

BÁO CÁO HÀNG TUẦN

Họ và tên: Trần Minh Hiếu

MSV : 21011601

Lớp học : K15-CNTT3

Ngày : 10/09/2023

MỤC ĐÍCH

Báo cáo hàng tuần này nhằm mục đích tìm hiểu về các loại dữ liệu liên quan đến tài chính, ngân hàng từ Kaggle và trình bày các phân tích cơ bản về chúng.

DANH SÁCH DỮ LIỆU

1. Dữ liệu 1

- Tên: Bank Customer Churn Dataset.
- Nguồn gốc: Kaggle - truy cập tại: <https://www.kaggle.com/datasets/gauravtopre/bank-customer-churn-dataset> (<https://www.kaggle.com/datasets/gauravtopre/bank-customer-churn-dataset>)
- Kích thước: 10000 dòng, 12 cột.
- Mục đích: Nhằm đến dự đoán churn cho ngân hàng ABC (ABC Bank).

2. Dữ liệu 2

- Tên: Africa Economic, Banking and Systemic Crisis Data.
- Nguồn gốc: Kaggle - truy cập tại: <https://www.kaggle.com/datasets/chirin/africa-economic-banking-and-systemic-crisis-data> (<https://www.kaggle.com/datasets/chirin/africa-economic-banking-and-systemic-crisis-data>)
- Kích thước: 1059 dòng, 14 cột.
- Mục đích: Nhằm nghiên cứu yếu tố nào liên quan nhiều nhất đến sự khủng hoảng trong hệ thống tài chính. Với mức lạm phát nào trở thành một cuộc khủng hoảng lạm phát thực tế?

PHÂN TÍCH DỮ LIỆU

Bank Customer Churn Dataset

1. Định nghĩa

Tập dữ liệu này là một tập phân tích và nghiên cứu tình trạng churn của khách hàng. Bank customer churn là sự mất mát hoặc chuyển đổi của khách hàng từ một ngân hàng sang ngân hàng khác hoặc hoàn toàn ngừng sử dụng các dịch vụ tài chính của một ngân hàng. Churn trong lĩnh vực ngân hàng có thể có nhiều hình thức, bao gồm:

- Chuyển ngân hàng: Khách hàng chuyển toàn bộ hoặc một phần tiền tiết kiệm, tài khoản thanh toán hoặc khoản vay sang ngân hàng khác.
- Hủy tài khoản: Khách hàng đóng tài khoản ngân hàng hoặc thẻ tín dụng của họ.

- Ngừng sử dụng dịch vụ: Khách hàng ngừng sử dụng các dịch vụ ngân hàng như internet banking, thẻ tín dụng, vay tiền, v.v.

2. Cấu trúc

- Trường 1: ID Khách hàng (customer_id) - mã số duy nhất cho mỗi khách hàng.
- Trường 2: Điểm tín dụng (credit_score) - con số đại diện cho sự đáng tin cậy tài chính của khách hàng.
- Trường 3: Nước (country) - quốc tịch của khách hàng.
- Trường 4: Giới tính (gender) - giới tính của khách hàng.
- Trường 5: Độ tuổi (age) - tuổi của khách hàng.
- Trường 6: Thời gian sử dụng dịch vụ (tenure) - thời gian mà một khách hàng đã sử dụng dịch vụ của ngân hàng.
- Trường 7: Số dư tài khoản (balance) - số dư tài khoản hiện tại của khách hàng trong ngân hàng.
- Trường 8: Số lượng sản phẩm (products_number) - số lượng sản phẩm hoặc dịch vụ mà một khách hàng đã mua hoặc sử dụng.
- Trường 9: Thẻ tín dụng (credit_card) - thẻ tín dụng của khách hàng.
- Trường 10: Trạng thái khách hàng (active_member) - trạng thái hoạt động, sử dụng dịch vụ của khách hàng.
- Trường 11: Lương dự đoán (estimated_salary) - dự đoán lương của khách hàng.
- Trường 12: Sự mất mát hoặc chuyển đổi (churn) - tình trạng khách hàng đã ngừng sử dụng dịch vụ hoặc chuyển đổi sang dịch vụ của ngân hàng khác.

3. Bảng thống kê

- Định dạng của tập dữ liệu:

#	Column	Non-Null	Count	Dtype
0	customer_id	10000	non-null	int64
1	credit_score	10000	non-null	int64
2	country	10000	non-null	object
3	gender	10000	non-null	object
4	age	10000	non-null	int64
5	tenure	10000	non-null	int64
6	balance	10000	non-null	float64
7	products_number	10000	non-null	int64
8	credit_card	10000	non-null	int64
9	active_member	10000	non-null	int64
10	estimated_salary	10000	non-null	float64
11	churn	10000	non-null	int64

dtypes: float64(2), int64(8), object(2)

- 5 dòng đầu của tập dữ liệu:

	customer_id	credit_score	country	gender	age	tenure	balance	products_number	credit_card	active_member	estimated_salary	churn
0	15634602	619	France	Female	42	2	0.00	1	1	1	101348.88	1
1	15647311	608	Spain	Female	41	1	83807.86	1	0	1	112542.58	0
2	15619304	502	France	Female	42	8	159660.80	3	1	0	113931.57	1
3	15701354	699	France	Female	39	1	0.00	2	0	0	93826.63	0
4	15737888	850	Spain	Female	43	2	125510.82	1	1	1	79084.10	0

- Thống kê tập dữ liệu:

	customer_id	credit_score	age	tenure	balance	products_number	credit_card	active_member	estimated_salary	churn
count	1.000000e+04	10000.000000	10000.000000	10000.000000	10000.000000	10000.000000	10000.000000	10000.000000	10000.000000	10000.000000
mean	1.569094e+07	650.528800	38.921800	5.012800	76485.889288	1.530200	0.70550	0.515100	100090.239881	0.203700
std	7.193619e+04	96.653299	10.487806	2.892174	62397.405202	0.581654	0.45584	0.499797	57510.492818	0.402769
min	1.556570e+07	350.000000	18.000000	0.000000	0.000000	1.000000	0.00000	0.000000	11.580000	0.000000
25%	1.562853e+07	584.000000	32.000000	3.000000	0.000000	1.000000	0.00000	0.000000	51002.110000	0.000000
50%	1.569074e+07	652.000000	37.000000	5.000000	97198.540000	1.000000	1.00000	1.000000	100193.915000	0.000000
75%	1.575323e+07	718.000000	44.000000	7.000000	127644.240000	2.000000	1.00000	1.000000	149388.247500	0.000000
max	1.581569e+07	850.000000	92.000000	10.000000	250898.090000	4.000000	1.00000	1.000000	199992.480000	1.000000

- Ma trận tương quan của tập dữ liệu:

	customer_id	credit_score	age	tenure	balance	products_number	credit_card	active_member	estimated_salary	churn
customer_id	1.000000	0.005308	0.009497	-0.014883	-0.012419	0.016972	-0.014025	0.001665	0.015271	-0.006248
credit_score	0.005308	1.000000	-0.003965	0.000842	0.006268	0.012238	-0.005458	0.025651	-0.001384	-0.027094
age	0.009497	-0.003965	1.000000	-0.009997	0.028308	-0.030680	-0.011721	0.085472	-0.007201	0.285323
tenure	-0.014883	0.000842	-0.009997	1.000000	-0.012254	0.013444	0.022583	-0.028362	0.007784	-0.014001
balance	-0.012419	0.006268	0.028308	-0.012254	1.000000	-0.304180	-0.014858	-0.010084	0.012797	0.118533
products_number	0.016972	0.012238	-0.030680	0.013444	-0.304180	1.000000	0.003183	0.009612	0.014204	-0.047820
credit_card	-0.014025	-0.005458	-0.011721	0.022583	-0.014858	0.003183	1.000000	-0.011866	-0.009933	-0.007138
active_member	0.001665	0.025651	0.085472	-0.028362	-0.010084	0.009612	-0.011866	1.000000	-0.011421	-0.156128
estimated_salary	0.015271	-0.001384	-0.007201	0.007784	0.012797	0.014204	-0.009933	-0.011421	1.000000	0.012097
churn	-0.006248	-0.027094	0.285323	-0.014001	0.118533	-0.047820	-0.007138	-0.156128	0.012097	1.000000

Africa Economic, Banking and Systemic Crisis Data

1. Định nghĩa

Tập dữ liệu này là một nguồn thông tin hữu ích để nghiên cứu về tình hình kinh tế, ngân hàng và các cuộc khủng hoảng hệ thống trong lục địa Châu Phi. Điều này có thể giúp ta hiểu rõ hơn về tình hình kinh tế và tài chính trong khu vực này và thực hiện các phân tích hoặc nghiên cứu về các vấn đề liên quan.

2. Cấu trúc

- Trường 1: Trường hợp (case) - mỗi case tương ứng với một con số đại diện cho một quốc gia.
- Trường 2: Mã quốc gia (cc3) - mã quốc gia gồm 3 ký tự.
- Trường 3: Quốc gia (country) - Tên quốc gia ở Châu Phi.
- Trường 4: Năm (Year): Năm quan sát các diễn biến, tình trạng ở các nước.
- Trường 5: Khủng hoảng hệ thống tài chính (systemic_crisis) - tình trạng khủng hoảng trong năm.
- Trường 6: Tỷ giá hối đoái (exch_usd) - tỷ giá hối đoái của một quốc gia với USD.
- Trường 7: Vỡ nợ nội địa (domestic_debt_in_default) - tình trạng nợ nội địa vỡ nợ diễn ra trong năm.
- Trường 8: Vỡ nợ ngoại vụ (sovereign_external_debt_default) - tình trạng vỡ nợ ngoại vụ của quốc gia trong năm.
- Trường 9: Tổng vỡ nợ so với GDP (gdp_weighted_default) - tổng vỡ nợ của quốc gia so với GDP trong năm.
- Trường 10: Tỷ lệ lạm phát - CPI (inflation_annual_cpi) - tỷ lệ lạm phát dựa trên chỉ số giá tiêu dùng trong năm.
- Trường 11: Độc lập (independence) - quốc gia độc lập, tự chủ.
- Trường 12: Khủng hoảng tiền tệ (currency_crises) - tình trạng khủng hoảng tiền tệ diễn ra trong năm.
- Trường 13: Khủng hoảng lạm phát (inflation_crises) - tình trạng khủng hoảng lạm phát diễn ra trong năm.

- Trường 14: Khủng hoảng ngân hàng (banking_crisis) - tình trạng khủng hoảng ngân hàng diễn ra trong năm.

3. Bảng thống kê

- Định dạng của tập dữ liệu:

#	Column	Non-Null	Count	Dtype
0	case	1059	non-null	int64
1	cc3	1059	non-null	object
2	country	1059	non-null	object
3	year	1059	non-null	int64
4	systemic_crisis	1059	non-null	int64
5	exch_usd	1059	non-null	float64
6	domestic_debt_in_default	1059	non-null	int64
7	sovereign_external_debt_default	1059	non-null	int64
8	gdp_weighted_default	1059	non-null	float64
9	inflation_annual_cpi	1059	non-null	float64
10	independence	1059	non-null	int64
11	currency_crises	1059	non-null	int64
12	inflation_crises	1059	non-null	int64
13	banking_crisis	1059	non-null	object
dtypes: float64(3), int64(8), object(3)				

- 5 dòng đầu của tập dữ liệu:

	case	cc3	country	year	systemic_crisis	exch_usd	domestic_debt_in_default	sovereign_external_debt_default	gdp_weighted_default	inflation_annual_cpi
0	1	DZA	Algeria	1870	1	0.052264	0	0	0.0	3.441456
1	1	DZA	Algeria	1871	0	0.052798	0	0	0.0	14.149140
2	1	DZA	Algeria	1872	0	0.052274	0	0	0.0	-3.718593
3	1	DZA	Algeria	1873	0	0.051680	0	0	0.0	11.203897
4	1	DZA	Algeria	1874	0	0.051308	0	0	0.0	-3.848561

- Thống kê tập dữ liệu:

	case	year	systemic_crisis	exch_usd	domestic_debt_in_default	sovereign_external_debt_default	gdp_weighted_default	inflation_annual_cpi
count	1059.000000	1059.000000	1059.000000	1059.000000	1059.000000	1059.000000	1059.000000	1.059000e
mean	35.613787	1967.767705	0.077432	43.140831	0.039660	0.152975	0.006402	2.084889e
std	23.692402	33.530632	0.267401	111.475380	0.195251	0.380133	0.043572	6.757274e
min	1.000000	1860.000000	0.000000	0.000000	0.000000	0.000000	0.000000	-2.850214e
25%	15.000000	1951.000000	0.000000	0.195350	0.000000	0.000000	0.000000	2.086162e
50%	38.000000	1973.000000	0.000000	0.868400	0.000000	0.000000	0.000000	5.762330e
75%	56.000000	1994.000000	0.000000	8.462750	0.000000	0.000000	0.000000	1.164405e
max	70.000000	2014.000000	1.000000	744.306139	1.000000	1.000000	0.400000	2.198970e

- Ma trận tương quan của tập dữ liệu:

	case	year	systemic_crisis	exch_usd	domestic_debt_in_default	sovereign_external_debt_default	gdp_weighted_default
case	1.000000	0.115574	0.010991	-0.231976	0.128358	-0.039262	-0.032981
year	0.115574	1.000000	0.197450	0.248757	0.136828	0.271890	-0.054670
systemic_crisis	0.010991	0.197450	1.000000	0.202687	0.122158	0.249850	0.005274
exch_usd	-0.231976	0.248757	0.202687	1.000000	0.005253	0.422890	-0.040726
domestic_debt_in_default	0.128358	0.136828	0.122158	0.005253	1.000000	0.464751	-0.029874
sovereign_external_debt_default	-0.039262	0.271890	0.249850	0.422890	0.464751	1.000000	0.345915
gdp_weighted_default	-0.032981	-0.054670	0.005274	-0.040726	-0.029874	0.345919	1.000000
inflation_annual_cpi	0.044762	0.037035	0.106452	-0.011947	0.151832	0.072609	-0.004535
independence	0.021858	0.407360	0.147083	0.126034	0.109120	0.228192	0.078936
currency_crises	0.095339	0.189390	0.112751	-0.056472	0.227585	0.199428	0.016970
inflation_crises	0.006405	0.098630	0.172562	-0.063783	0.224429	0.187930	0.017630

NHẬN XÉT VÀ KẾT LUẬN

Bank Customer Churn Dataset

1. Nhận xét

- Tập dữ liệu sạch, không có điểm dữ liệu nào null.
- 3 trường age, balance, active_member là có mối tương quan nhất với churn.
- Độ tuổi của khách hàng tập trung vào khoảng từ 36-40 tuổi.
- Số dư trong tài khoản tập trung vào khoảng 80000.
- 45.4% là khách hàng nữ, 54.6% là khách hàng nam
- 48.5% trạng thái hoạt động của khách hàng tạm thời không sử dụng dịch vụ
- 20.4 % Khách hàng ngừng hoặc chuyển đổi sử dụng dịch vụ

2. Kết luận

- Tình trạng churn của khách hàng là một vấn đề quan trọng trong ngành ngân hàng và dịch vụ tài chính. Sự mất mát hoặc chuyển đổi khách hàng có thể gây ra sự suy giảm doanh thu và lợi nhuận, và do đó, quản lý mối quan hệ khách hàng và giảm thiểu churn là một phần quan trọng của chiến lược kinh doanh của các tổ chức tài chính.
- Để hiểu và quản lý tốt tình trạng churn, ngân hàng cần thu thập và phân tích dữ liệu từ các nguồn khác nhau như hồ sơ khách hàng, giao dịch tài chính, và thông tin về sản phẩm. Phân tích dữ liệu giúp xác định các yếu tố ảnh hưởng đến tình trạng churn và tạo ra các chiến lược để giảm thiểu nó.

Africa Economic, Banking and Systemic Crisis Data

1. Nhận xét

- Tập dữ liệu sạch, không có điểm dữ liệu nào null.
- 3 trường systemic_crisis, domestic_debt_in_default, external_debt_default là có mối tương quan nhất với banking_crisis
- Tỷ giá hối đoái của tất cả quốc gia ở Châu Phi 1860-2014 vào khoảng 45 USD.
- Tỷ lệ vỡ nợ nội địa của tất cả quốc gia ở Châu Phi 1860-2014 là 3.97%
- Tỷ lệ vỡ nợ ngoại vụ của tất cả quốc gia ở Châu Phi 1860-2014 là 15.3%
- Tổng vỡ nợ so với gdp của tất cả quốc gia ở Châu Phi 1860-2014 rơi vào khoảng 0.0064

2. Kết luận

- Tập dữ liệu bao gồm thông tin về nhiều khía cạnh khác nhau của kinh tế và ngân hàng, cho phép nghiên cứu về các yếu tố tác động đến sự phát triển kinh tế và tài chính, biến đổi trong thời gian của các quốc gia châu Phi. Cụ thể hơn là nguyên nhân và kết quả dẫn đến các cuộc khủng hoảng.
- Có thể thấy rõ được cuộc khủng hoảng kinh tế năm 2008 tác động ra sao đến các quốc gia ở Châu Phi, đỉnh điểm là ở Zimbabwe.