**Team Members**

* Mohamed AlGhaly
* Abdallah Galal
* Radwa Mohamed
* Mohamed Shoeb

**ITI FINAL PROJECT**

Term Deposit Dataset

|  |
| --- |
|  |
|  |
| |  | | --- | | **Github:**  [**https://github.com/al-ghaly/ITI-Project**](https://github.com/al-ghaly/ITI-Project) | |

**Step 1 Data Wrangling**

**Data Validation & Definition:**

***The aim of this step is to get to know the data and validate each cell.***

**NOTE: We won’t deal with any quality issues here (unless a fatal issue presents itself), because the methodology for dealing with dirty data will depend on what we are aiming to achieve.**

**We already have a data definition file, but let’s dig into the data.**

* **We have 31,647 rows & 18 Columns.**
* **The data have Nulls but does not necessarily mean no missing values.**
* **IDs are unique.**
* **Age ranges from 18 to 95 with a mean value of 40.**
* **We have 11 different jobs and 206 missing jobs.**
* **We have (Married, Single, and Divorced) clients.**
* **We have Primary, Secondary, and College education with 1314 missing values.**
* **The default column is a perfect Yes | No column.**
* **The balance column ranges from -8k to 102k with a mean value of 1360, and after a deep investigation it appears that negative values are normal here.**
* **Housing & Loan are perfect Yes | No Columns.**
* **Contact column has two values (Cellular and telephone) with 9k missing values.**
* **Day & Month columns are perfect 31 | 12 columns.**
* **The duration column ranges from 0 to 4920 seconds with a mean value of 4 minutes.**
* **Campaign column ranges from 1 to 63.**
* **P days column have 25924 values of -1 which mean something, we can’t get.**
* **Previous column ranges from 0 to 275 with 25924 zeros.**
* **P outcome column has 27k missing values.**
* **Subscribed column is a perfect Yes | No column.**
* **After looking at some statistics about the data, all looks good and ready to go.**

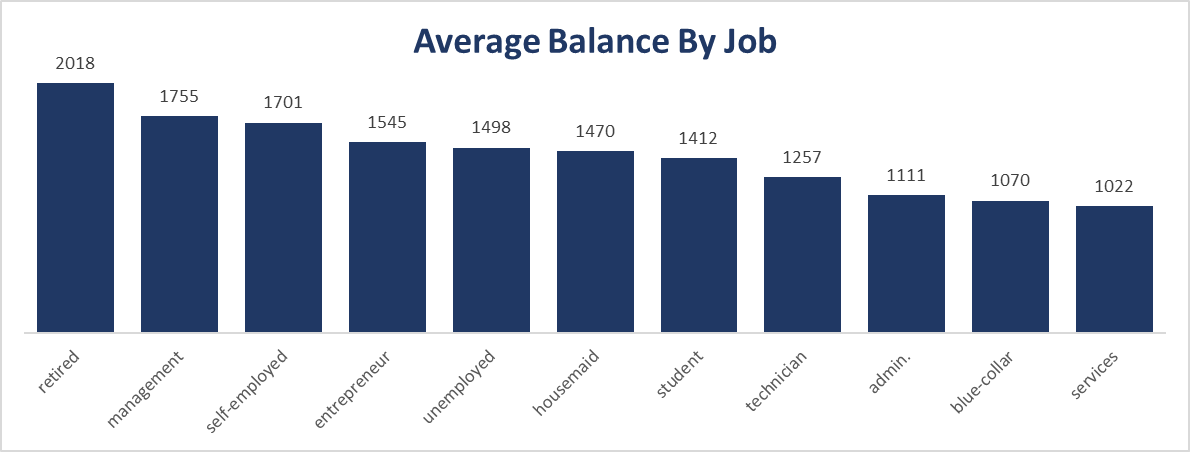
**Data cleaning:**

**This is the typica next step in the data wrangling process, but we will call that off for now, as we are going to analyze the data using various tools and each one needs its own data cleansing techniques.**

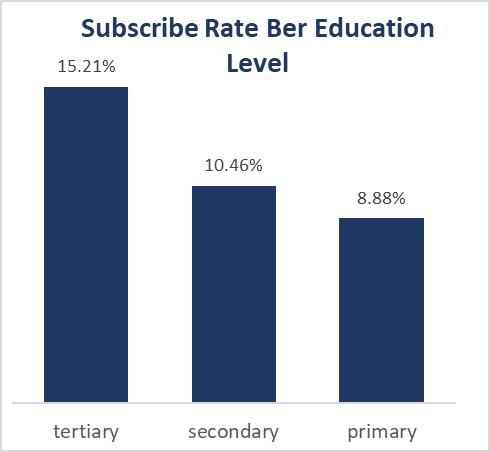
**Step 2 Answering Questions**

***We will answer some insightful business questions using various tools such as Excel, Python, SQL, and Power BI.***

1. **How does the average yearly balance vary based on the client's job type?**

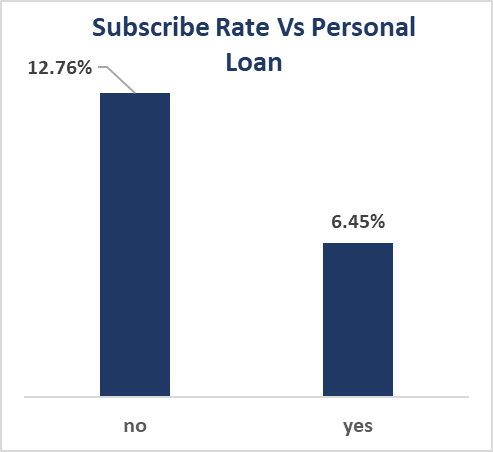
It looks like the average yearly balance **depends mainly** on the job type as:

* Retired clients have the largest balance which makes perfect sense.
* Management-Layer clients are rich.
* Services and Blue-Collar have the lowest average yearly balance.

1. **Is there a relationship between the client's education level and their decision to subscribe?**

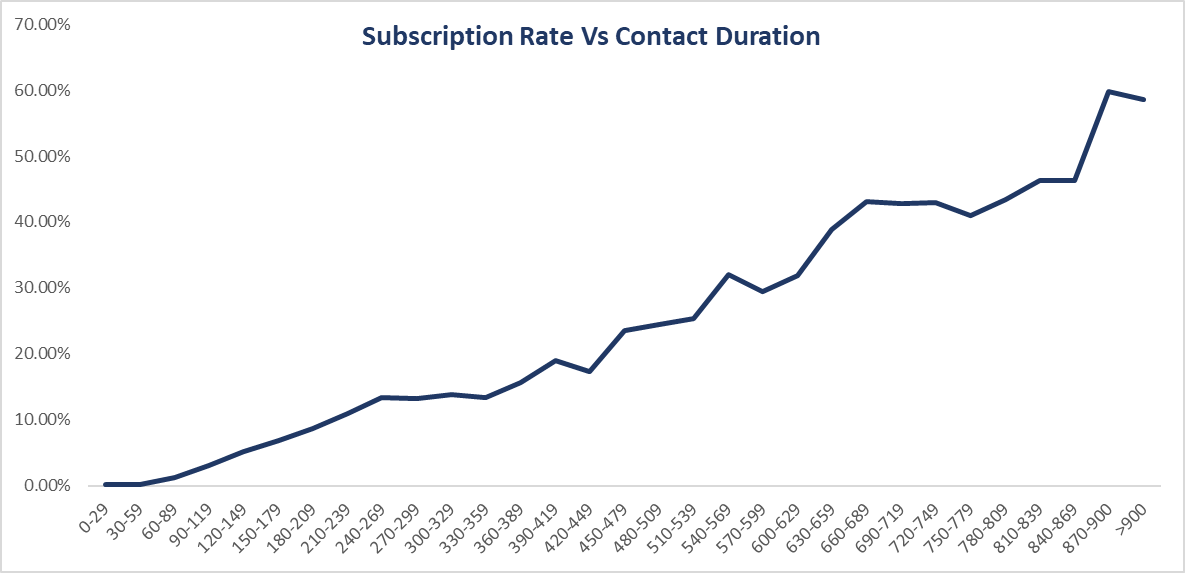
Clearly **higher-educated** clients are **more likely** to subscribe.

This insight can be very helpful as we may need to target highly educated clients in our campaign.

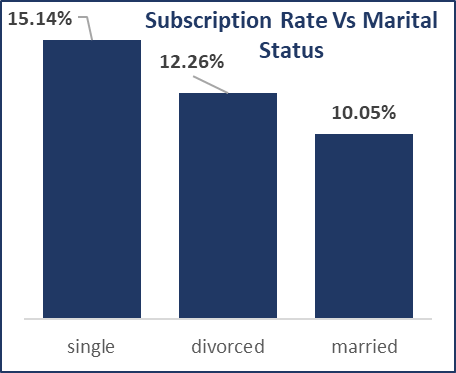
1. **Do clients with a personal loan tend to subscribe more or less frequently compared to those without a loan?**

Clearly **clients with a personal loan** are more likely to subscribe.

This insight can be very helpful as we may need to target clients with a personal loan in our campaign.

1. **Are there any notable differences in the contact duration for subscribed and non-subscribed clients?**

A blind man can see that as the duration of the contact increases the subscription **rate increases**.

1. **Is there a relationship between the client's Marital Status and their decision to subscribe?**

Clearly **single clients** are more likely to subscribe.

This insight can be very helpful as we may need to target single in our campaign.