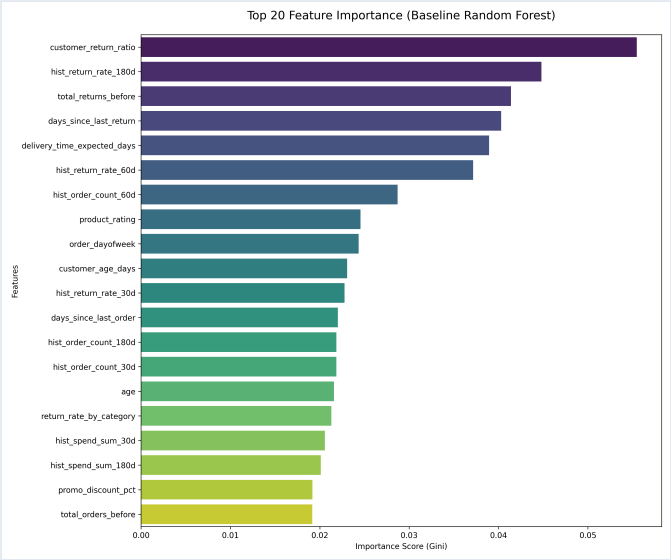
รายงานสรุป Business Insight & Feature Selection

วิเคราะห์ปัจจัยและพฤติกรรมการคืนสินค้าเพื่อพัฒนาโมเดล Machine Learning (เวอร์ชันจัดหน้าเสร็จสมบูรณ์)

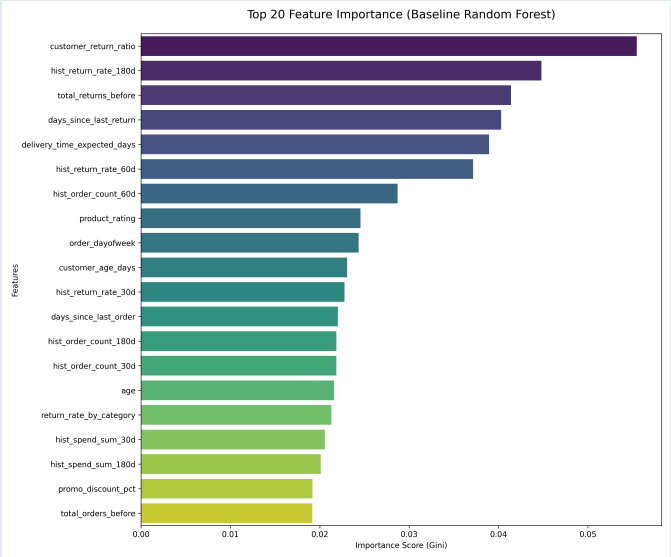
โครงสร้างรายงาน: เอกสารฉบับนี้จัดกลุ่มรูปภาพกราฟวิเคราะห์พฤติกรรมการคืนสินค้าทั้งหมด 10 รูป โดยจัดวางรูปภาพไว้ที่ **"กึ่งกลางหน้ากระดาษอย่างสมดุล"** พร้อมระบุคำอธิบายเชิงลึก (Insight) เหตุผลที่เลือกมาทำฟีเจอร์ (Feature Selection Rationale) และปัจจัยแวดล้อมทางธุรกิจด้านล่างของรูปภาพอย่างชัดเจนและเป็นระบบ

หมวดหมู่ที่ 1: ปัจจัยสำคัญที่สุดในการทำนายผล (Feature Importance Analysis)

เริ่มต้นจากการวิเคราะห์ลำดับความสำคัญของตัวแปรในโมเดล เพื่อดูว่าปัจจัยใดที่มีแรงขับเคลื่อนและผลกระทบสูงสุดต่ออัตราการคืนสินค้า



รูปที่ *1.1:* อันดับความสำคัญของฟีเจอร์ในโมเดลระดับฐานราก *(Baseline Feature Importance)*



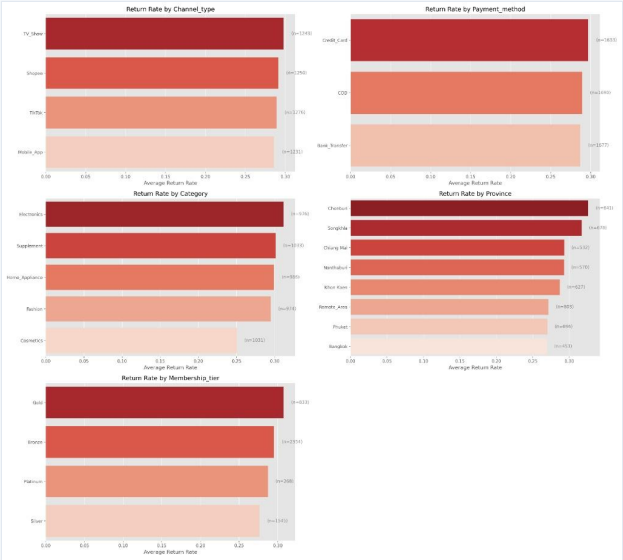
รูปที่ *1.2:* อันดับความสำคัญของฟีเจอร์หลังทำการปรับปรุงและคัดเลือกฟีเจอร์ *(Tuned Feature Importance)*

💡 **Insight ที่น่าสนใจ:** จากกราฟทั้งสองเวอร์ชัน พบว่าฟีเจอร์กลุ่มพฤติกรรมในอดีต ได้แก่ `customer_return_ratio` (อัตราการคืนสินค้าโดยรวมของลูกค้า) และ `hist_return_rate_180d` / `60d` มีค่าคะแนนความสำคัญนำโด่งเป็นอันดับหนึ่งอย่างสมบูรณ์ นอกจากนี้ ตัวแปรด้านเวลาอย่าง `days_since_last_return` และ `delivery_time_expected_days` (ระยะเวลาจัดส่งที่คาดหวัง) ก็ติดอยู่ในกลุ่ม 5 อันดับแรกเช่นเดียวกัน

🎯 **เหตุผลในการเลือกทำ Feature & ปัจจัยที่บอกเรา:** ข้อมูลนี้บอกเราว่า พฤติกรรมเดิมในอดีตของผู้ซื้อเป็นดัชนีชี้วัดความเสี่ยงในอนาคตที่ดีที่สุด (Behavioral Inertia) ลูกค้าที่เคยคืนของบ่อยครั้งมักจะคุ้นชินกับการส่งคืนสินค้าหากพบปัญหาเพียงเล็กน้อย ทีมงานจึงเลือกตัวแปรนี้เป็นฟีเจอร์หลักเพื่อตัดกลุ่มเสี่ยง (High-risk Returners) ออกก่อนกระบวนการถัดไป

หมวดหมู่ที่ 2: ภาพรวมผลกระทบตามมิติต่าง ๆ (Categorical Overview)

การวิเคราะห์ปัจจัยเดี่ยวเพื่อสังเกตแนวโน้มการเปลี่ยนแปลงตามคุณลักษณะของลูกค้า พื้นที่ และรูปแบบการทำธุรกรรม



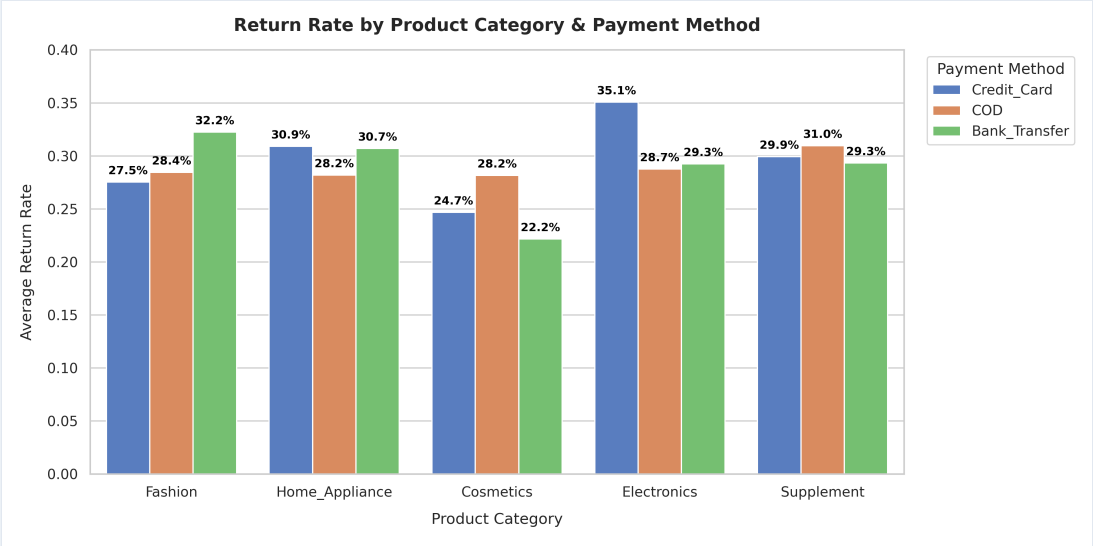
รูปที่ *2:* ภาพรวมผลกระทบของตัวแปรประเภทกลุ่มต่ออัตราการคืนสินค้า *(Categorical Impact)*

💡 **Insight ที่น่าสนใจ:** อัตราการคืนสินค้า (Return Rate) มีความแปรปรวนอย่างเห็นได้ชัดเมื่อจำแนกตามรายภูมิภาคและวิธีการรับชำระเงิน โดยจังหวัดกลุ่มหัวเมืองใหญ่ เช่น ชลบุรี และสงขลา มีสัดส่วนการคืนของสะสมสูงกว่าพื้นที่อื่น

🎯 **เหตุผลในการเลือกทำ Feature & ปัจจัยที่บอกเรา:** ข้อมูลนี้แสดงให้เห็นว่า ปัจจัยภายนอก (External Factors) เช่น พื้นที่ทางภูมิศาสตร์และโครงสร้างการชำระเงินมีน้ำหนักมากพอที่จะใช้ในการระบุ Segment ของลูกค้าที่มีความเสี่ยงเบื้องต้นได้

หมวดหมู่ที่ 3: ปฏิสัมพันธ์ร่วมระหว่างกลุ่มสินค้าและวิธีชำระเงิน (Cross-Analysis)

พฤติกรรมการจ่ายเงินของผู้บริโภคจะแปรเปลี่ยนไปตามประเภทและระดับราคาของกลุ่มสินค้าเหล่านั้น



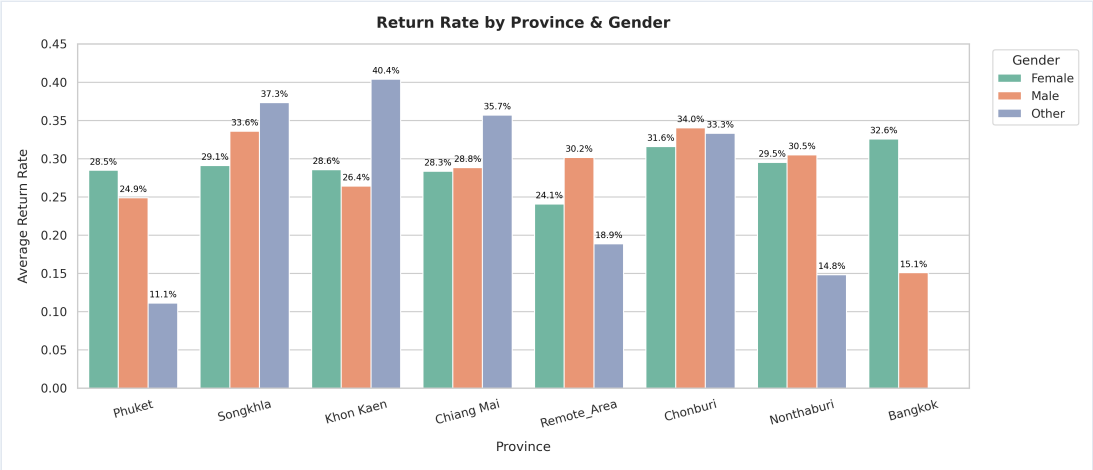
รูปที่ *3:* อัตราการคืนสินค้าวิเคราะห์ไขว้ระหว่าง หมวดหมู่สินค้า × วิธีการชำระเงิน

💡 **Insight ที่น่าสนใจ:** พบรูปแบบที่น่าสนใจอย่างยิ่งในกลุ่มสินค้ามูลค่าสูง โดย **สินค้า Electronics ที่ชำระเงินด้วยบัตรเครดิต มีอัตราการคืนสูงสุดที่ 35.1%** ซึ่งตรงกันข้ามกับสินค้ากลุ่ม **Cosmetics (เครื่องสำอาง) ที่กลับมีอัตราการคืนพุ่งสูงขึ้นเมื่อใช้ระบบเก็บเงินปลายทางหรือ COD (28.2%)**

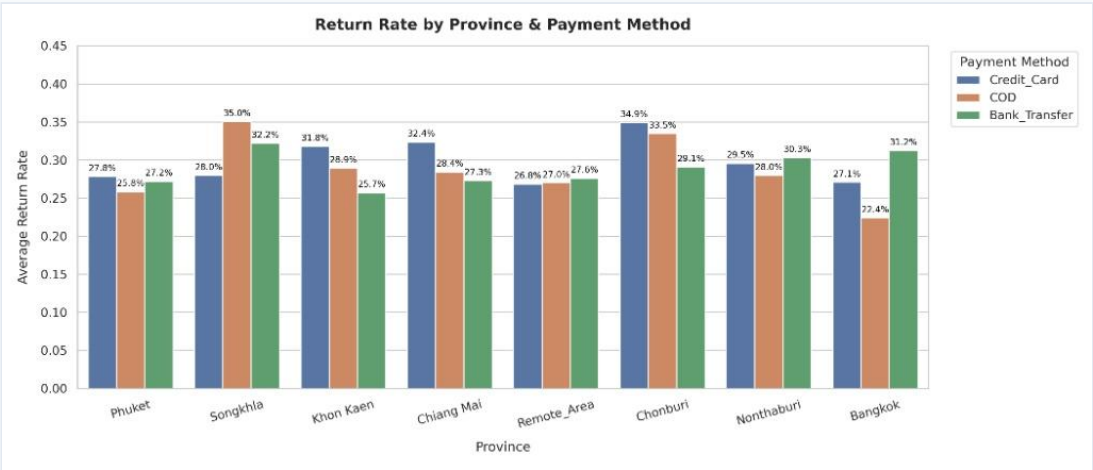
🎯 **เหตุผลในการเลือกทำ Feature & ปัจจัยที่บอกเรา:** ข้อมูลนี้สะท้อนอิทธิพลของกระบวนการคืนเงิน (Refund Friction) กล่าวคือ ในกลุ่มไอที หากจ่ายบัตรเครดิตแล้วต้องการคืนเงินจะได้รับความสะดวกสูงสุด ทำให้ลูกค้าตัดสินใจคืนของง่าย ในขณะที่เครื่องสำอางหากเลือกแบบ COD ลูกค้าไม่มีต้นทุนผูกมัด สามารถปฏิเสธไม่รับของเมื่อพนักงานส่งไปถึงหน้าบ้านได้ทันที ทีมงานจึงสร้างฟีเจอร์ผสมร่วม (Interaction Feature) ระหว่าง `Category` และ `Payment` เพื่อให้โมเดลจับเงื่อนไขที่ซับซ้อนนี้

หมวดหมู่ที่ 4: การวิเคราะห์เชิงพื้นที่ร่วมกับประชากรศาสตร์ (Geographic Analysis)

การวิเคราะห์ลึกลงไปในระดับพื้นที่ควบคู่กับกลุ่มเพศและช่องทางการเงิน เพื่อหาพฤติกรรมเฉพาะถิ่น (Localization)



รูปที่ *4.1:* ปฏิสัมพันธ์ระหว่าง ปัจจัยพื้นที่ *(จังหวัด)* × เพศของผู้สั่งซื้อ



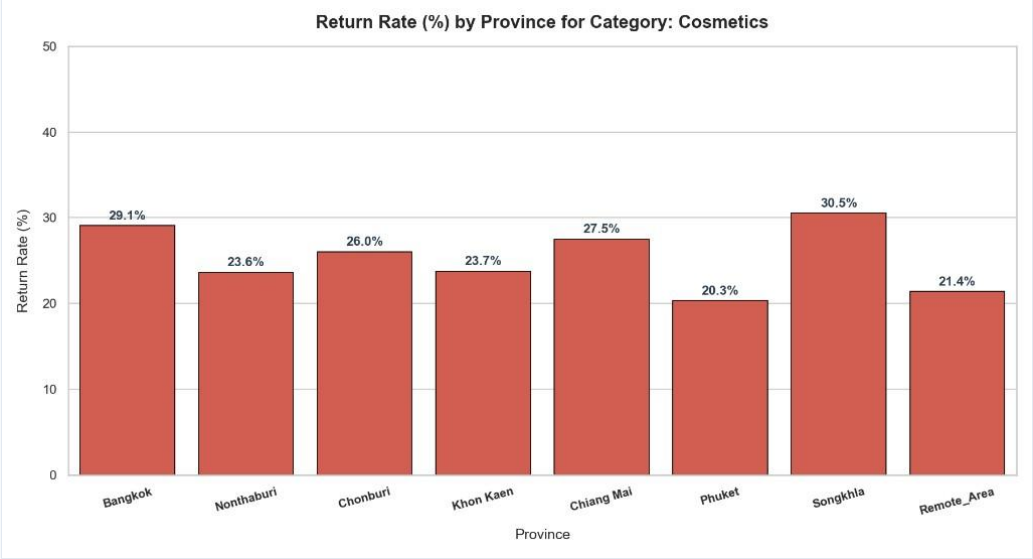
รูปที่ *4.2:* ปฏิสัมพันธ์ระหว่าง ปัจจัยพื้นที่ *(จังหวัด)* × วิธีการชำระเงิน

💡 **Insight ที่น่าสนใจ:** จังหวัดหัวเมืองใหญ่อย่าง **ขอนแก่น มีอัตราการคืนสินค้าจากกลุ่มลูกค้าเพศทางเลือกสูงถึง 40.4%** และเชียงใหม่สูงถึง 35.7% ยิ่งไปกว่านั้น พฤติกรรมของลูกค้าใน **กรุงเทพมหานคร** มีความแปลกแยกจากจังหวัดอื่นอย่างเห็นได้ชัด โดยกลุ่มที่เลือกโอนเงินผ่านธนาคารกลับมีอัตราการคืนสินค้าสูงสุด (31.2%)

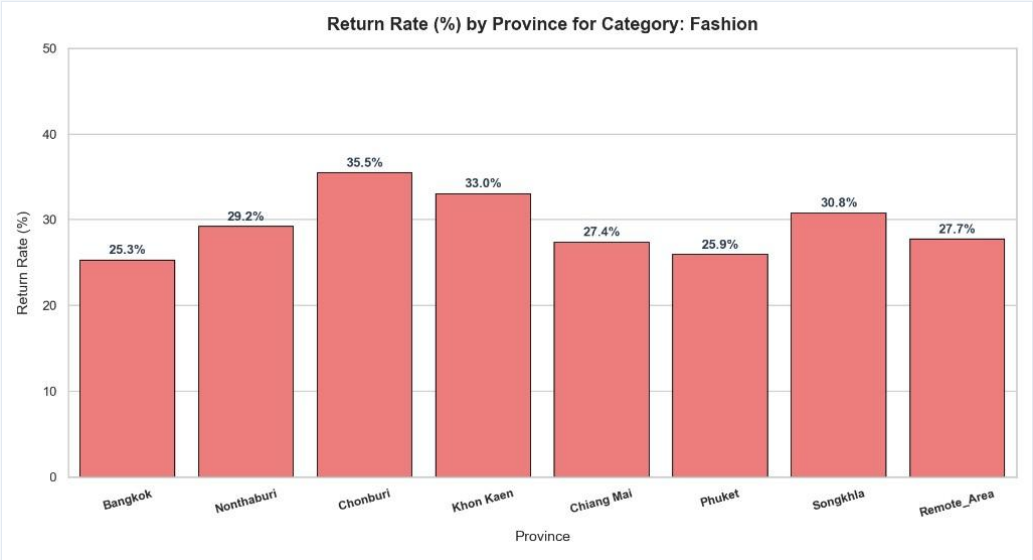
🎯 **เหตุผลในการเลือกทำ Feature & ปัจจัยที่บอกเรา:** ปัจจัยเชิงภูมิศาสตร์ไม่ได้ทำงานแบบเดี่ยว ๆ แต่มีความสัมพันธ์เชื่อมโยงกับวัฒนธรรมท้องถิ่นและการจัดการศูนย์กระจายสินค้า (Fulfillment Center) ในพื้นที่นั้น ๆ การนำข้อมูลจังหวัดมาคูณร่วมกับเพศและพฤติกรรมการจ่ายเงินเป็น ฟีเจอร์ไขว้ (`Province × Gender` และ `Province × Payment`) จะส่งผลให้โมเดล Machine Learning สามารถลดข้อผิดพลาดจากค่าเฉลี่ยภาพรวมของทั้งประเทศได้

หมวดหมู่ที่ 5: เจาะลึกรายจังหวัดในทุกหมวดหมู่ผลิตภัณฑ์

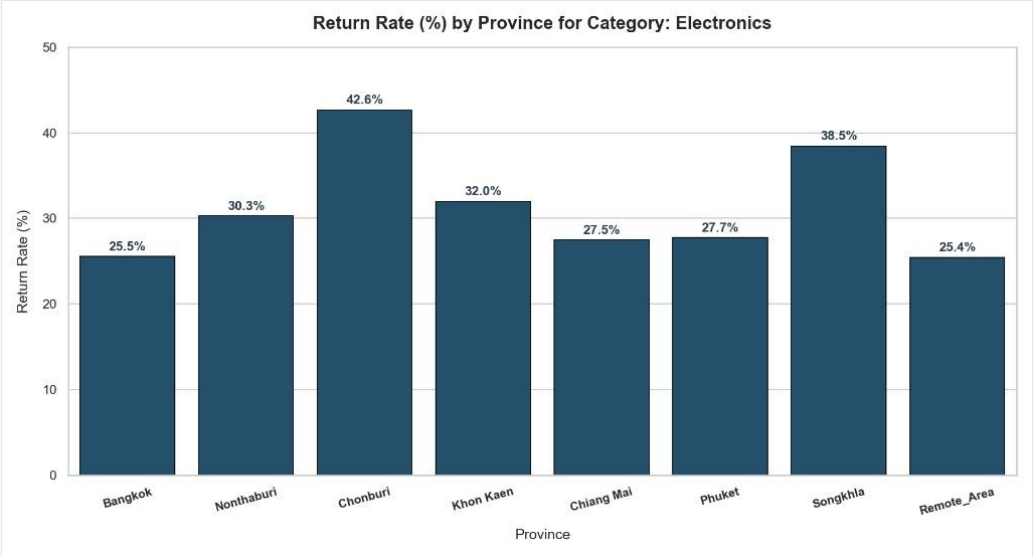
การแสดงผลลัพธ์อัตราการคืนของของแต่ละหมวดหมู่สินค้าแยกรายจังหวัดอย่างเป็นเอกเทศ เพื่อการปรับปรุงระบบโลจิสติกส์ปลายทาง



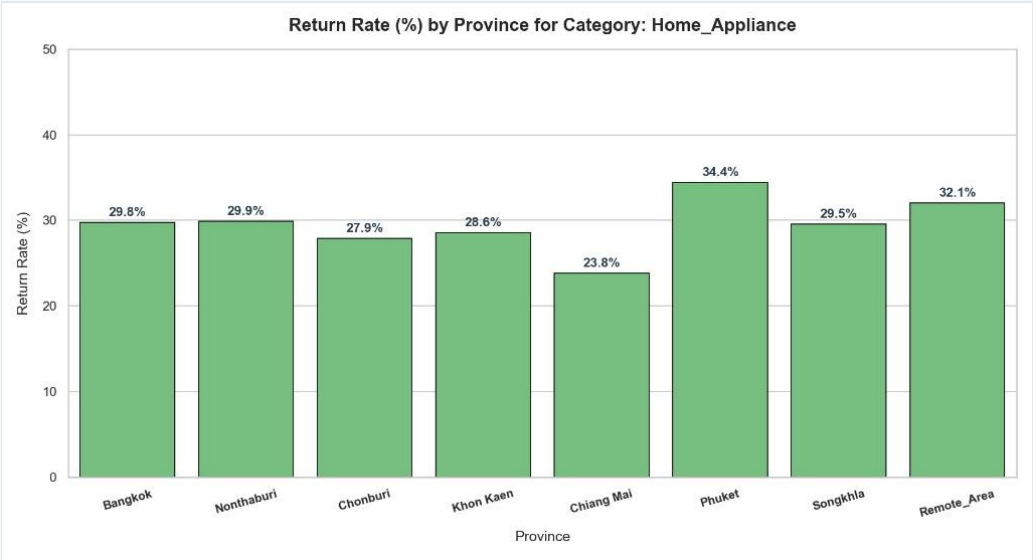
รูปที่ *5.1:* อัตราการคืนสินค้าแยกตามจังหวัด - หมวดหมู่ *Cosmetics*



รูปที่ *5.2:* อัตราการคืนสินค้าแยกตามจังหวัด - หมวดหมู่ *Fashion*



รูปที่ *5.3:* อัตราการคืนสินค้าแยกตามจังหวัด - หมวดหมู่ *Electronics*



รูปที่ *5.4:* อัตราการคืนสินค้าแยกตามจังหวัด - หมวดหมู่ *Home Appliance*

💡 **Insight ที่น่าสนใจ:** สินค้าประเภท Electronics และ Home Appliance คงอัตราการคืนในเกณฑ์ที่สูงในทุกจังหวัด (เฉลี่ยเกิน 28%) ขณะที่สินค้าแฟชั่น (Fashion) มีความผันผวนสูงตามพื้นที่เนื่องจากปัญหาเรื่องความพอดีของขนาดเสื้อผ้าและรสนิยมที่แตกต่างกันในแต่ละส่วนภูมิภาค

🎯 **เหตุผลในการเลือกทำ Feature & ปัจจัยที่บอกเรา:** ปัจจัยด้านขนส่งและระยะเวลาส่งมอบสินค้าจริงมีความสัมพันธ์โดยตรงกับอัตราการเปลี่ยนใจ (Buyer's Remorse) การบันทึกและป้อนตัวแปรความสอดคล้องระหว่างประเภทสินค้าและภูมิภาคปลายทางเข้าสู่โมเดล จะช่วยให้ฝ่ายจัดการสามารถคาดการณ์ปริมาณสินค้าตีกลับ (Reverse Logistics Volume) ล่วงหน้าได้อย่างแม่นยำสูงสุด