Column	Changes	Reason
Row number	removed	Not needed
Customer ID	no changes	
Last name	Removed column	Not needed
Credit Score	3 missing values replaced with N/A	Fill in blanks
Country	ES, DE, DE changed to respective Spain, Germany, France	uniform
Gender	Changed F, M to respective Female, Male	
Age	Values of 2 and NULL changed to N/A	Uniform, removed age mistakes
Tenure	No changes	
Balkance	No changes	
NumofProducts	No changes	
HasCRCard	No Changes	
Is Active member	No changes	
Estimated Salary	Found 1 missing value, replaced with N/A	uniform
Exited from bank	No changes	

1. Download Pig E. Bank's client data set (.xlsx). Open the data set in Excel and take a moment to familiarize yourself with the data.
2. To understand the data, you'll first need to assess the quality of the data, by checking for missing values, errors, and inconsistencies.
    - You'll also need to clean your data, using the techniques that you learned in previous Achievements. Fix any inconsistencies in the table and/or any errors, as far as it is possible.
    - Document your processes for assessing the data quality and cleaning the data, and note down any missing values or errors.

3. Now that you have cleaned the data, you're ready to calculate some basic descriptive statistics to understand the data. Remember, your goal is to identify the risk factors that have contributed to customers leaving the bank.

- Separate the clients into 2 groups: one for those who have left the bank and a second for those who have stayed (hint: "1" in the "ExitedFromBank" column represents customers who have left).

Exited from bank

Row Labels	Count of Gender	Average of Credit Score	Average of Age
Male	83	643	46
Female	121	632	45
Grand Total	**204**	**637**	**45**

Row Labels	Sum of ExitedFromBank?	Sum of IsActiveMember	Average of Age
France	77	21	46
Germany	75	25	45
Spain	52	15	45
Grand Total	**204**	**61**	**45**

Stayed

Row Labels	Count of Gender	Average of Credit Score	Average of Age
Female	341	652	38
Male	445	652	37
N/A	1	497	24
Grand Total	**787**	**652**	**37**

Row Labels	Sum of IsActiveMember	Average of Age
France	219	38
Germany	103	37
Spain	120	38
Grand Total	**442**	**37**

- Use pivot tables and other Excel functions to identify the top 3 to 4 factors that lead to clients leaving.
- Gather and analyze statistical information on both groups (e.g., find averages, means).
- Determine the leading factors that contribute to client loss, based on your analysis of the data provided.
- Document your results and how you reached them.

➢

Customer_ID	Credit Score	Age	Tenure	Balance	NumOfProducts	Estimated Salary
MAX	850	82	10	197042	3	$199,661.50
MIN	411	18	0	0	1	$371.05
MEAN	652	37	5	74831	2	$98,943.39

Exited from Bank

Customer_ID	Credit Score	Age	Tenure	Balance	NumOfProducts	Estimated Salary
MAX	850	69	10	213146	4	$199,725.39
MIN	376	22	0	0	1	$417.41
MEAN	636	45	5	90397	1	$97,183.38

- Based on the statistical information on averages, individuals with lower Credit score, higher age, and lower number of products left Pig E Bank.
- Based on the analysis of the data the determining factors for client loss are:
  1. Inactivity (inactive clients are more likely to leave Pig E Bank)
  2. Gender (Female client loss is disproportionate to males)
  3. Nationality, (75 out of 178 German clients left the bank, highest percentage).

4. Using the information you've uncovered so far, create a decision tree to determine the probability of customers leaving the bank.
    a. Pick which tool you'll use to create your decision tree. You can either create your own template using Excel or Powerpoint, for example, or download a decision-tree template.
    b. Determine which decision node will have the greatest impact and place it at top of the tree. For example, if you decide that an estimated salary below 15,000 USD is the biggest risk factor, then you would put this at the top and build your tree from there. Make sure that your decision tree includes the top 3 to 4 risk factors you identified in step 3.
5. Combine your decision tree and answers document into one PDF and upload it here for your tutor to review.