many customers we have in each country by hovering on the bar graph.

A screenshot of a computer

Description automatically generated

* File save as give the name as desired and choose .twbx and save it locally on desktop.

**\*\*Note 2- To add new worksheet, click on the square with a plus which is immediately next to sheet 1.**

**A screenshot of a computer

Description automatically generated**

**Data and Analytics pane**

* On left side pane we have data and analytics
* In data pane we have the data source and the tables and fields for the tables.
* Analytics pane can be used to add reference line, clustering etc.

**\*\*Note 3-**

* Visuals are created using rows and columns shelf.
* You can add multiple fields in rows and columns but make sure you are comparing meaningful fields.

**Title**

Double click and type the name you want to display for that particular visual.

---------------------------------------------------------------------------------------------------------------------**E.g.2**

* Open new worksheet which is sheet 2 as per note 2.
* Drag Country and city to columns & count to rows.

A screenshot of a computer

Description automatically generated

* You can remove a field from rows or columns by dragging it into the empty space or click on delete key on keyboard.

**Pages shelf**

Used to split current visual into pages. Good for step-by-step analysis.

**E.g.3**

* Drag Country to columns & count to rows.
* Drag count to pages shelf above filter on left side.

A screenshot of a computer

Description automatically generated

* On the right side we get a new window to control the pages

A screenshot of a computer

Description automatically generated

* Initially we have a bar which shows country with 1 customer. If we use the window on the right side and click on the arrow, we get countries with 2 customers and so on.

**Filters shelf**

* Drag Country to columns & count to rows.
* Drag count to pages shelf above filter on left side.
* Drag country to filter shelf.

A screenshot of a computer

Description automatically generated

* This window pops up we can choose which country we want to remove from display, we need to uncheck the box.
* Suppose we remove Germany we get the graph showing France and USA

A graph of a bar

Description automatically generated

**Show me pallet**

Located on the right side where we can choose different types of plots.

**Marks card**

1. **Color**

* Drag Country to columns & count to rows.
* Drag country to color in marks card.
* We can now see the bar graph is colored different for each country.

**A graph of a bar

Description automatically generated with medium confidence**

1. **Size**

* We can change size of marks by clicking on size in marks card and adjust slider as per the size we desire.

**A screenshot of a computer

Description automatically generated**

1. **Label**

* Click on it to add labels

1. **Detail**
2. **Tooltip**

**E.g.4**

* Drag to columns & score to rows.
* In show me pallet select pie chart
* Click on label in marks card select show mark labels

A pie chart with numbers and a pie chart

Description automatically generated

**\*\*Note 4**

At left bottom of tableau we have information such as the image below

A screenshot of a cell phone

Description automatically generated

* The graph has 3 marks which are 3 bars, we have one row of the graph and have total 5 customers.

**Chapter 3**

**Data modeling**

Creating relationships between tables

**3 types of data models**

1. **Conceptual Data model**
2. **Logical Data model**
3. **Physical Data model**

**Dimension**

Tables having information of a person or objects like employee, products are dimensions and are usually small tables.

**Fact**

Tables having events or transactions, time in it like sales, orders are facts and are usually big tables.

**Data modeling layers**

**Logical layer** we have relationships- Go to data source in tableau the default is logical layer. We can create relationships by dragging tables.

**Physical layer** we have joins and unions- In the logical layer double click on the table and it takes you to physical layer. If we drag tables from left panel tableau automatically creates joins and union.

**Viz we have data blending.**

**Joins**

**E.g.**

**A screen shot of a computer

Description automatically generated**

**E.g.5**

* Go to data source.
* Drag customers table and drop on logical layer.
* Double click on customers table to go to physical layer.
* In physical layer drag and drop orders table next to customer table. Tableau automatically creates inner join between the two tables.

A screenshot of a computer

Description automatically generated

* For modifying the join click on the join and choose which join you want
* For selecting the key which matches in both tables for e.g. in these tables it is the customer ID. We can change it by clicking on the arrow next to customer ID and choosing.

A screenshot of a computer

Description automatically generated

**Unions**

Combines rows of 2 tables. For this both tables must have same number of fields and fields must have same number of data types.

**E.g.**

A hand holding a brick

Description automatically generated

**E.g.6**

* Go to data source, drag orders table to the logical layer and double click to go to the physical layer.
* Drag Order\_Archive right below orders in physical layer, it will show union, we can place the table there, so it creates union.
* We can check in data we get new field called table name where it shows records are from Orders.csv and Orders\_Archive

**Relationships**

**E.g.7**

* In data source page drag orders table to the logical layer.
* Drag customers table next to the orders table on same line, tableau creates a relationship.
* Scroll down to see which keys are connected and the relationship between the tables, we can modify the relationships and add more fields to the keys matched between tables.

A screenshot of a computer

Description automatically generated

* If not sure leave the relationship as many to many cardinality.
* Choose many if you have duplicates in table else choose one.

**Data blending**

Method of combing data and visualization levels from two different data sources using a left join. Unique feature in Tableau, not found in any BI tool.

E.g.8

* We need two data sources, first one is csv, so choose text file on left panel and use the products.csv file as provided.
* For adding another data source click on the cylindrical icon as shown in image.

A screenshot of a computer

Description automatically generated

* Second data source is the products.json file. We need select the schema, check the box for data.
* For switching between the data sources click on the cylinder we can choose the data source we want to use
* We can do data blending in the worksheet by creating visualizations.
* On the left pane we can see two data sources and can switch between them to select the tables from them.
* Drag product ID and drop on rows, Products becomes the primary data source and has a blue tick on left pane.
* Click on data, edit blend relationship, if automatic tableau decides, else we need to specify which key are common for doing this we need to choose custom, and double click on the key which is as default and can choose.