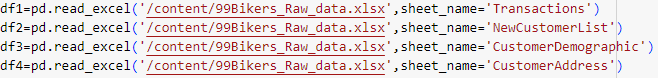
Tiền xử lý dữ liệu

Công cụ xử lý dữ liệu

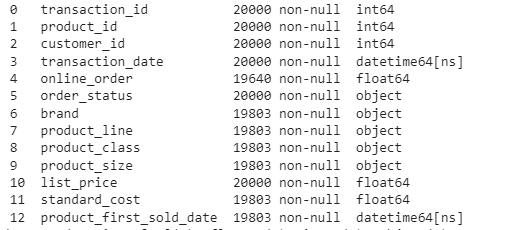




Import file dữ liệu gốc từ excel vào colab, ta sẽ có được 4 file, Transactions, New\_Customer\_List, Customer\_Demographic, Customer\_Adress:



Thông qua bảng Transactions, ta sẽ lấy bảng này làm main table để left join đến những bảng còn lại, đầu tiên xem qua info về các cột chứa giá trị null trong bảng Transactions:



Drop giá trị null trong cột brand vì không có brand đồng nghĩa các transaction\_id cũng sẽ không có product\_line,product\_class,product\_size,standard\_cost và product\_first\_sold\_date.



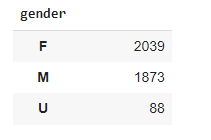
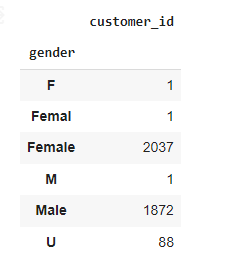
Tiếp đến là bảng Customer\_Demographic, thực hiện drop cột default vì cũng không có 1 thông tin gì phù hợp cho việc phân tích



Thực hiện nói 2 column first\_name,last\_name để tạo ra cột full\_name, Định dạng lại giá trị trong cột gender.



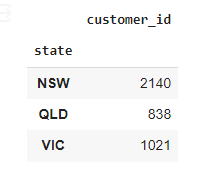
Kết quả:



Tiếp đến là bảng Customer\_Adress, định dạng lại giá trị trong column State:



Kết quả:



Thực hiện left join giữa bảng Customer\_Demographic và bảng Customer\_Adress nối bằng key là customer\_id và left join 1 lần nữa với bảng Transactions thông qua key cũng là customer\_id để có được cleaning table.



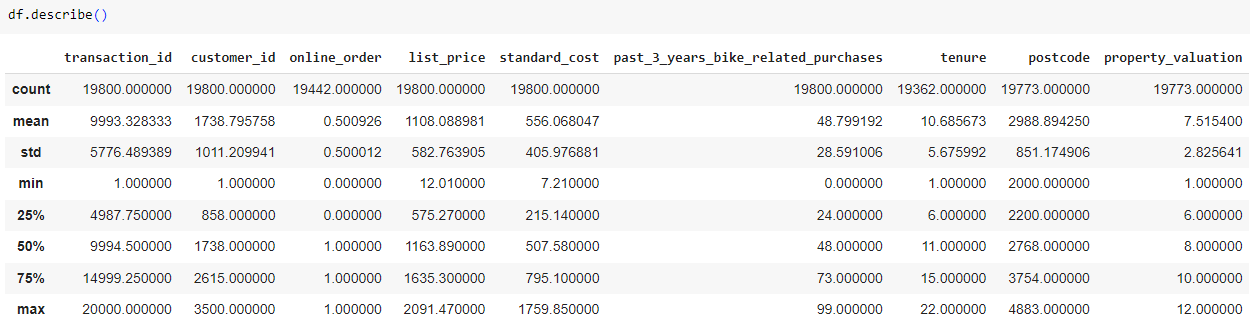


Sau khi có được cleaning table, xuất ra csv và thấy rằng còn rất nhiều giá trị null trong từng cột, do đó import 1 lần nữa vào colab để drop đi các giá trị null và đặt thresh=13 để có thể hiểu rằng phải có trên 13 cột không chứa giá trị null thì dữ liệu mới được accept và không được drop đi.



Phần 2: Phân tích dữ liệu

Thống kê mô tả:

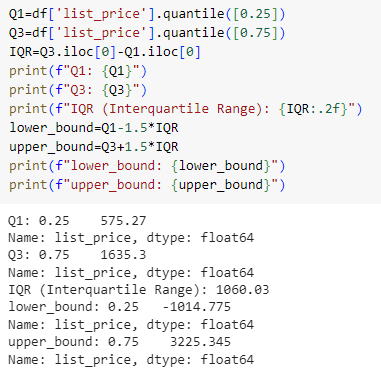


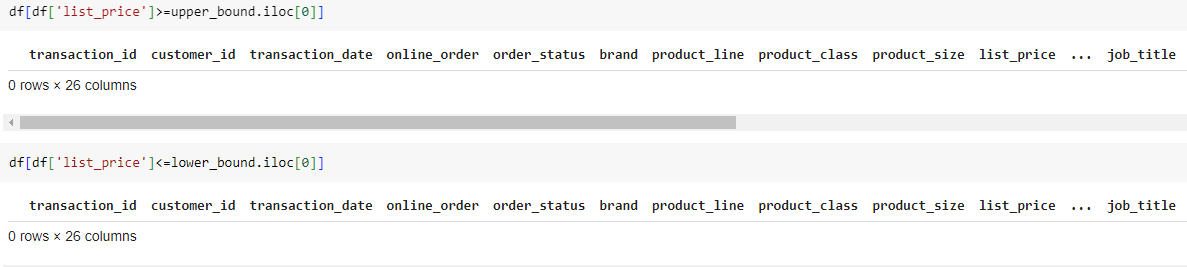
Kiểm tra outliers

Q3-Q1= IQR(Độ trải giữa)

UL (Giới hạn trên) = Q3 + IQR \* 1.5, LL (Giới hạn dưới) = Q1 – IQL \*1.5.

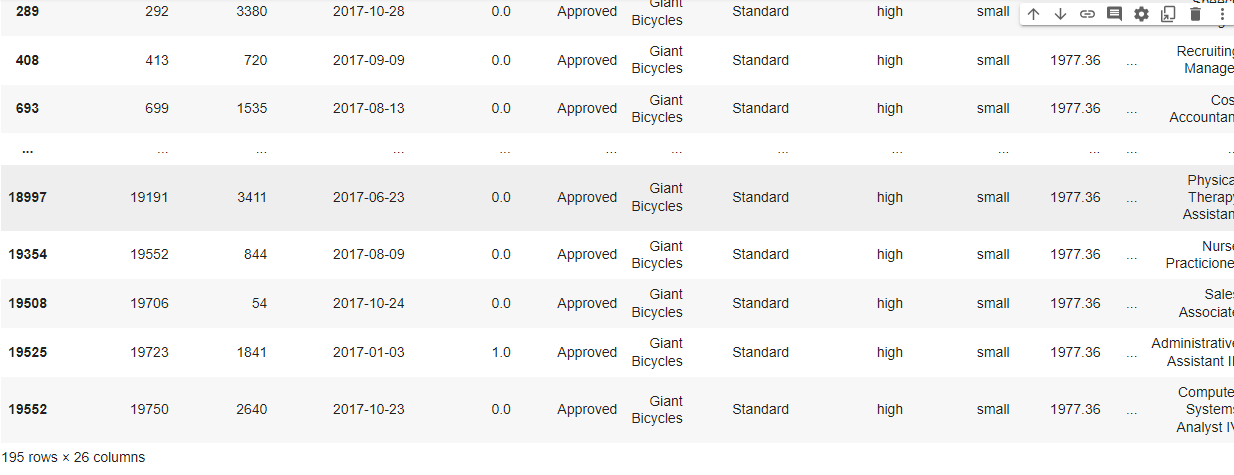
Ví dụ với lower bound và upper bound cho column list\_price:





Dựa vào công thức trên, ta thấy các cột:

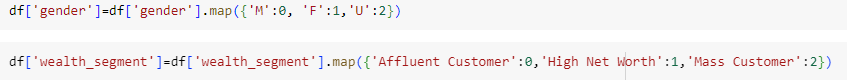
* List\_price có UL = 3225.345, LL = -1014.775 so với min,max thì không có giá trị ngoại lai
* Profit có UL = 1874.93, LL = -910.91 so với min,max thì không có giá trị ngoại lai, lợi nhuận công ty nhận được từ kết quả hoạt động kinh doanh ở mức vừa phải, không tăng đột biến và cũng không có sự giảm mạnh
* Standard cost có UL = 1665.04, LL = -654.8 so với min,max thì so với cận trên thì có giá trị ngoại lai lên đến 195 hàng, điều này cho thấy có rất nhiều sản phẩm có chi phí biến động khá cao, về giá trị cận dưới thì không có giá trị ngoại lai

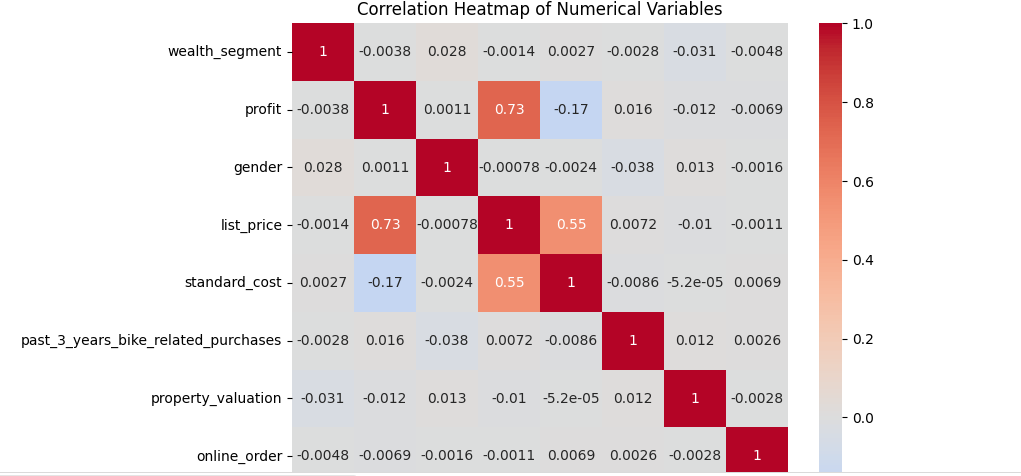


* Past\_3\_years\_related\_purchases có UL = 146.5, LL = -49.5 so với min,max thì không có giá trị ngoại lai, các sản phẩm so về số lượng bán hàng ở quá khứ thì cũng không có nhiều đơn hàng biến động tăng giảm về mặt số lượng
* Tenure có UL = 28.5, LL = -7.5 so với min,max thì không có giá trị ngoại lai, sự gắn bó mua hàng lâu dài với công ty ở mức độ vừa phải, không có khách hàng nào gắn bó hoặc rời bỏ quá lâu.
* Property valuation có 16 = , LL = 0 so với min,max thì không có giá trị ngoại lai.

Phân tích tương quan:

Map thêm giá trị 0,1,2 cho 2 column gender và wealth\_segment:



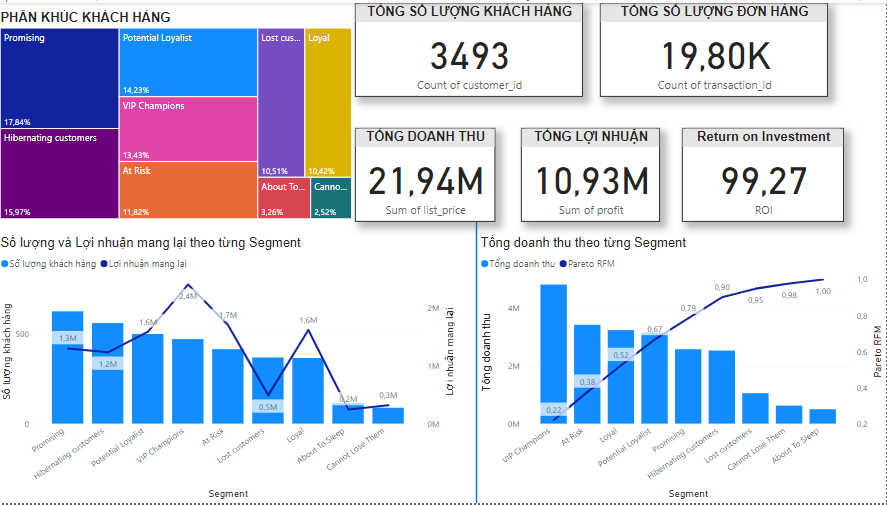


Nhận xét:

* List\_price có tác động dương khá mạnh đến profit 0.73
* List\_price có tác động dương yếu đến standard\_cost 0.55
* Standard\_cost có tác động âm kém đến profit -0.17

Phân tích RFM

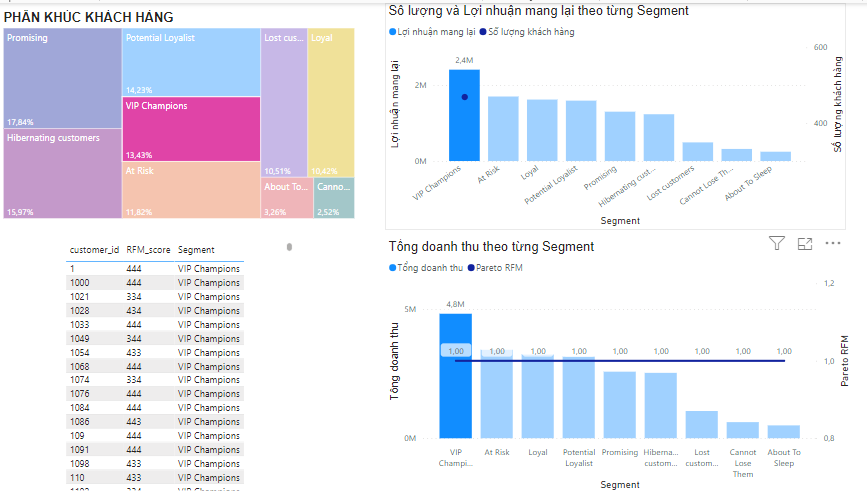
Tổng quát:



Ta thấy phân khúc của các từng customer segment phân bố đồng đều từ 10 %trở lên, phần lớn thuộc về nhóm Promising là 17,84% và Hibernating Customer 15,97%, không có biến động quá lớn về 1 bên, chỉ riêng 2 nhóm khách hàng About to Sleep và Cannot Loose Them chiếm số lượng rất ít dưới 3%.

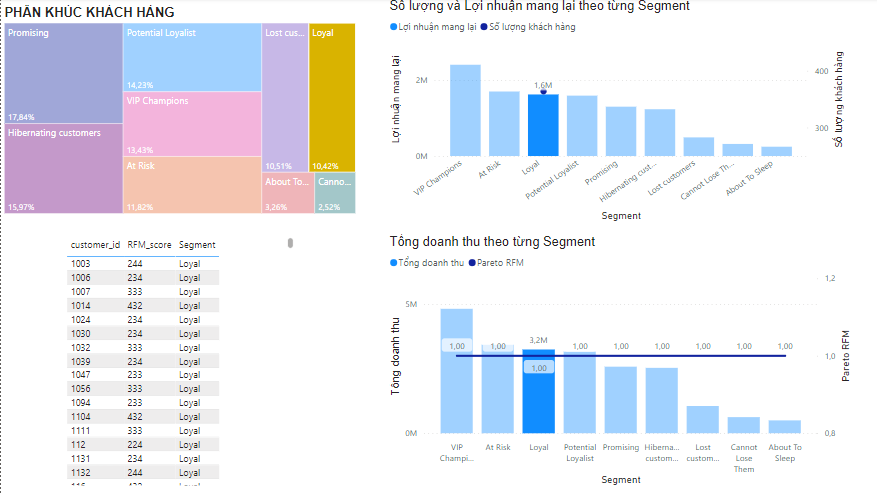
+ Vip Champion: nhóm Khách Hàng tốt nhất, chiếm số lượng 13,84% trong tổng số KH, doanh thu đóng góp 22%, lợi nhuận mang lại cao nhất đến 2,4 triệu đô. Họ đã mua gần đây, đặt hàng thường xuyên và chi tiêu nhiều nhất.

=> Cần chăm sóc nhóm KH này để tiếp tục duy trì ở phân khúc hiện tại nhằm mang lại doanh thu cho tương lai cao hơn, mặc dù đứng thứ 4 trong bảng xếp hạng về số lượng khách hàng nhưng lại mang tỉ lệ đóng góp lại cao nhất trong doanh thu, lợi nhuận mang lại cao nhất cho công ty, cần dự báo số lượng sản phẩm nhóm KH này hay tiêu thụ để stock kho, hàng hoá phải sẵn sàng để đáp ứng nhu cầu KH. Khi có sản phẩm mới, phải giới thiệu họ trước tiên, họ có thể là những người có nhu cầu được giới thiệu sản phẩm và tư vấn từ phía marketing rất tốt.



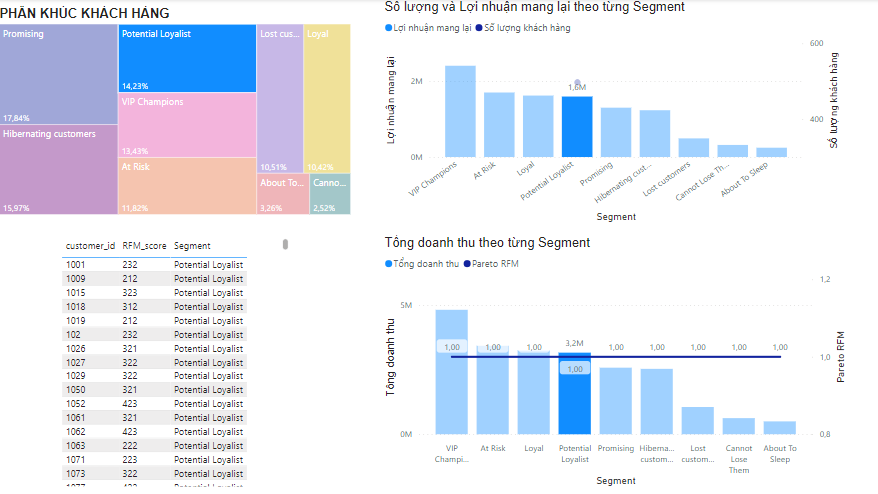
+Loyal: nhóm KH tốt thứ 2,mặc dù số lượng chiếm cũng ở mức khá (10,42%) ,nhưng doanh thu đóng góp lại đứng thứ 3 (14%), lợi nhuận mang lại cũng ở mức rất cao(1,6 triệu đô)

=>Họ đặt hàng thường xuyên có đóng góp 1 phần lớn trong doanh thu và có tương tác với công ty thường xuyên. Đưa ra những chương trình khuyến mãi phù hợp, tăng cường tương tác tạo mối quan hệ chặt chẽ với họ trong việc mua bán hàng để nâng cấp nhóm KH này lên champion.



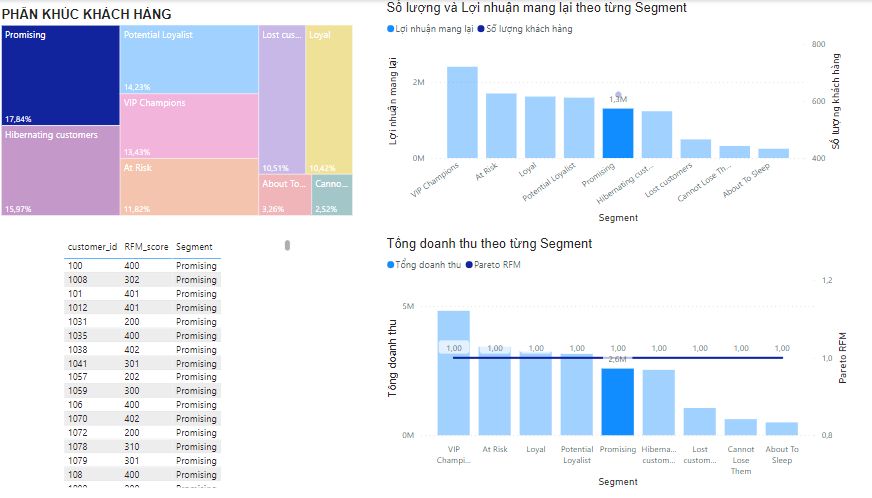
+ Potential Loyalist: nhóm KH tốt thứ 3, chiếm số lượng thứ 3 (14,23%), doanh thu đóng góp 15%, mặc dù chiếm số lượng nhiều hơn so với nhóm khách hàng loyal nhưng lợi nhuận của nhóm này đóng góp cũng ngang bằng với loyal, (1,6 triệu đô)

=> Họ đã mua hàng gần đây, số lượng hàng họ mua ở mức cao nhưng mức chi tiêu ở trung bình => tương tự nhóm Loyal, song với việc nhóm khách hàng này chiếm số lượng đông, cần đưa ra những chương trình khuyến mãi đã được cá nhân hoá cho phù hợp với mức tiêu dùng của họ nhằm tối ưu hóa đơn hàng để họ mua được nhiều đơn hàng và mang lại doanh thu nhiều hơn cho công ty, nhắm tới việc nâng cấp nhóm KH này lên loyal hoặc champion.



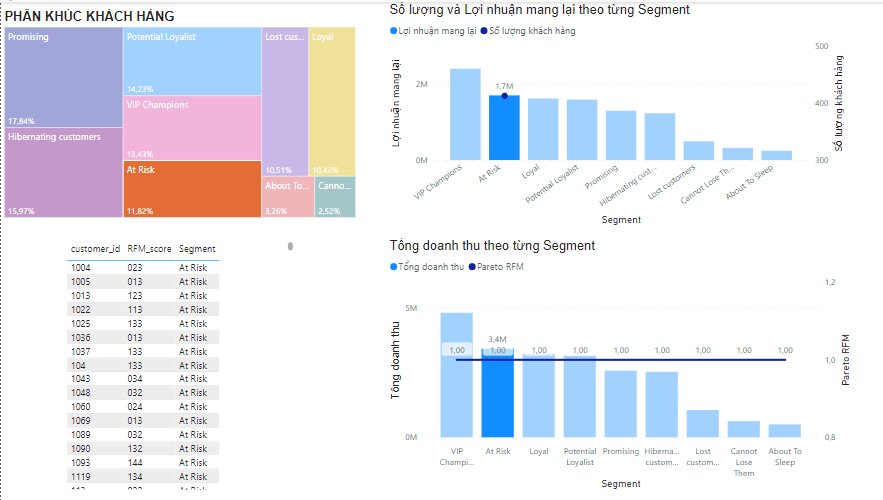
+ Promising: Đây là nhóm chiếm tỷ trọng nhiều nhất trong tổng số lượng KH (17,84%) nhưng doanh thu và lợi nhuận mang lại cũng chỉ đạt ở mức trung bình(đóng góp 12% doanh thu và lợi nhuận mang lại là 1,3 triệu đô). Họ là những mới mua hàng gần đây, vì mới sử dụng sản phẩm của công ty nên họ thường sẽ chi tiêu khá ít tiền và mua với số lượng rất ít để đánh giá xem sản phẩm của công ty có phù hợp với nhu cầu của họ hay không. Mặc dù là nhóm khách hàng mới mua hàng gần đây, nhưng lợi nhuận mà họ đóng góp cho công ty cũng ở mức khá tốt.

=> cung cấp phiếu giảm giá, đưa ra chương trình khuyến mãi đã được cá nhân hoá để giữ tương tác với họ, quản lý và phát triển sản phẩm thật tốt, đưa ra những sản phẩm mới phù hợp với nhu cầu sử dụng của phân khúc khách hàng này nhằm kích thích nhu cầu mua hàng của họ trong tương lai gần. Phải giữ mối quan hệ thật tốt với nhóm khách hàng này.



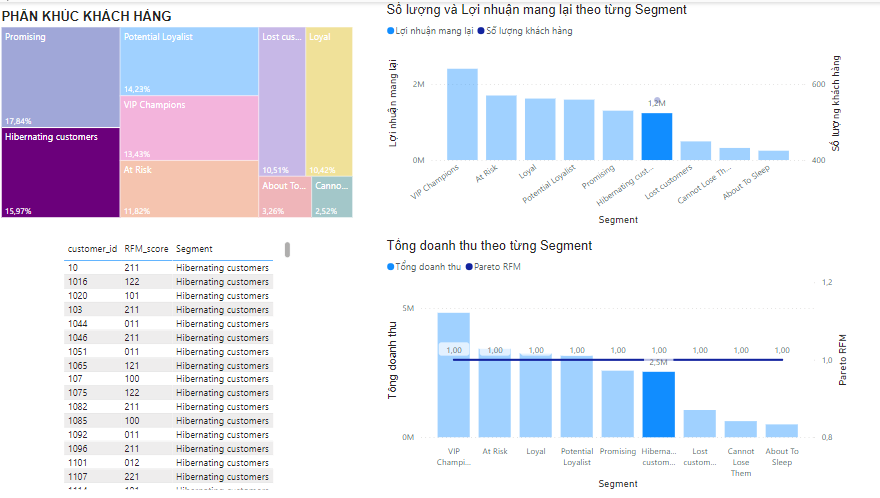
+ At Risk: Nhóm khách hàng chiếm tỉ trọng đứng thứ 5 về mặt số lượng(11,82%), doanh thu chỉ nằm sau nhóm khách hàng Vip Champions và lợi nhuận cho công ty nằm trong top 3 mức cao nhất. Họ là những khách hàng đã mua hàng của công ty từ rất lâu nhưng mỗi lần mà họ place 1 order thì sẽ là những đơn hàng với số lượng lớn là giá trị rất cao.

=> Đây là những khách hàng cốt lõi giúp công ty mang lại doanh thu cao hơn nhưng chỉ có 1 vấn đề là thời gian đặt hàng của họ có thể gây ra vấn đề tồn kho khá lâu cho cho công ty => đề xuất khuyến mãi và dự báo thời gian back-order của họ, đồng thời dự trữ số lượng hàng trong kho cho hợp lý với số lượng order của họ cho các lần order kế tiếp.



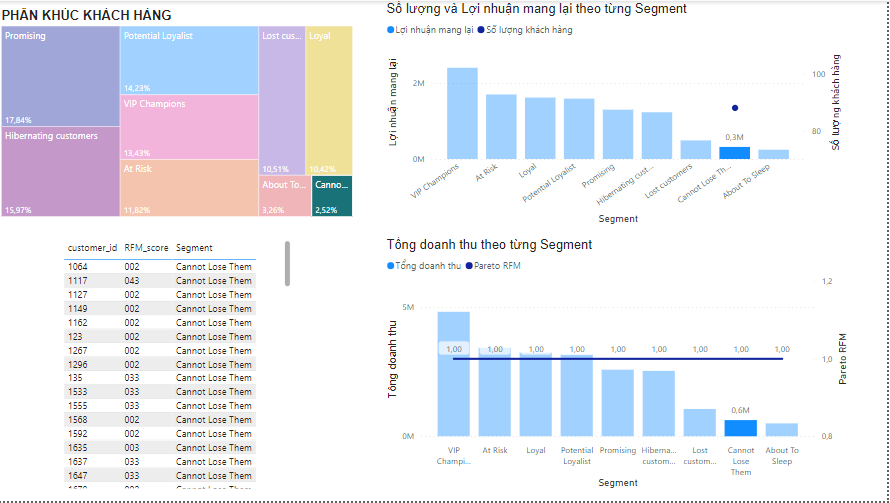
+ Hibernating Customers: Nhóm khách hàng ngủ đông chiếm số lượng đứng thứ 2(15,97%). Số lượng khách hàng này chiếm số lượng cũng thuộc đông nhất song với việc đó thì phần trăm doanh thu đóng góp (11%) và lợi nhuận mang lại cũng ở mức trung bình tương tự với nhóm Promising(1,2 triệu đô).

=>Đây là nhóm khách hàng đã mua hàng của công ty từ rất lâu và thông thường mỗi lần đặt hàng lại của họ đi kèm với số lượng rất ít và đơn hàng không được cao, có thể coi rằng những khách hàng này họ có thể đi mua hàng từ những công ty khác và khi công ty đó không còn đủ số lượng thì họ mới quay lại đặt hàng từ công ty mình. Cần phải kích nhu cầu tiêu dùng của họ và liên lạc để hỏi xem rằng họ có hứng thú với dịch vụ và sản phẩm bên mình hay không, ngoài ra cần đưa ra các chính sách phát triển về dịch vụ để họ cảm thấy yên tâm và vui vẻ cho những lần mua hàng đợt sau.



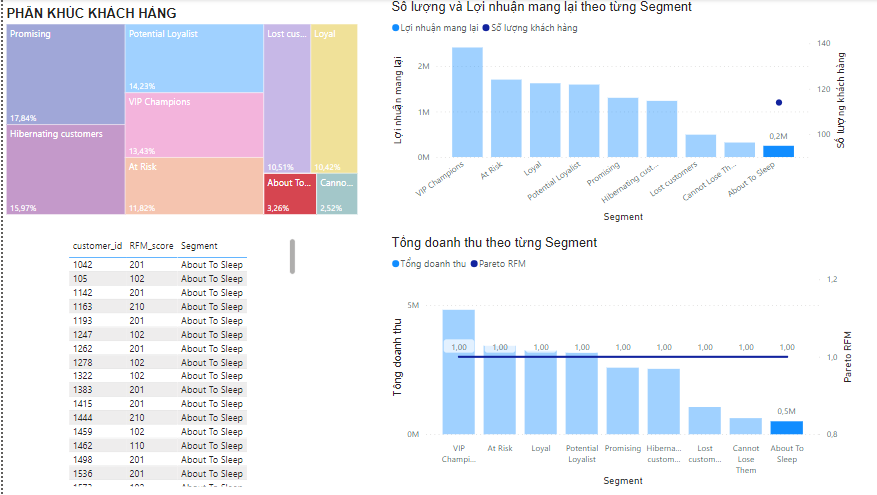
+ Can’t lose them: nhóm này chiếm số lượng ít nhất chỉ với 2,52% số KH, doanh thu đóng góp nằm trong nhóm thấp nhất 2%, lợi nhuận mang lại thì chỉ vỏn vọn 0,3 triệu đô. Họ đã từng thực hiện các đơn đặt hàng lớn nhất và thường xuyên nhưng đã lâu không trở lại.

=> Bộ phận sales/ chăm sóc KH cần liên hệ tìm hiểu nguyên nhân tại sao khách hàng không quay lại order đơn hàng từ mình để ra chương trình marketing cho phù hợp, ngoài ra đối với mỗi khi nhóm khách hàng này quay lại order, cần phải chuẩn bị hàng sẵn và đảm bảo hàng phải có chất lượng tốt nhất.



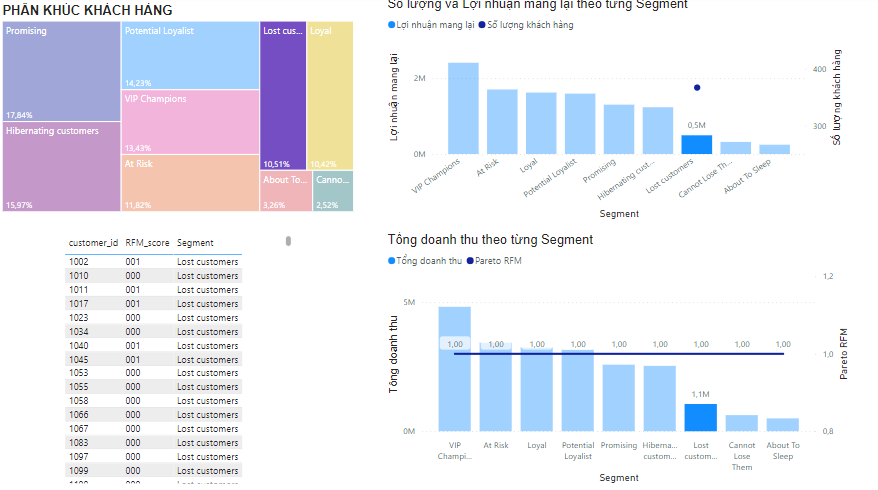
+ About to sleep: chiếm 3,26% tổng số KH, tưng tự với nhóm cannot loose them có số lượng khách hàng thấp nhất , doanh thu đóng góp chỉ ở mức kém 2%. Họ là những khách hàng phổ biến, mua hàng cách đây cũng không quá lâu nhưng mỗi lần order lại hầu như là những đơn hàng không có giá trị để góp nhiều vào trong doanh thu của công ty. Lợi nhuận mà nhóm khách hàng này mang lại ở mức thấp nhất chỉ có 0,2 triệu đô

=> Tương tự nhóm Need Attention, đưa ra chương trình khuyến mãi có thời hạn phù hợp với những đơn hàng có giá trị thấp mà nhóm khách hàng này hướng tới mua nhằm đẩy nhanh quá trình sales hàng và tăng thêm doanh thu.



Lost Customer: nhóm này cũng với số lượng khá cao là 10,51% tổng số lượng KH, doanh thu đóng góp 5%. Đối với nhóm khách hàng này, họ hầu như đã mất liên lạc mua hàng với công ty bởi thời gian quay lại order của họ hoàn toàn rất xa và order số lượng rất ít, các đơn hàng ít giá trị.

=> Bộ phận sales/chăm sóc KH cần liên hệ tìm hiểu nguyên nhân tại sao lại có 1 lượng khách hàng khá lớn bỏ đi và cần đề xuất ra các chương trình marketing cho phù hợp nhằm kéo nhóm khách hàng quay trở lại.



**THUYẾT TRÌNH TỪ KHÚC NÀY TRỞ ĐI.**

Ta thấy phân khúc của các từng customer segment phân bố đồng đều từ 10 %trở lên, phần lớn thuộc về nhóm Promising là 17,84% và Hibernating Customer 15,97%, không có biến động quá lớn về 1 bên, chỉ riêng 2 nhóm khách hàng About to Sleep và Cannot Loose Them chiếm số lượng rất ít dưới 3%.

+ Vip Champion: nhóm Khách Hàng tốt nhất, chiếm số lượng 13,84% trong tổng số KH, doanh thu đóng góp 22%, lợi nhuận mang lại cao nhất đến 2,4 triệu đô. Họ đã mua gần đây, đặt hàng thường xuyên và chi tiêu nhiều nhất.

Loyal: nhóm KH tốt thứ 2,mặc dù số lượng chiếm cũng ở mức khá (10,42%) ,nhưng doanh thu đóng góp lại đứng thứ 3 (14%), lợi nhuận mang lại cũng ở mức rất cao(1,6 triệu đô)

Potential Loyalist: nhóm KH tốt thứ 3, chiếm số lượng thứ 3 (14,23%), doanh thu đóng góp 15%, mặc dù chiếm số lượng nhiều hơn so với nhóm khách hàng loyal nhưng lợi nhuận của nhóm này đóng góp cũng ngang bằng với loyal, (1,6 triệu đô)

Promising: Đây là nhóm chiếm tỷ trọng nhiều nhất trong tổng số lượng KH (17,84%) nhưng doanh thu và lợi nhuận mang lại cũng chỉ đạt ở mức trung bình(đóng góp 12% doanh thu và lợi nhuận mang lại là 1,3 triệu đô). Họ là những mới mua hàng gần đây, vì mới sử dụng sản phẩm của công ty nên họ thường sẽ chi tiêu khá ít tiền và mua với số lượng rất ít để đánh giá xem sản phẩm của công ty có phù hợp với nhu cầu của họ hay không. Mặc dù là nhóm khách hàng mới mua hàng gần đây, nhưng lợi nhuận mà họ đóng góp cho công ty cũng ở mức khá tốt.

At Risk: Nhóm khách hàng chiếm tỉ trọng đứng thứ 5 về mặt số lượng(11,82%), doanh thu chỉ nằm sau nhóm khách hàng Vip Champions và lợi nhuận cho công ty nằm trong top 3 mức cao nhất. Họ là những khách hàng đã mua hàng của công ty từ rất lâu nhưng mỗi lần mà họ place 1 order thì sẽ là những đơn hàng với số lượng lớn là giá trị rất cao.

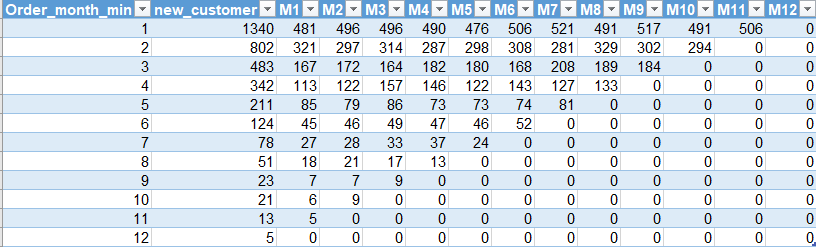
Hibernating Customers: Nhóm khách hàng ngủ đông chiếm số lượng đứng thứ 2(15,97%). Số lượng khách hàng này chiếm số lượng cũng thuộc đông nhất song với việc đó thì phần trăm doanh thu đóng góp (11%) và lợi nhuận mang lại cũng ở mức trung bình tương tự với nhóm Promising(1,2 triệu đô).

Can’t lose them: nhóm này chiếm số lượng ít nhất chỉ với 2,52% số KH, doanh thu đóng góp nằm trong nhóm thấp nhất 2%, lợi nhuận mang lại thì chỉ vỏn vọn 0,3 triệu đô. Họ đã từng thực hiện các đơn đặt hàng lớn nhất và thường xuyên nhưng đã lâu không trở lại.

About to sleep: chiếm 3,26% tổng số KH, tưng tự với nhóm cannot loose them có số lượng khách hàng thấp nhất , doanh thu đóng góp chỉ ở mức kém 2%. Họ là những khách hàng phổ biến, mua hàng cách đây cũng không quá lâu nhưng mỗi lần order lại hầu như là những đơn hàng không có giá trị để góp nhiều vào trong doanh thu của công ty. Lợi nhuận mà nhóm khách hàng này mang lại ở mức thấp nhất chỉ có 0,2 triệu đô

Lost Customer: nhóm này cũng với số lượng khá cao là 10,51% tổng số lượng KH, doanh thu đóng góp 5%. Đối với nhóm khách hàng này, họ hầu như đã mất liên lạc mua hàng với công ty bởi thời gian quay lại order của họ hoàn toàn rất xa và order số lượng rất ít, các đơn hàng ít giá trị.

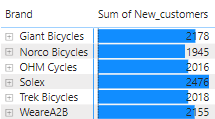
Phân tích Cohort



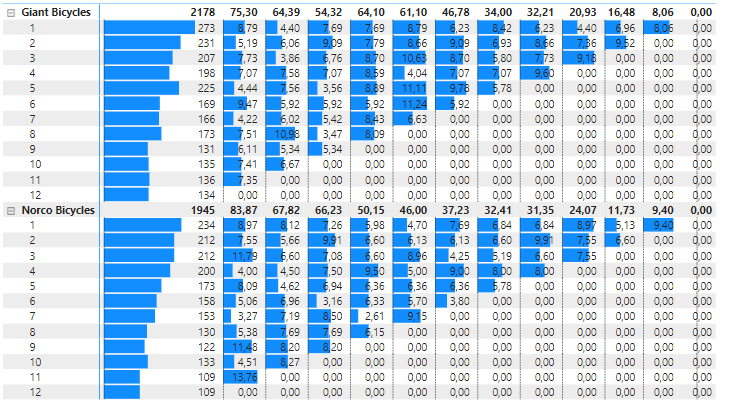
Dựa vào bảng Cohort qua theo từng Month, Số lượng khách hàng mới xuất hiện mua hàng rơi vào tháng 1 rất đông với là 1340 ,chiếm 38% so với các tháng còn lại trong năm 2017. Số lượng khách hàng cũ quay lại mua hàng qua từng tháng cũng ở mức đồng đều nhau và không có sự giảm mạnh, tỉ lệ quay lệ qua từng tháng tiếp theo cũng dao động ở mức khá cao trên 30%(lấy trung bình cộng của khách hàng quay lại sử dụng qua từng tháng tiếp theo chia cho new/customer của tháng đó)

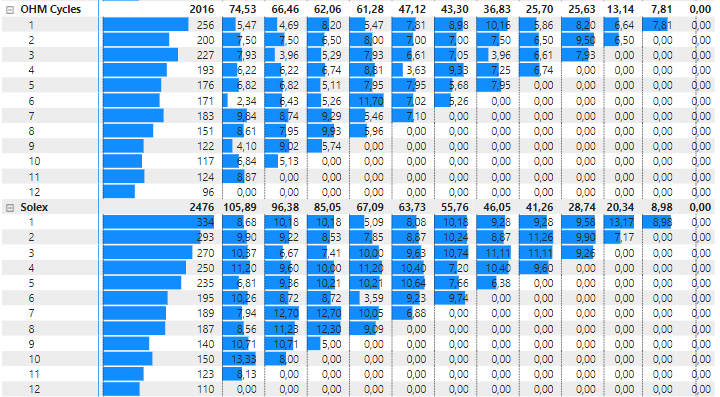
=>Điều này cho thấy công ty đã thành công trong việc triển khai các chính sách marketing và phát triển các mối quan hệ với khách hàng rất tốt. Mặc dù số lượng khách hàng mới có thể giảm theo từng tháng, nhưng số lượng khách hàng cũ quay lại mua hàng của họ sẽ bù đắp vào việc phát triển doanh thu 1 cách đáng kể. Chúng ta có thể thấy rằng tháng 6 có 124 khách hàng mới, nhưng qua 6 tháng tiếp theo, đã có tới 52 trong số 124 khách hàng đó quay trở lại mua hàng, tỉ lệ lên đến 40%.

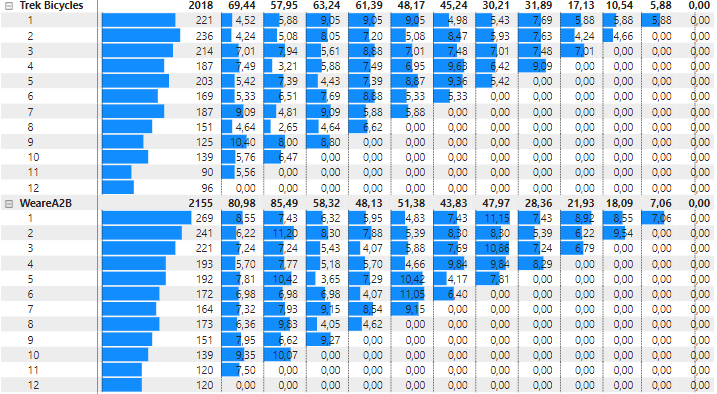
Cohort theo từng brand



Số lượng khách hàng mới cho từng hãng xe không quá lên lệch nhau, tỉ lệ khách hàng mới đồng đều cho các loại xe, cao nhất thuộc về hãng xe Solex => Điều này cho thấy sự phân bố khách hàng rộng rãi cho từng loại xe yêu thích của mình, không có khách hàng nào nghiêng về hẳn 1 brand.

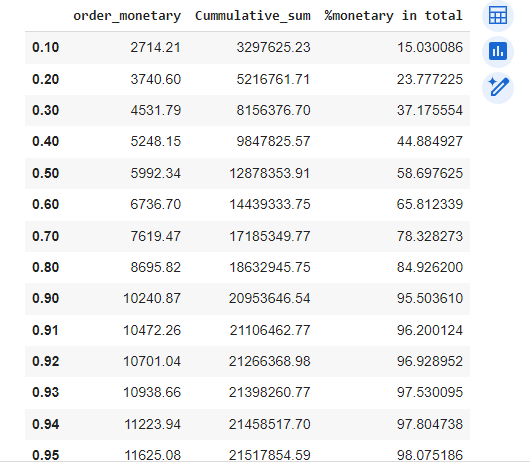






Dựa vào bảng Cohort\_Theo\_Từng\_Brand, ta thấy số lượng khách hàng mới đến mua các brand xe khác nhau đều ở mức ổn định, gia tăng mạnh vào các tháng 1,2,3,4,5 đều đạt các lượt khách hàng mới từ 170 khách hàng trở lên. Từ tháng 6 đến hết tháng 12, có thể do nhu cầu tiêu dùng xe không cao nên khách hàng giảm đi ít hơn chứ không quá đáng kể. Tỉ lệ quay trở lại mua hàng của khách hàng ở 6 hãng xe khách nhau vẫn ở mức thấp, dao động dưới 13% => Cần đưa ra các chính sách thu hút khách hàng,giảm giá xe hoặc tặng các item đi kèm khi mua xe,cải thiện marketing hoặc cần có các voucher mua hàng khi khách quay trở lại lần tiếp theo, nâng cao số lượng khách hàng đến mua hàng phù hợp với mỗi brand. Tuy số lượng khách hàng mới đến mua vẫn ở mức đồng đều cho thấy chính sách quảng cáo và nhu cầu tiêu dùng của khách hàng rất mạnh nhưng vẫn cần tăng cao tỉ lệ giữ chân của khách hàng nhằm không bị mất đi các khách hàng tiềm năng.

Phân tích thêm về tỉ lệ mang về lợi nhuận thông qua số lượng khách hàng, thông qua customer\_id và order\_monetary từ RFM table, ta tính được %monetary in total:



=>95% số lượng khách hàng đóng góp vào 98.1% doanh thu của công ty, chỉ có 5% khách hàng là đóng góp 1.9% còn lại. Điều này cho thấy số lượng khách hàng càng nhiều thì lợi nhuận càng tăng, số lượng các tệp khách hàng phân bố rất đồng đều, mỗi 1 tệp khách hàng đóng góp mỗi phần đồng đều cho lợi nhuận của công ty, không phải chỉ hướng về 1 tệp khách hàng vì có đến 95% số lượng người đóng góp vào lợi nhuận của công ty, phần lớn thuộc về nhóm Vip Champion như đã phân tích ở phần RFM. ROI nằm ở mức rất cao là 1 dấu hiệu tuyệt vời cho lợi nhuận của công ty (99,27%), có nghĩa là cứ mỗi 1 khách hàng place 1 order mặc dù chi phí bỏ ra là 100$ thì vẫn sẽ mang về lợi nhuận là 99,27$. => Công ty đang làm rất tốt về mặt dịch vụ chăm sóc khách hàng. Nên giữ vững về mặt services này và nâng cao hơn về chất lượng sản phẩm của mình.