|  |  |
| --- | --- |
| BUSINESS ANALYTICS SKILLS FOR THE FUTURE-PROOFS SUPPLY CHAINS | **BUSINESS INTELLIGENCE**  **Tableau desktop**  Exercsises handout |

# Exercise 1

## Introduction to Tableau and data preparation

**Creating a new workbook and connecting with data**

1. In the menu on the left, select the type of data set you want to use. In this case, choose **To a File > Microsoft Excel**. Load the **Global Superstore.xls** file.
2. The initial Tableau window opens with the **Data Source** tab open (at the bottom of the screen).
   * On the left is the **Connections** panel. The name of the file (Global Superstore) is visible here – by clicking on the name, the file can be renamed. Below the names are the worksheets (tables) from Excel that you have uploaded.
     + Excel worksheets are treated in the same way as tables in databases and it is possible to select one table or join several tables
     + To select a worksheet with the data on which you want to make visualizations, simply drag the worksheet (in this case **Orders**) into the white area (**Drag tables here**)

A red arrow pointing to a white background

Description automatically generated

* In the lower part of the screen, you will get a preview of the data contained in the data set:
  + To change the name of the columns, you need to click on the arrow that appears when you move the mouse over the column and select **Rename**.
  + To change the **data type** of a specific column, click on the icon above the column name (Abc, #...)
* In the upper right, two possible ways of connecting to the data set are visible: **Live** and **Extract**.
  + Live connection leaves the data in the database or source file (in this case Excel). This method is fast and changes are made to the visualizations instantly.
  + By extracting data, it is placed in Tableau's built-in memory. This mode is suitable for connecting to slower data sources.

A screenshot of a computer

Description automatically generated

1. At the bottom of the screen, click on the **Sheet 1** tab to create visualizations.

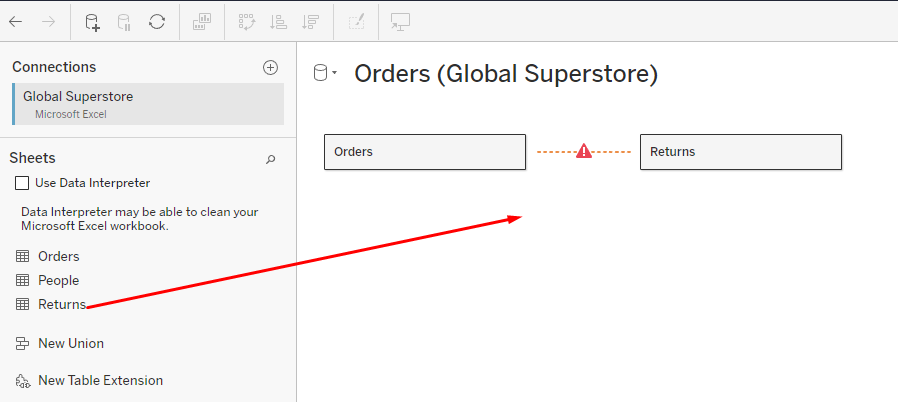
A screenshot of a computer

Description automatically generated

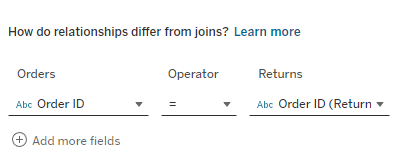
### **Connecting more tables (worksheets)**

Sometimes it is necessary to visualize data from several different tables (worksheets). To add attributes from other tables of the same data source (eg other worksheets of the Global Superstore file), you need to modify the existing data connection.

1. Open the **Data Source** tab
2. Drag the **Returns** worksheet to the right



1. In the lower part, it is necessary to determine which are the (common) fields through which these two worksheets will be connected
   * For both worksheets it will be the **Order ID** field



1. Returning to **Sheet 1**, fields from both worksheets (Orders and Returns) are visible on the left side.

**Setting up and formatting data**

Data preparation is a very important part of successful visualization. The data is very often not formatted in an ideal way and it is necessary to prepare it so that it is suitable for visualizations.

In this section, the **Flight Data.xls** file will be used, which shows the number of resolved monthly incidents per employee. Open the file and examine how the data is structured. The **Ideal** worksheet shows how the data should look. Tableau has built-in capabilities to transform data to be ready for analysis.

1. Create a new workbook and load the **Flight Data.xls** file located on your computer
2. Drag the **Resolved Incidents** worksheet to the right

A screen shot of a computer

Description automatically generated

In the data preview, you can see certain problems - there are no column names, headers from Excel have many null fields, etc. Tableau has a built-in tool for editing badly formatted data and preparing data for analysis called **Data Interpreter**.

1. Click on **Use Data Interpreter**
   * Null values ​​are now removed and columns are correctly identified
   * If you want to see in more detail what the data interpreter did, you can click on **Review the results**. An Excel file will open with a description of all the changes made.

### **Pivoting**

Ideal data should be longer, with more rows and narrower, with fewer columns. In order to achieve this, it is necessary:

1. Mark all date columns - click on the first column, hold the SHIFT key and mark the last column
2. Right-click and select **Pivot**. You now have data with 3 columns and a large number of rows

A screenshot of a computer

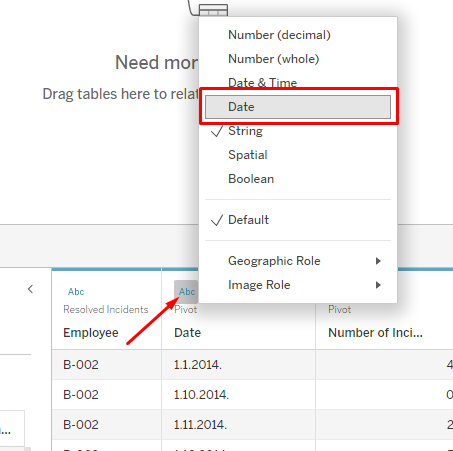
Description automatically generated

1. Rename the Pivot Field Names column to **Date**, and the Pivot Field Values ​​column to **Number of Incidents** (right click and Rename)

### **Spliting columns**

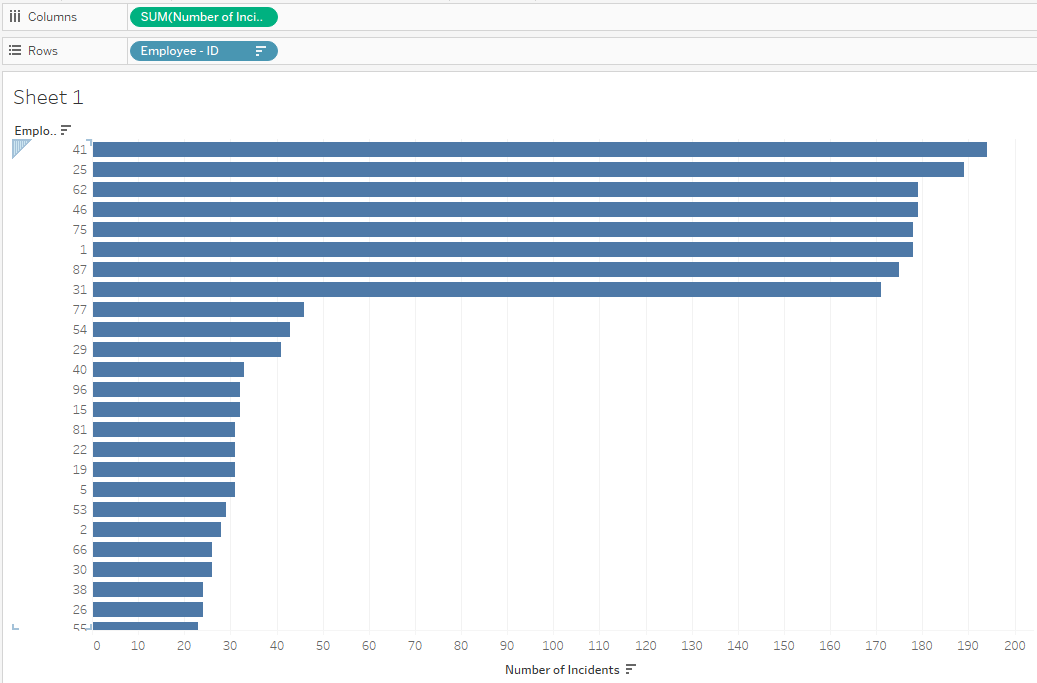
Notice how the "Employee" field consists of two pieces of data - the location code (A, B, C, D, E) and the employee code. That column can be split into two columns, since it has a separator (hyphen).

1. Right-click on the **Employee** column and select **Split**. You got two new columns - Split 1 and Split 2
2. Name the Split 1 field **Employee Location**, and the Split 2 field **Employee ID**
3. It can be seen that the Date field was marked by Tableau as text, but it should be a date field. Click on the field type label and change it to **Date**.



### **Visualisation example**

1. Open the **Sheet 1** tab and create a visualization that will show the number of resolved incidents for each employee (according to the employee code)
   * Drag the **Employee** **- ID** field to Rows, and the **Number of Incidents** field to Columns. Sort the data by size.



1. From the above visualization, it is obvious that there is a part of employees who have a very large number of resolved incidents. On the other hand, there is another part of the employees that has a rather low number of resolved incidents. It is possible to make the first part of the employee stand out visually:
   * Create a new field by right-clicking on the white area on the left (where the fields are) and select **Create Calculated Field**
   * Under the name of the new field, enter **Employee Rank**, and the formula should be:  
     IF SUM([Number of Incidents])>50 then 'good' ELSE 'bad' END

A screenshot of a computer

Description automatically generated

* + A newly created **Employee Rank** field appeared in the fields on the left. Drag that field to the **Color** tab
  + There is a legend on the right. Change the color of the 'good' mark in red, and the 'bad' mark in gray (by double-clicking on the color).
  + Click on the **Label** mark and check **Show mark labels**

A graph of numbers and a number of objects

Description automatically generated

## Dimensions and measures; Continuous and discrete data

### Tableau classifies each field into a dimension or a measure, depending on the type of data.

### **A screenshot of a computer Description automatically generatedDimensions and Measures**

**Dimensions** are fields used to group or break down data. Dimensions usually (but not always) contain qualitative values ​​(eg names, dates, geographic data, etc.) and affect the level of detail in the visualization.

**Measures** are usually numerical (quantitative) values ​​on which various calculation operations can be performed. The value of the measure depends on the context of the dimensions. For example, the sum of shipping costs is different if there is no dimension (if only all shipping costs are added up) and if the dimension is Shipping Priority (then shipping costs will be displayed for each individual shipping priority).

In Tableau, dimensions and measures are separated by a horizontal line.

At the bottom of the dimensions and measures are some predefined fields, created by Tableau. This is, for example, the **Orders (Count)** field, which serves us if we want to see how many rows (in this case, orders) are in a certain dimension.

1. Open Tableau, load the **Global Superstore** file (**Orders** worksheet)
2. Open a new worksheet and name it **Orders by Category**
3. Move the **Category** field to rows, and the **Orders (Count)** field to columns

The total number of orders by category is now displayed.

A screenshot of a computer

Description automatically generated

1. Open a new worksheet and name it **Dimensions and Measures**
2. Move the **Category** field into columns.

The Category field is the dimension by which the data was broken down (in this case into 3 categories).

1. Drag the **Profit** field to rows.

After adding the dimension and measure, a bar chart was created. A bar graph requires one dimension and one measure.

1. Open **Show Me** in the upper right corner and explore the different types of charts that can be made with one dimension and one measure

### **Continuous and discrete fields**

All fields have an icon on the left side that indicates the type of field (string, integer, date, etc.), and the icon is marked in blue or green. Green fields are **continuous** fields that represent continuous values ​​(without any breaks). The blue fields are **discrete** and represent values ​​that are unique and shown as individual.

1. Open a new worksheet and name it **Profit**
2. Drag the **Order Date** field into columns to get a table with years in the header. Adding a continuous variable to the rows will create an axis on the graph
3. Drag the **Profit** field to rows

A graph with a line

Description automatically generated

1. Open a new worksheet and name it **Shipping Cost Analysis**
2. Drag the Order Date field into columns to get a table with years in the header. If some discrete variable is added to the rows, it will create a table
3. Drag the **Market** field to rows
4. Now it is necessary to add some continuous variable to the table to display the values. Take the **Shipping Cost** field and drag it to the mark **Abc**

A screenshot of a computer

Description automatically generated A screenshot of a shipping data

Description automatically generated

### **Data filtering**

By filtering, users can find the necessary information more precisely. The result of filtering differs depending on the type of data.

1. Open the **Profit** worksheet
2. By filtering discrete fields, all possible values ​​are obtained by which it is possible to filter.   
   Drag the **Market** field into the **Filters** section.

A screenshot of a computer

Description automatically generated

1. In the window that appears, select **All** (or mark all possible values) and click **OK**
2. The **Market** field is visible in the Filter section. Right-click on that field and select **Show filter**. A filter appeared on the right side of the screen with all possible values ​​by which it is possible to filter data

A screenshot of a computer

Description automatically generated

1. Open the **Shipping Cost Analysis** worksheet
2. Filtering continuous fields provides a range of values ​​that can be filtered. Drag the **Discount** field into the **Filters** section.
3. In the window that appears, select **Next** and **OK**
4. In the Filter section, the Discount field is visible. Right-click on that field and select **Show filter**. A filter appeared on the right side of the screen with a range of values ​​by which it is possible to filter the data

A screenshot of a computer

Description automatically generated

### **Spatial data**

1. Open a new worksheet and name it **Profit by State**
2. Double-click on the **Country** field to create the map. Drag the **Profit** field (continuous field) to the **Color** mark and change its display type to **Average**:

A screenshot of a map

Description automatically generated

You will get a choropleth map showing the average profit by state.

A map of the world

Description automatically generated

1. Open a new worksheet and name it **Regions**
2. Double-click on the **Country** field to create the map. Drag the **Region** field (discrete field) to the **Color** mark. You will get a map with marks for each individual region

A screenshot of a computer screen

Description automatically generated

### **Continuous and discrete dates**

Date fields are specific in that they can behave both as continuous values ​​and as discrete values. If there is a blue date in the place of a column or row, then it is a discrete field, and if it is green, it is a continuous field.

1. Open the **Profit** worksheet. It can be seen that this is a discrete date (blue color). This means that each part of the date is treated as a category.

A screenshot of a computer

Description automatically generated

1. The date can be further broken down by clicking on the **'+' sign** to the left of the name. In this way, the date will be broken down into quarters (for each year), and then into months. Each part of the date (year, quarter, month...) is treated as an individual category. This way, for example, sales can be displayed by month or quarter.

A screenshot of a graph

Description automatically generated

1. Remove quarters and months from the columns (right click on **Quarter and Month > Remove**) to get only the profit display by years (the initial display)
2. Right-click on the **YEAR (Order Date)** field located in the columns. The date parts are divided into two sections. The first section (marked in blue in the image) represents discrete parts of the date, while the second section (marked in green in the image) represents continuous parts of the date. Select the year from the second, green, section.

A screenshot of a computer

Description automatically generated

1. Now the date line has become continuous and the date field in the columns is colored green. When you click on the '+' sign to drill down on a date, no additional fields for quarters and months will be created, but the drill down will be done on the current axis.

A graph on a white background

Description automatically generated

## Dashboards and storytelling

A dashboard is a collection of worksheets and other information displayed on one screen. When creating a dashboard, it is necessary to pay attention to the following:

* It is necessary to know who your audience is, to whom will you present the information?
* The most important content on the dashboard should be located in the upper left corner
* Limit the number of visualizations on the dashboard (no more than 5)
* Use filters to make the dashboard more interactive
* Emphasize important information

1. Open the Tableau file **Tableau - dashboard**
2. Create a new dashboard and call it **Sales KPI**
3. At the top, in the **Dashboard** menu, under the **Device Layouts** option, add **Desktop**

A screenshot of a computer

Description automatically generated

1. In the panel on the left, select **Default**
2. At the top of the menu, click on **Format > Dashboard**. In the panel on the left, change the **Dashboard Shading** to light gray

A screenshot of a computer

Description automatically generated

1. Close the **Format Dashboard** panel by clicking the **X**
2. Drag the **Sales by Salesperson** worksheet to the dashboard

A red arrow pointing to the right

Description automatically generated

1. Move the **Sales by delivery type** worksheet below that visualization

A screenshot of a graph

Description automatically generated

1. Move the **Sales by years** worksheet to the right of that visualization

A screenshot of a computer

Description automatically generated

1. Drag the **Text** object to the upper left corner

A screenshot of a computer

Description automatically generated

1. In the window that opened, type the text **Superstore sales** and increase the font size to 22
2. Adjust the size of the **Text** object so that it occupies approximately 1/3 of the width of the dashboard

A screenshot of a computer

Description automatically generated

After dragging all the worksheets, a dashboard was created. After creating the first initial appearance of the dashboard, it is now necessary to further customize it. By deleting a visualization from the dashboard, the appearance of the dashboard will automatically change. However, replacing one visualization on the dashboard with another will not affect the appearance of the dashboard.

1. Click on the chart **Sales by delivery type**
2. In the panel on the left, hover over the **Top 10 countries by sales** worksheet
3. Click on the curved arrow on the right side of the worksheet

A screenshot of a computer

Description automatically generated

### **Dashboard interactivity**

Depending on the context of the dashboard, it is possible to give the audience the ability to dig deeper into the data.

Question #1: What are the top 3 countries in terms of product sales in the 'Technology' category?

1. Click on the visualization **Sales by years**
2. Click on the arrow on the right. In the menu that opens, select **Use as Filter**

A screenshot of a graph

Description automatically generated

By selecting this option, the filter is also applied to other visualizations. When you click on a line on the line graph Sales by years, the other visualizations will also change according to the selection. You can do the same by clicking on the **funnel** (filter) icon.

After selecting the green line that indicates the Technology category, it can be seen that California, New York and Pennsylvania have the highest sales.

1. Using the **Blank** object, edit the dashboard by adding blank space between the visualizations to make them clearer

### Tableau story

Tableau story combines worksheets and dashboards with the goal of telling a story about visualizations.

Question #1: What is the sales situation in the last 5 years

1. Create a new story by clicking on the last icon at the bottom of the screen and name it **Superstore sales**
2. Drag the **Sales by Salesperson** sheet to the first story point
3. Change the title of the story point so that it says "3 out of 4 referrals have sales of more than $500,000"
4. Drag the **Sales by years** worksheet to the right of the name of the first story point and release
5. Change the name of the second story point to "Sales in all categories have increased over the years"

A screenshot of a computer

Description automatically generated

1. Drag the worksheet Top 10 states by sales to the right of the name of the second story point and release
2. Change the name of the third story point to "The lowest sales were achieved in the states of Virginia and Michigan"

Within each story point, it is possible to add annotations:

A graph of a line

Description automatically generated with medium confidence