**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.**

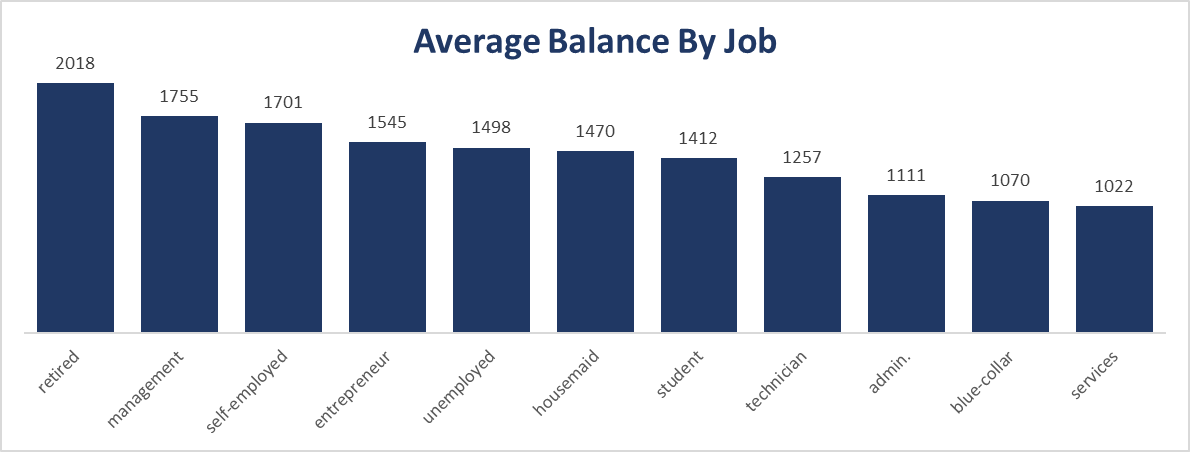
**Data cleaning:**

**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.**

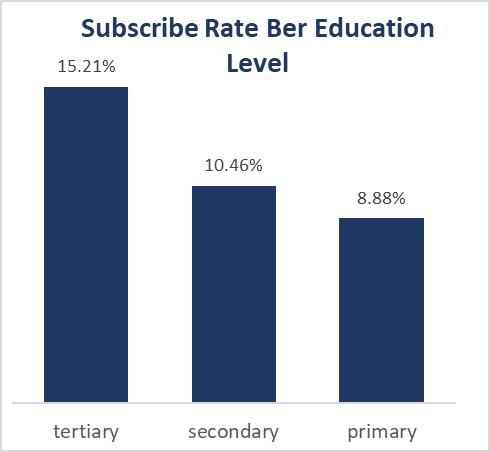
**Step 2 Answering Questions**

***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?**

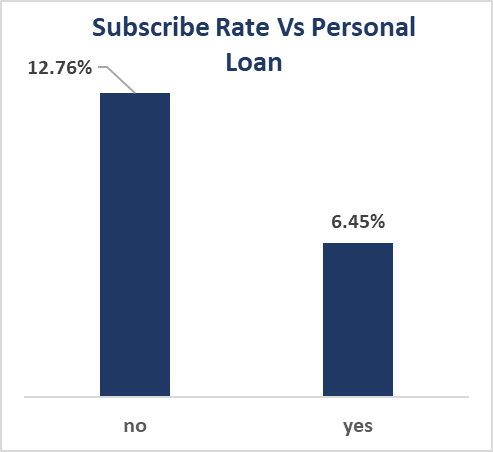
It looks like the average yearly balance **depends mainly** on the job type as:

* 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?**

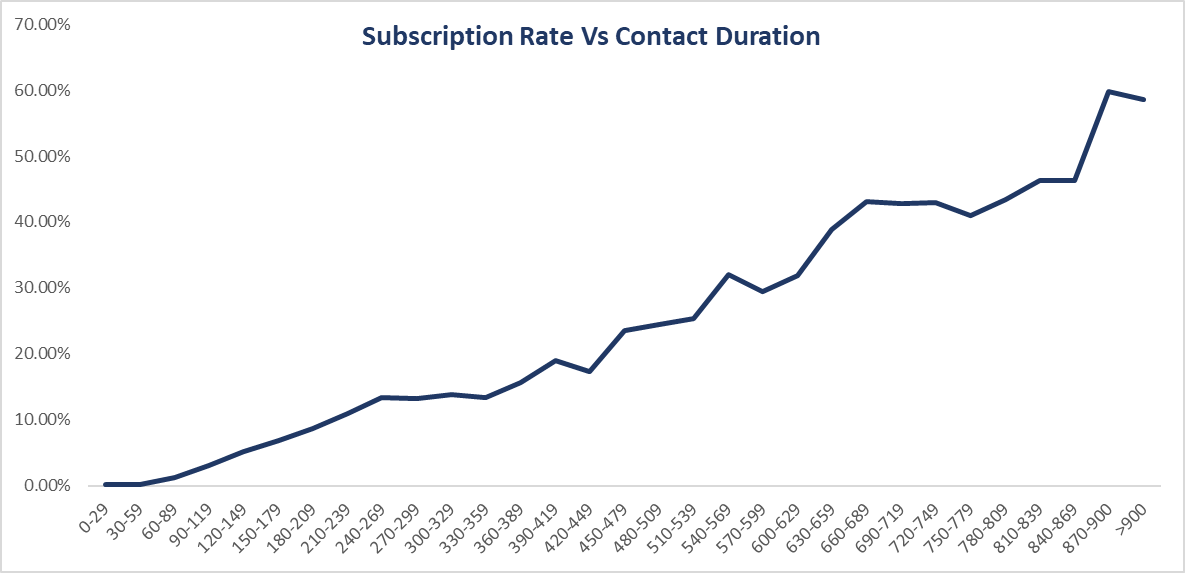
Clearly **higher-educated** clients are **more likely** to subscribe.

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

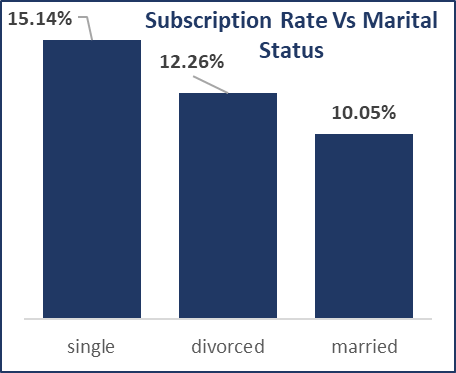
1. **Do clients with a personal loan tend to subscribe more or less frequently compared to those without a loan?**

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

This insight can be very helpful as we may need to target clients without 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.

**A picture containing text, screenshot, font, diagram

Description automatically generatedDashboard to Communicate the Insights**

**Step 3 Data Warehousing**

**Step one will be defining business process:**

1. **Marketing Campaign (Marketing Team)**

* In this process we are going to model the company’s marketing campaigns, including the telephonic marketing campaign.
* This process is mainly concerned with the marketing team of

the bank.

1. **Financial Analysis (Finance Team)**

* This process is mainly concerned with analyzing term deposits and loans.
* It will also involve Analyzing the behavior of clients.
* The main purpose of this process is to manage deposits and loans to maximize our profits.

1. **Customer Services: (Customer Support Team)**

* This process involves three main relative branches:
* Customer Inquiries.
* Customer Complaints.
* Customer Feedback.
* Mainly concerned with customer satisfaction.

**Step two**

**will be defining which questions we want our model to tackle:**

Questions (Will be answered along with a specific time period)

1. Which customers respond to our campaigns the most (Age/Gender/Education/and so on).
2. What factors help us increase the conversion/subscription rate.
3. What deposit types attract the most clients.
4. How often each client rolls over/cancels their deposit, and what factors affect its decision.
5. What are the most profitable deposit types.
6. What are the trends of the revenue/expenses of the deposits.
7. What are the most common deposit types.
8. What is the effect of changing the interest rate on the profit.
9. What is the cancellation rate of our term deposits.
10. Which customers are the most profitable (Age/Gender/Education/and so on).
11. What is the impact of investment amount, the duration of their investment, the interest rate they are offered on the revenue.
12. What are the factors that affect the decision of the clients on whether to roll over or not.
13. What is the channel of communication that is most lovable/hated.
14. How many complaints are we getting.
15. How to improve customer satisfaction.
16. How to improve our business.
17. What loans are most rewarding in terms of profit.
18. What are the possible ways to reduce expenses.
19. What are the best ways to maximize revenue.
20. Which loan category is most rewarding.
21. What is the annual net profit for the bank.
22. What are the factors that make us decline a specific loan.
23. Fdf
24. fdf
25. Way, way more, but just keep up with this.

**STEP Three**

**Step three will be defining granularity for each business process:**

**Marketing Campaign**

* The most detailed grain is each marketing action (Phone call).
* I preferred just for the simplicity of the modeling to separate all marketing related processes into a single business process, in which will analyze the marketing performance, by analyzing each individual action taken in response to a marketing campaign.

**Financial Analysis**

* The most detailed grain is each interaction or transaction made by or to any client, on a specific date.
* Transaction or interaction here refers to:
  + Opening a new deposit account.
  + Renewal of a deposit.
  + Cancellation of a deposit.
  + Loan request.
  + Approval or refusal of a loan.

**Customer Services**

* The most detailed grain is the combination of an individual customer care action (Inquiry, Feedback, Complaint), on a specific deposit or loan for a specific customer, on a specific branch, on a specific employee, at a given date.

**STEP Four**

**Step 4 will be the capstone for the project which involves determining both facts and dimensions.**

**Marketing Campaign**

* In this fact table we want to analyze the behavior of our clients who invest on term deposits or loans, and assess the marketing team performance, to maximize performance of our marketing campaigns.

**Measurements/Attributes**

* Subscribed

**Dimensions**

* Customer
* Transactions-Bridge
* Date

**Financial Analysis**

* In this fact table we want to analyze any interaction or transaction made by or to any client.
  + The term deposits we offer to our clients.
  + The loans we offer to our clients.
  + The process of assessing clients asking for loans, we want to be 101 % sure any loan given to any client would be PAID BACK.
  + The revenue analysis of the bank.
  + The profit analysis.
  + Monitoring our expenses.

**Measurements/Attributes**

* Interest Rate
* Renewed
* Cancelled
* Amount
* Expenses
* # TODO to be removed after usage Transaction Type (Loan or Deposit)
* Approved

**Dimensions**

* Customer
* Transactions-Bridge
* Date
* Branch

**Customer Services**

* This fact table helps us keep track of customer satisfaction and provides ways to improve the company’s performance by responding to clients’ needs.

**Measurements/Attributes**

* Channel (channel of interaction)
* Action Type (Inquiry, Feedback, or Complaint)
* Severity

**Dimensions**

* Customer
* Transactions-Bridge
* Branch
* Employee
* Date

**DIMENSIONS:**

1- Customer

• A dimension holding data about all the potential and current clients of our bank.

2- Transactions-Bridge

• Holding data about all term deposits and loans made by the bank.

3- Date

• Typical calendar dimension for any DWH.

4- Branch

• A dimension holding data about all the branches of the bank.

5- Employee

• A dimension holding data about all the employees in the bank.

**STEP Five**

A picture containing text, diagram, screenshot, font

Description automatically generated**Step 5 will be the schema modeling, and we will use a star schema.**

**STEP Six**

**Step eight is a discussion about step 5.**

**Why did we choose star schema modeling?**

**A star schema is perfect for our design for some reasons:**

1. Simplified query performance: Star schema modeling allows us for simplified and optimized query performance. Since the fact table is at the center of the schema and connected to the dimension tables through foreign keys, queries can be executed efficiently and quickly, without the need for complex joins or subqueries.
2. Improved data analysis: Star schema modeling provides a simplified and intuitive way to analyze data from multiple dimensions. Analysts can easily drill down into data by navigating through the dimension tables, allowing for more complex analysis and insights.
3. Easier maintenance: Star schema modeling is easier to maintain than other modeling approaches. Since each dimension table is connected directly to the fact table, changes to one dimension table will not affect other tables in the schema. This makes it easier to modify and update the data warehouse over time.
4. Scalability: Star schema modeling is highly scalable and can handle large amounts of data. By separating the data into smaller, more manageable tables, the schema can accommodate large amounts of data without impacting query performance or data analysis.

Overall, star schema modeling is a popular and effective approach for designing data warehouses. It provides a simple, intuitive, and scalable way to organize and analyze data, making it easier for analysts and business users to get the insights they need from their data.

**It mainly consists of 3 components:**

* + Dimensions: integral companions to a fact table. containing the textual context associated with a business process measurement event.
  + Facts: stores the performance measurements resulting from an organizations’ business process events.
  + Measurements: The actual measurements stored in the fact tables.

For a simple data warehouse just like the one we are modeling a star schema is the perfect fit as it covers all our business needs, with high performance, data integrity, and powerful possibilities.