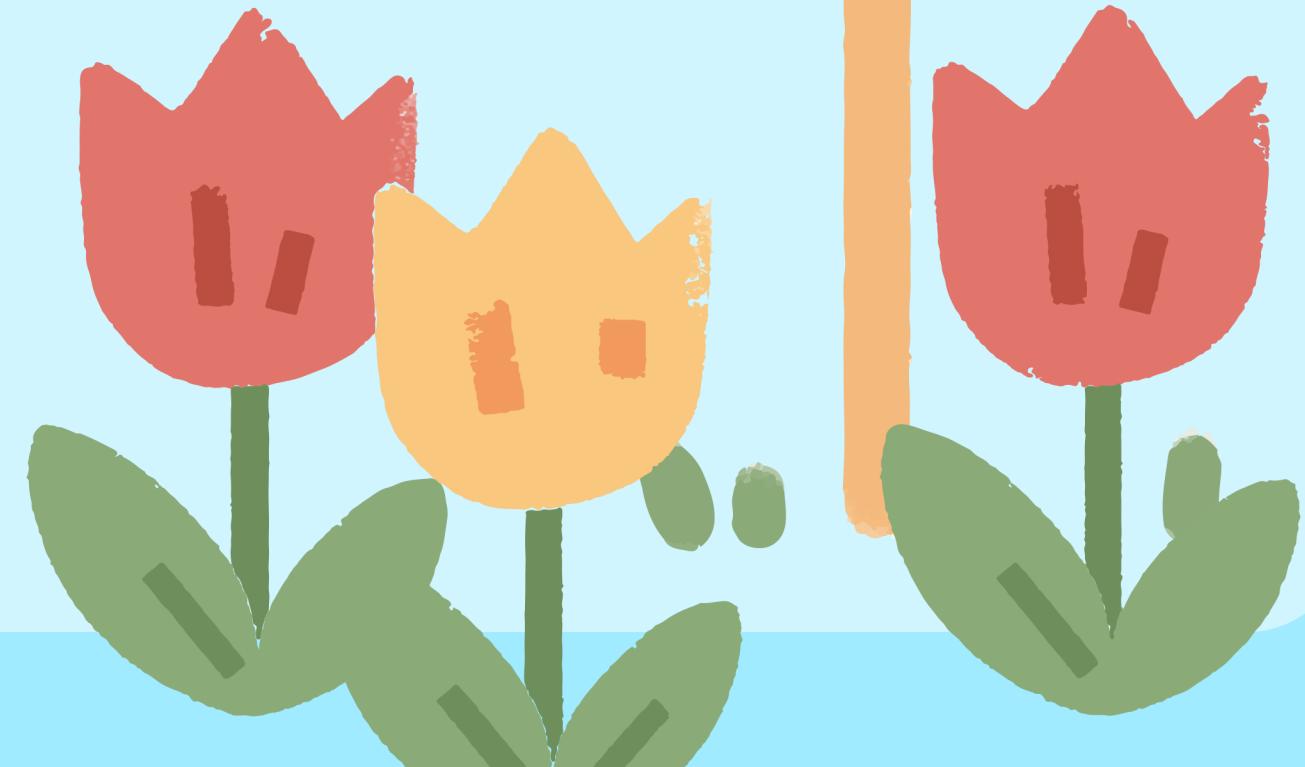


telecom analysis

Team 2



Meet the team 2



Olivia Wilson



Đồng Thị Lan



Nguyễn Thị Linh Chi



Nội dung

Phân I: Khảo sát dữ liệu

- 1.1 Tổng quan về chủ đề
- 1.2 Đặc điểm dữ liệu
- 1.3 Requirement

Phân II: Phân tích và xử lý dữ liệu

- 2.1 Khám phá dữ liệu
- 2.2 Kiến trúc Data Warehouse
- 2.3 Tiền xử lý dữ liệu ETL
- 2.4 Data models

Phân III: Xây dựng báo cáo

- 3.1 Dash 1: Tổng quan
- 3.2 Dash 2: Đặc điểm khách hàng
- 3.3 Dash 3: Tình hình sddv
- 3.4 Dash 4: Số liệu đáng chú ý

Phân IV: Rút ra bài học

- 4.1 Làm việc cá nhân
- 4.2 Làm việc đội nhóm

Phần I: Khảo sát dữ liệu

- 1.1 Tổng quan về chủ đề
- 1.2 Đặc điểm dữ liệu
- 1.3 Requirement



Phần I: Khảo sát dữ liệu

1.1 Tổng quan về chủ đề



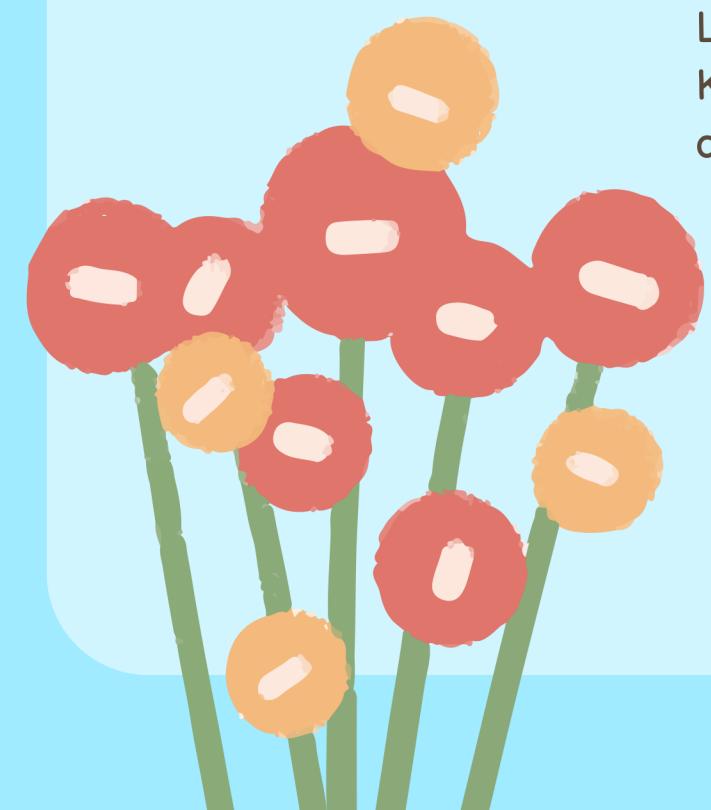
Tại: Công ty cung cấp dữ liệu mạng viễn thông

Tại sao cần phân định kỳ phân tích dữ liệu sử dụng dịch vụ của khách hàng?

01

ĐẶC THÙ SẢN PHẨM

Là sản phẩm dịch vụ/phi vật lý.
Khách hàng tiêu dùng và ở lại vì
dịch vụ



02

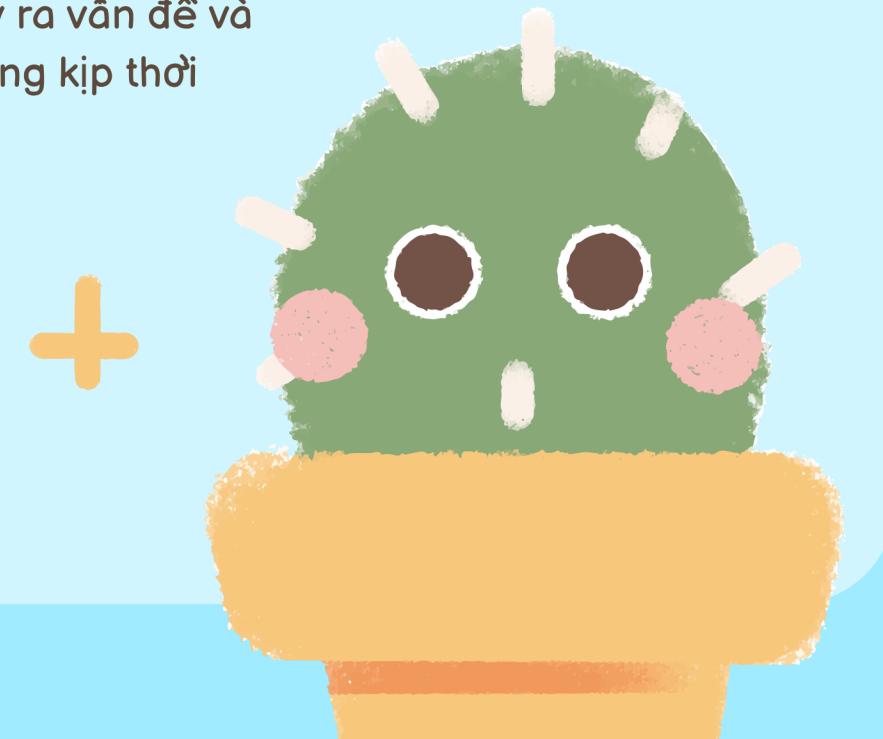
CHI PHÍ CHUYỂN ĐỔI

Duy trì khách hàng cũ có lợi nhuận
hơn so với tìm kiếm khách hàng mới

03

PHÁT HIỆN MỤC TIÊU QUẢN TRỊ

Phân tích khu vực xảy ra vấn đề và
xử lý/ quản trị bán hàng kịp thời



Phần I: Khảo sát dữ liệu

1.2 Đặc điểm dữ liệu

Dữ liệu gốc

Tên Data

Telecom_churn_me.csv

Thời gian

2019-2022

Tần suất cập nhật

Tháng

Số lượng bản ghi

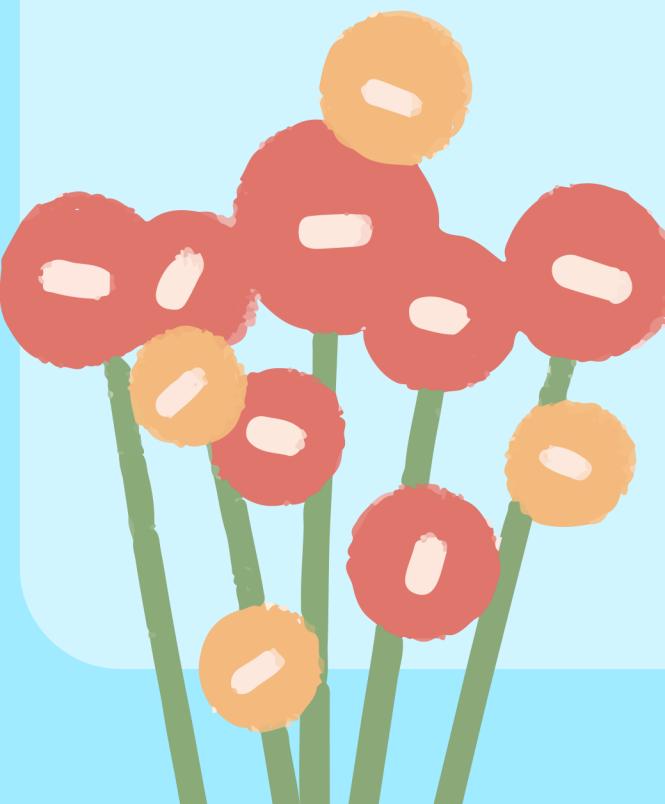
1 bản

Kích thước bộ dữ liệu

44 Mb

Kích thước dl cập nhật mới

1.08 Mb



Phần I: Khảo sát dữ liệu

1.2 Đặc điểm dữ liệu

Row

31,25k

Column

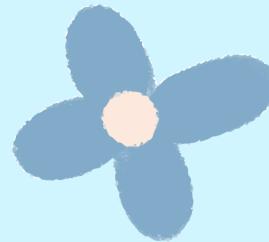
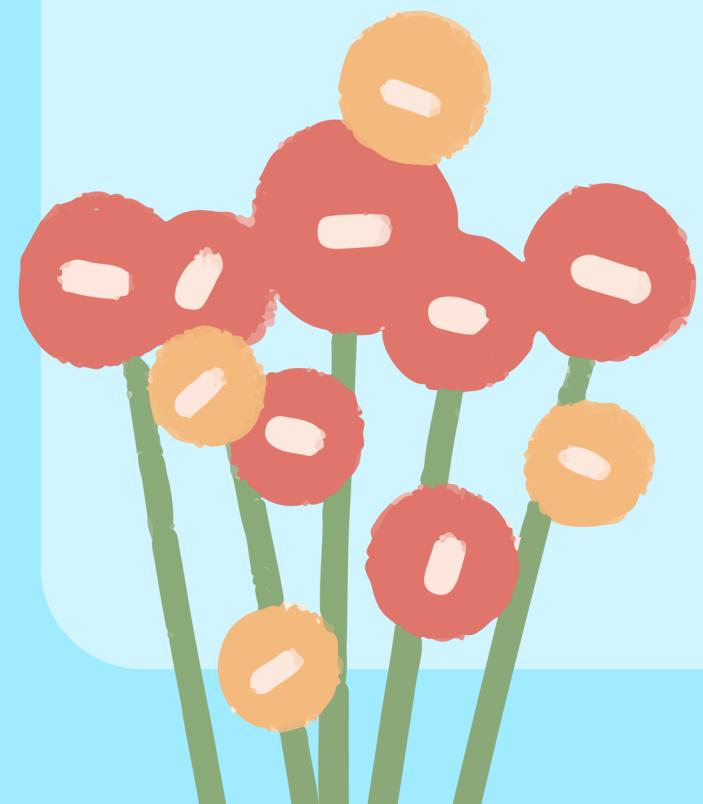
100

Thời gian

61 tháng
hơn 5 năm

Khu vực

4 Miền
19 khu vực



Phần I: Khảo sát dữ liệu

1.2 Đặc điểm dữ liệu

STT	Name old	Translate	New Name
1	rev_Mean	Doanh thu trung bình hàng tháng (số tiền thu)	TB DT/tháng
2	mou_Mean	Số phút sử dụng trung bình hàng tháng	TB phút gọi/tháng
3	totmrc_Mean	Tổng phí định kỳ trung bình hàng tháng	TB phí định kỳ/tháng
4	da_Mean	Số cuộc gọi hỗ trợ danh bạ trung bình	TB cuộc gọi hỗ trợ
5	ovrmou_Mean	Số phút sử dụng vượt mức trung bình	TB phút vượt mức
6	ovrrev_Mean	Doanh thu từ việc sử dụng vượt mức trung bình	TB DT vượt mức
7	vceovr_Mean	Doanh thu từ việc sử dụng thoại vượt mức trung bình	TB DT CGT vượt mức
8	datovr_Mean	Doanh thu từ việc sử dụng dữ liệu vượt mức trung bình	TB DT CGDL vượt mức
9	roam_Mean	Số cuộc gọi chuyển vùng trung bình	TBCG chuyển vùng
10	change_mou	Tỷ lệ thay đổi số phút sử dụng hàng tháng so với trung bình ba tháng trước	Tỷ lệ phút gọi/tháng vs TB 3 tháng trước
11	change_rev	Tỷ lệ thay đổi doanh thu hàng tháng so với trung bình ba tháng trước	Tỷ lệ DT/tháng vs TB 3 tháng trước
12	drop_vce_Mean	Số cuộc gọi thoại bị rớt (thất bại) trung bình	TBCGT thất bại
13	drop_dat_Mean	Số cuộc gọi dữ liệu bị rớt (thất bại) trung bình	TBCGDL thất bại
14	blk_vce_Mean	Số cuộc gọi thoại bị chặn (thất bại) trung bình	TBCGT bị chặn
15	blk_dat_Mean	Số cuộc gọi dữ liệu bị chặn (thất bại) trung bình	TBCGDL bị chặn
16	unan_vce_Mean	Số cuộc gọi thoại không được trả lời trung bình	TBCGT không trả lời
17	unan_dat_Mean	Số cuộc gọi dữ liệu không được trả lời trung bình	TBCGDL không trả lời
18	recv_vce_Mean	Số cuộc gọi thoại nhận được trung bình	TBCGT nhận được
19	recv_sms_Mean	Số tin nhắn thoại nhận được trung bình	TBTNT nhận được
20	comp_vce_Mean	Số cuộc gọi thoại hoàn thành trung bình	TBCGT hoàn thành

Phần I: Khảo sát dữ liệu

1.2 Đặc điểm dữ liệu

21	comp_dat_Mean	Số cuộc gọi dữ liệu hoàn thành trung bình	TBCGDL hoàn thành
22	custcare_Mean	Số cuộc gọi chăm sóc khách hàng trung bình	TBCG cskh
23	ccrndmou_Mean	Số phút sử dụng cuộc gọi chăm sóc khách hàng trung bình (đã làm tròn)	TBPG cskh (lt)
24	cc_mou_Mean	Số phút sử dụng cuộc gọi chăm sóc khách hàng trung bình (chưa làm tròn)	TB PG cskh (clt)
25	inonemin_Mean	Số cuộc gọi vào dưới một phút trung bình	TBCG <1p
26	threeway_Mean	Số cuộc gọi ba bên trung bình	TBCG 3 bên
27	mou_cvce_Mean	Số phút sử dụng cuộc gọi thoại hoàn thành trung bình (chưa làm tròn)	TBPGT hoàn thành (clt)
28	mou_cdat_Mean	Số phút sử dụng cuộc gọi dữ liệu hoàn thành trung bình (chưa làm tròn)	TBPGDL hoàn thành (clt)
29	mou_rvce_Mean	Số phút sử dụng cuộc gọi thoại nhận được trung bình (chưa làm tròn)	TB phút nhận (clt)
30	owylis_vce_Mean	Số cuộc gọi thoại không dây đi ra ngoài trung bình	TBCGKD ngoại mạng
31	mouowylisv_Mean	Số phút sử dụng cuộc gọi thoại không dây đi ra ngoài trung bình (chưa làm tròn)	TBPGKD ngoại mạng (clt)
32	iwylis_vce_Mean	Số cuộc gọi thoại không dây nội mạng	TBCGKD nội mạng
33	mouiwylysv_Mean	Số phút sử dụng cuộc gọi thoại không dây đến trung bình (chưa làm tròn)	TB phút nhận
34	peak_vce_Mean	Số cuộc gọi thoại đỉnh vào và ra trung bình	TBCGT đến-đi lâu nhất
35	peak_dat_Mean	Số cuộc gọi dữ liệu đỉnh trung bình	TBCGDL lâu nhất
36	mou_peav_Mean	Số phút sử dụng cuộc gọi thoại đỉnh trung bình (chưa làm tròn)	TBCGT lâu nhất
37	mou_pead_Mean	Số phút sử dụng cuộc gọi dữ liệu đỉnh trung bình (chưa làm tròn)	TBCGDL lâu nhất
38	opk_vce_Mean	Số cuộc gọi thoại ngoài giờ cao điểm trung bình	TBCGT ngoài giờ G
39	opk_dat_Mean	Số cuộc gọi dữ liệu ngoài giờ cao điểm trung bình	TBCGDL ngoài giờ G
40	mou_opkv_Mean	Số phút sử dụng cuộc gọi thoại ngoài giờ cao điểm trung bình (chưa làm tròn)	TBPGT ngoài giờ G

Phần I: Khảo sát dữ liệu

1.2 Đặc điểm dữ liệu

41	mou_opkd_Mean	Số phút sử dụng cuộc gọi dữ liệu ngoài giờ cao điểm trung bình (chưa làm tròn)	TBPGDL ngoài giờ G
42	drop_blk_Mean	Số cuộc gọi bị rớt hoặc bị chặn trung bình	TBCG rớt/chặn
43	attempt_Mean	Số cuộc gọi được thử trung bình	TBCG gọi thử
44	complete_Mean	Số cuộc gọi hoàn thành trung bình	TBCG hoàn thành
45	callfwdv_Mean	Số cuộc gọi chuyển tiếp trung bình	TBCG chuyển tiếp
46	callwait_Mean	Số cuộc gọi chờ trung bình	TBCG chờ
47	churn	Trường hợp rời bỏ dịch vụ trong khoảng 31-60 ngày sau ngày quan sát	BỎ DV trong tháng thứ 2
48	months	Tổng số tháng sử dụng dịch vụ	Tổng số tháng SDDV
49	uniqsubs	Số lượng thuê bao duy nhất trong hộ gia đình	STB trong gia đình
50	actvsubs	Số lượng thuê bao đang hoạt động trong hộ gia đình	STB trong gia đình active
51	new_cell	Người dùng điện thoại di động mới	Điện thoại mới
52	crlscod	Mã hạng tín dụng	Mã tín dụng
53	asl_flag	Giới hạn chi tiêu tài khoản	TK bị giới hạn
54	totcalls	Tổng số cuộc gọi trong suốt thời gian khách hàng sử dụng	TCG từ đầu dịch vụ
55	totmou	Tổng số phút sử dụng trong suốt thời gian khách hàng sử dụng	TPG từ đầu dịch vụ
56	totrev	Tổng doanh thu	Tổng doanh thu
57	adjrev	Tổng doanh thu điều chỉnh hóa đơn trong suốt thời gian khách hàng sử dụng	DT theo hóa đơn từ đầu
58	adjmou	Tổng số phút gọi điều chỉnh hóa đơn trong suốt thời gian khách hàng sử dụng	SPG theo hóa đơn từ đầu
59	adjqty	Tổng số cuộc gọi điều chỉnh hóa đơn trong suốt thời gian khách hàng sử dụng	SCG theo hóa đơn từ đầu
60	avorev	Doanh thu trung bình hàng tháng trong suốt thời gian khách hàng sử dụng	DTTB hàng tháng từ đầu

Phần I: Khảo sát dữ liệu

1.2 Đặc điểm dữ liệu

61	avgmou	Số phút sử dụng trung bình hàng tháng trong suốt thời gian khách hàng sử dụng	PGTB hàng tháng từ đầu
62	avgqty	Số cuộc gọi trung bình hàng tháng trong suốt thời gian khách hàng sử dụng	CGTB hàng tháng từ đầu
63	avg3mou	Số phút sử dụng trung bình hàng tháng trong ba tháng trước	PGTB 3 tháng gần nhất
64	avg3qty	Số cuộc gọi trung bình hàng tháng trong ba tháng trước	CGTB 3 tháng gần nhất
65	avg3rev	Doanh thu trung bình hàng tháng trong ba tháng trước	DTTB 3 tháng gần nhất
66	avg6mou	Số phút sử dụng trung bình hàng tháng trong sáu tháng trước	PGTB 6 tháng gần nhất
67	avg6qty	Số cuộc gọi trung bình hàng tháng trong sáu tháng trước	CGTB 6 tháng gần nhất
68	avg6rev	Doanh thu trung bình hàng tháng trong sáu tháng trước	DTTB 6 tháng gần nhất
69	prizm_social_one	Chỉ nhom xã hội	Nhóm xã hội
70	area	Khu vực địa lý	Khu vực
71	dualband	Băng tần kép	Bộ phát wifi
72	refurb_new	Thiết bị cầm tay: tân trang hoặc mới	Thiết bị di động: Cũ-Mới
73	hnd_price	Giá thiết bị cầm tay hiện tại	Giá di động
74	phones	Số lượng thiết bị cầm tay được phát hành	Số lượng di động
75	models	Số lượng mẫu được phát hành	Mẫu phát hành
76	hnd_webcap	Khả năng truy cập web của thiết bị cầm tay	Khả năng truy cập web của di động
77	truck	Chỉ số thiết bị lỗi	Thiết bị lỗi hay không
78	rv	Chỉ số dùng 5G/LTE	Dùng 4G-5G hay ko?
79	ownrent	Tình trạng chủ nhà/thuê nhà	Tình trạng cư trú
80	lor	Thời gian cư trú	Thời gian cư trú

Phần I: Khảo sát dữ liệu

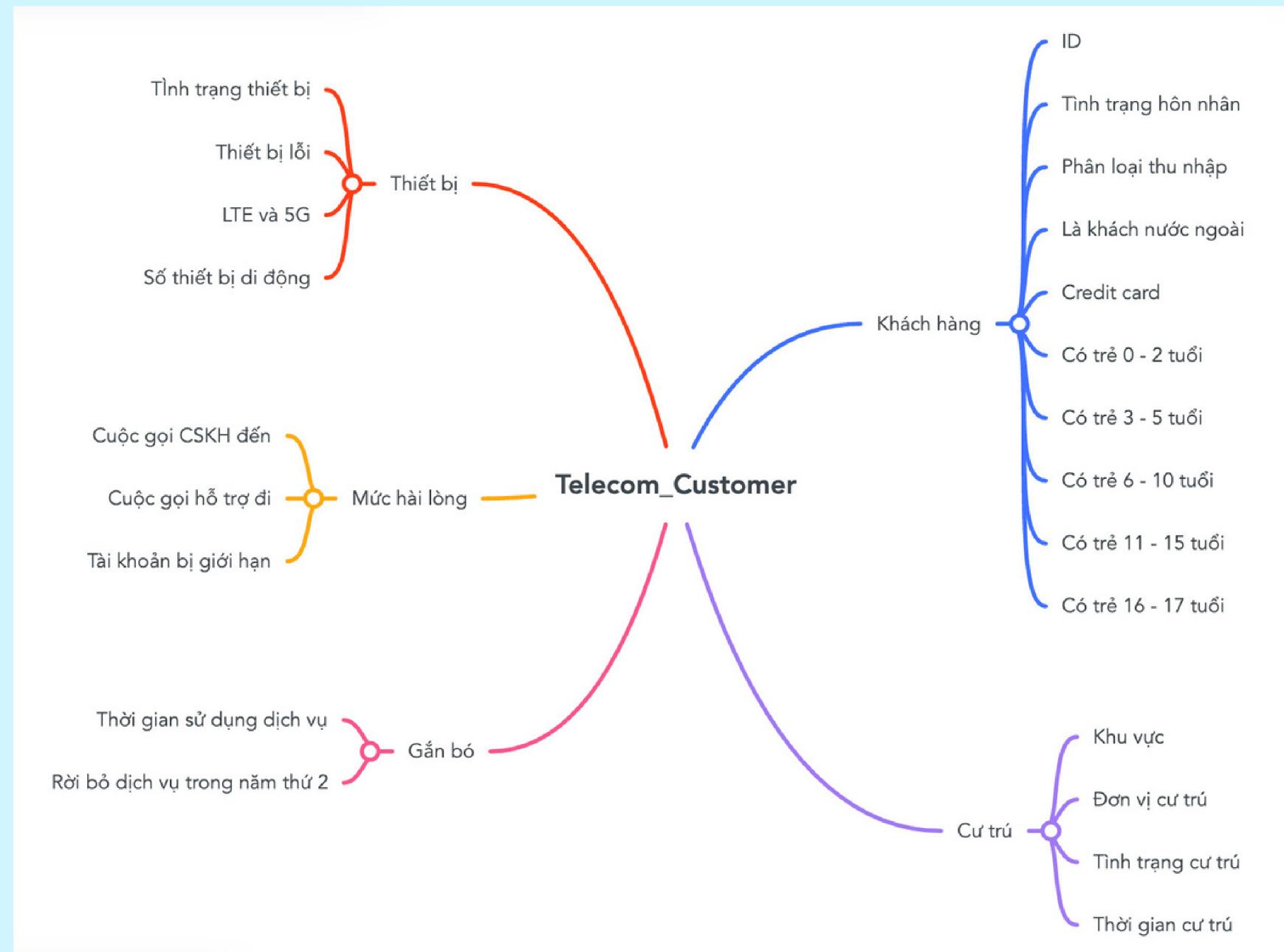
1.2 Đặc điểm dữ liệu

81	dwltype	Loại đơn vị cư trú	Đơn vị cư trú
82	marital	Tình trạng hôn nhân	Tình trạng hôn nhân
83	adults	Số lượng người lớn trong hộ gia đình	Số người lớn
84	infobase	Khớp với cơ sở dữ liệu thông tin	Khớp dl
85	income	Thu nhập ước tính	Nhóm thu nhập
86	numbcars	Số lượng xe được biết	Số xe
87	HHstatin	Chỉ số tình trạng hộ gia đình cao cấp	Mã HGD cao cấp
88	dwlsize	Kích thước đơn vị cư trú	Loại quy mô cư trú
89	forgntvl	Biển giả du lịch nước ngoài	Khách nước ngoài
90	ethnic	Mã tổng hợp dân tộc	Mã dân tộc
91	kid0_2	Trẻ em từ 0 - 2 tuổi trong hộ gia đình	Kid0-2
92	kid3_5	Trẻ em từ 3 - 5 tuổi trong hộ gia đình	Kid3-5
93	kid6_10	Trẻ em từ 6 - 10 tuổi trong hộ gia đình	Kid6-10
94	kid11_15	Trẻ em từ 11 - 15 tuổi trong hộ gia đình	Kid11-15
95	kid16_17	Trẻ em từ 16 - 17 tuổi trong hộ gia đình	Kid16-17
96	creditcd	Chỉ số thẻ tín dụng	Dùng thẻ tín dụng hay không
97	eqpdays	Số ngày (tuổi) của thiết bị hiện tại	Tuổi thiết bị
98	Customer_ID	Số ID của khách hàng	ID khách hàng

Phần I: Khảo sát dữ liệu

1.2 Đặc điểm dữ liệu

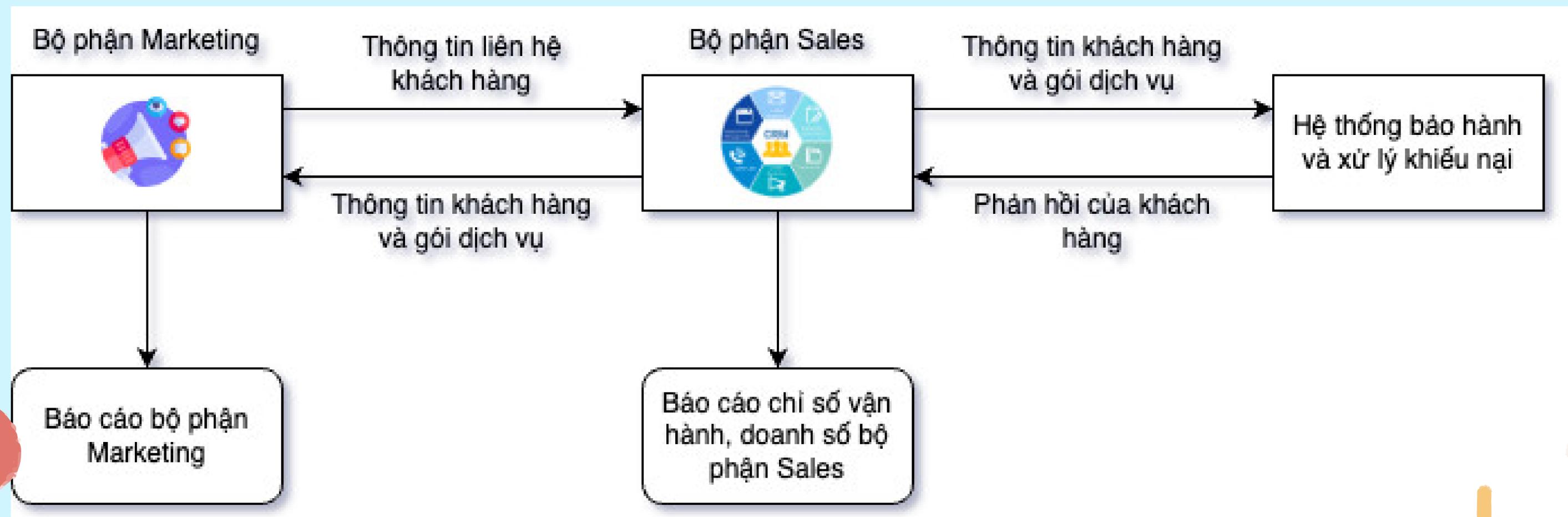
Cây phân loại dữ liệu



Phần I: Khảo sát dữ liệu

1.2 Đặc điểm dữ liệu

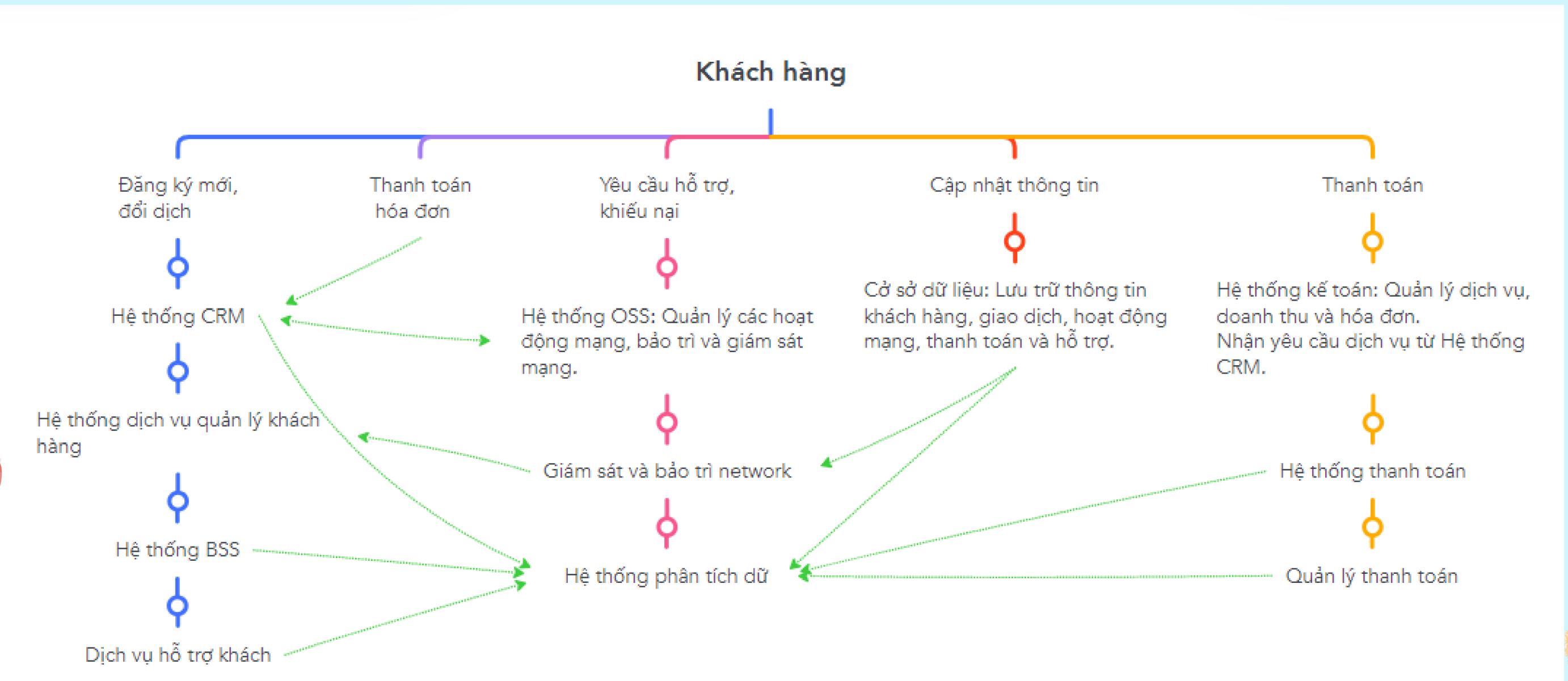
Luồng dữ liệu hệ thống CNTT bộ phận bán hàng



Phần I: Khảo sát dữ liệu

1.2 Đặc điểm dữ liệu

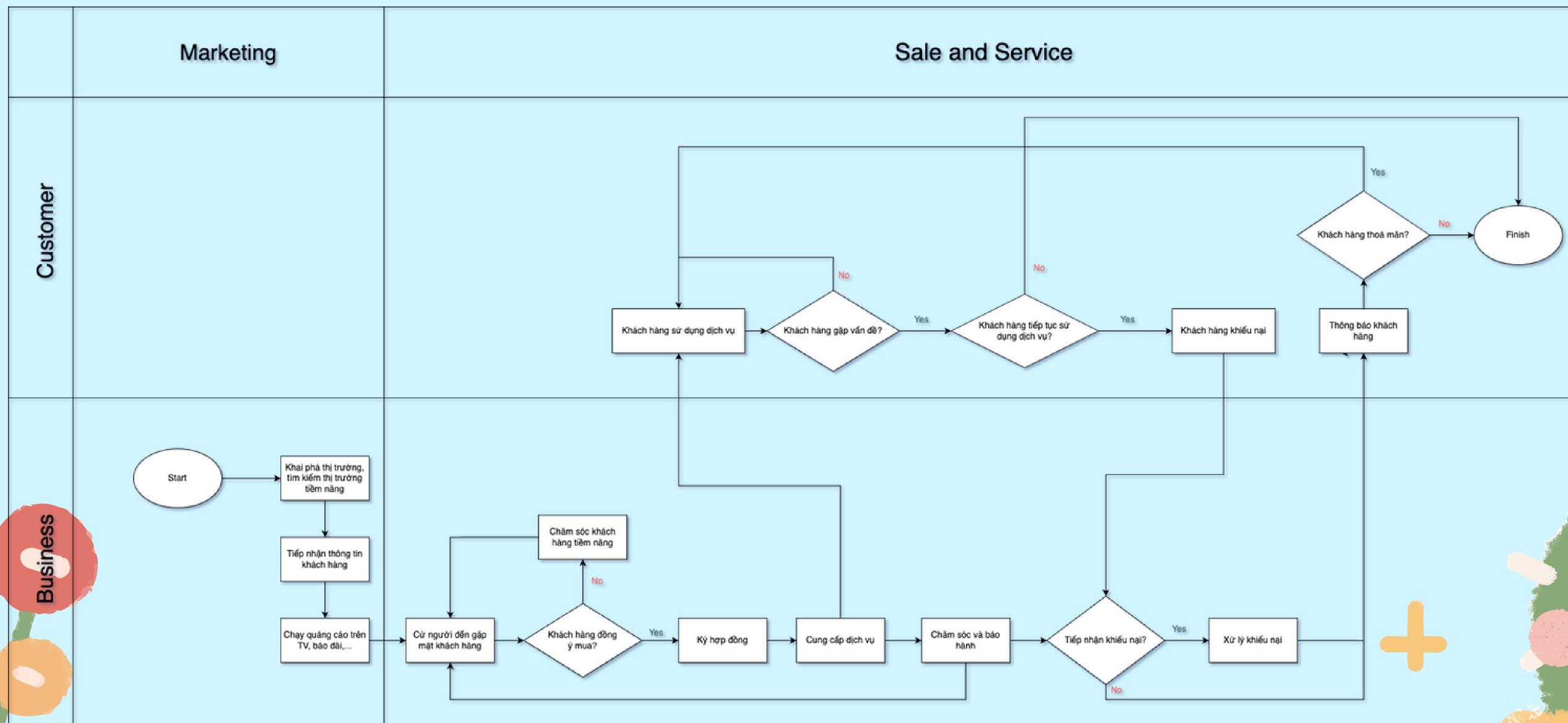
Luồng dữ liệu hệ thống CNTT bộ phận IT



Phần I: Khảo sát dữ liệu

1.3 Requirement

Quy trình bán dịch vụ



Phần I: Khảo sát dữ liệu

1.3 Requirement

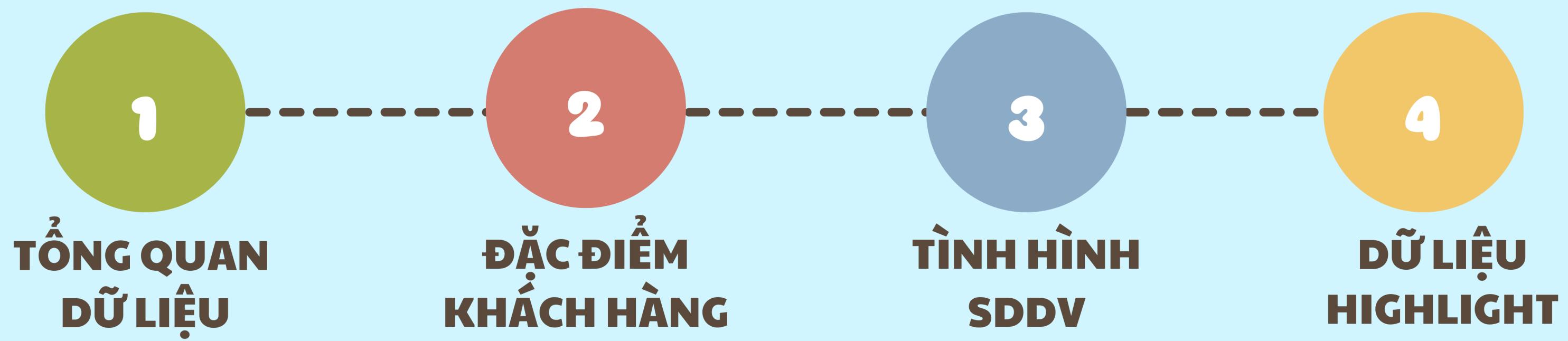
Mô hình kinh doanh Canvas

Đối tác chính <ul style="list-style-type: none">Nhà cung cấp thiết bị mạngBên cung cấp nội dung truy cập như nhà mạng,...Cơ quan quản lýĐối tác bán lẻ như đại lý, cửa hàng, siêu thị,..Đối tác công nghệBên cung cấp cơ sở hạ tầng: xây dựng, đường dẫn, kho bãi,...	Hoạt động chính <ul style="list-style-type: none">Quản lý và bảo trì mạngDịch vụ khách hàng và hỗ trợMarketing và bán hàngQuản lý hóa đơn và doanh thuNghiên cứu và phát triểnTuân thủ quy định Nguồn lực chính <ul style="list-style-type: none">Hàng hóaNhân sựThương hiệu và uy tínCông nghệ và phần mềmNền tảng tài chính	Giá trị cốt lõi <ul style="list-style-type: none">Phạm vi phủ sóng rộngKết nối đáng tin cậyBảo mậtTốc độ caoHỗ trợ khách hàngDịch vụ gia tăng giá trịGiải pháp sáng tạoGói dùng linh hoạt	Quan hệ khách hàng <ul style="list-style-type: none">Hỗ trợ cá nhânTự phục vụDịch vụ tự độngChương trình gắn kết cộng đồng người dùng	Phân khúc khách hàng <ul style="list-style-type: none">Người dùng cá nhânDoanh nghiệpChính phủ và khu hành chính côngCác bên viễn thông khácKhách hàng quốc tế
Cơ cấu chi phí <ul style="list-style-type: none">Chi phí hạ tầng mạngChi phí hoạt độngChi phí Marketing và bán hàngChi phí nghiên cứu và phát triểnPhi và chi phí tuân thủ điều khoản hợp đồng	Kênh phân phối <ul style="list-style-type: none">Cửa hàng bán lẻBán trực tuyếnHệ thống cửa hàngTrung tâm dịch vụ khách hàngĐại lý đối tácỨng dụng di động	Dòng doanh thu <ul style="list-style-type: none">Phi thuê baoPhi sử dụng trả theo lượtBán thiết bịDịch vụ gia tăng giá trịDoanh thu quảng cáoDịch vụ bán sỉ		

Phần I: Khảo sát dữ liệu

3. Requirement

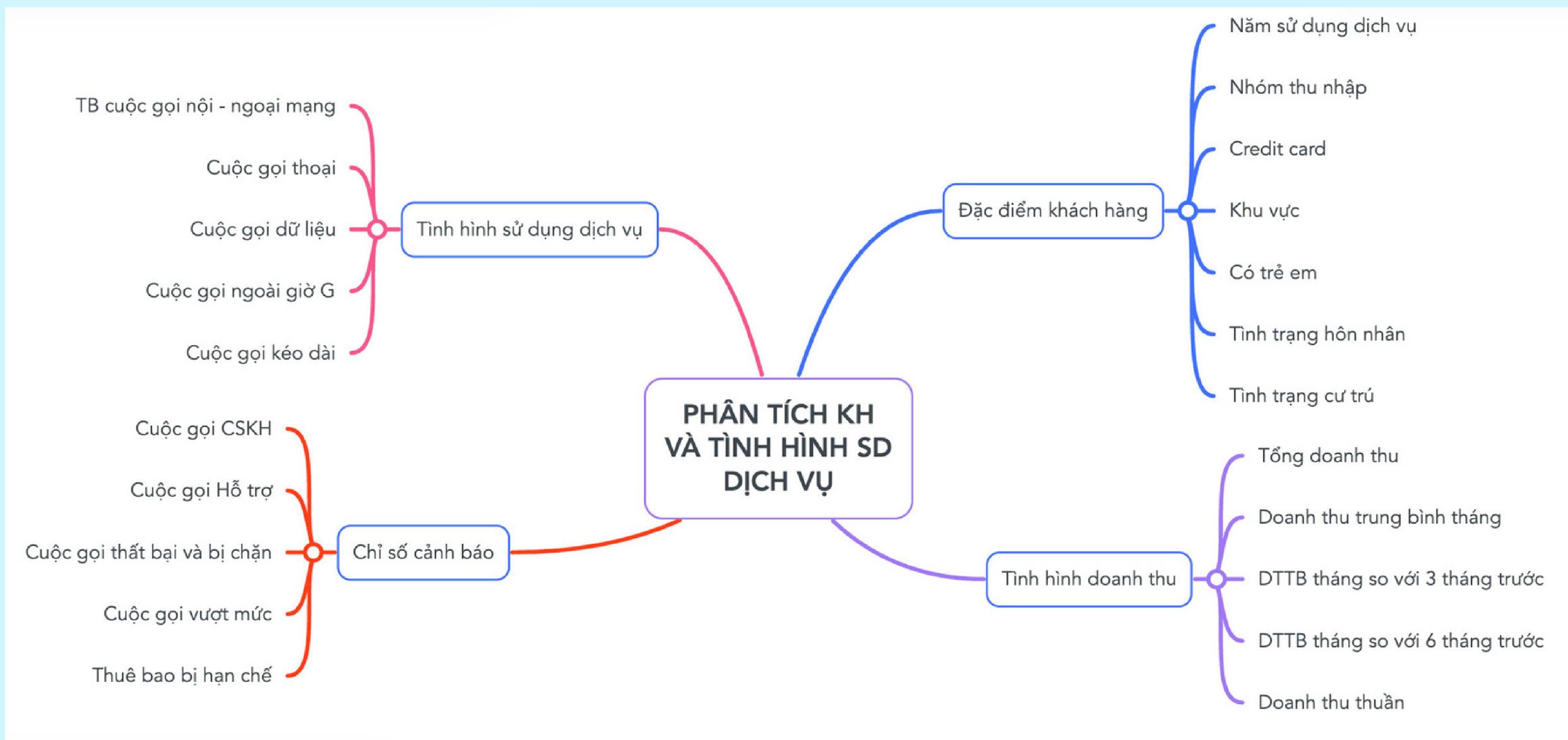
3.3. Chủ điểm phân tích



Phần I: Khảo sát dữ liệu

1.3 Requirement

Cây yêu cầu



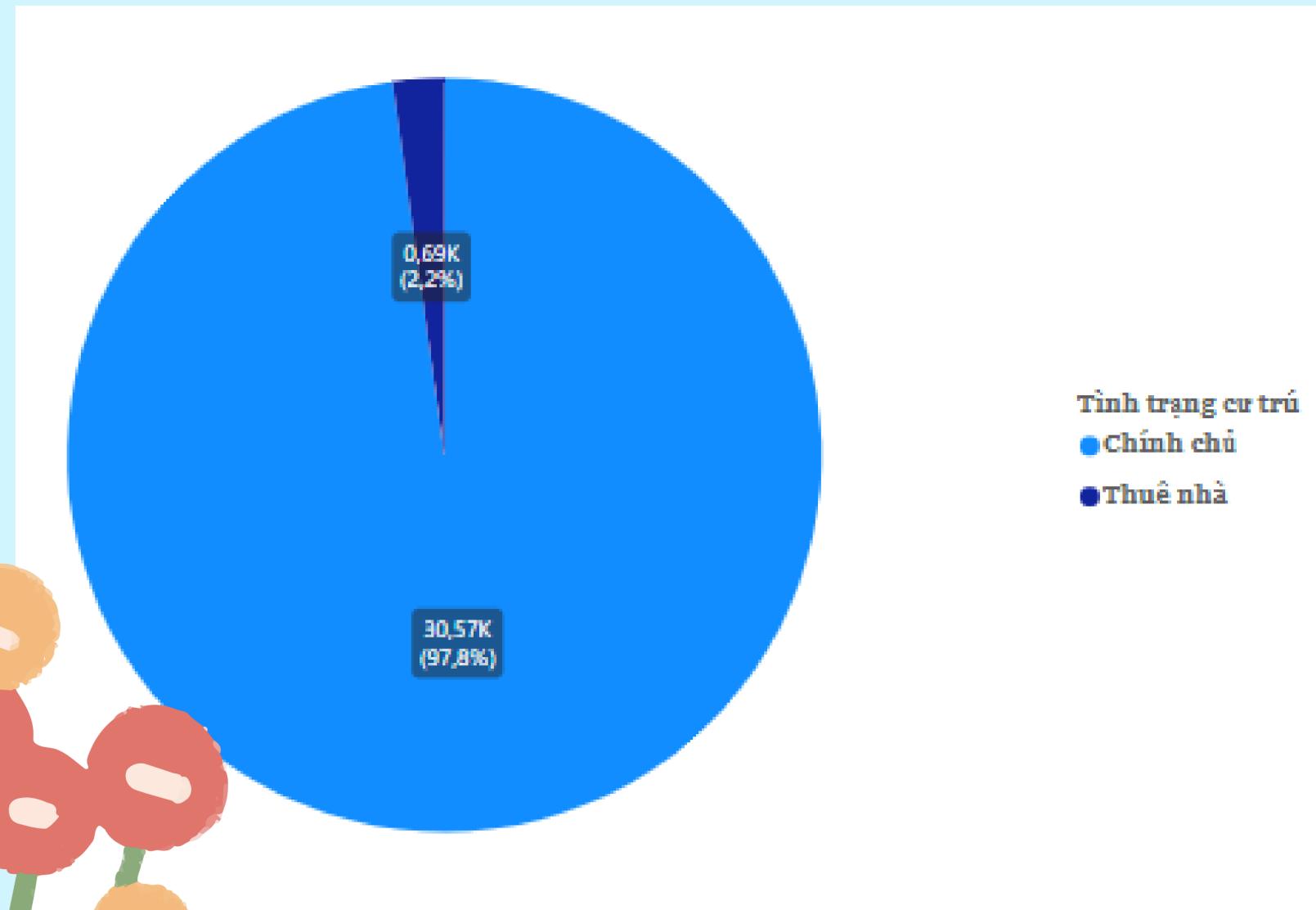
Phần II: Phân tích và xử lý dữ liệu ETL

- 2.1. Khám phá dữ liệu
- 2.2 Kiến trúc Data Warehouse
- 2.3 Tiền xử lý dữ liệu ETL
- 2.4 Dimension, Data models



Phần II: Phân tích và xử lý dữ liệu ETL

2.1 Khám phá dữ liệu

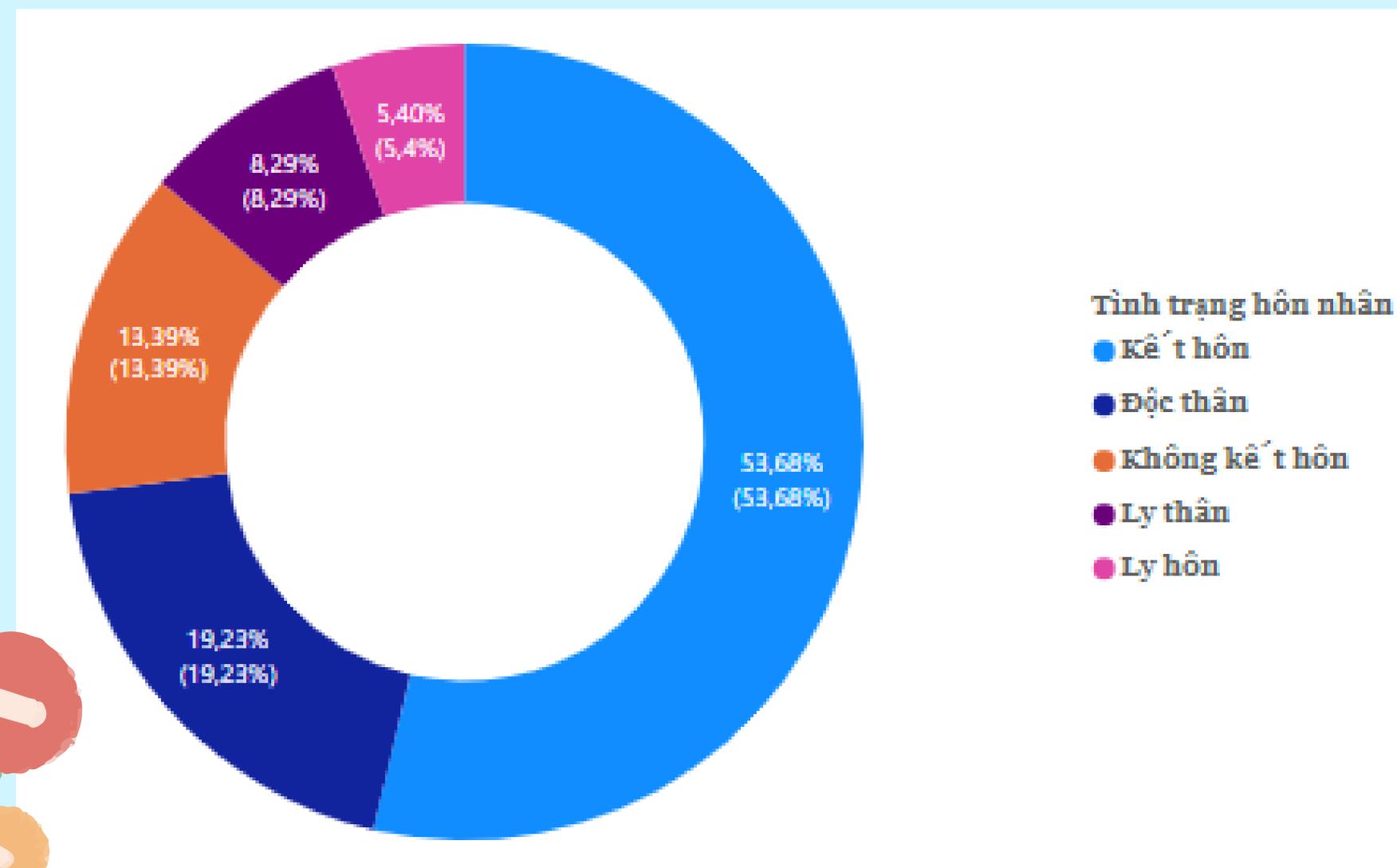


1. Tỷ lệ chính chủ cao (97%)

- **Sự ổn định:**
 - Tỷ lệ cư trú chính chủ cao cho thấy mức độ ổn định về nơi ở của dân cư. Điều này có thể phản ánh rằng đa số người dân có sở hữu bất động sản riêng, giúp họ có cuộc sống ổn định và an toàn hơn.
- **Sự đầu tư vào tài sản:**
 - Điều này cũng cho thấy rằng nhiều người dân đã đầu tư vào bất động sản. Sự đầu tư này có thể là một dấu hiệu của nền kinh tế phát triển, khi người dân có khả năng tài chính để mua nhà và không phải phụ thuộc vào việc thuê nhà.
- **Tính chất cộng đồng:**
 - Với tỷ lệ chính chủ cao, các khu dân cư có thể phát triển tính chất cộng đồng mạnh mẽ hơn. Người dân có xu hướng gắn bó với nơi họ sống và có thể tham gia nhiều hơn vào các hoạt động cộng đồng và giữ gìn môi trường sống.

Phần II: Phân tích và xử lý dữ liệu ETL

2.1 Khám phá dữ liệu

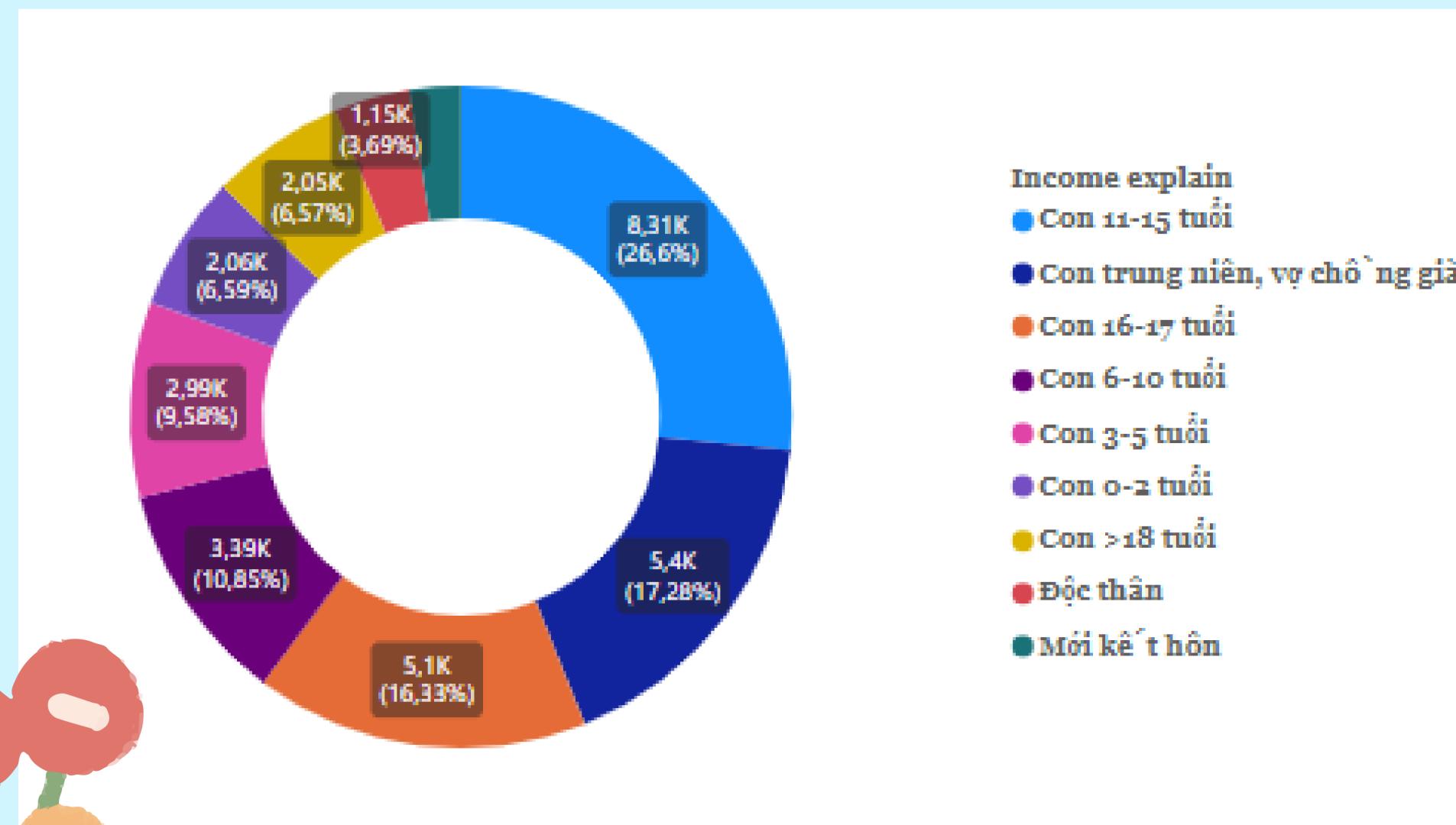


Kết hôn (53,68%)

- Phổ biến nhất: Hơn một nửa dân số đã kết hôn, cho thấy hôn nhân vẫn là hình thức quan hệ phổ biến nhất trong xã hội.
- Ổn định xã hội: Tỷ lệ này cho thấy mức độ ổn định và cam kết lâu dài trong quan hệ gia đình.
- Tình trạng hôn nhân trong xã hội rất đa dạng, với hầu hết dân số kết hôn nhưng cũng có sự hiện diện đáng kể của những người độc thân, không kết hôn, ly thân và ly hôn. Điều này cho thấy một xã hội với nhiều lựa chọn và quan điểm khác nhau về quan hệ cá nhân và gia đình.

Phần II: Phân tích và xử lý dữ liệu ETL

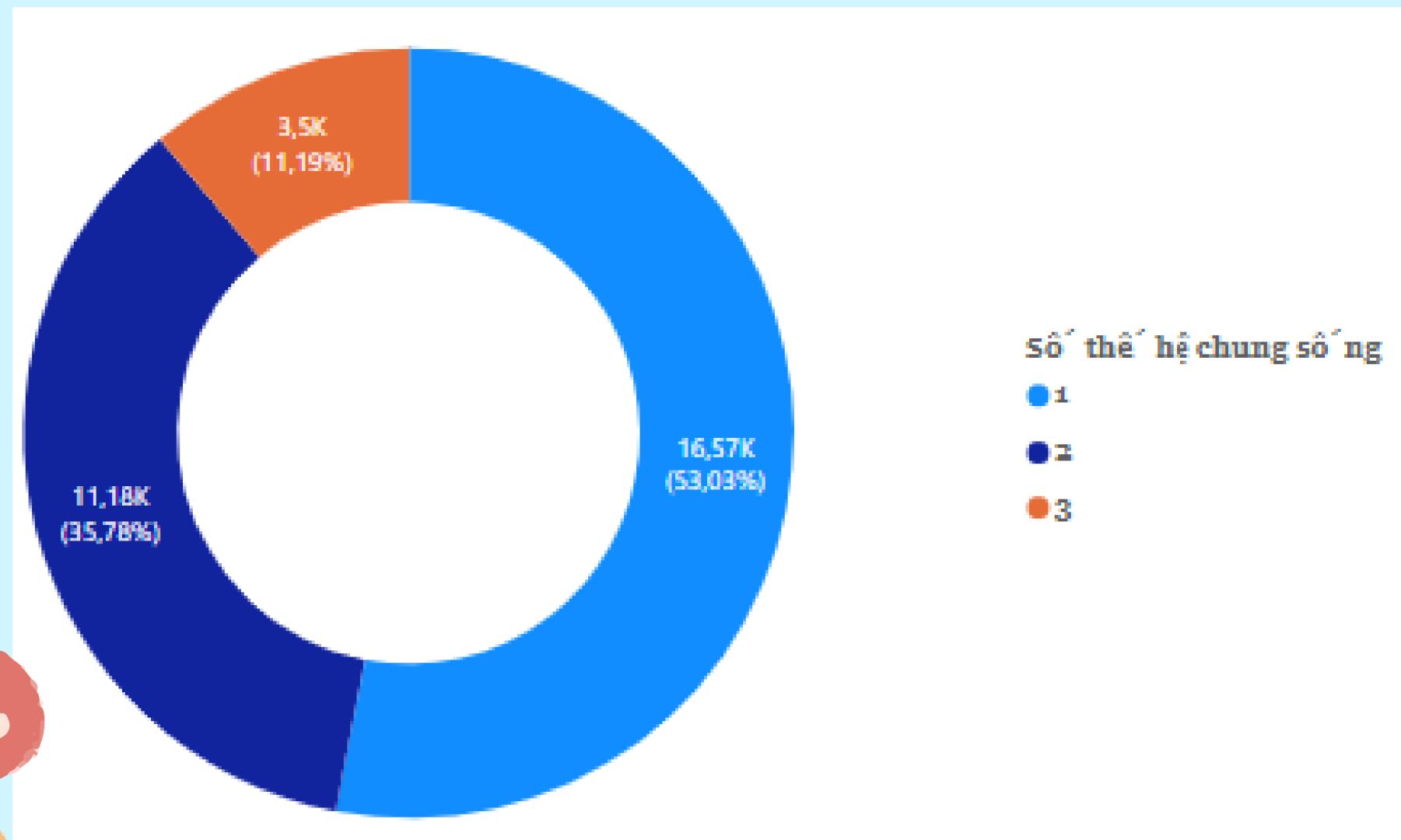
2.1 Khám phá dữ liệu



Đa phần cơ cấu năm về lứa tuổi từ 30 trở lên, lứa 18-29 ít hơn hẳn. Cho thấy sự già hóa dân số trong cơ cấu người dùng

Phần II: Phân tích và xử lý dữ liệu ETL

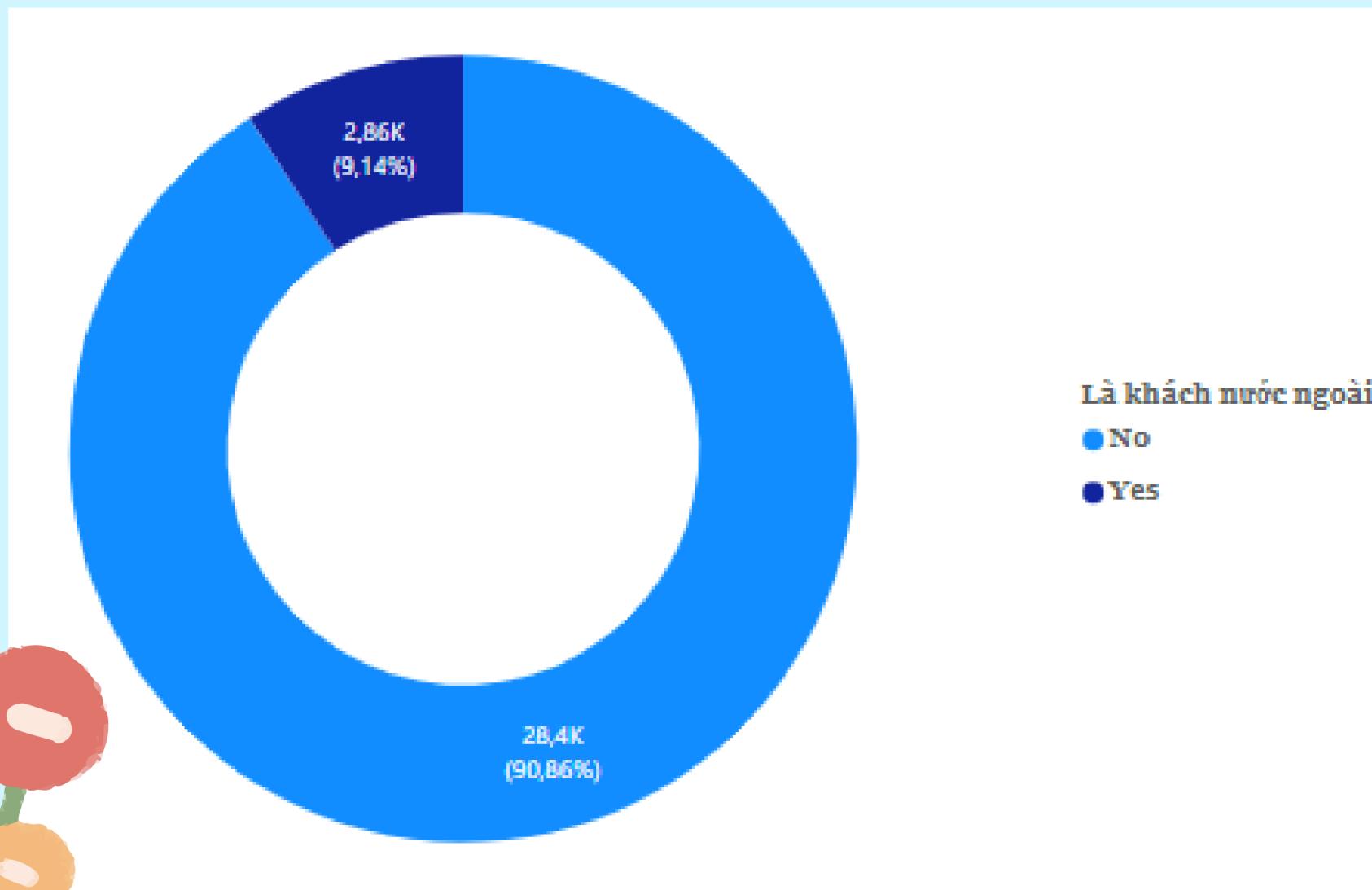
2.1 Khám phá dữ liệu



Cơ cấu các thể hộ sống chung cho thấy một sự pha trộn giữa mô hình gia đình truyền thống và hiện đại. Trong khi gia đình ba thế hệ vẫn chiếm ưu thế, có một phân đàng kể các gia đình hạt nhân và người sống độc lập, phản ánh sự đa dạng và thay đổi trong cấu trúc gia đình và lối sống trong xã hội hiện đại.

Phần II: Phân tích và xử lý dữ liệu ETL

2.1 Khám phá dữ liệu



Khách hàng trong nước chiếm ưu thế:

Phân lớn doanh nghiệp khách hàng là nội địa: Với hơn 90% khách hàng doanh nghiệp là trong nước, điều này cho thấy sự phụ thuộc lớn vào thị trường nội địa. Sự ổn định và quen thuộc: Doanh nghiệp có thể đang tập trung vào việc phục vụ các khách hàng nội địa, nơi họ có sự hiểu biết sâu sắc về thị trường và các mối quan hệ kinh doanh.

Khách hàng nước ngoài chiếm dưới 10%:

Thị trường quốc tế chưa phát triển: Tỷ lệ khách hàng nước ngoài thấp cho thấy doanh nghiệp chưa mở rộng mạnh mẽ ra thị trường quốc tế hoặc chưa thu hút được nhiều khách hàng quốc tế.

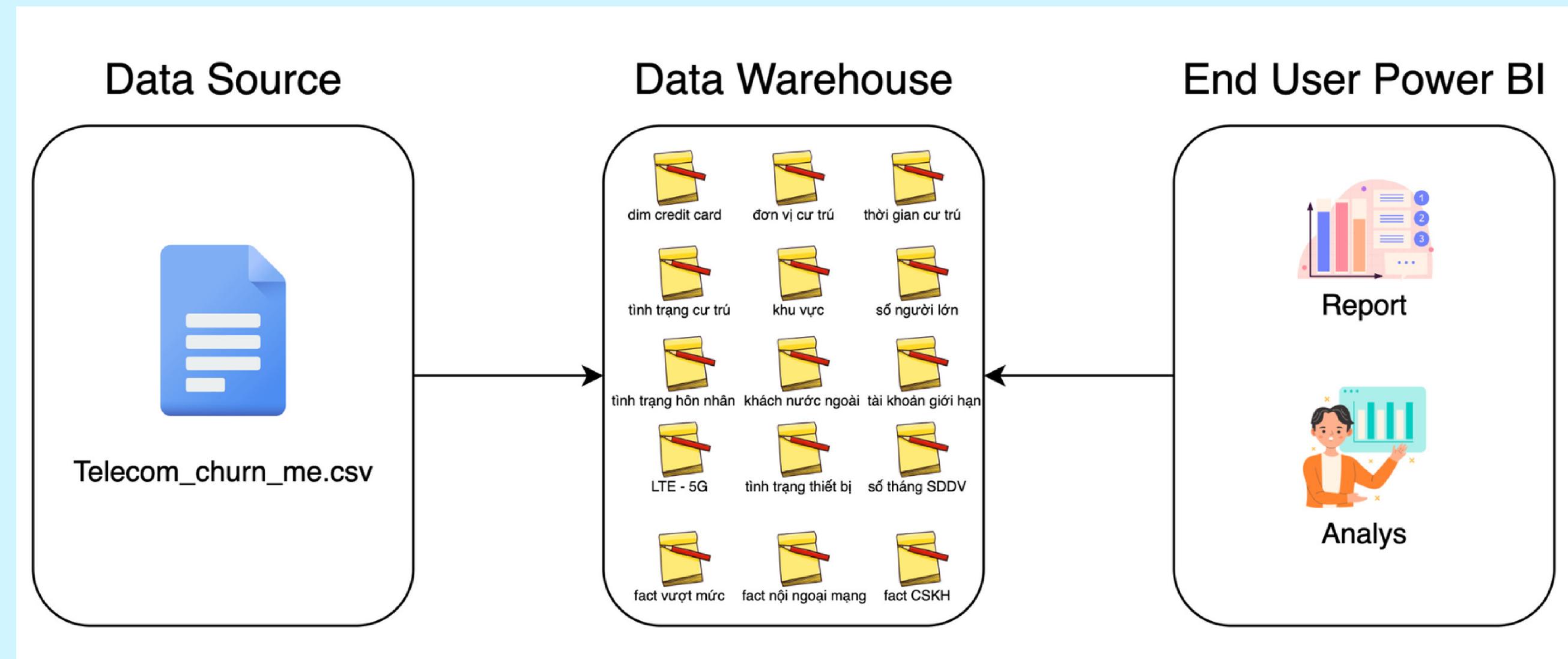
Tiềm năng tăng trưởng: Đây có thể là cơ hội để doanh nghiệp mở rộng và khai thác thị trường quốc tế, tăng cường sự hiện diện toàn cầu và giảm sự phụ thuộc vào thị trường nội địa.

Chiến lược thị trường: Doanh nghiệp có thể đang áp dụng chiến lược tập trung vào thị trường nội địa, tối ưu hóa dịch vụ và sản phẩm cho khách hàng trong nước.

Nguồn lực và chuyên môn: Việc tập trung vào thị trường nội địa có thể giúp doanh nghiệp tối ưu hóa nguồn lực và chuyên môn hóa trong các lĩnh vực mà họ mạnh nhất.

Phần II: Phân tích và xử lý dữ liệu ETL

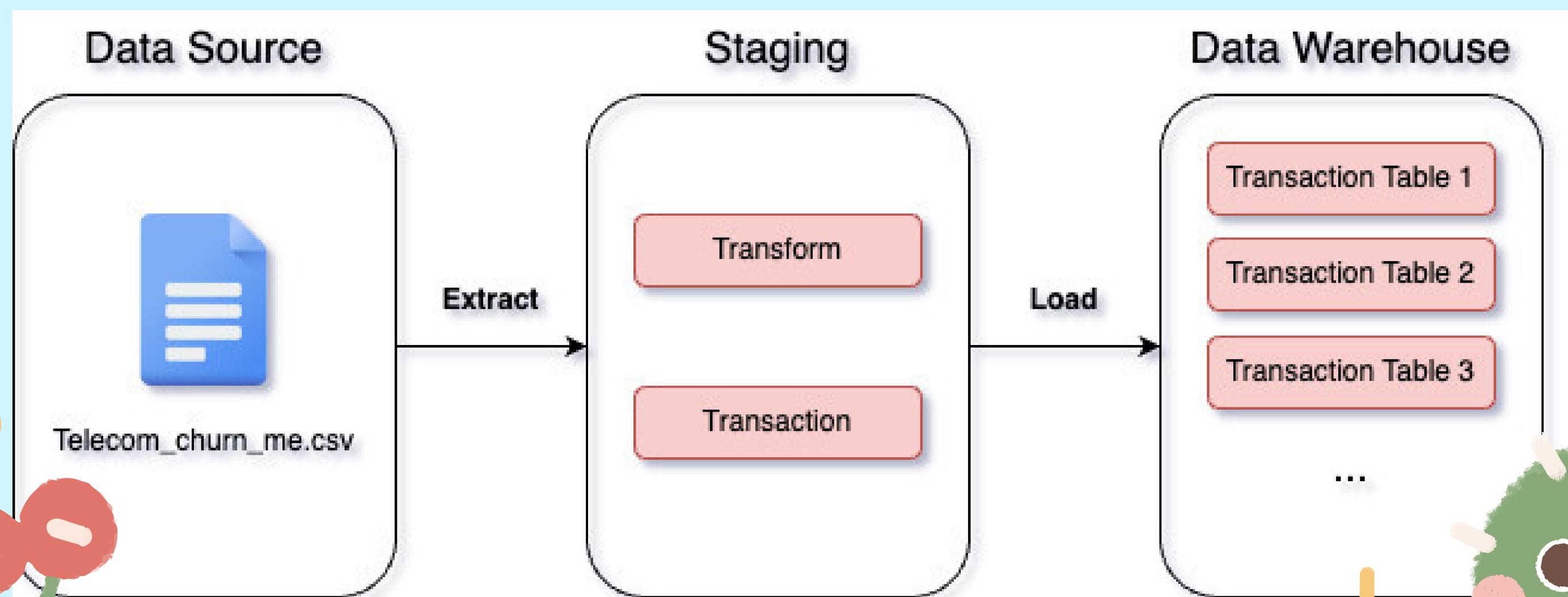
2.2 Kiến trúc Data Warehouse



Phần II: Phân tích và xử lý dữ liệu ETL

2.3 Tiền xử lý dữ liệu ETL

Sơ đồ ETL



Phần II: Phân tích và xử lý dữ liệu ETL

2.3 Tiền xử lý dữ liệu ETL

Xử lý ETL

Kiểu tra type, đổi type cột

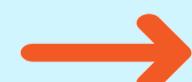
	1.2 DT bill/TDT	ABC 123 Discount KH
4429	0,959781468	0,0402185
4952	0,902412992	0,0975870
2789	0,954135414	0,0458645



	% DT bill/TDT	% Discount KH
429	95,98%	4,02%
952	90,24%	9,76%
789	95,41%	4,59%

Xóa dữ liệu trống/ lõi

	= Table.Rem
1	1 ² ₃ rev_Mean
2	239975
3	574925
4	1699
	38



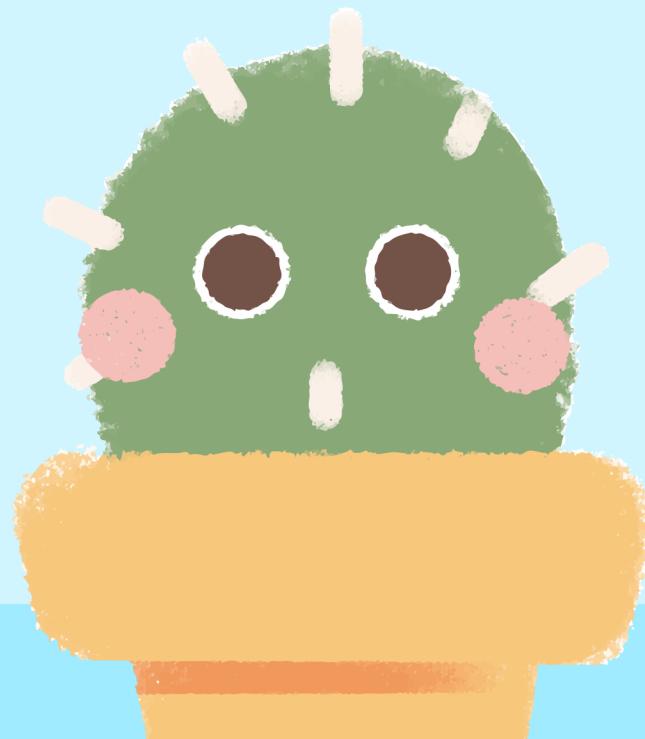
	= Table.Select
1	1 ² ₃ rev_Mean
2	239975
3	1699
4	5523
5	380525
	316625

Xóa cột

	Query	Manage Columns	Reduce Rows	Sort	Transform
e	.SelectRows(#"Filtered Rows4", each not List.IsEmpty(List.RemoveMatchingItems				
_dat_Mean	0	1 ² ₃ plcd_vce_Mean	1 ² ₃ plcd_dat_Mean	1 ² ₃ recv_vce_Mean	1 ²
	0	5233333333	0	4233333333	
	0	9	0	3333333333	
	0	2223333333	0	9466666667	
	0	6666666667	0	0	
	0	3446666667	0	2713333333	
	0	1383333333	0	53	
	0	3446666667	0	0	
	0	1383333333	0	2713333333	



	Table.RemoveColumns(#"Removed Blank Rows7",{"plcd_vce_Mean", "plcd_dat_Mean", "recv_vce_Mean", "recv_dat_Mean", "recv_sms_Mean", "recv_vce_Mean", "recv_dat_Mean", "recv_sms_Mean"}))	unan_dat_Mean	1 ² ₃ recv_vce_Mean	1 ² ₃ recv_sms_Mean
		0	4233333333	
		0	3333333333	
		0	9466666667	
		0	0	
		0	2713333333	



Phần II: Phân tích và xử lý dữ liệu ETL

2.3 Tiền xử lý dữ liệu ETL

Xử lý ETL

Tạo DIMS

A_C_kid16_17	A_C_credited	123_eqpdays
U	Y	361
U	Y	1504
U	Y	434
U	Y	601
U	Y	199



A_C_credited
Y
Y
Y
Y
Y
Y



A_C_credited
Y
N

Đổi tên

A_C_credited
1
2



A_C_Credit_card
Y
N

Sắp xếp

A_C_ethnic
N
I
S
Z
J
H
G
D
F
U
O
R
B
M
P
X



A_C_ethnic
B
D
F
G
H
I
J
M
N
O
P
R
S
U
X
Z

Phần II: Phân tích và xử lý dữ liệu ETL

2.3 Tiền xử lý dữ liệu ETL

Xử lý ETL

Tách cột tạo dims

The screenshots illustrate the process of splitting a single column into multiple dimensions:

- Screenshot 1:** Shows a table with a single column 'area' containing various geographical areas like 'NORTHWEST/ROCKY MOUNTAIN ...', 'GREAT LAKES AREA', etc.
- Screenshot 2:** A 'Split Column by Number of Characters' dialog box is open. It shows the 'Number of characters' input field set to 4, and the 'Once, as far right as possible' radio button selected. Preview rows show the data being split into three columns: 'area.1' (e.g., 'NORTHWEST/ROCKY MOUNTAIN'), 'area.2' (e.g., 'AREA'), and 'dualband' (e.g., 'Y').
- Screenshot 3:** The result of the split operation, showing three new columns: 'area.1', 'area.2', and 'dualband'. The 'area.1' column contains values like 'NORTHWEST/ROCKY MOUNTAIN', 'GREAT LAKES', 'NEW ENGLAND', etc. The 'area.2' column contains 'AREA' repeated for each row. The 'dualband' column contains 'Y' for most rows and 'N' for some.
- Screenshot 4:** A final table titled 'Changed Type1' showing the transformed data. The 'area.1' column now contains specific area names, and the 'area.2' column contains 'N' or 'Y' values.

Thêm cột điều kiện

The screenshots illustrate the creation of a conditional column:

- Screenshot 1:** An 'Add Conditional Column' dialog box. It shows a condition: 'If Credit_card equals Y Then 123 = Yes Else No'. The 'Output' dropdown is set to '123'.
- Screenshot 2:** The resulting table with a new column 'Custom'. The 'Credit_card' column has values 'Y' and 'N'. The 'Custom' column contains 'Yes' for rows where 'Credit_card' is 'Y' and 'No' for rows where 'Credit_card' is 'N'.

Phần II: Phân tích và xử lý dữ liệu ETL

2.3 Tiền xử lý dữ liệu ETL

Xử lý ETL

Thêm các cột điều kiện khác

= Table.AddColumn(#"Removed Duplicates", "kid11_15", Custom, {{"U", "Unknown"}, {"Y", "Yes"}})	
1	U
2	Y

= Table.RenameColumns(#"Added Condition", "new_cell", Explain)	
1	U
2	Y
3	N

= Table.AddColumn(#"Removed Duplicates", "ownrent", Custom, {{"O", "Chính chủ"}, {"R", "Thuê nhà"}})	
1	O
2	R

Add Conditional Column
Add a conditional column that is computed from the other columns or values.
New column name: Nội mảng ưu thế
Column Name: %Nội mảng/Ngoại mảng
Operator: is greater than or...
Value: 100
Then: Yes
Else: No
OK Cancel

Custom Column
Add a column that is computed from the other columns.
New column name: Discount KH
Custom column formula: = if [TBDT tháng từ đầu] >= 100 then 100 else 0
Available columns:
DT bill từ đầu
TPG từ đầu dịch vụ
TOG từ đầu dịch vụ
Tổng doanh thu
Số tháng SĐDV
TBDT tháng từ đầu
DT bill/TDT
OK Cancel

Duplicate Column
FORMAT Parse From Text
10³ Scientific
From Number
Query
PRC
NaN
Fac
All F
APF
= Table.RenameColumns(#"Inserted", "Percent Off", {"%Nội mảng/Ngoại mảng", "DT vượt/DT tháng"})
Customer_ID
1 1275 1035257 0,002815183
2 0 1045516 12,16039187
OK Cancel

Column
Parse
10³ Scientific
From Number
Query
PRC
NaN
Fac
All F
APF
= Table.AddColumn(#"Renamed Columns", "Nội mảng ưu thế", each if [#%"Nội mảng/Ngoại mảng"] >= 100 then "Yes" else "No")
TBCGKD nội mảng
19 7
OK Cancel

Phần II: Phân tích và xử lý dữ liệu ETL

2.3 Tiền xử lý dữ liệu ETL

Hệ thống Dimensions

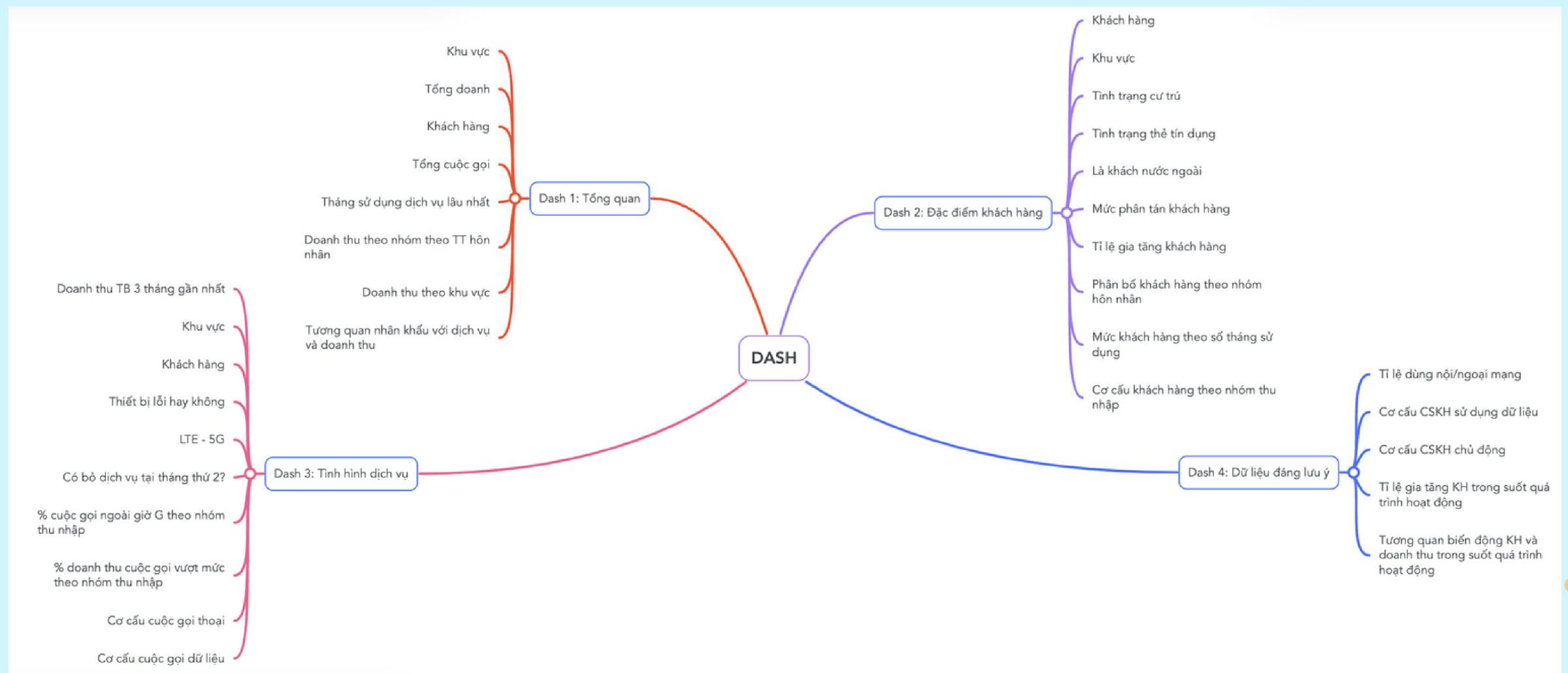
2	3	2	61	6	2	2	2	2
LTE_5G	Tình trạng thiết bị mới	Bổ dịch vụ	Tháng SDDV	Năm SDDV	Khách nước ngoài	Thiết bị báo lỗi	Tk bị giới hạn	
Yes	Yes	Yes	1	Năm đầu	Yes	Yes	Yes	
No	No	No	2	Năm thứ 2	No	No	No	
	Unknown		3	Năm thứ 3				
			...	Năm thứ 4				
			61	Năm thứ 5				
				Năm thứ 6				

2	4	19	5	2	2	15	6	3
Credit Card	Khu vực 1	Khu vực 2	Tình trạng hôn nhân	Đơn vị cư trú kiểu	Tình trạng cư trú	Thời gian cư trú	Số người lớn	Số thẻ hệ
Yes	Miền Đông	NORTHWEST/ROCKY MOUNTAIN	Ly thân	Single	Thuê nhà	1	1	1
No	Miền Bắc	HOUSTON	Ly hôn	Married	Chính chu	2	2	2
	Miền Nam	CENTRAL/SOUTH TEXAS	Kết hôn			3	3	3
	Miền Tây	MIDWEST	Độc thân			4	4	
		SOUTHWEST	Không kết hôn			5	5	
		DALLAS				6	6	
		CALIFORNIA NORTH				7		
		CHICAGO				8		
		NEW ENGLAND				9		
		GREAT LAKES				10		
		OHIO				11		
		PHILADELPHIA				12		
		NEW YORK CITY				13		
		SOUTH FLORIDA				14		
		ATLANTIC SOUTH				15		
		DC/MARYLAND/VIRGINIA						
		NORTH FLORIDA						
		TENNESSEE						
		LOS ANGELES						

Phần II: Phân tích và xử lý dữ liệu ETL

2.3 Tiền xử lý dữ liệu ETL

Cây phân tích dash board



Phần II: Phân tích và xử lý dữ liệu ETL

2.4. Data models

Sơ đồ mối quan hệ dữ liệu ERD



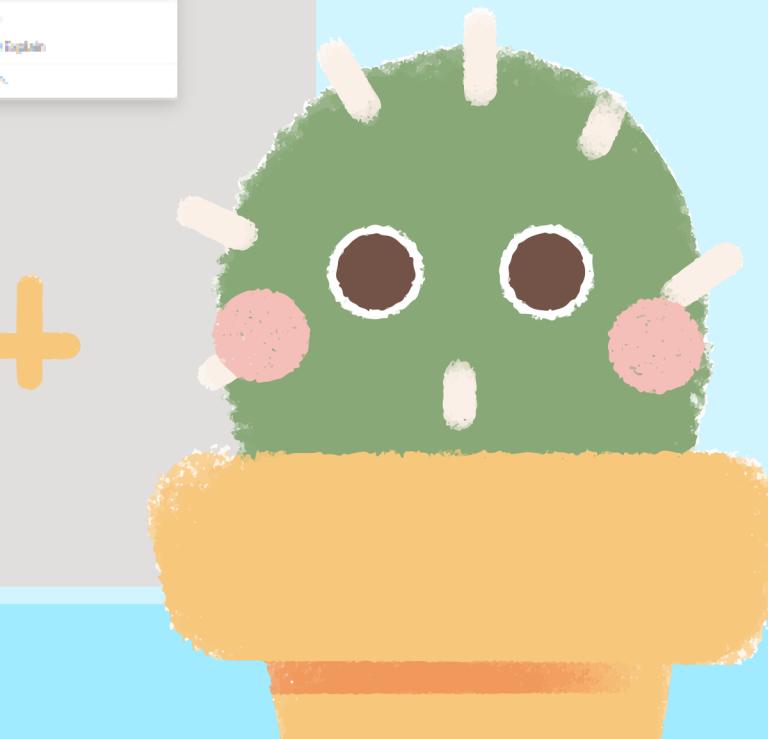
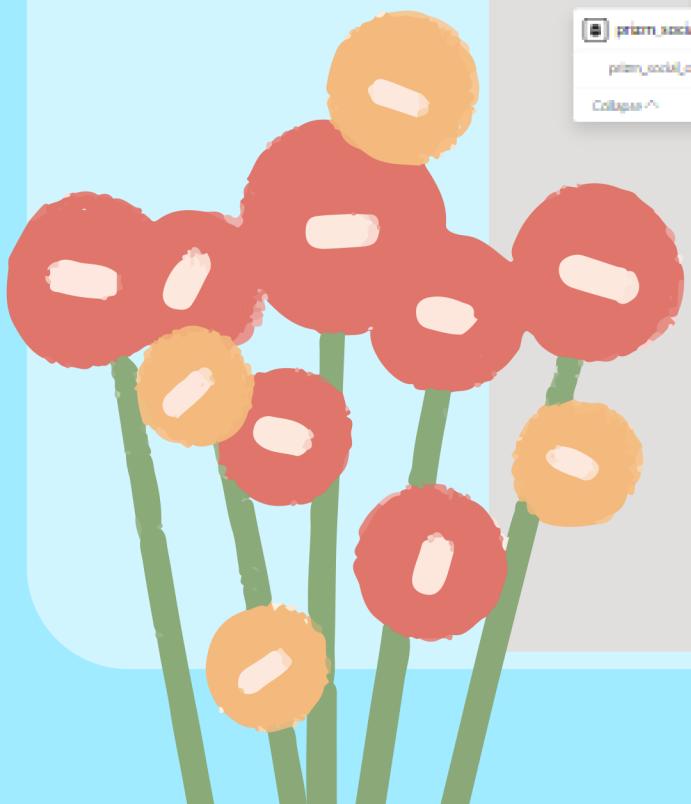
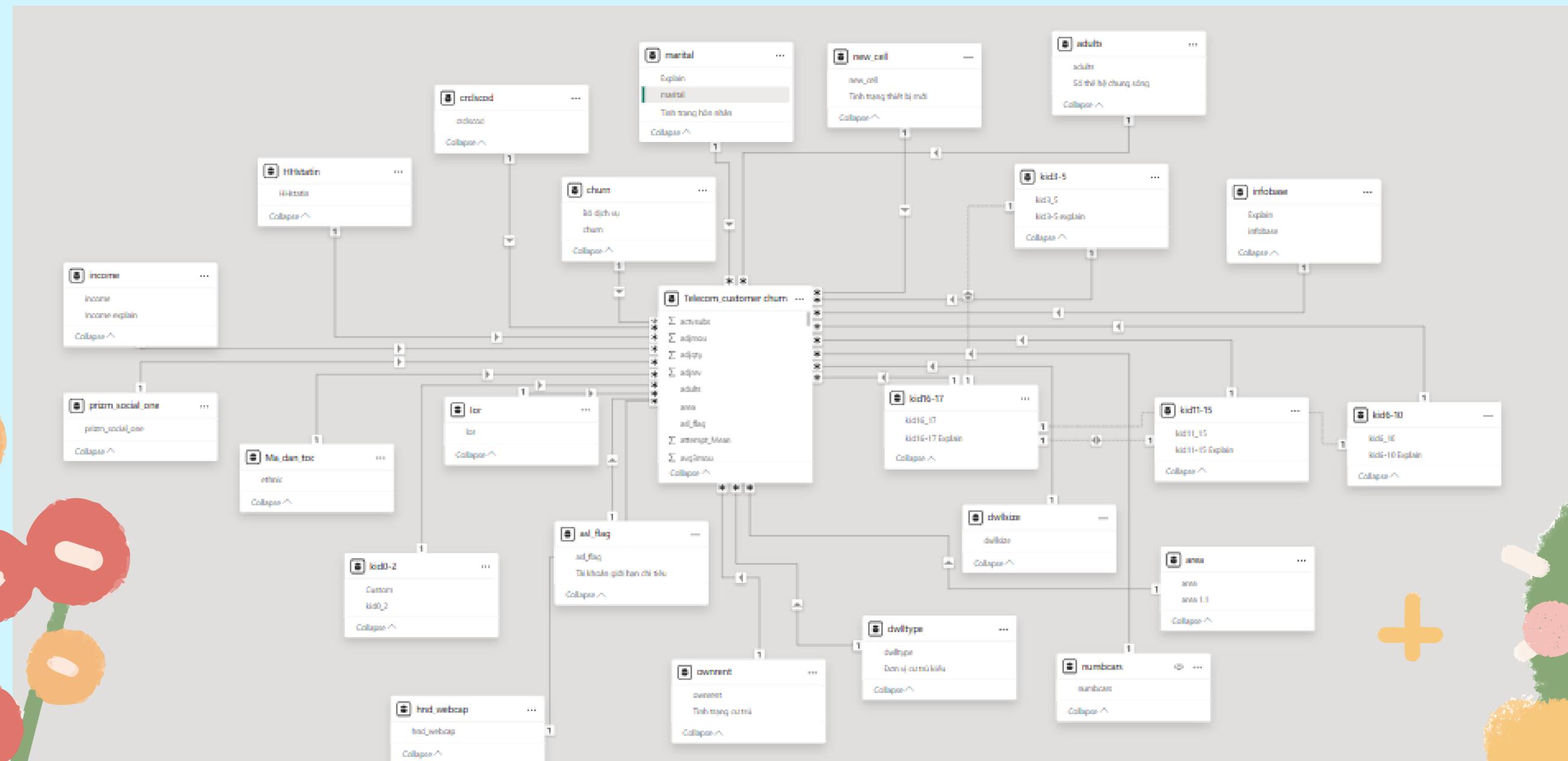
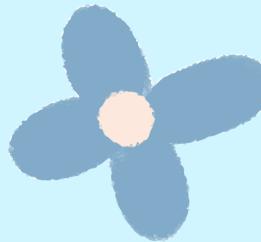


Phần II: Phân tích và xử lý dữ liệu ETL



2.4 Data models

Data models OLAP

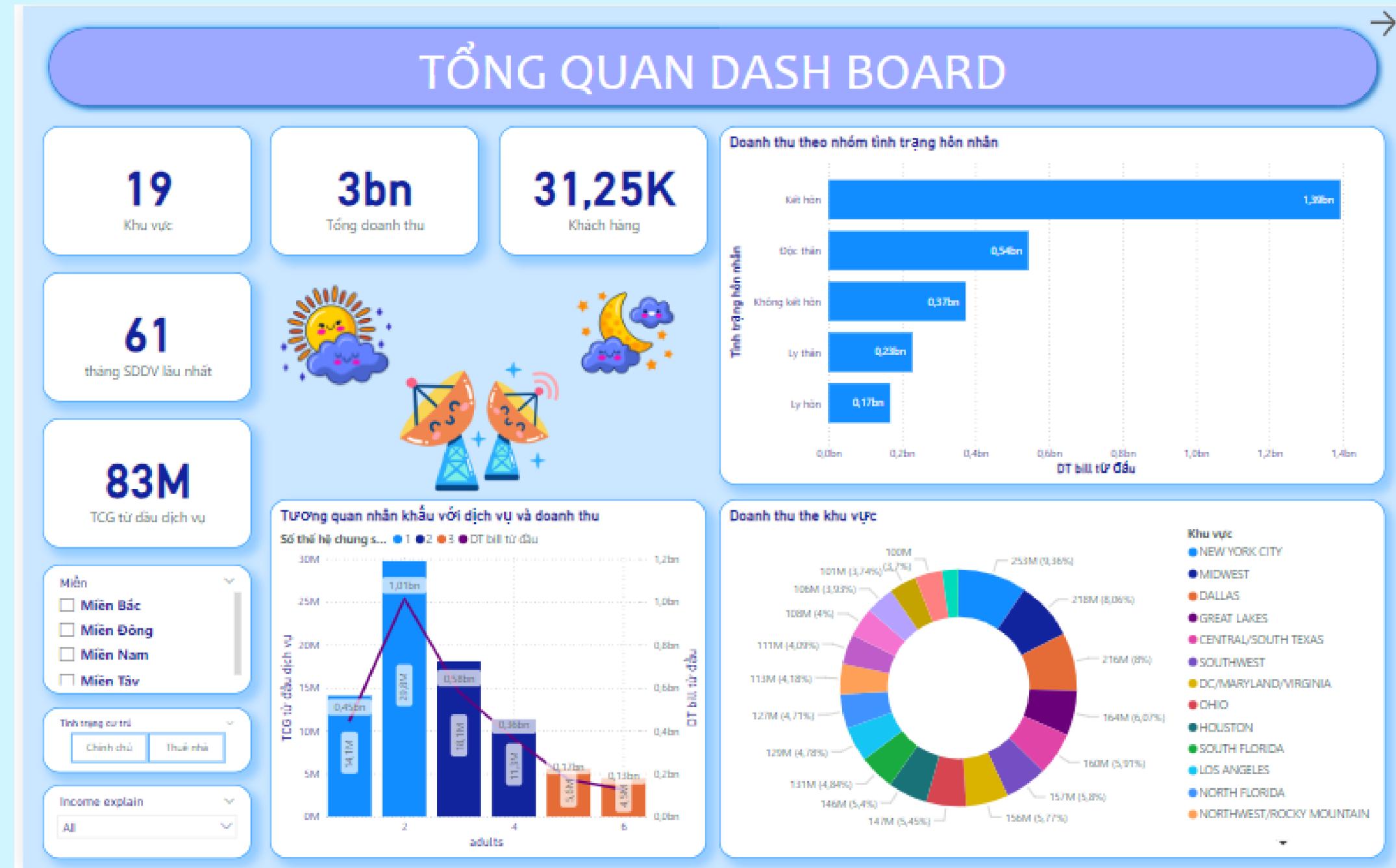


Phần III: Xây dựng báo cáo



Phần III: Xây dựng báo cáo

3.1 Dash 1: Tổng quan



Phần III: Xây dựng báo cáo

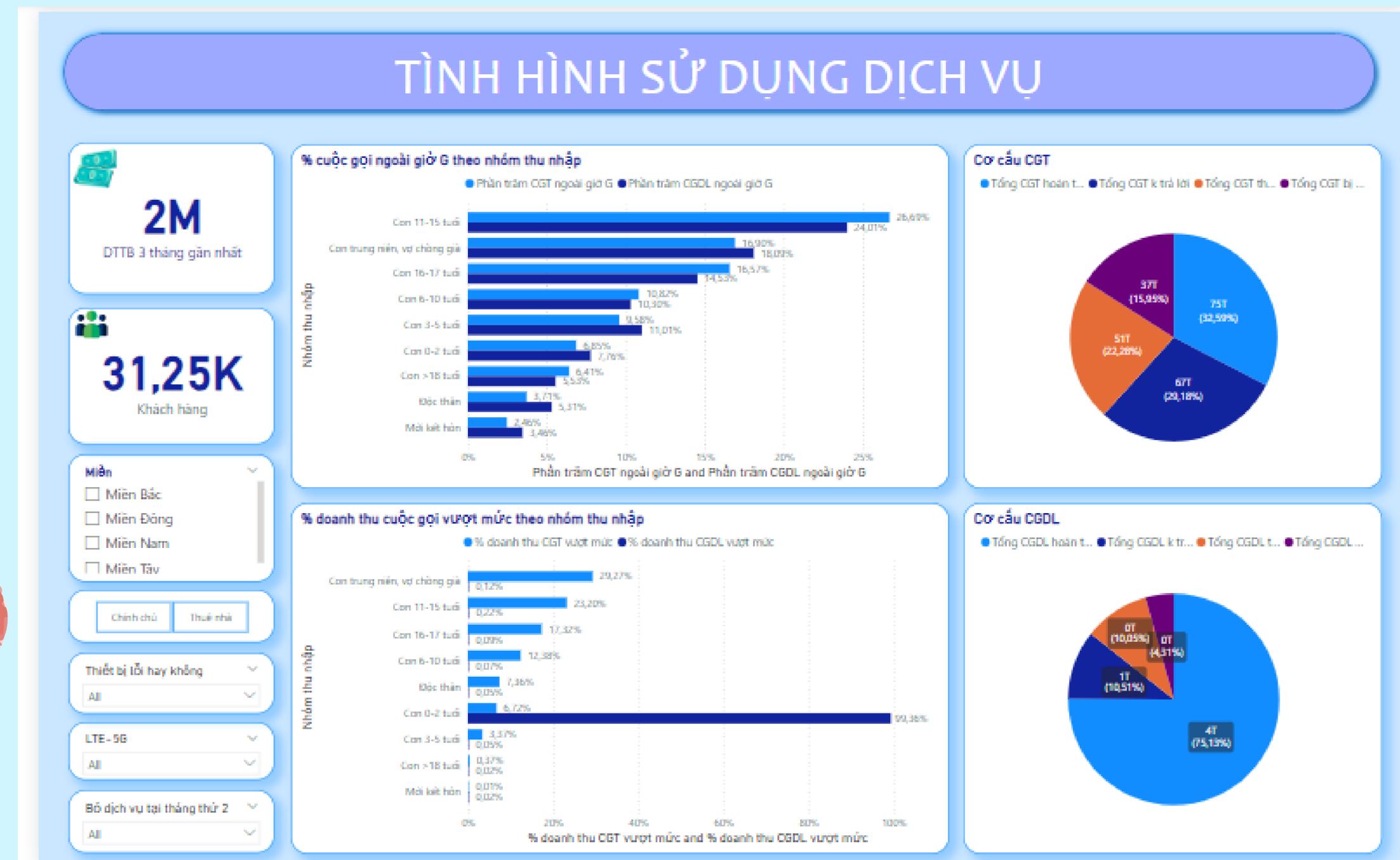
3.2. Dash 2: Đặc điểm khách hàng



Phần III: Xây dựng báo cáo

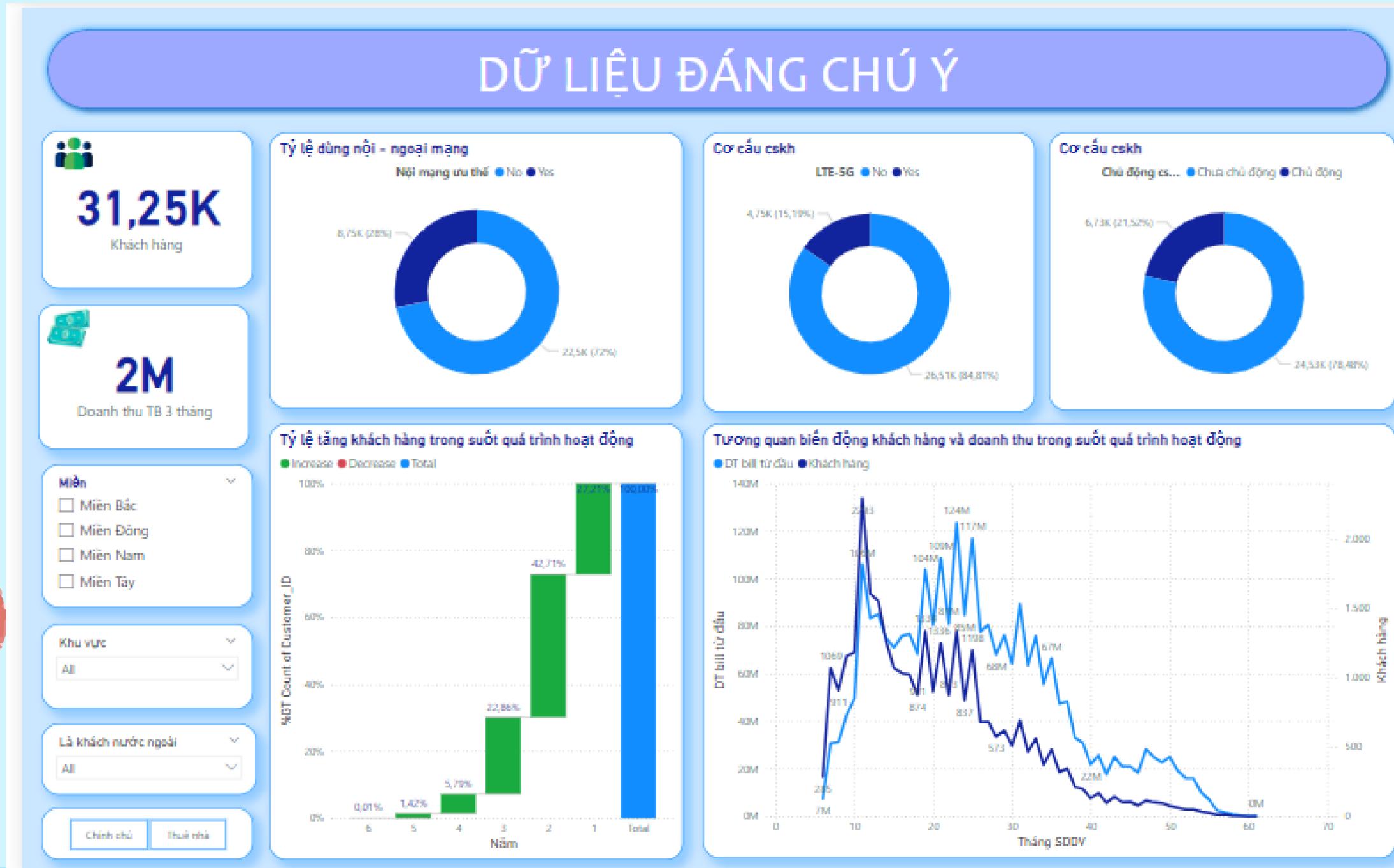
3.3. Dash 3: Tình hình sử dụng dịch vụ

TÌNH HÌNH SỬ DỤNG DỊCH VỤ



Phần III: Xây dựng báo cáo

3.4. Dash 4: DL đáng chú ý:



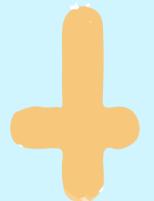
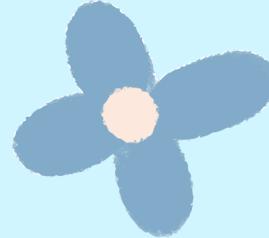
Phân IV: Rút ra bài học

1. Làm việc cá nhân
2. Làm việc đội nhóm





Phần IV: Rút ra bài học



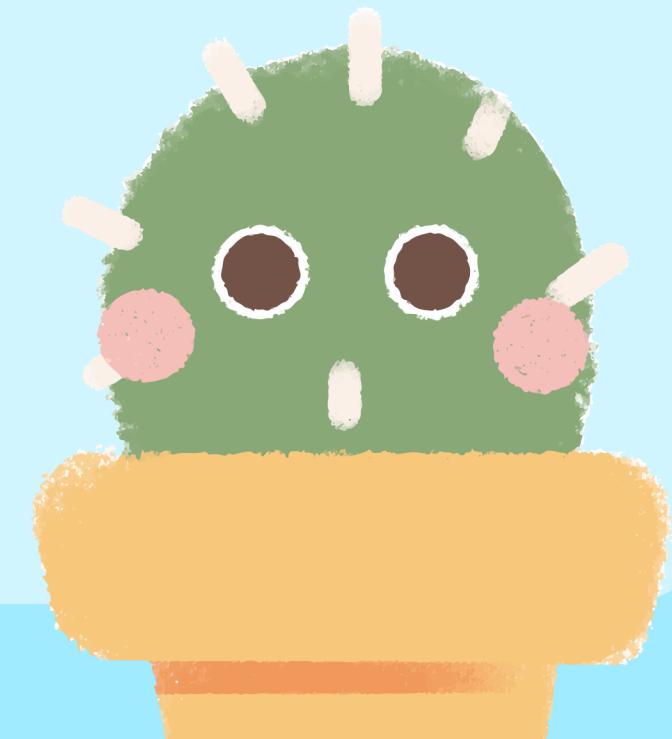
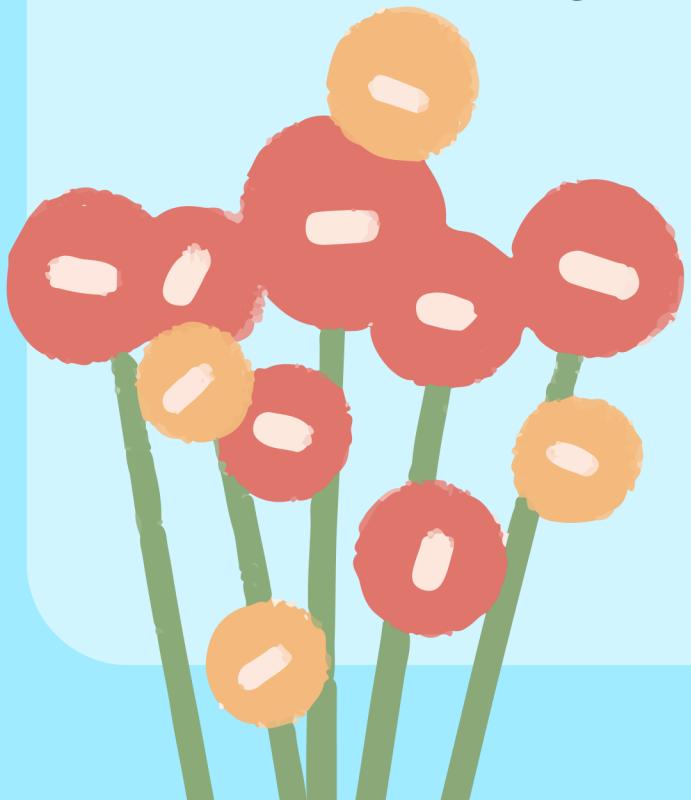
Hiểu hơn về các kiến thức về tiền xử lý dữ liệu, xây dựng Data Warehouse

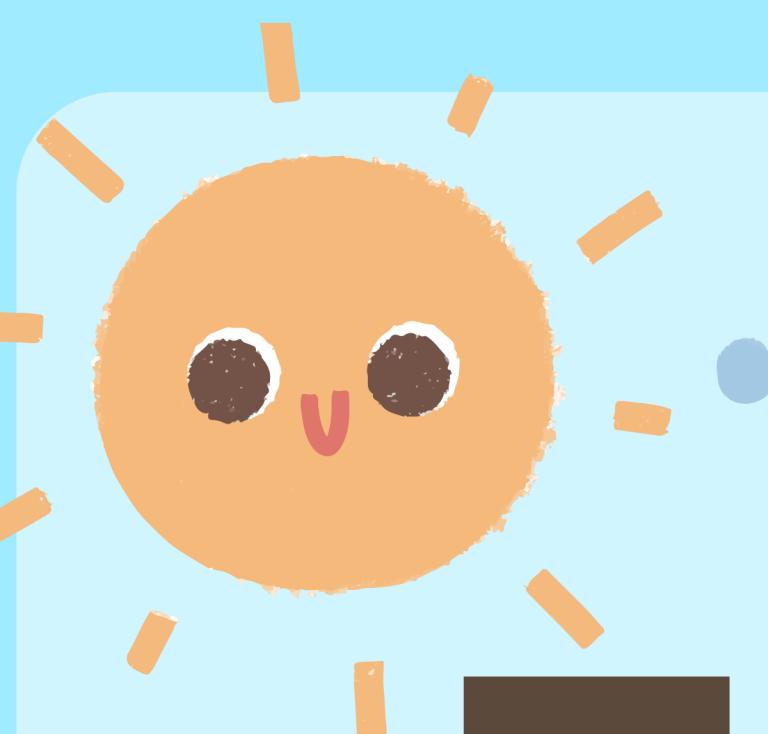
Những bước đầu có thể xử lý và phân tích dữ liệu

Hiểu hơn về cách thức sử dụng các công cụ để phân tích dữ liệu như Power Query, Power BI

Xây dựng cơ bản được hệ thống Dashboard hoàn chỉnh phục vụ cho việc phân tích dữ liệu

Tăng khả năng tư duy về dữ liệu và khai phá đưa ra dự báo cho công việc





Thank you!



