



Financial Report

Presentation by

Võ Minh Thành

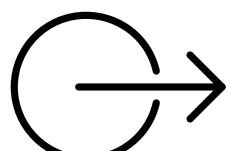
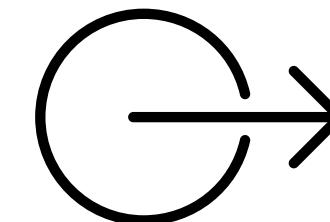


Table of

Contents



- 01.** Tổng quan về dữ liệu
- 02.** Xử lý dữ liệu
- 03.** Trả lời cho câu hỏi 1 và 2
- 04.** Trả lời cho câu hỏi 3



vexere

I. Overview về dữ liệu



1. Mô tả khái quát

Loại dữ liệu

Nhật ký đặt vé máy bay (mỗi dòng tương ứng 1 vé).

Quy mô

19,096 bản ghi, 20 cột.

Mục đích tiềm năng

Phân tích:

- Doanh thu
- Khách hàng
- Marketing - Strategy
- Chuyến bay

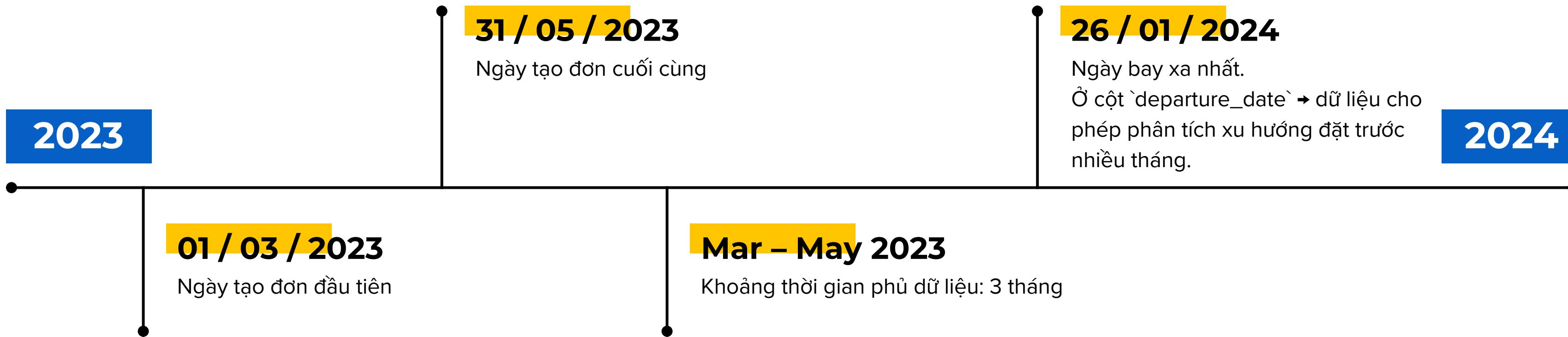
2. Cấu trúc & nhóm cột chính

Nhóm thông tin	Các trường tiêu biểu	Ghi chú
Đơn hàng	order_id, ticket_id, ticket_status, ticket_source_name	ID & trạng thái thanh toán
Hành khách	passenger_id, passenger_gender, passenger_age	ID đã băm ẩn danh – phù hợp phân tích nhân khẩu học
Thời gian	created_date, departure_date, departure_time, arrival_date, arrival_time	Tách rõ ngày & giờ
Hành trình	from_to_location, from_to, airline, seat_class	Tên tuyến & hãng bay
Giá trị giao dịch	ticket_price, baggage, insurance_fee, discount_amount	Cho phép phân tích doanh thu-chiết khấu

Thiếu dữ liệu:

- **passenger_gender**
(13 giá trị trống)
 - **passenger_age**
(5 giá trị trống)
 - **seat_class**
(120 giá trị trống)
- Tỷ lệ khuyết < 1 %.**

3. Phạm vi thời gian



4. Đặc trưng đáng chú ý

01. Nguồn đặt vé đa kênh

iOS, Android, GYL,... giúp so sánh hiệu quả kênh.

02. Giá vé & phụ phí tách riêng

thuận lợi tính doanh thu thuần và biên lợi nhuận.

03.Ẩn danh người dùng

tuân thủ bảo mật, nhưng vẫn hỗ trợ phân tích theo khách hàng.

04. Đa tuyến bay nội địa

trường from_to_location hiển thị cặp tỉnh → TPHCM/Hà Nội..., có thể gom cụm điểm đến.

II. Data Processing



Xử lý các giá trị bất thường

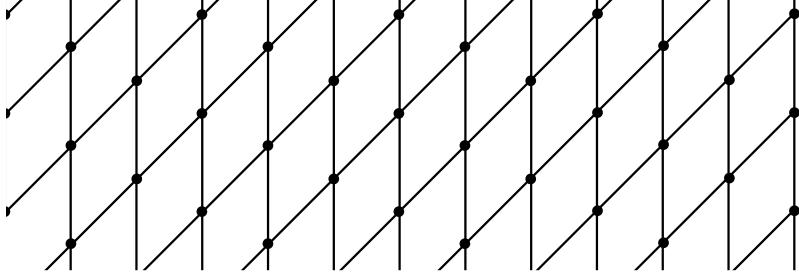
Data không có giá trị trùng lặp

Data bị null ở:

- passenger_gender
(13 giá trị trống)
- passenger_age
(5 giá trị trống)
- seat_class
(120 giá trị trống)
→ **Tỷ lệ khuyết < 1 %.**

Data có các giá trị âm hoặc sai logic tại:

- passenger_age: -1, 0, 115, 999999
- giá trị âm trong baggage, insurance_fee, discount_amount



passenger_gender

- **Phương pháp:** Ánh xạ passenger_id → passenger_gender để điền 11 null, sau đó loại 2 hàng còn lại.
- **Kết quả:** Từ 13 null ban đầu, điền được 11, loại bỏ 2 (0.01% tổng dữ liệu).

passenger_age

- **Phương pháp:** Ánh xạ passenger_id → passenger_age để điền 5 null, không cần loại bỏ.
- **Kết quả:** Từ 5 null ban đầu, điền được 5
- **Giá trị bất thường:** Ánh xạ passenger_id → passenger_age (loại bỏ các giá trị bất thường khỏi mapping) để điền 29 bản ghi.
- **Kết quả:** Tất cả 29 giá trị bất thường được sửa (0 còn lại; chiếm 0.15% tổng dữ liệu).

=> Ngoài ra trong bài này em cũng xác định các giá trị bị sai định dạng, chuyển lại cho đúng

seat_class

- **Phương pháp:** Loại bỏ thẳng 120 hàng null trong seat_class.
- **Kết quả:** Xóa 120/19094 hàng (0.63% tổng dữ liệu), không còn null.

baggage, insurance_fee, discount_amount

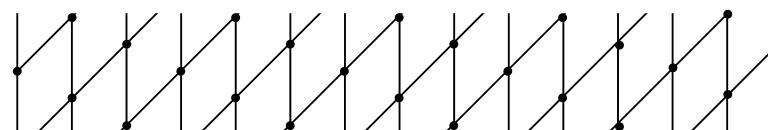
- **Phương pháp:** Thay tất cả giá trị âm thành 0 cho mỗi cột.
- **Kết quả:**
 - baggage: 2 giá trị âm (0.01%) → sửa hết thành 0.
 - insurance_fee: 3 giá trị âm (0.02%) → sửa hết thành 0.
 - discount_amount: 2 giá trị âm (0.01%) → sửa hết thành 0.

Feature

Enginnering

seat_class

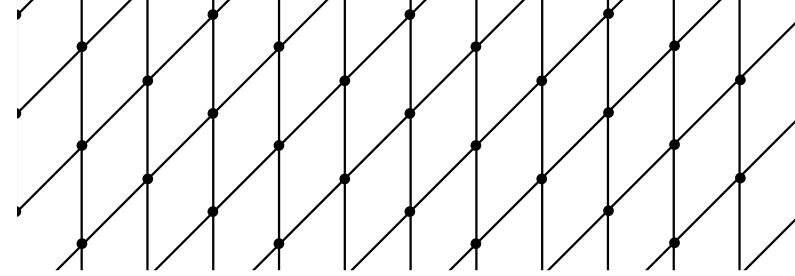
- **Phương pháp:** Gom nhóm 69 giá trị seat_class thành 4 hạng (First Class, Business, Deluxe, Economy) dựa trên chứa “SBoss”, “Business”, “DLX” hoặc còn lại.
- **Kết quả:** First Class: 4, Business: 16, Deluxe: 41, Economy: 18 913 (tổng 18 974, không còn Unknown).



Trả lời cho: Câu 1 và Câu 2



Phân tích theo chức năng



1. Revenue & Yield

- Đây là bước khởi đầu quan trọng để hiểu nguồn gốc doanh thu (Gross vs. Net) và hiệu suất giá vé (Yield) theo hạng ghế, từ đó xác định hạng ghế hoặc phân khúc nào đang đóng góp lớn nhất và phân bổ chiến lược giá phù hợp.

2. Customer Insights

- Phân tích hành vi và phân khúc khách hàng (qua độ tuổi, giới tính, kênh, tần suất, RFM, retention) để hiểu rõ ai đang mua vé, mua như thế nào và khả năng quay lại, từ đó định hướng chiến lược giữ chân và tối ưu trải nghiệm.

3. Marketing and Acquisition

- Phân tích xu hướng booking theo kênh và thời gian lead-time để tối ưu chiến dịch tiếp thị, ưu đãi và phân bổ tài nguyên nhằm tăng lượng đặt chỗ, thu hút khách mới và kích thích đặt sớm.

4. Route Performance

- Phân tích hiệu suất tuyến bay qua số lượt đặt, doanh thu, tỉ lệ hủy và tải theo giờ khởi hành để xác định “hot routes”, cân bằng doanh thu hai chiều, giảm hủy và tối ưu công suất chuyến.

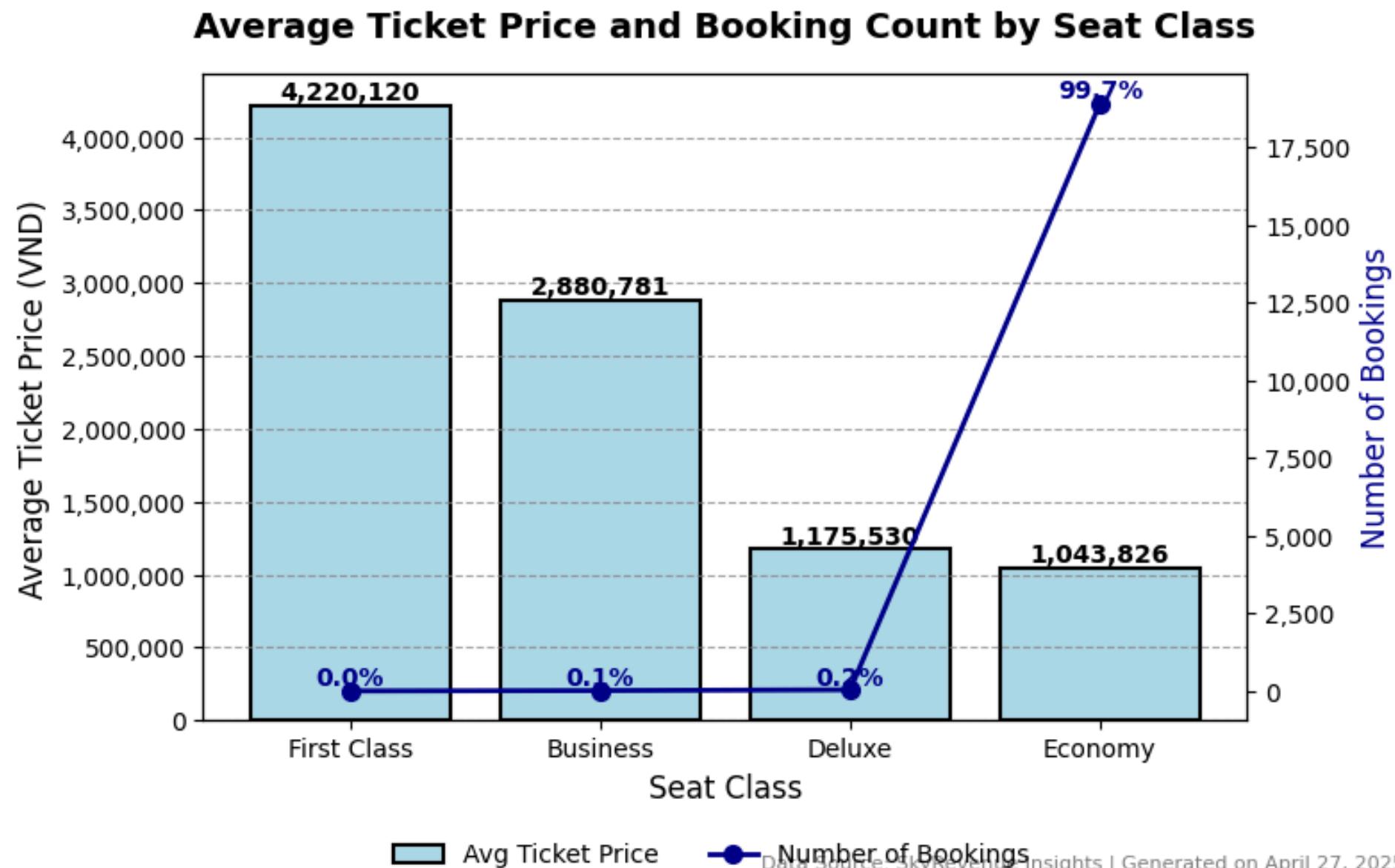
=> Ngoài ra trong bài này em cũng xác định các giá trị bị sai định dạng, chuyển lại cho đúng

1. Revenue & Yield

Gross Revenue: 19.85 tỉ VNĐ

Total Discount: 462,4 triệu VNĐ

Net Revenue: 19.39 tỉ VNĐ



Nhận định chính:

Phần lớn doanh thu của Sky được tạo ra từ lượng đặt chỗ hạng Economy khổng lồ – mặc dù giá vé **Economy** trung bình chỉ khoảng **1,04 triệu đ**, nhưng chiếm **99,7 % tổng số đặt chỗ** và đóng góp gần như toàn bộ doanh thu ròng (\approx 19,39 tỷ đ trong giai đoạn phân tích), trong khi các hạng cao cấp (First, Business, Deluxe) chiếm chưa đến 1 % lượng đặt chỗ dù giá vé cao gấp 2–4 lần.

Net Revenue by Booking Channel

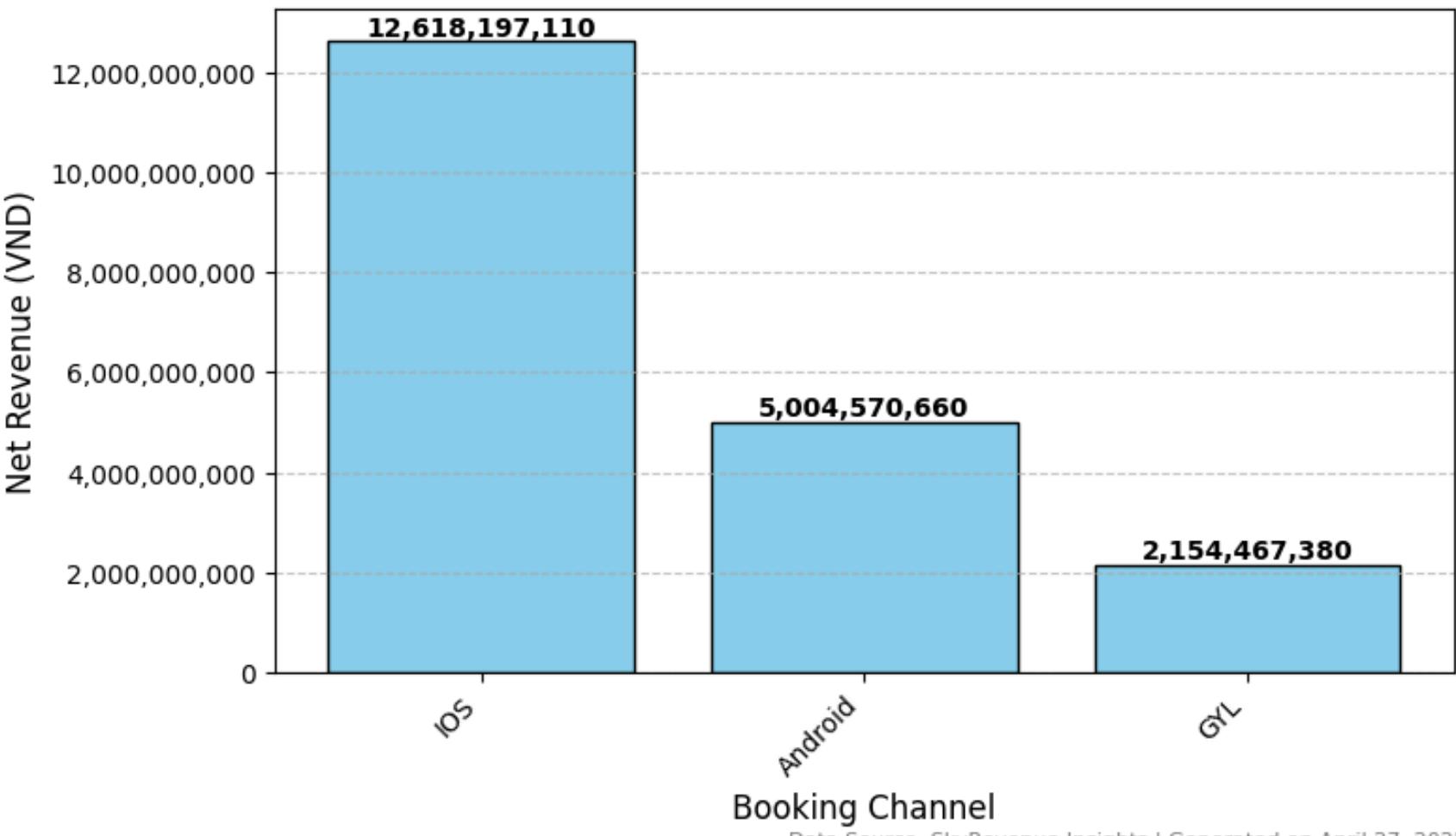


Chart 2

Nhận định chính:

- Chart 1 - Xu hướng doanh thu hàng tháng:** Doanh thu ròng tăng từ ~6,2 tỷ đ tháng 3 lên đỉnh ~8,2 tỷ đ tháng 4 rồi giảm về ~5,4 tỷ đ tháng 5.
- Chart 2 - Kênh đặt chỗ:** iOS chiếm ưu thế với ~12,62 tỷ đ (65 % tổng doanh thu), Android ~5 tỷ đ và GYL chỉ ~2,15 tỷ đ.
- Chart 3 - Tuyến bay chính:** HCM–Hà Nội và Hà Nội–HCM dẫn đầu với ~2,04 và ~1,77 tỷ đ, gấp hơn hai lần tuyến thứ ba.

→ Chú tâm:

- Ưu tiên tối ưu UX/UI, quy trình thanh toán và chương trình khuyến mãi cá nhân hóa trên iOS để gia tăng chuyển đổi và giá trị đơn hàng trung bình.**
- Triển khai các chương trình chuyên biệt (giá động, khuyến mãi theo mùa) và quảng bá mạnh trên tuyến này để duy trì và mở rộng thị phần.**

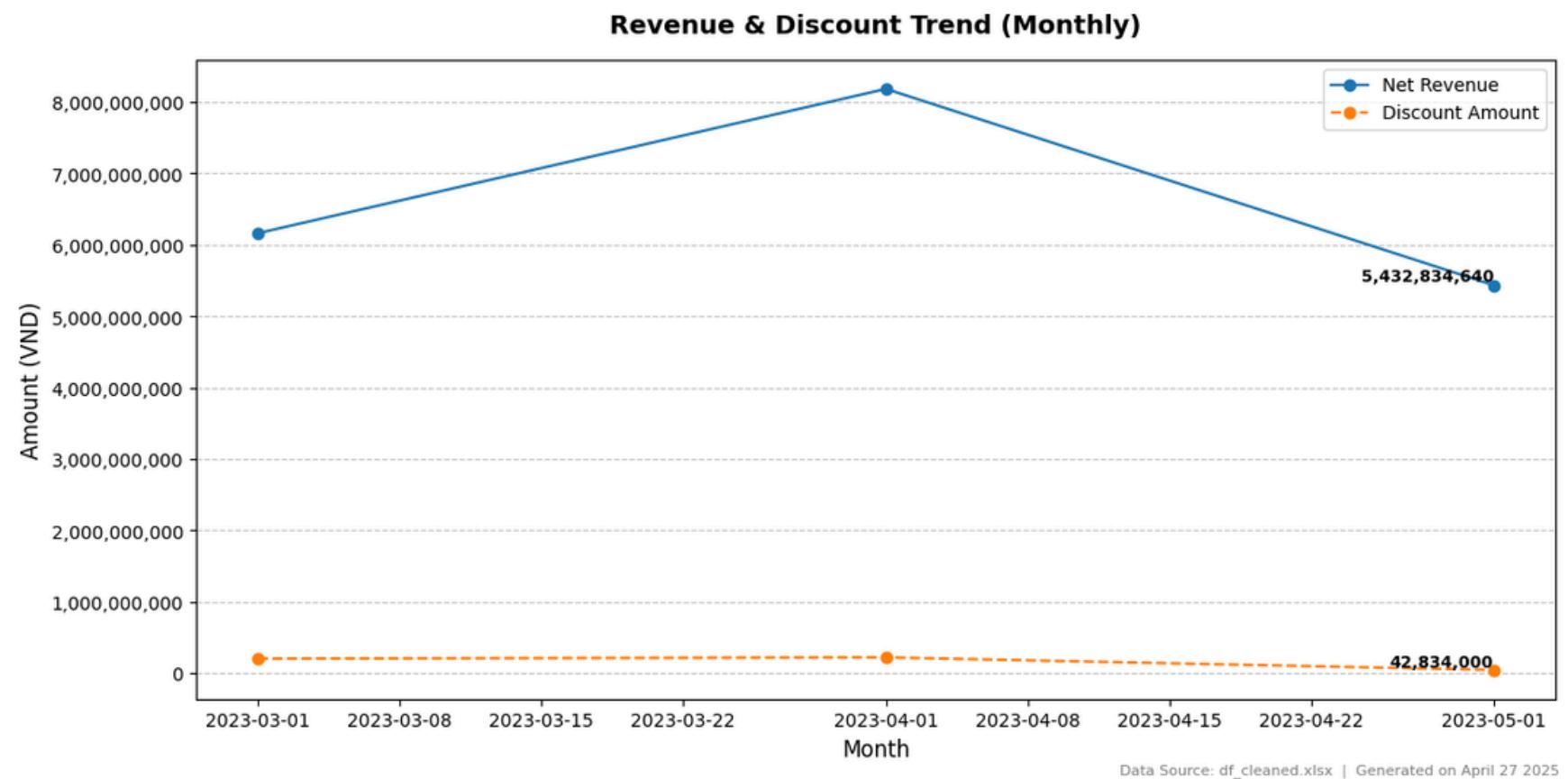


Chart 1
Top 10 Routes by Net Revenue

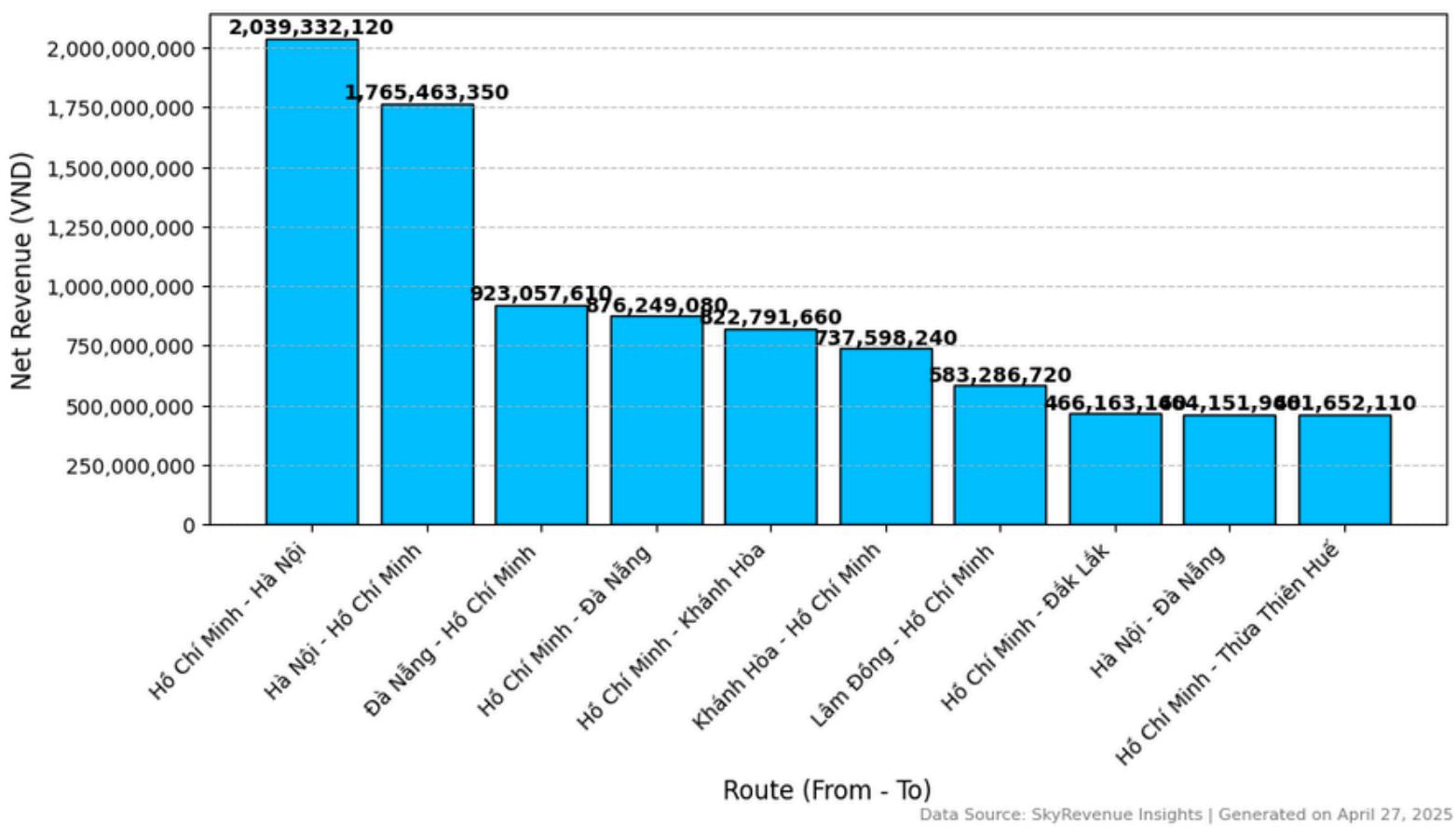
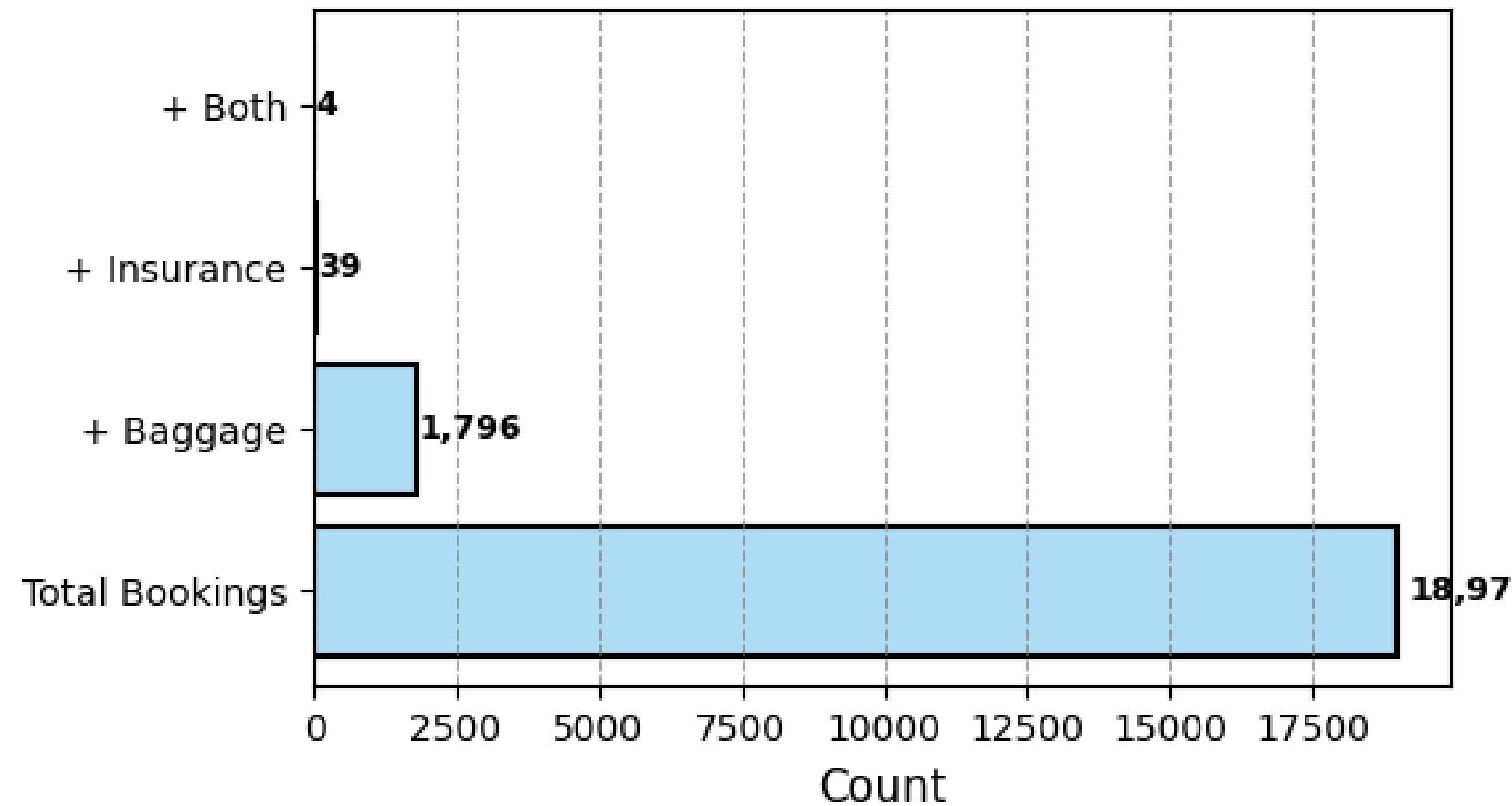


Chart 3

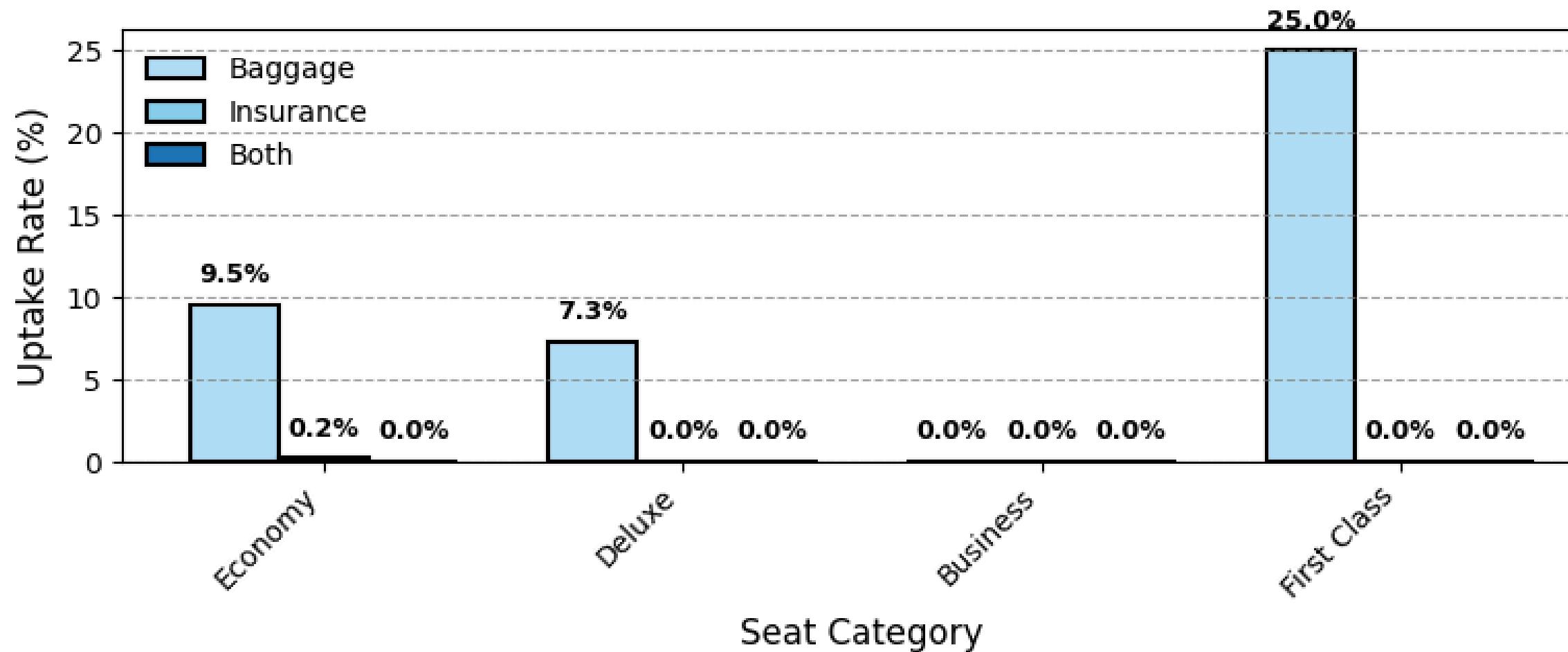
Ancillary Attach-rate Funnel



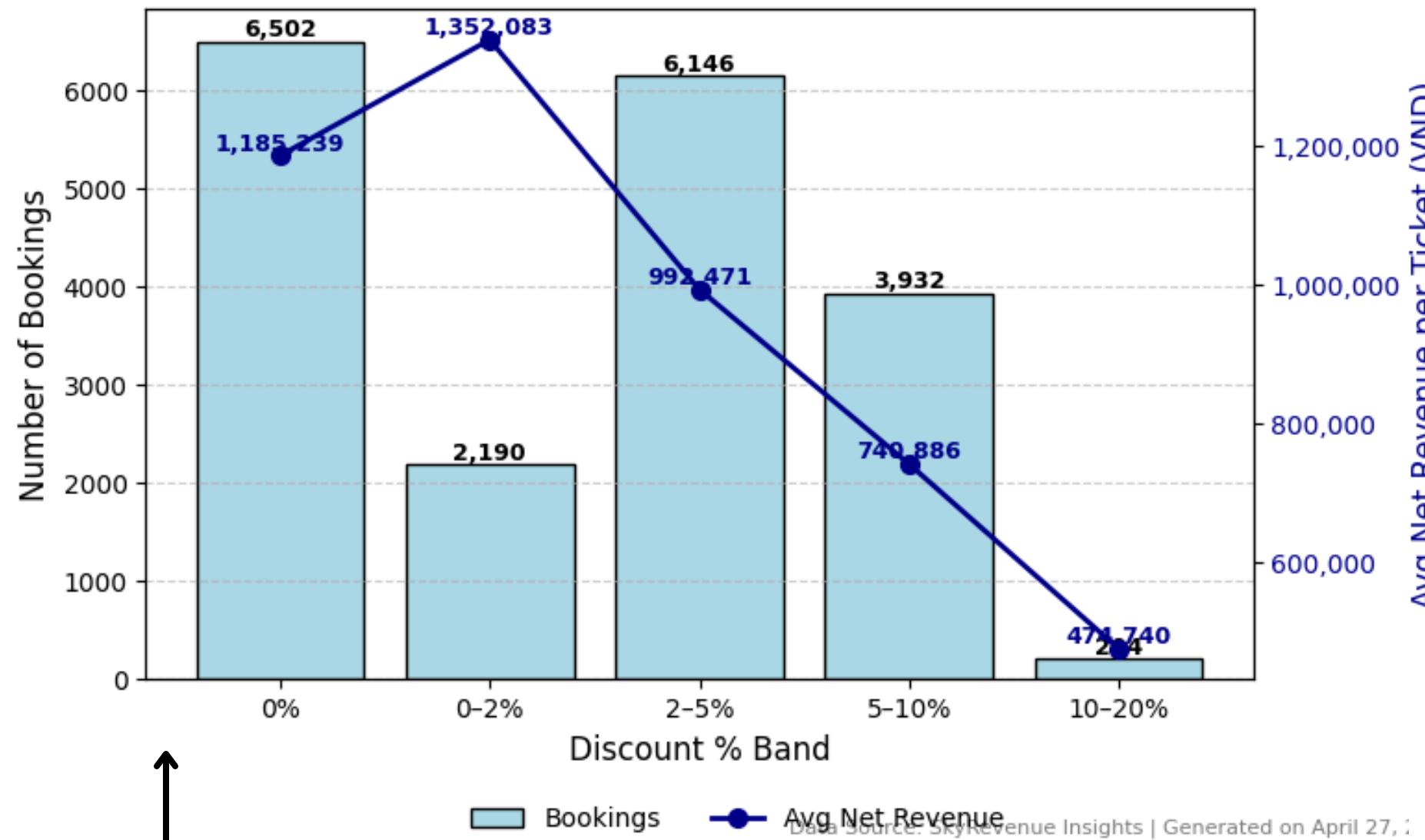
Nhận định chính:

- **Tỷ lệ mua thêm hành lý (baggage):** Khoảng 9,5 % tổng đặt chỗ, với hạng Economy ~9,5 %, Deluxe ~7,3 % và First Class cao nhất 25 %.
 - **Tỷ lệ mua bảo hiểm (insurance):** Rất thấp, chỉ 0,2 % tổng đơn, gần như không xuất hiện ở các hạng ngoài Economy.
 - **Giao dịch kết hợp cả hai (baggage + insurance):** Gần như không có (4 đơn).
- **Chú tâm:** Thiết kế gói upsell hành lý/bảo hiểm hoặc combo giá trị để tăng tỉ lệ mua thêm, đặc biệt nhắm tới khách Economy, Deluxe và đặc biệt là First Class

Ancillary Uptake by Seat Category



Booking Volume and Avg Net Revenue by Discount Band



Tối ưu khung chiết khấu:

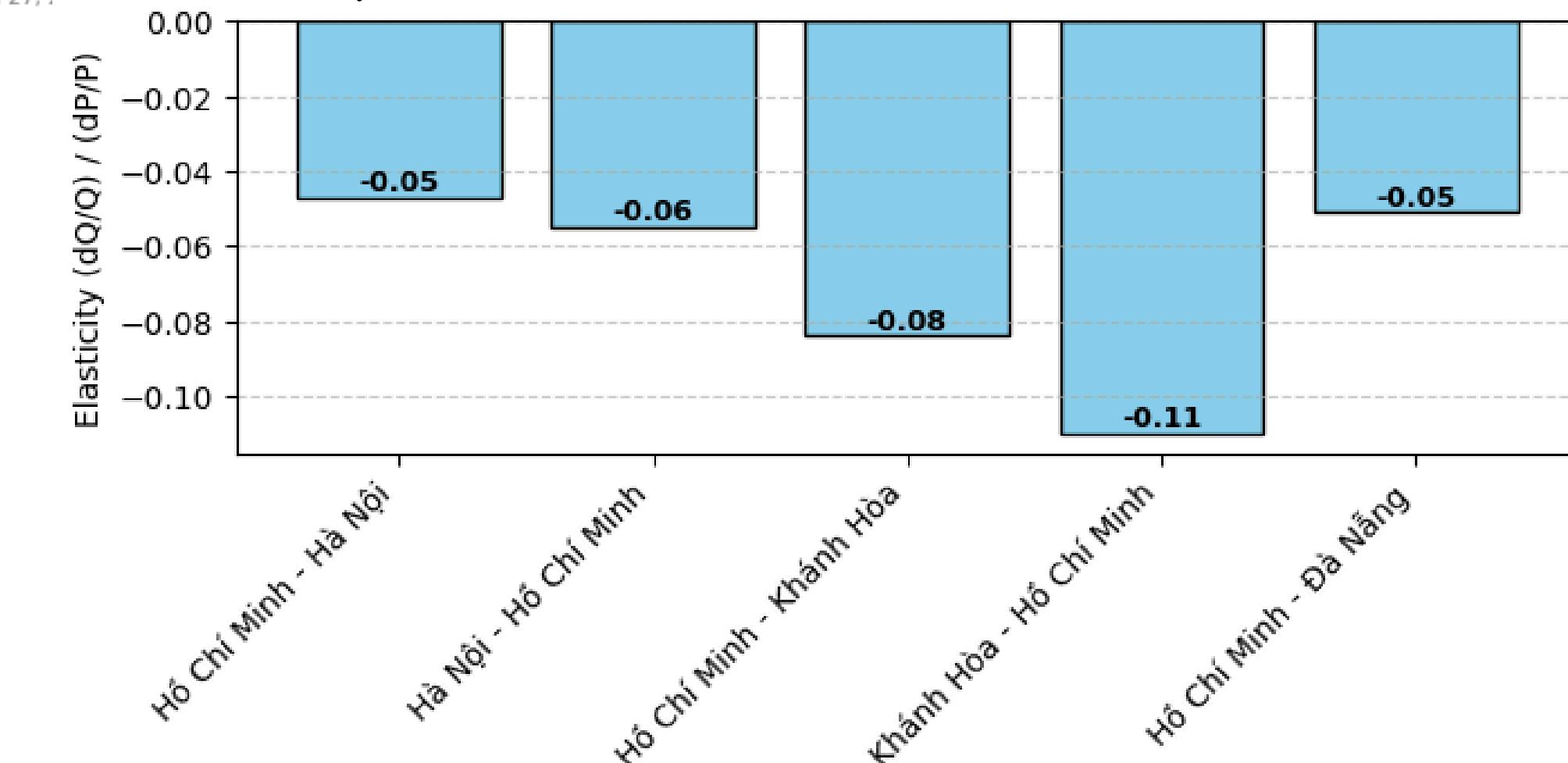
- Phân bổ khối lượng theo băng chiết khấu:** Nhóm không chiết khấu (0 %) chiếm ~6500 booking với giá vé ~1,18 triệu đ; băng 2–5 % cũng đạt ~6146 booking nhưng giá vé giảm còn ~0,99 triệu đ.
- Hiệu quả chiết khấu cao:** Chiết khấu 10–20 % chỉ thu hút ~472 booking, quá thấp so với mức giảm giá.
- **Chú tâm:** Giữ giá gốc làm mặc định, chỉ cân nhắc chiết khấu nhẹ (<2 %) cho các nhóm khách cụ thể hoặc áp dụng khuyến mãi giá trị gia tăng (combo, upsell) thay vì giảm trực tiếp.

Phân biệt chiến lược giá theo tuyến:

Độ co giãn giá theo tuyến: Khánh Hòa ↔ Hồ Chí Minh nhạy cảm nhất với giá ($\epsilon \approx -0,11$), trong khi Hà Nội ↔ Hồ Chí Minh ít nhạy cảm nhất ($\epsilon \approx -0,06$).

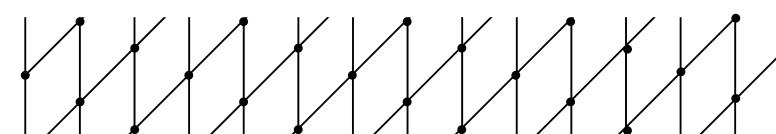
→ **Chú tâm:** Xây dựng cơ chế giá động theo tuyến, thử giảm nhẹ giá trên Khánh Hòa–HCM để kích cầu và có thể tăng giá trên HN–HCM mà không làm sụt lượng đặt chỗ.

Price Elasticity of Demand by Route



Kết luận

- **Đặc điểm chung:** Doanh thu chủ yếu đến từ hạng Economy với khối lượng đặt chỗ áp đảo dù giá vé thấp; các hạng cao cấp bán rất ít dù giá cao gấp 2–4 lần; kênh iOS chiếm ưu thế về doanh thu; khung chiết khấu hiện tại chủ yếu là giảm giá nhẹ (2–5 %) nên không tạo cú hích mạnh.
- **Hướng làm tiếp:** Tối ưu hóa yield bằng cách triển khai dynamic pricing dựa trên hạng ghế và tuyến, cá nhân hóa chiết khấu theo phân khúc (ví dụ chiết khấu sâu cho khách mới hoặc “At Risk”), đồng thời phát triển upsell dịch vụ phụ trợ (hành lý, bảo hiểm, bundle) để gia tăng giá trị đơn hàng trung bình mà không phụ thuộc thuần vào cắt giảm giá vé.



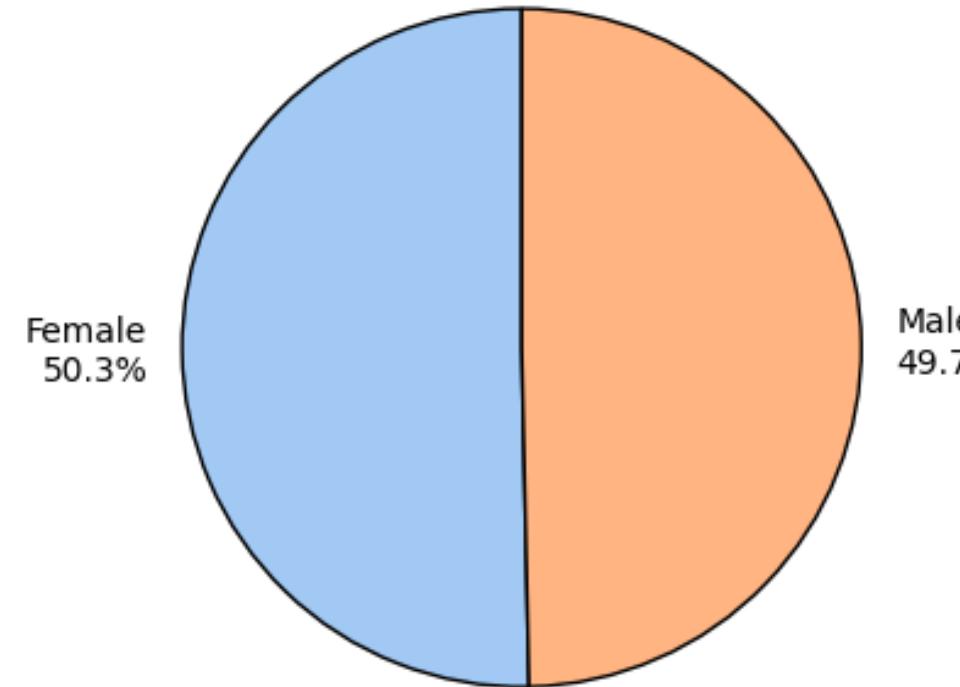
2. Customer Insights

Unique Passengers: 5252 người

Total Booking: 18974

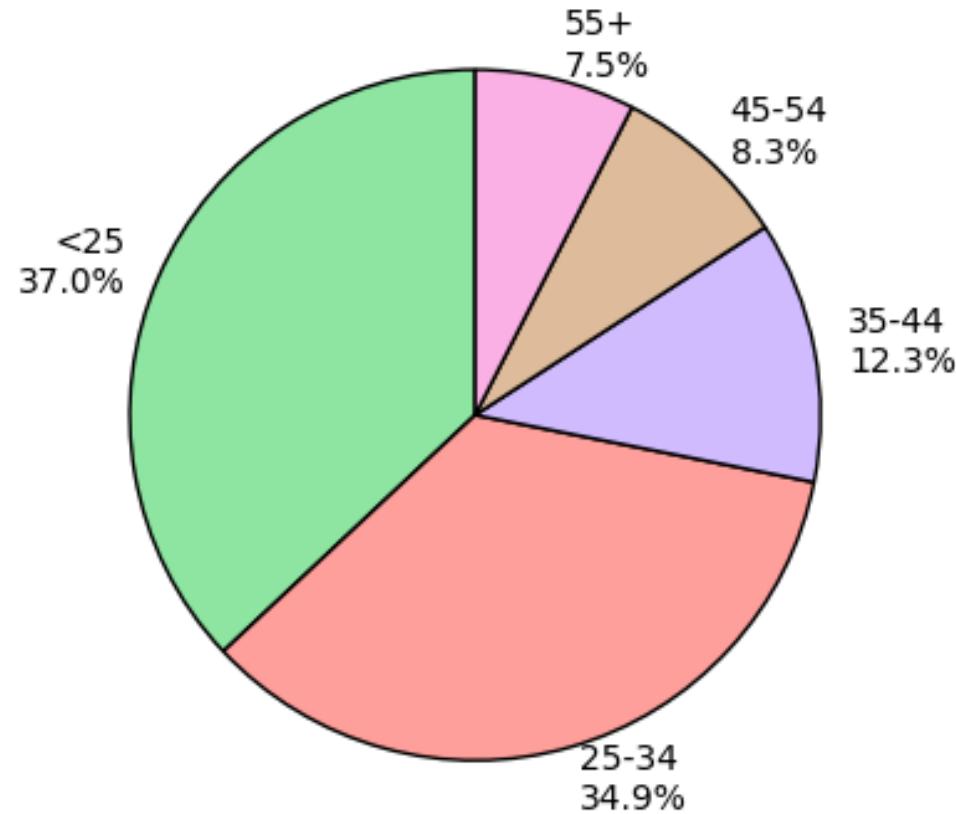
Avg Bookings per Passenger: 3.61

Passenger Gender Share



Source: SkyRevenue

Passenger Age Group Share



Source: SkyRevenue

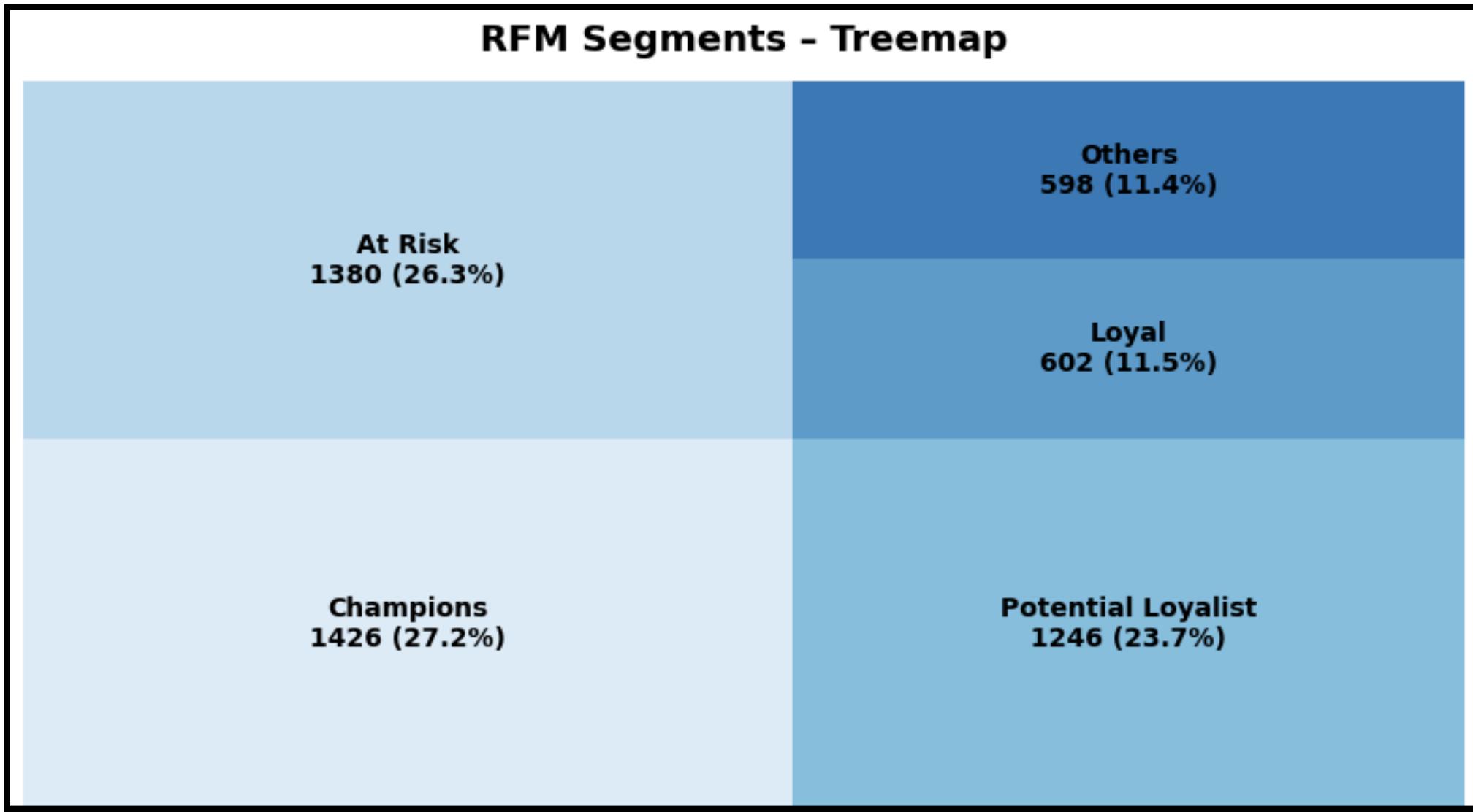
Nhận định chính:

- Tần suất đặt vé cao:** Mỗi hành khách trung bình đặt 3,61 lần trong 3 tháng, cho thấy mức độ quay lại tốt.
- Nhóm tuổi chủ lực (<35):** Chiếm 71,9 % hành khách, là lực lượng mua vé chủ yếu và thường đặt nhiều lần.
- Giới tính cân bằng:** Nam 49,7 % – Nữ 50,3 %, không cần ưu tiên riêng giới nào.
- Kênh di động ưa thích:** Nhóm trẻ dưới 35 tuổi ưu tiên đặt qua app.

Insights:

- Chương trình giữ chân & upsell cho khách quay lại:** Tạo “loyalty program” hoặc gói vé tích điểm, ưu đãi cá nhân hóa để khuyến khích hành khách mua tiếp (vì trung bình họ đã mua >3 vé).
- Ưu đãi cho nhóm trẻ (<35):** Phát triển gói giá sinh viên, last-minute deals và thông báo đẩy (push notification) qua app để kích thích đặt vé thêm.
- Chiến dịch marketing không phân biệt giới:** Thiết kế ưu đãi chung (ví dụ: gói “gia đình trẻ” hay “cặp đôi”) thu hút cả nam và nữ.
- Tối ưu trải nghiệm mobile:** Nâng cấp tính năng app (chatbot, đề xuất cá nhân hóa) để đẩy mạnh tương tác và đặt vé quay lại.

RFM Segments - Treemap

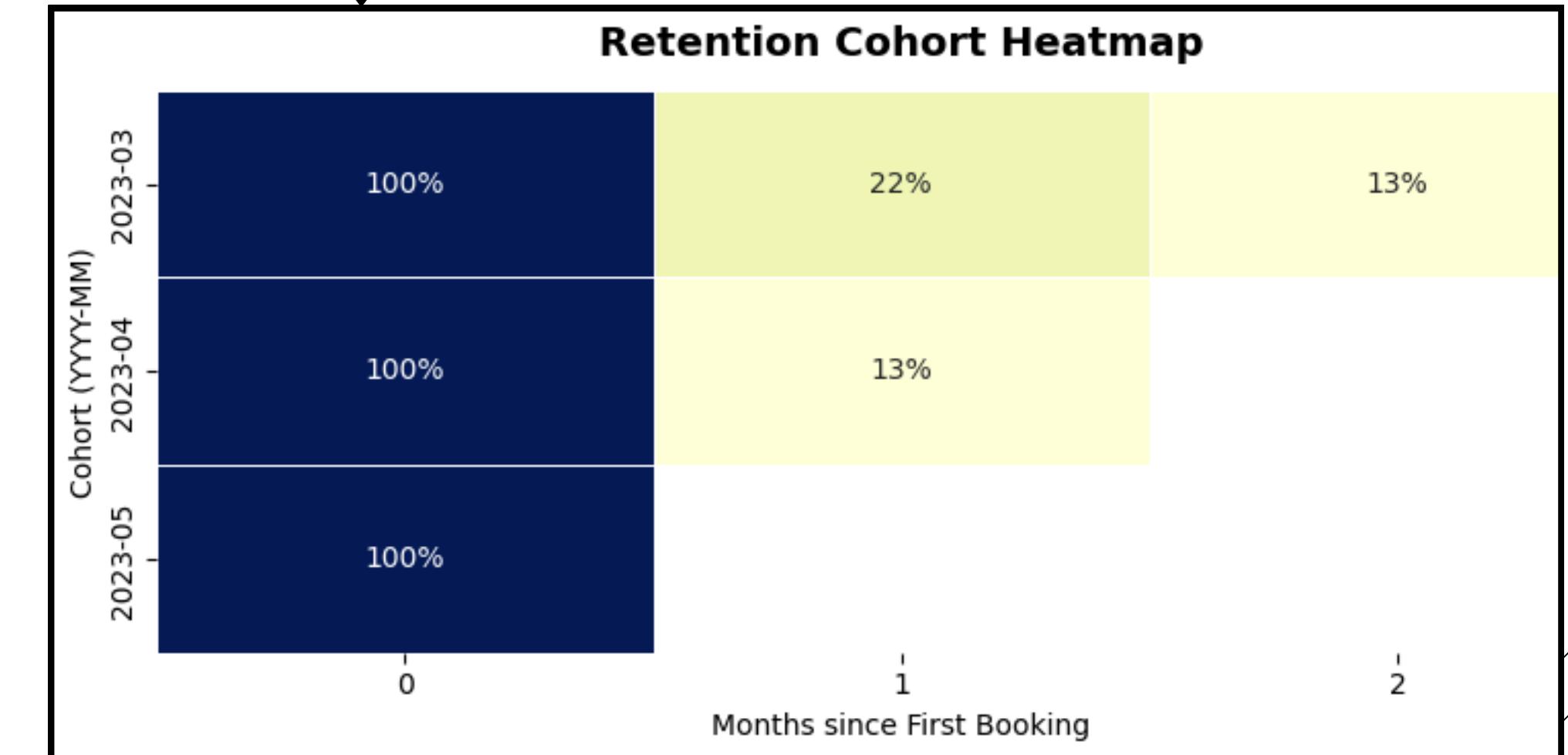


Insight: Sau 1 tháng, chỉ ~22 % khách từ cohort 03/2023 và ~13 % khách từ 04/2023 quay lại đặt vé; đến tháng thứ 2, retention chỉ còn ~13 % (03/2023) và ~0 % (04/2023).

→ **Chú tâm:** Cần triển khai chiến dịch chăm sóc và nhắc nhở (email/SMS/app push) ngay sau lần đặt đầu để cải thiện tỷ lệ mua lại trong tháng đầu, đặc biệt với nhóm mới.

Insight: “Champions” (27,2 %) và “At Risk” (26,3 %) chiếm hơn nửa cơ sở khách; “Loyal” chỉ 11,5 %.

→ **Chú tâm:** Ưu tiên giữ chân nhóm “At Risk” bằng ưu đãi cá nhân hóa; tăng chuyển đổi “Potential Loyalist” (23,7 %) thành “Loyal” qua chương trình tích điểm/khuyến mãi định kỳ; củng cố mối quan hệ với “Champions” thông qua lợi ích độc quyền.



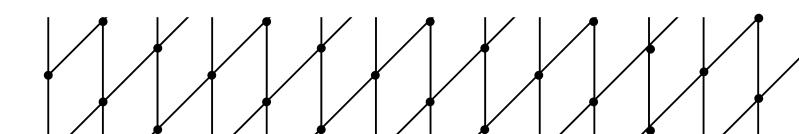
Kết luận

- **Đặc điểm chung:**

- Khách under 35 tuổi chiếm ~72 %, đặt trung bình 3,6 lần; tỉ lệ nam/nữ cân bằng; kênh di động (iOS/Android) ưu thế về cả booking lẫn retention.
- RFM cho thấy hai nhóm lớn nhất là Champions (27 %) và At Risk (26 %), nhưng nhóm Loyal chỉ 11 %, retention sau tháng đầu chỉ ~20 %.

- **Hướng tiếp theo:**

- *Xây dựng chương trình loyalty & chăm sóc cá nhân hóa:* Ưu tiên giữ chân nhóm At Risk và chuyển đổi Potential Loyalist thành Loyal thông qua điểm thưởng, voucher định kỳ.
- *Chiến dịch kích cầu nhóm mới:* Dùng ưu đãi “lần đầu” và nhắc nhở tức thì sau booking để nâng retention tháng đầu.
- *Cá nhân hóa nội dung kênh di động:* Đẩy mạnh push notification và in-app message phù hợp nhóm tuổi <35, tập trung vào các phân khúc có tần suất đặt cao.

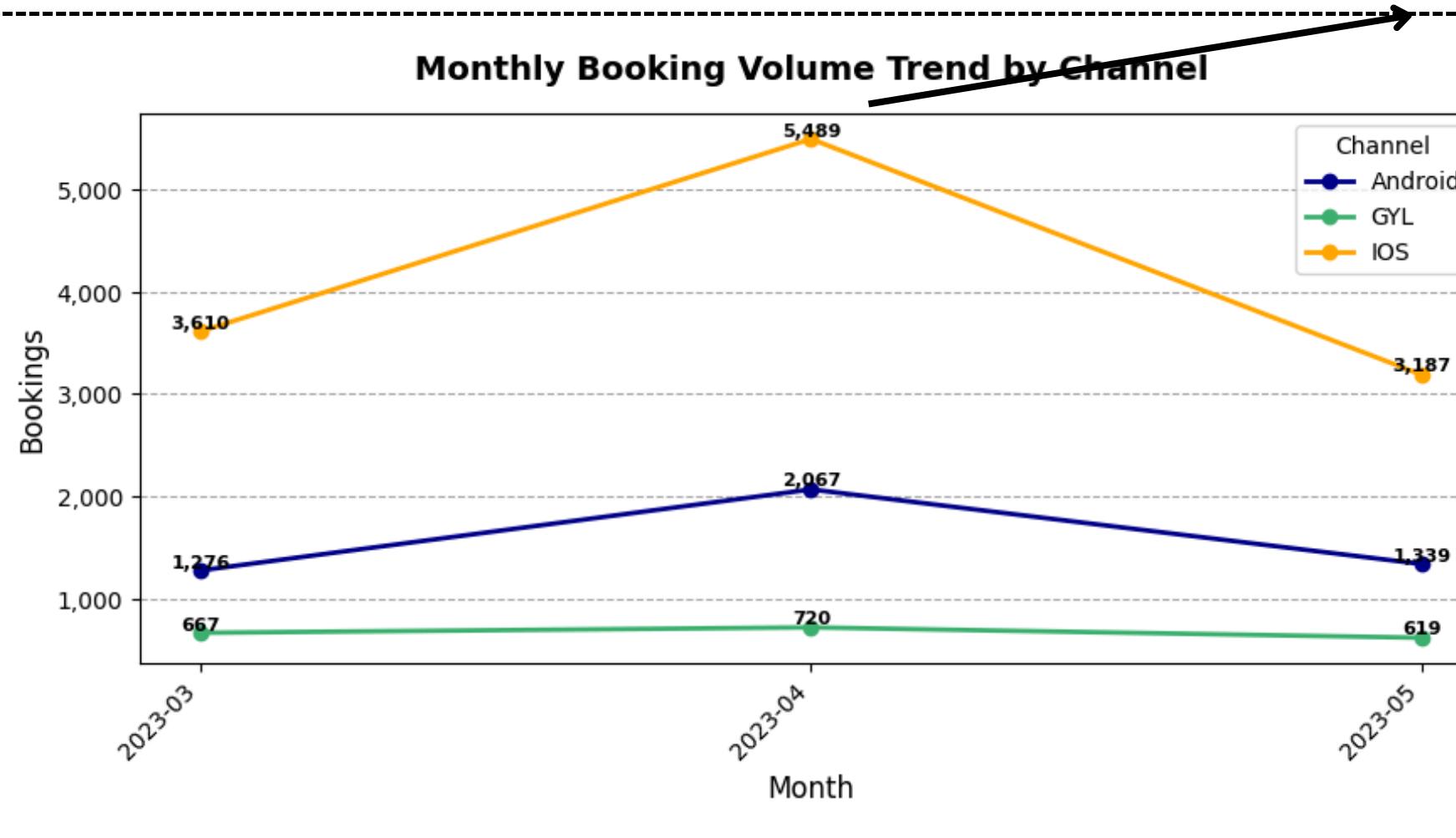


3. Marketing and Accquisition

Total April Bookings: 8276

Bookings Depart 28/4–3/5: 1623

Percentage: 19.6%



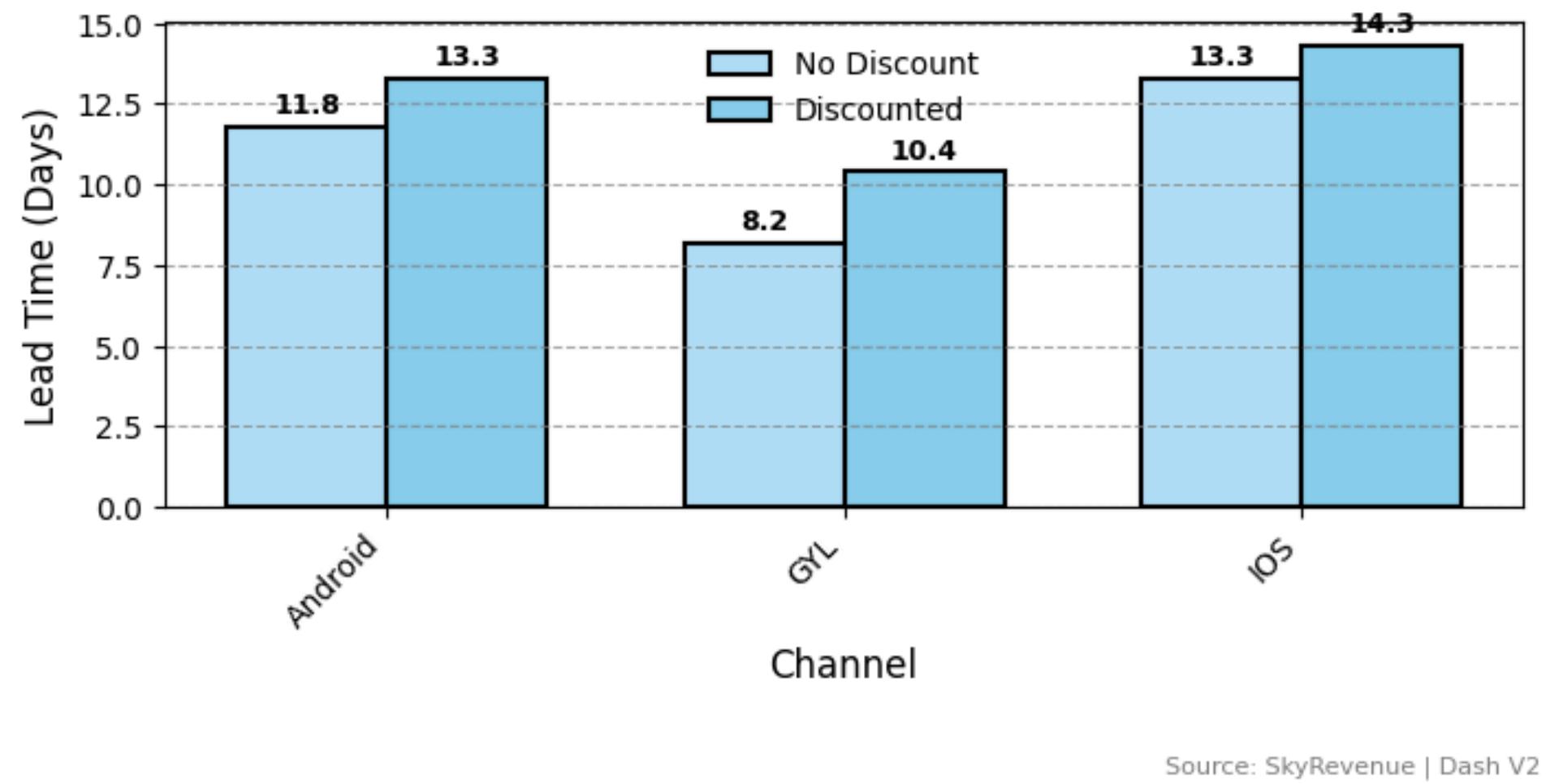
Nhận định chính:

- Xu hướng đặt chỗ theo kênh:** Tất cả kênh đều đạt đỉnh booking tháng 4 rồi giảm tháng 5, nhưng iOS tăng mạnh nhất (+52 % tháng 3→4) và vẫn cao gấp ~2.6× Android.
- Tính mùa vụ rõ rệt:** Booking tháng 5 trên mỗi kênh giảm ~30 % so với tháng 4.
- Kênh GYL ổn định thấp:** GYL chỉ tăng nhẹ tháng 4 và giảm tháng 5, luôn ở dưới 4 % tổng volume.

Chú tâm:

- Giữ đà tăng trên iOS** bằng chiến dịch khuyến mãi hoặc push notification vào tháng thấp điểm; đồng thời cải thiện trải nghiệm Android để giảm khoảng cách.
- Lên kế hoạch ưu đãi “giảm giá mùa thấp điểm”** hoặc khuyến khích đặt vé sớm (early bird) vào tháng 3 để ổn định dòng tiền hoặc tháng 5 cho mùa hè.

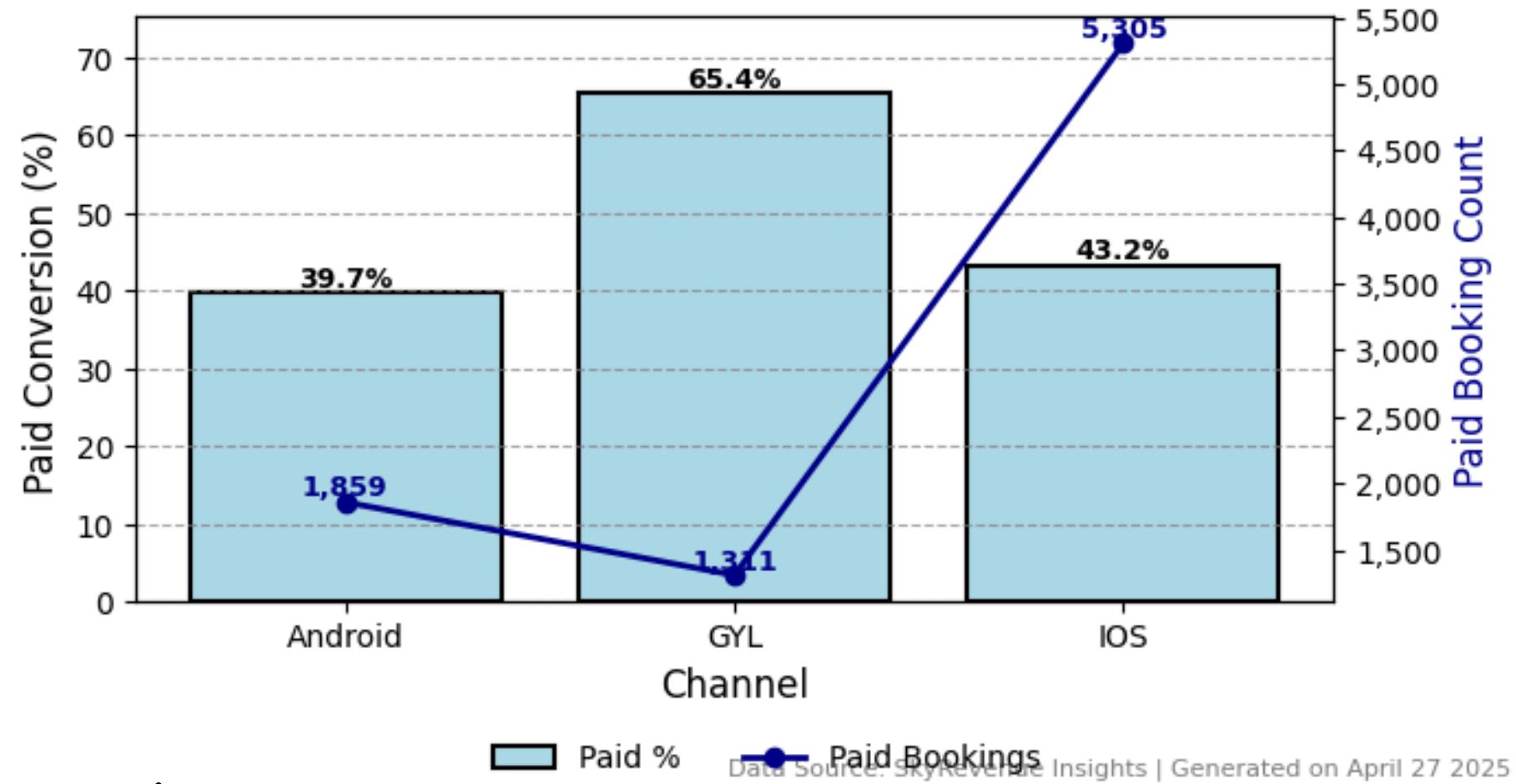
Avg Lead Time by Channel: Discounted vs No Discount



Nhận định chính và Chú tâm

- Insight chính:** Booking có chiết khấu có xu hướng đặt sớm hơn: trung bình sớm hơn 1–2 ngày trên mọi kênh (Android: 13,3 vs. 11,8 ngày; GYL: 10,4 vs. 8,2 ngày; iOS: 14,3 vs. 13,3 ngày).
- Chú tâm & Hành động:**
 - Ưu đãi cho khách đặt sớm (Early-Bird):** Đẩy mạnh chiết khấu nhỏ (<5 %) cho nhóm lead-time dài (>10 ngày) để khuyến khích khách đặt vé sớm, giúp cải thiện dự báo nhu cầu.
 - Cá nhân hóa thông điệp:** Gửi email/app push nhắc nhở khách có lịch trình bay dài về ưu đãi đặt sớm; tận dụng đặc tính kênh (iOS có lead-time dài nhất) để tối ưu ROI.
 - Phân khúc campaign:** Trên GYL, nơi khách thường đặt sát ngày bay (<10 ngày), tập trung “last-minute deals”; trên Android/iOS, ưu tiên “early bird” bundling dịch vụ phụ trợ (hành lý, bảo hiểm) kèm chiết khấu.

Paid Conversion Rate & Paid Bookings by Channel



Insight: GYL có tỉ lệ chuyển đổi thanh toán cao nhất (65,4 %) nhưng chỉ 1 371 giao dịch, trong khi iOS có tỉ lệ 43,2 % nhưng khối lượng thanh toán lớn nhất (5 305).

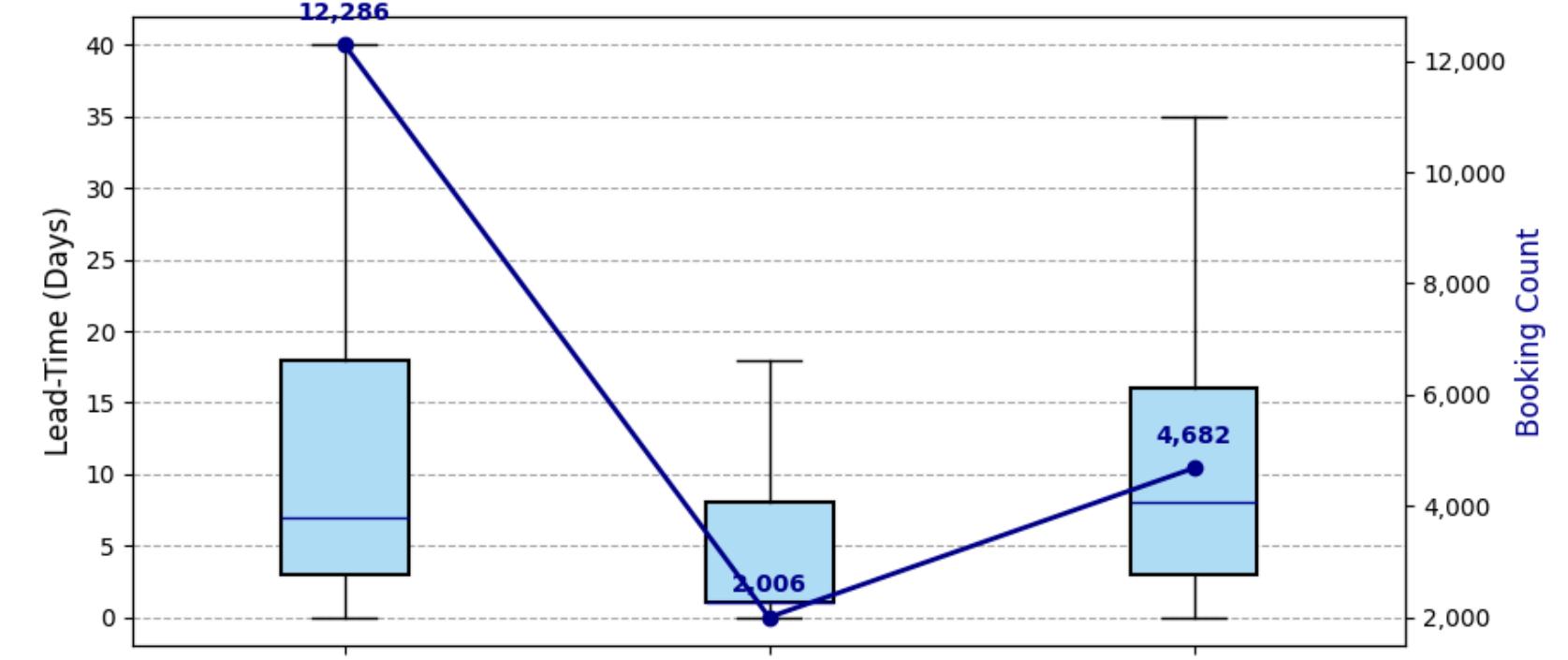
→ **Chú tâm:** **GYL hiệu quả về tỉ lệ nhưng thị phần nhỏ; nên đẩy mạnh tiếp thị và cải thiện UX/GYL để gia tăng số lượng giao dịch.** Trên iOS, ưu tiên phát triển các tính năng mua nâng cao (one-click checkout) để tận dụng tỉ lệ chuyển đổi ổn định với khối lượng lớn.

Insight: Trung bình khách đặt trước 13 ngày, nhưng 50 % booking xảy ra trong khoảng 2–17 ngày trước bay; nhóm GYL đặt gấp nhất (75 % <8 ngày), iOS dài hạn nhất (75 % <18 ngày).

→ **Chú tâm:**

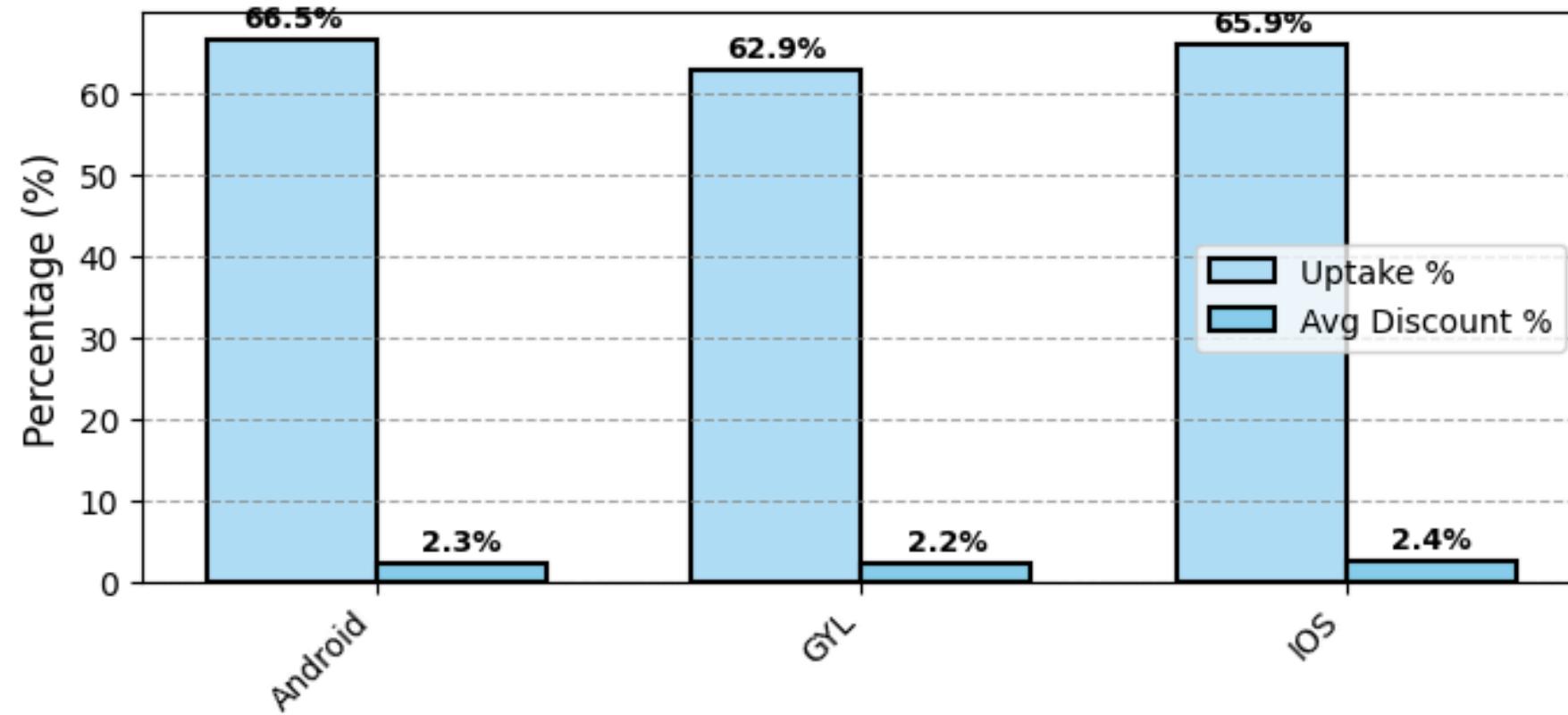
- iOS: Khách đặt sớm, thích dịch vụ trọn gói—có thể tung “early bird” ưu đãi đặc biệt.
- GYL: Đặt cận ngày, cần ưu đãi “last-minute” hoặc liên kết với thông tin chuyến khẩn cấp (ví dụ: cảnh báo giá giảm gần ngày bay).
- Android ở giữa: Cân bằng cả hai chiến lược.

Lead-Time Distribution & Booking Count by Channel



Data Source: SkyRevenue Insights | Generated on April 27 2025

Discount Uptake vs Avg Discount Rate by Channel



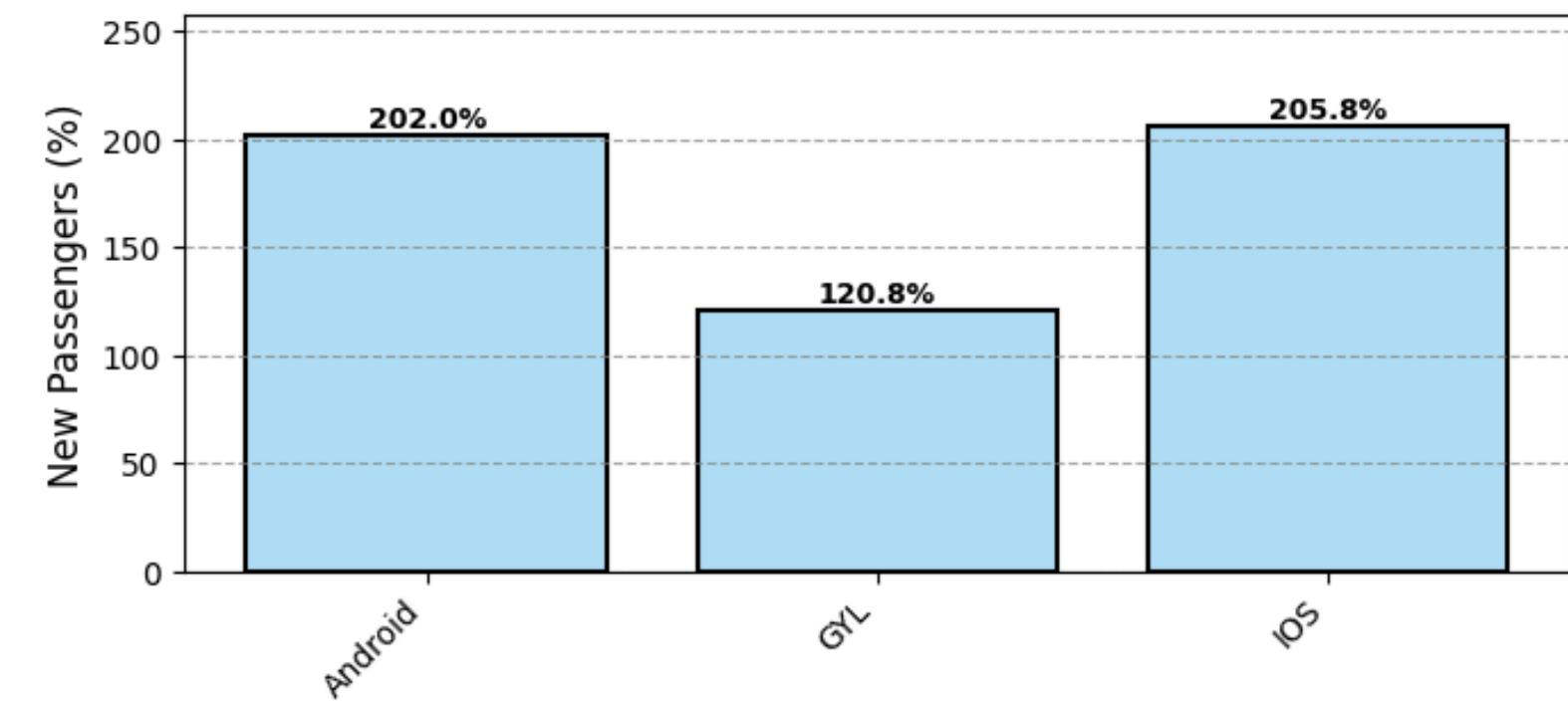
Data Source: SkyRevenue Insights | Generated on April 27 2025

Insight: Mặc dù có đến ~65–67 % đơn đặt chỗ qua ba kênh (Android, GYL, iOS) nhận chiết khấu, mức chiết khấu trung bình trên mỗi booking lại rất thấp (chỉ ~2,2–2,4 %) và chưa được phân bổ trọng tâm cho iOS(nền tảng cho nhiều khách và lợi nhuận nhất)

→ **Chú tâm:** *Phân chia tỉ lệ chiết khấu tùy theo nền tảng mang lại nhiều lợi nhuận và đặc điểm nhóm khách hàng đó*

- **Insight:** iOS và Android đều thu hút lượng khách mới vượt hơn 200 % so với số khách cũ, trong khi GYL chỉ ~120,8 %. Khách iOS và Android quay lại đặt nhiều lần hơn (“Avg_Bookings_per_Pax” ~3,49 và 3,38) so với GYL (~2,55).
- **Chú tâm:** Đẩy mạnh chiến dịch tiếp thị thu hút khách mới trên GYL, có thể qua referral program hoặc ưu đãi lần đầu, để cân bằng nguồn khách.

Share of New Passengers by Channel



Data Source: SkyRevenue Insights | Generated on April 27 2025

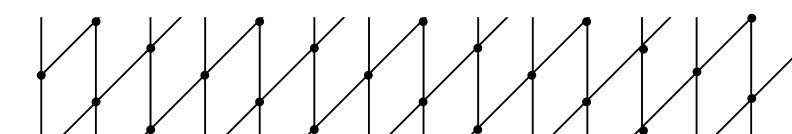
Kết luận

- **Đặc điểm chung:**

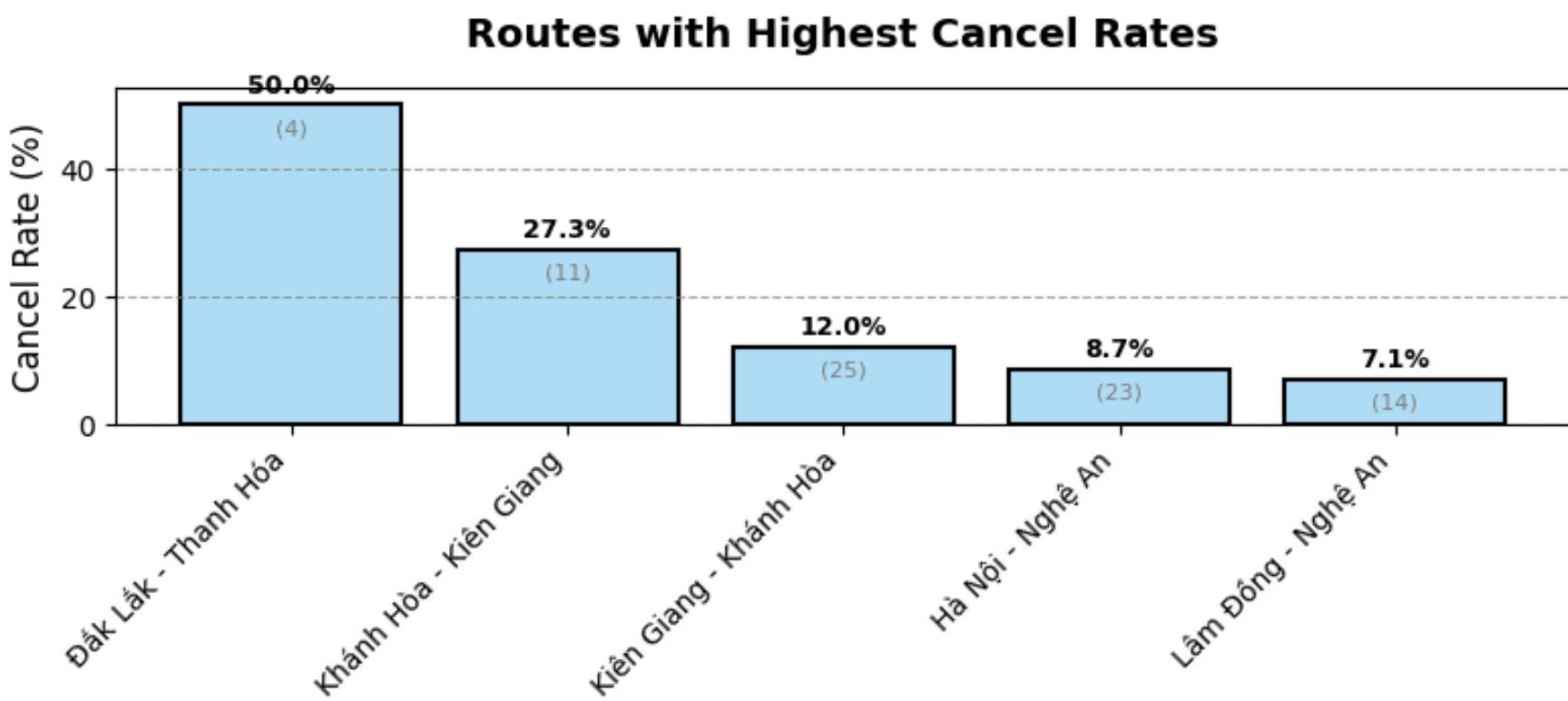
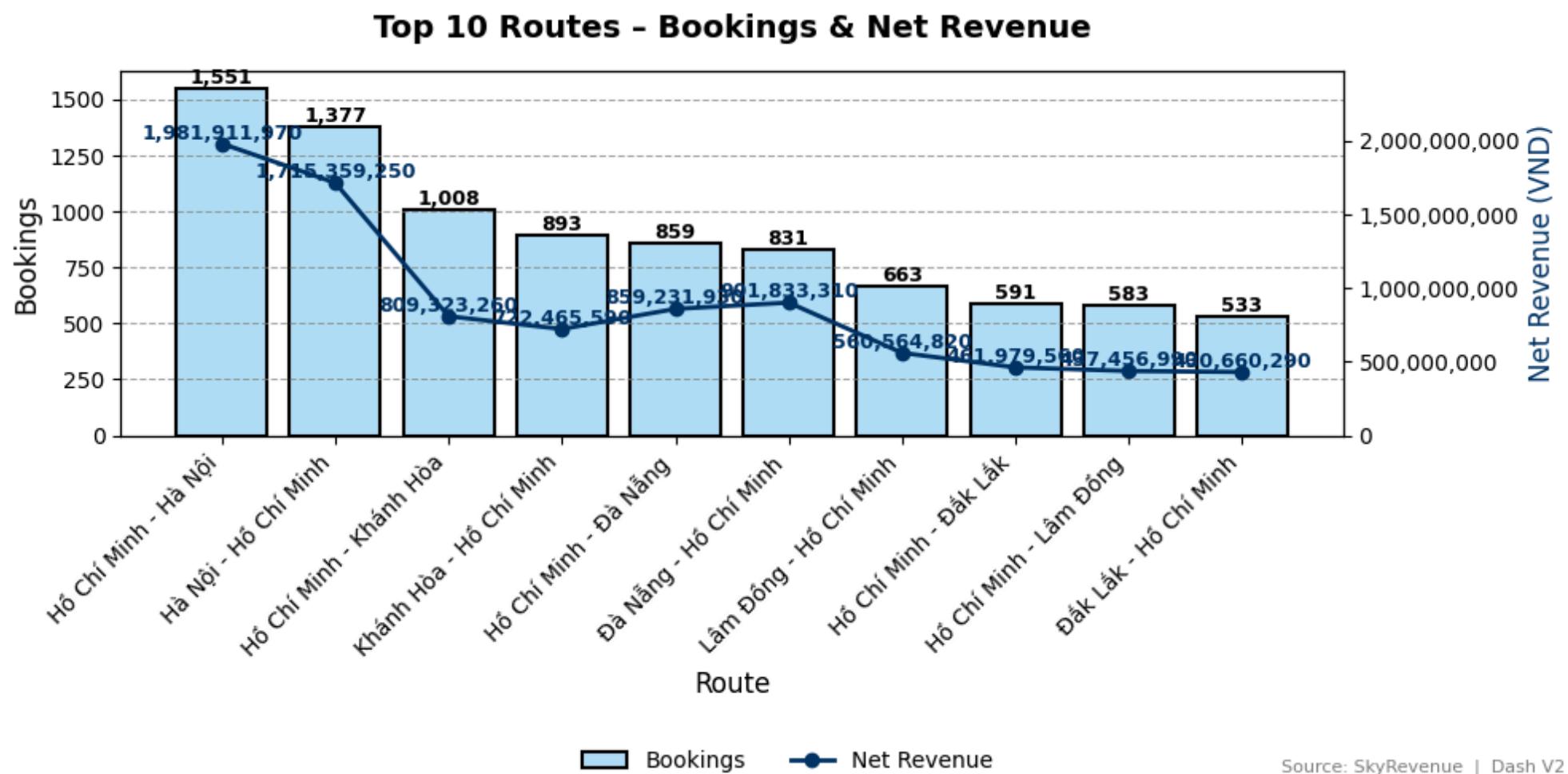
- iOS dẫn đầu với đỉnh booking tháng 4 và tỷ lệ tăng mạnh; Android đứng thứ hai; GYL luôn nhỏ dưới 4 % volume.
- Khách có chiết khấu thường đặt sớm hơn 1–2 ngày; GYL đặt sát ngày (<10 ngày), iOS đặt sớm nhất.
- GYL có tỉ lệ thanh toán cao (65 %) nhưng khối lượng thấp; iOS thanh toán lớn nhất.
- Android/iOS thu hút >200 % khách mới, GYL ~120 %.

- **Hướng tiếp theo:**

- **Chiến dịch mùa thấp điểm & early-bird:** Đẩy mạnh ưu đãi “giảm giá mùa thấp điểm” hoặc “early bird” chiết khấu nhẹ (<5 %) cho Android/iOS; triển khai “last-minute deals” cho GYL.
- **Cá nhân hóa kênh:** Sử dụng push notification/email targeting cho khách lead-time dài (iOS) và nhắc nhở gấp ngày (GYL); cải thiện UX/GYL để nâng lượng giao dịch.
- **Chương trình thu hút & giữ khách mới:** Referral program và ưu đãi lần đầu trên GYL; tiếp tục nuôi dưỡng khách mới bằng drip marketing và upsell phụ trợ.



4. Route Performance

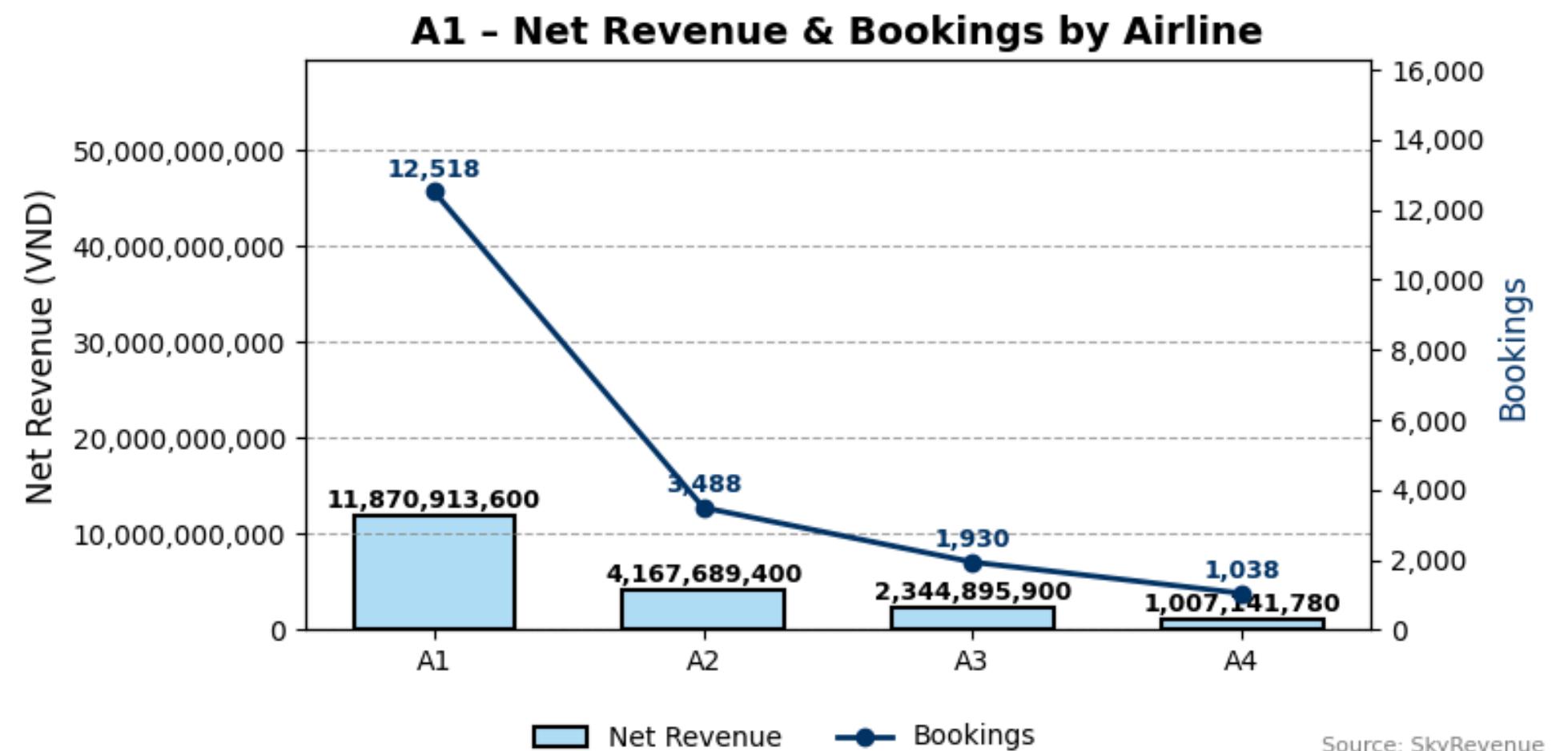


Nhận định chính:

- Insight:** Dù HCM–Hà Nội có số booking cao nhất (1 551) và doanh thu lớn (~1,98 tỷ đ), tuyến Hà Nội–HCM với booking ít hơn (1 377) lại chỉ kém ~10 % doanh thu, cho thấy giá vé chiều HN–HCM cao hơn trung bình.
- **Chú tâm:** Cân nhắc điều chỉnh giá chiều HCM→HN để cân bằng doanh thu hai chiều, hoặc khuyến khích upsell dịch vụ phụ trợ trên chiều ít booking (HN→HCM) nhằm tận dụng mức giá cao.

Tuyến có tỉ lệ hủy cao:

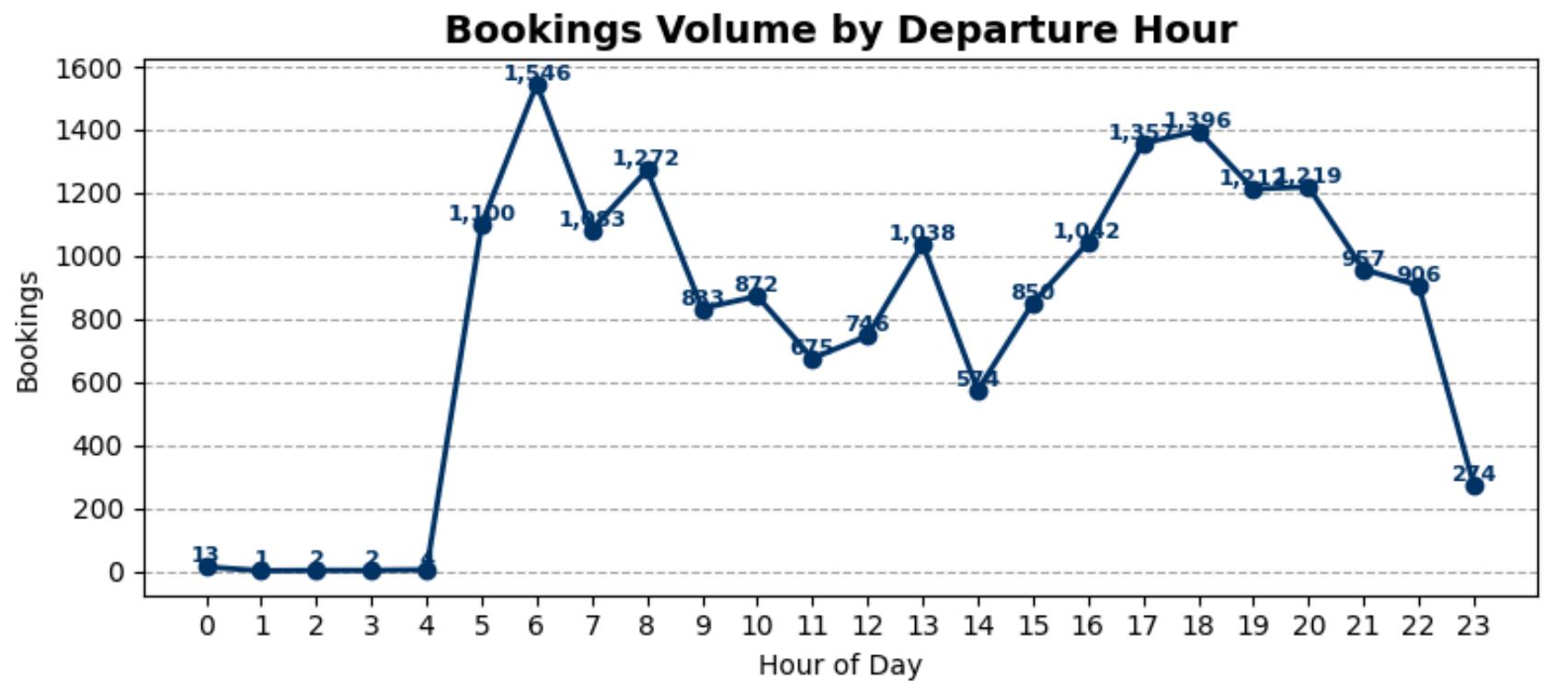
- Insight:** Đăk Lăk–Thanh Hóa có tỉ lệ hủy lên tới 50 % (4 booking, 2 hủy), Khánh Hòa–Kiên Giang 27,3 %.
- **Chú tâm:** Điều tra nguyên nhân hủy (thời tiết, khung giờ, linh hoạt vé) trên các tuyến này; xem xét đưa vào điều khoản giá vé linh hoạt (flex fare) hay chính sách đổi/hủy rõ ràng để giảm mất cân bằng đặt/hủy.



- **Insight:** A1 chiếm tới ~57 % tổng doanh thu ròng (≈ 11.87 tỷ đ) và 12 518 booking, gấp $\sim 3\times$ số booking của hãng kế tiếp (A2) và gấp $\sim 12\times$ so với A4.

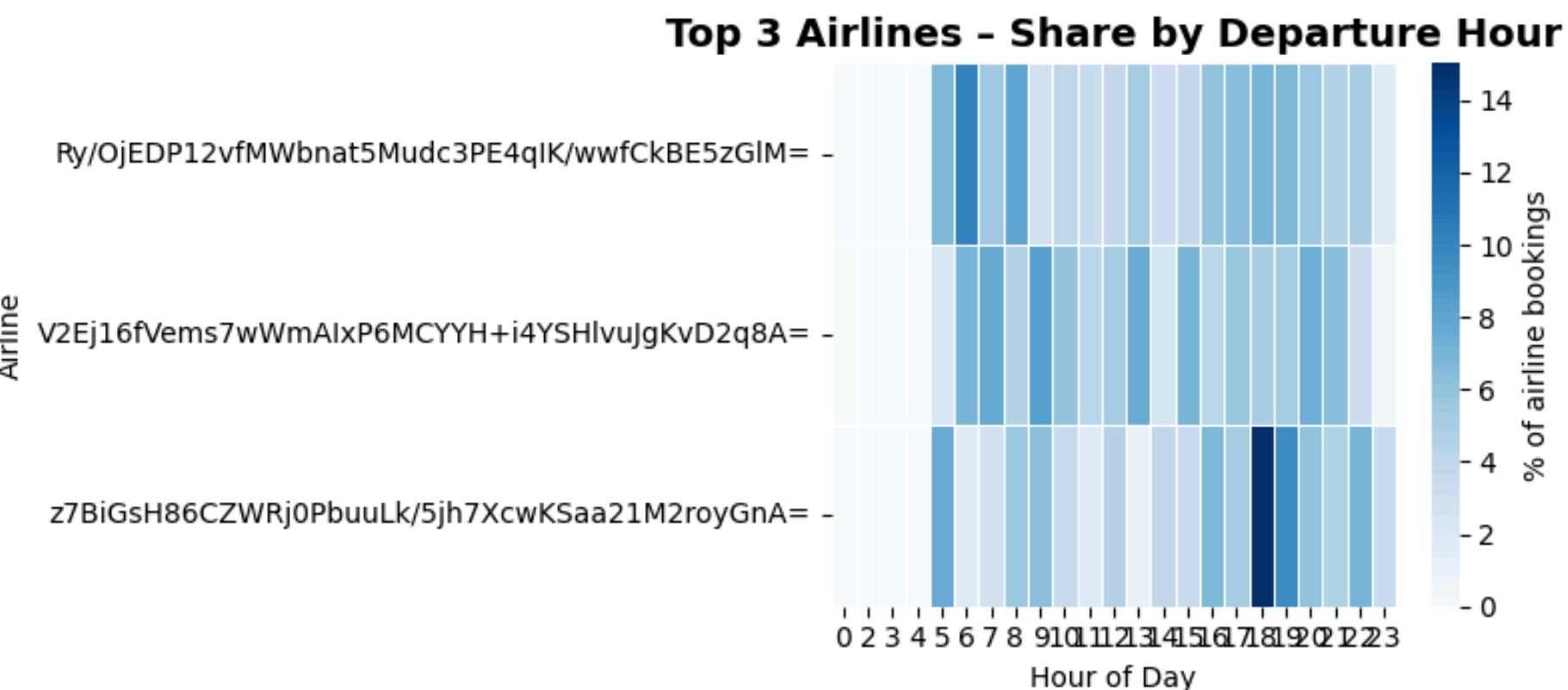
→ **Chú tâm:**

- **Giữ chân A1:** Duy trì mối quan hệ ưu tiên với A1 (gói hợp tác, phí hoa hồng thấp, ưu đãi marketing chung) để không bị mất thị phần chính.
- **Đa dạng hóa đối tác:** Đàm phán cải thiện lợi ích với A2–A4 (ví dụ tăng tần suất chuyến hoặc giờ bay phù hợp) nhằm cân bằng rủi ro phụ thuộc quá lớn vào một hãng duy nhất.
- **Tối ưu phân phối:** Xem xét chiến lược giá linh hoạt riêng cho từng hãng dựa trên lịch bay, hot route và tỷ lệ phủ chuyến của họ, đồng thời phân bổ ngân sách khuyến mãi hợp lý để thúc đẩy A2–A4.



- Tối ưu giá & ưu đãi khung giờ cao điểm:** Áp dụng giá linh hoạt (dynamic pricing) cao hơn hoặc upsell dịch vụ phụ trợ (hành lý, bảo hiểm) cho các chuyến khởi hành vào 5–8h sáng và 17–19h để tận dụng nhu cầu lớn.
- Khuyến khích đặt giờ ngoài peak:** Cung cấp “off-peak discount” nhẹ hoặc voucher ăn uống tại sân bay cho khách chọn giờ bay giữa ngày (9–12h) hoặc muộn (20–22h), giúp dàn trải tải và tăng công suất sử dụng.

- Insight:** Giờ bay “vàng” tập trung vào các khung 5–8h sáng và 17–19h chiều, với đỉnh cao nhất ở 6h sáng (~1 546 booking) và 18h chiều (~1 396 booking).



- Quảng bá các khung giờ ít được ưa thích:** Sử dụng email/SMS cá nhân hóa nhắc nhở khách có lịch linh hoạt, kèm ưu đãi gói combo dịch vụ phụ trợ, để thu hút thêm booking vào khung 0–4h và 13–16h, giảm áp lực cao điểm.

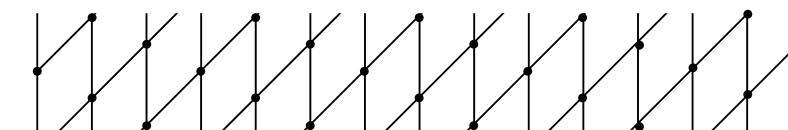
Kết luận

- **Đặc điểm chung:**

- Tuyến HCM–HN đứng đầu về booking và doanh thu; chiều ngược lại có giá cao hơn nên lượt đặt ít hơn.
- Một số tuyến nhỏ (Đắk Lăk–Thanh Hóa, Khánh Hòa–Kiên Giang) có tỉ lệ hủy rất cao (50 %, 27 %), gây lãng phí chỗ.
- Giờ bay “vàng” rơi vào 5–8h sáng và 17–19h chiều, tạo áp lực cao điểm.

- **Hướng tiếp theo:**

- **Điều chỉnh giá hai chiều:** Tăng giá hoặc upsell dịch vụ trên chiều HN→HCM để tận dụng mức giá cao, đồng thời cân bằng tải.
- **Giảm hủy tuyến yếu:** Điều tra và áp dụng chính sách vé linh hoạt (flex fare) hoặc ưu đãi đổi/hủy để giảm tỉ lệ hủy trên tuyến Đắk Lăk–Thanh Hóa và Khánh Hòa–Kiên Giang.
- **Quản lý giờ bay peak:** Áp dụng dynamic pricing cao điểm và khuyến khích giờ ngoài peak bằng off-peak discount hoặc voucher sân bay để dàn trải hành khách.



Vậy tập trung ưu tiên gì?



Phát triển Hệ thống Dynamic Pricing:

Tích hợp bộ cơ chế định giá linh hoạt theo tuyến & khung giờ (5-8h, 17-19h) dựa trên độ co giãn giá và nhu cầu thực tế, giúp tối ưu doanh thu hai chiều và dàn trải công suất.



Nâng cấp Ứng dụng Mobile (iOS/Android):

Bổ sung tính năng “early-bird deal” & “last-minute alerts”, one-click checkout, upsell hành lý/bảo hiểm ngay trong flow mua vé để tận dụng kênh di động, nơi chiếm >65 % doanh thu.



Ra mắt Chương trình Loyalty & Upsell Bundle:

Xây dựng loyalty program với điểm thưởng, ưu đãi cá nhân hóa cho nhóm “At Risk” và “Potential Loyalist”; đồng thời phát triển gói combo vé + hành lý/bảo hiểm để gia tăng giá trị đơn hàng.

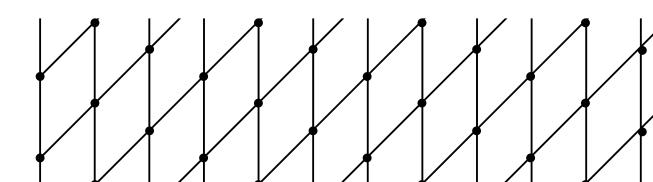
Trả lời cho: Câu 3



Define Datasets & Tools

Datasets cần thu thập

- User Interaction Logs: event_id, user_id, timestamp, screen_id, event_type, metadata (thời gian ở màn hình, click, error).
 - → **Conversion Rate: Đếm event_type=search_start vs payment_complete**
 - → **TTP: Lấy timestamp(payment_complete) – timestamp(search_start)**
 - → **Seat Success: Đếm event_type=seat_selected_success vs seat_selection_attempt**
- Booking Transactions: booking_id, user_id, created_at, payment_timestamp, status, price, discount_amount. Ví dụ:
(b456, u123, 2025-04-28T10:02:00, 10:05:00, "completed", 1438000, 20000)
 - → **Conversion Rate: chỉ tính các booking.status=completed trên tổng user search**
 - → **TTP: payment_timestamp – created_at**
- seat_selection: selection_id, booking_id, user_id, attempt_time, success_flag. Ví dụ: (s789, b456, u123, 2025-04-28T10:03:00, TRUE)
 - → **Seat Selection Success Rate: % success_flag=TRUE trên tổng attempts**
- User Profiles & Segmentation: session_id, user_id, start_time, end_time, channel VD: (sess1, u123, 2025-04-28T10:00, 10:10, "iOS")
 - **Giúp gộp event_logs trong một phiên để tính CR và TTP chính xác theo session**
 - **Phân tích CR/TTP theo kênh iOS/Android/Web**
- seat_availability: availability_id, train_id, seat_id, timestamp, is_available. vd: (a101, "SE4", "Car1-11", 2025-04-28T10:02:50, TRUE)
 - → **Xác định ngay thời điểm nào ghế vẫn trống để giảm lỗi "thiếu chỗ" trước khi user chọn (giúp nâng Seat Selection Success Rate)**



Define Datasets & Tools

Công cụ phân tích

1. Ingestion & Streaming

- Kafka: Hứng real-time event từ app/web.
- Kafka Connect: Đẩy dữ liệu vào Data Lake.

2. Storage & Processing

- Data Lake: GCP GCS (raw events, transactions).
- Data Warehouse: Snowflake hoặc BigQuery (cho SQL analytics nhanh, rẻ khi cần scale).

3. ETL/ELT & Orchestration

- Airflow: Lên lịch pipeline ingestion → transformation.
- dbt: Làm sạch dữ liệu (staging → marts), tracking version dữ liệu.

4. BI & Product Analytics

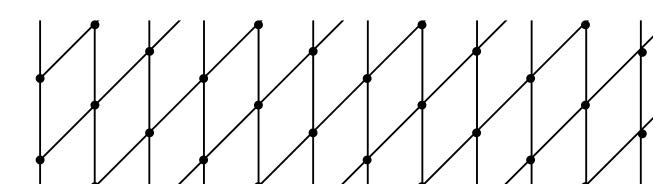
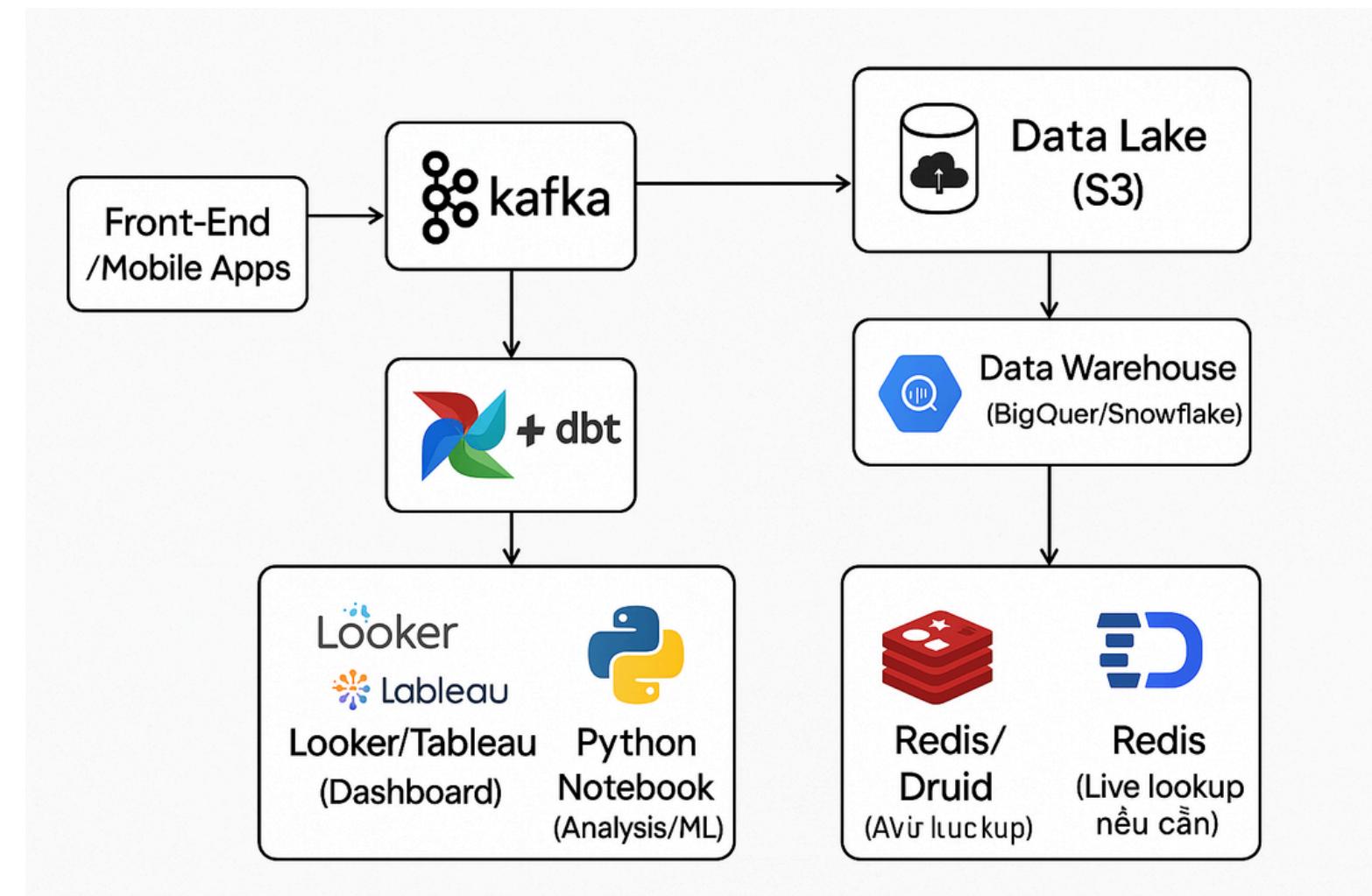
- Looker Studio (Google Data Studio free): Dashboard funnel, booking CR, cohort.
- Mixpanel hoặc Amplitude: Track user funnel, retention, A/B test.

5. Realtime Features (nếu cần)

- Redis hoặc Druid: Cache lookup real-time (số ghế trống, giá vé live).
- Kafka Streams: Nếu cần tính nhanh các chỉ số active user.

6. Machine Learning (giai đoạn sau)

- Python (pandas, scikit-learn): Chạy cohort analysis, A/B test, phân tích uplift.
- (Option) MLflow: Khi cần track & deploy mô hình như dynamic pricing.



Top 3 Success Metrics

Trước hết, cần tập trung vào tối ưu trải nghiệm end-to-end, giảm friction và “điểm chết” trong phễu từ lúc search đến khi hoàn tất thanh toán—nhất là bước chọn chỗ, vốn dễ làm người dùng bỏ dở. Vì vậy, ba chỉ số quan trọng nhất là:

Conversion Rate (CR) từ Search → Booking hoàn tất

- **Định nghĩa:** % user hoàn tất thanh toán trên tổng user bắt đầu search. → dùng để xác định điểm chưa tối ưu để tối ưu hóa
- **Tầm quan trọng:** Đo lường hiệu quả end-to-end của flow; ảnh hưởng trực tiếp đến doanh thu.

Time to Purchase (TTP)

- **Định nghĩa:** Thời gian trung bình (phút/giờ) từ lần search đầu đến confirm payment, tỷ lệ thuận với friction: càng nhanh, CR càng cao.
- **Tầm quan trọng:** Giảm friction, tăng sự hài lòng; UX nhanh gọn giúp nâng CR.

Seat Selection Success Rate

- **Định nghĩa:** % booking chọn được ghế không lỗi, là thước đo độ tin cậy và mượt mà của bước then chốt, tránh drop-off và hủy vé.
- **Tầm quan trọng:** Bước chọn chỗ là nút thắt; lỗi cao sẽ tạo drop-off và hủy vé.

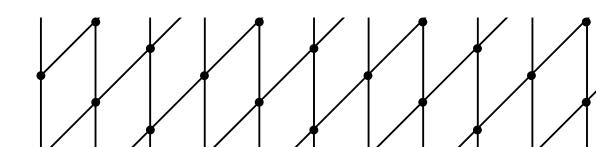
3.3 Improving One Key Metric

Cải thiện Conversion Rate

- **Lý do chọn:** CR là chỉ số “đầu phễu” quyết định tổng số booking, ảnh hưởng ngay đến doanh thu.
-> Kết quả mong đợi: Rút ngắn TTP, giảm lỗi chọn ghế, cá nhân hóa flow theo segment, từ đó gia tăng Conversion Rate và doanh thu tổng thể.

1. Funnel Analysis

- **Mục tiêu:** Xác định chính xác “điểm chết” trong quy trình booking.
- **Xác định các bước phễu**
 1. Search Start: user mở form nhập ga đi/ga đến.
 2. Date Selected: user chọn ngày (1 chiều hoặc khứ hồi).
 3. Passenger Info: user chọn số hành khách & ưu đãi.
 4. Show Results: user xem danh sách chuyến.
 5. Seat Selection: user attempt chọn chỗ.
 6. Checkout: user nhập thông tin cá nhân + phương thức thanh toán.
 7. Payment Complete: booking hoàn tất.
- **Cách làm**
 - **Dùng báo cáo funnel trên Looker/Tableau: tính tỷ lệ (%) user chuyển từ bước i → i+1.**
 - **Ví dụ: nếu 10.000 user Search Start, nhưng chỉ 5.000 user tới Checkout, ta có drop-off 50% ở bước giữa.**
- **Kết quả mong đợi:** Biết được bước nào drop-off > X% (ví dụ 30%), tập trung phân tích thêm tại đó.



2. Segmentation Analysis

- **Mục tiêu:** Hiểu xem nhóm user nào “mắc kẹt” nhiều nhất, để ưu tiên cải thiện.

- **Xác định các bước phễu**

- 1.Kênh: iOS / Android / Web
- 2.Nhóm tuổi: <25, 25–34, 35–44, v.v.
- 3.Hạng khách: student / VIP / union member
- 4.Route type: High-demand (HCM–HN) vs Low-demand (SGN–QNG)

- **Cách làm**

- Tạo bảng pivot / crosstab CR theo mỗi phân khúc (ví dụ CR_iOS = 20%, CR_Android = 15%).
- So sánh CR giữa phân khúc để thấy: “Android CR thấp hơn iOS 5 điểm % → ưu tiên UX Android.”
- Lọc tiếp theo route, theo nhóm tuổi để thấy đâu mới là bottleneck thực sự.

3. Seat Availability Audit

- **Mục tiêu:** Giảm lỗi “thiếu chỗ” – một nguyên nhân lớn khiến user bỏ.

- **Xác định tỷ lệ lỗi**

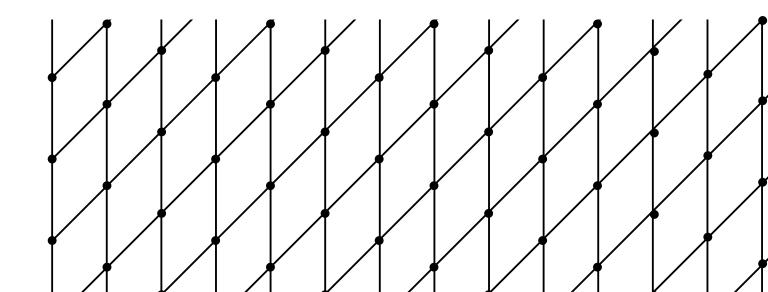
- Tính # seat_selection_attempts và # seat_selection_failures (metadata.error_code=“no_seat”).
- Seat Success Rate = (attempts – failures)/attempts.

- **Phân tích theo chiều**

- Theo train_id, car_number, giờ cao điểm (5–8h, 17–19h) để xem đâu xảy ra nhiều lỗi nhất.
- Ví dụ: Car2 trên tàu SE4 vào 18h có thất bại 10% lượt chọn → cần refresh availability nhanh hơn.

- **Hành động**

- Xác định hệ số TTL (time-to-live) của cache seat availability.
- Thống kê time lag giữa cập nhật DB vs phản hồi user; đặt threshold (ví dụ max 2s).

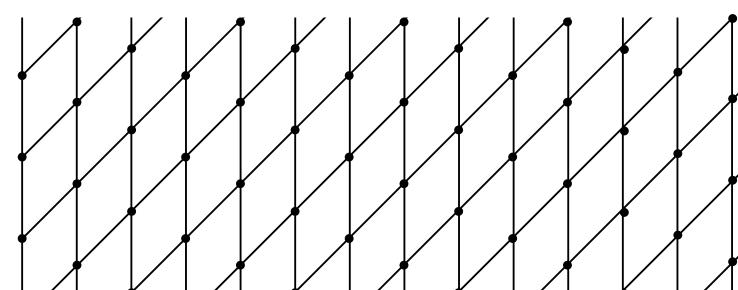


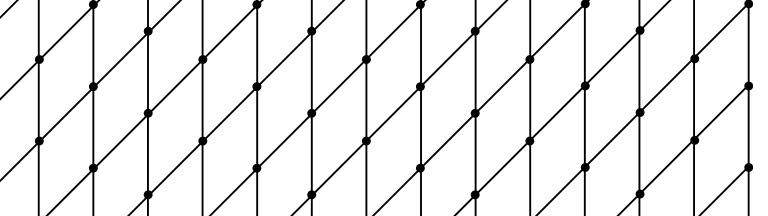
4. A/B Testing (UX Experiments)

- **Mục tiêu:** Đo đạc chính xác uplift trong CR khi thay đổi UI/UX.
- **Thiết kế test**
 - Control: Giao diện chọn ghế / form hiện tại.
 - Variant: Giao diện mới (ví dụ highlight chỗ trống, chỉ hiển thị ghế giá tốt trước).
- **Metric đo lường**
 - Primary: Conversion Rate (variant vs control).
 - Secondary: Seat Selection Success Rate, Time to Purchase.
- **Phương pháp**
 - Phân bổ ngẫu nhiên user vào hai nhóm; đảm bảo đủ sample size (ví dụ mỗi nhóm ≥ 5.000 user).
 - Dùng Mixpanel/Amplitude hoặc Looker để theo dõi kết quả real-time.
 - Chạy test cho đến khi đạt statistical significance ($p\text{-value} < 0.05$).

5. Dashboard & Alerting

- **Mục tiêu:** Giám sát liên tục, phát hiện sớm khi Conversion Rate tụt.
- **Dashboard Key Views**
 - Funnel Overview: % từng bước, so sánh theo ngày/tuần.
 - CR by Segment: biểu đồ bar tại từng kênh hoặc route.
 - Seat Success Heatmap: lỗi seat fail theo giờ và train.
 - TTP Distribution: box-plot thời gian đặt vé.
- **Alerting**
 - Đặt threshold (ví dụ $CR < 15\%$) \rightarrow tự động cảnh báo qua Slack/Email.
 - Giám sát latency của seat_availability API; nếu $> 2s \rightarrow$ alert.





Thank You

For Your Attention

