

Data Science CMU Project Contest 2024

รายงานสรุปข้อมูลยอดขนส่งในสองไตรมาสแรกปี 2024

Mini Project Report

จัดทำโดย

นาย ภัทรดนัย สอนสว่าง

นางสาว นิภัสสา ชัยนิวัฒนา

นาย พุฒิพงศ์ โยมะบุตร

เสนอ

ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่

บทคัดย่อ

โครงการนี้มีวัตถุประสงค์เพื่อวิเคราะห์และศึกษาข้อมูลการส่งออกสินค้าของประเทศไทยในสองไตรมาสแรกของปี 2024 โดยใช้เทคนิคการวิเคราะห์ข้อมูลและการประมวลผลทางสถิติ เพื่อระบุแนวโน้มการเปลี่ยนแปลงของการส่งออกในกลุ่มสินค้าต่าง ๆ รวมถึงการนำข้อมูลที่น่ามาวิเคราะห์ไปหาปัจจัยในการส่งออกสินค้าโดยข้อมูลที่ใช้ในการศึกษาครั้งนี้ได้รับจากแหล่งข้อมูลจาก Data Science CMU Project Contest และผ่านกระบวนการทำความสะอาดข้อมูล (Data Cleaning) เพื่อความแม่นยำในการวิเคราะห์ ผลการศึกษาจะช่วยให้เห็นภาพรวมของโครงสร้างการส่งออกสินค้า และสามารถนำไปใช้ในการวางแผนกลยุทธ์ของบริษัทในอนาคตได้อย่างมีประสิทธิภาพ

คำนำ

รายงานฉบับนี้เป็นส่วนหนึ่งของโครงการ Data Science CMU Project Contest 2024 จัดทำขึ้นเพื่อใช้อธิบายประกอบโครงการรายงานสรุปข้อมูลยอดขนส่งในสองไตรมาสแรกของปี 2024 ซึ่งประกอบไปด้วยขั้นตอนการทำความสะอาดข้อมูล การหาข้อมูลเชิงลึกจากข้อมูล และการนำเสนอข้อมูลในรูปแบบของ Dashboard ผ่านโปรแกรม Power BI เพื่อนำข้อมูลนี้ไปใช้วางแผนพัฒนาด้านการให้บริการขนส่งต่อไป

ทางคณะผู้จัดทำขอขอบพระคุณ ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่ ในการจัดกิจกรรมนี้สุดท้ายนี้ทางคณะผู้จัดทำหวังว่ารายงานฉบับนี้จะสามารถเป็นประโยชน์ไม่มากนักแก่ผู้อ่านทุกท่าน

ขอขอบพระคุณ

คณะผู้จัดทำ

วิธีการวิจัย

1. ใช้ Jupyter Notebook ในการเขียนโค้ด Python
2. Import Library ที่นำมาใช้งาน ได้แก่ pandas, numpy, re, thefuzz, spacy
3. โหลดชุดข้อมูล Dataset ผ่านไฟล์ CSV โดยใช้ library pandas
4. เปลี่ยนชื่อคอลัมน์ เพื่อง่ายต่อความเข้าใจ
5. เปลี่ยนชนิดของข้อมูลให้เหมาะสม
6. เช็คความสมบูรณ์ของข้อมูล ด้วยการเช็คค่าเฉลี่ย, จำนวนข้อมูล , จำนวนข้อมูลที่ซ้ำ, จำนวนข้อมูลที่เป็นค่าว่าง
7. ทำความสะอาดข้อมูลและปรับปรุงคุณภาพของข้อมูล โดยการใช้ RegEx กับ thefuzz ในการแปลงคำที่มีค่าความใกล้เคียงของคำใกล้เคียงกัน เพื่อให้ได้ข้อมูลที่มีคุณภาพและสามารถนำข้อมูลที่ถูกต้องนำไปวิเคราะห์ต่อไปได้อย่างมีคุณภาพ
8. สร้างคอลัมน์ Zone เพื่อใช้ในการวิเคราะห์ข้อมูลแบ่งตามภูมิภาค ในประเทศไทย
9. การตรวจสอบภูมิภาคจากจังหวัด (PROVINCE) ที่มีจำนวนข้อมูล ZIPCODE ที่ไม่ถูกต้อง โดยเลือกจังหวัดตามปริมาณการส่งสินค้า ที่ส่งผลกระทบต่อผลลัพธ์ ซึ่งคิดเป็น 1% จากข้อมูลทั้งหมด
10. นำข้อมูลที่ได้มาจากการทำความสะอาดข้อมูลไปใช้ในการสร้าง Dashboard เพื่อสรุปผลและสามารถวิเคราะห์ข้อมูล ผ่านโปรแกรม Power BI

ผลการวิจัย

1. ได้ข้อมูลที่มีคุณภาพและมีความสมบูรณ์ของข้อมูลมากขึ้น
2. สามารถสรุปผลยอดจำนวนการขนส่งในแต่ละจังหวัด, ภาค และแต่ละชนิดสินค้าที่สนใจ ผ่าน Dashboard ได้
3. สามารถเปรียบเทียบยอดจำนวนการส่งสินค้าที่เปลี่ยนแปลงในแต่ละเดือนหรือวัน ผ่าน Dashboard ได้

การอภิปรายผล สรุปและข้อเสนอแนะ

1. จังหวัดที่มียอดการขนส่งมากที่สุด 5 อันดับ ได้แก่ เชียงใหม่, กรุงเทพมหานคร, เชียงราย, ลำปาง, ลำพูน ตามลำดับ จึง เสนอแนะให้เพิ่มจำนวนสาขา และพนักงานในจังหวัดเหล่านี้
2. จังหวัดที่มียอดการขนส่งน้อยที่สุด 5 อันดับ ได้แก่ อุทัยธานี, ยะลา, ระนอง, ตราด, สตูล ตามลำดับ
3. วันอาทิตย์มียอดการส่งน้อยที่สุด โดยในวันอาทิตย์ยอดการส่งสูงที่สุดที่จังหวัดกรุงเทพมหานคร
4. จากการสรุปผลพบว่าในจังหวัดเชียงใหม่มีการส่งในวันอาทิตย์ต่ำ แต่จะมียอดการส่งสูงมากในวันจันทร์ จึง เสนอแนะให้มีการเปิดรับส่งเพิ่มในวันอาทิตย์
5. เดือนที่มียอดการส่งมากที่สุดคือเดือนมีนาคม และเดือนที่มียอดการส่งน้อยที่สุดคือเดือนเมษายน
6. สินค้าที่มียอดการส่งมากที่สุด 5 อันดับได้แก่ ผ้า, ดอกไม้สด, อุปกรณ์, อะไหล่, เสื้อผ้า ตามลำดับ จึง เสนอแนะให้เตรียมอุปกรณ์ที่เหมาะสมกับสินค้านั้นๆในปริมาณที่เหมาะสมกับทีวีเคราะห์
7. พื้นที่ที่มียอดการส่งมากได้แก่ Upper Northern Zone (ภาคเหนือ) กับ Central Zone (ภาคกลาง)
8. เสนอแนะให้มีการคำนวณระยะทางที่ส่งในแต่ละสาขา เพื่อลดค่าใช้จ่ายทางการขนส่ง เช่น น้ำมัน

รายการอ้างอิง

1. Postal codes in Thailand. (2024). [ออนไลน์]. ได้จาก:

https://en.wikipedia.org/wiki/Postal_codes_in_Thailand [สืบค้นเมื่อ 22 สิงหาคม 2567].

2. Python RegEx. (2024). [ออนไลน์]. ได้จาก

https://www.w3schools.com/python/python_regex.asp [สืบค้นเมื่อ 22 สิงหาคม 2567]