People Analytics: Employee Absenteeism

Problem Description:
You and your team are working for a company in Canada called *Canada Stores Limited*. The company owns chain stores in multiple locations, selling products in categories such as Bakery, Dairy, Processed food, and Meat.

The HR department approached you to help them solve the issue of absenteeism at work. For HR, understanding absenteeism and its reasons can help HR effectively do its job. Also, it saves the organization enormous expenses.

Your task is to assist the HR department in understanding and predicting the factors that influence the levels of employee absenteeism (i.e., predictors of absence).

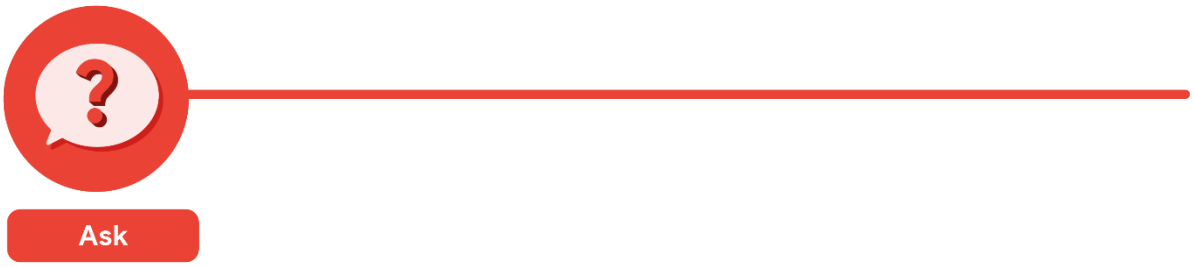
Data Description:
Each observation in the dataset represents an employee. There are 8335 observations and 12 columns in the dataset.

Variable	Type	Description
EmployeeNumber	Numeric	Employee ID number
Surname	Text	Employee last/family name
GivenName	Text	Employee first name
Gender	Categorical	Employee Gender (M = Male; F = Female)
JobTitle	Text	Employee job title
DepartmentName	Categorical	Department to which the employee belongs
StoreLocation	Text	Location of the store with which the employee works
Division	Categorical	Business divisions that contain one or many departments
Age	Numeric	Employee age (year)
LengthService	Numeric	Employee tenure in the organization (years)
AbsentHours	Numeric	Total number of absent hours for every employee
BusinessUnit	Categorical	Organization units (Head Office and Stores) that contain divisions.

Deliverables:

- The main deliverables are:
    1. R code with understandable comments. The code should be written in a clear matter. Use self-explanatory names for the variables (e.g., do not use x or df to name variables or data structures).
    2. A complete report that follows the instructions below.
- The dataset, R code, and report must all be part of an R project.
- I will ask each group to report their progress every week after the lab. You must show evidence of progress.
    - A member representing each group must come to the office every Wednesday and show the progress in R Code and report writing.
    - Not showing up will deduct 0.5% of the project grade every time a group misses a Thursday briefing.
- The submission deadline is **the End of Week 14**.
- Presentations will be in Week 15.

Instructions:
Your role as an analyst requires that you turn the issue into action. There are six business analysis phases that will help you make seamless decisions: ask, prepare, process, analyze, share, and act. You are asked to follow these steps in your report. Let's walk through the steps to see how they can help you solve problems.



Step 1: Ask
It's impossible to solve a problem if you don't know what it is. These are some things to consider:

- Define the problem you're trying to solve
- Make sure you fully understand the stakeholder's expectations
- Focus on the actual problem and avoid any distractions
- Collaborate with stakeholders and keep an open line of communication
- Take a step back and see the whole situation in context

Questions to ask yourself in this step:

1. What are my stakeholders saying their problems are?
2. Now that I've identified the issues, how can I help the stakeholders resolve their questions?

Step 2: Prepare
You will decide what data you need to collect in order to answer your questions and how to organize it so that it is useful. You might use your business task to decide:

- What metrics to measure
- Locate data in your database
- Create security measures to protect that data

Questions to ask yourself in this step:

1. What do I need to figure out how to solve this problem?
2. What research do I need to do?

Assume that the dataset provided to you satisfies this step in the business analysis job. But read more about the U.S. Census Bureau, how they collect and prepare data, and what data the Bureau usually collects.

Step 3: Process

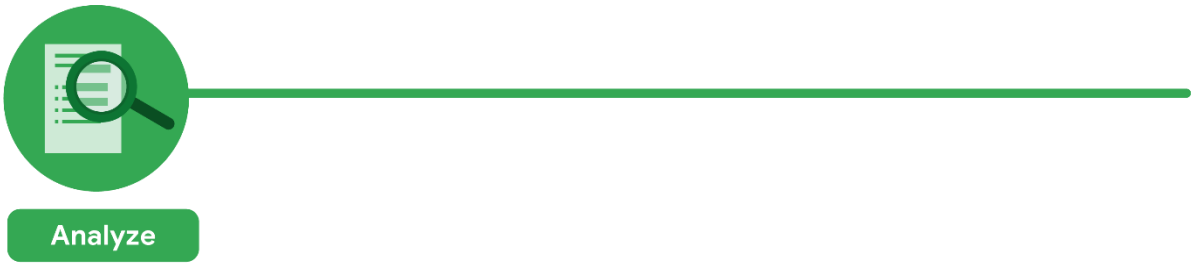
Clean data is the best data, and you will need to clean up your data to get rid of any possible errors, inaccuracies, or inconsistencies. This might mean:

- Using spreadsheet functions to find incorrectly entered data
- Using SQL functions to check for extra spaces
- Removing repeated entries
- Checking as much as possible for bias in the data

Questions to ask yourself in this step:

1. What data errors or inaccuracies might get in my way of getting the best possible answer to the problem I am trying to solve?
2. How can I clean my data, so the information I have is more consistent?

Since this course is about learning how to use R in business analytics, learn how to use it in process and cleaning data.



Step 4: Analyze

You will want to think analytically about your data. At this stage, you might sort and format your data to make it easier to:

- Perform calculations
- Combine data from multiple sources
- Create tables with your results

Questions to ask yourself in this step:

1. What story is my data telling me?
2. How will my data help me solve this problem?
3. Who needs my company's product or service? What type of person is most likely to use it?

You are expected to perform all the needed analyses:

- Run all the needed Descriptive Analytics. Make sure that you interpret the results you get.
- Use Tableau for data visualizations.
- To estimate the median housing price, make sure that you use the appropriate statistical estimation technique.
- As for the predictive analytics part:
  - Build a model of housing prices to predict median house.
  - Train the model to learn from the data to predict the median housing price.
  - Predict housing prices.

Step 5: Share

Everyone shares their results differently, so be sure to summarize your results with clear and enticing visuals of your analysis using data viz tools like graphs or dashboards. This is your chance to show the stakeholders you have solved their problems and how you got there. Sharing will certainly help your team:

- Make better decisions
- Make more informed decisions
- Lead to stronger outcomes
- Successfully communicate your findings

Questions to ask yourself in this step:

1. How can I make what I present to the stakeholders engaging and easy to understand?
2. What would help me understand this if I were the listener?

Step 6: Act

Now it's time to act on your data. You will take everything you have learned from your data analysis and put it to use. This could mean providing your stakeholders with recommendations based on your findings so they can make data-driven decisions.

Questions to ask yourself in this step:

1. How can I use the feedback I received during the Share phase (step 5) to actually meet the stakeholder's needs and expectations?

These six steps can help you to break the data analysis process into smaller, manageable parts, which is called **structured thinking**. This process involves four basic activities:

1. Recognizing the current problem or situation
2. Organizing available information
3. Revealing gaps and opportunities
4. Identifying your options

When you are starting out in your career as a business analyst, it is normal to feel pulled in a few different directions with your role and expectations. Following processes like the ones outlined here and using structured thinking skills can help get you back on track, fill in any gaps and let you know exactly what you need.