

2024

Data Got Talent

BÁO CÁO KINH DOANH & ĐỀ XUẤT GIẢI PHÁP WA TAB

Thực hiện bởi: **Pentagon**



Western Australia Totalisator Agency Board

Báo cáo này nhằm mục đích cung cấp thông tin về tình hình kinh doanh, đề xuất mô hình cải thiện doanh thu và dự đoán doanh số của WA TAB.

I.	GIỚI THIỆU VÀ ĐÁNH GIÁ VỀ DOANH NGHIỆP	5
	Giới thiệu về doanh nghiệp	5
	Dữ liệu kinh doanh của doanh nghiệp	5
	Tình hình kinh doanh	14
	Đánh giá vấn đề của doanh nghiệp	27
II.	ĐỀ XUẤT GIẢI PHÁP CẢI THIỆN	28
III.	XÂY DỰNG VÀ ĐÁNH GIÁ MÔ HÌNH	34
	Data Preprocessing	34
	Mô hình ARIMA	35
	Mô hình học máy RNN theo LSTM và GRU	39
	Dự đoán doanh thu trong tương lai	42
IV.	TỔNG KẾT	43
	PHỤ LỤC	43

MỤC LỤC HÌNH ẢNH

Hình 1. Biểu đồ phân phối khách hàng theo độ tuổi.....	9
Hình 2. Tỷ lệ giới tính của tập khách hàng	10
Hình 3. Mức độ phân phối nơi cư trú của khách hàng	10
Hình 4. Tỷ lệ phân phối các ngày trong tuần mà khách hàng tham gia cá cược.....	11
Hình 5. Tổng thu nhập, chi phí, lợi nhuận và tổng số vé đã bán vào các ngày trong tuần.....	11
Hình 6. Tổng thu nhập, chi phí, lợi nhuận và tổng số vé đã bán vào các ngày	12
Hình 7. Heatmap thể hiện mối tương quan giữa các biến	13
Hình 8. Tổng quan về tình hình hoạt động	14
Hình 9. Doanh thu cá cược và lợi nhuận theo quý	14
Hình 10. Biểu đồ doanh thu và lợi nhuận theo tháng	14
Hình 11. Doanh thu cá cược tháng 12/2022.....	15
Hình 12. Doanh số bán vé theo thời gian.....	16
Hình 13. Doanh thu theo ngày.....	16
Hình 14. Số tiền chia cho người chơi thắng theo thời gian.....	17
Hình 15. Sự đóng góp của từng loại hình cá cược vào tổng doanh thu cá cược	18
Hình 16. Doanh thu của các loại hình cá cược theo thời gian.....	18
Hình 17. Doanh thu của các loại hình cá cược phân bổ vào các ngày trong tuần	19
Hình 18. Doanh thu của các loại hình cá cược theo nhóm tuổi	20
Hình 19. Doanh thu của các loại hình cá cược theo bang	20
Hình 20. Tỷ lệ đóng góp của từng loại hình cá cược vào doanh thu theo giới tính.....	21
Hình 21. Tổng quan khách hàng của WA TAB.....	21
Hình 22. Số người tham gia cá cược theo thời gian	22
Hình 23. Số khách hàng và doanh thu của từng nhóm tuổi	22
Hình 24. Doanh thu của các loại hình cá cược theo giới tính.....	23
Hình 25. Doanh thu theo tenure in days.....	23
Hình 26. Phân nhóm khách hàng dựa trên mô hình RFM	25
Hình 27. Doanh thu theo phân khúc khách hàng.....	25
Hình 28. Tỷ lệ lạm phát ở Úc 2021 - 2023 (Nguồn: SBSNews)	27
Hình 29. Tỷ lệ khách hàng trong nhóm qua 2 năm	29
Hình 30. Biểu đồ số lượng người chơi trò chơi điện tử trên toàn thế giới từ năm 2015 đến năm 2023 (Đơn vị: Tỷ người).....	33
Hình 31. Tỷ lệ thắng cược theo phân khúc khách hàng và trung bình	34
Hình 32. Doanh số theo thời gian.....	35

Hình 33. Biểu đồ dải trung bình và độ lệch chuẩn.....	35
Hình 34. Kết quả kiểm định tính dừng.....	36
Hình 35. Phân tích tính mùa vụ, xu hướng của dữ liệu doanh số.....	36
Hình 36. Mô hình ARIMA dự để dự đoán doanh số.....	37
Hình 37. Đặc điểm của mô hình	38
Hình 38. Kết quả dự đoán trên tập dữ liệu kiểm tra.....	38
Hình 39. Kết quả trực quan hóa phép vào toàn bộ dữ liệu.....	39
Hình 40. Kết quả dự đoán trên tập kiểm tra với LSTM.....	41
Hình 41. Kết quả dự đoán trên tập kiểm tra với GRU	41
Hình 42. Kết quả dự đoán tổng doanh thu của WA TAB trong quý 1 năm 2023.....	42

MỤC LỤC BẢNG

Bảng 1. Bảng miêu tả dữ liệu của doanh nghiệp	6
Bảng 2. Kết quả sau khi kiểm tra các giá trị trùng lặp và giá trị rỗng.....	7
Bảng 3. Bảng kết quả RFM.....	24
Bảng 4. Bảng thống kê mô tả kết quả RFM.....	24
Bảng 5. Kế hoạch thu hút và giữ chân cho từng phân khúc khách hàng	30
Bảng 6. Dữ liệu chuỗi thời gian sau khi chuyển đổi	34
Bảng 7. Bảng đánh giá các chỉ số của mô hình ARIMA	39
Bảng 8. Bảng đánh giá các chỉ số của 2 mô hình LSTM, GRU	42
Bảng 9. Kết quả chi tiết	42

I. GIỚI THIỆU VÀ ĐÁNH GIÁ VỀ DOANH NGHIỆP

Giới thiệu về doanh nghiệp

Racing and Wagering Western Australia (RWWA) là một tổ chức hoạt động với vai trò là cơ quan quản lý đua ngựa và đua chó săn trong phạm vi bang Tây Úc. Trong đó, RWWA giám sát các hoạt động đặt cược ngoài trường đua liên quan đến lĩnh vực đua ngựa. Hoạt động này được thực hiện thông qua **WA TAB**, một doanh nghiệp nhà nước quản lý hơn 300 địa điểm bán lẻ cá cược và vận hành dịch vụ cá cược trực tuyến **TABtouch**. WA TAB cho phép người chơi đặt cược vào các sự kiện đua và các môn thể thao khác nhau, bao gồm cả các giải đấu trong nước và quốc tế.

TAB cung cấp 2 hình thức cá cược gồm Fixed-odds (cá cược với tỷ lệ cố định) và Pari-mutuel (cá cược chung) cho cả 2 thể loại Racing và Sports:

- **Fixed-odds Betting (Cược Cố Định):** Trong cược cố định, người chơi đặt cược vào kết quả của một sự kiện cụ thể (ví dụ: ngựa chiến thắng, chó về nhất,... đối với thể loại Racing hay cược đội chơi thắng, đoán tỉ số bàn ghi,... đối với Sports) nhưng với một tỷ lệ cược được xác định trước bởi nhà cái. Tỷ lệ cược này là cố định và không thay đổi dù số tiền đặt cược có thay đổi. Với hình thức này, thu nhập của nhà cái phụ thuộc vào việc tính toán tỷ lệ cược sao cho có lợi nhuận bất kể kết quả thắng hay thua.
- **Pari-mutuel Betting (Cược Pari-mutuel):** Trái ngược với cược cố định, ở đây, toàn bộ tiền cược từ người chơi được tập hợp lại thành một nhóm, gọi là pool. Sau đó, tiền cược được chia đều giữa những người chơi đặt cược đúng. Tỷ lệ trả thưởng phụ thuộc vào tổng số tiền đặt cược cho mỗi lựa chọn và số người thắng cược. Một phần trăm hoa hồng được trích từ tổng tiền cược để làm lợi nhuận cho nhà cái.

Với tập dữ liệu được cung cấp bao gồm dữ liệu về khách hàng và giao dịch trong 2 năm từ ngày 01/01/2021 đến 31/12/2022, việc phân tích dữ liệu và đưa ra các báo cáo về tình hình kinh doanh giúp doanh nghiệp có cái nhìn tổng quan về hiệu quả hoạt động cũng như là hỗ trợ doanh nghiệp ra các quyết định mang tính chiến lược. Bên cạnh đó, mục tiêu hiện tại của doanh nghiệp cũng đề cập đến việc dự báo chính xác cho tổng số tiền đặt cược của khách hàng trong tương lai.

Dữ liệu kinh doanh của doanh nghiệp

Mô tả dữ liệu

Dữ liệu được cung cấp gồm 12.364.101 giao dịch của 144.869 khách hàng trong khoảng thời gian từ ngày 01/01/2021 đến ngày 31/12/2022 với 16 thuộc tính được mô tả chi tiết ở bảng sau:

Bảng 1. Bảng miêu tả dữ liệu của doanh nghiệp

STT	Thuộc tính	Mô tả	Kiểu dữ liệu	Các giá trị
1.	DATE_DIM	Ngày thực hiện giao dịch	object	730 ngày kể từ ngày 01/01/2021 đến ngày 31/12/2022
2.	DAY_OF_WEEK	Tên ngày thực hiện giao dịch trong tuần	object	Mon, Tue, Wed, Thu, Fri, Sat, Sun
3.	BET_ACCOUNT_NUM_HASH	Mã định danh tài khoản cá cược của khách hàng	int64	144869 khách hàng
4.	AGE	Tuổi của khách hàng	float64	Từ 18 đến 104 tuổi
5.	AGE_BAND	Nhóm tuổi của khách hàng	object	18-24, 25-34, 35-44, 45-54, 55-64, 65+
6.	GENDER	Giới tính của khách hàng	object	M: nam, F: nữ, U: không xác định
7.	TENURE_IN_DAYS	Số ngày kể từ ngày tạo tài khoản đến ngày thực hiện giao dịch cá cược	int64	
8.	RESIDENTIAL_STATE	Tiểu bang mà khách hàng cư trú	object	WA: Tây Australia, OTH: khu vực khác
9.	FOB_RACING_TURNOVER	Tổng số tiền đã cược đã được sử dụng cho các sự kiện đua với tỷ lệ cược cố định	float64	
10.	FOB_SPORT_TURNOVER	Tổng số tiền đã cược đã được sử dụng cho các sự kiện thể thao với tỷ lệ cược cố định	float64	
11.	PARI_RACING_TURNOVER	Tổng số tiền mà người chơi đã đặt cược vào các sự kiện đua mà tiền thắng cuộc sẽ được chia đều cho những người giành chiến thắng sau khi trừ các phí và thuế	float64	
12.	PARI_SPORT_TURNOVER	Tổng số tiền mà người chơi đã đặt cược vào các sự kiện thể thao mà tiền thắng cuộc sẽ được chia đều cho những người giành chiến thắng sau khi trừ các phí và thuế	float64	

13.	TOTAL_ TURNOVER	Tổng số tiền mà người chơi đã đặt cược trong suốt một ngày	float64
14.	DIVIDENDS _PAID	Số tiền mà một người chơi nhận được sau khi thắng cược	float64
15.	GROSS _MARGIN*	Biên lợi nhuận gộp mà nhà cung cấp cá cược đã kiếm được từ hoạt động cá cược	float64
16.	TICKETS	Tổng số vé mà khách hàng đã mua trong ngày	int64

* Trường dữ liệu được cung cấp có tên là `GROSS_MARGIN`, được mô tả là biên lợi nhuận gộp mà nhà cung cấp cá cược đã kiếm được từ hoạt động cá cược. Tuy nhiên, sau khi phân tích, nhóm nghiên cứu nhận thấy rằng các giá trị trong tập dữ liệu không phản ánh đúng công thức và khoảng giá trị (0-100%) của biên lợi nhuận gộp. Do đó, trong ngữ cảnh bài toán, nhóm sẽ nhận định rằng đây là trường dữ liệu về lợi nhuận gộp `GROSS_PROFIT`, tức là tổng doanh thu trừ đi chi phí sản xuất hoặc cung cấp dịch vụ.

Công thức tính doanh thu:

$$\text{Total_turnover} = \text{Fob_racing_turnover} + \text{Fob_sport_turnover} + \text{Pari_racing_turnover} + \text{Pari_sport_turnover}$$

Nhìn chung thì dữ liệu gồm có các loại thông tin sau:

- Thông tin ngày giao dịch: Ngày thứ bao nhiêu trong tuần.
- Thông tin giao dịch: Thời gian giao dịch, khách hàng giao dịch, tổng số tiền của mỗi loại hình giao dịch, tổng số tiền cho tất cả giao dịch, ...
- Thông tin khách hàng: Giới tính, độ tuổi, nhóm tuổi, tiểu bang mà khách hàng cư trú, ...

Làm sạch và khai phá dữ liệu

Trong quá trình khai phá dữ liệu cho thấy dữ liệu không tồn tại cái giá trị trùng lặp tuy nhiên có tồn tại các giá trị null như bảng bên dưới:

Bảng 2. Kết quả sau khi kiểm tra các giá trị trùng lặp và giá trị rỗng

STT	Thuộc tính	Số giá trị null
1.	DATE_DIM	0
2.	DAY_OF_WEEK	0
3.	BET_ACCOUNT_NUM_HASH	0
4.	AGE	2676

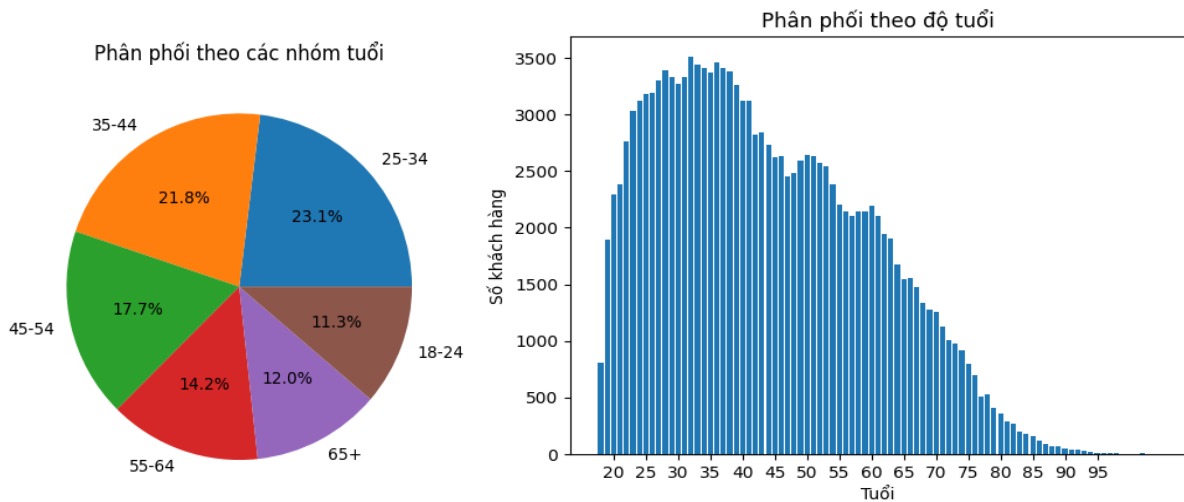
5.	AGE_BAND	0
6.	GENDER	0
7.	TENURE_IN_DAYS	0
8.	RESIDENTIAL_STATE	0
9.	FOB_RACING_TURNOVER	3.787.445
10.	FOB_SPORT_TURNOVER	10.584.521
11.	PARI_RACING_TURNOVER	3.836.612
12.	PARI_SPORT_TURNOVER	12.348.423
13.	TOTAL_TURNOVER	0
14.	DIVIDENDS_PAID	0
15.	GROSS_MARGIN	0
16.	TICKETS	0

Lý giải cho dữ liệu bị null thì các thuộc tính FOB_RACING_TURNOVER, FOB_SPORT_TURNOVER, PARI_RACING_TURNOVER, PARI_SPORT_TURNOVER tồn tại các giá trị null bởi vì có khách hàng có thể chỉ chọn duy nhất một loại cá cược trong một ngày giao dịch dẫn đến các loại cá cược khác có giá trị null.

Đối với thuộc tính AGE có 2676 dòng có giá trị null, tương ứng với 10 khách hàng riêng biệt có giá trị thuộc tính AGE là null trên tổng số 144.869 khách hàng, điều này làm ảnh hưởng đến mô hình RFM, vậy nên cần loại bỏ để hỗ trợ việc xây dựng mô hình mà không ảnh hưởng nhiều đến kết quả thực tế.

Dữ liệu vẫn tồn tại một số giá trị không phù hợp đó là các thuộc tính FOB_RACING_TURNOVER và FOB_SPORT_TURNOVER có giá trị nhỏ hơn 0. Giả định ban đầu đặt ra là tổ chức hoạt động cá cược dưới hình thức nợ thanh toán. Tuy nhiên, sau khi kiểm chứng từ kênh online - website doanh nghiệp thì hình thức này không tồn tại, người chơi bắt buộc phải thanh toán sau khi cược. Do đó, đây được xem như những dữ liệu cần loại bỏ vì không có ý nghĩa thực tế trong thời gian hiện tại.

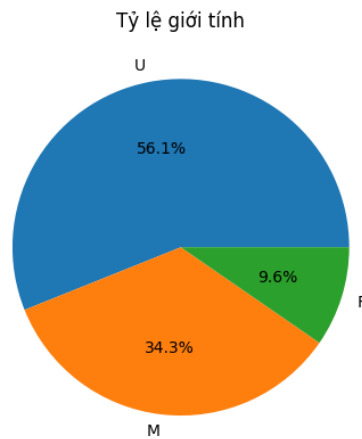
Sau khi xử lý dữ liệu cần trực quan hóa để mô tả chi tiết hơn và phân tích chẩn đoán dữ liệu. Đối với dữ liệu nhân khẩu học của khách hàng, đầu tiên đi vào phân tích thành phần khách hàng theo độ tuổi.



Hình 1. Biểu đồ phân phối khách hàng theo độ tuổi

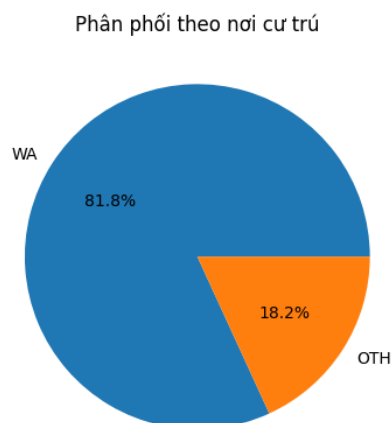
Dữ liệu về số lượng khách hàng cá cược cho thấy một số điểm quan trọng về độ tuổi của khách hàng:

- Sự ít quan tâm ở nhóm trẻ: Nhóm tuổi 18-24 có số lượng người chơi thấp nhất, chỉ với hơn 11% số lượng người chơi. Điều này chỉ ra rằng nhóm tuổi này ít quan tâm đến cá cược, hoặc họ có ít tiền hơn để dành cho việc này, chú trọng hơn vào các hoạt động giải trí khác.
- Sự phổ biến nhất là nhóm 25-34 và 35-44 tuổi: Với gần 45% số lượng người chơi thuộc nhóm này, nhóm tuổi này là nhóm có số lượng người chơi cá cược nhiều nhất. Điều này cho thấy rằng nhóm tuổi này có sự quan tâm mạnh mẽ đối với cá cược, nguyên nhân do họ đang ở độ tuổi có thu nhập ổn định và sẵn lòng tham gia các hoạt động giải trí có tính rủi ro. Đặc biệt nhất là nhóm 25-34 vì đa phần họ sẽ trong độ tuổi độc thân và tự chủ tất cả vấn đề tài chính.
- Sự ổn định trong nhóm tuổi trung niên: Nhóm tuổi 35-44 và 45-54 vẫn duy trì một lượng người chơi ổn định, cho thấy rằng cá cược vẫn là một hình thức giải trí được ưa chuộng trong những nhóm tuổi này.
- Sự quan tâm của người cao tuổi: nhóm tuổi 55-64 và 65+ vẫn có số lượng người chơi đáng kể, cho thấy rằng cá cược vẫn có sức hấp dẫn đối với những người ở độ tuổi này, có thể do họ tìm kiếm các hình thức giải trí trong thời gian rảnh rỗi của mình.
- Sự giảm dần theo độ tuổi: Có một xu hướng giảm dần số lượng người chơi từ nhóm tuổi 35-44 đến nhóm 65+. Điều này phản ánh sự thay đổi trong khả năng tài chính, sở thích, và thời gian rảnh rỗi khi mọi người lớn tuổi hơn.



Hình 2. Tỷ lệ giới tính của tập khách hàng

Về giới tính, phần lớn người chơi có giới tính không xác định chiếm hơn ½ trên tổng số khách hàng với 56.1%. Tiếp theo là nhóm khách hàng nam giới với 34.3% và cuối cùng là nhóm nữ giới với 9.6%. Nhóm người không xác định giới tính có thể là giới tính khác hoặc chưa cung cấp thông tin về giới tính. Doanh nghiệp cần phải nghiên cứu thêm về nhóm giới tính không xác định và lý do chênh lệch của về tỷ lệ giới tính khách hàng vì điều này đóng góp vào quá trình tiếp thị đến khách hàng từ việc phân nhóm đến cách tiếp thị phù hợp cho từng đối tượng khách hàng.

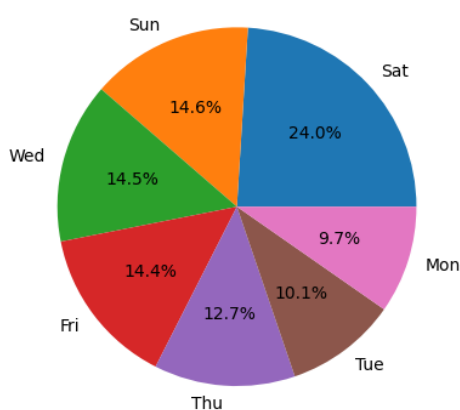


Hình 3. Mức độ phân phối nơi cư trú của khách hàng

Đa số khách hàng của WA TAB là người dân Tây Australia với khoảng hơn 80% và gần 20% còn lại là cư trú ở khu vực khác. Sự không cân đối này do WA TAB là một doanh nghiệp ở phía Tây Australia nên đa số khách hàng của họ sẽ ở đây.

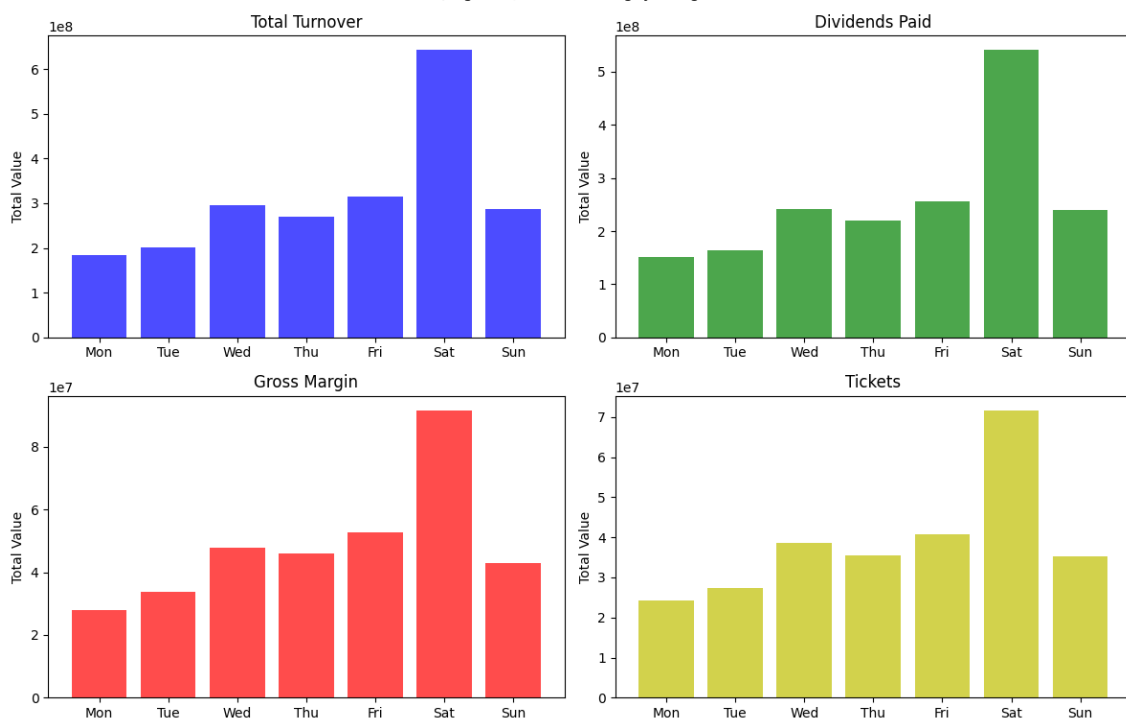
Đối với dữ liệu giao dịch cá cược, tỷ lệ phân phối số khách hàng cá cược vào các ngày trong tuần như sau:

Phân phối các ngày trong tuần



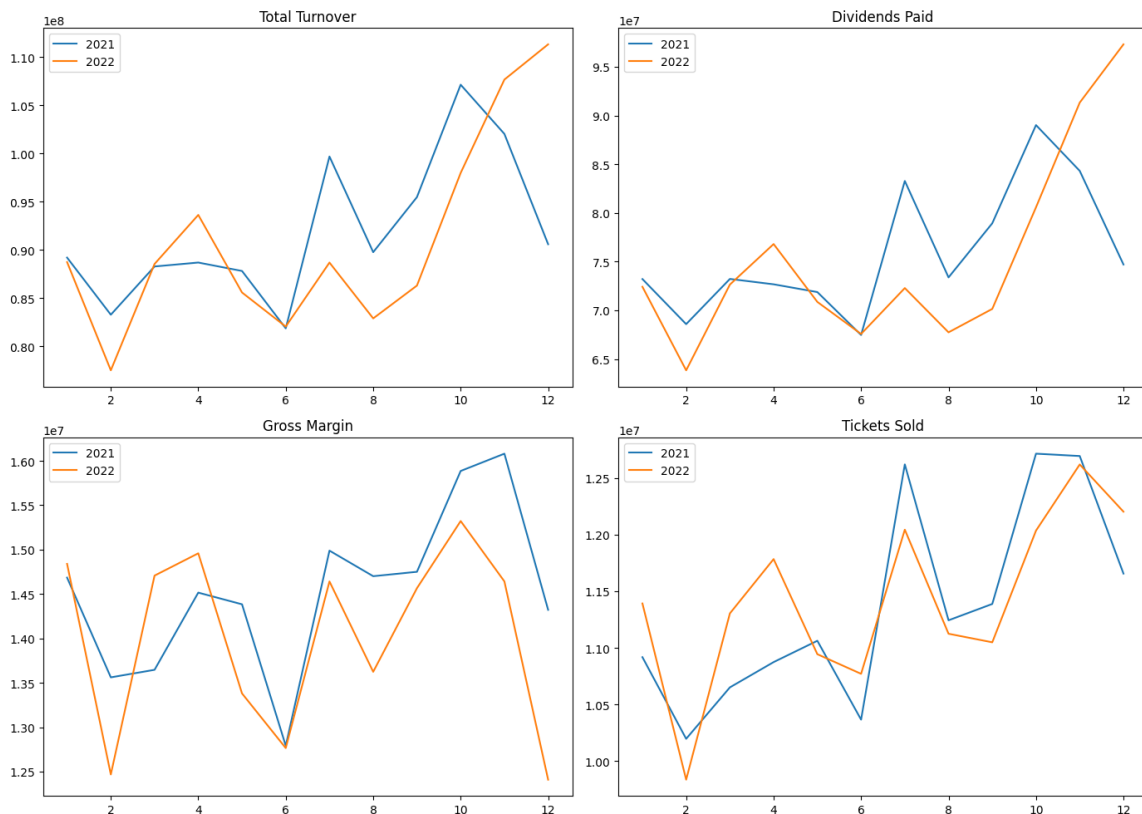
Hình 4. Tỷ lệ phân phối các ngày trong tuần mà khách hàng tham gia cá cược

Dữ liệu giao dịch theo các ngày trong tuần



Hình 5. Tổng thu nhập, chi phí, lợi nhuận và tổng số vé đã bán vào các ngày trong tuần

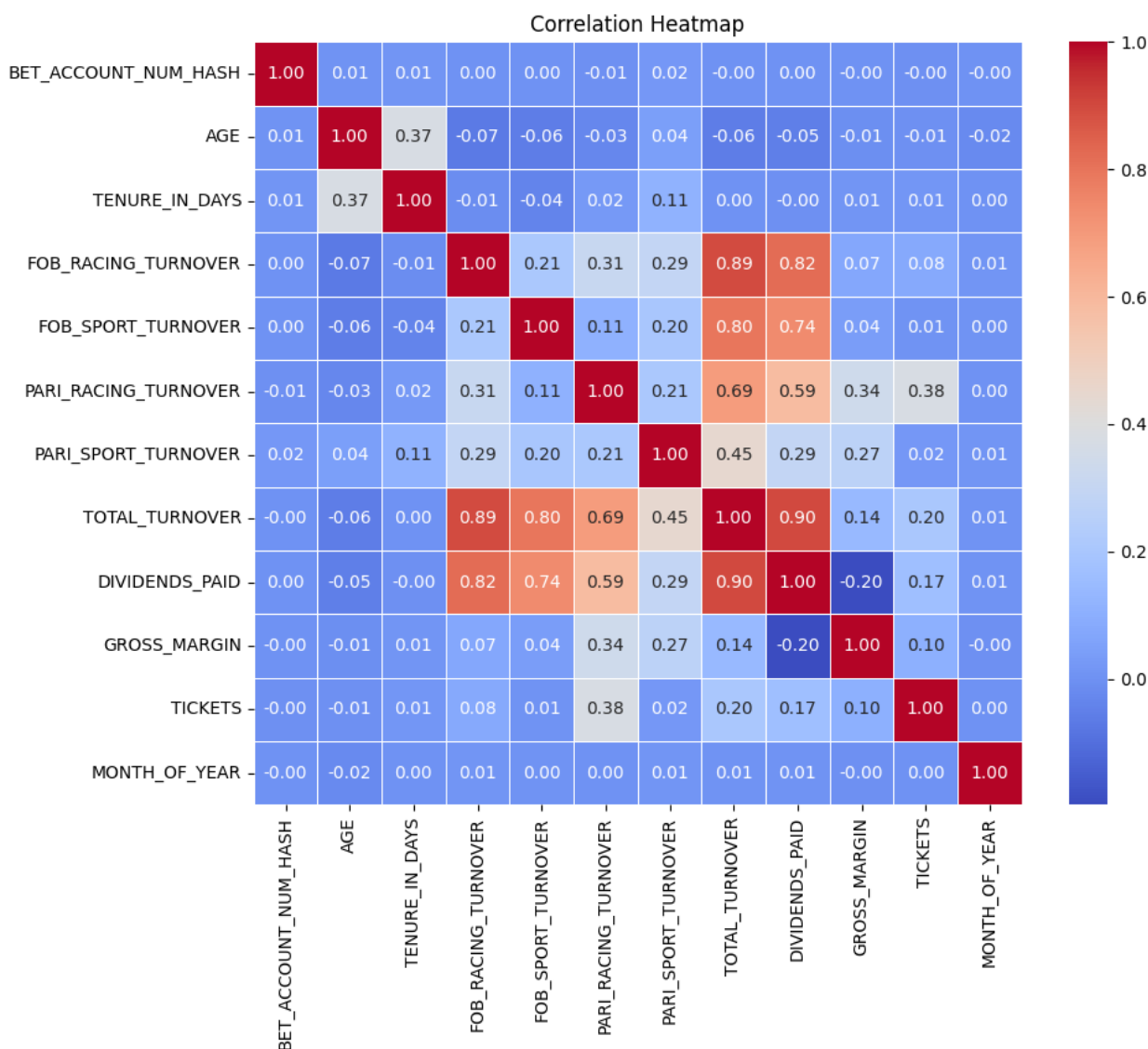
Từ các biểu đồ phân phối có thể thấy khách hàng của WA TAB thường thực hiện các giao dịch nhiều nhất vào thứ 7 bởi vì các ngày cuối tuần thường là các ngày nghỉ của đa số người làm việc, do đó họ có thời gian nhiều hơn để tham gia các hoạt động giải trí như cá cược sự kiện đua hoặc thể thao. Chủ nhật tuy cũng là cuối tuần nhưng đa số họ sẽ dành thời gian để tham dự các nghi lễ, đi lễ nhà thờ, hoặc thực hiện các hoạt động tôn giáo khác, thay vì tham gia các hoạt động giải trí cá cược.



Hình 6. Tổng thu nhập, chi phí, lợi nhuận và tổng số vé đã bán vào các ngày

- Doanh số tổng cộng (TOTAL_TURNOVER): Có sự tăng nhẹ trong doanh số từ năm 2021 đến 2022, với tháng cao điểm là tháng 12 năm 2022, đạt khoảng 111 triệu. Tháng có doanh số thấp nhất trong năm 2022 là tháng 2.
- Số tiền trả lại cho người chơi thắng (DIVIDENDS_PAID): Số tiền trả lại cho người chơi thắng có xu hướng giảm nhẹ từ năm 2021 sang 2022. Thể hiện xu hướng mùa vụ của việc kinh doanh này.
- Lợi nhuận: Tháng 10 và 11 của cả hai năm đều có lợi nhuận cao, cho thấy có thể có các sự kiện hoặc mùa cao điểm ảnh hưởng đến lợi nhuận.
- Vé đã bán (TICKETS_SOLD): Số lượng vé bán ra có sự tăng trong năm 2022 so với năm 2021, đặc biệt là vào các tháng cuối năm. Sự tăng này có thể phản ánh sự quan tâm tăng lên từ phía khách hàng hoặc các chiến dịch tiếp thị hiệu quả và sự trở lại sau dịch của WA TAB.

Phân tích tương quan giữa các biến thông qua heatmap



Hình 7. Heatmap thể hiện mối tương quan giữa các biến

Khi khai thác mối quan hệ giữa các biến, nhóm nhận thấy rằng TOTAL_TURNOVER bị ảnh hưởng rất nhiều bởi FOB và PARI nguyên nhân là: Tổng doanh số bằng tổng của từng hình thức và thể loại.

DIVIDENDS phụ thuộc rất nhiều vào FOB_SPORT, FOB_RACING bởi vì giá trị cược lớn sẽ mang lại số tiền lớn khi chiến thắng tuy nhiên còn phụ thuộc vào số lượng vé thắng và tỷ lệ thắng của các vé cược, nó cũng phụ thuộc nhiều vào PARI_RACING và PARI_SPORT tuy nhiên còn phụ thuộc vào số lượng vé của người chơi chiến thắng và tổng giá trị cược của tất cả người chơi vào hình thức này.

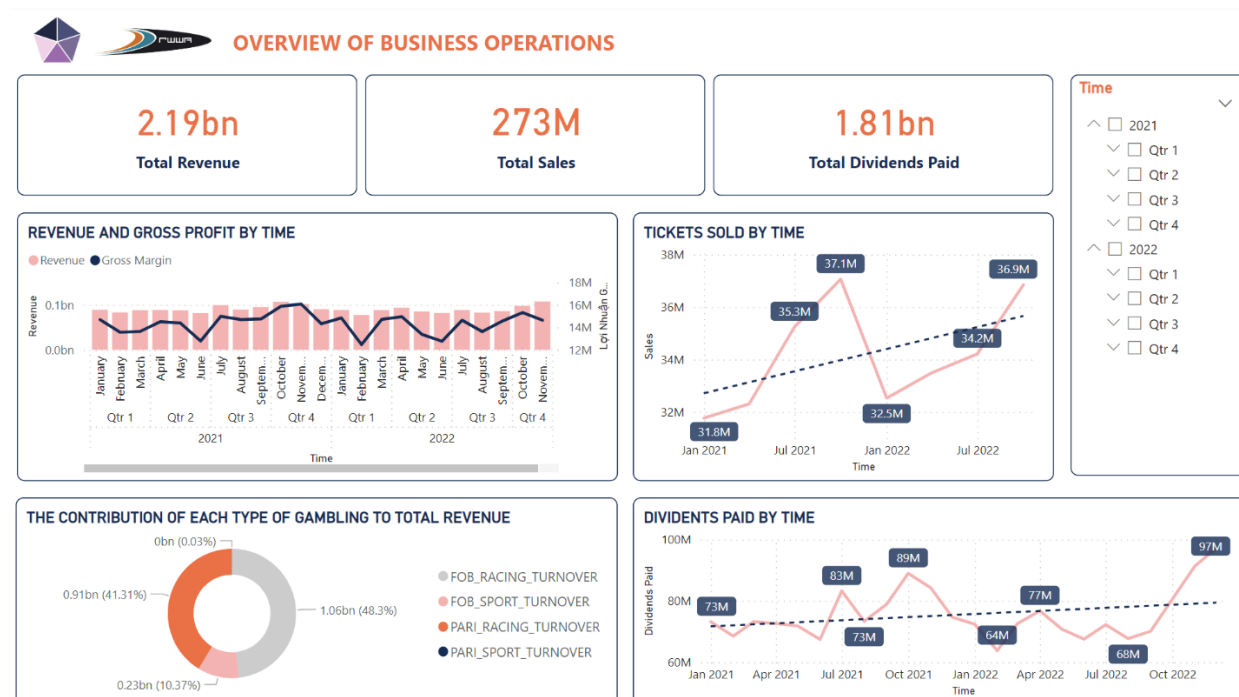
Về GROSS_MARGIN, nhóm nhận thấy dữ liệu này được gán nhãn chưa phù hợp, với ngữ cảnh bài toán, nhóm sẽ nhận định đây là biến GROSS_PROFIT bởi vì:

$$gross\ margin = \frac{(revenue - cost)}{revenue}$$

Vậy nên gross margin phải luôn nhỏ hơn hoặc bằng 1 hoặc nhỏ hơn hoặc bằng 100 với đơn vị %

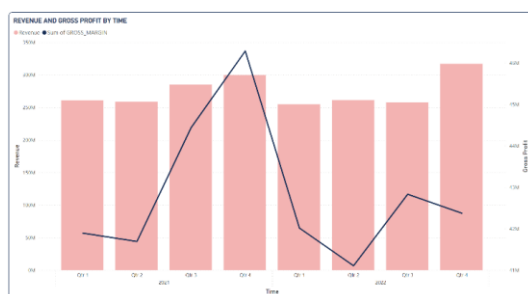
Tình hình kinh doanh

Tổng quan

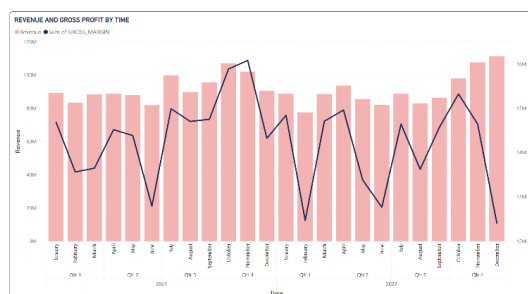


Hình 8. Tổng quan về tình hình hoạt động

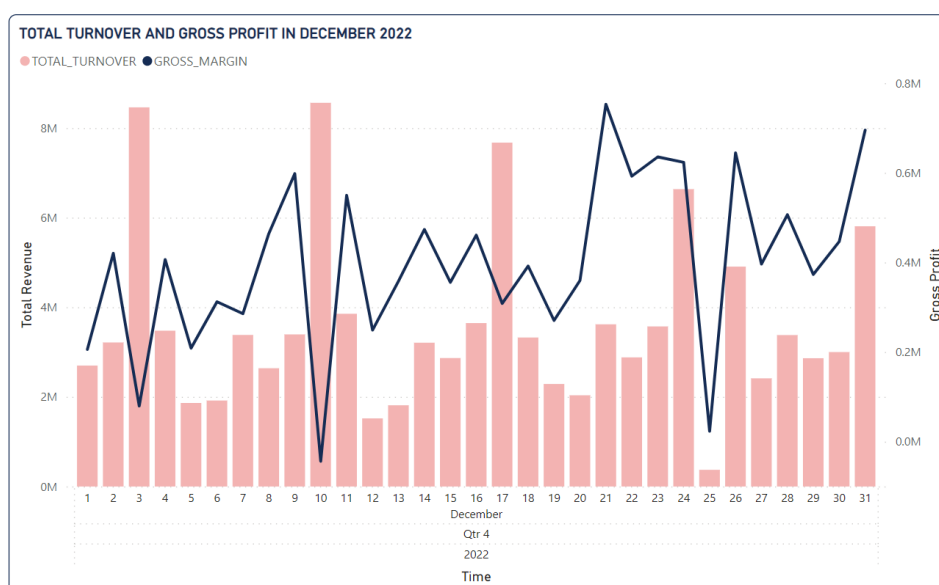
Biểu đồ tổng quan hoạt động kinh doanh của WA TAB trong khoảng thời gian phân tích từ ngày 1/1/2021 đến ngày 31/12/2022. Tổng doanh thu của các loại hình cá cược là 2.19 tỷ AUD với số vé bán ra là 273 triệu vé. Tổng số tiền phải chia cho người chơi thắng là 1.81 tỷ AUD. Nhìn chung, doanh thu trong 2 năm không có sự chênh lệch nhiều so với tổng thể, doanh số bán vé có xu hướng tăng, tuy nhiên lợi nhuận lại giảm đáng kể.



Hình 9. Doanh thu cá cược và lợi nhuận theo quý



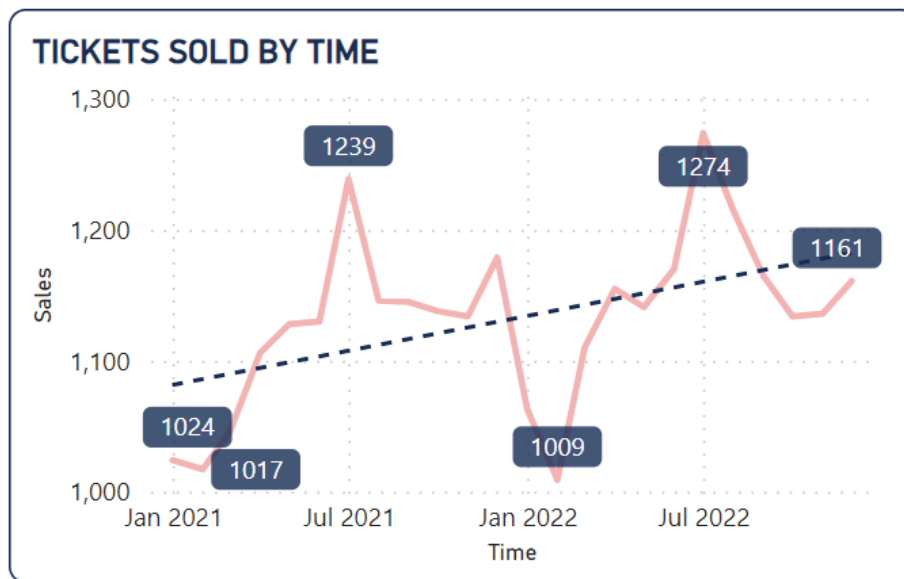
Hình 10. Biểu đồ doanh thu và lợi nhuận theo tháng



Hình 11. Doanh thu cá cược tháng 12/2022

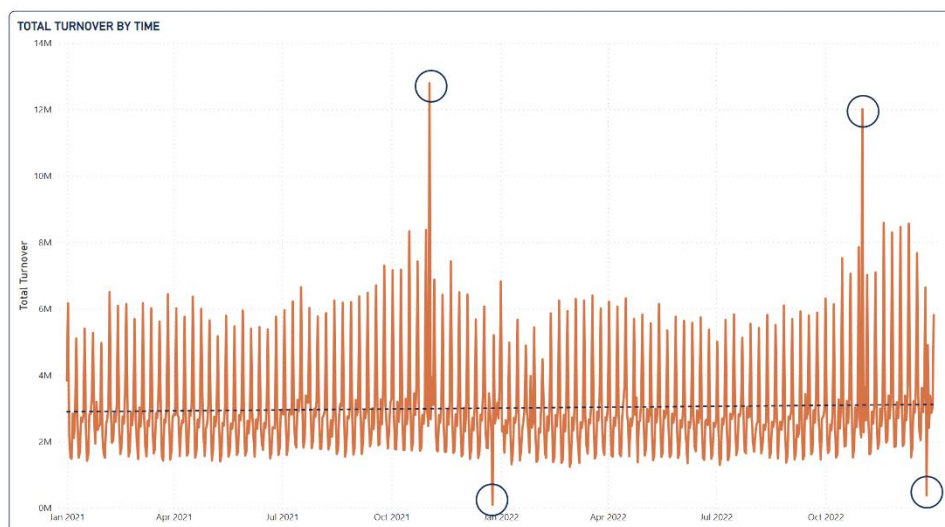
Doanh thu cá cược có xu hướng tăng từ đầu năm đến cuối năm dù có sự biến động mạnh trong giai đoạn phân tích. Tháng có doanh thu cao nhất là tháng 12/2022. Trong khoảng thời gian từ tháng 8 đến tháng 12/2022, tổng doanh thu đã tăng mạnh lên đến 34.25% (tăng 2,839,3324.34 AUD). Mặc dù vậy, trong giai đoạn tháng 10 đến tháng 12/2022, doanh thu tăng lên nhưng lợi nhuận lại giảm xuống đột ngột và chạm đáy ở tháng 12/2022 với 12,408,767.04 AUD, giảm 13,36% so với cùng kỳ năm trước. Tháng 2 có doanh thu thấp nhất có thể do doanh nghiệp chịu ảnh hưởng của một sự kiện bất thường trong thời điểm này.

Đối với biến động lợi nhuận giữa 2 năm các quý ghi nhận sự giảm đáng kể. Quý 2, quý 3 và quý 4 ghi nhận có sự giảm trong lợi nhuận lần lượt là 1.41%, 3.62% và 8.46% so với cùng kỳ năm 2021. Ngoài ra, lợi nhuận cũng ghi nhận những tháng có sự giảm so với tháng trước bao gồm tháng 6/2021, tháng 2/2022, tháng 6/2022 và tháng 12/2022.



Hình 12. Doanh số bán vé theo thời gian

Doanh số bán vé từ năm 2021 đến cuối năm 2022 có chiều hướng tăng. Giai đoạn tăng trưởng dài nhất của doanh số là từ tháng 1 đến tháng 7/2022 với 1,685,833 vé. Trong năm 2022, doanh số bán vé cũng tiếp tục tăng và đạt đỉnh điểm vào quý 4. Doanh số thường giảm vào các thời điểm là tháng 2, 6, 8 và 12, với mức thấp nhất ghi nhận là vào tháng 2/2022.

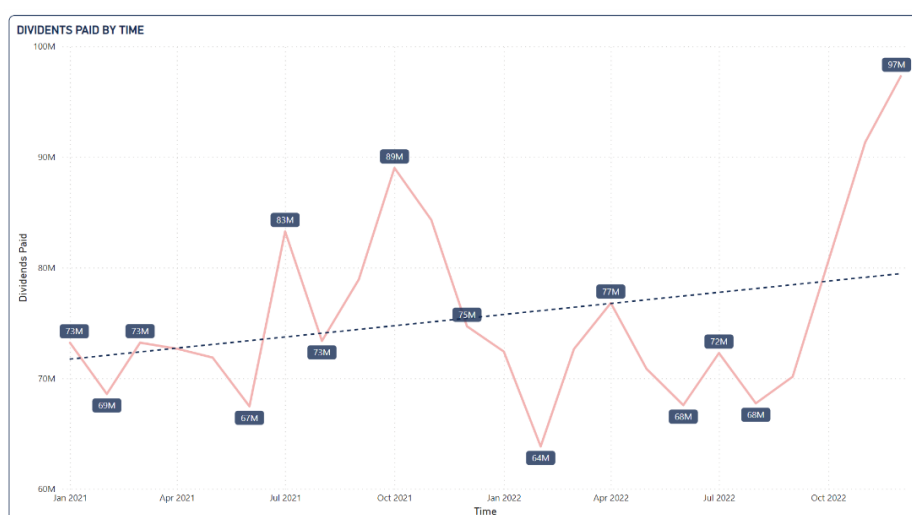


Hình 13. Doanh thu theo ngày

Doanh thu theo ngày xuất hiện 1 chu kỳ nhất định với doanh thu thấp vào các ngày đầu tuần (thứ 2, 3) và cao nhất vào thứ 7, các ngày còn lại trong tuần có doanh thu khá đồng đều với những biến động nhỏ. Đáng chú ý, ngày 2/11/2021 và 1/11/2022 lại là hai ngày có doanh thu và doanh số cao nhất trong năm, nhờ vào sự kiện Melbourne Cup (thứ 3 đầu tiên của tháng 11) - một trong những giải đua ngựa danh giá nhất thế giới. Melbourne Cup thu hút sự chú ý của một lượng lớn người

hâm mộ và giới truyền thông, làm tăng cao nhu cầu tham gia cá cược đua ngựa, và đặc biệt là góp phần quan trọng vào việc tạo ra doanh thu cao nhất trong năm cho WA TAB. Điều này cũng được minh chứng thông qua việc doanh thu và doanh số bán vé của cả hai loại hình racing tăng đáng kể vào 2 ngày này. Ngược lại, ngày 25/12 hàng năm thường là thời điểm có doanh thu và doanh số thấp nhất trong năm, bởi vì đó là ngày Lễ Giáng Sinh (Christmas Day). Mặc dù Christmas Eve được tổ chức khá lớn trên toàn thế giới, tại Úc, Christmas Day được coi là sự kiện lớn nhất trong mùa và mọi người thường tập trung vào việc tổ chức các bữa tiệc cả ngày và sum vầy bên gia đình thay vì tham gia các hoạt động giải trí như cá cược.

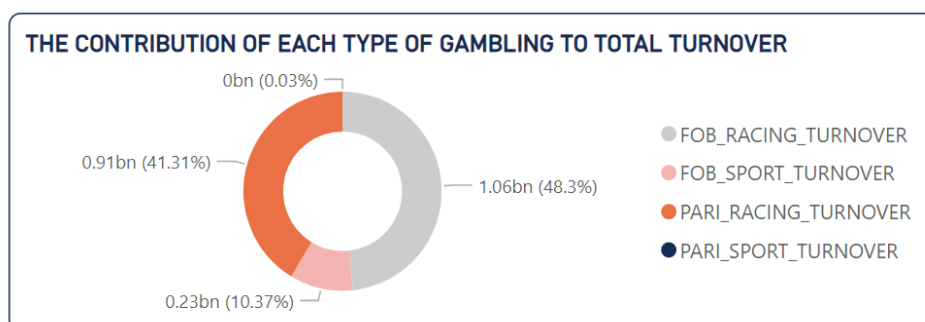
Nhìn chung, giai đoạn tăng trưởng tốt là khoảng tháng 9 đến tháng 11 hàng năm. Điều này được giải thích qua việc, đây là thời điểm mùa xuân diễn ra ở nước Úc, có thời tiết dễ chịu, không khí ấm áp và lượng mưa trung bình thấp, phù hợp với việc tổ chức các hoạt động ngoài trời như các cuộc đua động vật và các trận thi đấu thể thao. Ngoài ra, vào mùa hè từ tháng 12 đến tháng 2, doanh thu và doanh số bán vé có xu hướng giảm do đây là cao điểm du lịch ở Úc với những ngày lễ đặc biệt như ngày đầu năm mới (1/1) Lễ hội pháo hoa ở Perth (thủ phủ của miền Tây nước Úc) nhân ngày Quốc Khánh 26/1, Lễ hội Sydney Mardi Grass và rất nhiều hoạt động dành cho mùa hè khác. Mùa thu và mùa đông, doanh thu có sự biến động do xuất hiện các ngày mưa nhiều và tuyết rơi, cản trở việc tiến hành các cuộc đua và các sự kiện thể thao, cũng như hạn chế trong việc di chuyển và tham gia của người chơi trực tiếp.



Hình 14. Số tiền chia cho người chơi thắng theo thời gian

Số tiền chia cho người chơi thắng có chiều hướng tăng trong năm, Quý 1 và 2 của cả hai năm đều ghi nhận số tiền chia cho người chơi thắng thấp hơn đáng kể so với quý 4. Các tháng có số tiền chia cho người chơi thắng thấp thường là tháng 2, 6 và 8. Đỉnh điểm của số tiền chia cho người chơi thắng là vào tháng 10 cho năm 2021 và tháng 12 cho năm 2022. Đặc biệt là vào cuối năm, với

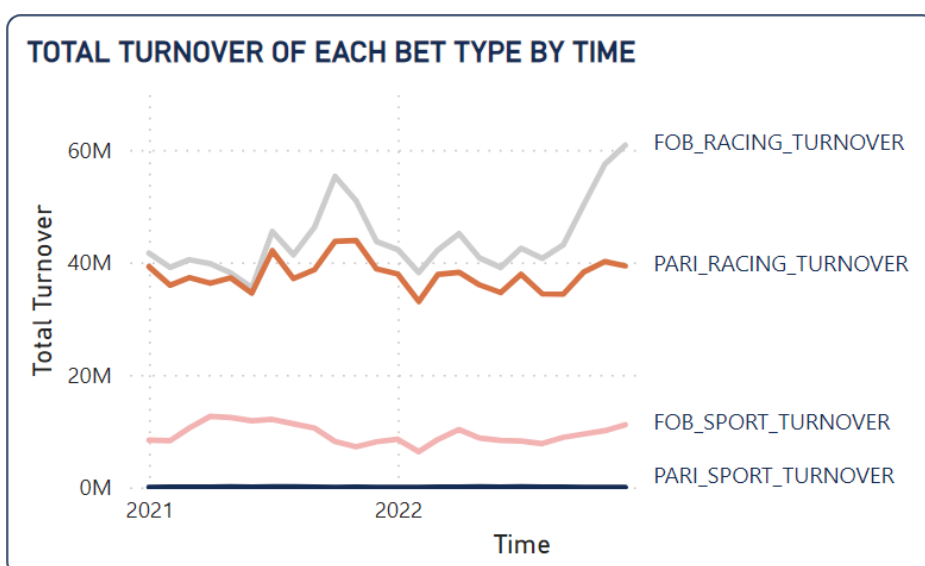
mức tăng cao nhất diễn ra từ tháng 8 đến tháng 12/2022, tăng lên 29,562,993.07 AUD (43.64%). Đặc biệt, tháng 12 năm 2022 tăng 30.35% so với cùng kỳ năm trước.



Hình 15. Sự đóng góp của từng loại hình cá cược vào tổng doanh thu cá cược

Về đóng góp của các loại hình cá cược vào doanh thu, có thể thấy fixed-odds racing chiếm tỷ lệ cao nhất với 48.3%, tiếp theo là pari-mutuel racing với 41.31%. Loại cá cược có tỷ lệ đóng góp thấp nhất là pari-mutuel sports chỉ với 0.03%. Cá cược fixed-odds được ưa chuộng hơn so với pari-mutuel có thể do hình thức pari-mutuel rủi ro hơn khi người chơi không biết được số tiền họ sẽ được chia sau khi thắng là bao nhiêu. Về thể loại, người chơi đặt cược nhiều vào thể loại racing hơn sports, điều này được giải thích qua việc WA TAB có chiến lược tập trung vào cung cấp các dịch vụ cá cược cho các sự kiện đua động vật.

Sự đóng góp của các thể loại cá cược



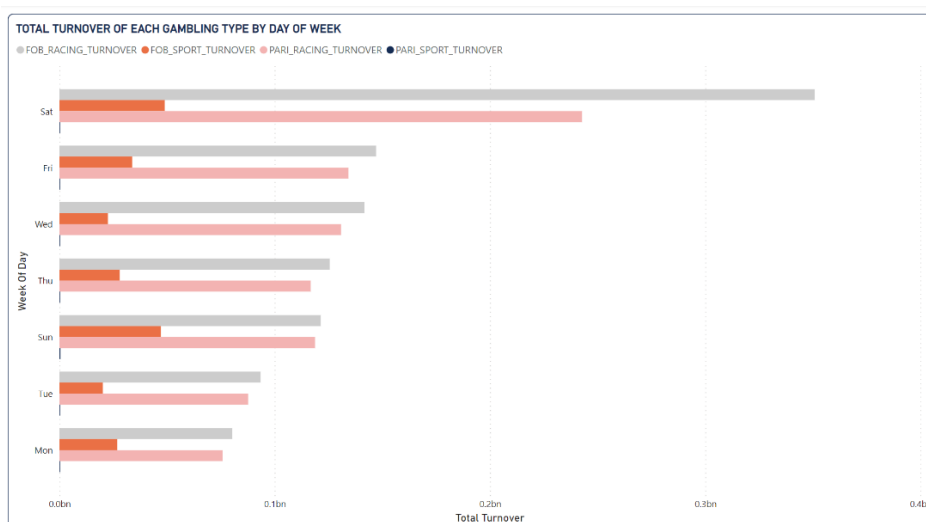
Hình 16. Doanh thu của các loại hình cá cược theo thời gian

Như phân tích ở trên, thể loại cá cược pari-mutuel sports có doanh thu thấp nhất trong 4 thể loại mà WA TAB cung cấp. Sau khi phân tích, có thể nhận thấy thể loại này có xu hướng giảm trong giai

đoạn năm 2021 đến năm 2022. Trong năm, thể loại này có xu hướng tăng từ đầu năm đến quý 2, giữ ổn định đến quý 3 và bắt đầu giảm đến cuối năm. Ghi nhận được giai đoạn tháng 1-2 và tháng 10-12 có doanh thu thấp và chênh lệch nhiều so với các tháng giữ năm. Ngày 25/6/2021 và ngày 7/8/2022 có doanh thu cao nhất trong năm.

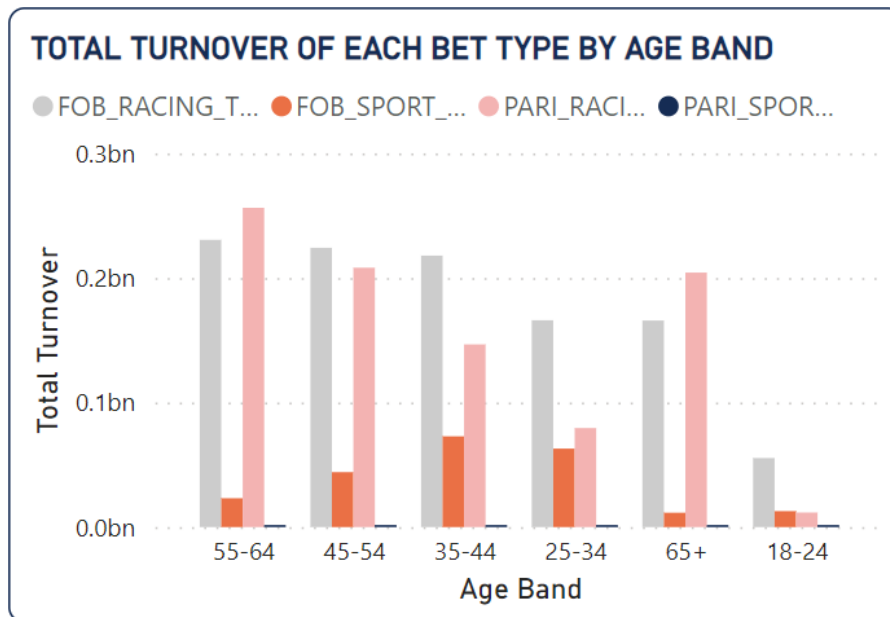
Đối với thể loại cá cược fixed-odds sports, tình hình hoạt động có chiều hướng không tốt khi doanh thu trong năm 2022 giảm 12.35% so với 2021. Ghi nhận doanh thu cao nhất vào tháng 4/2021 và thấp nhất là vào tháng 2/2022. Ngày 25/9/2021 và ngày 24/9/2022 là các ngày có điểm dữ liệu cao nhất trong năm.

Đối với thể loại cá cược fixed-odds racing, ghi nhận xu hướng tăng nhẹ trong giai đoạn phân tích. Trong năm, doanh thu thường tăng vào các tháng cuối năm từ tháng 9 đến tháng 12. Trong khi đó, các tháng có doanh thu thấp là vào tháng 2, 6 và tháng 8. Đối với thể loại cá cược pari-mutuel racing, doanh thu có xu hướng giảm. Doanh thu cao nhất là vào tháng 11/2021 và thấp nhất là vào tháng 2/2022.



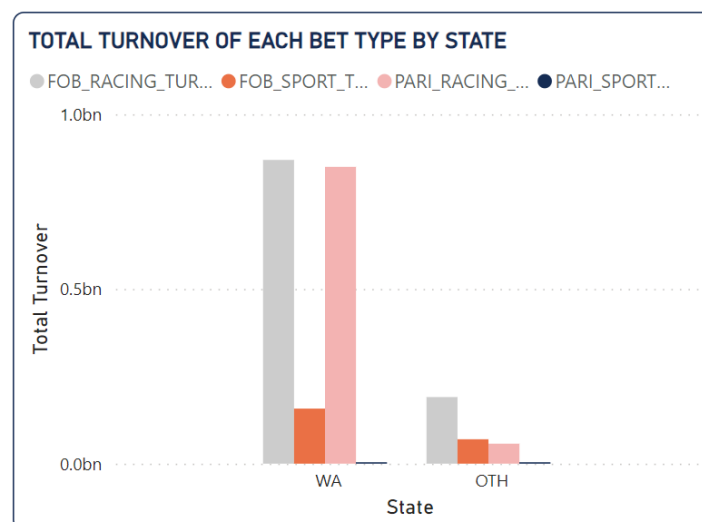
Hình 17. Doanh thu của các loại hình cá cược phân bố vào các ngày trong tuần

Về biểu đồ phân bố doanh thu các loại hình cá cược theo các ngày trong tuần có thể thấy doanh thu lớn nhất vào ngày thứ 7 và thấp nhất vào các ngày đầu tuần là thứ 2 và thứ 3. Đối với các ngày còn lại trong tuần, doanh thu phân bố khá đồng đều. Doanh thu thể loại cá cược fixed-odds và pari-mutuel racing có doanh thu cao vào các ngày thứ 7. Như giải thích ở trên, tuy chủ nhật cũng là ngày cuối tuần nhưng Úc được biết đến như một quốc gia Công giáo nên mọi người ở đây thường sẽ đi Lễ vào chủ nhật.



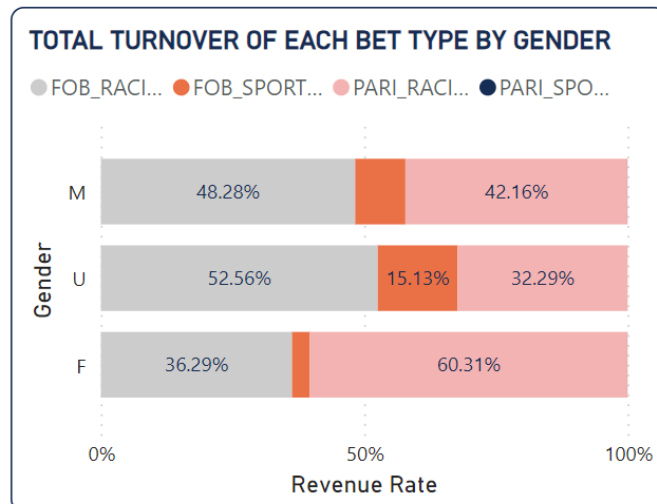
Hình 18. Doanh thu của các loại hình cá cược theo nhóm tuổi

Về nhóm tuổi, khoảng cách về doanh thu của cá cược fixed-odds racing và pari-mutuel racing dần rút ngắn khi nhóm tuổi tăng lên đến nhóm 55 tuổi trở lên, doanh thu của pari-mutuel đã vượt qua fixed-odds. Điều này cho thấy những người lớn tuổi thích mạo hiểm hơn vì cược theo hình thức pari-mutuel, người chơi sẽ không biết số tiền mình được chia khi họ thắng và mất đi nếu bị thua. Đối với thể loại fixed-odds sports, nhóm tuổi 35-44 chơi đóng góp nhiều nhất.



Hình 19. Doanh thu của các loại hình cá cược theo bang

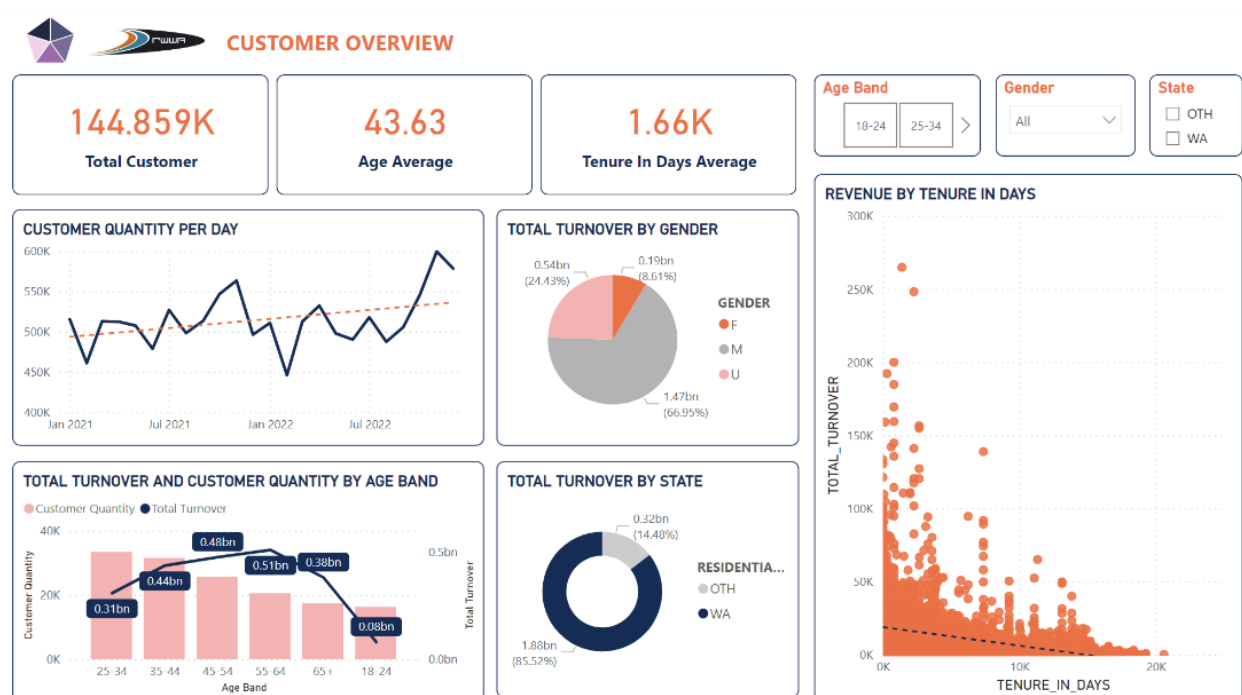
Về khu vực, có thể nhận thấy các khách hàng ở khu vực phía Tây thường chọn thể loại cá cược racing ở cả 2 hình thức, trong khi khách hàng đến từ khu vực khác thì thường chọn cá cược cược theo hình thức fixed-odds ở 2 thể loại racing và sports.



Hình 20. Tỷ lệ đóng góp của từng loại hình cá cược vào doanh thu theo giới tính

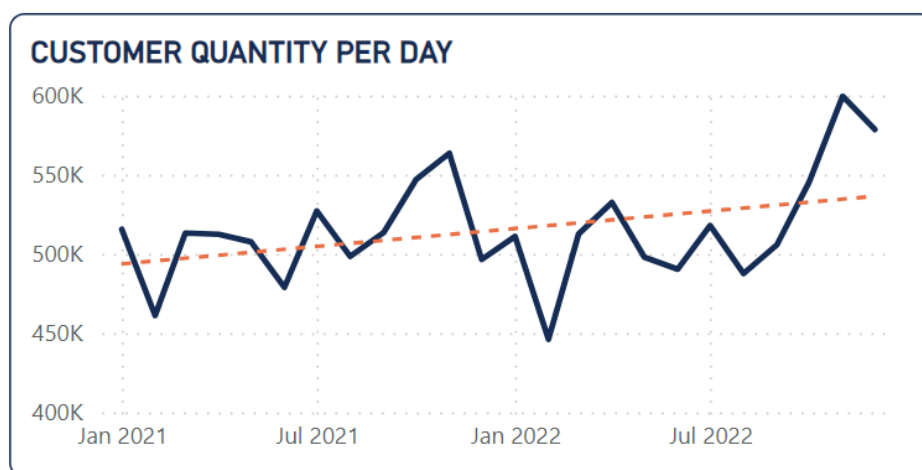
Về độ yêu thích của các giới tính đối với từng loại hình cá cược, giới tính không xác định thường chọn cá cược fixed-odds racing và nữ giới thường chọn cá cược pari-mutuel racing. Sự phân bố doanh thu đối với các hình thức cá cược racing ở giới tính nam khá đồng đều. Ngoài ra giới tính không xác định cũng cá cược vào thể loại sports nhiều hơn so với 2 giới tính còn lại.

Phân tích khách hàng



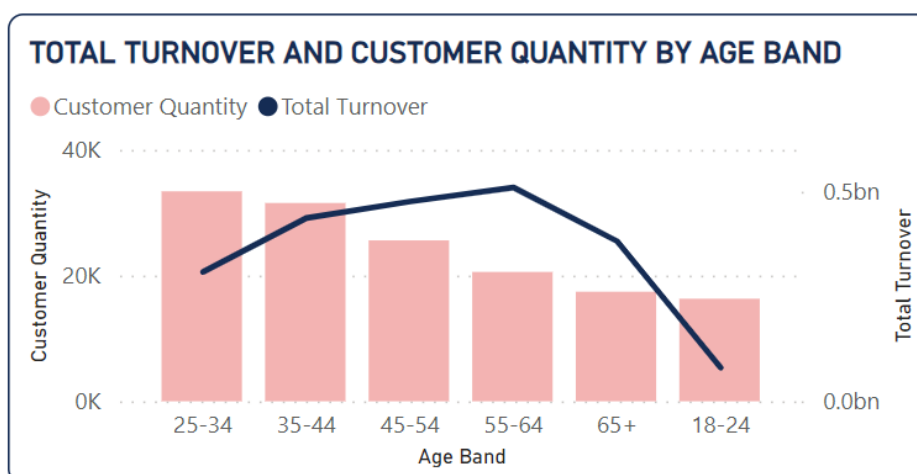
Hình 21. Tổng quan khách hàng của WA TAB

Biểu đồ tổng quan về khách hàng cung cấp thông tin về tổng số khách của doanh nghiệp trong 2 năm (2021-2022) là gần 145,000 khách hàng với độ tuổi trung bình là khoảng 44 tuổi. Số ngày tính từ ngày khách hàng mở tài khoản đến ngày đặt cược trung bình là 1,660 ngày.



Hình 22. Số người tham gia cá cược theo thời gian

Trong giai đoạn phân tích, số khách hàng tham gia cá cược trong 1 ngày có xu hướng tăng lên. Tháng 11 hàng năm số lượng người tham gia cá cược trong tháng tăng cao và ngược lại tháng 2 có số lượng người tham gia cá cược trong tháng giảm thấp nhất trong năm. Đối với số người chơi tăng cao trong tháng 11 được giải thích qua sự kiện Melbourne Cup. Các tháng có số người chơi trong 1 ngày giảm, chẳng hạn như tháng 12 đến tháng 1 có thể do đây là mùa cao điểm du lịch tại đất nước này và mọi người sẽ có xu hướng đi chơi xa thay vì tham gia vào các trò chơi giải trí.

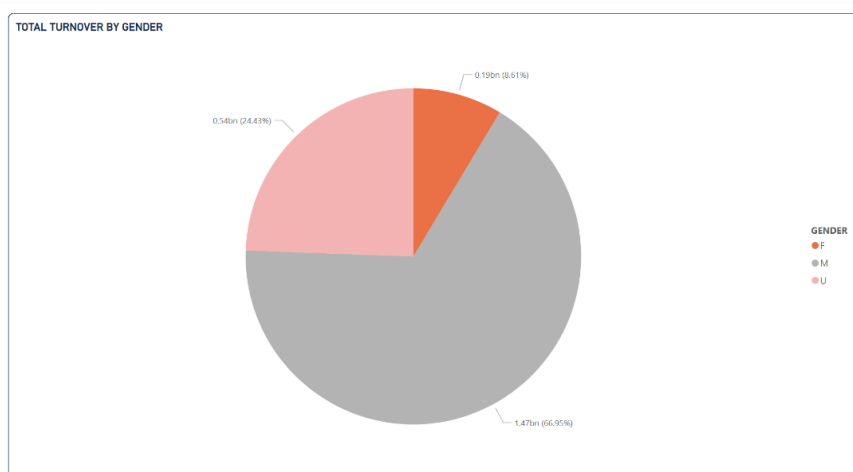


Hình 23. Số khách hàng và doanh thu của từng nhóm tuổi

Đối với biểu đồ các nhóm tuổi của khách hàng, nhóm tuổi tập trung nhiều khách hàng nhất là 25-34 tuổi và thấp nhất là 18-24 tuổi. Tuy nhiên nhóm tuổi đem lại doanh thu nhiều nhất là nhóm 55-

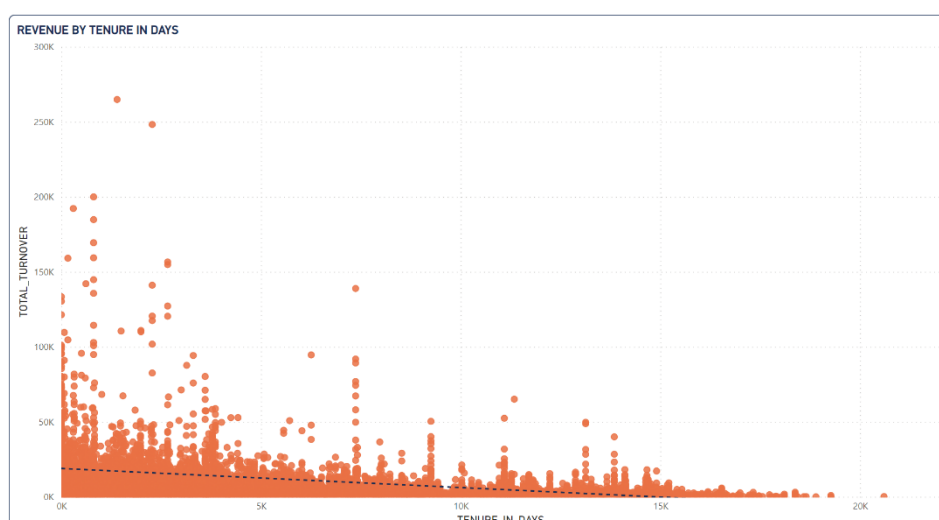
64 tuổi (23.25%), theo sau là 45-54 tuổi (21.72%) , 35-44 tuổi (19.93%) và nhóm 65+ (17.4%). Nhóm đóng góp thấp hơn là 25-34 tuổi (14.06%) và 18-24 tuổi (3.64%)

Mặc dù, giới tính không xác định có số lượng khách hàng nhiều nhất nhưng nam giới lại đóng góp lớn nhất vào doanh thu của doanh nghiệp với gần ⅔ trên tổng doanh thu. Vì vậy cần phải xem xét lại về sự phân bố giới tính và điều tra lại về tập khách hàng có giới tính không xác định.



Hình 24. Doanh thu của các loại hình cá cược theo giới tính

Giới tính không xác định có số lượng người tham gia hơn 50% số lượng khách hàng trong 2 năm của WA TAB nhưng nhóm người chơi giới tính nam lại đóng góp nhiều nhất vào doanh thu với tỷ lệ là 66.95%. Điều này cho thấy doanh nghiệp đang không hiệu quả trong việc thu hút doanh thu từ nhóm khách hàng đông nhất. Vì vậy, doanh nghiệp cần phải tiến hành điều tra nguyên nhân dẫn đến tình trạng này và nghiên cứu sâu hơn về tập khách hàng của mình để đưa ra những hành động kịp thời và hiệu quả cho việc cân bằng doanh thu theo giới tính người chơi.



Hình 25. Doanh thu theo tenure in days

Biểu đồ REVENUE BY TENURE IN DAYS cho thấy khách hàng có số ngày từ ngày mở tài khoản tính đến ngày đặt cược càng lớn thì khả năng họ đem lại doanh thu càng thấp. Và ngược lại, nếu ngày đặt cược gần ngày mở tài khoản thì giá trị doanh thu họ mang lại càng cao. Doanh nghiệp cần phải có những chiến dịch cụ thể để tối thiểu tenure in days của mỗi khách hàng để giúp khả năng tham gia cá cược của và đem đến doanh thu của khách hàng tăng lên.

Với những đặc điểm trên, kết hợp với việc xác định các chỉ số RFM cho từng khách hàng như sau:

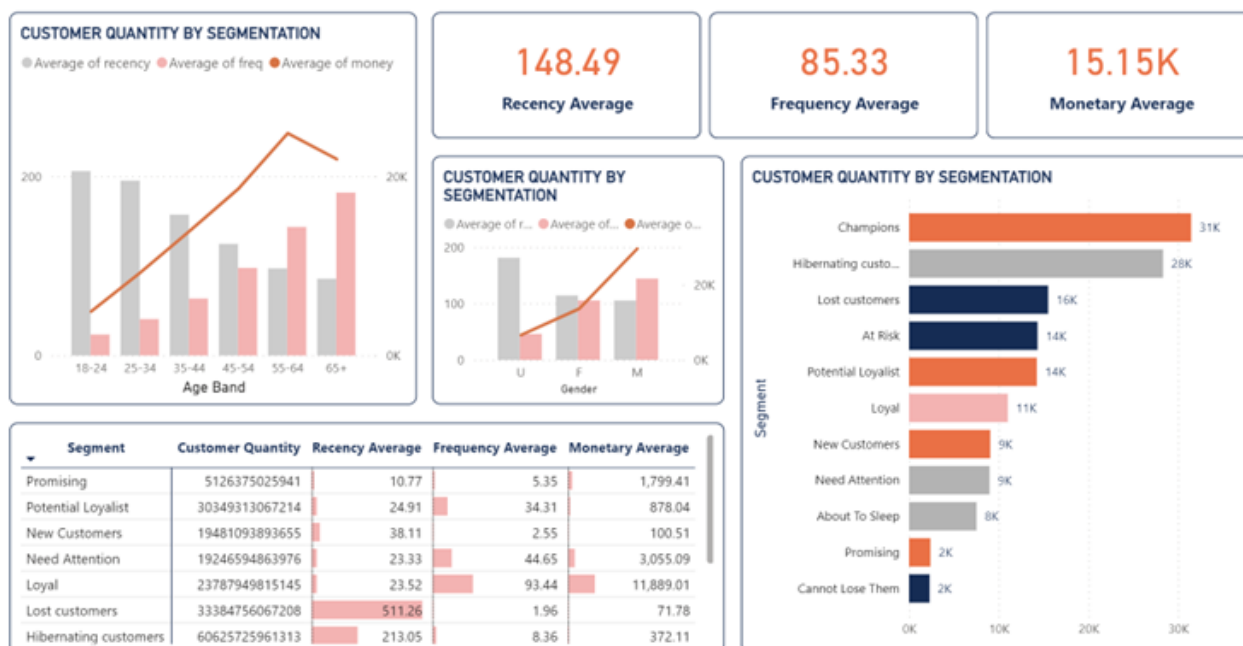
Bảng 3. Bảng kết quả RFM

BET_ACCOUNT_NUM_HASH	recency	money	freq
13154	1.0	366186.08	53.0
18379	1.0	32033.00	199.0
37251	61.0	100.00	1.0
129273	13.0	155.00	16.0
348318	99.0	30.00	4.0

Bảng 4. Bảng thống kê mô tả kết quả RFM

	BET ACCOUNT NUM HASH	recency	money	freq
count	1.448590e+05	1.448590e+05	1.448590e+05	1.448590e+05
mean	2.143528e+09	148.494936	1.515111e+04	85.332296
std	1.239167e+09	201.312452	6.188234e+04	142.057642
min	1.315400e+04	1.000000	0.000000e+00	1.000000
25%	1.071472e+09	5.000000	1.630000e+02	3.000000
50%	2.145177e+09	57.000000	9.920000e+02	18.000000
75%	3.211517e+09	246.000000	7.066750e+03	98.000000
max	4.294965e+09	730.000000	5.544861e+06	730.000000

Sau khi tính toán 3 giá trị R, F và M dựa trên dữ liệu thu được của khách hàng, recency ghi nhận khách hàng gần nhất là đặt cược cách đây 1 ngày và xa nhất là 730 ngày, 50% khách hàng đã gần 2 tháng chưa quay lại đặt cược. Bên cạnh đó, chỉ số Frequency thể hiện 25% khách hàng chỉ mới cá cược 3 lần, quá thấp so với mức trung bình là 85 lần. Xét về chi tiêu của các khách hàng, trung bình tổng chi tiêu rơi vào khoảng 15,151 và khách hàng chi tiêu nhiều nhất lên đến hơn 5 triệu AUD. Từ kết quả RFM, các nhóm khách hàng có thể được phân loại thành 11 nhóm với 4 phân khúc.



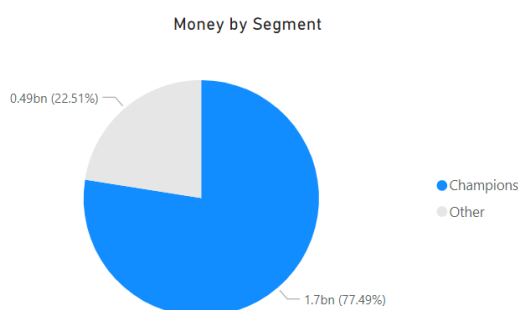
Hình 26. Phân nhóm khách hàng dựa trên mô hình RFM

○ Phân khúc khách hàng VIP

- Gồm những khách hàng thuộc nhóm Champion.
- Đặc điểm:

Những khách hàng thường xuyên cá cược gần đây nhất, có tần suất cá cược cao nhất và có giá trị tiêu dùng lớn nhất. Họ không chỉ cá cược thường xuyên mà còn chi tiêu nhiều tiền cho các giao dịch.

Lượng khách Champion rất đông, dẫn đầu về số lượng khách hàng, chiếm khoảng 20% lượng khách.



Hình 27. Doanh thu theo phân khúc khách hàng

Không những thế, nhóm này mua với lượng lớn tiền, lượng tiền trung bình lên tới 54.000 AUD, rất cao so với các nhóm khác. Các chỉ số Recency, Frequency cũng tốt nhất, thậm chí vượt trội hơn tất cả. Doanh thu của nhóm này đem lại lên tới hơn 77%, doanh nghiệp thực sự có một nhóm khách hàng rất chất lượng và chịu chi. Tỷ lệ 20/80 này thực sự rất ấn tượng.

- Phân khúc khách hàng trung thành, tiềm năng trung thành

- Gồm những khách hàng thuộc nhóm Loyal và Potential Loyalist.
- Đặc điểm:

Khách hàng thuộc phân khúc này những khách hàng thường xuyên cá cược, có giao dịch trong 1 tháng trở lại đây và mức chi tương đối lớn.

Nhóm khách hàng tiềm năng trung thành là những khách hàng cũng có giao dịch trong 1 tháng trở lại, chi tiêu trung bình khá và tần suất ghi nhận cũng khá nhiều.

- Phân khúc khách hàng cần quan tâm và có nguy cơ

- Gồm những khách hàng thuộc 5 nhóm Need Attention, Promising, New Customers, At Risk và Can't Lose Them.
- Đặc điểm:

Need Attention là những khách hàng có tần suất cá cược ở mức khá, mua gần đây, tuy nhiên lượng tiền bỏ ra không nhiều. Promising là những khách hàng có lần cá cược gần đây nhưng số lần cá cược ít. Hai nhóm trên cho thấy những giao dịch đầu có kết quả thắng cao. Về mặt tích cực rằng họ sẽ trở lại cá cược tiếp, nhưng họ phải duy trì mức thắng cao vì vốn dĩ họ là những người đang khảo sát, số tiền họ chi cũng không quá nhiều nhưng cũng khá hứa hẹn.

New Customers là những khách hàng mới cá cược gần đây nhất, giá trị bỏ ra còn thấp và vẫn còn ít số lần chi tiêu.

At Risk là khách hàng tần suất cá cược và giá trị tiêu dùng ở mức khá tốt. Tuy nhiên những khách hàng đã khá lâu không quay lại, ghi nhận khoảng thời gian lên tới 9 tháng và họ đã từng có lịch sử cá cược rất thường xuyên với giá trị giỏ hàng ở mức khá trở lên.

Can't Lose Them là những khách hàng đã rất lâu không quay lại và từng cá cược thường xuyên, với giá trị cá cược lớn. Doanh nghiệp có thể đánh mất những khách hàng này nếu không có hoạt động kích thích họ quay lại.

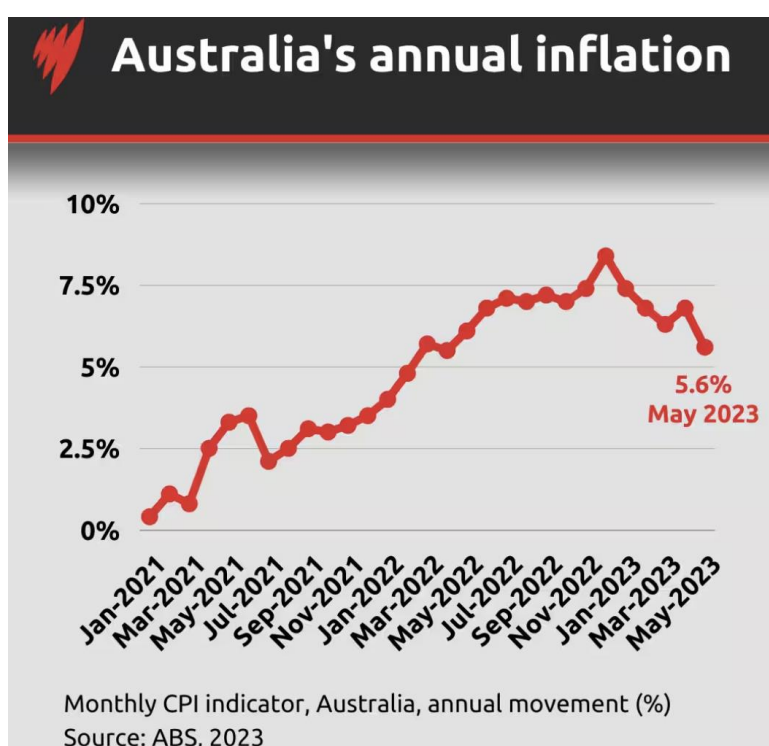
- Phân khúc khách hàng phổ thông

- Gồm những khách hàng thuộc nhóm Hibernating, Lost và About to sleep.
- Đặc điểm:

Khách hàng thuộc phân khúc này là những khách hàng đã rất lâu không quay lại, tần suất mua và giá trị giỏ hàng cũng rất thấp. Nhóm này thường là những khách hàng có hành vi mua tìm kiếm sự đa dạng hoặc chỉ mua suy nhất một lần để trải nghiệm và so sánh với các sản phẩm/dịch vụ khác. Họ là những khách hàng đã khá lâu chưa cá cược, trước đó cá cược với tần suất thấp và giá trị thấp. Tỷ lệ thắng cược của 3 nhóm này cũng rất thấp so với mặt bằng chung.

Đánh giá vấn đề của doanh nghiệp

So sánh tình hình kinh doanh của WA TAB năm 2022 có sự suy giảm so với năm 2021. Một trong những nguyên nhân có thể gây ra điều này là do sự giảm sút trong tình hình kinh doanh sau ảnh hưởng tiêu cực từ đại dịch Covid-19. Cùng với đó là sự lạm phát tăng cao từ năm 2021 đến hết năm 2022 tại Úc. Nền kinh tế trải qua giai đoạn suy thoái hoặc giảm tốc độ tăng trưởng trong năm 2022 sau khi đã phục hồi từ đại dịch Covid-19. Ngoài ra các yếu tố khác như tăng giá năng lượng và nguyên liệu cũng có thể làm tăng chi phí kinh doanh và giảm lợi nhuận của WA TAB.



Hình 28. Tỷ lệ lạm phát ở Úc 2021 - 2023 (Nguồn: SBSNews)

WA TAB có những điểm mạnh rõ ràng, trong đó có sự tăng trưởng của doanh thu qua các năm ổn định. WA TAB xác định được điểm mạnh và phát triển điểm mạnh về các hình thức cá cược hiện có của doanh nghiệp. Hoạt động cá cược sự kiện đua theo hình thức fix-odds và pari-mutuel đóng vai trò quan trọng trong việc tạo ra doanh thu chính và mạnh mẽ cho công ty. Tuy nhiên, sự mất cân bằng trong cơ cấu sản phẩm khi quá phụ thuộc vào hai hình thức cá cược sự kiện đua fix-odds

và pari-mutuel trong khi các hình thức cá cược thể thao fix-odds và pari-mutuel chiếm tỷ lệ rất nhỏ vừa là điểm mạnh vừa là vấn đề mà doanh nghiệp cần phải cải thiện, biện pháp khắc phục sẽ được cập nhật trong phần tiếp theo.

Bên cạnh đó, WA TAB cũng phải đối mặt với một số điểm yếu và thách thức. Một trong những thách thức lớn nhất mà WA TAB đang phải đối diện đó là tính thời vụ cao của hoạt động kinh doanh. Đặc điểm của tính thời vụ cao là sự biến động mạnh về doanh thu theo chu kỳ tuần, với doanh số cao vào thứ 7 mỗi tuần. Tính thời vụ cao làm cho doanh thu trở nên không ổn định, gây khó khăn trong việc phân bổ và quản lý hiệu quả nguồn lực. Cụ thể, vào ngày cao điểm có thể dẫn đến quá tải, thiếu hụt nguồn cung, ảnh hưởng chất lượng dịch vụ. Vào ngày thấp điểm, doanh thu không cao nhưng vẫn phải duy trì hệ thống đường đua, mặt bằng cỏ và bảo trì trang thiết bị. Đồng thời, WA TAB cũng phải chi trả chi phí thức ăn, chăm sóc sức khỏe cho ngựa và chó dẫn đến mất cân bằng trong việc quản lý, kiểm soát và tối ưu doanh thu. Thứ hai, sự phụ thuộc vào nhóm khách hàng trung thành và các hình thức cá cược truyền thống cũng giới hạn khả năng tiếp cận với các phân khúc thị trường mới. Đáng chú ý là trong 3 nhóm khách hàng chiếm số lượng cao nhất trong các phân khúc khách hàng hiện có, chỉ có 1 nhóm thuộc phân khúc khách hàng trung thành, trong khi 2 nhóm còn lại thuộc phân khúc khách hàng có nguy cơ rời bỏ doanh nghiệp cao. Điều này đặt ra câu hỏi đáng quan tâm về hiệu quả của các chiến lược giữ chân khách hàng hiện tại của doanh nghiệp. Ngoài ra, khoảng cách giữa doanh thu và lợi nhuận khá lớn chứng tỏ hiệu quả hoạt động của các doanh nghiệp trong ngành này đang gặp phải nhiều khó khăn. Có thể do chi phí hoạt động quá cao hoặc tỷ lệ hoàn trả khách hàng quá cao, dẫn đến việc lợi nhuận bị ảnh hưởng nghiêm trọng.

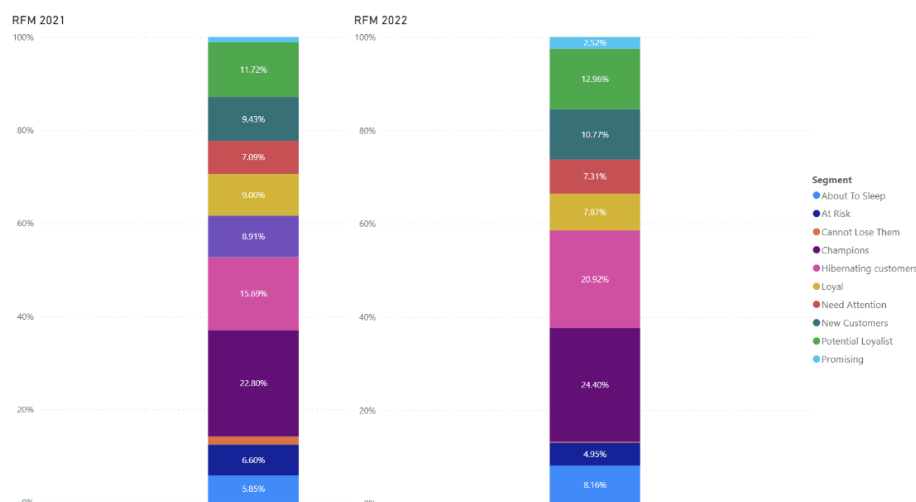
Với những ưu điểm và hạn chế mà doanh nghiệp đang có các vấn đề đặt ra yêu cầu doanh nghiệp cần giải quyết:

- Doanh thu phụ thuộc vào nhóm khách hàng trung thành và các hình thức cá cược của nhóm (Doanh thu phụ thuộc vào nhóm khách hàng trung niên - cao tuổi).
- Tính thời vụ cao của hoạt động kinh doanh.
- Tỷ lệ tăng trưởng doanh thu chưa ổn định.
- Sự mất cân bằng trong cơ cấu sản phẩm (các loại hình cá cược).

II. ĐỀ XUẤT GIẢI PHÁP CẢI THIỆN

Vấn đề 1: Doanh thu phụ thuộc vào nhóm khách hàng trung thành và các hình thức cá cược của nhóm (Doanh thu phụ thuộc vào nhóm khách hàng trung niên - cao tuổi).

Sử dụng mô hình RFM để phân khúc khách hàng theo 11 nhóm hành vi. Từ đó sẽ làm rõ sự chuyển hóa phân khúc giữa 2 năm hoạt động và tìm được mối quan hệ giữa sự phân bổ doanh thu đối với các nhóm khách hàng chiếm tỷ trọng cao trong doanh nghiệp để có thể tiếp thị hiệu quả và tối ưu tỷ lệ giữ chân khách hàng.



Hình 29. Tỷ lệ khách hàng trong nhóm qua 2 năm

Nhóm khách hàng đã không sử dụng sản phẩm hoặc dịch vụ trong một thời gian dài và có thể đang xem xét các đối thủ chiếm tỷ trọng lớn thứ 2. Cho thấy việc cải thiện khách hàng Hibernating Customers là cấp thiết. Bên cạnh đó, nhóm khách hàng đóng góp nhiều nhất cho doanh nghiệp cũng là nhóm chiếm tỷ trọng cao nhất. Tuy nhiên, qua sơ đồ chuyển hóa khách hàng trong 2 năm có thể thấy rằng tỷ trọng nhóm Champions không có sự thay đổi đáng kể trong khi các nhóm tiêu cực lại có sự gia tăng cao trong phần trăm phân bổ. Điều này có thể đặt giả thiết rằng doanh nghiệp chưa thật sự làm tốt trong công cuộc chuyển đổi khách hàng tiềm năng.

Vì thế, nếu doanh nghiệp muốn tăng tỷ lệ chuyển đổi nhanh chóng thì nên tập trung đầu tiên vào việc cải thiện Hibernating, tìm kiếm nguyên nhân rời đi của nhóm Lost Customer và duy trì Champions, cần triển khai các giai đoạn giải pháp song song trong ngắn hạn như sau:

- **Cải thiện nhóm khách hàng đã bị bỏ lỡ Hibernating Customer và tìm nguyên nhân rời đi của nhóm Lost Customer.**

Đối với vấn đề này, WA TAB cần thu thập thêm các thông tin về hành trình của khách hàng, các hoạt động mà họ tham gia cũng như khảo sát về độ hài lòng của họ đối với các dịch vụ. Từ đó, dùng thuật toán để tìm hiểu lý do vì sao họ từng mua hàng và tương tác tích cực với công ty, cố gắng kết nối lại với họ. Cung cấp các dịch vụ chăm sóc khách hàng tốt hơn để tăng độ hài lòng của họ, khuyến khích họ mua hàng tại công ty trở lại. Bên cạnh đó, rất khó để khiến nhóm Lost Customer quay trở lại, thay vào đó cần tìm hiểu, khai thác dữ liệu

từ nhóm này để xác định nguyên nhân rời đi của họ. Từ đó, có những hành động ngăn ngừa sự rời bỏ của các nhóm khách hàng thuộc top nguy cơ.

○ **Duy trì doanh thu từ nhóm Champions.**

Nhóm Champions hiện đang là nhóm khách hàng quan trọng bậc nhất và cũng là nhóm chiếm tỉ trọng cao nhất của doanh nghiệp, họ tạo ra doanh thu lớn và duy trì điều đó đều đặn. Thay vì công ty đi tìm kiếm khách hàng mới mỗi ngày thì việc giữ chân nhóm khách hàng này là cấp thiết hơn. Một số đề xuất có thể kể đến như đẩy mạnh sản phẩm, dịch vụ mà họ hứng thú - cá cược đua ngựa bằng cách đổi mới và phát triển nó như 1 hình thức mới. Điều này sẽ được đề cập ở phần sau.

Ngoài ra, trong dài hạn để có sự chuyển đổi tốt trong các nhóm phân khúc khách hàng, doanh nghiệp cần xem xét thực hiện các giai đoạn chiến dịch sau:

Bảng 5. Kế hoạch thu hút và giữ chân cho từng phân khúc khách hàng

Giai đoạn	The Winner's Circle (VIP Champion)	Wake The Champion (Loyal & Potential Loyalist)	Reignite The Passion (Need Attention & Promising...)	Return To Glory Phổ thông
Giai đoạn 1	- Thiết kế chương trình ưu đãi, hệ thống thẻ VIP - Khảo sát, đàm phán điểm (2 với đội ngũ chuyên gia tháng)	- Xác định cấu trúc, ưu đãi cho chương trình tích điểm/khuyến mãi - Xây dựng chiến lược - Nghiên cứu địa điểm, quảng cáo các chương trình concept cho sự kiện VIP trình.	- Lên danh sách sự kiện/khuyến mãi làm chiến dịch quảng nội dung nhắc nhở - Lên ý tưởng thiết kế, đồng trực tuyến	- Lên kế hoạch chiến dịch quảng cáo tiếp cận - Lập bản khảo sát nhân viên không hài lòng
Giai đoạn 2	- Hoàn thiện thiết kế, hệ thống thẻ VIP sẵn sàng triển khai - Hoàn tất hợp đồng, lên trình tích điểm (2 kế hoạch dịch vụ tư vấn tháng)	- Thiết kế giao diện, hoàn thiện tính năng chương trình - Sản xuất nội dung quảng năng đầu tiên cho cộng đồng - Chuẩn bị nội dung, cáo, hình ảnh cho các chương trình sự kiện VIP chiến dịch Quý 1	- Thiết kế khuyến mãi, quảng cáo cho nội dung nhắc nhở - Triển khai các tính năng đầu tiên cho cộng đồng	- Sản xuất tài liệu chiến dịch tiếp - Chuẩn bị ưu đãi, khuyến mãi thu hút khách hàng

Giai đoạn	The Winner's Circle (VIP Champion)	Wake The Champion (Loyal & Potential Loyalist)	Reignite The Passion (Need Attention & Promising...)	Return To Glory Phổ thông
Giai đoạn 3 (2 vấn cá nhân tháng)	- Triển khai chương trình, cấp thẻ VIP - Khởi động dịch vụ tư vấn cá nhân - Gửi thư mời, thông báo sự kiện VIP Quý 1	- Thử nghiệm chương trình tích điểm với nhóm khách hàng - Đăng tải quảng cáo khởi chạy các chiến dịch	- Bắt đầu gửi nội dung chạy chiến dịch nhắc nhở, khuyến khích mãi mãi - Mở cộng đồng, mời khách hàng tham gia	- Đăng tải và khởi chạy chiến dịch - Triển khai đãi thu hút khách hàng

Vấn đề 2: Tối ưu tính thời vụ trong xu hướng doanh thu

Trong bối cảnh kinh doanh ngày càng cạnh tranh, việc sử dụng công nghệ để tối ưu hóa quyết định kinh doanh là yếu tố then chốt giúp doanh nghiệp phát triển vượt bậc trên thị trường. Đặc biệt dự đoán và phân tích dữ liệu sẽ giúp doanh nghiệp đưa ra các quyết định kinh doanh một cách chính xác hơn. Điều này có thể giúp tối ưu hóa tính thời vụ trong xu hướng doanh thu và chiến lược kinh doanh, tăng cường hiệu suất và tăng trưởng kinh doanh. Vì vậy, chúng tôi đề xuất sử dụng mô hình phân tích dữ liệu chuỗi thời gian truyền thống ARIMA và mô hình học sâu RNN, cụ thể là LSTM (Long Short-Term Memory) và GRU (Gated Recurrent Unit) để dự đoán và phân tích dữ liệu doanh thu trong 3 tháng tiếp theo cho doanh nghiệp đối với tập dữ liệu này. Với những lý do sau:

- **Đặc điểm của tập dữ liệu:** Với tập dữ liệu chỉ ghi nhận trong khoảng thời gian 2 năm cùng với tính thời vụ cao khi biến động lên xuống có chu kỳ vào những dịp đặc biệt và có sự khác biệt với năm tài chính (bắt đầu từ đầu tháng 8 và kết thúc vào cuối tháng 7 năm sau - nguồn báo cáo tài chính do WA TAB cung cấp). Dự báo ngắn hạn sẽ là sự lựa chọn chính xác và an toàn hơn cho tổ chức khi đưa ra quyết định kinh doanh hiệu quả trong thời gian ngắn, ví dụ như điều chỉnh tỷ lệ cược, chiến lược marketing.
- **Đặc điểm ngành:** Doanh thu của ngành cá cược đua có thể thay đổi mạnh theo từng giải đua, phong độ ngựa, tâm lý người chơi. Ngoài ra, cũng rất khó dự đoán chính xác trong dài hạn do chịu ảnh hưởng bởi nhiều yếu tố như thời tiết, tình hình sức khỏe ngựa, chiến thuật thi đấu.

Mô hình LSTM và GRU có khả năng học các mối quan hệ phức tạp và dài hạn trong dữ liệu chuỗi thời gian, điều mà các mô hình truyền thống như ARIMA có thể gặp khó khăn. LSTM và GRU cũng có khả năng xử lý dữ liệu chuỗi thời gian không tuyến tính và không ổn định, điều này rất quan

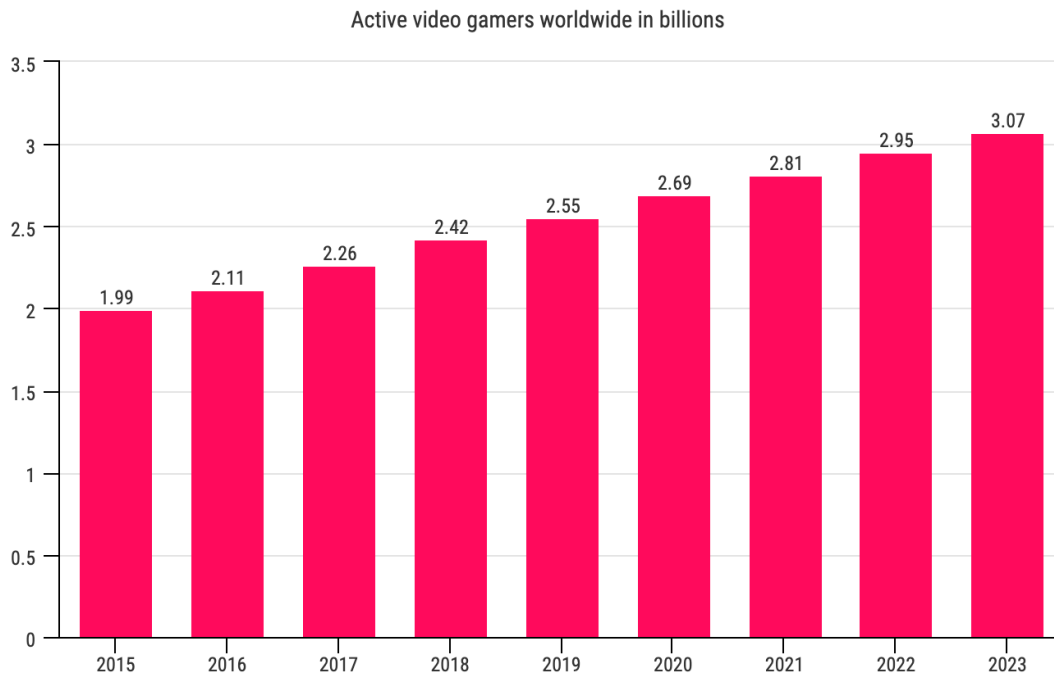
trọng đối với dữ liệu thực tế. Ngoài ra, LSTM và GRU có thể tự động học và nhớ các đặc trưng từ dữ liệu, giảm bớt công việc tiền xử lý dữ liệu đặc biệt trong việc cân nhắc loại bỏ các giá trị ngoại lai và lựa chọn đặc trưng. Điều này không chỉ giúp tiết kiệm thời gian và công sức, mà còn có thể dẫn đến kết quả dự đoán chính xác hơn.

Vấn đề 3: Hành động tiếp theo của doanh nghiệp sau khi dự đoán được doanh thu trong tương lai. Sau khi xây dựng và lựa chọn mô hình cho ra kết quả dự đoán doanh thu tốt nhất, kết quả dự đoán được chia theo các trường hợp sau:

- Trường hợp 1: Giá trị dự đoán doanh thu có xu hướng tăng hoặc ổn định.
- Trường hợp 2: Giá trị dự đoán doanh thu có xu hướng giảm.

Với 3 trường hợp kết quả có thể xảy ra, chúng tôi đề xuất hành động cải thiện như sau:

Trường hợp 1: Phát triển hình thức cá cược mới thu hút khách hàng. WA TAB có thể tận dụng nhu cầu ở thị trường ngành này để phát triển thêm một thể thức mới của cá cược đua ngựa - đua ngựa ảo NFT. Trong đó, mỗi con ngựa ảo đều được chứng thực và tồn tại dưới dạng NFT (Non-fungible token - token không thể thay thế). Về bản chất thì phương thức này dựa trên Blockchain nên vẫn đảm bảo được các đặc điểm về độ đáng tin cậy cũng như giao dịch thuận tiện. Với hình thức này, tổ chức có thể thực hiện tạo một nền tảng đua ngựa ảo ngay trên website của mình, một cộng đồng nơi người chơi có thể mua và lai tạo ngựa NFT, đặt cược vào các cuộc đua ngựa thời gian thực và tổ chức các cuộc đua để kiếm lợi nhuận thực tế. Bên cạnh đó, những nỗ lực số hóa môn đua ngựa trước đây đã được khán giả đón nhận nồng nhiệt: từ chương trình phát sóng cuộc đua trong các trang web đến hệ thống cá cược trực tuyến. Ngoài ra, theo dữ liệu khảo sát từ Statista, có hơn khoảng 40% dân số thế giới tham gia các trò chơi điện tử, và tỉ lệ này đang được tăng dần theo thời gian bởi vì sự tiện lợi của nó. Do đó, hình thức này vô cùng có tiềm năng để doanh nghiệp phát triển. Tuy nhiên, do chi phí phát sinh từ việc triển khai nền tảng cho hình thức này là khá cao và cần thời gian dài hạn nên tổ chức có thể cân nhắc hoạt động đua ngựa ảo này như 1 giải pháp phát triển trong dài hạn.



Hình 30. Biểu đồ số lượng người chơi trò chơi điện tử trên toàn thế giới từ năm 2015 đến năm 2023 (Đơn vị: Tỷ người)

WA TAB có thể hợp tác với các nhà cung cấp phần mềm uy tín như BetConstruct, SBTech, EveryMatrix để phát triển nền tảng đua ngựa ảo và đề ra đa dạng các thể loại cá cược như cược thắng, cược về đích trong top 3, cược thời gian hoàn thành. Vào những mùa cao điểm, có thể thu hút thêm người chơi bằng cách tổ chức giải đấu đua ngựa ảo với giải thưởng hấp dẫn hoặc cũng có thể dùng các giải đấu để tạo điểm nhấn, làm cân bằng lượng khách hàng vào những mùa thấp điểm doanh thu. Một trong những cách đó là vừa cung cấp dịch vụ mới vừa thay đổi chiến lược về giá trong thời điểm này để tạo cảm giác sinh lời cho người chơi. Một hình thức cá cược mới trên nền tảng như đua ngựa ảo cũng góp phần làm cải thiện trải nghiệm người dùng trực tuyến, tăng tính tò mò qua giao diện website và ứng dụng di động, tăng cường tính tương tác và cá nhân hóa dịch vụ.

Trường hợp 2: Doanh nghiệp nên sử dụng quy luật số lớn để quản lý rủi ro dựa trên xác suất thống kê. Xem xét lại nguyên nhân sâu xa dẫn tình hình kinh doanh của doanh nghiệp bị dự đoán là giảm sút, tăng cường quản lý và điều chỉnh các yếu tố ảnh hưởng đến doanh thu, như giảm chi phí hoặc tối ưu hóa chiến lược tiếp thị. Tuy nhiên, với quy luật số lớn trong cá cược thì doanh thu của doanh nghiệp chỉ có thể bị giảm nhẹ hoặc hầu như không giảm. Quy luật số lớn trong lĩnh vực cá cược là một nguyên lý trong xác suất và thống kê, cho rằng khi kích thước mẫu tăng lên, giá trị trung bình của nó sẽ gần với mức trung bình của toàn bộ tổng thể. Điều này áp dụng cho việc cá cược, khi số lượng cược tăng lên, kết quả trung bình sẽ tiệm cận giá trị xác suất lý thuyết. Điều này giúp dự đoán kết quả của các cược dựa trên xác suất thống kê và là cơ sở quan trọng để

quản lý rủi ro và đưa ra quyết định trong lĩnh vực cá cược. Ví dụ: khi một người chơi đánh bạc quyết định đặt cược trên một trò chơi có xác suất thắng là 50%. Nếu người chơi này tiến hành đặt cược một số lần đủ lớn, theo Quy luật số lớn, tỷ lệ chiến thắng trung bình sẽ tiệm cận với xác suất thực tế là 50%. Dù có những lần thua liên tiếp, nhưng khi số lần đặt cược tăng lên, kết quả trung bình sẽ phản ánh đúng xác suất lý thuyết, giúp người chơi quản lý rủi ro và đưa ra quyết định dựa trên xác suất thống kê.



Hình 31. Tỷ lệ thắng cược theo phân khúc khách hàng và trung bình

III. XÂY DỰNG VÀ ĐÁNH GIÁ MÔ HÌNH

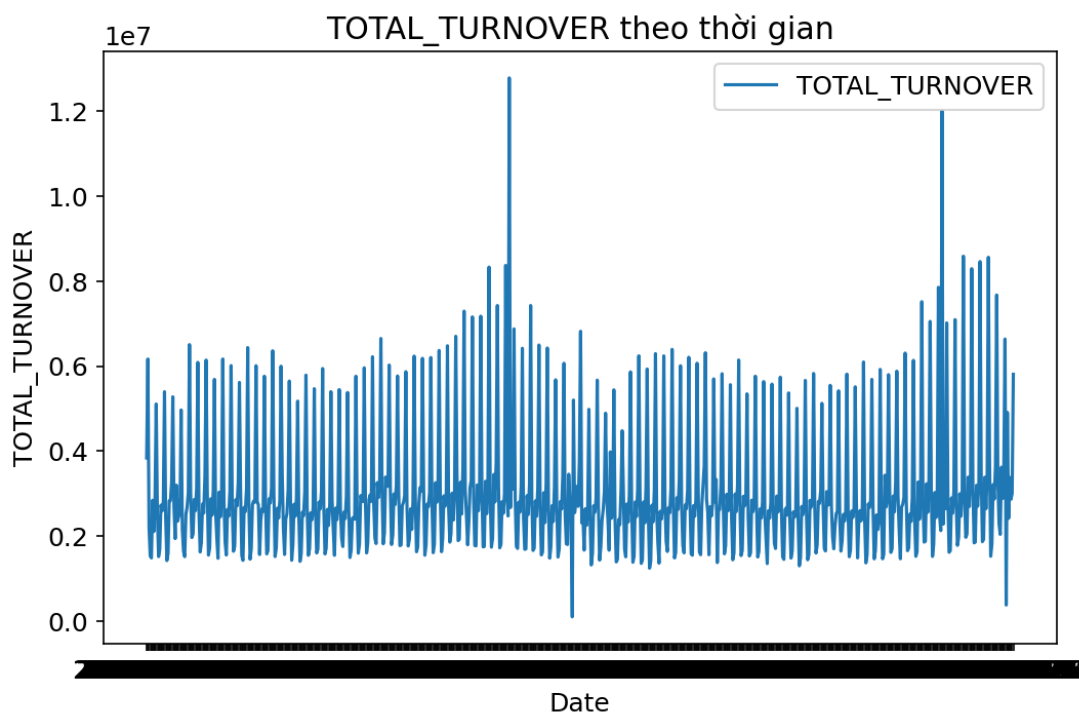
Data Preprocessing

Với mục tiêu xây dựng mô hình dự đoán doanh thu cho WA TAB. Từ dữ liệu ban đầu, nhóm quyết định chuyển đổi dữ liệu sang chuỗi thời gian để thuận tiện và tăng tính chính xác của mô hình. Tất cả dữ liệu giao dịch sẽ được gom nhóm và tính tổng lại theo ngày. Dữ liệu chuỗi thời gian sau khi được chuyển đổi có dạng như sau:

Bảng 6. Dữ liệu chuỗi thời gian sau khi chuyển đổi

DATE_DIM	FOB_RACING_TURNOVER	FOB_SPORT_TURNOVER	PARI_RACING_TURNOVER	PARI_SPORT_TURNOVER	TOTAL_TURNOVER	DIVIDENDS_PAID	GROSS_MARGIN	TICKETS
1/1/2021	1740318.55	160299.45	1934935.51	0	3835553.51	3114653.42	6.122.205.292	488231
1/2/2021	3329675.95	257613.55	2578252.53	2	6165544.03	4950145.3	1108986.05	698207
1/3/2021	930583.79	331455.71	935813.69	0	2197853.19	1826918.53	3.290.953.996	280680
1/4/2021	586806.57	321699.93	614867.17	183.5	1523557.17	1287351.59	1.998.129.026	194823
1/5/2021	602362.79	224912.71	653355.82	147.5	1480778.82	1213737.41	2.517.490.252	201668

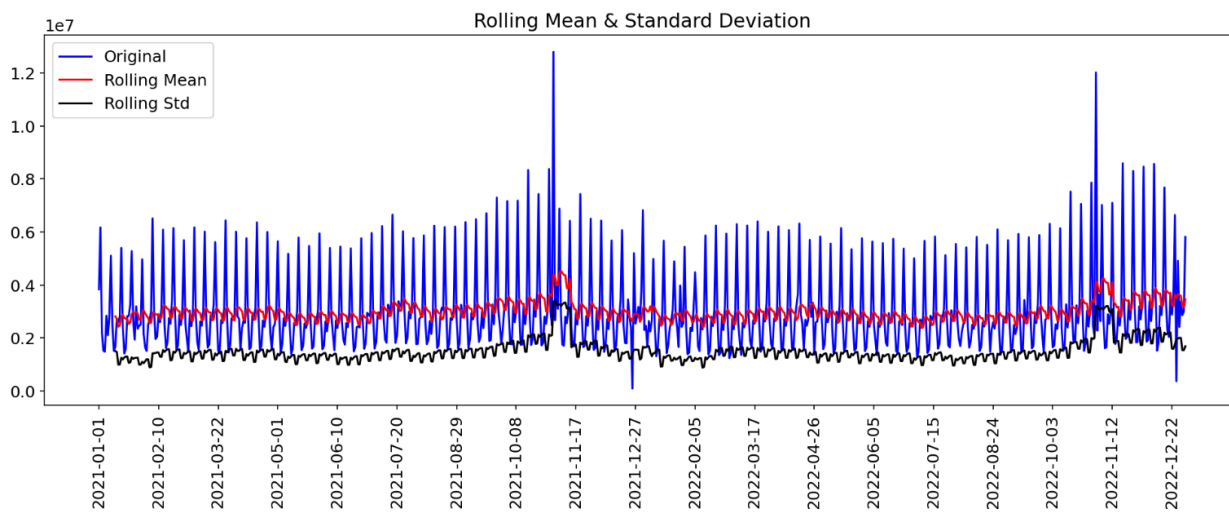
Dữ liệu gồm 720 dòng tương đương với 720 ngày từ 1/1/2021 đến 31/12/2022. Với mục tiêu dự đoán doanh thu vậy nên cột TOTAL_TURNOVER sẽ là đặc trưng được chọn để đưa vào mô hình dự đoán.



Hình 32. Doanh số theo thời gian

Mô hình ARIMA

Quá trình xây dựng mô hình



Hình 33. Biểu đồ dải trung bình và độ lệch chuẩn

Dựa vào đồ thị hình 33 có thể rút ra nhận xét là dữ liệu đã có tính dừng tuy nhiên để đảm bảo tính chính xác, nhóm thực hiện kiểm tra 2 kiểm định ADF và KPSS. Theo kết quả của hình bên dưới thì dữ liệu đã đầy đủ điều kiện để xác định là dữ liệu dừng.

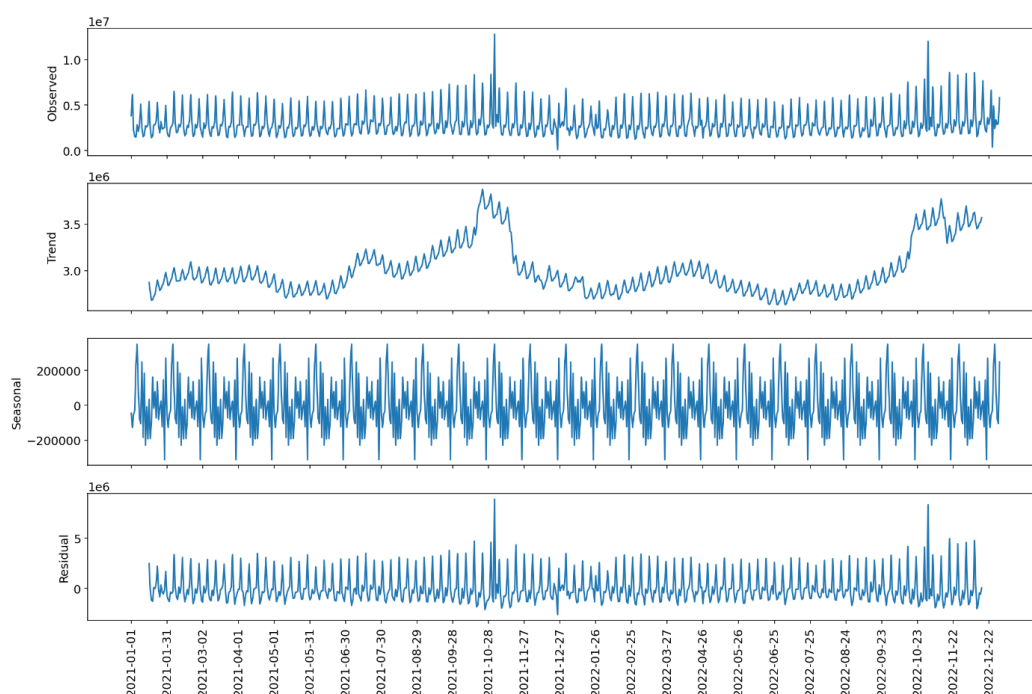
```

Augmented Dickey-Fuller Test:
Test Statistic      -2.758498
p-value             0.064451
No. Lags Used       20.000000
Number of Observations Used  709.000000
Critical Value (1%)  -3.439607
Critical Value (5%)  -2.865625
Critical Value (10%) -2.568945
dtype: float64
-----
KPSS Test:
Test Statistic      0.138584
p-value             0.100000
No. Lags Used       174.000000
Critical Value (10%) 0.347000
Critical Value (5%)  0.463000
Critical Value (2.5%) 0.574000
Critical Value (1%)  0.739000
dtype: float64

```

Hình 34. Kết quả kiểm định tính dừng

Kết quả theo kiểm định KPSS cho thấy chuỗi đã dừng tuy nhiên ADF cho thấy dữ liệu không dừng. Theo statsmodel trường hợp này dữ liệu có xu hướng dừng và cần loại bỏ xu hướng. Nhóm tiếp tục phân ra dữ liệu để có thể thấy xu hướng tăng, giảm của dữ liệu, cho thấy dữ liệu không có xu hướng vậy chuỗi thời gian đã đủ điều kiện để xác định tính dừng. Tiếp theo nhóm thực hiện phân rã dữ liệu để xem được các thành phần cấu thành nên chuỗi bao gồm: mùa vụ (seasonal), phần dư (residual) như hình sau:



Hình 35. Phân tích tính mùa vụ, xu hướng của dữ liệu doanh số

Dựa vào kết quả trên có tính mùa vụ, và chu kỳ tương ứng với 1 năm, doanh số thường sẽ tăng lên nhiều vào gia đoạn cuối tháng 10 đầu tháng 11.

Quá trình huấn luyện mô hình

Nhóm thực hiện xây dựng mô hình bằng cách chia dữ liệu làm 2 phần huấn luyện (train_data) và kiểm tra (test_data):

- Dữ liệu huấn luyện là chuỗi thời gian từ 1/1/2021 đến 19/10/2022.
- Dữ liệu kiểm tra là chuỗi thời gian từ 20/10/2022 đến 31/10/2022.

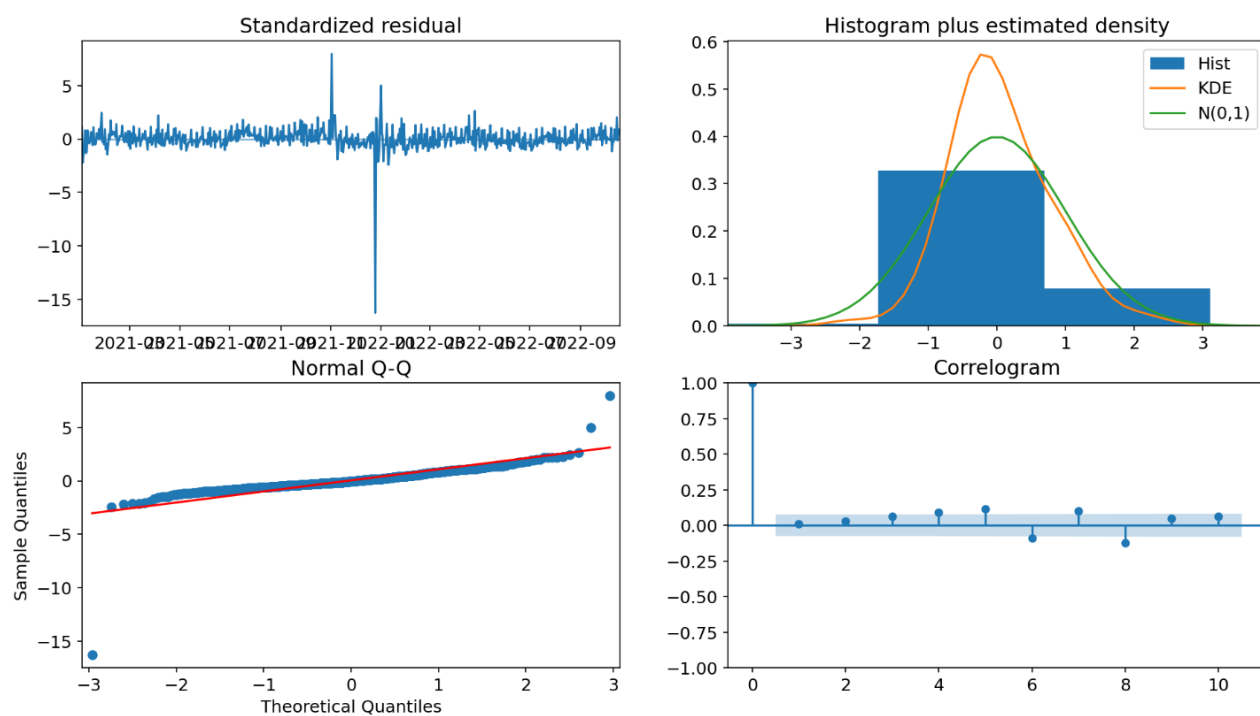
Áp dụng dữ liệu đã có vào phương thức auto_arima, có được mô hình tối ưu nhất như hình sau:

```
Best model: ARIMA(5,1,5)(0,0,0)[0] intercept
Total fit time: 42.697 seconds
```

SARIMAX Results						
=====						
Dep. Variable:	y	No. Observations:	657			
Model:	SARIMAX(5, 1, 5)	Log Likelihood	-41.588			
Date:	Sat, 23 Mar 2024	AIC	107.175			
Time:	12:07:09	BIC	161.009			
Sample:	01-01-2021	HQIC	128.047			
	- 10-19-2022					
Covariance Type:	opg					
=====						
	coef	std err	z	P> z	[0.025	0.975]

intercept	-0.0003	0.000	-0.820	0.412	-0.001	0.000
ar.L1	0.4958	0.037	13.508	0.000	0.424	0.568
ar.L2	-1.1873	0.030	-39.205	0.000	-1.247	-1.128
ar.L3	0.3796	0.048	7.925	0.000	0.286	0.473
ar.L4	-0.7334	0.030	-24.205	0.000	-0.793	-0.674
ar.L5	-0.2784	0.032	-8.780	0.000	-0.341	-0.216
ma.L1	-1.5192	0.030	-50.919	0.000	-1.578	-1.461
ma.L2	1.7224	0.043	40.347	0.000	1.639	1.806
ma.L3	-1.8020	0.033	-54.451	0.000	-1.867	-1.737
ma.L4	1.2897	0.038	34.101	0.000	1.216	1.364
ma.L5	-0.6724	0.025	-26.434	0.000	-0.722	-0.623
sigma2	0.0609	0.001	68.829	0.000	0.059	0.063
=====						
Ljung-Box (L1) (Q):	0.09	Jarque-Bera (JB):	262745.86			
Prob(Q):	0.77	Prob(JB):	0.00			
Heteroskedasticity (H):	0.92	Skew:	-4.97			
Prob(H) (two-sided):	0.54	Kurtosis:	100.54			
=====						

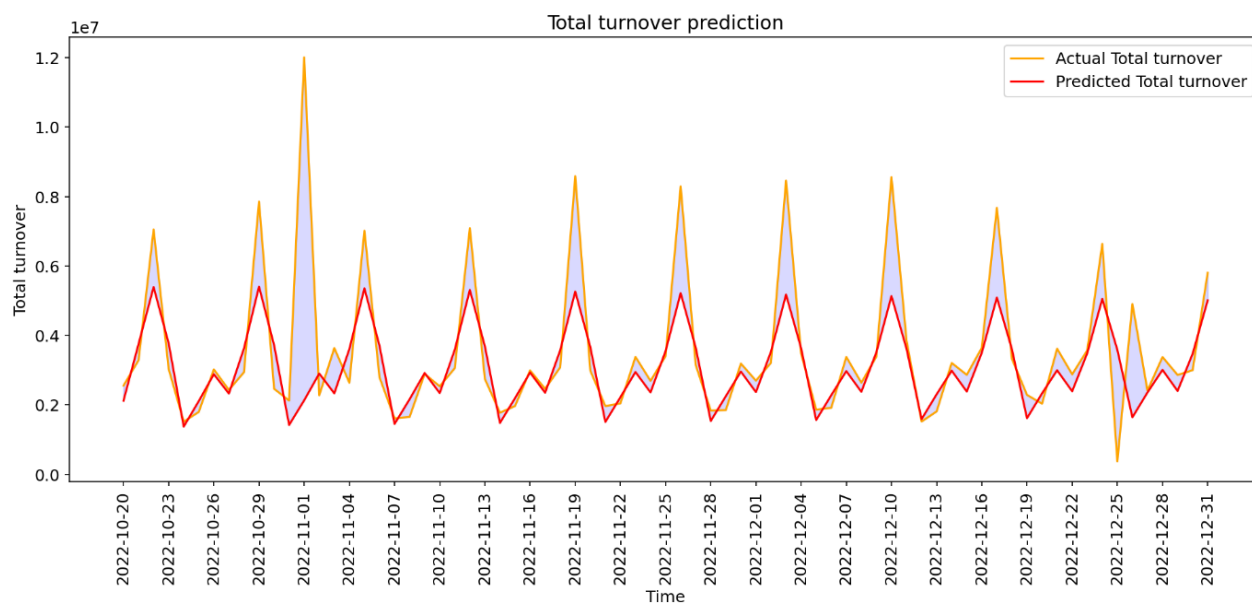
Hình 36. Mô hình ARIMA dự để dự đoán doanh số



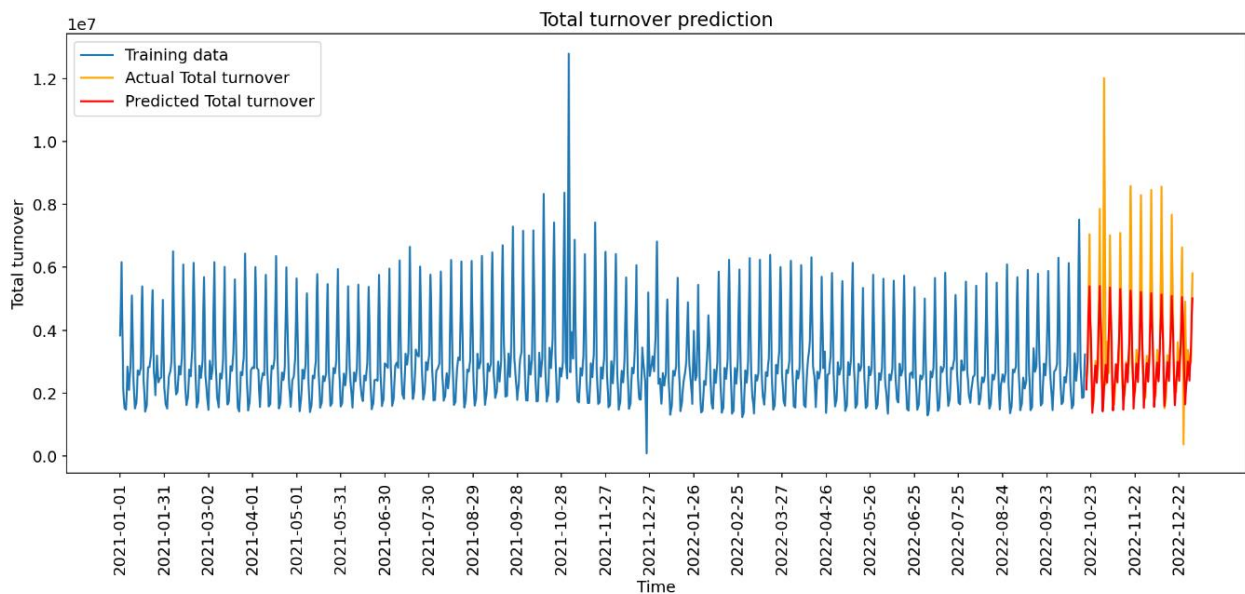
Hình 37. Đặc điểm của mô hình

Biểu đồ cho thấy đã tìm ra được tính mùa vụ, xu hướng phù hợp và đã loại bỏ được dữ liệu bị nhiễu, phần dư được phân phối chuẩn và có sự tương quan thấp với chính nó.

Đánh giá mô hình ARIMA



Hình 38. Kết quả dự đoán trên tập dữ liệu kiểm tra



Hình 39. Kết quả trực quan hóa phép vào toàn bộ dữ liệu

Bảng đánh giá các chỉ số của mô hình ARIMA:

Bảng 7. Bảng đánh giá các chỉ số của mô hình ARIMA

Metric	Value
MSE	0.180398
RMSE	0.424733
MAPE	1.692027
MAE	0.253592

Mô hình học máy RNN theo LSTM và GRU

Xây dựng mô hình và huấn luyện mô hình

Nhằm đảm bảo rằng các đặc trưng (features) trong dữ liệu có cùng phạm vi giá trị, giúp mô hình học máy hội tụ nhanh hơn và giảm thiểu ảnh hưởng của các giá trị ngoại lai (outliers), nhóm thực hiện quá trình chuẩn hóa dữ liệu với kỹ thuật MinMaxScaler. Điều này giúp dữ liệu chuỗi thời gian được chuyển đổi linh hoạt để nằm trong khoảng từ 0 đến 1.

$$X_{scaled} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Sau khi chuẩn hóa dữ liệu, chúng tôi thực hiện việc chia dữ liệu để huấn luyện (train) và kiểm tra (test):

- Dữ liệu huấn luyện là chuỗi thời gian từ 1/1/2021 đến 19/10/2022.
- Dữ liệu kiểm tra là chuỗi thời gian từ 20/10/2022 đến 31/10/2022

Xây dựng cấu trúc mô hình LSTM gồm 3 layer:

- Input Layer: Layer đầu tiên của mô hình là một layer LSTM với 50 units (được xác định bởi tham số đầu tiên trong LSTM(50)). Đây là lớp đầu tiên của mô hình, nơi mà dữ liệu đầu vào được chuyển đến mạng.
- Hidden Layers: Mô hình có hai layer LSTM có kích thước 50 units mỗi layer. Cả hai layer đều được đặt với `return_sequences=True`, cho phép chúng trả về output tại mỗi bước thời gian, không chỉ ở bước cuối cùng. Điều này giúp mô hình học được thông tin từ mỗi bước thời gian trong chuỗi đầu vào.
- Output Layer: Layer cuối cùng của mô hình là một layer Dense với một unit (được đặt với `Dense(1)`). Điều này làm cho layer này trở thành một layer tuyến tính đơn giản, chỉ có một đầu ra. Hàm kích hoạt mặc định của layer này là hàm tuyến tính, tức là không có hàm kích hoạt được áp dụng trên đầu ra của nó.
- Optimizer và Loss Function: Mô hình được biên dịch với optimizer là 'adam'. Adam là một thuật toán tối ưu hóa phổ biến được sử dụng để cập nhật trọng số của mô hình dựa trên gradient của hàm mất mát. Hàm mất mát được sử dụng để đo lường sai số giữa các dự đoán và giá trị thực tế là 'mean_squared_error' (được chỉ định trong `loss='mean_squared_error'`). Nhóm sử dụng MSE để đo lường sự khác biệt giữa dự đoán và giá trị thực tế, bằng cách tính trung bình bình phương của sai số.

Xây dựng cấu trúc mô hình GRU bao gồm 3 layer:

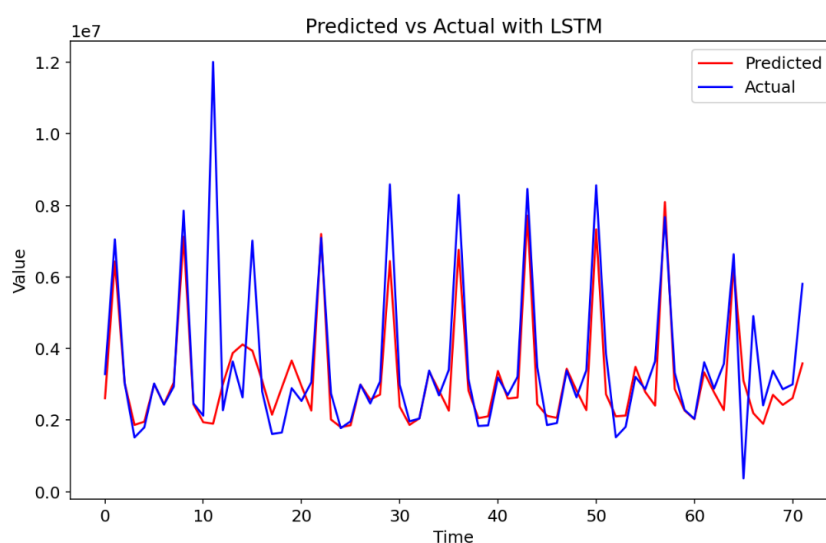
- Input Layer: Layer đầu tiên của mô hình là một layer GRU với 50 đơn vị (được xác định bởi tham số đầu tiên trong GRU(50)). Điều này có nghĩa là chúng ta sẽ có 50 neuron trong layer này. `input_shape` là kích thước của dữ liệu đầu vào, có dạng `(sequence_length, len(selected_features))`, trong đó `sequence_length` là độ dài của chuỗi thời gian mà chúng ta sử dụng để dự đoán và `len(selected_features)` là số lượng đặc trưng đã chọn cho mỗi bước thời gian.
- Hidden Layer: Mô hình có một layer GRU duy nhất với 50 đơn vị. Các layer GRU sau cùng sẽ trả về output tại mỗi bước thời gian, không chỉ ở bước cuối cùng. Điều này giúp mô hình học được thông tin từ mỗi bước thời gian trong chuỗi đầu vào.
- Output Layer: Layer cuối cùng của mạng neural là một layer Dense với một đơn vị (`Dense(1)`). Điều này có nghĩa là chúng ta sẽ có một đầu ra duy nhất từ layer này. Hàm kích hoạt mặc định của layer này là hàm tuyến tính, tức là không có hàm kích hoạt được áp dụng trên đầu ra của nó.
- Optimizer và Loss Function: Mô hình được biên dịch với optimizer là 'adam'. Adam là một thuật toán tối ưu hóa phổ biến được sử dụng để cập nhật trọng số của mô hình dựa trên

gradient của hàm mất mát. Hàm mất mát được sử dụng để đo lường sai số giữa các dự đoán và giá trị thực tế là 'mean_squared_error' (được chỉ định trong `loss='mean_squared_error'`). Độ đo này tính trung bình bình phương của sai số, là cách phổ biến để đo lường hiệu suất của mô hình dự báo.

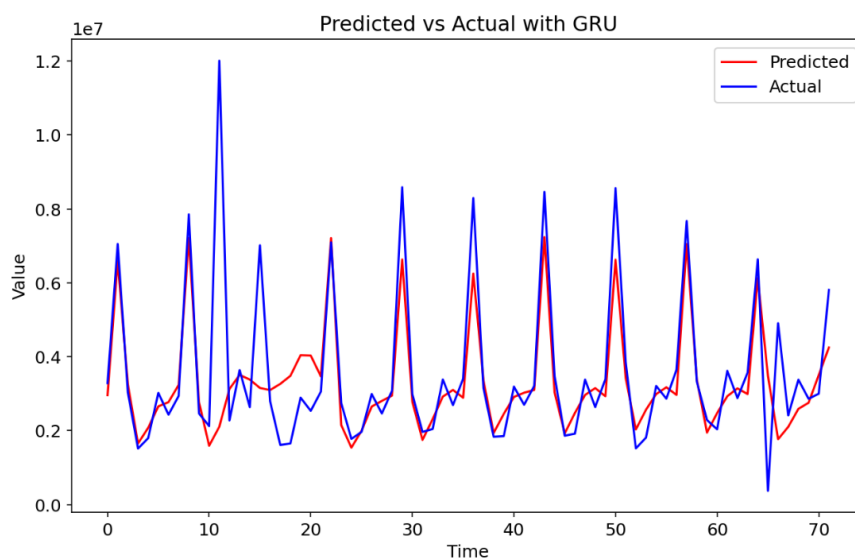
Sau khi xây dựng nhóm thực hiện huấn luyện mô hình dựa vào chuỗi dữ liệu huấn luyện được đề cập ở trên.

Đánh giá mô hình học máy RNN theo LSTM và GRU

Kết thúc giai đoạn dự đoán, nhóm thực hiện đánh giá mô hình dựa vào chuỗi dữ liệu kiểm tra (Từ 20/10/2022 đến 31/12/2022 tương đương với 72 ngày.), nhóm thu được kết quả như sau:



Hình 40. Kết quả dự đoán trên tập kiểm tra với LSTM



Hình 41. Kết quả dự đoán trên tập kiểm tra với GRU

Bảng đánh giá các chỉ số của 2 mô hình LSTM, GRU

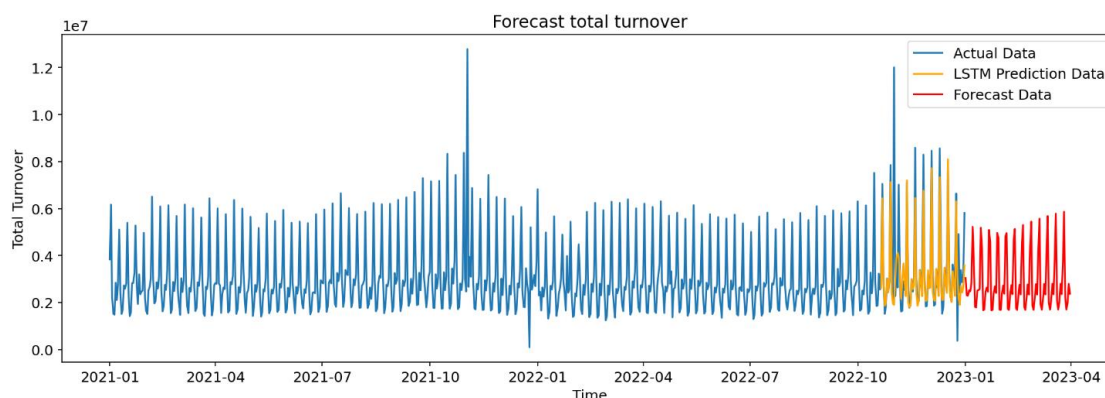
Bảng 8. Bảng đánh giá các chỉ số của 2 mô hình LSTM, GRU

Metrics	LSTM	GRU
MSE	2225744617441.1655	2343543392535.415
RMSE	1491892.964472038	1530863.610036967
MAPE	26.88046181893749	31.45873252476823
MAE	727570.2497992987	795862.209475643

Dựa vào các chỉ số đánh giá mô hình ARIMA bảng 5 và các chỉ số đánh giá mô hình LSTM và GRU kết luận được: Mô hình học sâu RNN có độ chính xác cao hơn mô hình truyền thống ARIMA và trong RNN trường hợp này mô hình LSTM có độ chính xác cao hơn GRU.

Dự đoán doanh thu trong tương lai

Sau khi chọn được mô hình tối ưu nhất là LSTM thì nhóm thực hiện dự đoán doanh thu tương lai trong vào quý 1 năm 2023.



Hình 42. Kết quả dự đoán tổng doanh thu của WA TAB trong quý 1 năm 2023

Kết quả chi tiết hơn được nhóm thể hiện ở bảng sau đây:

Bảng 9. Kết quả chi tiết

TOTAL_TURNOVER	
count	90.00
mean	2913282.00
std	1250804.00
min	1667945.00
25%	2006185.00
50%	2410126.00
75%	3831198.00
max	5865324.00

Từ dữ liệu bảng trên, có thể thấy so với 2 năm trước đó thì doanh số quý 1 năm 2023 theo dự đoán sẽ không thay đổi nhiều. Vì vậy công ty cần đưa ra các quyết định phù hợp để tăng trưởng trở lại mạnh mẽ hơn, phục hồi kinh tế sau đại dịch và tránh ảnh hưởng của suy giảm kinh tế toàn cầu.

IV. TỔNG KẾT

Về mặt hạn chế, do thiếu thông tin về dữ kiện phân vùng địa lý, nghề nghiệp cũng như thu nhập của khách hàng nên rất khó để có chiến lược cụ thể về cách tiếp cận cho từng tập khách hàng ở mảng nhân khẩu học. Bên cạnh đó, một số điểm chuyển đổi từ cao xuống thấp trong chu kỳ tăng trưởng của doanh thu, cụ thể là vào tháng 2,6,8 chưa thể tìm ra dẫn cứ khách quan, xác thực để bổ sung thêm cho luận điểm nguyên nhân mà nhóm đã đề ra.

Về những phần có thể phát triển trong tương lai, doanh nghiệp có thể tiến hành thu thập dữ liệu về hành vi của khách hàng khi sử dụng nền tảng online để có thể khai thác sâu hơn về các nhân tố ảnh hưởng đến quyết định rời bỏ của khách hàng.

PHỤ LỤC

Bảng từ viết tắt

Từ viết tắt	Ý nghĩa
RWWA	Racing and Wagering Western Australia
R	Rececy
F	Frequency
M	Monetary
ARIMA	Autoregressive Intergrated Moving Average
GRU	Gated Recurrent Unit
LSTM	Long short-term memory
WA TAB	Western Australia Totalisator Agency Board

Báo cáo trực quan về tình hình kinh doanh của WA TAB: [Tại đây](#).