Data Analytics Immersive

Introduction to Tableau: Labs
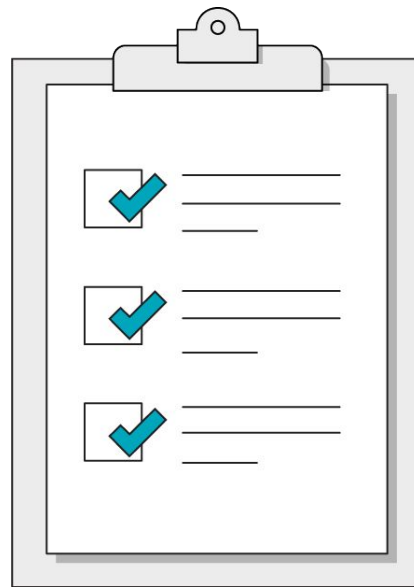
Unit: Data analysis and communication with Tableau and Power BI

GENERAL ASSEMBLY

Learning Objectives

In this lesson, we'll:

- <mark>Practice</mark> the foundational Tableau skills we have learned so far on two data sets:

  - A company's HR attrition data

  - World Cup soccer game results

Any questions?

Agenda

● Mini Lab 1: Investigating the attributes of people leaving a company

● Mini Lab 2: World Cup results

● Wrap up and Q&A

Introduction to Tableau: Labs

Mini Lab 1: Investigating the Attributes of People Leaving a Company

GA

What Factors Are Different in People Who Leave?

Your company wants to launch a new employee retention program and as a part of this project they have tasked you with the following:

1. Review HR data and determine how factors such as education, marital status, and years since last promotion differ between those who have left the company and those who are still employed there.

2. Visualize how the characteristics of individuals who leave the organization differ from those who stay.

You are going to create some graphics of employee attrition data.

1. Load and prepare the data.
2. Walk through how to create some charts that we haven't yet tried out.
3. Time for you to explore the data and try out visualizations.

Output:

A Tableau file containing at least ten graphs showing insights into the HR attrition data.

1. Open Tableau.
2. Create a new Tableau file.
3. Load the **HR_Attrition.csv** data.
4. Explore the data and have a look at the **data dictionary***, which defines what each field represents.
5. What are the limitations of the data? What is missing, or what don't you understand? What assumptions might you need to make?
6. Think about what the main drivers of attrition might be. Create a list of some graphics that you think it would be interesting to see.

* The data is originally from [here](#), and the data dictionary is from [here](#).

Load and Prepare the Data (cont.)

Name	Description
AGE	Numerical Value
ATTRITION	Employee leaving the company (0=no, 1=yes)
BUSINESS TRAVEL	(1=No Travel, 2=Travel Frequently, 3=Tavel Rarely)
DAILY RATE	Numerical Value - Salary Level
DEPARTMENT	(1=HR, 2=R&D, 3=Sales)
DISTANCE FROM HOME	Numerical Value - THE DISTANCE FROM WORK TO HOME
EDUCATION	Numerical Value
EDUCATION FIELD	(1=HR, 2=LIFE SCIENCES, 3=MARKETING, 4=MEDICAL SCIENCES, 5=OTHERS, 6= TECHNICAL)
EMPLOYEE COUNT	Numerical Value
EMPLOYEE NUMBER	Numerical Value - EMPLOYEE ID
ENVIRONMENT SATISFACTION	Numerical Value - SATISFACTION WITH THE ENVIRONMENT
GENDER	(1=FEMALE, 2=MALE)
HOURLY RATE	Numerical Value - HOURLY SALARY
JOB INVOLVEMENT	Numerical Value - JOB INVOLVEMENT
JOB LEVEL	Numerical Value - LEVEL OF JOB
JOB ROLE	(1=HC REP, 2=HR, 3=LAB TECHNICIAN, 4=MANAGER, 5= MANAGING DIRECTOR, 6= RESEARCH DIRECTOR, 7= RESEARCH SCIENTIST, 8=SALES EXECUTIVE, 9= SALES REPRESENTATIVE)

JOB SATISFACTION	Numerical Value - SATISFACTION WITH THE JOB
MARITAL STATUS	(1=DIVORCED, 2=MARRIED, 3=SINGLE)
MONTHLY INCOME	Numerical Value - MONTHLY SALARY
MONTHLY RATE	Numerical Value - MONTHLY RATE
NUMCOMPANIES WORKED	Numerical Value - NO. OF COMPANIES WORKED AT
OVER 18	(1=YES, 2=NO)
OVERTIME	(1=NO, 2=YES)
PERCENT SALARY HIKE	Numerical Value - PERCENTAGE INCREASE IN SALARY
PERFORMANCE RATING	Numerical Value - ERFORMANCE RATING
RELATIONSHIP SATISFACTION	Numerical Value - RELATIONSHIP SATISFACTION
STANDARD HOURS	Numerical Value - STANDARD HOURS
STOCK OPTIONS LEVEL	Numerical Value - STOCK OPTIONS
TOTAL WORKING YEARS	Numerical Value - TOTAL YEARS WORKED
TRAINING TIMES LAST YEAR	Numerical Value - HOURS SPENT TRAINING
WORK LIFE BALANCE	Numerical Value - TIME SPENT BETWEEN WORK AND OUTSIDE
YEARS AT COMPANY	Numerical Value - TOTAL NUMBER OF YEARS AT THE COMPANY
YEARS IN CURRENT ROLE	Numerical Value -YEARS IN CURRENT ROLE
YEARS SINCE LAST PROMOTION	Numerical Value - LAST PROMOTION
YEARS WITH CURRENT MANAGER	Numerical Value - YEARS SPENT WITH CURRENT MANAGER

A note on the data source

Please use the HR_Attrition.csv file rather than downloading the file from Kaggle as we have modified it slightly to make it easier to analyse.

For those of you who are interested, the differences are:

- We have changed the order of 2 columns: `EmployerNumber` and `Gender`
- We have removed `EmployeeCount` as we don't need it for this analysis (it has a value of 1 for every row)
- We have added 2 columns: `StartDate` and `TermDate`, which give the start date and termination dates for employees. This is to help with visualisations.

The total number of rows of data is the same (1471 excluding header row), and all other data is exactly the same in the two files.

Let's create a simple bar chart that visualizes the average incomes of people who have left the company as well as people who have stayed.

1. Drag the Dimension 'Attrition' into Columns.
2. Drag the Measure 'Monthly Income' into Rows.
3. Change the aggregation for the measure from 'SUM' to 'AVG.'
4. Format the chart (change the title, format the text, change the size of the chart to fill the canvas, show labels of the categories and format them to be $, remove the y axis tick marks).

Let's create a pie chart that visualizes the biological gender breakdown of both individuals who have left the company as well as people who have stayed.

1. Drag the Dimension 'Attrition' into Columns.
2. Using the marks card change the visualization into a 'Pie.'
3. Drag the Dimension 'Gender' onto Color in the Marks card.
4. Drag Dimension 'Employee Number' onto Angle in the Marks card and change the dimension into a Measure(Count).
5. Drag Dimension *HR_Attrition.csv(Count)* onto Label in the Marks card, and then using Quick Table Calculations change to 'Percent of total.'
6. Go to Analysis, Percentage Of and Column.
7. Format the chart (change the title, change the colours, format the labels, show labels of the categories and format them to be $).
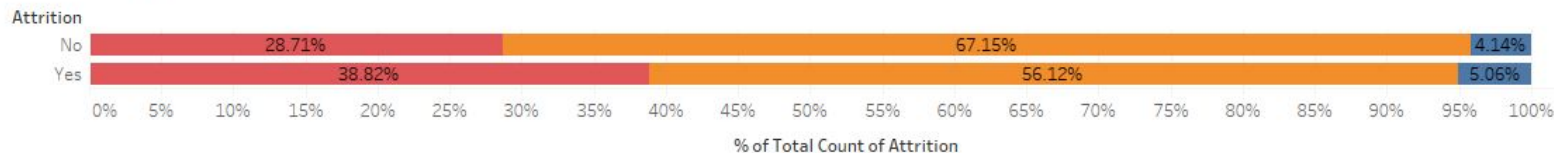
Create a bar chart that calculates the percentage of employees in each department for both employees that have left and those that have stayed. Use color to differentiate departments and make sure that the bar is labeled with percentages. Your completed product should look like this:
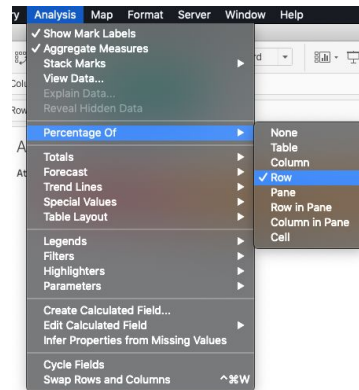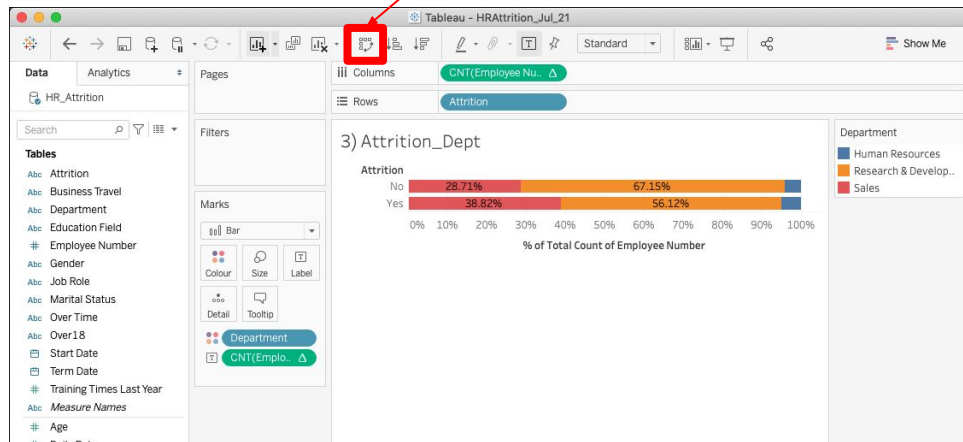
Chart 3: Attrition Split by Department

Swaps Columns and Rows

Investigate the data and charts, and make more to show your insights.

You should aim to have a workbook of at least ten different graphs before finishing, as we will use this workbook for a future lab.

You will need to submit your workbooks for this and the next Lab at the end of class.

Introduction to Tableau: Labs

Mini Lab 2: World Cup Results

You have started working for a sports analytics company. They want you to look at the World Cup results since the tournament began and share your insights. To get started:

1. Load the **Soccer_World_Cup_results.csv** data file into a new workbook.
2. Investigate the data and show some insights from it. Visualize the following:
   - Which has been the highest scoring tournament so far?
   - Which have been the top-scoring countries so far?
   - Which countries have got to the Final the most?
   - Which countries have reached the knockout stage the most? (Knockout stage consists of Round of 16, Quarter Finals, Semi Finals and Finals.)
   - *Optional:* Any more insights that you find and think are interesting

Hint: You may need to create a Calculated Field for some of your data. We'll cover it in a future lesson, but you can read up about them here.
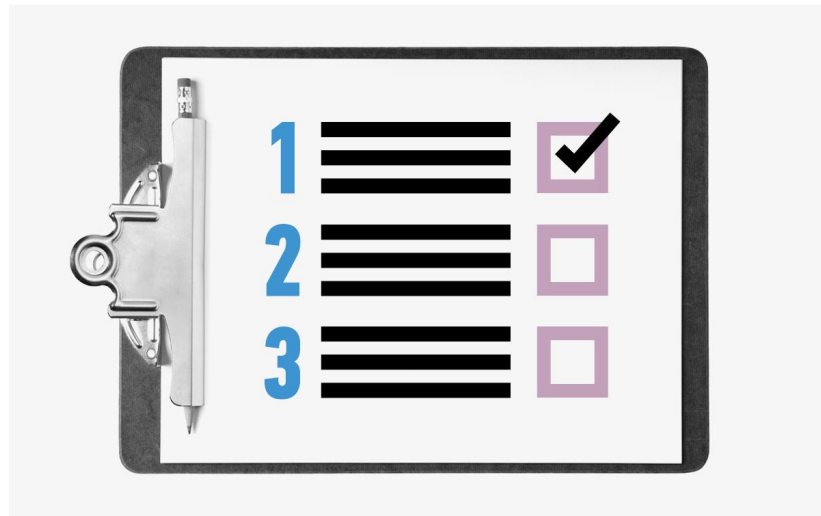
Introduction to Tableau: Labs

Wrapping Up

Recap

Today, we:

- Practiced using the foundational Tableau skills we have learned so far on two data sets:

    - A company's HR attrition data

    - World Cup game results

Q&A