Keanu Gomes
Answers 5.4
January 3, 2024

Data Ethics & Applied Analytics
5.4: Intro to Data Mining

Assess the quality of the data, by checking for missing values, errors, and inconsistencies.

Columns	Missing values	Missing values treatment
last_name	1 NULL value, customer_id: 15752047	*PII Security Risk, entire column to be removed.*
credit_score	3 blank values, customer_id: 15627801, 15785542, and 15570060	*left-as-is.*
gender	1 NULL value, customer_id: 15737173	*left-as-is.*
age	1 NULL value, customer_id: 15699309	*left-as-is.*
est_salary	2 blank values, customer_id: 15597945, and 15785542	*left-as-is.*

note: no duplicates found the Pig E. Bank client data set.

Columns	Inconsistencies	Inconsistent values treatment
country	DE=Germany ES=Spain FR=France	*Find and replace country ID values with their full country names.* DE-->Germany = 23 replacements ES-->Spain = 118 replacements FR-->France = 480 replacements
gender	F=Female M=Male	*Find and Replace gender values with full gender identity.* F-->Female = 19 replacements M-->Male = 49 replacements
age	age=2 found in data set 11 times. customer_id: 15745307, 15692463, 15718673, 15633648,15603323, 15745621,15607040, 15810978,15668886, 15603436, and 15807609.	*replaced with avg age of 20 year olds in data set.*

Keanu Gomes
Answers 5.4
January 3, 2024

Clean and prepare the data.

Columns dropped	Columns renamed	Columns' type changed	Comment/Reason
Row_Number			Column is irrlevant to analysis.
Last_Name			PII Security Risk, column removed from data set analysis.
	Tenure		Tenure = the duration of the customer's relationship with the bank.
	{Customer_ID:customer_id}, {Credit Score: credit_score}, {Country:country}, {Gender:gender}, {Age:age}, {Balance:balance}, {NumOfProducts:num_of_products}, {HasCrCard?:has_credit_card}, {IsActiveMember:is_active_member}, {Estimated Salary:est_salary}, {ExitedFromBank:exited_from_bank}		Lowercasing and dashing implemented for smoother analysis.
		balance	inconsistent float numbers, change data type to float decimals at .00. Reformatted to account number format since these are balances of bank accounts.
		has_credit_card	changing binary numeric to string obj, 1=yes, 2=no (991 total replacements)
		is_active_member	changing binary numeric to string obj, 1=yes, 2=no (991 total replacements)
		est_salary	Reformatted to account number format since these are balances of bank accounts.
		exited_from_bank	changing binary numeric to string obj, 1=yes, 2=no (991 total replacements)

Dataset	New column	Column/s it was derived from	Conditions
Pig E. Bank client data_clean			
note: no column derivations were made for this analysis.			

Keanu Gomes
Answers 5.4
January 3, 2024

Exited from bank: Group 1

Values	
Min of credit_score	376
Max of credit_score	850
Average of credit_score	637
Min of age	20
Max of age	69
Average of age	45
Min of account_age	0
Max of account_age	10
Average of account_age	5
Min of balance	0
Max of balance	213146.2
Average of balance	$ 90,239.22
Min of num_of_products	1
Max of num_of_products	4
Average of num_of_products	1
Min of est_salary	$ 417.41
Max of est_salary	$199,725.39
Average of est_salary	$ 97,155.20
Count of exited_from_bank	204

Stayed at bank: Group 2

Values	
Min of credit_score	411
Max of credit_score	850
Average of credit_score	652
Min of age	18
Max of age	82
Average of age	37
Min of account_age	0
Max of account_age	10
Average of account_age	5
Min of balance	0
Max of balance	197041.8
Average of balance	$ 74,830.87
Min of num_of_products	1
Max of num_of_products	3
Average of num_of_products	2
Min of est_salary	$ 371.05
Max of est_salary	$199,661.50
Average of est_salary	$ 98,943.39
Count of exited_from_bank	787

Group 1 and 2 descriptive statistics

Count of exited_from_bank			
Country	**no**	**yes**	**Grand Total**
France	83.96%	16.04%	100.00%
Germany	70.82%	29.18%	100.00%
Spain	79.53%	20.47%	100.00%
Grand Total	**79.41%**	**20.59%**	**100.00%**
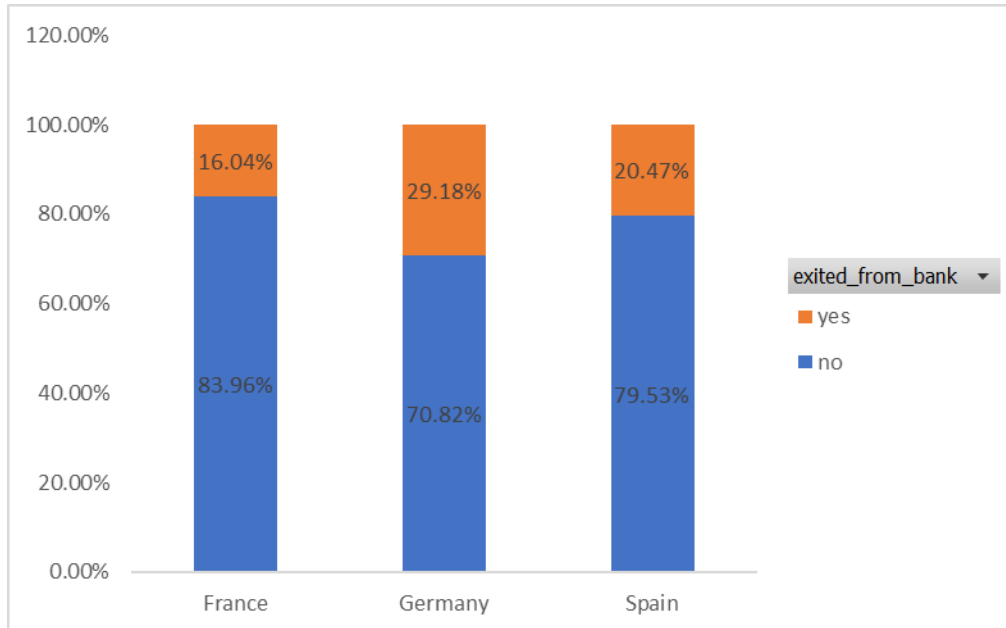
Which country had the highest exit rate? -->

Germany had the highest exit rate from the bank at *29%*, while *France* had the lowest exit rate at *16%.*

Client loss decision tree

Which clients are most likely to exit from Pig E. Bank?
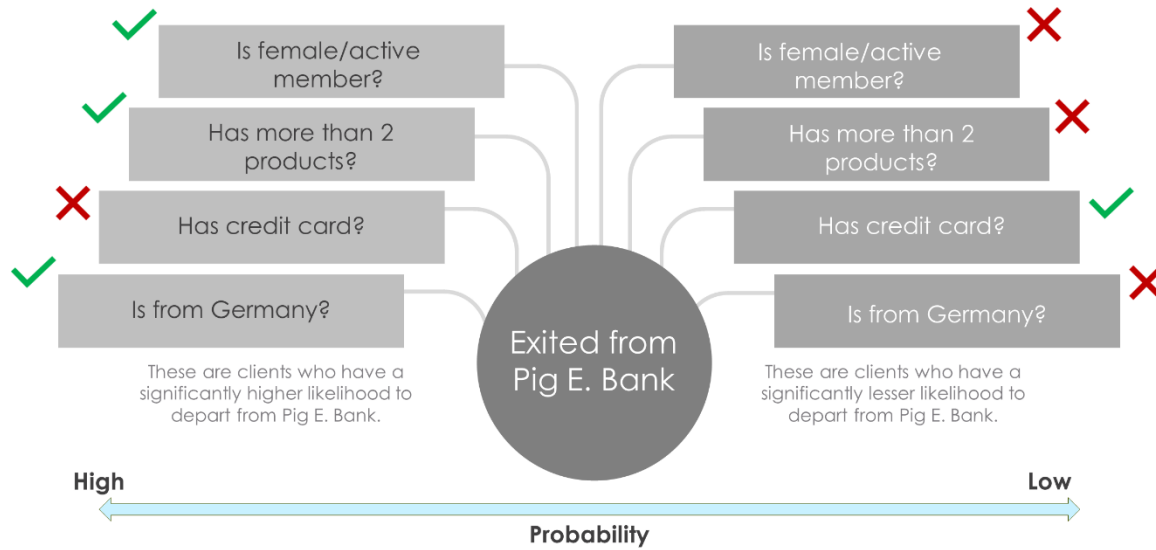
Which country had the highest exit rate? ---> **note:** Germany had the highest exit rate from the bank at **29%**, while France had the lowest exit rate at **16%**.
Active members by country ---> **note:** Active German members have the highest exit rate from the bank at **20%**, while France has the lowest exit rate with **9%**.
Non-active members by country ---> **note:** Among non-active members, Germany exhibits the highest exit rate from the bank at **39%**, whereas France has the lowest exit rate at **23%**.

Female-active members by country --> **note:** Among female active members, Germany had the highest exit rate from the bank at **19%**, while France has the lowest exit rate at **14%**.
Male-active members by country --> **note:** Among male active members, Germany had the highest exit rate from the bank at **20%**, while Spain has the lowest exit rate at **4%**.

Female-non-active members by country ---> **note:** Among female non-active members, Germany had the highest exit rate from the bank at **42%**, while France has the lowest exit rate at **32%**.
Male-non-active members by country ---> **note:** Among male non-active members, Germany had the highest exit rate from the bank at **35%**, while France has the lowest exit rate at **15%**.

Female-active members by country, and with credit card ------------------------> **note:** Among female active members with a credit card, Germany had the highest exit rate from the bank at **19%**, while France has the lowest exit rate at **12%**.
Female-active members by country, and without credit card --------------------> **note:** Among female active members without a credit card, Spain had the highest exit rate from the bank at **26%**, while France has the lowest exit rate at **19%**.

Male-active members by country, and with credit card --------------------------> **note:** Among male active members with a credit card, Germany had the highest exit rate from the bank at **19%**, while Spain has the lowest exit rate at **2%**.
Male-active members by country, and without credit card ----------------------> **note:** Among male active members without a credit card, Germany had the highest exit rate from the bank at **22%**, while France has the lowest exit rate at **3%**.

Female-non-active members by country, with credit card ------------------------> **note:** Among female non-active members with a credit card, Germany had the highest exit rate from the bank at **39%**, while Spain has the lowest exit rate at **28%**.
Female-non-active members by country, without credit card --------------------> **note:** Among female non-active members without a credit card, Spain had the highest exit rate from the bank at **53%**, while France has the lowest exit rate at **28%**.

Male-non-active members by country, without credit card ----------------------> **note:** Among male non-active members with a credit card, Germany had the highest exit rate from the bank at **37%**, while Spain has the lowest exit rate at **19%**.
Male-non-active members by country, without credit card ----------------------> **note:** Among male non-active members without a credit card, Germany had the highest exit rate from the bank at **31%**, while France has the lowest exit rate at **6%**.