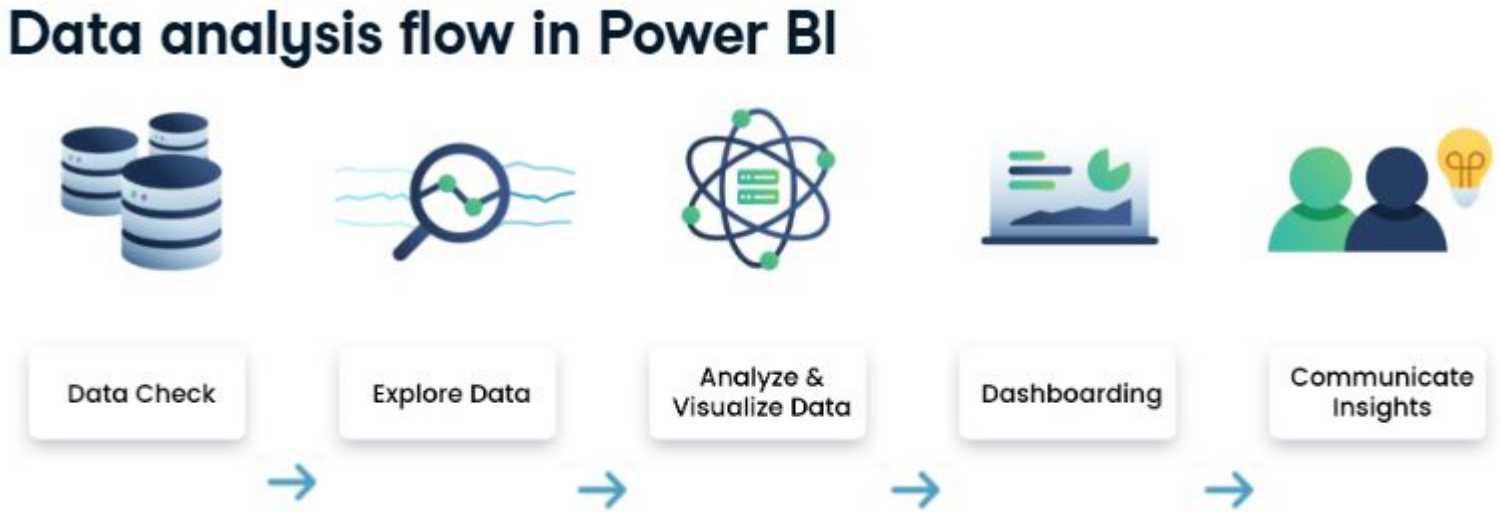
End to End with PowerBI Part I

Dr. Lee

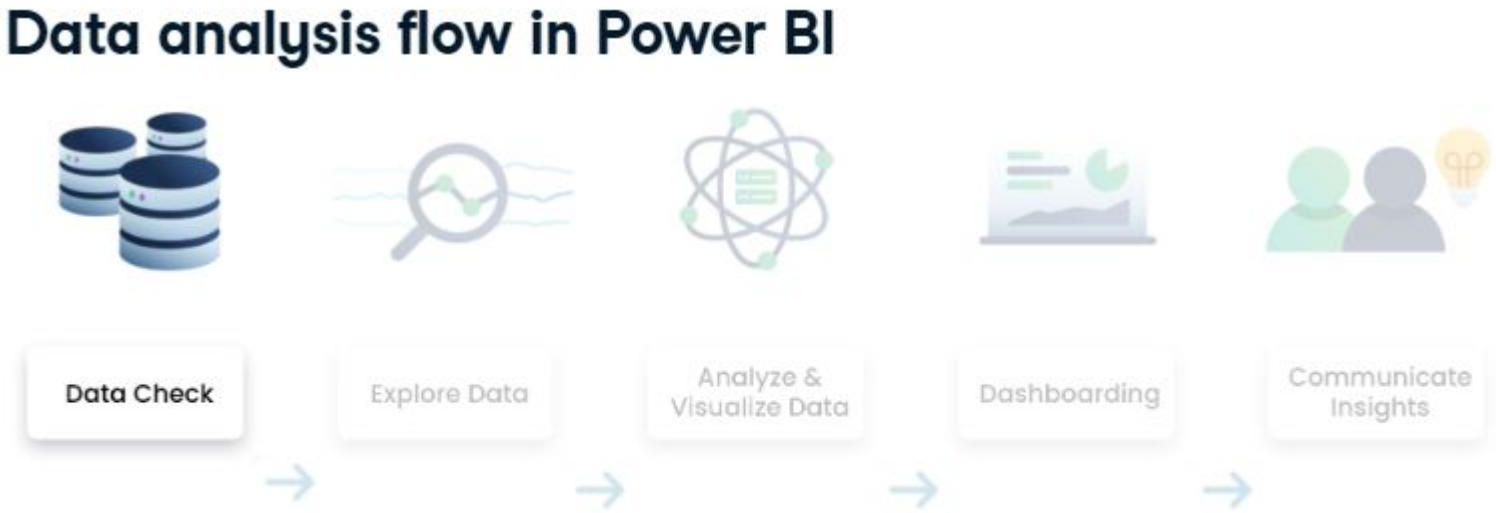
What is a case study?

- Apply your skills
- No new concepts
- Solve a real-world problem

Data Analysis with Power BI

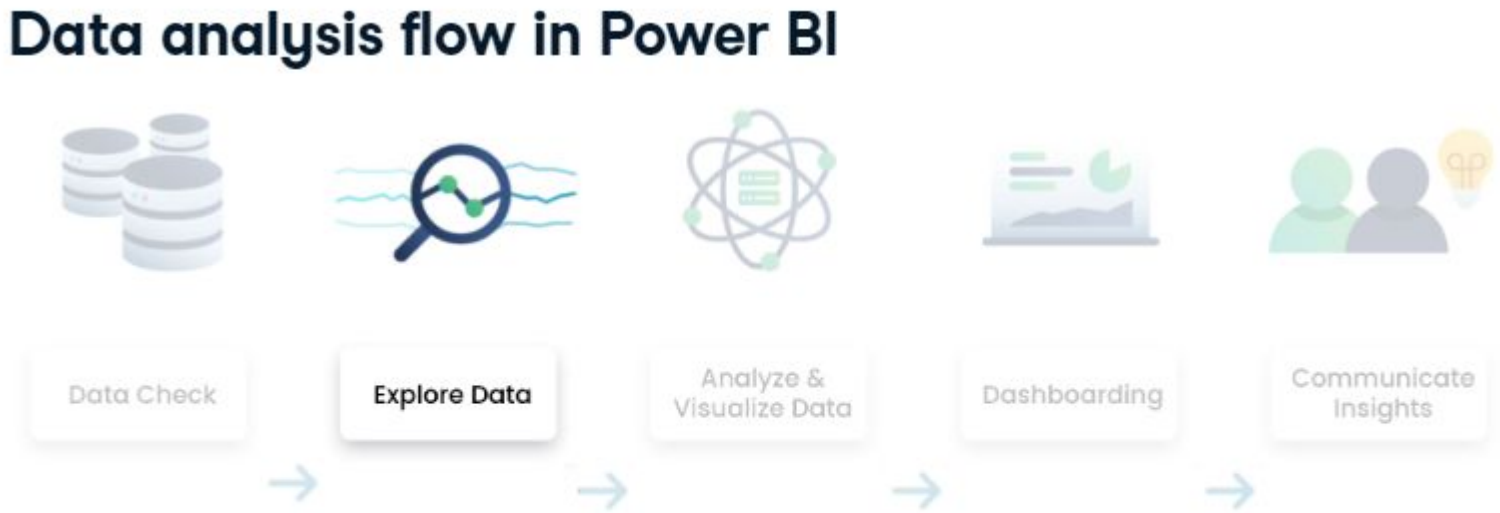


Data Check



- Check for duplicate or missing values
- Do a sense check with other internal data sources

Explore Data



- Ask yourself the right questions

Data analysis flow in Power BI

- Choose the right visualization to convey a message
- Perform more advanced analysis

Data analysis flow in Power BI

- Combine visualizations in one or more dashboards

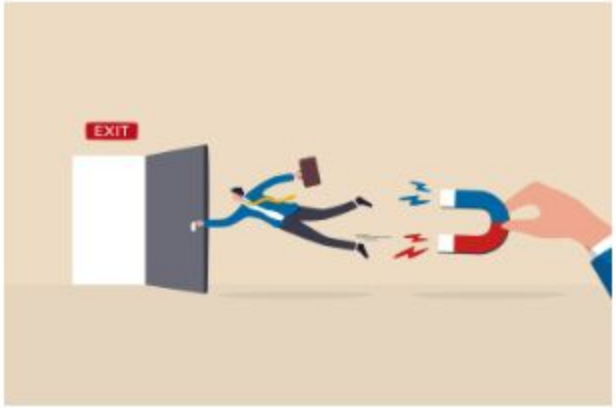
Data analysis flow in Power BI

- Communicate your insights to stakeholders

The problem

Solving customer churn

- A fictitious dataset about churn from a Telecom provider (Databel)
- Your task: discover why customers are churning



Defining churn

> The **churn rate**, also known as the rate of attrition or customer churn, is the rate at which customers stop doing business with an entity

- Leaky bucket problem
- Keeping customers is easier than getting new customers
- Reducing churn is a priority for many companies

[1] https://www.investopedia.com/terms/c/churnrate.asp

Calculating churn

Simplified formula

```
Churn rate = customers lost / total number of customers
Churn rate = 10 / 100
10 / 100 = 10%
```

There are **multiple ways** to calculate churn

- Varies by industry and revenue model
- An e-commerce platform could e.g. define a churner as someone who hasn't made a purchase in the last 12 months

The data

Key characteristics

- Databel, a fictitious Telecom provider
- One big table containing 29 columns
- One row per **customer**
- Snapshot of the database at a specific moment in time

The data

Dimensions

Column name	Description
Customer_id	The unique ID that identifies a customer
Churn Label	Contains 'Yes' or 'No' to indicate if a customer churned.
Demographic fields	Age, Gender, State, ...
Premium plans	Unlimited Data, International Plan, ...
...	...

The data

Measures

Column name	Description
Total charges	Sum of all monthly charges
Monthly charge	The average of all monthly charges billed to the customer
Extra data charges	Extra charges for data downloads above the specified customer plan
Extra international charges	Extra charges for international calls for customers not on an international plan
Customer service calls	Number of calls made to customer service
...	...

Metadata sheet

Demo

The first step in any analysis is doing a data check. In this exercise you'll create two measures to check if the count of customer ids is equal to the count of **unique** customer ids.

This check is particularly important, because in case there are duplicate rows we might double-count costs later.

- Open the CSV file `Databel - Data.csv` from the Datasets folder.
- Open the **metadata information** in a separate page.

Create two new measures and name them:

- `Number of Customers`
- `Number of Unique Customers`

You can add your two new measures to an empty page using the visualizations of your choice, and name the page "Data Check". Moving forward it's your job to find correct page names for your analysis. Having appropriate names for the page will be useful when you start building a dashboard in the final part of the case study.

Does the count of unique customers match the count of customers? Indicate also the number of unique costumers.

- Yes, 6687
- No, 6687
- Yes, 2687
- No, 6678

Calculating Churn

It will be extremely useful to have a measure that calculates churn before deep-diving into the analysis. There is a column called `Churn Label` that indicates "Yes" or "No", but this column isn't the easiest to work with.

You'll convert this column to a binomial column indicating if the customer churned or not. You need to use that to calculate the churn rate.

The column `Churn Label` contains "Yes" and "No" values. Use an `IF()` statement to convert the `Churn Label` column into a `Churned` column. It should contain a `1` if a customer churned, and `0` in case a customer didn't churn.

Create a measure for `Number of Churned Customers` using your previously created `Churned` column.

Calculate the `Churn Rate` and format it as a percentage. Creating a new visualization that will display the `Churn Rate` in a new page called "Churn Rate".

What's the total churn rate for "Databel"? (Answer format: XX.XX%)