DATASCI200 - Project 2

December 2024, **Ankita Mathur**

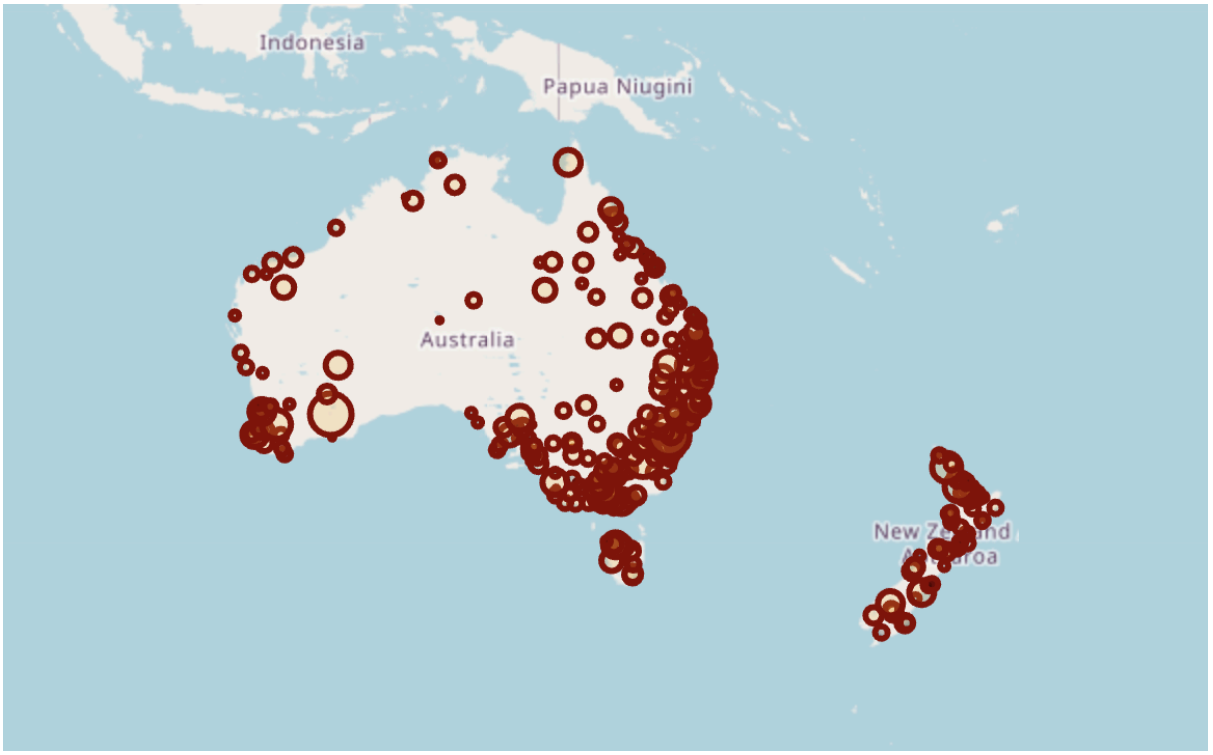
Dataset:

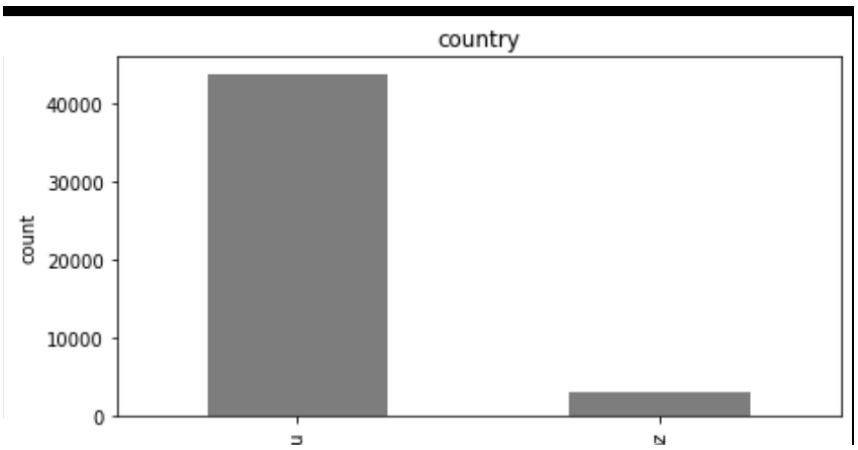
The dataset is a leading Fashion retailer in Australia. MiQ runs a display advertising campaign for this brand, where it shows ads to users leading them to make a purchase on the brand's website. The given dataset is the Sales data for all users who made a purchase online in the first half of October '17.

Dataset structure:
The dataset contains ~50k entries with 15 columns, describing: -

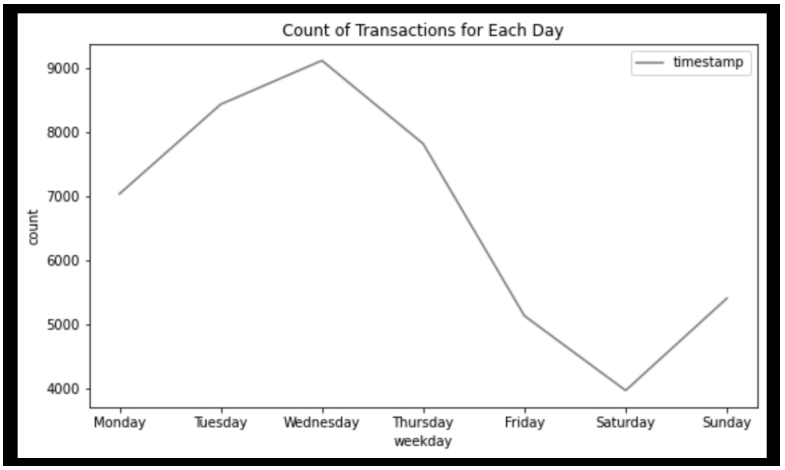
- Shopping details (timestamp, ip_address, country etc.)
- User details (user_id , user_gender, country_province, user_birthday etc.)
- Product details (product _id, product_name, number_of_products, order_coupon_code etc.)
- revenue
- Purchase details (payment_type, is_first_order etc.)

Initial exploration: - Total # of user by regions: we can observe significant higher # of customers mainly in Australia.





Line Chart-Total # of transactions by weekday



Total # of transactions by weekday: Higher number of transactions in the weekdays such as Tuesday and Wednesday exceeds the weekend transactions.

Questions to ponder over

What we plan to cover in the final report the total # of users over the regions.

1. An overview of the brand's sales by the following attributes:
    a. Overall - Total Sales and Revenue in the given time period
    b. Basket - Avg. unique quantity, revenue per order.
    c. Attributes - Time of Day, Day of Week, Geography, Payment Type
    d. Frequency - How many are single/multiple purchasers? What is the frequency of multi-purchase? Any typical attributes?

2. Product Affinity – Which products are more likely to sell together?

3. Finally, I'll arrive at a statistical segmentation of the brand's audience based on Revenue.